

# SCIENTIFIC REPORTS

OPEN

## Efficient network disintegration under incomplete information: the comic effect of link prediction

Suo-Yi Tan<sup>1,\*</sup>, Jun Wu<sup>1,\*</sup>, Linyuan Lü<sup>2,3,\*</sup>, Meng-Jun Li<sup>1</sup> & Xin Lu<sup>1,4</sup>

Received: 11 December 2015

Accepted: 24 February 2016

Published: 10 March 2016

The study of network disintegration has attracted much attention due to its wide applications, including suppressing the epidemic spreading, destabilizing terrorist network, preventing financial contagion, controlling the rumor diffusion and perturbing cancer networks. The crux of this matter is to find the critical nodes whose removal will lead to network collapse. This paper studies the disintegration of networks with incomplete link information. An effective method is proposed to find the critical nodes by the assistance of link prediction techniques. Extensive experiments in both synthetic and real networks suggest that, by using link prediction method to recover partial missing links in advance, the method can largely improve the network disintegration performance. Besides, to our surprise, we find that when the size of missing information is relatively small, our method even outperforms than the results based on complete information. We refer to this phenomenon as the “comic effect” of link prediction, which means that the network is reshaped through the addition of some links that identified by link prediction algorithms, and the reshaped network is like an exaggerated but characteristic comic of the original one, where the important parts are emphasized.

Complex networks describe a wide range of systems in nature and society<sup>1–3</sup>. Examples include the Internet, metabolic networks, electric power grids, supply chains, urban road networks, and the world trade web among many others. The study of complex networks has become an important area of multidisciplinary research involving physics, mathematics, biology, social sciences, informatics, and other theoretical and applied sciences. Due to its broad applications, research on the structural robustness of complex networks, i.e., the ability to endure threats and survive accidental events, has received growing attention<sup>4–9</sup> and has become one of the central topics in the complex network research.

In the majority of cases, networks are beneficial, such as power grids and Internet, where we want to preserve their function. Many studies have considered methods for maximizing the structural robustness of these beneficial networks<sup>10–16</sup>. In another situation by which this paper is motivated, however, we want to disintegrate a network if it is harmful, such as immunizing a population in social networks or suppressing the virus propagation in computer networks. The immunization problem is mathematically equivalent to asking how to disintegrate a given network with a minimum number of node removals<sup>17</sup>, which is very important since in most cases the number of immunization doses is limited or very expensive. Other examples of network disintegration include destabilizing terrorist networks<sup>18</sup>, preventing financial contagion<sup>19</sup>, controlling the rumor diffusion<sup>20</sup>, and perturbing cancer networks<sup>21</sup>. Although the problem of network disintegration attracts less attention than the problem of network protection, some related works have been devoted to the study of the disintegration strategy. For example, Holme *et al.*<sup>22</sup> compared the effect of four different targeted disintegration strategies: high degree and betweenness centrality, and their corresponding adaptive versions where the degree (betweenness) of the remaining node is recomputed after each node removal. They found that the removals by the two adaptive methods outperform the two original static methods. Chen *et al.*<sup>23</sup> developed a new immunization strategy, called the “equal graph partitioning” (EGP) strategy. The main idea of the EGP method is to fragment the network into many connected clusters of equal size, which requires 5% to 50% fewer immunization doses compared to the classical

<sup>1</sup>College of Information System and Management, National University of Defense Technology, Changsha, Hunan, 410073, P. R. China. <sup>2</sup>Alibaba Research Center for Complexity Sciences, Alibaba Business College, Hangzhou Normal University, Hangzhou, Zhejiang, 311121, P. R. China. <sup>3</sup>Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, China. <sup>4</sup>Department of Public Health Sciences, Karolinska Institutet, Stockholm, 17177, Sweden. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.W. (email: junwu@nudt.edu.cn)

targeted strategy. Schneider *et al.*<sup>24</sup> developed an immunization approach based on optimizing the susceptible size, which outperforms the best known strategy based on immunizing the highest-betweenness links or nodes.

In the early works on network disintegration, it was usually assumed that the attacker can obtain perfect information on the network structure, in other words, they assumed that the observed networks are complete. However, the complete information of network structure is not always available in realistic cases. Growing attention has been paid to the study of network disintegration with imperfect information. Dezső *et al.*<sup>25</sup> proposed a biased treatment strategy against viruses spreading based on uncertain information, in which the likelihood of identifying and administering a cure to an infected node depends on its degrees as  $k^\alpha$ . Li *et al.*<sup>26</sup> studied the optimal attack problem based on incomplete information, which means that one can obtain the information of partial nodes, when the information is certain. Moreover, many researches<sup>27–30</sup> focused on the disintegration strategy based on local information, i.e. the knowledge of the neighborhood.

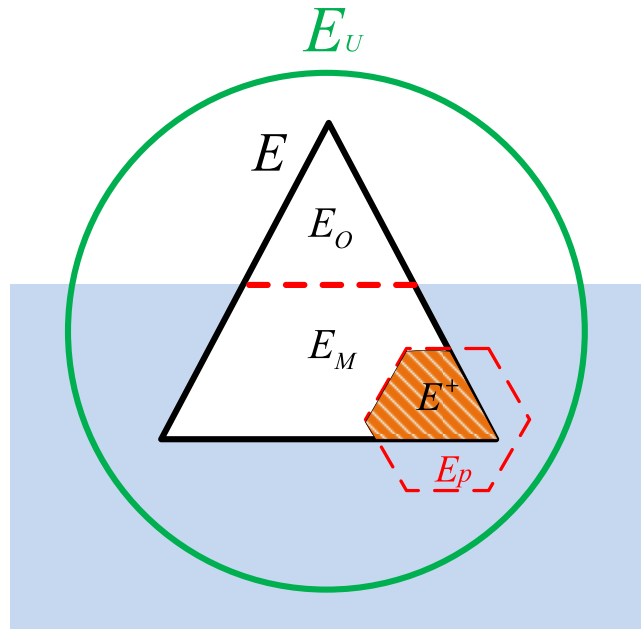
Different from the above studies which consider either uncertain information or partial information of individual level, in this paper we focus on another important and frequent scenario of imperfect information, such that part of links (i.e., interactions between nodes) are missing in the observed network. In many real networks, such as food webs<sup>31</sup>, terrorist networks<sup>32</sup>, sexual contact networks<sup>33</sup>, protein-protein interaction networks<sup>34</sup>, and disease relationship networks<sup>35</sup>, it is easy to obtain the information of nodes, but difficult to detect the relations or interactions between nodes, which is usually costly or even infeasible. The missing links may reduce the network disintegration performance. To address this problem, a potential approach is to recover the missing links (or part of the missing links), which remind us the so-called “link prediction” problem<sup>36</sup>. Link prediction algorithms aim at estimating the likelihood of the existence of a link between two nodes based on the observed network structure and the attributes of nodes. Therefore, before the attack we can use one of the link prediction algorithms to recover parts of the missing links and then identify the targets based on the “improved” network. Experiments on both synthetic and real networks show that with the assistance of link prediction the performance of disintegration can be largely improved.

## Results

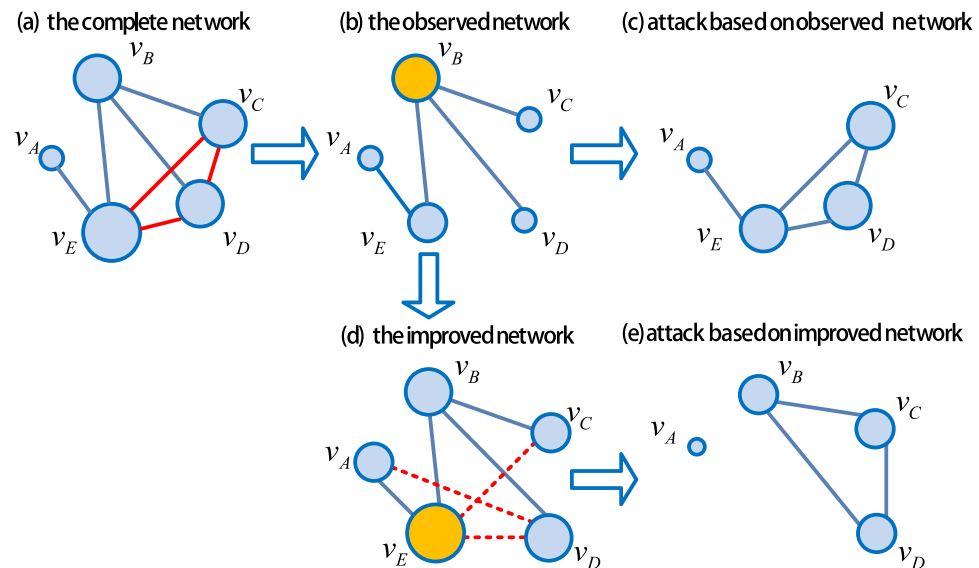
**Network disintegration model based on link prediction.** A network can be presented by a simple undirected graph  $G = (V, E)$ , where  $V$  is the set of nodes, and  $E$  is the set of links. Multiple links and self-loops are not allowed. Let  $N = |V|$  and  $W = |E|$  be the number of nodes and number of links, respectively. Let  $k_i$  be the degree of node  $v_i$ , which equals the number of links connected to node  $v_i$ . We assume that all nodes are known but partial link information is missing. Denote by  $E_O$  and  $E_M$  the set of observed links and missing links, respectively. Clearly, we have  $E_O \cup E_M = E$ . Therefore, the observed network can be presented by  $G_O = (V, E_O)$ . We define  $\alpha = |E_M|/W \in [0, 1]$  as the proportion of missing link. Denote by  $E_U = V \times V$  the universal set containing all  $N(N-1)/2$  possible links. The task of link prediction is to reveal the set of missing links  $E_M$  from the space of link prediction  $\Omega_P = E_U - E_O$ . Denote by  $G_P = (V, E_O \cup E_P)$  the improved network by adding the predicted links  $E_P (\subseteq \Omega_P)$ . We define the ratio  $\beta = |E_P|/|E_O|$  as the magnitude of additional link information. In general, we have  $E_P \neq E_M$  due to the error predictions. Denote by  $E^+ = E_P \cap E_M$  the set of links that are correctly predicted. We use the true positive rate (recall or sensitivity)  $R_{TPR} = |E^+|/|E_M|$  to measure the proportion of links that are correctly predicted among the missing links set  $E_M$ , and the ratio  $R_{ppv} = |E^+|/|E_P|$ , i.e., the positive predictive value (precision), to measure the proportion of links that are correctly predicted among the predicted links set  $E_P$ . To express the mathematical description of link prediction intuitively, we give the iceberg diagram for link prediction problem in Fig. 1. In a manner of speaking, the network is like an iceberg. We can only see the part above sea level but do not know the rest under the sea. Link prediction is a technique to infer the invisible part based on the knowledge of observed part.

We identify the targets based on the improved network  $G_P$  and then carry out the attack in the original complete network  $G$ . Note that if a node is attacked, its attached links will be removed together with its removal. Denote by  $\hat{V} \subseteq V$  the set of nodes that are attacked (i.e., targets) and  $\hat{E} \subseteq E$  the set of removed links, then the network obtained after node attacks is  $\hat{G} = (V - \hat{V}, E - \hat{E})$ . We define the ratio  $f = |\hat{V}|/N \in [0, 1]$  as the strength coefficient of node attacks. Among the many attack strategies<sup>28</sup> we apply the most used “high degree strategy” in this paper. In this strategy, nodes are attacked according to their rank of degree. i.e., high degree nodes will be attacked firstly. Let  $k_i^O$  be the degree of node  $v_i$  in  $G_O$  and  $k_i^P$  be the degree of node  $v_i$  in  $G_P$ . Without link prediction, we remove nodes in the descending order of the node degree  $k_i^O$ . With link prediction, we remove nodes in the descending order of the node degree  $k_i^P$ . As the attack strength coefficient  $f$  increases, the network will eventually collapse at a critical value  $f_c$  which is generally used to measure the structure robustness of a complex network from the view of defenders. The larger the  $f_c$  is, the more robust the network is. Here we employ  $f_c$  to evaluate the performance of network disintegration strategy from the view of attackers. Smaller  $f_c$  implies more efficient network disintegration.

Figure 2 presents a simple example of how our method works. The complete network contains  $N = 5$  nodes and  $W = 7$  links. The initial degrees of the five nodes in the complete network are  $k_A = 1, k_B = 3, k_C = 3, k_D = 3$ , and  $k_E = 4$ , respectively. We assume that three links are missing, namely  $E_M = \{e_{CD}, e_{CE}, e_{DE}\}$ . The observed network contains four links,  $E_O = \{e_{AE}, e_{BC}, e_{BD}, e_{BE}\}$ . Then the magnitude of missing link information is  $\alpha = |E_M|/W = 3/7$  and the space of link prediction is  $\Omega_P = \{e_{AB}, e_{AC}, e_{AD}, e_{CD}, e_{CE}, e_{DE}\}$ . Assume we add three links, i.e.,  $E_P = \{e_{AD}, e_{CE}, e_{DE}\}$ , predicted by one link prediction algorithm<sup>37</sup>. Then the magnitude of link prediction information is  $\beta = |E_P|/|E_O| = 3/4$ . Among the three links in  $E_P$ , only  $e_{CE}$  and  $e_{DE}$  are predicted right, i.e.,  $E^+ = E_P \cap E_M = \{e_{CE}, e_{DE}\}$ . Thus we obtain the sensitivity  $R_{TPR} = |E^+|/|E_M| = 2/3$  and the precision  $R_{ppv} = |E^+|/|E_P| = 2/3$ . The degrees of the five nodes in the observed network  $G_O$  are  $k_A^O = 1, k_B^O = 3, k_C^O = 1, k_D^O = 1$  and  $k_E^O = 2$ , respectively. After the addition of three predicted links, their degrees in the improved network  $G_P$  (see Fig. 2(d)) become  $k_A^P = 2, k_B^P = 3, k_C^P = 2, k_D^P = 3$  and  $k_E^P = 4$ , respectively. Based on the observed



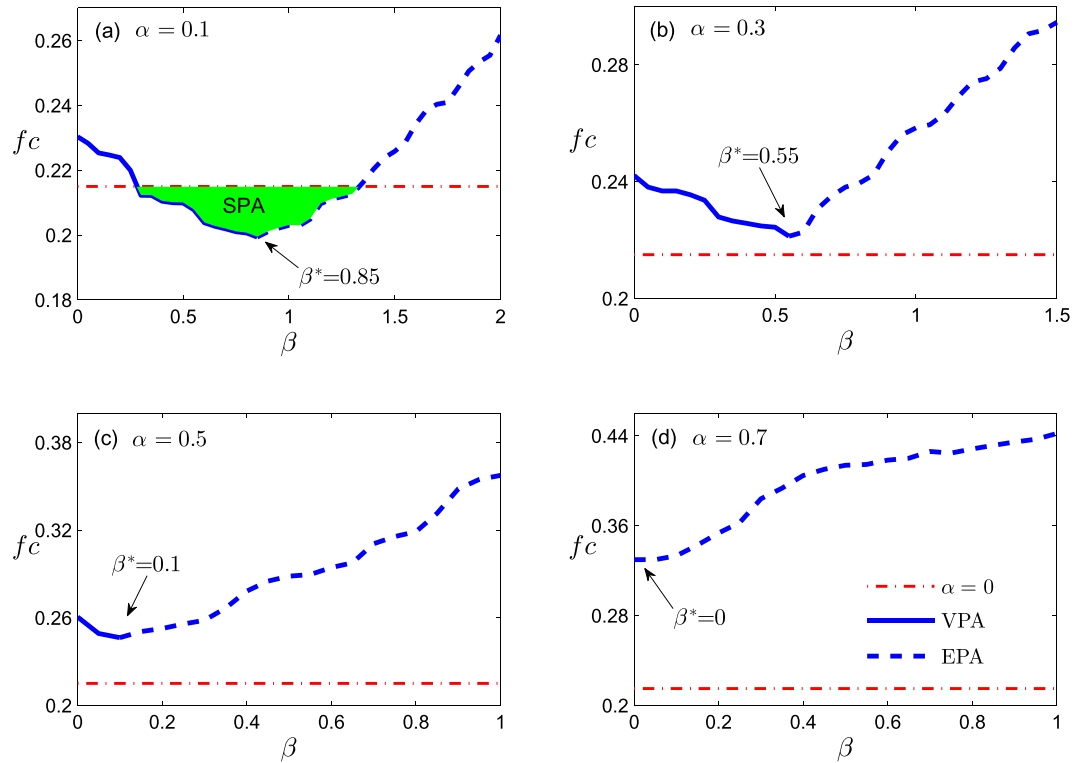
**Figure 1. Iceberg diagram for link prediction problem.** The triangle represents the set of links  $E$ , i.e., the complete information, which is divided into two parts: above the sea level is the observed part  $E_O$ , below the sea level is the invisible (missing) part  $E_M$ . The hexagon represents the set of predicted links, namely  $E_p$ . The polygon filled by stripes represents the set of links that are predicted right, namely  $E^+$ . The circle represents the universal set containing all possible links  $E_U$ .



**Figure 2. Illustration of network disintegration model based on link prediction.** (a) The complete network  $G$ . (b) The observed network  $G_O$  with three missing links. (c) The network obtained after removing the node  $v_B$  based on the observed network. (d) The improved network  $G_P$  with three predicted links added (dotted lines). (e) The network obtained after removing the node  $v_E$  based on the improved network. The size of each node is proportional to its degree in the current network.

network  $G_O$ , the node  $v_B$  with the largest degree  $k_B^O = 3$  will be removed preferentially as shown in Fig. 2(c), and the network  $\hat{G}$  obtained after removing the node  $v_B$  is still connected. While based on the improved network, the node  $v_E$  with the largest degree  $k_E^P = 4$  will be removed preferentially as shown in Fig. 2(e), and the network  $\hat{G}$  obtained after removing the node  $v_E$  is disintegrated into two components.

**Comic effect of link prediction.** To analyze the impact of link prediction on network disintegration, we firstly perform experiments on synthetic networks. Due to the ubiquity of scale-free networks with a power-law degree distribution  $p(k) \sim k^{-\lambda}$  in real life world, our studies first focus on the network disintegration in scale-free

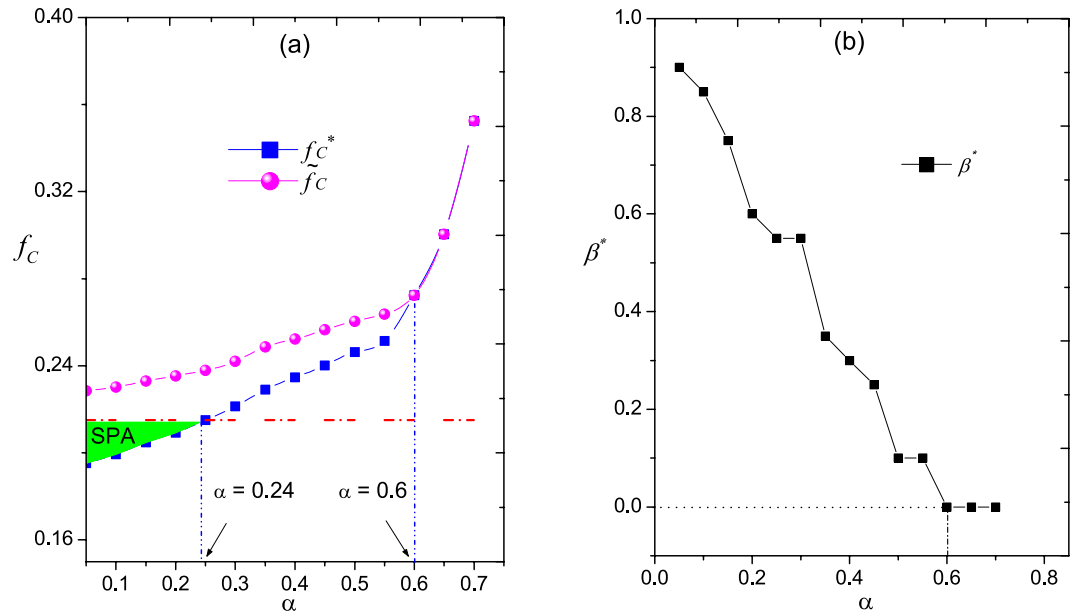


**Figure 3.** The critical attack strength coefficient  $f_c$  versus the magnitude of link prediction information  $\beta$  with various magnitude of missing link information  $\alpha$  in a random scale-free networks. The degree distribution follows  $p(k) = (\lambda - 1)m^{\lambda-1}k^{-\lambda}$ , where  $N = 1000$ ,  $\lambda = 2.5$ , and  $m = 2$ . The results are averaged over 100 independent realizations. The solid lines represent the “valid prediction area” (VPA) and the dash lines represent the “excessive prediction area” (EPA). The dash dotted lines are the reference lines, which represent the case of complete link information, namely  $\alpha = 0$ . The filled area represents the “surpassing prediction area” (SPA) where  $f_c$  is even lower than the case of complete link information.

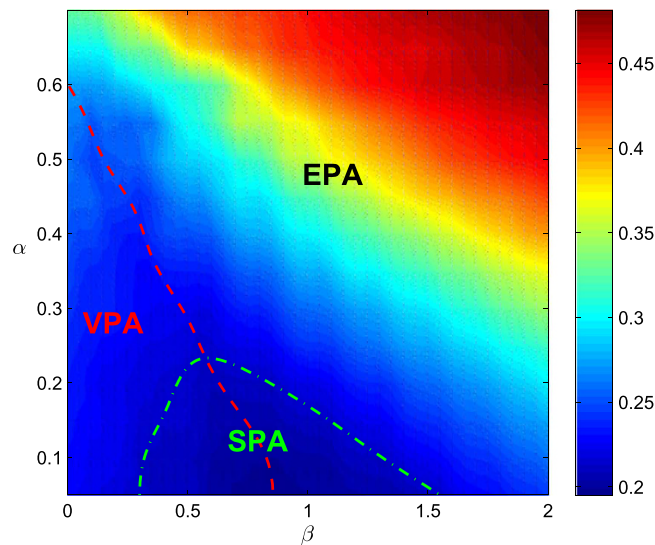
networks. The random scale-free networks with degree distributions  $p(k) = (\lambda - 1)m^{\lambda-1}k^{-\lambda}$  are generated by using the method proposed in ref. 38. In Fig. 3, we report the dependence of critical attack strength coefficient  $f_c$  on the magnitude of link prediction information  $\beta$ . We use resource allocation (RA) link prediction algorithm<sup>37</sup> to predict the missing links. For comparison, we also show the case of complete link information, i.e.  $\alpha = 0$ , which is usually considered as the ideal case.

From Fig. 3, we can see that with the increasing number of missing links, the  $f_c$  curve shifts gradually to top-left. For  $\alpha = 0.1$ ,  $\alpha = 0.3$  and  $\alpha = 0.5$ ,  $f_c$  first decreases with  $\beta$  and then increases after  $\beta > \beta^*$ . We call the region  $[0, \beta^*]$  the “valid prediction area” (VPA) and the region  $(\beta^*, \beta_{max})$  the “excessive prediction area” (EPA) where the inclusion of any additional predicted links will bring negative effects on the performance of network disintegration. To our surprise, we find an area in which the performance of our method is even better than the “ideal case” where the critical attack strength coefficient is  $f_c^0$ . We call the area “surpassing prediction area” (SPA), see Fig. 3(a). Figure 4(a) shows the performance of network disintegration under the optimal magnitude of link prediction information (i.e.,  $f_c^*$ ), along with the performance of network disintegration without link prediction (i.e.,  $\tilde{f}_c$  when  $\beta = 0$ ). The difference between  $f_c^*$  and  $\tilde{f}_c$  indicates the contribution of the additional links predicted by link prediction algorithm. We find that when  $\alpha < 0.24$ ,  $f_c^*$  is lower than  $\tilde{f}_c$ , which corresponds to the SPA. It can be explained that the link prediction amplifies the heterogeneity of node importance and reshape the network structure like drawing an exaggerated and characteristic comic. We refer to this phenomenon as the “comic effect” of link prediction. The values of  $f_c^*$  and  $\tilde{f}_c$  meet at  $\alpha = 0.6$ , implying that in some cases we can reconstruct the original network to improve the performance of network disintegration even when the network has about 60% links are missing.

It is worth pointing out that, when  $\alpha$  is large enough, see in Fig. 3(d) when  $\alpha = 0.7$ , there is no “valid prediction area” and  $\beta^* = 0$ . It suggests that link prediction will be counterproductive for the network disintegration performance if overmuch links are missing. The reason is that the link prediction accuracy is usually very low if the prediction based on the observed network with many missing links<sup>39</sup>. These results show that when the link information is not complete, a proper number of additional links can efficiently improve the performance of network disintegration and even obtain better performance (i.e., lower  $f_c$ ) than the case with complete information. It is true that the added links by link prediction may connect to wrong nodes and thus we may not recover the original network completely. However, through link prediction, we partly recover the ranking of node importance, which is really critical in network disintegration.



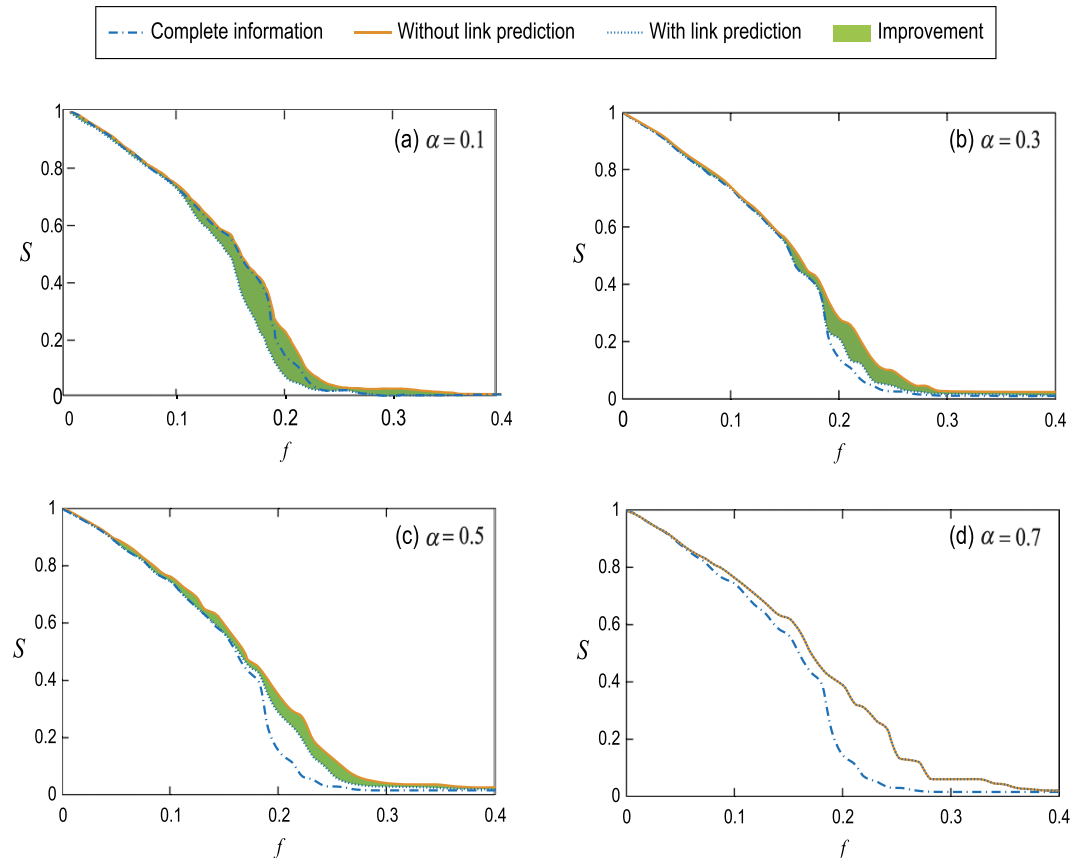
**Figure 4. The contribution of link prediction to the network disintegration.** (a) The optimal critical attack strength coefficient  $f_c^*$  (squares), comparing with the critical attack strength coefficient without link prediction  $f_c^0$  ( $\beta = 0$ ) (circles). The horizontal dash dot line presents the value of  $f_c^0 = 0.215$  obtained under complete information. (b) The optimal magnitude of link prediction information  $\beta^*$ . The original network is the same as the one we used in Fig. 3. The results are averaged over 100 independent realizations of link prediction.



**Figure 5. The critical attack strength coefficient  $f_c$  in the  $(\alpha, \beta)$  plane.** The original network is the same as the one we used in Fig. 3. The red dash line presents the optimal magnitude of link prediction information  $\beta^*$ . The left region and the right region of the red dash line are corresponding to the valid prediction area (VPA) and excessive prediction area (EPA), respectively. The area under the green dash dot line is the surpassing prediction area (SPA). The results are averaged over 100 independent realizations of link prediction.

We also show in Fig. 4(b) the optimal magnitude of link prediction information  $\beta^*$  as a function of the magnitude of missing link information  $\alpha$ . We find that  $\beta^*$  monotonically decreases with  $\alpha$  and eventually reaches to zero at about  $\alpha = 0.6$ , which suggests that the less links are missing, the more predicted links (usually with high accuracy) are required to be added to obtain the best effect. On the contrary, if more links are missing, the less predicted links are added because adding more links will lead to more mistakes due to the low accuracy of link prediction. The dependence of the critical attack strength coefficient  $f_c$  on parameter  $\alpha$  and  $\beta$  is shown in Fig. 5, where the VPA, EPA and SPA can be clearly partitioned.

The measure  $f_c$  is the critical fraction of nodes at which the network completely collapses. However, sometimes we are also interest in the case when the network suffers a big damage without completely collapsing. Figure 6



**Figure 6.** The relative size of giant component  $S$  versus attack strength coefficient  $f$  under attacks with complete link information (dash dot lines), attacks without link prediction (solid lines) and attacks with optimal link prediction information (dotted lines). The filled area demonstrates the improvement of the effect of network disintegration due to link prediction. The original network is the same network as in Fig. 3. For different  $\alpha$ , we set  $\beta = \beta^*$  as shown in Fig. 3. The results are averaged over 100 independent realizations of link prediction.

reports the fraction of nodes in the giant component after node attacks  $S$  as a function of attack strength coefficient  $f$  with various magnitude of missing link information  $\alpha$ . Here we set  $\beta = \beta^*$  for corresponding  $\alpha$ , namely  $\beta = 0.85$  for  $\alpha = 0.1$ ,  $\beta = 0.55$  for  $\alpha = 0.3$ ,  $\beta = 0.1$  for  $\alpha = 0.5$  and  $\beta = 0$  for  $\alpha = 0.7$ . The effect of network disintegration can be characterized by the area under the curve of  $S$ . The smaller the area is, the more efficient the network disintegration is. Therefore, the area between the curve of  $S$  with link prediction (dotted lines) and without link prediction (solid lines) demonstrates the improvement of the performance of network disintegration with the assistance of link prediction. The improvement of our method is significant for small  $\alpha$  and the “comic effect” of link prediction appears in the case of  $\alpha = 0.1$ , see Fig. 6(a).

**Experiments on real networks.** The study of disintegration is important for many real-world systems such as rumor spreading in online social networks, disease transmission through airlines and foodweb. To evaluate the performance of our method, we investigate four real-world networks: (i) the Political blogosphere network (PB)<sup>40</sup>, (ii) the network of the US air transportation system (USAir) (<http://toreopsahl.com/datasets/#usairports>), (iii) the Foodweb of south Florida during the wet season (Foodweb)<sup>41</sup> and (iv) the collaboration network between Jazz musicians (Jazz)<sup>42</sup>. Basic statistics of these networks are shown in Table 1. As we can see, all networks are well connected, with high clustering coefficients and short average distances.

We simulate the prediction and disintegration process on these networks, and results are shown in Fig. 7. All four networks exhibit similar pattern with the synthetic networks: the critical attack strength coefficients,  $f_c$ , all decrease at the beginning as the ratio of additional links increase, after an optimal ratio, the performance of disintegration degenerates while more links are added. It is interesting to observe that, all the four networks have a large “surpassing prediction area”, where  $f_c$  decreases to even below the value obtained under complete information. The SPA for the four networks are  $\beta \in [0.05, 2.35]$  for PB,  $\beta \in [0.15, 1.45]$  for USAir,  $\beta \in [0.15, 1.75]$  for Foodweb and  $\beta \in [0.05, 1.55]$  for Jazz.

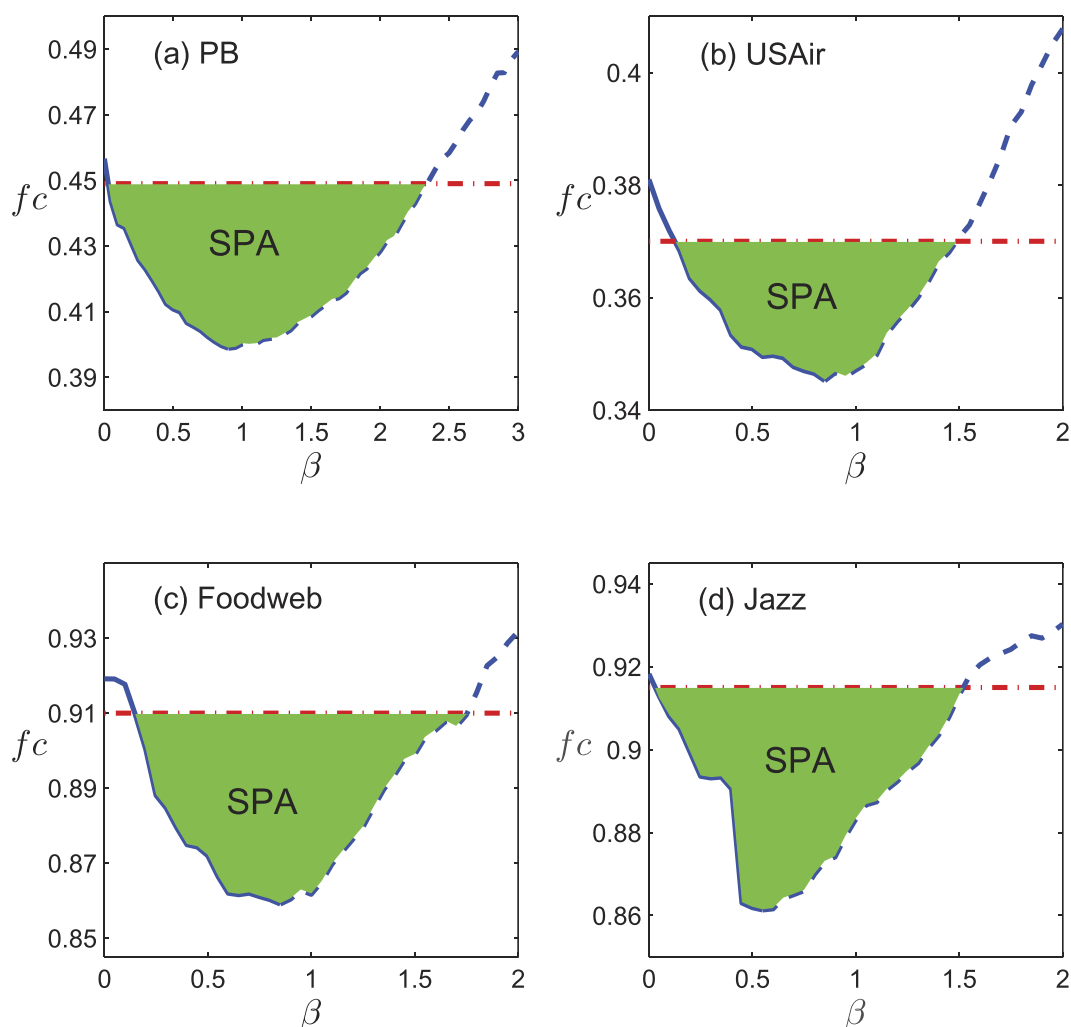
## Discussion

Network disintegration with incomplete link information is an important and challenging problem. In this paper, we introduced the link prediction as a strategy for attackers to improve the performance of network



Networks	$N$	$W$	$\langle k \rangle$	$C$	$r$	$\langle l \rangle$
PB	1222	16714	27.36	0.36	-0.221	3.65
USAir	1574	28236	35.8	0.384	-0.113	3.14
Foodweb	128	2106	32.906	0.312	-0.111	1.73
Jazz	198	2742	27.7	0.618	0.020	2.23

**Table 1. Basic statistics of four real networks.**  $N$  and  $W$  are the number of nodes and links.  $\langle k \rangle$  is the average degree;  $C$  is the clustering coefficient;  $r$  is the assortativity;  $\langle l \rangle$  is the average shortest distance.



**Figure 7. The critical attack strength coefficient  $f_c$  versus  $\beta$  with a certain missing information accuracy  $\alpha = 0.1$  in four real networks.** The solid lines represent the “valid prediction area” (VPA) and the dash lines represent the “excessive prediction area” (EPA). The dash dot lines are the reference lines, which represent the case of complete link information, namely  $\alpha = 0$ . The filled area represents the “surpassing prediction area” (SPA) where  $f_c$  is even lower than the case of complete link information. For each  $\beta$ , the result is averaged over 100 simulations.

disintegration. We showed that although the missing of link information harms the effect of network disintegration, link prediction can help to improve the performance remarkably. We found with surprise that if the magnitude of missing link information is not too large, the effect of network disintegration with the assistance of link prediction even can be better than the case of complete link information. We called this phenomenon the “comic effect” of link prediction. Although, the link prediction does not recover the missing information completely, but it reshapes the network just like an exaggerated but characteristic comic. As a result, the importance of the key nodes is emphasized by adding a number of predicted links. We believe that the comic effect of link prediction may exist in many backgrounds, not only in the network disintegration. For example, link prediction can not only help to improve the classification accuracy of partially labeled networks<sup>43</sup> but also be used in recommender

systems<sup>44</sup>. These useful applications demonstrate that hidden information revealed by link prediction can help to improve the accuracy of information filtering algorithms.

Moreover, we exposed the area of excessive prediction where the addition of more predicted links will give negative contribution. An optimal magnitude of link prediction information is obtained when the critical attack strength coefficient reaches the minimum. Beyond the optimal magnitude of link prediction information, the contribution of link prediction to the network disintegration will decrease and can even be negative. In addition, we found that the optimal magnitude of link prediction information decreases with the increasing of missing link information, indicating that when there are many missing links it should be very cautious to add new links. For real applications, how to obtain the optimal magnitude of link prediction information for real networks is still an open and challenging problem, as we usually don't know the portion of missing links and thus it's difficult to evaluate the algorithm's performance. According to the results in this paper, by adding a small number of predicted links is usually beneficial when the number of missing links is moderate. Future studies are required to evaluate the choice of appropriate link prediction algorithms to achieve better network disintegration performance<sup>45</sup>.

## Methods

**Algorithms for link prediction.** The link prediction problem has been a long-standing challenge in modern information era. Its main goal is to estimate the existence likelihood of nonobserved links based on the known topology and node attributes. The simplest index of link prediction is the common neighbors (CN) index which in common sense, two nodes,  $x$  and  $y$ , are more likely to have a link if they have many common neighbors<sup>46</sup>.

$$S_{xy}^{CN} = |\Gamma(x) \cap \Gamma(y)|. \quad (1)$$

where  $\Gamma(t)$  denotes the set of neighbors of node  $t$ .

Resource Allocation (RA) index<sup>37</sup> is an improved index based on CN, which assign less-connected neighbors more weight. The index is motivated by the resource allocation dynamics on networks. Consider a pair of nodes,  $x$  and  $y$ , which are not directly connected. The node  $x$  can send some resource to  $y$ , with their common neighbors being transmitters. The similarity between  $x$  and  $y$  can be defined as the amount of resource  $y$  received from  $x$ . The mathematical expressions are

$$S_{xy}^{RA} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{|\Gamma(z)|}, \quad (2)$$

**Performance measurement of network disintegration.** In the context of complex networks, the critical removal fraction of nodes  $f_c$  for the disintegration of networks is generally used to characterize the network robustness from the view of defenders. The larger  $f_c$  is, the more robust the network is. This measure emerged from the random graph theory and was stimulated by Albert *et al.*<sup>4</sup>. Instead of a strict extreme property, it considers statistically how the removal of nodes leads to a deterioration of network performance, and eventually to the collapse of the network at a given critical removal fraction  $f_c$ . The most common performance measurements include the diameter, the size of the largest component and the average path length. We choose  $\kappa \equiv \langle k^2 \rangle / \langle k \rangle^2$  as the criterion for the collapse of networks<sup>47,48</sup>, where the angular brackets  $\langle \cdot \rangle$  denote an ensemble average. After each node is removed, we calculate  $\kappa$ . When  $\kappa$  becomes less than 2, we record the number of nodes  $t$  removed up to that point. The threshold  $f_c$  is calculated as  $f_c = \langle t \rangle / N$ . Here we employ  $f_c$  to measure the effect of network disintegration strategy from the view of attackers. Smaller  $f_c$  implies more efficient network disintegration.

## References

1. Albert, R. & Barabási, A. L. Statistical mechanics of complex networks. *Reviews of Modern Physics* **74**, 47–51 (2002).
2. Newman, M. E. J. The structure and function of complex networks. *SIAM Review* **45**, 167–256 (2003).
3. Wang, X. F. Complex networks: Topology, dynamics and synchronization. *International Journal of Bifurcation and Chaos* **12**, 885–916 (2002).
4. Albert, R., Jeong, H. & Barabási, A. L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
5. Bollobás, B. & Riordan, O. Robustness and vulnerability of scale-free random graphs. *Internet Mathematics* **1**, 1–35 (2003).
6. Wu, J., Tan, Y.-J., Deng, H.-Z. & Zhu, D.-Z. Vulnerability of complex networks under intentional attack with incomplete information. *Journal of Physics A* **40**, 2665–2671 (2007).
7. Wu, J., Barahona, M., Tan, Y.-J. & Deng, H.-Z. Natural connectivity of complex networks. *Chinese physics letters* **27**, 078902 (2010).
8. Wu, J., Barahona, M., Tan, Y.-J. & Deng, H.-Z. Spectral measure of structural robustness in complex networks. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* **41**, 1244–1252 (2011).
9. Wu, Z.-X. & Holme, P. Onion structure and network robustness. *Physical Review E* **84**, 026106 (2011).
10. Shargel, B., Sayama, H., Epstein, I. R. & Bar-Yam, Y. Optimization of robustness and connectivity in complex networks. *Physical Review Letters* **90**, 068701 (2003).
11. Paul, G., Tanizawa, T., Havlin, S. & Stanley, H. E. Optimization of robustness of complex networks. *European Physical Journal B* **38**, 187–191 (2004).
12. Valente, A. X. C. N., Sarkar, A. & Stone, H. A. Two-peak and three-peak optimal complex networks. *Physical Review Letters* **92**, 118702 (2004).
13. Liu, J. G., Wang, Z. T. & Dang, Y. Z. Optimization of robustness of scale-free network to random and targeted attacks. *Modern Physics Letters B* **19**, 785–792 (2005).
14. Tanizawa, T., Paul, G., Cohen, R., Havlin, S. & Stanley, H. E. Optimization of network robustness to waves of targeted and random attacks. *Physical Review E* **71**, 047101 (2005).
15. Beygelzimer, A., Grinstein, G. E., Linsker, R. & Rish, I. Improving network robustness by edge modification. *Physica A* **357**, 593–612 (2005).
16. Schneider, C., Moreira, A., JS, A., Havlin, S. & Herrmann, H. Mitigation of malicious attacks on networks. *Proceedings of the National Academy of Science of the United States of America* **108**, 3838 (2011).
17. Morone, F. & Makse, H. A. Influence maximization in complex networks through optimal percolation. *Nature* **524**, 65–68 (2015).
18. Raab, J. & Milward, H. B. Dark networks as problems. *Journal of Public Administration Research and Theory* **13**, 413–439 (2003).



19. Kobayashi, T. & Hasui, K. Efficient immunization strategies to prevent financial contagion. *Scientific Reports* **4**, 3834 (2014).
20. Tripathy, R. M., Bagchi, A. & Mehta, S. A study of rumor control strategies on social networks. In *International Conference on Information and Knowledge Management*, 1817–1820 (2010).
21. Quayle, A. P., Siddiqui, A. S. & Jones, S. J. M. Preferential network perturbation. *Physica A* **371**, 823–840 (2006).
22. Holme, P., Kim, B. J., Yoon, C. N. & Han, S. K. Attack vulnerability of complex networks. *Physical Review E* **65**, 056109 (2002).
23. Chen, Y., Paul, G., Havlin, S., Liljeros, F. & Stanley, H. E. Finding a better immunization strategy. *Physical Review Letters* **101**, 058701 (2008).
24. Schneider, C., Mihaljev, T., Havlin, S. & Herrmann, H. Suppressing epidemics with a limited amount of immunization units. *Physical Review E* **84**, 061911 (2011).
25. Dezső, Z. & Barabási, A.-L. Halting viruses in scale-free networks. *Physical Review E* **65**, 055103 (2002).
26. Li, J., Li, Y., Deng, H.-Z. & Tan, Y.-J. Optimal attack strategy in random scale-free networks based on incomplete information. *Chinese Physics Letters* **28**, 068902 (2011).
27. Cohen, R., Havlin, S. & ben Avraham, D. Efficient immunization strategies for computer networks and populations. *Physical Review Letters* **91**, 247901 (2003).
28. Holme, P. Efficient local strategies for vaccination and network attack. *Europhysics Letters* **68**, 908–914 (2004).
29. Gomez-Gardenes, J., Echenique, P. & Moreno, Y. Immunization of real complex communication networks. *European Physical Journal B* **49**, 259–264 (2006).
30. Gallos, L. K., Liljeros, F., Argyrakis, P., Bunde, A. & Havlin, S. Improving immunization strategies. *Physical Review E* **75**, 045104 (2007).
31. Estrada, E. Food webs robustness to biodiversity loss: The roles of connectance, expansibility and degree distribution. *Journal of Theoretical Biology* **244**, 296–307 (2007).
32. Carley, K. M. Destabilization of covert networks. *Computational and Mathematical Organization Theory* **2**, 51–66 (2006).
33. Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E. & Aberg, Y. The web of human sexual contacts. *Nature* **411**, 907–908 (2001).
34. Shen, J. W. *et al.* Predicting protein-protein interactions based only on sequences information. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 4337–4341 (2007).
35. Menche, J. *et al.* Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 1257601 (2015).
36. Lü, L. & Zhou, T. Link prediction in complex networks: A survey. *Physica A* **390**, 1150–1170 (2011).
37. Zhou, T., Lü, L. & Zhang, Y.-C. Predicting missing links via local information. *European Physical Journal B* **71**, 623–630 (2009).
38. Molloy, M. & Reed, B. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms* **6**, 161–179 (1995).
39. Liu, W. & Lü, L. Link prediction based on local random walk. *Europhysics Letters* **89**, 58007 (2010).
40. Adamic, L. A. & Glance, N. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of 3rd International Workshop on Link Discovery*. 36–43 (ACM Press, 2005).
41. Ulanowicz, R. E., Bondavalli, C. & Egnotovich, M. S. Network analysis of trophic dynamics in south florida ecosystem, fy 97: The florida bay ecosystem. *Technical Report Ref. No. [UMCES/CBL 98–123* (1998).
42. Gleiser, P. M. & Danon, L., Community structure in Jazz. *Advances in Complex Systems* **4**, 565–573 (2003).
43. Zhang, Q.-M., Lü, L. & Ming-Sheng, S. Similarity-based classification in partially labeled networks. *International Journal of Modern Physics C* **21**, 813–824 (2010).
44. Lü, L. *et al.* Recommender systems. *Physics Reports* **519**, 1–49 (2012).
45. Lü, L., Pan, L., Zhou, T., Zhang, Y.-C. & Stanley, H. E. Toward link predictability of complex networks. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 2325–2330 (2015).
46. Liben-Nowell & D, K. J. The link prediction problem for social networks. *Journal of the American Society for Information Science and Technology* **58**, 1019–1031 (2007).
47. Cohen, R., Erez, K., ben Avraham, D. & Havlin, S. Breakdown of the internet under intentional attack. *Physical Review Letters* **86**, 3682–3685 (2001).
48. Cohen, R., Erez, K., ben Avraham, D. & Havlin, S. Resilience of the internet to random breakdowns. *Physical Review Letters* **85**, 4626–4628 (2000).

## Acknowledgements

J.W. acknowledges the Natural Science Foundation of China under Grant No. 71371185 and the Program for New Century Excellent Talents in University under Grant No. NCET-12-0141. L.L. acknowledges the Natural Science Foundation of China under Grant No. 11205042, Research Start-up Fund of Hangzhou Normal University under Grant PE13002004039, Zhejiang Provincial Natural Science Foundation of China under Grant No. LR16A050001 and the EU FP7 Grant 611272 (project GROWTHCOM). X.L. acknowledges the Natural Science Foundation of China under Grant Nos. 71301165 and 71522014. We thank Qiang Luo for useful discussions and comments.

## Author Contributions

J.W. and L.L. designed research; S.-Y.T., J.W., L.L., X.L. and M.-J.L. performed research; S.-Y.T. conducted the experiments; J.W., S.-Y.T., L.L. and X.L. analysed the results; J.W., S.-Y.T., L.L. and X.L. wrote the paper.

## Additional Information

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Tan, S.-Y. *et al.* Efficient network disintegration under incomplete information: the comic effect of link prediction. *Sci. Rep.* **6**, 22916; doi: 10.1038/srep22916 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>