

RESEARCH

Open Access



Coupling effects on turning points of infectious diseases epidemics in scale-free networks

Kiseong Kim¹, Sangyeon Lee¹, Doheon Lee^{1,2*} and Kwang Hyung Lee^{1*}

From DTMBIO 2016: The Tenth International Workshop on Data and Text Mining in Biomedical Informatics Indianapolis, IN, USA. 24-28 October 2016

Abstract

Background: Pandemic is a typical spreading phenomenon that can be observed in the human society and is dependent on the structure of the social network. The Susceptible-Infected-Recovered (SIR) model describes spreading phenomena using two spreading factors; contagiousness (β) and recovery rate (γ). Some network models are trying to reflect the social network, but the real structure is difficult to uncover.

Methods: We have developed a spreading phenomenon simulator that can input the epidemic parameters and network parameters and performed the experiment of disease propagation. The simulation result was analyzed to construct a new marker VRTP distribution. We also induced the VRTP formula for three of the network mathematical models.

Results: We suggest new marker VRTP (value of recovered on turning point) to describe the coupling between the SIR spreading and the Scale-free (SF) network and observe the aspects of the coupling effects with the various of spreading and network parameters. We also derive the analytic formulation of VRTP in the fully mixed model, the configuration model, and the degree-based model respectively in the mathematical function form for the insights on the relationship between experimental simulation and theoretical consideration.

Conclusions: We discover the coupling effect between SIR spreading and SF network through devising novel marker VRTP which reflects the shifting effect and relates to entropy.

Keywords: Epidemics, Social network structure, Scale-free, Susceptible-infected-recovered, Value of recovered on turning point, Spreading phenomena, Contagiousness, Recovery rate

Background

Epidemics, information, memes, cultural fads are representative spreading phenomena observed in human society. The pattern of spreading differs with the structure of the social network. SIR is one of models describing spreading phenomena suggested by A. G. McKendrick et al. [1] in 1924. The model expresses spreading in the form of differential equation among population compartments; susceptibles, infected, and removed. However, this model cannot reflect individual interactions. The network theory emerged since random graph model of Erdős–Rényi model

(ER) [2] in the 1960s. Milgram showed the small world structure separated into 6 step distance through the experiment of mail forward which is reflecting interpersonal connection [3]. The interaction of each person can be represented by nodes and edges via network theory. There are three major network models of different features; scale-free network (SF) by Barabasi [4], small-world network (SW) by Strogatz [5] and ER random network. There are many types of research of spreading phenomena reflecting individual interactions through random network models. Keeling et al. [6] review of this research with the basis of epidemiological theory and network theory and suggest how the two fields of network theory and epidemiological modeling can deliver an improved understanding of

* Correspondence: dhlee@biosoft.kaist.ac.kr; khlee@biosoft.kaist.ac.kr

¹Department of Bio and Brain Engineering, KAIST, Daejeon, South Korea
Full list of author information is available at the end of the article



disease dynamics and better public health through effective disease control. Shirley et al. [7] also compared epidemiological properties of some networks with different levels of heterogeneity in connectedness and mentioned that scale-free was fastest and reached largest in size, then random graph and the small world. Spreading phenomena are highly dependent on the network structure and understanding the structure is important to figure out and predict spreading phenomena. However, understanding the network structure is difficult because of its large scale, privacy, and difficulty in control. We devised a novel marker and observed changes in patterns of SIR epidemic spreading consequently on SF network model by simulation with various epidemic parameters and network parameters. It enables us to find an interesting aspect of coupling between the structure of the social network and the spreading phenomena.

Network model

The scale-free model was suggested by A. Barabasi in 1999 [4]. This model shows fast, large-scale spreading because networks following this model are made up of many nodes of small degree and few nodes of large degree, "Hubs" (Fig. 1).

Many types of research on spreading phenomena used scale-free networks such as finding reproduction number [8], meme spreading [9] and analyzing computer virus data [10].

The small-world model was presented by Strogatz et al. in 1998 [5]. This network model has the possibility of the bridging link between distant nodes in spreading. Researches like percolation [11] [12] and transition to oscillation in epidemics [13] are based on the small-world network model. In this study, we used scale-free networks to reflect the fast and large-scale spreading.

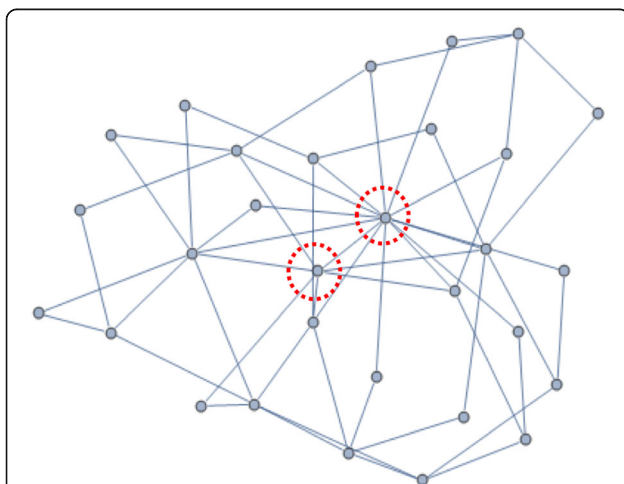


Fig. 1 Scale-Free network with hubs. An example of a scale-free network. Highlighted two nodes are hub nodes, whose degree is larger than non-hub nodes

Close nodes are connected each other while some bridging links which connect between far nodes appear (red stroke) (Fig. 2).

Epidemic model

Many models are describing epidemic spreading. Those models are different in how they define population compartments. In this study, we used simple SIR model which considers two spreading factors; infection and recovery. SIR model uses three compartments; susceptible, infected, and recovered. In 1927, Kermeck and McKendrick presents three differential equation describing the relationship among three compartments.

$$\frac{ds}{dt} = -\beta si, \quad \frac{di}{dt} = \beta si - \gamma i, \quad \frac{dr}{dt} = \gamma i \quad (1)$$

where s , i and r represents susceptible, infected, and recovered respectively [14]. Solving these nonlinear differential equations by the numerical approach, we can get the solution with the form of the time-series function of each compartment.

Figure 3 shows the time-series change of the population of three compartments. The blue, red and green curve represent the change of Susceptible, Infective, and Recovered population respectively. The number of population of susceptible decreases steadily while population of infected increases in the early part and decreases after the turning point (TP) and population of recovered increases continuously. The Eq. (1) was modeled and derived on the assumption of the fully-mixed model, and it cannot reflect the epidemic spreading by the individual contacts.

In this study, as we applied the SF network model to our computational epidemic simulator, we devised a novel marker and observed the changes in each epidemic spreading in the networks. Figure 4 shows a snapshot of network spreading situation with the

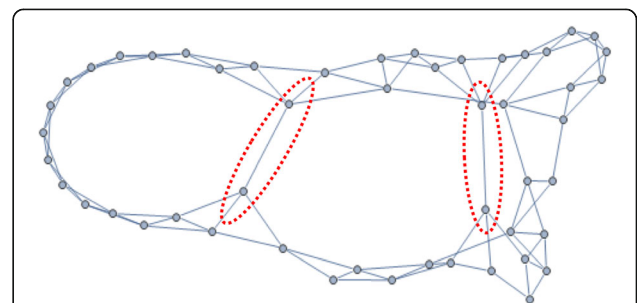
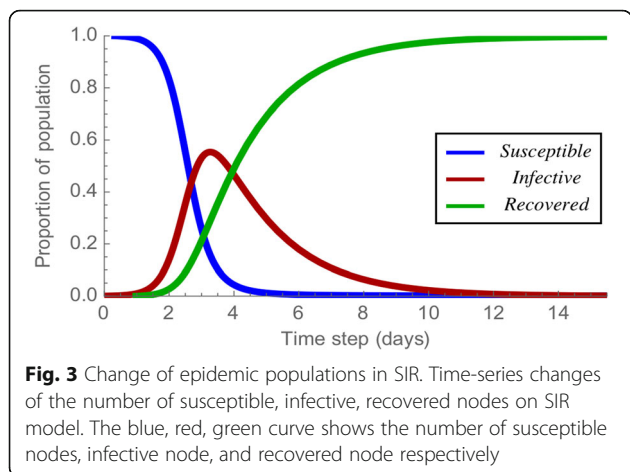


Fig. 2 Small world network with bridging links. An example of small world network. Nodes in small world network are mainly connected with physically adjacent nodes. Based on probability, small world network has bridging edges which connect distant nodes. Two bridging edges are highlighted



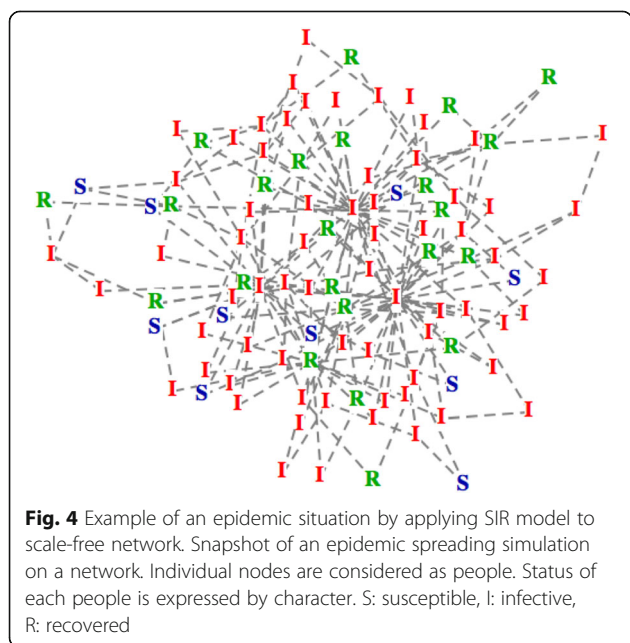
representation of each individual in compartments by using the graph theory with vertex and edge.

The blue “S”, red “I” and green “R” represent susceptible, infected, recovered individuals respectively, and the gray dashed line represents the contact relationship between individuals. Newman [9] calculated the SIR model solution in many different network models in his study. Trying to represent the real aspect of the social network, we used the scale-free network. We observed some features of the novel marker through the simulation of the epidemic spreading on SF network for different spreading parameters and network parameters.

Methods

Devising novel marker VRTP

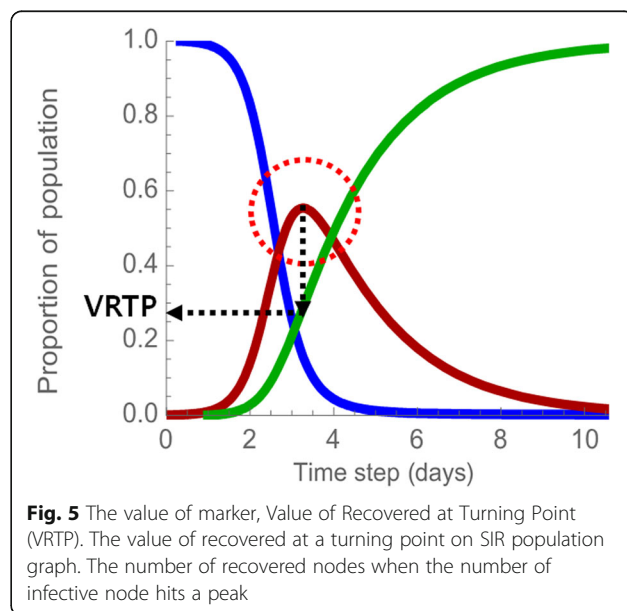
Here, we devise a marker VRTP which is the value of recovered population on the turning point. The turning



point means the time point at the peak of the infected population. The turning point exists in the curves of the infected population in the SIR model. The number of the infected increases till the time is at turning point and after this point, the number of infected nodes decreases. We chose to observe the number of the recovered population as VRTP, the value of recovered population in turning point (Fig. 5). In general, many epidemics researches focused on the number of the infected population instead of the number of the susceptible and the recovered. As we see the relationship in the SIR model, the number of infected changes depending on the recovery parameter, which means that both curves are not independent. However, to understand and predict spreading in the social network in the other aspect, it is necessary to observe the number of S and R also. Through mathematical consideration and simulation for acquiring the maps of VRTP by the parameters, we found a coupling relationship between spreading parameter and network parameter.

Epidemic simulation overview

We built an epidemic simulator for observing the spreading phenomena on the network. For the development of simulator and the simulation, we devise a simulation algorithm. The Inputs of the simulation algorithm include the network (adjacency matrix), β (contagiousness), γ (recovery factor), T (epidemic duration), and q (initial infect ratio). Epidemic duration T is the number of time steps of the simulation, and initial infection ratio q is the ratio of infective vertices to the whole population. First, it generates an SF network for simulation with given number of population and network parameter. It creates an array z, containing the status of the



n^{th} vertex in time t . For each vertex, the value of z at time t can be one of 0, 1, and 2 respectively represents that vertex is susceptible, infective, or recovered. We set every z to 0 at $t=0$ because all vertices are susceptible before the epidemic spreading. It places the infective vertices randomly according to the initial infect ratio q . After the initial adoption, epidemic spreading simulation repeatedly works during epidemic spreading. Then the vertex falls into two kinds of random process stages, the recovery and immunization stage or the infection stage through contacts. In every cycle of epidemic spreading, firstly we search and find the infected vertices and then find infected nodes and their neighboring nodes. We adopted the Monte-Carlo probability experiment using β to determine whether the adjacent node becomes infected or not. In the recovery and immunization stage, through the Monte-Carlo probability experiment using γ again, we decide to make those infected vertices to be recovered or not. With this kind of process, the infected vertex in time t ($z[n, t] = 1$) become recovered vertex in time $t + 1$ ($z[n, t + 1] = 2$) if $1/\gamma$ is bigger than a random real number between 0 and 1. After this stage, in the stage of infection, we find susceptible vertices adjacent to infective vertices in time t ($z[n, t] = 0$, with the adjacent infective node). Among those susceptible vertices, a vertex becomes an infective vertex in time $t + 1$ ($z[n, t + 1] = 1$), which represents epidemic transmission, if β is bigger than a random real number between 0 and 1. For each time step, we recorded the number of susceptible, infective, recovered vertices during epidemic spreading process (Fig. 6).

Construction of VRTP distribution

From the result of the epidemic simulation, we can get the time series data of populations of each compartment. To make the time-series function smooth, we interpolate the discrete time-series data with the cubic spline function. Because the randomness exists in every trial of the epidemic simulation, we gather the result of every simulation trial by each parameter and calculate VRTP value in each trial and construct the distribution of the VRTPs to select the representative value. With Kolmogorov-Smirnov (K-S) statistical test [15], we figure out the distribution be the parametric Gaussian or not. Each mean value of VRTPs distribution from each simulation result by network and epidemic parameters was calculated for further analysis.

Results

Derivation of VRTP formula

For mathematical consideration of VRTP, we derive the VRTP formula with three theoretical assumption model. Those are the fully-mixed model (mean field model), the pairwise approximation model and the degree-based model.

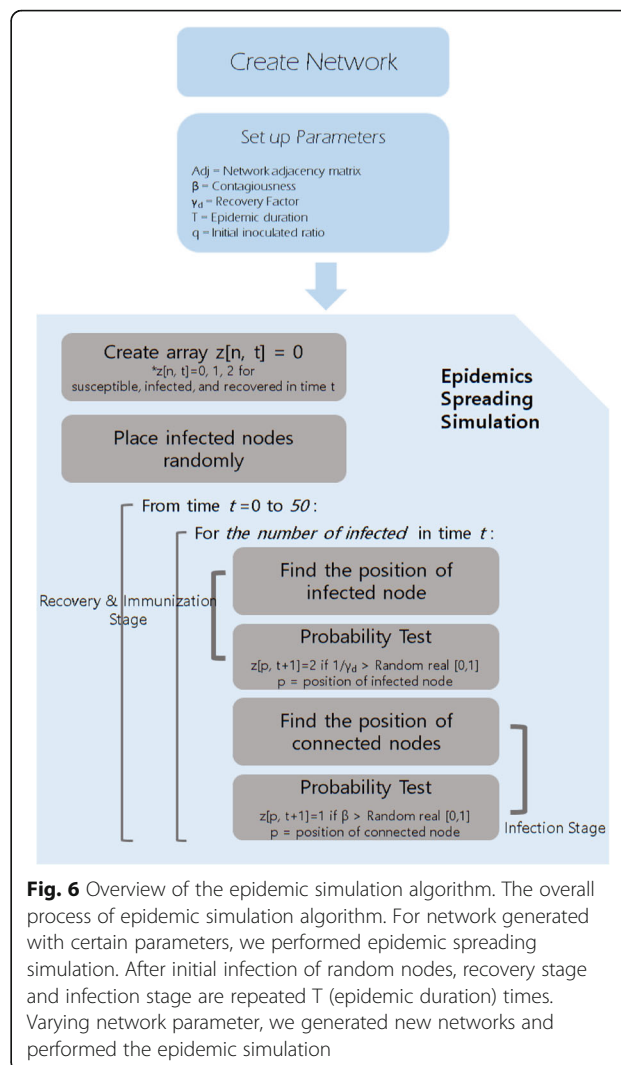


Fig. 6 Overview of the epidemic simulation algorithm. The overall process of epidemic simulation algorithm. For network generated with certain parameters, we performed epidemic spreading simulation. After initial infection of random nodes, recovery stage and infection stage are repeated T (epidemic duration) times. Varying network parameter, we generated new networks and performed the epidemic simulation

For fully-mixed model based on ‘mass action principle’[16], we solved the nonlinear differential Eq. (1) algebraically remaining r at TP. The exact form of r at TP is like following,

$$r_{TP} = \frac{\ln(s_0 R_0)}{R_0} \tag{2}$$

where r_{TP} is VRTP and R_0 is reproduction number, s_0 is the initial value of the susceptible population.

Likewise, we solve the differential equations in the pairwise approximations model or moment closure method [17] (configuration model),

$$\langle r_i(t = TP) \rangle = \frac{\ln(\langle s_i(0) \rangle R_0 k_i)}{R_0 k_i} \tag{3}$$

where k_i is the degree of the i^{th} node.

And in the degree-based model [18],

$$\sum_{k=0}^{\infty} q_k r_k(t) = \frac{\ln(R_0 k s_k(0))}{R_0 k} \tag{4}$$

where q_k is the probability that a vertex with degree k is present.

As we can see, the whole form of the functions is like $\ln(x)/x$ for $s_0 \sim 1$ and x is the reproduction number. The following figure (Fig. 7) shows the change of VRTP values by reproduction number.

From the result of theoretical consideration, finally, we can guess the range of the VRTP values in the SIR spreading model. In observation of the function form of VRTP, we can see that VRTP is low than $1/e \sim 37\%$ at $R_0 = e$.

So, the range of VRTP is $[0, 1/e]$ and epidemic characteristics is divided into two regions, lower ($R_0 < e$) and upper ($R_0 > e$) region. The VRTP increases by R_0 till the $R_0 = e$ and after that point, R_0 decreases by R_0 . From the value of recovered population has the upper bound of 30% of the whole population when the infected population is maximum, we concluded that before the recovered being under 37%, and infected population would be decreased.

VRTP surface and curves

As far as we know, the reproduction number R_0 consists of two parameters, contagiousness β and recovery rate γ and specifically $R_0 = \beta/\gamma$. We did the epidemic simulation by the parameters of two epidemic parameters, contagiousness β and recovery rate γ and of one network structure parameter k . From the simulation result, we constructed the distribution of VRTP and calculated the representative mean value of VRTP. Then we constructed the surface of VRTP and observed the change of VRTP varying those epidemic parameters.

With both β and γ , VRTP increased rapidly from 0 to maximum value and decreased. The surface of VRTP shows some fluctuations while k increases, and the location of the peak of VRTP moves toward $\beta = 0, \gamma = 0$. Also, VRTP always has its maximum value below 30% of the whole population (Fig. 8).

If we magnify the surface of $k = 2$ and $k = 10$, we can observe the area of the low beta area, we can see the smooth change between two curves of low β with sustaining same function form (Figs. 9 and 10).

With network parameter k , VRTP surface changes drastically in the region of low γ_d . If we magnify the surface and observe the curve of low γ_d by k , we do not miss that the change between VRTP curves of low γ_d with changing the function form (Figs. 11 and 12).

We can observe same findings again in the VRTP surfaces with varying k (Fig. 13), which we find out in the previous curves.

Discussion

Derivation of VRTP formula Speed of spreading under fixed γ

k and β affect epidemic spreading speed under fixed γ . β is the probability of infection between the infective vertex and its neighboring susceptible vertex, so higher β results in the increase of speed. Likewise, parameter k used to decide the overall structure of network influences to the speed, vertices in networks with higher k shows higher network density than vertices in networks with lower k , which means more chances to spread epidemic to neighbors. γ_d is the inverse of γ and represents the recovery time after infection. Epidemics with high γ_d results in large number of infective vertices because they can stay in infective vertex for long time steps and infect neighboring susceptible vertices. These

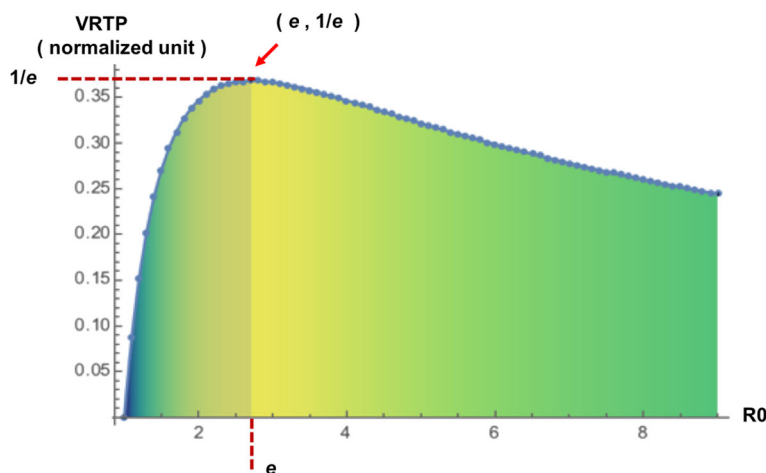


Fig. 7 Theoretical VRTP values by reproduction number. The theoretical result which is calculated using the differential equation of SIR model without networks. The result of theoretical calculation shows the approximate range of VRTP, which is lower than $1/e \sim 37\%$ at $R_0 = e$

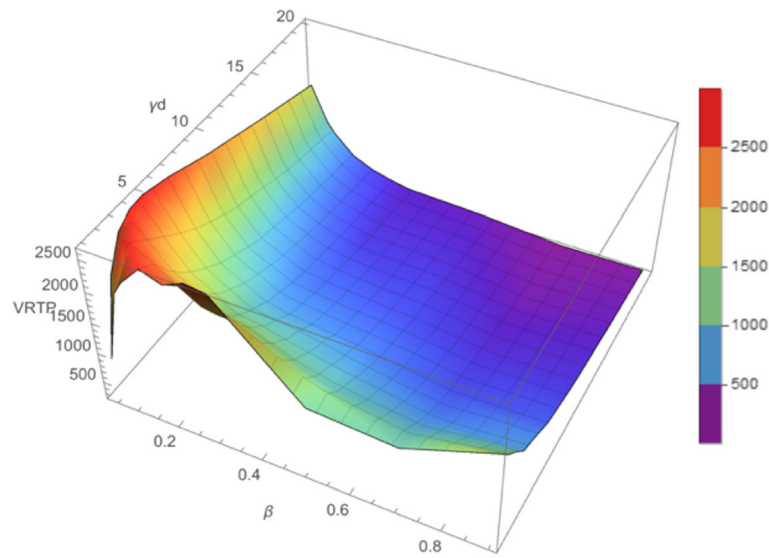


Fig. 8 Example of VRTP surface. An example of VRTP surface. It shows some fluctuations while k increases, and the location of the peak of VRTP moves toward $\beta = 0, \gamma = 0$. VRTP always has its maximum value below about 30% of the whole population as we calculated theoretically

large γ cases are applicable to diseases like acquired immune deficiency syndrome (AIDS) in real-world. On the other hand, the number of recovered vertices increases rapidly in the event of epidemics with low γ_d . In this case, the epidemic spreading is obstructed by recovered vertices because they are considered as disconnected vertices from the network. As a result, epidemics subsided because the recovery speed exceeds the speed of infection.

Shifting the location of saddle point of VRTP curve

VRTP curves normally have two peaks and a saddle point (Fig. 14). With the observation of VRTP curves,

we can conclude that two independent factors influencing the shape of VRTP curve, γ and $\beta-k$. Changes of β and k result similar effects. The increase of these values results in the faster spreading of epidemics. That makes a slight shift to the left of VRTP curve. On the contrary, the increase of γ results shift of VRTP curve to the right. Because the speed of epidemic spreading goes slower while connections of the networks become disconnected quickly as γ increases. However, VRTP curves with parameters $k = 4$ or larger γ do not show fluctuation. In the case of former, the speed of epidemic spreading becomes too slow to make fluctuation although β increases to 0.95. The latter case, in large γ , the

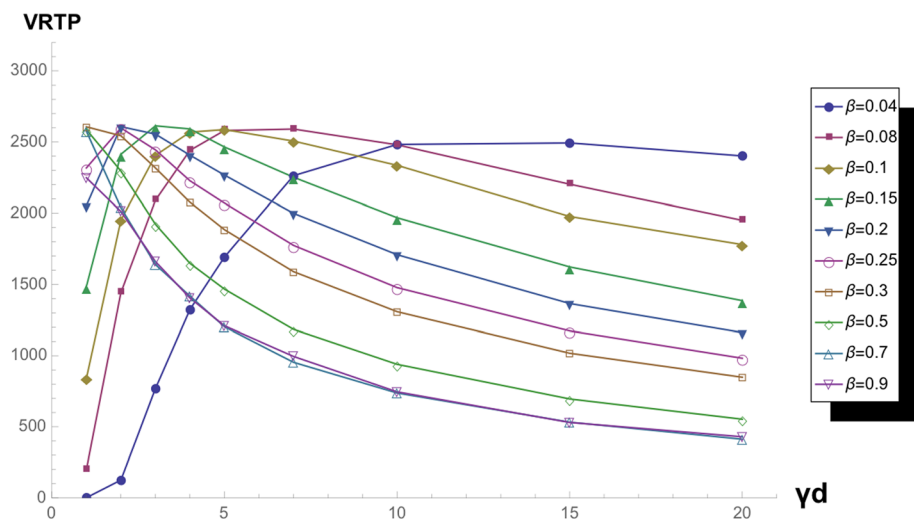
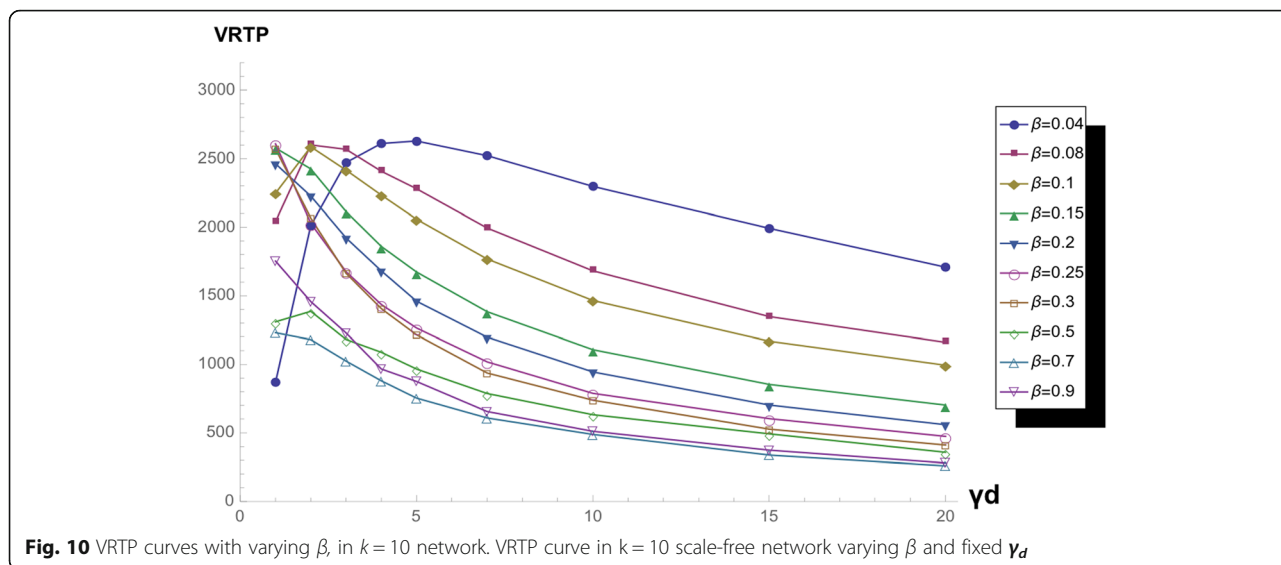


Fig. 9 VRTP curves with varying β , in $k = 2$ network. VRTP curve in $k = 2$ scale-free network varying β and fixed γ_d



obstructing power of γ by disconnecting vertices goes too weak to make fluctuation. We concluded that the saddle point appears to the point that satisfying $4 \sim \gamma \times \beta \times k$.

Network structure change by k and scaling effect

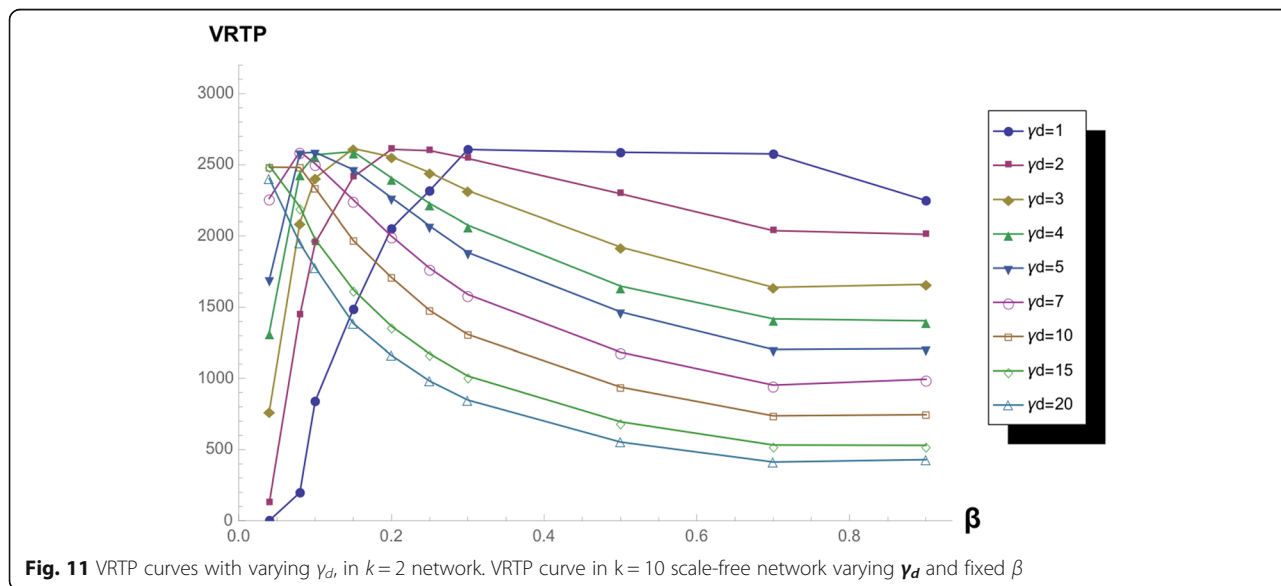
Eqs. (3) and (4) shows the characteristics of VRTP depends on R_0 and also on k which is related the density of the network. In the case of low k , Eqs. (3) and (4) converse to Eq. (2). But in the case of high k , the $\ln(x)/x$ part can be considered as $1/x$. And if we consider the effect of the network parameter, the reproduction number R_0 should be adjusted as like $R_0 k$. That is a scaling effect on VRTP. It is coincident with the result of Bartlett [19].

Some aspect of VRTP function form

All of the VRTP formula has the function $\ln(x)/x$. If we set $s_0 = 1$ and substitute the inverse of reproduction number $1/R_0$ as the probability P , the function form would be the form $-P \ln(P)$. It is the form of Gibbs Entropy [20]. So we can infer that the VRTP may be related to the epidemic system information. And it is necessary for investigating more in the future work.

Needs of the number of the recovered

There are not many epidemics spreading situations that we can draw the VRTP surface. For to draw the surface, we must know the VRTP values of whole epidemic parameters. But if get them, we figure out the network characteristics. We need the data which



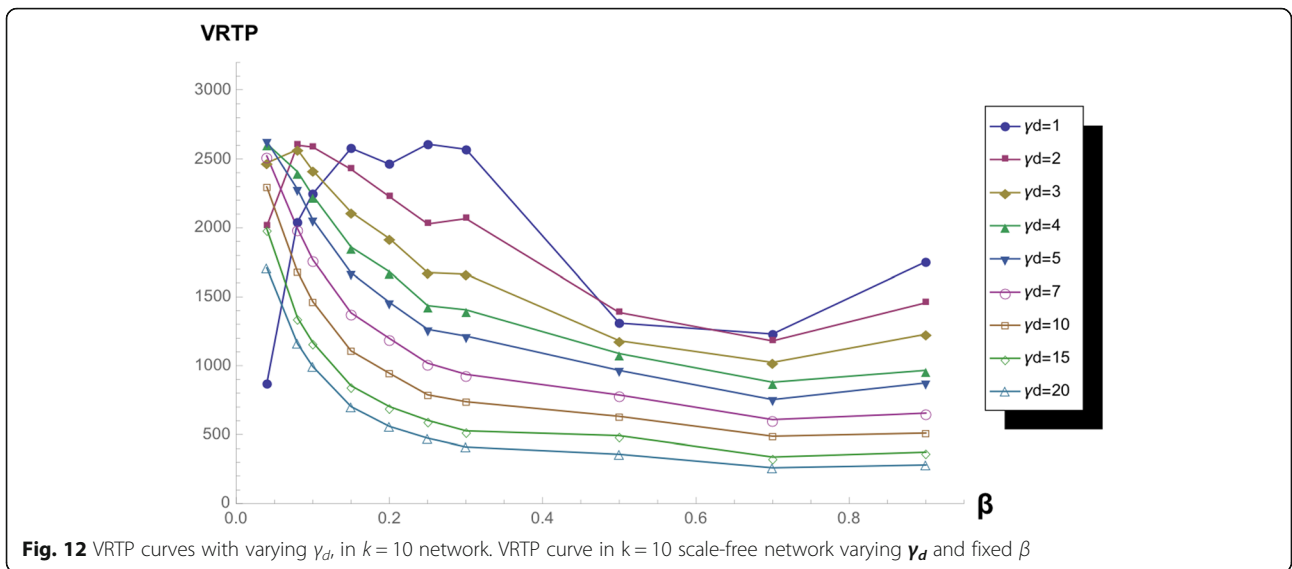


Fig. 12 VRTP curves with varying γ_d , in $k = 10$ network. VRTP curve in $k = 10$ scale-free network varying γ_d and fixed β

contains the time-series number of the recovered population with the infected population simultaneously. That makes us understand the characteristics of coupling effects in VRTP between the network and the SIR epidemics. For an example, in the prevalence of Influenza-Like-Illness (ILI), we must gather not only

the data of the number of the infected but also of the recovered.

Conclusions

We developed an epidemic simulator for the SIR spreading on the SF network. It has a handy ability to parameterize

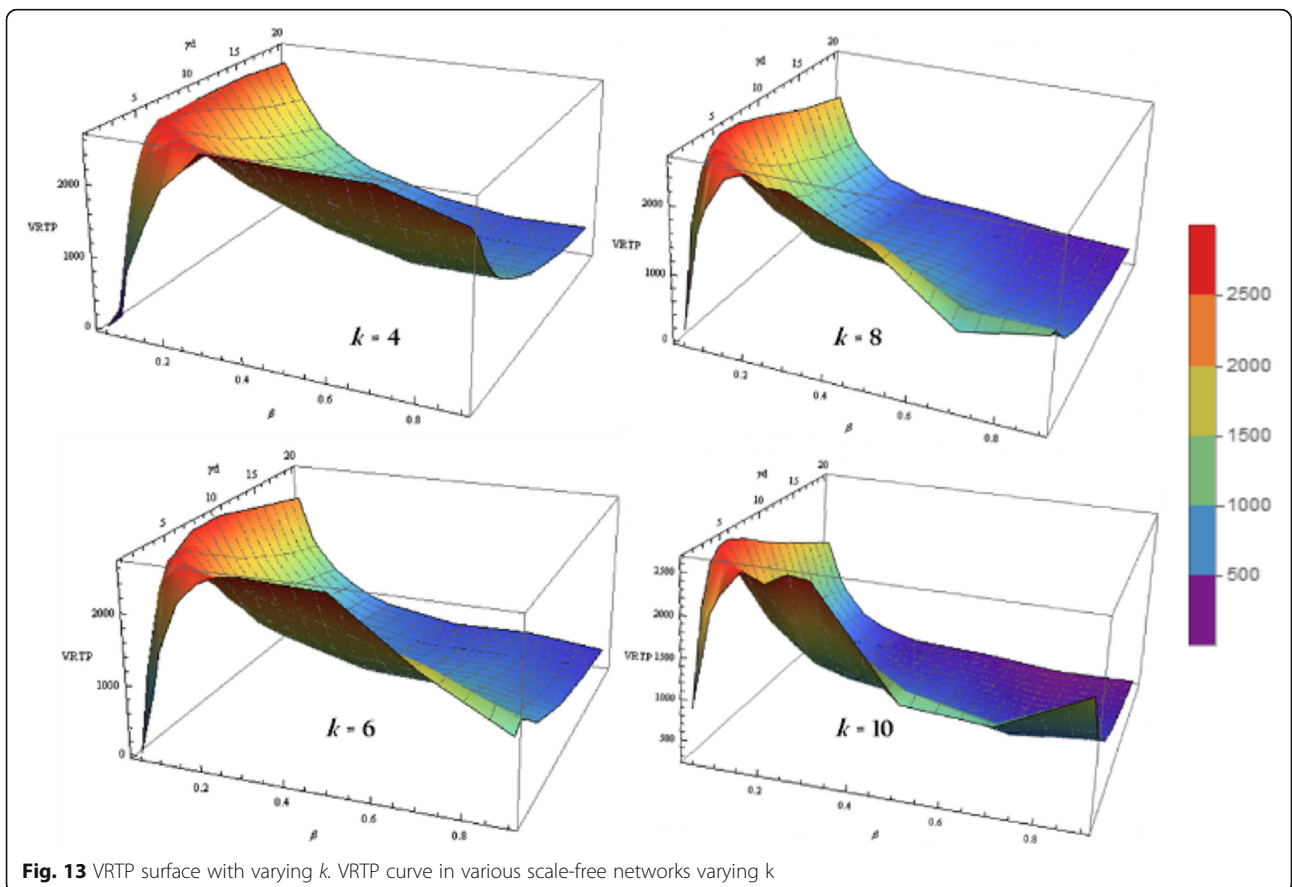


Fig. 13 VRTP surface with varying k . VRTP curve in various scale-free networks varying k

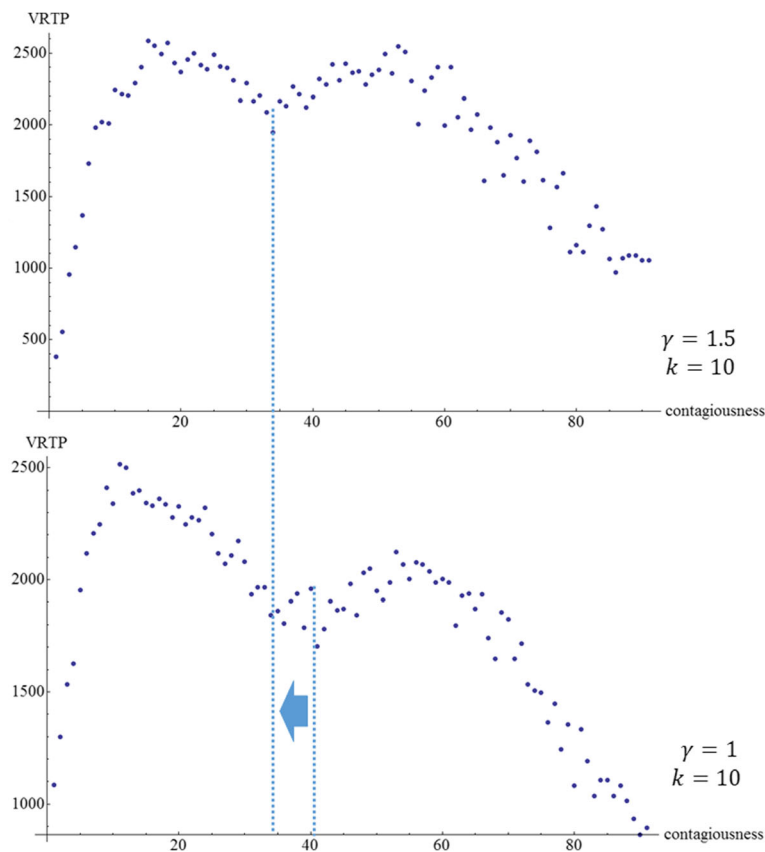


Fig. 14 The location shift of the saddle point of VRTP curve along decreasing γ . The shifting of the saddle point of VRTP curve. Normally there are two peaks and a saddle point in VRTP curve. The curve is influenced by two factors, γ and (β, k) . Changes of β and k increase the speed of epidemic spreading. As a result, the slight shift of VRTP curve to left happens. On the contrary, the increase of γ results in the shift to the right because the speed of spreading decreases

the epidemic processes, network types, node characteristics. And we devise the marker VRTP to reflect the epidemic turning points coupled to the recovered population and to discover the coupling effect between SIR spreading and SF network with the function form of the rough estimation among the parameters k , γ , β . We derive the analytic formulation of VRTP in the fully mixed model, the configuration model, and the degree-based model respectively in the form of entropy.

Abbreviations

ER: Erdos-Renyi; ILI: Influenza-like-illness; K-S: Kolmogorov-Smirnov; SF: Scale-free; SIR: Susceptibles-Infectives-Recovered; TP: Turning point; VRTP: Value of recovered on turning point

Acknowledgements

Not applicable.

Funding

The publication charges for this article was funded by the Bio-Synergy Research Project (NRF-2012M3A9C4048758) of the Ministry of Science, ICT and Future Planning through the National Research Foundation.

Availability of data and materials

The epidemic simulation codes written in Wolfram Mathematica [21] and used during the current study are available in the GitHub (<https://github.com/BioBrain-KAIST/VRTP>).

Authors' contributions

KS, SY designed the method, conducted the experiments, and wrote the manuscript. DH supervised the project. KH gave the ideas and supervised the project. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

About this supplement

This article has been published as part of BMC Bioinformatics Volume 18 Supplement 7, 2017: Proceedings of the Tenth International Workshop on Data and Text Mining in Biomedical Informatics. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-18-supplement-7>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Bio and Brain Engineering, KAIST, Daejeon, South Korea.

²Bio-Synergy Research Center, Daejeon, South Korea.

Published: 31 May 2017

References

- Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proc R Soc London Ser A*. 1927;115:700–21.
- Erdos P, Renyi A. On the evolution of random graphs. *Publ Math Inst Hungar Acad Sci*. 1960;5:17–61.
- Milgram S. The small world problem. *Psychol Today*. 1967;1:61–7.
- Barabasi AL, Albert R. Emergence of scaling in random networks. *Science*. 1999;286:509–12.
- Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature*. 1998;393:440–2.
- Keeling MJ, Eames KTD. Networks and epidemic models. *J R Soc Interface*. 2005;2:295–307.
- Shirley MDF, Rushton SP. The impacts of network topology on disease spread. *Ecol Complexity*. 2005;2:287–99.
- Zhang H, Guan ZH, Li T, Zhang XH, Zhang DX. A stochastic SIR epidemic on scale-free network with community structure. *Phys A*. 2013;392:974–81.
- Saxena A, Iyengar SRS, Gupta Y. Understanding spreading patterns on social networks based on network topology. In: 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). 2015. p. 1616–7.
- Pastor-Satorras R, Vespignani A. Epidemic spreading in Scale-free networks. *Phys Rev Lett*. 2001;86:3200–3.
- Moore C, Newman MEJ. Exact solution of site and bond percolation on small-world networks. *Phys Rev E*. 2000;62:7059–64.
- Kenah E, Robins JM. Network-based analysis of stochastic SIR epidemic models with random and proportionate mixing. *J Theor Biol*. 2007;249:706–22.
- Kuperman M, Abramson G. Small world effect in an epidemiological model. *Phys Rev Lett*. 2001;86:2909–12.
- Hethcote HW. The mathematics of infectious diseases. *SIAM Rev*. 2000;42:599–653.
- DeGroot MH, Schervish MJ. *Probability and Statistics*. Boston: Pearson Education; 2011. pp. 657.
- Hamer WH. The Milroy Lectures on epidemic disease in England the evidence of variability and persistence of type. *Lancet*. 1906;1:733–739.
- Andersson H, Britton T. *Stochastic epidemic models and their statistical analysis*, Springer Lecture Notes in Statistics, vol. 151. New York: Springer; 2000. p. 66–70.
- Barthélemy M, Barrat A, Pastor-Satorras R, Vespignani A. Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *J Theor Biol*. 2005;235(2):275–88.
- Bartlett J, Plank MJ. Epidemic dynamics on random and scale-free networks. *ANZIAM J*. 2012;54:3–22.
- Reichl LE. *A Modern Course in Statistical Physics*. Weinheim: Wiley-VCH; 2009. pp. 344.
- Stephen Wolfram Research Inc. *Mathematica Edition: Version 10.0*. Champaign: Stephen Wolfram Research Inc; 2014.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

