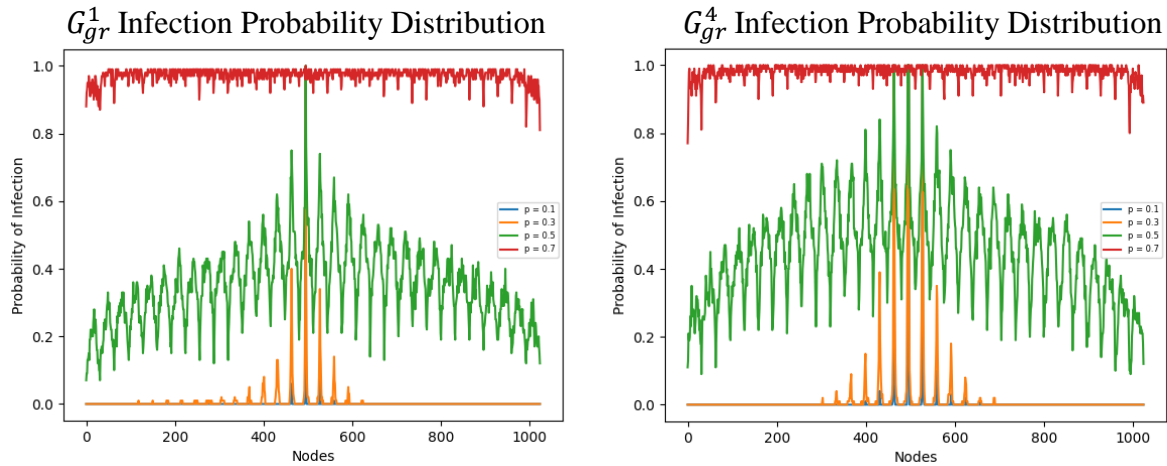


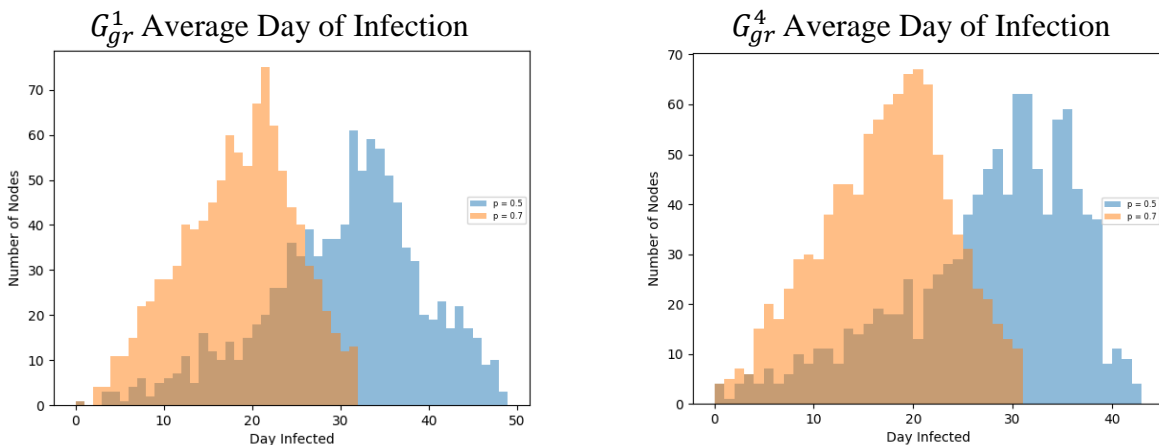
CS330 Programming Assignment 1 – Outbreak Analysis

Adam Clay – U31446255

I computed the probability of infection and the average day of infection (of the days the node was infected at all) for every node in each of the six graphs provided by iterating through the nodes of the graph using BFS, and storing the day of infection (degree of separation from the source) and a bool (true if the nodes was infected, false if not) associated with each node. This is repeated 100 times for each graph, and the probability of infection is computed by dividing the number of times the node was infected by 100, and the average day of infection is computed by dividing the accumulation of days of infection by the number of days the node was infected.

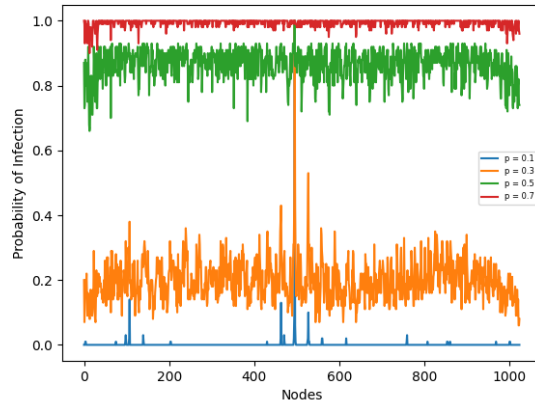


The graphs G_{gr}^1 and G_{gr}^4 are identical, other than the fact that grid4 has 4 source nodes, while G_{gr}^1 only has one. Over the course of 100 trials, for low probabilities of infection ($p = 0.1, 0.3$), both graphs infected 20-30 nodes at least once, and for high probabilities of infection ($p = 0.5, 0.7$), both graphs infected every node at least once. Since grid graphs only have connections to the nodes immediately adjacent to them, low probabilities of infection cause the disease to quickly peter out and not spread very far from the source node(s): in both graphs, nodes near the source(s) have relatively high probability of infection, but nodes far from the source(s) were never infected. On the other hand, in cases of high probability of infection, the disease was able to propagate through the population easily due to the many paths between two nodes. The multiple sources in G_{gr}^4 made the probabilities of nearby nodes increase, but largely did not have an effect on the nodes far from the source(s).

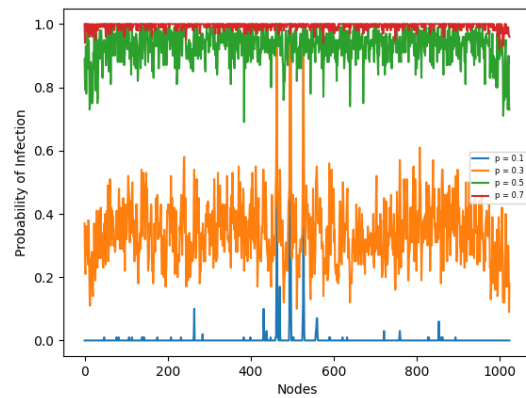


Shown above is the average day of infection for G_{gr}^1 and G_{gr}^4 for $p = 0.5$ and 0.7 . I did not include the distributions for $p = 0.1$ and 0.3 because they largely followed the same trends as shown above. With 4 source nodes instead of 1, the infection is able to spread more quickly with more nodes being infected in earlier days. The peak of the infection is also earlier in G_{gr}^4 , and more nodes are infected in G_{gr}^4 on those peak days, and the virus is able to propagate through all nodes more quickly. These results would become more drastic as more source nodes are added to the graph.

G_{gs}^1 Infection Probability Distribution

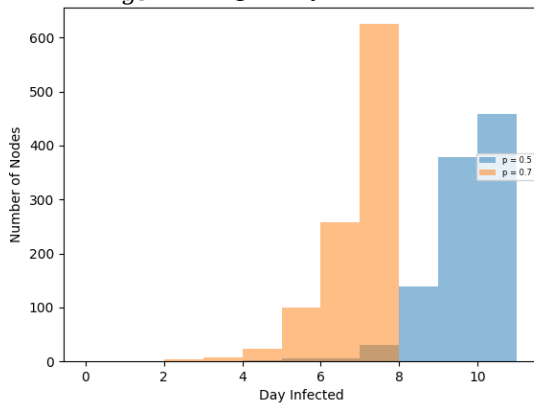


G_{gs}^4 Infection Probability Distribution

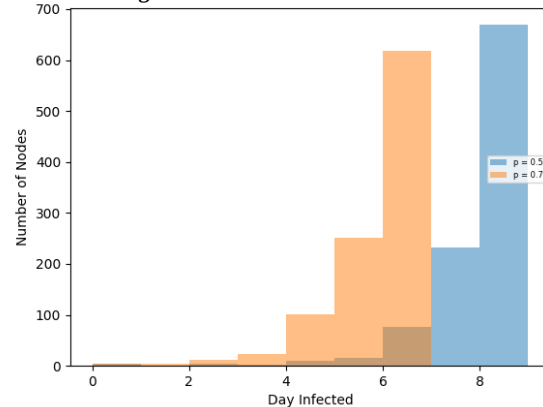


These 2 graphs are identical to G_{gr}^1 and G_{gr}^4 except each node connects to another node chosen uniformly at random. For $p = 0.1$, G_{gs}^1 was able to infect only 20 nodes, G_{gs}^4 infected 90, and for all other probabilities both graphs infected all nodes at least once. In these graphs, due to the random shortcuts, all nodes have nearly the same probability of being infected because, instead of the vanilla grid that may have many edges between two nodes, there is a possibility in these shortcut grids that two distant nodes have just one edge between them. In the vanilla grid it is clear that the nodes closest (fewer edges between) to the source node(s) have a much higher infection probability than nodes farther away. This is not the case with shortcuts, as nodes are not limited to infecting only nodes adjacent to them in the grid.

G_{gs}^1 Average Day of Infection

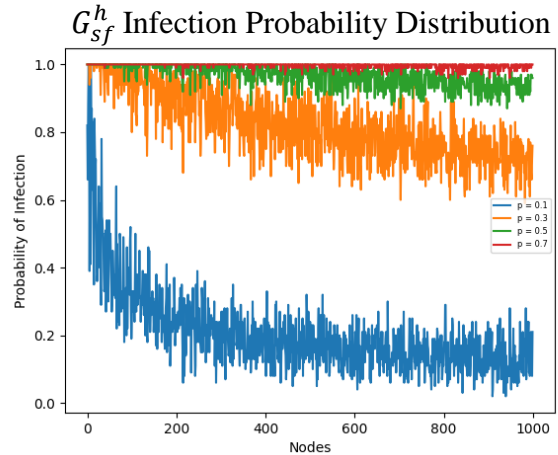
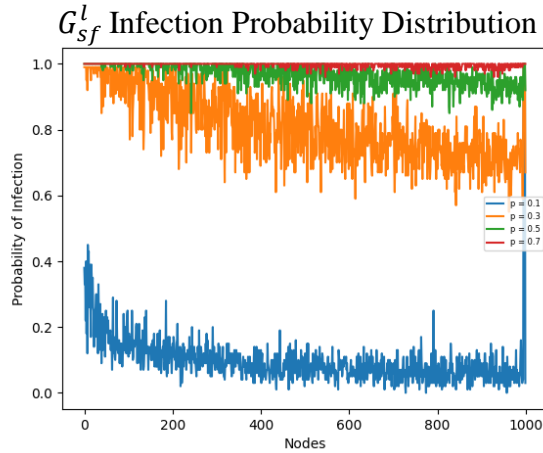


G_{gs}^4 Average Day of Infection

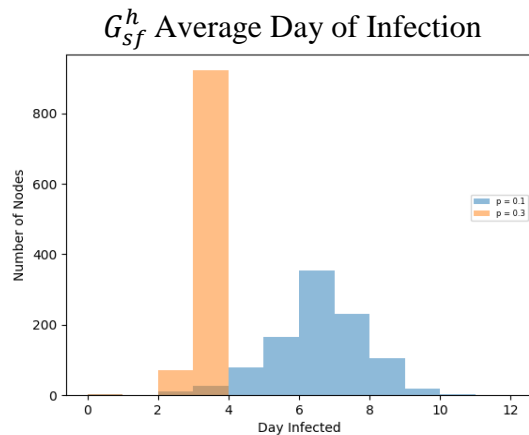
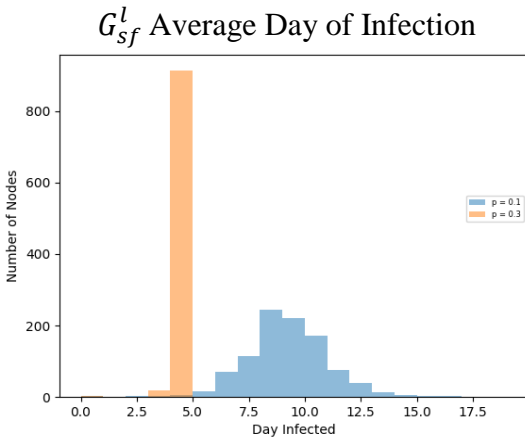


As was the case in the grids without shortcuts, G_{gs}^4 exhibits more infections earlier, more infections on each day, and infects all nodes more quickly than G_{gs}^1 . As compared to the grids without shortcuts, the grids with shortcuts infect far more nodes each day (hundreds of nodes are infected per day in the graph with shortcuts, while only dozens are without), and infect all of the nodes in the graph far more quickly (8-10 days in the grids with shortcuts, and 40-50 days without). Clearly, the infection probability of nodes G_{gs}^1 and G_{gs}^4 , and the number of nodes infected per day, is far greater than nodes in G_{gr}^1 and G_{gr}^4 .

G_{gr}^4 had an infection probability distribution with the highest probabilities focused around the source node, and the probability of infection decreased for nodes further from the sources, this is not the case for G_{gs}^4 . While nodes in G_{gr}^4 had the probability of infecting another node n edges away of $n * p$ (where $p = 0.1, 0.3, 0.5, 0.7$), G_{gs}^4 had a uniformly random chance of infecting that node with probability p . So, all nodes in G_{gs}^4 have a much greater chance of being infected, as is shown in the probability distributions for G_{gs}^4 ; all nodes (other than source nodes) have nearly an equal chance of being infected in each iteration. The same comparison is true for G_{gr}^1 and G_{gs}^1 . The fact that paths between nodes may be far shorter in G_{gs} than in G_{gr} drastically increases the probability of each node being infected, especially those far from the source node(s), and because each node has one extra neighbor in G_{gs} than in G_{gr} the infection spreads far more quickly for all four infection rates. When $p < 0.5$, in both G_{gr} and G_{gs} , the probabilities of infection are far lower than for values of $p \geq 0.5$. This is due to the edge-layout of both grids: since each node will have 2, 3, 4, or 5 (for G_{gs}) neighbors, the probability of any node infecting another with infection rate ≥ 0.5 is far greater than the probability of the node not infecting others (e.g. the probability of a node with 4 neighbors infecting at least one other node with $p = 0.7$ is $1 - (0.3)^4 = .99$), while the probability of a node infecting any other is much less likely for rates < 0.5 (e.g. a node with 4 neighbors for $p = 0.1$ is $1 - (0.9)^4 = .34$). After this next node is infected the same probability applies to its neighbors, showing that infection rates ≥ 0.5 propagate the infection further, and more quickly through G_{gr} and G_{gs} .



G_{sf}^l had four source nodes of low degrees, and for G_{sf}^h had 2 source nodes of high degrees. Over the 100 iterations, G_{sf}^h infected each node at least once for all values of p , and G_{sf}^l infected all but 7 for $p = 0.1$, and infected all nodes at least once for all other values of p . Other than the source nodes, the two graphs are identical, and other than the probabilities of 1.0 for the different sources in the two graphs, the probability distributions are largely the same. For infection rate of 0.1, the probability for infection for nearly all nodes is greater, but for all other values of p they are largely the same.



Due to the very high connectivity of these graphs, for $p = 0.5, 0.7$ all nodes were infected within 2-3 days, so the contrast in the average day of infection for G_{sf}^l and G_{sf}^h is better displayed for $p = 0.1$ and 0.3 , as shown above. Nodes are infected earlier in G_{sf}^h than in G_{sf}^l because the high degrees of connectivity for the source nodes in G_{sf}^h infect many more nodes right away. The amount of infections per day is comparable between the two, however the peak of infections for G_{sf}^l when $p = 0.1$ are days 8 and 9, while in G_{sf}^h for $p = 0.1$ very few nodes are being infected at that time: G_{sf}^l infects more nodes on later days than G_{sf}^h .

In both experiments, for all values of p , there is a clear downward trend in probability of infection from nodes closer to node 0 to nodes closer to node 1000. Thus, the earlier nodes are infected far more often than the later nodes. The reason for this is clear in the adjacency lists of nodes in G_{sf} : nodes 0-2, and 8-13 all have dozens of neighbors, while nodes 990-1000 all only have four neighbors. Since the earlier nodes have such high degrees of connectivity, the chances that one of their neighbors has been infected is very high, and thus the probability of them getting infected themselves is greater than the value of p , as shown in the probability of infection graphs. The more neighbors a particular node has, the greater the probability of infection for that node.