Communications in Information and Systems
Volume 0, Number 0, 1, 2022

Compression and Symmetry of Small-World Graphs and Structures

Ioannis Kontoyiannis*, Yi Heng Lim,
Katia Papakonstantinopoulou, and Wojtek Szpankowski†

For various purposes and, in particular, in the context of data compression, a graph can be examined at three levels. Its structure can be described as the unlabelled version of the graph; then the labelling of its structure can be added; and finally, given then structure and labelling, the contents of the labels can be described. Determining the amount of information present at each level and quantifying the degree of dependence between them requires the study of symmetry, graph automorphism, entropy, and graph compressibility. In this paper, we focus on a class of small-world graphs. These are geometric random graphs where vertices are first connected to their nearest neighbours on a circle and then pairs of non-neighbours are connected according to a distance-dependent probability distribution. We establish the degree distribution of this model, and use it to prove the model's asymmetry in an appropriate range of parameters. Then we derive the relevant entropy and structural entropy of these random graphs, in connection with graph compression.

AMS 2000 subject classifications: Primary 60Fxx, 94A15; secondary 58R01, 05C80.

Keywords and phrases: Entropy, Information, Lossless compression, Random graph, Random network, Small-world graph, Graphical structure, Symmetry.

*Supported in part by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant," project number 1034.

†Supported in part by NSF Center for Science of Information (CSoI) Grant CCF-0939370, and in addition by NSF Grants CCF-1524312, CCF-2006440, and CCF-2007238.

1. Introduction

Our main aim in this work is to develop rigorous results on structural properties that are fundamental to statistical and information-theoretic problems involving the information shared between the labels and the structure of a random graph, specifically within the class of *small-world graphs*. For various statistical and signal processing tasks and, in particular, in the context of data compression, the information present in a graph can be examined at three levels. First, its *structure* can be described, that is, the unlabelled version of the graph. Second, its labelling can be described given its structure. And third, the actual contents of the labels can be described, given the structure and the labelling [41]. In some problems, for example in recovering the node arrival order of dynamic networks [30], the goal is to first *recover* label information by examining a graph structure, and then to explain the structural properties (such as symmetry) involved in their analysis.

More formally, the labelled and unlabelled graph compression problems can be described as follows. Fix a graph model on the collection $\mathcal{G}(n)$ of all simple, undirected, labelled graphs on n vertices. First, we aim to understand the best achievable performance of efficiently computable *source codes* for this model [16]. A source code (C_n, D_n) here consists of an encoder C_n mapping graphs in $\mathcal{G}(n)$ to finite-length bit strings, and of a decoder D_n that inverts C_n. The goal is to make the (expected) length of the output bit string as short as possible. Of particular interest to us here is the related problem of the compression of graph *structures*. In this case, the encoder C_n is presented with a graph G_n isomorphic to a sample from $\mathcal{G}(n)$, and $D_n(C_n(G_n))$ is only required to be a labelled graph isomorphic to G_n, so that only the structural information is preserved. We again seek to characterize efficient source codes with minimal code lengths. This optimal compression performance is characterized by the entropy of the distribution on unlabelled graphs induced by the model, which we call its *structural entropy*.

Structural properties. Several interesting structural properties and quantities arise naturally in connection with graph compression. As we describe next, determining the structural entropy often involves computing the size of the automorphism group of a graph, as well as the typical number of positive-probability *labelled representatives* (re-labellings or permutations) of a given structure.

In general, given a labelled graph G_n generated by some model on $\mathcal{G}(n)$, all $n!$ label permutations lead to the same *structure* $S_n := S_n(G_n)$; *however*, not all permutations may be permissible under the model, and some permutations may lead to the exact same graph. The latter property is well characterized by the automorphism group, $\mathrm{Aut}(G_n)$, of G_n. When the cardinality of the automorphism group is one, then the graph is *asymmetric* since every feasible permutation is distinct (in term of the labelled graph) and gives the same structure. In some cases, such as the Erdős-Rényi (ER) model [12] and preferential attachment graphs (PAG) [8, 28], all permutations lead to the same graph with high probability. In other words, these models are invariant under isomorphism. Furthermore, in the ER model every permutation is feasible, unlike under the PAG model. For PAG graphs, the number of distinct re-labellings can be computed as the ratio of the number of feasible permutations, $|\Gamma(G_n)|$, and the size of the automorphism group, $|\mathrm{Aut}(G_n)|$. As a consequence, the structural entropy of the unlabelled graph is a function of $\log |\Gamma(G_n)|/|\mathrm{Aut}(G_n)|$ as well as of the (labelled) graph entropy. However, when the model is not invariant under isomorphism, we need to actually estimate the *conditional entropy* of the (labelled) graph under a given structure.

One such class is the family of small-world graphs [44, 25] on the circle. In this paper we focus on the symmetry, entropy and automorphism properties of graphs generated by this model.

Contributions. We study the small-world model [44, 25] where n vertices are arranged on the circle in increasing order, and each node is connected to its two nearest neighbours. Then different pairs of nodes are (independently) connected with probability proportional to $1/k^a$, where k is their distance and $a \in (0, 1)$ is fixed parameter; precise definitions are given in Section 3. Such a model does not satisfy the two properties discussed above: It is not invariant under isomorphism, and not every permutation is feasible.

For the small-word model we first compute the mean degree of a node (Proposition 3.1), and in Theorem 3.4 we prove that it is asymmetric with high probability. This allows us to derive very accurate asymptotic estimates for the graph entropy and structural entropy; these are presented in Theorem 4.2 and Corollary 4.3. Finally, in Theorem 4.4 we give a precise upper bound on the conditional entropy of a small-world graph given its structure, a result which is of independent interest from both the combinatorial and information-theoretic points of view.

Prior work. There is a long history of very detailed results on the problem of determining the fundamental limits of the best achievable compression performance for sequential data; see, e.g., [40, 26, 27] and the references therein. However, the study of the compression problem for graph and tree models, in both the information theory and the computer science literature, is more recent [45, 7, 1, 15, 17, 13]. In 1990, Naor [34] proposed an efficiently computable representation for unlabelled graphs (answering Turán's [42] open question), and showed that this representation is optimal up to the second leading term of the entropy when all unlabelled graphs are equally likely. Naor's result is, asymptotically, a special case of corresponding expansions developed later in [15], where general ER graphs were analyzed. Further extensions to PAG graphs were derived in [29].

An approach based on automata, was used in [33] to design an optimal graph compression scheme. Recently, the authors of [18] proposed a general universal lossless source coding algorithm for graphs, and there are also a number of heuristic methods for real-world graph compression, including a grammar-based scheme for data structures [13, 32, 38]. Efficient compression algorithms were developed in [10], leveraging symmetry properties of graphs arising in connection with deep neural networks. A comprehensive survey of lossless graph compression algorithms can be found in [11].

There are a number of studies of the compression problem for trees [21, 23, 45, 31, 22]. For binary, plane-oriented trees, information-theoretic results were rigorously obtained in [31], and a universal, grammar-based lossless coding scheme was proposed in [22].

In the computer science literature, the focus has been almost exclusively on algorithmic complexity, and very little attention seems to have been given to comparisons with fundamental information-theoretic compression measures – which is the main focus of this paper. Also, work in both communities has largely been restricted to labelled graphs, or graphs with strong edge independence assumptions (with the exception of [2, 31]). As we show, interesting additional complications arise when the goal is to compress graphical *structures*.

Paper organization. In the next section, after some technical preliminaries, we review some known symmetry and structural entropy properties of the ER and PAG models. The small-world graph model is introduced in Section 3, where its degree distribution is determined and its asymmetry established. Our main results on the graph entropy and structural entropy of small-world graphs are stated and proved in Section 4.

2. Preliminaries: Random Graphs and Entropy

2.1. Graphs, structures, labels, and symmetry

Let $\mathcal{G}(n)$ denote the class of all (undirected, simple, labelled) graphs $G = (V, E)$ on $n = |V|$ vertices, where for simplicity we take $V = \{1, 2, \dots, n\}$

throughout. Let P_n denote a *model* for such graphs, that is, a discrete probability mass function (PMF) on $\mathcal{G}(n)$. For example, under the classical Erdős-Rényi (ER) model [20, 12] with parameter p, for $G \in \mathcal{G}(n)$,

$$P_n(G) = p^{|E|}(1-p)^{\binom{n}{2}-|E|},$$

where $|E|$ is the number of edges of $G = (V, E)$. On the other hand, for the preferential attachment graphs $\mathrm{PA}(m, n)$ studied, e.g., in [8], where a new node connects to m existing nodes with probability which is proportional to their degree, the probability $P_n(G)$ does not depend only on $|E|$.

Any such model P_n induces a probability distribution Q_n on structures. Let $\mathcal{S}(n)$ denote the class of all unlabelled graphs of n vertices. Then the induced probability of a structure $S \in \mathcal{S}(n)$ is the sum of the probabilities of all graphs G with the same structure S,

$$Q_n(S) = \sum_{G \in \mathrm{Iso}(S)} P_n(G), \tag{1}$$

where $\mathrm{Iso}(S) \subset \mathcal{G}(n)$ is the isomorphism equivalence class consisting of all graphs in $\mathcal{G}(n)$ with structure S.

Some standard models P_n, such as the simple ER model and $\mathrm{PA}(m, n)$ graphs [28], are invariant under isomorphism, that is, $P_n(G) = P_n(G')$, whenever G and G' are both permissible (i.e., they have nonzero probability under P_n) and there is an S such that both $G, G' \in \mathrm{Iso}(S)$. In such cases we simply have,

$$Q_n(S) = P_n(G) \cdot |\mathrm{Iso}(S) \cap \Pi_n|,$$

where G is any graph in $\mathrm{Iso}(S)$ and Π_n denotes the collection of all permissible graphs in $\mathcal{G}(n)$, i.e., the support of P_n. If, in addition, every permutation is permissible by a given model – as in ER model – then the number of graphs isomorphic to a given G is equal to the number of permutations of the labels, $n!$, divided by the number of such permutations that lead to exactly the same graph, namely, the size of the automorphism group Aut(G) of G. Therefore,

$$Q_n(S) = P_n(G) \cdot \frac{n!}{|\mathrm{Aut(G)}|}. \tag{2}$$

More generally, in cases like the $\mathrm{PA}(m, n)$ model [28], where not all permutations are permissible, we have,

$$Q_n(S) = P_n(G) \cdot \frac{|\Gamma(G)|}{|\mathrm{Aut(G)}|}. \tag{3}$$

where $\Gamma(G)$ is the set of permissible permutations. For example, for $\mathrm{PA}(m,n)$ we know that $E[\log|\Gamma(G)|] = n\log n - O(n\log\log n)$ [28]. [Throughout the paper, log denotes the natural logarithm $\log_e$.] As we will see later, the small-world model considered here is not invariant under isomorphism, and not every permutation is permissible.

It is of interest to know how much symmetry a given graph has. In particular, in some applications one needs to know if a graph is asymmetric, as defined below.

Definition 2.1. *A graph G is called* asymmetric *if* $|\mathrm{Aut(G)}| = 1$.

It is known that, in appropriate parameter ranges, the ER [24] and PAG [28] models generate asymmetric graphs with high probability:

Theorem 2.2 (ER asymmetry [24])**.** (i) *For a sequence of random graphs $\{G_n\}$ under the ER model with parameters $\{p_n\}$, such that, as $n \to \infty$,*

$$p_n \gg \frac{\log n}{n} \quad \textit{and} \quad 1 - p_n \gg \frac{\log n}{n},$$

we have, for any $t > 0$,

$$\Pr(G_n \textit{ is symmetric}) = O(n^{-t}),$$

as $n \to \infty$.
(PAG asymmetry [28]) (ii) *For a sequence of random graphs $\{G_n\}$ under the* $\mathrm{PA}(m,n)$ *model with $m \geq 3$, we have that, for some $\delta > 0$,*

$$\Pr(G_n \textit{ is symmetric}) = O(n^{-\delta}),$$

as $n \to \infty$.

One of our main results below will be the development of a statement analogous to Theorem 2.2 for a class of small-world random graphs.

2.2. Entropy and compressibility

Detailed asymptotic expansions for the graph entropy $H(G_n)$ under the ER and PAG models are known as we review below. First we note, without proof, a simple expression for the binary entropy function.

Lemma 2.3. *The binary entropy function* $h(p) = -p \log p - (1-p)\log(1-p)$ *satisfies, as* $p \to 0$,

$$h(p) = p \log\left(\frac{1}{p}\right) + p - \frac{1}{2}p^2 + O(p^3).$$

Moreover, the error term always satisfies $-(1/2)p^3 \le O(p^3) \le 0$*, and for* $p \le 1/4$ *it also satisfies* $O(p^3) \le -(1/10)p^3$.

Lemma 2.4. (i) **(ER graph entropy)** *For a sequence of ER random graphs* $\{G_n\}$ *with parameters* $\{p_n\}$,

$$H(G_n) = \frac{n(n-1)}{2} h(p_n),$$

and if $p_n \to 0$ *as* $n \to \infty$,

$$H(G_n) = \frac{n(n-1)}{2}\left[p_n \log\left(\frac{1}{p_n}\right) + p_n - \frac{1}{2}p_n^2 + O(p_n^3)\right].$$

(ii) **(PAG graph entropy [39, 28])** *For a sequence of* $\mathrm{PA}(m, n)$ *random graphs* $\{G_n\}$*, we have, as* $n \to \infty$,

$$H(G_n) = mn \log n + m\left(\log 2m - 1 - \log m! - A\right)n + o(n), \tag{4}$$

where,

$$A = \sum_{d=m}^{\infty} \frac{\log d}{(d+1)(d+2)}.$$

Proof. We only sketch the proof for the ER model. By definition, $G_n \sim P_n$ describes $\binom{n}{2}$ independent Bern(p_n) random variables, so, $H(G_n) = \binom{n}{2}h(p_n)$ and using Lemma 2.3 gives the claimed result. The proof of (4) can be found in [28]. $\square$

For a random graph G_n with structure S_n, the chain rule for entropy implies that,

$$H(S_n) = H(G_n) - H(G_n|S_n). \tag{5}$$

Using this identity together with relation (2) in combination with Theorem 2.2, Choi and Szpankowski [15] for the ER model and Luczak et al. [28]

for PAG graphs, establish the following asymptotic expansions for the entropy of ER and PAG random structures. An analogous expansion for a class of small-world graphs is established in this paper.

Theorem 2.5 (ER structural entropy)**.** (i) *For a sequence of ER random graphs* $\{G_n\}$ *with parameters* $\{p_n\}$ *that satisfy,*

$$p_n \gg \frac{\log n}{n} \qquad \text{and} \qquad 1 - p_n \gg \frac{\log n}{n},$$

as $n \to \infty$*, we have, for some* $\beta > 0$*,*

$$H(S_n) = \frac{n(n-1)}{2} h(p_n) - \log n! + O\Big(\frac{\log n}{n^\beta}\Big).$$

(PAG structural entropy) (ii) *For a sequence of* PA(m, n) *random graphs* $\{G_n\}$ *with* $m \geq 3$ *we have, as* $n \to \infty$*,*

$$H(S_n) = (m-1)n \log n + R_n, \tag{6}$$

where R_n *satisfies,*

$$Cn \leq |R_n| \leq O(n \log \log n),$$

for some nonzero constant $C = C(m)$*.*

Definition 2.6. *The* compressibility *of a random graph* $G_n = (V_n, E_n)$ *is measured by the average number of bits (or, rather, nats) per edge used in its best possible description, that is,* $C_n = H(G_n)/E(|E_n|)$*. We say that the sequence of random graphs* $\{G_n\}$ *is* compressible*, if* $C_n = O(1)$*.*

Recent studies indicate that many real-world examples of large graphs, including web graphs and social media graphs, are compressible. For an extensive discussion of compressibility in different models see [14].

For the ER model we note that each node has $\text{Bin}(n-1, p_n)$ edges, which is $\approx \text{Po}(np_n)$ for large n, as long as $p_n = o(1)$; to see this, recall Theorem 1 of [9]. Also, $E(|E_n|) = n(n-1)p_n/2$, so by Lemma 2.4 in this case,

$$C_n = \frac{H(G_n)}{E(|E_n|)} \sim -\log p_n,$$

which is unbounded. Therefore, in the above sense, the ER model with parameters $p_n = o(1)$ is incompressible. Similarly, for PA(m, n) graphs, we have $C_n \sim \log n$, for $m \geq 3$.

3. A small-world model

Here we examine a small-world model, similar to those introduced in [44, 25]. More specifically, it is a Newman-Watts-type model [37]; also see [35, 14, 2]. It is a model of *geometric* random graphs, with high-clustering properties that differentiate them from ER and PAG models [19, 36].

Consider the vertex set $V = \{1, 2, \ldots, n\}$ arranged on the circle, with each vertex connected by an edge to its two nearest neighbours. For each one of the remaining $\binom{n}{2} - n$ pairs of vertices (u, v), we add an edge between them with probability $p(|u - v|)$, where $|u - v|$ is the discrete distance on the circle and $p_n(k) = c_n k^{-a}$, for some $a \in (0, 1)$ and with,

$$c_n = b_n(1 - a)\left(\frac{2}{n}\right)^{1-a},$$

where $\{b_n\}$ is a nondecreasing, unbounded sequence of positive real numbers, with $b_n = o(n^{1-a})$, as $n \to \infty$. In all the results and discussion below we implicitly assume that n is large enough so that all the $p_n(k)$ are less than one, which is always possible by the assumptions on b_n.

The graph shown in Figure 1 on is an example of a small world graph with $n = 16$; nearest neighbour edges are shown in blue and random edges are green. Note that there are more edges between nearby nodes and fewer between distant ones.

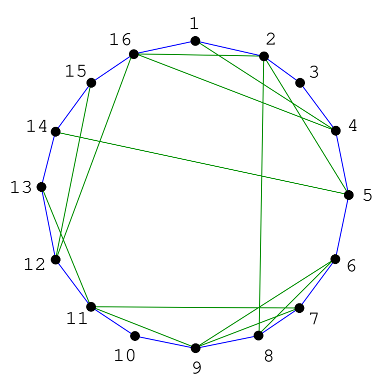

Figure 1: Example of a small world graph on $n = 16$ vertices.

We call a random undirected graph G_n generated by this model a *random small-world graph with parameters* a *and* b_n, and we write $G_n \sim \mathrm{SW}(a, b_n)$. It is assumed throughout that $a \in (0, 1)$ and $b_n = o(n^{1-a})$.

Before examining the $\mathrm{SW}(a, b_n)$ class further, we remark that small-world models are an important class of *geometric* random graphs in that, unlike in the ER model, the connectivity of a small-world graph depends on the actual locations of the nodes. Although it will not play a role in our analysis, we mention that another important characteristic of such graphs is the "small-world property." This means that the graph distance between any two nodes is much smaller than in a purely random graph, with high probability; see the above references or the texts [36, 43] for details.

3.1. Degree distribution

Proposition 3.1 (SW mean degree)**.** *The mean degree μ_n of an arbitrary node in a random graph $G_n \sim \mathrm{SW}(a, b_n)$ satisfies,*

$$\mu_n = 2b_n + 2 + O\Big(\frac{b_n}{n^{1-a}}\Big),$$

as $n \to \infty$.

For the proof we need the following lemmas. The expansions in Lemma 3.2 follow from straightforward applications of Euler-Maclaurin summation; see, e.g., [5, 6].

Lemma 3.2. *As $n \to \infty$,*

$$\sum_{k=1}^{n} \frac{1}{k} = \log n + \gamma + \frac{1}{2n} + O\Big(\frac{1}{n^2}\Big),$$

where γ is Euler's constant and the error term is bounded in absolute value by $\frac{1}{6n^2}$, for all $n \geq 2$. Also, as $n \to \infty$, for any $s > 0$, $s \neq 1$,

$$\sum_{k=1}^{n} \frac{1}{k^s} = \frac{1}{(1-s)n^{s-1}} + \zeta(s) + \frac{1}{2n^s} + O\Big(\frac{1}{n^{s+1}}\Big),$$

where ζ is the Riemann zeta function,

$$\zeta(s) := \begin{cases} \displaystyle\sum_{k=1}^{\infty} \frac{1}{k^s}, & \text{if } s > 1, \\ \displaystyle\lim_{M\to\infty} \left[\sum_{k=1}^{M} \frac{1}{k^s} - \frac{M^{1-s}}{(1-s)}\right], & \text{if } s \in (0,1), \end{cases}$$

and the error term is bounded in absolute value by $\frac{s}{6n^{s+1}}$.

Next we will apply Lemma 3.2 to get some simple estimates regarding the probabilities $p_n(k)$. The proof of Lemma 3.3 is given in the Appendix.

Lemma 3.3. *For odd n, let:*

$$S_{n,1} = 2 \sum_{k=2}^{(n-1)/2} p_n(k), \qquad S_{n,2} = 2 \sum_{k=2}^{(n-1)/2} p_n(k)^2.$$

Similarly, for even n, let:

$$S'_{n,1} = 2 \sum_{k=2}^{(n-2)/2} p_n(k) + p_n(n/2), \qquad S'_{n,2} = 2 \sum_{k=2}^{(n-2)/2} p_n(k)^2 + p_n(n/2)^2.$$

Then, as $n \to \infty$,

$$\begin{aligned} S_{n,1} &= 2b_n + O\Big(\frac{b_n}{n^{1-a}}\Big), \\ S_{n,2} &= \begin{cases} \Big(\frac{4(1-a)^2}{1-2a}\Big)\frac{b_n^2}{n} + O\Big(\frac{b_n^2}{n^{2-2a}}\Big), & a \in (0,1/2), \\ \frac{b_n^2 \log n}{n} + O\Big(\frac{b_n^2}{n}\Big), & a = 1/2, \\ 2^{3-2a}(1-a)^2\zeta(2a)\frac{b_n^2}{n^{2-2a}} + O\Big(\frac{b_n^2}{n}\Big), & a \in (1/2,1), \end{cases} \end{aligned}$$

and the same results hold with $S'_{n,1}$ in place of $S_{n,1}$, and $S'_{n,2}$ in place of $S_{n,2}$.

Proof of Proposition 3.1. The edges of G_n can be described as $\binom{n}{2} - n$ independent Bernoulli random variables. Choose and fix $n \geq 5$ be arbitrary.

Suppose n is odd. Considering, without loss of generality, the node $u = 1$, let X_k, Y_k, for $k = 2, 3, \ldots, (n-1)/2$, denote binary random variables, where each X_k and each Y_k describe whether node $u = 1$ is connected to a different node at distance k from node $u = 1$. Then $\{X_k, Y_k\}$ are independent Bernoulli random variables with corresponding parameters $\{p_n(k)\}$, and the degree of node $u = 1$, W_n, say, can be expressed as $W_n = [2 + \sum_k (X_k + Y_k)]$. Therefore, the mean degree of any vertex is,

$$\mu_n := E(W_n) = 2 + 2 \sum_{k=2}^{(n-1)/2} p_n(k) = 2 + S_{n,1}. \tag{7}$$

Similarly, if n is even, there are n possible edges between pairs of nodes at each distance $k = 2, 3, \ldots, \frac{n-2}{2}$, and $n/2$ possible edges between pairs of nodes at distance $n/2$. Here, the mean degree of a vertex is,

$$\mu_n = 2 + 2\sum_{k=2}^{(n-2)/2} p_n(k) + p_n(n/2) = 2 + S'_{n,1}. \tag{8}$$

Combining (7) and (8) with Lemma 3.3 completes the proof. □

3.2. Asymmetry

Let $G = (V, E)$ be an arbitrary undirected graph on $V = \{1, 2, \ldots, n\}$, with no self loops. We first make a series of definitions following the terminology of [24].

The set of neighbours of a vertex $u \in V$ is denoted,

$$N(u) = \{v \in V \ : \ (u, v) \in E\}.$$

Let π be any permutation on V. The *defect of a vertex* $u \in V$ under the permutation π is,

$$D_\pi(u) = |N(\pi(u)) \triangle \pi(N(u))|,$$

which can also be expressed as,

$$\begin{aligned} D_\pi(u) = \sum_{v \neq \pi(u)} \Big[&\mathbb{I}\{(\pi(u), v) \in E,\, (u, \pi^{-1}(v)) \notin E\} \\ &+ \mathbb{I}\{(\pi(u), v) \notin E,\, (u, \pi^{-1}(v)) \in E\}\Big]. \end{aligned} \tag{9}$$

The *defect of the graph G under π* is, $D_\pi(G) = \max_{u \in V} D_\pi(u)$, and the *total defect of G* is,

$$D(G) = \min_{\pi \neq \mathsf{id}} D_\pi(G),$$

where id denotes the identity permutation. Note that G is asymmetric iff $D(G) \neq 0$.

Theorem 3.4 (SW asymmetry)**.** *Let $\{G_n\}$ be a sequence of small-world random graphs, $G_n \sim \mathrm{SW}(a, b_n)$, $n \geq 1$. If,*

$$b_n = o(n^{1-a}), \qquad \textit{and} \qquad \frac{b_n}{\log n} \to \infty, \qquad \textit{as } n \to \infty,$$

then, for any $t > 0$, as $n \to \infty$:

$$\Pr(G_n \text{ is symmetric}) = O(n^{-t}).$$

Following [24], the proof of Theorem 3.4, given in the Appendix, is based in part on an application of the following simple concentration bound.

Proposition 3.5. [3, 4] *Let $Z = f(\xi_1, \xi_2, \ldots, \xi_m)$ be a function of the independent Bernoulli random variables $\{\xi_i\}$, and suppose that f has the bounded difference property that, for some $c > 0$,*

$$\begin{aligned} \max_{j,\{\xi_i\}} \big| f(\xi_1, \ldots, \xi_{j-1}, \xi_j, \xi_{j+1}, \ldots, \xi_m) & \\ - f(\xi_1, \ldots, \xi_{j-1}, 1-\xi_j, \xi_{j+1}, \ldots, \xi_m) \big| &\leq c. \end{aligned} \tag{10}$$

Let $p_i = E(\xi_i)$ for each i, and $\sigma^2 = c^2 \sum_i p_i(1-p_i)$. Then, for all $0 < t < 2\sigma/c$:

$$\Pr\left[|Z - E(Z)| > t\sigma \right] \leq 2e^{-t^2/4}.$$

4. Entropy of the small-world model

4.1. Graph entropy

As with Lemma 3.2, the expansions in Lemma 4.1 below are easy applications of Euler-Maclaurin summation [5, 6]. It will be used in the proof of Theorem 4.2, given in the Appendix.

Lemma 4.1. *As $n \to \infty$,*

$$\sum_{k=1}^{n} \frac{\log k}{k} = \frac{1}{2}(\log n)^2 + \gamma' + \frac{1}{2}\frac{\log n}{n} + O\Big(\frac{\log n}{n^2}\Big),$$

where γ' is defined, in analogy to Euler's constant, as,

$$\gamma' = \lim_{n \to \infty} \left[\sum_{k=1}^{n} \frac{\log k}{k} - \frac{1}{2}(\log n)^2 \right],$$

and the error term is bounded in absolute value by $\frac{1+\log n}{6n^2}$, for all $n \geq 2$. Also, as $n \to \infty$, for any $s > 0$, $s \neq 1$,

$$\sum_{k=1}^{n} \frac{\log k}{k^s} = \frac{\log n}{(1-s)n^{s-1}} - \frac{1}{(1-s)^2 n^{s-1}} - \zeta'(s) + \frac{\log n}{2n^s} + O\Big(\frac{\log n}{n^{s+1}}\Big),$$

where the error term is bounded in absolute value by $\frac{1+s\log n}{6n^{s+1}}$, for all $n \geq 2$.

Theorem 4.2 (SW graph entropy)**.** *Let $G_n \sim \mathrm{SW}(a, b_n)$, $n \geq 1$, be a sequence of small-world random graphs with,*

$$b_n = o(n^{1-a}), \qquad \text{and} \qquad \frac{b_n}{\log n} \to \infty, \qquad \text{as } n \to \infty.$$

The entropy of this small-world model is,

$$H(G_n) = nb_n\big[\log n - \log b_n - C_a + o(1)\big], \tag{11}$$

where,

$$C_a = a\Big(\frac{1 + \log 2}{1 - a}\Big) + \log\Big((1-a)2^{1-a}\Big) - 1. \tag{12}$$

Remark. Note that, combining the above expansion for the entropy $H(G_n)$ with the expression for the mean degree of an arbitrary node in G_n given in Proposition 3.1, we have that, as $n \to \infty$,

$$\frac{H(G_n)}{E(|E_n|)} = \frac{nb_n \log n(\delta_n + o(1))}{2nb_n + O(n)} \sim \frac{\delta_n}{2} \log n,$$

where the positive sequence $\{\delta_n\}$ is bounded above and bounded away from zero. Therefore, the average number of "bits per edge" in G_n is unbounded, so in the terminology of [14] the $\mathrm{SW}(a, b_n)$ model here is incompressible.

4.2. Structural entropy

Having an estimate for the entropy of a random graph $G_n \sim \mathrm{SW}(a, b_n)$ in Theorem 4.2, it is easy to get a corresponding estimate for the entropy of the random structure S_n associated with G_n. Since, given S_n, there are at most $n! \leq n^n$ possible graphs G_n with structure S_n, we have that,

$$H(G_n|S_n) \leq n \log n. \tag{13}$$

And since $H(S_n) = H(G_n) - H(G_n|S_n)$ as noted in (5), combining (13) with (11) immediately yields:

Corollary 4.3 (SW structural entropy)**.** *Under the assumptions of Theorem 4.2, the entropy of the structures S_n associated with the small-world random graphs $G_n \sim \mathrm{SW}(a, b_n)$ satisfies,*

$$H(S_n) = nb_n\big[\log n - \log b_n - C_a + o(1)\big],$$

where the constant C_a is given in (12).

Finally we examine the conditional entropy $H(G_n|S_n)$, which describes the degree of uncertainty that remains about the graph G_n after knowing its structure S_n. In Theorem 4.4 we obtain a slightly more refined estimate than the crude upper bound in (13), which gives a tighter result when $b_n = o(n^t)$ for all $t > 0$.

Theorem 4.4 (SW conditional entropy)**.** *Let $G_n \sim \mathrm{SW}(a, b_n)$ be a sequence of small-world random graphs with associated structures S_n, $n \geq 1$. Suppose that,*

$$b_n = o(n^{1-a}), \qquad \textit{and} \qquad \frac{b_n}{\log n} \to \infty, \qquad \textit{as } n \to \infty.$$

Then the conditional entropy of the graph G_n given its structure S_n has:

$$H(G_n|S_n) \leq n \log b_n + (\log 5)n + \log\left(\frac{n}{b_n}\right) + O(1).$$

First we establish a simple, general upper bound. As in Section 2, we write P_n for the PMF of G_n on $\mathcal{G}(n)$ and similarly Q_n for the induced PMF of S_n in $\mathcal{S}(n)$. We also write $\mathcal{G}_a(n) \subset \mathcal{G}(n)$ for the support of P_n, and we call graphs $G \in \mathcal{G}_a(n)$ *admissible.*

Lemma 4.5. *For any graph $G \in \mathcal{G}_a(n)$ with structure S, let $\tau(G)$ denote the number of admissible graphs G' that are isomorphic to G,*

$$\tau(G) = |\mathrm{Iso}(S) \cap \mathcal{G}_a(n)|.$$

Then:

$$H(G_n|S_n) \leq \sum_{G \in \mathcal{G}(n)} P_n(G) \log \tau(G).$$

Proof. First observe that $\tau(G)$ is the same for all $G \in \mathrm{Iso}(S) \cap \mathcal{G}_a(n)$. Therefore, with only a slight abuse of notation, we will write $\tau(S)$ for $\tau(G)$ if S is the structure of some $G \in \mathcal{G}_a(n)$. In analogy with $\mathcal{G}_a(n)$, let $\mathcal{S}_a(n)$ denote the set of admissible structures, i.e., those $S \in \mathcal{S}(n)$ for which there is an admissible G with structure S. Then we have,

$$\begin{aligned} H(G_n|S_n) &= \sum_{S \in \mathcal{S}_a(n)} Q_n(S) H(G_n|S_n = S) \\ &\overset{(a)}{\leq} \sum_{S \in \mathcal{S}_a(n)} Q_n(S) \log \tau(S), \end{aligned}$$

so that,

$$\begin{aligned} H(G_n|S_n) &\overset{(b)}{=} \sum_{S\in\mathcal{S}_a(n)} \sum_{G\in\mathrm{Iso}(S)} P_n(G)\log\tau(S) \\ &= \sum_{G\in\mathcal{G}(n)} P_n(G)\log\tau(G), \end{aligned}$$

where (a) follows from the elementary fact that the entropy of a random variable with m values is at most $\log m$, and (b) follows from the basic observation (1). □

Next we obtain a simple bound on the tails of the degree of the nodes of $G_n \sim \mathrm{SW}(a, b_n)$. Its proof is given in the Appendix.

Proposition 4.6. *Under the assumptions of Theorem 4.4, the probability that there is at least one node in G_n with degree greater than $9b_n/2$ is $O(n^{-t})$, for any $t > 0$.*

We are now in a position to prove the theorem.

PROOF OF THEOREM 4.4. Let $\mathcal{B}_n$ be the collection of 'bad' graphs $G \in \mathcal{G}_a(n)$ in the sense of Proposition 4.6, that have at least one node with degree greater than $d := 9b_n/2$. For any 'good' graph $G \in \mathcal{B}_n^c$, we can estimate $\tau(G)$ as follows. Suppose G has structure S and let $G' \in \mathrm{Iso}(S) \cap \mathcal{G}_a(n)$ be not identical to G. Let $\pi \neq \mathsf{id}$ be the permutation on $V = \{1, 2, \ldots, n\}$ that maps G to G'. For G' to be admissible it must contain the cycle of edges $1 - 2 - \cdots - n - 1$, which means that G must contain the cycle,

$$\pi^{-1}(1) - \pi^{-1}(2) - \cdots - \pi^{-1}(n) - \pi^{-1}(1).$$

So to bound $\tau(G)$ it suffices to get an upper bound on the number of permutations π with this property.

Fix an arbitrary $i \in V$ as $i = \pi^{-1}(1)$. Since $G \in \mathcal{B}_n^c$, the degree of i is at most d, so there are at most d choices for the node $\pi^{-1}(2)$, and similarly, there are then at most d choices for $\pi^{-1}(3)$. Continuing this way, there are at most a total of d^{n-1} choices for the values of $\pi^{-1}(j)$, for $j = 2, 3, \ldots, n$, and an additional n choices for the initial value of $i = \pi^{-1}(1)$. Therefore, there are at most nd^{n-1} possible such permutations, and so,

$$\tau(G) \leq n(5b_n)^{n-1}.$$

Finally, we can substitute this in Lemma 4.5 to obtain that,

$$\begin{aligned}
H(G_n|S_n) &\leq \sum_{G\in\mathcal{B}_n^c} P_n(G)\log\tau(G) + \sum_{G\in\mathcal{B}_n} P_n(G)\log\tau(G) \\
&\leq \log\left[n(5b_n)^{n-1}\right] + P_n(\mathcal{B}_n)\log n!\,,
\end{aligned}$$

and using the elementary bound $n! \leq n^n$, and Proposition 4.6 with $t = 2$, we obtain,

$$H(G_n|S_n) \leq n\log b_n + (\log 5)n + [\log n - \log b_n] + O(1),$$

as claimed. □

5. Conclusions

This works examines the degree of compressibility of random graphs and structures generated by a one-dimensional version of the Newman-Watts small-world model. First, it is shown that graphs from that model are asymptotically asymmetric with high probability, and the graph entropy of the model is computed. Then, using this symmetry, it is established that the structure entropy is asymptotically equal to the graph entropy – with equality proved for the first three (and most significant) terms in their asymptotic expansion. Finally, a more accurate bound is given on the conditional entropy of the random graph itself given its structure.

Potential applications of this work can be developed in areas where large graphs naturally arise, with characteristics similar to those in the model examined here; e.g., see [44, 37, 25, 35, 36] for references to empirical studies involving graphical data sets with high clustering and other small-world properties. In particular, our results can provide theoretical guidelines for designing effective compression algorithms for such data sets, as well as benchmark values for the fundamental limits of the best compression ratios that can be achieved theoretically.

An interesting and important direction for future work is the design of efficient, near-optimal compression algorithms for random-world random structures. These could have important applications for the communication and storage of many real-world data sets, including, e.g., metabolite processing networks, neuronal brain networks, and social influence networks.

Finally we note that all the basic estimates in Lemmas 2.3, 3.2, and 4.1 are given in nonasymptotic form with closed-form expressions for the error terms. Therefore, we expect that all the asymptotic results in this paper can, with some additional work, be turned into precise, nonasymptotic, finite-n bounds with explicit constants.

Appendix: Proofs

PROOF OF LEMMA 3.3. We only give the proof for odd n; the case of even n is similar.

For $S_{n,1}$, by the definition of the $p_n(k)$ we have,

$$S_{n,1} = 2b_n(1-a)\Big(\frac{2}{n}\Big)^{1-a}\left[\sum_{k=1}^{(n-1)/2}\frac{1}{k^a} - 1\right],$$

and by Lemma 3.2 this is,

$$S_{n,1} = 2b_n(1-a)\Big(\frac{2}{n}\Big)^{1-a}\left[\frac{[(n-1)/2]^{1-a}}{1-a} + O(1)\right] = 2b_n + O\Big(\frac{b_n}{n^{1-a}}\Big).$$

For $S_{n,2}$, we similarly have,

$$S_{n,2} = 2b_n^2(1-a)^2\Big(\frac{2}{n}\Big)^{2-2a}\left[\sum_{k=1}^{(n-1)/2}\frac{1}{k^{2a}} - 1\right],$$

and we apply Lemma 3.2 in three cases.

For $a \in (0,1/2)$,

$$\begin{aligned}
S_{n,2} &= 2b_n^2(1-a)^2\Big(\frac{2}{n}\Big)^{2-2a}\left[\frac{[(n-1)/2]^{1-2a}}{1-2a} + O(1)\right]\\
&= \frac{2(1-a)^2}{1-2a}b_n^2\Big(\frac{2}{n}\Big)\Big(\frac{2}{n}\Big)^{1-2a}\left[[(n-1)/2]^{1-2a} + O(1)\right]\\
&= \Big(\frac{4(1-a)^2}{1-2a}\Big)\frac{b_n^2}{n} + O\Big(\frac{b_n^2}{n^{2-2a}}\Big).
\end{aligned}$$

For $a = 1/2$,

$$S_{n,2} = \frac{b_n^2}{n}\left[\sum_{k=1}^{(n-1)/2}\frac{1}{k} - 1\right] = \frac{b_n^2\log n}{n} + O\Big(\frac{b_n^2}{n}\Big).$$

And for $a \in (1/2, 1)$,

$$\begin{aligned} S_{n,2} &= 2b_n^2(1-a)^2\Big(\frac{2}{n}\Big)^{2-2a}\Big[\zeta(2a) + O\Big(\frac{1}{n^{2a-1}}\Big)\Big] \\ &= 2^{3-2a}(1-a)^2\zeta(2a)\frac{b_n^2}{n^{2-2a}} + O\Big(\frac{b_n^2}{n}\Big), \end{aligned}$$

as claimed. □

Proof of Theorem 3.4. In view of the discussion preceding the theorem, if $G_n \sim \mathrm{SW}(a, b_n)$, the probability that it is symmetric can be bounded above as,

$$\begin{aligned} \Pr(G_n \text{ is symmetric}) &= \Pr(D(G_n) = 0) \\ &\le \sum_{\pi \neq \mathsf{id}} \Pr(D_\pi(G_n) = 0) \\ &= \sum_{\pi \neq \mathsf{id}} \Pr\Big(\max_{u \in V} D_\pi(u) = 0\Big), \end{aligned}$$

and defining, for any $\pi \neq \mathsf{id}$,

$$Z_\pi = \sum_{u : u \neq \pi(u)} D_\pi(u),$$

we have,

$$\Pr(G_n \text{ is symmetric}) \le \sum_{\pi \neq \mathsf{id}} \Pr(Z_\pi = 0). \tag{14}$$

To further bound the probability that $Z_\pi = 0$, we will use Proposition 3.5. To that end, first observe that, after ignoring the first term in (9), we have, for any π and any u such that $u \neq \pi(u)$,

$$E[D_\pi(u)] \ge \sum_{v \neq u, \pi(u)} \Pr\big((u, \pi^{-1}(v)) \in E,\ (\pi(u), v) \notin E\big).$$

Under the assumptions that $u \neq \pi(u)$ and $v \neq u, \pi(u)$, the two events in the last probability above always refer to two distinct edges, so they are independent, and hence,

$$E[D_\pi(u)] \ge \sum_{v \neq u, \pi(u)} \Pr\big((u, \pi^{-1}(v)) \in E\big)\big[1 - \Pr\big((\pi(u), v) \in E\big)\big]. \tag{15}$$

Each term in the last sum is of the form $p_n(k)[1 - p_n(k')]$ for some k, k'. Therefore, since $p_n(k)$ is decreasing in k for each n, for odd n, $E[D_\pi(u)]$ is bounded below by,

$$[1 - p_n(2)] \sum_{v \neq u, \pi(u)} \Pr\big((u, \pi^{-1}(v)) \in E\big) \geq [1 - p_n(2)][S_{n,1} - p_n(2)], \tag{16}$$

with $S_{n,1}$ defined in Lemma 3.3. Note that in the sum that appears in (15) and in (16) we ignore terms that correspond to edges between nodes at distance $k = 1$, and only sum over pairs at distance k between 2 and $(n-1)/2$. So, by the result of the lemma, we have,

$$\begin{aligned} E[D_\pi(u)] &\geq [1 - p_n(2)]\Big[2b_n + O\Big(\frac{b_n}{n^{1-a}}\Big) - p_n(2)\Big] \\ &= \Big[1 - O\Big(\frac{b_n}{n^{1-a}}\Big)\Big]\Big[2b_n + O\Big(\frac{b_n}{n^{1-a}}\Big)\Big] \\ &= 2b_n[1 + o(1)], \end{aligned}$$

since $b_n = o(n^{1-a})$. A similar computation shows that the same result holds for even n. And letting $d(\pi)$ denote the degree of a permutation π, i.e., the number of u such that $\pi(u) \neq u$, we have, by the above bound and the definition of Z_π, that:

$$E(Z_\pi) \geq 2d(\pi)b_n[1 + o(1)]. \tag{17}$$

Recall that all $D_\pi(u)$ and Z_π can be expressed as functions of the independent Bernoulli random variables introduced in the proof of Proposition 3.1. From the expression in (9) it is clear that, changing the value of any one of the edges corresponding to these random variables can only change the value of $D_\pi(u)$ by at most 2, and adding or deleting any such edge only affects at most four of the terms in the sum Z_π. Therefore, Z_π considered as a function of these Bernoulli random variables satisfies the bounded difference property (10) with $c = 8$.

For the variance σ^2 we note that each of the $d(\pi)$ many terms in the sum defining Z_π depends on $(n - 3)$ of the corresponding binary variables. Therefore, since some of them may influence $D_\pi(u)$ for more than one u, we can bound, for odd n,

$$\sigma^2 \leq 8^2 \times 2 \times d(\pi) \times \sum_{k=2}^{(n-1)/2} p_n(k)[1 - p_n(k)] = \bar{\sigma}^2 := 64d(\pi)[S_{n,1} - S_{n,2}],$$

where $S_{n,2}$ is defined in Lemma 3.3. By the result of the lemma, under the present assumptions we have $S_{n,2} = o(b_n)$ for all $a \in (0,1)$, and hence,

$$\sigma^2 \le \bar{\sigma}^2 := 128 d(\pi) b_n [1 + o(1)]. \tag{18}$$

On the other hand, for each u in the definition of Z_π, considering the influence on $D_\pi(u)$ of only those $v \neq \pi(u)$ that lie on the "right" of $\pi(u)$ on the circle (in order to avoid double-counting edges), and arguing exactly as above, we obtain a corresponding lower bound,

$$\sigma^2 \ge \underline{\sigma}^2 := 64 d(\pi) b_n [1 + o(1)]. \tag{19}$$

Analogous computations show that the bounds (18) and (19) also hold for even n, and we are now in a position to apply Proposition 3.5.

Let N be large enough so that, for all $n \ge N$, we have $120 d(\pi) b_n \le \bar{\sigma}^2 \le 132 d(\pi) b_n$ by (18), $\underline{\sigma}^2 \ge 60 d(\pi) b_n$ by (19), and the lower bound in (17) is at least $d(\pi) b_n$. Note that N can be chosen independently of π, since the $o(1)$ terms in each of these bounds do not depend on π.

Let $s = \lambda \bar{\sigma}$, for a fixed $\lambda \in (0, 1/132)$. Then, for any $\pi \neq \mathsf{id}$ and all $n \ge N$, we have, by the choice of λ and the upper bound on $\bar{\sigma}^2$,

$$\begin{aligned} \Pr(Z_\pi = 0) &\le \Pr(Z_\pi < d(\pi) b_n (1 - 132\lambda)) \\ &\le \Pr(Z_\pi < d(\pi) b_n - \lambda \bar{\sigma}^2) \\ &= \Pr(Z_\pi < d(\pi) b_n - s \bar{\sigma}). \end{aligned}$$

And by the definition of $\bar{\sigma}^2$ and the lower bound on $E(Z_\pi)$,

$$\Pr(Z_\pi = 0) \le \Pr(Z_\pi < d(\pi) b_n - s\sigma) \le \Pr(Z_\pi < E(Z_\pi) - s\sigma).$$

Therefore, by the bound in Proposition 3.5, we obtain,

$$\begin{aligned} \Pr(Z_\pi = 0) &\le \Pr(|Z_\pi - E(Z_\pi)| > s\sigma) \\ &\le 2e^{-s^2/4} = 2e^{-\lambda^2 \bar{\sigma}^2/4} \le 2e^{-33\lambda^2 d(\pi) b_n}, \end{aligned} \tag{20}$$

as long as,

$$\lambda < \frac{1}{8} < \frac{\sqrt{5}}{4\sqrt{11}} = \frac{\sqrt{60 d(\pi) b_n}}{4\sqrt{132 d(\pi) b_n}} \le \frac{\underline{\sigma}}{4\bar{\sigma}} \le \frac{\sigma}{4\bar{\sigma}},$$

which implies $s < \sigma/4 = 2\sigma/c$.

Finally, we will sum all the probabilities in (20) as in (14). Since there are no more than $n!/(n-d)! \le n^d$ permutations that fix $(n-d)$ vertices, we have, that,

$$\begin{aligned}
\Pr(G_n \text{ is symmetric}) &\le \sum_{\pi\neq\mathsf{id}} \Pr(Z_\pi = 0) \\
&\le 2\sum_{d=1}^{n} n^d e^{-33\lambda^2 d b_n} \\
&= 2\sum_{d=1}^{n} e^{d[\log n - 33\lambda^2 b_n]},
\end{aligned}$$

and since $b_n/\log n \to \infty$, the right-hand side above is $O(n^{-t})$, for any $t > 0$, as claimed. □

Proof of Theorem 4.2. In the notation of the proof of Proposition 3.1, the edges connecting each node on the circle is described by a collection of independent Bernoulli random variables. Therefore, considering all n nodes and accounting for double-counting, when n is odd (the case when n is even is similar),

$$H(G_n) = \frac{n}{2} H(\{X_k, Y_k\}) = n \sum_{k=2}^{(n-1)/2} h(p_n(k)).$$

A weaker version of Lemma 2.3 is that, $h(p) = p\log(1/p) + p - O(p^2)$, for small p, where the error term is between 0 and p^2 for $p < 1/2$. Therefore, taking n large enough so that all $p_n(k) < 1/2$, we have that,

$$\begin{aligned}
H(G_n) = &-n \sum_{k=2}^{(n-1)/2} p_n(k)\log p_n(k) \\
&+ n \sum_{k=2}^{(n-1)/2} p_n(k) - n \sum_{k=2}^{(n-1)/2} \Delta_{n,k} p_n(k)^2 \\
= &\; anc_n \sum_{k=2}^{(n-1)/2} \frac{\log k}{k^a} - \frac{n}{2}(\log c_n) S_{n,1} + \frac{n}{2} S_{n,1} - \Delta_n \frac{n}{2} S_{n,2}, \qquad (21)
\end{aligned}$$

for appropriate constants $\Delta_{n,k}, \Delta_n$ in $[0,1]$, where $S_{n,1}$ and $S_{n,2}$ are defined as in Lemma 3.3.

Using Lemma 4.1, the first term in (21) can be expressed as,

$$
\begin{aligned}
&a(1-a)2^{1-a}n^a b_n \times \\
&\quad\times\left[\frac{\log((n-1)/2)}{(1-a)((n-1)/2)^{a-1}} - \frac{1}{(1-a)^2((n-1)/2)^{a-1}} + O(1)\right] \\
&\quad= n^a b_n\left[a(n-1)^{1-a}\log((n-1)/2) - a(n-1)^{1-a}\frac{1}{(1-a)} + O(1)\right] \\
&\quad= n^a b_n\left[an^{1-a}\log n - an^{1-a}\Big(\frac{1+\log 2}{1-a}\Big) + O(1)\right] \\
&\quad= ab_n n\log n - ab_n n\Big(\frac{1+\log 2}{1-a}\Big) + o(n). \qquad (22)
\end{aligned}
$$

The sum of the second and third terms in (21), using Lemma 3.3, is,

$$
\begin{aligned}
&-\frac{n}{2}\Big[\log\Big((1-a)2^{1-a}\Big) + \log b_n - (1-a)\log n - 1\Big]\times \\
&\quad\times\left[2b_n + O\Big(\frac{b_n}{n^{1-a}}\Big)\right] \\
&= n^a b_n\Big[(1-a)n^{1-a}\log n + n^{1-a} - n^{1-a}\log b_n \\
&\qquad - n^{1-a}\log\Big((1-a)2^{1-a}\Big)\Big] + o(n\log n) \\
&= (1-a)b_n n\log n - b_n n\log b_n \\
&\qquad - b_n n\log\Big((1-a)2^{1-a}\Big) + b_n n + o(nb_n). \qquad (23)
\end{aligned}
$$

And the last term in (21), by Lemma 3.3, is $o(nb_n)$ for all $a \in (0,1)$. Substituting this together with (22) and (23) into (21), yields,

$$
nb_n\left\{\log n - \log b_n - \left[a\Big(\frac{1+\log 2}{1-a}\Big) + \log\Big((1-a)2^{1-a}\Big) - 1\right] + o(1)\right\},
$$

as required. □

Proof of Proposition 4.6. We give the proof for odd n; the case of even n is similar.

Let $W_n(i)$ denote the (random) degree of node i in G_n, so that $W_n(1) = W_n$ as in the proof of Proposition 3.1. We will apply Proposition 3.5 to bound the tails of W_n. Note that, $E(W_n) = 2 + S_{n,1} = 2b_n + 2 + o(1)$, by Lemma 3.3. Also, as a function of the Bernoulli variables $\{X_k, Y_k\}$ introduced in the proof of Proposition Proposition 3.1, W_n satisfies the assumptions of

Proposition 3.5 with $c = 1$. And in this case, in the notation of Lemma 3.3, the variance σ^2 is,

$$\sigma^2 = 2 \sum_{k=2}^{(n-1)/2} p_n(k)(1 - p_n(k)) = S_{n,1} - S_{n,2} = b_n[2 + o(1)].$$

Now consider N large enough such that, for all $n \geq N$,

$$2b_n \leq E(W_n) \leq 3b_n, \qquad \text{and} \qquad b_n \leq \sigma^2 \leq 3b_n.$$

Then by the union bound and symmetry, we have that, for any $\lambda \in (0, 1/2)$ and $n \geq N$,

$$\begin{aligned}
\Pr\Big(\max_{1\leq i\leq n} W_n(i) > 2(\lambda+2)b_n\Big) &\leq n \Pr\big(W_n > 2(\lambda+2)b_n\big) \\
&\leq n \Pr\big(W_n > E(W_n) + (2\lambda+1)b_n\big) \\
&\leq n \Pr\big(|W_n - E(W_n)| > (2\lambda+1)b_n\big) \\
&= n \Pr\big(|W_n - E(W_n)| > s\sigma\big),
\end{aligned}$$

where we took $s = (2\lambda + 1)b_n/\sigma$. Since $0 < s < 2\sigma$ for $n \geq N$ by our assumptions, we can apply Proposition 3.5 with $\lambda = 1/4$ to obtain that,

$$\begin{aligned}
\Pr\Big(\max_{1\leq i\leq n} W_n(i) > 9b_n/2\Big) &\leq 2n \exp(-s^2/4) \\
&= 2 \exp\Big\{\log n - \frac{9b_n^2}{16\sigma^2}\Big\} \\
&\leq 2 \exp\Big\{\log n - \frac{3b_n}{16}\Big\},
\end{aligned}$$

and since $b_n/\log n \to \infty$ as $n \to \infty$, the result follows. □

References

[1] E. Abbe. Graph compression: The effect of clusters. In *Proc. of the Fifty-Fourth Annual Allerton Conference*, 2016.

[2] D.J. Aldous and N. Ross. Entropy of some models of sparse random graphs with vertex-names. *Probability in the Engineering and Informational Sciences*, 28:145–168, April 2014.

[3] N. Alon, J.-H. Kim, and J. Spencer. Nearly perfect matchings in regular simple hypergraphs. *Israel Journal of Mathematics*, 100(1):171–187, 1997.

[4] N. Alon and J.H. Spencer. *The probabilistic method.* John Wiley & Sons, New York, 2004.

[5] T.M. Apostol. *Introduction to analytic number theory.* Springer-Verlag, New York, 1976.

[6] T.M. Apostol. An elementary view of Euler's summation formula. *Amer. Math. Monthly*, 106(5):409–418, May 1999.

[7] A.R. Asadi, E. Abbe, and S. Verdú. Compressing data on graphs with clusters. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 1583–1587, June 2017.

[8] A. L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[9] A.D. Barbour and P. Hall. On the rate of Poisson convergence. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 95, pages 473–480, 1984.

[10] S. Basu and L. R. Varshney. Universal and succinct source coding of deep neural networks. *arXiv*: 1804.02800, 2018.

[11] M. Besta and T. Hoefler. Survey and taxonomy of lossless graph compression and space-efficient graph representations. *arXiv*: 1806.01799, 2018.

[12] B. Bollobás. *Random graphs.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2001.

[13] F. Chierichetti, R. Kumar, S. Lattanzi, M. Mitzenmacher, A. Panconesi, and P. Raghavan. On compressing social networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 219–228, New York, NY, USA, 2009. ACM.

[14] F. Chierichetti, R. Kumar, S. Lattanzi, A. Panconesi, and P. Raghavan. Models for the compressible web. *SIAM J. Comput.*, 42(5):1777–1802, 2013.

[15] Y. Choi and W. Szpankowski. Compression of graphical structures: Fundamental limits, algorithms, and experiments. *IEEE Trans. Inform. Theory*, 58(2):620–638, February 2012.

[16] T.M. Cover and J.A. Thomas. *Elements of information theory.* Wiley-Interscience, 2006.

[17] P. Delgosha and V. Anantharam. Universal lossless compression of graphical data. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 1578–1582, June 2017.

[18] P. Delgosha and V. Anantharam. Universal lossless compression of graphical data. *arXiv*: 1909.09844, 2019.

[19] D. Easley and J. Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, Cambridge, UK, 2010.

[20] P. Erdős and A. Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

[21] Z. Golebiewski, A. Magner, and W. Szpankowski. Entropy and optimal compression of some general trees. *ACM Transaction on Algorithms*, 15, 2018.

[22] J. C. Kieffer J. Zhang, E.-H. Yang. A universal grammar-based code for lossless compression of binary trees. *IEEE Transactions on Information Theory*, 60:1373–1386, 2014.

[23] J. C. Kieffer, E. H. Yang, and W. Szpankowski. Structural complexity of random binary trees. In *2009 IEEE International Symposium on Information Theory*, pages 635–639, June 2009.

[24] J.H. Kim, B. Sudakov, and V.H. Vu. On the asymmetry of random regular graphs and random graphs. *Random Structures & Algorithms*, 21(3-4):216–224, October 2002.

[25] J.M. Kleinberg. Navigation in a small world. *Nature*, 406(6798):845–845, 2000.

[26] I. Kontoyiannis. Second-order noiseless source coding theorems. *IEEE Trans. Inform. Theory*, 43(4):1339–1341, July 1997.

[27] I. Kontoyiannis and S. Verdú. Optimal lossless data compression: Non-asymptotics and asymptotics. *IEEE Trans. Inform. Theory*, 60(2):777–795, February 2014.

[28] T. Łuczak, A. Magner, and W. Szpankowski. Asymmetry and structural information in preferential attachment graphs. *Random Structures &Algorithms*, 55(3):696–718, 2019.

[29] T. Luczak, A. Magner, and W. Szpankowski. Compression of preferential attachment graphs. In *IEEE International Symposium on Information Theory*, 2019.

[30] A. Magner, J. Sreedharan, A. Grama, and W. Szpankowski. Inferring temporal information from a snapshot of a dynamic network. *Nature Scientific Reports*, 9:3057–3062, 2019.

[31] A. Magner, K. Turowski, and W. Szpankowski. Lossless compression of binary trees with correlated vertex names. *IEEE Trans. Inform. Theory*, 64:6070–6080, 2018.

[32] S. Maneth and F. Peternek. Grammar-based graph compression. *Information Systems*, 76, 2018.

[33] M. Mohri, M. Riley, and A. T. Suresh. Automata and graph compression. In *IEEE International Symposium on Information Theory*, 2015.

[34] M. Naor. Succinct representation of general unlabelled graphs. *Discrete Applied Mathematics*, 28:303–307, 1990.

[35] M.E.J. Newman. Models of the small world. *J. Stat. Phys.*, 101(3-4):819–841, 2000.

[36] M.E.J. Newman. *Networks*. Oxford university press, Oxford, UK, 2018.

[37] M.E.J. Newman and D.J. Watts. Renormalization group analysis of the small-world network model. *Physics Letters A*, 263(4-6): 341-346, 1999.

[38] L. Peshkin. Structure induction by lossless graph compression. In *Data Compression Conference*, 2007.

[39] M. Sauerhoff. On the entropy of models for the web graph. *In submission*, 2016.

[40] V. Strassen. Asymptotische Abschätzungen in Shannons Informationstheorie. In *Trans. Third Prague Conf. Information Theory, Statist. Decision Functions, Random Processes (Liblice, 1962)*, pages 689–723. Publ. House Czech. Acad. Sci., Prague, 1964.

[41] L. Tray and I. Kontoyiannis. The feature-first block model. *arXiv*: 2105.13762, 2021.

[42] G. Turán. On the succinct representation of graphs. *Discrete Applied Mathematics*, 8:289–294, 1984.

[43] R. van der Hofstad. *Random graphs and complex networks*, volume 1. Cambridge University Press, Cambridge, UK, 2016.

[44] D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[45] J. Zhang, E.-H. Yang, and J.C. Kieffer. A universal grammar-based code for lossless compression of binary trees. *IEEE Trans. Inform. Theory*, 60(3):1373–1386, 2014.

Ioannis Kontoyiannis
Statistical Laboratory
Centre for Mathematical Sciences
University of Cambridge
Wilberforce Road
Cambridge CB3 0WB
United Kingdom
E-mail address: yiannis@maths.cam.ac.uk

Yi Heng Lim
Department of Engineering
University of Cambridge
Trumpington Street
Cambridge CB2 1PZ
United Kingdom
E-mail address: limjohnyh@gmail.com

Katia Papakonstantinopoulou
Department of Informatics
Athens University of Economics and Business
Patission 76
Athens 10434
Greece
E-mail address: katia@aueb.gr

Wojtek Szpankowski
Department of Computer Science
Purdue University
305 N. University Street
West Lafayette, IN 47907-2107
U.S.A.
E-mail address: szpan@purdue.edu