Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Time dependent correlations between the probability of a node being infected and its centrality measures

Semra Gündüç *, Recep Eryiğit

Ankara University Computer Engineering Department, Ankara, Turkey

A R T I C L E   I N F O

Article history:
Received 15 May 2020
Received in revised form 3 September 2020
Available online 21 October 2020

A B S T R A C T

Pandemics are a growing world-wide threat for all societies. Throughout history, various infectious diseases presented widely spread damage to human life, economic viability and general well-being. The scale of destruction of the most recent pandemic, COVID-19, has yet to be seen. This work aims to introduce intervention methodology for the prevention of global scale spread of infectious diseases. The proposed method combines time-dependent infection spreading data with the social connectivity structure of the society. SIR model simulations provided the dynamic of contamination spread in different sets of network data. Seven centrality measures parameterized the local and global importance of each node in the underlying network. At each time step the calculated values of the correlations between node infection probability and node centrality values are analyzed. Calculations show that correlations increase at the beginning of infection spread and reaches its highest value when spreading starts to become an epidemic. The peak is at the very early stages of the spreading; and with this analysis, it is possible to predict the node infection probability from time-dependent correlations data.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

The dynamics of diffusion, contagious disease spread, rumor, a new idea, or computer viruses exhibit common spreading mechanisms [1–6]. In all of these phenomena, the transmission requires contact interaction within neighboring nodes. All real-world systems, such as social, technological, communication, transportation, are organized in a connectivity structure that results in a complex network of nodes (vertex) connected by links (edges) [7]. Therefore, understanding of the underlying connectivity structure of the system of concern is essential for a realistic model of the contact driven spreading phenomena [8,9]. Not only are such models vital to estimating or understanding the behavior but also to manipulate the system into a preferred state [10]. The effects of connectivity structure on the evolution of the dynamic processes on the network are well studied [11–13]. Power network breakouts, failures in communication networks, crisis in financial networks, and many cascading processes may be related to negligible changes or failures in the connectivity structure of a system [14]. In all these works, detailed knowledge of the network structures is the crucial element on the success or failure in taking precautions to protect society from the spread of disease. A method less dependent on the detailed underlying connectivity will be of great benefit to complex network science.

Some network nodes are more effective than others in accelerating the spreading process [15]. Their strength in spreading disease is related to their local and global positioning in the network. Mathematical formulation of the relation between the network structure and local and global weights of each network node is the centrality measures. There exist

* Corresponding author.
E-mail address: gunduc@ankara.edu.tr (S. Gündüç).

https://doi.org/10.1016/j.physa.2020.125483
0378-4371/© 2020 Elsevier B.V. All rights reserved.
more than 200 centrality measures, with each classifying different aspects of the weight of member nodes according to their importance and inter-relational strength on the network. Despite the complexity of modern systems such as social, technological and communication networks, centrality measures are useful to quantify the local and global positions of the nodes. Each centrality measure aims another aspect of the role of the nodes. The centrality measure, which is related to the number of neighbors, is called degree centrality. Similarly, betweenness, closeness, eigenvector, clustering and PageRank centrality measures are among the most commonly used centrality measures.

Despite abundance of centrality measures, only a limited number of centrality measures are totally independent and are commonly used in the literature [16]. Correlations between centrality measures are indication of degree of dependence among the centrality measures. Moreover, the centrality measure correlations also depend on network topology. Network modularity is the source of this cross-network variations [17] and is also affects dynamic process factors. For this reason, the relation between network centrality measures and the spreading phenomena is far from clear [17].

In controlling the diffusion processes such as an epidemic outbreak, the optimum performance can be achieved only with complete knowledge of the network [12]. There are a considerable number of recent articles which studied the efficient methods of immunization and isolation. In these studies, the observed fact is that, in real-world networks, the hub nodes – the nodes with a high degree of connectivity – play a crucial role in the spreading phenomena [18]. Real-world networks are far more complicated than a simple network. News diffusion on a communication network may affect the other diffusion processes proceeding on other layers. In multi-layer networks, the news spreading on one layer increases awareness, hence can prevent disease from spreading on another [2,4].

Fighting the spread of contagious disease is a race against time. The main tools to slow the spread of an infection are vaccination and quarantine. In most cases, the means of immunization may not be readily available, or time-consuming, which makes large-scale application impossible. Quarantine is an extreme prevention method; it can be efficient only if used selectively and for a limited period. Either or both of these methods on selected nodes decisively control the diffusion. Two questions thus emerge: (i) Is the spread at its initial stages? (ii) which nodes are the most vulnerable? In most cases, even the identification of the spread requires good knowledge of the status of the nodes. The local and global position nodes determine the effect of the spreading process. Centrality measures parameterize the vulnerability of the node during a diffusion process. If a node is vulnerable, it has many connections, remains between large communities, its centrality measures reflect the vulnerability level. Therefore, the knowledge of the correlations between local and global node importance and the spreading dynamics becomes crucial for prevention planning. At the initial stages of the spread, almost all nodes are susceptible. At this stage, there are no correlations between the node infection probability and the node centrality measure values. As the infection spreads, the first nodes to be infected are the ones that have high centrality values. The correlation between node infection probability and high centrality values increases as the infection spread widens. At the final stages, the number of infected nodes is large, and the relative importance of the node position is lost, indicating a lower correlation between the centrality measures and node infection probability. Hence, there exists a time-dependent relation between infection probability and node centrality values. Here, two questions, which centrality measures are more relevant for a given network, and knowing the centrality value of a node can one predict the possible time of being infected.

The present work aims to introduce a systematic scheme to identify the relevant network centrality measures to predict the infection probability of a node given the time interval. The relevance of different centrality measures depends on network typologies This knowledge can be a guide to prevent the spread of a contagious disease at an early stage of an outbreak. The time dependencies of correlation between the node centrality measures and spreading dynamics have not been considered in the literature yet.

In the proposed scheme, the correlations between the node centrality values and the node infection probability at a given time slice are calculated for selected centrality measures. Hence studying the relation between the spreading of disease and centrality measures cannot be abstract from the correlations among the centrality measures [17]. Correlations among the centrality measures and the correlations between node centrality values and node infection probability give evidence for the relevant centrality measures given the network topology. The main aim is to make model predictions on the real-world systems. The existing observations show that there are two different types of artificial networks, Erdős–Renyi [19] and scale free [20] networks, which have the same characteristics as the real-world systems. These network structures have their distinct characteristic connectivity patterns, $P(k)$, which gives the probability of a node having $k$ neighbors. To test the proposed scheme, two artificial networks, namely, Barabási–Albert and Erdős–Renyi are employed. Later, Airline [21] and Facebook [22] network data are used as real-world examples. Susceptible–Infected–Recovered (SIR) model [23] is employed for the dynamics of the spreading of the contagious disease. Barabási–Albert and Erdős–Renyi network models are used as model networks. The observations are tested on real-world networks.

For the preparation of the network connectivity structure and the calculation of the network parameters NetworkX python package [24] is employed. The created network topologies are used as the input for Idlib python library [25] which is used for the SIR model spreading dynamics. The python libraries, pandas [26], and NumPy [27] are used for data analysis and correlation calculations.

The work is organized as follows. Details of the methodology are discussed in the following section. Section 3 is devoted to the results and discussions. Finally, conclusions are presented in the fourth section.
2. The method

SIR epidemic model is based on the scenario in which those susceptible $S$ become infected, $I$ denote infected nodes, $R$ denote nodes recovered. The probabilities $\beta$, gamma $\gamma$ control the number of infected and recovered individuals. A third parameter, $\lambda$ is the ratio of initial infected nodes which determines initially infected individuals. During a dynamic process, the number of individuals does not change. Hence, the total number of individuals,

$$N = S + I + R$$

remains constant. At the initial stages, all individuals except a small number of infected individuals, $I(0)$, are susceptible. Since the individuals are grouped under three categories, such models are also called compartmental models. By using $s = S/N$, $i = I/N$, $r = R/N$ to denote the fraction of the population in each compartment, the SIR model becomes

$$\frac{ds}{dt} = -\beta si$$
$$\frac{di}{dt} = \beta si - \gamma i$$
$$\frac{dr}{dt} = \gamma i. \tag{1}$$

SIR model sets the dynamics of the spreading of the contamination. If the number of infected increases more than the recovered, an epidemic occurs. Hence the epidemic condition:

$$\beta si - \gamma i > 0 \tag{2}$$

The spreading of contamination is a function of three parameters: transmissibility, duration of illness, and the average rate of contact between susceptible and infected individuals. The first two, transmissibility and duration of contamination, are characteristics of the contaminating agent, while the third parameter is directly related to the topology of the interaction network. The reproduction number $R_0$, is

$$R_0 = \tau \tilde{c} d. \tag{3}$$

If $R_0$ has a non-vanishing value, then the spreading becomes an epidemic. In Eq. (3) $\tau$ is the transmissibility, $\tilde{c}$ is the average rate of contact between susceptible and infected individuals, and $d$ is the duration of illness. SIR model assumes a well-mixed population with an equal probability of contacting any susceptible individual. In this model, the basic reproduction number is simply

$$R_0(SIR) = \frac{\beta}{\gamma} \tag{4}$$

Comparing Eqs. (3) and (4), the relation, $\beta = \tau \tilde{c}$ and $d = 1/\gamma$ can be obtained.

The SIR model is an aggregate (compartmental) model where each interacts with all others. The model in which all interact with all others is not realistic for modern societies. An agent-based SIR model with an interaction pattern constitutes a more realistic model of the infection spreading. Complex networks or real-world networks represent the connectivity structure of a society. The links of the determine the node–node interaction pattern. Two variables, the node index and the state $(S, I$ or $R$), represent the position and the state of the node. At time $t$, the individual living at site $i$, is denoted by the variable $X_i(t)$ where $i = 1, \ldots, N$. The node variable $X_i$ can take three values, $S$, $I$ or $R$. The transition probabilities of the changing state of an individual are given by

$$\Pr(X_i(t + \Delta t) = l|X_i(t) = S) = \beta \sum_{j=1}^{NN_i} \delta_{X_j(t),I} \Delta t + o(\Delta t), \tag{5}$$
$$\Pr(X_i(t + \Delta t) = R|X_i(t) = I) = \gamma \Delta t + o(\Delta t). \tag{6}$$

where $\delta_{X_j(t),I}$ is the Kronecker delta. The sum over the nearest neighbors $(NN_i)$, $\sum_{j=1}^{NN_i} \delta_{X_j(t),I}$, counts the number of infected neighbors. Eqs. (5) and (6) are the probabilities of the agent living on the node $i$ being infected (if it is susceptible at time $t$) and recovered (if it is infected at time $t$) at next time step $(t = t + \Delta t)$. On the regular networks as $N \to \infty$ and $\Delta t \to 0$ the probabilistic model gives the same results as expected from the augmented model.

$$\Pr(X_i(t + \Delta t) = l|X_i(t) = S) = \beta \sum_{j=1}^{NN_i} \delta_{X_j(t),I} \Delta t \tag{7}$$
$$\Pr(X_i(t + \Delta t) = R|X_i(t) = I) = \gamma \Delta t \tag{8}$$

The number of neighbors differs according to the connectivity structure of the nodes. If a susceptible node has a large number of neighbors, it is more likely to be infected. Similarly, if an infected node has a large number of neighbors, it is
more likely it may cause someone to be infected, which means the probability of spreading the illness is also high. The connectivity structure depends on the topology of the network. In this work, three different categories are considered: Barabási–Albert, Erdős–Renyi, Real-World networks.

Regardless of the network structure, nodes possess characteristic parameters which uniquely identify its weight and importance in the network. Nodes can be associated with many different measurable quantities. One of the most commonly used is the degree of a node, which determines the number of connected neighboring nodes. Among over 200 centrality measures the quantities

- degree centrality,
- betweenness centrality,
- closeness centrality,
- eigenvector centrality,
- clustering coefficient,
- pagerank centrality

are the network parameters which are used widely in network analysis. All of the network parameters are designed to give special information on the connectivity structure of a complex network. Each is designed to measure a different aspect of the complex relations among the nodes. Despite centrality, measures emphasize a different aspect of complex relations; they are correlated and some of them have strong correlations. These correlations are also reflected in the spreading dynamics. The relation between the speed of the spreading phenomena and the network parameters of the nodes gives a clear indication of the correlations between network parameters and the dynamics of spreading.

The model is based on the comparison between the correlations among the centrality measures and the spreading dynamics on the same network topology. To address the problem of predicting the relationship between the centrality measures and the spreading dynamics on a given network, the model is tested on well-known Erdős–Renyi and Barabási–Albert network topologies. The SIR Model dynamics are simulated on both artificially generated and real-world network structures. The replicas of the networks with an equal average degree are used to follow the spreading dynamics for 50-time steps. In order to eliminate the artifact effects of the seed nodes of the contamination, the contamination started at the randomly chosen seed sites, and the averages are calculated over a large number of statistically independent samples. The correlations between the network parameters and the probability of a node being contaminated are calculated at each time step.

3. Results and discussions

In this section, the SIR model simulation on artificial and real-world network results will be presented. First, two artificial networks, scale-free and random networks, will be used to introduce the hypothesis of the present work. The main aim of the simulation work is to show the time-dependent changes in the correlations between node centrality measures and the probability of a node being infected. For a given network topology, the density of the connections determines the time scale of the spreading. For artificial networks, the average number of connections is a crucial control parameter. Hence, the average number of neighbors per node is taken 3 and 6 for time scale comparisons.

Artificial societies (scale-free and random networks) consist of $N = 10{,}000$ nodes. The same initial set-up, (the same number of nodes, the same average number of neighbors) results in different connectivity patterns — replicas. Replicas have the same characteristics in the average, but exhibit spreading variations. Hence, for each artificial network setting, 50 independent replicas are created for simulations.

For every replica, 500 independent configurations of randomly selected seed spreader sites are created. The time evolution of the spreading (Eq. 7) is followed for 50 time steps. At each time step, the averages are taken over 50 replica and on each replica 500 new runs started from statistically independent initial configurations.

For all network data sets the same SIR model parameters are used, $\beta = 0.01$, $\gamma = 0.005$ and the infection probability, $\lambda = 0.01$. The initial infected individuals (The seed spreaders) are randomly selected.

The time dependent correlations between the network centrality measures and the infection probability of the nodes vary in time. At the beginning of the spread there exist no correlation, since all nodes share the same infection probability until the spread of infection start to grow (initial stages of spreading). At a certain point, topology of the network come into play where a peak in the correlations appear. As the infection spreads to all off the nodes, the correlations start to die off. Both the speed of the spread and the duration of the contamination are effective on the strength of the correlations. The height and the time of the correlation peak is closely related with the topology of the network. In order to observe the relations between the correlations and the topology of the network the hypothesis are tested on the artificial networks.

3.1. Artificial networks

Topological differences between scale-free and random networks create opportunities to compare spreading dynamics on the real-world networks. Hence, the correlations between centrality measures and the spreading phenomena are tested on both Barabási–Albert (scale-free) and Erdős–Renyi (random) networks.
Fig. 1. Correlations among the centrality measures for Barabási–Albert Network with average degrees 6 and 12.

Fig. 2. Correlations between the infection probability and the centrality measures of the nodes. The spread of infection on Barabási–Albert network with $N = 10000$ and average degrees 6 and 12.

3.1.1. Barabási–Albert Networks

Barabási–Albert network is well known for its characteristic power-law degree distribution, which indicates the existence of a small number of nodes with high connectivity (hub nodes), while a large number of nodes with a small number of neighbors. First, the correlations among the centrality measures are calculated to check the dependencies. In calculating correlations among the centrality measures of nodes, 50 statistically independent Barabási–Albert networks are used.

Fig. 1 show the correlations between selected centrality measures. The correlations are obtained as the averages over 50 replicas of Barabási–Albert networks with 6 (Figs. 1(a)) and 12 (1(b)) average connections per site. The comparison of the subfigures, 1(a) and 1(b) indicate that increasing average number of connections also increases the existing correlations. Moreover, pagerank and degree centrality measures are highly correlated, give the same information with the same strength. The clustering centrality measure does not correlate with any other measure, which does not change with the increasing connectivity of the system. Similarly, the closeness centrality exhibits a relatively weak correlation with the other measures. Its correlations increase with increasing connectivity.

Fig. 2 shows the correlations between the node centrality measures and the node infection probability. The node infection probability denotes the probability of being infected if an infection starts to spread at any node in the network. For every independent graph, 500 different runs are performed with statistically independent initial configurations. During 50 time steps, the spread of infection is followed. At every time step, the infected nodes are labeled as susceptible, infected, or recovered. The node infection probabilities are calculated over 500 sample runs at each time step. If a node is infected once, this node is considered as infected for all the time steps.

After calculation of the node infection probabilities, the correlations between infection probabilities and centrality measures are obtained and the averages are taken over all 50 independent graphs.

Correlations among the centrality measures are reflected in the centrality-infection probability correlations. Fig. 2, shows that degree (dgr) and page rank (pgr) follow closely each other within the accuracy of the plot. Eigenvector (eig)
Fig. 3. The spread of infection on Barabási–Albert networks with average degrees 6 and 12. The straight lines indicate the critical slope changes at the initial stages of epidemics.

Fig. 4 show the correlation between different centrality measures using 50 statistically independent graphs. The correlations of centrality measures of Erdős–Rényi networks for 6, Fig. 4(a) and 12, Fig. 4(b) average degree per site indicate that increasing average number of connections also increases the existing correlations. In the random network case, the correlations among the centrality measures are stronger than the scale-free network. Page rank and degree centrality measures are highly correlated as it is for the scale-free networks (Figs. 4 and 1). Clustering centrality measures do not have a correlation with any other measure for both scale-free and random networks. This fact does not change with the increasing connectivity of the system.

The effect of the network topology on the spreading dynamics can be seen in Fig. 5. The most distinct feature of the Erdős–Renyi network results is the observed smooth rise in the correlations between the node centrality measures and the node infection probability (Fig. 2). The peak, seen in the Barabási–Albert network case is lost. Hence the identification and Betweenness (btw) centrality measures follow a similar pattern with the page rank and degree centralities but they remain slightly lower as expected (Fig. 1). Remaining two centrality measures, closeness (clo) and clustering (cls), behave independently from the other.

Fig. 3 show the percentage changes of susceptible, infected and recovered populations for 50 time steps. If Figs. 2 and 3 are considered together, the relation between the peaks and the spreading dynamics can be better interpreted. Both Figs. 3 and 2 contain two subfigures for 6 and 12 average degree respectively. Increased connectivity increases the speed of the infection spread. Hence, from Figs. 3(a) and 3(b) it can be seen that between the average degrees 6 and 12 the speed off the spread of infection approximately doubles. A characteristic feature of SIR-like epidemics, at the initial stages, the slope of the increase in the number of infected individuals is small up to a critical percentage. After the initial stage, a new phase with a higher slope starts which may lead to an epidemic situation. The correlations between the infection probability and centrality measures reflect the observed initial stages of the contamination spread. Correlations start to grow at the initial stage where the slope of the infection spread is low, reach to a peak at the point where the slope changes. The position of the peak is the position of the point where the slope changes. The slopes are calculated by fitting the infected population data for both average degrees 6 and 12. As the infection spreads, after the peak of the correlations, a large number of nodes have the same probability of being infected, hence the correlations start to decrease.

The position of the peak gives an indication of the timing of any planned action. The peak position of the correlations is much earlier than the peak of the infection spread. At the peak, the spread can be controlled by pointing the nodes with the highest centrality values. Immunizations, isolation, or quarantine are most useful before or at the peak of the correlations. The degree and page rank centralities are the decisive factors for the control of infection spreading (Fig. 2).

The role of topology on the spreading phenomena can be better understood by comparing the results of scale-free and random networks. In the following subsection spread of SIR type infection on Erdős–Rényi networks will be subject to discussion.

3.1.2. Erdős–Rényi Networks

Erdős–Rényi random networks are one of the oldest and best-studied models of node connectivity. However, its resemblance to the real-world networks such as the Internet, social networks or biological networks is limited since it lacks network clustering or transitivity, and its Poisson degree distribution does not overlap with the degree distributions of the real-world networks. Despite that, the real-world networks are mostly scale-free networks, depending on the network size and the relation among the nodes, some parts or the whole network may exhibit random connectivity.

Fig. 4 show the correlation between different centrality measures using 50 statistically independent graphs. The correlations of centrality measures of Erdős–Rényi networks for 6, Fig. 4(a) and 12, Fig. 4(b) average degree per site indicate that increasing average number of connections also increases the existing correlations. In the random network case, the correlations among the centrality measures are stronger than the scale-free network. Page rank and degree centrality measures are highly correlated as it is for the scale-free networks (Figs. 4 and 1). Clustering centrality measures do not have a correlation with any other measure for both scale-free and random networks. This fact does not change with the increasing connectivity of the system.

The effect of the network topology on the spreading dynamics can be seen in Fig. 5. The most distinct feature of the Erdős–Rényi network results is the observed smooth rise in the correlations between the node centrality measures and the node infection probability (Fig. 2). The peak, seen in the Barabási–Albert network case is lost. Hence the identification
of the infection spread by observing the peak in the early stages of the infection spread (Fig. 6) cannot be easily done. Moreover, apart from the clustering centrality measure, all centrality measures follow closely with each other.

Figs. 5 and 6 contain information on the affect of the average degree, 6 and 12 respectively. Increased connectivity increases the speed of the infection spread. Hence, it can be seen that the speed of the spread of infection approximately doubles as the average connectivity increases from 6 to 12. Comparisons between the Figs. 3(a) with 6(a) and 3(b) with 6(b) that the influencer (hub) nodes of the scale free networks play an important role in increasing the speed of the spread.

3.2. Real-world networks

To compare scale-free and random networks with the real-world network data sets Airline Data [21] (3425 nodes (vertex) and 19257 links (edges)) and Facebook ego network [22] (4039 nodes (vertex) and 88234 links (edges)) are used as test ground.

Fig. 7 show the Airline and Facebook network diagrams. Despite the fact that both networks possess power-law degree distribution, Airline Network is a uni-center network while Facebook has more than one center. In the Facebook network case the local centers of dense connectivity [28,29] are sparsely connected with the remaining centers. Figs. 7(a) and 7(b) show the node distributions and connection of the Airline and Facebook networks.

The topological differences in the connectivity structures between the airline and the Facebook networks are reflected in the correlations among the centrality measures. Fig. 8 shows the correlations between centrality measures of airline network (Fig. 8(a)) and Facebook Network (Fig. 8(b)). The correlations among the airline network centrality measure closely follow the Barabási–Albert network as expected. For the airline network page rank, eigenvalue, and degree centralities are closely related. Similarly, betweenness also shows a correlation with page rank and degree centralities.
Fig. 6. The spread of infection on Erdős–Renyi networks with average degrees 6 and 12. The straight lines indicate the critical slope changes at the initial stages of epidemics.

Fig. 7. Distinct network structures of Airline and Facebook networks.

Clustering centrality shows no correlation with any other centrality measures. For the Facebook case, degree and betweenness centralities correlate with the page rank centrality. Apart from these correlations non of the centrality measure exhibit significant correlation.

The effect of the correlations between the centrality measures can be seen on the centrality and the infection-spread probability correlations. Fig. 9 show the time-dependent changes in the centrality and the infection spread correlations and the spreading dynamics in the airline network. As one can see, highly correlated centrality measures exhibit similar correlations (Fig. 9(a)). The curves, corresponding to the highly correlated centrality measures (degree, page rank, eigenvalue centralities) follow each other closely. Betweenness, have the same shape, the peak position but the height of the peak, and in general, the whole curve is lower than the group of highly correlated three centrality measures. All these four centrality measures resemble closely the behavior of the Barabási–Albert case (Fig. 2). The closeness centrality measure curve has no peak, similar to the case of centrality measures of random networks, increases smoothly, and reaches a saturation level. Finally, the behaviors of the clustering centrality correlations very low and carry no significant information. The peak value and the correlation curves correspond to the point where the infection spreading slope changes. Fig. 9(b) show the time-dependent changes in susceptible, infected, and recovered individuals. The initial slow-spreading regime remains only for a very short period. After the initial stage, the slope of the curve changes. During the initial spreading phase the correlations between the probability of being infected and the node centrality measures show a strong correlation. After the slope change, these strong correlations become weaker.

The Facebook network, due to its multi-centered connectivity structure, exhibits no strong correlations among the centrality measures (Fig. 8(b)). Hence, Fig. 10(a) show that the correlations between the node centrality measures and the infection probability curves separated from each other. However, degree and eigenvector centralities have similar shapes.
Fig. 8. Correlations among the centrality measures of Airline and Facebook networks.

Fig. 9. Comparative study of correlations between the infection probability and the centrality measures and the spread of infection on Airline network.

4. Conclusions

Advanced technologies increased the quality and standards of life. One of the most significant impacts of the new technologies is on the pharmaceutical and medical research which opened the possibilities for a better and longer lifespan for large populations living in the developed and also developing countries. Another impact of technological advancement is on communication and transportation. All of the goods are available globally, while the individuals travel to see the natural, historical sites and exotic lifestyles. On the communication side, worldwide information networks made it possible to reach information globally. Despite advanced technologies used for vaccines, medicine, and medical equipment, the spreading of contagious diseases is one of the problems. Starting with the Spanish Flu, many virus-based diseases cost a great many lives. Today, COVID-19 is a striking example of a new type of virus infection. It is seen in all of the countries and peak positions. The peak positions also in accord with the slope change point of the infection spread (Fig. 10(b)). The curves corresponding to page rank and betweenness centrality measures have similar behavior, but the peak positions and the peak heights indicate little correlation with the infection spreading phenomena.

At the beginning of the contamination spread, separated centers of high activity (Fig. 10) act like small communities with a high degree of connectivity, which causes the fast spread of infection. The spread of infection and the correlations between the node centrality values and the infection probability are presented by Fig. 10. Comparing with the previous cases, relatively fast initial contamination spread (Fig. 10(b)) is reflected to the peak positions of the centrality correlation curves (Fig. 10(a)).
globally. It has no vaccine or a specific method of treatment. A unique tool to fight with this highly contagious and deadly virus is isolation or quarantine. It is almost impossible to impose long term personal isolation globally, reducing travel, identifying and isolating the most vulnerable groups or individuals, and the possible super spreaders seemingly the most effective method for fighting the spread. At this point, technology comes into play. The advanced communication technologies give a clear indication for the connectivity structure of the nodes: interactions and movements of the individuals in the societies. Hence, the knowledge of the connectivity structure of society is a tool for disease prevention attempts.

In this work, the observed facts are:

- The correlations among the centrality measures reflect the topological structure of the network.
- On the complex networks, the contamination starts from the seed sites. The majority (particularly for BA and real-world networks) of these seed sites are nodes with low connectivity. When the spreading starts from such a relatively isolated site, contamination shows a slow pace until it reaches the highly connected nodes. After this initial transient regime, the system enters a second phase that has a higher slope. Hence, two distinct phases can be linearized separately (Figs. 3, 6, 9, and 10). Hence, the infectious disease spreading curve exhibits a short-flat initial stage (low slope) followed by a relatively faster spreading regime (higher slope).
- There exists a close, time-dependent relation between the node infection probability and the node centrality values. The correlations between the centrality measures and the node infection probability increase until the end of the flat initial stage, reach its peak at the crossing point of the fitted lines indicating two separate phases.
- The correlation behaves differently after the initial-stage for scale-free and random networks.

1. For scale-free networks, the increase in the correlations reaches a peak value at the point where the initial slow-spreading regime leaves its place to a fast-spreading regime.
2. Random networks show a different behavior, the correlation between the centrality measures and node infection probability increases in the initial slow-spreading stage, and it levels off at the time where the spreading speed increases.

- The strength of the correlation between node infection probability and the node centrality measures depends on the centrality measure and the topology of the network.
- The selected real-world networks, namely, airline and Facebook networks, both have power-law degree distribution, but exhibit different centrality-infection probability correlation patterns.
- The difference is due to the connectivity structure of the networks. Airline networks have uni-centered structures, while in the Facebook case, nodes are centered around more than one loosely connected large communities.
- Among the studied centrality measures, degree and pagerank provide the most reliable information.

The world has experienced various influenza outbreaks in the recent past. In light of the observed facts, the existing infection spread data can be combined with the connectivity information of the individuals (nodes) for the disease prevention planning. For efficient planning, the network structure of the society must be sketched out, the centrality measures of each node must be calculated, correlations with the existing infection spread data, and centrality measures can be used to identify the society specific prevention planning.

Fig. 10. Comparative study of correlations between the infection probability and the centrality measures and the spread of infection on Facebook network.
CRediT authorship contribution statement

Semra Gündüç: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Simulation. Recep Eryiğit: Data curation, Writing - original draft, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Q. Wu, X. Fu, M. Small, X.J. Xu, The impact of awareness on epidemic spreading in networks, Chaos (2012) 22.
[2] S. Gunduc, The effects of diffusion of information on epidemic spread - A multilayer approach, Acta Phys. Polon. B 50 (2019) 179.
[3] Y. Zhang, Y. Su, L. Weigang, H. Liu, Interacting model of rumor propagation and behavior spreading in multiplex networks chaos, Solut. Fractals 121 (2019) 168–177.
[4] S. Gunduc, A study on the effects of diffusion of information on epidemic spread, Int. J. Model. Simul., Sci. Comput. 10 (2019) 1950015, (14 pages).
[5] L. Yang, X. Yang, J. Liu, Q. Zhu, C. Gan, Epidemics of computer viruses: A complex-network approach, Appl. Math. Comput. 219 (2013) 8705–8717.
[6] R.P. Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Phys. Rev. Lett. 86 (2001) 14, 3200-3203.
[7] M.C. Gonzalez, A.-L. Barabási, Complex networks - from data to models, Nat. Phys. 3 (2007) 224–225.
[8] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Phys. Rep. Vol. 424 (45) (2006) 175–308.
[9] M. Opsuiko, J. Ruhlman, Impact of the network structure on the SIR model spreading phenomena in online networks, in: Conference: The Eighth International Multi-Conference on Computing in the Global Information Technology ICCGI, Nice, France, 2013.
[10] J. Liua Q. Xiongbc, W. Shia, X. Shia, K. Wang, Physica A 452 (2016) 209–219.
[11] F.A. Rodrigues, T. Peron, C. Connaughton, J. Kurths, Y. Moreno, A machine learning approach to predicting dynamical observables from network structure, 2019.
[12] H. Cherifi, G. Palla, B.K. Szymanski, X. Lu, On community structure in complex networks: challenges and opportunities, Appl. Netw. Sci. 4 (2019) 1–17.
[13] K. Li, H. Zhang, G. Zhu, M. Small, X. Fu, Suboptimal control and targeted constant control for semi-random epidemic networks, IEEE Trans. Syst., Man, Cybern.: Syst. (2019).
[14] Y.C. Lai, A.E. Motter, T. Nishikawa, Attacks and cascades in complex networks, Lect. Notes Phys. 650 (2004) 299–310.
[15] J. Bae, S. Kim, Identifying and ranking influential spreaders in complex networks by neighborhood coreness, Physica A 395 (2014) 549–559.
[16] M. Jalihi, A. Salehzadeh-Yazdi, Y. Asgari, S. Arab, M. Yaghmaie, A. Ghavamzadeh, et al., CentiServer: A comprehensive resource, web-based application and R package for centrality analysis, PLoS One 10 (2015) 1–8.
[17] S. Oldham, B. Fulcher, L. Parkes, A. Arnatkevic, C. Suo, A. Fornito, Consistency and differences between centrality measures across distinct classes of networks, PLoS One 14 (2019) 7.
[18] G.F. De Arruda, A.L. Barbieri, P.M. Rodrigues, F.A. Rodrigues, Y. Moreno, L. da Fontoura Costa, Phys. Rev. E 90 (2014) 032812.
[19] P. Erdős, A. Rényi, On random graphs, in: I Publicationes Mathematicae, Vol. 6, 1959, pp. 290–297.
[20] J.P. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, A. Barabasi, Structure and tie strengths in mobile communication networks, Proc. Nat. Acad. Sci. 104 (2007) 7332–7336.
[21] http://openflight.org.
[22] J. McAuley, J. Leskovec, Learning to discover social circles in ego networks, NIPS (2012).
[23] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 115 (1927) 700–721.
[24] Aric A. Hagberg, Daniel A. Schult, Pieter J. Swart, Exploring network structure, dynamics, and function using networkx, in: Proceedings of the 7th Python in Science Conference, SciPy, 2008.
[25] G. Rossetti, L. Milli, S. Rinzivillo, A. Sirbu, D. Pedreschi, F. Giannotti, Ndlib: a python library to model and analyze diffusion processes over complex networks, J. Data Sci. Anal. (2017).
[26] W. McKinney, Data structures for statistical computing in python, in: Proceedings of the 9th Python in Science Conference, 2010, pp. 51-56.
[27] T.E. Oliphant, A Guide to NumPy, Trelgol Publishing USA, 2009.
[28] M. Girvan, M.E.J. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. USA. 99 (2020) 7821–7826.
[29] M.E.J. Newman, Networks: An Introduction, Oxford University Press, Oxford UK, 2010.