

Predicting the size and probability of epidemics in a population with heterogeneous infectiousness and susceptibility

Joel C. Miller

*Mathematical Modeling & Analysis Group and Center for Nonlinear Studies, MS B284, Los Alamos National Laboratory**

(Dated: February 6, 2007)

We analytically address disease outbreaks in large, random networks with heterogeneous infectivity and susceptibility. The transmissibility T_{uv} (the probability that infection of u causes infection of v) depends on the infectivity of u and the susceptibility of v . Initially a single node is infected, following which a large-scale epidemic may or may not occur. We use a generating function approach to study how heterogeneity affects the probability that an epidemic occurs and, if one occurs, its attack rate (the fraction infected). For fixed average transmissibility, we find upper and lower bounds on these. An epidemic is most likely if infectivity is homogeneous and least likely if the variance of infectivity is maximized. Similarly, the attack rate is largest if susceptibility is homogeneous and smallest if the variance is maximized. We further show that heterogeneity in infectious period is important, contrary to assumptions of previous studies. We confirm our theoretical predictions by simulation. Our results have implications for control strategy design and identification of populations at higher risk from an epidemic.

The spread of infectious disease is a problem of great interest [1]. Much work has focused on how diseases spread in networks of human, animal, or computer interactions [2, 3, 4, 5, 6, 7, 8, 9]. The transmissibility, the probability that an edge transmits infection, has a network-dependent threshold (which can be zero) corresponding to a second order phase transition above which an epidemic may happen and below which epidemics are not found. Ideally an intervention reduces the transmissibility or modifies the network to raise the threshold so that epidemics cannot occur. Most study has focused on determining the threshold value under varying assumptions [6, 7, 8, 9, 10, 11] in order to design an optimal intervention.

For many diseases and networks, it is impractical to reduce transmissibility sufficiently to eliminate the possibility of an epidemic. An intervention strategy should therefore optimize competing goals: minimize social cost, reduce the probability a large-scale epidemic occurs, and reduce the attack rate (fraction infected) if an epidemic does occur. Recently the probability and attack rate have been investigated [2, 3, 4, 12, 13], but none of these has systematically investigated the effect of heterogeneity in transmissibilities. Heterogeneities can result from variations in the application of interventions or from natural differences in the population such as variation in recovery time. It is often assumed that this special case can be mapped without loss of generality to recovery of all individuals after a single time step [4, 6, 7, 8, 9, 14] and so the number of new cases from a single case is distributed binomially. However, it may be inferred from [15] that this assumption is false. We have recently become aware of independent work [16] which shows that recovery time heterogeneity reduces the epidemic probability, but has no effect on the attack rate. In this Letter, we consider how generic heterogeneities affect the epidemic probability and the attack rate if an epidemic occurs.

The epidemics we study spread on random networks of N nodes with degree distribution given by $P(k)$ where k is the degree. We use the SIR model [1]: nodes are divided into susceptible, infectious, and recovered classes. We modify the model to include heterogeneities. An infectious node u with infectivity \mathcal{I}_u connected to a susceptible node v with susceptibility \mathcal{S}_v infects v with probability equal to the transmissibility $T_{uv}(\mathcal{I}_u, \mathcal{S}_v)$ of the edge. Infectious nodes recover and are no longer susceptible. The outbreak begins with a single infection (the *index case*) which spreads to neighboring nodes. If an epidemic occurs, the eventual number infected is $\mathcal{O}(N)$, otherwise the outbreak is localized. \mathcal{I} and \mathcal{S} can be quite arbitrary, *e.g.*, \mathcal{I} may be a vector representing time of infection, level of virus shedding, frequency of handwashing, etc. The form of T_{uv} can also be quite general: it need only be integrable and bounded in $[0, 1]$.

The spread of an epidemic on a network with heterogeneous infectivity and susceptibility is equivalent to a special case of directed percolation for which the probability of retaining an edge depends on both the base and target node. In this formalism, infection spreads to the out-component of the index case [2, 3]. If the disease has sufficiently high transmissibility a single giant strongly connected component G_{scc} exists [17], occupying a fixed fraction of the network as $N \rightarrow \infty$. The set of nodes not in G_{scc} , but from which G_{scc} can be reached is denoted G_i , while the set of nodes not in G_{scc} but reachable from G_{scc} is denoted G_o [i for ‘in’ and o for ‘out’] as demonstrated in figure 1. If the index case is in $G_i \cup G_{scc}$ an epidemic occurs, infecting all of $G_o \cup G_{scc}$ and very few other nodes. In the limit $N \rightarrow \infty$ the probability of an epidemic is the probability the index case is in $G_i \cup G_{scc}$ and the attack rate is the fraction of nodes in $G_o \cup G_{scc}$ [23]. We use \mathcal{P} and \mathcal{A} to denote the limiting epidemic probability and attack rate. In general the sizes of G_i and G_o may differ significantly so $\mathcal{P} \neq \mathcal{A}$.

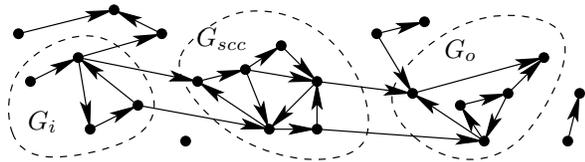


FIG. 1: Schematic representation of G_i , G_{scc} , and G_o . All nodes in G_{scc} can reach any other node in G_{scc} .

This contrasts with the case of homogeneous transmissibility where the problem can be mapped to undirected bond percolation [4, 18, 19] and $\mathcal{P} = \mathcal{A}$.

We develop a general theory to find \mathcal{P} allowing both infectivity and susceptibility to be heterogeneous. Generating function approaches [20] have been used to study disease spread both inside the body [21] or in society [3, 4]. We modify these approaches to calculate \mathcal{P} based on the distribution of \mathcal{I} and \mathcal{S} . Holding the average transmissibility fixed, we then use Jensen's inequality to find distributions which give upper and lower bounds on \mathcal{P} . Because G_o and G_i interchange roles if edge directions are reversed, \mathcal{A} is calculated in the same manner. Our predictions are confirmed through simulations on a large Erdős-Rényi network.

Each node u has an infectivity \mathcal{I}_u and a susceptibility \mathcal{S}_u chosen from independent distributions given by $P(\mathcal{I}_u)$ and $P(\mathcal{S}_u)$. Given the infectivity \mathcal{I}_u of u , the relation $T_{uv}(\mathcal{I}_u, \mathcal{S}_v)$, and the distribution $P(\mathcal{S})$, we define the out-transmissibility of u as

$$T_o(u) = \int T_{uv}(\mathcal{I}_u, \mathcal{S}_v) P(\mathcal{S}_v) d\mathcal{S}_v. \quad (1)$$

From (1) and the distribution $P(\mathcal{I})$, we know the distribution $P_o(T_o)$. We similarly define the in-transmissibility T_i and its distribution $P_i(T_i)$. P_o and P_i must yield the same average, but not all pairs P_o and P_i with the same average are consistent. For each P_o there exists at least one P_i and *vice versa*. Henceforth we consider just P_i and P_o , and do not use $P(\mathcal{S})$ and $P(\mathcal{I})$.

We choose the index case u_0 uniformly from the population. We classify infected cases by their generation, measuring the number of infectious contacts in the chain between them and u_0 (generation 0). We note that the generation time need not be fixed: generations may overlap in time, changing the temporal dynamics but not affecting our results.

Our class of random networks is defined by the Molloy-Reed algorithm [22]. Short cycles are rare. The neighborhood of u_0 is tree-like on successively longer length-scales as $N \rightarrow \infty$. Consequently, \mathcal{P} equals the probability that the transmission chains in an infinite tree are infinite.

We define a probability generating function $f(x)$ for the number of infected nodes in generation 1:

$$f(x) = p_0 + p_1 x + \dots + p_j x^j + \dots,$$

where p_j is the probability that the index case directly infects j neighbors. The index case has degree k with probability $P(k)$ and thus p_j is given by

$$p_j = \sum_{k=j}^{\infty} P(k) \int_0^1 \text{Bi}(k, j, T_o) P_o(T_o) dT_o,$$

where $\text{Bi}(k, j, T_o)$ is the likelihood of j successful trials from k attempts, each with probability T_o . Note that p_j depends on the distribution P_o but not P_i .

In subsequent generations, the probability that a node is infected is proportional to its degree. Early in the epidemic an infected node with degree k has $k-1$ susceptible neighbors because the source of its infection cannot be re-infected. As such the probability q_j that this individual infects j neighbors is

$$q_j = \frac{1}{\langle k \rangle} \sum_{k=j+1}^{\infty} k P(k) \int_0^1 \text{Bi}(k-1, j, T_o) P_o(T_o) dT_o.$$

where $\langle \cdot \rangle$ denotes the expected value. We let $h(x) = \sum q_j x^j$ be the generating function for the number of new cases caused by a non-index case. The generating function for the number of infections caused by n non-index cases is $[h(x)]^n$. Consequently it may be shown that the generating function for the number of infections in generation $g > 0$ is given by

$$f(h^{g-1}(x)),$$

where h^{g-1} denotes composition of h with itself $g-1$ times. For later use we rearrange f and h as

$$f(x) = \int_0^1 P_o(T_o) \sum_{k=0}^{\infty} [1 + T_o(x-1)]^k P(k) dT_o, \quad (2)$$

$$h(x) = \int_0^1 \frac{P_o(T_o)}{\langle k \rangle} \sum_{k=1}^{\infty} [1 + T_o(x-1)]^{k-1} k P(k) dT_o. \quad (3)$$

The extinction probability is $\lim_{g \rightarrow \infty} f(h^{g-1}(0))$. To calculate this, we find $\lim_{g \rightarrow \infty} h^{g-1}(0)$ which is a solution to $x = h(x)$. At most two solutions exist in the interval $[0, 1]$, one of which is $x = 1$. If no other solution exists then $x = 1$ is a stable fixed point and $\mathcal{P} = 0$. Otherwise the iteration converges to $x_0 < 1$ and

$$\mathcal{P} = 1 - f(x_0).$$

Because f and h are independent of P_i , \mathcal{P} is unaffected by heterogeneities in susceptibility.

We now seek distributions P_o maximizing or minimizing \mathcal{P} subject to $\langle T \rangle = T^*$. In their investigation of recovery time heterogeneities, [16] showed that the probability is maximized if recovery times are identical. We generalize this to arbitrary sources of heterogeneities in

infectivity, using a similar proof. For notational convenience we use $\delta^*(T)$ to denote the δ -function $\delta(T - T^*)$, set

$$\hat{h}(T, x) = \frac{1}{\langle k \rangle} \sum_{k=1}^{\infty} [1 + T(x-1)]^{k-1} k P(k),$$

and rewrite (3) to explicitly show that h depends on P_o

$$h[P_o](x) = \int_0^1 \hat{h}(T_o, x) P_o(T_o) dT_o.$$

We similarly define $f[P_o](x)$. Because \hat{h} is a convex function of T , Jensen's inequality shows $P_o = \delta^*$ minimizes $h[P_o](x)$. We denote the smallest root of $x = h[\delta^*](x)$ by x_1 . For $x < x_1$ and any P_o , we have $x < h[\delta^*](x) \leq h[P_o](x)$. Thus the root x_0 of $x = h[P_o](x)$ satisfies $x_1 \leq x_0$, so x_0 is minimized if $P_o = \delta^*$.

Similar calculations show $f[\delta^*](x) \leq f[P_o](x)$ for all P_o . Further, $f[\delta^*](x)$ is an increasing function of x . Thus the extinction probability $f[P_o](x_0)$ is minimized by $P_o = \delta^*$. So homogeneous infectivity maximizes \mathcal{P} .

In addition, we find a new lower bound. Jensen's inequality also implies that fixing $\langle T \rangle = T^*$ but increasing $\langle T^2 \rangle$ reduces \mathcal{P} . Consequently, \mathcal{P} is minimized by $P_o(T_o) = (1 - T^*)\delta(T_o) + T^*\delta(T_o - 1)$.

Thus we have shown that an epidemic is most likely if T_o is homogeneous and least likely if its variance is maximized. Analogously the attack rate is largest if T_i is homogeneous, and smallest if its variance is maximized.

We expect that a threshold value of $\langle T \rangle$ exists above which epidemics can occur [$x = h(x)$ has two roots] and below which they cannot. Allowing $\langle T \rangle$ to vary by continuously changing P_o , the fixed point $x = 1$ of $x = h(x)$ bifurcates into two when $h'(1) = 1$. We find

$$h'(1) = \frac{\langle T \rangle \langle k^2 - k \rangle}{\langle k \rangle}.$$

So the epidemic threshold is $\langle T \rangle = \langle k \rangle / \langle k^2 - k \rangle$, generalizing the results of [3, 4].

We confirm our predictions by comparison with simulations on an Erdős-Rényi network with 100000 nodes and $\langle k \rangle = 4$. As $N \rightarrow \infty$, Erdős-Rényi networks with fixed average degree have Poisson degree distribution and

$$h(x) = f(x) = \int_0^1 \exp[\langle k \rangle T_o (x-1)] P_o(T_o) dT_o.$$

For our first comparison, we consider the effect of varying infection time. We discretize time, and take different models of recovery time given in the caption of figure 2. For each time step, the probability of infecting a susceptible neighbor is p , and so an infected individual with recovery time τ has $T_o = 1 - (1 - p)^\tau$. As a reference we consider the case where all infectious individuals recover after exactly five time steps. We vary p

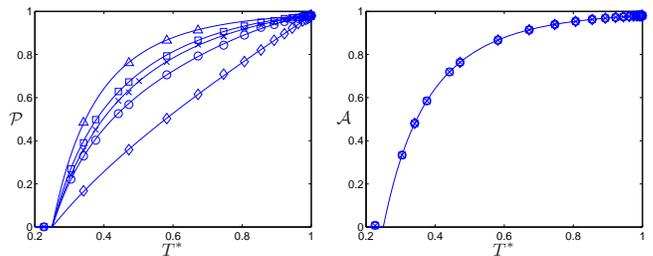


FIG. 2: Comparison of theory (lines) with simulation (symbols). For the different distributions of infectivity (with susceptibility constant), \mathcal{P} changes, but \mathcal{A} does not. We use constant recovery time $\tau = 5$ (\triangle), $\tau = 0$ or ∞ (\diamond), $\tau = 2$ or 8 (\square), $\tau = 1$ or 10 (\circ), and finally a constant recovery rate (\times).

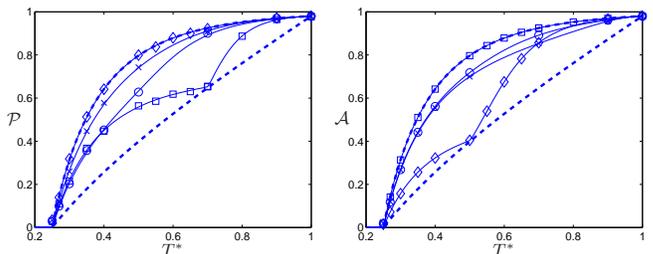


FIG. 3: Comparison of theory (curves) with simulation (symbols) for $T_{uv} = 1 - \exp(-\alpha \mathcal{I}_u \mathcal{S}_v)$. The theoretical bounds are in dashed bold. The distributions are \diamond : $P(\mathcal{I}) = \delta(\mathcal{I} - 1)$, $P(\mathcal{S}) = 0.5\delta(\mathcal{S} - 0.001) + 0.5\delta(\mathcal{S} - 1)$; \times : $P(\mathcal{I}) = 0.5\delta(\mathcal{I} - 0.3) + 0.5\delta(\mathcal{I} - 1)$, $P(\mathcal{S}) = 0.2\delta(\mathcal{S} - 0.1) + 0.8\delta(\mathcal{S} - 1)$; \circ : $P(\mathcal{I}) = 0.5\delta(\mathcal{I} - 0.1) + 0.5\delta(\mathcal{I} - 1)$, $P(\mathcal{S}) = 0.2\delta(\mathcal{S} - 0.1) + 0.8\delta(\mathcal{S} - 1)$; \square : $P(\mathcal{I}) = 0.3\delta(\mathcal{I} - 0.001) + 0.7\delta(\mathcal{I} - 1)$, $P(\mathcal{S}) = \delta(\mathcal{S} - 1)$.

in order to change the average transmissibility T^* . The fraction of nodes with each recovery time is chosen such that $\sum P(\tau)[1 - (1 - p)^\tau] = 1 - (1 - p)^5 = T^*$.

The results of several examples are shown in figure 2. Each data point represents 10000 simulations. Away from the epidemic threshold, there is a clear distinction between an epidemic and a non-epidemic outbreak. For definiteness, we define an epidemic to occur if over 500 nodes are infected. Theory and simulations are in good agreement. The upper bound for epidemic probability is realized by the case where all infections last exactly five time steps. The lower bound is realized by the case where some infections last forever and infect all neighbors, while the rest recover before infecting anyone. Because susceptibility is homogeneous, \mathcal{A} does not vary.

For our second comparison we perform calculations for systems with both \mathcal{I} and \mathcal{S} heterogeneous. We use the same Erdős-Rényi network, but assume that $T_{uv} = 1 - \exp[-\alpha \mathcal{I}_u \mathcal{S}_v]$ where the distributions of \mathcal{I} and \mathcal{S} are fixed and α is varied to tune T^* . We find good agreement between theory and simulations in figure 3.

We have shown in this Letter that the effect of general heterogeneity in infectivity and susceptibility on epidemic probability \mathcal{P} and attack rate \mathcal{A} can be accurately

modeled using a generating function approach. We find that \mathcal{P} and \mathcal{A} may differ substantially. We have further shown that heterogeneity in recovery time has a significant effect on \mathcal{P} and cannot be ignored.

For fixed average transmissibility we have found upper and lower bounds for both \mathcal{P} and \mathcal{A} . Further we have found distributions realizing these bounds. For fixed average transmissibility, increasing the variance of P_o decreases \mathcal{P} and increasing the variance of P_i decreases \mathcal{A} .

These results can be used to assist in designing control strategies. For example, if choosing between a strategy that reduces infectivity or susceptibility by half for all of the population or one that reduces infectivity or susceptibility completely for half the population, it is better to choose the latter. As another example, consider a strategy which attempts to locate and isolate infecteds compared with a strategy which attempts to provide susceptible individuals with protection. Both may be affected by inability to reach everyone. The first strategy has a heterogeneous impact on infectivity, while the second strategy has a heterogeneous impact on susceptibility. If the strategies have the same average impact on T then the first reduces the probability of an epidemic more while the second reduces its size more. Which strategy is optimal depends on whether the outbreak is small enough that an epidemic can be prevented.

Our results can also be used to identify populations most at risk from epidemics. Populations with low genetic diversity are already known to be at particularly high risk from an outbreak because the lack of heterogeneity allows the transmissibility to be higher. However, our results show that even for a fixed average transmissibility, a population with lower genetic variation will be more severely affected by a disease.

For heterogeneous infectivity but homogeneous susceptibility, Newman [4] anticipated that \mathcal{A} follows from the formulae derived under the assumption of homogeneous T . He did not address the effect on \mathcal{P} . We have shown that \mathcal{A} is independent of heterogeneity in infectiousness, and so for this special case the prediction is valid. However, it fails if susceptibility is also heterogeneous.

The theory developed here can be generalized in a number of ways. Most simply, we can introduce edge weights to represent some details of the contact between u and v . The same theory will hold, but the calculation of T_o and T_i as in (1) must incorporate the edge weight distribution. We can also introduce correlations between the distributions of \mathcal{I} , \mathcal{S} , and k in an individual without significant theoretical difficulties, though the conclusions may change. It is more complicated to introduce correlations of \mathcal{I} , \mathcal{S} , or k between neighbors.

We have assumed that the network has few short cycles. More realistic social network models incorporate significant clustering. However, at high transmissibilities, if any neighbors are infected, an epidemic is very likely. \mathcal{P} is close to the probability that the initial node

infects any neighbors and loops may be ignored. At low transmissibilities loops are not traced out by the infection and again may be neglected. Loops affect our results only at intermediate transmissibilities. The generating function approach becomes difficult because even early in an outbreak an infected node may have multiple infected neighbors.

This work was carried out under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy at Los Alamos National Laboratory under Contract No. DE-AC52-06NA2539. The author thanks Lindi Wahl, J. Mac Hyman, Anja C. Slim, Luís M. A. Bettencourt, Eduardo López, Shweta Bansal, and Lauren A. Meyers for useful discussions.

-
- * Electronic address: jomiller@lanl.gov; Los Alamos Report LA-UR-06-8193
- [1] R. M. Anderson and R. M. May, *Infectious Diseases of Humans* (Oxford Science Publications, 1991).
 - [2] L. A. Meyers, M. Newman, and B. Pourbohloul, *Journal of Theoretical Biology* **240**, 400 (2006).
 - [3] L. A. Meyers, *Proceedings of the American Mathematical Society* **44**, 63 (2007), ISSN 0002-9939.
 - [4] M. E. J. Newman, *Physical Review E* **66** (2002).
 - [5] R. R. Kao, L. Danon, D. M. Green, and I. Z. Kiss, *Proceedings of the Royal Society B* **274**, 1999 (2006).
 - [6] N. Madar, T. Kalisky, R. Cohen, D. ben Avraham, and S. Havlin, *The European Physical Journal B* **38**, 269 (2004).
 - [7] M. Serrano and M. Boguña, *Physical Review Letters* **97**, 088701 (2006).
 - [8] M. Boguña, R. Pastor-Satorras, and A. Vespignani, *Physical Review Letters* **90**, 028701 (2003).
 - [9] R. Pastor-Satorras and A. Vespignani, *Physical Review Letters* **86**, 3200 (2001).
 - [10] P. S. Dodds and D. J. Watts, *Physical Review Letters* **92**, 218701 (2004).
 - [11] P. S. Dodds and D. J. Watts, *Journal of Theoretical Biology* **232**, 587 (2005), ISSN 0022-5193.
 - [12] L. A. Meyers, B. Pourbohloul, M. E. J. Newman, D. M. Skowronski, and R. C. Brunham, *Journal of Theoretical Biology* **232**, 71 (2005).
 - [13] M. E. J. Newman, *Physical Review Letters* **95**, 108701 (2005).
 - [14] M. E. J. Newman, *SIAM Review* **45**, 167 (2003).
 - [15] M. B. Hastings, *Physical Review Letters* **96**, 148701 (2006).
 - [16] E. Kenah and J. Robins, Arxiv q-bio/0610057 (2006).
 - [17] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, *Computer Networks* **33**, 309 (2000).
 - [18] P. Grassberger, *Mathematical Biosciences* **63**, 157 (1983).
 - [19] J. L. Cardy and P. Grassberger, *Journal of Physics A: Mathematics and General* **18**, L267 (1985).
 - [20] H. S. Wilf, *generatingfunctionology* (A K Peters, Ltd, 2005), 3rd ed.
 - [21] J. E. Hubbarde, G. Wild, and L. M. Wahl, in preparation

(2007).

- [22] M. Molloy and B. Reed, Random structures & algorithms **6**, 161 (1995), ISSN 1042-9832.
- [23] This statement is true for the probability without requir-

ing $N \rightarrow \infty$, but if the initial infection is in G_i , a few nodes outside of $G_o \cup G_{scc}$ may also be infected, leading to a $o(1)$ correction to the attack rate.