

IDENTIFYING TOXIC SUBREDDITS

HELPING REDDIT MODERATORS TO FIND SUBREDDITS
THAT INFLUENCE OTHER SUBREDDITS IN A NEGATIVE
WAY.

FANGZENG LYI

PABLO MARTÍN CALVO

CHARLOTTE SMOOR

NATHALIE VAN VEEN

MEREL WEMELSFELDER

INFORMATION VISUALIZATION – UvA

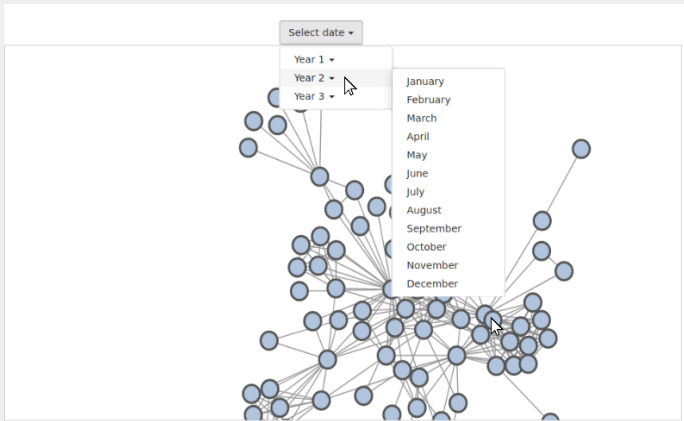
FRONT-END

We already have:

- A visual representation of Subreddit clusters and their interactions
- Drop-down menu to select a time frame to investigate
- Possibility to select a Subreddit cluster to zoom in on
- A heat map to be shown when the deepest level of the graph is reached. It shows the degree of positive/negative influence between subreddits.

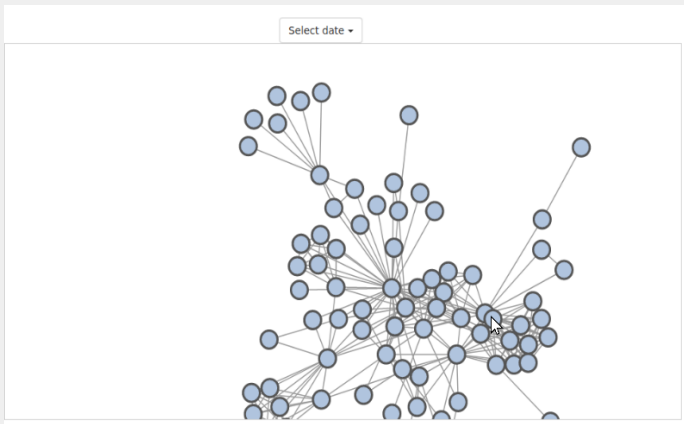
FRONT-END

An initial cluster of subreddits is shown. The dots currently represent clusters of subreddits, since we start on the highest level of the visualization. The user can select a period in time that (s)he wants to investigate.

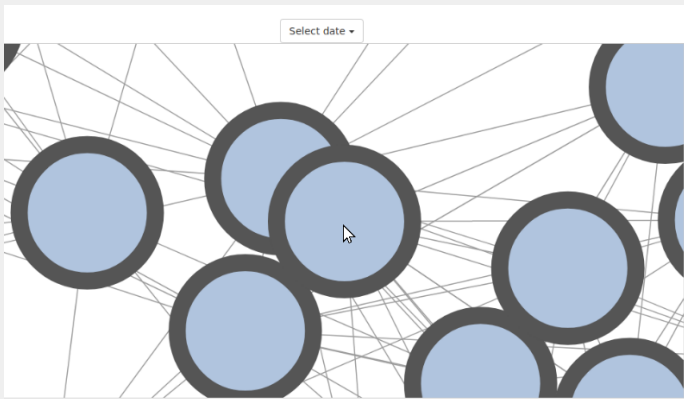


FRONT-END

The user can zoom in on a particular cluster of subreddits by clicking on the corresponding node.

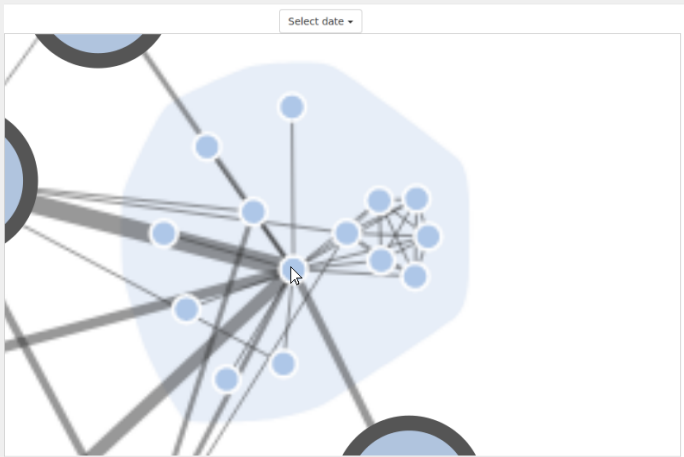


If the user clicks on a node, the graph will zoom in on this node...



FRONT-END

... and expand the node, so the user can see the clusters/subreddits that are inside.



Planned implementations:

- Arrow in the top-left corner to go back to a higher cluster level
- More information shown next to the cluster graph, so the user can deliberately make a choice on which cluster to zoom in.
- Possibly select multiple Subreddit clusters for zooming in
- Number in the top-right corner indicating the percentage of Subreddits that is currently shown.
- Maybe replace the drop-down menu for selecting a time frame by a slider, this requires user testing

BACK-END

Input:

- A date filter, containing the selected time frame.
- Selection of subreddits that should be shown.

Output:

- *success*: True if everything went well, False otherwise.
- *metadata*
- *networks*: One or more networks. Each network is composed by a list of nodes and a list of edges.
- *dendogram*: Structure of the cluster hierarchy.

Louvain Clustering Algorithm:

This algorithm produces hierarchical, non-overlapping clustering. This means that several levels of clustering are performed, with higher levels being composed by the nodes of previous levels. Typically, the algorithm produces 2 to 3 levels.

The number of produced levels can be varying, and can not manually be set. The number of resulting levels after clustering can be found in the metadata. Level 0 is always there, being the raw network without having applied the clustering.

KEY CHALLENGES

KEY CHALLENGES

The main required action to go 'live' is to connect the back-end to the front-end, since the front-end has been using dummy data so far. Other challenges include the implementation of earlier mentioned additional features, such as showing more detailed information next to the graph so the user can make a deliberate choice which cluster to investigate.