

# Epidemic Spreading on Weighted Contact Networks

P. Schumm, C. Scoglio, D. Gruenbacher  
EECE  
Kansas State University  
{pbschumm, caterina, grue}@ksu.edu

T. Easton  
IMSE  
Kansas State University  
teaston@ksu.edu

## ABSTRACT

The study of epidemics is a crucial issue to several areas. An epidemic can have devastating economic and social consequences. A single crop disease in Kansas could destroy the yearly income of many farmers. Previous work using graph theory has determined a universal epidemic threshold found in the graph topology for a binary contact network in the compartmental Susceptible-Infected (SI) analysis. We expand this threshold to a more realistic measure. A binary uniform level of contact within a society is too idealistic and an improved threshold is found in allowing a spectrum of contact within a contact network. The expanded contact network also allows for asymmetric contact such as a mother caring for her child. Further study in this area should lead to improved simulators, disease modeling, policies and control of infectious diseases and viruses.\*

## Keywords

Epidemiology, agent based, aggregate, dynamic parallel network simulation, weighted contact networks

## 1. INTRODUCTION

Compartmental models are the most widely studied class of epidemics models. The *Susceptible-Infected* (SI) model assumes no immunity or recovered state. This allows the agents to be re-infected repeatedly, thus switching back and forth between the two states. This could occur with certain malicious viruses or in an agricultural setting where either there is no immunity associated with the disease or the agent represents a crop that can be re-infected again. In the SIR compartmental model, first formulated (though never published) by Reed and Frost in the 1920s, each node represents an individual who can belong, in a given time, to one of three compartments, *Susceptible*, *Infected*, or *Recovered* (SIR). Key parameters in this model are the infection rate  $\beta$ , which captures the aetiology of the infection process, and the recovery rate  $\delta$  of an infected node [2]. The SEIR model adds an *Exposed* compartment increasing its effectiveness at modeling biological diseases.

In graph theory, a graph consists of a set of points called nodes

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Bionetics'07, December 10–13, 2007, Budapest, Hungary.  
Copyright 2007 ICST 978-963-9799-11-0

or vertices. The interconnections among the nodes are known as links or edges, and in a contact network, they represent a form of contact or relation. The neighbors of a node are those connected to the node. A node's degree is the number of neighbors it has. We can define a node to represent a single individual or group, a small location or county, or a computer or computing group. A graph can be described mathematically by an adjacency matrix  $\mathbf{A}$  where row  $i$  contains the neighbors  $j$  of node  $i$  by a 1 in the appropriate  $j^{\text{th}}$  columns with the remaining entries 0.

In this paper, we expand an SI epidemic threshold to accommodate asymmetric and variable contact by weighting the links of the contact networks. The introduced threshold brings to light two interesting topics. First, the assignment of the weights and how they are distributed can have significant effects on the epidemic resistance of the topology. A single topology can obtain a spectrum of virus resistance for different weight placements. Secondly, as weights are assigned they can form regions of higher weights that are much more conducive to epidemic spreading. These regions behave as disease "highways" across the network.

We take and expand the binary-contact threshold of [22] to include weighted asymmetric contact in Section 3.1, where we utilize weighted links to take the contact out of the disease parameter  $\beta$  and place it on the graph. Section 3.2 discusses the simulation tools we are constructing to model epidemics spreading and transfer, and their application to this work. We expound on the weights in Section 3.3 and investigate how they differentiate topologies in Section 4, with further insight and conclusions in Section 5.

## 2. RELATED WORK

The epidemic threshold of an SI system is the line between the existence of infected nodes and their non-existence as time tends to infinity. Spectral analysis is used in [22], where the authors present a model of virus propagation on a given realistic topology to demonstrate that the topology of the underlying network has a strong impact on the speed of propagation. Using a probabilistic model and by some approximations, they found the epidemic threshold

$$\frac{\beta_a}{\delta} \leq \frac{1}{\lambda_{1,A}} \quad (1)$$

where  $\beta$  is the average rate of infection along a link,  $\delta$  is the curing rate on an infected node, and  $\lambda_{1,A}$  is the largest eigenvalue of the adjacency matrix of the network topology. If the above condition is met, there is no epidemic. Using this model, they

\* This paper is best viewed in a color format.

confirmed the result in [20] stating that an infinite power law graph has a threshold which tends to zero. More rigorous results on the same topic are presented in [14], where the question of what makes an epidemic either weak or potent is addressed. More precisely, the authors have identified topological properties of the graph that determine the persistence of epidemics. In particular, they showed that if the ratio of cure to infection rates is smaller than the spectral radius of the graph, then the mean epidemic lifetime is of order  $\log n$ , where  $n$  is the number of nodes, so the epidemic dies out quickly. Conversely, if this ratio is bigger than a generalization of the isoperimetric constant of the graph, then the mean epidemic lifetime is of order  $e^a$  for a positive constant  $a$ , resulting in the epidemic dying out slowly.

The main focus of this paper is a study of the effects of weighting the network's links. The application of an infection rate that varies among the links has been investigated previously with a few fixed  $\beta$ 's [17][8],  $\beta_{ij}(\text{distance})$  [12][16],  $\beta(\text{time})$  [15], randomly generated  $\beta_{ij}$  [19], and  $\beta_{ij}$  from a bimodal Gaussian distribution [21]. In all of the above, the contact measures are part of the respective  $\beta$  and not the topology. Historically, the components of disease related contact have been included in the  $\beta$  parameter and not as part of the graph and its links.

Researchers have developed a number of simulators to model and predict the spread of diseases. Two notable ones are the Epidemiological Simulation System (EpiSims) [9][11] from Los Alamos National Laboratory and the Spatiotemporal Epidemiological Modeler (STEM) [13] from IBM. EpiSims is an exhaustive simulation tool that provides precise results; however this creates an extremely large simulation. EpiSims provides details about every simulated person and the significant event that happens to each person during the simulation—including infection, incapacitation, and treatment—along with a time stamp and current location [9]. STEM is a dual of EpiSims as an aggregate simulator that divides the details down to a county level, yet can provide disease predictions for an entire country.

### 3. MODEL AND SIMULATION

#### 3.1 Model

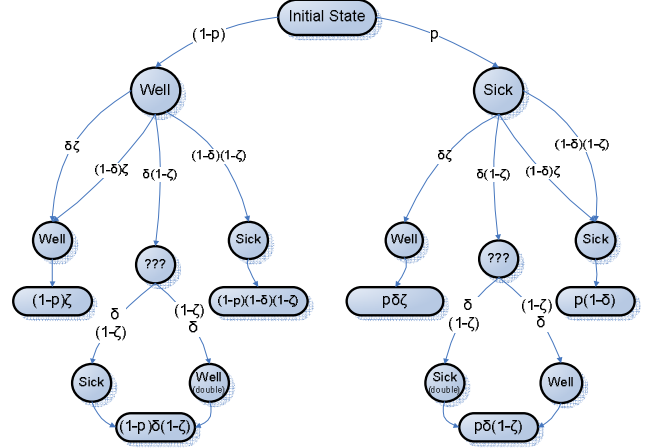
We develop an initial SI model based on the one of [22]. With the two states, *Susceptible* and *Infected*, we have the three probabilities for the state  $p$ , for not receiving an infection  $\zeta$ , and for a cure and return to the susceptible state  $\delta$ . Starting with the probability  $\zeta$  of a node  $v_i$  not receiving infections from its neighbors  $v_j$ ,

$$\zeta_{i,t} = \prod_j (1 - \beta_a p_{j,t-1}) \quad (2)$$

Where  $p_{i,t}$  is the probability of node  $v_i$  being sick at time  $t$ , we follow a probability tree to form the core SI model in Equation 3.

$$p_{i,t} = 1 - (1 - p_{i,t-1})\zeta_{i,t} - \delta p_{i,t-1}\zeta_{i,t} - \frac{1}{2}\delta p_{i,t-1}(1 - \zeta_{i,t}) \quad (3)$$

This selection of probabilities produces the same initial equations as those of [22]. They derive Equations 2 and 3, and compare these to a simulation to prove their modeling accuracy.



**Figure 2. The Probability Structure of an SI Model**

**We neglect the terms which switch states twice and reduce the probability of the sequenced terms.**

Using small-value approximations and converting to a matrix notation, they work out Equation 1. Our initial model differs from theirs by adding a weight  $\omega$  in Equation 2.

$$\zeta_{i,t} = \prod_j (1 - \omega_{ij} \beta_f p_{j,t-1}) \quad (4)$$

Equation 3 remains unchanged except in the value of  $\zeta$ . The adjacency matrix now becomes a weighted matrix  $\mathbf{W}$ , where a  $\omega_{ij}$  between 0 and 1 replaces every 1 in  $\mathbf{A}$ . When we absorb the average weight  $\bar{\omega}$  into  $\mathbf{W}$ , we define  $\mathbf{W}' = \mathbf{W}/\bar{\omega}$ . Deriving a threshold in a similar manner, we have the expanded threshold

$$\frac{\beta_a}{\delta} \leq \frac{1}{\lambda_{1,W'}} \quad (5)$$

where  $\lambda_{1,W'}$  is the largest eigenvalue of the weighted and scaled matrix  $\mathbf{W}'$ .

#### 3.2 Simulation

The thresholds are verified by parallel simulations run on our SI simulator. We are designing algorithms and software for parallel network simulation and topology analysis with mobile agents that allow dynamic environmental inputs. We work to design efficient, realistic simulations that allow real-time implementation of virtually any contagious virus. Simulation topics include networks composed of humans, animals, biological cells, computers, and crops with disease related spreading. For rural graphs, we plan to interface them with GIS software. We test the thresholds on several un-weighted topologies with the SI model to verify their accuracy. Using the road structure of a northeast Kansas Indian reservation, we designed a simple farm contact graph (PIR) to investigate rural epidemics. We collected a layout of the Abilene [1] backbone network for an internet analysis. The largest graph we analyzed comes from the Cooperative Association for Internet Data Analysis (CAIDA) [6] and their analysis of autonomous systems from March 12, 2007 with over 24,000 nodes. We also utilize classic topologies: star, mesh, Erdős-Rényi ER [10], scale free SF [4], and a ring with

node degree 100. Of these, the ER and SF are designed to model realistic networks. We depict these graphs in the appendix. Note that the use of these graphs is not an attempt to model a realistic system, but to verify the accuracy and universal property of the model on several different types of graphs. We intend to study the relationships between other graph metrics and the differences observed among the disease propagation in the graphs. Although the graphs tested here are static, we are working on “intelligent” graphs that can demonstrate the effects of various policies in

response to a disease or virus, even as simple as nodes that reduce contact with sick nodes.

Figures 3 through 5 demonstrate how the model aligns with the SI threshold for each un-weighted contact network respectively. Initially, a fourth of the nodes are infected, and then simulated over 1000 discrete time segments. The thresholds are depicted in yellow. The simulation is the oscillating lines, while the model predicts the smooth ones. The threshold is easy to identify in the SI model, being the lowest line not reaching zero.

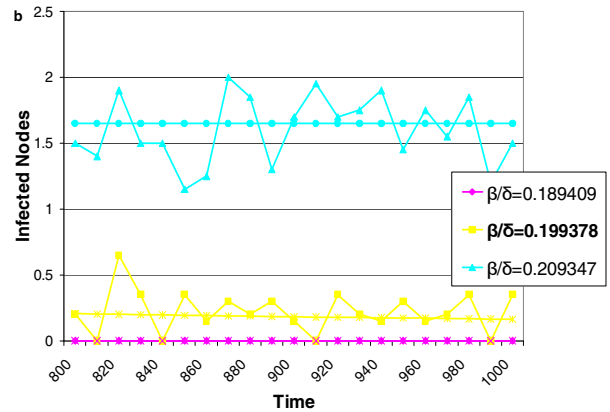
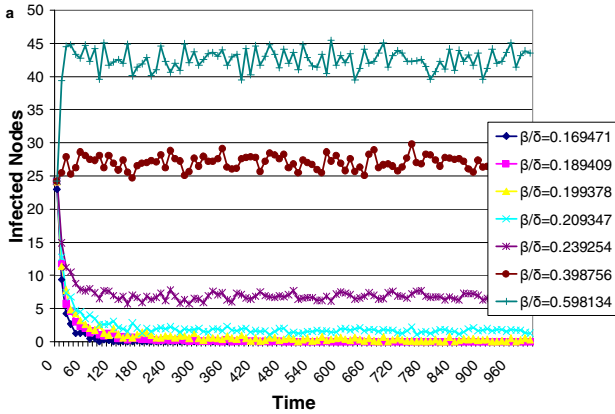


Figure 3. Simulation and threshold of Abilene Backbone network (93 Nodes) with binary links.

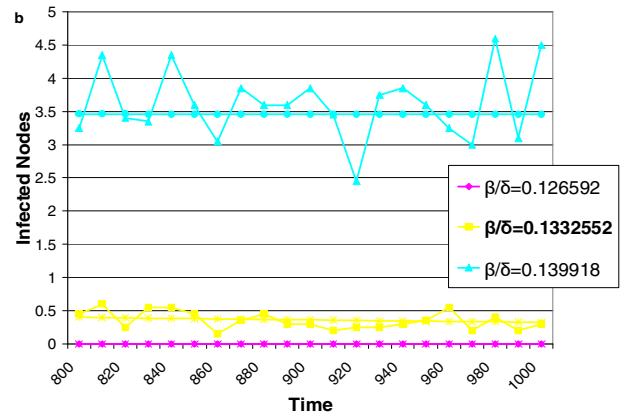
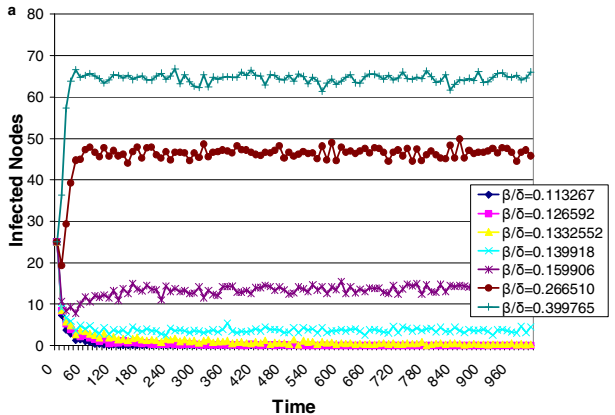


Figure 4. Simulation and threshold of PIR farms network (102 Nodes) with binary links.

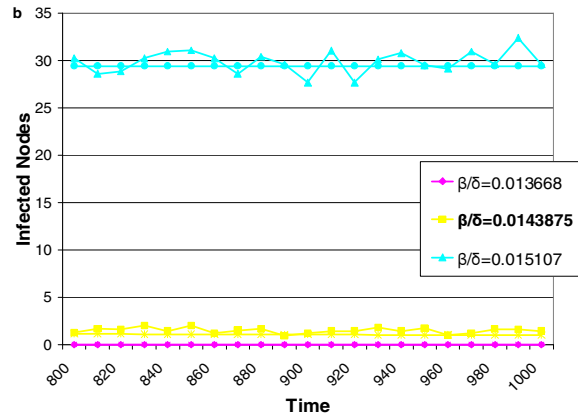
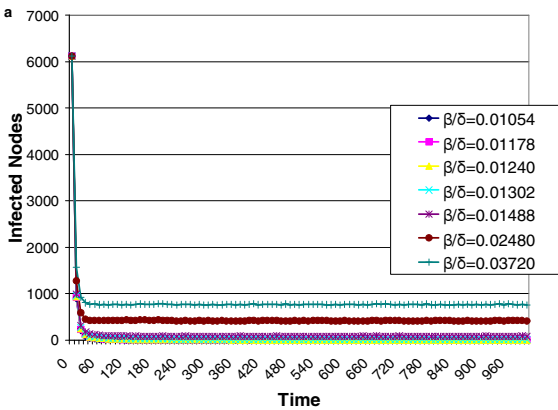


Figure 5. Simulation and threshold of CAIDA autonomous systems network (24491 Nodes) with binary links.

### 3.3 Weights

The weight  $\omega$  gives the ability to model varying levels of contact [18]. When we assign  $\omega$ , we change slightly the definition of  $\beta_a$ . Initially,  $\beta_a$  was a probability that described the average infection rate along a link. It contains within it an assumed average contact, but now we place this contact on the topology. We define  $\beta_f$  as a full contact measure of  $\beta$  as

$$\beta_f \bar{\omega} = \beta_a \quad (6)$$

So now  $\beta_a$  has an adjusted upper bound described in Equation 7.

$$\begin{aligned} \beta_f \omega_{ij} &\leq 1 \\ \frac{\beta_a}{\bar{\omega}} \omega_{ij} &\leq 1 \\ \omega_{ij} &\leq 1 \\ \beta_a &\leq \bar{\omega} \leq 1 \end{aligned} \quad (7)$$

Without the  $\bar{\omega}$  introduced into the model, we would always produce a higher threshold, indicating that the weights are improving the graph's disease resistance, since when entries  $a_{ij} \leq b_{ij}$  for all  $i$  and  $j$ , it follows  $\lambda_{1,A} \leq \lambda_{1,B}$ , where  $\lambda_{1,A}$  and  $\lambda_{1,B}$  are the largest eigenvalues of graphs A and B respectively [3]. We assigned the contact weights to PIR by the presence of a fence or road and the length of the adjacent boundary. For Abilene's weights, we assigned them by the links' physical capacity and where recorded, the traffic. Other graphs' weights are assigned randomly from a normal distribution of weights to investigate their thresholds as weighted graphs. For homogeneous graphs, it is noted that the weighted graphs appear to have at best the same threshold as the corresponding adjacency graph. We observed that as the standard deviation of the weight distribution tends to zero on homogeneous graphs, the threshold comes down to the binary contact threshold. Homogeneous weighted graphs were never seen to have a better threshold than their binary contact counterparts in the trials run by the authors.

With the introduction of weights, there arises a new form of clustering, namely, weight clusters. A weight cluster occurs as a connected group of highly infectious agents, without requiring these agents to be a topological cluster. Such a cluster can serve as a virus super-highway through the network. Figures 6 through 9 are colored by the weights on each link and the normalized weight-degree (the sum of the weights on all adjacent links) of the nodes. The color order starts with purple as the highest, running down through the reds to greens with the black/grey as the lowest weights. This representation allows us to see clusters that are topology independent, such as a string of highly weighted nodes through a community or region. In the Abilene cluster map, the link colors range over a smaller set of the possible colors due to a tighter weight distribution from the hardware capacities. With the addition of the authors' selection of contact weights, the threshold of Abilene changes by only 8%, while the PIR threshold swings 30%. A network connection between two large groups of nodes could be poorly connected in graph theoretic terms; yet if those links comprise some of the highest weights, the connection could be easily crossed resulting in a larger epidemic. Both types of clustering should be considered in future epidemic control and prevention policies.

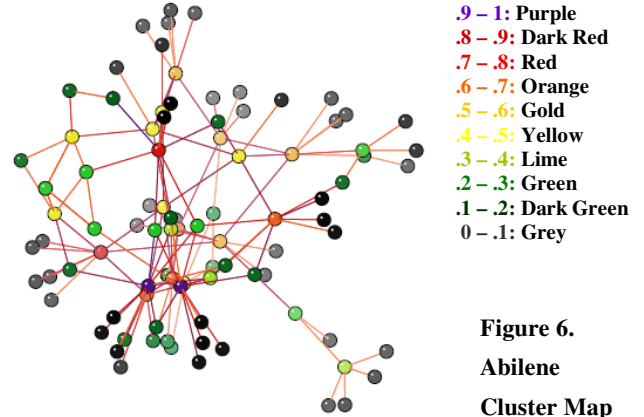


Figure 6.  
Abilene  
Cluster Map

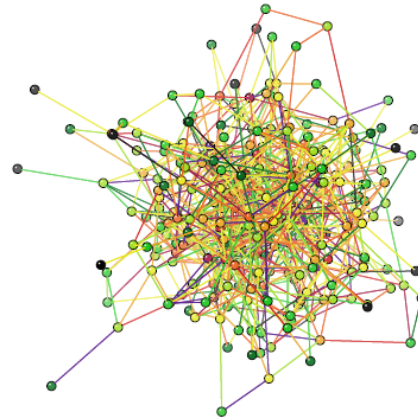


Figure 7.  
Erdős-Rényi  
Cluster Map

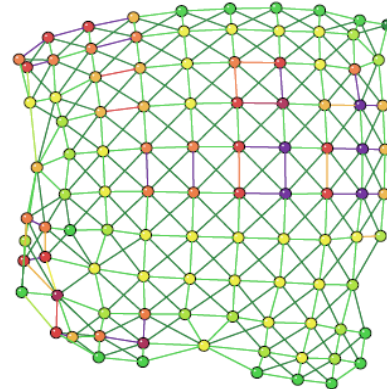


Figure 8.  
PIR Farms  
Cluster Map

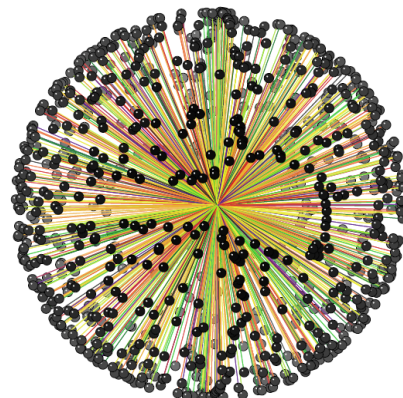
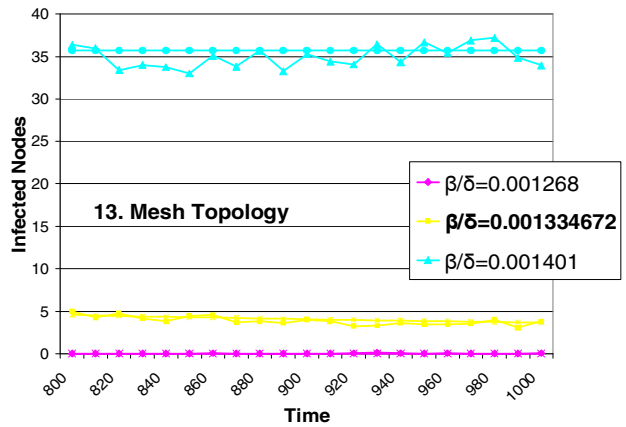
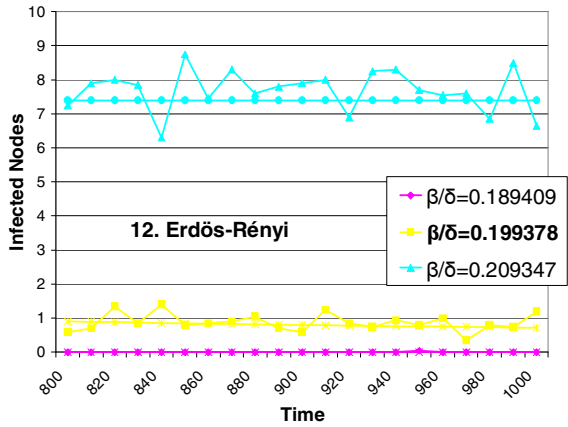
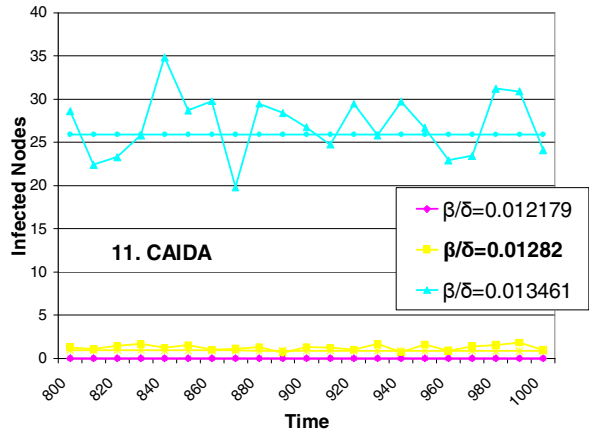
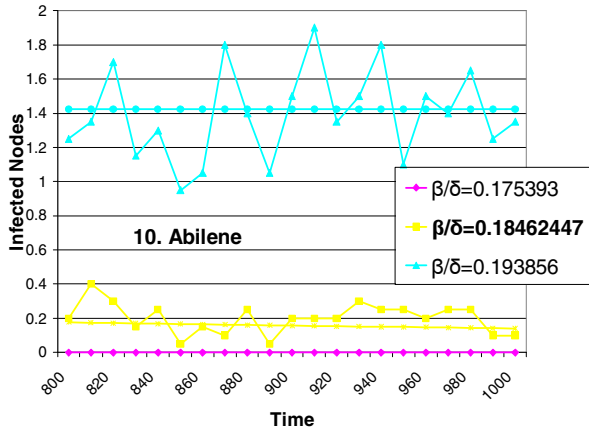


Figure 9.  
Star Topology  
Cluster Map

**Figure 10-13 Thresholds with Weighted Links**

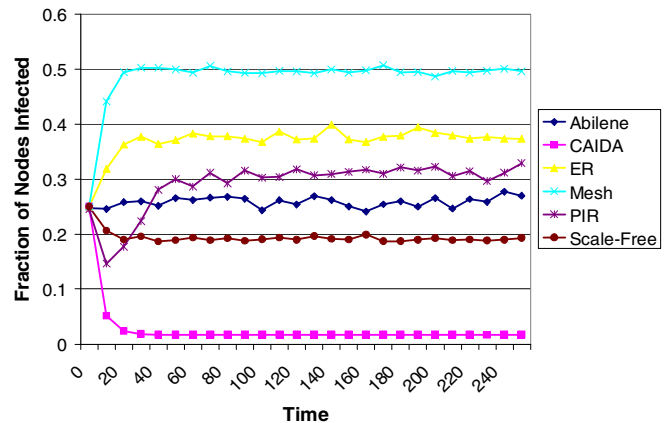


**4. COMPARING TOPOLOGIES**

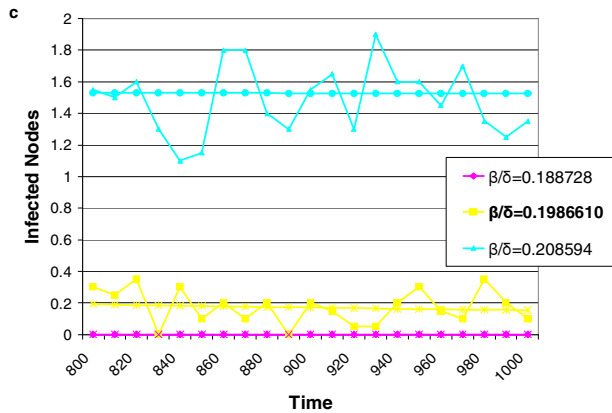
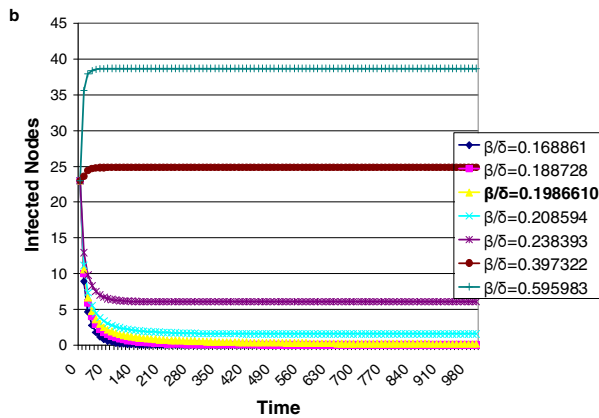
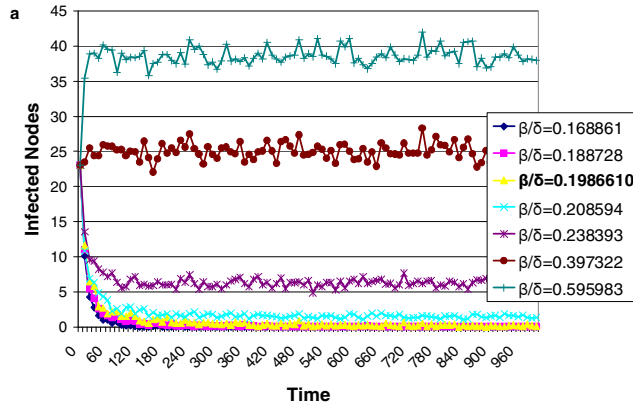
We tested the results of the threshold derived for weighted graphs and can see how the weights affect the graph. We record the eigenvalues of our graphs in the appendix. Pictured in figures 10 through 13 are the results for several threshold simulations. Each data set is averaged over 20 runs for the simulation for 1000 discrete time segments, while the model produces identical results for each of the 20 runs. For the threshold depictions, we show three values of the ratio  $\beta_d/\delta$  with one 5% above the threshold and one 5% below. Each simulation is started with approximately one fourth of the nodes randomly selected to be infected state. They contain the same initial setup as figures 3-5, all with a  $\delta = 0.2$ . These results demonstrate the accuracy of the expanded threshold. While some of the calculated thresholds are lower, seemingly worse, they are merely more realistic. There are significant differences among the networks as the factor of the threshold changes. Figure 14 contains the simulation until  $t = 250$  for 6 of the networks when the  $\beta_d/\delta$  ratio is twice the threshold. They are graphed by the fraction of the nodes infected. For the CAIDA graph, the fraction is very low, seemingly the most robust against a virus. Yet Abilene, which has been designed to be an optimal network, is fairly constant. The Mesh topology is the highest as it intuitively should be. When above the threshold, PIR drops to an unusual and early minimum that is

not seen in the other topologies. This might be an aspect of rural contact networks. It cannot be attributed to the weight distribution because it is found also in the binary simulations seen in figure 4.

We also include an analysis of asymmetric weights on the contact networks. As shown in figures 15 and 16, the asymmetric threshold is as precise and accurate as the previous threshold.

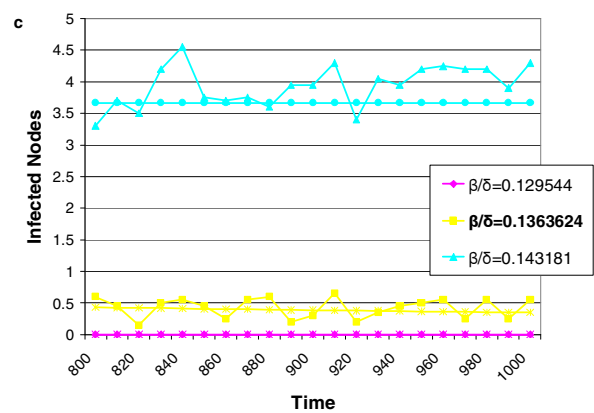
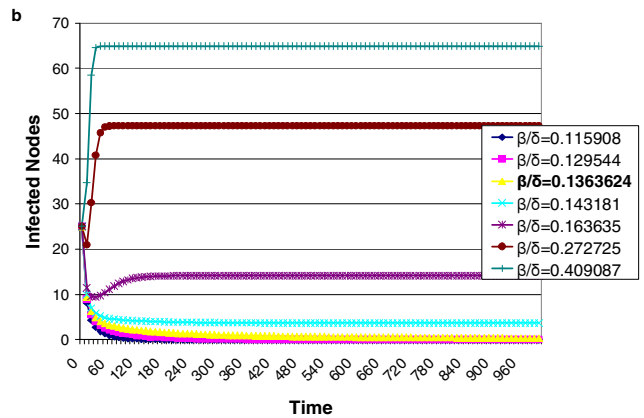
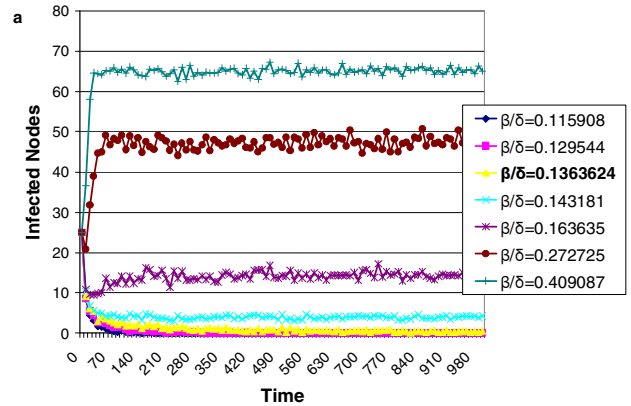


**Figure 14. Fraction of Graph Infected against time for  $\beta_a/\delta = 2/\lambda_{1,W}$**



**Figure 15. Simulation (a), Model (b), and Threshold (c) for an asymmetric Abilene (93) graph**

The asymmetric PIR graph is more robust in the sense that it has a 2.3% and 33% higher epidemic threshold than the **A** and **W'** PIR graphs respectively. Contact weights have a strong ability to alter a graph without changing the topology. When a government wishes to implement an epidemic prevention policy, it can reduce the “weights” and generate similar results without having to either destroy several farms and businesses or having to completely quarantine entire cities.



**Figure 16. Simulation (a), Model (b), and Threshold (c) for an asymmetric PIR (102) graph**

## 5. CONCLUSIONS

Implementing asymmetric weights on a contact network add a significant amount of realism to the contact distribution. They can significantly change your contact network levels of disease resistance. Figures 17 and 18 display the differences among the thresholds of Abilene and PIR, for the different weighting schemes. The SI threshold, while useful for demonstrating the results of weighting methods, is limited in application and is not well suited to be a basis for a simulation. Still, significant differences are readily demonstrated through an SI model,

implying that proper weighting is essential to accurate contact modeling. Interestingly, when the only asymmetrical properties associated with a given contact network are completely reversed, they produce identical thresholds when found by an eigenvalue threshold. So if a rural contact network has only wind as an asymmetrical factor, then wind from the southwest produces an identical threshold as wind from the northeast produces as a property of the transpose of a matrix.

We introduced asymmetric weights with epidemic spreading and verified them with one of the simulation tools we have designed. We defined weight clusters as a crucial element of disease management policy. Future work exists in the area of topology related thresholds for these models, in the study of the effects of the weight distribution and assignment on the epidemic resistance of a graph, and, of course, in optimal universal simulation technologies.

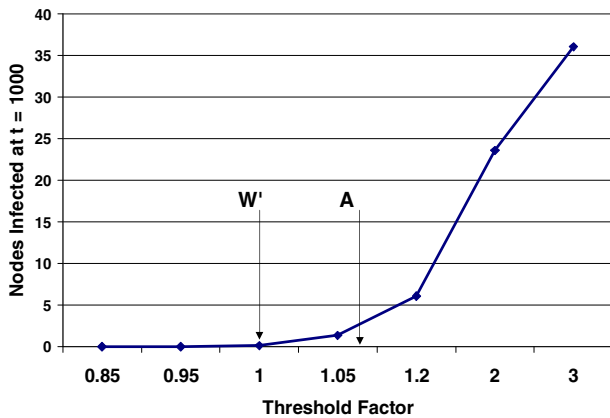


Figure 17. Nodes Infected at Time = 1000 on Abilene W'

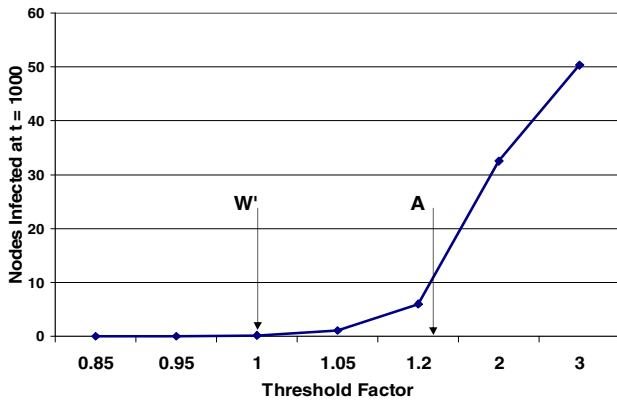


Figure 18. Nodes Infected at Time = 1000 on PIR W'

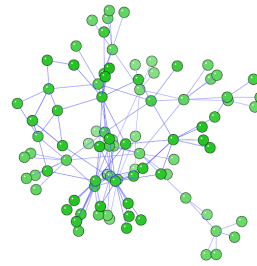
## 6. ACKNOWLEDGMENTS

Our gratitude to the developers of Pajek [5] and KiNG [7] for the use of their innovative software and Dr. Ganesh for his insightful advice.

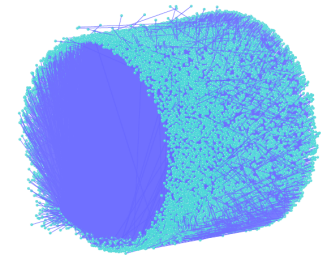
## 7. APPENDIX

Network layouts created by Pajek [5]. Topology figures produced through KiNG imaging from Duke University [7]. Distortion in CAIDA topology is due to the large number of nodes.

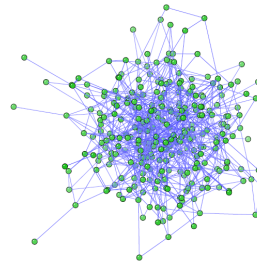
## 7.1 Network Figures



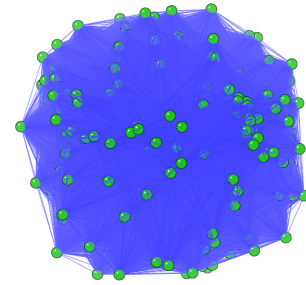
**Abilene Backbone**  
93 Nodes



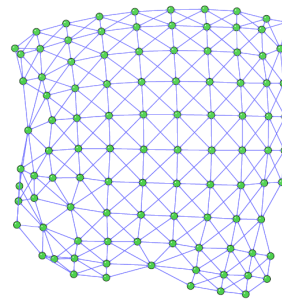
**CAIDA 3/12/2007**  
24,491 Nodes



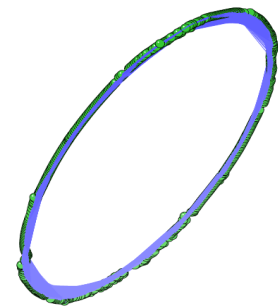
**Erdős-Rényi**  
300 Nodes



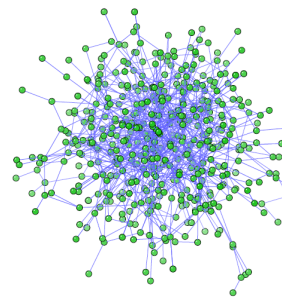
**Mesh Topology**  
750 Nodes



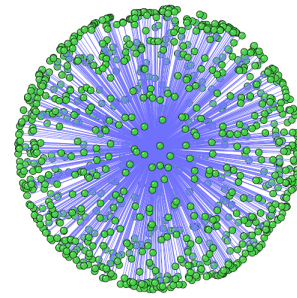
**PIR Farms**  
102 Nodes



**Ring Topology**  
500 Nodes



**Scale-Free Topology**  
500 Nodes



**Star Topology**  
1000 Nodes

## 7.2 Network Eigenvalues

**Table 1. Largest Eigenvalues for each Graph**

Symmetric	Adjacency A	Weighted, Scaled W'
Abilene	5.0156	5.4164
CAIDA	69.50	78.00
Erdős-Rényi	6.4111	6.8166
Mesh	749.00	749.2478
PIR	7.5044	9.7529
Ring	100.00	100.3050
Scale-Free	7.5568	7.9739
Star	31.6070	35.3921
Asymmetric		
Abilene	—	5.0337
PIR	—	7.3334

## 8. REFERENCES

- [1] Abilene Network. Detailed information about the objectives, organization, and development of the Abilene network are available from <http://www.internet2.edu/abilene>.
- [2] Bailey, N. T. J. *The Mathematical Theory of Infectious Diseases and its Applications*. Hafner Press, New York (1975).
- [3] Bapat, R. B. Comparing the Spectral Radii of Two Nonnegative Matrices. *The American Mathematical Monthly* 96, 2 (Feb. 1989): 137-139.
- [4] Barabási, A. and Albert, R. Emergence of Scaling in Random Networks *Science* 286, 509-512, 1999.
- [5] Batagelj, V. and Mrvar, A. Pajek: Large Network Analysis. University of Ljubljana, Slovenia, Information available at <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>.
- [6] Cooperative Association for Internet Data Analysis (CAIDA), AS Relationships. Available at <http://www.caida.org/data/active/as-relationships/>.
- [7] Davis, I. and Chen, V. Kinemage Next Generation. Information available at <http://kinemage.biochem.duke.edu>.
- [8] Duerr, H.P. Schwehm, M. Leary, C. C. De Vlas, S. J. Eichner, M. The impact of contact structure on infectious disease control: influenza and antiviral agents. *Epidemiology and Infection* 135, 1124-1132, 2007.
- [9] EpiSims. Los Alamos National Laboratory. Detailed information about the objectives, organization, and development of EpiSims are available from <http://www.ccs.lanl.gov/ccs5/apps/epid.shtml>
- [10] Erdős, P. and Rényi, A. On Random Graphs *Publicationes Mathematicae*. 6, 290–297, 1959.
- [11] Eubank, S. Scalable, Efficient Epidemiological Simulation. *In Proceedings of the 2002 ACM symposium on Applied computing (SAC '02)* (Madrid, Spain, March 11-14, 2002). ACM Press, New York, NY, 2002, 139-145.
- [12] Ferguson, N. M. Cummings, D. Cauchemez, S. Fraser, C. Riley, S. Meeyai, A. Iamsirithaworn, S. Burke, D. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 437, 209–214, 2005.
- [13] Ford, D., Kaufman, J., Eiron, I. An Extensible Spatial and Temporal Epidemiological Modeling System. *International Journal of Health Geographics*, 5, 4 (Jan. 2006).
- [14] Ganesh, A. Massoulie, L. Towsley, D. The effect of network topology on the spread of epidemics. *IEEE Infocom 2005*, Miami, USA, 2005.
- [15] Li Y, et al. Predicting Super Spreading Events during the 2003 Severe Acute Respiratory Syndrome Epidemics in Hong Kong and Singapore. *American Journal of Epidemiology* 160, 719–728, 2004.
- [16] Longini, Jr., I. Nizam, A. Xu, S. Ungchusak, K. Hanshaworakul, W. Cummings, D. Halloran, M. E. Containing Pandemic Influenza at the Source. *Science* 309, 1083–1087, 2005.
- [17] Meyers, L.A. Newman, M.E. Pourbohloul, B. Predicting Epidemics on Directed Contact Networks. *Journal of Theoretical Biology* 240, 400–418, 2006.
- [18] Newman, M. E. Analysis of Weighted Networks. *Physical Review (E)*, 70, 056131, 2004.
- [19] Newman, M. E. Spread of Epidemic Disease on Networks. *Physical Review (E)*, 66, 016128, 2002.
- [20] Pastor-Satorras, R. Vespignani, A. Epidemic Spreading in Scale-Free Networks *Physical Review Letters*, 86 (14), 3200-3203, April 2001.
- [21] Pourbohloul, B. Meyers, L. A. Skowronski, D. M. Krajden, M. Patrick, D. M. Brunham, R. C. Modeling Control Strategies of Respiratory Pathogens. *Emerging Infectious Diseases* 11, 1249–1256, 2005.
- [22] Wang, Y., Chakrabarti, D., Wang, C., and Faloutsos, C. Epidemic Spreading in Real Networks: An Eigenvalue Viewpoint. *In Proceedings of the Symposium on Reliable Distributed Computing*, 25–34, Florence, Italy, October 2003.