Small Maximal Matchings in Random Graphs

Michele Zito*

Department of Computer Science, University of Liverpool, Liverpool L69 7ZF, UK

Abstract. We look at the minimal size of a maximal matching in general, bipartite and d-regular random graphs. We prove that the ratio between the sizes of any two maximal matchings approaches one in dense random graphs and random bipartite graphs. Weaker bounds hold for sparse random graphs and random d-regular graphs. We also describe an algorithm that with high probability finds a matching of size strictly less than $n/2$ in a cubic graph. The result is based on approximating the algorithm dynamics by a system of linear differential equations.

1 Introduction

A matching in a graph is a set of disjoint edges. Several optimisation problems are definable in terms of matchings. If G is a graph and M is a matching in G, we count the number of edges in M and the goal is to maximise this value, then the corresponding problem is that of finding a maximum cardinality matching in G. This problem has a glorious history and an important place among combinatorial problems [2, 5, 8]. However few other matching problems share its nice combinatorial properties. If $G = (V, E)$ is a graph, a matching $M \subseteq E$ is *maximal* if for every $e \in E \setminus M$, $M \cup e$ is not a matching; $V(M) = \{v : \exists u \{u, v\} \in M\}$. Let $\beta(G)$ denote the minimum cardinality of a maximal matching in G. The minimum maximal matching problem is that of finding a maximal matching in G with $\beta(G)$ edges. The problem is NP-hard [10]. The size of any maximal matching is at most $2\beta(G)$ [6] in general graphs and at most $\left(2 - \frac{1}{d}\right)\beta(G)$ [11] in regular graphs of degree d. Some negative results are known about the approximability of $\beta(G)$ [11].

In this paper we abandon the pessimistic point of view of worst-case algorithmic analysis by assuming that each input graph G occurs with a given probability. Nothing seems to be known about the most likely value of $\beta(G)$ or the effectiveness of any approximation heuristics in this setting. In Section 2 we prove that the most likely value of $\beta(G)$ can be estimated quite precisely, for instance, if G is chosen at random among all graphs with a given number of vertices. Similar results are proved in Section 3 for dense random bipartite graphs. Also, simple algorithms exist which, *with high probability* (w.h.p.), that is with probability approaching one as $n = |V(G)|$ tends to infinity, return matchings of size $\beta(G) + o(n)$. Lower bounds on $\beta(G)$, improving the ones presented above, are proved also in the case when higher probability is given to graphs with few edges. Most of the bounds on $\beta(G)$ are obtained by exploiting a simple relation between maximal matchings and independent sets. In Section 4 we investigate the possibility of applying a similar reasoning if G is a random d-regular graph. After

* Supported by EPSRC grant GR/L/77089.

showing a number of lower bounds on $\beta(G)$ for several values of d, we present an algorithm that finds a maximal matching in a d-regular graph. We prove that with high probability it returns a matching of size asymptotically less than $n/2$ if G is a random cubic graph.

In what follows $\mathcal{G}(n,p)$ ($\mathcal{G}(K_{n,n},p)$) denotes the usual model of random (bipartite) graphs as defined in [1]. Also $\mathcal{G}(n, d\text{-reg})$ denotes the following model for random d-regular graphs [9, Section 4]. Let n urns be given, each containing d balls (with dn even): a set of $dn/2$ pairs of balls (called a *configuration*) is chosen at random among those containing neither pairs with two balls from the same urn nor couples of pairs with balls coming from just two urns. To get a random $G \in \mathcal{G}(n, d\text{-reg})$ let $\{i,j\} \in E(G)$ if and only if there is a pair with one ball belonging to urn i and the other belonging to urn j. If $\mathcal{G}$ is a random graph model, $G \in \mathcal{G}$ means that G is selected with a probability defined by $\mathcal{G}$. The random variable $X = X_k(G)$ counts the number of maximal matchings of size k in G. The meaning of the sentences "*almost always* (a.a.)", "for *almost every* (a.e.) graph" is defined in [1, Ch. II].

2 General Random Graphs

Let $q = 1 - p$. If U is a random indicator $\Pr[U]$ will denote $\Pr[U = 1]$.

Theorem 1. *If* $G \in \mathcal{G}(n,p)$ *then* $\mathrm{E}(X) = \binom{n}{2k} \frac{(2k)!}{k!} \left(\frac{p}{2}\right)^k q^{\binom{n-2k}{2}}$.

Proof. Let M_i be a set of k independent edges, assume that G is a random graph sampled according to the model $\mathcal{G}(n,p)$ and let $X^i_{p,k}$ be the random indicator equal to one if M_i is a maximal matching in G. $\mathrm{E}(X^i_{p,k}) = \Pr[X^i_{p,k}] = p^k q^{\binom{n-2k}{2}}$. Then by linearity of expectation

$$\mathrm{E}(X) = \sum_{|M_i|=k} \mathrm{E}(X^i_{p,k}) = |\{M_i : |M_i| = k\}| \cdot p^k q^{\binom{n-2k}{2}}$$

The number of matchings of size k is equal to the possible ways of choosing $2k$ vertices out of n times the number of ways of connecting them by k independent edges divided by the number of orderings of these chosen edges. □

A lower bound on $\beta(G)$ is obtained by bounding $\mathrm{E}(X)$ and then using the Markov inequality to prove that $\Pr[X > 0]$ approaches zero as the number of vertices in the graph becomes large. Assuming $2k = n - 2\omega$

$$\mathrm{E}(X) \le \frac{n^{\frac{n}{2}-\omega}}{(2\omega)!} \left(\frac{p}{2}\right)^{\frac{n}{2}-\omega} q^{2\omega^2-\omega} \le \left(\frac{pn}{2}\right)^{\frac{n}{2}} \left(\frac{\mathrm{e}}{npq\omega}\right)^{\omega} q^{2\omega^2}$$

and this goes to zero only if $\omega = \Omega(\sqrt{n})$. However a different argument gives a considerably better result.

Theorem 2. $\beta(G) > \frac{n}{2} - \frac{\log n}{\log(1/q)}$ *for a.e.* $G \in \mathcal{G}(n,p)$ *with* p *constant.*

Proof. If M is a maximal matching in G then $V \setminus V(M)$ is an independent set. Let $Z = Z_{p,2\omega}$ be the random variable counting independent sets of size $2\omega = \frac{2\log n}{\log(1/q)}$ in a random graph G. If X counts maximal matchings of size $k = \frac{n}{2} - \omega$,

$$\begin{aligned}\Pr[X>0] &= \Pr[X>0 \mid Z>0]\Pr[Z>0] + \Pr[X>0 \mid Z=0]\Pr[Z=0] \\ &\le \Pr[X>0 \mid Z>0]\Pr[Z>0] + 0\cdot 1 \le \Pr[Z>0] \to 0\end{aligned}$$

The last result follows from a theorem in [4] on the independence number of dense random graphs. Thus $\beta(G) > \frac{n}{2} - \frac{\log n}{\log 1/q}$ for a.e. $G \in \mathcal{G}(n,p)$. □

The argument before Theorem 2 is weak because even if $\mathrm{E}(Z_{p,2\omega})$ is small $\mathrm{E}(X)$ might be very large. The random graph G might have very few independent sets of size 2ω but many maximal matchings of size $\frac{n}{2} - \omega$.

Results in [4] also have algorithmic consequences. Grimmett and McDiarmid considered the simple greedy heuristic which repeatedly places a vertex v in the independent set I if there is no $u \in I$ with $\{u,v\} \in E(G)$ and removes it from G. It is easily proved that $|I| \sim \frac{\log n}{\log(1/q)}$.

Theorem 3. $\beta(G) < \frac{n}{2} - \frac{\log n}{2\log(1/q)}$ *for a.e.* $G \in \mathcal{G}(n,p)$ *with* p *constant.*

Proof. Let $\mathcal{IS}$ be an algorithm that first finds a maximal independent set I in G using the algorithm above and then looks for a perfect matching in the remaining graph. With probability approaching one $|I| \ge (1-\delta)\frac{\log n}{\log(1/q)}$ for all $\delta > 0$. Also, $\mathcal{IS}$ does not expose any edge in $G - I$. Hence $G - I$ is a completely random graph on about $n - |I|$ vertices, each edge in it being chosen with constant probability p. Results in [3] imply that a.a. such graphs contain a matching with at most one unmatched vertex. □

Independent sets are useful also for sparse graphs. If $p = \frac{c}{n}$ a lower bound on $\beta(G)$ can be obtained again by studying $\alpha(G)$, the size of a largest independent set of vertices in G.

Theorem 4. $\beta(G) > \frac{n}{2} - \frac{n \log c}{c}$ *for a.e.* $G \in \mathcal{G}(n, c/n)$, *with* $c > 2.27$.

Proof. $\alpha(G) < \frac{2n\log c}{c}$ for a.e. $G \in \mathcal{G}_{n,c/n}$ for $c > 2.27$ [1, Theorem XI.22]. The result follows by an argument similar to that of Theorem 2. □

If $p = \frac{c}{n}$ for c sufficiently small, the exact expression for $\mathrm{E}(X)$ in Theorem 1 gives an improved lower bound on $\beta(G)$. Roughly, if c is sufficiently small and U is a large independent set in G then the graph induced by $V \setminus U$ very rarely contains a perfect matching.

Theorem 5. $\beta(G) > \frac{n}{3}$ *for a.e.* $G \in \mathcal{G}_{n,c/n}$, *with* $c \in (2.27, 16.99]$

Proof. Let $k = \frac{n}{2} - \frac{dn}{c}$. If $d \in \left(\frac{c}{6}, \frac{c}{2}\right)$ then $k < n/3$. Hence $\frac{n!}{(n-2k)!\ k!} \le n^{2k}/k!$ and

$$\mathrm{E}(X) \le O(1)\cdot\sqrt{\frac{c}{\pi(c-2d)n}}\left(\frac{c^2\mathrm{e}}{(c-2d)n}\right)^{\frac{n}{2}-\frac{dn}{c}} \mathrm{e}^{-\frac{2d^2n}{c}+d}$$

which goes to zero for every d in the given range. The best choice of d is the smallest and the theorem follows by noticing that $\frac{\log c}{c} < \frac{1}{6}$ if $c > 16.9989$. □

3 Bipartite Graphs

The results in the last section can be extended to the case when $G \in \mathcal{G}(K_{n,n}, p)$. Again $\beta(G)$ is closely related to a graph parameter whose value, at least in dense random graphs, can be estimate rather well. Given a bipartite graph $G = (V_1, V_2, E)$ with $|V_1| = |V_2| = n$, a *split independent set* in G is a set of 2ω independent vertices S with $|S \cap V_i| = \omega$. Let $\sigma(G)$ be the size of a largest split independent set in G. If M is a maximal matching in a bipartite graph G then $V \setminus V(M)$ is a split independent set.

Theorem 6. *If $G \in \mathcal{G}(K_{n,n}, p)$ then*

1. $\mathrm{E}(X) = \binom{n}{k}^2 k! p^k q^{(n-k)^2}$.
2. *If $Z = Z_{p,n-k}$ is the random variable counting split independent sets of size $n-k$ and $Y = Y_{p,k}$ is the random variable counting perfect matchings in $H \in \mathcal{G}(K_{k,k}, p)$ then* $\mathrm{E}(X) = \mathrm{E}(Z) \cdot \mathrm{E}(Y)$.

Proof. Let M_i be a set of k independent edges and $G \in \mathcal{G}((K_{n,n}, p)$ and let $X^i_{p,k}$ be the random indicator equal to one if M_i is a maximal matching in G. $\mathrm{E}(X^i_{p,k}) = \Pr[X^i_{p,k}] = p^k q^{(n-k)^2}$. Then

$$\mathrm{E}(X) = \sum_{|M_i|=k} \mathrm{E}(X^i_{p,k}) = |\{M_i : |M_i| = k\}| \cdot p^k q^{(n-k)^2}$$

The number of matchings of size k is given by the possible ways of choosing k vertices out of n on each side times the number of permutations on k elements. □

If p is constant, it is fairly easy to bound the first two moments of Z and get good estimates on the value of $\sigma(G)$.

Theorem 7. $\sigma(G) \sim \frac{4 \log n}{\log 1/q}$ *for a.e. $G \in \mathcal{G}(K_{n,n}, p)$ with p constant.*

Proof. The expected number of split independent sets of size 2ω is $\binom{n}{\omega}^2 q^{\omega^2}$. Hence, by the Markov inequality, and Stirling's approximation to the factorial $\Pr[Z > 0] < \left(\frac{n^\omega}{\omega!}\right)^2 q^{\omega^2}$ and the right side tends to zero as n grows if $2\omega = 2\left\lceil \frac{2 \log n}{\log 1/q} \right\rceil$.

Let $2\omega = 2\left\lfloor \frac{2(1-\epsilon)\log n}{\log 1/q} \right\rfloor$ for any $\epsilon > 0$. The event "$Z = 0$" is equivalent to "$\sigma(G) < 2\omega$" because if there is no split independent set of size 2ω then the largest of such sets can only have less than 2ω elements. By the Chebyshev inequality $\Pr[Z = 0] \leq \mathrm{Var}(Z)/\mathrm{E}(Z)^2$. Also $\mathrm{Var}(Z) = \mathrm{E}(Z^2) - \mathrm{E}(Z)^2$. There are $s_\omega = \binom{n}{\omega}^2$ ways of choosing ω vertices from two disjoint sets of n vertices. If Z^i is the random indicator set to one if S^i is a split independent set in G then $Z = \sum Z^i$ and $\mathrm{E}(Z^2) = \sum_{i,j} \Pr[Z^i \wedge Z^j] = \sum_{i,j} \Pr[Z^j] \sum_i \Pr[Z^i \mid Z^j]$ where the sums are over all $i, j \in \{1, \ldots, s_\omega\}$. Finally by symmetry $\Pr[Z^i \mid Z^j]$ does not actually depend on j but only on the amount of intersection between S^i and S^j. Thus, if $S^1 = \{1, \ldots, 2\omega\}$,

$$\mathrm{E}(Z^2) = \left(\sum_j \Pr[Z^j]\right) \left(\sum_i \Pr[Z^i \mid Z^1]\right) = \mathrm{E}(Z) \cdot \mathrm{E}(Z|Z^1).$$

Thus to prove that $\Pr[Z = 0]$ converges to zero it is enough to show that the ratio $\mathrm{E}(Z|Z^1)/\mathrm{E}(Z)$ converges to one. By definition of conditional expectation

$$\mathrm{E}(Z|Z^1) = \sum_{0 \leq l_1, l_2 \leq \omega} \binom{\omega}{l_1}\binom{\omega}{l_2}\binom{n-\omega}{\omega-l_1}\binom{n-\omega}{\omega-l_2} q^{\omega^2 - l_1 l_2}$$

Define T_{ij} (generic term in $\mathrm{E}(Z|Z^1)/\mathrm{E}(Z)$) by

$$T_{ij}\binom{n}{\omega}^2 = \binom{w}{i}\binom{w}{j}\binom{n-\omega}{\omega-i}\binom{n-\omega}{\omega-j}q^{-ij}$$

Tedious algebraic manipulations prove that $T_{00} \le 1 - \frac{2\omega^2}{n-\omega+1} + \frac{\omega^3(2\omega-1)}{(n-\omega+1)^2}$, and, for sufficiently large n, $T_{ij} \le \frac{\omega^2}{n-\omega+1}$ for $i+j=1$, and $T_{ij} \le T_{10}$ for all $i,j \in \{1,\ldots,\omega\}$. From these results it follows that

$$\begin{aligned}\Pr\left[\sigma < 2\left\lfloor \frac{2(1-\epsilon)\log n}{\log 1/q}\right\rfloor\right] &\le T_{00} + T_{10} + T_{01} + \omega^2 T_{10} - 1 \\ &\le 1 - 2\omega^2/n + \frac{\omega^3(2\omega-1)}{(n-\omega+1)^2} + \frac{2\omega^2}{n} + \frac{\omega^4}{n} - 1 \\ &\le \frac{\omega^3(2\omega-1)}{(n-\omega+1)^2} + \frac{\omega^4}{n}\end{aligned}$$

□

Theorem 8. *$\beta(G) > n - \frac{2\log n}{\log 1/q}$ for a.e. $G \in \mathcal{G}(K_{n,n},p)$ with p constant.*

The similarities between the properties of independent sets in random graphs and those of split independent sets in random bipartite graphs have some algorithmic implications. A simple greedy heuristic almost always produces a solution whose cardinality can be predicted quite tightly. Let I be the independent set to be output. Consider the process that visits the vertices of a random bipartite graph $G(V_1,V_2,E)$ in some fixed order. If $V_i = \{v_1^i,\ldots,v_n^i\}$, then the algorithm will look at the pair (v_j^1, v_j^2) during step j. If $\{v_j^1, v_j^2\} \notin E$ and if there is no edge between v_j^i and any of the vertices which are already in I then v_j^1 and v_j^2 are inserted in I. Let $\sigma_g(G) = |I|$.

Theorem 9. *$\sigma_g(G) \sim \frac{\log n}{\log 1/q}$ for a.e. $G \in \mathcal{G}(K_{n,n},p)$ with p constant.*

Proof. Suppose that $2(k-1)$ vertices are already in I. The algorithm above will add two vertices v_1 and v_2 as the kth pair if $\{v_1,v_2\} \notin E$ and there is no edge between either v_1 or v_2 and any of the vertices which are already in I. The two events are independent in the given model and their joint probability is $(1-p)\cdot(1-p)^{2(k-1)} = (1-p)^{2k-1}$. Let W_k (for $k \in \mathbb{N}^+$) be the random variable equal to the number of pairs considered before the kth pair is added to I. W_k has geometric distribution with parameter $P_k = (1-p)^{2k-1}$. Moreover the variables $W_1, W_2, \ldots$ are all independent. Let $Y_\omega = \sum_{k=1}^{\omega} W_k$. The event "$Y_\omega < n$" is implied by "$\sigma_g(G) > 2\omega$": if the split independent set returned by the greedy algorithm contains more than 2ω vertices that means that the algorithm finds ω independent pairs in strictly less than n trials. Also if $Y_\omega < n$ then certainly each of the W_k cannot be larger than n. Hence

$$\Pr[Y_\omega < n] \le \Pr[\cap_{k=1}^{\omega}\{W_k \le n\}] = \prod_{k=1}^{\omega}\{1 - [1-(1-p)^{2k-1}]^n\}$$

Let $\omega = \left\lceil \frac{(1+\epsilon)\log n}{2\log 1/q}\right\rceil$ and, given $\epsilon > 0$ and $r \in \mathbb{N}$, choose $m > r/\epsilon$. For sufficiently large n, $\omega - m > 0$. Hence

$$\Pr[Y_\omega < n] \le \prod_{k=\omega-m}^{\omega}\{1 - [1-(1-p)^{2k+1}]^n\}$$

that is at most $\{1-[1-(1-p)^{2(\omega-m)+1}]^n\}^m$. Since $(1-x)^n \ge 1-nx$, we also have $\Pr[Y_\omega < n] \le \{n(1-p)^{2(\omega-m)+1}\}^m = o(n^{-r})$. The event "$Y_\omega > n$" is equivalent to "$\sigma_g(G) < 2\omega$". Let $\omega = \left\lfloor \frac{(1-\epsilon)\log n}{2\log 1/q}\right\rfloor$. If $Y_\omega > n$ then there must be at least one k for which $W_k > n/\omega$. Hence $\Pr[Y_\omega > n] \le \Pr[\cup_{k=1}^{\omega}\{W_k > n/\omega\}]$ and this is at most

$$\sum_{k=1}^{\omega} \Pr[W_k > n/\omega] \leq \omega[1-(1-p)^{2\omega-1}]^{\lfloor n/\omega \rfloor}.$$

By the choice of ω, $(1-p)^{2\omega-1} > \frac{n^{-(1-\epsilon)}}{1-p}$. Hence

$$\Pr[Y_\omega > n] \leq \omega \left[1 - \frac{n^{-(1-\epsilon)}}{1-p}\right]^{\lfloor n/\omega \rfloor} \leq \omega \ \exp\left\{-\frac{n^{-(1-\epsilon)}}{1-p} \left\lfloor \frac{n}{\omega} \right\rfloor\right\}$$

Finally $\Pr[Y_\omega > n] \leq \omega \exp\left\{-\frac{n^\epsilon}{(1-p)\omega} - o(1)\right\}$ since $\lfloor n/\omega \rfloor > n/\omega - 1$, and the result follows from the choice of ω. □

The greedy algorithm analysed in Theorem 9 does not expose any edge between two vertices that are not selected to be in I. Therefore $G - I$ is a random graph. Classical results ensure the existence of a perfect matching in $G - I$, and polynomial time algorithms exist which find one such a matching. We have proved the following.

Theorem 10. *$\beta(G) < n - \frac{\log n}{2 \log 1/q}$ for a.e. $G \in \mathcal{G}(K_{n,n}, p)$ with p constant.*

4 Regular Graphs

In this section we look at the size of the smallest maximal matchings in random regular graphs. Again known upper bounds on the independence number of such graphs imply, in nearly all interesting cases, good lower bounds on $\beta(G)$.

Theorem 11. *For each $d \geq 3$ there exists a constant $\gamma(d)$ such that $\beta(G) \geq \gamma(d)n$ for a.e. $G \in \mathcal{G}(n, d\text{-reg})$.*

Proof. It is convenient to use the configuration model described in the introduction. Two pairs of balls in a configuration are *independent* if each ball is chosen from a distinct urn. A matching in a configuration is a set of independent pairs. The expected number of maximal matchings of size k in a random configuration is

$$\frac{n!\ d^{2k}}{k!\ (n-2k)!} \frac{[2k(d-1)]!}{[k(2d-1)-nd/2]!} \frac{(dn/2)!}{(dn)!} 2^{d(n-2k)}$$

If $k = \gamma n$, using Stirling's approximation to the factorial, this is at most

$$f(\gamma, d)^n = \left\{ \left(\frac{1}{1-2\gamma}\right) \left[\frac{d(1-2\gamma)^2}{\gamma}\right]^\gamma \frac{[\gamma(d-1)]^{2\gamma(d-1)}}{[\gamma(2d-1)-d/2]^{\gamma(2d-1)}} \left[\frac{\gamma(2d-1)-d/2}{d}\right]^{\frac{d}{2}} \ 2^{\frac{d}{2}-2\gamma} \right\}^n$$

For every d there exists a unique $\gamma_1(d) \in \left(\frac{d}{2(2d-1)}, \frac{1}{2}\right)$ for which $f(\gamma, d) \geq 1$, for $\gamma \in (\gamma_1(d), 0.5)$. Since the probability that a random configuration corresponds to a d-regular graph is bounded (see for example [1, Chap 2]), the probability that a random d-regular graph has a maximal matching of size γn is at most $f(\gamma, d)^n$. If $d > 6$ a better bound is obtained by using $\gamma(d) = (1 - \alpha_3(d))/2$ where $\alpha_3(d)$ is the smallest value in $(0, 1/2)$ such that $\alpha(G) < \alpha_3(d)n$ for a.a. $G \in \mathcal{G}(n, d\text{-reg})$ [7]. □

The relationship between independent sets and maximal matchings can be further exploited also in the case where $G \in \mathcal{G}(n, d\text{-reg})$, but random regular graphs are rather sparse graphs and the approach used in the previous sections cannot be easily applied in this context. However, a simple greedy algorithm which finds a large independent in a d-regular graph can be modified and incorporated in a longer procedure that finds a small maximal matching in a random regular graph. Consider the following algorithm $\mathcal{A}$.

Input: Random d-regular graph with n vertices

(1) $M \leftarrow \emptyset$;

(2) **while** there is a vertex of degree d **do**

    choose v in V u.a.r. among the vertices of degree d;

    $M \leftarrow M \cup \{v, u_1\}$; /* Assume $N(v) = \{u_1, \ldots, u_{\deg_G v}\}$ */

    $V \leftarrow V \setminus \{v\}$;

(3) **for** $j = 1$ **to** $d - 1$ **do**

    choose v u.a.r. among the vertices of degree $d - j$ in $V(M)$;

    $V \leftarrow V \setminus \{v\}$;

(4) find a maximal matching M' in what is left of G;

(5) make $M \cup M'$ into a maximal matching for G.

Step (2) essentially mimics one of the algorithms presented in [9], with the only difference that instead of selecting an independent set of vertices, the process selects a set of edges. Step (4) can be clearly performed in polynomial time. In general the set $M \cup M'$ is an *edge dominating set* (each edge in G is adjacent to some edge in $M \cup M'$) but it is not necessarily a matching. However [10] any edge dominating set F can be transformed in polynomial time into a maximal matching M of G with $|M| \le |F|$. Let $D_i = \{v : \deg_G v = i\}$. In the remaining part of this section we will analyse the evolution of $|D_i|$ for $0 \le i \le d$, as the algorithm goes through step (2) and (3). Step (3) is performed in a number of iterations. For $j \ge 0$, let $V_i^j(t)$ be the size of D_i at stage t of iteration j, with the convention that iteration 0 refers to the execution of step (2).

Step (2) *for d-regular graphs.* Theorem 4 in [9] implies that step (2) proceeds for asymptotically $x_1 = \frac{1}{2} - \frac{1}{2}\left(\frac{1}{d-1}\right)^{\frac{2}{d-2}}$ stages, adding an edge to M at every stage. Let $V_i^{j+}(t) = |D_i \cap V(M)|$ at stage t of iteration j and set $V_i^{j-}(t) = V_i^j(t) - V_i^{j+}(t)$. Let $\Delta V_i^{0\ \mathrm{sign}}(t)$ denote the expected change of $V_i^{0\ \mathrm{sign}}(t)$ (with sign $\in$ {"", "+", "−"}) moving from stage t to $t+1$, of step (2), conditioned to the history of the algorithm's execution up to stage t. Let v be the chosen vertex of degree d. We assume a given fixed ordering among the vertices adjacent to v. The edge $\{v, u_1\}$ is added to M and edges $\{v, u_l\}$ (for $l = 2, \ldots, \deg_G v$) are removed from G. Vertex v becomes of degree zero and the expected reduction in the number of vertices of degree i that are (not) in $V(M)$ is $\frac{iV_i^{0+}(t)}{n-2t}$ (resp. $\frac{iV_i^{0-}(t)}{n-2t}$), that is the probability that a vertex in $D_i \cap V(M)$ (resp. $D_i \cap (V \setminus V(M))$) is hit over d trials. The "loss" of a vertex of degree i implies the "gain" of a vertex of degree $i-1$. Moreover if $u_1 \in D_{i+1} \cap (V \setminus V(M))$ at stage t, then $u_1 \in D_i \cap V(M)$ at stage $t+1$. Let $\delta_{r,s} = 1$ if $r = s$ and zero otherwise. In what follows $i \in \{1, \ldots, d-1\}$. Also $V_d^{0-}(t) = V_d^0(t)$. We have

$$\Delta V_d^0(t) = -1 - \frac{dV_d^0(t)}{n-2t}$$
$$\Delta V_i^{0+}(t) = -\frac{iV_i^{0+}(t)}{n-2t} + (1 - \delta_{d-1,i})\frac{(i+1)V_{i+1}^{0+}(t)}{n-2t} + \frac{(i+1)V_{i+1}^{0-}(t)}{nd-2dt}$$
$$\Delta V_i^{0-}(t) = -\frac{iV_i^{0-}(t)}{n-2t} + \frac{(i+1)(d-1)V_{i+1}^{0-}(t)}{nd-2dt}$$
$$\Delta V_0^0(t) = 1 + \frac{V_1^{0-}(t)}{n-2t} + \frac{V_1^{0+}(t)}{n-2t}$$

Setting $x = t/n$, $V_i^{1\text{ sign}}(t) = n v_i^{1\text{ sign}}(t/n)$, we can consider the following system of differential equations:

$$\begin{aligned}
{v_d^0}'(x) &= -1 - \frac{d v_d^0(x)}{1-2x} && v_d^0(0) = 1\\
{v_i^{0+}}'(x) &= -\frac{i v_i^{0+}(x)}{1-2x} + (1-\delta_{d-1,i})\frac{(i+1)v_{i+1}^{0+}(x)}{1-2x} + \frac{(i+1)v_{i+1}^{0-}(x)}{d(1-2x)} && v_i^{0+}(0) = 0\\
{v_i^{0-}}'(x) &= -\frac{i v_i^{0-}(x)}{1-2x} + \frac{(i+1)(d-1)v_{i+1}^{0-}(x)}{d(1-2x)} && v_i^{0-}(0) = 0\\
{v_0^0}'(x) &= 1 + \frac{v_1^{0-}(x)}{1-2x} + \frac{v_1^{0+}(x)}{1-2x} && v_0^0(0) = 0
\end{aligned}$$

In each case $|V_i^{1\text{ sign}}(t+1) - V_i^{1\text{ sign}}(t)|$ is bounded by a constant. Also, the system of differential equations above is sufficiently well-behaved, so that the hypotheses of Theorem 1 in [9] are fulfilled and thus for large n, $V_i^{1\text{ sign}}(t) \sim n v_i^{1\text{ sign}}(t/n)$ where $v_i^{1\text{ sign}}(x)$ are the solutions of the system above.

Lemma 1. *For each $d \in \mathbb{N}^+$ and for each $i \in \{1,\ldots,d\}$, there is a number A_i, two sequences of real numbers $\{B_{i,0}^j\}_{j=1,\ldots,\lceil\frac{d-i+1}{2}\rceil}$, $\{B_{i,1}^j\}_{j=1,\ldots,\lceil\frac{d-i}{2}\rceil}$, and a number C_i such that the system of differential equation above admits the following unique solutions:*

$$\begin{aligned}
v_i^{0-}(x) &= A_i(1-2x) + (1-2x)^{\frac{i}{2}}\left[C_i\log(1-2x) + \textstyle\sum_{j=0}^{\lceil\frac{d-i+1}{2}\rceil} B_{i,0}^j x^j\right]\\
&\quad + (1-2x)^{\frac{i+1}{2}}\textstyle\sum_{j=0}^{\lceil\frac{d-i}{2}\rceil} B_{i,1}^j x^j\\
v_i^{0+}(x) &= v_i^{0-}(x)\left[\left(\frac{d}{d-1}\right)^{d-i} - 1\right]\\
v_0^0(x) &= f_0(x) - f_0(0)
\end{aligned}$$

where $f_0(x) = x + \left(\frac{d}{d-1}\right)^{d-1}\int \frac{v_1^{0-}(x)}{1-2x}\mathrm{d}x$.

Proof. We sketch the proof of the first two results (which can be formally carried out by induction on $d-i$). For $i = d$, $v_d^0(x) = v_d^{0-}(x) = -\frac{1}{d-2}(1-2x) + \frac{d-1}{d-2}(1-2x)^{d/2}$. Assuming the result holds for $v_{i+1}^{0-}(x)$ and letting $D_i = (i+1)\left(\frac{d-1}{d}\right)$, we have

$$v_i^{0-}(x) = D_i(1-2x)^{\frac{i}{2}}\int_0^x \frac{v_{i+1}^{0-}(s)}{(1-2s)^{\frac{i}{2}+1}}\mathrm{d}s$$

and the result follows by integration (in particular, the logarithmic terms are present only if $i \le 2$).

Let $I_{i+1}^{0-}(x) = \int_0^x \frac{v_{i+1}^{0-}(s)}{(1-2s)^{\frac{i}{2}+1}}\mathrm{d}s$; then $v_i^{0-}(x) = D_i I_{i+1}^{0-}(x)$. Therefore

$$\begin{aligned}
v_i^{0+}(x) &= (i+1)\int_0^x \frac{v_{i+1}^{0+}(s)}{(1-2s)^{1+\frac{i}{2}}}\mathrm{d}s + \frac{i+1}{d}I_{i+1}^{0-}(x)\\
&= (i+1)\int_0^x \frac{v_{i+1}^{0+}(s)}{(1-2s)^{1+\frac{i}{2}}}\mathrm{d}s + \frac{v_i^{0-}(x)}{d-1}\\
&= (i+1)\left[\left(\frac{d}{d-1}\right)^{d-i-1} - 1\right]\int_0^x \frac{v_{i+1}^{0-}(s)}{(1-2s)^{1+\frac{i}{2}}}\mathrm{d}s + \frac{v_i^{0-}(x)}{d-1}\\
&= (i+1)\left[\left(\frac{d}{d-1}\right)^{d-i-1} - 1\right]\frac{d v_i^{0-}(x)}{(i+1)(d-1)} + \frac{v_i^{0-}(x)}{d-1}\\
&= \left[\left(\frac{d}{d-1}\right)^{d-i-1} - 1\right]\frac{d v_i^{0-}(x)}{d-1} + \frac{v_i^{0-}(x)}{d-1}
\end{aligned}$$

The third result follows by replacing the expression for v_i^{0+} in

$$v_0^{0\prime}(x) = 1 + \frac{v_1^{0-}(x)}{1-2x} + \frac{v_1^{0+}(x)}{1-2x}$$

□

Lemma 2. *Let x_1 be the smallest root of $v_d^0(x) = 0$. After Step (2) is completed the size of M is asymptotically $x_1 n$ for a.e. $G \in \mathcal{G}(n, d\text{-reg})$.* □

Step (3.j) for cubic graphs. During this step the algorithm chooses a random vertex in $D_{3-j} \cap V(M)$ and removes it from G (all edges incident to it will not be added to M). Let $c_j(3-j)n/2$ be the number of edges left at the beginning of iteration j. If iteration $j-1$ ended at stage $x_j n$, the parameter c_j satisfies the recurrence:

$$c_j = \frac{c_{j-1}(4-j)}{3-j} - \frac{2(4-j)x_j}{3-j} = \left(1 + \frac{1}{3-j}\right)(c_{j-1} - 2x_j)$$

(with $c_0 = 1$) where x_1 has been defined above and x_2 and x_3 will be defined later. For all $i \in \{1, 2, 3\}$ the expected decrease in the number of vertices of degree i in $V(M)$ (resp. not in $V(M)$) is $\frac{iV_i^{j+}(t)}{c_j n - 2t}$ ($\frac{iV_i^{j-}(t)}{c_j n - 2t}$). The following set of equations describes the expected change in the various $V_i^{j\ \mathrm{sign}}(t)$. In what follows $i \in \{1, \ldots, 3-j\}$. Notice that $V_i^{j+}(t) = 0$ for all $i > 3-j$ during iteration j so there are only $3-j$ equations involving $V_i^{j+}(t)$ but there are always two involving $V_i^{j-}(t)$.

$$\begin{aligned} \Delta V_0^j(t) &= 1 + \frac{V_1^{j-}(t)}{c_j n - 2t} + \frac{V_1^{j+}(t)}{c_j n - 2t} \\ \Delta V_i^{j+}(t) &= -\delta_{3-j,i} - \frac{iV_i^{j+}(t)}{c_j n - 2t} + \frac{(i+1)V_{i+1}^{j+}(t)}{c_j n - 2t} \\ \Delta V_i^{j-}(t) &= -\frac{iV_i^{j-}(t)}{c_j n - 2t} + (1 - \delta_{2,i})\frac{(i+1)V_{i+1}^{j-}(t)}{c_j n - 2t} \end{aligned}$$

Leading to the d.e.'s

$$\begin{aligned} v_0^{j\prime}(x) &= 1 + \frac{v_1^{j-}(x)}{c_j - 2x} + \frac{v_1^{j+}(x)}{c_j - 2x} & v_0^j(0) &= 0 \\ v_i^{j+\prime}(x) &= -\delta_{3-j,i} - \frac{iv_i^{j+}(x)}{c_j - 2x} + \frac{(i+1)v_{i+1}^{j+}(x)}{c_j - 2x} & v_i^{j+}(0) &= v_i^{(j-1)+}(x_j) \\ v_i^{j-\prime}(x) &= -\frac{iv_i^{j-}(x)}{c_j - 2x} + (1 - \delta_{2,i})\frac{(i+1)v_{i+1}^{j-}(x)}{c_j - 2x} & v_i^{j-}(0) &= v_i^{(j-1)-}(x_j) \end{aligned}$$

Theorem 12. *Let x_j be the smallest positive root of $v_{4-j}^{(j-1)+}(x) = 0$, for $j \in \{1, 2, 3\}$. For a.e. $G \in \mathcal{G}(n, 3\text{-reg})$ algorithm $\mathcal{A}$ returns a maximal matching of size at most*

$$\beta_u(G) \sim n\left(x_1 + \frac{v_1^{2-}(x_3) + v_2^{2-}(x_3)}{2}\right)$$

Proof. The result follows again by applying Theorem 1 in [9] to the random variables $V_i^{1\ \mathrm{sign}}(t)$. Notice that all functions $v_i^{1\ \mathrm{sign}}(x)$ have a simple expression which can be derived by direct integration and, in particular,

$$x_2 = \frac{c_1}{2}\left[1 - \exp\left(-\frac{2v_2^0(x_1)}{c_1}\right)\right] \qquad x_3 = \frac{c_2}{2}\left[1 - 4\left(1 - \frac{v_1^{1+}(x_2)}{c_2}\right)^2\right]$$

□

5 Conclusions

In this paper we presented a number of results about the minimal size of a maximal matching in several types of random graphs. If the graph G is dense, with high probability $\beta(G)$ is concentrated around $|V(G)|/2$ (both in the general and bipartite case). Moreover simple algorithms return an asymptotically optimal matching. We also gave simple combinatorial lower bounds on $\beta(G)$ if $G \in \mathcal{G}(n, c/n)$. Finally we presented combinatorial bounds on $\beta(G)$ if $G \in \mathcal{G}(n, d\text{-reg})$ and an algorithm that finds a maximal matching of size asymptotically less than $|V(G)|/2$ in G. The complete analysis was presented for the case when $G \in \mathcal{G}(n, 3\text{-reg})$. In such case the bound in Theorem 11 and the algorithmic result in Theorem 12 imply that $0.3158n < \beta(G) < 0.47563n$. Results similar to Theorem 12 can be proved for random d-regular graphs, although some extra care is needed to keep track of the evolving degree sequence. Our algorithmic results exploit a relationship between independent sets and maximal matchings. In all cases the given minimisation problem is reduced to a maximisation one, and the analysis is completed by exploiting a number of techniques available to deal with the maximisation problem. The weakness of our results for sparse graphs and for regular graphs leaves the open problem of finding a more direct approach which might produce better results.

References

1. B. Bollobás. *Random Graphs*. Academic Press, 1985.
2. J. Edmonds. Paths, Trees and Flowers. *Canadian Journal of Math.*, 15:449–467, 1965.
3. P. Erdős and A. Rényi. On the Existence of a Factor of Degree One of a Connected Random Graph. *Acta Mathematica Academiae Scientiarum Hungaricae*, 17(3–4):359–368, 1966.
4. G. R. Grimmett and C. J. H. McDiarmid. On Colouring Random Graphs. *Mathematical Proceedings of the Cambridge Philosophical Society*, 77:313–324, 1975.
5. J. Hopcroft and R. Karp. An $n^{5/2}$ Algorithm for Maximal Matching in Bipartite Graphs. *SIAM Journal on Computing*, 2:225–231, 1973.
6. B. Korte and D. Hausmann. An Analysis of the Greedy Heuristic for Independence Systems. *Annals of Discrete Mathematics*, 2:65–74, 1978.
7. B. D. McKay. Independent Sets in Regular Graphs of High Girth. *Ars Combinatoria*, 23A:179–185, 1987.
8. S. Micali and V. V. Vazirani. An $O(v^{1/2}e)$ Algorithm for Finding Maximum Matching in General Graphs. In *Proceedings of the 21st Annual Symposium on Foundations of Computer Science*, pages 17–27, New York, 1980.
9. N. C. Wormald. Differential Equations for Random Processes and Random Graphs. *Annals of Applied Probability*, 5:1217–1235, 1995.
10. M. Yannakakis and F. Gavril. Edge Dominating Sets in Graphs. *SIAM Journal on Applied Mathematics*, 38(3):364–372, June 1980.
11. M. Zito. *Randomised Techniques in Combinatorial Algorithmics*. PhD thesis, Department of Computer Science, University of Warwick, 1999.