Optimizing the Administration of COVID-19 Vaccines

Rohan Bhushan
rsbhusha@ucsd.edu

Kunal Jain
kujain@ucsd.edu

Shivam Lakhotia
slakhotia@ucsd.edu

Sothyrak Tee Srey
ssrey@ucsd.edu

Abstract

The COVID-19 pandemic has resulted in millions of deaths around the world and various effective vaccines are being developed and manufactured throughout the world. But given that the production of vaccines is still catching up with the demand, it is imperative to optimize the distribution to the general public to effectively check the spread of COVID. In this work, we aim to use convex optimization to realize the maximum impact of administering the vaccines. We formulate this as an optimization problem where we find a subset of people who if given the vaccine will minimize the transmission of the virus. Since not everyone is at equal fatality risk from COVID, we also factor in the vulnerability of various groups of people. We constructed multiple distinct simulations to test our formulation, and experiment with different social situations to identify which individuals to administer vaccines, given a limited amount. Our findings show that in situations where the vulnerability of all individuals is equal, the vaccine should be prioritized to that group of people who interact with maximum number of people such as grocery store employees, teachers etc. When we factor in the the vulnerability of people, then most vulnerable should be given the vaccine first.

1 Introduction

COVID-19 has infected over 121 million people and claimed the lives of over 2.67 million people in the past year across the globe (who). Fortunately, clinical research has finally developed effective vaccines. Back in November of 2019, the preliminary efficacy results for their COVID-19 vaccine reported by Pfizer/BioNTechand Moderna was at least 90%. Oxford and AstraZeneca group also confirmed the safety of their vaccines across variety of groups (Ramasamy et al., 2020).

Currently, these vaccines are administered in a phase based strategy in that Phase 1A includes the front line workers, Phase 1B includes people over 65 years of age. Phase 1C includes people who are have high-risk medical conditions so and so forth (pha). In this work we want to verify that this strategy is indeed a good strategy not only in terms of saving individual lives which, of course, is paramount but also is able to curb the community transmission of the virus which is also vital.

There has been some evidence that fully vaccinated people are less likely to spread COVID, further investigation is still under progress (cdc). In this work, we are working under the assumption that fully vaccinated people are far less likely to spread the COVID. We want to find a subset of people who if given the vaccine will minimize the spread of COVID. In terms of strategy, this is more of a community level strategy than an individual-level strategy. Even though our results agree with the current strategy to fully vaccinate people who have higher exposure and to people over the age of 65 and with serious medical conditions, our method is able to provide a more granular subset of people who should be vaccinated first instead of blanket groups which we believe will result in better prevention of the transmission of COVID.

We plan to use an exposure graph where every person is a node and the edges between two nodes represent if they are exposed to each other or not. In May'20, Apple and Google partnered to build an exposure notification system which was able to track the nearby devices and notify people if they have been exposed to COVID (Team, 2020). By simple modifications to this system, we can build the exposure graph and subsequently find out the subset of nodes/people who should be administered the vaccine in order to minimize the spread of COVID. Since this information is not available to general public, we are currently working with simulated graphs which shows the efficacy of our method.

In the next section, we look at some previous works that are relevant to figuring out how to effectively distributing COVID-19 vaccines and various strategies for selecting significant nodes in a network. In section 3, we formulate the primal and dual problems, Karush–Kuhn–Tucker conditions. In section 4, we further discuss the two methods used in this paper based on the formulations in section 3. In section 5, we introduce the graphs and why they are relevant for our experiments. In section 6, we show the results of our two methods on the graphs introduced in section 5. In the final section, we have conclusions and the future directions of work.

2 Related Work

Since the beginning of the COVID-19 pandemic, various works were done to to keep the spread under control and prevent any potential outbreak in the future. (Guo et al., 2020) showed a effective method to control epidemics spread by analyzing a complex social network and using information entropy to identify influential nodes which plays a crucial part in target advertisement, epidemics and rumor control.

(Santini, 2020) addressed the lack of resource and propose a strategy on how to overcome this by performing real time simulation of a social network consisting of older and younger people. In their work, they experiment with two approaches, greedily providing vaccines to the nodes with the most neighbors (highest degree) and administering vaccines to young people with the most connections to the elders, who are more vulnerable. The paper suggests that the most effective strategy is to give priority to those who are more heavily connected to the vulnerable people(caregivers, cleaning personnel, etc). But when taking distribution of vaccines over time into account, the optimal strategy switches to prioritizing a fraction of elderly and then giving it to young people as the number of victims increases.

Similarly, (Cheng et al., 2021) also performed a similar experiment on similar model and concluded that high-risk individuals should be given priority. However, there are still many assumptions and factors to consider in their paper as well as (Santini, 2020). Most importantly, a crucial component that is missing in these two paper is the mechanism to decide who exactly out of those people in each group should be selected to receive the vaccine assuming that we know which group should be prioritized and how many out of each group are eligible.

Identifying influential nodes in a network is not a novel research topic. (Zhang and Zhang, 2020) also published a paper that proposed a modified greedy algorithm as a way to select the top-k most influential nodes in a given graph (Zhang and Zhang, 2020). (Elkin et al., 2013) on the other hand, used convex relaxation to identify those nodes. Even though their work focuses solely on social network, the use of convex optimization in their work provide great insights for figuring out an optimized way of allocating the limited COVID-19 vaccines for a given community. As a result, this paper uses the optimization problem used it their paper as a reference point. With convex optimization, we could easily add or remove linear constraints or modify the objective function to solve other problems related to resource allocation to maximize effect to the entire networks given the limitation.

Before selecting a set of people from a given community, it may be beneficial to be able to identify the sub-communities in a bigger network of people. The work of (Voss et al., 2016) on spectral clustering may also be incorporated to our current work to further improve the quality of vaccine distribution. In their paper, they proposed an algorithm to cluster nodes in a given graph. This can be useful in a bigger graph since this algorithm attempts to divide the entire graph into smaller community which allows us to allocate our limited resource more efficiently.

Given previous works on strategies for COVID-19 vaccines distribution as well as methodologies for selecting influential nodes in a graph, our main contribution of this paper to these ongoing researches is to figure out exactly who among the people in the community should be vaccinated. This is done by introducing some new key ideas such as taking into account the number of hops between nodes rather than the reachability between two nodes, considering the vulnerability of each individual, clustering nodes into sub-communities before nodes selection, and experimenting on more realistic models that resemble real life scenarios (grocery stores, nursing homes, schools, and the combination of all three).

3 Formulation

We have an undirected graph G consisting of vertices V and edges E where each node is corresponding to a person and there is an edge between two nodes if they come in frequent contact. We are also given a diagonal vulnerability matrix, C, where C_{ii} denotes how likely is person i likely to contract COVID. We also know the total available doses of the vaccines denoted by k. Under this setting, we wish to find a subset V^* of V such that $|V^*| = k$ such that this minimizes the distance between the subset of nodes given the vaccine and the remaining nodes.

We create $R \in \mathbb{R}^{|V| \times |V|}$, where r_{ij} denotes the minimum number of hops required to reach node j from node i. If node j is not reachable from node i then R_{ij} is ∞.

3.1 Primal

Let $x \in \mathbb{R}^{|V|}$ be the solution vector such that if $v_i \in V^*$ then x_i = 1 else x_i = 0. Let $t \in \mathbb{R}^{|V|}$ such that t_i = 1 if v_i is reachable from at least one node in V^* and t_i = 0 otherwise. Let $e \in \mathbb{R}^{|V|}$ where e_i = 1 for all i.

Now, for discrete values of x, the problem is NP-hard (Elkin et al., 2013). Therefore, following (Elkin et al., 2013) we relax the problem to continuous values of x.

The primal formulation of the stated optimization thus becomes:

$$\min_{x,t} e^T R^T C x$$

such that

$$t \leq R^T x$$

$$t \leq e$$

$$e^T x = k$$

$$0 \leq x \leq e$$

Here we are minimizing the sum of the total cost which is computed using $R^T C x$ for all nodes. Currently, we are not using t because we are assuming that the entire graph is reachable. In case we have disconnected components we can add t to the objective.

In the first constraint, t which should be less than $R^T x$ is the distance of all the nodes from the selected set. If a node is not reachable from the selected set then $t = 0$. Condition $e^T x = k$ ensures that we have utmost k nodes in the selected set. And the final constraint is the relaxation on the values of x.

3.2 Dual

We define the Lagrangian associated with the primal form as:

$$\begin{aligned} L(x,t,\lambda,\nu) = e^T R^T C x + \lambda_1^T (t - R^T x) \\ + \lambda_2^T (t - e) + \lambda_3^T (x - e) \\ - \lambda_4^T x + \nu * (e^T x - k) \end{aligned}$$

Next, we define the Lagrange dual function as:

$$g(\lambda,\nu) = \inf_{x,t} L(x,t,\lambda,\nu)$$

When we solve the dual function we get the following dual formulation of the above problem is as follows:

$$\min_{\lambda,\nu} e^T \lambda + \nu k$$

such that

$$\lambda \geq 0$$

$$e^T R^T C + \lambda^T + \nu e^T \geq 0$$

R and C are defined in Section 3.1

3.3 KKT conditions

Let (x^*, t^*) and (λ^*, ν^*) be any of the optimal points of the primal and dual problems. Then the first-order KKT conditions that lead of zero duality gap are as follows:

$$t^* \leq R^T x*$$

$$t^* \leq e$$

$$e^T x^* = k$$

$$0 \leq x^* \leq e$$

$$\lambda^* \geq 0$$

$$\lambda^* (x^* - e) = 0$$

$$C^T R e + \lambda + \nu * e = 0$$

4 Methodology

We tried two different approaches(both using the formulation described in the previous section but differently) for finding the most important nodes (to give vaccines to) given a graph. These are as follows:

4.1 Method 1

In this method, given graph G and the number of vaccines to administer k, we directly apply primal objective as introduced in section 3.1. As stated in the previous section, we need to construct R (measure of number of hops required to reach a vertex from another) from the given graph G. If we have the additional data of vulnerabilities of people in form of C as described in the previous section, this method can take that into account while calculating the important nodes. If we don't have the vulnerability data, we can just treat C to be an identity matrix (implying everyone has same vulnerability). This primal objective is then a convex optimization problem (as all the constraints, inequalities and optimization function is convex) and can be solved via DCP solver. We used cvxpy library of python to solve it.

4.2 Method 2

As an alternative to previous method, we also tried an approach that uses the combination of the popular spectral clustering method (Shen, 2021) and the formulation we introduced in Section 3. In this method, given graph G and the number of vaccines to administer k, we first find k clusters. Clustering is performed using spectral clustering (Shen, 2021) on the adjacency matrix of the given graph. Once we have found k clusters, we then use our primal objective(Section 3.1) with $k = 1$ within each cluster to find the most influential node within that cluster. There are several issues that this approach suffers from in comparison to our approach. Since spectral clustering uses k-means as one of the steps, there is some randomization involved with initialization. Thus, there is no globally optimal clustering and each run of the algorithm may give different results. Another shortcoming is its inability to incorporate vulnerability data prior to clustering to find better clusters. With this approach, we can include vulnerability only after the clustering is done and we need to find a single most important node.

We test both of the above methods on different kinds of real life relevant graphs which are introduced in next section.

5 Experiments

During our experimentation using the primal form of the problem, we identified three distinct community models representative of specific real-world situations. In addition to these experiments, we analyzed the results of the algorithm on various random graphs simulating interactions between people. When the vulnerability is not mentioned, it is assumed that all nodes have equal vulnerability (the diagonal of C is all 1's). We show results for both the approaches(Section 3.1 & Section 3.4) on three distinct community models.

5.1 Grocery Store

The grocery store example is an important simulation, due to its vital role no matter the severity of the virus. Let n be the number of employees, and m be the total number of customers. The grocery store can be visualized as a number of "aisles", each belonging to an employee. Each of the n employees is connected to the employees in the neighboring aisles, and to a random fraction of the m customers. We also include a probability of interaction, p, which determines the probability of interaction between customers within each aisle, where 0 results in no interactions, and 1 results in a guaranteed interaction for each customer with another customer. This model can be expanded to similar situations, like indoor/outdoor dining or retail.

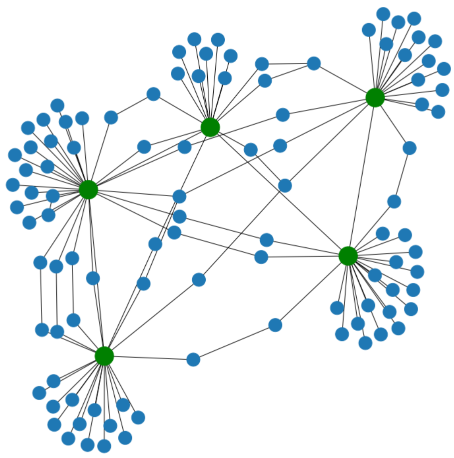

Figure 1: A grocery store simulation with 5 employees (Green) and a total of 100 customers (Blue) with a probability of interaction of 0.2

5.2 Nursing Home

The nursing home simulation looks at how a fully connected community, with a certain number of people with connections to other communities, can impact vaccine distribution. Specifically, we know that during the pandemic, nursing homes have been closed off to visitors, so this model attempts to analyze how the connections of employees affect the network. Let n be the number of employees at the nursing home, m be the number of residents at the nursing home, and g be the number of people in the outside community, randomly distributed for each employee. The graph is constructed by creating a fully connected graph with the employees and residents, and connecting each employee to a random graph with g nodes (this subgraph is not necessarily fully connected). The resulting graph is similar to another important simulation of gatherings, where the members of the gathering form a fully connected network, but also have connections to communities outside of the gathering.

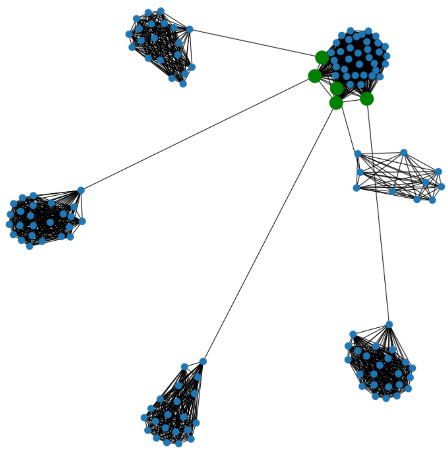

Figure 2: A nursing home simulation with 5 employees (Green), 35 residents (inner component), and 100 people randomly distributed among the outside communities (outer components)

5.3 School

As school reopenings become increasingly common around the U.S., the analysis of vaccine distribution among this community is vital. We constructed a school model to identify how vaccines can be effectively distributed among school populations. Let n be the number of teachers, m be the size of each class, and f be the size of each family. The constructed graph is a fully connected subgraph with n nodes representing the teachers, where each node is connected to a fully connected subgraph with m nodes, representing a class of students. Each student is connected to a fully connected subgraph with f nodes, representing a family.

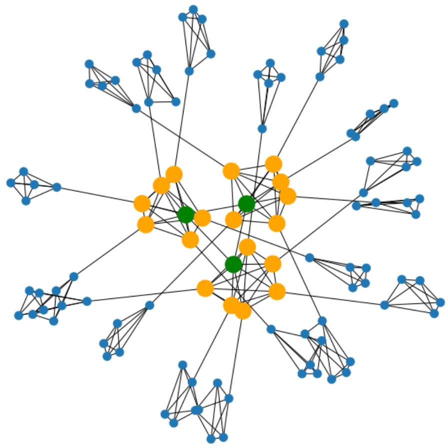

Figure 3: A school simulation with 3 teachers (Green), a class size of 6 (Orange), and a family size of 5 (Blue)

5.4 Community

Utilizing the three developed models simulating different critical aspects of society w.r.t. vaccine distribution, we built a community model. The simulation consists of grocery stores, nursing homes, and schools all interconnected as applicable. For example, nursing home employee nodes have connections to students and families within the school model, and employees at the grocery store. Additionally, the "unnamed" nodes, such as those connected to the employees at the grocery store are randomly connected with each other, creating a large network of interconnectivity.

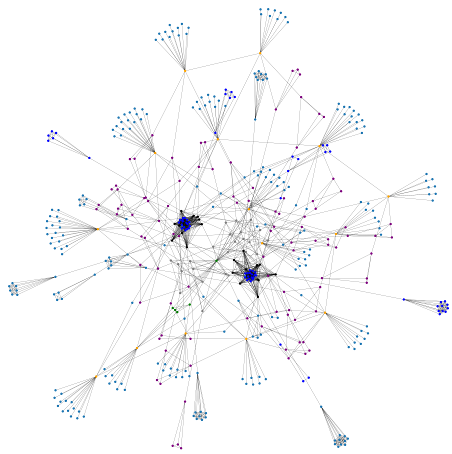

Figure 4: A community model consisting of 1 school, 3 grocery stores, and 2 nursing homes

6 Results

After developing our models, we ran the algorithm on each graph with different parameters and vulnerability distributions to identify the optimal vaccine administration.

6.1 Grocery Store

Analyzing the results of the algorithm on the grocery store graph, we observe the following:

- The employees are always given preference in the case of the number of vaccines being less than or equal to the number of employees, even in increasing probabilities of customer interaction.

- If the probability of customer interaction is 0 and the number of vaccines is more than the number of employees, then the extra vaccines should be distributed to the customers within the "middle" aisles, rather than customers within the "edge" aisles.

- When factoring in the vulnerability in customer population, we saw that the algorithm (based on primal objective) would prefer customers who were more vulnerable over employees, specifically in cases where other customers within the same aisle were also vulnerable.

- We also observe higher preference to vulnerable customers over the edge employees, and vulnerable customers closer to the center aisle would be given preference, since these nodes are closer to the other nodes.

Overall, the results show the importance of vaccinating the employees over the customers due to the large connections not only with the customers in the corresponding aisle, but as well as the connections to neighboring aisles. Also, vulnerable people that interact with more people should be given higher preference as expected.

In the case of spectral clustering, we find similar results as the original algorithm, where employees are given preference.

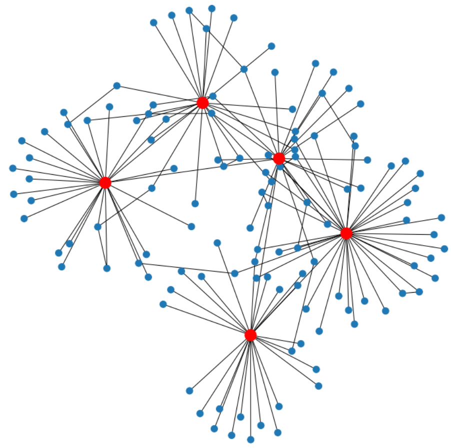

Figure 5: The solution based on method 1 to the grocery store simulation with 5 employees, 100 customers, and 5 vaccines

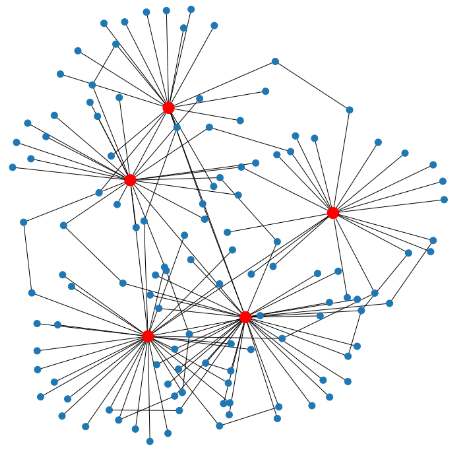

Figure 6: The solution based on method 2 approach to the grocery store simulation with 5 employees, 100 customers, and 5 vaccines

6.2 Nursing Home

For the nursing home scenario, we observe that results are similar to the grocery store case:

- When the number of vaccines is less than or equal to the number of employees, the employees of the nursing home model are always given preference.
- Interestingly, when the size of outside communities is much lower than the size of the nursing home, the employees are still given preference, as any extra contact gives them preference over the nursing home residents.
- Incorporating vulnerability, which is especially pertinent for older populations, such as

the elderly residents in nursing homes, we obtain a vaccine administration focused around them.

- If the vulnerability of the residents of the nursing home is increased, they are often given the vaccine over the employees. The cases where this does not happen are if the outside communities are incredibly large, or if the number of vulnerable residents is low.

- A point of importance is that even if the vulnerability of the outside community is increased drastically and with large frequency, the residents and employees are given preference.

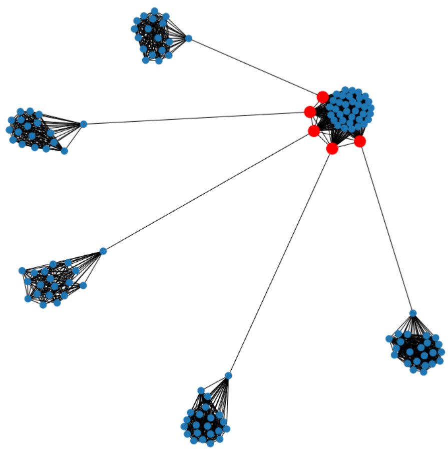

Figure 7: The solution based on method 1 to the nursing home simulation with 5 employees, 35 residents, 100 outsiders randomly distributed among the employees, and 5 vaccines

Approaching this model with spectral clustering, we see that the results are inconsistent with the results of the original algorithm. Specifically, we often see preference given to the individuals forming the connection between the employee and outside community.

6.3 School

The results for the school scenario are as follows:

- When the vaccines are limited, teachers are give more importance than students which is expected as teachers are the link between two fully connected classrooms.

- Even in extreme cases with family sizes or class sizes, the teachers are still given preference over students and their families.

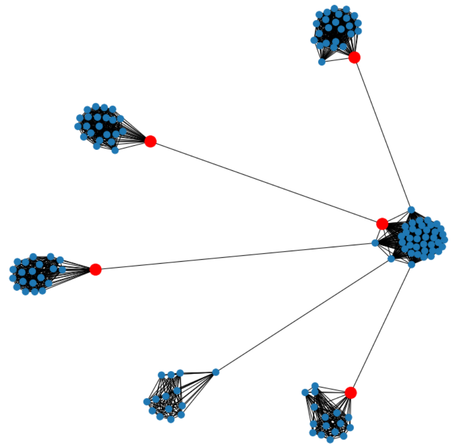

Figure 8: The solution based on method 2 approach to the nursing home simulation with 5 employees, 35 residents, 100 outsiders, and 5 vaccines

- The vulnerability of students or families presents a twist on the results. If the number of vaccines is limited to less than the number of teachers, the vaccines will be distributed to the teachers who have vulnerable students or to the students with vulnerable family members first, before other teachers. Additionally, in extreme cases of vulnerability, it is possible for the student to outweigh the teacher.

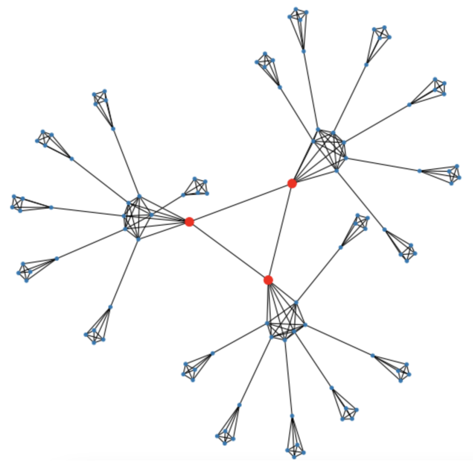

Figure 9: The solution based on method 1 to the school simulation with 3 teachers, class size of 6, family size of 5, and 3 vaccines

Utilizing the alternative approach of spectral clustering, we find similar conclusions of administering the limited vaccines to the teachers.

From the results so far, comparing method 1 and method 2(as introduced in Section 4), we see that method 1 approach is more robust as method 2

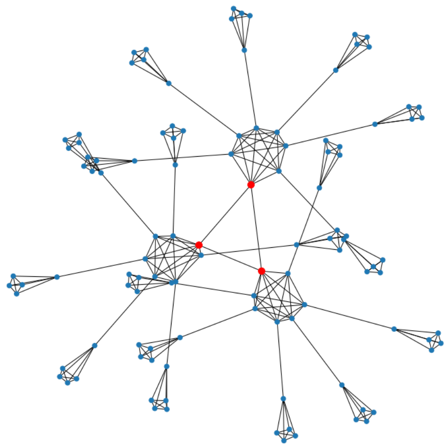

Figure 10: The solution based on method 2 approach to the school simulation with 3 teachers, class size of 6, family size of 5, and 3 vaccines

gives undesirable results on nursing home graph. We also identified a real life scenario conveying method 2 being less robust than method 1. The scenario being if there are two weakly connected clusters and one cluster being full of highly vulnerable people and other cluster being full of very low vulnerable people. In that case, method 2 will select one person each from the two groups but method 1 takes vulnerability into account. Thus, for the community graph as introduced earlier, we run the more robust method 1 instead of method 2.

6.4 Community

Altering the various parameters of each model gives us different results for the community simulation, but the consensus when the number of vaccines is limited is to give preference to the nodes with the strongest influence in the graph, while minimizing the distance from the subset to the nodes not selected. Specifically, teachers and employees (grocery store and nursing home) were almost always given preference, except in the case the graph randomly created other nodes, such as a family member, with a large number of connections.

For the simulation in 4, we obtain results containing a majority of high influence nodes (employees, nurses). For 25 vaccines, we see that the limited vaccines should be distributed to the following individuals, as visualized in 11:

Node	Count
Teacher	1
Student	6
Family Member	3
Employee	3
Resident	0
Nurse	10
Unnamed	2

The nurses are given the largest preference in this simulation, with the majority of the remaining vaccines distributed to students, families, and employees. The other nodes selected can be attributed to the random construction of the graph, whereupon a student or unnamed node may have a high influence.

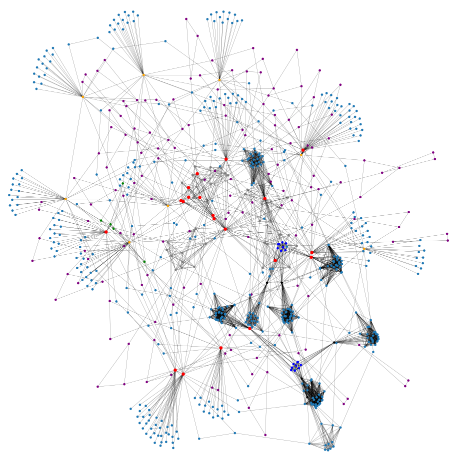

Figure 11: The solution based on method 1 to the community model with 1 school, 3 grocery stores, and 2 nursing homes

When the vulnerability is altered for specific nodes, we see that the algorithm tends to favor those in extreme cases, but in general distributes the vaccine to those with a large influence on the graph. For the simulation in 4, with the following vulnerability distribution:

Node	Vulnerability
Teacher	0.98
Student	1.1
Family Member	0.93
Employee	0.97
Resident	0.93
Nurse	1
Unnamed	1

We obtain the following vaccine distribution for 25 vaccines:

Node	Count
Teacher	0
Student	1
Family Member	1
Employee	11
Resident	6
Nurse	5
Unnamed	1

These greatly different results can be attributed to the distribution of the vulnerability introduced to the problem. Specifically, by increasing the vulnerability of older populations, such as the residents and family members, our algorithm gives preference to those more vulnerable to the virus, but also weighs the importance of network influence, as seen by the number of employees and nurses. A significant point is also the number of nurses chosen by the algorithm due to their close proximity to those more vulnerable.

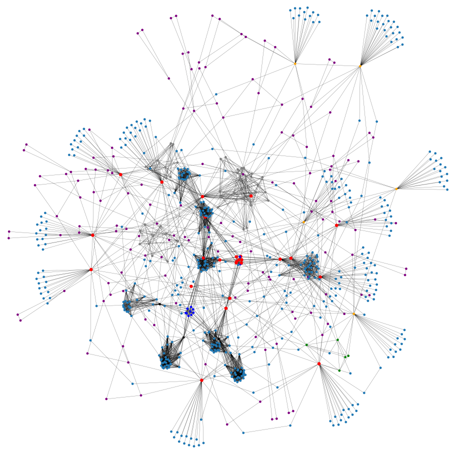

Figure 12: The solution based on method 1 to the community model with 1 school, 3 grocery stores, and 2 nursing homes. Accounts for vulnerability, with residents in nursing homes being most vulnerable, and students being least vulnerable

7 Conclusion & Future Work

Currently, companies like Google and Apple have developed and released ways for users to opt-in to programs tracking COVID exposure using Bluetooth signals from nearby phones, which could produce the data for the algorithm (Gebhart and Gennie, 2020). Similarly, vulnerability for individuals can be assessed by people's health-care profile. By using this data we have formulated a convex optimization problem which minimizes the transmission of COVID. We show results of experimentation with different simulations and randomized graphs display the importance of effective distribution to individuals who come into contact with large numbers of people, even over more vulnerable populations. Groups such as grocery store employees, nursing home workers, and teachers are examples of people who, when administered the vaccine, would have the largest reach. This presents a complementary method which can be implemented along with the current COVID-19 vaccine distribution efforts prioritizing the vulnerable populations.

The applications of this project are multifold, and not solely limited to COVID-19, but rather a wide variety of fields, such as pyramid scheme optimization, knowledge distribution, social network influencers, and resource distribution.

Other future work may include analyzing non-symmetric vulnerabilities between two nodes. For example, in our current algorithm, we consider an interaction between two people to be equal, but in reality there are various factors that could impact the effect of the exchange, such as the absence of a mask for one person. Adding this aspect could make the algorithm more robust and convert the algorithm from undirected edges to weighted directed edges. Additionally, we have not taken into account inherent biases such as people working at hospitals. Although this could be induced using the vulnerability, future work in this area could improve this optimization project.

8 Task Assignments

- Literature Survey - Kunal Jain, Shivam Lakhotia, Sothyrak Tee Srey, Rohan Bhushan
- Data ingestion - Kunal Jain, Shivam Lakhotia, Sothyrak Tee Srey, Rohan Bhushan
- Algorithm implementation - Kunal Jain, Shivam Lakhotia, Sothyrak Tee Srey, Rohan Bhushan
- Experiments - Kunal Jain, Shivam Lakhotia, Sothyrak Tee Srey, Rohan Bhushan
- Report - Kunal Jain, Shivam Lakhotia, Sothyrak Tee Srey, Rohan Bhushan

References

Covid-19 vaccine phases.

Interim public health recommendations for fully vaccinated people.

Who coronavirus (covid-19) dashboard.

Sibo Cheng, Arcucci Rossella, Christopher C Pain, and Yi-Ke Guo. 2021. Optimal vaccination strategies for covid-19 based on dynamical social networks with real-time updating. *arXiv preprint arXiv:2103.00485*.

Lisa Elkin, Ting Kei Pong, and Stephen Vavasis. 2013. Convex relaxation for finding planted influential nodes in a social network.

Bennett Cyphers Gebhart and Gennie. 2020. Apple and google's covid-19 exposure notification api: Questions and answers.

Chungu Guo, Liangwei Yang, Xiao Chen, Duanbing Chen, Hui Gao, and Jing Ma. 2020. Influential nodes identification in complex networks via information entropy. *Entropy*, 22(2):242.

Maheshi N Ramasamy, Angela M Minassian, Katie J Ewer, Amy L Flaxman, Pedro M Folegatti, Daniel R Owens, Merryn Voysey, Parvinder K Aley, Brian Angus, Gavin Babbage, et al. 2020. Safety and immunogenicity of chadox1 ncov-19 vaccine administered in a prime-boost regimen in young and old adults (cov002): a single-blind, randomised, controlled, phase 2/3 trial. *The Lancet*, 396(10267):1979–1993.

Simone Santini. 2020. Covid-19 vaccination strategies with limited resources – a model based on social network graphs.

T Shen. 2021. The mathematics behind spectral clustering and the equivalence to pca.

Keyword Team. 2020. Exposure notification api launches to support public health agencies.

James Voss, Mikhail Belkin, and Luis Rademacher. 2016. The hidden convexity of spectral clustering.

Yu Zhang and Yan Zhang. 2020. Top-k influential nodes in social networks: A game perspective.