Random growth networks with exponential degree distribution

Cite as: Chaos **30**, 113120 (2020); https://doi.org/10.1063/5.0022840
Submitted: 27 July 2020 . Accepted: 19 October 2020 . Published Online: 09 November 2020

Fei Ma, Xudong Luo, Ping Wang, and Renbo Zhu

Chaos ARTICLE scitation.org/journal/cha

Random growth networks with exponential degree distribution

Cite as: Chaos **30**, 113120 (2020); doi: 10.1063/5.0022840
Submitted: 27 July 2020 · Accepted: 19 October 2020 ·
Published Online: 9 November 2020

Fei Ma,[1,a)] Xudong Luo,[2] Ping Wang,[3,4,5,b)] and Renbo Zhu[4]

AFFILIATIONS

[1]School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China
[2]College of Mathematics and Statistics, Northwest Normal University, Lanzhou 730070, China
[3]National Engineering Research Center for Software Engineering, Peking University, Beijing 100871, China
[4]School of Software and Microelectronics, Peking University, Beijing 102600, China
[5]Key Laboratory of High Confidence Software Technologies (PKU), Ministry of Education, Beijing 100871, China

[a)]**Electronic mail:** mafei123987@163.com
[b)]**Author to whom correspondence should be addressed:** pwang@pku.edu.cn

ABSTRACT

A great variety of complex networks can be well represented as random graphs with some constraints: for instance, a provided degree distribution, a smaller diameter, and a higher clustering coefficient. Among them, the degree distribution has attracted considerable attention from various science communities in the last few decades. In this paper, we focus mainly on a family of random graphs modeling complex networks that have an exponential degree distribution; i.e., $P(k) \sim \exp(\alpha k)$, where k is the degree of a vertex, $P(k)$ is a probability for choosing randomly a vertex with degree equal to k, and α is a constant. To do so, we first introduce two types of operations: type-A operation and type-B operation. By both the helpful operations, we propose an available algorithm $\mathscr{A}$ for a seminal model to construct exactly solvable random graphs, which are able to be extended to a graph space $\mathfrak{S}(p, p^c, t)$ with probability parameters p and p^c satisfying $p + p^c = 1$. Based on the graph space $\mathfrak{S}(p, p^c, t)$, we discuss several topological structure properties of interest on each member $N(p, p^c, t)$ itself and find model $N(p, p^c, t)$ to exhibit the small-world property and assortative mixing. In addition, we also report a fact that in some cases, two arbitrarily chosen members might have completely different other topological properties, such as the total number of spanning trees, although they share a degree distribution in common. Extensive experimental simulations are in strong agreement with our analytical results.

It is well known that random graphs have widely been studied in the realm of complex networks mainly because they can be selected to mimic various complex systems. In the past few years, some random graphs of great interest have been proposed. Most of them turn out to be scale-free. There is little attention paid to random graphs that follow an exponential degree distribution. The goal of this paper is to generate a family of random graphs obeying an exponential degree distribution via an algorithm $\mathscr{A}$. In addition, as will be shown later, all the proposed random graphs have an identical degree distribution, a phenomenon that is seldom considered in exponential random graphs. Next, we analytically study some structural parameters on the proposed random graphs and find that each member exhibits the small-world property and has assortative mixing. Finally, using the derived analytic solution to spanning tree number, we stress that to some extent, all members behave completely different topological structures from one another while they share a degree distribution in common.

I. INTRODUCTION

Random graph in its simplest form is a set of vertices connected together in pairs by edges. In fact, the relevant study on random graphs has a long history. It dates back to the influential work of Erdös and Rényi[1] in the 1950s and 1960s. Since then, the random graph theory has attracted more attention from various disciplines and hence gradually developed into one of the members of modern discrete mathematics.

In general, almost all studies correlated to random graphs have a common assumption, i.e., connecting a pair of vertices selected randomly from the whole vertex set with a given probability p independently. This leads to a fact that the resulting random graph

$\mathscr{G}(V, E)$ has a vertex degree distribution $P(k)$ obeying

$$P(k) = \binom{|V|}{k} p^k (1-p)^{|V|-k}, \tag{1}$$

which will become the Poisson distribution in the limit of large graph size. Initially, people thought that most of the complex systems are organized in a random manner; therefore, random graphs can be naturally adopted as a useful tool to help us understand some universal and intriguing phenomena behind such systems, particularly, epidemiology on systems. Until the end of the last century, with enormous numbers of data, researchers found that a significant body of complex networks follows power law or a right-skewed degree distribution in terms of their own outlines in a log–log plot, which is completely different from that of the random graph, and thus are called scale-free networks often. This suggests that these complex systems appear to be not entirely random.[2] After that, the next almost two decades have seen an increasing interest of studying complex networks with the scale-free feature, including the World Wide Web (WWW), the Internet, sexual contact networks, scientist cooperation networks, friendship networks, protein interaction networks, and metabolic networks.[2–7] In reality, another probability distribution, namely, an exponential degree distribution, prevails in many situations, such as survival times for decaying atomic nuclei or the Boltzmann distribution of energies in statistical mechanics.[8] The goal of this paper is to produce a family of random graphs whose degree distributions are neither Poisson nor power-law but exponential.

The reasons for choosing this distribution are twofold. First, it can be seen in a number of small-world network models including the well-known Watts–Strogatz (WS) model.[3] Second, based on such a degree distribution, one can generate many other network models of great interest, for instance, power-law networks according to the weight distribution. To see why this is so, we here provide an illustrative example as follows. Consider a network model with an exponential degree distribution, that is, the fraction of degree k vertices follows an exponential distribution $P(k) \sim \exp(ak)$ in which parameter a indicates whether this distribution is normalizable or not. Now, we are able to think, in some cases, that the weight value w_k of the degree k vertex is exponentially related to degree value k; i.e., $w_k \sim \exp(bk)$, where symbol b represents another parameter. Then, the probability distribution of weight w_k obeys the following expression:

$$P(w_k) = P(k)\frac{dk}{dw_k} \sim \frac{w_k^{-1+\frac{a}{b}}}{b}. \tag{2}$$

Obviously, this means a power law with exponent $\gamma = 1 - \frac{a}{b}$. Hence, it is of prominent interest to study random graphs having an exponential degree distribution in terms of both realistic and theoretical sense. In this paper, we focus mainly on the simplest and most widely studied random graphs whose vertex degree distributions are exponential in form.

This paper is organized into four sections: in Sec. I, in order to smoothly build up our graph space $\mathfrak{S}(p, p^c, t)$ whose each member behaves a provided degree distribution $P(k) \sim \exp(ak)$, we will make use of two types of operations, type-A operation and type-B operation, which have been introduced in our previous work in detail.[9,10] In some sense, type-B operation is a generalized form of that defined in Ref. 10. At the same time, many basic yet important definitions are also recalled here. Based on both operations and definitions, Sec. II aims to establish graph space $\mathfrak{S}(p, p^c, t)$ using an algorithm $\mathscr{A}$. After that, for our purpose, the topic of Sec. III is made up of Subsections III A–III F in which we study many topological structure properties in depth, such as the degree distribution, diameter, clustering coefficient, spanning tree number, and Pearson's correlation coefficient. Finally, we close this paper by making an elaborated conclusion and bringing some heuristic discussions associated with our graph space $\mathfrak{S}(p, p^c, t)$ for future work in Sec. IV.

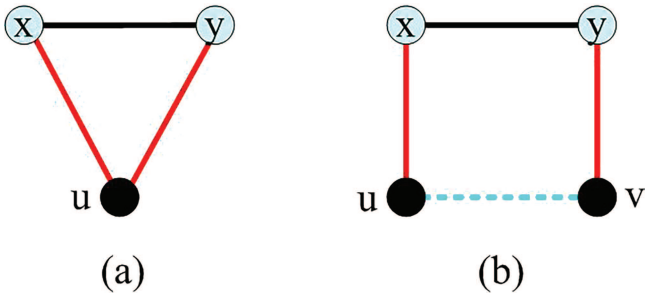

FIG. 1. The diagrams of type-A operation and type-B operation. As illustrated in panel (a), the type-A operation is also referred to as the inverted triangle operation as in Ref. 9. The right panel (b) shows the type-B operation, also named the generalized rectangle operation compared with that in Ref. 10.

A. Operations and definitions

In this section, we will introduce two types of operations used later, type-A operation and type-B operation, and some useful definitions in the language of graph theory.

1. Type-A operation

For a given vertex u, bringing an edge xy on vertices x and y and then connecting vertex pair x and y with vertex u (where newly created edges are colored in red) together produces a cycle C_3, shown in Fig. 1(a). Such a process is viewed as the type-A operation, also defined as the inverted triangle method in our previous work.[9]

2. Type-B operation

For a given active edge uv (which is colored in indigo) with two vertices u and v, bringing two vertices x and y linked by an edge xy, connecting vertex x with one end point u of edge uv by a new edge and similar implementation for vertex y and the other end point v together produces a cycle C_4. Such a process is type-B operation, also commonly defined as a rectangle method as in Ref. 10; see Fig. 1(b). In fact, it is natural to extend such an operation to be a generalization form, regarded as the generalized rectangle method in which that active edge uv may be replaced by an active path $\mathscr{P}$ of length l, namely, $\mathscr{P} : u(v_0)v_1 \cdots v_{l-1}(v_l)v$. It is so evident to the eye that the active path $\mathscr{P}$ will degenerate into an active edge when the path length is equal to 1. Hence, we in Fig. 1(b) denote an active path $\mathscr{P}$ by a dashed line on two distinct end vertices, say, u and v.

Definition 1 (Ref. 11). A spanning subgraph $\mathscr{G}(V', E')$ of a graph $\mathscr{G}(V, E)$ is a subgraph having the same vertex set as $\mathscr{G}(V, E)$ and a number of edges $|E'|$ such that $|E'| \leq |E|$. A spanning tree $\mathscr{T}(V', E')$ of a connected graph $\mathscr{G}(V, E)$ is a subgraph, which is a tree with $|E'| = |V| - 1$.

Definition 2 (Ref. 11). A given graph $\mathscr{G}(V, E)$ is considered simple if it has no loops and no two edges joining the same pair of vertices. Throughout this paper, we just discuss random graphs of this kind.

II. ALGORITHM $\mathscr{A}$

In view of both operations mentioned above, let us first put our insights into creating graph model $N(p, p^c, t)$ that is a member of graph space $\mathfrak{S}(p, p^c, t)$. Initially, one should choose a seminal seed, i.e., an arbitrary graph $\mathscr{G}(V, E)$ having cycle decomposition, which might be connected or not. For an unconnected seminal seed, one has to analyze each component belonging to the resulting random graph $N(p, p^c, t)$ built via both type-A and type-B operations after t time steps. Without loss of generality, for convenience and brevity, we here take into account this case where seminal model $\mathscr{G}(V, E)$ is connected and contains a cycle $C_{|V|}$ to which all vertices in $\mathscr{G}(V, E)$ belong. In practice, a similar analysis is suitable for the unconnected case.

Now, algorithm $\mathscr{A}$ will be proposed to generate our graph space $\mathfrak{S}(p, p^c, t)$. Before processing the task, we need to introduce a few terminologies to clarify the running process of algorithm $\mathscr{A}$. Given a connected graph $\mathscr{G}(V, E)$ mentioned above and a predefined probability parameter p $(0 < p < 1)$, one will encounter the following four cases plotted in Fig. 2 after implementing type-A operation exactly once. For brevity, we only make an elaborated description of *1-1-configuration* here, and the remaining three can be easily understood with the help of both similar explanation to *1-1-configuration* and illustration in Fig. 2. For a path $\mathscr{P}$ of length 2, i.e., $\mathscr{P} : uwv$, vertex w will be considered completely saturated after applying type-A operation to vertex w; that is to say, its degree is changed by plus 2. The other vertices, however, are referred to as incompletely saturated ones due to keeping their own degrees unchanged. On the other hand, to make them completely saturated ultimately, one need to take useful advantage of type-B operation for vertices u and v exactly once, which is highlighted using a blue dashed cycle in Fig. 2(a). (In essence, such an implementation is not taken immediately after using type-A operation but performed in the final step of algorithm $\mathscr{A}$ shown later. For details, see algorithm $\mathscr{A}$ and its corresponding example that we will describe shortly.) As said above, we omit the detailed clarifications for other configurations here and then pay our attention to algorithm $\mathscr{A}$ (i.e., Algorithm 1) as follows. It is worth mentioning that the first step of $\mathscr{A}$ is manipulated based on a cycle contained in graph $\mathscr{G}(V, E)$.

To make algorithm $\mathscr{A}$ more concrete and understandable, we have to introduce an example shown in Fig. 3, which can serve as a guideline to the eye. Consider a given seminal graph $\mathscr{G}$ plotted in Fig. 3(a) and suppose that probability parameter $p = 1/3$ in our algorithm $\mathscr{A}$, we feed model $\mathscr{G}$ into algorithm $\mathscr{A}$ and obtain at random a model in graph space $\mathfrak{S}(1/3, 2/3, t)$ at each specified time step t. Here, we only take a member in graph space $\mathfrak{S}(1/3, 2/3, 1)$ as a representative as shown in Fig. 3(b) to make a description about how algorithm $\mathscr{A}$ runs at time step $t = 1$. First, algorithm $\mathscr{A}$ will choose uniformly at random two distinct vertices according to probability $p = 1/3$ after gaining seminal graph $\mathscr{G}$ and then type-A operation is implemented on both the vertices, which are labeled with label 1 (5–10). Then, algorithm $\mathscr{A}$ goes into *phase 1* (11–14) and finds that there is a *2-1-configuration*. This suggests that the vertex with label 2 needs to perform type-A operation exactly once. After that, some portion of completely unsaturated vertices, i.e., vertices marked by number 3, will be handled using type-B operation (15–17). After accomplishing the three procedures mentioned above, algorithm $\mathscr{A}$ goes into *phase 2* (18–20) and discovers that there exists a *1-2-configuration* and thus again applies type-B operation to two incompletely saturated vertices with number 4. By far, algorithm $\mathscr{A}$ outputs a representative as in Fig. 3(b), will automatically terminate at present, and then returns to the next time step.

Our discussion and algorithm $\mathscr{A}$ up to this point are general and applicable to an admitted connected graph as stated above, which is used as a seminal one. We in this paper preferably select a cycle C_4 as a seminal graph so as to better describe implementation of algorithm $\mathscr{A}$. After t time steps, we can derive a graph space $\mathfrak{S}(p, p^c, t)$, which contains various members (models) $N(p, p^c, t)$. Then, our goal is to uncover difference and similarity among these members (models) $N(p, p^c, t)$ both analytically and experimentally. Figure 4 shows an illustration when setting probability parameter $p = 0.5$.

III. TOPOLOGICAL PROPERTIES

Along the line of recent research,[12] we in this section devote to studying some widely discussed topological structure characteristics, for instance, average degree, degree distribution, diameter, and clustering coefficient. By the discussions about these characteristics shown shortly, it will be clear to see that while all the members $N(p, p^c, t)$ in our graph space $\mathfrak{S}(p, p^c, t)$ may exhibit some similar structural parameters to one another, some distinguishable features among them are also found clearly.

A. Average degree

As an index indicating whether a given network model is sparse or not, the average degree has been popularly adopted in the complex network study. Here, it is clear due to algorithm $\mathscr{A}$ that for an arbitrary probability parameter p, each member $N(p, p^c, t)$ of graph space $\mathfrak{S}(p, p^c, t)$ has a pair of identical equations, i.e., vertex number (order) $|\mathfrak{V}(p, p^c, t)| = 4 \times 3^t$ and edge number (size) $|\mathfrak{E}(p, p^c, t)| = 2 \times 3^{t+1} - 2$; hence, they all share an average degree $\langle k(p, p^c, t)\rangle = 2|\mathfrak{E}(p, p^c, t)|/|\mathfrak{V}(p, p^c, t)| \cong 3$ in common. It is so obvious to the eye that each member $N(p, p^c, t)$ exhibits sparsity, a feature that is commonly seen in a large number of real-life and artificial complex networks.[2]

B. Diameter

Almost all complex networks that have been previously investigated have a smaller diameter. The diameter is defined as the maximum of all distances between any two vertices in a connected graph. Notice that the distance between vertex pair u and v is the

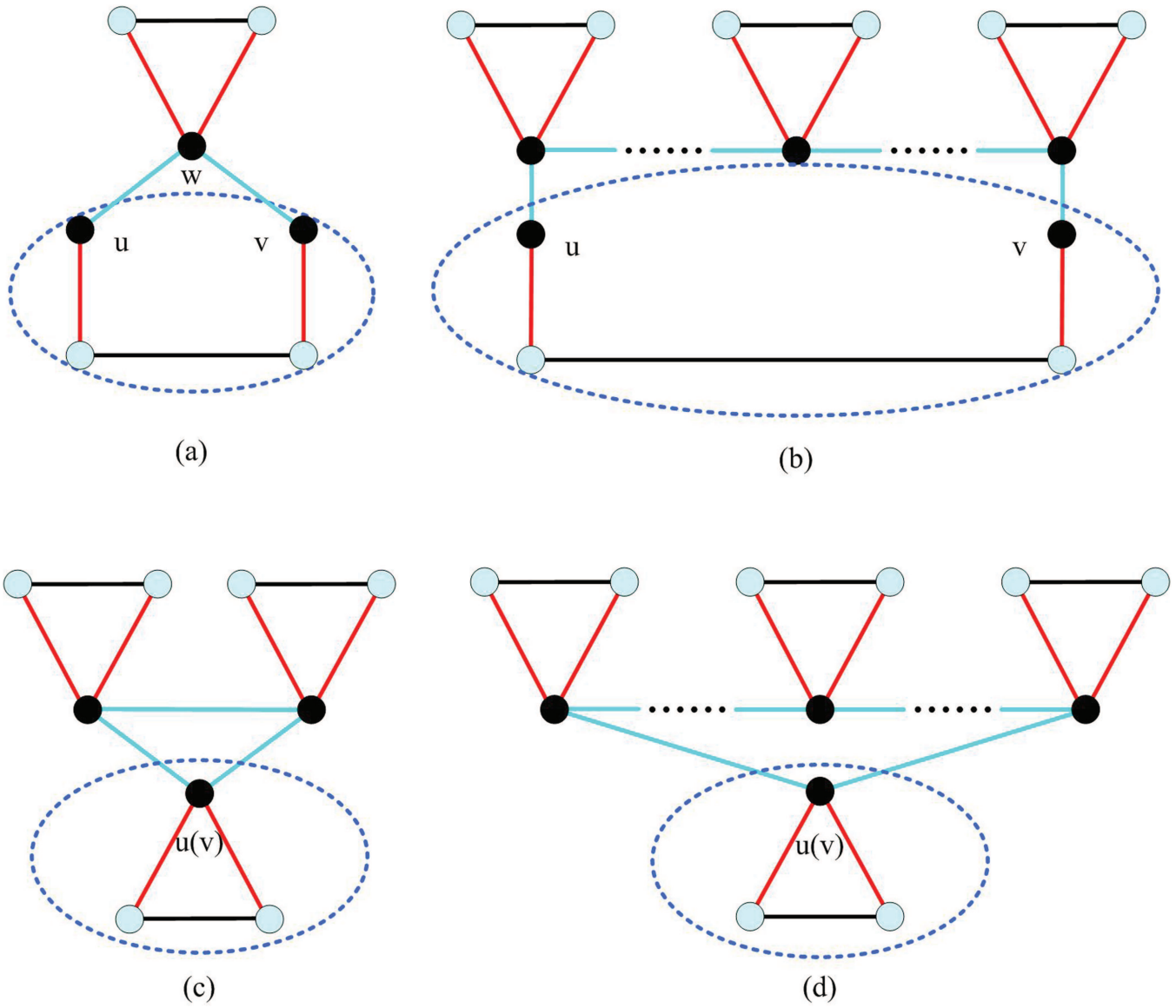

FIG. 2. The diagrams of four different types of configurations after applying type-A operation at each time step in which vertices u and v are incompletely saturated. For our purpose, we think of panels (a), (b), (c), and (d) as *1-1-configuration*, *1-2-configuration*, *2-1-configuration*, and *2-2-configuration*, respectively. Broadly speaking, the *1-2-configuration* is generalization to a saturated path of the *1-1-configuration* in which all the internal vertices are saturated except for two incompletely saturated end points. Similarly, demonstration is suitable for both *2-2-configuration* and *2-1-configuration*.

minimum of lengths of all paths connecting them in the language of graph theory. More generally, diameter can be utilized as an index measuring information delay on the complex network as a whole. The smaller the diameter is, the more quickly the information is disseminated. In the past several decades, one has proven both analytically and empirically that a great number of complex networks appear to display the small-world property, i.e., a smaller diameter and a higher clustering coefficient[3] (discussed in Subsection III C). Motivated by this, we will make a detailed discussion about diameter $\mathfrak{D}(p, p^c, t)$ of network models $N(p, p^c, t)$ when varying probability parameter p. Due to randomness brought by probability parameter p, let us first study two extreme examples, namely, probability parameter p equal to either 0 or 1, which are deterministic and easily coped with.

1. Case 1

Suppose parameter p equal to 1, our graph space $\mathfrak{S}(p, p^c, t)$ contains a unique member $N(1, 0, t)$ generated solely using type-A

Algorithm 1. Algorithm $\mathscr{A}$ is to construct a family of random graphs with an identical exponential degree distribution based on a seminal model $\mathscr{G}(V, E)$

Require: initial circle $C_{|V|}$; probability p; end time T

Ensure G_T
1: $G(V, E) \leftarrow C_{|V|}$
2: **for** $t = \{1, 2..., T\}$ **do**
3: $E_{t-1} \leftarrow$ all edges added at time step $t-1$
4: $\Psi \leftarrow \emptyset$ âŰů Ψ is used to save all vertices applied by type-A operation
5: **for each** $u \in V$ **do**
6: **if** $rand(0,1) < p$ **then**
7: update G by applying type-A operation to vertex u
8: add u to Ψ
9: **end if**
10: **end for**
11: **for each** $V_{u\mathscr{P}_u} \subseteq V$ s.t. $u \notin \Psi$ and $V_{\mathscr{P}} \subseteq \Psi$ **do**
12: update G by applying type-A operation to vertex u âŰů 1: $u\mathscr{P}u$ denotes a circle, $V_{u\mathscr{P}_u}$ is the vertex set of cycle $u\mathscr{P}u$
13: add u to Ψ
14: **end for**
15: **for each** $uv \in E_{t-1}$ s.t. $u, v \notin \Psi$ **do**
16: update G by applying type-B operation to edge uv
17: **end for**
18: **for each** $V_{u\mathscr{P}_v} \subseteq V$ s.t. $u, v \notin \Psi$ and $V_{\mathscr{P}} \subseteq \Psi$ **do**
19: update G by applying type-B operation on two end-vertices u and v of path $u\mathscr{P}v$
20: **end for**
21: **end for**
22: **return** G

operation. For such a model, we can derive an exact solution to its diameter $\mathfrak{D}(1, 0, t)$ by induction on time step t. In the initial condition, cycle C_4 has diameter $\mathfrak{D}(1, 0, 0)$ equivalent to 2. Next, model $N(1, 0, 1)$ is built by applying type-A operation to each vertex of cycle C_4; therefore, diameter $\mathfrak{D}(1, 0, 1)$ will be added by two compared to $\mathfrak{D}(1, 0, 0)$, namely, $\mathfrak{D}(1, 0, 1) = 2 + 2 = 4$, which can be easily checked by hand. Now, let $\mathfrak{D}(1, 0, j) = 2(j + 1)$ hold for all $j(< t)$. Taking into account both the functions of type-A operation, namely, connecting two end points of a new edge with one vertex of model $N(1, 0, j)$, and self-similarity of model $N(1, 0, j)$, one can capture a closed-form solution to diameter $\mathfrak{D}(1, 0, j + 1)$. In particular, an arbitrary path of length $\mathfrak{D}(1, 0, j)$ in model $N(1, 0, j)$ will become a distinct path of length $\mathfrak{D}(1, 0, j + 1)$ of model $N(1, 0, j + 1)$ with respect to type-A operation. Therefore, diameter $\mathfrak{D}(1, 0, j + 1)$ is equal to $1 + \mathfrak{D}(1, 0, j + 1) + 1 = 2((j + 1) + 1)$ and induction is complete. Diameter $\mathfrak{D}(1, 0, t) = 2(t + 1)$ holds for $t \geq 0$.

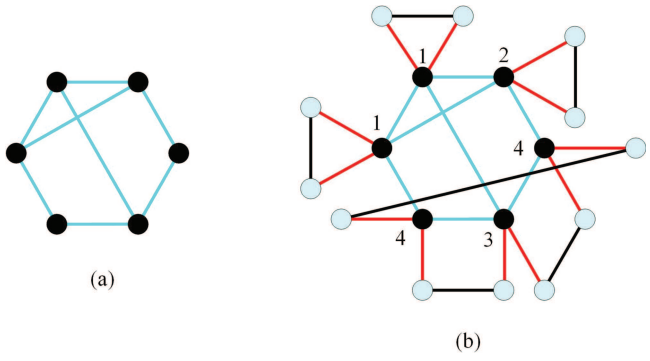

FIG. 3. The diagram of application of both type-A and type-B operations to a given seminal graph $\mathscr{G}(V, E)$ in panel (a). The resulting graph is shown in panel (b). In fact, this is only a description of our algorithm $\mathscr{A}$ implemented on a given seminal graph $\mathscr{G}(V, E)$ at the first time step.

2. Case 2

In the other extremal situation, $p = 0$, there is only a deterministic member $N(0, 1, t)$ in graph space $\mathfrak{S}(p, p^c, t)$. We directly give the solution to diameter $\mathfrak{D}(0, 1, t)$, i.e., $\mathfrak{D}(0, 1, t) = 2(t + 1)$, and omit computation with respect to similar explanation as above.

After all, deterministic models are easy to manage. Below, let us put insight into investigation about diameter $\mathfrak{D}(p, p^c, t)$ subject to $0 < p < 1$. Consider the procedure described by algorithm $\mathscr{A}$ carefully, triangles will be added into model $N(p, p^c, t)$ until *phase 1* accomplishes and then algorithm $\mathscr{A}$ implements type-B operation for the rest of vertices. In fact, we can obtain a great number of cycles with various lengths induced by new edges generated at time step t (commonly called an edge-induced subgraph in the jargon of graph theory). These such cycles are disjoint with each other because both type-A operation and type-B operation are performed only

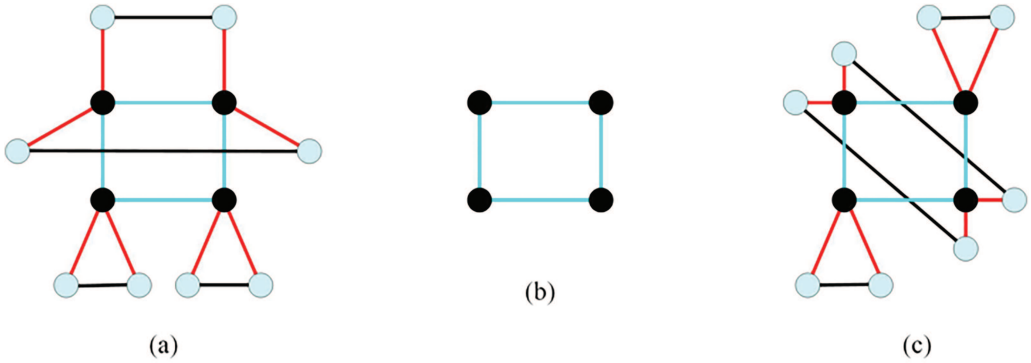

FIG. 4. The diagrams of model $N(p, p^c, t)$ with probability parameter $p = 0.5$ at time step $t = 1$. Here, the seminal model $N(0.5, 0.5, 0)$ is a cycle C_4 allocated at the middle position, namely, panel (b). The leftmost panel (a) and the rightmost panel (c) are two representatives $N(0.5, 0.5, 1)$, respectively, generated by applying both type-A operation to seminal model $N(0.5, 0.5, 0)$ with probability $p = 0.5$ and type-B operation with its corresponding complementary probability $p^c = 1 - p = 0.5$.

on vertices. For the first several time steps, it is not hard to calculate the precise expression of $\mathfrak{D}(p, p^c, t)$ by enumerating all possible cases, and we can obtain $\mathfrak{D}(p, p^c, t) = 2(t + 1)$. Hence, we make a plausible assertion: $\mathfrak{D}(p, p^c, t) = 2(t + 1)$ is true for all time steps t. Enumeration, however, will fail to work as time goes by. Different from the descriptions mentioned above, we do not adopt induction on time step t to compute diameter $N(p, p^c, t)$ but give a reasonable proof by means of computer simulations shown in Fig. 5. It is no doubt that there is a perfect agreement between our assertion to diameter $\mathfrak{D}(p, p^c, t)$ and the experimental solution shown by computer simulations. This implies that both type-A operation and type-B operation have no considerable effect on diameters of all the members in graph space $\mathfrak{S}(p, p^c, t)$.

By far, for all parameters $0 \le p \le 1$, each member $N(p, p^c, t)$ of graph space $\mathfrak{S}(p, p^c, t)$ has a smaller diameter compared with the corresponding order, $\mathfrak{D}(p, p^c, t) = 2(t + 1) \approx O(\ln |\mathfrak{V}(p, p^c, t)|)$ shown in Fig. 5 and captures one of both features of small-world network models.[3] In order to further turn out member $N(p, p^c, t)$ to be indeed small world, Subsection III C is going to study the clustering coefficient of member $N(p, p^c, t)$.

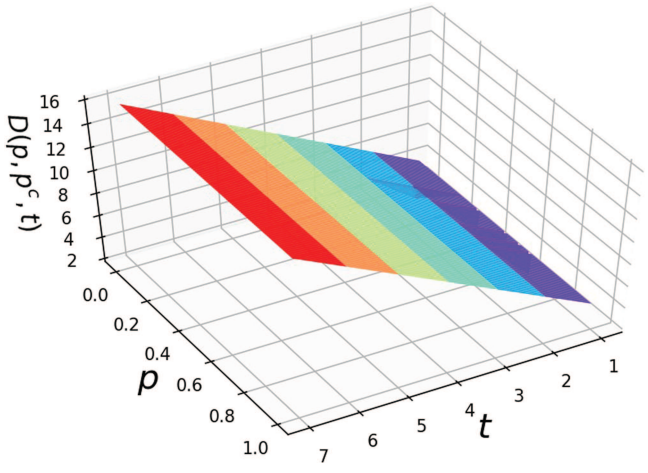

FIG. 5. The diagrams of diameter $\mathfrak{D}(p, p^c, t)$ of network models $N(p, p^c, t)$ at time steps $1 \le t \le 10$.

C. Clustering coefficient

First, there, in fact, is a significant difference between type-A operation and type-B operation in the topology terms. One of the most important reasons for this is as below. Type-A operation is to add triangle (cycle C_3) into graph and type-B operation, however, brings other cycles with various sizes, such as cycle C_4. For a degree k vertex, its clustering coefficient c can be defined as $c = \frac{n_k}{k(k-1)/2}$, where n_k is the total number of edges attached to vertex itself and the term in the denominator represents the number of all possible edges in its first-order neighbor set. By the definition of the clustering coefficient above, each vertex of cycle C_3 has the highest clustering coefficient equal to unity. On the other hand, all vertices of the latter, for instance, cycle C_4, have the lowest clustering coefficient being zero. As explained here, type-A operation and type-B operation will take a significant influence on the clustering coefficient of the growth process of member $N(p, p^c, t)$, respectively. Apparently, type-A operation plays a positive role. On the contrary, type-B operation decreases the clustering coefficient at a large degree.

As known, the triadic closure process is of ubiquitous interest in many real-life complex networks, for instance, individuals trending to form ties to the friends of their friends in a friendship network. In practice, not all individuals would like to do this. To portray such real-world phenomena well, we denote the former by type-A operation and the latter by type-B operation. In some sense, our graph space $\mathfrak{S}(p, p^c, t)$ can serve as a candidate for helping us to better understand dynamic functions and rich properties on many complex networks. From now on, we will analyze the clustering coefficient of each vertex of member $N(p, p^c, t)$, then distinguish the contribution from each vertex to the clustering coefficient $\langle \mathfrak{C}(p, p^c, t) \rangle$ over the whole model $N(p, p^c, t)$, and finally

derive a closed-form solution to $\langle \mathfrak{C}(p,p^c,t)\rangle$. As before, let us first proceed by two deterministic members $N(1,0,t)$ and $N(0,1,t)$.

1. Case 1

Since vertex v added at time step t_i into model $N(1,0,t)$ will have degree $2(t-t_i+1)$ and $t-t_i+1$ cycle C_3, its clustering coefficient $C(t_i,t)$ is equal to $1/(2t-2t_i+1)$. Therefore, the solution to the clustering coefficient $\langle \mathfrak{C}(1,0,t)\rangle$ can be precisely written as

$$\langle \mathfrak{C}(1,0,t)\rangle = \frac{\sum_{v\in N(1,0,t)} C(t_i,t)}{|\mathfrak{V}(p,p^c,t)|} = \sum_{t_i=0}^{t} P(k(t_i,t))C(t_i,t) \approx \sum_{t_i=0}^{t} \frac{(\sqrt{3})^{-k(t_i,t)}}{(2t-2t_i+1)}, \tag{3}$$

where $P(k(t_i,t))$ means a probability for choosing a degree $k(t_i,t)$ vertex, which will be discussed in Subsection III D in detail.

2. Case 2

Due to type-B operation no cycle C_3 is created. Without doubt, $\langle \mathfrak{C}(0,1,t)\rangle$ will approach to zero in the limit of large graph size.

3. Case 3

For parameter p varying in the interval $(0,1)$, an arbitrary vertex is performed using the type-A operation with probability p independently when one obtains model $N(p,p^c,t)$ from the preceding model $N(p,p^c,t-1)$ $(t\geq 1)$. With the same calculation as that of *Case 1*, the closed-form solution to the clustering coefficient $\langle \mathfrak{C}(p,p^c,t)\rangle$ can be expressed as

$$\begin{aligned}\langle \mathfrak{C}(p,p^c,t)\rangle &= \frac{1}{|\mathfrak{V}(p,p^c,t)|}\frac{4pt}{(t+1)(2t+1)} + \frac{1}{|\mathfrak{V}(p,p^c,t)|}\left(\sum_{t_i=1}^{t} p|\Delta\mathfrak{V}(t_i,t)|\frac{1+p(t-i)}{(t+1-t_i)(2(t-t_i)+1)}\right)\\ &\quad + \frac{1}{|\mathfrak{V}(p,p^c,t)|}\left(\sum_{t_i=1}^{t} p^c|\Delta\mathfrak{V}(t_i,t)|\frac{p(t-t_i)}{(t+1-t_i)(2(t-t_i)+1)}\right)\\ &= \frac{1}{4\times 3^t}\left(\frac{4pt}{(t+1)(2t+1)} + \sum_{t_i=1}^{t}\frac{8p\times 3^{t_i-1}}{2(t-t_i)+1}\right),\end{aligned} \tag{4}$$

in which symbol $|\Delta\mathfrak{V}(t_i,t)|$ denotes the total number of vertices generated at time step t_i.

With the help of both *Cases 1–3* and Fig. 6, we can conclude that for reasonable parameter p, each model $N(p,p^c,t)$ in graph space $\mathfrak{S}(p,p^c,t)$ has a higher clustering coefficient in the limit of large system size. At the same time, there must exist an inequality as follows:

$$0 = \mathfrak{C}(0,1,t) \leq \mathfrak{C}(p,p^c,t) \leq \mathfrak{C}(1,0,t),$$

meaning that there is a phase transition point p_α above which $\mathfrak{C}(p,p^c,t)$ has a tendency to a nonzero constant in the limit of large graph size and below which $\mathfrak{C}(p,p^c,t)$ will converge to zero.

In a nutshell, almost all models $N(p,p^c,t)$ in graph space $\mathfrak{S}(p,p^c,t)$ show small-world property according to a small diameter in Subsection III B and a high clustering coefficient addressed in this subsection.

D. Degree distribution

From the topological structure point of view, the degree distribution may encapsulate many simplest and most important characters of a graph under consideration. As mentioned in Ref. 2, the World Wide Web (WWW) and a great variety of other complex networks have the scale-free feature with respect to their own power-law or right-skewed degree distributions. After that, more and more followers try their best to turn out whether various types of complex networks have the scale-free feature using degree distribution. As previously mentioned in Sec. I, relatively little attention has been paid to complex networks with exponential distribution in the past few decades. In essence, the exponential distribution is very popular in our living world. Indeed, as we will show below, all the models in our graph space $\mathfrak{S}(p,p^c,t)$ obey exponential degree distributions in form.

Taking into consideration the growth process of model $N(p,p^c,t)$, we should employ the accumulative method in a discrete case to capture the degree distribution,[14] as follows:

$$P_{cum}(k\geq k(t_i,t)) = \frac{\sum_{k\geq k(t_i,t)}|N(k)|}{|\mathfrak{V}(p,p^c,t)|}, \tag{5}$$

where $|N(k)|$ is the number of vertices with degree k.

It is worth noticing that at time step t, vertices with the degree greater than or equal to $k(t_i,t)$ are that ones generated at time step t_i or early. Hence, $\sum_{k\geq k(t_i,t)} N(k)$ in Eq. (5) is equivalent to $|\mathfrak{V}(p,p^c,t_i)|$; therefore, Eq. (5) can be simplified to

$$P_{cum}(k\geq k(t_i,t)) \approx 3^{t_i-t}. \tag{6}$$

Substituting $k(t_i,t) = 2(t-t_i+1)$ into Eq. (6) yields

$$P_{cum}(k\geq k(t_i,t)) \approx (\sqrt{3})^{-k(t_i)} = \exp(-\ln\sqrt{3}\times k(t_i)), \tag{7}$$

suggesting that our model $N(p,p^c,t)$ has an exponential degree distribution with parameter $a = -\ln\sqrt{3}$.

Until now, we have produced a class of random graphs with a unique exponential degree distribution. Some widely focused topological properties are shared by all the random graphs, for

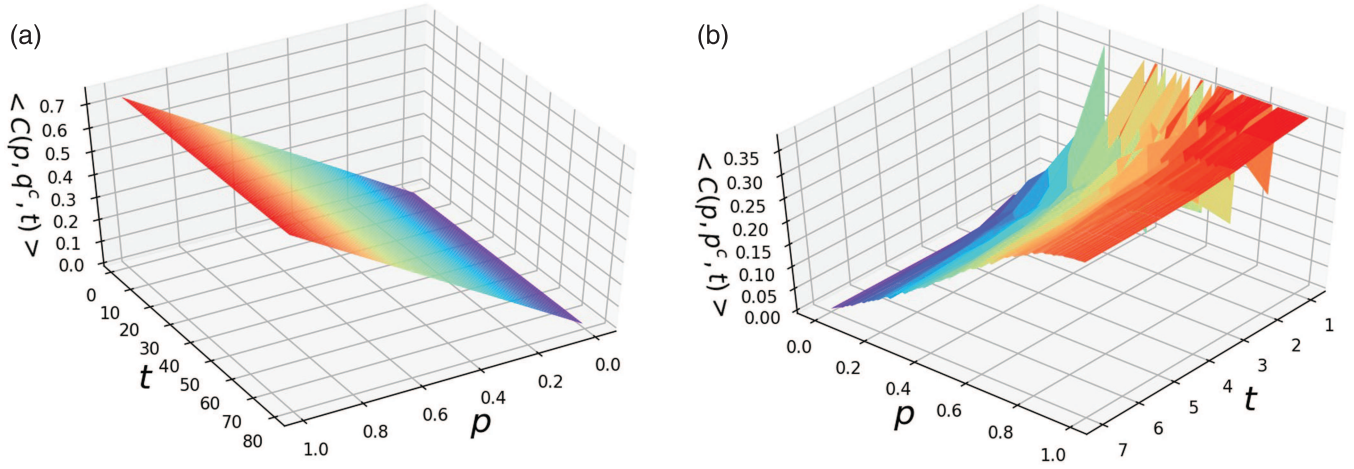

FIG. 6. The diagrams of two types of measures for computing clustering coefficients $\langle \mathfrak{C}(p, p^c, t) \rangle$ of models $N(p, p^c, t)$. The left panel (a) shows results derived by Eq. (4) and the right panel (b) illustrates computer simulation results based on the other definition of the clustering coefficient in Ref. 13 at time steps $1 < t < 10$. In the right panel, there is a stronger fluctuation for small parameters p, but such phenomena will vanish as t becomes large. It appears to display a considerable difference between both the panels and further reveals that our analytical result is in poor agreement with simulations. In reality, the method adopted in this paper and that in Ref. 13 are to unveil a fact, a rich "cluster" in complex networks, from two different viewpoints. Therefore, either the left or the right panel can be employed to prove that each model $N(p, p^c, t)$ in graph space $\mathfrak{S}(p, p^c, t)$ has a higher clustering coefficient in the limit of a large system size.

example, sparsity feature, smaller diameter, and higher clustering coefficient. In some sense, both operations appear to have no noticeable influence on topological properties except for the clustering coefficient.

E. Pearson's correlation coefficient

Recent studies have shown that for many real-world networks, the degrees of vertices at either end point of an edge chosen uniformly at random are not independent but correlated with one another. To put it another way, individuals tend to build friendships with those like them in a social network. Newman and Park[15] proposed a measure r based on the Pearson correlation coefficient to depict such a phenomenon. To be self-contained, we reorganize the appearance of Pearson's correlation coefficient r using notations in this paper as

$$r(p, p^c, t) = \frac{|\mathfrak{E}(p, p^c, t)|^{-1} \sum_{e_{ij} \in N(p, p^c, t)} k_i k_j - \left[|\mathfrak{E}(p, p^c, t)|^{-1} \sum_{e_{ij} \in N(p, p^c, t)} \frac{1}{2}(k_i + k_j) \right]^2}{|\mathfrak{E}(p, p^c, t)|^{-1} \sum_{e_{ij} \in N(p, p^c, t)} \frac{1}{2}(k_i^2 + k_j^2) - \left[|\mathfrak{E}(p, p^c, t)|^{-1} \sum_{e_{ij} \in N(p, p^c, t)} \frac{1}{2}(k_i + k_j) \right]^2}. \tag{8}$$

With the help of Pearson's correlation coefficient $r(p, p^c, t)$, a great deal of social networks turn out to possess positive value r and hence are assortatively constructed. On the other hand, non-social networks, such as technological and biological networks, have a disassortative mixing pattern.[15]

As described in algorithm $\mathscr{A}$, both type-A operation and type-B operation will connect a couple of degree two young vertices to one already existing vertex at each time step. The only difference between both the operations is that two young vertices are connected by a new edge in type-A operation, but they are not connected to each other in type-B operation. In other words, either type-A operation or type-B operation has no influence on Pearson's correlation coefficient $r(p, p^c, t)$ of model $N(p, p^c, t)$ as a whole. Based on these discussions mentioned above, we are likely to obtain an exact solution to Pearson's correlation coefficient $r(p, p^c, t)$ using an available classification over all the vertex pairs, i.e., grouping different edges incident to two vertices with distinct degrees into different families. For convenience and brevity, we take advantage of some helpful notations, for instance, $\mathfrak{E}_{ij}(p, p^c, t)$ denoting a set of edges e_{ij} linking degree i vertex with degree j vertex of model $N(p, p^c, t)$, shortened as $\mathfrak{E}_{i,j}(p, p^c, t) \stackrel{def}{=} \{e_{ij} | e_{ij} \in N(p, p^c, t)\}$.

Now, let us turn our attention to edge classifications of model $N(p,p^c,t)$. At first sight, there are $t+1$ kinds of vertices with degrees $2(t+1-t_i)$ $(0 \le t_i \le t)$ connected to each of the initial four vertices of cycle C_4; therefore, we can write

$$|\mathfrak{E}_{2(t+1),2(t+1-t_i)}(p,p^c,t)| = \begin{cases} 4, & t_i = 0, \\ 8, & 1 \le t_i \le t. \end{cases} \tag{9}$$

Similarly, for $1 \le t_i \le t$, a group of equations can be obtained as follows:

$$|\mathfrak{E}_{2(t+1-t_i),2(t+1-t_i-t_j)}(p,p^c,t)| = \begin{cases} |\Delta\mathfrak{V}(t_i,t)|/2, & t_j = 0, \\ 2|\Delta\mathfrak{V}(t_i,t)|, & 1 \le t_j \le t-t_i, \end{cases} \tag{10}$$

where $|\Delta\mathfrak{V}(t_i,t)|$ is the number of vertices added at time step t_i into model $N(p,p^c,t)$. Armed with Eqs. (9) and (10), we can gain

$$\begin{aligned} \sum_{e_{ij}\in N(p,p^c,t)} k_i(t)k_j(t) = {} & \sum_{i=0}^{t} |\mathfrak{E}_{2(t+1-i),2(t+1-i)}(t)| \times 2(t+1-i) \\ & \times 2(t+1-i) + \sum_{i=1}^{t+1}\sum_{j=1}^{i-1} |\mathfrak{E}_{2i,2(i-j)}(t)| \\ & \times 2i \times 2(i-j), \end{aligned} \tag{11}$$

$$\begin{aligned} \sum_{e_{ij}\in N(p,p^c,t)} k_i(t) + k_j(t) = {} & \sum_{i=0}^{t} |\mathfrak{E}_{2(t+1-i),2(t+1-i)}(t)| \\ & \times [2(t+1-i) + 2(t+1-i)] \\ & + \sum_{i=1}^{t+1}\sum_{j=1}^{i-1} |\mathfrak{E}_{2i,2(i-j)}(t)| \times [2i + 2(i-j)], \end{aligned} \tag{12}$$

and

$$\begin{aligned} \sum_{e_{ij}\in N(p,p^c,t)} k_i^2(t) + k_j^2(t) = {} & \sum_{i=0}^{t} |\mathfrak{E}_{2(t+1-i),2(t+1-i)}(t)| \\ & \times [4(t+1-i)^2 + 4(t+1-i)^2] \\ & + \sum_{i=1}^{t+1}\sum_{j=1}^{i-1} |\mathfrak{E}_{2i,2(i-j)}(t)| \times [4i^2 + 4(i-j)^2]. \end{aligned} \tag{13}$$

Substituting Eqs. (11)–(13) into Eq. (8) produces the analytical solution to Pearson's correlation coefficient $r(p,p^c,t)$. Here, the numerical behavior of parameter $r(p,p^c,t)$ is shown in Fig. 7.

As seen in Fig. 7, for a given parameter p, both type-A operation and type-B operation do not have a significant influence on Pearson's correlation coefficient $r(p,p^c,t)$, implying that analytical results are strongly consistent with computer simulations. Different from many scale-free models in Refs. 10 and 15, all the members of graph space $\mathfrak{S}(p,p^c,t)$ have assortative mixing because of their Pearson correlation coefficients $r(p,p^c,t)$ more than zero.

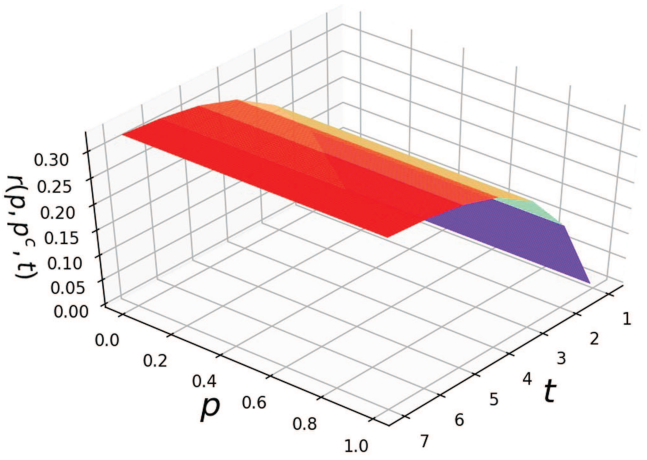

FIG. 7. The diagram of Pearson's correlation coefficient $r(p,p^c,t)$ of network model $N(p,p^c,t)$ at time steps $1 \le t \le 10$.

F. Spanning trees

Counting the number of spanning trees of any finite graph is of great interest and has a long history. In theory, this issue had been addressed by the well-known Kirchhoff's matrix-tree theorem (the total number of the spanning tree number is equivalent to the product of all nonzero eigenvalues of the Laplacian matrix of a graph in question). However, this does not diminish researchers' exploration for seeking some available and simple methods for many interesting graphs with specific structures, such as the Apollonian network[16] and Sierpinski networks.[17] In addition, the spanning tree number is in general treated as an important invariant correlated to several kinds of dynamical functions on models, for instance, reliability,[18] synchronization capability,[19] and random walks,[20–22] to name just a few. Taking into account randomness introduced by parameter p, there is no explicit solution to this counting issue on models $N(p,p^c,t)$. Hence, we in this subsection only enumerate spanning trees of two deterministic models $N(1,0,t)$ and $N(0,1,t)$ because of their own special topological structures and give an upper bound and a lower bound about the spanning tree number on graph space $\mathfrak{S}(p,p^c,t)$. First, let us turn our attention to the simplest model $N(1,0,t)$.

1. Case 1

As $p = 1$, type-A operation is only deployed in which degree $k_i = 2(t+1-t_i)$ $(0 \le t_i \le t)$ vertex occupies $t+1-t_i$ cycle C_3; therefore, the total number of spanning trees of model $N(1,0,t)$ is

$$S(1,0,t) = 4\prod_{t_i=0}^{t-1} 3^{|\Delta\mathfrak{V}(t_i,t)|}. \tag{14}$$

2. Case 2

As $p = 0$, no type-A operation is deployed at all in the growth process of model $N(0,1,t)$. To count the total number of spanning trees of model $N(0,1,t)$, we need to introduce two concepts,[23]

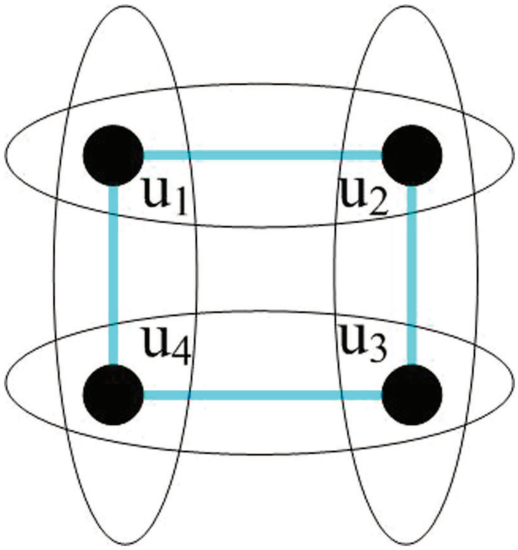

FIG. 8. The diagram of model $N(0, 1, t)$ in which the initial cycle C_4 on vertices $u_i (1 \le i \le 4)$ is lightened in color.

$\alpha(0, 1, t)$ representing the spanning tree number on edge u_1u_2 and $\beta(0, 1, t)$ being the number of *2-(u; v)-forest* on edge u_1u_2 where we already take useful advantage of the other illustration of model $N(0, 1, t)$ plotted in Fig. 8. The symmetric and self-similar structure of model $N(0, 1, t)$ suggests that it is sufficient to just focus on edge u_1u_2, and then, we can have

$$\begin{cases} \alpha(0,1,t) = 3\alpha^2(0,1,t-1)\beta(0,1,t-1) + \alpha^3(0,1,t-1), \\ \beta(0,1,t) = 3\alpha^2(0,1,t-1)\beta(0,1,t-1), \end{cases} \tag{15}$$

with the initial condition $\alpha(0, 1, 1) = 4$ and $\beta(0, 1, 1) = 3$.

From Eq. (15), it is easy to find out that there exists a bridge between $\alpha(0, 1, t)$ and $\beta(0, 1, t)$ as

$$\alpha(0,1,t) = \left[-\lambda + (4/3+\lambda)\left(\frac{1}{3}\right)^{t-1}\right]\beta(0,1,t), \tag{16}$$

where parameter λ satisfies $\frac{\alpha(0,1,t)}{\beta(0,1,t)} + \lambda = \frac{1}{3}\left(\frac{\alpha(0,1,t-1)}{\beta(0,1,t-1)} + \lambda\right)$. Substituting Eq. (16) into Eq. (15) yields

$$\beta(0,1,t) = 3^{3^{t-1}} \times \prod_{i=0}^{t-2}\left(3\left[-\lambda + (4/3+\lambda)\left(\frac{1}{3}\right)^{t-2-i}\right]^2\right)^{3^i}. \tag{17}$$

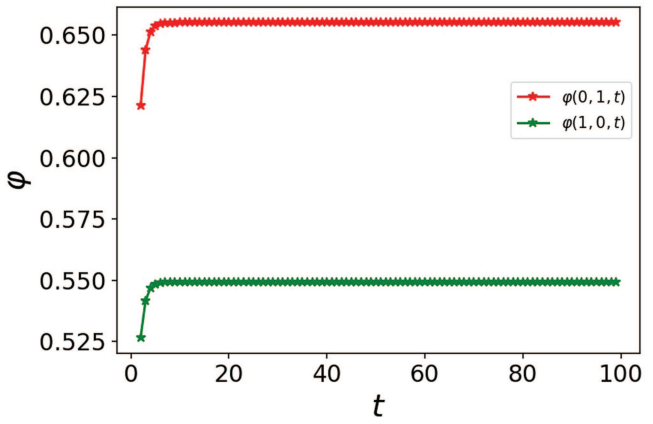

FIG. 9. The diagrams of spanning tree number entropies $\varphi(0, 1, t)$ and $\varphi(1, 0, t)$ of both network models $N(0, 1, t)$ and $N(1, 0, t)$, respectively, at time steps $1 \le t \le 100$.

On the basis of Eqs. (15) and (17), the solution to the total number of spanning trees of model $N(0, 1, t)$ can be written as

$$S(0,1,t) = 4 \times \left[-\lambda + (4/3+\lambda)\left(\frac{1}{3}\right)^{t-1}\right]^3 \times \left[3^{3^{t-1}} \times \prod_{i=0}^{t-2}\left(3\left[-\lambda + (4/3+\lambda)\left(\frac{1}{3}\right)^{t-2-i}\right]^2\right)^{3^i}\right]^4. \tag{18}$$

By both Eqs. (15) and (18), it is so transparent to the eye that type-B operation will generate more spanning tress than type-A operation. To further clarify the difference between $S(1, 0, t)$ and $S(0, 1, t)$, we make use of spanning tree entropy for models $N(1, 0, t)$ and $N(0, 1, t)$, respectively, defined as $\varphi(1, 0, t) = \ln S(1, 0, t)/|\mathfrak{V}(1, 0, t)|$ and $\varphi(0, 1, t) = \ln S(0, 1, t)/|\mathfrak{V}(0, 1, t)|$. The results shown in Fig. 9 coincide with our assertion that the spanning tree entropy corresponding to each member of graph space $\mathfrak{S}(p, p^c, t)$ must fall within a region limited by both red line and green line.

IV. CONCLUSION

In conclusion, we study a family of solvable random graphs with an exponential degree distribution. By contrast with those widely discussed random graphs, our graph models are generated by a novel algorithm $\mathscr{A}$, which is based on both types of operations, called type-A operation and type-B operation, respectively, and a probability parameter p. By the available tuning value p, we can obtain a great diversity of random graphs that constitute a graph space $\mathfrak{S}(p, p^c, t)$. Next, we consider some topological structure parameters on each member in graph space $\mathfrak{S}(p, p^c, t)$. Both analytical results and computer simulations state that all the members $N(p, p^c, t)$ of graph space $\mathfrak{S}(p, p^c, t)$ show some common topological

structure properties, including the small-world property, an identical exponential degree distribution, and assortative mixing. This suggests a fact that random graphs with an identical degree distribution are in most cases similar to one another, at least reasonable for our random graphs proposed in this paper because of their own similar diameter, clustering coefficient, and Pearson's correlation coefficient. On the other hand, one can stress using both Eqs. (15) and (18) that random graphs with an identical degree distribution are very likely to show completely different outlines from each other with respect to spanning tree number. This means that spanning tree number is a more sensible topological structural index than the other three mentioned above and hence can be exploited to better distinguish the difference among a number of random graphs with the same degree distribution. In addition, topics of interest will be discussed in graph space $\mathfrak{S}(p, p^c, t)$ in the near future, including random walks, synchronization, and percolation. Below are two interesting problems:

Problem 1. Is there a critical point at which the solution to an average path length takes place a transition? One reason for this is that many long-distance edges are added by *phase 2* into the resulting model.

Problem 2. What type of probability distribution does the sizes of those young edge-induced cycles in model $N(p, p^c, t)$ follow? One reason for this is that they are disjoint with each other.

ACKNOWLEDGMENTS

This research was supported by the National Key Research and Development Plan under Grant No. 2017YFB0800603 and the National Natural Science Foundation of China (NNSFC) under Grant No. 61662066.

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

REFERENCES

[1] P. Erdös and A. Rényi, Publ. Math. **6**, 290 (1959).
[2] A.-L. Barabási and R. Albert, Science **5439**, 509–512 (1999).
[3] D. J. Watts and S. H. Strogatz, Nature **393**, 440–442 (1998).
[4] B. Karrer and M. E. J. Newman, Phys. Rev. E **84**, 036106 (2011).
[5] L. Y. Zhang and J. L. Ren, J. Stat. Mech. **2019**, 033204).
[6] K. Hu, J. B. Hu, L. Tang *et al.*, J. Stat. Mech. **2018**, 100001).
[7] W. Liu, L. Y. Ma, B. Jeon, L. Chen, and B. Chen, J. Theor. Biol. **455**, 26–38 (2018).
[8] M. E. J. Newman, Contemp. Phys. **46**, 323–351 (2005).
[9] F. Ma, J. Su, Y. X. Hao, B. Yao, and G. H. Yan, Physica A **492**, 1194–1205 (2018).
[10] F. Ma, X. M. Wang, and P. Wang, Chaos **30**, 013136 (2020).
[11] J. Adrian Bondy and U. S. R. Murty, *Graph Theory with Application* (The MaCmillan Press Ltd., London, 1976).
[12] X. M. Wang and F. Ma, Chaos **30**, 043120 (2020).
[13] In a mathematical notation, for a given graph G, its clustering coefficient C can be denoted by the next relation $C = \frac{3\times\text{triangle number on graph } G}{\text{number of connected triples of vertices on graph } G}$. Here, a triangle is a cycle C_3 on three vertices and a connected triple means a vertex, which is connected to a pair of other vertices.
[14] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys. Rev. E **65**, 066122 (2002).
[15] M. E. J. Newman and J. Park, Phys. Rev. E **68**, 036122 (2003).
[16] Z. Z. Zhang, B. Wu, and F. Comellas, Discret. Appl. Math. **169**, 206–213 (2014).
[17] T. Hasunuma, Electron. Notes Discret. Math. **60**, 47–54 (2017).
[18] G. J. Szabó, M. Alava, and J. Kertész, Physica A **330**, 31–36 (2003).
[19] N. Takashi and A. E. Motter, Phys. Rev. E **73**, 065106 (2006).
[20] P. Marchal, Electron. Commun. Probab. **5**, 39–50 (2000).
[21] J. D. Noh and H. Rieger, Phys. Rev. Lett. **92**, 118701 (2004).
[22] Z. Z. Zhang, T. Shan, and G. Chen, Phys. Rev. E **87**, 012112 (2013).
[23] F. Ma and B. Yao, Theor. Comput. Sci. **708**, 46–57 (2018).