Tabriz Journal of Electrical Engineering (TJEE), vol. 55, no. 1, 2025 Serial no. 111
DOI: 10.22034/tjee.2024.59608.4777

Signed Social Network Vulnerability Analysis in Terms of Clustering coefficient and Balance Theory

Mansooreh Mirzaie[1*], Maryam Nooraei Abadeh[2]
[1]Department of Electrical and Computer Engineering, Golpayegan College of Engineering, Isfahan University of Technology, Golpayegan, Iran, m.mirzaei@iut.ac.ir
[2]Department of Computer Engineering, Abadan Branch, Islamic Azad University, Abadan, Iran, ma.nooraei@iau.ac.ir
* Corresponding Author

Received:15/12/2023, Revised:29/05/2024, Accepted: 01/08/2024.

Abstract. The robustness of a social network in response to unexpected events is still challenging for real-world networks. This paper addresses the challenge of evaluating social network robustness against unexpected events, particularly in real-world signed networks. We analyze the clustering vulnerability and balance of Signed Social Networks (SSNs) under the failure of important nodes. The main objective is to identify the critical nodes whose removal disrupts the network by weakening its clustering. It is evaluated by the Average Local Clustering Coefficient guided by the balance degree of the networks. To identify critical nodes, we propose parameter-based greedy strategies that remove nodes based on specific criteria. We conduct a comprehensive analysis of the real and synthetic SSNs generated by different well-known models and also online datasets. Our experiments demonstrate that removing a small percentage of nodes with the highest "Fans Minus Freaks (FMF)" value significantly reduces the network's clustering coefficient. Interestingly, centrality and PageRank metrics also play a role, but to a lesser extent, ranking second and third in terms of critical impact, respectively.

Keywords: Signed Social Networks, Clustering coefficient, Vulnerability, Network measures, Balance theory.

1. Introduction

The flexibility of a network against attacks and shortcomings has been a growing concern in recent years. The robustness may be one of the desirable features for complex networks such as social networks and tries to determine how the network performance is affected by external disturbances. In other words, robustness determines the flexibility of a network in response to unexpected events, including defensive attacks and stochastic defects [1]. Complex systems with an ability to maintain organizational structure, performance, and responsiveness under such unexpected disturbances are more robust [2]. The concept of vulnerability is generally used to understand and describe the lack of capability and flexibility of complex systems [3, 4]. The robustness of social networks and other complex systems depends on the network structure and how the network clusters are connected.

The difference between the signed and unsigned networks can be considered in two main aspects; one in terms of the complexity of the signs and another due to adding signs to the edges [5]. The signed edges provide a unique interpretation for the signed networks concerning the inherent characteristics of *SSNs* like the structural balance defined by balance theory. Moreover, the property of transitivity is defined in the unsigned networks, and a high value of local clustering in the form of triangles is significant. According to balance theory [6], some triangles are more seemingly formed in the balanced states than others (i.e., unbalanced) in the signed networks. Subsequently, analyzing the signed networks needs to capture both unique properties of signed networks, e.g., the distribution of structurally formed triangles. The interpretation of positive and negative signs is different from these settings. This is the challenge of utilizing theories of signed edges to evaluate how the failure of signed edges affects the clustering of an *SSN*.

Clustering is a fundamental feature in a network associated with a wide range of topics by considering the diffusion of information in a social network (e.g., rumor propagation) [7]. Local Clustering Coefficient (*LCC*) [8] evaluates node connectivity. A node has a high clustering coefficient if its' neighbors are directly connected. When nodes are removed from a network, some other nodes may be separated from their original clusters.

Finding a solution to this emerging problem is still a big challenge because (a) Average Local Clustering Coefficient (*ALCC*) behavior is not uniform in node elimination, and it is unpredictable even in response to a minor variation, and (b) most of the real social networks have large sizes; therefore, they do not have an exact computation and a unique solution. In this paper, we analyze the vulnerability of clustering, in particular for the signed social networks. We also reply to this question for an *SSN*, how can we find a set of vertices that maximally degrade the network-clustering coefficient? To answer this question, the clustering coefficient is evaluated by removing the nodes with the highest amount of *FMF* [9], centrality [10], and PageTrust measures as the important criteria in an *SSN* analysis to detect critical nodes concerning the vulnerability. To verify the results, balance theory, as an influential concept in the structural clusterability of a network, is evaluated under critical node failure. The main contributions of this paper are:

- Analyzing the complex *SSN* vulnerability using the clustering coefficient guided by the balance of clustering,
- Finding a subset of an *SSN* to evaluate the vulnerability of clustering, which maximizes the changes of the clustering coefficient by focusing on centrality measures, and
- Experimental analysis of the artificial and real networks under the failure of critical nodes.

The remainder of the paper is structured as follows. The problem statement is discussed in Section 2. The third section contains related work. Section 4 introduces the structural balancing theory of the signed graphs. The details of the suggested technique and the key parameters of the *SSN*s clustering coefficient measures are discussed in Section 5. Section 6 assesses the findings of specific experiments that used artificial and actual *SSN*s to identify crucial nodes in terms of vulnerability. We conclude in Section 7.

2. Balance theory and Problem statements

Balance and clusterability are widely studied in social networks, such as [11-18]. The concept of balancedness has been applied to some research areas of social sciences and many other fields. To verify the existence of stable clusters in a graph, the concept of separability of signed graphs and the clusterability issues, such as the local clustering coefficient distribution, and the triangle distribution of signed graphs, should be examined. The triangle theorem is the cornerstone of structural balance and explains why the signed networks have conflicts. In defining structural balance, Heider [6] stated that when all of the triangles in a generic signed network have an even number of negative edges, the network is balanced.

Balanced theory establishes a viewpoint on a triangle with three positive signs (a) and those with one positive sign (c) that are more plausible and common in real *SSNs* than triangles with two positive signs (b) or none (d). The balanced triangles with three positive edges serve as an example that “the friend of my friend is my friend”. In contrast, those with one positive and two negative edges capture the notions that “the enemy of my enemy is my friend”, “the friend of my enemy is my enemy”, and “the enemy of my friend is my enemy”. In other words, a signed graph is balanced in terms of clusterability if and only if there exists a divider to separate its nodes set into two subsets [۱۹]. One of them is perhaps empty, such that every positive edge connects two nodes in the same subset and every negative edge connects two nodes from different subsets. There is a simple algorithm to detect balance [20]. The goal states that all communications within each part are positive and any connection between two parts is negative. To study balance in real-world signed networks, a precise way is needed to quantify how balanced they are [21]. Based on the clustering coefficient in a network, we consider the parameters that quantify crucial measures in terms of vulnerability. These measures either determine the strong or the weak sense of balance, i.e., have an odd number of negative edges (strong balance) or precisely one negative edge (weak balance) [22].

To analyse the vulnerability of clustering in *SSNs*, the main challenge is finding how the behaviour of a network is changed when elements (nodes or edges), in particular elements with high-value centrality measures, are removed. Node attributes play an important role in shaping network structures. A node's fault, specifically in valuable nodes, could affect the network performance. For this purpose, the adjacency matrix of an *SSN* named *A* is considered *N* * *N*, in which *N* is the number of nodes. Each entry of *A* is a connection $(i,j) \in$ [23] to present negative edges, no edges, and positive edges between node *i-th* and *j-th* in graph *G'(V,E)*. The most important parameters, which are used in vulnerability analysis of *SSNs,* include the degree of centrality (called the *FMF* centrality measure), PageTrust, and the centrality-based measure. At the same time, the network balance values characterize the stability of the signed graph. A formal research question based on the definitions expressed in the clustering vulnerability of *SSNs* is: How can we find a subset $S^* \subseteq V$, $|S^*|=k$, $k\leq N$ to evaluate the vulnerability of clustering, which maximizes the changes of the clustering coefficient of *G'(V, E)* by focusing on the centrality measures?

3. Related works

The vulnerability analysis has sparked a lot of interest among network scientists. This research topic is divided into two categories. The first is evaluating the graph robustness by finding the critical nodes, and the second is manipulating a network's robustness from the network vulnerability perspective [24]. In the first category, different criteria and metrics were proposed, including the connectivity value of a graph [25], the typical size of the largest cluster of a graph and the corresponding size of the isolated cluster [26], centrality measures like betweenness and the geodesic length [27], eigenvector [28], a comparative approach between node shortest path [29], overall pair-wise connectivity [25] or new parameters to describe how positive and negative connections interact in a signed network [30].

To manipulate the robustness, various techniques have been provided, such as [26] and [31] or applying graph percolation [32]. Veremyev et al. [33]

and Veremyev et al. ([34] created frameworks for integer programming to find critical nodes that minimize a connectivity metric subjected to budgetary constraints. For additional network vulnerability analyses, the reader is motivated to study the surveys of [35] and [36].

The concept of the clustering coefficient was firstly presented in [8] and was extended to the positive weighted edges in [37]. The analysis of the vulnerability based on Average Local Clustering is rarely investigated [38], [39], and [40].

In a study [41], the authors analysed the vulnerability of community structure by an exploratory approach. They also discussed heuristic approaches to discover the critical nodes in a modularity-based graph structure. The authors in [42] discussed nodes whose failure critically harms the network by clustering reduction. Alim et al. [23] presented a technique to produce the community edges to determine the critical components. Ertem et al. [43] investigated how to detect node clusters in social networks with a high clustering coefficient; however, their work did not take into account the vulnerability analysis of a network considering the clustering coefficient. The authors in [44] studied the vulnerability of networks and its impact on the network performance.

Most of the available approaches for social network analysis consider unsigned networks [٤٥]; therefore, edges are considered unsigned or only with positive value [46]. In [47, 48] negative edges are used to explore communities (clusters); however, these edges are not used to measure popularity or similarity. In [9], several methods have been compared. These analyses focus mainly on defining global trust sizes using the path length or adapting PageRank [49] and [50].

According to the author's best knowledge, no research has been done on the vulnerability of signed social networks in which an edge’s sign reflects user’ positive or negative attitude toward another user. In this paper, we investigate the impact of removing critical nodes on the network clustering structure. Our main objective is to identify the critical nodes whose failure damages the network due to the weakening of the cluster. This is evaluated by the Average Local Clustering Coefficient guided by the balance degree of networks. We define the vulnerability of *SSN*s based on the clustering coefficient guided by the balance of the networks as an optimization problem. The approach finds critical nodes using proposed parameter-based greedy techniques while removing critical nodes of an *SSN* to analyse the clustering coefficient.

To the best knowledge of the authors, the strategy that we propose provides a number of advantages over previous studies. On the target dataset, our approach achieves significantly higher accuracy than the previously discussed methods. This overcomes the shortcomings of earlier techniques and offers a notable increase in performance. Compared to existing methods, our suggested method has a faster inference time and is computationally more efficient. This improved performance is significant for applications in real life where speed is an essential component. Moreover, compared to previous approaches, our method is more flexible and manages difficult situations (noisy inputs, changing environmental circumstances) better.

In comparison with the architecture of the model to the previous complex designs, it is much more straightforward. This simplification can result in less resource usage, faster training, and simpler interpretation. Beyond the topic that is addressed in this study, similar problem areas may benefit from the application of the suggested strategy. This versatility enhances the partnership's significance and total influence.

4. Proposed approach

In this section, we introduce the primary definitions of the proposed approach. Table 1 lists the symbols used in the problem definition. The proposed approach is shown in Fig. 1. As the input parameter of the proposed method, the adjacency matrix is utilized. The clustering coefficient is analyzed based on the normalized data matrix of the centrality criteria. The selected measures are defined in Section 5.1. At each step, the behavior of the clustering coefficient is captured by removing the highest elements concerning the selected measures. The objective functions are defined in Section 5.2 to find a maximum subset $S^* \subseteq V$, which maximizes the changes in the clustering coefficient. The pseudocode for the proposed approach is provided in the following pseudocode.

The pseudocode of the proposed approach

```
# Normalize data matrix
normalized_matrix = normalize(matrix)
# Define centrality measures
centrality_measures = [measure1, measure2, ...]
# Maximum change in clustering coefficient
while True:
    # Find element with highest value for current centrality measure
    max_index = find_max_index(normalized_matrix, centrality_measures[0])
    # Calculate change in clustering coefficient
    delta_cc = calculate_delta_cc(matrix, max_index)
    # Check if improvement
    if delta_cc <= max_delta_cc:
        break
    # Update variables
    S.append(max_index) max_delta_cc = delta_cc
    # Remove element from consideration
    normalized_matrix[max_index] = 0
# Return maximum subset
```

```
8.    return S
```

The method is extendable to different centrality measures. Furthermore, different multi-optimization techniques can be used for solving this problem in future work.

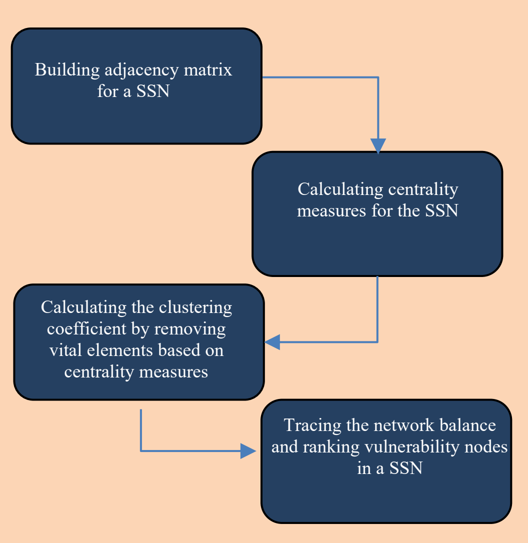

Building adjacency matrix for a SSN
Calculating centrality measures for the SSN
Calculating the clustering coefficient by removing vital elements based on centrality measures
Tracing the network balance and ranking vulnerability nodes in a SSN

Fig.1. The architecture of the proposed approach

Table 1: Notations

Descriptions	**Notations**
Signed graph	$G = (V, E^{+}, E^{-})$
Negative edge	E^{-}
Positive edge	E^{+}
Negative in-degrees	$d_{in}^{-}(v_i)$
Positive in-degrees	$d_{in}^{+}(v_i)$
Negative out-degrees	$d_{out}^{-}(v_i)$
Positive out-degrees	$d_{out}^{+}(v_i)$
The out-degree of positive links of u_j	O_j^{+}
A matrix full of ones of the specified size	$J_{n\times n}$
The teleportation parameter	$0 < \alpha < 1$
Unsigned adjacency matrix	$\bar{A}$
Adjacency matrix	A
The absolute diagonal degree matrix	$\bar{H}$
Represents the sign of the edge between i and j	s_{ij}
Degree of node u(unsigned)	k_i
Number of nodes	N
Local clustering coefficient of node u	$C(u)$
Sign local clustering coefficient of node u	$\hat{C}(u)$
Average local clustering coefficient of graph	$C(G)$
A positive integer for the number of subsets	k
The weighted edge between i , j	$z(i,j)$
Degree of balance	DB
The indegree of positive links of ui	$\lvert I_i^{+}\rvert$
The indegree of negative links of ui	$\lvert I_i^{-}\rvert$

4.1 Measures

In a social network, the centrality measure of each node is determined by a value. This value represents the node's importance in the network. For a signed network, a centrality measure involves the combination of two values and the interaction between positive and negative edges. The most common measures may have both positive and negative values. Positive and negative edges are consistent with the “friend” and “foe” relationships. The interplay between significant relationships and the inherent imbalance of positive and negative edges in real signed networks poses a significant challenge for determining centrality measures. A centrality measure for the signed social networks, which considers the degree of positive and negative nodes, FMF (Fans Minus Freaks), has been proposed by [51]. Other measures, such as PageRank and centrality-based measures [52] have also been considered to discuss the vulnerability of signed networks based on the proposed method in this paper. A modified version of PageRank, called PageTrust [53], is also considered.

Definition 1 Signed graph and corresponding adjacency matrix. We show a directed signed network by $G = (V, E^{+}, E^{-})$. In this network, each edge $(a, b) \in E^{+}$ means a trust b, and each edge $(a, b) \in E^{-}$ means a distrust b. The relevant adjacency matrix entry of a graph contains 0, -1, and 1 for unconnected nodes, negative, and positive edges between two nodes.

Definition 2 Positive and negative in/out degrees. $d_{in}^{+}(V)$ and $d_{in}^{-}(V)$ are used for the positive and negative in-degrees, and $d_{out}^{+}(V)$ and $d_{out}^{-}(V)$ stand for the corresponding out-degrees. This parameter can also be extended to neighbour sets as $N_{in}^{+}(V)$ and $N_{in}^{-}(V)$ applied for the positive and negative in-neighbor sets, and $N_{out}^{+}(V)$ and $N_{out}^{-}(V)$ are considered for the corresponding out-neighbor sets.

Definition 3 Degree of centrality (*FMF*). Nodes with positive edges are introduced as Fans, and those with negative edges are Freaks. Degree centrality could be defined by this simple generalization to the signed networks. More formally,

$$FMF(v_i) = d_{in}^{+}(v_i) - d_{in}^{-}(v_i) \tag{1}$$

Definition 4 PageRank-based Algorithm Parameters. PageRank is defined on directed graphs with non-negative edge weights [54]. It depicts the path of a random “surfer”, which follows the directed edges in a random manner and “teleporting” to a random node at chosen intervals. It is the dominant left eigenvector of the Google matrix G, which is given by Eq. (2).

$$G_s = (1-\alpha)\bar{H}^{-1}\bar{A} + (\alpha/n)J_{n\times n} \tag{2}$$

where, $J_{n\times n}$ is a matrix full of one with a specified size, and $0 < \alpha < 1$ represents the teleportation parameter. G is a left-stochastic matrix (it means that in G each row sums to one), and $\bar{H}$ is the definite diagonal degree matrix. PageRank only applies to non-negative edge weights; so, the unsigned adjacency matrix $\bar{A}$ must be used instead of A. Therefore, the final rank is not a sign of popularity but a sign of centrality, indicating the desire of users to be central without the recognition of friends and foe links. The original PageRank algorithm computes the credit rating for the *i-th* node:

$$Pr_i = \sum_{u_j \in I_i^+} \frac{Pr_j}{|O_j^+|} \tag{3}$$

where, $|O_j^+|$ is the out-degree of positive links of u_j. Pr_i could be computed iteratively by Eq. (4):

$$Pr_i^{t+1} = \alpha \sum_{u_j \in I_i^+} \frac{Pr_j^t}{|O_j^+|} + (1-\alpha)\frac{1}{N} \tag{4}$$

PageTrust [53], as a modified version of PageRank, is used in the signed networks, in particular in ranking nodes in the signed social networks [55]. In PageTrust, the ranking value of node i is computed using PR_i as follows:

$$PR_i^{t+1} = \left(1 - Q_{ii}(t)\right)[\alpha \sum_{u_j \in I_i^+} \frac{Pr_j^t}{|O_j^+|} + (1-\alpha)\frac{1}{N}] \tag{5}$$

where Q is a matrix computed as:

$$Q(t+1) = T(t)P(t) \tag{6}$$

and $T(t)$ is the transition matrix at time t. PR_i^{t+1} and also PR_i^t are matrixes with $N*1$ dimensions.

In each iteration step, P is calculated as follows [53]:

$$P_{ij}(t+1) = \begin{cases} 1 & if\ (i,j) \in G^- \\ 0 & if (i=j) \\ Q_{ij}(t+1) & otherwise \end{cases} \tag{7}$$

where, G^- denotes the positive links of G. The initial values are defined as $P(0) = Q(0) = A^-$, where the adjacency matrix of each subgraph is represented by A^- while considering just negative links.

Definition 5 Centrality-based measure Centrality Measure Matrix. By sorting the nodes in a network according to m selected centrality measurement and the centrality value, the centrality matrix is represented by matrix D, where a_{ij} is the normalized node i value based on the centrality measure$j$ shown by s_{ij}. The entry of D is normalized to a normal integrated range based on Eq. (9).

$$CC(u_i) = \frac{|d_{in}^+| - |d_{in}^-|}{|d_{in}^+| + |d_{in}^-|} \tag{8}$$

As mentioned previously, $|d_{in}^+|$ is the in-degree of positive links of u_i and $|d_{in}^-|$ is the in-degree of negative links of u_i.

Definition 6 Centrality Measure Matrix. By sorting the nodes in a network according to m selected centrality measurement and the centrality value, the centrality matrix is represented by matrix D where a_{ij} is the normalized value of node i according to the centrality metric j shown by s_{ij}. The entry of D is normalized to a normal integrated range based on Eq. (9).

$$a_{ij} = \frac{s_{ij}}{\sqrt{\sum_{i=1}^{m} {s_{ij}}^2}} \qquad i = 1,2 \ldots m \ ; \ j = 1,2 \ldots n \tag{9}$$

4.2 Objective functions

4.2.1 The Cluster Vulnerability Analysis of SSNs

Definition 7. Local Clustering Coefficient. For a node $i \in V$, there are $k_i$adjacent vertices of i in G. $C(i)$ is the local clustering coefficient, which is characterized as the likelihood that two random neighbors of i are linked in the same way. It quantifies how a neighbor subgraph is close to a clique. The local clustering coefficient C(i) is defined as follows ([8]).

$$C(i) = \frac{\sum_{j,q}(z_{(j,i)}\ z_{(i,q)}\ z_{(j,q)})}{k_i(k_i-1)} \tag{10}$$

where, k_i is the degree of node i by considering the unsigned values, and $z(i,j)$ is the weighted edge between i , j. This index requires a binary network to compare the un-weighted degree in the denominator and considers the weight of all edges in the triangles. The clustering coefficient can be applied to both signed and unsigned networks. It is possible to replace the unsigned adjacency values instead of the signed values. The signed local clustering $\hat{C}(i)$ is computed using Eq. (11):

$$\hat{C}(i) = \frac{\sum_{j,q}\sum_{j,q}(z_{s(j,i)}\ z_{s(i,q)}\ z_{s(j,q)})}{k_i(k_i-1)} \tag{11}$$

The index $\hat{C}(i)$ varies in [-1, 1]. It takes the value 1 when all i neighbors are linked in pairs, and these pairs only form positive triangles with i. The value -1 is taken in the same way and only forms negative triangles. Zero value indicates that i has the same number of positive and negative triangles or i neighbors are not connected.

Definition 8 Average Local Clustering Coefficient (*ALCC*). In graph theory, *ALCC* is a measure used to determine how much vertices of a graph tend to cluster together. *ALCC* for graph G is denoted by *C(G)*. This metric is computed as the *LCC average* over all vertices in the network and is calculated using Eq. (12):

$$C(G) = \frac{1}{N}\sum_{u \in V} C(i) \tag{12}$$

where, $-1 \leq C(i) \leq 1$ for every node $i \in V$, and *C(G)* is normalized and only takes values between [-1, 1] inclusively.

Definition 9 Clustering Structure Analysis (*CSA*). For a signed graph $G = (V, E^+, E^-)$ and a positive number of $k \leq N$, the problem is finding a subset S^* of V with maximum cardinality k (number of members) to maximize the change of the clustering coefficient**.**

$$\Delta C(S) = C(G) - C(G[V \backslash S]) \tag{13}$$

$$S^* = argmax\ \Delta C(S), \qquad S \subseteq V, |S| \leq k \tag{14}$$

The purpose of the *CSA* problem is to identify the network critical vertices to change the average value of the *LCC*. Input parameter k is interpreted as the maximum number of the lost node, which keeps the resistance of the normal network performance against defensive or random attacks. So, the state of

$|S| = k$ identifies the critical vertices accurately to investigate the worst scenarios of failing a network.

4.2.2 Network Energy Based on Balance Theory

Structural balance theory is a fundamental framework for interpreting the interactions between signs. It is also used to investigate their effects on dynamic signed networks. This theory discusses every possible relationship between individuals in such networks. This measure has its motivation in balance theory to determine the stability of a signed graph. Balanced triads (in which the product of signs is positive) are considered stable and unbalanced triads are deemed unstable.

Definition 10. Strong Balance Shapley Value. This measure, defined as $v(C)$, is considered the number of balanced triads minus the number of unbalanced triads.

$$v(C) = \sum_{\{v_j,v_j,v_k\}\in T(C)} -s_{ij}\, s_{jk} s_{kj} \qquad (15)$$

where, the set of triads in C is $T(C)$, and s_{ij} is the edge between i and j. If two nodes of a triad, to which i belong are already present in C, i can only contribute marginally through the triad. This only happens in one-third of the permutations for a pair of adjacent neighbors j and k of node i. Thus, the sharp value $SV(v_i)$ is provided by:

$$SV(v_i) = \sum_{\{v_j,v_k\},v_j\in N(v_k),v_j,v_k\in N(v_i)} \frac{1}{3}(-s_{ij}\, s_{jk} s_{kj}) \qquad (16)$$

Definition 11 Degree of Balance. This measure is defined as DB. Where, G is a signed graph, c(G) stands for the number of cycles of G, and $c_+(G)$ represents the number of positive cycles of G. DB is given by Eq. (17) [56]:

$$DB = \frac{c_+(G)}{c(G)} \qquad (17)$$

In this work, we use this measure for triads (cycles of size 3). We analyze how removing valuable nodes can change the relationships among nodes over time. This analysis is based on relationships with common friends. It is also examined which of these changes converge to a less balanced social structure. Our simulations have been run with the original signs of each selected dataset and discuss the result in the next section.

5. Analysis and Results

As mentioned, ALCC is one of the most popular metric used for network clustering evaluation [٥٧]. A higher ALCC of a network result in better network clustering. ALCC represents several modular network features, such as global-scale phenomena, modular structure, and small diameter (or community structure). ALCC is significant in connected and unconnected diagrams, including dense and sparse graphs. Small networks are expected to have a low clustering coefficient; however, the existing complex networks have a high clustering coefficient.

5.1 Datasets and analysis methods

In order to analyze the relationship between clustering and network vulnerability, empirically, the dimensional generalization of the Watts-Stroggats diagrams was applied. The performance analysis of the proposed approach was conducted in two parts of a synthetic network and three real signed networks. Table 2 shows the characteristics of the synthetic datasets. The columns are the number of nodes, edges, positive edges, negative edges, the maximum value of the Global Clustering Coefficient (CG), and the minimum value of CG. The synthetic graphs were produced using the adjacency matrix of simple Watts-Stroggats graphs.

Table 2. Six signed networks used in our analysis

#Nodes	#edges	#positive edges	#negative edges	Max C_G	Min C_G
300	779	515	264	0.204	0.165
500	1425	947	478	0.67	0.563
1000	3246	2660	586	0.67	0.37
1500	4608	3252	1356	0.93	0.58
2000	6018	4720	1298	0.53	0.3
3000	9335	6929	2406	0.127	0.086

In the analysis phase, the values of *FMF*, *CC*, and *PR* were computed and normalized for the network nodes. Then for each network, k-nodes with the largest values of features are extracted from the network. To analyze the vulnerability, the following strategies are considered:

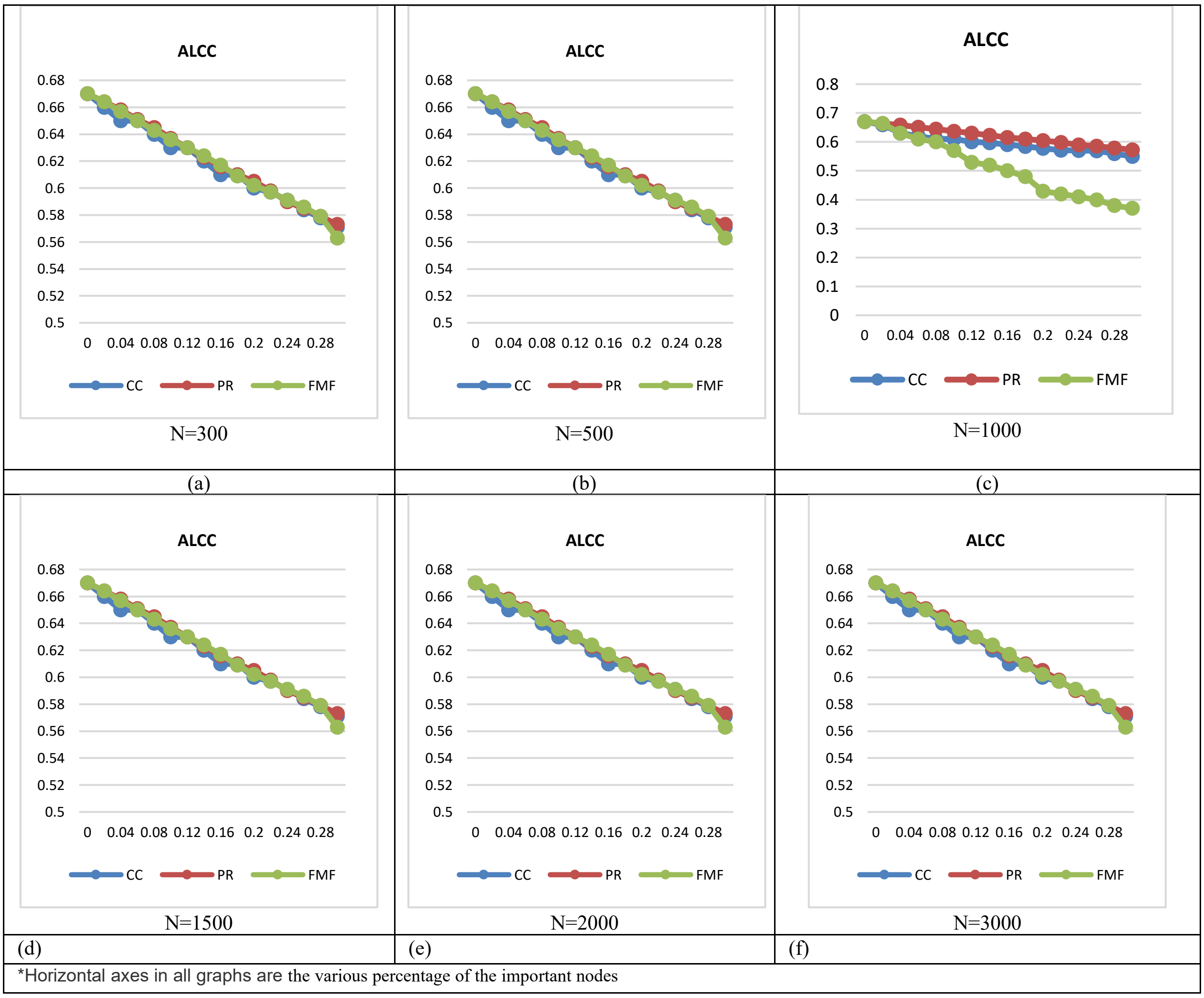
- *FMF-greedy*: Removing nodes using a greedy method concerning the highest values of *FMF*.
- *PR-greedy*: Removing nodes using a greedy method concerning the highest values of *PR*.
- *CC-greedy*: Removing nodes using a greedy method concerning the highest values of *CC*.

One possible algorithm to implement these greedy strategies is the simple-greedy algorithm. The main novelty in this paper is a developed parameter-based greedy algorithm for finding the changes in the clustering coefficient for the CSA problem. The *FMF-greedy* algorithm is shown in Algorithm 1. The other two strategies can be written similar to the same algorithm by modifying line 2.

Algorithm 1 FMF_greedy algorithm

```
1. S←∅;
2. For each $u \in V$ with the max value of FMF do
3.     $\Delta C(S) = C(G) - C(G[V\backslash\{u\}])$
4. End for
5. S←k vertices with highest $\Delta C(S)$ values.
6. Return S
```

The evaluation results of the given strategies for the synthetic datasets are shown in Fig. 2. Horizontal values in the graphs represent the removed fraction of the total nodes, and the vertical values represent the average value of the clustering coefficient.



N=300 (a) N=500 (b) N=1000 (c)

N=1500 (d) N=2000 (e) N=3000 (f)

*Horizontal axes in all graphs are the various percentage of the important nodes

Fig. 2. Changes of *ALCC* based on removing the important nodes

Fig. 2. based on removing the various percentage of the important nodes (in the horizontal axis). The values are based on eliminating significant nodes according to the proposed strategies. The results reveal that by removing the nodes with the highest amount of *FMF*, *ALCC* experiences a drastic reduction, and the graph converges to lower values. For example, in Fig. 2(c), after removing 0.08 nodes, the *FMF* diagram shows a value of about 0.5 for *ALCC* with a significant reduction compared to the other strategies. This reduction difference is increased by removing more percentage of nodes. Considering that *ALCC* computes the average coefficient, removing critical nodes in smaller networks has a more significant impact on this value.

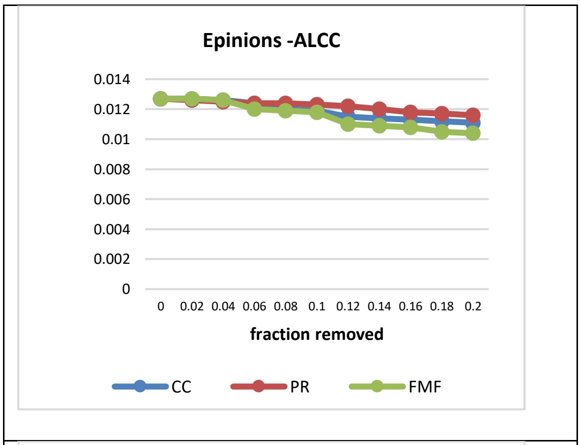
The *ALCC* reduction is more significant in the *FMF-greedy* rather than in the other strategies, in particular when the data size is increased. For example, see Fig. 2(c) to Fig. 2(f), in which the *FMF* chart is located below the *CC* and *PR* charts for the most percentage of the points. This result indicates that nodes with high *FMF* values are critical nodes and should be considered a main problem in the vulnerability issues. On the other hand, removing nodes based on *FMF-greedy* causes the maximum reduction in the clustering coefficient by minimizing the removed nodes. Therefore, the network vulnerability is more affected.

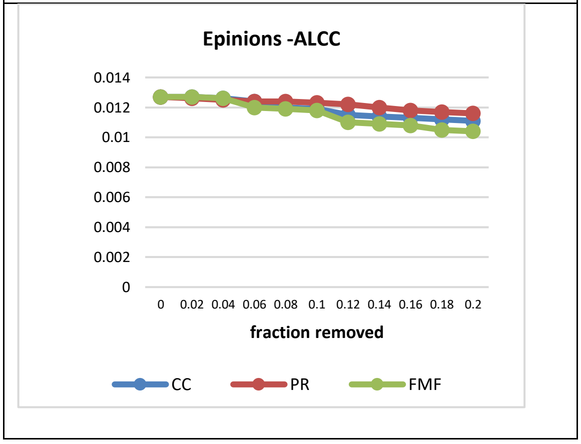
As mentioned, the purpose of the *CSA* problem is to identify the critical vertices of the network to affect the *ALCC* values. We can regularize the required network performance against defensive or random attacks by input parameter *k* (Eq. (14)). As the results show in the *FMF-greedy*, with a smaller value of *k*, the network clustering goes out of the normal mode. For *CC-greedy* and *PR-greedy* algorithms, the value of *k* is higher.

The online signed social networks in this paper include the trusted network of the Epinions, the social network of the blog Slashdot, and the voting network of Wikipedia (Table 3). Moreover, these sub-networks were directed, and we made them undirected. We have two convincing reasons for this process. The reasons are: (i) These datasets have a tiny fraction of reciprocated edges with different signs (0.0032 % for Epinions, 0.0037 % for Slashdot, and 0.0273 % for Wikipedia); so, the number of deleted edges is fewer, (ii) In balance theory, the excessive edges should not have a conflict with the structural balance theory of the signed networks. Therefore, it is reasonable to expect that the reciprocated edges have the same signs.

Table 3. Three real signed networks used in our analysis

	Epinions	Slashdot	Wikipedia
#Nodes	119,217	82,144	7,118
#Edges	841,200	549,202	103,747
#positive edges	85.0%	77.4%	78.7%
#Negative edges	15.0%	22.6%	21.2%
Triad	13,375,407	1,508,105	790,532





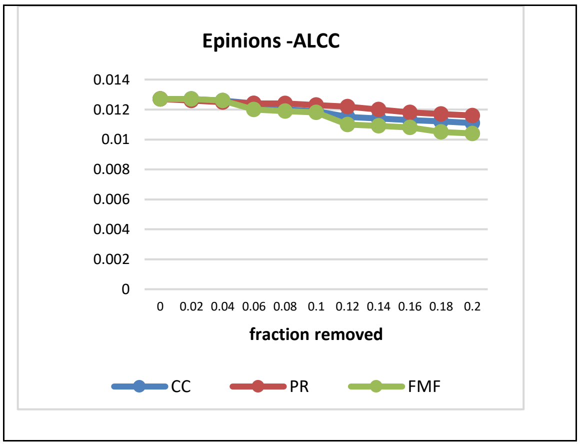


Fig. 3. Changes of ALCC based on removing the important nodes in real signed networks

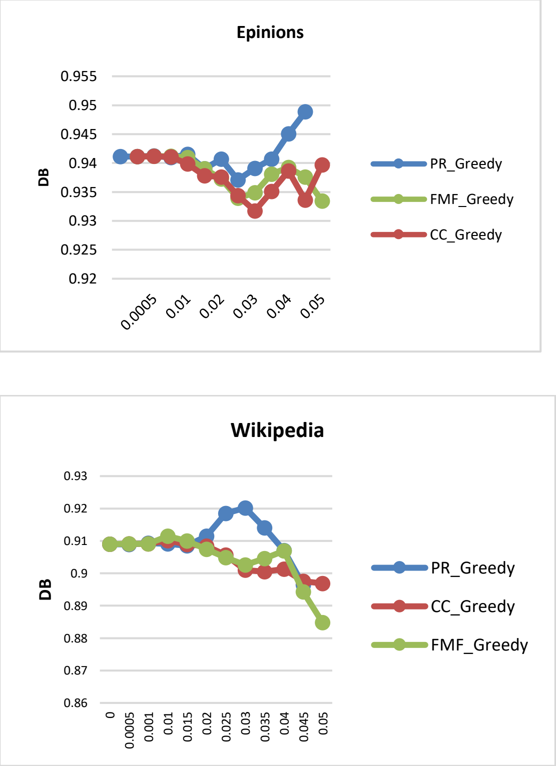
Figure 3 represents the behavior of the clustering coefficient based on the removal of critical centrality nodes by the applied strategies. The results reveal that removing nodes based on the *FMF-greedy* strategy creates the maximum reduction in the clustering coefficient with the least number of eliminated nodes, as in the case of synthetic datasets. Furthermore, the nodes with the greatest *FMF* are critical, and they strongly affect the network's vulnerability. In other words, removing the critical nodes significantly damages the network due to reducing the clustering coefficient. The parameter *CC* and *PR* measure significantly reduce the clustering coefficient when a considerable number of network nodes are removed. Deleting a large percentage of nodes is not possible for assessing the behavior of the clustering coefficient in vulnerability analysis.

In the analysis phase, we investigated and verified the reasons for changes in the *ALCC*. For this purpose, we utilized important properties and principles of the signed network clustering guided by structural balance theory.

Facchetti et al. [58] showed that real signed networks are almost balanced. It is due to the distribution of the signs of edges on the node in these types of networks. In the following, we specifically focus on the proportion of balanced/unbalanced triangles suggested by balance theory. We analyze what happens if important centrality nodes, including *FMF*, *PR*, and *CC*, are removed in the real datasets.

Our results are the averaged results of 10 generated networks for each method on each dataset. *FMF*, *PR*, and *CC* perform near identically on the balanced values distribution and the local clustering coefficient. In the experiments, we assess the variance of balance degrees after removing nodes in each *greedy* strategy. We verify the influence of removing nodes on the degree of balance for the Epinion dataset. In each removal step, the number of positive and negative triangles are decreased unpredictably. Therefore, the values of balance may increase or decrease. Dominant triads are positive for all datasets at first. When some nodes are removed from a network, the number of triads decreases with a random distribution. However, removing based on the *FMF-greedy* strategy decreases positive triads more than other strategies. It means that the removal strategy has a significant effect on how triads evolve. It also affects the degree of balance.

In the *FMF-greedy* algorithm, the number of positive triads is decreased more than the number of negative triads. In other strategies, *DB* increases or decreases in each removing step. To show the phenomenon more clearly, the final values of *DB* are compared with other strategies on real-world datasets in Fig. 4. As shown, the results of *DB* in *FMF-greedy* are slightly less than the *DB* values in other strategies with some exceptions.



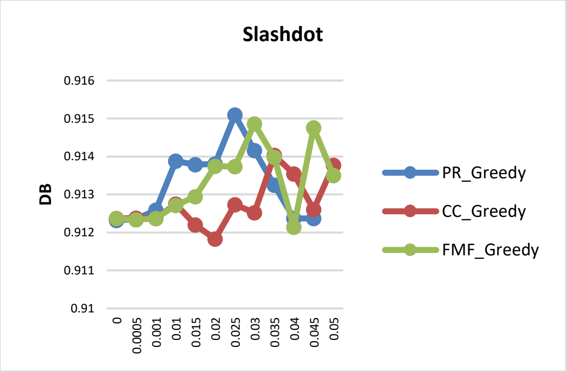


Fig. 4. Changes of DB based on removing the important node in real signed networks

6. Conclusions and Future work

Very few works have focused on modeling signed networks, and most have been carried out for unsigned networks. Generative network models have provided a deep insight into the underlying network structures. With this summary, we attempted to gather and analyze the scientific activities in the field of social balance. We provided a systematic approach to capture the vulnerability of *SSNs* in terms of important properties and principles of the signed networks. We proposed a novel model for vulnerability detection in the signed networks guided by structural balance theory. Empirical experiments on three real-world signed networks demonstrate that the clustering coefficient was significantly reduced while removing the lower percentage of the nodes with the highest amount of *FMF*. In other words, *FMF* was a more critical node in the vulnerability of *SSNs*, and *CC* and *PR* took the following ranks, respectively. To verify the results, the structural balance of the *SSN* was evaluated in each step. The clustering coefficient distribution and the balance degree in triangle distribution after removing nodes in each *greedy* method were changed to agree with each other for the *FMF*, *PR*, and *CC* measures. We will further investigate both directed and weighted signed networks in future work. For future work, we plan to extend this study by applying other sign distributions with the probabilities being proportional to the positive/negative counts of edges.

References

[1] S. Wang, Y. Du, and Y. Deng, "A new measure of identifying influential nodes: Efficiency centrality," *Communications in Nonlinear Science and Numerical Simulation,* vol. 47, pp. 151-163, 2017.

[2] X. Wang, B. Jiang, and B. Li, "Opinion dynamics on social networks," *Acta Mathematica Scientia,* vol. 42, no. 6, pp. 2459-2477, 2022.

[3] L. Fei, H. Mo, and Y. Deng, "A new method to identify influential nodes based on combining of existing centrality measures," *Modern Physics Letters B,* vol. 31, no. 26, p. 1750243, 2017.

[4] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, "Vital nodes identification in complex networks," *Physics Reports,* vol. 650, pp. 1-63, 2016.

[5] S. Kumar, A. Mallik, A. Khetarpal, and B. Panda, "Influence maximization in social networks using graph embedding and graph neural network," *Information Sciences,* vol. 607, pp. 1617-1636, 2022.

[6] D. Cartwright and F. Harary, "Structural balance: a generalization of Heider's theory,"

Psychological review, vol. 63, no. 5, p. 277, 1956.

[7] X. He, R. Zhang, and B. Zhu, "A generalized modularity for computing community structure in fully signed networks," *Complexity,* vol. 2023, 2023.

[8] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature,* vol. 393, no. 6684, p. 440, 1998.

[9] J. Kunegis, A. Lommatzsch, and C. Bauckhage, "The slashdot zoo: mining a social network with negative edges," in *Proceedings of the 18th international conference on World wide web*, 2009: ACM, pp. 741-750.

[10] L. S. Alla and A. S. Kare, "Opinion Maximization in Signed Social Networks Using Centrality Measures and Clustering Techniques," in *International Conference on Distributed Computing and Intelligent Technology*, 2023: Springer, pp. 125-140.

[11] E. Estrada, "Rethinking structural balance in signed social networks," *Discrete Applied Mathematics,* vol. 268, pp. 70-90, 2019.

[12] T. Minh Pham, I. Kondor, R. Hanel, and S. Thurner, "The effect of social balance on social fragmentation," *Journal of the Royal Society Interface,* vol. 17, no. 172, p. 20200752, 2020.

[13] S. Aref, L. Dinh, R. Rezapour, and J. Diesner, "Multilevel structural evaluation of signed directed social networks based on balance theory," *Scientific reports,* vol. 10, no. 1, pp. 1-12, 2020.

[14] F. Adriaens and S. Apers, "Testing properties of signed graphs," *arXiv preprint arXiv:2102.07587,* 2021.

[15] L. Dinh, R. Rezapour, L. Jiang, and J. Diesner, "Enhancing structural balance theory and measurement to analyze signed digraphs of real-world social networks," *Frontiers in Human Dynamics,* vol. 4, p. 1028393, 2023.

[16] A. Arya, P. K. Pandey, and A. Saxena, "Balanced and Unbalanced Triangle Count in Signed Networks," *IEEE Transactions on Knowledge and Data Engineering,* 2023.

[17] ا. مهدی پور, س. مرادی, ج. باقرزاده, and "Modeling homogeneous contact distribution of nodes and its application in routing in Mobile Social Networks," مجله مهندسی برق دانشگاه تبریز, vol. 51, no. 4, pp. 431-441, 2021. [Online]. Available: https://tjee.tabrizu.ac.ir/article_14667_46e8382c34f8e83deb2d6ba7a1c3cd74.pdf.

[18] ش. اسماعیلی, ه. پروین, پ. مرادی and "ترکیب تجزیه نامنفی ماتریسی با روابط اعتماد برای توصیه در شبکه‌های اجتماعی," مجله مهندسی برق دانشگاه تبریز, vol. 50, no. 2, pp. 605-618, 2020. [Online]. Available: https://tjee.tabrizu.ac.ir/article_10954_7ce1afbc4106f8245f364156e217c4ae.pdf.

[19] S. Dhelim, N. Aung, M. T. Kechadi, H. Ning, L. Chen, and A. Lakas, "Trust2Vec: Large-scale IoT trust management system based on signed network embeddings," *IEEE Internet of Things Journal,* vol. 10, no. 1, pp. 553-562, 2022.

[20] H. Du, X. He, and M. W. Feldman, "Structural balance in fully signed networks," *Complexity,* vol. 21, no. S1, pp. 497-511, 2016.

[21] D. Feng, R. Altmeyer, D. Stafford, N. A. Christakis, and H. H. Zhou, "Testing for balance in social networks," *Journal of the American Statistical Association,* vol. 117, no. 537, pp. 156-174, 2022.

[22] L. Shi, W. Li, M. Shi, K. Shi, and Y. Cheng, "Opinion Polarization Over Signed Social Networks With Quasi Structural Balance," *IEEE Transactions on Automatic Control,* 2023.

[23] M. A. Alim, N. P. Nguyen, T. N. Dinh, and M. T. Thai, "Structural vulnerability analysis of overlapping communities in complex networks," in *Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 01*, 2014: IEEE Computer Society, pp. 5-12.

[24] M. N. Abadeh and M. Mirzaie, "Ranking Resilience Events in IoT Industrial Networks," in *2021 5th International Conference on Internet of Things and Applications (IoT)*, 2021: IEEE, pp. 1-5.

[25] T. N. Dinh, Y. Xuan, M. T. Thai, P. M. Pardalos, and T. Znati, "On new approaches of assessing network vulnerability: hardness and approximation," *IEEE/ACM Transactions on Networking,* vol. 20, no. 2, pp. 609-619, 2012.

[26] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *nature,* vol. 406, no. 6794, p. 378, 2000.

[27] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, "Attack vulnerability of complex networks," *Physical review E,* vol. 65, no. 5, p. 056109, 2002.

[28] S. Allesina and M. Pascual, "Googling food webs: can an eigenvector measure species' importance for coextinctions?," *PLoS computational biology,* vol. 5, no. 9, p. e1000494, 2009.

[29] T. H. Grubesic, T. C. Matisziw, A. T. Murray, and D. Snediker, "Comparative approaches for assessing network vulnerability," *International Regional Science Review,* vol. 31, no. 1, pp. 88-112, 2008.

[30] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Signed networks in social media," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2010: ACM, pp. 1361-1370.

[31] T. P. Peixoto and S. Bornholdt, "Evolution of robust network topologies: Emergence of central backbones," *Physical review letters,* vol. 109, no. 11, p. 118703, 2012.

[32] D. S. Callaway, M. E. Newman, S. H. Strogatz, and D. J. Watts, "Network robustness and

fragility: Percolation on random graphs," *Physical review letters,* vol. 85, no. 25, p. 5468, 2000.

[33] A. Veremyev, O. A. Prokopyev, and E. L. Pasiliao, "Critical nodes for distance-based connectivity and related problems in graphs," *Networks,* vol. 66, no. 3, pp. 170-195, 2015.

[34] A. Veremyev, O. A. Prokopyev, and E. L. Pasiliao, "An integer programming framework for critical elements detection in graphs," *Journal of Combinatorial Optimization,* vol. 28, no. 1, pp. 233-273, 2014.

[35] X. Chen, "System vulnerability assessment and critical nodes identification," *Expert Systems with Applications,* vol. 65, pp. 212-220, 2016.

[36] T. Gomes *et al.*, "A survey of strategies for communication networks to protect against large-scale natural disasters," in *Resilient Networks Design and Modeling (RNDM), 2016 8th International Workshop on*, 2016: IEEE, pp. 11-22.

[37] G. Kalna and D. J. Higham, "A clustering coefficient for weighted networks, with application to gene expression data," *Ai Communications,* vol. 20, no. 4, pp. 263-271, 2007.

[38] A. Kuhnle, N. P. Nguyen, T. N. Dinh, and M. T. Thai, "Vulnerability of clustering under node failure in complex networks," *Social Network Analysis and Mining,* vol. 7, no. 1, pp. 1-15, 2017.

[39] H. Liu, Z. Tian, A. Huang, and Z. Yang, "Analysis of vulnerabilities in maritime supply chains," *Reliability Engineering & System Safety,* vol. 169, pp. 475-484, 2018.

[40] Y. E. Malashenko, I. A. Nazarova, and N. y. M. Novikova, "Analysis of cluster damages in network systems," *Computational Mathematics and Mathematical Physics,* vol. 60, no. 2, pp. 341-351, 2020.

[41] N. P. Nguyen, M. A. Alim, Y. Shen, and M. T. Thai, "Assessing network vulnerability in a community structure point of view," in *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on*, 2013: IEEE, pp. 231-235.

[42] A. Kuhnle, N. P. Nguyen, T. N. Dinh, and M. T. Thai, "Vulnerability of clustering under node failure in complex networks," *Social Network Analysis and Mining,* vol. 7, no. 1, p. 8, 2017.

[43] Z. Ertem, A. Veremyev, and S. Butenko, "Detecting large cohesive subgroups with high clustering coefficients in social networks," *Social Networks,* vol. 46, pp. 1-10, 2016.

[44] E. J. Bienenstock and P. Bonacich, *Balancing efficiency and vulnerability in social networks*. na, 2002.

[45] A. Kumari, R. K. Behera, K. S. Sahoo, A. Nayyar, A. Kumar Luhach, and S. Prakash Sahoo, "Supervised link prediction using structured-based feature extraction in social network," *Concurrency and Computation: practice and Experience,* vol. 34, no. 13, p. e5839, 2022.

[46] J. Scott and P. J. Carrington, *The SAGE handbook of social network analysis*. SAGE publications, 2011.

[47] B. Yang, W. Cheung, and J. Liu, "Community mining from signed social networks," *IEEE transactions on knowledge and data engineering,* vol. 19, no. 10, pp. 1333-1348, 2007.

[48] J. Huang *et al.*, "Negative can be positive: Signed graph neural networks for recommendation," *Information Processing & Management,* vol. 60, no. 4, p. 103403, 2023.

[49] W. Xing and A. Ghorbani, "Weighted pagerank algorithm," in *Proceedings. Second Annual Conference on Communication Networks and Services Research, 2004.*, 2004: IEEE, pp. 305-314.

[50] M. Pasquinelli, "Google’s PageRank algorithm: A diagram of cognitive capitalism and the rentier of the common intellect," *Deep search: The politics of search beyond Google,* pp. 152-162, 2009.

[51] J. Kunegis, A. Lommatzsch, and C. Bauckhage, "The slashdot zoo: mining a social network with negative edges," presented at the Proceedings of the 18th international conference on World wide web, Madrid, Spain, 2009.

[52] P. Bonacich and P. Lloyd, "Calculating status with negative relations," *Social networks,* vol. 26, no. 4, pp. 331-338, 2004.

[53] C. d. Kerchove and P. V. Dooren, "The pagetrust algorithm: How to rank web pages when negative links are allowed?," in *Proceedings of the 2008 SIAM International Conference on Data Mining*, 2008: SIAM, pp. 346-352.

[54] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the web," Stanford InfoLab, 1999.

[55] M. Shahriari and M. Jalili, "Ranking nodes in signed social networks," *Social network analysis and mining,* vol. 4, no. 1, p. 172, 2014.

[56] A. Teixeira, F. C. Santos, and A. P. Francisco, *Emergence of Social Balance in Signed Networks*. 2017.

[57] D. Schoch, "signnet: An R package for analyzing signed networks," *Journal of Open Source Software,* vol. 8, no. 81, p. 4987, 2023.

[58] G. Facchetti, G. Iacono, and C. Altafini, "Computing global structural balance in large-scale signed social networks," *Proceedings of the National Academy of Sciences,* vol. 108, no. 52, pp. 20953-20958, 2011.