

# Cluster #0 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 0, from importance channel 0 (negative), represents a motif consisting of 2.1 ( $\pm 0.7$ ) nodes. The concept is generally associated with an impact of -3.0 ( $\pm 1.3$ ) on the prediction outcome.

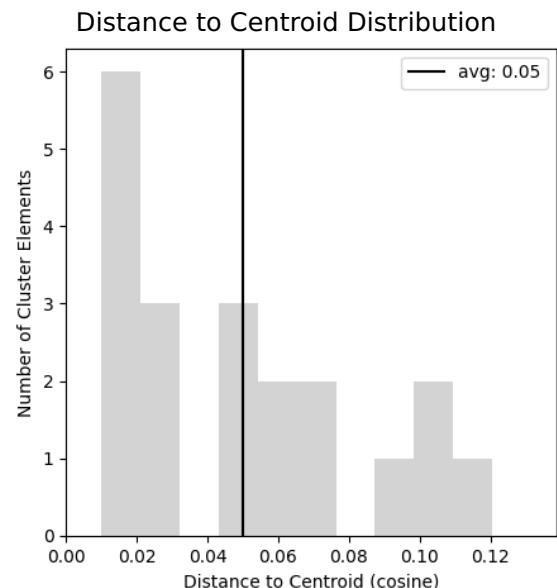
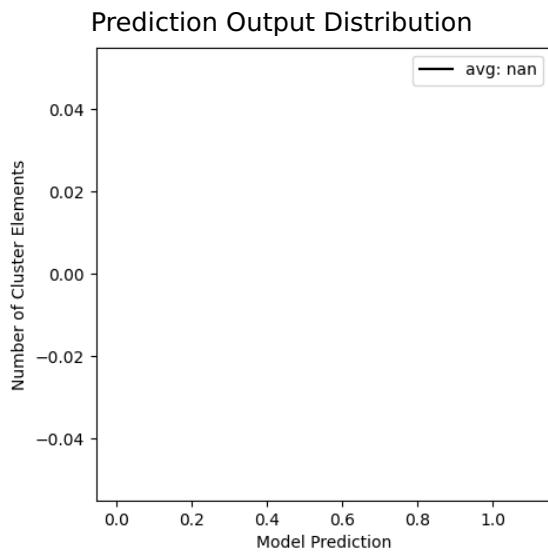
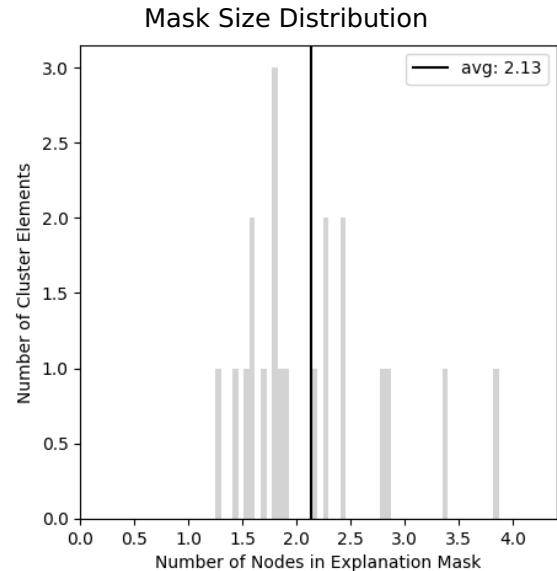
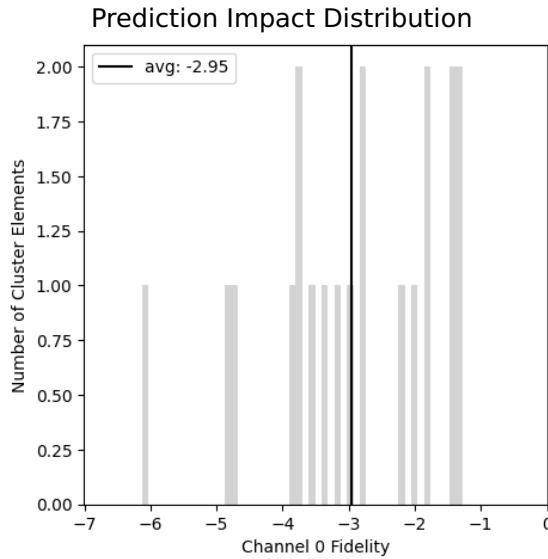
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	20
Channel Index	0.0 (0.0)

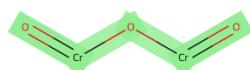
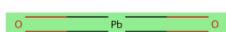
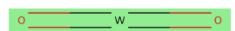
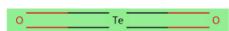
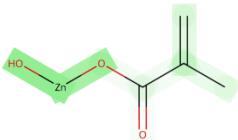
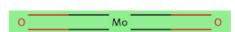
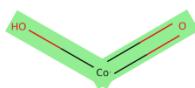
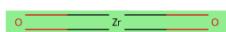
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

① This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES representation "B#N" denotes a boryl (B) connected to a nitrogen (N) atom by a triple bond, indicating the presence of a boron nitride fragment. This structure is highly polar due to the significant electronegativity difference between boron and nitrogen, which can affect water solubility. Polarity is a crucial factor influencing a molecule's ability to dissolve in water, as water is a polar solvent. However, boron nitride in bulk form is known for its low solubility in water, which suggests that the presence of the B#N fragment could increase the rigidity and planarity of the molecule, potentially hindering interactions with water molecules and thus reducing solubility.

**Hypothesis:** Molecules containing the "B#N" substructure are likely to be less soluble in water. The high polarity of the boron nitride bond is offset by the characteristics of bulk boron nitride, which bears low aqueous solubility due to its rigidity and planarity, impeding sufficient interaction with water molecules.

# Cluster #1 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 1, from importance channel 0 (negative), represents a motif consisting of 2.4 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of -1.5 ( $\pm 0.3$ ) on the prediction outcome.

## Properties

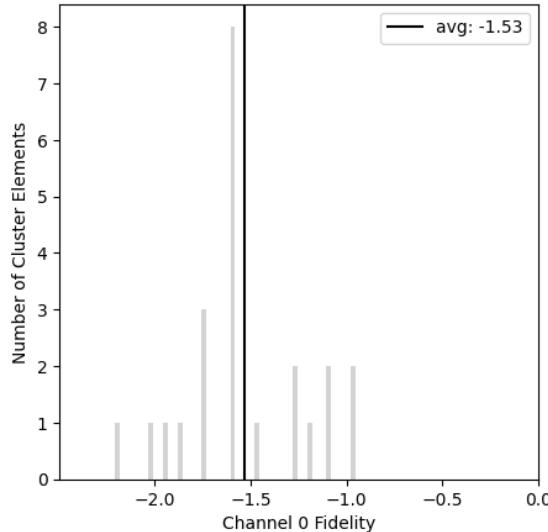
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	23
Channel Index	0.0 (0.0)

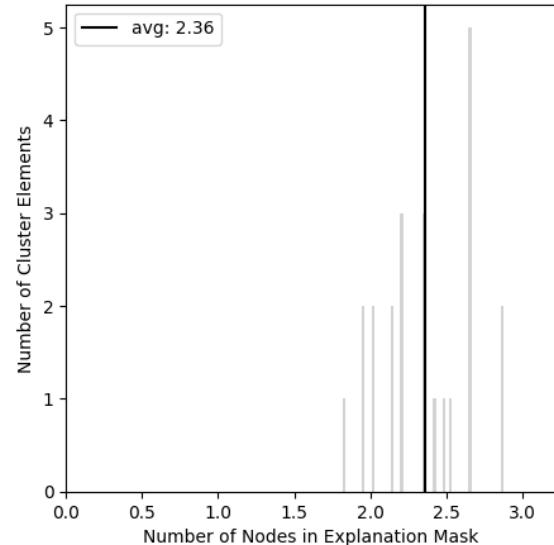
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

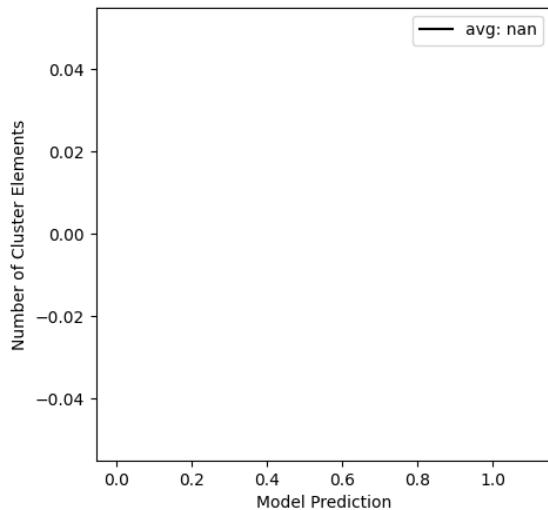
Prediction Impact Distribution



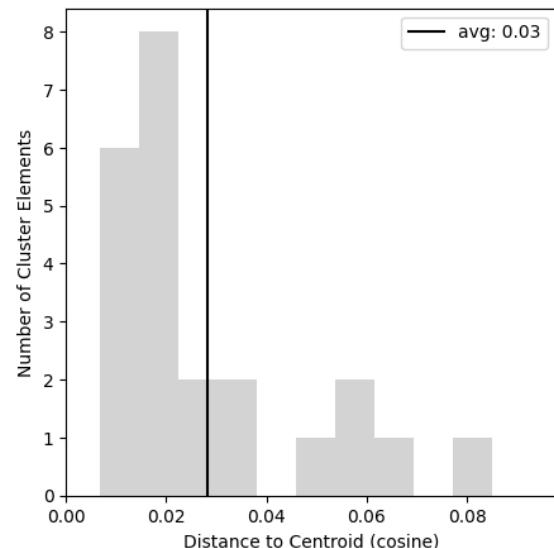
Mask Size Distribution



Prediction Output Distribution

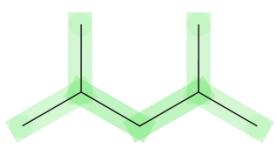
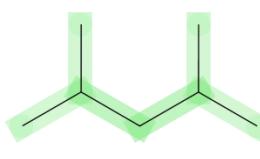
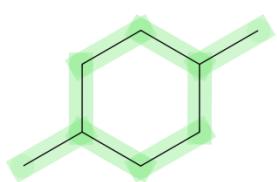
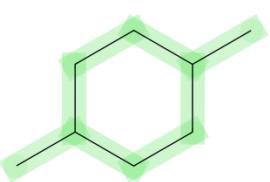
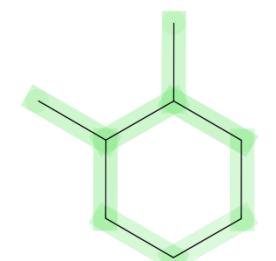
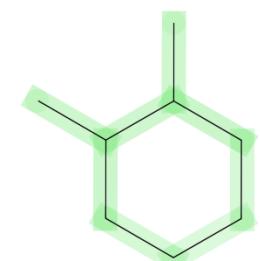
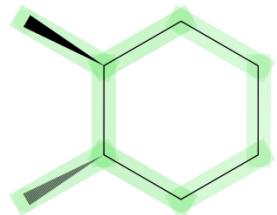
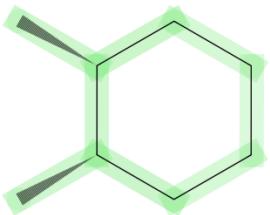
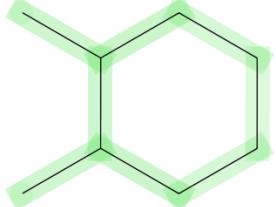
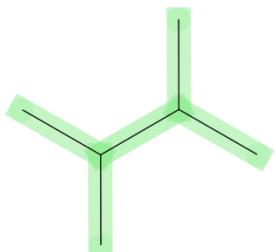
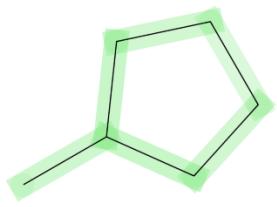
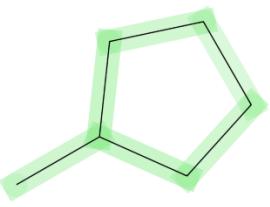
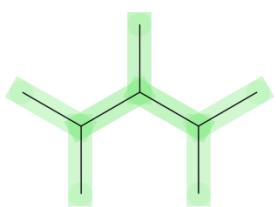
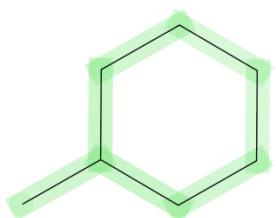
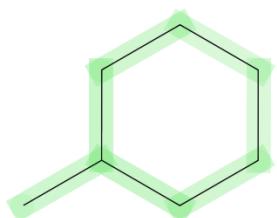
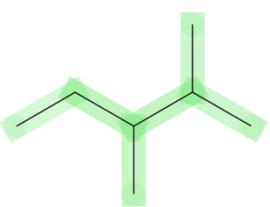


Distance to Centroid Distribution



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The linear alkane fragment represented by "C-C-C" indicates a hydrocarbon chain, which lacks any functional groups that could engage in hydrogen bonding with water molecules. The absence of polar or hydrogen bonding groups in this structure makes it hydrophobic, decreasing its affinity for the polar solvent water, thus reducing solubility.

**Hypothesis:** Alkane chains decrease water solubility due to their hydrophobic nature. The non-polar "C-C-C" structure cannot interact effectively with water molecules, leading to a negative influence on solubility.

# Cluster #2 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 2, from importance channel 0 (negative), represents a motif consisting of 2.8 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of -1.2 ( $\pm 0.3$ ) on the prediction outcome.

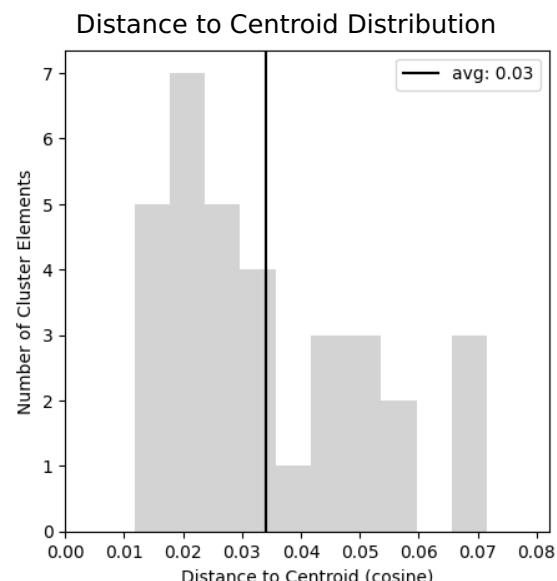
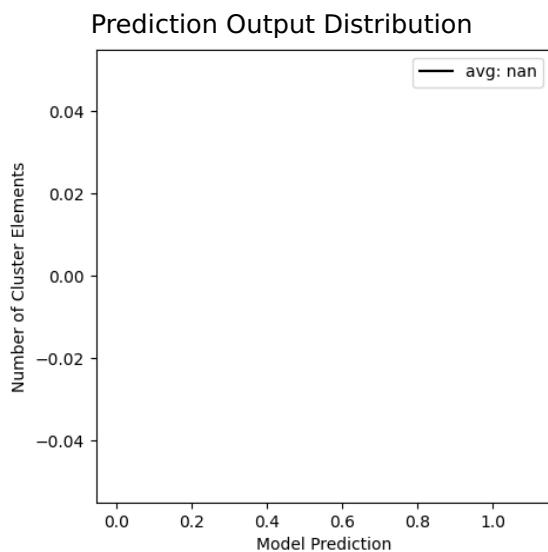
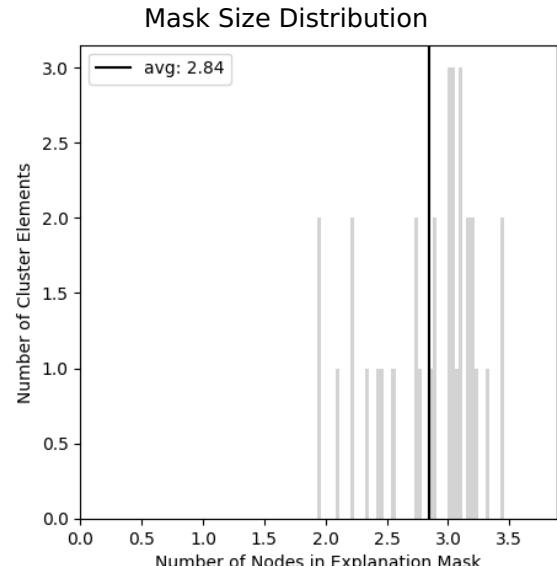
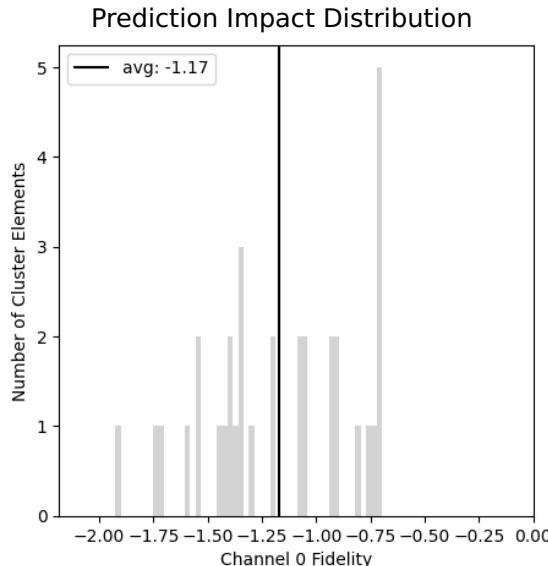
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	33
Channel Index	0.0 (0.0)

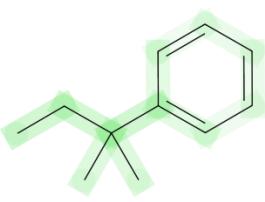
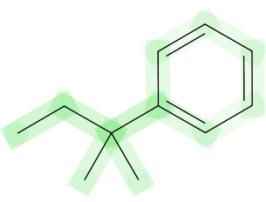
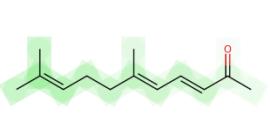
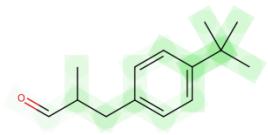
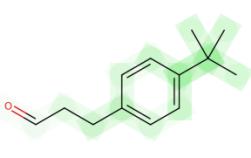
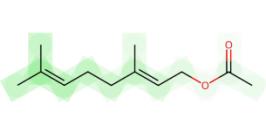
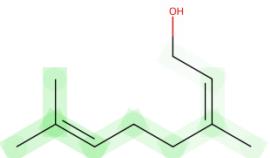
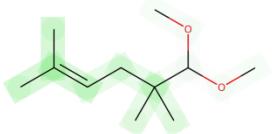
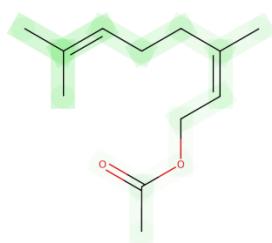
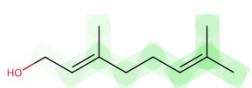
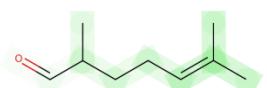
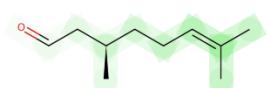
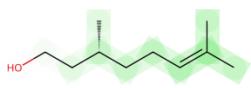
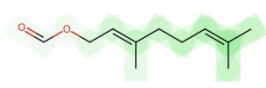
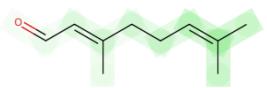
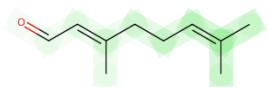
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



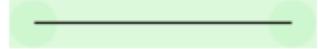
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The substructure represented by "C-C" in SMILES is indicative of a simple alkane linkage, which typically comprises of nonpolar carbon-carbon bonds. Nonpolar compounds have limited interaction with water, which is a polar solvent. The inability of these two nonpolar carbon atoms to form significant interactions, such as hydrogen bonds, with water molecules results in a decrease in solubility.

**Hypothesis:** The presence of an alkane linkage "C-C" in a molecule is associated with a decrease in water solubility. The nonpolar nature of the carbon-carbon bond does not allow for effective interactions with polar water molecules, leading to reduced solubility.

# Cluster #3 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 3, from importance channel 0 (negative), represents a motif consisting of 3.1 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of -1.8 ( $\pm 0.3$ ) on the prediction outcome.

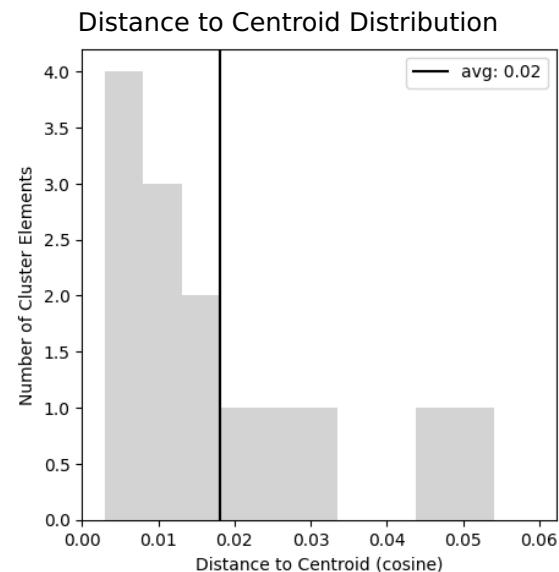
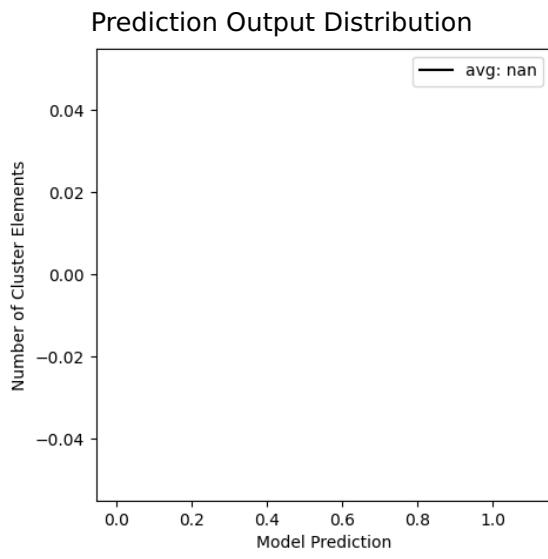
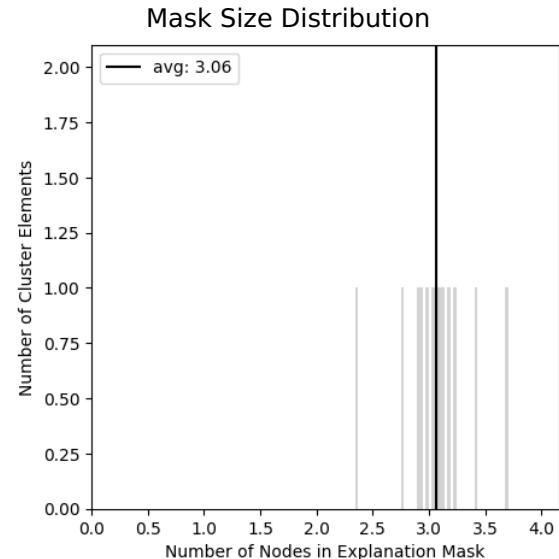
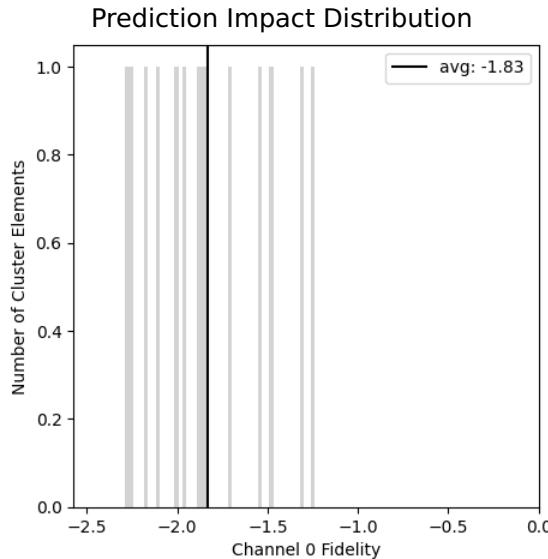
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	14
Channel Index	0.0 (0.0)

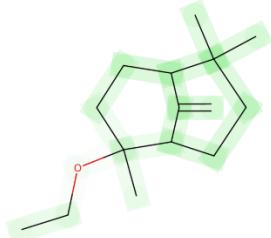
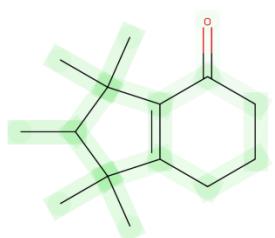
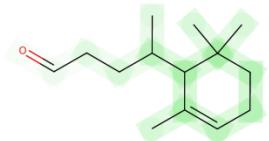
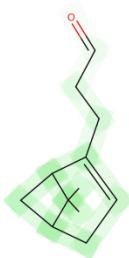
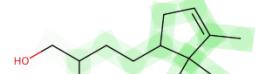
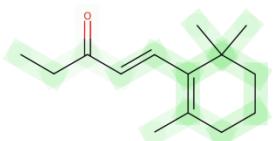
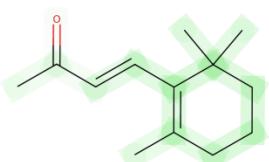
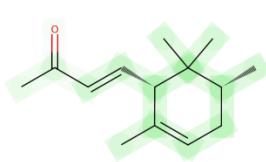
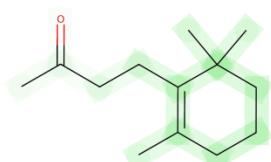
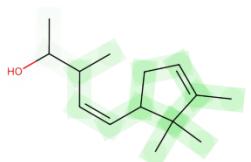
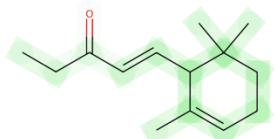
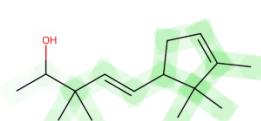
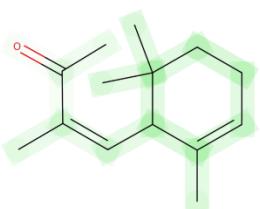
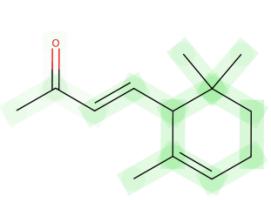
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



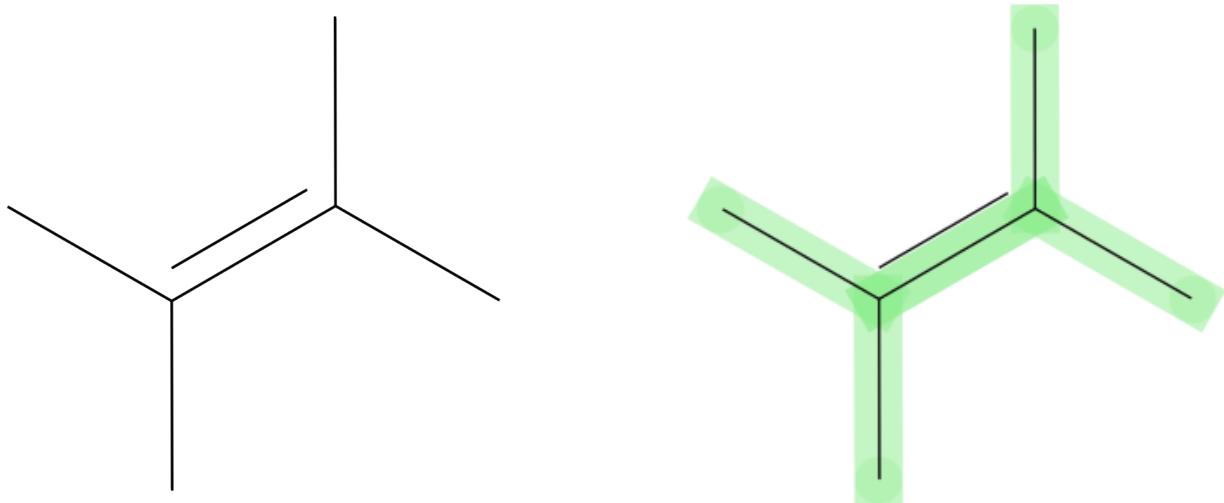
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a hydrophobic hydrocarbon chain with a degree of unsaturation, indicated by the double bond " $=C$ ". Hydrocarbons are typically nonpolar and have low affinity for polar solvents like water due to the lack of polar functional groups or charge distribution that would facilitate interaction with water molecules. The double bond introduces some degree of rigidity and planarity in the structure, potentially reducing solvation effect as it lessens the conformations available for the molecule to adapt to the polar water environment.

**Hypothesis:** The structure " $C-C(-C)=C(-C)-C$ " has a negative influence on water solubility. The hydrophobic nature of hydrocarbons and the presence of a double bond contribute to decreased solubility in water due to poor molecular interactions and reduced structural flexibility.

# Cluster #4 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 4, from importance channel 0 (negative), represents a motif consisting of 3.3 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of -2.3 ( $\pm 0.4$ ) on the prediction outcome.

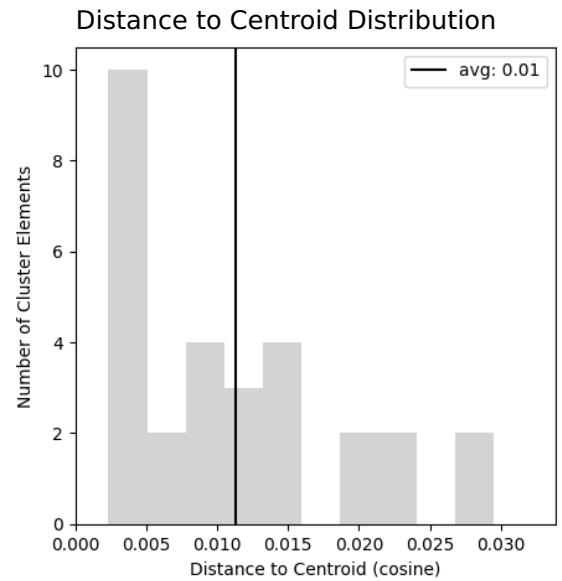
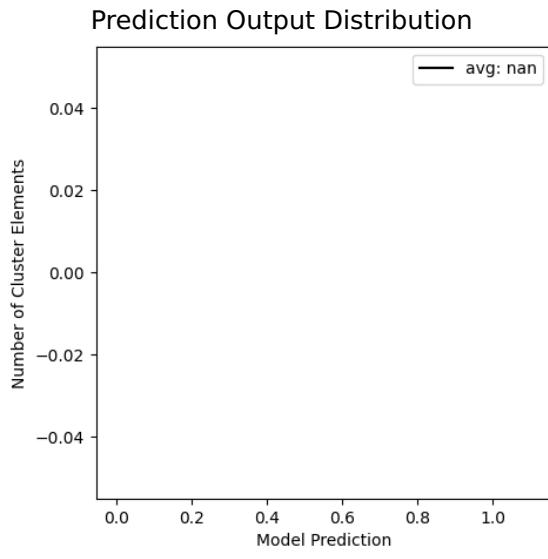
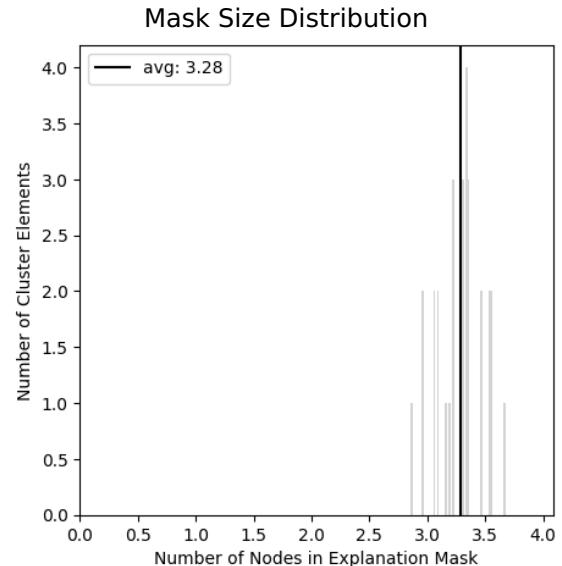
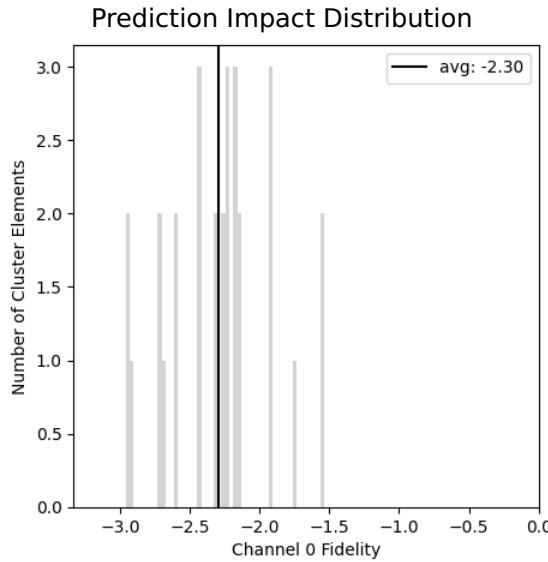
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	29
Channel Index	0.0 (0.0)

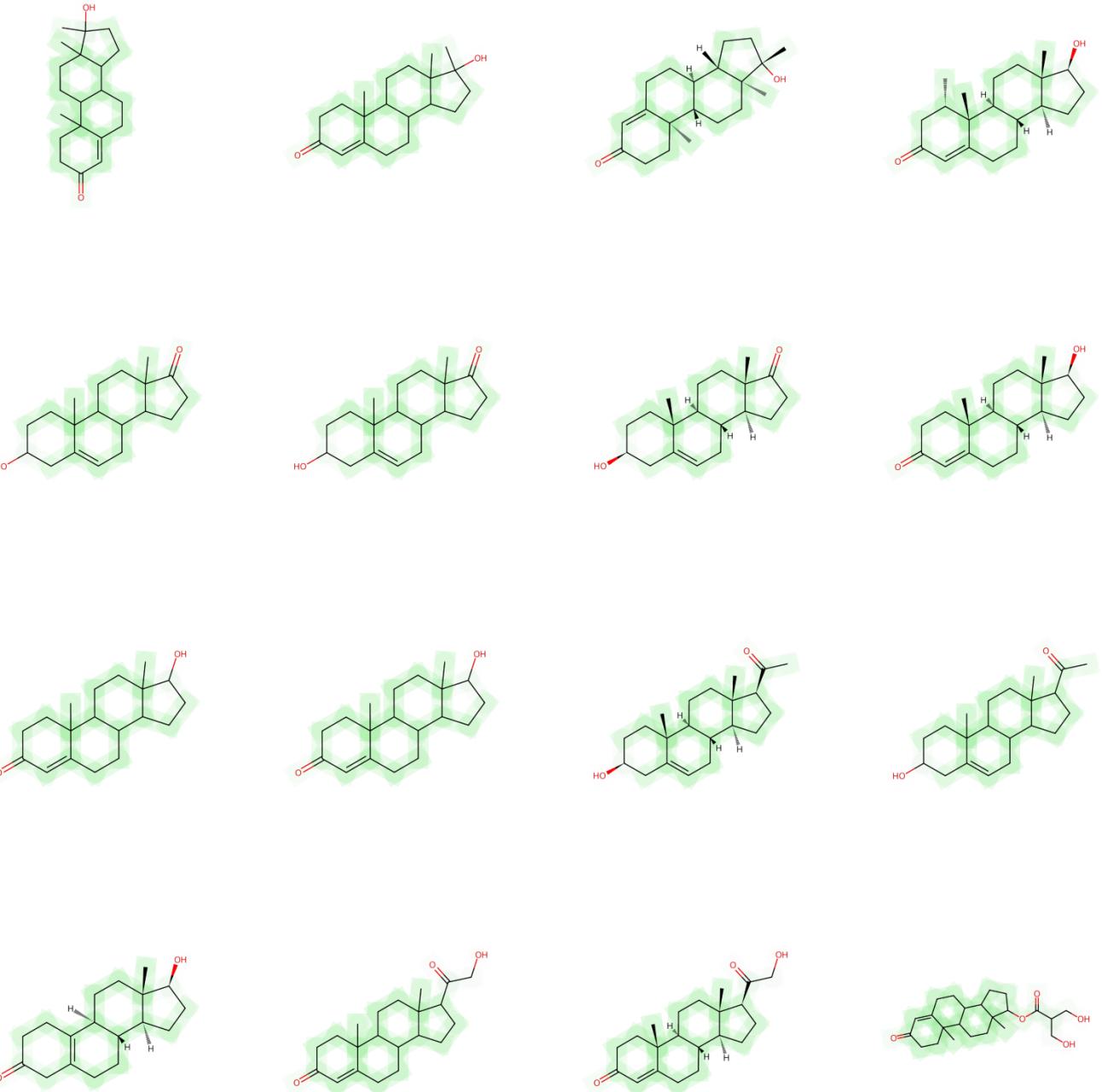
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



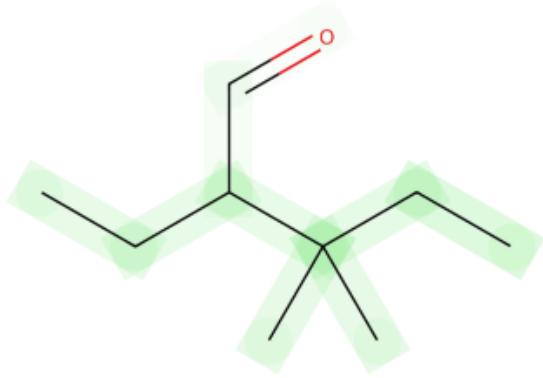
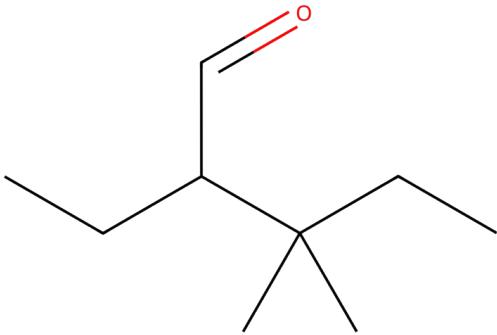
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure given resembles a hydrocarbon chain with a carbonyl group ( $C=O$ ). The hydrocarbon part of the molecule is non-polar, which usually reduces water solubility due to the inability to form hydrogen bonds with water molecules. However, the presence of a carbonyl group could introduce a site for potential hydrogen bonding due to the oxygen's electronegativity. The influence of -2.30 towards water solubility suggests that the non-polar character of the hydrocarbon chain has a more significant effect than the polar carbonyl group, likely due to the larger size of the non-polar region that impedes interaction with water.

**Hypothesis:** The presence of a long hydrocarbon chain with a single carbonyl group leads to a decrease in water solubility. The hydrocarbon part of the molecule diminishes water solubility because it cannot effectively participate in hydrogen bonding and is not energetically favorable in the polar environment of water. Contrastingly, the carbonyl group is polar, but its influence on solubility is overshadowed by the larger non-polar hydrocarbon chain.

# Cluster #5 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 5, from importance channel 0 (negative), represents a motif consisting of 2.8 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of -1.8 ( $\pm 0.2$ ) on the prediction outcome.

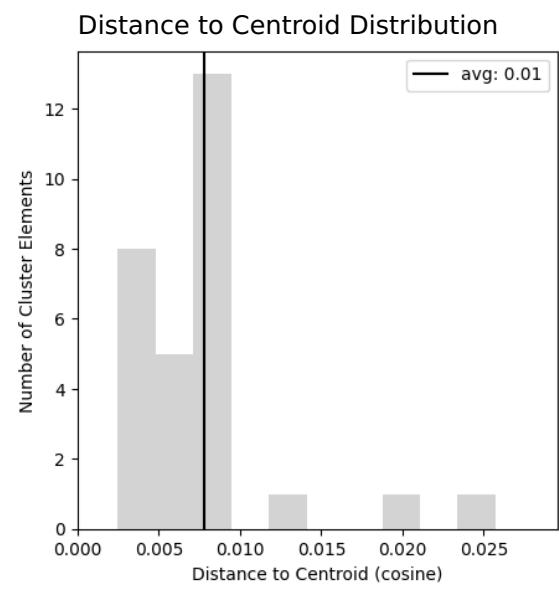
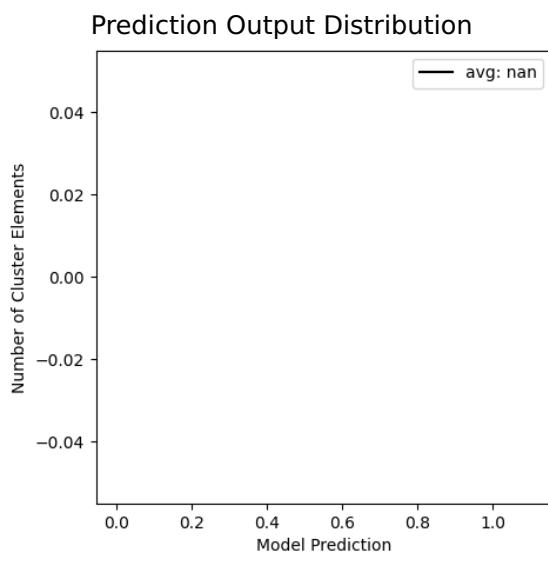
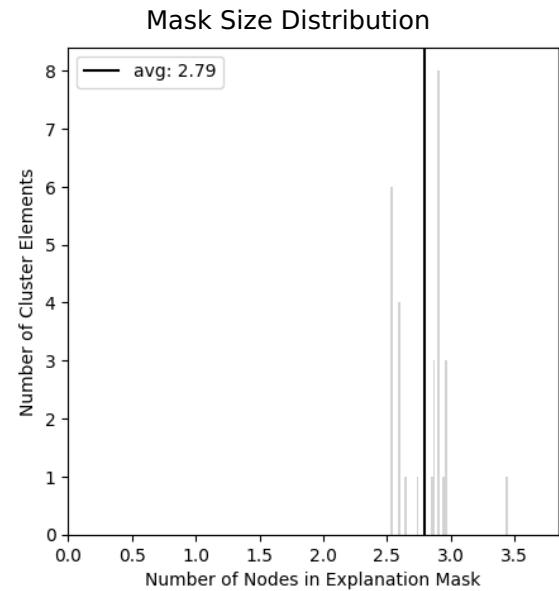
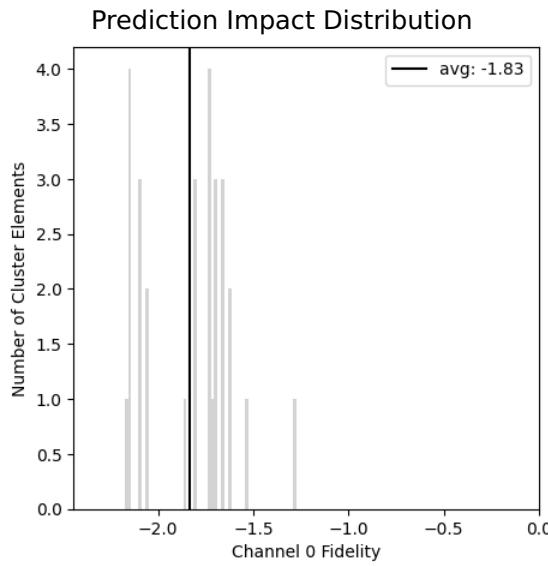
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	29
Channel Index	0.0 (0.0)

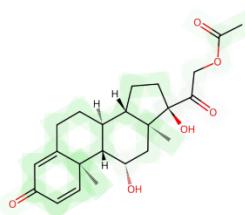
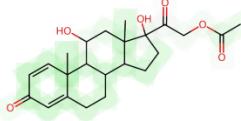
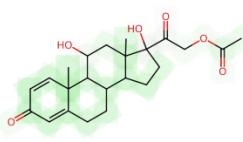
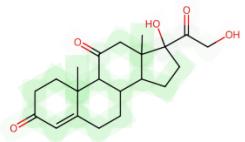
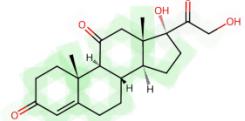
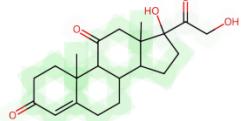
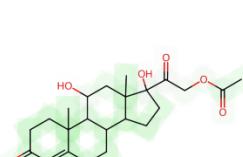
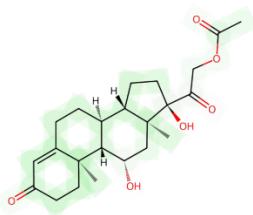
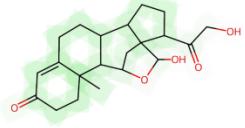
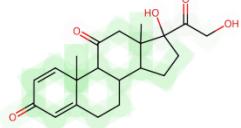
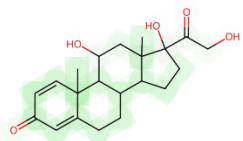
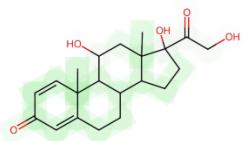
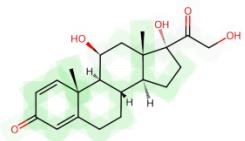
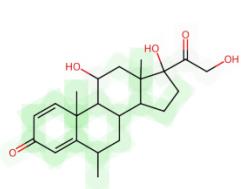
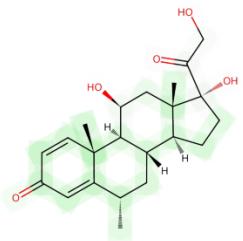
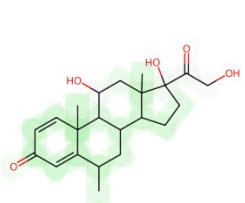
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



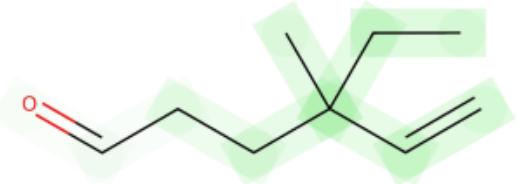
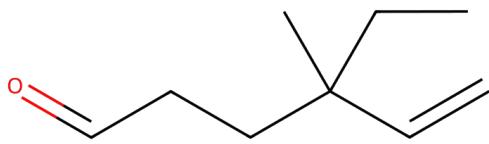
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a hydrocarbon chain with a terminal aldehyde group. The aldehyde group is polar due to the presence of a carbonyl ( $C=O$ ) bond, which can form hydrogen bonds with water molecules, increasing solubility. However, the long hydrocarbon chain is non-polar and hydrophobic, which typically reduces solubility in water. The net effect on water solubility for such a molecule would thus depend on the balance between the hydrophilic aldehyde group and the hydrophobic hydrocarbon chain. The negative influence on water solubility suggests that the hydrocarbon chain's hydrophobic effect outweighs the aldehyde group's hydrophilic effect.

**Hypothesis:** Molecules with the substructure " $C=C-C(-C)(-C-C)-C-C-C=O$ " are less soluble in water than similar molecules without this substructure. The long hydrophobic hydrocarbon chain diminishes significantly the water solubility despite the presence of a polar aldehyde group, indicating that the hydrophobic interactions are predominant in this case.

# Cluster #6 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 6, from importance channel 0 (negative), represents a motif consisting of 2.9 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of -2.5 ( $\pm 0.4$ ) on the prediction outcome.

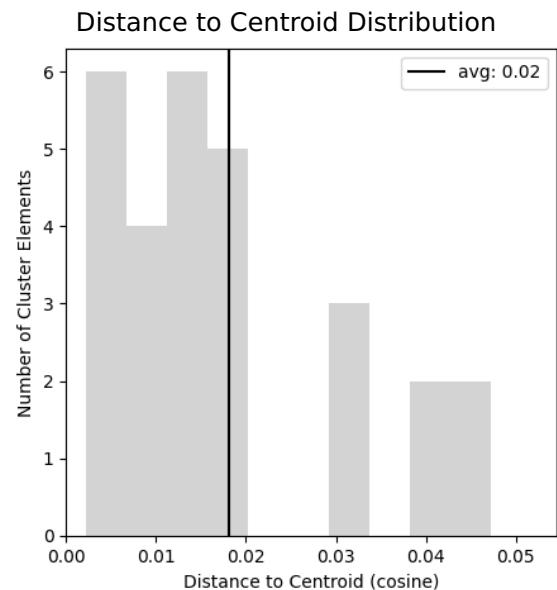
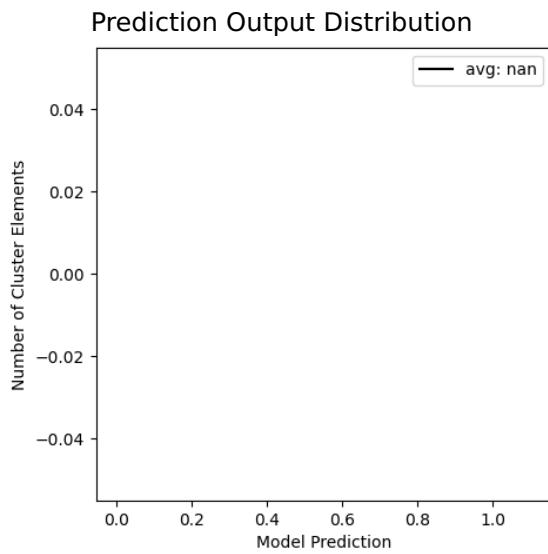
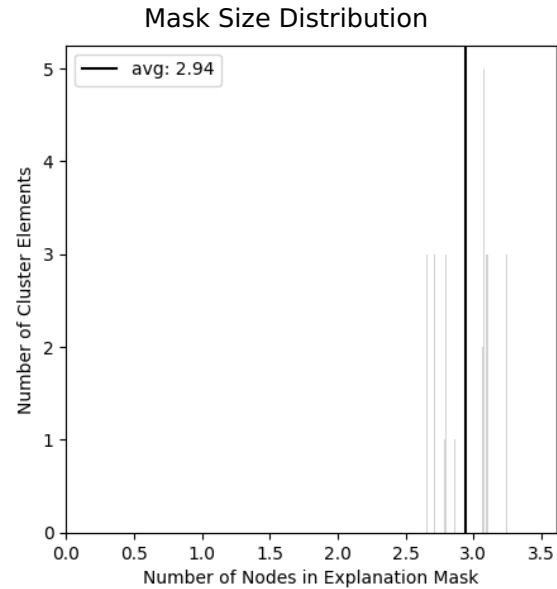
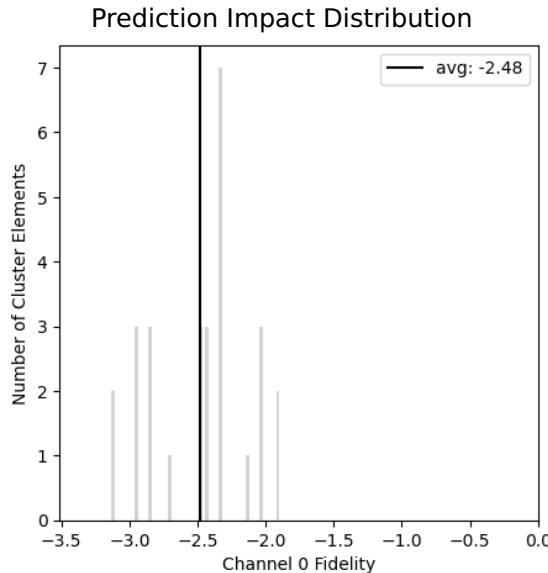
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	28
Channel Index	0.0 (0.0)

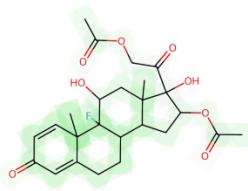
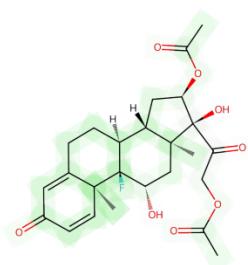
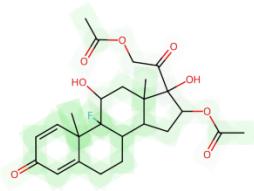
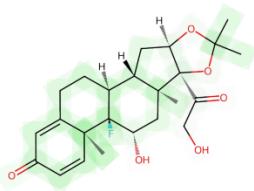
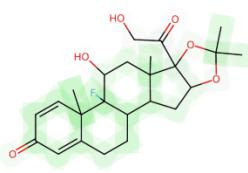
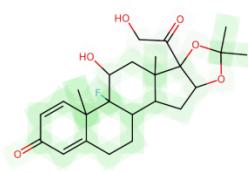
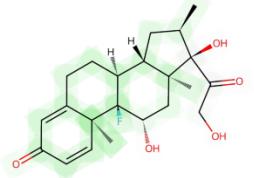
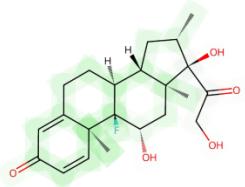
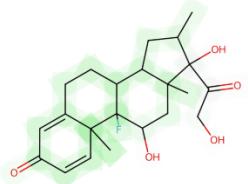
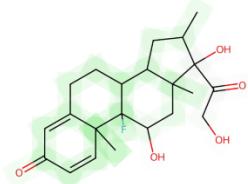
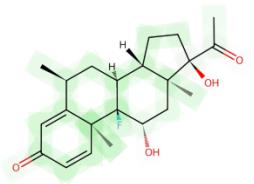
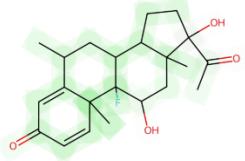
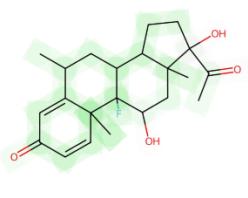
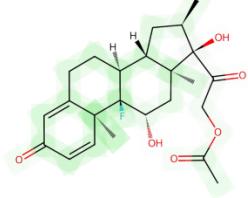
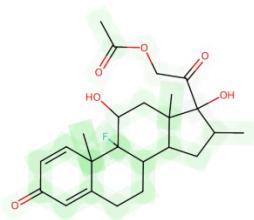
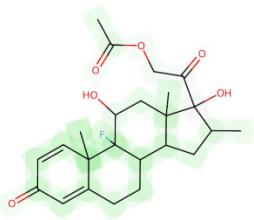
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES representation "C-C1-C-C(-C-C-O)-C-C-C-1(-C)-F" describes a molecular structure that includes a fluorine atom (F) connected to a hydrocarbon ring and an alcohol group (-C-C-O). The presence of the fluorine atom is likely responsible for decreasing the water solubility of the molecule due to its electronegative nature which leads to strong carbon-fluorine bonds that are not easily disrupted by the polar water molecules. However, the presence of an alcohol group could potentially increase solubility due to its ability to form hydrogen bonds with water. The hydrocarbon ring and chain contribute to hydrophobic character, and thus could further reduce solubility. Therefore, the overall influence on water solubility likely results from the interplay between these functional groups, with the fluorine possibly having a dominant effect.

**Hypothesis:** The presence of a fluorine atom in a hydrocarbon ring structure with an adjacent alcohol group results in decreased water solubility of the molecule. The strong carbon-fluorine bonds resist interaction with water, while the alcohol group's hydrogen bonding capacity is insufficient to counterbalance the effect of the fluorine and the hydrophobicity of the hydrocarbon chains.

# Cluster #7 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 7, from importance channel 0 (negative), represents a motif consisting of 3.2 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of -2.7 ( $\pm 0.2$ ) on the prediction outcome.

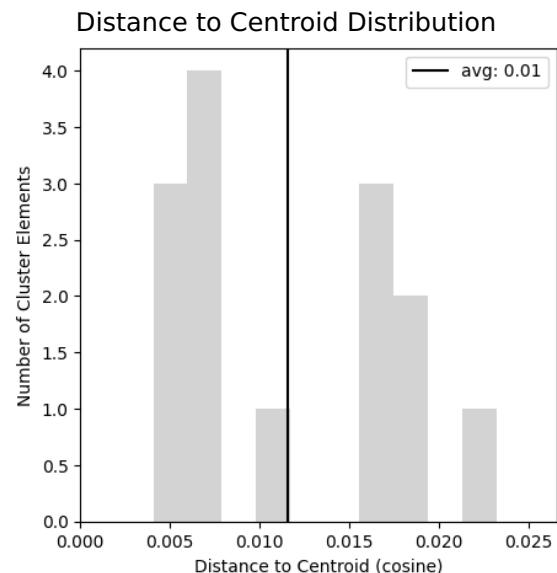
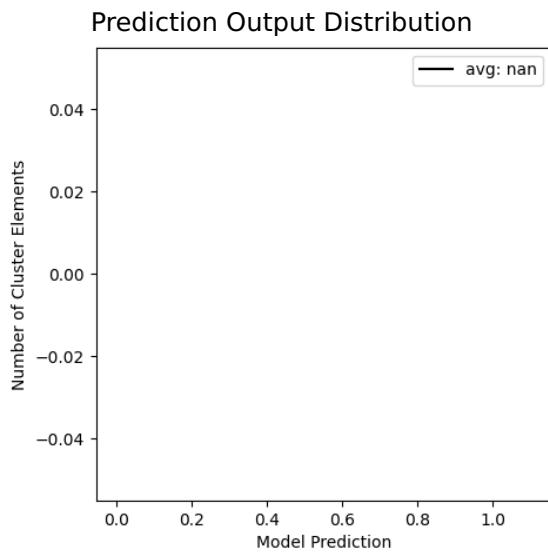
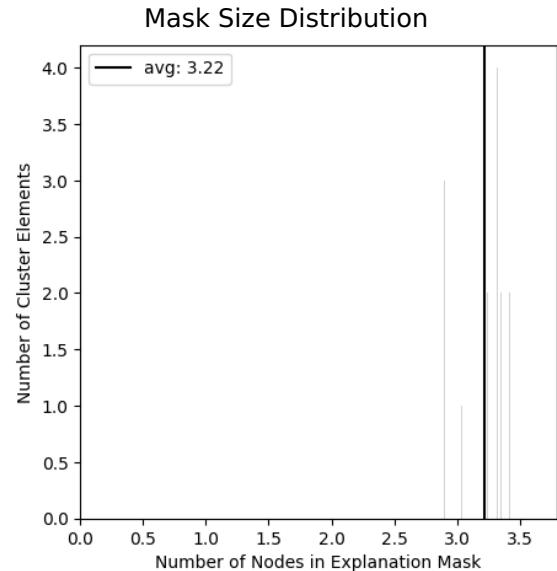
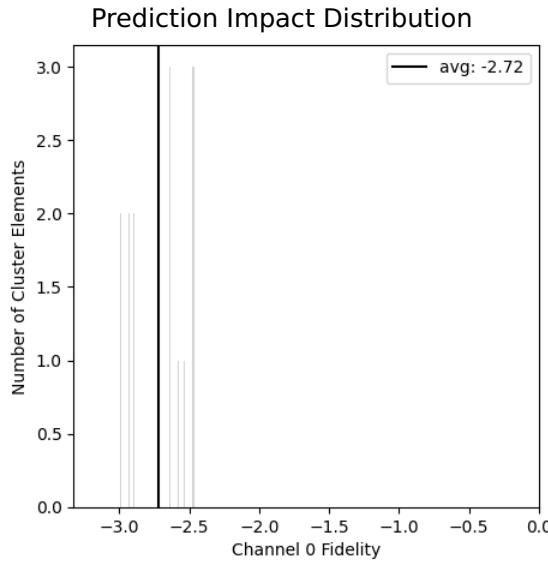
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	14
Channel Index	0.0 (0.0)

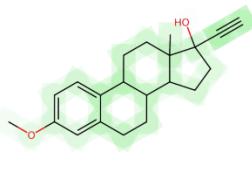
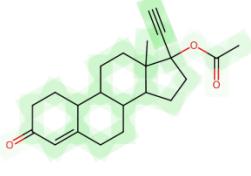
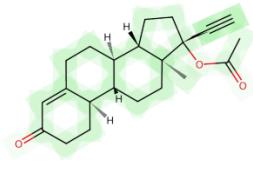
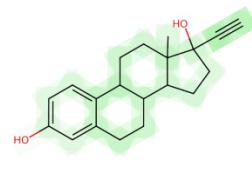
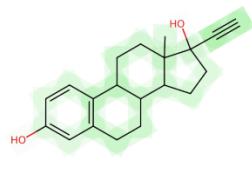
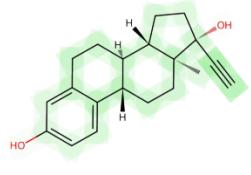
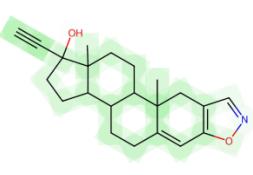
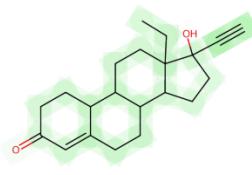
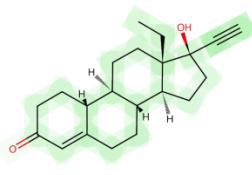
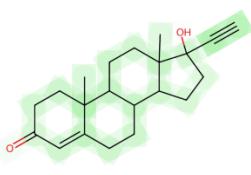
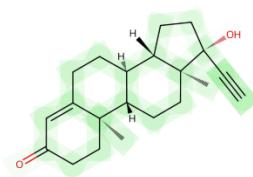
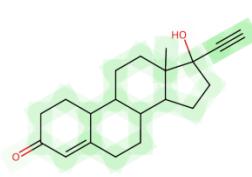
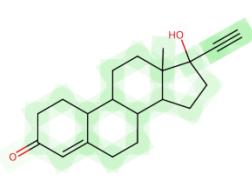
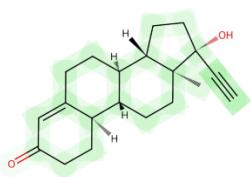
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



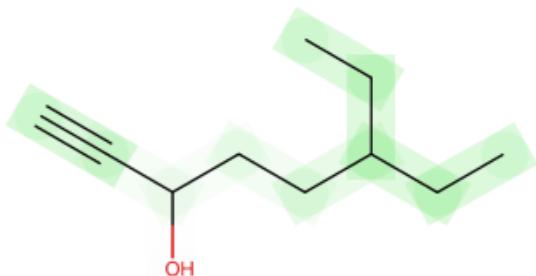
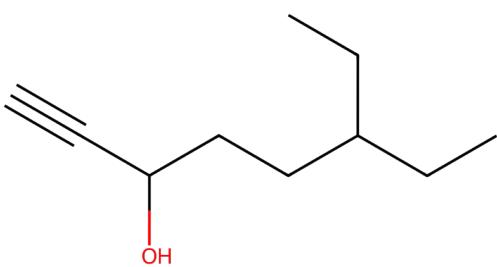
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure "C#C-C(-O)-C-C-C(-C-C)-C-C" contains a hydroxyl group (-OH) attached to a carbon that is part of a larger hydrocarbon chain. Hydroxyl groups are polar due to the electronegativity difference between oxygen and hydrogen atoms, which makes compounds containing them more likely to be soluble in water, a polar solvent. However, the overall influence on water solubility is negative, suggesting that while the hydroxyl group promotes solubility, the effect is overwhelmed by the long hydrocarbon chains present in the structure, which are nonpolar and hydrophobic. They decrease the overall solubility of the molecule in water due to their tendency to avoid interaction with polar molecules like water.

**Hypothesis:** The presence of the hydroxyl group suggests enhanced solubility in water, but is counteracted by the extensive nonpolar hydrocarbon content in the molecule leading to a net decrease in water solubility. The polar hydroxyl group allows for hydrogen bonding with water, which should increase solubility, but the long nonpolar chains inhibit this process, resulting in a compound that is less soluble in water than one might expect from the presence of a hydroxyl group alone.

# Cluster #8 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 8, from importance channel 0 (negative), represents a motif consisting of 3.1 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of -2.8 ( $\pm 0.7$ ) on the prediction outcome.

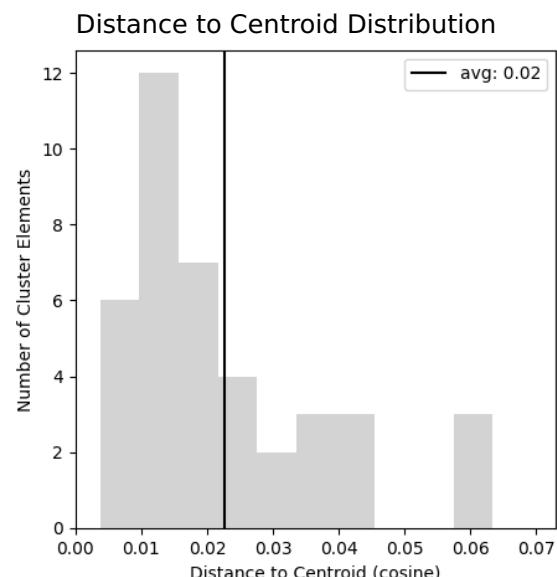
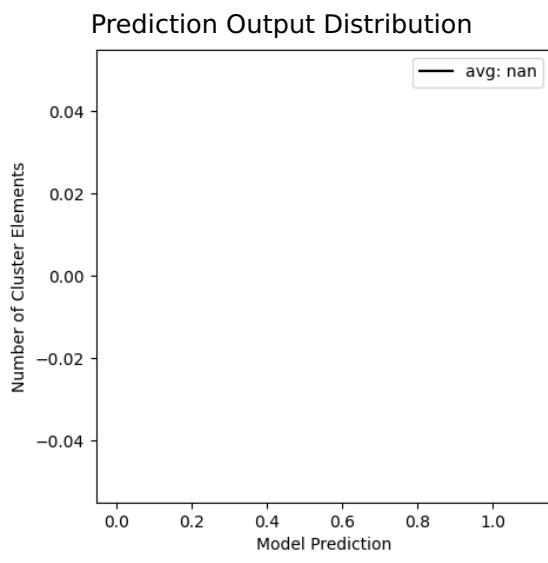
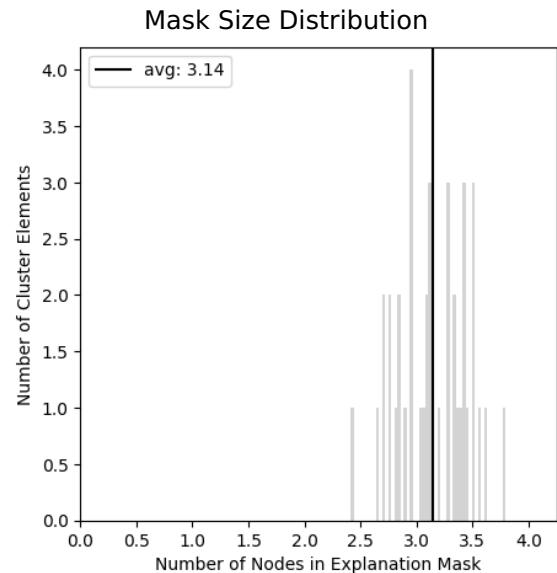
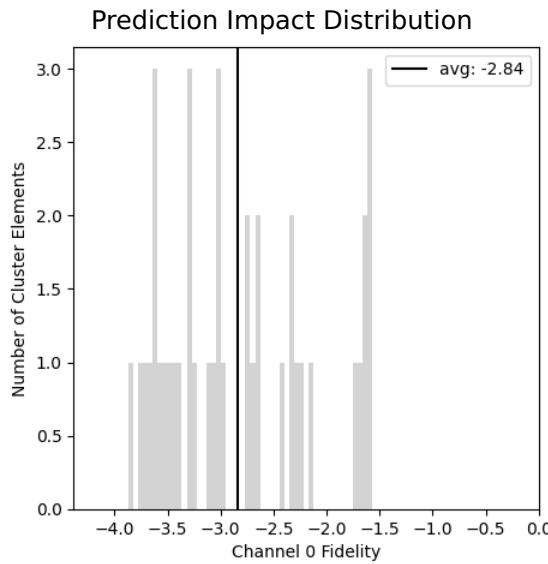
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	40
Channel Index	0.0 (0.0)

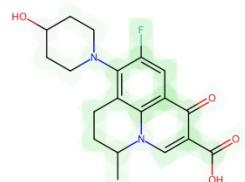
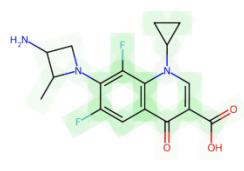
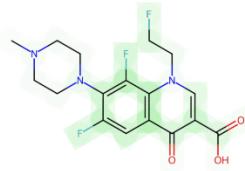
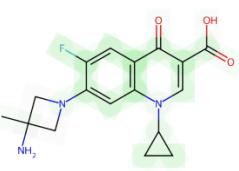
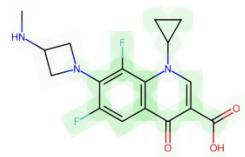
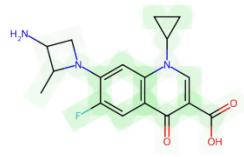
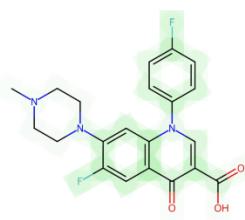
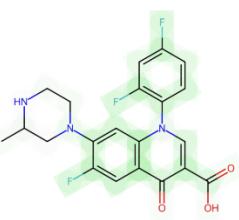
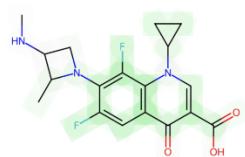
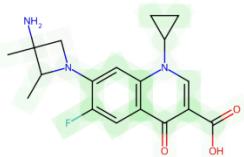
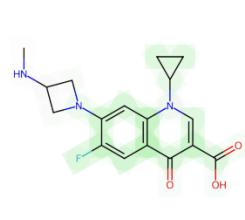
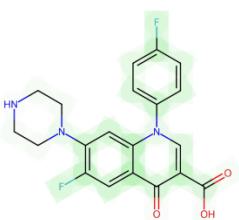
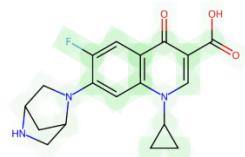
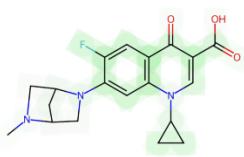
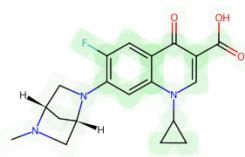
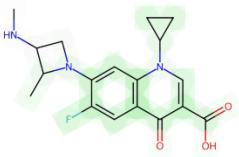
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



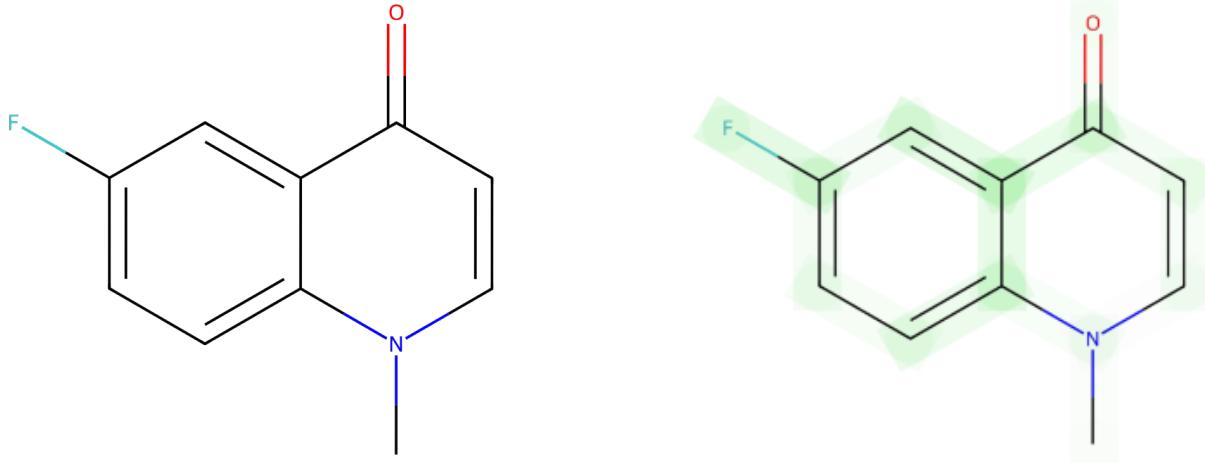
# Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure represented by the given SMILES notation corresponds to a heteroaromatic compound containing a nitrogen atom, a carbonyl group, and a fluorine substituent on a fused benzene ring system. The presence of the electronegative fluorine atom can increase the overall polarity of the molecule due to its strong inductive effect, thereby enhancing its water solubility. Additionally, the nitrogen atom in the ring can engage in hydrogen bonding with water, a key interaction that facilitates solubility. However, the overall aromaticity of the system and the carbonyl group could contribute to a planar, rigid structure that may pose some limitations to solubility due to decreased entropy upon dissolution and possible intermolecular pi-pi interactions that favor the solid state.

**Hypothesis:** The identified molecular structure, with a nitrogen-containing aromatic ring, a carbonyl group, and a fluorine atom, demonstrates a tendency towards increased water solubility. The enhanced polarity from the fluorine atom, as well as the potential for hydrogen bonding due to the nitrogen, are likely contributing factors to the solubility, while the extensive aromatic system could have a counterbalancing effect, limiting the overall solubility increase.

# Cluster #9 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 9, from importance channel 0 (negative), represents a motif consisting of 2.8 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of -3.6 ( $\pm 0.6$ ) on the prediction outcome.

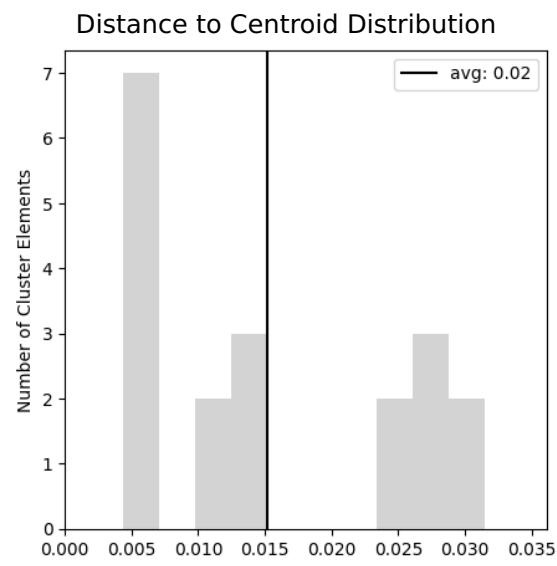
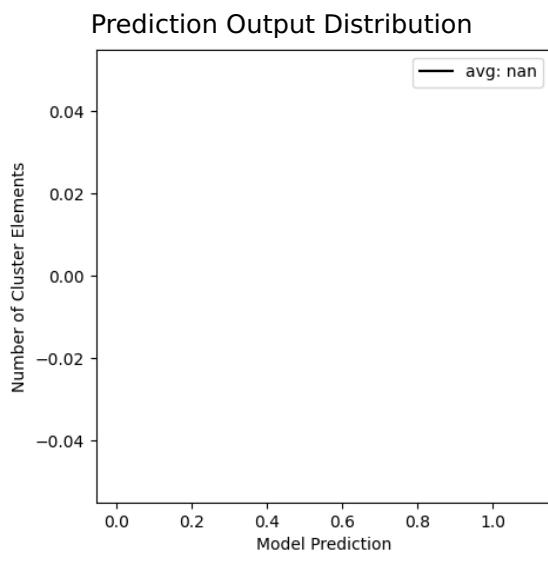
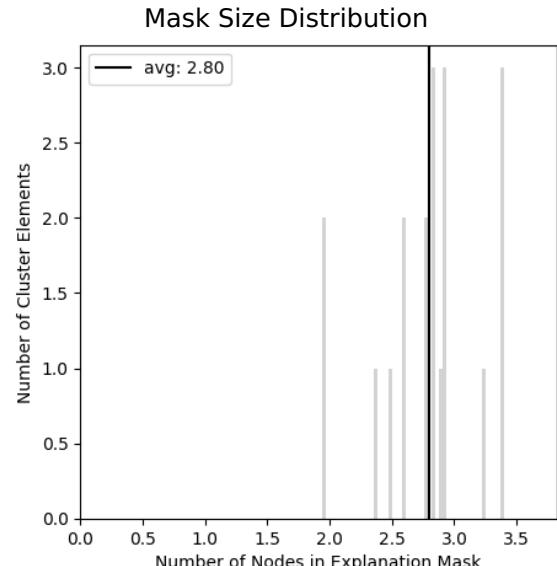
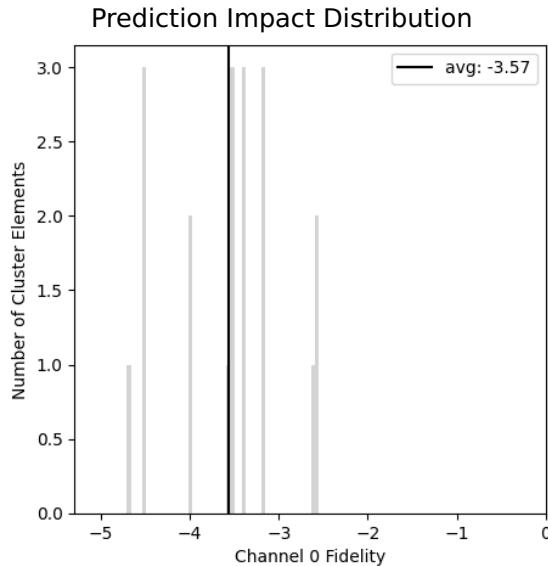
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	19
Channel Index	0.0 (0.0)

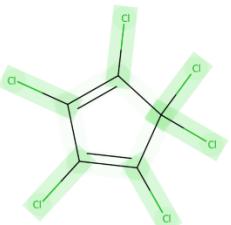
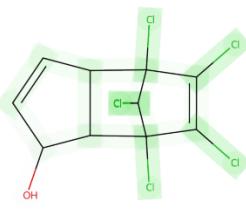
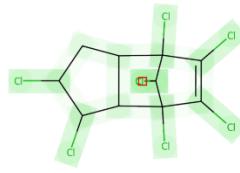
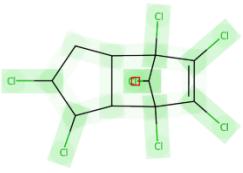
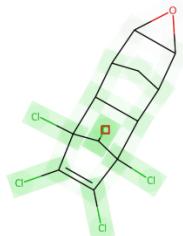
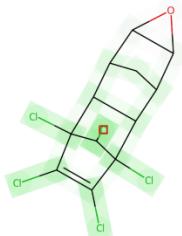
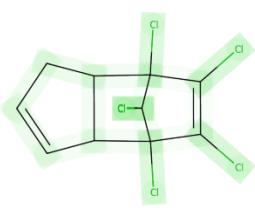
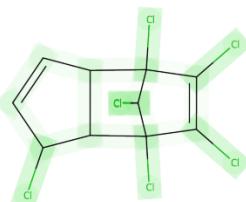
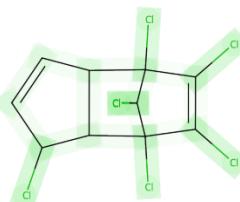
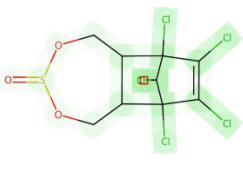
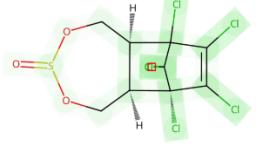
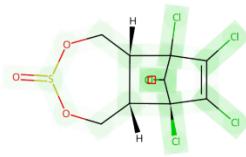
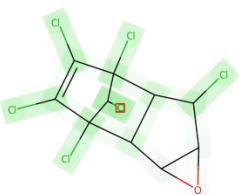
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES "Cl-C1=C-C(-Cl)=C(-Cl)-C-1-Cl" represents a molecule with a benzene ring where each carbon atom is substituted with a chlorine atom, known as hexachlorobenzene. The high degree of halogenation on the benzene ring increases hydrophobic character and steric bulk, making the molecule less likely to interact with water molecules. The electron-withdrawing nature of the chlorine atoms also decreases the polarity of the molecule, further reducing its ability to dissolve in water.

**Hypothesis:** Molecules containing highly chlorinated benzene rings demonstrate decreased water solubility. The increased hydrophobicity and reduced polarity arising from chlorine substitution on the aromatic ring are the likely reasons for this negative impact on solubility in water.

# Cluster #10 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 10, from importance channel 0 (*negative*), represents a motif consisting of 2.9 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of -1.5 ( $\pm 0.3$ ) on the prediction outcome.

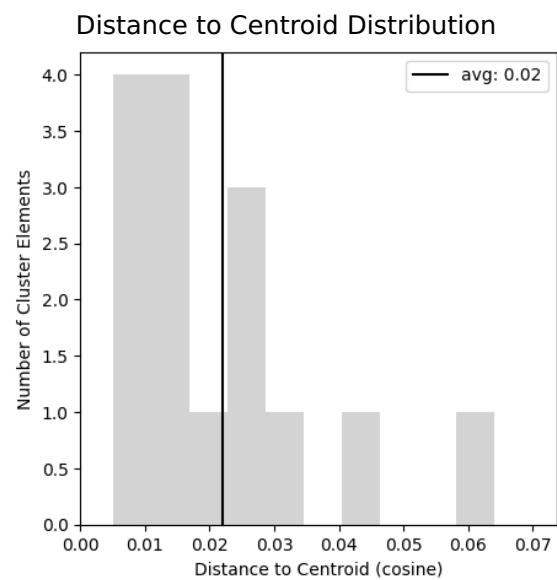
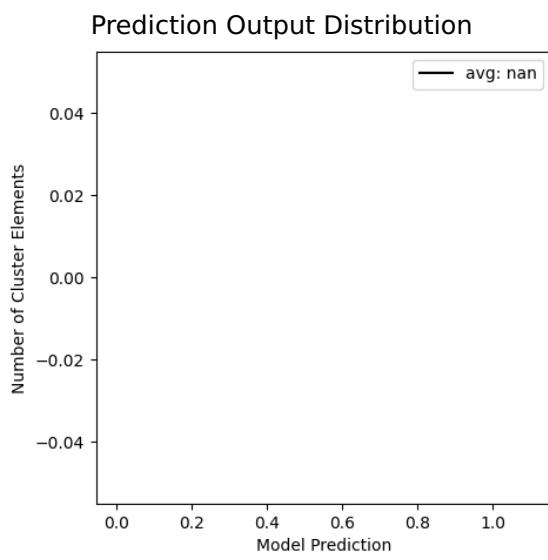
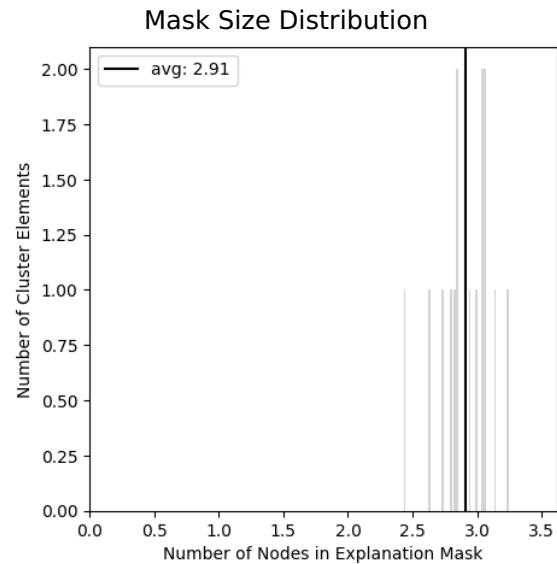
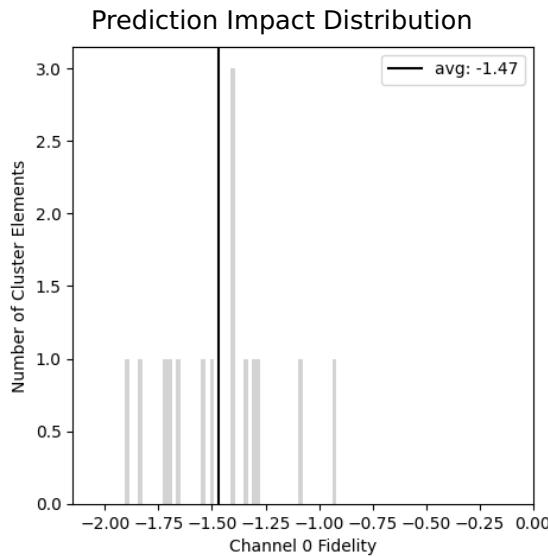
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	15
Channel Index	0.0 (0.0)

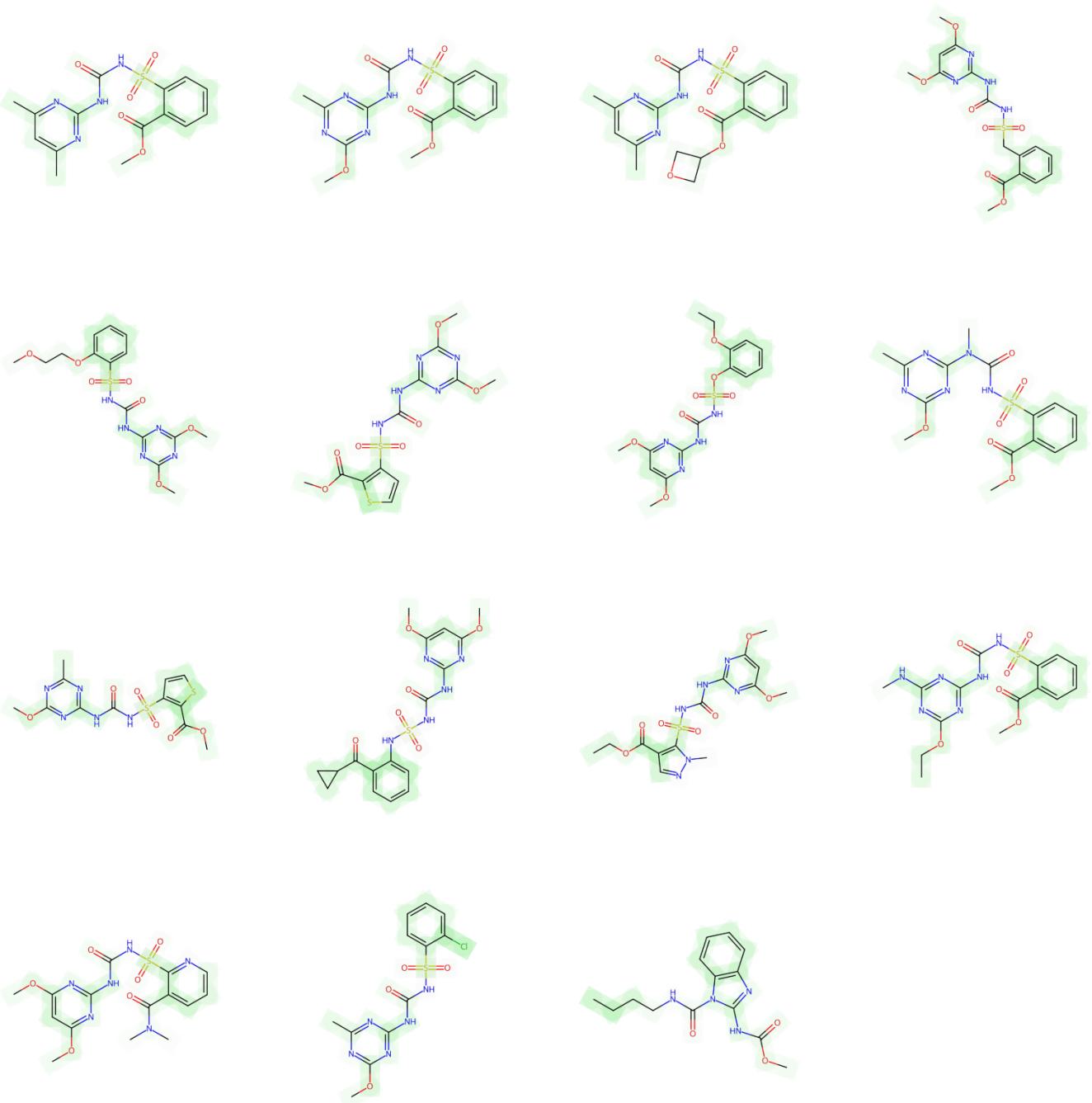
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



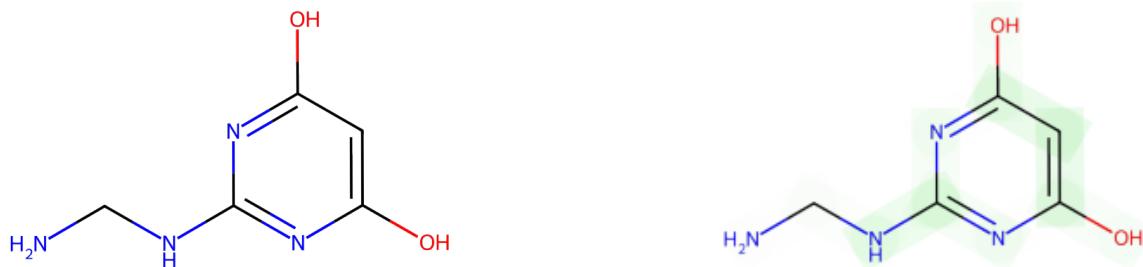
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a molecule with two amine groups ( $\text{N-C-N}$ ) connected to a pyridine-like ring with hydroxyl substituents (-O). The amine groups are generally hydrophilic due to their ability to donate a pair of electrons, forming hydrogen bonds with water molecules. The hydroxyl groups also contribute to water solubility for the same reason. However, the nitrogen in the ring (as part of the pyridine-like structure) can engage in pi-electron delocalization, which can reduce polar character and thus water solubility slightly. But overall, the presence of the amine and hydroxyl groups suggests an increase in water solubility compared to non-functionalized ring structures.

**Hypothesis:** The " $\text{N-C-N-c1:n:c(-O):c:c(-O):n:1}$ " structure tends to increase the water solubility of a molecule. Amines and hydroxyl-substituted aromatic rings can form hydrogen bonds, a key interaction for water solubility. The delocalization of electrons in the pyridine-like ring has a marginal negative effect, which is vastly outweighed by the hydrophilic amine and hydroxyl groups.

# Cluster #11 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 11, from importance channel 0 (*negative*), represents a motif consisting of 3.1 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of -1.1 ( $\pm 0.2$ ) on the prediction outcome.

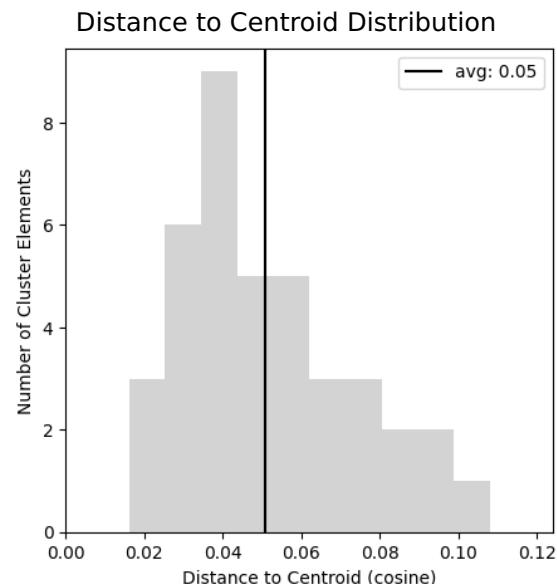
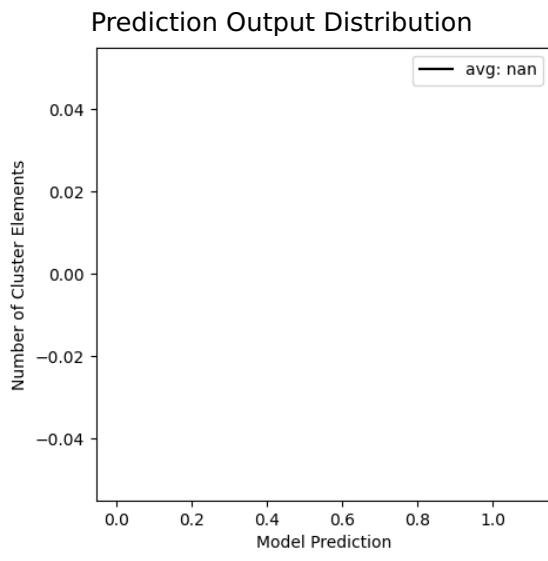
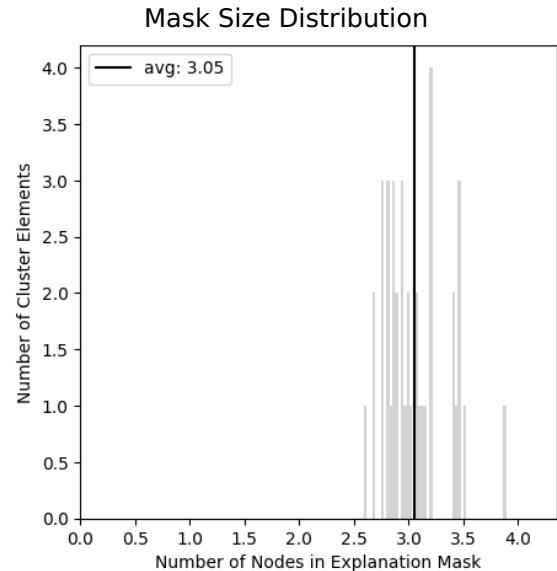
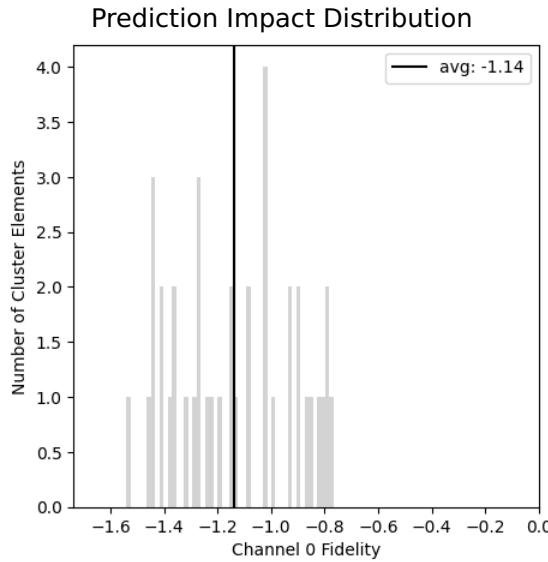
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	39
Channel Index	0.0 (0.0)

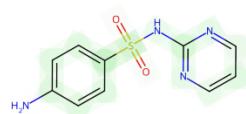
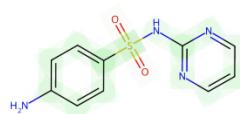
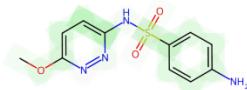
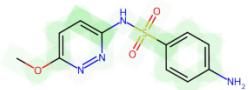
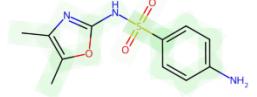
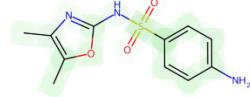
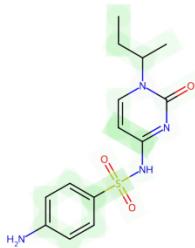
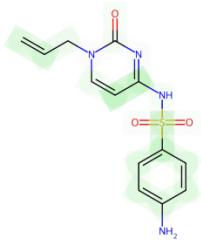
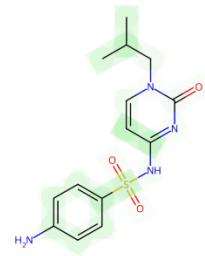
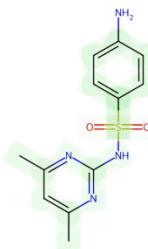
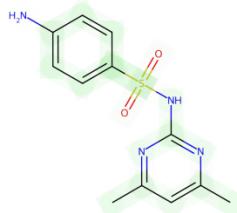
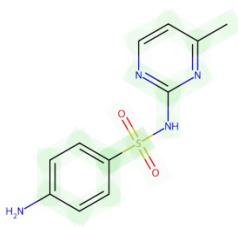
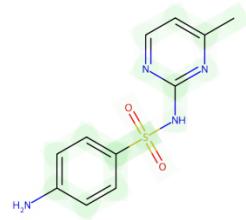
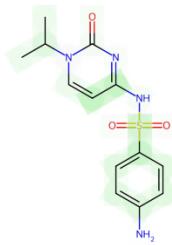
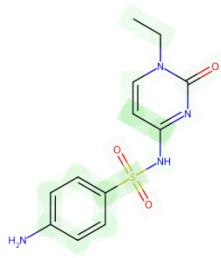
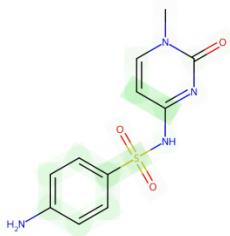
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES representation given indicates the presence of a sulfonamide group linked to a pyridone ring. Sulfonamides generally increase water solubility due to their ability to form hydrogen bonds with water molecules, as the sulfur atom is typically bonded to two oxygen atoms capable of acting as hydrogen bond acceptors. However, the negative influence on water solubility observed here could be reasoned by the intramolecular hydrogen bonding between the pyridone nitrogen and the sulfonamide hydrogen, making the compound less available for intermolecular interactions with water. Moreover, the aromatic nature of the pyridone ring also tends to decrease solubility due to its hydrophobic character and the delocalization of electrons which can lead to less favorable interactions with the polar water molecules.

**Hypothesis:** The combination of the sulfonamide group with a pyridone ring in this molecular structure leads to decreased water solubility. Intramolecular hydrogen bonding and the hydrophobicity of the pyridone ring are likely reducing the molecule's ability to interact with water, therefore decreasing its solubility despite the presence of a group that typically enhances solubility.

# Cluster #12 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 12, from importance channel 0 (*negative*), represents a motif consisting of 3.1 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of -1.2 ( $\pm 0.4$ ) on the prediction outcome.

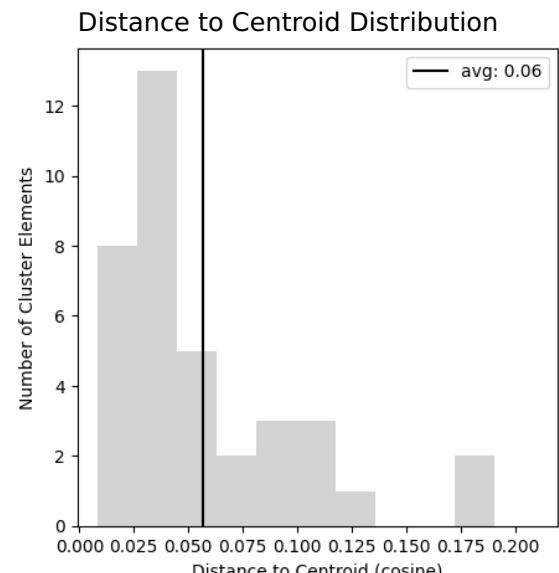
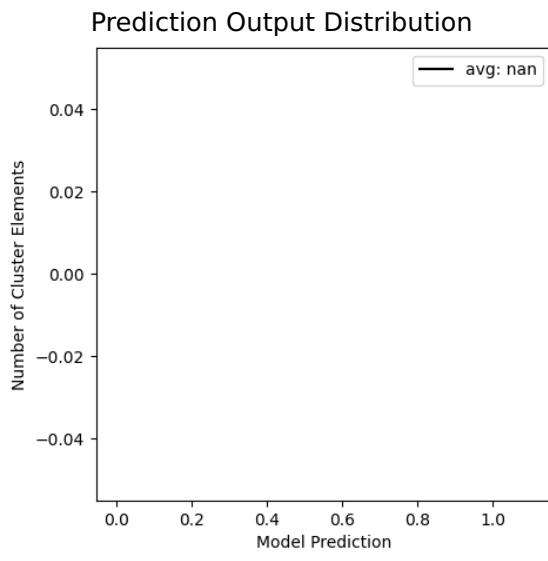
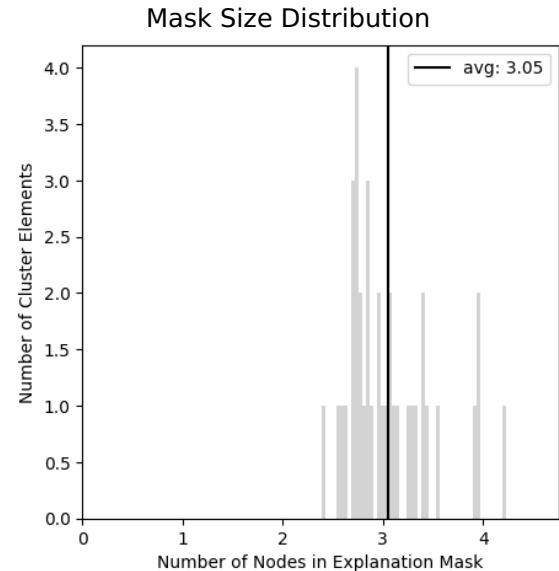
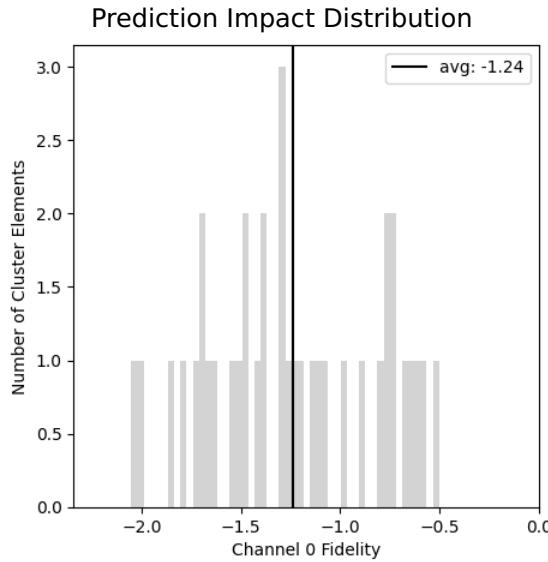
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	37
Channel Index	0.0 (0.0)

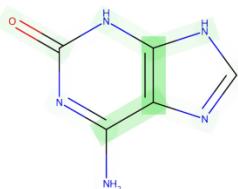
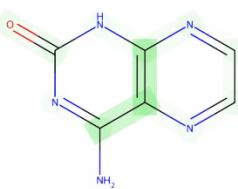
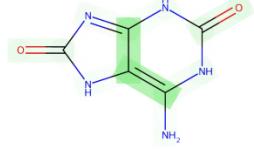
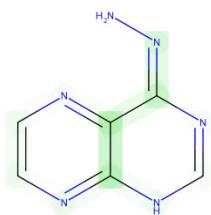
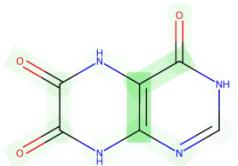
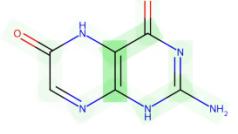
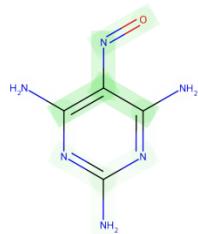
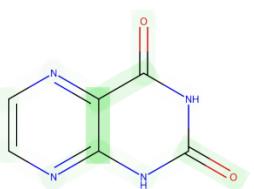
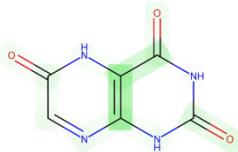
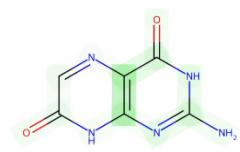
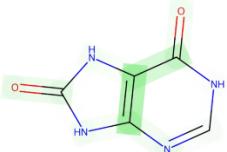
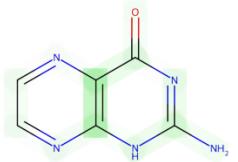
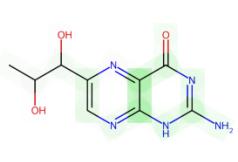
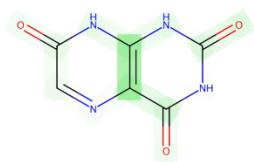
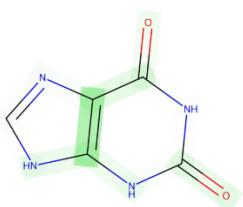
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



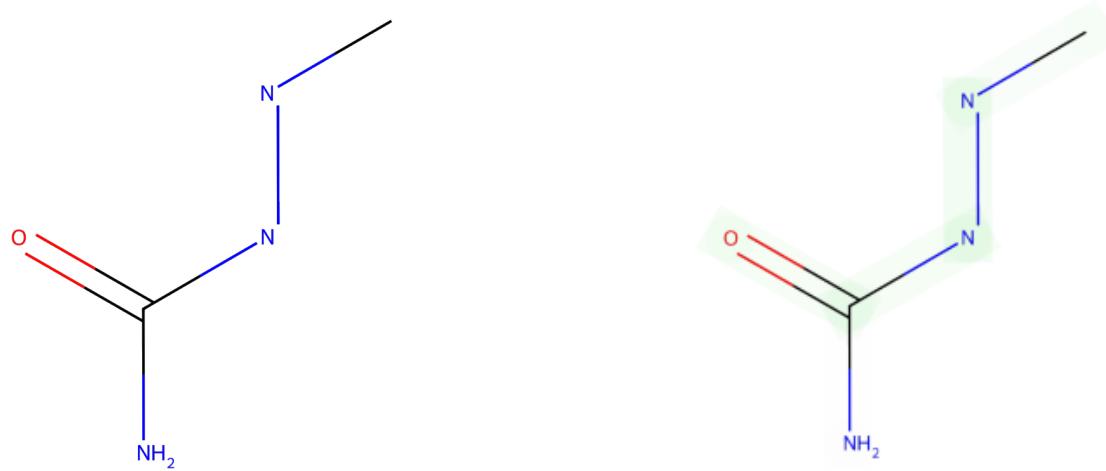
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The substructure "C-N=N-C(-N)=O," features an azo group (-N=N-) flanked by two carbon-containing groups, one of which is bonded to a nitro group (-N=O). This configuration may affect the water solubility due to potential polarity and ability to form hydrogen bonds. The azo group itself is not particularly polar and does not readily engage in hydrogen bonding. However, the presence of a nitro group could introduce some polarity to the molecule due to the electronegative oxygen atoms. These oxygen atoms can act as hydrogen bond acceptors. Despite this, the overall influence of the combination of these groups seems to exert a negative influence on water solubility, possibly because the hydrophobic character of the azo linkage outweighs the hydrophilic influence of the nitro group.

**Hypothesis:** The molecular structure "C-N=N-C(-N)=O" decreases the water solubility of the molecule. It is hypothesized that the hydrophobicity of the azo linkage reduces the water solubility more than the nitro group's polar character can increase it. Furthermore, the potential for the nitro group to engage in hydrogen bonding is insufficient to overcome the overall nonpolar nature of the molecule led by the azo group.

# Cluster #13 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 13, from importance channel 0 (*negative*), represents a motif consisting of 3.1 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of -0.8 ( $\pm 0.3$ ) on the prediction outcome.

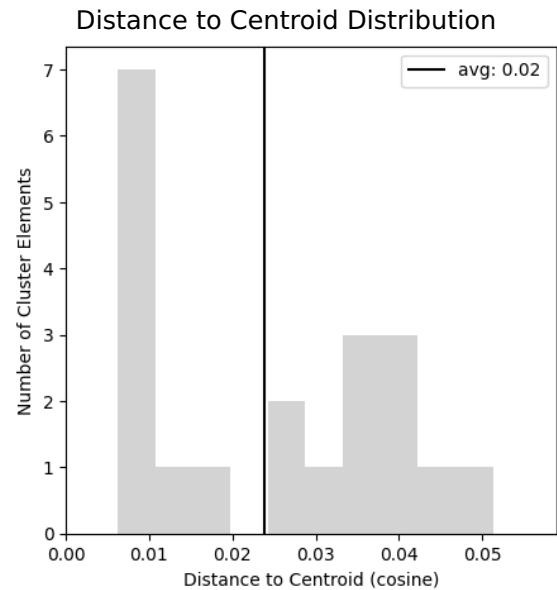
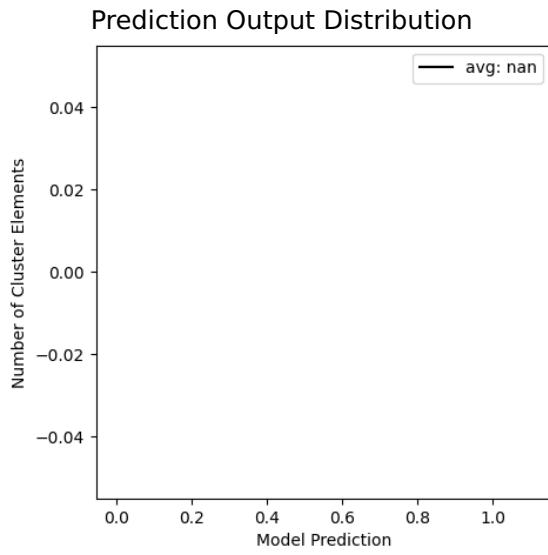
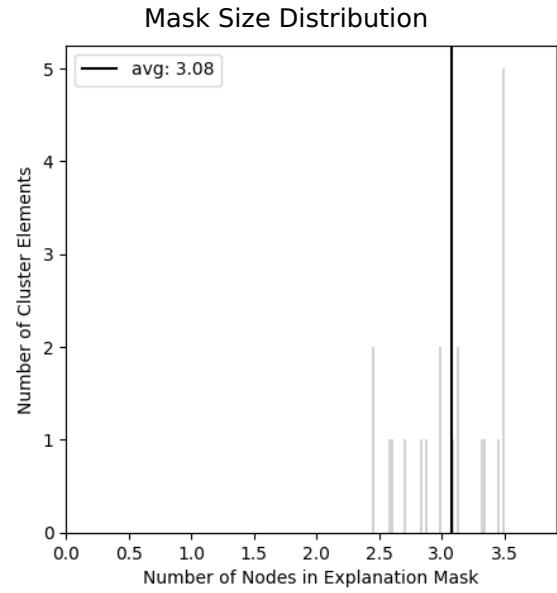
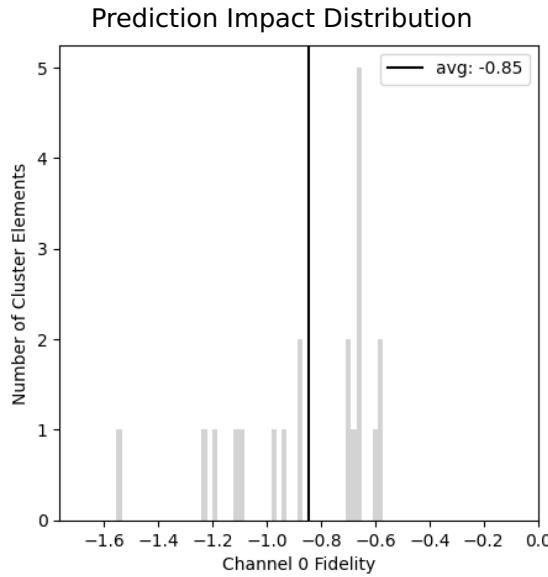
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	20
Channel Index	0.0 (0.0)

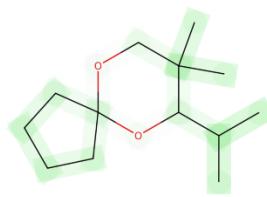
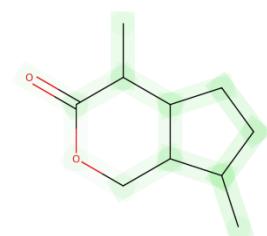
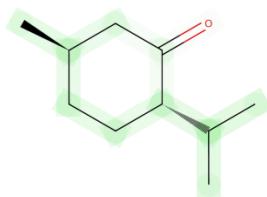
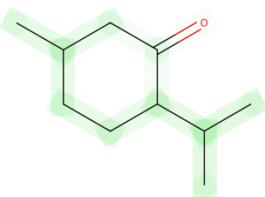
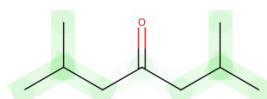
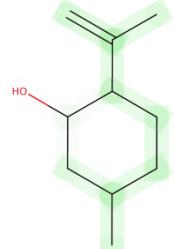
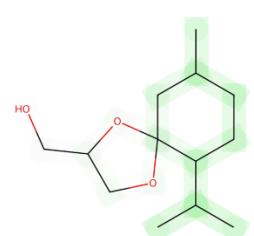
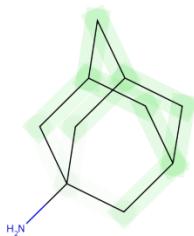
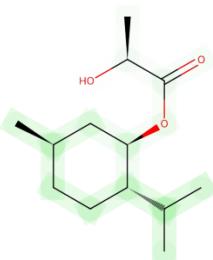
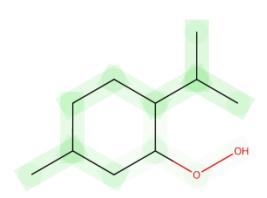
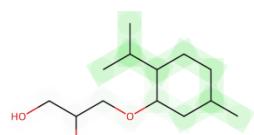
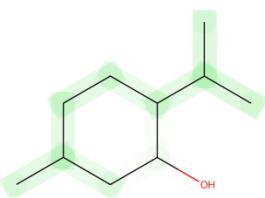
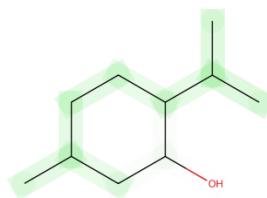
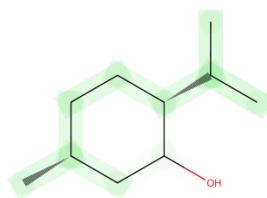
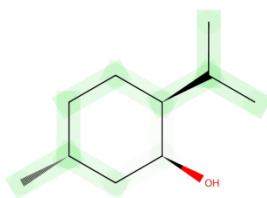
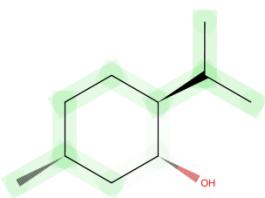
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The fragment "C-C-C(-C)-C-O" contains a non-polar hydrocarbon chain (C-C-C(-C)-C) and a polar hydroxyl group (-O). The hydrocarbons, being non-polar, tend to decrease water solubility because they do not interact favorably with water, a polar solvent. However, the presence of the hydroxyl group, which can form hydrogen bonds with water, can increase solubility. The negative influence on water solubility suggests that in this fragment, the non-polar interactions of the hydrocarbon chain outweigh the polar interactions of the hydroxyl group.

**Hypothesis:** The molecular fragment "C-C-C(-C)-C-O" slightly decreases the water solubility of molecules. The long non-polar hydrocarbon chain likely impairs solubility more than the hydroxyl group can enhance it due to unfavorable interactions with water. The presence of a single polar functional group is not enough to overcome the solubility-reducing effect of the non-polar carbon chain.

# Cluster #14 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 14, from importance channel 0 (*negative*), represents a motif consisting of 2.8 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of -0.8 ( $\pm 0.3$ ) on the prediction outcome.

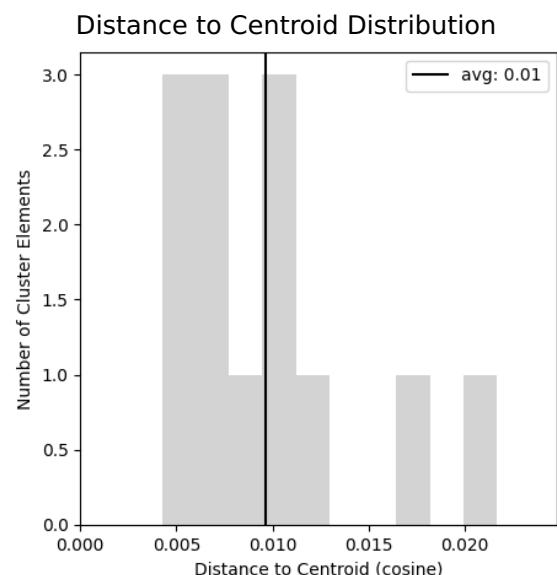
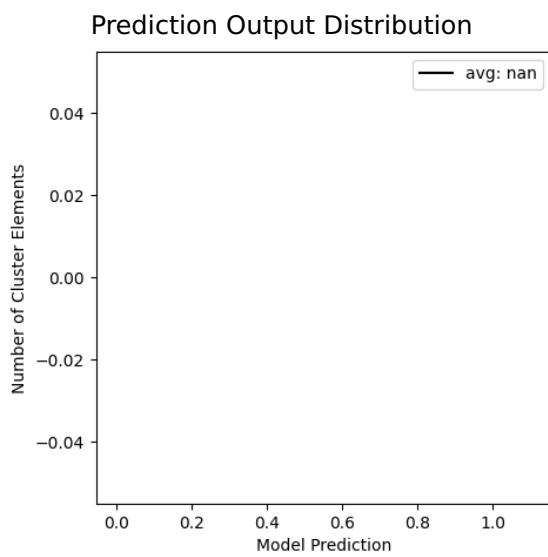
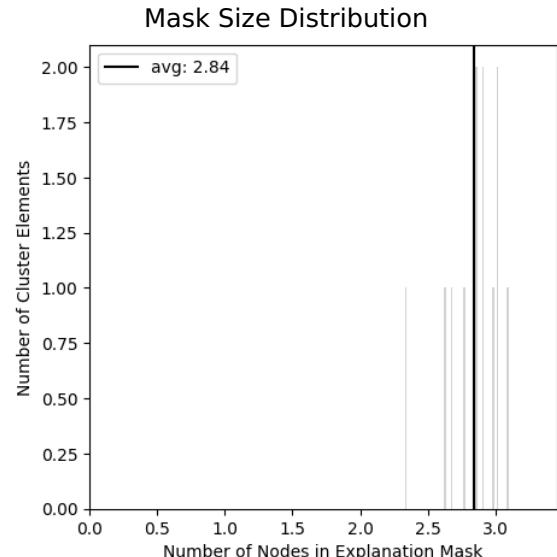
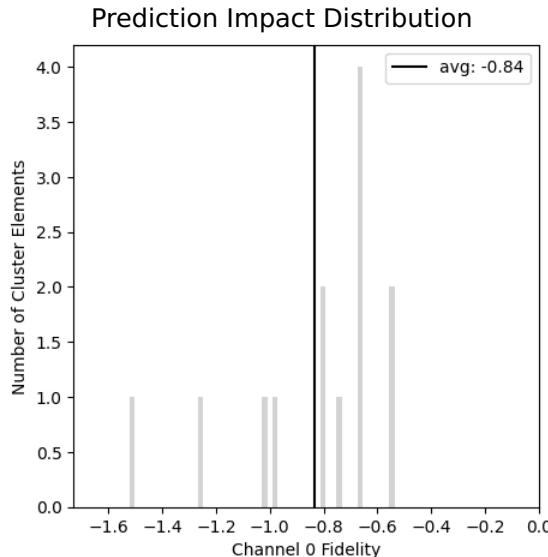
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	13
Channel Index	0.0 (0.0)

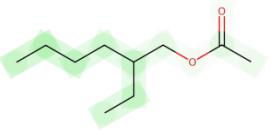
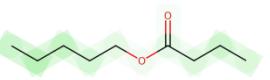
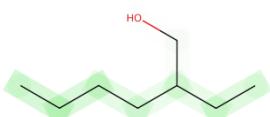
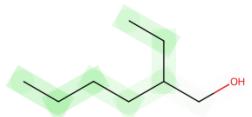
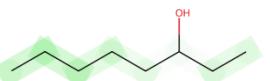
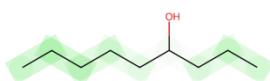
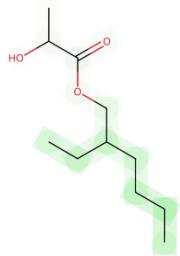
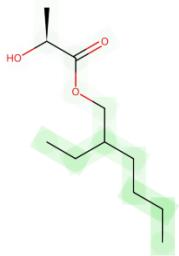
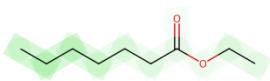
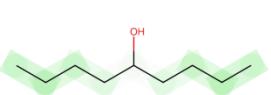
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



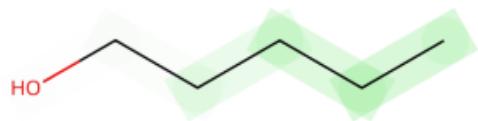
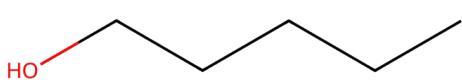
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a linear alkyl chain with a terminal hydroxyl group. The alkyl chain (C-C-C-C) is hydrophobic and typically decreases solubility in water due to its nonpolar nature. However, the presence of a hydroxyl group (-OH) at one end introduces a polar functional group that can form hydrogen bonds with water molecules. The hydrogen bonding capability of the hydroxyl group partially mitigates the hydrophobicity of the alkyl chain and can enhance the solubility of the molecule in water up to a point. The evidence suggests that despite the presence of the hydroxyl group, the negative influence indicates that the hydrophobic interactions due to the alkyl chain dominate, thus reducing overall water solubility.

**Hypothesis:** Alkyl chains with a terminal hydroxyl group have reduced water solubility due to the predominance of hydrophobic interactions over the hydrogen bonding capability of the hydroxyl group. The balance between increased solubility due to the polar hydroxyl group and decreased solubility because of the nonpolar alkyl chain results in an overall negative influence on water solubility. The longer the hydrophobic alkyl chain, the more substantial its influence on reducing water solubility, despite the hydrophilic contributions of the hydroxyl group.

# Cluster #15 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 15, from importance channel 0 (*negative*), represents a motif consisting of 3.1 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of -1.7 ( $\pm 0.8$ ) on the prediction outcome.

## Properties

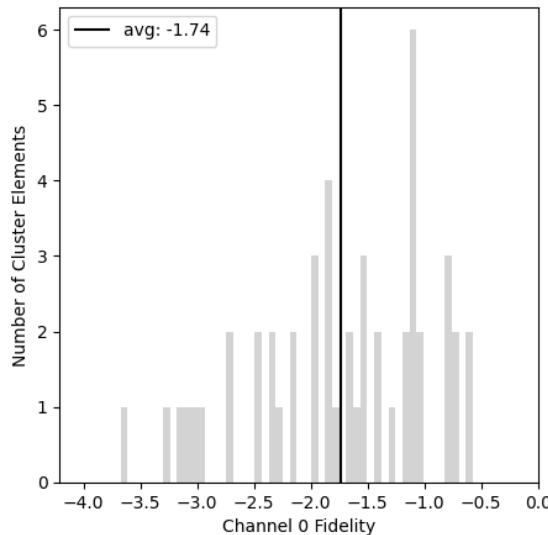
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	49
Channel Index	0.0 (0.0)

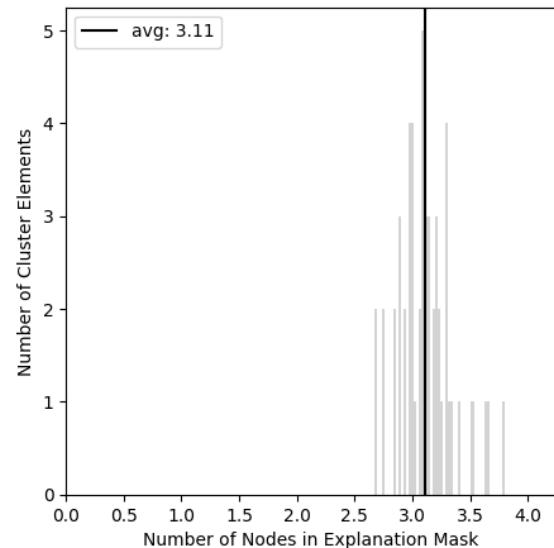
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

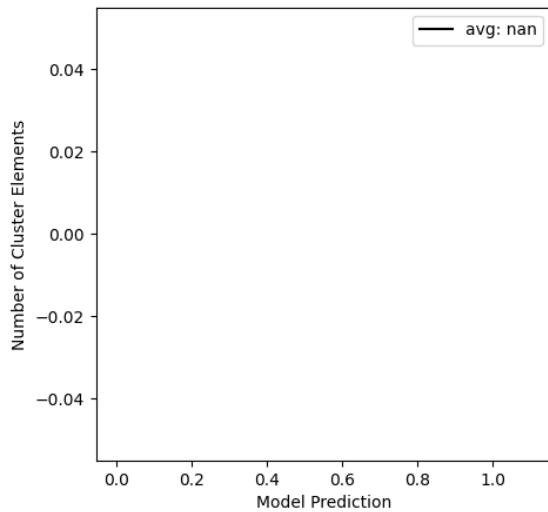
Prediction Impact Distribution



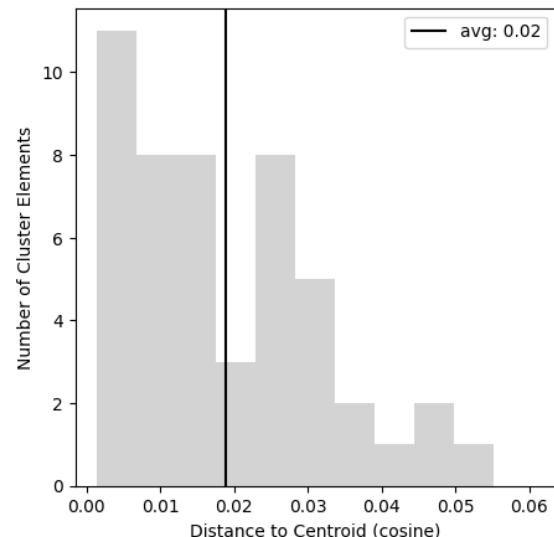
Mask Size Distribution



Prediction Output Distribution

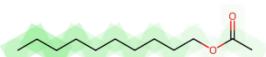
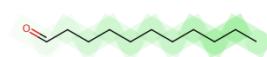
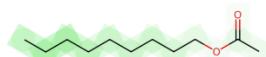
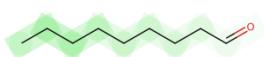
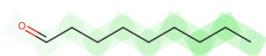
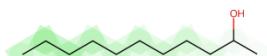
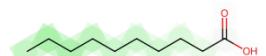
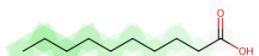
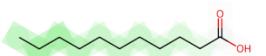
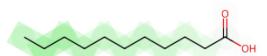
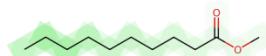
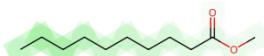
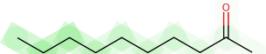
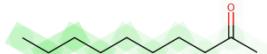
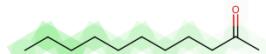
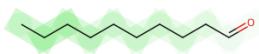


Distance to Centroid Distribution



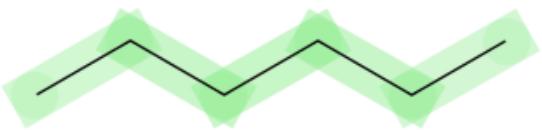
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES representation "C-C-C-C-C" depicts a hydrocarbon chain, which is a non-polar molecule. The length of the chain suggests that it is a significant hydrophobic fragment, which decreases water solubility. Water solubility is often due to the ability of a molecule to hydrogen bond or interact with water molecules, and non-polar molecules do not have this capability.

**Hypothesis:** A straight chain hydrocarbon structure is negatively correlated with water solubility. The increased length of the carbon chain enhances hydrophobic interactions, and the lack of polar functional groups fails to provide any affinity towards water molecules, both factors leading to a decrease in solubility in water.

# Cluster #16 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 16, from importance channel 0 (*negative*), represents a motif consisting of 3.0 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of -1.9 ( $\pm 0.4$ ) on the prediction outcome.

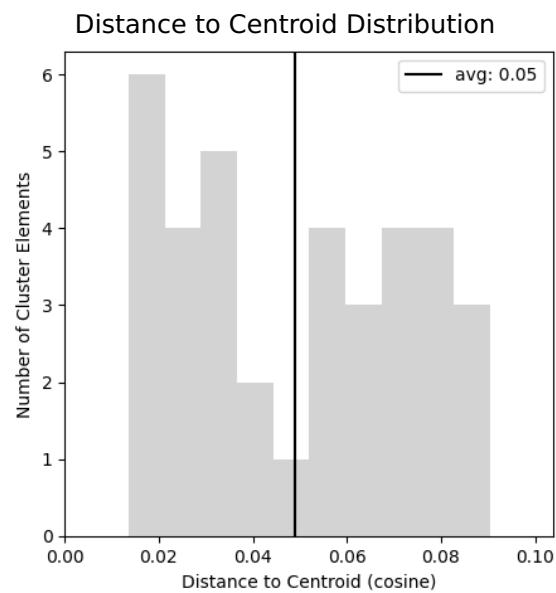
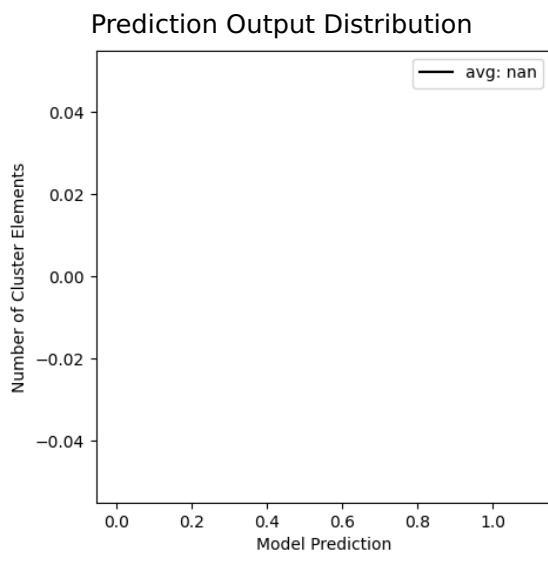
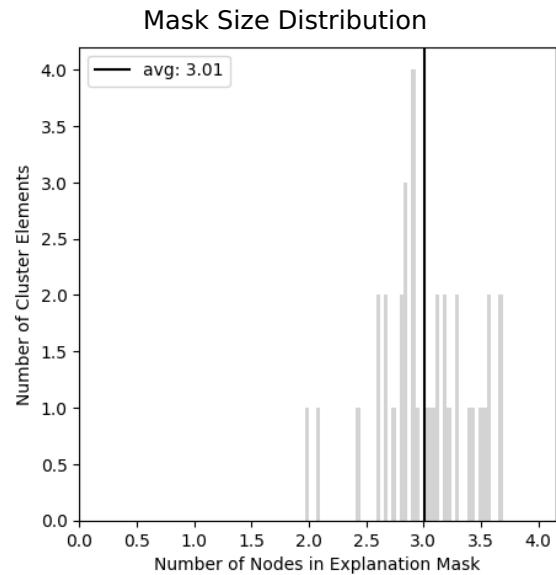
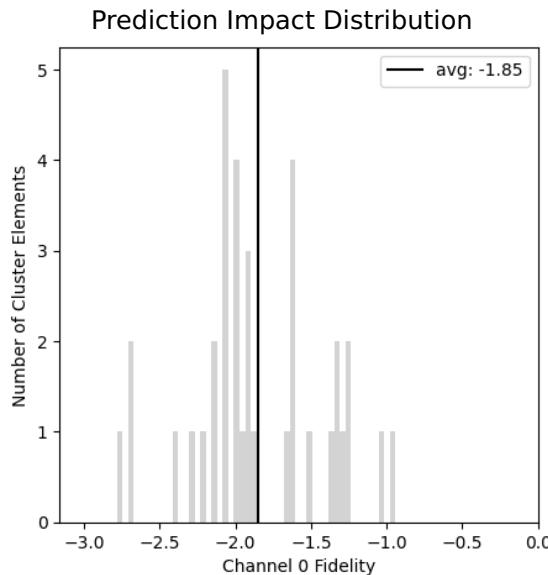
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	36
Channel Index	0.0 (0.0)

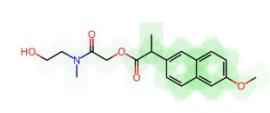
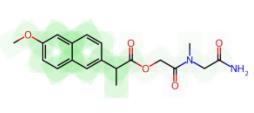
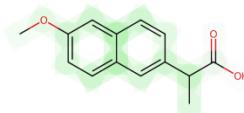
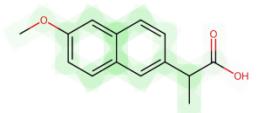
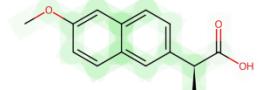
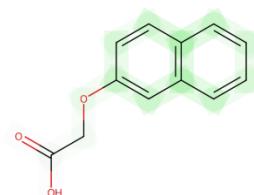
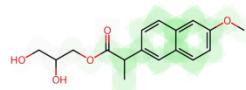
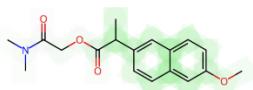
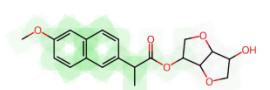
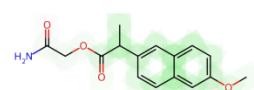
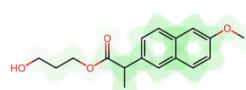
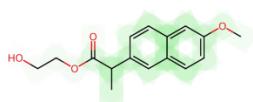
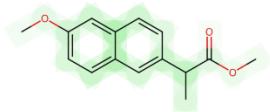
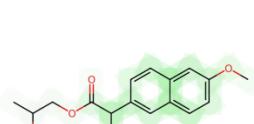
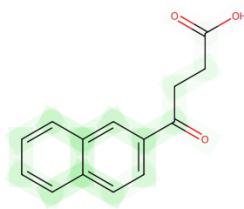
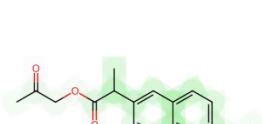
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



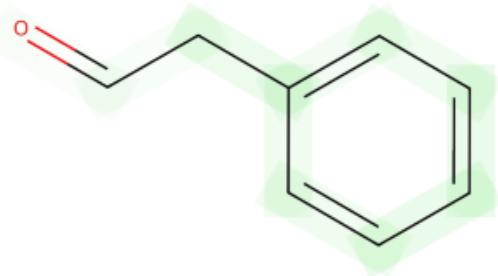
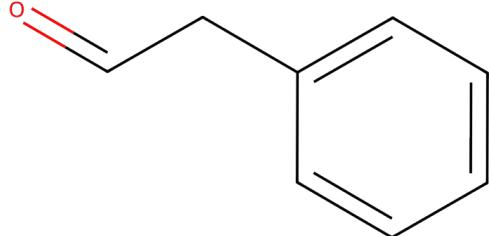
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES representation corresponds to a structure with a phenyl ring (c1ccccc1) attached to a vinyl group (-C=C-) which is in turn connected to a carbonyl group (O=C). The phenyl ring is a hydrophobic moiety, reducing water solubility, while the vinyl group adds some rigidity to the structure, which may hinder effective interactions with water molecules. Meanwhile, the carbonyl group can engage in hydrogen bonding due to its polar nature, which could increase water solubility. In this case, the hydrophobic character of the phenyl ring seems to outweigh the polar interactions provided by the carbonyl group, resulting in a negative influence on water solubility.

**Hypothesis:** A molecule containing a phenyl ring attached to a vinyl group and a carbonyl group has a tendency to exhibit lower water solubility. The hydrophobic interactions from the phenyl ring dominate over the potential hydrogen bonding from the carbonyl group, thereby reducing overall solubility in water.

# Cluster #17 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 17, from importance channel 0 (*negative*), represents a motif consisting of 3.2 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of -2.0 ( $\pm 0.2$ ) on the prediction outcome.

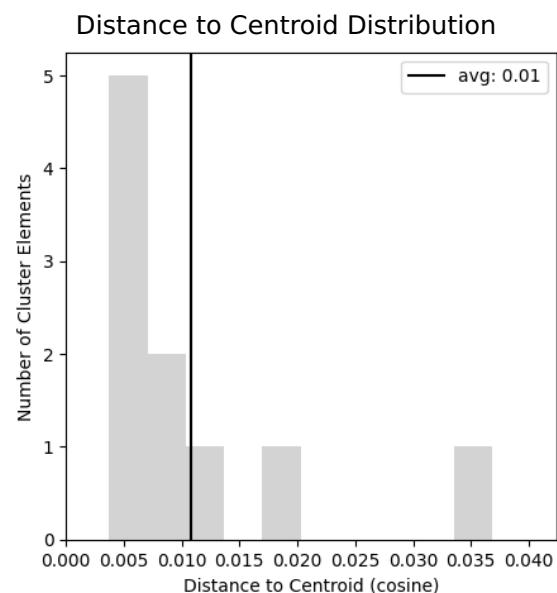
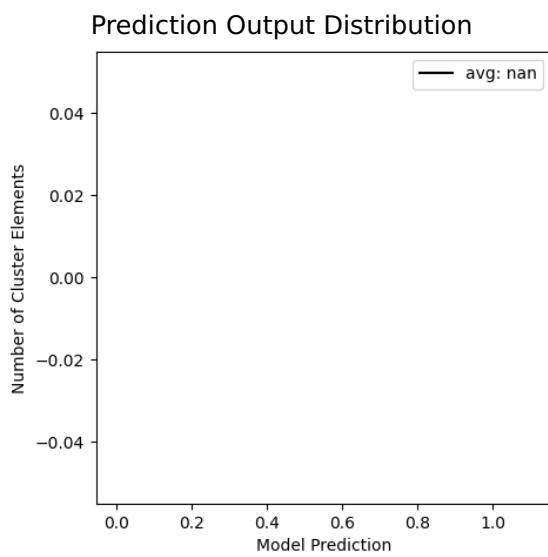
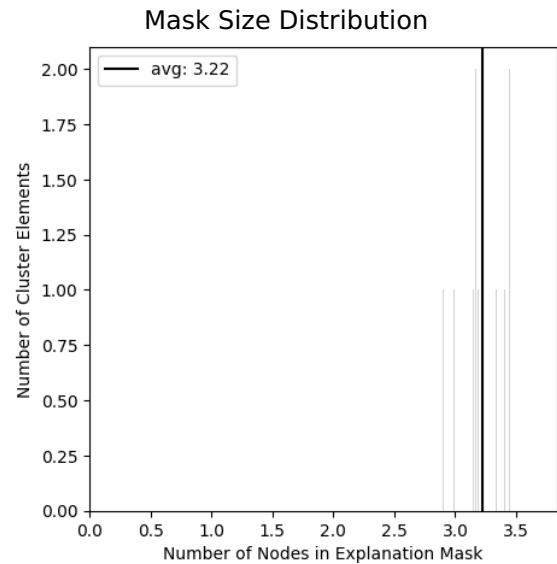
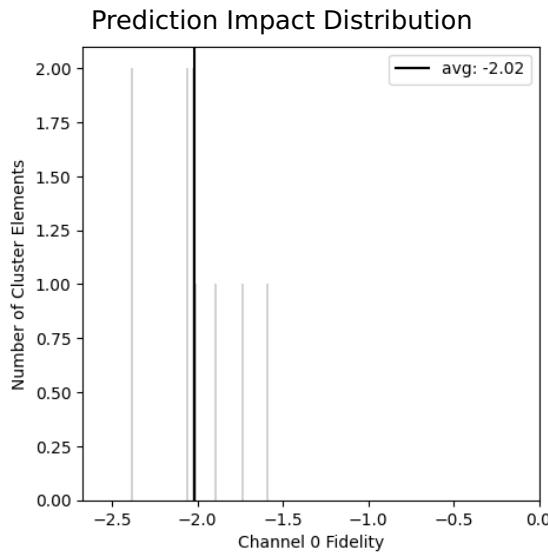
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	10
Channel Index	0.0 (0.0)

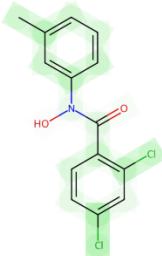
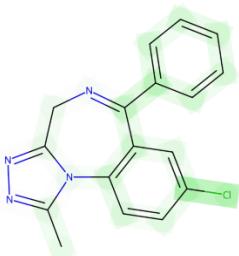
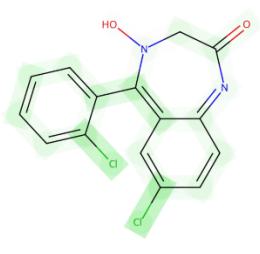
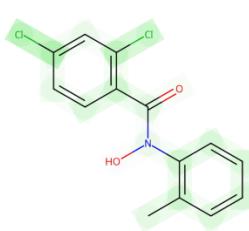
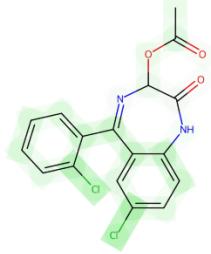
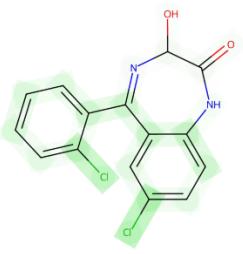
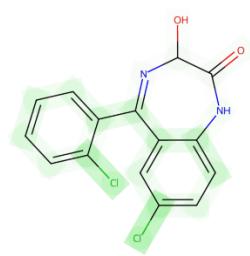
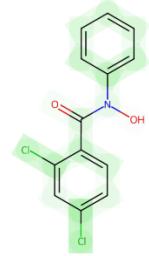
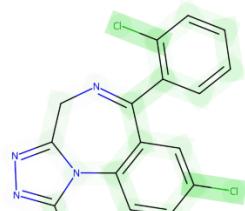
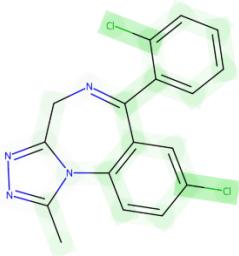
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



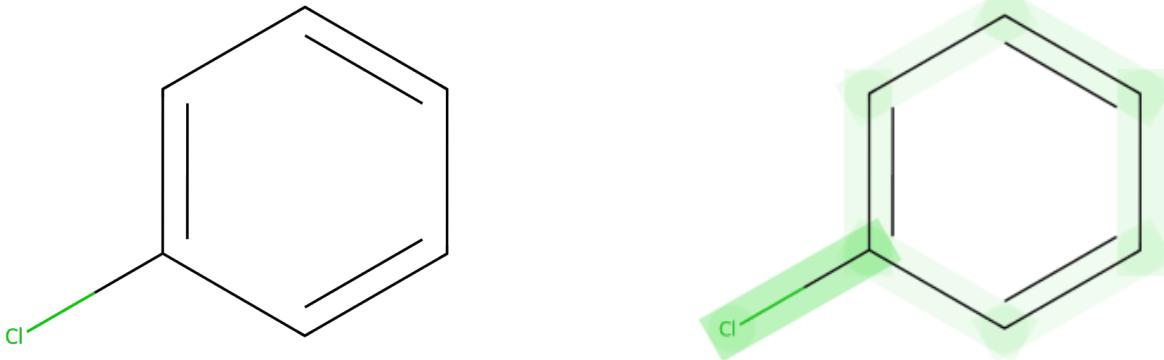
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure "Cl-c1:c:c:c:c:c:1" represents a benzene ring with a chlorine substituent, known as chlorobenzene. The presence of the chlorine atom, which is highly electronegative, slightly polarizes the molecule. However, the benzene ring is hydrophobic due to its non-polar nature, making it resistant to mixing with polar solvents like water. The chloro-substitution is not sufficient to overcome the intrinsic hydrophobic character of the benzene ring, resulting in an overall decrease in water solubility.

**Hypothesis:** Chlorobenzene substructures tend to decrease water solubility due to the benzene ring's hydrophobic nature. The electronegative chlorine atom does not significantly enhance solubility since the molecule's overall polarity is low and cannot favorably interact with water molecules.

# Cluster #18 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 18, from importance channel 0 (*negative*), represents a motif consisting of 2.8 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of -1.2 ( $\pm 0.5$ ) on the prediction outcome.

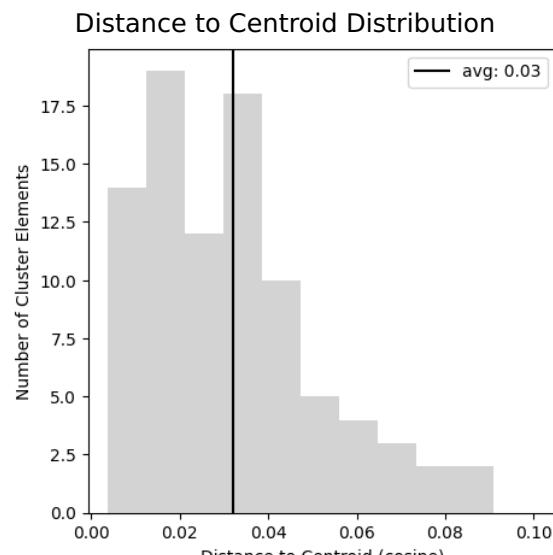
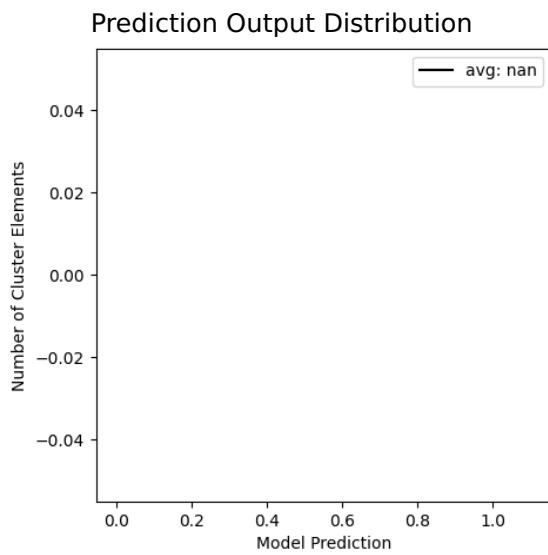
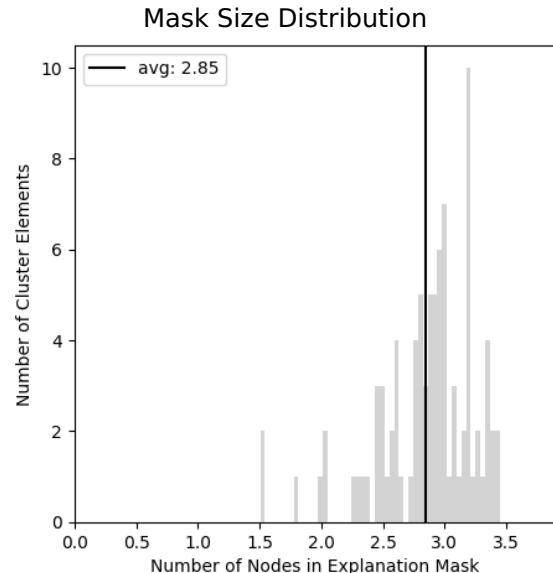
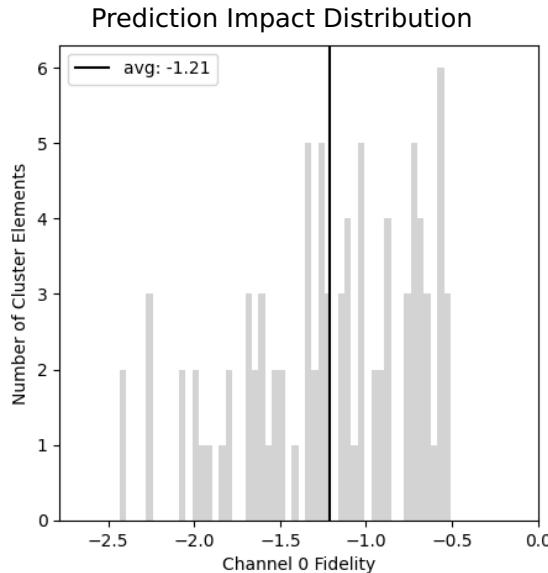
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	89
Channel Index	0.0 (0.0)

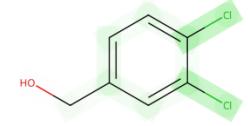
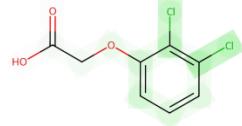
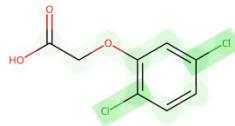
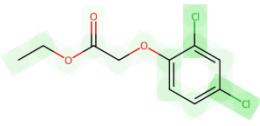
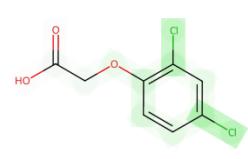
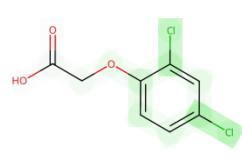
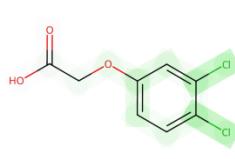
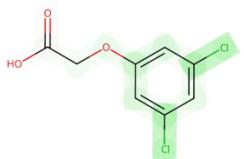
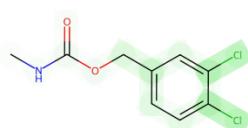
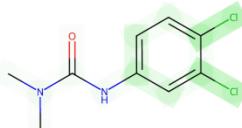
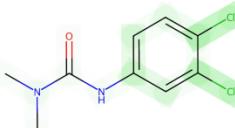
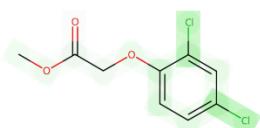
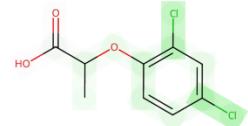
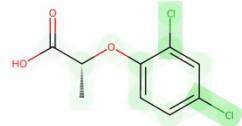
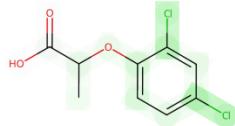
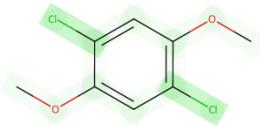
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



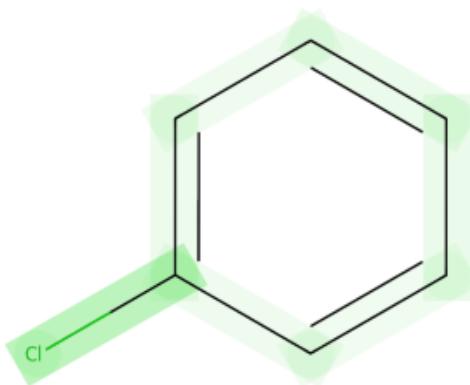
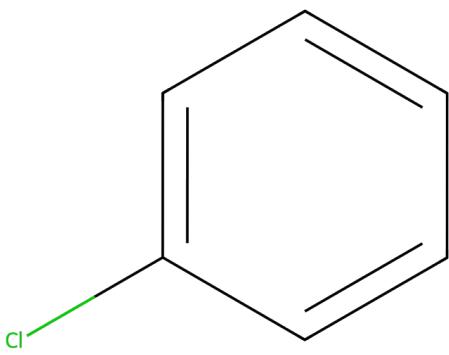
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES represents a benzene ring where one hydrogen atom has been replaced by a chlorine atom, creating chlorobenzene. The presence of chlorine, being an electronegative atom, can increase the polarity of the molecule to a small extent, allowing for some interaction with the polar solvent water. However, the benzene ring itself is non-polar and hydrophobic due to its stable system of delocalized pi electrons, which resists solubilization in water. The hydrophobic character of the benzene ring tends to overpower the slight increase in polarity due to the chlorine substituent, leading to an overall decrease in the water solubility of the molecule.

**Hypothesis:** Chlorobenzene substructure decreases water solubility of molecules. The chloro group, although polar, cannot sufficiently counterbalance the hydrophobic nature of the benzene ring which dominates the interaction with water, leading to decreased solubility.

# Cluster #19 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 19, from importance channel 0 (*negative*), represents a motif consisting of 2.4 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of -2.6 ( $\pm 0.3$ ) on the prediction outcome.

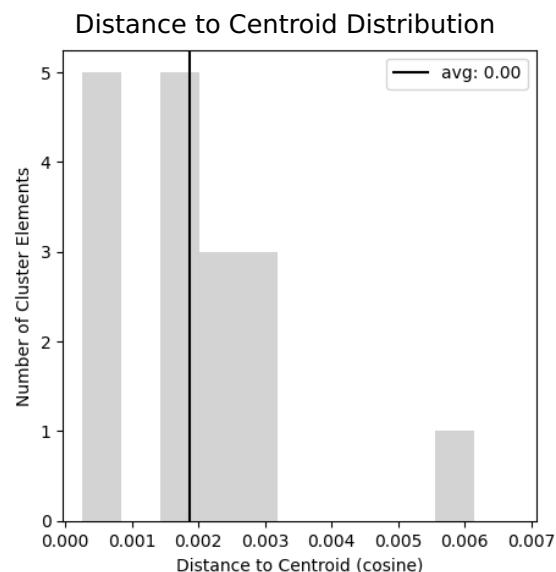
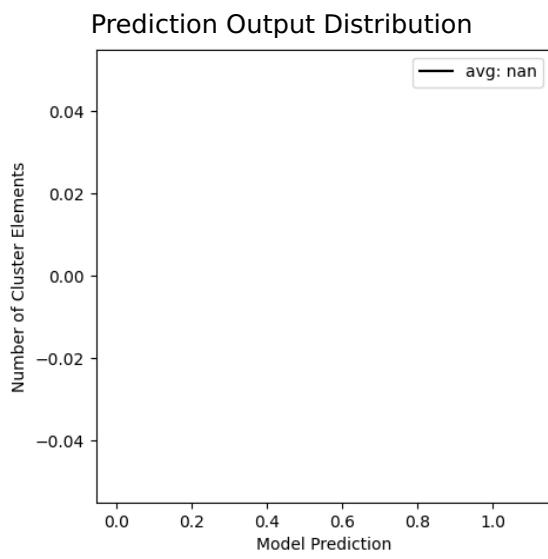
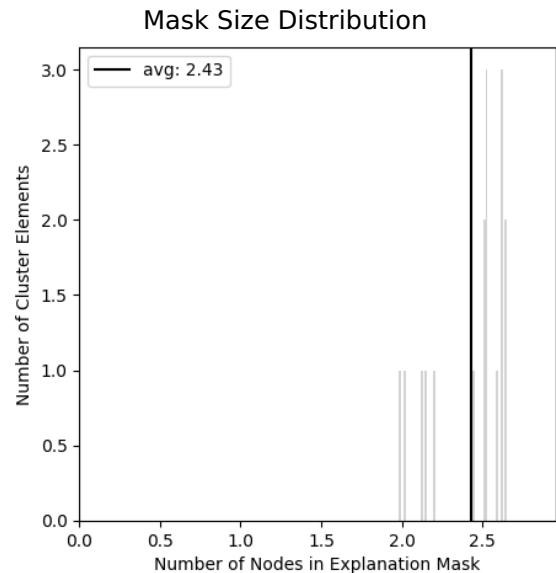
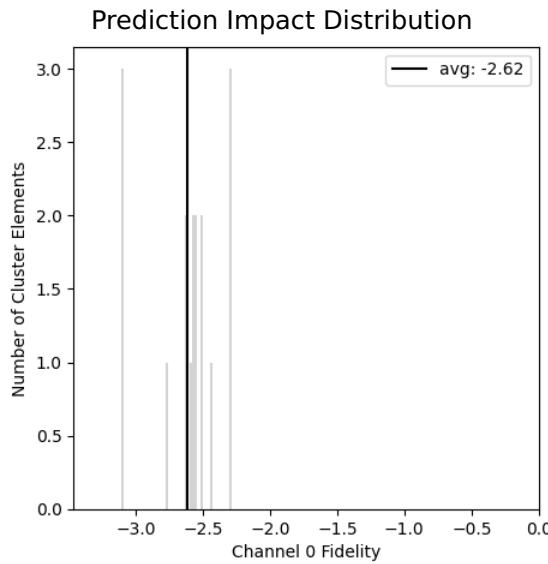
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	17
Channel Index	0.0 (0.0)

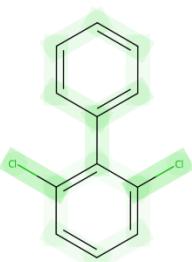
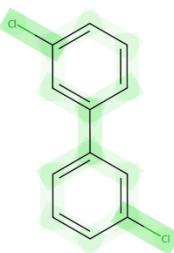
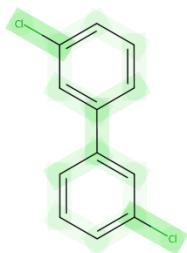
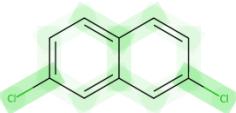
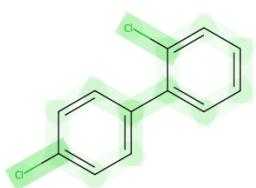
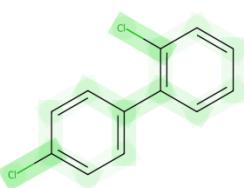
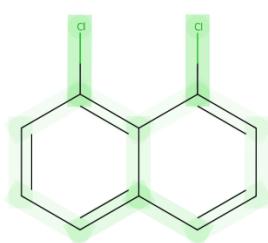
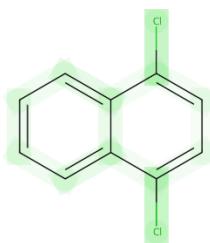
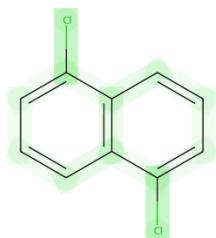
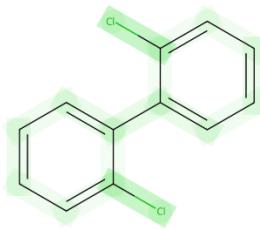
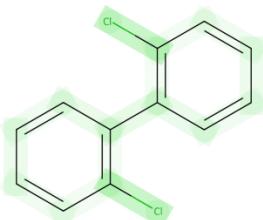
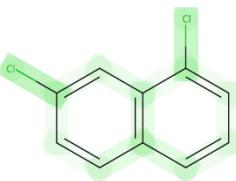
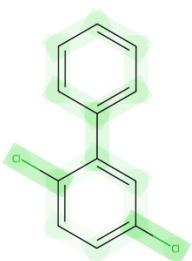
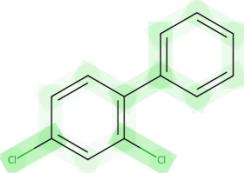
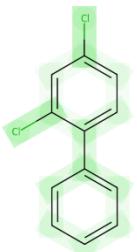
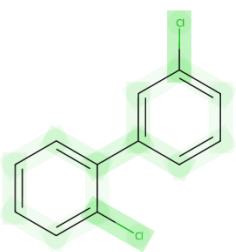
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



# Example Elements

① This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES structure "Cl-c1:c:c:c:(-Cl):c:c:1" represents a dichlorobenzene ring, a benzene ring substituted with two chlorine atoms. The presence of halogens like chlorine usually decreases water solubility because halogens are less polar than water and increase the hydrophobic character of the molecule. Additionally, chlorine atoms are relatively large substituents which can hinder the interaction of the molecule with water molecules. Consequently, the structure's planarity and aromatic nature reduce its potential to form hydrogen bonds with water, a key factor in solubility.

**Hypothesis:** The presence of two chlorine atoms on a benzene ring significantly reduces water solubility of the molecule. This is due to the increased hydrophobicity from the chlorine substituents and the reduced capacity to engage in hydrogen bonding with water, both of which are important factors in the solvation process.

# Cluster #20 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 20, from importance channel 0 (*negative*), represents a motif consisting of 3.1 ( $\pm 0.0$ ) nodes. The concept is generally associated with an impact of -5.1 ( $\pm 0.1$ ) on the prediction outcome.

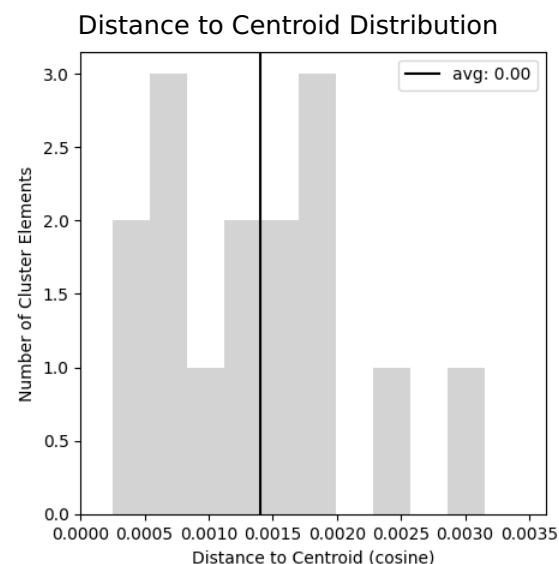
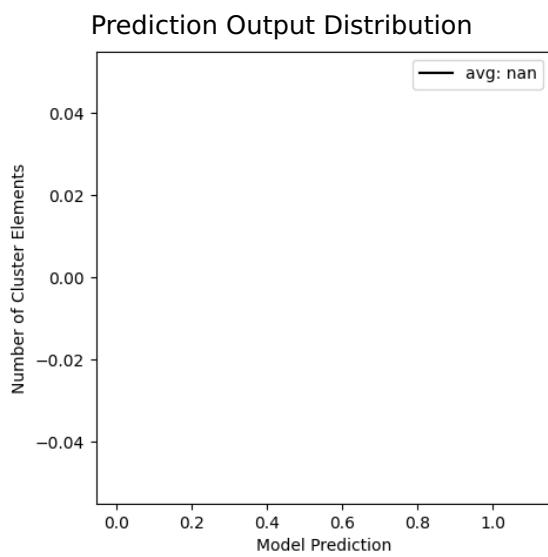
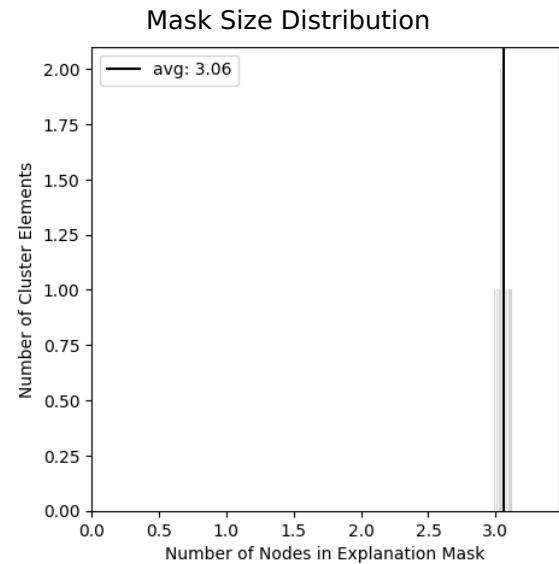
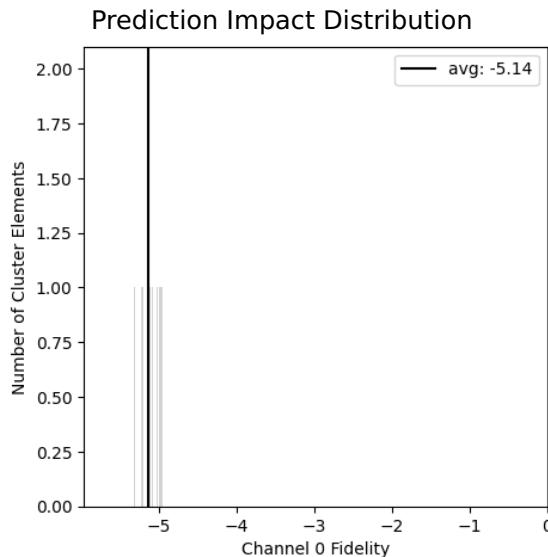
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	15
Channel Index	0.0 (0.0)

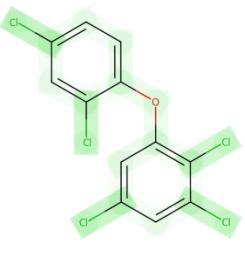
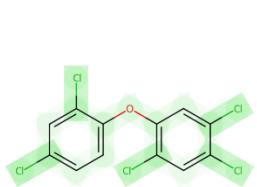
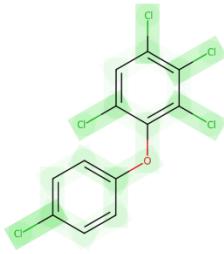
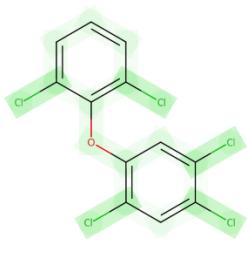
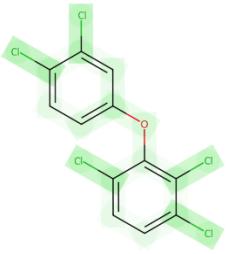
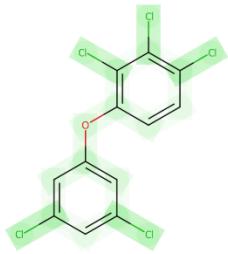
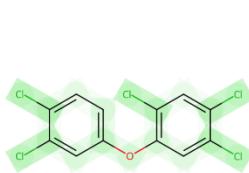
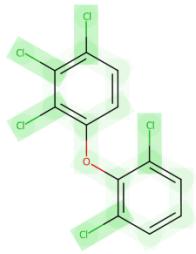
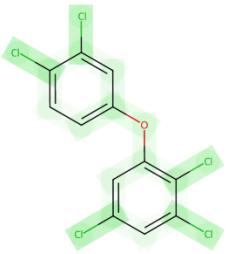
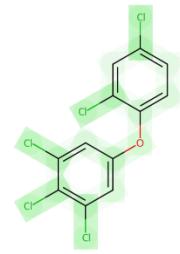
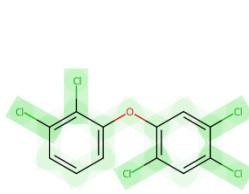
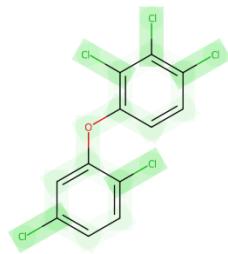
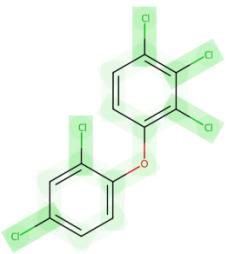
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



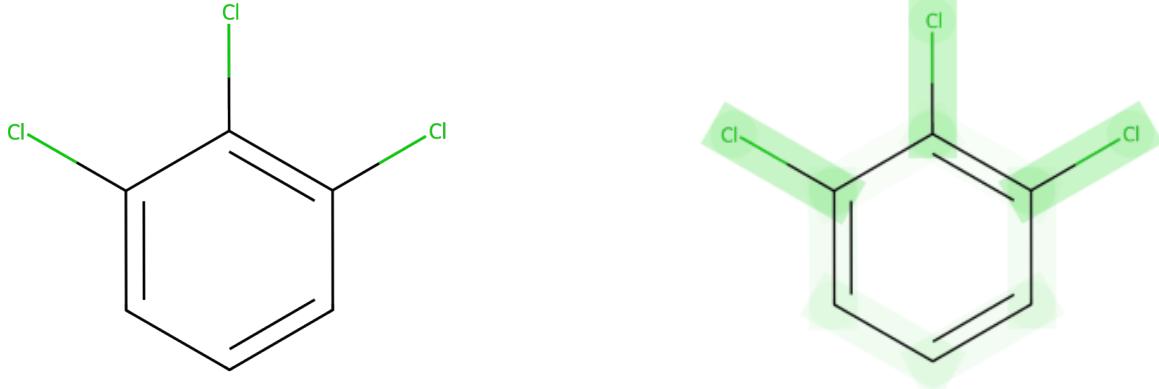
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The presence of chlorine atoms attached to a benzene ring as depicted in the SMILES "Cl-c1:c:c:c:c(-Cl):c:1-Cl" indicates a tri-chlorinated benzene molecule, which is hydrophobic in nature. Chlorine atoms are halogens and generally render the molecule more lipophilic, decreasing its ability to hydrogen bond with water molecules. As a result, the solubility of this molecule in water would be expected to be low.

**Hypothesis:** A tri-chlorinated benzene molecule is expected to have low water solubility. The electron-withdrawing effect of the chlorine atoms decreases the polarity of the molecule, making it less likely to interact with water. Additionally, the size and the nature of the chlorine decrease the molecule's ability to participate in hydrogen bonding, further reducing its solubility in water.

# Cluster #21 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 21, from importance channel 0 (*negative*), represents a motif consisting of 3.0 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of -5.2 ( $\pm 0.4$ ) on the prediction outcome.

## Properties

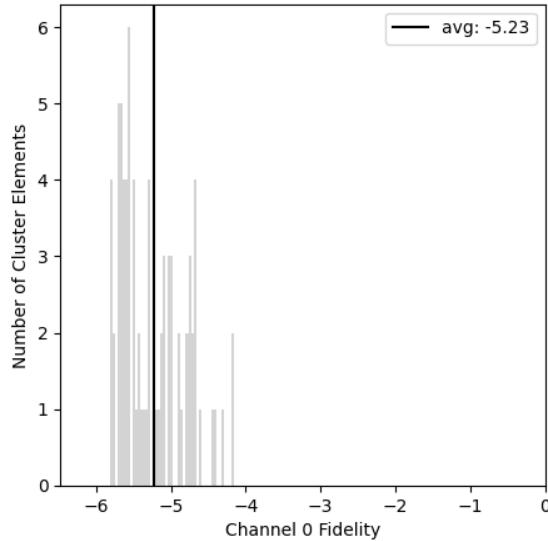
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	82
Channel Index	0.0 (0.0)

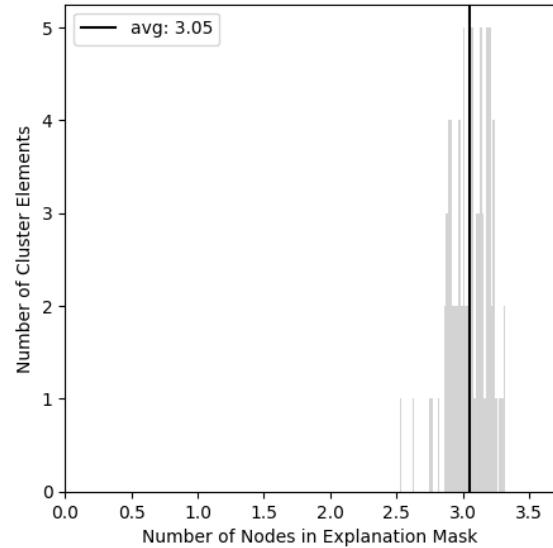
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

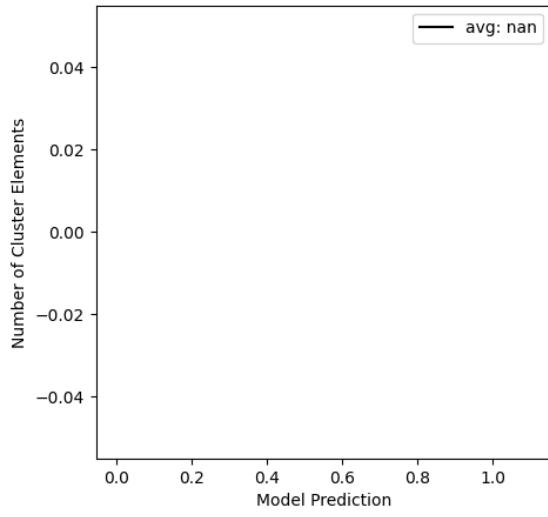
Prediction Impact Distribution



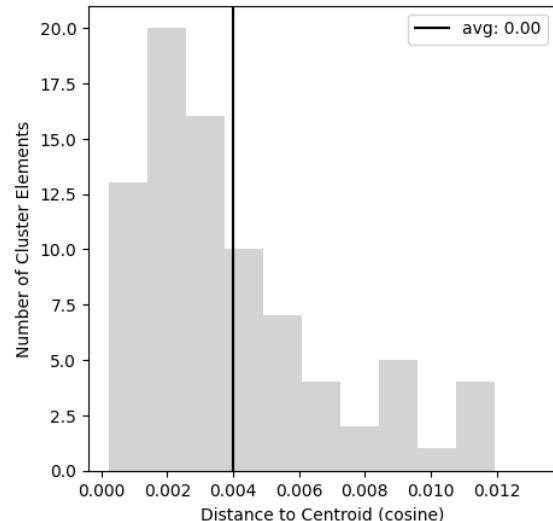
Mask Size Distribution



Prediction Output Distribution

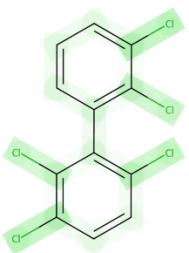
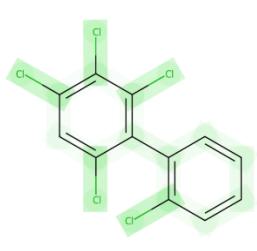
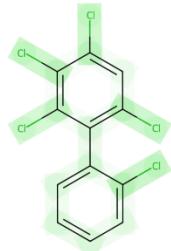
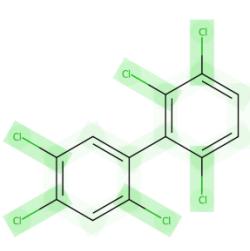
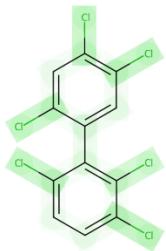
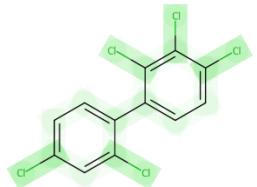
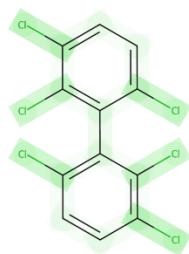
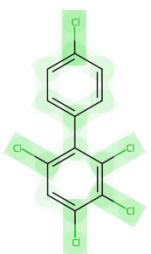
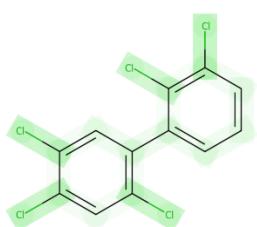
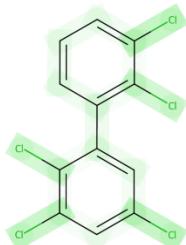
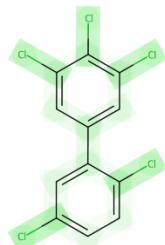
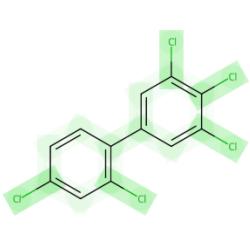


Distance to Centroid Distribution



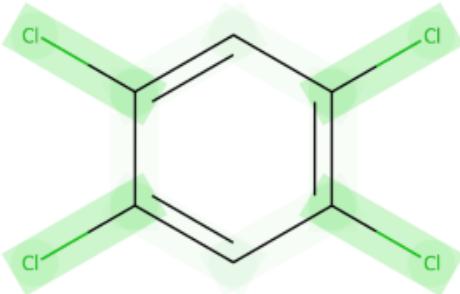
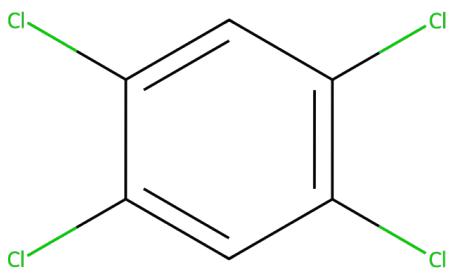
# Example Elements

① This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES "Cl-c1:c:c(-Cl):c(-Cl):c:c1-Cl" represents a symmetrical molecule where a benzene ring is substituted with four chlorine atoms. Chlorine is significantly more electronegative than carbon, which suggests that the electrons in the C-Cl bonds will be pulled towards the chlorine atoms, polarizing these bonds. However, chlorobenzenes exhibit decreased solubility in water compared to benzene due to the hydrophobic nature of the aromatic ring and the overall size and symmetry of the molecule which makes it less energetically favorable for water molecules to solvate and disrupt the structured network of hydrogen bonds.

**Hypothesis:** A benzene ring substituted with four chlorine atoms decreases water solubility. The high electronegativity of chlorine polarizes the C-Cl bonds but the molecule's hydrophobic character and steric hindrance outweigh the potential increase in solubility due to polarity, leading to a decrease in water solubility.

# Cluster #22 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 22, from importance channel 0 (*negative*), represents a motif consisting of 3.4 ( $\pm 0.1$ ) nodes. The concept is generally associated with an impact of -6.2 ( $\pm 0.1$ ) on the prediction outcome.

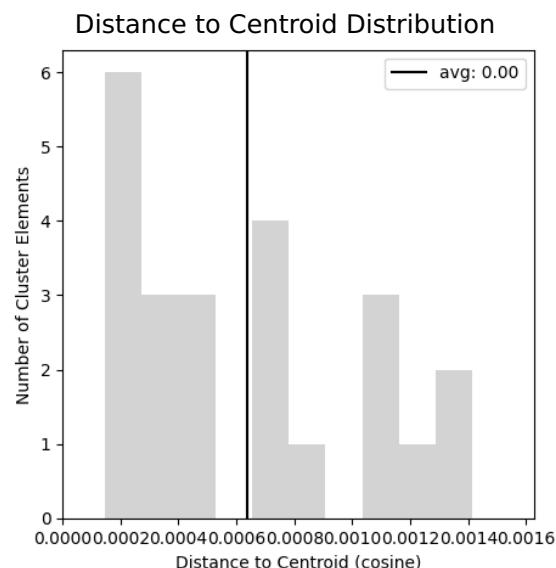
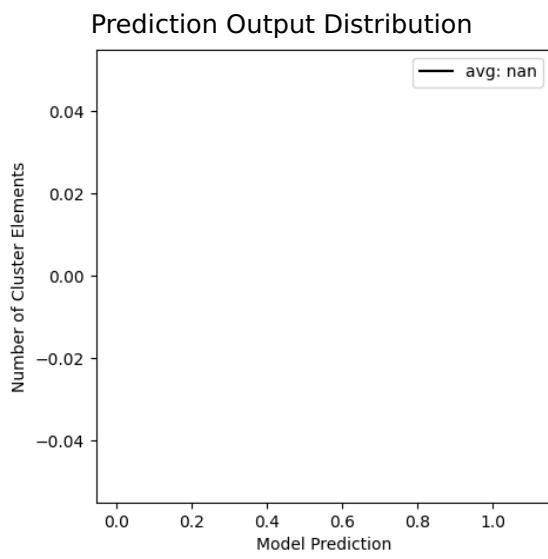
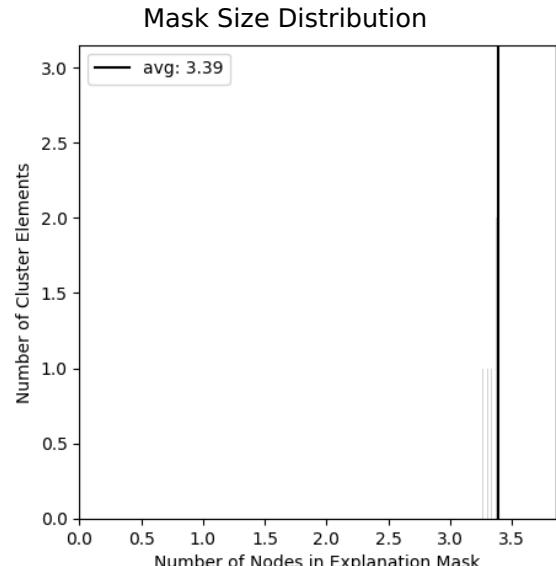
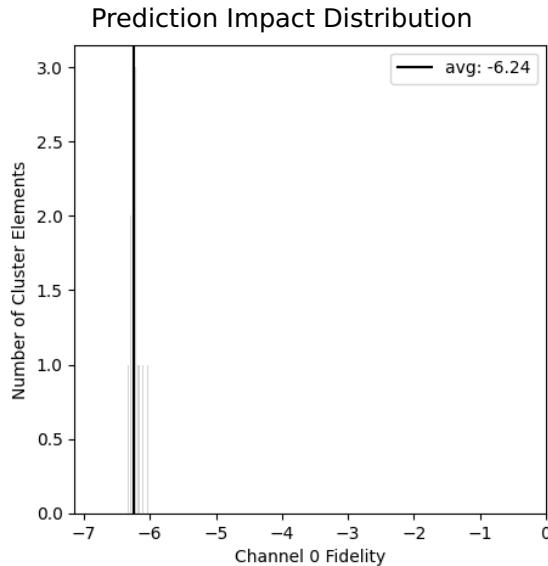
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	23
Channel Index	0.0 (0.0)

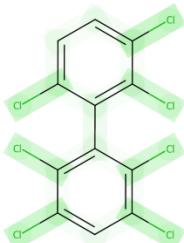
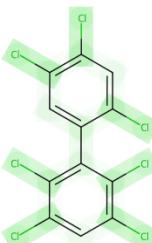
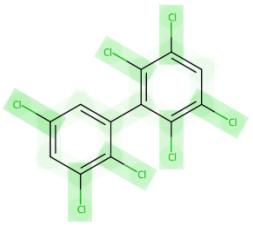
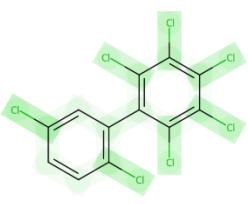
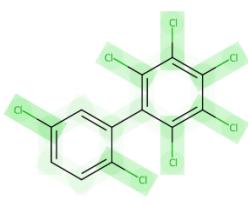
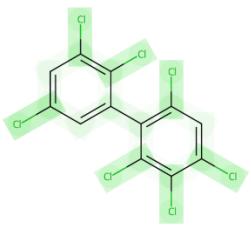
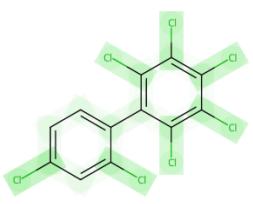
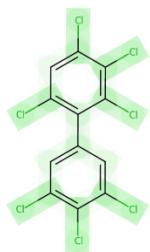
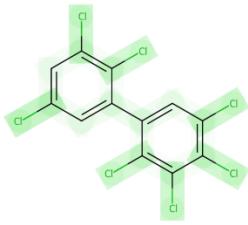
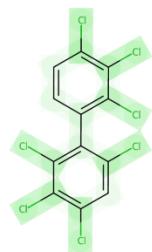
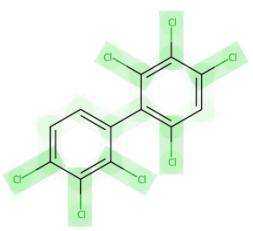
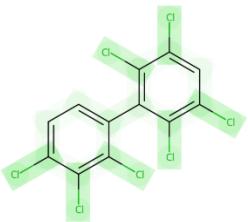
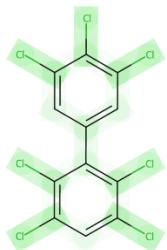
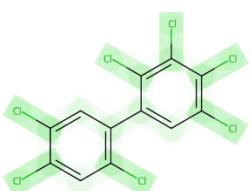
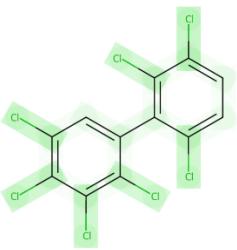
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



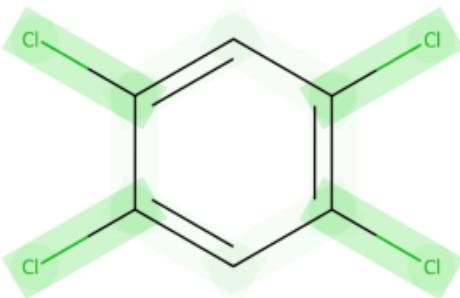
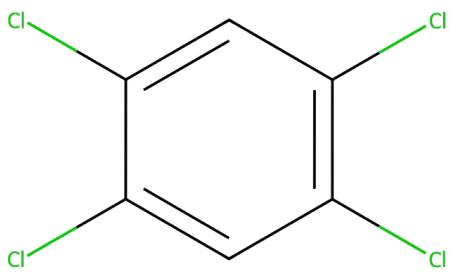
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure provided corresponds to pentachlorobenzene, which is a benzene ring with five chlorine atoms attached to it. Chlorine is a halogen, and while small organic molecules containing a single halogen may have increased solubility due to dipole interactions with water, the presence of multiple halogens – especially five, as in this case – tends to significantly decrease water solubility. This decrease is due to the increase in hydrophobic character, making the compound less likely to interact with polar solvent molecules like water. Additionally, the increased steric bulk and electronic effects of multiple chlorine atoms on the benzene ring can disrupt solvation processes.

**Hypothesis:** Pentachlorobenzene is less soluble in water due to its heavily halogenated structure. The high number of chlorine atoms contributes to a hydrophobic nature and overall bulkiness, which impedes interaction with water molecules and thereby decreases solubility.

# Cluster #23 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 23, from importance channel 0 (*negative*), represents a motif consisting of 3.1 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of -4.5 ( $\pm 0.8$ ) on the prediction outcome.

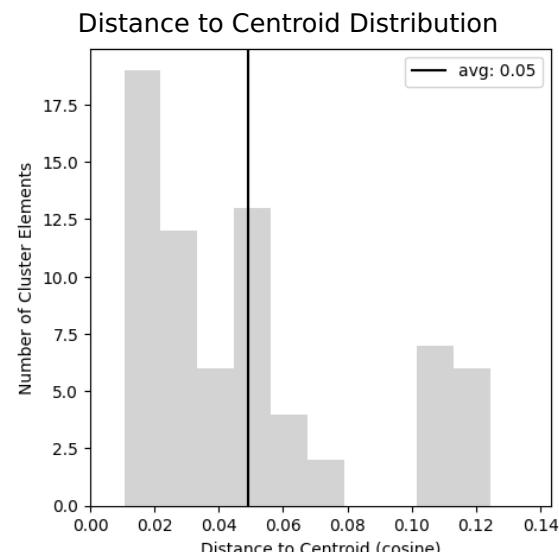
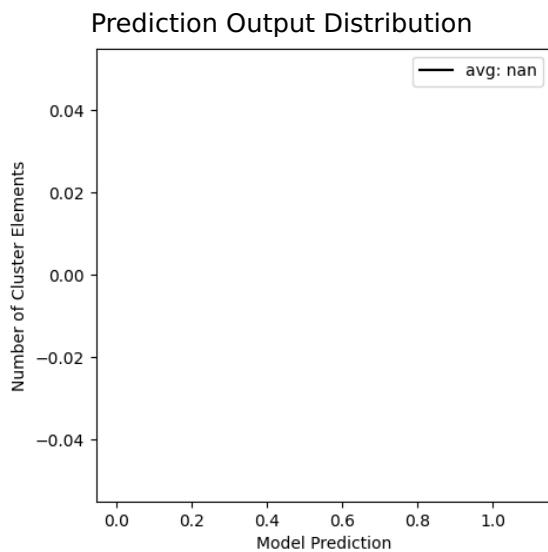
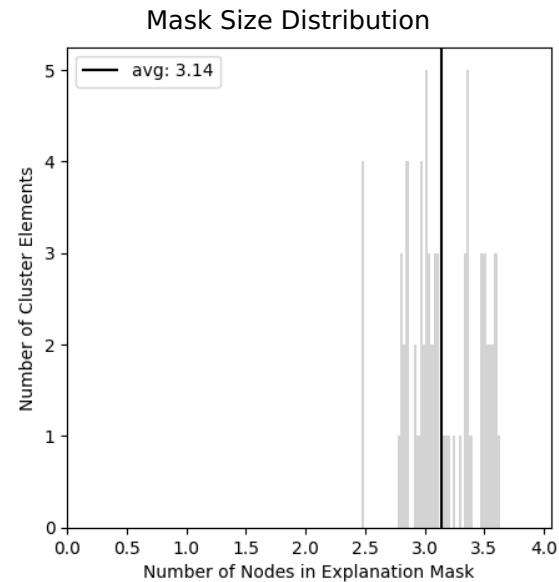
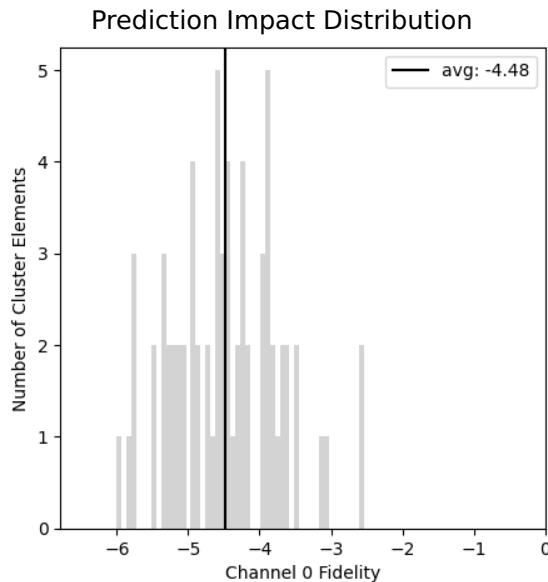
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	69
Channel Index	0.0 (0.0)

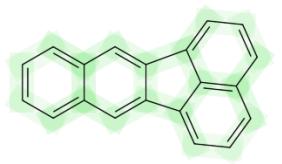
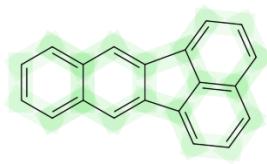
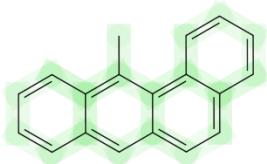
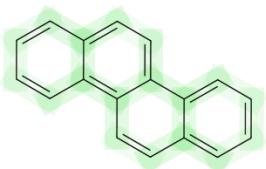
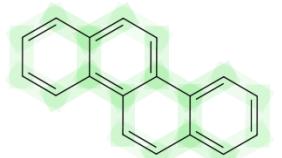
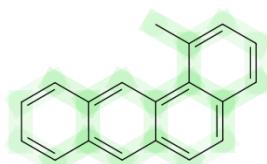
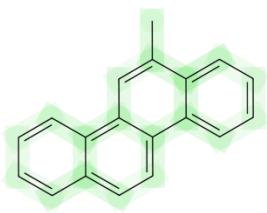
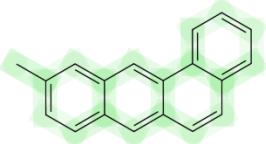
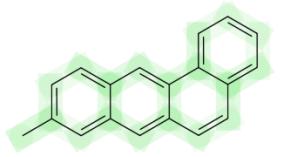
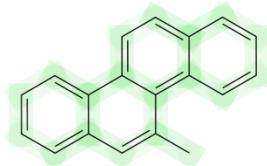
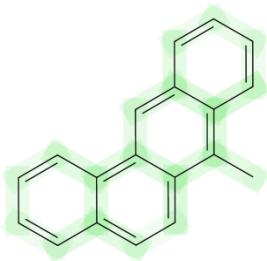
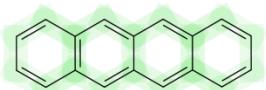
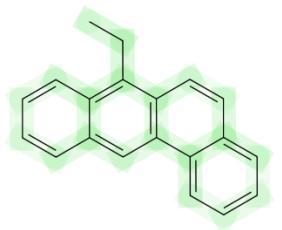
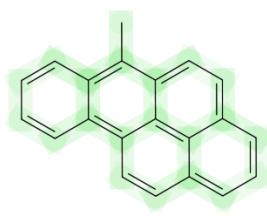
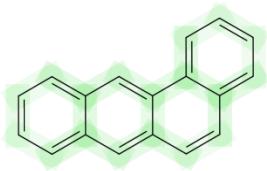
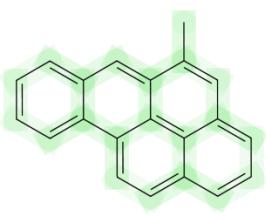
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



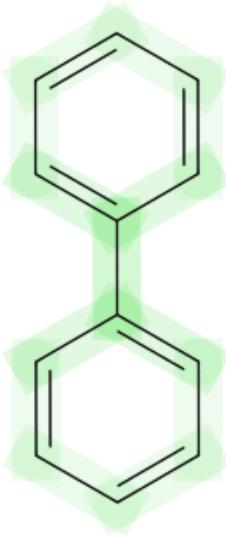
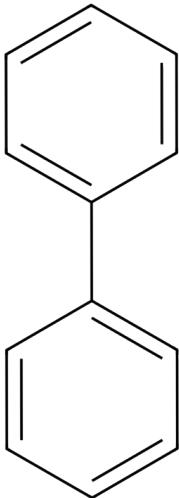
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a polycyclic aromatic hydrocarbon framework, specifically a biphenyl substituent. Polycyclic aromatic hydrocarbons have a tendency to exhibit low solubility in water due to their large hydrophobic surface area and the lack of functional groups that can engage in hydrogen bonding with water molecules. The biphenyl moiety, comprised of two benzene rings connected by a single bond, further enhances this effect by increasing the planarity and  $\pi$ - $\pi$  stacking interactions, which stabilizes the hydrophobic interactions and discourages the solute from dissolving in a polar solvent like water.

**Hypothesis:** Molecules containing biphenyl substructures are likely to be less soluble in water. The extended aromatic system of biphenyl leads to increased hydrophobic interactions and reduced interaction with polar water molecules due to the lack of polar functional groups, resulting in a negative influence on water solubility.

# Cluster #24 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 24, from importance channel 0 (*negative*), represents a motif consisting of 2.6 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of -1.6 ( $\pm 0.3$ ) on the prediction outcome.

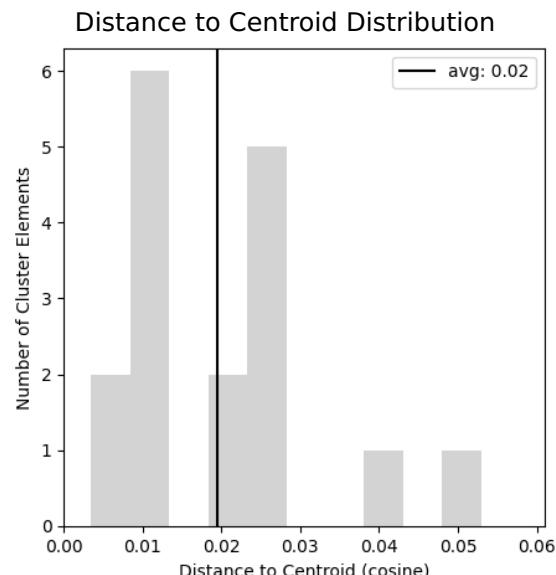
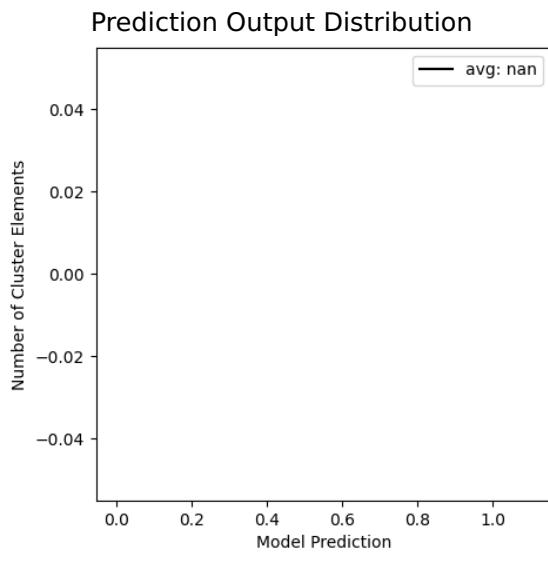
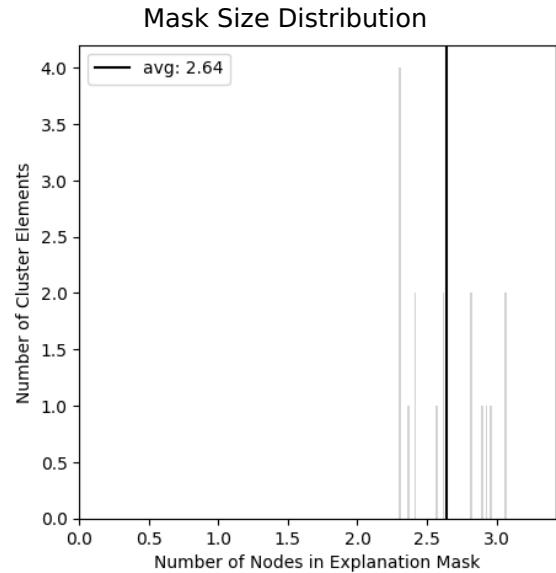
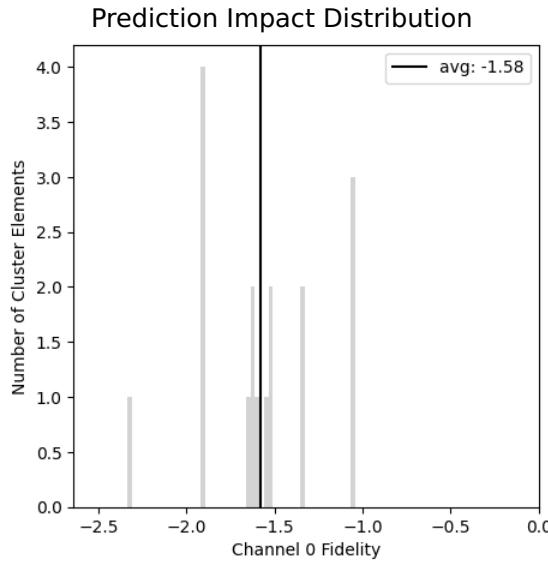
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	17
Channel Index	0.0 (0.0)

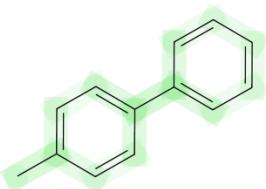
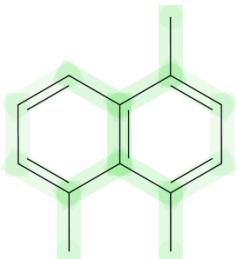
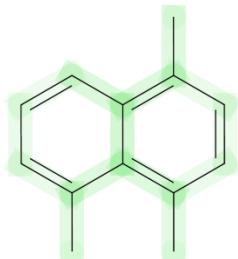
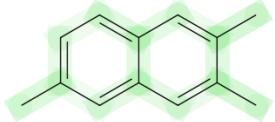
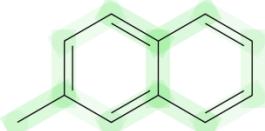
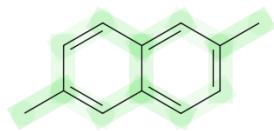
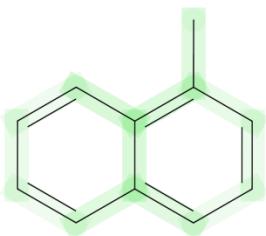
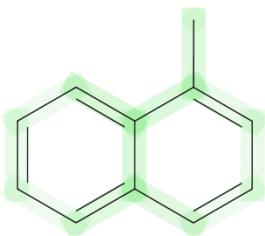
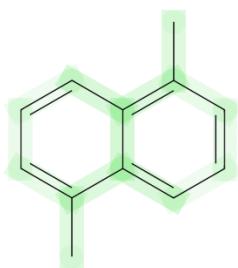
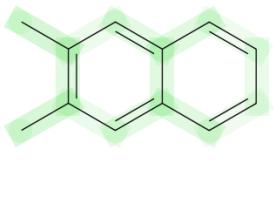
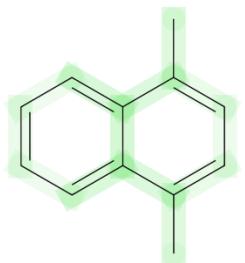
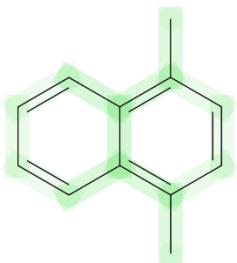
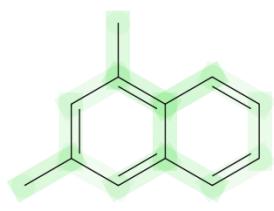
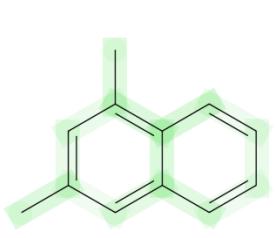
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



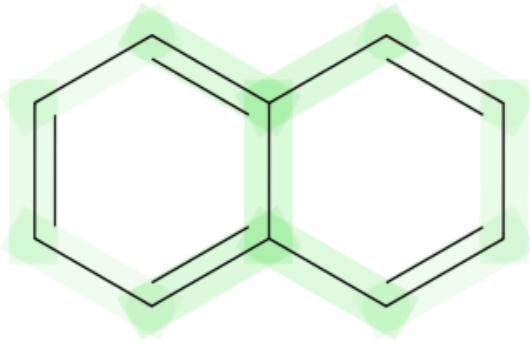
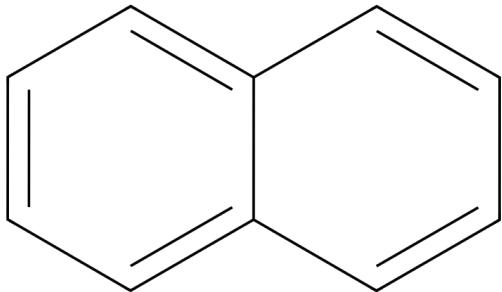
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The provided SMILES representation illustrates a polyaromatic hydrocarbon structure, specifically naphthalene. Such molecules are known for their hydrophobic nature due to the extensive delocalized pi-electron system which does not favor interaction with polar solvent like water. Additionally, the lack of polar or ionizable groups within this structure further diminishes its water solubility.

**Hypothesis:** The structure "c1:c:c:c2:c:c:c:c2:c:1" leads to decreased water solubility. The pi-electron system across the fused aromatic rings reduces interaction with polar water molecules, while the absence of functional groups capable of hydrogen bonding does not support solvation in water.

# Cluster #25 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 25, from importance channel 0 (*negative*), represents a motif consisting of 3.1 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of -0.8 ( $\pm 0.3$ ) on the prediction outcome.

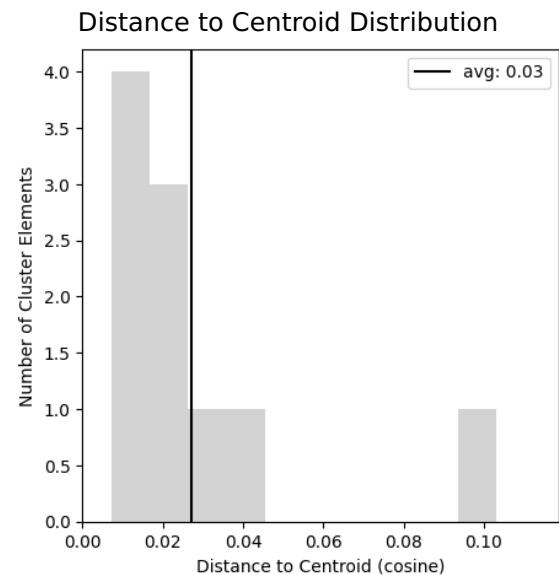
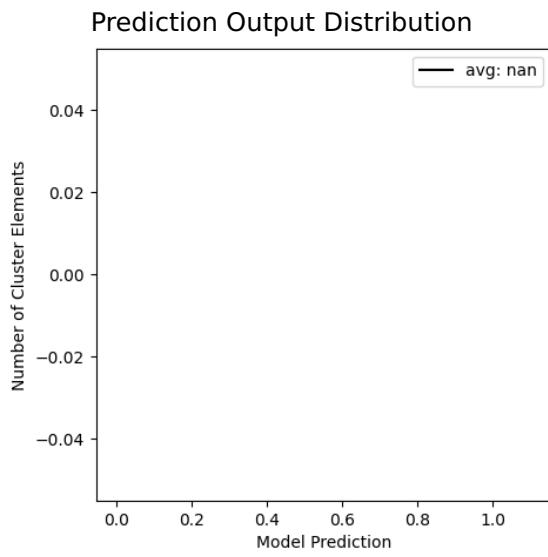
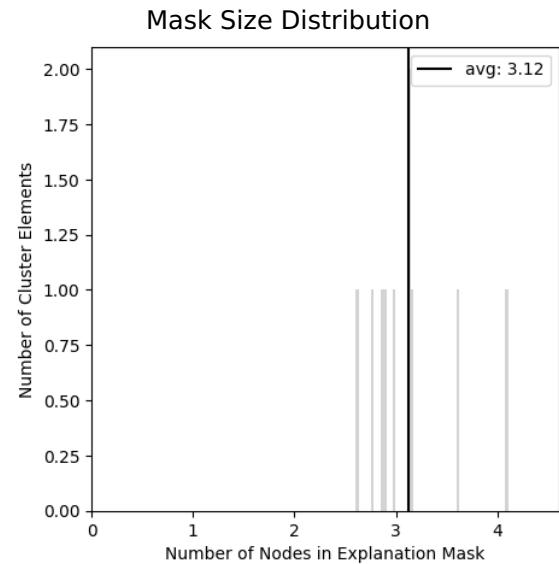
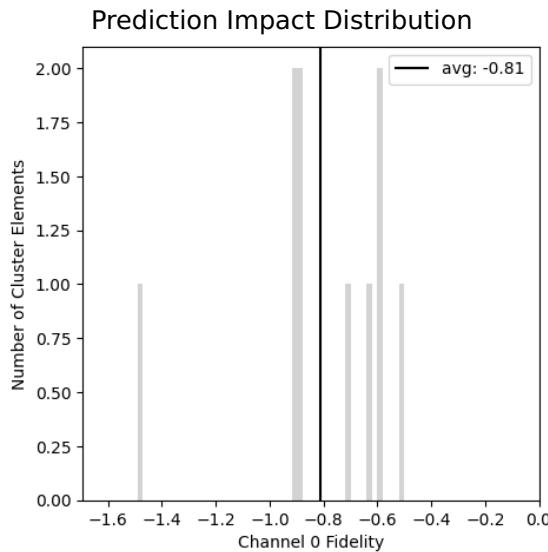
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	10
Channel Index	0.0 (0.0)

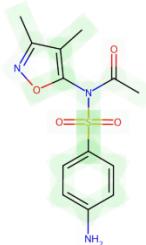
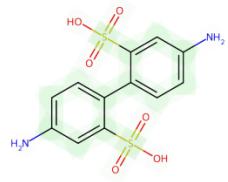
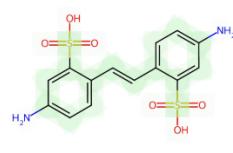
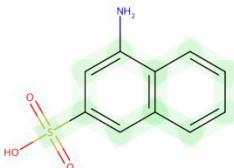
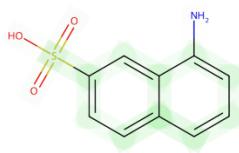
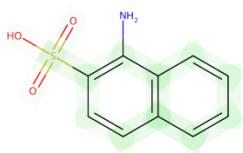
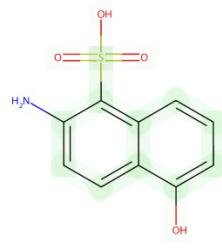
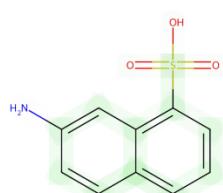
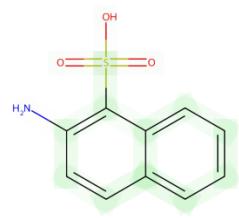
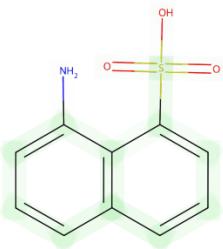
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



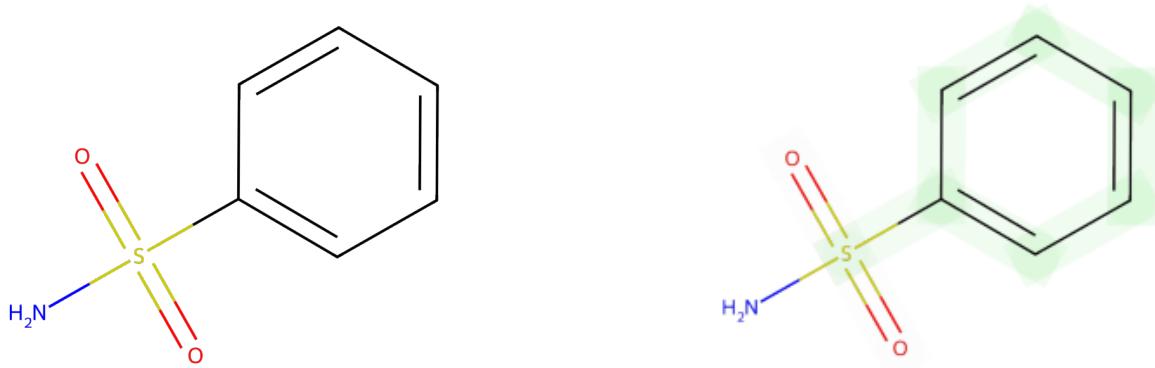
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a molecule containing a sulfonamide group (N-S(=O)(=O)-) attached to a benzene ring. Sulfonamide groups are known for their polar nature due to the presence of sulfur dioxide (S(=O)2) moieties. The polarity of this group enhances hydrogen bonding with water molecules, thereby increasing solubility. Additionally, the benzene ring, which is non-polar, can decrease solubility due to its tendency to resist mixing with water. However, in this specific structure, the influence of the sulfonamide group towards water solubility appears to outweigh that of the benzene ring.

**Hypothesis:** The presence of the sulfonamide group in the molecule with a benzene ring tends to increase the molecule's water solubility. The increased solubility is likely due to enhanced hydrogen bonding capabilities of the sulfonamide group that overcomes the hydrophobic effect of the benzene ring. Thus, we posit that the sulfonamide group's polarity and ability to form hydrogen bonds with water molecules results in a net increase in water solubility for the molecule.

# Cluster #26 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 26, from importance channel 0 (*negative*), represents a motif consisting of 3.3 ( $\pm 0.1$ ) nodes. The concept is generally associated with an impact of -1.2 ( $\pm 0.3$ ) on the prediction outcome.

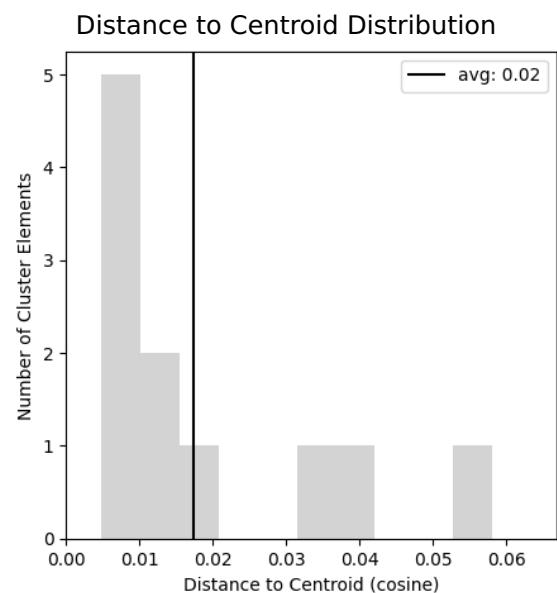
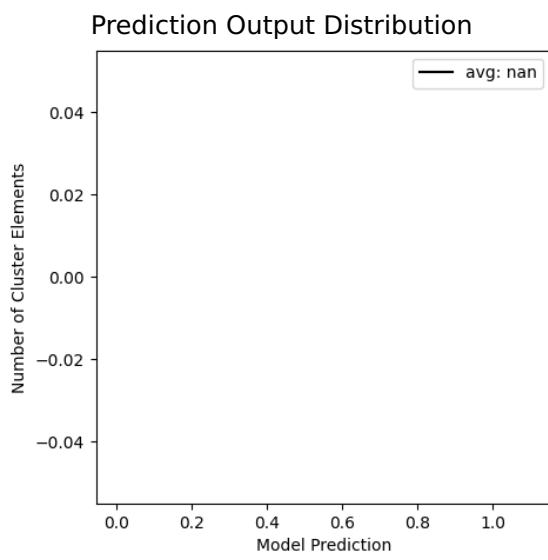
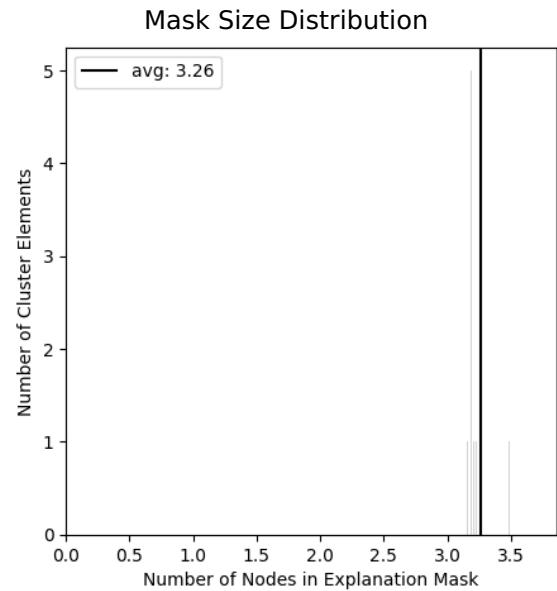
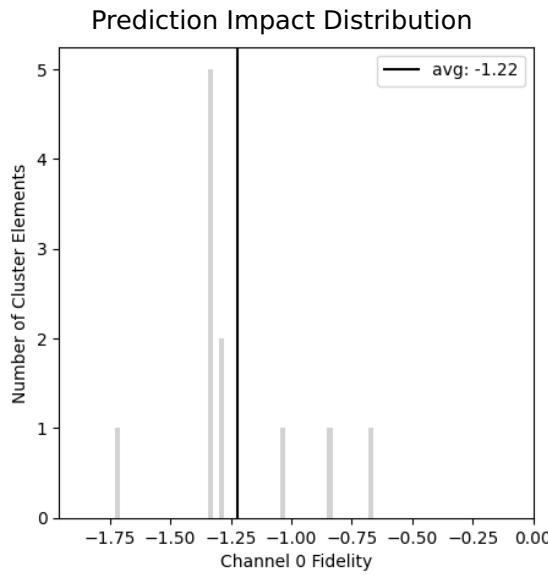
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	11
Channel Index	0.0 (0.0)

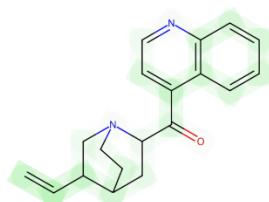
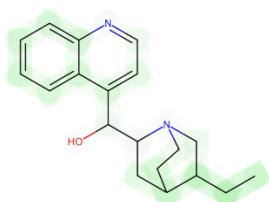
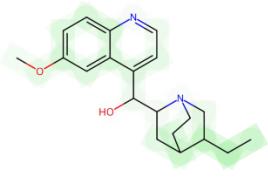
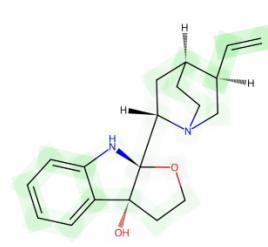
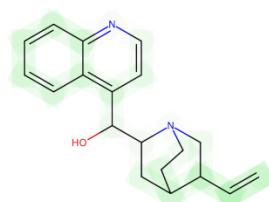
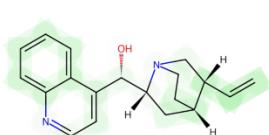
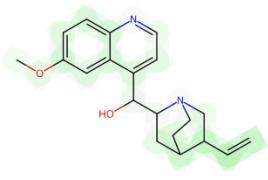
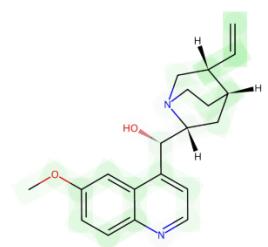
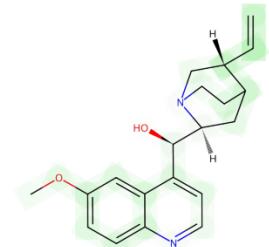
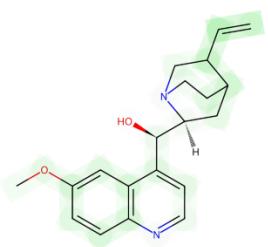
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



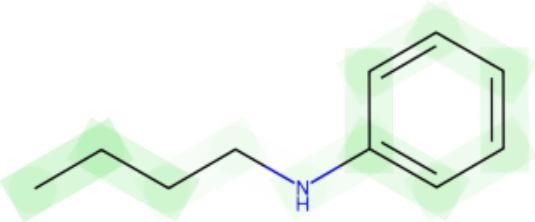
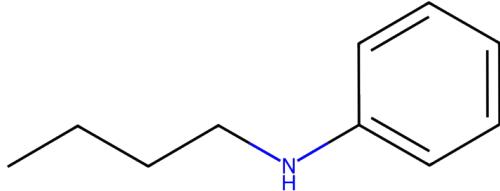
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES representation describes a nonpolar alkyl chain (C-C-C-C-) attached to an aromatic ring via a nitrogen atom. Nonpolar chains tend to decrease water solubility due to their hydrophobic nature, meaning they do not form favorable interactions with water molecules. The aromatic ring can contribute to this effect but may also engage in  $\pi$ - $\pi$  interactions or hydrogen bonding depending on its substitution pattern, which can sometimes increase solubility. However, the connection through nitrogen, a more polar atom, might offer some polar character to the molecule. Since the influence is negative, it seems the nonpolar effects outweigh any polar interactions in this case.

**Hypothesis:** Molecules with an alkyl chain attached to an aromatic ring via a nitrogen atom are likely to have reduced water solubility. The nonpolar alkyl chain reduces solubility by preventing favorable interactions with water molecules. Concurrently, any increase in solubility afforded by the aromatic ring and nitrogen atom is insufficient to overcome the hydrophobic influence of the alkyl chain.

# Cluster #27 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 27, from importance channel 0 (*negative*), represents a motif consisting of 3.0 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of -1.2 ( $\pm 0.4$ ) on the prediction outcome.

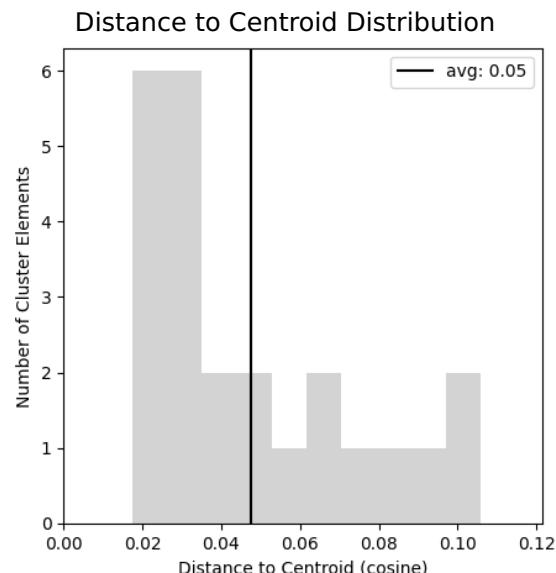
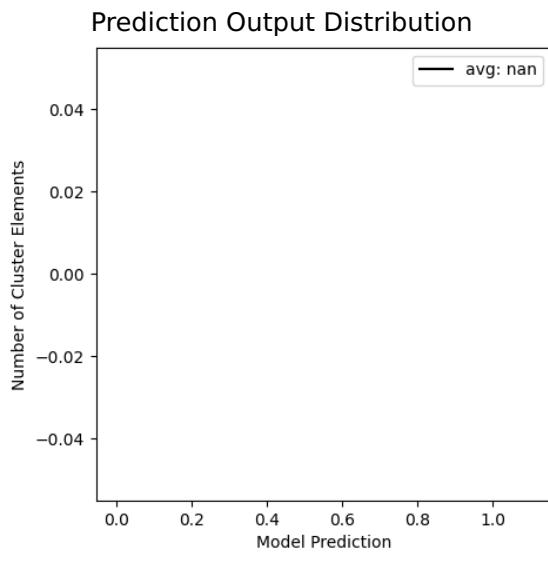
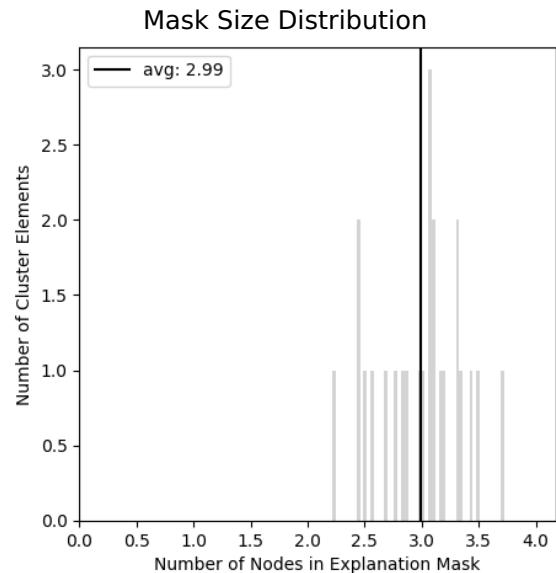
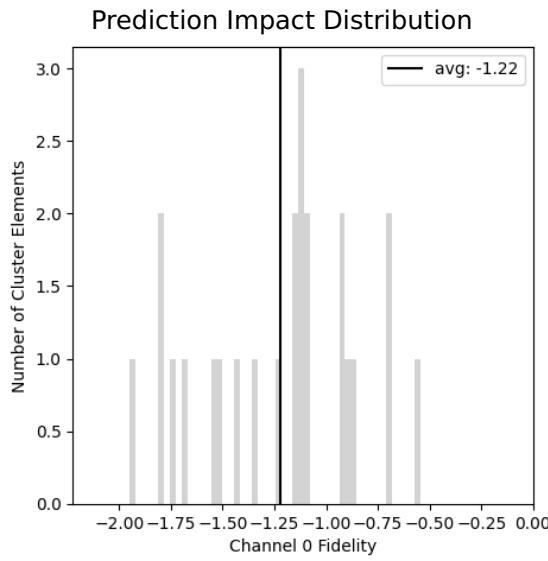
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	24
Channel Index	0.0 (0.0)

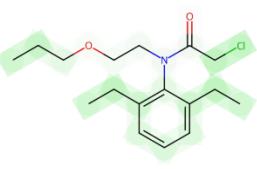
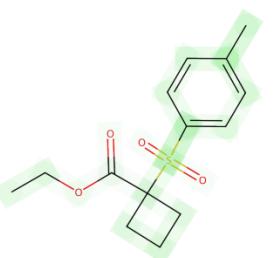
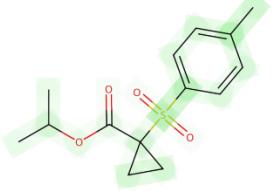
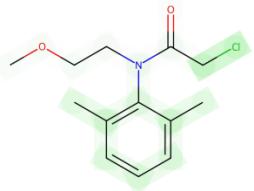
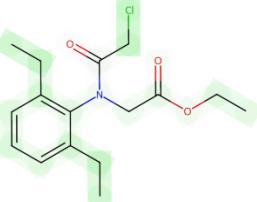
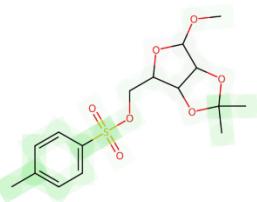
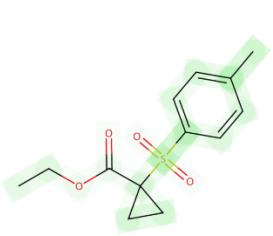
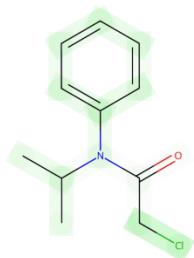
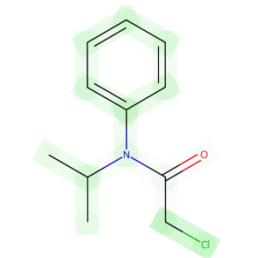
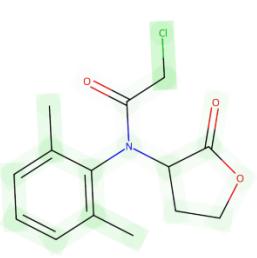
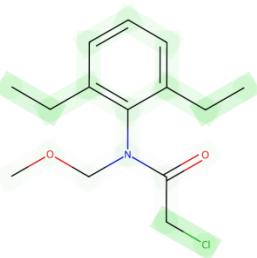
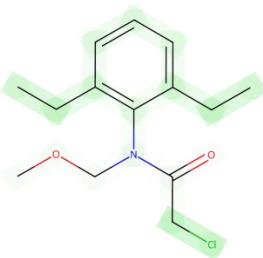
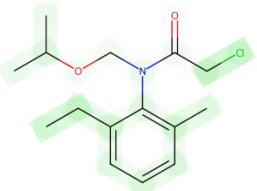
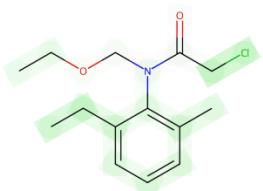
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



# Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The presence of an oxygen atom directly bonded to a carbon, and a sulfur atom in a cyclic structure as indicated by the SMILES "O-C-C1(-S)-C-C-1", suggests potential for hydrogen bonding interactions with water molecules due to the polarity of these atoms. Cyclic structures often introduce rigidity, which could affect how the molecule interacts with the solvent. In water solubility, polarity and the molecule's ability to form hydrogen bonds are critical factors, as they enable the molecule to better interact with the polar water molecules, thus increasing solubility. However, the negative influence suggests that the specific arrangement of the atoms or the rigidity introduced by the ring structure hinders the solubility of the molecule in water.

**Hypothesis:** Molecules with the substructure "O-C-C1(-S)-C-C-1" are likely to have decreased solubility in water. The presence of oxygen and sulfur atoms suggests that the molecule has the potential for hydrogen bonding, which usually increases water solubility, but the specific arrangement of these elements within a cyclic structure may restrict the molecule's interaction with water, thus reducing solubility.

# Cluster #28 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 28, from importance channel 0 (*negative*), represents a motif consisting of 2.8 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of -1.4 ( $\pm 0.4$ ) on the prediction outcome.

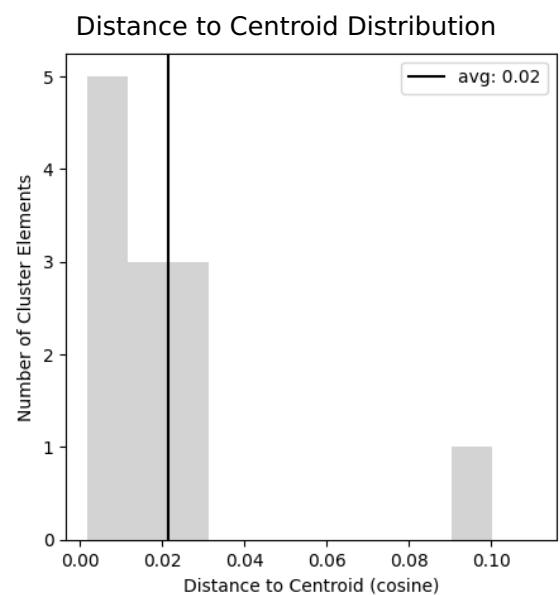
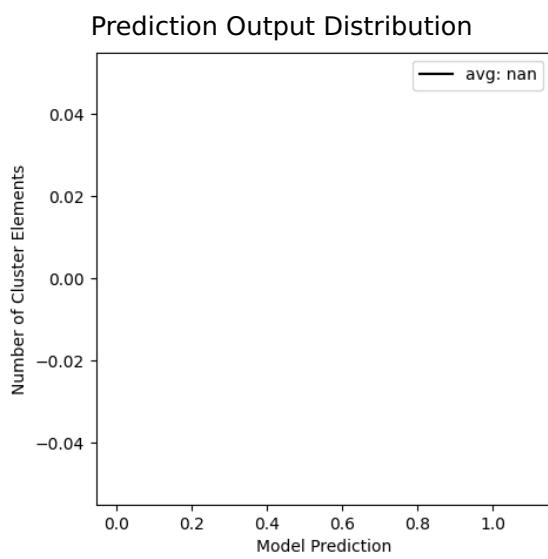
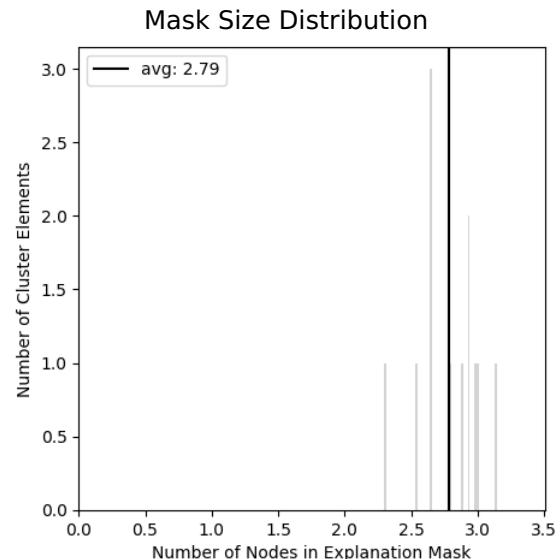
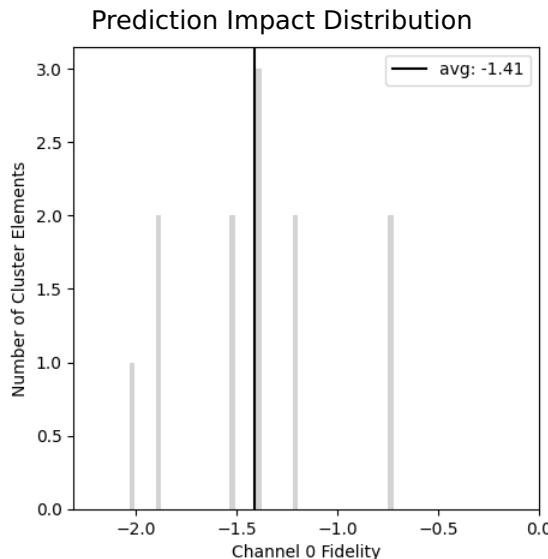
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	12
Channel Index	0.0 (0.0)

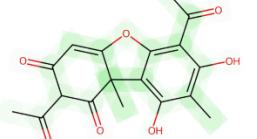
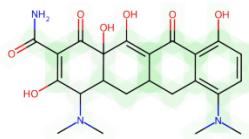
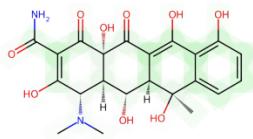
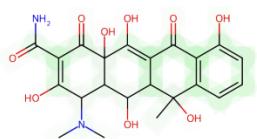
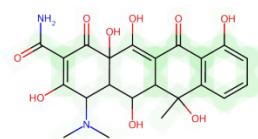
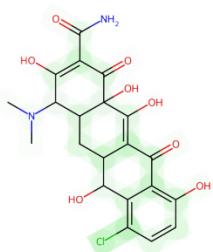
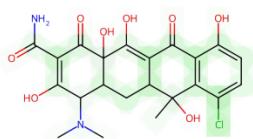
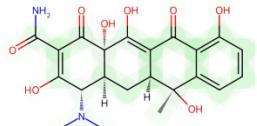
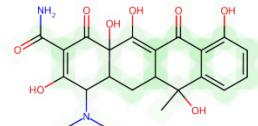
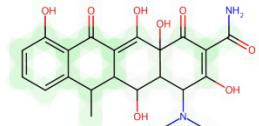
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



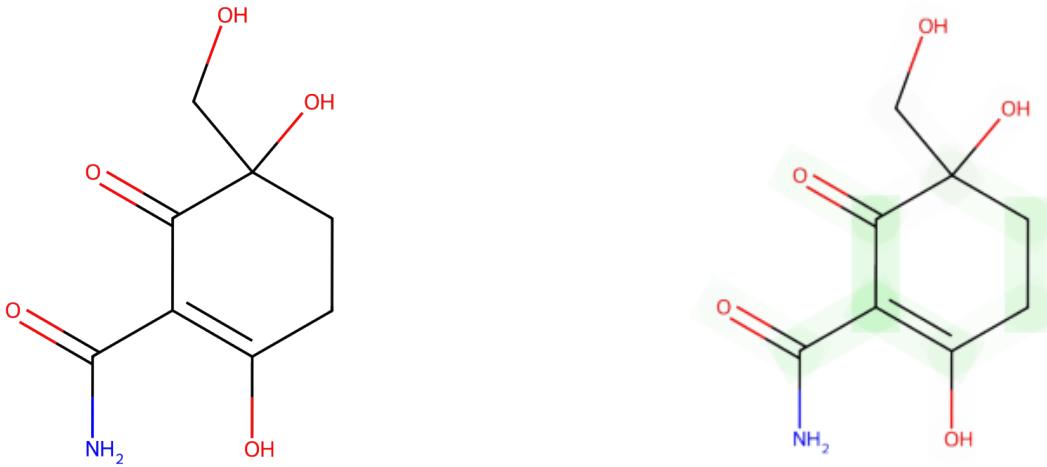
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES representation indicates a molecule featuring an amide group (N-C(=O)) and a benzene ring with two hydroxyl groups (-C(-O) and -C-O) and another ketone group (=O) attached to it. Amide groups can form hydrogen bonds with water molecules due to the presence of both a hydrogen atom and a lone pair of electrons on the nitrogen atom, which could enhance water solubility. However, the presence of the benzene ring, which is hydrophobic, typically reduces water solubility. The hydroxyl groups increase the molecule's polarity, enhancing solubility, but this effect can be diminished by the benzene ring's hydrophobic character. Additionally, the ketone group adds to polarity and enhances water solubility through potential hydrogen bonding. The influence score of -1.41 suggests that overall, this structure may reduce water solubility when attached to a molecule, indicating the hydrophobic properties, possibly due to the benzene ring, may have a greater effect than the polar groups in this case.

**Hypothesis:** The SMILES substructure "N-C(=O)-C1=C(-O)-C-C-C(-O)(-C-O)-C-1=O" is associated with a decrease in water solubility. This is likely due to the hydrophobic effects of the benzene ring overriding the potential hydrogen bonding capabilities of the amide and hydroxyl groups, as well as the polarity of the ketone group.

# Cluster #29 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 29, from importance channel 0 (*negative*), represents a motif consisting of 2.6 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of -2.7 ( $\pm 1.2$ ) on the prediction outcome.

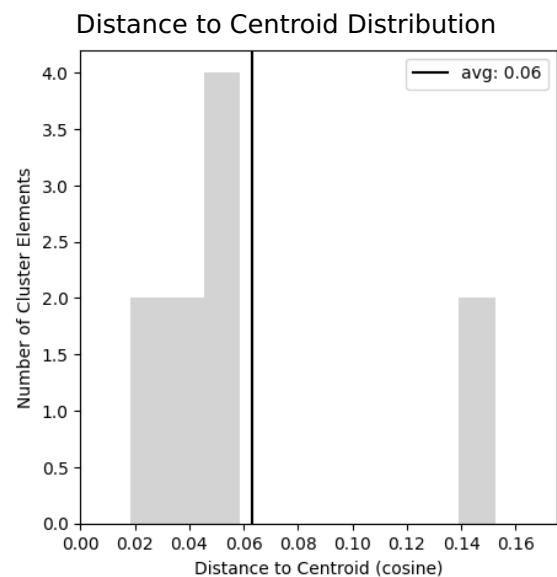
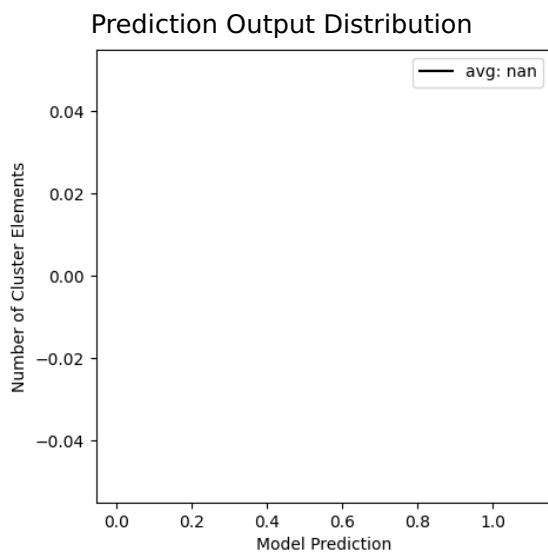
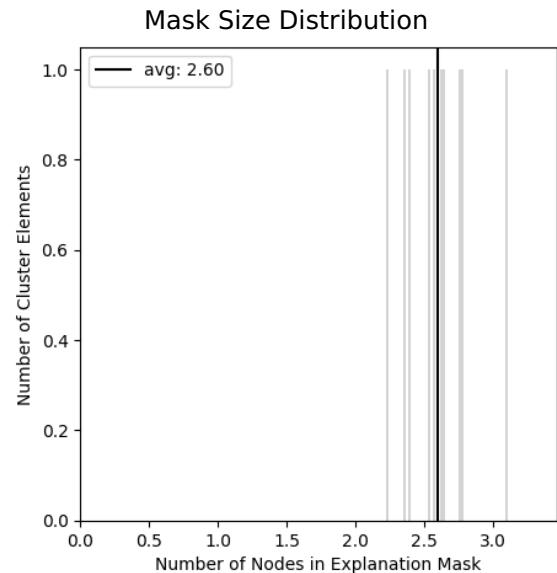
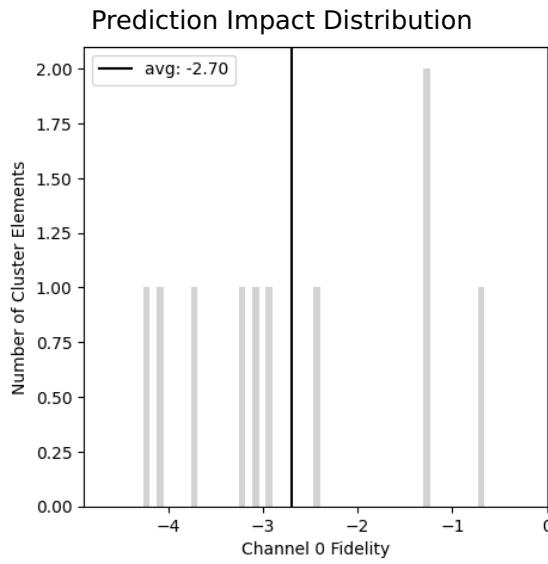
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	10
Channel Index	0.0 (0.0)

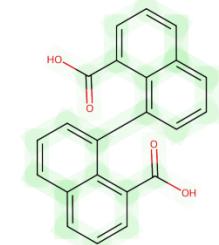
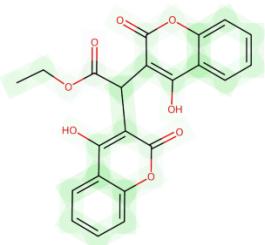
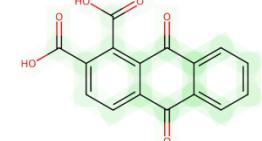
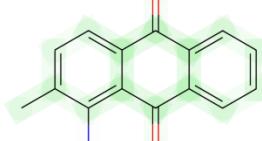
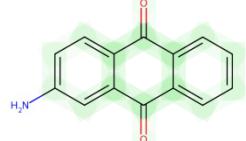
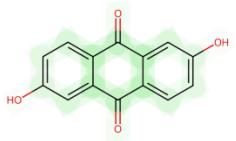
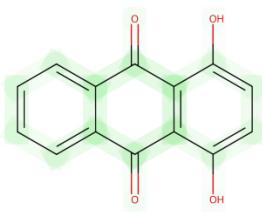
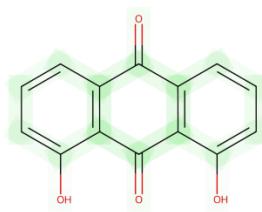
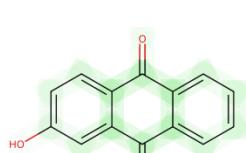
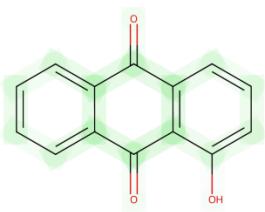
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



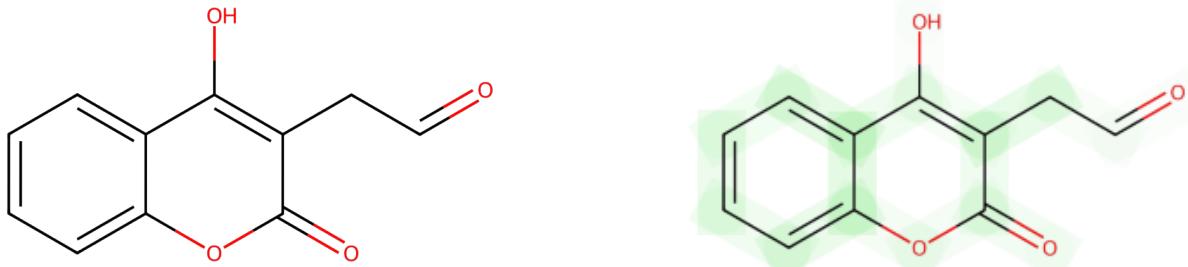
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES representation "O=C-C-c1:c(-O):c2:c:c:c:c:2:o:c:1=O" depicts a structure with several oxygen-containing functional groups, including hydroxyl (-OH) groups and carbonyl (C=O) groups. These functional groups are known to form hydrogen bonds with water, which increases a molecule's solubility in water. The ring structure also contains conjugated double bonds, which could add some level of rigidity to the molecule; however, in this case, the polar functional groups take precedence in affecting solubility.

**Hypothesis:** The substructure shown by the given SMILES has a negative influence on water solubility due to the presence of multiple oxygen-containing functional groups that are capable of hydrogen bonding with water. The conjugated system might contribute to some rigidity, but overall, the polar functional groups are likely the dominant factor in enhancing the molecule's water solubility through enhanced interactions with water molecules.

# Cluster #30 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 30, from importance channel 0 (*negative*), represents a motif consisting of 2.8 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of -1.7 ( $\pm 0.7$ ) on the prediction outcome.

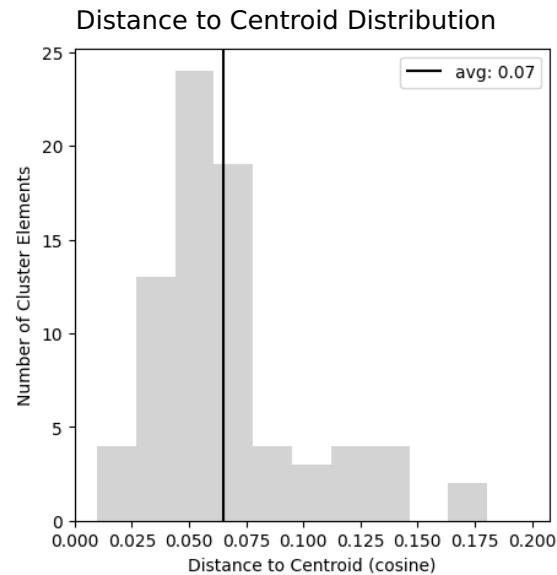
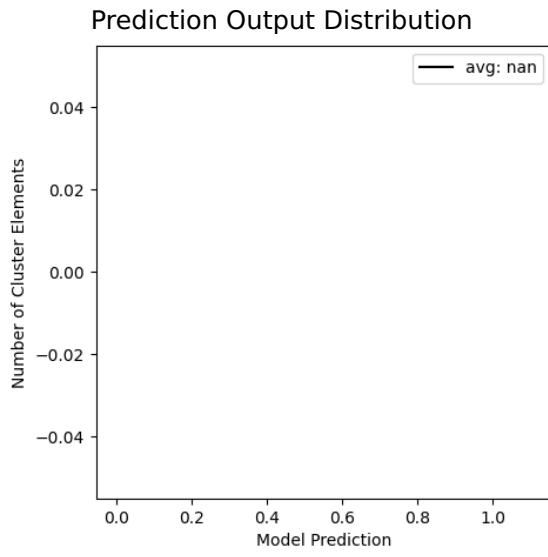
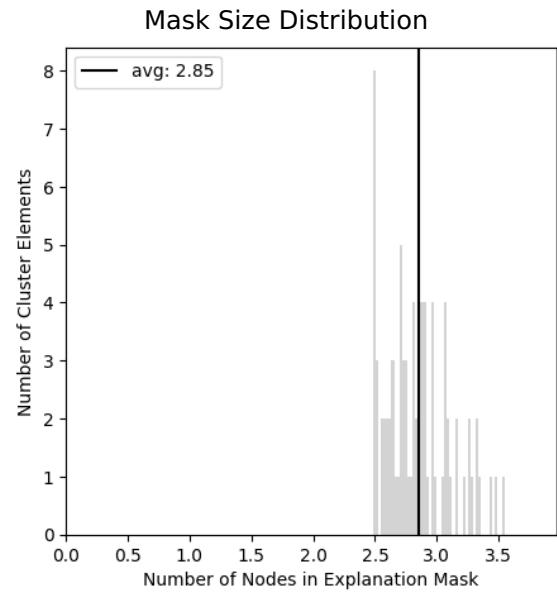
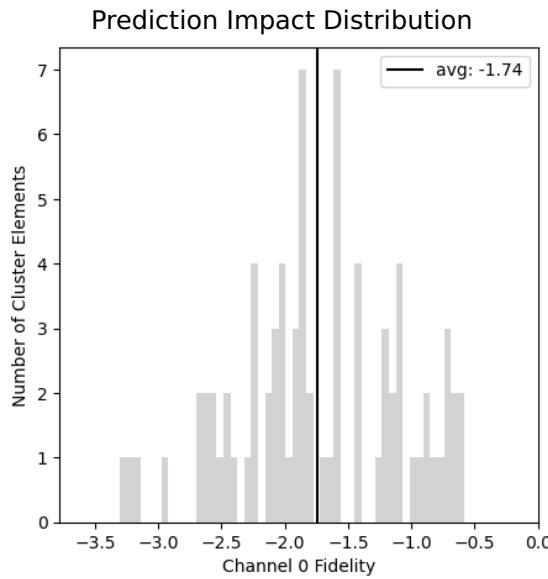
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	77
Channel Index	0.0 (0.0)

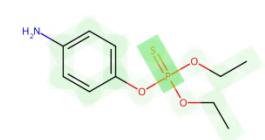
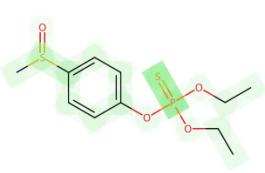
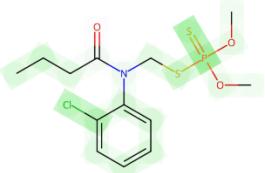
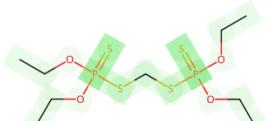
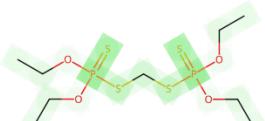
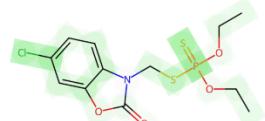
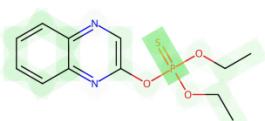
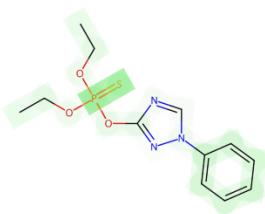
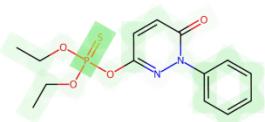
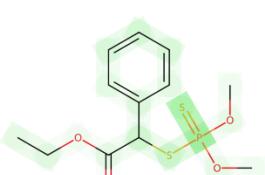
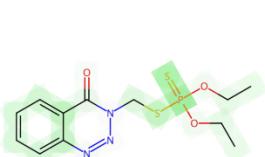
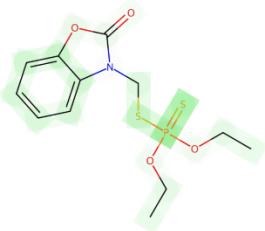
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



# Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The presence of the chlorine atoms in the given structure plays a significant role in decreasing water solubility. Chlorines are relatively large, electron-withdrawing atoms that increase the hydrophobic character of the molecule, reducing its affinity for water. Furthermore, the carbon-chlorine bonds are covalent and do not readily ionize or form hydrogen bonds with water molecules, which further decreases solubility. The carbon-nitrogen-sulfur backbone of the structure may potentially participate in hydrogen bonding due to the heteroatoms, but the overall effect of the chlorines dominates the water solubility profile.

**Hypothesis:** Molecules that contain the "C-N-S-C(-Cl)(-Cl)-Cl" substructure are likely to be poorly soluble in water. The chlorine atoms confer hydrophobicity and limit interactions with water molecules, thus reducing the compound's ability to dissolve in aqueous environments. The influence of the carbon-nitrogen-sulfur segment is overshadowed by the presence of the multiple chlorine atoms.

# Cluster #31 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 31, from importance channel 0 (*negative*), represents a motif consisting of 2.8 ( $\pm 0.6$ ) nodes. The concept is generally associated with an impact of -1.4 ( $\pm 0.6$ ) on the prediction outcome.

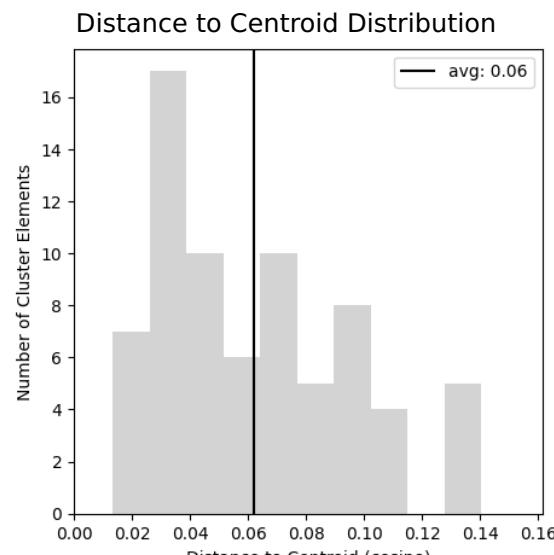
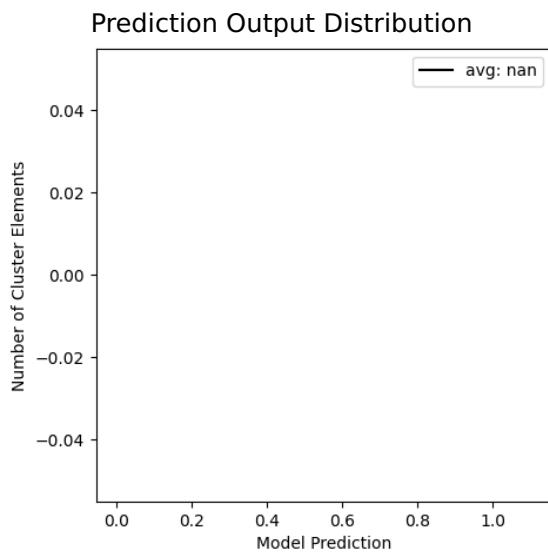
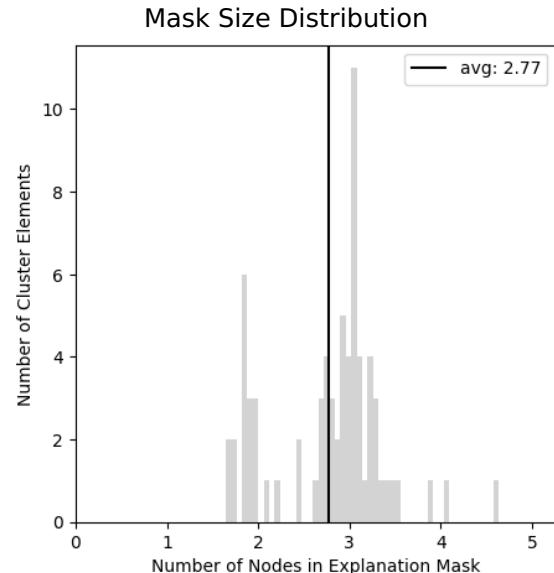
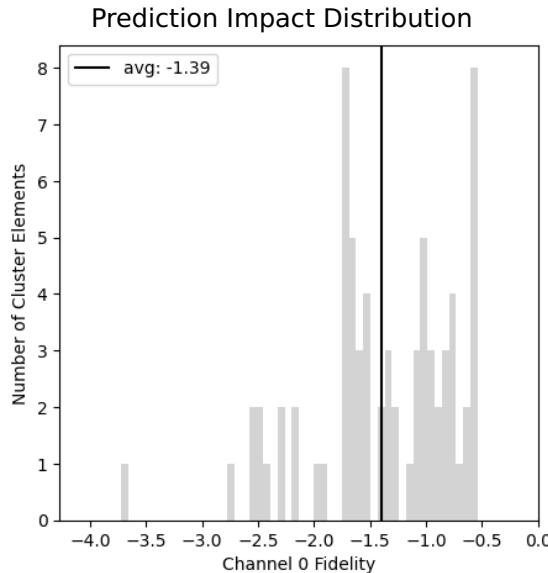
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	72
Channel Index	0.0 (0.0)

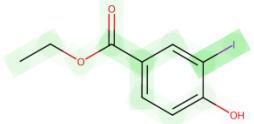
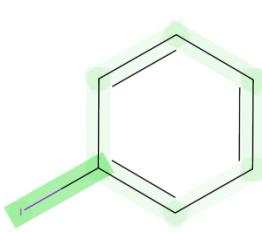
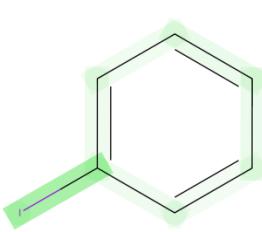
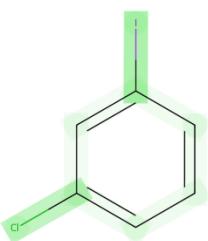
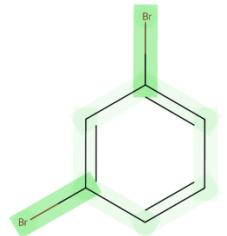
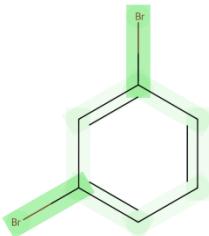
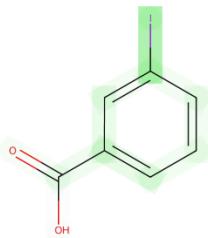
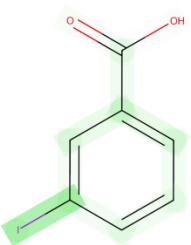
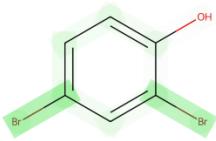
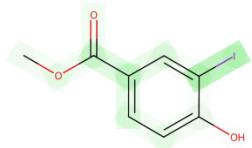
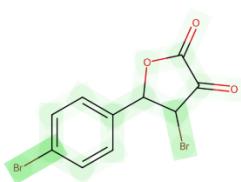
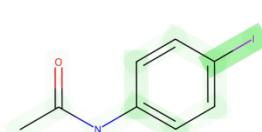
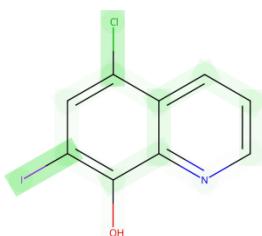
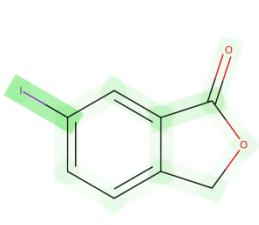
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



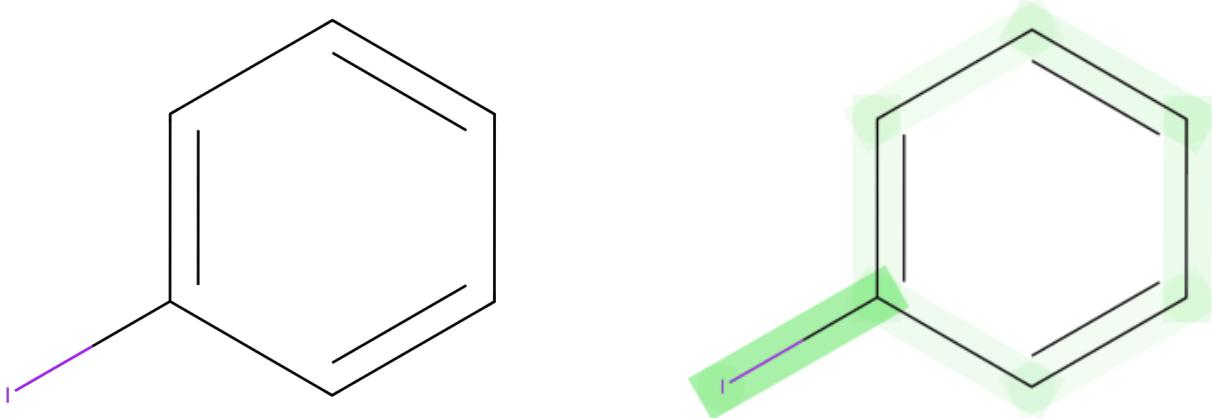
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES representation "I-c1ccccc1" describes a mono-iodinated benzene ring. This structure indicates a large, hydrophobic iodine atom attached to an aromatic benzene ring, which typically reduces water solubility due to the non-polar nature of the iodine and the ring. The iodine atom's larger size and lower electronegativity compared to other possible substituents, like hydroxyl or amino groups, make it less likely to engage in hydrogen bonding with water molecules. Moreover, the benzene ring itself is hydrophobic due to its delocalized  $\pi$ -electron system, which does not favor interactions with the polar water molecules.

**Hypothesis:** The presence of an iodine atom on a benzene ring reduces water solubility. The large, hydrophobic iodine atom decreases the molecule's overall polarity, and the aromatic ring system does not engage effectively with the polar solvent, resulting in a decrease in solubility in water.

# Cluster #32 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 32, from importance channel 0 (*negative*), represents a motif consisting of 2.9 ( $\pm 0.1$ ) nodes. The concept is generally associated with an impact of -0.9 ( $\pm 0.2$ ) on the prediction outcome.

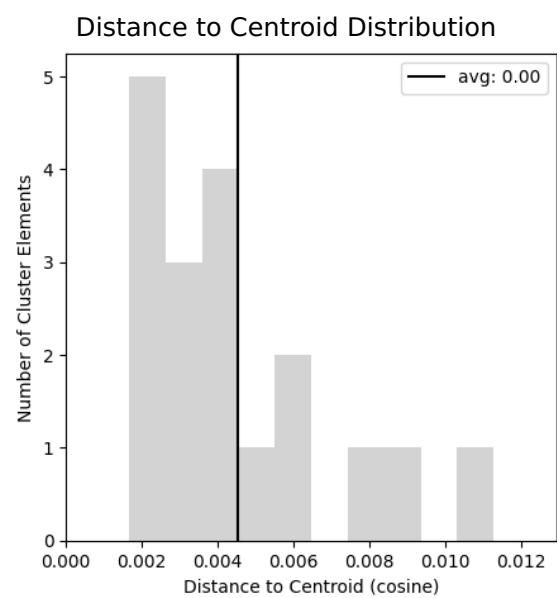
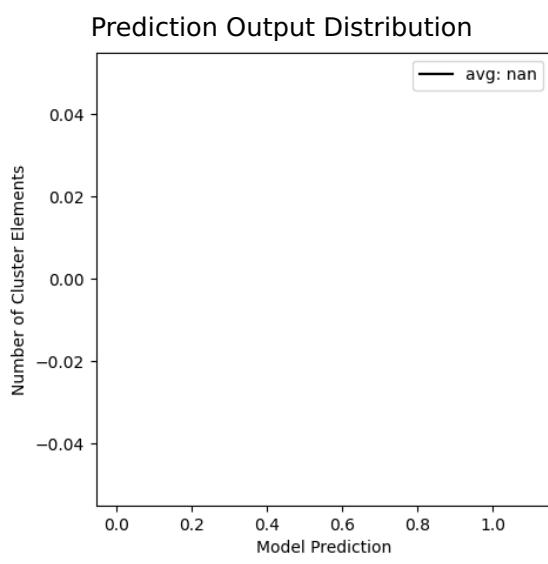
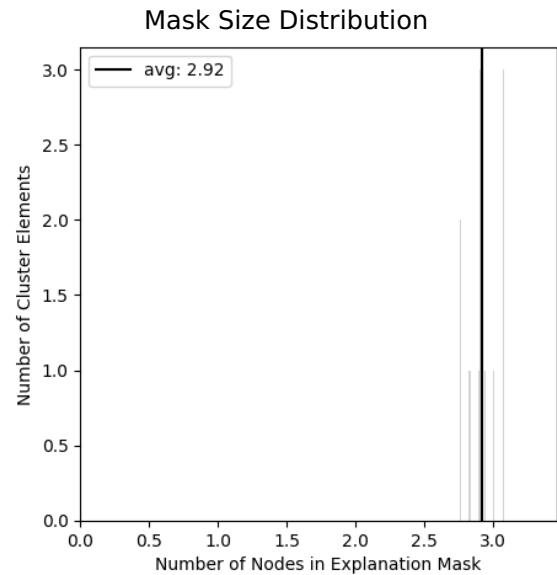
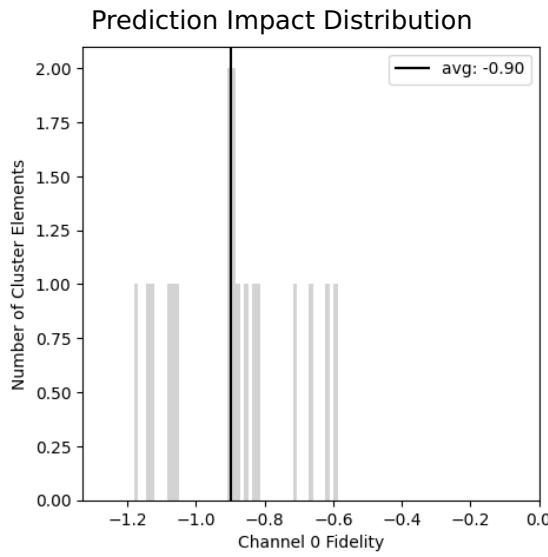
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	18
Channel Index	0.0 (0.0)

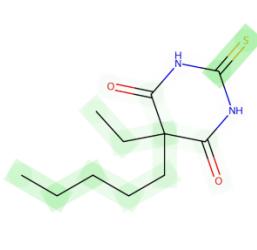
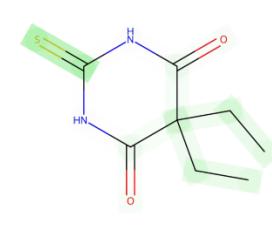
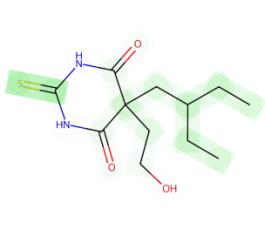
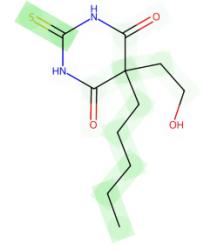
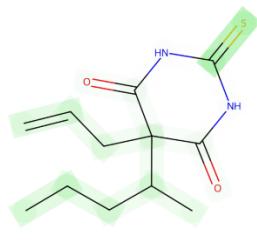
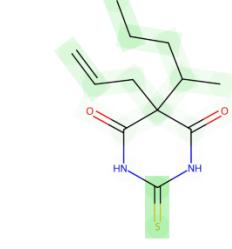
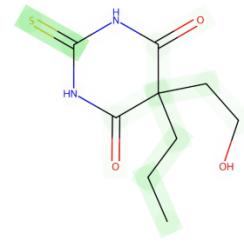
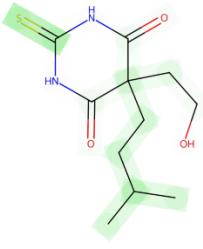
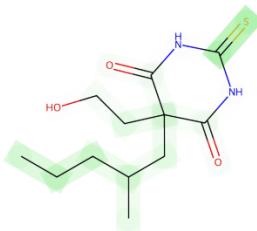
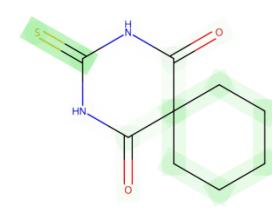
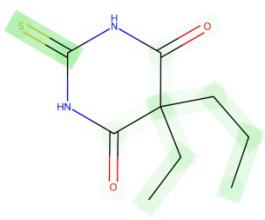
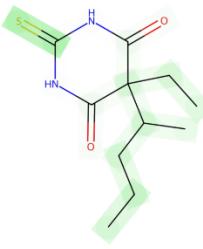
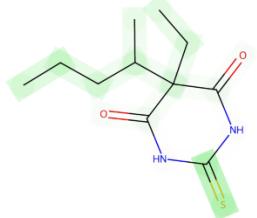
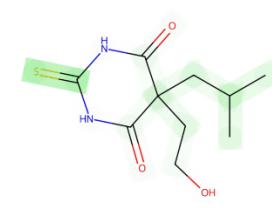
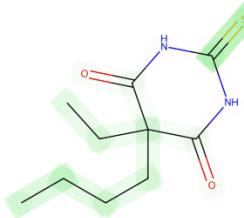
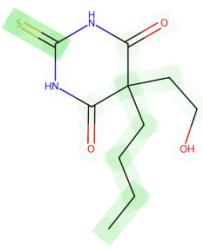
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The substructure " $\text{N}-\text{C}(=\text{S})-\text{NH}_2$ " indicates the presence of a sulfonamide group attached to an amide group. Sulfonamides are generally polar due to the presence of nitrogen and sulfur atoms which can form hydrogen bonds with water molecules, enhancing solubility. However, the presence of the attached amide group, while also polar and capable of hydrogen bonding, might hamper the overall solubility due to its potential to engage in internal hydrogen bonding, which makes the molecule more self-associated rather than associating with water molecules.

**Hypothesis:** The presence of a sulfonamide group attached to an amide group in a molecule reduces its water solubility. The sulfonamide's inherent polarity and ability to hydrogen bond with water is counteracted by the amide group's tendency to form internal hydrogen bonds, leading to a decrease in solubility in water.

# Cluster #33 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 33, from importance channel 0 (*negative*), represents a motif consisting of 2.9 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of -1.0 ( $\pm 0.2$ ) on the prediction outcome.

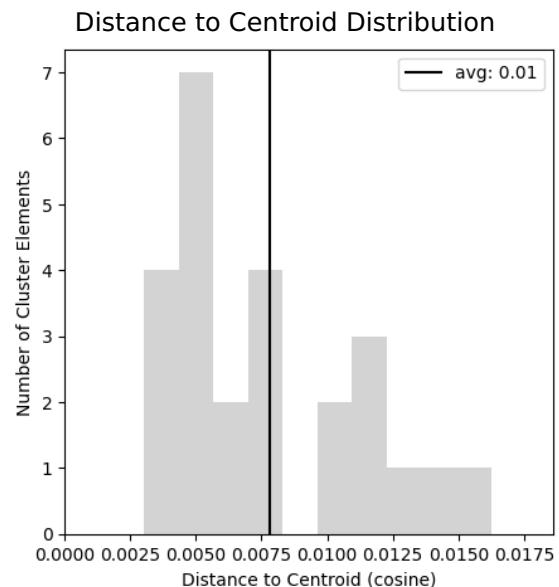
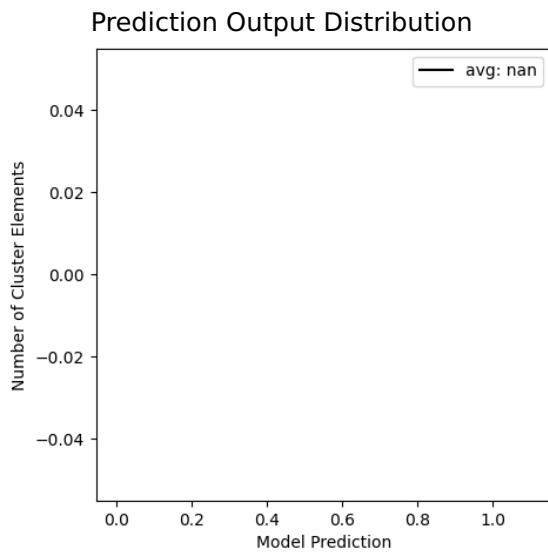
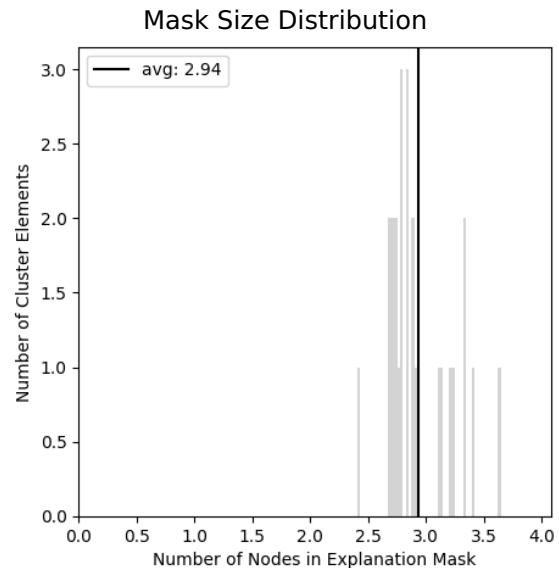
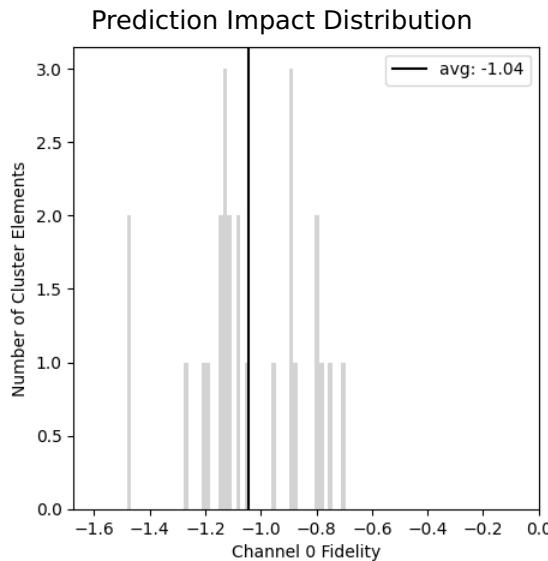
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	25
Channel Index	0.0 (0.0)

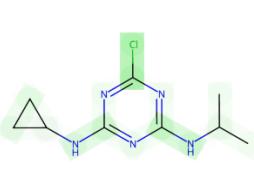
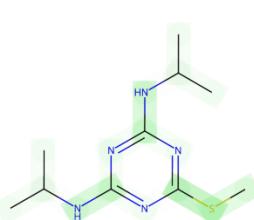
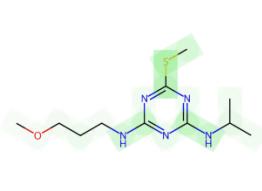
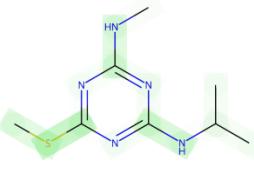
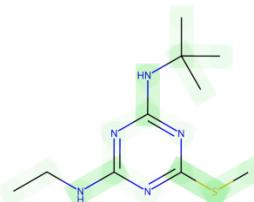
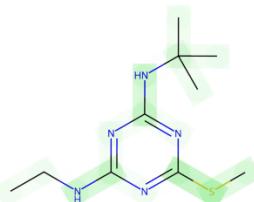
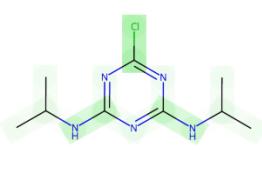
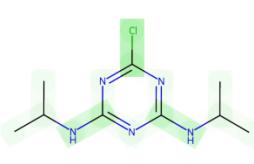
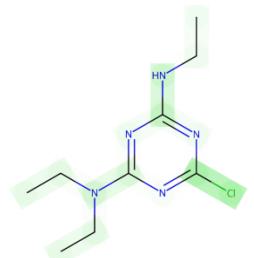
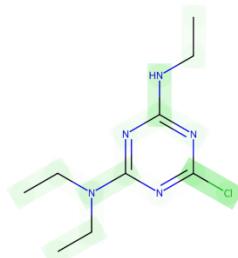
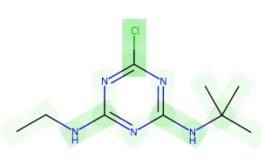
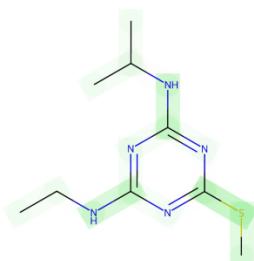
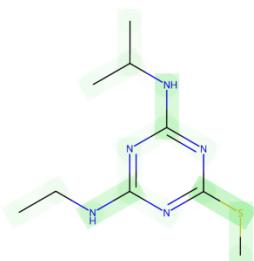
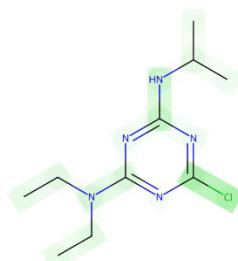
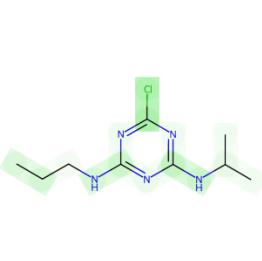
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



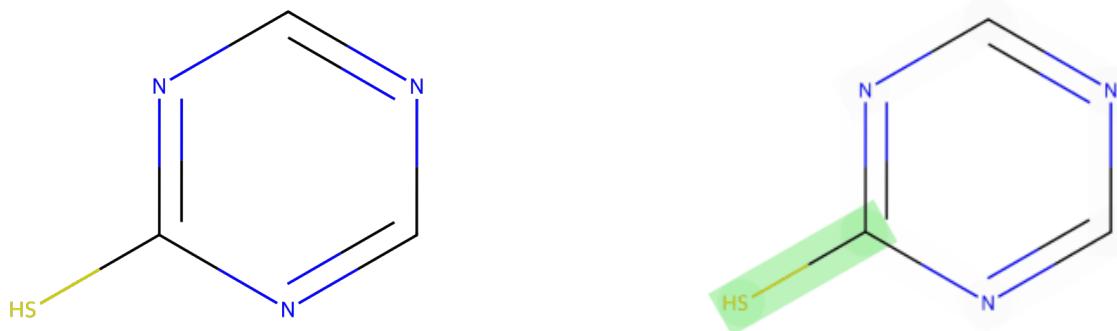
# Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure represented by "S-c1:n:c:n:c:n:1" is a thiophene-like heterocycle containing sulfur and nitrogen. Sulfur atoms are typically more hydrophobic due to their low electronegativity compared to oxygen. However, the presence of adjacent nitrogen atoms within the ring increases the polarity because of nitrogen's higher electronegativity and its ability to engage in hydrogen bonding. The influence of these nitrogen atoms could be insufficient to overcome the overall hydrophobic character contributed by the sulfur atom, thus leading to a decrease in water solubility.

**Hypothesis:** A thiophene-like structure with sulfur and nitrogen atoms within the ring decrease water solubility, likely due to the relative hydrophobic character of sulfur being dominant. Despite the potential for nitrogen to enhance solubility through polarity and hydrogen bonding, its effect might not be strong enough to counteract the hydrophobic influence of the sulfur atom.

# Cluster #34 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 34, from importance channel 0 (*negative*), represents a motif consisting of 4.2 ( $\pm 0.8$ ) nodes. The concept is generally associated with an impact of -4.9 ( $\pm 0.6$ ) on the prediction outcome.

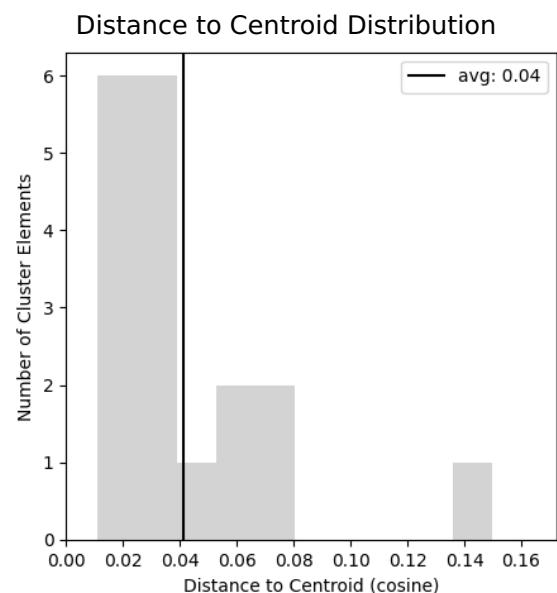
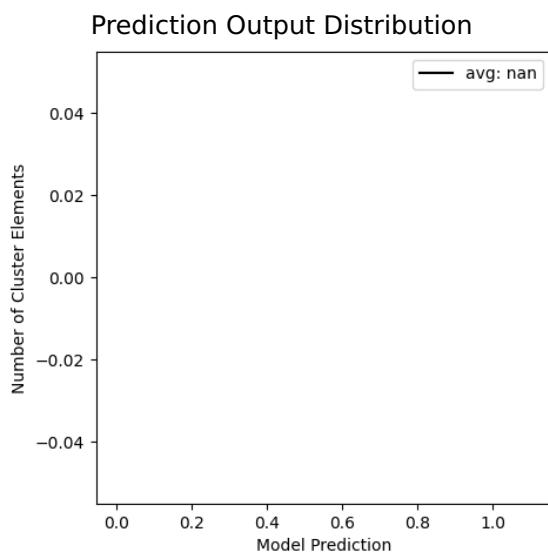
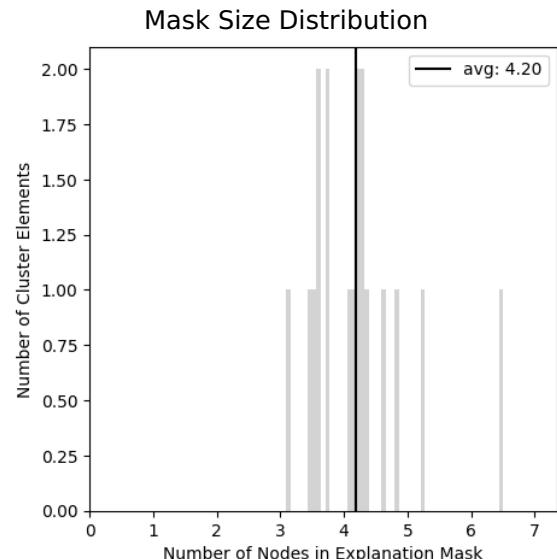
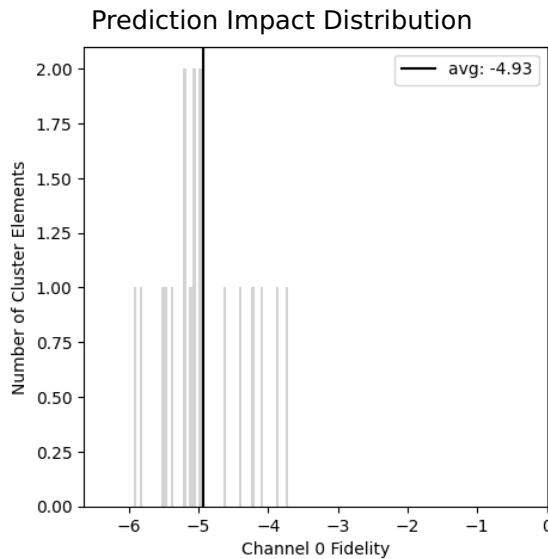
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	18
Channel Index	0.0 (0.0)

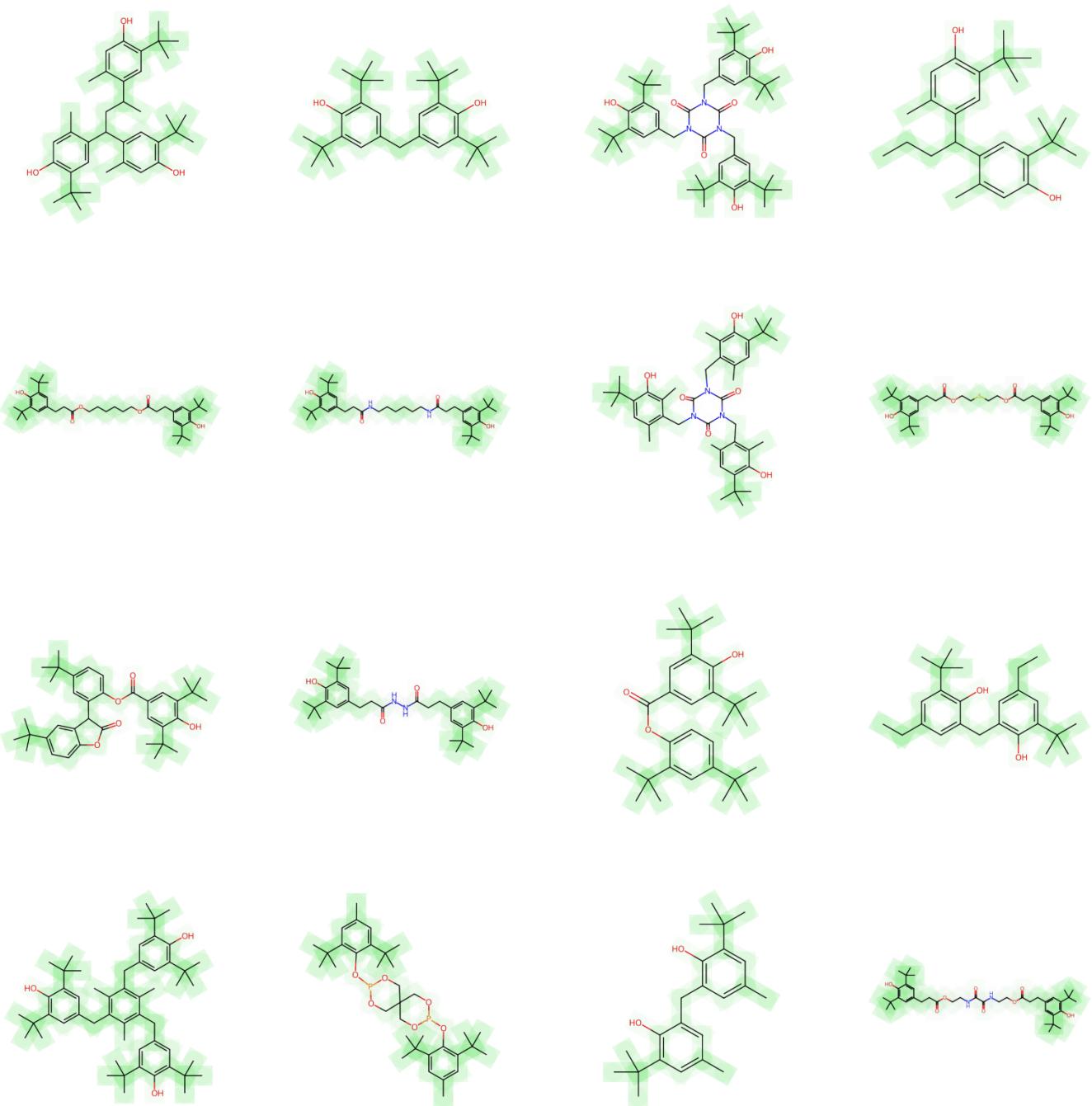
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



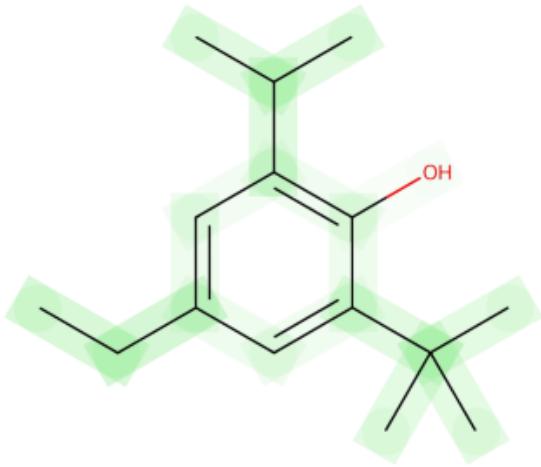
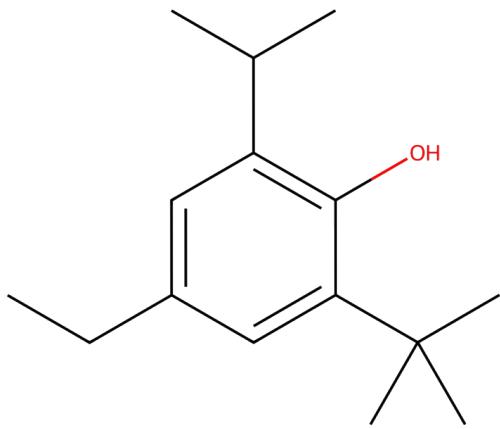
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The presence of the hydrophobic alkyl chains (represented by 'C-C-' and '-C(-C)-C') in the molecular substructure can significantly decrease water solubility as they tend to avoid interaction with water molecules, leading to a hydrophobic effect. Whereas the presence of a hydroxyl group ('-O'), which is a polar functional group, generally increases water solubility due to its ability to form hydrogen bonds with water. In the given structure, the influence of the hydrophobic alkyl chains seems to be stronger than the influence of the single hydroxyl group, leading to an overall decrease in water solubility.

**Hypothesis:** The given molecular fragment with hydrophobic alkyl chains has a negative influence on water solubility. The preponderance of hydrophobic components in the structure suggests they are more determinant in reducing solubility than the solitary hydroxyl group is in enhancing it, which leads to poor solubility in water. If a molecule's structure has more hydrophobic than hydrophilic groups, it likely decreases water solubility.

# Cluster #35 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 35, from importance channel 0 (*negative*), represents a motif consisting of 3.1 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of -4.3 ( $\pm 0.8$ ) on the prediction outcome.

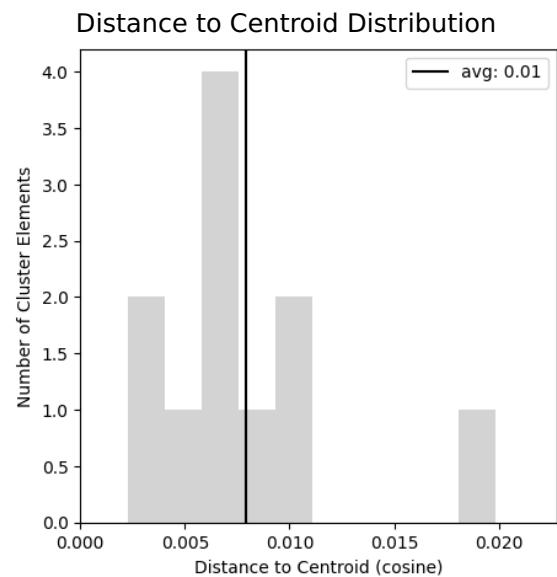
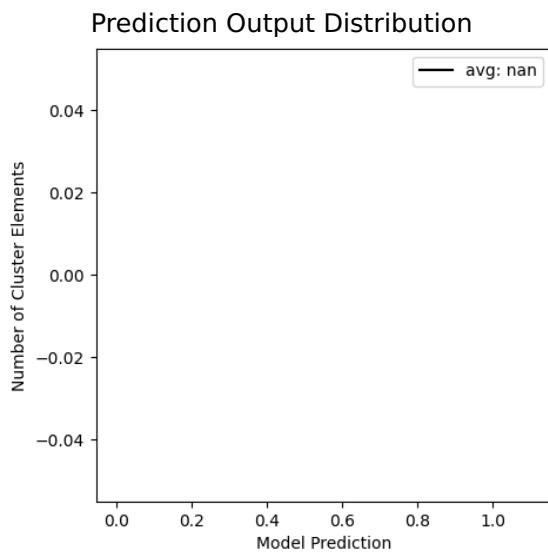
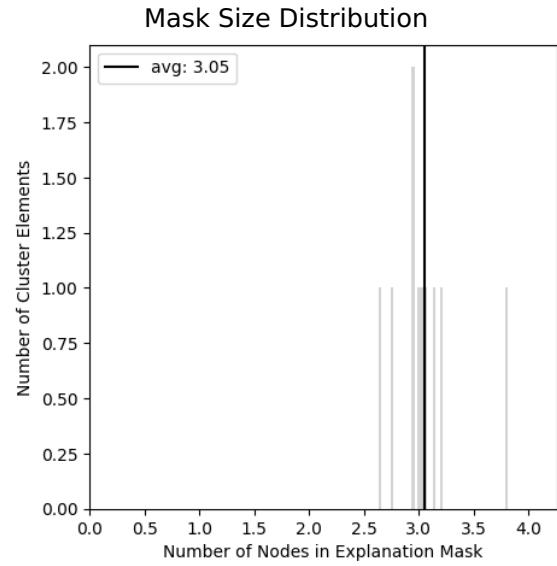
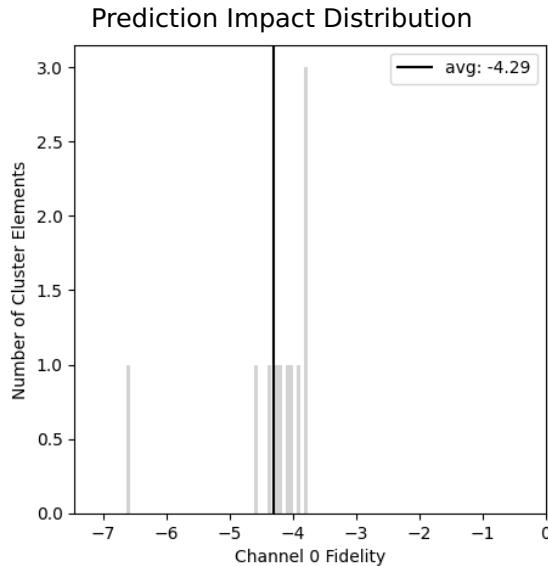
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	11
Channel Index	0.0 (0.0)

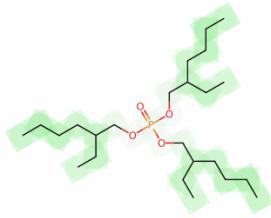
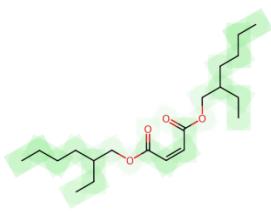
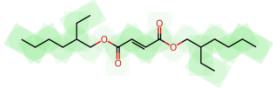
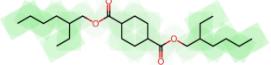
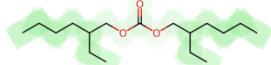
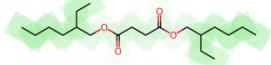
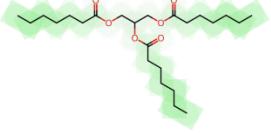
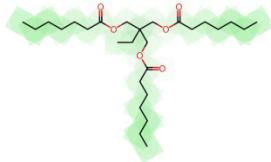
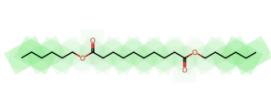
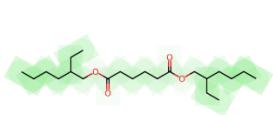
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



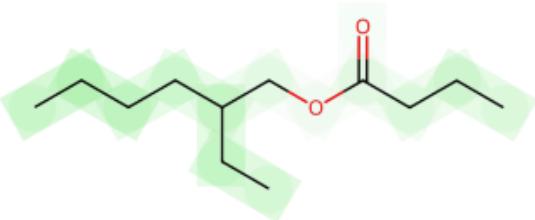
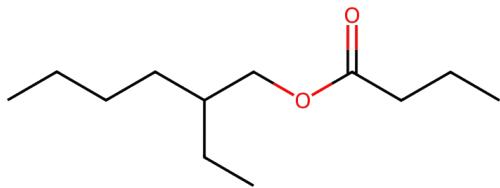
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES representation "C-C-C-C-C(-C-C)-C-O-C(=O)-C-C-C" indicates the presence of a hydrophobic alkyl chain (the repeated "C-C" units) and a hydrophilic ester group ("-O-C(=O)-"). The long alkyl chain can decrease water solubility due to its nonpolar characteristics, which do not interact favorably with polar water molecules. On the other hand, the ester group can interact with water through dipole-dipole interactions and potential hydrogen bonding, increasing solubility. The influence value of -4.29 indicates that the overall effect of this substructure may reduce water solubility, suggesting the hydrophobic interactions of the alkyl chain predominate over the hydrophilic interactions of the ester group.

**Hypothesis:** The molecule with the substructure "C-C-C-C-C(-C-C)-C-O-C(=O)-C-C-C" is hypothesized to have reduced water solubility. This is likely due to the dominating effect of the long hydrophobic alkyl chain impeding solvation by water molecules, despite the presence of a hydrophilic ester group that slightly mitigates this influence.

# Cluster #36 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 36, from importance channel 0 (*negative*), represents a motif consisting of 4.9 ( $\pm 1.4$ ) nodes. The concept is generally associated with an impact of -4.8 ( $\pm 0.7$ ) on the prediction outcome.

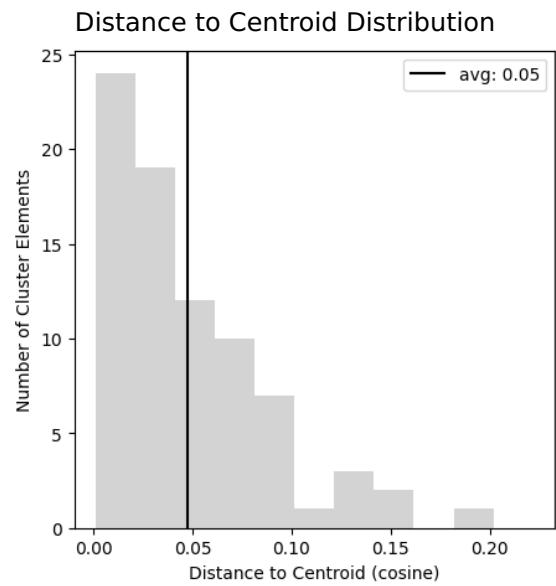
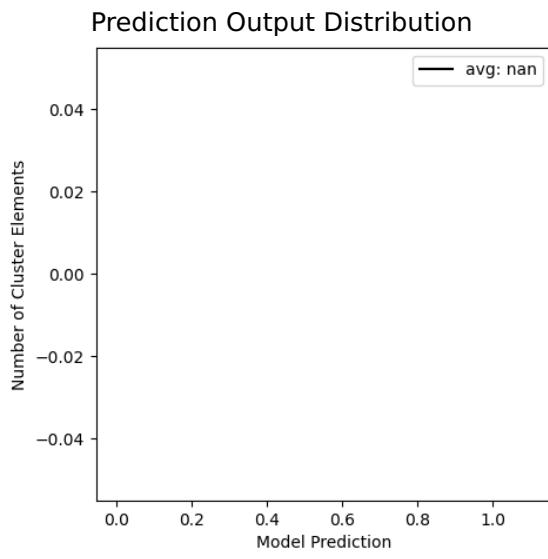
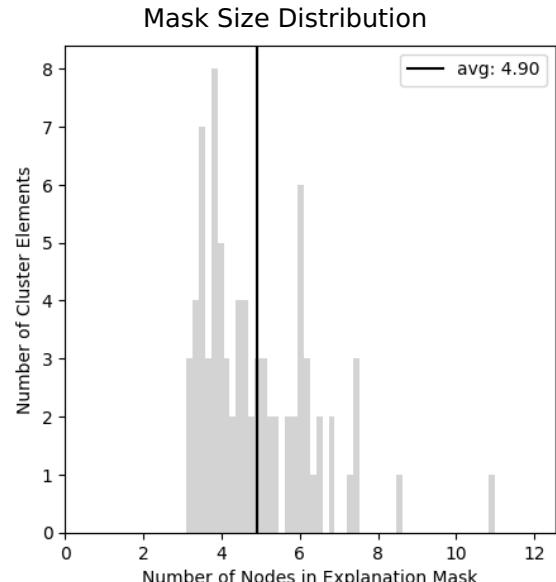
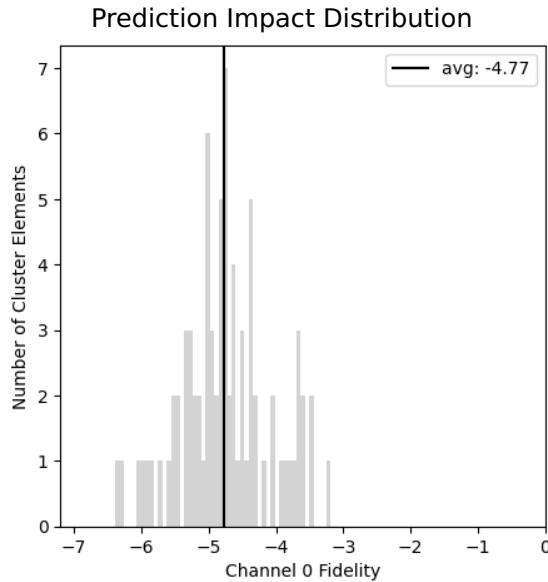
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	79
Channel Index	0.0 (0.0)

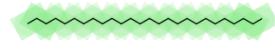
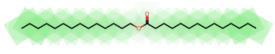
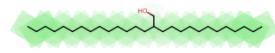
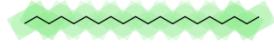
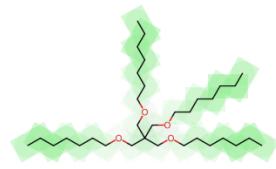
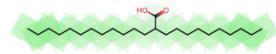
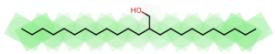
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



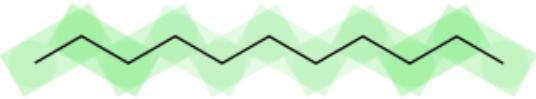
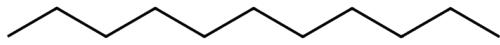
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES string denotes a chain of eleven carbon atoms connected by single bonds, representing a linear alkane hydrocarbon. The solubility of such hydrocarbons in water is very low due to their nonpolar nature, as water is a polar solvent and "like dissolves like". In other words, nonpolar substances do not interact favorably with polar solvents like water because there is little to no affinity between the molecules. The influence value of -4.77 suggests a significant decrease in water solubility, which can be attributed to the increase in nonpolar character as the carbon chain length increases. This decreasing solubility trend is consistent with the understanding that longer alkane chains have more pronounced hydrophobic characteristics, thus resisting solvation by water.

**Hypothesis:** Molecules with a substructure of a long linear alkane chain have a decreased water solubility. This is likely due to the hydrophobic nature of the carbon-hydrogen backbone of alkanes which does not interact favorably with the polar water molecules, leading to poor solubility. The influence value of -4.77 reflects this strong hydrophobic character in the context of water solubility.

# Cluster #37 - negative

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 37, from importance channel 0 (*negative*), represents a motif consisting of 5.2 ( $\pm 1.0$ ) nodes. The concept is generally associated with an impact of -5.8 ( $\pm 0.6$ ) on the prediction outcome.

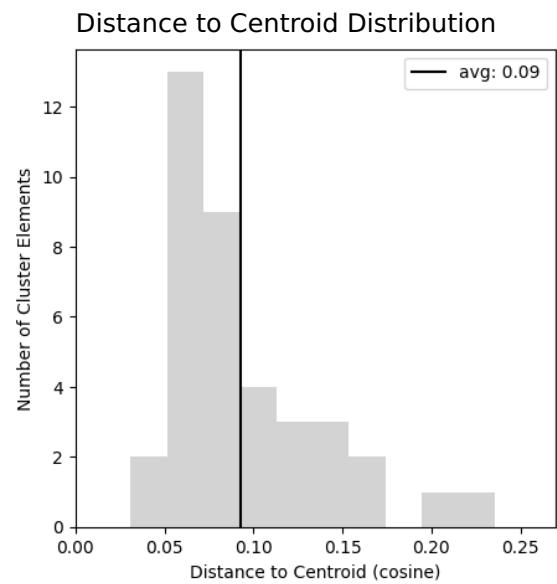
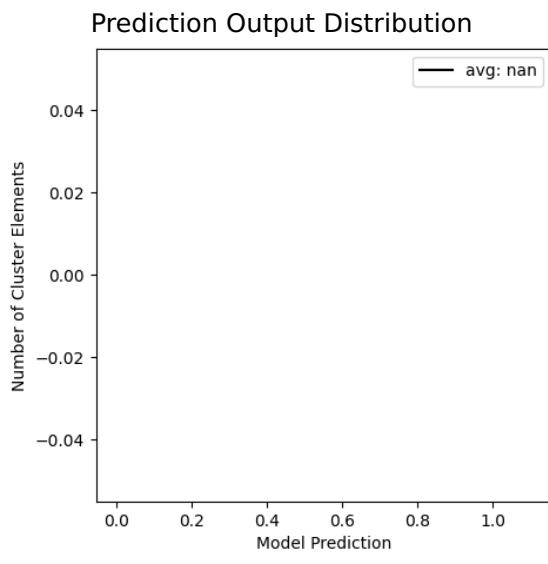
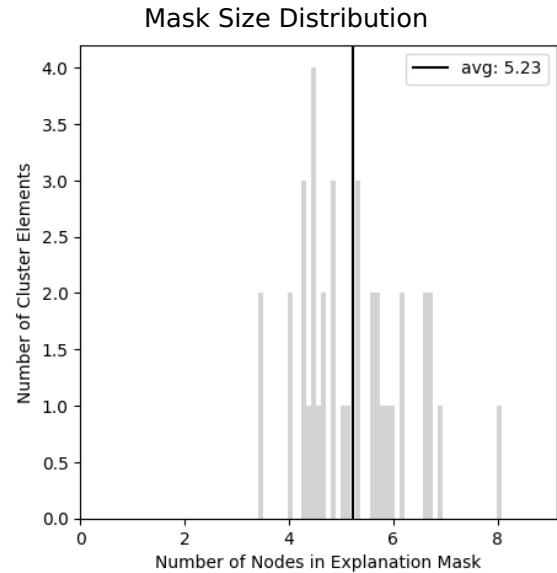
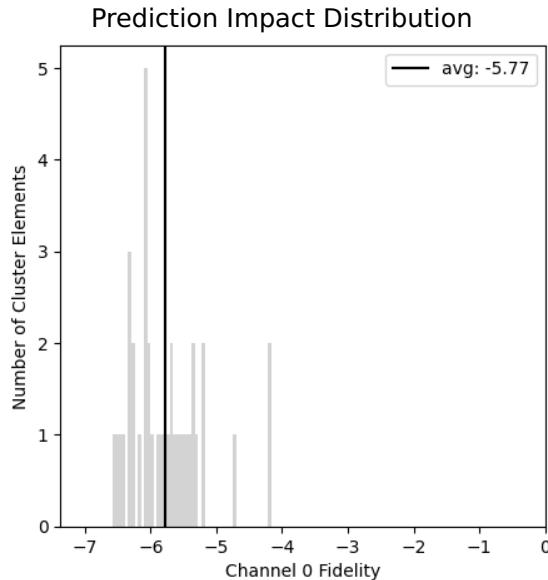
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	38
Channel Index	0.0 (0.0)

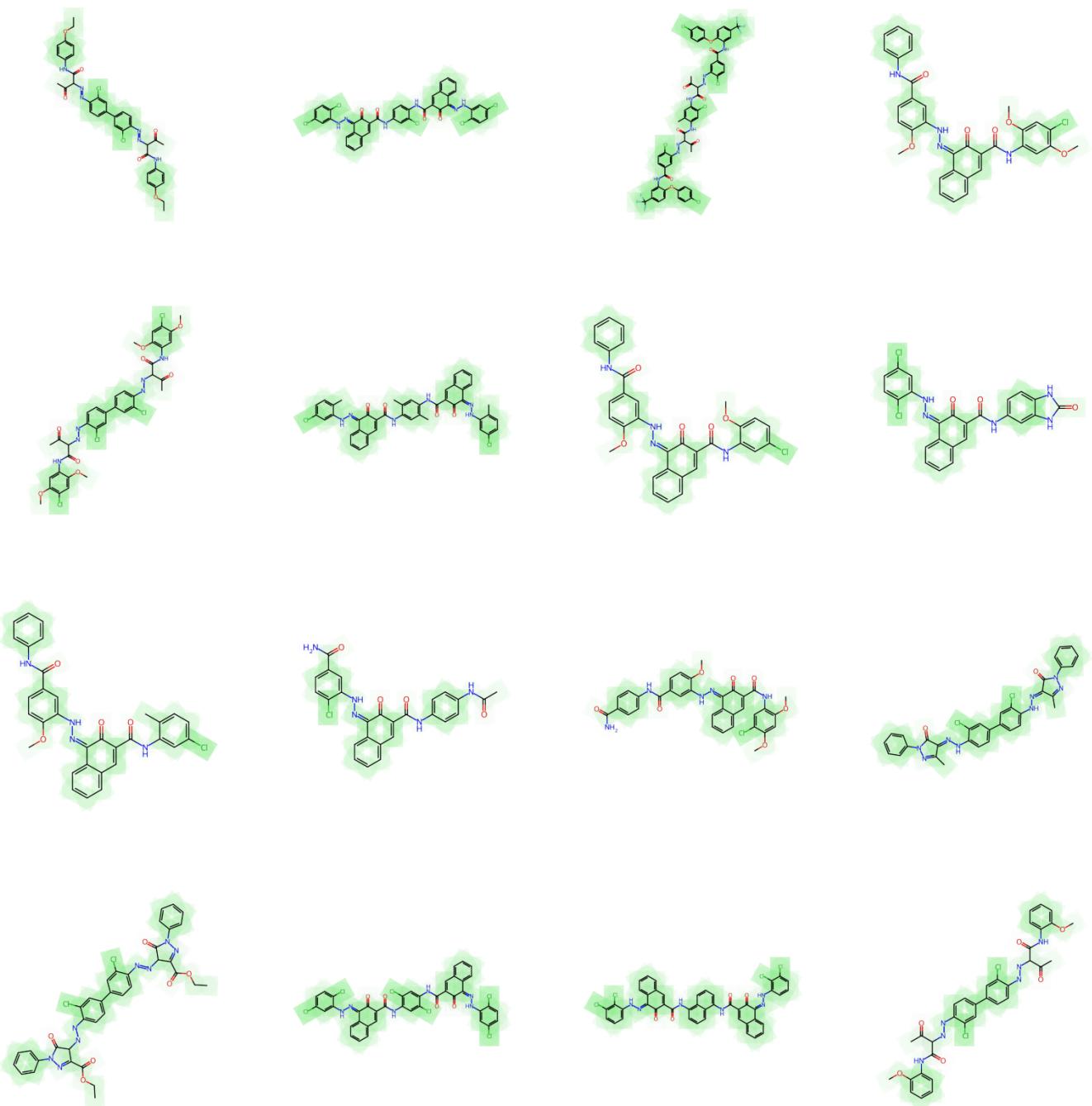
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



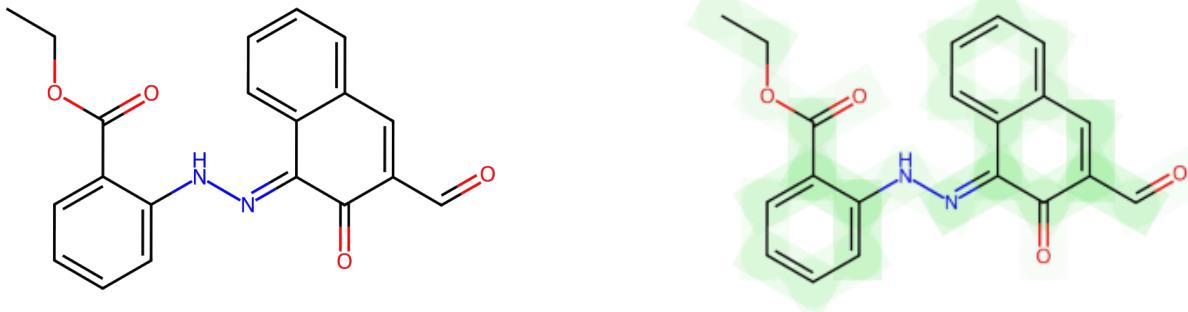
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The molecular substructure in question includes both hydrophobic (e.g., aromatic rings denoted by "c1:c:c:c:c:c:1") and hydrophilic (e.g., hydroxyl "-O-", nitro "-N=N-", and carbonyl "C=O" groups) components. However, the presence of large hydrophobic aromatic rings likely overwhelms the solubility-enhancing effects of the polar functional groups, ultimately leading to a decrease in water solubility. Additionally, the molecular rigidity introduced by these aromatic rings could hinder the molecule's ability to adapt and interact favorably with the aqueous environment, further depressing solubility.

**Hypothesis:** The identified molecular substructure is related to decreased water solubility due to the dominance of hydrophobic aromatic moieties. Although polar and potentially ionizable groups are present in the structure, the overall size and hydrophobic character of the aromatic portions significantly reduce the molecule's affinity for the aqueous phase, as evidenced by a negative influence on water solubility.

# Cluster #38 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 38, from importance channel 1 (*positive*), represents a motif consisting of 3.6 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 0.8 ( $\pm 0.2$ ) on the prediction outcome.

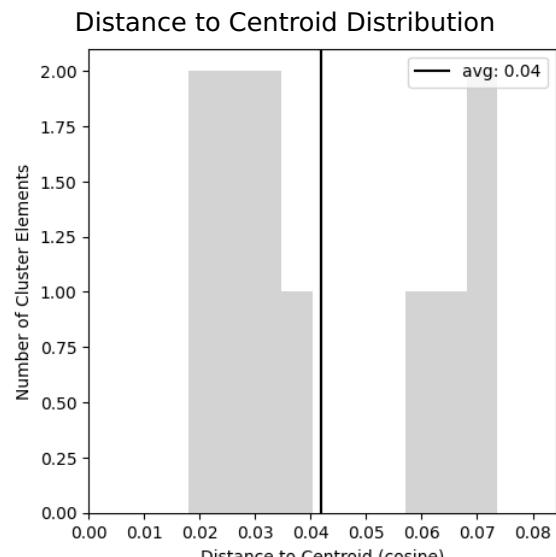
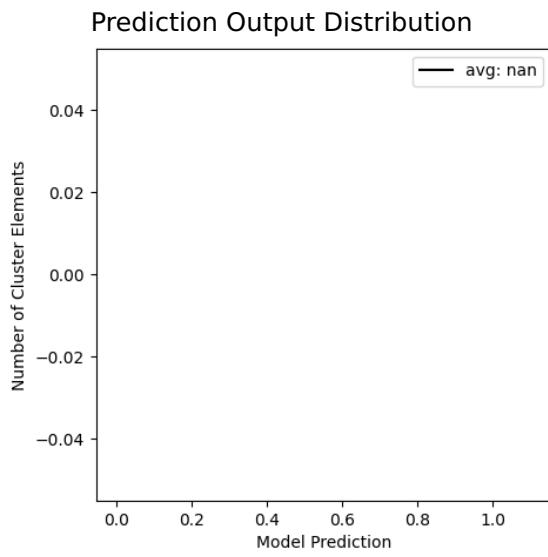
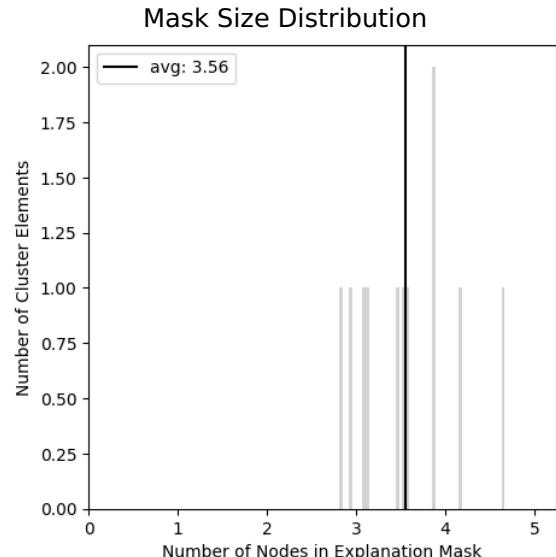
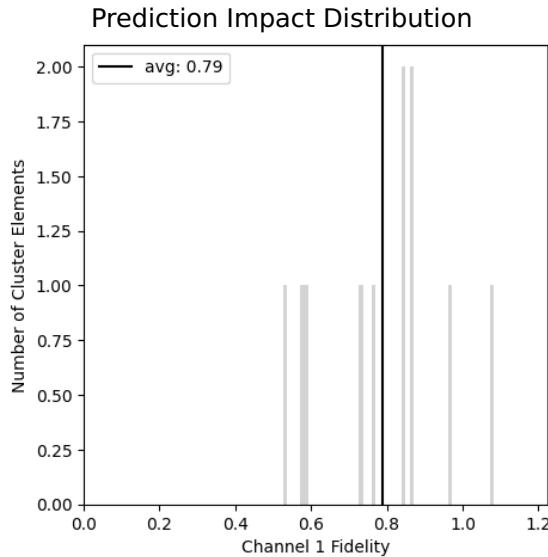
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	11
Channel Index	1.0 (0.0)

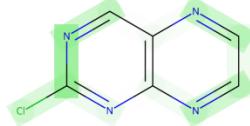
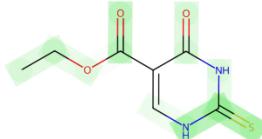
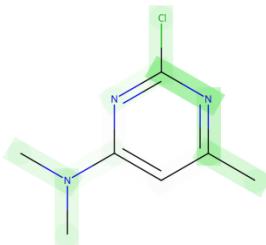
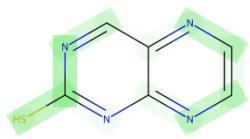
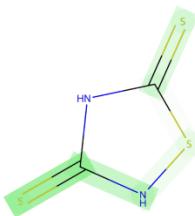
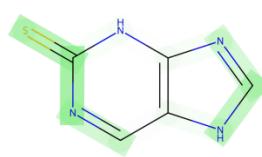
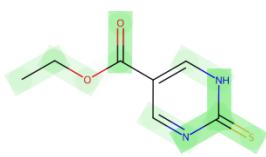
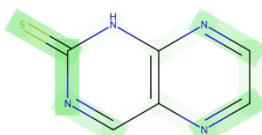
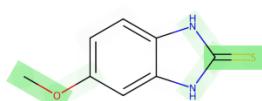
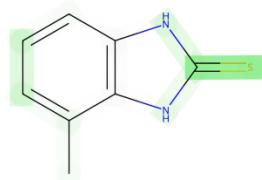
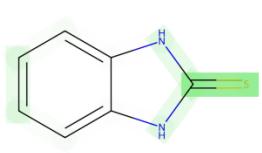
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



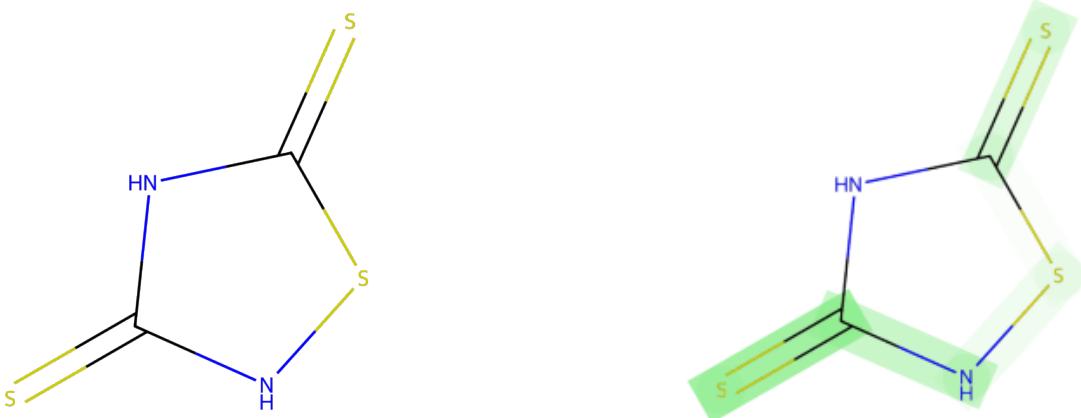
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES structure "S=C1NSC(=S)N1" describes a heterocyclic compound containing sulfur and nitrogen atoms within a ring system. The presence of heteroatoms (i.e., atoms other than carbon and hydrogen) like sulfur (S) and nitrogen (N) in a molecule increases its polarity. Increased polarity enhances a compound's ability to engage in hydrogen bonding with water molecules, which is a key interaction for solubility in water. Additionally, the double bond character suggested by "=C" and "=S" in the ring can also affect solubility due to potential resonance structures that may distribute charge across the molecule, further increasing water-molecule interactions.

**Hypothesis:** Molecules with the "S=C1NSC(=S)N1" substructure have enhanced water solubility. This is likely due to the increased polarity and potential hydrogen-bonding capabilities conferred by the sulfur and nitrogen heteroatoms, as well as resonance structures that can arise from the double bonds, which allow for better interaction with the polar solvent water.

# Cluster #39 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 39, from importance channel 1 (*positive*), represents a motif consisting of 2.7 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 0.5 ( $\pm 0.0$ ) on the prediction outcome.

## Properties

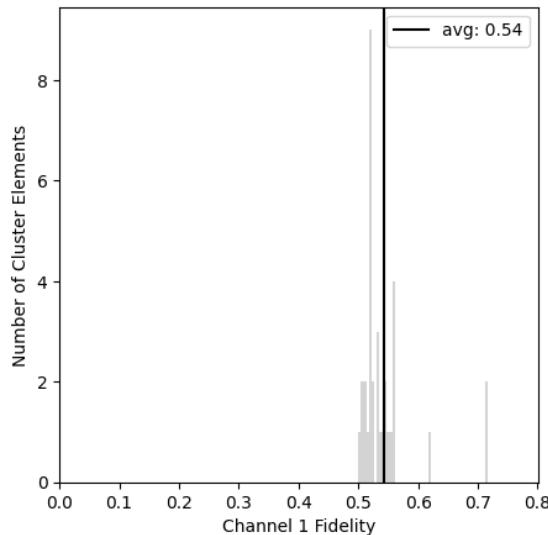
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	33
Channel Index	1.0 (0.0)

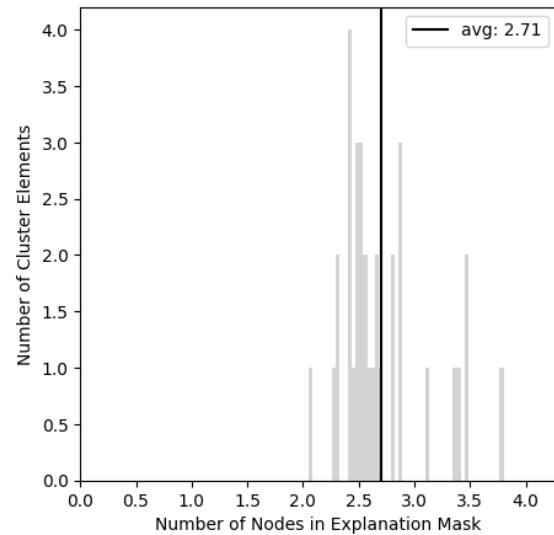
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

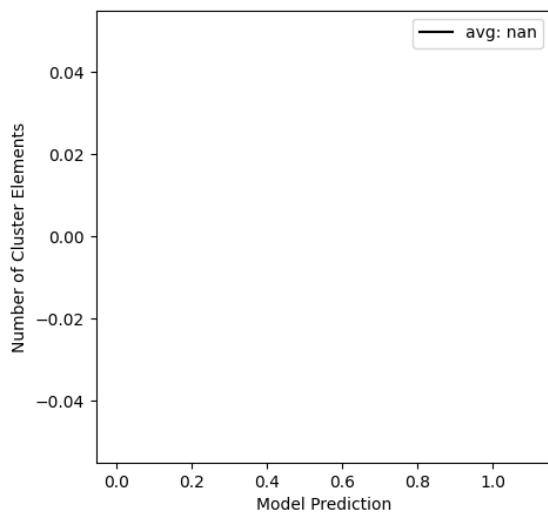
Prediction Impact Distribution



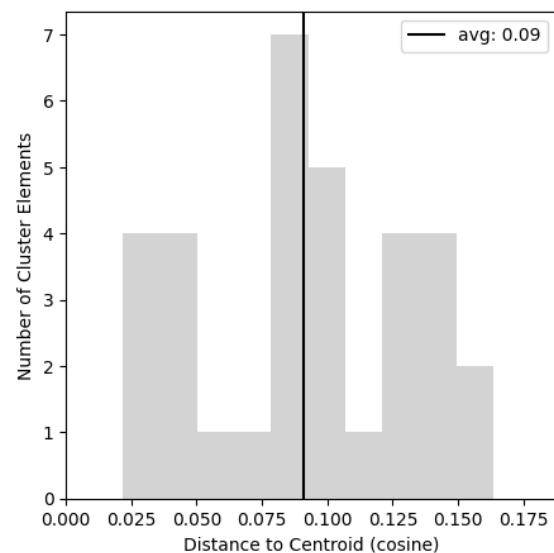
Mask Size Distribution



Prediction Output Distribution

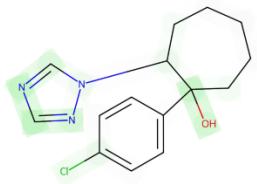
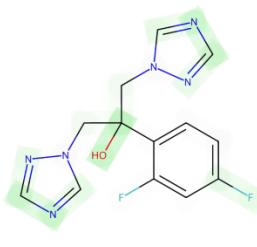
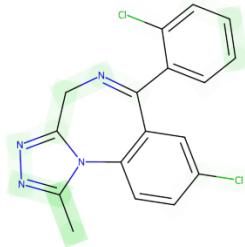
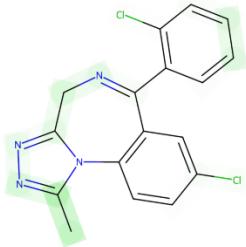
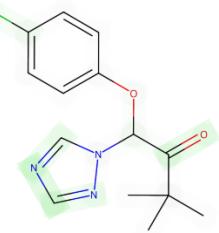
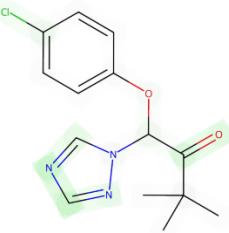
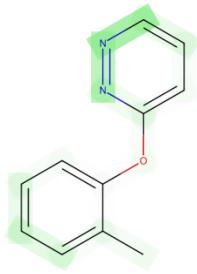
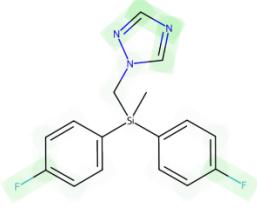
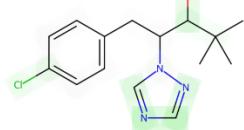
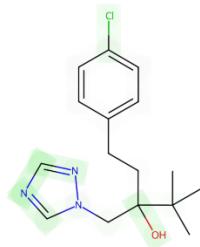
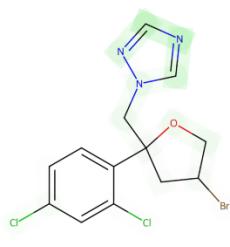
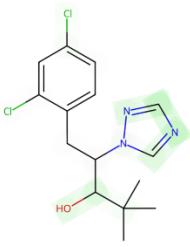
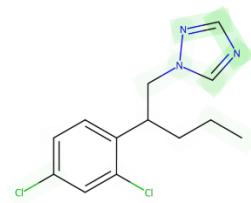
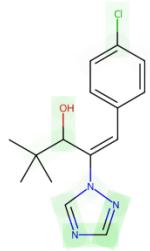
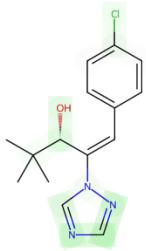
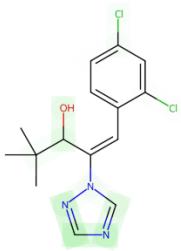


Distance to Centroid Distribution



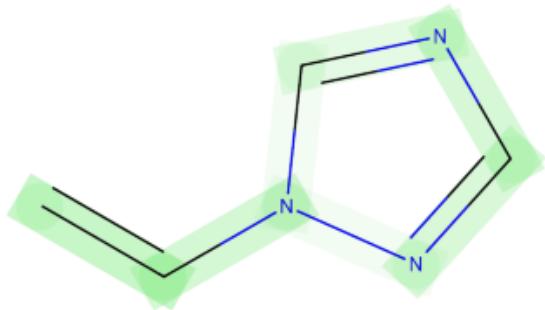
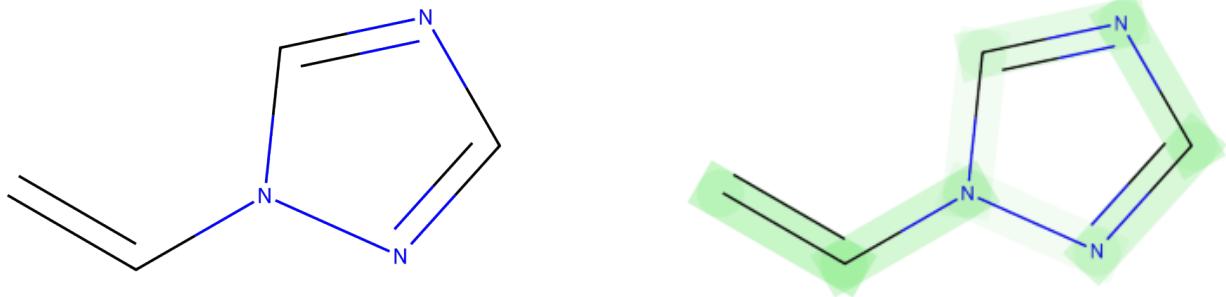
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES representation "C=C-n1:c:n:c:n:1" depicts a conjugated system that includes a heterocyclic aromatic ring with alternate nitrogen and carbon atoms. This structure introduces polarity due to the electronegativity difference between nitrogen and carbon atoms. The polarity and ability to engage in hydrogen bonding with water molecules increase the solubility of such a molecule in water. Additionally, the presence of a double bond (indicated by "C=C") may increase the rigidity of the structure, which could potentially decrease solubility by making the molecule less able to adapt to the water structure.

**Hypothesis:** Molecules containing the substructure "C=C-n1:c:n:c:n:1" will generally exhibit moderate water solubility. This is due to the polar nature of the nitrogen atoms in the aromatic ring, which can form hydrogen bonds with water, countered slightly by the influence of the double bond which may reduce solubility by increasing molecular rigidity.

# Cluster #40 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 40, from importance channel 1 (*positive*), represents a motif consisting of 3.2 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 0.8 ( $\pm 0.2$ ) on the prediction outcome.

## Properties

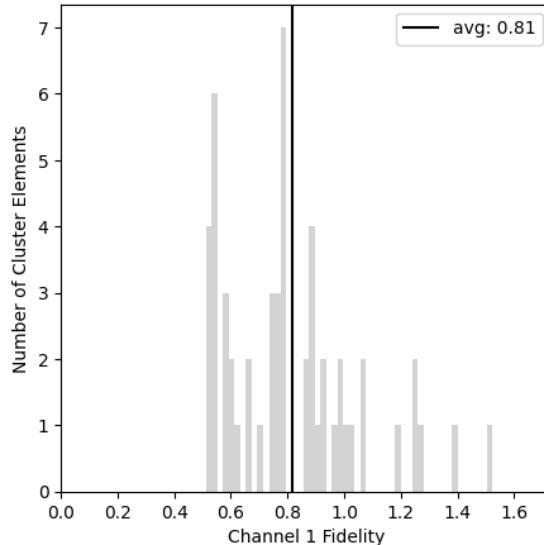
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	54
Channel Index	1.0 (0.0)

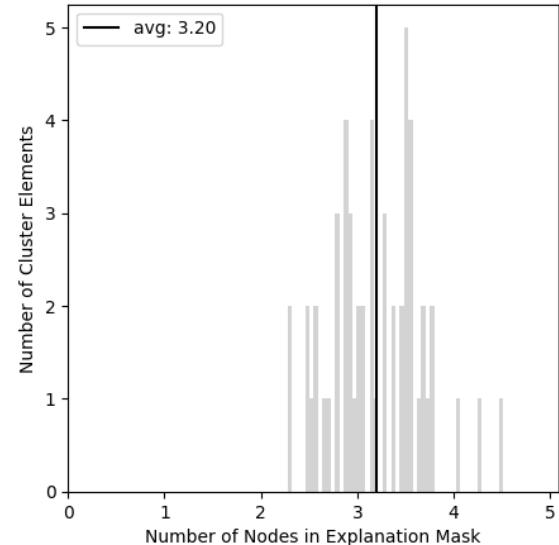
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

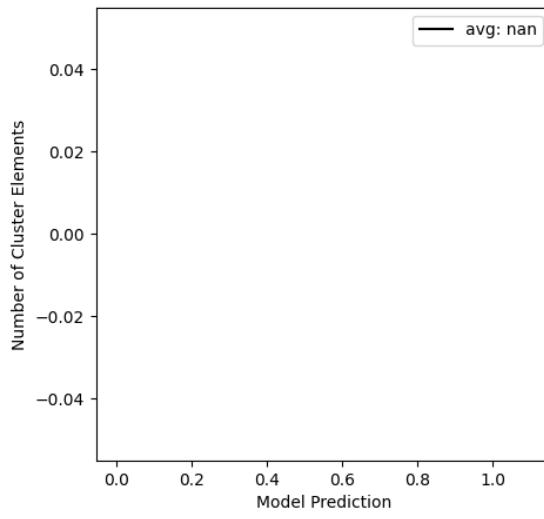
Prediction Impact Distribution



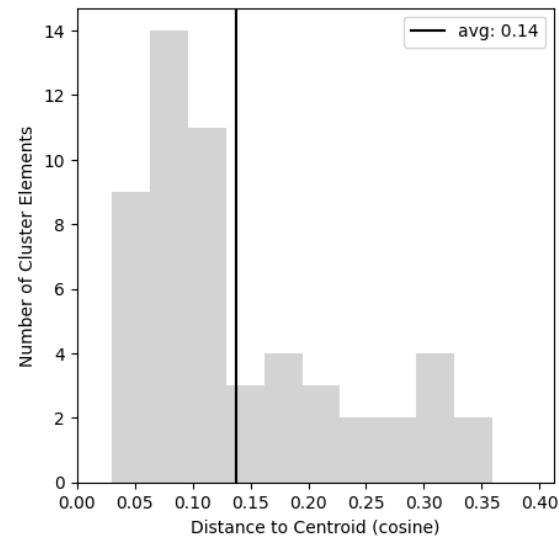
Mask Size Distribution



Prediction Output Distribution

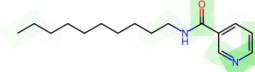
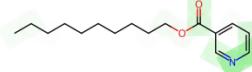
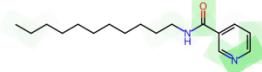
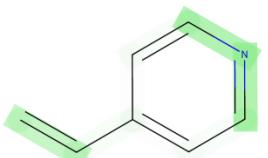
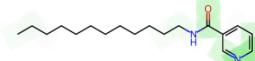
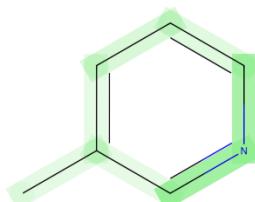
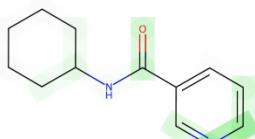
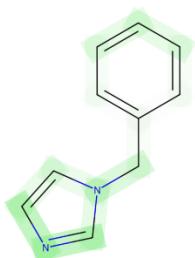
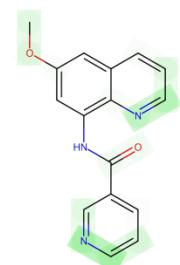
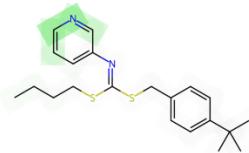
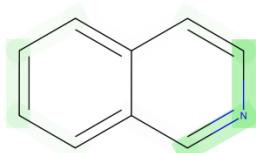
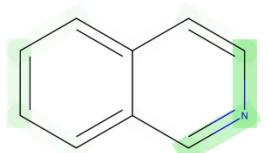
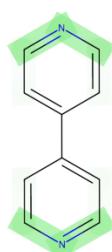
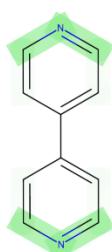
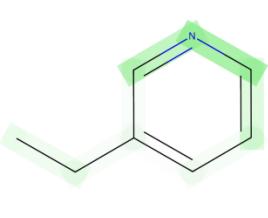


Distance to Centroid Distribution



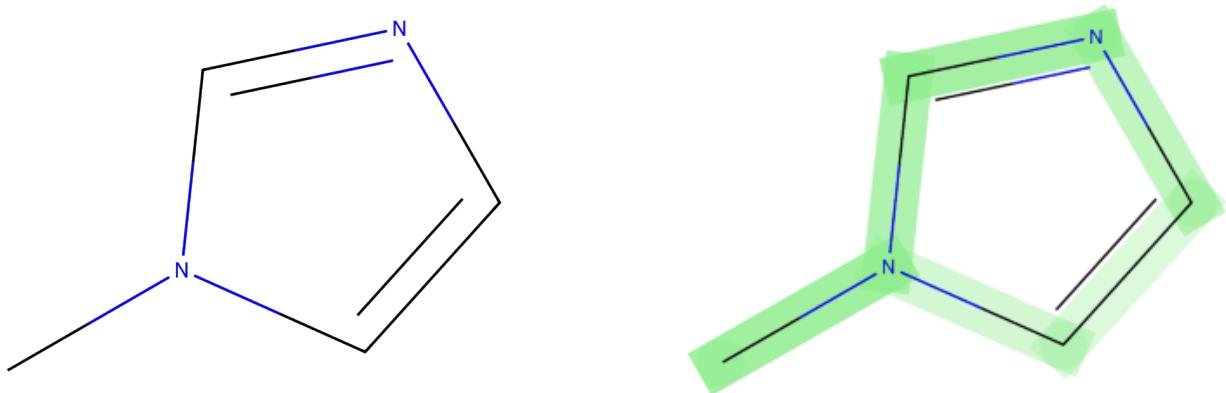
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure represented by the SMILES "CN1C=CN=C1" corresponds to a pyrimidine ring with an amino group attached. Pyrimidine is an aromatic heterocycle that, while less polar than pyridine, still presents some polarity due to the presence of nitrogen atoms. The amino group (-NH<sub>2</sub>) increases the overall polarity of the molecule, enhancing its hydrogen bonding capability with water molecules. Amines are generally recognized to increase water solubility because the -NH<sub>2</sub> group can serve as a hydrogen bond donor, forming favorable interactions with the hydrogen bond acceptor molecules in water. Additionally, the double bonds in the ring system could affect the water solubility due to possible π-π stacking interactions with water, but their impact is likely less significant than that of the amino group.

**Hypothesis:** The presence of an amino group on a pyrimidine ring increases water solubility. The nitrogen in the pyrimidine and the amino group both contribute to improved hydrogen bonding with water, while the aromatic nature of the pyrimidine might slightly decrease solubility; however, the positive inductive effect of the amino group is presumed to play a more dominant role in enhancing water solubility.

# Cluster #41 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 41, from importance channel 1 (*positive*), represents a motif consisting of 2.8 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 0.8 ( $\pm 0.2$ ) on the prediction outcome.

## Properties

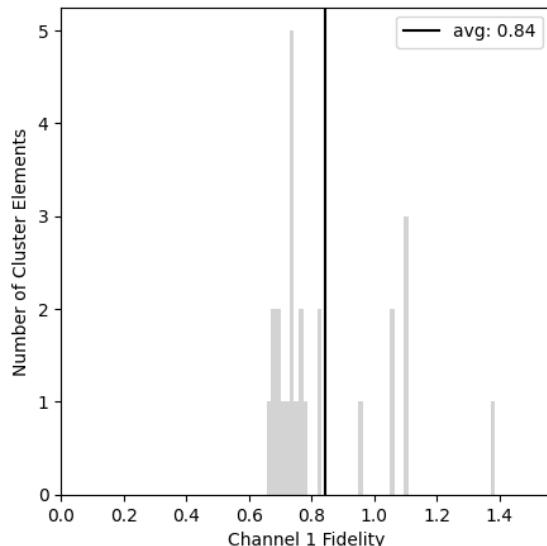
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	25
Channel Index	1.0 (0.0)

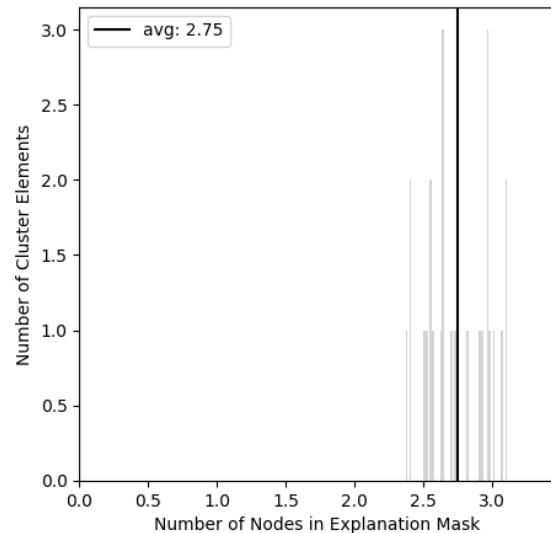
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

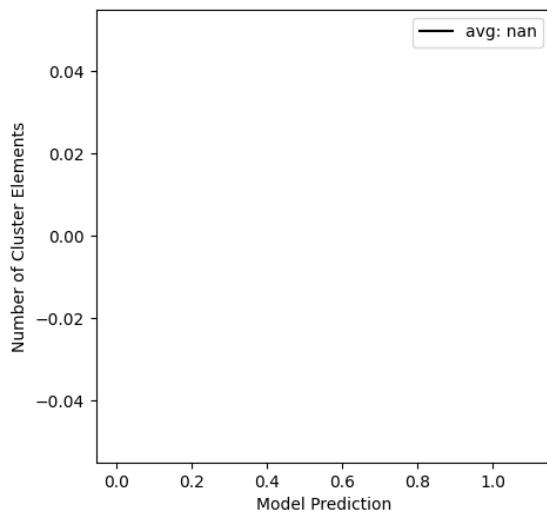
Prediction Impact Distribution



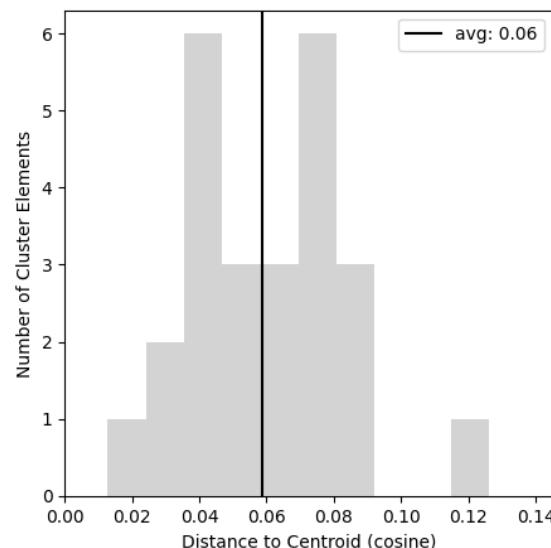
Mask Size Distribution



Prediction Output Distribution

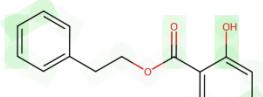
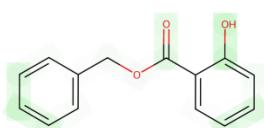
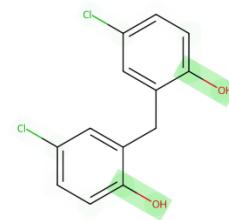
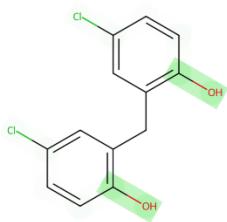
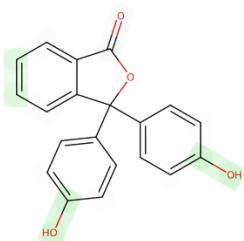
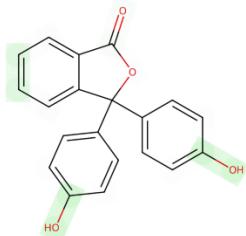
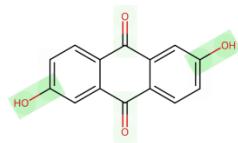
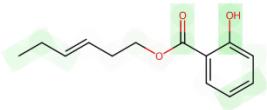
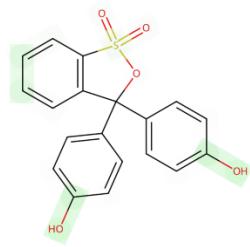
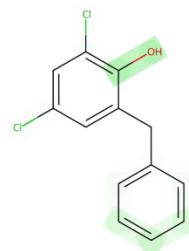
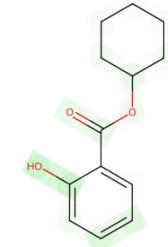
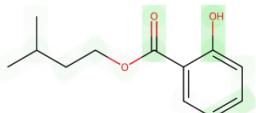
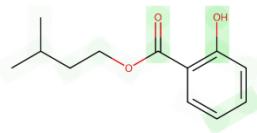
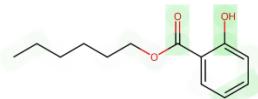
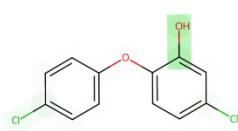
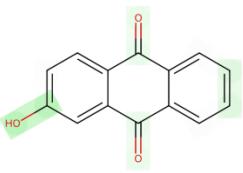


Distance to Centroid Distribution



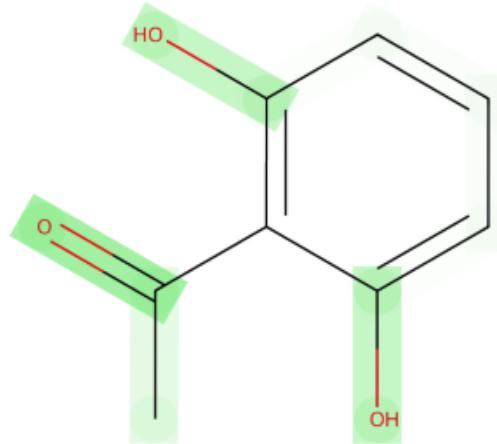
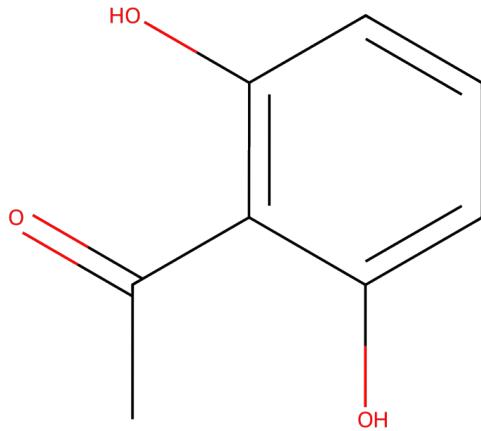
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given structure represents a molecule with a phenolic group (-c1:c(-O):c:c:c:c1-O), which includes two hydroxyl groups (-OH) attached to a benzene ring, and a ketone group (-C(=O)-). The hydroxyl groups can form hydrogen bonds with water molecules, which significantly enhances water solubility. The ketone group also has a polar carbonyl (-C=O) that can engage in dipole-dipole interactions with water. However, the presence of the non-polar carbon chain (C-C) might slightly reduce solubility compared to compounds with only polar functional groups.

**Hypothesis:** Molecules containing phenolic and ketone groups connected to non-polar chains have an increased tendency to be soluble in water. The hydroxyl groups offer sites for hydrogen bonding, while the carbonyl group can interact with water through dipole-dipole forces, both promoting water solubility. The hydrophobic carbon chain could slightly decrease solubility, but the polar groups' influence is more pronounced in this case, leading to a high overall water solubility.

# Cluster #42 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 42, from importance channel 1 (*positive*), represents a motif consisting of 3.3 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 0.9 ( $\pm 0.1$ ) on the prediction outcome.

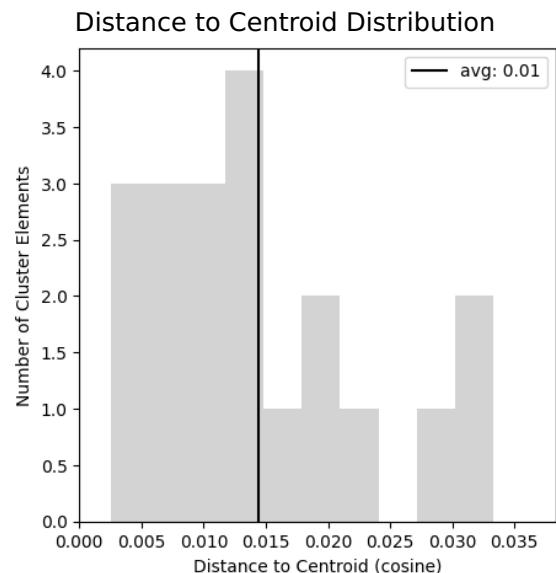
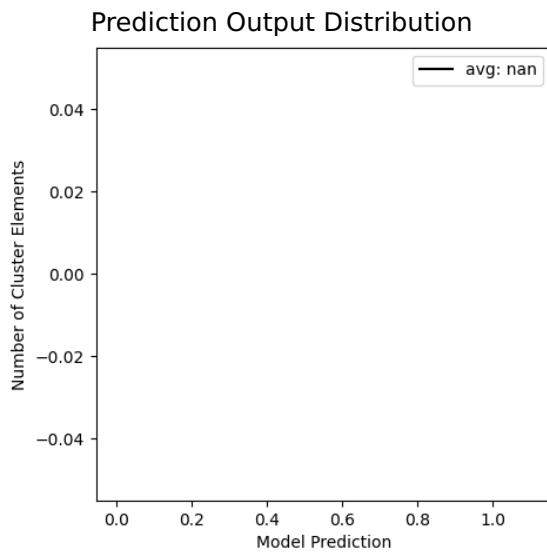
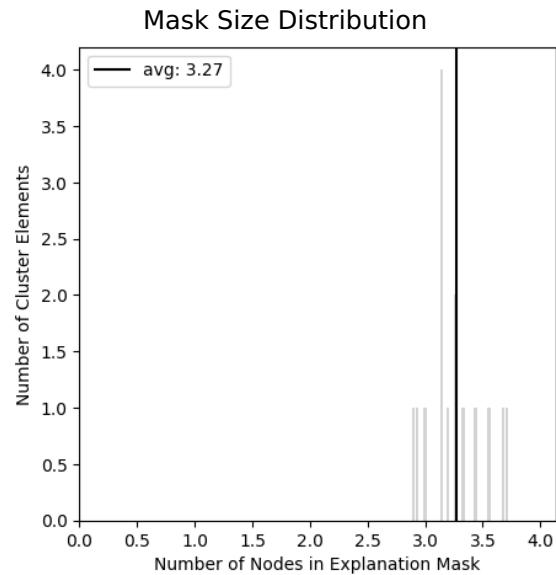
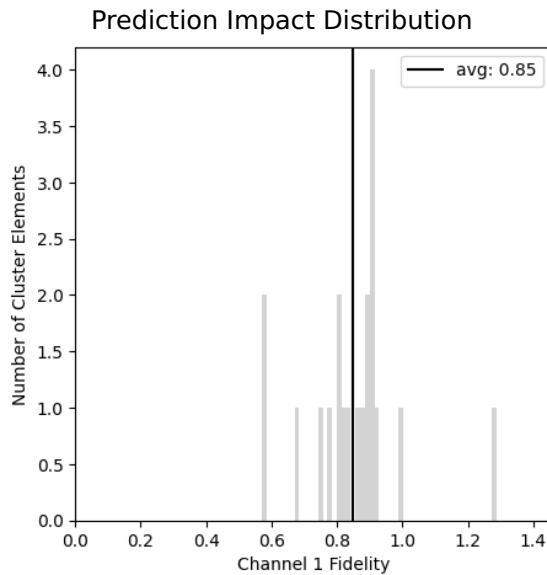
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	20
Channel Index	1.0 (0.0)

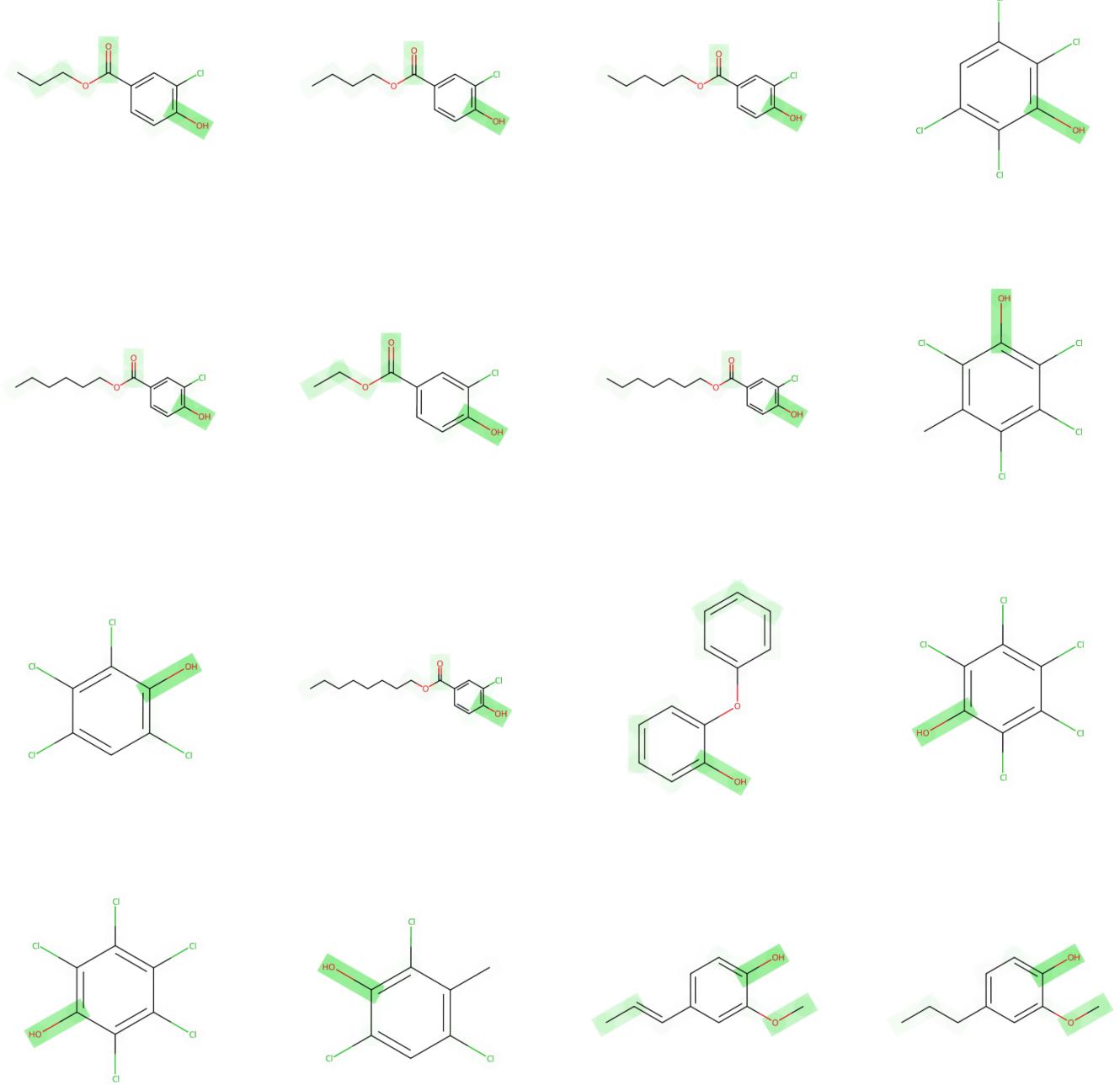
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



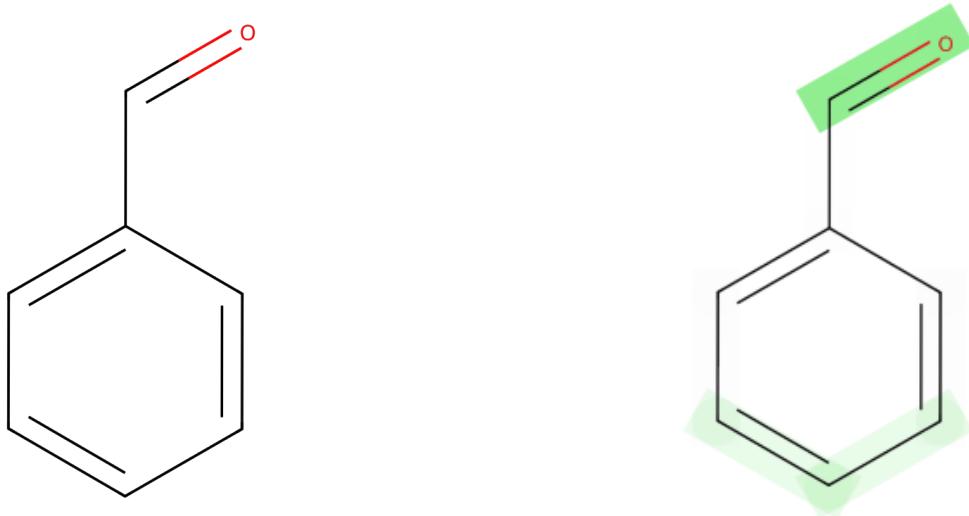
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a benzene ring with an attached carbonyl group (a carbon double-bonded to an oxygen). The presence of the benzene ring suggests a hydrophobic character due to the stable, nonpolar nature of the aromatic system, which typically decreases water solubility. However, the carbonyl group is polar and can engage in hydrogen bonding with water molecules, which would increase solubility. The balance between these two opposing influences (hydrophobic aromaticity and polar carbonyl) is likely to determine the net solubility in water.

**Hypothesis:** The presence of a carbonyl group attached to an aromatic benzene ring in a molecule correlates with moderate water solubility. The benzene ring reduces solubility due to its hydrophobicity, whereas the carbonyl group enhances it through polarity and hydrogen bonding potential. The empirical evidence indicates the polar carbonyl effect slightly outweighs the hydrophobic aromatic effect, leading to the observed positive influence on water solubility.

# Cluster #43 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 43, from importance channel 1 (*positive*), represents a motif consisting of 3.2 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 1.0 ( $\pm 0.1$ ) on the prediction outcome.

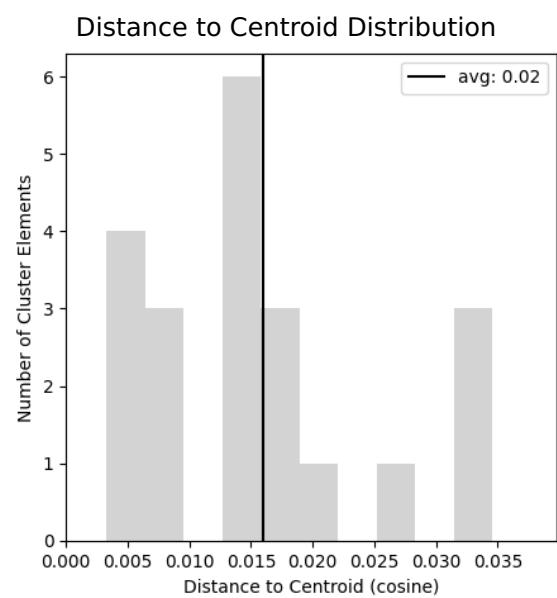
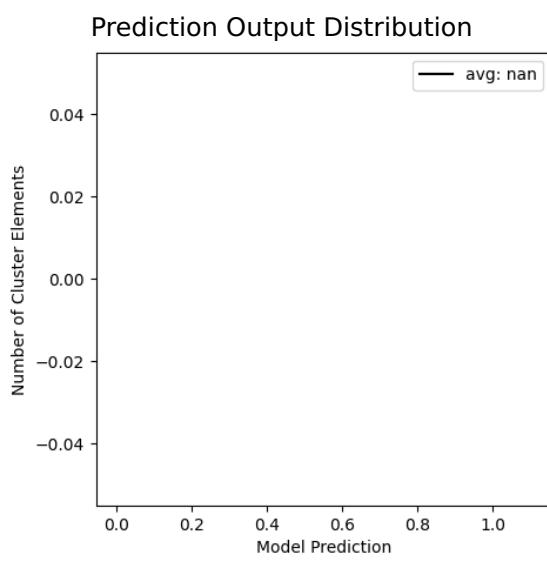
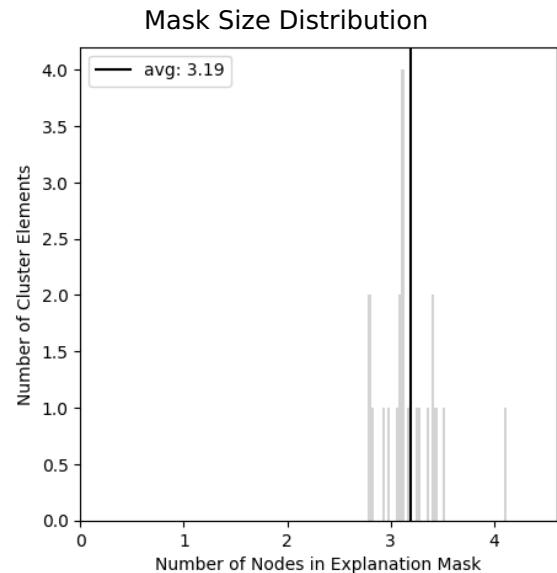
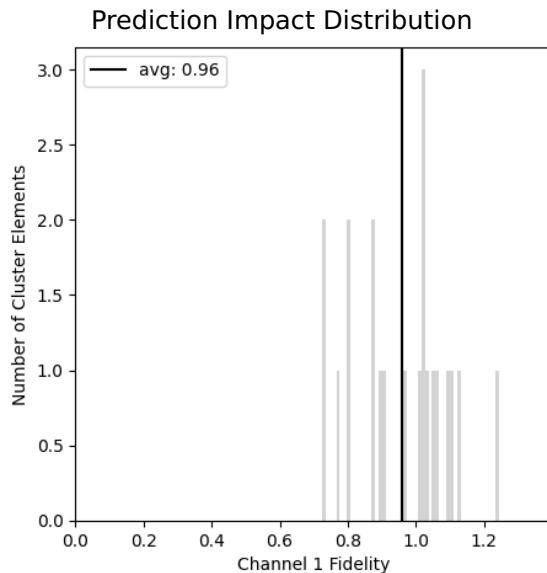
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	21
Channel Index	1.0 (0.0)

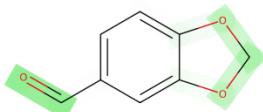
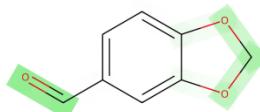
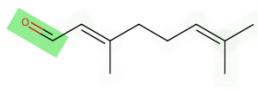
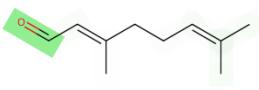
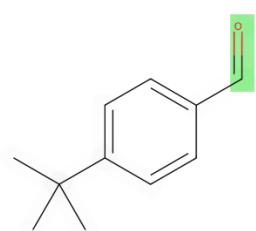
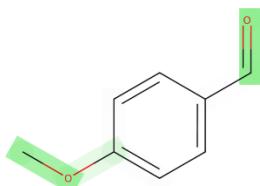
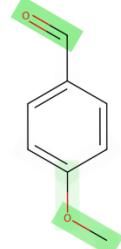
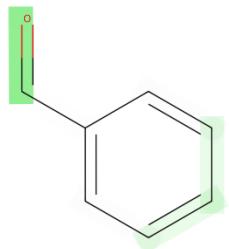
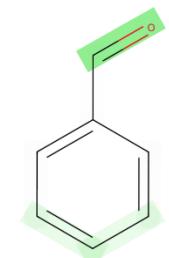
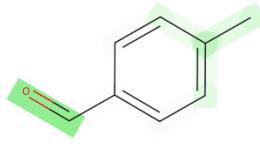
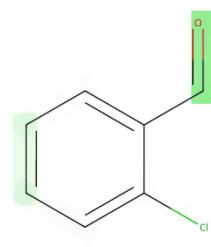
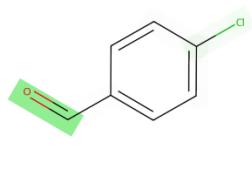
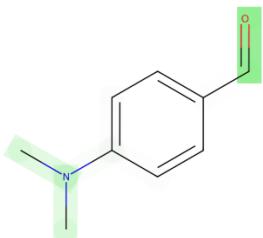
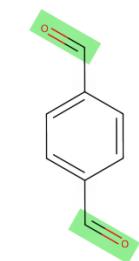
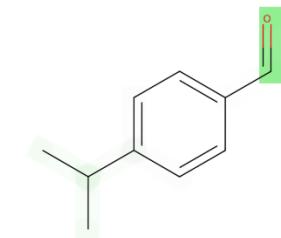
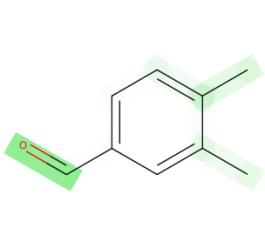
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



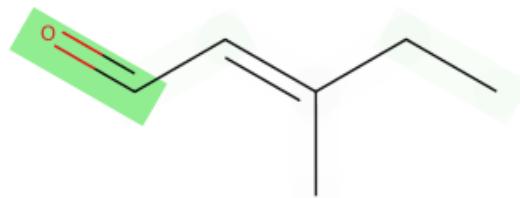
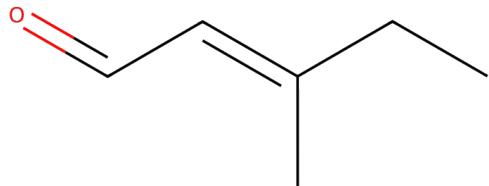
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure "C-C-C(-C)=C-C=O" represents a hydrocarbon chain with a terminal aldehyde group. The hydrocarbon part is nonpolar and hydrophobic, which would typically decrease water solubility. However, the presence of a carbonyl group (C=O) introduces a polar functionality which can engage in hydrogen bonding with water molecules, thus increasing the water solubility of the molecule.

**Hypothesis:** The molecular fragment represented as "C-C-C(-C)=C-C=O" tends to increase water solubility due to the polar aldehyde group. While the nonpolar hydrocarbon chain inherently resists solubility in water, the aldehyde's ability to form hydrogen bonds with water overcomes this resistance, making the molecule as a whole more soluble.

# Cluster #44 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 44, from importance channel 1 (*positive*), represents a motif consisting of 3.1 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 0.9 ( $\pm 0.1$ ) on the prediction outcome.

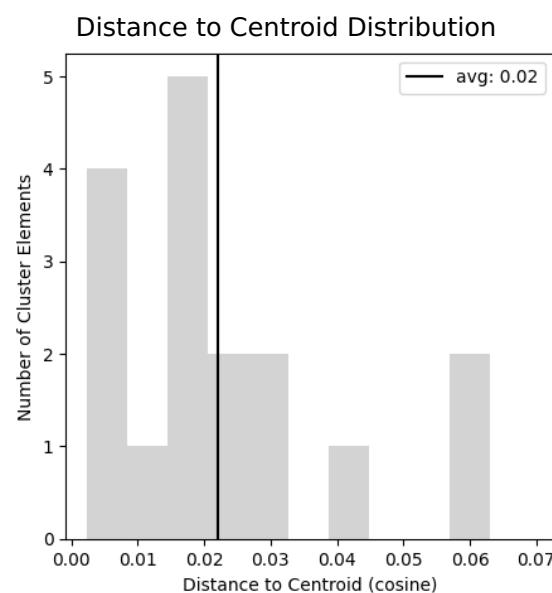
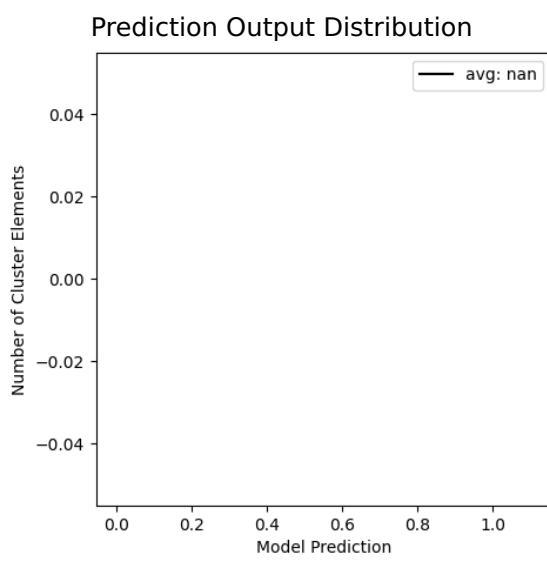
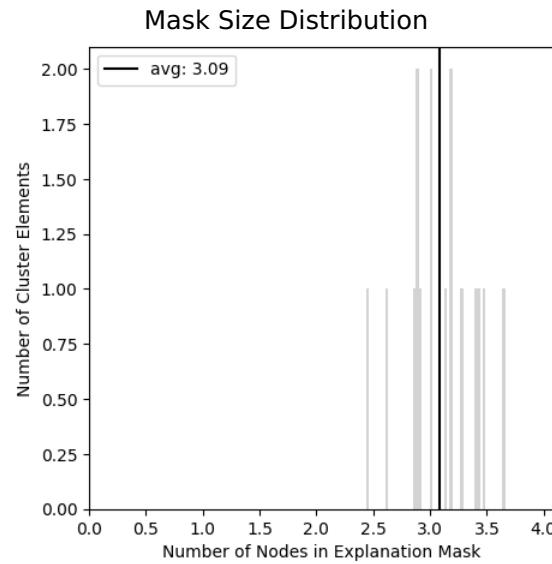
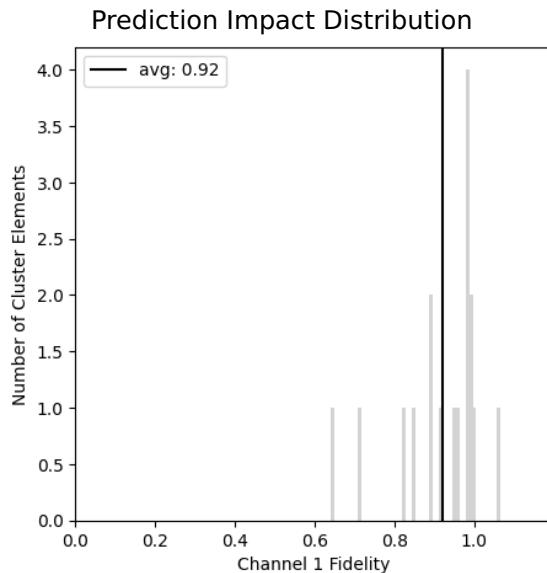
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	17
Channel Index	1.0 (0.0)

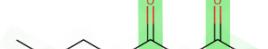
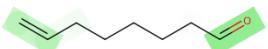
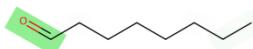
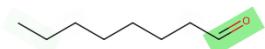
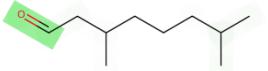
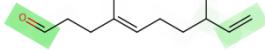
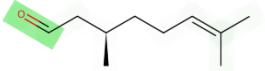
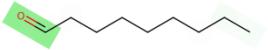
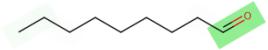
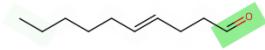
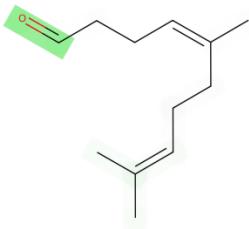
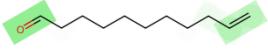
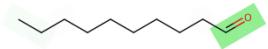
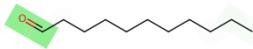
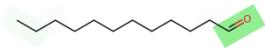
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a straight-chain alkyl group terminated with a carbonyl group (a ketone). The alkyl group contributes to hydrophobicity, decreasing water solubility, while the carbonyl group can form hydrogen bonds with water molecules, increasing water solubility. The presence of the carbonyl group likely exerts a stronger influence towards solubility due to its ability to engage in hydrogen bonding, which is a strong intermolecular force in aqueous environments.

**Hypothesis:** The molecular structure "C-C-C-C=O" is associated with increased water solubility, likely due to the hydrogen bonding capability of the carbonyl group. The alkyl chain's effect is overshadowed by the polarity and hydrogen-bonding ability of the carbonyl, resulting in an overall positive influence on water solubility.

# Cluster #45 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 45, from importance channel 1 (*positive*), represents a motif consisting of 3.0 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 1.0 ( $\pm 0.2$ ) on the prediction outcome.

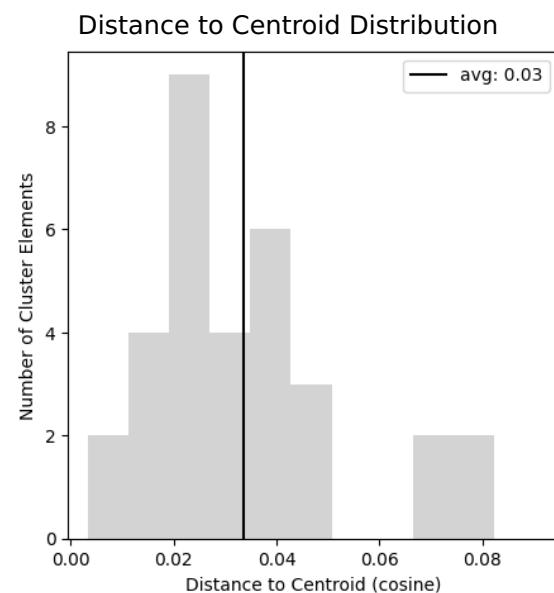
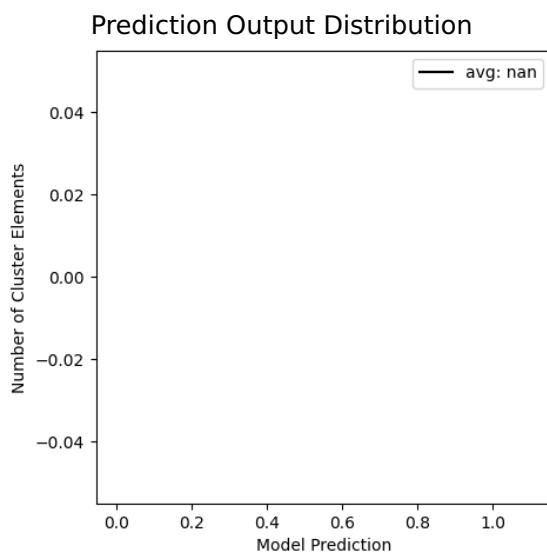
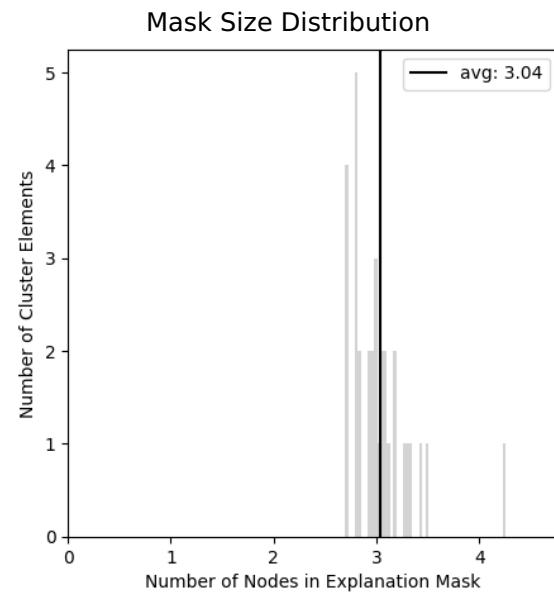
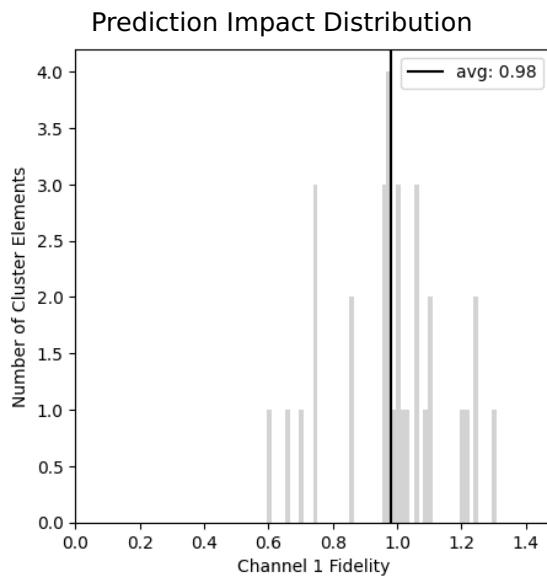
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	32
Channel Index	1.0 (0.0)

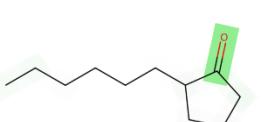
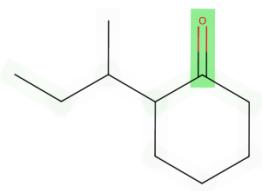
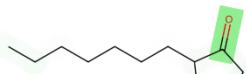
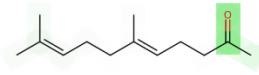
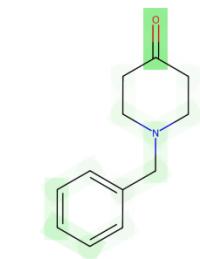
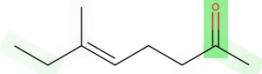
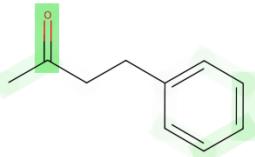
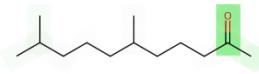
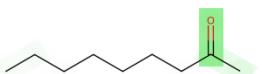
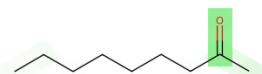
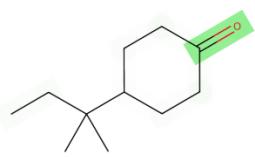
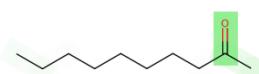
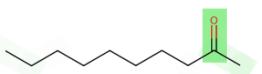
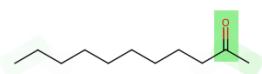
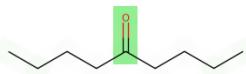
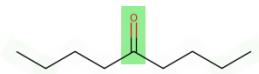
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The molecular substructure represented by "C-C-C-C(-C)=O" suggests a hydrocarbon chain with an aldehyde functional group. Aliphatic hydrocarbon chains, being nonpolar, typically decrease water solubility due to the hydrophobic effect. However, the presence of the aldehyde group with the carbonyl function (=O) at the end of the chain can increase the molecule's polarity, making it more likely to interact with water molecules through hydrogen bonding. This dual characteristic could lead to a moderate influence on water solubility.

**Hypothesis:** Molecules containing the substructure "C-C-C-C(-C)=O" have a moderate influence on water solubility. The hydrocarbon chain exhibits hydrophobic characteristics reducing water solubility, whereas the aldehyde group increases solubility due to its polar nature and the potential for hydrogen bonding with water.

# Cluster #46 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 46, from importance channel 1 (*positive*), represents a motif consisting of 3.2 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 1.0 ( $\pm 0.2$ ) on the prediction outcome.

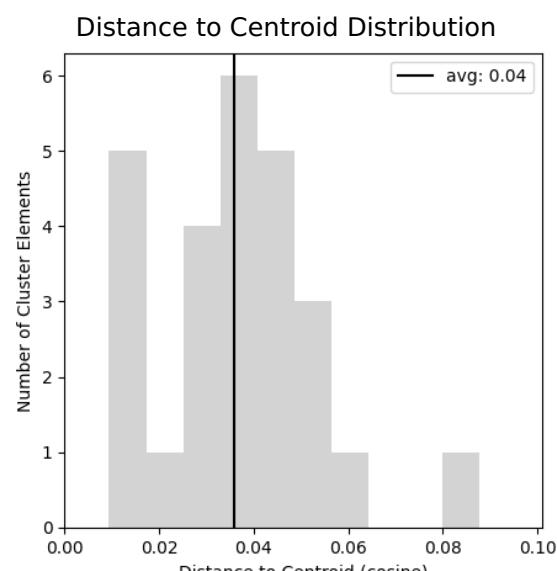
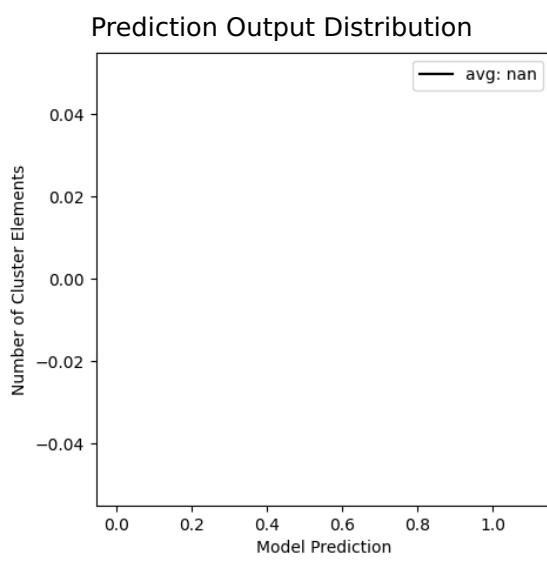
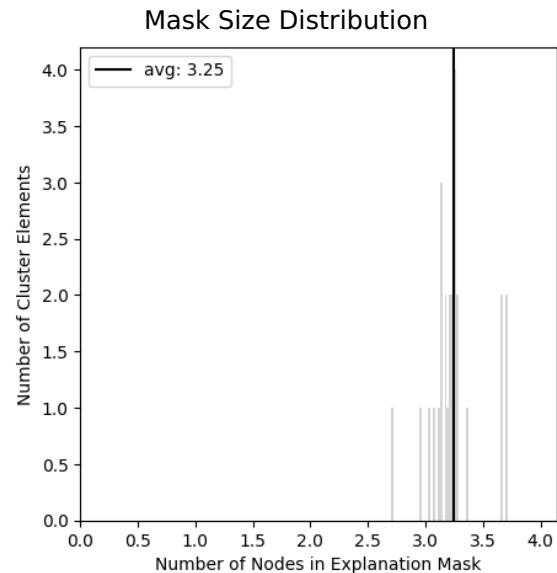
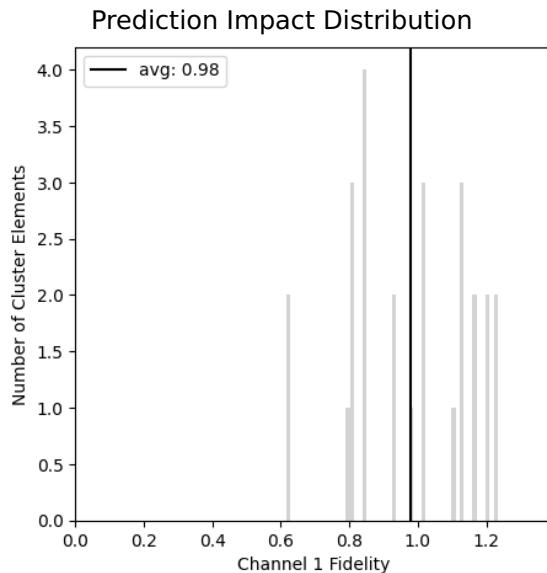
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	26
Channel Index	1.0 (0.0)

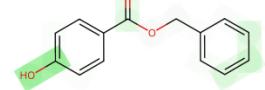
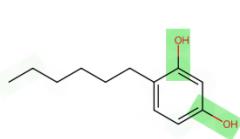
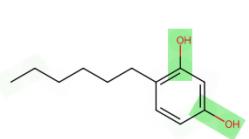
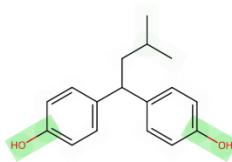
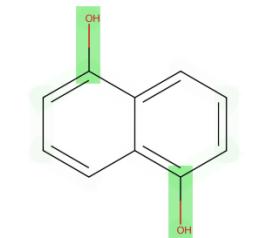
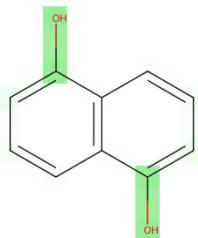
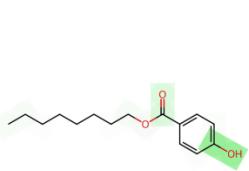
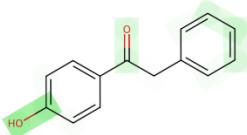
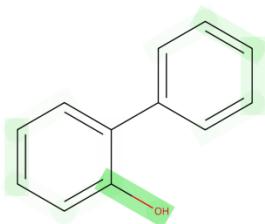
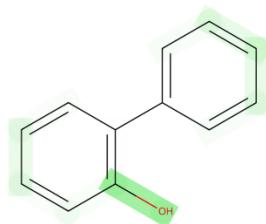
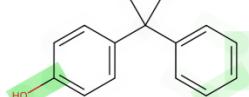
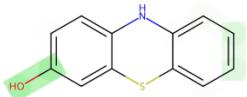
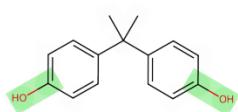
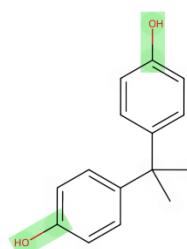
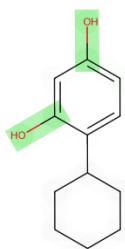
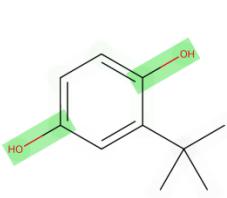
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



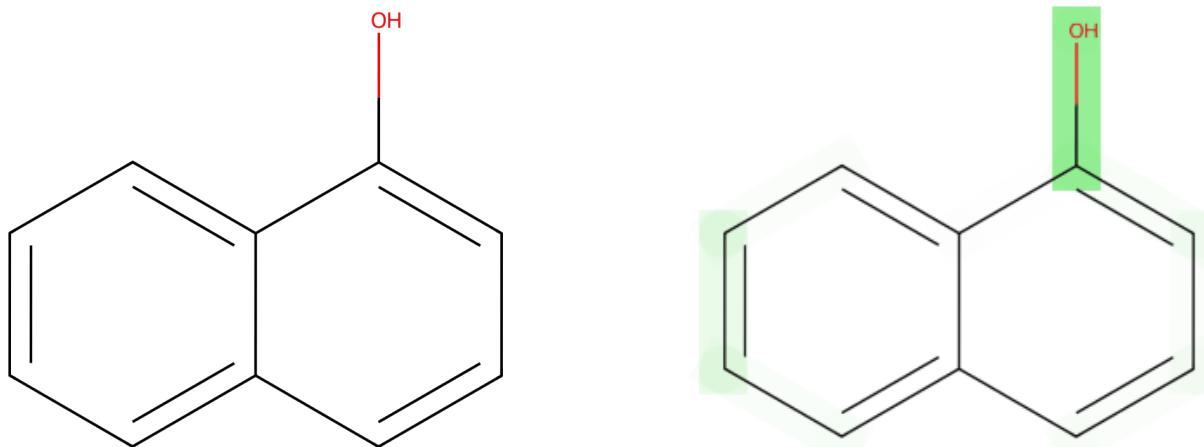
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES representation "O-c1:c:c:c:c2:c:c:c:c1:2" describes a naphthalene derivative with a hydroxyl group attached to it. The hydroxyl group (-OH) is known to increase water solubility due to its ability to form hydrogen bonds with water molecules. Naphthalene itself is hydrophobic and not readily soluble in water; however, the addition of a polar functional group like -OH can enhance its solubility by facilitating interaction with the polar solvent, in this case, water.

**Hypothesis:** The presence of a hydroxyl group on a naphthalene ring system increases water solubility. The hydroxyl group forms hydrogen bonds with water, which outweigh the hydrophobic character of the fused aromatic rings, thus enhancing the overall solubility of the molecule in water.

# Cluster #47 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 47, from importance channel 1 (*positive*), represents a motif consisting of 3.1 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 1.0 ( $\pm 0.1$ ) on the prediction outcome.

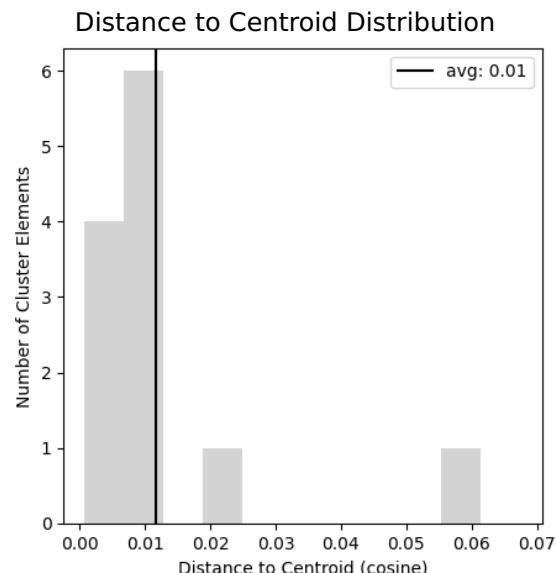
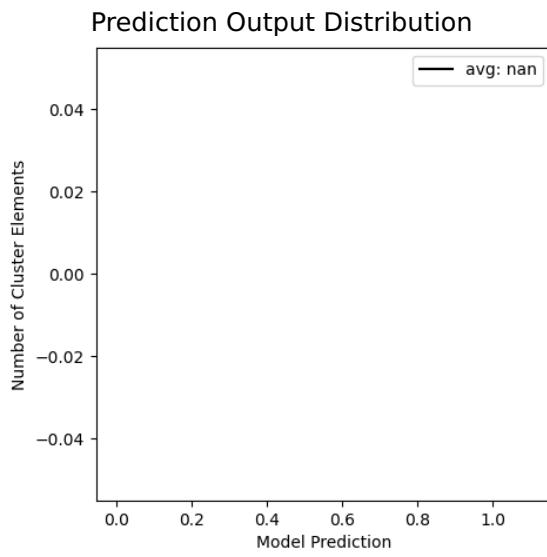
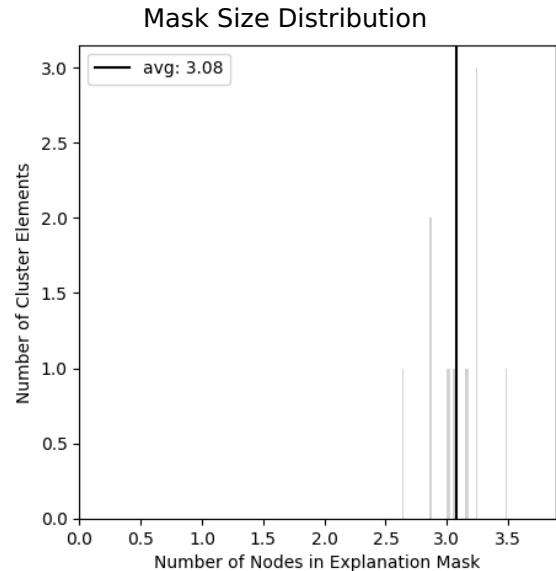
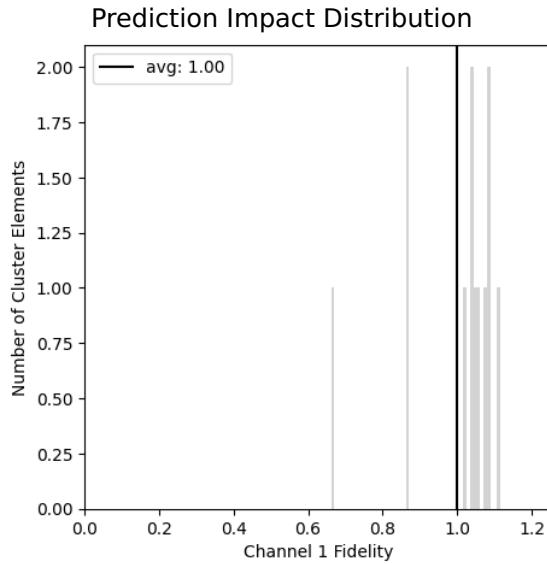
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	12
Channel Index	1.0 (0.0)

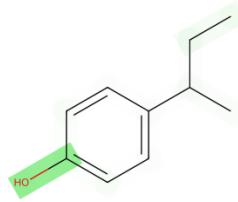
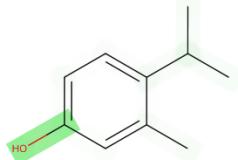
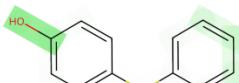
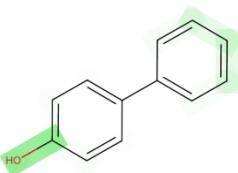
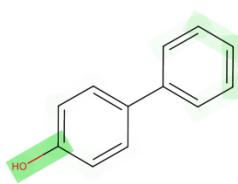
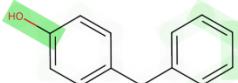
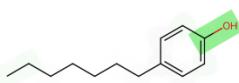
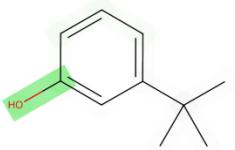
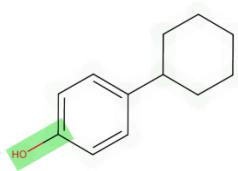
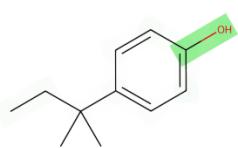
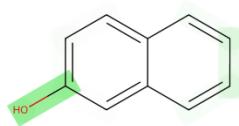
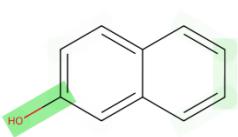
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The designated substructure suggests the presence of both hydrophobic (carbon chain 'C-C-') and hydrophilic (phenol 'c1:c:c:c:c(-O):c:1') groups within the same molecule. The presence of the phenolic hydroxyl group ('-O') attached to an aromatic ring provides the potential for hydrogen bonding with water molecules, which can enhance the solubility of the molecule in water. Additionally, the aromatic ring can partake in π-π interactions, which do not significantly hinder solubility and may sometimes aid in solvation processes, depending on the context of the surrounding substructures. The influence of the carbon chain, on the other hand, is generally to decrease water solubility due to its nonpolar nature.

**Hypothesis:** The molecular structure "C-C-c1:c:c:c:c(-O):c:1" indicates a possible increase in water solubility. This is hypothesized to be due to the hydroxyl group bound to an aromatic ring, which can form hydrogen bonds with water, overcoming the hydrophobic effect of the carbon chain to a certain extent.

# Cluster #48 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 48, from importance channel 1 (*positive*), represents a motif consisting of 3.1 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 1.0 ( $\pm 0.1$ ) on the prediction outcome.

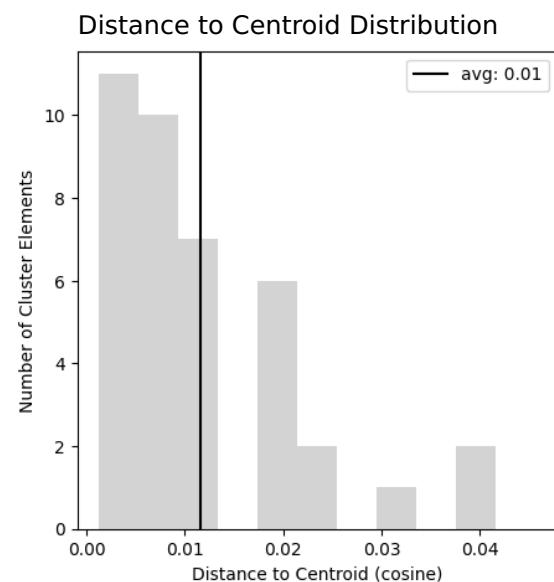
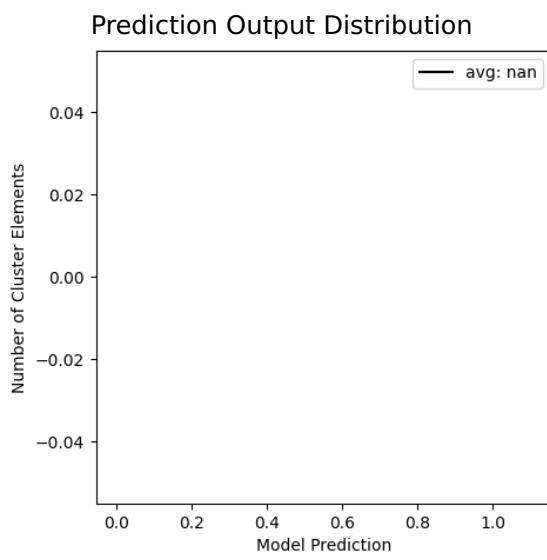
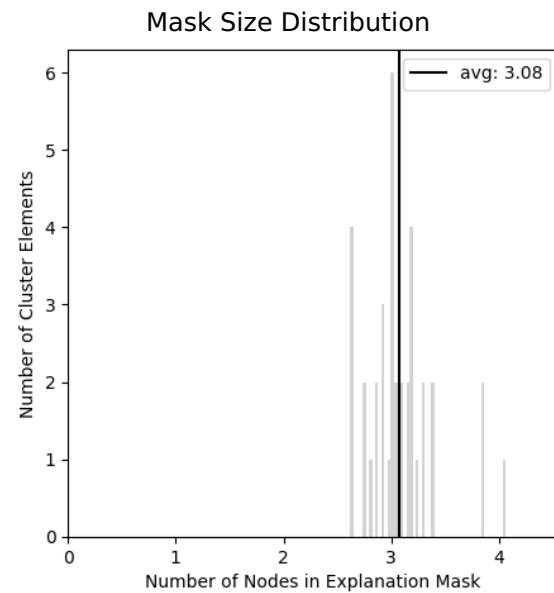
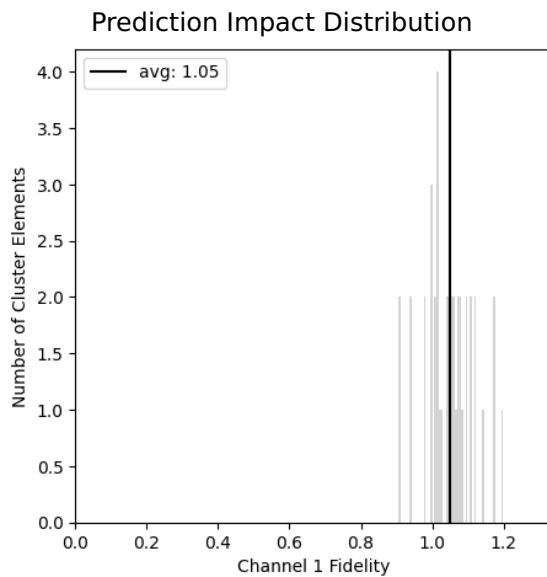
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	39
Channel Index	1.0 (0.0)

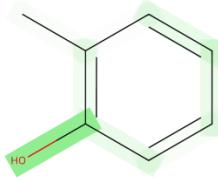
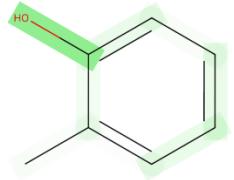
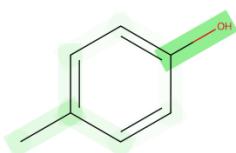
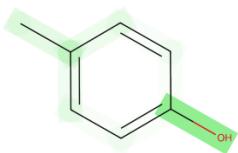
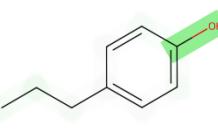
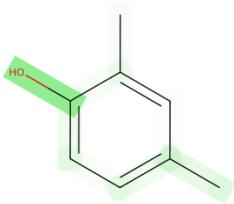
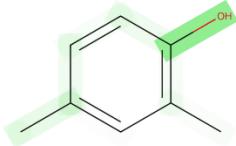
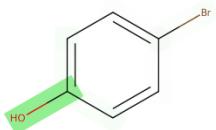
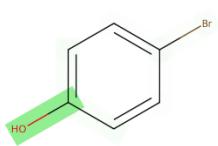
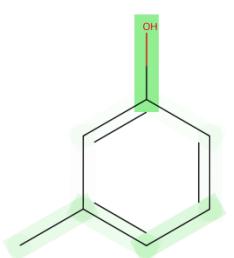
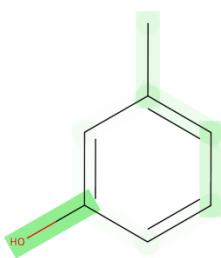
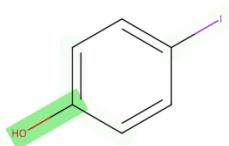
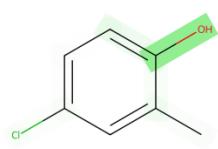
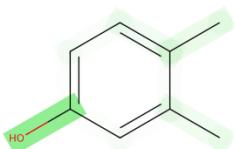
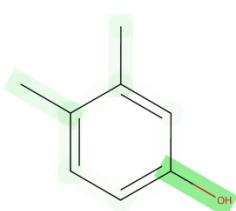
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



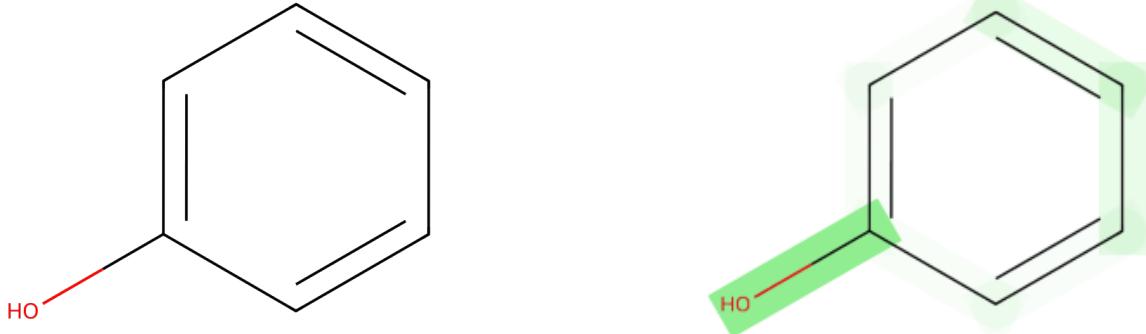
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES represents a phenol substructure, where a hydroxyl group is directly attached to an aromatic benzene ring. The presence of the hydroxyl (-OH) group in phenols increases water solubility due to its ability to form hydrogen bonds with water molecules. The ring itself is hydrophobic but the -OH group's polarity dominates the interaction with water, making the compound more soluble than the benzene ring alone.

**Hypothesis:** The phenol substructure increases water solubility due to the hydroxyl group. The aromatic ring provides a relatively hydrophobic backdrop, but the presence of the polar hydroxyl group enables the formation of hydrogen bonds with water, enhancing solubility.

# Cluster #49 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 49, from importance channel 1 (*positive*), represents a motif consisting of 3.2 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 0.6 ( $\pm 0.0$ ) on the prediction outcome.

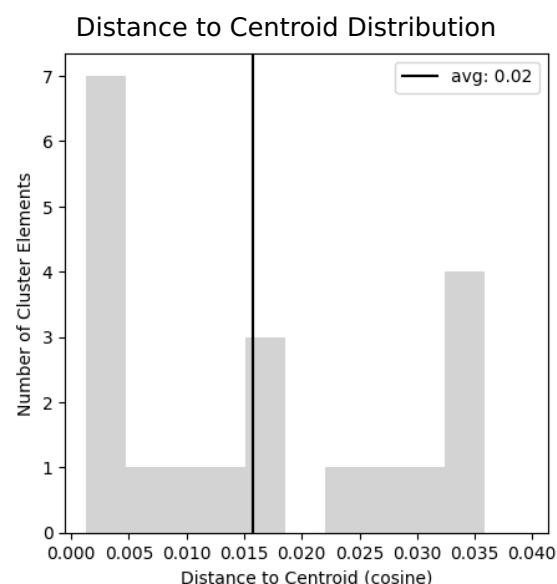
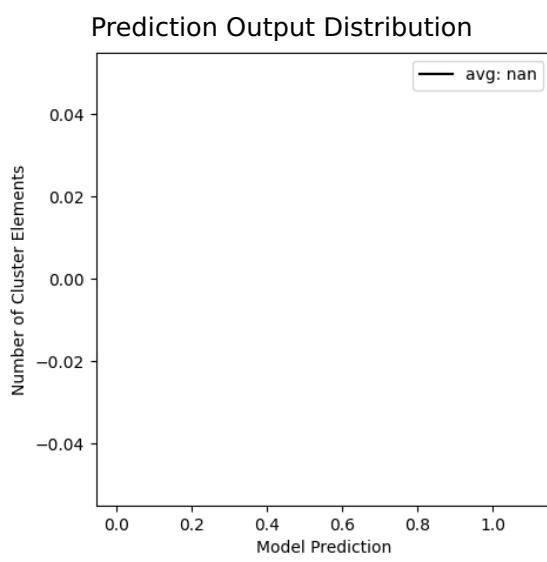
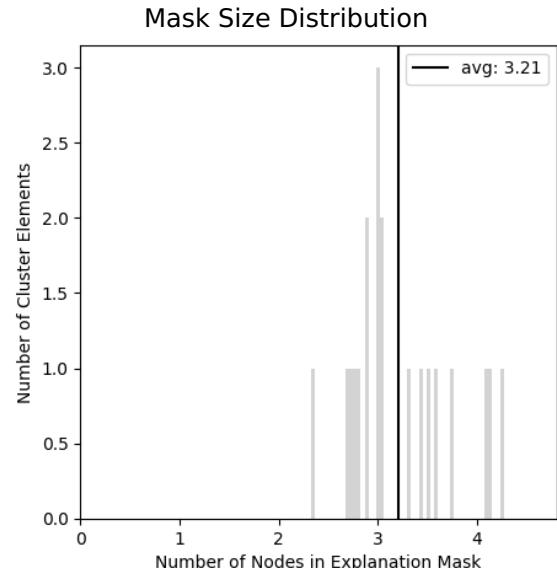
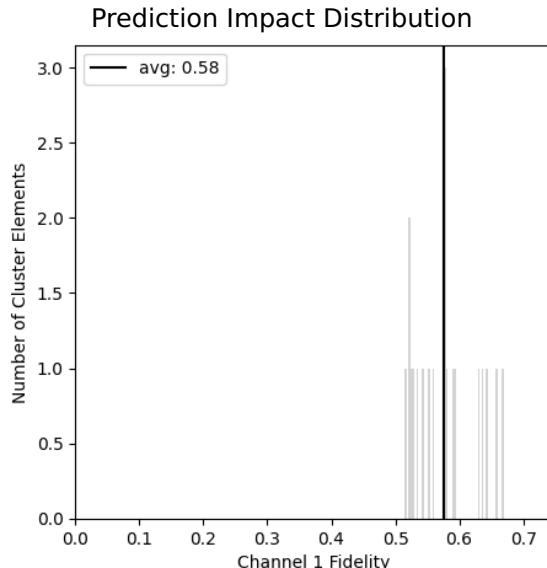
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	20
Channel Index	1.0 (0.0)

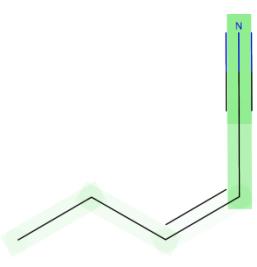
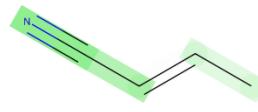
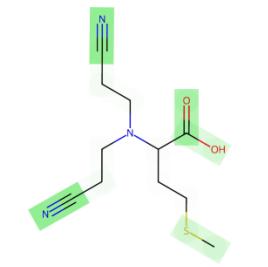
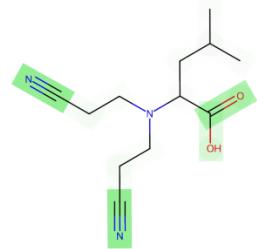
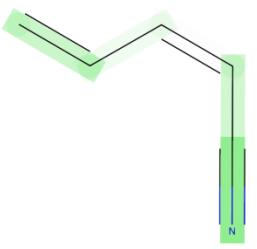
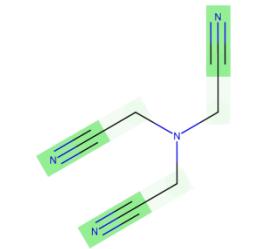
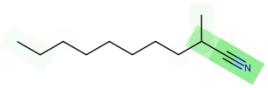
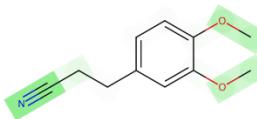
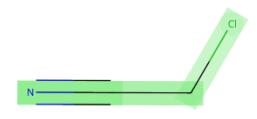
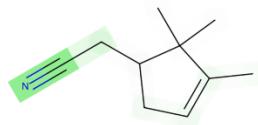
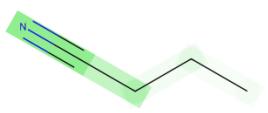
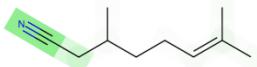
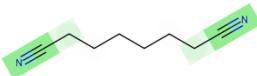
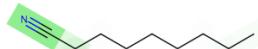
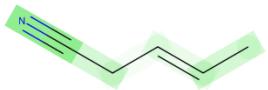
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The "C#N" structure represents a nitrile group, which is a carbon triple-bonded to a nitrogen. Nitrile groups are polar due to the electronegativity difference between carbon and nitrogen, which can form dipole-dipole interactions with water molecules. However, the influence towards water solubility is moderate because, while the nitrile group can engage in some polarity-based interactions, it lacks hydrogen bond donors that would enhance solubility more significantly.

**Hypothesis:** The nitrile group's moderate polarity contributes moderately towards water solubility. The polar carbon-nitrogen triple bond allows for dipole-dipole interactions with water, partially overcoming hydrophobic tendencies and increasing solubility, but the absence of hydrogen bonds limits this effect.

# Cluster #50 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 50, from importance channel 1 (*positive*), represents a motif consisting of 2.7 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 0.6 ( $\pm 0.1$ ) on the prediction outcome.

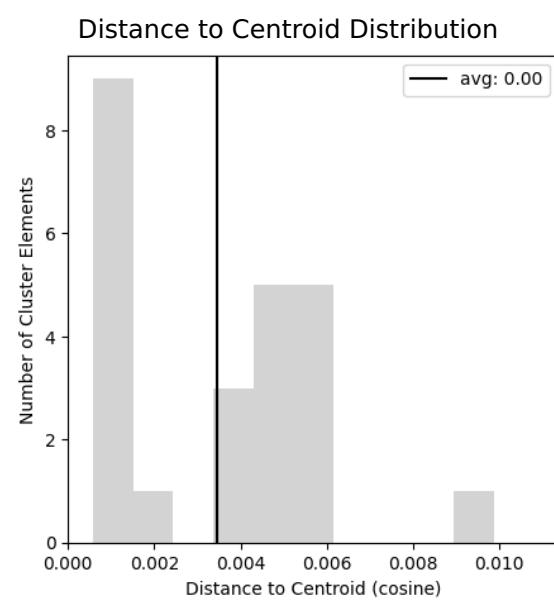
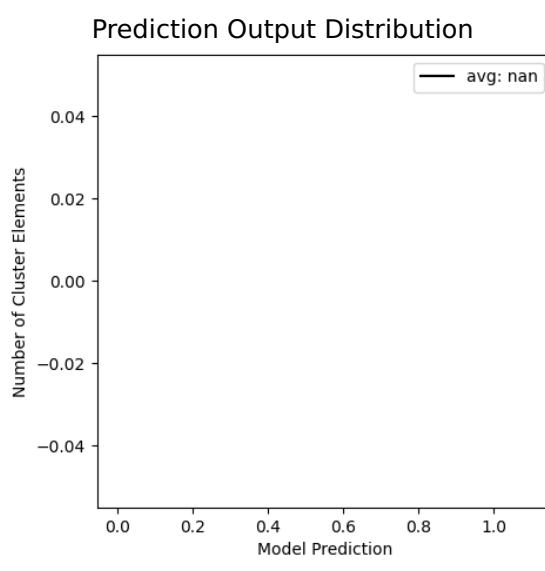
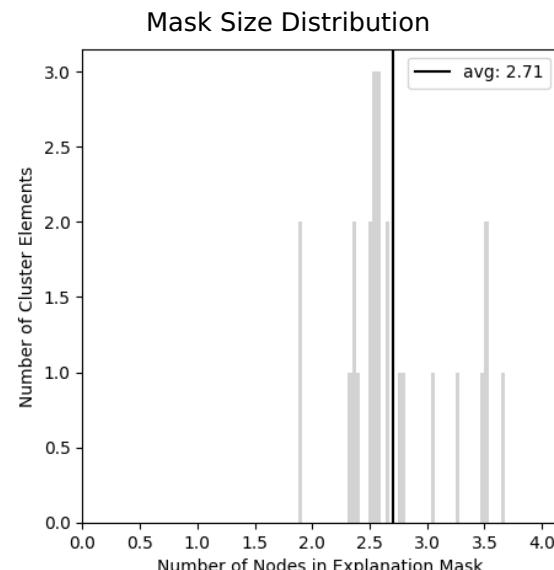
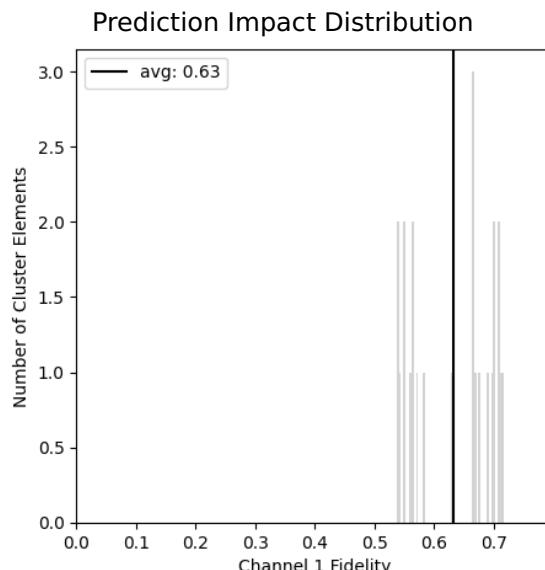
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	24
Channel Index	1.0 (0.0)

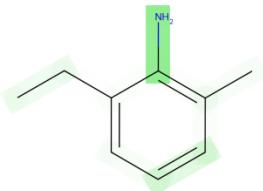
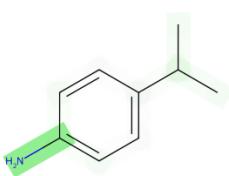
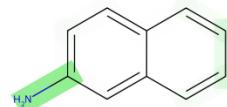
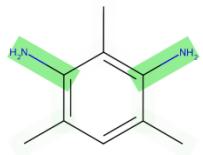
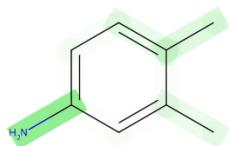
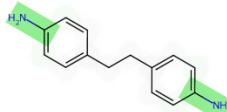
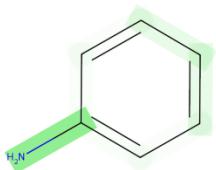
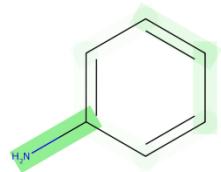
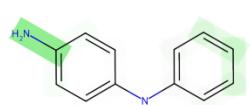
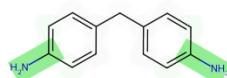
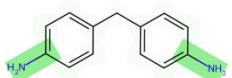
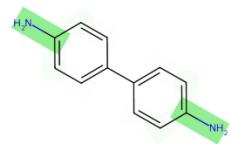
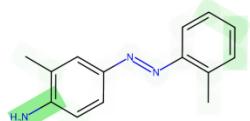
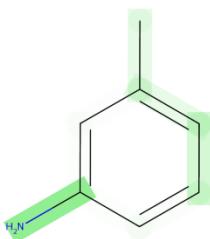
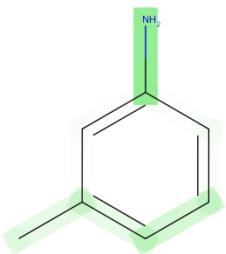
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



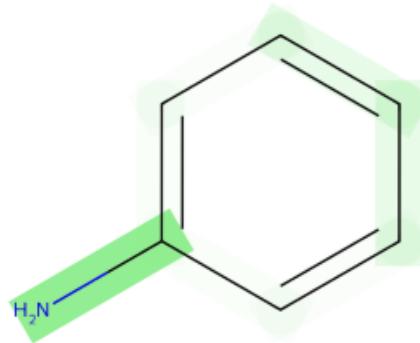
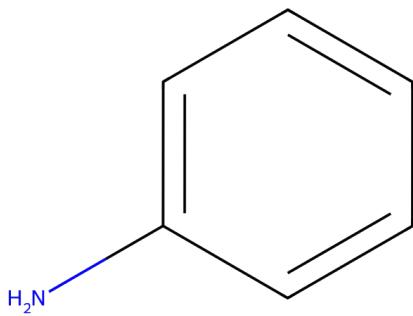
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The presence of a nitrogen atom directly connected to an aromatic ring, as depicted by the SMILES representation "N-c1:c:c:c:c:1", can play a significant role in enhancing the water solubility of a molecule. This is due to the ability of the nitrogen atom to engage in hydrogen bonding interactions with water molecules. The aromaticity of the ring provides a delocalized system which can further interact with water via  $\pi-\pi$  interactions or dipole-induced dipole interactions, although these are generally weaker compared to hydrogen bonds.

**Hypothesis:** Molecules containing a nitrogen atom connected to an aromatic ring have increased water solubility compared to those without this structural motif. The electron-donating nature of nitrogen allows for hydrogen bond formation with water, while the extended  $\pi$ -system of the aromatic ring offers additional, albeit weaker, interactions with water molecules.

# Cluster #51 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 51, from importance channel 1 (*positive*), represents a motif consisting of 2.6 ( $\pm 0.6$ ) nodes. The concept is generally associated with an impact of 0.6 ( $\pm 0.0$ ) on the prediction outcome.

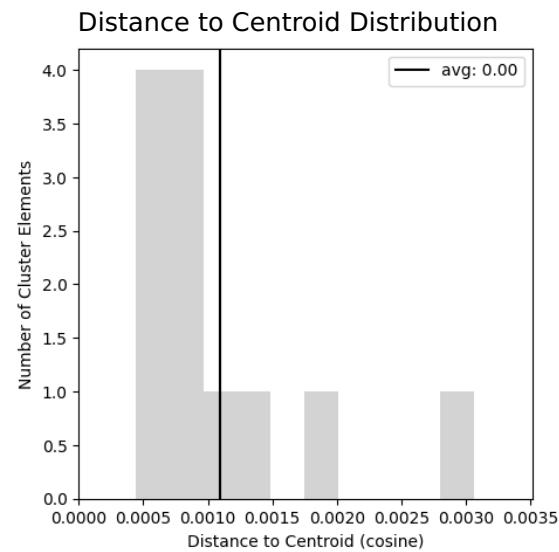
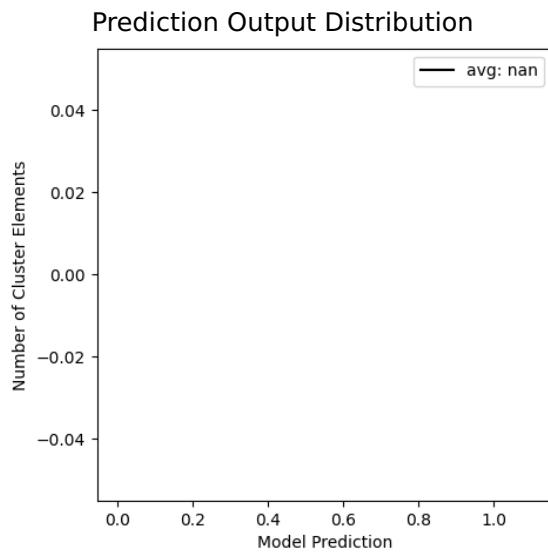
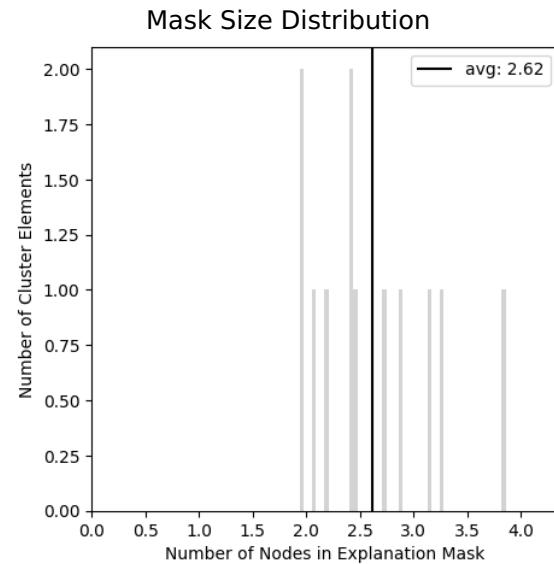
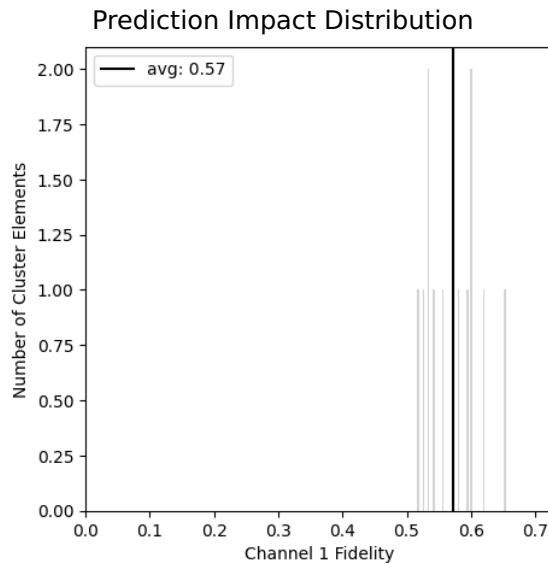
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	12
Channel Index	1.0 (0.0)

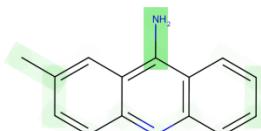
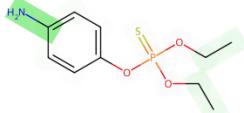
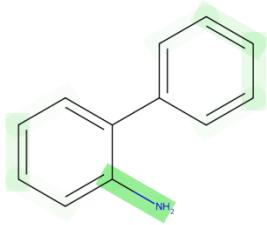
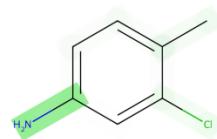
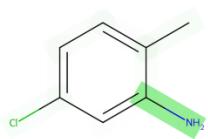
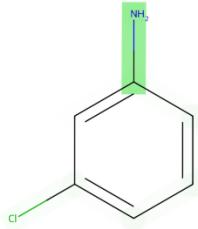
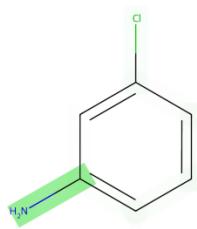
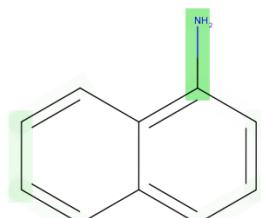
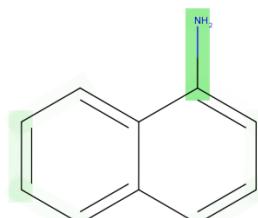
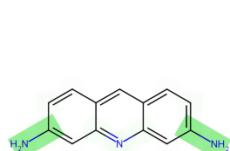
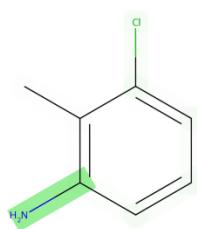
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



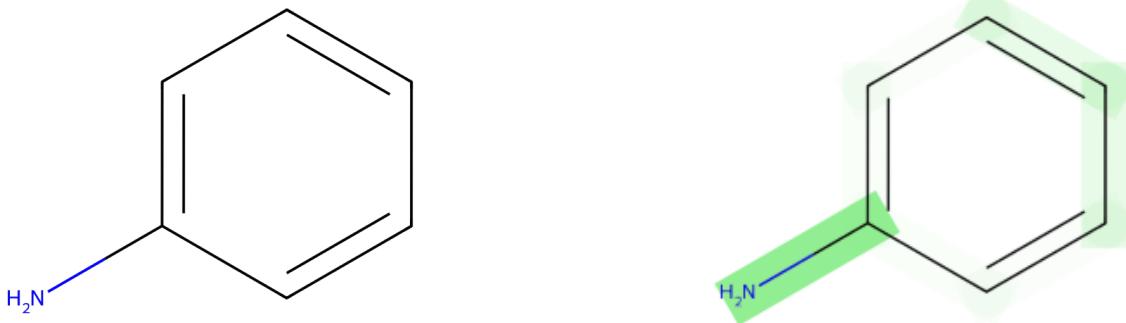
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a pyridine ring, a six-membered aromatic ring with one nitrogen atom substituted for one of the carbon atoms. The presence of a heteroatom, in this case nitrogen, with a lone pair of electrons, increases the polarity of the molecule. This polarity makes the molecule capable of forming hydrogen bonds with water molecules, which is a key interaction that enhances solubility in water. Additionally, the aromatic nature of the pyridine ring may somewhat hinder solubility due to the system's inherent stability and stacking interactions, but the effect of the nitrogen's polarity is evidently stronger in promoting solubility.

**Hypothesis:** Molecules featuring the "N-c1:c:c:c:c:1" substructure have an increased tendency to dissolve in water compared to purely hydrocarbon aromatic rings. The presence of the nitrogen atom contributes to increased polarity and the ability to form hydrogen bonds with water, thereby enhancing the solubility. Although the aromatic character of the ring could have a countering effect, the empirical evidence suggests that the influence of the nitrogen is greater, resulting in a positive influence on water solubility.

# Cluster #52 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 52, from importance channel 1 (*positive*), represents a motif consisting of 3.0 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 0.6 ( $\pm 0.0$ ) on the prediction outcome.

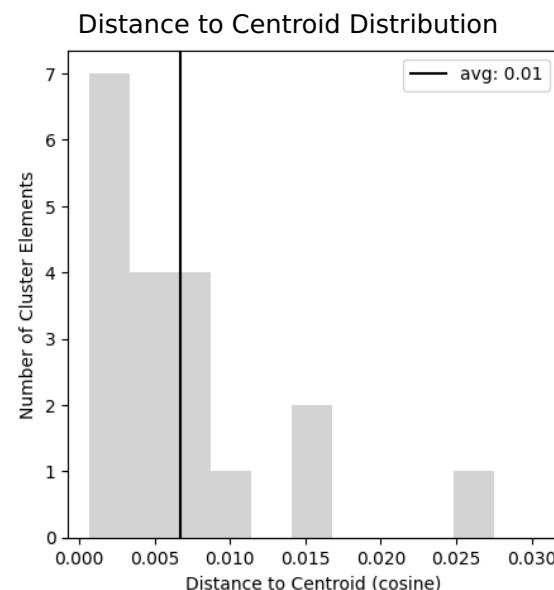
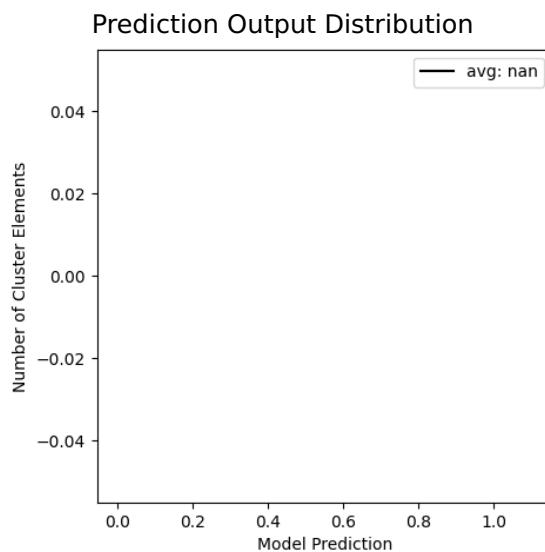
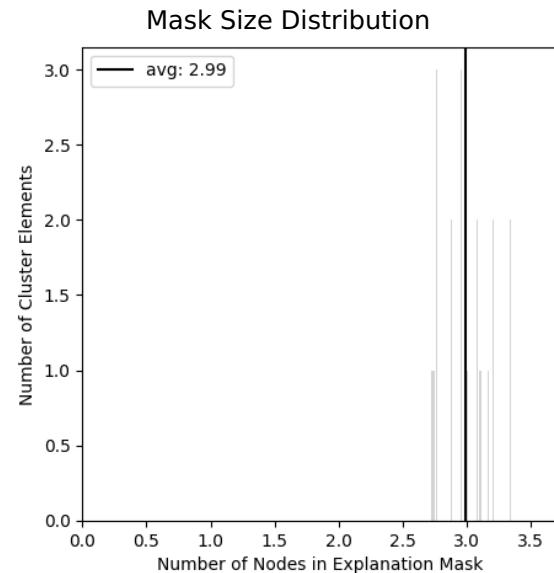
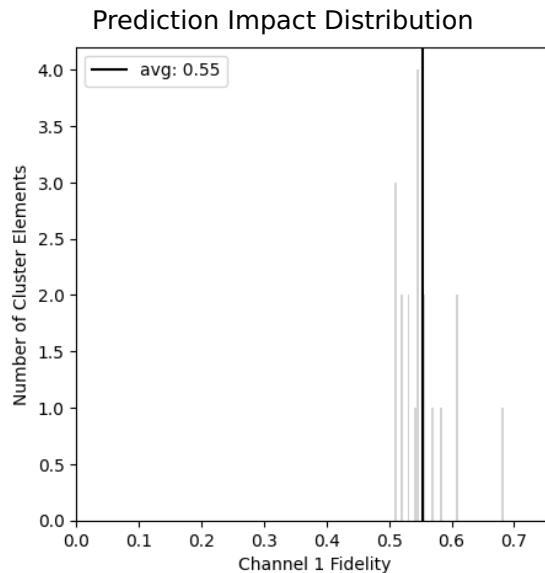
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	19
Channel Index	1.0 (0.0)

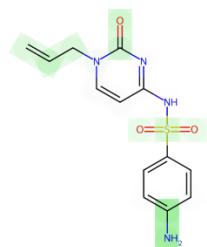
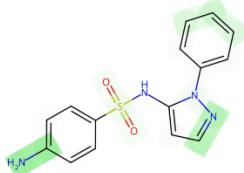
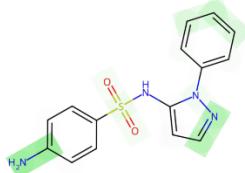
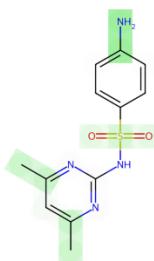
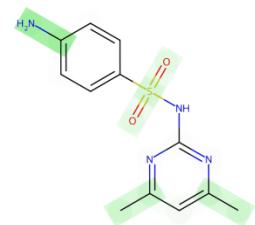
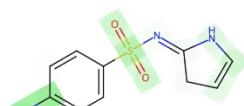
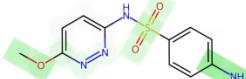
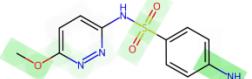
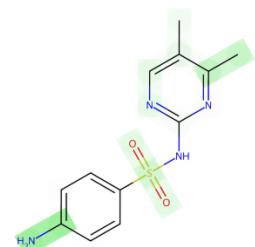
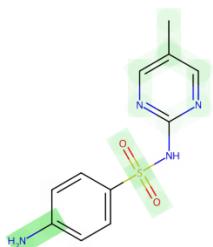
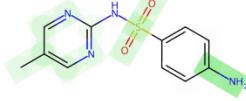
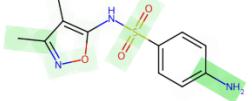
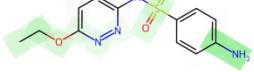
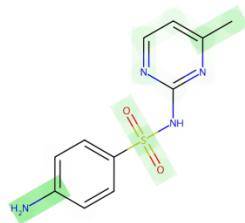
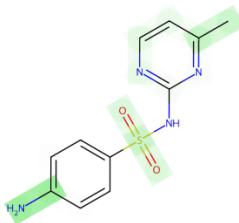
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



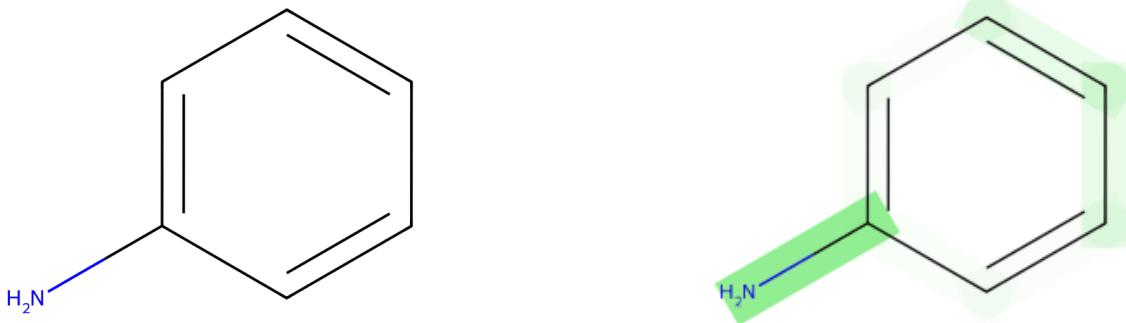
# Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES notation "N-c1:c:c:c:c:1" represents an aromatic ring with a nitrogen atom directly attached to it, which signifies the presence of a pyridine ring. In Pyridine, the nitrogen atom contributes to water solubility due to its ability to engage in hydrogen bonding with water molecules. Nitrogen, being an electronegative atom, draws electron density away from the attached hydrogen atoms in the water, which can form hydrogen bonds with the lone pair electrons of the nitrogen. Additionally, the hybridization of nitrogen in pyridine affords a partial positive charge that can interact with the partial negative charge of the oxygen atoms in water. Aromatics typically decrease water solubility but the presence of a heteroatom like nitrogen in the ring improves solubility compared to benzene due to these polar interactions.

**Hypothesis:** Pyridine substructures increase water solubility due to the presence of the electronegative nitrogen capable of hydrogen bonding. The heteroatom's electronegativity combined with the aromatic system's planarity provides a balance which allows for interaction with water without significantly disrupting solubility like purely hydrocarbon aromatic systems would.

# Cluster #53 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 53, from importance channel 1 (*positive*), represents a motif consisting of 3.1 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 0.6 ( $\pm 0.1$ ) on the prediction outcome.

## Properties

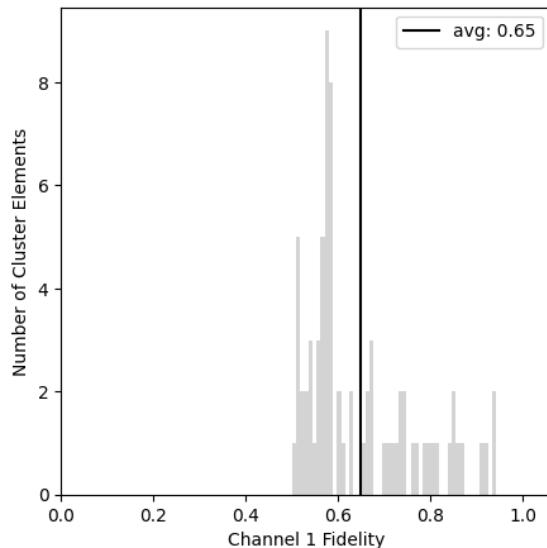
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	73
Channel Index	1.0 (0.0)

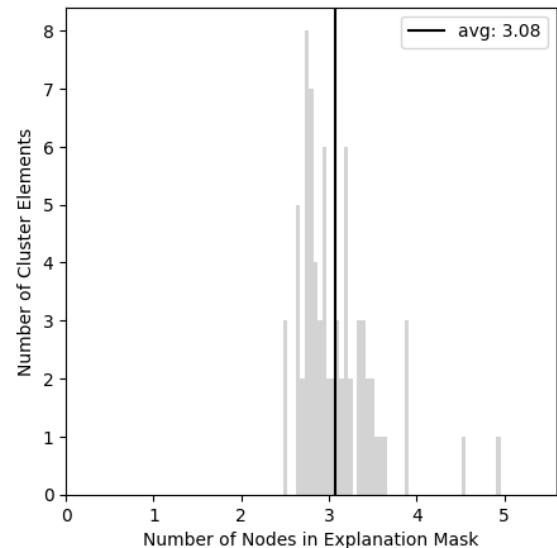
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

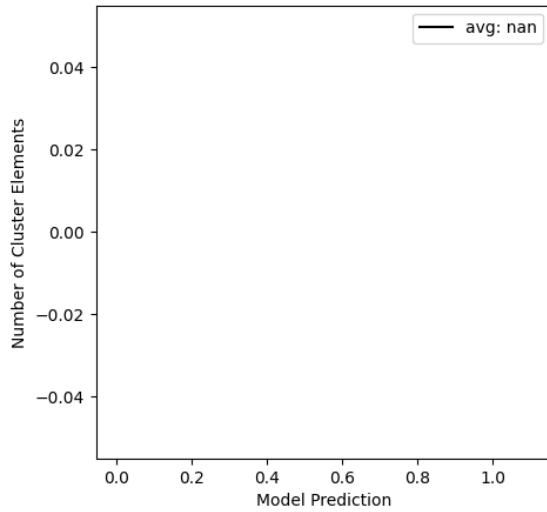
Prediction Impact Distribution



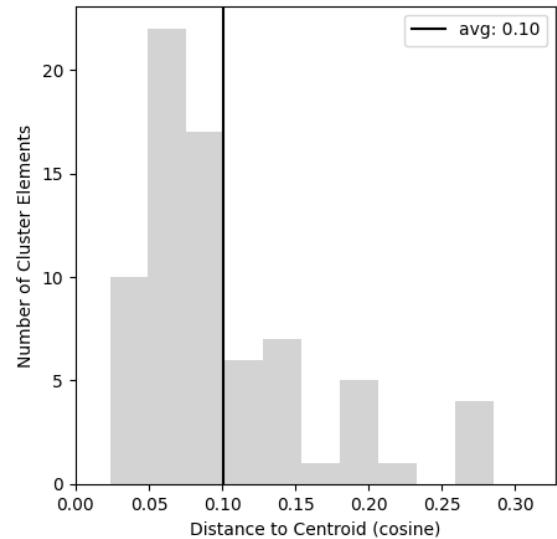
Mask Size Distribution



Prediction Output Distribution

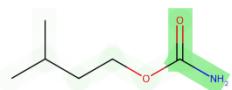
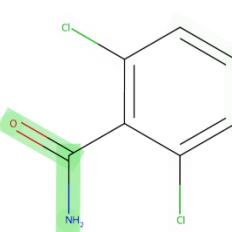
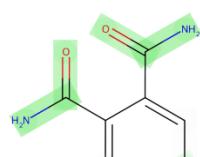
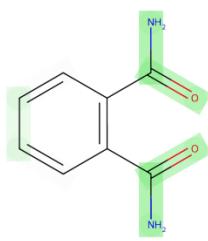
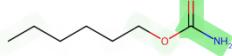
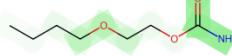
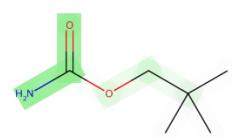
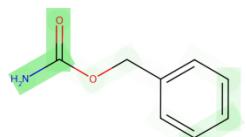
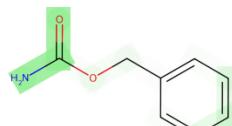
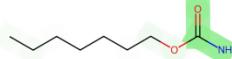
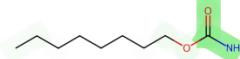
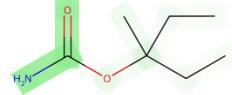
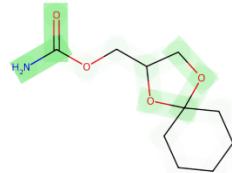
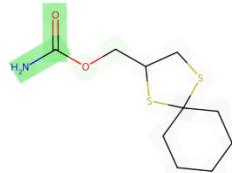
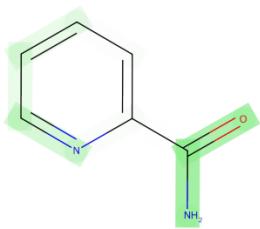


Distance to Centroid Distribution



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES representation "N#CN" indicates a molecule with a cyano group (-C≡N) attached to a nitrogen atom. The cyano group is polar due to the significant electronegativity difference between the carbon and nitrogen atoms, leading to a dipole moment. Additionally, nitrogen atoms can engage in hydrogen bonding with water molecules due to their lone pair of electrons, which significantly enhances water solubility. The combination of these two factors likely explains the relatively high influence on water solubility.

**Hypothesis:** Molecules containing the "N#CN" fragment show enhanced water solubility due to the polar nature of the cyano group and the ability of the nitrogen to form hydrogen bonds with water. This suggests that the electronegativity difference and hydrogen bonding capabilities are the driving forces behind the increased solubility.

# Cluster #54 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 54, from importance channel 1 (*positive*), represents a motif consisting of 3.3 ( $\pm 0.9$ ) nodes. The concept is generally associated with an impact of 0.9 ( $\pm 0.1$ ) on the prediction outcome.

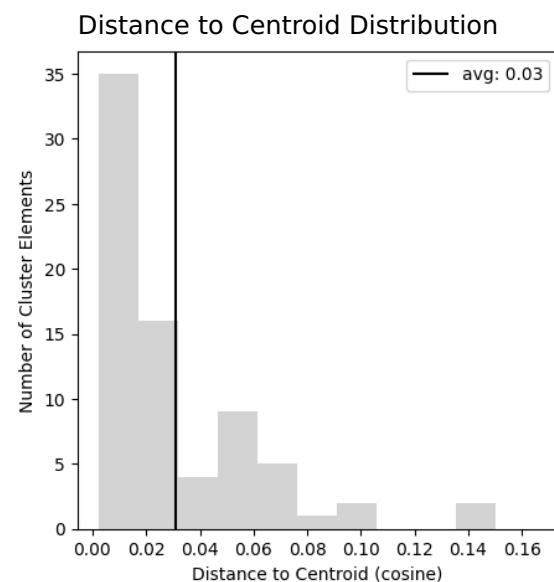
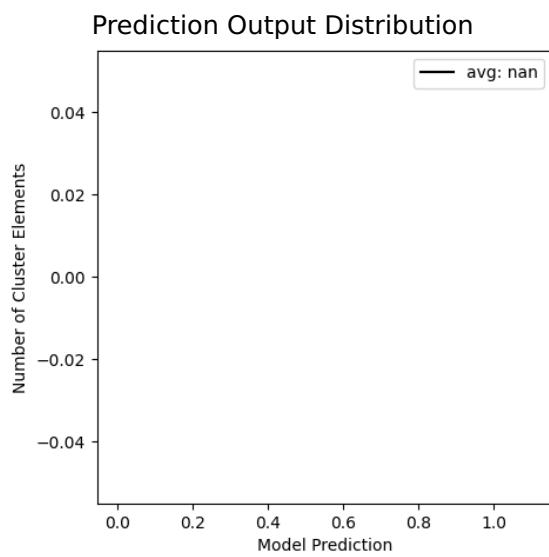
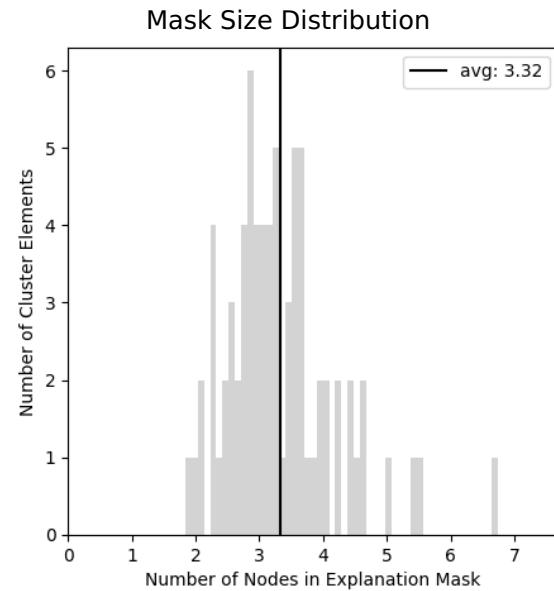
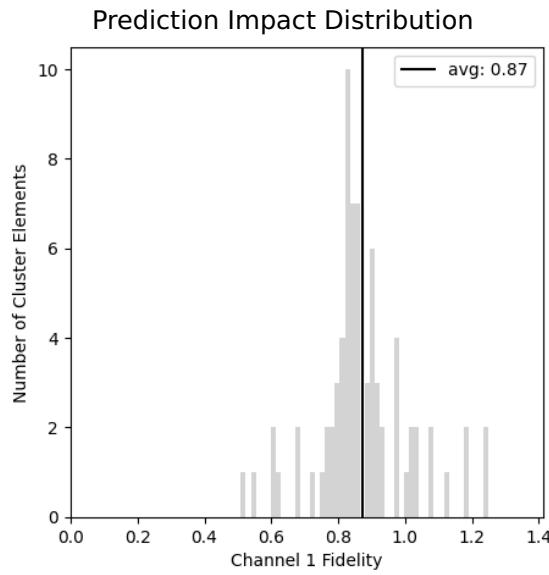
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	74
Channel Index	1.0 (0.0)

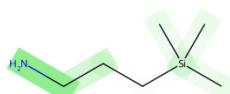
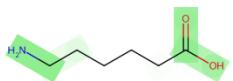
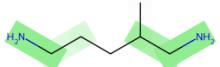
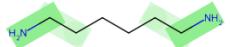
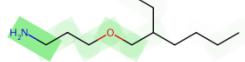
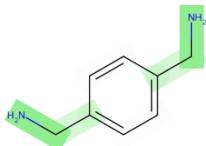
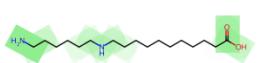
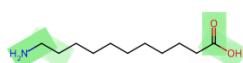
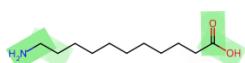
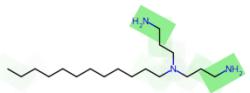
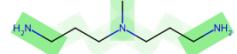
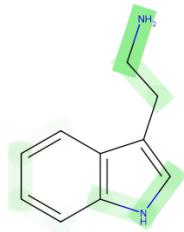
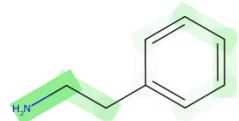
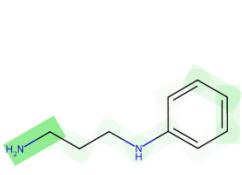
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES fragment "CN" represents a molecule with a cyano group, comprised of a carbon atom triple-bonded to a nitrogen atom. This functional group is known for its polarity due to the significant difference in electronegativity between carbon and nitrogen, which can enhance water solubility because water is also a polar solvent. The polarity allows for favorable dipole-dipole interactions and hydrogen bonding with water molecules, making it energetically favorable for the molecule to dissolve. Furthermore, the small size of the cyano group does not significantly obstruct solvation.

**Hypothesis:** Molecules containing the cyano group ("CN") are likely to exhibit increased water solubility. This is attributed to the cyano group's strong polarity and ability to engage in dipole-dipole interactions as well as hydrogen bonding with water, thereby enhancing its dissolution in the aqueous phase.

# Cluster #55 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 55, from importance channel 1 (*positive*), represents a motif consisting of 3.1 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 1.0 ( $\pm 0.1$ ) on the prediction outcome.

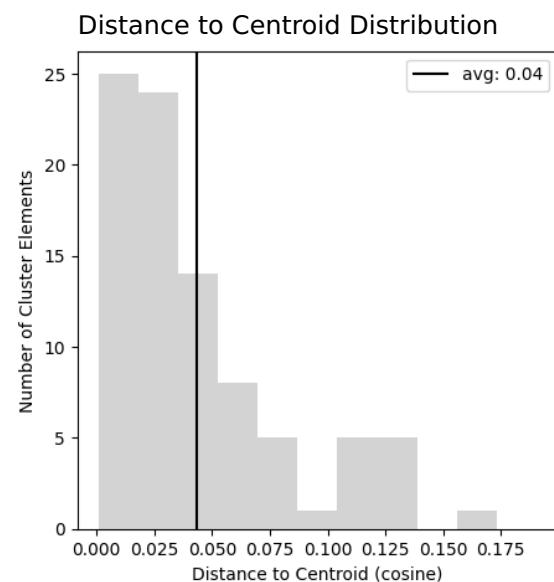
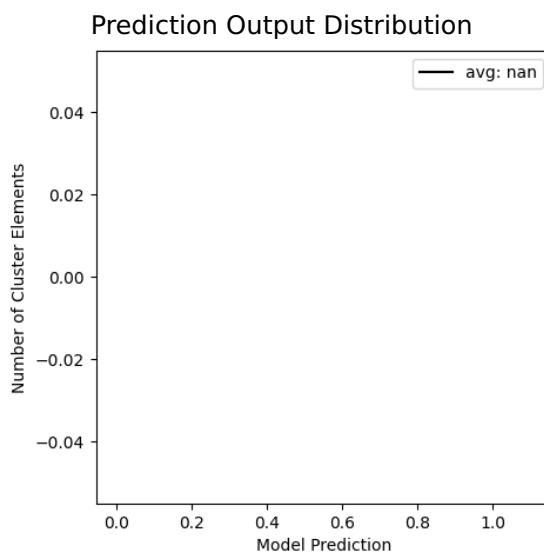
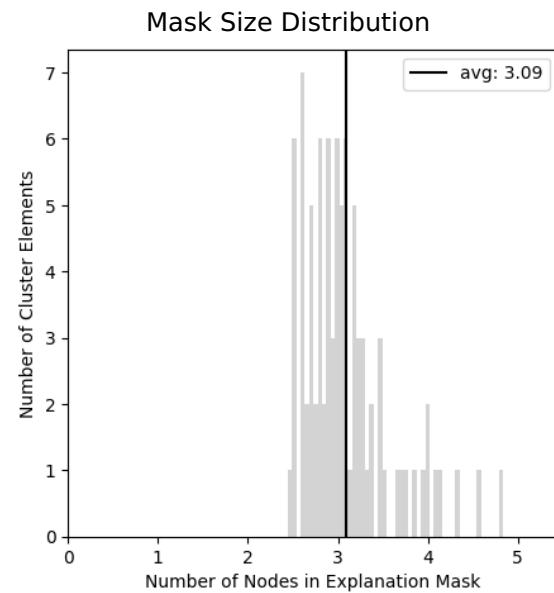
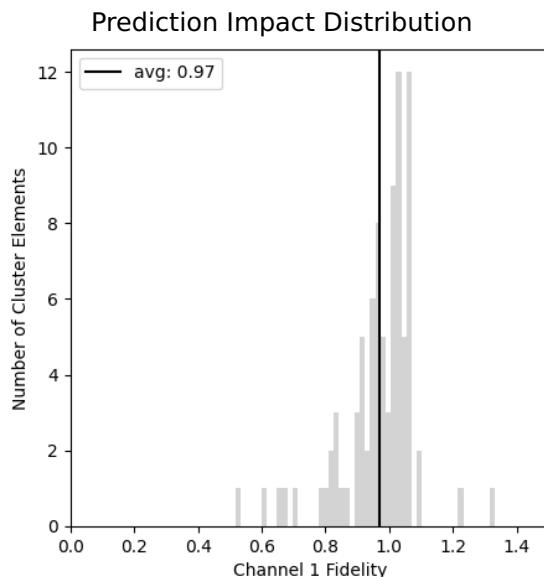
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	88
Channel Index	1.0 (0.0)

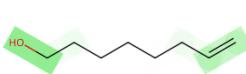
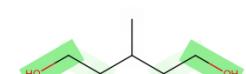
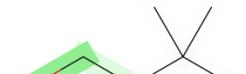
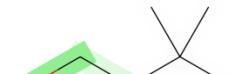
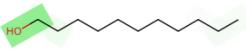
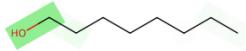
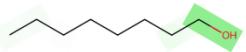
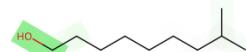
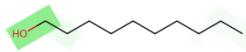
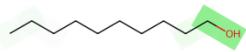
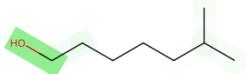
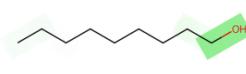
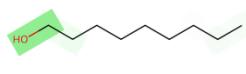
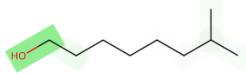
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure depicted by the SMILES notation "C-O" represents a carbon-oxygen single bond, typically found in alcohols, ethers, and esters. This bond is polar due to the electronegativity difference between carbon and oxygen atoms, rendering the molecule soluble in water. The presence of the oxygen atom allows for the formation of hydrogen bonds with water molecules, which is a key interaction that promotes solubility in water.

**Hypothesis:** Molecules containing the "C-O" substructure have a tendency to be soluble in water. The polar nature of the carbon-oxygen bond and the ability to form hydrogen bonds with water molecules are hypothesized to be the driving forces behind the high influence on water solubility.

# Cluster #56 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 56, from importance channel 1 (*positive*), represents a motif consisting of 3.0 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 0.9 ( $\pm 0.1$ ) on the prediction outcome.

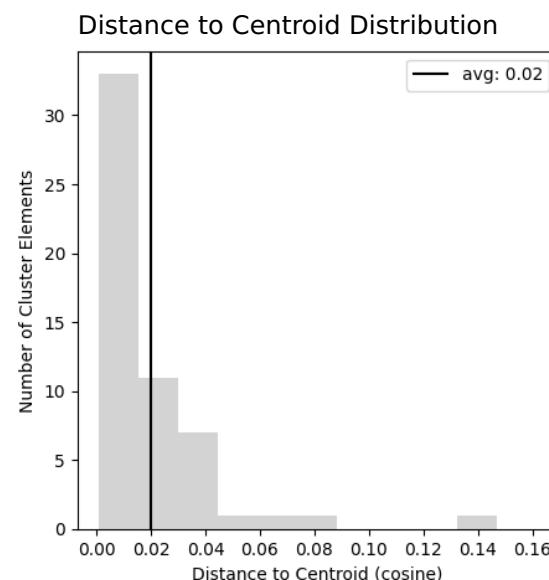
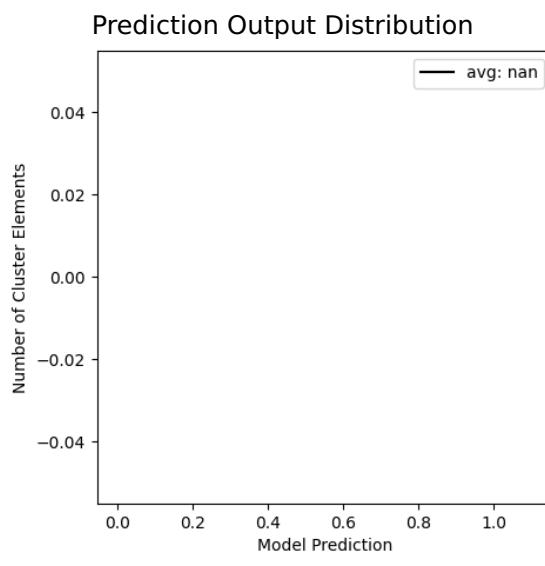
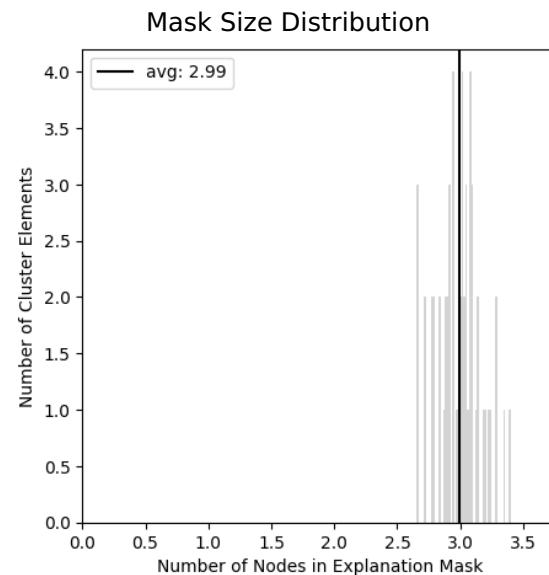
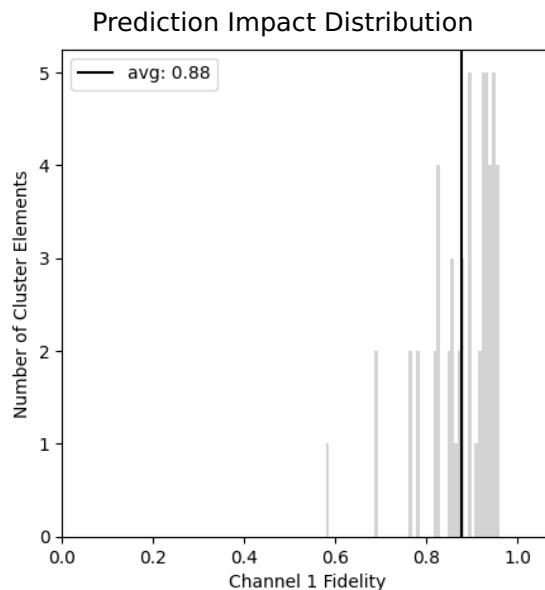
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	55
Channel Index	1.0 (0.0)

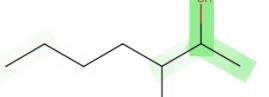
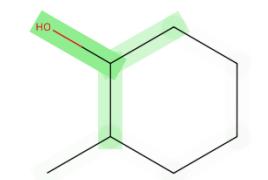
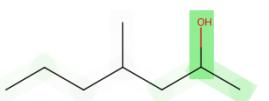
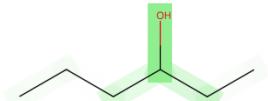
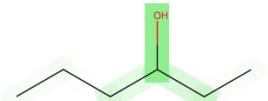
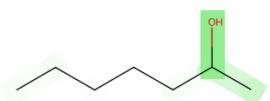
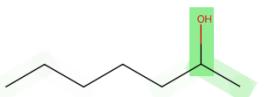
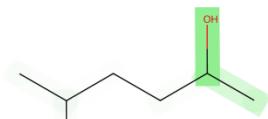
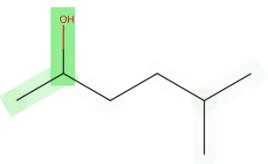
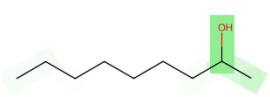
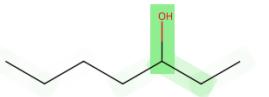
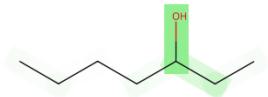
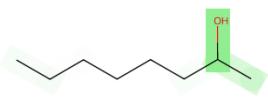
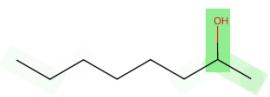
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The "C-C-O" substructure represents a simple alkyl chain ending with a hydroxy group. The alkyl chain (C-C) is nonpolar and would typically decrease solubility in polar solvents like water. However, the presence of the hydroxy group (-O) introduces a polar characteristic to the molecule, which significantly enhances its water solubility. The hydroxy group can form hydrogen bonds with water molecules, which facilitates the dissolution process.

**Hypothesis:** Molecules with the substructure "C-C-O" tend to have increased water solubility due to the hydroxy group's presence. The hydrophobic effect of the alkyl chain is outweighed by the polar, hydrophilic nature of the hydroxy group that enables stronger interaction with water through hydrogen bonding.

# Cluster #57 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 57, from importance channel 1 (*positive*), represents a motif consisting of 2.9 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 0.8 ( $\pm 0.1$ ) on the prediction outcome.

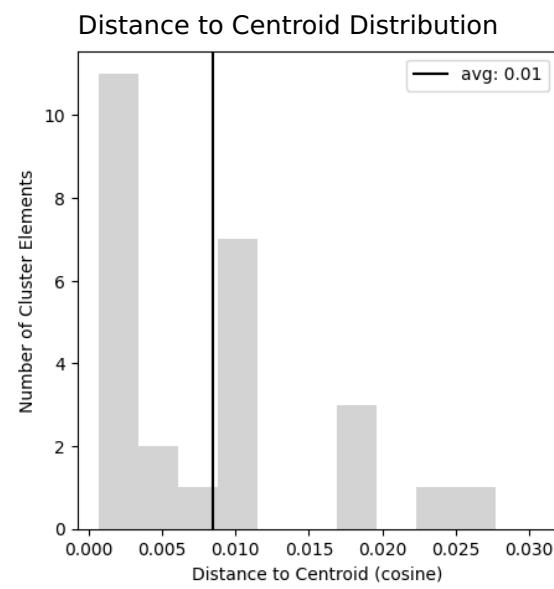
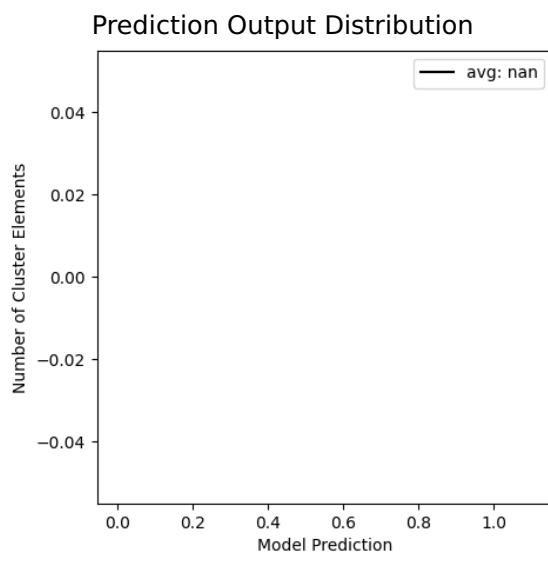
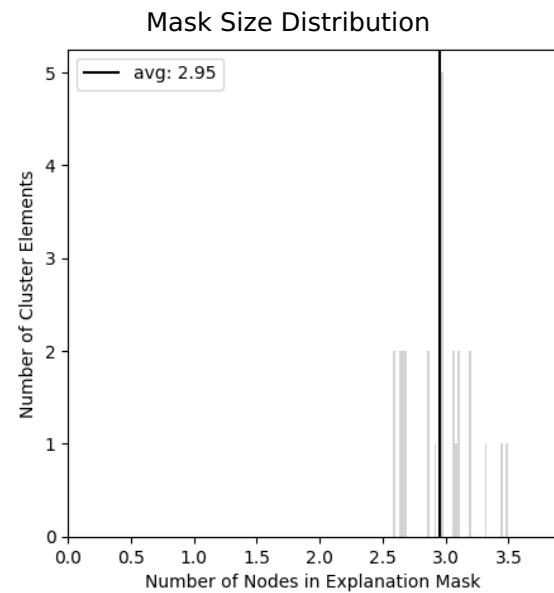
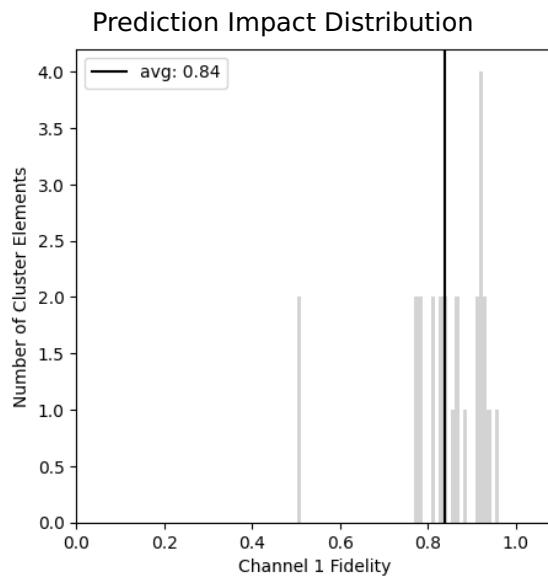
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	26
Channel Index	1.0 (0.0)

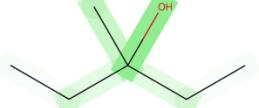
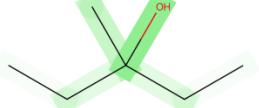
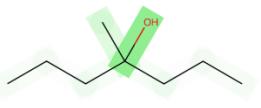
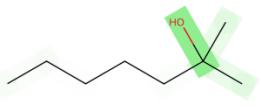
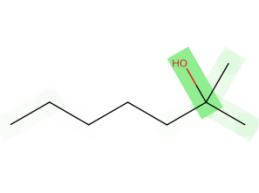
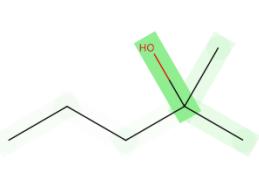
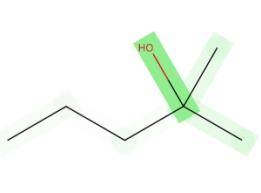
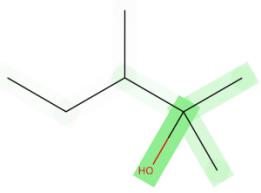
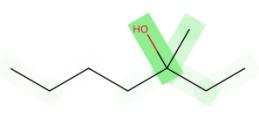
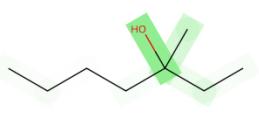
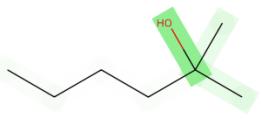
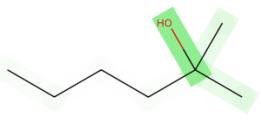
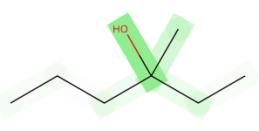
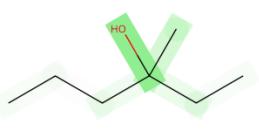
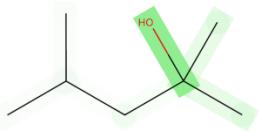
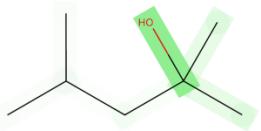
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



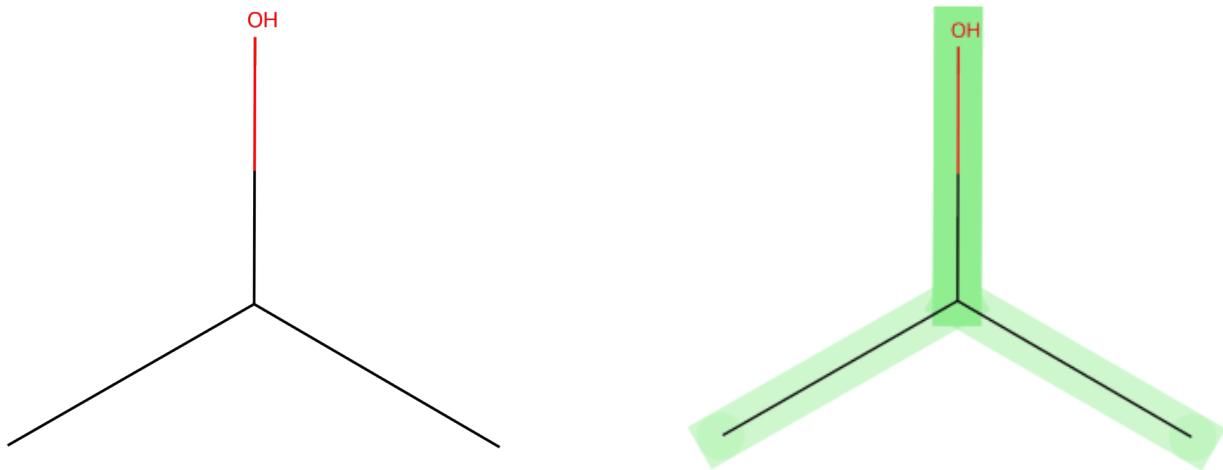
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The presence of -OH (hydroxyl) group in the given molecular fragment plays a critical role in the water solubility of the molecule. This is because the hydroxyl group is capable of forming hydrogen bonds with water molecules, a key interaction that facilitates solubility. Additionally, the branching at the secondary carbon (connected to three other carbons) slightly reduces steric hindrance around the hydroxyl group, potentially allowing for more effective interaction with water molecules compared to a primary carbon which might be more hindered.

**Hypothesis:** Molecules containing the substructure "C-C(-C)-O" are likely to have enhanced water solubility due to the presence of the hydroxyl functional group, which can participate in hydrogen bonding with water. The secondary carbon atom linked to the hydroxyl group likely further assists in this solubility by offering less steric hindrance, facilitating better interaction with water molecules.

# Cluster #58 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 58, from importance channel 1 (*positive*), represents a motif consisting of 3.3 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 0.8 ( $\pm 0.2$ ) on the prediction outcome.

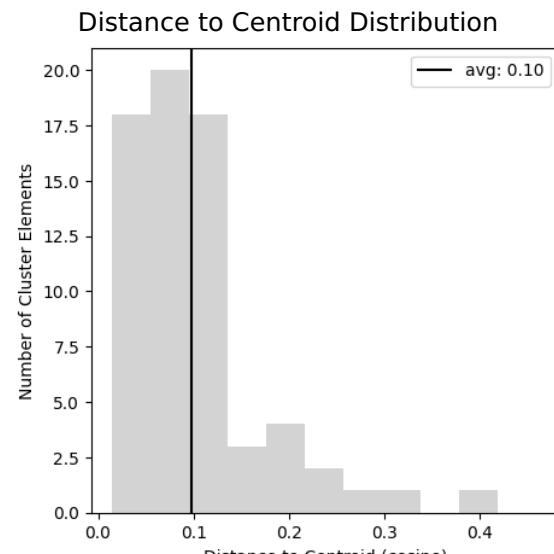
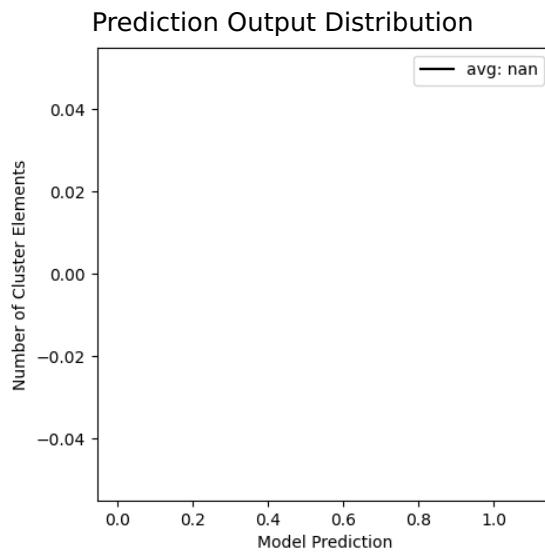
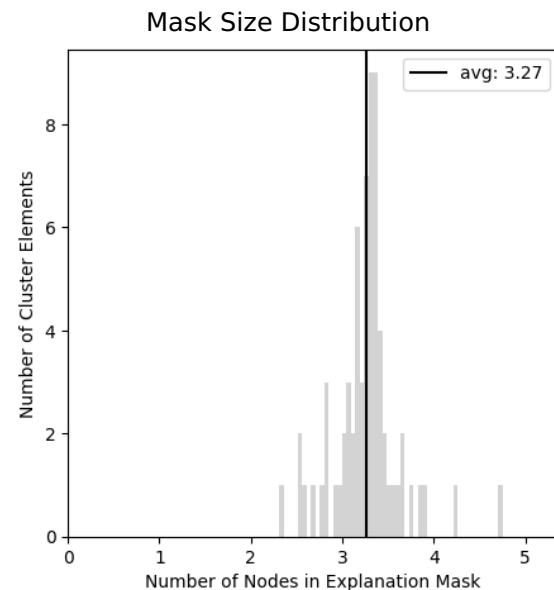
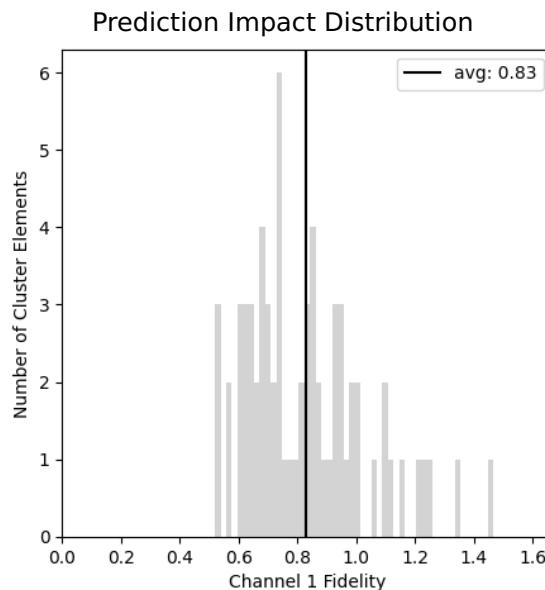
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	68
Channel Index	1.0 (0.0)

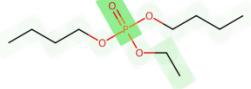
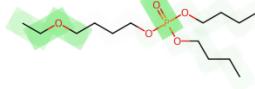
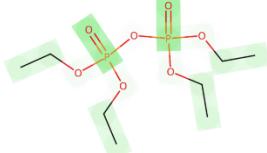
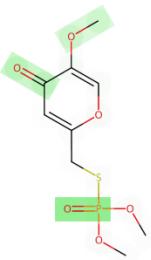
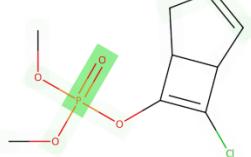
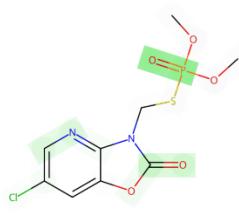
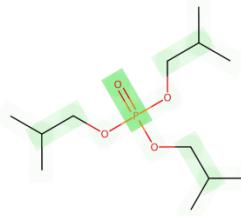
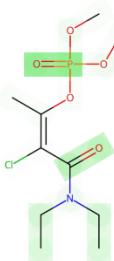
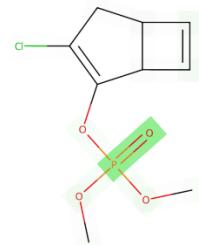
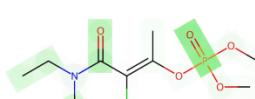
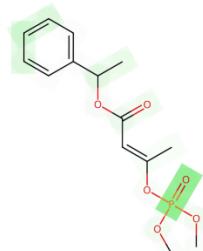
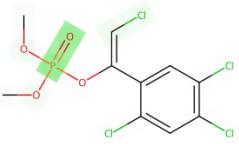
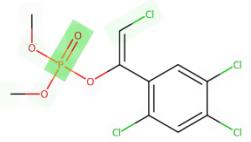
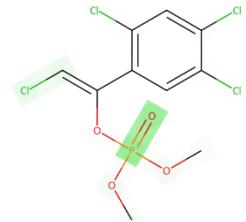
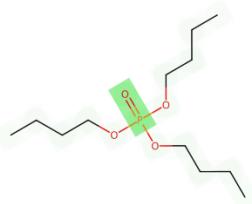
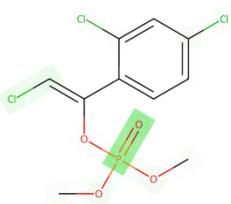
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



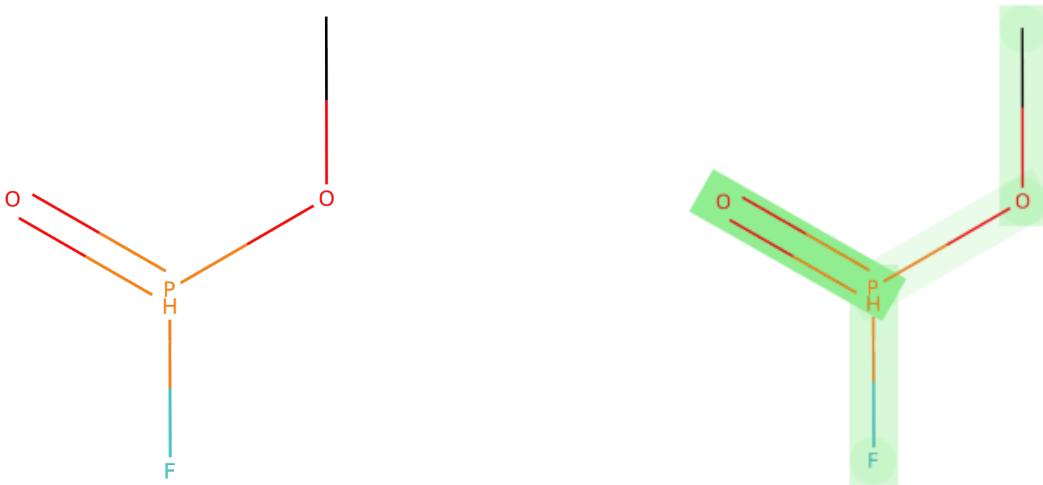
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure "C-O-[PH](=O)-F" contains a hydroxyl group (-OH) attached to a carbon which is in turn connected to a phosphorus atom double-bonded to oxygen and single-bonded to fluorine. Hydroxyl groups are known to engage in hydrogen bonding with water molecules, which significantly enhances water solubility. Additionally, the presence of a phosphoric group adjacent to a fluorine atom may also contribute as it introduces polarity to the molecule, facilitating its interaction with the polar water molecules.

**Hypothesis:** Molecules containing the substructure "C-O-[PH](=O)-F" tend to have increased water solubility. The hydroxyl group likely promotes interactions with water through hydrogen bonding. The phosphorus atom double-bonded to oxygen and single-bonded to fluorine adds polarity to the molecule, thus further increasing its affinity to dissolve in water.

# Cluster #59 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 59, from importance channel 1 (*positive*), represents a motif consisting of 3.1 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 0.8 ( $\pm 0.2$ ) on the prediction outcome.

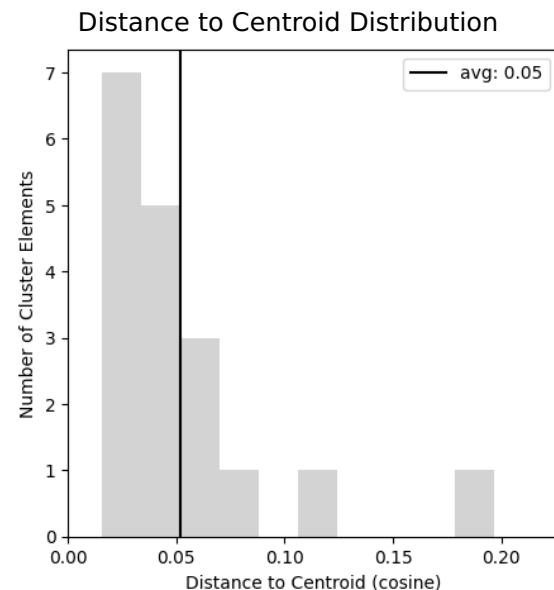
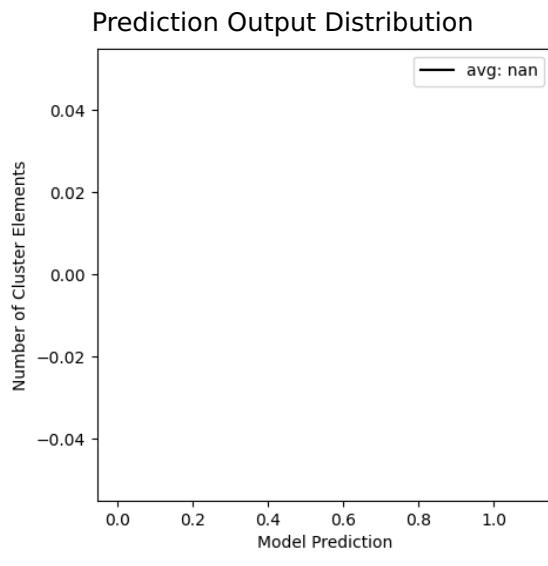
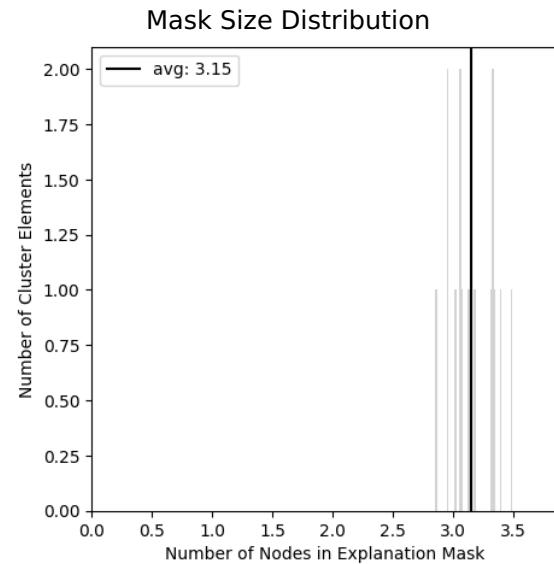
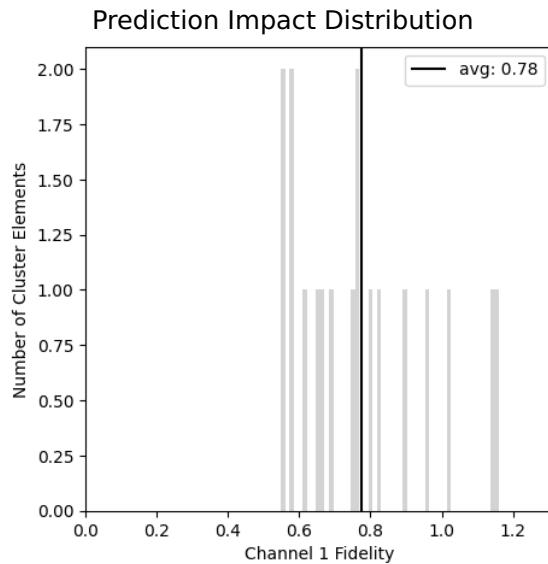
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	18
Channel Index	1.0 (0.0)

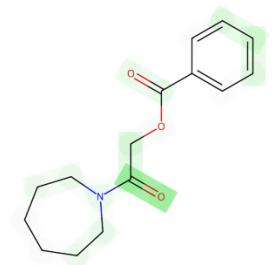
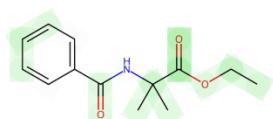
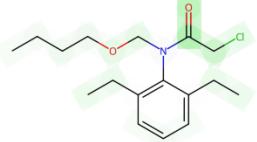
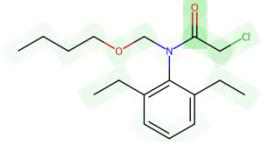
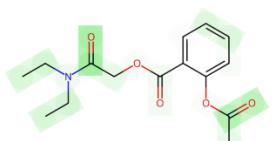
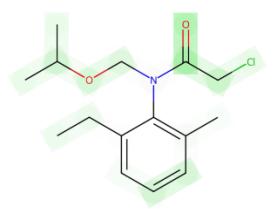
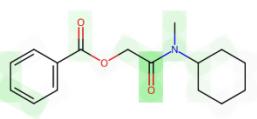
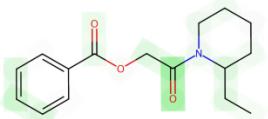
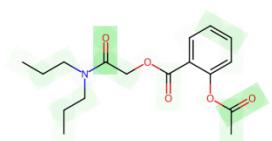
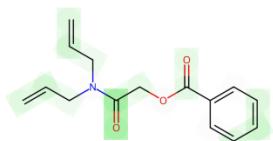
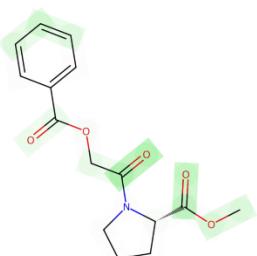
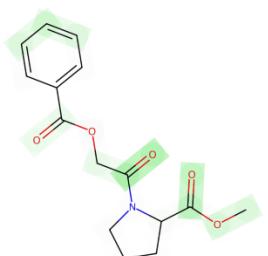
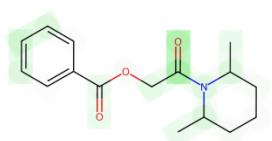
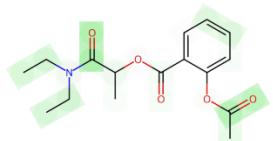
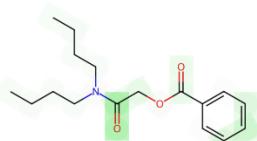
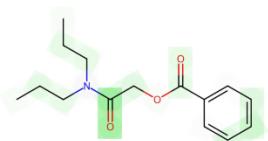
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



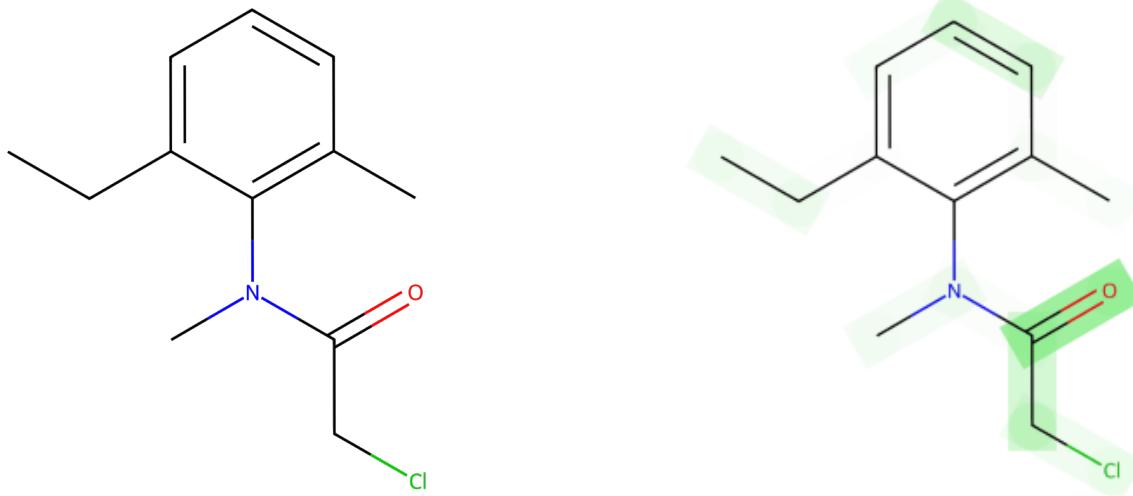
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The substructure presented with the SMILES code "C-C-c1:c:c:c:c(-C):c:1-N(-C)-C(=O)-C-Cl" suggests a chlorinated aromatic compound with an amide functional group. The aromatic ring may provide a level of hydrophobicity, but the electronegative nitrogen and oxygen atoms in the amide group will have the capability to form hydrogen bonds with water molecules. Additionally, the chloroalkyl side chain can also engage in dipole interactions, albeit to a lesser extent due to the lower electronegativity of chlorine compared to oxygen and nitrogen. The overall water solubility is a balance between these contrasting contributions, where the polar amide likely plays a significant role in enhancing solubility.

**Hypothesis:** Molecules containing a chlorinated aromatic ring connected to an amide group (as in the given substructure) will have an intermediate influence on water solubility. The presence of the amide group is expected to enhance solubility due to hydrogen bonding and dipole-dipole interactions, while the hydrophobic aromatic ring and chloro substituent may reduce solubility. The empirical influence of 0.78 towards water solubility suggests that the solubility-enhancing effects of the amide outweigh the hydrophobic effects, but not to the extent of making the molecule highly soluble.

# Cluster #60 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 60, from importance channel 1 (*positive*), represents a motif consisting of 4.3 ( $\pm 1.5$ ) nodes. The concept is generally associated with an impact of 1.0 ( $\pm 0.4$ ) on the prediction outcome.

## Properties

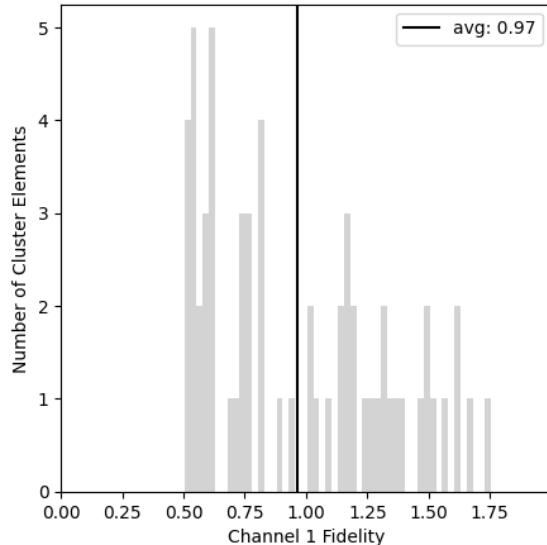
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	61
Channel Index	1.0 (0.0)

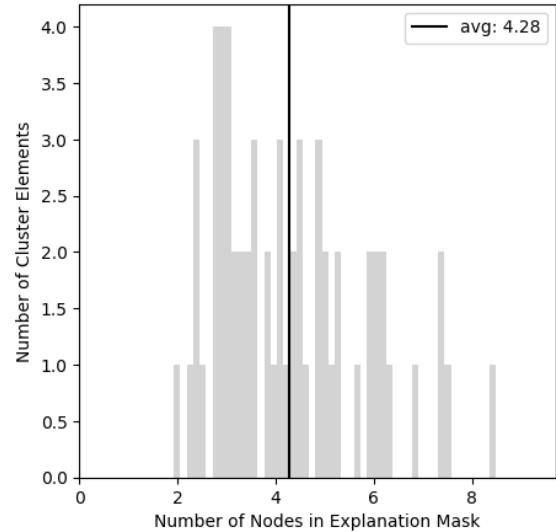
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

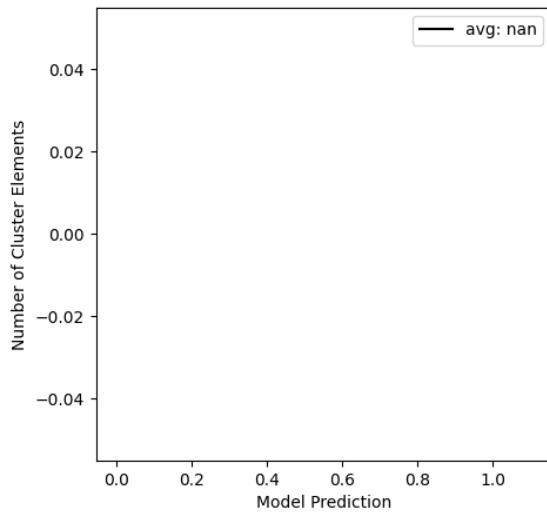
Prediction Impact Distribution



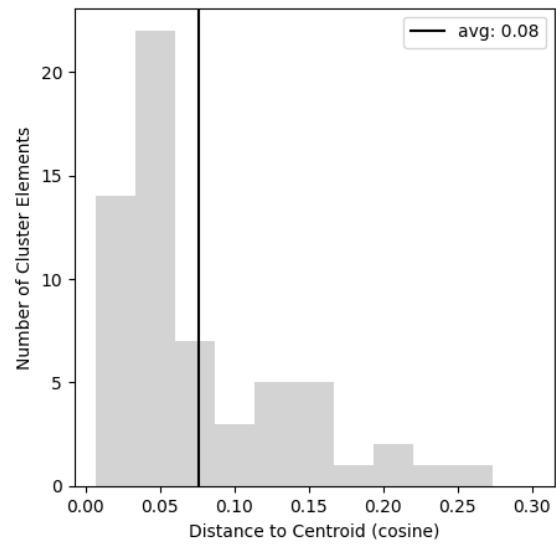
Mask Size Distribution



Prediction Output Distribution

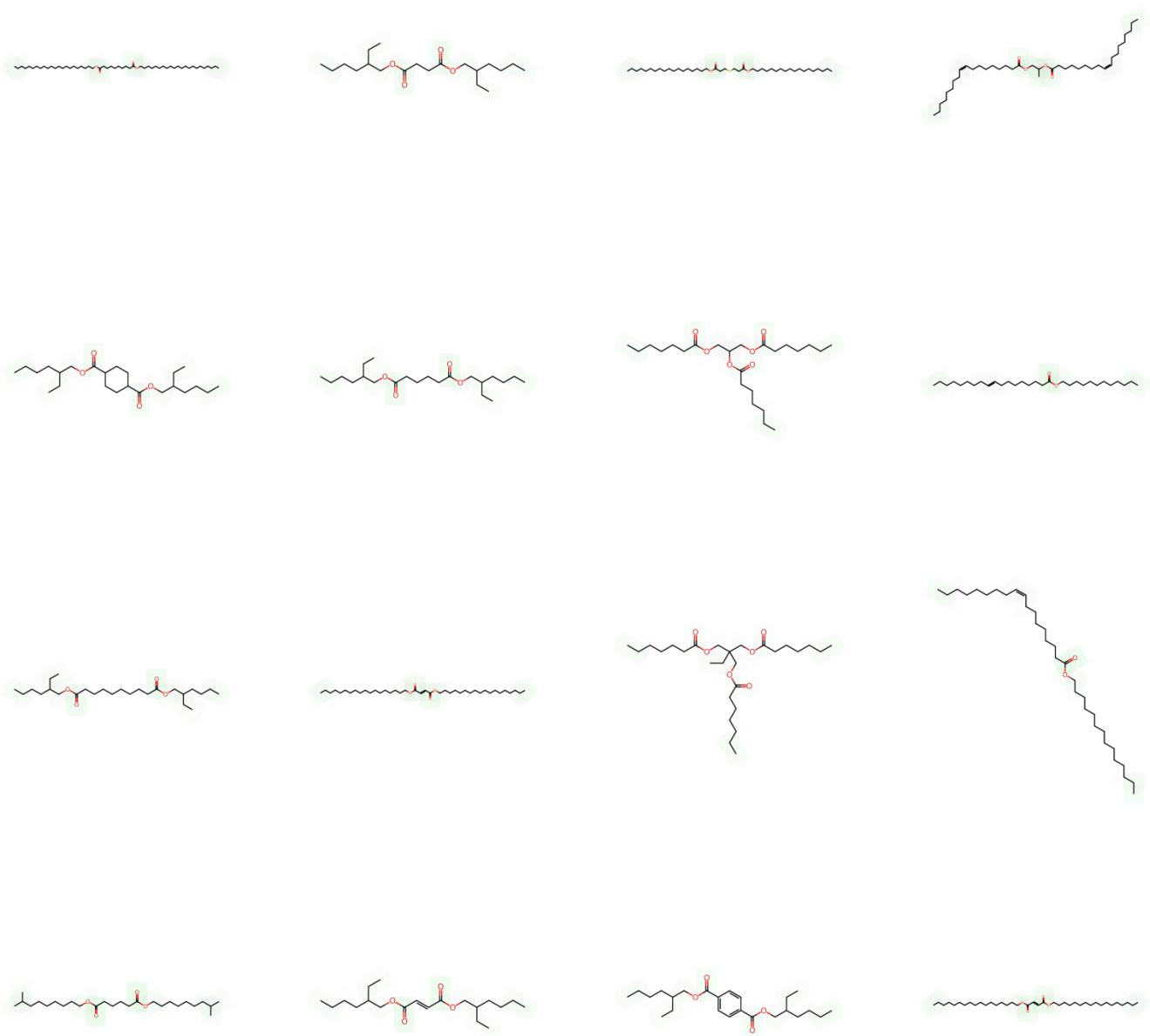


Distance to Centroid Distribution



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The substructure provided suggests an alkane chain with a carboxylic acid moiety. The long hydrocarbon tail (alkane part) is hydrophobic and generally not soluble in water. However, the presence of a carboxylic acid group (-C(=O)OH) indicates potential for hydrogen bonding due to the polar nature of the COOH group. The hydrogen bonding is a key interaction between the molecule and water, which can enhance solubility.

**Hypothesis:** The presence of a carboxylic acid group in a hydrocarbon chain increases water solubility. The hydrophobic alkane chain would tend to decrease solubility, but the carboxylic acid group can engage in hydrogen bonding with water molecules, which typically increases solubility despite the hydrophobic alkane chain.

# Cluster #61 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 61, from importance channel 1 (*positive*), represents a motif consisting of 3.1 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 0.9 ( $\pm 0.3$ ) on the prediction outcome.

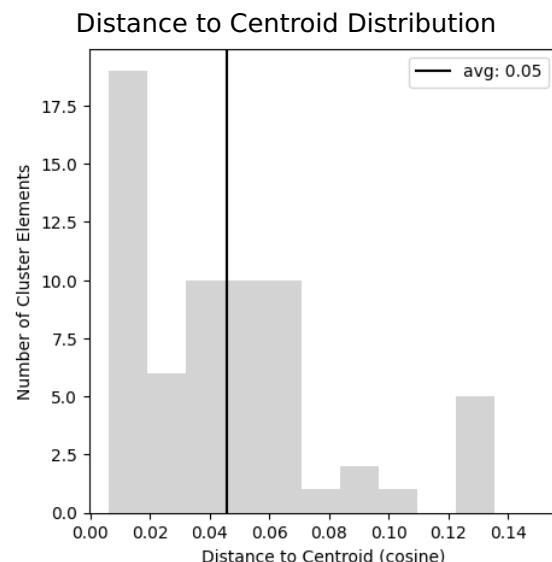
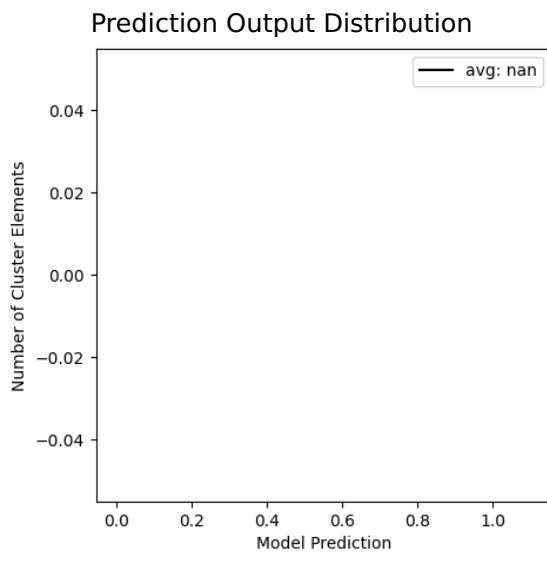
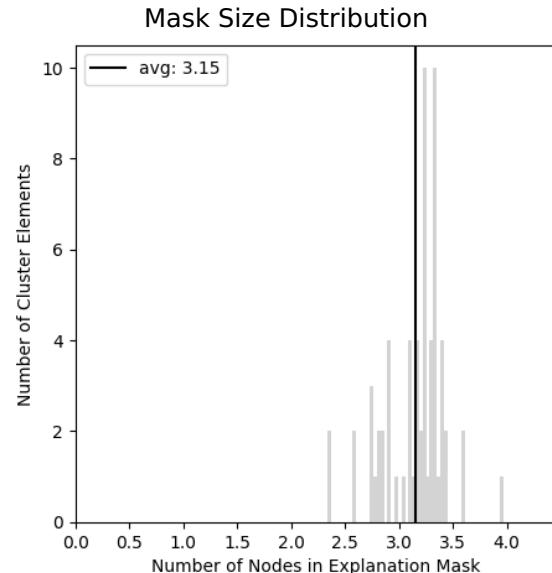
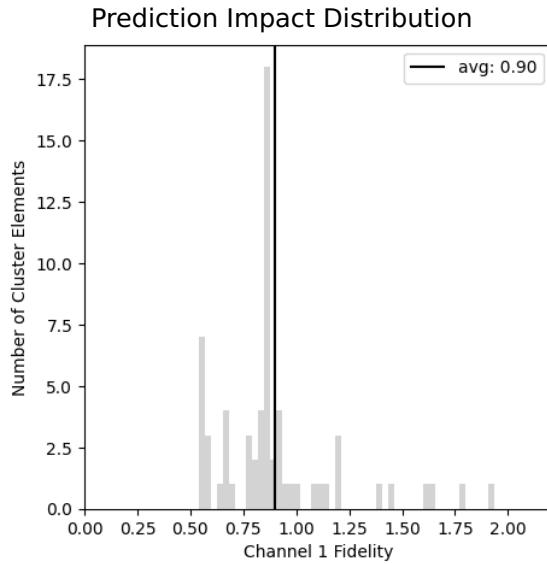
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	64
Channel Index	1.0 (0.0)

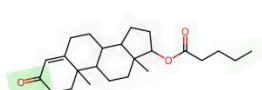
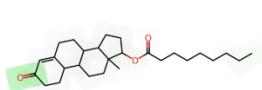
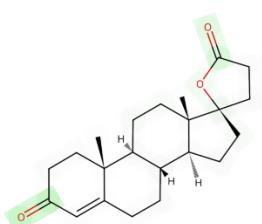
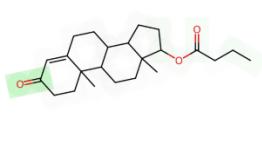
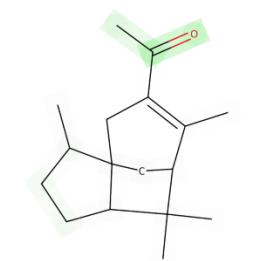
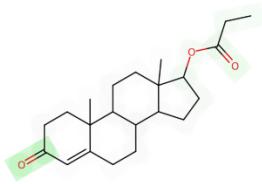
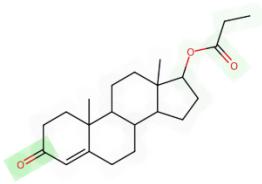
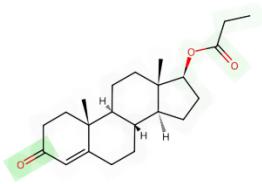
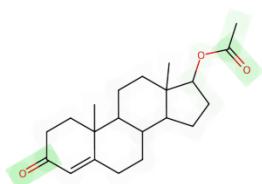
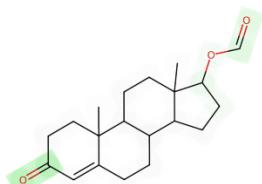
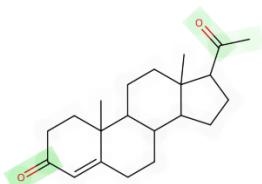
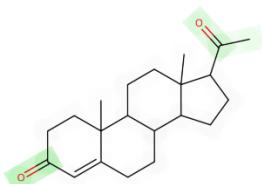
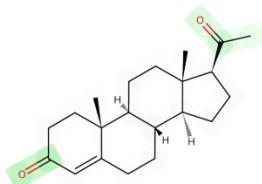
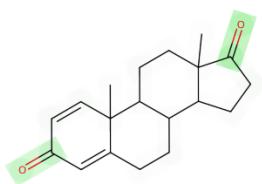
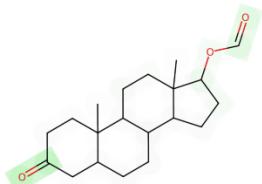
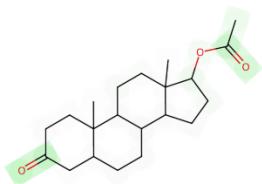
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



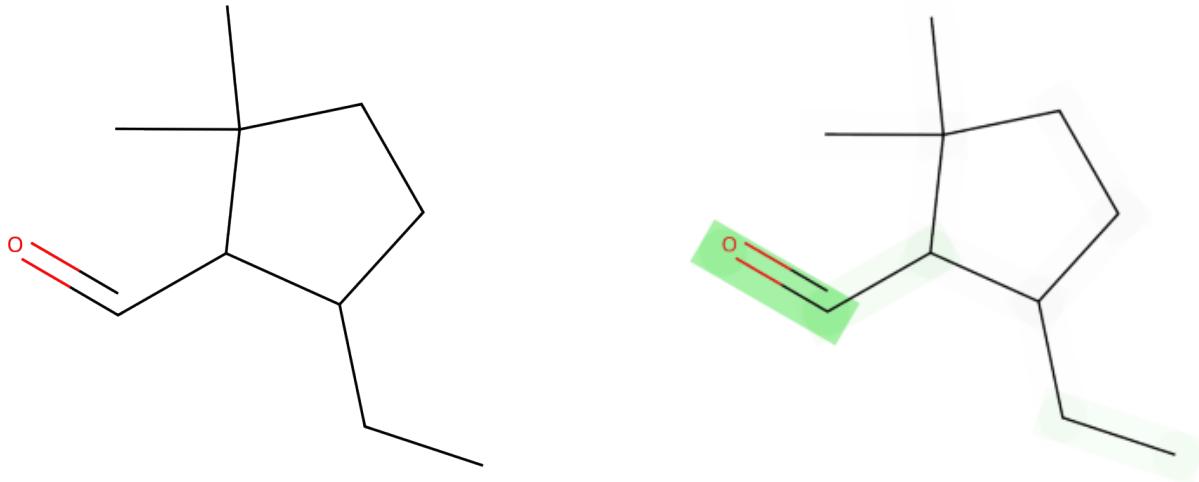
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The provided SMILES representation indicates a molecular substructure with a cyclic and aliphatic character combined with a ketone group. Such a configuration suggests the presence of both hydrophobic (aliphatic) and hydrophilic (ketone) domains within the molecule. The ketone group ( $-C=O$ ) can engage in hydrogen bonding with water molecules due to its polar carbonyl function, enhancing solubility. Meanwhile, the cyclic and branched aliphatic parts might hinder solubility due to their hydrophobic nature. However, given the high influence towards water solubility, it seems the polar interactions with water afforded by the ketone function are the dominant factor in this context.

**Hypothesis:** Molecules with the substructure "C-C-C1-C-C-C(-C)(-C)-C1-C=O" are likely to be relatively soluble in water. The hydrophilic ketone functional group contributes to water solubility through hydrogen bonding interactions. Conversely, the aliphatic chains and cyclic structure may reduce solubility but are overridden by the solubilizing effects of the ketone.

# Cluster #62 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 62, from importance channel 1 (*positive*), represents a motif consisting of 3.3 ( $\pm 0.6$ ) nodes. The concept is generally associated with an impact of 0.8 ( $\pm 0.1$ ) on the prediction outcome.

## Properties

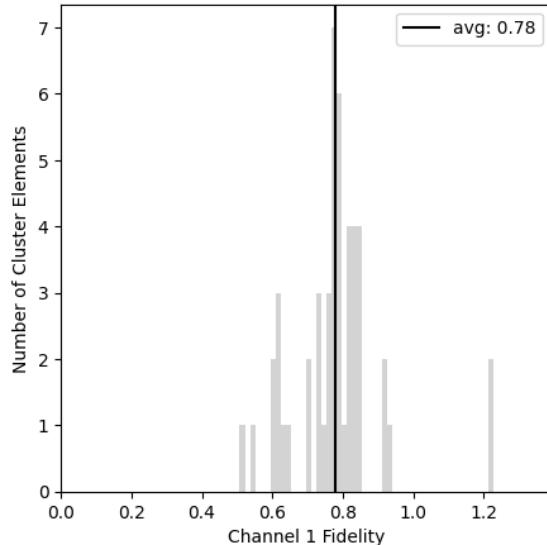
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	49
Channel Index	1.0 (0.0)

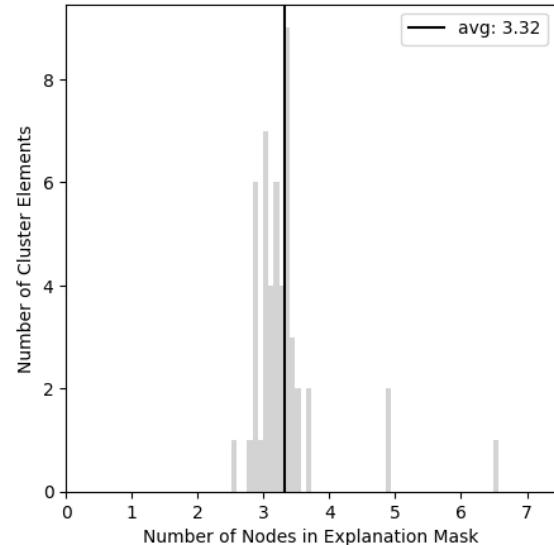
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

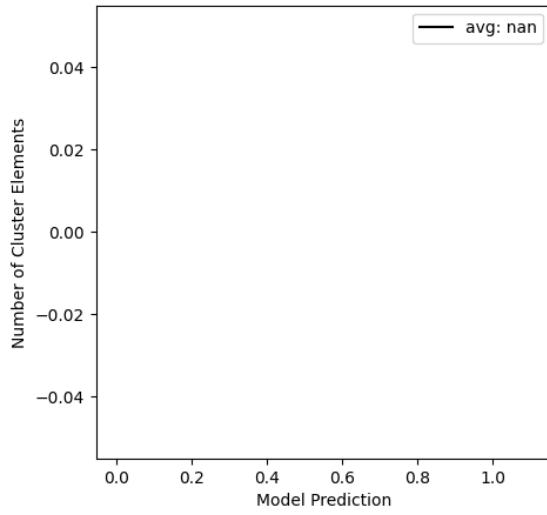
Prediction Impact Distribution



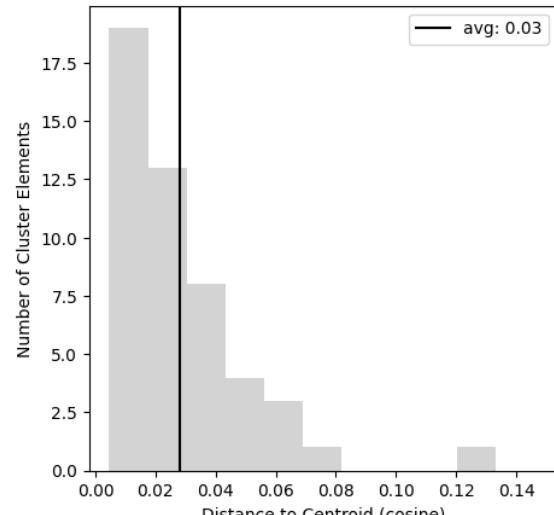
Mask Size Distribution



Prediction Output Distribution

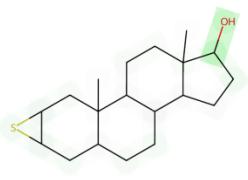
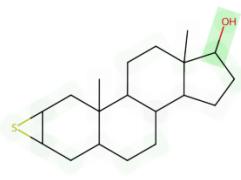
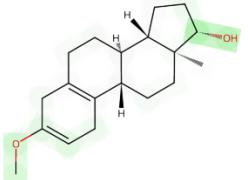
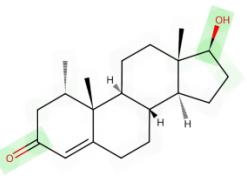
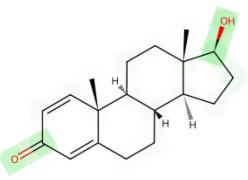
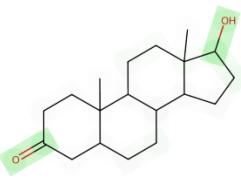
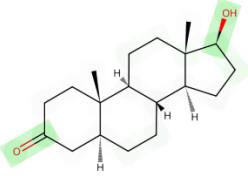
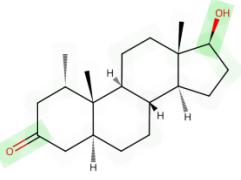
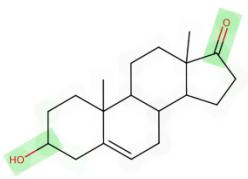
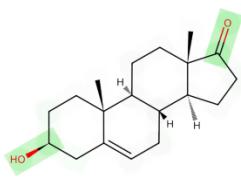
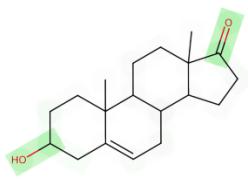
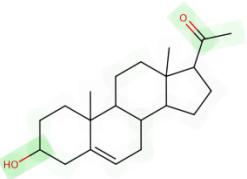
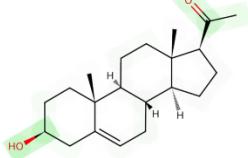
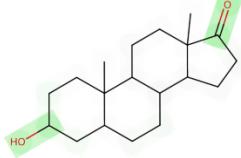
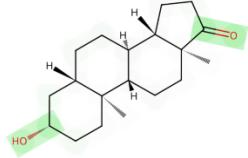
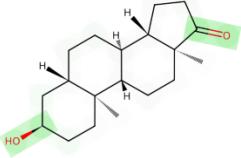


Distance to Centroid Distribution



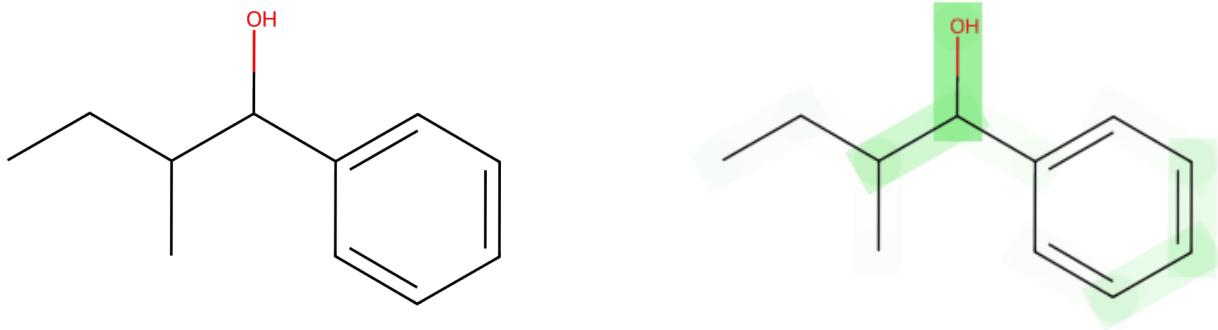
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES structure represents a hydrocarbon backbone with a phenyl group (benzene ring) and a hydroxyl group (-OH) as substituents. The hydroxyl group is known to increase water solubility due to its polarity and ability to form hydrogen bonds with water molecules. Meanwhile, the hydrocarbon part of the molecule is typically hydrophobic and decreases solubility in water. The benzene ring may also decrease solubility due to its nonpolar nature, although its effects could be mitigated by the substituents it carries. In this case, the presence of the hydroxyl group may be the key contributor to the observed influence on water solubility, as it introduces a polar functional group that can interact favorably with water.

**Hypothesis:** Molecules containing the substructure "C-C-C(-C)-C(-O)-c1:c:c:c:c:c:1" have increased water solubility compared to similar molecules lacking this substructure. The hydroxyl (-OH) group attached to the hydrocarbon chain provides a site for hydrogen bonding with water molecules, thus enhancing the molecule's water solubility. Conversely, the benzene ring and the hydrocarbon backbone may counteract this effect to some extent due to their nonpolar nature, but the overall influence of the hydroxyl group is significant.

# Cluster #63 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 63, from importance channel 1 (*positive*), represents a motif consisting of 2.9 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 1.3 ( $\pm 0.1$ ) on the prediction outcome.

## Properties

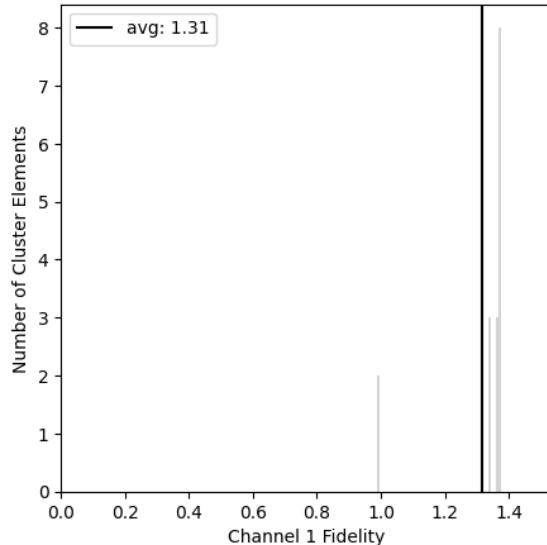
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	16
Channel Index	1.0 (0.0)

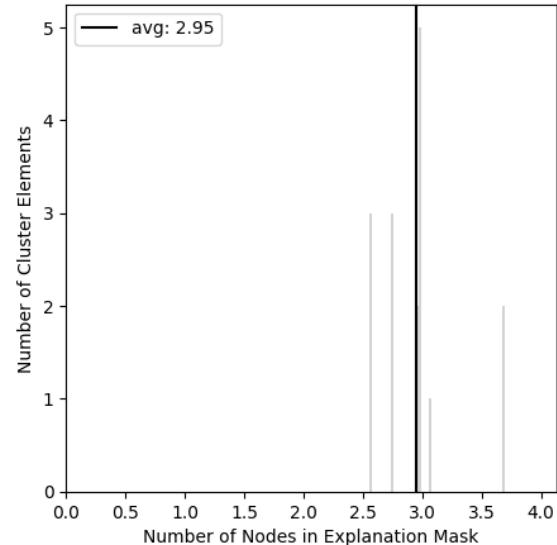
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

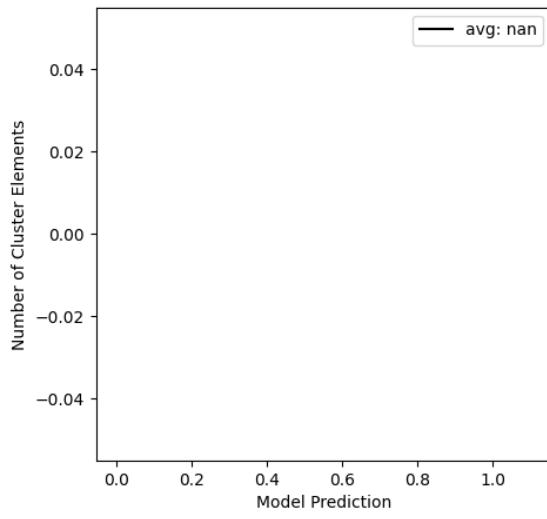
Prediction Impact Distribution



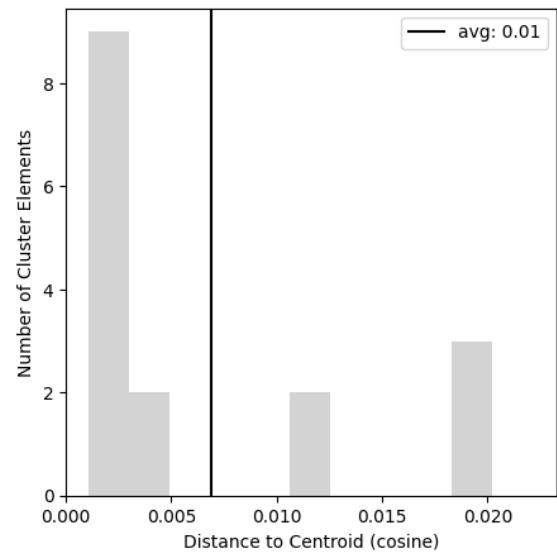
Mask Size Distribution



Prediction Output Distribution

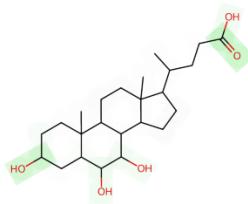
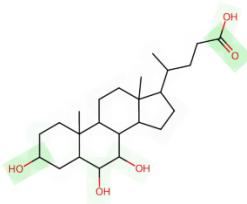
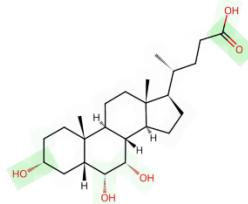
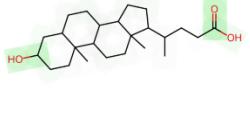
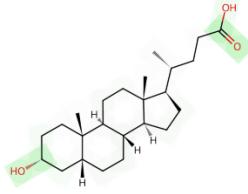
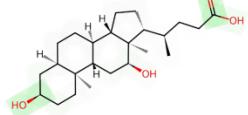
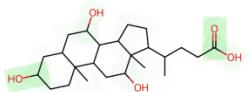
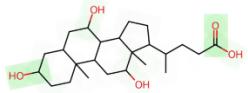
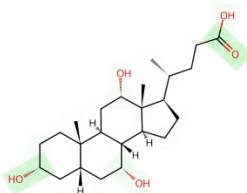
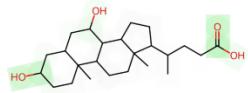
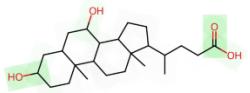
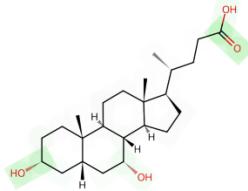
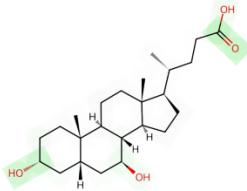
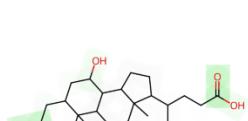
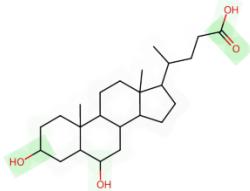


Distance to Centroid Distribution



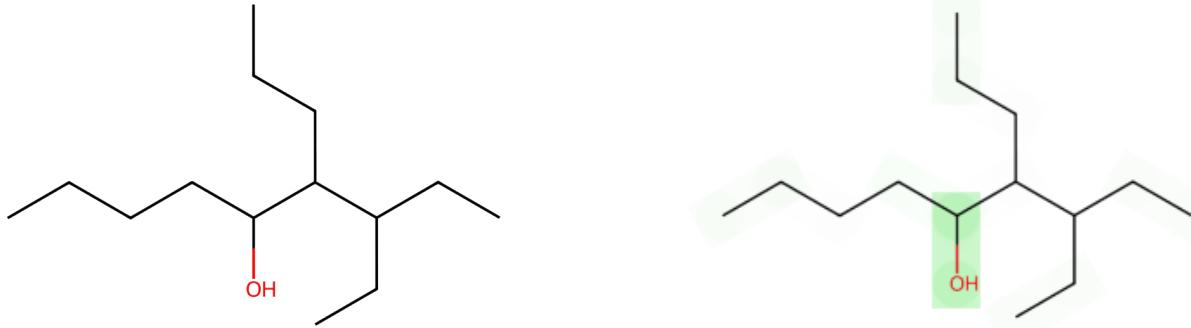
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES representation describes a hydrocarbon chain with a pendant alcohol group (-C(-O)-). The alcohol group is a hydrophilic moiety due to the presence of the polar O-H bond, which can form hydrogen bonds with water molecules. The long hydrocarbon chains, on the other hand, are hydrophobic and do not interact favorably with water. However, the presence of the alcohol group can increase the solubility of the molecule in water compared to a hydrocarbon without any functional groups, as it provides a site for interaction with water molecules.

**Hypothesis:** Molecules with the structure "C-C-C-C-C(-O)-C(-C-C-C)-C(-C-C)-C-C" exhibit increased water solubility compared to similar length hydrocarbons without functional groups. The -C(-O)- group facilitates interaction with water through hydrogen bonding, resulting in an increased tendency to dissolve. Despite the overall hydrophobic nature of the molecule, this functional group contributes enough polarity to enhance solubility.

# Cluster #64 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 64, from importance channel 1 (*positive*), represents a motif consisting of 3.2 ( $\pm 0.6$ ) nodes. The concept is generally associated with an impact of 1.0 ( $\pm 0.1$ ) on the prediction outcome.

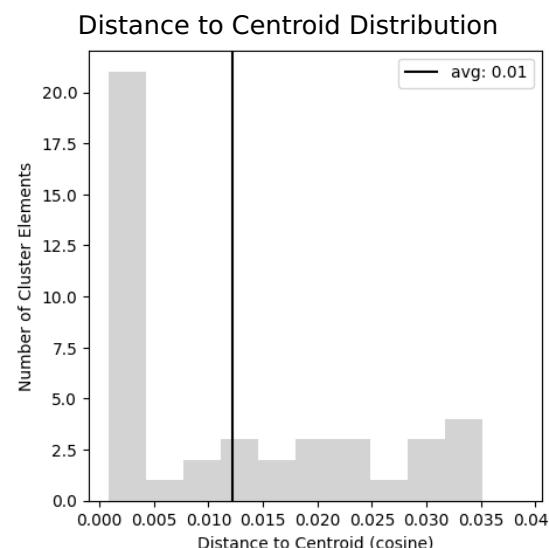
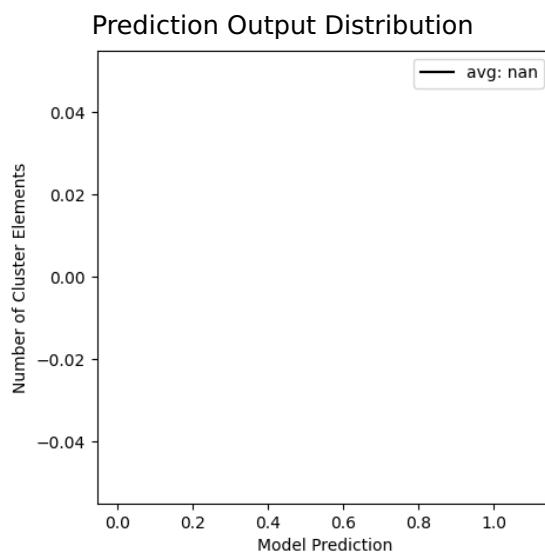
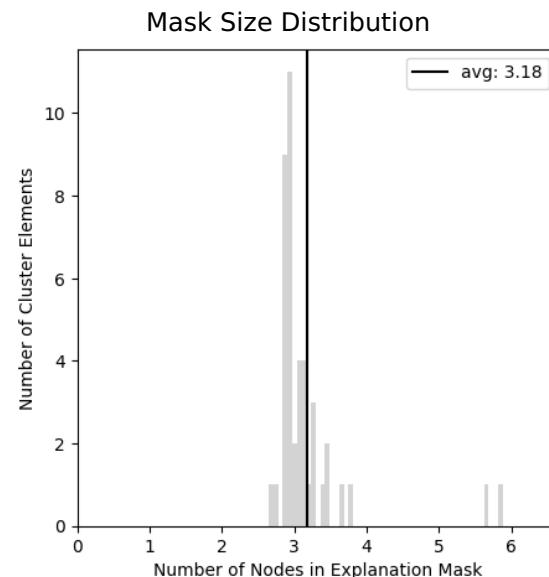
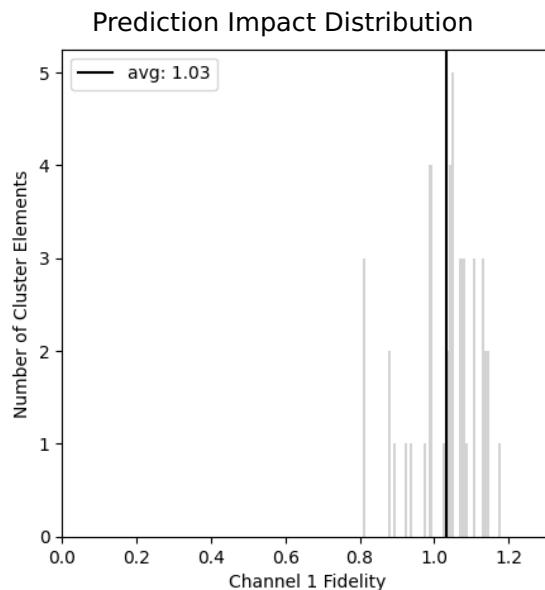
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	43
Channel Index	1.0 (0.0)

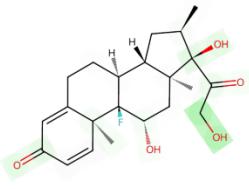
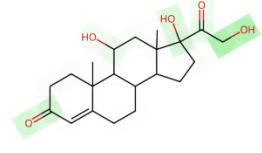
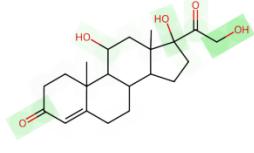
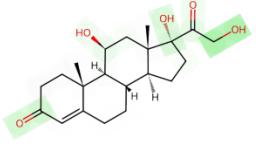
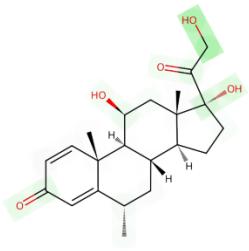
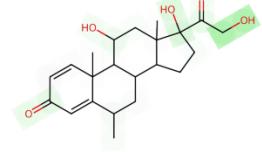
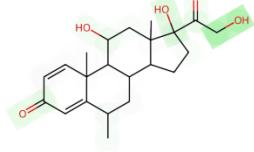
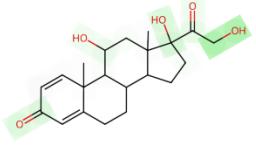
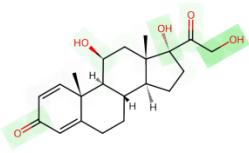
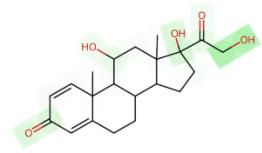
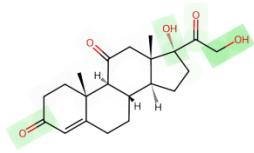
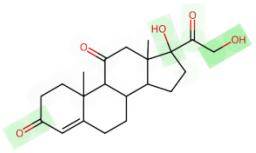
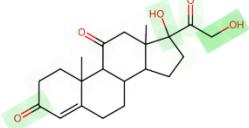
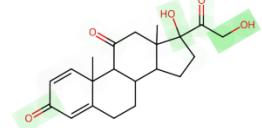
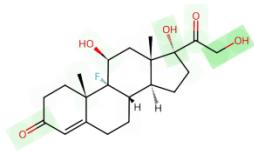
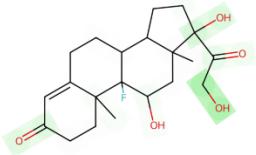
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



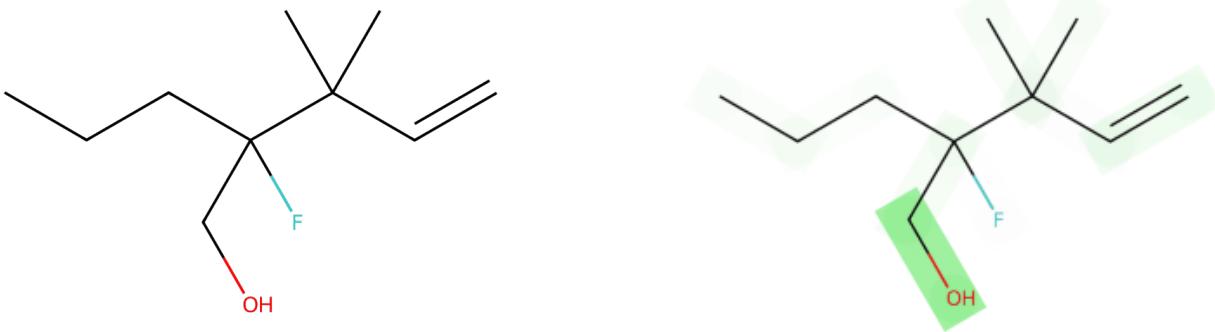
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The presence of a fluorine atom (denoted by "-F" in the SMILES string) within the structure likely increases the water solubility of the molecule due to fluorine's high electronegativity. The fluorine atom creates a dipole within the molecule, thus increasing its ability to form hydrogen bonds with water molecules. Additionally, the presence of a hydroxyl group (denoted by "-O") suggests the potential for hydrogen bonding with water, as this group can serve as both a hydrogen bond donor and acceptor. Conversely, the multiple carbon-carbon bonds and branched alkyl chains (indicated by the "C-C" and multiple "-C" in the SMILES string) typically have a hydrophobic character, thus reducing water solubility. However, the influence of the electronegative fluorine and the hydroxyl group seems to have a more dominant effect in this structure.

**Hypothesis:** The molecule with the given substructure has an increased water solubility due to the presence of a fluorine atom and a hydroxyl group. The electronegativity of fluorine creates polarity in the molecule aiding in its interaction with water, while the hydroxyl group can form hydrogen bonds, both of which enhance solubility despite the presence of hydrophobic alkyl chains.

# Cluster #65 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 65, from importance channel 1 (*positive*), represents a motif consisting of 3.2 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 1.0 ( $\pm 0.2$ ) on the prediction outcome.

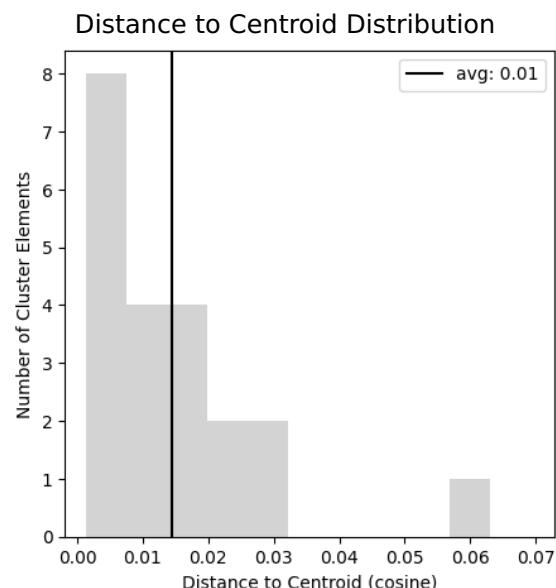
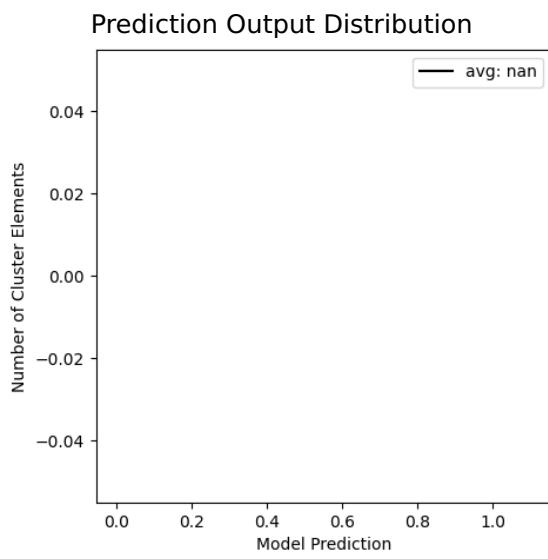
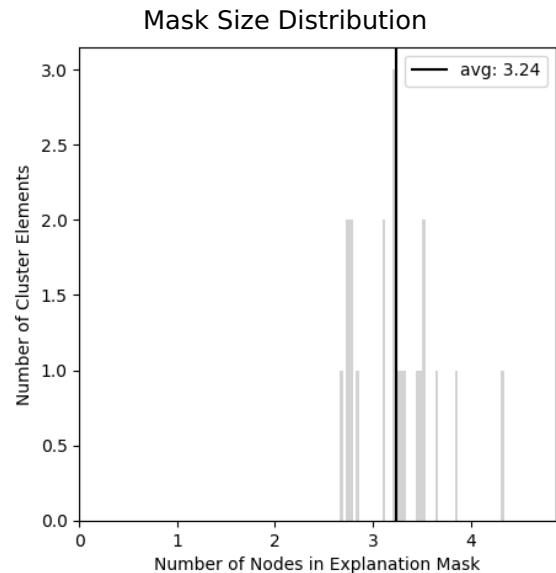
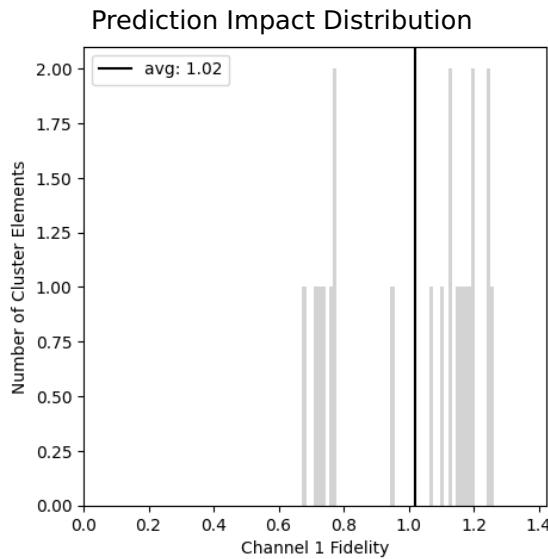
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	21
Channel Index	1.0 (0.0)

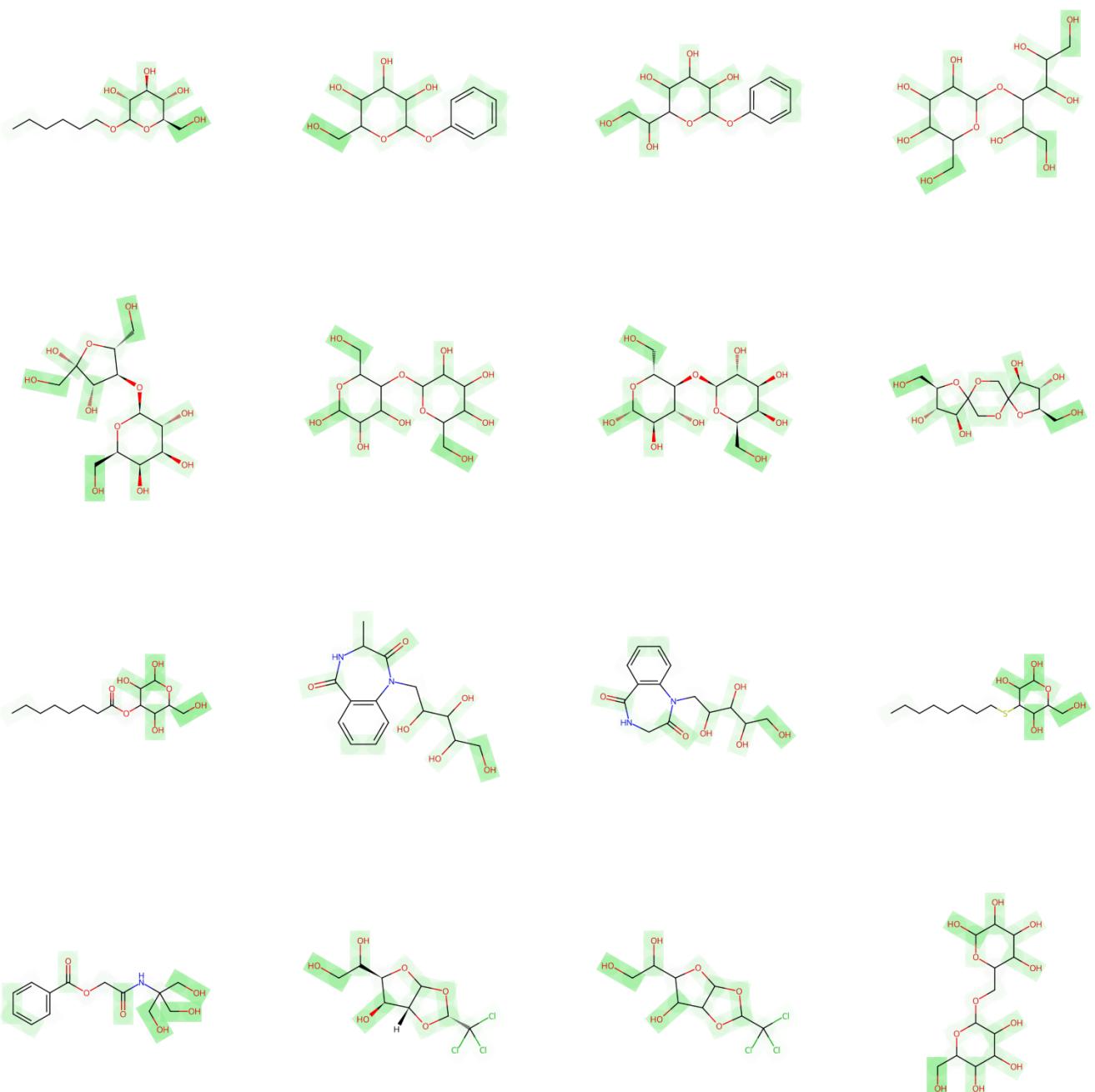
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



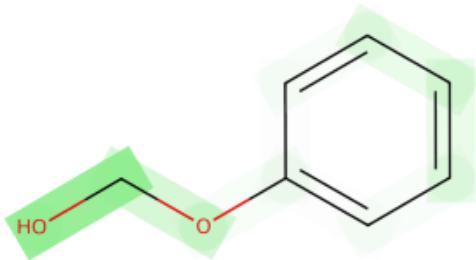
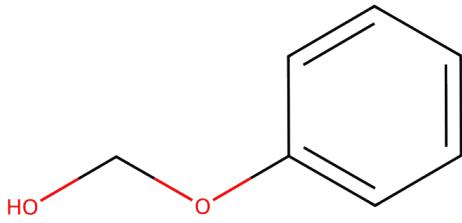
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given structure consists of an anisole derivative, a benzene ring with a methoxy group (-OCH<sub>3</sub>) attached. The oxygen atoms in the methoxy group and in the ether linkage (-O-) can engage in hydrogen bonding with water molecules, enhancing the solubility of the molecule in water. The aromatic ring, while typically hydrophobic, is countered by the polar methoxy group, which increases the molecule's overall polarity and, consequently, its water solubility.

**Hypothesis:** The presence of an ether linkage and a methoxy group attached to an aromatic ring increases water solubility. These groups introduce polarity and potential sites for hydrogen bonding, which helps overcome the hydrophobic nature of the benzene ring, leading to a higher tendency of the molecule to dissolve in water.

# Cluster #66 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 66, from importance channel 1 (*positive*), represents a motif consisting of 3.6 ( $\pm 0.6$ ) nodes. The concept is generally associated with an impact of 1.1 ( $\pm 0.2$ ) on the prediction outcome.

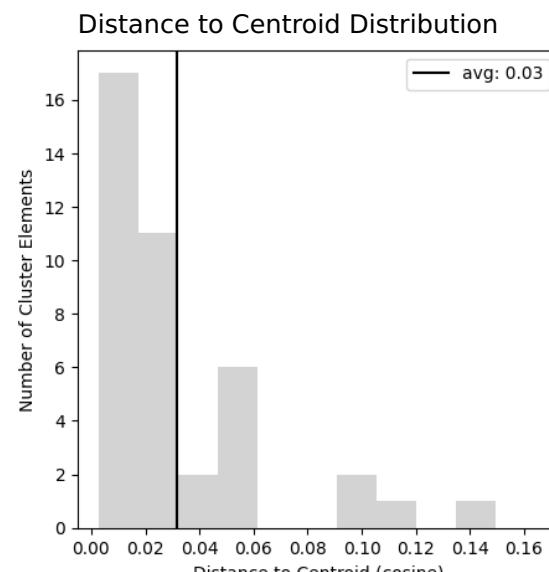
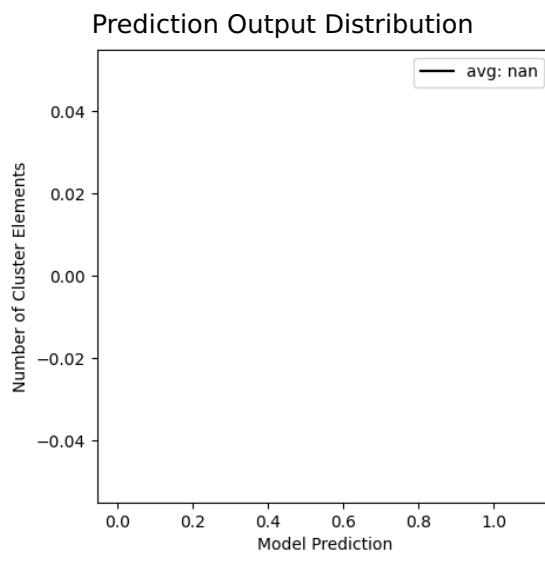
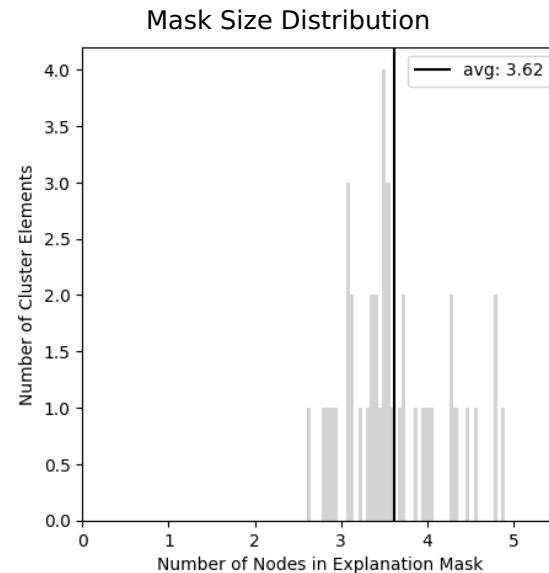
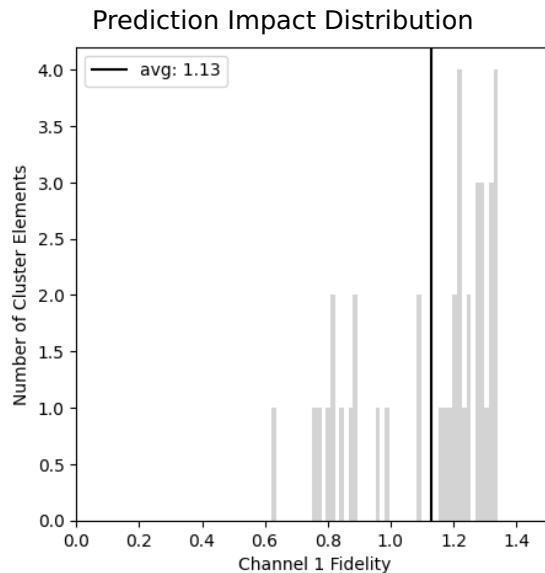
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	40
Channel Index	1.0 (0.0)

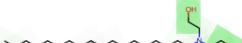
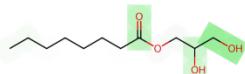
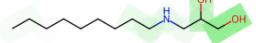
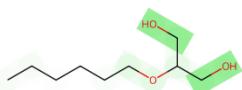
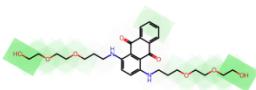
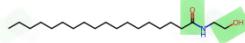
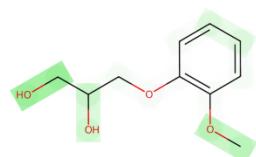
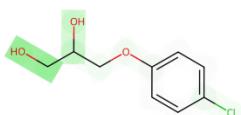
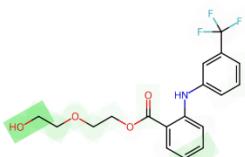
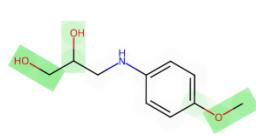
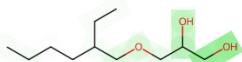
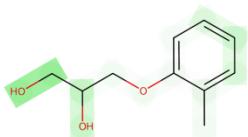
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



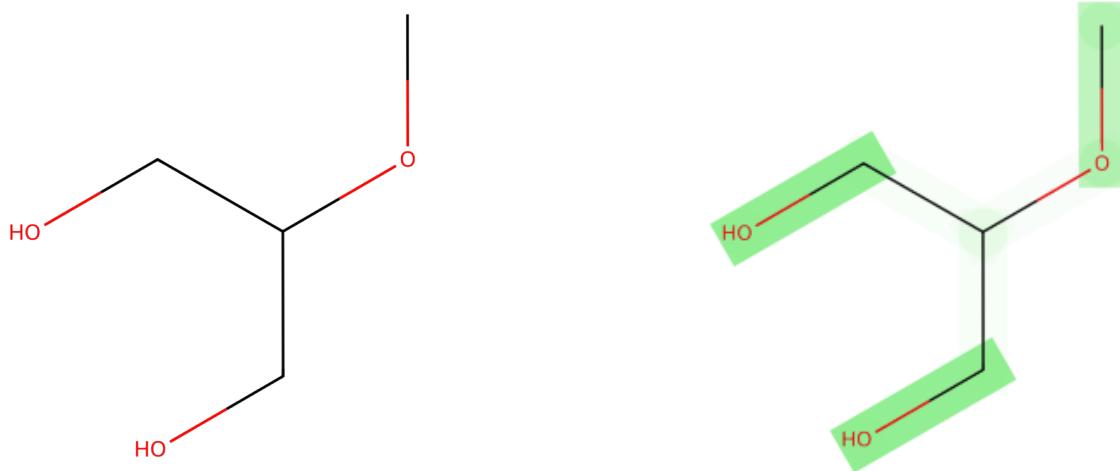
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure "C-O-C(-C-O)-C-O" suggests the presence of multiple hydroxyl groups (-OH) attached to a carbon chain. Hydroxyl groups are polar in nature due to the difference in electronegativity between oxygen and hydrogen atoms. This polarity allows hydroxyl groups to form hydrogen bonds with water molecules, which is a key interaction that promotes solubility in water. Additionally, the structure implies that the molecule is branched, providing more sites for water interaction. The branching and presence of hydroxyl groups combined are likely to enhance the solubility of the molecule since it can interact more effectively with the polar water molecules.

**Hypothesis:** Molecules containing the substructure "C-O-C(-C-O)-C-O" are more soluble in water than those without it. The presence of multiple hydroxyl groups increases the molecule's ability to form hydrogen bonds with water, which enhances solubility. Furthermore, the branched structure of the molecule provides additional sites for interaction with water, facilitating even greater solubility.

# Cluster #67 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 67, from importance channel 1 (*positive*), represents a motif consisting of 3.5 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 0.9 ( $\pm 0.2$ ) on the prediction outcome.

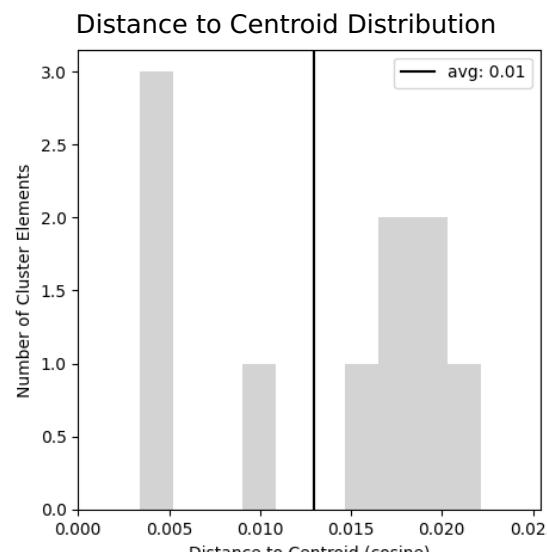
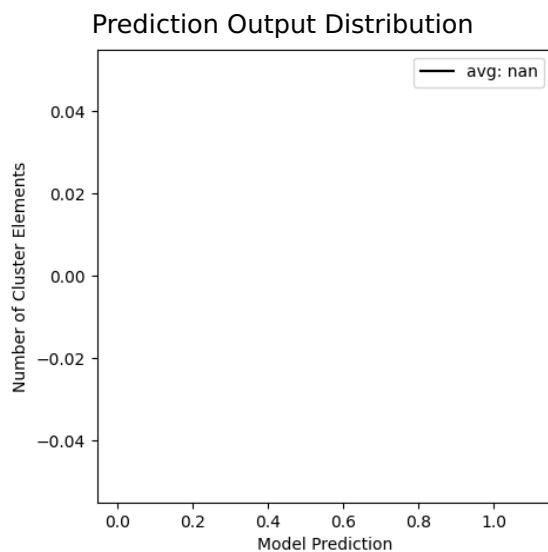
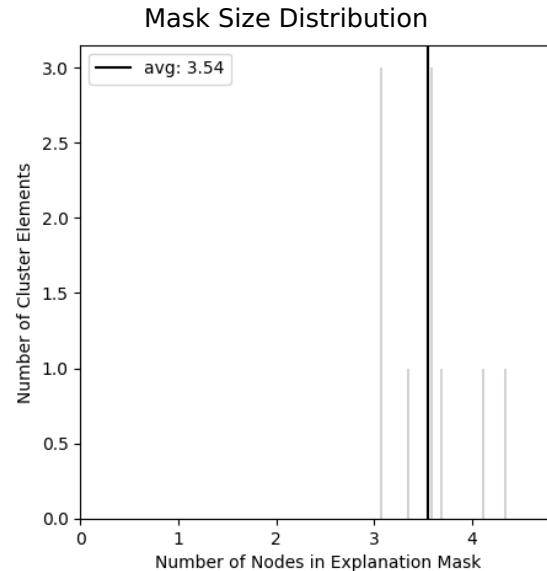
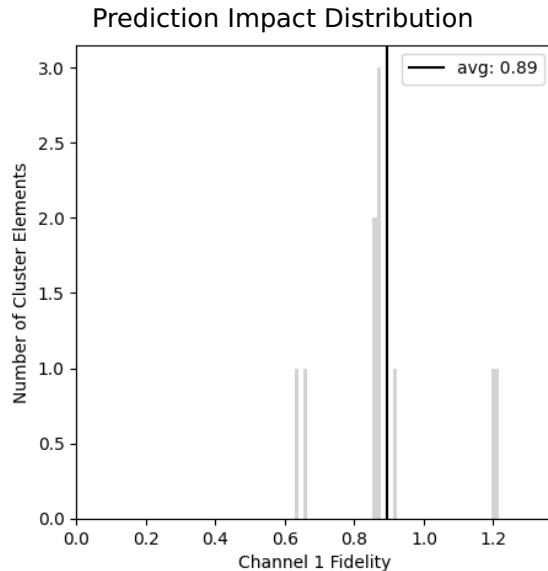
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	10
Channel Index	1.0 (0.0)

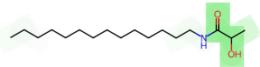
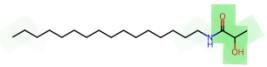
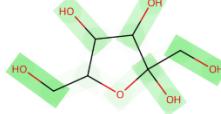
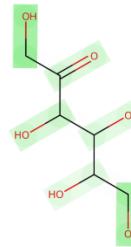
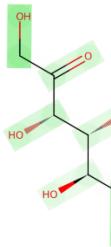
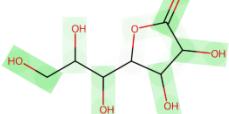
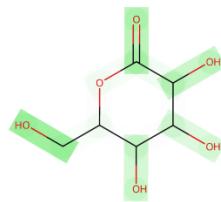
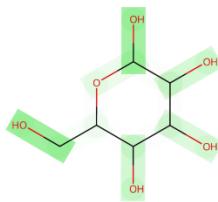
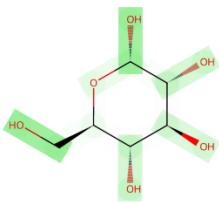
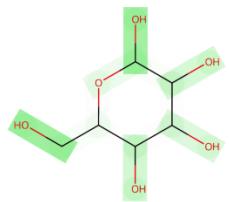
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



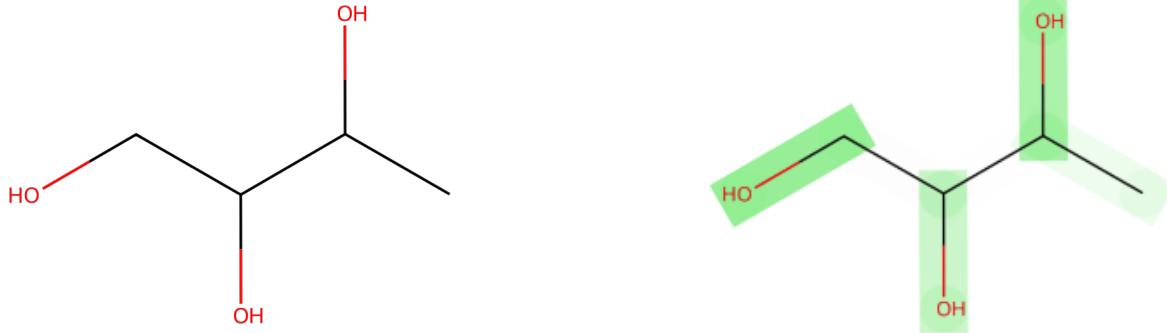
## Example Elements

① This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The "-O" groups in the given structure are hydroxyl groups, which are highly polar due to the electronegativity difference between oxygen and hydrogen. This polarity allows for strong hydrogen bonding with water molecules, which is a key factor in determining water solubility. Moreover, the fact that there are multiple hydroxyl groups likely increases the solubility because more sites are available for hydrogen bonding with water. The carbon chain ("C-C") acts as a hydrophobic backbone, but its effect on water solubility is less than the effect of the hydroxyl groups.

**Hypothesis:** Molecules featuring the substructure "C-C(-O)-C(-O)-C-O" have a high tendency to dissolve in water due to the presence of multiple hydroxyl groups. These groups enhance water solubility through their capacity for hydrogen bond formation with water molecules, despite the presence of a hydrophobic carbon backbone.

# Cluster #68 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 68, from importance channel 1 (*positive*), represents a motif consisting of 3.5 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 1.4 ( $\pm 0.5$ ) on the prediction outcome.

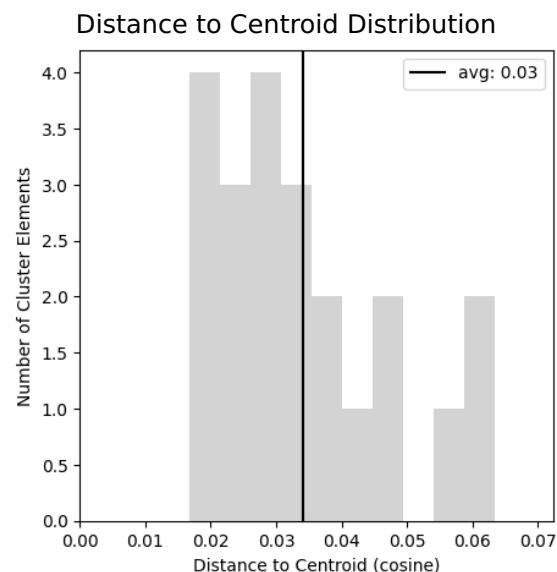
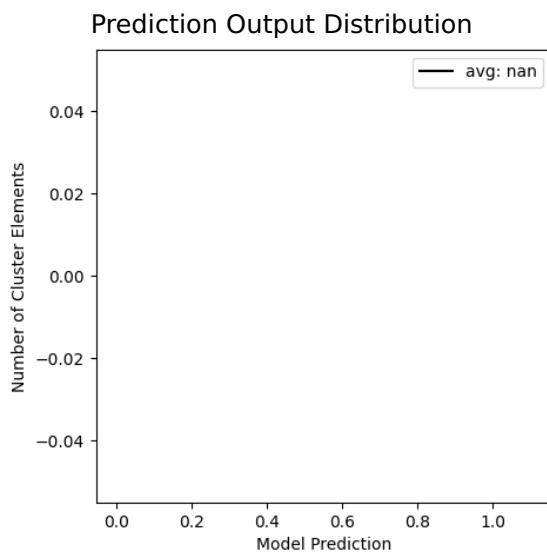
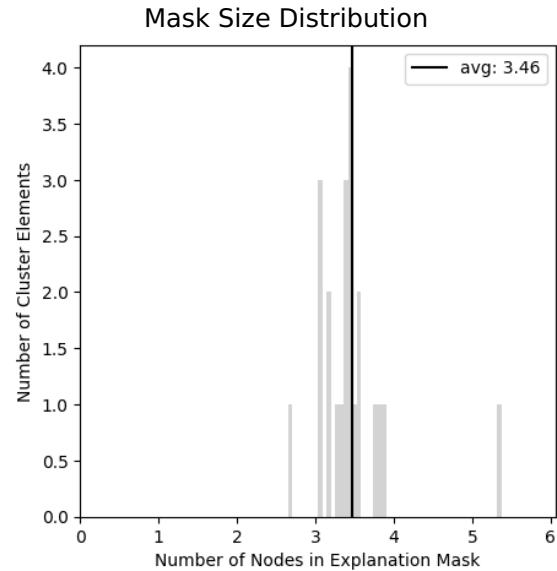
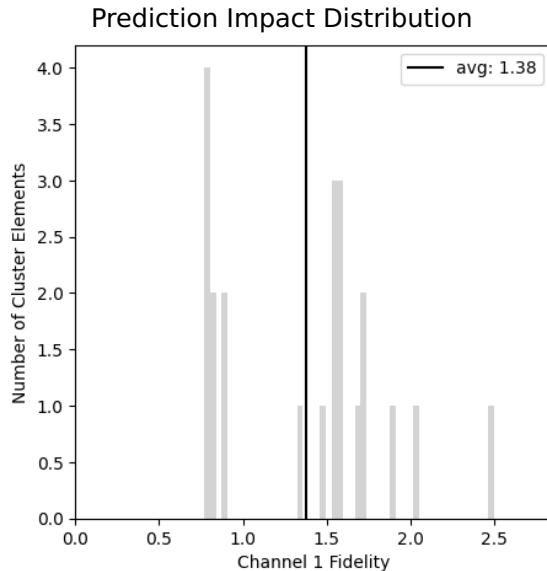
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	22
Channel Index	1.0 (0.0)

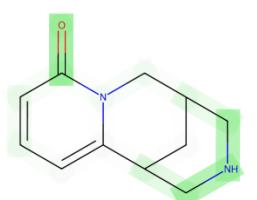
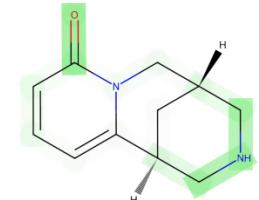
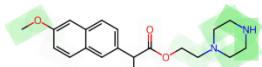
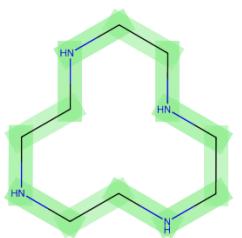
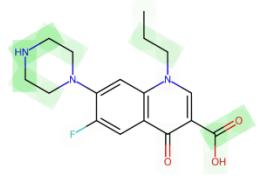
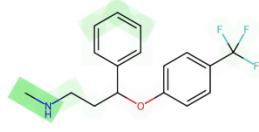
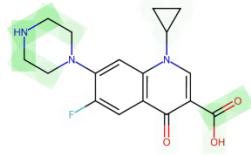
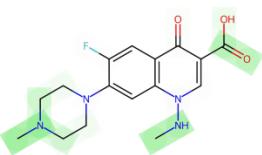
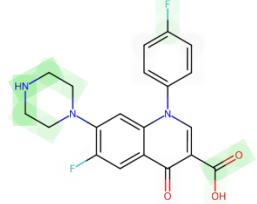
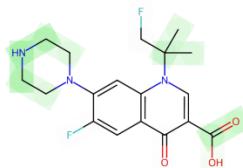
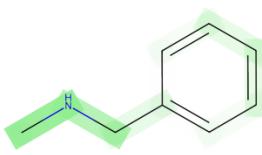
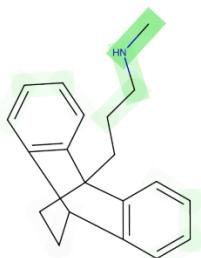
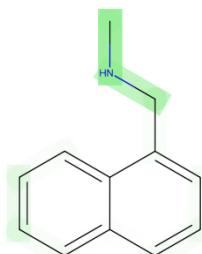
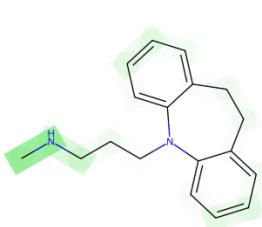
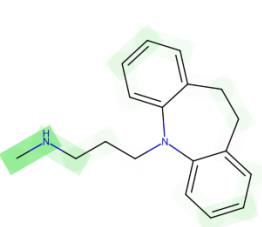
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The presence of the nitrogen atom in the "-C-N-" substructure within the C-C-N-C backbone suggests an increased polarity due to the electronegativity of nitrogen, which can lead to hydrogen bonding interactions with water. These interactions are critical for solubility in water because they enable the molecule to better integrate into the aqueous environment by replacing water-water hydrogen bonds with solute-water hydrogen bonds.

**Hypothesis:** Molecules containing the "-C-N-" substructure are likely to have enhanced water solubility. The electronegative nitrogen atom can engage in hydrogen bonding with water, leading to improved integration of the molecule into the aqueous environment. This interaction is particularly significant when the nitrogen is part of a small alkyl chain, as in the "C-C-N-C" fragment, where it can influence the solubility of the molecule by increasing its overall polarity.

# Cluster #69 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 69, from importance channel 1 (*positive*), represents a motif consisting of 3.4 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 1.2 ( $\pm 0.1$ ) on the prediction outcome.

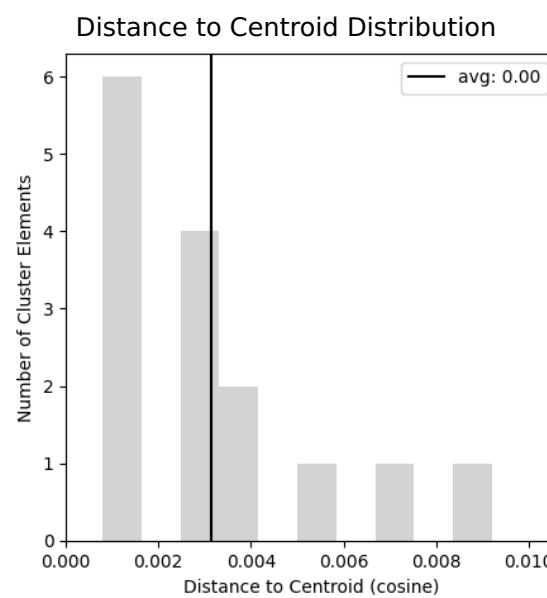
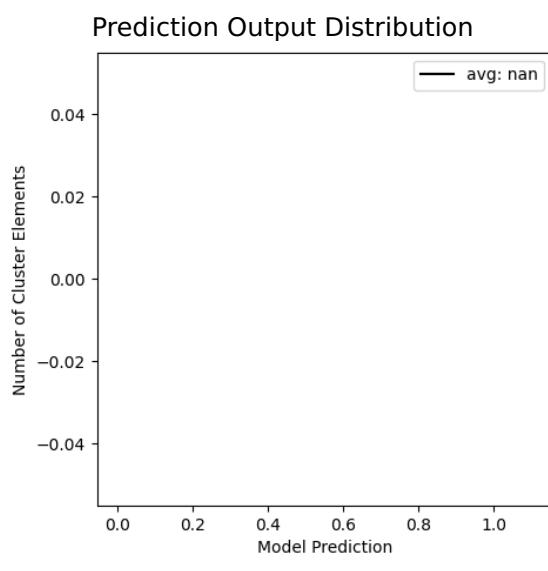
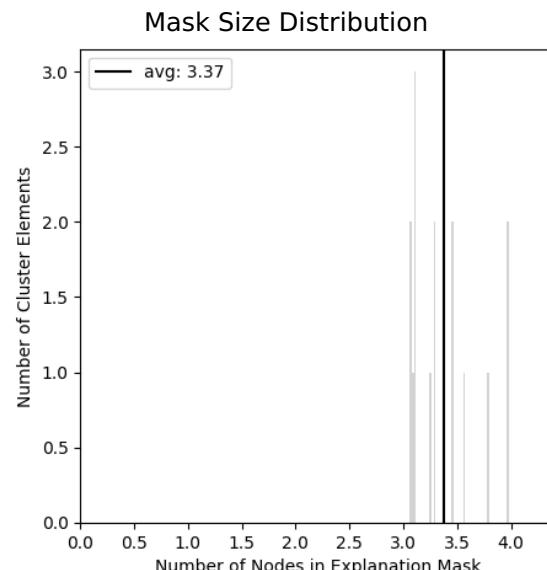
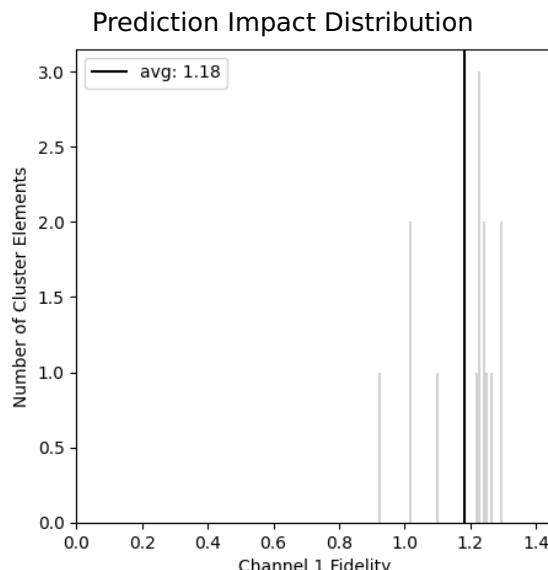
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	15
Channel Index	1.0 (0.0)

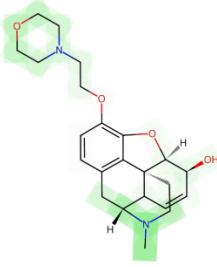
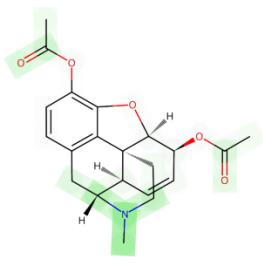
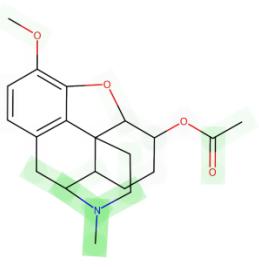
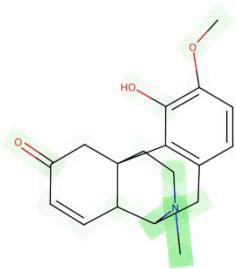
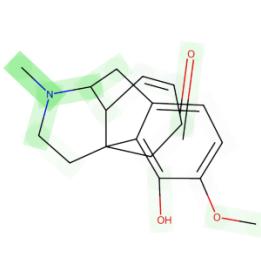
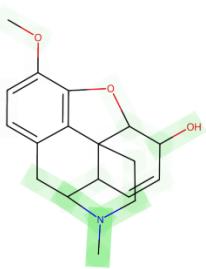
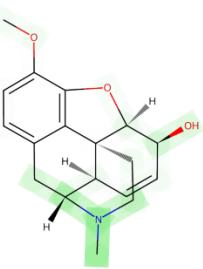
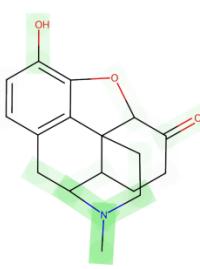
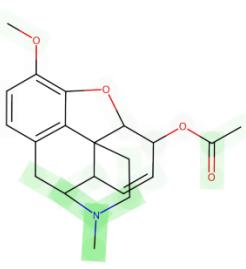
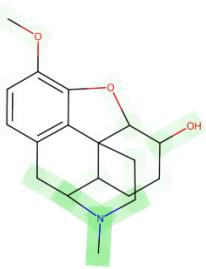
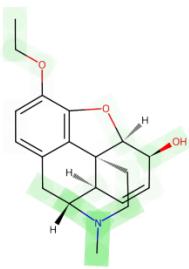
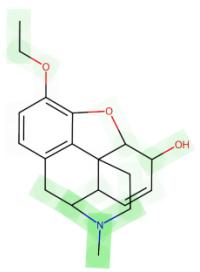
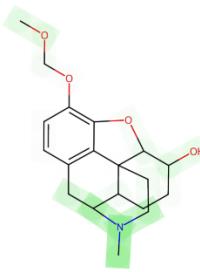
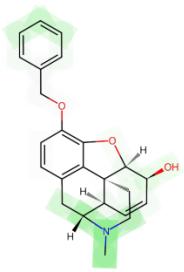
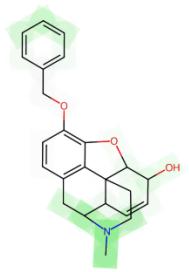
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



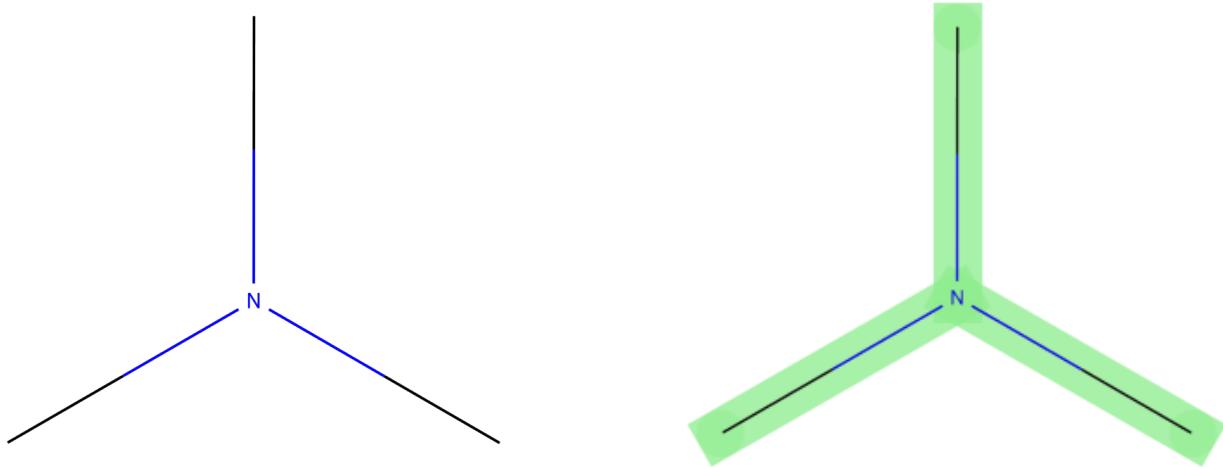
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The substructure "C-N(-C)-C" represents a tertiary amine where a nitrogen atom is bonded to three carbon atoms. Tertiary amines are known to be capable of forming hydrogen bonds with water due to the lone pair of electrons on the nitrogen. However, the alkyl groups attached to the nitrogen can contribute to hydrophobic interactions, which may lessen water solubility. The balance between these opposing effects—the ability to hydrogen bond and the hydrophobic character of the alkyl groups—determines the overall solubility in water.

**Hypothesis:** Molecules containing the substructure "C-N(-C)-C" have a moderate tendency to be water-soluble. This is likely because the hydrogen-bonding capability of the tertiary amine nitrogen increases solubility, while the alkyl groups can reduce it by promoting hydrophobic interactions.

# Cluster #70 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 70, from importance channel 1 (*positive*), represents a motif consisting of 3.6 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 0.8 ( $\pm 0.1$ ) on the prediction outcome.

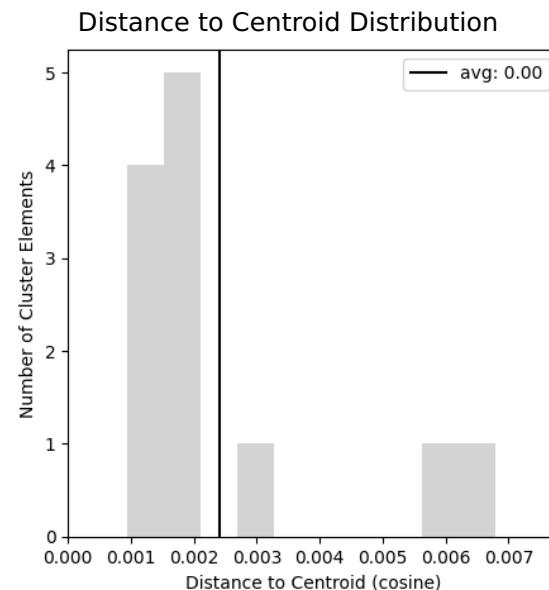
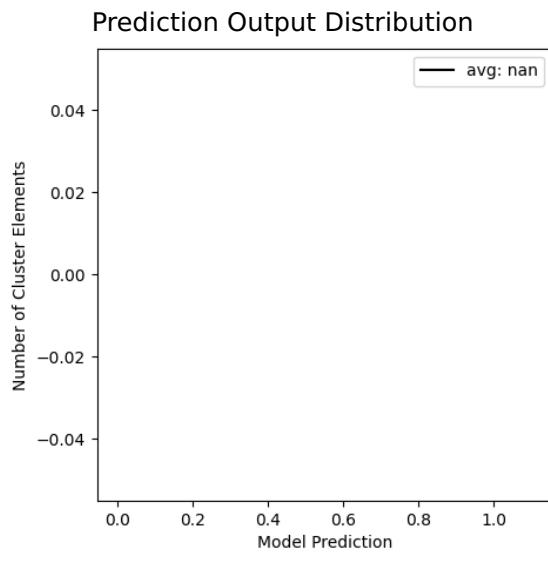
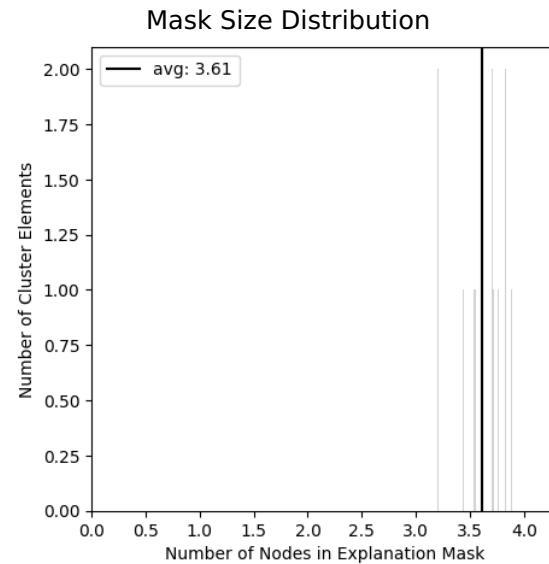
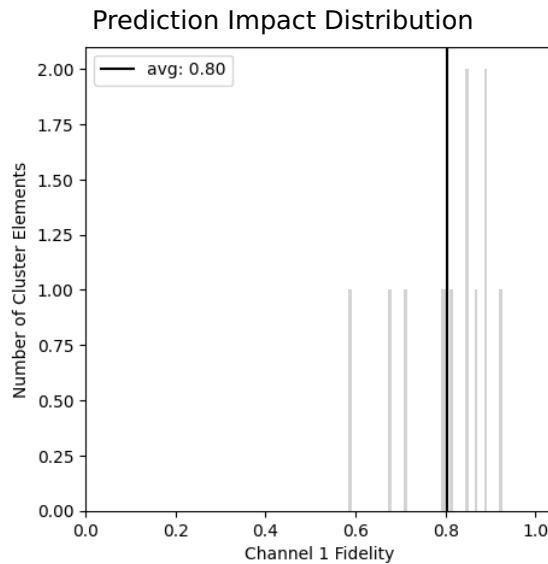
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	12
Channel Index	1.0 (0.0)

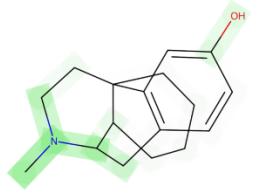
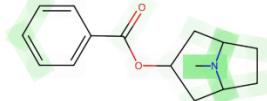
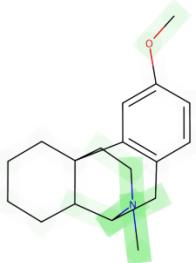
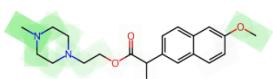
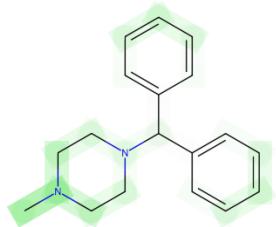
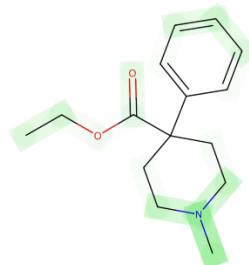
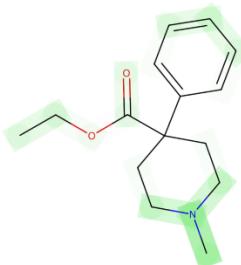
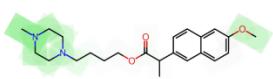
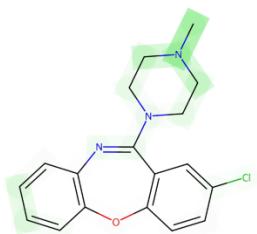
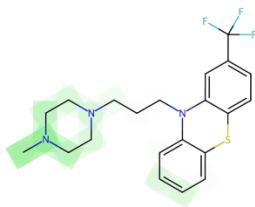
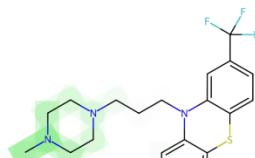
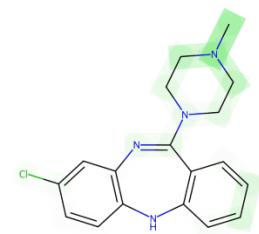
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The linkage of carbon atoms (C-C) that leads to a nitrogen atom (N) followed by another carbon atom (C) in the molecular structure suggests the presence of an amine group (C-N-C). Amines are known for their polar nature due to the electronegativity difference between nitrogen and carbon, which can form hydrogen bonds with water molecules. This polar characteristic and hydrogen-bonding ability greatly enhance the likelihood of the molecule to interact with and dissolve in water.

**Hypothesis:** A molecular substructure containing a C-C-N-C sequence promotes water solubility. The amine group within this sequence has a polar character that allows for favorable interactions with water molecules, while its ability to form hydrogen bonds further enhances water solubility.

# Cluster #71 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 71, from importance channel 1 (*positive*), represents a motif consisting of 3.4 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 0.7 ( $\pm 0.2$ ) on the prediction outcome.

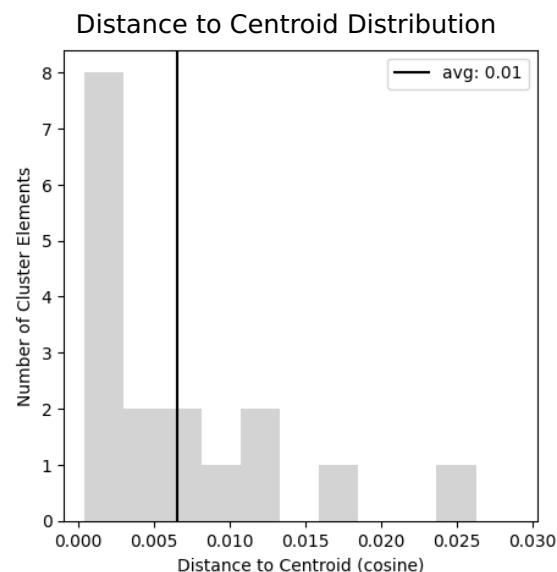
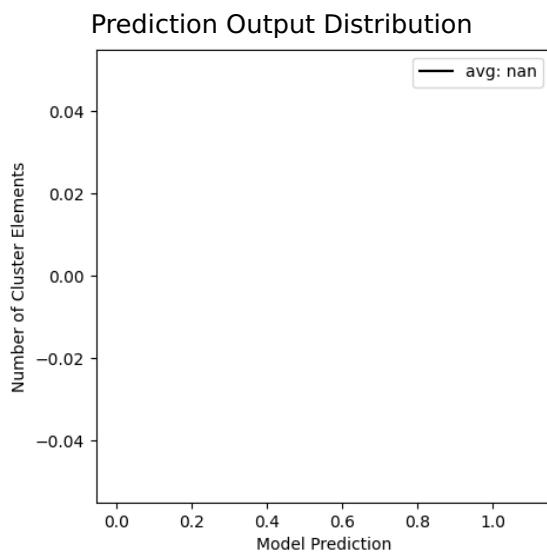
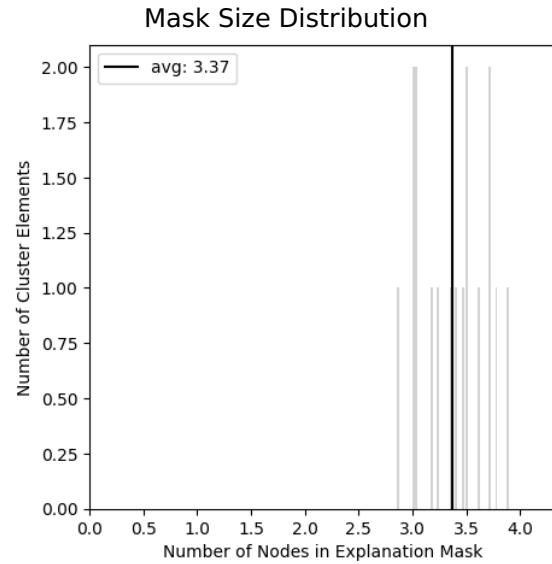
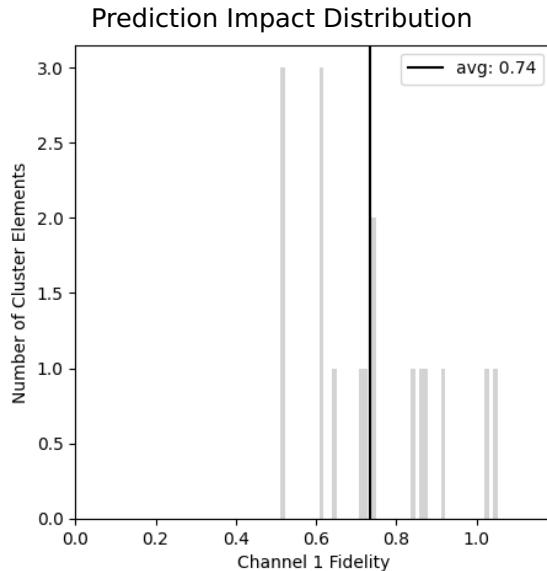
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	17
Channel Index	1.0 (0.0)

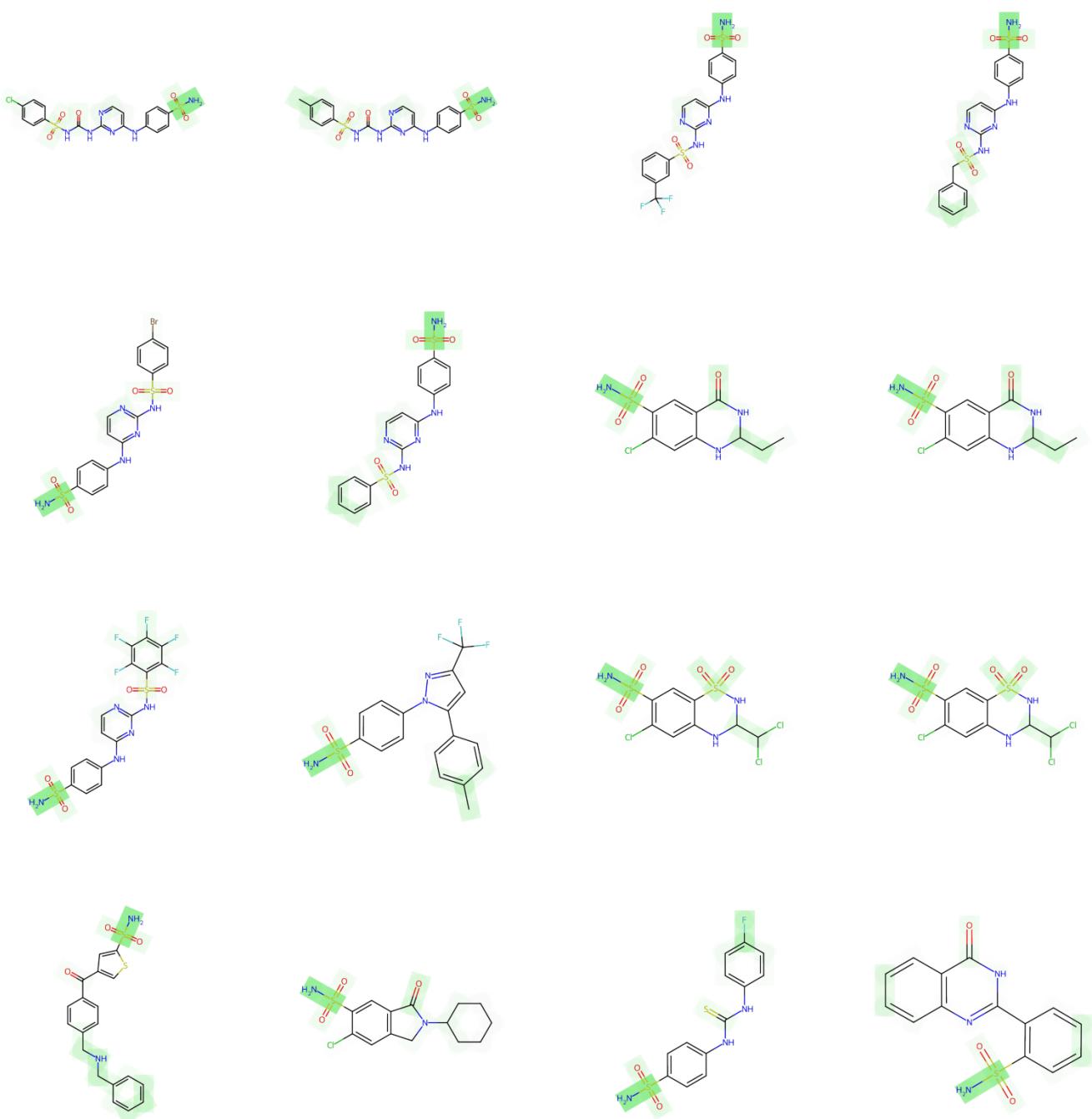
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



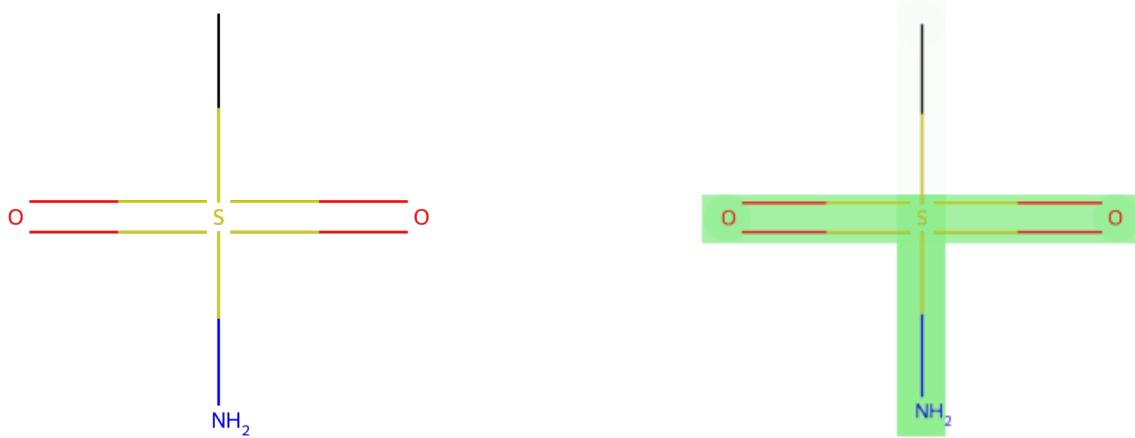
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The presence of a sulfonamide group in a molecule, as depicted by the SMILES "C-S(-N)(=O)=O", increases the water solubility due to its ability to form strong hydrogen bonds with water molecules. The sulfonamide functional group contains a sulfur atom double bonded to two oxygen atoms and single bonded to a nitrogen atom, this arrangement allows the oxygen and nitrogen atoms to act as hydrogen bond acceptors. Water molecules, being polar and having a bent shape, can easily align themselves around these sites through the hydrogen bonding framework, effectively solvating the molecule and increasing its solubility in water.

**Hypothesis:** Molecules containing the sulfonamide substructure are likely to exhibit enhanced water solubility. The polar nature of the sulfonamide group and its ability to engage in hydrogen bonding interactions are the primary factors contributing to this property. This hypothesis is supported by the reported 0.74 influence that the sulfonamide functional group has on water solubility, signifying a moderate to strong effect on this property.

# Cluster #72 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 72, from importance channel 1 (*positive*), represents a motif consisting of 3.2 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 0.7 ( $\pm 0.2$ ) on the prediction outcome.

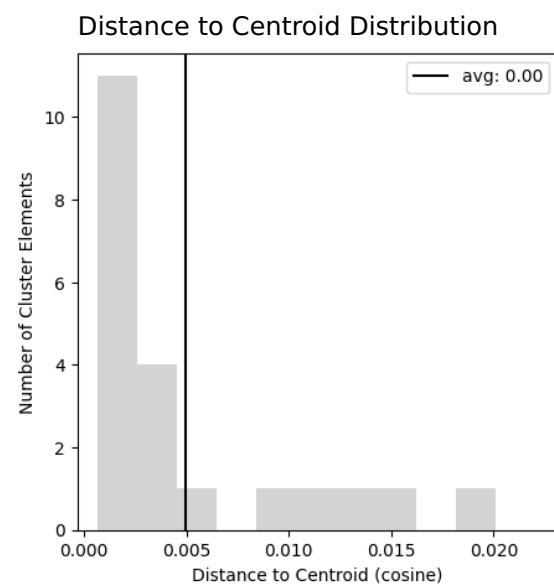
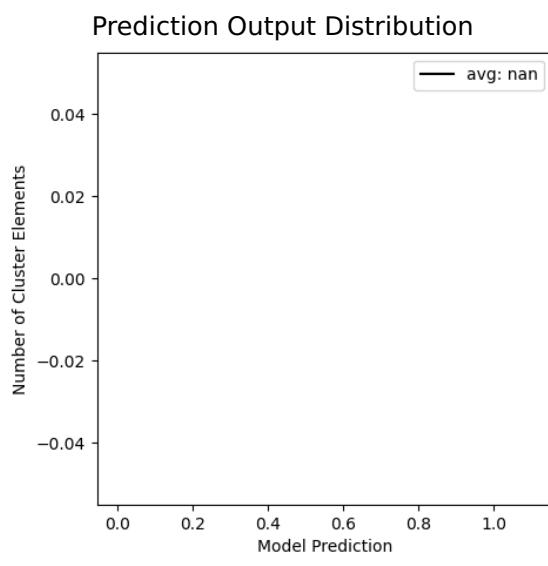
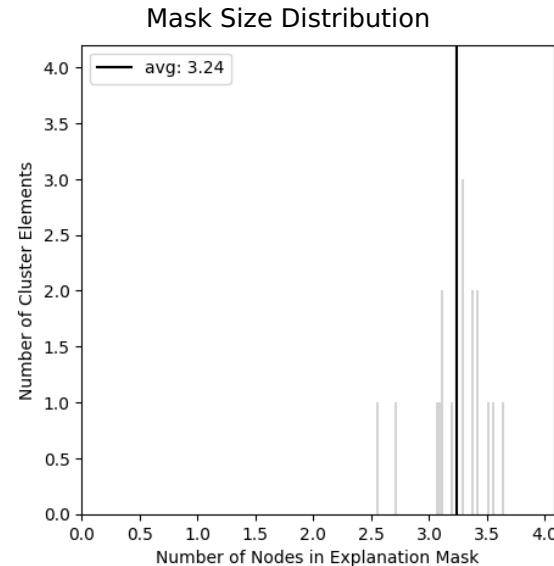
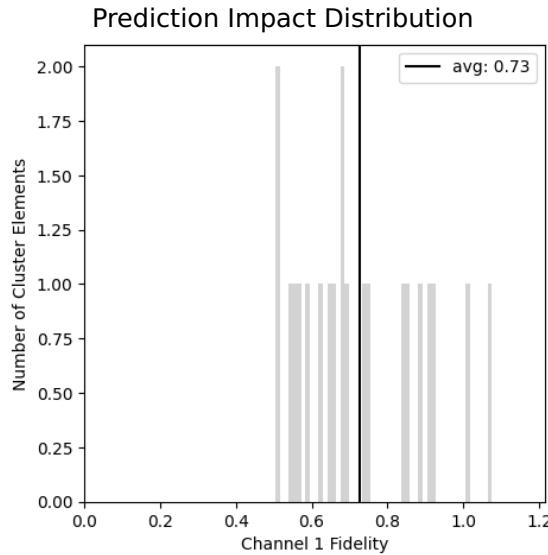
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	21
Channel Index	1.0 (0.0)

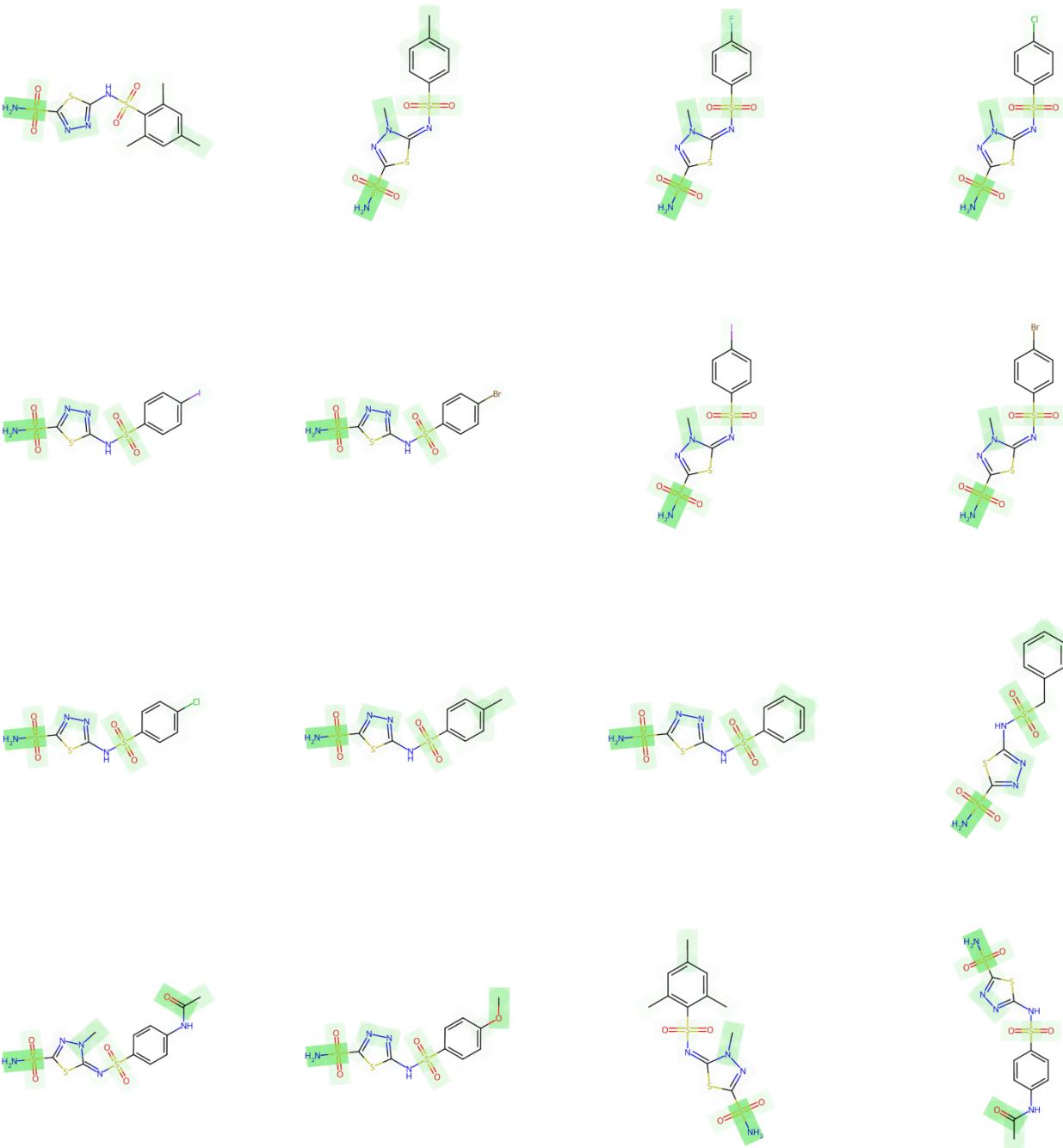
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The substructure "C=C-S(=O)-N-C-N" contains several functional groups that can interact with water molecules, potentially enhancing solubility. The S(=O) moiety suggests the presence of a sulfoxide group, which is polar and can engage in hydrogen bonding with water, increasing solubility. Additionally, the presence of two nitrogen atoms indicates potential sites for hydrogen bonding or ionic interactions, if they are part of amine or amide groups. The double bond, "C=C", does not directly promote water solubility but its position between polar or hydrogen-bonding groups might not significantly hinder the overall water solubility due to the compound's potential to create a strong interaction with water molecules.

**Hypothesis:** The substructure "C=C-S(=O)-N-C-N" is likely to increase water solubility. The sulfoxide group and nitrogen-containing groups are capable of strong interactions with water through hydrogen bonding or potential ionic interactions, while the non-polar C=C double bond's position is mitigated by the surrounding polar entities.

# Cluster #73 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 73, from importance channel 1 (*positive*), represents a motif consisting of 3.3 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 1.0 ( $\pm 0.1$ ) on the prediction outcome.

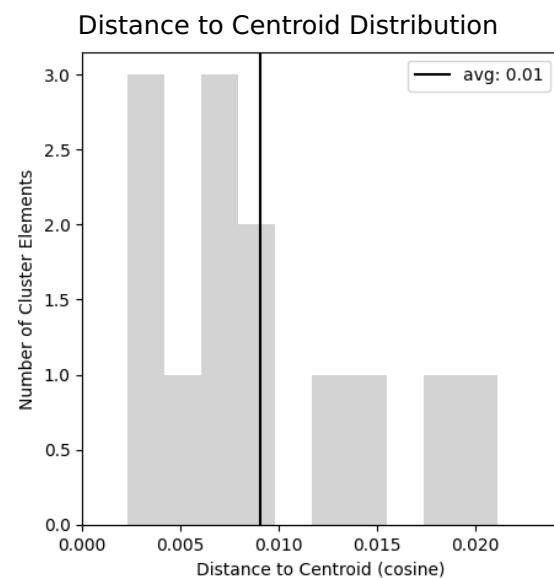
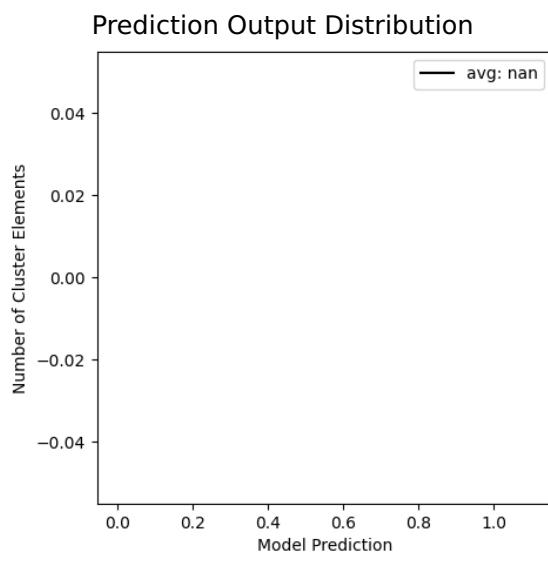
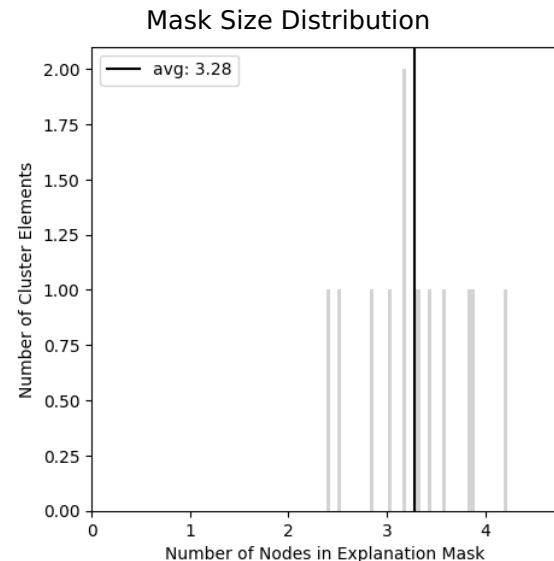
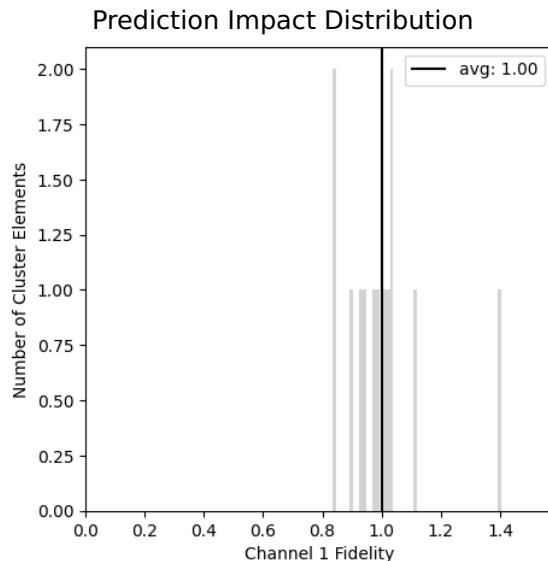
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	13
Channel Index	1.0 (0.0)

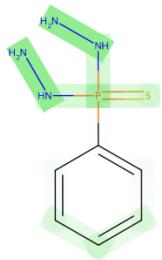
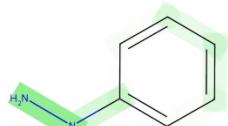
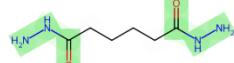
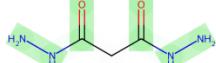
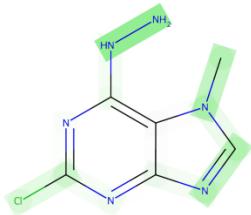
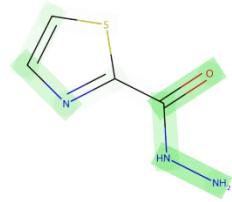
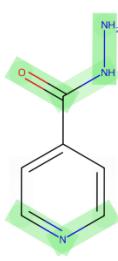
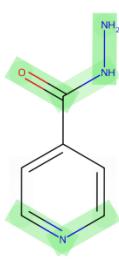
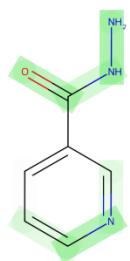
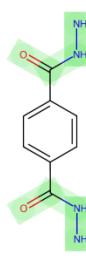
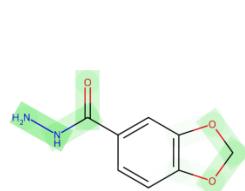
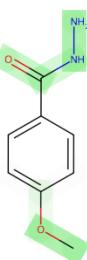
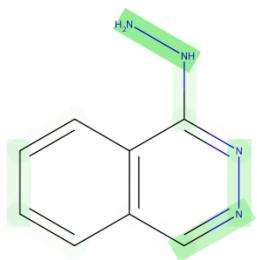
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The substructure "C-N-N", indicative of a carbon atom bonded to two nitrogen atoms in a chain, implies the presence of an amine (-NH-) group connected to a nitrile (-CN) or an azo (-N=N-) group. Amines are known to be polar and can easily form hydrogen bonds due to the presence of a lone pair of electrons on the nitrogen atom. This would increase the affinity of the molecule for water molecules, contributing to its solubility. The nitrile group, due to its triple bond, is also polar and can act as a hydrogen bond acceptor, further enhancing solubility in water.

**Hypothesis:** Molecules with the "C-N-N" structure are likely to have increased water solubility. This is due to the polar nature of the amine and nitrile groups that enable hydrogen bonding with water, as well as the possibility of the azo linkage contributing to polarity and solubility.

# Cluster #74 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 74, from importance channel 1 (*positive*), represents a motif consisting of 3.1 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 0.7 ( $\pm 0.1$ ) on the prediction outcome.

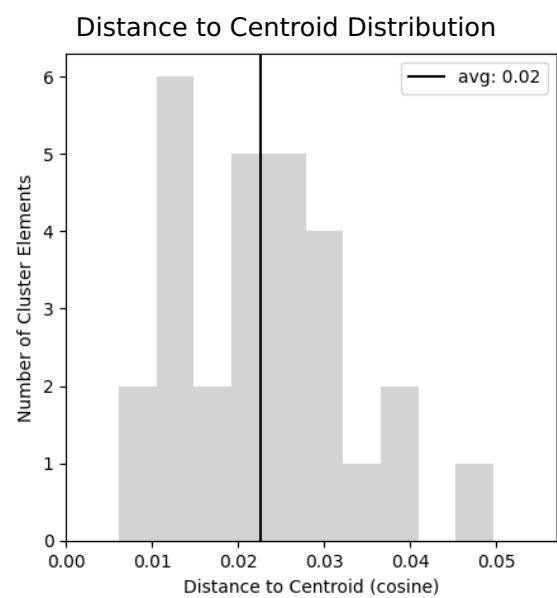
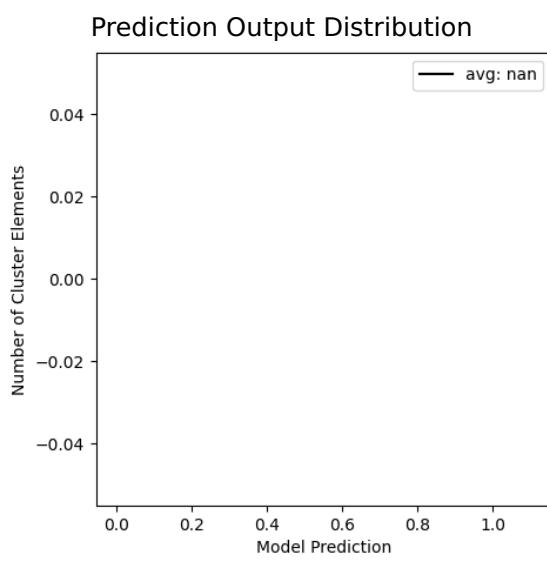
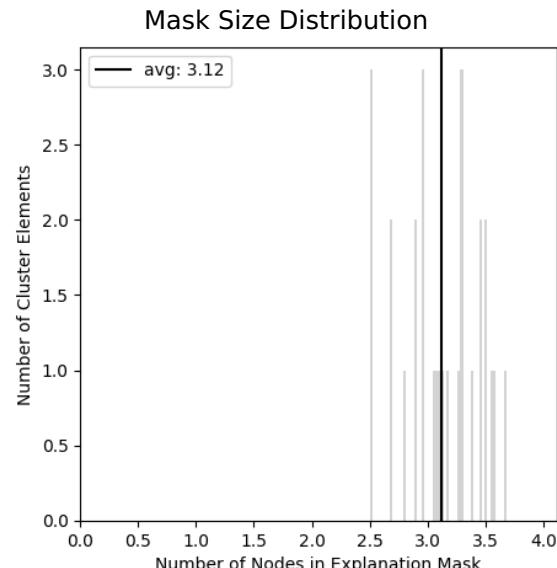
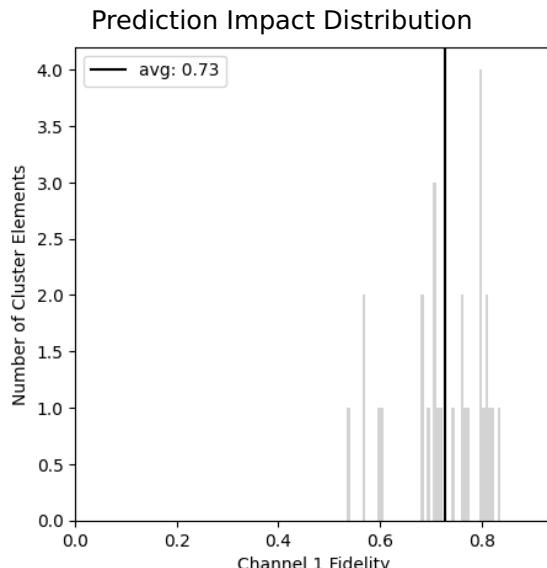
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	28
Channel Index	1.0 (0.0)

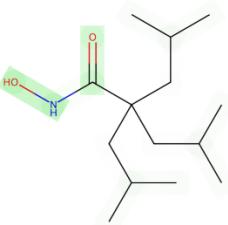
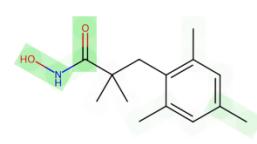
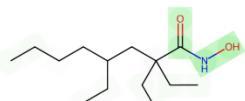
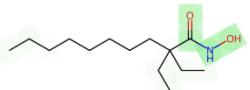
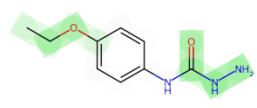
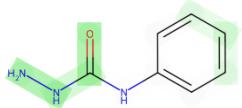
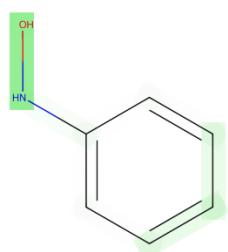
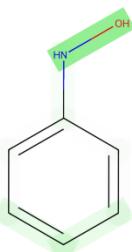
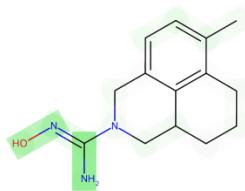
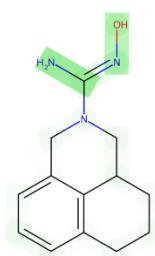
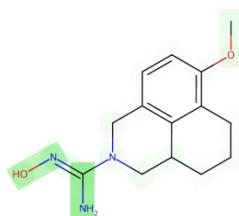
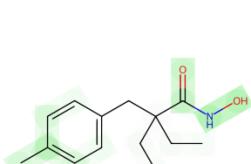
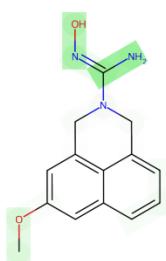
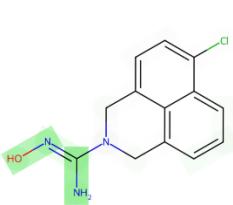
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



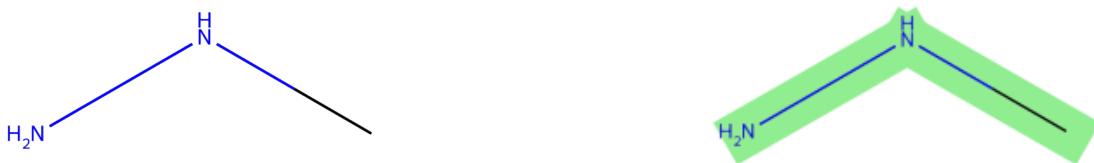
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The presence of the functional group "-N-N-", known as an azo group, within the molecular structure indicated by "C-N-N" suggests enhanced solubility in water. This is due to the presence of nitrogen atoms which can act as hydrogen bond acceptors. Hydrogen bonding is a key determinant of water solubility; molecules that can form hydrogen bonds with water are generally more water-soluble. Azo groups, being composed of two nitrogen atoms, provide sites for such interactions, thus increasing the molecule's affinity for water. Moreover, the carbon ("C-") attached to the azo group may also play a role, especially if it is part of a hydrophilic (water-attracting) functional group.

**Hypothesis:** Molecules containing the "C-N-N" substructure are likely to be water-soluble due to their ability to interact with water molecules through hydrogen bonding. The nitrogen atoms in the azo group act as sites of hydrogen bond acceptance which enhances water solubility, and the attached carbon could further influence this property depending on its own substituents or functional groups.

# Cluster #75 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 75, from importance channel 1 (*positive*), represents a motif consisting of 3.2 ( $\pm 0.4$ ) nodes. The concept is generally associated with an impact of 1.3 ( $\pm 0.3$ ) on the prediction outcome.

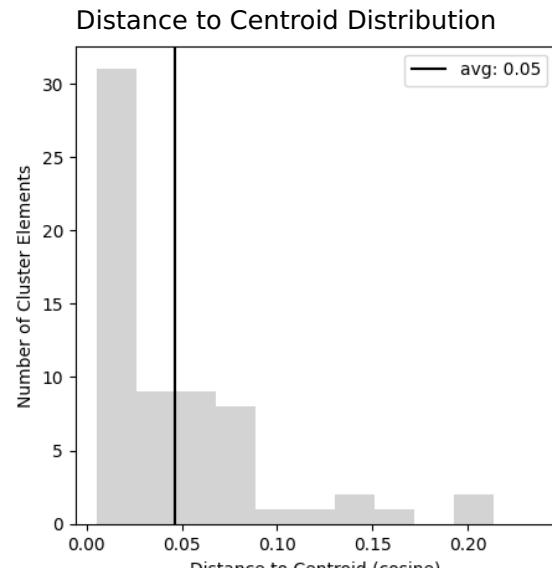
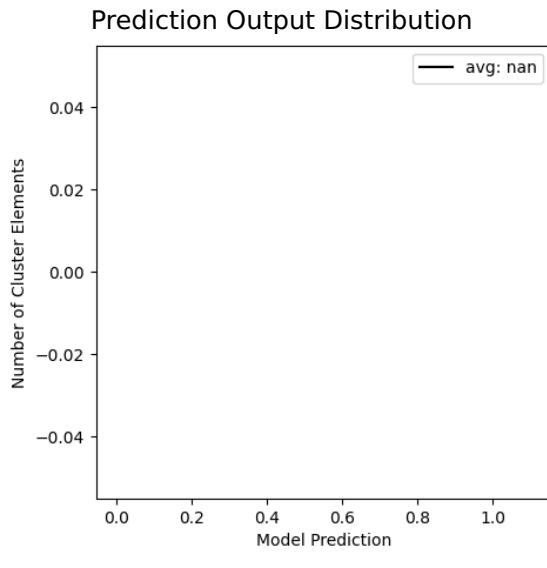
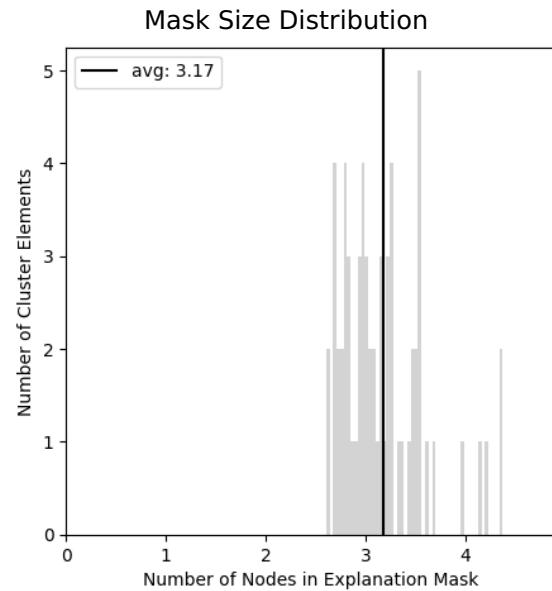
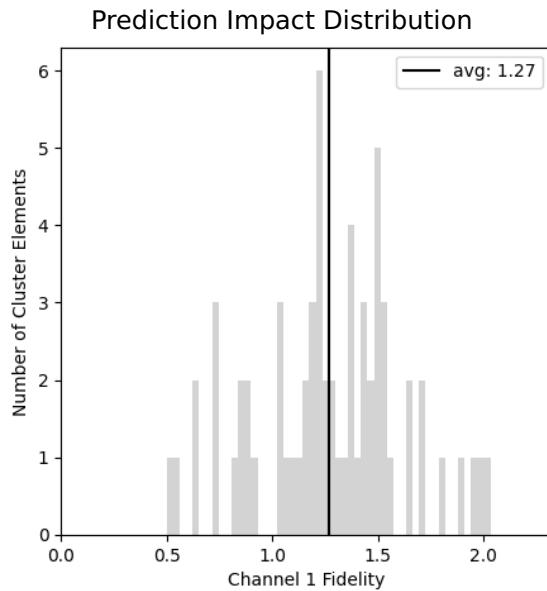
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	64
Channel Index	1.0 (0.0)

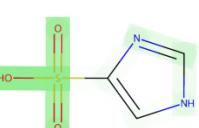
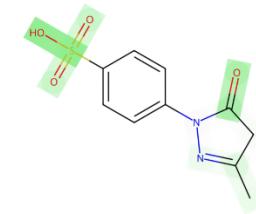
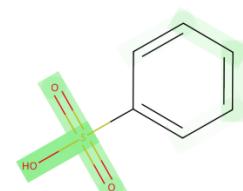
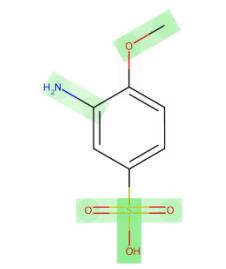
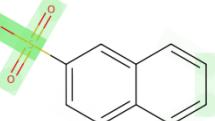
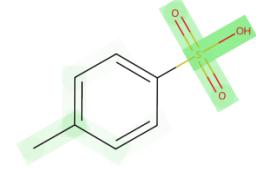
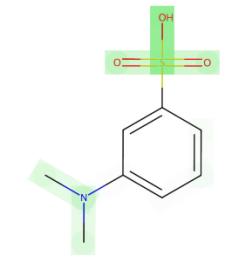
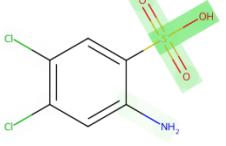
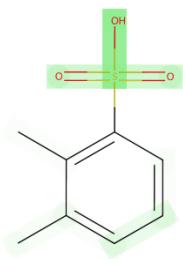
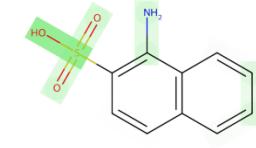
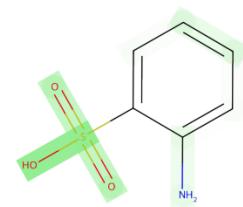
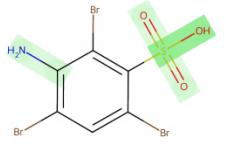
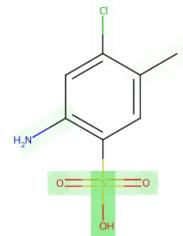
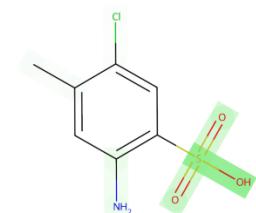
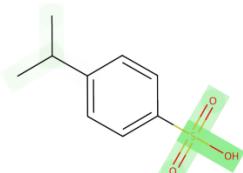
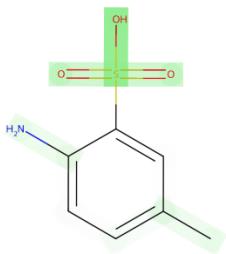
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure corresponds to a sulfonate group, which is a strong polar, ionic group that can form hydrogen bonds with water molecules. The presence of an  $\text{S}(\text{=O})_2\text{-OH}$  moiety suggests high hydrophilicity due to the strongly electronegative oxygen atoms that can engage in hydrogen bonding. The sulfonate group, being an excellent leaving group, can also enhance the solubility by undergoing reactions that yield more soluble products.

**Hypothesis:** Substructures incorporating a sulfonate group are likely to increase water solubility. The polar nature of the sulfonate group and its ability to form hydrogen bonds with water contribute to this tendency. Additionally, the potential for facile reactions in aqueous environments that lead to soluble products also helps explain the observed influence on water solubility.

# Cluster #76 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 76, from importance channel 1 (*positive*), represents a motif consisting of 2.6 ( $\pm 0.1$ ) nodes. The concept is generally associated with an impact of 0.8 ( $\pm 0.1$ ) on the prediction outcome.

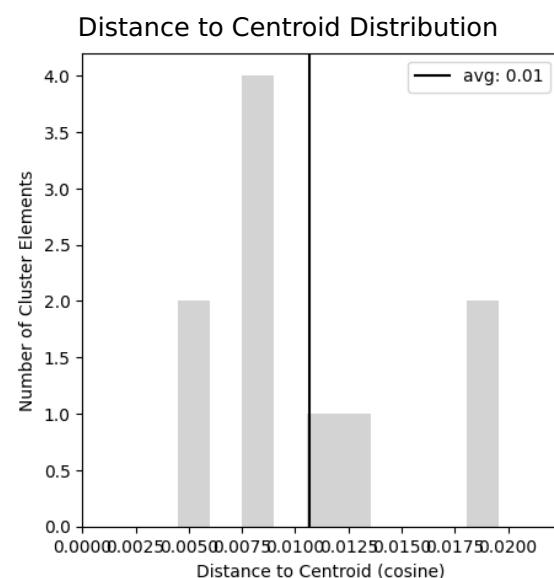
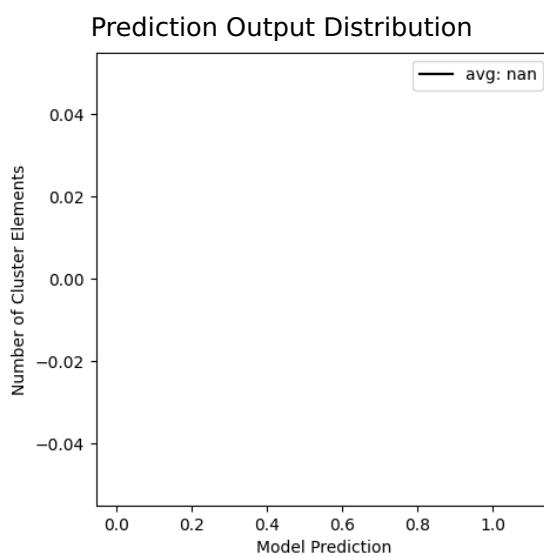
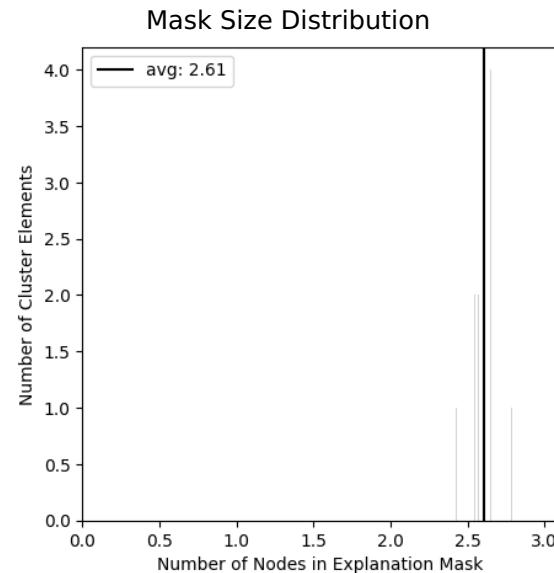
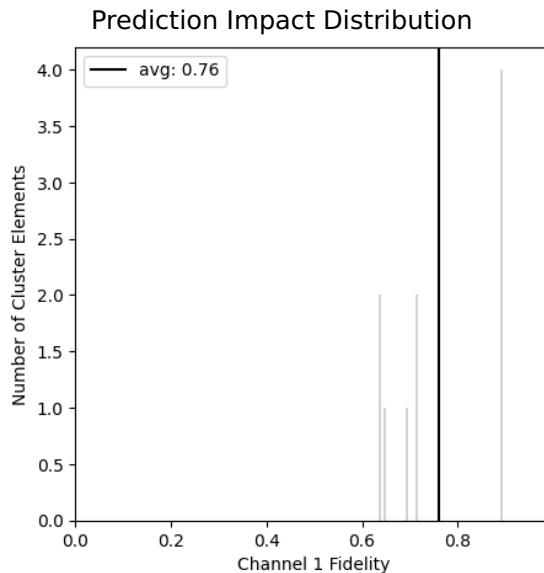
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	10
Channel Index	1.0 (0.0)

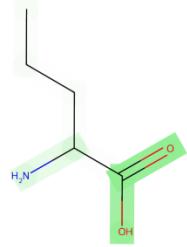
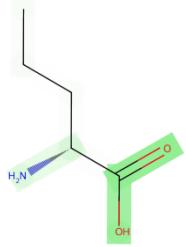
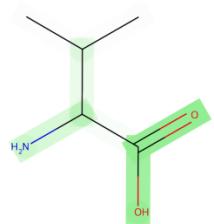
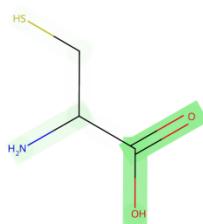
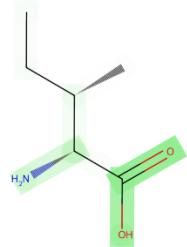
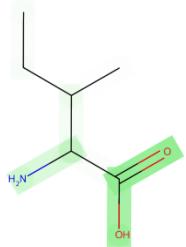
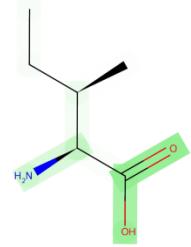
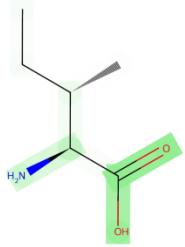
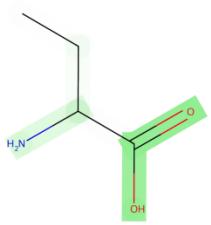
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The molecular fragment " $O=C-O$ " represents a carboxyl group, which is highly polar and can participate in hydrogen bonding with water molecules, thus increasing the solubility of the molecule in water. The carbon double-bonded to oxygen creates a partial positive charge on the carbon and a partial negative charge on the oxygen. This polar nature along with the -OH group allows for hydrogen bonding where the hydrogen atoms of water can interact with the carbonyl oxygen, and the oxygen atom of the carboxyl group can form hydrogen bonds with the hydrogen atoms of water molecules. These interactions are a key factor in the dissolution process as they can overcome the cohesive forces within the water and the compound, resulting in increased solubility.

**Hypothesis:** Molecules containing the " $O=C-O$ " substructure have a tendency to be more soluble in water. This is because the carboxyl group forms favorable hydrogen bond interactions with water molecules, and the polar nature of the carboxyl group enhances the molecule's affinity for the aqueous environment.

# Cluster #77 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 77, from importance channel 1 (*positive*), represents a motif consisting of 3.1 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 1.0 ( $\pm 0.1$ ) on the prediction outcome.

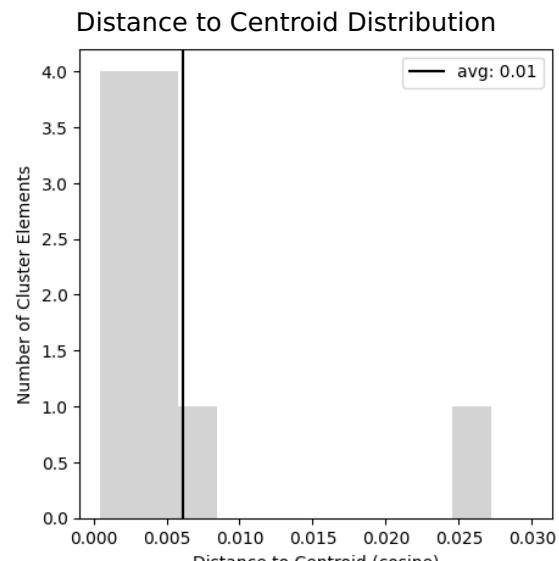
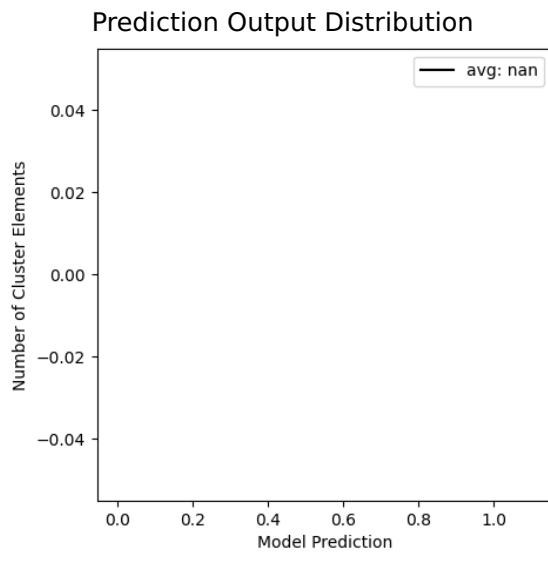
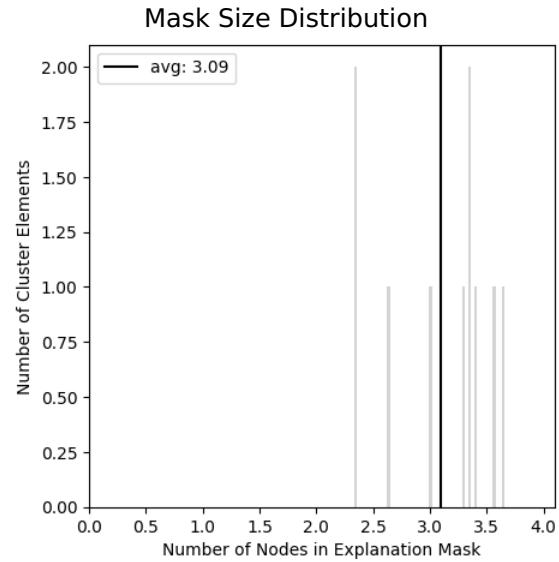
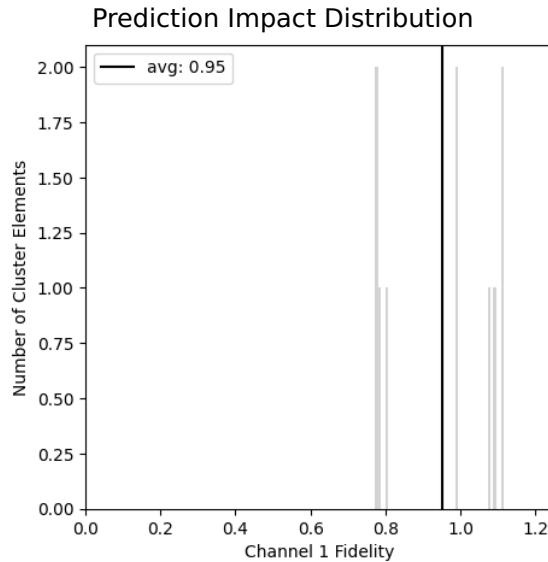
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	10
Channel Index	1.0 (0.0)

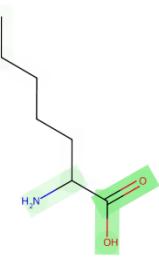
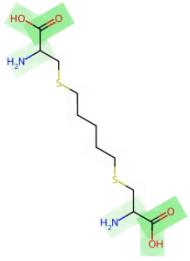
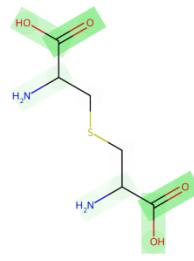
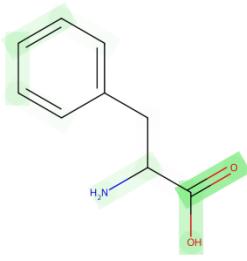
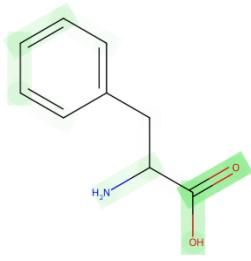
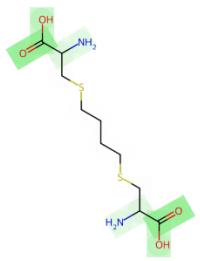
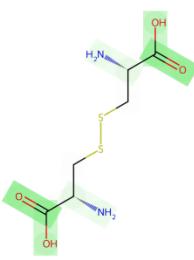
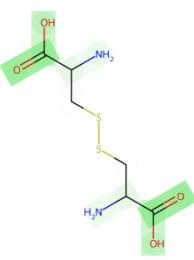
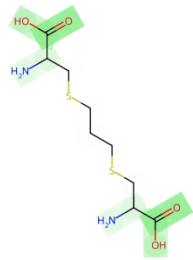
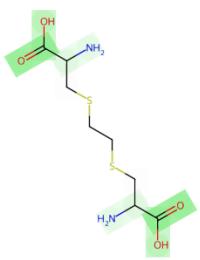
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



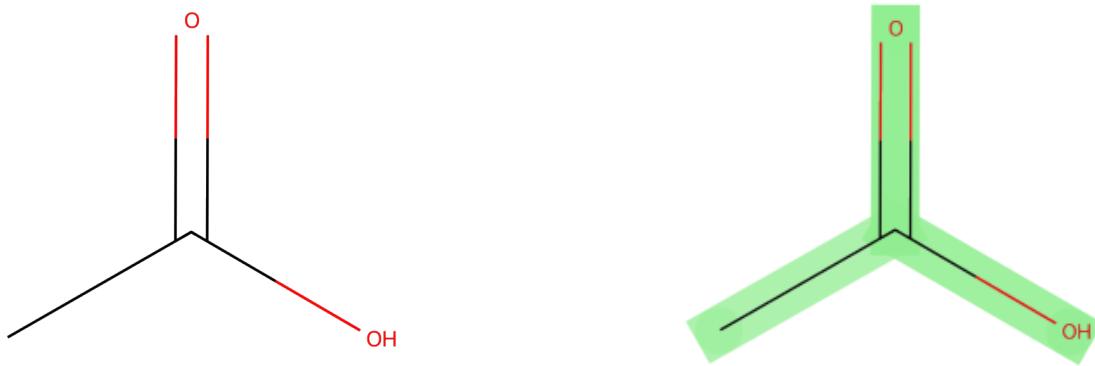
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure represented by the SMILES string "C-C(=O)-O" includes a carboxyl group, which is known to impart high water solubility to molecules. The carboxyl group consists of a carbon double-bonded to an oxygen (carbonyl group) and single bonded to a hydroxyl group. The high electronegativity of the oxygen atom in both carbonyl and hydroxyl groups allows for the formation of hydrogen bonds with water molecules. Hydrogen bonding is a key factor in determining water solubility, as it allows for the solute to be better integrated into the hydrogen bond network of water.

**Hypothesis:** Molecules containing the substructure "C-C(=O)-O" are likely to be highly soluble in water. This is because the carboxyl group can participate in hydrogen bonding, which facilitates the dissolution of the molecule in the aqueous environment. Consequently, the presence of this functional group greatly enhances the water-attracting characteristics of a molecule due to these potential hydrogen bond interactions.

# Cluster #78 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 78, from importance channel 1 (*positive*), represents a motif consisting of 3.2 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 0.9 ( $\pm 0.2$ ) on the prediction outcome.

## Properties

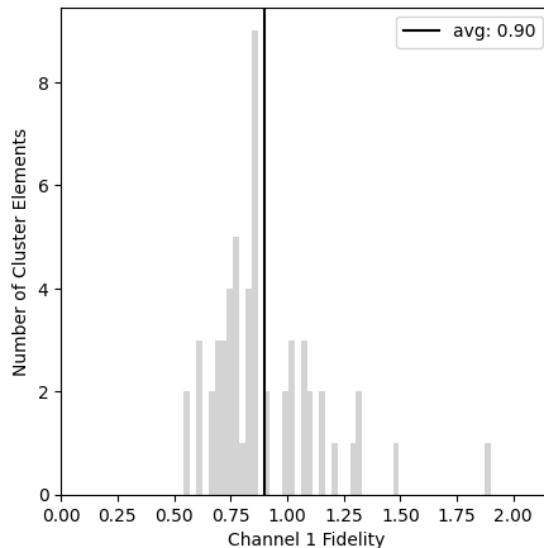
ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	56
Channel Index	1.0 (0.0)

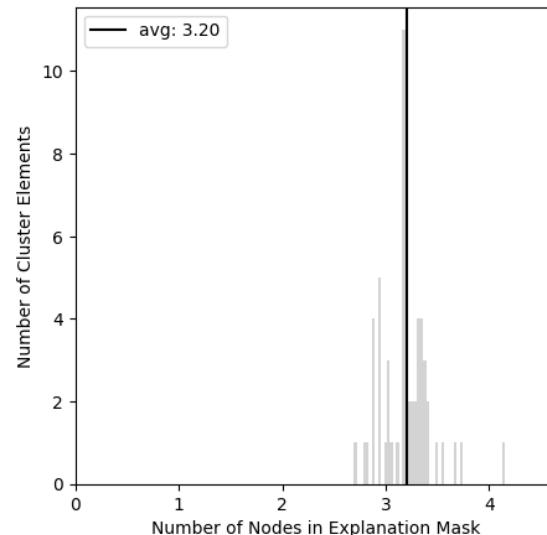
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.

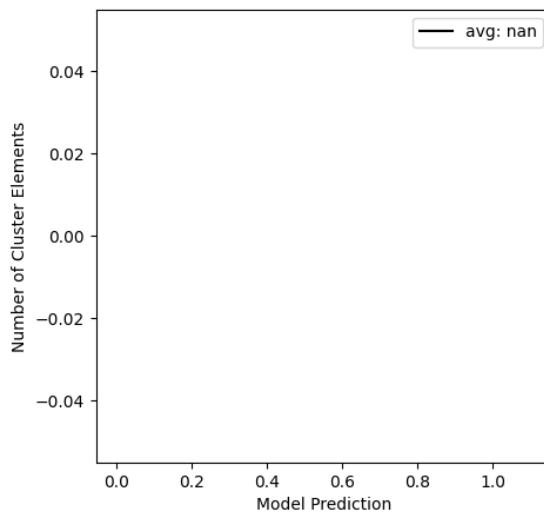
Prediction Impact Distribution



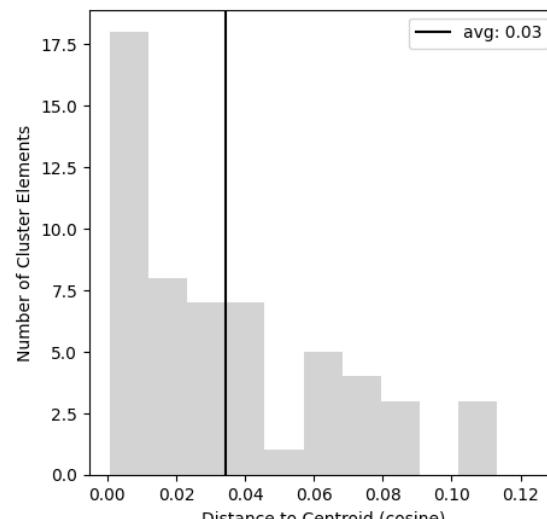
Mask Size Distribution



Prediction Output Distribution

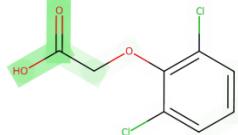
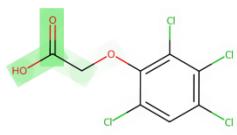
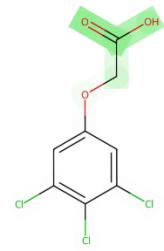
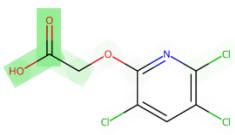
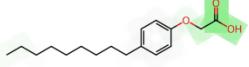
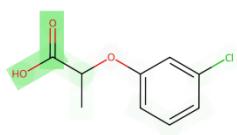
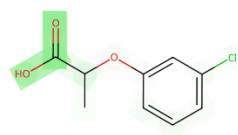
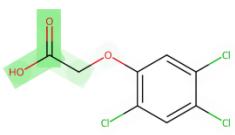
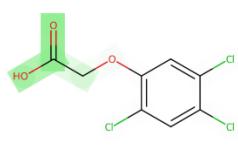
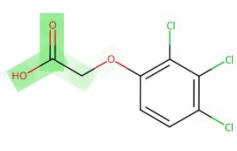
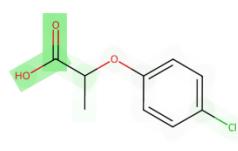
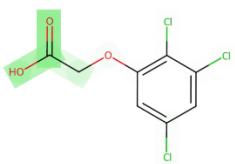
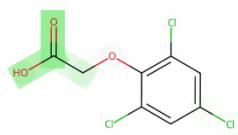
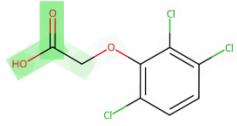
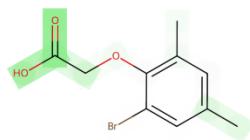
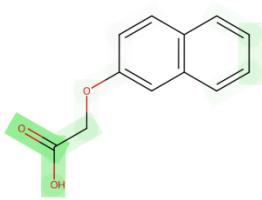


Distance to Centroid Distribution



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a di-carboxylic acid, a molecule that contains two carboxyl functional groups (-COOH). Carboxyl groups are known to enhance water solubility due to their ability to form hydrogen bonds with water molecules. Hydrogen bonding increases the interaction between the solute and solvent, facilitating dissolution. Additionally, the presence of two such groups increases the molecule's polarity, which is another factor that enhances solubility in water, a polar solvent.

**Hypothesis:** Molecules with the substructure -C(=O)-OH have an increased water solubility. The presence of two carboxyl groups in the molecule, as indicated by the SMILES "O=C(-O)-C-C(=O)-O", suggests a high potential for hydrogen bonding and increased polarity, leading to the hypothesized enhanced water solubility.

# Cluster #79 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 79, from importance channel 1 (*positive*), represents a motif consisting of 2.8 ( $\pm 0.5$ ) nodes. The concept is generally associated with an impact of 0.9 ( $\pm 0.3$ ) on the prediction outcome.

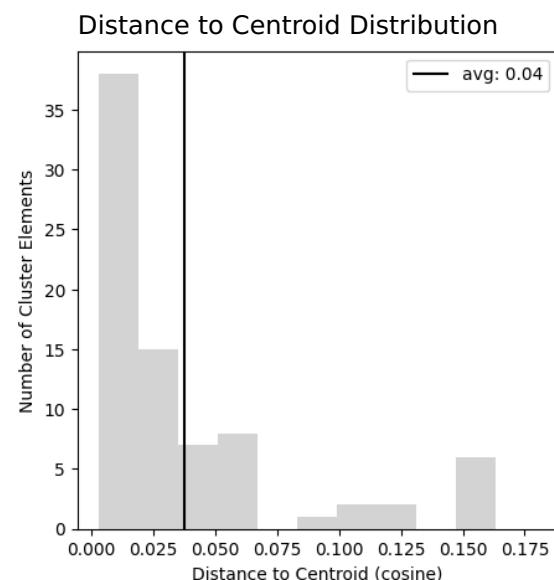
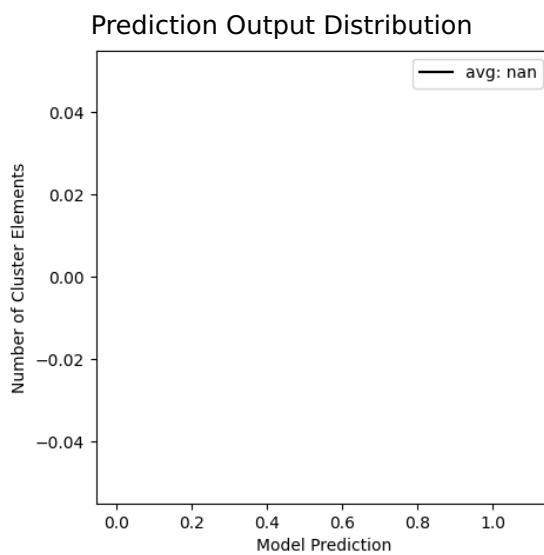
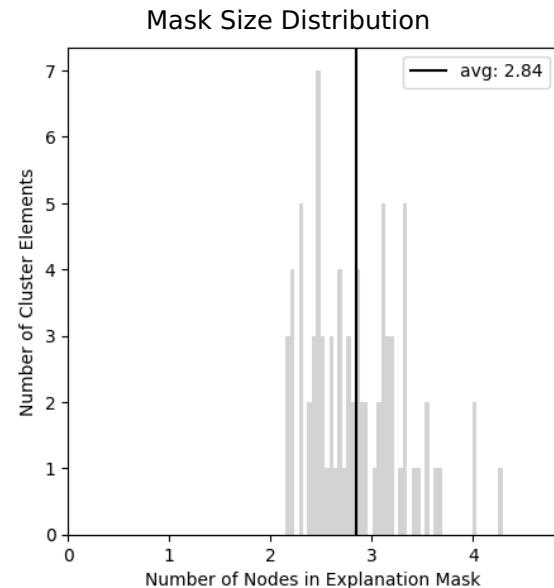
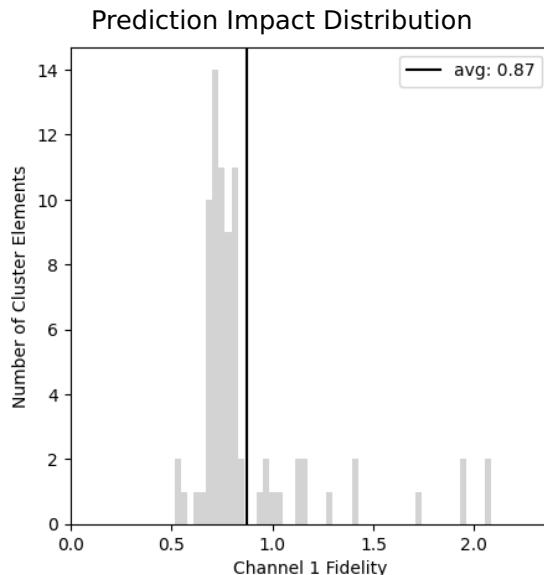
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	79
Channel Index	1.0 (0.0)

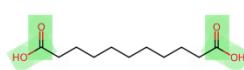
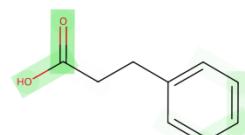
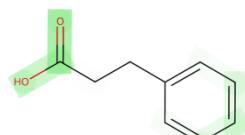
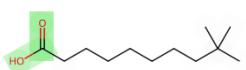
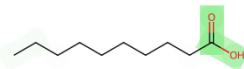
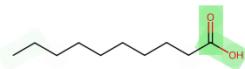
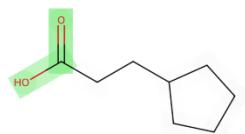
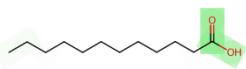
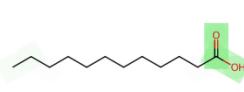
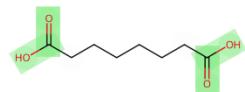
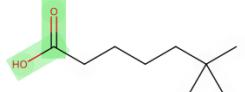
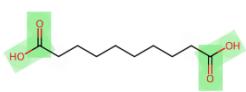
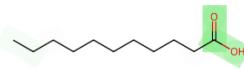
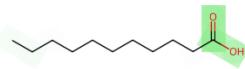
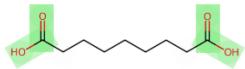
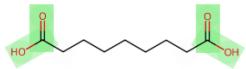
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given molecular substructure "O=C(-O)-C-S" suggests the presence of polar functional groups, such as the carboxylic acid (-COOH) and thioether (-S-). Polar functional groups are known to form hydrogen bonds with water molecules, which is a key factor in enhancing solubility. The carboxylic acid group, with both a carbonyl (C=O) and a hydroxyl (O-H) group, can donate and accept hydrogen bonds, making it highly soluble in water. The thioether group, while less polar than the hydroxyl or carbonyl, still has some polarity due to the presence of sulfur and can also contribute to solubility through dipole-dipole interactions.

**Hypothesis:** Molecules containing the substructure "O=C(-O)-C-S" are likely to have increased water solubility. This is attributed to the capability of the carboxylic acid group to engage in hydrogen bonding and the polarity of the thioether group, which can both interact with water to enhance solubility.

# Cluster #80 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 80, from importance channel 1 (*positive*), represents a motif consisting of 2.8 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 0.6 ( $\pm 0.1$ ) on the prediction outcome.

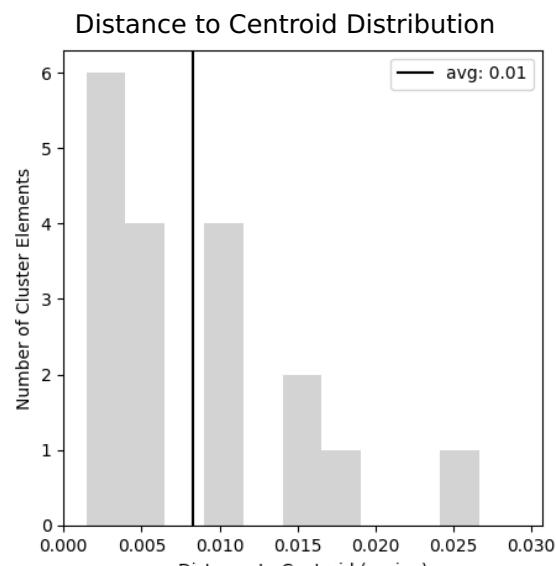
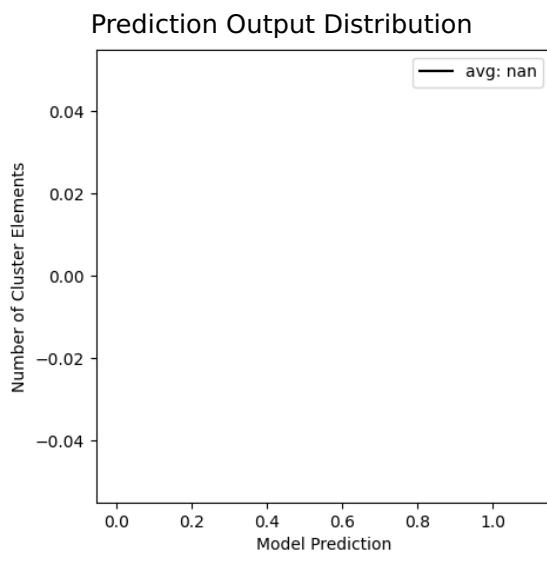
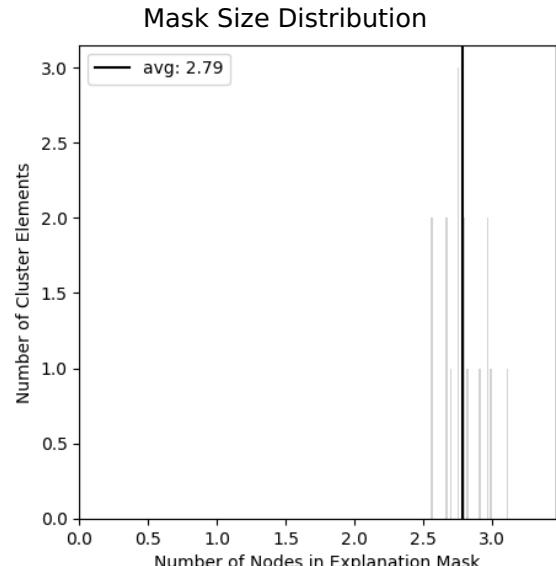
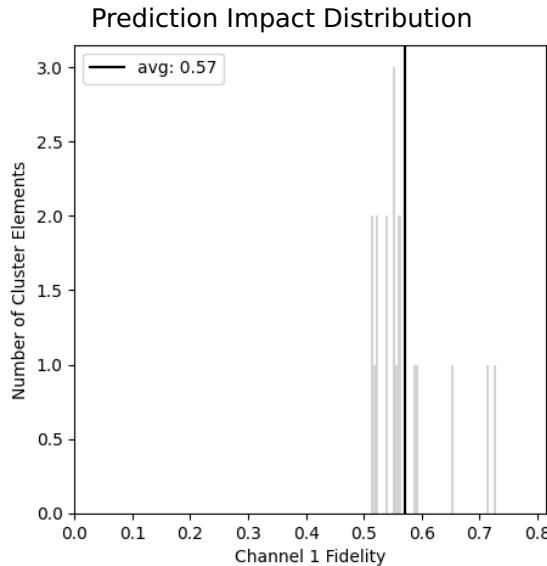
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	18
Channel Index	1.0 (0.0)

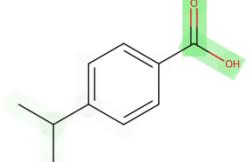
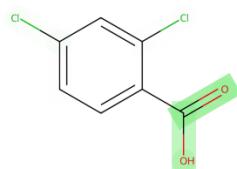
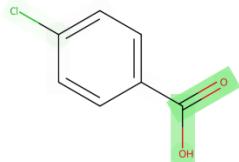
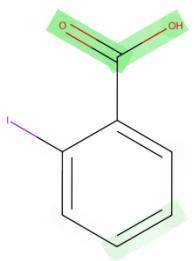
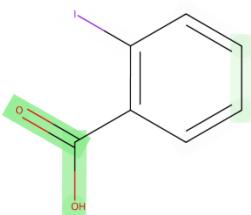
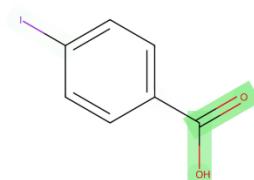
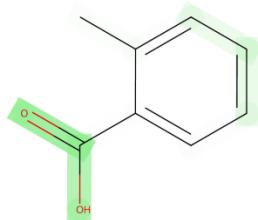
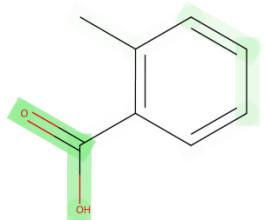
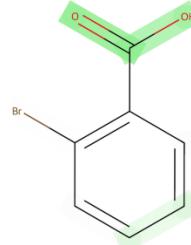
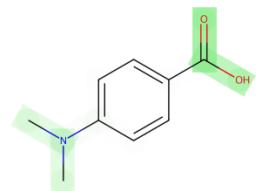
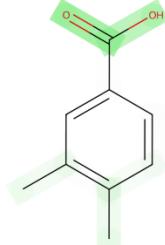
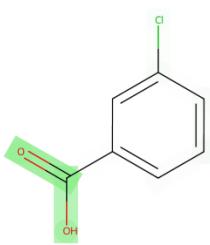
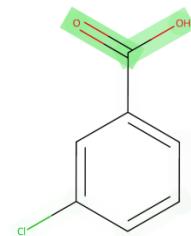
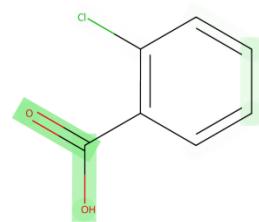
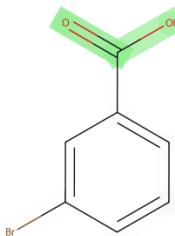
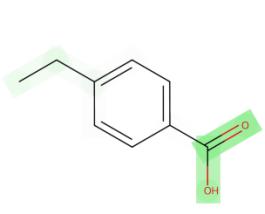
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



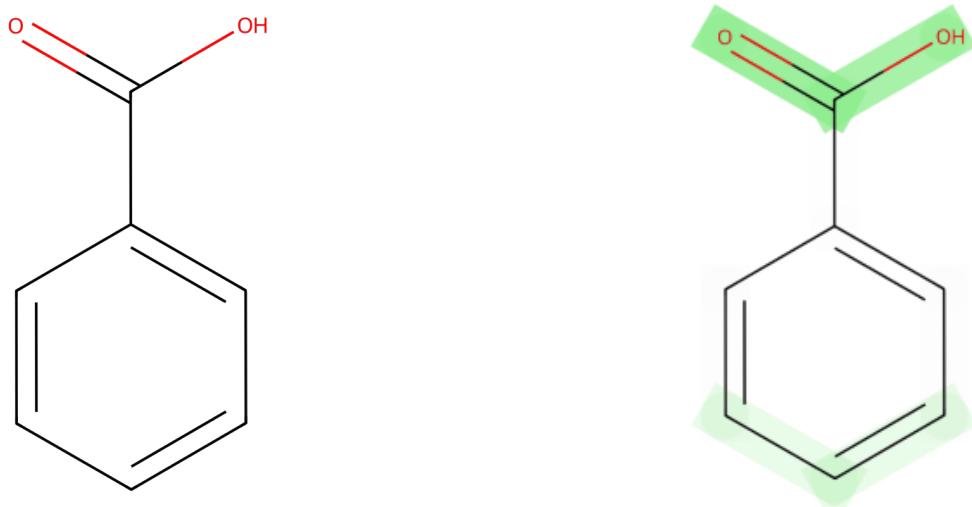
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES representation "O=C(-O)-c1ccccc1" describes a structure related to benzene with a carboxylic acid functional group (COOH) attached to it. Carboxylic acids are known to enhance water solubility due to their ability to donate a proton to water, thereby forming a carboxylate ion, which is more soluble. Additionally, the benzene ring is hydrophobic, which typically decreases solubility in water, but the presence of the carboxylic acid group can offset this through hydrogen bonding with water molecules.

**Hypothesis:** Molecules with a benzene ring attached to a carboxylic acid group will show a moderate increase in water solubility. The carboxylic acid is a solubilizing group due to its capacity to engage in hydrogen bond formation with water, while the benzene ring slightly reduces this effect due to its hydrophobic nature. However, the overall structure leans towards increased solubility in water because of the presence of the carboxylic acid group.

# Cluster #81 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 81, from importance channel 1 (*positive*), represents a motif consisting of 3.0 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 0.8 ( $\pm 0.1$ ) on the prediction outcome.

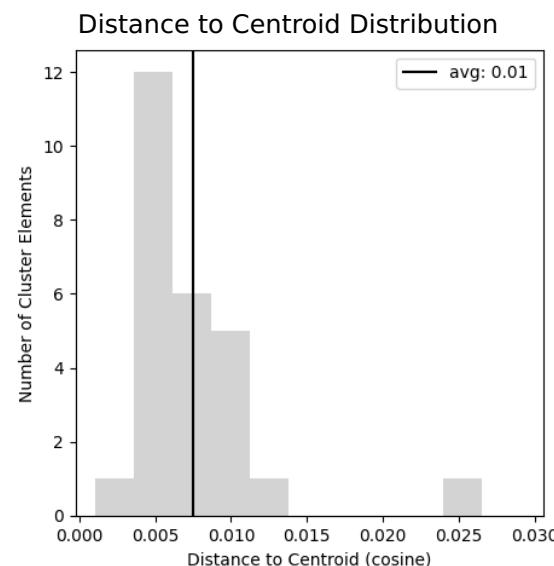
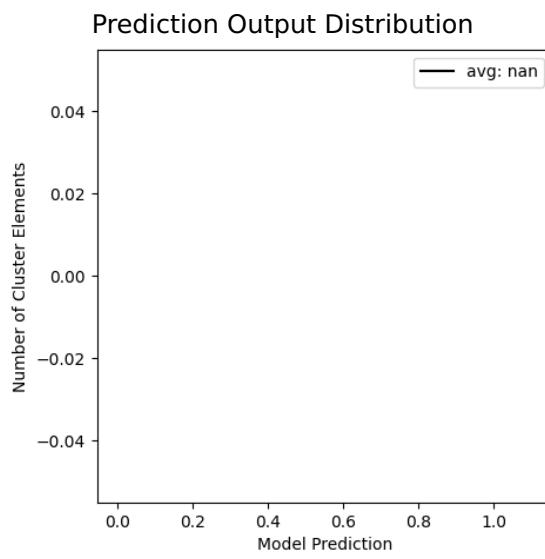
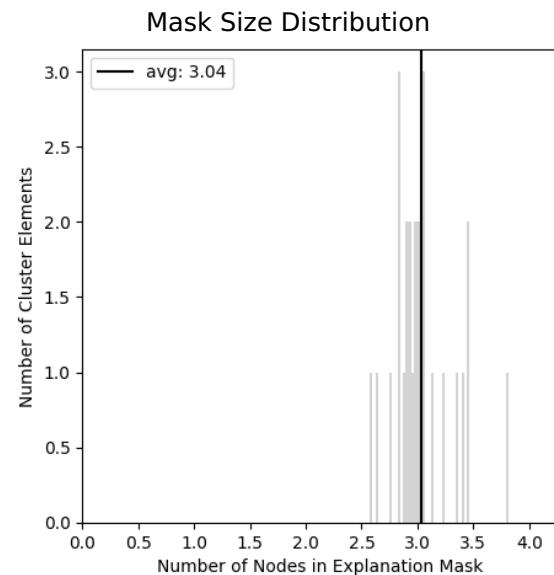
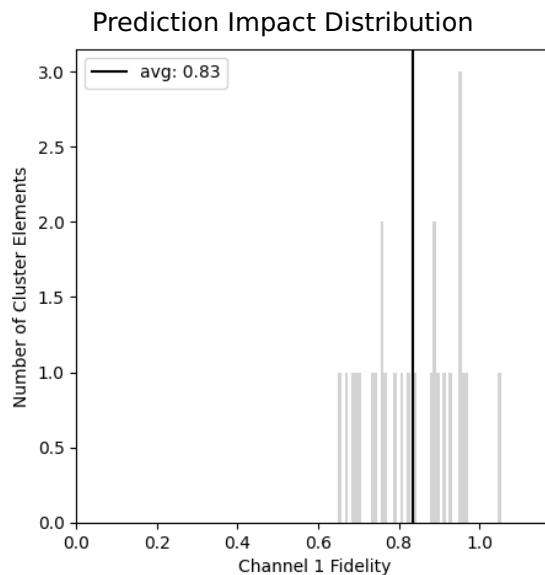
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	26
Channel Index	1.0 (0.0)

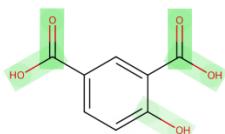
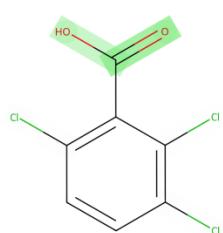
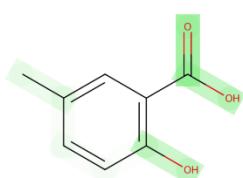
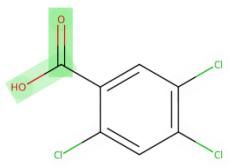
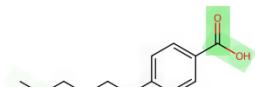
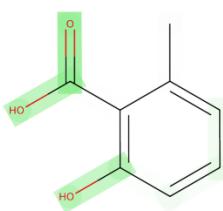
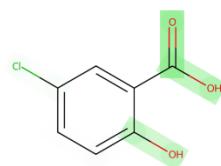
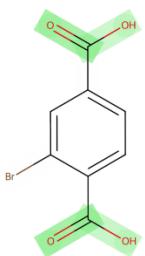
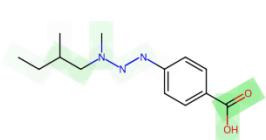
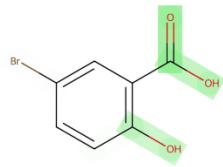
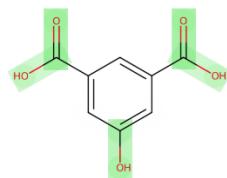
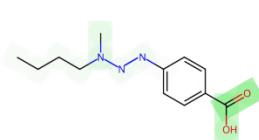
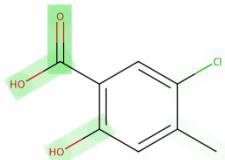
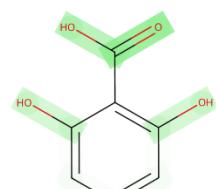
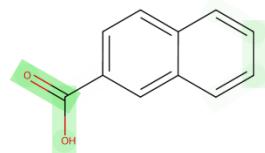
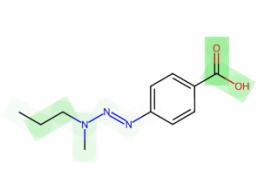
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



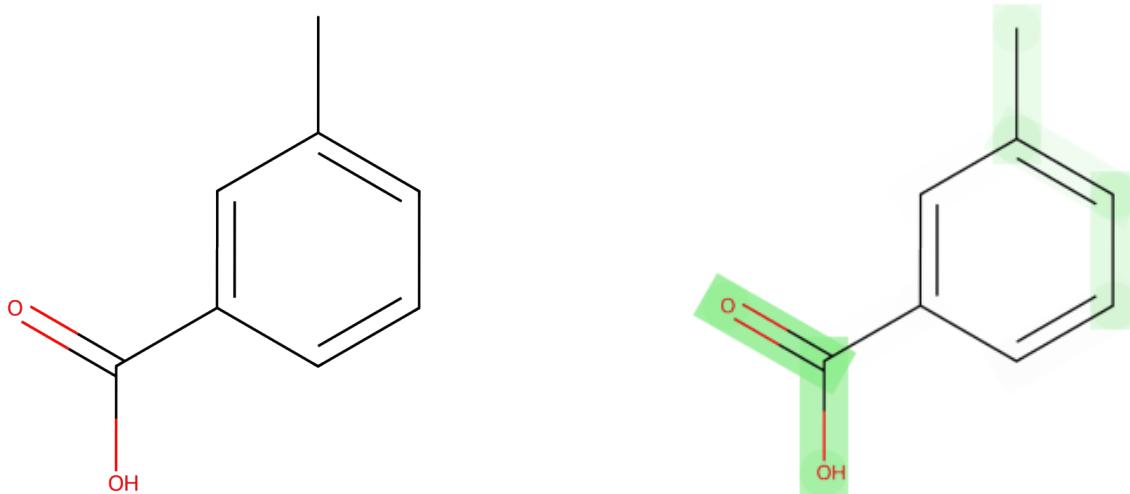
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a phenyl ring substituted with an aliphatic chain bearing a carboxylic acid group. The carboxylic acid group, noted as "-C(=O)-O", is a strong contributor to water solubility due to its ability to form hydrogen bonds with water molecules. Hydrogen bonding increases the interaction between the molecule and water, thereby enhancing solubility. The overall non-polar nature of the benzene ring may somewhat reduce solubility, but the effect of the polar carboxylic acid is significant enough to have a net positive influence on the molecule's solubility in water.

**Hypothesis:** Molecules containing a carboxylic acid group attached to a phenyl ring are likely to be more soluble in water due to the hydrogen-bonding capability of the carboxyl group. The polar carboxyl function can interact strongly with water, overcoming the relatively lower solubility contribution of the non-polar benzene ring, which results in a structure that enhances water solubility overall.

# Cluster #82 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 82, from importance channel 1 (*positive*), represents a motif consisting of 2.9 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 0.9 ( $\pm 0.3$ ) on the prediction outcome.

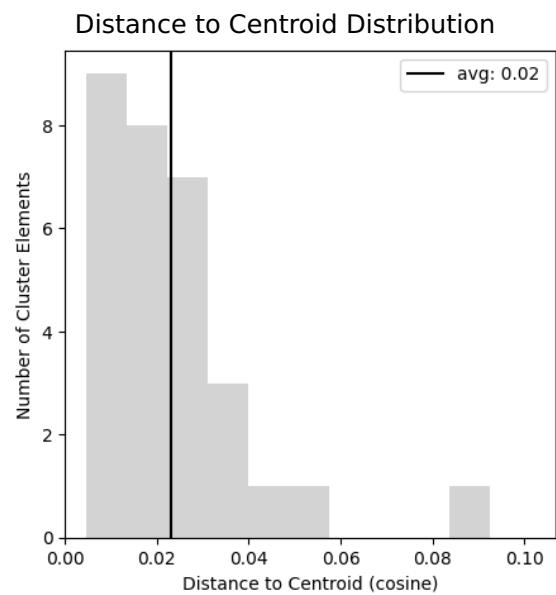
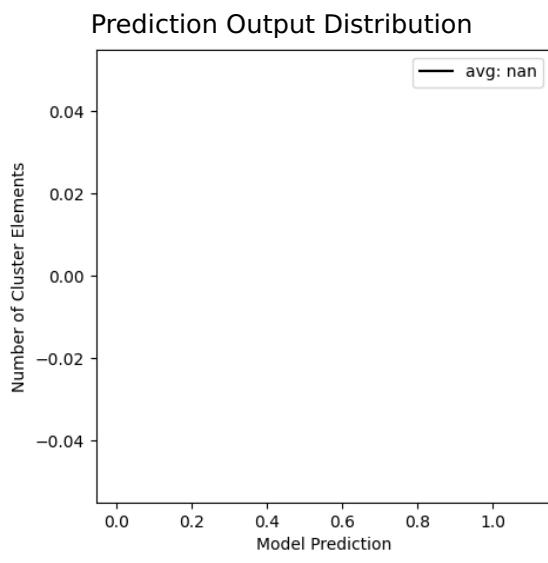
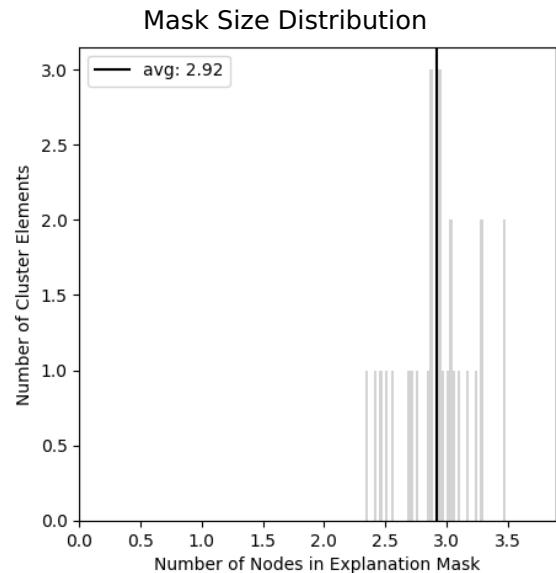
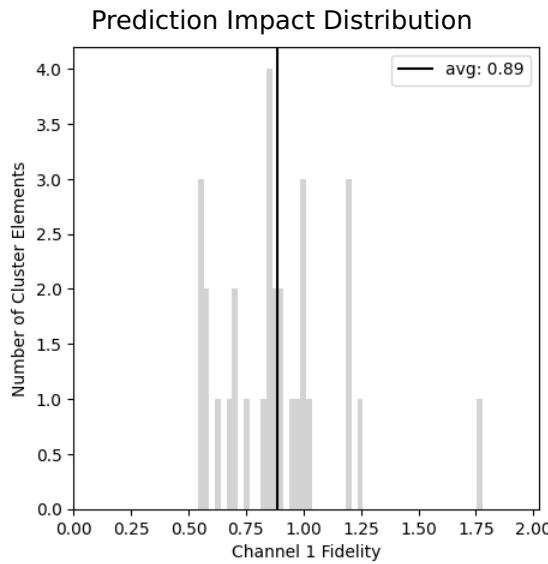
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	30
Channel Index	1.0 (0.0)

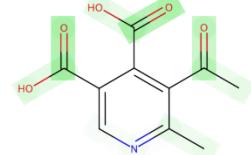
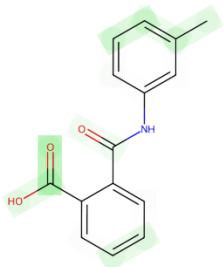
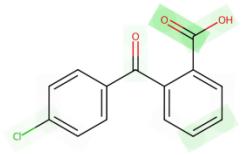
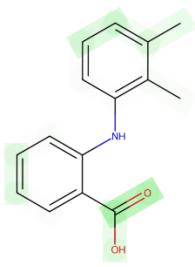
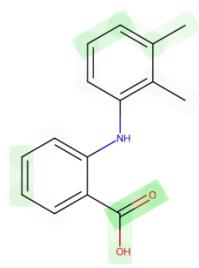
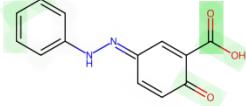
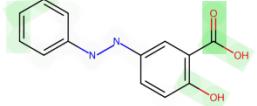
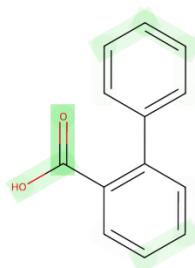
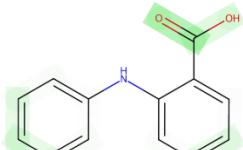
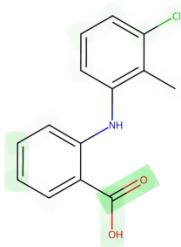
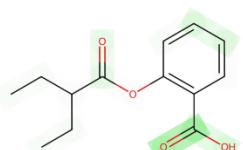
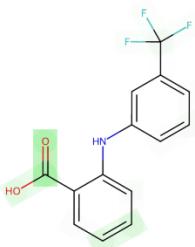
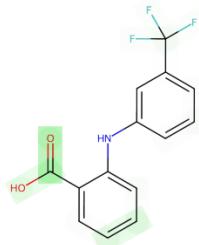
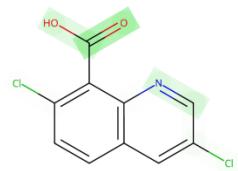
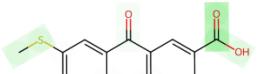
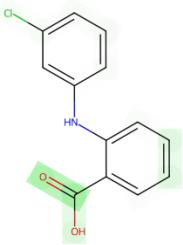
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



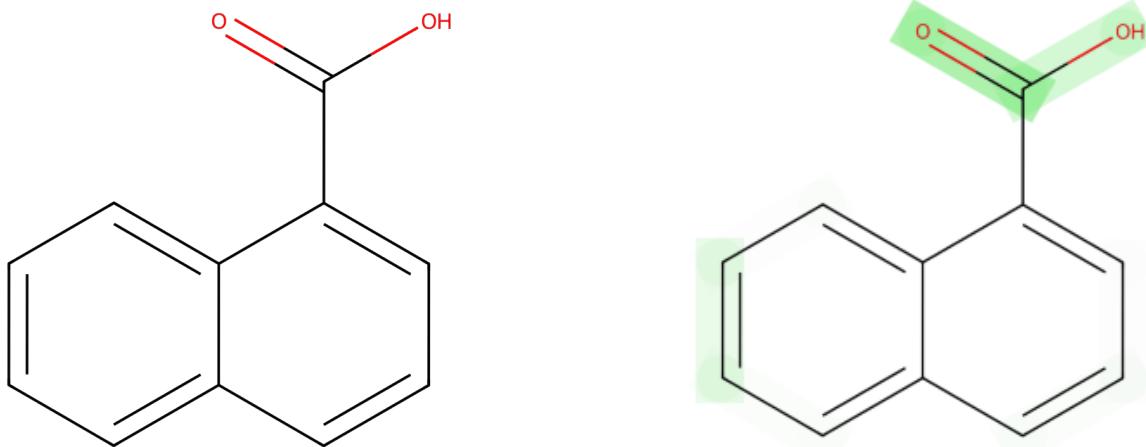
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The SMILES representation "O=C(-O)-c1:c:c:c:c2:c:c:c:c:1:2" depicts a structure with a carboxylic acid (-COOH) functional group attached to a fused aromatic ring system. The carboxylic acid group is known for its high solubility in water due to its ability to donate a hydrogen ion ( $H^+$ ) to the water, forming a charged carboxylate ion that interacts well with the polar water molecules. Additionally, the presence of the fused aromatic ring system provides a planar and relatively non-polar region that typically reduces water solubility. However, the influence of the polar carboxylic acid group in this case seems to be significantly overriding the effect of the fused rings.

**Hypothesis:** Molecules containing a carboxylic acid group attached to a fused aromatic ring system generally have a high tendency to dissolve in water due to the hydrophilic nature of the carboxylic acid. The formation of a charged species when the carboxylic acid ionizes enhances water's ability to solvate the molecule, while the fused aromatic system's contribution to solubility is less pronounced in the presence of the strongly polar carboxyl group.

# Cluster #83 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 83, from importance channel 1 (*positive*), represents a motif consisting of 2.9 ( $\pm 0.1$ ) nodes. The concept is generally associated with an impact of 0.7 ( $\pm 0.1$ ) on the prediction outcome.

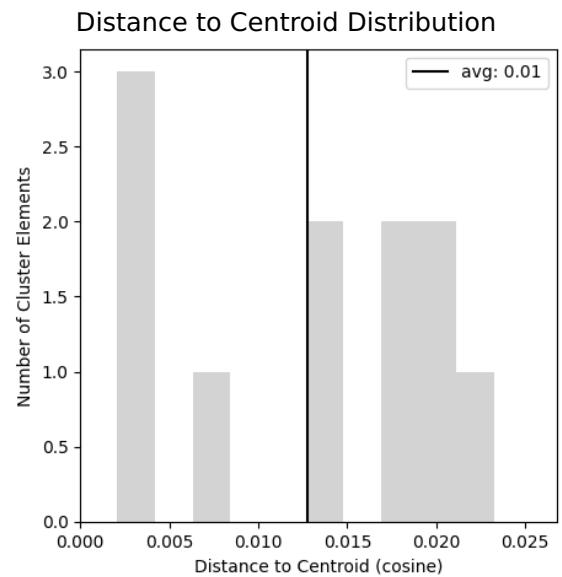
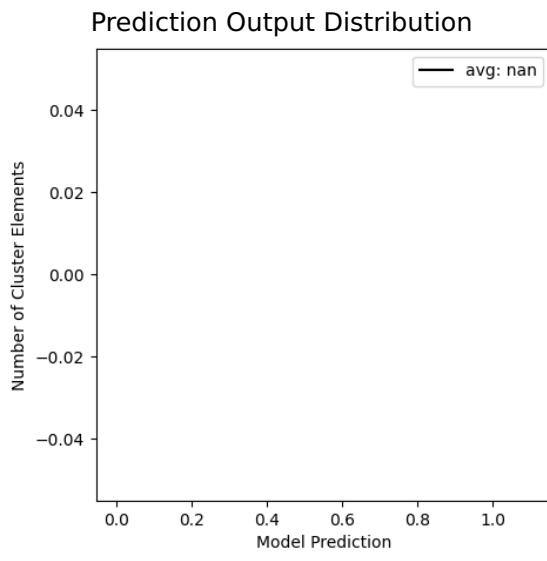
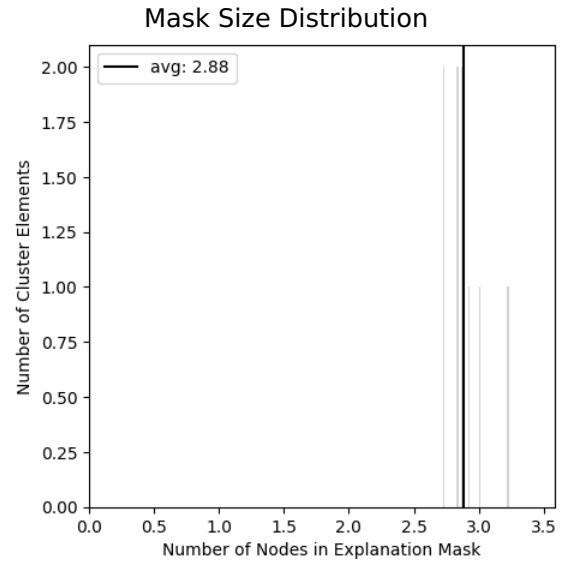
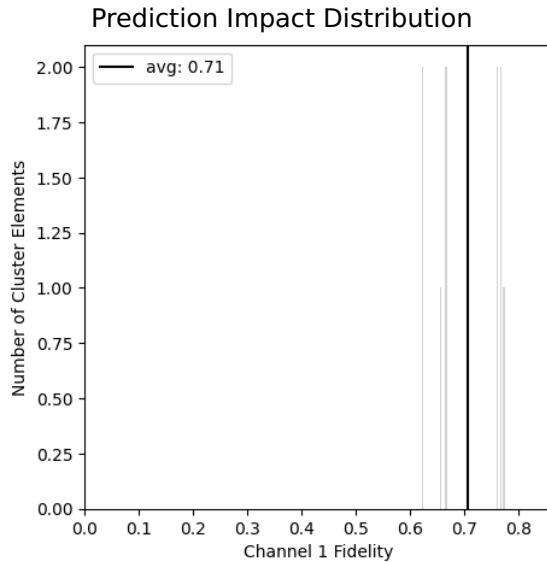
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	11
Channel Index	1.0 (0.0)

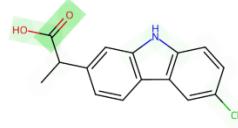
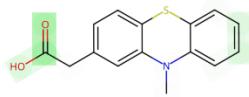
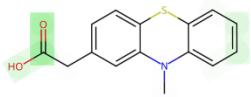
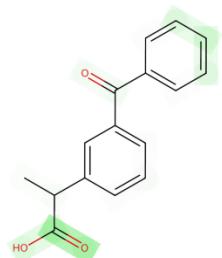
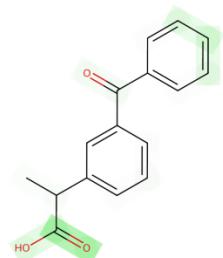
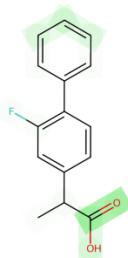
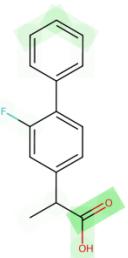
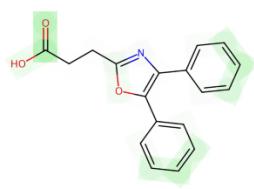
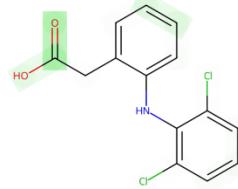
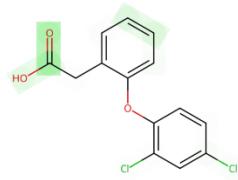
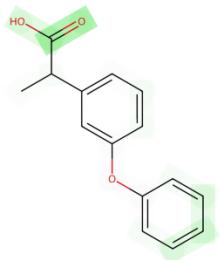
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



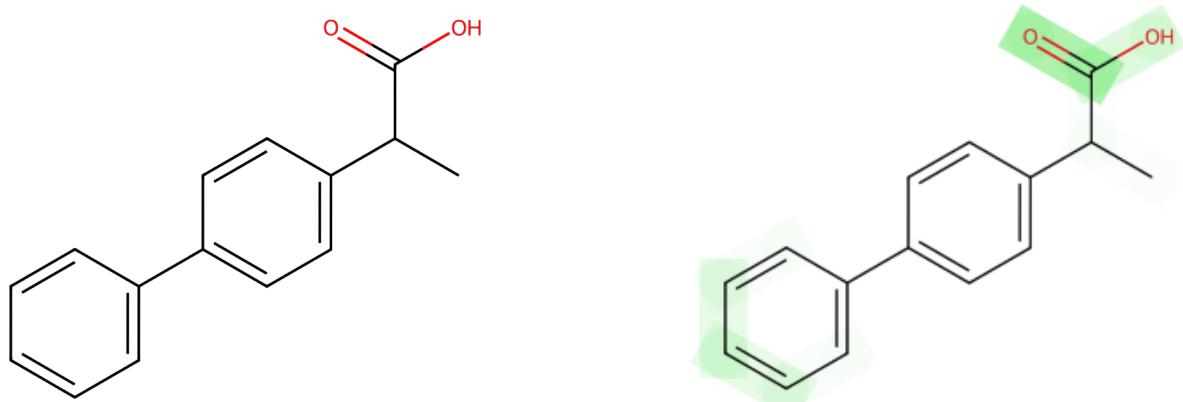
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The molecular substructure represented by this SMILES string features a benzene ring linked to a carboxylic acid group through a bridge. The carboxylic acid group, which comprises a carbonyl (C=O) and a hydroxyl (O-H) group, is known for its ability to form hydrogen bonds with water molecules due to the polarity of the O-H bond and the partially negative charge on the oxygen atom. This feature significantly contributes to a molecule's solubility in water. Moreover, the benzene ring, while generally hydrophobic, possesses a certain degree of solubility in water when attached to hydrophilic functional groups like the carboxylic acid, due to an overall polar character being introduced to the aromatic system.

**Hypothesis:** Molecules containing a benzene ring attached to a carboxylic acid group are expected to have an increased solubility in water. The carboxylic acid group enhances solubility through hydrogen bonding with water, while the benzene ring's hydrophobic nature is mitigated by the polar nature of the attached substituent, thus allowing the molecule to better interact with the aqueous environment.

# Cluster #84 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 84, from importance channel 1 (*positive*), represents a motif consisting of 3.1 ( $\pm 0.2$ ) nodes. The concept is generally associated with an impact of 0.8 ( $\pm 0.2$ ) on the prediction outcome.

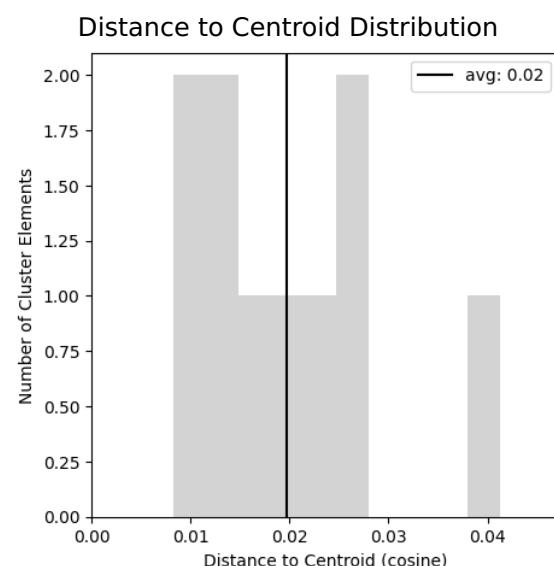
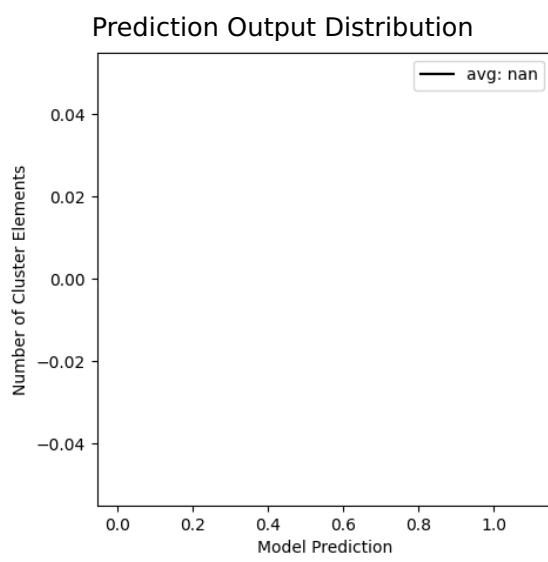
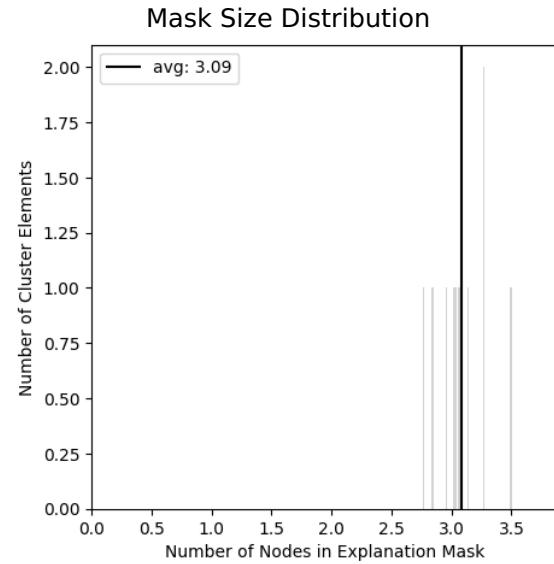
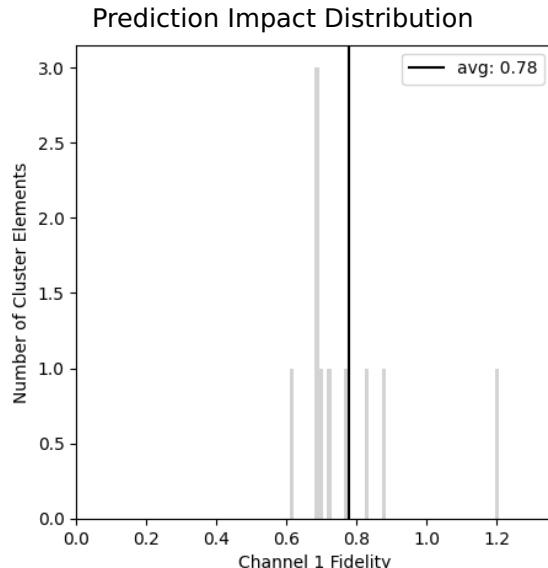
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	10
Channel Index	1.0 (0.0)

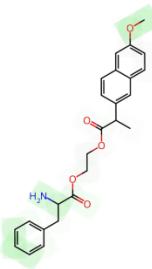
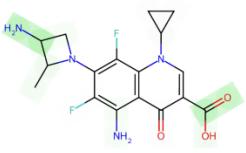
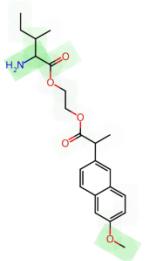
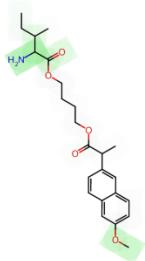
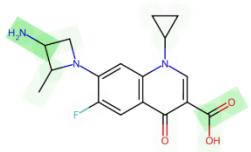
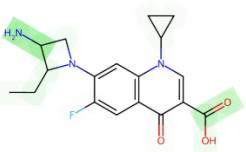
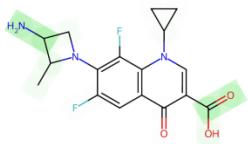
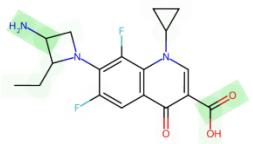
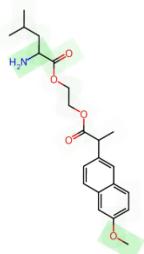
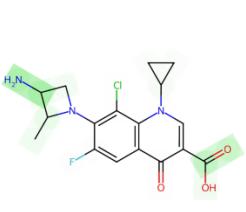
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



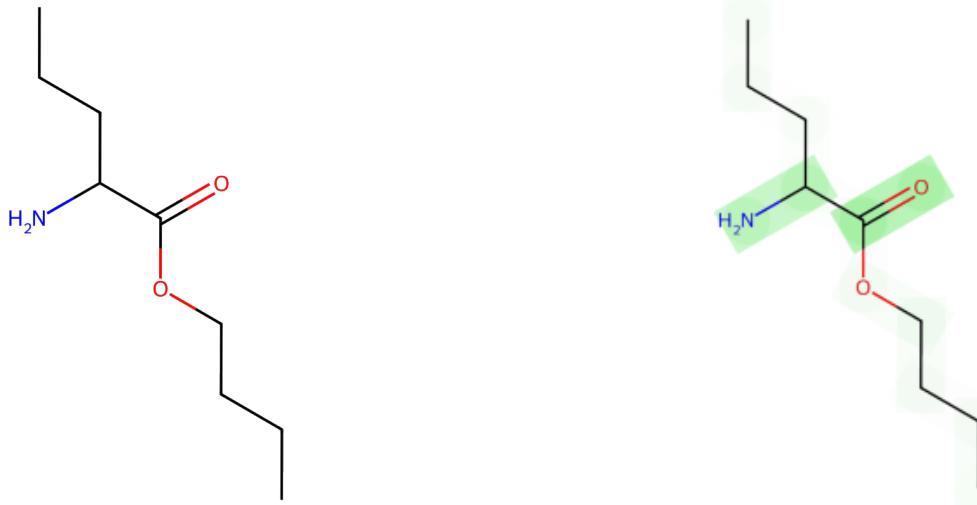
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The molecular structure represented by the given SMILES notation "C-C-C-O-C(=O)-C(-N)-C-C-C" indicates the presence of both hydrophobic (carbon chains) and hydrophilic (oxygen and nitrogen containing) functional groups. The hydroxyl group (-OH from the oxygen) and the carboxyl group (-C(=O)) along with the amine group (-N) can form hydrogen bonds with water, increasing the solubility. The carbon chains, on the other hand, are nonpolar and tend to decrease water solubility. The position of the hydrophilic groups within the molecular framework is such that they can interact favorably with water molecules, overcoming the hydrophobic tendencies of the carbon chains to some extent.

**Hypothesis:** The SMILES structure "C-C-C-O-C(=O)-C(-N)-C-C-C" is moderately influential towards increasing water solubility. The oxygen and nitrogen in the structure can form hydrogen bonds with water, enhancing solubility, while the hydrocarbon tail remains a limiting factor due to its hydrophobic nature, which compromises some of the molecule's overall water solubility.

# Cluster #85 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 85, from importance channel 1 (*positive*), represents a motif consisting of 3.7 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 0.9 ( $\pm 0.0$ ) on the prediction outcome.

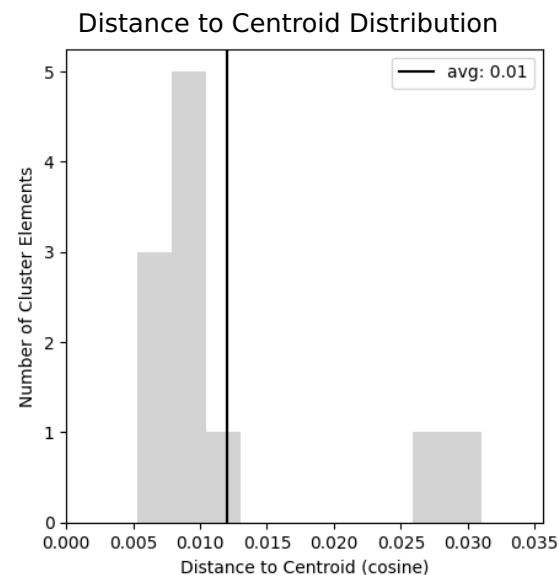
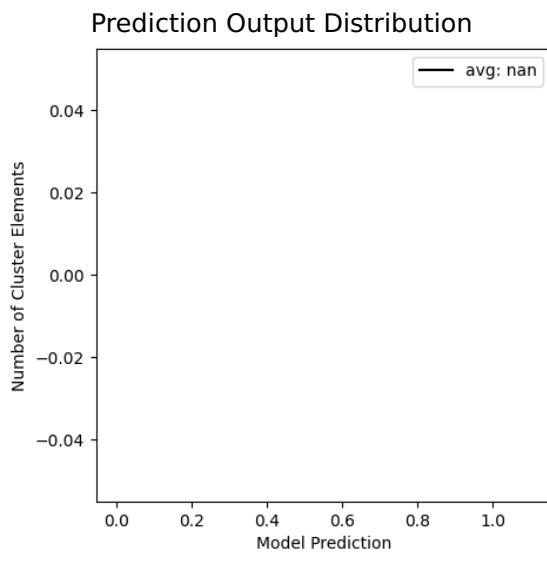
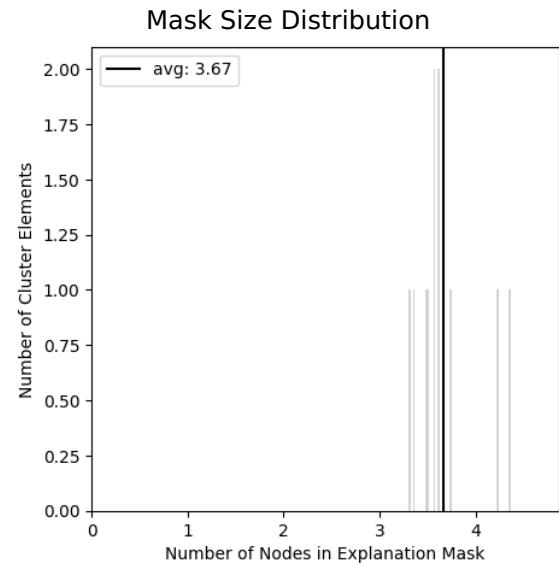
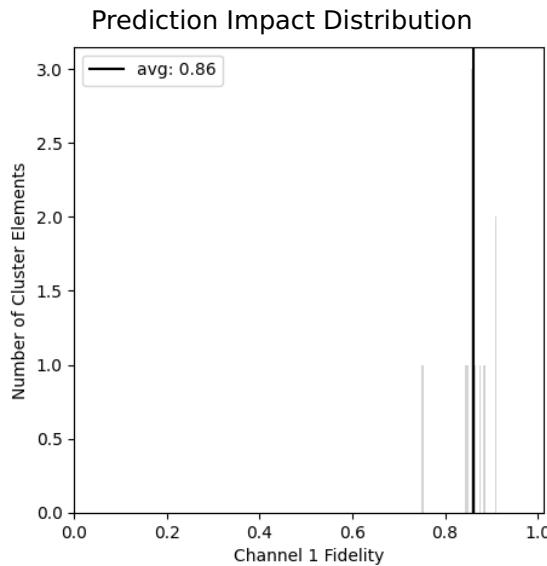
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	11
Channel Index	1.0 (0.0)

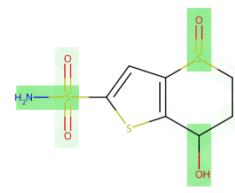
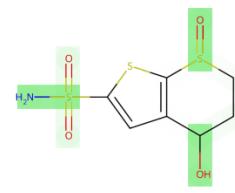
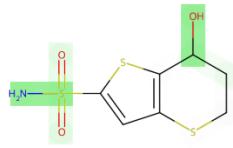
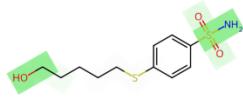
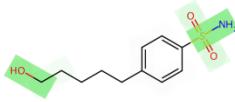
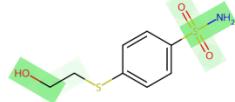
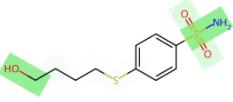
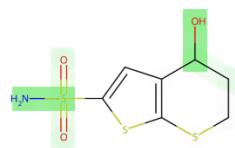
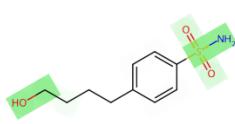
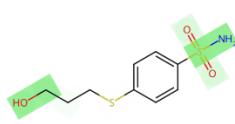
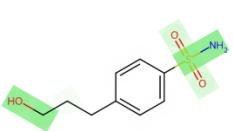
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



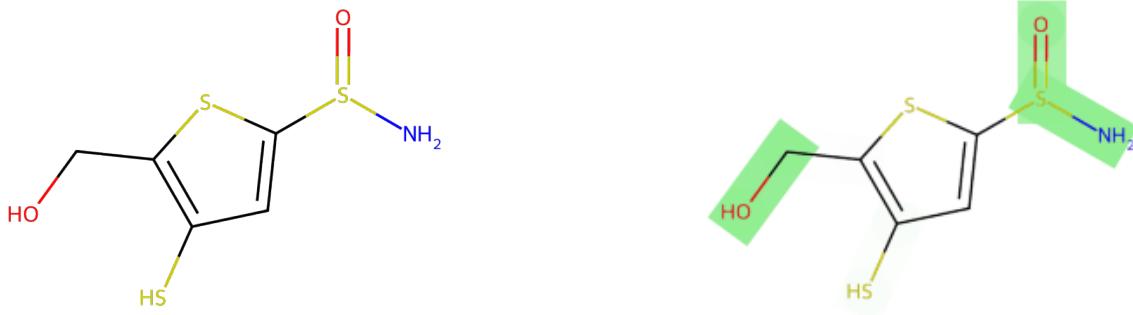
## Example Elements

① This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The structure given by the SMILES representation "N-S(=O)-c1:c:c(-S):c(-C-O):s:1" includes a variety of functional groups that are known to impact water solubility. The presence of a sulfonamide group (N-S(=O)), a thiol group (-S), and an alcohol group (-C-O) in the aromatic system suggests an ability to form hydrogen bonds with water molecules. Hydrogen bonding is a key interaction that enhances solubility in water. The aromatic system itself usually decreases solubility due to its hydrophobic nature, but the substituents in this case provide polar character to the molecule, mitigating the hydrophobic effect.

**Hypothesis:** Molecules with the "N-S(=O)-c1:c:c(-S):c(-C-O):s:1" substructure are likely to have increased water solubility due to multiple polar functional groups capable of hydrogen bonding with water. The sulfonamide and alcohol groups can serve as hydrogen bond donors or acceptors, while the thiol may also contribute to solubility through polarity and potential ionization. Despite the hydrophobic character of the aromatic ring, the overall effect of the substituents leads to a molecule more amenable to interaction with water, hence the observed positive influence on solubility.

# Cluster #86 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 86, from importance channel 1 (*positive*), represents a motif consisting of 4.7 ( $\pm 1.4$ ) nodes. The concept is generally associated with an impact of 0.9 ( $\pm 0.5$ ) on the prediction outcome.

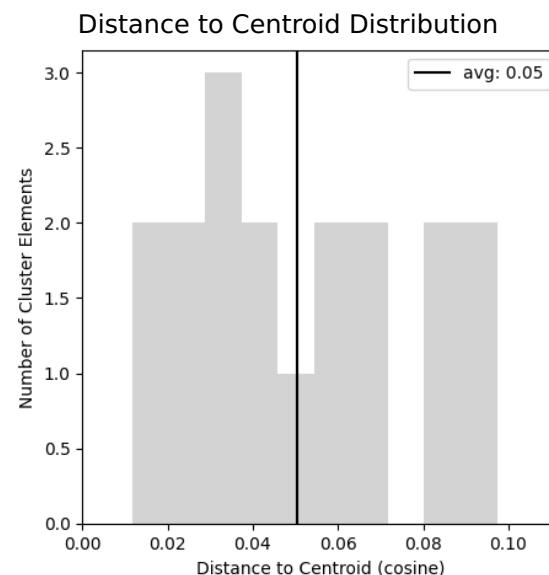
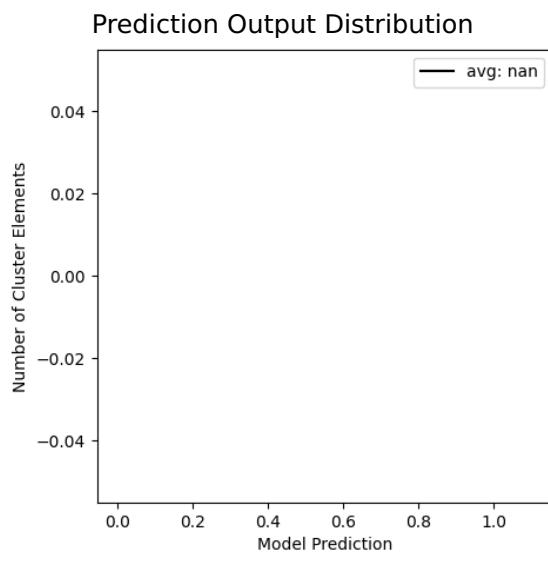
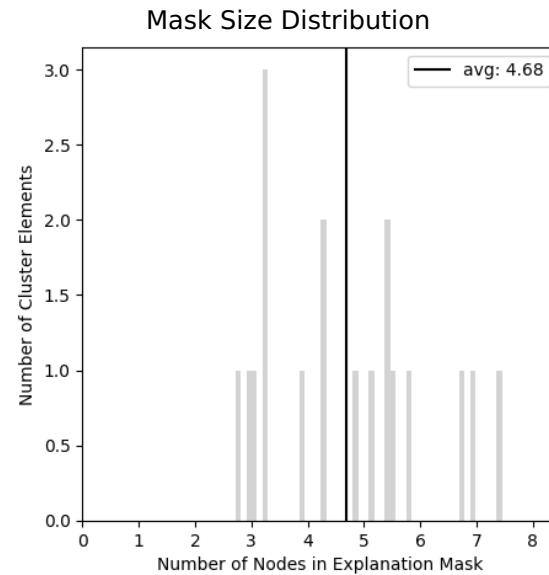
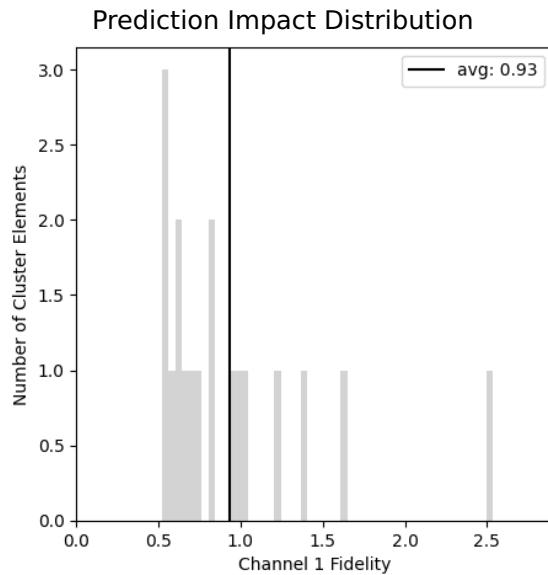
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	18
Channel Index	1.0 (0.0)

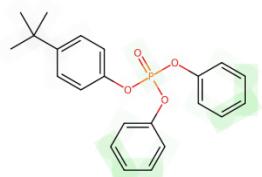
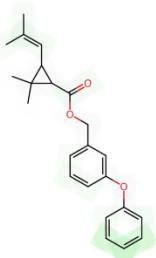
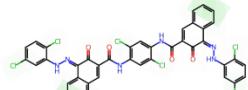
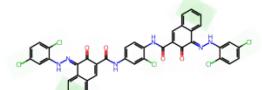
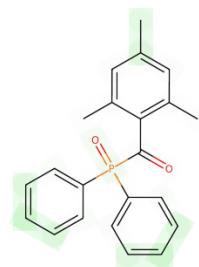
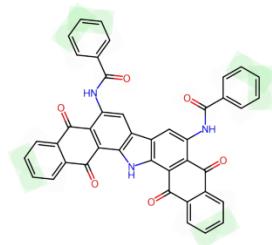
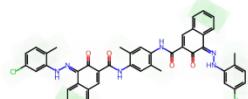
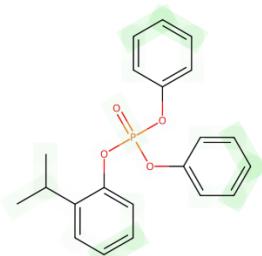
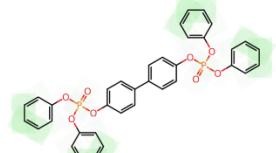
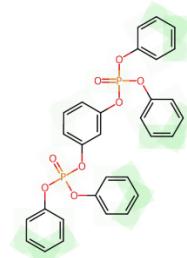
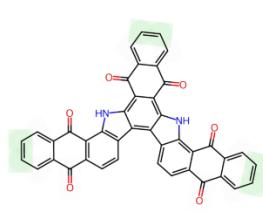
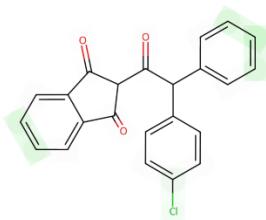
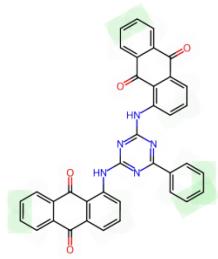
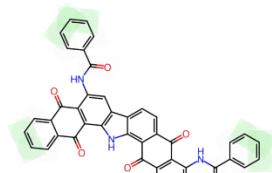
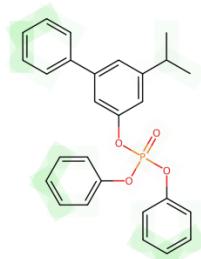
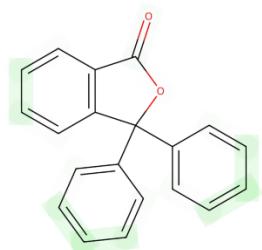
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



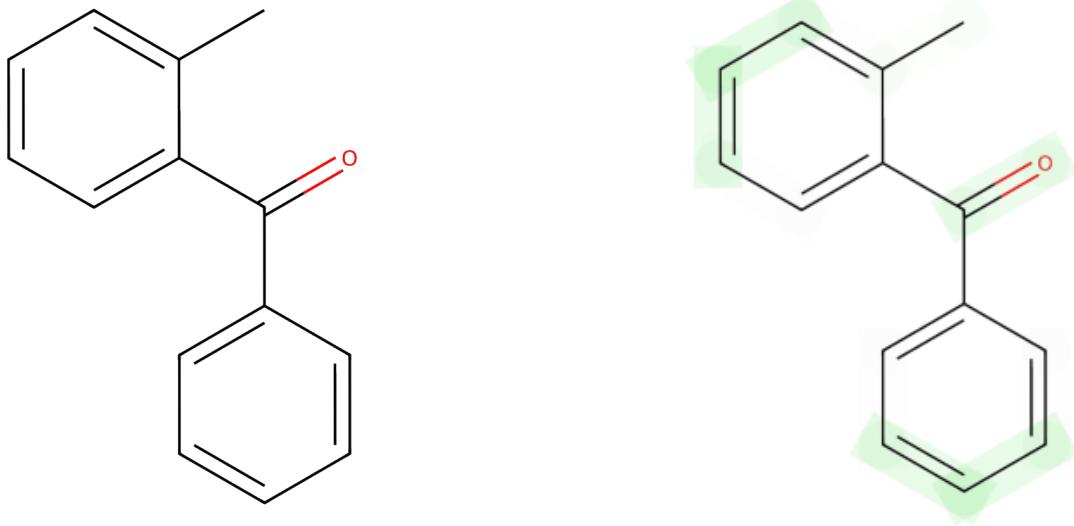
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES structure represents a molecule with two phenyl rings (aromatic hydrocarbons) attached to a central carbonyl group. The presence of aromatic rings typically decreases water solubility due to the nonpolar nature of these rings, which do not interact well with water. However, the carbonyl group is polar and can engage in hydrogen bonding with water, which would increase solubility. The influence towards water solubility suggests that in this case, the effect of the carbonyl group on the solubility outweighs the hydrophobic effect of the aromatic rings.

**Hypothesis:** Molecules with the substructure of two phenyl rings connected by a carbonyl group have an increased tendency towards water solubility. The carbonyl group's ability to engage in hydrogen bonding with water likely enhances the solubility despite the hydrophobic nature of the aromatic rings. This leads to the conclusion that the polar functional groups attached to aromatic systems can significantly influence the solubility of the molecule in water.

# Cluster #87 - positive

## Summary

ⓘ This is a summary about the most important properties of the cluster. It lists which explanation channel the cluster is from, what the average size of the explanation is as well as the average impact of this clusters members to the prediction outcome of the model.

Cluster 87, from importance channel 1 (*positive*), represents a motif consisting of 3.0 ( $\pm 0.3$ ) nodes. The concept is generally associated with an impact of 0.7 ( $\pm 0.2$ ) on the prediction outcome.

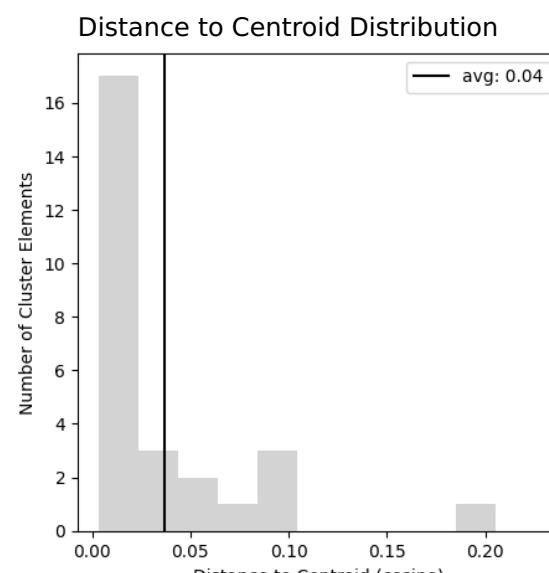
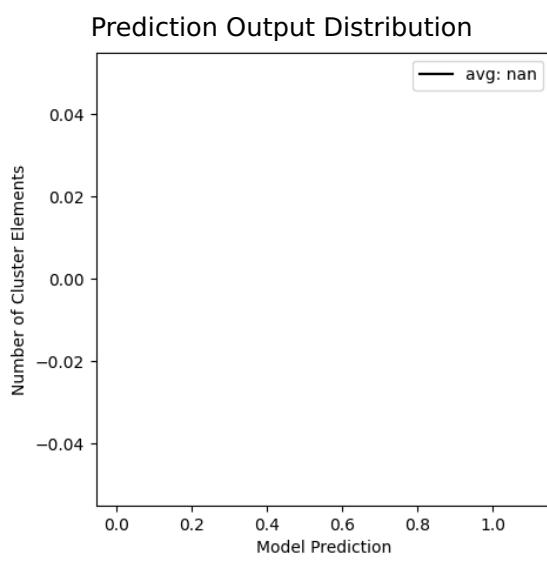
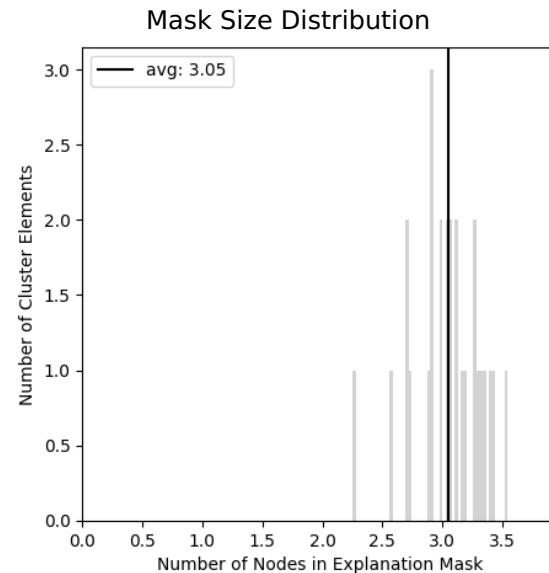
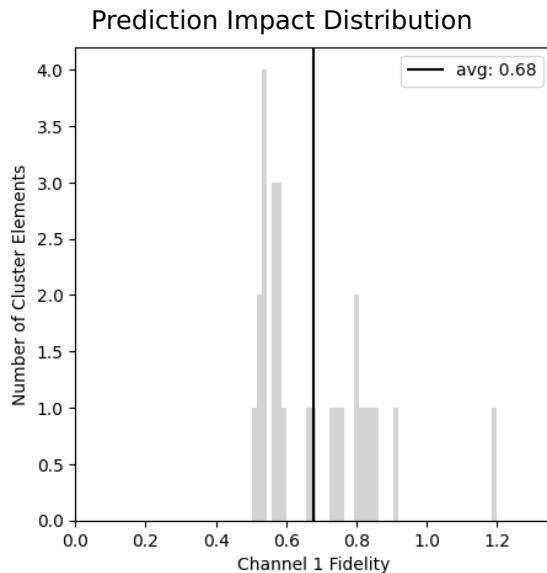
## Properties

ⓘ This section shows some numeric information about the cluster in a tabular format. The left column is a description of the property and the right column shows the value of that property.

No. Cluster Members:	27
Channel Index	1.0 (0.0)

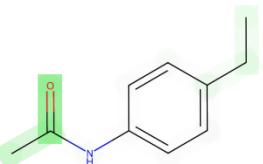
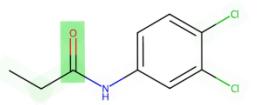
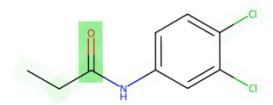
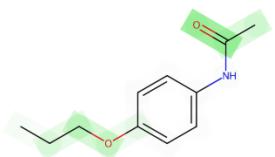
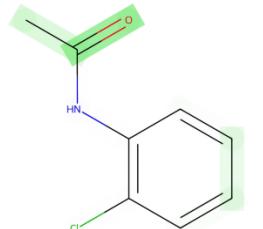
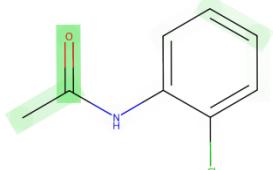
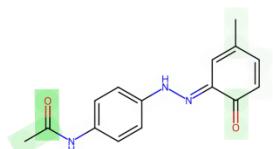
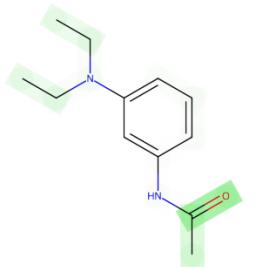
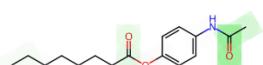
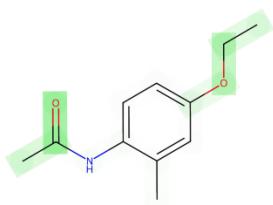
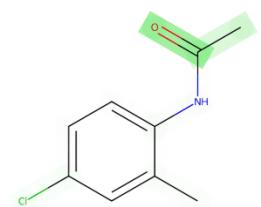
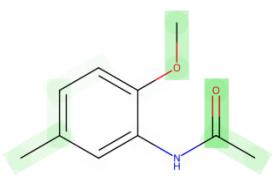
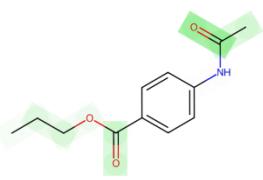
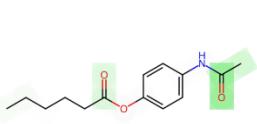
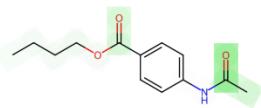
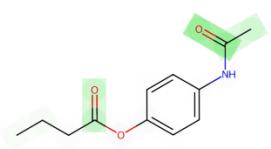
## Member Statistics

ⓘ This section shows various statistical distributions about some aspects of the various cluster members. As distributions over the cluster members, the vertical axis always shows the number of elements associated with a certain properties and the horizontal axis shows the different values that the properties can take.



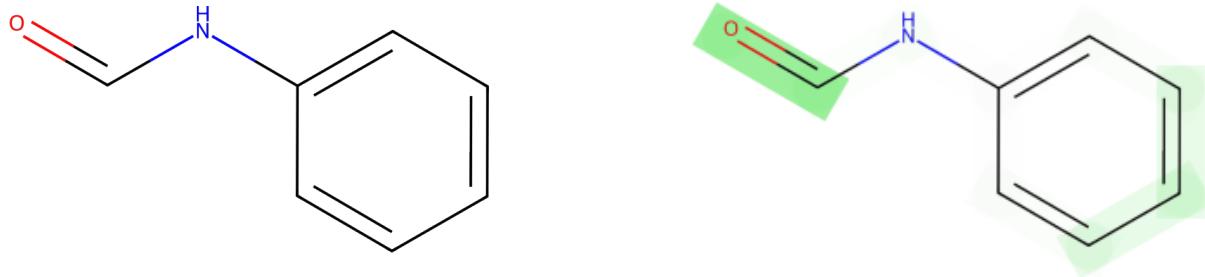
## Example Elements

ⓘ This section shows the visualization of some example members of the cluster. A cluster member is a combination of a base graph and an explanation mask, whereby the masks indicates what kind of motif the cluster represents. In the best case, the explanation masks (highlighted nodes and edges) show some sort of common pattern even though the base graphs may differ.



## Prototype

ⓘ This section shows the cluster prototype. A "prototype" in this case is a graph which is meant to be representative of the cluster as a whole. In that sense it is supposed to be the subgraph motif which is the common theme among the cluster members individual explanations. The prototype consists of a visualization and a natural language description. The prototype is automatically approximated through a non-deterministic process and thus may not be completely accurate.



## Prototype Hypothesis

ⓘ This section displays a hypothesis about a possible underlying causal reasoning behind the identified concept and its contribution to the prediction outcome. This hypothesis is automatically generated by a language model which receives a reduced representation of the concept prototype. Therefore, the hypothesis is likely not entirely accurate and should be interpreted with caution.

**Detailed Explanation:** The given SMILES representation describes a molecular structure that consists of an amide functional group ( $\text{O}=\text{C}-\text{N}$ ) attached to a benzene ring ( $\text{c}1:\text{c}:\text{c}:\text{c}:\text{c}:\text{c}:\text{c}1$ ). The amide group comprises a carbonyl ( $\text{C}=\text{O}$ ) bonded to a nitrogen atom, which can participate in hydrogen bonding. Hydrogen bonds play a crucial role in increasing water solubility, as they can interact with water molecules. However, the presence of a benzene ring, which is hydrophobic due to its nonpolar nature, can reduce water solubility. Therefore, the water solubility of this molecule may result from a balance between the polar, hydrogen bonding-capable amide group and the non-polar, hydrophobic benzene ring.

**Hypothesis:** Molecules featuring the  $\text{O}=\text{C}-\text{N}-\text{c}1:\text{c}:\text{c}:\text{c}:\text{c}:\text{c}:\text{c}1$  structure have moderate influence towards water solubility, due to the competing solubility effects of the amide group and the benzene ring. The amide is expected to increase water solubility through hydrogen bonding, while the benzene ring's hydrophobicity can reduce it.