Several mathematical methods for identifying crucial nodes in networks

WenJun Zhang
School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China; International Academy of Ecology and Environmental Sciences, Hong Kong
E-mail: zhwj@mail.sysu.edu.cn, wjzhang@iaees.org

Received 9 June 2012; Accepted 15 September 2012; Published online 1 December 2012
IAEES

Abstract
Crucial nodes in a network refer to those nodes that their existence is so important in preserving topological structure of the network and they independently determine the network structure. In this study I introduced and proposed several mathematical methods for identifying crucial nodes in networks. They fall into three categories, node perturbation, network analysis, and network dynamics. Node perturbation methods include adjacency matrix index, degree or flow change index, node perturbation index, etc. Network dynamics methods include network evolution modeling, etc. Network analysis methods include node degree, criticality index, branch flourishing index, node importance index, etc. Advantages and advantages of these methods were discussed. Finally, I suggested that some of these methods may also be used to identify crucial links (connections) in networks. In this case, the change of a link refers to presence/absence of a link, or change of flow in the link, etc.

Keywords networks; crucial nodes; identification; node perturbation; network dynamics; network analysis; crucial links (connections); mathematical methods.

1 Introduction
Crucial nodes are a few of nodes that govern the structure of a network (Junker, 2006). Their missing or even small changes will substantially change the network. Identification of crucial nodes is a fundamental problem in network analysis (Pimm et al., 1991; Montoya et al., 2006; Butts, 2009; Ding, 2012). In this study, I introduced or proposed several mathematical methods for identifying crucial nodes in networks based on previous studies. These methods can be further used in a wider area of network science, such as cancer networks, metabolic networks, etc (Krogan et al., 2006; Ibrahim et al., 2011; Tacutu et al., 2011; Budovsky and Fraifeld, 2012).

2 Methods
Firstly I define the crucial nodes in a network as that with the following features:
(1) Their existence is so important in preserving topological structure of the network (Zhang, 2012a);
(2) They independently determine network structure;
(3) These nodes are closely related to other nodes in the network.
Several mathematical methods for identifying crucial nodes in networks are described as follows.

2.1 Node perturbation index
I define node perturbation index (NP) as
\[ NP = \frac{dN}{dn}/N \]

or

\[ NP = dN/dn \]

where \( N \): measure of network structure; \( n \): state value or proportion of a known node in the network. Theoretically, \( NP \) of all nodes in the network are normally distributed, i.e., \( NP \approx 0 \) for most nodes. Crucial nodes have \( NP \) much larger or less than 0.

There are many measures of network structure, i.e., total links, total number of nodes, network flow (Latham, 2006), degree distribution (Zhang, 2011; Zhang and Zhan, 2011), aggregation index, coefficient of variation, entropy (Zhang and Zhan, 2011; Zhang, 2012a), and other measures (Paine, 1992; Power et al., 1996; Dunne et al., 2002; Montoya and Sole, 2003; Allesina et al., 2005; Barabasi, 2009).

Another definition of \( NP \) is

\[ NP = \frac{(N_t - N_0)}{N_0/n_0} \]

or

\[ NP = \frac{(N_t - N_0)}{N_0} \]

where \( N_t, N_0 \): measure of network structure after and before a node is completely removed from the network, respectively; \( n_0 \): state value or proportion of the node in the network before the node is removed from the network; \( NP \approx 1 \), if the functionality of the node is positively proportional to its state value or proportion in the network; \( NP \approx -1 \), if the functionality of the node is negatively proportional to its state value or proportion in the network; \( NP \gg 1 \), if the node is a crucial node.

Node perturbation index, \( NP \), is a general index that can be further materialized in various ways.

2.2 Criticality index

Criticality index is defined as

\[ CI_i = \sum_{c=1}^{n} \left( 1 + C_{bc} \right) d_c + \sum_{e=1}^{m} \left( 1 + C_{fe} \right) f_e \]

where \( CI_i \): value of criticality index of node \( i \); \( n \): the number of source nodes directing to target node \( i \); \( d_c \): the number of target nodes of the \( c \)-th source node, and \( C_{bc} \): the backward-oriented criticality index of the \( c \)-th source node. Similarly, \( m \): the number of target nodes of source node \( i \); \( f_e \): the number of source nodes of the \( e \)-th target node, and \( C_{fe} \): the forward-oriented criticality index of the \( e \)-th target node.

The nodes with larger \( CI \) tend to be crucial nodes. This index is characterized by the following features: (1) considering both forward- and backward-oriented between-node relations; (2) only nodes within the same network can be compared for their relative importance.

I define this index based on the keystone index, etc (Jordán et al., 1999, 2006; Jordán, 2001; Zhang, 2012a).

2.3 Degree change index

I define degree change index as

\[ DC_i = \sum_{j=1}^{n} \left| \frac{(O_{ij} - O_{ij_0})}{O_{ij_0}} \right| + \left| \frac{(I_{ij} - I_{ij_0})}{I_{ij_0}} \right| \]
or

$$DC_i = \sum_{j=1}^{v} (|O_{jt} - O_{0j}| + |I_{jt} - I_{0j}|)$$

where $DC_i$: value of degree change index of node $i$; $v$: total number of nodes in the network; $O_{jt}, O_{0j}$: out-degree of node $j$ after and before node $i$ is changed respectively; $I_{jt}, I_{0j}$: in-degree of node $j$ after and before node $i$ is changed respectively.

The nodes with larger $DC$ tend to be crucial nodes.

2.4 Flow change index

I define flow change index as

$$FC_i = \sum_{j=1}^{v} \left( \frac{|O_{jt} - O_{0j}|}{O_{0j}} + \frac{|I_{jt} - I_{0j}|}{I_{0j}} \right)$$

or

$$FC_i = \sum_{j=1}^{v} (|O_{jt} - O_{0j}| + |I_{jt} - I_{0j}|)$$

where $FC_i$: value of flow change index of node $i$; $v$: total number of nodes in the network; $O_{jt}, O_{0j}$: outflow of node $j$ after and before node $i$ is changed respectively; $I_{jt}, I_{0j}$: influx of node $j$ after and before node $i$ is changed respectively. The nodes with larger $FC$ tend to be crucial nodes.

Another flow change index is defined as

$$FC_k = \sum_{i,j} |f_{ijt} - f_{ij0}|$$

where $f_{ijt}, f_{ij0}$: flow between node $i$ and $j$ after and before node $k$ is changed. The nodes with larger $FC$ tend to be crucial nodes.

2.5 Adjacency matrix index

Following the definition of Zhang (2012a), suppose the adjacency matrix of a network with $v$ nodes is $D = (d_{ij})_{v \times v}$. If $d_{ij} = d_{ji} = 0$, then there is no connection from $v_i$ to $v_j$; if $d_{ij} = -d_{ji}$, and $|d_{ij}| = 1$, then there is only a directed connection from $v_i$ to $v_j$; if $d_{ij} = d_{ji} = 1$, then there is only an undirected connection from $v_i$ to $v_j$; if $d_{ij} = d_{ji} = 2$, then there are two parallel connections from $v_i$ to $v_j$; if $d_{ij} = 3$, then $v_i$ has a loop; if $d_{ii} = 4$, then $v_i$ is an isolated node; if $d_{ii} = 5$, then $v_i$ is an isolated node and it has a loop. $i, j = 1, 2, \ldots, v$.

I define adjacency matrix index as

$$AD_k = \sum_{i,j} |d_{ijt} - d_{ij0}|$$

where $d_{ijt}, d_{ij0}$: value of the element $d_{ij}$ after and before node $k$ is changed. The nodes with larger $AD$ tend to be crucial nodes.

2.6 Centrality index

Centrality indices are widely used (Scardoni and Laudanna, 2012). The first centrality index is betweenness centrality (Navia et al., 2010). It measures how central a given node is in terms of being adjacent to many shortest paths in the network. It is based on quantifying how often node $i$ is on the shortest path between each
pair of nodes \( j \) and \( k \). The standardized centrality index for node \( i \) is

\[
C_i = 2 \sum_{j<k} g_{jk}(i)/g_{jk}/[(v-1)(v-2)]
\]

where \( i \neq j \) and \( k \), \( g_{jk} \) is the number of equally shortest paths between nodes \( j \) and \( k \), and \( g_{jk}(i) \) is the number of these shortest paths to which node \( i \) is adjacent, \( v \) is the total number of nodes. The denominator is twice the number of pairs of nodes without node \( i \). If \( C_i \) is large for trophic group \( i \), the loss of this node will have many rapidly spreading effects in the network.

The second centrality index is closeness centrality. It measures how close a node is to the rest of nodes. It is based on the proximity principle and quantifies how short the minimal paths from a given node to all other nodes are (Wassermann and Faust, 1994). The standardized form is

\[
CC_i = (v-1)/\sum_{j=1}^{v-1} d_{ij}
\]

where \( i \neq j \), and \( d_{ij} \) is the length of the shortest path between nodes \( i \) and \( j \) in the network. The smallest value of \( CC_i \) will be for that trophic group that upon being removed will affect the majority of other groups.

2.7 Branch flourishing index

I define the branch flourishing index of a node as

\[
BF_i = \sum_{j \neq i} (n_{ij} \times m_{lij})
\]

where \( BF_i \): branch flourishing index of the node \( i \); \( n_{ij} \): the total number of paths (chains) between nodes \( i \) and \( j \). \( m_{lij} \): the mean path (chain) length of all paths (chains) between nodes \( i \) and \( j \), \( j \neq i \); \( v \): the total number of nodes in the network.

The nodes with larger \( NS \) tend to be crucial nodes.

2.8 Node degree

Node degree (number of connections of node) is always treated as the simplest index for measuring node importance. The nodes with more links tend to be crucial nodes.

2.9 Connections and between-node connection strength

Various measures on strength (e.g., correlation such as linear correlation, partial correlation, Spearman correlation) and number of connections (interactions) can be used to determine crucial nodes (Paine, 1980; Zhang, 2007, 2011, 2012b; Ding, 2012). For the statistic networks (Zhang, 2012b), a node with more connections (\( d_i \)) and larger mean correlation (\( mc_i \)) tends to be a crucial node. For example, we may judge the nodes with both connections and mean correlation larger than that of 95% of other nodes as crucial nodes in the network. A simple index for this criterion is

\[
CS_i = d_i \times mc_i
\]

Here I propose another index, node importance index, for identifying crucial nodes in statistic networks

\[
SC_i = \sum_{j \neq i} d_{ij}
\]
where $SC_i$: node importance index of the node $i$; $d_{ij}$: the path (chain) strength between nodes $i$ and $j$ in the network, $j \neq i$; $v$: the total number of nodes in the network. $d_i$ can be defined in different ways. For example,

$$d_i = \max \prod \left| r_{kl} \right|$$

where $r_{kl}$: the correlation between nodes $k$ and $l$ in the path (chain) $t$ between nodes $i$ and $j$, $t = 1, 2, \ldots, n_{ij}$; $n_{ij}$: the total number of paths (chains) between nodes $i$ and $j$. The nodes with larger $SC$ tend to be crucial nodes.

### 2.10 Network evolution method

Network evolution modeling (Zhang, 2012c) can be used to find crucial nodes. The nodes that cause greater changes of network structure during network evolution are crucial nodes. Sensitivity analysis can be conducted in network evolution modeling to find crucial nodes. For example, we may change the sequence and time of a node joining the network to investigate its impact on the network.

Other evolution (or succession) methods can also be used (Bond, 1989; Rossberg et al., 2005).

### 3 Discussion

Above methods fall into three categories, node perturbation, network analysis, and network dynamics. Node perturbation methods, such as adjacency matrix index, degree or flow change index, node perturbation index, etc., identify crucial nodes by comparing structural changes of the network resulted from changes of each node. Therefore these methods need a large amount of experiments. From the view of definition of crucial node, however, they are highly reliable methods. Network dynamics methods include network evolution modeling (e.g., community assembly modeling), etc. These methods need to have a deep insight into mechanism of network dynamics and need to build an ideal model for network evolution. They are also high reliable but a lot of works should be done before they can normally function. Network analysis methods, like node degree, criticality index, centrality index, branch flourishing index, etc., need the information of network itself only, and thus cost much less than other methods. Nevertheless, they identify crucial nodes only by analyzing static connection structure of nodes and are thus less reliable than other methods. Connection strength-connection number method (e.g., node importance index) above is mainly a network analysis method. However, if the connection strength is measured by between-node correlation in the process of network evolution, it is then a network dynamics method.

Some of these methods may also be used to identify crucial links (connections) in networks. In this case, the change of a link refers to presence/absence of a link, or change of flow in the link, etc.

### References


Junker BH, Koschutzki D, Schreiber F. 2006. Exploration of biological network centralities with CentiBiN. BMC Bioinformatics, 7: 219
Navia AF, Cortés E, Mejia-Falla PA. 2010. Topological analysis of the ecological importance of elasmobranch fishes: A food web study on the Gulf of Tortugas, Colombia. Ecological Modelling, 221: 2918-2926