Soc. Netw. Anal. Min. (2013) 3:939–952
DOI 10.1007/s13278-013-0118-8

ORIGINAL ARTICLE

Quantifying topological robustness of networks under sustained targeted attacks

Mahendra Piraveenan · Gnana Thedchanamoorthy · Shahadat Uddin · Kon Shing Kenneth Chung

Received: 14 December 2012 / Revised: 28 March 2013 / Accepted: 20 May 2013 / Published online: 7 June 2013

Abstract In this paper, we introduce a measure to analyse the structural robustness of complex networks, which is specifically applicable in scenarios of targeted, sustained attacks. The measure is based on the changing size of the largest component as the network goes through disintegration. We argue that the measure can be used to quantify and compare the effectiveness of various attack strategies. Applying this measure, we confirm the result that scale-free networks are comparatively less vulnerable to random attacks and more vulnerable to targeted attacks. Then we analyse the robustness of a range of real world networks, and show that most real world networks are least robust to attacks based on betweenness of nodes. We also show that the robustness values of some networks are more sensitive to the attack strategy as compared to others. Furthermore, robustness coefficient computed using two centrality measures may be similar, even when the computational complexities of calculating these centrality measures may be different. Given this disparity, the robustness coefficient introduced potentially plays a key role in choosing attack and defence strategies for real world networks. While the measure is applicable to all types of complex networks, we clearly demonstrate its relevance to social network analysis.

Keywords Complex networks · Robustness · Social networks

M. Piraveenan (✉) · G. Thedchanamoorthy · S. Uddin · K. S. K. Chung
Centre for Complex Systems Research, Faculty of Engineering and IT, The University of Sydney, Sydney , NSW 2006, Australia
e-mail: mahendrarajah.piraveenan@sydney.edu.au

G. Thedchanamoorthy
e-mail: gthe3845@uni.sydney.edu.au

S. Uddin
e-mail: shahadat.uddin@sydney.edu.au

K. S. K. Chung
e-mail: kenneth.chung@sydney.edu.au

1 Introduction

The study of complex networks is a dominant trend in recent research that transcends domain boundaries. The ability of a network to perform its intended function depends on how it responds to pressures—both internal and external. Such pressures could include errors, random attacks, targeted attacks based on some criteria, and malevolent and sustained attacks which remove nodes in sequence. The ability of a network to withstand such pressures has been variously called error tolerance, attack tolerance, resilience or robustness of a network, depending on the context (Albert et al. 2000; Crucittia et al. 2004; Venkatasubramanian et al. 2004; Dekker and Colbert 2004; Ng and Efstathiou 2006; Costa et al. 2007). In this paper, we are interested in the ability of a network to resist complete topological disintegration in the face of random or targeted node removals. We will call this topological/structural robustness, or simply robustness of a network. It has been shown that such ability depends on the topological structure of the network. For example, scale-free networks are more resilient against random attacks, but more vulnerable to targeted attacks, compared to Erdos–Renyi random networks (Albert et al. 2000). Small-world structure (Milgram 1967; Zaidi(2013) found in networks, hierarchical structure (Gilbert et al. 2011) as well as community structure (Rees and Gallagher 2012; Cazabet et al. 2012) may also influence resilience. It can be

immediately seen that quantifying such resilience (robustness) of a network is vital in a number of disciplines. For example, computer networks should be designed in such a way that they should function properly when some nodes (routers or hosts) fail, by technical faults or under attack. On the other hand, in a network of terrorist cells, we might be interested in the best strategy to attack the network so as it is disabled and disintegrated as quickly as possible. Therefore, measuring and comparing the robustness of networks under various failure and attack scenarios is of vital importance.

Albert et al. (2000) first studied the robustness of networks by comparing random and scale-free networks. Let us briefly recall the definition of scale-free networks here. Scale-free networks are those networks that display similar topological features irrespective of scale. Such networks are described by power law degree distributions, formally specified as

$$p_k = Ak^{-\gamma}U(k/k_{\max}) \quad (1)$$

where U is a step function specifying a cut off at $k = k_{\max}$ (Dorogovtsev and Mendes 2003). Albert et al. considered a number of metrics, including network diameter, the size of the largest component, and the average size of the rest of the components, to study network robustness. Since then, a host of measures have been proposed and used to understand the structural robustness of networks against attacks (Crucittia et al. 2004; Costa et al. 2007; Dekker and Colbert 2004; Ng and Efstathiou(2006). However, as we will discuss in the next section, these measures have some drawbacks. They are either (1) based upon the averaging of single-node removals, or (2) only consider the point in time where a phase transition occurs (in terms of quantities such as the size of the largest component) before the network begins to disintegrate[1]. However, as we will show, it is important to consider the entire history of the disintegration process to understand the structural robustness of networks. Albert et al. (2000) consider this history by using plots of network diameter etc., versus fraction of nodes removed. In this paper, we introduce a single measure which can capture information from such plots. Therefore, our measure is suitable to analyse the topological robustness of networks under sequential node removals. We are particularly interested in applying this measure to sustained targeted attack scenarios.

Since topological robustness is a concept applicable in many domains of complex networks, we will take a generic approach in this paper and draw examples from all possible domains. However it is easy to see that the measure introduced will be relevant to a number of social networks. As mentioned above, we might be interested in disintegrating a terrorist cell network, and for this purpose might want to consider its robustness against different attack strategies. We might want to break up a 'contact network' of people in an epidemiological scenario, where 'attacking' a node simply means vaccinating a person and thereby removing that person from the contact network, so that the spread of infection is slowed down. On the other hand, there are scenarios where a social network might have to be defended against malicious forces trying to disintegrate it: for example, an undemocratic government trying to break up an online social network, to prevent vital news from spreading. In all these social network scenarios, it is important to measure the topological robustness of the social network, so that best attack and defence strategies might be devised.

The rest of the paper is organised as follows: In the next section, we will review the existing structural robustness measures, and analyse their utility in understanding sequential node removal. In the following section we will introduce our robustness measure. We will then apply this measure to synthesised and real world networks. Confirming the result of Albert et al. (2000) regarding scale-free and random networks, we will then consider networks under a range of sequential node removal strategies, including random node removal and targeted attacks, choosing nodes based on degree, betweenness centrality and closeness centrality. Finally, we will present our conclusions.

2 Existing measures of topological robustness

Albert et al. (2000) considered error and attack tolerance of complex networks in the following manner. They removed nodes from complex networks one by one until all nodes are extracted (we will call this 'sustained attack', as opposed to an attack where only a portion of the nodes are ever removed), and studied the variation of topological properties in networks due to these removals. They removed nodes in two separate orders (1) random order (2) and ordered by degree (highest degree first). They analyzed the following three topological properties:

1. Network diameter
2. The size of the largest component
3. The average size of the rest of the components

The analysis of (Albert et al. 2000) was undertaken by constructing plots of the above three properties against the proportion of nodes removed. By doing this, they showed that random topologies are more vulnerable to random node removals, whereas scale-free networks showed remarkable resilience to such random removals. Therefore,

[1] We will sometimes refer to this phenomena simply as 'phase transition', when the context is clear.

in the case of scale-free networks, the size of the largest component decreased slowly, the average network diameter increased slowly, and the average size of the rest of the components converged towards unity slowly, compared to random networks. However, when networks are subjected to targeted sustained attacks, the outcome differed. Random networks performed very similarly under random attacks and targeted attacks. However, scale-free networks under sustained targeted attacks demonstrated signs of quick disintegration: such as rapidly decreasing size of the largest component, rapidly increasing average diameter, and the average size of the rest of the components converging quickly towards unity. Therefore, compared to random networks, scale-free networks were shown to have more resilience towards random attacks and less resilience towards targeted attacks.

Albert et al. (2000), however, relied on plotting metrics versus fraction of nodes removed, rather than a single robustness measure, to demonstrate these facts. Following their work, a plethora of metrics have been proposed to measure the topological robustness of networks as a single measure (Costa et al. 2007; Venkatasubramanian et al. 2004; Dekker and Colbert 2004; Ng and Efstathiou (2006). However, they typically calculate averaged effects of single-node removals, rather than effects of sequential removals, or are too simplistic. For example, the *network efficiency* has been defined as the average of inverted shortest path lengths (Crucittia et al. 2004), and used for quantifying the robustness of a network. Node removals are not explicitly considered in this measure. Another measure derived from network efficiency is *network vulnerability* (Costa et al. 2007). The vulnerability V_i of a given node i is calculated as $(E - E_i)/E$, where E is the efficiency of the network and E_i is the efficiency after a given node i has been removed. Network vulnerability can be calculated as an average of individual vulnerability values of nodes, and therefore relies on averaging over individual node removals. Similarly, Venkatasubramanian et al. (2004) define the *structural robustness* measure for networks. They arrive at this by defining a series of other metrics. First they define the *accessibility* of a node as the number of nodes that are reachable from this node. The *effective accessibility* of a graph is then defined, inter alia, as the sum of accessibilities of nodes in that graph. Then they define the *structural robustness* with respect to vertex j of a graph as the ratio of the effective accessibility of the graph S_j obtained by deleting vertex j from the original graph to the maximum possible effective accessibility. Finally, *average-case structural robustness* of a graph is defined as the average of the structural robustness values computed over all the vertices (Venkatasubramanian et al. 2004). While it is not our purpose here to explain the measure, it suffices to say that this measure relies on averaging over single-node removals.

Meanwhile, Dekker and Colbert (2004) introduced two concepts of connectivity for a graph which can be used to model network robustness: the *node connectivity* and *link connectivity*, which are the smallest number of nodes and links respectively, whose removal results in a disconnected or single-node graph. While these measures are applicable for networks under sustained attack, they are measured at a single time point when network disintegration (phase transition in terms of the size of the largest component) occurs. Ng and Efstathiou (2006) use two measures, average shortest path and network diameter, to quantify network robustness.

Indeed there is a substantial body of work which introduces and analyses structural robustness measures, and it is beyond the scope of this paper to review them all. However, it can be seen that none of these measures are ideal for our purpose of understanding a network's resilience under sustained attack (sequential node removal). Some of these measures, such as network efficiency, do not explicitly consider node removal at all. Other measures consider averages of single-node removals, such as vulnerability (Costa et al. 2007). Yet other measures, such as those proposed by (Dekker and Colbert 2004), consider sustained node removals, but are concerned with only the time point when the phase transition occurs and network disintegrates. However, it could be argued that how well the network resists further disintegration after this phase transition is also important in determining its structural robustness. We are therefore interested in proposing a measure which is applicable to analysing network robustness under sustained attacks, and while being a single measure, holistically captures the fragmenting behaviour of the network not just at phase transition but before and after that as well.

3 Definition of the robustness coefficient

To propose a robustness measure which is a single measure, let us first observe the following: Among the metrics that were used by Albert et al. to construct their plots, network diameter has the disadvantage that, as soon as the network fragments, it becomes infinity. Therefore it could not be further used to quantify the robustness of a fragmenting network. The average size of the rest of the components, $\langle s \rangle$, is small (or even zero) initially, and increases and peaks at the time when the network starts to disintegrate. If the network is attacked further, it decreases again and converges towards unity. As we will demonstrate later, it is however desirable to consider a metric which shows a similar trend throughout the fragmentation process. The size of the largest component, denoted as S, is such a metric. It is always finite, and always decreases

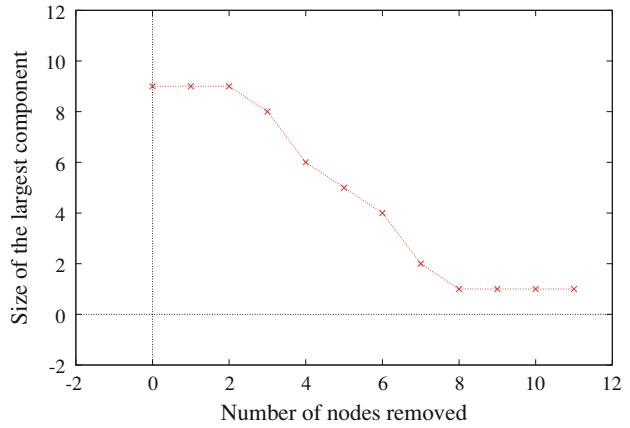

Fig. 1 Size of largest component against the number of nodes removed for a network under sustained targeted attack

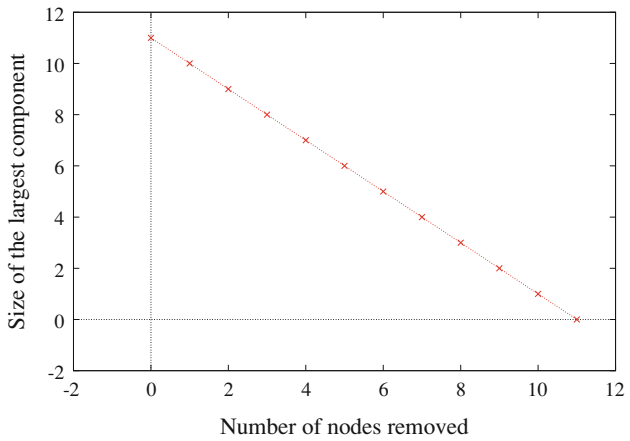

Fig. 2 Size of largest component against the number of nodes removed for an ideally robust network under targeted attack

during a persistent attack. It starts from N the network size, or a value $S_0 < N$, and decreases and converges towards unity. We will therefore use this to define our robustness coefficient.

Now let us consider the size of the largest component of a network which is under persistent attack. The largest component size S versus the number of nodes removed (or number of time steps) plot may look like Fig. 1 for a small network[2] with $N = 11$.

Now consider an 'ideal' network in terms of robustness under sustained attack. If the network resists disintegration, the S of such network should decrease by unity each time a node is removed. That is, the size of largest component decreases only by the removed node, while all other nodes remain part of the largest (single) component until they are themselves removed. S will become unity only when all nodes except one have been removed/destroyed. The S versus nodes removed plot of such an ideal network with $N = 11$ may look like Fig. 2.

We propose that the ratio of areas under such two plots define the topological robustness under sustained attack for any network. The reasoning behind this formulation is that, for an ideally robust network the size of the largest component will decrease linearly, while the more non-robust the network is, the quicker it will collapse, and the change in the size of the largest component will reflect this collapse.

By considering trapeziums of unitary width along the x axis, the area under first curve could be estimated as:

$$A_1 = 0.5(S_0 + S_1) + 0.5(S_1 + S_2) + \cdots + 0.5(S_{N-1} + S_N) = 0.5S_0 + \sum_{k=1}^{N-1} S_k + 0.5S_N$$

where S_k is the size of the largest component after k nodes are removed.

Here S_0 is the initial largest component size. Since S_N, the size of the largest component after N nodes are removed, is by definition zero, we may say that

$$A_1 = \sum_{k=0}^{N} S_k - 0.5S_0 \tag{2}$$

Meanwhile, the area under the second curve would be given (by considering a triangle of base N and height N) by

$$A_2 = (1/2)N^2 \tag{3}$$

because after each node removal, the size of the largest component will reduce only by a single node.

Therefore, we define robustness coefficient R as:

$$R = \frac{A_1}{A_2} = \frac{2\sum_{k=0}^{N} S_k - S_0}{N^2} \tag{4}$$

Since R is always less than unity and often is a very small (non-negative) quantity, it is suitable to define it as a percentage. That is:

$$R = \frac{A_1}{A_2} = \frac{200\sum_{k=0}^{N} S_k - 100S_0}{N^2} \tag{5}$$

It can be verified that the above definition gives $R = 100$ % for a fully connected network of any size, as expected. It should be noted here that similar area under the curve (AUC) measures are used in a number of disciplines. For example, in signal detection theory, the area under a receiver operating characteristic (ROC) curve (Hanley and Mcneil 1982) denotes the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. The curve is generated by plotting the fraction of true positives out of the positives

[2] Of course, the exact size of the largest component at each time step will depend on the network topology and the type of attack. The figure only shows a typical case, to be contrasted with Fig. 2.

against the fraction of false positives out of the negatives, therefore both quantities are fractions. In mechanics, the area under curve of a velocity versus time plot of a moving object denotes the distance that it has travelled (Kreyszig 2005). In a number of other disciplines, the AUC measure of a plot has physical meaning or interpretation.

It can be easily noted that the quicker a network begins to disintegrate, the smaller the robustness coefficient will be. However, it also captures the largest component curve before and after the phase transition point by considering the area under this curve. Therefore this measure is superior to the phase-transition time T_{pt} which only indicates the time-steps (or number of nodes) needed before a network begins to disintegrate.

4 Measuring attack tolerance using the robustness coefficient

In this section we demonstrate the use of our measure, by applying it to a number of synthesized and real-world networks.

4.1 Robustness coefficient of synthesized networks

First of all we confirm the results of (Albert et al. 2000) by comparing synthesized random and scale-free networks using our robustness measure. For this purpose, we used 10 synthesized scale-free networks, and 10 synthesized Erdos–Renyi random networks, each with 500 nodes and 1,200 links, but wired uniquely. We undertook sequential node removal in (1) random order (2) and ranked by node degree (highest degree first). We measured the robustness coefficient in each scenario. Our results are shown in Fig. 3.

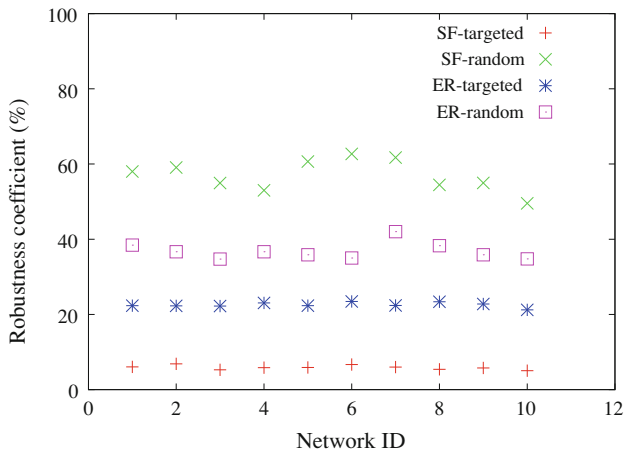

Fig. 3 Error and attack tolerance of random and scale-free networks, measured using the robustness coefficient. The *squares* denote Erdos–Renyi random networks under random node removals, whereas the *stars* denote Erdos–Renyi random networks under ordered node removals. The *crosses* denote scale-free networks under random node removals, where as the *plusses* denote scale-free networks under ordered node removals. It could be seen that the robustness coefficient is comparatively high for scale-free networks under random node removals. However, when ordered attacks are considered, the robustness coefficient for scale-free networks is comparatively low

We may see from Fig. 3 that scale-free networks have comparatively higher robustness coefficient values for random node removal. However, when ordered node removals (targeted attacks) based on node degree are considered, the robustness coefficient values of scale-free networks are lower as compared to random networks. The average R for Erdos–Renyi networks are 36.83 % under random attacks and 22.54 % under targeted (hub-based) attacks, whereas the average R for scale-free networks are 56.91% under random attacks and 5.88 % under targeted (hub-based) attacks. This simple example confirms the results obtained by (Albert et al. 2000) using plots of networks metrics, and demonstrates the utility of using the robustness coefficient to better quantify such results.

4.2 Robustness coefficient of real-world networks

Now we consider the robustness coefficient of some real world networks. We consider both directed and undirected networks, including gene regulatory networks, transcription networks, cortical networks, neural networks, food webs, Internet Autonomous Systems (AS) networks and citation networks. An explanation is necessary to some of these types of networks, since the usage of names can be ambiguous. In our transcription networks, nodes are regulatory genes and regulated proteins, and the links are the interactions between them Kepes (2007). These are bipartite and directed networks. On the other hand, by gene regulatory networks we mean networks where nodes are genes, and the links are the inhibitory or inducing effects of one gene on the expression of another gene Alon (2007). Note the subtlety that unlike transcription networks, only genes are considered as nodes in these directed networks. Similarly, by *cortical networks* we denote the networks of dependencies between various regions of the cerebral cortex (in a set of primates) (Tang et al. 2008, 4). The nodes are regions in the cortex, and the links are functional dependencies. Note that the nodes are *not* individual neurons. On the other hand, neural networks are networks where nodes are individual neurons belonging to an organism's neural system and links are anatomical connections between neurons (Junker and Schreiber 2008). In citation networks, nodes are research papers (or other citable documents) and links denote citations between these documents. In food webs, nodes are organisms in an ecosystem and the links represent predator–prey relationships between them (Solé and Valverde 2004). These networks

can be considered undirected or directed (prey–predator). In internet AS networks, the nodes represent an autonomous system present in the Internet and the edges represent a commercial agreement between two Internet Service Provides (who own the two ASs). Such an agreement defines whether they agree to exchange data and how to charge each other (Piraveenan et al. 2009).

4.3 Ordering of targeted attacks

It has been demonstrated that most of the networks described above, as well as a vast number of other real world networks, are scale-free (Dorogovtsev and Mendes 2003; Kepes 2007; Alon 2007; Albert and Barabási 2002; Solé and Valverde 2004). Therefore, it could be expected that they will all be quite resilient against random attacks, and display high robustness coefficients under sequential random removal of nodes. For this reason, we have mainly considered targeted attacks in determining their robustness coefficients, in the following analysis. However, we have considered three different orderings for such targeted attacks, namely (1) degree-based ordering (2) betweenness-centrality-based ordering (3) and closeness-centrality-based ordering. By considering these three orderings, we intend to demonstrate that the robustness coefficient can be a valuable tool in choosing a strategy to attack/defend a network.

The three metrics mentioned above are properties of individual nodes. The degree of a node, as is well known, is the number of links it possesses. The Betweenness Centrality measures the fraction of shortest paths that pass through a given node, averaged over all pairs of node in a network. It is formally defined, for a directed graph, as

$$\mathrm{BC}(v) = \frac{1}{(N-1)(N-2)} \sum_{s \neq v \neq t} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \tag{6}$$

where $\sigma_{s,t}$ is the number of shortest paths between source node s and target node t, while $\sigma_{s,t}(v)$ is the number of shortest paths between source node s and target node t that pass through node v. On the other hand, closeness centrality is a measure of how long it will take for information to spread from a given vertex to all others in the network (Newman 2005). It essentially measures the average geodesic distance between a given node and all other nodes in the network It is defined as

$$\mathrm{CC}(v) = \frac{1}{\sum_{i \neq v} d_g(v,i)} \tag{7}$$

where d_g (v, i) is the shortest path (geodesic) distance between nodes v and i. Note that the average is 'inverted' so that the node which is 'closest' to all other nodes will have the highest measure of closeness centrality. We used these three metrics to order nodes which are subjected to targeted sequential removal, and measured the robustness coefficient under each scenario.

As an example of our results, we show the disintegration profiles for the six food webs that we analysed in Figs. 4, 5, and 6 for the three types of attack orderings that we considered (degree-based, betweenness-based, and closeness-based orderings). It can be seen that, the higher the area under these plots are, the larger the robustness coefficient is. However, it should be noted that the network sizes are different, and some plots have more points in them compared to others.

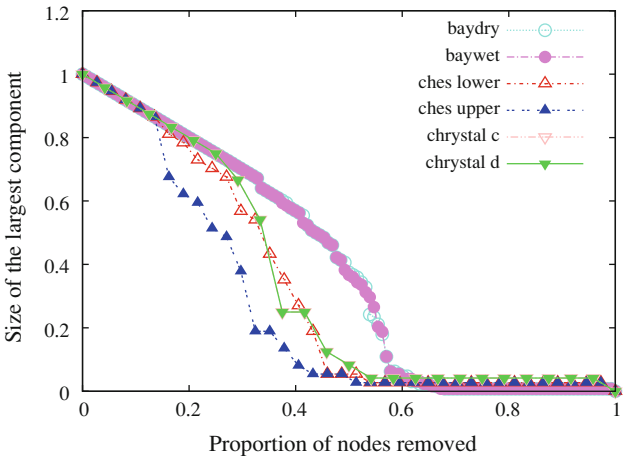

Fig. 4 The size of the largest components (as a proportion of network size) against the number of nodes removed (also as a proportion of network size) for six food webs. The nodes to be removed were chosen in the order of node degree. Note that the *plots* for Chrystal C and Chrystal D foodwebs overlap

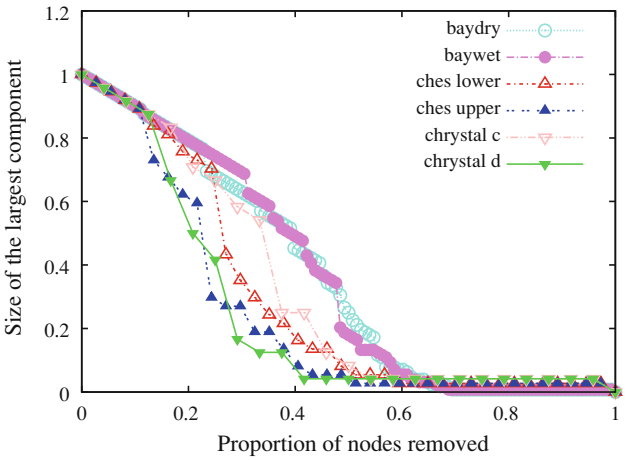

Fig. 5 The size of the largest components (as a proportion of network size) against the number of node removed (also as a proportion of network size) for six food webs. The nodes to be removed were chosen in the order of betweenness centrality

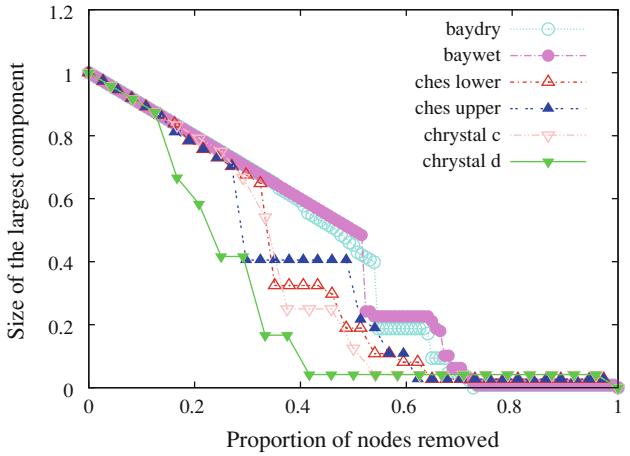

Fig. 6 The size of the largest components (as a proportion of network size) against the number of nodes removed (also as a proportion of network size) for six food webs. The nodes to be removed were chosen in the order of closeness centrality

4.4 Summary of results

A list of real world networks that we have studied is shown in Table 1. The robustness coefficient values, for the three types of attacks mentioned above, for all considered networks is also summarised in Table 1. We may note from this table, in general, that cortical networks, neural networks, food webs, and a few miscellaneous networks including primary school interactions, show the highest robustness, while Internet AS networks are among the least robust to targeted attacks. This itself is a significant result, showing that man-made or rapidly evolving networks are likely to be less robust compared to biological networks evolved over millennia. In some instances, such as in cortical networks, the high robustness could be explained by the large links to node (M/N) ratio, or network density. For example, the human cortical network has $N = 994$ nodes and $M = 27{,}040$ links, giving it a density of 27.20 links per node. Similarly, we may see that the network density is quite high for the primary school interactions (30.00 and 23.27 for the two networks, respectively). In other cases, topological design is responsible for the robustness. For example, the Chrystal C food web has a density of 5.12 links per node, yet shows a robustness coefficient of above 60% for all types of attack, which is remarkable. We postulate that this is probably the case if the network has a low scale-free exponent.

In Table 2, we show the amount of physical time taken (in seconds) to simulate complete decomposition of each of these networks under sustained targeted attacks. Obviously, this time depends on the computer system used to simulate the decomposition process. However, given that we used the same computing machine to simulate all decomposition, this time can be an indicator of how well the network

Table 1 Robustness coefficients of real world networks against targeted persistent (sequential) attacks

No	Network	R(D) (%)	R(BC) (%)	R(CC) (%)
	Neural networks			
1	*C. elegans*	58.83	41.79	55.92
	Internet AS networks			
2	Internet AS-1998	1.4	1.8	12.99
3	Internet AS-1999	1.71	1.91	13.34
4	Internet AS-2000	2.14	2.26	14.45
	Transcription networks			
5	*C. glutamicum*	1.52	1.35	13.12
6	*E. coli*	1.64	1.46	11.95
	Cortical connectivity networks			
7	Human	83.50	51.38	71.01
8	Cat	88.14	69.21	81.33
9	Macaque	69.27	52.65	69.19
	Gene regulatory networks			
10	*R. norvegicus*	53.64	22.52	50.45
11	*C. elegans*	24.45	14.25	32.93
12	*A. thaliana*	8.35	5.91	31.18
	Foodwebs			
13	Bay dry	77.72	71.73	83.26
14	Bay wet	77.64	72.07	86.33
15	Chess upper	60.92	55.08	75.38
16	Chess lower	49.23	45.58	52.30
17	Chrystal C	62.50	60.42	71.18
18	Chrystal D	44.79	44.10	47.57
	Citation networks			
19	Smart grid	28.49	24.80	37.03
20	Small world	20.71	14.09	25.70
	Metabolic networks			
21	Human	4.93	1.94	35.68
22	Rhesus monkey	3.54	1.56	32.54
23	Chimpanzee	3.83	1.57	33.34
24	*Acholeplasma laidlawii*	2.30	1.29	31.95
25	*Ashbya gossypii*	5.46	3.67	35.42
26	*Acaryochloris marina*	4.44	3.98	16.27
	PPI			
27	*H.Pylori*	14.43	13.59	27.69
28	Human	2.45	1.31	26.89
29	Mouse	1.35	1.22	7.85
	US Airlines			
30	US Air lines-1997	22.06	15.02	35.45
	Collaboration networks			
31	Netscience	1.68	1.79	27.21
	Cell signalling networks			
32	CA 1	20.93	16.66	32.44
	Miscellaneous networks			
33	Dolphins	45.32	32.57	53.43
34	Jazz Musicians	78.60	50.08	76.72

Table 1 continued

No	Network	R(D) (%)	R(BC) (%)	R(CC) (%)
35	Karate	26.30	24.05	38.75
36	Pharmaceutical	31.58	18.86	53.57
37	Primary school interaction 1	94.04	82.68	95.21
38	Primary school interaction 2	94.22	78.33	88.90
39	Sixteen story hospital	30.69	23.27	59.39
40	Software	17.12	14.63	30.71
41	Vehicle	48.06	40.68	53.90

Three modes of node removal are used, namely (a) degree-based removal (b) betweenness-based removal and (c) closeness-based removal. It can be seen that some networks have nearly equal robustness against all removal strategies, whereas the robustness of other nodes can vary greatly. The network data is taken from (Collections of connectivity data on the Macaque brain 2009; Baumbach 2007; Michigan Molecular Interaction Database (2008; Watts and Strogatz 1998; Pajek datasets (2007; Zachary 1977; Gleiser and Danon 2003; Primary school cumulative networks 2011; Lusseau et al. 2003; New England complex systems institute research projects. URL:http://necsi.edu/projects/braha/largescaleengineering.html) *R(D)* robustness under degree-based targeted attacks, *R(BC)* robustness under betweenness-based targeted attacks, *R(CC)* robustness under closeness-based targeted attacks

resists sustained targeted attacks. We used a MacBook Pro i5 machine with 4GB RAM to simulate the decomposition process.

Obviously, the time it takes to completely destroy the network depends on its initial size as well. Therefore, we normalized the time by network size, computing the average time it took (in seconds) for a node to get destroyed. While this measure is still not a sound indicator of a network's robustness because the proxy measure used here is the execution time of node destruction, it can be obtained without additional computation and gives useful supplementary information. The Table 2 shows that the following networks have relatively high 'destruction time' per node: Human cortical, Rat GRN, Smart Grid citation, Sixteen hospital, and the primary school interactions. We may see that there is strong correlation between networks of high robustness and networks that have high per-node destruction times.

However, it is of more interest to us, in demonstrating the utility of robustness coefficient, to compare the robustness of the same network under different attack types. It could be seen from Table 1 that in most cases, betweenness-centrality-based attack is a better strategy to disintegrate networks quickly and thoroughly. However, some networks are less sensitive to the type of attacks compared to others, and the robustness coefficient enables us to quantify this difference. For example, *C. elegans* neural network is only about two-thirds as robust against betweenness-centrality-based attack as compared to degree-based attack. The *E. coli* transcription network offers an even bigger contrast, where the robustness to degree-based attack is 1.64 %, yet the robustness to closeness-based attack is 11.95 %. In other networks, such as food webs, the mode of attack makes less difference. This information is important because any centrality measure is much more computationally expensive compared to node degrees, and require 'global' (network level) information whereas node degree only requires local (node level) information. In any case, there are some networks, such as Internet AS networks, where the degree centrality-based attack seems a marginally better strategy. Therefore, being able to quantify and compare the effectiveness of various strategies is vital, and the robustness coefficient enables us to do this.

We can also observe that closeness centrality often performs worse even than degree centrality. This is mostly the case in metabolic networks, transcription networks, foodwebs, and many others. However there are cases, such as in *H. sapien* cortical network, where degree centrality is by far the worst strategy (robustness against degree centrality-based attack is 83.5 % and robustness against closeness centrality-based attack is 71.01 %). It would be an interesting research question to analyse what topological characteristics make a network far less robust to any type of centrality-based attack as compared to degree-based attack. It can be postulated that this would be related to the assortativity of the networks (Newman 2002; Piraveenan et al. 2008). However, this is clearly subject to future research and verification.

In terms of destruction time per node, we may however note that the betweenness and closeness centrality measures perform typically worse than node degree. This is presumably because calculating shortest paths takes up computation time. In Table 2, we can note that in some cases, such as human cortical network and the primary school interaction networks, the betweenness-based attack takes almost double the time to disintegrate networks as compared to the degree-based attack. In other networks, the difference is not as pronounced. This observation stresses the fact that, while most networks may be least robust to betweenness-based attacks, it may not be worthwhile to utilise a betweenness-based strategy if the robustness difference is small, because it may take more computational effort. This again highlights the need to exactly quantify network robustness against different types of attacks, which we have done in this paper.

As an example, we show in Fig. 7 the attack profile of the *C. elegans* neural networks in terms of physical time. We show both the amount of time it took for each node removal (Fig. 7a), and the cumulative time (Fig. 7b), for

Table 2 Times taken to simulate disintegration of real world networks under targeted persistent (sequential) attacks

No	Network	R(D)		R(BC)		R(CC)	
		Tot.	Avg.	Tot.	Avg.	Tot.	Avg.
	Neural networks						
01	*C. elegans*	61.07	0.21	89.25	0.30	95.44	0.32
	Internet AS networks						
2	Internet AS-1998	10,526	3.27	21,627	6.72	38,688	12.03
3	Internet AS-1999	23,647	5.24	41,682	9.23	58,672	13.00
4	Internet AS-2000	27,657	4.27	47,681	7.36	64,980	10.03
	Transcription networks						
5	*C. glutamicum*	17.54	0.03	20.44	0.04	53.59	0.10
6	*E. coli*	100.54	0.09	197.77	0.17	1,088.06	0.95
	Cortical connectivity networks						
7	Human	14,766	14.86	26,224	26.38	84,329	84.84
8	Cat	3.180	0.05	4.172	0.06	4.356	0.06
9	Macaque	2.93	0.04	3.28	0.04	3.37	0.04
	Gene regulatory networks						
10	*R. norvegicus*	4,609	5.63	4,662	5.69	9,810	11.98
11	*C. elegans*	155.9	0.27	216.6	0.37	362.7	0.62
12	*A. thaliana*	23.41	0.06	26.66	0.07	91.28	0.23
	Foodwebs						
13	Bay dry	13.27	0.10	18.87	0.15	17.78	0.14
14	Bay wet	13.47	0.10	18.69	0.15	18.27	0.14
15	Chess upper	1.15	0.03	1.248	0.03	0.943	0.03
16	Chess lower	1.05	0.03	1.16	0.03	0.762	0.03
17	Chrystal C	0.66	0.04	0.73	0.04	0.75	0.04
18	Chrystal D	0.39	0.04	0.38	0.04	0.392	0.04
	Citation networks						
19	Smart grid	1,579	1.54	2,719	2.66	3,120	3.05
20	Small world	12.89	0.06	14.81	0.06	15.07	0.06
	Metabolic networks						
21	Human	264.2	0.21	297.8	0.23	3,175	2.47
22	Rhesus monkey	152.0	0.12	189.8	0.16	2,052.3	0.17
23	Chimpanzee	152.9	0.13	180.6	0.15	1,944.1	1.59
24	*Acholeplasma laidlawii*	162	0.16	103	0.10	1,026	1.02
25	*Ashbya gossypii*	195.6	0.18	318.9	0.30	1,983.1	1.86
26	*Acaryochloris marina*	4.11	0.205	4.23	0.025	3.16	0.019
	PPI						
27	*H.Pylori*	132	0.19	245	0.35	384	0.54
28	Human	223	0.15	286	0.19	2,761	2.53
29	Mouse	13.43	0.03	15.36	0.03	13.55	0.03
	US Airlines						
30	US Air lines-1997	27.6	0.083	34.9	0.105	50.2	1.51
	Collaboration networks						
31	Netscience	198.98	0.14	301.96	0.21	2,943.09	2.01
	Cell signalling networks						
32	CA1	83.9	0.15	140	0.26	179	0.33
	Miscellaneous networks						
33	Dolphins	1.81	0.029	1.94	0.031	1.43	0.029
34	Jazz Musicians	37.136	0.187	51.56	0.260	59.54	0.260

Table 2 continued

No	Network	R(D)		R(BC)		R(CC)	
		Tot.	Avg.	Tot.	Avg.	Tot.	Avg.
35	Karate	0.971	0.029	0.895	0.026	0.534	0.016
36	Pharmaceutical	290.2	0.501	394.3	0.681	637.2	1.100
37	Primary school interaction 1	127.4	0.540	224.8	0.953	235.1	0.996
38	Primary school interaction 2	115	0.483	212	0.890	212	0.891
39	Sixteen Story Hospital	1,408	1.72	2,436	2.97	3,807	4.64
40	Software	27.93	0.068	40.35	0.099	59.11	0.144
41	Vehicle	5.33	0.044	6.87	0.044	5.99	0.050

Time measured in Seconds. Both the total time, and the average time per node are shown. Three modes of node removal are used, namely (a) degree-based removal (b) betweenness-based removal and (c) closeness-based removal. The network data is taken from (Collations of connectivity data on the Macaque brain 2009; Baumbach 2007; Michigan Molecular Interaction Databse 2008; Watts and Strogatz 1998; Pajek datasets 2007; Zachary 1977; Gleiser and Danon 2003; Primary school cumualtive networks 2011; Lusseau et al. 2003; New England complex systems institute research projects. URL:http://necsi.edu/projects/braha/largescaleengineering.html

R(D) robustness under degree-based targeted attacks, *R(BC)* robustness under betweenness-based targeted attacks, *R(CC)* robustness under closeness-based targeted attacks

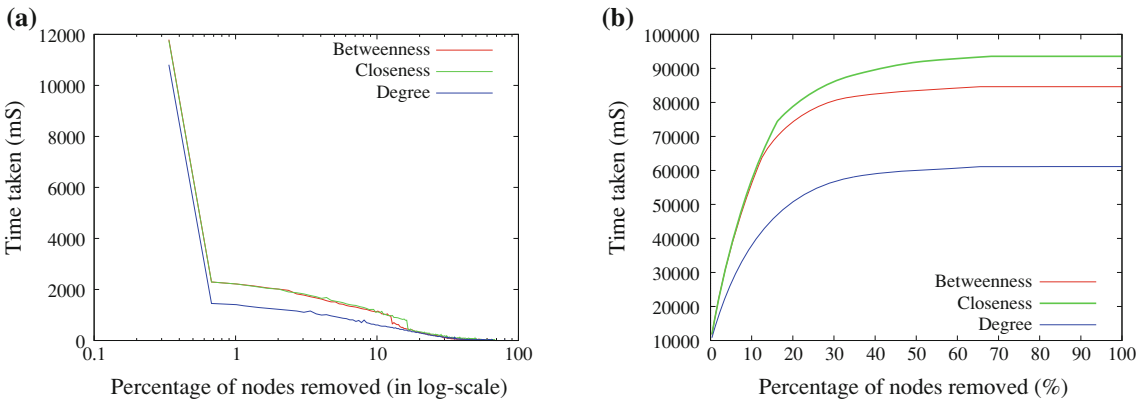

Fig. 7 The disintegration profile of the *C. elegans* neural network in terms of physical time taken in a MacBook Pro i5 machine. It could be seen that betweenness and closeness based attacks are more time consuming as compared to a degree-based attack, even though they might be more efficient in making the network disintegrate with fewer node removals. **a** Time for individual node removals. **b** Cumulative time for node removals

all three types of attacks we considered. It could be seen that the betweenness and closeness centrality plots remain 'above' the degree-based attack profile, highlighting that these attacks take more computational time in comparison.

Let us also note here that, it is tempting to use the phase transition time in an attack profile as an indicator to measure the robustness of networks under sustained attacks. The phase transition time T_{ps} could be defined, in one possible way, as the number of node removals (which could be also calculated as a percentage compared to network size) it takes before the largest component begins to reduce in size. However, such a measure does not capture what happens after the transition to the largest component. A network may have a better ability to resist total decomposition even after it begins to disintegrate, and such a network has to be classified as having better robustness. To illustrate this, we show the disintegration profiles of Small world citation network and *C. glucamitum* transcription network in Fig. 8. The nodes are sequentially removed in closeness centrality order. We can see that the largest component of the citation network immediately begins to decrease in size, and thus the phase transition time T_{ps} = 1 node removal = 0.1 % of network size. However, for the transcription network, the size of the largest component is stable for T_{ps} = 28 node removals = 5.19 % of network size, before it begins to decrease. However, Table 1 shows that the robustness of the citation network is 25.7 %, whereas the robustness of the transcription network is only 13.12 %, against closeness-centrality-based targeted attacks. Considering the plots in

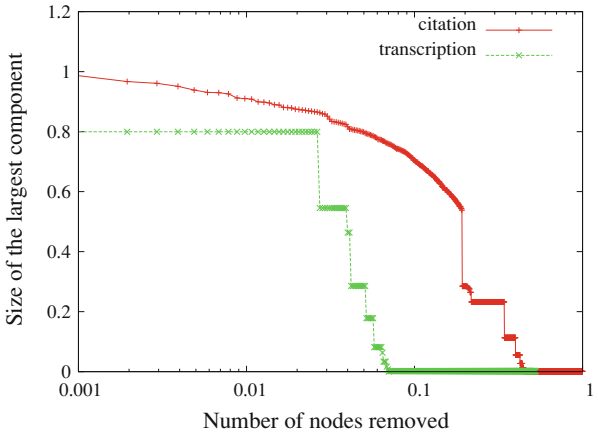

Fig. 8 The size of the largest components (as a proportion of network size) against the number of nodes removed (also as a proportion of network size) for *C. glucamitum* transcription and Small world citation networks. The nodes to be removed were chosen in the order of closeness centrality. It could be observed that while the size of the largest component begins to decrease immediately for the Small world citation network, it stays stable for a short while for the *C. glucamitum* transcription network

Fig. 8, it is clear this is the case because the citation network resists total disintegration better, even after it begins to disintegrate. This example makes clear that simple measures such as the phase transition time cannot give the full picture about a networks' disintegration profile, and the robustness coefficient proposed here is necessary.

5 Comparison with existing robustness measures

In this section, we compare our robustness coefficient measure with some of the existing measures described in Sect. 2 From the set of measures which compute the average of individual node removals, we consider the average vulnerability (Costa et al. 2007), which is derived from the *network efficiency* measure, and is convincing in its logic and easy to implement. Among the group of measures which do not compute average of node removals, we consider the average of shortest path lengths, and network diameter, as suggested by (Ng and Efstathiou 2006). We computed these metrics for all the real world networks we considered and the results are shown in Table 3. For comparison, the robustness coefficient based on betweenness-based attack is also shown.

We can immediately see from the table that there are significant differences in the ordering of networks based on each of these metrics. Let us consider the comparison between average vulnerability and robustness coefficient. It could be assumed that networks which have relatively high vulnerability would also have low robustness, and vice versa. Indeed, this is true for some networks in the table. For example, all the metabolic networks have relatively high vulnerability and relatively low robustness coefficient. Similarly, the primary school interaction networks have relatively low vulnerability and relatively high robustness coefficient. However, there are many cases where there is no such correlation. A prominent example is the case of Internet AS networks, which have very low vulnerability (they are among the least vulnerable of all the networks we studied in terms of average vulnerability, as the Table 3 shows), yet their robustness coefficient is very low also. Therefore, we might surmise that while these networks are very resilient to individual node removals, they are not at all robust to targeted sequential attacks. We might postulate that since the Internet displays the so-called rich-club phenomenon (Colizza et al. 2006; Zhou and Mondragón 2004), once the rich club has been removed by a sequential attack, the rest of the network very quickly disintegrates. The average vulnerability measure, which does not consider sequential removals, is not able to capture this. Similarly, we may see that Chesapeake upper, Chesapeake lower, Chrystal C and Chrystal D foodwebs all have relatively high average vulnerability, yet they also have high robustness coefficient. This means that even though the first node removal, especially if targeted, may severely effect the networks' integrity, they are able to better cope with consequential node removals.

Similarly, if we consider the average shortest path length, and assume that a smaller average shortest path length signifies higher robustness (the average path lengths of the networks studied vary from 1.6 to 13.72), we can see that in many cases such as the foodwebs, it is true that networks with lower average path lengths have higher robustness coefficient, and vice versa. However, there are cases such as the Internet AS Networks, transcription networks and PPI (protein–protein interaction) networks, where networks with low average path length have very low robustness coefficient; therefore the average path length fails to capture robustness against sequential targeted attacks. Furthermore, in the case of network diameter, if we consider that networks with lower diameter are more robust (the diameter values we studied range from 2 to 42), we may see that there is not always correlation between robustness measured by network diameter and our robustness coefficient. For example, *E. coli* transcription network and Rat gene regulatory network have very similar network diameters (9 and 10), but significantly different robustness coefficients (1.46 and 22.52 %). Similarly, *A. laidlawii* metabolic network and house mouse protein–protein interaction network have similar robustness coefficients (1.29 and 1.22 %), but significantly different network diameters (42 and 9). Therefore there is no correlation between the two metrics.

Rather than belabouring this point further, we leave it to the interested reader to rank the networks according to each

Table 3 Comparing robustness coefficient with other robustness measures. The average vulnerability (Costa et al. 2007) is given in units of 10^{-4}

No	Network	Avg. vulnerability	Avg. short. path	Diameter	R(BC)
	Neural networks				
1	*C. elegans*	9.94	1.99	2	41.79
	Internet AS networks				
2	Internet AS-1998	3.76	3.76	9	1.8
3	Internet AS-1999	2.43	3.74	10	1.91
4	Internet AS-2000	1.39	3.70	9	2.26
	Transcription networks				
5	*C. glutamicum*	39.14	4.69	10	1.35
6	*E. coli*	11.18	3.60	9	1.46
	Cortical connectivity networks				
7	Human	2.43	3.12	6	51.38
8	Cat	43.45	1.7	3	69.21
9	Macaque	40.96	2.24	5	52.65
	Gene regulatory networks				
10	*R. norvegicus*	5.85	3.78	10	22.52
11	*C. elegans*	15.49	4.23	12	14.25
12	*A. thaliana*	29.92	4.12	15	5.91
	Foodwebs				
13	Bay dry	8.60	1.77	3	71.73
14	Bay wet	9.03	1.78	3	72.07
15	Chess upper	53.38	1.70	3	55.08
16	Chess lower	56.63	1.75	3	45.58
17	Chrystal C	99.31	1.60	3	60.42
18	Chrystal D	97.98	1.72	3	44.10
	Citation networks				
19	Smart grid	6.45	2.98	6	24.80
20	Small world	22.44	2.37	4	14.09
	Metabolic networks				
21	Human	30.24	10.51	32	1.94
22	Rhesus monkey	27.41	10.17	25	1.56
23	Chimpanzee	30.95	9.49	29	1.57
24	*Acholeplasma laidlawii*	47.59	13.72	42	1.29
25	*Ashbya gossypii*	47.59	8.77	24	3.67
26	*Acaryochloris marina*	80.85	3.50	9	3.98
	PPI				
27	*H.Pylori*	15.13	4.14	9	13.59
28	Human	19.99	6.50	20	1.31
29	Mouse	37.41	3.36	9	1.22
	US Airlines				
30	US Air lines-1997	27.21	2.74	6	15.02
	Collaboration networks				
31	Netscience	19.72	5.82	17	1.79
	Cell signalling networks				
32	CA 1	15.85	4.21	10	16.66
	Miscellaneous networks				
33	Dolphins	80.85	3.36	8	32.57
34	Jazz Musicians	20.40	2.23	6	50.08
35	Karate	202.5	2.41	5	24.05

Table 3 continued

No	Network	Avg. vulnerability	Avg. short. path	Diameter	R(BC)
36	Pharmaceutical	6.51	2.63	5	18.86
37	Primary school interaction 1	5.20	1.86	3	82.68
38	Primary school interaction 2	10.96	1.94	3	78.33
39	Sixteen story hospital	4.85	3.12	9	23.27
40	Software	26.09	3.70	9	14.63
41	Vehicle	28.12	2.88	6	40.68

The network data is taken from (Collations of connectivity data on the Macaque brain 2009; Baumbach 2007; Michigan Molecular Interaction Databse 2008; Watts and Strogatz 1998; Pajek datasets 2007; Zachary 1977; Gleiser and Danon 2003; Primary school cumulative networks 2011; Lusseau et al. 2003; New England complex systems institute research projects. URL:http://necsi.edu/projects/braha/largescaleengineering.html

R(BC) robustness coefficient calculated in terms of betweenness

metric of robustness in Table 3, and compare the rankings. We have done this and seen that there are significant differences on the ranking based on each of the other metrics and our robustness coefficient, highlighting that our robustness coefficient particularly captures a network's ability to maintain topological integrity against*sequential* attacks, which the other metrics do not.

6 Conclusions

In this paper, we introduced a new measure for quantifying robustness of networks under sustained targeted attacks. The robustness coefficient measure has the advantage of providing information about the entire decomposition profile of the network, at the same time being a single measure. While we took a generic approach, we explained how the measure is relevant to social network analysis.

Using this measure and synthesized networks, we confirmed the result that scale-free networks are more resilient to random attacks and more vulnerable to targeted attacks compared to random topologies. We then analysed the robustness of a number of real world networks under sustained targeted attacks, comparing various attack strategies. We specifically considered attack strategies based on (1) node degree (2) node betweenness (3) node closeness. We highlighted that most networks are least robust against betweenness-centrality-based attack, and most robust against closeness-centrality-based attack, as shown in Table 2. However, there were networks for which a degree-based attack was the most effective. Furthermore, we pointed out that, since betweenness analysis is computationally expensive, it may not make sense to adapt a betweenness-based attack strategy if the robustness of the network is only marginally lower against betweenness-based attacks. We illustrated this point by computing and comparing the physical time taken for simulating network disintegration. This analysis demonstrated the utility of our robustness measure as a tool in designing and comparing attack strategies.

Sustained attacks based on other node properties could be conceived. Topologically, there exist a host of other centrality measures, such as Eigenvector centrality and information centrality (Bonacich 2001; Noh and Rieger 2004). Attacks also could be designed based on other topological properties such as local assortativity (Piraveenan et al. 2010). Non-topological attributes of nodes, such as the age or importance of people in social networks, or the amount of data stored in computers in computer networks etc., also could be used to rank nodes to be chosen for attacks. It is clear that the robustness coefficient we have introduced could be effectively used to compare all these attack strategies.

We need to be mindful here that network vulnerability cannot be considered totally in isolation from the nature of the network. The structural elements explain robustness, other things being equal. Furthermore, the overall procedure of generating the largest component size against number of nodes removed plot and calculating the robustness coefficient from it has high time complexity, and research is ongoing in search of algorithms to improve this. That said, we believe that this measure will be extensively utilised in analysing structural robustness of networks by the scientific community.

References

Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74:47–97

Albert R, Jeong H, Barabási AL (2000) Error and attack tolerance of complex networks. Nature 406:378–382

Alon U (2007) Introduction to systems biology: design principles of biological circuits. Chapman and Hall, London

Baumbach J (2007) Coryneregnet 4.0-a reference database for corynebacterial gene regulatory networks. BMC Bioinf 8

Bonacich P (2001) Eigenvector-like measures of centrality for asymmetric relations. Soc Netw 23(3):191–201
Cazabet R, Takeda H, Hamasaki M, Amblard F (2012) Using dynamic community detection to identify trends in user-generated content. Soc Netw Anal Min 2(4):361–371
Colizza V, Flammini A, Serrano MA, Vespignani A (2006) Detecting rich-club ordering in complex networks. Nat Phys 2:110–115
Collations of connectivity data on the Macaque brain (2009) URL: http://www.cocomac.org/
Costa LDF, Rodrigues FA, Travieso G, Villas Boas PR (2007) Characterization of complex networks: a survey of measurements. Adv Phys 56(1):167–242
Crucittia P, Latora V, Marchiori M, Rapisarda A (2004) Error and attack tolerance of complex networks. Phys A 340:388–394
Dekker AH, Colbert BD (2004) Network robustness and graph topology. In: Proceedings of the 27th Australasian conference on computer science. ACSC '04, vol 26, Australian Computer Society Inc., Darlinghurst, pp 359–368
Dorogovtsev SN, Mendes JFF (2003) Evolution of networks: from biological nets to the internet and WWW. Oxford University Press, Oxford
Gilbert F, Simonetto P, Zaidi F, Jourdan F, Bourqui R (2011) Communities and hierarchical structures in dynamic social networks: analysis and visualization. Soc Netw Anal Min 1(2):83–95
Gleiser P, Danon L (2003) Adv Complex Syst 6:565
Hanley JA, Mcneil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143(1):29–36
Junker BH, Schreiber F (2008) Analysis of biological networks (Wiley Series in Bioinformatics). Wiley, New York
Kepes F (2007) (ed) Biological networks. World Scientific, Singapore
Kreyszig E (2005) Advanced engineering mathematics, 9th edn. Wiley, New York
Lusseau D, Schneider K, Boisseau OJ, Haase P, Slooten E, Dawson SM (2003) Dolphin social network. Behav Ecol Sociobiol 54
Michigan Molecular Interaction Database (2008) University of Michigan URL:http://mimi.ncibi.org/MimiWeb/main-page.jsp
Milgram S (1967) The small world problem. Psychol Today 1:61
Newman MEJ (2002) Assortative mixing in networks. Phys Rev Lett 89(20):208–701
Newman MEJ (2005) A measure of betweenness centrality based on random walks. Soc Netw 27(1):39–54
Ng AKS, Efstathiou J (2006) Structural robustness of complex networks. Phys Rev 3:175–188
Noh JD Rieger H (2004) Random walks on complex networks. Phys Rev Lett 92:118–701
Pajek datasets (2007). URL: http://vlado.fmf.uni-lj.si/pub/networks/data
Piraveenan M, Prokopenko M, Zomaya AY (2008) Local assortativeness in scale-free networks. Europhys Lett 84(2):28–002
Piraveenan M, Prokopenko M, Zomaya AY (2009) Assortativity and growth of Internet. Eur Phys J B 70:275–285
Piraveenan M, Prokopenko M, Zomaya AY (2010) Local assortativeness in scale-free networks—addendum. Europhys Lett 89(4):49–901
Primary school cumulative networks (2011) URL:http://www.sociopatterns.org/datasets/primary-school-cumulative-networks/
Rees BS, Gallagher KB (2012) Overlapping community detection using a community optimized graph swarm. Soc Netw Anal Min 2(4):405–417
Solé RV, Valverde S (2004) Information theory of complex networks: on evolution and architectural constraints. Complex networks In: Ben-Naim E, Frauenfelder H, Toroczkai Z (eds) Lecture notes in physics, vol 650. Springer, Berlin
Tang A, Honey C, Hobbs J, Sher A, Litke A, Sporns O, Beggs J (2008) Information flow in local cortical networks is not democratic. BMC Neurosc 9(Suppl 1):O3
Venkatasubramanian V, Katare S, Patkar PR, Mu F (2004) Spontaneous emergence of complex optimal networks through evolutionary adaptation. CoRR nlin.AO/0402046
Watts DJ, Strogatz SH (1998) Collective dynamics of small-world networks. Nature 393(6684):440–442
Zachary W (1977) An information flow model for conflict and fission in small groups. J Anthropol Res 33:452–473
Zaidi F (2013) Small world networks and clustered small world networks with random connectivity. Soc Netw Anal Min 3(1):51–63
Zhou S, Mondragón RJ (2004) The rich-club phenomenon in the internet topology. IEEE Commun Lett 8:180–182