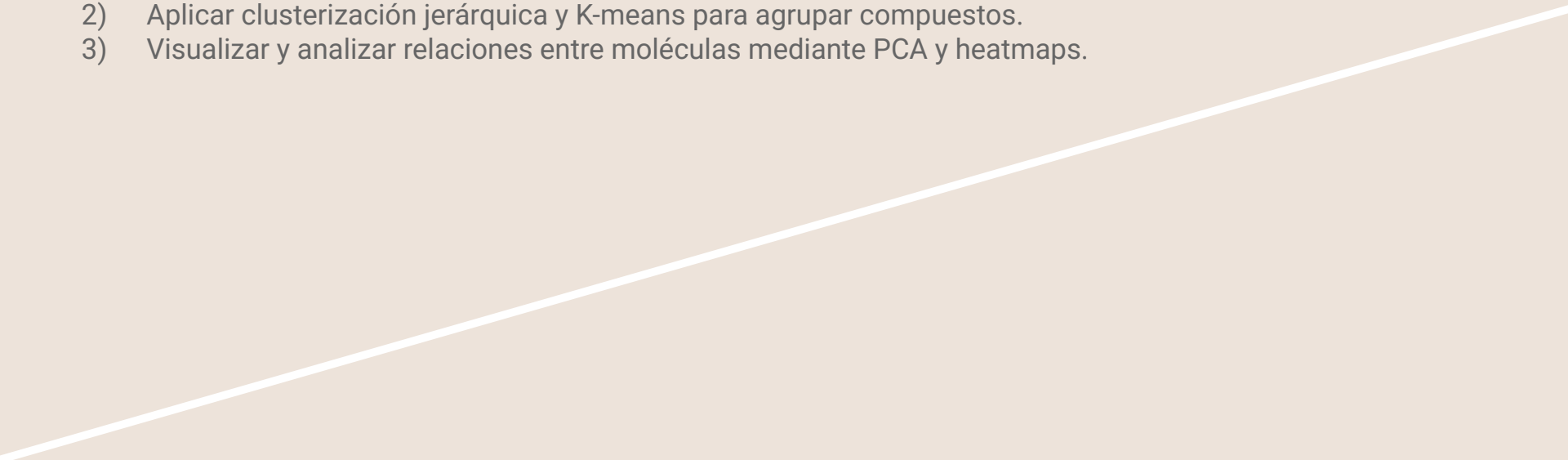


TP4

Similitud de compuestos y clusterización



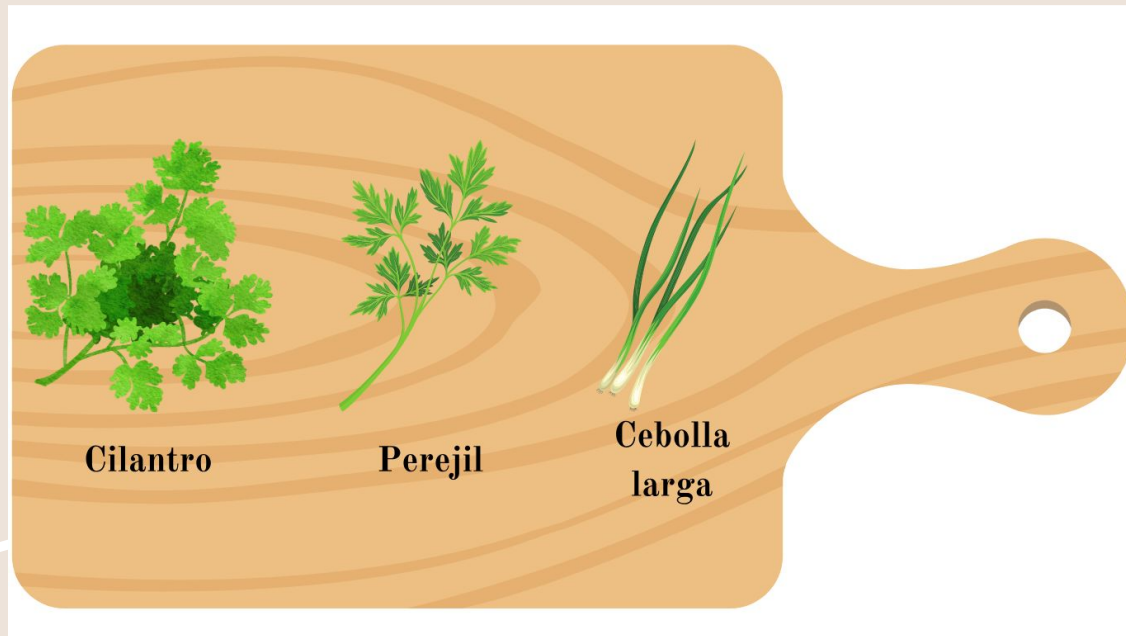
Objetivos

- 1) Aprender a calcular similitud molecular usando fingerprints y métricas de RDKit.
 - 2) Aplicar clusterización jerárquica y K-means para agrupar compuestos.
 - 3) Visualizar y analizar relaciones entre moléculas mediante PCA y heatmaps.
- 

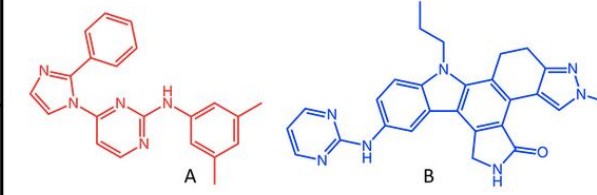
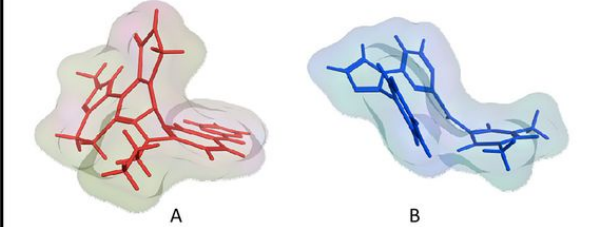
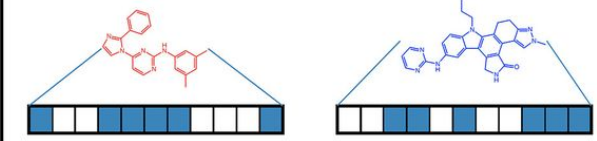
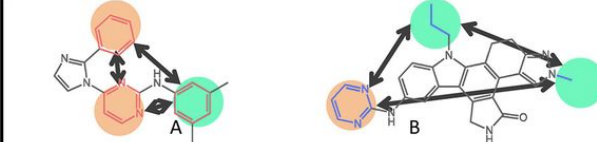
Organización de la clase

9:00 a 9:30	Introducción al TP
9:30 a 10:30	Trabajo en la guía de ejercicios (Parte 1: Clusterización por Fingerprints)
10:30 a 11:00	Recreo
11:00 a 12:00	Trabajo en la guía de ejercicios (Parte 2: Clusterización por PCA)
12:00 a 13:00	Lectura de paper y puesta en común

Similitud



Similitud

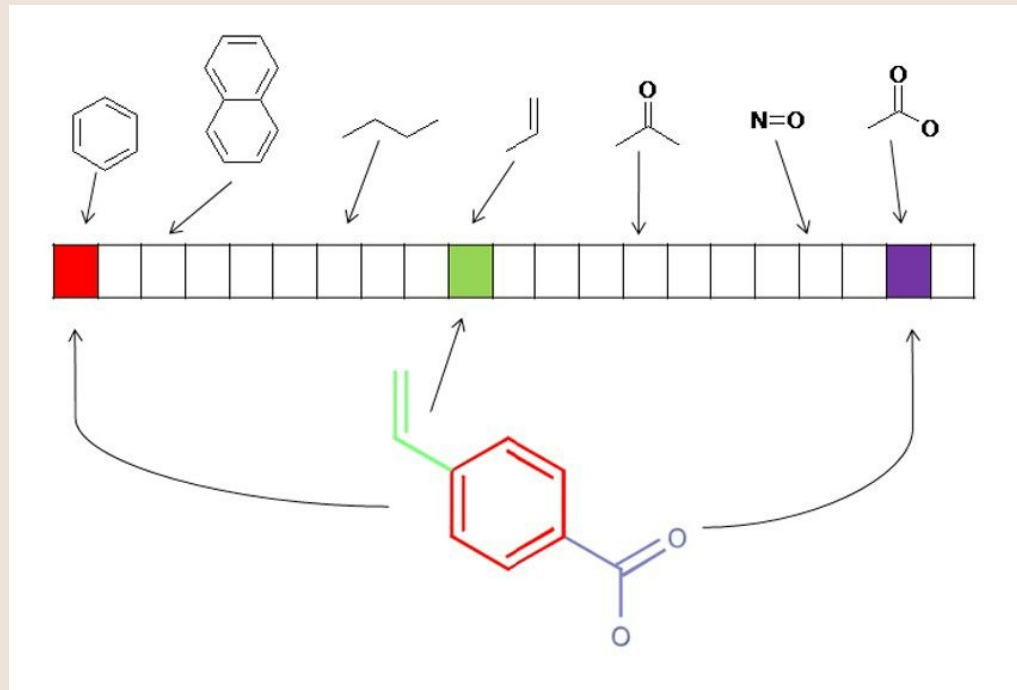
Chemical similarity	<table><tr><td></td><td>Mol. weight</td><td>LogP</td><td>Rotatable bonds</td><td>Aromatic rings</td><td>Heavy atoms</td></tr><tr><td>A</td><td>341.4</td><td>5.23</td><td>4</td><td>4</td><td>26</td></tr><tr><td>B</td><td>463.5</td><td>4.43</td><td>4</td><td>5</td><td>35</td></tr></table>							Mol. weight	LogP	Rotatable bonds	Aromatic rings	Heavy atoms	A	341.4	5.23	4	4	26	B	463.5	4.43	4	5	35
	Mol. weight	LogP	Rotatable bonds	Aromatic rings	Heavy atoms																			
A	341.4	5.23	4	4	26																			
B	463.5	4.43	4	5	35																			
Molecular similarity																								
2D similarity																								
3D similarity																								
Biological similarity	<table><tr><td></td><td>Vascular endothelial growth factor receptor 2</td><td>Tyrosine-protein kinase TIE-2</td></tr><tr><td>A</td><td>active</td><td>inactive</td></tr><tr><td>B</td><td>active</td><td>active</td></tr></table>							Vascular endothelial growth factor receptor 2	Tyrosine-protein kinase TIE-2	A	active	inactive	B	active	active									
	Vascular endothelial growth factor receptor 2	Tyrosine-protein kinase TIE-2																						
A	active	inactive																						
B	active	active																						
Global similarity																								
Local similarity																								

Parte 1

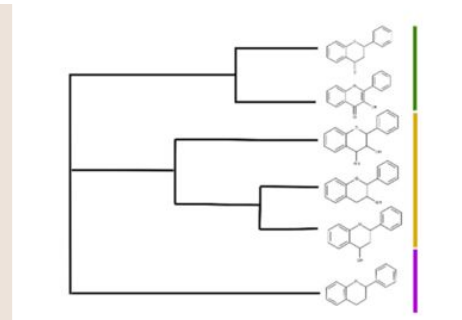
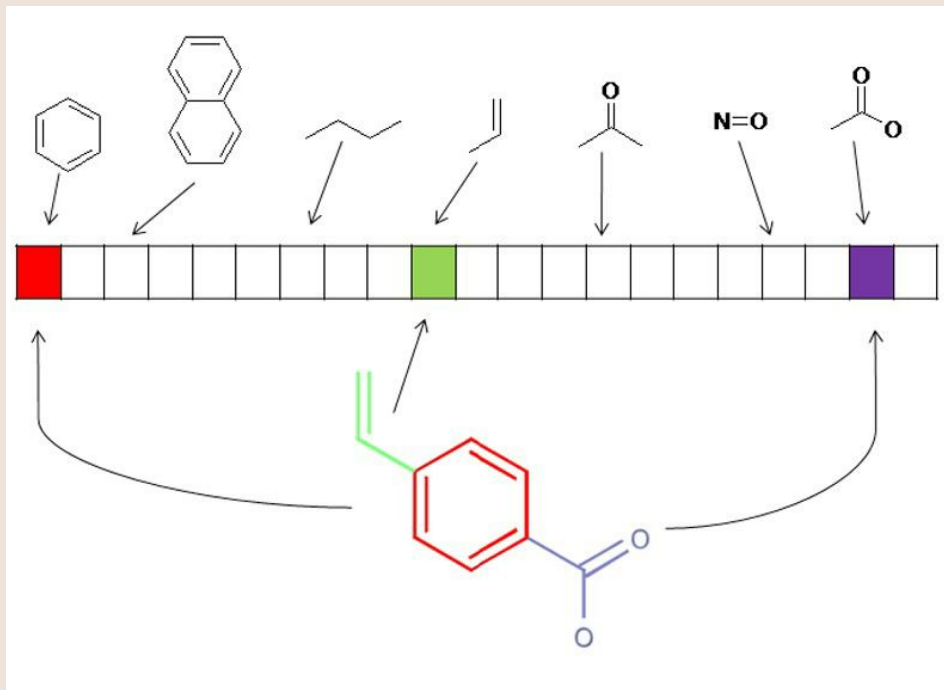
Clusterización por Fingerprints



Clusterización por Fingerprints

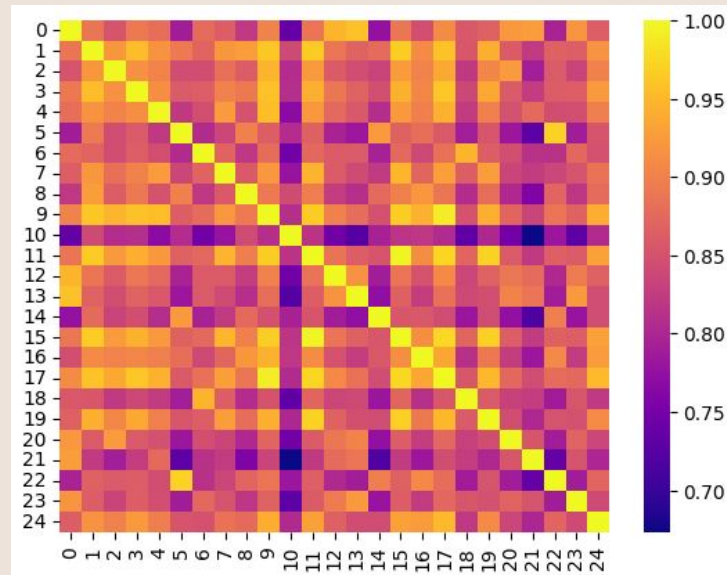


Clusterización por Fingerprints

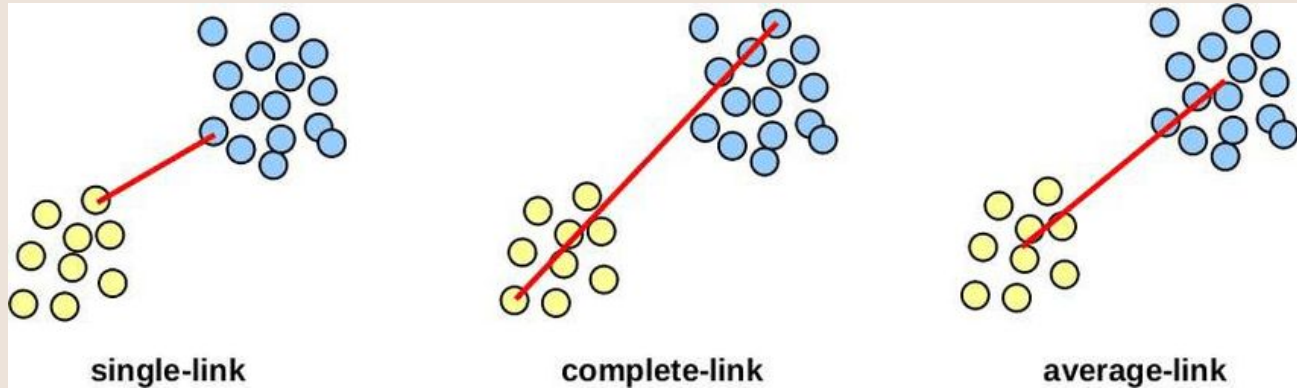


Clusterización por Fingerprints

- 1) Calcular Mol de cada molécula con RDKit
- 2) Calcular el FingerprintMol de cada molécula
- 3) Armar la matriz de similitud

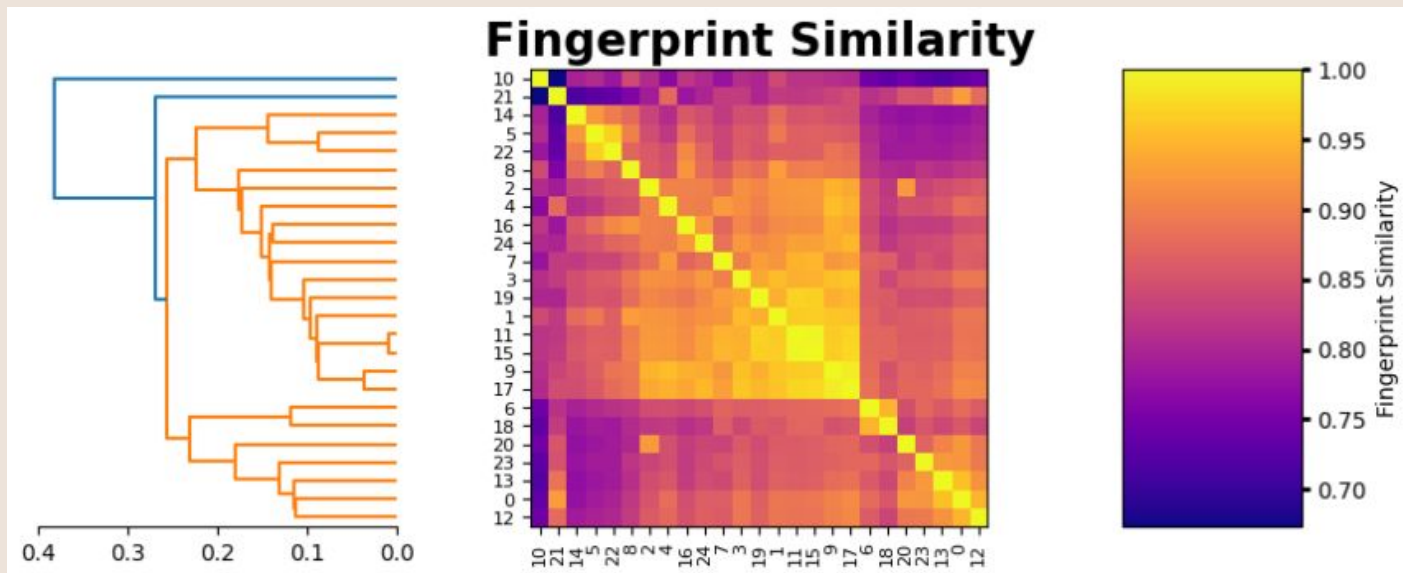


Clusterización por Fingerprints



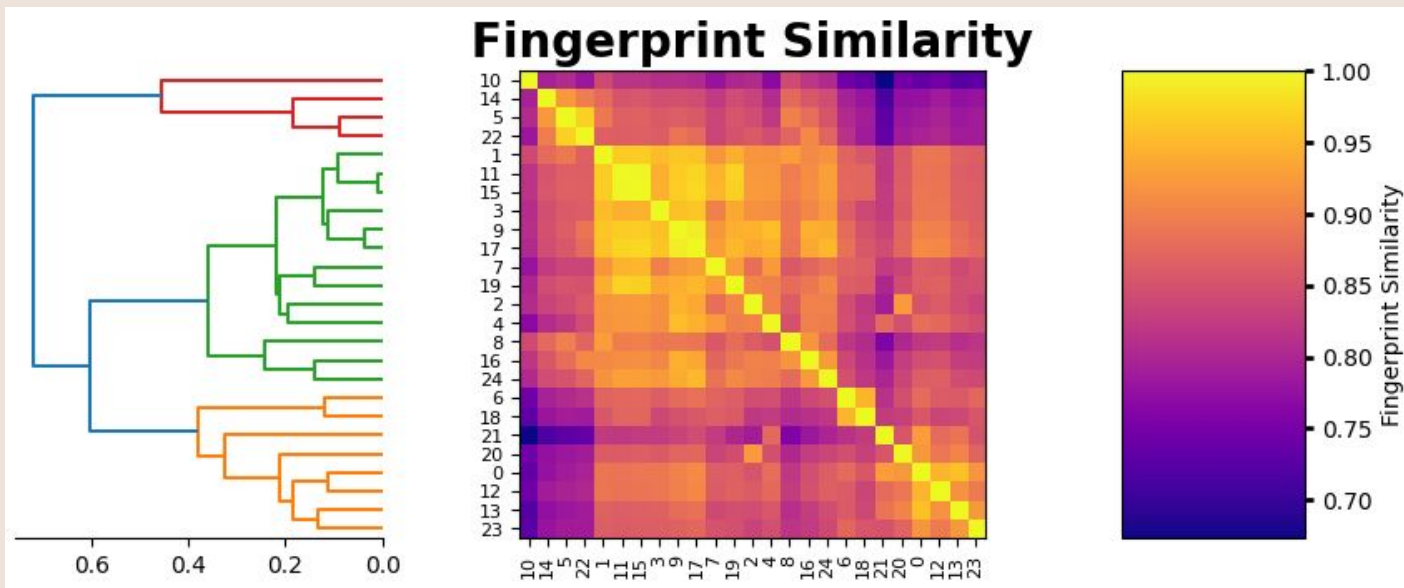
Clusterización por Fingerprints

Clusterizar con single



Clusterización por Fingerprints

Clusterizar con complete

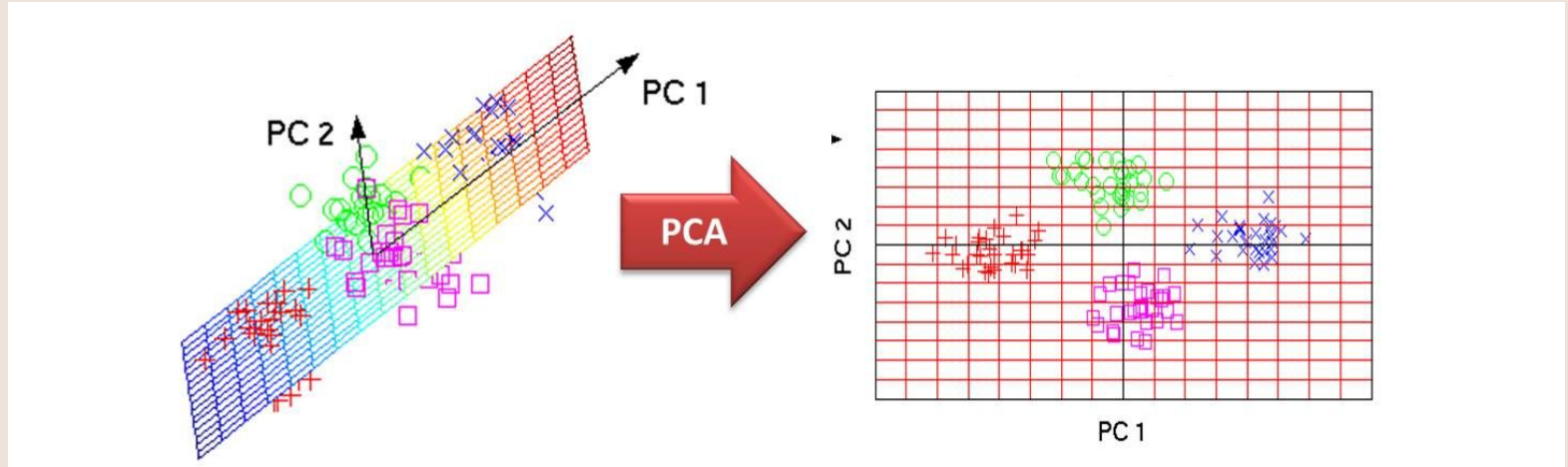


Parte 2

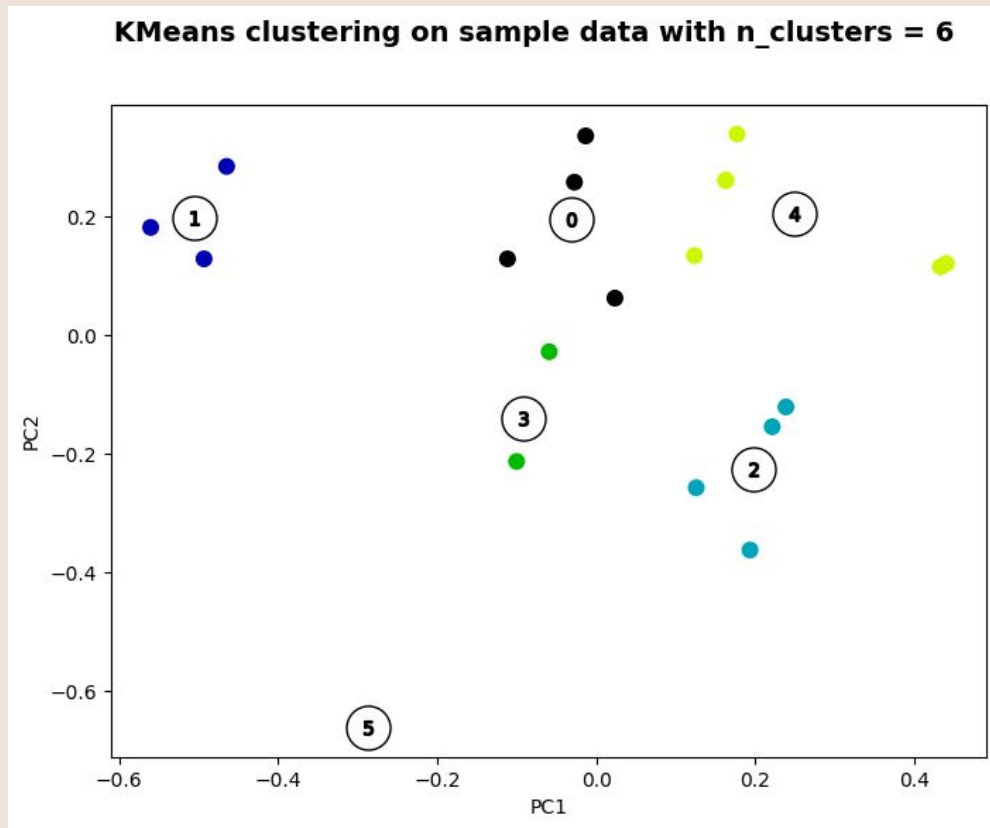
Clusterización por PCA



Clusterización por PCA



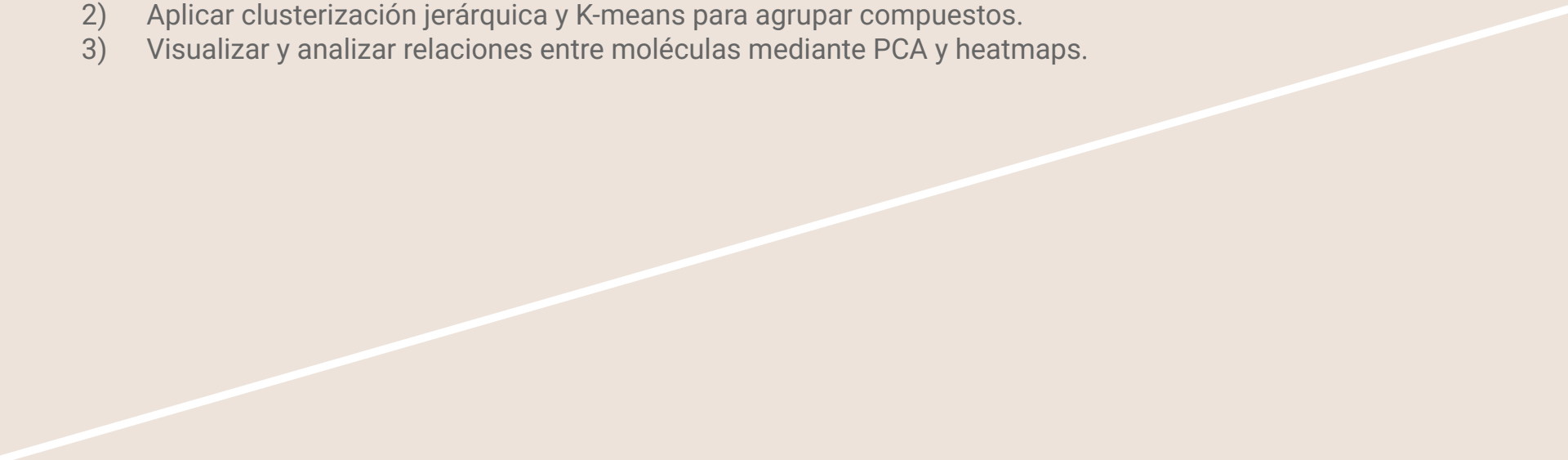
Clusterización por PCA



En resumen!

	Criterios para clusterizar
Fingerprints	Similitud estructural
PCA	Similitud de propiedades fisicoquímicas

Objetivos

- 1) Aprender a calcular similitud molecular usando fingerprints y métricas de RDKit.
 - 2) Aplicar clusterización jerárquica y K-means para agrupar compuestos.
 - 3) Visualizar y analizar relaciones entre moléculas mediante PCA y heatmaps.
- 

Cierre

Improving Measures of Chemical Structural Similarity Using Machine Learning on Chemical–Genetic Interactions

Hamid Safizadeh, Scott W. Simpkins, Justin Nelson, Sheena C. Li, Jeff S. Piotrowski, Mami Yoshimura, Yoko Yashiroda, Hiroyuki Hirano, Hiroyuki Osada, Minoru Yoshida, Charles Boone, and Chad L. Myers*



Cite This: *J. Chem. Inf. Model.* 2021, 61, 4156–4172



Read Online

<https://pubs.acs.org/doi/10.1021/acs.jcim.0c00993?ref=pdf>
<https://github.com/csbio/VS-SVM>