

Clique coloring of dense random graphs

Noga Alon *

Michael Krivelevich †

Abstract

The clique chromatic number of a graph $G = (V, E)$ is the minimum number of colors in a vertex coloring so that no maximal (with respect to containment) clique is monochromatic. We prove that the clique chromatic number of the binomial random graph $G = G(n, 1/2)$ is, with high probability, $\Omega(\log n)$. This settles a problem of McDiarmid, Mitsche and Prałat who proved that it is $O(\log n)$ with high probability.

1 The main result

A *clique* in an undirected graph $G = (V, E)$ is *maximal* if it is not properly contained in a larger clique. A *clique coloring* of G is a vertex coloring so that no maximal clique (with at least two vertices) is monochromatic. Let $\chi_c(G)$ denote the minimum possible number of colors in a clique coloring of G . This invariant is called the *clique chromatic number* of G and has been studied in a considerable number of papers, see [3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15]. McDiarmid, Mitsche and Prałat [13] initiated the study of $\chi_c(G)$ for the random binomial graph $G = G(n, p)$. While for sparse random graphs their upper and lower bounds for the typical behavior of $\chi_c(G(n, p))$ are rather close to each other, for the dense case they do not have any nontrivial upper bound. In particular, for the random graph $G = G(n, 1/2)$ they proved that with high probability (whp, for short), that is, with probability tending to 1 as n tends to infinity, $\chi_c(G) \leq (1/2 + o(1)) \log_2 n$ and raised the problem of proving any nontrivial lower bound. In this note we show that the logarithmic estimate is tight, up to a constant factor.

Theorem 1.1. *There exists an absolute positive constant c so that whp the random graph $G = G(n, 1/2)$ satisfies $\chi_c(G) \geq c \log_2 n$. Therefore, whp $\chi_c(G) = \Theta(\log n)$.*

*Sackler School of Mathematics and Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel. Email: nogaa@tau.ac.il. Research supported in part by an ISF grant and by a GIF grant.

†Sackler School of Mathematics, Tel Aviv University, Tel Aviv 69978, Israel. Email: krivelev@post.tau.ac.il. Research supported in part by USA-Israeli BSF grant 2014361.

The proof appears in the next two sections. Throughout the proof we assume, whenever this is needed, that n is sufficiently large. All logarithms are in base 2, unless otherwise specified. To simplify the presentation, we omit all floor and ceiling signs whenever these are not crucial, and make no attempt to optimize the absolute constants in our estimates.

2 Preliminaries

This section includes the main technical part of the proof. It introduces several typical properties of the random graph $G(n, 1/2)$, established in the following three lemmas. The proof of the first two is straightforward, while that of the third one requires some work.

Lemma 2.1. *Let $G = G(n, 1/2) = (V, E)$. Then whp for every set S of at most $\frac{1}{2000} \log n$ vertices of G there are more than $n^{0.999} \log n$ vertices in $V - S$ that are not connected to any vertex of S .*

Lemma 2.2. *The following holds for the random graph $G = G(n, 1/2)$ whp. For every set Y of $|Y| = y \geq n^{0.999}$ vertices of G , the number of vertices in $V - Y$ that have less than $0.41y$ non-neighbors in Y is smaller than $\frac{1}{4} \log n$.*

The third lemma is a bit more technical. It is convenient to define first the following property of a set of vertices Y .

Definition 2.1. *Let $Y \subsetneq V$ be a set of vertices of an n -vertex graph $G = (V, E)$. Call Y significant if it satisfies the following two conditions:*

- *Every vertex $v \in V - Y$ has at least $n^{0.999}$ non-neighbors in Y ;*
- *The number of vertices $v \in V - Y$ that have at most $0.41y$ non-neighbors in Y is at most $\frac{1}{4} \log n$.*

Lemma 2.3. *Let $G = G(n, 1/2) = (V, E)$. Then whp every significant set Y in G contains a clique K of size $k = 1.9 \log n$ so that every vertex $v \in V - Y$ has at least one non-neighbor in K .*

We proceed with the proofs of the three lemmas above. The proofs of the first two are very easy.

Proof of Lemma 2.1: Fix a set S of $s \leq \frac{1}{2000} \log n$ vertices. The number of vertices in $V - S$ that are not connected to any member of S is a binomial random variable with parameters $n - s$ and $1/2^s$ whose expectation is $(n - s)2^{-s} \geq (1 - o(1))n^{0.9995}$. By the

standard estimates for binomial distributions (c.f., e.g., [2], Theorem A.1.13) the probability that this number is smaller than $n^{0.999} \log n$, which is less than half its expectation, is smaller than $e^{-n^{0.9995}/8}$. As the number of possible sets S is less than $n^{\log n}$ the desired result follows by the union bound. \square

Proof of Lemma 2.2: If G contains a set Y of size $|Y| = y \geq n^{0.999}$ violating the lemma's claim, then there is a subset $X \subset V \setminus Y$ of size $|X| = x = \frac{1}{4} \log n$ such that every vertex $x \in X$ sends more than $0.59y$ edges to Y . This implies that G has more than $0.59xy$ edges crossing between X and Y . For two given sets X, Y as above, the number of edges $e(X, Y)$ between X and Y in $G(n, 1/2)$ is distributed binomially with parameters xy and $1/2$. Using, again, the known estimates for binomial distributions (c.f. [2], Theorem A.1.1), we obtain that the probability that $e(X, Y) \geq 0.59xy$ is at most $e^{-2 \cdot 0.09^2 xy} < e^{-0.016xy}$. Summing over all possible choices of X and Y , it follows that the probability of the existence of a set Y violating the lemma's claim is at most

$$\begin{aligned} & \sum_{y \geq n^{0.999}} \binom{n}{y} \binom{n-y}{x} e^{-0.016xy} \leq \sum_{y \geq n^{0.999}} \left(\frac{en}{y}\right)^y n^x e^{-0.016xy} \\ & \leq \sum_{y \geq n^{0.999}} n^x \cdot \left(en^{0.001} \cdot e^{-\frac{0.016 \log n}{4}}\right)^y \leq \sum_{y \geq n^{0.999}} n^x \cdot n^{-0.015 \cdot n^{0.999}} = o(1), \end{aligned}$$

completing the proof of the lemma. \square

Proof of Lemma 2.3: Fix a set Y of $y \geq n^{0.999}$ vertices of $G = G(n, 1/2) = (V, E)$ and expose all edges of G between Y and $V - Y$. If Y is not significant then there is nothing to prove, we thus assume that Y is significant. Put $r = \frac{1}{4} \log n$ and $s = k - r = 1.65 \log n$. Let B be a set of r vertices in $V - Y$ containing all vertices in $V - Y$ that have less than $0.41y$ non-neighbors in Y . Put $m = n^{0.99}$ and choose, for each $v \in B$, a subset of Y of size m consisting of non-neighbors of v , where all these subsets are pairwise disjoint. (It is easy to choose these sets sequentially, as each $v \in B$ has at least $n^{0.999}$ non-neighbors in Y .) This defines r subsets which we denote by Z_1, Z_2, \dots, Z_r . Put $Y' = Y - \cup_{i=1}^r Z_i$, $|Y'| = y'$, and note that each vertex $v \in V - (Y \cup B)$ has at least $0.41y - rn^{0.99} > 0.405y'$ non-neighbors in Y' .

Claim: There are s pairwise disjoint subsets $Z_{r+1}, Z_{r+2}, \dots, Z_k$ of Y' , each of size exactly m , so that every vertex $v \in V - (Y \cup B)$ has at least $0.4m$ non-neighbors in each of the subsets Z_j , $r + 1 \leq j \leq k$.

Proof of claim: Choose the sets randomly and apply the standard estimates for hypergeometric distributions (c.f. [9], Theorem 2.10). \square

Let \mathcal{F} be the family of all subsets of size k of Y that contain exactly one element in each set Z_i and contain at least one non-neighbor of each vertex $v \in V - Y$. Note that by the definition of the first sets Z_1, \dots, Z_r , each set that contains an element from each Z_i has at least one non-neighbor of each vertex $v \in B$. On the other hand, for each fixed $v \in V - (Y \cup B)$, when we choose randomly one member from each Z_j for $r + 1 \leq j \leq k$, the probability that we do not choose any non-neighbor of v is at most $0.6^s = 0.6^{1.65 \log n} < \frac{1}{n^{1.1}}$. Therefore, by the union bound, almost all these choices do include at least one non-neighbor of each such v and hence

$$|\mathcal{F}| \geq (1 - o(1))m^k = (1 - o(1))n^{0.99 \cdot 1.9 \log n}.$$

We now expose the edges in the induced subgraph of G on $\cup_j Z_j$ and show that the probability that none of the members of \mathcal{F} is a clique is much smaller than 2^{-n} . This can be proved in several ways, either by using martingales (see [1], Section 4.1 for a similar argument), or by using Talagrand's Inequality, or by using the extended Janson's Inequality (c.f., [2], Theorem 8.1.2). The last alternative seems to be the shortest, and we proceed with its detailed description.

For each member K of \mathcal{F} , let x_K denote the indicator random variable whose value is 1 iff K is a clique in G and let $X = \sum_{K \in \mathcal{F}} X_K$. Our objective is to show that $X > 0$ with probability that is close enough to 1 to enable applying the union bound over all relevant sets Y . The expectation of each X_K is clearly

$$E(X_K) = 2^{-\binom{k}{2}}.$$

Thus, by linearity of expectation,

$$E(X) = |\mathcal{F}|2^{-\binom{k}{2}} = (1 - o(1))m^k 2^{-\binom{k}{2}} = (1 - o(1))(m2^{-(k-1)/2})^k > n^{0.03k} > n^{10}$$

with (a lot of) room to spare.

Put $\mu = E(X)$, and define $\Delta = \sum_{K, K'} \text{Prob}[X_K = X_{K'} = 1]$ where the summation is over all (ordered) pairs K, K' of members of \mathcal{F} that satisfy $2 \leq |K \cap K'| \leq k - 1$. By the extended Janson Inequality the probability that $X = 0$ is at most $e^{-\mu^2/2\Delta}$.

Note that $\Delta = \sum_{i=2}^{k-1} \Delta_i$ where Δ_i is the contribution of pairs K, K' with $K, K' \in \mathcal{F}$, $|K \cap K'| = i$. Thus

$$\Delta_i \leq |\mathcal{F}|2^{-\binom{k}{2}} \binom{k}{i} (m-1)^{k-i} 2^{-\binom{k}{2} + \binom{i}{2}} \leq m^k 2^{-2\binom{k}{2} + \binom{i}{2}} \binom{k}{i} m^{k-i}.$$

We next prove that

$$\Delta = \sum_{i=2}^{k-1} \Delta_i \leq (1 + o(1)) \frac{k^2}{m^2} \mu^2. \quad (1)$$

To do so, consider the following cases.

Case 1: $i = 2$. In this case

$$\frac{\Delta_2}{\mu^2} \leq (1 + o(1)) \frac{\binom{k}{2} \cdot 2}{m^2} \leq (1 + o(1)) \frac{k^2}{m^2}.$$

Case 2: $3 \leq i < 100$. Here

$$\frac{\Delta_i}{\mu^2} \leq (1 + o(1)) \frac{\binom{k}{i} 2^{\binom{i}{2}}}{m^i} < (1 + o(1)) \left(\frac{k 2^{i/2}}{m} \right)^i \leq \left(\frac{k 2^{50}}{m} \right)^3 = \frac{1}{m^{3-o(1)}}.$$

Case 3: $100 \leq i \leq k - 2$. In this case

$$\frac{\Delta_i}{\mu^2} \leq (1 + o(1)) \frac{\binom{k}{i} 2^{\binom{i}{2}}}{m^i} < (1 + o(1)) \left(\frac{k 2^{i/2}}{m} \right)^i \leq \left(\frac{1}{n^{0.04-o(1)}} \right)^{100} = \frac{1}{n^{4-o(1)}}.$$

Summing the contributions for all i , $2 \leq i \leq k - 1$ ($< \log n$), the inequality (1) follows.

By the extended Janson Inequality this implies that the probability that $X = 0$ is at most

$$e^{-\mu^2/2\Delta} \leq e^{-(1+o(1))m^2/2k^2} < e^{-n^{1.98-o(1)}}.$$

As the number of possible significant sets Y is smaller than 2^n , the assertion of the lemma follows, by the union bound. \square

3 Completing the proof

In this section we prove Theorem 1.1. By the results in the previous section it suffices to prove the following deterministic statement.

Proposition 3.1. *Let $G = (V, E)$ be a graph on n vertices satisfying the assertions of Lemma 2.1, Lemma 2.2 and Lemma 2.3. Then $\chi_c(G) > \frac{1}{2000} \log n$.*

Proof: Assume this is false, and let Y_1, Y_2, \dots, Y_s be a partition of the vertex set V into disjoint non-empty sets, each containing no maximal clique of G , where $s \leq \frac{1}{2000} \log n$. It is easy to verify that the conclusions of Lemmas 2.1, 2.2 guarantee in particular that G contains at least one edge, and thus $s > 1$. For each i , $1 \leq i \leq s$, let $v_i \in V - Y_i$ be a vertex with the minimum number of non-neighbors in Y_i among all vertices in $V - Y_i$, and let t_i denote the number of these non-neighbors. Therefore, the number of vertices of G which

are not connected to any vertex of $S = \{v_1, v_2, \dots, v_s\}$ is at most $\sum_{i=1}^s t_i$, and since G satisfies the assertion of Lemma 2.1 this number exceeds $n^{0.999} \log n$. By averaging there exists an index i so that $t_i \geq n^{0.999}$. Fix such an i and note that $Y = Y_i$ is a significant set. Indeed, by the definition of t_i each $v \in V - Y_i$ has at least $t_i \geq n^{0.999}$ non-neighbors in Y_i , and as G satisfies the conclusion of Lemma 2.2 the number of vertices $v \in V - Y_i$ that have at most $0.41|Y_i|$ non-neighbors in Y_i is at most $\frac{1}{4} \log n$.

Since G satisfies the conclusion of Lemma 2.3, the set $Y := Y_i$ contains a clique K of size $1.9 \log n$ which contains at least one non-neighbor of each vertex $v \in V - Y_i$. This clique is contained in a maximal clique of G , call it K' . However, K' is contained in Y_i , as K has a non-neighbor of each vertex $v \in V - Y_i$. Thus K' is a maximal clique of G which is contained in Y_i , contradiction. This completes the proof of the proposition and hence of Theorem 1.1. \square

4 Concluding remarks and open problems

We have shown that the clique chromatic number $\chi_c(G)$ of the random graph $G = G(n, 1/2)$ is, whp, $\Theta(\log n)$. The same proof applies to binomial random graphs with any constant edge probability bounded away from 0 and 1. Together with the upper bound proved in [13] we conclude that for any fixed p , $0 < p < 1$, the random graph $G = G(n, p)$ satisfies, whp, $\chi_c(G) = \Theta_p(\log n)$.

We have made no attempt to optimize the absolute constants in our estimates. The constant $\frac{1}{2000}$ can certainly be improved, but it seems that the method as it is does not suffice to determine the tight constant here. It seems plausible that

$$\chi_c(G(n, 1/2)) = (1/2 + o(1)) \log n$$

whp.

Acknowledgments

We thank Colin McDiarmid for helpful discussions, and Andrzej Dudek, Tomasz Łuczak and Paweł Prałat for their input. This project was initiated during a workshop in Monash University, Melbourne, in June 2016. We thank the organizers of this workshop for their hospitality.

References

- [1] N. Alon, V. Asodi, C. Cantor, S. Kasif and J. Rachlin, Multi-node graphs: a framework for multiplexed biological assays, *Journal of Computational Biology* 13, 2006, 1659–1672.
- [2] N. Alon and J. H. Spencer, *The Probabilistic Method*, Fourth Edition, Wiley, 2016.
- [3] T. Andreae, M. Schughart and Zs. Tuza, Clique-transversal sets of line graphs and complements of line graphs, *Discrete Math.* 88, 1991, 11–20.
- [4] G. Bacsó, S. Gravier, A. Gyárfás, M. Preissmann and A. Sebó, Coloring the maximal cliques of graphs, *SIAM Journal on Discrete Mathematics* 17, 2004, 361–376.
- [5] C.N. Campos, S. Dantas and C.P. de Mello, Colouring clique-hypergraphs of circulant graphs, *Electron. Notes Discret. Math.* 30, 2008, 189–194.
- [6] M.R. Cerioli and A.L. Korenchender, Clique-coloring circular-arc graphs, *Electron. Notes Discret. Math.* 35, 2009, 287–292.
- [7] D. Défossez, Clique-coloring some classes of odd-hole-free graphs, *J. Graph Theory* 53, 2006, 233–249.
- [8] S. Gravier, C. Hoáng and F. Maffray, Coloring the hypergraph of maximal cliques of a graph with no long path, *Discrete Math.*, 272, 2003, 285–290.
- [9] S. Janson, T. Łuczak and A. Ruciński, *Random graphs*, Wiley, 2000.
- [10] S. Klein and A. Morgana, On clique-colouring with few P4s, *J. Braz. Comput. Soc.* 18, 2012, 113–119.
- [11] J. Kratochvíl and Zs. Tuza, On the complexity of bicoloring clique hypergraphs of graphs, *J. Algorithms*, 45, 2002, 40–54.
- [12] Z. Liang, E. Shan and L. Kang, Clique-coloring claw-free graphs, *Graphs and Combinatorics*, 2015, 1–16.
- [13] C. McDiarmid, D. Mitsche and P. Prałat, Clique coloring of binomial random graphs, arXiv 1611.01782
- [14] B. Mohar and R. Skrekovski, The Grötzsch Theorem for the hypergraph of maximal cliques, *Electronic J. Comb.* 6, 1999, R26.

- [15] E. Shan, Z. Liang and L. Kang, Clique-transversal sets and clique-coloring in planar graphs, *European J. Combin.* 36, 2014, 367–376.