Clustering Coefficient Analysis Through Small-Scale Diffusion Testing

Claire Kelling*, Ashton Verdery**   Departments of Statistics* and Sociology/Criminology**

Motivation

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.

Definitions

- **Isomorphic**: Two graphs G and G', each of n points, are called **isomorphic** if there exists a 1-1 correspondence between the points of G and G' which preserves adjacency.
- **Regular**: A **k-regular graph** is a graph in which each node has exactly degree k [4].
- **Susceptible-Infected-Recovery**: This model incorporates behavior where individuals are susceptible and may be infected with a certain probability (p). After they are infected, they have a certain probability of recovery.
  - In our case, we use a very simple SIR Model where the probability of recovery is 100% after one time interval, and they cannot be infected again from other nodes.
- **Susceptible-Infected**: In this model, there is also a given probability of infection, but there is no chance for those who have been infected to recover.
  - In both cases, we let the probability of infection, p, be 50%.

Literature

- 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 [1].
- Keeling et al. (2005) found that increases in clustering increases epidemic thresholds [2].
- 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" [3].

Results

We consider a graphical representation where we can compare the average time of diffusion, the average number of nodes that are infected, the number of nodes in the network, and the clustering coefficient of the network. This is mostly in order to address the concerns of [3], in order to control for other variables.
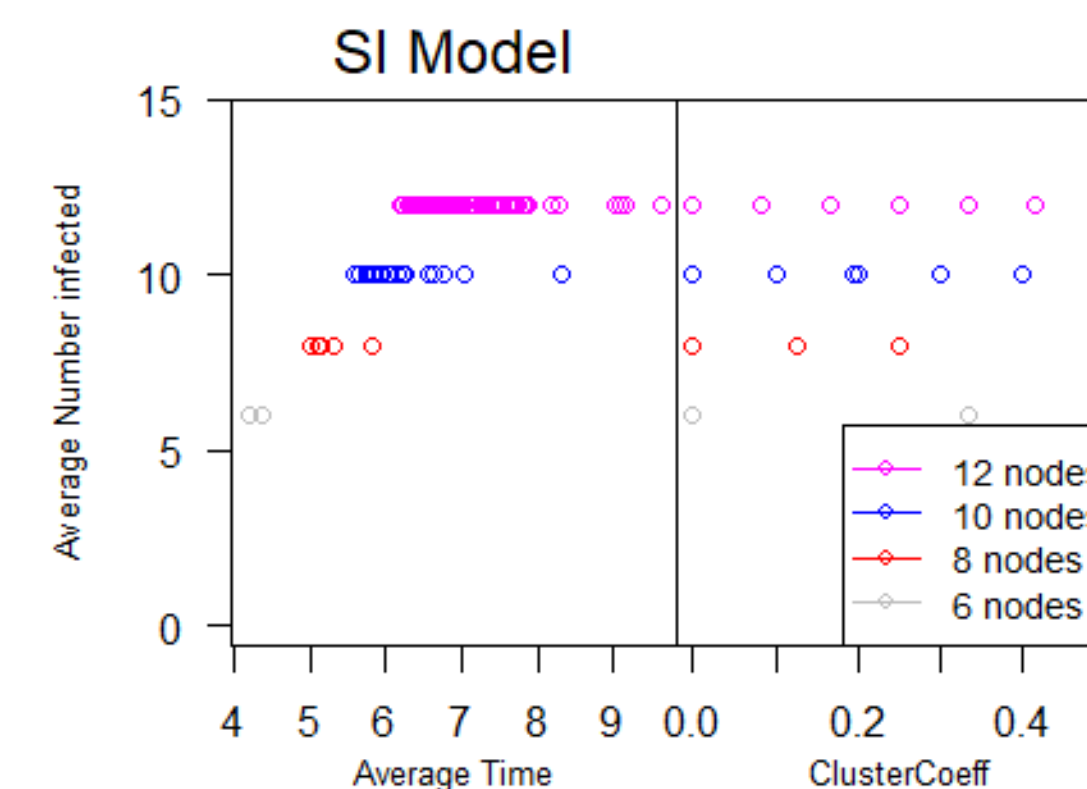

Figure 1: In the SI model, the number of nodes that are infected after the diffusion process is simply the total number of nodes in the network. Also, as the number of nodes increases, so does the the time of infection.

Figure 2: 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 [1], the size of the epidemic (number infected) declines as well as the speed of the realization.
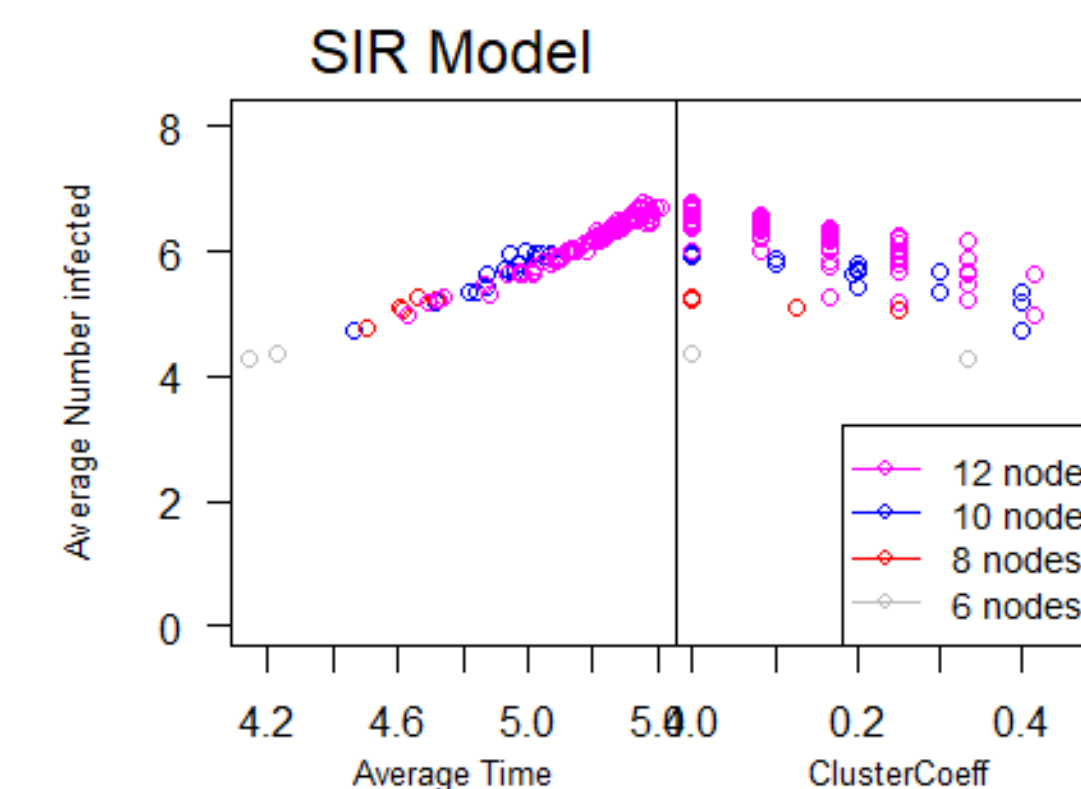
Methods and Measures

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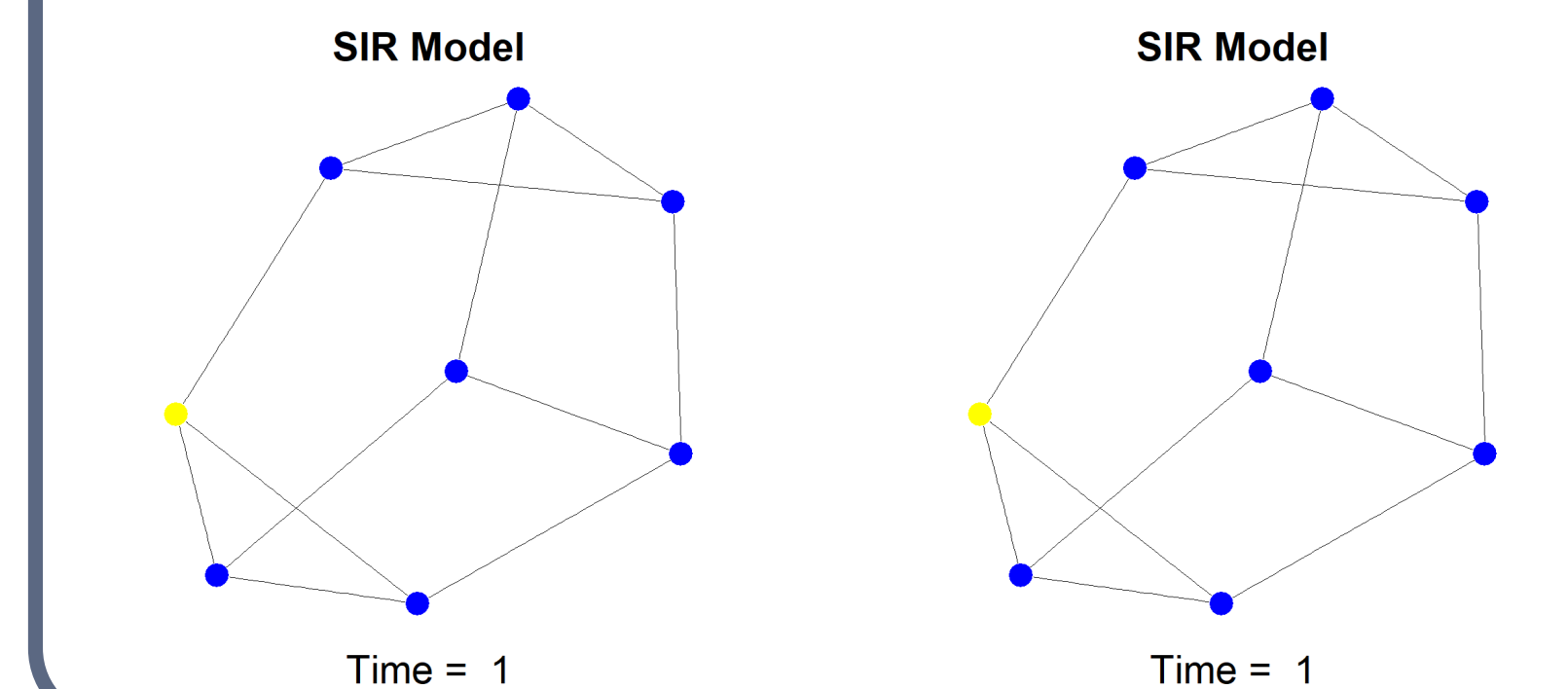- Complete set of isomorphic networks

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

SI/SIR Model Examples

References

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

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

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

[4] Markus Meringer. Fast generation of regular graphs and construction of cages. *Journal of Graph Theory*, 30(2):137–146, 1999.