

Effects of Contact Network Models on Stochastic Epidemic Simulations

Rehan Ahmad and Kevin S. Xu

EECS Department, University of Toledo, Toledo, OH 43606, USA
Rehan.Ahmad@rockets.utoledo.edu, Kevin.Xu@utoledo.edu

1 Introduction

The importance of modeling the spread of epidemics through a population has led to the development of mathematical models for infectious disease propagation. Such models are heavily dependent on the interaction between individuals. A number of empirical studies have provided estimates of frequency and duration of interactions between individuals [3,4], which affect the disease propagation. Other work has considered modeling disease propagation on random contact networks with different properties [1,2].

The objective of our work is to compare different models for contact networks and to find the best model to use to simulate contact networks that are close to the actual network. We do this by comparing the disease dynamics of a stochastic Susceptible-Infectious-Recovered (SIR) model over the simulated network with the disease dynamics over the actual network, since our goal is to find the model that best preserves disease propagation dynamics.

We consider a stochastic SIR model applied to a range of network contact models that differ in their complexity levels over six datasets of interactions between people in a variety of settings. Our results demonstrate that the choice of network model can have a significant effect on how closely the outcomes of the epidemic simulation on a simulated network match the outcomes on the actual network. In particular, preserving degrees of nodes appears to be much more important than preserving cluster structure for accurate epidemic simulations.

2 Datasets and Methods

We consider a variety of contact network datasets in this study. Table 1 shows summary statistics for each dataset. We evaluate the quality of a contact network model for simulations of epidemics by conducting the following steps for each dataset:

1. Simulate 5,000 epidemics over the actual network.
2. Fit contact network model to actual network.
3. Simulate 100 networks from contact network model. For each simulated network, simulate 50 epidemics over the network for 5,000 epidemics total.
4. Compare the results of the epidemic simulations over the actual network with those over the simulated networks.

	HYCCUPS	Friends & Family	High School	Infectious	Primary School	HOPE
Number of nodes	43	123	126	201	242	1178
Sensor type	Wi-Fi	Bluetooth	RFID	RFID	RFID	RFID
Proximity range	N/A	5 m	1–1.5 m	1–1.5 m	1–1.5 m	Room
Graph density	0.326	0.228	0.217	0.0328	0.285	0.569
Clustering coefficient	0.604	0.496	0.522	0.459	0.480	0.748
Average degree	14.0	27.8	27.1	6.56	68.7	671
Maximum degree	28	73	55	21	134	1072

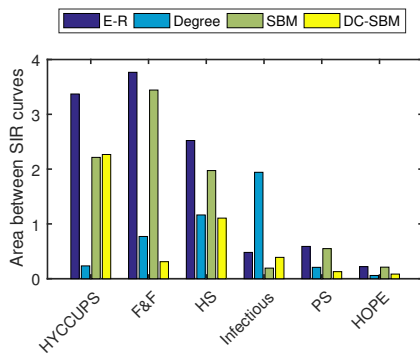
Table 1: Summary statistics from datasets used in this study.

These steps are repeated for each contact network model that we consider: Erdős-Rényi Model, degree Model, stochastic block model, and degree-corrected stochastic block model. To get a fair evaluation of the dynamics of epidemics spreading over different contact network models, all of the parameters which are not related to the contact network model are kept constant. In our experimental setup, we have set the probability of infection for every interaction between people to be equal to 0.025. The probability of recovery is also set to be 0.025.

3 Results and Discussion

Unlike most prior studies such as [3], we measure the quality of a network model by the area between its mean SIR curve over all epidemics on the simulated networks compared to the mean SIR curve for epidemics on the actual network. By measuring the area between the curves (e.g. the fraction of recovered people after the disease dies out as in [3]), we capture the difference in transient dynamics (e.g. the rate at which the infection spreads) rather than just the difference in final outcomes. Our findings suggest that the degree-corrected stochastic block model (DC-SBM) is the best choice of contact network model in epidemic simulations because it resulted in the minimum average area between SIR curves. Interestingly, using the degree model resulted in an average area between SIR curves only slightly larger than the DC-SBM despite having less than half as many parameters. The SBM (without degree correction) also has half as many parameters as the DC-SBM, but has over twice the area between SIR curves. We note that the difference between the degree model and the SBM *cannot* be observed using log-likelihood as the quality measure, as both models are very close in log-likelihood. This leads us to believe that preserving degree has a greater effect on accuracy of epidemic simulations than preserving community structure. Furthermore, this finding demonstrates that one cannot simply evaluate the accuracy of a contact network model for epidemic simulations only by examining goodness-of-fit on the actual contact network!

In this study, we estimated the parameters for each contact network model using the contact network itself, which we cannot do in practice because the contact network is often unknown. As a result, one would have to estimate the



(a)

Model	Area	Log-lik	# param
E-R	1.82	-0.597	1
Degree	0.73	-0.496	319
SBM	1.43	-0.504	328
DC-SBM	0.71	-0.385	647

(b)

Fig. 1: Comparison of (a) area between SIR curves of each model with respect to actual network for each dataset and (b) shows quality measures (higher is better for log-likelihood; lower for others) averaged over all datasets for each model.

model parameters from prior knowledge or partial observation of the contact network, which introduces additional error that was not studied in this paper. It would be of great interest to perform this type of sensitivity analysis to identify whether the DC-SBM and degree model are still superior even when presented with less accurate parameter estimates. Also, there is a risk of overfitting in more complex models which should be examined in a future extension of this work. Another limitation of this study is our consideration of static unweighted networks. Indeed, prior work [3] has shown that it is important to consider the time duration of contacts between people, which can be reflected as weights in the contact network, as well as the times themselves, which can be accommodated by using models of dynamic rather than static networks. We plan to expand this work in the future by incorporating models of weighted and dynamic networks to provide a more thorough investigation.

References

1. Eames, K.: Modeling disease spread through random and regular contacts in clustered populations. *Theoretical Population Biology* 73(1), 104–111 (2008)
2. Larson, R.C.: Simple models of influenza progression within a heterogeneous population. *European Journal of Operational Research* 55(3), 399–412 (2007)
3. Machens, A., Gesualdo, F., Rizzo, C., Tozzi, A.E., Barrat, A., Cattuto, C.: An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices. *BMC Infectious Diseases* 13(1), 185 (2013)
4. Smieszek, T., Fiebig, L., Scholz, R.W.: Models of epidemics: when contact repetition and clustering should be included. *Theoretical Biology and Medical Modelling* 6(1), 11 (2009)