IDENTIFYING SUPERSPREADERS FOR EPIDEMICS USING R0-ADJUSTED NETWORK CENTRALITY

Taesik Lee
Hyun-Rok Lee
Kyosang Hwang

Industrial & Systems Engineering
Korea Advanced Institute of Science and Technology
Daejeon 157-032, REPUBLIC OF KOREA

ABSTRACT

Identifying "superspreaders" in a network is a key problem to designing an effective mitigation strategy against a spread of an epidemic disease. Superspreaders are a set of nodes that play a hub role in a disease spread network, and classical network centrality measures are often used to identify such hubs. In this research, we test a hypothesis that a node's intrinsic property plays a role in the dynamics of disease spreading in a network. Specifically, we test whether spreading of an epidemic disease is affected by a node's property of being an amplifier or attenuator. Using GEM (Global Epidemic Model), we conducted experiments for epidemic spreading on a hypothetical, global network of 155 cities. We find that node's intrinsic property plays a significant role in disease spreading dynamics. Based on these findings we propose a new metric, R0-adjusted centrality.

1 INTRODUCTION

Designing an effective intervention strategy is a critical problem for mitigating the spread of infectious diseases. Among many intervention strategies, targeted vaccination strategies are recognized as the most efficient strategy(Holme et al. 2002; Cohen et al. 2003; Holme 2004; Chen et al. 2008; Schneider et al. 2011). Because there is limited supply and time for administering vaccine, distributing vaccines to the most effective targets is important. A central problem in designing a targeted vaccination strategy is how to identify those targets that, when immunized, would generate the maximum mitigation effect. These targets are often referred to as a superspreader. This problem of identifying superspreaders is an instance of a key player problem in a social network (Borgatti 2006). Use of network centrality measures is among the prominent approaches in a key player problem, and much research on targeted vaccination uses network centrality measures to identify superspreaders (Colizza et al. 2007; Colizza et al. 2006; Colizza and Vespignani 2007; Kitsak et al. 2010; Freeman 1979; Friedkin 1991; Sikic et al. 2011).

In the field of social network analysis, the idea of centrality has been widely recognized and extensively studied since the late 1940s. Particularly well-known measures of centrality include degree centrality, closeness centrality, betweenness centrality (Freeman 1979) and eigenvector centrality (Bonacich 1987). These centrality measures are typically defined for a binary network, where a link between nodes either exists or does not exist. As a natural extension, there have been a number of proposals to extend the Freeman's centrality measures to incorporate weights of links (Barrat et al. 2004; Brandes 2001; Newman 2001; Opsahl et al. 2010). In the context of social network, these weights often indicate a tie strength between the two nodes, such as frequency of communications.

Spread of pandemic disease shares common aspects of a generic flow behavior on a network, and as such, research on social network's centrality analysis sheds a light on the study of pandemic disease spread.

Thus, it is not surprising that there is a large volume of research that identify *central* nodes in a disease spreading network by using network centrality measures. One of the important considerations in designing a study for disease spread on a network is to ensure the measures used in the study appropriately capture the dynamics of the spreading process.

In the context of designing targeted immunization strategies, apparently there are few research works that use centrality measures in weighted networks. As recognized by Opsahl, Agneessens, and Skvoretz (2010), *"by dichotomizing the network, much of the information contained in a weighted network's datasets is lost"*. This could lead to the identification of suboptimal set of target nodes. While the practical feasibility of measuring tie strength for a contact network of individuals is arguable, there exist some networks that we can reasonably model as a weighted network. A global air transportation network linking cities around the world is one such example. Evidently, an edge linking two nodes, Washington D.C. and Chicago, has a significantly different tie strength - number of people traveling directly between the two cities - from an edge between Washington D.C. and Dayton in Ohio.

Another factor that may affect the outcome of analysis is the use of centrality measures that are not suitable for the flow characteristics of infectious disease spreading. Borgatti (2005) argues that the effectiveness of centrality measures is determined by the inherent assumptions made in the definition of each measure. These assumptions can be characterized along two dimensions - flow trajectory and transmission mechanism. Table 1 shows a typology of flow processes on a network (Borgatti 2005). According to his typology, spreading of an infectious disease is a serial duplication process on paths, and none of the four well-known centrality measures is appropriate for identifying vaccination targets.

Table 1: Typology of flow processes with relevant centrality measures shown in parenthesis [adapted from Borgatti (2005)]

	Parallel duplication	Serial duplication	Transfer
Geodesics	No process (None)	Mitotic reproduction (Freeman closeness)	Package delivery (Freeman closeness, Freeman betweenness)
Paths	Internet name-server (Freeman closeness, Freeman degree)	Viral infection (None)	Mooch (None)
Trails	E-mail broadcast (Freeman closeness, Freeman degree)	Gossip (None)	Used goods (None)
Walks	Attitude influencing (Freeman closeness, Freeman degree, Bonacich eigenvector)	Emotional support (None)	Money exchange (None)

Finally, we notice that the four centrality measures do not consider a node's intrinsic property in transferring diseases. Each node in a global air transportation network represents a city, and each city has different degree of vulnerability to infectious disease spread. In some cities, influx of a new infection quickly evolves into an epidemic, while in others, it vanishes without causing much trouble. Thus, a node in this network does not simply transfer the disease to its neighboring cities, but in doing so, it may attenuate or amplify its intensity. Given an influx of an epidemic disease, some cities shows higher "disease prevalence" than others - i.e., some cities suffer from higher number of infected population than others. We hypothesize that capturing these aspects of a node in a centrality measure will improve its effectiveness in identifying superspreaders on a disease spreading network.

Since disease prevalence is a time-varying quantity, thus difficult to capture *a priori* when computing a centrality measure, we may consider *likelihood* or *susceptibility* of a city with respect to disease spreading. It seems that a reasonable candidate parameter to capture such likelihood/susceptibility is the basic reproductive number, $R0$. When a novel disease is introduced to a community, how fast it will propagate in the community and how prevalent it will be varies significantly for different communities. Using the classic *SIR* model, they are determined by the basic reproductive number, $R0$, which is a function of infectivity α, contact rate β and recovery rate γ. SIR model is a type of an epidemic model where the spread of an infectious disease is modeled by a coupled set of ordinary differential equations. Disease spreading is modeled as a diffusion process from susceptible (S) to infectious (I) to recovered (R) population subgroup. Even if we assume α and γ are largely invariable across different communities as being a disease property, β is certainly a community-specific quantity. For SARS in 2003 for an example, the estimated $R0$ for Hongkong is 3.1~4.2, while for Vietnam, it is 1.8~3.1 (Wallinga and Teunis 2004).

In this paper, we first examine a hypothesis that a node's intrinsic property plays a role in disease spreading dynamics of a network. Specifically, we test whether spreading of an epidemic disease is affected by a node's property of being an amplifier or attenuator. To test this hypothesis, we construct a hypothetical network of 155 cities and examine two scenarios by varying $R0$ values for the cities. In one scenario, cities with high network centrality are assigned high $R0$ values, making those cities a topologically-central node as well as an amplifier. Cities with low network centrality are assigned low $R0$ values. In the other scenario, we use the opposite setting: cities with high network centrality are assigned low $R0$ values, and vice versa. We find that for a given network topology, the two scenarios exhibit a significant difference in a peak time and peak volume for disease spreading. Based on this finding, we propose a new metric, $R0$-adjusted centrality, as a means to rank order and identify superspreader nodes in a disease spreading network. The proposed $R0$-adjusted centrality incorporates $R0$ value of a node to the existing network centrality measure to quantify a node's importance in two aspects - network topology and amplifying/attenuating the intensity of disease spreading.

2 CENTRALITY MEASURES

In this section, we first present four well-known centrality measures - degree, closeness, and betweenness centrality by Freeman (1979), and eigenvector centrality by Bonacich (1987). Then, we briefly discuss the extensions of the four centrality measures for a weighted network.

2.1 Degree, Closeness, Betweenness and Eigenvector Centrality for a Binary Network

Degree centrality of a node i is probably the most intuitive centrality measure. It is defined as the number of nodes that the node i is connected to. Formally, it can be written as

$$C_D(i) = \sum_{j=1}^{n} x_{ij}, \tag{1}$$

where x_{ij} is 1 if node i and j are connected. Degree centrality can be thought of as a measure of immediate effects by a node when an infections disease is spreading on a network. A node with high degree centrality can play a role of a key player for certain type of flow processes such as a parallel duplication process (Borgatti 2005):

Closeness centrality measures how *close* a node is to all other nodes in a network. It is defined as the inverse of the sum of the length of the shortest paths from a node to all others.

$$C_C(i) = \left(\sum_{j=1}^{n} d(i,j) \right)^{-1}, \tag{2}$$

where $d(i,j)$ is the length of the shortest path linking node i and j. For a binary network, $d(i,j)$ is the minimum number of links connecting the two nodes. One interpretation of a node's closeness centrality is that it measures the expected time for an interaction that node i makes with other nodes. It should be noted that closeness centrality implicitly assumes that *the intermediary nodes increases the cost of an interaction* (Opsahl, Agneessens, and Skvoretz 2010) and thus what flows on a network follows shortest paths (Borgatti 2005). In fact, Borgatti (2005) argues that using closeness centrality to identify central nodes may be inappropriate if the underlying flow characteristics do not conform to the shortest path assumption.

Betweenness centrality is also based on the notion of shortest paths on a network. Freeman (1979)'s definition of betweenness centrality can be written as

$$C_B(i) = \sum_j \sum_k \frac{g_{jk}(i)}{g_{jk}}, \tag{3}$$

where g_{jk} it the number of shortest paths between two nodes j and k, and $g_{jk}(i)$ is the number of these paths that contain node i. Conventional interpretation of betweenness centrality is that it measures how much *control* a node has over interactions/flows on a network and if neutralized, it significantly hampers the flow. Again, Borgatti (2005) points out that betweenness centrality is also based on the shortest path assumption and that it further assumes what flows on a network is indivisible.

Bonacich (1972) proposed a centrality measure based on a notion that *a node's centrality is its summed connections to others, weighted by their centralities.* This notional definition can be written as

$$\lambda \boldsymbol{e} = \boldsymbol{A}\boldsymbol{e}, \tag{4}$$

where $\mathbf{A}$ is the adjacency matrix of a graph representing a network. Thus, $\mathbf{e}$ is an eigenvector of the adjacency matrix $\mathbf{A}$, and λ is its associated eigenvalue. Compared to degree centrality, which measures a node's immediate effects, eigenvector centrality measures a long-term effects including both direct and indirect effects (Borgatti 2005). Like degree centrality, it can be a suitable measure of centrality when the underlying flow characteristics follow a parallel duplication process.

2.2 Generalized Centrality for a Weighted Network

Centrality definitions presented above are defined for a binary network, where a relationship between two nodes is either connected or not connected. Naturally, these measures have been extended to a weighted network. For degree centrality, binary variable x_{ij} in (1) is simply replaced with the weights of each link, w_{ij} (Barrat et al. 2004):

$$C_D^w(i) = \sum_{j=1}^{n} w_{ij}. \tag{5}$$

Eigenvector centrality is extended along the similar line of argument: instead of a binary adjacency matrix $\mathbf{A}$, it uses the adjacency matrix whose entries represent the weights on the links (Newman 2004). (4) can be rewritten as:

$$e_i = \lambda^{-1} \sum_j A_{ij}^w e_j, \tag{6}$$

where A_{ij}^w is a weight for a link connecting node i and j.

Since closeness centrality and betweenness centrality are based on the shortest paths, their extension to a weighted network comes from the shortest path in a weighted network. A well-known algorithm to find the shortest path in a weighted network, where weight on a link represents cost, was proposed by Dijkstra (1959). In Newman (2001) and Brandes (2001), the inverse of the weight for each link is defined as the distance l_{ij} between two nodes: $l_{ij} = 1/w_{ij}$. Use of the inverse makes sense for many weighted networks

because weights often indicate a tie strength, e.g. frequency of contacts. Thus, the higher a weight of a link is, the *closer* the two nodes are. With this, closeness and betweenness centrality for a weighted network is computed from (2) and (3) with $d(i, j)$ and g_{jk} replaced by $d(i, j)^w$ and g_{jk}^w.

3   EFFECTS OF NODE PROPERTY ON DISEASE SPREADING IN A NETWORK

This paper addresses the following key hypothesis: an epidemic's spreading on a network is affected not only by the network's topological structure but also by the nodes' property. In the context of epidemic spreading on a network, a relevant node property is whether a node amplifies or attenuates the flow intensity of a disease. For example, a city in a global air transport network can play a role of epidemic amplifier or attenuator. We conjecture that $R0$ level of a city determines the type of a role that a city plays during an epidemic spreading.

When there is an influx of epidemic disease cases from other cities, a city with high $R0$ will experience quick spreading of the disease within the city. Assuming no travel restriction for the infected individuals, this implies a higher number of infected individuals traveling to other cities. In other words, a city with high $R0$ augments the intensity of disease flow on a network, and broadcasts the disease to a larger number of neighboring cities at a higher intensity. On the other hand, in a city with low $R0$, the degree of augmentation will be less. If $R0$ of a city is less than 1, the magnitude of a subsequent broadcast will not be greater than the amount of influx. Thus, the flow intensity is attenuated as an infectious disease transmits through a low $R0$ city. This leads us to a conjecture that, even when multiple nodes have the same structural centrality, their overall impact in disease spreading on a network is likely to be different.

In this section, we present results from an experiment on a hypothetical network of 155 cities. Results from the experiment verifies that a node property - $R0$ of each city - does indeed affect a epidemic's spreading on a network.

3.1 Experiment Using Global Epidemic Model

To simulate an epidemic disease spreading on a network, we use the Global Epidemic Model (GEM). GEM is a publicly available simulation model developed by a group of researchers participating in MIDAS (Models of Infectious Disease Agent Study)(Epstein et al. 2007). It models population mix among 155 cities around the world according to the real airline travel data(Guimera et al. 2005). In GEM, an intercity travel network of 155 cities is built in. An edge connecting two nodes represents an airline route connecting two cities. Daily travel volume between two cities defines a weight assigned to an edge linking the two cities. The 155 cities in the model exchange a subset of their population on a daily basis as defined in the network. Total population size in the model is six hundred million, and spreading of an epidemic disease within a city is modeled by an equation-based *SEIR* model. SEIR model is a type of an epidemic model, where an additional population subgroup, called exposed (E), is included, where infected individuals are not yet infectious themselves.

There are a few modifications and assumptions we implemented to the original model to better suit the purpose of our experiment. While GEM allows to use stochastic traveling mode in the model, we use it in a deterministic mode to reduce complications in analyzing our experimental results. We ignore the seasonality effect on the spreading of a pandemic disease. For a weight of an edge between two cities, we use an average of two edges of opposing direction (city *A* to *B* and *B* to *A*) to keep the balance of travel volumes between the two cities. This assumption stabilizes the population size of each city in the network. Finally, we assume all 155 cities have equal population size, so each city's population is approximately four million.

Our experiment compares two scenarios as follows: we first rank-order the 155 cities according to their centrality value in the global air travel network. We then group them into three classes - 50 high centrality nodes, 55 medium centrality nodes, and 50 low centrality nodes. Each of these groups represents a relative importance of the nodes in a network topological sense. Since the network has weighted edges, we use a

generalized centrality measure on a weighted network: C_D^w, C_C^w, and C_E^w. 155 nodes are then rank-ordered with respect to each centrality measure. Note that we exclude the betweenness centrality because more than 100 nodes have zero C_B^w, thus not suitable for the experiment. Next, we assume three levels of $R0$ values - high, medium, and low - to assign to each group. We assign a high $R0$ value to the high centrality group, a medium $R0$ to the medium centrality group, and a small $R0$ value to the low centrality group. This is the first scenario. In the second scenario, we assign $R0$ values in reverse order: low $R0$ to the high centrality group, medium $R0$ to the medium centrality group, and high $R0$ to the low centrality group.

The first scenario depicts a situation where "topologically important" nodes act as amplifier nodes. In the second scenario, topologically important nodes act as attenuator nodes (or amplification with less intensity). Since the population size of all cities is equal, the *average* $R0$ in the two scenarios can be thought of as being equivalent. Also recall that the underlying network topology is the same for both scenarios. Thus, the two scenarios are expected to highlight the effect from a different pairing configuration of two properties for a node: its topological centrality, which comes from a network structure, and an amplifying/attenuating characteristic, which is an intrinsic property of a node.

Initial number of infected individuals is set to one hundred in a city of disease outbreak, and we run 155 simulations by making each of the 155 cities an originating node. Each of the two scenarios are tested under various settings for $R0$ values from a large range of (0.8, 1.7, 2.6) to a small range of (1.6, 1.7, 1.8). We compare the two scenarios in each setting for their global peak time and volume of infected population, averaged from simulation runs finished as pandemic case.

3.2 Results

Table 2 shows the peak time and peak volume from the two scenarios: (high centrality)-(high $R0$) vs. (high-centrality)-(low $R0$) where high $R0$ = 2.6 and low $R0$ = 0.8. This result is obtained by using degree centrality C_D^w as a centrality measure. Although not reported here, we have similar results when using the other two centrality measures.

Table 2: Global peak time and volume for the two scenarios: High $R0$ = 2.6, Low $R0$ = 0.8

	Global Peak Time (*day*)			Global Peak Volume ($10^7 ppl$)		
	Min	Median (Mean)	Max	Min	Median (Mean)	Max
High Centrality, High $R0$	63	90 (89.15)	120	2.57	3.68 (3.54)	4.46
High Centrality, Low $R0$	118	203 (211.28)	353	1.29	1.56 (1.60)	2.23

Result shown in Table 2 displays a significant difference between the two scenarios in both peak time and peak volume. When high centrality nodes are assigned a high $R0$ value, an epidemic turns out to spread much faster (mean peak time of 89.15 days) than in the case of (high centrality)-(low $R0$) case (mean peak time of 211.28 days). It is not only that a peak of epidemic comes faster, but the volume of the infected at the peak time is higher ($3.54 * 10^7$ vs. $1.60 * 10^7$).

This result agrees with our intuitive expectation. Recall again that in both scenarios, the topological structure of the underlying network is exactly the same. Thus, the difference comes solely from whether high centrality nodes act as an amplifier or an attenuator. In the first scenario, the network is configured such that nodes with high centrality amplify the intensity of a disease flow by *incubating* a large number of infected individuals. Consequently, not only does it broadcast the disease to a larger number of neighboring cities (due to high centrality), but also at a larger magnitude. On the other hand, in the second scenario, nodes with high centrality are not as effective in spreading the disease to their neighbors because the intensity of disease flow is weakened. By a same token, high $R0$ nodes are limited in their ability to spread the disease by their disadvantageous topological position (i.e. low centrality).

To further examine the effect of $R0$ on disease spreading, Figure 1 shows the peak time and peak volume for several $R0$ settings. For both graphs, bars on the left of a reference case of uniform $R0$ - (1.7, 1.7, 1.7) - are the results from (high centrality) - (high $R0$) scenarios, and on the right are the results from

(low centrality) - (high $R0$) scenarios. It shows that (high centrality) - (high $R0$) scenarios cause a stronger disease spreading with a shorter peak time and a higher peak volume, and the opposite is true for (high centrality) - (low $R0$) scenario.
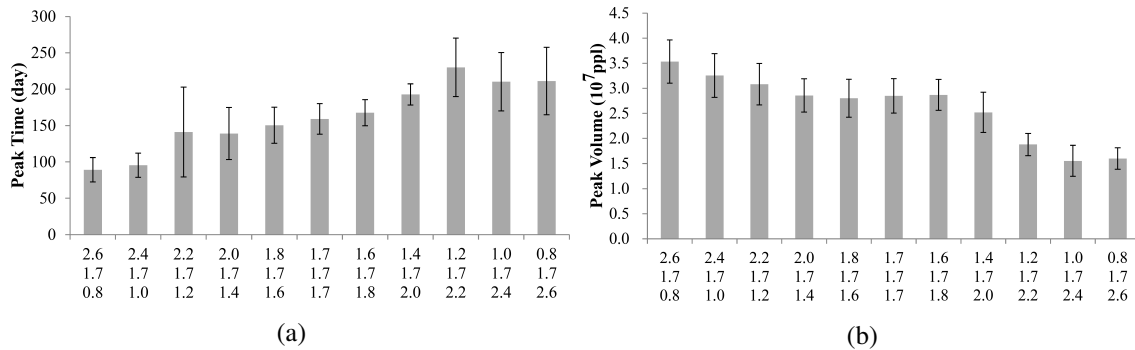

(a)   (b)

Figure 1: Effect of centrality-$R0$ pairing: (a) peak time (b) peak volume. Three numbers on x-axis category denotes, from top to bottom, $R0$ values used for high-, medium-, and low-centrality node group. Bar indicates one-standard deviation.

Figure 1 also shows that as the gap between high-$R0$ and low-$R0$ gets wider (i.e. farther away from the middle, reference case), the effect of non-uniform $R0$ among the nodes is more pronounced. This trend suggests that x-axis is subject to a potentially meaningful interpretation. One possible interpretation is a "degree of synergy" between the two aspects of a node: its centrality and a spread of $R0$ values among the nodes in a network. The degree of synergy that promotes disease spreading is maximum in the left-most case, and gradually decreases to a minimum for the right-most case.

All these results indicate that $R0$, a node's intrinsic property, plays an important role in assessing a disease spreading dynamics on a network. It verifies our initial hypothesis that an epidemic's spreading on a network is affected not only by the network's topological structure but also by the nodes' property.

An implication of the above finding on targeted vaccination strategy design is that we need a metric that takes into account a node's network topological importance and its flow amplifying property. In other words, superspreaders on a epidemic spreading network cannot be determined by network centrality alone and a node's property should be incorporated. This gives us a motivation to develop a new centrality measure in a specific context of epidemic spreading networks.

4 CENTRALITY WITH NODE PROPERTY

This section discusses a proposed scheme of incorporating node property into the degree and closeness centrality. We illustrate how they are calculated in the global city network we use in this paper. We begin this section by turning our attention to the flow characteristics of infectious disease spread in a network of cities where the cities are connected through global air transport network.

First, this network is clearly a weighted network, and the weight of a link between two cities represents the volume of travelers. Thus, it warrants a use of centrality measures developed for a weighted network. Second, according to Borgatti's typology of flow processes (Borgatti 2005), infectious disease falls under the category of serial duplication on paths in a network. However, disease spreading on a network of city deviates in both dimensions. When a city hosts a number of infected individuals and they travel to other cities, these individuals carry the disease to their destination cities. A node *broadcasts* the disease rather than transmits it to a single node per time. Thus, unlike individual's contact network, it follows a parallel duplication process. In terms of the flow trajectory, it takes unrestricted walks rather than paths. It is stated in Borgatti (2005) that *the infection spreads from person to person by duplication, ..., but does not re-infect anyone who already has had it because they become immune.* This is the reason why a spread of

infection is said to follow a path: it does not visit the same node more than once. It is true for individuals' contact network, but for a city network, it is possible that a city is affected by inflow of infected individuals repeatedly. In the midst of disease flow, the same nodes can be visited multiple times and the same edge can be traversed multiple times as well. These differences make it worthwhile to test which centrality measures would be the best indicator of superspreader nodes.

Another characteristic of this network is the role of individual node's intrinsic property. When we use weights for disease spread network, what we intend to capture is the exchange volume of infected individuals. This would properly represent the strength of a tie in the context of disease spread. This quantity, in turn, is affected not only by travel volume between two cities, but also by how prevalent the disease is in the cities that we are looking at. For a given travel volume from city *A* to city *B*, the fraction of infected individuals in the influx to city *B* will be higher when city *A* has a large volume of infected population. As mentioned in section 3, a node can act as an amplifier or attenuator as it transmits a disease, and *R*0 of a node is an important factor to determine a node's role. This leads us to a conjecture that, with topological centrality being equal, a node with high *R*0 is more *central* than the ones with low *R*0.

In (5), $\sum_{j=1}^{n} w_{ij}$ can be interpreted as the total outflux of traveling people from node *i*. Here, we assume the links represent outgoing travels for the sake of simplicity. Note that what we are truly interested in is the outflux of infected individuals per each link. Then, we can modify the summation as

$$\sum_{j=1}^{n} \rho_i w_{ij} = \rho_i \sum_{j=1}^{n} w_{ij}, \tag{7}$$

where ρ_i is the infected population in node *i* as a fraction of the total number of population in node *i*. ρ_i is a function of time. In general, it increases with time, reaches its peak, and gradually drops down. Thus evaluating (7) in a way that suits our intention is not straightforward. However, we can use the fact that ρ_i is proportional to *R*0, at least up to certain point (Figure 2).
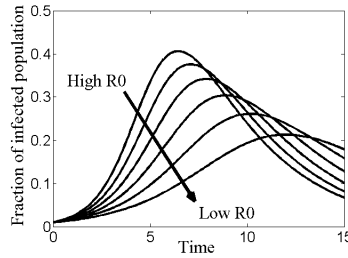

Figure 2: $\rho_i(t)$ is proportional to *R*0 up to certain point.

Thus, we use *R*0 as a surrogate for ρ_i to rewrite (7) as:

$$\rho_i \sum_{j=1}^{n} w_{ij} \propto R0(i) \sum_{j=1}^{n} w_{ij} = R0(i) \times C_D^w(i). \tag{8}$$

As in Opsahl, Agneessens, and Skvoretz (2010), we use a tuning parameter β to allow some room to account for the uncertainty associated with the use of *R*0 as a proxy for ρ_i. Then, *R*0-adjusted degree centrality is defined as:

$$C_D^{w,\beta}(i) = R0(i)^{\beta} \times C_D^w(i). \tag{9}$$

Likewise, we can modify the eigenvector centrality for a weighted network in (6) by multiplying ρ_i and subsequently approximating by *R*0 with a tuning parameter:

$$e_i = \lambda^{-1} \sum_{j} \rho_i A_{ij}^w e_j \propto \lambda^{-1} \sum_{j} R0(i)^{\beta} \cdot A_{ij}^w e_j, \tag{10}$$

which defines $R0$-adjusted eigenvector centrality, $C_E^{w,\beta}$.

When defining the closeness and betweenness centrality for weighted network, the inverse of the weights is used to represent a length of each link, $l_{ij} = 1/w_{ij}$. Following the same logic we use for $R0$-adjusted degree centrality, we redefine a length of each link as:

$$l_{ij}^* = \frac{1}{\rho_i w_{ij}} \propto \frac{1}{R0(i)^\beta \times w_{ij}}. \tag{11}$$

With (11), a new shortest path can be identified, and we can compute $R0$-adjusted closeness and betweenness centrality. This modification can be interpreted as adding a penalty to a path that contain nodes with low $R0$.

5 DISCUSSIONS AND CONCLUSION

We study the effects of a node's intrinsic property and structural centrality in infectious disease spreading using a hypothetical global air transport network. Compared to a contact network of individuals, disease spreading on a network of cities connected by air transport presents differences in two aspects. First, its flow process follows a parallel-duplication process on unrestricted walks, rather than a serial-duplication on paths. Second, each node is not a passive transmitter of diseases, but can augment or attenuate the intensity of the flow. These differences motivate us to test impacts of node's intrinsic property, $R0$ value, within the same network topology, and we find that different $R0$ settings change overall epidemicity.

To take into account a city's role as an amplifier or attenuator of disease spreading, we propose a new epidemic centrality measure, $R0$-adjusted centrality, by incorporating the basic reproduction number $R0$ of each city into the weighted centrality measures. When assessing *disease spreading centrality* of a node on a network, incorporating edge weights and $R0$ of nodes makes logical sense. Especially for degree-based centrality measures, these additional considerations have a relatively clear interpretation. Including edge weights reflects the volume of population exchange between cities, and thus seems a more accurate representation for disease transmission on an air travel network. It can be further refined by estimating the volume of infective population exchange along the edges. This is done by using $R0$ of each city as a proxy for the fraction of infective population during the disease spreading. We plan to test the effectiveness of of the proposed centrality measure in terms of identifying key nodes in a disease spreading network.

There are some limitations that need to be addressed in future work. First, in deriving $R0$-adjusted centrality, we only consider the outflux of infected travelers and attach the originating node's $R0$ to the edge. In our case of intercity network, $\rho_i(t)$ in (8) and (11) is also affected by the influx of infected individuals from other cities. This simplification needs careful investigation. Second, effects from our assumption of equal population size for all nodes need to be examined. Third, use of mean values for interpreting the results from the 155 simulation runs of a different origin of a disease outbreak may not be the best interpretation.

In addition to the further research issues mentioned above, there is another interesting topic for investigation. In a recent study by Opsahl, Agneessens, and Skvoretz (2010), they point out that degree-based centrality measure has two components: strength of an edge and the number of degree itself. Likewise, for closeness and betweenness centrality, the shortest path on a binary network only looks at minimizing intermediary nodes and it ignores the number of intermediary nodes on a weighted network. They argue that while the weighted centrality measures are a useful extension, ignoring the sheer number of links may fail to capture subtleties in the underlying flow process. Incorporating $R0$-adjusted generality into their measure is not readily evident, and requires some modification.

REFERENCES

Barrat, A., M. Barthelemy, R. Pastor-Satorras, and A. Vespignani. 2004. "The architecture of complex weighted networks". *Proceedings of the National Academy of Sciences of the United States of America* 101 (11): 3747–3752.

Bonacich, P. 1972. "Factoring and weighting approaches to status scores and clique identification". *The Journal of Mathematical Sociology* 2 (1): 113–120.

Bonacich, P. 1987. "Power and Centrality: A Family of Measures". *American Journal of Sociology* 92 (5): pp. 1170–1182.

Borgatti, S. P. 2005. "Centrality and network flow". *Social Networks* 27 (1): 55 – 71.

Borgatti, S. P. 2006. "Identifying sets of key players in a social network". *Computational & Mathematical Organization Theory* 12:21–34.

Brandes, U. 2001. "A faster algorithm for betweenness centrality*". *The Journal of Mathematical Sociology* 25 (2): 163–177.

Chen, Y., G. Paul, S. Havlin, F. Liljeros, and H. E. Stanley. 2008, Jul. "Finding a Better Immunization Strategy". *Phys. Rev. Lett.* 101:058701.

Cohen, R., S. Havlin, and D. ben Avraham. 2003, Dec. "Efficient Immunization Strategies for Computer Networks and Populations". *Phys. Rev. Lett.* 91:247901.

Colizza, V., A. Barrat, M. Barthelemy, A.-J. Valleron, and A. Vespignani. 2007, 01. "Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions". *PLoS Med* 4 (1): e13.

Colizza, V., A. Barrat, M. Barthelemy, and A. Vespignani. 2006. "The role of the airline transportation network in the prediction and predictability of global epidemics". *Proceedings of the National Academy of Sciences of the United States of America* 103 (7): 2015–2020.

Colizza, V., and A. Vespignani. 2007, Oct. "Invasion Threshold in Heterogeneous Metapopulation Networks". *Phys. Rev. Lett.* 99:148701.

Dijkstra, E. W. 1959. "A note on two problems in connexion with graphs". *Numerische Mathematik* 1:269–271.

Epstein, J. M., D. M. Goedecke, F. Yu, R. J. Morris, D. K. Wagener, and G. V. Bobashev. 2007, 05. "Controlling Pandemic Flu: The Value of International Air Travel Restrictions". *PLoS ONE* 2 (5): e401.

Freeman, L. C. 1979. "Centrality in social networks conceptual clarification". *Social Networks* 1 (3): 215 – 239.

Friedkin, N. E. 1991. "Theoretical Foundations for Centrality Measures". *American Journal of Sociology* 96 (6): pp. 1478–1504.

Guimera, R., S. Mossa, A. Turtschi, and L. A. N. Amaral. 2005. "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles". *Proceedings of the National Academy of Sciences* 102 (22): 7794–7799.

Holme, P. 2004. "Efficient local strategies for vaccination and network attack". *Europhysics Letters* 68 (6): 908–914.

Holme, P., B. J. Kim, C. N. Yoon, and S. K. Han. 2002, May. "Attack vulnerability of complex networks". *Phys. Rev. E* 65:056109.

Kitsak, M., L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse. 2010. "Identification of influential spreaders in complex networks". *Nature Physics* 6:888–893.

Newman, M. E. J. 2001. "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality". *Physical Review E* 64 (1): 016132+.

Newman, M. E. J. 2004, Nov. "Analysis of weighted networks". *Phys. Rev. E* 70:056131.

Opsahl, T., F. Agneessens, and J. Skvoretz. 2010. "Node centrality in weighted networks: Generalizing degree and shortest paths". *Social Networks* 32 (3): 245 – 251.

Schneider, C. M., T. Mihaljev, S. Havlin, and H. J. Herrmann. 2011, Dec. "Suppressing epidemics with a limited amount of immunization units". *Phys. Rev. E* 84:061911.

Sikic, M., A. Lancic, N. Antulov-Fantulin, and H. Stefancic. 2011, October. "Epidemic centrality and the underestimated epidemic impact on network peripheral nodes". *ArXiv e-prints*.

Wallinga, J., and P. Teunis. 2004. "Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures". *American Journal of Epidemiology* 160 (6): 509–516.

AUTHOR BIOGRAPHIES

TAESIK LEE is an associate professor of Industrial & Systems Engineering Department at KAIST, Korea. His research interests include system design, modeling and simulation. For the past few years, he has been working on designing various aspects of healthcare delivery systems. Dr. Lee obtained his MS and PhD degree from MIT in the Dept. of Mechanical Engineering. His email address is taesik.lee@kaist.edu.

HYUN-ROK LEE is a MS candidate in Industrial & Systems Engineering Department at KAIST. His research interests included modeling and simulation. His email address is hyunrok@kaist.ac.kr.

KYOSANG HWANG is a MS candidate in Industrial & Systems Engineering Department at KAIST. His research interests included modeling and simulation. His email address is jpoo02@kaist.ac.kr.