Quantifying the Robustness of Network Controllability

Sun, Peng; Van Mieghem, Piet; Kooij, R.E.; He, Zhidong; Van Mieghem, Piet

DOI
10.1109/ICSRS48664.2019.8987628

Publication date
2019

Document Version
Accepted author manuscript

Published in
2019 4th International Conference on System Reliability and Safety, ICSRS 2019

Quantifying the Robustness of Network Controllability

Peng Sun*, Robert E. Kooij*†, Zhidong He*, Piet Van Mieghem*
*Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands
†iTrust Centre for Research in Cyber Security, Singapore University of Technology and Design, Singapore

***Abstract*—In this paper, we propose closed-form analytic approximations for the minimum number of driver nodes needed to fully control networks, where links are removed according to both random and targeted attacks. Our approximations rely on the concept of critical links. A link is called critical if its removal increases the required number of driver nodes. We validate our approximation on both real-world and synthetic networks. For random attacks, the approximation is always very good, as long as the fraction of removed links is smaller than the fraction of critical links. For some cases, the approximation is still accurate for larger fractions of removed links. The approximation for an attack, where first the critical links are removed, is also accurate, as long as the fraction of removed links is sufficiently small. Finally, we show that the critical link attack is the most effective among 4 considered attacks, as long as the fraction of removed links is smaller than the fraction of critical links.**

I. Introduction

Our society nowadays depends critically on the proper functioning of a variety of infrastructures, such as the Internet, the power grid, water management networks and mobile communication networks. It is common practice to model such infrastructures as complex networks. Research over the last decades has led to a deep understanding of structural and robustness properties of complex networks [1], [2]. In recent years, the emphasis has shifted to understanding the controllability of such networks [3] [4] [5] [6]. Controllability is an essential property for the safe and reliable operation of real-life infrastructures. A system is said to be controllable if it can be driven from any initial state to any desired final state by external inputs in finite time [6]. Merging classical control theory with network science [7] introduced the notion of structural controllability. Let the $N \times N$ matrix A represent the network's wiring diagram, while the connection of M input signals to the network is described by the $N \times M$ input matrix B, where $M \leq N$. Then, the system characterized by (A, B) is said to be structurally controllable, if it is possible to fix the non-zero parameters in A and B in such a way that the obtained system (A, B) is controllable in the classical sense of satisfying Kalman's rank condition. Liu *et al.* [3] found a method that gives the minimum number of driver nodes, which are driven by external inputs, that are needed to achieve structural controllability of a directed network. As was pointed out by Cowan *et al.* [8], the results reported in Liu *et al.* [3] critically depend on the assumption that the network has no self-links, i.e. a node's internal state can only be changed upon interaction with a neighbor. In this paper, we will also assume this condition. Ruths *et al.* [9] developed a theoretical framework for characterizing control profiles of networks. Yuan *et al.* [4] further proposed the concept of exact controllability based on the maximum multiplicity of all eigenvalues of the adjacency matrix A to find the driver nodes in networks. Jia *et al.* [5] classified each node into one of three categories, based on its likelihood of being included in a minimum set of driver nodes and discovered bimodal behaviour for the fraction of redundant nodes, when the average degree of the networks is high. Nepusz *et al.* [6] indicated that most real-world networks are more controllable than their randomized counterparts. Yan *et al.* [10] investigated the relation between the maximum energy needed for controllability and the number of driver nodes.

Real-world networks are often confronted with topological perturbations such as link-based random failures or targeted malicious attacks. For instance, in power grids, the breakdown of connections between different substations in some cases can be interpreted as random failures due to circuit aging or natural disasters. Malicious, and targeted attacks can seriously degrade the network performance [11]. In transportation networks, betweenness centrality-based targeted attacks can have a significant impact on normal operation [12].

Network robustness under topological perturbations has been widely investigated. The effective graph resistance [13], the viral conductance [14], the size of giant component [15], betweenness and eigenvector centrality are computed to measure the robustness of networks under topological perturbations. Wang *et al.* [16] investigated two interconnection topologies for interdependent networks and proposed the derivative of the largest mutually connected component as a new robust metric, which addresses the impact of a small fraction of failed nodes. Trajanovski *et al.* [17] studied the robustness envelope and concluded that centrality-based targeted attacks are sufficient for studying the worst-case behavior of real-world networks. Koc *et al.* [18] found that increasing the effective graph resistance of synthetic power systems results in decreased grid robustness against cascading failures by targeted attacks.

The robustness of the network controllability can be as-

sessed by quantifying the increase in the minimum number of driver nodes N_D, under perturbation of the network topology. The impact of topological perturbations on the controllability of networks has been investigated extensively in recent years. Pu *et al.* [19] found that the degree-based node attack is more efficient than a random attack for degrading the controllability in directed random and scale-free networks. Nie *et al.* [20] found that the controllability of Erdős-Rényi random graphs with a moderate average degree is less robust, whereas a scale-free network with moderate power-law exponent shows a stronger ability to maintain its controllability, when these networks are under intentional link attack. Thomas *et al.* [21] identified that the potency a degree-based attack is directly related (on average) to the betweenness centrality of the edges being removed. Lu *et al.* [22] discovered that a betweenness-based strategy is quite efficient to harm the controllability of real-world networks. Mengiste *et al.* [23] introduced a new graph descriptor, 'the cardinality curve', to quantify the robustness of the control structure of a network to progressive link pruning.

The previous works on the robustness of network controllability listed above, have been mainly based upon simulations. In this paper we quantify the robustness of network controllability by deriving analytical expressions, approximating the increase of the number of driver nodes, upon random and targeted link removals. Based upon methods from statistical physics, Liu *et al.* [3] already found analytical approximations for the number of driver nodes N_D, as a function of the nodes in- and out-degree distributions. However, the obtained expressions are an implicit set of equations, which are derived under the assumptions of $N \to \infty$ (thermodynamic limit) and sufficiently large average node degree.

We propose an analytical approximation to quantify the robustness of network controllability, based upon the concept of critical links, introduced in [3]. A link is said to be critical if its removal increases the minimum number of driver nodes N_D. We derive analytical approximations for the minimum number of driver nodes N_D as a function of the fraction of removed links, both for random and targeted attacks. We show the performance of our approximations in both real-world and synthetic networks. Finally, we compare an attack based upon crictial links, to attacks based upon topological properties, such as the out-in degree-based attack.

This paper is organized as follows. In Section II, we introduce some basic concepts and definitions in network controllability proposed in [3]. In Section III and IV, we propose analytic approximations for the minimum number of driver nodes N_D when the network is under random attacks and targeted attacks, respectively. In Section V, we compare the robustness of controllability under four different attack methods. Section VI concludes the paper.

II. Controllability of networks and driver nodes

A. Controllability of networks

A system is controllable if it can be driven from any initial state to any desired final state by proper variable inputs in finite time [24]. Most real systems are driven by nonlinear processes, but the controllability of nonlinear systems is in many aspects structurally similar to that of linear systems [3]. We consider a linear, time-invariant dynamics on a directed network, which is described by:

$$\frac{dx(t)}{dt} = Ax(t) + Bu(t) \tag{1}$$

where the $N \times 1$ vector $x(t) = (x_1(t), x_2(t), ..., x_N(t))^T$ denotes the state of the system with N nodes at time *t*. The weighted matrix A is an $N \times N$ matrix which describes the network topology and the interaction strength between the components. The $N \times M$ input matrix B identifies the $M \leq N$ nodes controlled by input signals. The $M \times 1$ vector $u(t) = (u_1(t), u_2(t), ..., u_M(t))^T$ is the input signal vector.

The linear system defined by equation (1) is controllable, if and only if the $N \times NM$ controllability matrix:

$$C = (B, AB, A^2B, ..., A^{N-1}B) \tag{2}$$

has full rank, i.e., $rank(C) = N$. This criterion is called Kalman's controllability rank condition [25]. The rank of matrix C provides the dimension of the controllable subspace of the system. We need to choose the right input matrix B consisting of a minimum number of driver nodes to assure that the controllability matrix C has full rank. System (1) is said to be structurally controllable if it is possible to fix the non-zero parameters in A and B in such a way that the obtained system (A, B) satisfies Kalman's rank condition. We assume that A has no self-loops, i.e. all entries in the diagonal of A are zero.

B. Driver nodes and critical links

Liu *et al.* [3] proved that the minimum number of driver nodes needed for structural controllability, where the input signals are injected to control the directed network, can be obtained through the "maximum matching" of the network. Define the source node of a directed link as the node from which the link originates and the target node as the node where the link terminates. A maximum matching of a directed network is a maximum set of links that do not share source or target nodes [26], which is illustrated in Figure 1(a). Such links are coined "matching links". Target nodes of matching links are matched nodes and the other nodes are unmatched nodes. For a given maximum matching, connecting driver nodes with unmatched nodes gives a minimum number of driver nodes N_D needed for controlling the network.

In order to find the maximum number of matching links, so as to determine the minimum number of driver nodes N_D, a directed network G with N nodes and L links can be converted into a bipartite graph $B_{N,N}$ with $2N$ nodes and L links, as shown in Figure 1(b). A maximum matching in a bipartite graph can be obtained efficiently by the Hopcroft-Karp algorithm [27]. The unmatched nodes in a maximum matching constitute a minimum set of driver nodes. It is worth noting that a minimum set of driver nodes is not necessarily unique. The Hopcroft-Karp algorithm guarantees to return the minimum number of driver nodes to completely control the

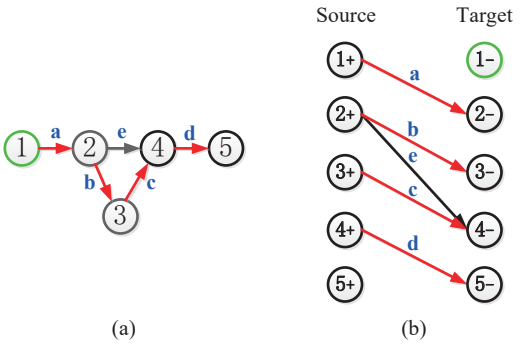


Fig. 1. Driver nodes and critical links in a directed network G. (a) An example network G with $N = 5$ nodes and $L = 5$ directed links. Since link a, b, c and d are the maximum set of links that do not share source and target nodes, these links are matching link. Target nodes of these matching links, i.e., node 2, 3, 4 and 5, are matched nodes. Node 1 is an unmatched node. (b) The bipartite graph with $2N$ nodes and L links. Matching links are highlighted in red in the bipartite graph. Driver nodes are highlighted in green. To create the bipartite graph, each node V in the original network G will be translated into source node V_+ and target node V_- in the bipartite graph. The first column of the bipartite graph are all possible source nodes, whereas the nodes in the second column are all possible target nodes. Links in the bipartite graph are determined by the directed links in the original network G. By using the Hopcroft-Karp algorithm, a maximum set of matching links can be found in the bipartite graph. None of the matching links share a common source or target node. Then, the target nodes of matching links are matched nodes. Other target nodes are unmatched nodes, which are also driver nodes.

network. The computational complexity of the Hopcroft-Karp algorithm to find all driver nodes is $O(\sqrt{N}L)$.

Links can be classified into three categories: critical, redundant, and ordinary [3]. A link is critical if its removal increases the number of driver nodes to remain in full control of the system. A link is redundant if it never belongs to a maximum matching. A link is ordinary if it is neither critical nor redundant. In this paper, we want to derive analytical expressions for the increase in the minimum number of driver nodes, upon link removal. We will use the concept of critical links to construct such approximations, both for random link removals and targeted attacks.

III. Number of Driver Nodes under Random Attacks

In this section, we assume that links are removed from the network uniformly at random. We derive an analytical approximation for the minimum number of driver nodes N_D for random attacks and show the performance of the approximation for real-world and synthetic networks.

A. The fraction l of removed links is less than the fraction of critical links l_c

For a network with N nodes and L links, denote the minimum number of driver nodes by N_{D0}. The number of critical links L_C can be determined by applying the Hopcraft-Karp algorithm L times, by considering all L networks that are obtained by removing exactly one link from the original network. If we denote the number of removed links by m, then the fraction of removed links $l = \frac{m}{L}$, while the fraction of critical links l_c satisfies $l_c = \frac{L_c}{L}$. We consider the case $l \leq l_c$, i.e. m links are removed uniformly at random, under the condition that the number of removed links $m \leq L_c$. Now assume that of these m links i links are critical ($i \leq m$) and, hence, $m - i$ links are non-critical. We assume that the set of critical links is nearly unchanged when the fraction of removed links is small. Invoking the fact that after removing a critical link, the minimum number of driver nodes N_D increases by one [3], thus, when i critical links are iteratively removed one by one, the minimum number of driver nodes N_D increases by one in each iteration. For the $m - i$ removed non-critical links, the minimum number of driver nodes N_D remains the same. Since there are $\binom{L_c}{i}$ possible ways to choose i critical links from L_c critical links and there are $\binom{L-L_c}{m-i}$ possible ways to choose $m - i$ non-critical links from $L - L_c$ non-critical links, the contribution to the increase in N_D for each i is $i\binom{L_c}{i}\binom{L-L_c}{m-i}$. The expectation of the increase N_D^* of the minimum number of driver node N_D after randomly removing m links, is the sum of this expression for all $i = 1, 2, \ldots, m$ and divide it by $\binom{L}{M}$.

$$N_D^* = \frac{\sum_{i=1}^{m} i\binom{L_c}{i}\binom{L-L_c}{m-i}}{\binom{L}{m}} \tag{3}$$

We first rewrite the numerator of the right hand site of Eq. (3):

$$\begin{aligned}\sum_{i=1}^{m} i\binom{L_c}{i}\binom{L-L_c}{m-i} &= \sum_{i=1}^{m} \frac{L_c!}{(i-1)!(L_c-i)!}\binom{L-L_c}{m-i} \\ &= L_c \sum_{i=1}^{m} \binom{L_c-1}{i-1}\binom{L-L_c}{m-i} \\ &= L_c \sum_{i=0}^{m-1} \binom{L_c-1}{i}\binom{L-L_c}{m-i-1}\end{aligned}$$

By using Vandermonde's formula: $\sum_{j=0}^{k}\binom{a}{j}\binom{b}{k-j} = \binom{a+b}{k}$, we obtain $L_c\sum_{i=0}^{m-1}\binom{L_c-1}{i}\binom{L-L_c}{m-1-i} = L_c\binom{L-1}{m-1}$. Finally, dividing this expression by $\binom{L}{m}$, we obtain

$$N_D^* = lL_c \tag{4}$$

When the fraction of removed links is less than, or equal to l_c, we obtain

$$N_D = N_{D0} + lL_c \tag{5}$$

For convenience, we normalize the minimum number of driver nodes N_D to the fraction of the minimum number of driver nodes, i.e., $\frac{N_D}{N}$ and denote the obtained approximation as $n_{D,rand}$.

$$n_{D,rand} = \frac{N_{D0} + lL_c}{N} \tag{6}$$

1) Validation for real-world networks: We evaluate the performance of the approximation $n_{D,rand}$ in (6) for 8 real-world networks. Table I presents the properties of the 8 real-world networks: the number of nodes (N), the number of links (L), the initial minimum number of driver nodes (N_{D0}) and the number of critical links (L_c).

TABLE I
PROPERTIES OF THE 8 REAL-WORLD NETWORKS

Networks	N	L	N_{D0}	L_c
Amazon network [28]	105	441	25	29
Berlin traffic network [29]	224	523	14	123
IEEE118 power grid [30]	118	179	38	36
Illinois students network [30]	70	366	3	8
Hagy Chesapeake Bay ecosystem [31]	37	215	9	4
INSNA social network [32]	60	94	35	5
s838 [3]	512	819	119	179
TRN-Yeast-2 [3]	688	1079	565	23

Figure 2 shows the comparison between our approximation Eq. (6) and simulation results in the considered real-world networks. For each figure, the right-most point at the horizontal axis denotes the fraction of critical links l_c. We use 10000 realizations and obtain mean values for the fraction of minimum number of driver nodes n_D, together with the $95\%-$ confidence interval, for each fraction l. Visual inspection of Figure 2 confirms that our approximation (6) is close to the simulation results for the 8 real-world networks, when the fraction of removed links l satisfies $l \leq l_c$.

To further quantify the accuracy of the approximation $n_{D,rand}$, Table II gives two performance indicators. K different values of the fraction of removed links, i.e., $c_1, c_2, ..., c_K$, are evenly determined in the interval $[0, l_c]$. Let $n_D^*(c_i)$ and $n_D(c_i)$ denote the mean simulated n_d and the approximation (6) at the fraction of removed links $l = c_i$, respectively. The performance indicator γ denotes the fraction of the interval $[0, l_c]$ for which the absolute value of the relative error between the approximation and the mean simulated value, does not exceed 5%.

$$\gamma = \frac{\sum_{i=0}^{K} \mathbf{1}_{\left| \frac{n_D^*(c_i) - n_D(c_i)}{n_D^*(c_i)} \right| \leq 5\%}}{K}$$

Finally, r denotes the absolute value of the relative error between the approximation and the mean value obtained through simulation, at $l = l_c$. Table II shows for all real-world networks that the approximation (6) for $n_{D,rand}$ performs very well for $l \leq l_c$. For 5 out of the 8 considered networks, the absolute value of the relative error at $l = l_c$ is less than 0.5%.

TABLE II
PERFORMANCE INDICATORS FOR THE APPROXIMATION $N_{D,rand}$ FOR THE 8 REAL-WORLD NETWORKS; $l \leq l_c$

Networks	γ	r
Amazon	100%	0.11%
Berlin traffic	100%	4.82%
IEEE118 power grid	100%	2.31%
Illinois students	100%	0.35%
Hagy Chesapeake Bay	100%	0.07%
INSNA	100%	0.20%
s838	100%	4.80%
TRN-Yeast-2	100%	0.01%

2) Synthetic networks : Next we test our approximation Eq. (6) on two types of synthetic networks. When generating the directed Erdős-Rényi random network $G_p(N)$ with N nodes, the probability that every node has an outbound link to the other nodes is p. We generate the scale-free network $BA(N, M_0, M)$ by using the Barabási-Albert (BA) model, where N is the number of nodes, M is the number of out-going links for each new node added to the current network. We assume that initially the network consists of a complete digraph on M_0 nodes, where M_0 equals M. In the initial complete digraph, every pair of distinct nodes is connected by a pair of unique links (one in each direction). New nodes are added to the network one at a time. Each new node is connected to M existing nodes with a probability that is proportional to the number of links that the existing nodes already have. Figure 3 shows that both for Erdős-Rényi and Barabási-Albert (BA) networks, our analytic approximation (6) for $n_{D,rand}$ fits well with simulation results, when the fraction of removed links l is less than the fraction of critical links l_c. For the results depicted in Figure 3, Table III reports the performance indicators γ and r introduced in the previous subsection. Table III shows that also for the considered synthetic networks, the approximation $n_{D,rand}$ performs very well for $l \leq l_c$.

The overall conclusion of this subsection is that our approximation $n_{D,rand}$ in Eq.(6) gives a very good estimation for the minimum number of driver nodes, if the fraction of randomly removed links l is smaller than, or equal to, the fraction of critical links l_c.

TABLE III
PERFORMANCE INDICATORS FOR THE APPROXIMATION $n_{D,rand}$ FOR THE 4 SYNTHETIC NETWORKS; $l \leq l_c$

Networks	γ	r
ER: $G_{0.07}(50)$	100%	2.08%
ER: $G_{0.04}(100)$	100%	1.80%
BA: N=200, $E[D]$=4	100%	0.29%
BA: N=500, $E[D]$=8	100%	0.09%

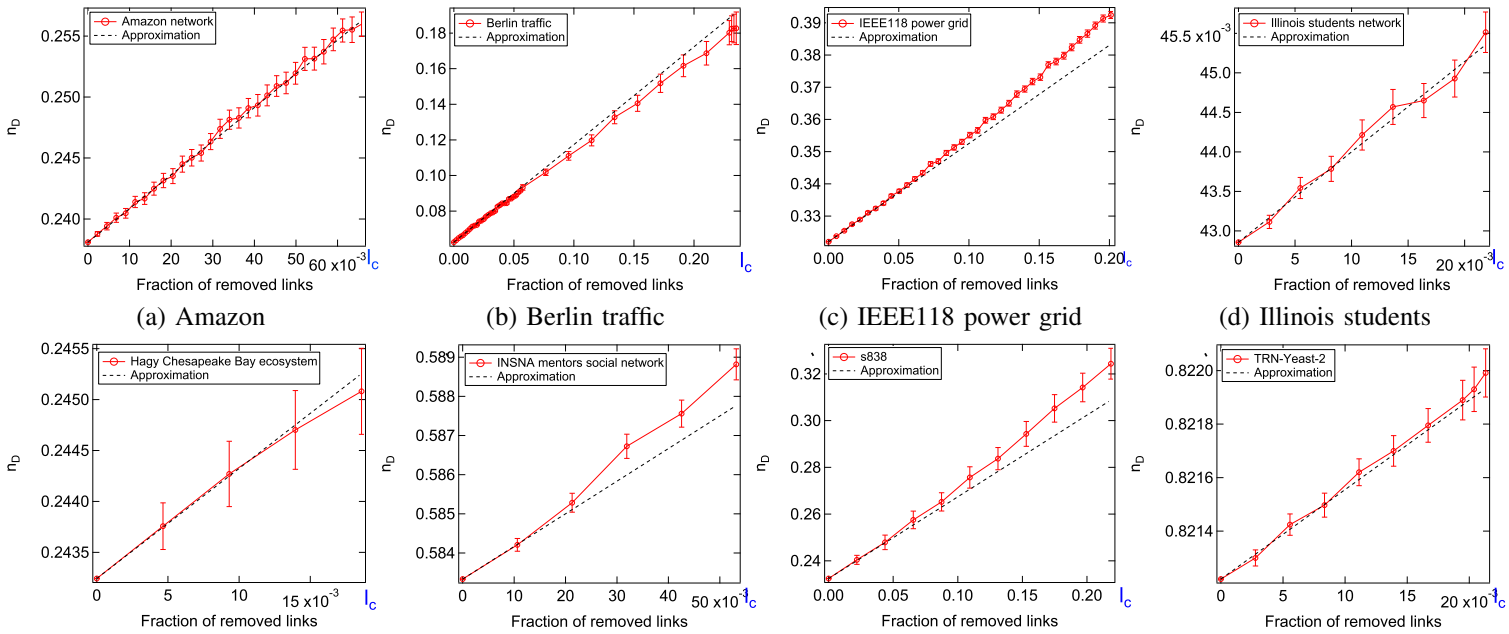
B. The fraction l of removed links is larger than the fraction of critical links l_c

Because in most cases l_c is quite small, we also estimate the normalized minimum number of driver nodes n_D when the fraction l of removed links is larger than the fraction l_c of critical links. Therefore, for $l \geq l_c$, we propose a simple closed-form approximation for n_D:

$$n_D = al^2 + bl + c \tag{7}$$

where the parameters a, b and c will be determined by some boundary conditions. For the first two boundary conditions we assume that, for $l = l_c$, Eq.(7) has the same value and the same derivative as Eq. (6). This leads to the equations $N_{D0} + l_c L_c = N(al_c^2 + bl_c + c)$ and $L_c = N(2al_c + b)$, respectively. Finally, if we remove all links, i.e. $l = 1$, all nodes need to be controlled. This gives the boundary condition $1 = a + b + c$. Solving for a, b and c and combining with the approximation Eq.(6), we obtain the following approximation for n_D for all values of l:

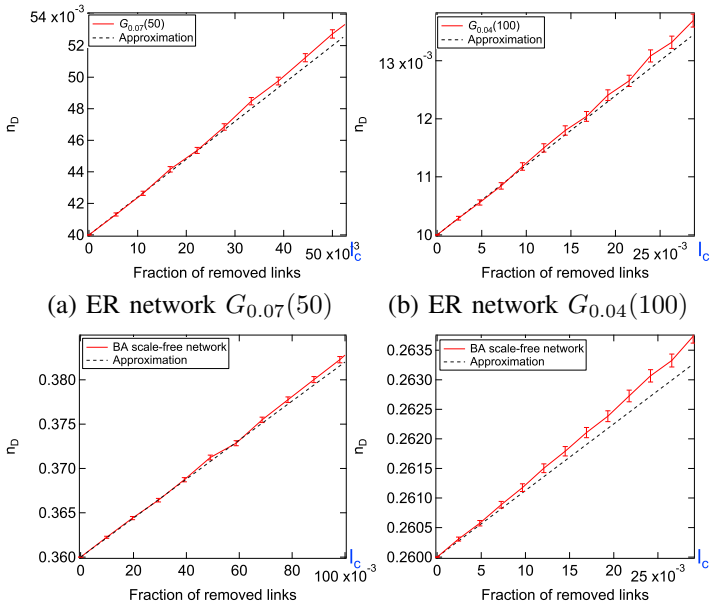
$$n_{D,rand} = \begin{cases} \frac{N_{D0} + lL_c}{N} & l \leq l_c \\ \frac{al^2 + bl + c}{N} & l \geq l_c \end{cases} \tag{8}$$



(a) Amazon (b) Berlin traffic (c) IEEE118 power grid (d) Illinois students

(e) Hagy Chesapeake Bay (f) INSNA (g) s838 (h) TRN-Yeast-2

Fig. 2. Performance of the approximation (6) for the normalized minimum number of driver nodes n_D as a function of the fraction of removed links l in real-world networks under random attacks. The results for each fraction l is based on 10000 simulations.



(a) ER network $G_{0.07}(50)$ (b) ER network $G_{0.04}(100)$

(c) N=200, $E[D]$=4 (d) N=500, $E[D]$=8

Fig. 3. The normalized minimum number of driver nodes n_D as a function of the fraction of removed links l in synthetic networks under random attacks. In each sub-figure, we generate 100 corresponding synthetic networks and calculate the average fraction of critical links l_c and the average value of n_D for each fraction of removed links. For each network, the value of n_D for each fraction l is based on 10000 simulations.

with, $a = \frac{N-N_{D0}-L_c}{N(l_c-1)^2}$, $b = L_cN - 2al_c$, and $c = 1 - L_cN + a(2l_c - 1)$. Eq.(8) respresents a closed-form approximation for n_D, which only depends on N, L, N_{D0} and L_C. The computational complexity of the approximation is $O(\sqrt{N}L^2)$, which is needed for the computation of L_C.

We compare the approximation (8) with simulation results for the 8 real-world networks and two types of synthetic networks. Figure 4 shows that for moderate values of the fraction of removed links, the approximation exhibits a very good fit for the real-world networks. This is quantified in Table IV where we show two performance indicators: r which denotes the relative error at $l = 0.2$ and l^*, which represents the smallest value of l, where the relative error between the approximation and the simulated mean exceeds 5%.

TABLE IV
PERFORMANCE INDICATORS FOR THE APPROXIMATION $n_{D,rand}$ FOR THE 8 REAL-WORLD NETWORKS

Networks	r	l^*
Amazon	3.12%	0.32
Berlin traffic	3.15%	0.24
IEEE118 power grid	2.31%	0.29
Illinois students	30.20%	0.12
Hagy Chesapeake Bay	5.22%	0.19
INSNA	1.50%	0.68
s838	4.15%	0.23
TRN-Yeast-2	0.39%	0.72

Figure 4 illustrates that the approximation both under- and overestimates the value of n_D. Table IV shows that the approximation is the most accurate for the INSNA social network and the TRN-Yeast-2 network, while the least accurate for Illinois students network and Hagy Chesapeake Bay ecosystem. According to Table IV, for 6 out of the 8 real-world networks, for random link removals up to 20%, the absolute value of the relative error of the approximation (8)

does not exceed 5%. For the worst performing network, Hagy Chesapeake Bay, 12% of the links can be removed before the absolute relative error exceeds 5%.

Finally, Figure 5 shows that the comparison for Erdős-Rényi and Barabási-Albert networks, leads to the same conclusions as above. The performance indicators r and l^* for the 4 synthetic networks are given in Table V.

The overall conclusion of this subsection is that our approximation $n_{D,rand}$ in (8), in most cases, also gives a good estimation for the minimum number of driver nodes, if the fraction of randomly removed links l is larger than the fraction of critical links l_c, but still sufficiently small.

TABLE V
PERFORMANCE INDICATORS FOR THE APPROXIMATION (6) FOR THE 4 SYNTHETIC NETWORKS

Networks	r	l^*
ER: $G_{0.07}(50)$	2.32%	0.47
ER: $G_{0.04}(100)$	23.56%	0.08
BA: N=200, $E[D]$=4	1.47%	0.57
BA: N=500, $E[D]$=8	3.25%	0.28

IV. DRIVER NODES UNDER TARGETED ATTACKS

In this section, we quantify the impact of targeted link attacks on the minimum number of driver nodes. We assume that the attacker knows the critical links, which will be attacked first. We consider two scenarios. In the first scenario, the attacker removes critical links uniformly at random. We call this a random critical link attack. For the second scenario, we rank the critical links according to some network property. Inspired by the degree-based attack methods adopted in [21], we will rank the critical links in ascending order of their out-in degree $\delta_{i,j}$, which is defined as the sum of the out-degree of its source node d_i^{out} and the in-degree of its target node d_j^{in}, i.e., $\delta_{i,j} = d_i^{\text{out}} + d_j^{\text{in}}$. We refer to the second case as a targeted critical link attack. For both scenarios, we first remove critical links in the original networks. After all critical links are removed, the other links are removed uniformly at random. Attacks based upon critical links removal were also suggested by Mengiste et al. [23], however, only simulations results were reported.

A. The fraction l of removed links is less than the fraction of critical links l_c

Again, we will derive an approximation for the minimum number of driver nodes. We assume that, as long as the number of removed links $m \leq L_c$, the removal of each link increases the minimum number of driver nodes N_D by one. Consequently, when the number of removed links is smaller than L_c (the fraction of removed links l is smaller than l_c), the approximation for the minimum number of driver nodes N_D increases linearly with the fraction of removed links l. When the number of removed links equals the number of critical links L_c, the minimum number of driver nodes N_D equals $N_{D0} + L_c$. Thus, when the fraction l of removed links is no more than the fraction l_c of critical links, we obtain the following approximation for n_D:

$$n_{D,crit} = \frac{N_{D0} + lL}{N} \tag{9}$$

We evaluate the performance of (9) in our 8 real-world networks. Figure 6 shows that the targeted critical link attack is slightly more efficient than the random critical link in increasing the minimum number of driver nodes. Considering the small difference between the two scenarios, in the remainder of the paper, we will only consider random critical link attack, and simply refer to it as critical link attack. For all cases the approximation (9) is a good fit for sufficiently small l, while in some cases this holds for all $l \leq l_c$. We also observe that the approximation (9) provides a worst-case estimate for the number of needed driver nodes. Comparing with the critical link attack, we quantify the performance of the approximation (9) in Table VI. We use γ, the fraction of the interval $[0, l_c]$ where the absolute value of the relative error does not exceed 5%, and the absolute value of the relative error r at $l = l_c$, as the performance indicators.

TABLE VI
PERFORMANCE INDICATORS FOR THE APPROXIMATION $n_{D,crit}$ FOR THE 8 REAL-WORLD NETWORKS; $l \leq l_c$

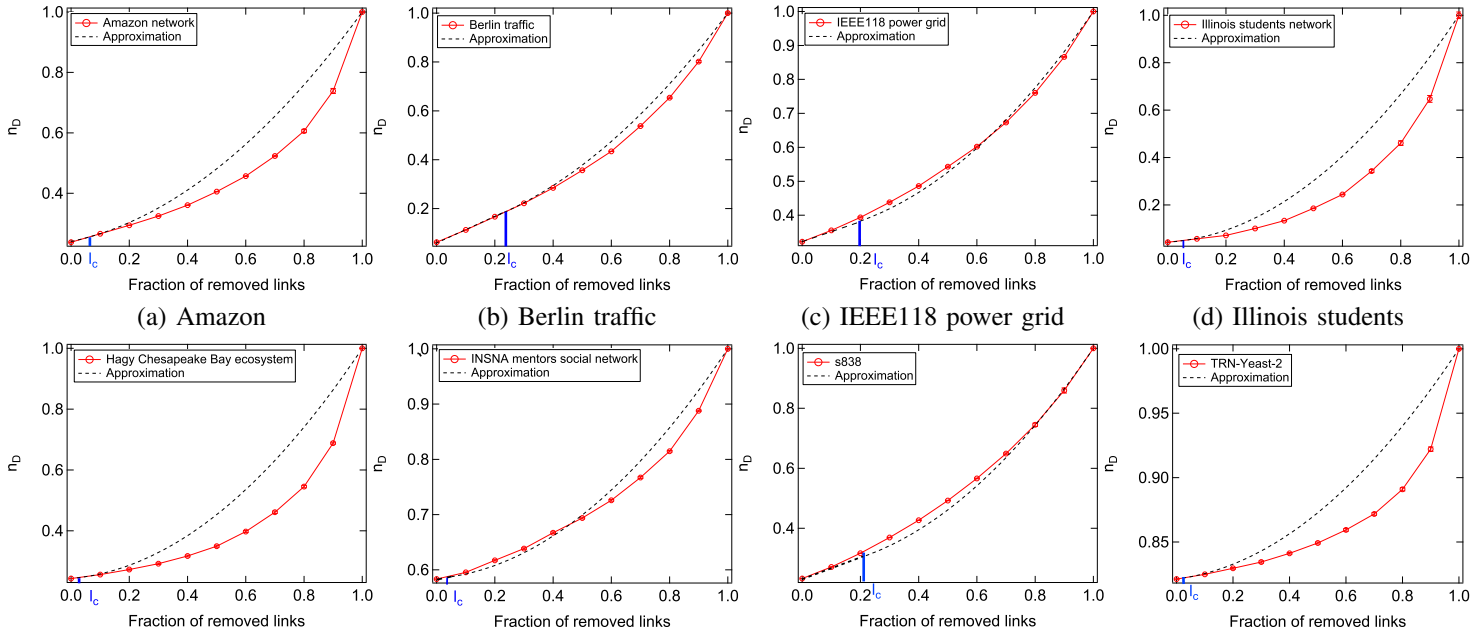
Networks	γ	r
Amazon	100%	0.68%
Berlin traffic	6.38%	79.60%
IEEE118 power grid	78.58%	7.25%
Illinois students	33.33%	22.22%
Hagy Chesapeake Bay	100%	0%
INSNA	100%	0%
s838	70%	9.88%
TRN-Yeast-2	100%	0.68%

While for 4 of the 8 considered real-world networks the approximation (9) for $n_{D,crit}$ is very good, the approximation is reasonable for two networks (IEEE118 power grid and s838) and rather poor for the remaining two (Berlin traffic and Illinois students). However, approximation (9) always seems to overestimate the normalized minimum number of driver nodes n_D and, hence, approximation (9) can be considered a worst-case approximation.

Next we evaluate the performance of (9) in synthetic networks. Figure 7 shows that our approximation Eq. (9) fits well with the simulation results in the first few removal steps. Qualitatively we observe the same behaviour as in Figure 6.

B. The fraction l of removed links is larger than the fraction of critical links l_c

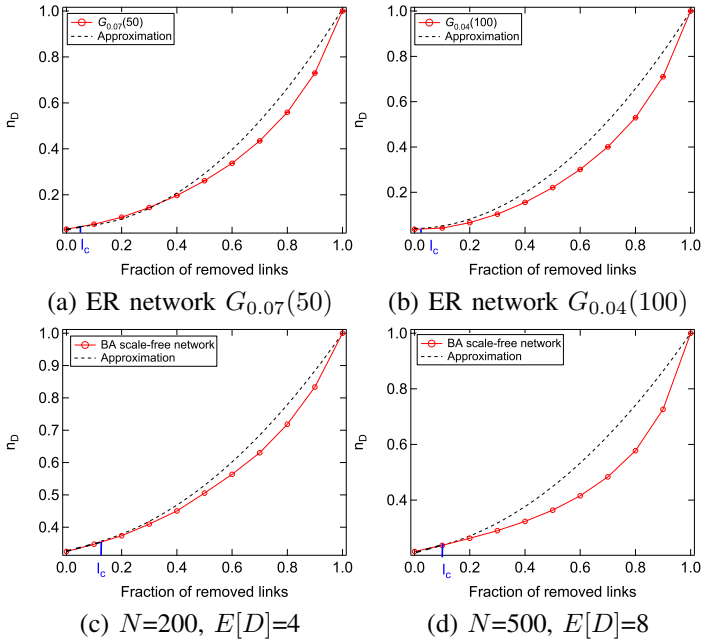
We now construct an approximation when the number of removed links is larger than L_c (the fraction of removed links l is larger than l_c), in a similar way as in the previous section. Again assuming that for $l \geq l_c$ it holds that n_D is quadratic in l, we obtain $N_D = dl^2 + el + f$. Boundary conditions are now obtained from the assumptions that the parabola passes through $(1, 1)$ and $(l_c, N_{D0} + L_C N)$ and has a zero derivative



(a) Amazon (b) Berlin traffic (c) IEEE118 power grid (d) Illinois students

(e) Hagy Chesapeake Bay (f) INSNA (g) s838 (h) TRN-Yeast-2

Fig. 4. The normalized minimum number of driver nodes n_D as a function of the fraction of removed links l in real-world networks under random attacks. In each plot, the dashed line shows the simulation results and the solid line shows our approximation. The simulation results for each fraction l is based on 10000 simulations.



(a) ER network $G_{0.07}(50)$ (b) ER network $G_{0.04}(100)$

(c) N=200, $E[D]$=4 (d) N=500, $E[D]$=8

Fig. 5. The normalized minimum number of driver nodes n_D as a function of the fraction of removed links l in synthetic networks under random attacks. The results for each fraction l is based on 10000 simulations.

at the latter point. This leads to the following approximation for n_D for all values of l:

$$n_{D,crit} = \begin{cases} \frac{N_{D0}+lL}{N} & l \le l_c \\ \frac{dl^2+el+f}{N} & l \ge l_c \end{cases} \quad (10)$$

with, $d = \frac{N-N_{D0}-l_cL}{N(l_c-1)^2}$, $e = -2dl_c$, and $f = 1 + d(2l_c - 1)$.

From Figure 8 and Figure 9, we can find the approximation $n_{D,crit}$ fits well with simulation results when the fraction of removed links is sufficiently small. When the fraction of removed links is getting larger, the difference between our approximation and simulation results is relatively large. However, in all cases the approximation seems to serve as a worst-case estimate for the number of required driver nodes. This implies that approximation (10) can have value in risk assessment studies.

V. Comparison of n_D under different attack strategies

In this section, we compare the minimum number of driver nodes for link removals under four attack strategies: (a) critical link attack (targeted attack), (b) out-in degree-based attack, (c) betweenness-based attack and (d) random attack. In the out-in degree-based attack, we remove links one by one in the ascending order of the out-in degree using the recalculated out-in degree distribution at every removal step. In the betweenness-based attack, we remove links one by one in the descending order of the betweenness using the recalculated betweenness distribution at every removal step.

Figure 10 and Figure 11 show that, for most values of l, the out-in degree-based attack is the most harmful attack strategy. In other words, the out-in degree-based attack strategy is more efficient than other attack strategies in increasing the minimum number of driver nodes N_D, and, thus, degrading the controllability of the networks. However, if the fraction of removed links is small ($l \le l_c$), the critical link attack is more

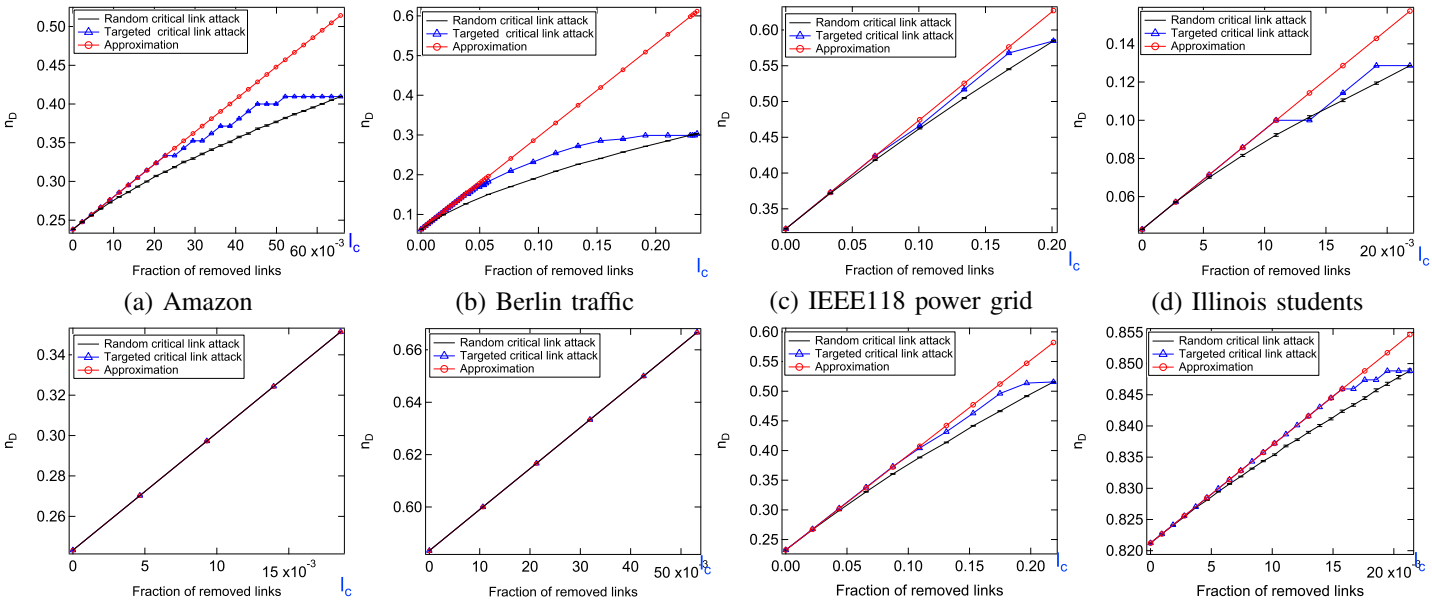


Fig. 6. Performance of the approximation for the normalized minimum number of driver nodes n_D as a function of the fraction of removed links l in real-world networks under targeted attacks. The results for each fraction l is based on 10000 simulations.

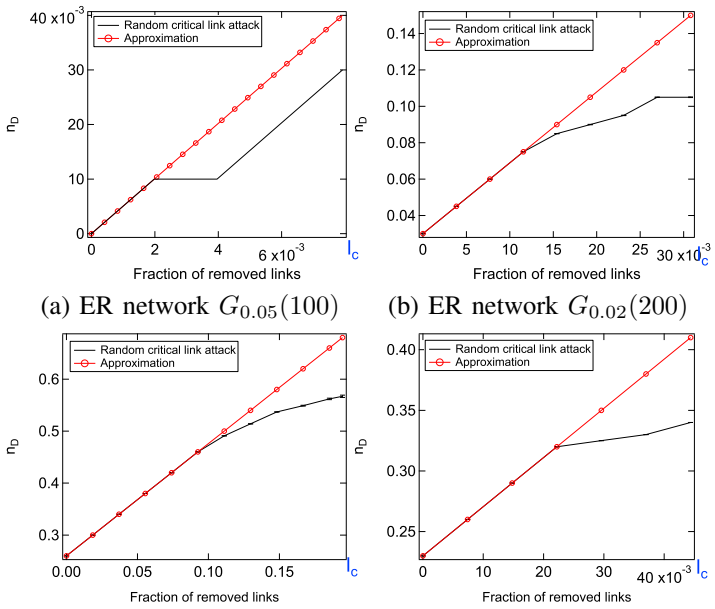


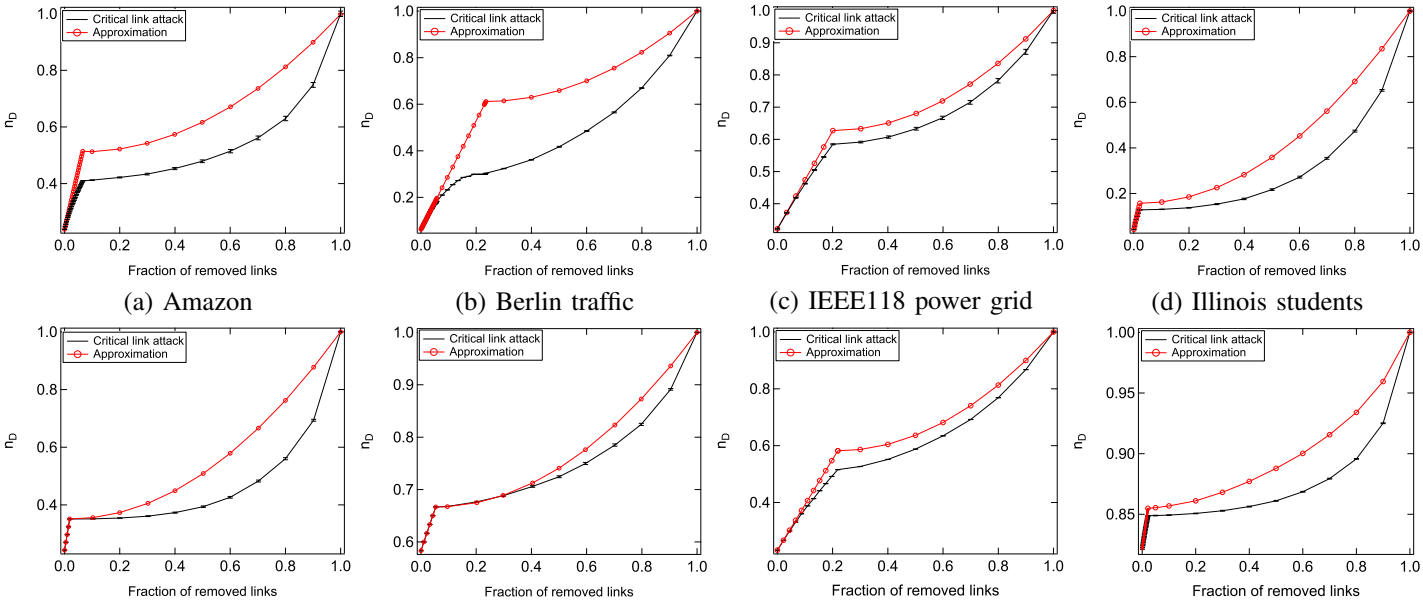
Fig. 7. Performance of the approximation for the normalized minimum number of driver nodes n_D as a function of the fraction of removed links l in synthetic networks under targeted attacks.

effective than the out-in degree based attack. The most obvious case where this happens is for the TRN-Yeast-2 network, see Figure 10(h). When the fraction of removed links becomes larger, the critical link attack becomes less effective than the out-in degree-based attack. For large values of l, the targeted attack approaches the random attack. The random attack is the least effective attack strategy.

From results in Figure 10 and Figure 11, we can deduce that the links with a small out-in degree have a strong tendency to be critical links, whose removal increases the minimum number of driver nodes N_D more efficiently. The maximum matching, which is used to determine N_D, can explain this phenomenon. As shown in Figure 1, link a and link d have a small out-in degree which equals 2. The number of matching links will decrease by 1 after removing either link a or link d. Consequently, the number of unmatched (driver) nodes will increase by one. Thus, link a and link d are critical links. Link e has a larger out-in degree which equals 4. The number of matching links is unchanged after removing link e. Link e is not a critical link. As a result, the link with a larger out-in degree is less likely to be a critical link since after removing this link, other links which share the same source or target node with this link, can also be alternative matching links.

VI. Conclusion

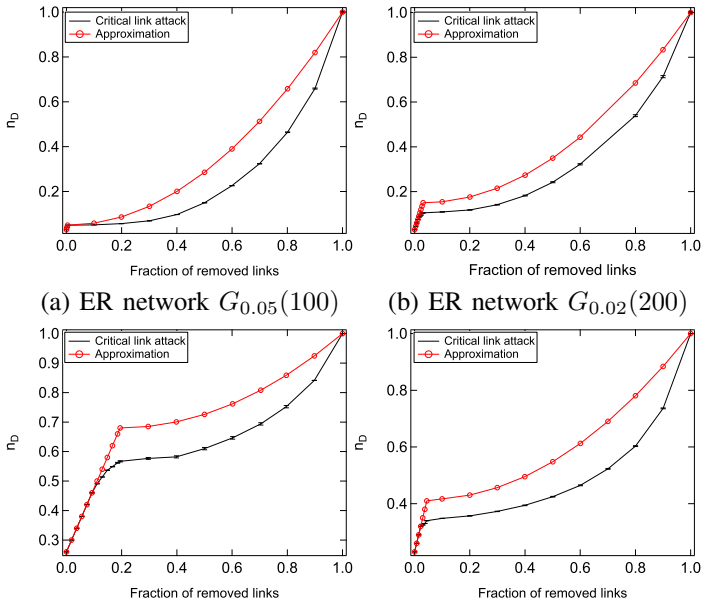
In this study, we derived analytical closed-form approximations for the minimum number of driver nodes N_D needed to control networks, as a function of the fraction of removed links, both for random and targeted attacks. Our approximations rely on the notion of critical links. As targeted attack we consider the case, where first critical links are removed. Both for random and targeted attacks, our approximation is linear in the fraction of removed links l, as long as this fraction is smaller than the fraction of critical links. For fractions of removed links larger than the fraction of critical links, our approximation is quadratic in l. We validated our approximation through simulations on real-world and synthetic networks. For random attacks, the approximation is always very good, as long as the fraction of removed links is smaller than the



(a) Amazon (b) Berlin traffic (c) IEEE118 power grid (d) Illinois students

(e) Hagy Chesapeake Bay (f) INSNA (g) s838 (h) TRN-Yeast-2

Fig. 8. Performance of the approximation for the normalized minimum number of driver nodes n_D as a function of the fraction of removed links l in real-world networks under targeted attacks.



(a) ER network $G_{0.05}(100)$ (b) ER network $G_{0.02}(200)$

(c) BA network with $E[D]$=4 (d) BA network with $E[D]$=8

Fig. 9. Performance of the approximation for the normalized minimum number of driver nodes n_D as a function of the fraction of removed links l in synthetic networks under targeted attacks.

fraction of critical links. For some cases, the approximation is still accurate for larger fractions of removed links. The approximation for attacks targeting the critical links is also accurate, as long as the fraction of removed links links is sufficiently small. The approximation for the targeted attack always serves as a worst-case estimate. Finally, we showed that the critical link attack is the most effective among 4 considered attacks, as long as the fraction of removed links is smaller than the fraction of critical links.

ACKNOWLEDGEMENT

This research was partially supported by the National Research Foundation (NRF), Prime Ministers Office, Singapore, under its National Cybersecurity R & D Programme (Award No. NRF 2014NCR-NCR001-40) and administered by the National Cybersecurity R & D Directorate. This research was also supported by the China Scholarship Council (No. 201706220113).

REFERENCES

[1] M. E. J. Newman, "The structure and function of complex networks," *SIAM REVIEW*, vol. 45, pp. 167–256, 2003.

[2] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, "Resilience of the internet to random breakdowns," *Phys. Rev. Lett.*, vol. 85, pp. 4626–4628, Nov 2000.

[3] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Controllability of complex networks," *Nature*, vol. 473, no. 7346, p. 167, 2011.

[4] Z. Yuan, C. Zhao, Z. Di, W.-X. Wang, and Y.-C. Lai, "Exact controllability of complex networks," *Nature Communications*, vol. 4, p. 2447, 2013.

[5] T. Jia, Y.-Y. Liu, E. Csóka, M. Pósfai, J.-J. Slotine, and A.-L. Barabási, "Emergence of bimodality in controlling complex networks," *Nature Communications*, vol. 4, p. 2002, 2013.

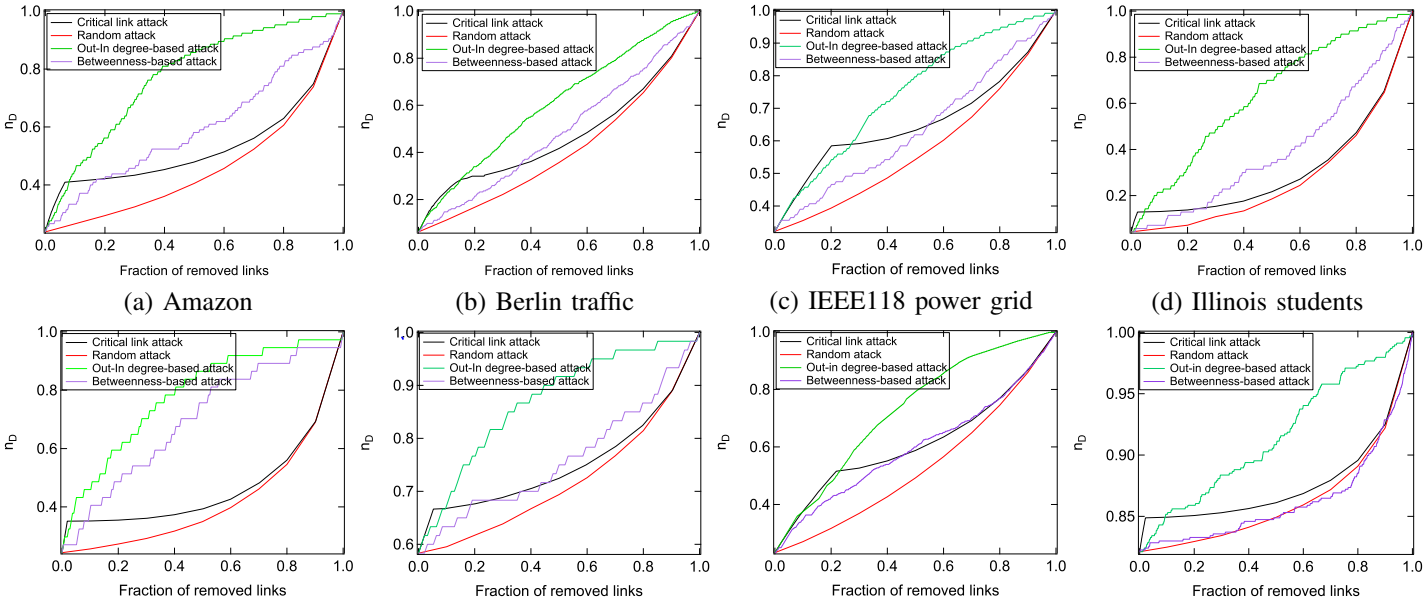
[6] T. Nepusz and T. Vicsek, "Controlling edge dynamics in complex networks," *Nature Physics*, vol. 8, no. 7, p. 568, 2012.

[7] Ching-Tai Lin, "Structural controllability," *IEEE Transactions on Automatic Control*, vol. 19, no. 3, pp. 201–208, June 1974.

[8] N. J. Cowan, E. J. Chastain, D. A. Vilhena, J. S. Freudenberg, and C. T. Bergstrom, "Nodal dynamics, not degree distributions, determine the structural controllability of complex networks," *PloS one*, vol. 7, no. 6, pp. 1–5, 06 2012.

[9] J. Ruths and D. Ruths, "Control profiles of complex networks," *Science*, vol. 343, no. 6177, pp. 1373–1376, 2014.

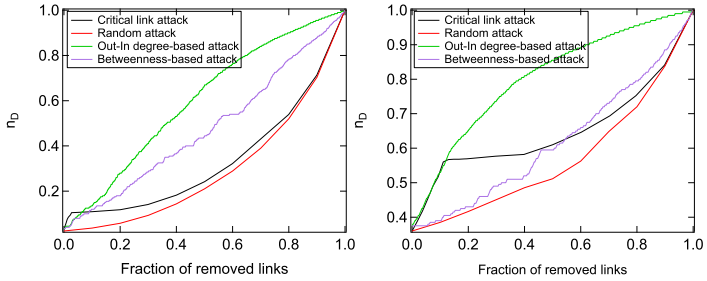
[10] G. Yan, G. Tsekenis, B. Barzel, J.-J. Slotine, Y.-Y. Liu, and A.-L. Barabási, "Spectrum of controlling and observing complex networks," *Nature Physics*, vol. 11, no. 9, p. 779, 2015.



(a) Amazon (b) Berlin traffic (c) IEEE118 power grid (d) Illinois students

(e) Hagy Chesapeake Bay (f) INSNA (g) s838 (h) TRN-Yeast-2

Fig. 10. Performance of different attack strategies in real-world networks



(a) ER network $G_{0.02}(200)$ (b) BA network with $E[D]$=4

Fig. 11. Performance of different attack strategies in synthetic networks

[11] H. Cetinay, K. Devriendt, and P. Van Mieghem, "Nodal vulnerability to targeted attacks in power grids," *Applied Network Science*, vol. 3, no. 1, p. 34, 2018.

[12] B. Berche, C. von Ferber, T. Holovatch, and Y. Holovatch, "Resilience of public transport networks against attacks," *The European Physical Journal B*, vol. 71, no. 1, pp. 125–137, 2009.

[13] X. Wang, E. Pournaras, R. E. Kooij, and P. Van Mieghem, "Improving robustness of complex networks via the effective graph resistance," *The European Physical Journal B*, vol. 87, no. 9, p. 221, 2014.

[14] A. Socievole, F. De Rango, C. Scoglio, and P. Van Mieghem, "Assessing network robustness under SIS epidemics: The relationship between epidemic threshold and viral conductance," *Computer Networks*, vol. 103, pp. 196–206, 2016.

[15] X. Wang, Y. Koç, S. Derrible, S. N. Ahmad, W. J. Pino, and R. E. Kooij, "Multi-criteria robustness analysis of metro networks," *Physica A: Statistical Mechanics and its Applications*, vol. 474, pp. 19–31, 2017.

[16] X. Wang, R. E. Kooij, and P. Van Mieghem, "Modeling region-based interconnection for interdependent networks," *Physical Review E*, vol. 94, no. 4, p. 042315, 2016.

[17] S. Trajanovski, J. Martín-Hernández, W. Winterbach, and P. Van Mieghem, "Robustness envelopes of networks," *Journal of Complex Networks*, vol. 1, no. 1, pp. 44–62, 2013.

[18] Y. Koc, M. Warnier, P. Van Mieghem, R. E. Kooij, and F. M. Brazier, "The impact of the topology on cascading failures in a power grid model," *Physica A: Statistical Mechanics and its Applications*, vol. 402, pp. 169–179, 2014.

[19] C.-L. Pu, W.-J. Pei, and A. Michaelson, "Robustness analysis of network controllability," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 18, pp. 4420–4425, 2012.

[20] S. Nie, X. Wang, H. Zhang, Q. Li, and B. Wang, "Robustness of controllability for networks based on edge-attack," *PloS one*, vol. 9, no. 2, p. e89066, 2014.

[21] J. Thomas, S. Ghosh, D. Parek, D. Ruths, and J. Ruths, "Robustness of network controllability to degree-based edge attacks," in *International Workshop on Complex Networks and their Applications*. Springer, 2016, pp. 525–537.

[22] Z.-M. Lu and X.-F. Li, "Attack vulnerability of network controllability," *PloS one*, vol. 11, no. 9, p. e0162289, 2016.

[23] S. Abebe Mengiste, A. Aertsen, and A. Kumar, "Effect of edge pruning on structural controllability and observability of complex networks," *Scientific Reports*, vol. 5, p. 18145, 12 2015.

[24] A. Lombardi and M. Hörnquist, "Controllability analysis of networks," *Physical Review E*, vol. 75, no. 5, p. 056110, 2007.

[25] R. E. Kalman, "Mathematical description of linear dynamical systems," *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, vol. 1, no. 2, pp. 152–192, 1963.

[26] Y. Yang and G. Xie, "Mining maximum matchings of controllability of directed networks based on in-degree priority," in *2016 35th Chinese Control Conference (CCC)*. IEEE, 2016, pp. 1263–1267.

[27] J. E. Hopcroft and R. M. Karp, "An n^5/2 algorithm for maximum matchings in bipartite graphs," *SIAM Journal on Computing*, vol. 2, no. 4, pp. 225–231, 1973.

[28] R. Rossi and N. Ahmed, "The network data repository with interactive graph analytics and visualization." in *Proc. of the 29th AAAI Conference on Artificial Intelligence*, vol. 15, 2015, pp. 4292–4293.

[29] O. Jahn, R. H. Möhring, A. S. Schulz, and N. E. Stier-Moses, "System-optimal routing of traffic flows with user constraints in networks with congestion," *Operations research*, vol. 53, no. 4, pp. 600–616, 2005.

[30] P. Crucitti, V. Latora, and M. Marchiori, "A topological analysis of the italian electric power grid," *Physica A: Statistical mechanics and its applications*, vol. 338, no. 1-2, pp. 92–97, 2004.

[31] M. A. Evans and D. Scavia, "Forecasting hypoxia in the Chesapeake Bay and Gulf of Mexico: Model accuracy, precision, and sensitivity to ecosystem change," *Environmental Research Letters*, vol. 6, no. 1, p. 015001, 2010.

[32] D. R. White, "Rethinking the role concept," *Research methods in social network analysis*, p. 429, 2017.