SIR epidemics in dynamic contact networks: Electronic Supplement

Erik Volz * Lauren Ancel Meyers †

June 17, 2007

1 Degree distribution for susceptible nodes

The multivariate generating function for the number of contacts from a susceptible to a node in $\mathcal{S}, \mathcal{I}$, or $\mathcal{R}$ will be a function of three dummy variables, $x_{\mathcal{S}}, x_{\mathcal{I}}$, and $x_{\mathcal{R}}$. Each dummy variable corresponds to the number of contacts to a node of the corresponding type. Assuming each contact from a susceptible node node has independent probabilities p_S and p_I of going to a susceptible or infectious node respectively, and given a degree k node, the composition of contacts will have a multinomial distribution generated by $(x_S p_S + x_I p_I + x_R(1-p_S-p_I))^k$. Then given a probability θ^k of a degree k node remaining susceptible up to a time t, the degree distribution among susceptibles is retrieved by summing over k:

$$\begin{aligned} g_S(x_S, x_I, x_R) &= \sum_k p_k \theta^k (x_S p_S + x_I p_I + x_R(1-p_S-p_I))^k / \sum_k p_k \theta^k \\ &= g(\theta(x_S p_S + x_I p_I + x_R(1-p_S-p_I))/g(\theta) \end{aligned} \tag{1}$$

The mean of the multivariate distribution generated by equation 1 is retrieved by differentiation with respect to the corresponding dummy variable. For example, differentiating with respect to x_I gives the average number of contacts from a susceptible to an infectious node. We denote the average number of contacts among susceptible nodes to nodes of type X as $\delta_S(X)$. For example, the average number of contacts to infectious nodes is

$$\begin{aligned} \delta_S(I) &= [\frac{d}{dx_I} g_S(x_S, x_I, x_R)]_{x_S = x_I = x_R = 1} \\ &= \theta p_I g'(\theta)/g(\theta) \end{aligned} \tag{2}$$

*Department of Integrative Biology, University of Texas-Austin, 1 University Station, C0930, Austin, TX 78712. To whom correspondence should be addressed. E-mail: erik.volz@mail.utexas.edu

†Department of Integrative Biology, University of Texas-Austin, 1 University Station, C0930, Austin, TX 78712.

Usually we don't want to work with the distribution generated by equation 1, but rather with the distribution for excess degree as described in the text. That is, suppose we select an edge between susceptibles and a node of type X, and follow the edge to the susceptible node, whereupon we enumerate the contacts emanating from the node we arrived at. This is equivalent to selecting nodes with probability proportional to $X-$degree, so the distribution is retrieved by weighting each term of the PGF by the corresponding $X-$degree. For example, the excess degree distribution for susceptibles selected with probability proportional to $I-$degree is generated by

$$\begin{aligned} g_{SI}(x_S, x_I, x_R) &= \frac{d}{dx_I} g_S(x_S, x_I, x_R) / [\frac{d}{dx_I} g_S(x_S, x_I, x_R)]_{x_S=x_I=x_R=1} \\ &= g'(\theta(x_S p_S + x_I p_I + x_R(1 - p_S - p_I)))/g'(\theta) \end{aligned} \tag{3}$$

Due to the simplicity of the multinomial distribution of contacts among sets $\mathcal{A}_{SS}, \mathcal{A}_{SI}, \mathcal{A}_{SR}$, we have that $g_{SI}(x_S, x_I, x_R) = g_{SS}(x_S, x_I, x_R) = g_{SR}(x_S, x_I, x_R)$, as can be verified by repeating the above calculation.

As before, the mean of the distribution yields the quantities of interest. We have

$$\begin{aligned} \delta_{S,I}(I) = [\frac{d^2}{dx_I^2} g_S(x_S, x_I, x_R)]_{x_S=x_I=x_R=1} / \\ [\frac{d}{dx_I} g_S(x_S, x_I, x_R)]_{x_S=x_I=x_R=1} = \\ = p_I \theta g''(\theta)/g'(\theta) \end{aligned} \tag{4}$$

and,

$$\begin{aligned} \delta_{S,I}(S) = [\frac{d^2}{dx_S^2} g_S(x_S, x_I, x_R)]_{x_S=x_I=x_R=1} / \\ [\frac{d}{dx_I} g_S(x_S, x_I, x_R)]_{x_S=x_I=x_R=1} = \\ = p_S \theta g''(\theta)/g'(\theta). \end{aligned} \tag{5}$$

If we wish to find the mean excess degree counting contacts to all node-types, we set $x_S = x_I = x_R$ and conduct the calculation as usual.

$$\begin{aligned} \delta_{S,I} = [\frac{d}{dx}[\frac{d}{dx_I} g_S(x_S, x_I, x_R)]_{x_S=x_I=x_R=x}]_{x=1} / \\ [\frac{d}{dx_I} g_S(x_S, x_I, x_R)]_{x_S=x_I=x_R=1} = \\ = \theta g''(\theta)/g'(\theta) \end{aligned} \tag{6}$$

2 Estimation of epidemiological quantities from Atlanta syphillis data

An individual in the study, ego_i, reports m_i contacts over a duration T_i, which is the time interval from his or her first reported contact to the date of the

interview. We use t_{ij} to denote the duration of the j'th contact of individual ego_i, that is, the time between the first and last interaction with that individual. We assume that contacts change at a constant rate ρ, so that the durations t_{ij} should be exponentially distributed with a mean $1/\rho$. We can then estimate the mixing rate ρ as the reciprocal average duration of a sample contact. Let n be the sample size. Then,

$$\hat{\rho} = 1/ < t_{ij} >= n \left(\sum_i \sum_j t_{ij} \right)^{-1} \tag{7}$$

In this sexual network, the characteristic contact duration was $< t_{ij} >= 31$ days, and $\hat{\rho} = .032$.

Given a concurrent degree k_i for individual i, the expected number of unique contacts over a duration T_i is $m_i = \rho k_i T_i$. This can be reversed to estimate the concurrent degree for individual i.

$$\hat{k}_i = \frac{m_i}{\hat{\rho} T_i} \tag{8}$$

It can be difficult to generalize properties of the sample to the population at large for link-tracing sample designs. Nevertheless, suppose the likelihood of being included in the sample depends on the probability of being traced by a sexual contact from someone who is already in the sample. Then as a simple approximation, let the sample inclusion probability of individual i be proportional to k_i. This allows us to estimate the population degree distribution using an RDS-type estimator (refer to citations in text). Let n_k denote the number of sample elements with estimated degree k. Then the probability of an individual having degree k is estimated as

$$\hat{p}_k = \frac{n_k}{k} / \sum_i \frac{n_i}{i}. \tag{9}$$

Assuming a transmissibility of 62.7% and the recovery and mixing rates given above, the transmission rate can be estimated using the equation for τ.

$$\hat{r} = \frac{\tau}{1-\tau}(\hat{\mu} + \hat{\rho}) = 0.0649 \tag{10}$$

We estimate the prevalence of syphilis in the 1996 outbreak using an estimator similar to equation 9, which yields $\hat{J}_\infty$ =35%:

$$\hat{J}_\infty = \sum_j \frac{y_j}{d_j} / \sum_j \frac{1}{d_j} \tag{11}$$

where d_j is the degree of sample unit j and $y_j = 1$ if unit j is infected and $y_j = 0$ otherwise.

Figure 1 shows a colorized version of the final-size diagram found in the text.

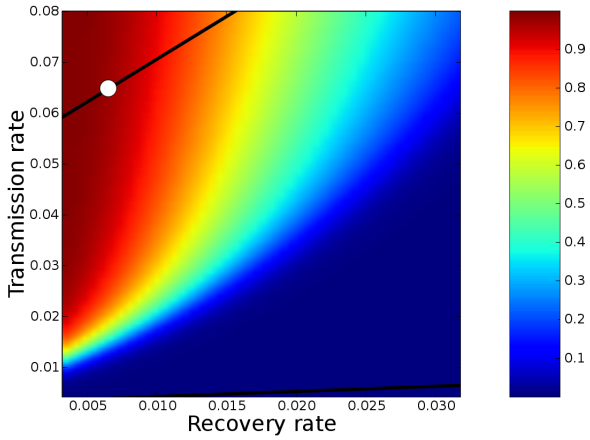

Figure 1: The final epidemic size as predicted by the NE model is shown with respect to the transmission rate r and recovery rate μ for the Atlanta syphilis data. Lighter colors correspond to larger final size, as given by the color bar on right. The thick black line corresponds to the ratio r/μ that gives the expected transmissibility of $\tau = 0.627$. The large data point indicates the expected recovery rate of $\mu = 1/154$.

3 Dynamics of p_S

$$\dot{p}_S = \frac{d}{dt}\frac{M_{\mathcal{SS}}}{M_{\mathcal{S}}} = \frac{\dot{M}_{\mathcal{SS}}}{M_S} - \frac{M_{\mathcal{SS}}\dot{M}_S}{M_S^2} \tag{12}$$

As $-\dot{S}$ nodes become infectious a number of contacts proportional to $-\dot{S}\delta_{S,I}(S)$ move from set $\mathcal{A}_{\mathcal{SS}}$ to $\mathcal{A}_{\mathcal{SI}}$. Thus

$$\dot{M}_{\mathcal{SS}} = -2 \times (-\dot{S})\delta_{S,I}(S)/g'(1) \tag{13}$$

where the factor of two accounts for two contacts corresponding to each edge between two susceptible nodes. Combining equation 13 with the equations for M_S and $\dot{M}_S$ yields the result reported in main text.

4 Comparison of deterministic NE model to related MA and static network models

A natural question is whether there already exist simpler models that can adequately reproduce the important features of the NE model (table 4, main text). In fact, there are static network models that reasonably approximate the final size of an epidemic in the NE model, but cannot address epidemiological dynamics. Standard compartmental SIR models, however, cannot reproduce any of the important aspects of the NE model dynamics.

Figure 2 shows epidemic final size predicted by the deterministic NE model as the mixing rate ρ is varied, as well as by three other models: a standard compartmental SIR model (i.e. a mass-action model) and two static network approximations. The mass action model is based on the common approximation

$$\dot{S} = -\mu R_0 SI \tag{14}$$

$$\dot{I} = \mu R_0 SI - \mu I. \tag{15}$$

In this approximation, R_0 is calculated as

$$R_0 = \frac{\tau}{\mu}\left(\frac{g''(1)}{g'(1)}(\mu+\rho)+\rho\right) \tag{16}$$

The derivation of this value of R_0 is found in a paper currently in preparation [4]. Since this approximation presupposes no network structure at all, it fails to reproduce NE dynamics for any level of random mixing.

The two static network approximations are also derived in the manuscript in preparation [4]. Given a static network, bond percolation methods can be used to calculate the final size [Newman(2002)]. These values are graphed in figure 2. The "Static net-concurrent" approximation naively assumes the concurrent degree distribution and therefore does not take random mixing into account at all. This approximation matches the NE model at $\rho = 0$, but quickly diverges. The "Static net-implicit" approximation assumes that each node has degree equal to its expected number of contacts during an infectious period. This approximation much more closely tracks the NE model over the entire range of ρ.

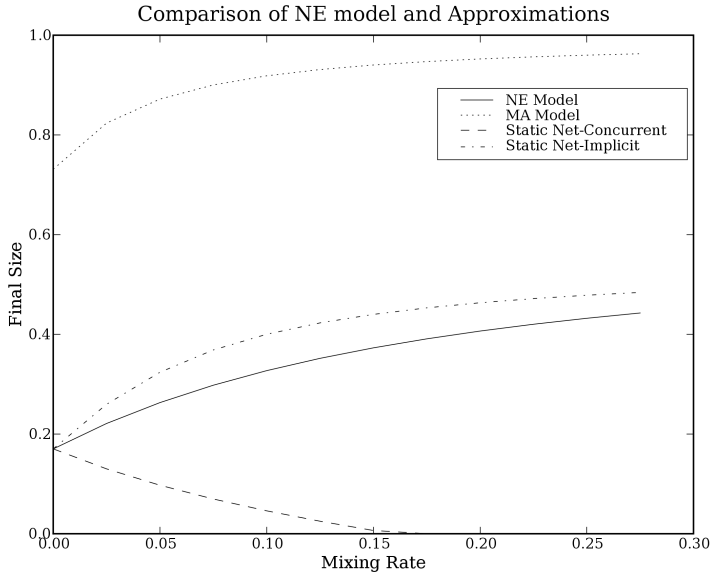

Figure 2: Comparison of final epidemic size versus the mixing rate ρ. The deterministic NE model is compared to two static network approximations and a simple compartmental SIR model (MA).

5 Comparison of deterministic NE model to PA model

Pair approximation (PA) [] is an alternative framework for modeling epidemic dynamics in random networks. A network consists of nodes with various degrees and edges between nodes. PA works by explicitly modeling the number of degree k nodes in each epidemic state for degree k, as well as the number of edges between each possible pair. PA is highly intuitive and flexible, and has been used to model a variety of complex epidemiological scenarios. Nevertheless, PA has some disadvantages relative to our methods. PA is high-dimensional, with the number of dynamic variables increasing as $O(n^2)$ with n distinct degree classes. In our experience, this can cause numerical instability of solutions.

For example, consider the power law degree distribution used in the main text. This distribution was truncated at $k_{max} = 75$. PA requires a dynamic variable for every degree value 1-75 for each state S, I, R.

- S_n : 75 variables for susceptibles of degree n
- I_n : 75 variables for infecteds of degree n
- $[S_n S_m]$: $75^2 = 5625$ variables for number of pairs between susceptibles
- $[S_n I_m]$: $75^2 = 5625$ variables for number of pairs between susceptibles and infecteds
- $[I_n I_m]$: $75^2 = 5625$ variables for number of pairs between infecteds

Thus the total number of variables in this PA model is 17025. In contrast, the NE model will always have four dynamic variables. When $\rho = 0$, this reduces to three variables; and in the limit of $\rho \to \infty$, it reduces to two variables.

The predictions of our model are fairly consistent with those made by comparable PA models. In figure 3 we have compared the trajectories of identical SIR epidemics as solved by the PA model and NE model. To make the comparison as fair as possible, the same integration method (Runge-Kutta) and time step (0.01) was used for both models. The degree distribution was Poisson ($\lambda = 2.5$) truncated at $k_{max} = 8$. Unfortunately, our implementation of PA became unstable for maxiumum degree $k_{max} > 8$. The final size is very close, but, there is some discrepancy in the predicted trajectories. This may result partly from the forward-looking integration method, which was chosen for its simplicity.

6 Robustness of NE model

Most of the results presented in this manuscript assume parameter values that yield R_0 well above unity. Many models that behave well at high values of R_0, show noticeable errors near the epidemic treshold. We have therefore compared the NE model to PA over a range of transmission rates. As a baseline, we

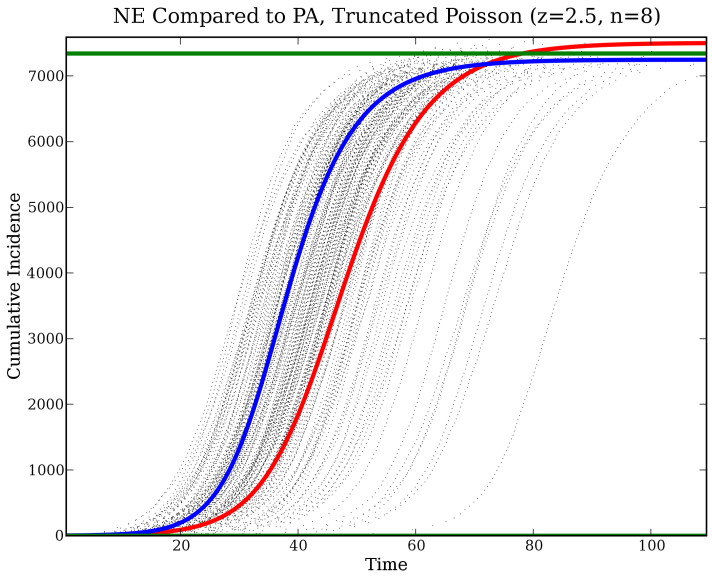

Figure 3: Prevalence is shown versus time for 100 simualtions (dots), the NE model(blue), and the PA model (red). The green line shows the expected final size based on bond-percolation. The degree distribution is Poisson (2.5) truncated at 8. $\mu = .1, r = .2$.

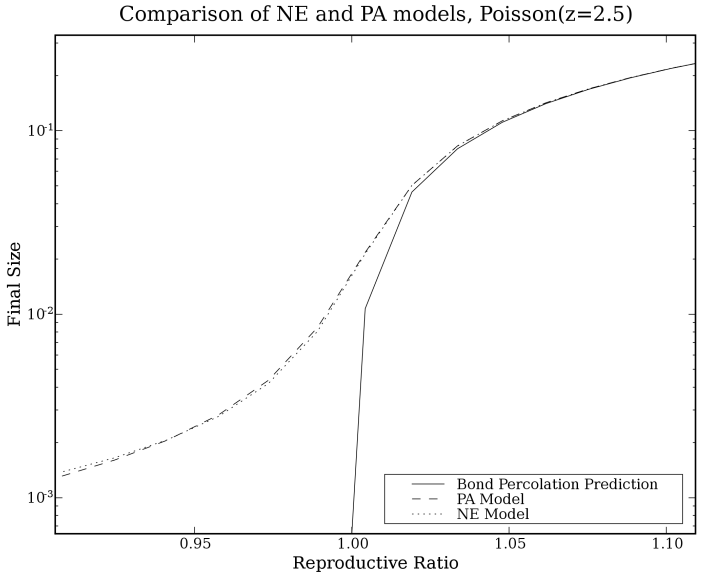

Figure 4: The final size as predicted by the NE and PA models is compared to the bond-percolation solution over a range of R_0. The degree distribution is Poisson(2.5) truncated at 6. $\mu = .1$. r is varied to give the corresponding R_0 value.

have computed the final size (FS) using bond-percolation techniques, which are known to give asymptotically correct results.

Figure 4 shows that the final size as predicted by both the NE model and the PA model are in close agreement with BP. Although it is difficult to draw generalizations from this experiment, PA is closer to BP below the epidemic threshold, and NE is closer to BP above it. NE and PA diverge around $R_0 = 1$. Some discrepancy is inevitable given that the initial conditions $\epsilon = 1-^{-4}-$ FS can never be less than ϵ, and both NE and PA converge to ϵ as $R_0 \to 0$. It is likely that the discrepancy between NE, PA, and BP could be made arbitrarily small for both NE and PA by 1. decreasing ϵ, 2. using a smaller time step for integration, and 3. using a better integration method (Runge-Kutta).

6.1 Clustering

The NE model applies to a simple class of random networks that do not include transitivity (clustering) or assortativity. This new framework is, however, quite flexible and can be extended to consider more complex network structures in the future.

To address this possible limitation of the model, we have evaluated its per-

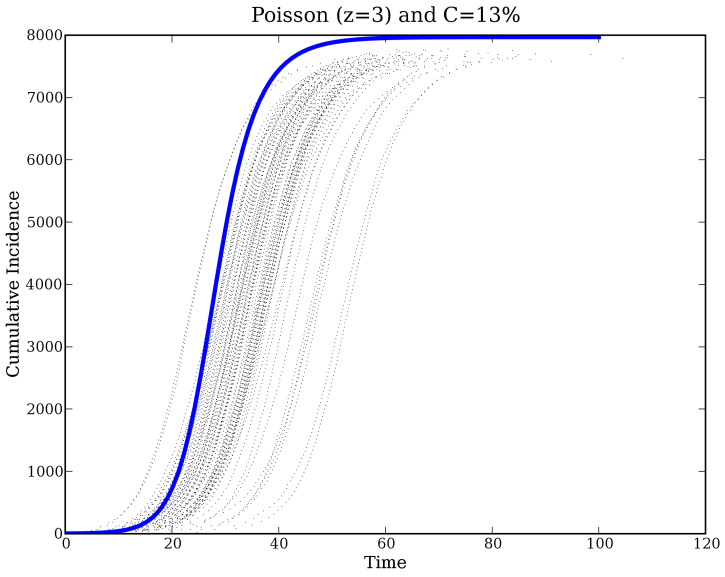

Figure 5: The NE model is compared to several hundred simulated SIR epidemics in networks with clustering of approximately 13%. The degree distribution is Poisson (3). $r = .2, \mu = .1, \rho = 0$.

formance on networks with non-trivial amounts of clustering. We conducted SIR simulations on static networks with variable amounts of clustering generated using the algorithm in [3], and compared the results to solutions of the NE model with $\rho = 0$.

Figure 5 compares simulations to the NE model at $C = 13\%$, where C is the clustering coefficient [3]. This level of clustering produces a noticable error in the NE model, which may not be acceptable for some applications. Figure 6 shows simulations and the NE model at $C = 3.5\%$. There is no noticable discrepancy in this case.

Fortunately, the NE model is applicable to a number of populations that are likely to have low clustering. Many sexual networks have very low clustering (¡4%). Heterosexual populations have no transitive closure at all, by virtue of their bipartite structure. In that case, one can consider the a measure similar to the clustering coefficient, namely the frequency of 4-cycles.

A study [2] of 250 MSMs in Colorado Springs revealed a clustering coefficient of 2.98%. A study of a heterosexual population [5, 2] with 82 people found that the frequency of 4-cycles was 0.486%. And, a study using data from Add Health [1] showed that when individuals choose sex-partners, they tend to avoid former or current partners of their friends. This prevents the creation of short

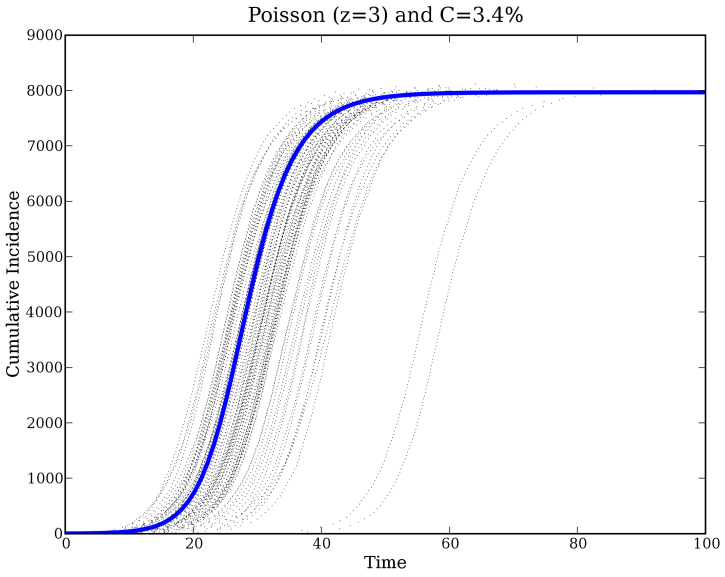

Figure 6: The NE model is compared to several hundred simulated SIR epidemics in networks with clustering of approximately 3.4%. The degree distribution is Poisson (3). $r = .2, \mu = .1, \rho = 0$.

cycles (e.g. triads and 4-cycles) or cliques. Since network models are most easily applied to sexual and drug-use networks, the NE model should be suitable for a large range of applications.

References

[1] Bearman, P.S., Moody, J., Stovel, K. 2004. Chains of affection: The structure of adolescent romantic and sexual networks. American Journal of Sociology, 110,1,44–91.

[2] Lind, P.G., Gonzalez, M.C., Herrman, H.J. 2005. Cycles and clustering in bipartite networks. Phys. Rev. E, 72, 56127.

[Newman(2002)] Newman, M., 2002. The spread of epidemic disease on networks. Phys. Rev. E 66, 016128.

[3] Volz, E. 2004. Random networks with tunable degree distribution and clustering. 70, 56115.

[4] Volz, E., Meyers, L.A. 2007. Static network approximations for SIR epidemics in dynamic contact networks. In preparation.

[5] J. L. Wylie and A. Jolly, Sex Transm. Dis. 28, 14 2001.