RAIRO-Oper. Res. 51 (2017) 1055–1076
DOI: 10.1051/ro/2017008

RAIRO Operations Research
www.rairo-ro.org

THE VERTEX ATTACK TOLERANCE OF COMPLEX NETWORKS

John Matta[1], Gunes Ercal[1] and Jeffrey Borwey[1]

Abstract. The purpose of this work is four-fold: (1) We propose a new measure of network resilience in the face of targeted node attacks, *vertex attack tolerance*, represented mathematically as $\tau(G) = \min_{S\subset V} \frac{|S|}{|V-S-C_{\max}(V-S)|+1}$, and prove that for d-regular graphs $\tau(G) = \Theta(\Phi(G))$ where $\Phi(G)$ denotes conductance, yielding spectral bounds as corollaries. (2) We systematically compare $\tau(G)$ to known resilience notions, including integrity, tenacity, and toughness, and evidence the dominant applicability of τ for arbitrary degree graphs. (3) We explore the computability of τ, first by establishing the hardness of approximating unsmoothened vertex attack tolerance $\hat{\tau}(G) = \min_{S\subset V} \frac{|S|}{|V-S-C_{\max}(V-S)|}$ under various plausible computational complexity assumptions, and then by presenting empirical results on the performance of a betweenness centrality based heuristic algorithm applied not only to τ but several other hard resilience measures as well. (4) Applying our algorithm, we find that the random scale-free network model is more resilient than the Barabasi−Albert preferential attachment model, with respect to all resilience measures considered.

Mathematics Subject Classification. 68R10, 90B15, 05C50, 68Q25.

Received May 19, 2016. Accepted January 27, 2017.

1. Introduction

While networks arise ubiquitously from countless and varied disciplines, an important problem relevant to all types of networks is that of measuring and generating *resilience* characteristics. Any notion of resilience must express the relative size of a most critical set of target edges or target vertices whose removal, upon an attack, would be detrimental to the remaining network and attempt to quantify the amount of resulting damage. The graph-theoretic network resilience problem in its utmost generality is not new. But, the most well-studied theoretical machinery to discuss it, namely conductance [9] is more meaningful for networks with homogeneous degree distributions because it is a fundamentally edge-based notion. Conductance, in its various normalized and non-normalized forms, is a fundamental property of graphs independent of its applicability to edge-based resilience or resilience in general and is intimately related to both the mixing times of random walks on graphs [35] and the eigenvalues of the graph's adjacency matrix (or normalized adjacency matrix as befits the situation) [9]. The fundamental importance of conductance to the resilience of homogeneous degree networks

Keywords. Graph theory, resilience, Scale-Free networks, spectral Gap, approximation Hardness, Heuristic Algorithms.

[1] Southern Illinois University, Edwardsville, Illinois, USA. gercal@siue.edu

Article published by EDP Sciences

is well illustrated by the literature on *regular expanders* [3] which are defined in terms of high conductance and spectral gap and have become synonymous with extremally resilient families of regular graphs.

However, there is a growing body of work [31] indicating that many real world networks, ranging from biological networks to online social networks, often exhibit heterogeneous and scale-free degree distributions. Even in situations where the scale-free nature of a degree distribution may be questioned [12], an assumption of degree homogeneity would be even more suspect in many networks [2]. Some notable exceptions to this are, of course, networks whose connection are purely defined by distances in low dimensional space (such as random ad-hoc networks) as well as networks that have very hard and frugal constraints on the number of connections their nodes are allowed. In any case, there is a gap between much of the beautiful existing graph theoretical machinery to analyze resilience in networks based on conductance and the actual resilience, particularly against *node* attacks in many actual networks, which have heterogeneous degrees. To illustrate why degree heterogeneity poses such a problem to the application of any edge-based resilience notion in general and conductance in particular, consider the case of the *Star* network of n nodes, an undirected graph $G = (V, E)$ on $n = |V|$ nodes with a unique designated central node v_0 such that node v_0 is connected to every other node $v \in V - \{v_0\}$, and there are no other edges between any other pair of nodes. The conductance of $Star(n)$ is maximally 1, whereas the star is in fact the *least* resilient network with respect to targeted *node* attacks as the removal of a single node v_0 results in $n - 1$ isolated nodes. A motivation for this work is the search for a measure that accurately captures resilience against targeted node attacks in both homogeneous degree and heterogeneous degree networks. In particular, such a measure should behave similarly to conductance for regular networks, particularly assigning expander high resilience, but should diverge from conductance with respect to heterogeneous degree networks, particularly assigning the star network the lowest possible resilience. The other primary motivation for this work is answering *what is the actual comparative resilience of different generative scale-free models?*

This work consists of four logical modules. In the first module, we propose our resilience measure, called *vertex attack tolerance* (shortened VAT) denoted mathematically by $\tau(G) = \min_{S \subset V} \frac{|S|}{|V - S - C_{\max}(V-S)|+1}$, where $C_{\max}(V - S)$ is the largest connected component in $V - S$, and we establish a rigorous relationship between VAT and conductance and spectral gap in the case of *regular graphs*. In particular, we prove that $\frac{1}{d}\Phi(G) \leq \tau(G) \leq \mathrm{d}\Phi(G)$ if $\Phi(G) \leq \frac{1}{d^2}$, and $\frac{1}{d}\Phi(G) \leq \tau(G) \leq d^2\Phi(G)$ otherwise. Moreover, applying Cheeger's inequality, we obtain the following eigenvalue bounds as a corollaries: given a connected d-regular graph $G = (V, E)$, let λ_2 denote the second largest eigenvalue of G's normalized adjacency matrix. Then, $\frac{\tau(G)^2}{2d^4} \leq 1 - \lambda_2 \leq 2\mathrm{d}\tau(G)$. Moreover, if $\Phi(G) \leq \frac{1}{d^2}$, then $\frac{\tau(G)^2}{2d^2} \leq 1 - \lambda_2 \leq 2\mathrm{d}\tau(G)$.

In our second module, we are concerned with systematic comparison of VAT with several existing resilience measures in addition to conductance [9, 35] for general graphs: vertex expansion [18, 26], integrity [5], toughness [10], tenacity [27] and scattering number [24]. We exhibit sequences of graph families for which VAT uniquely captures the inuitive relative ranking of resilience. Our initial comparisons are based on the star, barbell (two identical cliques joined by a single edge), and apple (a clique adjacent to a single node by a single edge) family of graphs, for which we desire a measure respecting the following resilience ordering based on the component size distribution resulting from any frugal attack: Resilience(Star) < Resilience(Barbell) < Resilience(Apple). We perform further experimental comparisons on several other graphs, including two graphs generated *via* random expander families, and present results of our comparisons.

In the third module of our paper we explore computational issues regarding VAT and other resilience measures. We first establish the approximation-hardness of unsmoothened vertex attack tolerance $\hat{\tau}(G) = \min_{S \subset V} \frac{|S|}{|V - S - C_{\max}(V-S)|}$ under four separate, plausible computational complexity assumptions. We then present a greedy, betweenness-centrality (Greedy-BC) based heuristic algorithm to compute τ that is generalizable to other node-based resilience measures considered. Existing NP-Hardness results for resilience measures considered in this work include [13] for integrity, [6] for toughness, [28] for tenacity, [8] for scattering number, [34] for conductance, and [26] for vertex expansion. We perform various empirical comparisons demonstrating the effectiveness of Greedy-BC in computing VAT, integrity, tenacity, and toughness for realistic graph classes despite known hardness results for general graphs. Our experiments also show that hill-climbing with a neighborhood defined by Hamming distance up to two does not significantly improve accuracy despite being significantly more inefficient.

The fourth logical module of our paper consists of extensive calculations of $\tau(G)$ and other resilience measures for two important generative models of scale-free graphs: preferential attachment (PA) graphs captured *via* the Barabasi−Albert model [4], and random scale-free graphs captured *via* the PLOD model [33]. We have also computed the vertex attack tolerance of *heuristically-optimized trade-offs* graphs referred to as the HOTNet model [17]. HOTNets, which exhibited the worst vertex attack tolerance on various parameter settings considered even when controlling for average degree. However, we wish to control for degree distribution in our comparisons, and HOTNets cannot be tweaked to follow an exact degree distribution. We present extensive simulation results comparing PA graphs to PLODs of identical degree distribution. For small graph sizes, PA graphs exhibit better resilience than both PLOD and HOTNet models. However, it is the asymptotic behavior of a graph family that is statistically relevant. Our results indicate that PLOD graphs with degree distributions identical to PA graphs of the same size exhibit increasingly better resilience than the PA type graphs, although both graph types appear surprisingly resilient when the generative PA parameter is $m = 2$. Our results may be compared and contrasted with previous claims on the attack tolerance of scale-free networks [1, 2, 12].

Prior to proceeding, we note that this work contains partial results from the authors' previous conference paper [16], preprints [14, 15, 30], and thesis [29]. In particular, VAT was first proposed by the authors in conference paper [16]. Parts of the section comparing resilience measures are also contained in author thesis [29]. Spectral bounds for VAT and relations with conductance are also contained in preprints [14, 30]. Computational complexity results are also in [15]. Preliminary results on the vertex attack tolerance of scale-free models are found in preprint [30], though the algorithm used there is less accurate and efficient.

2. Definitions and preliminaries

2.1. Definitions

In all following definitions, when given a directed graph $G = (V, E)$ and considering a resilience measure r on G, the *connected components* of G refer to the *strongly connected components* of G. If weakly connected components are desirable with respect to an application, then one may define the resilience upon the underlying undirected graph G' of G instead.

Given a graph $G = (V, E)$, we define vertex attack tolerance (VAT) of G as

$$\tau(G) = \min_{S\subset V, S\neq\emptyset} \left\{ \frac{|S|}{|V - S - C_{\max}(V - S)| + 1} \right\}$$

where $C_{\max}(A)$ is defined as the largest connected component of $A \subset V$. Note that τ is a smoothened version of the measure of vertex attack tolerance previously introduced by the authors in [16]. It will also be convenient to refer to this unsmoothened version of vertex attack tolerance, which we denote as follows:

$$\hat{\tau}(G) = \min_{S\subset V, S\neq\emptyset} \left\{ \frac{|S|}{|V - S - C_{\max}(V - S)|} \right\}$$

Combinatorial conductance or edge based conductance [9, 35] is defined as

$$\begin{aligned} \Phi(G) &= \min\nolimits_{S\subset V, \mathrm{Vol}(S)\leq \mathrm{Vol}(V)/2} \left\{ \frac{|\mathrm{Cut}(S, V - S)|}{\mathrm{Vol}(S)} \right\} \\ &= \min\nolimits_{S\subset V, \mathrm{Vol}(S)\leq \mathrm{Vol}(V)/2} \left\{ \frac{|\mathrm{Cut}(S, V - S)|}{\delta_S |S|} \right\} \end{aligned}$$

where $|\mathrm{Cut}(S, V-S)|$ is the size of the cut separating S from $V-S$, $\mathrm{Vol}(S)$ is the sum of the degrees of vertices in S, and δ_S is the *average* degree of vertices in S. Conductance is a normalized form of edge expansion, which is defined as

$$\epsilon^E(G) = \min_{S\subset V, |S|\leq |V|/2} \left\{ \frac{|\mathrm{Cut}(S, V-S)|}{|S|} \right\}$$

On the other hand, vertex expansion [18, 26] is defined as

$$\epsilon^V(G) = \min_{S\subset V} \left\{ n\frac{|N(S)|}{|S||V-S|} \right\}$$

where $N(S)$ denotes the outer boundary of S, namely $N(S) = \{i \in V - S \mid \exists u \in S, _\ni \{u, v\} \in E\}$. Integrity [5] is defined as

$$\hat{I}(G) = \min_{S\subset V} \{|S| + C_{\max}(V-S)\}$$

However, to compare with other measures, we will also use the normalized version of integrity. Normalized integrity is defined as

$$I(G) = \min_{S\subset V} \left\{ \frac{|S| + C_{\max}(V-S)}{|V|} \right\}$$

Toughness [10] is defined as

$$t(G) = \min_{S\subset V} \left\{ \frac{|S|}{\omega(V-S)} \right\}$$

where $\omega(V-S)$ is the number of connected components in V-S. Tenacity [27] is defined as

$$T(G) = \min_{S\subset V} \left\{ \frac{|S| + C_{\max}(V-S)}{\omega(V-S)} \right\}$$

Scattering number [24] is defined as

$$sn(G) = \max_{S\subset V} \{\omega(V-S) - |S|\} \tag{2.1}$$

However, as higher scattering numbers corresponds to worse resilience, it will be more convenient to discuss inverse scattering, which we denote

$$h(G) = \frac{1}{sn(G)} = \min_{S\subset V} \left\{ \frac{1}{\omega(V-S) - |S|} \right\}$$

It is clear that all above notions are defined based on finding some *worst case* target set of nodes. Therefore, it will be convenient to directly denote the referenced target set as well as the measure defined conditional upon a particular target set. For the case of vertex attack tolerance, we denote set-vertex tolerance as

$$\tau_S(G) = \frac{|S|}{|V - S - C_{\max}(V-S)| + 1}$$

so that clearly $\tau(G) = \min_{S\subset V} \tau_S(G)$ and correspondingly

$$S(\tau(G)) = \mathrm{argmin}_{S\subset V} \tau_S(G)$$

Similarly for set-conductance:

$$\Phi_S(G) = \frac{|\mathrm{Cut}(S, V-S)|}{\delta_S |S|}$$

so that clearly $\Phi(G) = \min_{S\subset V,|\mathrm{Vol}(S)|\leq|\mathrm{Vol}(V)|/2} \Phi_S(G)$ and correspondingly

$$S(\Phi(G)) = \mathrm{argmin}_{S\subset V,|\mathrm{Vol}(S)|\leq|\mathrm{Vol}(V)|/2}\tau_S(G)$$

The pattern is clear for set edge expansion, set vertex expansion, set integrity, set toughness, set tenacity, and smoothed inverse scattering number, respectively denoted $I_S(G), t_S(G), T_S(G)$, and $h_S(G)$. For all such set resilience measures f_S, we may generally denote the target set function as $S(f(G)) = \mathrm{argmin} f_S(G)$.

Having just defined set-vertex attack tolerance and set-resilience in general, we may define *vertex attack tolerance in expectation*, with respect to given probability distribution function P over all possible attack sets $S \subset V$:

$$\tau^P(G) = \sum_{S\subset V} \tau_S(G)P(S)$$

For any resilience measure $r(G)$ (*e.g.* $r(G) = I(G)$ in the case of integrity), similarly define for any probability distribution P over subsets of vertices:

$$r^P(G) = \sum_{S\subset V} r_S(G)P(S)$$

It will be convenient to denote the subgraph of a graph $G = (V, E)$ that is induced by a vertex set $S \subset V$ as $G_S = (S, E_S)$.

The *normalized adjacency matrix* of a graph $G = (V, E)$ is the $|V|$ by $|V|$ matrix A where $A_{u,v} = 0$ if $\{u, v\} \notin E$ and $A_{u,v} = \frac{1}{d_u}$ where d_u is the degree of u otherwise if $\{u, v\} \in E$. Note that the normalized adjacency matrix of a graph is identical to the probability transition matrix, or Markov chain, of the natural random walk.

There are some special infinite families of graphs which we shall refer to in comparing resilience notions. The *star graphs* are denoted $Star(n)$ and defined as follows:

Definition 2.1. The *star graph* $Star(n)$ is an undirected graph $G = (V, E)$ on $n = |V|$ nodes with a unique designated central node q such that node q is connected to every other node $u \in V - \{q\}$, and there are no other edges between any other pair of nodes. Namely, $E = \{\{q, u\} \mid u \in V - \{q\}\}$. When the labeling of the graph is arbitrary, we may without loss of generality assume that $q = 1$.

The *barbell graph* denoted $Barbell(2n)$ are defined as:

Definition 2.2. The *barbell graph* $Barbell(2n)$ is an undirected graph $G = (V, E)$ on $2n = |V|$ nodes created by joining two K_n cliques by a single edge, which we may refer to as the central edge. We may also sometimes refer to the two cliques as the *left clique* and the *right clique*.

The *apple graph* denoted $\mathrm{Apple}(n)$ is defined as:

Definition 2.3. The *apple graph* $\mathrm{Apple}(n)$ is an undirected graph $G = (V, E)$ on $n = |V|$ nodes created by joining a K_{n-1} clique to a single designated node q by a single edge. When the labeling of the graph is arbitrary, we may without loss of generality assume that $q = 1$.

2.2. Generalizations and variations of VAT

A generalization of vertex attack tolerance which allows us to discuss a type of vertex-weighted graph G which has costs $c(x)$ and values $v(x)$ associated with each vertex $x \in V$. Specifically, the graph context concerned here is that of graphs $G = (V, E, c, v)$ such that c, v are positive real-valued functions on the vertex set V. If c is specified without v being specified, then we will assume that $c = v$, and similarly for the case that v is specified without c being specified. If neither c nor v are specified, then the assumption is that $c(x) = v(x) = 1$ for all

$x \in V$. Having clarified the context of such cost-value node-weighted graphs G, the following is the appropriate generalization of VAT:

$$\tau(G) = \min_{S \subset V, S \neq \emptyset} \left\{ \frac{\sum_{x \in S} c(x)}{1 + \sum_{y \in V} v(y) - \sum_{y \in S + C_{\max}(V-S)} v(y)} \right\}$$

A parametrized generalization of vertex attack tolerance which allows for linear weighting parameters α, β is the following, which we call (α, β)-vertex attack tolerance:

$$\begin{aligned}\tau(G, \alpha, \beta) &= \min_{S \subset V, S \neq \emptyset} \left\{ \frac{\alpha|S| + \beta}{|V - S - C_{\max}(V-S)| + 1} \right\} \\ &= \min_{S \subset V, S \neq \emptyset} \left\{ \frac{\beta + \alpha \sum_{x \in S} c(x)}{1 + \sum_{y \in V} v(y) - \sum_{y \in S + C_{\max}(V-S)} v(y)} \right\}\end{aligned}$$

Clearly VAT is identical to $(1,0)$-VAT. There exist situations in which $(1,1)$-VAT, which nominally appears quite similar to VAT, may be more appropriate [29]. The relative appropriateness of a weighted measure depends upon the relative frugality one wishes to assign to vertex attacks.

In our last variation of vertex attack tolerance, we consider that there may be applications in which we wish to assign a non-zero resilience to an already disconnected network by considering the resilience of each individual component. Recall that otherwise, the resilience of a disconnected graph G is zero, *for all measures considered*. Therefore, we may define the vector $\boldsymbol{\tau}(G)$ as follows. Let $G = (V, E)$ be a graph with k strongly connected components $C_1, C_2, \ldots C_k$ such that $|C_1| \geq |C_2| \geq \ldots |C_{k-1}| \geq |C_k|$. Then, letting G_i denote the subgraph of G induced by component C_i, we may define the vector $\boldsymbol{\tau}(G)$ as follows:

$$\boldsymbol{\tau}(G) = \langle \tau(G_1), \tau(G_2), \ldots, \tau(G_k) \rangle$$

Note that this variation naturally extends to all other resilience measures considered in this work as well, so that one may define vectors $\boldsymbol{I}(G), \boldsymbol{t}(G), \boldsymbol{T}(G), \boldsymbol{h}(G)$ similarly when given disconnected G and non-zero resilience is desired.

2.3. Mathematical preliminaries

Conductance is arguably the most important and well-studied resilience notion in the context of d-regular connected graphs. Part of the reason for the importance of conductance is in the intimate relationship it shares with spectral gap and mixing time. One of the most important such relationships is *Cheeger's inequality* [9,35]:

Theorem 2.4. *Given a connected d-regular, undirected graph $G = (V, E)$, let λ_2 denote the second largest eigenvalue of G's normalized adjacency matrix. Then,*

$$\frac{\Phi(G)^2}{2} \leq 1 - \lambda_2 \leq 2\Phi(G) \tag{2.2}$$

Let us note the following preliminary observation:

Remark 2.5. For undirected, connected $G = (V, E)$ with $|V| \geq 2$, $0 < \tau(G) \leq 1$, and $C_{\max}(V - S(\tau(G))) \neq \emptyset$.

The first bound follows from the non-emptiness of S by definition of τ, in addition to the fact that, for any vertex $v \in V$, $\tau(G) \leq \tau_{\{v\}}(G) = \frac{1}{|V - \{v\} - C_{\max}| + 1} \leq 1$. The non-emptiness of the largest remaining connected component follows from the fact that the only way that $C_{\max}$ can be empty is by taking $S = V$, but such S cannot achieve as low a set-VAT as that achieved by a single node, and therefore cannot be the set corresponding to VAT.

Moreover,

Lemma 2.6. *For any connected, undirected graph $G = (V, E)$ on $n = |V| \geq 3$ nodes, $\tau(G) \geq \frac{1}{n-1}$*

Proof of Lemma 2.6. Because G is connected, at least one node must be attacked in order to result in any disconnection, so that the minimum achievable value of the numerator of the τ function is 1. Considering the maximum possible denominator of τ, namely the maximum achievable $|V - S - C_{\max}| + 1$, we claim that it is $n - 1$. Because we already argued that $|S| \geq 1$, it suffices to show that $C_{\max} \neq 0$ for this situation, as $n - 1 - 1 + 1 = n - 1$ then yields the desired lower bound. But, $C_{\max}$ cannot be zero unless $V = S$ (or else some node would remain), and taking $V = S$ gives a VAT of $1 >> \frac{1}{n-1}$ for $n \geq 3$. □

The following bound is well-known for conductance:

Remark 2.7. For undirected, connected $G = (V, E)$ with $|V| \geq 2$, $0 < \Phi(G) \leq 1$.

For the proofs in the next section, the following inequality will prove useful:

$$\forall a, b, x, y > 0,\ \frac{a}{x} < \frac{b}{y}\ \rightarrow\ \frac{a}{x} < \frac{a+b}{x+y} < \frac{b}{y} \tag{2.3}$$

Even more useful is a corollary of this inequality that follows by induction:

Corollary 2.8. *Let $n > 0$ be a natural number, and for each natural number i from 1 to n, let positive numbers a_i and b_i be given. Moreover, let c be any real number that satisfies $c \leq \min_{1 \leq i \leq n} \frac{a_i}{b_i}$. Then, the following is true:*

$$c \leq \frac{\sum_{1 \leq i \leq n} a_i}{\sum_{1 \leq i \leq n} b_i} \tag{2.4}$$

3. Spectral bounds for VAT in regular graphs

As stated, conductance is a very important and well-studied resilience notion in the context of d-regular connected graphs. The two main results of this section are the following theorems which state that the vertex attack tolerance and conductance of the same graph are within a factor d of each other as long as the conductance is not too high and the unsmoothened VAT exists. The first theorem upper bounds VAT *via* conductance, and the second theorem lower bounds the unsmoothened version of VAT *via* conductance:

Theorem 3.1. *For undirected, connected, d-regular $G = (V, E)$ with $|V| \geq 2$, if $\Phi(G) \leq \frac{1}{d^2}$ then $\tau(G) < \mathrm{d}\Phi(G)$, and $\tau(G) < d^2\Phi(G)$ otherwise.*

Theorem 3.2. *For undirected, connected, d-regular $G = (V, E)$ with $|V| \geq 2$, $\Phi(G) < \mathrm{d}\tau(G)$.*

Applying Cheeger's inequality to these theorems results in the following corollary:

Corollary 3.3. *Given a connected d-regular graph $G = (V, E)$, let λ_2 denote the second largest eigenvalue of G's normalized adjacency matrix. Then,*

$$\frac{\tau(G)^2}{2d^4} \leq 1 - \lambda_2 \leq 2\mathrm{d}\tau(G) \tag{3.1}$$

Furthermore, if $\Phi(G) \leq \frac{1}{d^2}$, then

$$\frac{\tau(G)^2}{2d^2} \leq 1 - \lambda_2 \leq 2\mathrm{d}\tau(G) \tag{3.2}$$

In proving Theorem 3.1 we shall use the following lemma regarding conductance in d-regular graphs:

Lemma 3.4. *Given a connected, undirected d-regular graph $G=(V,E)$, there exists a set S such that $S = S(\Phi(G))$ and the induced subgraph G_S is connected.*

Now we demonstrate the proofs of the theorems and lemma:

Proof of Theorem 3.1. First note that the bound for the case that $\Phi(G) > \frac{1}{d^2}$ simply follows from both conductance and VAT being normalized measures in $(0,1]$. So, WLOG assume that $\Phi(G) \leq frac1d^2$. Let set $S = S(\Phi(G))$ such that the induced subgraph G_S is connected, realizing the existence of such non-empty S *via* Lemma 3.4. Let S^{out} denote the vertex boundary of S that is outside of S, also called the *outer vertex boundary of* S and precisely being $S^{\text{out}} = \{v \in V - S \mid \exists e = \{u,v\} \in \text{Cut}(S, V-S)\}$. We may lower bound and upper bound S^{out} as follows:

(1) $|\text{Cut}(S, V-S)| \leq d|S^{\text{out}}|$.
(2) $|S^{\text{out}}| \leq |\text{Cut}(S, V-S)|$.

Combining bound (2) above with the fact that $\frac{|\text{Cut}(S,V-S)|}{d|S|} = \Phi(G) \leq \frac{1}{d^2}$ we have that

$$|S^{\text{out}}| \leq |\text{Cut}(S, V-S)| \leq \frac{|S|}{d}$$

Now, consider the structure of G upon the removal of nodes S^{out}. As $S^{\text{out}} \subset V - S$, and removal of S^{out} also removes edges along $\text{Cut}(S, V-S)$, we know that at least S would remain as a connected component. There are two possible situations: either (i) S would be the largest connected component in which case $|V - S^{\text{out}} - C_{\max}(V - S^{\text{out}})| = |V - S - S^{\text{out}}|$ or, (ii) S is not the largest connected component in which case $|V - S^{\text{out}} - C_{\max}(V - S^{\text{out}})| \geq |S|$.

First consider case (i): by definition of Φ, S cannot contain a strict majority of nodes of V, and therefore, $|V - S| \geq |S|$ is known. Combining this with $|S^{\text{out}}| \leq |S|\frac{1}{d}$ implies that

$$|V - S^{\text{out}} - C_{\max}(V - S^{\text{out}})| = |V - S - S^{\text{out}}| \geq |S|\frac{d-1}{d}$$

Further combining this with bound (2) above, we have that

$$\begin{aligned}\tau_{S^{\text{out}}} &= \frac{|S^{\text{out}}|}{|V - S^{\text{out}} - C_{\max}(V - S^{\text{out}})| + 1} \\ &\leq \frac{(d-1)|\text{Cut}(S, V-S)|}{d|S| + 1} \\ &< (d-1)\Phi(G) < \text{d}\Phi(G)\end{aligned}$$

finishing the proof for case (i), because $\tau(G) \leq \tau_{S^{\text{out}}} < \text{d}\Phi(G)$.

In the second case (ii), we also obtain that $\tau_{S^{\text{out}}}(G) \leq \text{d}\Phi(G)$: due to the previously established facts $|S^{\text{out}}| \leq |\text{Cut}(S, V-S)|$ and $|V - S^{\text{out}} - C_{\max}(V - S^{\text{out}})| \geq |S|$ we have

$$\begin{aligned}\tau_{S^{\text{out}}} &= \frac{|S^{\text{out}}|}{|V - S^{\text{out}} - C_{\max}(V - S^{\text{out}})| + 1} \\ &< \frac{|\text{Cut}(S, V-S)|}{|S|} \\ &= \text{d}\Phi(G)\end{aligned}$$

finishing the proof. □

Proof of Theorem 3.2. Let $S = S(\tau(G))$, which is non-empty by definition. Furthermore, let $T = C_{\max}(V - S)$, which is also non-empty by Remark 2.5. Consider the set conductance of T, namely $\Phi_T(G)$. There are two situations which we must consider separately:

(1) $|T| > \frac{|V|}{2}$.

(2) $|T| \leq \frac{|V|}{2}$.

Let us first consider the first situation. In that case, $\Phi_T(G) = \frac{|\mathrm{Cut}(T,V-T)|}{d|V-T|}$. However, because all edges of $\mathrm{Cut}(T, V - T)$ must be adjacent to S, we know that $|\mathrm{Cut}(T, V - T)| \leq d|S|$. Moreover, clearly, $|V - T| \geq |V - S - T + 1|$. Combining these facts, we obtain that:

$$\Phi(G) \leq \Phi_T(G) \leq \frac{d|S|}{|V - S - T + 1|} = \mathrm{d}\tau(G) \tag{3.3}$$

This finishes the proof for the case that T is the majority of the nodes. Therefore, let us finally consider the case that T does not comprise the majority of the vertices:

Let us denote q as the number of connected components of the induced subgraph G_{V-S-T}, and, furthermore, denote each such remaining component by $C_i, \forall 1 \leq i \leq q$. Because T itself is the *largest* connected component of induced subgraph G_{V-S} by definition, and $T < \frac{|V|}{2}$ by assumption, we know further that each $|C_i| < \frac{|V|}{2}$. It will be convenient to denote $C_{q+1} = T$. Therefore,

$$\forall 1 \leq i \leq q + 1, \Phi_{C_i}(G) = \frac{|\mathrm{Cut}(C_i, V - C_i)|}{d|C_i|} \tag{3.4}$$

Now, note the following facts due to the action of S and the definition of connected components:

(1) $d|S| \geq |\mathrm{Cut}(T, V - T)| + \sum_{i=1}^{q} |\mathrm{Cut}(C_i, V - C_i)| = \sum_{i=1}^{q+1} |\mathrm{Cut}(C_i, V - C_i)|$.

(2) $|V - S - T + 1| \leq |V - S| = |T| + \sum_{i=1}^{q} |C_i| = \sum_{i=1}^{q+1} |C_i|$.

The second fact is clear from the definition of connected components as a partition. The first fact follows from the action of S: all edges crossing the boundary of a connected component must be adjacent to S. Combining the two facts, we obtain that

$$\tau(G) = d\frac{|S|}{d|V - S - T + 1|} \geq \frac{\sum_{i=1}^{q+1} |\mathrm{Cut}(C_i, V - C_i)|}{\sum_{i=1}^{q+1} d|C_i|} \tag{3.5}$$

Now, note that each term of the sum in the numerator is a numerator of $\Phi_{C_i}(G)$ whereas each term of the sum in the denominator is a corresponding denominator of $\Phi_{C_i}(G)$. Applying Corollary 2.8 with $c = \Phi(G)$, with $a_i = |\mathrm{Cut}(C_i, V - C_i)|$, and with $b_i = d|C_i|$, we obtain that $\tau(G) \geq \Phi(G)$, completing the proof. □

Proof of Lemma 3.4. Let $S = S(\Phi(G))$, and assume that G_S can be partitioned into two subgraphs that are disconnected from each other, namely $G_1 = (S_1, E_1)$ and $G_2 = (S_2, E_2)$ such that $S_2 = S - S_1$ and $E_2 = E_S - E_1$. Therefore, the edge boundary or S is correspondingly partitioned such that

$$|\mathrm{Cut}(S, V - S)| = |\mathrm{Cut}(S_1, V - S)| + |\mathrm{Cut}(S_2, V - S)|$$

For convenience, denote $a = |\mathrm{Cut}(S_1, V - S)| = |\mathrm{Cut}(S_1, V - S_1)|$, $b = |\mathrm{Cut}(S_2, V - S)| = |\mathrm{Cut}(S_2, V - S_2)|$, and $c = |\mathrm{Cut}(S, V - S)| = a + b,$. Furthermore, denote $x = |S_1|$, $y = |S_2|$, and $z = |S| = x + y$. Now, consider the relative set conductances of S_1 and S_2, and without loss of generality, assume that $\Phi_{S_1}(G) \leq \Phi_{S_2}(G)$. Then, because the normalization factor d is lost from both sides we have:

$$\frac{a}{x} \leq \frac{b}{y}$$

But, due to the minimality of the conductance achieved by S, we have

$$\frac{a+b}{x+y} \leq \frac{a}{x} \leq \frac{b}{y}$$

But, by equation 2.3, this can only be satisfied when

$$\frac{a+b}{x+y} = \frac{a}{x} = \frac{b}{y}$$

And, in that case, S_1 or S_2 also achieve the conductance of G. However, we did not assume that either of them are connected. Nonetheless, it is clear that as long as one of them is disconnected, the same argument as above can be applied to get another partition into two smaller sets yet which each achieve the conductance Φ. However, due to the finiteness of G, this cannot continue ad infinitum, and there must exist a set S_i such that $\Phi_G = \Phi_{S_i}(G)$ and induced subgraph G_{S_i} is connected. □

4. Comparison of resilience notions for general graphs

In contrast to the situation of regular graphs, here we first consider the extent to which conductance fails to capture node-based resilience in the case of *heterogeneous degree* graphs. We note the following immediately: there exists an infinite graph family, namely the star graphs Star(n) such that a targeted attack against the central node results in a maximal disconnection of the graph into $n-1$ isolated nodes. Yet, the following are the conductance and vertex attack tolerance of Star(n) respectively, for $n \geq 3$:

(1) $\Phi(\text{Star}(n)) = 1\cdot$
(2) $\tau(\text{Star}(n)) = \dfrac{1}{n-1}\cdot$

In other words, there is an infinite graph family which is maximally intolerant against a targeted node attack. The conductance of this family is maximal, whereas the vertex attack tolerance of this family is minimal, amongst all possible graphs. We prove this introductory observation as follows: let q be the designated central node in Star(n). First, let us bound $\Phi(\text{Star}(n))$: let $S = S(\Phi(\text{Star}(n)))$. First consider the case that $q \ni S$. In that case, all degrees of nodes in S are one, and each node in S is adjacent only to q which is not in S. Therefore, $|\text{Cut}(S, V-S)| = |S| = |\text{Vol}(S)|$, proving that $\Phi(\text{Star}(n)) = 1$ for this case. In the other case that $q \ni S$, note that the volume constraint in the definition of conductance restricts us to sets S that do not exceed half of the total volume of V. The volume of q alone is proportional to its degree $n-1$, whereas there are exactly $n-1$ other nodes whose total volume is exactly $n-1$. Therefore, in the case that $q \in S$, it must be that $S = \{q\}$ alone. In this situation, $\text{Cut}(S, V-S)$ is exactly all the $n-1$ edges of G, and so calculating for conductance again yields $\Phi_S(\text{Star}(n)) = \frac{n-1}{n-1} = 1$. Now we consider $\tau(\text{Star}(n))$. As intuitively clear, first consider $\tau_S(\text{Star}(n))$ for $S = \{q\}$, namely a targeted attack of the central node. Because removal of q results in only isolated nodes, $|C_{\max}| = 1$. Moreover, $|S| = 1$. Therefore, $\tau_S(\text{Star}(n)) = \frac{1}{n-1}$. By Lemma 2.6, no other set can achieve a lower VAT value, and so $\tau(\text{Star}(n)) = \frac{1}{n-1}$.

Having established that the minimally resilient graph with respect to targeted vertex attacks indeed achieves minimal VAT as well while achieving maximal conductance, we wish to further compare both against other resilience notions and against the barbell and apple families as well. We explicitly axiomatize the following in our comparisons:

Remark 4.1. Any measure of network resilience Resilience(G) which is based upon an attacker model that targets a *critical set of nodes* whose removal causes a proportionally severe disconnection in the resulting network must be consistent with the following relative ranking of resilience for any even $n > 3$:

$$\text{Resilience}(\text{Star}(n)) < \text{Resilience}(\text{Barbell}(n)) < \text{Resilience}(\text{Apple}(n)) \tag{4.1}$$

TABLE 1. Resilience measures for 3 families of graphs.

Graph Type	VAT	I(G)	t(G)	T(G)	h(G)	$\epsilon^V(G)$	$\Phi(G)$
Star	$\frac{1}{n-1}$	$\frac{2}{n}$	$\frac{1}{n-1}$	$\frac{2}{n-1}$	$\frac{1}{n-1}$	$\frac{4}{n}$, n even; $\frac{4n}{n^2-1}$, n odd	1
Barbell	$\frac{2}{n}$	$\frac{1}{2}+\frac{1}{n}$	$\frac{1}{2}$	$\frac{n+2}{4}$	$\frac{1}{2}$	$\frac{4}{n}$	$\frac{2}{n}$
Apple	$\frac{1}{2}$	$\frac{n-1}{n}$	$\frac{1}{2}$	$\frac{n-1}{2}$	$\frac{1}{2}$	$\frac{n}{n-1}$	1

TABLE 2. Comparison of measures on 7 graphs.

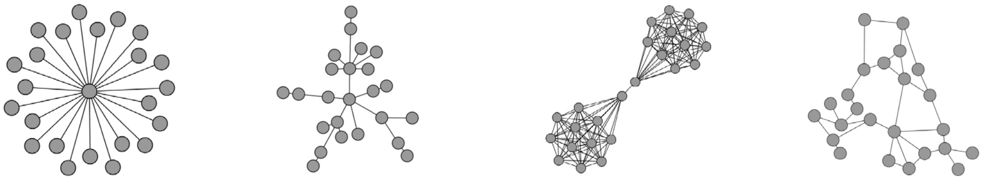
Graph Type	VAT	I(G)	T(G)
Star	.043	.083	.087
Tree	.063	.292	.467
Barbell	.083	.541	6.500
C3	.200	.417	.833
Apple	.500	.958	11.500
3-Regular	.538	.500	1.400
Cayley	1.000	.583	2.333

We have already established why the star graph is maximally vulnerable against a targeted node attack: the size of the critical node set attacked is minimal (1) while the resulting disconnection is maximal in that all components are isolated nodes. For the case of both the apple and the barbell, it is clear that a single node attack results in two cliques: for apple, those two cliques are of size $n-1$ and 1, whereas for barbell they are each of size $\frac{n}{2}$. Because a clique is a maximally connected graph, there can be no further attack that is *critical* in the sense that the cost of the attack justifies the amount of damage that can result. Therefore, the unique critical attack for the apple results in almost all nodes remaining connected to each other, whereas the barbell attack creates the worst case separation into *two* components possible as it minimizes the total number of connected pairs of nodes possible subject to the total node and component constraints. Therefore, we have justified Remark 4.1.

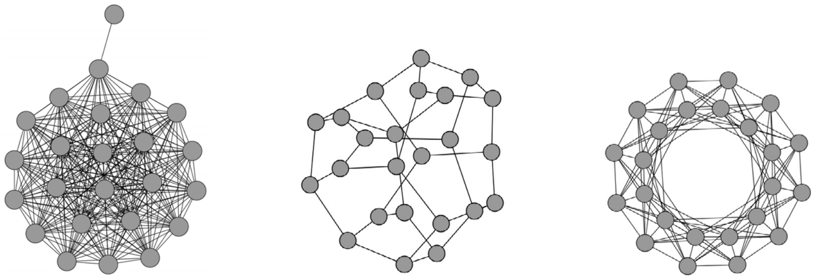
In Table 1 we exhibit our resilience computations for the star, barbell, and apple, for the following measures: VAT, normalized integrity ($I(G)$), toughness ($t(G)$), tenacity ($T(G)$), smoothed inverse scattering ($h(G)$), vertex expansion ($\epsilon^V(G)$), and conductance ($\Phi(G)$). Explanations for the computations of Table 1 may be found in author John Matta's thesis [29]. Amongst the resilience notions considered, it can be seen that only VAT, integrity, and tenacity respect the ordering stated by Remark 4.1. Therefore, in further experimental comparisons, we restrict ourselves to VAT, integrity, and tenacity.

Experiments were conducted on the following seven graphs of Figure 1, each having exactly 24 nodes: the tree is generated randomly according to the $HOTNet$ algorithm given in [17], resulting in a graph with a scale-free degree distribution. The C3 graph was adapted from a graph used in [11], with some nodes removed to yield 24 nodes, and is chosen to represent the case where a multiple-vertex bottleneck would intuitively divide a graph into three components. Both the random 3-regular graph [21, 22] and the random Cayley graph [32] are representative of *expander families* of graphs, whose extremely high resilience is well-established. Furthermore, the number of generators for the random Cayley graph are chosen to guarantee greater expansion properties than the random 3-regular graph with high probability.

Table 2 summarizes the resilience measures on the seven representative graphs. All measures are exactly computed *via* branch-and-bound methods. A striking aspect of the results are in the relatively *high* values



(a) Star (b) Tree (c) Barbell (d) C3



(e) Apple (f) Random 3-Regular Expander (g) Random Cayley Expander

FIGURE 1. The graphs used in our comparisons.

accorded to the barbell graph *via* both integrity and tenacity. In particular, the tenacity of the barbell graph is significantly higher than *both* expander graphs, while the integrity of the barbell is higher than the 3-regular expander and very close to the integrity of the Cayley expander. In fact, the integrity of the Cayley expander and the integrity of the barbell are closer to each other than any other pair of integrity values for the seven graphs considered. Moreover, the second two closest pair of integrity values is for the C3 graph and the 3-regular expander. Vertex attack tolerance, on the other hand, is unique in its ranking of the expander families as exhibiting the highest resilience, clearly separated from all other graphs except possibly for the apple graph. We understand that different measures may certainly be more useful depending on other contexts. However, VAT is unique amongst all measures considered in simultaneously satisfying Remark 4.1 *and* the according relatively high resilience to expander families, particularly when compared against a maximally bottlenecked graph such as the barbell.

5. Computational aspects of VAT

The results of the previous section used an exact branch and bound algorithm that is infeasible for significantly larger graphs. In this section we first prove the hardness of approximating unsmoothened VAT (UVAT) $\hat{\tau}$ under various plausible computational complexity hypothesis, and then empirically investigate the performance of a greedy betweenness centrality based algorithm to compute not only VAT but also other resilience measures.

5.1. Hardness of approximating UVAT

Our reduction extends the techniques for the NP-Hardness proof for the vertex integrity of co-bipartite graphs presented in [13]. Similarly, our computational hardness results for UVAT involves reductions with the Balanced Complete Bipartite Subgraph problem (BCBS). The BCBS problem is defined as:

Definition 5.1. *Instance:* A balanced bipartite graph $G = (V_1, V_2, E)$ with $n = |V_1| = |V_2|$ and an integer $0 < k \leq n$. *Question:* Does there exist $A \subset V_1$ and $B \subset V_2$ such that $|A| = |B| = k$ and (A, B) form a $k \times k$ complete bipartite graph?

The maximization version of the problem can be referred to as MAX-BCBS. The following three theorems regard the hardness of approximating MAX-BCBS under various plausible complexity theoretic assumptions:

Theorem 5.2 [20]**.** *It is NP-hard to approximate the MAX-BCBS problem within a constant factor if it is NP-hard to approximate the maximum clique problem within a factor of* $n/2^{c\sqrt{\log n}}$ *for some small enough* $c > 0$.

Theorem 5.3 [25]**.** *Let* $\epsilon > 0$ *be an arbitrarily small constant. Assume that SAT does not have a probabilistic algorithm that runs in time* $2^{n^{\epsilon}}$ *on an instance of size* n. *Then there is no polynomial time (possibly randomized) algorithm for MAX-BCBS that achieves an approximation ratio of* $N^{\epsilon'}$ *on graphs of size* N *where* $\epsilon' = \frac{1}{2^{O(1/\epsilon \log(1/\epsilon))}}$.

Theorem 5.4 [23]**.** *MAX-BCBS is R4SAT-Hard*[2] *to approximate within a factor of* n^{δ} *where* n *is the number of vertices in the input graph, and* $0 < \delta < 1$ *is some constant. More specifically, under the random 4-SAT hardness hypothesis: there exists two constants* $\epsilon_1 > \epsilon_2 > 0$ *such that no efficient algorithm is able to distinguish between bipartite graphs* $G = (V_1, V_2, E)$ *with* $|V_1| = |V_2| = n$ *which have a clique of size* $\geq (n/16)^2(1+\epsilon_1)$ *and those in which all bipartite cliques are of size* $\leq (n/16)^2(1+\epsilon_2)$.

As the hardness of random 3-SAT would imply the hardness of random 4-SAT, it follows that MAX-BCBS is R3SAT-Hard to approximate with same setting of parameters. Nonetheless, we mention the following previous result of [19] in which some parameters differ:

Theorem 5.5 [19]**.** *MAX-BCBS is R3SAT-Hard to approximate within a factor of* n^{δ} *where* n *is the number of vertices in the input graph, and* $0 < \delta < 1$ *is some constant. For every* $\epsilon > 0$, *it is R3SAT-Hard to approximate MAX-BCBS within a factor of* $1/2 + \epsilon$. *More specifically, it is R3SAT-hard to distinguish between the cases* $k > (1/4 - \epsilon)n$ *and* $k < (1/8 + \epsilon)n$.

Our main hardness result is as follows:

Theorem 5.6. *All of the following statements hold even when UVAT is restricted to co-bipartite graphs.*

(I) *It is NP-Hard to approximate UVAT within a constant factor if it is NP-hard to approximate the maximum clique problem within a factor of* $n/2^{c\sqrt{\log n}}$ *for some small enough* $c > 0$.
(II) *Let* ϵ, ϵ' *be as in Theorem 5.3. If SAT has no probabilistic algorithm that runs in time* $2^{n^{\epsilon}}$ *on instances of size* n, *then there is no polynomial time (possibly randomized) algorithm for UVAT that achieves an approximation ratio of* $N^{\epsilon'}$ *on graphs of size* N
(III) *UVAT is R4SAT-Hard (and, therefore, R3SAT-Hard) to approximate within a factor of* n^{δ} *where* n *is the number of vertices in the input graph, and* $0 < \delta < 1$ *is some constant.*

The theorem follows directly from part (III) of the following Lemma and Theorems 5.2–5.5.

Lemma 5.7. *Let* $G = (V_1, V_2, E)$ *with* $|V_1| = |V_2| = n$ *be a bipartite graph with* $E \neq \emptyset$, *and let* $\overline{G} = (V_1, V_2, \overline{E})$ *be the co-bipartite complement of* G. *Let* $BK(G) = \{(A,B) | A \times B$ *is a bipartite clique in* G *with* $|A| \leq |B|\}$. *Moreover, let* $BBK(G) = \{(A,B) \in V_1 \times V_2 | A \times B$ *is a bipartite clique of* G *with* $|A| = |B|\}$, *and let* $(\hat{A}, \hat{B}) = argmax_{(A,B)\in BBK(G)} |A|$ *be the maximum balanced bipartite clique of* G *with corresponding size* $k = |\hat{A}|$. *Then, the following hold:*

(I) $\hat{\tau}(\overline{G}) = \min_{(A,B)\in BK(G)} \frac{2n-|A|-|B|}{|A|} = \min_{(A,B)\in BK(G)} \frac{2n-|B|}{|A|} - 1$
(II) $\frac{n}{k} - 1 \leq \hat{\tau}(\overline{G}) \leq 2(\frac{n}{k} - 1)$
(III) *If UVAT can be approximated to factor* α *in polynomial time, then MAX-BCBS can be approximated to factor* 2α *in polynomial time, even when restricted to co-bipartite graphs.*

[2] R4SAT-Hard and R3SAT-Hard refer to hardness under the assumption that random 4-SAT and random 3-SAT are NP-Hard respectively.

Proof of Lemma 5.7. Let $S = S(\tau(\overline{G}))$, $U = S(\hat{\tau}(\overline{G}))$, $R = S(I(\overline{G}))$, and $C = S(T(\overline{G}))$ be the critical attack sets corresponding to τ, $\hat{\tau}$, I, and T for $\overline{G}$, respectively. Furthermore, let $S_i = V_i \cap S$, $U_i = V_i \cap U$, $R = V_i \cap R$, and $C_i = V_i \cap C$. For $X \in \{S, U, R, C\}$, let $A_X = \min\{V_1 - X_1, V_2 - X_2\}$ and $B_X = \max\{V_1 - X_1, V_2 - X_2\}$.

Note that $\overline{G}$ is not a clique as $E \neq \emptyset$. Moreover, because V_1 and V_2 must both be cliques in $\overline{G}$, A_X and B_X must each be cliques in $\overline{G}$, for any $X \in \{S, U, R, C\}$. Namely, the removal of X results in exactly two cliques A_X and B_X in $\overline{G}$. Clearly, there can be no edge between A_X and B_X in $\overline{G}$ as such an edge would have remained upon the removal of X. Therefore, (A_X, B_X) forms a bipartite clique in G. Part (I) of the lemma now follows from the definitions of $\hat{\tau}$ and the fact that $|A_X| \leq |B_X|$.

Now note that for any $(A, B) \in BK(G)$, any subset $B_A \subset B$ such that $|B_A| = |A|$ forms a balanced bipartite clique with A. Also clearly, $BBK(G) \subset BK(G)$. Therefore, by (I) and fact that $|A_X| \leq |B_X| \leq n$, (II) follows as well.

For part (III): let M be an algorithm that gives a constant factor approximation for UVAT with approximation factor $\alpha > 1$. Let q such that $\hat{\tau}(\overline{G}) \leq q \leq \alpha\hat{\tau}(\overline{G})$ be the approximation to $\hat{\tau}$ computed *via* M on the input. Simplifying and rearranging Lemma 5.7 part (II.b):

$$\frac{n}{q+1} \leq k \leq \frac{n}{1 + q/(2\alpha)} \tag{5.1}$$

Similarly, let $r = (\frac{n}{q+1})/(\frac{n}{1+q/(2\alpha)})$ denote the ratio between the right hand side and left hand side of the inequality, so:

$$r = \frac{q+1}{1 + q/(2\alpha)} \tag{5.2}$$

If $r > 2\alpha$ then $1 > 2\alpha$ resulting in a contradiction. Therefore,

$$\frac{n}{q+1} \leq k \leq 2\alpha\frac{n}{q+1} \tag{5.3}$$

And, $\frac{n}{q+1}$ is thus a $\frac{1}{2\alpha}$ approximation for the MAX-BCBS problem with corresponding approximation ratio 2α. □

Given the varied evidence for the approximation hardness of UVAT based on plausible complexity assumptions, we conjecture that both VAT and UVAT are at least NP-hard to compute:

Conjecture 5.8. *VAT and UVAT are NP-Hard.*

5.2. An algorithmic paradigm: Performance of Greedy-BC

Based on the approximation hardness of UVAT, the following conjectured hardness of VAT, and previously existing NP-Hardness results for all other resilience measures considered in this work, we investigate use of a greedy heuristic to approximately compute values not only for VAT, but also for integrity, toughness, tenacity and scattering number. The algorithm, called Greedy-BC, is implemented using fast betweenness centrality [7] for unweighted graphs. The algorithm is shown in Figure 2. With the fast betweenness centrality implementation, the overall running time of Greedy-BC is $O(|V|^2|E|)$.

VAT, integrity, toughness, tenacity and smoothed inverse scattering number are all minimization problems, whose objective functions involve similar calculations of component orders and numbers. The main difference between these resilience measures is the objective itself. Therefore, our Greedy-BC algorithm can be used to calculate any of the resilience measures. At each step of the algorithm we choose the remaining node with the highest betweenness centrality, add it to the attack set S, and remove the node from the graph. We calculate and record the resilience measure's value at that configuration. At the end of the algorithm we choose the set S that minimizes the desired resilience measure. We then *optionally* run a hill climbing procedure on the resulting S, to attempt further convergence to the true value of the resilience measure.

```
heuristic GREEDY-BC(G, res) is
    input: G, a graph with vertices, V
        res, a resilience measure
    output: S, a critical attack set of nodes
    S_0 ← betweennessApprox(G, res)
    S ← hillclimb(G, S_0, res)
    return S
end heuristic
procedure betweennessApprox(G)
    res_best ← ∞
    S_best ← S_cur ← ∅
    for n = 0 to |V| do
        BC[ ] ← BrandesBetweennessCentrality(G)
        V_max ← max(BC)
        S_cur ← S_cur ∪ V_max
        if res(S_cur) < res_best then
            S_best ← S_cur
            res_best ← res(S_cur)
        end if
        G ← G − V_max
    end for
    return S_best
end procedure
```

FIGURE 2. Greedy-BC Method with optional hill-climbing to approximate different resilience Measures.

TABLE 3. Correctness of Greedy-BC algorithm on 24-node graphs before hill climbing. The number shown is the percentage error of the calculation. Zero indicates that the true value was determined by the algorithm. A score greater than 0 indicates a percentage over the true value.

Graph Type	VAT	I(G)	t(G)	T(G)	h(G)
Star	0	0	0	0	0
Tree	0	0	17%	0	0
Barbell	0	0	0	0	0
C3	7%	0	71%	2%	20%
Apple	0	0	0	0	0
3-Regular	0	0	60%	34%	0
Cayley	9%	0	200%	0	0

We investigated the accuracy of the greedy heuristic compared to both optimal and hill-climbing results, for each resilience measure. The comparison with the hill-climbing results was performed due to the extra computational overhead of hill-climbing, so that we could determine if hill-climbing was worthwhile empirically. To compare to the exact optimal results, Greedy-BC was first run on the seven original 24-node graphs, for which the exact values of all resilience measures are known.

As can be seen in Table 3 the heuristic did very well in calculating the resilience measures for the 24-node graphs considered. Values for Star, Barbell and Apple were calculated correctly for all measures. The largest error in calculating VAT was 9% for the Cayley expander graph. The worst results were obtained for toughness, where four out of seven results were off by more than 10%. Overall, results are perfect for integrity, and very good for VAT, tenacity and scattering number.

```
procedure HILLCLIMB(G, S, res, hd)
    input: G, a graph with vertices, V
        res, a resilience measure
        hd, a Hamming distance
    output: S, a critical attack set of nodes
    changed = true
    while changed do
        changed = false
        for each S_0 a Hamming distance of hd from S do
            if res(S_0) < res(S) then
                S ← S_0
                changed = true
            end if
        end for
    end while
    return S
end procedure
```

FIGURE 3. 1D Hill climbing.

TABLE 4. Improvement from 1-D hillclimbing on 24-node graphs. The number shown is the percentage of improvement over the original calculation. Zero indicates no improvement.

Graph Type	VAT	I(G)	t(G)	T(G)	h(G)
Star	0	0	0	0	0
Tree	0	0	0	0	0
Barbell	0	0	0	0	0
C3	0	0	14%	0	0
Apple	0	0	0	0	0
3-Regular	0	0	7%	7%	0
Cayley	0	0	0	0	0

The second question considered is whether results are improved by hill climbing. Specifically, results must be improved enough to warrant the increased computational complexity. We tested two different forms of hill climbing, 1-D and 2-D. 1-D hill climbing is based on a neighborhood Hamming distance of 1. The membership status of one vertex in $|S|$ is changed, and results are moved in the direction of the largest improvement. 2-D hill climbing uses a Hamming distance of two: the membership of two vertices is changed. All combinations with Hamming distance of 2 are tested. In both cases, if a change will improve the score, the modified set with the lowest resilience measure is chosen and the process is repeated. Algorithms for hill climbing are given in Figure 3.

As shown in Table 4, improvement from 1-D hill climbing is very small. In most cases the algorithm has calculated the correct measure without hill climbing, and therefore no improvement is possible. Improvement was only possible in 9 of 35 cases, and only actually occurred in 3 of the nine. As hill climbing potentially imposes a high cost in terms of computational complexity, the improvement from hill climbing does not seem to be worth the cost. Improvement from 2-D hill climbing also remains small, and is shown in Table 5. 2-D hill climbing caused an improvement of results in only 5 cases, at a high computational cost.

In order to further examine the usefulness of hill climbing we tested it on 18 larger graphs, 9 preferential attachment (BA) graphs, and 9 PLOD graphs with the same degree structure. These graphs are large enough that calculating the resilience measures with branch and bound methods is infeasible. For each graph, we have three values: the initial resilience calculation using only betweenness centrality, the improvement from 1-D hill

TABLE 5. Improvement from 2-D hillclimbing on 24-node graphs. The number shown is the percentage of improvement over the original calculation. Zero indicates no improvement.

Graph Type	VAT	I(G)	t(G)	T(G)	h(G)
Star	0	0	0	0	0
Tree	0	0	17%	0	0
Barbell	0	0	0	0	0
C3	0	0	71%	2%	19%
Apple	0	0	0	0	0
3-Regular	0	0	0	20%	0
Cayley	0	0	0	0	0

TABLE 6. Correctness of Greedy-BC algorithm on 24-node graphs. Three numbers given are the percentage error without hill climbing, the percentage improvement from 1-D hill climbing, and the percentage improvement from 2-D hill climbing. An initial zero indicates that the true value was determined by the algorithm. Improvement scores greater than 0 indicates a percentage over the original computation.

Graph Type	VAT	I(G)	t(G)	T(G)	h(G)
Star	0	0	0	0	0
Tree	0	0	17%/0/17%	0	0
Barbell	0	0	0	0	0
C3	7%/0/0	0	71%/14%/71%	2%/0/2%	20%/0/19%
Apple	0	0	0	0	0
3-Regular	0	0	60%/7%/0	34%/7%/20%	0
Cayley	9%/0/0	0	200%/0/0	0	0

TABLE 7. Improvement from 1-D and 2-D hillclimbing on larger graphs. The number shown is the percentage of improvement over the original calculation. Zero indicates no improvement.

	VAT		integrity		toughness		tenacity		scattering	
BA graphs	1D	2D	1D	2D	1D	2D	1D	2D	1D	2D
Barabasi−Albert Graphs										
AVG	2	2	2	3	6	6	2	3	0	5
MAX	18	19	7	7	11	14	11	11	4	29
MED	0	0	0	3	5	5	1	3	0	1
PLOD Graphs										
AVG	2	3	0	0	4	4	1	3	7	27
MAX	9	9	1	1	7	7	5	11	39	89
MED	3	3	0	0	3	3	1	3	0	25

climbing, and the improvement from 2-D hill climbing. The results of the average, maximum and median improvements for 1-D and 2-D hill climbing are shown in Figure 7.

Average improvement for VAT was only 2−3%, for both 1-D and 2-D hill climbing. In fact, five of nine Barabasi−Albert graphs did not show any improvement from hill climbing, and only one showed an improvement of more than 2%. The same-degree PLOD graphs showed equally unimpressive improvement. Hill climbing did not improve any by PLOD result by more than 9%. Combined with the previous results showing the heuristic's good accuracy in calculating VAT, we determine that the improvement resulting from hill climbing is not worth the cost.

As can be seen in Table 7, hill climbing improvements for integrity were a maximum of 7%, and had the same low 2−3% average as VAT. The max improvement for toughness was only 14%, and the max improvement for tenacity was only 11%. These results imply a large robustness for Greedy-BC in calculating all these resilience measures. Hill climbing was more effective with scattering number. The PLOD graphs showed an average improvement of 29%, and a maximum improvement of 89%. The average improvement with the Barabasi−Albert graphs, even with 2-D hill climbing, was still only a small 5%.

Given the aforementioned hardness results concerning every one of the resilience measures considered, our Greedy-BC heuristic gives an effective way to estimate their value and determine the corresponding attack set.

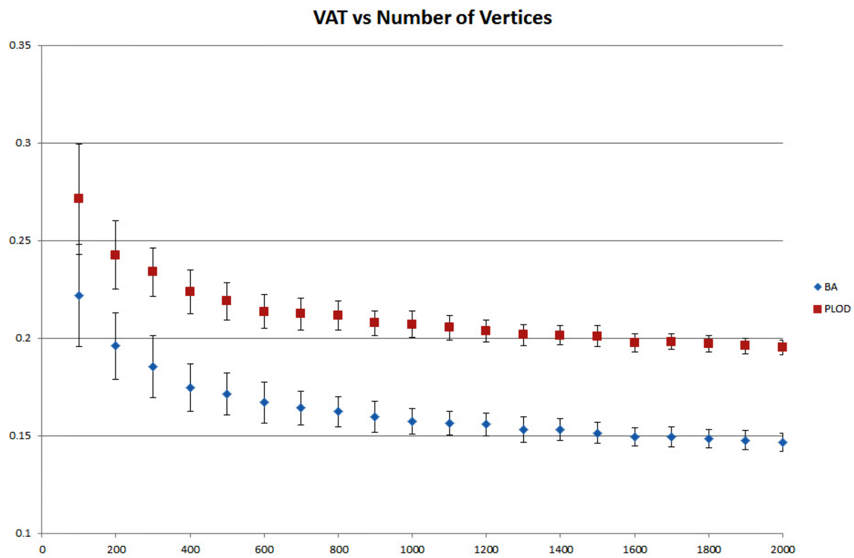
6. VAT of scale-free networks

In this final module, we investigate the vertex attack tolerance of *scale-free networks*. The specific models under consideration are the Barabasi−Albert preferential attachment (PA or BA) model and the random scale free PLOD model. The results obtained by running our hill-climbing (1-D) algorithm on the preferential attachment based Barabasi−Albert graphs [4] (with $m = 2$ neighbors chosen for each entering vertex) and the random scale-free graphs of the same degree distribution can be seen in the plots of Figure 4. The *degree distribution* of the random scale-free graphs were *generated* by the corresponding preferential attachment graph. Because we are interested in how the topological generative properties affects the resilience of different scale-free models, we felt it is particularly important to control for degree distribution as much as possible. Moreover, because a random graph of any given degree distribution can be defined, it is appropriate in this case to first generate the PA graph and then feed its degree distribution into the random scale-free graph generator. We note further that our random scale-free graph generator is thus identical to the PLOD model [33] except for our explicit input of the degree distribution. Unfortunately, for HOTNets we cannot guarantee generating a HOTnet with an exactly pre-specified degree distribution.

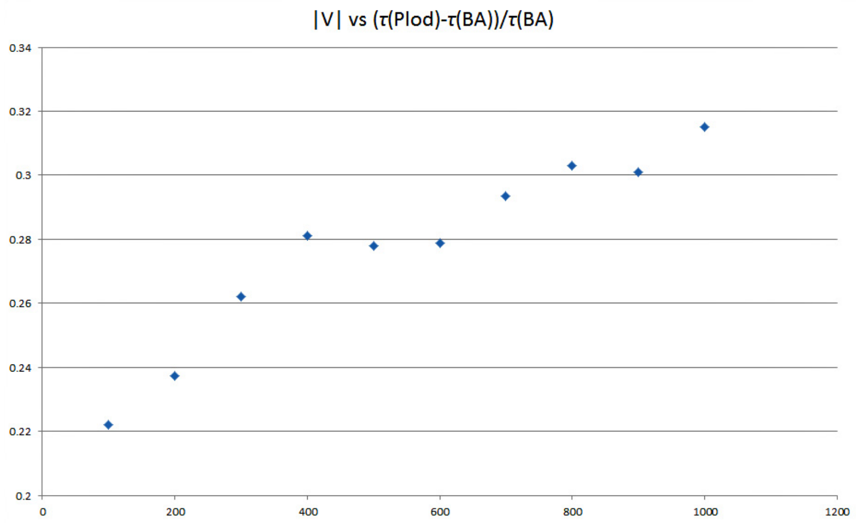
It can be seen that the random scale-free graphs have consistently better vertex attack tolerance than the preferential attachment based Barabasi−Albert graphs for the parameter $m = 2$. We also computed VAT values for various HOTnets, and all appeared significantly less resilient than the PA and random scale-free graphs of corresponding size and same average degree. We computed PA graphs with parameter $m = 1$ as well and discovered that such PA graphs were significantly non-resilient, with easily identifiable hubs whose removal causes severe disruption, yielding very low VAT values. We had difficulty in generating identical degree PLOD graphs from PA graphs with $m = 1$ due to the preponderance of degree one nodes. Additionally the generated PLOD graphs tended not to be connected, so that the VAT value was trivially zero[3]. Therefore, we include the results for the $m = 2$ PA graphs and their corresponding degree PLOD counterparts only, noting that such graphs remain relatively resilient despite doubling of graph sizes. However, the corresponding PLOD graphs remain *consistently more* resilient than the PA graphs of identical degree distribution, provided they are connected (which happens in likelihood given $m = 2$).

In addition to the actual computed VAT values, one may view the worst-case disruptions caused by the critical attack sets of the PA graph and the identically sized and degree distributed PLOD graph of 1000 nodes respectively in Figure 5. The severity of the disruption caused despite the relatively high VAT values is due to the large size of the attack set required. The ability to discover such a large and effective attack set is further indicative of the quality of our algorithm. Indeed, the attack sets discovered by other algorithms we have previously attempted, such as genetic algorithms and simulated annealing approaches, on the same graphs resulted in slightly higher VAT values and much smaller attack sets yielding far less disruption, particularly for the PLOD graph. The critical attack of the PA graph resulted in 142 components, the largest component of which had 64 nodes, whereas the critical attack of the PLOD graph of identical degree distribution resulted in 66 components, the largest component of which had 69 nodes, illustrating that the disruption caused to the PA graph was more severe. Moreover, the size of the attack set required for the PLOD was significantly higher

[3]While the vector $\boldsymbol{\tau}$ over VAT values over all components could be considered, different component size distributions would also have affected the meaningfulness of such comparisons.



(a) VAT as a function of Graph Size

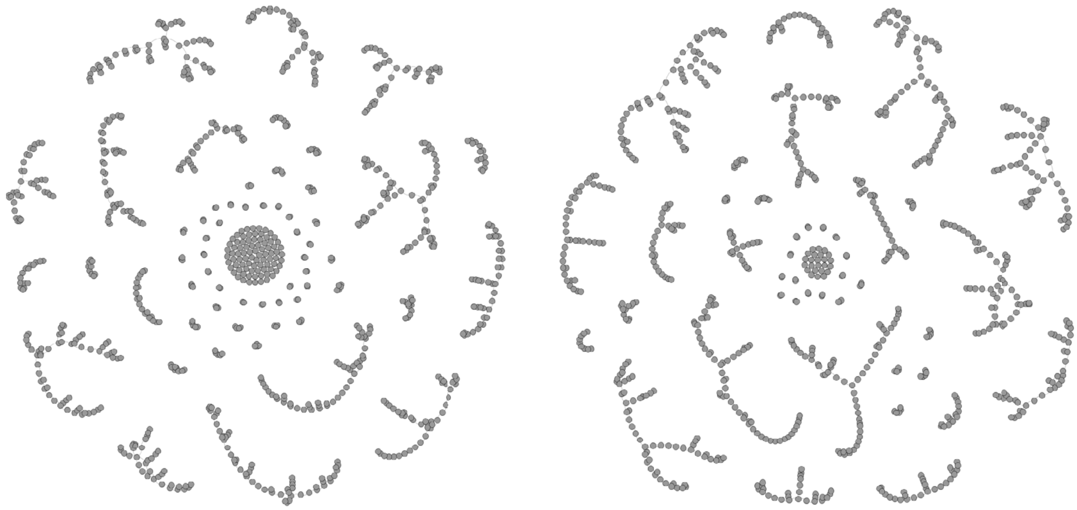


(b) Relative improvement of PLOD VAT vs. PA VAT

FIGURE 4. VAT Comparisons of PA and PLOD graphs.

than the size of the attack set required for the PA graph, further illustrating the higher resilience of the random scale free networks. We reiterate, in the spirit of [2], that the resilience of a scale-free graph is highly dependent on the exact generative model, even when two graphs have *identical degree distributions*. And, just as many expander families known to exhibit high resilience are based on *random* constructions, *random* scale-free graphs also appear to be amongst the more resilient graphs satisfying the scale-free property.

Finally, we confirm our results concerning the relatively greater resilience of PLOD graphs compared to BA graphs by measuring resilience with respect to integrity, toughness, tenacity, and scattering number as well.



(a) PA Attack (b) PLOD Attack

Figure 5. PA and PLOD attack visualization. Nodes in the center of the visualization form a group of isolated nodes not belonging to any non-trivial component.

Table 8. Resilience of Barabasi−Albert preferential attachment graphs as compared to same-degree randomly generated PLOD graphs. Number shown is B-A resilience/ PLOD resilience. A number greater than one indicates that the resilience of the PLOD graph is greater.

Nodes	VAT	I(G)	t(G)	T(G)	h(G)
35	0.833	1.000	1.072	1.000	0.751
40	1.110	1.000	1.310	1.303	2.252
45	1.130	1.133	1.460	1.199	4.000
50	1.144	1.067	1.610	1.491	5.495
100	1.152	1.192	1.425	1.332	3.571
250	1.209	1.164	1.782	1.677	13.157
500	1.315	1.123	1.601	1.585	4.500
1000	1.209	1.159	1.581	1.570	3.800
2500	1.322	1.235	1.678	1.664	6.141

This pattern of greater resilience for random PLOD graphs continues with the other resilience measures, as well. In Table 8, we give the ratio of each resilience measure for PA versus PLOD, for graphs of increasing size. It can be seen that, in almost every case, the PLOD graph has a higher resilience, no matter which resilience measure is used. It is also interesting to notice that the ratio of the resiliences of the two graphs generally increases as the size of the graph increases. That is to say, the PLOD graphs become not only more resilient, but also proportionally more resilient, as the size of the graph increases.

7. Conclusion

In this work we have considered mathematically robust notions of node-based resilience for complex, scale-free networks with the aim of comparing the resilience of different scale-free models. We have sought a resilience measure that not only acts provably similar to conductance in the case of regular degree graphs but also correctly captures intuitive resilience orderings comparing various graph families of arbitrary degree distributions, including the strict inequalities Resilience(Star) $<$ Resilience(Barbell) $<$ Resilience(Apple). With such motivation, we have proposed vertex attack tolerance $\tau(G)$: we have proved that indeed $\tau(G) = \Theta(\Phi(G))$ for undirected d-regular graphs, yielding spectral bounds as corollaries. Furthermore, not only is $\tau(\text{Star}) < \tau(\text{Barbell}) < \tau(\text{Apple})$, but also τ is unique in capturing the resilience orderings for the graphs of Table 2 when compared with other well-known measures such as integrity, tenacity, toughness, conductance, vertex expansion, and scattering number.

Towards the purpose of actual measurement, we have explored computational questions concerning vertex attack tolerance with mixed results: On the one hand, we have proved that unsmoothened vertex attack tolerance $\hat{\tau}(G)$ is hard to approximate under any of four different plausible computational complexity hypotheses even when restricted to co-bipartite graphs. On the other hand, we have also proposed a novel greedy betweenness centrality based heuristic Greedy-BC to empirically approximate not only vertex attack tolerance, but all other considered resilience measures, which are known to be NP-Hard in general graphs. The Greedy-BC heuristic did very well in calculating the resilience measures for all 24-node graphs considered, with perfect results achieved for integrity, and approximation factors within 1.1 achieved for VAT, tenacity, and scattering number. Furthermore, for the much larger scale-free graphs for which it was infeasible to calculate the measures exactly, we found that hill-climbing with a neighborhood defined by Hamming distance up to two does not significantly improve accuracy despite being significantly more inefficient. These results support the practical use of Greedy-BC for all of the resilience measures considered on various realistic network classes.

Finally, we applied our Greedy-BC and hillclimbing algorithms to compute every resilience measure for the preferential attachment Barabasi−Albert [4] (BA) networks, random scale-free [33] (PLOD) networks, and *heuristically-optimized trade-offs* [17] (HOTNet) networks. HOTNets exhibited the worst vertex attack tolerance on various parameter settings considered even when controlling for average degree. As we wished to control for degree distribution in our comparisons, and HOTNets cannot be tweaked to follow an exact degree distribution, we focused on extensively comparing BA networks to PLOD networks of identical degree distributions: our results indicate that PLOD graphs with degree distributions identical to PA graphs of the same size exhibit increasingly better resilience than the PA type graphs asymptotically and *across all resilience measures*, although both graph types appear surprisingly resilient when the generative PA parameter is $m = 2$. Our results may be compared and contrasted with previous claims [1, 2, 12] regarding the resilience or lack thereof of various scale-free networks.

References

[1] R. Albert, H. Jeong and A.-L. Barabasi, The internet's achilles' heel: Error and attack tolerance of complex networks. *Nature* **406** (2000) 200–0.

[2] D.L. Alderson, L. Li, W. Willinger and J.C. Doyle, Understanding internet topology: principles, models, and validation. *IEEE/ACM Trans. Netw.* **13** (2005) 1205–1218.

[3] N. Alon, Eigenvalues and expanders. *Combinat.* **6** (1986) 83–96.

[4] A.-L. Barabási and R. Albert, Emergence of scaling in random networks. *Sci.* **286** (1999) 509–512.

[5] C.A. Barefoot, R. Entringer and H. Swart, Vulnerability in graphs-a comparative survey. *J. Comb. Math. Comb. Comput.* **1** (1987) 12–22.

[6] D. Bauer, S.L. Hakimi and E. Schmeichel, Recognizing tough graphs is np-hard. *Discrete Appl. Math.* **28** (1990) 191–195.

[7] U. Brandes, A faster algorithm for betweenness centrality. *J. Math. Soc.* **25** (2001) 163–177.

[8] H. Broersma, J. Fiala, P.A. Golovach and T. Kaiser, Daniël Paulusma and Andrzej Proskurowski. Linear-time algorithms for scattering number and hamilton-connectivity of interval graphs. *J. Graph Theory* **79** (2015) 282–299.

[9] F. Chung, Spectral Graph Theory. American Mathematical Society (1997).

[10] V. Chvatal, Tough graphs and hamiltonian circuits. *Discrete Math.* **306** (2006) 910 – 917.

[11] L.F. Costa, F.A. Rodrigues, G. Travieso and P.R. Villas Boas, Characterization of complex networks: A survey of measurements. In *Adv. Phys.* **56** (2007) 167–242.

[12] J.C. Doyle, D.L. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka and W. Willinger, The robust yet fragile nature of the internet. *Proc. of the National Academy of Sciences of the United States of America* **102** (2005) 14497–14502.

[13] P.G. Drange, M.S. Dregi and P. vant Hof, On the computational complexity of vertex integrity and component order connectivity. In Algorithms and Computation. Springer International Publishing (2014) 285–297.

[14] G. Ercal, On vertex attack tolerance of regular graphs. Preprint arXiv:1409.2172 (2014).

[15] G. Ercal, A note on the computational complexity of unsmoothened vertex attack tolerance. Preprint arXiv:1603.08430 (2016).

[16] G. Ercal and J. Matta, Resilience notions for scale-free networks. In Complex Adaptive Systems (2013) 510–515.

[17] A. Fabrikant, E. Koutsoupias and Ch.H. Papadimitriou, Heuristically optimized trade-offs: A new paradigm for power laws in the internet. In *ICALP* (2002) 110–122.

[18] U. Feige, M. Hajiaghayi and J. Lee, Improved Approximation Algorithms for Minimum-Weight Vertex Separators. In *Proc. of the Thirty-seventh Annual ACM Symposium on Theory of Computing*. ACM, New York (2005) 563–572.

[19] U. Feige, Relations between average case complexity and approximation complexity. In *Proc. of the Thiry-fourth Annual ACM Symposium on Theory of Computing*. ACM (2002) 534–543.

[20] U. Feige and Sh. Kogan, *Hardness of approximation of the balanced complete bipartite subgraph problem.* Technical Report MCS04-04 (2004).

[21] J. Friedman, J. Kahn and E. Szemeredi, On the second eigenvalue of random regular graphs. In *Proc. of the twenty-first annual ACM symposium on Theory of computing*, *STOC '89*. New York, NY, USA, ACM (1989) 587–598.

[22] Z. Furedi and J. Komlos, The eigenvalues of random symmetric matrices. *Combinatorica* **1** (1981) 233–241.

[23] A. Goerdt and A. Lanka, An approximation hardness result for bipartite clique (2004) 048.

[24] H.A. Jung, On maximal circuits in finite graphs. *Ann. Discrete Math.* **3** (1978) 129–144.

[25] S. Khot, Ruling out ptas for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM J. Comput.* **36** (2006) 1025–1071.

[26] A. Louis, P. Raghavendra and S. Vempala, The complexity of approximating vertex expansion. Preprint arXiv:1404.0103 (2013).

[27] S. Stueckle M. Cozzens, D.M. The tenacity of a graph. In *Seventh International Conference on the Theory and Applications of Graphs*, New York, NY, USA Wiley (1995) 1111–1122.

[28] D.E. Mann, *The Tenacity of Trees*. Ph.D. thesis, Northeastern University, Boston, MA (1993).

[29] J. Matta, *Comparing the Effectiveness of Resilience Measures*. Master's thesis, Southern Illinois University, Edwardsville (2014).

[30] J. Matta, J. Borwey and G. Ercal, Comparative resilience notions and vertex attack tolerance of scale-free networks. Preprint arXiv:1404.0103 (2014).

[31] M. Newman, A.-L. Barabasi and D.J. Watts, The Structure and Dynamics of Networks. Princeton University Press (2006).

[32] I. Pak, Random cayley graphs with o(log[g]) generators are expanders. In *Proc. of the 7th Annual European Symposium on Algorithms*, ESA '99, London, UK. Springer Verlag (1999) 521–526.

[33] C.R. Palmer and J.G. Steffan, Generating network topologies that obey power laws. In *Global Telecommunications Conference (2000). GLOBECOM '00. IEEE*, Vol. 1 (2000) 434–438.

[34] J. Šíma and S.E. Schaeffer, On the np-completeness of some graph cluster measures. In *SOFSEM: Theory and Practice of Comput. Sci.* Springer (2006) 530–537.

[35] A. Sinclair and M. Jerrum, Approximate counting, uniform generation and rapidly mixing markov chains. *Inf. Comput.* **82** (1989) 93–133.