Prediction of a Pandemic Outbreak in a Community: A Novel approach based on Complex Networks

Chejarla Santosh Kumar
Dept. of Computer Science and Engineering
Amrita School of Engineering, Amrita Vishwa Vidyapeetham
Amritapuri, Kerala
Ch.santoshkumar211@gmail.com

Nimisha Dughyala
Dept. of Computer Science and Engineering
Amrita School of Engineering, Amrita Vishwa Vidyapeetham
Amritapuri, Kerala
nimishadugyala@gmail.com

Akula Sree Deepthi
Dept. of Computer Science and Engineering
Amrita School of Engineering, Amrita Vishwa Vidyapeetham
Amritapuri, Kerala
deepthideepu015@gmail.com

Balaji Vamshi Pasala
Dept. of Computer Science and Engineering
Amrita School of Engineering, Amrita Vishwa Vidyapeetham
Amritapuri, Kerala
Balajivamshi.pasala@gmail.com

Abstract— **Every pandemic has a calculated value associated with it called "basic reproduction number", which represents on average to how many healthy persons can get infected from an infected person, but how far are the people from each other in a community is a question. The answer can be approximated by calculating the "Average Degree" and "Degree of Separation" of a community network, especially friendship network because the probability of a person to meet the other person is high if the other person is a friend or colleague of him compared to his neighbor even the neighbor is nearer to him. On comparing the "Average Degree" and "Degree of Separation" with the "basic reproduction number" value we can calculate how far an individual in a community is from being infected from the pandemic and how many steps needed by the pandemic to infect the entire community. In this paper, we will present a new approach to predict the outbreak of a pandemic in a community and will also analyze the outbreak of COVID-19 with other pandemics happened in history.**

Keywords—Social Networks; Degree of Separation; Small world; Pandemic; COVID-19; R naught

I. INTRODUCTION

Since the advent of social networks, we have made numerous analyses on the data from these networks and made many conclusions and predictions. If we consider a friendship network, with each node representing a person and an edge between two persons, represents the friendship between them, the "Average Degree" of the network represents how many friends on average a certain person have in the community, Fig. 1. Illustrates "Degree" [1] of nodes in a network and "Degree of Separation" [2 - 5] is another measure we can calculate using the network which represents – chosen two random nodes in a network, how many intermediate edges are present in the shortest path between the nodes. Fig. 2. Represents a six degree

of separation i.e. 'A' is connected to 'B' by a chain of five intermediate nodes/persons which represents a chain of friends.

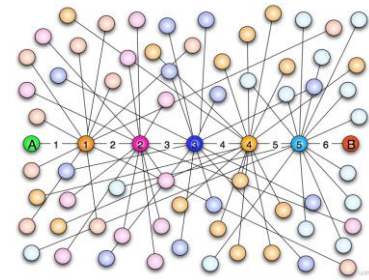Fig. 1. Degree of Nodes in a Network (quantamagazine.org)

Fig. 2. Degree of Separation (Wikimedia.org)

On comparing the values with "basic reproduction number" [6] denoted by "R0" we can predict the outbreak of a pandemic in any community. "R0" value represents how many cases can be infected from a single case that is infected with the infection. An infection with "R0" < 1 represents that the disease/infection can't spread but infection with "R0">1 represents that the infection will spread in the community if measures were not taken.
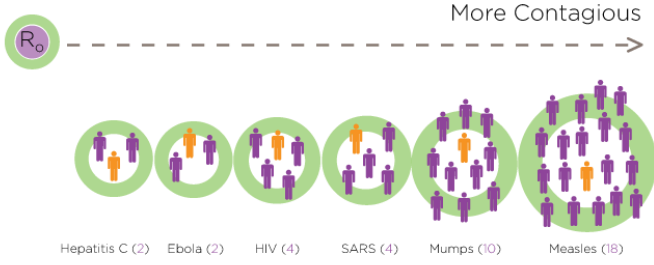
Fig.3. R0 values of Different Pandemics (healthline.com)

With this paper we would like to present an approach to predict the outbreak of a pandemic in a community based on the parameters "Average Degree", "Degree of Separation" and "R0" and will also perform analysis on data collected from Facebook from different University communities to simulate the outbreak and compare the outbreak of COVID-19 with other pandemics.

II. METHODOLOGY

To predict the outbreak of a pandemic in a community we suggest to calculate "n-steps" and "Infection Distance" based on "Average Degree" (let it be denoted by 'θ'), "Degree of Separation" (let it be denoted by 'α') and "R0".

1. *n-steps*: It is the ratio of "Average Degree" (θ) of the network to the "R0", which represents how many steps required by the pandemic to spread across the entire community.

$$n - steps\ = θ/R0$$

Average Degree (θ): Average Degree of the network is the ratio of the sum of the degree of all the nodes in the network to the total number of nodes.

$$θ = \frac{\sum_{i=0}^{n} \deg(i)}{n}$$

deg(i) – Degree of Node i
n – Number of nodes in the network

2. Infection Distance: It is the difference between 'α' and 'R0', it represents on average how many persons away is the healthy person from an infected person.

$$Infection\ Distance = α - R0$$

Degree of Separation (α): Degree of Separation in a network is the ratio of the sum of shortest path lengths from every node to every other node in the network to 'nC2'. Here the shortest path length represents the number of intermediate edges from the origin node to the target node and 'nC2' represents how many connections/friendships can be possible in the network with 'n' nodes.

$$α = \frac{\sum_{i=0}^{n} \sum_{j=0}^{n} d(i,j)}{nC2}$$

d(i,j) = shortest path length from node i to node j

n = no. of nodes in the network

i, j ∈ nodes in the network

Relatively low values of "*n-steps*" and "Infection Distance" compared with the other pandemics in the history, suggest that the rate of spread of the new pandemic in the community is high and Relatively high values of "*n-steps*" and "Infection Distance" compared with the other pandemics in the history, suggest that the rate of spread of the new pandemic in the community is low.

III. ANALYSIS

We made the analysis on two Facebook datasets collected from Stanford and Auburn University which are available in networkrepository.com [7] completely anonymized by replacing the usernames with numbers. The nodes in the dataset represent persons and an edge between two persons represents that both the persons are friends.

1. STANFORD3:
   On analyzing this network built from this dataset consisting of 11586 nodes and 568309 edges, which is illustrated in Fig. 7, the following results were obtained:
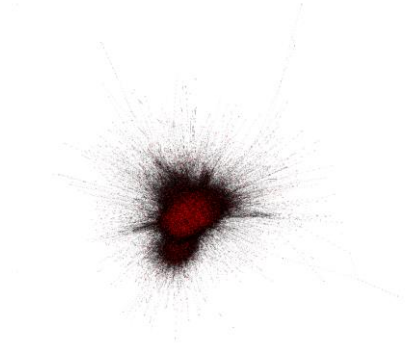   - θ = 98.102
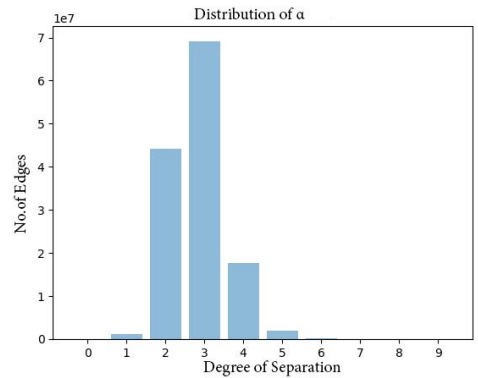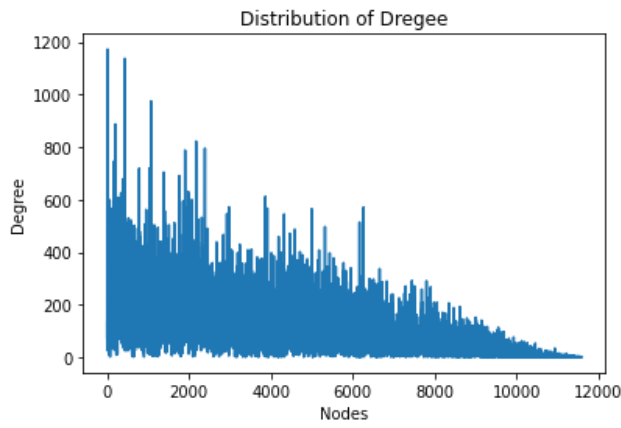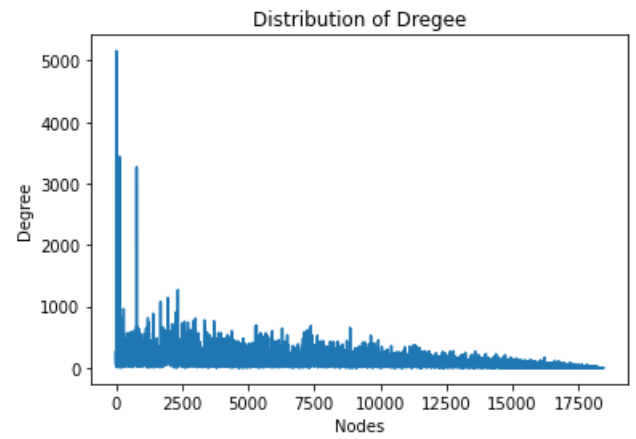   - α: 2.818

Fig. 7 Network

Fig. 8

Fig. 9

Fig. 13

2. AUBURN71:

On analyzing this network built from this dataset consisting of 18448 nodes and 973918 edges, which is illustrated in Fig. 10, the following results were obtained:
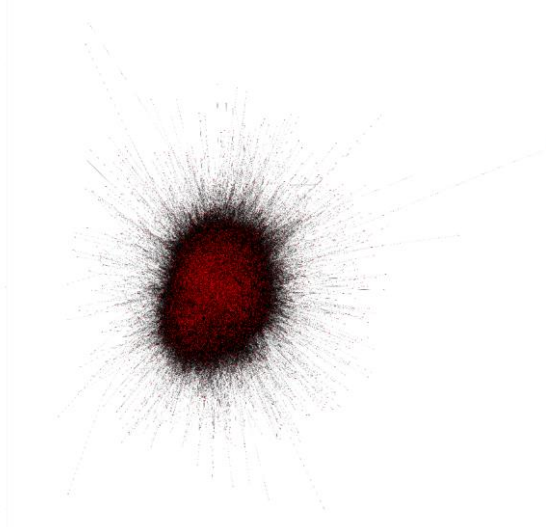
- θ = 105.585
- α: 3.286

Fig. 10 Network

Fig. 12
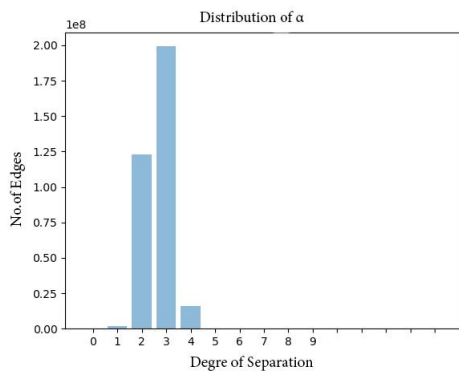
Pandemic	R0 [6][8]	Dataset	*n-steps*	Infection Distance
MERS	0.3-0.8	1	327-122	2.518-2.018
		2	351-131	2.986-2.485
Flu	1.3	1	75.463	1.518
		2	81.219	1.986
Ebola	2	1	49.051	0.818
		2	52.792	1.286
HIV	3.6	1	27.250	-0.782
		2	29.329	-0.314
COVID-19	2.2	1	44.591	0.617
		2	47.993	1.085

Table. 1 Summary of Datasets

All the R0 values are from the period of their outbreak, the R0 values might have changed now as the vaccinations of many are available now, as for a new pandemic outbreak prediction we need to consider the R0 values at the times of their outbreak, so, these values are relevant for the prediction.

From Table 1. We can observe that the "*n-steps*" and "Infection Distance" values of major pandemics are relatively low compared to not so widespread pandemics, the table clearly explains the outbreak of COVID-19, if we have a person infected with COVID-19 in the community then on average each person one in the community is not more than 0.617-1.085 persons away from the infected person and COVID-19 needs just approximately 44.5 - 47.9 steps to infect the entire community if no measures were taken and if a person in the community is infected with MERS then on average each person in the community is 2.0-2.9 persons away from the infected person and MERS need approximately 122-351 steps to infect the entire community if no measures were taken. The comparison of "*n-steps*" and "Infection Distance" between COVID-19 and MERS shows the wide outbreak of COVID-19 and also supports our approach to predict the outbreak of pandemics in any community.

IV. CONCLUSION

In this paper, we have proposed a new approach to predict the outbreak of a pandemic in a community based on "Average Degree", "Degree of Separation" and "R0" values by analyzing the community networks and also implemented the approach on the datasets collected from Facebook and supported our approach with the obtained results.

V. REFERENCES

[1] "Degree distribution", En.wikipedia.org, 2020. [Online]. Available: https://en.wikipedia.org/wiki/Degree_distribution. [Accessed: 03- Apr- 2020].

[2] "Six degrees of separation", En.wikipedia.org, 2020. [Online]. Available: https://en.wikipedia.org/wiki/Six_degrees_of_separation.

[3] "Three and a half degrees of separation - Facebook Research", Facebook Research, 2020. [Online]. Available: https://research.fb.com/blog/2016/02/three-and-a-half-degrees-of-separation/.

[4] Zhang, Lei & Tu, Wanqing. (2009). Six Degrees of Separation in Online Society.

[5] Bakhshandeh, Reza & Samadi, Mehdi & Azimifar, Zohreh & Schaeffer, Jonathan. (2011). Degrees of Separation in Social Networks. Proceedings of the 4th Annual Symposium on Combinatorial Search, SoCS 2011.

[6] "Basic reproduction number", En.wikipedia.org, 2020. [Online]. Available: https://en.wikipedia.org/wiki/Basic_reproduction_number.

[7] Ryan A. Rossi and Nesreen K. Ahmed, The Network Data Repository with Interactive Graph Analytics and Visualization, AAAI, 2015. Available: http://networkrepository.com

[8] "How contagious is Zika compared to other diseases? Here's how researchers figured that out.", Vox, 2020. [Online]. Available: https://www.vox.com/2016/4/19/11450396/zika-virus-contagious-ebola-measles. [Accessed: 03- Apr- 2020].