Limit Distribution for the Existence of Hamiltonian Cycles in Random Bipartite Graphs

A. M. FRIEZE

A random bipartite graph D with 2n vertices is generated by allowing each of the n^2 possible edges to occur with probability $p = (\log n + \log \log n + c_n)/n$. We show that

$$\lim_{n\to\infty} P(D \text{ contains a Hamiltonian cycle}) = \begin{cases} 0 & c_n \to -\infty \\ e^{-2e^{-c}} & c_n \to c \\ 1 & c_n \to \infty \end{cases}$$

Introduction

Komlós and Szemerédi [2] showed that if the edges of a random labelled graph G(n, p) on n vertices are drawn independently with probability $p = p_n = (\log n + \log \log n + c_n)/n$ and HAM denotes the event that G(n, p) has a Hamiltonian cycle, then

$$\lim_{n \to \infty} P(\text{HAM}) = \begin{cases} 0, & c_n \to -\infty, \\ e^{-e^{-c}}, & c_n \to c, \\ 1, & c_n \to +\infty \end{cases}$$

$$= \lim_{n \to \infty} P(\text{D2})$$
(1.1)

where D2 is the necessary event that each vertex of G(n, p) has degree at least 2.

Independently Korsunov [3] proved the same result for $c_n \to +\infty$.

This tightened Posá's result [5] that $p = a \log n/n$ for a sufficiently large is enough to ensure that G(n, p) is almost surely Hamiltonian.

An elegant result of McDiarmid [4] shows that if D(n, p) is a random vertex labelled digraph with n vertices in which each arc is drawn independently with probability p then

$$P(D(n, p) \text{ is Hamiltonian}) \ge P(G(n, p) \text{ is Hamiltonian})$$
 (1.2)

from which one can, for example, show that D(n, p) is almost surely Hamiltonian if $c_n \to +\infty$ above.

In this paper we look at random vertex labelled bipartite graphs B(n, p) in which there are 2n vertices partitioned into 2 sets V_1 and V_2 of size n and in which the edges are drawn independently with probability p. It is very pleasing, though perhaps not surprising, that a result similar to (1.1) can be proved.

2. MAIN RESULT

For ease of reference we next list some notation and define some events needed later.

NOTATION. Let G be a graph. V(G), E(G) denote the vertex and edge sets of G, respectively.

For $S \subseteq V(G)$, $d_G(S) = |\{w \notin S: (v, w) \in E(G) \text{ for some } v \in S\}| \text{ and for } v \in V, d_G(v) = \text{the degree of } v \text{ in } G.$

A path P of G has no repeated edges, distinct endpoints and l(P) edges. A cycle Chas I(C) edges.

$$L(G) = \max(l(P): P \text{ is a path of } G)$$

We will be concerned with a bipartite graph BG with vertex partition V_1 , V_2 where $n = |V_1| = |V_2|$ always.

We assume that the vertices in V_1 are painted black and that the vertices in V_2 are painted white. Terms like black sets, white sets and unichromatic sets have their obvious

The following lemma describes some properties of BG that hold almost surely.

Lemma 1. Assume $\log \log n + c_n \to \infty$. Let a vertex be small if $d_{BG}(v) \le \log n/10$ and large otherwise.

The following hold almost surely:

- (i) n log n ≤ |E(BG)| ≤ 2n log n,
- (ii) $d_{BG}(v) \leq 4 \log n$ for all $v \in V_1 \cup V_2$,
- (iii) BG contains fewer than $n^{1/2}$ small vertices.
- (b) Let $F \subseteq E(BG)$, be such that no small vertex is incident with an edge of F and no large vertex is incident with more than $\log n/200$ edges of F. Then
 - (iv) $H = (V_1 \cup V_2, E(BG) F)$ is connected,
 - (v) $d_H(S) \ge 2|S| n_1$ for all unichromatic S, $1 \le |S| \le 2n/5$ $|\{v \in V_1 \cup V_2 : d_{BG}(v) \le 1\}|.$

PROOF (OUTLINE). (a) can be proved by routine calculation and (b) follows easily once one has established properties analogous to Section 1 of Komlós and Szemerédi [2].

Let N denote the set of BG satisfying the conditions of Lemma 1. We now define the following events:

 $BG \in D2$ if and only if $d_{BG}(v) \ge 2$, for all $v \in V$.

 $BG \in LC$ if and only if a longest cycle of BG has as many vertices as a longest path of BG.

 $BG \in ODD$ if and only if L(BG) is odd.

 $BG \in EVEN$ if and only if L(BG) is even.

 $BG \in HAM$ if and only if BG contains a Hamiltonian cycle.

We now give some lemmas, whose proofs are either omitted or left until after the proof of the main theorem.

In the following lemmas, the probability of an edge being included in BG is p = $(\log n + \log \log n + c_n)/n$.

LEMMA 2.

$$\lim_{n\to\infty} P(\mathrm{BG}\in\mathrm{D2}) = \begin{cases} 0, & c_n\to-\infty,\\ \mathrm{e}^{-2\epsilon^{-c}}, & c_n\to c,\\ 1, & c_n\to+\infty. \end{cases}$$

PROOF. Use inclusion-exclusion as in Erdős and Rényi [1].

LEMMA 3.

$$\lim_{n\to\infty}P(\mathrm{BG}\in N\cap\mathrm{D2}\cap\overline{\mathrm{LC}})=0,\qquad\text{if }\log\log\,n+c_n\to+\infty.$$

PROOF. Given later.

The main result follows easily from these lemmas.

THEOREM.

$$\lim_{n\to\infty} P(\mathrm{BG}\in\mathrm{HAM}) = \begin{cases} 0, & c_n\to-\infty, \\ \mathrm{e}^{-2e^{-c}}, & c_n\to c, \\ 1, & c_n\to+\infty. \end{cases}$$

PROOF. We first note that the probabilities in Lemma 2 are obviously upper bounds for the probability of BG being Hamiltonian. We can therefore assume $c_n \not = -\infty$. On the other hand $LC \cap N \cap D2 \subseteq HAM$ because

(1) $BG \in N \cap D2$ implies that BG is connected. (put $F = \emptyset$ and $n_1 = 0$ in Lemma 1(b)).

(2) Any connected subgraph in LC is Hamiltonian, for if a longest cycle C of BG was not a Hamiltonian cycle then we could derive a longer path from the fact that C is connected to the rest of the graph.

The rest then follows easily from Lemmas 1, 2, 3:

$$P(BG \in HAM) \ge P(BG \in LC \cap N \cap D2)$$
, by the above,
= $P(BG \in N \cap D2) - o(1)$, by Lemma 3,
= $P(BG \in D2) - o(1)$, by Lemma 1.

Now use Lemma 2.

We turn now to the proof of Lemma 3. We first give a form of a result of Posá on the endpoints of a set of longest paths in a graph that has been known to T. I. Fenner and the author for some time, but has not, as yet, found any application.

Let $P = (v_0, v_1, \ldots, v_k)$ be a longest path in a graph G. Then if $(v_k, v_t) \in E(G)$ where t < k-1 then we find that $P' = (v_0, v_1, \ldots, v_t, v_k, v_{k-1}, \ldots, v_{t+1})$ is also a longest path.

We say that P' is obtained from P by a flip. There may be several ways of flipping P and we can obviously generate many longest paths by sequences of flips.

Starting with $P_0 = P$ above we derive a sequence of longest paths P_0 , P_1 , P_2 , ... all with v_0 as one endpoint. The other endpoint w_i of P_i is the one distinct from v_0 . At any stage of our procedure we will have produced a sequence $\sigma_m = (P_0, P_1, P_2, \ldots, P_m)$, the first s of which will have been scanned. Initially we have $\sigma_0 = (P_0)$ with P_0 unscanned. In general we take the first unscanned path P_{s+1} ; if, however, s = m we terminate this process. Let Q_1, Q_2, \ldots, Q_r be the paths that can be generated from P_{s+1} by flipping. We add to the sequence σ_m any path whose other endpoint is not a member of $W_m = \{w_0, w_1, \ldots, w_m\}$. Let $END(v_0) = W_m$ when we terminate, which must happen eventually as W_m cannot grow indefinitely.

LEMMA 4. Let v_t be a vertex of P, $v_t \notin END(v_0)$ and suppose that there exists $w \in END(v_0)$ such that $(v_t, w) \in E(G)$.

Then
$$\{v_{t-1}, v_{t+1}\} \cap END(v_0) \neq \emptyset$$
 (assume $v_{-1} = v_1$).

PROOF. Let $s = \min(r; (w_r, v_t) \in E(G))$. If $w_s \notin \{v_{t-1}, v_{t+1}\}$ then clearly $\{v_{t-1}, v_{t+1}\} \cap W_s = \emptyset$. But this means that the edge (w_s, v_t) can be used to flip P_s . But the neighbour(s) of v_t on P_s must be v_{t-1}, v_{t+1} . For if not, the sequence of flips used to obtain P_s must have deleted one of the edges (v_{t-1}, v_t) , (v_t, v_{t-1}) . But when an edge is deleted one of its vertices becomes an endpoint. Thus in this case one of v_{t-1}, v_t, v_{t+1} has already been an endpoint. The lemma follows from this contradiction.

COROLLARY.

$$d_G(\text{END}(v_0)) < 2|\text{END}(v_0)|$$

Now for any $v \in \text{END}(v_0)$ there is a longest path of G with v as an endpoint. We may clearly carry out the same construction keeping v as a fixed endpoint, thus creating a set END(v) of other endpoints to v.

In summary, we create a set $END = \{v_0\} \cup END(v_0)$ such that for each $v \in END$ there is a set END(v) satisfying

$$d_G(END(v)) < 2|END(v)|$$
 (2.1a)

If $v \in END$ and $w \in END(v)$ then v and w are the endpoints of some longest path (2.1b)

In the next lemma we have a bipartite graph $BG \in EVEN$. Let END etc. be as above. For $v \in END$ let $\Phi(v)$ be the vertex adjacent to v on all of the paths produced during the construction of END(v). The definition of flip and the fact that these longest paths have an even number of edges justifies this definition.

LEMMA 5. If $X = \Phi(\text{END})$ then $x \in X$ implies $|\Phi^{-1}(x)| \le 2$ and hence $|X| \ge |\text{END}|/2$.

PROOF. Let P be the original longest path used to start our construction. Let $v \in END$ and $x = \Phi(v)$. Our result will follow if we show that v and x are adjacent on P. If $v = v_0$ then (v_0, x) is a terminal edge of P. If $v \neq v_0$ let $Q = (v_0, \ldots, x, v)$ be the first path generated that has v as an endpoint. If (x, v) is not an edge of P then (x, v) was added during a previous flip. But then one out of x and v was already an endpoint at this stage. Since I(P) is even, all vertices in END are the same colour as v_0 , which implies $x \notin END(v_0)$ and hence that v has already been an endpoint—contradiction. Thus (x, v) is an edge of P and our result follows.

The arguments used in previous work depend on showing that END is large and that for each $v \in \text{END}$, END(v) is large and that there are enough edges to ensure that with high probability there is an edge of the form (v, w) where $w \in \text{END}(v)$. However in the bipartite case if L(BG) is even then obviously this cannot be done. Overcoming this difficulty is the main problem solved in this paper. In fact it suffices to prove

LEMMA 6.

$$\lim_{n\to\infty} P(BG \in EVEN \cap D2 \cap N) = 0.$$

PROOF. Let $p_2 = a/(n \log n)$ where a = 305 (it is preferable to carry a around in formulae so that one can easily see later why a particular value was chosen) and let $p_1 = (p - 2p_2 + p_2^2)/(1 - p_2)^2$. We generate the edges of BG as follows: E_b is a random subset of $V_1 \times V_2$ where $e \in V_1 \times V_2$ is independently included in E_b with probability p_1 and excluded with probability $1 - p_1$; E_g is a random subset of $V_1 \times V_2 - E_b$ with inclusion probability p_2 ; E_y is a random subset of $V_1 \times V_2 - (E_b \cup E_g)$ with inclusion probability p_2 . $E(BG) = E_b \cup E_g \cup E_y$. E_b , E_g , E_y are referred to as blue, green and yellow edges respectively. E(BG) is the graph $(V_1 \cup V_2, E_b)$.

One can easily confirm that the edge probability of BG is p as required. Such a colouring of BG is said to be good and BG is said to be well-coloured if

$$L(BG_b) = L(BG).$$
 (2.2a)

no large vertex has more than $\log n/200$ incident edges that are coloured green or yellow. (2.2c)

Let GOOD denote the event that the colouring chosen is good. The crux of the proof is the following pair of inequalities which hold for large n:

$$P(GOOD|BG \in N \cap D2 \cap EVEN) \ge \left(1 - \frac{3a}{(\log n)^2}\right)^n.$$
 (2.3)

$$P((BG \in N \cap D2 \cap EVEN) \cap GOOD) \le e^{-a^2n/76 \log^2 n}$$
 (2.4)

It follows immediately from (2.5) and (2.6) that

$$P(BG \in EVEN \cap D2 \cap N) \le \left(1 - \frac{3a}{(\log n)^2}\right)^{-n} e^{-a^2/76 \log^2 n}$$

 $\le e^{(4a - a^2/76)n/\log^2 n}$

for large n. The lemma follows immediately.

PROOF of 2.3. We shall prove the stronger result that for any $BG_0 \in N$.

$$P(GOOD|BG = BG_0) \ge \left(1 - \frac{3a}{(\log n)^2}\right)^n$$
 (2.5)

from which (2.3) follows easily.

Let P be any longest path of BG_0 . Routine calculations show that the r.h.s. of (2.5) is a lower bound for the probability that the edges of P are blue and (2.2b), (2.2c) hold.

PROOF OF 2.4. We first note that condition (2.2b) and (2.2c) of a good colouring ensure via Lemma 1 that if $BG \in N \cap D2$ is well-coloured then for large $n, H = BG_b$ satisfies

$$S \subseteq V$$
, S unichromatic and $|S| \le 2n/5$ implies $d_H(S) \ge 2|S|$. (2.6a)

Let us now write

 $P(GOOD \cap (BG \in N \cap D2 \cap EVEN))$

$$= \sum_{H \in \Omega} P(GOOD \cap (BG \in N \cap D2 \cap EVEN) | BG_b = H) P(BG_b = H), \quad (2.7)$$

where Ω is the set of graphs with n vertices which can be derived from a graph in $N \cap D2 \cap EVEN$ by deleting edges.

Now let H be fixed member of Ω . We will show that for large n

$$P_H = P(GOOD \cap (BG \in N \cap D2 \cap EVEN) | BG_b = H) \le e^{-a^2n/76 \log^2 n}$$
(2.8)

from which (2.4) follows, on using (2.7).

We next describe the probability P_H in the following way: given $H \in \Omega$, let BG be obtained by adding random edges $X = E_g \cup E_y$ to H. Then

$$P_H = P((a) \text{ BG} \in N \cap D2 \cap \text{EVEN},$$

(b)
$$L(H) = L(BG)$$
,

Now clearly $P_H = 0$ if $H \notin EVEN$, using conditions (a) and (b), and by the above, $P_H = 0$

also if H does not satisfy (2.6). So assume now that $H \in EVEN$ and H satisfies (2.6). Let

$$Q_H = P(L(H) = L(BG))$$

Clearly $P_H \leq Q_H$ and we shall show that

$$Q_H \leq e^{-a^2n/76 \log^2 n}$$
(2.9)

from which (2.8) and the lemma follows.

Now instead of adding X to H all at once, we add random edges E_g to H to create a graph H'' and then add further random edges to H'' to create BG.

Let OUT = $\{v: \text{ there exists a longest path } P \text{ of } H'' \text{ such that (i) } v \text{ is not }$ a vertex of P, (ii) the endpoints of P are both coloured differently to v}

We show next that

$$P((|OUT| < an/30 \log n) \cap (L(H'') = L(H)) \le e^{-an/120 \log n}$$
 (2.10)

$$P((L(BG) = L(H)) \cap (|OUT| \ge an/30 \log n)) \le e^{-a^2n/75 \log^2 n}.$$
(2.10)

Since L(BG) = L(H) implies L(H'') = L(H) we deduce from this pair of inequalities that

$$P(L(BG) = L(H)) \le e^{-an/120 \log n} + e^{-a^2n/75 \log^2 n}$$

which implies (2.9) for large n.

PROOF OF (2.10). Let P be a longest path of H and let END, END(v) for $v \in END$ and the function Φ be as defined in Lemmas 4 and 5. As (2.6) holds we know from (2.1a) that $|\text{END}| \ge 2n/5$ and further from Lemma 5 that $|X| \ge n/5$ where $X = \Phi(\text{END})$.

For $x \in X$ let $FIN(x) = \bigcup_{v \in \Phi^{-1}(x)} END(v)$ and let A(x) be the event: there exists $w \in FIN(x)$ such that $(x, w) \in E(H'')$. The important point to note is that if A(x) occurs and L(H'') = L(H) then $\Phi^{-1}(x) \subseteq OUT$. To see this, suppose $x \in X$, $y \in \Phi^{-1}(x)$ and L(H'') = L(H). Let $Q = (v_0, v_1, \dots, v_k)$ be a longest path of H obtained from P by a sequence of flips such that $v_0 = y$, $v_1 = x$ and $(x, v_k) \in E(H'')$. As l(Q) is even, there is a vertex z of a different colour to y and not lying in Q. As H is connected there is a path R from z to some vertex v_t of Q not containing any other vertex of Q. Now $t \neq 0$ otherwise Q is not a longest path of H''. If C is the cycle $(v_1, v_2, \ldots, v_k, v_1)$ let Q_1 be the path obtained by deleting the edge (v_i, v_{i+1}) of Q and let Q_2 be the catenation of Q_1 and R. Clearly $l(Q_2) \ge l(Q)$ and so Q_2 is a longest path of H''. As v_0 is not on Q_2 and the endpoint z of Q_2 is of a different colour to v_0 we have $v_0 \in OUT$ as was to be shown.

We show next that with high probability, A(x) occurs for a large number of x.

Suppose first $x \in X$ and there does not exist $(x, w) \in E(H)$ with $w \in FIN(x)$. Since $|FIN(x)| \ge 2n/5$ we find that $Pr(A(x)) \ge 1 - (1 - p_2)^{2n/5} \ge a/3 \log n$ for n large. On the other hand for $x \in X$ with $(x, w) \in E(H)$, $w \in FIN(x)$ we have Pr(A(x)) = 1. These events are all independent and so using $|X| \ge n/5$ and standard inequalities for the tails of the binomial distribution we have

$$\Pr(|\{x \in X : A(x) \text{ occurs}\}| \leq an/30 \log n) \leq e^{-an/120 \log n}$$

which proves (2.14). To complete the proof of our lemma it only remains now to give the proof of (2.11):

PROOF OF (2.11). Assume L(H'') = L(H). If $x \in OUT$ let P(x) denote some longest path of H'' not passing through x. Let B(x) denote the set of endpoints obtainable by a sequence of flips as in Lemma 4. Note that $|B(x)| \ge 2n/5$ because of (2.1a) and (2.6). Now L(BG) = L(H'') only if none of the edges added to H'' join an $x \in OUT$ to some $y \in B(x)$. But the probability of this occurring is clearly no more than

$$(1-p_2)^{2an^2/150\log n} < e^{-a^2n/75\log^2 n}$$

which proves (2.11) and completes the lemma

The proof of our theorem can now be completed with:

PROOF OF LEMMA 2. Because of Lemma 6 we need only prove

$$\lim_{n\to\infty} P(BG \in N \cap D2 \cap ODD \cap \overline{LC}) = 0$$
 (2.12)

We use an edge colouring argument as in Lemma 6, but things are fortunately much simpler and much of the proof can be lifted from the previous proof. We construct BG as in Lemma 6, but now we can absorb E_y into E_g to make a blue-green graph with approximately twice as many green edges as before. Let BG_b be as before and let a good colouring be as defined in (2.2). The proof of (2.5) goes through as before.

To prove Lemma 3 we have only to prove

$$P(GOOD \cap (BG \in N \cap D2 \cap ODD \cap \overline{LC})) \leq e^{-an/4 \log n}$$
 (2.13)

and then (2.12) will follow immediately. Now

 $P(GOOD \cap (BG \in N \cap D2 \cap ODD \cap \overline{LC}))$

$$= \sum_{H \in \Omega'} P(GOOD \cap (BG \in N \cap D2 \cap ODD \cap \overline{LC}) | BG_b = H) P(BG_b = H)$$
 (2.14)

where Ω' is the set of graphs with n vertices which can be derived from a graph in $N \cap D2 \cap ODD \cap \overline{LC}$ by deleting edges. Now let H be a fixed member of Ω' . We will show that for large n

$$P_{H} = P(GOOD \cap (BG \in N \cap D2 \cap ODD \cap \overline{LC})|BG_{b} = H) < e^{-an/4 \log n}$$
 (2.15)

from which (2.13) follows on using (2.14).

We next describe the probability of P_H in the following way: given $H \in \Omega'$, let BG be obtained from H by adding random edges E_g to H. Then

$$P_H = P((a) \text{ BG} \in N \cap D2 \cap ODD \cap \overline{LC},$$

(b)
$$L(H) = L(BG)$$
,

Now clearly $P_H = 0$ if $H \notin ODD$ and also $P_H = 0$ if H does not satisfy (2.6). So assume now that $H \in ODD$ and that H satisfies (2.6). Let

$$Q_H = P((a) L(H) = L(BG) \text{ and } (b) BG \in \overline{LC}) \ge P_{H}$$

Now let P be any longest path of H and let the sets END, END(v) for $v \in END$ be as defined in Lemma 4. As (2.6) holds we deduce that these sets are all at least 2n/5 in size. Now in order to have L(H) = L(BG) and $BG \in \overline{LC}$ the following event must occur: no edge of X joins a vertex $v \in END$ to a vertex $w \in END(v)$.

But the probability of this happening is less than or equal to $(1-p_2)^{2(4n^2/25-2n\log n)}$ and the lemma follows.

We note that McDiarmid's results apply equally well to random bipartite graphs and hence (1.2) is valid when D(n, p) is a random bipartite digraph. We note also that it is

straightforward to modify this proof to give one for (1.1). In particular we do not need to prove Lemma 6 in this case.

REFERENCES

- 1. P. Erdős and A. Rényi, On random matrices, Publ. Math. Inst. Hungar. Acad. Sci. 8 (1964), 455-461.
- 2. J. Komlós and E. Szemerédi, Limit distribution for the existence of hamiltonian cycles in a random graph, Discrete Math. 43 (1983), 55-63.
- 3. A. D. Korsunov, Solution of a problem of Erdős and Rényi on hamiltonian cycles in non-oriented graphs, Soviet Math. Dokl. 17 (1976), 760-764.
- 4. C. McDiarmid, General percolation and random graphs Adv. Appl. Prob. 13 (1981), 40-60.
- 5. L. Posá, Hamiltonian circuits in random graphs, Discrete Math. 14 (1976), 359-364.

Received 12 November 1982 and in revised form 16 July 1984

A. M. FRIEZE

Department of Computer Science and Statistics, Queen Mary College, Mile End Road, London E1 4NS, England