Sparse graphs usually have exponentially many optimal colorings

Michael Krivelevich *

Department of Mathematics, Sackler Faculty of Exact Sciences,

Tel Aviv University, Tel Aviv 69978, Israel.

krivelev@post.tau.ac.il

Submitted: September 6, 2001; Accepted: June 8, 2002
MR Subject Classification: 05C80, 05C15

Abstract

A proper coloring of a graph $G = (V, E)$ is called optimal if the number of colors used is the minimal possible; i.e., it coincides with the chromatic number of G.

We investigate the typical behavior of the number of distinct optimal colorings of a random graph $G(n, p)$, for various values of the edge probability $p = p(n)$. Our main result shows that for every constant $1/3 < a < 2$, most of the graphs in the probability space $G(n, p)$ with $p = n^{-a}$ have exponentially many optimal colorings.

Given a graph $G = (V, E)$, an *unordered* partition $V = V_1 \cup \ldots \cup V_k$ is called a k-coloring, if each of the color classes V_i is an independent set of G. It is important to observe that we consider unordered partitions only, and therefore two k-colorings $(V_1, \ldots, V_k)$ and $(U_1, \ldots, U_k)$, for which there exists a permutation $\sigma \in S_k$ satisfying $V_i = U_{\sigma(i)}$, $1 \leq i \leq k$, are considered to be indistinguishable. A k-coloring $(V_1, \ldots, V_k)$ of G is *optimal*, if the number of colors is the minimal possible, i.e. $k = \chi(G)$, where $\chi(G)$ denotes as usually the chromatic number of G. Here are two simple examples to illustrate the above definitions: a) the graph $G = K^n - e$ has chromatic number $\chi(G) = n - 1$ and a unique optimal coloring; b) Define $G = (V, E)$ as follows: $V = A \cup B$, $A \cap B = \emptyset$, $|A| \geq 1$, $|B| = n \geq 2$; fix two distinct vertices $u, v \in B$ and define $E(G) = \{(a, b) : a \in A, b \in B\} \cup \{(u, v)\}$. Then it is easy to see that $\chi(G) = 3$ and G has exactly 2^{n-2} optimal colorings, where each optimal coloring has the following form: (V_1, V_2, V_3), where $V_1 = A$, $u \in V_2$, $v \in V_3$.

How many optimal colorings does a typical graph G on n vertices with given density $p = |E(G)|/\binom{n}{2}$ have? In order to address this question quantitatively we need to introduce a probability space of graphs on n vertices to make the notion of a "typical graphs"

*Supported by a USA-Israel BSF grant, by a grant from the Israel Science Foundation and a Bergmann Memorial Award

meaningful. We will make use of the probability space $G(n,p)$ of binomial random graphs. $G(n,p)$ is a random graph on n labeled vertices $\{1,\ldots,n\}$, where each pair $1 \le i < j \le n$ is chosen to be an edge independently and with probability $p = p(n)$. Sometimes with some abuse of notation we will use $G(n,p)$ to denote also a random graph on n vertices chosen according to the distribution induced by $G(n,p)$. As customary we will study *asymptotic* properties of the random graph $G(n,p)$. This means in particular that the number of vertices n will be assumed as large as necessary.

Equipped with this notation we can now reformulate our main question as follows: what is a typical behavior of the number of optimal colorings of a random graph drawn from $G(n,p)$? As our main result shows this number is exponentially large in n for small and moderate values of the edge probability $p = p(n)$. For simplicity we assume here that $p(n)$ has the form $p(n) = n^{-a}$ for a constant $a > 0$.

Theorem 1 *Let $\epsilon > 0$. Let $p(n) = n^{-a}$ for a constant $a > 0$.*

1. *If $\frac{1}{3} < a \le \frac{1}{2}$, then with probability at least $1-\epsilon$ a random graph $G(n,p)$ has at least $\exp\left\{\frac{\epsilon^2}{10} n^{\frac{3a-1}{2}}\right\}$ optimal colorings;*

2. *If $\frac{1}{2} < a < 1$, then with probability at least $1-\epsilon$ a random graph $G(n,p)$ has at least $\exp\left\{\frac{(1-a)\epsilon^2}{20} n^{\frac{a}{2}} \ln n\right\}$ optimal colorings.*

Thus for $1/3 < a < 1$ we get exponentially many optimal colorings in a typical graph from $G(n, n^{-a})$, where the exponent in the estimate of the number of optimal colorings grows with a. To complement the result observe that for all $1 \le a < 2$ the graph $G(n,p)$ contains almost surely (i.e. with probability tending to 1 as n tends to infinity) $\Theta(n)$ isolated vertices and is non-empty. These two conditions imply easily that the number of optimal colorings is $e^{\Theta(n)}$. With some effort Theorem 1 can be strengthened to the "almost sure" form, i.e. the graph $G(n,p)$ will have exponentially many optimal colorings not only with probability at least $1-\epsilon$, but also almost surely; this would result in some loss in the exponent.

Now we will prove our main result, Theorem 1. Denote

$$N = \binom{n}{2} .$$

Let $1 \le t \le n$, $T > t$ be integers.
For $0 \le i \le N$ we denote

$$\begin{aligned} a_i &= Pr[|E(G)| = i] , \\ b_i &= b_i(t) = Pr[\chi(G) \le t / |E(G)| = i] , \\ c_i &= c_i(t,T) = Pr[G \text{ has at least } \tfrac{1}{2} e^{\frac{T}{t} - \frac{T}{n}} \ t\text{-colorings} / |E(G)| = i, \chi(G) \le t] . \end{aligned}$$

Proposition 2 *With the above notation, if $i \le N - T$ and $b_i > 0$, then $c_i \ge \frac{2b_{i+T}}{b_i} - 1$.*

Proof. Define an auxiliary bipartite graph $H = (X \cup Y, F)$. The vertex set of H is a disjoint union of sets X and Y, where

$$\begin{aligned} X &= \{G : |V(G)| = n, |E(G)| = i, \chi(G) \le t\}, \\ Y &= \{G : |V(G)| = n, |E(G)| = i + T, \chi(G) \le t\}, \end{aligned}$$

and two graphs $G \in X$, $G' \in Y$ are connected by an edge in H if $E(G) \subset E(G')$. The definition of b_i implies that $|X| = b_i\binom{N}{i}$, $|Y| = b_{i+T}\binom{N}{i+T}$. As the property of being t-colorable is monotone decreasing, for every graph $G \in Y$ every subgraph of G with i edges is t-colorable. Hence we obtain:

$$|F| = |Y|\binom{i+T}{i} = b_{i+T}\binom{N}{i+T}\binom{i+T}{i}.$$

Let $X_0 = \left\{G \in X : deg_H(G) \le \frac{1}{2}\binom{N-i}{T}\right\}$. Then

$$\begin{aligned} |F| &\le |X_0|\frac{1}{2}\binom{N-i}{T} + |X \setminus X_0|\binom{N-i}{T} = |X|\binom{N-i}{T} - \frac{1}{2}|X_0|\binom{N-i}{T} \\ &= b_i\binom{N}{i}\binom{N-i}{T} - \frac{1}{2}|X_0|\binom{N-i}{T}. \end{aligned}$$

It follows from the above two estimates on $|F|$ that

$$|X_0| \le \frac{2b_i\binom{N}{i}\binom{N-i}{T} - 2b_{i+T}\binom{N}{i+T}\binom{i+T}{i}}{\binom{N-i}{T}} = 2(b_i - b_{i+T})\binom{N}{i}. \qquad (1)$$

Now we will prove that every graph in $X \setminus X_0$ has the required number of t-colorings. Indeed, let $G \in X \setminus X_0$. Clearly, if $E(G) \subset E(G')$ and G' is t-colorable, then some t-coloring of G is a valid coloring of G' as well. For a fixed t-coloring $V(G) = V_1 \cup \ldots \cup V_t$ of G, the number of graphs G' with $i+T$ edges and with $E(G) \subset E(G')$, for which $(V_i)_{i=1}^t$ is a proper coloring, is at most

$$\binom{N - i - \sum_{i=1}^t \binom{|V_i|}{2}}{T} \le \binom{N - i - \frac{n}{2}\left(\frac{n}{t} - 1\right)}{T}$$

by convexity of the function $f(x) = \binom{x}{2}$. As by the definition of X_0 the number of t-colorable graphs G' with $i + T$ edges for which $E(G) \subset E(G')$ is at least $\frac{1}{2}\binom{N-i}{T}$, we derive that the number of proper t-colorings of G is at least

$$\begin{aligned} \frac{\frac{1}{2}\binom{N-i}{T}}{\binom{N-i-\frac{n}{2}\left(\frac{n}{t}-1\right)}{T}} &\ge \frac{1}{2}\exp\left\{\frac{\frac{nT}{2}\left(\frac{n}{t}-1\right)}{N-i}\right\} \ge \frac{1}{2}\exp\left\{\frac{\frac{nT}{2}\left(\frac{n}{t}-1\right)}{N}\right\} = \frac{1}{2}\exp\left\{\frac{T\left(\frac{n}{t}-1\right)}{n-1}\right\} \\ &\ge \frac{1}{2}e^{\frac{T}{t}-\frac{T}{n}} \end{aligned}$$

(we used the bound $\binom{x}{y} \Big/ \binom{x-z}{y} \geq e^{yz/x}$ in the first inequality above). It follows then from the definition of c_i and from bound (1) that

$$c_i \geq \frac{|X \setminus X_0|}{|X|} \geq \frac{b_i\binom{N}{i} - 2(b_i - b_{i+T})\binom{N}{i}}{b_i\binom{N}{i}} = 1 - \frac{2(b_i - b_{i+T})}{b_i} = \frac{2b_{i+T}}{b_i} - 1 \ .$$

The proposition is proven. □

Recalling the above mentioned monotonicity of the property of being t-colorable, we obtain that if $b_i = 0$ for some i, then also $b_j = 0$ for all $j > i$. Therefore, the conclusion of Proposition 2 can be rewritten in the following form.

Corollary 3 *If $i \leq N - t$, then $b_i c_i \geq 2b_{i+T} - b_i$.*

Lemma 4 *Let $1 \leq t \leq n$, $t < T = o(n^2 p)$ be integers. Denote $\mu = Pr[\chi(G) \leq t]$. Then*

$$Pr[G \text{ has at least } \tfrac{1}{2}e^{\frac{T}{t}-\frac{T}{n}} \ t\text{-colorings}/\chi(G) \leq t] \geq 1 - (1 + o(1))\frac{2T}{\mu n\sqrt{p}} \ .$$

Proof. We will use again the notation defined before Proposition 2. Observe first that it follows from the definitions of μ, a_i's and b_i's that

$$\mu = \sum_{i=0}^{N} a_i b_i \ .$$

Also,

$$\sum_{i=N-T+1}^{N} a_i b_i \leq \sum_{i=N-T+1}^{N} a_i = Pr[|E(G)| > N - T] = o\left(\frac{T}{n\sqrt{p}}\right) \ ,$$

due to the standard estimates on the upper tail of a binomial random variable. Hence,

$$\sum_{i=0}^{N-T} a_i b_i \geq \mu - o\left(\frac{T}{n\sqrt{p}}\right) \ . \tag{2}$$

The definitions of a_i, b_i, c_i imply that

$$Pr[\Big(G \text{ has at least } \tfrac{1}{2}e^{\frac{T}{t}-\frac{T}{n}} \ t\text{-colorings}\Big) \&(\chi(G) \leq t)] = \sum_{i=0}^{N} a_i b_i c_i.$$

Therefore,

$$\begin{aligned}
& Pr[G \text{ has at least } \tfrac{1}{2}e^{\frac{T}{t}-\frac{T}{n}} \ t\text{-colorings}/\chi(G) \leq t] \\
= \ & Pr[\Big(G \text{ has at least } \tfrac{1}{2}e^{\frac{T}{t}-\frac{T}{n}} \ t\text{-colorings}\Big) \&(\chi(G) \leq t)](Pr[\chi(G) \leq t])^{-1} \qquad (3) \\
= \ & \frac{1}{\mu}\sum_{i=0}^{N} a_i b_i c_i \geq \frac{1}{\mu}\sum_{i=0}^{N-T} a_i b_i c_i \ . \qquad (4)
\end{aligned}$$

Our aim is to estimate from below the sum $\sum_{i=0}^{N-T} a_i b_i c_i$. Recall that by Corollary 3, $b_i c_i \geq 2b_{i+T} - b_i$. Therefore

$$\begin{aligned}\sum_{i=0}^{N-T} a_i b_i c_i &\geq \sum_{i=0}^{N-T} a_i(2b_{i+T} - b_i) = \sum_{i=0}^{N-T} (a_i b_i - 2a_i(b_i - b_{i+T})) \\ &\geq \sum_{i=0}^{N-T} a_i b_i - \sum_{i=0}^{N-T} \left(2 \max_{0 \leq i \leq N} a_i\right)(b_i - b_{i+T})\end{aligned}$$

Using well known estimates of the binomial coefficients, one can easily prove that $\max_{i=0}^{N} a_i \leq (n\sqrt{p})^{-1}$. Then applying (2), we can bound the last expression from below by:

$$\begin{aligned}\mu - o\left(\frac{T}{n\sqrt{p}}\right) - \frac{2}{n\sqrt{p}} \sum_{i=0}^{N-T} (b_i - b_{i+T}) &= \mu - o\left(\frac{T}{n\sqrt{p}}\right) - \frac{2}{n\sqrt{p}} \left(\sum_{i=0}^{T-1} b_i - \sum_{i=N-T+1}^{N} b_i\right) \\ &\geq \mu - (1 + o(1))\frac{2T}{n\sqrt{p}}\,.\end{aligned}$$

Substituting the above estimate into (4), we obtain

$$Pr[G \text{ has at least } \tfrac{1}{2} e^{\frac{T}{t} - \frac{T}{n}}\ t\text{-colorings}/\chi(G) \leq t] \geq 1 - (1 + o(1))\frac{2T}{\mu n\sqrt{p}}\,,$$

as promised. □

We are now in position to prove our main result, Theorem 1. The key ingredients in the proof are results on the concentration on the chromatic number of $G(n,p)$, due to Shamir and Spencer [6], Łuczak [4], and Alon and Krivelevich [1].

Recall that $p(n) = n^{-a}$ for $a > 1/3$. Consider first the case $1/3 < a \leq 1/2$. Set $\epsilon_0 = \epsilon/4$. The above mentioned result of Shamir and Spencer and its proof imply that in this case $\chi(G(n,p))$ is concentrated in width $n^{1/2} p \ln(np)$, or specifically, for large enough n there exists a $t_0 = t_0(n,p)$ so that:

1. $Pr[\chi(G) \leq t_0] \geq \epsilon_0$;
2. $Pr[t_0 \leq \chi(G) < t_0 + n^{1/2} p \ln(np)] \geq 1 - \epsilon_0$.

Let t_0 be as above. Set $I = [t_0, t_0 + n^{1/2} p \ln(np))$. Notice that as the asymptotic value of $\chi(G(n,p))$ is concentrated in I, due to the results on the asymptotic behavior of $\chi(G(n,p))$ ([2], [3]) every $t \in I$ satisfies $t = (1 + o(1))np/(2\ln(np))$.

Now, set

$$T = \frac{\epsilon_0^2 n^{1/2}}{p^{1/2} \ln(np)}\,.$$

Then it is immediate that for all $t \in I$,

$$\begin{aligned}\frac{1}{2}\exp\left\{\frac{T}{t}-\frac{T}{n}\right\} &= \frac{1}{2}\exp\left\{(1+o(1))\frac{\frac{\epsilon_0^2 n^{1/2}}{p^{1/2}\ln(np)}}{\frac{np}{2\ln(np)}}\right\} = \frac{1}{2}\exp\left\{(1+o(1))\frac{2\epsilon_0^2}{n^{1/2}p^{3/2}}\right\} \\ &> \exp\left\{\frac{\epsilon^2}{10}n^{\frac{3a-1}{2}}\right\} .\end{aligned}$$

Set $K = e^{\frac{\epsilon^2}{10}n^{\frac{3a-1}{2}}}$. Let A denote the event "G has less than K optimal colorings". Let also A_t be the event "G has less than K t-colorings". Notice that for graphs G with $\chi(G) = t$ the events A and A_t coincide. Then it follows from Lemma 4, the choice of t_0 and the above estimate:

$$\begin{aligned}Pr[A] &\leq \sum_{t\notin I} Pr[\chi(G)=t] + \sum_{t\in I} Pr[\chi(G)=t]Pr[A|\chi(G)=t] \\ &= \sum_{t\notin I} Pr[\chi(G)=t] + \sum_{t\in I} Pr[\chi(G)=t]Pr[A_t|\chi(G)=t] \\ &\leq \epsilon_0 + \sum_{t\in I} Pr[\chi(G)=t]\frac{Pr[A_t|\chi(G)\leq t]}{Pr[\chi(G)=t|\chi(G)\leq t]} \\ &\leq \epsilon_0 + \sum_{t\in I} Pr[\chi(G)=t]\frac{Pr[A_t|\chi(G)\leq t]}{Pr[\chi(G)=t]} \\ &\leq \epsilon_0 + \sum_{t\in I}\frac{(1+o(1))2T}{np^{1/2}Pr[\chi(G)\leq t]} \\ &\leq \epsilon_0 + \frac{(1+o(1))2|I|T}{\epsilon_0 np^{1/2}} = \epsilon_0 + \frac{(1+o(1))2n^{1/2}p\ln(np)}{\epsilon_0 np^{1/2}}\frac{\epsilon_0^2 n^{1/2}}{p^{1/2}\ln(np)} < 4\epsilon_0 = \epsilon .\end{aligned}$$

The case $1/3 < a \leq 1/2$ is completed.

Now we treat the remaining case $1/2 < a < 1$. The argument here is quite similar, with only significant difference being the availability of a stronger concentration result for the chromatic number.

Set $\epsilon_0 = \epsilon/6$. Alon and Krivelevich proved in [1] that for large enough n there exists a $t_0 = t_0(n,p)$ so that

1. $Pr[\chi(G) \leq t_0] \geq \epsilon_0$;
2. $Pr[\chi(G) \in \{t_0, t_0+1\}] \geq 1-\epsilon_0$.

(We would like to mention that Shamir and Spencer [6] and Łuczak [4] proved somewhat weaker results, still showing concentration of $\chi(G(n,p))$ in an interval of a fixed length. Their results would also suffice for our purposes here.)

Let t_0 be as above. Set $I = \{t_0, t_0+1\}$. Again due to the results on the asymptotic value of $\chi(G(n,p))$ one gets $t_0 = (1+o(1))np/(2\ln(np))$.

Set this time

$$T = \epsilon_0^2 np^{1/2} .$$

Then for both $t \in I$,

$$\begin{aligned}\frac{1}{2}\exp\left\{\frac{T}{t}-\frac{T}{n}\right\} &= \frac{1}{2}\exp\left\{(1+o(1))\frac{\epsilon_0^2 np^{1/2}}{\frac{np}{2\ln(np)}}\right\} = \frac{1}{2}\exp\{(1+o(1))2\epsilon_0^2 p^{-1/2}\ln(n^{1-a})\} \\ &> \exp\left\{\frac{(1-a)\epsilon^2}{20}n^{a/2}\ln n\right\} .\end{aligned}$$

Set $K = e^{\frac{(1-a)\epsilon^2}{20}n^{a/2}\ln n}$. Let A be the event "G has less than K optimal colorings". Let also A_t be the event "G has less than K t-colorings". Then, similarly to the previous case, Lemma 4 provides:

$$\begin{aligned}Pr[A] &\leq \sum_{t\notin I} Pr[\chi(G)=t] + \sum_{t\in I} Pr[\chi(G)=t]Pr[A_t|\chi(G)=t] \\ &\leq \epsilon_0 + \frac{(1+o(1))2|I|T}{\epsilon_0 np^{1/2}} = \epsilon_0 + \frac{(1+o(1))4}{\epsilon_0 np^{1/2}}\epsilon_0^2 np^{1/2} < \epsilon_0 + 5\epsilon_0 = 6\epsilon_0 = \epsilon .\end{aligned}$$

The theorem is proven. □

We have thus proven that most of the graphs in the probability space $G(n,p)$ with $p = n^{-a}$, $1/3 < a < 2$, have exponentially many optimal colorings. A close examination of the proof reveals that tighter concentration results for the chromatic number of $G(n,p)$ would translate immediately to better bounds on the number of optimal colors for the case $1/3 < a \leq 1/2$ and possibly would enable to extend the result to higher values of the edge probability $p(n)$, i.e. to smaller values of a. So far the dense case remains completely open. We conjecture that almost surely the random graph $G(n,p)$ has at least superpolynomially many in n optimal colorings as long as the edge probability $p(n)$ satisfies $p(n) \leq 1-\epsilon$ for a constant $\epsilon > 0$. A particularly appealing case is that of the edge probability $p = 0.5$ – the most studied random graph.

In the opposite direction, it would be quite interesting to bound from *above* a typical number of optimal colorings. We believe that the bounds presented in Theorem 1 are very far from being tight, but at present we are unable to improve them.

One may also study the structure of optimal colorings in $G(n,p)$. In particular, how many pairs of unconnected vertices are typically *rigid*, i.e. are in the same color class in every optimal coloring of G? Partial results can be obtained applying the ideas similar to those used in the proof of Theorem 1, but in general this remains open.

Information about the number of optimal colorings of a graph G is encoded by the *chromatic polynomial* of G, $p_G(x)$ (see, e.g. [5]). By definition, for a positive integer $x \geq 1$, the value $p_G(x)$ is equal to the number of x-colorings of G (this time ordered ones, for example $p_{K_n}(x) = x(x-1)\ldots(x-n+1)$). Studying coefficients and values of the chromatic polynomial of a random graph $G(n,p)$ appears to be an attractive task, see [7] for some related results.

Finally, one can study similar quantitative questions about other graph parameters such as optimal independent sets or optimal matchings.

References

[1] N. Alon and M. Krivelevich, *The concentration of the chromatic number of random graphs*, Combinatorica 17 (1997), 303–313.

[2] B. Bollobás, *The chromatic number of random graphs*, Combinatorica 8 (1988), 49–55.

[3] T. Łuczak, *The chromatic number of random graphs*, Combinatorica 11 (1991), 45–54.

[4] T. Łuczak, *A note on the sharp concentration of the chromatic number of random graphs*, Combinatorica 11 (1991), 295–297.

[5] R. C. Read and W. T. Tutte, *Chromatic polynomials*, in: Selected Topics in Graph Theory (L. W. Beineke and R. M. Wilson, Eds.), pp. 15-42, Academic Press, New York, 1988.

[6] E. Shamir and J. Spencer, *Sharp concentration of the chromatic number of random graphs $G_{n,p}$*, Combinatorica 7 (1987), 124–129.

[7] D. J. A. Welsh, *Counting colourings and flows in random graphs*, in: Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), pp. 491–505, Bolyai Soc. Math. Stud., 2, János Bolyai Math. Soc., Budapest, 1996.