Frederick Jones Week 4

[**Week Four - Assignment Centrality Measures**](https://bbhosted.cuny.edu/webapps/assignment/uploadAssignment?content_id=_81616025_1&course_id=_2364492_1&group_id=&mode=view)

Data Source The data utilized in this analysis is sourced from the Stanford Network Analysis Project (SNAP). Specifically, we focus on the "Bitcoin OTC trust weighted signed network" dataset,found here [click here](https://snap.stanford.edu/data/soc-sign-bitcoinotc.html). This dataset comprises 5,881 nodes and 35,592 edges, with edge weights ranging from -10 to 10, representing trust ratings among members.

As described on the website, this dataset represents a network of trust among individuals who engage in Bitcoin transactions on the Bitcoin OTC platform. Due to the anonymity of Bitcoin users, maintaining a reputation record is crucial to mitigate risks associated with fraudulent transactions. Members rate each other on a scale from -10 (complete distrust) to +10 (complete trust), facilitating trust evaluation in the Bitcoin community. This dataset is the first explicit weighted signed directed network available for research (Bitcoin, 2018).

The data is stored in a compressed ".csv.gz" file, downloadable from the website. Each line in the file includes: SOURCE, TARGET, RATING, and TIME.

* SOURCE: Node ID of the rater (user providing the rating).
* TARGET: Node ID of the rated user.
* RATING: The trust rating assigned by the source to the target, ranging from -10 to 10 (integer values only).
* TIME: Timestamp of the rating in seconds since Epoch.

For analysis purposes, we categorize the RATING variable into Positive (RATING >= 0) and Negative (RATING < 0), treating it as a binary categorical variable.

High-Level Plan The high-level plan involves downloading the .csv.gz file, extracting it using tools like winzip, and loading the .csv file into a Jupyter notebook. Subsequently, we employ the network package for network analysis.

Our hypothesis posits that TARGET nodes with positive trust ratings will exhibit, on average, higher degree centrality compared to those with negative trust ratings. This conjecture is based on the assumption that a SOURCE user establishes connections only with TARGET users they trust. Hence, well-trusted TARGET users are likely to have a higher degree centrality due to their numerous connections.

Citation Bitcoin OTC trust weighted signed network (2018). Retrieved from <https://snap.stanford.edu/data/soc-sign-bitcoinotc.html>.

S. Kumar, F. Spezzano, V.S. Subrahmanian, C. Faloutsos. Edge Weight Prediction in Weighted Signed Networks. IEEE International Conference on Data Mining (ICDM), 2016.