Article

Not peer-reviewed version

Random Walks Based Node Centralities to Attack Complex Networks

Massimiliano Turchetto , Michele Bellingeri * , Roberto Alfieri , Ngoc-Kim-Khanh Khan , Quang Nguyen , Davide Cassi

Posted Date: 1 November 2023

doi: 10.20944/preprints202310.2066.v1

Keywords: real-world networks; node centrality; random walk processes; network robustness; network random walks

Article

Random Walks Based Node Centralities to Attack Complex Networks

Massimiliano Turchetto [1,2], Michele Bellingeri [1,2,*], Roberto Alfieri [1,2], Ngoc-Kim-Khanh Nguyen [3], Quang Nguyen [4,5] and Davide Cassi [1,2]

[1] Dipartimento di Scienze Matematiche, Fisiche e Informatiche, Università di Parma, via G.P. Usberti, 7/a, 43124 Parma, Italy
[2] INFN, Gruppo Collegato di Parma, I-43124 Parma, Italia
[3] Faculty of Basic Science, Van Lang University, Ho Chi Minh City, Vietnam
[4] Department of Physics, International University, Linh Trung, Thu Duc, Ho Chi Minh City, Vietnam
[5] Vietnam National University Ho Chi Minh City, Linh Trung, Thu Duc, Ho Chi Minh City, Vietnam
* Correspondence: michele.bellingeri@unipr.it

Abstract: Investigating the network response to node removal and the efficacy of the node removal strategies are related and fundamental questions in network science. Research studies proposed many node centralities based on the network structure ranking node to remove. The random walk (RW) on networks describes a stochastic process in which a walker travels among nodes. RW can be a model of transport, diffusion, and search on networks, and an essential tool for studying the importance of network nodes. In this manuscript, we propose four new measures of node centrality based on RW. Then, we compare the efficacy of the new RW node removal strategies to network dismantle with effective node removal strategies from the literature, such as betweenness and closeness node removal over synthetic and real-world networks. We evaluate the network dismantle along node removal using the size of the largest connected component (LCC). We find that, hence the betweenness nodes attack is the best strategy overall, the new node removal strategies based on RW show the highest efficacy in peculiar network topology. Specifically, RW strategies based on covering times emerge as the most effective strategy on a synthetic lattice network and two real-world road networks. Our results may be useful in selecting the best node attack strategies in a specific class of networks and in building more robust network structures.

Keywords: real-world networks; node centrality; random walk processes; network robustness; network random walks

1. Introduction

Numerous studies have been conducted in the last years to explore the response of real-world networks to the removal of nodes [1–7]. These investigations simulate the consequences of node removal (attack) on the network and have applications in diverse scientific fields such as ecology [5], transportation [8], informatics [9], neural [10,11], and social networks [12,13].

The main objectives of these studies have been twofold. Firstly, they aim to assess networks' robustness, which measures the system's ability to maintain functionality after link and node removal. Secondly, they seek to identify the link and node removals that cause the most significant damage to the network, thereby uncovering the key players that significantly influence network functioning.

Analyzing attack strategies provides valuable insights into enhancing network resilience by anticipating threats and identifying elements requiring protection [5,6].

An attack strategy refers to the identification and implementation of methods or techniques aiming at disrupting or dismantling a network [5–7]. It also plays a crucial role in situations where network disruption is necessary, such as halting the spread of a disease or a computer virus or impeding the growth of a cancer cell [14–16]. Many centralities' measurements have been proposed to select important nodes to remove. See [17] for a summary. Methods to measure node centralities

 doi:10.20944/preprints202310.2066.v1

are generally based on the topological structure of the network, such as removing nodes accounting for their degree and betweenness [5,17,18]. The betweenness node removal strategy, which removes nodes according to their recalculated betweenness centrality, yields the best attack in 70-80% of the cases [17].

The random walk (RW) on networks describes a stochastic process in which a walker travels among nodes along network links [19,20]. RW can be a model of transport, diffusion, and search on networks [21,22], a handy tool for studying the structure of networks [19], and the importance of network nodes [23–25].

In this manuscript, we join network attack simulation and random walk processes on networks. Here, we propose four new measures of node centrality based on RW. The new removal strategies focus on important notions in RW walks theory, such as the covering time, start, and stop nodes. Then, we test the proposed node centralities as effective strategies to rank nodes to remove to dismantle the network over synthetic and real-world networks. We compare the efficacy of the new node removal strategies based on random walks with effective node removal strategies from the literature, such as betweenness and closeness node removal.

2. Methods

2.1. Basic Notions

In this work, we consider binary and undirected networks $G(V,E)$ where V and E are the sets of nodes (vertices) and links (edges). $N = |V|$ indicates the number of nodes. $L = |E|$ indicates the number of edges. We assume G to be undirected. The symbol A denotes the $N \times N$ adjacency matrix of G, having entries a_{ij}, for $i,j = 1,\dots,N$, such that $a_{ij} = 1$ if $(i,j) \in E$, and $a_{ij} = 0$ otherwise (27). A '*path*' between two nodes u and w is a sequence of nodes $\langle v_1,\dots,v_k\rangle$ with $v_1 = u$ and $v_k = w$, such that $(v_i, v_{i+1}) \in E$, for $i = 1,\dots,k-1$. The length of the path equals the number of edges it contains. The '*distance*' d_{ij} is the shortest path length between node i and j [44]. In this work, all considered networks are connected, i.e., a path exists between each pair of nodes in V (28).

2.2. Synthetic Networks

1. **ER**: classical Erdös-Rényi (ER) random graph [26]. In the ER model, each edge has a fixed probability of being present or absent, independently of the other edges. The ER graph is defined by two parameters only, the number of nodes N and the probability of drawn links p. We indicate $ER(N,p)$ of N nodes, and probability p of link between each pair of vertices. We investigate ER network with $N = 80$ and $p = 0.15$.
2. **LTC**: rectangular (or square) lattice (LTC) complex network. In graph theory, a lattice graph is called a mesh or grid graph. The LTC is a specific lattice graph where nodes form a grid with square meshes. The LTC can be defined with two parameters, x, and y, indicating the number of nodes along each side. We simulate two $LTC(x,y)$ networks by choosing $x = 20$ and $y \in 5{,}20$ [27].
3. **BBT**: balanced binary tree [28–30]. A balanced binary tree is a tree data structure in which the difference in height between the left and right subtrees of any node is at most one. A reduced version (for space constraints) with 50 nodes is shown in Figure 1.

For statistical relevance of results obtained on ER random graphs, we performed 10^3 graph generations.

Figure 1 depicts examples of the synthetic networks used in this research.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2023 doi:10.20944/preprints202310.2066.v1

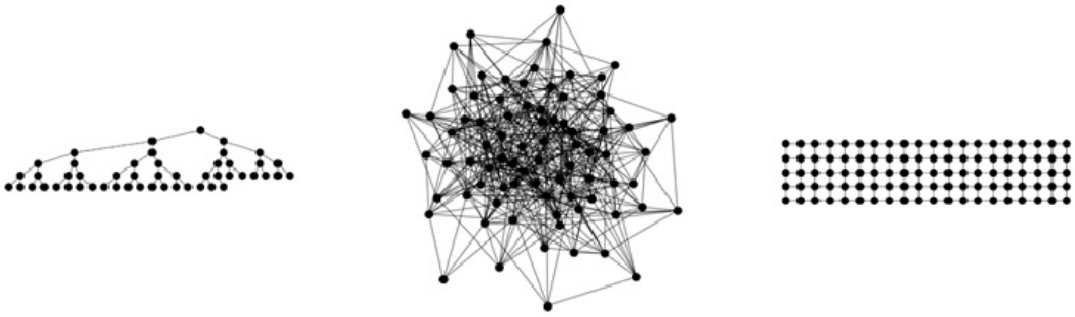

Figure 1. The picture displays examples of the synthetic networks used in this study. From left to right, the reported networks are: BBT (50 nodes), ER(N=80, p=0.15) and LTC(20,5).

2.3. Real-World Complex Networks

1. **Air Control**: This network was constructed from the USA's FAA (Federal Aviation Administration) National Flight Data Center (NFDC), Preferred Routes Database (Preferred Routes Database: http://www.fly.faa.gov/). Nodes in this network represent airports or service centers, and links are created from strings of preferred routes recommended by the NFDC [32].
2. **Arenas Email**: email communications among people working within a medium-sized university (i.e., Universitat Rovira i Virgily, Spain) with about 1700 employees [21]. Nodes are employees, and links describe mailing among them.
3. **Barcelona Flow**: models the traffic flow in Barcelona (Spain). Nodes represent intersections among roads, and links represent roads (Transportation Networks, https://github.com/bstabler/TransportationNetworks).
4. **Uk Faculty**: personal friendship network within a faculty at a university in the UK. This network comprises 81 vertices representing individuals and edges representing their friendship relations [34].
5. **Netscience**: a coauthorship network focusing on scientists involved in network science. The network represents collaborations among these scientists [25]. Nodes are scientists, and links depict the coauthorship in scientific papers.
6. represents the second ring road of Beijing city, China's capital. Nodes and links represent road intersections and roads, respectively [35].
7. **Beijing 3^{rd}**: represents the third ring road of Beijing city, China's capital. Nodes and links represent road intersections and roads, respectively [35].
8. **Beijing 4^{th}**: represents the fourth ring road of Beijing city, China's capital. Nodes and links represent road intersections and roads, respectively [35].
9. **Beijing 5^{th}**: represents the fifth ring road of Beijing city, China's capital. Nodes and links represent road intersections and roads, respectively [35].
10. **Euroroad**: a topological representation of international European roads in which nodes represent intersections among roads and links represent roads [36].
11. **Littlerock food-web**: a model of trophic interactions among species of the Little Rock Lake ecosystem in Wisconsin. In this ecological network, nodes represent living species, and links represent the transfer of nutrients between them [37].
12. **Olocene**: the Olocene food web ecological network is the basis of the 48 million years old uppermost early Eocene Messel Shale food web. Nodes are biological species, and links represent trophic relationships among them [38].
13. **San-Francisco Reduced**: represents a reduced version of the San Francisco road network [31] (Real Datasets for Spatial Databases, https://users.cs.utah.edu/~lifeifei/SpatialDataset.htm), obtained by applying a simple spatial-partitioning algorithm, resulting in a smaller, computationally affordable graph for the scope of this work.
14. **Road Minnesota**: the road map of Minnesota (US) [40]. Nodes represent intersections among roads, and links represent roads.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2023 doi:10.20944/preprints202310.2066.v1

15. **San Joaquin County**: California (US) city road map [31](Real Datasets for Spatial Databases, https://users.cs.utah.edu/~lifeifei/SpatialDataset.htm). Nodes are the intersections among roads, and links represent roads.

2.4. Network Structural Indicators

In Table 1, we report network structural indicators useful to compare the structure of the networks considered in this work. The network diameter $Diam$ is the maximum length among all shortest paths between each pair of nodes (44); the average node degree is the average number of links to the node $\overline{k}$ [41]; the average clustering coefficient CC is the number of closed triplets (or triangles) over the total number of triplets (both open and closed) [42,43]; the average node distance $\overline{\delta}$ is the average length of the shortest path among node pairs (44); and the network density (or connectance) ρ, i.e., the fraction of realized edges among all possible edges that can be drawn in the network [44,45].

Table 1. Network structural indicator values for the synthetic and real-world networks analyzed.

Network	$\|V\|$	$\|E\|$	$Diam$	$\overline{k}$	$\overline{\delta}$	CC	ρ
air-control	1226	2410	17	3.931	5.924	0.064	0.003
arenas-email	1133	5451	8	9.622	3.603	0.166	0.009
barcelona-flow	930	1798	27	3.867	12.721	0.084	0.004
beijing-2th	144	233	19	3.236	7.813	0.011	0.023
beijing-3rd	322	544	27	3.379	11.030	0.018	0.011
beijing-4th	547	926	33	3.386	13.904	0.019	0.006
beijing-5th	815	1308	48	3.210	17.246	0.024	0.004
euroroad	1039	1305	62	2.512	18.377	0.035	0.002
littlerock-foodweb	183	2452	4	26.798	2.135	0.332	0.147
netscience	379	914	17	4.823	6.026	0.431	0.013
olocene-foodweb	700	6425	6	18.357	2.629	0.074	0.026
road-minnesota	2641	3303	100	2.501	35.349	0.028	0.001
san-francisco-reduced	435	440	41	2.023	17.461	0.000	0.005
san-joaquin-county	7087	9793	50	2.764	13.939	0.000	0.000
uk-faculty	81	577	4	14.247	2.072	0.473	0.178
LTC(20,5)	100	175	23	3.500	8.250	0.000	0.035
LTC(20,20)	400	760	38	3.800	13.300	0.000	0.010
BBT	100	99	12	1.980	7.654	0.000	0.020
ER(N=80,p=0.15)	80.0	474.52	3.1	11.863	1.969	0.148	0.150

2.5. Node Removal Strategies

Node Removal (NR), also called node attack [46,47], refers to the process of selectively removing nodes from a network to study the impact on the structural properties of the network [13]. The removal strategy refers to how nodes are chosen to be removed from the network by assigning a value to each node and then defining an order to perform NR.

We define a series of RW-based node NR strategies and investigate their effectiveness in dismantling the networks. We compare their efficacy against two well-known centrality measures from the literature: closeness and betweenness node removals. We quantify the network dismantling after NR using the largest connected component (LCC) size. Node centrality rank is computed at the beginning of the simulation, i.e., before the first node removal. NR is performed by following the order of node centrality and computing the LCC after each removal. In the case of ties, i.e., nodes with equal centrality values, we randomly sort the nodes. The node centralities and the simulation analyses are performed using the complex network analysis (CNA) library Graph Tool (Tiago P. Peixoto), which consists of Python bindings for C++ and is highly performant as it is based on the Boost Graph Library [49].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2023 doi:10.20944/preprints202310.2066.v1

In the following, we define the NR strategies used in this work.

2.6. Betweenness Centrality

The betweenness centrality of a node $v \in V$ is defined as:

$$btw(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from s to t, and $\sigma_{st}(v)$ is the number of shortest paths from s to t that pass through node v [50].

2.7. Closeness Centrality

The closeness centrality of a node $v \in V$ is defined as

$$cls(v) = \frac{1}{\sum_{u \neq v} d_{u,v}}$$

where $d_{u,v}$ represents the distance between node u and node v in the network [51].

2.8. Random Walk Based Strategies

A simple Random Walk (RW) on G is a graph traversal in which an agent moves from node u to node v, such that v is chosen with uniform probability among the (first) neighbors of u [52]. Formally, the probability of transition from u to v can be defined as [53]:

$$p_{uv} = \frac{A_{uv}}{\sum_{w \in \tau} A_{uw}}$$

where τ is the neighbor's node set of u, and A_{uv} is the element of the adjacency matrix of G. The walk ends when all vertices have been visited at least once. We call the vertex from which the walk starts the 'start node'. For statistical relevance of analysis, we averaged results from 10^3 RWs for each start node $v \in V$. In the following part of this section, we define four RW-based strategies to perform node removals.

2.9. Recurrence Number

The Recurrence Number (RN) of a node is the number of times a random walker passes through the node during the covering process. Since the random walker covers all graph nodes, the simulation stops with a vector of RNs, one for each node. We call this vector of length $|V|$, the Recurrence Vector (RV), and each RN is > 0. In this node attack strategy, we remove nodes in decreasing order of RN.

2.10. Stop Node

The Stop Node (SN) is the last node encountered by a RW, or in other terms, the node where the RW stops its travel. We call Stop Vector (SV), the vector of length $|V|$ in which the entry i accounts for how many times the node i acted as a SN. Since we iterate 10^3 RW simulations, the sum of the SV entries is 10^3. In this node attack strategy, we remove nodes in ascending order of SN.

2.11. Cover Time

Given a vertex $v \in V$, we call the time step the action of passing from v to a (randomly chosen) neighbor. The notion of cover time refers to the number of time steps needed to visit all graph nodes [54]. We call Cover Time Vector (CTV), the vector of length $|V|$ in which the entry i accounts the CT when i is the starting node. The CTV accounts for each source node, the corresponding CT. In this node attack strategy, we remove nodes in decreasing order of CT, i.e., starting nodes producing higher CT are removed first.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2023 doi:10.20944/preprints202310.2066.v1

2.11. Stop Distance

Given a random walk on G, the Stop Distance (SD) is the distance $d_{s,t}$, for $s, t \in V$, where s and t are, respectively, the start and the stop node of the random walk. We call Stop Distance Vector (SDV) the vector of length $|V|$ in which the entry i accounts for the SD when i is the starting node. The SDV stores the corresponding SD for each source node. In this node attack strategy, we remove nodes in ascending order of SD, i.e., starting nodes near the stop node are removed first.

See Algorithm 1 for an explanation of the RW simulation analysis.

Algorithm 1: Methodology of the RW analysis.

```
RW(G(V,E), start_node):
    rec_number[v] ← 0, ∀ v ∈ V
    rec_number[start_node] ← 1
    cov_time ← 1
    stop_node ← start_node
    v ← start_node
    while ∃ x ∈ V | rec_num[x] == 0 do
            u ← randomly chose a neighbor of v
            rec_num[u] ← rec_num[u] + 1
            stop_node ← u
            cov_time ← cov_time + 1
            v ← u
    end while
    stop_distance← d(start_node, stop_node)
```

2.12. Network Robustness Indicator

2.12.1. Largest Connected Component

The Largest Connected Component (LCC), also called the giant component [29], indicates the connected subgraph of G having the largest set of nodes. In literature, it has often been used as a network robustness indicator to evaluate the effectiveness of node or link removal strategies [55] [56] [14] by observing the decreasing trends of LCC after such removals.

2.12.2. Robustness

The robustness value R represents the area under the curve of a decreasing trend of LCC [57] [17] [55]. The lower R, the higher the efficacy of a NR to dismantle the network. On the other hand, the higher R, the lower the efficacy of a NR to dismantle the network. For sake of clarity, we also define the inverse of robustness R^{-1}. In this manner, higher R^{-1} values denote more effective NR strategies.

Furthermore, given a fixed network, this value is normalized by the maximum value obtained among all NR strategies. This procedure allows us to compare the different robustness values obtained on a network while varying the different strategies. Additionally, given a fixed strategy, we denote R^{-1}_{avg} as the average value of R^{-1} obtained across all networks, allowing us to rank the average performance of each NR strategy across all networks.

In Table 2, we furnish a list of abbreviations used in this manuscript.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2023 doi:10.20944/preprints202310.2066.v1

Table 2. List of the abbreviations used in this manuscript.

LCC	network largest connected component
$\|\boldsymbol{V}\|$	number of nodes in the network
$\|\boldsymbol{E}\|$	Number of links in the network
Diam	Diameter of the network
$\bar{k}$	Average node degree
$\bar{\delta}$	Average length of shortest path among all node pairs
$\boldsymbol{CC}$	Clustering coefficient, i.e., number of closed triples
$\boldsymbol{\rho}$	Network density, i.e., fraction of realized links in the network among all possible links
R	Robustness of the network
R^{-1}	Inverse of the network robustness
R_{avg}^{-1}	Average inverse robustness R^{-1} among all networks
RN	Recurrence Number
CT	Covering Time
SN	Stop Node
SD	Stop Distance
BTW	Betweenness Centrality
CLS	Closeness Centrality

3. Results and Discussion

In this study, we simulated random walk processes to cover the networks and evaluate node importance. We introduced four node attack strategies based on the simulated random walks process to assign each node a ranking (a value or score). Subsequently, we utilized these scores to define new node centrality measures. The introduced strategies include recurrence number, stop node, stop distance, and covering time. Then, by attacking 19 networks, four of which are synthetic, and the rest are real-world networks, we compared the efficacy of dismantling the network of the new node centralities with two well-known competitors from literature, namely Betweenness (BTW) and Closeness (CLS) node removals.

In Figure 2, we show the LCC decrease as a function of the node removal fraction for real-world networks and in Figure 3 for the synthetic networks. Figure 4 displays the inverse of robustness, R^{-1}, normalized per row (i.e., per network), where each cell in the table is assigned a darker color as the strategy becomes more effective than the others. We report the average inverse robustness value across all networks in the last row.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2023 doi:10.20944/preprints202310.2066.v1

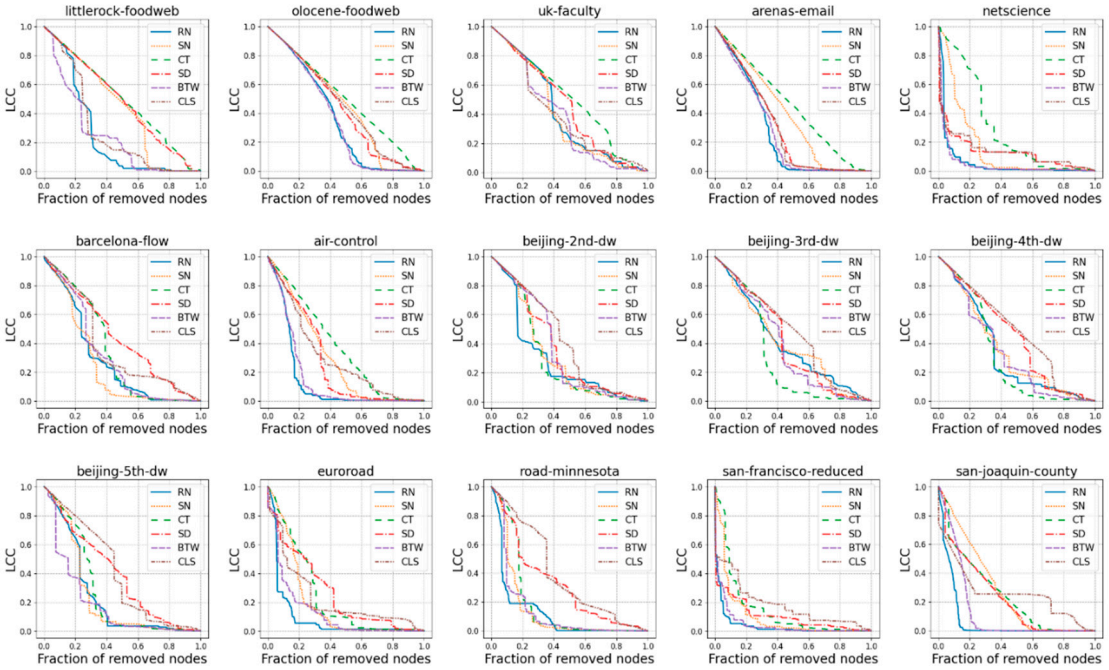

Figure 2. Impact of NR strategies on real-world networks: LCC (y-axis) as a function of the fraction of removed nodes (x-axis).

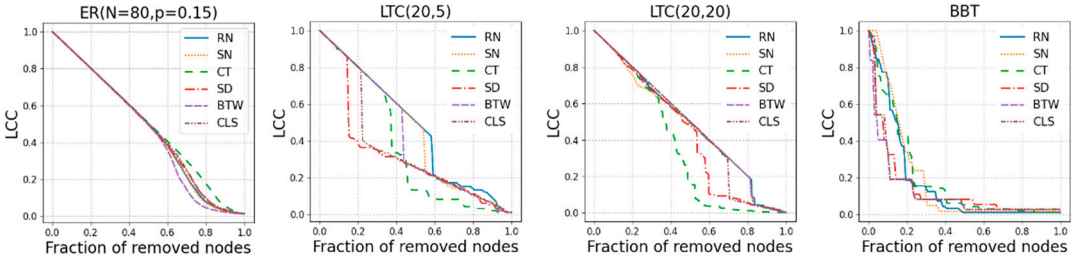

Figure 3. Impact of NR strategies on synthetic networks: LCC (y-axis) as a function of the fraction of removed nodes (x-axis).

	RN	SN	CT	SD	BTW	CLS
littlerock-foodweb	0.958	0.56	0.488	0.504	1.0	0.851
olocene-foodweb	0.98	0.78	0.692	0.807	1.0	0.772
uk-faculty	0.932	0.963	0.75	0.843	1.0	0.962
arenas-email	1.0	0.668	0.566	0.864	0.977	0.87
netscience	0.846	0.299	0.161	0.375	1.0	0.352
barcelona-flow	0.878	1.0	0.712	0.556	0.798	0.65
air-control	1.0	0.469	0.371	0.522	0.893	0.463
beijing-2nd-dw	1.0	0.926	0.972	0.798	0.807	0.699
beijing-3rd-dw	0.759	0.762	1.0	0.75	0.795	0.646
beijing-4th-dw	0.924	0.883	1.0	0.702	0.928	0.647
beijing-5th-dw	0.774	0.792	0.696	0.481	1.0	0.46
euroroad	1.0	0.492	0.388	0.351	0.721	0.406
road-minnesota	1.0	0.838	0.659	0.372	0.86	0.307
san-francisco-reduced	1.0	0.305	0.211	0.297	0.984	0.264
san-joaquin-county	1.0	0.255	0.28	0.304	0.715	0.266
ER(N=80,p=0.15)	0.968	0.972	0.922	0.963	1.0	0.953
LTC(20,5)	0.685	0.71	0.876	1.0	0.757	0.909
LTC(20,20)	0.728	0.742	1.0	0.832	0.723	0.756
BBT	0.271	0.377	0.235	0.85	1.0	0.905

	RN	SN	CT	SD	BTW	CLS
avg	0.96	0.767	0.683	0.726	1.0	0.712
overall ranking	**2**	**3**	**6**	**4**	**1**	**5**

Figure 4. Inverse network robustness R^{-1} for each network analyzed. To compare the efficacy of the node attack strategies, we normalize R^{-1} with its maximum value for each network. In this way, the maximum R^{-1} for each network equals 1. The higher R^{-1}, the more effective is the attack strategies

 doi:10.20944/preprints202310.2066.v1

to dismantle the network. In the last row, we depict the average R^{-1} value for all networks. Darker cell color indicates higher R^{-1}.

Moreover, in Figures A1 and A2 in the Appendix, we furnish the scatterplots of the random walk based node centralities *vs.* the betweenness node centrality for the real-world networks, and in Figures A3 and A4 in the Appendix, we depict the scatterplots of the random walk based node centralities *vs.* the node degree centrality for the real-world networks.

In the following, we summarize and discuss the outcomes for each NR strategy.

BTW: Our results show that the well-known betweenness nodes attack (BTW) is the best strategy overall as $R_{avg}^{-1} = 1$ (Figure 4). BTW was the most effective on both food webs, Uk Faculty, Arenas email, and Beijing 5^{th}. It has also achieved good results on synthetic networks, particularly ER networks and BBT. The performance of BTW remains quite good because $R^{-1} > 0.7$ for all other networks. These results confirm previous studies indicating that betweenness is a very effective strategy for dismantling complex networks [17] [5].

CLS: The closeness nodes attack (CLS) performs poorly on most road maps, while it is particularly effective on Uk Faculty, Little Rock Food-web, and Arenas Email regarding real-world networks. Regarding synthetic networks, it exhibits fairly good performance overall, especially on ER, BBT and LTC(20,20). CLS ranks fifth among the examined strategies as $R_{avg}^{-1} \approx 0.71$.

SN: The Stop Node (SN) has notable effectiveness on the ER random graph. Regarding real-world networks, SN is the most effective on Barcelona Flow and the second most effective on UK Faculty. It also demonstrates good effectiveness on Beijing 2^{nd} and 4^{th}, as well as on Road Minnesota. SN is the third strategy regarding average effectiveness, with $R_{avg}^{-1} \approx 0.76$.

We defined a 'stop node', the node where the RW stops its travel. For this reason, nodes acting many times as stop nodes are likely to be peripheral nodes, with a very low probability of encountering an RW. On the contrary, nodes that never (or rarely) acted as a stop node are likely to be central in the network and encounter an RW. The SN strategy removes nodes in ascending order of stop node, thus removing first the central nodes.

CT: The cover time (CT) emerges as the most effective strategy on LTC(20,20), and two real-world networks, Beijing 3^{rd} and 4^{th}. It also exhibits noteworthy effectiveness on Beijing 2^{nd} and ER. CT ranks last in terms of average effectiveness, with $R_{avg}^{-1} \approx 0.68$. The covering time is the number of time steps the RW needs to pass over all nodes in the network [54]. The CT node attack strategy removes nodes in decreasing order of their covering time when they are the start node. This way, start nodes producing higher covering times are removed first. The CT strategy returns peculiar results; on the one hand, CT shows the worst average efficacy (lowest R_{avg}^{-1}); on the other hand, it carried out the best performance in dismantling one synthetic and two real-world networks. The synthetic network is the square grid LTC, i.e., the model network with a planar structure and highly homogeneous node degree. In Figure 5 we depict the twenty most central nodes for each node removal strategy for the LTC networks of different size. The twenty most central nodes selected by the CT strategy are distributed over the entire network, whereas for all the other strategies, the most central nodes reside in a central part of the network. Therefore, for example, if we remove the highest BTW nodes from the LTC network, it will survive a large LCC composed by the peripheral nodes of the network (See Figure 5). In other terms, CT selects nodes covering the whole network structure, and for this, removing nodes according to the CT strategy may cause a faster LCC dismantling.

The two real-world networks where CT is highly effective are the road networks of the Beijing ring. Further, CT performs well in dismantling the Minnesota road networks (R^{-1}>0.85, Figure 4). These road networks show a planar-like structure and a narrow range of node degrees (see Figure 6). Therefore, it emerges an interesting ability of the CT node attack strategies to dismantle the networks with the specific characteristics of the planar-like structure and homogeneous node degree. In Figure 7 we depict the fifty most central nodes for each node removal strategy for the Minnesota and the Beijing 3^{rd} road networks. Like what was observed for the LTC, the fifty most central nodes, according to the CT strategy, are distributed over the entire network. In contrast, for all the other strategies, most central nodes reside in a part of the network. Therefore, this CT-specific node rank

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2023 doi:10.20944/preprints202310.2066.v1

property may result in the effective dismantling of real-world networks with a planar-like structure and homogeneous node degree, such as road networks.

SD: The stop distance (SD) performs well on synthetic graphs, particularly on LTC(20,5), proving the most effective strategy. As for real networks, it demonstrates a solid performance on Olocene food-web, UK Faculty, Arenas email, and Beijing 2^{nd}. SD is the fourth strategy in terms of average effectiveness, with $R_{avg}^{-1} = 0.72$. We defined the stop distance for a pair of nodes s and t, the shortest path length between the start node s and the stop node t of the random walk. SD attack strategy removes nodes in ascending order of SD.

For this reason, the SD remove first the start nodes that are a small distance from the respective stop node. This strategy emerges as particularly effective node removal over the synthetic network square lattice LTC of lower dimension. As we can see in Figure 5, the SD strategy can select nodes whose removals trigger the disruption of the LCC network in two parts. Therefore, removing nodes very near to their stop nodes can be a good method to dismantle this kind of model network and consequently select important nodes for its network robustness.

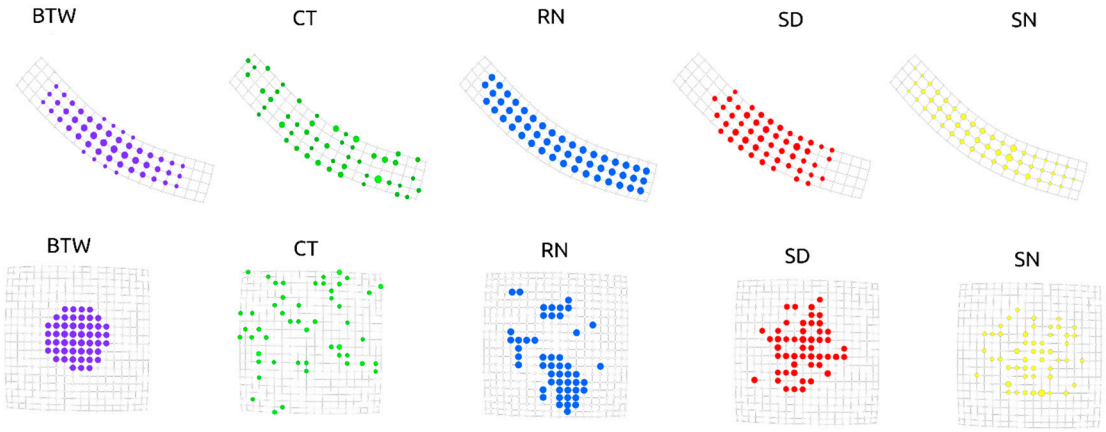

Figure 5. (Top row) the twenty most central nodes according to each node removal strategy for the LTC(20,5). (Bottom row) the twenty most central nodes according to each node removal strategy for the LTC(20,20).

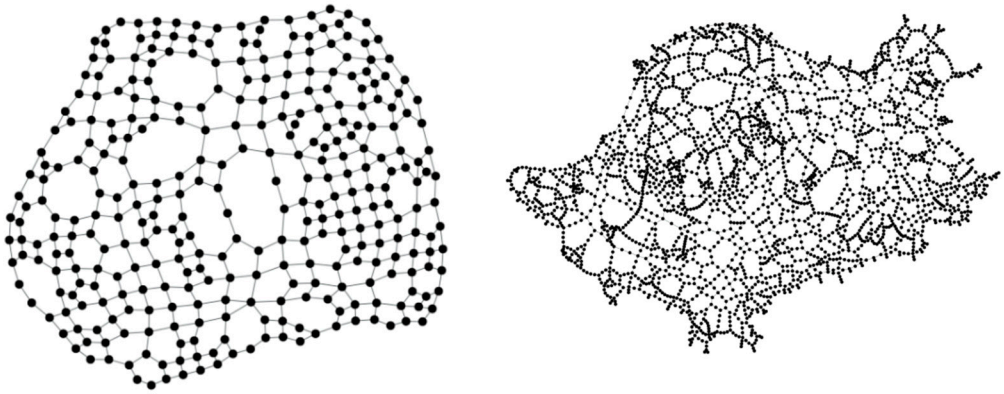

Figure 6. An illustrative example showcasing the structural approximation of a lattice by specific portions of a road map. The left network is the Beijing 3^{rd} ring road network. The right network is the Minnesota road network.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2023 doi:10.20944/preprints202310.2066.v1

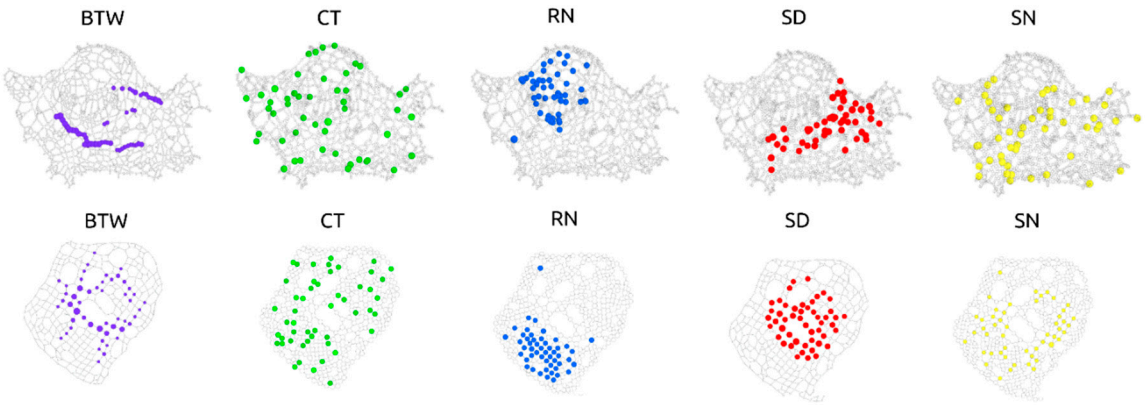

Figure 7. In this figure, we show the fifty most central nodes according to each node removal strategy. The upper row network is the Minnesota road network. The lower row network is the Beijing 3^{rd} ring road network.

RN: For a sufficiently large number of iterations, the recurrence number (RN) approximates very well the degree (See Figures A3 and A4 in the Appendix). As easily verified, the degree vector is the eigenvector of the transition matrix corresponding to the eigenvalue 1 (Perron-Frobenius eigenvector)[58]. Given this property, the RN is a degree-like node removal strategy and can be generally effective on most networks. Specifically, RN is the top strategy for San Francisco (reduced), San Joaquin County, and Beijing 2^{nd} road networks. Additionally, it maintains an average level of effectiveness R^{-1} greater than 0.75 on all other real networks. As for synthetic networks, it is less effective on lattices (LTC) where $R^{-1} < 0.75$ and ineffective on BBT. RN is the second most effective strategy among the tested networks with $R_{avg}^{-1} \approx 0.96$. Remove node based on their degree requires local information only, and for this, the node degree attack is a strategy with a low computational cost. The low computational cost and the good performance confirm this strategy as a good candidate for network dismantling.

4. Conclusions

Finding the best node attack strategy to dismantle the network is a paramount problem in network science [18] [5] [3] [17]. In this manuscript, we proposed four new node removal strategies based on a simulated random walk on the network and compared them with well-known strategies from the literature. The well-known node removal strategy based on the node betweenness resulted in the best strategy, on average, to dismantle the networks, confirming previous research [5]. Nonetheless, the random walk-based node removal proposed here presents peculiar and high effectiveness on specific networks. The CT strategy of removing nodes in decreasing order of the covering time they produce when they start node is highly effective in dismantling network planar-like structures and homogenous node degrees, such as road and square lattice networks. The methodology presented here can open future research. On the one hand, the node removal strategy proposed here can be helpful for another significant network science problem, such as finding the most influential spreader nodes in the network [59]. On the other hand, it will be interesting to investigate the efficacy of the random walk-based node attack strategies proposed here to lower other network robustness indicators, such as network efficiency [51].

A possible shortcoming of the proposed node removal strategies based on a simulated random walk can be the simulation cost. Nonetheless, we can say that dynamic processes based on random walks have become computationally more accessible than two decades ago. It is now possible to establish a series of statistically significant simulations using tools such as [48] that are adequately optimized for conducting small-scale simulations like the ones presented in this study. Our objective in further investigating these topics is to migrate our codes, making them suitable for harnessing parallel hardware in the HPC environment and enabling the simulation of such processes on large-

scale graphs. Moreover, we aim to introduce new strategies that facilitate the exploration of novel properties of real networks, which are often challenging to access solely through theoretical analysis.

Acknowledgments: The authors acknowledge the Italian Ministry of Foreign Affairs and International Cooperation. This research is funded by a grant from the Italian Ministry of Foreign Affairs and International Cooperation. This research is funded by Ecosister project, funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.5 - Call for tender No. 3277 of 30/12/2021 of Italian Ministry of University and Research funded by the European Union – NextGenerationEU Award Number: Project code ECS00000033, Concession Decree No. 1052 of 23/06/2022 adopted by the Italian Ministry. This work is supported by the Vietnam's Ministry of Science and Technology (MOST) under the Vietnam-Italy scientific and technological cooperation program for the period 2021-2023. This work is supported by the Vietnam National University Ho Chi Minh City (VNU-HCM), Ho Chi Minh city, Vietnam under grant number B2018-42-01. We are greatly thankful to Van Lang University, Vietnam, for providing the budget for this study. This research has benefited from the high-performance computing (HPC) cluster of the Università degli Studi di Parma.

References

1. Cohen, R.; Havlin, S. *Complex Networks: Structure, Robustness and Function*; Cambridge university press, 2010;
2. Cohen, R.; Erez, K.; Ben-Avraham, D.; Havlin, S. Resilience of the Internet to Random Breakdowns. *Phys Rev Lett* **2000**, *85*, 4626.
3. Morone, F.; Makse, H.A. Influence Maximization in Complex Networks through Optimal Percolation. *Nature* **2015**, *524*, 65–68.
4. Callaway, D.S.; Newman, M.E.J.; Strogatz, S.H.; Watts, D.J. Network Robustness and Fragility: Percolation on Random Graphs. *Phys Rev Lett* **2000**, *85*, 5468.
5. Bellingeri, M.; Cassi, D.; Vincenzi, S. Efficiency of Attack Strategies on Complex Model and Real-World Networks. *Physica A: Statistical Mechanics and its Applications* **2014**, *414*, 174–180.
6. Huang, X.; Gao, J.; Buldyrev, S. V; Havlin, S.; Stanley, H.E. Robustness of Interdependent Networks under Targeted Attack. *Phys Rev E* **2011**, *83*, 65101.
7. Nie, T.; Guo, Z.; Zhao, K.; Lu, Z.-M. New Attack Strategies for Complex Networks. *Physica A: Statistical Mechanics and its Applications* **2015**, *424*, 248–253.
8. Pagani, A.; Mosquera, G.; Alturki, A.; Johnson, S.; Jarvis, S.; Wilson, A.; Guo, W.; Varga, L. Resilience or Robustness: Identifying Topological Vulnerabilities in Rail Networks. *R Soc Open Sci* **2019**, *6*, 181301.
9. Cohen, R.; Erez, K.; Ben-Avraham, D.; Havlin, S. Breakdown of the Internet under Intentional Attack. *Phys Rev Lett* **2001**, *86*, 3682.
10. Bassett, D.S.; Bullmore, E.D. Small-World Brain Networks. *The neuroscientist* **2006**, *12*, 512–523.
11. Bullmore, E.; Sporns, O. Complex Brain Networks: Graph Theoretical Analysis of Structural and Functional Systems. *Nat Rev Neurosci* **2009**, *10*, 186–198.
12. Borgatti, S.P.; Mehra, A.; Brass, D.J.; Labianca, G. Network Analysis in the Social Sciences. *Science (1979)* **2009**, *323*, 892–895.
13. Boldi, P.; Rosa, M.; Vigna, S. Robustness of Social and Web Graphs to Node Removal. *Soc Netw Anal Min* **2013**, *3*, 829–842.
14. Sartori, F.; Turchetto, M.; Bellingeri, M.; Scotognella, F.; Alfieri, R.; Nguyen, N.-K.-K.; Le, T.-T.; Nguyen, Q.; Cassi, D. A Comparison of Node Vaccination Strategies to Halt SIR Epidemic Spreading in Real-World Complex Networks. *Sci Rep* **2022**, *12*, 21355.
15. Nguyen, N.-K.-K.; Nguyen, T.-T.; Nguyen, T.-A.; Sartori, F.; Turchetto, M.; Scotognella, F.; Alfieri, R.; Cassi, D.; Nguyen, Q.; Bellingeri, M. Effective Node Vaccination and Containing Strategies to Halt SIR Epidemic Spreading in Real-World Face-to-Face Contact Networks. In Proceedings of the 2022 RIVF International Conference on Computing and Communication Technologies (RIVF); 2022; pp. 1–6.
16. Michele Bellingeri Massimiliano Turchetto, D.B. Modeling the Consequences of Social Distancing over Epidemics Spreading in Complex Social Networks: From Link Removal Analysis to SARS-CoV-2 Prevention. *Front Phys* **2021**, *9*, 681343.
17. Wandelt, S.; Sun, X.; Feng, D.; Zanin, M.; Havlin, S. A Comparative Analysis of Approaches to Network-Dismantling. *Sci Rep* **2018**, *8*, 1–15.
18. Iyer, S.; Killingback, T.; Sundaram, B.; Wang, Z. Attack Robustness and Centrality of Complex Networks. *PLoS One* **2013**, *8*, e59613.
19. Burioni, R.; Cassi, D. Random Walks on Graphs: Ideas, Techniques and Results. *J Phys A Math Gen* **2005**, *38*, R45.
20. Masuda, N.; Porter, M.A.; Lambiotte, R. Random Walks and Diffusion on Networks. *Phys Rep* **2017**, *716*, 1–58.

21. Guimer, R.; Danon, L.; Diaz-Guilera, A.; Giralt, F.; Arenas, A. Self-Similar Community Structure in a Network of Human Interactions. *Phys. Rev. E* **2003**, *68*, 65103.
22. Agliari, E. Exact Mean First-Passage Time on the T-Graph. *Phys Rev E* **2008**, *77*, 11128.
23. Noh, J.D.; Rieger, H. Random Walks on Complex Networks. *Phys Rev Lett* **2004**, *92*, 118701.
24. Rocha, L.E.C.; Masuda, N. Random Walk Centrality for Temporal Networks. *New J Phys* **2014**, *16*, 63023.
25. Newman, M.E.J. Analysis of Weighted Networks. *Phys Rev E* **2004**, *70*, 56131.
26. RENYI, E. On Random Graph. *Publicationes Mathematicate* **1959**, *6*, 290–297.
27. Acharya, B.D.; Gill, M.K. On the Index of Gracefulness of a Graph and the Gracefulness of Two-Dimensional Square Lattice Graphs. *Indian J. Math* **1981**, *23*, 14.
28. Cormen, T.; Leiserson, C.; Rivest, R.; Stein, C. Book: Introduction to Algorithms 2009.
29. Van Steen, M. Graph Theory and Complex Networks. *An introduction* **2010**, *144*.
30. Chen, G.; Wang, X.; Li, X. Fundamentals of Complex Networks: Models, Structures and Dynamics. **2014**, 96.
31. Preferred Routes Database.
32. Kunegis, J. Konect: The Koblenz Network Collection. In Proceedings of the Proceedings of the 22nd international conference on world wide web; 2013; pp. 1343–1350.
33. Transportation Networks.
34. Nepusz, T.; Petróczi, A.; Négyessy, L.; Bazsó, F. Fuzzy Communities and the Concept of Bridgeness in Complex Networks. *Phys Rev E* **2008**, *77*, 16107.
35. Guo, X.-L.; Lu, Z.-M. Urban Road Network and Taxi Network Modeling Based on Complex Network Theory. *J. Inf. Hiding Multim. Signal Process.* **2016**, *7*, 558–568.
36. Šubelj, L.; Bajec, M. Robust Network Community Detection Using Balanced Propagation. *Eur Phys J B* **2011**, *81*, 353–362.
37. Martinez, N.D. Artifacts or Attributes? Effects of Resolution on the Little Rock Lake Food Web. *Ecol Monogr* **1991**, *61*, 367–392.
38. Dunne, J.A.; Labandeira, C.C.; Williams, R.J. Highly Resolved Early Eocene Food Webs Show Development of Modern Trophic Structure after the End-Cretaceous Extinction. *Proceedings of the Royal Society B: Biological Sciences* **2014**, *281*, 20133280.
39. Real Datasets for Spatial Databases.
40. Rossi, R.A.; Ahmed, N.K. The Network Data Repository with Interactive Graph Analytics and Visualization. In Proceedings of the AAAI; 2015.
41. Bellingeri, M.; Montepietra, D.; Cassi, D.; Scotognella, F. The Robustness of the Photosynthetic System I Energy Transfer Complex Network to Targeted Node Attack and Random Node Failure. *J Complex Netw* **2021**, *10*, cnab050, doi:10.1093/comnet/cnab050.
42. Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D.-U. Complex Networks: Structure and Dynamics. *Phys Rep* **2006**, *424*, 175–308.
43. Gleeson, J.P.; Melnik, S.; Hackett, A. How Clustering Affects the Bond Percolation Threshold in Complex Networks. *Phys Rev E* **2010**, *81*, 66114.
44. Dunne, J.A.; Williams, R.J.; Martinez, N.D. Network Structure and Biodiversity Loss in Food Webs: Robustness Increases with Connectance. *Ecol Lett* **2002**, *5*, 558–567.
45. Bellingeri, M.; Vincenzi, S. Robustness of Empirical Food Webs with Varying Consumer's Sensitivities to Loss of Resources. *J Theor Biol* **2013**, *333*, 18–26.
46. Nguyen, Q.; Pham, H.-D.; Cassi, D.; Bellingeri, M. Conditional Attack Strategy for Real-World Complex Networks. *Physica A: Statistical Mechanics and its Applications* **2019**, *530*, 121561.
47. Albert, R.; Barabási, A.-L. Statistical Mechanics of Complex Networks. *Rev Mod Phys* **2002**, *74*, 47.
48. Tiago P. Peixoto Graph-Tool, Efficient Network Analysis, Https://Graph-Tool.Skewed.De/.
49. Siek, J.G.; Lee, L.-Q.; Lumsdaine, A. *The Boost Graph Library: User Guide and Reference Manual, The*; Pearson Education, 2001;
50. Freeman, L.C. A Set of Measures of Centrality Based on Betweenness. *Sociometry* **1977**, 35–41.
51. Marchiori, M.; Latora, V. Harmony in the Small-World. *Physica A: Statistical Mechanics and its Applications* **2000**, *285*, 539–546.
52. Campari, R.; Cassi, D. Random Collisions on Branched Networks: How Simultaneous Diffusion Prevents Encounters in Inhomogeneous Structures. *Phys Rev E* **2012**, *86*, 21110.
53. Xia, F.; Liu, J.; Nie, H.; Fu, Y.; Wan, L.; Kong, X. Random Walks: A Review of Algorithms and Applications. *IEEE Trans Emerg Top Comput Intell* **2019**, *4*, 95–107.
54. Lovász, L. Random Walks on Graphs. *Combinatorics, Paul erdos is eighty* **1993**, *2*, 4.
55. Bellingeri, M.; Bevacqua, D.; Scotognella, F.; Cassi, D. The Heterogeneity in Link Weights May Decrease the Robustness of Real-World Complex Weighted Networks. *Sci Rep* **2019**, *9*, 10692.
56. Zhang, Y.; Ng, S.T. Identification and Quantification of Node Criticality through EWM–TOPSIS: A Study of Hong Kong's MTR System. *Urban Rail Transit* **2021**, *7*, 226–239.

 doi:10.20944/preprints202310.2066.v1

57. Schneider, C.M.; Moreira, A.A.; Andrade Jr, J.S.; Havlin, S.; Herrmann, H.J. Mitigation of Malicious Attacks on Networks. *Proceedings of the National Academy of Sciences* **2011**, *108*, 3838–3841.
58. Levin, D.A.; Peres, Y. *Markov Chains and Mixing Times*; American Mathematical Soc., 2017; Vol. 107;.
59. Kitsak, M.; Gallos, L.K.; Havlin, S.; Liljeros, F.; Muchnik, L.; Stanley, H.E.; Makse, H.A. Identification of Influential Spreaders in Complex Networks. *Nat Phys* **2010**, *6*, 888–893.