Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 969–970, February 1988
Mathematics

Quasi-random graphs

F. R. K. Chung[†], R. L. Graham[‡], and R. M. Wilson[§]

[†]Bell Communications Research, Morristown, NJ 07960; [‡]AT&T Bell Laboratories, Murray Hill, NJ 07974; and [§]California Institute of Technology, Pasadena, CA 91125

Contributed by R. L. Graham, October 23, 1987

ABSTRACT We introduce a large equivalence class of graph properties, all of which are shared by so-called random graphs. Unlike random graphs, however, it is often relatively easy to verify that a particular family of graphs possesses some property in this class.

Introduction

Perhaps the simplest model of generating a random graph G on n vertices is the process that successively considers each pair $\{v, v'\}$ of vertices of G and independently with probability 1/2 defines $\{v, v'\}$ to be an edge of G. More precisely, this process induces a uniform probability distribution on the set $\mathcal{G}(n)$ of all ordered graphs on n vertices, with each particular graph having probability $2^{-\binom{n}{2}}$. It often happens that for some graph property **P**, it is true that

$$\text{Prob}\{G \in \mathcal{G}(n) : G \text{ satisfies } \mathbf{P}\} \to 1 \text{ as } n \to \infty.$$

In this case, a "typical" graph in $\mathcal{G}(n)$, which we denote by $G_{1/2}(n)$, will have property **P** with overwhelming probability as $n \to \infty$.

Random graphs have been used extensively in a variety of contexts in graph theory and combinatorics and, in particular, are an integral component of the powerful "probability method" of Erdös (e.g., see refs. 1–3). Their effectiveness derives in part from the fact that it is often easier to prove that a given property actually holds for almost all graphs on n vertices (with this measure) than to prove that it holds for some specific graph. Indeed, there are numerous properties that are known to be possessed by almost all graphs but for which no explicit graph possessing them can (yet) be given. A well-known example is the property of having no complete or independent set of size $\log n/\log 2$.

The main thrust of this note is to announce the equivalence of a number of different graph properties, all shared by almost all $G \in \mathcal{G}(n)$, in the following sense: any graph satisfying *any one* of the properties must of necessity satisfy *all* of them. We call such graphs *quasi-random*. Moreover, many of these (equivalent) properties are quite easy to verify for specific graphs, thus allowing us to deduce random-like behavior for these graphs (we mention several such families at the end). We point out that these results share the same philosophy with some recent work of Thomason (4, 5) [who studied properties of "(p, α)-jumbled" graphs]; N. Alon and F.R.K.C. (unpublished results); Rödl (6); and P. Frankl, V. Rödl, and R.M.W. (unpublished results). Our motivation stemmed from early research of Wilson (7, 8).

Notation

Let $G = (V, E)$ denote a graph with vertex set V and edge set E (in general, we use the graph terminology in ref. 9). We use the notation $G(n)$ to denote a graph with n vertices, and we denote the number of edges of G by $e(G)$. For $X \subseteq V$, let $X|_G$ denote the subgraph induced by X, and let $e(X)$ denote the number of edges of $X|_G$. For $v \in V$, define

$$nd(v) : = \{x \in V : \{v, x\} \in E\}, \ \deg(v) : = |nd(v)|.$$

Further, if $G' = (V', E')$ is some other graph, we let $N^*_G(G')$ denote the number of labeled occurrences of G' as an induced subgraph of G. In other words,

$$N^*_G(G') = \{\lambda : V' \to V : \lambda(V)|_{G'} \cong G\}.$$

The related quantity $N_G(G')$ is defined to be the number of occurrences of G' as a (not necessarily induced) subgraph of G. Thus, $N_G(G') = \Sigma_H N^*_G(H)$, where the sum is taken over all $H = (V', E_H)$ with $E_H \supseteq E'$. Finally, the *adjacency matrix* $A(G) = [a(v, v')]_{v,v' \in V}$ of G is defined by setting $a(v, v') = 1$, if $\{v, v'\} \in E$, and 0 otherwise. We ordinarily order the eigenvalues λ_i of $A(G)$ [which must be real, since $A(G)$ is symmetric] so that $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$.

The Main Results

We next list a set of graph properties that a graph $G(n)$ might satisfy. Each property will contain in its statement occurrences of the asymptotic "little-oh" notation $o(\cdot)$. The dependence of different $o(\cdot)$ on the different properties they refer to will usually be suppressed, however. The use of these $o(\cdot)$ can be viewed in two essentially equivalent ways.

In the first way, suppose we have two properties **P** and **P′**, each with occurrences of $o(1)$, say. Thus, $\mathbf{P} = \mathbf{P}(o(1))$, $\mathbf{P}' = \mathbf{P}'(o(1))$. The implication "$\mathbf{P} \Rightarrow \mathbf{P}'$" then means that if each $o(1)$ in $\mathbf{P}(o(1))$ is replaced by a fixed (but arbitrary) function $f(n) = o(1)$ [i.e., $f(n) \to o$ as $n \to \infty$], then there is some other function $f'(n) = o(1)$ (depending on f) so that if $G(n)$ satisfies $\mathbf{P}(f(n))$ then it must also satisfy $\mathbf{P}'(f'(n))$. The particular choice made for f depends on the context, common ones being $n^{-1/2}$ and $1/\log n$ [when $f(n) = o(1)$].

In the second way, we can think instead of considering a family $\mathcal{F}$ of graphs $G(n)$ with $n \to \infty$. In this case, the interpretation of $o(1)$ is the usual one as $G = G(n)$ ranges over $\mathcal{F}$.

$\mathbf{P}_1(s)$: For all graphs $H(s)$ on s vertices,

$$N^*_G(H(s)) = [1 + o(1)]n^s 2^{-\binom{s}{2}}.$$

Thus, $\mathbf{P}_1(s)$ asserts that all $2^{\binom{s}{2}}$ labeled graphs $H(s)$ on s vertices occur asymptotically the same number of times in G [just as is the case for $G_{1/2}(n)$].

$$\mathbf{P}_2(t)\text{: } e(G) \geq [1 + o(1)]\frac{n^2}{4}, \ N_G(C_t) \leq [1 + o(1)]\left(\frac{n}{2}\right)^t,$$

where C_t denotes the cycle with t vertices.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

P$_3$: $e(G) \geq [1 + o(1)]\frac{n^2}{4}$, $\lambda_1 = [1 + o(1)]\frac{n}{2}$, $\lambda_2 = o(n)$.

P$_4(\varepsilon)$: For each subset $S \subseteq V$ with $|S| \geq \varepsilon n$,

$$e(S) = [1 + o(1)]\,\frac{|S|^2}{4}.$$

P$_5$: For each subset $S \subseteq V$ with $|S| = \lfloor n/2 \rfloor$,

$$e(S) = [1 + o(1)]\,\frac{n^2}{16}.$$

P$_6$: $$\sum_{v,v' \in V} \left| s(v, v') - \frac{n}{2} \right| = o(n^3),$$

where $s(v, v') := |\{y \in V : a(v, y) = a(v', y)\}|$ for $v, v' \in V$.

P$_7$: $$\sum_{v,v' \in V} \left| |nd(v) \cap nd(v')| - \frac{n}{4} \right| = o(n^3).$$

THEOREM. *The following properties are equivalent for a graph* G = G(n):

(a) P_1(s), *fixed* s $\geq$ 4;
(b) P_2(4);
(c) P_2(t), *fixed even* t $\geq$ 4;
(d) P_3;
(e) $P_4(\varepsilon)$, *fixed* $\varepsilon > 0$;
(f) P_5;
(g) P_6;
(h) P_7.

What is perhaps most surprising is how strong the (apparently weak) condition **P**$_2$(4) actually is. It implies in particular that if a graph G has about the same number of edges and 4-cycles that a random graph of the same size has, then in fact all fixed-size subgraphs must occur as induced subgraphs of G asymptotically equally often (as the size of G becomes large). Graphs having any (and therefore, all) of the above properties will be called quasi-random.

On the other hand, the requirements that t is even and that $s \geq 4$ are necessary, as shown by the following graph $G^*(4n)$. The vertex set of $G^*(4n)$ consists of four disjoint sets V_i, $1 \leq i \leq 4$. On V_1 and V_2 we have complete graphs. Between V_3 and V_4 we have a complete bipartite graph. Between $V_1 \cup V_2$ and $V_3 \cup V_4$ we place a random graph with edge probability 1/2. It is easily seen that $G^*(4n)$ satisfies **P**$_1$(3) and **P**$_2(2t + 1)$ for any fixed t but is not quasi-random. Similar considerations show the necessity of each of the other conditions in the statements of the properties.

Let us call a family $\mathcal{F}$ of graphs *forcing* if it is true that

$$N_{G(n)}(F) = [1 + o(1)]n^v 2^{-e}, \text{ for all } F = F(n, e) \in \mathcal{F},$$

implies $G(n)$ is quasi-random [where $F(n, e)$ denotes that F has n vertices and e edges]. For example, if P_t denotes the path with t vertices, and $K_{s,t}$ denotes the complete bipartite graph on vertex sets of sizes s and t, then it follows from the *Theorem* that $\{P_2, C_4\}$ is forcing, as are $\{C_{2s}, C_{2t}\}$, $s \neq t$; $\{P_2, K_{2,t}\}$, $t \geq 2$; and $\{K_{2,s}, K_{2,t}\}$, $s \neq t \geq 2$. On the other hand, $\{P_2, P_3, C_3\}$ is not forcing. It seems to be a challenging problem to characterize forcing families.

The proofs of the preceding results are rather lengthy and hinge on eigenvalue arguments and careful use of the so-called second moment method, and will be presented elsewhere. The same techniques can be used to establish the corresponding results for quasi-random graphs that imitate random graphs generated with a more general edge probability $p = p(n)$ (see also ref. 5). On the other hand, it would be of great interest to understand the corresponding situation for hypergraphs. Preliminary evidence indicates that in this case many of the analogous results no longer hold and that fundamental new insight will be needed in order to make significant progress here.

1. Erdös, P. & Spencer, J. (1974) *Probabilistic Methods in Combinatorics* (Akadémiai Kiadó, Budapest).
2. Bollobás, B. (1985) *Random Graphs* (Academic, New York).
3. Palmer, E. M. (1985) *Graphical Evolution* (Wiley, New York).
4. Thomason, A. (1987) in *Surveys in Combinatorics 1987*, LMS Lecture Notes Series 123, ed. Whitehead, C. (Cambridge Univ. Press, Cambridge, U.K.), pp. 173–196.
5. Thomason, A. (1987) *Ann. Discrete Math.* **33,** 307–331.
6. Rödl, V. (1986) *Discrete Math.* **59,** 125–134.
7. Wilson, R. M. (1972) *J. Number Theory* **4,** 17–47.
8. Wilson, R. M. (1974) *Math. Centre Tracts* **55,** 18–41.
9. Bondy, J. A. & Murty, U. S. R. (1976) *Graph Theory with Applications* (American Elsevier, New York).