

Comparing Vaccination Strategies Across Facebook Friend Networks

Florian Frick, Kevin Veno

Introduction

During the global pandemic, one of the most crucial factors for thwarting the spread of COVID-19 was having an effective distribution of the vaccine when there was not enough for everyone. For our project, we set out to investigate the differences in vaccination distribution strategies, based on which types of people vaccinating would help to mitigate the spread of a given disease the most. A popular model for epidemiology, and one that models the transmission of COVID-19 quite well, is the Susceptible-Infectious-Recovered or SIR model. We implemented an SIR model using the NetworkX library in Python to simulate the spread of a disease in a population. To achieve this, we focused on three different vaccination strategies: random vaccination (control group), vaccination based on the highest k-means nodes, and vaccinations on notable bridge nodes between clusters. For this model, the goal was to see which strategies did the best at mitigating the overall total number of

infected individuals and the maximum number of people infected at one time in the simulation. The maximum number of infected people at a given time in the simulation helps us to also understand different aspects of this pandemic simulation, as a major issue with COVID-19 was hospital overcrowding. This will give us an idea of how many people would need to take up hospital resources.

Methods

For our model simulation, We utilized the Facebook100 dataset, specifically the 'America75' subset, which comprised 6,386 nodes, 217,662 edges, a mean degree (kmean) of 68.168, a clustering coefficient (C) of 0.164, and an average shortest path length (ellmean) of 2.77. In our SIR model, we set both the Beta and Gamma values to 0.3 for this simulation.

To determine the number of clusters needed for our given dataset, we employed the Louvain method for extracting the number of communities for our network, using a built-in Python package from the

community library. We utilized the Girvan-Newman algorithm, a built-in NetworkX function, to partition the network and assign each node to its appropriate community cluster. We realized, too late, that this method of community detection was not well suited for our purposes because our networks are large (200k edges) and the Girvan-Newman algorithm has runtime proportional to the number of edges squared.

Before running the main simulation, we compiled two lists: a degree list sorted from lowest to highest degree nodes, and a list with the least to most notable bridge nodes in the network. We developed a custom algorithm to identify prominent bridge nodes, which calculated a value for each node based on the number of clusters it interacts with (excluding its own), multiplied by the number of neighboring nodes in a different cluster, divided by the total number of neighboring nodes within its own cluster. This approach yielded the best results for identifying prominent bridge nodes in our simulation, in comparison to some of the built in networkx functions as this method takes into account if a node

makes multiple bridge connections to more than just one cluster.

Each strategy type in the simulation was run 100 times, and the results were displayed in a histogram to ensure that our findings were not due to chance and could be replicated. Before each trial in the simulation, a fixed number of individuals were vaccinated based on the chosen vaccination strategy. Once a node was vaccinated, it could no longer receive or transmit the virus for the remainder of the simulation.

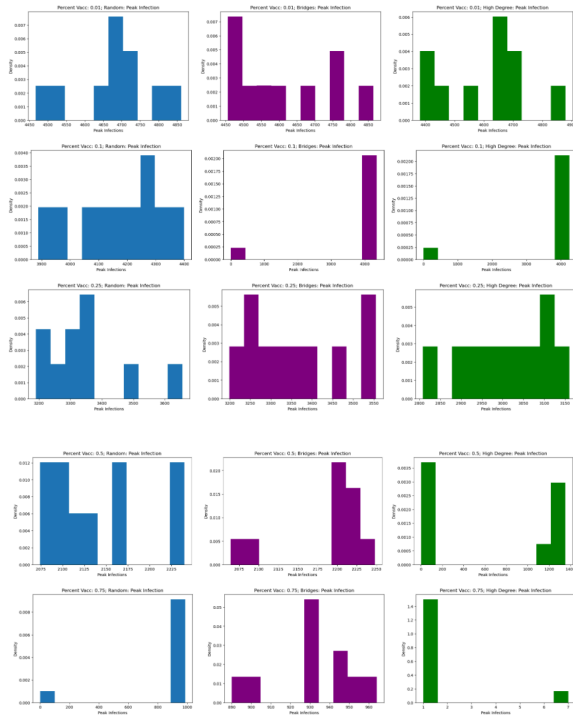
We used a fixed number of vaccines for all three vaccination strategies, running each method with the fixed number of vaccinations over multiple iterations and averaging the results. For our control method, we randomly vaccinated nodes in the network. For the other strategies, we selected nodes with the highest degree and most prominent bridge nodes. In each trial, we recorded the total number of people infected and the highest number of people infected at any given time in the simulation.

After collecting this data for each of the 100 trials per vaccination strategy, we converted the data into histograms to

normalize it and illustrate the worst-case scenario within the simulation.

Results

The following graphs show the histogram of peak infections across 15 trials with varying percentages of initial vaccination. The blue graphs show the strategy of random vaccination, the purple of vaccinating bridges, and the green of vaccinating high degree nodes. Initial vaccination increases from 1% of the population at the top to 75% at the bottom.

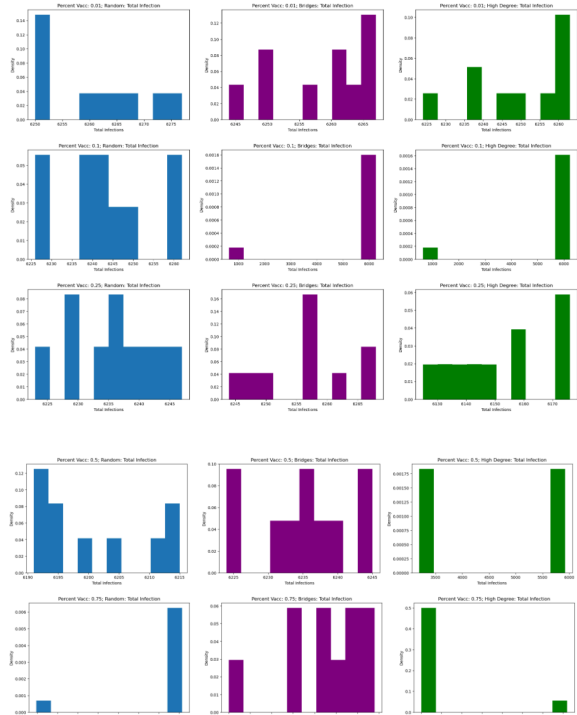


It is clear that the strategy of vaccinating high degree nodes becomes best at lowering the peak infection as the percent of initial vaccination increases. This is

reflected in the average peak infection across the trials as shown in the table below.

Average Peak Infections:	Average Total Infections:
Percent Vaccinated = 0.01	Percent Vaccinated = 0.01
Random 4683	Random 6260
Bridge 4622	Bridge 6258
High deg. 4609	High deg. 6249
Percent Vaccinated = 0.1	Percent Vaccinated = 0.1
Random 4168	Random 6243
Bridge 3727	Bridge 5694
High deg. 3657	High deg. 5660
Percent Vaccinated = 0.25	Percent Vaccinated = 0.25
Random 3348	Random 6236
Bridge 3360	Bridge 6256
High deg. 3006	High deg. 6152
Percent Vaccinated = 0.5	Percent Vaccinated = 0.5
Random 2140	Random 6200
Bridge 2187	Bridge 6235
High deg. 641	High deg. 4548
Percent Vaccinated = 0.75	Percent Vaccinated = 0.75
Random 849	Random 6076
Bridge 933	Bridge 6242
High deg. 1	High deg. 4791

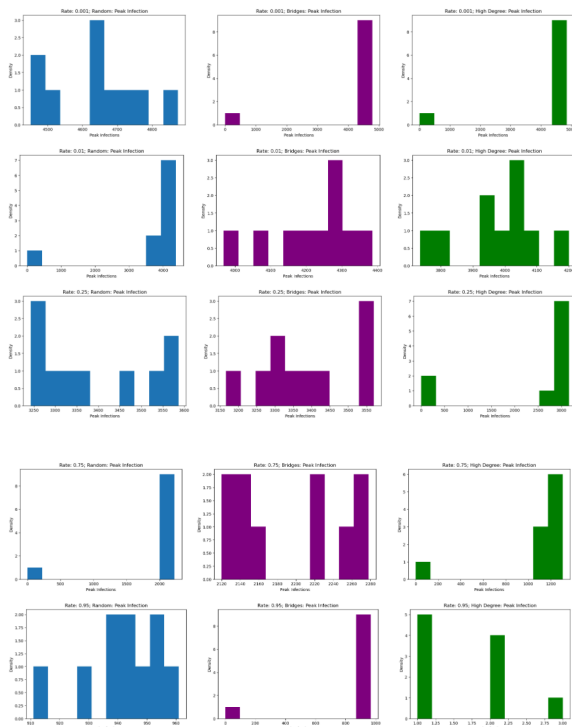
Similarly, the strategy with the lowest average peak total infections was the high degree node strategy. This is also reflected in the following histograms



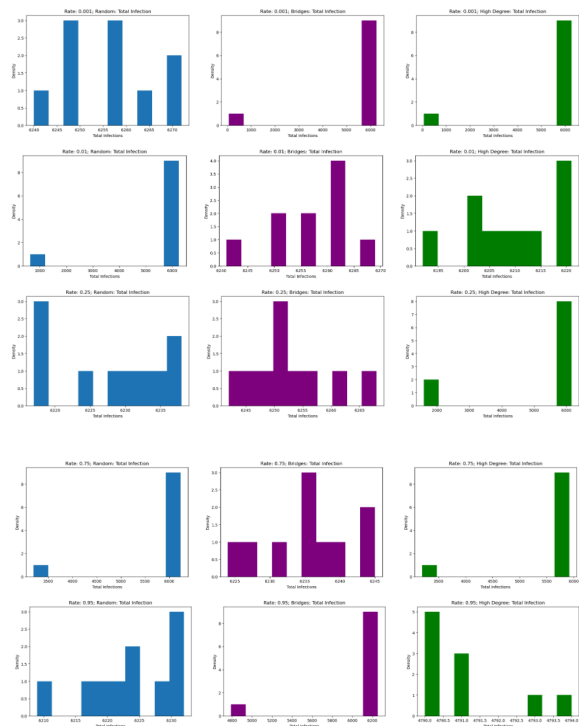
Next, the results of running an SIRV model with rates of vaccination at each timestep, rather than an initial vaccination are as follows. With 10 trials of rates from 0.1% to 95% vaccination at each timestep, we see that the average peak and total both favor the high degree node strategy at all rates.

Average Peak Infections:	Average Total Infections:
Vaccination Rate = 0.001	Vaccination Rate = 0.001
Random 4635	Random 6256
Bridge 4194	Bridge 5629
High deg. 4182	High deg. 5630
Vaccination Rate = 0.01	Vaccination Rate = 0.01
Random 3671	Random 5680
Bridge 4217	Bridge 6257
High deg. 3979	High deg. 6209
Vaccination Rate = 0.25	Vaccination Rate = 0.25
Random 3394	Random 6228
Bridge 3388	Bridge 6252
High deg. 2410	High deg. 5238
Vaccination Rate = 0.75	Vaccination Rate = 0.75
Random 1929	Random 5920
Bridge 2196	Bridge 6235
High deg. 1105	High deg. 5627
Vaccination Rate = 0.95	Vaccination Rate = 0.95
Random 941	Random 6223
Bridge 844	Bridge 6095
High deg. 1	High deg. 4791

The corresponding histograms for the peak infections are:



And the histograms for total infections are:



Again, these show the high degree strategy out performing bridges and random. As expected, bridges are better than random in our results, however, which justifies it as a somewhat valid strategy nonetheless, especially with tweaking in the future.

Discussion

Based on the result, it appears that given a higher rate of vaccination of at least .1-.25 vaccination ratio the high degree nodes starts to outperform both the random vaccination, and vaccinating prominent bridge nodes, this gap is only expanded with higher vaccination percentages. Overall

at higher vaccination rates the vaccinating high degree nodes still performed best when taking into account the total number of people infected in the model but where vaccinating high degree nodes performs even better with this model is when taking into account the total number of people infected at a given time, as this is most important to the simulations overall goal which is to find a strategy that would mitigate the total infections at any given time as this was a big issue during covid as there was not enough hospital beds to go around for everyone as there were such large spikes cases throughout the pandemic.

In the future, evaluating the efficacy of more strategies, with more metrics would lead to better ways of leveraging network structure in dealing with a pandemic. Further, using a model of epidemic spread that is more realistic, such as the Watt's or

hierarchical metapopulation model would give more conclusive results about the best strategy. Similarly, including metadata such as age or immune strength to the model and combining structural and nonstructural information to a more holistic strategy will give a more applicable strategy to implement into the real world.

Bibliography

Kate M. Bubar et al. ,Model-informed COVID-19 vaccine prioritization strategies by age and serostatus.Science 371, 916-921(2021).

D.H. Morris et al., Optimal, near-optimal, and robust epidemic control. Communications Physics 4, 78 (2021).

Traud, Amanda L., et al., "Social Structure of Facebook Networks." SSRN Electronic Journal, 2011. Crossref, <https://doi.org/10.2139/ssrn.1470768>.