


## Basic reproduction number of epidemic models on sparse networks

Satoru Morita <sup>\*</sup>

*Department of Mathematical and Systems Engineering, Shizuoka University, Hamamatsu 432-8561, Japan*

 (Received 30 December 2021; revised 25 May 2022; accepted 6 September 2022; published 22 September 2022)

The basic reproduction number  $\mathcal{R}_0$  is a standard indicator of infection control in epidemiology. Although  $\mathcal{R}_0$  has been studied extensively for deterministic epidemic models, it has not been formulated accurately for models adopting network structures. Here, we extend a four-compartment model that includes commonly used epidemic models to a Markov process on networks. By examining the Markov process in detail, we derive a canonical formula for  $\mathcal{R}_0$  involving two probability values. Numerical calculations show that the derived formula is a better approximation than the conventional formula when the network is very sparse. We propose this as a standard formula for controlling infections that can only be transmitted through intimate contact, where contacts between individuals can be described as a sparse network.

DOI: [10.1103/PhysRevE.106.034318](https://doi.org/10.1103/PhysRevE.106.034318)

### I. INTRODUCTION

Despite advances in medical science and public health, infectious diseases pose serious threats to humanity, as exemplified by the outbreaks of emerging infectious diseases such as SARS, MERS, and COVID-19 [1,2]. In epidemiology, the basic reproduction number  $\mathcal{R}_0$  is an important indicator of the transmission potential of infectious diseases [3–6].  $\mathcal{R}_0$  gives the average number of secondary cases for a typical infection in a completely susceptible population. When  $\mathcal{R}_0 > 1$ , the infection can spread in the host population. By contrast, when  $\mathcal{R}_0 < 1$ , the infection does not spread. Higher values of  $\mathcal{R}_0$  indicate epidemics that are difficult to control, and infection prevention measures are focused on reducing  $\mathcal{R}_0$ . The features of  $\mathcal{R}_0$  have been investigated thoroughly for compartmental or mean-field models based on ordinary differential equations [3–6]. Despite the fact that human contact networks are a key factor in the spread of infection [7,8], no unified formula for  $\mathcal{R}_0$  has been obtained for the epidemic models on networks [9–13].

In this paper, we consider an SEIRS model with four compartments—susceptible (S), exposed (E), infectious (I), and recovered (R)—that encompasses several commonly used epidemic models, and we investigate the impact of population disorder as represented by a network on the spreading processes. We assume a network that is sparse and has no loops. By examining the stochastic process of this model in detail, we clarify the importance of two probability values,  $c_1$  and  $c_2$ , defined below. Consider a pair of individuals  $i$  and  $j$ . When an infection transmission occurs from individual  $j$  to individual  $i$ , individual  $i$  recovers before individual  $j$  with probability  $c_1$  [Fig. 1(b)]. By contrast, when an infection transmission occurs from individual  $i$  to individual  $j$ , individual  $i$  recovers before individual  $j$  with probability  $c_2$  [Fig. 1(c)]. We derive a formula for  $\mathcal{R}_0$  based on  $c_1$  and  $c_2$  for an SEIRS model

on networks. This approach better approximates the number of secondary infections, which is crucial given that neighbors with waned immunity can be reinfected.

The remainder of this paper is organized as follows. Section II redefines the SEIRS model as a Markov process on networks. Section III presents a method for approximating the basic reproduction number for the SEIRS model on networks. Section IV compares the theoretical results with numerical calculations for three types of networks. Section V discusses the results and their implications.

### II. MODEL

Here, we consider the SEIRS model. A susceptible individual is not infected and not immune, and an exposed individual is infected but is in the latent noninfectious period. An infectious individual has completed the latency period and is infectious, and a recovered individual is immunized after recovering. The current model considers the process of returning from R to S, assuming that immunity is lost over time. Let  $S$ ,  $E$ ,  $I$ , and  $R$  be the populations of susceptible, exposed, infectious, and recovered individuals, respectively. In the framework of a deterministic ordinary differential equation model, the SEIRS model is as follows,

$$\begin{aligned} \frac{dS(t)}{dt} &= -\beta S(t)I(t) + \omega R(t), \\ \frac{dE(t)}{dt} &= \beta S(t)I(t) - \sigma E(t), \\ \frac{dI(t)}{dt} &= \sigma E(t) - \gamma I(t), \\ \frac{dR(t)}{dt} &= \gamma I(t) - \omega R(t), \end{aligned} \quad (1)$$

where  $\beta$  represents the rate of transmission between a susceptible and an infectious individual, exposed individuals acquire infectivity at the rate  $\sigma$  (moving to I), infectious

<sup>\*</sup>morita.satoru@shizuoka.ac.jp

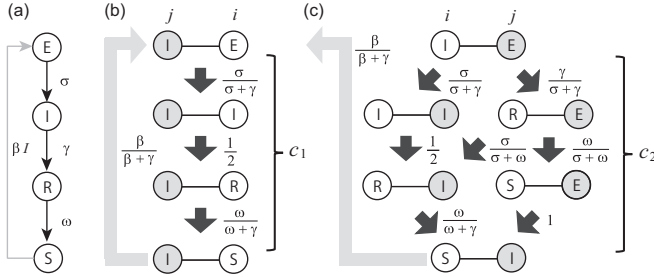


FIG. 1. Illustration of the calculation of probabilities  $c_1$  and  $c_2$  in terms of the rates of infection transmission  $\sigma$ ,  $\gamma$ , and  $\omega$  between individuals in susceptible (S), exposed (E), infectious (I), and recovered (R) states. Consider a pair of individuals  $i, j$  who become I or E as a result of infection transmission between them. (a) shows the rates of transition between states E, I, R, and S. (b) shows the processes required for a newly exposed individual  $i$  to recover and become susceptible again before the original infectious individual  $j$  recovers. The probability of each process is shown on the right-hand side of the corresponding arrow. The product of the three probabilities gives probability  $c_1$  in Eq. (12). (c) shows the processes required for the original infectious individual  $i$  to recover and become susceptible again before a newly exposed individual  $j$  recovers. There are multiple possible paths, and summing their probabilities yields probability  $c_2$  in Eq. (16). The lightly shaded arrows indicate the process by which the I-S pair becomes I-E (i.e., infection occurs), and the probability that this occurs is given by Eq. (11) if there are no other infectious individuals adjacent to individual  $i$ . In the main text, we calculate the average number of secondary infections from individual  $j$ , indicated by the shaded circles.

individuals recover at the rate  $\gamma$  (moving to R), and recovered individuals lose immunity at the rate  $\omega$  (moving to S). Although several studies have used  $\beta/N$  (where  $N$  is the total population) instead of  $\beta$  in Eq. (1) [3–6], we adopt this form to facilitate the introduction of the network structure, setting  $\beta$  to be the rate at which an infectious individual infects a neighboring susceptible individual. Here, we assume that the outbreak occurs on a short timescale (i.e., that the epidemic dynamics is substantially faster than the demographic dynamics). Thus, births and deaths are neglected, and therefore the total population  $N = S(t) + E(t) + I(t) + R(t)$  remains constant. As summarized in Table I, the model represented as Eq. (1) includes various commonly used models as limits.

TABLE I. Several popular models expressed as parameter limits of Eq. (1).  $c_1$  and  $c_2$  are the probability coefficients derived in this paper.

Model	$\sigma$	$\omega$	$c_1$	$c_2$
SIS	$\rightarrow \infty$	$\rightarrow \infty$	1/2	1/2
SEIS	$(0, \infty)$	$\rightarrow \infty$	$\frac{1}{2} \frac{\sigma}{\sigma+\gamma}$	$1 - \frac{1}{2} \frac{\sigma}{\sigma+\gamma}$
SIRS	$\rightarrow \infty$	$(0, \infty)$	$\frac{1}{2} \frac{\omega}{\omega+\gamma}$	$\frac{1}{2} \frac{\omega}{\omega+\gamma}$
SIR	$\rightarrow \infty$	$\rightarrow 0$	0	0
SEIR	$(0, \infty)$	$\rightarrow 0$	0	0

For the model in Eq. (1), the basic reproduction number is [4]

$$\mathcal{R}_0 = \frac{\beta}{\gamma} N. \tag{2}$$

This is because the average length of the infectious period is  $1/\gamma$ , and the rate of new infections during the infectious period is  $\beta N$ . A more rigorous method for calculating  $\mathcal{R}_0$  uses the next-generation matrix [3,14,15]. Note that Eq. (2) is independent of  $\sigma$  and  $\omega$ . If the population is well mixed, the above statement is adequate. In reality, however, individuals are in contact with only a very small portion of the entire population. A simple way to model such a situation is to describe human interactions as a network that is relatively sparse [16]. Given a population of  $N$  individuals, the network is represented using the  $N \times N$  adjacency matrix in which entry  $(i, j)$  represents the link between individuals  $i$  and  $j$ , as follows:

$$A_{ij} = \begin{cases} 1 & \text{if there is a link between } i \text{ and } j, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Here, we assume that the network is undirected ( $A_{ij} = A_{ji}$ ) and that the weights of the connections are uniform. In this case, the degree of individual  $i$  is defined as

$$k_i = \sum_{j=1}^N A_{ij}, \tag{4}$$

and the average degree of the network is given as

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N A_{ij}. \tag{5}$$

When the SEIRS model given by Eq. (1) is extended to a population network of size  $N$ , the  $4^N$ -state Markov process should be considered [17]. This is because each individual is in one of the four compartments (S, E, I, R). Individuals in compartments E, I, and R independently transition to compartments I, R, and S at rates  $\sigma$ ,  $\gamma$ , and  $\omega$ , respectively, as shown in Fig. 1(a).

### III. APPROXIMATION THEORY

By applying the conventional individual-based mean-field approximation [7,17], we obtain approximated differential equations instead of Eq. (1),

$$\begin{aligned} \frac{dp_S(i, t)}{dt} &= -\beta \sum_{j=1}^N A_{ij} p_S(i, t) p_I(j, t) + \omega p_R(i, t), \\ \frac{dp_E(i, t)}{dt} &= \beta \sum_{j=1}^N A_{ij} p_S(i, t) p_I(j, t) - \sigma p_E(i, t), \\ \frac{dp_I(i, t)}{dt} &= \sigma p_E(i, t) - \gamma p_I(i, t), \\ \frac{dp_R(i, t)}{dt} &= \gamma p_I(i, t) - \omega p_R(i, t), \end{aligned} \tag{6}$$

where  $p_S(i, t)$ ,  $p_E(i, t)$ ,  $p_I(i, t)$ , and  $p_R(i, t)$  are the probabilities that individual  $i$  is in states S, E, I, and R, respectively.

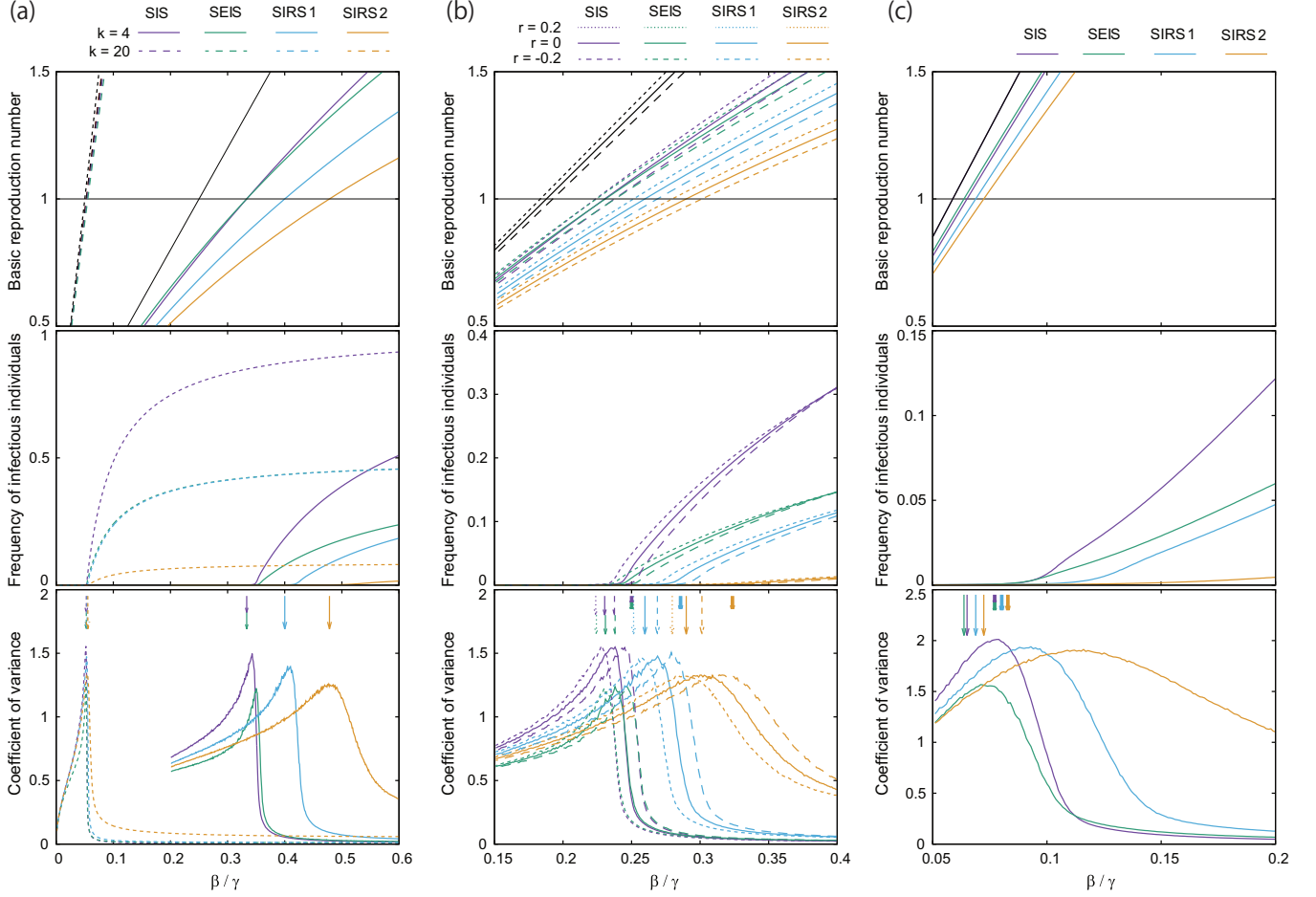


FIG. 2. Comparison of theoretical and numerical calculations for the following networks: (a) two regular graphs with  $k = 4$  (solid lines) and  $k = 20$  (dotted lines); (b) random networks with 5000 nodes having degree 2 and 5000 nodes having degree 6, where the degree correlation is  $r = 0.2$  (dotted lines), 0 (solid lines), and  $-0.2$  (dashed lines); (c) Barabási-Albert network with  $\langle k \rangle = 4$ . We set  $N = 10\,000$  and  $\gamma = 1$ . The graphs in the top row plot the theoretical calculated value of  $\mathcal{R}_0$  as a function of  $\beta/\gamma$ . Here, the black curve corresponds to  $\mathcal{R}_0 = k\beta/\gamma$  as obtained by Eq. (10). The purple, green, blue, and orange curves represent Eq. (21) or (20) for SIS, SEIS, SIRS 1, and SIRS 2, respectively, where the values of  $c_1$  and  $c_2$  are as listed in Table II. The graphs in the middle row plot the relative frequency of infectious individuals ( $\langle I \rangle/N$ ) as a function of  $\beta/\gamma$  by numerical simulation. The graphs in the bottom row plot the coefficient of variation  $\text{Std}(I)/\langle I \rangle$ , the maximum of which corresponds to the epidemic threshold. Here, the long downward arrows represent the epidemic thresholds  $\beta_c$  given by Eq. (22), and the short downward arrows represent those given by Eq. (25); these values are likewise shown in Table II. Note that in the middle graph of (a), the results for SEIS and SIRS 1 with  $k = 20$  nearly overlap.

By performing a stability analysis of the disease-free state, the  $2N$ -dimensional Jacobi matrix of Eq. (6) can be obtained,

$$J = \begin{pmatrix} -\sigma I & \beta A \\ \sigma I & -\gamma I \end{pmatrix}. \quad (7)$$

where  $I$  is the  $N$ -dimensional unit matrix. Then, by dividing the Jacobi matrix into the parts related to infections and the rest of the transitions

$$J = T - \Sigma = \begin{pmatrix} 0 & \beta A \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} \sigma I & 0 \\ -\sigma I & \gamma I \end{pmatrix}, \quad (8)$$

the next-generation matrix can be constructed [7,17],

$$T\Sigma^{-1} = \frac{\beta}{\gamma} \begin{pmatrix} A & A \\ 0 & 0 \end{pmatrix}. \quad (9)$$

Thus, the basic reproduction number would be calculated as follows,

$$\mathcal{R}_0 = \frac{\beta}{\gamma} \Lambda_{1,A}, \quad (10)$$

where  $\Lambda_{1,A}$  is the largest eigenvalue of the adjacency matrix  $A$ . For a complete graph ( $A_{ij} = 1$  for all  $i, j$ ), Eq. (10) is equivalent to Eq. (2). In what follows, we derive a deviation from this approximation for the infection-spreading process in sparse networks. Consider a pair consisting of neighboring individuals  $i$  and  $j$  who are susceptible and infectious, respectively. The probability that individual  $i$  becomes exposed before individual  $j$  recovers is

$$p_{\geq 1} = \frac{\beta}{\beta + \gamma}. \quad (11)$$

Immediately after this infection occurs, the pair is in the state shown at the top of Fig. 1(b). The probability that individual

$i$  recovers and becomes susceptible again before individual  $j$  recovers is

$$c_1 = \frac{1}{2} \frac{\sigma}{\sigma + \gamma} \frac{\omega}{\omega + \gamma}, \quad (12)$$

because three independent events must occur, as shown in Fig. 1(b). The values of  $c_1$  for some well-known models are summarized in Table I. Assuming that individual  $i$  has no other infectious neighbors, it is possible for individual  $j$  to infect individual  $i$  again with a probability of

$$p_{\geq 2} = c_1 \left( \frac{\beta}{\beta + \gamma} \right)^2. \quad (13)$$

In the generalized form, the probability that individual  $i$  is repeatedly infected at least  $h$  times is

$$p_{\geq h} = c_1^{h-1} \left( \frac{\beta}{\beta + \gamma} \right)^h. \quad (14)$$

Thus, we can compute the average number of times a neighbor is infected by a focal node as follows:

$$\begin{aligned} \bar{h} &= \sum_{h=1}^{\infty} h(p_{\geq h} - p_{\geq h-1}) = \sum_{h=1}^{\infty} p_{\geq h} \\ &= \frac{\beta}{\beta(1 - c_1) + \gamma}. \end{aligned} \quad (15)$$

Conversely, consider the case of infection from individual  $i$  to  $j$ . As the exposed individual  $j$  can cause infection in  $k_j - 1$  individuals other than individual  $i$ , the average number of secondary infections is  $(k_j - 1)\bar{h}$ . In addition, there is a possibility of reinfection if individual  $i$  has recovered. If we approximate that individual  $i$  has no other infectious neighbors, the probability that this neighbor recovers before individual  $i$  can be calculated as follows:

$$c_2 = \frac{1}{2} \frac{\sigma}{\sigma + \gamma} \frac{\omega}{\omega + \gamma} + \frac{\gamma}{\sigma + \gamma} \left( \frac{\sigma}{\sigma + \omega} \frac{\omega}{\omega + \gamma} + \frac{\omega}{\sigma + \omega} \right). \quad (16)$$

The derivation of Eq. (16) is illustrated in Fig. 1(c), and the values of  $c_2$  for some well-known models are given in Table I. Clearly,  $c_1 \leq c_2$ . If there is no recovery stage ( $\omega \rightarrow \infty$ ),  $c_1 + c_2 = 1$ , and if there is no exposure stage ( $\sigma \rightarrow \infty$ ),  $c_1 = c_2$ . The average number of times the infection is transferred back to the source neighbor is  $c_2\bar{h}$ . By summing the above results, we obtain the average number of infections from individual  $j$  as

$$n_{\text{infection}}^j = (k_j - 1 + c_2) \frac{\beta}{\beta(1 - c_1) + \gamma}. \quad (17)$$

Now, to calculate the basic reproduction number, we develop a new individual-based mean-field approximation using Eq. (17). The average number of times that infectious individual  $j$  infects one of its neighbors is approximated by dividing Eq. (17) by  $k_j$ :

$$\frac{n_{\text{infection}}^j}{k_j} \simeq \frac{\beta[1 - (1 - c_2)/k_j]}{\beta(1 - c_1) + \gamma}. \quad (18)$$

Here, we do not specify which neighbor is the original source of infection but approximate it by a uniform distribution.

Thus, the average number of infections from individual  $j$  to individual  $i$  can be approximated by

$$B_{ij} = \frac{\beta[1 - (1 - c_2)/k_j]}{\beta(1 - c_1) + \gamma} A_{ij}, \quad (19)$$

which provides the next-generation matrix. Clearly,  $B_{ij} \leq (\beta/\gamma)A_{ij}$ , and  $B_{ij} = (\beta/\gamma)A_{ij}$  if  $c_1 = c_2 = 1$ . Thus, the basic reproduction number is given by

$$\mathcal{R}_0 = \frac{\beta}{\gamma} \Lambda_{1,B}, \quad (20)$$

where  $\Lambda_{1,B}$  is the largest eigenvalue of  $B_{ij}$ .

## IV. EXAMPLES

### A. Regular graph

As a simple example, we consider the case of a regular graph in which all nodes are of the same degree,  $k_i = k$ . In this case, it is apparent that  $(1, 1, \dots, 1)^T$  is an eigenvector of  $B_{ij}$ , and according to the Perron-Frobenius theorem, it gives the largest eigenvalue. Thus, the basic reproduction number is given as

$$\mathcal{R}_0 = \frac{\beta(k - 1 + c_2)}{\beta(1 - c_1) + \gamma}, \quad (21)$$

and the epidemic threshold (i.e.,  $\beta$  satisfying  $\mathcal{R}_0 = 1$ ) is therefore given as

$$\beta_c = \frac{\gamma}{k - 2 + c_1 + c_2}. \quad (22)$$

In the limit of large  $k$ , we obtain  $\mathcal{R}_0 = k\beta/\gamma$ , which is consistent with the result from Eq. (10). If  $k$  is finite,  $\mathcal{R}_0$  and  $\beta_c$  depend on  $c_1$  and  $c_2$ . In the case of  $k = 2$ , for SIR and SEIR ( $c_1 = c_2 = 0$ ),  $\beta_c$  diverges to infinity, which implies that the infection cannot spread. This is consistent with the fact that a disease cannot spread when  $k < 2$ . The top graph in Fig. 2(a) shows plots of Eq. (21) for four cases: SIS, SEIS, and two types of SIRS (short and long immunity duration), as summarized in Table II. The results for  $k = 20$  show that the difference between models is small for large  $k$ ; thus, Eq. (10) is acceptable when the network is not sparse. SIS and SEIS have the same epidemic threshold  $\beta_c = \gamma/(k - 1)$  because  $c_1 + c_2 = 1$ , but they have different values of  $\mathcal{R}_0$ , in contrast to the results from the deterministic model. For SIRS,  $c_1 = c_2$ , and this value increases with  $\omega$ . Thus, the longer the period of R, the smaller  $\mathcal{R}_0$  becomes, and the larger  $\beta_c$  becomes. To validate the above theoretical results, we performed numerical calculations of the Markov process on a network with  $N = 10\,000$  [the middle graph in Fig. 2(a)]. Here, the network was built using a configuration model that is designed to eliminate self-loops and multilinks [18]. The Gillespie algorithm was used for the infection simulation [19], and the average was taken over  $10^{10}$  Monte Carlo steps, where a new infection was introduced every time all infections became extinct. To estimate the epidemic threshold numerically, we plotted the coefficient of variation  $\text{Std}(I)/\langle I \rangle$ , which is the standard deviation of the number of individuals infected divided by the average [bottom graph in Fig. 2(a)]. The maximum of  $\text{Std}(I)/\langle I \rangle$  is considered to correspond to the epidemic

TABLE II. Parameters ( $\sigma$ ,  $\omega$ ,  $c_1$ ,  $c_2$ ) and epidemic threshold ( $\beta_c$ ) for the four types of compartment models shown in Table I.  $N = 10\,000$  and  $\gamma = 1$ . Here,  $\beta_c^{(\text{RG}4)}$  and  $\beta_c^{(\text{RG}20)}$  represent  $\beta_c$  values for a regular graph with  $k = 4$  and  $k = 20$ , respectively;  $\beta_c^{(2,6^+)}$ ,  $\beta_c^{(2,6)}$ , and  $\beta_c^{(2,6^-)}$  are those for networks having two degrees ( $k = 2$  and  $k = 6$ ), where  $r = 0.2$ ,  $r = 0$ , and  $r = -0.2$ , respectively; and  $\beta_c^{(\text{BA})}$  is that for the Barabási-Albert network with  $\langle k \rangle = 4$ . The upper rows show the theoretically determined threshold [Eq. (20) or (22)] and the lower rows show the numerically estimated value based on the maximum of coefficient of variation  $\text{Std}(I)/\langle I \rangle$ . The figures in parentheses are calculated by Eq. (25). The bottom line shows the results using the conventional formula, Eq. (10).

Model	$\sigma$	$\omega$	$c_1$	$c_2$	Threshold	$\beta_c^{(\text{RG}4)}$	$\beta_c^{(\text{RG}20)}$	$\beta_c^{(2,6^+)}$	$\beta_c^{(2,6)}$	$\beta_c^{(2,6^-)}$	$\beta_c^{(\text{BA})}$
SIS	$\rightarrow \infty$	$\rightarrow \infty$	1/2	1/2	Theoretical	0.333	0.053	0.224	0.230 (0.250)	0.237	0.065 (0.077)
					Numerical	0.343	0.052	0.228	0.238	0.245	0.078
SEIS	1	$\rightarrow \infty$	1/4	3/4	Theoretical	0.333	0.053	0.224	0.231 (0.250)	0.238	0.064 (0.077)
					Numerical	0.350	0.052	0.230	0.238	0.246	0.071
SIRS 1	$\rightarrow \infty$	1	1/4	1/4	Theoretical	0.400	0.054	0.251	0.260 (0.286)	0.269	0.069 (0.080)
					Numerical	0.405	0.053	0.258	0.269	0.279	0.093
SIRS 2	$\rightarrow \infty$	1/10	1/22	1/22	Theoretical	0.478	0.055	0.280	0.290 (0.324)	0.301	0.072 (0.083)
					Numerical	0.477	0.054	0.289	0.300	0.314	0.116
Eq. (10)			1	1		0.250	0.050	0.184	0.188 (0.200)	0.193	0.059 (0.072)

threshold. It can be seen that the present theory agrees well with the numerical calculations.

**B. Network with two types of nodes**

Next, as an example of a case where the degree is distributed, we consider the case of a network in which half of the nodes have degree 2 and the other half have degree 6. Here, we also consider the effect of degree correlation, which is measured by the assortativity coefficient  $r$  [20]. We created three networks ( $r = 0.2, 0$ , and  $-0.2$ ) using the configuration model method [18], where the probability of connection between tips emerging from a node changes to match the degree correlation. Figure 2(b) shows the results of calculations similar to those described in Sec. IV A. The top graph shows the basic reproduction number, as given by Eq. (20). Owing to the degree fluctuations,  $\mathcal{R}_0$  is smaller than in the case of a regular graph and increases with the degree correlation. These trends are consistent with those in the previous reports [20,21]. We find that  $\beta_c$  for SIS is slightly smaller than that for SEIS (see also Table II). The bottom graph shows that the epidemic thresholds obtained from Eq. (20) are a little smaller than the numerical results, but considerably better than the results from Eq. (10) in Table II. In particular, the effect of degree correlation is qualitatively well reproduced.

**C. Barabási-Albert model**

A final example is the Barabási-Albert model [22], in which the degree distribution follows a power law,  $\rho(k) \sim k^{-3}$ . For this case, the epidemic threshold  $\beta_c$  is known to converge to zero when  $N \rightarrow \infty$  [23]. Figure 2(c) shows the results of calculations similar to those described in Secs. IV A and IV B. The discrepancy between the theoretical and numerical results appears to be larger than in the previous two cases. In addition,  $\beta_c$  in SEIS is slightly smaller than  $\beta_c$  in SIS, indicating that the incubation period can increase the likelihood of the spread of infection. This may be because  $c_2$  is larger for SEIS than for SIS, and its effect outweighs that of the smaller  $c_1$  value.

**V. DISCUSSION**

The epidemic threshold  $\beta_c$  was obtained from the basic reproduction number given by Eq. (20). Although the numerical calculations do not give exact values of  $\beta_c$  because of the finite size of the population ( $N = 10\,000$ ), it is clear that these results are an improvement over those given by Eq. (10). However, the theoretical values of  $\beta_c$  tend to be underestimated. This is because the calculations in Eqs. (15) and (17) ignore the possibility that after becoming susceptible, individual  $i$  may be infected by a neighbor other than individual  $j$  [bottom graphs in Figs. 1(b) and 1(c)]. As this effect is large for hubs in the Barabási-Albert network, it would explain the large discrepancy seen in that example. The above approximation theory can be coarse-grained by using the degree-based mean-field approximation [7,11,23,24]. For the case with no degree correlation ( $r = 0$ ), the average number of secondary infected individuals with degree  $k'$  transmitted from an infectious individual with degree  $k$  is

$$B_{k'k} = \frac{\beta(k-1+c_2)k'p(k')}{\beta(1-c_1)+\gamma\langle k \rangle}, \tag{23}$$

which gives the next-generation matrix by classifying the individuals by degree [24]. Calculating its largest eigenvalue according to Ref. [24], the basic reproduction number is given by

$$\mathcal{R}_0 = \frac{\beta}{\beta(1-c_1)+\gamma\langle k \rangle} \left( \frac{\langle k^2 \rangle}{\langle k \rangle} - 1 + c_2 \right). \tag{24}$$

Thus, the epidemic threshold for  $\mathcal{R}_0 = 1$  is given as

$$\beta_c = \frac{\gamma\langle k \rangle}{\langle k^2 \rangle - (2-c_1-c_2)\langle k \rangle}. \tag{25}$$

For a regular graph,  $\langle k^2 \rangle = k^2$ ; thus, Eq. (25) becomes Eq. (22). The numbers in parentheses in Table II are the epidemic threshold values calculated by Eq. (25) and shown by the short arrows in Figs. 2(b) and 2(c). These numbers tend to be larger than those given by Eq. (20); this is due to the fact that there are fluctuations in the probability of infection for individuals of the same degree.  $\beta_c = \gamma\langle k \rangle/\langle k^2 \rangle$  and  $\beta_c = \gamma\langle k \rangle/(\langle k^2 \rangle - \langle k \rangle)$  are well-known theoretical formulas



for SIS and SIR [7,16], respectively, but Eq. (25) indicates that

$$\beta_c = \frac{\gamma \langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \quad \text{for SIS and SEIS,}$$

$$\beta_c = \frac{\gamma \langle k \rangle}{\langle k^2 \rangle - 2\langle k \rangle} \quad \text{for SIR and SEIR,} \quad (26)$$

because  $c_1 + c_2 = 1$  for SIS and SEIS, and  $c_1 = c_2 = 0$  for SIR and SEIR. These equations deviate from the well-known degree-based mean-field approximation results [7,11,23] by the amount of  $-\langle k \rangle$  in the denominator. For SIS, the same equation as that in Eq. (26) was obtained by approximations that consider dynamical correlation [25,26]. Note, however, that such degree-based approximations are not always accurate because they do not reproduce the vanishing threshold even when  $\gamma > 3$  [27].

Note that our derivation of the theoretical equation takes into account the effect of reinfection between the same pair but ignores the effect of dynamical correlations between adjacent pairs and network loops. In other words, even without considering the effects of such higher-order correlations, our results are more accurate than those of the conventional mean-field theory. The next-generation matrix in Eq. (19) generally differs from those obtained using heterogeneous pair approximation [28]; however, oddly, they coincide in the case of regular graphs. Furthermore, we performed numerical calculations on networks with larger clustering coefficients,

which have many loops, and the epidemic threshold increased with the clustering coefficient as expected (not shown in this paper). Theoretical formulation of  $\mathcal{R}_0$  by higher-order approximation is a subject for future study.

In summary, we derived a unified formula for  $\mathcal{R}_0$  for epidemic models of networks. The next-generation matrix is represented by Eq. (19), which contains two theoretically calculated probability values,  $c_1$  and  $c_2$ . In reality, reinfection may be much less frequent, and therefore  $c_1$  and  $c_2$  may be considered to have smaller values. Recently, Markov process models that consider dynamical correlations among the states of two or more nodes have been used to study infectious disease transmission [29,30]. Although such models can more accurately determine the epidemic threshold, the computational cost is enormous. By contrast, the formula derived in this paper for the basic reproduction number is simple and easy to use. For complex models that can be attributed to the Markov process, we expect that our formula can be used if  $c_1$  and  $c_2$  can be estimated. With the increasing availability of high-resolution data on human contacts, this result for  $\mathcal{R}_0$  can be used as a reference quantity for controlling the spread of infectious diseases.

#### ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI (Grants No. 21K03387 and No. 21H01575).

- 
- [1] N. C. Peeri, N. Shrestha, M. S. Rahman, R. Zaki, Z. Tan, S. Bibi, M. Baghbanzadeh, N. Aghamohammadi, W. Zhang, and U. Haque, *Int. J. Epidemiol.* **49**, 717 (2020).
  - [2] J. Guarner, *Am. J. Clin. Pathol.* **153**, 420 (2020).
  - [3] O. Diekmann, J. A. P. Heesterbeek, and J. A. Metz, *J. Math. Biol.* **28**, 365 (1990).
  - [4] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, New York, 1991).
  - [5] H. W. Hethcote, *SIAM Rev.* **42**, 599 (2000).
  - [6] O. Diekmann and J. A. P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*, Wiley Series in Mathematical & Computational Biology (Wiley, West Sussex, UK, 2000).
  - [7] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, *Rev. Mod. Phys.* **87**, 925 (2015).
  - [8] I. Z. Kiss, J. C. Miller, and P. L. Simon, *Mathematics of Epidemics on Networks* (Springer, Cham, 2017).
  - [9] M. Boguñá and R. Pastor-Satorras, *Phys. Rev. E* **66**, 047104 (2002).
  - [10] K. T. D. Eames and M. J. Keeling, *Proc. Natl. Acad. Sci. USA* **99**, 13330 (2002).
  - [11] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Rev. Mod. Phys.* **80**, 1275 (2008).
  - [12] S. Yuan, P. van den Driessche, F. H. Willeboordse, Z. Shuai, and J. Ma, *J. Math. Biol.* **73**, 665 (2016).
  - [13] J. Yang and F. Xu, *IEEE Access* **7**, 26474 (2019).
  - [14] O. Diekmann, J. A. P. Heesterbeek, and M. G. Roberts, *J. R. Soc. Interface* **7**, 873 (2010).
  - [15] P. van den Driessche, *Infect. Dis. Model.* **2**, 288 (2017).
  - [16] A.-L. Barabási and M. Pósfai, *Network Science* (Cambridge University Press, Cambridge, UK, 2016).
  - [17] P. Van Mieghem, J. Omic, and R. Kooij, *IEEE/ACM Trans. Netw.* **17**, 1 (2009).
  - [18] M. Newman, *Networks* (Oxford University Press, Oxford, UK, 2018).
  - [19] D. T. Gillespie, *Annu. Rev. Phys. Chem.* **58**, 35 (2007).
  - [20] M. E. J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
  - [21] S. Morita, *Physica A* **563**, 125419 (2021).
  - [22] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
  - [23] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).
  - [24] S. Morita, *Physica A* **587**, 126514 (2022).
  - [25] A. S. Mata, R. S. Ferreira, and S. C. Ferreira, *New J. Phys.* **16**, 053006 (2014).
  - [26] C.-R. Cai, Z.-X. Wu, M. Z. Q. Chen, P. Holme, and J.-Y. Guan, *Phys. Rev. Lett.* **116**, 258301 (2016).
  - [27] S. Chatterjee and R. Durrett, *Ann. Probab.* **37**, 2332 (2009).
  - [28] A. S. Mata and S. C. Ferreira, *Europhys. Lett.* **103**, 48003 (2013).
  - [29] J. T. Matamalas, A. Arenas, and S. Gómez, *Sci. Adv.* **4**, eaau4212 (2018).
  - [30] G. Burgio, A. Arenas, S. Gómez, and J. T. Matamalas, *Commun. Phys.* **4**, 111 (2021).