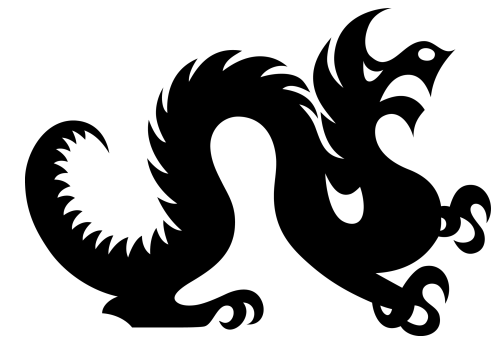
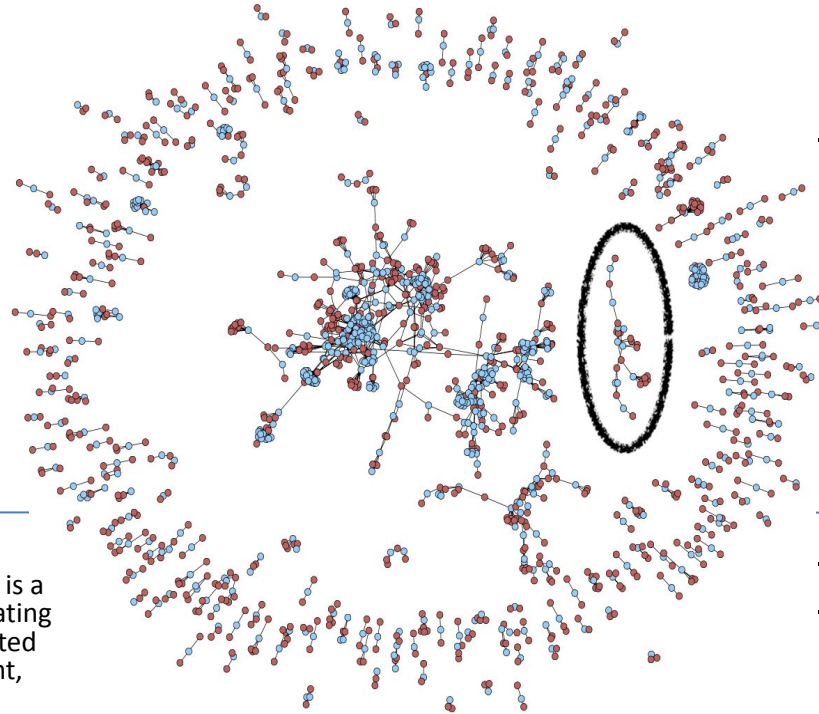


CRII: SaTC: Moderating Effects of Automation on Information Transmission in Social Forums



Challenge:

- Design multi-platform data streams for monitoring the information environment.
- Develop diagnostic applications that allow users and operators to navigate their environments more effectively by utilizing machine learning (ML)-based software and minimalist, publicly-available data and analytic features.



Solution:

- [NewsTweet](#) (example graph at center) is a publicly-accessible data stream integrating social and mass media, whose connected subgraphs form hierarchies of coherent, topical stories.
- Collect all of [PolitiFact](#)'s fact checks in real time in parallel to NewsTweet to generalize content veracity features.
- Broader-scope, platform-specific data streams are collected in real-time from Twitter, Facebook, Reddit, and 4Chan to link/expand from cross-platform content.
- Analysis utilizing mixed data and features to handle a diversity of automation types and coordination strategies.

Operational Landscape:

The bipartite, article-tweet network that we refer to as *NewsTweet*, draws edges between an article and each of the tweets it embeds. The graph shows structure from the week of March 10, 2020: red nodes are articles and blue nodes are embedded tweets, and subgraphs (such as the circled) are composed of highly-coherent topical clusters

Scientific Impact:

- Developed novel data of streams, rich for study, as well as software for their processing and analysis.
- Mathematical analysis has optimized operationally minimalist (text only) bot detection features.
- We've produced an ML task—now in pilot experimentation—via an REU supplement for ad hominem attack detection software.

Broader Impact and Broader Participation:

- The first distribution of our conversational analysis controller is released as the [pyconversations](#) module, on Python package index.
- Operating under the theory that NLP algorithms converge towards maximum-likelihood solutions, our work is spinning off mathematical methodologies for best-possible NLP models that can be computed cheaply on a CPU—*not a GPU*.