SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Final size of network epidemic models: Properties and connections

Yi WANG[1,2*] & Jinde CAO[2,3*]

[1]*School of Mathematics, Southeast University, Nanjing* 210096, *China;*
[2]*School of Mathematics and Physics, China University of Geosciences, Wuhan* 430074, *China;*
[3]*Jiangsu Provincial Key Laboratory of Networked Collective Intelligence, Southeast University, Nanjing* 210096, *China*

Appendix A Connections between SIR models in annealed networks and classical SIR models in homogeneous populations

One implicit assumption of classical SIR models is the homogeneous Poisson process, that is, the waiting times for next infection event and recovery event occurring are distributed exponentially. The following conservation equality:

$$\frac{d}{dt}\{S(t)+I(t)-\frac{\gamma}{\beta}\ln(S(t))\}\equiv 0,\ t\geqslant 0$$

is employed to derive the classical final size relation. Lemma 1 in main text shows that if the population size is sufficiently large and the population is homogeneous mixing, the basic reproduction number R_0 and the final epidemic size Z are strongly connected by a simple relationship in the limit of negligible proportion of infectious individuals.

In [1], Ma and Earn showed that the simple relationship $Z=1-e^{-R_0 Z}$ was invariant to a range of population structures. In particular, they gave a sufficient condition under which the relationship held for multi-patch SIR epidemic model. Moreover, they provided two examples illustrating social heterogeneities that altered the final size relation. One important class of social heterogeneities is the contact heterogeneity, which can be described by networks or graphs. In other words, each individual is denoted by a node or vertex and each contact or some sort of interaction between individuals is denoted by a link or edge. Thus, networks provide a substrate for disease spreading in it.

Pastor-Satorras and Vespignani [2] proposed the first model describing SIS epidemic in annealed networks. The model divides the population into different classes based on the number of contacts (degree) each individual has, and assumes homogeneous mixing in the same degree class. For SIR dynamics, May and Lloyd [3] first discussed properties of infection processes in annealed scale-free networks, such as the final size of an epidemic in an infinite closed population and the dependence of infection probability on an individual's connectivity degree within the population. The basic reproduction number R_0 of model (2) is dependent on epidemiological parameters and network heterogeneity, and it determines whether a disease can invade a population [4].

For networks of regular type (for example, complete graph), $P(k)$ follows the Delta distribution, i.e., $P(k)=1$ for $k=k_c$ and $P(k)=0$ for $k\neq k_c$. Let $\beta=\tau k_c$, then model (2) reduces to model (1). For networks of exponential type (for example, Erdös-Rényi random graph), $P(k)$ peaks at an average degree $\langle k\rangle$ and decays exponentially fast for $k\gg\langle k\rangle$ and $k\ll\langle k\rangle$; that is, the average degree $\langle k\rangle$ characterizes a network. Let $\beta=\tau\langle k\rangle$, then model (2) reduces to model (1). These two cases correspond to the standard incidence rate in classical epidemiology [5], which shows that model (2) is a refinement of classical model (1) since it uses a general degree distribution $P(k)$ to describe the contact heterogeneity within a population.

On the other hand, population contact networks usually present connectivity cutoffs due to physical constraints or finite population sizes. That is, there exists a maximum degree $k_{max}=n$. Let $\beta_{kj}=\tau kjP(j)/\langle k\rangle$, $k,j=1,\ldots,n$, then system (2) becomes

$$\begin{cases}\frac{dS_k(t)}{dt}=-S_k(t)\sum_{j=1}^{n}\beta_{kj}I_j(t),\\ \frac{dI_k(t)}{dt}=S_k(t)\sum_{j=1}^{n}\beta_{kj}I_j(t)-\gamma I_k(t),\\ \frac{dR_k(t)}{dt}=\gamma I_k(t),\ k=1,\ldots,n,\end{cases}\tag{A1}$$

which is the same as classical multi-group SIR model without population demography [5].

* Corresponding author (email: wangyi-mail@163.com, jdcao@seu.edu.cn)

Appendix B SIR epidemic in annealed networks

We present three methods to find final size equations, analyze their qualitative behavior, and clarify their equivalence.

Appendix B.1 Self-consistent equation approach

Looking at $\dot{S}_k(t)$-equation of (2) and integrating from 0 to t, we have

$$S_k(t) = (1-\epsilon_k)e^{-\tau k\psi(t)}, \tag{B1}$$

where $\psi(t)$ is a auxiliary function defined as

$$\psi(t) = \int_0^t \Theta(s)ds = \int_0^t \sum_k \frac{kP(k)I_k(s)}{\langle k\rangle}ds = \frac{1}{\gamma\langle k\rangle}\sum_k kP(k)R_k(t). \tag{B2}$$

To find a closed equation for the total density of infected nodes, we first give the differential equation for $\psi(t)$:

$$\frac{d\psi(t)}{dt} = \frac{1}{\langle k\rangle}\sum_k kP(k)I_k(t) = 1-\gamma\psi(t) - \frac{1}{\langle k\rangle}\sum_k kP(k)(1-\epsilon_k)e^{-\tau k\psi(t)}, \tag{B3}$$

where we have used $I_k(t) = 1 - S_k(t) - R_k(t)$ and Eq. (B1).

Similar to the derivation of classical final size relation, it is straightforward to verify that $\lim_{t\to+\infty} S_k(t)$, $\lim_{t\to+\infty} I_k(t)$, and $\lim_{t\to+\infty} R_k(t)$ exist, that $\lim_{t\to+\infty} I_k(t) = 0$, and that $\lim_{t\to+\infty}(S_k(t)+R_k(t)) = 1$. Then, we obtain the total density of infected nodes at the end of an epidemic:

$$Z = R(+\infty) = \lim_{t\to+\infty}\sum_k P(k)R_k(t) = 1 - \sum_k P(k)(1-\epsilon_k)e^{-\tau k\psi_\infty}, \tag{B4}$$

where $\psi_\infty = \psi(+\infty)$ for brevity.

Note that $\lim_{t\to+\infty} I_k(t) = 0$, then it follows that $\lim_{t\to+\infty} d\psi(t)/dt = 0$, i.e.,

$$\psi_\infty = \frac{1}{\gamma} - \frac{1}{\gamma\langle k\rangle}\sum_k kP(k)(1-\epsilon_k)e^{-\tau k\psi_\infty}. \tag{B5}$$

Equations (B4) and (B5) determine the final size of an SIR infection in annealed networks. Let

$$f(x) = \frac{1}{\gamma} - \frac{1}{\gamma\langle k\rangle}\sum_k kP(k)(1-\epsilon_k)e^{-\tau kx}.$$

It can be seen that $f(0) = \frac{1}{\gamma} - \frac{1}{\gamma\langle k\rangle}\sum_k kP(k)(1-\epsilon_k) > 0$, $f(1/\gamma) < 1/\gamma$, and that

$$\frac{df(x)}{dx} = \frac{\tau}{\gamma\langle k\rangle}\sum_k k^2P(k)(1-\epsilon_k)e^{-\tau kx} < \frac{\tau}{\gamma}\frac{\langle k^2\rangle}{\langle k\rangle} = R_0$$

when $x \in (0, 1/\gamma)$. If $R_0 < 1$, then according to Brouwer fixed point theorem there exists a unique $\psi_* \in (0, 1/\gamma)$ such that $\psi_* = f(\psi_*)$, and the total density of nodes affected by an epidemic is $R_* = 1 - \sum_k P(k)(1-\epsilon_k)e^{-\tau k\psi_*}$. In particular, in the limit $\epsilon_k \to 0$, Eq. (B5) becomes

$$\psi_\infty = \frac{1}{\gamma} - \frac{1}{\gamma\langle k\rangle}\sum_k kP(k)e^{-\tau k\psi_\infty}. \tag{B6}$$

Obviously, $\psi_\infty = 0$ is always a solution, corresponding to the disease-free equilibrium.

On the other hand, it can be seen that $f'(x) > 0$, $f''(x) < 0$, and that $\lim_{x\to+\infty} f'(x) = 0$. Then, the initial infectious seeds may cause a big outbreak if and only if the condition $f'(0) > 1$ is satisfied, i.e.,

$$R_0 > 1 + \frac{\tau}{\gamma\langle k\rangle}\sum_k k^2P(k)\epsilon_k. \tag{B7}$$

In particular, in the limit $\epsilon_k \to 0$, Eq. (B7) is equivalent to $R_0 > 1$. In other words, there exists a unique positive solution $\psi^* \in (0, 1/\gamma)$ of Eq. (B6), and correspondingly the final epidemic size Z is given by

$$Z = 1 - \sum_k P(k)e^{-\tau k\psi^*}. \tag{B8}$$

It is worth mentioning that ψ^* can be obtained from the iteration method and Z can be estimated by inserting ψ^* into Eq. (B8). Moreover, it is worth remarking that for SIR epidemic without demography, there are two ways to minimize the total number of infected or outbreak size at the end of an epidemic, i.e., decreasing the transmission rate τ oriented to the susceptible (e.g., washing hands, getting vaccines, etc.) or increasing the recovery/remove rate γ oriented to the infected (e.g., receiving treatment, stimulating quarantine, etc.).

Appendix B.2 Classical multi-group approach

Noting the equivalence with classical multi-group epidemic model, we can investigate the qualitative behavior of final size equations by means of the method in [6]. To see this, rewriting the S_k-equations of (A1), we have

$$\frac{d\ln(S_k(t))}{dt} = -\sum_{j=1}^{n}\beta_{kj}I_j(t)$$

so that

$$\ln(S_k(t)) - \ln(1-\epsilon_k) = -\sum_{j=1}^{n}\beta_{kj}\int_0^t I_j(s)ds. \tag{B9}$$

On the other hand, summing the S_k-equation and the I_k-equation of system (A1), we obtain

$$\frac{d(S_k+I_k)(t)}{dt} = -\gamma I_k(t),\ t>0$$

so that

$$(S_k+I_k)(t) - 1 = -\gamma\int_0^t I_k(s)ds. \tag{B10}$$

From (B9) and (B10), we conclude that

$$\ln(S_k(t)) - \ln(1-\epsilon_k) = \sum_{j=1}^{n}\frac{\beta_{kj}}{\gamma}(S_j(t)+I_j(t)-1),\ t>0,$$

or equivalently

$$\sum_{j=1}^{n}\frac{\beta_{kj}}{\gamma}(S_j(t)+I_j(t)) - \ln(S_k(t)) = \sum_{j=1}^{n}\frac{\beta_{kj}}{\gamma} - \ln(1-\epsilon_k).$$

In other words, the following conservation equation:

$$\frac{d}{dt}\Big(\sum_{j=1}^{n}\frac{\beta_{kj}}{\gamma}(S_j(t)+I_j(t)) - \ln(S_k(t))\Big) = 0,\ \forall\ t>0 \tag{B11}$$

holds for each degree class $k = 1, 2, \ldots, n$.

Integrating (B11) from 0 to $+\infty$, we get

$$\begin{cases} S_1(+\infty) = (1-\epsilon_1)\exp\big\{\sum\limits_{j=1}^{n}\frac{\beta_{1j}}{\gamma}(S_j(+\infty)-1)\big\}, \\ \vdots \\ S_n(+\infty) = (1-\epsilon_n)\exp\big\{\sum\limits_{j=1}^{n}\frac{\beta_{nj}}{\gamma}(S_j(+\infty)-1)\big\}. \end{cases} \tag{B12}$$

To study the asymptotic behavior of Eqs. (B12), consider the map $G: R^n \to R^n$ read as

$$G(X) = (G_1(X), \ldots, G_n(X))^T,\ X \in R^n,$$

where

$$G_l(X) = (1-\epsilon_l)\exp\big\{\sum_{j=1}^{n}\frac{\beta_{lj}}{\gamma}(X_j-1)\big\},\ 1 \leqslant l \leqslant n.$$

Moreover, for $X, Y \in R^n$, introduce the following notations:

$$X \leqslant Y \ \Leftrightarrow\ X_k \leqslant Y_k,\ 1 \leqslant k \leqslant n,$$

$$X < Y \ \Leftrightarrow\ X \leqslant Y \text{ and } X_k < Y_k,\ \text{for some } k,\ 1 \leqslant k \leqslant n,$$

$$X \ll Y \ \Leftrightarrow\ X_k < Y_k,\ 1 \leqslant k \leqslant n.$$

Then, it is easy to verify that G is monotone increasing, which means that $X \leqslant Y \Rightarrow G(X) \leqslant G(Y)$. Therefore, if for all k, $0 \leqslant \epsilon_k < 1$, i.e., $S_k(0) > 0$ and $I_k(0) \geqslant 0$, it follows that

$$0 \ll G(0) \leqslant G(S_0) \leqslant S_0,$$

where $S_0 = (S_1(0), \ldots, S_n(0))^T = (1-\epsilon_1, \ldots, 1-\epsilon_n)^T \gg 0$.

Furthermore, using the mathematical induction for each $q \geqslant 1$ yields

$$0 \ll G(0) \leqslant \cdots \leqslant G^q(0) \leqslant G^{q+1}(0) \leqslant G^{q+1}(S_0) \leqslant \cdots \leqslant G(S_0) \leqslant S_0.$$

According to the criterion of monotone bounded sequence, we deduce that

$$0 \ll \lim_{q\to+\infty} G^q(0) \doteq S^- \leqslant S^+ \doteq \lim_{q\to+\infty} G^q(S_0) \leqslant S_0.$$

Noting that G is continuous, we obtain

$$0 \ll G(S^-) = S^-,\ G(S^+) = S^+.$$

Hence, we immediately obtain the following result.

Lemma 1. Assume that $\gamma > 0$ and $0 \leqslant \epsilon_k < 1$, $k = 1, 2, \ldots, n$. Then, the interval $[S^-, S^+]$ contains all the fixed points of G in $[0, S_0]$.

Noting that G is continuously differentiable, we obtain

$$\frac{\partial G_l(X)}{\partial X_j} = \frac{\beta_{lj}}{\gamma} G_l(X),\ \forall\ X \in R^n,\ 1 \leqslant l, j \leqslant n,$$

which can be rewritten as a compact matrix form

$$DG(X) = diag(G(X))BQ^{-1},\ X \in R^n, \tag{B13}$$

where $B = (\beta_{lj})_{1 \leqslant l,j \leqslant n}$, $Q = diag(\gamma, \ldots, \gamma)$ and $Q^{-1} = diag(1/\gamma, \ldots, 1/\gamma)$.

Based on the monotony of G, we conclude that DG has the property of monotony. In particular, for every $0 \leqslant X \leqslant Y \leqslant S_0$ and a vector $A \geqslant 0$, we have

$$DG(X)A \leqslant DG(Y)A.$$

Using the fact that $G(S^-) = S^-$ and $G(S^+) = S^+$ yields $DG(S^-) = diag(S^-)BQ^{-1}$ and $DG(S^+) = diag(S^+)BQ^{-1}$.

Theorem 1. Assume that $\gamma > 0$ and $0 \leqslant \epsilon_k < 1$, $k = 1, 2, \ldots, n$. Assume in addition that matrix B is irreducible, i.e., the node of degree l can be reached from any other node of degree j ($j \neq l$). Then, it holds that

(1) $G(S_0) = S_0$ if and only if $\epsilon_k = 0$ for $k = 1, 2, \ldots, n$;

(2) If $\epsilon_k > 0$ for some $k = 1, 2, \ldots, n$, then there exists a unique fixed point $S(+\infty)$ of G, which lies in the interval $(0, S_0)$.

Proof. Noting that matrix B is irreducible, $S^+ \gg 0$ and $Q^{-1} = diag(1/\gamma, \ldots, 1/\gamma)$, we deduce that matrix $DG(S^+) = diag(S^+)BQ^{-1}$ is nonnegative and irreducible.

Proof of (1): We will show that $S_0 \gg 0$ is a fixed point of G if and only if $\epsilon_k = 0$ for $k = 1, 2, \ldots, n$. To this end, $G(S_0) = S_0$ implies that

$$\begin{cases} (1 - \epsilon_1) = (1 - \epsilon_1) \exp\big\{ - \sum\limits_{j=1}^{n} \frac{\beta_{1j}}{\gamma} \epsilon_j \big\}, \\ \vdots \\ (1 - \epsilon_n) = (1 - \epsilon_n) \exp\big\{ - \sum\limits_{j=1}^{n} \frac{\beta_{nj}}{\gamma} \epsilon_j \big\}, \end{cases} \tag{B14}$$

which is satisfied if and only if

$$\sum_{j=1}^{n} \frac{\beta_{lj}}{\gamma} \epsilon_j = 0,\ l = 1, \ldots, n \Leftrightarrow BQ^{-1} I_0 = 0.$$

Combining the fact that matrix BQ^{-1} is irreducible with $I_0 = (\epsilon_1, \ldots, \epsilon_n)^T \geqslant 0$, we have

$$BQ^{-1} I_0 = 0 \Leftrightarrow \epsilon_k = 0,\ k = 1, 2, \ldots, n.$$

Proof of (2): Suppose that $I_0 > 0$. Then, it follows that $0 \ll G(S_0) < S_0$ and $S^+ < S_0$. Also, using the monotony of G leads to $S^+ = G(S^+) \leqslant G(S_0)$.

Next, to show that there exists a unique fixed point $S(+\infty)$ of G whenever $I_0 > 0$, it is enough to verify that $S^+ = S^-$. To see this, we prove it by contradiction, that is, assuming $S^- < S^+$. Then, it follows that

$$S^+ - S^- = G(S^+) - G(S^-) = \int_0^1 DG(S^- + \chi(S^+ - S^-))(S^+ - S^-) d\chi.$$

Noting that DG is monotone increasing and for all $\chi \in [0, 1]$

$$DG(S^- + \chi(S^+ - S^-))(S^+ - S^-) \leqslant DG(S^+)(S^+ - S^-),$$

we have

$$S^+ - S^- \leqslant DG(S^+)(S^+ - S^-).$$

Noting that $DG(S^+)$ is a nonnegative irreducible matrix, from the Perron-Frobenius theorem we have

$$\omega^T(S^+ - S^-) \leqslant \omega^T DG(S^+)(S^+ - S^-) = \rho(DG(S^+))\omega^T(S^+ - S^-),$$

where $\omega \gg 0$ is a left eigenvector of $DG(S^+)$ related to the spectral radius $\rho(DG(S^+))$. Using assumption $S^- < S^+$, we obtain $\rho(DG(S^+)) \geqslant 1$.

On the other hand, looking at that

$$G(S_0) - S^+ = G(S_0) - G(S^+) = \int_0^1 DG(S^+ + \chi(S_0 - S^+))(S_0 - S^+) d\chi,$$

and for all $\chi \in [0, 1]$

$$DG(S^+ + \chi(S_0 - S^+))(S_0 - S^+) \geqslant DG(S^+)(S_0 - S^+),$$

we deduce that

$$\omega^T(G(S_0) - S^+) \geqslant \rho(DG(S^+))\omega^T(S_0 - S^+).$$

Combing with $\rho(DG(S^+)) \geqslant 1$ yields

$$\omega^T(G(S_0) - S^+) \geqslant \omega^T(S_0 - S^+) \Rightarrow \omega^T G(S_0) \geqslant \omega^T S_0,$$

which contradicts with $G(S_0) < S_0$. □

Remark 1. Theorem 1 has two elements of biological and mathematical implications. First, it clarifies the asymptotic behavior of final size equations (B12). The biological consequence is that the disease will not appear if and only if there are no initial infectious seeds. On the contrary if there are initial infectious seeds of any degree, then nodes of all degree classes can be infected independent of the initial distribution of disease. Second, it tells us that the final size of an epidemic can be calculated as $Z = 1 - S(+\infty) = 1 - \sum_{k=1}^{n} P(k)S_k(+\infty)$, while the final density of susceptible nodes is estimated numerically as

$$S(+\infty) = \lim_{q\to+\infty} G^q(S_0)$$

independent of the initial distribution I_0 of infectious nodes. More precisely, if $I_0 > 0$, we may choose the formula

$$S(+\infty) = \lim_{q\to+\infty} G^q(0)$$

to estimate the final density of susceptible nodes. Moreover, if the network is not connected, i.e., matrix B is not irreducible, Theorem 1 can be used to each isolated connected component such that the matrix associated with it is irreducible.

Noting that $\beta_{kj} = \tau kjP(j)/\langle k\rangle$, $k, j = 1, \ldots, n$, motivated by Ma and Earn [1] we know that if $\sum_{j=1}^{n} \beta_{kj}$ is a constant, i.e., the transmission rate is the same in all degree classes, then the classical final size relation holds. In fact, we conclude the following result.

Theorem 2. Assume that the transmission rate in each degree class is the same, i.e., $\sum_{j=1}^{n} \beta_{kj} = \beta$ for $k = 1, 2, \ldots, n$. Then, in the limit of infinitesimal initial infectious seeds, i.e., $I_0 \to 0$, the final size Z of an epidemic is given by $Z = 1 - e^{-R_0 Z}$, where $R_0 = \beta/\gamma$.

Remark 2. The sufficient condition $\sum_{j=1}^{n} \beta_{kj} = \beta$ reduces to $k = k_c = \frac{\beta}{\tau}$, which means that each node in the network should have exactly the same degree. Noting that the final size Z for the whole network is the weighted sum of final size Z_k for each degree class k, i.e.,

$$Z = \sum_{k=1}^{n} P(k)Z_k, \tag{B15}$$

where $Z_k = R_k(+\infty) = 1 - S_k(+\infty)$ is given by

$$1 - Z_k = \exp\big\{ -\sum_{j=1}^{n} \frac{\beta_{kj}}{\gamma} Z_j \big\}.$$

As a result, there will always exist a particular degree distribution $P^*(k)$ such that Z in Eq. (B15) is a solution of the classical final size equation $Z = 1 - e^{-R_0 Z}$. However, it should be pointed out that since contact networks are reconstructed from network statistics and they are given a priori, it is more interesting to find sufficient or necessary conditions on the transmission matrix B such that Eq. (B15) holds.

Appendix B.3 Edge-based approach

Rewriting Eq. (B1) yields

$$\frac{S_k(t)}{S_k(0)} = \frac{S_k(t)}{1 - \epsilon_k} = e^{-\tau k \psi(t)},$$

which signifies the probability that initial susceptible nodes of degree k still remains susceptible by time t. Letting $\theta(t) = e^{-\tau\psi(t)}$, we interpret $\theta(t)$ as the probability that a stub or half edge has never transmitted infection by t to the node under consideration, usually called test node. Then, we have $S_k(t) = (1 - \epsilon_k)\theta^k(t)$ and $S(t) = \sum_k (1 - \epsilon_k)P(k)\theta^k(t)$.

Let ϕ_S, ϕ_I and ϕ_R be the probability that a stub has never traversed infection to the test node and in addition the current neighbor is susceptible, infectious and recovered, respectively. Noting that in annealed networks, a node chooses a new neighbor at each moment, it follows that the probability of attaching to a node of a given status equals the proportion of all stubs emanating from nodes of that status. Hence, we must give the proportion of stubs that emanate from susceptible, infectious and recovered nodes η_S, η_I and η_R, respectively. Due to the rapid turnover of neighbors, we immediately obtain the following relation: $\phi_S = \theta\eta_S$, $\phi_I = \theta\eta_I$ and $\phi_R = \theta\eta_R$.

To close the equation of $S(t)$, we need $\theta(t)$. The probability θ decreases due to the infection traversing across a stub. Then,

$$\frac{d\theta(t)}{dt} = -\tau\phi_I(t). \tag{B16}$$

Since the neighbors attached to a single stub any two moments are independent, there are no direct relation between ϕ_S and ϕ_I or between ϕ_I and ϕ_R. Substituting $\phi_I(t) = \theta(t)\eta_I(t)$ into Eq. (B16) yields

$$\frac{d\theta(t)}{dt} = -\tau\theta(t)\eta_I(t). \tag{B17}$$

Now we find the equation satisfied by $\eta_I(t)$. Stubs emanating from infectious nodes change into stubs emanating from recovered nodes at rate γ, thus

$$\frac{d\eta_R(t)}{dt} = \gamma\eta_I(t) = \gamma\frac{\phi_I(t)}{\theta(t)} = -\frac{\gamma}{\tau}\frac{d\theta(t)}{\theta(t)dt}.$$

Integrating from 0 to t yields (noting that $\eta_R(0) = 0$ and $\theta(0) = 1$)

$$\eta_R(t) = -\frac{\gamma}{\tau}\ln(\theta(t)). \tag{B18}$$

The equation satisfied by $\eta_S(t)$ can be computed directly. Using the fact that the probability a stub emanating from a node of degree k is $kP(k)/\langle k\rangle$ and that the probability the initial susceptible node keeping susceptible is $(1-\epsilon_k)\theta^k(t)$, we have

$$\eta_S(t) = \frac{1}{\langle k\rangle}\sum_k kP(k)(1-\epsilon_k)\theta^k(t). \tag{B19}$$

Thus, $\eta_I(t) = 1-\eta_S(t)-\eta_R(t) = 1-\frac{1}{\langle k\rangle}\sum_k kP(k)(1-\epsilon_k)\theta^k(t) + \frac{\gamma}{\tau}\ln(\theta(t))$.

Substituting it into Eq. (B17), we obtain

$$\frac{d\theta(t)}{dt} = -\tau\theta(t) + \frac{\tau\theta(t)}{\langle k\rangle}\sum_k kP(k)(1-\epsilon_k)\theta^k(t) - \gamma\theta(t)\ln(\theta(t)). \tag{B20}$$

The final size of an epidemic can be obtained by setting $d\theta(t)/dt = 0$. Solving it, we have

$$\theta(+\infty) = \exp\left\{-\frac{\tau}{\gamma}\left(1-\frac{\sum_k kP(k)(1-\epsilon_k)\theta^k(+\infty)}{\langle k\rangle}\right)\right\}. \tag{B21}$$

Thus,

$$Z = 1-S(+\infty) = 1-\sum_k (1-\epsilon_k)P(k)\theta^k(+\infty). \tag{B22}$$

Remark 3. Noting that $\theta(+\infty) = e^{-\tau\psi(+\infty)} = e^{-\tau\psi_\infty}$, thus Eq. (B22) is the same as Eq. (B4). Then, substituting $\psi_\infty = -\ln(\theta(+\infty))/\tau$ into Eq. (B5), we obtain Eq. (B21). In particular, in the limit $\epsilon_k \to 0$ and all nodes having the same number of stubs k_c, we find

$$\theta(+\infty) = \exp\left\{-\frac{\tau}{\gamma}\big(1-\theta^{k_c}(+\infty)\big)\right\}$$

and

$$\theta(+\infty) = S^{1/k_c}(+\infty) = (1-R(+\infty))^{1/k_c} = (1-Z)^{1/k_c}.$$

Taking k_c'th power on both sides yields

$$1-Z = \exp\big(-\frac{\tau}{\gamma}k_cZ\big),$$

which retrieves the classical final size relation $Z = 1-e^{-R_0Z}$ (noting that $R_0 = \tau k_c/\gamma$). Moreover, the basic reproduction number R_0 can be explained directly using edge-based approach. The probability that a newly infectious node early in the epidemic has degree k is $kP(k)/\langle k\rangle$. Since its neighbors changes rapidly, it may have k new neighbors at all times and cause new infections at rate τk for the whole infectious period. Then,

$$R_0 = \sum_k \frac{kP(k)}{\langle k\rangle}\frac{\tau k}{\gamma} = \frac{\tau}{\gamma}\frac{\langle k^2\rangle}{\langle k\rangle}.$$

Appendix C SIR epidemic in quenched networks

We present three methods to derive final size equations in quenched networks, and discuss their equivalence.

Appendix C.1 Bond percolation approach

Given a degree distribution $P(k)$, the generating function of it is defined as

$$G_0(x) = \sum_{k=0}^{\infty} P(k)x^k, \tag{C1}$$

where x is a dummy variable. Note that $G_0(1) = \sum_k P(k) = 1$, and the degree distribution $P(k)$ can be "generated" by repeated differentiation:

$$P(k) = \frac{1}{k!}\frac{d^kG_0(x)}{dx^k}\big|_{x=0}. \tag{C2}$$

The generating function has some nice properties, for example, the mean degree $\langle k\rangle$ of a network is calculated as

$$\langle k\rangle = \sum_k kP(k) = \frac{dG_0(x)}{dx}\big|_{x=1} \doteq G_0'(1). \tag{C3}$$

Another useful quantity is the distribution of degrees of nodes arrived at by following randomly chosen edges. Since edges are more likely to be attached to nodes of higher degree, the probability is proportional to $kP(k)$, and thus the generating function for this distribution is

$$\sum_k \frac{kP(k)}{\langle k\rangle}x^k = x\sum_k \frac{kP(k)}{\langle k\rangle}x^{k-1} = x\frac{G_0'(x)}{G_0'(1)}. \tag{C4}$$

In actual applications, we are more interested in the number of edges emanating from a node excluding the edge we follow, which is one less than the degree, usually called the "excess degree" of a node. To obtain the generating function for excess degree distribution, we simply divide Eq. (C4) by x on both sides and have

$$G_1(x) \doteq \frac{G_0'(x)}{G_0'(1)} = \frac{1}{\langle k\rangle}G_0'(x). \tag{C5}$$

Consider a random network of arbitrary degree distribution, the percolation problem focuses on the component size distribution with bond occupation probability T ($0 \leqslant T \leqslant 1$). To this end, the generating functions $G_0(x,T)$ and $G_1(x,T)$, also as functions of the occupation probability T, for the distribution of the number of occupied edges emanating from a node are needed. More precisely, the probability that a node of degree k having exactly m edges occupied follows the binomial distribution $C_k^m T^m (1-T)^{k-m}$, and then the generating function is

$$\begin{aligned} G_0(x,T) &= \sum_{m=0}^{\infty}\sum_{k=m}^{\infty} P(k) C_k^m T^m (1-T)^{k-m} x^m = \sum_{k=0}^{\infty} P(k) \sum_{m=0}^{k} C_k^m (xT)^m (1-T)^{k-m} = \sum_{k=0}^{\infty} P(k)(xT+1-T)^k \\ &= G_0(1+(x-1)T). \end{aligned} \tag{C6}$$

Note that, the following equalities hold

$$G_0(x,1) = G_0(x),\ G_0(1,T) = G_0(1) = 1,\ \frac{\partial G_0(x,T)}{\partial x}\Big|_{x=1} = TG_0'(1) = T\langle k\rangle.$$

Similarly, the generating function for the excess degree distribution with occupation probability T is

$$G_1(x,T) = G_1(1+(x-1)T), \tag{C7}$$

and the equalities hold for $G_1(x,1)$, $G_1(1,T)$ and $\frac{\partial G_1(x,T)}{\partial x}\big|_{x=1}$.

Denote by $P_s(T)$ the distribution of size s of clusters of nodes linked together by occupied edges. Denote by $H_0(x,T)$ the generating function for this distribution, then

$$H_0(x,T) = \sum_{s=0}^{\infty} P_s(T) x^s. \tag{C8}$$

Let $H_1(x,T)$ be the generating function for the distribution of size of clusters of connected nodes that are arrived at by following a randomly chosen edge. It follows that [7] $H_1(x,T)$ satisfies a Dyson-equation-like self-consistent equation

$$H_1(x,T) = xG_1(H_1(x,T),T). \tag{C9}$$

Similarly, the generating function $H_0(x,T)$ for the size of the whole component reached from a randomly chosen node is

$$H_0(x,T) = xG_0(H_1(x,T),T). \tag{C10}$$

If functions $G_0(x)$ and $G_1(x)$ are known, one can solve Eq. (C9) for $H_1(x,T)$ and substitute it into Eq. (C10) to find $H_0(x,T)$. Then, the probability distribution $P_s(T)$ of component of size s is given by the s-th derivative of $H_0(x,T)$.

In general, it is impossible to find a closed-form solution of Eq. (C9), for example, if it is a transcendental equation. However, assuming no giant component exists in the network, one can use Eqs. (C9) and (C10) to find the average component size in closed form. Similar to Eq. (C3), we get

$$\langle s(T)\rangle = \sum_s sP_s(T) = \frac{\partial H_0(x,T)}{\partial x}\Big|_{x=1} = 1 + \frac{\partial G_0(x,T)}{\partial x}\Big|_{x=1} \cdot \frac{\partial H_1(x,T)}{\partial x}\Big|_{x=1}, \tag{C11}$$

using the fact $H_1(1,T) = 1$ and $G_0(1,T) = 1$.

Moreover, differentiating Eq. (C9) with respect to x yields

$$\frac{\partial H_1(x,T)}{\partial x}\Big|_{x=1} = 1 + \frac{\partial G_1(x,T)}{\partial x}\Big|_{x=1} \cdot \frac{\partial H_1(x,T)}{\partial x}\Big|_{x=1}, \tag{C12}$$

and hence

$$\frac{\partial H_1(x,T)}{\partial x}\Big|_{x=1} = \frac{1}{1 - \frac{\partial G_1(x,T)}{\partial x}\big|_{x=1}}. \tag{C13}$$

Substituting Eq. (C13) into Eq. (C11) and using Eqs. (C6) and (C7) lead to

$$\langle s(T)\rangle = 1 + \frac{\partial G_0(x,T)}{\partial x}\Big|_{x=1} \cdot \frac{1}{1 - \frac{\partial G_1(x,T)}{\partial x}\big|_{x=1}} = 1 + \frac{TG_0'(1)}{1 - TG_1'(1)}. \tag{C14}$$

Note that Eq. (C14) can be used to calculate the mean component size for any occupation probability T and arbitrary degree distribution $P(k)$. It is easy to see that Eq. (C14) diverges if $TG_1'(1) = 1$, which signifies the first appearance of a giant component. In other words, there exists a critical occupation probability $T_c = 1/G_1'(1)$, above which a giant component appears. Using Eq. (C5), we have

$$T_c = \frac{\langle k\rangle}{G_0''(1)} = \frac{\langle k\rangle}{\langle k^2\rangle - \langle k\rangle}. \tag{C15}$$

If there is a giant component in the network ($T > T_c$), $H_0(1,T) = 1$ does not hold since it generates the probability distribution of size of components excluding the giant component. Then, by definition $H_0(1,T) + \mathcal{S}(T) = 1$, where $\mathcal{S}(T)$ denotes the density of nodes taken by the giant component. Thus,

$$\mathcal{S}(T) = 1 - H_0(1,T) = 1 - \sum_{s=0}^{\infty} P_s(T). \tag{C16}$$

Substituting Eqs. (C9) and (C10) into Eq. (C16) gives

$$\mathcal{S}(T) = 1 - G_0(u, T), \tag{C17}$$

where $u \equiv H_1(1, T)$ satisfies the following self-consistent equation:

$$u = G_1(u, T). \tag{C18}$$

Now the disease threshold and the final epidemic size can be derived by mapping SIR model to percolation model. Note that the disease-causing contacts in SIR model correspond to the occupied edges in percolation model. Since the probability that an infectious node makes disease-causing contacts is $\frac{\tau}{\tau+\gamma}$, substituting $T = \frac{\tau}{\tau+\gamma}$ into Eq. (C15) we obtain the disease threshold

$$\frac{\tau}{\tau+\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle} = 1,$$

which gives the basic reproduction number

$$R_0 = \frac{\tau}{\tau+\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle} = \frac{\tau}{\tau+\gamma}\left(\langle k\rangle - 1 + \frac{Var[k]}{\langle k\rangle}\right). \tag{C19}$$

If $R_0 > 1$, there is an epidemic, and the final epidemic size Z is given by

$$Z = \mathcal{S}(T) = 1 - G_0(1 - T + Tu) = 1 - \sum_{k=0}^{\infty} P(k)(1 - T + Tu)^k, \tag{C20}$$

where u is given by

$$u = G_1(1 - T + Tu) = \frac{G_0'(1 - T + Tu)}{G_0'(1)} = \frac{1}{\langle k\rangle}\sum_{k=0}^{\infty} kP(k)(1 - T + Tu)^{k-1}. \tag{C21}$$

The generating functions for some specific degree distributions are calculated as follows. For networks of Delta degree distribution, $G_0(x) = x^{k_c}$ and $G_1(x) = x^{k_c-1}$. For networks of Poisson degree distribution, $G_1(x) = G_0(x) = e^{\langle k\rangle(x-1)}$. For networks of pure power-law degree distribution, $G_0(x) = Li_l(x)/\zeta(l)$ and $G_1(x) = Li_{l-1}(x)/(x\zeta(l-1))$. For networks of general degree distribution, the generating function for $P(k)$ is given in terms of a polynomial with the finite power, i.e.,

$$G_0(x) = \sum_{k=0}^{n} P(k)x^k = \sum_{k=0}^{n} \frac{N_k}{\sum_{k=0}^{n} N_k}x^k = \frac{\sum_{k=0}^{n} N_k x^k}{N}. \tag{C22}$$

It immediately follows that

$$G_1(x) = \frac{\sum_{k=0}^{n} kN_k x^{k-1}}{\sum_{k=0}^{n} kN_k}. \tag{C23}$$

Substituting Eqs. (C22) and (C23) into Eqs. (C20) and (C21), we can numerically calculate the final epidemic size Z.

Appendix C.2 Edge-based approach

The key idea is as follows [8]. Let $\theta(t)$ be the probability that a randomly chosen neighbor w of v has not transmitted infection to v by time t. Let $\phi_S(t)$, $\phi_I(t)$ and $\phi_R(t)$ be the probability that w is susceptible, infectious and recovered, and has not transmitted infection to v by time t, respectively. Then, $\theta(t) = \phi_S(t) + \phi_I(t) + \phi_R(t)$.

Consider a susceptible node of degree k at $t = 0$, then the probability it remains susceptible by time t is $(1 - \epsilon_k)\theta^k(t)$. Hence, the probability that nodes of any degree remain susceptible at time t is the weighted sum of the product of the degree distribution $P(k)$ with the probability a node of degree k is susceptible by time t, i.e.,

$$S(t) = \sum_k S_k(0)P(k)\theta^k(t) = \sum_k (1 - \epsilon_k)P(k)\theta^k(t). \tag{C24}$$

To close Eq. (C24), we have to find $\theta(t)$. The probability θ may decrease because of the infection traversing across an edge, which is only possible if it is connected to an infectious node. Then,

$$\frac{d\theta(t)}{dt} = -\tau\phi_I(t). \tag{C25}$$

An edge does not satisfy the definition of $\phi_I(t)$ either because infection crosses it (with rate τ), or because the infectious neighbor recovers (with rate γ). An edge begins to satisfy the definition because the susceptible neighbor becomes infectious by receiving infection other than the edge under consideration. Let $h(t)$ be the probability that a neighbor of an edge is susceptible, then it becomes infectious with rate $-dh(t)/dt$. Thus,

$$\frac{d\phi_I(t)}{dt} = -(\tau+\gamma)\phi_I(t) - \frac{dh(t)}{dt}.$$

Note that the probability the neighbor at the end of a randomly chosen edge is of degree k is $kP(k)/\langle k\rangle$, and that the probability it is susceptible is $(1 - \epsilon_k)\theta^{k-1}(t)$, excluding the edge under consideration. Thus,

$$h(t) = \sum_k \frac{kP(k)}{\langle k\rangle}(1 - \epsilon_k)\theta^{k-1}(t) = \frac{1}{\langle k\rangle}\sum_k (1 - \epsilon_k)kP(k)\theta^{k-1}(t).$$

Then, it follows that

$$\frac{d\phi_I(t)}{dt} = -(\tau+\gamma)\phi_I(t) + \frac{\tau\phi_I(t)}{\langle k\rangle}\sum_k (1-\epsilon_k)k(k-1)P(k)\theta^{k-2}(t). \tag{C26}$$

Furthermore, using Eq. (C25) and integrating Eq. (C26) from 0 to t yields

$$\phi_I(t) = \theta(t) - \frac{\gamma}{\tau}(1-\theta(t)) - \frac{1}{\langle k\rangle}\sum_k (1-\epsilon_k)kP(k)\theta^{k-1}(t),$$

where the initial conditions $\phi_I(0) = \sum_k kP(k)\epsilon_k/\langle k\rangle$ and $\theta(0) = 1$ are employed. As a result, it follows that

$$\frac{d\theta(t)}{dt} = -\tau\theta(t) + \gamma(1-\theta(t)) + \frac{\tau}{\langle k\rangle}\sum_k (1-\epsilon_k)kP(k)\theta^{k-1}(t). \tag{C27}$$

From Eq. (C27), we can calculate the density of nodes in each state at any time t directly. For example, the density of infectious nodes at time t is given by

$$\frac{dI(t)}{dt} = -\frac{dS(t)}{dt} - \gamma I(t) = -\sum_k (1-\epsilon_k)kP(k)\theta^{k-1}(t)\frac{d\theta(t)}{dt} - \gamma I(t). \tag{C28}$$

In the limit $\epsilon_k \to 0$, it is not difficult to obtain the basic reproduction number R_0 using the next generation matrix method [9]. Specifically, linearizing Eqs. (C26)-(C28) around the disease-free equilibrium ($\phi_I = 0$, $\theta = 1$ and $I = 0$) and calculating the dominant eigenvalue of the next generation matrix gives

$$R_0 = \frac{\tau}{\tau+\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}, \tag{C29}$$

which is the same as (C19).

To obtain the final size relation, letting $d\theta(t)/dt = 0$ yields

$$\theta(+\infty) = \frac{\frac{\tau}{\langle k\rangle}\sum_k (1-\epsilon_k)kP(k)\theta^{k-1}(+\infty) + \gamma}{\tau+\gamma}. \tag{C30}$$

Then, the final epidemic size is given by

$$Z = R(+\infty) = 1 - S(+\infty) = 1 - \sum_k (1-\epsilon_k)P(k)\theta^k(+\infty). \tag{C31}$$

In the limit $\epsilon_k \to 0$, Eq. (C30) can be rewritten as

$$\theta(+\infty) = \frac{\tau\frac{G_0'(\theta(+\infty))}{G_0'(1)} + \gamma}{\tau+\gamma}. \tag{C32}$$

Similarly, Eq. (C31) can be rewritten as

$$Z = 1 - \sum_k P(k)\theta^k(+\infty) = 1 - G_0(\theta(+\infty)). \tag{C33}$$

Recall that in the percolation model, the quantity u measures the probability that the node at the end of a randomly chosen edge belongs to one of the finite components rather than the giant component. In terms of epidemiology, it is the probability that the node at the end of a randomly chosen edge escaping from the infection finally. Then, the probability that a node does not receive infection through one of its edges is the sum of two parts: first, the probability that no transmission occurs along the edge is $1-T$; second, the probability that the transmission may occur along the edge but it is connected to a susceptible node is Tu. Thus, the total probability that a node of degree k remains susceptible is $(1-T+Tu)^k$. Letting $\theta(+\infty) = 1-T+Tu$, we know Eq. (C33) is equivalent to Eq. (C20). Substituting $\theta(+\infty) = 1-T+Tu$ and $T = \frac{\tau}{\tau+\gamma}$ into Eq. (C32) leads to

$$1 - T + Tu = T\frac{G_0'(1-T+Tu)}{G_0'(1)} + 1 - T, \tag{C34}$$

which is equivalent to Eq. (C21).

Remark 4. The model presented here takes large proportion of initial infectious seeds into consideration, and the results are more general. In the limit $\epsilon_k \to 0$, the basic reproduction number R_0 and the final epidemic size Z obtained are equivalent to those based on the percolation theory.

Appendix C.3 Stochastic process approach

This approach is motivated by the edge-based approach in subsection C.2, and the key idea is similar [10]. Consider a test individual v, preventing it from transmitting the infection to direct neighbors. Denote by θ_∞ the probability that a neighbor w of v never transmits to v, then the probability that a node of degree k never receives infection is θ_∞^k. Noting that the probability that a node of degree k is susceptible at $t = 0$ is $1-\epsilon_k$, we obtain the total fraction of susceptible nodes across a network:

$$S(+\infty) = \sum_k (1-\epsilon_k)P(k)\theta_\infty^k. \tag{C35}$$

To obtain the final size relation, one must calculate θ_∞. Let ϖ be the probability that the neighbor w of v is susceptible at the end of an epidemic. Recalling that $T = \tau/(\tau+\gamma)$ is the probability a randomly chosen infectious neighbor transmits infection to v before recovery, we know the probability v receives infection is $T(1-\varpi)$. Thus, the probability of v not receiving infection across the chosen edge is

$$\theta_\infty = 1 - T(1-\varpi) = 1 - T + T\varpi. \tag{C36}$$

Noting that the probability w is of degree k by following a randomly chosen edge is $kP(k)/\langle k\rangle$, and the probability w is susceptible at the end of an epidemic is $(1-\epsilon_k)\theta_\infty^{k-1}$ given that v does not transmit infection to w, we have

$$\theta_\infty = \sum_k \frac{kP(k)}{\langle k\rangle}(1-\epsilon_k)\theta_\infty^{k-1}.$$

Hence, Eq. (C36) becomes

$$\theta_\infty = 1 - T + T\frac{\sum_k (1-\epsilon_k)kP(k)\theta_\infty^{k-1}}{\langle k\rangle}, \tag{C37}$$

and the final size Z is given by

$$Z = 1 - S(+\infty) = 1 - \sum_k (1-\epsilon_k)P(k)\theta_\infty^k.$$

Remark 5. One implicit assumption of the above approach is that the number of initial infectious seeds is large enough to avoid stochastic extinction of an epidemic. An advantage of this approach is that it is not dependent on the spreading process of an epidemic, but just the probability that a node will not receive infection from any neighbor finally. The final size equation (C37) is equivalent to Eq. (C30). In the limit $\epsilon_k \to 0$, Eqs. (C21), (C32) and (C37) are equivalent. However, it should be pointed out that the stochastic process approach can not tell us whether a well-defined final size relation exists. If a well-defined final size relation can be derived through integro-differential equations, the stochastic process approach could be effective.

References

1 Ma J, Earn D J D. Generality of the final size formula for an epidemic of a newly invading infectious disease. Bull Math Biol, 2006, 68: 679-702
2 Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. Phys Rev Lett, 2001, 86: 3200-3203
3 May R M, Lloyd A L. Infection dynamics on scale-free networks. Phys Rev E, 2001, 64: 066112
4 Jin Z, Li S P, Zhang X G, et al. Epidemiological modeling on complex networks. In: Lü J, Yu X, Chen G, et al. (eds) Complex Systems and Networks. Berlin, Heidelberg: Springer, 2016. 51-77
5 Anderson R M, May R M. Infectious diseases of humans: dynamics and control. Oxford: Oxford University Press, 1992
6 Magal P, Seydi O, Webb G. Final size of a multi-group SIR epidemic model: irreducible and non-irreducible modes of transmission. Math Biosci, 2018, 301: 59-67
7 Newman M E J. Spread of epidemic disease on networks. Phys Rev E, 2002, 66: 016128
8 Miller J C, Kiss I Z. Epidemic spread in networks: existing methods and current challenges. Math Model Nat Phenom, 2014, 9: 4-42
9 van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math Biosci, 2002, 180: 29-48
10 Miller J C. A note on the derivation of epidemic final sizes. Bull Math Biol, 2012, 74: 2125-2141