Threshold functions for H-factors

Noga Alon *
Department of Mathematics
Raymond and Beverly Sackler Faculty of Exact Sciences
Tel Aviv University, Tel Aviv, Israel
and
Raphael Yuster
Department of Mathematics
Raymond and Beverly Sackler Faculty of Exact Sciences
Tel Aviv University, Tel Aviv, Israel

Abstract

Let H be a graph on h vertices, and let G be a graph on n vertices. An *H-factor* of G is a spanning subgraph of G consisting of n/h vertex disjoint copies of H. The *fractional arboricity* of H is $a(H) = max\{\frac{|E'|}{|V'|-1}\}$, where the maximum is taken over all subgraphs (V', E') of H with $|V'| > 1$. Let $\delta(H)$ denote the minimum degree of a vertex of H. It is shown that if $\delta(H) < a(H)$ then $n^{-1/a(H)}$ is a sharp threshold function for the property that the random graph $G(n, p)$ contains an H-factor. I.e., there are two positive constants c and C so that for $p(n) = cn^{-1/a(H)}$, almost surely $G(n, p(n))$ does not have an H-factor, whereas for $p(n) = Cn^{-1/a(H)}$, almost surely $G(n, p(n))$ contains an H-factor (provided h divides n). A special case of this answers a problem of Erdös.

*Research supported in part by a United States Israeli BSF grant

1 Introduction

All graphs considered here are finite, undirected and simple. If G is a graph of order n and H is a graph of order h, we say that G has an *H-factor* if it contains n/h vertex disjoint copies of H. Thus, for example, a K_2-factor is simply a perfect matching.

Let $G = G(n,p)$ denote, as usual, the random graph with n vertices and edge probability p. In the extensive study of the properties of random graphs, (see [5] for a comprehensive survey), many researchers observed that there are sharp *threshold functions* for various natural graph properties. For a graph property A and for a function $p = p(n)$, we say that $G(n,p)$ satisfies A *almost surely* if the probability that $G(n,p(n))$ satisfies A tends to 1 as n tends to infinity. We say that a function $f(n)$ is a *sharp threshold function* for the property A if there are two positive constants c and C so that $G(n, cf(n))$ almost surely does not satisfy A and $G(n, Cf(n))$ satisfies A almost surely.

Let H be a fixed graph with h vertices. Our concern will be to find the threshold function for the property that $G(n,p)$ contains an H-factor, (assuming, of course, that h divides n). In case $H = K_2$ this has been established by Erdös and Rényi in [7]. They showed that $\frac{log(n)}{n}$ is a sharp threshold function in this case, and there are many subsequent papers by various authors that supply more detailed information regarding this problem. In the general case, however, it is much harder to determine the threshold function. Even for the case $H = K_3$ the threshold is not known (cf. [3], Appendix B). In [3], page 243 P. Erdös raised the question of determining the threshold function in case $H = H_6$ is the graph on the 6 vertices $a_1, a_2, a_3, b_1, b_2, b_3$ whose 6 edges are a_1b_1, a_2b_2, a_3b_3 and a_1a_2, a_2a_3, a_1a_3. It turns out that in this case $n^{-2/3}$ is a sharp threshold function for the existence of an H-factor. In fact, the graph H_6 is just an element of a large family of graphs H for which we can determine a sharp threshold function for the existence of an H-factor. In order to define this family we need the following definition.

For a simple undirected graph H that contains edges, define the *fractional arboricity* of H as

$$a(H) = max\{\frac{|E'|}{|V'| - 1}\},$$

where the maximum is taken over all subgraphs (V', E') of H with $|V'| > 1$. Observe that by the well known theorem of Nash-Williams [13], $\lceil a(H) \rceil$ is just the arboricity of H, i.e., the minimum number of forests whose union covers all edges of H. Denote by $\delta(H)$ the minimum degree of a vertex of H, and let $\mathcal{F}$ be the family of all graphs H for which $a(H) > \delta(H)$ (or, equivalently, the family of all graphs with arboricity bigger than the minimum degree). Our main result is the following

Theorem 1.1 *Let H be a fixed graph in $\mathcal{F}$. Then the following two statements hold*

1. *There exists a positive constant c such that if $p = cn^{-1/a(H)}$ then almost surely $G(n,p)$ does not contain an H-factor.*

2. *There exists a positive constant C such that if $p = Cn^{-1/a(H)}$ then almost surely $G(n,p)$ contains an H-factor, assuming h divides n.*

Thus Theorem 1.1 asserts that for every $H \in \mathcal{F}$, $n^{-1/a(H)}$ is a sharp threshold function for the property that G contains an H-factor. In particular the theorem shows that $n^{-2/3}$ is a sharp threshold function in the special case $H = H_6$ mentioned above.

Our method yields the following extension of Theorem 1.1 as well.

Theorem 1.2 *Define the set $\mathcal{G}$ recursively as follows*

1. *$\mathcal{F} \subset \mathcal{G}$.*

2. *If C_1, C_2 are two of the connected components of some $H' \in \mathcal{G}$ and H is obtained from H' by adding to it less than $a(H')$ edges between C_1 and C_2 then $H \in \mathcal{G}$.*

If $H \in \mathcal{G}$ then $n^{-1/a(H)}$ is a sharp threshold function for the existence of an H-factor.

The proofs are presented in the next section. They rely on the Janson inequalities (cf. [6] and [8]), and on a method used by Alon and Füredi in [2].

The last section contains some concluding remarks and open problems.

2 The proofs

We begin by establishing the first statement in Theorem 1.1 which is not difficult and, in fact, holds even when H is not a member of $\mathcal{F}$.

Lemma 2.1 *Let H be any fixed graph that contains edges, and let $a = a(H)$. There exists a positive constant $c = c(H)$ such that almost surely $G(n,p)$ does not contain an H-factor for $p = cn^{-1/a}$.*

Proof Let $H' = (V', E')$ be any subgraph of H for which $\frac{|E'|}{|V'|-1} = a$. Denote $|E'| = e'$ and $|V'| = v'$. Let $\{A_i \ : \ i \in I\}$ denote the set of all distinct labeled copies of H' in the complete labeled graph on n vertices. Let B_i be the event that $A_i \subset G(n,p)$, and let X_i be the indicator random variable for B_i. Let $X = \sum_{i \in I} X_i$ be the number of distinct copies of H' in $G(n,p)$. It suffices to show that almost surely $X < \frac{n}{h}$. The expectation of X clearly satisfies

$$E[X] \leq \binom{n}{v'} v'!(cn^{-\frac{v'-1}{e'}})^{e'} \leq c^{e'} n$$

and yet clearly $E[X] = \Omega(n) \to \infty$. Choosing an appropriate constant c we obtain $E[X] < \frac{n}{2h}$. We next show that $Var[X] = o(E[X]^2)$. This suffices since by Chebyschev's inequality it implies that almost always $X < \frac{n}{h}$. For two copies A_i and A_j we say that $i \sim j$ if they share at least one edge. Let $\Delta = \sum_{i\sim j} Pr[B_i \wedge B_j]$, the sum taken over ordered pairs. Since $Var[X] \leq E[X] + \Delta$ and $E[X] \to \infty$, it remains to show that $\Delta = o(E[X]^2)$. The intersection of any A_i and A_j is a subgraph $H'' = (V'', E'')$ of H' (not necessarily an induced subgraph). We can therefore partition Δ into partial sums Δ'' corresponding to the various possible H''. It suffices to show that for each typical term Δ'' , $\Delta'' = o(E[X]^2)$. Denote $|V''| = v''$ and $|E''| = e''$. Then

$$\Delta'' \leq \binom{n}{v'} v'! \binom{n-v'}{v'-v''} (v'-v'')!(cn^{-\frac{v'-1}{e'}})^{(2e'-e'')}$$

$$\leq c^{2e'-e''} n^{2v'-v''-\frac{v'-1}{e'}(2e'-e'')} \leq c^{2e'-e''} n$$

since $\frac{v''-1}{e''} \geq \frac{v'-1}{e'}$. Hence $\Delta'' = o(E[X]^2)$. $\Box$

Note that by considering the minimal H' for which $\frac{|E'|}{|V'|-1} = a(H)$ we could obtain $\Delta = o(E[X])$ but our estimate suffices.

In order to prove the second part of Theorem 1.1 we need to state the Janson inequalities in our setting. Let Ω be a finite universal set and let R be a random subset of Ω where $Pr[r \in R] = p_r$, these events mutually independent over $r \in \Omega$. Let $\{A_i \ : \ i \in I\}$ be subsets of Ω, I a finite index set. Let B_i be the event $A_i \subset R$, and let X_i be the indicator random variable for B_i, and $X = \sum_{i \in I} X_i$ the number of $A_i \subset R$. For $i, j \in I$ we write $i \sim j$ if $i \neq j$ and $A_i \cap A_j \neq \emptyset$. We define

$$\Delta = \sum_{i \sim j} Pr[B_i \wedge B_j]$$

where the sum is taken over ordered pairs. Note that if the B_i were all independent then we would have $\Delta = 0$. The Janson inequalities state that when the B_i are "mostly" independent, then X is still close to a Poisson distribution with mean $\mu = E[X]$. The first inequality (cf. [6]) applies when Δ is small relative to μ,

Lemma 2.2 *Let B_i, Δ, μ be as above and assume that $Pr[B_i] \leq \epsilon$ for all i. Then*

$$Pr[X = 0] \leq exp(-\mu + \frac{1}{1-\epsilon}\frac{\Delta}{2})$$

When $\frac{\Delta}{2} \geq \mu(1-\epsilon)$ the bound in Lemma 2.2 is worthless. Even for Δ slightly less it is improved by the second Janson inequality (cf. [6])

Lemma 2.3 *Under the assumptions of Lemma 2.2 and the further assumption that $\Delta \geq \mu(1-\epsilon)$*

$$Pr[X = 0] \leq exp(-\frac{\mu^2(1-\epsilon)}{2\Delta})$$

The Janson inequalities play a crucial role in the proof of the next lemma.

Lemma 2.4 *If H is any fixed graph with $h-1$ vertices and fractional arboricity $a = a(H)$ then there exists a constant $C = C(H)$ such that almost surely $G(n,p)$ contains n/h vertex disjoint copies of H where $p = Cn^{-1/a}$.*

Proof It suffices to show that almost surely every subset of n/h vertices contains a copy of H. Fix such a subset of vertices, $A \subset \{1, 2, \ldots n\}$. We use a similar notation to that in the proof of Lemma 2.1. That is, X denotes the number of labeled copies of H in A, and Δ'' denotes the sum on all ordered pairs of copies of H whose intersection corresponds to a fixed subgraph H'' of H. However, this time we need to bound $\mu = E[X]$ from below;

$$\mu = E[X] \geq \binom{n/h}{h-1}(Cn^{-1/a})^e \geq (h(h-1))^{1-h}C^e n^{-e/a+h-1}$$

where e is the number of edges of H. Note that $E[X] = \Omega(n)$ since $\frac{e}{h-2} = \frac{e}{(h-1)-1} \leq a$. We now bound Δ'' from above,

$$\Delta'' \leq \binom{n/h}{h-1}(h-1)!\binom{n/h-(h-1)}{h-1-v''}(h-1-v'')!(Cn^{-1/a})^{2e-e''}$$
$$\leq C^{2e-e''}n^{2h-2-v''-(1/a)(2e-e'')}.$$

Claim: If $C > (h(h-1))^{2h-2}2^{h^2+1}$ then

$$\frac{\mu^2}{2\Delta} > n \tag{1}$$

Proof Note that $e'' \geq 1$ in each term Δ''. Hence

$$\frac{\mu^2}{\Delta''} \geq \frac{(h(h-1))^{2-2h}C^{2e}n^{-2e/a+2h-2}}{C^{2e-e''}n^{2h-2-v''-(1/a)(2e-e'')}}$$
$$\geq (h(h-1))^{2-2h}Cn^{v''-e''/a} \geq (h(h-1))^{2-2h}Cn \geq 2^{h^2+1}n.$$

There are less than 2^{h^2} subgraphs of H, so the last inequality (which holds for any Δ'') implies (1). This completes the proof of the claim.

Returning to the proof of the Lemma with C selected as in the above claim, we proceed as follows. If $\Delta < \mu$ we use Lemma 2.2. Note that in our case we may pick ϵ as an arbitrary small constant, and the lemma implies (since $\mu > 3n$ for our C),

$$Pr[X = 0] \leq exp(-\mu/3) \leq exp(-n)$$

If $\Delta > \mu$ we use Lemma 2.3. Picking $\epsilon = 0.1$ and using the above claim we obtain

$$Pr[X = 0] \leq exp(-0.9n)$$

In any case, $\binom{n}{n/h} Pr[X = 0]$ tends (even exponentially) to zero when n tends to infinity, and this completes the proof of the Lemma. $\Box$

Armed with Lemma 2.4 we can now complete the proof of Theorem 1.1. Given $H \in F$, let d be a vertex of minimal degree in H and denote the set of its neighbors by $N(d)$. Set $H' = H \setminus \{d\}$. Note that $a(H') = a(H)$ since $H \in \mathcal{F}$. We apply Lemma 2.4 by first setting $p' = C'n^{-1/a(H)}$ where C' is chosen as in Lemma 2.4. Almost every $G(n, p')$ will have n/h vertex disjoint copies of H'. We now need to match every remaining vertex in G to a copy of H' in such a way that there is an edge between the assigned vertex and each vertex in the set corresponding to $N(d)$ in the matched copy. We use (a modified version of) the method from [2] to do so. We choose the edges of $G(n, p)$ once again (but still keeping the edges of the first selection) with probability $p'' = n^{-1/a(H)}$. Note that this is the same probability space as $G(n, p)$ where $(1 - p) = (1 - p')(1 - p'')$. We define a random bipartite graph with one side being the n/h pairwise disjoint copies of H' and the other side being the remaining vertices of G. There is an edge of the bipartite graph between a copy and a remaining vertex if the vertex can be matched to the copy using only the new randomly chosen edges. The edge probability in this bipartite graph is $n^{-\delta/a}$ where δ is the degree of d. Moreover, crucially, the edges of this bipartite graph are chosen independently, since their existence is determined by considering pairwise disjoint subsets of edges of our random graph. Since $\delta < a$ it follows from the result in [7] that almost always there is a perfect matching. Note also that $p = p' + p'' - p'p''$ and since $p'' < p'$ we may generously set $C = 2C'$. This completes the proof of Theorem 1.1. $\Box$

Proof of Theorem 1.2: The fact that the threshold function for the existence of an H-factor is at least $n^{-1/a(H)}$ follows directly from Lemma 2.1. Let $H \in \mathcal{G}$. We must show that there is a constant $C = C(H)$ such that almost always $G(n, Cn^{-1/a(H)})$ contains an H-factor. We prove this by induction on the minimal number of applications of rule 2 in the definition of $\mathcal{G}$ needed to demonstrate the membership of H in $\mathcal{G}$. If no such application is needed then $H \in \mathcal{F}$ and the result follows from Theorem 1.1. Otherwise, there is an $H' \in \mathcal{G}$ and two connected components C_1 and C_2 of it such that H is obtained from H' by adding a set R of edges between C_1 and C_2, where $r = |R| < a(H')$. It is easy to check that this implies that $a(H) = a(H')$. By the induction hypothesis, there exists a constant C' such that almost surely $G(n, C'n^{-1/a(H)})$ contains an H'-factor. As in the proof of Theorem 1.1, we choose the edges of $G(n, p)$ once again with probability $n^{-1/a(H)}$. We define a random bipartite graph with one side being the n/h pairwise disjoint copies of C_1 and the other side being the n/h pairwise disjoint copies of C_2 (that are also pairwise disjoint with the copies of C_1). There is an edge in the bipartite graph between a vertex corresponding to a copy of C_1 and a vertex corresponding to a copy of C_2 if the edges corresponding to R exist in $G(n,p)$ among the freshly selected edges. The edge probability in this bipartite graph is $n^{-r/a(H)}$ and the choices of distinct edges are mutually independent. Since $r < a(H)$ it follows, as in Theorem 1.1, that $G(n,p)$ almost surely contains an H-factor, where $(1-p) = (1-C'n^{-1/a(H)})(1-n^{-1/a(H)})$. We can now set $C = C' + 1$ and $p = Cn^{-1/a(H)}$ to complete the proof. □

3 Concluding remarks

Somewhat surprising is the fact that there are many regular graphs H that fall into the category of Theorem 1.2. As an example, consider three arbitrary cubic graphs, and subdivide an edge in each of them. Add a new vertex and connect it to the vertices of degree 2 that were introduced by the subdivisions. The resulting graph H is cubic, and satisfies the properties of the graphs in Theorem 1.2, which supplies the appropriate threshold function for the existence of an H-factor.

Our theorems raise a natural algorithmic question. Suppose H is a graph in the family $\mathcal{G}$ defined in Theorem 1.2 and $a(H) = a$. Then, by the theorem, there is a positive constant $C = C(H)$ such that for $p(n) = Cn^{-1/a}$ the random graph $G(n,p)$ contains, almost surely, an H-factor, provided $|V(H)|$ divides n. Can we find such an H-factor efficiently? The proof easily supplies a polynomial time algorithm for every fixed H. Moreover, this algorithm can be parallelized. To see this observe that in the first step of the proof it suffices to find a *maximal* set of vertex disjoint copies of an appropriate graph H' in our random graph $G(n,p)$, where the maximality is with respect to containment. Such a set can be found in NC (i.e., in polylogarithmic time, using a polynomial number of parallel processors) using any of the known NC-algorithms for the maximal independent set problem (see, e.g., [10], [11], [1]). The rest of the algorithm only has to find perfect matchings in appropriately defined graphs, and this can be done in (randomized) NC by the results of [9] or [12]. Thus, the H-factors whose existence is guaranteed almost surely in Theorems 1.1 and 1.2 can actually be found, almost surely, efficiently (even in parallel).

The methods used in the proofs of the theorems can be used to compute the thresholds for the existence of spanning graphs other than H-factors. For example, let H be the 4 vertex graph consisting of a vertex of degree 1 joined to a triangle. Let Q be the graph obtained from $\frac{n}{4}$ pairwise disjoint copies of H, $H_1, \ldots, H_{\frac{n}{4}}$, where a_i is the vertex of degree 1 in H_i, by adding a cycle of length $\frac{n}{4}$ on the vertices a_i. By Theorem 1.1, $p = Cn^{-2/3}$ is a threshold for the existence of an H-factor. Suppose we now draw edges again with probability of $n^{-2/3}$ (which is much more than needed) in the subgraph of the $n/4$ vertices of degree 1. By the result of Pósa in [14] we will almost always have a Hamilton cycle in this subgraph. Therefore $n^{-2/3}$ is a sharp threshold function for the property that Q is a spanning subgraph of $G(n,p)$. Various similar examples can be given. Here, too, the proof is algorithmic, by applying the result of [4].

The following conjecture seems plausible.

Conjecture 3.1 *Let H be an arbitrary fixed graph with edges. Then the threshold for the property that $G(n,p)$ contains an H-factor, (if h divides n) is $n^{-1/a(H)+o(1)}$.*

We note that Lemma 2.1 shows that the above threshold is at least $n^{-1/a(H)}$. Also, the $o(1)$ term cannot be omitted entirely because, for example, $log(n)/n$ is the threshold for a K_2-factor, although $a(K_2) = 1$. Similarly, the threshold for a K_3-factor is at least $log(n)^{1/3}n^{-2/3}$ since as proved by Spencer [15] this is the threshold for every vertex to lie on a triangle, which is an obvious necessary condition in our case.

Note added in proof: We have recently learned that A. Ruciński, in *Matching and covering the vertices of a random graph by copies of a given graph*, Discrete Math. 105 (1992), 185-197, proved, independently (and before us), Theorem 1.1, using similar techniques. He did not prove the more general Theorem 1.2.

References

[1] N. Alon, L. Babai and A. Itai, *A fast and simple randomized parallel algorithm for the maximal independent set problem*, J. Algorithms 7 (1986), 567-583.

[2] N. Alon and Z. Füredi, *Spanning subgraphs of random graphs*, Graphs and Combinatorics 8 (1992), 91-94.

[3] N. Alon and J. H. Spencer, *The Probabilistic Method*, John Wiley and Sons Inc., New York, 1991.

[4] D. Angluin and L. Valiant, *Fast probabilistic algorithms for Hamilton circuits and matchings*, J. Computer Syst. Sci. 18, 155-193.

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

[6] R. B. Boppana and J. H. Spencer, *A useful elementary correlation inequality*, J. Combinatorial Theory Ser. A 50 (1989), 305-307.

[7] P. Erdös and A. Rényi, *On the existence of a factor of degree one of a connected random graph*, Acta Math. Acad. Sci. Hungar. 17 (1966), 359-368.

[8] S. Janson, *Poisson approximation for large deviations*, Random Structures and Algorithms 1 (1990), 221-230.

[9] R. M. Karp, E. Upfal and A. Wigderson, *Constructing a perfect matching in random NC*, Combinatorica 6 (1986), 35-48.

[10] R. M. Karp and A. Wigderson, *A fast parallel algorithm for the maximal independent set problem*, J. ACM 32 (1985), 762-773.

[11] M. Luby, *A simple parallel algorithm for the maximal independent set problem*, Siam J. Computing 15 (1986), 1036-1053.

[12] K. Mulmuley, U. V. Vazirani and V. V. Vazirani, *Matching is as easy as matrix inversion*, Proc. 19^{th} ACM STOC (1987), 345-354.

[13] C. St. J. A. Nash-Williams, *Decomposition of finite graphs into forests*, J. London Math. Soc. 39 (1964), 12.

[14] L. Pósa, *Hamiltonian circuits in random graphs*, Discrete Math. 14 (1976), 359-364.

[15] J. H. Spencer, *Threshold functions for extension statements*, J. Combinatorial Theory Ser. A, 53 (1990), 286-305.