

# Appendix to “Improving Epidemic Modeling with Networks,” Economic Commentary, 2020-23

Ben Craig, Tom Phelan, Jan-Peter Siedlarek, and Jared Steinberg  
Federal Reserve Bank of Cleveland

This appendix provides details on the simulations contained in the main text. For the benefit of the reader we briefly review some foundational material relating to networks and graphs before turning to the systems of equations characterizing heterogeneous mean field networks and homogeneous networks with clustering. For brevity we focus only on those details necessary for the replication of the models given in the main text, and refer the reader to Kiss, Miller and Simon [1] for further details.

First recall the notions of graphs, edges and degree.

**Definition 1.** A *graph* is an ordered pair  $G = (V, E)$ , where  $V$  is a set of *vertices* and  $E \subseteq V \times V$  is the set of *edges* between vertices. The *degree* of a node is the number of edges connected to it.

Implicit in the above definition is the notion that edges are imbued with *direction*, since  $E$  lies in the direct product  $V \times V$  of ordered pairs. For this reason some authors refer to the above as a *directed* graph. We will only be concerned with undirected graphs, which may be viewed as graphs of the above form in which for any  $v, w \in V$  we have  $(v, w) \in E$  if and only if  $(w, v) \in E$ . If  $V$  is a finite set and ordered according to  $V = (v_1, \dots, v_N)$  then each graph  $G$  is associated with an *adjacency* matrix with  $(i, j)$  value denoted  $g_{ij}$ , which is equal to 1 if  $(v_i, v_j) \in E$  and zero otherwise.

Throughout our analysis we focus exclusively on SIR epidemics (as opposed to, say, SIS epidemics, in which recovery does not confer immunity) on different networks. In an SIR model the nodes in question may be one of three possible types,  $S, I$  and  $R$ , indicating whether or not the node is *susceptible, infected* or *recovered*, respectively. A susceptible node that is adjacent to  $n$  distinct infected nodes becomes infected at (Poisson) rate  $\tau n$ , for some constant  $\tau$ , while an infected node recovers at rate  $\gamma$ . The arrival processes are all independent of one another, and a recovered node is assumed to remain recovered indefinitely.

Although this description is in principle sufficient to simulate an epidemic on an arbitrary network, it becomes intractable if the number of nodes grows beyond a small size. For this reason we adopt the *mean-field* approach outlined in Chapters 4 and 5 of Kiss, Miller and Simon [1] to approximate the laws of motion of certain population-level, or aggregate, quantities of interest. Intuitively, we are primarily concerned with the numbers of infected and susceptible agents, rather than their identity, and so are interested in the network structure only insofar as it aids our understanding of these population aggregates.

To understand this approach we first introduce some more notation. Fix a network with  $N$  nodes and adjacency matrix  $(g_{ij})_{i,j=1}^N$  and write  $X_i(t)$  for the health status of node  $i$  at date  $t$ . For any possible health

status  $A$  the expected number of nodes of type  $i$  at date  $t$  is then

$$[A](t) = \sum_{i=1}^N P(X_i(t) = A). \quad (1)$$

For any pair of health statuses  $A$  and  $B$ , we write

$$[AB](t) = \sum_{i=1}^N \sum_{j=1}^N g_{ij} P(X_i(t) = A, X_j(t) = B) \quad (2)$$

for the number of pairs in the network connected by a link in which the first node has status  $A$  and the second has status  $B$ . Similarly, for any three statuses  $A, B$  and  $C$  we write

$$[ABC](t) := \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N g_{ij} g_{jk} P(X_i(t) = A, X_j(t) = B, X_k(t) = C) \quad (3)$$

for the expected number of arrangements of triples in which the statuses of the first, second and third nodes are  $A, B$  and  $C$ , respectively. The first observation necessary for the simulation of an SIR epidemic on networks is the following result characterizing the law of motion of quantities of the form (1) and (2).

**Theorem 1** (Theorem 4.4 of Kiss [1]). *For an SIR model on an arbitrary network, we have*

$$\begin{aligned} [\dot{S}] &= -\tau[SI] \\ [\dot{I}] &= \tau[SI] - \gamma[I] \\ [\dot{R}] &= \gamma[I] \\ [\dot{SI}] &= -\gamma[SI] + \tau([SSI] - [ISI] - [SI]) \\ [\dot{SS}] &= -2\tau[SSI] \\ [\dot{II}] &= 2\tau([SI] + [ISI]) - 2\gamma[II]. \end{aligned} \quad (4)$$

*Proof sketch.* We will only provide a heuristic derivation, and refer the reader to the literature for details. The first-order terms  $[\dot{S}]$ ,  $[\dot{I}]$  and  $[\dot{R}]$  are all standard and so are omitted. For the  $[SI]$  term note that

$$[SI](t + \Delta_t) = [SI](t) - \Delta_t \gamma[SI](t) - \Delta_t \tau([ISI] + [SI]) + \Delta_t \tau[SSI] + o(\Delta_t). \quad (5)$$

One can interpret the terms on the right-hand side of (5) as follows: the mass of  $(S, I)$  pairs at date  $t$ , minus the mass of  $(S, I)$  pairs for which the second (infected) node recovers, minus the mass of  $(S, I)$  pairs for which the first (susceptible) node is infected by the first node of a  $(I, S, I)$  triple, minus the mass of  $(S, I)$  for which the latter infects the former, plus the mass of new  $(S, I)$ , that were previously the first two pairs of a  $(S, S, I)$  triple. For the  $[\dot{SS}]$  term note that  $[SS](t + \Delta_t) = [SS](t) - \Delta_t \tau[SSI](t) - \Delta_t \tau[ISS](t) + o(\Delta_t)$ , which we interpret as the preceding number of  $(S, S)$  pairs minus the mass of  $(S, S, I)$  triples for which the third infects the second, minus the mass of  $(I, S, S)$  triples for which the first infects the second. We then use the identity  $[SSI] = [ISS]$  in order to derive the expression in (4). Finally, for the law of  $[II]$ , first note that there are four ways one may construct new  $(I, I)$  pairs: from  $(S, I)$  and  $(I, S)$  pairs, in which one infects the other, or from  $(I, S, I)$  triples in which either the first or the third infects the middle node. Second, the number of  $(I, I)$  pairs is reduced whenever one of the two nodes recovers. The final expression in (4) follows from these observations together with the identity  $[SI] = [IS]$ .  $\square$

## Homogeneous networks with no clustering

The law of motion given in Theorem 1 holds for an arbitrary network. However, it is not immediately suitable to generate simulations of an SIR pandemic because it involves the terms  $[SSI]$  and  $[ISI]$  for which no law of motion is specified. In the terminology of the literature of epidemics on networks, we say that the system is not *closed*, because the law of motion of the expected numbers of nodes and pairs of given health statuses depends upon variables outside of this set. We therefore seek a *closure* of the system by approximating some of the higher-order moments in terms of lower-order quantities.

This procedure can be applied at any level of the approximation. For instance, one can obtain a closed system of equations by taking the first three equations in (4) and using the approximation  $[SI] \approx (n/N)[S][I]$  for some integer  $n$ . Since the sole second-order moment  $[SI]$  is replaced with the first-order moments  $[S]$  and  $[I]$  the system is said to be “closed at the level of pairs”. This would produce the familiar triple of equations corresponding to the baseline SIR model, and amounts to assuming that infected nodes are uniformly distributed throughout the network. Since we are primarily interested in illustrating *departures* from the baseline SIR model, we instead pursue a closure “at the level of triples”, by considering the whole system in (4) and positing approximations to  $[SSI]$  and  $[ISI]$ .

The choice of this approximation will depend upon the phenomenon we wish to capture. First suppose that the graph is homogeneous, meaning that every node has the same degree  $n$ . The total number of edges emanating from susceptible nodes is denoted  $n[S]$ , and the total number of edges connecting susceptible nodes to infected is  $[SI]$ . The proportion of edges emanating from a susceptible node connected to an infected node is  $[SI]/(n[S])$ , while the proportion of edges connected to a susceptible node is  $[SS]/(n[S])$ . If the edges connecting a susceptible and an infected node are approximately uniformly distributed amongst all edges in the network, then an approximation to  $[SSI]$  may be obtained by noting that for a given susceptible node  $u$ , the probability that any two neighbour nodes  $v$  and  $w$  together form an  $(S, S, I)$  triple is  $[SS][SI]/(n^2[S]^2)$ . Since there are  $n(n-1)$  such pairs of  $v$  and  $w$  for any susceptible node and  $[S]$  susceptible nodes, we obtain the approximation

$$[SSI] = \frac{(n-1)}{n} \frac{[SS][SI]}{[S]}. \quad (6)$$

Similarly, under the same assumption that edges are approximately uniformly distributed we obtain

$$[ISI] = \frac{(n-1)}{n} \frac{[SI]^2}{[S]}. \quad (7)$$

The closure given by the pair of equations (6) and (7) is a natural starting point but is unable to capture the possibility that infected nodes are likely to be “clustered” together. Since the distribution of edges was assumed to be uniform, as  $n$  is held fixed and  $N$  grows large the probability that for a given  $(S, I)$  pair, the  $n-1$  remaining neighbors of the susceptible node are connected to the infected node approaches zero. For this reason we augment the above equations in a parsimonious way to incorporate *clustering*. The clustering coefficient for a network is defined to be the ratio of open triangles to closed triangles, and so represents the probability that two neighbors of a given node are connected to one another. Writing  $\|G\| := \sum_{i=1}^N \sum_{j=1}^N g_{ij}$  and  $\text{Tr}(G) = \sum_{i=1}^N g_{ii}$ , one may show that this ratio is given by

$$\phi := \frac{\text{Tr}(G^3)}{\|G^2\| - \text{Tr}(G^2)}. \quad (8)$$

Now fix an  $(S, I)$  and note that of the  $n - 1$  additional nodes connected to the susceptible node, a fraction  $1 - \phi$  of them are not connected to the infected node, and a fraction  $\phi$  are connected. For the former set, we reason as above and observe that the number of such nodes that are susceptible (and hence form a  $(S, S, I)$ ) triple is  $(1 - \phi)(n - 1)[SS]/(n[S])$ . For the remaining  $\phi(n - 1)$  nodes that are also connected with the infected node in the given pair, we suppose that the expected number of susceptible nodes is multiplied by a factor  $C_{SI}$ . In this way we obtain the approximation

$$[SSI] \approx \frac{(n - 1)}{n} \frac{[SI][SS]}{[S]} (1 - \phi + \phi C_{SI}). \quad (9)$$

Similarly, the number of  $[ISI]$  triples is approximately given by

$$[ISI] \approx \frac{(n - 1)}{n} \frac{[SI]^2}{[S]} (1 - \phi + \phi C_{II}) \quad (10)$$

for a correlation coefficient  $C_{II}$ . To completely specify the dynamics of the system it remains to choose  $C_{SI}$  and  $C_{II}$ . There are several choices that are common in the literature. One natural choice is

$$(C_{SI}, C_{II}) = \left( \frac{N}{n} \frac{[SI]}{[S][I]}, \frac{N}{n} \frac{[II]}{[I]^2} \right) \quad (11)$$

as this choice quantifies the extent to which the number of  $(S, I)$  pairs differs from that which would prevail if the nodes were uniformly distributed,  $[SI] = [S][I]n/N$ , and similarly for  $(I, I)$  pairs. One technical issue with (11) is that the implied closed system will violate the consistency requirements  $[SS] + [SI] = n[S]$  and  $[II] + [IS] = n[I]$  that must hold for any graph. House and Keeling [2] show that one may correct this problem by dividing  $(C_{SI}, C_{II})$  in (9) and (10) by  $C_{SI}[SS]/(n[S]) + C_{II}[SI]/(n[S])$ . Simplification gives the closure

$$\begin{aligned} [SSI] &\approx \frac{(n - 1)}{n} \frac{[SS][SI]}{[S]} \left( 1 - \phi + \phi \frac{n[S][I][SI]}{([SS][I] + [S][II])[SI]} \right) \\ [ISI] &\approx \frac{(n - 1)}{n} \frac{[SI]^2}{[S]} \left( 1 - \phi + \phi \frac{n[S]^2[II]}{([SS][I] + [S][II])[SI]} \right) \end{aligned}$$

To see this, note that the first coefficient becomes

$$\begin{aligned} \frac{n[S]C_{SI}}{C_{SI}[SS] + C_{II}[SI]} &= \frac{n[S] \frac{N}{n} \frac{[SI]}{[S][I]}}{\frac{N}{n} \frac{[SI]}{[S][I]} [SS] + \frac{N}{n} \frac{[II]}{[I]^2} [SI]} \\ &= \frac{n[SI]/[I]}{\frac{[SI]}{[S][I]} [SS] + \frac{[II]}{[I]^2} [SI]} \\ &= \frac{n[SI][S][I]}{[SI][I][SS] + [II][S][SI]}. \end{aligned}$$

For the second equation we have

$$\begin{aligned} \frac{n[S]C_{II}}{C_{SI}[SS] + C_{II}[SI]} &= \frac{n[S]C_{SI}(C_{II}/C_{SI})}{C_{SI}[SS] + C_{II}[SI]} \\ &= \frac{n[SI][S][I]}{[SI][I][SS] + [II][S][SI]} \frac{\frac{N}{n} \frac{[II]}{[I]^2}}{\frac{N}{n} \frac{[SI]}{[S][I]}} \\ &= \frac{n[SI][S][I]}{[SI][I][SS] + [II][S][SI]} \frac{[II][S]}{[I][SI]} \\ &= \frac{n[S]^2[II]}{([I][SS] + [II][S])[SI]}. \end{aligned}$$

When combined with Theorem 1 the above expressions provide us with a closed system of equations that we may use to produce the graphs pertaining to clustering in the main text.

## Heterogeneous networks with no clustering

The clustering model of the previous section assumed that all nodes in the network had the same degree. We now generalize this to allow for the possibility that different nodes have different degrees, which necessitates some additional notation. Suppose again that there are  $N$  nodes in the network and that the degrees of these nodes assume  $M$  distinct values. For each  $k = 1, \dots, M$  write  $N_k$  for the number of nodes of degree  $k$ , so that  $N_1 + \dots + N_M = N$ , and for each node  $i = 1, \dots, N$  write  $\text{deg}(i)$  for the degree of the  $i$ th node. For any  $k = 1, \dots, M$  and health status  $A$ , write

$$[A_k](t) = \sum_{i:\text{deg}(i)=k} P(X_i(t) = A) \quad (12)$$

for the expected number of nodes with degree  $k$  and status  $A$ . Similarly, for any integers  $1 \leq k, l \leq M$  and health statuses  $A$  and  $B$ , write

$$[A_k B_l](t) = \sum_{i:\text{deg}(i)=k} \sum_{j:\text{deg}(j)=l} g_{ij} P(X_i(t) = A, X_j(t) = B) \quad (13)$$

for the expected number of  $(A, B)$  pairs in which the first node has status  $A$  and the second has status  $B$ . If one of the indices in (13) is omitted this indicates that index is summed over, e.g.  $[A_k B](t) = \sum_{l=1}^M [A_k B_l](t)$ . The following is then a straightforward generalization of Theorem 1.

**Theorem 2** (Theorem 5.2 of Kiss, Miller and Simon [1]). *For an SIR model on an arbitrary network with  $M$  distinct node degrees, we have*

$$\begin{aligned} [\dot{S}_k] &= -\tau[S_k I] \\ [\dot{I}_k] &= \tau[S_k I] - \gamma[I_k] \\ [\dot{R}_k] &= \gamma[I_k] \end{aligned} \quad (14)$$

for all  $k = 1, \dots, M$ .

Note that for each  $k = 1, \dots, M$  we may sum the terms in (14) to obtain  $[\dot{S}_k] + [\dot{I}_k] + [\dot{R}_k] = 0$ , and so the expected number of nodes of each degree is constant over time. For simplicity, the main text supposes that  $M = 2$ , so that the population can be divided ex-ante into two distinct groups, named hermits and friendlies, who differ in the number of links they have with other people. The hermits and friendlies will be denoted with subscripts 1 and 2, respectively, so that  $S_1(t), S_2(t), I_1(t)$  and  $I_2(t)$  for the masses of agents at time  $t$  who are susceptible and hermetic, susceptible and friendly, infected and hermetic, and infected and friendly, respectively. If the (constant) masses of each type are denoted by  $M_1$  and  $M_2$ , then the number of recovered agents of each type may be determined residually by the identities

$$\begin{aligned} R_1(t) &= M_1 - S_1(t) - I_1(t) \\ R_2(t) &= M_2 - S_2(t) - I_2(t). \end{aligned}$$

Note that analogously to Theorem 1, the system in Theorem 2 is exact but not closed. To obtain a system of equations suitable for simulation, for  $k, l = 1, 2$  we close the system by choosing

$$[S_k I_l] = \frac{n_k^2}{(M_k n_k + M_l n_l)} [S_k][I_l] \quad (15)$$

where  $n_1$  and  $n_2$  denote the degree (i.e. number of links emanating from) associated with hermits and friendlies, respectively. The specification in (15) may be viewed as the simplest departure from the baseline

SIR model that incorporates heterogeneity of links. To see this, note that if one of the populations vanishes, then the resulting closed system reduces to a homogeneous SIR model closed at the level of pairs. e.g. if  $M_2 = 0$  then the expression in (15) becomes  $[S_1 I_2] = (n_1/M_1)[S_1][I_1]$  and  $[S_2 I_2] = [S_1 I_2] = [S_1 I_2] = 0$ , which is exactly the familiar baseline SIR model.

## References

- [1] Kiss, István Z., Joel C. Miller, and Péter L. Simon. "Mathematics of epidemics on networks." Cham: Springer 598 (2017).
- [2] House, Thomas, and Matt J. Keeling. "The impact of contact tracing in clustered populations." PLoS Comput Biol 6.3 (2010): e1000721.