Article

Average reachability: A new metric to estimate epidemic growth considering the network structure and epidemic severity

Bita Shams, **Mohammad Khansari**
Faculty of New Sciences and Technologies, University of Tehran, Amir Abad, North Kargar Street, Tehran 14399-57131, Iran
E-mail: m.khansari@ut.ac.ir

Received 12 May 2019; *Accepted 15 June 2019*; *Published 1 September 2019*

Abstract
It is a fundamental issue to find a small subset of individuals in a complex network such that their immunization (i.e. removal) minimizes epidemic growth in the network. Though some network topological metrics have been proposed to estimate the effect of individual immunization or epidemic growth of the network, none of them considered the severity of the current epidemic. This paper proposes a new metric, called average reachability (AR) to estimate epidemic growth in a network. AR incorporates infection rate of epidemics to make a trade-off between network local connectivity and global reachability. Moreover, we intend to generalize stochastic hill-climbing immunization (SHCI) algorithm to minimize network epidemic growth regarding all estimation criteria. SIR simulation on immunized networks shows that the combination of AR and SHCI results in minimal epidemic growth compared to immunization algorithms that minimize density or sum of square partitions.

Keywords complex network; epidemic spreading; immunization; stochastic hill-climbing algorithm; average reachability.

Network Biology
ISSN 2220-8879
URL: http://www.iaees.org/publications/journals/nb/online-version.asp
RSS: http://www.iaees.org/publications/journals/nb/rss.xml
E-mail: networkbiology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Many real systems such as social interactions, internet, and so on can be modeled as complex networks. In such networks, nodes represent system components (e.g. people, computers, web pages, etc.) and edges represent the relationship between each pair of components (Gao et al., 2011; Zhang, 2012, 2018). The threat of epidemics spreading (e.g., infection, computer viruses, and rumors spreading) upon contact of infected nodes has resulted in the definition of the network immunization problem. Network immunization problem refers to identify a limited subset of nodes whose immunization results in minimal network's susceptibility (Aspnes et al., 2006a; Chai et al., 2011; Chen et al., 2008; Schneider et al., 2011, 2012; Ventresca and Aleman, 2013).

During the last decade, many research have been conducted to develop new immunization strategies. These researches can be categorized into two groups: the first group, called targeted immunization that try to estimate individuals' roles in epidemic spreading. The family of targeted immunization algorithms first rank nodes based on an importance factor and then immunize b nodes with the highest priority (Gao et al., 2011; nodes based on an importance factor and then immunize *b* nodes with the highest priority (Gao et al., 2011; Hébert-Dufresne et al.,

2013; Masuda, 2009; Schneider et al., 2011, 2012) where *b* is the number of available resources for immunization. Several measures have been proposed to assess the node importance, such as node degree (Cohen et al., 2003; Dezső and Barabási, 2002; Holme et al., 2002; Hu and Tang, 2012; Zhang, 2012, 2018; Zhang and Zhan, 2011), betweenness (Chen et al., 2008; Holme et al., 2002; Salathé and Jones, 2010; Schneider et al., 2011; Zhang, 2016b), eigenvector (Restrepo et al., 2006; Zhang, 2018), PageRank (Miller and Hyman, 2007; Ventresca and Aleman, 2013), and, intercommunal centrality (Hébert-Dufresne et al., 2013; Masuda, 2009; Salathé and Jones, 2010; Yamada and Yoshida, 2012). Unlike advantages of targeted immunization algorithms over random immunization, their performance is still so far from optimal solution since nodes are greedily selected for immunization (Borgatti, 2006).

To address this shortcoming, the second family of immunization algorithms has been introduced based on optimization approach. These algorithms take advantage of relations between network topological properties and epidemics (Aspnes et al., 2006a; Chen et al., 2008; Funk et al., 2010; Peng et al., 2010; Schneider et al., 2011, 2012; Shams and Khansari, 2013; Ventresca and Aleman, 2013). A group of research has been conducted based on minimizing the largest connected component (LCC) of networks as an estimator of the worst-case epidemic (Chen et al., 2008; Schneider et al., 2011, 2012; Shams and Khansari, 2013). Moreover, a branch and bound immunization algorithm has been proposed to reduce the largest eigenvalue of the adjacency matrix of network which is inverse of network epidemic threshold (Peng et al., 2010). Some other researches emphasize on the fact that immunization algorithms should inoculate nodes whose removal minimize network density (Funk et al., 2010; Ventresca and Aleman, 2013) or sum of square partitions (SSP) (Aspnes et al., 2006a) which are important factors in network epidemic growth.

A recent study (Shams and Khansari, 2015) shows that the effectiveness of these algorithms highly relies on the severity of epidemics. Density-based algorithms are more effective in suppression of low-infection rate epidemics, and, LCC-based algorithms are more successful to control epidemics with high infection-rate. The reason is that they optimize a metric that cannot accurately estimate epidemic growth in all cases of epidemics. More clearly, density is an important factor in the case of low infection-rate epidemics, yet, size of connected component is getting a better estimator as infection-rate is increasing (Shams and Khansari, 2015).

Furthermore, these algorithms do not take into account the network connectivity structure, and, only consider the macroscopic properties such as the number of nodes (e.g., SSP, LCC) and edges (e.g. density). Intuitively, these metrics are not worthy estimators in different structure of networks. For instance, the number of infected individuals during a certain epidemic might be totally different in small-world and Scale-Free network with equal density and average degree (Shams and Khansari, 2015).

These shortcomings motivated us to introduce a new measure, called Average Reachability (AR), which calculates average epidemic growth by considering the infection rate of epidemics. Moreover, AR takes into account connectivity structure of network by computing a weighted average of nodes' distance as a virtual measure of network connectivity.

The contributions of this paper are as follows:

- We provide a unified theoretical framework to analyze the relations between network properties and epidemic grow at different levels of epidemic severity.
- We propose a new metric, called average reachability (AR), to estimate network epidemic growth. AR takes advantages of epidemic infection rate and distance to address the shortcomings of previous measures (i.e. network density and sum of square partitions)
- Stochastic hill-climbing immunization algorithm is generalized to optimize all criteria of epidemic growth including network density, sum of square partitions and average reachability.
- Our theoretical analysis is supported by numeric result of SIR simulation on immunized networks

The subsequent materials of this paper are organized as follows: In section 2, theoretical analysis is provided to explain how network properties influence epidemic growth. In section 3, a general immunization algorithm is proposed to minimize all criteria for epidemic growth approximation. In section 4, simulation settings and numeric results are discussed. Finally, in section 5, we conclude the whole paper with points to the future work.

Table 1 Mathematical notations

Notation	Description
G	Network
N	Number of nodes (i.e. size)
E	Number of edges
b	Number of immunization resources
S	Immunization solution (i.e. The set of nodes chosen by immunization algorithm for vaccination)
G/S	The residual network after removal of immunization resources.
W	Number of infected individuals
$E[w]$	Expected number of infected individuals
$W(i)$	Number of infected individuals if *i* is the infection seed
α	Infection rate
k_i	Degree of *i-th*node
ρ	Network density
$C(i)$	Size of the component containing *i-th* node
c_j	Size of *j-th* component
$P(i,j)$	The probability that infection is propagated to *j*-th node from *i*-th node
$d(i,j)$	The distance between node *i* and node *j*
SSP	Sum of square partition
AR	Average reachability

2 Approximation of Epidemic Growth Using Network Properties

Network epidemic growth is defined as the total number of infected individuals during an epidemic. Recently, several research have been conducted to estimate epidemic growth regarding network properties (Aspnes et al., 2006b; Danon et al., 2011; Funk et al., 2010; Schneider et al., 2011, 2012; Shams and Khansari, 2013, 2014; Ventresca and Aleman, 2013). Network density and sum of square partitions are the most important criteria. In case of single source of infection, the expected number of infected individuals (i.e. epidemic growth) is obtained by

$$E[W] = \sum_{i=1}^{N} W(i).f(i) \quad (1)$$

where $f(i)$ is the probability of initiating epidemic at node i, $W(i)$ is the expected number of infected individuals by node i and N is number of network vertices. Note that $f(i)$ is $1/N$ in case of uniformly random outbreak. Therefore, the expected number of infected individuals by node i can be obtained by

$$E[W] = 1/N\ (\sum_{i=1}^{N} W(i)) \quad (2)$$

In the following, we provide theoretical analysis on how $W(i)$ can be formulated using three different network metrics: density, sum of square partitions, and average reachability. Table 1 briefly introduces the mathematical notations.

2.1 Density

It is a known fact that the more neighbor a node has, the more likely it is to cause a large number of infected cases (Danon et al., 2011). We assume that an infected node can independently infect each of its neighbors with probability of $p(j)=\alpha$. The number of immediate infected individuals by node i is obtained by

$$E[W(i)] = \sum_{j=1}^{k_i} p(j) * 1 = \sum_{j=1}^{k_i} \alpha = k_i \alpha \quad (3)$$

where k_i is degree of ithnode. Therefore, Eq.2 is can be written as

$$E[W] = \frac{1}{N} \sum_{i=1}^{N} E[W(i)] = \frac{1}{N} \sum_{i=1}^{N} k_i \alpha = \frac{1}{N} \alpha \sum_{i=1}^{N} k_i \quad (4)$$

Sincethe total number of network edges is $E = 1/2(\sum_{i=1}^{N} k_i)$, we have

$$E[W] = \frac{2}{N} \alpha E = (N-1)\alpha\rho \quad (5)$$

where $\rho = E/\binom{n}{2}$)is the network density. Eq.5 proves that all immunization algorithms have to reduce network density as an important factor in the growth rate of epidemics.

Remark.1. *Given a Graph G = ⟨N, E⟩ and number of immunization resources (b), immunization problem refers to the optimization problem of finding a subset of b nodes whose removal minimize network density that is formulated in Eq.6*

$$S_{opt} = argmin\,(\,\rho(G/S) |\, S \subseteq N \text{ and } |S| = b) \quad (6)$$

where S is a subset of b nodes, G/S represent the residual network after removal of S, and ρ(G/S)network density in G/S.

Though infection rate appears in Eq.5, it does not affect immunization algorithms. In other words, immunization algorithms should only minimize the network density regardless of the current epidemic as stated in Theorem.1.

Theorem. 1. *Immunization algorithms, optimizing Eq.5, obtain a specific solution for all epidemics regardless of their infection rate (α). In other word. If S_1 is an optimal immunization solution for epidemic $e_1$with α_1, then $S_1$is an optimal solution for epidemic e_2 with $\alpha_2 \neq \alpha_1$*

Proof*: if S_1 is an optimal solution for e_1, then*

$$\forall\, solution\ S_i\ E[W|S_1] \leq E[W|S_i] \rightarrow$$

$$\forall\, solution\ S_i (N-1) * \alpha_1 * \rho(G/S_1) \leq (N-1) * \alpha_{1*}\, \rho(G/S_i) \rightarrow$$

$$\forall\, solution\ S_i \rho(G/S_1) \leq \rho\left(G/_{S_i}\right) \rightarrow$$

$$\forall\, solution\ S_i (N-1) * \alpha_2 * \rho(G/S_1) \leq (N-1) * \alpha_2 * \rho\left(G/_{S_i}\right).$$

Hence, S_1 is optimal solution for e_2.

2.2 Sum of square partitions (SSP)

Though network density strongly influences epidemic growth, it only estimates the first waves of infected individuals and fails to estimate the number of infected individuals in long-term. To overcome this shortcoming, a new measure has been proposed based on connected components of networks (Aspnes et al., 2006b; Chen et al., 2008; Gallos et al., 2007; Hadidjojo and Cheong, 2011; Masuda, 2009; Niu et al., 2009; Restrepo et al., 2006; Schneider et al., 2011, 2012; Ventresca and Aleman, 2013; Yamada and Yoshida, 2012; Yoshida and Yamada, 2012).

A network connected component is defined as a set of nodes which all are reachable from each other (Zhang, 2012, 2016a, 2018). Since there is no reachable path between components, the epidemic is suppressed in the component where it is started (Chen et al., 2008; Gallos et al., 2007; Hadidjojo and Cheong, 2011; Masuda, 2009; Niu et al., 2009; Restrepo et al., 2006; Schneider et al., 2011, 2012; Shams and Khansari, 2015; Ventresca and Aleman, 2013; Yamada & Yoshida, 2012; Yoshida and Yamada, 2012). Therefore, an infection seed *i* can maximally infect all other nodes in the component contains *i* (i.e. $C(i)$), and, as a consequence, $W(i)$ is lower than or equal $C(i)(E[W(i)] \leq |C(i)|)$. Hence, the expected number of infected individuals is obtained by

$$E[W] = \frac{1}{N}\sum_{i=1}^{N} W(i) \leq 1/N \sum_{i=1}^{N} |C(i)| \quad (7)$$

where $C(i)$is the size of connected component that contains node i.

As nodes have equal probability to initiate infection, the probability of an outbreak occurring in *j*-th component is obtained by $(|c_j|/N)$ where $|c_j|$ is size of *j*-th component and *N* is network size. Therefore, the expected mean number of infected individuals (i.e. mean worst epidemic size) is $(1/N)\sum_{j=1}^{N_C} |c_j|^2$ called sum of square partition (Aspnes et al., 2006a).

$$E\,[W] \leq 1/N \sum_{i=1}^{N} C(i) = \sum_{j=1}^{N_C} |c_j| * \left(|c_j|/N\right) = (1/N)\sum_{j=1}^{N_C} |c_j|^2 = SSP \quad (8)$$

where N_c is the number of connected components in the network. According to Eq.8, sum of square partitions of the network is an upper bound for the estimated number of infected individuals.

Remark.3. *Given a Graph ($G = (N, E)$) and number of immunization resources (b), immunization problem refers to the optimization problem of finding a subset of b nodes whose removal network sum of square partitions (SSP), as is formulated in Eq.9*

$$S_{opt} = argmin\,(\,SSP(G/S)|\,S \subseteq N\ and\ |S| = b) \quad (9)$$

where S is a subset of b nodes, G/S represents the residual network after removal of S, and SSP(G/S) sum of square partitions in G/S.

Note that all immunization algorithms minimizing SSP, converge to a unique solution regardless of epidemic severity since SSP does not rely on the infection rate (See Eq.8), and so on, the infection rate does not affect the optimal solution of Eq.9 just like Eq.6.

2.3 Average reachability (AR): A new criterion for network epidemic growth

Estimation of epidemic growth using aforementioned metrics are not precise since the network density only covers the immediate impact of infected nodes and SSP is an upper bound for the number of infected individuals. Moreover, there are two other motivations to define a new metric to estimate the number of infected individuals: First, current criteria neglect the epidemic parameters such as infection rate between two individuals. Clearly speaking, they estimate the same epidemic growth regardless of variations of infection rate for different type of epidemics such as influenza, smallpox, SARS. Second, previous metrics do not consider the connectivity structure of networks. In other word, they are not able to distinguish between networks with similar density and component size. For instance, in Fig.1, SSP gives a similar estimation of epidemic growth in all networks in spite of high connectivity and consequent epidemic growth of network (A). On the other hand, density cannot discriminate between networks (C) and (D), while, epidemic growth in network (C) is obviously more than network (B) and (D) due to its lower average distance.

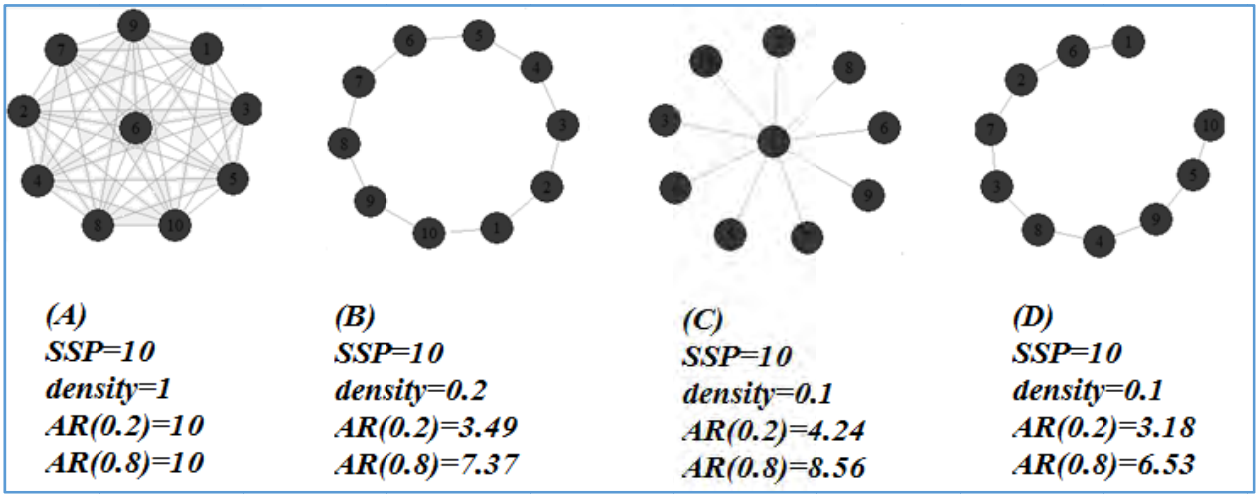

Fig. 1 Sample networks to illustrate shortcoming of density and SSP metrics in including infection rate and network connectivity. (A) Full graph. (B) Ring graph. (C) Star graph. (D) Line graph.

To address these shortcomings, we introduce a new measure to approximate the network epidemic growth (i.e. number of infected individuals). The new measure, called average reachability (AR(α)), calculate network epidemic growth with regards to infection rate of the epidemic (i.e. α). Additionally, it is an implicit combination of network density and sum of square partitions.

Fig.1 sample networks to illustrate shortcoming of density and SSP metrics in including infection rate and network connectivity: (A) Full graph. (B) Ring graph. (C) Star graph. (D) Line graph.

Here, a more precise measure is introduced to estimate $E\,[W(i)]$. In contrast to Eq.3 which only considers the transmission probability from a node to its neighbors, we calculate the transmission probability to all nodes of networks

$$E[W(i)] = \sum_{j \neq i} P_\alpha(i,j) \tag{10}$$

where $P_\alpha(i,j)$ is infection probability of node *j* by node *i*.

Infection probability between node *i* and *j* depends on the length and number of paths between node *i* and node *j*. Trivially, the probability that node *i* infect node j through the path with length of *l* is equal to α^l. As infection probability exponentially decreases when *l* increases, the shortest path between two nodes has the

main impact on the value of infection probability between them. That is why shortest path-based immunization algorithms (e.g. between immunization) outperforms other network immunization algorithm

Based on this fact and also, high computational complexity to find all connecting paths of node *i* and *j*. we assume that viruses are only propagated upon the shortest path between nodes (Hadidjojo and Cheong, 2011; Schneider et al., 2011; Shams and Khansari, 2014, 2015). Therefore, $P_\alpha(i,j)$ can be obtained by

$$P_\alpha\ (i,j) = \alpha^{d(i,j)} \tag{11}$$

where α is the transmission probability over a link and $d(i,j)$ is the distance (i.e. the length of shortest path) between node *i* and *j*. Obviously, α is lower than or equal to one. Therefore, the expected number of infected individuals in duration of epidemics is obtained by

$$AR(\alpha) = E[W] = (1/N)\sum_{i=1}^{N} W(i) = 1/N \sum_{i=1}^{N}\sum_{j\neq i} P_\alpha(i,j) \tag{12}$$

which we call it average reachability (AR) and is simply calculated through Eq.13

$$AR(\alpha) = (1/N)\ \sum_{i=1}^{N}\sum_{j\neq i} \alpha^{d(i,j)} \tag{13}$$

AR converges to sum of square partitions when $\alpha\rightarrow 1$ as proved in Theorem.2. It makes sense due to the impact of infection rate on the epidemic reproduction number. In other words, high infection rate results in fast epidemic spreading through the whole network (Hartvigsen et al., 2007; Salathé and Jones, 2010; Shams and Khansari, 2015; Ventresca and Aleman, 2013). Therefore, immunization algorithms should fragment network to small-components such that the epidemic will be suppressed in one component and not be propagated to other components.

Theorem.2. *Average reachability is equal to the sum of square partitions when* $\alpha \rightarrow 1$.

Proof: *It is well-established that the distance of nodes in two different components is infinite. Since* $lim_{d(i,j)\rightarrow\infty}\ \alpha^{d(i,j)} = 0$, *inner summation of Eq.13 is limited to nodes in C(i)(i.e. the component containing node i).*

$$E[W] = \left(\frac{1}{N}\right)\sum_{i=1}^{N}\sum_{j\in C(i)} P(i,j) \tag{14}$$

Now, assuming$\alpha \rightarrow 1$, *we have*

$$lim_{\ \alpha\rightarrow 1}\ E[W] = \left(\frac{1}{N}\right) lim_{\ \alpha\rightarrow 1}\sum_{i=1}^{N}\sum_{j\in C(i)}\alpha^{d(i,j)} = \left(\frac{1}{N}\right)\sum_{i=1}^{N}\sum_{j\in C(i)} 1 = \left(\frac{1}{N}\right)\sum_{i=1}^{N}|C(i)| \tag{15}$$

As Eq.15 is equivalent of Eq.7, so its result will be $(\frac{1}{N}\sum_{j=1}^{N_C}|c_j|^2)$ *which is sum of square partitions.*

Moreover, AR converges to network density when $\alpha\rightarrow 0$ (See Theorem.3). In other words, weak epidemics are suppressed in the local neighborhood of infection seed as they die out in a short time. Therefore, local connectivity is more important than global reachability in case of small infection rate.

Theorem 3. *Average reachability is reduced to network density when* $\alpha \rightarrow 0$.

Proof: *Average reachability sum up the distance (i.e. length of shortest path)) between all pairs of the*

nodes in the network. Therefore, AR takes into accounts the distance between each pair two times, Eq.13 can be written as

$$E[W] = (1/N)\ \sum_{i=1}^{N}\sum_{j\neq i}\alpha^{d(i,j)} = (2/N)\ \sum_{L=1}^{L_{max}}\alpha^{L} * N_l \quad (16)$$

where N_l is the number of shortest path with length L in the network and L_{max} is the maximum distance in the network.Now, for small α, we have:

$$lim_{\alpha\to 0}\, E[W] = lim_{n\to\infty}\,(2\alpha/N)\ \sum_{L=1}^{L_{max}}\alpha^{L-1} * N_l = (2\alpha/N)(N_1 + lim_{\alpha\to 0}\sum_{L=2}^{L_{max}}\alpha^{L-1} * N_l) \quad (17)$$

where N_1 is number of shortest-paths with length one (i.e. numbers of edges in the network). Since$lim_{\alpha\to 0\,\&\, l>2}\,\alpha^{L-1} = 0$, we have $lim_{\alpha\to 0}\sum_{L=2}^{L_{max}}\alpha^{L-1} * N_l) = 0$. Therefore, we have:

$$lim_{\alpha\to 0}\, E[W] = (2/N)\ (N_1) \quad (18)$$

$$lim_{\alpha\to 0}\, E[W] = \frac{2E}{N} = (N-1)\rho\alpha \quad (19)$$

where $\rho = E/\binom{n}{2}$ is the network density. Note that as mentioned in Theoreom.1, algorithms minimizing $(N-1)\rho\alpha$ are only required to minimize network density regardless of infection rate.

Remark.3. *Given a Graph (G = ⟨N, E⟩), infection rate (α), and number of immunization resources (b), immunization problem refers to the optimization problem of finding a subset of b nodes whose removal minimizes the average reachability of network that is formulated in Eq.20*

$$S_{opt} = argmin\ (\ AR(G/S\ ,\alpha)|\ S\ \subseteq\ N\ and\ |S|\ =\ b) \quad (20)$$

where S is a subset of b nodes, α is infection rate of the epidemic, G/S represent the residual network after removal of S, and $AR(G/S)$average reachability in G/S.

3 Stochastic Hill-Climbing Algorithm For Immunization Problem

Aspen et al. (2006a) proved that minimizing sum of square partition problem is reduced to a vertex cover problem which is NP-hard. Similarly, the problem of minimizing network density and average reachability can be proved to be NP-hard.

To solve these problems, we generalize previously proposed immunization algorithm, called SHCI, for minimizing the worst case epidemic-size (i.e. the largest connected component of network) (Shams and Khansari, 2013). SHCI first starts with a feasible random binary solution. The feasible solution is presented as a binary string (*S*) in which $S_i = 1$ if *i*th node is immunized and $S_i = 0$ otherwise. Trivially, the number of 1's should be equal to *b* where *b* is the number of immunization resources.

After initiating a random solution, we randomly select a node u among immunized group (i.e.,$\{i|S_i = 1\}$) and a node *v* from the other group (i.e.,$\{i|S_i = 1\}$). Then, fitness of solution (i.e., *F*(*S*,*G*,*α*)) is recalculated in case of switching *u* and *v*. If the new solution achieved higher fitness, the new values are accepted (See algorithm 1). Fitness of *S*(*F*(*G*,*S*, *α*)) can be defined in any form of Eq.21, Eq.22, and Eq.23

$$F_1(G,S,\alpha) = -\rho(G/\{i|s_i = 1\}) \quad (21)$$

$$F_2(G,S,\alpha) = -SSP(G/\{i|s_i = 1\}) \quad (22)$$

$$F_3(G,S,\alpha) = -AR(G/\{i|s_i = 1\},\alpha) \quad (23)$$

The immunization algorithms minimizing $F_1(G,S,\alpha)$,$F_2(G,S,\alpha)$, and $F_3(G,S,\alpha)$ are represented by

D-SHCI, SSP-SHCI and AR-SHCI, respectively.

Table 2 Stochastic hill-climbing immunization algorithm (Algorithm 1).

Input: G: a network , k: number of immunization resources, t: stopping time steps

generate a random candidate solution S from solution space, F:

Repeat

Choose a random bit u from immunized groups ($\{i|S_i = 1\}$)

Choose a random bit v from non-immunized groups ($\{i|S_i = 0\}$)

ΔF = the difference of fitness function $F(G, S, \alpha)$ if the values of u and v are switched

If ΔF >=0 switch values of u and v in S

Until F has not changed for t iteration

4 Experiments and Results

To evaluate effectiveness of AR and stochastic hill-climbing algorithm, we run SIR model on several artificial networks and calculate the average fraction of infected individuals *(W)* on immunized networks with immunization rate $V \in \{0.1,0.2, \dots ,0.9\}$.

Table 3 Simulation settings.

	Approach	**Parameters**	**Value**
Epidemic Model	SIR	Infection rate	$\alpha\epsilon\{0.2,0.5,0.8\}$
		Infection Time	T=1
Immunization algorithms	D-SHCI SSP-SHCI AR-SHCI	Immunization rate	$V \in \{0.1,0.2, \dots ,0.9\}$.
Dataset	Small-World (SW) Scale-Free (SF) Erdös-Renyi (ER)	Size	100
		Average degree	$< d > \epsilon\{2,4, \dots ,10\}$)

4.1 Simulation settings

To evaluate immunization algorithms under different condition of epidemic severity, we simulate our SIR model at various infection rates $\alpha\epsilon\{0.2,0.5,0.8\}$ indicating different severity levels (Shams and Khansari, 2015). We set infection Time equal to one in order to make emphasize on the impact of infection rate.

To simulate SIR model, we consider model presented by (Christley et al., 2005; Shams and Khansari, 2013, 2015). In SIR model, nodes are categorized as Susceptible (i.e. healthy but susceptible to subsequent infection),

Infection and Infective (i.e. host and spreader of the virus), and Recovered (i.e. healthy and immune to subsequent infection). Initially, all nodes are susceptible. First, a random node is infected to initiate the epidemic. At each time step, every infected node will infect their susceptible neighbors with probability α. Besides, an infected node will be recovered after T time steps. The procedure is iterated until there are no more infected nodes in the network (Christley et al., 2005; Shams and Khansari, 2013, 2015).

We generate several artificial networks in our experiments. These networks include Small-World (SW), Scale-Free (SF), and Erdös-Renyi (ER) networks of 100 nodes. To generate SW network, we use Watts-Strogatz algorithm (Watts and Strogatz, 1998) with switching probability of 0.01. SF networks are constructed based on the algorithm proposed in (Cho et al., 2009; Goh et al., 2001) with γ=2.5. Finally, ER networks are generated using $G(n,m)$ implementation introduced in (Erdos and Renyi, 1959). Since network epidemic reproduction rate R_0 is highly related to the average number of immediate contacts (i.e. degree) to infected individuals (Hartvigsen et al., 2007; Shams and Khansari, 2015), we study the performance of algorithms while varying average degrees ($i.e. < d > \epsilon\{2,4, \dots,10\}$).

Table 2 summarizes our simulation settings. Due to stochastic dynamics of network generation, immunization algorithms and epidemic model, we compute average fraction of infected individuals over 5 runs of SIR model on 5 runs of immunization algorithm and 5 runs of network generation. Fig.2, Fig.3, Fig.4 depicts the performance of immunization algorithms in Small-World, Scale-Free, and, Erdös-Renyi networks, respectively.

4.2 Results

We plot the fraction of infected individuals (W) over fraction of vaccinated nodes V. In Small-World (SW) networks, AR-SHCI obviously outperforms other immunization algorithms (Fig. 2). It shows improvement up to 45 % compared to SSP-SHCI and D-SHCI (e.g., Fig.2.O). SSP-SHCI performs. D-SHCI exhibits the worst performance in small-world network especially in case of high infection rate. That can be explained by homogenous degree distribution of network. More clearly, as all nodes have the same degree in small-world network, the fitness function of D-SHCI cannot differentiate the superiority of different solution, and, performs just like a random algorithm. This behavior has been also seen for degree immunization algorithms (Hu and Tang, 2012) that greedily minimize network density by removal of the nodes with higher degree (Shams and Khansari, 2015). On the other hand, SSP-SHCI performs as well as AR-SHCI in low average degree (e.g., Fig. 2 B), while, its performance is up to 35% lower than AR-SHCI in case of higher network density (e.g., Fig. 2 N).The reason is that SSP-SHCI minimizes sum of square partitions by fragmenting network to small component regardless of their internal connectivity. Therefore, it might result in highly dense components that are potentially fast epidemic spreading and high epidemic growth. The results emphasize on the fact that both network size and density have a great impact on epidemic growth. So, D-SHCI and SSP-SHCI that seek to focus on reduction of the network size or network density cannot individually minimize the epidemic growth.

Unlike Small-World networks, D-SHCI exhibits high performance in Scale-Free networks (Fig. 3). This result can be justified based on skewed and power-law degree distribution of Scale-Free networks (Barabási, 1999). Scale-free networks contain a small- number of hub nodes connecting a large number of low-degree nodes. Accordingly, hub nodes play critical roles in network connectivity. Therefore, D-SHCI results in minimal network density and connectivity by removing hub nodes. As shown in Fig. 3. Col1, AR-SHCI and D-SHCI show higher performance in case of α=0.2 due to its short epidemic period. In other words, such epidemics are suppressed in neighborhood of initial seeds of infection. So an efficient immunization algorithm should minimize immediate impact of node infection (i.e. node degree) just like D-SHCI. Additionally, AR-SHCI makes more significance on close nodes in case of small α (See Theorem.3).

Fig. 4 illustrates the performance of immunization algorithms on Erdös-Renyi (ER) networks. SSP-SHCI performs worse than other two immunization algorithms in case of α=0.2 and 0.5 (Fig. 4 Col1 and Col2). It can be explained in term of high percolation threshold of ER networks. In other word, a large fraction of nodes should be removed to fragment ER networks (Albert et al., 2000). Consequently, SSP-SHCI cannot reduce network epidemic growth by fragmentation. Therefore, an immunization algorithm is required to consider internal connectivity of component. Hence, AR-SHCIand D-SHCI could reduce network epidemic up to 25%

compared to SSP-SHCI (e.g., Fig.4 J). On the other hand, SSP-SHCI performs better than D-SHCI in case α=0.8 (Fig3. Col3). This can be justified by Theorem.2 which emphasize on the fact that epidemic growth is strongly related to size of network component when $\alpha \rightarrow 1$. Therefore, those immunization algorithms that minimize size of component (e.g. SSP-SHCI) outperform other immunization algorithms.

5 Conclusion

In this paper, we provided theoretical analysis of epidemic growth and immunization with regards to network properties. A new metric, called average reachability (AR), was proposed to give better estimation of network epidemic growth. AR incorporates epidemic infection rate which has not been addressed with previous measures. AR is a metric which estimate epidemic growth based on shortest paths among nodes. Though this approach ignores some available information about other paths, it is still reliable as infection probability exponentially decreases as paths get longer and so, the value of infection probability over shortest paths are dominant over that value over other paths.

Additionally, it should be noted that AR makes a trade-off between local connectivity and global reachability with regard to epidemic infection rate. AR emphasize on the local connectivity in case of weak epidemics in order to reduce immediate impact of infection. On the other hand, it considers network global reachability in case of severe epidemics which are disseminated to the whole networks rapidly. Theoretical analysis was provided to demonstrate generality of AR compared to other network metrics (i.e., density and sum of square partitions).

Furthermore, stochastic hill-climbing immunization was generalized to optimize all evaluation criteria. Immunization algorithm was called D-SHCI, SSP-SHCI and AR-SHCI regarding their objective function. Generally, experimental results show that SCHI algorithm could converge to the global optimum as its performance slightly over in different runs. Moreover, SIR simulation on immunized network showed that AR-SHCI exhibits the best performance in all networks and all infection rate. SSP-SHCI shows better performance in case of sever epidemics, while, D-SHCI shows better performance in case of weak epidemics. The high performance of AR-SHCI in Small-World networks suggests its high efficiency in modular networks. D-SHCI performs similar to AR-SHCI in Scale-Free networks. Therefore, D-SHCI is more recommended in these networks due to its lower time complexity.

For future work, average reachability can be used as an evaluation metric to compare efficiency of immunization algorithms in various networks and epidemics. Moreover, it is valuable to propose a new immunization algorithm for optimizing AR. Finally, introducing a more accurate measure regarding infection rate and other epidemic parameters such as infection time, could be a new research line.

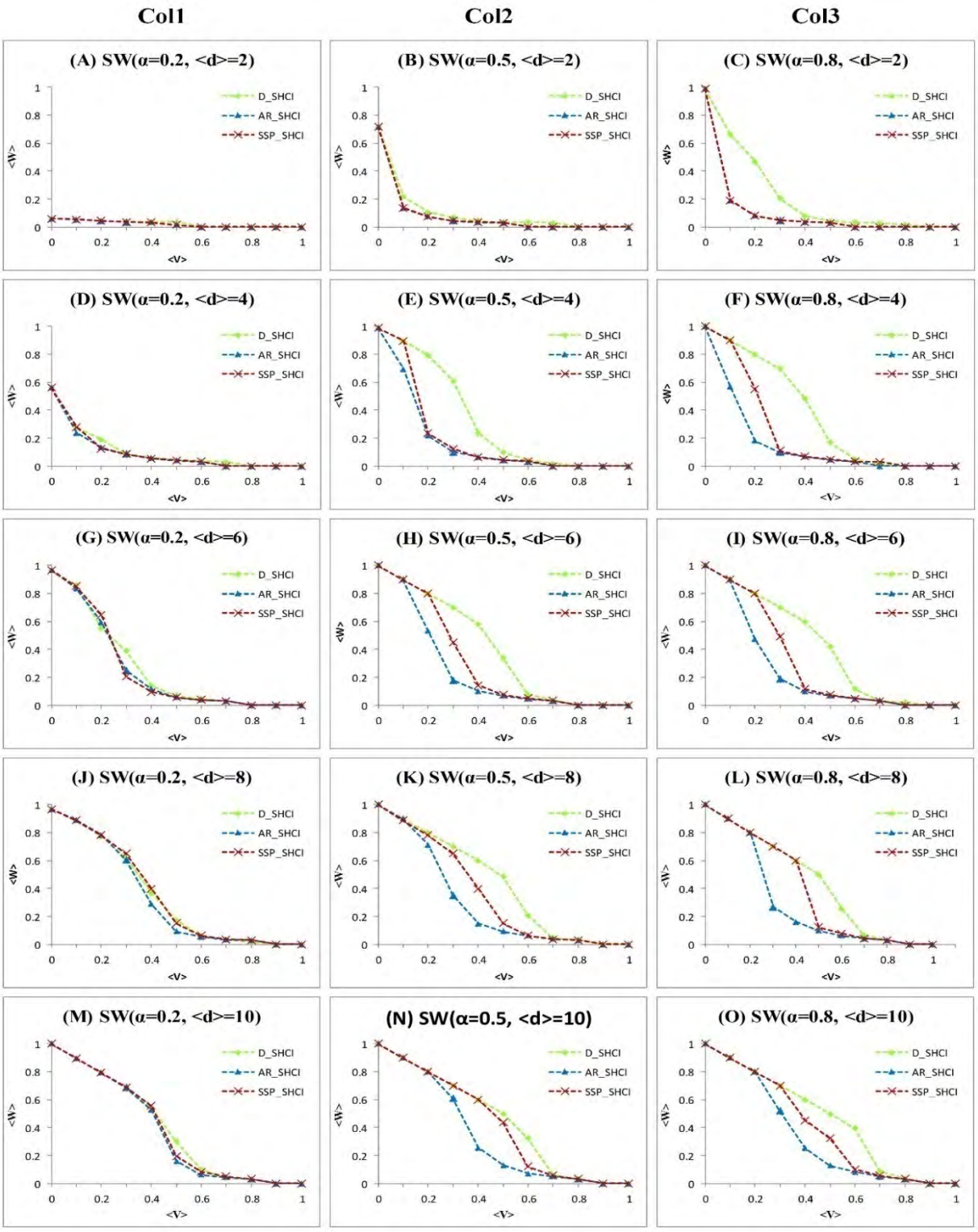

Fig. 2 Fraction of infected individuals (*W*) over fraction of immunized nodes (*V*) in Small-World networks.

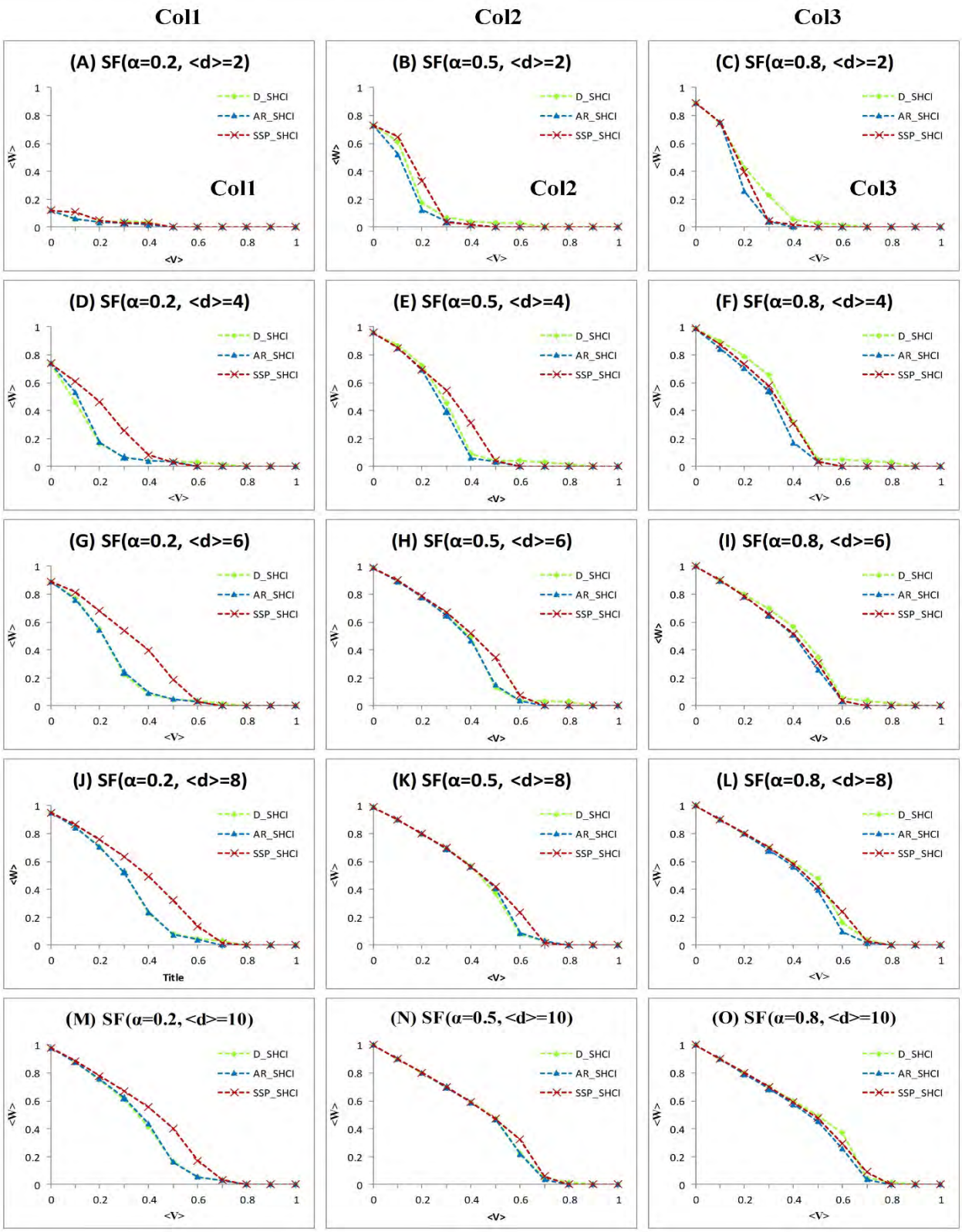

Fig. 3 Fraction of infected individuals (*W*) over fraction of immunized nodes (*V*) in Scale-Free networks.

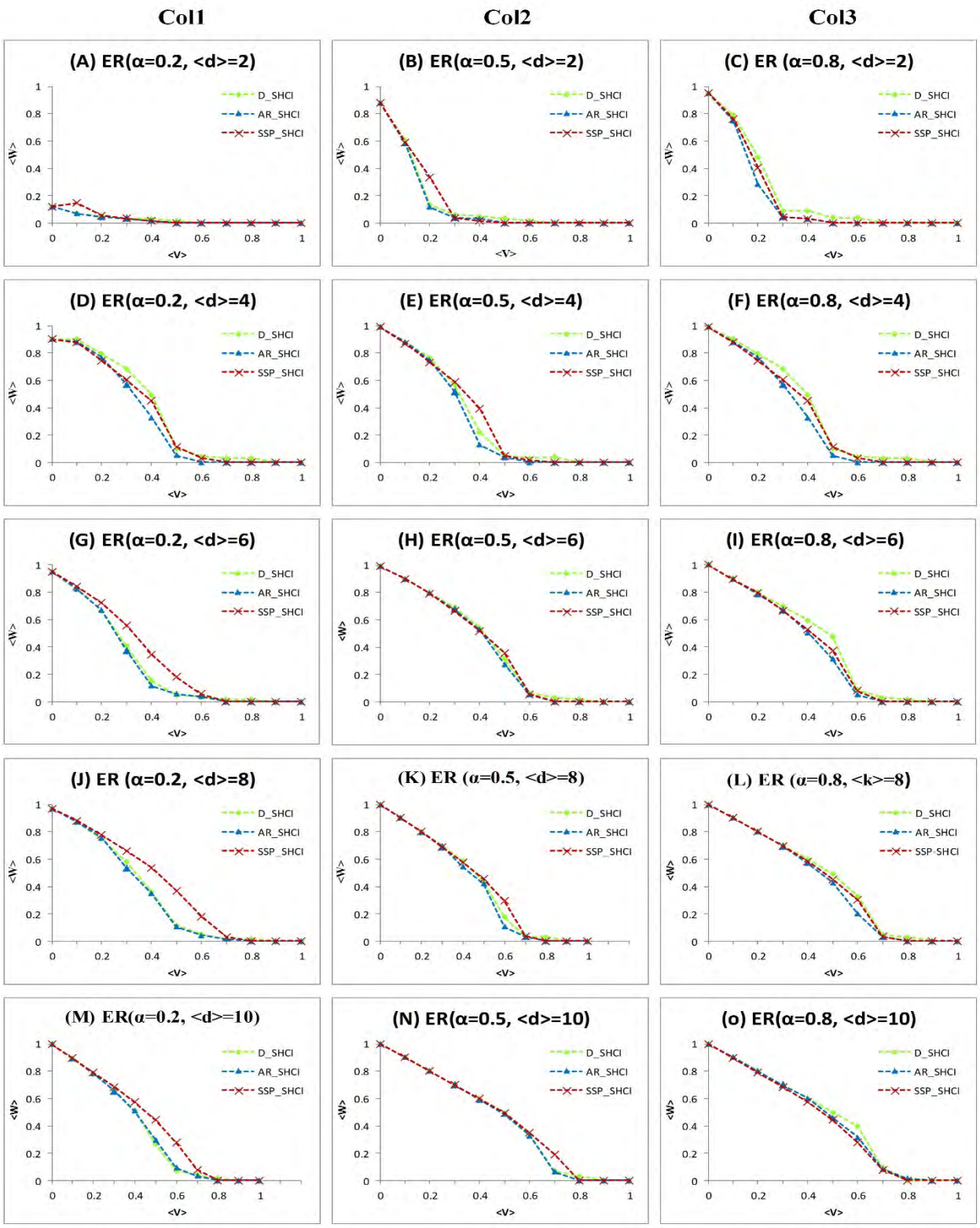

Fig. 4 Fraction of infected individuals (*W*) over fraction of immunized nodes (*V*) in Erdös-Renyi networks.

References

Albert R, Jeong H, Barabási A. 2000. Error and attack tolerance of complex networks. Nature, 406(6794): 378

Aspnes J, Chang K, Yampolskiy A. 2006a. Inoculation strategies for victims of viruses and the sum-of-squares partition problem. Journal of Computer and System Sciences, 72(6): 1077-1093

Barabási A. 1999. Emergence of scaling in random networks. Science, 286(5439): 509-512

Borgatti SP. 2006. Identifying sets of key players in a social network. Computational and Mathematical Organization Theory, 12(1): 21-34

Chai H, Liu Y, Lin S, Yu H, Zhu Z. 2011. A new immunization strategy for complex network. 2011 Fourth International Workshop on Chaos-Fractals Theories and Applications, 152-155

Chen Y, Paul G, Havlin S, Liljeros F, Stanley H. 2008. Finding a better immunization strategy. Physical Review Letters, 101(5): 06870

Cho YS, Kim JS, Park J, Kahng B, Kim D. 2009. Percolation transitions in scale-free networks under the achlioptas process. Physical Review Letters, 103(13): 135702

Christley RM, Pinchbeck GL, Bowers RG, Clancy D, French NP, Bennett R, Turner J. 2005. Infection in social networks: using network analysis to identify high-risk individuals. American Journal of Epidemiology, 162(10): 1024-1031

Cohen R, Havlin S, Ben-Avraham D. 2003. Efficient immunization strategies for computer networks and populations. Physical Review Letters, 91(24): 2-5

Danon L, Ford AP, House T, Jewell CP, Keeling MJ, Roberts GO, Ross JV, Vernon MC, 2011. Networks and the epidemiology of infectious disease. Interdisciplinary Perspectives on Infectious Diseases, 2011

Dezső Z, Barabási AL. 2002. Halting viruses in scale-free networks. Physical Review E, 65(5): 055103

Erdos P, Renyi,A. 1959. On random graphs. Publicationes Mathematica, 6: 290-297

Funk S, Salathé M, Jansen VA. 2010. Modelling the influence of human behaviour on the spread of infectious diseases: a review. Journal of the Royal Society Interface, 7(50): 1247-1256

Gallos L, Liljeros F, Argyrakis P, Bunde A, Havlin S. 2007. Improving immunization strategies. Physical Review E, 75(4): 045104

Gao C, Liu J, Zhong N. 2011. Network immunization with distributed autonomy-oriented entities. IEEE Transactions on Parallel and Distributed Systems, 22(7): 1222-1229

Goh KI, Kahng B, Kim D. 2001. Universal behavior of load distribution in scale-free networks. Physical Review Letters, 87(27): 278701

Hadidjojo J, Cheong SA. 2011. Equal graph partitioning on estimated infection network as an effective epidemic mitigation measure. PloS One, 6(7): e22124

Hartvigsen G, Dresch JM, Zielinski L, Macula J, Leary CC. 2007. Network structure, and vaccination strategy and effort interact to affect the dynamics of influenza epidemics. Journal of Theoretical Biology, 246(2): 205-213

Hébert-Dufresne L, Allard A, Young JG, Dubé LJ. 2013. Global efficiency of local immunization on complex networks Scientific Reports, 3: 2171

Holme P, Kim BJ, Yoon CN, Han SK., 2002. Attack vulnerability of complex networks. Physical Review E, 65(5): 056109

Hu K, Tang Y. 2012. Immunization for complex network based on the effective degree of vertex. International Journal of Modern Physics B, 26(6): 1250052

Masuda N. 2009. Immunization of networks with community structure. New Journal of Physics, 11(12): 123018

Miller JC, Hyman JM. 2007. Effective vaccination strategies for realistic social networks. Physica A:

Statistical Mechanics and Its Applications, 386(2): 780-785

Niu C, Li L, Xu D. 2009. A new network immunization strategy better than High Degree First. 2009 IEEE International Conference on Communications Technology and Applications, 821-825

Peng C, Jin X, Shi M. 2010. Epidemic threshold and immunization on generalized networks. Physica A: Statistical Mechanics and Its Applications, 389(3): 549-560

Restrepo J, Ott E, Hunt B. 2006. Characterizing the dynamical importance of network nodes and links. Physical Review Letters, 97(9): 4

Salathé M, Jones JH. 2010. Dynamics and control of diseases in networks with community structure. PLoS Computational Biology, 6(4): e1000736

Schneider CM, Mihaljev T, Havlin S, Herrmann HJ. 2011. Suppressing epidemics with a limited amount of immunization units. Physical Review E, 84(6): 061911

Schneider CM, Mihaljev T, Herrmann HJ. 2012. Inverse targeting —An effective immunization strategy. Europhysics Letters, 98(4): 46002

Shams B, Khansari M. 2013. Immunization of complex networks using stochastic hill-climbing algorithm. ICCKE 2013: 283-288

Shams B, Khansari M. 2014. Using network properties to evaluate targeted immunization algorithms. Network Biology, 4(3):74-94

Shams B, Khansari M. 2015. On the impact of epidemic severity on network immunization algorithms. Theoretical Population Biology, 106: 83-93

Ventresca M, Aleman D. 2013. Evaluation of strategies to mitigate contagion spread using social network characteristics. Social Networks, 35(1): 75-88

Watts D J, Strogatz SH. 1998. Collective dynamics of “small-world” networks. Nature, 393(6684): 440-442

Yamada Y, Yoshida T. 2012. A comparative study of community structure based node scores for network immunization. Active Media Technology, 328-337

Yoshida T, Yamada Y. 2012. Community structure based node scores for network immunization. PRICAI 2012: Trends in Artificial Intelligence, 7458: 899-902

Zhang WJ. 2012. Computational Ecology: Graphs, Networks and Agent-based Modeling. World Scientific, Singapore

Zhang WJ. 2016a. Detecting connectedness of network: A Matlab program and application in tumor pathways and a phylogenic network. Selforganizology, 3(4): 117-120

Zhang WJ. 2016b. Screening node attributes that significantly influence node centrality in the network. Selforganizology, 3(3): 75-86

Zhang WJ. 2018. Fundamentals of Network Biology. World Scientific Europe, London, UK

Zhang WJ, Zhan CY. 2011. An algorithm for calculation of degree distribution and detection of network type: with application in food webs. Network Biology, 1(3-4): 159-170