Generating function technique in complex networks

Ming Li and Bing-Hong Wang*
Department of Modern Physics, University of Science and Technology of China, Hefei 230026,
People’s Republic of China
E-mail: *bhwang@ustc.edu.cn

Abstract. Complex network is a useful tool in describing the interactions between different agents of a complex system, and has attracted considerable interest recently. There are various approaches to studying the structure and dynamic of complex networks for the diverse and complicated connectivity structure. One more applicable theoretical approach is the generating function technique, which is useful for the studies of network structure and dynamic, especially for tree-like networks. In this paper, we will give a summary of the basic ideas of this approach, and explore the structure of networks. A cascading failures model we have proposed before will also be presented as an application of this approach.

1. Introduction
Network is a broad concept which can come into various aspects of our lives, such as communication network, Internet, relationship network, neural network and metabolic network. Complex network theory is such a uniform research platform for the topology structure and dynamic of networks in different domains[1]. For example, epidemic disease among people, spread of rumours online and cascading failures on the transport network can all be studied by the network percolation model[2, 3, 4].

As an interdiscipline subject, the methods and models available for complex networks are varied. One of the important theoretical approaches is the generating function technique[5, 6, 7], which is based on the branching process. This approach is often used to study the structure of tree-like networks, and the propagating processes that can be mapped to a network percolation problem[6, 7], such as epidemic and cascading failures. In other words, the generating function technique is a mean field method for the percolation on tree-like random networks. In addition, this approach is concerned with a network ensemble with a given degree distribution, not a network with given connection. The results obtained by this technique are the average of the ensemble, which will be consistent with the simulation (or empirical) results on a special network with size \( N \to \infty \).

In the paper, we will give a brief review and discussion of the generating function technique with some examples.

2. Method
2.1. generating function
Generating function sometimes refers to as z-transform is widely used in mathematics, and plays an important role in probability theory[5]. Considering a sequence \( a_n, n = 0, 1, 2, \ldots \), the
The corresponding generating function is defined as

\[ G(x) = \sum_{n=0}^{\infty} a_n x^n. \]  

(1)

For many well-defined sequences, \( G(x) \) can be written in a closed form, and the coefficient \( a_n \) in the sequence can be recovered by Taylor expand,

\[ a_n = \frac{1}{n!} \left. \frac{d^n G(x)}{dx^n} \right|_{x=0}. \]  

(2)

Considering a discrete random variable \( A \), taking non-negative integer values. If \( a_n \) in eq.(1) is the probability that \( A \) takes the value \( n (= 0, 1, 2, \ldots) \), \( G(x) \) is also called as probability generating function. In this way, the expected value of \( A \) can be expressed as \( G'(1) \). In addition, a probability generating function \( G(x) \) must satisfy \( G(1) = 1 \). Taking Poisson distribution \( a_n = \lambda^n e^{-\lambda} / n! \) as an example, the probability generating function is \( G(x) = e^{\lambda(x-1)} \). Obviously, we have \( G(1) = e^{\lambda(x-1)}|_{x=1} = 1 \) and \( G'(1) = \lambda e^{\lambda(x-1)}|_{x=1} = \lambda \).

### 2.2. generating function in complex networks

Next, we turn our attention to complex networks. In a network, the focus is usually concentrated on the states of nodes, which generally depend on the states of the links connected to them. For example, if all the links of a node are not connected to the giant component of the network, the node must not belong to the giant component.

For convenience, let \( x \) be the probability that a link is in state \( X \), where \( X \) is relevant to the problem we considered. Further, we hypothesize that the states of links are independent of each other. Then, the probability that all the links of a node with degree \( k \) are in state \( X \) is \( x^k \). Averaging this probability over the degree distribution of the network, we can obtain the fraction of the nodes with all their links in state \( X \), in other words, the probability that all the links of a randomly chosen node are in state \( X \). According to the definition of the generating function (1), this fraction (or probability) can be written as

\[ G_0(x) = \sum_{k=0}^{\infty} p_k x^k. \]  

(3)

Let \( x = 1 \), thus \( G_0(1) = \sum_{k=0}^{\infty} p_k = 1 \). This can be understood as that when the state of link is sure to \( X \), this fraction (or probability) must be 1. Furthermore, the average degree of the network can be expressed as

\[ G'_0(1) = \sum_{k=0}^{\infty} p_k k = \langle k \rangle. \]  

(4)

In this way, the average degree could be understood as the rate of change of the node state with link.

More generally, for a node with degree \( k \), the probability that there are only \( i (\leq k) \) links are in state \( X \) can be written as

\[ \binom{k}{i} x^i (1-x)^{k-i}. \]  

(5)

Using the degree distribution of the network, the probability that a randomly chosen node are in this state (or the fraction of nodes in this state) can be written as

\[ \sum_{k=i}^{\infty} p_k \binom{k}{i} x^i (1-x)^{k-i} = \frac{(-x)^i}{i!} \frac{d^i}{dx^i} G_0(1-x). \]  

(6)
This function is more general than eq.(3) to represent the state of a node. However, it must be pointed that eq.(5) can reduce to $x^k$ by letting $i = k$, but eq.(6) can not reduce to eq.(3).

Functions (3) and (6) describe the state of a randomly chosen node in a network. However, the nodes considered in practical network problems are not always chosen randomly, but rather ones reached by following randomly chosen links. So, we need to further explore the state of such a node.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{The generating functions for node states. Each link in the network is in state $X$ with probability $x$. In the figure, links of state $X$ are represented by solid lines, and empty lines for those not in state $X$. (a) All links of a randomly chosen node are in state $X$. (b) Only $i$ links of a randomly chosen node are in state $X$. (c) All the other links of a node reached by following a randomly chosen link are in state $X$. (d) Only $i$ links in the other links of a node reached by following a randomly chosen link are in state $X$.}
\end{figure}

Supposing the degree of the node reached by following a randomly chosen link is $k$, then, the probability that all the other links of this node are in state $X$ is $x^{k-1}$. This probability can be used to describe the state of this node. Note that the link used for reaching this node must be ruled out. To obtain the average probability as eq.(3), we need to know the distribution of the node degree reached by following a randomly chosen link, the so-called excess degree distribution $q_k$. Obviously, $q_k$ depends on the degree distribution $p_k$, the larger $p_k$ is, the larger $q_k$ is. In addition, following a link, it is easier to reach a node with larger $k$. In consideration of the above two points, we have

$$q_k \propto p_k k.$$  \hspace{1cm} (7)

The normalized form of $q_k$ is

$$q_k = \frac{p_k k}{\sum_k p_k k} = \frac{p_k k}{\langle k \rangle}.$$ \hspace{1cm} (8)

Therefore, the average probability that all the other links of a node reached by following a
randomly chosen link are in state \( X \) is

\[
G_1(x) = \sum_{k=1}^{\infty} q_k x^{k-1} = \sum_{k=1}^{\infty} \frac{p_k}{(k)} x^{k-1}.
\]  

(9)

Note that the smallest degree of the node reached by following a link is 1. Obviously, eq.(9) satisfies \( G_1(1) = 1 \) and \( G_1(x) = G_0'(x)/G_0'(1) \).

Furthermore, similar with eq.(6), the average probability that only \( i(\leq k - 1) \) links in the other links of a node reached by following a randomly chosen link are in state \( X \), can be written as

\[
\sum_{k=i+1}^{\infty} q_k \binom{k-1}{i} x^i (1-x)^{k-1-i} = \frac{(-x)^i}{i!} \frac{d^i}{dx^i} G_1(1-x).
\]  

(10)

Now that we have represented the common node states in a network shown graphically in fig.1. It must be pointed that all these functions are for the case that the states of links are independent of each other. For networks with correlated links, these functions do not apply. However, an approximate expression can be obtained by revising the functions shown in fig.1 for some special cases[8, 9, 10], this point is not considered here.

3. Network structure

3.1. hierarchical structure of networks

Hierarchical structure is one of the important characters of real networks[11]. In this section, we will give a brief discussion on the hierarchical structure of networks using generating function.

![Figure 2. Hierarchical structure of networks. The number of nodes in the \((m + 1)\)-th layer is equal to the total excess degrees of the nodes in the \(m\)-th layer.](image)

As shown in fig.2, we choose a node randomly as the center of the network, and then the direct neighbors of this node are the first layer of the network, and the neighbors’ neighbors are the second layer, and so on. Since the central node is chosen randomly, its state can be described by the generating function \( G_0(x) \). Here, state \( X \) means the connection of links. Following the links of the central node, we will get the first layer of the network. As mentioned last section, the generating function \( G_1(x) \) can represent the state of the node reached by
following a link. So, replacing \(x\) in \(G_0(x)\) with \(G_1(x)\), we will obtain the state of the first layer \(G^{(1)}(x) = G_0(G_1(x))\). Similarly, the state of the second layer can be represented as \(G^{(2)}(x) = G_0(G_1(G_1(x)))\). Generally, the state of the \(m\)-th layer can be written as 

\[
G^{(m)}(x) = G_0(G_1(...G_1(x)...)) = G^{(m-1)}(G_1(x)), \quad m > 1. \tag{11}
\]

Note that the state of links must be independent of each other, i.e., loop is forbidden.

For the state function \(G^{(m)}(x)\) of the \(m\)-th layer, the integer power exponent of \(x\) must be the number of links that connect to the nodes in the \((m+1)\)-th layer, since \(x\) represents the states of the excess links of the \(m\)-th layer. Therefore, the derivative of eq.(12) with respect to \(x\) at \(x = 1\) must equal the total excess degrees of the nodes in the \(m\)-th layer, that is, the number of nodes in the \((m+1)\)-th layer (see fig.2),

\[
\langle k_{m+1} \rangle = \left. \frac{dG^{(m)}(x)}{dx} \right|_{x=1} = \left. \frac{dG^{(m-1)}(x)}{dx} \right|_{x=1} G'_1(1) = \langle k_m \rangle G'_1(1), \quad m \geq 1. \tag{13}
\]

From eq.(15), we can also find that

\[
\frac{\langle k_{m+1} \rangle}{\langle k_m \rangle} = G'_1(1), \quad m \geq 1. \tag{16}
\]

This indicates that the growth rate of the node number of the \(m\)-th layer with \(m\) is a constant \(G'_1(1)\). For an infinite network, there must be an infinite number of layers, so the node number can not decrease with \(m\), i.e.,

\[
G'_1(1) = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \geq 1. \tag{17}
\]

This is the existence condition of the giant component in a tree-like network[6, 12].

For a network with finite size, the average number of layers \(\bar{l}\) must be a finite number. Next, we will show how to obtain \(\bar{l}\). Using eq.(16), the node number of the \(m\)-th layer can be written as

\[
\langle k_m \rangle = \langle k_1 \rangle \left( \frac{\langle k_2 \rangle}{\langle k_1 \rangle} \right)^{m-1} = \langle k \rangle \left( \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \right)^{m-1}, \quad m \geq 1. \tag{18}
\]

Summing over all the layers, we will obtain the network size

\[
N = 1 + \sum_{m=1}^{\bar{l}} \langle k_m \rangle. \tag{19}
\]

This yields

\[
\bar{l} = \frac{\ln \left[ (N - 1)(\langle k^2 \rangle - 2\langle k \rangle) + \langle k \rangle^2 \right] - 2\ln \langle k \rangle}{\ln(\langle k^2 \rangle - \langle k \rangle) - \ln \langle k \rangle}. \tag{20}
\]
For a real network, \( N \gg \langle k \rangle \) and \( \langle k^2 \rangle \gg \langle k \rangle \), thus \( \bar{l} \) can be approximately represented as

\[
\bar{l} = \frac{\ln N - \ln \langle k \rangle}{\ln \langle k^2 \rangle - \ln \langle k \rangle} + 1. \tag{22}
\]

This result can be seen as an approximation of the average distance of a real network[13, 14], which indicates that \( \bar{l} \sim \ln N \).

3.2. \( k \)-shell structure
As we pointed above, the hierarchical structure shown in fig.2 has an arbitrary center. In complex network, there is another hierarchical structure with a determined center, i.e., \( k \)-shell structure[15]. For a given network, the \( k \)-shell structure is determined.

To illustrate \( k \)-shell structure of networks, we need to take a moment here to give the definition of \( k \)-core of network. \( k \)-core is a subnetwork of a network, in which nodes are connected to one another by at least \( k \) paths. \( k \)-core can be obtained by removing repeatedly all the nodes of degree less than \( k \). The result, if it exists, is the \( k \)-core of the network (see fig.3). It is easy to know that the size of \((k+1)\)-core \( S_{k+1} \) is smaller than that of \( k \)-core \( S_k \). The nodes that belong to \( k \)-core and not belong to \((k+1)\)-core are called as \( k \)-shell of the network.

Next, we will give the size of \( k \)-core and \( k \)-shell of a network ensemble with given degree distribution. It must be pointed that \( k \)-core and \( k \)-shell can not exist in a finite tree-like network, so all the networks discussed here are infinite. For a node in \( k \)-core, it must have at least \( k \) links that connect to other nodes in \( k \)-core. Therefore, using the generating functions (6) and (10), we can get the size of \( k \)-core (or the fraction of nodes in \( k \)-core) directly,

\[
S_k = \sum_{n=k} \frac{(-R_k)^n}{n!} \frac{d^n G_0(1 - R_k)}{dR_k^n} \tag{23}
\]

\[
R_k = \sum_{n=k-1} \frac{(-R_k)^n}{n!} \frac{d^n G_1(1 - R_k)}{dR_k^n} \tag{24}
\]

![Figure 3. k-core and k-shell structure of networks. In k-core, nodes are connected to one another by at least k paths. The nodes that belong to k-core and not belong to (k+1)-core are called as k-shell of the network.](image-url)
Here, $R_k$ means the probability that the node reached by following a randomly chosen link belongs to $k$-core. Note that the sum $\sum_n$ in eq.(25) begins with $k-1$, since the link used to reach the node has already provided a link for $k$-core. Solving the two equations, we can get the size of $k$-core $S_k$. Thus, the size of $k$-shell can be obtained,

$$L_k = S_k - S_{k+1}. \tag{27}$$

Figure 4. The size of $k$-shell of ER networks for different average degrees. In the simulation, the size of network is set as 10000. The solid lines are the corresponding theoretical predictions of eqs.(28) and (29).

For Erdős-Rényi (ER) network[12], the degree distribution is a Poisson distribution $p_k = e^{-\langle k \rangle} \langle k \rangle^k / k!$, thus eqs.(24) and (26) can be rewritten as

$$S_k = 1 - \Gamma(k, \langle k \rangle R_k), \tag{28}$$

$$R_k = 1 - \Gamma(k - 1, \langle k \rangle R_k). \tag{29}$$

Here, $\Gamma(n, x)$ is the regularized incomplete gamma function. Figure 4 compares these theoretical results with the simulations. One can find the simulation results agree well with this theoretical analysis. When the average degree of the network is small, it is clear the size of $k$-shell is small. When $\langle k \rangle < 1$, there is no giant component in the network, so $k$-shell is obviously inexistent. With the increasing of the average degree of the network, $k$-shell with larger $k$ will grow up and replace that of small $k$. Therefore, with the increase of the average degree, $k$-shell increases first and then decreases (see fig.4).

4. Robustness of networks

The robustness of networks is one of the important topics in complex networks[4, 16], which refers to the ability of a network to resist change without adapting its initial stable configuration. The typical robustness evaluation method is to consider the largest connected component in a network after a fraction $1 - p$ of nodes are removed, the so-called giant component. The birth point $p_c$ of a non-zero size of giant component $S$ is often used to evaluate the robustness of the network. A small birth point $p_c$ means the network is robust and resistant to the removal of a large number of nodes.

When we consider the robustness of networks, there are two link states in the network, i.e., connecting to the giant component or not. Let $R$ be the probability that a link connects to the giant component. Thus, substituting $x = 1 - R$ into eqs.(3) and (9), we can obtain the probability that a randomly chosen node is not in the giant component $G_0(1 - R)$, and the probability that a randomly chosen link does not connect to the giant component $G_1(1 - R)$. 

Here, $R_k$ means the probability that the node reached by following a randomly chosen link belongs to $k$-core. Note that the sum $\sum_n$ in eq.(25) begins with $k-1$, since the link used to reach the node has already provided a link for $k$-core. Solving the two equations, we can get the size of $k$-core $S_k$. Thus, the size of $k$-shell can be obtained,
Next, we consider the cascading failures model proposed in ref. [17], which focuses on the robustness of networks with overlapping of connectivity and dependence links. Here, the connectivity link just is the ordinary link of network, and the dependence link represents the dependence of two nodes in a network. Specifically, a dependence link between node \( i \) and node \( j \) means that if node \( i \) fails, node \( j \) also fails, and vice versa. We consider a network connected by connectivity links with a degree distribution \( p_k \) and each node has exactly one dependence link. For a network with size \( N \), there are \( N/2 \) dependence links. To represent the relation of the connectivity links and the dependence links, we assume that a fraction \( \beta \) of the dependence links overlaps with the connectivity links, and the other dependence links are set randomly. That means a fraction \( \beta \) of nodes will be adjacent to their dependence partners connected by the dependence links.

The cascading failure process begins with removing a fraction \( 1 - p \) of nodes randomly. We assume that all the nodes not belong to the giant component will fail. As the model setting, the node with a failed dependence partner will also fail. So, two dependence nodes can be preserved after the cascading failures, only when they belong to the giant component simultaneously.

To obtain the size of the giant component \( S \), we need the generating function to express the state of a dependence pair, instead of that of a node. For nonadjacent dependence nodes, the dependence partner is a randomly chosen one, whose state can be described by generating function \( G_0(x) \). So, the states of nonadjacent dependence pairs can be described by the joint state of two nodes. As shown in fig.5, the states for the adjacent dependence pairs can be described by generating functions \( G_1(x) \) and \( G_2(x)G_1(x) \), where \( G_2(x) = G_0''(x)/G_0'(1) \).

For an arbitrary degree distribution \( p_k \) and the fraction of initial removal \( 1 - p \), we can solve eq.(31) to obtain \( R \), and then insert it into eq.(30) to obtain the order parameter \( S \). In fig.6, we give the simulation results for ER networks, which agree with the analytical results well. In addition, we can find that there are two types of phase transition in this model, which can also be predicted by eq.(31) (see ref. [17] for details). For ER networks, we can also find that the tricritical point is a constant \( \beta_c = \frac{1}{3} \).

\[ S = (1-\beta)p^2 \left[ 1 - G_0(1-R) \right]^2 + \beta p^2 \left[ 1 - [G_1(1-R)]^2 \right], \quad (30) \]
\[ R = (1-\beta)p^2 \left[ 1 - G_1(1-R) \right] \left[ 1 - G_0(1-R) \right] + \beta p^2 \left[ 1 - G_2(1-R)G_1(1-R) \right]. \quad (31) \]
5. Conclusion
In this paper, we have summarized the generating function technique used in the studying of complex networks, and explored the hierarchical and \( k \)-shell structures of networks using this technique. A cascading dynamic model has also been studied in this paper as an example that how to use this technique to solve the network dynamic model. In addition to the topics covered in this paper, there are a number of problems can be solved by the generating function technique, such as spread of epidemic disease on networks\[7\], observability transition\[18\], catastrophic cascade of failures in interdependent networks\[19\], clique percolation\[20\]. In the other hand, as mentioned in this paper, this technique can only give an exact result for tree-like random networks, and further research is needed to clustering and correlated networks.

Acknowledgments
This work is funded by the National Natural Science Foundation of China (Grant Nos.: 11275186 and 91024026) and the Open Funding Programme of Joint Laboratory of Flight Vehicle Ocean-based Measurement and Control (Grant No.: FOM2014OF001).

References
[1] Newman M E J 2010 Networks: An introduction (New York: Oxford University Press)
[2] Albert R and Barabási A L 2002 Statistical mechanics of complex networks Rev. Mod. Phys. 74 47
[3] Dorogovtsev S N, Goltsev A V and Mendes J F F 2008 Critical phenomena in complex networks Rev. Mod. Phys. 80 1275
[4] Cohen R and Havlin S 2010 Complex networks: Structure, robustness and function (Cambridge: Cambridge University Press)
[5] Wilf H S 1994 Generatingfunctionology (London: Academic Press)
[6] Newman M E J, Strogatz S H and Watts D J 2001 Random graphs with arbitrary degree distributions and their applications Phys. Rev. E 64 026118
[7] Newman M E J 2002 Spread of epidemic disease on networks Phys. Rev. E 66 016128
[8] Newman M E J 2003 Ego-centered networks and the ripple effect Soc. Netw. 25 83
[9] Serrano M A and Boguñá M 2006 Clustering in complex networks: II. Percolation properties Phys. Rev. E 74 056115
[10] Berchenko Y, Artzy-Randrup Y, Teicher M and Stone L 2009 Emergence and size of the giant component in clustered random graphs with a given degree distribution Phys. Rev. Lett. 102 138701
[11] Shao J, Buldyrev S V, Braunstein L A, Havlin S and Stanley H E 2009 Structure of shells in complex networks. Phys. Rev. E 2009 80 036105
[12] Bollobás B 2001 Random graphs (Cambridge: Cambridge University Press)
[13] Newman M E J 2001 Scientific collaboration networks: I. Network construction and fundamental results
Phys. Rev. E 64 016131
[14] Newman M E J 2001 Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality
Phys. Rev. E 64 016132
[15] Dorogovtsev S N, Goltsev A V and Mendes J F F 2006 k-core organization of complex networks Phys. Rev. Lett. 96 040601
[16] Albert R, Jeong H and Barabási A L 2000 Error and attack tolerance of complex networks Nature 406 378
[17] Li M, Liu R-R, Jia C-X and Wang B-H 2013 Critical effects of overlapping of connectivity and dependency
links on percolation of networks New J. Phys. 15 093013
[18] Yang Y, Wang J and Motter A E 2012 Network observability transitions Phys. Rev. Lett. 109 258701
[19] Buldyrev S V, Parshani R, Paul G, Stanley H E and Havlin S 2010 Catastrophic cascade of failures in
interdependent networks Nature 464 1025
[20] Palla G, Derényi I and Vicsek T 2007 The critical point of k-clique percolation in the Erdős-Rényi graph
J. Stat. Phys. 128 219