SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
PHASE TRANSITIONS AND CRITICAL PHENOMENA
STATISTICAL PHYSICS
COMPUTER SCIENCE
COMPLEX NETWORKS

Received
8 October 2014

Accepted
14 January 2015

Published
9 February 2015

Correspondence and requests for materials should be addressed to F.M. (molnaf@rpi.edu)

Building Damage-Resilient Dominating Sets in Complex Networks against Random and Targeted Attacks

F. Molnár, Jr.[1,2], N. Derzsy[1,2], B. K. Szymanski[2,3] & G. Korniss[1,2]

[1]Department of Physics, Applied Physics, and Astronomy, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY, 12180-3590 USA, [2]Social Cognitive Networks Academic Research Center, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY, 12180-3590 USA, [3]Department of Computer Science, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY, 12180-3590 USA.

We study the vulnerability of dominating sets against random and targeted node removals in complex networks. While small, cost-efficient dominating sets play a significant role in controllability and observability of these networks, a fixed and intact network structure is always implicitly assumed. We find that cost-efficiency of dominating sets optimized for small size alone comes at a price of being vulnerable to damage; domination in the remaining network can be severely disrupted, even if a small fraction of dominator nodes are lost. We develop two new methods for finding flexible dominating sets, allowing either adjustable overall resilience, or dominating set size, while maximizing the dominated fraction of the remaining network after the attack. We analyze the efficiency of each method on synthetic scale-free networks, as well as real complex networks.

Dominating sets play a critical role in complex networked systems by providing efficient sources of influence and information dispersal, or hubs of surveillance[1–4], and are applied in social, infrastructure, and communication networks[5–7]. Most recently, dominating sets were employed to controllability in complex networks[8–11], observability of the power-grid[12], and to finding high-impact optimal subsets in protein interaction networks[13]. While finding the smallest, most efficient dominating set has gained significant interest, it is also important to understand how robust these dominating sets are against various forms of network damage[14].

By definition, a dominating set is a subset of nodes in a network, such that every node not in the dominating set is adjacent to at least one node in this set; in other words, every node has at least one neighbor (or itself) in the dominating set. The smallest cardinality dominating set is the minimum dominating set (MDS), which is of particular interest, because it provides the most cost-efficient solution for network control, assuming a constant per-node cost of implementing control, in fixed or slowly evolving networks. Research has been focused on finding bounds for the size of MDS[1,15], finding approximations to the MDS[16,17], understanding its expected scaling behavior in complex networks[3,4], and studying the impact of assortativity[36,37] on network domination[2,18].

Attacks on complex networks, fault tolerance, and defense strategies against damage of nodes and edges have also gained significant interest in network science[19–22]. Networks with scale-free topologies have been found to be resilient against random node damage, but vulnerable to targeted removal of high degree nodes[23–25]. Research has also focused on improving the robustness of these networks against various combinations of attacks[26–28], and on studying the dynamically progressing effects of an initial damage, such as cascading failures[29,30].

The connectivity of the surviving network structures and the fraction of the remaining set of nodes still dominated following failures or attacks are both essential for sustainable network operations and carrying out network functions. While the former (structural integrity) has been studied in great detail over the past two decades[19–25], the latter (domination stability) has not received any attention.

We assume that the network damage is relatively small, and although the network may become fragmented due to the loss of nodes, we assume it remains functional. In such cases efficient domination over the network is still important and desirable, just as it is in undamaged networks. However, considering that most dominating set search methods aim for the smallest possible set size (and corresponding cost) in a fixed topology network, even a small damage could severely disrupt the complete domination "coverage". Our goal is to understand how fragile dominating sets are, how to improve them, and ultimately to provide new methods for selecting dominating sets with adjustable balance between resilience and cost.

The resilience of a dominating set against network damage is measured by *domination stability*, which we define as the fraction of the network still dominated after some nodes (which may include nodes from the dominating set) are removed from the network:

$$s(f) := \frac{\left|\bigcup_{j\in \mathrm{DS}} \mathrm{N}^{+}(j)\right|}{N(1-f)}, \qquad (1)$$

where DS is the subset of the original network's dominating set that remains after network damage, f is the fraction of nodes removed from the network, and $\mathrm{N}^{+}(j)$ is the surviving closed neighborhood of node j following network damage. In order to measure stability, we need to simulate network damage by actually removing nodes from the network and calculating the remaining dominated fraction.

Domination stability depends not only on the fraction of removed nodes, but also on the order in which nodes have been removed from the network. Similarly to many studies in the literature, we consider two damage scenarios: random and targeted node removals. The random node removal strategy models network damage produced by natural causes or errors, while the targeted node removal method reflects the impact of intentional, targeted attacks on a network. In the random damage scenario nodes are removed with equal probability, in random order. In case of targeted attacks, the nodes are removed in degree-ranked order, with highest degrees being removed first. We indicate which strategy we consider in the subscript of stability: s_{rand} denotes the stability against random damage, and s_{deg} corresponds to the stability against targeted attack (interchangeably denoted as degree-ranked removal).

Results

Stability of Various Fixed Dominating Sets. We start our analysis by measuring the stability of three different dominating sets, that we use for baseline comparison with our new methods. These are the following:

- greedy minimum dominating set (MDS)[1,4,31], where nodes are selected by a sequential greedy search algorithm in order to approximate the actual (NP-hard) smallest dominating set,
- "cutoff" dominating set (CDS)[18], where all nodes above a degree threshold are selected into set X, and the nodes not dominated by any nodes in set X are selected into set Y. The dominating set is then given by $X \cup Y$. The degree threshold is selected such that it minimizes the size of the resulting dominating set,
- degree-ranked dominating set (DDS), where we select all nodes in decreasing order of degree (with random tie-breaking) as dominators until the selected set dominates the entire network.

Our first choice is MDS, due to its importance in cost-efficient control of complex networks, and because it provides a high-quality approximation to the actual smallest dominating set. The other methods we have chosen are potentially useful when finding the greedy MDS or solving the binary integer programming equivalent is impractical, e.g., when the adjacency information of the network is incomplete, or the network is too large to run these algorithms in a reasonable amount of time. In these cases heuristic algorithms, such as CDS or DDS can find suboptimal (not the smallest possible), yet small enough dominating sets that are still useful for practical applications. In particular, the excess nodes selected by these methods may help to increase domination stability.

Figure 1 shows the stability against the fraction of removed nodes for MDS, CDS and DDS in the entire remaining network [Fig. 1(a), (b)] and in the remaining giant component [Fig. 1(c), (d)]. It is clear that the degree-ranked node removal reduces the dominated fraction much faster than the random node removal, because high-degree nodes are more likely to be dominator nodes than low degree nodes. The giant component itself also breaks down much faster, as shown in the insets of Fig. 1(c) and (d). However, as long as a giant component exists, it has higher domination stability than the entire network, in both scenarios. The slight increase of stability at high damage rates is a side effect caused by removal of nodes that had lost domination by earlier removals. When the network damage is high, it becomes more likely that these nodes are deleted, causing the dominated fraction of the remaining network to increase. At this point, however, the network is almost completely destroyed and domination stability becomes meaningless.

The stability curves show much more disturbed shapes in degree-ranked removal than random removal, due to the differences in the degree structure of each dominating set. In MDS, there is no preference toward any particular node degree during selection of dominators (besides the natural effect of the greedy selection, where the high-degree nodes provide a larger increase in the number of dominated nodes, hence they are more likely to be selected), which means that removal of high-degree nodes has a smooth (albeit strong) impact on stability. In CDS, we can see a fast initial drop as we remove the very high degree nodes that were specifically selected for dominators (in set X), then continuing at a more gentle slope as the dominators from the Y set are removed, since any node that was not dominated by X, regardless of degree, may be in set Y. Although the Y set may seem wasteful in CDS construction, with the right degree threshold the size of the CDS is actually very close to the MDS[18], and the excess nodes provide a fair increase in stability. DDS is the simplest but most inefficient method for finding a dominating set because it selects *all* nodes starting from the highest degrees until all nodes are dominated. However, the resulting redundancy of dominators in the network is providing the highest stability of all three methods.

We can also observe the general tendency that a larger dominating set provides higher stability. At any given fraction of removed nodes, there is a positive correlation between stability and the size of the original dominating set, in both random [Fig. 1(a)] and degree-ranked [Fig. 1(b)] node removals. We clearly illustrate this correlation in Fig. 1(e) and (f), where we show stability as a function of the dominating set size, at various damage levels. This means that the MDS, which is the smallest (most cost-efficient) dominating set, is also the most vulnerable, to both random damage and targeted attacks.

Note, that Fig. 1 only shows the stability for networks with a certain degree exponent that are uncorrelated (i.e., with Spearman's[38–40] $\rho = 0$). Stabilities at different values of these parameters are presented in Supplementary Figures S1–S5.

We have also included supplementary videos to illustrate the evolution of domination stability as the network disintegrates, during random node removal (Supplementary Movie 1) and degree-ranked node removal (Supplementary Movie 2).

The main conclusion we can draw is that the larger number of dominating nodes selected by heuristic methods CDS and DDS, compared to the smaller and more optimal MDS, can effectively increase the stability of domination. However, all three methods are "fixed" in the sense that they give only a single possible dominating set size (and corresponding stability) for a given network.

Flexible-Redundancy Dominating Set (frDS). In order to overcome the limitations of fixed methods, we must analyze in detail how domination is lost when the network is damaged. First, we realize that loss of domination occurs locally at each node: those nodes that lose all dominators will reduce the domination stability of the network. Therefore, stability can be expressed locally, as the domination *redundancy* of each node. This quantity simply counts how many dominating nodes are within the closed neighborhood of a given node. A large dominating set can successfully increase domination stability, if the extra nodes are distributed in a way that they increase domination redundancy on many nodes. This seems to occur naturally for CDS and DDS, however we cannot guarantee that redundancy was increased in the most optimal way (relative to MDS), nor can we control the number of selected nodes.

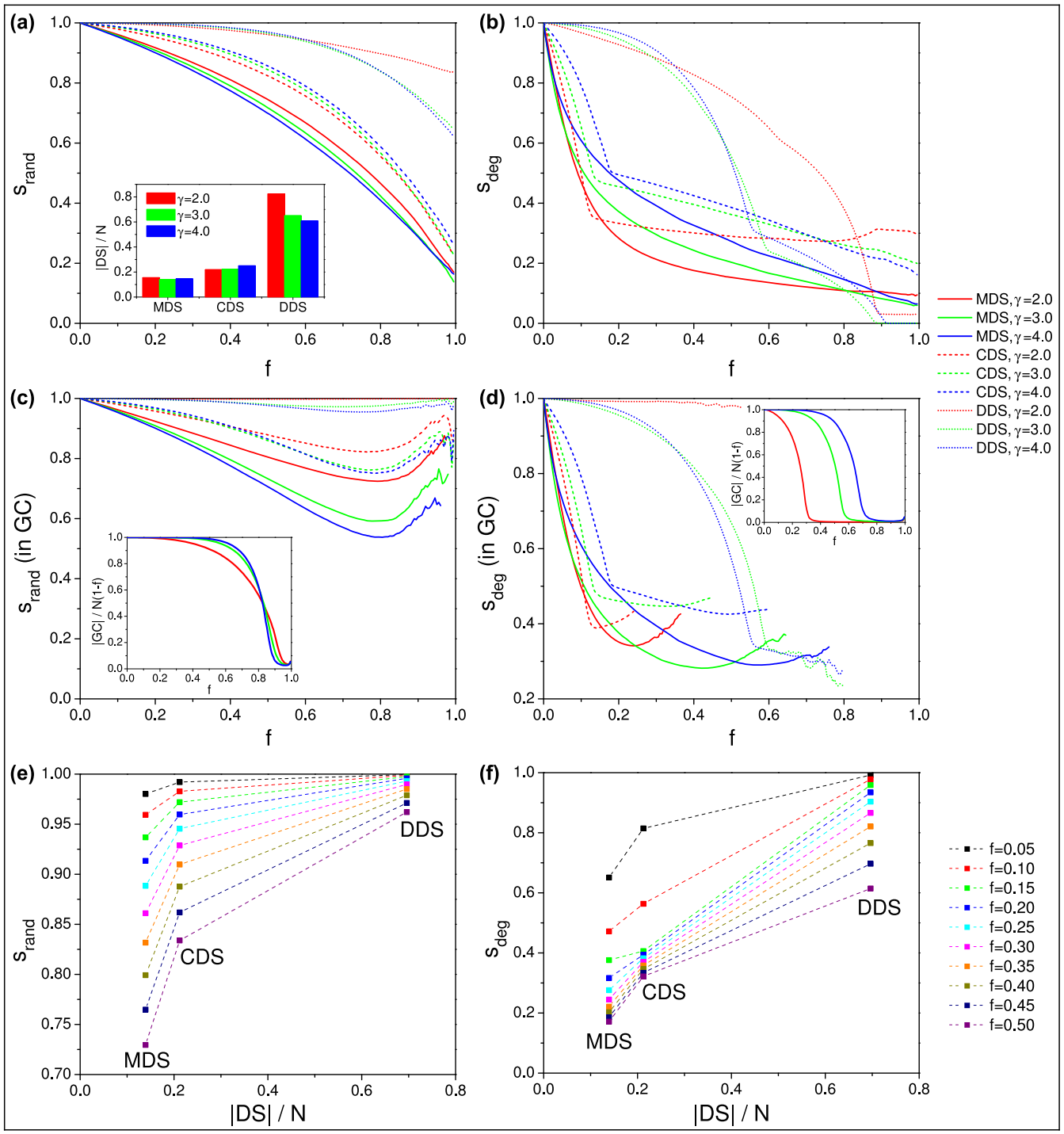

Figure 1 | Stability of various dominating sets against random and degree-ranked node removal. Subfigures (a), (c), and (e) show random node removal, (b), (d), and (f) show degree ranked node removal. Subfigures (a) and (b) show stability in the entire network, while (c) and (d) show stability within the remaining giant component. The inset in (a) shows the corresponding sizes of dominating sets, and insets in (c) and (d) show the size of the corresponding giant component. Subfigures (e) and (f) show a correlation between set size and stability, at $\gamma = 2.5$. All plots show synthetic scale-free networks, $N = 5000$, $\langle k \rangle = 8$, averaged over 200 network samples.

We introduce the flexible-redundancy dominating set (frDS) to solve these problems. We explicitly set an average domination redundancy in the network, denoted by r, that must be guaranteed by frDS, while aiming for minimum set size. Note, that $r = 1$ is equivalent to the minimum dominating set (MDS), and when r is an integer, the frDS is identical to the *h-dominating set* (with $h = r$) studied by Cooper, et al.[41]. Finding an frDS is most likely NP-hard, since it is also NP-hard to find an MDS[42] or an h-dominating set[43], but we can use a modified greedy algorithm to find an approximation.

The steps of finding an frDS are as follows. First, we assign a domination redundancy requirement, $r(i)$ for each node i as an integer value indicating at least how many dominators node i must have in the dominating set. Given the desired average (non-integer) r value for the entire network, we assign the nearest integer values

$\lfloor r \rfloor$ and $\lceil r \rceil$ to each node randomly, such that the network average will be r (the probability of assigning $\lceil r \rceil$ is $r - \lfloor r \rfloor$, which is analogous to a biased coin toss). For the greedy selection we define a dominating potential $p(i)$ as the number of nodes in the closed neighborhood of i that have not yet reached their domination requirement, and therefore selecting node i can help them advance toward their goal. (Note, by definition, the potential of an already selected node is zero.) At each greedy step we select one node with maximum dominating potential (with random tie-breaking), until the requirements of all nodes have been fulfilled. Note, that since dominating potential is an integer number between 0 and N, nodes can be sorted according to their potential in $O(N)$ steps, and it is possible to maintain sortedness after changing the potential of a node in $O(1)$ step (see Supplementary Note 1 for further details and pseudocode). This results in the same computational time complexity as for the greedy MDS approximation, $O(E)$. Also note, that if $r > N$, then the node requirements can never be satisfied, in which case the greedy selection naturally falls back to selecting nodes in degree-ranked order, because at every step every neighbor of a node may be advanced toward its goal.

Flexible-Cost Dominating Set (fcDS). When we aim for a desired dominating set size (cost level, i.e., having a limited budget), we can, in principle, aim for the necessary redundancy level in frDS to achieve that desired cost. However, we can further improve stability by considering the expected attack pattern on the network (if the information is available), and optimize the selected dominating set accordingly. For example, if the attack is expected at high-degree nodes, we should avoid selecting many of those nodes as dominators, despite their ability to cover large fractions of the network.

We can optimize our choice of dominators by including the probability of losing each node into the calculation of local stability, which we aim to maximize. First, we assign a strength value $s(i) \in (0, 1)$ to each node i, which represents the a-priori estimated probability for not losing that node after the attack (i.e., the anticipated attack pattern). Then, we calculate the current domination stability of node i as follows:

$$\text{stability}(\text{DS}, i) = \begin{cases} 0 & \text{if } \text{DS} \cap N^+(i) = \varnothing \\ 1 - \prod_{j \in \text{DS} \cap N^+(i)} (1 - s(j)) & \text{otherwise,} \end{cases} \quad (2)$$

which is the probability that node i will remain dominated (not lose all dominators), assuming nodes will be deleted independently; DS denotes the currently selected dominating set. For selecting the next dominator, we choose one that increases the total stability of the network maximally. The total potential increase of stability can be calculated for each node as follows:

$$\text{potential}(i) = \sum_{j \in N^+(i)} \text{stability}(\text{DS} \cup \{i\}, j) - \text{stability}(\text{DS}, j) \quad (3)$$

$$= \sum_{j \in N^+(i)} (1 - \text{stability}(\text{DS}, i)) \cdot s(i). \quad (4)$$

Therefore, we always select a node with maximum potential (with random tie-breaking). Note, that unlike in frDS, the potential here is a non-integer value, thus we can only use comparative sorting to order nodes by potential, which needs $O(N \log N)$ steps. In addition, after selecting each dominator, the stability values have to be recomputed in the selected node's closed neighborhood, and the potentials for nodes with distance up to two from this node. This involves $O(d^2)$ nodes, where d is the average degree. Thus, maintaining sortedness of nodes by their potential requires $O(d^2 \log N)$ steps after selecting each dominator.

In order to compare stability of fcDS with frDS and other dominating sets, we calculate the "a-priori" node strength values as follows: $s(i) = 0.5$ for random node removal, and $s(i) = 1 - d(i)/N$ for degree-ranked node removal. Here, we assume the size of the anticipated damage is unknown, thus strength values are expressing relative probabilities only. The strength value for a random damage is arbitrary, as long as it is uniform among the nodes, and it is inversely proportional to node degree in a degree-ranked attack. Further details of fcDS and pseudocode are included in Supplementary Note 2.

Stability Comparison of Dominating Sets. We seek to answer two main questions in our analysis. First, we want to see how much stability we can achieve by selecting various sizes of dominating sets (in other words, how does the stability scale with larger invested cost of domination). Second, we want to know how much more efficient our methods are compared to the fixed dominating sets, that is, given the same size of dominating set as MDS, CDS, or DDS, how much higher stability can our methods provide.

Figures 2 and 3 show domination stability achieved by frDS and fcDS as a function of redundancy and dominating set size, respect-

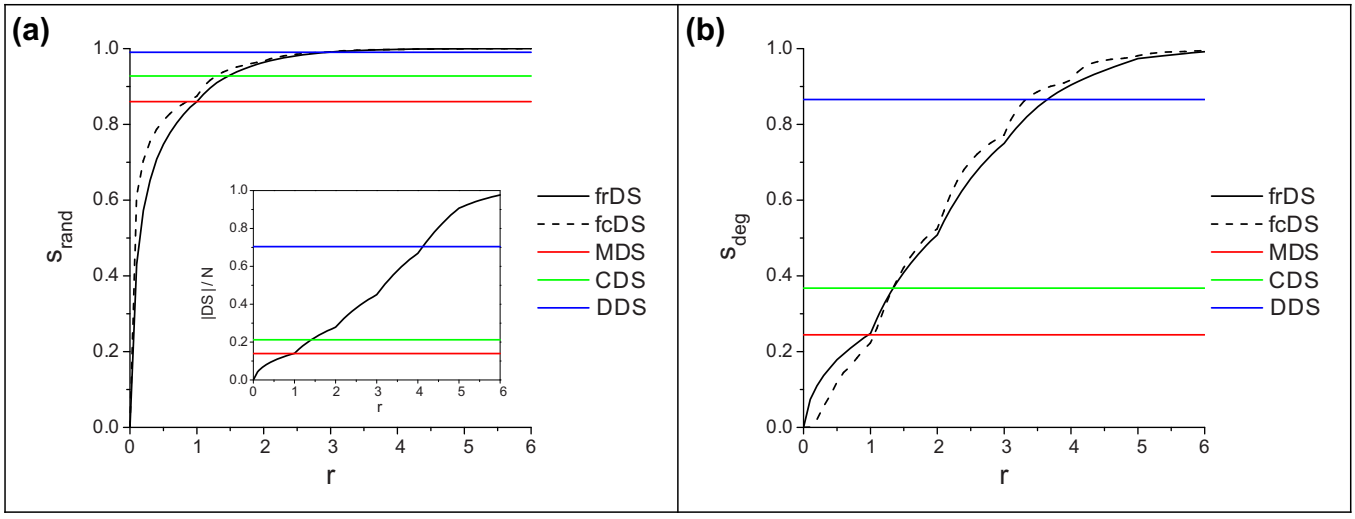

Figure 2 | Domination stability in frDS and fcDS as a function of domination redundancy. (a) shows random node removal, (b) shows degree-ranked node removal. The inset shows the sizes of the corresponding dominating sets. The size of fcDS is set to match frDS at any given r value. Synthetic scale-free networks, $N = 5000$, $\langle k \rangle = 8$, $\gamma = 2.5$, $f = 0.3$, averaged over 200 network samples.

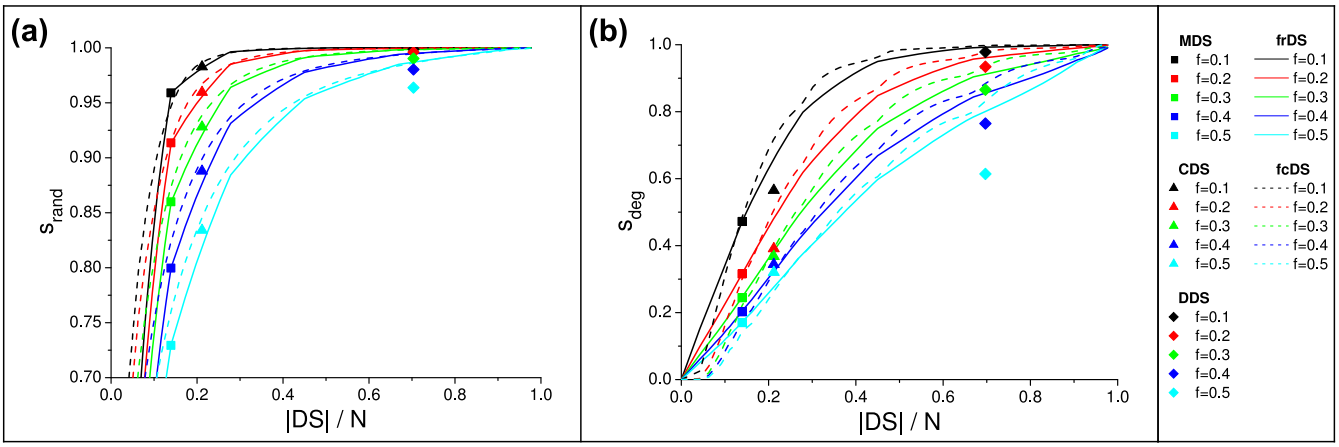

Figure 3 | Stability of frDS and fcDS as a function of dominating set size (cost) for various network damage fractions. Stabilities of MDS, CDS, and DDS are presented at their corresponding cost values. Subfigure (a) shows random node removal, (b) shows degree-ranked node removal, for synthetic scale-free networks, $N = 5000$, $\langle k \rangle = 8$, $\gamma = 2.5$, averaged over 200 network samples.

ively. Stability achieved by the fixed methods (MDS, CDS, DDS) are also shown at their corresponding cost values for comparison. The general shape of the curves in both figures are similar, since the dominating set size is roughly proportional to redundancy (see Fig. 2 inset and Supplementary Fig. S6). In case of random damage, the stability rapidly increases with cost, until the size of MDS is reached, then the curve saturates. There is little advantage in selecting a dominating set larger than approximately twice the size of MDS, because stability is already very close to 1, even at large damage values. However, in case of degree-ranked damage, there is a steady increase in stability as more nodes are selected as dominators. In both cases, fcDS provides somewhat higher stability than frDS at moderate damage levels, but frDS is more stable at small damage levels. These observations hold across a wide range of network parameters, see Supplementary Figs. S7 and S8. It is also clear that both frDS and fcDS can provide great flexibility in adjusting the size of the dominating set and stability.

The stability of frDS and fcDS at cost levels identical to MDS, CDS, and DDS are presented in Fig. 4. Our results show that frDS provides stability values very similar to the fixed methods (in case of MDS, it is identical by definition, thus it is not shown), while fcDS shows a minor improvement in stability. On the other hand, both frDS and fcDS show significant improvement over the fixed methods against degree-ranked attacks, at low network damage fractions. MDS and CDS show a tipping point in damage, where these methods become slightly more effective than frDS or fcDS, but the difference is minimal, and it occurs only at moderate to high network damage ($f \gtrsim 0.3$).

Stability in Real Networks. We analyze stability of frDS and fcDS, as well as other dominating sets, in several real complex networks, listed in Table 1. These include an internet peer-to-peer network (p2p-Gnutella08)[44], the power transmission network of continental Europe (ENTSO-E power-grid)[45,46], and one brain graph extracted from MRI data (KKI21-KKI2009-19)[47,48]. Note, that we only use the giant component of these networks. A brief analysis of the degree distribution of Gnutella08 is provided in Supplementary Figs. S17–S19; degree distribution of the powergrid is provided in Supplementary Figs. S20–S23.

The brain graph we analyze here (KKI-21-KKI2009-19) is one of 200 graphs available from[47]. These graphs have peculiar structural properties, and are very similar to each other. In particular, all brain graphs are very dense: $\langle k \rangle \approx 150$ (Supplementary Fig. S23); they are all very assortative[36,37]: $\rho \approx 0.6$ (Supplementary Fig. S24); and they have very similar degree distributions (see Supplementary Figs. S25–S27). It is also interesting that the size of MDS is very small, only 3–4% network size, while the size of CDS and DDS is very large, around 60% and 100% of network size, respectively (Supplementary Fig. S28). We attempt to separate the effects of density and assortativity in order to identify their impact on domination stability.

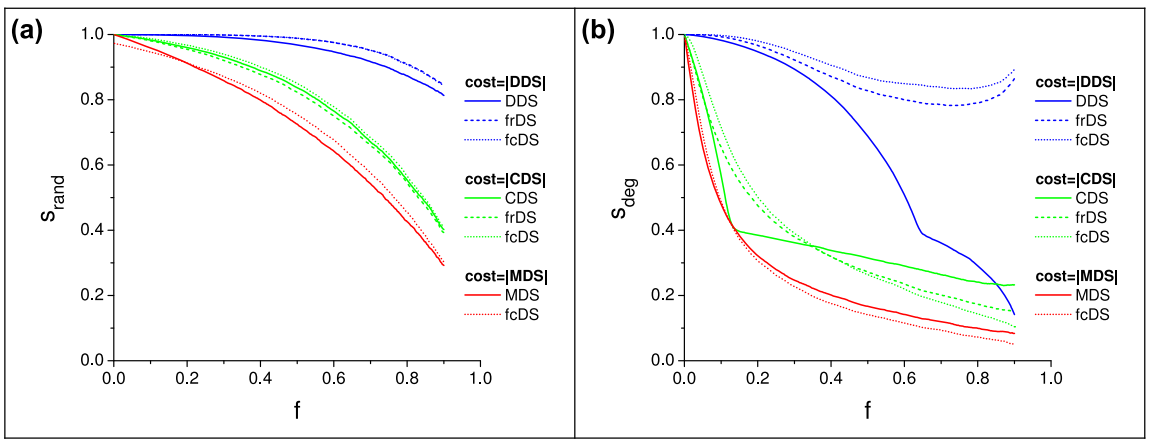

Figure 4 | Comparison of domination stability at fixed cost levels, as a function of network damage fraction. Stability of frDS and fcDS are plotted at cost values identical to MDS, CDS and DDS. Subfigure (a) shows random node removal, (b) shows degree-ranked node removal, for synthetic scale-free networks, $N = 5000$, $\langle k \rangle = 8$, $\gamma = 2.5$, averaged over 200 network samples.

Table 1 | Parameters of real networks used in our analysis. The data refers exclusively to the giant component

Name	Source	N	k_{min}	k_{max}	$\langle k \rangle$	Spearman's ρ[38–40]
Gnutella08	[44]	6299	1	97	6.60	0.03
powergrid	[45, 46]	1494	1	13	2.89	−0.18
KKI-21-KKI2009-19	[47, 48]	712098	1	6505	138.2	0.62

Figure 5 shows domination stability as a function of dominating set size for the real network samples. In general, we see that stability of frDS and fcDS matches the stability of MDS, and exceeds the stability of CDS and DDS, at identical set sizes. In case of Gnutella08 and the powergrid, the stability curves saturate slowly, and the curve shapes are not as smooth as for synthetic scale-free networks, due to having more disturbed (non-scale-free) degree distributions. However, the brain graph shows very high domination

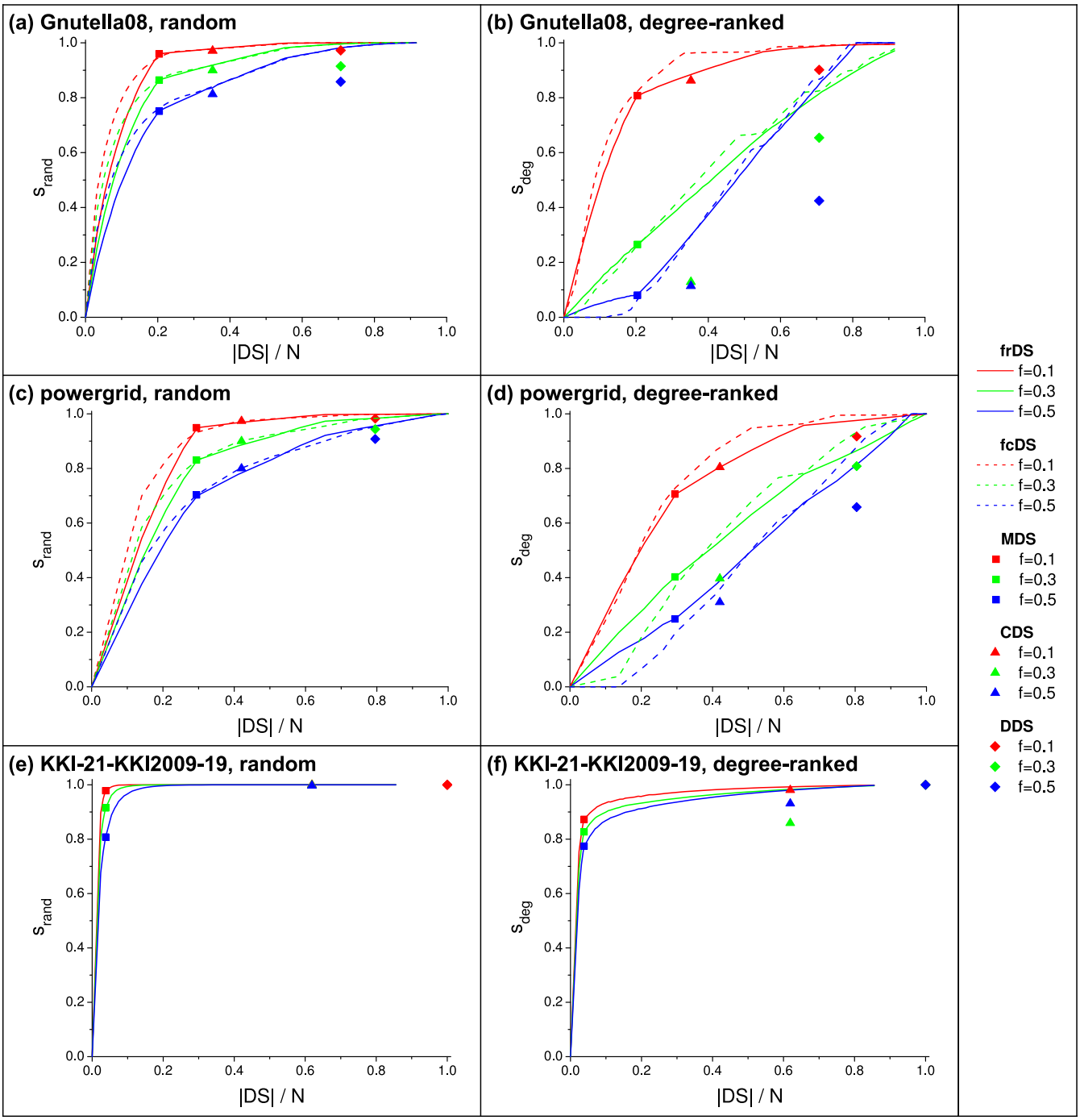

Figure 5 | Stability of frDS, fcDS and other dominating sets in real networks against random and degree-ranked attacks, for various damage fractions: (a,b) Gnutella peer-to-peer network; (c,d) ENTSO-E powergrid; (e,f) Brain (MRI) network. Data is averaged over 20 independent runs of node removal. See Table 1 for network parameters.

stability against both random and targeted attacks. In all cases, the relative advantage of frDS and fcDS over CDS and DDS (i.e., cost-efficiency) remains as high as in synthetic scale-free networks.

We can observe the effects of assortativity separately from other structural properties by artificially changing the network's assortativity, using a biased edge-mixing method (see in Ref. 18 and Supplementary Note 3), which rewires the edges in the graph, while keeping the degree sequence unchanged. Using this method we present a brief analysis of dominating set size vs. assortativity in Supplementary Figs. S29–S31. In general, we see the expected behavior that dominating sets tend to become larger in more assortative networks[18]. Note, that the size of DDS in the brain graph (Supplementary Fig. S31) being 100% of nodes regardless of assortativity is the result of a particular topological feature; there are a small number of leaves (degree 1 nodes) connected to degree 2 nodes, thus DDS has to select al nodes down to degree 2 (essentially all nodes) to dominate these off-hanging leaves — a feature left unchanged by edge-mixing.

Figure 6 presents the effects of assortativity on domination stability. We see an unexpected behavior: as assortativity increases, domination stability decreases against random damage, but increases against an attack on high-degree nodes. We can understand this

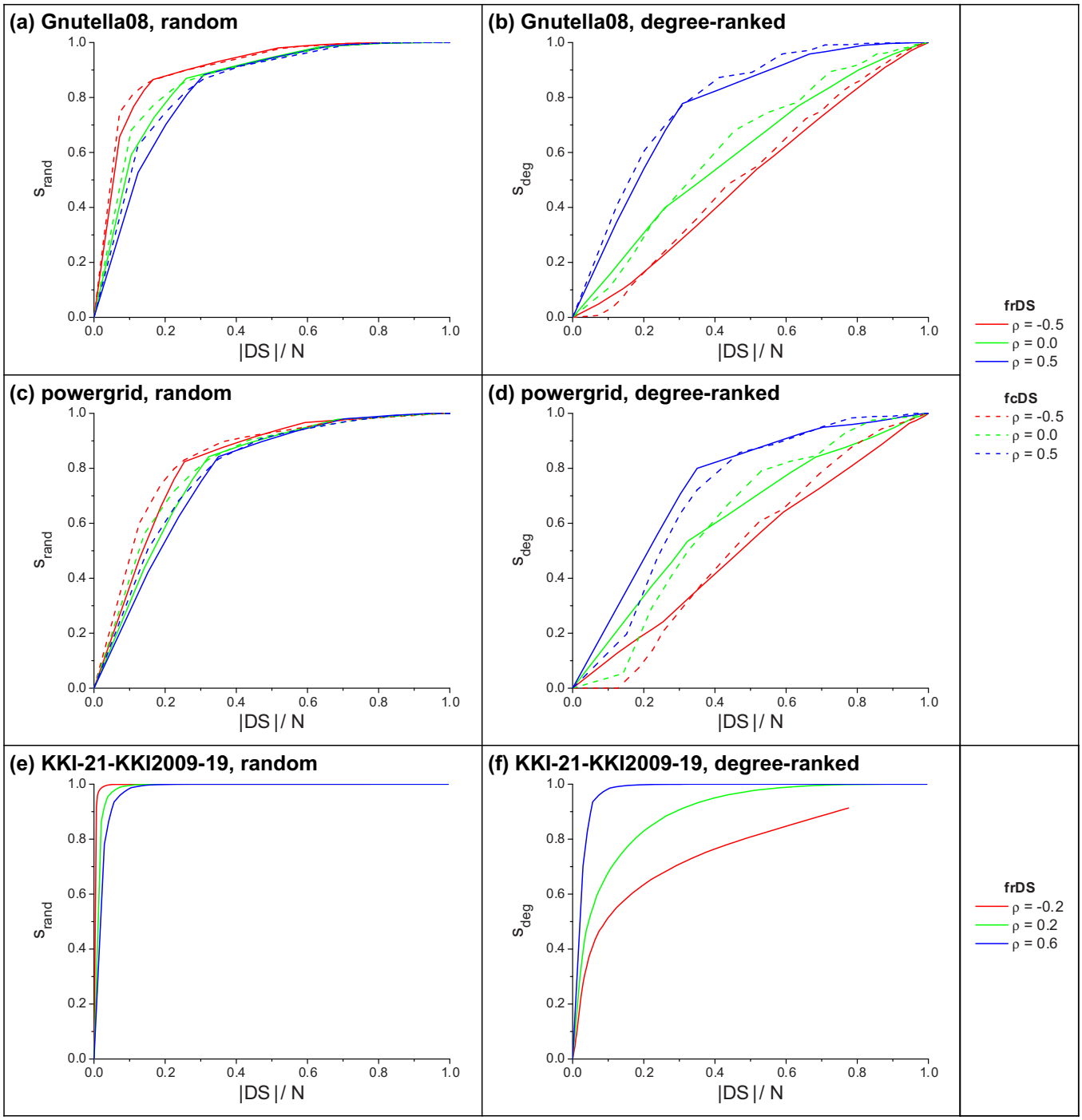

Figure 6 | Stability of frDS and fcDS in edge-mixed real networks against random and degree-ranked attacks, for various assortativity levels: (a,b) Gnutella peer-to-peer network; (c,d) ENTSO-E powergrid; (e,f) Brain (MRI) network. Network damage fraction $f = 0.3$. For (a-d) data is averaged over 50 independent runs edge mixing and node removal; (e,f) is from a single run. See Table 1 for parameters of the original networks.

behavior by considering the effects of assortativity on dominator node degrees. In disassortative networks dominators are mostly high-degree hubs, while in assortative networks dominators have a full range of degrees. Thus, when the network is disassortative and the damage is random, it is less likely to remove high-degree hubs and more likely to remove low degree nodes, the latter rarely being a dominator, leading to increased stability. On the other hand, the result is reversed when high-degree nodes are targeted, in which case we are more likely removing dominators, leading to decreased stability.

Finally, we can conjecture that the outstandingly high domination stability in brain graphs can be attributed to both their high average degree and high assortativity. High average degree results in a highly redundant dominating set (regardless of method) which resists random damage successfully, while high assortativity guarantees that an attack targeted at high degrees leaves the network with plenty of lower-degree dominators.

Partial Flexible-Redundancy Dominating Sets. There are two possible ways to achieve a certain desired cost (dominating set size) with frDS. Either we aim for the lowest r value that provides the desired cost, or we may choose a larger r value, and use only a fraction of the larger dominating set it provides. In the latter case we would select nodes in the same order as the greedy algorithm picked them. In other words, we can either select a full frDS with small r or a partial frDS with the same size but larger r. Figure 7 shows the comparison of these two cases (see Supplementary Figures S11–S16 for analysis over a wide range of network parameters). The contour curves of fixed stability values are monotonically increasing for larger r values, indicating that the cost for a certain stability level increases if we use partial frDS with higher r values. This also means that using full frDS with the smallest possible r value provides the highest possible stability.

In order to find the needed r value for a desired cost we must look at the relationship between r and the size of the resulting dominating set (see Fig. 2(a) inset, and Supplementary Figure S6). The frDS size curve has a complex shape, but it is always monotonically increasing. Therefore, we can use a bisection method for finding the desired r value. Without any assumptions (other than monotonicity) about the size of frDS we must calculate the full frDS for every tested r, each taking $O(E)$ time, leading to $O(E \log N)$ time complexity for the entire procedure.

It is also interesting to note that the cost of stability increases slightly for smaller r values when $r < 1$, in case of a random damage [in Fig. 7(a)]. In this case even the full frDS is providing only a partial dominating set (dominating only a fraction of nodes in the undamaged network). This indicates that r should never be smaller than 1; if a smaller cost is needed than the one provided by frDS with $r = 1$ (which is the MDS by definition), then a partial MDS (given by the greedy MDS algorithm) is a more optimal solution.

Effects of Incorrectly Estimated Damage in fcDS. For practical applications of fcDS, it is necessary to understand how stability is affected, when the network damage is estimated incorrectly. We can check this effect for a degree-ranked attack by using the following sigmoid strength function for a node with degree k:

$$s(k) = \frac{1}{1 + e^{\alpha(k-\kappa(\alpha,f))}}. \quad (5)$$

There are two control parameters for the anticipation. The slope parameter $\alpha \in (-\infty, \infty)$ describes the attack distribution: it expresses whether low degrees ($\alpha < 0$) or high degrees ($\alpha > 0$) are targeted, and how sharp the difference is between targeted and non-targeted node strengths; parameter f is the anticipated damage fraction. The $\kappa(\alpha, f)$ function gives the threshold for the sigmoid, such that the expected number of lost nodes equals the anticipated damage, $\sum_k (1 - s(k))p(k) = f$ (where $p(k)$ is the degree distribution). Note, that $\alpha = \infty$ gives a sharp cutoff selecting all nodes above κ, corresponding to the actual attack; $0 < \alpha \lesssim 5$ corresponds to an uncertain transition point but correct anticipation; $a \approx 0$ corresponds to a random guess; $-5 \lesssim \alpha < 0$ corresponds to an incorrect anticipation (i.e., anticipating attack on low degree nodes, when the attack occurs at high-degree nodes); and $\alpha \ll -5$ is the complete opposite of the actual attack.

Figure 8 shows the landscape of stability as a function of the control parameters. As expected, we obtain the highest stability when the attacked degrees and the size of the attack are correctly estimated. For small damage fractions ($f = 0.1$) we lose stability mostly for overestimating the size of the attack, while for moderate ($f = 0.3$) and large ($f = 0.5$) damages we lose stability for incorrectly anticipating which degrees are targeted.

Discussion

Our study of domination stability on real networks reveals the importance of average degree and assortativity in network domination. While the effect of the average degree alone is difficult to observe in real networks with unique topologies, experiments with synthetic scale free networks (Supplementary Figs. S9 and S10) show that increasing the average degree results in higher stability, simply because a node, on average, dominates more neighbors. Assortativity

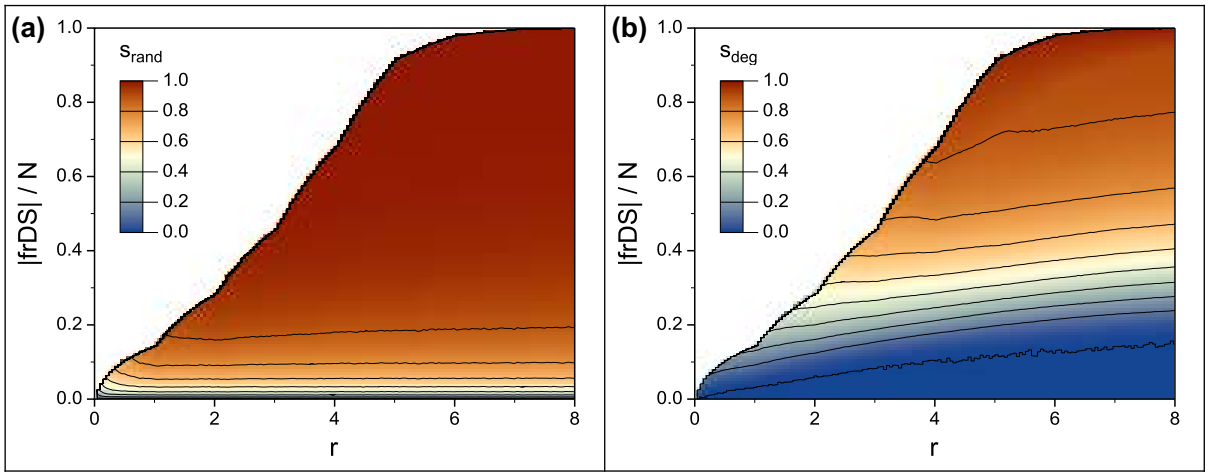

Figure 7 | Domination stability of partial frDS as a function of domination redundancy and dominating set size. The plotted area is bounded by the size of the full frDS at any given r. Subfigure (a) shows random node removal, (b) shows degree-ranked node removal, for synthetic scale-free networks, $N = 5000$, $\langle k \rangle = 8$, $\gamma = 2.5$, $f = 0.3$, averaged over 50 network samples.

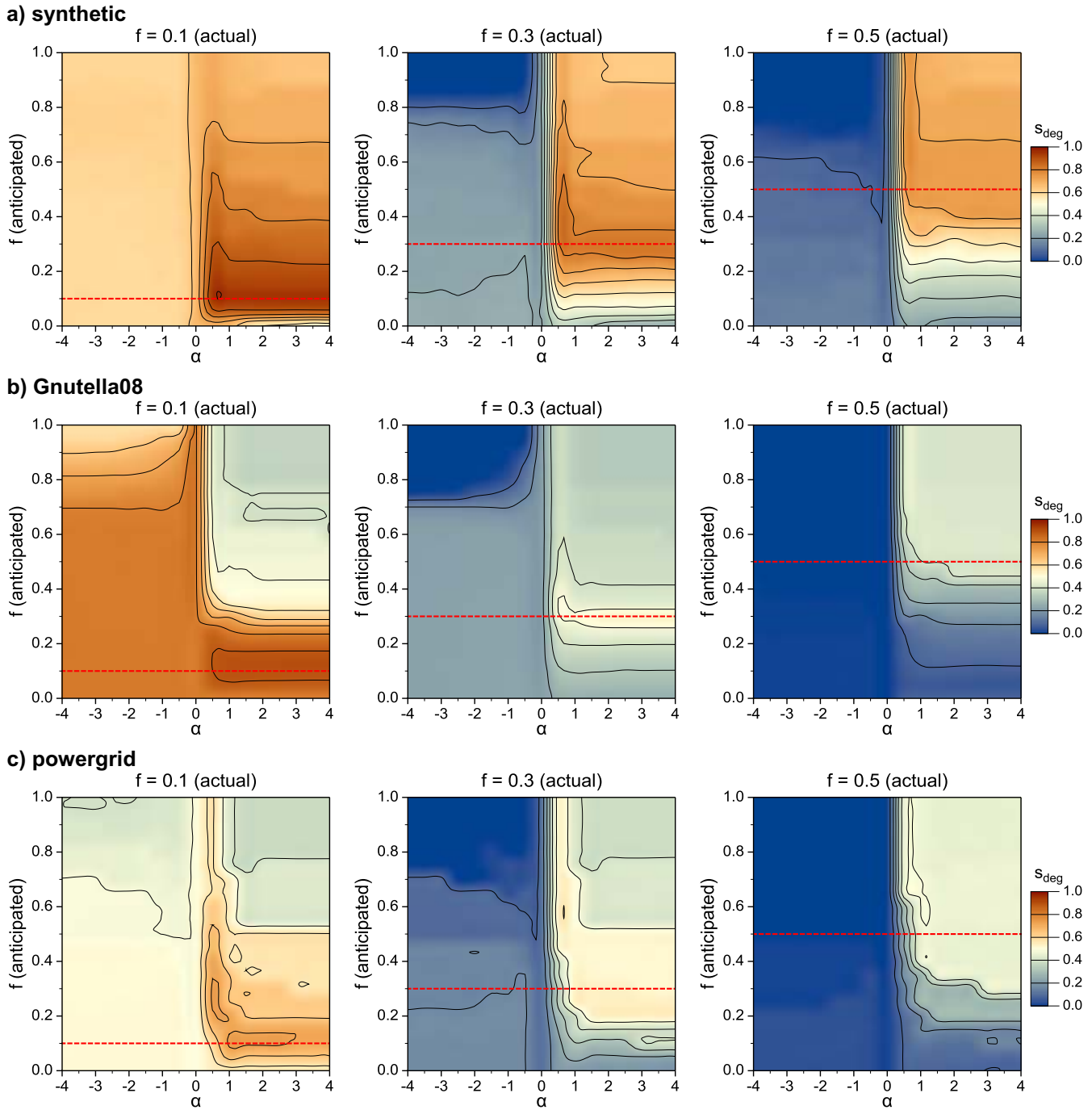

Figure 8 | Stability of fcDS against degree-ranked node removal as a function of the damage anticipation accuracy: (a) synthetic scale-free network with $N = 5000$, $\langle k \rangle = 8$, $\gamma = 2.5$; (b) Gnutella peer-to-peer network; (c) ENTSO-E powergrid. The actual damage fraction is indicated above the plots and marked by red dashed lines; the actual degree distribution of the damage corresponds to $\alpha \geq 4$ values.

has similar effects in real and in synthetic networks (increasing assortativity results in lower stability against random damage, and higher stability against a targeted attack), but the scope of these effects depends on the average degree. Assortativity has larger impact on stability against targeted attacks, while the average degree has larger impact on stability against random damage. Therefore, the degree structure of both the network and the node damage must be considered carefully when building optimal damage-resilient dominating sets.

We must clarify and make a distinction between the prescribed domination redundancy and the actual achieved domination redundancy in a network, when using frDS. The former is the one denoted by the r parameter, while the latter (i.e., the actual number of dominators in the closed neighborhood of a node) can be easily calculated for any given dominating set (not just frDS), and its average always exceeds the prescribed value. For example, even an MDS could have an actual average redundancy of 2.5 in certain networks, although most nodes would have only one dominator. However, an frDS with $r = 2.5$ would guarantee not only that the actual redundancy is at least 2.5, but also that no nodes will have less than 2 dominators.

The usage of frDS against degree-ranked or any other targeted attacks seems counter-intuitive, since in frDS, we aim for an overall increased redundancy that is most effective against random damage.

However, the greedy algorithm has no preference toward selecting low-degree or high-degree dominators when trying to fulfill domination requirements, and in general, we observe empirically that the selected dominators have a large variability in degrees. This indicates that dominators of a given node may have significantly different degrees, which helps to keep the node dominated even if high degree nodes are targeted by an attack.

In the calculation of node stability in fcDS we assumed that nodes are deleted independently. In a realistic scenario, an attack may have between-node correlations, especially, in spatial graphs (e.g., clustered attack on a power grid). Taking this into count would add more complexity to the calculations, which we postpone for future work. However, it is important to emphasize that even without correlations, the fcDS algorithm can use arbitrary node strength values, irrespective of node degrees, therefore its applicability goes much beyond our studied scenario of a degree-ranked attack.

Currently, the time complexity of fcDS is $O(d^2 \log N)$ for selecting each dominator node, which makes it prohibitive for very large graphs. In order to speed up the algorithm, the only obstacle we need to overcome is maintaining the sortedness of nodes by their potentials efficiently, which takes $O(\log N)$ steps after each change with comparative sorting. In principle, the potentials could be discretized and assigned to bins (the same optimization we use in frDS), which would lead to $O(E)$ complexity, as long as the bin count remains $O(N)$. However, the effects of such discretization on the dominating set and its stability is unclear, and it would require a thorough analysis to test the method's viability.

We can easily explain that fcDS has a slightly lower stability than frDS at low damage fractions, which we can observe in all graphs, by looking at the effects of incorrect attack anticipation. When the actual damage is very small, we overestimate the damage with our degree-dependent strength formula ($s(i) = 1 - d(i)/N$), because we assign nonzero probabilities to losing nodes with medium to low degrees. In reality, these nodes will not be deleted in a small targeted attack, thus the overestimated damage causes fcDS to lose stability, dropping slightly below the levels of frDS. This also underlines the need to correctly estimate the size as well as the distribution of the expected attack to achieve optimal domination stability.

Finally, we can provide a simple guide for selecting one of our two methods for practical applications. If we have no detailed information about a potential attack, or the network is very large, then frDS is a good choice for providing a dominating set with decent stability against any form of damage (mostly against random damage originating from natural sources), with a short computational time. However, if there is a fixed budget for dominators, or detailed (and reliable) information is available about potential attacks, then fcDS can be used to optimize the selected dominating set for the highest possible stability.

Methods

We measure domination stability as an averaged value over an ensemble of networks, using the following procedure. First, a network sample is generated, and its dominating set is calculated by one of the preselected dominating set search algorithms. Then, m nodes are removed from the network, according to a predetermined node removal strategy, where $m/N = f$ is the desired fraction removed from a network with N nodes. Finally, stability is evaluated using Eq. 1 in the remaining network.

Each node removal strategy is implemented using a sorted list of all nodes in the network; nodes are sorted such that the first m nodes will be removed. For random node removal the list of nodes is shuffled (a random permutation is computed) by the Fisher-Yates algorithm[32]. For degree-ranked node removal the nodes are sorted in decreasing order of degrees (with random tie-breaking).

We generate scale-free network samples using the configuration model[33–35]. First, a discrete power-law degree distribution is constructed for given network size N, degree exponent γ, and average degree $\langle k \rangle$. The degree sequence is then sampled from the degree distribution, and treated as a set of half-links for each node to be connected. Links are realized by randomly (uniformly) selecting any two unconnected half-links, until no more links can be formed. This may result in multiple links between some nodes, but they are treated only as single links, resulting in a small loss of total links. However, the loss is negligible, since we only focus on networks with $\gamma > 2$.

The average degree is controlled by adjusting the minimum degree cutoff $k_{\min}$ of the degree distribution, while the maximum degree cutoff $k_{\max} = \sqrt{N}$. The correct $k_{\min}$ value that yields the desired average degree for the network is obtained from a precomputed lookup table. We have used the same technique in our previous work[4] where we have shown the high level of accuracy achievable with this method. According to our previous notation in[4], the networks we use here are cCONF networks (abbreviation for configuration model with structural cutoff $k_{\max} = \sqrt{N}$).

1. Haynes, T. W., Hedetniemi, S. T. & Slater, P. J. *Fundamentals of Domination in Graphs.* New York: Marcel Dekker (1998).
2. Echenique, P., Gómez-Gardeñes, J., Moreno, Y. & Vázquez, A. Distance-*d* covering problems in scale-free networks with degree correlations. *Phys. Rev. E* **71**, 035102(R) (2005).
3. Nacher, J. C. & Akutsu, T. Dominating scale-free networks with variable scaling exponent: heterogeneous networks are not difficult to control. *New J. Phys.* **14**, 073005 (2012).
4. Molnár, F. Jr., Sreenivasan, S., Szymanski, B. K. & Korniss, G. Minimum dominating sets in scale-free network ensembles. *Sci. Rep.* **3**, 1736 (2013).
5. Kelleher, L. & Cozzens, M. Dominating Sets in Social Network Graphs. *Math. Soc. Sci.* **16**, 267–279 (1988).
6. Wang, F. *et al.* On positive influence dominating sets in social networks. *Theo. Comp. Sci.* **412**, 265–269 (2011).
7. Eubank, S., Anil Kumar, V. S., Marathe, M. V., Srinivasan, A. & Wang, N. Structural and algorithmic aspects of massive social networks. In *Proc. of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, Philadelphia: Society for Industrial and Applied Mathematics, 718–727 (2004).
8. Nacher, J. C. & Akutsu, T. Analysis on critical nodes in controlling complex networks using dominating sets. In *2013 International Conference on Signal-Image Technology & Internet-Based Systems*, New York: IEEE, 649–654 (2013).
9. Nacher, J. C. & Akutsu, T. Structural controllability of unidirectional bipartite networks. *Sci. Rep.* **3**, 1647 (2013).
10. Nacher, J. C. & Akutsu, T. Analysis of critical and redundant nodes in controlling directed and undirected complex networks using dominating sets. *J. Complex Networks* **2**, 394–412 (2014).
11. Jia, T. *et al.* Emergence of bimodality in controlling complex networks. *Nat. Commun.* **4**, 2002 (2013).
12. Yang, Y., Wang, J. & Motter, A. E. Network Observability Transitions. *Phys. Rev. Lett.* **109**, 258701 (2012).
13. Wuchty, S. Controllability in protein interaction networks. *Proc. Natl. Acad. Sci. USA* **111**, 7156–7160 (2014).
14. Nacher, J. C. & Akutsu, T. Structurally Robust Control of Complex Networks. *Phys. Rev. E* (in press, 2015); arXiv: 1410.2949 [physics. soc-ph].
15. Cooper, C., Klasing, R. & Zito, M. Lower bounds and algorithms for dominating sets in web graphs. *Internet Math.* **2**, 275–300 (2005).
16. Potluri, A. & Singh, A. Two Hybrid Meta-heuristic Approaches for Minimum Dominating Set Problem. *Lect. Notes Comput. Sc.* **7077**, 97–104 (2011).
17. Hedar, A. R. & Ismail, R. Hybrid Genetic Algorithm for Minimum Dominating Set Problem. *Lect. Notes Comput. Sc.* **6019**, 457–467 (2010).
18. Molnár, F. Jr. *et al.* Dominating Scale-Free Networks Using Generalized Probabilistic Methods. *Sci. Rep.* **4**, 6308 (2014).
19. Albert, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
20. Duch, J. & Arenas, A. Effect of random failures on traffic in complex networks. *Proc. SPIE* **6601**, 66010O (2007).
21. Gallos, L. K., Cohen, R., Argyrakis, P., Bunde, A. & Havlin, S. Stability and topology of scale-free networks under attack and defense strategies. *Phys. Rev. Lett.* **94**, 188701 (2005).
22. Holme, P. & Kim, B. J. Attack vulnerability of complex networks. *Phys. Rev. E* **65**, 056109 (2002).
23. Cohen, R., Erez, K., ben-Avraham, D. & Havlin, S. Resilience of the Internet to Random Breakdowns. *Phys. Rev. Lett.* **85**, 4626–4628 (2000).
24. Cohen, R., Erez, K., ben-Avraham, D. & Havlin, S. Breakdown of the Internet under Intentional Attack. *Phys. Rev. Lett.* **86**, 3682–3685 (2001).
25. Callaway, D. S., Newman, M. E. J., Strogatz, S. H. & Watts, D. J. Network Robustness and Fragility: Percolation on Random Graphs. *Phys. Rev. Lett.* **85**, 5468–5471 (2000).
26. Tanizawa, T. Structural robustness and transport efficiency of complex networks with degree correlation. arXiv: 1209.4897 [physics. soc-ph].
27. Paul, G., Tanizawa, T., Havlin, S. & Stanley, H. E. Optimization of robustness of complex networks. *Eur. Phys. J. B* **38**, 187–191 (2004).
28. Tanizawa, T., Paul, G., Cohen, R., Havlin, S. & Stanley, H. E. Optimization of network robustness to waves of targeted and random attacks. *Phys. Rev. E* **71**, 047101 (2005).
29. Hayashi, Y. & Miyazaki, T. Emergent rewirings for cascades on correlated networks. arXiv:cond-mat/0503615 [cond-mat. dis-nn].
30. Asztalos, A., Sreenivasan, S., Szymanski, B. K. & Korniss, G. Cascading failures in spatialy-embedded random networks. *PLoS ONE* **9**, e84563 (2014).
31. Alon, N. & Spencer, J. H. *The Probabilsitic Method. 2nd ed.* New York: Willey (2000).
32. Knuth, D. *The Art of Computer Programming 2: Seminumerical Algorithms. 3rd ed.* Boston: Addison-Wesley, 145–146 (1998).
33. Molloy, M. & Reed, B. A critical point for random graphs with a given degree sequence. *Random Struct. Algor.* **6**, 161–180 (1995).

34. Britton, T., Deijfen, M. & Martin-L of, A. Generating simple random graphs with prescribed degree distribution. *J. Stat. Phys.* **124**, 1377–1397 (2005).
35. Viger, F. & Latapy, M. Efficient and simple generation of random simple connected graphs with prescribed degree sequence. In *11th Intl. Comp. and Combin. Conf.*, Berlin: Springer, 440–449 (2005).
36. Newman, M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).
37. Newman, M. E. J. Mixing patterns in networks. *Phys. Rev. E* **67**, 026126 (2003).
38. Spearman, C. The Proof and Measurement of Association between Two Things. *Amer. J. Psychol.* **15**, 72–101 (1904).
39. Borkowf, C. B. Computing the nonnull asymptotic variance and the asymptotic relative efficiency of Spearman's rank correlation. *Comput. Stat. & Data Anal.* **39**, 271–286 (2002).
40. Litvak, N. & van der Hofstad, R. Uncovering disassortativity in large scale-free networks. *Phys. Rev. E* **87**, 022801 (2013).
41. Cooper, C., Klasing, R. & Zito, M. Lower bounds and algorithms for dominating sets in web graphs. *Internet Math.* **2**, 275–300 (2005).
42. Raz, R. & Safra, S. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proc. of the 29th Annual ACM Symposium on Theory of Computing,* New York: ACM, 475–484 (1997).
43. Klasing, R. & Laforest, C. Hardness results and approximation algorithms of k-tuple domination in graphs. *Inform. Process. Lett.* **89**, 75–83 (2004).
44. Stanford Network Analysis Project (SNAP), http://snap.stanford.edu/data, Accessed 02/12/2013.
45. Hutcheon, N. & Bialek, J. W. Updated and validated power flow model of the main continental European transmission network. In *Proc. of the IEEE PowerTech*, Grenoble, IEEE, 1–5 (2013).
46. Continental European Transmission Network (2009 winter data), http://www.powerworld.com/bialek, Accessed 01/08/2014.
47. Open Connectome Project, http://mrbrain.cs.jhu.edu/disa/download, Accessed 03/07/2014.
48. Roncal, W. G. *et al.* MIGRAINE: MRI Graph Reliability Analysis and Inference for Connectomics. In *1st IEEE Global Conf. on Signal and Info. Proc.*, Austin, TX, IEEE, 313–316 (2013).

Acknowledgments

We thank Tao Jia for valuable discussion. This work was supported in part by grant No. FA9550-12-1-0405 from the U.S. Air Force Office of Scientific Research (AFOSR) and the Defense Advanced Research Projects Agency (DARPA), by the Defense Threat Reduction Agency (DTRA) Award No. HDTRA1-09-1-0049, by the National Science Foundation (NSF) Grant No. DMR-1246958, by the Army Research Laboratory (ARL) under Cooperative Agreement Number W911NF-09-2-0053, by the Army Research Office (ARO) grant W911NF-12-1-0546, and by the Office of Naval Research (ONR) Grant No. N00014-09-1-0607. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies either expressed or implied of the Army Research Laboratory or the U.S. Government.

Author contributions

F.M., N.D., B.K.S. and G.K. designed the research; F.M. and N.D. implemented and performed numerical experiments and simulations; F.M., N.D., B.K.S. and G.K. analyzed data and discussed results; F.M., N.D., B.K.S. and G.K. wrote and reviewed the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

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

How to cite this article: Molnár, F., Derzsy, N., Szymanski, B.K. & Korniss, G. Building Damage-Resilient Dominating Sets in Complex Networks against Random and Targeted Attacks. *Sci. Rep.* 5, 8321; DOI:10.1038/srep08321 (2015).