DOI: http://dx.doi.org/10.26483/ijarcs.v11i5.6649

Volume 11, No. 5, September-October 2020

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

NEIGHBOR SCORE: IMPACT OF NEIGHBORING REGIONS IN COVID-19 OUTBREAK

Dipok Chandra Das
Institute of Information Technology
Noakhali Science and Technology University
Noakhali, Bangladesh

Dipanita Saha
Institute of Information Technology
Noakhali Science and Technology University
Noakhali, Bangladesh

Tahrim Kabir
Institute of Information Technology
Noakhali Science and Technology University
Noakhali, Bangladesh

Abstract:COVID-19 has affected people from all over the globe. This highly contagious disease is spreading from one region to another by people through traveling. The authorities are now trying to identify high-risk and low-risk regions considering each region as isolated. In reality, the regions of a country are connected through buses, trains and, airplanes, etc. and peoples are taking steps to go elsewhere to earn their livelihood. So, preventive and protective measures like lock-down fail to be effective. In this paper, we propose a metric named *Neighbor Score* to determine the actual invaded scenario of a region considering its adjacent i.e. neighbor regions. As a part of this research, an experiment is conducted with different *COVID-19* affected regions of Bangladesh. The experiment has found that the coefficient correlation value of neighboring regions' total *COVID-19* cases is positive (0.54 for 50% regions). We have also ranked and zoned the different regions of Bangladesh applying the *Neighbor Score.*

Keywords:covid-19; outbreak; neighbor; region; network; metric; Bangladesh

I. INTRODUCTION

The world is going through the *COVID-19* pandemic declared by the World Health Origination (WHO) on March 11 [1]. ***Co***rona **VI**rus**D**isease 20**19** (COVID-19) was first confirmed at Wuhan in China on December 2019. The reproduction value, R_0 of *COVID-19* is greater than 2.0 which indicates itself as a highly contagious disease [2], [3]. Till June 22 2020, *COVID-19* has infected 8,860,331 people and caused death of 465,740 over more than 215 countries [4]. People are the primary vector of this disease. *COVID-19* is spreading at a great speed through people from one country to another for advanced facilities of modern globalization i.e. travel comfortableness and ease connection among countries [5]. On February 2020, many countries impose international travel restriction as an effective action to avoid case importation of *COVID-19* from highly infected countries [6]. Alongside, applying lock-down all over the country is a procedure of containment of *COVID-19* from further spread of existing cases. But due to the bad impact on the economy, the authorities are applying partially lock-down. The regions in countries are ranked (high-to-low) and zoned (high, intermediate and low-risk zones) [7], [8]. In ranking and zoning, the information of the neighbor region is not considered. As a result, travel availability such as bus, train, etc. for work or other necessary tasks can spread COVID-19 from higher to its adjacent lower infected regions [9]. The information about neighboring regions' can also help us to understand better the situation of a region and take proper action for that region. In this paper, we are proposing a metric naming *Neighbor Score* to measure a region's *COVID-19* situation considering its own and neighboring regions.

II. LITERATURE REVIEW

There are established mathematical models and metrics in the field of infectious disease like *COVID-19*. The classic epidemic model is Susceptible-Infected-Removed (**SIR**) [10], [11]. The entire population is divided into three states: Susceptible(**S**), Infected (**I**) and Removed (**R**). Every individual of population is susceptible by birth for non-vaccine disease like *COVID-19*. An individual goes from state **S** to **I** and then to **R** with immunity or dead. The popular variant of **SIR** model applied on **COVID-19** is **SEIR** (Susceptible-Exposed-Infected-Recovered) model. Here, **E** refers to a condition that susceptible individual has come contact with infected. **R** state indicates recovered with immunity. **SEIR** is applied to assess the effectiveness of measures taken by the authorities [3]. The other popular variants of **SIR** are **MSEIR** and **SIS** [10], [12]. State **M** refers to temporal immunity through birth of an individual. After end of temporal immunity, that individual becomes susceptible. **SIS** model (Susceptible-Infected-Susceptible) is applied on disease with no immunity.

The works on spread of epidemic disease on network is recognized [12], [13], [14]. The works have found that spread rate of epidemic depends largely on structure of social network. Construction of *contact network* compliments the existing models. In *contact network*, the vertex set is of the people and edge set is of the connection among people. The degree of a vertex refers to occurrence number of that vertex on edge set. The researchers have found that person with higher degree increases the chance of being infected than person with lower degree.

Reproduction number, R_0 is the most popular metric for any infectious disease. The basic reproduction number, R_0 measures the average number of secondary infections by an infected [10], [15]. A disease with R_0 less than 1 will die out but R_0 greater than 1 indicating will spread. The higher value of R_0 refers to higher spread of disease. The researchers have found that R_0 of *COVID-19* is higher than SARS [2], [3].The R_0 of *COVID-19* value varies in different countries because of geographical location, preventive measures, etc. The authorities take actions accordingly for each country like lock-down, partial lock-down, etc. depending on value of that country's R_0. For estimating the reproduction number R_0 early in an epidemic, a Bayesian statistical framework is proposed in [16]. Their studies estimate the temporal patterns R_0 and were applied during the *SARS* pandemic for the early monitoring of the effect of control measures. Another metric named contact number, σ calculates the number of contacts by an infected [10]. σ determines the number of exposed population. The value of σ remains low with quarantine, physical or social distancing, etc.

Ranking is used to rank high to low risk regions. Pluchino*et. al.* use epidemic data, health-care and environmental information like air pollution, housing information etc. to rank the high-risk cities of Italy [7]. Ren *et. al.* try to identify high risk regions in China mega-cities using epidemic data [8].

Zoning concept is used to divide different regions of a country into several zones like red, yellow/orange and green by two interval points, $\delta_{g,y}$ and $\delta_{y,r}$ [17], [18]. Zoning is generally applied with total *COVID-19* cases of regions, $r_i.s$ and its function is described below:

$$z(r_i) = \begin{cases} green & r_i.s < \delta_{g,y} \\ yellow\ or\ orange & \delta_{g,y} \le r_i.s < \delta_{y,r} \\ red & r_i.s \ge \delta_{y,r} \end{cases} \quad (1)$$

The principle metric used in ranking and zoning is total *COVID-19* cases of a region. Sometimes, $r_i.s$ is normalized into number of cases per 100,000 people of region. The neighbor regions cases alongside own region's cases can be more practical in ranking and zoning.

III. NEIGHBOR SCORE

Let there are n number of considered regions and region set, $R = \{r_1, r_2, r_3, \ldots, r_n\}$. The adjacent regions are called neighbor. The neighbor set N is the subset of Cartesian product of $R \times R$. The region and neighbor set construct un-directed weighted graph i.e. *region network*, $G(R, N)$ where vertex set is R and edge set is N. There is an attribute $r_i.s$ representing total number of *COVID-19* cases of region, r_i. The weight function $\omega(r_i, r_j)$ of each neighbor edge r_i to r_j is defined as,

$$\omega(r_i, r_j) = \begin{cases} 0 & if\ (r_i = r_j) \\ \omega_{i,j} & if\ (r_i, r_j) \in N \vee (r_j, r_i) \in N \\ \infty & if\ (r_i, r_j) \notin N \end{cases} \quad (2)$$

There is a transmission function, $\lambda(r_i, r_j)$ for all neighbor edges representing the rate at which a *COVID-19* case is transmitted from r_i to r_j.

$$\lambda(r_i, r_j) = \begin{cases} 1 & if\ (r_i = r_j) \\ \lambda_{i,j} & if\ (r_i, r_j) \in N \vee (r_j, r_i) \in N \\ 0 & if\ (r_i, r_j) \notin N \end{cases} \quad (3)$$

There are two matrices of equal dimension ($n \times n$), net weight, $\delta_{i,j}$ and net transmission, $\mu_{i,j}$. Net weight matrix, $\delta_{i,j}$ compute weight from region, r_i to region, r_j and net transmission matrix $\mu_{i,j}$ compute transmission rate from region r_i to r_j. The *shortest-path*, p for all-pair regions can be found applying appropriate algorithms.

For the shortest-path $p = (r'_0, r'_1, r'_1, \ldots, r'_k)$ from region r_i to r_j, we compute $\delta_{i,j}$ and $\mu_{i,j}$.

$$\delta_{i,j} = \begin{cases} \sum_{l=1}^{k} \omega(r'_{l-1}, r'_l) & if\ p\ exists\ from\ r_i\ to\ r_j \\ \infty & otherwise \end{cases} \quad (4)$$

$$\mu_{i,j} = \begin{cases} \prod_{l=1}^{k} \lambda(r'_{l-1}, r'_l) & if\ p\ exists\ from\ r_i\ to\ r_j \\ 0 & otherwise \end{cases} \quad (5)$$

Now, we define neighbors according to the following (6).

$$Neighbor, N_d(r_i) = \{r_j : r_j \in R \wedge \delta_{i,j} \le d\} \quad (6)$$

According to (6), $N_0(r_i) = \{r_i\}$ as only $\delta_{i,j} = 0$ and $N_1(r_i)$ is the set of regions with net weight less or equal to 1. Now, we compute neighbor score.

Neighbor Score,

$$N_d S(r_i) = \sum_{r_j \in N_d(r_i)} r_j.s \times \mu_{i,j} \quad (7)$$

According to (7), we can compute $N_0 S(r_i)$, $N_1 S(r_i)$ and so on.

IV. EXPERIMENT

To establish the concept of Neighbor Score, there is a research question to answer: **RQ - Does there exist any relationship among neighbor regions in *COVID-19* outbreak?**

Popular evaluation metric, Correlation Coefficient (**CC**) is used to answer **RQ**. ***CC*** measures a degree to which two variables are statistically related. CC is computed using (8).

$$CC(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)\,Var(Y)}} \quad (8)$$

In **(8)**, both X and Y refer to total *COVID-19* cases of a region. ***CC*** value provides a sign. The absolute value determines the strength of the relation and sign indicates whether the relation is positive or negative. In this study, we have selected Bangladesh, a densely populated country of South-Asia. Institute of Epidemiology Disease Control and Research (**IEDCR**) of Bangladesh has reported the first confirmed case of *COVID-19* on March 08 and first death on March 18 [19]. **WHO-Bangladesh** has recorded 115,786 infected cases and 1,502 deaths till June 22, 2020 [20]. Fig.1. shows the complete picture of *COVID-19* in Bangladesh in terms of tests (Fig. 1(a)), confirmed cases (Fig. 1(a)), deaths (Fig. 1(b)) and recoveries (Fig. 1(b)) till June 08.

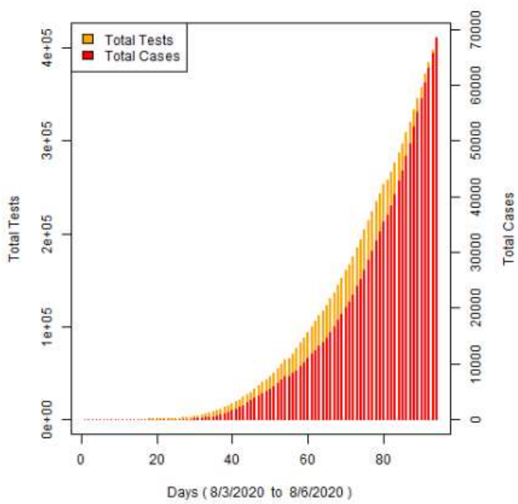

Figure. 1(a). Tests and Cases of COVID-19in Bangladesh

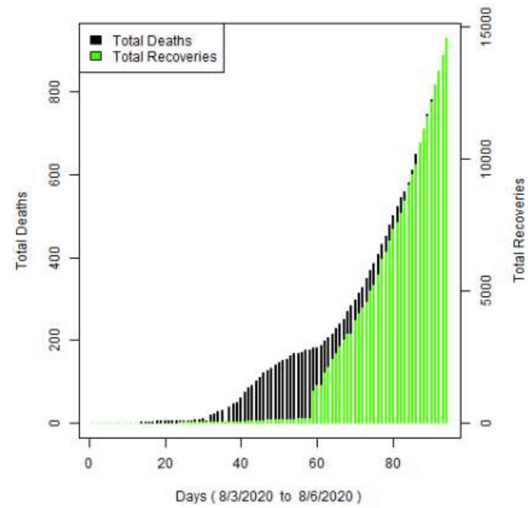

Figure. 1(b). Deaths and Recoveries of COVID-19 in Bangladesh

Figure. 1. COVID-19 Outbreak in Bangladesh

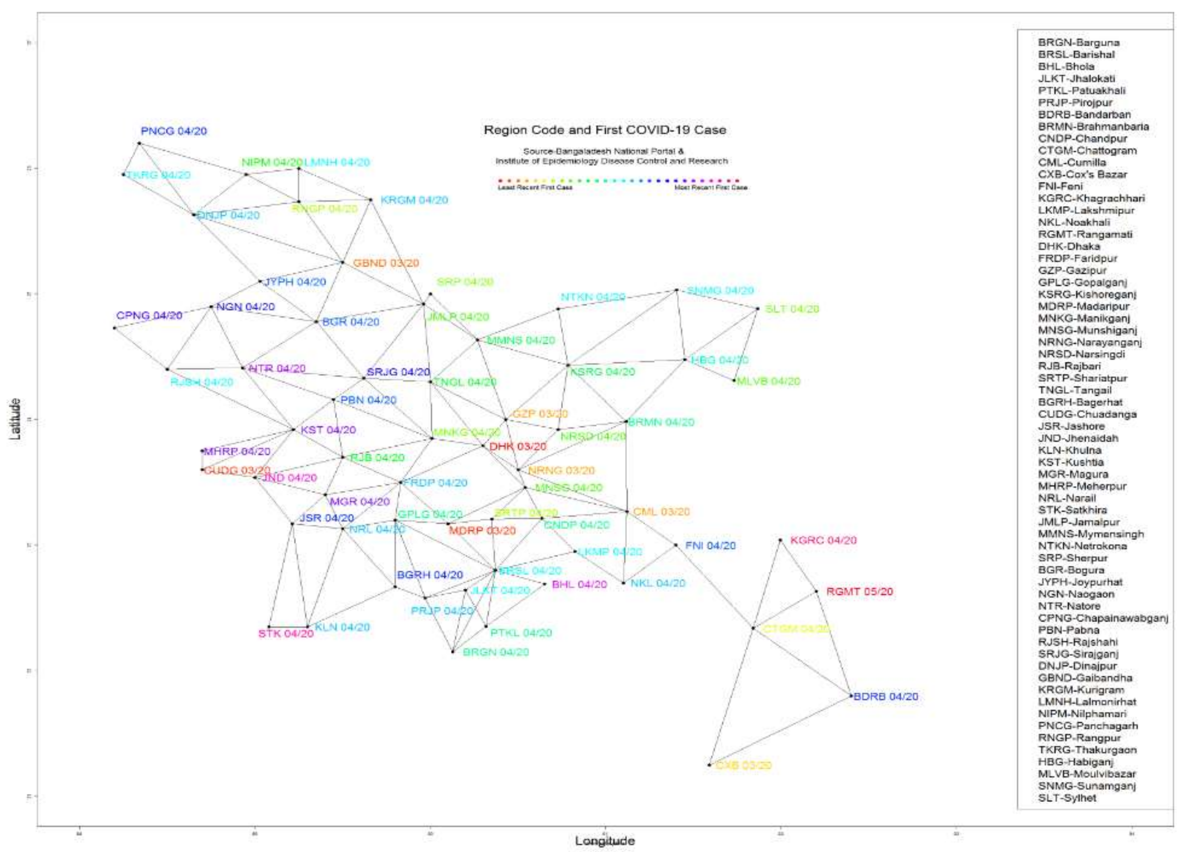

Figure. 2. Region Network of Bangladesh

DOI: http://dx.doi.org/10.26483/ijarcs.v11i5.6649
ISSN No. 0976-5697
Volume 11, No. 5, September-October 2020
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info

A. Data Collection & Experimental Settings

Bangladesh is constituted of sixty four autonomous administrations, named district [21]. The data about districts are collected from website of Bangladesh National Portal [22]. We have considered each district as a region. In this paper, we are experimenting with $\mathbf{N_0S}$ and $\mathbf{N_1S}$. For simplicity, we assume weight of each edge, $\boldsymbol{\omega_{i,j} = \omega = 1.0}$ and transmission rate of each edge, $\boldsymbol{\lambda_{i,j} = \lambda = 1.0}$. The regions (i.e. districts) constructed a graph i.e. region network, ***RN*** (see Fig. 2.). $\mathbf{N_1(r_i)}$ is the union set of $\mathbf{N_0(r_i) = \{r_i\}}$ and its adjacent regions. The degree of a region, $\mathbf{deg(r_i)}$ in ***RN*** is determined by **(9)**.

$$deg(r_i) = \frac{|\{(r_m, r_n) \in N \wedge (r_m = r_i \vee r_n = r_i)\}|}{2} \quad (9)$$

Here, $\mathbf{deg(r_i)}$ determines the number of adjacent regions. The degree of region network is defined by **(10)**.

$$deg(RN) = \frac{\sum_{r_i \in R} deg(r_i)}{|R|} \quad (10)$$

The $\mathbf{deg(RN)}$ of Bangladesh's region network is 4.25. The regions are plotted in the graph according to their latitudes (y-axis) and longitudes (x-axis). The dates of first confirmed cases of regions are also labeled. To answer the question ***RQ***, we have collected region wise *COVID-19* distribution from the daily report of IEDCR [23] dated from April 13, 2020 to May 15, 2020. We explore shortest-path for all pairs of regions. Here, the region network is un-directed equi-weight graph, we apply **Breadth First Seacrh** (BFS) to find out the shortest paths. Then, we compute net weight matrix, $\boldsymbol{\delta_{i,j}}$ and net transmission matrix, $\boldsymbol{\mu_{i,j}}$.

B. Experimental Results

The correlation values are plotted in Fig. 3. The lowest value is positive indicating the relationship between a region and its neighboring regions is positive in *COVID-19* spreading. The correlation values of more than 50% districts are above 0.54. So, research question **RQ** finds out that there exists relationship among neighboring regions in *COVID-19* outbreak. ***Neighbor Score*** can be used in ranking and zoning application. These two use total *COVID-19* cases of a region, $r_i.s$. In those applications, *Neighbor Score* rather than $\mathbf{r_i.s}$ can produce effective output in infectious disease like *COVID-19*. Based on data till June 13 2020, regions of Bangladesh are ranked and zoned applying $\boldsymbol{N_0S}$ and $\boldsymbol{N_1S}$ in Fig. 4. and Fig. 5. respectively. Applying $\mathbf{N_0S}$, the top five high scoring regions are *Dhaka*, *Chittagong*, *Narayanganj*, *Cumilla* and *Munsiganj*. But the top five regions in $\mathbf{N_1S}$ are *Narayanganj*, *Munshiganj*, *Dhaka*, *Gazipur* and *Tangail*. The degrees of these regions are, deg(Dhaka)=5, deg(Chittagong)=5, deg(Narayanganj)=6, deg(Cumilla)=6, deg(Munsiganj)=7, deg(*Gazipur*)=6 and, deg(Tangail)=6 and higher than average degree of the region network (4.25). *Chittagong* has higher number of cases but its adjacent (i.e. $\mathbf{N_1}$) regions have lower cases. *Munshiganj* has highest degree of 7 in region network resulting in placed on top five in both scoring. *Dhaka* and *Narayanganj* have higher number of cases individually. Though *Tangail* is lower-risk region in $\boldsymbol{N_0S}$ but take place top five in $\mathbf{N_1S}$ score due to higher degree and being adjacent of *Dhaka*. Only one region has crossed the 5000 mark in $\boldsymbol{N_0S}$ but there are 10 regions crossed that mark in $\boldsymbol{N_1S}$ because of higher average degree (4.25).

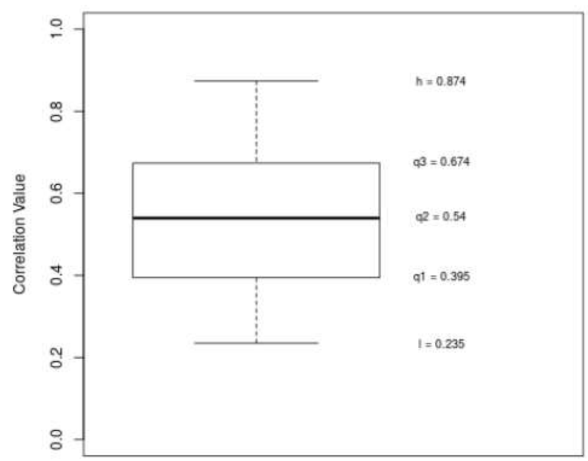

Figure 3. Box-plot of Correlation Coefficient Values

The regions are colored red, yellow and green according to ***'Bangladesh Risk Zone-Based COVID-19 Containment Implementation Strategy/Guide'*** [18]. In (1), $\mathbf{r_i.s}$ is normalized into number of total cases per 100,000 populations. According to [18], we set $\boldsymbol{\delta_{g,y} = 3}$ and $\boldsymbol{\delta_{y,r} = 60}$ per 100,000 people based on. The three regions are colored red, rest sixty-one regions are yellow and two regions are green in $\mathbf{N_0S}$. But 30 regions are red and rest 34 are yellow in $\mathbf{N_1S}$. The number of red zones increases significantly. *Tangail* is one of lower and green region in $\mathbf{N_0S}$. But it becomes a red zone in $\mathbf{N_1S}$ due to higher degree, $\deg(Tangail) = 6$.

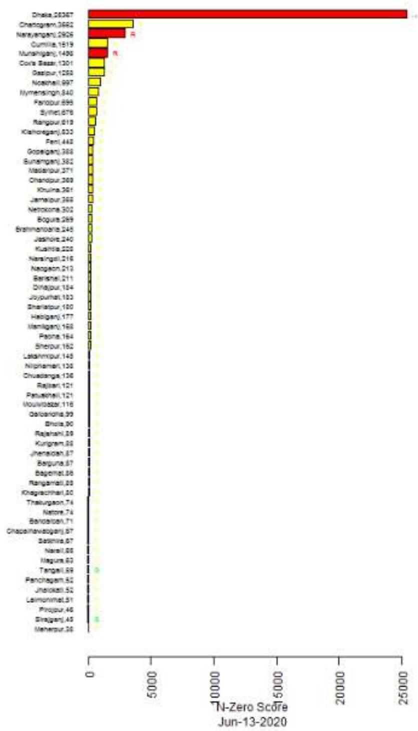

Figure. 4. Ranking and Zoning applying $\mathbf{N_0 S}$

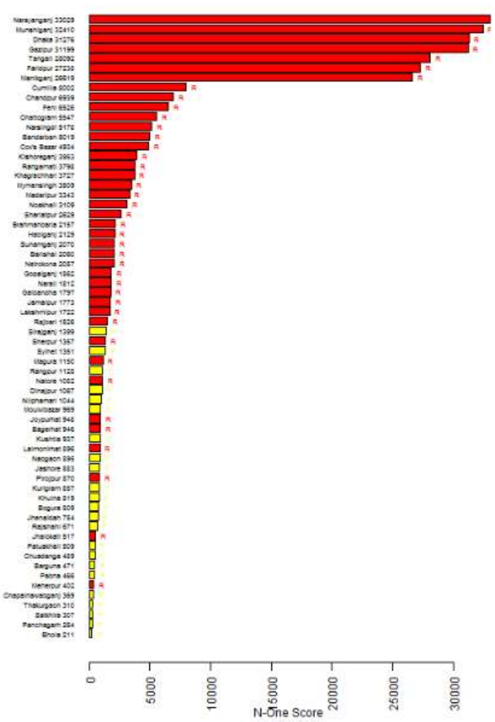

Figure. 5. Ranking and Zoning applying $\mathbf{N_1 S}$

V. CONCLUSION AND FUTURE WORKS

Now, the different regions form a connected region network due to advanced facilities of globalization like bus, train, or air-plane. Due to the high reproduction number of COVID-19, it is spreading from high-risk to adjacent low-risk regions through people. So, the authorities should take measures considering not only own regions but also neighboring regions. Our proposed Neighbor Score metric can capture the situation of a region better and help to take proper measures. In future studies, Neighbor Score will be revised with epidemic (like number of deaths, a number of recoveries, etc.), geographic (like temperature, air quality, etc.), demographic (like gender, age, etc.) and health-care (like a number of hospitals) data of regions.

ACKNOWLEDGMENT

The authors wish to thank all the frontline workers of Bangladesh fighting *COVID-19*.

REFERENCES

[1] World Health Organization, "Coronavirus disease (COVID-19): situation report 51.", 2020.

[2] Ying and Gayle, Albert A and Wilder-Smith, Annelies and Rockl, "The reproductive number of COVID-19 is higher compared to SARS coronavirus," *Journal of travel medicine*, 2020.

[3] Biao and Wang, Xia and Li, Qian and Bragazzi, Nicola Luigi and Tang, Sanyi and Xiao, Yanni and Wu, Jianhong Tang, "Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions," *Journal of clinical medicine*, vol. 9, no. 2, p. 462, 2020.

[4] World Health Organization, "Coronavirus disease (COVID-19): situation report 154.," 2020.

[5] Santiago and Jones, Malcolm K and Marty, Aileen M Mas-Coma, "COVID-19 and globalization," *One Health*, vol. 9, 2020.

[6] Matteo and Davis, Jessica T and Ajelli, Marco and Gioannini, Corrado and Litvinova, Maria and Merler, Stefano and y Piontti, Ana Pastore and Mu, Kunpeng and Rossi, Luca and Sun, Kaiyuan and others Chinazzi, "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Science*, vol. 368, no. 6489, pp. 395-400, 2020.

[7] A and Inturri, G and Rapisarda, A and Biondo, AE and Moli, R Le and Zappala, C and Giuffrida, N and Russo, G and Latora, V Pluchino, "A Novel Methodology for Epidemic Risk Assessment: the case of COVID-19 outbreak in Italy," *arXiv preprint arXiv:2004.02739*, 2020.

[8] Hongyan and Zhao, Lu and Zhang, An and Song, Liuyi and Liao, Yilan and Lu, Weili and Cui, Cheng Ren, "Early forecasting of the potential risk zones of COVID-19 in China's megacities," *Science of The Total Environment*, vol. 729, p. 138995.

[9] Alfonso J and Gallego, Viviana and Escalera-Antezana, Juan Pablo and M, "COVID-19 in Latin America: The implications of the first confirmed case in Brazil," *Travel medicine and infectious disease*, 2020.

[10] Herbert W Hethcote, "The mathematics of infectious diseases," *SIAM review*, vol. 42, no. 4, pp. 599-653, 2000.

[11] Andrew Atkeson, "What will be the economic impact of covid-19 in the us? rough estimates of disease scenarios," National Bureau of Economic Research, 2020.

[12] Jennifer and Ma, Junling and Van den Driessche, P and Willeboordse, Frederick H Lindquist, "Effective degree network disease models," *Journal of mathematical biology*, vol. 62, no. 2, pp. 143-164.

[13] Mark EJ Newman, "Spread of epidemic disease on networks," *Physical review E*, vol. 66, no. 1, p. 016128, 2002.

[14] Romualdo and Vespignani, Alessandro Pastor-Satorras, "Epidemic dynamics in finite size scale-free networks," *Physical Review E*, vol. 65, no. 3, p. 035108, 2002.

[15] Klaus Dietz, "The estimation of the basic reproduction number for infectious diseases," *Statistical methods in medical research*, vol. 2, no. 1, pp. 23-41, 1993.

[16] Simon and Bo, "Real-time estimates in early detection of SARS," *Emerging infectious diseases*, vol. 12, no. 1, p. 110, 2006.

[17] India Ministry of Home Affairs, "MHA Order No. 40-3/2020-DM-I," May 01, 2020.

[18] Bangladesh Ministry of Health and Family Welfare, "Bangladesh Risk Zone-Based COVID-19 Containment Implementation Strategy/Guide," June 09, 2020.

[19] World Health Organization | Bangladesh, "COVID-19, situation report 4.," 2020.

[20] World Health Organization | Bangladesh, "COVID-19, situation report 17," 22 June 2020.

[21] Bangladesh Bureau of Statistics, "Population and Housing Census-2011," 2011.

[22] Bangladesh National Portal. https://bangladesh.gov.bd, [Online; accessed 10-Jul-2020].

[23] Institute of Epidemiology Disease Control and Research. https://iedcr.gov.bd,[Online; accessed 10-Jul-2020].