

# Social Computing (CSC 555)

## Community Detection for Epidemic Dynamics

Piyush Tiwari (ptiwari), Kavita Kumari (kkumari)

November 16, 2020

### 1 Introduction

Our idea is to detect communities in social networks to prevent the spread of infectious diseases. Our idea is mainly based on [3], [2] and [1]. Community is defined as a group of users who interact with each other more frequently than with people outside their group. We can accomplish multiple tasks such as friends recommendation, network analysis and visualization, etc. by identifying community structure. In the real-world network, we have the content information of the network along with their topological structure. Our idea is to detect the community in the social network that would help us detect the spread of infectious diseases like COVID-19. We will try to represent the usage of community structure with respect to epidemic dynamics and how we can control the spread using these communities detected through social networks. Since COVID-19 is a virus that spreads very quickly when people come in close contact, our aim is to find if social network communities can help us in identifying the spread.

For any new pandemic, it will take time to find the medicine. So the best possible way to stop the spread is by social distancing but in the large scale to follow the social distancing is not easy. So we thought about the idea to detect the community and minimize the effect of Pandemic as much as possible. For example as per the current scenario the entire world is impacted by COVID and the most proven idea to control its spread is social distancing, our aim is to understand if maintaining social distancing within the social network community would help by detecting the disease spread within the community[1]. Another idea that we have implemented is how vaccine availability can contribute towards reducing the death rate due to infection spread.

### 2 Description of data

We will be using the CALTECH36 dataset for our experiment[5]. This dataset is extracted from Facebook. People are represented by nodes and friendship is represented by edges. There are 769 nodes and 16.7K edges in the dataset. It has a max-

imum degree of 248 with a degree average of 43 (<http://networkrepository.com/socfb-Caltech36.php>) . Apart performing our experiment on CALTECH36, we will scale our model for larger datasets which are as follows-

Dataset	No. of nodes	No. of edges	Max Degree	Average Degree
PRINCETON12	6.6K	293.3K	628	88
OKLAHOMA97	17.4k	892.5K	2.6K	102
GEORGETOWN15	9.4k	425.6k	1.2k	90
CMU	6.6K	250K	840	75

### 3 Related Work

Our idea is mainly based on [3]. The paper talks about how the structure of community affects the epidemic dynamics. It also explains the importance of community structure to control infectious diseases. This paper is based on the standard SIR model. The results show the impact of change in community structure connectivity on the average epidemic size, duration and the peak. The paper also talks about the importance of immunization and social distancing in controlling the spread of an epidemic. [1] explains the impact overlapping communities on the susceptible-infected-susceptible (SIS) epidemic spreading model. Overlapping communities contribute significantly to the spread of epidemic and sometimes leads to a peak of the spread velocity [1]. [2] talks about different applications of community detection and we have referred it mainly to understand the scope of community detection for real issues like epidemic spread, recommendation systems, etc.

We are trying to identify ways in which we can reduce or stop the spread of the virus by utilizing social network. Our first step is to identify the communities in a social network. We will identify weak ties (critical connections) and then see if removing weak ties would help in controlling the virus spread.

For influence propagation, we have referred [6]. This paper describes social influence as the situation when individuals adopt a new action because of their social network (people in their network have adopted). The experiments are based on Polly data. Polly is an application which allows senders to record a short message. Poly was mainly used as a platform for voice-based information services for low-skilled, low-literate people. The paper focuses on two things - 1. Pattern Discovery - Given who influences whom and when, find a general influence pattern obeyed by the network. Patterns can help us understand how influence propagates in social networks and also to spot anomalies, like spammers, or faulty equipment. The authors have described and experimented on 3 patterns - Fizzle, Rendezvous, and Dispersion. 2. Generator - given a friendship social network, design a simple, local propagation mechanism, to generate realistic looking influence networks. For this, the paper proposes a propagation mechanism that simulates an influence graph on top of existing social networks.

## 4 Baseline

We created a cluster of networks using K-means clustering. We will affect random people in any community and evaluate the time steps taken by the K-means algorithm to achieve maximum propagation. Our vaccine simulation model is based on the standard susceptible-infected-resistant (SIR) model as the baseline.

## 5 Hypothesis

- Community structure affects the spread of diseases in networks. Mostly Epidemic spreading takes place between overlapping communities whose members are highly linked.
- By identifying and removing critical connections, we can reduce the spread of the disease.
- Approximately 90% death rate will reduce after vaccine introduction.
- Approximately 80% spread of Corona Virus will be controlled after vaccine introduction.

## 6 Experiment and Setup

### 6.1 Community Detection

As per our understanding we have to identify the weak ties between the communities and represent if we can separate the communities with these weak ties, we can control the spread. There are  $n$  nodes in the network connected by undirected friendship connection, forming a network where  $\text{connection}(u,v)$  represents a connection between person  $u$  and  $v$ . Any person of one community can be connected by any other person directly or indirectly through the connection. We define a critical connection as a connection, that if we remove, it will result in a situation in which some communities won't be able to reach some other community. We will come up with the critical connections in the network and identify if their removals help in controlling the spread.

In real life, if 2 people know each other and a third person, who knows 1 of them, ends up knowing both of them. So if any member of one closed group is connected to any member of another small group, both the groups in a long run ends up in knowing each other and they can meet each other on some occasion. This spreads the pandemic quite fast. Our aim is to find such critical connections and stop them to get mingled up all together. This will reduce the spread of infection in a large amount [1].

We mainly focus on external connections of community instead of internal connections. Our input is a graph where students are the vertices and friendship is the edges between them. Our approach is based on the modification of

Newman's well known modularity function. If a graph has  $n$  nodes and  $m$  edges,  $G_{i,j}$  is the link between  $(i,j)$ ,  $d_i$  is the degree of node  $i$  and graph is partitioned into  $K$  clusters, then Newman's modularity[4] is

$$Q_{Newman} = \sum_{l=1}^K \sum_{i \in C_l, j \in C_l} S(i, j) \quad (1)$$

where  $S(i,j)$  is the link strength between two nodes  $i$  and  $j$  and is measured by comparing the two network interaction  $G_{i,j}$  with the expected number of connections  $(d_i \cdot d_j) / 2m$

$$S(i, j) = \frac{1}{2m} (G_{i,j} - \frac{d_i \cdot d_j}{2m}) \quad (2)$$

The Next step is find a appropriate optimization of  $Q$ . To do this we start with each node belongs to separated community. A node is then chosen randomly. We will try to move this node from its current community. If a positive gain is found then the node is placed with the community with maximum gain. otherwise it will remain in its original community. We will reiterate this steps until no more gain is found. When moving node  $x$  to community  $C$ , the composite modularity gain[4] is calculated as

$$\delta Q_{Newman} = \sum_{i,j \in C \cup x} S(i, j) - \sum_{i,j \in C} S(i, j) = \frac{1}{2m} (\sum_{i \in C} G_{i,x} - \frac{d_x}{2m} \sum_{i \in C} d_i) \quad (3)$$

After finding the communities we have broken all the ties between any 2 communities and evaluate the time steps to propagate the virus throughout the networks via influence propagation algorithm. We have identified the communities in a network and identified weak ties.

---

**Algorithm 1:** Structure based Community Detection Algorithm

---

**Result:** A set of Communities

initialization: A Graph  $G(V, E)$ ;

**while**  $i \in V$  **do**

**while**  $j \in V$  **do**

        Remove  $i$  from its community, place to  $j$ 's community;

        compute the composite modularity gain  $\delta Q$  **end**

**end**

    Choose  $j$  with maximum modularity gain(if exists) and move  $i$  to  $j$ 'th community;

    Otherwise  $i$  stays in its community;

---

## 6.2 Influence Propagation

We have implemented influence propagation by targeting influential users in the social network. We have infected an entity in each community and measured how

fast the infection propagates over the entire network. We have taken the number of influenced nodes at every 15 time steps for our analysis. We have compared our approach with k-means clustering and observed significant reduction in spread.

### 6.3 Covid Vaccine Simulation

In covid Vaccine simulation we tried to find the effect of vaccine in the spread of Covid-19 as well as the effect of vaccine on the casualties due to Covid-19. We modeled the spread of virus as given in Fig1.

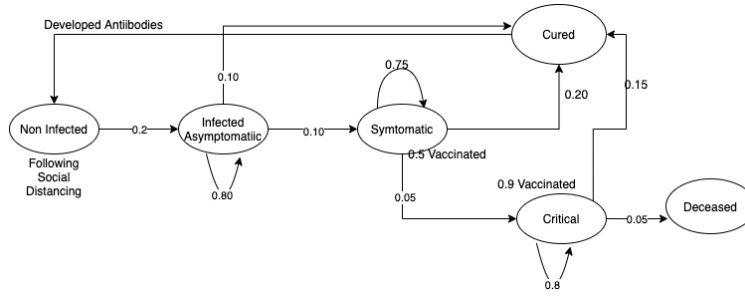


Figure 1: Spread of Virus

We experimented taking assumptions that after the vaccine would arrived, it will not be available to everybody immediately. So the availability of the vaccine is one of the variable that we experimented with, in the same time the effectiveness of vaccine is also not fixed so this is another variable on which we have experimented. Below is the state chart of the patient pathway of a person-agent in the agent-based COVID-19 model.

In simulation we tried to copy the current world scenario where people are following the Social Distance Norm and there is Quarantine centers as well as Home Quarantine. A person agent can go to social gathering and shopping and the probability of getting affected is 0.25 and 0.15 respectively. After that whomever the infected agent will meet, the infected rate we have experimented with 20%, 30% and 50% respectively. For asymptomatic agents we have not vaccinated them but for symptomatic agent because of availability of vaccine we experiment with 50% and 75% vaccine given. The effectiveness of medicine is around 90%. So some patient can go to critical stage and if vaccine is available all the agents got the vaccine in this stage. Still some needs Hospitalization and quarantine themselves in Quarantine Center. The success rate for critical patient to get cured after vaccine was taken around 90% so some of them deceased. After get cured the patient has developed the antibodies so the probability of again getting infected is very less now. We have experimented with this scenario.

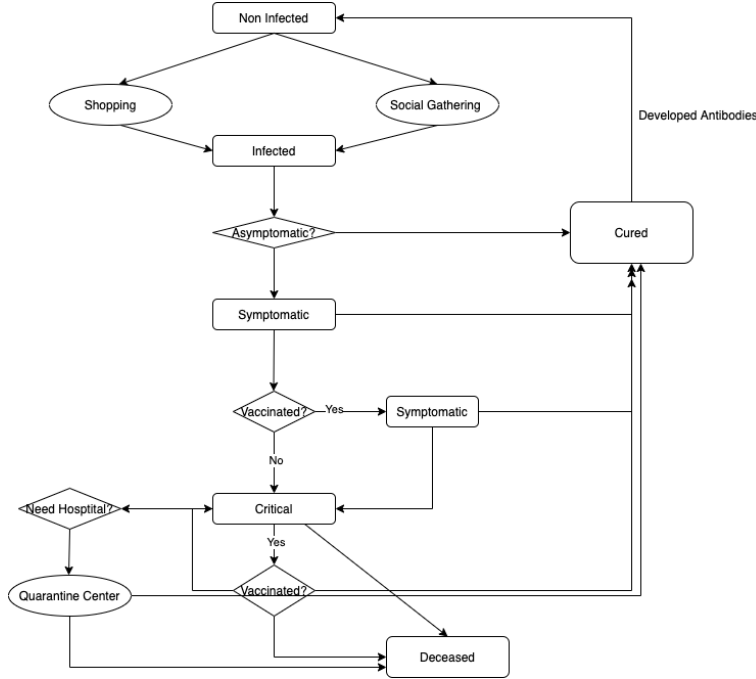


Figure 2: State chart of the patient pathway of a person-agent in the agent-based COVID-19 model.

## 7 Evaluation of Hypothesis and Results

We have used K-means algorithm to detect community as the baseline. We have evaluated our method by influence a person in each community and measured how fast the influence propagates over the entire network. The faster that influence propagates through the entire network, the more likely it will impact the broader population. We have measured and compared the time steps taken by the influence propagation algorithm to achieve maximum propagation and compare this with our baseline. We have run our model for caltech data and run the model on different different data mentioned in dataset. The graph is shown in fig-1 demonstrates that the propagation of virus when we used K-means clustering spreads more quickly than detecting the weak ties between communities through the algorithm we proposed.

We found that removing the weak ties from the network helps in controlling the spread of the virus. So from the result we can verify both the hypothesis we proposed earlier are true.

The simulation results can be seen in the below graphs(Fig 5, Fig 6 and Fig 7). In simulation without vaccine in 50 steps for same infection rate as for vaccine we got 245 causalities while with vaccine when available for 50% patients we got 45 causalities and when available for 70% user we got 38 causalities for

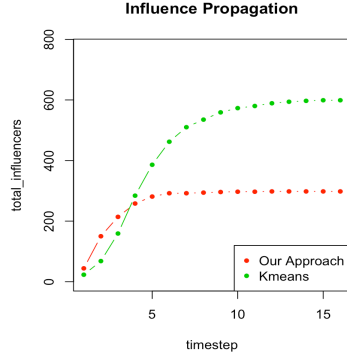


Figure 3: Community Spread with 25% initial spread

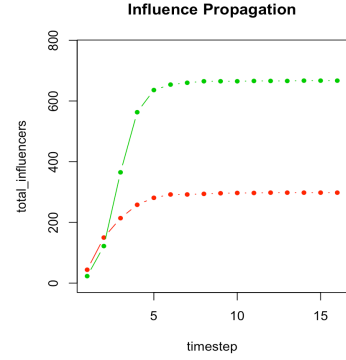


Figure 4: Community Spread with 50% initial spread

taking average over 1000 simulations. So from our experiment we got around 81% and 85% decrease in death when vaccine is available. So our hypothesis holds true in our case.

For the second hypothesis, the spread of Covid 19 will decrease by 80% has not hold true as after introducing although the death rate will decrease drastically still the spread of covid 19 will not decrease much in near future immediately. We have run the experiment on the above mentioned datasets and found similar results.

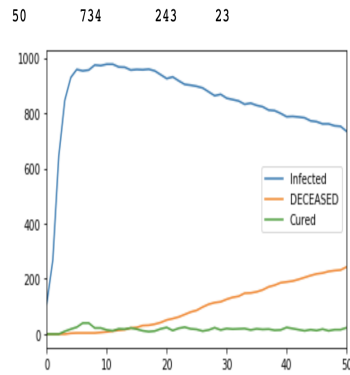


Figure 5: Death Simulation without vaccine

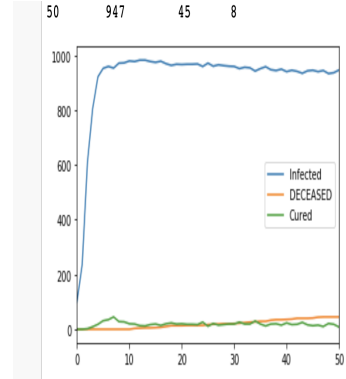


Figure 6: Simulation with vaccine availability 50% and effectiveness 90%

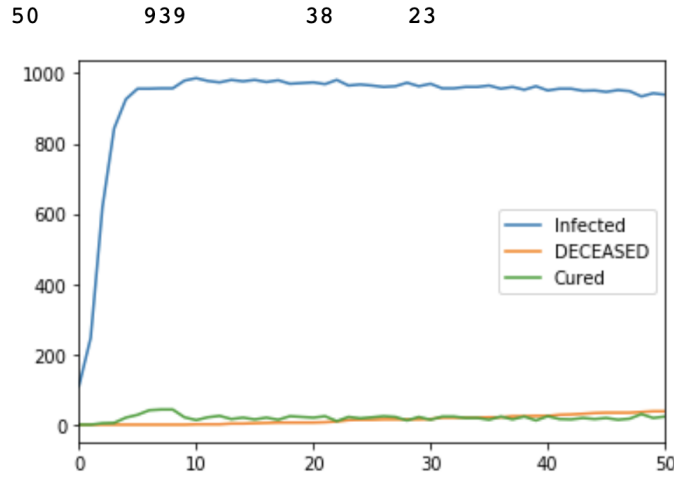


Figure 7: Simulation with vaccine availability 70% and effectiveness 90%

## References

- [1] Zhi-Hong Guana Jiancong Chena, Huiling Zhanga and Tao Li. *Epidemic spreading on networks with overlapping community structure*. 2011.
- [2] Neha Mini Singh Ahuja, Jatinder Singh. *Practical Applications of Community Detection*. 2016.
- [3] Marcel Salath and James H. Jones. *Dynamics and Control of Diseases in Networks with Community Structure*. 2010.
- [4] Emmanuel Viennet The Anh Dang. *Community Detection based on Structural And Attribute Similarities*. 2012.
- [5] Mucha P Porter M Traud A, Kelsic E. (2009) *Community Structure in Online Collegiate Social Networks. Proceedings of the 2009 APS March Meeting (16–20 March 2009)*. 2009.
- [6] Jay-Yoon Lee Danai Koutra Roni Rosenfeld Christos Faloutsos Yibin Lin, Agha Ali Raza. *Influence Propagation: Patterns, Model and a Case Study*.