Acta Mathematica Sinica, English Series
Jan., 2013, Vol. 29, No. 1, pp. 199–208
Published online: September 20, 2012
DOI: 10.1007/s10114-012-0387-6
Http://www.ActaMath.com

Acta Mathematica Sinica, English Series

The Clustering Coefficient and the Diameter of Small-world Networks

Lei GU
School of Mathematical Sciences, Fudan University, Shanghai 200433*, P. R. China*
E-mail: *gehirn724@gmail.com*

Hui Lin HUANG
College of Mathematics and Information Science, Wenzhou University,
Wenzhou 325035*, P. R. China*
E-mail: *huilin_huang@wzu.edu.cn*

Xiao Dong ZHANG
Department of Mathematics and MOE-LSC,
Shanghai Jiao Tong University, Shanghai 200240*, P. R. China*
E-mail: *xiaodong@sjtu.edu.cn*

Abstract The small-world network, proposed by Watts and Strogatz, has been extensively studied for the past over ten years. In this paper, a generalized small-world network is proposed, which extends several small-world network models. Furthermore, some properties of a special type of generalized small-world network with given expectation of edge numbers have been investigated, such as the degree distribution and the isoperimetric number. These results are used to present a lower and an upper bounds for the clustering coefficient and the diameter of the given edge number expectation generalized small-world network, respectively. In other words, we prove mathematically that the given edge number expectation generalized small-world network possesses large clustering coefficient and small diameter.

Keywords Generalized small-world network, clustering coefficient, diameter, Laplacian spectra

MR(2000) Subject Classification 05C82, 05C50

1 Introduction

It was discovered that there are two important characterizations: i.e., large clustering coefficient, and the small-world phenomenon (small diameter, or short average distance) in the real complex networks. The small-world effect may be dated back the famous experiments carried out by Milgram [1] in the 1960s, in which letters passed from person to person were able to reach a designated target individual in only a small number of steps. In order to describe complex networks with the two properties, Watts and Strogatz [2] in 1998 proposed a famous small-world network model which possesses both small average distance and large clustering

Received July 20, 2010, revised August 23, 2011, accepted October 14, 2011
Supported by National Natural Science Foundation of China (Grant Nos. 10971137 and 11271256), National Basic Research Program of China 973 Program (Grant No. 2006CB805900) and the Grant of Science and Technology Commission of Shanghai Municipality (STCSM No. 09XD1402500)

coefficient by simulation if the wired probability $p \in [p_1, p_2]$. Their paper inspired empirical studies of complex networks by physicists, engineering researchers, biologists, mathematicians (for example, see [3–7] and the references therein). In 1999, Newman and Watts [5] proposed a variation of small-world network model. Moreover, Newman et al. [8] showed that their network model possessed small average distance and large clustering coefficient by mean field theory and numerical results. Barbour and Reinert [9, 10] extensively investigated many properties, such as the average distance and the clustering coefficient, of the small-world networks (including continuous and discrete small-world networks) from the probabilistic prospective. In an earlier model, Bollobás and Chung [11] showed that a graph combined a lattice with a random matching has small diameter. Dynamical properties on small-world networks [12–17] also has attracted much interest since their behavior is much different from lattice. Recently, Gu et al. [18] analyzed the dynamical property on small-world networks by estimating graphical and spectral properties.

In this paper, we are motivated by several small-world models and proposed a generalized small-world network, which may be regarded as a generalization of some known small-world networks. Motivated by the Watts–Strogatz model, we focus on a special type of generalized small-world network with given expectation of edge number.

The main result of this paper presents a lower and an upper bounds for the clustering coefficient and the diameter of the given edge number expectation generalized small-world network, respectively. Let $\mathcal{S}(n, k, p)$ be the generalized small-world network with edge number expectation $cn \log n$ (see Section 2). Then we have

Theorem 1.1 *Let G be any graph in $\mathcal{S}(n, k, p)$ with $k = c \log n$ and $c \geq 1.2$. Then almost surely the clustering coefficient has lower bound $\frac{1}{6}(1-p)^3$ while the diameter has upper bound $\frac{72}{p} \log^2 n$.*

In other words, we proved mathematical rigorously that the given edge number expectation generalized small-world network possesses large clustering coefficient and small diameter. The rest of the paper is organized as follows: In Section 2, a generalized small-world network is proposed and some preliminary results are presented. In Section 3, the connectivity, the isoperimetric number and the Laplacian spectral gap of the given edge number expectation generalized small-world network have been investigated. In Section 4, the obtained results are used to present a lower and an upper bounds for the clustering coefficient and the diameter of the given edge number expectation generalized small-world network, respectively.

2 The Generalized Small-world Network

The *small-world network* proposed by Watts and Strogatz can be described as follows: Let $C_{n,k}$ be a $2k$-circulant graph (or network) with vertex set $V(C_{n,k}) = \{v_1, v_2, \ldots, v_n\}$ and edge set $E(C_{n,k}) = \{(v_i, v_{i+t}) : 1 \leq i \leq n, 1 \leq t \leq k\}$ $(i + t \equiv j \bmod (n)$ if $i + t > n)$. Then one rewires every edge of $C_{n,k}$ with probability p by changing one end of an edge uniformly at random. No loops and multi-edges are allowed. The importance of the Watts and Strogatz's model is due to the fact that this model, roughly speaking, possesses both large clustering coefficient and small diameter. From a mathematical point of view, the experimental result in [2] was far from surprising. For the sake of better analysis of the small-world model, Newman et al. [8] introduced a modified form of the small-world network model which can be described as the addition of a $2k$-circulant graph $C_{n,k}$ plus a random graph of order n. However, there may be multiedges in the small-world model.

We are motivated by extensive study of varies small-world networks from different views. In order to analyze the small-world network from mathematical point, we present a mathematical definition of a generalized small-world network.

Definition 2.1 *For a given labeling vertex set* $V(G) = \{v_1, \dots, v_n\}$ *and* $0 \le p, q \le 1$, *let* $\mathcal{S}(n, k, p, q)$ *be a probability space with* $\mathbb{P}(v_i \sim v_{i+t}) = 1 - p$ *for* $i = 1, \dots, n, t = 1, \dots, k$; *otherwise* $\mathbb{P}(v_i \sim v_{i+t}) = q$, *where* $v_i \sim v_j$ *is the event of* v_i *adjacent to* v_j. *Any graph of* $\mathcal{S}(n, k, p, q)$ *is called a generalized small-world network.*

In other words, this probability space is regarded as an inhomogeneous bond percolation on $2k$-circulant graph $C_{n,k}$ with the probability to preserve the edge is $1 - p$; while for the complement graph of $C_{n,k}$ in the complete graph K_n with the probability to preserve the edge is q. When $i + t > n$, then $v_{i+t} = v_j$ for $i + t \equiv j \mod (n)$. This model may be applied to describe networks of a global company business model where local companies have a probability to cooperate and cross national companies have another probability to cooperate through Internet and other communication methods.

Remark Firstly, if $q = 1 - p$, it is exactly classical Erdös–Rényi [19] random graphs $G_{n,1-p}$.

Secondly, if $q = \frac{2kp}{n-2k-1}$, then the expectation of edge number is kn, which is the same as Watts–Strogatz's model. Denote the special type of generalized small-world network by $\mathcal{S}(n, k, p) \doteq \mathcal{S}(n, k, p, \frac{2kp}{n-2k-1})$. In the rest of the paper, we will focus on this type of generalized small-world network with given expectation of edge number, since it is easier and more convenient to study than Watts–Strogatz's model in some cases but highlights the structure of Watts–Strogatz's one.

Thirdly, if $p = 0$, any small-world network in $\mathcal{S}(n, k, 0, q)$ is just Newman–Watts model which is the union of a $2k$-circulant graph and a classical ER model random graph $G_{n,q}$. Hence in a sense, $\mathcal{S}(n, k, p, q)$ is a generalization of many small-world networks.

Let $G = (V, E)$ be a simple and undirected graph with vertex set $V = \{1, \dots, n\}$ and edge set E. Let $A(G) = (a_{ij})$ be the *adjacency matrix* of G whose entry

$$a_{ij} = \begin{cases} 1, & \text{if } i \sim j, \\ 0, & \text{otherwise.} \end{cases}$$

Let $d_i = \sum_j a_{ij}$ be the *degree* of the vertex i. Thus $D = \text{diag}(d_1, \dots, d_n)$ is called the *degree diagonal matrix* of G. The *combinatorial Laplacian matrix* of a graph G is then defined by

$$L(G) = D - A.$$

In this paper, if $\{a_n\}$ and $\{b_n\}$ are two series, we denote $a_n = o(b_n)$ as $\lim_{n\to\infty} \frac{a_n}{b_n} \to 0$.

Let d_i be the random variable of the degree of vertex i. Moreover, let $d_i^{(1)} = \sum_{j=i-k}^{i+k} a_{ij}$, where $i \pm k \equiv j \mod (n)$ if $i - k < 0$ or $i + k \ge n$ and $d_i^{(2)} \doteq d_i - d_i^{(1)}$. Obviously, $d_i^{(1)}$ and $d_i^{(2)}$ are independent random variables. Moreover, both $d_i^{(1)}$ and $d_i^{(2)}$ are binomial distributed random variables and

$$\mathbb{P}(d_i = s) = \sum_{s_1+s_2=s} \mathbb{P}(d_i^{(1)} = s_1, d_i^2 = s_2)$$

$$= \sum_{s_1+s_2=s} \binom{2k}{s_1}(1-p)^{s_1}p^{2k-s_1}\binom{n-2k-1}{s_2}q^{s_2}(1-q)^{n-2k-1-s_2}.$$

In order to present precise distribution, we need the following known result.

Theorem 2.2 (DeMoivre–Laplace [20, p. 13]) *Let $B_{n,p}$ be the random variable with distribution* $\mathrm{Bin}(n,p)$. *Suppose $0 < p < 1$ depends on n in such a way that $pqn = p(1-p)n \to \infty$ as $n \to \infty$. Suppose $h_1 < h_2, h_2 - h_1 \to \infty$ as $n \to \infty$ and $|h_1| + |h_2| = o\{(pqn)^{2/3}\}$. Put $h_i = x_i(pqn)^{1/2}$, $i = 1, 2$. Then*

$$\mathbb{P}(pn + h_1 \le B_{n,p} \le pn + h_2) \sim \Phi(x_2) - \Phi(x_1),$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_\infty^x \mathrm{e}^{-t^2/2}dt$. Moreover, when $h = x(pqn)^{1/2}$ and $h = o(pqn)^{2/3}$, we have

$$\mathbb{P}(B_{n,p} \ge pn + h) \sim \frac{1}{x\sqrt{2\pi}}\mathrm{e}^{-x^2/2}.$$

Hence the distribution of $B_{n,p}$ tends to normal distribution if $pqn = p(1-p)n \to \infty$ as $n \to \infty$. The generating function of $B_{n,p}$ also tends to the generating function of Gaussian random variable which implies the central limit theorem.

Corollary 2.3 (Central limit theorem) *Let $B_{n,p}$ be the random variable with distribution* $\mathrm{Bin}(n,p)$ *satisfying $pqn = p(1-p)n \to \infty$ as $n \to \infty$, $\mu = np$ and $\sigma^2 = np(1-p)$. Then* $\lim_{n\to\infty} \frac{B_{n,p}-\mu}{\sigma} = N(0,1)$ *in law.*

Let d_i be the random variable of the degree of vertex i from $\mathcal{S}(n,k,p)$. Thus $d_i = d_i^{(1)} + d_i^{(2)}$, where $d_i^{(1)} \sim \mathrm{Bin}\big(2k, 1-p\big)$ and $d_i^{(2)} \sim \mathrm{Bin}(n-2k-1, \frac{2kp}{n-2k-1})$. Assuming that $k \to \infty$ as $n \to \infty$ and $k \ll n$ which means that $k = o(n)$, it is easy to see

Lemma 2.4 *Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables. Suppose that they are independent of each other, $Z_n = a_nX_n + b_nY_n$, where a_n and b_n are two sequences of real numbers satisfying $a_n^2 + b_n^2 = 1$. If $X_n \Rightarrow N(0,1)$ and $Y_n \Rightarrow N(0,1)$ in law as $n \to \infty$, then $Z_n \Rightarrow N(0,1)$ in law.*

Theorem 2.5 *Let d_i be the random variable of the degree of vertex v from $\mathcal{S}(n,k,p)$ as before and $k \to \infty$ as $n \to \infty$ and $k \ll n$. Then the distribution of $\frac{d_i-2k}{\sqrt{2kp(2-p-\delta(n))}}$ tends to normal distribution $N(0,1)$, when $n \to \infty$, where $\delta(n) = \frac{2kp}{n-2k-1}$.*

Proof Set $X_n = d_i^{(1)} \sim \mathrm{Bin}(2k, 1-p)$ and $Y_n = d_i^{(2)} \sim \mathrm{Bin}\big(n-2k-1, \frac{2kp}{n-2k-1}\big)$. Then by Corollary 2.3, $\frac{X_n-2k(1-p)}{\sqrt{2kp(1-p)}} \Rightarrow N(0,1)$ and $\frac{Y_n-2kp}{\sqrt{2kp(1-\delta(n))}} \Rightarrow N(0,1)$ in law. Furthermore, let $a_n = \sqrt{\frac{2kp(1-p)}{2kp(2-p-\delta(n))}}$ and $b_n = \sqrt{\frac{2kp(1-\delta(n))}{2kp(2-p-\delta(n))}}$. The assertion follows from Lemma 2.4 and the independence of $\{X_n\}$ and $\{Y_n\}$. □

Combining Theorem 2.5 and Lemma 2.3 and since $\delta(n) \to 0$ as $n \to \infty$, we have

Corollary 2.6 *Let d_i be the random variable of the degree of vertex i from $\mathcal{S}(n,k,p)$ and $k \to \infty$ as $n \to \infty$ and $k \ll n$. Then the limit distribution of d_i is the limit distribution of* $\mathrm{Bin}(\frac{2k}{(1-p)^2}, (1-p)^2)$, *when $k(n) \to \infty$.*

The next result follows from Corollary 2.6 and is important in following sections.

Lemma 2.7 *The maximum degree Δ of any graph G in $\mathcal{S}(n,k,p)$ with $k = c\log n$ almost surely has an upper bound $2c\log n + (\frac{1}{3} + \sqrt{\frac{1}{9} + 4c})\log n$.*

Proof We will prove for any $\alpha > 1$, $\mathcal{S}(n,k,p)$ has an upper bound $2k + f(\alpha, c, p)$, where

$$f(\alpha, c, p) = \left(\frac{\alpha}{3} + \sqrt{\frac{\alpha^2}{9} + 4c\alpha}\right)\log n.$$

By Corollary 2.6, the limiting distribution of vertex degree is the same as $\mathrm{Bin}(\frac{2k}{(1-p)^2}, (1-p)^2)$. Then let $\mu = 2k$, $\lambda = s - \mu$. By Chernoff's bounds (see [21, p. 26, Theorem 2.1]), we have for vertex i,

$$\mathbb{P}(d_i \geq s) \leq \exp\left(\frac{-\lambda^2}{2(\mu + \lambda/3)}\right).$$

Setting $\lambda = \left(\frac{\alpha}{3} + \sqrt{\frac{\alpha^2}{9} + 4c\alpha}\right)\log n$, it is easy to see

$$\mathbb{P}(d_i \geq s) \leq n^{-\alpha}.$$

Hence

$$\mathbb{P}\left(\Delta \leq 2k + \left(\frac{\alpha}{3} + \sqrt{\frac{\alpha^2}{9} + 4c\alpha}\right)\log n\right) \leq n^{-\alpha+1} \to 0.$$

Thus the maximum degree of $\mathcal{S}(n,k,p)$ has an upper bound

$$2k + \left(\frac{\alpha}{3} + \sqrt{\frac{\alpha^2}{9} + 4c\alpha}\right)\log n.$$

Since it is true for any $\alpha > 1$, the maximum degree of $\mathcal{S}(n,k,p)$ has an upper bound

$$2c\log n + \left(\frac{1}{3} + \sqrt{\frac{1}{9} + 4c}\right)\log n. \qquad \square$$

3 The Connectivity and the Isoperimetric Number of Networks

In this section, we first investigate how the connectivity of any graph in $\mathcal{S}(n,k,p)$ depends on k or p, and then present a lower bound for the isoperimetric number of any graph in $\mathcal{S}(n,k,p)$.

Theorem 3.1 *Any graph G in $\mathcal{S}(n,k,p)$ with $k = c\log n$ and $c < 0.5$ is almost surely disconnected for all $0 < p < 1$.*

Proof It is easy to see that

$$\begin{aligned}\mathbb{P}(d_i = 0) &= p^{2k}\left(1 - \frac{2kp}{n-2k-1}\right)^{n-2k-1} \sim \exp(2k(\log p - p)) \\ &= \exp(2c\log n(\log p - p)) = n^{-\alpha},\end{aligned}$$

where $\alpha = 2c(p - \log p)$. Let $X = \sum_{i\in V} \mathbf{1}_{d_i=0}$. Since $c < 0.5$ and $\log p - p$ is monotonous increasing on $(0,1)$, $\alpha < 1$, we have $EX = n^{1-\alpha}$ and

$$\begin{aligned}EX(X-1) = {}& 2nkp^{4k-1}\left(1 - \frac{2kp}{n-2k-1}\right)^{2(n-2k-1)} \\ &+ n(n-2k-1)p^{4k}\left(1 - \frac{2kp}{n-2k-1}\right)^{2(n-2k-1)-1}.\end{aligned}$$

Moreover,

$$\begin{aligned}\mathrm{Var}(X) &= EX(X-1) + EX - (EX)^2 \\ &\leq np^{2k}\left(1 - \frac{2kp}{n-2k-1}\right)^{n-2k-1} + 2knp^{4k-1}(1-p)\left(1 - \frac{2kp}{n-2k-1}\right)^{2(n-2k-1)}\end{aligned}$$

$$+ n(n-2k)p^{4k}\left(1-\frac{2kp}{n-2k-1}\right)^{2(n-2k-1)-1}\frac{2kp}{n-2k-1}.$$

Let $\beta = p^{2k}\left(1-\frac{2kp}{n-2k-1}\right)^{n-2k-1}$ and n be large enough. Thus

$$\mathrm{Var}(X) \le n\beta + \frac{1-p}{p}2kn\beta^2 + 4kpn\beta^2.$$

Hence by Chybeshev inequality, we have

$$\mathbb{P}(X=0) < \frac{\mathrm{Var}(X)}{(EX)^2+\mathrm{Var}(X)} \to 0.$$

Therefore any graph is almost surely disconnected. □

In order to present the assertion on any graph in $\mathcal{S}(n,k,p)$, we need the following lemma. A walk on a graph means a series of vertexes where each pair of neighbor in the series has an edge in the graph.

Lemma 3.2 *Let $G=(V,E)$ be any graph in $\mathcal{S}(n,k,p)$ with $k=c\log n$ and $p=\exp(-\alpha/c)$, where $V=\{1,2,\dots,n\}$ and $\alpha>1$. Let $A_1=\{1,\dots,k\},\dots,A_i=\{(i-1)k+1,\dots,ik\},\dots,A_s=\{(s-1)k+1,\dots,sk\}, A_{s+1}=\{sk+1,\dots,n\}$. Then for any vertex v_i there exists almost surely a walk $W(v_i)=\{v_i,v_{i_1},\dots,v_{i_j},\dots\}$ starting with v_i in $\mathcal{S}(n,k,p)$ such that there exists a subset $S=\{u_1,u_2,\dots,u_s\}\subset W(v_i)$ with $|S|=s$, and $u_i\in A_i$ for $i=1,\dots,s$.*

Proof Let G be a graph in $\mathcal{S}(n,k,p)$ and $G'=G\cap C_{n,k}$. Then for any vertex $i\in V$, we can define $>_i=\{j, j\sim i \text{ in } C_{n,k}, j>i \text{ if } i\le n-k \text{ and } j+n>i \text{ if } i>n-k\}$. For any $j\in >_i$, we denote $j\succ i$. For any vertex i in G', let d_i^+ be the number of edges with ij connected in G' and $j\succ i$. Hence $\mathbb{P}(d_i^+)$ is $p^k=\exp(k(\log p))$. Furthermore,

$$\mathbb{P}(\forall\, i, d_i^+>0) = (1-\exp(k(\log p))^n = \left(1-\frac{1}{n^\alpha}\right)^n \to 1.$$

And almost surely from any vertex v_i, we can define a walk $W(v_i)$ from vertex v_i as follows: we begin with v_i, let v_{i_1} be the closest neighbor in $G\in\mathcal{S}(n,k,p)$. Then we find the closest neighbor of v_{i_1} to be v_{i_2} and to v_{i_j}, and $W(v_i)=\{v_i,v_{i_1},\dots,v_{i_j},\dots\}$. Now letting s denote $\lfloor n/k\rfloor$, we partition $\{1,2,\dots,n\}$ into $s+1$ parts as $A_1=\{1,\dots,k\}, A_2=\{k+1,\dots,2k\}, A_3=\{2k+1,\dots,3k\},\dots,A_s=\{(s-1)k+1,\dots,sk\}, A_{s+1}\{sk+1,\dots,n\}$. Remembering that walk just refers to the nearest neighbor which is only available to choose from the k neighbors in $C_{n,k}$, we can choose a subset $S\subset W(v_i)$ with $|S|=s$ and the elements lies in disjoint sets $A_1,\dots,A_{s+1}$. □

Theorem 3.3 *Any graph G in $\mathcal{S}(n,k,p)$ with $k=c\log n$ and $c\ge 1.2$ is almost surely connected for all $0<p<1$.*

Proof We consider the following two cases:

Case 1 $p>\frac{\log n}{2k}=\frac{1}{2c}$. Any graph in the probability space $\mathcal{S}(n,k,p)$ can be coupled to random graph probability space $\mathcal{G}_{n,\frac{2pk}{\log n}\frac{\log n}{n}}$, where for every $G\in\mathcal{G}_{n,\frac{2pk}{\log n}\frac{\log n}{n}}$ is a subgraph of $G'\in\mathcal{S}(n,c\log n,p)$. Both $\mathcal{G}_{n,\frac{2pk}{\log n}\frac{\log n}{n}}$ and $\mathcal{S}(n,c\log n,p)$ means probability space here. By [3, p. 65, Theorem 2.8.3], $\mathcal{G}_{n,\beta\frac{\log}{n}}$ is almost surely connected for $\beta>1$. Hence any graph G in $\mathcal{S}(n,c\log n,p)$ is connected in the first case.

Case 2 $p\le\frac{1}{2c}$. By $c>1.2$, we have $c\log 2c>1$. Let $\alpha\equiv c\log\frac{1}{p}$. Then $\alpha\ge c\log 2c>1$. Hence p can be written as follows: $p=\exp(-\frac{\alpha}{c})$ with $\alpha>1$. Hence by Lemma 3.2, we choose a pair of vertex v_i and $v_{i'}$ such that at least $s-1$ elements in two walks $W(v_i)$ and $W(v_{i'})$, with $S \subset W(v_i)$ and $S' \subset W(v_{i'})$ respectively, share a common set in $A_1 \cdots A_{s+1}$. Hence the probability that there is no edge between $W(v_i)$ and $W(v_{i'})$ is less than $p^{s-1} = o(n^{-2})$ and there are $n(n-1)/2$ such pairs. Hence each pair of vertexes is connected almost surely, and any graph G in $\mathcal{S}(n, c\log n, p)$ is connected for all $0 < p < 1$ in this case. □

Let $G = (V, E)$ be a graph with n vertexes and the adjacency matrix $A = (a_{ij})$. The Cheeger constant (or isoperimetric number [22]) of the graph G is defined by

$$\iota(G) = \min\left\{\frac{\sum_{i\in S, j\notin S} a_{ij}}{|S|}, S \subset V, 0 < |S| \le \frac{n}{2}\right\},$$

where the minimum is taken over all nonempty subsets S of V satisfying $|S| \le \frac{n}{2}$. The number $\iota(G)$ is also called isoperimetric number of the graph G. Given a subset S, denote by r_S the number of components of $\mathcal{S}(n, k, p)$. The method here to estimate the isoperimetric number was introduced by Durrett [3] (see p. 171). And the following Lemma requires connectivity of $\mathcal{S}(n, k, p)$.

Lemma 3.4 *For any $K, \epsilon > 0$, there is an N, such that for all s and $n > N$, the number of elements in the set*

$$\Gamma(s) = \{S \subset V,\ |S| = s,\ r_S \le r = Ks\}$$

is at most $\exp((1+\epsilon)Ks\log n)$.

Proof Let $\sharp\Gamma(s)$ be the cardinality of $\Gamma(S)$.

$$\sharp\Gamma(s) \le \binom{n}{r}\binom{s-1}{r-1} \le \binom{n}{r}\binom{s}{r}.$$

By Stirling's formula, we have

$$\sharp\Gamma(s) \le \left(\frac{ne}{r}\right)^r \left(\frac{se}{r}\right)^r = \exp[r(2 + \log n + \log s - 2\log r)]$$
$$\le \exp\left(Ks(2 + \log n + 2\log(1/K))\right) \le \exp\left((1+\epsilon)Ks\log n\right).$$

This completes the proof. □

Lemma 3.5 *Let G be any graph in $\mathcal{S}(n, k, p)$ with $k = c\log n$ and $c \ge 1.2$. Then almost surely*

$$\iota(\mathcal{S}(n, k, p)) \ge \frac{pc}{3}.$$

Proof Let $A(G) = (a_{ij})$ be the adjacency matrix of G. For $1 \le s = |S| \le \frac{n}{2}$, we consider the following two cases.

Case 1 $r_S \ge \frac{pks}{3\log n}$. By Theorem 3.3, G is almost surely connected and

$$\frac{\sum_{i\in S, j\notin S} a_{ij}}{s} \ge \frac{r_S}{s} \ge \frac{pk}{3\log n} = \frac{pc}{3}.$$

Case 2 $r_S \le \frac{pks}{3\log n}$. By Lemma 3.4, the number of such S is at most $\exp\left(\frac{25pks}{72}\right)$. For a fixed S, let $N_k(S)$ be the neighborhood in $C_{n,k}$ of set S.

$$X_S = \sum_{i\in S, j\notin S} a_{ij} = \sum_{i\in S, j\notin S, j\in N_k(S)} \xi_{ij} + \sum_{i\in S, j\notin S\cap N_k(S)} \eta_{ij},$$

where $\xi_{i,j}$ is the Bernoulli random variable with expectation $1-p$ and $\eta_{i,j}$ is the Bernoulli random variable with expectation $\frac{2kp}{n-2k-1}$. Let Y_S be binomial with distribution $\mathrm{Bin}\big(s(n -$

$s), \frac{2pk}{n-2k-1})$. Since $(1-p) > \frac{2kp}{n-2k-1}$ for sufficient large n, $\mathbb{P}(X_S < a) \le \mathbb{P}(Y_S < a)$. Let $\mu = s(n-s)\frac{2pk}{n-2k-1}$ be the expectation of Y_S. Applying Chernoff's bound, we have

$$\mathbb{P}\bigg(Y_S \le \frac{(95-\sqrt{8593})\mu}{108}\bigg) \le \exp\bigg(-\frac{13\mu}{36}\bigg).$$

On the other hand,

$$\mu = s(n-s)\frac{2pk}{n-2k-1} \ge pks.$$

Hence

$$\mathbb{P}\bigg(\frac{Y_S}{s} \le \frac{(95-\sqrt{8593})pk}{108}\bigg) \le \exp\bigg(-\frac{13pks}{36}\bigg).$$

Therefore, for any fixed $1 \le s \le \frac{n}{2}$,

$$\mathbb{P}\bigg(\frac{Y_S}{s} \le \frac{(95-\sqrt{8593})pk}{108}\bigg| |S| = s\bigg) \le \exp\bigg(-\frac{pks}{72}\bigg).$$

Thus for sufficient large n, we have $\frac{pk}{3\log n} \le \frac{(95-\sqrt{8593})pk}{108}$ and

$$\mathbb{P}\bigg(\iota(\mathcal{S}) < \frac{pk}{\log n}\bigg) \le \sum_{s=1}^{\frac{n}{2}} \exp\bigg(-\frac{pks}{72}\bigg) \le \bigg(1-\exp\bigg(-\frac{pk}{72}\bigg)\bigg)^{-1} \exp\bigg(-\frac{pk}{72}\bigg) \to 0.$$

We complete the proof. □

4 The Clustering Coefficient and the Diameter

In this section, we give our main results which are a lower and an upper bounds for the clustering coefficient and the diameter of the given edge number expectation generalized small-world network, respectively. Clustering is an important feature for small-world networks and the clustering coefficient is used to measure quantity of clustering.

Definition 4.1 *Let $G = (V, E)$ be a simple graph with vertex degree sequence $\{d_1, d_2, \ldots, d_n\}$. Denote by s the number of triangles in G. Then the clustering coefficient* [7] *C is defined as follows*

$$C = \frac{6s}{\sum_{i=1}^{n} d_i(d_i-1)}.$$

Lemma 4.2 *Let X be the number of triangles in any graph G in $\mathcal{S}(n,k,p)$. Then for any $0 < \epsilon < 1$, almost surely*

$$X \ge \frac{1}{2}(1-\epsilon)k(k-1)(1-p)^3 n.$$

Proof It is easy to see that the number of triangles in $C_{n,k}$ is equal to $\frac{1}{2}n(k-1)k$. Denote by γ the number of triangles in $\mathcal{S}(n,k,p) \cap C_{n,k}$. Then $\gamma = \sum_{i=1}^{n(k-1)k/2} \xi_i$, where ξ_i is Bernoulli random variable with expectation $(1-p)^3$. Obviously,

$$E\gamma = (k-1)k(1-p)^3 n/2.$$

On the other hand, $E\gamma(\gamma-1)$ is the expected number of 2-tuples. In other words, the expected number of two distinct triangles. If two triangles share one edges, the expectation of pairs of triangles are less than $\frac{3}{2}k^2(k-1)n(1-p)^5$. If two triangles share no edges, the expectation of pairs are less than $(E\gamma)^2$. If two triangles share more than one edges, they must be the same. Then

$$\mathrm{Var}(\gamma) = E\gamma(\gamma-1) + E\gamma - (E\gamma)^2 < \frac{3}{2}k^2(k-1)n(1-p)^5 + \frac{1}{2}(k-1)k(1-p)^3 n.$$

Hence by the Chybeshev's inequality,

$$\begin{aligned}\mathbb{P}(X < (1-\epsilon)E\gamma) &\leq \mathbb{P}(\gamma < (1-\epsilon)E\gamma) \\ &\leq \frac{\frac{3}{2}k^2(k-1)n(1-p)^5 + \frac{1}{2}(k-1)k(1-p)^3 n}{\epsilon^2[\frac{1}{4}(k-1)^2k^2(1-p)^6n^2]}.\end{aligned}$$

Therefore, $\mathbb{P}(X < (1-\epsilon)E\gamma) \to 0$ when $n \to \infty$. So the assertion holds. □

Theorem 4.3 *Let C be the clustering coefficient of any graph G in $\mathcal{S}(n,k,p)$ with $k = c\log n$. Then for any $0 < \epsilon < 1$, almost surely*

$$C \geq \frac{3(1-\epsilon)c^2(1-p)^3}{(2c+\frac{1}{3}+\sqrt{\frac{1}{9}+4c})^2},$$

which is bounded away from zero when $n \to \infty$.

Proof It is easy to see that the assertion follows from Lemmas 4.2 and 2.7. □

In order to present an upper bound for the diameter of any graph in $\mathcal{S}(n,k,p)$, we also need the following results.

Lemma 4.4 ([23]) *Let G be a graph with n vertexes. Denote by $\lambda_2(G)$ and $\iota(G)$ the second smallest eigenvalue (which is called the algebraic connectivity) and the Cheeger constant of G, respectively. Then $\lambda_2(G) \geq \frac{\iota(G)^2}{2\Delta(G)}$, where $\Delta(G)$ is the maximum degree.*

Theorem 4.5 *The algebraic connectivity $\lambda_2(G)$ of any graph G in $\mathcal{S}(n,k,p)$ with $k = c\log n$ and $c \geq 1.2$ is almost surely bounded below by*

$$\frac{p^2c^2}{18(2c+\frac{1}{3}+\sqrt{\frac{1}{9}+4c})\log n}.$$

Proof The assertion follows from Lemmas 3.5, 2.7 and 4.4. □

Lemma 4.6 ([22]) *Let G be a connected graph with n vertexes and the algebraic connectivity $\lambda_2(G)$. Then the diameter $D(G)$ of G is at most*

$$D(G) \leq 2[\sqrt{(2\Delta(G)/\lambda_2(G))}\log_2 n],$$

where $\Delta(G)$ is the maximum degree.

Then we are able to prove the following theorem.

Theorem 4.7 *Let G be any graph in $\mathcal{S}(n,k,p)$ with $k = c\log n$ and $c \geq 1.2$. Then the diameter $D(G)$ is almost surely upper bounded below that*

$$D(G) \leq \frac{12(2c+\frac{1}{3}+\sqrt{\frac{1}{9}+4c})}{cp}\log n\log_2 n.$$

Proof The assertion follows from Lemmas 4.6, 4.5 and 2.7. □

Remark The righthand side of inequality in Theorem 4.3 besides $3(1-p)^3$ can be written as $\frac{1-\epsilon}{(2+\frac{1}{3c}+\sqrt{\frac{1}{9c^2}+\frac{4}{c}})^2}$, which is monotonous increasing with respect to c. The righthand side of inequality in Theorem 4.7 besides $\log n\log_2 n$ is $\frac{24+\frac{4}{c}+12\sqrt{\frac{1}{9c^2}+\frac{4}{c}}}{p}$, which is monotonous decreasing with respect to c. Combining these together with the connectivity requirement $c \geq 1.2$, we finally arrive at the main theorem.

Theorem 1.1 *Let G be any graph in $\mathcal{S}(n,k,p)$ with $k = c\log n$ and $c \geq 1.2$. Then almost surely the clustering coefficient has lower bound $\frac{1}{6}(1-p)^3$ while the diameter has upper bound $\frac{72}{p}\log^2 n$.*

We can see that the given edge number expectation generalized small-world network $\mathcal{S}(n,k,p)$ still possesses the two importance characterizations: the small diameter and large clustering coefficient for $k \geq 1.2\log n$ and for all constant $0 < p < 1$ independent of n. Therefore, we prove mathematically that the small-world networks have the large clustering coefficient and small diameter.

Acknowledgements The authors wish to thank the referees for their valuable comments and suggestions.

References

[1] Milgram, S.: The small world problems. *Psychology Today*, **2**, 67–70 (1967)
[2] Watts, D. J., Strogatz, S. H.: Collective dynamics of "small world networks". *Nature*, **393**, 440–442 (1998)
[3] Durett, R.: Random Graph Dynamics, Cambridge University, Cambridge, 2006
[4] Newman, M. E. J.: The structure and function of complex netowrks. *SIAM Rev.*, **45**, 167–256 (2003)
[5] Newman, M. E. J., Watts, D. J.: Renormalization group analysis of the small-world network model. *Phys. Lett. A*, **263**, 341–346 (1999)
[6] Watts, D. J.: Six Degrees, Norton, New York, 2003
[7] Newman, M. E. J., Watts, D. J., Stogatz, S. H.: Random graph models of social networks. *Proc. Natl. Acad. Sci. USA*, **99**, 2566–2572 (2002)
[8] Newman, M. E. J., Moore, C., Watts, D. J.: Mean-field solution of the small world network model. *Phys. Rev. Lett.*, **84**, 3201–3204 (2000)
[9] Barbour, A. D., Reinert, G.: Small worlds. *Random Structures and Algorithms*, **19**, 54–74 (2001)
[10] Barbour, A. D., Reinert, G.: Discrete small world networks. *Electron. J. Combin.*, **11**(47), 1234–1283 (2006)
[11] Bollobás, B., Chung, F.: The diameter of a cycle plus a random matching. *SIAM J. Discrete Math.*, **1**, 328–333 (1988)
[12] Barahona, M., Pecora, L. M.: Synchronization in small-world systems. *Phys. Rev. Lett.*, **89**, 054101 (2002)
[13] Gade, P. M., Hu, C. K.: Synchronous chaos in coupled map with small-world interactions. *Phys. Rev. E*, **62**, 6409–6413 (2000)
[14] Hong, H., Choi, M. Y., Kim, B. J.: Synchronization on small-world networks. *Phys. Rev. E*, **65**, 026139 (2002)
[15] Lago–Fernandez, L. F., Huerta, R., Corbacho, et al.: Fast response and temporal coherent oscillations in small-world networks. *Phys. Rev. Lett.*, **84**, 2758–2761 (2000)
[16] Olfati–Saber, R.: Ultrafast consensus in the small-world networks. In: Proceedings of American Control Conference, 2005, 2371–2378
[17] Wang, X. F., Chen, G.: Synchronization in small-world dynamical networks. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.*, **12**, 187–192 (2002)
[18] Gu, L., Zhang, X. D., Zhou, Q.: Consensus and synchronization problems on small-world networks. *J. Math. Phys.*, **51**(8), 082701 (2010)
[19] Erdös, P., Rényi, A.: On the evolution of random graphs. *Publication of Mathematical Institute of the Hungarian Academy of Sciences*, **5**, 17–61 (1960)
[20] Bollobás, B.: Random Graphs, second edition, Cambridge University, Cambridge, 2001
[21] Janson, S., Luczak, T., Rucinski, A.: Random Graphs, John Wiley, New York, 2000
[22] Alon, N., Milman, V. D.: λ_1, Isoperimetric inequalities for graphs, and superconcentrators. *J. Combin. Theory Ser. B*, **38**, 73–88 (1985)
[23] Berman, A., Zhang, X. D.: Lower bounds for the eigenvalues of Laplacian matrices. *Linear Algebra Appl.*, **316**, 13–20 (2000)