Clustering Coefficient Analysis Through Small-Scale Diffusion Testing

Claire Kelling*, Ashton Verdery**

Penn State University
Department of Statistics*, Department of Sociology and Criminology**

Women in Statistics and Data Science Conference

Motivation and Current Literature

In order to conduct an appropriate intervention for a given population, policymakers must know the vulnerability of that population's social network. We conduct an analysis of **diffusion over social networks** to determine network vulnerability.

- Neman et al. (2003) found that as clustering is increased, the size of the ultimate epidemic declines but the epidemic threshold, the level of infectivity needed for the epidemic to take off (hence, the speed of epidemic realization), is decreased [3].
- Keeling et al. (2005) found that increases in clustering increases epidemic thresholds [1].
- Finally, Kiss et al. (2008) attempt to resolve this debate by noting "[t]o study the theoretical effects of varying one network property (e.g., clustering), one would ideally like to generate multiple networks with all properties identical, except the property of interest" [2].

Methods and Measures

Network Characteristics:

- 6, 8, 10, and 12 node regular networks
- Degree 3
- Complete set of isomorphic networks

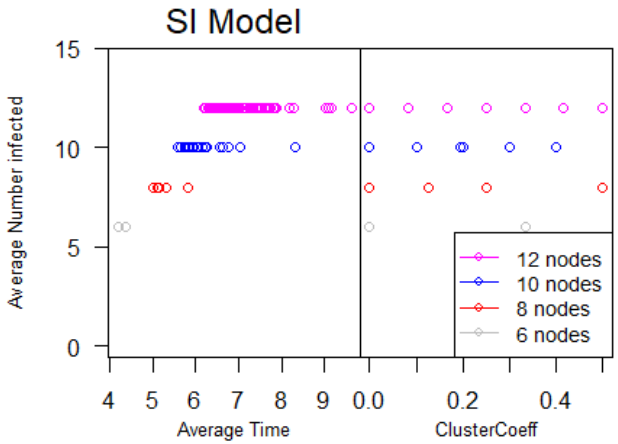
Diffusion Models:

- SI (Susceptible Infected) Model
- SIR (Susceptible Infected Recovered) Model

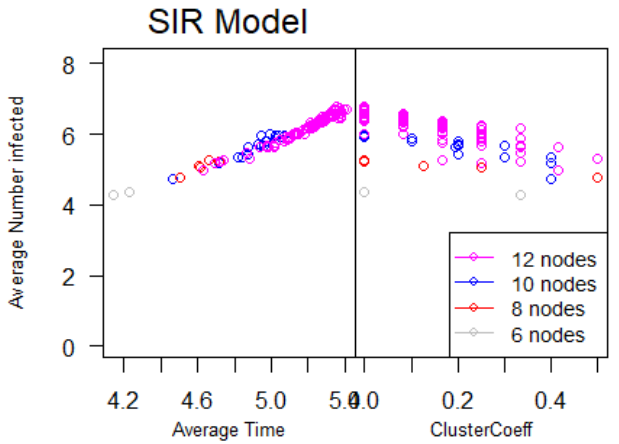
Outcomes:

- average number of time periods until the network is either saturated or no other node is infected
- total number of nodes that are infected at the end of the diffusion process

Results



(a) In the SI model, the number of nodes that are infected after the diffusion process is simply the total number of nodes in the network. As the number of nodes increases, so does the the time of infection.



(b) In the SIR model, in general, as the time increases, so does the number of nodes that are infected. There is also a linear relationship between the clustering coefficient and the number of people infected.

So, in our case for the SIR model, by using the complete set of isomorphic graphs and by controlling for many possible network features, we agree with [3], the size of the epidemic (number infected) declines as well as the speed of the realization.

References

[1] Matt Keeling. The implications of network structure for epidemic dynamics. *Theoretical population biology*, 67(1):1–8, 2005.

[2] Istvan Z Kiss and Darren M Green. Comment on âĂIJproperties of highly clustered networksâĂİ. *Physical Review E*, 78(4):048101, 2008.

[3] Mark EJ Newman. Properties of highly clustered networks. *Physical Review E*, 68(2):026121, 2003.