SMALL-WORLD GRAPHS: CHARACTERIZATION AND ALTERNATIVE CONSTRUCTIONS

RAMA CONT,* *Columbia University*

EMILY TANIMURA,** *Ecole des Hautes Etudes en Sciences Sociales*

Abstract

Small-world graphs are examples of random graphs which mimic empirically observed features of social networks. We propose an intrinsic definition of small-world graphs, based on a probabilistic formulation of scaling properties of the graph, which does not rely on any particular construction. Our definition is shown to encompass existing models of small-world graphs, proposed by Watts (1999) and studied by Barbour and Reinert (2001), which are based on random perturbations of a regular lattice. We also propose alternative constructions of small-world graphs which are not based on lattices and study their scaling properties.

Keywords: Random graph; small-world model

2000 Mathematics Subject Classification: Primary 05C80

1. Introduction

1.1. Complex networks and small-world graphs

Empirical studies of a wide variety of social and biological networks [2], [20], [21] [23], [24], [25] have revealed that many of them share several interesting properties:

- links among network nodes are globally sparse, i.e. the network is far from saturating the total number of possible links;

- high *local clustering* of links, i.e. the link structure displays a high local density, as measured for instance by the clustering coefficient (see below);

- although the network may contain a large number of nodes, a pair of nodes in the network is typically linked by a path whose length is orders of magnitude smaller than the network size and grows slowly with the number of nodes.

These properties distinguish real networks from simple models such as regular lattices or the Erdős–Rényi random graph model [13], which possess one or two of these properties but not all of them, and have inspired the development of a new class of random graph models, called *small-world networks* [20], [25], [26], which have generated, in turn, a host of applications and new mathematical problems [2], [10], [25].

Intuitively, a small-world graph is a random graph which possesses the three properties above. The prototype of the small-world graph, given by Watts [25], is a crossover between a regular lattice and an Erdős–Rényi random graph. The properties of the lattice-based model in [25] has been extensively studied using Monte Carlo simulation and via a mean-field approximation [22].

In the economics literature, Jackson [18] studied graphs that result from link formation in a game theoretic setting. The Nash equilibrium in this network formation game is a deterministic graph which exhibits high clustering and short distances between nodes for a wide range of parameters in the game. These properties are shown to hold at least in a high connectivity range, where the average degree grows rapidly with network size.

These small-world graphs are complex graph models whose mathematical study can be challenging. Barbour and Reinert [4] have proposed a rigorous analysis of a slightly modified version of the original construction of Watts. Aldous [1] proposed a growth model in which new nodes are added one at a time and form links to previous nodes according to given probabilistic rules: local weak convergence methods are used to show that parameters in this model can be chosen so that the resulting graph is sparse and exhibits high clustering. Bollobás *et al.* [8] proposed a framework for studying a general class of large inhomogeneous random graphs. For a recent review on the mathematical study of complex networks, see [10].

While these studies focus on properties of specific graph models, a common mathematical framework encompassing all these models and enabling us to compare them seems to be lacking in the literature: such a framework could emphasize the common features of different constructions and shed light on the mechanisms leading to the emergence of the small-world property in real networks. As noted by Watts [25], the small-world property should be stated in terms of scaling of the graph properties with the size of the graph. A reasonable definition of the small-world property should apply to the lattice-based construction [25], but could potentially include qualitatively different constructions. In particular, applications in social sciences are not naturally based on lattice models, so lattice-free constructions are desirable.

1.2. Outline

We propose here an intrinsic definition of small-world graphs which does not rely on an underlying lattice nor on any particular construction. Our definition is based on a probabilistic formulation of scaling properties of graph-theoretical quantities and encompasses existing models of small-world graphs, proposed by Watts and studied by Barbour and Reinert [4], which are based on random perturbations of a regular lattice.

Such lattice-based constructions are not natural for applications to social networks, where regular lattices are nowhere to be found. In the second part of the paper we propose two examples of small-world graphs which are based on a *community* structure to which random links are then added. As in Watts' model, the proposed models lead to large inhomogeneous random graphs, which match various empirical properties of social networks.

The paper is structured as follows. In Section 2 we recall some basic notions on (random) graphs and define a mathematical setting suitable for our purpose. In Section 3 we define the notion of *scaling behavior* in a way which is meaningful for random graphs.

Based on these definitions, we propose in Section 4 a mathematical definition of small-world graphs which is intrinsic in the sense that it does not rely on a particular construction. In Section 5 we show that this definition applies to Watts' randomly perturbed lattice model, in the setting of [4].

In Sections 6 and 7 we propose two examples of random graph models which satisfy the definition of a small world but whose construction is not based on an underlying lattice.

2. Definitions and notation

Let us start by defining a mathematical framework, which will allow us to make precise statements about networks, their statistical properties, and the scaling properties of various graph-theoretical quantities with the size of the network. In particular, our framework should allow for

- varying the number of nodes of the graph, in order to study large-size asymptotics and scaling;

- randomness in the structure of the network: many network models, including Erdős–Rényi graphs and the small-world construction of Watts [25], are instances of random graphs. Allowing for randomness is especially relevant in applications since properties of large networks can only be described in statistical terms.

2.1. Graphs and graph properties

A graph with nodes labeled $i = 1, \ldots, N$ is defined by the set of its links, which can be viewed as a subset Γ of $\{1, \ldots, N\} \times \{1, \ldots, N\}$. Since we have in mind models for social networks, we exclude links between a node and itself: $(i, i) \notin \Gamma$. Such a graph is conveniently represented by an $N \times N$ adjacency matrix M defined by

$$M(i, j) = \begin{cases} 1 & \text{if } (i, j) \in \Gamma, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The set of nodes of a graph Γ will be denoted by $[\Gamma]$ or, with an abuse of notation when the context is clear, by Γ. In the sequel we consider undirected graphs, i.e. such that $(i, j) \in \Gamma$ if and only if $(j, i) \in \Gamma$, although the definitions in Section 3 also apply to directed graphs. The set G_N of undirected graphs with N nodes is in one-to-one correspondence with symmetric adjacency matrices:

$$\{M \in M_{N \times N} \mid M(i, j) = M(j, i) \in \{0, 1\}, \ M(i, i) = 0\}. \tag{2}$$

We will denote by

$$G_\infty = \bigcup_{N \geq 1} G_N$$

the set of all (undirected) graphs, endowed with the cylindrical Borel σ-field $\mathcal{B}$: a function $Q \colon G_\infty \mapsto \mathbb{R}$ is $\mathcal{B}$-measurable if, for all $N \geq 1$, its restriction $Q_N \colon G_N \mapsto \mathbb{R}$ is measurable. Examples of interest are the degree of a node, the average degree, the (average) clustering coefficient, the typical interpoint distance, and the diameter of a graph.

The degree of a node i is defined as the number of nodes it is linked to:

$$\deg(i) = \sum_{n=1}^{N} \mathbf{1}_{M(i,n)=1} \,.$$

The average degree of a graph $\Gamma \in G_N$ is defined as

$$\overline{\deg}(\Gamma) = \frac{1}{N} \sum_{i=1}^{N} \deg(i). \tag{3}$$

The neighborhood $V(i)$ of a node i is defined as

$$V(i) = \{j \in [1, N] \mid M(i, j) = 1\}$$

and the set $V^k(i)$ as

$$V^k(i) = \{j \in [1, N] \mid M^k(i, j) > 0\}.$$

Note that $V^1(i) = V(i)$.

Empirical studies on social networks indicate that they are characterized by high local density, as measured by the notion of the *clustering coefficient*. The local clustering coefficient for a node i is defined as

$$\gamma_i = \frac{\text{card}\{(j, k) \in \Gamma \mid j \in V(i) \text{ and } k \in V(i)\}}{\text{card}(V(i))(\text{card}(V(i)) - 1)/2}.$$

Expressed as a function of the adjacency matrix M,

$$\gamma_i = \frac{\sum_{j \neq k \mid M(i,j)=M(i,k)=1} M(j, k)}{\sum_{l=1}^{N} M(i, l)(\sum_{l=1}^{N} M(i, l) - 1)/2}.$$

The average clustering coefficient γ is defined as

$$\gamma = \frac{1}{N} \sum_{i=1}^{N} \gamma_i.$$

Another quantity of interest is the 'size' of the graph as measured by the distance between the nodes. The distance between two nodes i and j is defined as the length of the shortest path linking them in the graph:

$$d_\Gamma(i, j) = \inf\{k \geq 1, M^k(i, j) > 0\}.$$

When the context is clear, we will simply denote the distance by $d(i, j)$. The average interpoint distance is given by

$$\bar{d}(\Gamma) = \frac{1}{N(N-1)} \sum_{i \neq j} d_\Gamma(i, j).$$

Another measure of distance in the graph is the diameter defined as

$$D(\Gamma) = \max_{i \neq j} d_\Gamma(i, j).$$

If the graph is not connected, $D(\Gamma) = \infty$.

2.2. Random graphs

A random graph of size N is a random variable Γ defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in G_N. The graph-theoretical quantities defined above (diameter, clustering coefficient, average interpoint distance, etc.) are then measurable functions of Γ and define random variables on $(\Omega, \mathcal{F}, P)$.

Many common random graph models result in sample graphs which may have more than one connected component with nonzero probability. In this context, measures of graph size such as the diameter are not finite valued. Also, computation of the diameter requires exhaustive

knowledge of all the links in the graph, so it is not an observable quantity for an observer who has access to a 'representative sample' of the whole network, such as a social scientist conducting a survey. The *typical interpoint distance*, defined as the distance between a pair of nodes chosen at random, is a more flexible notion of 'size' in this case. The typical interpoint distance of a graph Γ is the random variable defined as

$$T(\Gamma) = d_\Gamma(U),$$

where U is uniformly drawn among the links $\{(i, j) \mid 1 \leq i \leq N, \ 1 \leq j \leq N, \ i \neq j\}$ and independent from Γ. It is readily observed that the law of the typical interpoint distance depends only on the graph Γ itself.

Note that the typical interpoint distance is a random variable, even if the underlying graph is deterministic. In a graph that consists of several large connected components or a single one with a few isolated nodes, the typical interpoint distance provides more relevant information than other measures of distance.

Example 1. Consider a sequence S_n of graphs with diameters bounded by a constant C, and consider the graph Γ^N obtained by adding an isolated point to S_n. For any $N \geq 2$, $\bar{d}(\Gamma^N) = \infty$, while the typical interpoint distance verifies

$$P(T(\Gamma^N) \leq C) = 1 - \frac{2}{N};$$

so, as $N \to \infty$, we have

$$\limsup P(T(\Gamma^N) \geq C + \varepsilon) \to 0 \quad \text{for all } \varepsilon > 0,$$

i.e. $T(\Gamma^N)$ may converge in probability to a finite-valued random variable, whereas the average distance verifies $P(\limsup \bar{d}(\Gamma_N) = \infty) = 1$.

3. Scaling behavior of graph properties

A property often discussed in the literature on social networks is the scaling of various graph-theoretical quantities with the graph size [2], [23], [25]. Consider a sequence $(\Gamma^N)_{N \geq 1}$ of graphs with $\Gamma^N \in G_N$ and Q, a graph-theoretical quantity, defined as a measurable function on $Q \colon G_\infty \mapsto \mathbb{R}$. We can define in the following way the concept of scaling behavior for the quantity $Q(\Gamma^N)_{N \geq 1}$.

Definition 1. (*Upper-scaling bound.*) Let $(\Gamma^N)_{N \geq 1}$ be a sequence of graphs with $\Gamma^N \in G_N$. A deterministic sequence $(g(N))_{N \geq 1}$ is called an *upper-scaling bound* for $(Q(\Gamma^N))_{N \geq 1}$ if

$$\limsup_{N \to \infty} \frac{Q(\Gamma^N)}{g(N)} \leq 1.$$

The notion of the lower-scaling bound can be defined analogously.

Definition 2. (*Lower-scaling bound.*) Let $(\Gamma^N)_{N \geq 1}$ be a sequence of graphs with $\Gamma^N \in G_N$. A deterministic sequence $(g(N))_{N \geq 1}$ is called a *lower-scaling bound* for $(Q(\Gamma^N))_{N \geq 1}$ if

$$\limsup_{N \to \infty} \frac{Q(\Gamma^N)}{g(N)} \geq 1.$$

In some cases, upper and lower scaling bounds may coincide up to a multiplicative constant. This is the case if there is an $f(N)$ that belongs to the upper scaling bounds, a $g(N)$ belonging to the lower scaling bounds, and a $C > 0$ such that $\lim_{N \to \infty} f(N)/g(N) = C$. In this case we will say that '$Q(\Gamma^N)$ scales like $f(N)$'.

Example 2. (*Scaling behavior in nearest-neighbor lattices.*) Consider the d-dimensional lattice $\{-n, \ldots, 0, \ldots, n\}^d$ with nearest-neighbor links. The corresponding graph Γ^N has $N = (2n + 1)^d$ nodes and it is readily observed that, as $n \to \infty$,

- the average degree converges to $2d$;

- the diameter (which, in this case, is the length of the diagonal) scales as $cN^{1/d}$;

- the clustering coefficient is 0: $\gamma_i = 0$ for any node i.

In this case, these quantities give both the upper and lower scaling bounds, which coincide.

When $(\Gamma^N)_{N \geq 1}$ is a sequence of random graphs, the above definitions must be interpreted probabilistically. There are several possibilities, depending on the mode of convergence considered.

Definition 3. (*Scaling.*) Let $(\Gamma^N)_{N \geq 1}$ be a sequence of random graphs, defined on the probability space $(\Omega, \mathcal{F}, P)$. Consider a measurable function $Q \colon G_\infty \to (0, \infty)$. A deterministic sequence $(g(N))_{N \geq 1}$ is said to be

1. an upper-scaling bound *in expectation* of $Q(\Gamma^N)$ if

$$\limsup_{N \to \infty} \frac{\mathrm{E}[Q(\Gamma^N)]}{g(N)} \leq 1;$$

2. an upper-scaling bound *in probability* if

$$\limsup_{N \to \infty} \mathrm{P}\left(\frac{Q(\Gamma^N)}{g(N)} \leq 1 \right) = 1;$$

3. an *almost-sure* upper-scaling bound if

$$\mathrm{P}\left(\limsup_{N \to \infty} \frac{Q(\Gamma^N)}{g(N)} \leq 1 \right) = 1.$$

The above possibilities are not equally restrictive. It is thus possible, for example, that the smallest almost-sure upper bound is orders of magnitude greater than the upper bound in probability or in expectation.

Example 3. (*Erdős–Rényi random graph.*) The simplest random graph model studied by Erdős and Rényi [13] is one where, in a graph with N nodes, each link is established independently from the others, with probability $p(N)$. A graph configuration $G \in G_N$ is thus drawn with probability

$$\mathrm{P}(\Gamma^N = G) = p(N)^{l(G)}(1 - p(N))^{N(N-1)/2 - l(G)},$$

where $l(G)$ is the total number of links in G. We usually consider the case where $p(N) \to 0$. Erdős–Rényi graphs exhibit the following scaling properties [5], [13].

- The expected degree of a given node scales as $Np(N)$ when $N \to \infty$.

- (From Theorem 7.3 of [5].) When the expected degree of a link $Np(N)$ verifies

$$Np(N) = \ln N + f(N) \quad \text{with} \quad \lim_{N \to \infty} f(N) = -\infty,$$

  then

$$P(\{\text{there exists } N_0 \geq 1, \text{ for all } N \geq N_0, \Gamma^N \text{ is disconnected}\}) = 1.$$

  If $\lim_{N \to \infty} f(N) = +\infty$ then $\ln N / \ln(\ln N)$ is an almost-sure upper-scaling bound of the diameter. In an even higher connectivity range, $p(N)N = N^{1/d}$, the constant $d + 1$ defines an almost-sure upper-scaling bound.

- The expected average clustering coefficient is $p(N)$. Since the degree of any node has expectation $p(N)N$, the expected average clustering coefficient cannot have a lower-scaling bound greater than 0 unless the expected degree of a node tends to ∞ as $N \to \infty$, not a very realistic situation for social networks.

4. Small-world graphs: an intrinsic definition

Having defined scaling, we will now attempt to cast in our framework the characterization of small-world graphs, as formulated by Watts [25].

Watts' starting point is a set of empirical observations on the structure of social networks. Social networks are complex networks which are characterized by a number of common features:

- a large number of nodes;

- sparsity of links, i.e. the number of existing links is far from saturating the total number of possible links;

- a high degree of clustering, as measured for instance by the clustering coefficient; and

- length scales, such as the typical interpoint distance, which are orders of magnitude smaller than the number of nodes.

Comparing these properties with those of two well-known classes of graphs—regular lattices and Erdős–Rényi random graphs—Watts noted that, while neither of these possessed all of the desired properties, the Erdős–Rényi graph can exhibit realistic distance scaling properties, while regular lattices can exhibit high local clustering as the network size grows. This led to the following conjecture.

Conjecture 1. ([25].) *There exists a class of graphs that are highly clustered yet whose characteristic length and diameter scales are similar to Erdős–Rényi random graphs. These graphs are called small-world graphs.*

As we have seen from previous examples, the scaling of quantities in the graph depends on the total number of links. Comparing different graphs in terms of the scaling of quantities, such as clustering or distance, is only meaningful when the average degree is similar. It is thus necessary to impose a condition on the scaling of the average degree. By doing so we also exclude trivial examples; indeed, in the highest connectivity range, when the average degree scales as cN, even a lattice would have a diameter bounded by a constant. Thus, the cases of interest are those where the average degree is orders of magnitude smaller than N.

When this is the case, obtaining small distances between nodes is not trivial. As an example, a one-dimensional kth-nearest-neighbor lattice with an average degree as high as $2k = N^{1/d}$

would still have a large diameter scaling as $N^{1-1/d}$. Thus, even in this high connectivity range, efficient organization of links is required for the distances in the graph to be short. Bearing in mind the relevance to social networks, we will limit ourselves to lower connectivity ranges, although the existence of clustered graphs that have short diameter when the average degree scales as $N^{1/d}$ is mathematically nontrivial. Most models in the literature, including Watts' model, assume that the average degree is bounded by a constant. We think it is reasonable to impose a slightly more lenient condition, allowing also for an average degree scaling as $C \ln N$. As we have seen, this is the case for the connected Erdős–Rényi graph. In practice, a degree scaling as $C \ln N$, although unbounded, remains small even for large N.

Based on this discussion, we propose the following definition, which allows us to capture these features.

Definition 4. (*Small worlds.*) A network model $(\Gamma^N)_{N \geq 1}$ is said to be a small world if the following three conditions are verified.

- There is a constant $C_1 \geq 0$ such that $C_1 \ln N$ is an upper-scaling bound of the average degree $\overline{\deg}(\Gamma^N)$.

- There is a constant $C_2 \geq 0$ such that $C_2 \ln N$ is an upper-scaling bound for the typical interpoint distance.

- The clustering coefficient is bounded away from 0.

Each of these properties may hold almost surely, in probability or in expectation.

Remark 1. A related class of complex networks which has received much attention is that of *scale-free* networks [3], [7], in which the degree distributions have Pareto tails. The Barabasi–Albert model [3], which is the prototypical example of a scale-free graph, has been shown by Bollobás *et al.* [9] to exhibit small distances between nodes [6], but not to exhibit clustering: the clustering coefficient decreases to 0 with the size of the graph [7]. Thus, scale-free graphs need not be small worlds in the sense of Definition 4. It is also obvious that Watts' small-world graph is not a scale-free graph.

5. Lattice-based constructions of small worlds

Let us now recall the original small-world construction of Watts [25], [26] and show that it verifies Definition 4.

The Watts model starts from an initial graph L, chosen to be a regular lattice of low dimension. In this graph, each of the existing links is randomly rewired with small probability, resulting in a random perturbation of the original lattice. This model typically leads to a disconnected graph with nonzero probability, but this problem can be avoided by adding additional links to the initial graph instead of rewiring, as in [4].

We now describe the variation of the Watts model proposed by Barbour and Reinert [4]. Barbour and Reinert constructed a graph by superposing a ring lattice L with N nodes and an Erdős–Rényi random graph E. In the ring lattice, each node is connected to the $2k$ neighbors within distance k in the lattice. In the Erdős–Rényi graph, each link exists with probability p/N. The resulting graph, Γ_N, has links defined by $(i, j) \in \Gamma_N$ if and only if $(i, j) \in L$ or $(i, j) \in E$.

Proposition 1. *In the perturbed lattice model described above, with $\frac{1}{3} \leq p < 1$ and $k \geq 2$, we have*

- *$2 \ln N$ is an upper-scaling bound of the typical interpoint distance;*

- *the average degree is almost surely bounded by $2k + 1$, where $2k$ is the degree of the lattice; and*

- *the average clustering coefficient is almost surely bounded below by $\bar{\gamma}_L(2k-1)/(2k+1)$, where $\bar{\gamma}_L$ is the average clustering coefficient in the lattice.*

Let us recall some results shown in [4]. To approximate the typical interpoint distance between two nodes, we consider independent branching processes originating from each other and branching through nearest neighbors on the small-world graph, with branching rate $2k$. The typical interpoint distance is evaluated by estimating the probability that two independent branching processes of this type do not intersect before a given time. While the actual process is complicated due to the possible overlap of intervals, it can be approximated using a pure-growth process whose behavior is easier to characterize.

This growth process must take into account two types of intervals with different branching rates, those that have just been created and consist of only one point which can be an endpoint for a random link, and those of more than one point where $2k$ new deterministic neighbors at each time are potential endpoints of random links. Denote by $M^1(n)$ the number of one point intervals at time n, and denote by $M^2(n)$ the number of intervals containing more than one point at time n. Then the process

$$M(n) = \begin{pmatrix} M^1(n) \\ M^2(n) \end{pmatrix}$$

is defined recursively by

$$M^1(n) = \text{Bin}\left((M^1(n-1) + 2kM^2(n-1))N, \frac{p}{N} \right), \qquad M^1(0) = 1,$$
$$M^2(n) = M^1(n-1) + M^2(n-1), \qquad M^2(0) = 0.$$

The expected evolution of the process verifies

$$\text{E}[M(n) \mid \mathcal{F}_{n-1}] = AM(n-1),$$

where $\mathcal{F}_{n-1} = \sigma(M(0), \ldots, M(n-1))$ and

$$M = \begin{pmatrix} p & 2kp \\ 1 & 1 \end{pmatrix}.$$

The matrix A has two eigenvalues, λ and λ', where $|\lambda| > |\lambda'|$ and $\lambda > 1$. Thus, the eigenvalue $\lambda = \frac{1}{2}(p + 1 + \sqrt{(p+1)^2 + 4p(2k-1)})$ dominates in the long-run behavior of the process. If we denote by f^1 the standardized eigenvector associated with the eigenvalue λ, we have

$$\text{E}[(f^1)^\top M(n) \mid \mathcal{F}_{n-1}] = (f^1)^\top AM(n-1) = \lambda(f^1)^\top M(n-1).$$

Thus, $(W(n))_{n \geq 0}$, where $W(n) = \lambda^{-n}(f^1)^\top M(n)$ defines a martingale whose limit $W_{k,p} := \lim_{n \to \infty} W(n)$ appears in the characterization of the typical interpoint distance.

If the processes started from i and j have not intersected at time n, the distance between nodes i and j is at least $2m$. In [4], the time of intersection is expressed in relation to the typical time at which the first intersection is likely to occur. This time, n_d, is defined by the relation $\lambda^{n_d} \leq \sqrt{Np} < \lambda^{n_d+1}$, implying that $\lambda^{n_d} = \phi\sqrt{Np}$ for $1/\lambda < \phi \leq 1$.

The main result in [4] is stated as follows.

Proposition 2. ([4, Theorem 5.9, p. 1271].) *There is a random variable V verifying, for any $x \in \mathbb{Z}$,*

$$P(V \geq 2n_d + x) = E\left[\exp\left\{-\frac{\lambda^2}{\lambda - \lambda'}(\lambda - p)\phi^2\lambda^x W_{k,p}\tilde{W}_{k,p}\right\}\right], \tag{4}$$

where λ, n_d, ϕ, and $W_{k,p}$ have been defined above and $\tilde{W}_{k,p}$ is an independent copy of $W_{k,p}$.

The total variation distance between the distributions of $T(\Gamma^N)$, the typical interpoint distance, and the random variable V verifies

$$D_{\mathrm{TV}}(\mathcal{L}(T(\Gamma^N)), \mathcal{L}(V)) = O(\ln(Np)(Np)^{\gamma/(4-\gamma)}), \tag{5}$$

where γ verifies $0 < \gamma \leq \frac{1}{2}$ for any fixed kp. Hence, $D_{\mathrm{TV}}(\mathcal{L}(T(\Gamma^N)), \mathcal{L}(V)) \to 0$ if $Np \to \infty$ and kp is fixed.

The right-hand side of (4) can be used to obtain a detailed characterization of the typical interpoint distance. Here we are interested in results that bound its upper tail. For our purposes, it was convenient to denote by $f(p, k)$ any bounded strictly positive quantity (not always the same one) that does not depend on N when recalling the following bound from [4].

Proposition 3. ([4, p. 1280].) *The random variable V verifies*

$$P\left(\frac{\lambda - 1}{2}(V - 2n_d) \geq z\right)$$

$$\leq f(k, p)e^{-f(k,p)z}\log(f(k, p) + f(k, p)e^{f(k,p)z}) \quad for\ z \in \frac{\lambda - 1}{2}\mathbb{Z} \tag{6}$$

whenever $pk < 1$.

From the above we derive $\lim_{N\to\infty} P(V \geq 2\ln N) = 0$. We have $\lambda = \frac{1}{2}(p + 1 + \sqrt{(p+1)^2 + 4p(2k - 1)})$. If we choose p such that $\frac{2}{3} \leq pk < 1$, we have $\lambda > \frac{7}{4}$, and, thus,

$$2n_d \leq 2\frac{\ln(\sqrt{Np})}{\ln \lambda} \leq \frac{\ln(Np)}{\ln(7/4)} \leq \alpha \ln N,$$

where $\alpha < 2$.

Define the sequence $(x_N)_{N\geq N_0}$ by

$$x_N = \left\{\max_{x \in ((\lambda-1)/2)\mathbb{Z}} \left| x \leq \frac{\lambda - 1}{2}(2\ln N - 2n_d)\right.\right\},$$

so that $\lim_{N\to\infty} X_N = \infty$. Then

$$\lim_{N\to\infty} P(V \geq 2\ln N)$$

$$\leq \lim_{N\to\infty} P\left(\frac{\lambda - 1}{2}(V - 2n_d) \geq x_N\right)$$

$$\leq \lim_{N\to\infty} f(k, p)\exp\{-f(k, p)x_N\}\log(f(k, p) + f(k, p)\exp\{f(k, p)x_N\})$$

$$= 0.$$

Remark 2. Although (3) assumes that $pk < 1$, it is obvious that the upper bound we obtained for the typical interpoint distance would be valid for $pk > 1$ also, since the probability of every link in Γ^N is increasing in p and k.

By (5) we conclude that

$$\lim_{N \to \infty} P(T(\Gamma^N) \geq 2 \ln N) \leq \lim_{N \to \infty} [P(V \geq 2 \ln N) + O(\ln(Np)(Np)^{\gamma/(4-\gamma)})] = 0.$$

5.1. Degree

Clearly, when the probability of random links is small, the degrees are close to that in the lattice. Label the elements of $\{1, \ldots, N\} \times \{1, \ldots, N\}$ in an arbitrary way, $i = 1, \ldots, N(N-1)/2$, and define $Y_i = 1$ if the link i exists in the Erdős–Rényi subgraph E and 0 otherwise: the Y_i are then independent and identically distributed (i.i.d.) Bernoulli variables and we can apply the following lemma.

Lemma 1. (Bernstein's inequality.) *Let $S_k = \sum_{i=1}^{m} X_i$, where $(X_i)_{i=1}^{m}$ is a sequence of i.i.d. variables such that $E[X_i] = 0$, $E[X_i^2] = \sigma^2$, and $|X_i| \leq L$. Then*

$$P(|S_m| \geq t\sigma\sqrt{m}) \leq 2 \exp\left\{-\frac{t^2}{1 + \alpha/3}\right\},$$

where $\alpha = Lt/\sigma\sqrt{m}$.

Define $X_i = Y_i - E[Y_i]$, and note that $\sigma = p/N$. With $m = N(N-1)/2$ and summing over N, we obtain, for any $\varepsilon > 0$,

$$\sum_N P\left(\frac{1}{N} \sum_{i=1}^{N(N-1)/2} Y_i \geq p + \varepsilon\right) < \infty; \tag{7}$$

so, by the Borel–Cantelli theorem,

$$\limsup_N \frac{1}{N} \sum_{i=1}^{N(N-1)/2} Y_i < p + \varepsilon \quad \text{P-almost surely.}$$

The average degree of a node is given by

$$2k + \frac{1}{N} \sum_{i=1}^{N(N-1)/2} Y_i,$$

where $2k$ is the degree in the underlying lattice. Since $p < 1$, we can choose $\varepsilon > 0$ so as to ensure that, almost surely, the average degree is bounded by $2k + 1$.

5.2. Clustering

When the number of random links added is almost surely bounded, we can infer a bound on the average clustering coefficient. We define $\Delta_i = \operatorname{card}\{(j, l) \in \Gamma \mid j \in V_L(i) \text{ and } l \in V_L(i)\}$, where $V_L(i)$ refers to i's neighbors in the lattice. (In a regular lattice, $\Delta_i = \Delta$ is the same for all i.) The average clustering coefficient in the lattice $\bar{\gamma}_L$ is then given by

$$\bar{\gamma}_L = \frac{\Delta}{2k(2k-1)/2}.$$

If $2k$ is the degree of each node in the lattice and X_i the degree of i in the Erdős–Rényi graph, then the total degree is $2k + X_i$ in the perturbed lattice, and the average clustering coefficient in the small-world graph $\bar{\gamma}$ verifies

$$\bar{\gamma} \geq \frac{1}{N} \sum_{i=1}^{N} \frac{\Delta}{(2k + X_i)(2k - 1 + X_i)/2}.$$

By (7), the average clustering coefficient is almost surely bounded below by the solution to

$$\min_{\{\sum_{i=1}^{N} X_i = N\}} \frac{1}{N} \sum_{i=1}^{N} \frac{\Delta}{(2k + X_i)(2k - 1 + X_i)/2}. \tag{8}$$

The problem above has a unique solution, $X_1 = X_2 = \cdots = X_N = 1$. Indeed, if we define

$$f(X_i) = \frac{\Delta}{(2k + X_i)(2k - 1 + X_i)/2}$$

then, for any $S > 0$, the unique solution to the problem

$$\min_{X_1 + X_2 = S} f(X_1) + f(X_2)$$

is $X_1 = X_2 = S/2$. Suppose that, for $N > 2$, $(x_i)_{i=1}^{N}$ is a solution to (8). Then we cannot have $x_l \neq x_m$ for some m and l, since terms in the sum we minimize could be made smaller by putting $X_l = X_m = (x_l + x_m)/2$. Thus, we have

$$\bar{\gamma} \geq \frac{\Delta}{(2k + 1)(2k - 1 + 1)/2},$$

which implies that

$$\bar{\gamma} \geq \frac{\Delta}{(2k + 1)(2k - 1 + 1)/2} = \frac{2k - 1}{2k + 1} \bar{\gamma}_L.$$

6. A lattice-free small-world model

The construction discussed in the previous section is based on random perturbations of a regular lattice. Although this construction satisfies the required scaling properties of small worlds, it is not a plausible model for the genesis of small-world phenomena in the context of social or biological networks, since regular lattices are not natural underlying structures in such contexts. We now propose a lattice-free small-world model, verifying Definition 4, which is more natural for such applications. The idea, which can be seen as a variation on Granovetter's [16] notion of *weak links* between social communities, is to start from an initial set of disjoint fully connected graphs, representing tightly-knit communities, and to introduce random links.

We partition the nodes $1, \ldots, N$ into M disjoint sets of nodes $\mathcal{G}_1, \ldots, \mathcal{G}_M$: these *clusters* can be seen as 'communities of origin' of the nodes. For ease of exposition, we will consider the case where $\mathcal{G}_i$ contains δ nodes with $N = M\delta$, but this can be easily relaxed. We then associate to every node $i = 1, \ldots, N$ a cluster X_i uniformly drawn among $\mathcal{G}_1, \ldots, \mathcal{G}_M$, and link i to all nodes in the cluster X_i. Here X_i can be viewed as the 'secondary community' of the node i. Finally, we link all pairs of nodes i and j which share a common secondary community, i.e. $X_i = X_j$.
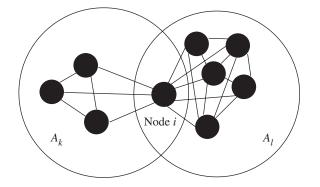
FIGURE 1: Local view. Node i is the only common member of communities A_k and A_l.
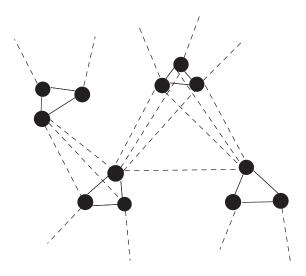

FIGURE 2: Solid lines link members of the same initial cluster. Dashed lines represent links created after the realization of the random variables.

The construction can be summarized as follows: define, for $m = 1, \ldots, M$, the set of nodes

$$A_m = \mathcal{G}_m \cup \{i \in \{1, \ldots, N\} \mid X_i = \mathcal{G}_m\};$$

Γ_N can then be characterized by

$$(i, j) \in \Gamma_N \iff i \neq j \text{ and } (i, j) \in \bigcup_{m=1}^{M} (A_m \times A_m).$$

The resulting graph contains fully connected subgraphs $A_m \times A_m$, linked together in a random manner. Typically, each node belongs to exactly two clusters. Figures 1 and 2 illustrate the structure of the resulting graph. In modeling terms we may think of the graph as a social

network where each individual belongs to two communities, such as, for example, a group of colleagues and a group of family members; thus, connecting two otherwise disjoint groups. We will see that these multiple community affiliations give rise to the small-world property. Communities are also natural in other contexts. In the context of the World Wide Web, these clusters would correspond to web domains or closely interlinked online communities. The presence of community structures in social networks is well documented in empirical studies [12] and algorithms have been proposed for detecting underlying community structures in networks [15], [17], [19], but such structures have rarely been integrated into network models.

We will now study the scaling properties of the model defined above.

Proposition 4. *The above construction with $\delta \geq 6$ and $d = 1$ satisfies Definition 4. More precisely,*

- *$6 \ln N$ is an almost-sure upper-scaling bound for the diameter of the graph;*

- *the expected average degree is bounded from above by $4\delta - 1$;*

- *the local clustering coefficient at each node is almost surely bounded from below by $\frac{1}{4}$.*

Remark 3. The choice $\delta \geq 6$ is for analytical convenience and the bound is not tight. It is likely that δ could be chosen smaller and/or that one could obtain a smaller scaling bound for the chosen value of δ. However, the cluster size δ must be greater than or equal to three to ensure almost surely that no cluster is isolated.

6.1. Scaling behavior of the average degree

Defining $S_m := \text{card}\{i \mid X_i = \mathcal{G}_m\}$, we have $\text{card}(A_m) \leq \delta + S_m$ and the total number of links in the graph is then bounded by

$$\sum_{m=1}^{M} \frac{\text{card}(A_m)(\text{card}(A_m) - 1)}{2} \leq \sum_{j=1}^{M} \frac{(S_j + \delta)(S_j + \delta - 1)}{2}.$$

Since the N variables $(X_i)_{i=1}^{N}$ are independently and uniformly drawn among $\mathcal{G}_1, \ldots, \mathcal{G}_M$, the variables $(S_j)_{j=1,\ldots,M}$ follow a multinomial law

$$P(S_1 = s_1, \ldots, S_M = s_M) = \begin{cases} \dfrac{N!}{s_1 \cdots s_M} \left(\dfrac{1}{M}\right)^N & \text{if } \sum_{i=1}^{M} s_i = N, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $E[S_j] = N/M = \delta$ and $\text{var}(S_j) = (N/M)(1 - 1/M) = \delta(1 - 1/M)$, so that

$$E[S_j^2] = \text{var}(S_j) + (E[S_j])^2 = \delta\left(1 - \frac{1}{M}\right) + \delta^2.$$

We have

$$E\left[\sum_{j=1}^{M} \frac{(S_j + \delta)(S_j + \delta - 1)}{2}\right] = \frac{1}{2} \sum_{j=1}^{M} (E[S_j^2] + (2\delta - 1)E[S_j]) + \delta(\delta - 1)$$

$$\leq \frac{M(4\delta^2 - \delta)}{2}.$$

The average degree in a graph of N nodes equals the total number of links divided by $N/2$. Thus, $(4\delta - 1)$ is an upper-scaling bound for the expected average degree.

6.2. Clustering behavior

We need to prove that the scaling of the clustering coefficient is bounded below by a constant. We first prove a lemma on local clustering in graphs that we then apply to our construction.

Lemma 2. *Consider a graph Γ and the subgraph $\Gamma(i) = \Gamma \cap [V(i) \times V(i)]$ formed by the neighbors $V(i)$ of i and the links between them. If $\Gamma(i)$ can be partitioned into K disjoint complete subgraphs then the local clustering coefficient γ_i verifies $\gamma_i \geq 1/K - 1/h$, where h is the number of nodes of the smallest subgraph.*

Proof. Consider such a partition of $\Gamma(i)$, and let $C_1, C_2, \ldots, C_l$ be the number of nodes in each complete subgraph. Then there are at least $\sum_{j=1}^{l} C_j(C_j - 1)/2$ links between i's neighbors. The maximal number of possible links between i's neighbors is $\sum_{j=1}^{l} C_j (\sum_{j=1}^{l} C_j - 1)/2$. Thus, we have

$$\gamma_i \geq \frac{\sum_{j=1}^{l} C_j(C_j - 1)}{(\sum_{j=1}^{l} C_j)^2} = \frac{\sum_{j=1}^{l} C_j^2}{(\sum_{j=1}^{l} C_j)^2} - \frac{\sum_{j=1}^{l} C_j}{(\sum_{j=1}^{l} C_j)^2}.$$

Using $2C_iC_j \leq C_i^2 + C_j^2$, we obtain

$$\left(\sum_{j=1}^{l} C_j \right)^2 = \sum_{j=1}^{l} C_j^2 + \sum_{j \neq k} C_j C_k \leq \sum_{j=1}^{l} C_j^2 + (l-1) \sum_{j=1}^{l} C_j^2 = l \sum_{j=1}^{l} C_j^2.$$

Noting that

$$\frac{\sum_{j=1}^{l} C_j}{(\sum_{j=1}^{l} C_j)^2} \leq \max_j \frac{1}{C_j},$$

the result follows.

Now we apply the Lemma 2 to our model. Since all nodes in the same A_m, $m = 1, \ldots, M$, are connected, any subset of A_m is a complete subgraph. Each i belongs to at most two such sets, A_j and A_k. These sets may have common elements, but we can always obtain a disjoint union $V(i) = (A_j - \{i\} - [\mathcal{G}_k]) \cup (A_k - \{i\} - [\mathcal{G}_j])$. The sets in the union contain at least card($[\mathcal{G}_j]$) $- 1$ and card($[\mathcal{G}_k]$) $- 1$ elements, respectively. Thus, $h \geq \delta - 1 \geq 5$. Applying the Lemma 2, we obtain as the lower bound for the clustering coefficient $\frac{1}{2} - 1/(\delta - 1) \geq \frac{1}{4}$, as stated in Proposition 4.

6.3. Scaling behavior of the diameter

In this section we will show that

$$\mathrm{P}\left(\limsup_N \frac{D(G_N)}{6 \ln N + 5} \leq 1 \right) = 0.$$

Our proof consists in showing that, for all $K \geq K_0$,

$$\mathrm{P}\left(\frac{D(G_K)}{6 \ln K + 5} \geq 1 \right) \leq \frac{1}{K^{1+\alpha}}, \qquad \alpha > 0.$$

Using this bound, we have, from the Borel–Cantelli theorem,

$$\mathrm{P}\left(\limsup_{N} \frac{D(G_N)}{6\ln N} \ge 1\right) = \mathrm{P}\left(\limsup_{N} \frac{D(G_N)}{6\ln N + 5} \ge 1\right)$$

$$\le \lim_{N\to\infty} \sum_{K\ge N} \mathrm{P}\left(\frac{D(G_k)}{6\ln K + 5} \ge 1\right)$$

$$\le \lim_{N\to\infty} \sum_{K\ge N} \frac{1}{K^{1+\alpha}}$$

$$= 0.$$

In order to show that

$$\mathrm{P}\left(\frac{D(G_N)}{6\ln N + 5} \ge 1\right) \le \frac{1}{N^{1+\alpha}}$$

for sufficiently large N, we will proceed in several steps. We start from an arbitrary subgraph G. First we will estimate the probability that there is a set S within distance $\ln N$ from G that contains at least $M/2$ clusters or, equivalently, $N/2$ nodes. Then we estimate the probability that all nodes in S^c are close to S.

We construct the set S by considering the clusters that we can reach from a cluster G_1 in a given number of steps. In what follows, we will assume that the number δ of nodes in each cluster verifies $\delta \ge 6$. We start from the set $E_0 = \{G_1\}$ and then define E_k and F_k recursively in the following way:

$$F_k = \{i \in [1,\dots,N] \mid \text{there exists } G \in E_k,\ i \in G\},$$

$$E_k = \left\{ G \notin \bigcup_{l=1}^{k-1} E_l \ \middle|\ \text{there exists } j \in F_{k-1},\ X_j = G \right\}.$$

We then let

$$\bar{F}_k = \mathrm{card}\{F^k\} = \mathrm{card}\{i \in [1,\dots,N] \mid \text{there exists } G \in E_k,\ i \in G\}.$$

Thus, E_k is the set of clusters that can be reached after exactly k steps from G_1, and $\bar{F}_k$ is the number of nodes that belongs to the clusters in E_k. Define the events

$$\mathcal{A}^k = \{\mathrm{card}(E_k) \ge \mathrm{e}\,\mathrm{card}(E_{k-1})\}.$$

We note that $\mathcal{A}^k$ occurs if $\sum_{j \in F_{k-1}} \mathbf{1}_{A_j} \ge \mathrm{e}\,\mathrm{card}(E_{k-1})$. Here A_j denotes the event $[X_j \notin \bigcup_{m<k-1} E_m$ and $X_j \neq X_a$ for $a \in F_{k-1}$ such that $a \le j]$. The event $\mathcal{A}^k$ implies that not too many of the successors of the nodes in F_{k-1} belong to the sets that have been found previously.

Define, for $u \ge 0$, the stopping times

$$\tau_u = \min\left\{ l \ge 1, \mathrm{card}\left(\bigcup_{j=0}^{l} E_j \right) \ge u \right\}.$$

For a given T, we want to estimate the probability of the event

$$\mathrm{P}\left(\bigcap_{j=1}^{T} \mathcal{A}^j\right) = \prod_{j=1}^{T} \mathrm{P}\left(\mathcal{A}^j \ \middle|\ \bigcap_{m=1}^{j-1} \mathcal{A}^m\right).$$

We will now establish the lower bound $P(\mathcal{A}^j \mid \bigcap_{m=1}^{j-1} \mathcal{A}^m) \geq 1 - c/M^3$. We obtain this bound in a different manner depending on whether $j = 1$, $2 \leq j \leq \tau_{\sqrt{M}}$, or $\tau_{\sqrt{M}} \leq j < \tau_{M/2}$.

Consider $j = 1$. We are on $\mathcal{A}^1$ if E_1 contains at least $3 > e$ different elements that are not in E^0. The probability of drawing an element among the previous sets is bounded by C/M; thus, whenever at least three of the $\delta \geq 6$ elements in N^0 have new successors, $\mathcal{A}^1$ is verified. The probability that at least three nodes have successors found previously is smaller than C/M^3.

Now we consider $2 \leq j \leq \tau_{\sqrt{M}}$. In fact, the probability of the event $\mathcal{A}^j$ only depends on $\bar{F}_{j-1}$ and $\sum_{m=0}^{j-1} \text{card}(E_m)$. While $j \leq \tau_{\sqrt{M}}$, the latter set necessarily contains less than $\sqrt{M}$ elements. Thus, the probability of drawing a set that is found previously is smaller than $1/\sqrt{M}$. Conditionally on the events $\bigcap_{m<j} \mathcal{A}^m$, we also know that $\bar{F}_{j-1} \geq 3^{j-1} \bar{F}_0$. Thus, for every $2 \leq j \leq \tau_{\sqrt{M}}$, we have $18 \leq \bar{F}_1 \leq \bar{F}_{j-1} \leq \sqrt{M}$. Since $\bar{F}_j = \delta \, \text{card}(E_j)$ with $\delta \geq 6$, we see that we are on $\mathcal{A}^j$ if at most $\frac{1}{2} \bar{F}_{j-1}$ of the variables $(X_l)_{l \in F^{j-1}}$ are among the previously found sets. Since the probability of intersecting previous sets is bounded by $1/\sqrt{M}$, we have

$$P\left(\mathcal{A}^k \,\Big|\, \bigcap_{j=1}^{k-1} \mathcal{A}^j\right) \geq 1 - \sum_{m=\bar{F}_{k-1}/2}^{\bar{F}_{k-1}} \binom{\bar{F}_{k_1}}{m} \left(\frac{1}{\sqrt{M}}\right)^m.$$

The binomial coefficient

$$\binom{\bar{F}_{k_1}}{m} = \frac{(\bar{F}_{k-1} - m + 1) \cdots (\bar{F}_{k-1})}{m!}.$$

The numerator contains m terms, each smaller than $\bar{F}_{k-1}$. Since $k - 1 \leq \tau_{\sqrt{M}}$, $\bar{F}_{k-1} \leq \sqrt{M}$.

We distinguish between two cases.

Case 1: $2 \ln M < \bar{F}_{k-1} \leq \sqrt{M}$. Then we have

$$
\begin{aligned}
\sum_{m=\bar{F}_{k-1}/2}^{\bar{F}_{k-1}} \binom{\bar{F}_{k_1}}{m} \left(\frac{1}{\sqrt{M}}\right)^m &\leq \sum_{m=\bar{F}_{k-1}/2}^{\bar{F}_{k-1}} \frac{(\bar{F}_{k-1})^m}{m!} \left(\frac{1}{\sqrt{M}}\right)^m \\
&\leq \sum_{m=\bar{F}_{k-1}/2}^{\bar{F}_{k-1}} \frac{(\sqrt{M})^m}{m!} \left(\frac{1}{\sqrt{M}}\right)^m \\
&\leq \sum_{m=\bar{F}_{k-1}/2}^{\bar{F}_{k-1}} \frac{1}{(\ln M)!} \\
&\leq \bar{F}_{k-1} \frac{1}{(\ln M)!} \\
&\leq \frac{\sqrt{M}}{(\ln M)!} \\
&\leq \frac{1}{M^3}.
\end{aligned}
\tag{9}
$$

Case 2: *conditionally on* $\mathcal{A}^1$, $18 \leq \bar{F}_1 \leq \bar{F}_{k-1} \leq 2 \ln M$. We have the bound

$$
\sum_{m=\bar{F}_{k-1}/2}^{\bar{F}_{k-1}} \binom{\bar{F}_{k_1}}{m} \left(\frac{1}{\sqrt{M}} \right)^m \leq \sum_{m=\bar{F}_{k-1}/2}^{\bar{F}_{k-1}} \frac{(\bar{F}_{k-1})^m}{m!} \left(\frac{1}{\sqrt{M}} \right)^m
$$

$$
\leq \sum_{m=\bar{F}_{k-1}/2}^{\bar{F}_{k-1}} (2 \ln M)^m \left(\frac{1}{\sqrt{M}} \right)^m \tag{10}
$$

$$
\leq \sum_{m=\bar{F}_{k-1}/2}^{\bar{F}_{k-1}} \left(\frac{2 \ln M}{\sqrt{M}} \right)^9
$$

$$
\leq \bar{F}_{k-1} \left(\frac{2 \ln M}{\sqrt{M}} \right)^9
$$

$$
\leq (2 \ln M)^{10} \frac{1}{M^4}
$$

$$
\leq \frac{1}{M^3}. \tag{11}
$$

We used the fact that there is an M_0 such that, for $M \geq M_0$, the term $(2 \ln M / \sqrt{M})^m$ is decreasing in m.

Now we establish a lower bound for $\mathrm{P}(\mathcal{A}^j \mid \bigcap_{m=1}^{j-1} \mathcal{A}^m)$, when $\tau_{\sqrt{M}} < j \leq \tau_{M/2}$. For such a j, conditionally on $\bigcap_{m=1}^{j-1} \mathcal{A}^m$, there is a $0 \leq \beta < 1$ such that $\bar{F}_j \geq M^\beta$. Indeed, conditionally on $\bigcap_{m=1}^{j} \mathcal{A}^m$, we have $\bar{F}_j \geq \mathrm{e}^j \bar{F}_0 \geq \exp\{\tau_{\sqrt{M}}\} \bar{F}_0$. Since, for every $T \geq 0$, $\sum_{l \leq T} \mathrm{card}(E_l) \leq \sum_{l=0}^{l=T} \delta^l \leq \delta^{T+1}$, we have

$$
\tau_{\sqrt{M}} = \min\left\{ T \geq 1 \;\middle|\; \sum_{j=0}^{T} \mathrm{card}(E_j) \geq \sqrt{M} \right\} \geq \min\{ T \geq 1 \mid \delta^{T+1} \geq \sqrt{M} \}. \tag{12}
$$

It follows that there must be a $\beta > 0$ such that $\tau_{\sqrt{M}} \geq \ln(M^\beta)$, which implies that $\bar{F}_j \geq \exp\{\tau_{\sqrt{M}}\} \bar{F}_0 \geq M^\beta \bar{F}_0 \geq M^\beta$.

For $\tau_{\sqrt{M}} < k \leq \tau_{M/2}$, we obtain a lower bound for the probability of $(\mathcal{A}^k \mid \bigcap_{j<k} \mathcal{A}^j)$ using Bernstein's inequality, (1). Conditionally on $\bigcap_{j=1}^{k-1} \mathcal{A}^j$, $\mathcal{A}^k$ occurs if at least $\mathrm{e} \, \mathrm{card}(E_{k-1})$ out of the $\bar{F}_{k-1} = \delta \, \mathrm{card}(E_{k-1}) \geq 6 \, \mathrm{card}(E_{k-1})$ independent variables reach new successors. By the definition of $\tau_{M/2}$, for $k \leq \tau_{M/2}$, the probability that the uniformly distributed variables $(X_i)_{i \in F_{k-1}}$ reach one of the previously attained elements is smaller than $\frac{1}{2}$, and, thus, the probability of finding a new successor is greater than $\mathrm{P}(Y_i = 1)$, where Y_i follows a $\mathrm{Ber}(\frac{1}{2})$. We fix an $\varepsilon \geq 0$ such that $3 - \varepsilon \geq \mathrm{e}$. If we let (Y_i) denote such a sequence of i.i.d. $\mathrm{Ber}(\frac{1}{2})$, we have

$$
\mathrm{P}\left(\mathcal{A}^k \;\middle|\; \bigcap_{j<k} \mathcal{A}^j \right) \geq 1 - \mathrm{P}\left(\left| \sum_{i=1}^{\bar{F}_{k-1}} \left(Y_i - \frac{1}{2} \right) \right| \geq \varepsilon \bar{F}_{k-1} \right).
$$

We will bound the right-hand side by again using the Bernstein inequality, (1).

For a sequence of i.i.d. $\mathrm{Ber}(\frac{1}{2})$, we apply (1) to $(X_i - \frac{1}{2})_{i=1}^{N}$ with $\sigma = \frac{1}{2}$, $L = 1$, and $t = \sqrt{\varepsilon}\sqrt{N}/\sigma$, which gives

$$
\alpha = \frac{\sqrt{\varepsilon}}{\sigma} \leq 3 \quad \text{and} \quad \frac{t^2}{1 + (\alpha)/3} \geq \frac{\varepsilon N}{2\sigma^2} \geq \varepsilon N.
$$

We have

$$P\left(\left|\sum_{i=1}^{\bar{F}_{k-1}}\left(X_i - \frac{1}{2}\right)\right| \geq \varepsilon \bar{F}_{k-1}\right) \leq 2\exp\{-\varepsilon \bar{F}_{k-1}\}.$$

As we saw before, conditionally on $\bigcap_{j\leq k-1}\mathcal{A}^j$, we have $\bar{F}^{k-1} \geq M^\beta$, with $\beta > 0$.

Thus, by the Bernstein inequality, we obtain, for $\tau_{\sqrt{M}} \leq k \leq \tau_{M/2}$,

$$P\left(\mathcal{A}^k \,\middle|\, \bigcap_{j<k}\mathcal{A}^j\right) \geq 1 - 2\exp\{-\varepsilon M^\beta\} \geq 1 - \frac{1}{M^3},$$

where the last inequality holds for $M \geq M_0$.

Let $L := \min[l \mid e^l \geq M/2] < \ln M$. If we are on $\mathcal{A}^j$ until L, $\mathrm{card}(E^L) \geq e^L\,\mathrm{card}(E_0) \geq M/2$. Thus, if the events $\mathcal{A}^j$ hold until $\tau_{M/2}$, necessarily $\tau_{M/2} \leq L$. We define the bounded stopping time $T := \min(L, \tau_{M/2})$. We note that we have

$$\left(\bigcap_{l=1}^{T}\mathcal{A}^l\right) \subset (\tau_{M/2} \leq L).$$

Since we have shown that, for $1 \leq j \leq \tau_{M/2}$, $P(\mathcal{A}^j \mid \bigcap_{m=1}^{j-1}\mathcal{A}^m) \geq 1 - 1/M^3$, and since the stopping time T is bounded by $\ln M$, we have

$$\begin{aligned}
P(\tau_{M/2} \leq L) &\geq P\left(\bigcap_{l=1}^{T}\mathcal{A}^l\right) \\
&= \prod_{j=1}^{T}P\left(\mathcal{A}^j \,\middle|\, \bigcap_{l=1}^{j-1}\mathcal{A}^l\right) \\
&\leq \left(1 - \frac{1}{M^3}\right)^T \\
&\leq \left(1 - \frac{1}{M^3}\right)^{\ln M}.
\end{aligned}$$

To conclude, we use the following limit results.

Lemma 3. *Let $f(N)$ be a function verifying $\lim_{N\to\infty} f(N) = \infty$ and let $g(N)$ be a function verifying $\lim_{N\to\infty} g(N) = \infty$ such that $h(N) := g(N)/f(N)$ verifies $\lim_{N\to\infty} h(N) = 0$. Then there is an N_0 such that we have, for all $N \geq N_0$,*

$$\left(1 - \frac{1}{f(N)}\right)^{g(N)} \geq 1 - \frac{g(N)}{f(N)}.$$

Proof. The assumptions imply that we have

$$\lim_{N\to\infty}\left(1 - \frac{1}{f(N)}\right)^{f(N)} = e^{-1}$$

and, for any $\Phi(n)$ with $\lim_{n\to\infty}\Phi(n) = 0$, we have

$$\lim_{N\to\infty}\frac{e^{\Phi(N)} - 1}{\Phi(N)} = 1. \tag{13}$$

Fix $\varepsilon \geq 0$. There is an N_0 such that, for $N \geq N_0$, we have $(1 - 1/f(N))^{f(N)} \geq e^{-(1+\varepsilon)}$, and $e^{\Phi(N)} - 1 \geq (1 - \varepsilon)\Phi(N)$ is equivalent to $e^{\Phi(N)} \geq 1 + (1 - \varepsilon)\Phi(N)$. Applying (13) with $\Phi = -(1 + \varepsilon)h(N)$ we obtain

$$\left(1 - \frac{1}{f(N)}\right)^{g(N)} \geq e^{-(1+\varepsilon)h(N)} \geq 1 - (1 - \varepsilon)(1 + \varepsilon)h(N) \geq 1 - \frac{g(N)}{f(N)}.$$

This completes the proof.

Applying Lemma 3 to $f(M) = M^3$ and $g(M) = \ln M$, we have

$$P(\tau_{M/2} \geq \ln M) \leq 1 - \left(1 - \frac{1}{M^3}\right)^{\ln M} \leq \frac{\ln M}{M^3}. \qquad (14)$$

We define the set $S := \bigcup_{j=0}^{\tau_{M/2}} E_j$.

We note that, when $\tau_{M/2} \leq \ln M$, for all $\mathcal{G}, \mathcal{G}' \in S$,

$$d_\Gamma(\mathcal{G}, \mathcal{G}') \leq d_\Gamma(\mathcal{G}, S) + d_\Gamma(\mathcal{G}', S) \leq 2\ln M.$$

Thus,

$$P(d_\Gamma(\mathcal{G}, \mathcal{G}') \leq 2\ln M) \geq P(\tau_{M/2} \leq \ln M) \geq 1 - \frac{\ln M}{M^3}.$$

Also, $\text{card}(S) \geq M/2$, by definition of $\tau_{M/2}$.

At this point, it is very easy to show that, almost surely, the typical interpoint distance scales logarithmically. With some additional work, we show the same to be true for the diameter.

We use a construction that divides the clusters in S^c into disjoint connected sets containing at most $\ln M$ elements and satisfying one of the following properties.

(H1) $[I \subset S^c \mid$ there exists $\mathcal{G}_m \in I$ such that $d(\mathcal{G}_m, S) = 1$ or $\ln M - 1 \leq \text{card}(I)]$.

(H2) $[I \subset S^c \mid$ there exist $\mathcal{G} \in I$ and I' satisfying property (H1) such that $d(\mathcal{G}, I') = 1]$.

These properties will be used later to show that the constructed sets are likely to be close to S.

We construct disjoint sets $(I_j)_{j=1}^{k_{\max}}$ in the following way. We define $I^0 = \varnothing$. For $k \geq 1$, we then take an arbitrary element $\mathcal{G} \in S^c - \bigcup_{l \leq k} I^l$ and define $\mathcal{G} =: E_0^k$. Then we define the following sets recursively

$$F_l^k = \{i \in [1, \ldots, N] \mid \text{ there exists } \mathcal{G} \in E_l^k, i \in \mathcal{G}\},$$

and then denote by $\tilde{F}_l^k$ the $\text{card}(F_l^k)/2$ nodes of smallest index in F_l^k and define

$$E_l^k = \left\{ \mathcal{G} \notin \bigcup_{j=1}^{l-1} E_j^k \;\middle|\; \text{there exists } a \in \tilde{F}_{l-1}^k, X_a = \mathcal{G} \right\}.$$

This is essentially similar to what was done previously, except that now we only use the successors of half of the nodes that we find at each step. For the process $(E_j^k)_j$, we define the

following stopping times:

$$\tau_{\ln M}^k = \inf\left\{ t \;\middle|\; \sum_{j=0}^t \operatorname{card}(E_j^k) \geq \ln M \right\},$$

$$\tau_S^k = \inf\{ t \mid E_t^k \cap S \neq \varnothing \},$$

$$\tau_{H_1}^k = \inf\{ t \mid E_t^k \cap I_s \neq \varnothing, \; s \leq t, \; I_s \text{ verifies property (H1)} \},$$

$$\tau_{H_2}^k = \min\{ t \mid E_t \cap I_s \neq \varnothing, \; s \leq t, \; I_s \text{ verifies property (H2)} \},$$

$$\tau^k = \min\{ \tau_{\ln M}^k, \tau_S^k, \tau_{H_1}^k, \tau_{H_2}^k \}.$$

We will show that τ^k is almost surely bounded. We construct (E_j^k) and (F_j^k) for $j \leq \tau^k$. Then we define I^k in the following way: if $\tau^k = \tau_S^k$, we define $I^k := \bigcup_{j=0}^{\tau^k} E_j^k$; if $\tau^k = \tau_{\ln M}^k$, we define $I^k := \bigcup_{j=0}^{\tau^k} E_j^k - A$, where $A \subset E^{\tau_{\ln M}}$ is chosen so that

$$\ln M - 1 \leq \operatorname{card}\left(\bigcup_{j=0}^{\tau^k} E_j^k - A \right) \leq \ln M.$$

The I^k defined this way verifies property (H1).

If $\tau^k = \tau_{H_1}^k$, we also define $I^k := \bigcup_{j=0}^{\tau^k} E_j^k$ and I^k verifies property (H2). If $\tau^k = \tau_{H_2}^k$, there is an $s < k$ such that $I^s \cap E_{\tau^k} \neq \varnothing$. In this case we consider the set

$$I^s \cup \left(\bigcup_{j \leq \tau^k} E_j^k \right).$$

If it has less than $\ln M$ elements, we call it I^k and it satisfies property (H2). If it has more than $\ln M$ elements, we split it into a set satisfying property (H1) and one or several sets satisfying property (H2) in the following way. The set $I^s \cup (\bigcup_{j \leq \tau^k} E_j^k)$ is connected by construction. Take an arbitrary cluster $\mathcal{G}$ in the set. Consider the nodes at a maximal distance d from $\mathcal{G}$ and remove them one by one as long as more than $\ln M$ nodes remain. Then do the same at distance $d - 1$, and so on. Nodes are removed one by one until the remaining connected set contains no more than $\ln M$ clusters and, thus, verifies property (H1). We call this set I_1^k. We partition the removed nodes into connected components, $(I_l^k)_{l=2,\dots,L}$. (Some components may consist of one element.) Since the division into connected components is a partition of the set of removed nodes, each element is in exactly one component. Each of these connected components contains an element linked to I_1^k and, thus, verifies property (H2). Indeed, a removed node at distance d from $\mathcal{G}$ is linked to a node at distance $d - 1$ from $\mathcal{G}$. Either this node is in I_1^k or it is linked to a node at distance $d - 2$ from $\mathcal{G}$. Since there is a C such that all elements at a distance less than or equal to C from $\mathcal{G}$ belong to I_1^k, each removed node belongs to a chain and, thus, to a connected component including an element at distance 1 from I_1^k. When I^k is constructed, we repeat the procedure for $k + 1$. We continue this way until a $k_{\max}$ such that

$$S^c - \bigcup_{l \leq k_{\max}} I_l = \varnothing.$$

Necessarily, $k_{\max} \leq \operatorname{card}(S^c) \leq M$. We have constructed disjoint sets $(I^k)_{k=1}^{k_{\max}}$ such that every element in S^c belongs to exactly one set I^k.

Now we will show that τ^k is almost surely bounded by $\ln M$ for $k = 1, 2, \dots, k_{\max}$. In this section we say that the event $\mathcal{A}^k$ occurs if $\operatorname{card}(E_k) \geq 1$. This guarantees that at least one new

cluster is reached at each step. We define the bounded stopping times $T^k := \min\{\ln M, \tau^k\}$. Then $(\bigcap_{j=1}^{T^k} Q^j) \subset (\tau^k \leq \ln M)$. The event $\mathcal{A}^j$ occurs when at least one new cluster is found at step j. Before T^k, fewer than $\ln M$ groups have already been reached. Since we use half of the successors of the nodes in a cluster, that is, $\delta/2 \geq 3$ successors,

$$P\left(\mathcal{A}^j \mid \bigcap_{l=1}^{j-1} \mathcal{A}^l\right) \geq \left(1 - \left(\frac{\ln M}{M}\right)^3\right).$$

We have

$$\begin{aligned}
P(\tau^k \leq \ln M) &\geq P\left(\bigcap_{l=1}^{T^k} \mathcal{A}^l\right) \\
&= \prod_{j=1}^{T^k} P\left(\mathcal{A}^j \mid \bigcap_{l=1}^{j-1} \mathcal{A}^l\right) \\
&\geq \left(1 - \left(\frac{\ln M}{M}\right)^3\right)^{T^k} \\
&\geq \left(1 - \left(\frac{(\ln M)}{M}\right)^3\right)^{\ln M}.
\end{aligned} \tag{15}$$

Moreover, the constructions at steps $k = 1, 2, \ldots, k_{\max}$ are independent. They involve different nodes since constructions stop when intersecting previous constructions. Thus, we have

$$P\left(\bigcap_{k=1}^{k_{\max}} \tau^k \leq \ln M\right) = \prod_{k=1}^{k_{\max}} P(\tau^k \leq \ln M) \geq \left(1 - \left(\frac{\ln M}{M}\right)^3\right)^{M \ln M}.$$

We can then apply (3) to obtain

$$\begin{aligned}
P\left(\bigcap_{k=1}^{k_{\max}} \tau^k \leq \ln M\right) &\geq \left(1 - \left(\frac{\ln M}{M}\right)^3\right)^{M \ln M} \\
&> 1 - M \ln M \frac{(\ln M)^3}{M^3} \\
&> 1 - \frac{1}{M^{1+\varepsilon}} \quad \text{for } \varepsilon > 0.
\end{aligned} \tag{16}$$

Now we will give a lower bound for the probability that $d(I^j, S) = 1$ when I^j is a set verifying property (H1). If I^j verifies property (H1), it either contains an element linked to S or contains $\ln M$ subgraphs. In the second case, since we only used half of the successors of the nodes in I^j to construct the set, there are still $(\delta/2) \ln M \geq 3 \ln M$ nodes whose successors are unexplored. We condition on the event $\{\tau_{M/2} \leq \ln M\}$, so that the set S contains at least $M/2$ subgraphs and on $\{\bigcap_{k=1}^{k_{\max}} \tau^k \leq \ln M\}$, which guarantees that each set $(I^k)_{k=1}^{k_{\max}}$ verifies property (H1) or property (H2). The probability that a successor is in S is at least $\frac{1}{2}$. Since there are at most $k_{\max} \leq M$ sets verifying property (H1), and these are disjoint, the conditional probability that

they are all at distance 1 from S is bounded by

$$\mathrm{P}\left(\bigcap_{\{j \mid I^j \text{ verifies property (H1)}\}} (d(I^j, S) = 1) \,\middle|\, \bigcap_{k=1}^{k_{\max}} \tau^k \le \ln M, \ \tau_{M/2} \le \ln M\right)$$

$$= \prod_{\{j \mid I^j \text{ verifies property (H1)}\}} \mathrm{P}\left(d(I^j, S) = 1 \,\middle|\, \bigcap_{k=1}^{k_{\max}} \tau^k \le \ln M, \ \tau_{M/2} \le \ln M\right)$$

$$\ge \left(1 - \left(\tfrac{1}{2}\right)^{3\ln M}\right)^M$$

$$\ge \left(1 - \frac{1}{M^{2+\alpha}}\right)^M$$

$$\ge 1 - \frac{1}{M^{1+\alpha}}, \tag{17}$$

where $\alpha > 0$. The last inequality follows by applying (3). If all sets verifying property (H1) contain an element at distance 1 from S then no element in a set verifying property (H1) is further than $\ln M + 1$ from S, since sets verifying property (H1) are connected and contain at most $\ln M$ elements. It follows that all sets verifying property (H2) must be at a distance of at most $2 \ln M + 2$ from S, since they contain an element at distance 1 from a set verifying property (H1), are connected, and have no more than $\ln M$ elements. Thus, the previous estimate implies that

$$\mathrm{P}\left(\max\{d_\Gamma(x, S), \ x \in S^c\} \le \ln M + 2 \,\middle|\, \bigcap_{k=1}^{k_{\max}} \tau^k \le \ln M, \ \tau_{M/2} \le \ln M\right)$$

$$\ge 1 - \frac{1}{M^{1+\alpha}}. \tag{18}$$

To conclude, we consider any two clusters $\mathcal{G}$ and $\mathcal{G}'$. When $\max\{d_\Gamma(x, S), \ x \in S^c\} \le \ln M + 2$, there are $\tilde{\mathcal{G}} \in S$ and $\hat{\mathcal{G}} \in S$ such that $d(\mathcal{G}, \hat{\mathcal{G}}) \le 2\ln M + 2$ and $d(\mathcal{G}', \tilde{\mathcal{G}}) \le 2\ln M + 2$. Also, we saw previously that if $\tilde{\mathcal{G}}$ and $\hat{\mathcal{G}}$ are in S, and $\tau_{M/2} \le \ln M$, then $d(\tilde{\mathcal{G}}, \hat{\mathcal{G}}) \le 2\ln M$. Thus, the probability that

$$d(\mathcal{G}, \mathcal{G}') \le d(\mathcal{G}, \hat{\mathcal{G}}) + d(\mathcal{G}', \tilde{C}) + d(\tilde{\mathcal{G}}, \hat{\mathcal{G}}) \le 6\ln M + 4$$

is bounded by the product of the estimations (14), (16), and (18) is

$$\mathrm{P}(d_\Gamma(\mathcal{G}, \mathcal{G}') \le 6\ln M + 4)$$

$$\ge \mathrm{P}\left(\max\{d_\Gamma(x, S), \ x \in S^c\} \le \ln M + 2 \,\middle|\, \bigcap_{k=1}^{k_{\max}} \tau^k \le \ln M, \ \tau_{M/2} \le \ln M\right)$$

$$\times \mathrm{P}\left(\bigcap_{k=1}^{k_{\max}} \tau^k \le \ln M\right) \mathrm{P}(\tau_{M/2} \le \ln M)$$

$$\ge \left(1 - \frac{1}{M^{1+\varepsilon}}\right)\left(1 - \frac{1}{M^{1+\tilde{\varepsilon}}}\right)\left(1 - \frac{\ln M}{M^3}\right)$$

$$\ge 1 - \frac{C}{N^{1+\varepsilon'}}.$$

Now, for any two nodes x and y, if $x \in \mathcal{G}$ and $y \in \mathcal{G}'$, we have $d_\Gamma(x, y) \leq d(\mathcal{G}, \mathcal{G}') + 1$, since all the clusters are complete subgraphs. Thus, we have shown that

$$\mathrm{P}\left(\frac{D(G_N)}{6 \ln N + 5} \geq 1 \right) \leq \frac{1}{N^{1+\varepsilon'}}.$$

By the arguments in the beginning of this section, it follows that

$$\mathrm{P}\left(\limsup_N \frac{D(G_N)}{6 \ln N} \geq 1 \right) = 0,$$

or, equivalently, that $6 \ln N$ is almost surely an upper-scaling bound of $D(G_N)$.

7. A randomized community-based small world

In this section we propose another random graph model with small-world properties, which can be seen as a randomized version of the preceding construction. We start from a collection of complete graphs representing *communities*, but allow each node to link to a random number of secondary communities. Allowing the number of communities to be random is both natural from a modeling perspective [11] and simplifies the study of scaling properties, as detailed below.

We will show that this construction verifies Definition 4, with a logarithmic scaling of the expected average degree, whereas the expected average degree was bounded independently of N in the previous construction.

We consider a set of N nodes partitioned into M 'communities' (clusters) $\mathcal{G}_1, \ldots, \mathcal{G}_M$ of approximate size $\ln N$. To simplify the exposition, we will assume that

$$\ln N - 1 \leq \mathrm{card}(\mathcal{G}_m) \leq \ln N + 1 \quad \text{for } m = 1, \ldots, M.$$

The graph Γ_N is constructed as follows. First, we link all nodes in each of the clusters $\mathcal{G}_1, \ldots, \mathcal{G}_M$. We then introduce links among communities by representing these clusters as nodes of an auxiliary Erdős–Rényi (hyper)graph E with M nodes and link probability $p = r \ln M / M$, with $r > 1$. If two clusters $\mathcal{G}_m$ and $\mathcal{G}_n$ are linked in E, we randomly pick a node in either $\mathcal{G}_m$ or $\mathcal{G}_n$ and link it to all points in the other cluster. Finally, we link all nodes i, j which belong to a common cluster. The graph Γ_N thus obtained differs from the one described in Section 6 in that the number of clusters to which a node is connected is now random.

Proposition 5. *The above construction is a small world in the sense of Definition 4:*

- *$2 \ln N$ is an almost sure upper-scaling bound for the diameter;*

- *$5 \ln N$ is an upper-scaling bound of the expected average degree;*

- *the expected clustering coefficient is bounded below by $\frac{1}{3}$.*

7.1. Behavior of the diameter

Since $p = r \ln M / M$, with $r = 1 + \delta$, Theorem 7.3 of [5] guarantees that $\ln M / \ln(\ln M)$ is an almost sure upper-scaling bound of the diameter of E. Since the nodes in $(\mathcal{G}_m)_{m=1,\ldots,M}$ are completely connected, it follows that $2 \ln M / \ln(\ln M) + 1$ is an upper-scaling bound for the diameter of Γ^N.

7.2. Behavior of the average degree

To determine the expected average degree, we bound the expected number of links in the graph. Recall that $G_m = \{(i, j) \in \mathcal{G}_m \times \mathcal{G}_m, \ i \neq j\}$ is a complete subgraph of Γ_N. Define C_m as the maximal complete subgraph containing G_m. From the above construction we observe that

$$\Gamma_N = \bigcup_{m=1}^{M} C_m.$$

The number of links in Γ^N is therefore bounded by

$$\sum_{m=1}^{M} \frac{\mathrm{card}([C_m])(\mathrm{card}[C_m] - 1)}{2}.$$

The set $[C_m]$ contains the elements in the cluster $\mathcal{G}_m$ (at most $\ln N + 1$) plus a number of nodes that is bounded by the degree of m in the auxiliary Erdős–Rényi graph, $\deg_E(m)$. Thus, the expected number of total links is bounded by

$$\mathrm{E}\left[\sum_{m=1}^{M} \frac{(\deg_E(m) + \ln N + 1)(\deg_E(m) + \ln N)}{2} \right]$$

$$= \sum_{m=1}^{M} \frac{(\mathrm{E}[(\deg_E(m))^2] + (2\ln N + 1)\,\mathrm{E}[\deg_E(m)] + (\ln N + 1)\ln N)}{2}.$$

Since the degree of every node m in the Erdős–Rényi graph follows a binomial law,

$$\mathrm{Bin}\left(\frac{r \ln M}{M}, M - 1 \right),$$

we obtain $\mathrm{E}[\deg_E(m)] \leq r \ln M$ and $\mathrm{E}[(\deg_E(m))^2] = \mathrm{var}[\deg_E(m)] + (\mathrm{E}[\deg_E(m)])^2 \leq r \ln M + (r \ln M)^2$. Thus, the expected total number of links in the system is less than

$$\frac{M}{2}(r \ln M + (r \ln M)^2 + (2\ln N + 1)r \ln M + (\ln N + 1)\ln N)$$

$$\leq \frac{M}{2}((r^2 + 2r + 1)(\ln N)^2 + (2r + 1)\ln N).$$

Dividing by $N/2 \geq (M \ln N)/2$, the expected average degree is bounded by $(2r + r^2 + 1)\ln N + 2r + 1 \leq 5 \ln N$.

7.3. Behavior of the clustering coefficient

We will now show that the total number of links in the auxiliary Erdős–Rényi graph converges in probability to a bounded random variable. We will use this fact together with Proposition 2 to determine a lower bound for the clustering coefficient.

We apply Theorem 1 to the possible links in the Erdős–Rényi graph. There are $K = M(M-1)/2$ such links. Since each link exists with probability $r \ln M/M$, we can apply the theorem to $(X_i - r \ln M/M)_{i=1}^{K}$, where $(X_i)_{i=1}^{K}$ is an i.i.d. sequence of $\mathrm{Ber}(r \ln M/M)$. We have

$$\frac{r \ln M}{2M} \geq \sigma^2 = \frac{r \ln M}{M} - \left(\frac{r \ln M}{M} \right)^2 \leq \frac{r \ln M}{M} \quad \text{and} \quad t = \frac{\sqrt{\varepsilon}}{\sigma},$$

which gives

$$\alpha = \frac{Lt}{\sigma\sqrt{K}} = \frac{2\sqrt{\varepsilon}}{\sigma^2\sqrt{K}} \leq \frac{4M\sqrt{\varepsilon}}{r\ln M\sqrt{M(M-1)/2}} \leq 1 \quad \text{and} \quad 1 + \frac{\alpha}{3} \leq 2$$

for sufficiently large M and small ε. Since

$$\frac{t^2}{1+\alpha/3} \geq \frac{\varepsilon}{2\sigma^2} \geq \frac{\varepsilon M}{2r\ln M} \geq \sqrt{M},$$

we obtain

$$P\left(\left|\sum_{i=1}^{K}\left(X_i - \frac{r\ln M}{M}\right)\right| \geq \sqrt{\varepsilon K}\right) \leq 2\exp\left\{-\left(\frac{t^2}{1+\alpha/3}\right)\right\} \leq 2\exp\{-\sqrt{M}\}.$$

This estimates the probability that the total number of links in the Erdős–Rényi graph is smaller than

$$\frac{M(M-1)}{2}\frac{r\ln M}{M} + \sqrt{\varepsilon}\sqrt{\frac{M(M-1)}{2}} \leq \frac{(M-1)r\ln M}{2} + \frac{\sqrt{\varepsilon}M}{\sqrt{2}} \leq N$$

for ε chosen sufficiently small and r close to 1. When this is the case,

$$P\left(\limsup \sum_{k,l} \mathbf{1}_{(k,l)\in E} \leq N\right) = 1.$$

Now, we would like to divide the neighbors of each node i into a disjoint, fully connected sets, in order to apply Proposition 2. Let $S_i := \text{card}\{m \mid i \in [C_m]\}$. Each subgraph C_m is fully connected. Thus, we have $V(i) = \bigcup_{\{m \mid i \in [C_m]\}}[C_m] - \{i\}$. This is a union of S_i fully connected sets. By removing elements that belong to more than one set in the union from all but one of the $[C_m]$, we obtain a partition $(A_k)_{k=1,\ldots,S_i}$ of $V(i)$ such that $\mathcal{G}_k - \{i\} \subset A_k \subset [C_k] - \{i\}$. Thus, $\text{card}(A_k) \geq \text{card}(\mathcal{G}_k - \{i\}) \geq \ln N - 2$ and, by Proposition 2, i's clustering coefficient verifies $\gamma_i \geq 1/S_i - 1/(\ln N - 2)$. We have the following bound on $\sum_{i=1}^{N} S_i$:

$$\sum_{i=1}^{N} S_i = \sum_{i=1}^{S_i}\sum_{m=1}^{M} \mathbf{1}_{\{i\in A_m\}}$$

$$\leq \sum_{m=1}^{M}\sum_{i=1}^{N} \mathbf{1}_{\{i\in\mathcal{G}_m\}} + \sum_{l,k} \mathbf{1}_{\{(k,l)\in E\}}$$

$$\leq N + \sum_{l,k} \mathbf{1}_{\{(k,l)\in E\}}.$$

Then

$$P\left(\sum_{i=1}^{N} S_i > 2N\right) \leq P\left(\sum_{l,k} \mathbf{1}_{\{(k,l)\in E\}} > N\right) \leq 2e^{-\sqrt{M}}. \tag{19}$$

We consider the solution to the problem

$$\min_{\{\sum_{i=1}^{N} S_i = 2N\}} \frac{1}{N}\sum_{i=1}^{N} \frac{1}{S_i}.$$

The same type of argument as the one used for (8) shows that the unique solution is $S_1 = S_2 = \cdots = S_N = 2$. Thus, with a probability bounded in (19), the average clustering coefficient is

bounded below by $\frac{1}{2} - 1/(\ln N - 2) \geq \frac{1}{3}$. Consequentially, by Borel Cantelli,

$$P\left(\limsup_N \frac{1}{N} \sum_{i=1}^{N} \gamma_i \leq \frac{1}{3}\right)$$

$$\geq P\left(\limsup_N \frac{1}{N} \sum_{i=1}^{N} \gamma_i \leq \frac{1}{3} \,\Bigg|\, \sum_{i=1}^{N} S_i \leq 2N\right) P\left(\limsup \sum_{i=1}^{N} S_i \leq 2N\right)$$

$$= 1.$$

References

[1] ALDOUS, D. (2004). A tractable complex network model based on the stochastic mean-field model of distance. In *Complex Networks* (Lecture Notes Phys. **650**), Springer, Berlin, pp. 51–87.

[2] BARABASI, A., NEWMAN, M. E. J. AND WATTS, D. (2005). *The Structure and Dynamics of Networks*. Princeton University Press.

[3] BARABASI A. L. AND ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286,** 509–512.

[4] BARBOUR, A. D. AND REINERT, G. (2006). Discrete small world networks. *Electron. J. Prob.* **11,** 1234–1283.

[5] BOLLOBÁS, B. (2001). *Random Graphs*. Academic Press, New York.

[6] BOLLOBÁS, B. AND RIORDAN, O. (2004). The diameter of a scale-free random graph. *Combinatorica* **24,** 5–34.

[7] BOLLOBÁS, B., JANSON, S. AND RIORDAN, O. (2002). Mathematical results on scale-free random graphs. *Handbook of Graphs and Networks – From the Genome to the Internet*, eds S. Bornholdt and H. Schuster, John Wiley, New York, pp. 1–34

[8] BOLLOBÁS, B., JANSON, S. AND RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31,** 3–122.

[9] BOLLOBÁS, B., RIORDAN, O., SPENCER, J. AND TUSNÁDY, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18,** 279–290.

[10] CHUNG, F. AND LU, L. (2006). *Complex Graphs and Networks*. American Mathematical Society, Providence, RI.

[11] CONT, R. AND LOEWE, M. (2008). Social distance and social interactions: a discrete choice model. To appear in *J. Math. Econom.*

[12] DEGENNE, A. AND FORSÉ, M. (1999). *Introducing Social Networks*. Sage Publications, London.

[13] ERDŐS, P. AND RÉNYI, A. (1960). On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutato Int. Közl.* **5,** 17–61.

[14] FERNHOLZ, D. AND RAMACHANDRAN, V. (2007). The diameter of sparse random graphs. *Random Structures Algorithms* **31,** 482–516.

[15] GIRVAN, M. AND NEWMAN, M. E. J. (1999). Community structure in social and biological networks. *Proc. Nat. Acad. Sci. USA* **99,** 7821–7826.

[16] GRANOVETTER, M. (1973). The strength of weak ties. *Amer. J. Sociology* **78,** 1287–1303.

[17] HOFMAN, J. AND WIGGINS, C. (2008). A Bayesian approach to network modularity. *Phys. Rev. Lett.* **100,** 258701.

[18] JACKSON, M. AND ROGERS, B. (2008). The economics of small worlds. *J. Europ. Econom. Assoc.* **3,** 617–627.

[19] JOHNSON, S. (1967). Hierarchical clustering schemes. *Psychometrika* **2,** 241–254.

[20] NEWMAN, M. (2003). The structure and function of complex networks. *SIAM Rev.* **45,** 167–256.

[21] NEWMAN, M., BARABASI, A. AND WATTS, D. (2006). *The Structure and Dynamics of Networks*. Princeton University Press.

[22] NEWMAN, M. E. J., MOORE, C. AND WATTS, D. (2000). Mean-field solution of small world networks. *Phys. Rev. Lett.* **84,** 3201–3204.

[23] NEWMAN, M. E. J., WATTS, D. AND STROGATZ, S. H. (2002). Random graph models of social networks, *Proc. Nat. Acad. Sci. USA* **99,** 2566–2572.

[24] TRAVERS, J. AND MILGRAM, S. (1969). An experimental study of the small world problem. *Sociometry* **32,** 425–443.

[25] WATTS, D. J. (1999). *Small Worlds*. Princeton University Press.

[26] WATTS, D. J. AND STROGATZ, S. H. (1998). Collective dynamics of small world networks. *Nature* **393,** 440–442.