Constraining the clustering transition for colorings of sparse random graphs

Michael Anastos   Alan Frieze*   Wesley Pegden†

Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh PA 15213, U.S.A.

Submitted: May 23, 2017; Accepted: Mar 7, 2018; Published: Mar 29, 2018
Mathematics Subject Classifications: 05C80

Abstract

Let Ω_q denote the set of proper $[q]$-colorings of the random graph $G_{n,m}, m = dn/2$ and let H_q be the graph with vertex set Ω_q and an edge $\{\sigma, \tau\}$ where σ, τ are mappings $[n] \to [q]$ iff $h(\sigma, \tau) = 1$. Here $h(\sigma, \tau)$ is the Hamming distance $|\{v \in [n] : \sigma(v) \neq \tau(v)\}|$. We show that w.h.p. H_q contains a single giant component containing almost all colorings in Ω_q if d is sufficiently large and $q \geqslant \frac{cd}{\log d}$ for a constant $c > 3/2$.

Keywords: Random Graphs; Colorings; Clustering Transition

1 Introduction

In this short note, we will discuss a structural property of the set Ω_q of proper $[q]$-colorings of the random graph $G_{n,m}$, where $m = dn/2$ for some large constant d. That is, proper colorings using colors from $[q] = \{1, 2, \ldots, q\}$. For the sake of precision, let us define H_q to be the graph with vertex set Ω_q and an edge $\{\sigma, \tau\}$ iff $h(\sigma, \tau) = 1$ where $h(\sigma, \tau)$ is the Hamming distance $|\{v \in [n] : \sigma(v) \neq \tau(v)\}|$. In the Statistical Physics literature the definition of H_q may be that colorings σ, τ are connected by an edge in H_q whenever $h(\sigma, \tau) = o(n)$. Our theorem holds a fortiori if this is the case.

Heuristic evidence in the statistical physics literature (see for example [15]) suggests there is a *clustering transition* c_d such that for $q > c_d$, the graph H_q is dominated by a single connected component, while for $q < c_d$, an exponential number of components are required to cover any constant fraction of it; it may be that $c_d \approx \frac{d}{\log d}$. (Here $A(d) \approx B(d)$ is taken to mean that $A(d)/B(d) \to 1$ as $d \to \infty$. We do not assume $d \to \infty$, only that d

*Research supported in part by NSF grant DMS1362785
†Research supported in part by NSF grant DMS1363136

is a sufficiently large constant, independent of n.) Recall that $G_{n,m}$ for $m = dn/2$ becomes q-colorable around $q \approx \frac{d}{2\log d}$ or equivalently when $d \approx 2q \log q$, [3, 7]. In this note, we prove the following:

Theorem 1.1. *If $q \geqslant \frac{cd}{\log d}$ for constant $c > 3/2$, and d is sufficiently large, then w.h.p. H_q contains a giant component that contains almost all of Ω_q.*

In particular, this implies that the clustering transition c_d, if it exists, must satisfy $c_d \leqslant \frac{3}{2}\frac{d}{\log d}$.

Theorem 1.1 falls into the area of "Structural Properties of Solutions to Random Constraint Satisfaction Problems". This is a growing area with connections to Computer Science and Theoretical Physics. In particular, much of the research on the graph H_q has been focussed on the structure near the *colorability threshold*, e.g. Bapst, Coja-Oghlan, Hetterich, Rassman and Vilenchik [4], or the *clustering threshold*, e.g. Achlioptas, Coja-Oghlan and Ricci-Tersenghi [2], Molloy [13]. Other papers heuristically identify a sequence of phase transitions in the structure of H_q, e.g., Krząkala, Montanari, Ricci-Tersenghi, Semerijan and Zdeborová [12], Zdeborová and Krząkala [15]. The existence of these transitions has been shown rigorously for some other CSPs. One of the most spectacular examples is due to Ding, Sly and Sun [8] who rigorously showed the existence of a sharp satisfiability threshold for random k-SAT.

An obvious target for future work is improving the constant in Theorem 1.1 to 1. We should note that Molloy [13] has shown that w.h.p. there is no giant component if $q \leqslant \frac{(1-\varepsilon_d)d}{\log d}$, for some $\varepsilon_d > 0$. Looking in another direction, it is shown in [9] that w.h.p. $H_q, q \geqslant d + 2$ is connected. This implies that Glauber Dynamics on Ω_q is ergodic. It would be of interest to know if this is true for some $q \ll d$.

Before we begin our analysis, we briefly explain the constant 3/2. We start with an arbitrary $[q]$-coloring and then re-color it using only approximately $\approx d/\log d$ of the given colors. We then use a disjoint set of approximately $d/2\log d$ colors to re-color it with a target $\chi \approx \frac{d}{2\log d}$ coloring τ. We will assume that τ uses colors from $\{q_0 + 1, \ldots, q_0 + \chi\}$.

2 Greedily Re-coloring

Our main tool is a theorem from Bapst, Coja-Oghlan and Efthymiou [5] on planted colorings. We consider two ways of generating a random coloring of a random graph. We will let $Z_q = |\Omega_q|$. The first method is to generate a random graph and then a random coloring. In the second method, we generate a random (planted) coloring and then generate a random graph compatible with this coloring.

Random coloring of the random graph $\boldsymbol{G_{n,m}}$: Here we will assume that m is such that w.h.p. $Z_q > 0$.

(a) Generate $G_{n,m}$ subject to $Z_q > 0$.

(b) Choose a $[q]$-coloring σ uniformly at random from Ω_q.

(c) Output $\Pi_1 = (G_{n,m}, \sigma)$.

Planted model:

1. Choose a random partition of $[n]$ into q color classes $V_1, V_2, \ldots, V_q$ subject to
$$\sum_{i=1}^{q} \binom{|V_i|}{2} \leqslant \binom{n}{2} - m.$$

2. Let $\Gamma_{\sigma,m}$ be obtained by adding m random edges, each with endpoints in different color classes.

3. Output $\Pi_2 = (\Gamma_{\sigma,m}, \sigma)$.

We will use the following result from [5]:

Theorem 2.1. *Let $d = 2m/n$ and suppose that $d \leqslant 2(q-1)\log(q-1)$. Then $\mathbf{Pr}(\Pi_2 \in \mathcal{P}) = o(1)$ implies that $\mathbf{Pr}(\Pi_1 \in \mathcal{P}) = o(1)$ for any graph+coloring property $\mathcal{P}$.*

Consequently, we will use the planted model in our subsequent analysis. Let
$$q_0 = \frac{q}{q-1} \cdot \frac{d}{\log d - 7\log\log d} \approx \frac{d}{\log d}. \tag{1}$$

The property $\mathcal{P}$ in question will be: "the given $[q]$-coloring can be reduced via single vertex color changes to a $[q_0]$-coloring".

In a random partition of $[n]$ into q parts, the size of each part is distributed as $Bin(n, q^{-1})$ and so the Chernoff bounds imply that w.h.p. in a random partition each part has size $\frac{n}{q}\left(1 \pm \frac{\log n}{n^{1/2}}\right)$.

We let Γ be obtained by taking a random partition $V_1, V_2, \ldots, V_q$ and then adding $m = \frac{1}{2}dn$ random edges so that each part is an independent set. These edges will be chosen from
$$N_q = \binom{n}{2} - (1+o(1))q\binom{n/q}{2} = (1-o(1))\frac{n^2}{2}\left(1 - \frac{1}{q}\right)$$
possibilities. So, let $\widehat{d} = \frac{mn}{N_q} \approx \frac{dq}{q-1}$ and replace Γ by $\widehat{\Gamma}$ where each edge not contained in a V_i is included independently with probability $\widehat{p} = \frac{\widehat{d}}{n}$. $V_1, V_2, \ldots, V_q$ constitutes a coloring which we will denote by σ. Now $\widehat{\Gamma}$ has m edges with probability $\Omega(n^{-1/2})$ and one can check that the properties required in Lemmas 2.2 and 2.3 below all occur with probability $1 - o(n^{-1/2})$ and so we can equally well work with $\widehat{\Gamma}$.

Now consider the following algorithm for going from σ via a path in Ω_q to a coloring with significantly fewer colors. It is basically the standard greedy coloring algorithm, as seen in Bollobás and Erdős [6], Grimmett and McDiarmid [10] and in particular Shamir and Upfal [14] for sparse graphs.

In words, it goes as follows. At each round of the algorithm, U denotes the set of vertices that have never been re-colored, by the start of the round. Having used $r-1$

colors to re-color some subset of vertices we start using color r. We let $W_j = V_j \cap U$ denote the unchanged vertices of V_j for $j \geqslant 1$. We then let k be the smallest index j for which $W_j \neq \emptyset$. During the re-coloring process, we will keep track of sets C_r and $U_r \subseteq U$, which are, respectively, the sets of vertices already re-colored r and the vertices of U not adjacent to any vertices in C_r. These sets are initially defined in the re-coloring process by the re-coloring all of W_k with color r so that $C_r = W_k$, and we proceed thereafter by choosing vertices from U_r and re-coloring them with color r (each time increasing $|C_r|$ by 1 and decreasing $|U_r|$ by at least 1). We finish when $U_r = \emptyset$, and in this way the terminal set C_r of vertices re-colored r will be a maximal independent subset of the set U. Note that in this construction, some or all of the vertices in V_r may be re-colored with $s < r$.

At any stage of the algorithm, U is the set of vertices whose colors have not been altered. The value of L in line D is $\left\lceil n/\log^2 \widehat{d} \right\rceil$.

ALGORITHM GREEDY RE-COLOR

begin
- Initialise: $r = 0, U = [n], C_0 \leftarrow \emptyset$;
- **repeat**;
- $r \leftarrow r + 1$, $C_r \leftarrow \emptyset$;
  - Let $W_j = V_j \cap U$ for $j \geqslant 1$ and let $k = \min\{j : W_j \neq \emptyset\}$;

A: $C_r \leftarrow W_k, U \leftarrow U \setminus C_r, U_r \leftarrow U \setminus \left\{\text{neighbors of } C_r \text{ in } \widehat{\Gamma}\right\}$;
  - If $r < k$, re-color every vertex in W_k with color r;

B: **repeat** (Re-color some more vertices with color r);

C: Arbitrarily choose $v \in U_r$, $C_r \leftarrow C_r + v$, $U_r \leftarrow U_r - v$;
  - $U_r \leftarrow U_r \setminus \left\{\text{neighbors of } v \text{ in } \widehat{\Gamma}\right\}$;
  - **until** $U_r = \emptyset$;
  - $U \leftarrow U \setminus U_r$;

D: **until** $|U| \leqslant L$;
- Suppose that at this point we have used r_0 colors;
- If possible, re-color U with colors $r_0 + 1, \ldots, r_0 + s_0$, where $s_0 = \left\lceil \frac{\widehat{d}}{\log^2 \widehat{d}} + 2 \right\rceil$;

end

2.1 Following a path in H_q

We first observe that each re-coloring of a single vertex v vertex in line C can be interpreted as moving from a coloring of Ω_q to a neighboring coloring in H_q. This requires us to argue that the re-coloring by GREEDY RE-COLOR is such that the coloring of $\widehat{\Gamma}$ is proper at all times. We argue by induction on r that the coloring at line A is proper. When $r = 1$ there have been no re-colorings. At the start of round r either all vertices previously colored r have been re-colored and C_r is a subset of vertices originally colored $k > r$ or is a subset of the vertices originally colored r. In the first case we can simply re-color W_k one vertex at a time with color r. Also, during the loop beginning at line B we only re-color vertices with color r if they are not neighbors of the set C_r of vertices colored r so far. This guarantees that the coloring remains proper until we reach line D. The following lemma shows that we can then reason as in Lemma 2 of Dyer, Flaxman, Frieze and Vigoda [9], as will be explained subsequently.

Lemma 2.2. *Let* $p = m/\binom{n}{2} = \Delta/n$ *where* Δ *is some sufficiently large constant. With probability* $1 - o(n^{-1/2})$*, every* $S \subseteq [n]$ *with* $s = |S| \leqslant n/\log^2 \Delta$ *contains at most* $s\Delta/\log^2 \Delta$ *edges.*

The above lemma, is Lemma 7.7(i) of Janson, Łuczak and Ruciński [11] and it implies that if $\Delta = \widehat{d}$ then w.h.p. $\widehat{\Gamma}_U$ at line D contains no K-core, $K = \frac{2\widehat{d}}{\log^2 \widehat{d}} + 1$. Here $\widehat{\Gamma}_U$ denotes the sub-graph of $\widehat{\Gamma}$ induced by the vertices U. For a graph $G = (V, E)$ and $K \geqslant 0$, the K-core is the unique maximal set $S \subseteq V$ such that the induced subgraph on S has minimum degree at least K. A graph without a K-core is K-*degenerate* i.e. its vertices can be ordered as $v_1, v_2, \ldots, v_n$ so that v_i has at most $K - 1$ neighbors in $\{v_1, v_2, \ldots, v_{i-1}\}$. To see this, let v_n be a vertex of minimum degree and then apply induction.

Suppose now that we have reached Line D and we find $|U| \leqslant L$. We claim that we can re-color the vertices in U with $K + 1$ new colors, all the time following some path in H_q. Let $v_1, \ldots, v_{|U|}$ denote an ordering of U such that the degree of v_i is less than K in the subgraph $\widehat{\Gamma}_i$ of $\widehat{\Gamma}$ induced by $\{v_1, v_2, \ldots, v_i\}$. We will prove the claim by induction on i, the inductive assertion being that we can re-color $v_1, v_2, \ldots, v_i$ ignoring conflicts caused by vertices $v_{i+1}, \ldots, v_{|U|}$. The asertion with $i = |U|$ shows the existence of the path we want. The claim is trivial for $i = 1$. Let σ_0 be the coloring of U at line D, when first we have $|U| \leqslant L$. By induction there is a path $\sigma_0, \sigma_1, \ldots, \sigma_r$ from the coloring σ_0 restricted to $\widehat{\Gamma}_{i-1}$, using only colors $r_0 + 1, \ldots, r_0 + s_0$ to do the re-coloring. Vertices outside of U will not be re-colored by this sequence.

Let (w_j, c_j) denote the $(vertex, color)$ change defining the edge $\{\sigma_{j-1}, \sigma_j\}$. We construct a path (of length $\leqslant 2r$) that re-colors $\widehat{\Gamma}_i$. For $j = 1, 2, \ldots, r$, we will re-color w_j to color c_j, if no neighbor of w_j has color c_j. Failing this, v_i must be the only neighbor of w_j that is colored c_j. This is because σ_r is a proper coloring of $\widehat{\Gamma}_{i-1}$. Since v_i has degree less than K in $\widehat{\Gamma}_i$, there exists a new color for v_i which does not appear in its neighborhood. Thus, we first re-color v_i to any new (valid) color, and then we re-color w_j to c_j, completing the inductive step. Note that because the colors used in Step D have not been used in Steps A,B,C, this re-coloring does not conflict with any of the coloring done in Steps A,B,C.

2.2 Bounding the number of colors used

We need to show that GREEDY RE-COLOR uses at most q_0 colors. To do this we show that w.h.p. each execution of Loop B re-colors a large number of vertices. Let $\alpha_1(G)$ denote the minimum size of a *maximal* independent set of a graph G. The round will re-color at least $\alpha_1(\Gamma_U)$ vertices, where U is as at the start of Loop B. The following result is from Lemma 7.8(i) of [11].

Lemma 2.3. *Let $p = m/\binom{n}{2} = \Delta/n$ where Δ is some sufficiently large constant. Then $\alpha_1(G_{n,m}) \geqslant \frac{\log\Delta - 3\log\log\Delta}{p}$ with probability $1 - o(n^{-1/2})$ (see Lemma 7.8(i)).*

Now the application of Step A and Loop B re-colors a maximal independent set C_r of the graph $\widehat{\Gamma}_U$ induced by U, as U stands at the beginning of Step A. This implies that the size of C_r stochastically dominates the size of a maximal independent set in $G_{|U|,\widehat{p}}$. This is because we can obtain $G_{|U|,\widehat{p}}$ by adding edges to $V_i \cap U, i \geqslant 1$ with probability $\widehat{p}$. Here we follow the usual analysis of greedy algorithms and argue that edges inside U are not conditioned by the process. This is often referred to as the method of deferred decisions. In this way we couple $\widehat{\Gamma}_U$ with $G_{|U|,\widehat{p}}$ so that every independent set in $G_{|U|,\widehat{p}}$ is contained in an independent set in $\widehat{\Gamma}_U$.

And so using Lemma 2.3 we see that w.h.p. each execution of Loop B re-colors at least

$$\frac{\log(\widehat{d}/\log^2\widehat{d}) - 3\log\log(\widehat{d}/\log^2\widehat{d})}{\widehat{d}} n \geqslant \frac{q-1}{q} \cdot \frac{\log d - 6\log\log d}{d} n$$

vertices, for d sufficiently large. We have replaced Δ of Lemma 2.3 by $\widehat{d}/\log^2\widehat{d}$ to allow for the fact that we have replaced n by $|U| \geqslant L$. Here we refer to the size of $|U|$ immediately after the update of r. Consequently, at the end of Algorithm GREEDY RE-COLOR we will have used at most

$$\frac{q}{q-1} \cdot \frac{d}{\log d - 6\log\log d} + \frac{\widehat{d}}{\log^2\widehat{d}} + 2 \leqslant \frac{q}{q-1} \cdot \frac{d}{\log d - 7\log\log d} = q_0$$

colors. The term $\frac{\widehat{d}}{\log^2\widehat{d}} + 2$ arises from the re-coloring of U at line D.

2.3 Finishing the proof:

Now suppose that $q \geqslant \frac{cd}{\log d}$ where d is large and $c > 3/2$. Fix a particular χ-coloring τ of $G_{n,m}$ that uses colors from $\{q_0+1, \ldots, q_0+\chi\}$. We prove that almost every $[q]$-coloring σ of $G_{n,m}$ can be transformed into τ changing one color at a time. It follows that for almost every pair of $[q]$-colorings σ, σ' we can transform σ into σ' by first transforming σ to τ and then reversing the path from σ' to τ.

We proceed as follows. Applying Theorem 2.1 with the property $\mathcal{P}$ as described following (1), we see that w.h.p., a uniformly random $[q]$-coloring σ of $G_{n,m}$ can be transformed one vertex at a time into a $[q_0]$-coloring θ. Then we process the vertices of the color classes of τ, re-coloring vertices to their τ-color. When we process a color class C of τ, we switch the color of vertices in C to their τ-color i_C one vertex at a time. We can do this because when we re-color a vertex v, a neighbor w will currently either have one of the q_0 colors used by θ and these are distinct from i_C or alternatively, w will have already been been re-colored with its τ-color and this will be distinct from i_C. This proves Theorem 1.1. □

Acknowledgement

We thank the referee for an exemplary sequence of reviews.

References

[1] D. Achlioptas and E. Friedgut, A Sharp Threshold for k-Colorability, *Random Structures and Algorithms*, 14 (1999) 63–70.

[2] D. Achlioptas, A. Coja-Oghlan and F. Ricci-Tersenghi, On the solution-space geometry of random constraint satisfaction problems, *Random Structures and Algorithms* 38 (2010) 251–268.

[3] D. Achlioptas and A. Naor, The Two Possible Values of the Chromatic Number of a Random Graph, *Annals of Mathematics* 162 (2005) 1333–1349.

[4] V. Bapst, A. Coja-Oghlan, S. Hetterich, F. Rassmann and D. Vilenchik, The condensation phase transition in random graph coloring, *Communications in Mathematical Physics* 341 (2016) 543–606.

[5] V. Bapst, A. Coja-Oghlan and C. Efthymiou, Planting colourings silently, *Combinatorics, Probability and Computing* 26 (2017) 338–366.

[6] B. Bollobás and P. Erdős, Cliques in random graphs, *Mathematical Proceedings of the Cambridge Philosophical Society* 80 (1976) 419–427.

[7] A. Coja-Oghlan and D. Vilenchik, Chasing the k-colorability threshold, *Proceedings of FOCS 2013*, 380–389.

[8] J. Ding, A. Sly and N. Sun, Proof of the satisfiability conjecture for large k, *Proceedings of STOC 2015* 59–68.

[9] M. Dyer, A. Flaxman, A.M. Frieze and E. Vigoda, Randomly coloring sparse random graphs with fewer colors than the maximum degree, *Random Structures and Algorithms* 29 (2006) 450–465.

[10] G. Grimmett and C. McDiarmid, On colouring random graphs, *Mathematical Proceedings of the Cambridge Philosophical Society* 77 (1975) 313–324.

[11] S. Janson, T. Łuczak and A. Ruciński, Random Graphs, Wiley 2000.

[12] F. Krzakąla, A. Montanari, F. Ricci-Tersenghi, G. Semerijian and L. Zdeborová, Gibbs states and the set of solutins of random constraint satisfaction problems, *Proceedings of the National Academy of Sciences* 104 (2007) 10318–10323.

[13] M. Molloy, The freezing threshold for k-colourings of a random graph, *Proceedings of STOC 2012*.

[14] E. Shamir and E. Upfal, Sequential and Distributed Graph Coloring Algorithms with Performance Analysis in Random Graph Spaces, *Journal of Algorithms* 5 (1984) 488–501.

[15] L. Zdeborová and F. Krzakąla, Phase Transitions in the Coloring of Random Graphs, *Physics Review E* 76 (2007).