A note on randomly colored matchings in random bipartite graphs

Alan Frieze
Department of Mathematical Sciences,
Carnegie Mellon University,
Pittsburgh PA15213,
USA.
alan@random.math.cmu.edu.*

October 3, 2020

Abstract

We are given a bipartite graph that contains at least one perfect matching and where each edge is colored from a set $Q = \{c_1, c_2, \ldots, c_q\}$. Let
$Q_i = \{e \in E(G) : c(e) = c_i\}$, where $c(e)$ denotes the color of e. The perfect matching color profile $mcp(G)$ is defined to be the set of vectors $(m_1, m_2, \ldots, m_q) \in [n]^q$ such that there exists a perfect matching M such that $|M \cap Q_i| = m_i$. We give bounds on the matching color profile for a randomly colored random bipartite graph.

1 Introduction

We consider the following problem: we are given a random bipartite graph G in which each edge is given a random color from a set $Q = \{c_1, c_2, \ldots, c_q\}$. An edge e is colored $c(e) = c_i$ with probability α_i where $\alpha_i > 0$ is a constant. Let $Q_i = \{e \in E(G) : c(e) = c_i\}$, where $c(e)$ denotes the color of e. The perfect matching color profile $mcp(G)$ is defined to be the set of vectors $(m_1, m_2, \ldots, m_q) \in [n]^q$ such that there exists a perfect matching M such that $|M \cap Q_i| = m_i$. We give bounds on the matching color profile for a randomly colored random bipartite graph.

Randomly colored random graphs have been studied recently in the context of (i) rainbow matchings and Hamilton cycles, see for example [2], [3], [7], [11]; (ii) rainbow connection see for example [5],

*Research supported in part by NSF grant DMS1661063

[9], [10], [13], [12]; (iii) pattern colored Hamilton cycles, see for example [1], [6]. This paper can be considered to be a contribution in the same genre. One can imagine a possible interest in the color profile via the following scenario: suppose that A is a set of tools and B is a set of jobs where edge $\{a, b\}$ indicates that b can be completed using a. If colors represent people, then one might be interested in equitably distributing jobs. I.e. determining whether $(n/q, n/q, \ldots, n/q) \in mcp(G)$. In any case, we find the problem interesting.

We will consider G to be the random bipartite graph $G_{n,n,p}$ where $p = \frac{\log n+\omega}{n}$, $\omega = \omega(n) \to \infty$ where $\omega = o(\log n)$. Erdős and Rényi [4] proved that G has a perfect matching w.h.p. We will prove the following theorem: let $\alpha_1, \alpha_2, \ldots, \alpha_q, \beta$ be positive constants such that $\alpha_1 + \alpha_2 + \cdots + \alpha_q = 1$ and $\beta < 1/q$. Let

$$\alpha_{\min} = \min\{\alpha_i : i \in [q]\}.$$

Theorem 1. *Let G be the random bipartite graph $G_{n,n,p}$ where $p = \frac{\log n+\omega}{n}$, $\omega = \omega(n) \to \infty$ where $\omega = o(\log n)$. Suppose that the edges of G are independently colored with colors from $C = \{c_1, c_2, \ldots, c_q\}$ where $\mathbb{P}(c(e) = c_i) = \alpha_i$ for $e \in E(G), i \in [q]$. Let $m_1, m_2, \ldots, m_q$ satisfy: (i) $m_1 + \cdots + m_q = n$ and (ii) $m_i \geq \beta n, i \in [q]$. Then w.h.p., there exists a perfect matching M in which exactly m_i edges are colored with $c_i, i = 1, 2, \ldots, q$.*

It is clear that w.h.p. $(n, 0, \ldots, 0) \notin mcp(G)$. This is because the bipartite graph induced by edges of color c_1 is distributed as $G_{n,n,\alpha_1 p}$ and this contains isolated vertices w.h.p. On the other hand, if $p \geq \frac{q(\log n+\omega)}{\alpha_{\min} n}$ then w.h.p. $mcp(G) = [n]^q$. To see this, suppose that $m_1 \leq m_2 \leq \cdots m_q \leq n$. Suppose we have found a matching that uses m_i edges of color c_i for $i \geq 0$. Let $n' = n - m_1 - \cdots - m_i$. Then the random bipartite graph induced by vertices not in M and having edges of color c_i has density at least $\frac{q\alpha_i n'}{\alpha_{\min} n} \cdot \frac{\log n+\omega}{n'} \geq \frac{\log n'+\omega/2}{n'}$ and so has a perfect matching w.h.p.

Open Question: What is the threshold for $mcp(G) = [0, n]^q$?

2 Structural Lemma

Suppose that the bipartition of $V(G)$ is denoted A, B. For sets $S \subseteq A, T \subseteq B$ we let $e_i(S, T)$ denote the number of $S : T$ edges of color c_i. We say that vertex u is c_i-adjacent to vertex v if the edge $\{u, v\}$ exists and has color c_i.

Lemma 2. *Let $p = \frac{\log n+\omega}{n}$, $\omega = \omega(n) \to \infty$ where $\omega = o(\log n)$. Then w.h.p.*

(a) $S \subseteq A, T \subseteq B$ and $\gamma_a \log n \leq |S| \leq n_0 = \gamma_a n/\log n$ and $|T| \leq \alpha_i \eta |S| \log n$ where $\gamma_a = \eta/(20\alpha_i)$ implies that $e_i(S : T) \leq 2\alpha_i \eta |S| \log n$ for $i = 1, 2, \ldots, q$.

(b) There do not exist sets $X \subseteq S \subseteq A, T \subseteq B$ and $i \in [q]$ such that $|S|, |T| \geq \beta n$ and $|X| = \gamma_b |S|/\log n, \gamma_b = 10\log(e/\beta)/\alpha_i$ and such that each $x \in X$ is c_i-adjacent to fewer than $\alpha_i \beta \log n/10$ vertices in T.

(c) There do not exist sets $X \subseteq S \subseteq A, T \subseteq B$ and $i \in [q]$ such that $|S|, |T| \geq \beta n$ and $|X| = |S|/\log n$ and a set $Z \subseteq T, |Z| = \gamma_b n/\log n$ such that each $x \in X$ is c_i-adjacent to $k = \frac{10 \log n}{\log\log n}$ vetices in Z.

(d) There do not exist sets $S \subseteq A, T \subseteq B$ and $i \in [q]$ such that $|S|, |T| \geq \beta n$ such that there are more than $\gamma_d n/\log n, \gamma_d = \frac{4}{\alpha_i} \log\left(\frac{e}{\beta}\right)$ vertices in T that not c_i-adjacent to a vertex in S.

(e) Fix $\gamma, \delta > 0$ constants. Then w.h.p. there do not exist sets S, T with $|S| = |T| = \gamma n/\log n$ such that $e_i(S,T) \geq \delta |S| \log n/\log\log n$.

(f) There do not exist sets $S \subseteq A, T \subseteq B$ and $i \in [q]$ such that $|S|, |T| \geq \beta n/10$ such that $e_i(S,T) = 0$.

Proof
(a) The probability that the condition is violated can be bounded by

$$\sum_{s=\gamma_a \log n}^{n_0} \sum_{t=1}^{\alpha_i \eta s \log n} \binom{n}{s}\binom{n}{t}\binom{st}{2\alpha_i \eta s \log n}(\alpha_i p)^{2\alpha_i \eta s \log n}$$

$$\leq \sum_{s=\gamma_a \log n}^{n_0} \sum_{t=1}^{\alpha_i \eta s \log n} \left(\frac{ne}{s}\right)^s \left(\frac{ne}{t}\right)^t \left(\frac{est\alpha_i p}{2\alpha_i \eta s \log n}\right)^{2\alpha_i \eta s \log n}$$

$$\leq \sum_{s=\gamma_a \log n}^{n_0} \sum_{t=1}^{\alpha_i \eta s \log n} \left(\frac{ne}{s}\right)^s \left(\frac{ne}{\alpha_i \eta s \log n}\right)^{\alpha_i \eta s \log n} \left(\frac{etp}{2\eta \log n}\right)^{2\alpha_i \eta s \log n}$$

$$\leq \sum_{s=\gamma_a \log n}^{n_0} \sum_{t=1}^{\alpha_i \eta s \log n} \left(\left(\frac{ne}{s}\right)^{1/2\alpha_i \eta \log n} \left(\frac{ne}{\alpha_i \eta s \log n}\right)^{1/2} \cdot \frac{e^{1+o(1)}\alpha_i s \log n}{2n}\right)^{2\alpha_i \eta s \log n}$$

$$\leq \sum_{s=\gamma_a \log n}^{n_0} \sum_{t=1}^{\alpha_i \eta s \log n} \left(\frac{s \log n}{n}\right)^{\alpha_i \eta s \log n - s} (\log n)^s \left(\frac{e^{3/2+o(1)}\alpha_i^{1/2}}{2\eta^{1/2}}\right)^{2\alpha_i \eta s \log n} = o(1).$$

(b) The probability that the condition is violated can be bounded by

$$\binom{n}{\beta n}^2 \binom{\beta n}{\gamma_b n/\log n} \left(e^{-\alpha_i \beta/4}\right)^{\gamma_b n} \leq \left(\left(\frac{e}{\beta}\right)^{(2\beta+o(1))} e^{-\alpha_i \beta \gamma_b/4}\right)^n = o(1).$$

The factor $e^{-\alpha_i \beta \gamma_b/4}$ comes from applying a Chernoff bound.

(c) We can assume w.l.o.g. that $|S| = |T| = \beta n$. The probability that the condition is violated can be bounded by

$$\binom{n}{\beta n}^2 \binom{\beta n}{n/\log n} \binom{\beta n}{\gamma_b n/\log n} \left(\binom{\gamma_b n/\log n}{k}(\alpha_i p)^k\right)^{n/\log n}$$
$$\leq \left(\frac{e}{\beta}\right)^{(2\beta+o(1))n} \left(\frac{e\gamma_b \alpha_i}{k}\right)^{kn/\log n} = o(1).$$

(d) The probability that the condition is violated can be bounded by

$$\binom{n}{\beta n}^2 \binom{\beta n}{\gamma_d n/\log n}(1-\alpha_i p)^{\beta n \gamma_d n/\log n} \leq \left(\left(\frac{e}{\beta}\right)^{2+o(1)} e^{-\alpha_i \gamma_d}\right)^{\beta n} = o(1).$$

(e) The probability that the condition is violated can be bounded by

$$\binom{n}{\gamma n/\log n}^2 \binom{\gamma^2 n^2/(\log n)^2}{\delta n/\log\log n} p^{\delta n/\log\log n} \leq$$

$$\left(\frac{e\log n}{\gamma}\right)^{2\gamma n/\log n} \left(\frac{\gamma^2 e \log\log n}{\delta \log n}\right)^{\delta n/\log\log n} = o(1).$$

(f) The probability that the condition is violated can be bounded by

$$2^{2n}(1-p)^{\beta^2 n^2/100} = o(1).$$

□

3 Proof of Theorem 1

Proof Assume from now on that the high probability conditions of Lemma 2 are in force. Let M be a perfect matching and let $\mu_i = |M \cap Q_i|$ for $i \in [q]$. Suppose that $\mu_1 > m_1 \geq \beta n$ and $\beta n \leq \mu_2 < m_2$. We show that we can find another matching M' such that $|M' \cap Q_1| = \mu_1 - 1$ and $|M' \cap Q_2| = \mu_2 + 1$. We do this by finding an alternating cycle with edge sequence $C = (e_1, f_1, \ldots, e_\ell, f_\ell)$ and vertex sequence $(x_1 \in A, y_1 \in B, x_2, \ldots, x_\ell, y_\ell, x_1)$ such that (i) $e_i = \{x_i, y_i\} \in M$, (ii) $f_i = \{y_i, x_{i+1}\} \notin M, i \in [\ell]$, (iii) $e_1 \in Q_1$ and (iv) $E(C) \setminus \{e_1\} \subseteq Q_2$. Repeating this for pairs of colors, one over-subscribed and one under-subscribed we eventually achieve our goal. It is sufficient to consider this case, seeing as we can always w.h.p. find a matching that has been randomly colored with $\approx \alpha_i n$ edges of color c_i, $i = 1, 2, \ldots, q$.

Next let $A_i = V(M \cap Q_i) \cap A$ and $B_i = V(M \cap Q_i) \cap B$ for $i \in [q]$ and for $S \subseteq A$ let $N_i(S) = \{b \in B : \exists a \in S \ s.t.\ . \{a, b\} \in Q_i\}$ and $N_i(a) = N_i(\{a\})$. Then let

$$D_0' = \left\{a \in A_2 : |N_2(a) \cap B_2| \geq \frac{\alpha_2 \beta \log n}{10}\right\}.$$

$$D_0 = \left\{a \in A_1 : |N_2(a) \cap M(A_2 \setminus D_0')| \leq k_0 = \frac{10\log n}{\log\log n}\right\}.$$

It follows from Lemma 2(b) that

$$|M(A_2 \setminus D_0')| \leq \frac{\gamma_b n}{\log n}.$$

It then follows from Lemma 2(c) that if $W_0 = A_1 \setminus D_0$ then

$$|W_0| \leq \frac{n}{\log n}. \qquad (1)$$

We now define a sequence of sets $W_0, W_1, \ldots$ where W_{j+1} is obtained from W_j by adding a vertex of $A_2 \setminus W_j$ for which $|N_2(a) \cap M(W_j)| \geq k_0$. Now consider $S = W_t, T = M(W_t)$ for some $t \geq 1$. Then we have

$$|S| = |T| \leq t + \frac{n}{\log n} \text{ and } e_2(S,T) \geq tk_0.$$

Given Lemma 2(e) with $\delta = 5, \gamma = 2$, we see that this sequence stops with $t = t^* \leq 4n/\log n$. So we now let $R_0 = A_2 \setminus W_{t^*}$. We note that

$$\begin{aligned} |R_0| &\geq \beta n - \frac{5n}{\log n} \\ a \in R_0 \text{ implies } |N_2(a) \cap M(R_0)| &\geq \frac{\alpha_2 \beta \log n}{10} - k_0. \end{aligned} \tag{2}$$

We now fix some $a_0 \in R_0$ and define a sequence of sets $X_0, Y_0, X_1, Y_1, \ldots$ where $X_j \subseteq R_0$ and $Y_j \subseteq B_2$. We let $X_0 = \{a_0\}$ and then having defined $X_i, i \geq 0$ we let

$$Y_i = N_2(X_i) \text{ and } X_{i+1} = \left(M^{-1}(Y_i) \setminus \bigcup_{j \leq i} X_j \right) \cap R_0.$$

We claim that for $i \geq 0$,

$$|X_i| \leq \frac{n}{200 \log n} \text{ implies that } |X_{i+1}| \geq \frac{\alpha_2 \beta \log n}{25} |X_i|. \tag{3}$$

We verify (3) below. Assuming its truth, there exists a smallest k such that

$$|X_k| \geq \frac{\alpha_2 \beta n}{5000}. \tag{4}$$

Starting with $\widehat{Y}_0 = \{b_0\}$ where $b_0 = M(a_0) \in \hat{R}_0$, we can similarly construct a sequence of sets $\widehat{Y}_1, \widehat{X}_1, \ldots$ where $\widehat{X}_j \subseteq M^{-1}(\hat{R}_0)$ and $\widehat{Y}_j \subseteq \hat{R}_0$. Here $\hat{R}_0$ is the equivalently defined set to R_0 in B_2. We can assume that $b_0 \in \hat{R}_0$, because of the sizes of the sets $R_0, \hat{R}_0$. More precisely, by (1), there will be $o(n)$ choices for a_0 for which $b_0 \notin \hat{R}_0$. Having defined $\widehat{Y}_i$ we let

$$\widehat{X}_i = N_2(\widehat{Y}_i) \text{ and } \widehat{Y}_{i+1} = \left(M(\widehat{X}_i) \setminus \bigcup_{j \leq i} \widehat{Y}_j \right) \cap \hat{R}_0.$$

and then let $\widehat{Y}_{i+1} = M(\widehat{X}_i)$. The equivalent of (3) will be

$$|\widehat{Y}_i| \leq \frac{n}{200 \log n} \text{ implies that } |\widehat{Y}_{i+1}| \geq \frac{\alpha_2 \beta \log n}{25} |\widehat{Y}_i|. \tag{5}$$

Assuming its truth, there exists ℓ such that

$$|\widehat{Y}_\ell| \geq \frac{\alpha_2 \beta n}{5000}. \tag{6}$$

It follows from Lemma 2(f) that at least 9/10 of the vertices of A_1 have a c_2-neighbor in $\hat{R}_0$ and at least 9/10 of the vertices of B_1 have a c_2-neighbor in R_0. We deduce from this that there is a pair $x_0 \in A_1, y_0 = M(x_0) \in B_1$ such that $N_2(x_0) \cap \hat{R}_0 \neq \emptyset$ and $N_2(y_0) \cap R_0 \neq \emptyset$. This defines an alternating cycle $x_0, u_0, P_1, b_0, a_0, P_2, v_0, y_0, x_0$. Here u_0 is a c_2-neighbor of x_0 in $\hat{R}_0$ and P_1 is (the reversal of) a path from u_0 to b_0 and P_2 is the path from a_0 to $v_0 \in X_k$, $v_0 \in N_2(y_0)$. This completes the proof of Theorem 1.

Verification of (3)**,** (5): We have by the assumption $a_0 \in R_0$ that

$$|X_1| = |Y_1| \geq \frac{\alpha_2 \beta \log n}{10} - o(\log n)$$

Now suppose that $1 \leq |X_i| \leq n/(200 \log n)$. Then, by (2),

$$e_2(X_i : (N_2(X_i) \setminus M(A_2 \setminus R_0))) \geq \frac{(\alpha_2 \beta \log n)|X_i|}{10 + o(1)}.$$

Applying Lemma 2(a) we see that

$$|N_2(X_i) \setminus M(A_2 \setminus R_0)| \geq \frac{(\alpha_2 \beta \log n)|X_i|}{20 + o(1)}. \tag{7}$$

Because the sets $X_1, X_2, \ldots$ expand rapidly, the total size of $\bigcup_{j \leq i} X_j$ is small compared with the R.H.S of (7) and (3) follows. The argument for (5) is similar. □

4 Concluding Remarks

We have established that w.h.p. *mcp*(G) is almost all of $[0, n]^q$ and posed the question of findng the exact threshold for *mcp*$(G) = [0, n]^q$. It seems technically feasible to extend our results to randomly colored $G_{n,p}$. We leave this for future research. It would be of some interest to analyse other spanning subgraphs from this point of view e.g. Hamilton cycles.

References

[1] M. Anastos and A.M. Frieze, Pattern Colored Hamilton Cycles in Random Graphs, *SIAM Journal on Discrete Mathematics* 33 (2019) 528-545.

[2] D. Bal and A.M. Frieze, Rainbow Matchings and Hamilton Cycles in Random Graphs, *Random Structures and Algorithms* 48 (2016) 503-523.

[3] C. Cooper and A.M. Frieze, Multi-coloured Hamilton cycles in random edge-coloured graphs, *Combinatorics, Probability and Computing* **11** (2002), 129-134.

[4] P. Erdős and A. Rényi, On random matrices, *Publ. Math. Inst. Hungar. Acad. Sci.* 8 (1964) 455-461.

[5] A. Dudek, A.M. Frieze and C. Tsourakakis, Rainbow connection of random regular graphs, *SIAM Journal on Discrete Mathematics* 29 (2015) 2255-2266.

[6] L. Espig, A.M. Frieze and M. Krivelevich, Elegantly colored paths and cycles in edge colored random graphs, *SIAM Journal on Discrete Mathematics* 32 (2018) 1585-1618.

[7] A. Ferber and M. Krivelevich, Rainbow Hamilton cycles in random graphs and hypergraphs. Recent trends in combinatorics, IMA Volumes in Mathematics and its applications, A. Beveridge, J. R. Griggs, L. Hogben, G. Musiker and P. Tetali, Eds., Springer 2016, 167-189.

[8] A.M. Frieze and P. Loh, Rainbow Hamilton cycles in random graphs, *Random Structures and Algorithms* 44 (2014) 328-354.

[9] A.M. Frieze and C. E. Tsourakakis, Rainbow connectivity of sparse random graphs, *Electronic Journal of Combinatorics* 19 (2012).

[10] A. Heckel and O. Riordan, The hitting time of rainbow connection number two, *Electronic Journal on Combinatorics* 19 (2012).

[11] S. Janson and N. Wormald, Rainbow Hamilton cycles in random regular graphs, *Random Structures Algorithms* 30 (2007) 35-49.

[12] N. Kamcev, M. Krivelevich and B. Sudakov, Some remarks on rainbow connectivity, *Journal of Graph Theory* 83 (2016), 372-383.

[13] M. Molloy, The rainbow connection number for random 3-regular graphs, *Electronic Journal of Combinatorics* 24 (2017).