Analyzing Power Grids' Cascading Failures and Critical Components using Interaction Graphs

Upama Nakarmi and Mahshid Rahnamay-Naeini
Electrical Engineering Department, University of South Florida, Tampa, Florida, USA
Email: unakarmi@mail.usf.edu, mahshidr@usf.edu

***Abstract*—An important property of cascading failures in power grids is that, in addition to the physical topology of the system, the physics of the power flow and functional dependencies among components largely affect the spatial distribution and propagation of failures. Various techniques have been proposed to capture the interactions among the components in the models of power grids, beyond the physical topology, during cascading failures. In this paper, we create the logical network of interactions among components atop the physical topology of the system using influence-based and correlation-based techniques. We will discuss that different techniques to derive the graph of interactions and various power grid operating settings can lead to different analyses of interactions and cascading behavior. We will further introduce a new measure based on the overlapped community structures in the graph of interactions to identify critical components in cascade of failures, whose protection can help in containing failures within a community and prevent the propagation of failures to the whole power grid. We have verified the important role of the identified critical components in our dataset of cascading failures based on power system simulations.**

I. Introduction

Large blackouts, such as the Northeast US blackout of 2003, result from cascade of failures in the transmission network of the power grid. Large number of factors and their complex interactions contribute in the cascading phenomenon. Intensive research efforts have been focused on understanding the underlying interactions in cascades, which can, for instance, help in predicting the propagation path of failures and the critical/vulnerable components in the power grid.

The physical topology of the power grids has been the focus of many studies in modeling and analyzing cascading failures, using graph theoretic [1], percolation theory [2] and epidemic models [3]. However, historical data and studies of the power grids have shown that the physical topology is not adequate in capturing the interactions among the components in the models of power grids. Real data confirm that failure of a critical transmission line can overload transmission lines that are 100 miles away and not physically connected together. Hence, the propagation of failures in the power grid is not local on its physical topology, instead there are influences/interactions among the components at distance due to physics of electricity governing the power flow dynamics as well as other functional dependencies controlled by the automated systems/operators. A study in [4] shows lack of strong connection between the physical topology properties and failure propagation process in the power grid. As such, instead of the physical topology, many recent studies have been focused on extracting and modeling the underlying logical graph of interactions for modeling power grids using data-driven techniques, such as influence-based and correlation-based approaches based on historical and simulation data [5]-[7]. Moreover, various studies have been conducted on the graph of interactions modeling power grids to identify critical components in power grids as well as structures embedded in the interactions [7], [8].

In light of the above, the study presented in this paper has two folds. *First*, we use two techniques in extracting the logical graph of interactions among the transmission lines of the power grid, specifically influence-based and correlation-based approaches, using a cascade dataset generated by our power-system simulations in MATPOWER [9]. Although failures of transmission lines, generators and cyber (control, monitoring, computation and communication) elements all contribute in the cascading failures, we focus on transmission line failures and their interactions due to their significant role in the cascade process (when the power flow redistributes due to failures it can cause transmission line overloads, which is a key mechanism in propagation of failures). A similar approach can be applied to other components and their interactions in the cascade. We will show that the resulted graph of interactions depends on the operating settings (e.g., loading level of the power grid) as well as the approach used to extract the interactions. In other words, we conjecture that each approach and each dataset (for different operating settings) will shed light on specific interactions in cascading failures.

Second, we study the structure of the derived graphs of interactions using a community detection technique. We focus on overlapped communities formed on the graph of interactions and use it to introduce a new centrality measure to specify the criticality of the components (transmission lines) in the cascade process. The main idea behind this new measure is that communities formed on interactions reveal the group of components that are likely to contain the failures within themselves during cascade (due to tight influences/interactions); however, if a component belongs to multiple communities (having tight interactions and influences with multiple groups of components) then it will be critical in the cascade process as it serves as a gateway to spread failures from one community to another. Identifying and protecting such components can help containing the failures within one community and prevent spreads over the whole system. We will verify the criticality of the identified components using our cascade dataset.

II. Related Work

In this section, we will review examples of cascade studies in power grids with a focus on extracting and analyzing the logical interactions among the components of the system.

A. Logical Graph of Interactions, Influences, and Correlations

To reveal the underlying interactions, influences and state correlations among various components of the power grid, various techniques including data-driven, statistical analysis and electrical interactions, based on Kirchhoff's and Ohm's laws have been used. For instance, Hines et al. in [10] introduced the line interaction graph, where vertices are the transmission lines in the power grid and the edges in the graph are inferred from the influences among the lines based on large amount of cascade data, electrical interaction considerations based on Line Outage Distribution Factors (LODFs) [11] and n-k line contingencies testing. Specifically, line outages and their orders were considered to statistically analyze the empirical probabilities of ordered pairs of lines in cascade data. They extended the study in [6], introducing a more systematic approach in analyzing cascade data by defining generations of failures within cascades and finding the interactions based on the influence model [12]. The characterized influences specify how failure in one line will affect the likelihood of failure in other transmission lines. In [13], Dobson et al. used historical transmission line outage data to form the network topology for statistical characterization of spread of failures in the power grid and evaluated the properties of generations in cascade data to define components' distances based on the formed topology.

In [8], Luo et al. formed the graph of interactions for transmission lines by introducing cascading failure chain and grouping failures in stages based on their order in the cascade to statistically define the interactions among the lines based on frequency and order of their appearance in cascade data. In [14], the authors used the cascade data and generations in each cascade to calculate the empirical probability of failure of a component causing failure in another component to define the graph of interactions and used it to find mitigation strategies by weakening interactions between key lines in the power grid.

Another approach in [7] forms the graph of interactions based on correlations among the failure/functional status of transmission lines in the cascade data to construct a network of positively correlated transmission lines. Also, electrical interactions among the components of a power grid can define the logical interactions in the cascade process. For instance, LODFs [11] that can measure the impact of a change in a line's status on the flows on other lines in the system can be used to define interaction links among transmission lines [10].

B. Analysis of Graph of Interactions

The graph of interactions has been used for various analysis in different studies. For instance, [6], [8] and [15] used the logical graph of interactions to identify critical transmission lines based on the interaction graph properties. In addition, some works are focused on predicting the distribution, path or size of the cascade based on the structures and properties in the graph of interactions [7].

III. Techniques to Form Graph of Interactions

We briefly review the influence- [6] and correlation-based [7] approaches to form the logical graph of interactions.

A. Influence-based Approach to Graph of Interactions

A combination of the branching process and influence model was used to derive the graph of interactions in [6]. First, strength of influences among the failures were defined, using the concept of generation in the branching process, where each generation of failures produces some dependent failures with a specific rate. However, to discriminate component outages resulting from a particular outage, they defined $\lambda_{i,m}$, which specifies the rate of failures (mean number of outages) in the next generation, m, for the outage of component i based on Poisson distribution. Conditional failure probability of component j, given failure of component i was defined as $g[j|i]$ and estimated by statistical analysis of number of times component j failed in the generation after component i failure (for details refer to [6]). The parameters of the Poisson distribution for failure rates and conditional probability of failures for pairs of components was estimated using simulation-based cascade data. Next, to characterize the influences that a component j receives from all other components, the influence model [12] was considered in which the summation of the influences received by a component should add up to 1. The propagation rates and conditional failure probabilities g, were combined to form an influence matrix *H*, which gives an overall status of the influences among the components. Specifically, the i, j element of matrix *H* is defined as $1 - e^{-\lambda_i g[j|i]}$ for $i \neq j$ and zero otherwise. The matrix *H* can be viewed as a weighted and directed graph of influences among components. For simplicity, we ignore the direction of links in our analyses.

B. Correlation-Based Approach to Graph of Interactions

While, the influence-based graph of interactions is formed based on conditional probabilities and failure propagation rates over cascade data grouped into generations, the correlation-based approach simply considers the pairwise correlation between component failures in cascades as done in [7] to define the graph of interactions. The dependance between failures are captured in the correlation matrix *CR*, which its i, j element is the Pearson correlation coefficient between the failure statuses of components i and j. This will result in a symmetric matrix and an undirected weighted graph, where the weights represent the correlation value among the components.

IV. Interaction Structures and Criticality

Studies in network science have shown that communities play important roles in defining the propagation behaviors in networks [16]. Particularly, the propagations tend to stay within communities due to tight internal interactions and weak external interactions. In this paper, we particularly focus on overlapped community structures in the interaction graph of the power grid as we believe they can provide a new perspective in defining the key players in the cascade process. Here, we review the community detection algorithm used in this paper and introduce a new centrality measure based on the community structure to identify critical components.

A. Overlapping Communities in Graphs of Interactions

We adopt the weighted, overlapped community detection algorithm presented in [17] (with minor modifications) to characterize the community structure in the graph of interactions. In the interaction graph, let $\mathcal{C} = \{C_1, C_2, ..., C_n\}$ denote the set of communities, where *n* is the number of communities. Since, communities are overlapped, $C_i \cap C_j \neq \emptyset$ for some of the i and j pairs of communities. This community detection uses the *belonging degree* and *conductance* as explained next.

1) Belonging Degree: The set of neighbors and the degree for node i is denoted by N_i and D_i, respectively. The degree of node i is defined as $D_i = \sum_{j \in N_i} w_{ij}$, where w_{ij} is the weight of the link from node i to node j. Thus, the belonging degree of node i to a community C_k is defined as $B(i, C_k) = (\sum_{j \in C_k} w_{ij})/D_i$.

2) Conductance: The conductance ϕ_{C_k} of a community $C_k \in \mathcal{C}$ is defined as $\phi_{C_k} = cut(C_k, G \backslash C_k)/w_{C_k}$, where $cut(C_k, G \backslash C_k)$ denotes the sum of the weights of edges adjacent to the nodes in the community except the edges inside the community itself and w_{C_k} denotes the sum of the weights of all the edges connected to the nodes in the community. Smaller values of conductance for communities are preferred.

The community detection algorithm presented in [17] is shown in Algorithm 1 with minor modifications to update the conductance as in line 8 and to remove the analyzed neighbors from the neighboring set as in line 9 and 11 thus, analyzing all neighbors of an initially identified community to either append/skip neighboring nodes. In this algorithm, the edges within community C are denoted by E_c. While, the appending process of a node to a community is different from the algorithm presented in [17], the identified communities are in agreement with the concept of tight interactions within communities and weak interactions outside communities.

B. A New Community-based Centrality Measure

The overlapped community structure in the graph of interactions reveals new players in the cascade process that cannot be identified using traditional centrality measures such as closeness, betweenness, eigenvector, and more [18]. These community structures are more likely to retain cascades inside each community. A component belonging to multiple communities (having tight interactions and correlations with multiple groups of components) serves as a gateway to spread failures from one community to another and thus is critical. Based on this concept, we propose a novel community-based centrality measure. The main idea is that nodes belonging to multiple communities are more important; however, there might be cases that a node belongs to multiple communities but the communities do not have a central role in propagation of failures themselves. To capture both of these factors, we first create an augmented graph of communities to define the importance of communities in the network and then add the importance of communities for the nodes (nodes belonging to multiple communities get the importance measure from all).

Algorithm 1 Community Detection Algorithm based on [17]

Input: Graph $G = (V, E)$
Output: Overlapped Communities $\mathcal{C}$
1: **Initialize:** $\mathcal{C} = \emptyset$
2: **while** $E \neq \emptyset$ **do**
3: $\quad C = \{i, j\}$ where $e(i, j) = \underset{(u,v) \in E}{\operatorname{argmax}} w_{ij}$
4: $\quad$ **while** $N_c \neq \emptyset$ **do**
5: $\quad\quad C' = C \cup \underset{w \in N_c}{\operatorname{argmax}} B(w, c)$
6: $\quad\quad$ **if** $\Phi(C') < \Phi(C)$ **then**
7: $\quad\quad\quad C = C'$
8: $\quad\quad\quad \Phi(C) = \Phi(C')$
9: $\quad\quad\quad N_c = N_c \backslash C$
10: $\quad\quad$ **else**
11: $\quad\quad\quad N_c = N_c \backslash C$
12: $\quad\quad$ **end if**
13: $\quad$ **end while**
14: $\quad E = E \backslash E_c$
15: $\quad \mathcal{C} = \mathcal{C} \cup C$
16: **end while**

The set of communities $\mathcal{C}$ found in Section IV-A is used to create an augmented graph $\mathcal{AG} = (C, E)$, where *C* represents the set of communities and *E* represents connections due to common links/nodes between communities. The common link e_{ij} has node i in community C_s at one end and node j in community C_r at the other end, while $s \neq r$. Similarly, a common node i belongs to both communities C_s and C_r. Next, we consider link weights among the nodes in $\mathcal{AG}$ to represent the strength of overlap among communities. Specifically, if $|C_i \cap C_j| \neq \emptyset$, we consider a weight inversely proportional to the overlap ratio *(OR)* between communities as $OR_{\{C_i, C_j\}} = |C_i \cap C_j|/|C_i \cup C_j|$, where $|C_i \cap C_j|$ and $|C_i \cup C_j|$ denote the number of overlapped nodes and the total number of unique nodes between communities C_i and C_j, respectively. In other words, communities with larger overlap ratios will be connected with smaller link weights (i.e., suggesting a smaller distance between the communities). The rest of the communities with no overlaps are assigned equal larger weights for their links than any overlapped communities (i.e., suggesting a larger distance between the communities). The weighted $\mathcal{AG}$ is then used to find closeness centrality (CL) for each community in the network of interactions. Finally, we define the community-based centrality of the node i as $I_i = \sum_{k \text{such that} i \in C_k} CL_k$, where CL_k is the closeness centrality of community k.

V. Cascade Studies and Results

We first introduce our cascade dataset generated for the IEEE 118 bus system using our power simulations in MATPOWER. Next, we use the influence-based and correlation-based approaches to derive the graph of interactions for the IEEE 118 bus system. We will then present our analyses of the graph of interactions and present the identified critical components based on the measure introduced in Section IV-B. Finally, we use the cascade dataset to verify the importance of the identified critical components.

A. Cascade Dataset and Operating Characteristics

We used MATPOWER [9], a package of MATLAB m-files, for solving the optimal power flow and simulating cascading failures based on a quasi-static approach that focuses on transmission line overloads as the mechanism for propagation of failures. Using our simulations, we generated a large dataset of cascade scenarios each triggered by two or three random initial failures in the system. We use the IEEE 118 test bus system with 118 buses, including substations and generators, and 186 transmission lines. We use the power grid loading level, r (the ratio of the total demand over total generation capacity of the system) to simulate cascades under different settings and evaluate the role of operating characteristics in cascade analyses. For each analysis, we have simulated at least 20,000 unique cascading failure scenarios.

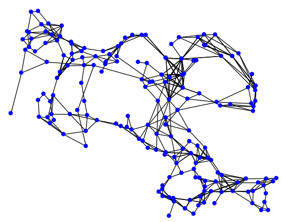
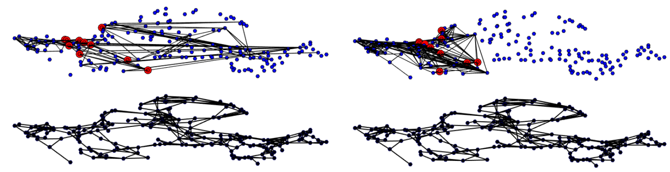


Fig. 1. IEEE 118 test bus case converted to a 186 node physical topology

B. Graph of Interactions

The simulated cascade dataset is used to find logical graphs of interactions, *H* and *CR* based on the techniques discussed in Sections III. The physical topology of the IEEE 118 is converted to a 186 node dual network (Fig. 1) where, the nodes represent the edges in the original topology and links represent common nodes between the lines. From now on, we refer to this dual graph as the physical topology of the system.

In our first study, we use all cascade scenarios without considering operating characteristics (i.e., r parameter). Excluding the self loops, total number of possible links in the dual graph is 34,410 for the undirected correlation *CR* and 68,820 for the directed influence *H* matrices. However, based on the dataset, the graph of interactions for *CR* and *H* consist of 34,396 and 32,504 links, respectively. As these approaches pick even the smallest interactions based on the data, to focus on the major interactions in the system we apply thresholds to only consider interactions with strength larger than a threshold value. Since, thresholds can result in islands in the graph, we focus on the Largest Connected Components (LCC). Identification of LCC in interactions graphs with major strength of interactions can be used in prediction of the largest cascade sizes. For comparisons between the interaction graphs based on influence- and correlation-based approaches, we choose thresholds such that the size of LCC is comparable in both networks. For instance, Fig. 2 shows the graph of interactions atop the physical topology for influence- and correlation-based approaches for threshold 0.7, which the size of LCC for *H* and *CR* are 57 and 59, respectively. The strong interactions on the same dataset of cascades is shown among different set of components in Fig. 2, which emphasize on the role of the technique in extracting the logical graph of interactions.



(a) *H* plot (Threshold = 0.7) (b) *CR* plot (Threshold = 0.7)

Fig. 2. Graph of interactions for *H* and *CR* over the physical topology.

TABLE I
OVERLAP RATIO *(OR)* OF COMMUNITIES IN *H* WITH THRESHOLD 0.7

Community	3	4	6	7	8	9	10	11	12
1	-	0.067	-	0.053	0.389	-	-	-	-
2	0.1	-	0.08	0.105	0.13	0.187	-	0.058	0.133
3	-	-	0.105	0.154	0.111	-	-	0.091	-
5	-	-	0.227	-	-	-	0.2	-	-
6	-	-	-	0.053	-	0.058	0.062	0.062	-
7	-	-	-	-	0.187	-	-	-	-
8	-	-	-	-	-	0.062	-	-	-
9	-	-	-	-	-	-	-	0.125	-

TABLE II
CRITICAL COMPONENTS BASED ON COMMUNITY-BASED CENTRALITY (I)

Rank	$H \geq 0.6$	$H \geq 0.7$	$CR \geq 0.4$	$CR \geq 0.7$
1	32	25	138	44
2	151	26	119	26
3	96	24	82	25
4	167	32	73	43
5	76	55	64	27

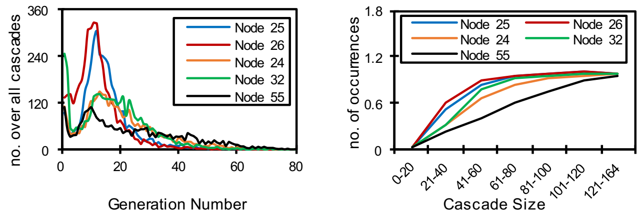
C. Analyses of Structures and Critical Components

To analyze the structure of interactions, we applied overlapped community detection algorithm discussed in Section IV-A over the LCC of the logical interaction graphs *H* and *CR* after applying threshold 0.7 and found 12 communities in each. However, the structure of communities and their *OR* are quite different. Our analysis show that some communities do not overlap suggesting that failures inside such communities rarely propagate to other communities. Table I shows the *OR* for the communities identified in *H* with threshold 0.7. Next, using the measure introduced in Section IV-B, we identified 5 most influential nodes in the cascade process in each of the networks with different thresholds but similar LCC size (Table II and marked in Fig. 2). Note that the size of LCC is similar for *H* and *CR* with threshold 0.6 and 0.4 respectively. For instance, nodes 32 and 151 in *H* with threshold 0.6 belong to 5 different communities, which contributed to their importance.

Similar to the different techniques that lead to different graph of interactions on the cascade process, operating settings, e.g. different loading levels for the power grid r, also result in different graphs of interactions and ranking of critical

TABLE III
CRITICAL COMPONENTS BASED ON COMMUNITY-BASED CENTRALITY (I)

	H (Threshold = 0.3)				*CR (Threshold = 0.3)*			
Rank	r=0.6	r=0.7	r=0.8	r=0.9	r=0.6	r=0.7	r=0.8	r=0.9
1	9	62	45	8	44	68	26	20
2	58	44	51	9	36	71	36	36
3	35	16	52	45	21	69	38	54
4	27	181	9	58	51	75	25	31
5	61	7	102	48	50	50	39	33



(a) (b)

Fig. 3. (a) Number of times critical components failed in various generations and (b) number of occurrences of the critical components in different cascade sizes for $H \geq 0.7$.

components. Table III, shows the top 5 critical components identified for the system for different r values. These results also show that depending on the condition and the operating settings of the power grid, the critical components of the system may vary. Therefore, it is important to perform criticality study with considerations about the power grid's conditions.

D. Verification of Criticality

In order to verify that the identified critical components based on the new community-based centrality measure (introduced in Section IV-B) reveal the actual influential components in the cascade process, we have done a set of analyses based on our cascade dataset. Particularly, we have focused on the 5 most critical components identified for H with a threshold of 0.7 (i.e., nodes 24, 25, 26, 32 and 55 according to Table II). Fig. 3a shows that these five nodes appear in the early generations of the cascades, which shows their contribution in the progress of cascade. Moreover, the results in Fig. 3b show that among all cascade sizes observed in the dataset, these nodes tend to be a part of larger cascades. In other words, the occurrence of these nodes in cascades increase with the size of cascade. These results confirm the criticality of the identified nodes using the new community-based centrality measure.

VI. CONCLUSION

We discussed that the underlying physical topology of power systems are not enough in capturing the complex interactions among components in models of cascading failures. We reviewed various techniques, with a focus on data-driven techniques, to derive the logical graph of interactions from simulation cascade data. We specifically studied the logical network of interactions for the IEEE 118 bus system using influence-based and correlation-based approaches and for various power grid operating settings, particularly its loading level using cascade dataset generated from our power system simulations. We also defined a novel measure based on the overlapped community structures in the graph of interactions to identify critical components in cascades, whose protection can help in containing failures within a community and prevent propagation to the whole power grid. Finally, we used the cascade dataset and showed that the identified critical components play important roles in cascading failures.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. 1541018. The authors would like to thank Mr. Sai Saravana Harish Ravindran for his assistance in some of the data preprocessing steps.

REFERENCES

[1] A. J. Holmgren, "Using graph models to analyze the vulnerability of electric power networks," *Risk analysis*, vol. 26, no. 4, pp. 955–969, 2006.

[2] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, no. 7291, pp. 1025–1028, 2010.

[3] P. Crucitti, V. Latora, and M. Marchiori, "Model for cascading failures in complex networks," *Physical Review E*, vol. 69, no. 4, p. 045104, 2004.

[4] P. Hines, E. Cotilla-Sanchez, and S. Blumsack, "Do topological models provide good information about electricity infrastructure vulnerability?" *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 20, no. 3, p. 033122, 2010.

[5] P. Hines and S. Blumsack, "A centrality measure for electrical networks," in *41st Hawaii International Conference on System Sciences*, 2008.

[6] P. D. Hines, I. Dobson, and P. Rezaei, "Cascading power outages propagate locally in an influence graph that is not the actual grid topology," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 958–967, 2017.

[7] Y. Yang, T. Nishikawa, and A. E. Motter, "Vulnerability and cosusceptibility determine the size of network cascades," *Physical Review Letters*, vol. 118, p. 048301, Jan 2017.

[8] C. Luo and J. Yang, "Identify critical branches with cascading failure chain statistics and hypertext-induced topic search algorithm," *arXiv preprint arXiv:1704.06917*, 2017.

[9] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011.

[10] P. D. Hines, I. Dobson, E. Cotilla-Sanchez, and M. Eppstein, ""Dual Graph" and "Random Chemistry" methods for cascading failure analysis," in *46th Hawaii International Conference on System Sciences*, 2013.

[11] J. Guo, Y. Fu, Z. Li, and M. Shahidehpour, "Direct calculation of line outage distribution factors," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1633–1634, 2009.

[12] C. Asavathiratham, S. Roy, B. Lesieutre, and G. Verghese, "The influence model," *IEEE Control Systems*, vol. 21, no. 6, pp. 52–64, 2001.

[13] I. Dobson, B. A. Carreras, D. E. Newman, and J. M. Reynolds-Barredo, "Obtaining statistics of cascading line outages spreading in an electric transmission network from standard utility data," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4831–4841, 2016.

[14] J. Qi, K. Sun, and S. Mei, "An interaction model for simulation and mitigation of cascading failures," *IEEE Transactions on Power Systems*, vol. 30, no. 2, pp. 804–819, 2015.

[15] J. Qi and S. Mei, "A cascading failure model by quantifying interactions," *arXiv preprint arXiv:1301.2055*, 2013.

[16] L. Weng, F. Menczer, and Y.-Y. Ahn, "Virality prediction and community structure in social networks," *Scientific reports*, vol. 3, p. 2522, 2013.

[17] Z. Lu, Y. Wen, and G. Cao, "Community detection in weighted networks: Algorithms and applications," in *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2013.

[18] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, and T. Zhou, "Identifying influential nodes in complex networks," *Physica a: Statistical mechanics and its applications*, vol. 391, no. 4, pp. 1777–1787, 2012.