Network Software Modelling Assignment 2

Students:	*Student #:*
Louis Carnec	15204934
Vijay Katta	15202724
Adedayo Adelowokan	15204151

April 22, 2017

Master of Science (Business Analytics)
UCD Michael Smurfit Graduate Business School
2016-2017

1 Read world phenomenon : Spread of disease through airports network

The world is closely connected than ever before, through transportation networks. Air, sea and land transportation continues to expand in reach, speed of travel and volume of passengers and goods carried. Pathogens and their vectors can now move further, faster and in greater numbers than ever before [3]. Epidemics can occur more easily and the emergence of novel pathogens exacerbates the situation [1]. Understanding the dispersal behaviour and identifying the outbreak of an epidemic in the preliminary stages enables critical response planning. Network topology properties surrounding the debut location explains much of the early stage variation in the spread of diseases [2].

In this simulation we will model the spread of disease through world airports network. It is inspired by the work of Yager and Taylor [4] which models the spread of pathogens through airports using a directed graph. We extend their work of simulating disease spread by selectively infecting the nodes to analyse the impact of graph/node properties such as diameter, average path length , centrality and clustering.

2 Describe briefly your choice of graph model and how it corresponds to the real-world phenomenon, e.g. your choice of directed versus undirected edges, edge weights, allowing or disallowing self-loops, etc.

Our real-world network graph is undirected where nodes represent airports and edges are routes.We used open source data from openflights.org along with population of the cities from unstats.un.org. Redundant edges and unconnected nodes are removed from the graph. Average of population of two cities where airpots are located is calculated and normalised before assigning it as edge weight. Airports where both origin and destination are with-in USA are used, and others are removed.

3 With the aid of a diagram, describe the simulation rules, e.g. how agents change state and what causes them to send messages.

To simulate the propagation of a disease through the airports network, we have selected four states - susceptible, infected, recovered and removed. Initial state of all nodes is susceptible and any node can be infected. There is no incubation period, a node once infected can infect neighbours with a probability of edge weight. An infected node recovers with a random probability which decreases as the number of steps increase. Alternatively a node with infection is removed from the network indefinitely. A node once infected and recovered is immune to infection and not susceptible any more.

Nodes are coloured based on the status - green is susceptible, yellow is infected, blue is recovered and red is removed/dead. Figure 1 below has 5 blocks each one represents the status of nodes at one time step. Status of nodes in step 1 is shown by block *(a)* , where node A is infected and is yellow , all the other nodes green showing that they are susceptible. In step 2 nodes B and D are infected with probabilities w_{AB} and w_{DA} respectively, as they are the neighbours of node A. Step 3 shows that all nodes are infected and in 4th time step, nodes B and D are red showing that they are dead/removed indefinitely. Node C is blue in the last step representing that it has recovered from infection.
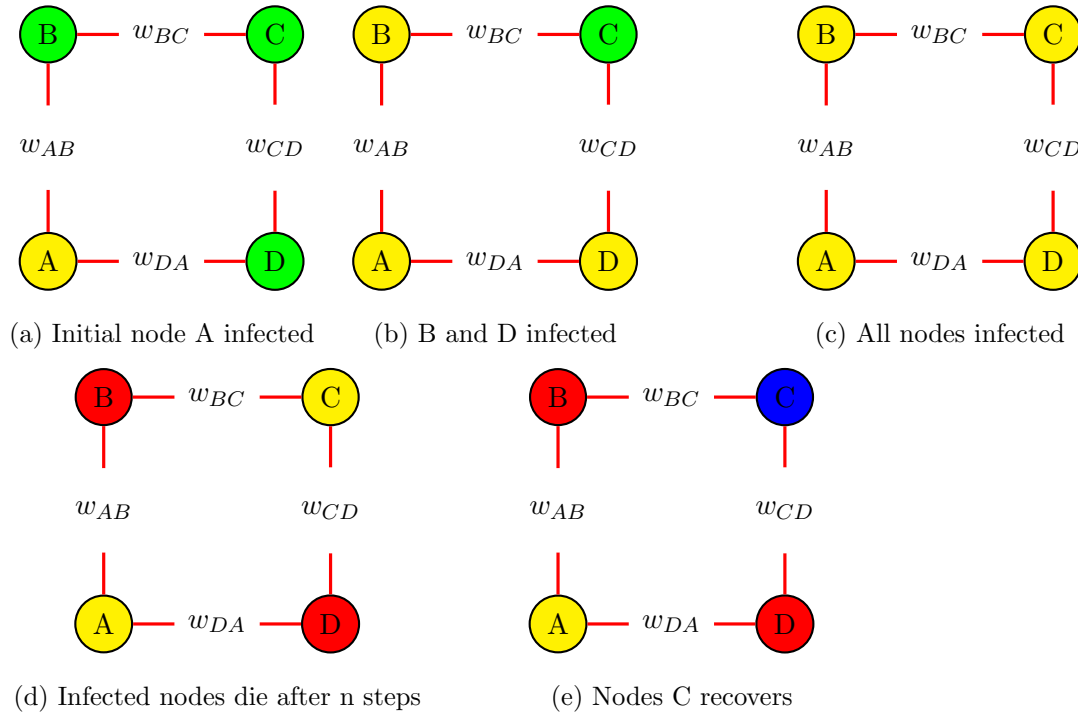

(a) Initial node A infected   (b) B and D infected   (c) All nodes infected

(d) Infected nodes die after n steps   (e) Nodes C recovers

Figure 1: Spread of Infection Simulation Rules

4   Program your simulation in Python.

5   State one or more simulation properties which you wish to investigate and which graph properties you hypothesize they may depend on.

As the size of the graph grows, it becomes increasingly difficult to visualise and understand it, thus quantitative performance measures are need. We will use the properties of graph/node such as Diameter and average path length, Clustering, Centrality and

Degree distributions, to analyse the impact of these metrics on the behaviour of the epidemic.

Diameter and average path length : If *l(u,v)* is the length of shortest path between nodes u and v , then the diameter of a graph(G) is the largest distances between any two nodes in G.

$$\text{diameter} = max \quad l(u,v)$$

The average path length is the average distance between any two nodes in the graph(G). Maximum value of average path length is diameter, but can be much smaller than diameter. If the graph is not connected then the diameter is infinity, but for this study all unconnected nodes are removed from the graph. We hypothesise that airports graph with shorter diameter and average path length will result in the spread of the quicker.

Clustering : Measures the extent to which a nodes' friends are friends with one another. This clustering measure is represented by the overall clustering coefficient *Cl(G)*, given by

$$Cl(G) = \frac{3 * \text{number of triangles in the graph}}{\text{number of connected triples of nodes}}$$

Another measure of clustering is defined for an individual node, it is given by

$$Cli(G) = \frac{\text{number of triangles connected to node i}}{\text{number of triples centered at i}}$$

We hypothesise that graphs and nodes with high clustering coefficients will spread the disease quicker and are potential candidates that need intervention, to stop the disease from becoming an epidemic.

Centrality: Importance of node's position in a graph is captured by the Centrality and there are different measures to do it.
Degree centrality: for node v of graph (G) with n nodes is given by,

$$\frac{d_v(G)}{n-1}, \text{ where } d_v(G), \text{ is the degree of node v}$$

Closeness centrality: Tracks how close a given node is to any other node: for node u, one such measure is given by

$$\frac{n-1}{\sum_{u \neq v} l(u,v)} \text{ , where l(u, v) is the distance between u and v}$$

3

Betweenness centrality: It measures the extent to which a node lies on paths between other nodes. If $n_{s,t}^u$ be the number of geodesic paths from s to t that pass through u and let $n_{s,t}$ be the total number of geodesic paths from s to t. Then the betweenness centrality of vertex u is:

$$b_u = \sum_{s,t} \frac{n_{s,t}^u}{n_{s,t}}$$

Nodes with high centrality tend to have considerable influence within a graph by virtue of their control over information passing between others. These nodes, when removed from the graph will most likely disrupt the spread of disease between other nodes as they lie on the largest number of routes.

Degree distributions: The degree of a node is the number of neighbours of the node. For any integer $k \geq 0$, the quantity p_k is the fraction of nodes having degree k. This is also the probability that a randomly chosen node in the graph has degree k. The quantities p_k, for $k \geq 0$, represent the degree distribution of the graph.

6 Carry out experiments to measure these properties in different types and sizes of graph (at least one real-world graph, and at least one scalable graph model such as the Erdos-Renyi random graph model, at multiple sizes). Present a table of data showing your results.

Questions:

How quickly should a given country stop letting flights in to avoid the disease reaching it? when is too late? Different results for iceland and uk for example?... Centrality of country: how quick does it get disease

simulation property: screening for disease - catch given percentage - effect on spread

simulation property - chance of receiving - centrality probability of country with disease of spreading to other countries weighted on population of that country probability increases at each timestep to account for fact that disease spreading within the country

Graph property tests - Tables of results

Results Diagrams/figures from Python code

Scalability test results along with tables and figures

Diameter: airport graph - take subgraph (random vs high degree nodes vs low degree nodes) and compare rate of infection and effect on stats for 3 simulations. random graph - check for different pn

Shortest Path - pick

7 State your conclusions, based on your data.

Discuss the results
    Statistics

References

[1] JONES, K. E., PATEL, N. G., LEVY, M. A., STOREYGARD, A., BALK, D., GIT-TLEMAN, J. L., AND DASZAK, P. Global trends in emerging infectious diseases. *Nature 451*, 7181 (2008), 990–993.

[2] LAWYER, G. Measuring the potential of individual airports for pandemic spread over the world airline network. *BMC infectious diseases 16*, 1 (2016), 70.

[3] TATEM, A. J., ROGERS, D. J., AND HAY, S. Global transport networks and infectious disease spread. *Advances in parasitology 62* (2006), 293–343.

[4] YAGER, N. A., AND TAYLOR, M. Edge-based control of disease propagation through the world-wide airport network. https://github.com/nicholasyager/airport-disease-modeling, 2014.