DESCRIPTIVE SOCIAL NETWORK ANALYSIS: THE INFORMATION NETWORK OF AUSTRALIAN TAFE EDUCATION AMONG TWITTER USERS

Cyrus Kwan

45200165 3/12/2020

METHOD:

RESEARCH PROBLEM:

The purpose of this report is to identify the driving Twitter users in terms of degrees, modularity, betweenness, and HITS metrics concerning the social network of Australian TAFE Education stakeholders.

PLANNING & DESIGN:

1. Find relevant social networking platform to conduct study
2. Collect data using open-source tools
3. Clean data using open-source tools
4. Perform descriptive data analysis
5. Report observations and insights based on descriptive analysis

COLLECTION:

The data for this study was collected using the open-source tool Netlytic. The tool uses the Twitter API to scrape a search for the most recent 2,500 tweets containing the query.

CLEANING:

Data cleaning was performed through Gephi. For the sake of simplicity, all nodes that were not connected to the main component were removed due to the presence of uninvolved cliques that are not related to TAFE Australia, i.e. Texas Association for Future Education. The resulting set of nodes were reduced from 1365 raw entries to 475 entries.

ANALYZING AND VISUALIZING DATA:

NETLYTIC ANALYSIS:

TEXT ANALYSIS:

It can be inferred from the text analysis of most frequently used words (see figure 1c) that the recent $3.2 million bid by Racing NSW to purchase the Hunter Valley TAFE centre in Scone (Patty, 2021) are highly relevant to the current landscape of Twitter users. Additionally, recent politics can be attributed to the rising involvement of Twitter users regarding the preservation of TAFE (see figure 1a).

NETWORK ANALYSIS:

Name network analysis was performed on the dataset to identify social hubs, communities, and I/O sources of information of which could not be achieved via single thread discussions in chain network analysis. The produced graph (see figure 1.d) shows a main component surrounded by multiple smaller cliques (to be removed as explained in method). The users @tafedefender and @tafeeducation are at the epicentre of discussion.

GEPHI ANALYSIS:

The nature of the dataset suggests that it is a directed network and was thus imported accordingly. Force Atlas was the layout of choice as it was the most appropriate option for viewing a large dataset (Cherven, 2015). Despite this, some basic information can still be gained from Fruchterman Reingold and Yiffan Hu layouts; the

clustering of nodes, contrast in relationships between cliques, and a general understanding of the most significant users in the dataset (see figures 2a & 2b).

Note: For all Gephi figures, measures dictating node appearance is in order of colour to size. I.e. figure 2a appears as [colour=modularity] to [size=degrees]

MEASURES:

DEGREES:

DEGREES

Degrees are necessary in portraying a general understanding of involvement within the network.

MODULARITY:

The partitions among Twitter users displays cliques of which further research can provide insights into the political landscape.

BETWEENNESS:

Required to determine most significant bridging nodes in the dataset that share relevant news and opinions.

HITS METRIC:

HUB:

Hubs are a necessary metric in identifying where in which discussion takes place in the network. Hubs are a central part of where many related authorities are linked (Kleinberg, 1999).

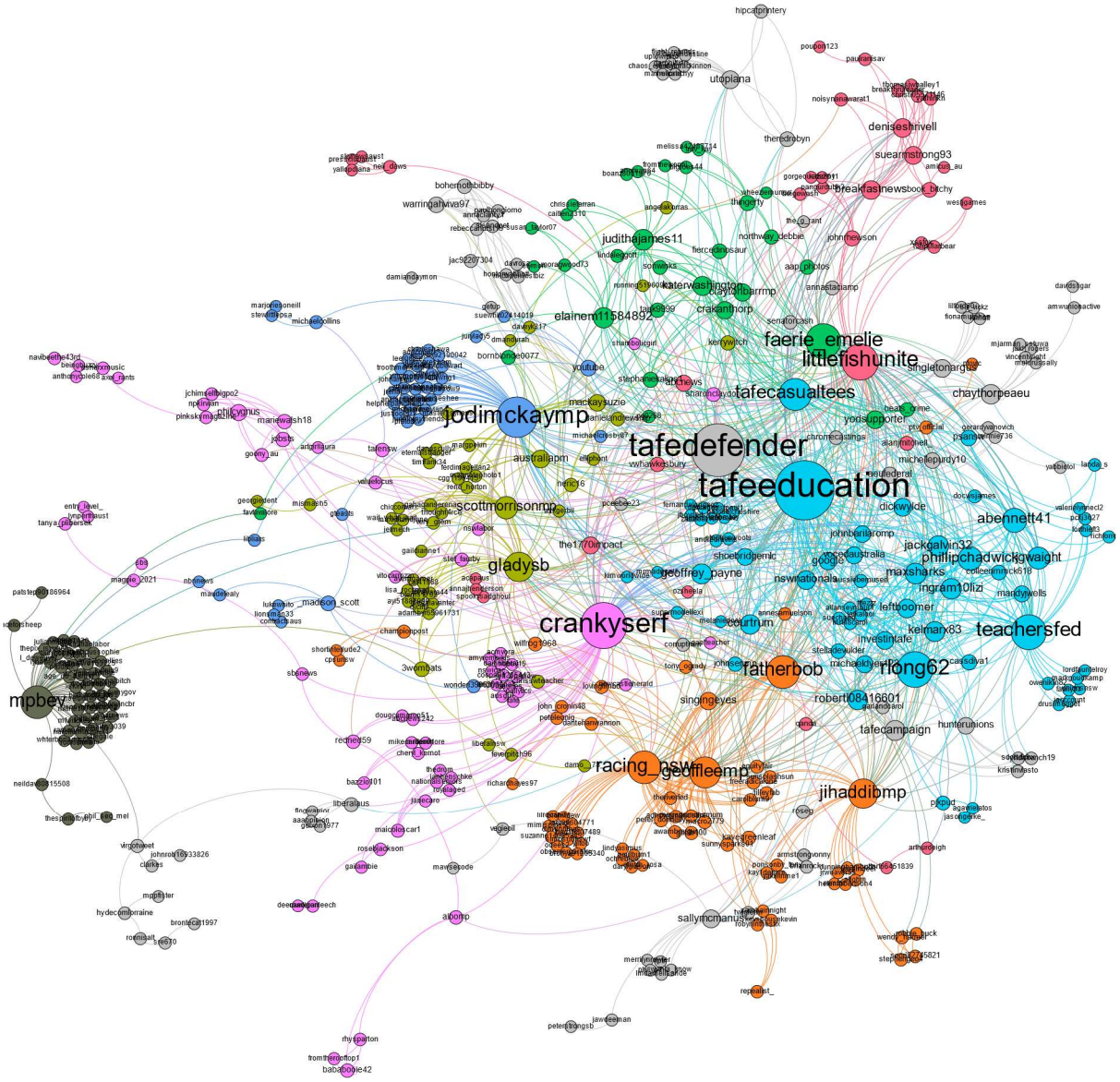
AUTHORITY:

Authority relates to the user's influence on the network and is largely attributed to their involvement with network hubs (Kleinberg, 1999). Due to the political nature of the dataset, authority is an important measure of the network.

CONCLUSION:

FIGURE 2C:

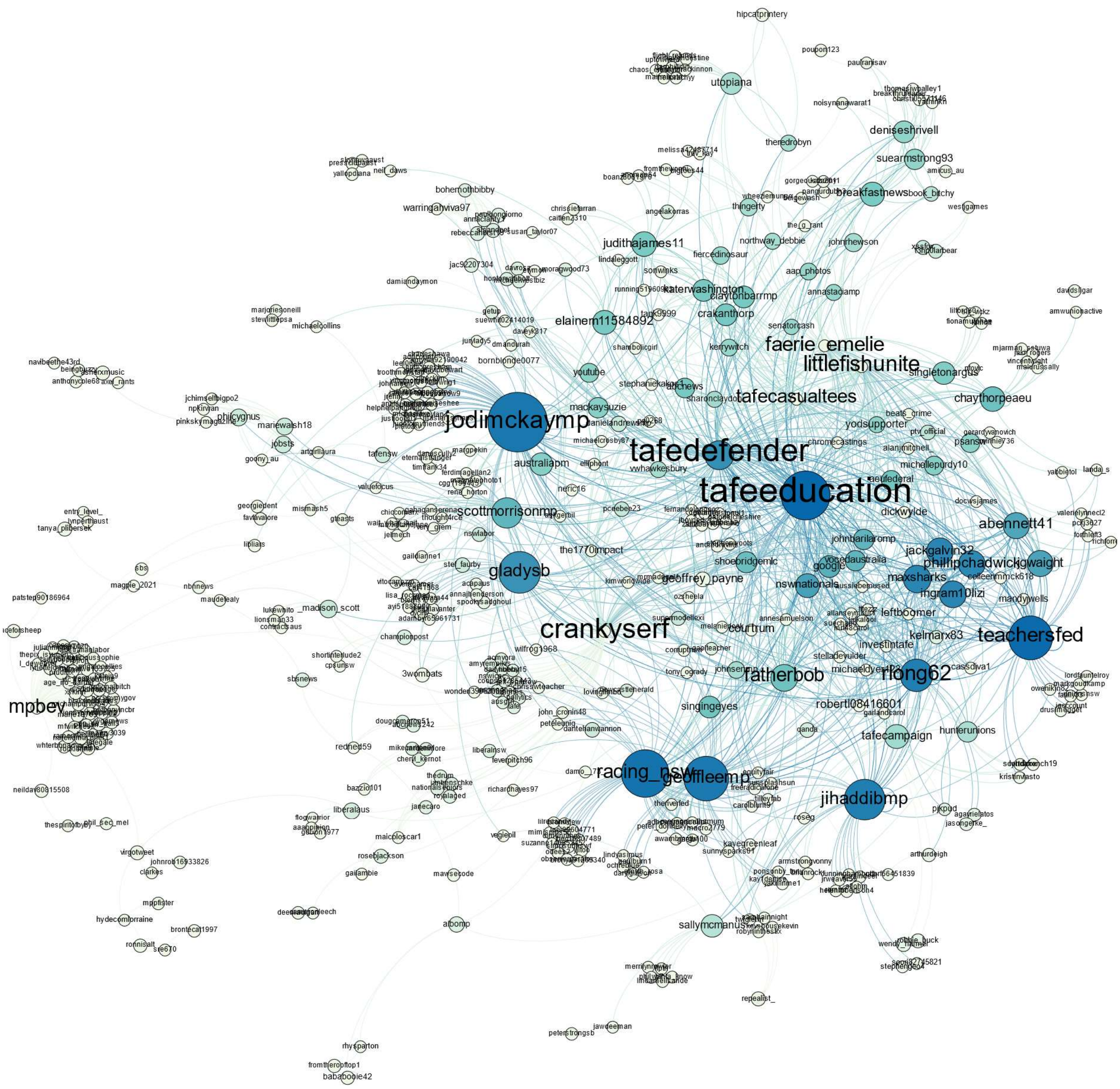
Modularity partition produced due to current events such as the involvement of @racing_nsw and @geoffleemp in their recent bid (Patty, 2021). Degrees shows total I/O by users.



Modularity to degrees

FIGURE 2D:

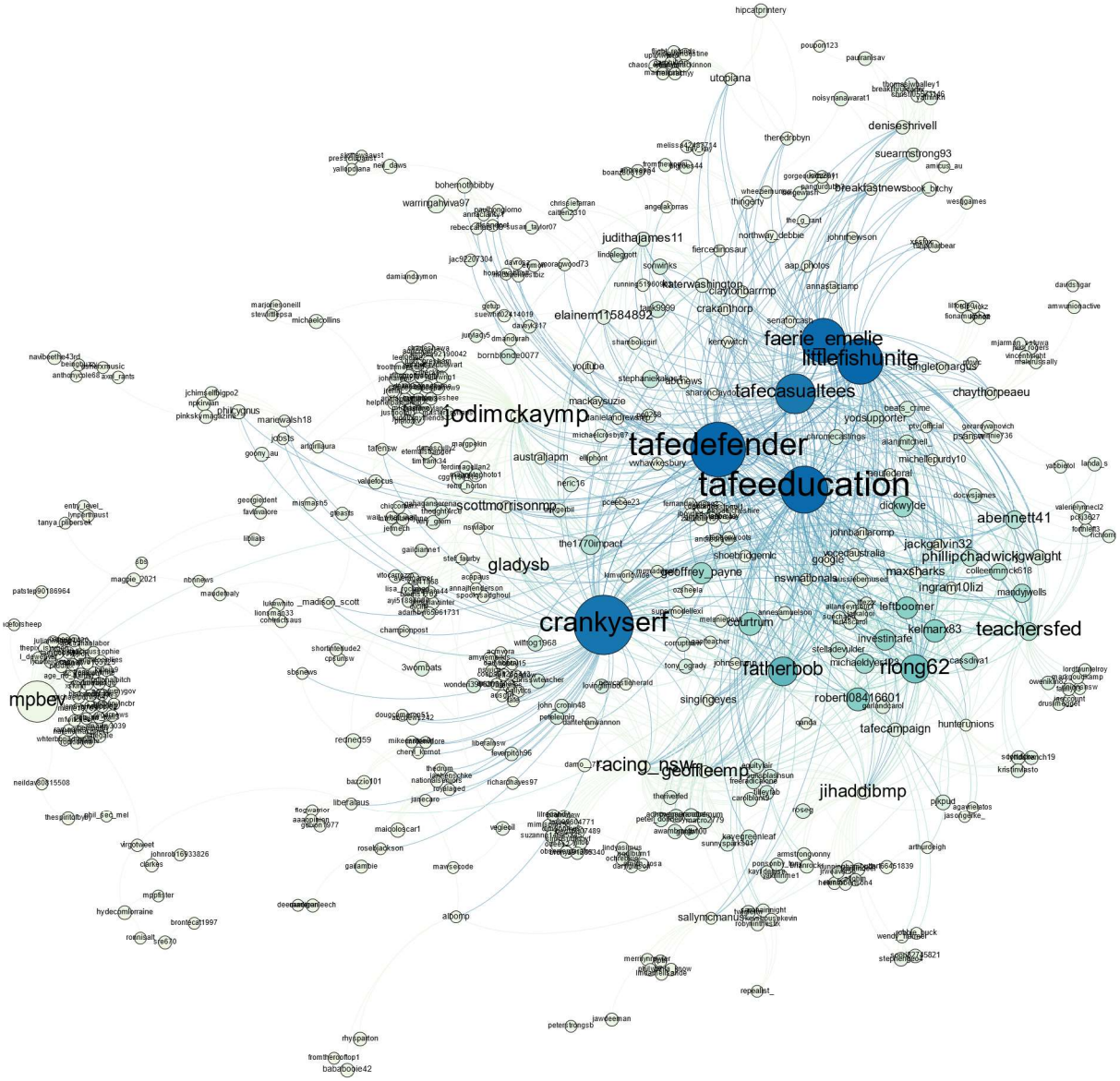
Although they are not at the forefront of TAFE discussion in terms of betweenness, user @jodimckaymp leads with the most in-degrees in the dataset. Thus, it can be determined that they are a significant figure in the network and are largely relevant.



Authority to in-degrees

FIGURE 2E:

Hubs are mostly comprised of information sharing accounts.



Hubs to out-degrees

CONCLUSION:

Users such as @tafeeducation show involvement in both consuming (in-degrees) and producing (out-degrees) user generated content (UGC). Some users' total degrees are significant purely from producing, sharing, or retweeting UGC. Users like @crankyserf output a lot of UGC however, they a mostly insignificant in terms of in-degrees. This can be interpreted as though they mostly retweet content rather that producing their own and ultimately do not have discussion directed toward them. Additionally, it can be interpreted that a positive correlation exists between authority and in-degrees, inversely a positive correlation also exists between hubs and out-degrees (compare figures 2d & 2e).

APPENDIX:

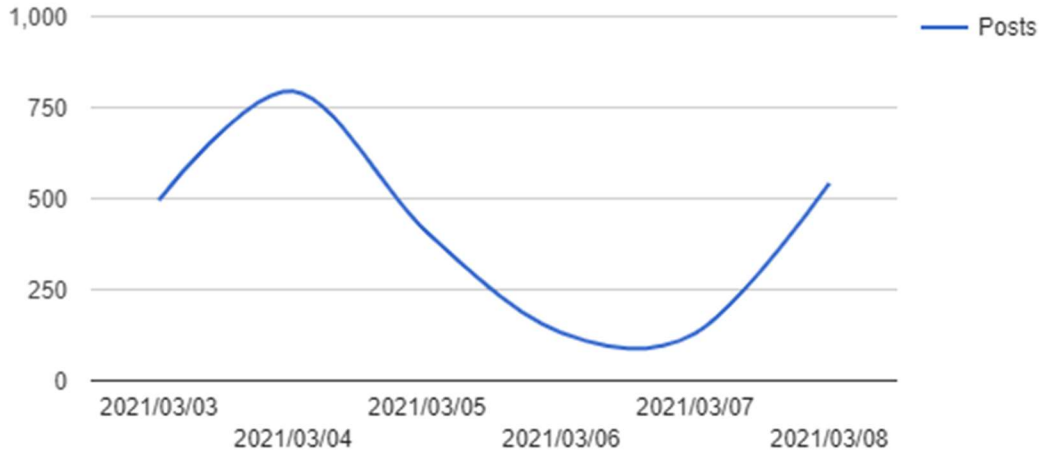


Figure 1a: Netlytic posts over time

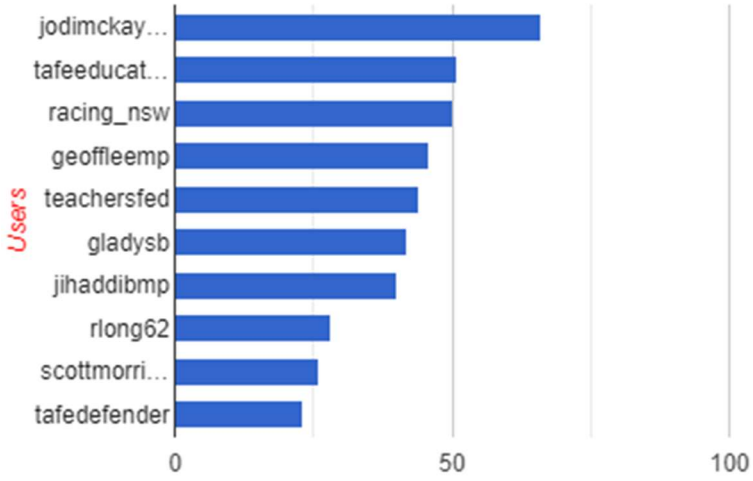


Figure 1b: Netlytic top 10 posters

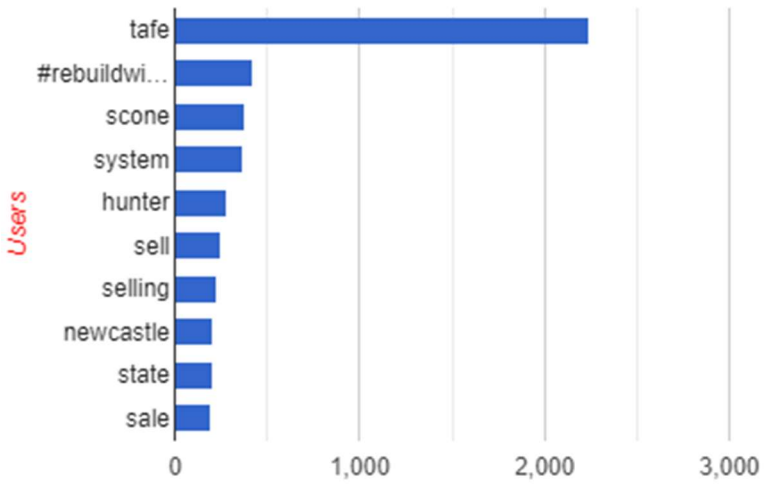


Figure 1c: Netlytic most frequently used words

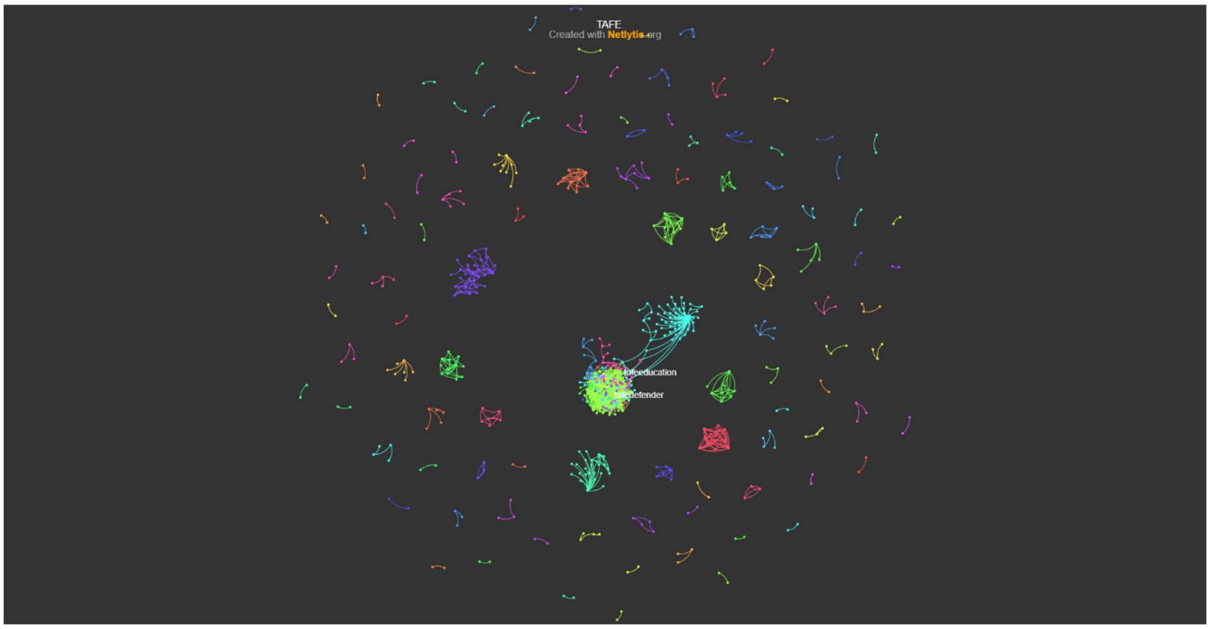


Figure 1d: Netlytic name network analysis

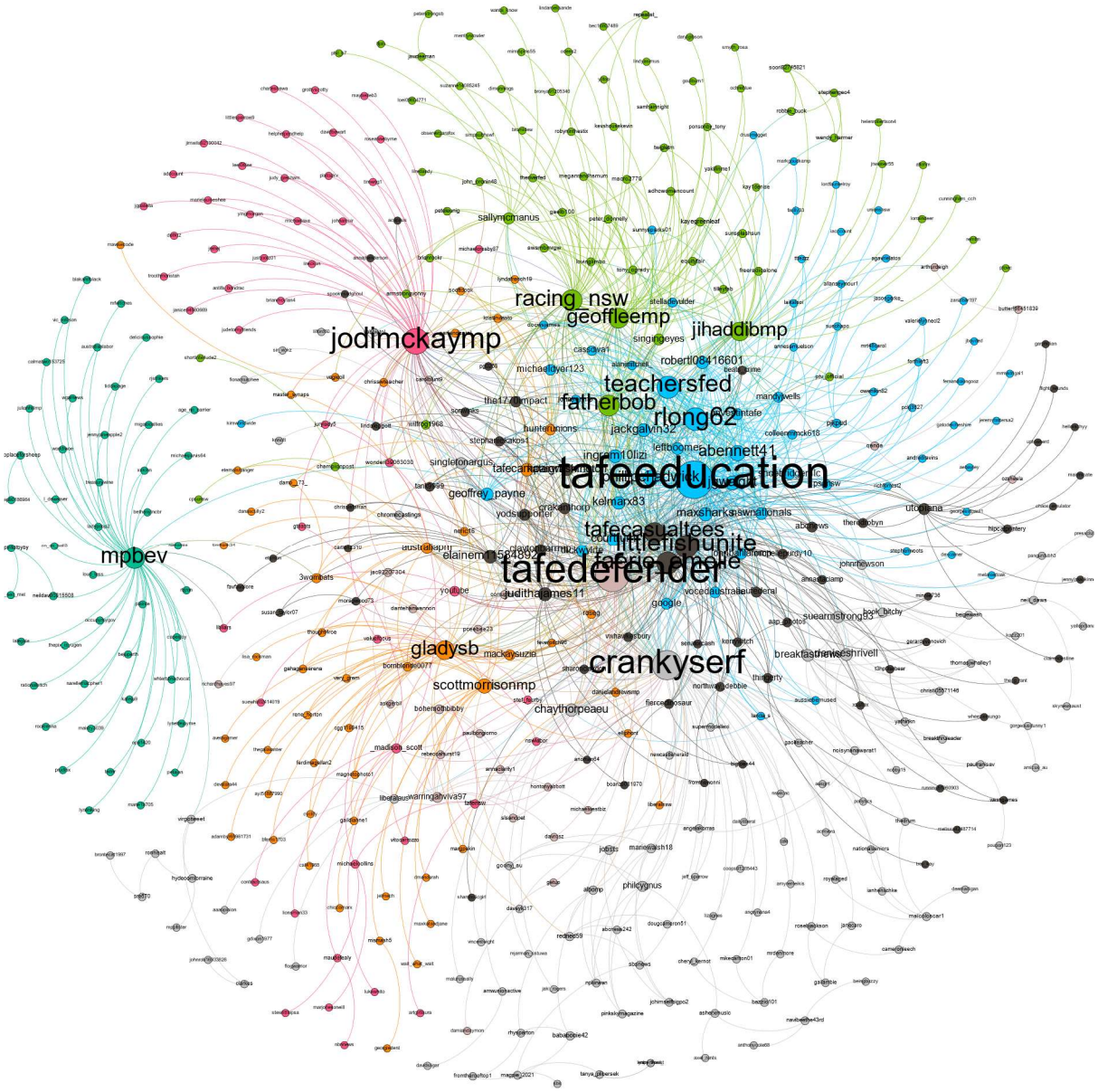


Figure 2a: Fruchterman Reingold modularity to degrees

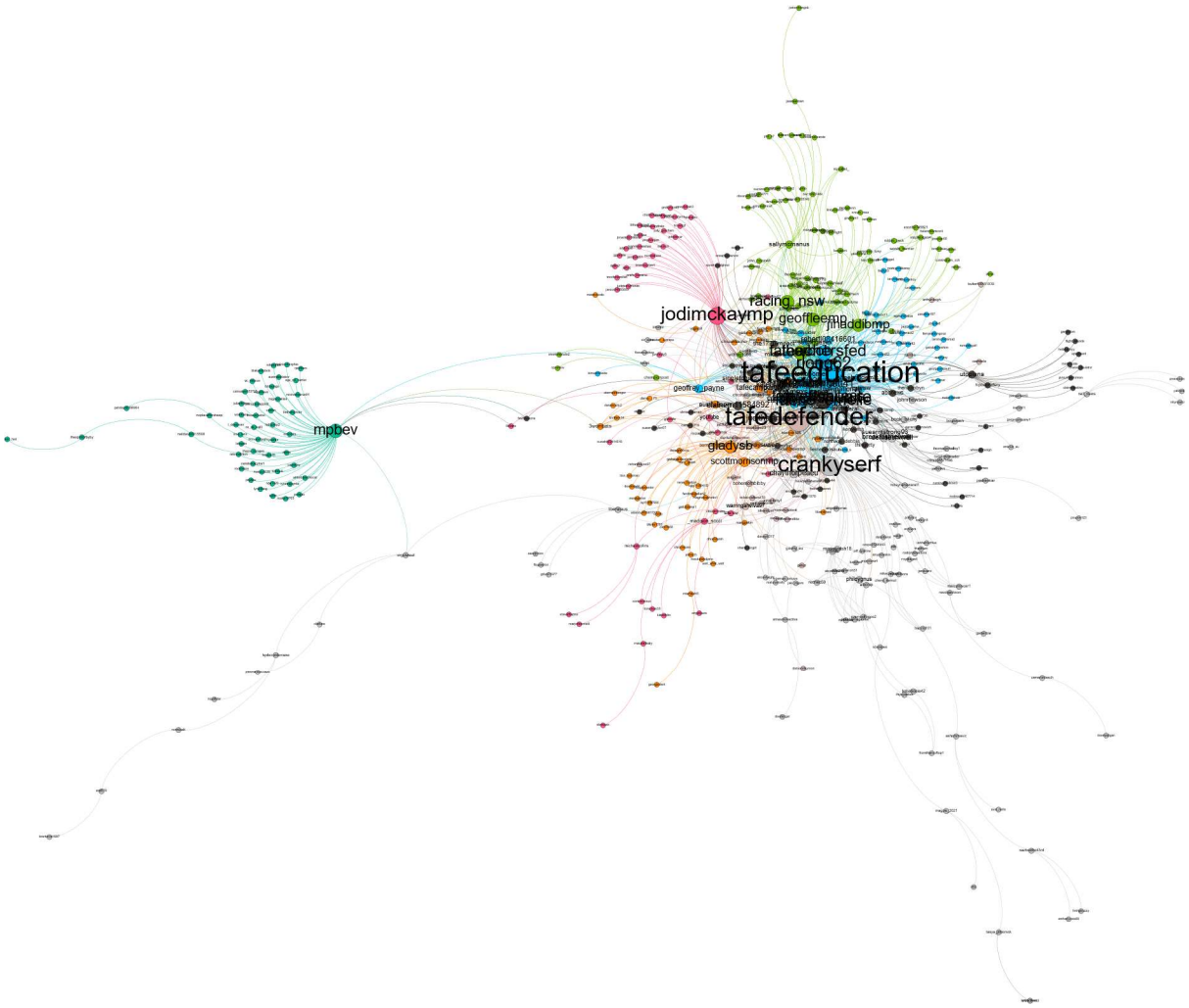


Figure 2b: Yiffan Hu modularity to degrees

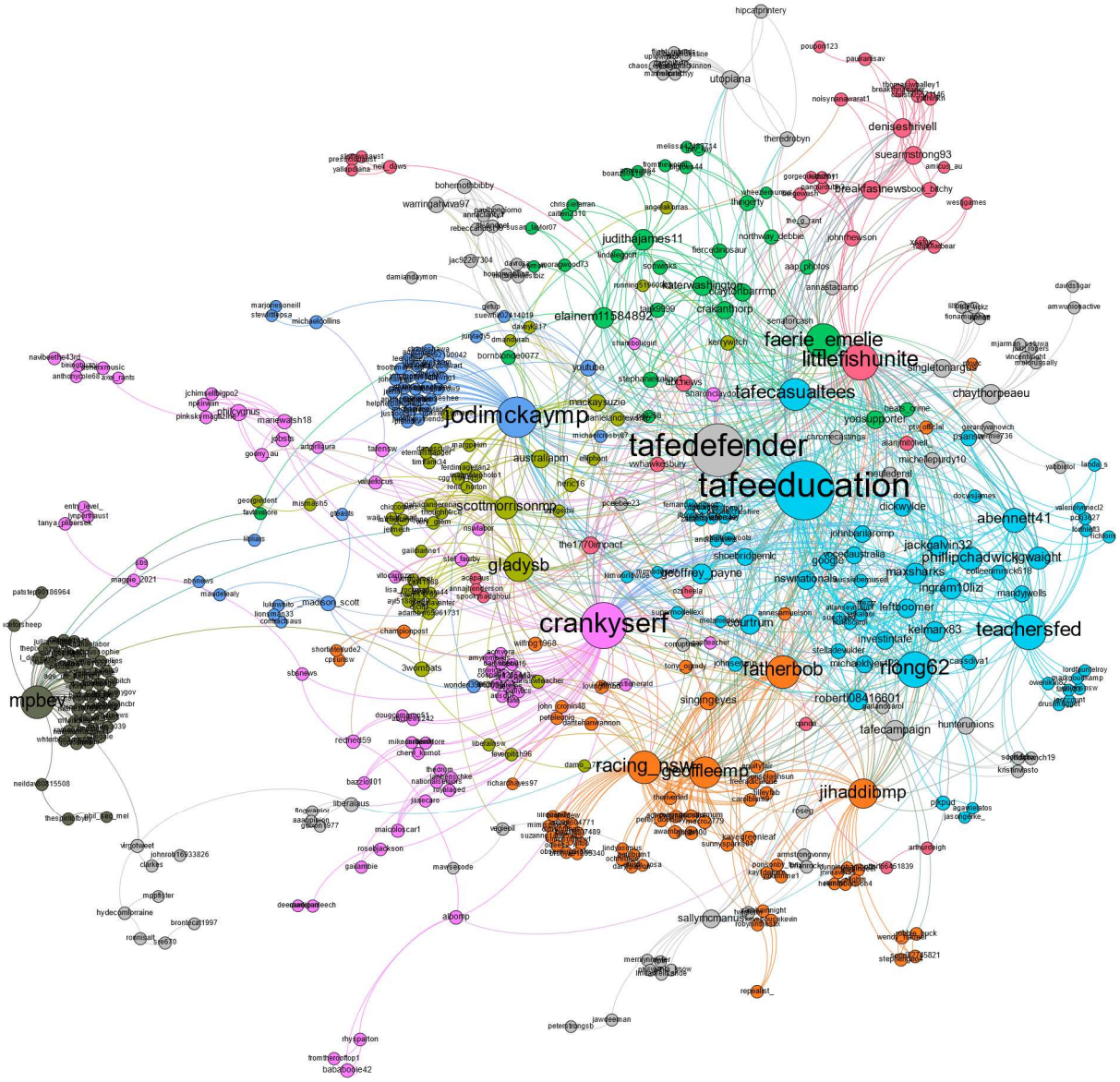


Figure 2c: Force Atlas modularity to degrees

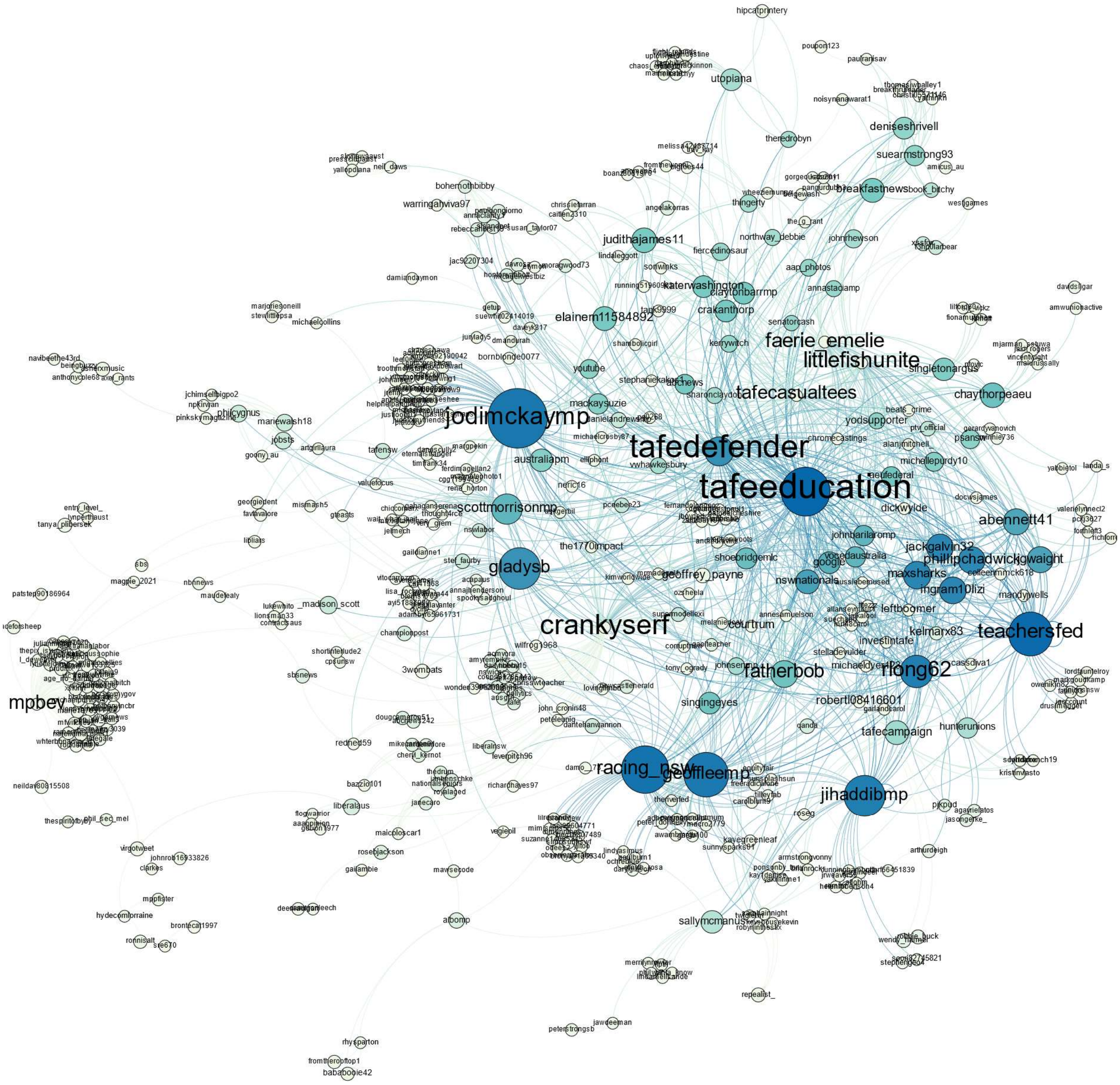


Figure 2d: Authority to in-degrees

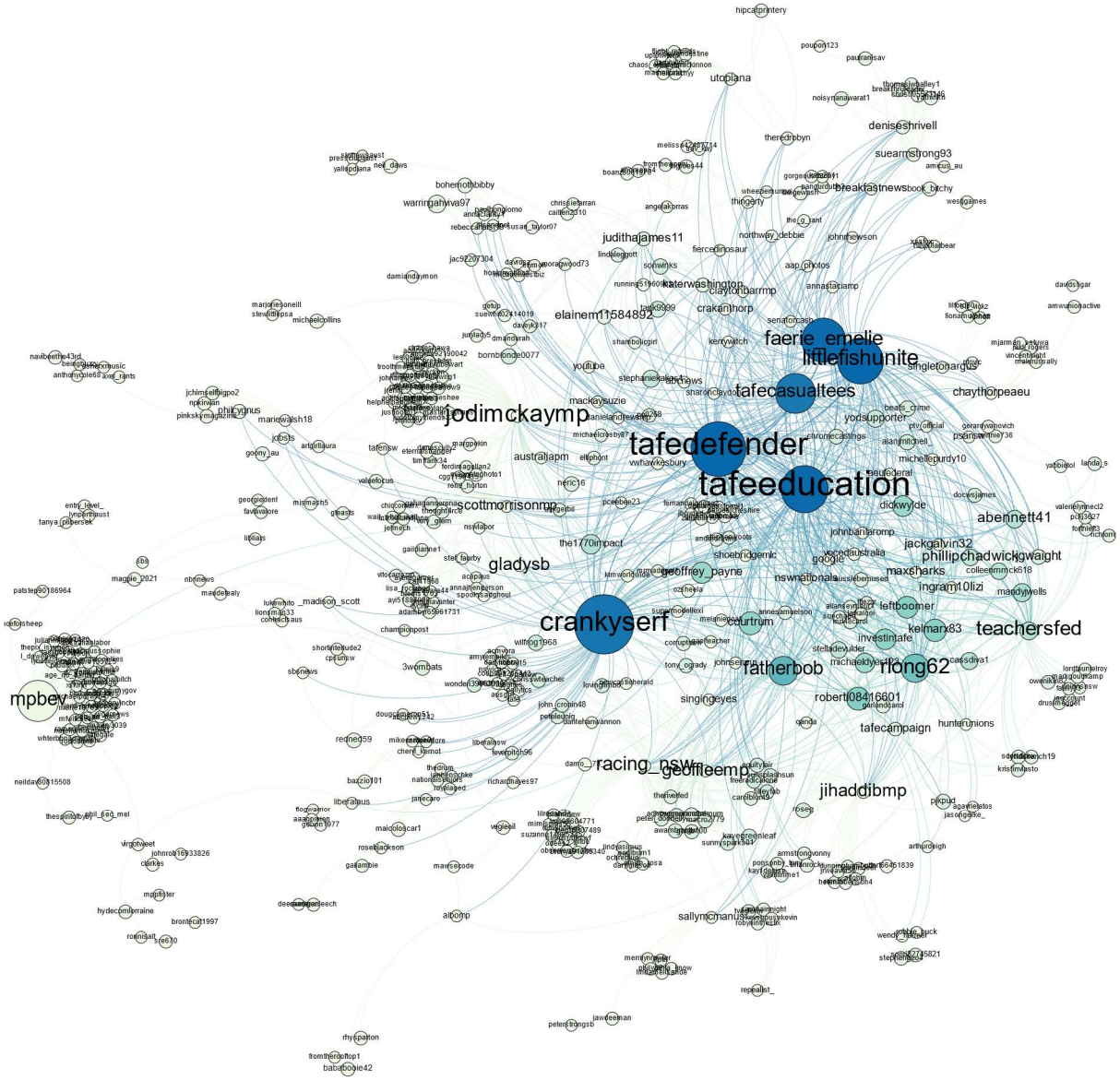


Figure 2e: Hub to out-degrees

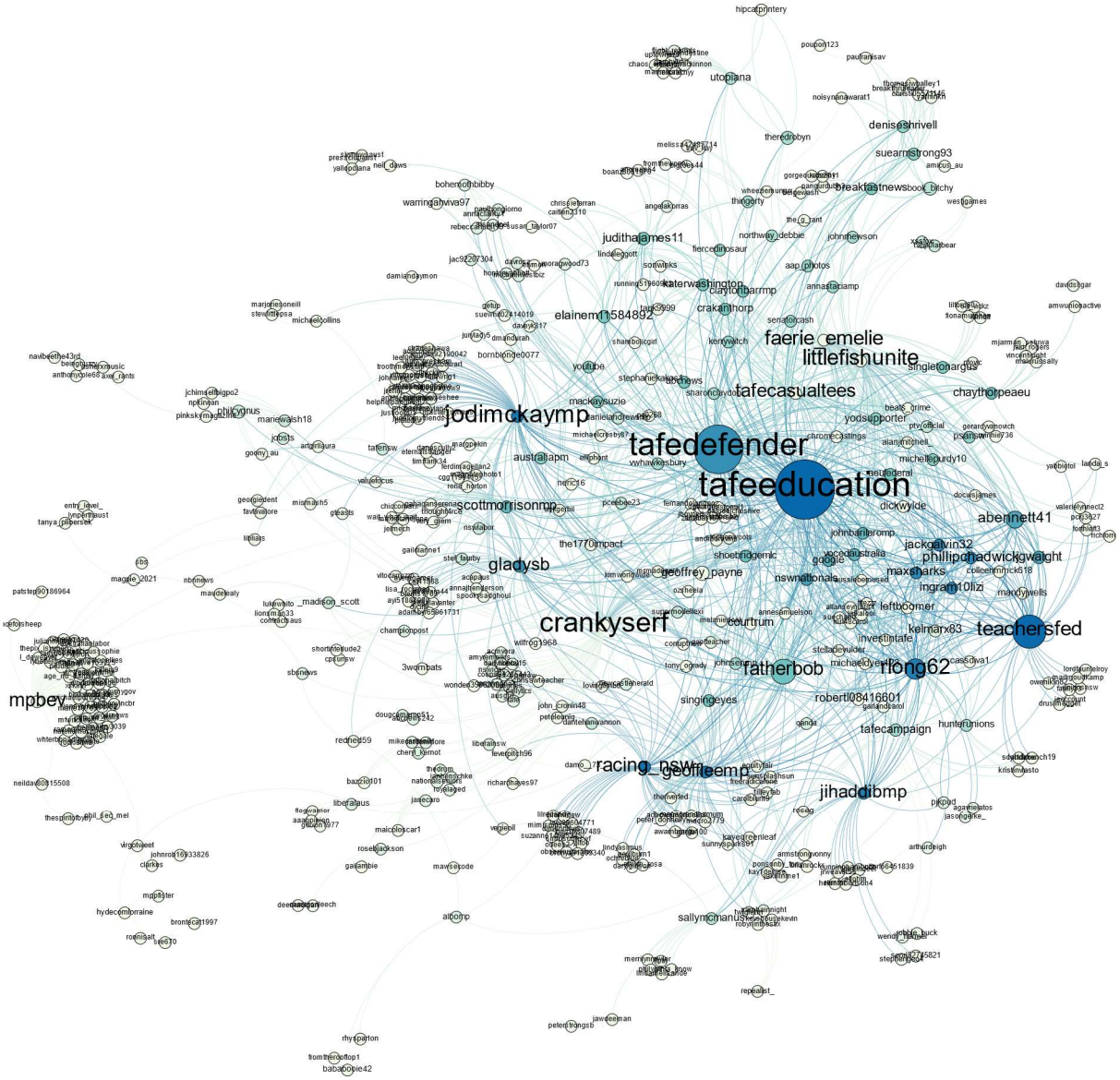


Figure 2f: Authority to betweenness

REFERENCES:

- Cherven, K., 2015. *Mastering Gephi network visualization*. Birmingham, U.K.: Packt Pub., pp.119-122.
- Kleinberg, J. (1999). *Authoritative sources in a hyperlinked environment*. Journal of the ACM, 46(5), 604–632. https://doi.org/10.1145/324133.324140
- Patty, A., 2021. *Racing NSW bids to buy TAFE site*. [online] The Sydney Morning Herald. Available at: <https://www.smh.com.au/national/nsw/racing-nsw-bids-to-buy-tafe-site-20210301-p576np.html> [Accessed 8 March 2021].