ISRAEL JOURNAL OF MATHEMATICS, Vol. 39, No. 4, 1981

ON FACTORS IN RANDOM GRAPHS

BY

E. SHAMIR AND E. UPFAL

ABSTRACT

The following result is proved: Let $G_{n,p}$ be a random graph with n vertices and probability p for an edge. If p is such that the random graph has min-degree at least r with probability 1, then any f-factor $1 \leqq f \leqq r$ exists with probability 1, as $n \to \infty$.

1. Definitions and results

Let G be a graph with vertex set $V(G)$, edge set $E(G)$. Given a map f of $V(G)$ into the set of non-negative integers, define an f-factor of G as a spanning subgraph of G in which the degree (valency) of x is $f(x)$. When $f \equiv 1$, a one-factor is obtained, which is also called a perfect matching, i.e. a set of non-intersecting edges covering $V(G)$. One may want to study a maximal matching when a perfect matching does not exist.

We shall study the existence of f-factors in the context of random graphs, following the works of Erdös–Renyi [2, 3, 4]. Let $1, \cdots, n$ be a fixed labelling of the vertices. Let $\{e_{ij}\}$, $1 \leqq i < j \leqq n$, be an array of independent random variables, each e_{ij} assuming the value 1 with probability p, 0 with probability $1-p$. This array determines a random graph on $\{1, \cdots, n\}$, where (ij) is an edge if and only if $e_{ij} = 1$. This probability space is denoted by $G_{n,p}$. We allow p to be a function of n, and study the asymptotic behavior of probabilities of events in $G_{n,p}$ as $n \to \infty$. In particular, here we study the property of having an f-factor.

The principal results about properties of $G_{n,p}$ were obtained by Erdös and Rényi [2, 3, 4]. They use a somewhat different space $G_{n,N(n)}$ of graphs with N random edges. The passage from $G_{n,N(n)}$ to $G_{n,p}$ with $p(n) \leftrightarrow 2N(n)/n^2$ is usually quite simple [1].

THEOREM 1 [4]. *Let n be even, $p = (1/n)(\log n + w(n))$, with $\lim_{n\to\infty} w(n) = \infty$. Consider the event $E : G \in G_{n,p}$, G has a 1-factor. Then*

Received September 1, 1980

(1.1) $$\lim_{n\to\infty} \text{Prob}\, E = 1 \quad (\textit{or } E \textit{ holds a.a.s.}),$$

where a.a.s. *is an abbreviation for "asymptotically almost surely".*

THEOREM 2 [3]. *Let* $p = (1/n)(\log n + (r-1)\log\log n + w(n))$, $r \geqq 1$, $\lim_{n\to\infty} w(n) = \infty$. *Then in* $G_{n,p}$, $\text{Min}\deg G \geqq r$, a.s.s.

Our main result is the following

THEOREM 3. *Let p be as in Theorem* 2. *Let* $1 \leqq f(x_i) \leqq r$, $\Sigma_{i=1}^{n} f(x_i)$ *even. Then G has an f-factor,* a.a.s.

REMARK. Our proof will hold also for the case $p = (1/n)(\log n + c)$, where a.a.s. G consists of a huge component and some isolated points. A.a.s. there is a perfect matching [1-factor] in the huge component [if it is even].

Notice that by Theorem 2 we may assume that $\deg(x) \geqq r$, $x \in V(G)$. Also it is easy to see that the graphic condition, assuring the existence of a graph with degree sequence $f(x_i)$, is satisfied for large n.

2. Alternating paths, trees and augmentation

Existence of factors can be approached by Tutte's characterization theorem [4, 8]. For one-factor, this approach was followed in [4, 7]. Here we shall follow another method, using augmentation of sub-factors by alternating paths, used extensively in the algorithmic studies of matching and flow problems [1, 5]. In this respect it is closer to Posa's proof for a Hamiltonian path in random graphs [6].

A *sub-f-factor* of G is a subgraph M such that $E(M) \subseteq E(G)$ and $\deg_M(x) \leqq f(x)$. In case of strict inequality, the vertex x is *unsaturated.* Edges in $E(G)\setminus E(M)$ are *free* edges.

Consider a path $x_0, x_1, \cdots, x_m$ such that

$$x_{i-1}x_i \text{ is a free edge for } i \text{ odd}, \quad \text{an edge of } M \text{ for } i \text{ even}, \quad 0 < i \leqq m.$$

This is an *alternating path* with respect to M. A vertex with an odd [even] index is an I-vertex [T-vertex]. If m is odd, x_0 and x_m are unsaturated, such a path can be used to augment M: Drop $x_{i-1}x_i$, i even, add the free edge $x_{i-1}x_i$, i odd, $0 < i \leqq m$. The resulting M' is still a sub-f-factor, with a degree increased by 1 at x_0, x_m but unchanged otherwise.

We outline the proof of Theorem 3. Consider the event

(2.1) $\mathcal{N}\mathcal{A}$: There exists in G a sub-f-factor M, with an unsaturated vertex, which admits no augmentation.

$\mathcal{NA}$ implies, by a construction which is the core of the proof, the existence of two (probably) large disjoint trees Γ_x, Γ_y with edges in $E(G)$. Being disjoint implies disconnection between two large sets of vertices. Hence the probability of finding such trees in G is small, in fact $o(1)$ as $n \to \infty$.

Assume the event $\mathcal{NA}$ holds for G. Without loss of generality, there are two unsaturated vertices x, y (if say $\deg(x) \leqq f(x) - 2$, we add some xu and drop some uy). The trees Γ_x, Γ_y are obtained by a greedy algorithm, trying to catch many alternating paths from x and from y, while keeping a balance between Γ_x and Γ_y.

3. The parallel construction of the trees Γ_x, Γ_y and the auxiliary set N

The construction proceeds in steps. Vertices added at even [odd] steps are T vertices [I vertices], respectively.

Step 0. x is the root of Γ_x, y is the root of Γ_y.

$$N = \{w \mid zw \in E(M) \text{ for } z = x \text{ or } z = y\}.$$

Step 1. $A = \{w \mid w \notin N,\ wx \text{ or } wy \text{ is a free edge}\} = A_x \cup A_y \cup B$ (disjoint union) where for $z = x, y$, $A_z \subseteq A$ is the set of w's which are connected to Γ_z only, B is the set connected to both. Let $\bar{A}_z \subseteq A_z$ be a maximal set such that two vertices in $\bar{A}_z$ are not connected by an edge of M. Extend Γ_z by connecting a set $\Delta_z \subset \bar{A}_z \cup B$ such that

$$\Delta_x \cap \Delta_y = \varnothing, \quad |\Delta_x| = |\Delta_y| = \lfloor \tfrac{1}{2}|B| \rfloor + \operatorname{Min}(|\bar{A}_x|, |\bar{A}_y|). \tag{3.1}$$

An edge of M connecting a vertex u in B with a vertex w in B, A_x or A_y, or an edge connecting a vertex u in A_x to a vertex w in A_y closes an alternating path between x and y, hence an augmentation for M. Thus the construction of Δ_x and Δ_y gives the maximum possible additions which are disjoint, equal and no edge of M connects two vertices in $\Delta_x \cup \Delta_y$.

$\vdots$

Step j, j even. Connect to Γ_z any w such that

$$(\exists u) \quad (u \text{ an } I\text{-vertex of } \Gamma_z,\ uw \in E(M)),$$

under the provision that the set of w's connected to both trees is split evenly between them (if a single one remains, it is connected to the smaller tree; if they are equal, it is connected to Γ_x).

Note: the vertex w is new (outside $\Gamma_x \cup \Gamma_y$). Indeed, w was not added in the $(j-1)$ step since vw is an edge of M.

If w had been added in a previous odd [even] step in $\Gamma_x \cup \Gamma_y$ as an I-vertex [T-vertex], then u would be a T-vertex [I-vertex with $u \in N$], which is impossible.

Add to N all vertices connected to the (newly introduced) T-vertices by edges of M.

Step j, j odd. $A = \{w \mid w \notin N$, w is new and connected to a T-vertex in $\Gamma_x \cup \Gamma_y$ by a free edge$\} = A_x \cup A_y \cup B$ (disjoint union).

The decomposition of A and how to extend Γ_z by new I-vertices is done precisely as described in Step 1 above.

The construction terminates when it is impossible to extend the trees and preserve equality in the number of I-vertices in Γ_x, Γ_y. Upon termination all T-vertices of one of the trees are connected only to vertices in $\Gamma_x \cup \Gamma_y \cup N$ (and perhaps one more vertex in case (3.1) was 0 at the terminating odd step, since $|\bar{A}_z| = 0$ if and only if $|A_z| = 0$).

4. The probable size of the trees

CLAIM 1. *$\mathcal{NA}$ implies that the trees Γ_z, $z = x, y$, each have at least two T-vertices,* a.a.s.

PROOF. The root z itself is one T-vertex. Being unsaturated, z has one free edge which is not xy (else M can be augmented). Each Γ_z has an I-vertex u_z. Otherwise x and y each has a single free edge, which is connected to the same vertex u (and each has $< r$ edges of M). The probability of this event is estimated by

$$\binom{n}{3}\binom{n}{2r-2}\left(\frac{\log n + w(n)}{n}\right)^{2r}\left(1 - \frac{\log n + w(n)}{n}\right)^{2(n-2r-1)}$$
$$\leqq n(\log n + w(n))^{2r} e^{-2(\log n + w(n))} = O(n^{-1}).$$

Now if Γ_z does not contain another T-vertex (connected to u_z), then all the I-vertices of $\Gamma_x \cup \Gamma_y$ are connected by M to one and the same vertex w. Only up to r vertices may be connected to w by M, hence $|\Delta_x \cup \Delta_y| \leqq r$ in the first step and $2 \leqq |A_x \cup A_y \cup B| \leqq r^2$ (since $r|\bar{A}_z| \geqq |A_z|$). Thus x and y are connected to at most $r^2 + 2r$ vertices and at least two of these vertices have a common neighbour. We can express the existence of such a configuration in a graph $G \in G_{n,p}$ and estimate its probability by

$$\sum_{k=2}^{r^2+2r} \binom{n}{2}\binom{n}{k}\binom{n}{1}\left(\frac{\log n + w(n)}{n}\right)^{k+2}\left(1-\frac{\log n + w(n)}{n}\right)^{2(n-k-1)}$$

$$\leqq \sum_k n(\log n + w(n))^{k+2} e^{-2\log n}$$

Let K be the number of I-vertices in Γ_z.

CLAIM 2. *Each tree has at least $K/2r$ and at most $rK+1$ T-vertices.*

PROOF. Each I-vertex is connected in M to at least one T-vertex. Each T-vertex is connected in M to at most r vertices. Thus after the splitting between trees, each tree has at least $K/2r$ T-vertices (including its root).

Each T-vertex (except for the root) is connected in M to at least one I-vertex. Each I-vertex is connected in M to at most r vertices. Hence there are at most $rK+1$ T-vertices.

CLAIM 3. *Each tree has at most $3rK$ vertices.*

PROOF. There are K I-vertices, at most $rK+1$ T-vertices.

CLAIM 4. $|N| \leqq 4r^2K$.

PROOF. Multiply the estimate for T-vertices in $\Gamma_x \cup \Gamma_y$ by r.

CLAIM 5. *Each tree has at least $n/80r^5$ T-vertices*, a.a.s.

PROOF. Let Γ_x be the tree which caused the construction to terminate. All its T-vertices are connected to $\Gamma_x \cup \Gamma_y \cup N$ and perhaps one more vertex s. Consider $\Gamma \subset (\Gamma_x \cup \Gamma_y \cup N \cup \{s\})$ which is the tree obtained from Γ_x by (i) reconnecting all vertices which are connected to a T-vertex of Γ_x, but went to Γ_y upon the splitting of B, (ii) connecting a $w \in N$ which is connected to a T-vertex by an edge of M, (iii) connecting s. Note that Γ is indeed a tree, we added vertices to the leaves of Γ_x with one connection each (as we did throughout).

$$|\Gamma| \leqq |\Gamma_x \cup \Gamma_y \cup N \cup \{s\}| \leqq 10r^2K.$$

Let t be the number of T-vertices in Γ_x. Since by Claim 1, $t \geqq 2$ and by Claim 2, $t \geqq k/2r$,

$$t \geqq \operatorname{Max}(2, \lceil |\Gamma|/20r^3 \rceil).$$

Consider the event

E: There is a tree Γ with $2 \leqq l \leqq n/2$ vertices in which t vertices, $t = \operatorname{Max}(2, \lceil l/20r^3 \rceil)$, are not connected in G to a vertex outside Γ.

$$\frac{1}{20r^3}\cdot \operatorname{Prob} E \leqq \sum_{2\leqq t\leqq n/40r^3} \binom{n}{t} t^t \left(\frac{\log n + w(n)}{n}\right)^{t-1} 2^t \left(1-\frac{\log n + w(n)}{n}\right)^{t(n-t)}$$

$$\leqq \sum_t n(2e\log n + w(n))^t \exp\left(-\frac{t(n-t)}{n}(\log n + w(n))\right)$$

$$\leqq \sum_{2\leqq t\leqq \log n} + \sum_{\log n\leqq t\leqq n/40r^3} = \sum{}' + \sum{}'';$$

$$\sum{}' \leqq n(2e\log n + w(n))^{40r^3} e^{-2\log n}, \quad \sum{}'' \leqq n^2(\log n)^{\log n} e^{-\log^2 n/2}.$$

Thus Γ_x has at least $n/40r^3$ T-vertices a.a.s. Hence, by Claim 2, $K \geqq n/40r^4$, and so Γ_y also has at least $n/80r^5$ T-vertices.

5. Conclusion of the proof

CLAIM 6. $\lim_{n\to\infty} \operatorname{Prob}(\mathcal{N}\mathcal{A}) = 0$.

PROOF. Let $q = n/\sqrt{\log n}$. Take a set A of T-vertices in Γ_x, $|A| = q$. It is connected by M to at most rq vertices. Thus Γ_y contains another set D of q vertices which (since $\mathcal{N}\mathcal{A}$ holds) has no connection in G to A. Indeed a free edge connecting a T-vertex b in Γ_x to a T-vertex c in Γ_y closes an alternating path between Γ_x and Γ_y, hence an augmentation for M. The probability that such A and D exist is bounded by

$$\binom{n}{q}\binom{n-q}{q}\left(1-\frac{\log n + w(n)}{n}\right)^{q^2} \leqq \left(\frac{ne}{q}\right)^{2q} \exp\left[-\frac{q^2}{n}(\log n + w(n))\right]$$

$$\leqq \left(\sqrt{\log n}\, e^{-\sqrt{\log n}}\right)^{n/\sqrt{\log n}},$$

which has a sub-exponential decrease, as $n \to \infty$.

REFERENCES

1. D. Angluin and L. Valiant, *Fast probabilistic algorithm for Hamiltonian circuits and matchings*, J. Comput. Syst. Sci. **18** (1979), 155–193.

2. P. Erdös and A. Rényi, *On the evolution of random graphs*, Publ. Math. Inst. Hungar. Acad. Sci. **5A** (1960), 17–61.

3. P. Erdös and A. Rényi, *On the strength of connectedness of a random graph*, Acta Math. Acad. Sci. Hungar. **12** (1961), 261–267.

4. 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.

5. E.L. Lawler, *Combinatorial Optimization: Networks and Matroids*, Holt, Rinehart and Winston, 1976.

6. L. Pósa, *Hamiltonian circuits in random graphs*, Discrete Math. **14** (1976), 359–364.

7. E. Shamir and E. Upfal, *One factor in random graphs based on vertex choice*, submitted, 1980.

8. W. T. Tutte, *The subgraph problem*, Ann. Discrete Math. 3 (1978), 289–295.

INSTITUTE OF MATHEMATICS
THE HEBREW UNIVERSITY OF JERUSALEM
JERUSALEM, ISRAEL

DEPARTMENT OF APPLIED MATHEMATICS
THE WEIZMANN INSTITUTE OF SCIENCE
REHOVOT, ISRAEL