Final size of network epidemic models: properties and connections

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

[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*

Citation Wang Y, Cao J D. Final size of network epidemic models: properties and connections. Sci China Inf Sci, 2021, 64(7): 179201, https://doi.org/10.1007/s11432-019-2656-2

Dear editor,

The classical final size relation of the standard susceptible-infected-recovered (SIR) epidemic model is based on homogeneous-mixing population [1,2]. Recent results have demonstrated the generality of the classical SIR final size relation [3]. However, even for an SIR epidemic in static configuration-type networks, the classical final size relation may be violated. Therefore, it is still an important problem to understand how the basic reproduction number, R_0, and the final size, Z, of disease related to each other on networks [4]. This study aims to clarify the relationship between them.

SIR epidemic in homogeneous populations. In 1927, Kermack and McKendrick [1] formulated the following standard SIR model:

$$
\begin{cases}
\dfrac{\mathrm{d}S(t)}{\mathrm{d}t} = -\beta S(t)I(t), \\[2mm]
\dfrac{\mathrm{d}I(t)}{\mathrm{d}t} = \beta S(t)I(t) - \gamma I(t), \\[2mm]
\dfrac{\mathrm{d}R(t)}{\mathrm{d}t} = \gamma I(t),
\end{cases}
\tag{1}
$$

where $S(t)$, $I(t)$, and $R(t)$ denote the proportion of susceptible, infected, and recovered individuals at time t, respectively. In addition, β and γ denote the infection and recovery rates of infected individuals, respectively. The initial conditions of system (1) are $S(0) = 1 - \epsilon$, $I(0) = \epsilon$ ($0 < \epsilon < 1$), and $R(0) = 0$.

The basic reproduction number, which is defined as the expected number of secondary cases produced by an index case during its infectious period in a totally susceptible population, of system (1) is $R_0 = \beta/\gamma$, while the final epidemic size, which is defined as the likely magnitude of an outbreak, is $Z = R(+\infty) \doteq \lim_{t \to +\infty} R(t)$. The following result summarizes the relationship between R_0 and Z [2].

Lemma 1. Let $(S(t), I(t), R(t))$ denote a solution vector of system (1). If $R_0 \leqslant 1$, then $I(t)$ directly decreases to zero as t approaches infinity. If $R_0 > 1$, then $I(t)$ first increases

up to a maximum value $I_{\max} = 1 - \frac{1}{R_0}\{1 + \ln(R_0(1 - \epsilon))\}$ and then decreases to zero as t approaches infinity. Moreover, $S(t)$ and $R(t)$ are nonincreasing and nondecreasing functions, respectively; in particular, there exists a unique limit $S(+\infty)$ in the interval $(0, 1/R_0)$ that satisfies the following equation:

$$
\ln\left(\frac{S(+\infty)}{1 - \epsilon}\right) = R_0\big(S(+\infty) - 1\big).
$$

If $\epsilon \to 0$, i.e., the proportion of infected individuals is infinitesimal, one obtains the well-known final size relation, $Z = 1 - \mathrm{e}^{-R_0 Z}$, which indicates that R_0 and Z are strongly connected by a simple relationship.

SIR epidemic in annealed networks. Annealed networks assume that the neighbors of each node rapidly change at every moment. By utilizing the degree-based mean-field theory, Moreno et al. [5] formulated the following model:

$$
\begin{cases}
\dfrac{\mathrm{d}S_k(t)}{\mathrm{d}t} = -\tau k S_k(t)\Theta(t), \\[2mm]
\dfrac{\mathrm{d}I_k(t)}{\mathrm{d}t} = \tau k S_k(t)\Theta(t) - \gamma I_k(t), \\[2mm]
\dfrac{\mathrm{d}R_k(t)}{\mathrm{d}t} = \gamma I_k(t),
\end{cases}
\tag{2}
$$

where $S_k(t)$, $I_k(t)$, $R_k(t)$ denote the relative density of susceptible, infected, and recovered nodes of degree k at time t, respectively. It is noteworthy that τ and γ denote the infection rate along each edge and the recovery rate of an infected node, respectively. Additionally, $\Theta(t) = \sum_k k P(k) I_k(t)/\langle k \rangle$ ($\langle k \rangle = \sum_k k P(k)$ denotes the average degree) denotes the probability that a randomly chosen stub or half-edge points to an infected node. System (2) is finished with the following initial conditions: $S_k(0) = 1 - \epsilon_k$, $I_k(0) = \epsilon_k$ ($0 < \epsilon_k < 1$), and $R_k(0) = 0$.

The infection threshold of system (2) is given by

$$
R_0 = \frac{\tau}{\gamma}\frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{\tau}{\gamma}\left(\langle k \rangle + \frac{\mathrm{Var}[k]}{\langle k \rangle}\right),
\tag{3}
$$

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

where $\epsilon_k \to 0$, for all k, and $\mathrm{Var}[k] = \langle k^2 \rangle - \langle k \rangle^2$ denotes the variance of the degree distribution $P(k)$. The connections between the SIR models in annealed networks and the classical SIR models in homogeneous populations are presented in Appendix A.

Using a self-consistent equation approach (see Appendix B.1 for details), in the limit $\epsilon_k \to 0$, Z satisfies the following equation:

$$Z = 1 - \sum_k P(k) \mathrm{e}^{-\tau k \psi^*}, \tag{4}$$

where $\psi^* \in (0, 1/\gamma)$ is unique if $R_0 > 1$, and satisfies the following equation:

$$\psi^* = \frac{1}{\gamma} - \frac{1}{\gamma \langle k \rangle} \sum_k k P(k) \mathrm{e}^{-\tau k \psi^*}. \tag{5}$$

To observe the relationship between Z and R_0, we analyze the following three cases.

Case 1. For networks of the regular type, $R_0 = \tau k_c / \gamma$. Moreover, it follows that $\psi^* = (1 - \mathrm{e}^{-\tau k_c \psi^*})/\gamma$ and $Z = 1 - \mathrm{e}^{-\tau k_c \psi^*}$. Hence, we obtain the classical final size relation.

Case 2. For networks of the Poisson type, i.e., $P(k) = \langle k \rangle^k \mathrm{e}^{-\langle k \rangle}/k!$, for $k = 0, 1, 2, \ldots$, $R_0 = \tau(\langle k \rangle + 1)/\gamma$. Therefore, Eq. (4) reduces to

$$Z = 1 - (1 - \gamma \psi^*) \mathrm{e}^{\tau \psi^*}.$$

Case 3. For networks of the scale-free type, i.e., $P(k) = Ak^{-l}$ $(2 < l \leqslant 3)$, where A denotes a normalization constant, $R_0 \to +\infty$ if $\langle k^2 \rangle \to +\infty$. Using the continuous approximation of degree k, we have

$$\psi^* = \frac{1}{\gamma} - \frac{A}{\gamma \langle k \rangle} \int_m^{+\infty} k^{-(l-1)} \mathrm{e}^{-\tau k \psi^*} \mathrm{d}k,$$

$$Z = 1 - A \int_m^{+\infty} k^{-l} \mathrm{e}^{-\tau k \psi^*} \mathrm{d}k,$$

where m denotes the minimum degree. The above two integrals can be addressed in terms of the incomplete Gamma function; therefore, Z can be approximated as a linear function of ψ^* for small ψ^*.

In general, we cannot find an explicit solution of (5). Meanwhile, we can still obtain some useful information on estimating Z. If τ is small or γ is large, i.e., ψ^* is small, then the exponential term in (5) can be given as follows:

$$\mathrm{e}^{-\tau k \psi^*} \approx 1 - \tau k \psi^* + \frac{1}{2} \tau^2 k^2 \psi^{*2}. \tag{6}$$

Substituting (6) into (5) yields

$$\psi^* \approx (R_0 - 1) \frac{2\gamma}{\tau^2} \frac{\langle k \rangle}{\langle k^3 \rangle}. \tag{7}$$

Besides, substituting (6) and (7) into (4) yields

$$Z \approx 2 \left(1 - \frac{1}{R_0}\right) \frac{\langle k \rangle \langle k^2 \rangle}{\langle k^3 \rangle} - 2 \left(1 - \frac{1}{R_0}\right)^2 \frac{\langle k^2 \rangle^3}{\langle k^3 \rangle^2}. \tag{8}$$

Remark 1. (1) Various methods are applied to obtain the final size equations. The asymptotic behavior of the equations is analyzed in Appendix B.2. The connections between them are presented in Appendix B.3. (2) Eq. (8) contains an epidemiological parameter, R_0, which can be estimated from real disease data, and a network parameter, $P(k)$, which can

be estimated from network statistics. (3) Eq. (8) can be utilized to estimate the effect of control effort on the spread of an epidemic [6].

SIR epidemic in quenched networks. Quenched networks assume that the neighbors of each node are fixed at all times. By utilizing the bond percolation theory, Newman [7] observed that Z (see Appendix C.1 for details) is given by

$$Z = \mathcal{S}(T) = 1 - \sum_{k=0}^{\infty} P(k)(1 - T + Tu)^k, \tag{9}$$

where $\mathcal{S}(T)$ denotes the density of nodes given by the giant component in a network, $T = \tau/(\tau + \gamma)$ denotes the probability that an infected node makes disease-causing contacts, and u satisfies the following self-consistent equation:

$$u = \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} k P(k)(1 - T + Tu)^{k-1}. \tag{10}$$

Moreover, the infection threshold is defined as

$$R_0 = T \left(\frac{\langle k^2 \rangle}{\langle k \rangle} - 1\right) = T \left(\langle k \rangle - 1 + \frac{\mathrm{Var}[k]}{\langle k \rangle}\right). \tag{11}$$

Its derivation is presented in Appendix C.1.

Next, we give some examples of the specific degree distribution for calculating Z.

Case 1. For networks of the delta degree distribution, the basic reproduction number is given by $R_0 = T(k_c - 1)$. From (9) and (10), we have

$$Z = 1 - (1 - T + Tu)^{k_c}, \tag{12}$$

and

$$u = (1 - T + Tu)^{k_c - 1}, \tag{13}$$

respectively.

Hence, u satisfies the following quadratic equation:

$$Tu^2 + (1 - T)u - (1 - Z) = 0. \tag{14}$$

Because $0 < T = \frac{\tau}{\tau + \gamma} < 1$ and $0 \leqslant Z < 1$, there exists a unique positive root, u^*, of (14):

$$u^* = \frac{-(1 - T) + \sqrt{(1 - T)^2 + 4T(1 - Z)}}{2T}.$$

By substituting it into (12), we obtain a self-consistent equation for Z:

$$Z = 1 - \left(\frac{1 - T + \sqrt{(1 - T)^2 + 4T(1 - Z)}}{2}\right)^{k_c}.$$

Case 2. For networks of the Poisson degree distribution, the basic reproduction number is given by $R_0 = T\langle k \rangle$. Subsequently, from (9) and (10), we obtain

$$Z = 1 - \mathrm{e}^{-T\langle k \rangle(1-u)}, \tag{15}$$

and

$$u = \mathrm{e}^{-T\langle k \rangle(1-u)}, \tag{16}$$

respectively.

It is noteworthy that $u = 1 - Z$, and by substituting it into (15), we obtain the classical final size relation as follows: $Z = 1 - \mathrm{e}^{-T\langle k \rangle Z} = 1 - \mathrm{e}^{-R_0 Z}$.

Case 3. For networks of the pure power-law degree distribution, the basic reproduction number is given by

$R_0 = T(\zeta(l-2)/\zeta(l-1) - 1)$, where $\zeta(s)$ denotes the Riemann ζ function. Subsequently, we obtain

$$Z = 1 - Li_l(1 - T + Tu)/\zeta(l), \tag{17}$$

and

$$u = \frac{Li_{l-1}(1 - T + Tu)}{(1 - T + Tu)\zeta(l-1)}, \tag{18}$$

where $Li_q(x)$ denotes the q-th polylogarithm of x.

For networks of the general degree distribution, the degree distribution can be approximated as $P(k) = N_k / \sum_{k=0}^{n} N_k$, where N_k denotes the number of nodes with degree k, and n denotes the maximum degree. By employing (9) and (10) (a polynomial with finite power), we can numerically calculate the final epidemic size, Z.

Remark 2. (1) The percolation approach is not suitable for demonstrating the explicit dynamical behavior of an epidemic on quenched networks, so an edge-based approach is provided in Appendix C.2. (2) The basic reproduction number, R_0, and final epidemic size, Z, based on edge-based approach are respectively equivalent to those based on percolation theory.

Conclusion. For the SIR epidemic on annealed networks, our results demonstrated that the classical final size relation holds only for the delta degree distribution, or alternatively, there exists a specific degree distribution, $P^*(k)$, such that the relation holds (given in Appendix B.2). For the SIR epidemic on quenched networks, it is observed that the classical final size relation holds only for the Poisson degree distribution. In complex networks with arbitrary degree distribution, these results are important for estimating the expected magnitude or an outbreak severity of an epidemic.

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Kermack W O, McKendrick A G. A contribution to the mathematical theory of epidemics. Proc R Soc A-Math Phys Eng Sci, 1927, 115: 700–721

2 Hethcote H W. The mathematics of infectious diseases. SIAM Rev, 2000, 42: 599–653

3 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

4 Pellis L, Ball F, Bansal S, et al. Eight challenges for network epidemic models. Epidemics, 2015, 10: 58–62

5 Moreno Y, Pastor-Satorras R, Vespignani A. Epidemic outbreaks in complex heterogeneous networks. Eur Phys J B, 2002, 26: 521–529

6 Bai T, Li S Y, Zou Y. Minimum input selection of reconfigurable architecture systems for structural controllability. Sci China Inf Sci, 2019, 62: 019201

7 Newman M E J. Spread of epidemic disease on networks. Phys Rev E, 2002, 66: 016128