Inferring Epidemic Parameters in Networked SIS Epidemics

Navid Azizan Ruhi
California Institute of Technology
ACM 158 Project
June 2, 2017

1 Introduction

The mathematical modeling and analysis of epidemic spread is of great importance in understating how epidemics behave and how to control them, and it has attracted significant interest from different communities. The study of epidemics applies to many areas such as epidemiology [1], rumor spread [2, 3], information propagation [4, 5], network security [6, 7], and viral marketing [8, 9]. Although there is a huge body of work on epidemic models, classical ones mostly neglect the underlying network structure and assume a uniformly mixed population, which is obviously far from reality. However, in recent years more realistic networked models have been introduced, and many interesting results are now known about them [10, 11, 12]. The most well-known of these is the networked Susceptible-Infected-Susceptible (SIS) model, in which each node is in one of the two states, susceptible (S) or infected (I). During any time interval, each susceptible (healthy) node has a chance of being independently infected by any of its infected neighbors (with probability $\beta$). Further, during any time interval, each infected node has a chance of recovering (with probability $\delta$) and becoming susceptible again. We discuss the problem of estimating epidemic parameters of the SIS model ($\beta$ and $\delta$) using observations.

2 Networked SIS Epidemics

Let $G = (V, E)$ be an arbitrary connected undirected network with $n$ nodes, and with adjacency matrix $A$. Each node can be in a state of health, represented by 0, or a state of infection, represented by 1. The state of the entire network can be represented by a binary $n$-tuple $\xi(t) = (\xi_1(t), \cdots, \xi_n(t)) \in \{0, 1\}^n$, where each of the entries represents the state of a node at time $t$, i.e. $i$ is infected if $\xi_i(t) = 1$ and it is healthy if $\xi_i(t) = 0$.

Given the current state $\xi(t)$, the next state that the network transitions to, $\xi(t+1)$, is independent of the past, which means that the system can be described as a Markov chain with $2^n$ states. Furthermore, the infection probability of each node in the next step is determined independently conditioned on the current state. Therefore the transition matrix $S$ of this Markov Chain has elements $S_{X,Y} = \Pr(\xi(t+1) = Y | \xi(t) = X)$ of the following form:

$$\Pr(\xi(t+1) = Y | \xi(t) = X) = \prod_{i=1}^{n} \Pr(\xi_i(t+1) = Y_i | \xi(t) = X),$$

for any two state vectors $X, Y \in \{0, 1\}^n$. 

1
As mentioned before, a healthy node can receive infection from any of its infected neighbors independently with probability $\beta$ per infected link, and an infected node can recover from the disease with probability $\delta$ (Fig. 1). Therefore, each term in the product in (1) can be written as

$$P(\xi_i(t+1) = Y_i | \xi(t) = X) = \begin{cases} 
(1 - \beta)^m_i & \text{if } (X_i, Y_i) = (0, 0), |N_i \cap S(X)| = m_i, \\
1 - (1 - \beta)^m_i & \text{if } (X_i, Y_i) = (0, 1), |N_i \cap S(X)| = m_i, \\
\delta & \text{if } (X_i, Y_i) = (1, 0), |N_i \cap S(X)| = m_i, \\
1 - \delta & \text{if } (X_i, Y_i) = (1, 1), |N_i \cap S(X)| = m_i, 
\end{cases}$$

where $S(X)$ is the support of $X \in \{0, 1\}^n$, i.e. $S(X) = \{i | X_i = 1\}$, and $N_i$ is the set of neighbors of node $i$.

It is known that this epidemic model exhibits a phase transition behavior at a certain threshold [13, 14], i.e., once the effective infection rate $\tau = \frac{\beta}{\delta}$ approaches a critical value $\tau_c$ [10] the epidemic appears not to die out. We should remark that this Markov chain has a unique absorbing state, which is the all-healthy state, because once the system reaches this state it remains there forever since there are no infected nodes to propagate infections. This means that if we wait long enough the epidemic will eventually die out, which may seem to be odd at first. However, what this means is that the question of the epidemic dying out is not interesting; what is interesting is the question of how long it takes for the epidemic to die out. In particular, if the mixing time of the Markov chain is exponentially large, one will not see it dying out in any reasonable time. Therefore the right question to ask is what is the mixing time of the Markov chain (or, equivalently, its mean time to absorption); it turns out that the threshold $\tau_c$ corresponds to the phase transition between “slow mixing” (exponential time) and “fast mixing” (logarithmic time) of the MC [15, 16, 17]. Fig. 2 demonstrates this phase transition for an Erdős-Rényi random graph.

One important implication of this result is that, if one can estimate the values of $\beta$ and $\delta$ using data, we will be able to predict the behavior of the epidemic and its eradication time. For this reason, we consider the problem of inferring the epidemic parameters using data.

### 3 Estimating Epidemic Parameters

In this section, we discuss the inference of the epidemic parameters, $\beta$ and $\delta$, based on a time-series observation of the states of the nodes. We assume that we only have the data for two consecutive time steps $t$ and $t+1$. Clearly, if we have access to more data in the time series, we can do the
same procedure as will be described here for any $t$, and average over them, which should provide more reliable estimates. Let us denote the state at time $t$ by $X \in \{0,1\}^n$, and at time $t+1$ by $Y \in \{0,1\}^n$, and drop the time index for simplicity. As mentioned before, conditioned on the current state $X$, the next state of the individual nodes $Y_i$ are independent. We define sets $\mathcal{N}_S$ and $\mathcal{N}_I$ as the set of susceptible and infected nodes at time $t$, respectively. Mathematically,

$$\mathcal{N}_S = \{i \mid X_i = 0\}, \quad \text{and} \quad \mathcal{N}_I = \{i \mid X_i = 1\}. \quad (3)$$

The likelihood of the data can then be written as

$$P(Y | \mathcal{N}_S \cup \mathcal{N}_I; \beta, \delta) = P(Y | \mathcal{N}_S; \beta)P(Y | \mathcal{N}_I; \delta)P(X), \quad (4)$$

where $Y_{\mathcal{N}_S}$ ($Y_{\mathcal{N}_I}$) denotes the elements of $Y$ indexed by $\mathcal{N}_S$ ($\mathcal{N}_I$). Therefore the maximum-likelihood (ML) estimate of $\beta$ and $\delta$ can be computed separately.

### 3.1 Maximum-likelihood Estimate of Recovery Rate $\delta$

Estimating $\delta$ is relatively straightforward, thus we do it first. By defining sets $\mathcal{N}_{IS}$ and $\mathcal{N}_{II}$ as

$$\mathcal{N}_{IS} = \{i \mid X_i = 1, Y_i = 0\},$$

$$\mathcal{N}_{II} = \{i \mid X_i = 1, Y_i = 1\}, \quad (6) (7)$$

we can write the likelihood as

$$P(Y_{\mathcal{N}_I} | X; \delta) = \delta^{\lvert \mathcal{N}_{IS} \rvert}(1-\delta)^{\lvert \mathcal{N}_{II} \rvert}. \quad (8)$$

To find the ML estimator, we need to find $\delta$ that maximizes this function. The log-likelihood can be simply written as

$$\log P(Y_{\mathcal{N}_I} | X; \delta) = \mathcal{N}_{IS} \log \delta + \mathcal{N}_{II} \log(1-\delta). \quad (9)$$
Taking its derivative with respect to \( \delta \) and setting to zero yields \( \frac{|N_{IS}|}{\delta} = \frac{|N_{II}|}{1-\delta} \), or

\[
\hat{\delta} = \frac{|N_{IS}|}{|N_{IS}| + |N_{II}|},
\]

which is what is expected intuitively.

### 3.2 Maximum-Likelihood Estimate of Infection Rate \( \beta \)

We now find the ML estimate of the infection rate \( \beta \), which is a bit more involved. Similar to our definitions earlier, let us define sets \( N_{SS} \) and \( N_{SI} \) as

\[
N_{SS} = \{i \mid X_i = 0, Y_i = 0\}
\]

(11)

\[
N_{SI} = \{i \mid X_i = 0, Y_i = 1\}.
\]

(12)

The other piece of the likelihood function (5), corresponding to \( Y_{N_S} \), can be expressed as

\[
P(Y_{N_S} | X; \beta) = \prod_{i \in N_{SS}} (1 - \beta)^{m_i} \prod_{j \in N_{SI}} (1 - (1 - \beta)^{m_j}),
\]

(13)

where \( m_i \) denotes the number of infected neighbors of node \( i \), i.e. \( m_i = |N_i \cap N_I| \). Equivalently, the log-likelihood is

\[
\log P(Y_{N_S} | X; \beta) = \sum_{i \in N_{SS}} m_i \log(1 - \beta) + \sum_{j \in N_{SI}} \log(1 - (1 - \beta)^{m_j}).
\]

(14)

Finding the optimizer of this expression is hard, because of the term inside the second summation. However, given that \( \beta \) is typically much smaller than 1, we can use the approximation \( (1 - \beta)^{m_j} \approx 1 - m_j \beta \). The log-likelihood can then be written as

\[
\log P(Y_{N_S} | X; \beta) \approx \log(1 - \beta) \sum_{i \in N_{SS}} m_i + \sum_{j \in N_{SI}} \log(m_j \beta).
\]

Setting the derivative, with respect to \( \beta \), equal to zero yields

\[
\frac{1}{1 - \beta} \sum_{i \in N_{SS}} m_i \approx \frac{|N_{SI}|}{\beta}
\]

which leads to

\[
\hat{\beta} \approx \frac{|N_{SI}|}{|N_{SI}| + \sum_{i \in N_{SS}} m_i}.
\]

(15)

Another approach to approximate the ML estimator would be to first take the derivative of (14) and then use the approximation \( (1 - \beta)^{m_j} \approx 1 - m_j \beta \). In other words, by setting the derivative to zero we have

\[
\frac{1}{1 - \beta} \sum_{i \in N_{SS}} m_i = \sum_{j \in N_{SI}} \frac{m_j (1 - \beta)^{m_j - 1}}{1 - (1 - \beta)^{m_j}},
\]

or

\[
\sum_{i \in N_{SS}} m_i = \sum_{j \in N_{SI}} \frac{m_j (1 - \beta)^{m_j}}{1 - (1 - \beta)^{m_j}}.
\]
Now using \((1 - \beta)^{m_j} \simeq 1 - m_j \beta\), this reduces to
\[
\sum_{i \in N_{SS}} m_i \simeq \sum_{j \in N_{SI}} \frac{1 - \beta m_j}{\beta}.
\]
Multiplying both sides by \(\beta\) and noting that \(N_{SI} \cup N_{SS} = N_S\), leads to
\[
\hat{\beta} \simeq \frac{|N_{SI}|}{\sum_{j \in N_{SI}} m_j + \sum_{i \in N_{SS}} m_i} = \frac{|N_{SI}|}{\sum_{i \in N_S} m_i},
\]
which is another approximation for the ML estimator (Notice the term \(\sum_{j \in N_{SI}} m_j\) in the denominator, instead of \(N_{SI}\) as in (15).)

4 Numerical Results

In this section, we compute the estimators derived above using observations from an SIS epidemic on an Erdős-Rényi random graph with \(n = 2000\) nodes. Table 1 shows the estimates for a number of different epidemic parameters. As one can see, the estimates are very close to the true values. We should remark that these values have been estimated using observations of only two time steps; as one observes more time steps, the estimator would become more accurate. The MATLAB code can be found in Appendix A.

Table 1: The values of estimates for a number of different epidemic parameters of an SIS epidemic on an Erdős-Rényi random graph with \(n = 2000\) nodes.

<table>
<thead>
<tr>
<th>True Values</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\delta)</td>
<td>(\hat{\delta})</td>
</tr>
<tr>
<td>0.8 0.05</td>
<td>0.778 0.0375 0.0595</td>
</tr>
<tr>
<td>0.8 0.01</td>
<td>0.831 0.0105 0.0095</td>
</tr>
<tr>
<td>0.5 0.02</td>
<td>0.513 0.0197 0.0165</td>
</tr>
<tr>
<td>0.2 0.01</td>
<td>0.190 0.0104 0.0094</td>
</tr>
<tr>
<td>0.1 0.005</td>
<td>0.102 0.0056 0.0053</td>
</tr>
</tbody>
</table>

5 Conclusion and Outlook

In summary, we found the maximum likelihood estimator of the epidemic parameters of the networked SIS model. This allows us to predict whether an epidemic dies out quickly or not. Numerical results confirm that the estimator is fairly accurate, even using limited observations. As future work, one may consider estimating the epidemic parameters using partial observations (not observing all the neighbors of a node) and also more complicated models such as SIRS, SIV, etc.
References


in exact Markovian SIR and SIS epidemics on networks,” in Decision and Control (CDC), 2014
A MATLAB Code

```matlab
1 clear

2

3 % Network
4 % Erdos–Renyi Random Graph
5 n = 2000;
6 p = 0.01; % 2*log(n)/n;
7 A = rand(n,n) < p;
8 A = triu(A,1);
9 A = A + A';
10 lambda = abs(eigs(A,1));

11 % Parameters
12 delta = 0.8;
13 beta = 0.05;

14 % Stochastic Process
15 % x is the state at time t.
16 % y is the state at time t+1.
17 x = [ones(n/2,1); % first half: infected
18      zeros(n/2,1)]; % second half: susceptible
19 y = zeros(n,1);

20 for i = 1:n
21    y(i) = x(i);
22    switch x(i)
23      case 0
24        m = sum(A(:,i).*x == 1);
25        if rand >((1-beta)^m)
26            y(i) = 1;
27        end
28      case 1
29        if rand >1-delta
30            y(i) = 0;
31        end
32    end
33 end

34 y1 = y(1:n/2); % state of the nodes that were infected
35 y2 = y(n/2+1:end); % state of the nodes that were susceptible

36 % Parameter Estimation (delta)
37 delta_est = length(find(y1==0))/length(y1)
```

8
Parameter Estimation ($\beta$)

```matlab
num = sum(A.*repmat(x',n,1),2); % number of infected neighbors of each node at time t
num2 = num(n/2+1:end); % number of infected neighbors of each susceptible node at time t

N_SI = find(y2); % (index of) the nodes that were susceptible and got infected
N_SS = find(1-y2); % (index of) the nodes that were susceptible and remained susceptible

beta_est1 = length(N_SI)/(length(N_SI)+sum(num2(N_SS)))
beta_est2 = length(N_SI)/(sum(num2(N_SI))+sum(num2(N_SS)))
```