

# M1 Writeup

March 10, 2015

## Contents

<b>1</b>	<b>Notation</b>	<b>2</b>
<b>2</b>	<b>System dynamics model</b>	<b>3</b>
2.1	Gravity model . . . . .	3
2.2	Model fit diagnostics . . . . .	3
<b>3</b>	<b>Structure generation</b>	<b>4</b>
3.1	Structures . . . . .	4
3.1.1	Random Network . . . . .	4
3.1.2	Starred Alley . . . . .	4
3.1.3	Ring . . . . .	6
3.1.4	Grid . . . . .	7

3.1.5	Bowtie . . . . .	8
3.1.6	Scalefree . . . . .	9
3.2	Covariates . . . . .	10
3.3	Network distance . . . . .	11
<b>4</b>	<b>Simulations</b>	<b>12</b>
<b>5</b>	<b>Results</b>	<b>12</b>
5.1	Simultaneous perturbation experiment . . . . .	12
5.2	Agent jitter and num chunks experiment . . . . .	12
5.3	Toy Structures . . . . .	12
5.4	WNS . . . . .	12

# 1 Notation

Each location  $i \in \mathcal{L}$  lies in a  $R^2$  at coordinate pair  $C_i$ . The matrix  $\Omega$  is a binary matrix indicating adjacent locations. For any two locations  $i$  and  $j$ ,  $\Omega_{i,j} = 1$  if location  $j$  is adjacent to location  $i$ . All networks defined will be undirected thus  $\Omega_{i,j} = \Omega_{j,i}$  for all  $i, j$ . The adjacency matrix is the defining component of the network. For some of the networks, the adjacency matrix can be defined without knowing the coordinate pairs, but for other networks the coordinate pairs are required to generate the adjacency matrix. Regardless of the network, we outline every step to generate the network in its entirety.

## 2 System dynamics model

### 2.1 Gravity model

To model the evolution of the disease, we model the probability that an infected location  $i$  infects an uninfected location  $j$  at time point  $t$ . We denote this probability as  $p_{i,j,t}$ . If we assume independence across events  $i$  infecting  $j$  and  $i'$  infecting  $j$ , the probability of  $j$  becoming infected at time  $t$  is

$$P(X_{j,t} = 1 | X_{\cdot,t}) = X_{j,t-1} \bigvee \left[ 1 - \prod_{i: X_{i,t-1}=1} (1 - p_{i,j,t}) \right].$$

To define  $p_{i,j,t}$  we use a form of the gravity model.

$$\text{logit } p_{i,j,t} = \beta_0 + \beta_1 U_j - \alpha \frac{d_{i,j}}{(c_i c_j)^\nu} - \eta g_j(\tau_{i,t}) - \rho A_{j,t-1} - \gamma A_{i,t-1}$$

where  $\tau_{i,t} = \sum_{t'=0}^t X_{i,t'}$  is the number of time points  $i$  has been infected. We compare different forms of  $g_j(\cdot)$  discussed later. The name of the model comes from  $d_{i,j}/(c_i, c_j)^\nu$  which is known as the gravity term. The effect of distance between two locations is diminished if the number of caves in the two locations is large. We assume as the number of caves increases, so does the number of bats effectively decreasing the effect of distance.

For model estimation, we use a Bayesian framework. The priors on the parameters are as follows  $\beta_0 \sim N(0, 100)$ ,  $\beta_1 \sim N(0, 10)$ ,  $\log(\alpha) \sim N(0, 1)$ ,  $\log(\nu) \sim N(0, 1)$ ,  $\rho \sim N(0, 1)$ ,  $\gamma \sim N(0, 1)$ .

### 2.2 Model fit diagnostics

When estimating the posterior distributions, we compared multiple forms for  $g(\cdot)$ .

- $g_j(x) = 0$
- $g_j(x) = x - 1$

- $g_j(x) = \exp(x - 1) - 1$
- $g_j(x) = \exp(\frac{\max_k c_k + 1}{c_j + 1}(x - 1)) - 1$

### 3 Structure generation

#### 3.1 Structures

##### 3.1.1 Random Network

The random network generates location coordinates uniformly on  $[0, 1] \times [0, 1]$ . Let  $C$  denote the set of coordinate pairs. For any  $i \in \{1, \dots, n\}$ ,  $C_i \in [0, 1] \times [0, 1]$ .

Let  $A$  be the  $n \times n$  adjacency matrix where  $A_{i,j} = 1$  if locations  $i$  and  $j$  are adjacent and  $A_{i,j} = 0$  otherwise. To determine which locations are connected, we must first define some notation. Let  $E_{i,j} = \|C_i - C_j\|_2$  and  $E_{i,[j]}$  denote the  $j^{th}$  largest norm excluding  $E_{i,i}$  which is zero. The set of  $k$  nearby locations to location  $i$  is  $N_i^{(k)} = \{j : E_{i,j} \leq E_{i,[k]}\}$ . Now, the adjacency matrix is defined as  $A_{i,j} = 1$  if  $i \in N_j^{(k)}$  or  $j \in N_i^{(k)}$ .

##### 3.1.2 Starred Alley

The starred alleyway was produced to incorporate the importance of network centrality metrics. An example of a starred alleyway can be found in figure 2. By looking at the structure, the goal is to identify the locations along the main horizontal line. If these locations are treated, it helps to contain the spread of the disease.

Distances between nodes are not defined in the usual way. For any node on the outer points, the distance to the closest node on the main horizontal line is 0.9 instead of 1.0.

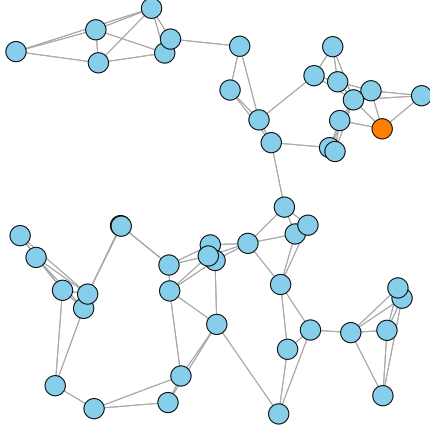


Figure 1: A random network with 50 locations

For a starred alley of size  $N$ , define

$$m_N = \arg \max_{m > 0} f(m) \mathbb{1}_{\{f(m) \leq N\}}$$

where

$$f(m) = m + \left\lceil \frac{m}{2} \right\rceil \left( \left\lceil \frac{m}{2} \right\rceil - (m \bmod 2) + 1 \right).$$

Let locations  $1, \dots, m_N$  index the locations on the main horizontal line. Next define disjoint sets of locations  $\mathcal{N}_i$  for  $i = 1, \dots, m_N$ . The cardinality for  $\mathcal{N}_i$  is defined as

$$|\mathcal{N}_i| = \left\lfloor \frac{i}{2} \right\rfloor + \mathbb{1}_{\{(m_N - i) < (N - f(m_N))\}}.$$

For each location  $j \in \mathcal{N}_i$  set  $A_{i,j} = A_{j,i} = 1$ .

To define the coordinates for each location, first set  $C_i = \{i, 0\}$  for  $i = 1, \dots, m_N$ . Then for each

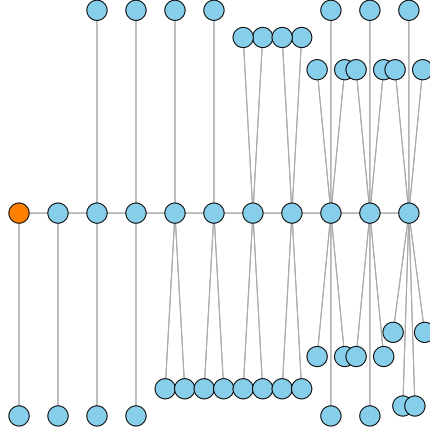


Figure 2: The starred alleyway network with 50 locations

### 3.1.3 Ring

The ring is a simple circle of locations with a concentrated group in one section. To form the ring, place the locations evenly on the circle then between one pair of locations insert more locations evenly. An example can be seen in figure 3.

Divide the number of locations  $N$  into two groups. Define  $N_1 = N - N_2$  where  $N_2 = \lceil N * 0.05 \rceil$ . Let  $\theta = 2\pi/(N_1 + 1)$  and  $r = 1/(2 \sin(\theta/2))$ . Place location 1 at  $C_1 = \{0, r\}$ . Given location  $i - 1$  at  $C_{i-1}$ , place location  $i$  at  $C_i = \text{rot}(\theta)C_{i-1}$  for  $i = 2, \dots, N_1$  and

$$\text{rot}(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

Note that  $C_1 \cdot C_{N_1} = \|C_1\| \|C_{N_1}\| \cos(2\theta)$ .

With the first  $N_1$  locations defined, we need to define the last  $N_2$  locations. Place location  $N_1 + i$  at  $C_{N_1+i} = \text{rot}(\theta/N_2)C_{N_1+i-1}$  for  $i = 1, \dots, N_2$ . Note that  $C_N \cdot C_{N_1} = \|C_N\| \|C_{N_1}\| \cos(\theta)$  and  $C_1 \cdot C_N = \|C_1\| \|C_N\| \cos(\theta)$ .

To define the adjacency matrix,  $A_{i,i+1} = A_{i+1,i} = 1$  for  $i = 1, \dots, N - 1$  and  $A_{1,N} = A_{N,1} = 1$ .

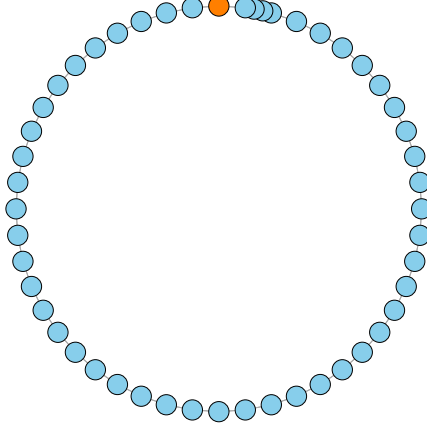


Figure 3: The ring network with 50 locations

### 3.1.4 Grid

The grid is a checkerboard-like structure. First define  $N_1$  to be the number of columns and  $N_2$  to be the number of rows. These values are defined by

$$\{N_1, N_2\} = \arg \min_{\substack{N_1, N_2 > 0 \\ N_1 N_2 = N \\ N_1 \leq N_2}} |N_1 - N_2|$$

Location  $i$  is positioned at  $C_i = \{(i \bmod N_2), \lfloor i/N_2 \rfloor\}$ . The adjacency matrix is defined as  $A_{i,j} = 1$  if  $\|C_i - C_j\| = 1$ .

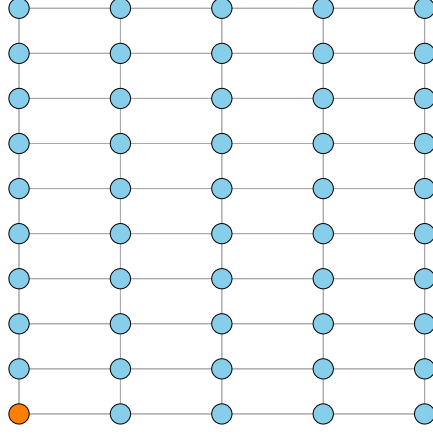


Figure 4: The grid network with 50 locations

### 3.1.5 Bowtie

The bowtie combines two grid networks with a random network in between. Spread of the disease begins in one grid and travels through the random network and into the other grid.

For a bowtie network of size  $N$ , create one random network of size  $N_R = \lceil 0.1N \rceil + ((N - \lceil 0.1N \rceil) \bmod 2)$  and two grid networks of size  $N_G = (N - N_R)/2$ .

Let  $i_R$  be the  $i^{th}$  location in the random network and  $i_{G_j}$  be the  $i^{th}$  location in the  $j^{th}$  grid.

To connect the three networks, we need to first identify which nodes in each network will connect to one of the other two networks. We will begin with the first grid. First, define  $N_1$  and  $N_2$  to be defined in the same way as in the grid description. Both grid networks are the same size, we do not need to distinguish between the two. Let  $M_G = \lceil N_2/3 \rceil$  be the number of points to connect from



the grid networks. Define the following four sets of locations

$$\begin{aligned}
\Lambda_{G_1} &= \{N_G - 2N_1M_G + kN_1 : k = 1, \dots, M_G\} \\
\Lambda_{G_2} &= \{N_G - 2N_1M_G + 1 + (k-1)N_1 : k = 1, \dots, M_G\} \\
\Lambda_{R_1} &= \{i : C_{i,1} \leq C_{[M_R],1}\} \\
\Lambda_{R_2} &= \{i : C_{i,1} \geq C_{[N_R-M_R+1],1}\}
\end{aligned}$$

where  $[\cdot]$  denotes the usual order statistics.

To define the adjacency matrix, combine the adjacency matrices from the three separate networks as a block diagonal matrix. To define the additional edges, set  $A_{i_{G_1},j_R} = A_{j_R,i_{G_1}} = 1$  for all  $i \in \Lambda_{G_1}$  and  $j \in \Lambda_{R_1}$  to connect the first grid to the random network. Then set  $A_{i_{G_2},j_R} = A_{j_R,i_{G_2}} = 1$  for all  $i \in \Lambda_{G_2}$  and  $j \in \Lambda_{R_2}$  to connect the second grid to the random network.

Now that the adjacency matrix is defined, we need to finish the network. The last step is scaling and adjusting the coordinates from the three networks. We define this as an ordered list of operations.

1. For  $i = 1, \dots, N_R$ :  $C_{i_R} = C_{i_R}/2$ .
2. Define  $S = (\max_i C_{i_R,1} - \min_i C_{i_R,1})/2$ .
3. For  $i = 1, \dots, N_R$ :  $C_{i_R,1} = C_{i_R,1} - \min_j C_{j_R,1} + \max_j C_{j_{G_1},1} + S$
4. For  $i = 1, \dots, N_G$ :  $C_{i_{G_2},1} = C_{i_{G_2},1} - \min_j C_{j_{G_2},1} + \max_j C_{j_R,1} + S$ .

### 3.1.6 Scalefree

The scalefree network is designed so that some locations are highly connected while others are only connected by a couple edges. This type of network was originally designed to mimic the network of webpages. The internet houses many websites that are highly connected (e.g. Google, Yahoo, etc.) and others that are not.

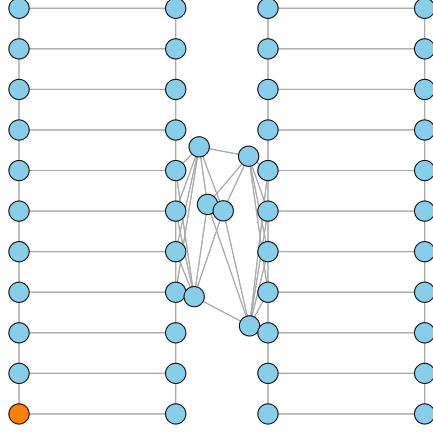


Figure 5: The bowtie network with 50 locations

To generate the network, we will define the process via induction. A network of size 1 is the trivial network. Given a network of size  $N - 1$ , attach the additional  $N^{th}$  node to existing node  $i \in \{1, \dots, N - 1\}$  with probability

$$\frac{\sum_{j \neq i} A_{j,i}}{\sum_{j,k : j > k} A_{j,k}}.$$

Generating a scale-free network in this fashion results in a Barabasi-Network. The defining feature of a scale-free network is the proportion of nodes with  $k$  edges is proportional to  $k^{-\gamma}$  for some  $\gamma$ . In the case of a Barabasi-Network  $\gamma = 3$ .

Given the network, we use the Fruchterman-Reingold algorithm to position the nodes in a two dimensional plane.

### 3.2 Covariates

Define  $X$  to be a  $n \times p$  centered and scaled covariate matrix. The covariates are multivariate normal  $MVN_p(\mu, \Sigma)$  where  $\mu_i = 2\tilde{C}_{i,1}^2 + 1\tilde{C}_{i,2}$  and  $\tilde{C}$  are the centered and scaled coordinates.

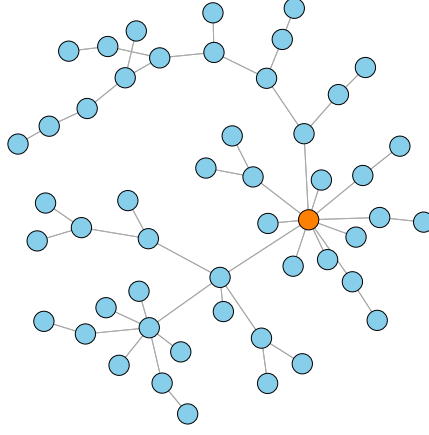


Figure 6: The scalefree network with 50 locations

The covariance matrix has the structure  $\Sigma_{ip+s,jp+t} = \rho e^{(-\tau E_{i,j} - \eta |s-t|)}$ . Thus, the variance of the individual covariates are 1 and the correlation decays as  $E_{i,j}$  increases and as  $|s - t|$  increases.

We need the mass of each individual node to compute the gravity term in the gravity model. In the case of White-Nose Syndrome, the mass is the number of caves. Arbitrarily set the first covariate to be the mass of the node. Define  $M_i = \lfloor X_{i,1} - X_{[1],1} \rfloor + 1$  to be the mass of location  $i$ .

### 3.3 Network distance

In order for the structure of the network to have a strong impact on the disease spread, we define distance as the geodesic distance raised to a power. The distance between adjacent nodes ( $i$  and  $j$  such that  $A_{i,j} = 1$ ) is defined to be 1. For non-adjacent nodes, the distance is the number of edges along the shortest path raised to a power. To determine the appropriate power, let  $p_{i,j}^{(1)}$  be the hypothetical probability node  $i$  infects node  $j$  where  $i$  and  $j$  have geodesic distance 1. Let  $p_{i,j}^{(2)}$  be the hypothetical probability of infection if  $i$  and  $j$  have geodesic distance 2. For their mass, use

the mean mass of the network. The log odds-ratio is equal to

$$-\frac{\alpha}{m^\rho} + \frac{\alpha 2^z}{m^\rho}.$$

Set this equation equal to  $\log(0.5)$  and solve for  $z$ .

## 4 Simulations

## 5 Results

### 5.1 Simultaneous perturbation experiment

### 5.2 Agent jitter and num chunks experiment

### 5.3 Toy Structures

### 5.4 WNS