A Network Model of Selective Exposure Using Community Detection

Subhayan Mukerjee, Ph.D. – National University of Singapore – email: mukerjee@nus.edu.sg – twitter: @wrahool

Motivation

Many studies use community detection techniques on audience co-exposure networks[1] to study selective exposure. However, their analytical choices are often arbitrary, rarely guided by systematic comparison of how different algorithms perform on audience co-exposure networks or grounded in network theory.

Model specification

M = $\{m_1, m_2, m_3, \ldots, m_{n1}\}$ media outlets with

R = $\{r_1, r_2, r_3, \ldots, r_{n1}\}$ reputation scores, drawn from a power-law distribution with exponent = α;

A = $\{a_1, a_2, a_3, \ldots, a_{n2}\}$ agents;

T = $\{t_1, t_2, t_3, \ldots, t_{n3}\}$ types of media outlets and agents;

Each agent a_i visits v_i outlets where v_i is drawn from a skewed normal distribution $N(\mu, \sigma, k)$;

Randomizing parameter $\rho \in [0, 1]$, a population level parameter that controls the selectivity in agent behavior. When $\rho = 0$, agents behave in a completely selective manner, visiting only outlets of the same type as themselves. When $\rho = 1$, agents behave completely randomly.

Network Construction: Baseline vs Augmented

- If the bipartite network between media outlets and agents is given by $G(M, A, E)$ with incidence matrix $\boldsymbol{B}$, the co-exposure network $G'(M, E')$ is defined by $\boldsymbol{B}^T\boldsymbol{B}$ with main diagonal elements set to 0[2].
- Prior work has shown that the main diagonal elements can affect the community structure revealed by community detection algorithms[3].
- So I define "baseline" vs "augmented" networks depending on whether the main diagonal elements are set to 0 or not.

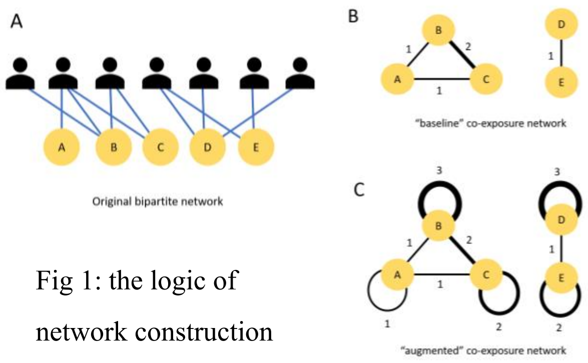

Fig 1: the logic of network construction

- Performance of algorithm is measured using the Normalized Mutual Information (NMI) score:

$$NMI(C,T) = \frac{-2\sum_{i=1}^{C}\sum_{j=1}^{T} N_{ij}\log\left(N_{ij}N/N_{io}N_{oj}\right)}{\sum_{i=1}^{C} N_{io}\log\left(N_{io}/N\right) + \sum_{j=1}^{T} N_{oj}\log\left(N_{oj}/N\right)}$$

Results on Simulated Networks

100 simulations for each value of ρ.

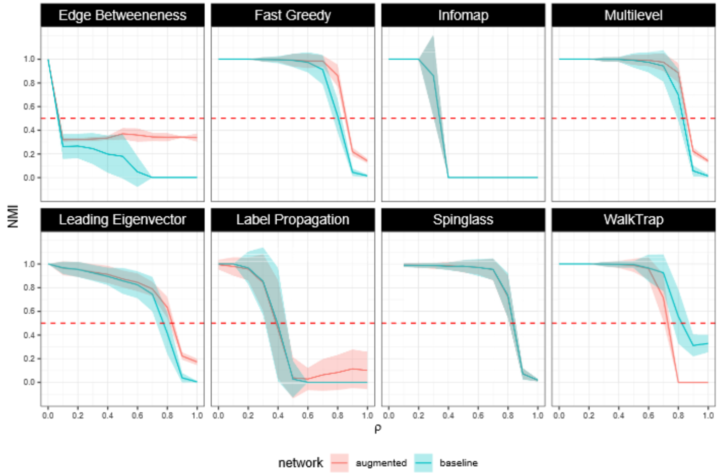

Fig 2: algorithm performance on simulated networks

Results on Empirical Network

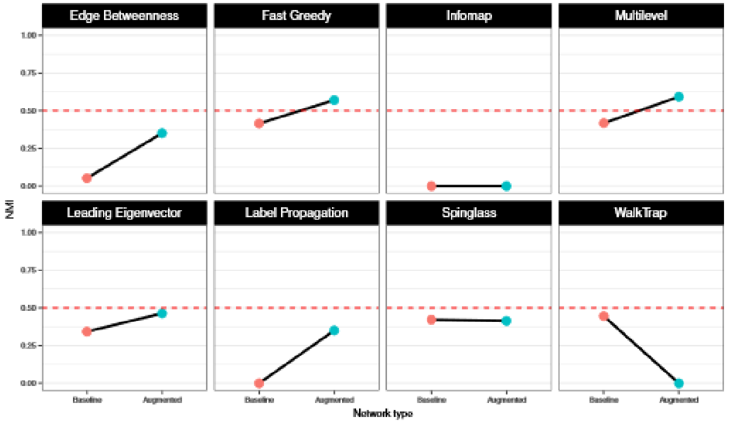

Fig 3: algorithm performance on empirical network

Conclusion

- Fast Greedy and Multilevel perform best.
- Their performances are further enhanced when the network is augmented
- Performances are replicated on an empirical network tracking a nationwide online population for nearly 4 years.

References

1. Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, Nir Grinberg, and David Lazer. Fake news on twitter during the 2016 U.S. presidential election.Science,363(6425):374–378, 2019.

2. M. E. J. Newman. Networks. Oxford University Press, Second Edition

3. Alex Arenas, Alberto Fernandez, and Sergio Gomez. Analysis of the structure of complex networks at different resolution levels. *New Journal of Physics*, 10(5), 2008