Modeling, Estimating, and Minimizing Average New Infections of COVID-19 Based on Information Theory and Social Networks

Hussein Al-Hamdani[1,2], Mustafa Alassad[1], Nitin Agarwal[1], and Seshadri Mohan[1]

[1]University of Arkansas at Little Rock, Little Rock AR, USA
[2]University of Kerbala, Kerbala, Iraq
{haalhamdani,mmalassad,nxagarwal,sxmohan}@ualr.edu

Abstract. This work proposes a novel model targets to utilize Shannon's channel capacity equation to formulate the spreading behavior of COVID-19 within a population and minimize this spreading by optimizing this equation using a proposed mobility model based on the focal structure analysis of social networks. This model facilitates the early detection of the hotspots and estimation of future infections. Based on the model, the average number of new infections reveals an interesting relationship with the total number of infected people when the population approaches infinity. Specifically, at the beginning of the pandemic, the model shows that the average number of new infections increases logarithmically. Meanwhile, the increment tends to flatten when the total number of infections exceeds a critical ratio of the population. The simulation results manifest that the approach leads to a massive reduction in the average number of new infections. Consequently, the proposed model can help healthcare systems to overcome the pandemic by identifying the hotspots early and facilitating the effective allocation of resources.

Keywords: COVID-19, focal structure analysis, information theory, virus' spread, mobility model, social networks, Shannon's channel capacity equation

1 Introduction

CORONAVIRUS disease of 2019 or (COVID-19) is a disease caused by the SARS Cov2 virus. On January 31st, 2020, the World Health Organization (WHO) declared that the entire world would be suffering a new pandemic caused by the COVID-19 virus. Since this declaration and at the time of writing this article, the virus has affected more than 20 million people worldwide, resulting in more than 800K deaths, of which more than five million cases and 163K deaths are from the USA alone [1]. Moreover, many reports around the world have predicted a second wave of the pandemic [2]. Therefore, restrictions have been issued by governments to attenuate the effects of the virus wave. So far, neither an ideal cure nor a vaccine or even a reliable road map for the treatment is available, although the number of cases has increased recently [3]. Many studies have modeled the spread of SARS, Ebola, and many other pandemics in the last two decades [4].

The well-known Shannon's channel coding theorem to achieve a reliable communicating for the information over the Additive White Gaussian Noise (AWGN) channel [8], [9] states that an information source can be transmitted with an arbitrarily low probability of error provided the rate of information transmission does not exceed channel capacity or equivalently the mutual information (the speed) in the channel should not be greater than the capacity of that channel.

$$C \geq I\left(X,Y\right) = B\log_2\left(1+\frac{S}{N_0 B}\right) \tag{1}$$

where C represents the channel capacity, B the bandwidth, N_0 the noise power spectral density, and S/N_0 the signal-to-noise power ratio. While applying equation (1) to model COVID-19, we let C or $I(X,Y)$ represents the average number of new infections in a population, S the virus power, and B the population of a city, state, or country; N_0, the noise power spectral density, now represents in some form the power of the mitigation techniques employed. Unlike in communication theory, the mutual information needs to be minimized such that the spreading of the virus is reduced or slowed. While the virus power S and the population B are uncontrollable factors, N_0 is controllable and is the critical variable to minimize the spreading of the virus.

To estimate the average number of new infections, we assume the population approaches infinity. In this case, as the analysis reveals, the curve of mutual information exhibits two distinct phases. While, during the beginning phase of the pandemic, the curve exhibits a logarithmic increment in the average number of new infections, beyond a certain threshold determined by the total number of infections exceeding a critical percentage of the population, a semi-stable new infections case can be reached.

This work would implement different theories that well-known methods utilized to track COVID-19 spread. The model presented here adapts an optimized mobility model based on social networks to break the spreading chain of the virus such that there is no virus particle carried through the community by maximizing N_0, alternatively, by minimizing mutual information (X,Y).

The model employs social network analysis to study the relationships among individuals and their implications and a combination of social networks and smartphone networks to examine the users' face-to-face interactions. While social networks provide information about the gatherings, events, and all kinds of information exchanges, smartphone networks will provide valuable information about the dynamics of social movements and face-to-face interactions.

Section 2 formulates the information-theoretic model of the virus spreading and the social mobility model for its mitigation. However, section 3 discusses the simulation results, and finally, section 4 provides the conclusions observed from this work.

2 System Model

2.1 Information-theoretic model for virus spreading behavior

Let $X = [x_1, x_2, x_3, \ldots\ldots, x_n]$ be a vector of n components that represents the sources of the virus, and each component is an independent and identically distributed (i.i.d) Gaussian random process. By considering the population approaches infinity [10], and denoting $\frac{S}{N_o B} = x$, the Shannon's equation (1) can be written as show below. Equation (2) shows that at the beginning of the pandemic, the number of new infections increases logarithmically. However, equation (3) shows that the average number of new infections is a semi-linear function depends on S and N_0 as the value of x increased. We address the question; what is the extent to which the new virus infection in the population, given by $I(X, Y)$, can be mitigated by optimizing N_0 by applying a social network-based mobility model? Section 2.2 proposes the used mobility model based on social networks to mitigate the spreading behavior of the virus that can save lives.

$$\lim_{x \to 0} I(X,Y) = \frac{S}{N_0} \lim_{x \to 0} \frac{1}{(1+x)\ln(2)} \tag{2}$$

$$\lim_{x \to 0} I(X,Y) = \frac{S}{N_0} \frac{1}{\ln(2)} \cong 1.44 \frac{S}{N_0} \tag{3}$$

2.2 Mobility model for optimizing virus spreading based on social networks

Social network analysis helps us to identify the relationships among people and find the patterns of social interactions and their implications. For this purpose, we introduce a Face-to-Face matrix to keep track of the interactions. The model proposes a way to identify the critical groups (or focal structures [11]) of nodes that can spread the virus across the network [12].

For carrying out social network analysis, a social network is represented by a graph with a set of nodes and edges. The nodes may be individuals or groups of individuals, and edges represent the relationship between nodes. In our analysis, the Face-to-Face matrix provides a way of defining the edges. This matrix, which specifies a dynamic weighted [0,1] undirected graph, is determined by measuring the connection strength between all pairs of nodes (i, j). The weight F_{ij} is a result of three critical parameters, Time (T) reflects the interaction time series, where $X_{ij}(t) = 1$, if node i at time t meets node j, otherwise 0. Distance (D) reflects the minimum distance between node i and node j, where $X_{ij}(d) = 1$, if the distance between node i and node j is less than a certain critical distance to expose both nodes to the virus. Accessibility (A) reflects the physical face-to-face interaction, where $X_{ij}(a) = 1$, if node i can physically access node j, otherwise 0. Table 1 considering the Face-to-Face matrix considers all the possible combinations of parameters to facilitate face-to-face physical interactions.

2.2.1 Proposed Dataset

The selected dataset is an undirected connected graph of a highly dense community consisting of 10000 nodes and more than 150,000 edges collected over twenty days from Facebook [13], as shown in Fig. 1. The users employed Facebook to manage, establish, and participate in big gatherings in Saudi Arabia in 2013. We utilized this network to map the social network platforms and smartphone networks when users used both to implement face-to-face interactions, where any edge represents an interaction between two nodes. Online social networks such as Twitter and Facebook would allow enabling the exact geolocation service on the user's mobile service [3].

2.2.2 Focal Structure Analysis

This section describes how the model works to find the intensive groups of nodes that can influence the entire network by considering both individuals' features and the communities' features [14]. For this purpose, a decomposition model has proposed to maximize the individuals' centrality to find the most central nodes in the network at the individual-level and maximize spectral modularity at the community-level [14]. The outcomes are an influential set of nodes that can influence the maximum number of nodes in the network.

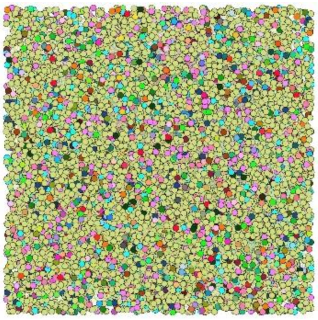

Fig. 1: Network representation.

Table 1: Users' Face-to-Face Interaction Logic.

A	D	T	Face-to-face Interaction $\left(F_{ij}\right)$
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	1

2.2.2.1 Individual-Level Analysis

The degree centrality is employed to measure the individuals' activity, and the number of links each node can visit every timestamp and find the most active nodes in the network.

$$max \sum_{i=1}^{n} \delta_i \tag{4}$$

Equation (4) id used to maximize the users' degree centrality value, where $\delta_i$refers to the degree centrality of node i, n is the number of users in the community.

2.2.2.2 Community- Level Analysis

In this section, the objective is to maximize the spectral modularity values and find the most central sets of nodes that can sparse the network the most. For this purpose, equation (5) has employed in the community-level, where ϱ_j^M refers to the graph modularity and employed to maximize the spectral modularity values

$$max \sum_{j=1}^{n} \varrho_j^M \tag{5}$$

Both levels will exchange information at both the local and network level to find the focal sets of nodes that can sparse the network the most.

3 Results

The spreading behavior of the virus has been modeled based on Shannon's equation in MATLAB; meanwhile, this equation has been optimized based on the mobility model in the Python environment. The simulation results have been earned based on the assumption that the population approaches infinity. Two different populations examples have been used in this work to simulate the proposed model, 10,000, and Arkansas population. The mobility model has optimized the values of N_0 used in section 2.2.2 to minimize the average number of new infections of the given populations. Moreover, the model identified intensive sets in the network that can maximize the network's sparsity at the first day of the data was collected, then the model did select the highest influential set that includes 34 nodes, based on set's and nodes' centrality values and their ability to sparse the network. We assume that this set is infected, or the nodes are the hotspots in the early stage of the outbreak, and all nodes should quarantine to stop the spread.

We consider all of them (34 nodes) to be the virus' power, which represents the number of nodes in the influential sets identified by the focal structure analysis identified on day one (early detection). Table 2 represents their influence in twenty days and how they interact with other nodes in the network.

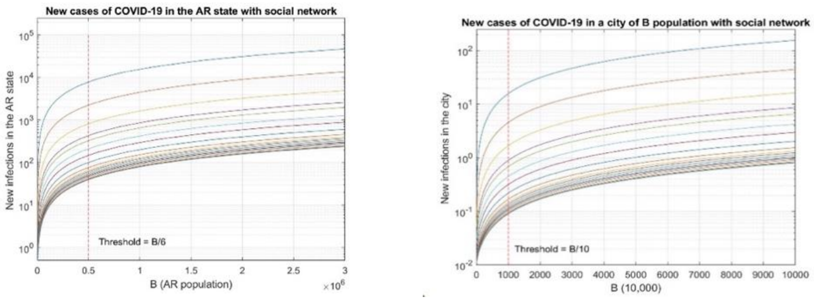

Fig. 2. Left side, the average number of new infections in the AR state. Right side, the average number of new infections in a city of 10,000 population.

Figure 2 shows the behavior of the virus spreading and the mitigation of the virus using the proposed model in the given populations. At the beginning of the pandemic, a massive number of people get infected by the virus. However, the new cases of the infected people attend to fatten when the total infected people exceed a certain threshold of the population. The threshold values are B/6 and B/10 for the State of Arkansas and the city of 10,000 population, respectively.

Table 2: Experimental results represent the influence of the most influential set of central nodes and their ability to spread the virus in the population over time.

<table>
<tr><th>Days</th><th>Number of connected nodes used in the model over time.</th><th>N_0</th><th rowspan="4">$I(X,Y)$
Readings as presented in curves in Fig. 2 based on the population (B)</th></tr>
<tr><td>1</td><td>413</td><td>327</td></tr>
<tr><td>2</td><td>779</td><td>611</td></tr>
<tr><td>3</td><td>1286</td><td>1013</td></tr>
</table>

4	1832	1393	
5	2181	1598	
6	2736	1989	
7	3254	2375	
8	3913	2877	
9	4575	3301	
10	5140	3548	
11	5371	3690	
12	5580	3843	
13	5753	3941	
14	5953	4065	
15	6173	4162	
16	6420	4292	
17	6638	4444	
18	6787	4519	
19	6876	4579	
20	6933	4604	

It is easy to note the effects of adopting the proposed model on both populations will reduce the average number of new infections of COVID-19. If the virus has been detected early within the population, the total number of infected people in Arkansas will be approximately more than 5000 cases. However, when a large separation is adopted based on the suggested communities, the average number of new infections will not be reduced accordingly. The same results with small reduction can be achieved based on the population of 10,000 people.

4 Conclusion

The information-theoretic model based on the mobility optimization model of social networks has been proposed for modeling and mitigating the spreading of COVID-19 virus within a population such that the health systems can tackle and overcome the current pandemic. The simulation results show that when the spreading techniques and early detections of COVID-19 cases have been adopted early, the proposed model ensures a massive of lives saving in the population; however, when the spreading techniques increased, lifesaving is not increased accordingly. This shows that not all the city or the state should go to shut down, but only the main hotspots such as (shopping centers, restaurants, subways, big groceries, and economic centers) in any city that people are using daily.

Acknowledgment

This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920, IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2675, N00014-17-1-2605, N68335-19-C-0359, N00014-19-1-2336, N68335-20-C-0540), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-17-S-0002, W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock, and the Australian

Department of Defense Strategic Policy Grants Program (SPGP) (award number: 2020-106-094). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

References

[1] "Coronavirus disease (COVID-19)." [Online]. Available: https://www.who.int cemergencies/diseases/novel-coronavirus-2019. [Accessed: 29-Aug-2020].

[2] "First and Second Waves of Coronavirus | Johns Hopkins Medicine." [Online]. Available: https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/first-and-second-waves-of-coronavirus. [Accessed: 28-Jul-2020].

[3] U. Qazi, M. Imran, and F. Ofli, "GeoCoV19: A Dataset of Hundreds of Millions of Multilingual COVID-19 Tweets with Location Information," pp. 1–3, 2020.

[4] R. D. Smith, "Responding to global infectious disease outbreaks: Lessons from SARS on the role of risk perception, communication and management," *Soc. Sci. Med.*, vol. 63, no. 12, pp. 3113–3123, Dec. 2006.

[5] A. Zlojutro, D. Rey, and L. Gardner, "A decision-support framework to optimize border control for global outbreak mitigation," *Sci. Rep.*, vol. 9, no. 1, pp. 1–14, Dec. 2019.

[6] M. Mandal, S. Jana, S. K. Nandi, A. Khatua, S. Adak, and T. K. Kar, "A model based study on the dynamics of COVID-19: Prediction and control," *Chaos, Solitons and Fractals*, vol. 136, p. 109889, 2020.

[7] F. Ndaïrou, I. Area, J. J. Nieto, and D. F. M. Torres, "Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan," *Chaos, Solitons and Fractals*, vol. 135, 2020.

[8] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, vol. 27, no. 4, pp. 623–656, 1948.

[9] J. B. Anderson and S. Mohan, Source and Channel Coding: An Algorithmic Approach, First edit. New York: Kluwer Academic, 1991.

[10] D. Tse and V. Pramod, *Fundamentals of wireless communication*, vol. 9780521845. 2005.

[11] M. Alassad, M. N. Hussain, and N. Agarwal, "Finding Fake News Key Spreaders in Complex Social Networks by Using Bi-Level Decomposition Optimization Method," in *International Conference on Modelling and Simulation of Social-Behavioural Phenomena in Creative Societies*, 2019, pp. 41–54.

[12] M. Alassad, M. N. Hussain, and N. Agarwal, "Developing Graph Theoretic Techniques to Identify Amplification and Coordination Activities of Influential Sets of Users," *Accept. Int. Conf. Soc. Comput. Behav. Model. Predict. Behav. Represent. Model. Simul.*, 2020.

[13] F. Şen, R. Wigand, N. Agarwal, S. Tokdemir, and R. Kasprzyk, "Focal structures analysis: identifying influential sets of individuals in a social network," *Soc. Netw. Anal. Min.*, vol. 6, no. 1, p. 17, 2016.

[14] R. Zafarani, M. A. Abbasi, and H. Liu, "Social media mining: An introduction," *Soc. Media Min. An Introd.*, vol. 9781107018, pp. 1–320, 2014.