Spring 2022 I606 Network Epidemiology

David Terner
Github Repository

April 5th, 2022

A note to the grader

I completed this assignment using R and three libraries: `ggplot2` for visualizations; `tidyverse` for general data wrangling; and `igraph` to create and to manipulate network/graph objects.

I chose R over Python for this assignment because: (i) I know how to vectorize in R but not in Python; (ii) testing and visualizing different parameterizations of β, μ, and/or i_0 is easier for me to do in R; and, (iii) I have limited time to complete this assignment, and while I would like to overcome issues (i) and (ii) with Python someday, today is not that day.

All of my code can be found here in my Github repository. I have two main files: `sis_functions.R`, which is where my SIS model is coded-up alongside a handful of helper functions, and `sis_script.R` which is used for different tests/analyses required by the assignment.

Question 1

Please refer to my `compartment_mod` in `sis_functions.R` for my SIS model. The code is largely inspired by the template provided in the assignment. A few comments are in order. Major differences include:

- I vectorize counting the number of infected neighbors for each susceptible node (line 44). I then simulate infection in one step (line 45).
- SIS models differ from SI models by allowing for recovery at rate μ. I simulate recovery via running p_t Bernouli trails where p_t is the number of infected nodes at time t (line 39).

Question 2

Plot your model's results using a sparse random graph and play with the parameters to get a feel for how μ, β, and i_0 change the equilibrium point of the system. The equilibrium point occurs when the system settles on a stable fraction of infected (see Fig 10.7 in Barabasi's textbook). Also take note of how long it takes for the system to reach equilibrium.

Solution

Barabasi's textbook provides a closed form solution for infection rate $i(t)$ for the SIS model given transmission rate β, recovery rate μ and initial infection share i_0:

$$i(t) = \left(1 - \frac{\mu}{\beta\langle k\rangle}\right)\left(\frac{\left(\frac{i_0}{1-i_0-\frac{\mu}{\beta\langle k\rangle}}\right)e^{(\beta\langle k\rangle-\mu)t}}{1+\left(\frac{i_0}{1-i_0-\frac{\mu}{\beta\langle k\rangle}}\right)e^{(\beta\langle k\rangle-\mu)t}}\right) \tag{1}$$

In the limiting case, we can show that:

$$\lim_{t\to\infty} i(t) = i(\infty) = 1 - \frac{\mu}{\beta\langle k\rangle} = 1 - \frac{1}{R_0} \geq 0 \iff R_0 \geq 1 \text{ otherwise } i(\infty) = 0 \tag{2}$$

Effects on equilibrium infection rates

Having a closed form expression allows us to see how changing i_0, β, and μ affect $i(\infty)$:

$$\begin{aligned}\frac{\partial i(\infty)}{\partial \beta} &= \frac{\mu}{(\beta\langle k\rangle)^2} > 0 \\ \frac{\partial i(\infty)}{\partial \mu} &= -\frac{1}{\beta\langle k\rangle} < 0 \\ \frac{\partial i(\infty)}{\partial i_0} &= 0\end{aligned}$$

Effect on time to equilibrium

One of the questions asks in this section asks how changes in i_0 impact changes in $i(\infty)$. By definition of equilibrium, given explicitly by (2), $i(\infty)$ doesn't depend on starting conditions. Trivially the greater the distance between starting conditions and equilibrium, $|i_0 - i(\infty)|$, the longer it will take for the disease to arrive at equilibrium. Denote time to equilibrium as $t(\infty)$. Further, define $\epsilon(t)$ as the distance between current infection rates and equilibrium infection rates:

$$\epsilon(t) = \Big(i(t) - i(\infty)\Big)^2 = \left(i(\infty)\frac{\frac{i_0}{i(\infty)-i_0}\exp\{(\beta\langle k\rangle - \mu)t\}}{1 + \frac{i_0}{i(\infty)-i_0}\exp\{(\beta\langle k\rangle - \mu)t\}} - i(\infty)\right)^2 \tag{3}$$

Solving for t:

$$t(\epsilon) = \frac{1}{\beta\langle k\rangle - \mu}\ln\Big(\frac{\big(\beta\epsilon\langle k\rangle + \sqrt{\beta^2\epsilon\langle k\rangle^2 - 2\beta\epsilon\langle k\rangle\mu + \epsilon\mu^2}\big)\big(\beta i_0\langle k\rangle - \beta\langle k\rangle + \mu\big)}{(\epsilon\langle k\rangle^2 i_0\beta^2)}\Big)$$

$t(\epsilon)$ tells us how long it will take to arrive at state infection that is $\epsilon > 0$ (squared) distance away from $i(\infty)$. Moreover, we can define time to equilibrium as:

$$t(\infty) = \lim_{\epsilon\to 0} t(\epsilon)$$

By setting ϵ arbitrarily small, we can use $t(\epsilon)$ to analyze how changes in our parameters affect time to equilibrium.
Whereas changing i_0, β, and μ have simple and predictable changes on $i(\infty)$, changing these same

parameters have complicated impacts on $t(\epsilon)$:

$$\begin{aligned}
\frac{\partial t(\varepsilon)}{\partial \beta} =& -\Bigg(\Big(\beta\langle k\rangle(\langle k\rangle-(i_0-1)\beta+\mu)\big(\langle k\rangle\beta\sqrt{\epsilon(\beta\langle k\rangle-\mu)^2}+(\beta\langle k\rangle-\mu)^2\big)\\
&\times\ln\Big[\frac{(\langle k\rangle(i_0-1)\beta+\mu)(\beta\epsilon\langle k\rangle+\sqrt{\epsilon(\beta\langle k\rangle-\mu)^2}}{\epsilon\langle k\rangle^2 i_0}\Big]+\beta\langle k\rangle\Big(\beta\langle k\rangle(\langle k\rangle(i_0-1)\beta+\mu)\ln(\beta^{-2})+\mu(\beta\langle k\rangle-\mu)\Big)\\
&\times\sqrt{\epsilon(\beta\langle k\rangle-\mu)^2}+\Big(\beta\langle k\rangle\big(\langle k\rangle(i_0-1)\beta+\mu\big)\ln(\beta^{-2})-\mu\big(\langle k\rangle(i_0-2)\beta+2\mu\big)\Big)(\beta\langle k\rangle-\mu)^2\Big)\varepsilon\Bigg)\\
&\times\Bigg(\Big(\sqrt{\epsilon(\beta\langle k\rangle-\mu)^2}\beta(\beta\langle k\rangle\epsilon+\sqrt{\epsilon(\beta\langle k\rangle-\mu^2}^{\,2})\Big)\big(\langle k\rangle(i_0-1)\beta+\mu\big)(\beta\langle k\rangle-\mu)^2\Bigg)^{-1}\\
\frac{\partial t(\epsilon)}{\partial \mu} =&\epsilon i_0\beta^2\langle k\rangle^2\Big(\big(\epsilon\beta\langle k\rangle+\sqrt{\epsilon\beta^2\langle k\rangle^2-2\beta\langle k\rangle\epsilon\mu+\epsilon\mu^2}\big)(\beta\langle k\rangle i_0-\beta\langle k\rangle+\mu)(\beta\langle k\rangle-)\Big)^{-1}\\
&\times\Bigg(\frac{(-2\beta\langle k\rangle\epsilon+2\epsilon\mu)(\beta\langle k\rangle i_0-\beta\langle k\rangle+\mu)}{2\epsilon i_0\beta^2\langle k\rangle^2\sqrt{\epsilon\beta^2\langle k\rangle^2-2\beta\langle k\rangle\epsilon+\epsilon\mu^2}}+\frac{\epsilon\beta\langle k\rangle+\sqrt{\epsilon\beta^2\langle k\rangle^2-2\beta\langle k\rangle\mu\epsilon+\epsilon\mu^2}}{\epsilon i_0\beta^2\langle k\rangle^2}\Bigg)\\
&+\frac{1}{(\beta\langle k\rangle-\mu)^2}\ln\Big(\frac{\big(\epsilon\beta\langle k\rangle+\sqrt{\epsilon\beta^2\langle k\rangle^2-2\beta\epsilon\mu+\epsilon^2}\big)(i_0\beta\langle k\rangle-\beta\langle k\rangle+\mu}{\epsilon i_0\beta^2\langle k\rangle^2}\Big)\\
\frac{\partial t(\epsilon)}{\partial i_0} =&\epsilon i_0\beta^2\langle k\rangle^2\Big(\frac{\varepsilon\beta\langle k\rangle\sqrt{\epsilon\beta^2\langle k\rangle^2-2\beta\langle k\rangle\mu\epsilon+\epsilon\mu^2}}{\epsilon i_0\beta\langle k\rangle}-\frac{\big(\epsilon\beta\langle k\rangle+\sqrt{\epsilon\beta^2\langle k\rangle^2-2\beta\langle k\rangle\mu\epsilon+\epsilon\mu^2}\big)(\beta\langle k\rangle i_0-\beta\langle k\rangle+\mu)}{\varepsilon\beta^2\langle k\rangle^2 i_0^2}\Big)\\
&\times\Big(\Big(\beta\langle k\rangle\epsilon+\sqrt{\epsilon\beta^2\langle k\rangle^2-2\beta\langle k\rangle\mu\epsilon+\epsilon\mu^2}\Big)(\beta\langle k\rangle i_0-\beta\langle k\rangle\mu)(\beta\langle k\rangle-\mu)\Big)^{-1}
\end{aligned}$$

The main point of these messy equations is that to argue that trying to make general statements about how changes in i_0, β, and/or μ affect $t(\epsilon)$/time to equilibrium is *fraught with difficulty*.

Equilibrium in practice

All preceding work, namely equations (1) and (2), are derived under the assumption that the network is sufficiently large. As a result, an ad-hoc definition/approach to computing (time to) equilibrium is necessary. In particular, I define a cumulative moving average of infection rates:

$$m(t) = \frac{1}{t}\sum_{\tau=0}^{t} i(t) \tag{4}$$

Time to equilibrium, t^* is thus defined as:

$$t^* \in \arg\min \text{ s.t. } |m(t) - m(t-1)| \leq \delta$$

where δ is set by the user.[1] Likewise, I define equilibrium as $m(t^*)$.
In practice, I use averaged values of $i(t)$ per t across 50 trials per (β, μ, i_o) for my definition of $m(t)$.

[1] In practice, I set $\delta = 10^{-3}$.

Coding Results

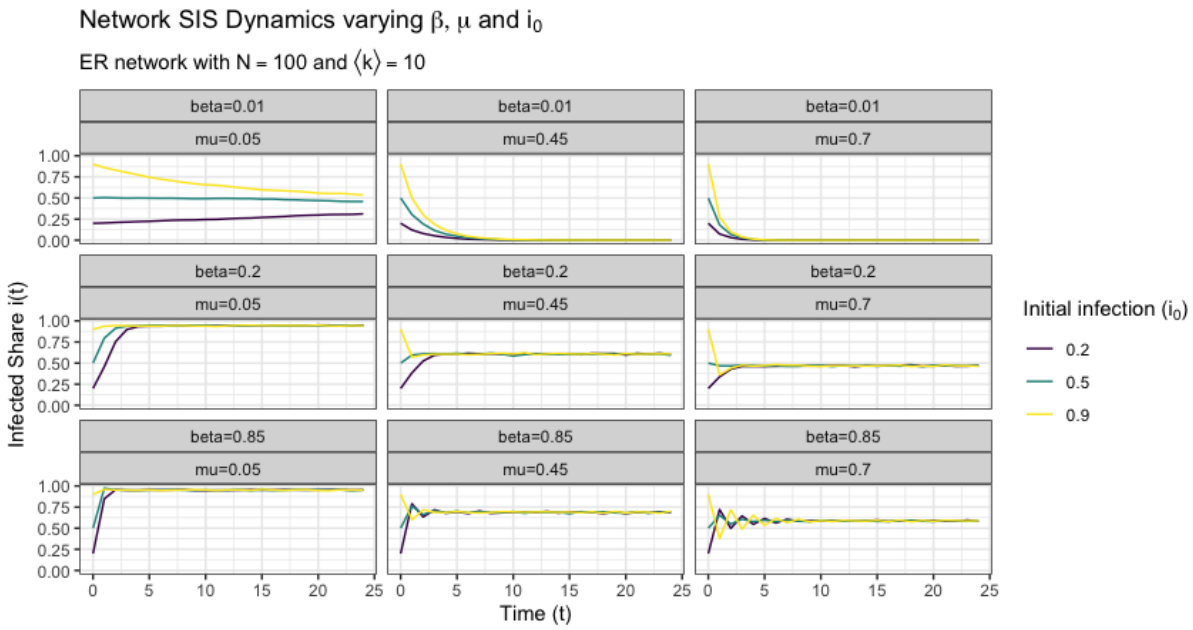


Figure 1: SIS Dynamics

Figure 1 shows how the SIS model works with different paramaterizations of β, μ and i_0. The top left panel ($\beta = 0.01$ and $\mu = 0.05$) suggests that even after 25 time steps that system is not in equilibrium. However, when β is increased to $\beta = 0.85$ (bottom left panel), then regardless of i_0, the $i(t)$ quickly converges to $i(\infty)$. In the case of $\beta = 0.01$ and $\mu = 0.45$ (or $\mu = 0.7$) since $\frac{\mu}{\beta\langle k\rangle} > 1$, the $i(\infty) = 0$.

Figure 2 reveals a common-sense truth: the closer a i_0 is to $i(\infty)$, the less time it takes for $i(t)$ to reach $i(\infty)$. When $i_0 < i(\infty)$, $i(t)$ grows with respect to time. In contrast, when $i_0 > i(\infty)$ as in the case of $i_0 = 0.95$, $i(t)$ decreases almost in a roughly monotonic fashion. However, since $|i_0 = 0.95 - i(\infty)| < |i_0 = 0.05 - i(\infty)|$, i_0 converges to equilibrium far faster.

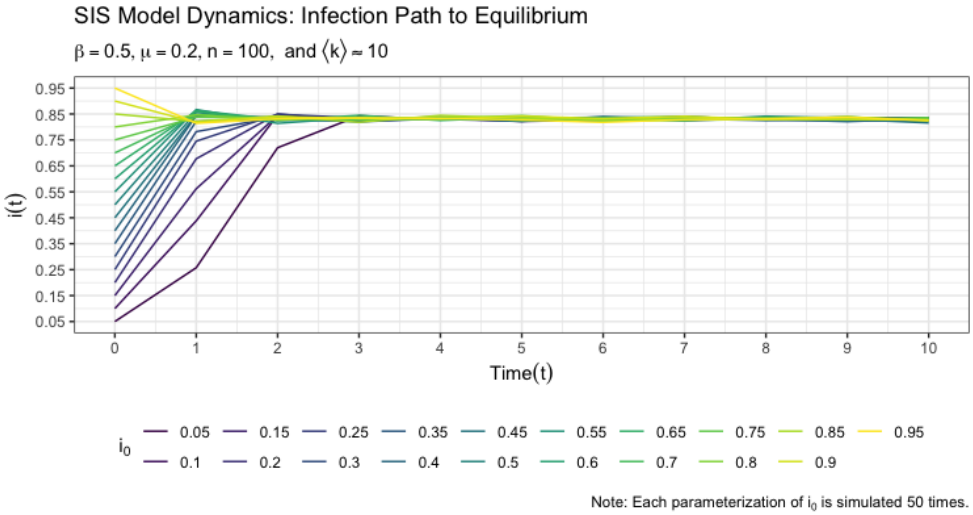


Figure 2: Infection Path Varying i_0

Lastly, Figure 3 provides some kind of estimate of equilibrium values based on the trial data displayed by Figure 1. What are the notable features of this chart? For the most part, across all β and μ, i_0 does not affect $m(t^*)$ (e.g. per panel, notice that there is little variation in $m(t^*)$/ vertical axis values). However, that farther i_0 is from $m(t^*)$, the higher $t^*(i_0)$ is. For example, consider Figure 3's top left panel: the $i_0 = 0.5$ configuration places the system virtually at equilibrium, whereas the $i_0 = 0.2$ configuration does not converge to $m(t^*)$ in under 100 iterations.

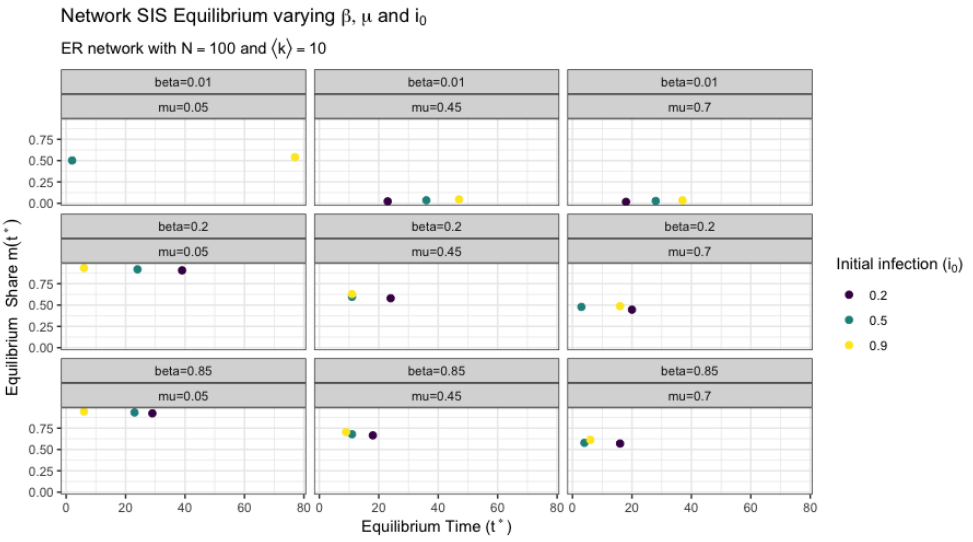


Figure 3: Equilibrium infection time t^* and share $m(t^*)$

Question 3

Solution

Appealing to the Barabasi's textbook section "SIS Model and the Vanishing Epidemic Threshold," define the characteristic time of an SIS epidemic, τ^{SIS}, or the time that it takes the epidemic to infect $1/e$ share of the population, as:

$$\tau^{SIS} = \frac{\langle k \rangle}{\beta \langle k^2 \rangle - \mu \langle k \rangle}$$

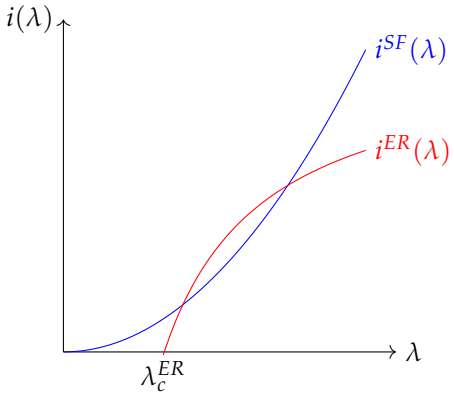
If we define λ as β/μ and solve for λ such that $\tau^{SIS} > 0$, then we will have found the epidemic threshold λ_c. Note that for a random graph that $\langle k^2 \rangle = \langle k \rangle (\langle k \rangle + 1)$. Thus, the epidemic threshold for an arbitrary network, such as a scale free graph, or an ER graph is:

$$\lambda_c^{SF} = \frac{\langle k \rangle}{\langle k^2 \rangle} \implies \lambda_c^{ER} = \frac{1}{\langle k \rangle + 1} \tag{5}$$

For the purposes of this assignment, we are supposed to generate a power law network with scaling coefficient $\gamma = 2.5$. Since $\gamma < 3$, this means that the scale-free network's $\langle k^2 \rangle$ is undefined, which in turn implies that $\lambda_c^{SF} = 0$.
We can plot the final infection share, $i(\infty)$, as function of λ. For the ER graph, $i^{ER}(\lambda)$ follows from (2),

$$i^{ER}(\lambda) = \begin{cases} 0 & \text{if } \lambda < \lambda_c^{ER} = \frac{1}{\langle k \rangle + 1} \\ 1 - \frac{1}{\lambda \langle k \rangle} & \text{otherwise} \end{cases}$$

Thankfully, Barabasi's textbook provides us with the fact that $i^{SF}(\lambda) \sim \lambda^{\frac{1}{3-\gamma}}$ when $\gamma < 3$.



Coding Results

To complete the final portion of the assignment, I first generate a sequence of 40-λ's from 0.01 to 0.40. I set $\beta_i = \beta = 0.01$ for all i. Then per each λ_i, I set $\mu_i = \beta/\lambda_i$. Per (β_i, μ_i), I then run 50 trials

on both an ER/random graph and scale-free graph. Each graph has $n = 100$ nodes, $\langle k \rangle \approx 2$, and in the case of the scale-free graph, I set $\gamma = 2.5$. As another practical matter, I compute $i(\infty)$ by averaging the last 20 (out of 100) time iterations across all trials for a given λ_i.

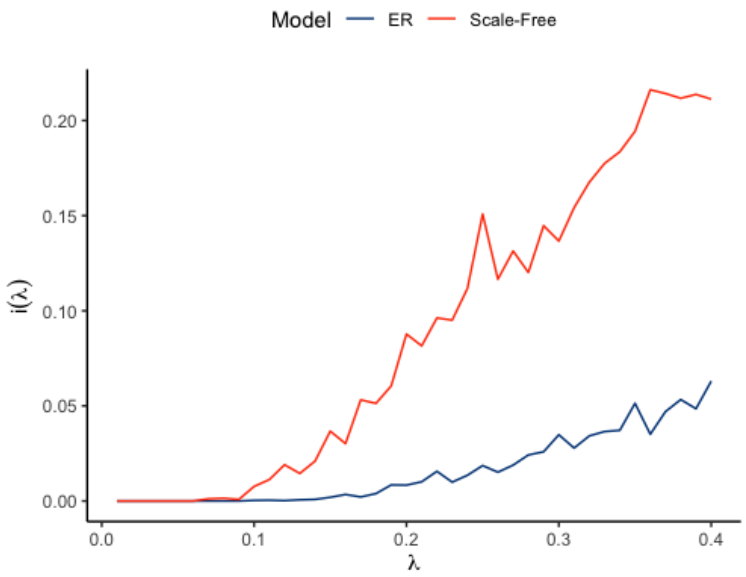


Figure 4: Barabasi's Image 10.11 Reconstructed

Figure 4 presents the results of the final exercise. While the difference between ER and SF epidemic thresholds is relatively small (due to working with small networks), the difference is nonetheless apparent. $\lambda = 0.07$ is the smallest value of λ such that $i(\lambda^{SF}) > 0$. By contrast, $\lambda = 0.10$ is the smallest value of λ such that $i(\lambda^{ER}) > 0$. Suppose that we take a looser view of what we mean by an epidemic threshold, namely that we want to find that smallest λ such that $i(\lambda) \geq 1\%$. With this in mind, $\lambda^{SF} = 0.11$ whereas $\lambda^{ER} = 0.21$.