Self-avoiding pruning random walk on signed network

Wang, Huijuan; Qu, Cunquan; Jiao, Chongze; Ruszel, Wioletta

DOI
10.1088/1367-2630/ab060f

Publication date
2019

Document Version
Final published version

Published in
New Journal of Physics

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG IOP Institute of Physics

PAPER • OPEN ACCESS

Self-avoiding pruning random walk on signed network

To cite this article: Huijuan Wang *et al* 2019 *New J. Phys.* **21** 035001

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
5 November 2018

REVISED
4 January 2019

ACCEPTED FOR PUBLICATION
11 February 2019

PUBLISHED
15 March 2019

PAPER

Self-avoiding pruning random walk on signed network

Huijuan Wang[1], **Cunquan Qu**[1,2], **Chongze Jiao**[1] **and Wioletta Rusze**[1]

[1] Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, Mekelweg 4, Delft, 2628 CD, The Netherlands
[2] School of Mathematics, Shandong University, Jinan, 250100, People's Republic of China

E-mail: c.qu-1@tudelft.nl (**Cunquan Qu**)

Keywords: random walk, signed network, self-avoiding walk

Abstract

A signed network represents how a set of nodes are connected by two logically contradictory types of links: positive and negative links. In a signed products network, two products can be complementary (purchased together) or substitutable (purchased instead of each other). Such contradictory types of links may play dramatically different roles in the spreading process of information, opinion, behaviour etc. In this work, we propose a self-avoiding pruning (SAP) random walk on a signed network to model e.g. a user's purchase activity on a signed products network. A SAP walk starts at a random node. At each step, the walker moves to a positive neighbour that is randomly selected, the previously visited node is removed and each of its negative neighbours are removed independently with a pruning probability r. We explored both analytically and numerically how signed network topological features influence the key performance of a SAP walk: the evolution of the pruned network resulted from the node removals, the length of a SAP walk and the visiting probability of each node. These findings in signed network models are further verified in two real-world signed networks. Our findings may inspire the design of recommender systems regarding how recommendations and competitions may influence consumers' purchases and products' popularity.

1. Introduction

The concept of multi-layer networks has been proposed in 2010 [1–5] to capture different types of relationships/links among the same set of nodes. For example, the rapid development of the Internet, smart phones and information technology has facilitated the boost of online platforms, such as Facebook and YouTube, for communications, creating and sharing information and knowledge. Users may participate in one or several online networks besides their physical contacts forming a multi-layer network where the nodes represent the users and the links in each layer represent a specific type of connections such as physical contacts and online follower-followee relationships. Such multi-layer networks support the spreading of e.g. information, behavioural patterns, opinions, fashion within each layer respectively and allow as well these spreading processes on different layers to interact, introducing new phenomena that dramatically differ from a single spreading process on a single network [6–19].

Signed networks is a special type of two-layer networks where the same set of nodes are connected by two logically contradictory types of links, so called positive and negative links. The positive and negative links may represent friendly and antagonistic interactions respectively in a signed social network [20] and represent the complementary (i.e. when a product e.g. a phone is purchased, the other product e.g. a phone charger is likely to be bought in addition) and substitutable (two products can be purchased instead of each other such as the phones from two competing brands) relationships respectively in a signed network of products [21–23].

Whereas all types of links in most multi-layer networks such as physical contact and online friendships are mostly positive thus facilitate the spread of information, opinion and etc the positive and negsative links in a signed network usually play dramatically different roles in a general spreading process. The random walk (RW) and self-avoiding walk have been used to model users' purchase activity on a recommendation network of

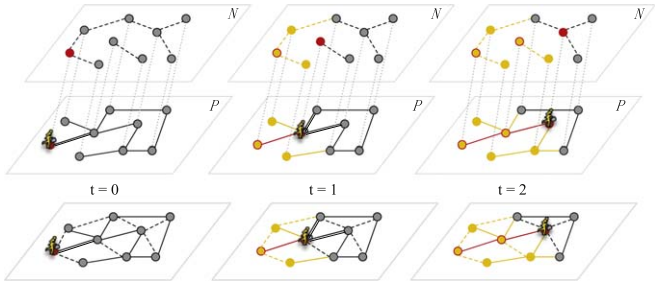


Figure 1. Schematic plot of SAP walk on a signed network with pruning probability $r = 1$. The signed network is represented on the top as a two-layer network a negative layer and a positive layer where dashed lines between the two layers emphasis that the nodes are the same individual across layers and represented at the bottom as a single network with two types of links: positive (solid lines) and negative (dotted lines). At $t = 0$, a walker visits a random node in the network, which is in red. At step 1 or $t = 1$, the walker moves to a random positive neighbour (in red) and the previously visited node and its negative neighbours (in yellow) are removed. Such steps repeat until the walker has no node to visit any more. The pruned network in grey is shrinking over time.

products where two products are connected if when a user is purchasing one product the other product is recommended[3] by the online retail platform like Amazon [24–26]. However, these models have not considered the substitute relations between products and the fact that once a product has been purchased, its substitutable products will be unlikely to be purchased afterwards. Hence, we propose in this work a self-avoiding pruning (SAP) walk on a signed network to model, e.g. a user's purchase behaviour on a signed network of products. As shown in figure 1, a SAP walk starts at a random node in a signed network at $t = 0$. At each step, the walker moves from its current location node i to a positive neighbour[4] j that is randomly selected, its previous location, i.e. node i is removed from the signed network[5] and each of node i's negative neighbours is removed independently with probability r. The walker repeats such steps until there is no new location to move to. Since each node pair can be connected by either a positive or negative link, but not both, the walker could equivalently, at each step, remove each negative neighbour of its current visiting node i independently with probability r, then move to a random positive neighbour j and afterwards remove the previously visited node i.

In the context of a signed product network, a SAP walk may model the purchase trajectory of a user on the network of products: initially, the user purchases a random product and afterwards buys a random complementary product of his/her previous purchase; however, the user will not buy the same product repetitively and unlikely buy the substitutable products of what he/she has bought. If the negative layer is empty, our model is equivalent to the self-avoiding walk model of user purchases on the recommendation network of products i.e. the positive network layer [24–26].

Consider the scenario of a signed social network. A SAP walk could model the trajectory that a walker (individual or company) recruits team members or collaborators aiming to recruit members with friendly relations.

In this paper, we aim to understand how features of a signed network influence: (1) the evolution of the pruned network topology resulted from the node removals in a SAP walk (2) the length/hopcount of a SAP walk, i.e. the number of positive links that a SAP walker traverses in total and (3) the visiting probability of each node by a SAP walk. Taking the product network as an example, we are actually going to explore how the initial signed product network features influence (1) the sub-network a user may further explore after several purchase actions and (2) the probability that a product is purchased and (3) how many purchases a user may perform in total in a SAP walk. The signed product network thus may affect both the user purchase behaviour and the popularity or market share of the products. These questions will be firstly explored on simple signed network models and afterwards on two real-world signed networks.

Our observations and analysis point out the significant influence of the negative/substitutable links on the number of purchases of a user, the distribution of the popularity of a product etc. These findings may inspire the design of future recommender systems: which complementary product(s) should be recommended in order to maximise the total number of purchases of a walker? With which products a product should not compete in order to maximise the total purchases of the product? How competition between products may affect market share of the products and users' purchases?

The SAP walk, as a starting model of users' purchase activities can be improved or extended from multi-perspectives, especially as more rich data becomes available. Section 6 illustrates one possible generalisation of

[3] If two products are frequently co-purchased by the same users, one product will be recommended when a user is buying the other product.

[4] A positive (negative) neighbour of a node is a neighbour that is connected with this node via a positive (negative) link.

[5] When a node is removed from the network, all links connected to the node, including both positive and negative links, are also removed.

the SAP model where the complementary products of a product are recommended or preferred with different strength.

The paper is organised as follows: we introduce the basic definitions related to signed network models and RWs in section 2. The influence of the signed network features on the aforementioned properties are studied in signed network models when the pruning probability is $r = 1$ in section 4, and when $r \neq 1$ in section 5. Our observations and understanding obtained in signed network models are further verified in two real-world signed networks in section 7. We summarise our findings and discuss promising future work is in section 8.

2. Definitions

In this section, we introduce basic definitions regarding to signed network representation and models, different types of RWs [26] and their relation to the SAP walk.

2.1. Signed network representation

In a signed network with N nodes, two $N \times N$ adjacency matrices A^+ and A^- can be used to represent the positive and negative connections respectively. Element $A^+_{i,j} = 1$ ($A^-_{i,j} = 1$) if node i and j are connected via a positive (negative) link. Or else $A^+_{i,j} = 0$ ($A^-_{i,j} = 0$). A signed network can be as well represented by a single adjacency matrix $A = A^+ - A^-$. Hence, each element

$$A_{i,j} = \begin{cases} 1, \text{ if } i \text{ and } j \text{ are connected via a positive link} \\ -1, \text{ if } i \text{ and } j \text{ are connected via a negative link} \\ 0, \text{ if } i \text{ and } j \text{ are not connected} \end{cases}$$

The positive degree $d^+_i = \sum_{j=1}^{N} A^+_{i,j}$ of a node i counts the number of positive neighbours of node i, whereas the negative degree $d^-_i = \sum_{j=1}^{N} A^-_{i,j}$ indicates the number of negative links incident to a node i. As shown in section 7, both the positive and negative degree of a node in real-world signed networks tend to follow a power law distribution. Since each node pair can be connected only by one type of links, positive or negative, but not both, $A^-_{i,j}A^+_{i,j} = 0$. An example of signed network is shown in figure 1, which is plotted both as a two-layer network (above) and a single network with two types of links (bottom).

2.2. Signed network models

The simplest signed networks can be constructed by generating the positive layer and negative layer independently from the same network model or two different network models respectively, such as the Erdős–Rényi ER and scale-free SF random network model.

Erdős–Rényi ER random network is one of the most studied random network models that allow many problems to be treated analytically [27, 28]. To generate an Erdős–Rényi random network with N nodes and average degree $E[D]$, we start with N nodes and place each link between two nodes that are chosen at random among the N nodes until a total number $L = \frac{NE[D]}{2}$ of links have been placed. All the links are bidirectional. In this paper, we choose $N = 1000$, $E[D] = 4$ for Erdős–Rényi random networks. Erdős–Rényi random networks are characterised by a Poisson degree distribution, $Pr[D = k] = \frac{(Np)^k e^{-Np}}{k!}$, where D is the degree of a random node in the network, and the link density $p = \frac{E[D]}{N-1}$.

We use the hidden parameter model [29–32] to generate scale-free networks which have a power-law degree distribution $Pr[D = k] = ck^{-\lambda}$ as observed in many real-world networks [33–35]. The hidden parameter model is considered because the degree distribution and the average degree of the generated scale-free networks are both controllable. We start with N isolated nodes and assign each node i a hidden parameter $\eta_i = \frac{1}{i^\alpha}$, $i = 1, 2, ..., N$. At each step, two nodes i and j are chosen randomly with a probability proportional to η_i and η_j and they are connected as a link if they were not connected previously. Such steps are repeated until $L = \frac{E[D]N}{2}$ links have been added. In this case, the generated random network has a power-law degree distribution $Pr[D = k] = ck^{-(1+\frac{1}{\alpha})}$. In this paper, we consider $N = 1000$, average degree $E[D] = 4$ and $\lambda = 3$, such that the ER and SF networks have the same average degree and a size of the largest connected component close to N.

The positive and negative degree of a node are possibly correlated, actually positively correlated as shown in the real-world signed networks in section 7. Moreover, triangles with an odd number of positive links, so called balanced triangles, have been shown to appear more frequently than the other types of signed triangles [36].

We focus on the simplest signed networks where the positive and negative connections are generated independently from either the same or different network models, i.e. ER or SF model. In this case, the positive and negative degree of a node are uncorrelated. We construct four types of signed networks: ER–ER, ER–SF, SF–SF and SF–ER, where $N = 1000$ and the average degree of both layers are 4. Moreover, we consider as well

ER–ER networks where $E[D^+] = 4$ and $E[D^-] = 0,\ 4$ and 8 to explore the influence of the density of the negative layer on SAP walks.

We explored as well the signed networks ER–ER and SF–SF where the positive and negative degree of a node are positively correlated with linear correlation coefficient $0 \leqslant \rho \leqslant 1$, $E[D^+] = E[D^-] = 4$ and $N = 1000$. Such networks are generated as follows. First, an ER (or SF) network is generated as the positive network layer. Second, set the negative degree of each node the same as its positive degree. Third, select randomly a fraction $1 - \rho$ of the nodes and shuffle randomly their negative degrees. After the shuffling, the generated degree sequences for the two layers are correlated with linear correlation coefficient ρ [37, 38]. Given the negative degree of each node, construct the negative network layer according to the configuration model [39].

3. Related work

Classic RW starts at a random node in an unsigned network. At each step, the walker moves from its current location node i to a neighbour that is selected uniformly at random. In this process, the walker can visit any node repeatedly if the network is connected. RW has been widely applied e.g. to model network routing protocol, users' visit at websites via hyper links and to detect network topology [40–44]. The self-avoiding random walk (SAW) is the same as the RW except that at each step the walker moves to a random neighbour that has not yet been visited. Hence, each node can be visited maximally once. A SAW stops when the walker has no further node to visit any more. The SAW was first introduced by chemist Froly to study the behaviour of polymers on lattice graph [45]. SAW has also been applied to detect protein–protein interaction [46], to detect network structure which is more efficient than classic RW by avoiding previously visited nodes in each step [47], and to detect unidentified network traffic [48].

Performance of these two types of RWs has been analytically studied [49]. The probability that a node is visited by a classic random walker has been shown to be proportional to the degree of that node. The path length of a SAW is the number of links that have been traversed in total in a SAW. The path length of SAW has been studied, especially regarding to the average and the probability distribution [50]. Tishby *et al* have found that the path length of SAW on an Erdős–Rényi random network follows the Gompertz distribution in the tail [51].

RW and SAW have been used to model users' purchase activity on a recommendation network [24, 25]. In contrast to RW and SAW, SAP walk addresses further that products can be substitutable to each other and are seldom or not purchased by the same user.

Jung *et al* considered signed RW, where the sign of the walker changes depending on the signs of the links that walker has traversed [52]. This work addresses, for the first time, that the signed links could influence the dynamics of the walk thus the walkers' trajectories. The SAP walk is equivalent to a self-avoiding walk on the positive network layer if the negative network layer is empty, i.e. no negative links exist.

Opinion diffusion (voter model) on a signed network has been proposed in [53]. Dynamics of influence diffusion and influence maximisation problem on signed networks have been explored [53] beyond the influence maximisation problem on single unsigned networks [54]. Viral spreading processes on signed networks have been studied in [55]. From an analytic point of view, SAP walk is more challenging to trace because the earlier trajectory of a walker influences its future moves, in contrast to spreading models where state transitions of each node depend only on the current states of neighbours and the local dynamic rule.

4. Influence of signed network topology on SAP walks when $r = 1$

4.1. Evolution of the pruned network structure

The SAP walk on a signed network is more complex than previous RW models. At each step, the walker moves to a random positive neighbour, and afterwards, not only the previously visited node but also its negative neighbours are removed/pruned from the signed network. As shown in figure 1, the pruned signed network (in grey) $G(t)$ is shrinking over time. The pruned signed network at a step t refers to the remaining signed network after the removal of nodes at step t. The initial signed network corresponds to $G(0)$.

The pruned positive network layer $G^+(t)$ suggests the potential sub-graph of the original signed network that the walker could further explore via a SAP walk. In this section, we will explore how the topology, especially the average degree, of the pruned positive network layer is changing over time. We start with the simpler case when both the initial positive and negative layers are ER networks, with possibly different average degree.

Firstly, we examine the case when the initial negative layer is an empty graph, i.e. the average degree $E[D^-(t = 0) = 0]$ initially is zero. The initial positive network layer possesses the binomial degree distribution, which approaches the Poisson distribution $Pr[D^+(0) = k] = \frac{c(0)^k}{k!}e^{-c(0)}$, under the condition that the network is sparse, i.e. the average degree $c(0) = E[D^+(0)]$ of the initial positive layer at $t = 0$ is a constant. The SAP walk

on such a signed network is equivalent to a SAW on the positive network layer $G^+(0)$. Initially, the network has $N(t=0)=N$ nodes. At any step t, the pruned positive network layer has $N-t$ nodes.

An insightful observation of a SAW walk on an ER random graph in [51] is as follows. A SAW walker has a higher probability to visit a neighbour with a higher degree. Take the step $t=1$ as an example. Starting from the random node that is visited at step $t=0$, the walker walks to a node with degree k in $G^+(0)$ at step $t=1$ with probability $kPr[D^+(0)=k]/c(0)$. A special property of the Poisson distribution is that $kPr[D^+(0)=k]/c(0)=Pr[D^+(0)=k-1]$. The probability to walk to a node with degree k in $G^+(0)$, thus with degree $k-1$ in $G^+(1)$ after the removal of the previously visited node and its links is $Pr[D^+(0)=k-1]$. The node to visit at $t=1$ is as if chosen randomly from $G^+(1)$.

Note that when a randomly selected node together with its links are removed from an ER network, the remaining network is again an ER network with the same link density, i.e. the probability that two nodes are connected. Hence, the network pruning resulting from a SAP walk on an ER positive network with an empty negative layer is statistically equivalent to the node removal process upon the initial ER network where at each step, a node is randomly selected and this node together with its links are removed from the network. The pruned positive network $G^+(t)$ at any step t is thus an ER network with $N-t$ nodes and link density $E[D^+(0)]/(N-1)$. The average degree[6] of the pruned positive network at step t is

$$E[D^+(t)]=\frac{E[D^+(0)]}{N-1}(N-t-1) \tag{1}$$

when the original signed network is sparse and the size N is large.

Furthermore, we consider the case where the negative network layer $G^-(0)$ is not empty but an ER network with its average degree $E[D^-(0)]>0$. At any step t of a SAP walk, the network is pruned by the removal of the previously visited node and its negative neighbours. Since the negative and positive ER networks are generated independently, the node to visit at each step is as if chosen randomly from the negative layer $G^-(t)$. From the view of the negative network layer, a random node and its negative neighbours together with all their negative links are removed from the negative layer at each step. The negative network layer $G^-(t)$ remains approximately an ER network with the same link density over time. This is an approximation because the neighbour of a random node tends to have a higher degree. The link density remains approximately the same $p^-\approx E[D^-(0)]/(N-1)$ over time, whereas at each step t on average, $N(t-1)-N(t)=1+(N(t-1)-1)p^-$ nodes are removed. The size $N(t)$ refers to the average size of the pruned network at step t over a large number realisations of the stochastic SAP walks. Hence, the size of the pruned negative layer, which is as well the size of the pruned positive layer follows

$$N(t)\approx\left(N-1+\frac{1}{p^-}\right)(1-p^-)^t-\frac{1}{p^-}+1. \tag{2}$$

From the prospective of the positive layer, at step t, the negative neighbours of the previously visited node are as if chosen randomly from positive layer $G^+(t-1)$. The positive layer remains as an ER network with the same link density $p^+=E[D^+(0)]/(N-1)$. The average degree of the pruned positive network layer at time t is

$$E[D^+(t)]=(N(t)-1)p^+\approx\left(N-1+\frac{1}{p^-}\right)p^+(1-p^-)^t-\frac{p^+}{p^-}. \tag{3}$$

When the original signed network is sparse and the size N is large

$$E[D^+(t)]\simeq\left(N+\frac{1}{p^-}\right)p^+(1-p^-)^t-\frac{p^+}{p^-}. \tag{4}$$

As shown in figure 2, the average degree $E[D^+(t)]$ of the positive pruned network as a function of the SAP walk step can be well approximated by our theoretical result equation (3), when the initial network is an ER–ER signed network. The average positive degree $E[D^+(t)]$ of the pruned network at each step t is proportional to the size of the pruned network as shown in equation (3). A denser negative layer with a large $E[D^-(0)]$, could more effectively prune the network, leading to a fast decrease of the average degree of the pruned positive layer.

ER(positive)-SF(negative) signed networks are pruned slightly less than ER–ER networks when both layers have the same average degree 4 (see figure 2). This can be explained as follows. If a visited node has a large negative degree, its removal will lead to the removal of many nodes, its negative neighbours. If a negative neighbour of a visited node has a large negative degree, however, the removal of such a negative neighbour together with its negative links will not remove extra nodes but makes the negative layer sparser, protecting the network from the pruning. In ER–SF networks, the visited nodes are as if randomly chosen from the negative layer, thus tend to have a low negative degree. Nodes with a high negative degree in the SF negative layer are likely

[6] The average degree of the pruned positive or negative network layer at a step t refers to the average degree at step t over all the nodes and over a large number of SAP walk realisations.

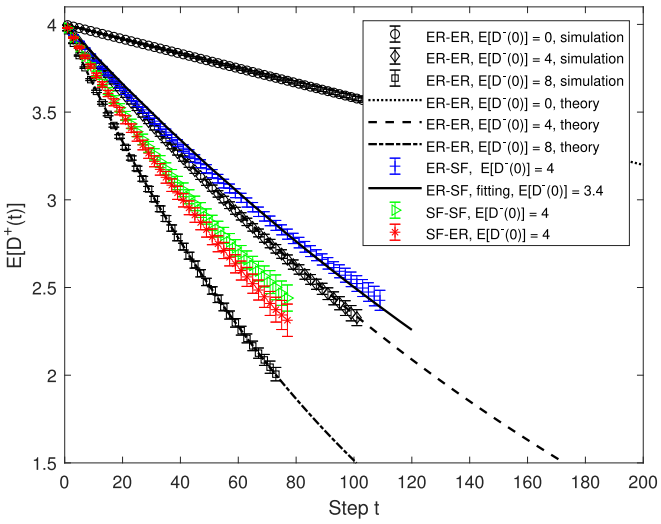


Figure 2. The average degree $E[D^+(t)]$ of the positive pruned network as a function of the step t of a SAP walk on a signed network with $N = 1000$ nodes. We simulate 100 independent realisations of a SAP walk on each of the independently generated 100 signed networks and plot the average and standard deviation (error bar) of $E[D^+(t)]$. The average degree of the initial positive layer is always $E[D^+(0)] = 4$. The SF network has a power exponent 3. The best curve fitting for ER–SF networks by theory equation (3) is obtained when $E[D^-(0)] = 3.4$.

to be removed as the negative neighbour of a visited node, which reduces the pruning. We expect that the SAP walks on ER–SF networks whose average degree in the negative layer is $E[D^-(0)] = 4$ perform similarly to the SAP walks on ER–ER networks but with a lower average degree $E[D^-(0)]$ in the negative layer. Hence, we fit the average degree $E[D^+(t)]$ of the positive pruned network as a function of the SAP walk step in ER–SF networks by our theory (3) for ER–ER networks. Figure 2 shows the optimal fit when $E[D^-(0)] = 3.4$. If we remove the top 2% of the nodes with the highest degree from the SF negative network layer whose original average degree is 4, the resultant average degree becomes 3.4.

Figure 2 shows that the pruned positive network, e.g. $E[D^+(t)]$, shrinks faster if the initial network is a SF–SF signed network than ER–SF signed network. This is mainly due to the fact that, a node with a large positive degree is likely to be visited in early steps and removed, significantly reducing the average degree of the positive pruned layer. However, the negative neighbours of a visited node are as if chosen randomly in the positive layer and tend to have a lower positive degree in a SF–SF network than in ER–SF network, slightly reducing the pruning effect.

Hence, a SF or in general a heterogeneous positive layer and a dense negative layer tend to facilitate the pruning of the network whereas a SF (heterogeneous) negative layer reduces the pruning effect.

4.2. Length of a SAP walk

The length or hopcount H of a SAP walk counts the total number of positive links, or the total number of move steps, a SAP walker traverses until it has no other node to move to. In the context of a signed produce network, $H + 1$ suggests the total number of purchases of a consumer. Signed networks leading to a large $H + 1$ promotes the purchases of more products. We would like to understand how the original signed network topology influences the length of a SAP walk.

The probability distribution of the length H of a SAP walk is shown in figure 3(a), for various types of signed networks. Intuitively, a SAP walk stops when the walker has no other node to move to, which is likely to happen if the current pruned positive network layer is less connected in the sense that no giant connected component exists but only small connected clusters exist. Hence, a dense initial negative network layer $G^-(0)$ leads to the removal of many nodes in each step and effectively reduces the connectivity of the positive layer, resulting in a small length H. This explains our observation in ER–ER signed network (figure 3(a)) that a dense $G^-(0)$ leads to a small length H on average.

The distribution of the length H of a SAP walk on an ER–ER network can be analytically derived. A SAP walk has a length $H = h$ requires that the node that the walker visits at step h has degree 0 in the pruned positive layer $G^+(h)$ and each node visited in a previous step t where $0 \leqslant t < h$, has a positive degree in the corresponding pruned positive network layer larger than 0. As discussed in section 4.2, the pruned positive layer remains an ER network with the same link density p^+ but with a shrinking size $N(t)$. Hence

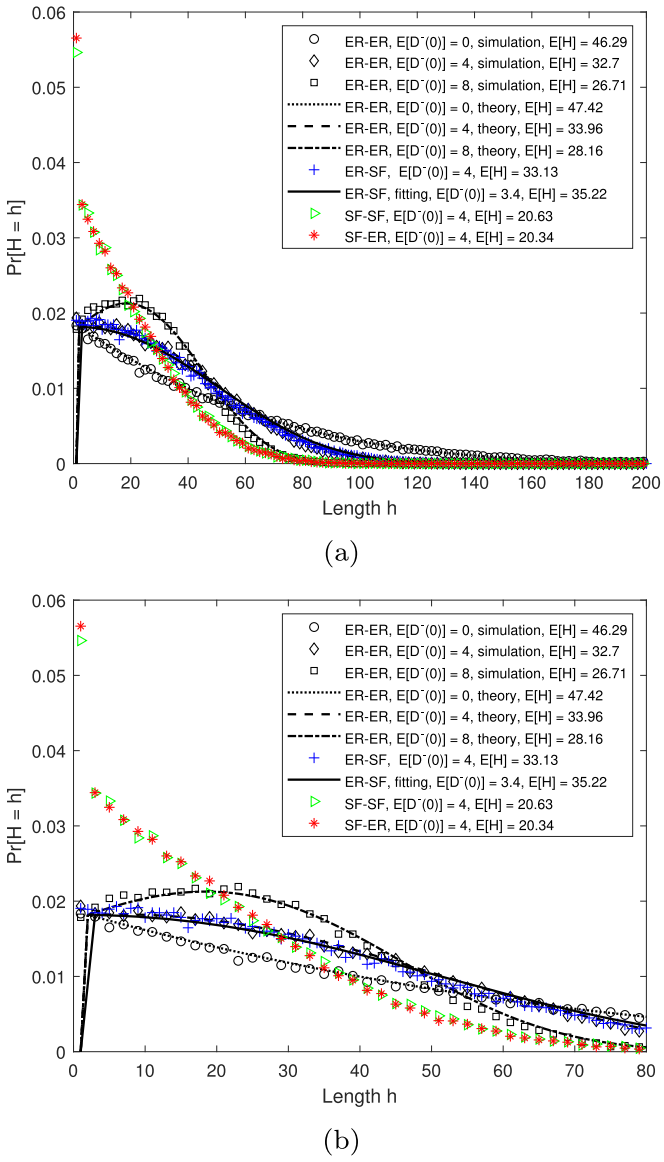


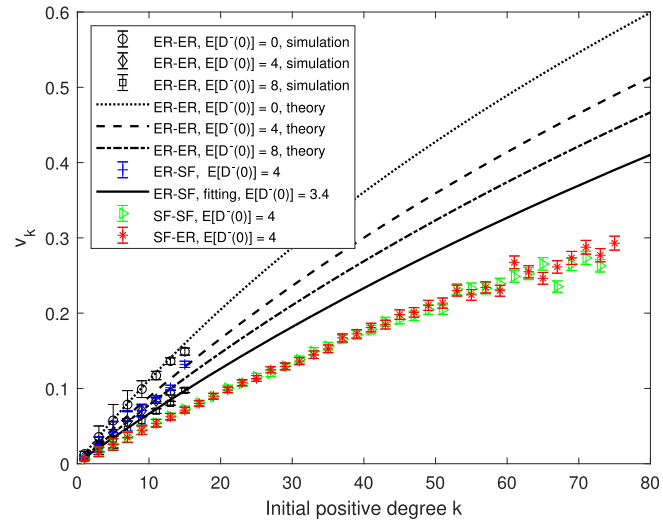
Figure 3. Probability distribution of the length of a SAP walk on a signed network.

$$Pr[H = h] \approx (1 - p^{+})^{N(h)-1} \prod_{t=0}^{h-1} [1 - (1 - p^{+})^{N(t)-1}]. \tag{5}$$

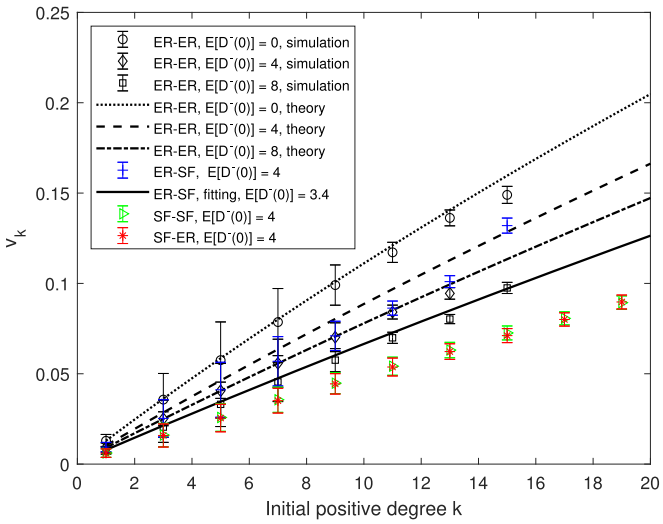
Figure 3(a) shows that our theory of the distribution of the length equation (5) well approximates the simulation results. The distribution of the length of a SAP walk on an ER–SF network can be well approximated by (5), our theory on an ER–ER network with $E[D^{-}(0) = 3.4]$.

Which type of signed networks tend to lead to a long length of a SAP walk? If we look at the average path length, the ordering of signed networks from the highest to the lowest follows: ER–ER($E[D^{-}(0) = 0]$), ER−SF, ER–ER($E[D^{-}(0) = 4]$), ER–ER($E[D^{-}(0) = 8]$), SF−SF, SF–ER($E[D^{-}(0) = 4]$). This ordering is consistent with our previous explanation: a heterogeneous positive layer such as SF network and a dense negative layer facilitate the pruning of the network leading to a short length of a SAP walk whereas a heterogeneous e.g. SF negative layer reduces the pruning effect attributing to a long length of a SAP walk.

The length of a SAP walk actually depends on not only the link density but as well the connectivity of the pruned positive layer. The negative neighbours of a visited node are as if chosen randomly in the positive layer since the two layers are independent in connections. Removal of such random nodes in the positive layer reduces less the connectivity of the positive layer if the original positive layer is a SF network since SF networks are robust against random node removals compared to ER networks. However, hubs in the positive layer are more likely to be visited and removed reducing more significantly the density of the SF positive layer.



(a)



(b)

Figure 4. The probability v_k that a node is visited by a SAP walk versus its degree $D^+(0) = k$ in the initial positive network is plotted for all possible degree values k in (a) and for small k in (b). Since many nodes may have the same initial positive degree $D^+(0) = k$, we plot their average visiting probability and the standard deviation.

4.3. Nodal visiting probability

The probability that a node is visited by a SAP walk implies a certain kind of importance of the node, e.g. the probability that a product is purchased when the signed network represents the network of products. Intuitively, a node with a higher positive degree in the initial signed network has a higher chance to be visited by a SAP walk. Hence, we examine the visiting probability of a node given its initial positive degree, which is shown in figure 4.

Firstly, we analytically derive the nodal visiting probability in ER–ER networks. Specifically, we compute the probability v_k that a random node j with degree $d_j^+(0) = k$ in the initial positive layer is visited by a SAP walk starting at a random node. We denote X_t as the node that is visited by a SAP walk at step t. Since the node j can be visited at any step $0 \leqslant t \leqslant h$ by a SAP walk of length h, we have

$$v_k = \sum_{h=0}^{N(0)-1} \sum_{t=0}^{h} Pr[X_t = j|H = h]Pr[H = h] \approx \sum_{t=0}^{N(0)-1} Pr[X_t = j|H \geqslant t]Pr[H \geqslant t], \quad (6)$$

assuming that the probability node j is visited at any step t is independent of the length H of the walk as long as $H \geqslant t$. The node j is visited at step t by a SAP walk that has a length $H \geqslant t$ requires that node j is not visited nor removed in the previous steps and j is connected with the node X_{t-1} visited at step $t - 1$. Hence

$$Pr[X_t = j|H \geqslant t] \approx \prod_{t'=0}^{t-1} (1 - Pr[X_{t'} = j]) \cdot \frac{\Delta_j^+(t-1)}{N(t-1)-1} \cdot \frac{1}{p^+(N(t-1)-1)}, \tag{7}$$

where $Pr[X_0 = j] = 1/N(0)$, $\Delta_j^+(t-1)$ is the degree of node j at step $t-1$ in the pruned positive network layer given that j is not visited in the first $t-1$ steps. The node X_t to be visited at step t as well as its negative neighbours are as if randomly chosen from the pruned positive layer G_{t-1}^+. The pruned network remains approximately (precisely if the negative layer is empty) an ER network with the same link density p^+ when the node visited and its negative neighbours are removed at each step. The ratio $\frac{\Delta_j^+(t-1)}{N(t-1)-1}$ is the probability that node j is connected with the node X_{t-1} visited in the previous step and $\frac{1}{p^+(N(t-1)-1)}$ is the probability that the walker choose node j out of the $p^+(N(t-1)-1)$ positive neighors of X_{t-1} to move to. We approximate the degree $\Delta_j^+(t')$ by its average using the same symbol, which follows the following recursion for $t' < t$ thus before the node is visited

$$\begin{aligned}\Delta_j^+(t') &\approx \frac{\Delta_j^+(t'-1)}{N(t'-1)-1}(\Delta_j^+(t'-1) - p^-(\Delta_j^+(t'-1)-1)-1) \\ &\quad + \left(1 - \frac{\Delta_j^+(t'-1)}{N(t'-1)-1}\right)(\Delta_j^+(t'-1) - p^-\Delta_j^+(t'-1)) \\ &= \Delta_j^+(t'-1)(1-p^-)\frac{N(t'-1)-2}{N(t'-1)-1},\end{aligned} \tag{8}$$

where $\Delta_j^+(0) = D_j^+(0) = k$. The first (second) term corresponds to the case that node j is (not) connected with the node visited at step $t'-1$. In the first case where j is connected with $X_{t'-1}$, the degree $\Delta_j^+(t')$ at step t' could be reduced from $\Delta_j^+(t'-1)$ due to the removal of $X_{t'-1}$ and its negative neighbours which happen to be a positive neighbour of node j. In the second case, the degree $\Delta_j^+(t')$ decreases from $\Delta_j^+(t'-1)$ due to the removal of $X_{t'-1}$'s negative neighbours which happen to be a positive neighbour of node j. Combining equations (2) and (5)–(8), we could derive the probability v_k that a random node j with degree $d_j^+(0) = k$ in the initial positive layer is visited by a SAP walk on an ER–ER signed network.

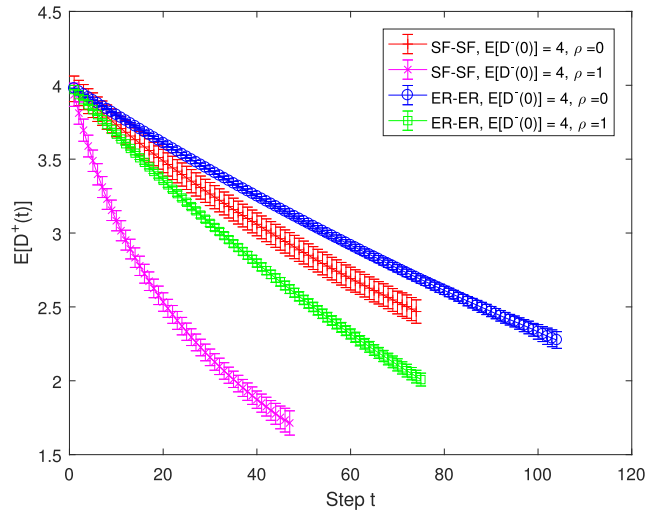
As shown in figure 4, our numerical solution of nodal visiting probability well approximates the simulation results especially when the initial negative ER network is sparse e.g. $E[D^-(0)] = 0$. When the initial negative ER network is denser, the actual visiting probability is lower than the prediction of the numerical solution. This is because our theoretical analysis assumes that the negative layer remains an ER network with the same link density after the removal of each visited node and its negative neighbours, as if all these nodes removed are chosen randomly. In fact, high negative degree nodes are more likely to be removed as a negative neighbour of a visited node. The actual $\Delta_j^+(t')$, thus also the visiting probability, is smaller than their corresponding analytic estimations.

Figure 4 shows that the visiting probability of a node grows approximately linearly with the initial positive degree of the node in each signed network model. A large slope of a curve in figure 4 features a fast growth of the nodal visiting probability as a function of the initial positive degree of a node, i.e. a high heterogeneity of nodal visiting probabilities. Interestingly, the order of the various signed network models in the heterogeneity of nodal visiting probabilities, the order of these networks in the average SAP walk length and the order of these signed networks in the average degree of the pruned positive layer at a given step are the same. A SAP walk that prunes the network slowly tends to have a long length and lead to a high heterogeneity in nodal visiting probabilities. A SAP walker tends to visit high degree nodes in the pruned positive layer at each step. A longer length of a SAP walk, thus, attributes to a higher visit probability of a node with a large initial positive degree, leading to more heterogeneity of nodal visiting probabilities.

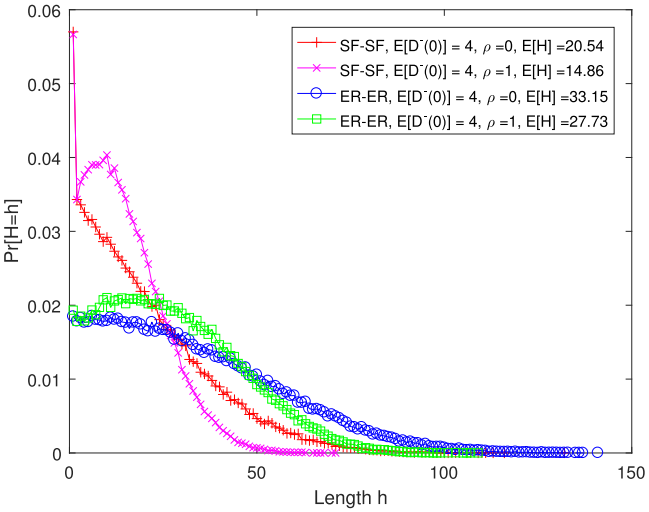
4.4. Influence of degree–degree correlation

The positive and negative degree of a node can be correlated. The real-world networks considered in section 7 have a positive correlation between the degrees of a node in the two layers. In this subsection, we explore how the degree–degree correlation between the positive and negative layers may influence the aforementioned performance of a SAP walk.

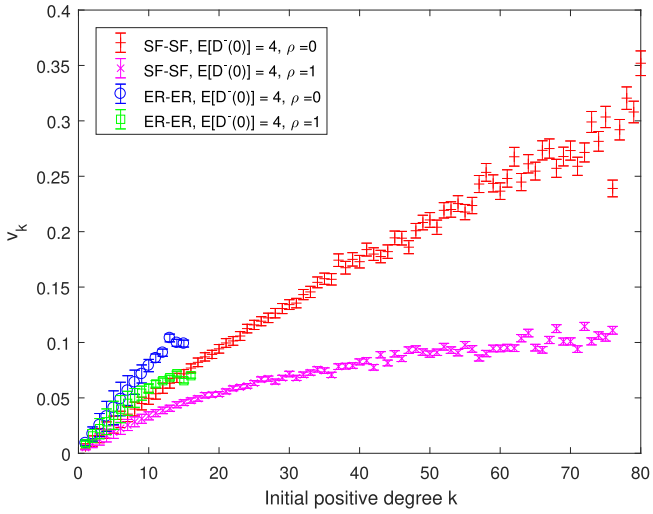
We consider ER–ER and SF–SF signed networks with $E[D^+] = E[D^-] = 4$ and $N = 1000$, where the degree–degree correlation ρ varies within [0,1]. We illustrate the three properties of a SAP walk using the two extreme case $\rho = 0$ and $\rho = 1$, whereas results from other ρ values within [0,1] lead to the same observations. We simulate 100 independent realisations of a SAP walk on each of the independently generated 100 signed networks to derive the three properties of SAP walks.



(a)



(b)



(c)

Figure 5. (a) The average degree $E[D^+(t)]$ of the positive pruned network as a function of the step t of a SAP walk (b) probability distribution of the length of a SAP walk and (c) the probability v_k that a node is visited by a SAP walk versus its degree $D^+(0) = k$ in the initial positive network, when the positive and negative degrees of a node are uncorrelated $\rho = 0$ and correlated $\rho = 1$ respectively.

In both ER–ER and SF–SF signed networks, we find that a positive degree–degree correlation evidently facilitates the pruning of the network, reduces the average path length and leads to a more homogeneous visiting probabilities among the nodes (see figures 5(a)–(c)). Such effects are more evident in SF–SF networks than in ER–ER networks and can be explained as follows. When the degree–degree correlation is positive, a high degree node in the positive layer tends to have a high degree in the negative layer. The high positive degree of such a node tends to let the walker visit the node in earlier steps. After being visited, the node together with its many negative neighbours, are removed, pruning the network significantly. The high negative degree of such a node tends to let the node be removed as a negative neighbour of a node that has been visited. The removal of such a node together with its many positive links significantly prunes the positive layer and reduces the connectivity.

4.5. Influence of the community structure in signed networks

Beyond the degree distribution and correlation in the two layers of a network, we explore further how the community structure of a signed network may influence the SAP walks. We consider the Girvan and Newman (GN) networks to model networks with a community structure [56]. Each GN network has $N = 1000$ nodes and average degree 4, the same as our ER and SF network models. The N nodes are divided into four groups, each with 250 nodes. Within each group, each of the $E[D_{\text{in}}] \cdot N/8$ links is placed between two nodes that are randomly selected from the group. On average, each node has $E[D_{\text{in}}]$ links connecting it to the nodes within the same group. We place further each of the $E[D_{\text{out}}] \cdot N/2$ links between two nodes that are randomly chosen from the N nodes but from different groups. The average out-degree $E[D_{\text{out}}]$ is the expected number of links that connect a node to the other nodes from a different group. The average degree 4 requires $E[D_{\text{in}}] + E[D_{\text{out}}] = 4$. When $E[D_{\text{out}}] = 0$, the GN network is composed of four isolated ER networks. When $E[D_{\text{out}}] = 3$, the GN network becomes an ER network. A GN network with $E[D_{\text{out}}] < 3$ represents a network with a community structure where nodes within the same group are more likely connected than nodes from different groups.

We focus on the case when the two layers of a signed network are independent. In this case, the SAP walks perform similarly on ER–GN (negative layer) networks and on ER–ER networks, since the visited nodes are as if randomly chosen from the negative layer. When the positive layer has the community structure, SAP walks perform quite differently. As shown in figure 6, as the GN–ER networks become more modular in the positive layer i.e. $E[D_{\text{out}}]$ becomes smaller, the average length of a SAP walk becomes smaller and the nodal visiting probabilities become less heterogeneous. Consider the case when $E[D_{\text{out}}] = 0$. The SAP walker walks only within the sub-graph of one group, limiting the length of a SAP walk. The length of a SAP walk in this case equals that of a SAP walk on an ER–ER network of $N/4$ nodes, where $E[D^+(0)] = 4$ and $E[D^-(0)] = 1$, because the sub-graph of the ER negative layer that corresponds to the group where the walker resides is again an ER network with the same link density. Hence, the distribution of the SAP walk length and the nodal visiting probabilities on GN–ER networks when $E[D_{\text{out}}] = 0$ can be analytically deduced by our theories for ER–ER networks. An interesting observation is that as $E[D_{\text{out}}]$ increases slightly from 0, the average length of a SAP increases significantly as shown in figure 6(a). The influence of the community structure of the positive layer on a SAP walk is evident when the community structure is significant. When the length of a walk becomes shorter as the positive layer becomes more modular, the nodal visiting probabilities become more homogeneous, which is consistent with our previous findings. The influence of the community structure in the positive layer on the pruning speed, i.e. the average degree $E[D^+(t)]$ of the positive pruned network as a function of the step t, is not evident. This is mainly due to the independence between the two layers: the negative neighbours of a visited node are as if chosen randomly from the positive layers.

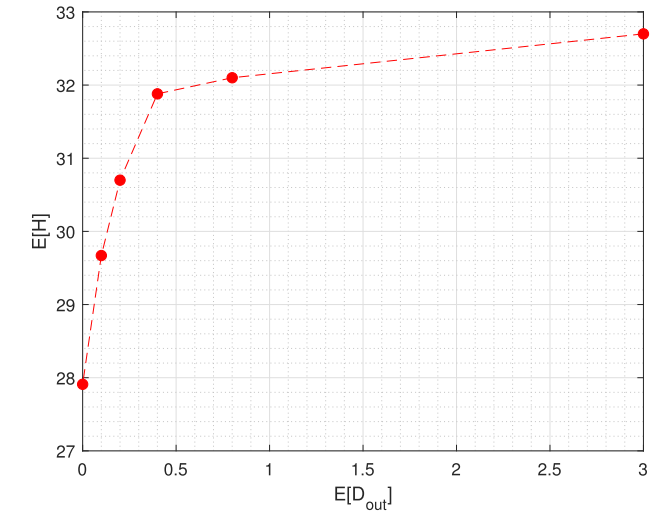
When the two layers are correlated, the influence of the community structure on SAP walks is non-trivial. Consider, for example, that both layers are GN networks with $E[D^-(0)] = 0.4$ and follow the same grouping (any two nodes that belong to the same group in one layer are also within the same group in the other layer). The average length of a SAP walk is further reduced to $E[H] = 28.31$ due to the correlated community structure in the negative layer.

5. Influence of signed network topology on SAP walk when $r \neq 1$

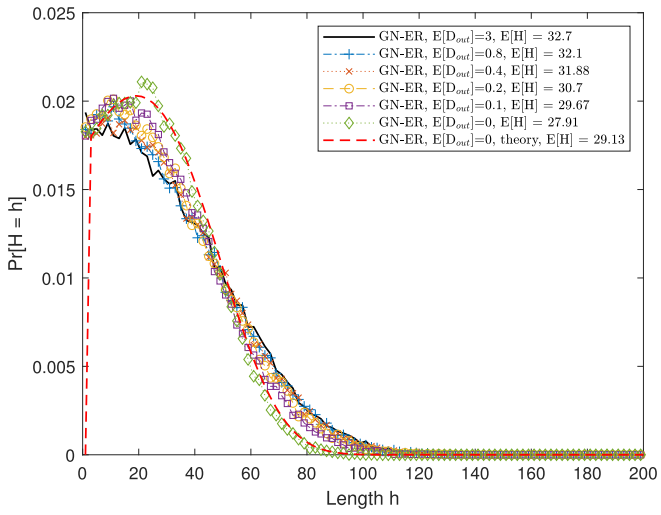
In this section, we consider the general case of the SAP walk that each negative neighbour of a visited node is removed independently with a probability r where $0 \leqslant r \leqslant 1$.

We first consider the ER–ER and SF–SF networks where the positive and negative layers are generated independently.

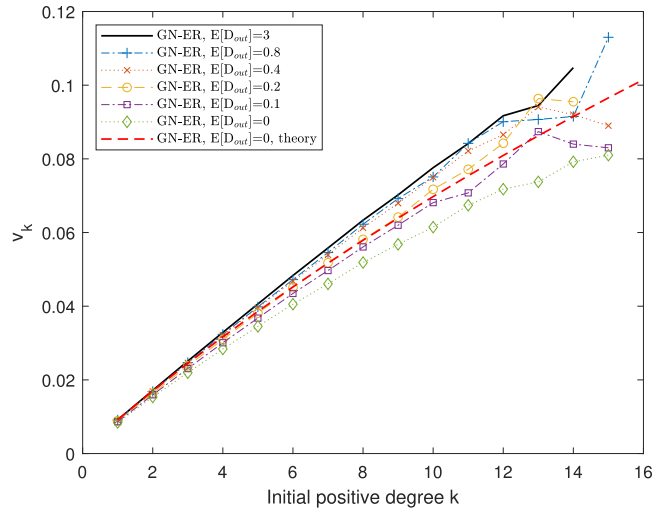
The SAP walk with a pruning probability r on an ER–ER signed network where the average degree of the two layers are $E[D^+]$ and $E[D^-]$ respectively is equivalent to the SAP walk with pruning probability 1 on an ER–ER signed network whose average degree in the two layers are $E[D^+]$ and $E[D^-] \cdot r$ respectively. Scaling the pruning probability by r in a SAP walk model is equivalent to scaling the link density of the negative layer by r.



(a)



(b)



(c)

Figure 6. (a) The average length of a SAP walk as a function of $E[D_{\text{out}}] = 3$ of the positive GN network layer, (b) the distribution of the length of a SAP walk and (c) the probability v_k that a node is visited by a SAP walk versus its degree $D^+(0) = k$ in the initial positive network. The SAP walk is on a GN–ER network with $N = 1000$ nodes and $E[D^+(0)] = E[D^-(0)] = 4$.

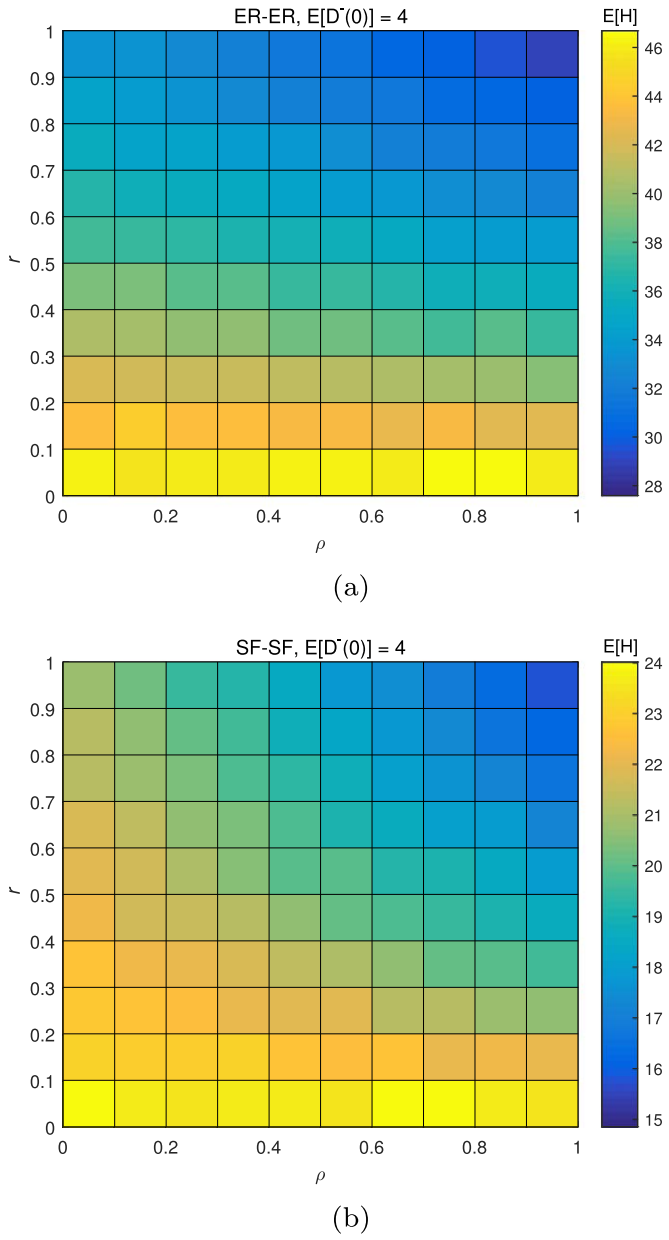


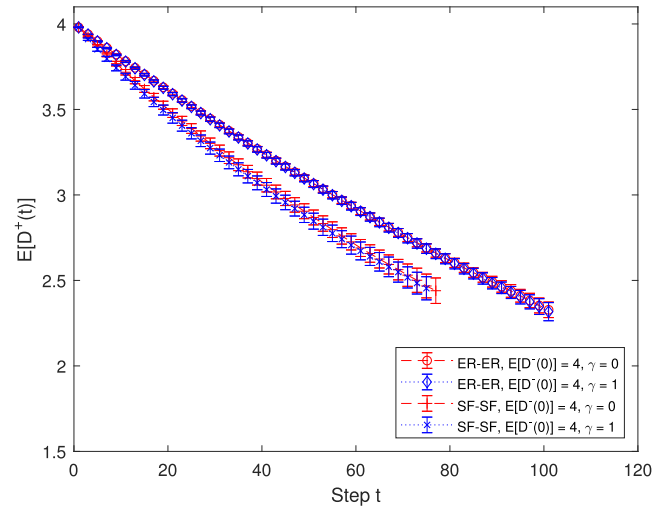
Figure 7. Phase diagram of the average path length $E[H]$ of a SAP walk. Dependence of the average path length on the pruning probability r and the degree–degree correlation ρ. The colour bar in the right represents the average path length. The underlying signed network is an ER–ER network in (a) and SF–SF in (b) with $E[D^+(0)] = E[D^-(0)] = 4$. For each set of parameters, the average path length is obtained as the average of 1000 SAP walks on each of the 100 signed networks.

Hence, all our theoretical results in section 4 for SAP walks with $r = 1$ on ER–ER networks can be extended to the SAP walks with an arbitrary pruning probability r on ER–ER networks.

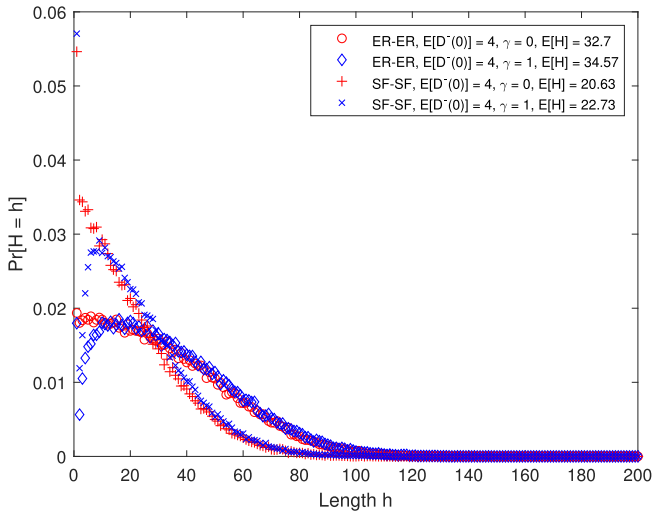
However, such equivalence does not hold when the positive and negative degrees of a node are correlated nor in SF–SF networks. We take the average path length of a SAP walk as an example and explore the effect of the pruning probability p and the degree–degree correlation ρ on the average path length. As shown in figure 7, the effect of the pruning probability on the average hopcount is more evident as the degree–degree correlation increases. When the degree–degree correlation is high, nodes with a high degree in both layers tend to be removed in early steps of a walk. In this case, a smaller pruning probability could effectively reduce the pruning.

6. Generalisation of the SAP walk model

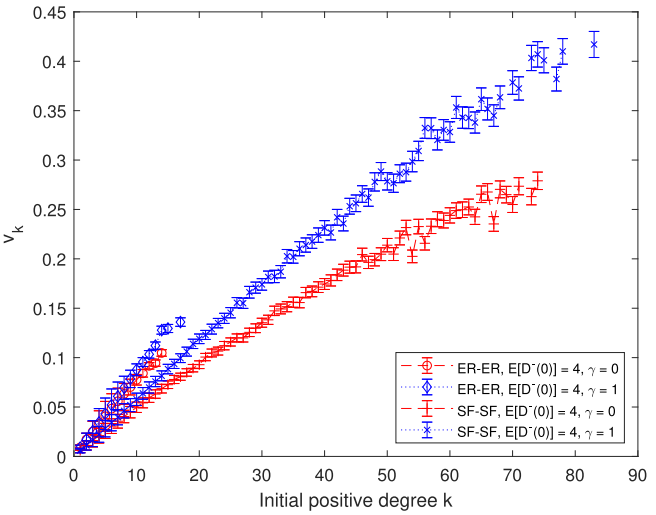
The SAP walk model can be generalised from multiple perspectives to better approximate real-world purchase behaviour of users. We illustrate one possible generalisation to take into account the heterogeneous preference of a user over the complementary/recommended products. It has been observed that customers tend to prefer popular projects and more engaged customers, i.e. those who have purchased/walked more, are more likely to



(a)



(b)



(c)

Figure 8. Comparison of the SAP walk when $\gamma = 0$ and the generalised SAP walk when $\gamma = 1$ in (a) the average degree $E[D^+(t)]$ of the positive pruned network as a function of the step t of a walk (b) the probability distribution of the length of a walk and (c) the probability v_k that a node is visited by a walk versus its degree $D^+(0) = k$ in the initial positive network. The two layers are independent in both ER–ER and SF–SF signed networks.

Table 1. Topological parameters of real networks.

Network	Nodes	Links	fraction of '+'links	fraction of '-'links	$E[D_+]$	$E[D_-]$	ρ_D
Wiki	6186	97 874	78 402 (80.11%)	19 472 (19.89%)	25.35	6.30	0.62
Epinions	6186	92 091	73 062 (79.33%)	19 029 (20.67%)	23.62	6.15	0.35

buy niche or less popular products [57]. We have shown that the popularity of a product, i.e. the visiting probability of the corresponding node, grows approximately linearly with its initial positive degree. Hence, we consider the generalised model where the probability for a walker at step t residing at node i to walk to a positive neighbour j of i is proportional to $d_j^+(0)^{\gamma/(t+1)}$ where $d_j^+(0)$ is the positive degree of node j in the initial signed network and $\gamma > 0$. In this case, a node with a high initial positive degree is preferred, whereas such preference becomes weaker as the walker moves more, i.e. becomes more engaged. Our classic SAP model discussed earlier corresponds to the case when $\gamma = 0$. In figure 8, we compare the three walk features of our classic model when $\gamma = 0$ and the generalised model when $\gamma = 1$. We find that such preference of visiting a node with a high initial positive degree leads to a longer walk on average and more heterogeneous nodal visiting probabilities. A node with a high initial positive degree is visited more thus becomes more popular when $\gamma = 1$.

The SAP model can be as well extended by letting the probability for a walker at step t residing at node i to walk to a positive neighbour j of i be proportional to the popularity of j, i.e. the number of walkers that have visited node j and such preference decreases as a walker moves more. We hypothesise that a node with a large initial positive degree may become more and more popular over time, based on our observations in the SAP walk with $\gamma = 1$.

7. SAP walks on real-world signed networks

Finally, we choose two real-world signed networks and explore their network features and how these features may influence the SAP walks on these networks. We consider the Wikipedia adminship election network and an Extracted Epinions social network [58]. In Wiki network, two nodes connected by a positive (negative) link suggest that the two users support (reject) each other to be an administrator. A positive (negative) link in Epinions network means that the corresponding two users trust (distrust) each other's reviews.

The Epinions network is far larger than Wiki. We have sampled the Epinions network by firstly removing all nodes with zero positive degree or zero negative degree and then randomly selecting the same number of nodes as in Wiki from the largest connected positive layer of Epinions together with the positive and negative links among these nodes. Basic topological features of these two networks of the same size are shown in table 1. The degree correlation ρ_D measures the linear correlation coefficient between the positive degree and negative degree of a node. The positive and negative layers tend to be positively correlated in their degrees, i.e. $\rho_D > 0$, instead of independent as assumed in our signed network models.

The degree distributions of the positive and negative layer in both Wiki and Epinions are shown in figure 9 are highly heterogeneous, closer to a scale-free distribution than a Poisson distribution.

Upon each real-world signed network, we simulate independently 10^5 SAP walks and investigate their key properties discussed earlier. The negative network layer has been missing in modelling the purchase behaviour of a user. Hence, we consider as well the SAP walk on these two real-world networks, where, however, the negative network layer is replaced as an empty network without any link.

Figure 10(a) shows that the positive layer is pruned or shrinks faster in Wiki than in Epinions network. Wiki has a shorter length of SAP walk on average than Epinions, as shown in figure 10(b). One explanation for both observations is that Wiki has a slightly denser initial negative layer (larger $E[D^-(0)]$) than Epinions as shown in table 1, which removes on average more nodes per step. Moreover, the high degree correlation ρ_D in Wiki contributes as well to a fast pruning in the positive layer and a short length of SAP walks.

When the negative layer is empty, i.e. $E[D^-(0) = 0]$, the positive layer is pruned far slower, the average path length $E[H]$ is far larger. A SAP walk on a signed network with an empty negative layer is equivalent to a self-avoiding walk on the positive network layer. In this case, the Epinion positive layer leads to a longer average path length than the Wiki positive layer. This is likely due to the higher standard deviation of the degree in the Wiki positive layer 49.55 than that in the Epinion 45.58 positive layer. A higher degree standard deviation implies an earlier visit of the hubs, whose removal may significantly prune the network and reduce the connectivity.

The visiting probability of a node versus its initial positive degree tends to have a larger slope in Epinions than that in Wiki. This is consistent with our observations in signed network models that the visiting probability v_k increases faster with k in a signed network that leads to a higher average degree $E[D^+(t)]$ of the pruned positive layer and a longer length $E[H]$ of SAP walks.

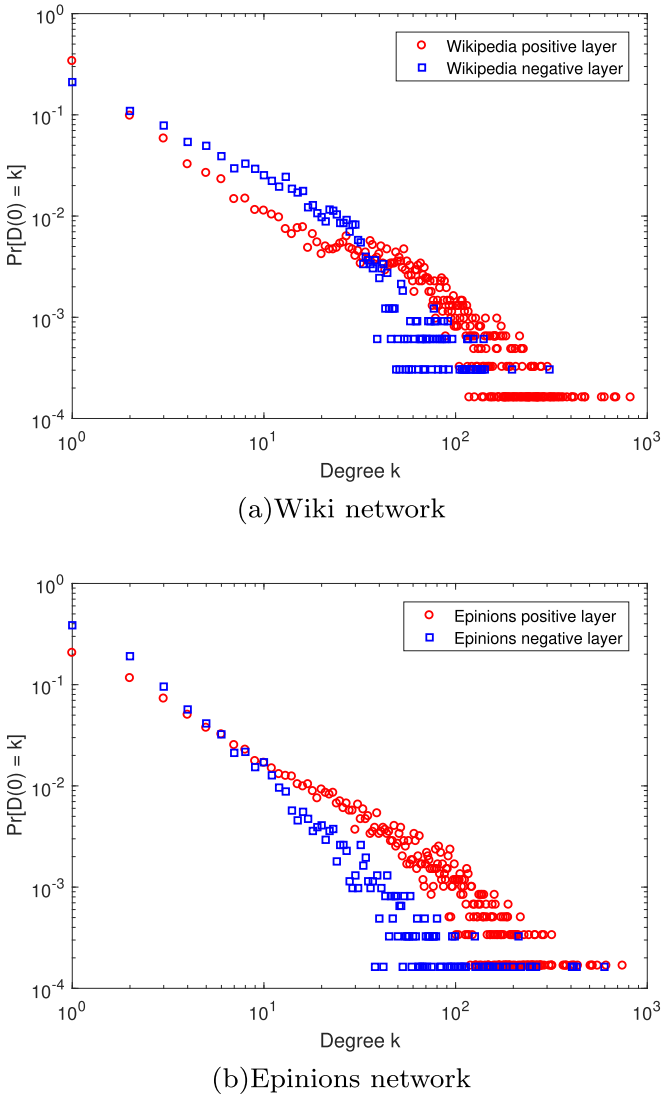
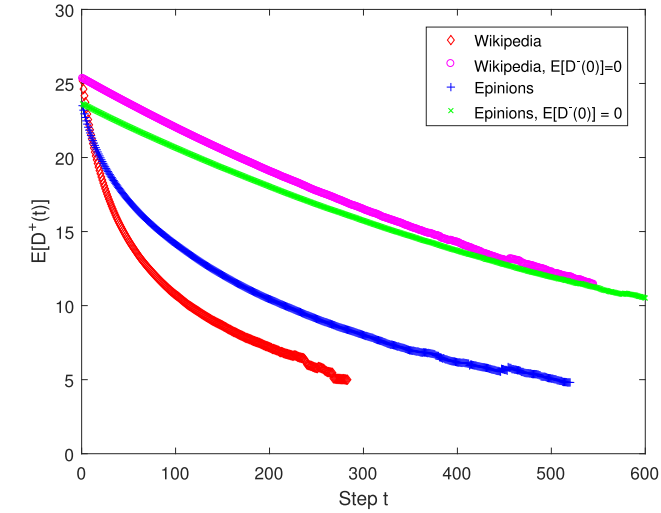


Figure 9. Degree distribution of real-world signed networks.

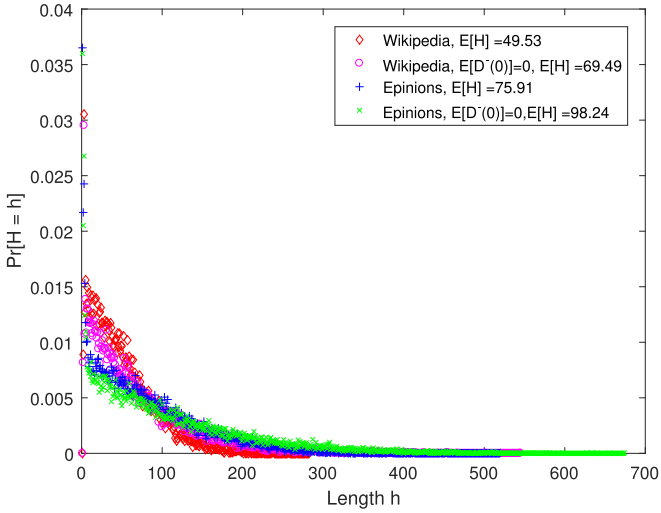
The negative network layer dramatically prunes a signed network and reduces length of a SAP walk. Moreover, a heterogeneous degree distribution in the positive layer and a positive degree–degree correlation between positive and negative layers may further enhance the pruning effect, shorten the SAP walk length and facilitate homogeneous visiting probabilities of nodes. These effects have been observed consistently in both network models and real-world networks.

8. Conclusion

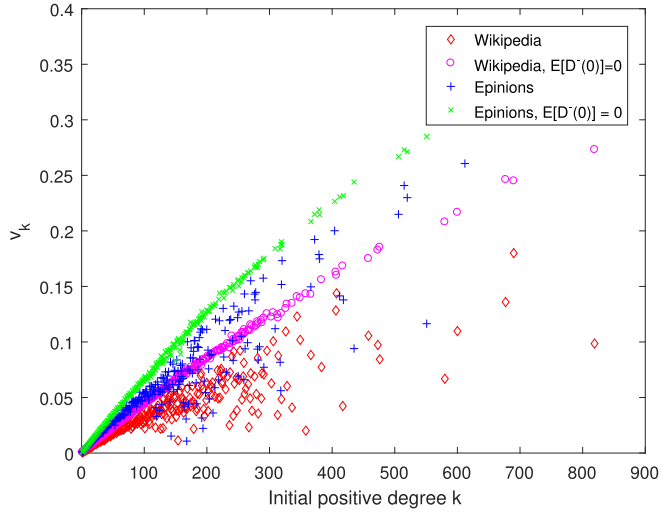
Classic spreading models assume that all network links are beneficial for information diffusion. However, the positive and negative links in a signed network may facilitate and prevent the contagion of information, opinion and behaviour etc respectively. As a start, we propose a SAP RW on a signed network to model, for example, a user's purchase activity on a signed network of products. We unravel the significant effect of the negative links and the signed network structure in general on SAP walks. We found that a more heterogeneous degree distribution of the positive network layer such as the power-law distribution, a denser negative layer and a high degree–degree correlation between the two layers tend to prune the network faster, suppress the length of SAP walks and reduce the heterogeneity in nodal visiting probabilities. When the two layers are independent, however, a more heterogeneous degree distribution of the negative network layer tends to slow down the pruning and contribute to a longer length of SAP walks and more heterogeneity in nodal visiting probabilities. These observations has been obtained from both signed network models and real-world signed network and analytically proved in signed ER–ER networks. Real-world networks tend to have a heterogeneous degree distribution in the positive layer and a positive degree–degree correlation, which reduce total purchases of users



(a)



(b)



(c)

Figure 10. (a) The average degree $E[D^+(t)]$ of the positive pruned network as a function of the step t of a SAP walk on a real-world signed network. (b) Probability distribution of the length of a SAP walk on a signed real-world network. (c) The probability v_k that a node is visited by a SAP walk versus its degree $D^+(0) = k$ in the initial positive network.

but increase the homogeneity of the popularity of products. Our findings point out the possibility to influence users' purchases and product popularity via recommendations and competitions.

It is interesting to explore further the influence of other key features. We have shown that the community structure in the positive network layer may reduce the length of a SAP walk and the heterogeneity of nodal visiting probabilities. Beyond the community structure, balanced triangles are shown to appear more frequently than unbalanced ones in real-world networks. The effect of the fraction of balanced triangles on SAP walks and other dynamic processes remains interesting to investigate. We could as well to improve the SAP walk towards a more realistic model of e.g. user's purchase activity, by taking into account, for example, the choice of the initial node to visit, the possibility that a walker/user may stop the walk earlier and the heterogeneity of the links preference over recommendations. Optimisation problems that are interesting to be further explored include how to add nodes to an existing signed network, how to add positive links via e.g. recommendations or how to recommend a product path/sub-graph to maximise the visiting probabilities of a group of nodes.

Acknowledgments

The authors would like to thank National Nature Science Foundation of China (Nos. 11601430, 11631014, 11871311) for support.

ORCID iDs

Huijuan Wang https://orcid.org/0000-0003-2684-4407
Cunquan Qu https://orcid.org/0000-0001-8090-2538

References

[1] Buldyrev S V, Parshani R, Paul G, Stanley H E and Havlin S 2010 *Nature* **464** 1025
[2] Kivelä M, Arenas A, Barthelemy M, Gleeson J P, Moreno Y and Porter M A 2014 *J. Complex Netw.* **2** 203
[3] De Domenico M, Solé-Ribalta A, Cozzo E, Kivelä M, Moreno Y, Porter M A, Gómez S and Arenas A 2013 *Phys. Rev.* X **3** 041022
[4] Sahneh F D, Scoglio C and Mieghem P Van 2013 *IEEE/ACM Trans. Netw.* **21** 1609
[5] Cardillo A, Gómez-Gardeñes J, Zanin M, Romance M, Papo D, Pozo F and Boccaletti S 2013 *Sci. Rep.* **3** 1344
[6] De Domenico M, Granell C, Porter M A and Arenas A 2018 *Nat. Phys.* **14** 523
[7] De Domenico M, Granell C, Porter M A and Arenas A 2016 *Nat. Phys.* **12** 901
[8] Solé-Ribalta A, De Domenico M, Gómez S and Arenas A 2016 *Physica* D **323** 73
[9] Saumell-Mendiola A, Serrano M A and Boguñá M 2012 *Phys. Rev.* E **86** 026106
[10] Sahneh F D, Scoglio C and Chowdhury F N 2013 *2013 American Control Conf.* 2307–12
[11] Wang H, Li Q, D'Agostino G, Havlin S, Stanley H E and Van Mieghem P 2013 *Phys. Rev.* E **88** 022801
[12] Cozzo E, Banos R A, Meloni S and Moreno Y 2013 *Phys. Rev.* E **88** 050801
[13] Liu M, Li D, Qin P, Liu C, Wang H and Wang F 2015 *PloS one* **10** e0120701
[14] Li D, Qin P, Wang H, Liu C and Jiang Y 2014 *Europhys. Lett.* **105** 68004
[15] Granell C, Gómez S and Arenas A 2014 *Phys. Rev.* E **90** 012808
[16] Granell C, Gómez S and Arenas A 2013 *Phys. Rev. Lett.* **111** 128701
[17] Li Q, Braunstein L A, Wang H, Shao J, Stanley H E and Havlin S 2013 *J. Stat. Phys.* **151** 92
[18] de Arruda G F, Cozzo E, Peixoto T P, Rodrigues F A and Moreno Y 2017 *Phys. Rev.* X **7** 011014
[19] Wang H, Chen C, Qu B, Li D and Havlin S 2017 *New J. Phys.* **19** 073039
[20] Leskovec J, Huttenlocher D and Kleinberg J 2010 *Proc. SIGCHI Conf. on Human Factors in Computing Systems, CHI '10* (New York, NY, USA: ACM) pp 1361–70
[21] McAuley J, Pandey R and Leskovec J 2015 *Proc. 21th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, KDD '15* (New York, NY, USA: ACM) pp 785–94
[22] Chiang K-Y, Hsieh C-J, Natarajan N, Dhillon I S and Tewari A 2014 *J. Mach. Learn. Res.* **15** 1177–1213
[23] Tang J, Chang Y, Aggarwal C C and Liu H 2016 *ACM Computing Surveys (CSUR)* **493** 42
[24] Hou L, Liu K and Liu J 2018 *Complex Networks and their Applications VI* ed C Cherifi *et al* (Cham: Springer International Publishing) pp 935–45
[25] Oestreicher-Singer G, Libai B, Sivan L, Carmi E and Yassin O 2013 *J. Mark.* **77** 1
[26] Masuda N, Porter M A and Lambiotte R 2017 *Phys. Rep.* **716** 1–58
[27] Erdős P and Rényi A 1960 *Publ. Math. Inst. Hung. Acad. Sci.* **6** 5117–5160
[28] Bollobás B 1998 *Modern Graph Theory* (Berlin: Springer) 215–52
[29] Caldarelli G, Capocci A, De Los Rios P and Muñoz M A 2002 *Phys. Rev. Lett.* **89** 258702
[30] Söderberg B 2002 *Phys. Rev.* E **66** 066121
[31] Boguñá M and Pastor-Satorras R 2003 *Phys. Rev.* E **68** 036112
[32] Barabási A-L 2016 *Network Science* (Cambridge: Cambridge University Press)
[33] Catanzaro M, Boguñá M and Pastor-Satorras R 2005 *Phys. Rev.* E **71** 027103
[34] Cohen R and Havlin S 2003 *Phys. Rev. Lett.* **90** 058701
[35] Barabási A-L and Albert R 1999 *Science* **286** 509
[36] Szell M, Lambiotte R and Thurner S 2010 *Proc. Natl Acad. Sci.* **107** 13636
[37] Schwartz N, Cohen R, Ben-Avraham D, Barabási A-L and Havlin S 2002 *Phys. Rev.* E **66** 015104
[38] Qu B, Li Q, Havlin S, Stanley H E and Wang H 2014 *Phys. Rev.* E **90** 052811

[39] Newman M E J, Strogatz S H and Watts D J 2001 *Phys. Rev.* E **64** 026118
[40] Li X, Tao X and Li N 2016 *IEEE Commun. Lett.* **20** 2280
[41] Tizghadam A and Leon-Garcia A 2010 *IEEE INFOCOM 2010 Conf. on Computer Communications Workshops* pp 1–6
[42] Aldrich M, Badshah A, Mayton B, Zhao N and Paradiso J A 2013 *2013 IEEE SENSORS* pp 1–4
[43] Noulas A, Scellato S, Lathia N and Mascolo C 2012 *2012 Int. Conf. on Privacy, Security, Risk and Trust (PASSAT), and 2012 Int. Conf. on Social Computing (SocialCom)* pp 144–53
[44] Saralegui U, la Sen M D and Alonso-Quesada S 2016 *2016 UKACC 11th Int. Conf. on Control (CONTROL)* pp 1–6
[45] Flory P J 1953 *Principles of Polymer Chemistry* (Ithaca, NY: Cornell University Press)
[46] Ou-Yang L, Dai D Q and Zhang X F 2015 *IEEE/ACM Trans. Comput. Biol. Bioinform.* **12** 1333
[47] Camilleri E, Rohde P P and Twamley J 2014 *Sci. Rep.* Sci. Rep. **4** 4791
[48] Nia M A, Atani R E, Fabian B and Babulak E 2016 *Secur. Commun. Netw.* **16** 3509–3526
[49] Lawler G F 1980 *Duke Math. J.* **47** 655
[50] Guttmann A J 2012 arXiv:1212.3448 [math-ph]
[51] Tishby I, Biham O and Katzav E 2016 *J. Phys. A: Math. Theor.* **49** 285002
[52] Jinhong Jung L S, Jin W and Kang U 2016 *ICDM* 973–978
[53] Li Y, Chen W, Wang Y and Zhang Z-L 2013 *Proc. 6th ACM Int. Conf. on Web Search and Data Mining, WSDM '13* (New York, NY, USA: ACM) pp 657–66
[54] Kempe D, Kleinberg J and Tardos E 2003 *Proc. 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, KDD '03* (New York, NY, USA: ACM) pp 137–46
[55] Li B 2016 *Master Thesis* Delft University of Technology
[56] Girvan M and Newman M E J 2002 *Proc. Natl Acad. Sci.* **99** 7821
[57] Goel E G S, Broder A and Pang B 2010 *WSDM 10: Proc. 3rd ACM Int. Conf. on Web Search and data Mining* 201–10
[58] Leskovec J and Krevl A 2014 SNAP Datasets: Stanford large network dataset collection