Probability in the Engineering and Informational Sciences, **28**, 2014, 145–168.
doi:10.1017/S0269964813000399

ENTROPY OF SOME MODELS OF SPARSE RANDOM GRAPHS WITH VERTEX-NAMES

DAVID J. ALDOUS
Department of Statistics
367 Evans Hall no. 3860
U.C. Berkeley, CA 94720
E-mail: aldous@stat.berkeley.edu; www.stat.berkeley.edu/users/aldous

NATHAN ROSS
Current Address: Department of Mathematics and Statistics
Richard Berry Building, University of Melbourne
VIC 3010
Australia
E-mail: nathan.ross@unimelb.edu.au; www.ms.unimelb.edu.au/∼rossn1

Consider the setting of sparse graphs on N vertices, where the vertices have distinct "names", which are strings of length $O(\log N)$ from a fixed finite alphabet. For many natural probability models, the entropy grows as $cN \log N$ for some model-dependent rate constant c. The mathematical content of this paper is the (often easy) calculation of c for a variety of models, in particular for various standard random graph models adapted to this setting. Our broader purpose is to publicize this particular setting as a natural setting for future theoretical study of data compression for graphs, and (more speculatively) for discussion of unorganized versus organized complexity.

1. INTRODUCTION

The concept *entropy* arises across a broad range of topics within the mathematical sciences, with different nuances and applications. There is a substantial literature (see Section 2.2) on topics linking entropy and graphs, but our focus seems different from these. In this paper, we use the word only with its most elementary meaning: for any probability distribution $\mathbf{p} = (p_s)$ on any finite set S, its *entropy* is the number

$$\mathrm{ent}(\mathbf{p}) = -\sum_s p_s \log p_s. \tag{1.1}$$

For an S-valued random variable X we abuse notation by writing $\mathrm{ent}(X)$ for the entropy of the distribution of X.

Consider an N-vertex undirected graph. Instead of the usual conventions about vertex-labels (unlabeled; labeled by a finite set independent of N; labeled by integers $1, \ldots, N$) our convention is that there is a fixed (i.e., independent of N) alphabet $\mathbf{A}$ of size $2 \le A < \infty$ and

© Cambridge University Press 2014 0269-9648/14 $25.00

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

that each vertex has a different "name", which is a length-$O(\log N)$ string $\mathbf{a} = (a_1, \ldots, a_m)$ of letters from $\mathbf{A}$.

We will consider probability distributions over such graphs-with-vertex-names, in the $N \to \infty$ "sparse graph limit" where the number of edges is $O(N)$. In other words, we study random graphs-with-vertex-names $\mathcal{G}_N$ whose average degree is $O(1)$. In this particular context (see Section 2.1 for discussion), one expects that the entropy should grow as

$$\text{ent}(\mathcal{G}_N) \sim cN \log N, \tag{1.2}$$

where c is thereby interpretable as an "entropy rate". Note the intriguing curiosity that the numerical value of the entropy rate c does not depend on the base of the logarithms, because there is a "log" on both sides of the definition (1.2), and indeed we will mostly avoid specifying the base.

In Section 4, we define and analyze a variety of models for which calculation of entropy rates is straightforward. In Section 5,we study one more complicated model which is conceptually more interesting in our framework and for which calculating the entropy rate is less straightforward. This is the *mathematical* content of the paper. Our motivation for studying entropy in this specific setting is discussed verbally in Section 2, and this discussion is the main *conceptual* contribution of the paper. The discussion is independent of the subsequent mathematics but may be helpful in formulating interesting probability models for future study. Section 3 gives some (elementary) technical background. Section 6 contains final remarks and open problems.

2. REMARKS ON DATA COMPRESSION FOR GRAPHICAL STRUCTURES

The well-known textbook [9] provides an account of the classical Shannon setting of data compression for sequential data, motivated by English language text modeled as a stationary random sequence. What is the analog for graph-structured data?

This is plainly a vague question. Real-world data rarely consists only of the abstract mathematical structure—unlabeled vertices and edges—of a graph; typically a considerable amount of context-dependent extra information is also present. Two illustrative examples:

(i) Phylogenetic trees on species: here part of the data is the names of the species and the names of clades;

(ii) Road networks: here part of the data is the names or numbers of the roads and some indication of the locations where roads meet.

Our setting is designed as one simple abstraction of "extra information", in which the (only) extra information is the "names" attached to vertices. Note that in many examples one expects some association between the names and the graph structure, in that the names of two vertices which are adjacent will on average be "more similar" in some sense than the names of two non-adjacent vertices. This is very clear in the phylogenetic tree example, because of the genus–species naming convention. So, when we study toy probability models later, we want models featuring such association.

Let us remind the reader of two fundamental facts from information theory [9].

(a) In the general setting (1.1), there exists a coding (e.g. Huffman code) $f_{\mathbf{p}} : S \to \mathbf{B}$ such that, for X with distribution $\mathbf{p}$,

$$\text{ent}(\mathbf{p}) \le \mathbb{E}\ \text{len}(f_{\mathbf{p}}(X)) \le \text{ent}(\mathbf{p}) + 1$$

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

and no coding can improve on the lower bound. Here $\mathbf{B}$ denotes the set of finite binary strings $\mathbf{b} = b_1 b_2 \cdots b_m$ and $\text{len}(\mathbf{b}) = m$ denotes the length of a string and entropy is computed to base 2. Recall that a coding is just a $1-1$ function.

(b) In the classical Shannon setting, one considers a stationary ergodic sequence $\mathbf{X} = (X_i)$ with values in a finite alphabet. Such a sequence has an *entropy rate*

$$H := \lim_{k\to\infty} k^{-1}\text{ent}(X_1, \ldots, X_k).$$

Moreover, there exist coding functions f (e.g., Lempel–Ziv) which are *universal* in the sense that for every such stationary ergodic sequence,

$$\lim_{m\to\infty} m^{-1}\mathbb{E}\ \text{len}(f(X_1, \ldots, X_m)) = H.$$

The important distinction is that in (a) the coding function $f_{\mathbf{p}}$ depends on the distribution of $\mathbf{X}$ but in (b) the coding function f is a function on finite sequences which does not depend on the distribution of $\mathbf{X}$.

In our setting of graphs with vertex-names, we can in principle apply (a), but it will typically be very unrealistic to imagine that observed real-world data is a realization from some *known* probability distribution on such graphs. At the other extreme, for many reasons one cannot expect there to exist, in our setting, "universal" algorithms analogous to (b). For instance, the vertex-names $(\mathbf{a}, \mathbf{a}^*)$ across some edges might be related by a deterministic cryptographic function. Also note it is difficult to imagine a definition analogous to "stationary" in our setting. So it seems necessary to rely on heuristic algorithms for compression, where *heuristic* means only that there is no good theoretical guarantee on compressed length. One could of course compare different heuristic algorithms at an empirical level by testing them on real-world data. As a theoretical complement, one could test an algorithm's efficiency by trying to prove that, for some wide range of qualitatively different probability models for $\mathcal{G}_N$, the algorithm behaves optimally in the sense of compressing to mean length $(c + o(1))N \log N$ where c is the entropy rate (1.2). And the contribution of this paper is to provide a collection of probability models for which we know the numerical value of c.

2.1. Remarks on the Technical Setup

The discussion above did not involve two extra assumptions made in Section 1, that the graphs are sparse and that the length of names is $O(\log N)$ (note the length must be at least order $\log N$ to allow the names to be distinct). These extra assumptions create a more focused setting for data compression that is mathematically interesting for two reasons. If the entropies of the two structural components—the unlabeled graph, and the set of names—were of different orders, then only the larger one would be important; but these extra assumptions make both entropies be of the same order, $N \log N$. So both of these two structural components and their association become relevant for compression. A second, more technical, reason is that natural models of sparse random graphs $\mathcal{G}_n$ invariably have a well-defined limit $\mathcal{G}_\infty$ in the sense of *local weak convergence* [3,4] of unlabeled graphs, and the limit $\mathcal{G}_\infty$ automatically has a property *unimodularity* directly analogous to stationarity for random sequences. This addresses part of the "difficult to imagine a definition analogous to stationary in our setting" issue raised above, but it remains difficult to extend this notion to encompass the vertex-names.

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

2.2. Related Work

We have given a verbal argument that the Section 1 setting of sparse graphs with vertex-names is a worthwhile setting for future theoretical study of data compression in graphical structures. It is perhaps surprising that this precise setting has apparently not been considered previously. The large literature on what is called "graph entropy", recently surveyed in [10], deals with statistics of a single unlabeled graph, which is quite different from our setting. Data compression for graphs with a fixed alphabet is considered in [12]. In a different direction, the case of *sequences* of length N with increasing-sized alphabets is considered in [13,15]. Closest to our topic is [7], discussing entropy and explicit compression algorithms for Erdős–Rényi random graphs. But all of this literature deals with settings that seem "more mathematical" than ours, in the sense of being less closely related to compression of real-world graphical structures involving extra information.

On the applied side, there is considerable discussion of heuristic compression algorithms designed to exploit expected features of graphs arising in particular contexts, for instance WWW links [5] and social networks [6]. What we proposed in the previous section as future research is to try to bridge the gap between that work and mathematical theory by seeking to devise and study general purpose heuristic algorithms.

On a more speculative note, we have a lot of sympathy with the view expressed by John Doyle and co-authors [1], who argue that the "organized complexity" one sees in real world evolved biological and technological networks is essentially different from the "disorganized complexity" produced by probability models of random graphs. At first sight it is unclear how one might try to demonstrate this distinction at some statistical level. But producing a heuristic algorithm that codes some class of real-world networks to lengths smaller than the entropy of typical probability models of such networks would be rather convincing.

3. A LITTLE TECHNICAL BACKGROUND

Here are some elementary facts [9] about entropy, in the setting (1.1) of a S-valued r.v. X, which we will use without comment.

$$\begin{aligned} \mathrm{ent}(X) &\le \log |S| \\ \mathrm{ent}(X, Y) &\le \mathrm{ent}(X) + \mathrm{ent}(Y) \\ \mathrm{ent}(X) &\ge \mathrm{ent}(h(X)) \text{ for any } h : S \to S'. \end{aligned}$$

These inequalities are equalities if and only if, respectively,

X has uniform distribution on S

X and Y are independent

h is $1-1$ on the essential range of X.

Also, if $\bar{\theta} = \sum_s q_s \theta_s$, where $q = (q_s)$ and each θ_s is a probability distribution, then

$$\mathrm{ent}(\bar{\theta}) \le \mathrm{ent}(q) + \sum_s q_s \, \mathrm{ent}(\theta_s) \tag{3.1}$$

with equality if and only if the supports of the θ_s are essentially disjoint. In random variable notation,

$$\mathrm{ent}(X) = \mathrm{ent}(f(X)) + \mathbb{E}\mathrm{ent}(X|f(X)) \tag{3.2}$$

where the random variable $\mathrm{ent}(X|Y)$ denotes entropy of the conditional distribution. (Note this is what a probabilist would call "conditional entropy", although information theorists

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

use that phrase to mean $\mathbb{E}\,\mathrm{ent}(X|Y)$). Write

$$\mathcal{E}(p) = -p\log p - (1-p)\log(1-p)$$

for the entropy of the Bernoulli(p) distribution. We will often use the fact

$$\mathcal{E}(p) \sim p\log\tfrac{1}{p} \text{ as } p \downarrow 0. \tag{3.3}$$

We will also often use the following three basic crude estimates. First,

$$\text{if } K_m \to \infty \text{ and } \tfrac{K_m}{m} \to 0 \text{ then } \log\binom{m}{K_m} \sim K_m \log\tfrac{m}{K_m}. \tag{3.4}$$

Second, for $X(n,p)$ with Binomial(n,p) distribution, if $0 \le x_n \le np$ and $x_n/n \to x \in [0,p]$ then

$$\log \mathbb{P}(X(n,p) \le x_n) = -n\Lambda_p(x_n/n) + \mathrm{O}(\log n), \tag{3.5}$$

where $\Lambda_p(x) := x\log(x/p) + (1-x)\log((1-x)/(1-p))$. The first-order term is standard from large deviation theory and the second-order estimate follows from finer but still easy analysis; see for example, Lemma 2.1 of [11]. Third, write $G[N,M]$ for the number of graphs on vertex-set $1,\ldots,N$ with at most M edges. It easily follows from (3.5) that

$$\text{if } \tfrac{M}{N} \to \zeta \in [0,\infty) \text{ then } \frac{\log G[N,M]}{N\log N} \to \zeta. \tag{3.6}$$

4. EASY EXAMPLES

Standard models of random graphs on vertices labeled $1,\ldots,N$ can be adapted to our setting of vertex-names in several ways. In particular, one could either

(i) rewrite the integer label in binary, that is as a binary string; or

(ii) replace the labels by distinct random strings as names.

These two schemes are illustrated in the first two examples below.

We present the results in a fixed format: a name for the model as a subsection heading, a definition of the model $\mathcal{G}_N$, typically involving parameters $\alpha,\beta,\ldots$, and a Proposition giving a formula for the entropy rate $c = c(\alpha,\beta,\ldots)$ such that

$$\mathrm{ent}(\mathcal{G}_N) \sim cN\log N \text{ as } N \to \infty.$$

Model descriptions and calculations sometimes implicitly assume N is sufficiently large.

These particular models are "easy" in the specific sense that independence of edges allows us to write down an exact expression for entropy; then calculations establish the asymptotics. We also give two general results, Lemmas 1 and 2, showing that graphs with short edges, or with similar names between connected vertices, have entropy rate zero.

4.1. Sparse Erdős–Rényi, Default Binary Names

4.1.1. Model N vertices, whose names are the integers $1,\ldots,N$ written as binary strings of length $\lceil \log_2 N \rceil$. Each of the $\binom{N}{2}$ possible edges is present independently with probability α/N, where $0 < \alpha < \infty$.

4.1.2. Entropy rate formula $c(\alpha) = \frac{\alpha}{2}$.

Proof. The entropy equals $\binom{N}{2}\mathcal{E}(\alpha/N)$; letting $N \to \infty$ and using (3.3) gives the formula.

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

4.2. Sparse Erdős–Rényi, Random A-ary Names

4.2.1. Model As above, N vertices, and each of the $\binom{N}{2}$ possible edges is present independently with probability α/N. Take $L_N \sim \beta \log_A N$ for $1 < \beta < \infty$ and take the vertex names as a uniform random choice of N distinct A-ary strings of length L_N.

4.2.2. Entropy rate formula $c(\alpha, \beta) = \beta - 1 + \frac{\alpha}{2}$.

PROOF: The entropy equals $\log \binom{A^{L_N}}{N} + \binom{N}{2}\mathcal{E}(\alpha/N)$. The first term $\sim (\beta - 1)N \log N$ by (3.4) and the second term $\sim \frac{\alpha}{2} N \log N$ as in the previous model. ■

Remark: One might have naively guessed that the formula would involve β instead of $\beta - 1$, on the grounds that the entropy of the sequence of names is $\sim \beta N \log N$, but this is the rate in a third model where a vertex name is a pair $(i, \mathbf{a})$, where $1 \le i \le N$ and $\mathbf{a}$ is the random string. This model distinction becomes more substantial for the model to be studied in Section 5.

4.3. Small Worlds Random Graph

4.3.1. Model Start with $N = n^2$ vertices arranged in an $n \times n$ discrete torus, where the name of each vertex is its coordinate-pair (i, j) written as two binary strings of lengths $\lceil \log_2 n \rceil$. Add the usual edges of the degree-4 nearest-neighbor torus graph. Fix parameters $0 < \alpha, \gamma < \infty$. For each edge (w, v) of the remaining set S of $\binom{N}{2} - 2N$ possible edges in the graph, add the edge independently with probability $p_N(||w - v||_2)$, where $p_N(r) = ar^{-\gamma}$ and $a := a_{N,\gamma}$ is chosen such that the mean degree of the graph $\mathcal{G}_N$ of these random edges $\to \alpha$ as $N \to \infty$ (see (4.3,4.4) for explicit expressions) and the Euclidean distance $||w - v||_2$ is taken using the torus convention.

4.3.2. Entropy rate formula

$$\begin{aligned} c(\alpha, \gamma) &= \alpha/2, \quad 0 < \gamma < 2 \\ &= \alpha/4, \quad \gamma = 2 \\ &= 0, \quad 2 < \gamma < \infty. \end{aligned}$$

Remark: The different cases arise because for $\gamma < 2$ the edge-lengths are order n whereas for $\gamma > 2$ they are $O(1)$.

PROOF: Write $r_{i,j} = \sqrt{i^2 + j^2}$ and $p_{i,j} = p_N(r_{i,j})$. The degree $D(v)$ of vertex v in $\mathcal{G}_N$ (assuming n is odd—the even case is only a minor modification) has mean

$$\begin{aligned} \mathbb{E}D(v) - 4 &= 4 \sum_{i,j=1}^{(n-1)/2} p_N(r_{i,j}) + 4 \sum_{i=2}^{(n-1)/2} p_N(r_{i,0}) \\ &= a \left(4 \sum_{i,j=1}^{(n-1)/2} (i^2 + j^2)^{-\gamma/2} + 4 \sum_{i=2}^{(n-1)/2} i^{-\gamma} \right). \end{aligned} \tag{4.1}$$

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

Similarly, the entropy of $\mathcal{G}_N$ is exactly

$$\mathrm{ent}(\mathcal{G}_N) = \frac{N}{2}\left(4\sum_{i,j=1}^{(n-1)/2}\mathcal{E}(p_N(r_{i,j})) + 4\sum_{i=2}^{(n-1)/2}\mathcal{E}(p_N(r_{i,0}))\right). \tag{4.2}$$

One can analyze these expressions separately in the three cases. First, consider the "critical" case $\gamma = 2$. Here, the quantity in parentheses in (4.1) is $\sim \int_1^{(n-1)/2} 2\pi r^{-1} dr \sim 2\pi \log n \sim \pi \log N$. We therefore take

$$a = a_{N,1} \sim \tfrac{\alpha}{\pi \log N} \tag{4.3}$$

so that $\mathcal{G}_N$ has mean degree $\to 4 + \alpha$. Evaluating the entropy similarly, where in the second line the "$\log a$" term is asymptotically negligible,

$$\begin{aligned}
\mathrm{ent}(\mathcal{G}_N) &\sim \frac{N}{2}\int_1^{(n-1)/2} 2\pi r\ \mathcal{E}(ar^{-2})dr \\
&\sim N\pi \int_1^{(n-1)/2} r \cdot ar^{-2} \cdot (-\log a + 2\log r)\ dr \\
&\sim 2N\pi a \int_1^{(n-1)/2} r^{-1}\log r\ dr \\
&\sim 2N\pi a \cdot \tfrac{1}{2}\log^2 n \\
&\sim \tfrac{\alpha}{4} N \log N
\end{aligned}$$

giving the asserted entropy rate formula in this case $\gamma = 2$.

In the case $\gamma < 2$, more elaborate though straightforward calculations (see the appendix) show that to have the mean degree $\to \alpha + 4$, we take

$$a = a_{N,\gamma} \sim \alpha\kappa_\gamma N^{-1+\gamma/2}; \quad \kappa_\gamma = \frac{2-\gamma}{2^{1+\gamma}\int_0^{\pi/4}\sec^{2-\gamma}(\theta)\,d\theta}. \tag{4.4}$$

and then establish the asserted entropy rate $\alpha/2$.

In the case $\gamma > 2$ the mean length of the edges of $\mathcal{G}_N$ becomes $O(1)$. One could repeat calculations for this case, but the asserted zero entropy rate follows from the more general Lemma 1 later, as explained in Section 4.6. ■

Remark: The case $\gamma < 2$ suggests a general principle that models with "long edges" should have the same entropy rates as if the edges were uniform random subject to the same degree distribution. But there seems no general formulation of such a result without explicit dependence assumption.

4.4. Edge-Probabilities Depending on Hamming Distance

We first describe a general model, then the specialization that we shall analyze.

4.4.1. General model Fix an alphabet $\mathbf{A}$ of size A. For each N choose L_N such that $N \le A^{L_N}$, and suppose $L_N \sim \beta((\log N)/(\log A))$ for some $\beta \in [1, \infty)$. Take N vertex-names as a uniform random choice of distinct length-L_N strings from $\mathbf{A}$. Write $d_H(\mathbf{a}, \mathbf{a}') = |\{i : a_i \ne a'_i\}|$ for Hamming distance between names. For each N let $\mathbf{w} = \mathbf{w}^N$ be a sequence

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

of decreasing weights $1 = w(1) \geq w(2) \geq \cdots \geq w(L_N) \geq 0$. We want the probability of an edge between vertices $(\mathbf{a}, \mathbf{a}')$ to be proportional to $w(d_H(\mathbf{a}, \mathbf{a}'))$. For each vertex $\mathbf{a}$, the expectation of the sum of $w(d_H(\mathbf{a}, \mathbf{a}'))$ over other vertices $\mathbf{a}'$ equals

$$\mu_N := \frac{N-1}{1-A^{-L_N}} \sum_{u=1}^{L_N} \binom{L_N}{u} \left(\frac{A-1}{A}\right)^u \left(\frac{1}{A}\right)^{L_N-u} w(u). \tag{4.5}$$

Fix $0 < \alpha < \infty$, and make a random graph $\mathcal{G}_N$ with mean degree α by specifying that, conditional on the set of vertex-names, each possible edge $(\mathbf{a}, \mathbf{a}')$ is present independently with probability $\alpha w(d_H(\mathbf{a}, \mathbf{a}'))/\mu_N$.

Note that in order for this model to make sense, we need $\mu_N \geq \alpha$, which is not guaranteed by the description of the model.

Intuitively, we expect that the lengths (measured by Hamming distance) of edges will be around the ℓ_N maximizing $\binom{L_N}{\ell_N} w^N(\ell_N)$, and that for all (suitably regular) choices of $\mathbf{w}^N$ with $\ell_N/L_N \to d \in [0,1]$ the entropy rate will involve $\mathbf{w}^N$ only via the limit d. Stating and proving a general such result seems messy, so we will study only the special case

$$w(u) = 1, \ 1 \leq u \leq M_N; \tag{4.6}$$
$$= 0, \ M_N < u \leq L_n$$
$$M_N/L_N \to d \in (0, 1 - \tfrac{1}{A}), \tag{4.7}$$
$$1 \leq \beta < \frac{\log A}{\Lambda_{1-1/A}(d)} \tag{4.8}$$

for $\Lambda_p(d)$ as at (3.5). Here condition (4.8) is needed, as we will see at (4.10), to make $\mu_N \to \infty$. Note that for the case $d = 0$, one could use Lemma 1 later and the accompanying conditioning argument in Section 4.6 to show that the entropy rate of $\mathcal{G}_N$ equals the rate $(\beta - 1)$ for the set of vertex-names. The opposite case $(1 - 1/A) \leq d \leq 1$ is essentially the model of Section 4.2 and the rate becomes $\beta - 1 + \alpha/2$: as expected, these rates are the $d \to 0$ and the $d \to 1 - 1/A$ limits of the rates in our formula below.

4.4.2. Entropy rate formula In the special case (4.6)–(4.8),

$$c(A, \alpha, \beta, d) = \beta - 1 + \frac{\alpha}{2}\left(1 - \frac{\beta \Lambda_{1-1/A}(d)}{\log A}\right).$$

To establish this formula, first observe that for Binomial $X(\cdot, \cdot)$ as at (3.5)

$$\mu_N = \frac{N-1}{1-A^{-L_N}} \mathbb{P}(1 \leq X(L_N, 1 - 1/A) \leq M_N), \tag{4.9}$$

and so by (3.5)

$$\frac{\log \mu_N}{\log N} \to 1 - \frac{\beta \ \Lambda_{1-1/A}(d)}{\log A}. \tag{4.10}$$

So condition (4.8) ensures that $\mu_N \to \infty$ and therefore the model makes sense. Write **Names** (to avoid overburdening the reader with symbols) for the random unordered set of vertex-names, and use (3.2) to write

$$\mathrm{ent}(\mathcal{G}_N) = \mathrm{ent}(\mathbf{Names}) + \mathbb{E}\,\mathrm{ent}(\mathcal{G}_N | \mathbf{Names}).$$

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

As in Section 4.2 the contribution to the entropy rate from the first term is $\beta - 1$. For the second term, write

$$\mathrm{ent}(\mathcal{G}_N|\mathbf{Names}) = \sum_{\mathbf{a}\neq\mathbf{a}'} \mathcal{E}\left(\tfrac{\alpha}{\mu_N}\right) \mathbb{1}_{(\mathbf{a}\in\mathbf{Names},\mathbf{a}'\in\mathbf{Names})} \mathbb{1}_{(d_H(\mathbf{a},\mathbf{a}')\le M_N)}$$

where the sum is over unordered pairs $\{\mathbf{a}, \mathbf{a}'\}$ in $\mathbf{A}^{L_N}$. Take expectation to get

$$\mathbb{E}\,\mathrm{ent}(\mathcal{G}_N|\mathbf{Names}) = \frac{\binom{N}{2}}{1 - A^{-L_N}} \sum_{u=1}^{M_N} \binom{L_N}{u} \left(\frac{A-1}{A}\right)^u \left(\frac{1}{A}\right)^{L_N-u} \mathcal{E}\left(\frac{\alpha}{\mu_N}\right).$$

But from definition (4.5) of μ_N this simplifies to $\frac{N}{2}\mu_N \mathcal{E}(\alpha/\mu_N)$, and then from (3.3)

$$\mathbb{E}\,\mathrm{ent}(\mathcal{G}_N|\mathbf{Names}) \sim \frac{\alpha N}{2} \log \mu_N.$$

Appealing to (4.10) establishes the entropy rate formula.

4.5. Non-Uniform and Uniform Random Trees

4.5.1. Model Construct a random tree $\mathcal{T}_N$ on vertices $1, \ldots, N$ as follows. Take $V_3, V_4, \ldots, V_N$ independent uniform on $\{1, \ldots, N\}$. Link vertex 2 to vertex 1. For $k = 3, 4, \ldots, N$ link vertex k to vertex $\min(k-1, V_k)$.

4.5.2. Entropy rate formula $c = 1/2$.

Proof:

$$\mathrm{ent}(\mathcal{T}_N) = \sum_{k=3}^{N} \mathrm{ent}(W_k)$$

where $W_k = \min(k-1, V_k)$ has entropy

$$\mathrm{ent}(W_k) = \tfrac{k-2}{N} \log N + \tfrac{N-k+2}{N} \log \tfrac{N}{N-k+2}$$

The sum of the first term $\sim \frac{1}{2} N \log N$ and the sum of the second term is of smaller order. ■

Remark: This tree arose in [2], where it was shown (by an indirect argument) that if one first constructs $\mathcal{T}_N$, then applies a uniform random permutation to the vertex-labels, the resulting random tree $\mathcal{T}_N^*$ is uniform on the set of all labeled trees. Cayley's formula tells us there are N^{N-2} labeled trees, so $\mathrm{ent}(\mathcal{T}_N^*) = \log N^{N-2}$and so $(\mathcal{T}_N^*)$ has entropy rate $c = 1$.

4.6. Conditions for Zero Entropy Rate

Here, we will give two complementary conditions under which the entropy rate is zero. Lemma 1 concerns the case where we start with deterministic vertex-names, and add random edges which mostly link a vertex to some of the "closest" vertices, specifically to vertices amongst the $(o(N^\varepsilon)$ for all $\varepsilon > 0)$ closest vertices. Lemma 2 concerns the case where we start with a deterministic graph on unlabeled vertices, and add random vertex-labels such that vertices linked by an edge mostly have names that differ in only $o(\log N)$ places. Note that these lemmas may then be applied conditionally. That is, if we start with a random

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

unordered set of names, and then (conditional on the set of names) add random edges in a way satisfying the assumptions of Lemma 1, then the entropy rate of the resulting $\mathcal{G}_N$ will equal the entropy rate of the original random unordered set of names. Similarly, if we start with a random graph on unlabeled vertices, then (conditional on the graph) add random names in a way satisfying the assumptions of Lemma 2, then the entropy rate of the resulting $\mathcal{G}_N$ will equal the entropy rate of the original random unlabeled graph.

LEMMA 1: *For each N, suppose we take N vertices with deterministic names (w.l.o.g. $1 \le i \le N$ written as binary strings, to fit our setup) and suppose for each i we are given an ordering $j(i,1), j(i,2), \ldots, j(i,N-1)$ of the other vertices. Say that an edge $(i, j = j(i,\ell))$ with $i < j$ has length ℓ. Consider a sequence of random graphs $\mathcal{G}_N$ whose distribution is arbitrary subject to*

(i) *The number E_N of edges satisfies $\mathbb{P}(E_N > N\beta) = o(N^{-1}\log N)$ for some constant $\beta < \infty$;*

(ii) *For some M_N such that $\log(M_N) = o(\log N)$, the r.v.*

$$X_N := \text{number of edges with length greater than } M_N$$

satisfies $\mathbb{P}(X_N > N\delta) = o(1)$ for all $\delta > 0$.

Then the entropy rate is $c = 0$.

REMARK: The lemma applies to the $\gamma > 2$ case of the "small worlds" model in Section 4.3. Take the ordering induced by the natural distance between vertices. In this case, E_N is a sum of independent indicators with $\mathbb{E}E_N \sim cN$ for some constant c. Standard concentration results (e.g. [8] Theorem 2.15) imply (i) for any $\beta > c$, and (ii) follows since for any sequence $M_N \to \infty$ we have $\mathbb{E}X_N = \mathrm{O}(NM_N^{1-\gamma/2}) = \mathrm{o}(N)$.

PROOF: We first show that the result holds with (i) replaced by

$$(\mathrm{i}')\ \mathbb{P}(E_N > N\beta) = 0,$$

and then use this modified statement to prove the lemma.

Assume now that $\mathcal{G}_N$ satisfies (i′) and (ii) and write $\mathcal{G}_N$, considered as an edge-set, as a disjoint union $\mathcal{G}'_N \cup \mathcal{G}''_N$, where $\mathcal{G}'_N$ consists of the edges of length $\le M_N$. Because $\mathcal{G}'_N$ contains at most βN edges out of a set of at most NM_N edges,

$$\begin{aligned}\mathrm{ent}(\mathcal{G}'_N) &\le \log(\beta N) + \log \binom{NM_N}{\beta N} \\ &= o(N\log N) \text{ by (3.4)}.\end{aligned}$$

Now fix $\delta > 0$ and condition on whether the number X_N of edges of $\mathcal{G}''_N$ is bigger or smaller than δN. Using (3.1) we obtain

$$\mathrm{ent}(\mathcal{G}''_N) \le \log 2 + \log G[N, \delta N] + \mathbb{P}(X_N > \delta N)\log G[N, \beta N].$$

Now $\mathbb{P}(X_N > \delta N) \to 0$ by assumption (ii), and then using (3.6) we obtain

$$\mathrm{ent}(\mathcal{G}''_N) \le (\delta + o(1))N\log N.$$

Because $\delta > 0$ is arbitrary we conclude

$$\mathrm{ent}(\mathcal{G}_N) \le \mathrm{ent}(\mathcal{G}'_N) + \mathrm{ent}(\mathcal{G}''_N) = o(N\log N).$$

Now assume that $\mathcal{G}_N$ satisfies the weaker hypotheses (i) and (ii). Defining $\widehat{\mathcal{G}_N}$ to have the conditional distribution of $\mathcal{G}_N$ given $E_N \le \beta N$, it is clear that $\widehat{\mathcal{G}_N}$ satisfies (i′). We will

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

show that it also satisfies (ii), implying (by the previous result) that the entropy rate of $\widehat{\mathcal{G}}_N$ is zero. Let $\delta > 0$. Conditioning on the event (A_N, say) that $E_N \le \beta N$,

$$\mathbb{P}(X_N > \delta N) = \mathbb{P}(X_N > \delta N | A_N)\mathbb{P}(A_N) + \mathbb{P}(X_N > \delta N | A_N^c)\mathbb{P}(A_N^c). \tag{4.11}$$

By (ii), the term on the left-hand side of (4.11) is o(1), and by (i), $\mathbb{P}(A_N^c) = \mathrm{o}(1)$, and so also $\mathbb{P}(A_N) \to 1$. Thus, $\mathbb{P}(X_N > \delta N | A_N)$ must be o(1), as desired.

To complete the proof, use (3.1) to write

$$\begin{aligned}\mathrm{ent}(\mathcal{G}_N) &\le \mathcal{E}(\mathbb{P}(A_N)) + \mathrm{ent}(\mathcal{G}_N | A_N)\mathbb{P}(A_N) + \mathrm{ent}(\mathcal{G}_N | A_N^c)\mathbb{P}(A_N^c) \\ &\le \log 2 + \mathrm{ent}(\widehat{\mathcal{G}}_N) + \mathbb{P}(A_N^c)\binom{N}{2}\log 2.\end{aligned}$$

The entropy rate of $\widehat{\mathcal{G}}_N$ is zero, and assumption (i) is exactly that $\mathbb{P}(A_N^c) = \mathrm{o}(N^{-1}\log N)$, so $\mathrm{ent}(\mathcal{G}_N) = \mathrm{o}(N \log N)$, as desired. ■

LEMMA 2: *Take a deterministic graph on N unlabeled vertices, and let c_N denote the number of components and e_N the number of edges. Construct $\mathcal{G}_N$ by assigning random distinct vertex-names $\mathbf{a}(v)$ of length $O(\log N)$ to vertices v, their distribution being arbitrary subject to*

$$\sum_{\text{edges}(\mathbf{a},\mathbf{a}')} d_H(\mathbf{a}, \mathbf{a}') = o(N \log N) \text{ in probability.}$$

If $e_N = O(N)$ and $c_N = o(N)$ then $\mathcal{G}_N$ has entropy rate zero.

PROOF: By a straightforward truncation argument we may assume there is a deterministic bound

$$\sum_{\text{edges } (\mathbf{a},\mathbf{a}')} d_H(\mathbf{a}, \mathbf{a}') \le s_N = \mathrm{o}(N \log N).$$

The name-lengths are $\le \beta \log N$ for some β. Consider first the case where there is a single component. Take an arbitrary spanning tree with arbitrary root, and write the edges of the tree in breadth-first order as $e_1, \ldots, e_{N-1}$. We can specify $\mathcal{G}_N$ by specifying first the name of the root; then for each edge $e_i = (v, v')$ directed away from the root, specify the coordinates where $\mathbf{a}(v')$ differs from $\mathbf{a}(v)$ and specify the values of $\mathbf{a}(v')$ at those coordinates. Write $\mathcal{S}$ for the random set of all these differing coordinates. Conditional on $\mathcal{S} = S$ the entropy of $\mathcal{G}_N$ is at most $(|S| + \beta \log N) \log A$, where the $\beta \log N$ term arises from the root name. So using (3.1)

$$\mathrm{ent}(\mathcal{G}_N) \le \mathrm{ent}(\mathcal{S}) + (s_N + \beta \log N) \log A.$$

With c_N components the same argument shows

$$\mathrm{ent}(\mathcal{G}_N) \le \mathrm{ent}(\mathcal{S}) + (s_N + c_N \beta \log N) \log A.$$

The second term is $\mathrm{o}(N \log N)$ by assumption, and

$$\mathrm{ent}(\mathcal{S}) \le \log\left(\sum_{i \le s_N} \binom{\beta N \log N}{i}\right) = \mathrm{o}(N \log N),$$

the final relation by, e.g. the $p = 1/2$ case of (3.5). ■

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

4.7. Summary

The reader will recognize the models in this section as standard random graph models, adapted to our setting in one of several ways. One can take a model of dynamic growth, adding one vertex at a time, and then assign the k'th vertex a name, e.g. the "default binary" or "random A-ary" used in the Erdős–Rényi models. Alternatively, as in the "Hamming distance" model, one can start with N vertices with assigned names and then add edges according to some probabilistic rule involving the names of end-vertices. Roughly speaking, for any existing random graph model where one can calculate anything, one can calculate the entropy rate for such adapted models. But this is an activity perhaps best left for future Ph.D. theses. We are more interested in models where the nature of the dependence of the graph structure and the name structure is less straightforward than starting by specifying one structure and having that influence the other. It is not so easy to devise tractable such models, but the next section shows our attempt.

5. A HYBRID MODEL

In this section, we study a model for which calculation of the entropy rate is less straightforward. It incidently reveals a connection between our setting and the more familiar setting of "graph entropy".

5.1. The Model

In outline, the graph structure is again sparse Erdős–Rényi $\mathcal{G}(N, \alpha/N)$, but we construct it inductively over vertices, and make the vertex-names copy parts of the names of previous vertices that the current vertex is linked to. Here are the details.

5.1.1. Model: Erdős–Rényi with hybrid names Take $L_N \sim \beta \log_A N$ for $1 < \beta < \infty$. Vertex 1 is given a uniform random length-L_N A-ary name. For $1 \le n \le N-1$:

> vertex $n+1$ is given an edge to each vertex $i \le n$ independently with probability α/N. Write $Q_n \ge 0$ for the number of such edges, and $\mathbf{a}^1, \ldots, \mathbf{a}^{Q_n}$ for the names of the linked vertices. Take an independent uniform random length-L_N A-ary string $\mathbf{a}^0$. Assign to vertex $n+1$ the name obtained by, independently for each coordinate $1 \le u \le L_N$, making a uniform random choice from the $Q_n + 1$ letters $a_u^0, a_u^1, \ldots, a_u^{Q_n}$.

See Figure 1. This model gives a family $(\mathcal{G}_N)$ parametrized by (A, β, α). Note that this scheme for defining "hybrid" names could be used with any sequential construction of a random graph, for instance preferential attachment models.

5.2. The Ordered Case

This model illustrates a distinction mentioned in Section 4.2. In the construction above, the nth vertex is assigned a name, say $\mathbf{a}^n$, during the construction, but in the final graph $\mathcal{G}_N$ we do not see the value of n for a vertex. The "ordered" model $(\mathcal{G}_N^{\mathrm{ord}})$ in which we do see the value of n for each vertex, by making the name be $(n, \mathbf{a}^n)$, is a different model whose analysis is conceptually more straightforward, so we will start with that model. We return to the unordered model in Section 5.6.

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

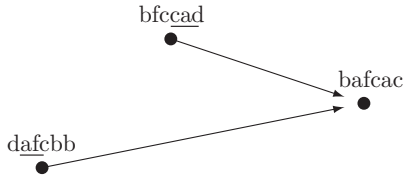


FIGURE 1. Schematic for the hybrid model. A vertex (right) arrives with some "original name" *bbdabc* and is attached to two previous vertices with names *dafcbb* and *bfccad*. The name given to the new vertex is obtained by copying for each position the letter in that position in a uniform random choice from the three names. Choosing the underlined letters gives the name shown in the figure.

5.2.1. *Entropy Rate Formula for* $(\mathcal{G}_N^{\mathrm{ord}})$

$$\frac{\alpha}{2} + \beta \sum_{k \geq 0} \frac{\alpha^k J_k(\alpha) h_A(k)}{k! \log A}, \tag{5.1}$$

where

$$J_k(\alpha) := \int_0^1 x^k e^{-\alpha x} dx$$

and the constants $h_A(k)$ are defined at (5.6).

Write $\mathcal{G}_{N,n}$ for the partial graph obtained after vertex n has been assigned its edges to previous vertices and then its name. We will show that, for deterministic $e_{N,n}$ defined at (5.8) below, as $N \to \infty$ the entropies of the conditional distributions satisfy

$$\max_{1 \leq n \leq N-1} \mathbb{E} |\mathrm{ent}(\mathcal{G}_{N,n+1} | \mathcal{G}_{N,n}) - e_{N,n}| = \mathrm{o}(\log N). \tag{5.2}$$

By the chain rule (3.2) this immediately implies

$$\mathrm{ent}(\mathcal{G}_N^{\mathrm{ord}}) - \sum_{n=1}^{N-1} e_{N,n} = \mathrm{o}(N \log N),$$

which will establish the entropy rate formula.

The key ingredient is the following technical lemma; note that the measures $\mu^{\mathbf{i}}$ below depend on the realization of $\mathcal{G}_N^{\mathrm{ord}}$ and are therefore random quantities. Write "ave" for average, and write

$$||\Theta||^{(k)} := \tfrac{1}{2} \sum_{\mathbf{a} \in \mathbf{A}^k} \left| \Theta(\mathbf{a}) - A^{-k} \right|$$

for the variation distance between a probability distribution Θ on $\mathbf{A}^k$ and the uniform distribution.

LEMMA 3: *Write* $(n, \mathbf{a}^n)$, $1 \leq n \leq N$ *for the vertex-names of* $\mathcal{G}_N^{\mathrm{ord}}$. *For each* $k \geq 1$ *and* $\boldsymbol{i} := (i_1, \ldots, i_k)$ *with* $1 \leq i_1 < \cdots < i_k \leq N$, *write* $\mu^{\boldsymbol{i}}$ *for the empirical distribution of*

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

$(a_u^{i_1}, \ldots, a_u^{i_k}), 1 \le u \le L_N$. *That is, the probability distribution on* $\mathbf{A}^k$

$$\mu^{\boldsymbol{i}}(x_1, \ldots, x_k) := L_N^{-1} \sum_{u=1}^{L_N} \mathbb{1}_{(a_u^{i_1} = x_1, \ldots, a_u^{i_k} = x_k)}.$$

Then

$$\Delta_N^{(k)} := \max_{2 \le n \le N} \mathbb{E} || \underset{1 \le i_1 < \cdots < i_k \le n}{\text{ave}} \mu^{\boldsymbol{i}} ||^{(k)} \le C \left(\frac{A^{k/2}}{\sqrt{\log N}} + \frac{k^2}{N} \right). \tag{5.3}$$

for a constant C *not depending on* k, N.

We defer the proof to Section 5.3.

Fix N and n, and consider $\text{ent}(\mathcal{G}_{N,n+1} | \mathcal{G}_{N,n})$, the entropy of the conditional distribution. Conditioning on the edges of vertex $n+1$ in $\mathcal{G}_{N,n+1}$, and using the chain rule (3.2), we find

$$\begin{aligned} \text{ent}(\mathcal{G}_{N,n+1} | \mathcal{G}_{N,n}) &= n\mathcal{E}(\alpha/N) \\ &+ \sum_{\substack{k=0,\ldots,n \\ 1 \le i_1 < \cdots < i_k \le n}} \left(\frac{\alpha}{N}\right)^k \left(1 - \frac{\alpha}{N}\right)^{n-k} \text{ent}(\mathbf{a}^{n+1} | \mathcal{G}_{N,n}, n+1 \to \{i_1, \ldots, i_k\}), \end{aligned} \tag{5.4}$$

where $n+1 \to \{i_1, \ldots, i_k\}$ denotes the event that vertex $n+1$ connects to vertices $i_1, \ldots, i_k$ and no others. The contribution to the entropy from the choice of edges is $n\mathcal{E}(\alpha/N)$, which as in previous models contributes (after summing over n) the first term $\alpha/2$ of the entropy rate formula, so in the following we need consider only the contribution from names, that is the sum in (5.4). Consider the contribution to the sum (5.4) from $k = 2$, that is on the event $\{Q_n = 2\}$ that vertex $n+1$ links to exactly two previous vertices. Conditional on these being a particular pair $1 \le i < j \le n$, with names $\mathbf{a}^i, \mathbf{a}^j$, the contribution to entropy is exactly

$$\text{ent}(\mathbf{a}^{n+1} | \mathcal{G}_{N,n}, n+1 \to \{i,j\}) = L_N \sum_{(a,a') \in \mathbf{A} \times \mathbf{A}} g_2(a,a') \, \mu^{(i,j)}(a,a'),$$

where

$$\begin{aligned} g_2(a,a') &= \mathcal{E}_A(\tfrac{A+1}{3A}, \tfrac{A+1}{3A}, \tfrac{1}{3A}, \tfrac{1}{3A}, \ldots, \tfrac{1}{3A}) \text{ if } a' \ne a \\ &= \mathcal{E}_A(\tfrac{2A+1}{3A}, \tfrac{1}{3A}, \tfrac{1}{3A}, \tfrac{1}{3A}, \ldots, \tfrac{1}{3A}) \text{ if } a' = a \end{aligned}$$

and where $\mathcal{E}_A(\mathbf{p})$ is the entropy of a distribution $\mathbf{p} = (p_1, \ldots, p_A)$. Now unconditioning on the pair (i,j), the contribution to $\text{ent}(\mathcal{G}_{N,n+1} | \mathcal{G}_{N,n})$ from the event $\{Q_n = 2\}$; that is the $k = 2$ term of the sum (5.4); equals

$$\begin{aligned} &L_N \sum_{1 \le i < j \le n} \frac{\alpha^2}{N^2} \left(1 - \tfrac{\alpha}{N}\right)^{n-2} \sum_{(a,a') \in \mathbf{A} \times \mathbf{A}} g_2(a,a') \, \mu^{(i,j)}(a,a') \\ &= \frac{L_N \alpha^2 \binom{n}{2}}{N^2} \left(1 - \tfrac{\alpha}{N}\right)^{n-2} \sum_{(a,a') \in \mathbf{A} \times \mathbf{A}} g_2(a,a') \underset{1 \le i < j \le n}{\text{ave}} \mu^{(i,j)}(a,a'). \end{aligned} \tag{5.5}$$

Lemma 3 now tells us that the sum in (5.5) differs from

$$h_A(2) := A^{-2} \sum_{(a,a') \in \mathbf{A} \times \mathbf{A}} g_2(a,a')$$

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

by at most $2g_2^*\Delta_N^{(2)}$ where $g_2^* \le \log A$ is the maximum possible value of $g_2(\cdot,\cdot)$ and $\Delta_N^{(2)}$ is as defined in Lemma 3. So to first order as $N \to \infty$, the quantity (5.5) is

$$e_{N,n,2} := \beta \log_A N \times \alpha^2 h_A(2) \tfrac{\binom{n}{2}}{N^2} \exp(-\alpha n/N),$$

with an error bounded by

$$(2\log A)\, L_N \binom{n}{2} \left(\frac{\alpha}{N}\right)^2 \left(1-\frac{\alpha}{N}\right)^{n-2} \Delta_N^{(2)}.$$

A similar argument applies to the terms in the sum (5.4) for a general number k of links. In brief, we define

$$e_{N,n,k} := \beta \log_A N \times \alpha^k h_A(k) \tfrac{\binom{n}{k}}{N^k} \exp(-\alpha n/N),$$

where

$$h_A(k) := A^{-k} \sum_{(a_1,\ldots,a_k)\in \mathbf{A}^k} \mathrm{ent}(\mathbf{p}^{[a_1,\ldots,a_k]}) \tag{5.6}$$

and where $\mathbf{p}^{[a_1,\ldots,a_k]}$ is the probability distribution $\mathbf{p}$ on $\mathbf{A}$ defined by

$$p^{[a_1,\ldots,a_k]}(a) = \frac{1 + A \times |\{i : a_i = a\}|}{(1+k)A}.$$

Also for $k = 0$ we set $h_A(0) = \log A$, the entropy of the uniform distribution on $\mathbf{A}$. Repeating the argument from the case $k = 2$, we find that (5.4) is, to first order, $\sum_{k\ge 0} e_{N,n,k}$, with error of order

$$L_N \sum_{k=0}^{n} \binom{n}{k} \left(\frac{\alpha}{N}\right)^k \left(1-\frac{\alpha}{N}\right)^{n-k} \Delta_N^{(k)}. \tag{5.7}$$

Applying Lemma 3 to bound $\Delta_N^{(k)}$ and then using simple properties of the binomial distribution yields that (5.7) is $o(\log N)$.

So, we are now in the setting of (5.2) with

$$e_{N,n} = \sum_{k\ge 0} e_{N,n,k}. \tag{5.8}$$

Because

$$\sum_{n=1}^{N-1} \tfrac{\binom{n}{k}}{N^k} \exp(-\alpha n/N) \sim \tfrac{N}{k!} J_k(\alpha)$$

calculating $\sum_{n=1}^{N-1} e_{N,n}$ gives the stated entropy rate formula.

5.3. Proof of Lemma 3

Fix N. Recall the construction of $\mathcal{G}_N^{\mathrm{ord}}$ involves an "original name process"—letters of the name of vertex n may be copies from previous names or may be from an "original name", independent uniform for different n. Consider a single coordinate, w.l.o.g. coordinate 1, of the vertex-names of $\mathcal{G}_N^{\mathrm{ord}}$. For each vertex n this is either from the original name of n or a copy of some previous vertex-name, so inductively the letter at vertex n is a copy of the letter originating at some vertex $1 \le C_1^N(n) \le n$; and similarly the letter at general coordinate

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

u is a copy from some vertex $C_u^N(n)$. Because the copying process is independent of the name origination process, it is clear that the (unconditional) distribution of each name $\mathbf{a}^n$ is uniform on length-L_N words. Moreover, it is clear that, for $1 \le i < j \le N$,

$$\begin{aligned} &\text{the two names } \mathbf{a}^i \text{ and } \mathbf{a}^j \text{ are independent uniform} \\ &\text{on the event } \{C_u^N(i) \ne C_u^N(j)\ \forall u\}. \end{aligned} \tag{5.9}$$

The proof of Lemma 3 rests upon the following lemma, whose proof we defer to the end of Section 5.4.

LEMMA 4: *For (I, J) uniform on $\{1 \le i < j \le n\}$, write $\theta_{N,n} = \mathbb{P}(C_1^N(I) = C_1^N(J))$. Then*

$$\max_{2 \le n \le N} \theta_{N,n} = O(1/N) \text{ as } N \to \infty.$$

We first use this lemma to prove Lemma 3 in the case where $k = 2$. For (I, J) as in Lemma 4,

$$\begin{aligned} \Delta_N^{(2)} &= \tfrac{1}{2} \max_{2 \le n \le N} \mathbb{E} \sum_{\mathbf{x} \in \mathbf{A}^2} \left| \mathbb{E}\left(L_N^{-1} \sum_{u=1}^{L_N} \mathbb{1}_{(a_u^I = x_1, a_u^J = x_2)} | \mathcal{G}_{N,n} \right) - A^{-2} \right| \\ &\le \tfrac{1}{2} \max_{2 \le n \le N} \sum_{\mathbf{x} \in \mathbf{A}^2} \mathbb{E} \left| L_N^{-1} \sum_{u=1}^{L_N} \mathbb{1}_{(a_u^I = x_1, a_u^J = x_2)} - A^{-2} \right|. \end{aligned} \tag{5.10}$$

By Lemma 4 and (5.9), the two names $\mathbf{a}^I, \mathbf{a}^J$ are independent uniform on $\mathbf{A}^{L_N}$ outside an event of probability $O(1/N)$. Under this event, we bound the total variation distance appearing in $\Delta_N^{(2)}$ by 1, leading to the second summand in the bound (5.3). If the two names are independent, then because the sum below has Binomial(L_N, A^{-2}) distribution with variance $< L_N A^{-2}$,

$$\mathbb{E} \left| L_N^{-1} \sum_{u=1}^{L_N} \mathbb{1}_{(a_u^I = x_1, a_u^J = x_2)} - A^{-2} \right| \le L_N^{-1/2} A^{-1}, \tag{5.11}$$

which contributes the first summand in the bound (5.3).

The proof of Lemma 3 for general k is similar. Taking $I_1, \ldots, I_k$ independent and uniform on the set $\{1 \le i_1 < \cdots < i_k \le n\}$, we have the analog of (5.10):

$$\Delta_N^{(k)} \le \tfrac{1}{2} \max_{2 \le n \le N} \sum_{\mathbf{x} \in \mathbf{A}^k} \mathbb{E} \left| L_N^{-1} \sum_{u=1}^{L_N} \mathbb{1}_{(a_u^{I_1} = x_1, \ldots, a_u^{I_k} = x_k)} - A^{-k} \right|.$$

The names $\mathbf{a}^{I_1}, \ldots, \mathbf{a}^{I_k}$ are independent outside of the "bad" event that some pair within k random vertices have the same $C_1^N(\cdot)$ value. But the probability of this bad event is bounded by $\binom{k}{2}$ times the chance for a given pair, which, after applying Lemma 4, leads to the second summand of the bound (5.3). And the upper bound for the term analogous to (5.11) becomes $L_N^{-1/2} A^{-k/2}$.

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

5.4. Structure of the Directed Sparse Erdős–Rényi Graph

In order to prove Lemma 4 and later results, we study the original name variables $C_u^N(i)$ defined at (5.9). It will first help to collect some facts about the structure of a directed sparse Erdős–Rényi random graph. Write (omitting the dependence on N)

$$\mathcal{T}_n = \{1 \le j \le n : \ \exists g \ge 0 \text{ and a path } n = v_0 > v_1 > \cdots > v_g = j \text{ in } \mathcal{G}_N\}.$$

We visualize $\mathcal{T}_n$ as the vertices of the tree of descendants of n although it may not be a tree. The next result collects two facts about the structure of $\mathcal{T}_n$ including that for large N it is a tree with high probability.

LEMMA 5: *For $m < n$ and $\mathcal{T}_n$ as above,*

(a) $\mathbb{P}(\mathcal{T}_n \cap \mathcal{T}_m \neq \emptyset) \le \frac{(\alpha e^\alpha)^2 + \alpha e^\alpha}{N}$.

(b) $\mathbb{P}(\mathcal{T}_n \textit{ is not a tree }) \le \frac{(\alpha e^\alpha)^3}{2N}$.

PROOF: First note that for $1 \le j < n \le N$ the mean number of decreasing paths from n to j of length $g \ge 1$ equals $\binom{n-j-1}{g-1}(\alpha/N)^g$. Because $n - j - 1 \le N$, this is bounded by $((\alpha)/(N))((\alpha^{g-1})/((g-1)!)$, and summing over g gives

$$\mathbb{P}(j \in \mathcal{T}_n) \le \mathbb{E}(\text{number of decreasing paths from } n \text{ to } j) \le \alpha e^\alpha / N. \tag{5.12}$$

We break the event $\mathcal{T}_n \cap \mathcal{T}_m \neq \emptyset$ into a disjoint union according to the largest element in the intersection: $\max \mathcal{T}_n \cap \mathcal{T}_m = j$ for $j = 1, \ldots, m$. Now note for $j \le m - 1$, we can write

$$\mathbb{P}(\max \mathcal{T}_n \cap \mathcal{T}_m = j) \le \mathbb{E} \sum_{x_n^j, y_m^j} \mathbb{1}_{(x_n^j \text{ is path in } \mathcal{G}_N)} \mathbb{1}_{(y_m^j \text{ is path in } \mathcal{G}_N)}, \tag{5.13}$$

where the sum is over edge-disjoint decreasing paths x_n^j from n to j and y_m^j from m to j. Since the paths are edge-disjoint, the indicators appearing in the sum (5.13) are independent and so we find

$$\begin{aligned}
\mathbb{P}(\max \mathcal{T}_n \cap \mathcal{T}_m = j) &\le \sum_{x_n^j, y_m^j} \mathbb{P}(x_n^j \text{ is path in } \mathcal{G}_N) \mathbb{P}(y_m^j \text{ is path in } \mathcal{G}_N) \\
&\le \sum_{x_n^j} \mathbb{P}(x_n^j \text{ is path in } \mathcal{G}_N) \sum_{y_m^j} \mathbb{P}(y_m^j \text{ is path in } \mathcal{G}_N) \\
&\le (\alpha e^\alpha / N)^2;
\end{aligned}$$

where the sums in the second line are over all paths from n (respectively m) to j, and the final inequality follows from (5.12). Now part (a) of the lemma follows by summing over $j < m$ and adding the corresponding bound (5.12) for the case $j = m$.

Part (b) is proved in a similar fashion. If $\mathcal{T}_n$ is not a tree then for some $j_2 \in \mathcal{T}_n$ and some $j_1 < j_2$ there are two edge-disjoint paths from j_2 to j_1. For a given pair (j_2, j_1) the mean number of such path-pairs is bounded by $\mathbb{P}(j_2 \in \mathcal{T}_n) \times (\alpha e^\alpha / N)^2$. By (5.12) this is bounded by $(\alpha e^\alpha / N)^3$, and summing over pairs (j_2, j_1) gives the stated bound. ■

REMARK: Note that for part (a) of the lemma we could also appeal to the more sophisticated inequalities of [14] concerning disjoint occurrence of events, which would give the stronger bound $\mathbb{P}(\max \mathcal{T}_n \cap \mathcal{T}_m = j) \le \mathbb{P}(j \in \mathcal{T}_m) \times \mathbb{P}(j \in \mathcal{T}_n)$.

PROOF OF LEMMA 4: Lemma 4 follows from Lemma 5(a) and the observation that $\{C_1^N(i) = C_1^N(j)\} \subseteq \{\mathcal{T}_i \cap \mathcal{T}_j \neq \emptyset\}$. ■

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

5.5. Making the Vertex Labels Distinct

In the ordered model studied above, the vertex-names are $(n, \mathbf{a}^n), 1 \le n \le N$. In order to study the unordered model described at the start of Section 5, we first must address the fact that the vertex-names $\mathbf{a}^n, 1 \le n \le N$ may not be distinct.

LEMMA 6: *Let $\mathcal{G}_N$ be random graphs-with-vertex-names, where (following our standing assumptions) the names have length $\log N/\log A \le L_N = O(\log N)$, and suppose that for some deterministic sequence $k_N = o(N)$, the number of vertices that have non-unique names in $\mathcal{G}_N$, say V_N, satisfies $\mathbb{P}(V_N \ge k_N) = o(1)$. Let $\mathcal{G}_N^*$ be a modification with unique names obtained by re-naming some or all of the non-uniquely-named vertices. Then $|\mathrm{ent}(\mathcal{G}_N^*) - \mathrm{ent}(\mathcal{G}_N)| = o(N \log N)$.*

PROOF: The chain rule (3.2) implies that

$$\mathrm{ent}(\mathcal{G}_N^*) \le \mathrm{ent}(\mathcal{G}_N) + \mathbb{E}\mathrm{ent}(\mathcal{G}_N^*|\mathcal{G}_N),$$

so we want to show that $\mathbb{E}\mathrm{ent}(\mathcal{G}_N^*|\mathcal{G}_N)$ is $o(N \log N)$. Considering the number of ways of relabeling V_N vertices,

$$\begin{aligned}
\mathbb{E}\mathrm{ent}(\mathcal{G}_N^*|\mathcal{G}_N) &\le \mathbb{E}\log\left(V_N!\binom{A^{L_N}}{V_N}\right),\\
&\le \mathbb{E}(\log A^{V_N L_N})\mathbb{1}_{(V_N<k_N)} + \mathbb{E}(\log A^{L_N V_N})\mathbb{1}_{(V_N\ge k_N)},\\
&\le \log(A) L_N[k_N + N\mathbb{P}(V_N \ge k_N)] = o(N\log N),
\end{aligned}$$

as desired. ∎

REMARK: The analogous lemma holds if instead we replace the labels of any random subset of vertices of $\mathcal{G}_N$ to form $\mathcal{G}_N^*$, provided the subset size satisfies the same assumptions as V_N.

LEMMA 7: *For $\mathcal{G}_N^{\mathrm{ord}}$ and $\mathcal{G}_N$,*

$$\mathbb{E}|\{n : \mathbf{a}^n = \mathbf{a}^m \text{ for some } m \ne n\}| = o(N). \tag{5.14}$$

PROOF: As in Lemma 4, the proof is based on studying the originating vertex $C_i(n)$ (now dropping the notational dependence on N) of the letter ultimately copied to coordinate i of vertex n through the "trees" $\mathcal{T}_n$. Given $\mathcal{T}_n$, the copying mechanism that determines the name $\mathbf{a}^n$ evolves independently for each coordinate, and this implies the *conditional independence property:* for $1 \le m < n \le N$, the events $\{C_i(n) = C_i(m)\}, 1 \le i \le L_N$ are conditionally independent given $\mathcal{T}_m$ and $\mathcal{T}_n$. Because

$$\begin{aligned}
&\mathbb{P}(a_i^n = a_i^m|\mathcal{T}_n, \mathcal{T}_m)\\
&\quad = \mathbb{P}(C_i(n) = C_i(m)|\mathcal{T}_n, \mathcal{T}_m) + \tfrac{1}{A}\mathbb{P}(C_i(n) \ne C_i(m)|\mathcal{T}_n, \mathcal{T}_m)
\end{aligned}$$

the conditional independence property implies

$$\begin{aligned}
&\mathbb{P}(\mathbf{a}^n = \mathbf{a}^m|\mathcal{T}_n, \mathcal{T}_m)\\
&\quad = \left[\mathbb{P}(C_1(n) = C_1(m)|\mathcal{T}_n, \mathcal{T}_m) + \tfrac{1}{A}\mathbb{P}(C_1(n) \ne C_1(m)|\mathcal{T}_n, \mathcal{T}_m)\right]^{L_N}.
\end{aligned} \tag{5.15}$$

Now we always have $C_1(n) \in \mathcal{T}_n$ so trivially

$$\mathbb{P}(C_1(n) = C_1(m)|\mathcal{T}_n, \mathcal{T}_m) = 0 \text{ on } \mathcal{T}_n \cap \mathcal{T}_m = \emptyset. \tag{5.16}$$

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

We show below that when the sets do intersect we have

$$\mathbb{P}(C_1(n) = C_1(m)|\mathcal{T}_n, \mathcal{T}_m) \le \tfrac{1}{2} \text{ on } \{\mathcal{T}_n \text{ and } \mathcal{T}_m \text{ are trees}\}. \tag{5.17}$$

Assuming (5.17), since $A \ge 2$, for $p \le 1/2$, we have $p + (1-p)/A \le 3/4$, and now combining (5.15)–(5.17), we find

$$\mathbb{P}(\mathbf{a}^n = \mathbf{a}^m|\mathcal{T}_n, \mathcal{T}_m) \le (\tfrac{3}{4})^{L_N} \mathbb{1}_{(\mathcal{T}_n \cap \mathcal{T}_m \neq \emptyset)} \text{ on } \{\mathcal{T}_n \text{ and } \mathcal{T}_m \text{ are trees}\}.$$

Now take expectation, appeal to part (a) of Lemma 5, and sum over m to conclude

$$\begin{aligned}\mathbb{P}(&\mathcal{T}_n \text{ is a tree}, \mathbf{a}^n = \mathbf{a}^m \text{ for some } m \neq n \text{ for which } \mathcal{T}_m \text{ is a tree})\\ &\le (\tfrac{3}{4})^{L_N} \ ((\alpha e^\alpha)^2 + \alpha e^\alpha) \to 0.\end{aligned}$$

Now any n for which the name $\mathbf{a}^n$ is not unique is either in the set of n defined by the event above, or in one of the two following sets:

$$\{n : \mathcal{T}_n \text{ is not a tree}\}$$

$$\begin{aligned}\{n : \mathcal{T}_n &\text{ is a tree, } \mathbf{a}^n = \mathbf{a}^m \text{ for some } m \neq n \text{ for which } \mathcal{T}_m \text{ is not a tree}\\ &\text{but } \mathbf{a}^n \neq \mathbf{a}^m \text{ for all } m \neq n \text{ for which } \mathcal{T}_m \text{ is a tree}\}.\end{aligned}$$

The cardinality of the final set is at most the cardinality of the previous set, which by part (b) of Lemma 5 has expectation $O(1)$. Combining these bounds gives (5.14).

It remains only to prove (5.17). For $v \in \mathcal{T}_n$ write $R_v(n)$ for the event that the path of copying of coordinate 1 from $C_1(n)$ to n passes through v. We may assume there is at least one edge from n into $[1, n-1]$ (otherwise we are in the setting of (5.16)). Given $\mathcal{T}_n$, the chance that vertex n adopts the label of any given neighbor in $\mathcal{T}_n$ is bounded by $1/2$, we see

$$\mathbb{P}(R_v(n)|\mathcal{T}_n) \le \tfrac{1}{2}, \quad v \in \mathcal{T}_n. \tag{5.18}$$

Similarly by (5.16) we may assume $\mathcal{T}_n \cap \mathcal{T}_m \neq \emptyset$. By hypothesis $\mathcal{T}_n$ and $\mathcal{T}_m$ are trees, and so there is a subset $\mathcal{M} \subseteq \mathcal{T}_n \cap \mathcal{T}_m$ of "first meeting" points v with the property that the path from v to n in $\mathcal{T}_n$ does not meet the path from v to m in $\mathcal{T}_m$ and

$$\{C_1(n) = C_1(m)\} = \cup_{v \in \mathcal{M}} [R_v(n) \cap R_v(m)]$$

with a disjoint union on the right. So

$$\mathbb{P}(C_1(n) = C_1(m)|\mathcal{T}_n, \mathcal{T}_m) = \sum_{v \in \mathcal{M}} \mathbb{P}(R_v(n)|\mathcal{T}_n) \times \mathbb{P}(R_v(m)|\mathcal{T}_m). \tag{5.19}$$

Now $v \to \mathbb{P}(R_v(n)|\mathcal{T}_n)$ and $v \to \mathbb{P}(R_v(m)|\mathcal{T}_m)$ are sub-probability distributions on $\mathcal{M}$ and the former satisfies (5.18). Now (5.19) implies (5.17). ■

5.6. The Unordered Model and its Entropy Rate

The model we introduced as $\mathcal{G}_N$ in section 5.1 does not quite fit our default setting because the vertex-names will typically not be all distinct. However, if we take the ordered model $\mathcal{G}_N^{\text{ord}}$ and then arbitrarily rename the non-unique names, to obtain a model $\mathcal{G}_N^{\text{ord}*}$ say, then

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

Lemmas 6 and 7 imply that only a proportion $o(1)$ of vertices are renamed and the entropy rate is unchanged:

$$(\mathcal{G}_N^{\mathrm{ord}*}) \text{ has entropy rate (5.1).}$$

Now we can "ignore the order", that is replace the names $\{(n, \mathbf{a}^n)\}$ by the now-distinct names $\{\mathbf{a}^n\}$, to obtain a model $\mathcal{G}_N^*$, say. In this section, we will obtain the entropy rate formula for $(\mathcal{G}_N^*)$ as

$$(\text{entropy rate for } \mathcal{G}_N^*) = (\text{entropy rate for } \mathcal{G}_N^{\mathrm{ord}*}) - 1. \tag{5.20}$$

The remainder of this section is devoted to the proof of (5.20). Write $\mathcal{H}_N$ for the Erdős–Rényi graph arising in the construction of $\mathcal{G}_N^{\mathrm{ord}*}$; that is, each vertex $n+1$ is linked to each earlier vertex i with probability α/N, and we regard the created edges as directed edges $(n+1, i)$. Now delete the vertex-labels; consider the resulting graph $\mathcal{H}_N^{\mathrm{unl}}$ as a random unlabeled directed acyclic graph. Given a realization of $\mathcal{H}_N^{\mathrm{unl}}$ there is some number $1 \le M(\mathcal{H}_N^{\mathrm{unl}}) \le N!$ of possible vertex orderings consistent with the edge-directions of the realization.

Lemma 8: *In the notation above,*

$$\mathrm{ent}(\mathcal{G}_N^{\mathrm{ord}*}) = \mathrm{ent}(\mathcal{G}_N^*) + \mathbb{E} \log M(\mathcal{H}_N^{\mathrm{unl}}). \tag{5.21}$$

Proof: According to the chain rule (3.2),

$$\mathrm{ent}(\mathcal{G}_N^{\mathrm{ord}*}) = \mathrm{ent}(\mathcal{G}_N^*) + \mathbb{E}\mathrm{ent}(\mathcal{G}_N^{\mathrm{ord}*} | \mathcal{G}_N^*).$$

We only need to show

$$\mathrm{ent}(\mathcal{G}_N^{\mathrm{ord}*} | \mathcal{G}_N^*) = \log M(\mathcal{H}_N^{\mathrm{unl}}),$$

which follows from two facts: given $\mathcal{G}_N^*$, all possible vertex orderings consistent with the edge-directions of $\mathcal{H}_N^{\mathrm{unl}}$ are equally likely and there are $M(\mathcal{H}_N^{\mathrm{unl}})$ of these orderings. The latter fact is obvious from the definition and to see the former, consider two such orderings; there is a permutation taking one to the other. Given a realization of $\mathcal{G}_N^{\mathrm{ord}*}$ associated with the realization of $\mathcal{H}_N^{\mathrm{unl}}$, applying the same permutation gives a different realization of $\mathcal{G}_N^{\mathrm{ord}*}$ associated with the same realization of $\mathcal{H}_N^{\mathrm{unl}}$. These two realizations of $\mathcal{G}_N^{\mathrm{ord}*}$ have the same probability, and map to the same element of $\mathcal{G}_N^*$, and (here we are using that the second part of the labels are all distinct) this is the only way that different realizations of $\mathcal{G}_N^{\mathrm{ord}*}$ can map to the same element of $\mathcal{G}_N^*$. ■

So it remains only to prove

Proposition 9:

$$\mathbb{E} \log M(\mathcal{H}_N^{\mathrm{unl}}) \sim N \log N.$$

Proof: Choose $K_N \sim N^\varepsilon$ for small $\varepsilon > 0$ and partition the labels $[1, N]$ into K_N consecutive intervals $I_1, I_2, \ldots$, each containing N/K_N labels. Consider a realization of the (labeled) Erdős–Rényi graph $\mathcal{H}_N$. The number V_i of edges with both end-vertices in I_i has Binomial$(\binom{N/K_N}{2}, \alpha/N)$ distribution with mean $\sim \frac{\alpha N}{2K_N^2}$, and from standard large deviation

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

bounds (e.g., [8] Theorem 2.15)

$$\mathbb{P}(V_i \le \tfrac{\alpha N}{K_N^2},\ \text{all } 1 \le i \le K_N) \to 1.$$

For a realization $\mathcal{H}_N$ satisfying these inequalities we have

$$M(\mathcal{H}_N^{\rm unl}) \ge \left(\left(\frac{N}{K_N} - \frac{2\alpha N}{K_N^2} \right)! \right)^{K_N}.$$

This holds because we can create permutations consistent with $\mathcal{H}_N^{\rm unl}$ by, on each interval I_i, first placing the (at most $\frac{2\alpha N}{K_N^2}$) labels involved in the edges with both ends in I_i in increasing order, then placing the remaining labels in arbitrary order. So

$$\begin{aligned} \mathbb{E} \log M(\mathcal{H}_N^{\rm unl}) &\ge (1 - o(1)) \log \left(\left(\frac{N}{K_N} - \frac{2\alpha N}{K_N^2} \right)! \right)^{K_N} \\ &\sim K_N \times \tfrac{N}{K_N} \log \tfrac{N}{K_N} \\ &\sim (1 - \varepsilon) N \log N \end{aligned}$$

establishing Proposition 9. ■

REMARK: Proposition 9 and Lemma 8 are in the spirit of the graph entropy literature, but we could not find these results there. As discussed in Section 2.2, this literature is largely concerned with the complexity of the structure of an unlabeled graph, or in the case of [7], the entropy of probability distributions on unlabeled graphs. A quantity of interest in these settings is the "automorphism group" of the graph which is closely related to $M(\mathcal{H}_N^{\rm unl})$ here. For example, an analog of (5.21) is shown in Lemma 1 of [7] and Theorem 1 there uses this lemma to relate the entropy rate between an Erdős–Rényi graph on N vertices with edge probabilities p_N with distinguished vertices and that of the same model where the vertex labels are ignored. Their result is very close to Proposition 9, but [7] only considers edge weights p_N satisfying $Np_N/\log(N)$ bounded away from zero, which falls outside our setting.

6. OPEN PROBLEMS

Aside from the (quite easy) Lemmas 1 and 2, our results concern specific models. Are there interesting "general" results in this topic? Here are two possible avenues for exploration.

Given a random graph-with vertex-names $\mathcal{G} = \mathcal{G}_N$, there is an associated random unlabeled graph $\mathcal{G}^{\rm unl}$ and an associated random unordered set of names **Names**, and obviously

$$\text{ent}(\mathcal{G}) \ge \max(\text{ent}(\mathcal{G}^{\rm unl}), \text{ent}(\textbf{Names})).$$

Lemmas 1 and 2, applied conditionally as indicated in Section 4.6, give sufficient conditions for $\text{ent}(\mathcal{G}) = \text{ent}(\mathcal{G}^{\rm unl})$ or $\text{ent}(\mathcal{G}) = \text{ent}(\textbf{Names})$. In general, one could ask "given $\mathcal{G}^{\rm unl}$ and **Names**, how random is the assignment of names to vertices?" The standard notion of *graph entropy* enters here, as a statistic of the "completely random" assignment, so the appropriate conditional graph entropy within a model constitutes a measure of relative randomness. Another question concerns measures of strength of association of names across edges. One could just take the space $\mathbf{A}^{L_N}$ of possible names, consider the empirical distribution across edges (v, w) of the pair of names $(\mathbf{a}(v), \mathbf{a}(w))$ as a distribution on the product space $\mathbf{A}^{L_N} \times \mathbf{A}^{L_N}$ and compare with the product measure using some quantitative measure

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

of dependence. But neither of these procedures quite gets to grips with the issue of finding conceptually interpretable quantitative measures of dependence between graph structure and name structure, which we propose as an open problem.

A second issue concerns "local" upper bounds for the entropy rate. In the classical context of sequences $X_1, \ldots, X_n$ from $\mathbf{A}$, an elementary consequence of subadditivity is that (without any further assumptions) one can upper bound $\mathrm{ent}(X_1, \ldots, X_n)$ in terms of the "size-k random window" entropy

$$\mathcal{E}_{n,k} := \mathrm{ent}(X_U, X_{U+1}, \ldots, X_{U+k-1}); \quad U \text{ uniform on } [1, n-k+1]$$

and this is optimal in the sense that for a stationary ergodic sequence the "global" entropy rate is actually equal to the quantity

$$\lim_{k\to\infty} \lim_{n\to\infty} k^{-1}\mathcal{E}_{n,k}$$

arising from this "local" upper bound. In our setting, we would like some analogous result saying that, for the entropy $\mathcal{E}_{N,k}$ of the restriction of $\mathcal{G}_N$ to some "size-k" neighborhood of a random vertex, there is always an upper bound for the entropy rate c of the form

$$c \le \lim_{k\to\infty} \lim_{N\to\infty} \frac{\mathcal{E}_{N,k}}{k \log N}$$

and that this is an equality under some "no long-range dependence" condition analogous to ergodicity. But results of this kind seem hard to formulate, because of the difficulty in specifying which vertices and edges are to be included in the "size-k" neighborhood.

Acknowledgments

The hybrid model arose from a conversation with Sukhada Fadnavis. The work for this project was completed when NR was at University of California, Berkeley with support from NSF grants DMS-0704159, DMS-0806118, DMS-1106999 and ONR grant N00014-11-1-0140. Aldous's research supported by N.S.F Grant DMS-1106998.

References

1. Alderson, D.L. & Doyle, J.C. (2010). Contrasting views of complexity and their implications for network-centric infrastructures. *Systems, Man and Cybernetics* 40: 839–852.
2. Aldous, D. (1991). The continuum random tree. I. *Annals of Probability* 19(1): 1–28.
3. Aldous, D. & Lyons, R. (2007). Processes on unimodular random networks. *Electronic Journal of Probability* 12(54): 1454–1508.
4. Aldous, D. & Michael Steele, J. (2004). The objective method: probabilistic combinatorial optimization and local weak convergence. *Probability on Discrete Structures*, Vol. 110, *Encyclopaedia of Mathematical Science*, Berlin: Springer, pp. 1–72.
5. Boldi, P. & Vigna, S. (2003). The webgraph framework I: compression techniques. In: *Proceedings of the 13th International World Wide Web Conference*, pp. 595–601. New York: ACM Press.
6. Chierichetti, F., Kumar, R., Lattanzi, S., Mitzenmacher, M., Panconesi A., & Raghavan P. (2009). On compressing social networks. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD '09), New York, NY, USA, ACM, pp. 219–228.
7. Choi, Y. & Szpankowski, W. (2012). Compression of graphical structures: Fundamental limits, algorithms, and experiments. *IEEE Transaction on Information Theory* 58: 620–638.
8. Chung, F. & Lu, L. (2006). *Complex graphs and networks*, Vol. 107 *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC.
9. Cover, T.M. & Thomas, J.A. (2006). *Elements of Information Theory*, 2nd edn, Hoboken, NJ, Wiley-Interscience [John Wiley & Sons].
10. Dehmer, M. & Mowshowitz, A. (2011). A history of graph entropy measures. *Information Science* 181(1): 57–78.

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

11. Kang, R.J. & McDiarmid, C. (2010). The t-improper chromatic number of random graphs. *Combinatorics, Probabability and Computing* 19(1): 87–98.
12. Kontoyiannis, I. (2003). Pattern matching and lossy data compression on random fields. *IEEE Transaction on Information Theory* 49(4): 1047–1051.
13. Orlitsky, A., Santhanam, N.P. & Zhang, J. (2004). Universal compression of memoryless sources over unknown alphabets. *IEEE Transaction on Information Theory* 50(7): 1469–1481.
14. van den Berg, J. & Kesten, H. (1985). Inequalities with applications to percolation and reliability. *Journal of Applied Probability* 22(3): 556–569.
15. Wagner A.B., Viswanath, P., & Kulkarni, S.R. (2011). Probability estimation in the rare-events regime. *IEEE Transaction on Information Theory* 57(6): 3207–3229.

APPENDIX A

A.1 Small Worlds Model: $0 < \gamma < 2$

Here, we complete the analysis of the graph entropy rate in the "small worlds" model of Section 4.3. First, we show that for a as in (4.4), the average degree tends to a constant. For $D_u = (n-1)/2 + b$ and $D_l = (n-1)/2 - b$, where b is constant with respect to N (and chosen large enough for the inequalities below to hold), we find using (4.1) that

$$\begin{aligned} 8a \int_0^{\pi/4} \int_1^{D_l \sec(\theta)} r^{-\gamma+1} dr\, d\theta &\le \mathbb{E}D(v) - 4 \\ &\le 8a \int_0^{\pi/4} \int_1^{D_u \sec(\theta)} r^{-\gamma+1} dr\, d\theta + \tfrac{4a}{2^{\gamma/2}}, \\ \tfrac{8aD_l^{2-\gamma}}{2-\gamma} \int_0^{\pi/4} \sec^{2-\gamma}(\theta)\, d\theta - \tfrac{2a\pi}{2-\gamma} &\le \mathbb{E}D(v) - 4 \\ &\le \tfrac{8aD_u^{2-\gamma}}{2-\gamma} \int_0^{\pi/4} \sec^{2-\gamma}(\theta)\, d\theta - \tfrac{2a\pi}{2-\gamma} + \tfrac{4a}{2^{\gamma/2}}. \end{aligned} \tag{A.1}$$

Taking a and κ_γ as in (4.4), the inequalities above imply $\mathbb{E}D(v) \to 4 + \alpha$.

To show the entropy rate is as claimed, take N large enough to make $a < 1/2$, so that $\mathcal{E}(ar^{-\gamma})$ is a decreasing function of r for $r > 1$. Using the inequality $-(1-x)\log(1-x) \le x$ for $0 < x < 1$,

$$\begin{aligned} &\frac{-8a}{\log(N)} \int_0^{\pi/4} \int_1^{D_l \sec(\theta)} \mathcal{E}(ar^{-\gamma}) r\, dr\, d\theta \\ &\quad \sim \frac{-8a}{\log(N)} \int_0^{\pi/4} \int_1^{D_l \sec(\theta)} r^{-\gamma+1} \log(ar^{-\gamma})\, dr\, d\theta, \end{aligned} \tag{A.2}$$

and we will show (A.2) tends to α as $N \to \infty$. From this point, similar arguments show the same is true with D_l replaced by D_u, so that following the arguments that established the convergence of the average degree and using (4.2), we find

$$\lim_{N\to\infty} \frac{\mathrm{ent}(\mathcal{G}_N)}{N \log(N)} = \frac{\alpha}{2},$$

as desired. To obtain the claimed asymptotic, note that we can write (A.2) as

$$-8a \log(a) \int_0^{\pi/4} \int_1^{D_l \sec(\theta)} r^{-\gamma+1} dr\, d\theta \tag{A.3}$$

$$+ 8a\gamma \int_0^{\pi/4} \int_1^{D_l \sec(\theta)} r^{-\gamma+1} \log(r)\, dr\, d\theta. \tag{A.4}$$

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press

From (A.1) above and the definition (4.4) of a, it is easy to see that (A.3) is

$$\alpha(1-\gamma/2)\log(N)+\mathrm{o}(\log(N)) \text{ as } N\to\infty. \tag{A.5}$$

Now, making the substitution $u=r^{2-\gamma}$, (A.4) is equal to

$$\begin{aligned}&\frac{8a\gamma}{(2-\gamma)^2}\int_0^{\pi/4}\int_1^{D_l^{2-\gamma}\sec^{2-\gamma}(\theta)}\log(u)\,du\,d\theta\\&\quad=\frac{8a\gamma}{(2-\gamma)^2}\int_0^{\pi/4}D_l^{2-\gamma}\sec^{2-\gamma}(\theta)[\log(D_l^{2-\gamma}\sec^{2-\gamma}(\theta))-1]\,d\theta.\end{aligned}$$

After simplification, the only term that is not $\mathrm{o}(\log(N))$ is

$$\frac{8a\gamma}{(2-\gamma)^2}D_l^{2-\gamma}\log(D_l^{2-\gamma})\int_0^{\pi/4}\sec^{2-\gamma}(\theta)\,d\theta,$$

which after simplification is equal to

$$\frac{\alpha\gamma}{2}\log(N)+\mathrm{o}(\log(N)). \tag{A.6}$$

Combining (A.5) and (A.6) with (A.3) and (A.4) implies (A.2) tends to α as $N\to\infty$.

https://doi.org/10.1017/S0269964813000399 Published online by Cambridge University Press