SADView - visualization of Twitter data

During the course of the project, we implemented and developed our network explorator and visualization named *SADView.* This tool is a Javascript web application, running in the user's browser, that displays collected Twitter data.
This tool can be accessed at the following location: https://imi-sad.pages.switch.ch/sadview/

This document briefly presents the functionalities of the SADView tool.

Principle

SADView displays data collected during the automatic exploration. In these files are recorded
- the graph created by collecting users and the interactions (mentions, retweets) between them
- the information about each user (name, short description, hashtag used, community he/she belongs to)
- the information about the communities (specific vocabulary used in the exchanges within the community)

For the live demo, these data files were collected separately and made available for the purpose of the visualization demo. SADView does not interact with a social network API directly. Upon prolongation of the project, it is planned to integrate all collection steps within a single web application.
In the current version, the data collection and processing is done in three automated steps:
- raw Twitter data acquisition and graph creation. For each account, the most recent tweets are considered. Each user mentions and retweets are used to connect users. The more mentions/retweets exist between two users, the stronger the edge in the graph is. After collection, weak edges (with weight lower than a given threshold) are discarded in order to reduce the "noise".
- Community detection (Louvain method) and lexical analysis are performed. For each community, a "corpus" of tweets is created and fed to the lexical analysis software *Iramuteq*
- Data is processed to a suitable format (JSON) for display. The graph layout is computed "offline" to allow a faster display and navigation. Final results are uploaded to online (S3) storage and fetched by SADView for display upon request.

Description of the options and parameters

Once data is loaded, the initial screen is displayed

Fig. 1: Main screen display. Every node in the graph corresponds to a user. Nodes of the same color belong to the same community.

The **controls** are located on the top row, and the **graph** generated from data collection is displayed below. Graph display allows zooming and panning. Controls allow to:

- **select data** collected from another day or using another set of seed nodes (in the public version, one month of daily collects are available)
- switch the graph layout between **ForceAtlas** (default) and **CirclePack** (cf. Fig. 2)
- filter nodes by **hop number**, to illustrate the propagation of data collection, nodes acquired at a later hop are then shown in transparency (cf. Fig. 3)
- search for a specific **username** or **hashtag** and display all nodes matching the search (cf. Fig. 4)
- switch the **node size** according to its full degree, in-degree or out-degree

Hovering on a node will highlight its connections. Clicking on a node will display the profile biography and the hashtags he used recently, as well as information about the community it belongs to (statistics and lexicon specific to the community).
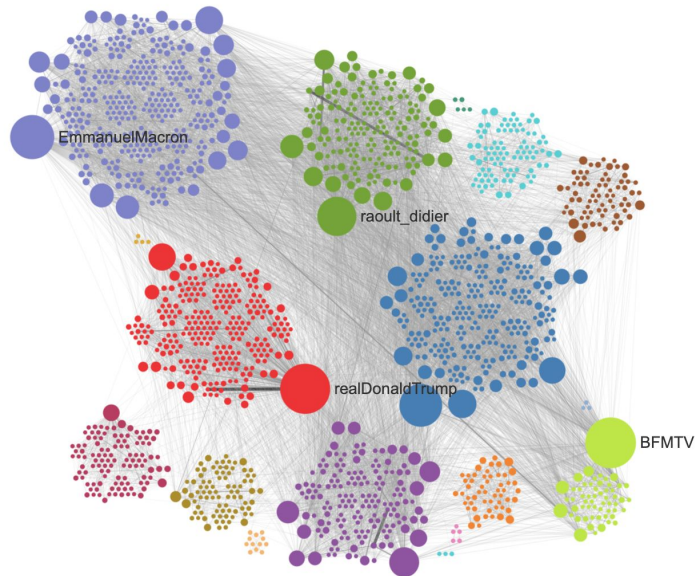
Fig. 2: Example of CirclePack layout. All nodes belonging to the same communities are kept together. Node size depends on the number of connections with other nodes in the graph (bigger = more connections).
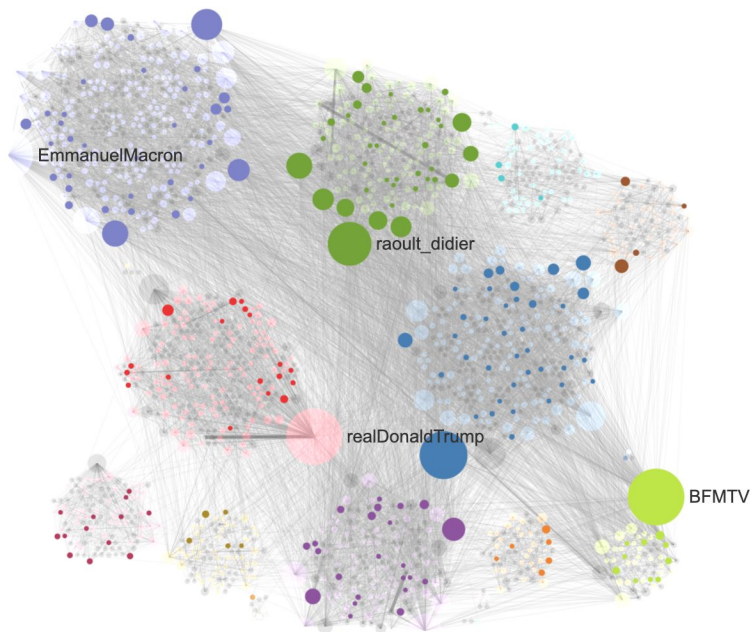
Hop number

Hashtag:

Search hashtag

Fig. 3: Display (in solid color) nodes collected at hops 0 and 1. Other nodes are shown in transparency.
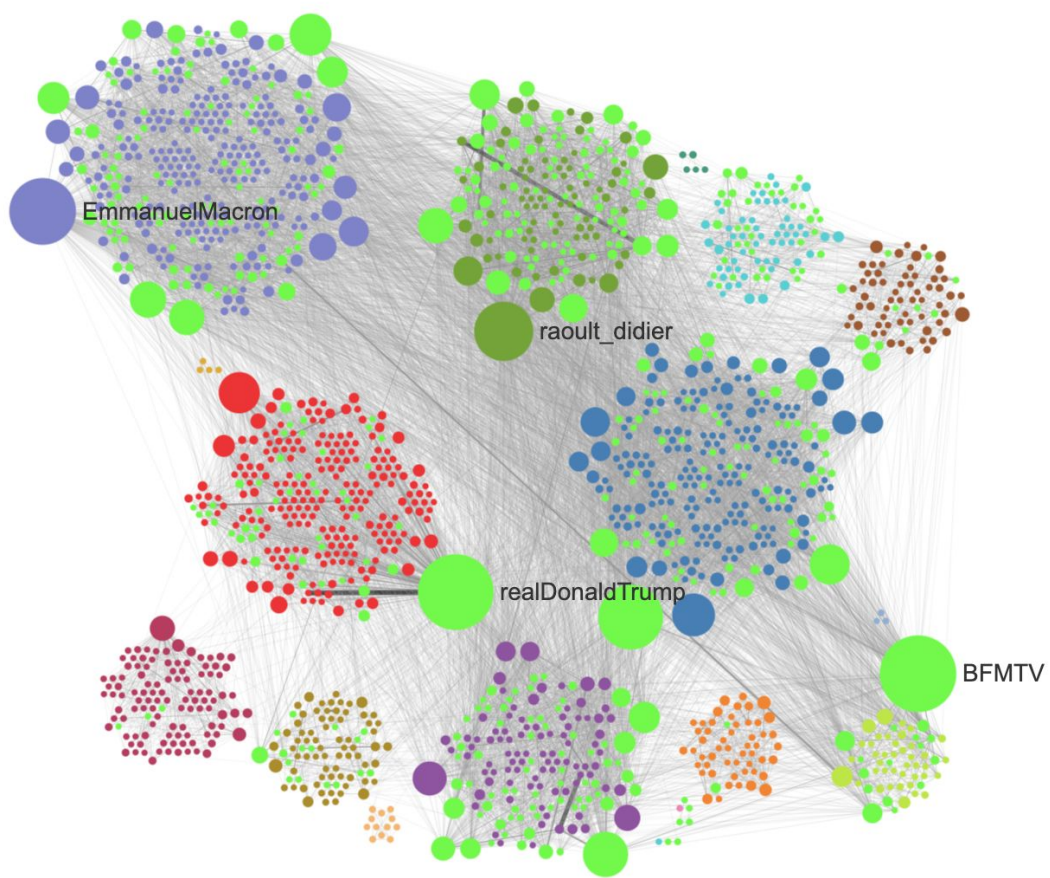
Hop number

Hashtag:

COVID19

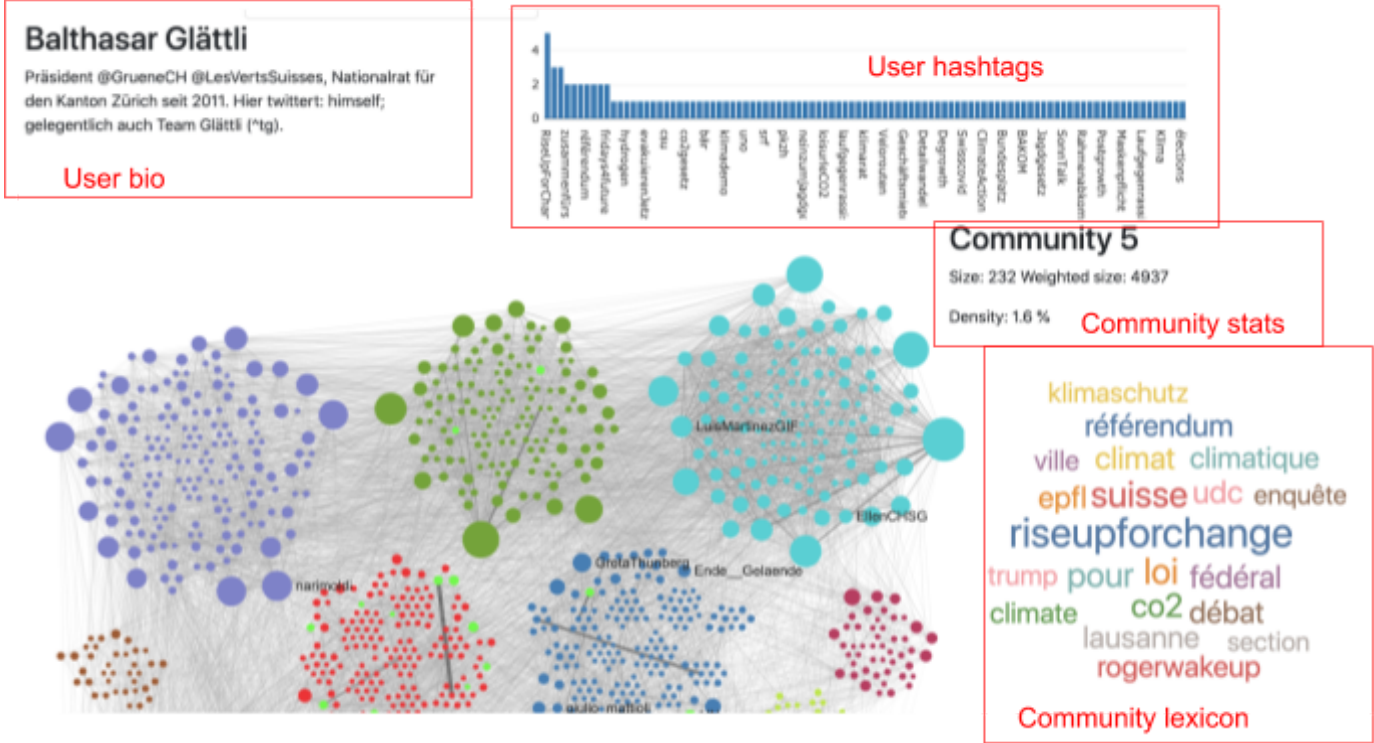Fig. 4: All users that used the hashtag #COVID19 are highlighted in green

Fig. 5: User and community information. Selected user is the president of the Swiss Green party. Lexicon (right-hand side) is related to climate issues debated in the Swiss parliament at the time of collection (Oct. 8 2020).