CS 556 – Spring 2018 Project 2 Modeling Worm Propagation in Large Networks

Project Due Dates:

Due before class on CANVAS by Tuesday, March 20, 2018

General Description

In this assignment, you will implement two different programs to <u>simulate</u> how a software worm propagates in a computer network (or Online Social Network - OSN) and infects machines. You will experiment with different types of large networks to understand if the topological structure of the network plays any role on how fast the worm propagates. You will also experiment with various curing rates and strategies. A report of the results of your experiments together with your analysis of the same is required for this project.

For this assignment, you will abstract a network (computer or OSN) as an undirected graph $G = \langle V, E \rangle$. The set of nodes, V, in the graph represents the different computers (for computer networks) or users (for OSNs). The set of edges, E, represents the connection between two machines (for computer networks) or the friendship / trust relationship between two users (for OSNs). An edge $e_i = (v_m, v_n)$, where $e_i \in E$ and $v_m, v_n \in V$ exists in the graph if v_m and v_n are directly connected (for computer networks) or are friends (for OSNs). For the rest of this discussion we will not make a distinction between computer networks and OSNs. We will assume that our graphs do not have parallel edges or self loops.

A network is characterized by its topology. For this project, we will work with three different types of networks: binomial networks (often called Erdös-Rényi network), scale-free networks (often called Barabási-Albert networks), and small-world networks (often called Watts-Strogatz networks). These three networks are specific instances of what are called random networks – some probability density function defines the probability that two nodes in the network are connected to each other by an edge. Do not assume that these are the only types of networks existing in the real world; however, they are representative. Erdös-Rényi networks are the classical networks that aptly describe many real-world phenomenon (such as propagation of virus in a biological population). Barabási-Albert networks accurately describe the topology of the world-wide web. Watts-Strogatz networks model many online social networks. As a first step, perform some research yourself

to develop a fair understanding of these networks. WolframAlpha and Wikipedia are good starting points for your research.

A worm spreads autonomously with some probability p from an infected node to a non-infected one that is connected to it. In the process, the worm infects the latter node. A worm will not spread to a node that is already infected. A worm continues to spread until either (i) it has infected all nodes in the network and has nowhere to spread to, or (ii) it is removed completely by applying a cure to nodes that removes the worm from infected nodes and prevents their re-infection or prevents an infection of a node to begin with.

Program 1

The purpose of this program is to study the propagation of worm on the three different types of networks through simulation when no cure (that is, worm defense) is applied. In other words, the worm will continue to spread until no uninfected node remains. While you can use any programming language of your choice but not mathematical packages such as Matlab, Mathematica, Sage or R, we suggest that you use Python with the graph manipulation library Networkx (http://networkx.github.io) or other similar libraries.

1. The first input to the program will be a CSV file for the graph. Each line in the file will be a comma-separated pair of nodes representing an edge. Use a separate program to generate the CSV file for the different types of networks.
2. The second input will be a probability value p that gives the probability of the worm infecting an un-infected node.
3. The third input will be the identity of the initial infected node from where the worm starts infecting.

The simulation works as follows. In every round, the program identifies an un-infected node that is adjacent to one of the infected nodes in the previous round, creates a list of nodes that are adjacent to this node and picks nodes to infect from this list with probability p. If the node selected is already infected, it is not infected. When all nodes in the adjacency list are checked (whether the nodes get infected or not) it completes one round. One such round is considered 1 time unit for the simulation. You need to compute the time it takes to infect all nodes. Keep track of the number of nodes that get infected in each round.

For each type of network, (Erdös-Rényi, Barabási-Albert, and Watts-Strogatz) experiment with different sizes of the network, such as 500 nodes / 1000 edges, 1000 nodes / 3000 edges and so on. We are not imposing any size requirements; it is for you to choose. However, you cannot choose network sizes that are too small (< 100 nodes and edges). Also, you may not want to use very large networks (such as > 100,000 nodes and edges) since graph algorithms can take time to execute. Choose wisely.

For each network type and size of network, run your program several times to compute the average time it takes for the worm to infect the specific network.

Program 2

The second program builds upon the first one by incorporating worm defense in the simulation. As the worm propagates through the network per program 1, a separate defense mechanism propagates simultaneously in the network either inoculating yet un-infected nodes or curing infected nodes. Inoculating an uninfected node means that the node cannot be infected by the malicious worm any further. Curing an infected node means that the worm infection is removed and the node is inoculated.

The inputs to the second program will be the same set as program 1 with two additional inputs:

1. The probability p' of a node getting inoculated or cured.
2. The identity of an initial node from which the curing process starts.

Deliverables

For each program, you need to provide a report of your experiments in terms of plots and analyses of the plots. The report will be evaluated based on the thoroughness with which the experiments are run and the maturity level of the analyses of the results. Run appropriate experiments to answer the following questions.

a) In general, the rate of spread of a worm in a network shows an S-shaped curve when there is no cure involved. Do your experiments validate this? Does the rate depend on the type of network chosen – Erdös-Rényi, Barabási-Albert, or Watts-Strogatz.
b) How does the application of a cure affect the rate of worm spread?
c) How does the type of the network affect, if at all, the average time it takes for the worm to spread without a cure being applied? With a cure?
d) If the result of (c) shows that the type of network does indeed have an effect on the spread of the worm can you suggest some strategy (other than applying the cure at a rate much faster than the spread of the worm) that will make the network more resilient against the spread of the worm. (Hint: This might involve identifying specific nodes on which to apply the cure than on other nodes. In which case, identify properties of these nodes.)