SIAM J. COMPUT.
Vol. 29, No. 2, pp. 590–599

KOLMOGOROV RANDOM GRAPHS AND THE INCOMPRESSIBILITY METHOD*

HARRY BUHRMAN†, MING LI‡, JOHN TROMP§, AND PAUL VITÁNYI¶

Abstract. We investigate topological, combinatorial, statistical, and enumeration properties of finite graphs with high Kolmogorov complexity (almost all graphs) using the novel incompressibility method. Example results are (i) the mean and variance of the number of (possibly overlapping) ordered labeled subgraphs of a labeled graph as a function of its randomness deficiency (how far it falls short of the maximum possible Kolmogorov complexity) and (ii) a new elementary proof for the number of unlabeled graphs.

Key words. Kolmogorov complexity, incompressibility method, random graphs, enumeration of graphs, algorithmic information theory

AMS subject classifications. 68Q30, 05C80, 05C35, 05C30

PII. S0097539797327805

1. Introduction.

The incompressibility of individual random objects yields a simple but powerful proof technique. The incompressibility method [9] is a new general purpose tool and should be compared with the pigeon hole principle or the probabilistic method. Here we apply the incompressibility method to randomly generated graphs and "individually random" graphs—graphs with high Kolmogorov complexity.

In a typical proof using the incompressibility method, one first chooses an individually random object from the class under discussion. This object is effectively incompressible. The argument invariably says that if a desired property does not hold, then the object can be compressed. This yields the required contradiction. Since a randomly generated object is *with overwhelming probability* individually random and hence incompressible, one usually obtains the property with high probability.

Results. We apply the incompressibility method to obtain combinatorial properties of graphs with high Kolmogorov complexity. These properties are parametrized in terms of a "randomness deficiency" function.[1] This can be considered as a parametrized version of the incompressibility method. In section 2 we show that for every labeled graph on n nodes with high Kolmogorov complexity (also called "Kolmogorov random graph" or "high-complexity graph"), the node degree of every vertex is about $n/2$ and there are about $n/4$ node-disjoint paths of length 2 between every

*Received by the editors September 23, 1997; accepted for publication (in revised form) October 9, 1998; published electronically November 10, 1999. A partial preliminary version appeared in *Proceedings of the Conference on Compression and Complexity of Sequences*, Salerno, Italy, 1997, IEEE Press, Piscataway, NJ.

http://www.siam.org/journals/sicomp/29-2/32780.html

†CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands (buhrman@cwi.nl). The work of this author was partially supported by the European Union through NeuroCOLT ESPRIT Working Group 8556, and by NWO through NFI Project ALADDIN NF 62-376 and SION grant 612-34-002.

‡Department of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada (mli@wh.math.uwaterloo.ca). The work of this author was supported in part by the NSERC Operating grant OGP0046506, CITO, a CGAT grant, and the Steacie Fellowship.

§CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands (tromp@cwi.nl).

¶CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands (paulv@cwi.nl). The work of this author was partially supported by the European Union through NeuroCOLT ESPRIT Working Group 8556 and by NWO through NFI Project ALADDIN NF 62-376.

[1] Randomness deficiency measures how far the object falls short of the maximum possible Kolmogorov complexity. It is formally defined in Definition 4.

pair of nodes. In section 2.2, we analyze "normality" properties of Kolmogorov random graphs. In analogy with infinite sequences one can call an infinite labeled graph normal if each finite ordered labeled subgraph of size k occurs in the appropriate sense (possibly overlapping) with limiting frequency $2^{-\binom{k}{2}}$. It follows from the Martin-Löf theory of effective tests for randomness [14] that individually random (high complexity) infinite labeled graphs are normal. Such properties cannot hold precisely for finite graphs, where randomness is necessarily a matter of degree: We determine close quantitative bounds on the normality (frequency of subgraphs) of high-complexity finite graphs in terms of their randomness deficiency.

Denote the number of unlabeled graphs on n nodes by g_n. In section 2.3 we demonstrate the use of the incompressibility method and Kolmogorov random graphs by providing a new elementary proof that $g_n \sim 2^{\binom{n}{2}}/n!$. This has previously been obtained by more advanced methods [12]. Moreover, we give a good estimate of the error term. Part of the proof involves estimating the order (number of automorphisms) $s(G)$ of graphs G as a function of the randomness deficiency of G. For example, we show that labeled graphs with randomness deficiency appropriately less than n are rigid (have but one automorphism: the identity automorphism).

Related work. Several properties (high degree nodes, diameter 2, rigidity) have also been proven by traditional methods to hold with high probability for randomly generated graphs [5, 4]. We provide new proofs for these results using the incompressibility method. They are actually proved to hold for the definite class of Kolmogorov random graphs—rather than with high probability for randomly generated graphs.

In [10] (also [9]) Li and Vitányi investigated topological properties of labeled graphs with high Kolmogorov complexity and proved them using the incompressibility method to compare ease of such proofs with the probabilistic method [7] and the entropy method.

In [8] it was shown that every labeled tree on n nodes with randomness deficiency $O(\log n)$ has maximum node degree of $O(\log n/\log\log n)$. Analysis of Kolmogorov random graphs was used to establish the total interconnect length of Euclidean (real-world) embeddings of computer network topologies [15] and the size of compact routing tables in computer networks [6]. Infinite binary sequences that asymptotically have equal numbers of 0s and 1s and, more generally, where every block of length k occurs (possibly overlapping) with frequency $1/2^k$ were called "normal" by E. Borel [2]. References [9, 11] investigate the quantitative deviation from normal as a function of the Kolmogorov complexity of a finite binary string. Here we consider an analogous question for Kolmogorov random graphs.[2] Finally, there is a close relation and genuine difference between high-probability properties and properties of incompressible objects; see [9, Section 6.2].

1.1. Kolmogorov complexity. We use the following notation. Let A be a finite set. By $d(A)$ we denote the *cardinality* of A. In particular, $d(\emptyset) = 0$. Let x be a finite binary string. Then $l(x)$ denotes the *length* (number of bits) of x. In particular, $l(\epsilon) = 0$, where ϵ denotes the *empty word*.

Let $x, y, z \in \mathcal{N}$, where $\mathcal{N}$ denotes the natural numbers. Identify $\mathcal{N}$ and $\{0,1\}^*$

[2] There are some results along these lines related to randomly generated graphs, but as far as the authors could ascertain (consulting Alan Frieze, Svante Janson, and Andrzej Rucinski around June 1996) such properties have not been investigated in the same detail as here. See, for example, [1, pp. 125–140]. But note that pseudorandomness also is different from Kolmogorov randomness.

according to the correspondence

$$(0, \epsilon), (1, 0), (2, 1), (3, 00), (4, 01), \ldots .$$

Hence, the length $l(x)$ of x is the number of bits in the binary string or number x. Let $T_0, T_1, \ldots$ be a standard enumeration of all Turing machines. Let $\langle \cdot, \cdot \rangle$ be a standard one-to-one mapping from $\mathcal{N} \times \mathcal{N}$ to $\mathcal{N}$, for technical reasons chosen such that $l(\langle x, y \rangle) = l(y) + O(l(x))$. An example is $\langle x, y \rangle = 1^{l(x)}0xy$. This can be iterated to $\langle \langle \cdot, \cdot \rangle, \cdot \rangle$.

Informally, the Kolmogorov complexity [13] of x is the length of the *shortest* effective description of x. That is, the *Kolmogorov complexity* $C(x)$ of a finite string x is simply the length of the shortest program, say in FORTRAN (or in Turing machine codes) encoded in binary, which prints x without any input. A similar definition holds conditionally in the sense that $C(x|y)$ is the length of the shortest binary program which computes x on input y. Kolmogorov complexity is absolute in the sense of being independent of the programming language up to a fixed additional constant term which depends on the programming language but not on x. We now fix one canonical programming language once and for all as reference and thereby $C()$. For the theory and applications, see [9]. A formal definition is as follows:

DEFINITION 1. *Let U be an appropriate universal Turing machine such that*

$$U(\langle \langle i, p \rangle, y \rangle) = T_i(\langle p, y \rangle)$$

for all i and $\langle p, y \rangle$. The conditional Kolmogorov complexity *of x given y is*

$$C(x|y) = \min_{p \in \{0,1\}^*} \{l(p) : U(\langle p, y \rangle) = x\}.$$

The unconditional Kolmogorov complexity of x is defined as $C(x) := C(x|\epsilon)$.

It is easy to see that there are strings that can be described by programs much shorter than themselves. For instance, the function defined by $f(1) = 2$ and $f(i) = 2^{f(i-1)}$ for $i > 1$ grows very fast, $f(k)$ is a "stack" of k twos. Yet for each k it is clear that $f(k)$ has complexity at most $C(k) + O(1)$. What about incompressibility?

By a simple counting argument one can show that whereas some strings can be enormously compressed, the majority of strings can hardly be compressed at all.

For each n there are 2^n binary strings of length n but only $\sum_{i=0}^{n-1} 2^i = 2^n - 1$ possible shorter descriptions. Therefore, there is at least one binary string x of length n such that $C(x) \geq n$. We call such strings *incompressible*. It also follows that for any length n and any binary string y, there is a binary string x of length n such that $C(x|y) \geq n$. Generally, for every constant c we can say a string x is *c-incompressible* if $C(x) \geq l(x) - c$. Strings that are incompressible (say, c-incompressible with small c) are patternless, since a pattern could be used to reduce the description length. Intuitively, we think of such patternless sequences as being random, and we use "random sequence" synonymously with "incompressible sequence."[3] By the same counting argument as before we find that the number of strings of length n that are c-incompressible is at least $2^n - 2^{n-c} + 1$. Hence there is at least one 0-incompressible string of length n, at least one-half of all strings of length n are 1-incompressible, at least three-fourths of all strings of length n are 2-incompressible, ..., and at least the $(1 - 1/2^c)$th part of all 2^n strings of length n are c-incompressible. This means

[3] It is possible to give a rigorous formalization of the intuitive notion of a random sequence as a sequence that passes all effective tests for randomness; see, for example, [9].

that for each constant $c \geq 1$ the majority of all strings of length n (with $n > c$) is c-incompressible. We generalize this to the following simple but extremely useful lemma.

LEMMA 1. *Let c be a positive integer. For each fixed y, every set A of cardinality m has at least $m(1 - 2^{-c}) + 1$ elements x with $C(x|y) \geq \lfloor \log m \rfloor - c$.*

Proof. The proof is by simple counting. □

As an example, set $A = \{x : l(x) = n\}$. Then the cardinality of A is $m = 2^n$. Since it is easy to assert that $C(x) \leq n + c$ for some fixed c and all x in A, Lemma 1 demonstrates that this trivial estimate is quite sharp. The deeper reason is that since there are few short programs, there can be only few objects of low complexity. We require another quantity: The prefix Kolmogorov complexity which is defined just as $C(\cdot|\cdot)$ but now with respect to a subset of Turing machines that have the property that the set of programs for which the machine halts is prefix-free; that is, no halting program is a prefix of any other halting program. For details see [9]. Here we require only the quantitative relation below.

DEFINITION 2. *The* prefix *Kolmogorov complexity of x conditional to y is denoted by $K(x|y)$. It satisfies the inequality*

$$C(x|y) \leq K(x|y) \leq C(x|y) + 2\log C(x|y) + O(1).$$

2. Kolmogorov random graphs.

Statistical properties of strings with high Kolmogorov complexity have been studied in [11]. The interpretation of strings as more complex combinatorial objects leads to a new set of properties and problems that have no direct counterpart in the "flatter" string world. Here we derive topological, combinatorial, and statistical properties of graphs with high Kolmogorov complexity. Every such graph possesses simultaneously all properties that hold with high probability for randomly generated graphs. They constitute "almost all graphs" and the derived properties a fortiori hold with probability that goes to 1 as the number of nodes grows unboundedly.

DEFINITION 3. *Each labeled graph $G = (V, E)$ on n nodes $V = \{1, 2, \ldots, n\}$ can be represented (up to automorphism) by a binary string $E(G)$ of length $\binom{n}{2}$. We simply assume a fixed ordering of the $\binom{n}{2}$ possible edges in an n-node graph, e.g., lexicographically, and let the ith bit in the string indicate presence* (1) *or absence* (0) *of the ith edge. Conversely, each binary string of length $\binom{n}{2}$ encodes an n-node graph. Hence we can identify each such graph with its binary string representation.*

DEFINITION 4. *A labeled graph G on n nodes has* randomness deficiency *at most $\delta(n)$ and is called $\delta(n)$-*random *if it satisfies*

$$C(E(G)|n) \geq \binom{n}{2} - \delta(n). \tag{2.1}$$

2.1. Some basic properties.

Using Lemma 1, with $y = n$, A the set of strings of length $\binom{n}{2}$, and $c = \delta(n)$ gives us the following lemma.

LEMMA 2. *A fraction of at least $1 - 1/2^{\delta(n)}$ of all labeled graphs G on n nodes is $\delta(n)$-random.*

As a consequence, for example, the $c\log n$-random labeled graphs constitute a fraction of at least $(1 - 1/n^c)$ of all graphs on n nodes, where $c > 0$ is an arbitrary constant.

Labeled graphs with high complexity have many specific topological properties, which seem to contradict their randomness. However, these are simply the likely properties, whose absence would be rather unlikely. Thus, randomness enforces strict statistical regularities—for example, to have diameter exactly 2.

We will use the following lemma (Theorem 2.6.1 in [9]).

LEMMA 3. *Let $x = x_1 \ldots x_n$ be a binary string of length n, and y a much smaller string of length l. Let $p = 2^{-l}$ and $\#y(x)$ be the number of (possibly overlapping) distinct occurrences of y in x. For convenience, we assume that x "wraps around" so that an occurrence of y starting at the end of x and continuing at the start also counts. Assume that $l \leq \log n$. There is a constant c such that for all n and $x \in \{0,1\}^n$ if $C(x) \geq n - \delta(n)$, then*

$$|\#y(x) - pn| \leq \sqrt{\alpha pn}$$

with $\alpha = [K(y|n) + \log l + \delta(n) + c]3l/\log e$.

LEMMA 4. *All $o(n)$-random labeled graphs have $n/4 + o(n)$ disjoint paths of length 2 between each pair of nodes i, j. In particular, all $o(n)$-random labeled graphs have diameter 2.*

Proof. The only graphs with diameter 1 are the complete graphs that can be described in $O(1)$ bits, given n, and hence are not random. It remains to consider an $o(n)$-random graph $G = (V, E)$ with diameter greater than or equal to 2. Let i, j be a pair of nodes connected by r disjoint paths of length 2. Then we can describe G by modifying the old code for G as follows:

- a program to reconstruct the object from the various parts of the encoding in $O(1)$ bits;
- the identities of $i < j$ in $2 \log n$ bits;
- the old code $E(G)$ of G with the $2(n-2)$ bits representing presence or absence of edges (j, k) and (i, k) for each $k \neq i, j$ deleted;
- a short program for the string $e_{i,j}$ consisting of the (reordered) $n - 2$ pairs of bits deleted above.

From this description we can reconstruct G in

$$O(\log n) + \binom{n}{2} - 2(n-2) + C(e_{i,j}|n)$$

bits, from which we may conclude that $C(e_{i,j}|n) \geq l(e_{i,j}) - o(n)$. As shown in [11] or [9] (here Lemma 3) this implies that the frequency of occurrence in $e_{i,j}$ of the aligned 2-bit block "11"—which by construction equals the number of disjoint paths of length 2 between i and j—is $n/4 + o(n)$. □

A graph is *k-connected* if there are at least k node-disjoint paths between every pair of nodes.

COROLLARY 1. *All $o(n)$-random labeled graphs are $(\frac{n}{4} + o(n))$-connected.*

LEMMA 5. *Let $G = (V, E)$ be a graph on n nodes with randomness deficiency $O(\log n)$. Then the largest clique in G has at most $\lfloor 2 \log n \rfloor + O(1)$ nodes.*

Proof. The proof is the same as the largest size transitive subtournament in a high-complexity tournament as in [9]. □

With respect to the related property of random graphs, in [1, pp. 86–87], it is shown that a random graph with edge probability 1/2 contains a clique on asymptotically $2 \log n$ nodes with probability at least $1 - e^{-n^2}$.

2.2. Statistics of subgraphs. We start by defining the notion of labeled subgraph of a labeled graph.

DEFINITION 5. *Let $G = (V, E)$ be a labeled graph on n nodes. Consider a labeled graph H on k nodes $\{1, 2, \ldots, k\}$. Each subset of k nodes of G induces a subgraph G_k of G. The subgraph G_k is an ordered labeled* occurrence *of H when we obtain H by relabeling the nodes $i_1 < i_2 < \cdots < i_k$ of G_k as $1, 2, \ldots, k$.*

It is easy to conclude from the statistics of high-complexity strings in Lemma 3 that the frequency of each of the two labeled two-node subgraphs (there are only two different ones: the graph consisting of two isolated nodes and the graph consisting of two connected nodes) in a $\delta(n)$-random graph G is

$$\frac{n(n-1)}{4} \pm \sqrt{\frac{3}{4}(\delta(n) + O(1))n(n-1)/\log e}.$$

This case is easy since the frequency of such subgraphs corresponds to the frequency of 1s or 0s in the $\binom{n}{2}$-length standard encoding $E(G)$ of G. However, to determine the frequencies of labeled subgraphs on k nodes (up to isomorphism) for $k > 2$ is a matter more complicated than the frequencies of substrings of length k. Clearly, there are $\binom{n}{k}$ subsets of k nodes out of n and hence that many occurrences of subgraphs. Such subgraphs may overlap in more complex ways than substrings of a string. Let $\#H(G)$ be *the number of times H occurs* as an ordered labeled subgraph of G (possibly overlapping). Let p be the probability that we obtain H by flipping a fair coin to decide for each pair of nodes whether it is connected by an edge or not:

$$p = 2^{-k(k-1)/2}. \tag{2.2}$$

THEOREM 1. *Assume the terminology above with $G = (V, E)$ a labeled graph on n nodes, k is a positive integer dividing n, and H is a labeled graph on $k \le \sqrt{2\log n}$ nodes. Let $C(E(G)|n) \ge \binom{n}{2} - \delta(n)$. Then*

$$\left|\#H(G) - \binom{n}{k}p\right| \le \binom{n}{k}\sqrt{\alpha(k/n)p}$$

with $\alpha := (K(H|n) + \delta(n) + \log\binom{n}{k}/(n/k) + O(1))3/\log e$.

Proof. A *cover* of G is a set $C = \{S_1, \ldots, S_N\}$ with $N = n/k$, where the S_i's are pairwise disjoint subsets of V and $\bigcup_{i=1}^{N} S_i = V$. According to [3], we have the following claim.

Claim 1. There is a partition of the $\binom{n}{k}$ different k-node subsets into $h = \binom{n}{k}/N$ distinct covers of G, each cover consisting of $N = n/k$ disjoint subsets. That is, each subset of k nodes of V belongs to precisely one cover.

Enumerate the covers as $C_0, C_2, \ldots, C_{h-1}$. For each $i \in \{0, 1, \ldots, h-1\}$ and k-node labeled graph H, let $\#H(G, i)$ be the number of (now nonoverlapping) occurrences of subgraph H in G occurring in cover C_i.

Now consider an experiment of N trials, each trial with the same set of $2^{k(k-1)/2}$ outcomes. Intuitively, each trial corresponds to an element of a cover, and each outcome corresponds to a k-node subgraph. For every i we can form a string s_i consisting of the N blocks of $\binom{k}{2}$ bits that represent presence or absence of edges within the induced subgraphs of each of the N subsets of C_i. Since G can be reconstructed from n, i, s_i, and the remaining $\binom{n}{2} - N\binom{k}{2}$ bits of $E(G)$, we find that $C(s_i|n) \ge l(s_i) - \delta(n) - \log h$. Again, according to Lemma 3 this implies that the frequency of occurrence of the aligned $\binom{k}{2}$-block $E(H)$, which is $\#H(G, i)$, equals

$$Np \pm \sqrt{Np\alpha}$$

with α as in the statement of Theorem 1. One can do this for each i independently, notwithstanding the dependence between the frequencies of subgraphs in different covers. Namely, the argument depends on the incompressibility of G alone. If the number of occurrences of a certain subgraph in *any* of the covers is too large or too small then we can compress G. Now,

$$\begin{aligned}\left|\#H(G) - p\binom{n}{k}\right| &= \sum_{i=0}^{h-1} |\#H(G,i) - Np| \\ &\leq \binom{n}{k}\sqrt{\alpha(k/n)p}. \qquad \square\end{aligned}$$

In [9, 11] we investigated up to which length l all blocks of length l occurred at least once in each $\delta(n)$-random string of length n.

THEOREM 2. *Let $\delta(n) < 2^{\sqrt{\frac{1}{2}\log n}}/4\log n$ and G be a $\delta(n)$-random graph on n nodes. Then for sufficiently large n, the graph G contains all subgraphs on $\sqrt{2\log n}$ nodes.*

Proof. We are sure that H on k nodes occurs at least once in G if $\binom{n}{k}\sqrt{\alpha(k/n)p}$ in Theorem 1 is less than $\binom{n}{k}p$. This is the case if $\alpha < (n/k)p$. This inequality is satisfied for an overestimate of $K(H|n)$ by $\binom{k}{2}+2\log\binom{k}{2}+O(1)$ (since $K(H|n) \leq K(H)+O(1)$), and $p = 2^{-k(k-1)/2}$, with k set at $k = \sqrt{2\log n}$. This proves the theorem. $\square$

2.3. Unlabeled graph counting. An unlabeled graph is a graph with no labels. For convenience we can define this as follows: Call two labeled graphs *equivalent* (up to relabeling) if there is a relabeling that makes them equal. An *unlabeled graph* is an equivalence class of labeled graphs. An *automorphism* of $G = (V, E)$ is a permutation π of V such that $(\pi(u), \pi(v)) \in E$ iff $(u, v) \in E$. Clearly, the set of automorphisms of a graph forms a group with group operation of function composition and the identity permutation as unity. It is easy to verify that π is an automorphism of G iff $\pi(G)$ and G have the *same binary string standard encoding*, that is, $E(G) = E(\pi(G))$. This contrasts with the more general case of permutation relabeling, where the standard encodings may be different. A graph is *rigid* if its only automorphism is the identity automorphism. It turns out that Kolmogorov random graphs are rigid graphs. To obtain an expression for the number of unlabeled graphs we have to estimate the number of automorphisms of a graph in terms of its randomness deficiency.

In [12] an asymptotic expression for the number of unlabeled graphs is derived using sophisticated methods. We give a new elementary proof by incompressibility. Denote by g_n the number of unlabeled graphs on n nodes—that is, the number of isomorphism classes in the set $\mathcal{G}_n$ of undirected graphs on nodes $\{0, 1, \ldots, n-1\}$.

THEOREM 3. $g_n \sim \frac{2^{\binom{n}{2}}}{n!}$.

Proof. Clearly,

$$g_n = \sum_{G \in \mathcal{G}_n} \frac{1}{d(\bar{G})},$$

where $\bar{G}$ is the isomorphism class of graph G. By elementary group theory,

$$d(\bar{G}) = \frac{d(S_n)}{d(Aut(G))} = \frac{n!}{d(Aut(G))},$$

where S_n is the group of permutations on n elements and $Aut(G)$ is the automorphism group of G. Let us partition $\mathcal{G}_n$ into $\mathcal{G}_n = \mathcal{G}_n^0 \cup \cdots \cup \mathcal{G}_n^n$, where $\mathcal{G}_n^m$ is the set of graphs for which m is the number of nodes moved (mapped to another node) by any of its automorphisms.

Claim 2. *For* $G \in \mathcal{G}_n^m$, $d(Aut(G)) \leq n^m = 2^{m \log n}$.

Proof. $d(Aut(G)) \leq \binom{n}{m} m! \leq n^m$. □

Consider each graph $G \in \mathcal{G}_n$ having a probability $\mathrm{Prob}(G) = 2^{-\binom{n}{2}}$.

Claim 3. $\mathrm{Prob}(G \in \mathcal{G}_n^m) \leq 2^{-m(\frac{n}{2} - \frac{3m}{8} - \log n)}$.

Proof. By Lemma 2 it suffices to show that if $G \in \mathcal{G}_n^m$ and

$$C(E(G)|n,m) \geq \binom{n}{2} - \delta(n,m)$$

then $\delta(n,m)$ satisfies

$$\delta(n,m) \geq m\left(\frac{n}{2} - \frac{3m}{8} - \log n\right). \tag{2.3}$$

Let $\pi \in Aut(G)$ move m nodes. Suppose π is the product of k disjoint cycles of sizes $c_1, \ldots, c_k$. Spend at most $m \log n$ bits describing π: For example, if the nodes $i_1 < \cdots < i_m$ are moved then list the sequence $\pi(i_1), \ldots, \pi(i_m)$. Writing the nodes of the latter sequence in increasing order we obtain $i_1, \ldots, i_m$ again; that is, we execute permutation π^{-1} and hence we obtain π.

Select one node from each cycle—say, the lowest numbered one. Then for every unselected node on a cycle, we can delete the $n-m$ bits corresponding to the presence or absence of edges to stable nodes, and $m-k$ half-bits corresponding to presence or absence of edges to the other, unselected cycle nodes. In total we delete

$$\sum_{i=1}^{k}(c_i - 1)\left(n - m + \frac{m-k}{2}\right) = (m-k)\left(n - \frac{m+k}{2}\right)$$

bits. Observing that $k = m/2$ is the largest possible value for k, we arrive at the claimed $\delta(n,m)$ of G (difference between savings and spendings is $\frac{m}{2}(n - \frac{3m}{4}) - m \log n$) of (2.3). □

We continue the proof of Theorem 3:

$$g_n = \sum_{G \in \mathcal{G}_n} \frac{1}{d(\bar{G})} = \sum_{G \in \mathcal{G}_n} \frac{d(Aut(g))}{n!} = \frac{2^{\binom{n}{2}}}{n!} E_n,$$

where $E_n := \sum_{G \in \mathcal{G}_n} \mathrm{Prob}(G) d(Aut(G))$ is the expected size of the automorphism group of a graph on n nodes. Clearly, $E_n \geq 1$, yielding the lower bound on g_n. For the upper bound on g_n, noting that $\mathcal{G}_n^1 = \emptyset$ and using the above claims, we find

$$\begin{aligned} E_n &= \sum_{m=0}^{n} \mathrm{Prob}(G \in \mathcal{G}_n^m) Avg_{G \in \mathcal{G}_n^m} d(Aut(G)) \\ &\leq 1 + \sum_{m=2}^{n} 2^{-m(\frac{n}{2} - \frac{3m}{8} - 2\log n)} \\ &\leq 1 + 2^{-(n - 4\log n - 2)}, \end{aligned}$$

which proves the theorem. □

The proof of the theorem shows that the error in the asymptotic expression is very small.

COROLLARY 2. $\frac{2^{\binom{n}{2}}}{n!} \leq g_n \leq \frac{2^{\binom{n}{2}}}{n!}(1+\frac{4n^4}{2^n})$.

The next corollary follows from (2.3) (since $m = 1$ is impossible).

COROLLARY 3. *If a graph G has randomness deficiency slightly less than n (more precisely, $C(E(G)|n) \geq \binom{n}{2} - n - \log n - 2$) then G is rigid.*

The expression for g_n can be used to determine the maximal complexity of an unlabeled graph on n nodes. Namely, we can effectively enumerate all unlabeled graphs as follows:

- Effectively enumerate all labeled graphs on n nodes by enumerating all binary strings of length n and for each labeled graph G do the following:
  If G cannot be obtained by relabeling from any previously enumerated labeled graph then G is added to the set of unlabeled graphs.

This way we obtain each unlabeled graph by precisely one labeled graph representing it. Since we can describe each unlabeled graph by its index in this enumeration, we find by Theorem 3 and Stirling's formula that if G is an unlabeled graph then

$$C(E(G)|n) \leq \binom{n}{2} - n\log n + O(n).$$

THEOREM 4. *Let G be a labeled graph on n nodes and let G_0 be the unlabeled version of G. There exists a graph G' and a label permutation π such that $G' = \pi(G)$ and up to additional constant terms $C(E(G')) = C(E(G_0))$ and $C(E(G)|n) = C(E(G_0), \pi|n)$.*

By Theorem 4, for *every* graph G on n nodes with maximum complexity there is a relabeling (permutation) that causes the complexity to drop by as much as $n \log n$. Our proofs of topological properties by the incompressibility method required the graph G to be Kolmogorov random in the sense of $C(E(G)|n) \geq \binom{n}{2} - O(\log n)$ or for some results $C(E(G)|n) \geq \binom{n}{2} - o(n)$. Hence by relabeling such a graph we can always obtain a labeled graph that has a complexity too low to use our incompressibility proof. Nonetheless, topological properties do not change under relabeling.

REFERENCES

[1] N. ALON, J.H. SPENCER, AND P. ERDÖS, *The Probabilistic Method*, Wiley, 1992,

[2] E. BOREL, *Leçons sur la Théorie des Functions*, 2nd ed., Gauthier-Villars, Paris, 1914, pp. 182–216.

[3] BARANYAI, ZS., *On the factorization of the complete uniform hypergraph,* in Infinite and Finite Sets, Proc. Coll. Keszthely, A. Hajnal, R. Rado, and V.T. Sós, eds., Colloq. Math. Soc. János Bolyai 10, North-Holland, Amsterdam, 1995, pp. 91–108.

[4] B. BOLLOBÁS, *Graph Theory*, Springer-Verlag, New York, 1979.

[5] B. BOLLOBÁS, *Random Graphs*, Academic Press, London, 1985.

[6] H. BUHRMAN, J.H. HOEPMAN, AND P. VITÁNYI, *Space-efficient routing tables for almost all networks and the incompressibility method,* SIAM J. Comput., 28 (1999), pp. 1414–1432.

[7] P. ERDÖS AND J. SPENCER, *Probabilistic Methods in Combinatorics*, Academic Press, New York, 1974.

[8] W.W. KIRCHHERR, *Kolmogorov complexity and random graphs*, Inform. Process. Lett., 41 (1992), pp. 125-130.

[9] M. LI AND P.M.B. VITÁNYI, *An Introduction to Kolmogorov Complexity and Its Applications*, 2nd ed., Springer-Verlag, New York, 1997.

[10] M. LI AND P.M.B. VITÁNYI, *Kolmogorov complexity arguments in Combinatorics*, J. Combin. Theory Ser. A, 66 (1994), pp. 226–236. Errata, J. Combin. Theory Ser. A, 69 (1995), p. 183.

[11] M. LI AND P.M.B. VITÁNYI, *Statistical properties of finite sequences with high Kolmogorov complexity*, Math. Systems Theory, 27 (1994), pp. 365–376.

[12] F. HARARY AND E.M. PALMER, *Graphical Enumeration*, Academic Press, New York, London, 1973.

[13] A.N. KOLMOGOROV, *Three approaches to the quantitative definition of information*, Problems Inform. Transmission, 1 (1965), pp. 1–7.

[14] P. MARTIN-LÖF, *On the definition of random sequences*, Inform. and Control, 9 (1966), pp. 602–619.

[15] P.M.B. VITÁNYI, *Physics and the new computation*, in Proceedings of the 20th International Symposium on Math. Foundations of Computer Science, Prague, 1995, Lecture Notes in Comput. Sci. 969, Springer-Verlag, Heidelberg, 1995, pp. 106–128.