How Infections Spread on Networks

CS 523: Complex Adaptive Systems

Assignment 4: Due Dec. 2, 2015 5:00 pm

1 Introduction

In this assignment you will investigate the spread of disease on networks. In class you learned about the Susceptible Infected Recovery (SIR) model of infection spread and how ordinary differential equations (ODE) can be used to characterize its behavior.

$$\begin{aligned}\dot{S} &= -\beta SI \\ \dot{I} &= \beta SI - \delta I \\ \dot{R} &= \delta I\end{aligned}$$

The ODE model assumes that populations are well mixed and that any agent in the system has an equal likelihood of interacting with any other. In some cases this may be a poor assumption, specifically in cases where agents are interacting through networks. For this assignment you will simulate SIR dynamics on three different networks and measure properties of the resulting infection. You will then compute some common network properties to determine which nodes in a network are most important for spreading infections.

Please hand in a printed copy of a 3-4 page paper describing your assignment, formatted according to the ACM SIG Proceedings format, available from `http://www.acm.org/sigs/publications/proceedings-templates`, together with the output, and the code you wrote to compete the assignment. I prefer to receive a hard copy, but if you submit your assignment electronically, please send the paper in pdf format rather than docx or tex.

2 Network Data

Three different undirected networks are available on the course website.

- `Celegans.g`: A metabolic network of the nematode *Caenorhabditis elegans*[4]
- `Jazz.g`: A network of collaborations among Jazz Musicians[6]
- `ASDec2014.g`: A network of connections between ISPs (Internet Service Providers), collected during the last week of December 2014.[5]

The files are formatted similarly to the files for Assignment three, except they omit the first line indicating the number of colors need to color the graph. Therefore the files simply contain a list of edges represented as ordered pairs (the two connecting nodes separated by a space) with one edge per line, e.g.,

```
1 3
0 2
2 1
```

You can assume that each node is labeled with a unique non-negative integer starting with 0.

It is highly recommended that you use existing network analysis software to conduct your investigations. There are many packages available for free, and you will want to spend some time investigating what is available before you begin working on the assignment. If you are using Python, consider NetworkX [3] or graph-tool [1]. If using R, consider i-graph [2]. Be sure to report what packages you used for your analysis.

Before investigating the network dynamics, first create Table 1, which reports some of the basic data about each network, including the number of nodes, the number of edges, and the average node degree.

3 SIR Dynamics on Networks

In this section you will implement and describe SIR dynamics on the three different networks. You should simulate the dynamics synchronously as follows:

1. Initialize all nodes as susceptible (S);
2. Select a single node to begin the infection;
3. While there are any infected nodes **do**:
   (a) For each infected node u in the previous step
      i. For each neighbor v of u: If v is susceptible set v to infected (on the next step) with probability β;
   (b) Set u to recovered on the next step;

Note: We assume that nodes are infected for exactly one time step before they move into a recovered state, i.e. $\delta = 1$.

3.1 Infections over time

For each of the three input graphs, simulate the SIR dynamics starting with a single node (choose a node that has close to the median degree for the network). Record the fraction of infected nodes in the network at each time point for three different values of β. Because the SIR dynamics are stochastic, you will want to simulate each infection multiple times with the same starting node. Plot the average of these runs over time, including the standard deviation using error bars for each value of β.

Discuss how the infection curves compare to the behavior seen in the Ordinary Differential Equation Model. First, you will want to refresh your memory about the behavior of the

ODE SIR model presented in class, either by reading the literature or by conducting some simulations of the ODE model. Elaborate on the differences between each network infection curve and what you think that it implies about the structure of the network.

3.2 Characterizing the Infection

Next, we will investigate how infections spread for various values of β. Select at least 20 different values of β. Simulate the SIR dynamics on the network starting with a random node, measuring the total proportion of the network that becomes infected, the time to clear the infection (all nodes are susceptible or removed), and the time to the largest number of infected nodes (time to peak). Be sure to simulate the infection enough times (each run from a different randomly chosen starting node) that you can reasonably estimate mean and standard deviation of each of these measures (at least 100 runs). For each measure plot the β values (horizontal axis) against the measure, including the mean and error bars for the standard deviation. To save space, make one plot for each measure, including a separate line (labeled appropriately) for each network. In your writeup, discuss what each measure tells you about the three different networks.

4 Influential Spreaders

It might be important to know which nodes in the network are most capable of spreading disease, for example to identify the most effective way to stop the spread of an infections, or conversely, the best way to spread information through a network. We will measure how influential a node is by measuring the average proportion of the network that becomes infected when the infection starts with that node. Using $\beta = 0.2$ measure the mean infection size when started from each node in each network. Once again be sure to run at least 100 simulations for each node. In a table, report the most and least influential nodes, and the average size of the infection they create.

4.1 Network Measures to Identify Influential Spreaders

Rather than simply running simulations it is often useful to study the topological structure of the network to identify influential spreaders. Using the software package of your choice, compute each of the following measures for each node in the network:

- **Degree:** Number of edges connected to the node;
- **Average Shortest Path Length:** For each node u compute the shortest path to all other nodes v, and take the average of their lengths (one average shortest path computed for each node);
- **Betweenness Centrality:** The fraction of all shortest paths a node in the network participates in.

For each network, make a scatter plot where each point is the (x, y) pair, where x is each of the above measures for a single node, and y is the average infection size when the infection starts at that node.

Which of these measures provides the best prediction of infection size? Why does each perform well or poorly? Investigate other network measures in the literature and see if you can find one that is better at identifying influential spreaders. Be sure to cite your sources.

5 Extra Credit

There are several opportunities for extra credit.

- Identify and implement other types of spreading dynamics on the network, for example SIS (Susceptible-Infected-Susceptible) or Rumour Dynamics [7]. Repeat part or all of the above analysis using these different dynamics.
- Investigate how spreading dynamics change when different nodes are removed from the networks. How many nodes do you have to remove from each network (and in what order) to achieve a significant reduction in infection? Define what you mean by significant reduction. Be sure to report all parameters.
- Some topological measures might be better at identifying influential spreaders for different values of β. Investigate other values of β and determine which measures are best for each.
- For assignment 2 you generated some random graphs. Y Investigate infection spreading on random graphs with various average degrees. Additionally, implement or find implementations of other random graph generators. Investigate spreading dynamics on these graphs.

6 What to hand in

Hand in a short report (not more than five pages) that describes your implementation and results. Remember to cite all of your sources using a consistent citation style and include a proper bibliography.

Please include a listing of your code and instructions for how to run it as an appendix to your report.

7 Late Policy

You are allowed three free "late days" to be used at your discretion throughout the semester. After you have used up your late days, I will deduct 10% per day from the grade you would have received on any late work. Since the projects will become more involved throughout the semester, I strongly recommend that you save your late days for when you really need them.

References

[1] graph-tool. `https://graph-tool.skewed.de/`.

[2] igraph. `http://igraph.org/redirect.html`.

[3] Networkx. `https://networkx.github.io/`.

[4] Jordi Duch and Alex Arenas. Community detection in complex networks using extremal optimization. *Physical review E*, 72(2):027104, 2005.

[5] Benjamin Edwards, Steven Hofmeyr, Stephanie Forrest, and Michel Van Eeten. Analyzing and modeling longitudinal security data: Promise and pitfalls. *Proceedings of ACSAC*, 2015.

[6] Pablo M Gleiser and Leon Danon. Community structure in jazz. *Advances in complex systems*, 6(04):565–573, 2003.

[7] Maziar Nekovee, Yamir Moreno, G Bianconi, and M Marsili. Theory of rumour spreading in complex social networks. *Physica A: Statistical Mechanics and its Applications*, 374(1):457–470, 2007.