FRCCS

From Random Perturbations To Localized Vulnerability Assessment in Interaction Graphs

Anas Zakroum[1,3✓], Roberto Interdonato[1,3], Pascal Degenne[1,3], Mathieu Roche[1,3], Danny Lo Seen [2,3]

[1] *CIRAD, UMR TETIS, F-34398, Montpellier, France. ; anas.zakroum@cirad.fr, roberto.interdonato@cirad.fr, pascal.degenne@cirad.fr, mathieu.roche@cirad.fr*
[2] *CIRAD, UMR TETIS, F-97455, Saint-Pierre, La Réunion, France.; danny_loseen@cirad.fr*
[3] *TETIS, Université de Montpellier, AgroParisTech, CIRAD, INRAE, Montpellier, France.*

[✓] *Presenting author*

Abstract. This study presents a framework to assess connectivity disruption by translating random perturbations to local vulnerabilities. Leveraging recent findings on random walks for local topology recollection, we introduce a method for quantifying changes within an (l)-local neighborhood. We derive vulnerability profiles for perturbed nodes and introduce a reachability index that aggregates neighboring nodes' vulnerabilities. Preliminary experiments on targeted attacks based on reachability scores shows performance on par with established strategies.

Keywords. *Network Robustness; Local Vulnerability; Anonymous walks*

1 Introduction

Networked systems can be subject to disturbances originating from internal failures or external perturbations [1]. These disturbances impact the inherent connectivity of the network through the loss of some of its nodes or the edges adjacent to them. Understanding the behavior of networked systems subject to such disturbances is central to the study of network robustness [1]. Connectivity disruption is usually investigated in the form of random failures or targeted attacks to which the response of the network is assessed at the large scale. For instance, a measure of common interest is the size of the remaining largest connected component subsequent to disturbances [3]. This measure is often used as a proxy to characterize the ability of a network to maintain its cohesion and function after an induced perturbation.

While connectivity disruption is often investigated at a large scale, we argue that the impact at a smaller granularity, e.g., on individual or groups of nodes, may be valuable enough to consider. In fact, recent studies suggest that even in networks exhibiting global robustness statistics, individual nodes may still be highly vulnerable [6]. Consequently, investigating the local impact of connectivity disruptions offers the potential for gaining detailed insights into vulnerable areas in networks.

In this work, we attempt to construct a framework under which connectivity disruption is assessed at the local scale, translating random perturbations into localized vulnerabilities. By

leveraging recent findings on random walks in capturing the topological information of a central node from its surroundings, we construct a metric that captures changes in the topology at l-hops around the nodes. We establish vulnerability profiles of the nodes against random perturbations and present the reachability index; a measure that accounts for the influence of neighboring nodes on a focus node.

We evaluate our approach by monitoring the largest component subsequent to targeted attacks in descending order of the reachability index (cf. paragraph 2). We compare our method to other to Our preliminary results show equivalent performances to established attack strategies in the network degradation task.

2 Methods

(A) Framework for assessing the impact of random perturbations. We consider failures occurring on edges. Our goal is to assess how the impact of such failures is reported on the nodes the graph. Let $G = (V,\ E)$ be a graph, $F \subset E$ a set of edges subject to random failures. Let $p \in (0, 1)$ be a parameter denoting the probability of failure of an edge. Each edge $e \in F$ is paired with a Bernoulli random variable $W \sim \mathcal{B}ern(p)$, and e is removed when its corresponding Bernoulli realization W_e succeeds, i.e., $W_e = 1$. According to the law of large numbers, for large networks, the fraction of removed edges goes with a high probability to the expected value of W, $\mathbb{E}(W) = p$. The expected number of edges removed is $\mathbb{E}\Big(\sum_{e\in F} W\Big) = |F|p$.

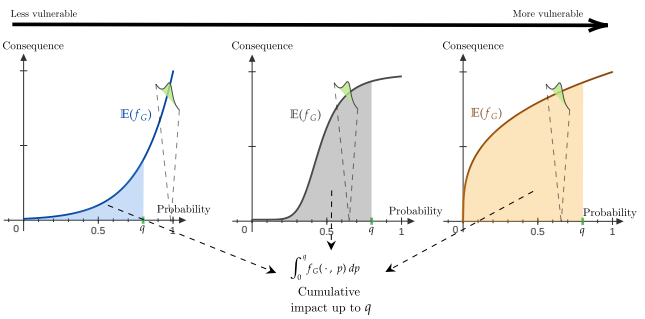
The consequences of the induced perturbations are aimed to be assessed at a nodal level. To this end, let's denote by $f_G(v, p) : V \times (0, 1) \longrightarrow \mathbb{R}_+$ a bounded non-negative real valued mapping that measures the magnitude of the induced perturbations on a node v. It is reported with respect original state of the graph. Since the induced perturbations are random, the impact function f_G is a random variable. Thus, we consider summarizing its randomness with the first order moment.

Depending on their connectivity, degree correlations and other characteristics, nodes are subject to impacts of different magnitudes. For instance, nodes exhibiting weak connectivity are usually more sensitive to disturbances and are expected to display higher magnitudes of the impact f_G. To account for this, we define the *vulnerability profile* of a node as the cumulative impact up to perturbation intensity q;

$$\begin{aligned} F_G \ :\ V \times (0,1) &\longrightarrow \ \mathbb{R}_+ \\ (\ \cdot\ , q) &\longmapsto \ \int_0^q \mathbb{E}\Big[f_G(\ \cdot\ ,\ p)\Big]\ \ \mathrm{d}p\ , \end{aligned} \tag{1}$$

where f_G is the impact function defined above. Figure 1 provides an illustration.

Figure 1: Illustration of a spectrum of possible vulnerability profiles.



Let $v \in V$ a node in G. When $q = 1$, we call $F(v) := F(v, 1)$ the expected risk on node v.

(B) Capturing the change in the surrounding topology of the nodes. As the induced perturbations occur on the edges, the network topology shifts in proportion to perturbation strength. Consequently, from a node's perspective, we hypothesize that quantifying changes in the subgraph formed by its surroundings might reveal insights into its vulnerability.

Recent results in [5, 4] show that under certain conditions, the distribution of the anonymized version of random walks (i.e dropped labels) starting from a node is sufficient to accurately reconstruct the topology of the ball centered around it. Ivanov et al. [4] propose an embedding framework along with a sampling scheme that allow for a vectorial representation of the nodes and the topology around them. Table 1 provides an illustration of how the embeddings are created.

Random walks	**Anonymous walks**	**Embedding**
Initial		For all $u \in V$, $q \in (0,1)$ $h_u^l = (p(w_1^l),\ \ldots,\ p(w_\eta^l))$
		$\boldsymbol{\eta}$ **:** No of **all anonymous walks** of length $\boldsymbol{l}$. $\boldsymbol{p(w_i^l)}$ **:** $\sum_{r \to a_i} \prod_{e \in r} p_e$
Perturbed		$g_u^l(q) = (p'(w_1^l),\ \ldots,\ p'(w_\eta^l))$

Table 1: Embeddings of nodes before and after introducing perturbations. $p(w_i^l)$ is the probability of observing anon. walk w_i^l of length l.

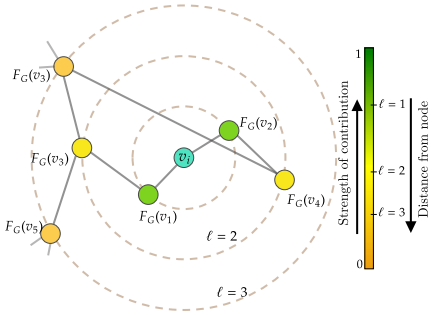


Figure 2: Strength of contribution (influence) of neighboring nodes.

Following [4], we represent each node v as a vector in $\mathbb{R}^\eta$ encoding a probabilistic representation of all possible anonymous walks of length l, thereby capturing the topology of the surrounding (l)-local neighborhood. Consequently, we define the impact function f_G as the cosine distance between the perturbed and original embeddings of the nodes

$$f_G(v, q) = 1 - \cos\Big(\ h_v^l\ ,\ g_v^l(q)\ \Big),$$

where $v \in V$ is a node of the graph and $q \in (0, 1)$ is the intensity of perturbations.

Sampling The true distribution of all anonymous walks starting from a given node is unknown. Thus, it is necessary to sample a sufficiently large number of walks m to approximate their true distribution. To this end, Ivanov et al. [4] adapts results from the problem of subgraph sampling in [8] to extract a sufficiently large number of walks m to guarantee small deviations of the empirical distribution of anonymous walks $\hat{D}^m$ from their true distribution D.

Let η be the number of all possible anonymous walks of length l. Then, for all $\epsilon > 0$ and $\delta > 0$, the number of walks m needed to satisfy the concentration bound $\mathbb{P}\left(\left\{\|D - \hat{D}^m\|_1 \geq \epsilon\right\}\right) \leq \delta$ is

$$m = \left\lceil \frac{2}{\epsilon^2}\Big(\log(2^\eta - 2) - \log(\delta)\Big)\right\rceil, \tag{2}$$

where $\|\cdot\|_1$ denotes the L_1 norm. The parameters ϵ and δ are parameters that control the concentration of the empirical distribution $\hat{D}$ around the true distribution D. By setting $\epsilon = 0.2$ and $\delta = 0.01$, we get $m = 7266$ walks to generate. In our case, we set $l = 6$, and use the parameters to ensure confidence in capturing the topological information around the nodes, both in the original and perturbed states.

(C) The contribution of neighboring nodes to the vulnerability Using Eq (1), each node $v \in V$ in the graph is assigned an expected risk $F_G(v)$ according to the impact defined in **(B)**. Our next step is to incorporate the influence of the vulnerability of neighboring nodes of

the graph. The influence of a node u on the vulnerability of node v is based on the distance between the two (the farther u from v, the lesser its contribution to the vulnerability of the latter). Nodes surrounded by vulnerable neighbors are expected to be less reachable in a perturbed network. Thus we define a reachability index on the nodes by considering all nodes within distance L as

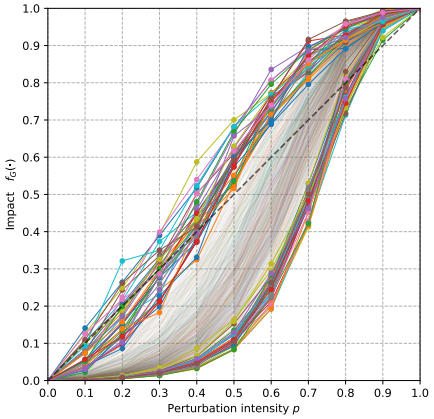
$$R_L(v) = F(v) + \sum_{u \in \mathcal{B}(v,L)} \beta^{\ell} F(u) \quad , \tag{3}$$

where $\mathcal{B}(v, L) = \{u \in V \mid u \neq v \ , \ d(v, u) \leq L\}$ is the set of nodes u of distance less than L from node v, $\ell = d(v_i, v_j)$, and $\beta \in (0, 1)$ is a discount factor. Figure 2 provides an illustration.

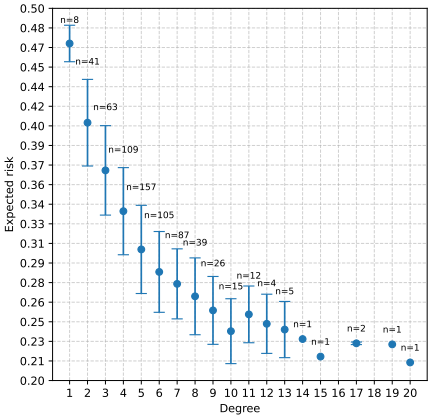
3 Preliminary results

We implement our method for a network of yeast propagation in agricultural plots (nodes). Graphs/networks in the agriculture domain enables modeling phenomena for public policies, e.g. modeling biomass exchanges to promote a circular economy, modeling pesticides in soils, etc [2] The edges are based on distances ($< 50m$). Edges represent the ability of the yeast to spread across plots. A summary of the network is given in Table 2.

We consider all anonymous walks of length $l = 6$ to create the embeddings of the nodes. For each node, we sample 7266 walks per node according to [4] to ensure an accurate estimation of the true distribution of anonymous walks of length l. By setting $q_i \in \{0.1, 0.2, \ldots, 0.9\}$, the impact function f_G defined in **(B)** is computed using Monte-Carlo estimation by repeating the graph perturbations $N = 100$ times (to get preliminary results) for all q_i. Reported in Figure 3, the vulnerability profiles and the distribution of the expected risk F_G with respect to node degree. In most cases, nodes of the yeast network showcase close to linear or sub-linearimpact to random perturbations (Fig. 3a) with an apparent dependency to the strength of their initial connections (Fig. 3b). We observe sharper increases in the impact around $p = 0.7$ indicating the percolation threshold.



(a) Vulnerability profiles of the nodes. Highlighted are the 5^{th} percentile of the expected risk.



(b) Distribution of the expected risk $F_G(\ \cdot\ , 1)$ with respect to node degree. We observe an inverse dependency.

Figure 3: Vulnerability profiles and distribution of the expected risk on the yeast network.

Next, we compute the reachability indexes of the nodes (Eq. (3)), by considering the discount factor β and L-distance as hyper-parameters. Our preliminary results show that a targeted attack scheme based on a descending order of reachability index can compare similarly or better to well-known efficient strategies as shown in Figure 4. Here, the penalization parameter β is set to 0.2 and the locality parameter L is set to 4.

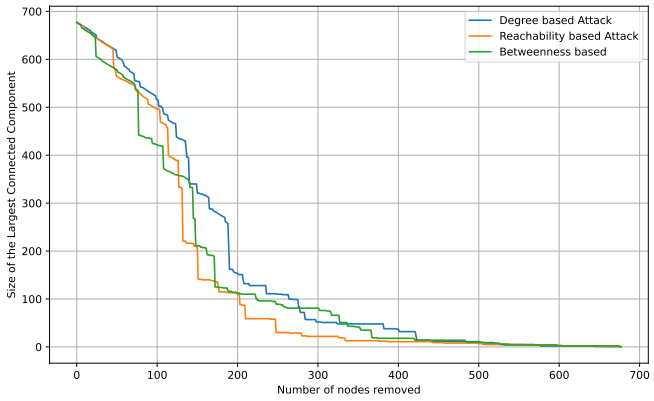


Figure 4: Size of the largest component with different attack strategies. $\beta = 0.2$ and $L = 4$.

More exhaustive experiments are needed to better evaluate the proposed framework. We believe our methodology has the potential to translate information from random perturbations into fine-grained insights about localized vulnerabilities while maintaining comparable performances with well known efficient attack strategies.

Appendix

Nodes	Edges	Density	Avg. deg.	Clust.
677	1891	0.0082	5.58	0.52

Table 2: Summary of the yeast propagation network.

References

[1] Oriol Artime, Marco Grassia, Manlio De Domenico, James P Gleeson, Hernán A Makse, Giuseppe Mangioni, Matjaž Perc, and Filippo Radicchi. Robustness and resilience of complex networks. *Nature Reviews Physics*, 6(2):114–131, 2024.

[2] Pascal Degenne and D Lo Seen. Ocelet: Simulating processes of landscape changes using interaction graphs. *SoftwareX*, 5:89–95, 2016.

[3] Scott Freitas, Diyi Yang, Srijan Kumar, Hanghang Tong, and Duen Horng Chau. Graph vulnerability and robustness: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 35(6):5915–5934, 2022.

[4] Sergey Ivanov and Evgeny Burnaev. Anonymous walk embeddings. In *International conference on machine learning*, pages 2186–2195. PMLR, 2018.

[5] Silvio Micali and Zeyuan Allen Zhu. Reconstructing markov processes from independent and anonymous experiments. *Discrete Applied Mathematics*, 200:108–122, 2016.

[6] Giannis Moutsinas and Weisi Guo. Node-level resilience loss in dynamic complex networks. *Scientific reports*, 10(1):3599, 2020.

[7] Mark Newman. *Networks*. Oxford university press, 2018.

[8] Nino Shervashidze, SVN Vishwanathan, Tobias Petri, Kurt Mehlhorn, and Karsten Borgwardt. Efficient graphlet kernels for large graph comparison. In *Artificial intelligence and statistics*, pages 488–495. PMLR, 2009.