Measures for Network Structural Dependency Analysis

Yordanos T. Woldeyohannes, Yuming Jiang

***Abstract*—A set of new measures for network structural dependency analysis is introduced. These measures are based on geodesic distance, which is the number of links in a shortest path. They capture the structural dependency effect at the path level, the node level and the overall network level, and hence can be used to index such dependencies. Unlike the related literature measures, a novel aspect of the proposed measures is that the impact of network fragmentation caused by a node failure is taken into explicit consideration in deciding the structural dependency effect. As a result, when applied to critical node identification in a network, the proposed measures give results that are more in line with intuition.**

I. INTRODUCTION

Networked systems such as communication networks have become an indispensable part of our daily life. As a consequence, failure of such a system even for short time, e.g. a few minutes or hours, could already be unacceptable let alone for longer time. However, network component failures (e.g. due to hardware, software and communication failures) are often. For a network, its inherent structural dependencies among nodes imply that the impact of one node's failure on the services provided by the network may significantly differ from that of another node's failure. Here arises a fundamental question, referred to as the *structural dependency impact problem* in this work, which is, *what measures may be used to assess the network structural dependency-caused impact*?

The purpose of this paper is to propose an answer to the *structural dependency impact problem*, following the idea that "the importance of a node is related to the ability of the network to respond to the deactivation of the node from the network" [1]. To this aim, a new set of *geodesic distance* based information centrality measures will be introduced, termed as *dependency indexes*. Specifically, these measures are the *path dependency index*, the *node dependency index* and the *network dependency index*. They respectively quantify the impact of a node's failure, at the path level on information communication from one node to another node in the network, at the node level on information communication from one node to other nodes in the network, and at the network level on information communication from any node in the network.

The dependency impact problem is related to the critical node detection problem, which is the problem of finding the most important nodes in a network and has applications in various fields [2]. In communication networks, such applications include network vulnerability analysis [3], critical node discovery [4] and robustness study [5]. In the literature, various measures have been proposed for critical node detection under the concept of *centrality* [6]. The classic centrality measures include node degree, closeness, betweenness and information [5] [7]. However, these measures are generally for the network level, where the impact of structural changes after the node failure at the path level and the node level is not focused.

The most related works are [1] and [8]. In [1], a new category of centrality measures, called *delta centrality*, are introduced. However, as implied by the definition, delta centrality measures only address the dependency impact problem at the network level. In [8], *absolute drop* in reciprocal geodesic distance is used as the basis to quantify the dependency impact, but only at the path level and the node level. In addition, as to be exemplified and analyzed, the dependency measures introduced in [1] and [8] have a strong limitation: While the removal of a node from a network may result in network fragmentation, this effect is not factored in these measures. Addressing this limitation and systematically quantifying the different level dependencies for the dependency impact problem constitute the novelty and contribution of this work.

The rest is organized as follows. First, the unification of the dependency measures in [1] and [8] is proved in Sec. II, where an example showing their limitation will also be given and discussed. Then, the set of new measures are proposed in Sec. III. In Sec. IV, results and *the application of the proposed measures for critical node identification* are demonstrated, compared and discussed. Finally, Sec. V gives the conclusion.

II. NETWORK MODEL AND EXISTING MEASURES

A. Network Model

We consider a network $G(\mathcal{N}, \mathcal{L})$, where $\mathcal{N}$ is the set of the N nodes and $\mathcal{L}$ is the set of the L links. We assume that nodes communicate through their shortest path.

The *information* measure [7] is used as the basis to quantify the influence of a node on a path, another node or the network, or in other words, how a path, another node or the network depends on the node. We use G^{-n} to denote the network that is the same as the original network G but with all links of node n removed, and $\mathcal{N}^{-n}$ to denote the remaining set of $\mathcal{N}$ after excluding node n. By definition, $\mathcal{N}^{-n}$ has $N-1$ nodes.

The concept of *information* between pairs of nodes was originally introduced in [7] as a centrality measure based on the theory of statistical estimation. For shortest path based communication, the *information* measure I_{ij} between node i and node j can be written as the reciprocal of the topological distance d_{ij} between the two nodes, i.e.

$$I_{ij} = \frac{1}{d_{ij}}, \tag{1}$$

where d_{ij} represents the *geodesic distance, i.e. the number of links in a shortest path*, between node i and node j. For a node to itself, it is defined that $d_{ii} = 0$ or $I_{ii} = \infty$; if there is no path between nodes i and j, $d_{ij} = \infty$ or $I_{ij} = 0$.

B. Existing Measures

In [1], to quantify the influence of a node n on the network G, a delta centrality measure, denoted as $\Delta(G|n)$, is introduced as a measure of the relative drop in the network efficiency caused by the deactivation of node n:

$$\Delta(G|n) = \frac{E(G) - E(G^{-n})}{E(G)}, \tag{2}$$

where $E(G)$ denotes the efficiency of the network G which, initially introduced in [9] based on the communication or information efficiency measure, is defined as:

$$E(G) = \frac{1}{N(N-1)} \sum_{\forall i \neq j \in \mathcal{N}} I_{ij}.$$

Note that the delta centrality measure $\Delta(G|n)$ only provides measure at the network level, i.e. how the failure of node n may impact the network.

In [8], a measure of the impact of node n on the path from node i to node j, denoted as $D(i \to j|n)$, is defined as

$$D(i \to j|n) = \frac{1}{d_{ij}} - \frac{1}{d_{ij}^{-n}} = I_{ij} - I_{ij}^{-n}, \tag{3}$$

where d_{ij}^{-n} denotes the geodesic distance between node i and node j in the network G^{-n} and $I_{ij}^{-n} = \frac{1}{d_{ij}^{-n}}$. By the definitions, it is clear that $d_{ij}^{-n} \geq d_{ij}$. Also in [8], based on the path level measure $D(i \to j|n)$, a node level measure, denoted as $D(i|n)$, is introduced which essentially measures the average influence or impact of node n on all the paths from node i to all other nodes, defined as:

$$D(i|n) = \frac{1}{N-1} \sum_{j \in \mathcal{N}^{-n}} D(i \to j|n),$$

which, with (3) applied, can be further written as

$$D(i|n) = \frac{1}{N-1} \sum_{j \in \mathcal{N}^{-n}} (I_{ij} - I_{ij}^{-n}). \tag{4}$$

C. Unification of the Existing Measures

At a first glance, the network level measure $\Delta(G|n)$ seems to be irrelevant to the two dependency measures $D(i \to j|n)$ and $D(i|n)$. In the following, we first extend the latter measures to a network level dependency measure, denoted as $D(G|n)$. Then, we prove the equivalency in ranking nodes between $\Delta(G|n)$ and the new network level dependency measure $D(G|n)$. Through this, the three measures, $\Delta(G|n)$, $D(i \to j|n)$ and $D(i|n)$ are unified.

As defined, $D(i \to j|n)$ is a measure of the influence or impact of node n on the path from node i to node j, based on which $D(i|n)$ essentially measures the average influence of node n on all the paths from node i to all other nodes. Since in (4), i can be any node in $\mathcal{N}$, then by taking average over all N such choices, we can extend the node level measure to a network level measure of the impact to all nodes as:

$$D(G|n) = \frac{1}{N} \sum_{i=1}^{N} D(i|n). \tag{5}$$

The following theorem summarizes the equivalency between $\Delta(G|n)$ and $D(G|n)$.

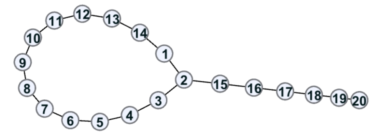

Fig. 1. Tadpole network

TABLE I
PATH DEPENDENCY MEASURES: TADPOLE NETWORK IN FIG. 1

j	3	4	5	8-14	15	16	17	18	19
$D(1 \to j\|2)$	0.42	0.24	0.15	0	0.5	0.66	0.25	0.2	0.17
$DI(1 \to j\|2)$	0.42	0.24	0.15	0	1	1	1	1	1

Theorem 1. *The ranking result of nodes based on $\Delta(G|n)$ is the same as that based on $D(G|n)$.*

Proof. Note that in the definition of $\Delta(G|n)$, $E(G)$ is the same for all nodes. Hence, the ranking result of nodes based on $\Delta(G|n)$ is the same as that based on $E(G) - E(G^{-n})$, which after applying the definition of efficiency becomes

$$E(G) - E(G^{-n}) = \frac{1}{N(N-1)} \sum_{\forall i \neq j} (I_{ij} - I_{ij}^{-n}). \tag{6}$$

For $D(G|n)$, by applying (4) and (3), it becomes:

$$D(G|n) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i; j=1}^{N} (I_{ij} - I_{ij}^{-n}). \tag{7}$$

A closer check on the right hand side of (6) and that of (7) reveals that they are indeed equal, because $\sum_{\forall i \neq j} (I_{ij} - I_{ij}^{-n}) = \sum_{i=1}^{N} \sum_{j \neq i; j=1}^{N} (I_{ij} - I_{ij}^{-n})$. Hence, we have

$$D(G|n) = E(G) - E(G^{-n}), \tag{8}$$

which concludes the proof. $\square$

D. The Limitation

As shown by their expressions, the three dependency impact measures $D(i \to j|n)$, $D(i|n)$ and $\Delta(G|n)$ (or equivalently $D(G|n) = E(G) - E(G^{-n})$ as discussed above) are all defined on $I_{ij} - I_{ij}^{-n}$. However, $I_{ij} - I_{ij}^{-n}$ inherently has a limitation, due to overlooking the possible fragmentation effect on the network after the deactivation of node n.

Specifically, if node j is unreachable to node i after failure or deactivation of node n, then the value of $I_{ij} - I_{ij}^{-n}$ or $D(i \to j|n)$ becomes I_{ij}, since in this case $I_{ij}^{-n} = 0$ by definition. As a result, for such cases, $D(i \to j|n) = I_{ij} - I_{ij}^{-n}$ only depends on how far node j is from node i in the presence of node n, i.e., on the value of d_{ij}, and if d_{ij} is large, $D(i \to j|n)$ will be small, giving an impression that the impact of node n on the path is small, contradicting to the fact that the path between i and j is unexistent after deactivation of n.

To demonstrate this limitation, consider a tadpole network shown in Fig. 1. In Table I, $D(1 \to j|2)$ for different j is shown, which indicates how each path starting with node 1 is impacted by the deactivation of node 2. Though simple, several surprising observations are revealed by the example.

First, as can be observed from the figure, the nodes 15 to 20 will all be unavailable to node 1 after the failure of node 2. However, Table I shows that for paths of $1 \to 15, \ldots, 1 \to 20$, their $D(1 \to j|2)$ values fall within the range $(0, 1)$ and differ from each other.

Second, $D(1 \to j|2)$ of the paths $1 \to 3$ and $1 \to 4$ are higher than of the paths $1 \to 18$ and $1 \to 19$, even though for the latter two, no path exists any more while for the former two, a path still exists after the deactivation of node 2.

Third, for $1 \to 8$ to $1 \to 14$, as easily verified from the figure, they are independent of node 2. This is reflected by their $D(1 \to j|2)$ being 0. In other words, the value 0 would naturally be considered as an indication of such independency. However, as implied by Table I, by adding m nodes sequentially to node 20, we would get for the new end node, a $D(1 \to j|2)$ value equal to $1/(7+m)$, which approaches 0 when m becomes large. In other words, the current way of calculating $D(1 \to j|2)$ gives an impression that the farther a node were from node 2, the less the node would be dependent on node 2, which is wrong.

The above observations assert that $I_{ij} - I_{ij}^{-n}$ has an evident limitation for being used as the basis in quantifying the dependency impact. Since $D(i \to j|n)$, $D(i|n)$, $D(G|n)$ and $\Delta(G|n)$ are all formulated based on $I_{ij} - I_{ij}^{-n}$, using them to index the dependency impact is void. To address this limitation, we propose a set of new measures for the purpose.

III. The Proposed Measures: Dependency Indexes

In order to take into account the fragmentation effect, the proposed measures quantify the dependency level not only based on the information measure but also on the availability of nodes. For this, a binary indicator variable A_{ij}^{-n} is used to measure the availability of node i to node j after failure of node n: $A_{ij}^{-n} = 1$ if node j is reachable to node i, and $A_{ij}^{-n} = 0$ if node j is unreachable.

The *path dependency index*, denoted as $DI(i \to j|n)$, which measures the dependency of the path $i \to j$ on node n is defined as:

$$DI(i \to j|n) \equiv \begin{cases} I_{ij} - I_{ij}^{-n} & \text{if } A_{ij}^{-n} = 1 \\ 1 & \text{if } A_{ij}^{-n} = 0. \end{cases} \quad (9)$$

There are three possible cases. One is, node j is unreachable to node i after failure of node n. In this case, the path $(i \to j)$ is totally dependent on node n and will be assigned the maximum dependency level of one. The second case is node j is reachable and there is no change in the path length. This implies that the path $(i \to j)$ is independent of node n, so $I_{ij} = I_{ij}^{-n}$ and $DI(i \to j|n) = 0$. The third case is that node j is reachable but the length of the path has increased, then the path dependency index will have a value in the range $(0, 1)$.

The *node dependency index*, denoted as $DI(i|n)$, measures the average level of dependency that node i has on node n for connecting to the other nodes, which is calculated from the path dependency index as:

$$DI(i|n) = \frac{1}{N-1} \sum_{j \in \mathcal{N}^{-n}/i \neq j} DI(i \to j|n). \quad (10)$$

There are also three possible cases. $DI(i|n) = 1$ means node i is totally dependent on node n: It is not able to connect to any other node in the network after failure of node n. $DI(i|n) = 0$ implies that node i is independent of node n, i.e. node i does not observe any connectivity change, both in terms of path length and availability. For $0 < DI(i|n) < 1$, it implies that the connectivity of node i to the rest of the nodes is affected but it can still reach at least one other node in the network.

TABLE II
Node dependency indexes: Tadpole network in Fig. 1

n \ i	1	2	6	9	15	17	19	20
2	0.38	X	0.34	0.33	0.72	0.72	0.72	0.72
9	0	0	0.02	X	0	0	0	0
15	0.27	0.27	0.27	0.27	X	0.78	0.78	0.78
19	0.05	0.05	0.05	0.05	0.05	0.05	X	1

The *network dependency index*, denoted as $DI(G|n)$, measures the average level of dependency the network G has on node n. That is $DI(G|n)$ measures the average dependency of the nodes in $\mathcal{N}^{-n}$ on node n. The network dependency index is hence calculated from the node dependency index as.

$$DI(G|n) = \frac{1}{N-1} \sum_{j \in \mathcal{N}^{-n}} DI(i|n). \quad (11)$$

Theorem 2. *If the failure of node n fragments the network G into M sub-networks where each sub-network G^m has $\alpha_m N$ number of nodes, $0 < \alpha_m < 1$, then:*

$$DI(G|n) \geq \sum_{m \in \{1..M\}} \alpha_m \Big(\sum_{k \in \{1..M\}/k \neq m} \alpha_k \Big). \quad (12)$$

Proof. A node i in a sub-network G^m will not be able to connect with the nodes in the other sub-networks after failure of node n. Let set $\mathcal{M} = \{1..M\}$. Thus,

$$DI(i|n) \geq \frac{1}{N-1} \sum_{k \in \mathcal{M} \setminus m} \alpha_k N, \quad (13)$$

where only the effect of the $\alpha_k N$ unavailable paths in each G^k, $(k \neq m)$, with $A_{ij}^{-n} = 0$ and hence $DI(i \to j|n) = 1$ is counted. Similarly, we have

$$DI(G|n) \geq \frac{1}{N-1} \sum_{m \in \mathcal{M}} \alpha_m N \Big(\frac{1}{N-1} \sum_{k \in \mathcal{M} \setminus m} \alpha_k N \Big), \quad (14)$$

which, with $N/(N-1) \geq 1$, gives (12) and concludes. □

If all the sub-networks have equal number of nodes, i.e., $\alpha_m = \alpha$, or α is a lower bound on α_m, $\forall m$, we get

$$DI(G|n) \geq M(M-1)\alpha^2. \quad (15)$$

As a special case with $M = N$, the lower bound becomes $(N-1)/N \approx 1$ for large N. An example is a star network, where the failure of the central node makes all other nodes disconnected, so the network fully depends on the central node, i.e. $DI(G|n) = 1$, and the lower bound is approached.

Remark: As implied by their formulations, the time complexity for calculating the proposed dependency indexes is the same as for calculating the corresponding measures (3), (4) and (2) proposed in [1] and [8].

IV. Results

This section presents results of the proposed dependency indexes and an application to critical node identification.

A. Dependency Indexes

For the path dependency index, $DI(1 \to j|2)$ is shown for the tadpole network also in Table I, in comparison with $D(1 \to j|2)$. From $DI(1 \to j|2)$, we see for paths $1 \to 15$, …, $1 \to 20$, its value is 1, implying dependency of these paths on node 2 as verified from the topology, which, however, is not reflected by $D(1 \to j|2)$ as discussed in Sec. II-D.

TABLE III
NETWORK DEPENDENCY AND RANKING OF CRITICAL NODES IN THE TADPOLE NETWORK (FIG. 1)

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$D(G\|n)$	0.022	**0.081**	0.022	0.018	0.014	0.012	0.011	0.010	0.010	0.010	0.011	0.012	0.014	0.018	**0.059**	**0.05**	**0.039**	**0.028**	0.015	0
Rank	(6)	**(1)**	(6)	(7)	(9)	(10)	(11)	(12)	(12)	(12)	(11)	(10)	(9)	(7)	**(2)**	**(3)**	**(4)**	**(5)**	(8)	(13)
$DI(G\|n)$	0.023	**0.467**	0.023	0.019	0.015	0.013	0.012	0.011	0.011	0.011	0.012	0.013	0.015	0.019	**0.41**	**0.35**	**0.28**	**0.19**	0.1	0
Rank	(7)	**(1)**	(7)	(8)	(9)	(10)	(11)	(12)	(12)	(12)	(11)	(10)	(9)	(8)	**(2)**	**(3)**	**(4)**	**(5)**	(6)	(13)
Degree	2	**3**	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1
Rank	(2)	**(1)**	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(3)

For the node dependency index, $DI(i|n)$, it is exemplified in Table II also using the tadpole network (Fig. 1). For example, the entry on column two ($i = 2$) and row three ($n = 15$) shows the value of the node dependency index $DI(2|15)$, i.e., the dependency level that node 2 has on node 15. In addition, the table shows $DI(20|19) = 1$, meaning node 20 is totally dependent on node 19, as implied by the topology.

For the network dependency index, $DI(G|n)$, it is shown and compared with $D(G|n)$ in Table III for the tadpole network. While not surprisingly, the values of $DI(G|n)$ and $D(G|n)$ are close for nodes 1, and 3 – 14, their difference is high for the other nodes. This is essentially due to the value differences in the underlying path level dependency indexes $DI(i|n)$ and $D(i|n)$ discussed above.

As a highlight, the network dependency index of node 2 is $DI(G|2) = 0.467$. Note that failure of node 2 fragments the network into two sub-networks. According to Theorem 2, the network dependency index of node two is lowered bounded by 0.44, which is very close to the actual value 0.467.

B. Application to Critical Node Identification

For the tadpole network shown in Fig. 1, Table III also compares the ranking results for critical node identification using $D(G|n)$, $DI(G|n)$ and node degree. Specifically, all these measures rank node 2 as the first, i.e. the most critical node, and node 20 the least. For the other nodes, node degree is unable to distinguish as it gives equal rank for them. $D(G|n)$ and $DI(G|n)$ place nodes 15-18 in the same order in the ranking, from 2^{nd} to 5^{th}. This ranking result is intuitive, since from the figure, their failure results in unavailability of some nodes, and the number of unavailable nodes becomes smaller in the same order. However, from node 19, the ranking becomes different. While $DI(G|n)$ still follows the same intuition and ranks node 19 at the 6^{th} place since its failure will make node 20 unavailable, $D(G|n)$ puts node 19 in the 8^{th} rank after several other nodes, which are nodes 1 and 3 (6^{th}) and nodes 4 and 14 (7^{th}) even though failure of any of these four nodes does not cause unavailability of others.

To examine the ranking difference further, a larger network as shown in Fig. 2 is considered. This network is a randomly generated scale-free network, i.e. a network with power-law degree distribution, with 500 nodes and 998 links using the Barabasi-Albert model [10]. The five most critical nodes, identified by $DI(G|n)$, are numbered from 1 to 5 corresponding to their criticality level and marked in Fig. 2. For presentation simplicity, these numbers are also used as their node numbers in the following discussion. Visually, this ranking follows the same intuition as discussed for the tadpole network, i.e. a node whose failure causes the unavailability of more nodes should be ranked higher. However, if $D(G|n)$ were used, the ranking order among the five nodes would become 1, 2, 5, 3 and 4, even though the failure of node 5 clearly leads to unavailability of fewer nodes compared to that of node 3 or 4.

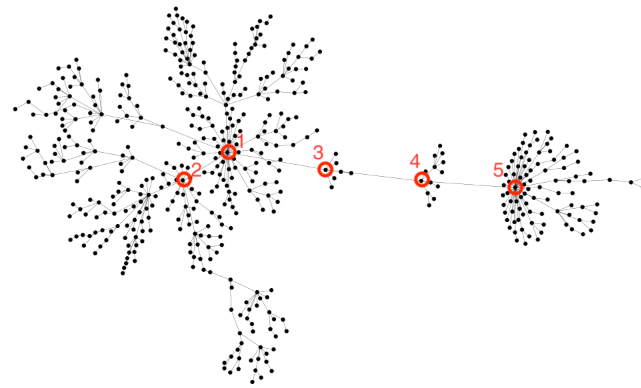

Fig. 2. Scale-free network: 500 nodes, 998 links

V. CONCLUSION

In this paper, the limitation of the related existing dependency measures is demonstrated and discussed. To address this limitation, a set of new measures are proposed, which assess structural dependencies at the path, node and network level. In particular, they capture fragmentation effects and hence it is possible to get, from their values, insights into the extent of fragmentation that the failure of a node will cause. The results also show that the proposed measures are better suitable for critical node identification reflecting the fragmentation effect.

VI. ACKNOWLEDGMENT

The work for this paper was performed in the context of the EU FP7 Marie Curie Actions project Grant Agreement No. 607584 (the Cleansky project).

REFERENCES

[1] V. Latora and M. Marchiori, “A measure of centrality based on network efficiency,” *New Journal of Physics*, vol. 9, no. 6, p. 188, 2007.
[2] M. Lalou *et al.*, “The critical node detection problem in networks: A survey,” *Computer Science Review*, vol. 28, pp. 92–117, 2018.
[3] T. N. Dinh *et al.*, “On new approaches of assessing network vulnerability: hardness and approximation,” *IEEE/ACM TON*, pp. 609–619, 2012.
[4] Y. Shen *et al.*, “On the discovery of critical links and nodes for assessing network vulnerability,” *IEEE/ACM TON*, vol. 21, pp. 963–973, 2013.
[5] D. F. Rueda *et al.*, “Robustness comparison of 15 real telecommunication networks: Structural and centrality measurements,” *Journal of Network and Systems Management*, vol. 25, pp. 269–289, 2017.
[6] A. Bavelas, “A mathematical model for group structures,” *Human organization*, vol. 7, no. 3, pp. 16–30, 1948.
[7] K. Stephenson and M. Zelen, “Rethinking centrality: Methods and examples,” *Social networks*, vol. 11, no. 1, pp. 1–37, 1989.
[8] D. Y. Kenett *et al.*, “Dependency network and node influence: Application to the study of financial markets,” *International Journal of Bifurcation and Chaos*, vol. 22, no. 07, p. 1250181, 2012.
[9] V. Latora and M. Marchiori, “Efficient behavior of small-world networks,” *Physical review letters*, vol. 87, no. 19, p. 198701, 2001.
[10] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks,” *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.