Appeared in *FOCS 97*

The Analysis of a List-Coloring Algorithm on a Random Graph (extended abstract)

Dimitris Achlioptas*† Michael Molloy*‡
(optas@cs.toronto.edu) (molloy@cs.toronto.edu)

Abstract

We introduce a natural k-coloring algorithm and analyze its performance on random graphs with constant expected degree c ($G_{n,p=c/n}$). For $k=3$ our results imply that almost all graphs with n vertices and $1.923\,n$ edges are 3-colorable. This improves the lower bound on the threshold for random 3-colorability significantly and settles the last case of a long-standing open question of Bollobás [5]. We also provide a tight asymptotic analysis of the algorithm. We show that for all $k \geq 3$, if $c \leq k \ln k - 3/2k$ then the algorithm almost surely succeeds, while for any $\epsilon > 0$, and k sufficiently large, if $c \geq (1+\epsilon) k \ln k$ then the algorithm almost surely fails. The analysis is based on the use of differential equations to approximate the mean path of certain Markov chains.

1 Introduction

Let $G_{n,p}$ denote the random graph on n vertices where each of the $\binom{n}{2}$ possible edges is present, independently of all others, with probability $p = p(n)$. The range of p that we will be interested in is $p = \Theta(n^{-1})$. For a graph property Q, we say that $G_{n,p}$ *almost surely (a.s.)* has Q if $\lim_{n\to\infty} \mathbf{Pr}(G_{n,p} \text{ has } Q) = 1$. One of the first observations made about $G_{n,p}$ by Erdös and Rényi [12] in their seminal work on random graph theory was the existence of *threshold* phenomena, the fact that for many interesting properties Q, the probability of $G_{n,p}$ having Q exhibits a sharp increase at a certain critical value of p. This notion of "sharpness" and "sudden appearance" is certainly not limited to random graphs and in fact it permeates numerous different settings of probabilistic product spaces, ranging from coin-flipping in distributed computing, to random-satisfiability, to statistical physics. Quantifying "sharp" and deriving general conditions for the "sudden appearance" of properties has been studied in [6, 7, 14, 15, 20, 31, 33], among numerous others.

Recently, Friedgut [14] made very significant progress in our understanding of this area by establishing necessary and sufficient conditions for a property to have a sharp threshold. Roughly speaking, he showed that a property has a "coarse" threshold (as opposed to a sharp one) if and only if it can be well-approximated by the property of containing some graph from a list of "small" graphs. Using the main theorem of [14], Achlioptas and Friedgut [1] showed that the chromatic number of a random graph exhibits a sharp threshold, in the sense that for $k \geq 3$, there exists $c_k(n) > 0$, such that for any $c < c_k(n)$, a.s. $\chi(G_{n,p=c/n}) \leq k$, while for any $c > c_k(n)$, a.s. $\chi(G_{n,p=c/n}) > k$. It is widely conjectured that $\lim_{n\to\infty} c_k(n)$ converges to an absolute constant c_k. Determining c_k, even for $k=3$, remains one of the most important open problems in the field of random graphs and it is closely related to the problem of determining the ratio of clauses to variables at which a random instance of k-SAT turns from a.s. satisfiable to a.s. unsatisfiable. For recent progress in this latter problem see [2, 14, 17, 22].

Łuczak [24] proved that asymptotically $c_k \sim 2k \ln k$ by showing the existence of suitably large, disjoint independent sets in $G_{n,p=c_k/n}$. Algorithmically, substantially less progress has been made. One of the first heuristics to be analyzed [18],[4], GIC, forms color classes by repeatedly removing greedily chosen independent sets until the remaining graph has no component with more than one cycle, at which point it is easy to 3-color it . Shamir and Upfal [32] and Fernandez de la Vega [13] showed that for any $\epsilon > 0$, there exists $k \geq k(\epsilon)$ such that if $c < (1+\epsilon) k \ln k$ then GIC a.s. succeeds in k-coloring $G_{n,p=c/n}$, while if $c > (1+\epsilon) k \ln k$ then GIC a.s. fails to k-color $G_{n,p=c/n}$. Different variations of this heuristic have been shown to have the same asymptotic performance [18]. (If we continue to remove independent sets greedily until the end of the execution, then the heuristic becomes quite inefficient [29] and for any $c > 1$, the number of colors it uses grows with n.) It is a major open question [16] whether for any $\epsilon > 0$ there is a polynomial time algorithm which will a.s. k-color, $G_{n,p=c/n}$ for arbitrarily high k ,when $c = (1+\epsilon) k \ln k$.

A different approach to determining c_k is motivated by the following: in many cases, a graph property P has the same threshold as a much weaker graph property P'. For example, the property "has a perfect matching" has the same threshold as the property "has minimum degree at least one", and the property "has a Hamilton cycle" has the same threshold as the property "has minimum degree at least two". Inspired by these, and other properties, Bollobás asked whether the property "has chromatic number greater than k" has the same threshold as the property "has

*Department of Computer Science, University of Toronto, Toronto, ON M5S 3G4, Canada.

§Supported by an NSERC (Natural Sciences and Engineering Research Council) PGS B Scholarship.

‖Supported in part by an NSERC grant.

a subgraph with minimum degree at least k". (It is a well-known fact that if $\chi(G) > k$ then G must have such a subgraph.) In 1991, Pittel, Spencer and Wormald [30] determined the threshold $p = \gamma_k/n$ for $G_{n,p}$ to have a subgraph with minimum degree at least k, for each $k \geq 3$, e.g. $\gamma_3 = 3.35\ldots$ (see [10] and [26] for some earlier bounds on this threshold, and hence c_k). Molloy [27] observed that this, along with some results of Bollobás and Thomasson [5] regarding GIC, imply that the answer to Bollobás' question is negative for $k \geq 4$. In particular, he showed that $c_4 \geq 5.16\ldots > 5.14\ldots = \gamma_4$.

In this paper, we introduce and analyze a greedy "list-coloring" algorithm. The algorithm runs by maintaining for each uncolored vertex v, a list of available colors L_v. If no list becomes empty, then the algorithm successfully finds a coloring of the graph. More specifically, the algorithm is defined as follows:

k-GREEDY-LIST (k-GL)

- `Initialize:` $U = V(G)$`, and` $L_v = \{1, ..., k\}$ `for every` $v \in U$.
- `While` $U \neq \emptyset$
  1. `Pick a vertex` v `uniformly at random from` $\{u \in U : |L_u|$ `is smallest` $\}$.
  2. `If` $|L_v| > 0$ `then assign to` v `a color` ω `chosen uniformly at random from` L_v `else exit (fail).`
  3. `For each` $u \in U$ `that is adjacent to` v `set` $L_u = L_u \backslash \{\omega\}$. `Set` $U = U - \{v\}$.

To aid the analysis, when $|L_v| = 0$ instead of exiting it will be more convenient to:
`assign to` v `a color` ω `chosen uniformly at random from` $\{1, 2, ..., k\}$. `Label` v `bad`.

Hence, the algorithm continues until all vertices receive colors. Clearly, if no vertex is labelled `bad` when $U = \emptyset$ then k-GL has k-colored G properly, and we say that k-GL *succeeds*. Let

$$d_k = \sup\{c \mid \mathbf{Pr}[k\text{-GL succeeds on } G_{n,p=c/n}] > 0\} \ .$$

We show that for $c < d_k$, k-GL succeeds with probability $\epsilon(c) > 0$, while for $c > d_k$, k-GL a.s. fails. Let $\mathbf{W}_{-1}$ denote the -1st branch of the $\mathbf{W}$ function [11], defined by $\mathbf{W}(x)\exp(\mathbf{W}(x)) = x$.

Theorem 1 $d_3 = \mathbf{W}_{-1}(-\exp(5/2)) = 3.847\ldots$

Let

$$f_k(c) = \lim_{n\to\infty} \mathbf{Pr}[\chi(G_{n,p=c/n}) \leq k] \ .$$

From [1], it follows that for all $c, \epsilon > 0$,

$$\text{if } f_k(c) \geq \epsilon \text{ then } f_k(c') = 1, \text{ for } c' < c \ .$$

This fact along with theorem 1 imply that $c_3 \geq d_3$. Since k-colorability is an edge-monotone property, a well-known equivalence between the two random graph models [3] implies

Corollary 1 *Almost all graphs with n vertices and $1.923\,n$ edges are 3-colorable.*

By theorem 1, we have $c_3 > 3.84 > 3.35\ldots = \gamma_3$ proving that the answer to Bollobás' question is negative for $k = 3$ as well. This settles the question completely and indicates that the gap is significantly greater than previously known.

Corollary 2 *For all $k \geq 3$, there is $\delta > 0$ such that for $\gamma_k < c < (1+\delta)\gamma_k$, a.s. $G_{n,p=c/n}$ is k-colorable and has a subgraph with minimum degree at least k.*

We will see that although we can determine analytic expressions for d_k for all $k \geq 3$, computing a good approximation of d_k for $k \geq 4$ seems to be a difficult numerical problem. In spite of this fact we will prove that the behavior of k-GL on $G_{n,p}$ exhibits a sharp threshold, in the following sense:

Theorem 2

a. For all $k \geq 3$, if $c \leq k \ln k - 3/2k$ then k-GL a.s. succeeds.

b. For any $\epsilon > 0$, there exists $k(\epsilon)$ such that for all $k \geq k(\epsilon)$, if $c \geq k \ln k(1+\epsilon)$ then k-GL a.s. fails.

Hence, although k-GL is quite good for small k, asymptotically it is "equivalent" to GIC.

1.1 Outline of the analysis

For $0 \leq i \leq k$, we denote by S_i the set of uncolored vertices v with $|L_v| = i$. Conditioning on the size of each S_i at a particular step, we can determine the expected change in each size during the next step. This allows us to model the behavior of these sizes throughout the algorithm's execution with a set of functions $s_1(x), ..., s_k(x)$ defined by a system of differential equations. It follows from the main theorem of [34] that for $0 < x < 1$, after roughly xn vertices have been colored, a.s. $|S_i| = s_i(x)n + \mathrm{o}(n)$. This is a common technique – for other applications see [8, 9, 17, 21, 28, 30, 34] – and we omit a detailed justification in this extended abstract.

We show that if $s_1'(x) < 0$ for all $0 \leq x \leq 1$, then the algorithm will succeed with constant positive probability. Intuitively, this means that if vertices enter S_1 slowly enough (at a rate less than one per step so that $s_1'(x) < 0$) then we can take care of them one at a time. On the other hand, if $s_1'(x) > 0$ for some $0 < x < 1$, then S_1 a.s. becomes quite large, (of size $\Theta(n)$) and it will a.s. eventually contain two adjacent vertices with the same unique color in their lists, so the algorithm will fail. Upon establishing these facts, determining d_k reduces to solving the associated system of differential equations. For $k \geq 4$ the system of differential equations along with its initial conditions yields a particularly unwieldy formula for d_k. To prove theorem 2, we will derive two new sets of initial conditions that are simpler to deal with. For theorem 2.a the set of initial conditions corresponds to analyzing a variation of k-GL where we initialize the lists of the vast majority of vertices to contain much fewer than k colors. For theorem 2.b the set of initial conditions does not have an algorithmic interpretation but we prove that the functions derived under these conditions lower bound the functions s_i for all i. Rather surprisingly, theorem 2.b indicates that the severe handicap imposed to prove theorem 2.a does not substantially affect the asymptotic performance of the algorithm.

We will omit floors and ceilings throughout, unless this leads to ambiguity.

2 Preliminaries

In our analysis, for all k, we assume that the random graph is not given to us in its entirety at the start of the algorithm. Instead we will learn about the graph as the algorithm proceeds. This scenario has been aptly named the *method of deferred decisions* by Knuth, Motwani and Pittel [23]. In particular, it is easy to see that at any step, the edges that have been examined (exposed) by k-GL are precisely the edges incident to colored vertices.

We say that k-GL is presently at time t (step t) if it is in the process of carrying out the $(t+1)$st iteration of the main loop. In other words, at time t, exactly t vertices have been colored thus far. Let $S_i(t)$ denote the random set of vertices whose lists have length i (contain i colors) at time t (for $i > k$, $S_i(t) = \emptyset$ for all t). We will often refer to the "lower" $i+1$ sets collectively, so we let

$$U_i(t) = \bigcup_{j=0}^{i} S_j(t) \ .$$

We will sometimes refer to U_i or S_i when no ambiguity arises. We define $\Delta S_i(t) = |S_i(t+1)| - |S_i(t)|$ and $\Delta U_i(t) = |U_i(t+1)| - |U_i(t)|$. For $\ell \subseteq \{1,\ldots,k\}$, $|\ell| = i$, let $S_i^\ell(t) = \{v \in S_i(t) : L_v = \ell\}$. We prove the following fact at the end of section 3.

Claim 1 *For every $0 \le i \le k$, if ℓ_1, ℓ_2 are i-subsets of $\{1,\ldots,k\}$, then for every $0 \le t \le n$,*

$$|S_i^{\ell_1}(t)| - |S_i^{\ell_2}(t)| = \mathrm{o}(n) \ .$$

¿From claim 1, we get

Lemma 1 *For every $1 \le i \le k$, and for every $1 \le t \le n$, the expected number of vertices which move from $S_i(t)$ to $S_{i-1}(t+1)$ during step t is $\frac{ic}{kn}|S_i(t)| + \mathrm{o}(1)$.*

Proof Suppose that the vertex being colored is $v \in S_j(t)$. Recall that until step $t+1$ we have exposed nothing about edges between v and U. Therefore, each $u \in S_i(t)$ is adjacent to v with probability $p = c/n$. For each $\omega \in L_u$, the probability that $\omega \in L_v$ is, by claim 1, j/k. By the definition of k-GL each color in L_v is assigned to v with probability $1/j$. Hence, since $|L_u| = i$, the probability that u moves to $S_{i-1}(t+1)$ is $\frac{ic}{kn}$ and the lemma follows from linearity of expectation. □

¿From lemma 1, we can determine the expected change in each S_i at each step, and thus in each U_i. As long as $U_i(t) \neq \emptyset$, by the definition of the algorithm, we will be removing exactly one vertex from U_i in each step. Since the list of each vertex shrinks by at most one in each step, the vertices entering U_i are precisely those moving from S_{i+1} to S_i. Hence, by lemma 1, when $U_i(t) \neq \emptyset$,

$$\mathbf{E}[\Delta U_i(t)] = \frac{(i+1)c}{kn}|S_{i+1}(t)| - 1 + \mathrm{o}(1) \qquad (1)$$

and similarly when $U_{i-1}(t) \neq \emptyset$

$$\mathbf{E}[\Delta S_i(t)] = \frac{(i+1)c}{kn}|S_{i+1}(t)| - \frac{ic}{kn}|S_i(t)| + \mathrm{o}(1) \ . \qquad (2)$$

The fact that (1), (2) are valid only if $U_i(t) \neq \emptyset$, $U_{i-1}(t) \neq \emptyset$, respectively will be the main complication of the analysis. We will often need to focus on the expected number of vertices entering S_i (and hence U_i) at step t so we let

$$\begin{aligned} \mathrm{In}_i(t) &= \frac{(i+1)c}{kn}|S_{i+1}(t)| \\ &= \frac{(i+1)c}{kn}\left(|U_{i+1}(t)| - |U_i(t)|\right) \ . \qquad (3) \end{aligned}$$

Lemma 2

(a) Let $t_0 \ge 0$ and $\delta > 0$ be such that $|U_i(t_0)| = 0$ and $\mathrm{In}_i(t) < 1 - \delta$, for $t \ge t_0$. Then with probability at least ρ, for some $\rho > 0$ independent of n, $\bigcup_{t \ge t_0} U_{i-1}(t) = \emptyset$, and a.s. $\bigcup_{t \ge t_0} U_{i-2}(t) = \emptyset$.

(b) Let $t_0 \ge 0$ and $\delta > 0$ be such that $|U_i(t_0)| = \mathrm{o}(n)$ and $\mathrm{In}_i(t) < 1 - \delta$, for $t \ge t_0$. Then a.s. $U_i(t) = \mathrm{o}(n)$, for $t \ge t_0$.

(c) Let $t_0 \ge 0$ and $\delta, \epsilon > 0$ be such that $\mathrm{In}_i(t) > 1 + \delta$, for $t_0 < t < t_0 + \epsilon n$. Then a.s. $\bigcup_{t \ge t_0} U_{i-1}(t) \neq \emptyset$.

Proof
(a): Let $\mathrm{Po}(\mu)$ denote a Poisson distributed variable with mean μ. At each step, the distribution of the number of vertices which move from S_{i+1} to S_i is asymptotic to $\mathrm{Po}(\mathrm{In}_i)$. Thus, for $t \ge t_0$, $|U_i(t)|$ is statistically dominated by the random variable $Z(t)$ defined by $Z(t_0) = 0$ and $Z_{t+1} = \max\{0, Z_t + J_t\}$ where $\mathbf{Pr}[J_t = j] = \mathbf{Pr}[\mathrm{Po}(1-\delta) = j+1]$, for each $j \ge -1$. Since $\mathbf{E}(J_t) < -\delta$, it is straightforward (but tedious) to show that

$$\text{a.s.} \qquad \sum_{t=t_0}^{n} |U_i(t)| = \mathrm{O}(n) \ . \qquad (4)$$

Thus, the probability that no vertices enter U_{i-1} after time t_0 is at least $\rho = (1 - \frac{c}{n})^{\mathrm{O}(n)} > 0$. The expected number of vertices which enter U_{i-1} after time t_0 is at most $\sum_{t \ge t_0} |U_i(t)| \times \frac{c}{n} = \mathrm{O}(1)$. ¿From Markov's inequality we get $|\bigcup_{t \ge t_0} U_{i-1}(t)| < \log n$ a.s. This last property implies that a.s. $\bigcup_{t \ge t_0} U_{i-1}(t)$ contains no two adjacent vertices which in turn implies that a.s. $\bigcup_{t\geq t_0} U_{i-2}(t) = \emptyset$.

(b): The proof of this fact follows similarly to the proof of part (a).

(c): In this case, for $t_0 < t < t_0 + \epsilon n$, $|U_i(t)|$ statistically dominates the random variable $Z(t)$ defined by $Z(t_0) = 0$ and $Z_{t+1} = \max\{0, Z_t + J_t\}$ where $\mathbf{Pr}[J_t = j] = \mathbf{Pr}[\mathrm{Po}(1+\delta) = j+1]$, for each $j \geq -1$. Since $\mathbf{E}(J_t) > \delta$, there a.s. exist t_b and $\zeta > 0$ such that $|U_i(t_b)| > i\zeta n$, implying that $U_i(t_b)$ contains ζn vertices with the same list. If there is an edge between any two of these vertices then $\bigcup_{t\geq t_0} U_{i-1}(t) \neq \emptyset$. The probability that there is no such edge is $(1-\frac{c}{n})^{\binom{\zeta n}{2}} = \mathrm{o}(1)$. $\square$

3 Three colors

It is easy to show that if all the components of a graph G have at most one cycle then 3-GL succeeds on G with probability one. For $G_{n,p=c/n}$, it is well-known (see e.g. [12, 19]) that if $c < 1$ then a.s. all its components are unicyclic. Thus, we will assume $c > 1$ for the rest of the paper.

Lemma 3 *There is some $\epsilon = \epsilon(c) > 0$ such that a.s. for every $\log^2 n \leq t \leq \epsilon n$, $|U_2(t)| > 0$.*

Proof We have that $\mathbf{E}[\Delta U_2(t)] \geq c/n|S_3(t)| - 1$, and hence for $\delta = \frac{c-1}{2} > 0$, if $S_3(t) > \frac{c+1}{2c}n$ then $\mathbf{E}[\Delta U_2(t)] > \delta$. Since $S_3(0) = n$ and $\mathbf{E}[\Delta S_3(t)] \geq -c/n|S_3(t)| - 1 \geq -(c+1)$ there exist $\epsilon > 0$ such that a.s. $\mathbf{E}[\Delta U_2(t)] > \delta$, for $0 \leq t \leq \epsilon n$. The claim follows immediately. $\square$

Definition 1 *Let $t_2^e = \min\{t \mid t > \epsilon n, U_2(t) = \emptyset\}$, where ϵ is as in lemma 3.*

Lemma 4 *A.s. no vertex will be labelled* bad *after time t_2^e.*

Proof Note that $\mathrm{In}_2(0) = c > 1$, $\mathrm{In}_2(n) = 0$ and that $\mathrm{In}_2(t)$ is non-increasing with t. Denote by t^* the first step such that $\mathrm{In}_2(t) < 1$ and note that, as we argued in the proof of lemma 3, a.s. $t^* > \epsilon n$, for some $\epsilon > 0$. Since $\mathbf{E}[\Delta U_2(t)] \geq \mathrm{In}_2(t) - 1$, it is straightforward to verify that there exists $\delta_1 > 0$ such that a.s. $|U_2(t)| > \delta_1 n$ for $\epsilon n \leq t < t^*$. Hence, since $\Delta U_2(t) \geq -1$, we have that a.s. $t_0 > t^* + \delta_1 n$. From this last fact, since $\mathbf{E}[\Delta S_3(t)] \leq -\mathrm{In}_2(t)$, it follows that there exists $\delta_2 > 0$ such that a.s. $\mathrm{In}_2(t_0) < 1 - \delta_2$. The lemma follows from applying (a) of lemma 2 with $i = 2$, $\delta = \delta_2$, $t_0 = t_2^e$. $\square$

Lemma 3 implies that for $\log^2 n \leq t \leq \epsilon n$, a.s. $\mathbf{E}[\Delta S_3(t)] = -\mathrm{In}_2(t)$ and $\mathbf{E}[\Delta U_2(t)] = \mathrm{In}_2(t) - 1$. Inspired by this fact, we introduce the following two differential equations

$$\begin{aligned} s_3'(x) &= -cs_3(x)\ , & s_3(0) = 1 \\ u_2'(x) &= cs_3(x) - 1\ , & u_2(0) = 0\ . \end{aligned}$$

The unique solutions are $s_3(x) = \exp(-cx), u_2(x) = 1 - x - \exp(-cx)$. Let x_0 denote the smallest root of $u_2(x)$ in $(0, 1]$. The following follows immediately from the main theorem of [34].

Fact 1 *For $0 \leq t \leq t_2^e$, $|S_3(t)| = s_3(t/n)n + \mathrm{o}(n)$, and $|U_2(t)| = u_2(t/n)n + \mathrm{o}(n)$.*

In view of lemma 2, our goal is to derive a condition on c such that for all $0 \leq t \leq n$,

$$\mathrm{In}_1(t) = \frac{2c}{3n}|S_2(t)| < 1 - \epsilon, \text{ for some } \epsilon > 0\ .$$

By (a) of lemma 2, such a condition would in fact imply $|U_1(t)| = \mathrm{o}(n)$ and hence $|U_2(t)| = |S_2(t)| + \mathrm{o}(n)$ for all $0 \leq t \leq n$. Therefore, the threshold for positive probability of success will be the value of c for which $\max_{x\in[0,x_0]} \frac{2c}{3}u_2(x) = 1$. It is easy to verify that in $[0, x_0]$, u_2 is unimodal (in fact u_2 is unimodal in $[0, 1]$) and in particular $u_2(x) \leq u_2(\ln c/c) = (c - \ln c - 1)/c$. For $c > 1$, $\frac{2}{3}(c - \ln c - 1) = 1$, for $c = \eta \equiv \mathbf{W}_{-1}(-\exp(5/2)) = 3.847\ldots$, where $\mathbf{W}_{-1}$ is the -1st branch of the $\mathbf{W}$ function, defined by $\mathbf{W}(x)\exp(\mathbf{W}(x)) = x$.

Theorem 1 $d_3 = \mathbf{W}_{-1}(-\exp(5/2)) = 3.847\ldots$

Proof of Theorem 1 For $c > \eta$, from fact 1, there exist $x \leq x_0$ and $\delta, \epsilon > 0$ such that $\mathrm{In}_1(t) > 1 + \delta$ for $(x - \epsilon)n \leq t \leq (x + \epsilon)n$. Thus, (c) of lemma 2 implies that a.s. 3-GL will fail. For $c < \eta$, from fact 1, there exists $\delta = \delta(c) > 0$ such that a.s. for all $0 \leq t \leq t_2^e$, $\mathrm{In}_1(t) < 1 - \delta$. Hence, applying (a) of lemma 2 for $0 \leq t \leq t_2^e$ and invoking lemma 3 for $t_2^e \leq t \leq n$ concludes the proof. $\square$

Proof of Claim 1 (sketch) We first sketch the proof for $k = 3$ and then for general k.

The claim is trivially true for the only list of length three. For the three lists of length two, first note that in establishing lemma 4, we only refer to $\mathbf{E}[\Delta U_2(t)]$, the derivation of which does *not* rely on claim 1 (vertices move from S_3 to U_2 iff they are adjacent to the vertex being colored). Also, for $t \geq t_2^e$ the claim is trivially true since a.s. $U_2(t) = \mathrm{o}(n)$ for all such t by (b) of lemma 2. To prove the claim for $t < t_2^e$, instead of developing one differential equation for $|S_2(t)|$, we develop a system of three differential equations, one for each $|S_2^\ell(t)|$. The three differential equations are "symmetric" and moreover, the initial conditions are identical for all ℓ, $s_2^\ell(0) = 0$. Hence, $s_2^{\ell_1} \equiv s_2^{\ell_2}$ for all ℓ_1, ℓ_2 and the claim follows the main theorem of [34].

For general k the ingredients of the proof are similar. The claim is trivially true for the only list of length k and in general the equidistribution of the $\binom{k}{i}$ lists in S_i implies the validity of the expression for $\mathbf{E}[\Delta U_{i-1}(t)]$. Arguing as we did for S_2 when $k = 3$, we inductively derive the equidistribution of the $\binom{k}{i-1}$ lists in S_{i-1} . $\square$

4 The analysis for general k

We define the following system of differential equations, which we denote by $\mathcal{S}$. We set $u_k(x) = 1 - x$ and for $1 \le i \le k-1$,

$$x_i^s = \begin{cases} \min\{x \mid c\frac{i+1}{k}u_{i+1}(x) \ge 1\}, \\ 1, \text{ if no } x \text{ as above exists} \end{cases}$$

$$u_i(x) = 0\,, \text{ for } x \in [0, x_i^s]$$

$$u_i'(x) = c\frac{i+1}{k}(u_{i+1}(x) - u_i(x)) - 1\,, \text{ for } x \ge x_i^s\ .$$

Note that the system is triangular and x_i^s is defined strictly in terms of u_{i+1}, so $\mathcal{S}$ is well-defined. For example, since $u_k(x) = 1 - x$ and $c > 1$, $x_{k-1}^s = 0$ (compare to lemma 3) and so $u_{k-1}'(x) = c\,(u_k(x) - u_{k-1}(x)) - 1$.

Definition 2 *If $x_i^s = 1$ then we say that u_i is* degenerate. *If u_i is non-degenerate let $x_i^e = \min\{x > x_i^s \mid u_i(x) = 0\}$, otherwise let $x_i^e = 1$.*

It is straightforward to show that if u_i is degenerate then so is u_j for all $j < i$. Furthermore, it is also straightforward to verify that for all non-degenerate u_i, $0 \equiv x_k^s \le ... \le x_i^s \le x_i^e \le ... \le x_k^e \equiv 1$. The following analogue of fact 1 holds:

Fact 2 *For any non-degenerate u_i, a.s. for every $x_i^s < x < x_i^e$, $|U_i(xn)| = u_i(x)n + \mathrm{o}(n)$.*

As in Section 3, k-GL a.s. fails if $\mathrm{In}_1(x)$ ever exceeds 1, and if $\mathrm{In}_1(x)$ is always less than 1, then it succeeds with probability at least ρ for some $\rho > 0$ independent of n. The following lemma, which we state without proof, is quite helpful.

Lemma 5 *If u_i is non-degenerate then $\mathrm{In}_i(t)$ is maximized for some $0 \le t \le x_i^e n$ and $u_i(x) \le 0$ for all $x \in [x_i^e, 1]$.*

Thus our analysis of the algorithm for any k and c is reduced to solving $\mathcal{S}$ and determining whether

$$\lambda(k,c) = \frac{2c}{k} \max_{x \in [x_2^s, x_2^e]} u_2(x) < 1\ .$$

It follows that for any $k \ge 3$, d_k is the unique solution for $c > 1$ of $\lambda(k,c) = 1$. Despite the elementary structure of $\mathcal{S}$, its exact solutions, which would allow us to determine d_k for all k, are very complicated due to the initial conditions of the system. For example, u_{k-2} can be written as

$$u_{k-2}(x) = 1 - x + (k-1)e^{-cx} - \left(\frac{k}{c(k-1)} + ke^{-cx_{k-2}^s}\right) e^{-\frac{c(k-1)}{k}(x - x_{k-2}^s)}\ .$$

It is crucial that x_{k-2}^s cannot be "eliminated" from the expression for u_{k-2}. Moreover, x_{k-2}^s has a closed form only in terms of the $\mathbf{W}$ function. For u_{k-3} we have to determine u_{k-3}^s, i.e., the solution of $u_{k-2}(x) = \frac{k}{c(k-2)}$. This is already very difficult. Thus, although we can write down analytic expressions for the exact solutions of $\mathcal{S}$, they involve solving increasingly harder exponential equations and the solution for u_2 which is of interest, involves solving $k-3$ such equations. It is not clear that such a formula would yield significant insight to the asymptotic behavior of k-GL.

5 Asymptotics for k

In the following we prove theorem 2 by analyzing two systems of equations which "approximate" $\mathcal{S}$.

5.1 Almost sure success

Theorem 2.a *For all $k \ge 3$, if $c \le k \ln k - 3/2k$ then k-GL a.s. succeeds.*

By (a) of lemma 2, if $\mathrm{In}_i(t) < 1 - \epsilon$, for all $0 \le t \le n$ then $\bigcup_{t \ge 0} U_{i-2}(t) = \emptyset$ a.s. Thus, if for all $0 \le t \le n$, $|U_3(t)|$ is such that $\mathrm{In}_2(t) \le \frac{3c}{kn}|U_3(t)| < 1 - \epsilon$, for some $\epsilon > 0$ then k-GL a.s. succeeds. Consequently, U_3 will be central in what follows. We will also need to show that initializing the lists to fewer than k colors cannot increase the probability that the algorithm succeeds. For this, we start with $L_v = \{1, \ldots, k\}$ for all v, and repeatedly argue as follows: fix an initialization of the lists, a vertex v and a color $\omega \in L_v$. By a straightforward monotonicity argument, the probability that the algorithm succeeds does not increase by removing ω from L_v. Finally, let $H_k = \sum_{i=1}^k 1/i$ denote the k-th Harmonic number and recall that $H_k > \ln k$.

To prove theorem 2.a we will prove that for $c < kH_k - 3/2k$, k-GL a.s. colors $G_{n,p=c/n}$, even when the lists of the vast majority of vertices contain much fewer than k colors initially. In particular, let $M = n/(H_k - 3/2)$ and assume that we initialize the lists of the vertices as follows:

1. Partition the vertices randomly into sets P_i, $i = 3, \ldots, k$, such that $|P_i|$ is $\lfloor M/i \rfloor$ or $\lceil M/i \rceil$.
2. Let $Q = |P_i|/\binom{k}{i}$. Partition P_i randomly into sets N_{ij}, $j = 1, \ldots, \binom{k}{i}$, such that $|N_{ij}|$ is $\lfloor Q \rfloor$ or $\lceil Q \rceil$, for all j.
3. For each vertex v: If $v \in N_{ij}$ then L_v is initialized to the j-th i-subset of $\{1, \ldots, k\}$.

For $c < k(H_k - 3/2)$ we will call this initialization of the S_i (and hence of the initial conditions for the differential equations) *harmonic loading.*

If we apply the algorithm after the above initialization, the vertex to be colored from $t = 0$ until $U_3(t) = \emptyset$, will be chosen from U_3. Since $|U_3(0)| = \Omega(n)$, there is $\epsilon > 0$, such that $|U_3(t)| > 0$ for $0 \le t \le \epsilon n$ and it is straightforward to prove that claim 1 holds here as well (note that for each i all $|S_i^\ell(0)|$ differ by at most one initially). Thus, for all $3 \le i \le k-1$, the differential equations for $|U_i|$ derived in section 4 are valid at $t = 0$ and hence only the initial conditions of $\mathcal{S}$ need to be modified. Finally, since the solutions of the differential equations with the new initial conditions correspond to an execution of the algorithm, we get that for all $3 \le i \le k$, $u_i(x) \le u_{i+1}(x) + \mathrm{o}(1)$, implying that $x_i^e \le x_{i+1}^e$ for all such i. In conclusion, $u_k(x) = 1 - x$, while for $3 \le i \le k-1$,

$$\begin{aligned} u_i(0) &= \frac{1}{n}\sum_{j=3}^{i} \frac{n}{j(H_k - 3/2)} = \frac{H_i - 3/2}{H_k - 3/2} \quad \text{and} \\ u_i'(x) &= c\frac{i+1}{k}\left(u_{i+1}(x) - u_i(x)\right) - 1 \ . \end{aligned} \tag{5}$$

Now note that under harmonic loading there exists some $\epsilon = \epsilon(c) > 0$, such that for all $2 \le i \le k-1$,

$$\begin{aligned} \mathrm{In}_i(0) &= c\frac{i+1}{k}(u_{i+1}(0) - u_i(0)) + \mathrm{o}(1) \\ &= \frac{c}{k(H_k - 3/2)} + \mathrm{o}(1) \\ &< 1 - \epsilon \ . \end{aligned} \tag{6}$$

The following lemma is central for the u_i under harmonic loading.

Lemma 6 *There is $\zeta = \zeta(c) > 0$, such that for $3 \le i \le k$, $u_i'(x) < -\zeta$, for $x \in [0, x_i^e]$.*

Proof To simplify notation, let $\gamma = c\frac{i+1}{k}$ and take $\zeta = \exp(-\gamma)\epsilon$, where ϵ is as in (6). The proof is by "backwards induction".
Base case: Trivial, since $u_k'(x) = -1$.
Inductive step: Assume that the lemma is true for u_{i+1}, $4 \le i+1 \le k$ and recall that $x_{i+1}^e \ge x_i^e$. Solving (5) for u_i, with the initial condition, and taking the derivative of the result we get that for all $x \in [0, x_i^e]$

$$\begin{aligned} u_i'(x) &= -\exp(-\gamma x)\cdot(\gamma^2 \int_0^x \exp(\gamma s) u_{i+1}(s) ds \\ &\quad - \gamma \exp(\gamma x) u_{i+1}(x) + \gamma u_i(0) + 1) \\ &\le -\exp(-\gamma x)\cdot(\gamma^2 u_{i+1}(x) \int_0^x \exp(\gamma s) ds \\ &\quad - \gamma \exp(\gamma x) u_{i+1}(x) + \gamma u_i(0) + 1) \\ &= -\exp(-\gamma x)\left(-\gamma u_{i+1}(x) + \gamma u_i(0) + 1\right) \\ &\le -\exp(-\gamma x)\left(-\gamma u_{i+1}(0) + \gamma u_i(0) + 1\right) \\ &= -\exp(-\gamma x)(-\mathrm{In}_i(0) + \mathrm{o}(1) + 1) \\ &< -\exp(-\gamma)\epsilon \ . \end{aligned}$$

□

Proof of Theorem 2.a Since $|U_2(0)| = 0$, by (a) of lemma 2, it suffices to show that for $0 \le t \le n$, $\mathrm{In}_2(t) < 1 - \delta$, for some $\delta > 0$. Take $\delta = \epsilon$, where ϵ is as in (6), and recall that $\mathrm{In}_i(t) = \frac{c(i+1)}{nk}(|U_{i+1}(t)| - |U_i(t)|)$ and that for $0 \le t \le x_i^e n$, $|U_i(t)| = u_i(t/n)n + o(n)$. Since $\mathrm{In}_2(0) = \frac{3c}{k}u_3(0) + \mathrm{o}(1) < 1 - \epsilon$ and, by lemma 6, u_3 is decreasing in $[0, x_3^e]$ we have that $\mathrm{In}_2(t) < 1 - \epsilon$, for $0 \le t \le x_3^e n$. For $x_3^e n \le t \le n$, we will show that in fact $\mathrm{In}_2(t) = \mathrm{o}(1)$. For all $3 \le i \le k$, equation (5) along with lemma 6 imply that $\mathrm{In}_i(x_i^e n) \le c\frac{i+1}{k}u_{i+1}(x_i^e) + \mathrm{o}(1) < 1 - \zeta$, and by lemma 6, u_{i+1} is decreasing in $[x_i^e, x_{i+1}^e]$. Hence, by (b) of lemma 2, $|U_i(t)| = \mathrm{o}(n)$ for $x_i^e n \le t \le x_{i+1}^e n$. Since $x_k^e = 1$, invoking this fact for $i = 3, \ldots, k-1$ concludes the proof. □

5.2 Almost sure failure

Theorem 2.b *For any $\epsilon > 0$, there exists $k(\epsilon)$ such that for all $k \ge k(\epsilon)$, if $c \ge k \ln k(1+\epsilon)$ then k-GL a.s. fails.*

Similarly to the proof of theorem 2.a, we will examine the differential equations of $\mathcal{S}$ subject to initial conditions different from the ones in $\mathcal{S}$. In particular, we will provide initial conditions such that the resulting functions are substantially easier to deal with. Unlike the proof of theorem 2.a, though, these initial conditions do not correspond to an initialization of the lists of the vertices, and the resulting functions do not have an algorithmic interpretation. In spite of this fact, we will be able to prove that they provide a lower bound to the solutions of $\mathcal{S}$ for all $x \in [0,1]$. Using the lower bound for $|U_2(t)|$, we will derive a condition on c such that k-GL a.s. fails.

Recall that $u_i(x) = 0$, for $x \in [0, x_i^s]$ and, by lemma 5, $u_i(x) \le 0$, for $x \in [x_i^s, 1]$. For all $2 \le i \le k$, we will define a function l_i, similar to u_i, and prove that $l_i(x) \le u_i(x)$, for all $x \in [0,1]$. Specifically, $l_k(x) = 1 - x$, while for $2 \le i \le k$,

$$\begin{aligned} l_i(0) &= 0 \quad \text{and} \\ l_i'(x) &= c\frac{i+1}{k}\left(l_{i+1}(x) - l_i(x)\right) - 1 \ . \end{aligned}$$

Lemma 7 *For $2 \le i \le k$,*

$$l_i(x) \le u_i(x) \ ,$$

for all $x \in [0,1]$.

Proof The proof will be by "backwards induction".
Base case: Trivial, since $l_k(x) = u_k(x) = 1 - x$.
Inductive step: Assume that the lemma is true for l_{i+1}, u_{i+1}, $3 \le i+1 \le k$ and let $\gamma = c\frac{i+1}{k}$. Then for all $x \in [0,1]$

$$\begin{aligned} l_i(x) &= \exp(-\gamma x)\int_0^x \exp(\gamma s)\left(\gamma l_{i+1}(s) - 1\right) ds \\ &\le \exp(-\gamma x)\int_0^x \exp(\gamma s)\left(\gamma u_{i+1}(s) - 1\right) ds \\ &= \exp(-\gamma x)(\int_0^{x_i^s} \exp(\gamma s)\left(\gamma u_{i+1}(s) - 1\right) ds \\ &\quad + \int_{x_i^s}^x \exp(\gamma s)\left(\gamma u_{i+1}(s) - 1\right) ds) \\ &\le u_i(x) \ . \end{aligned}$$

The last inequality is derived by taking cases: if $x \in [0, x_i^s]$ we note that $u_i(x) = 0$, while for all $s \in [0, x]$, $\gamma u_{i+1}(s) \leq 1$. Hence, we keep both summands and get $l_i(x) \leq 0 = u_i(x)$. If $x \in [x_i^s, 1]$ we use that for all $s \in [0, x_i^s]$, $\gamma u_{i+1}(s) \leq 1$. This allows us to drop the (negative) first summand and obtain $u_i(x)$. □

Lemma 8

$$l_2(x) > (1 - \exp(-cx/k))^k - x \ .$$

Proof Since all differential equations in the system for the l_i "start" at $x = 0$, solving the system is straightforward. Let $S_i(x)$ denote

$$\sum_{j=0}^{k-i-1} (-1)^{k-i+j} \binom{k}{j} \binom{k-j-1}{i} \exp(c\frac{j-k}{k}x) \ .$$

As the system is triangular, one can verify that

$$l_i(x) = 1 - x + S_i(x) \tag{7}$$

by induction. Now, $S_i(x)$ has a "closed" form with roughly $\min\{i, k-i\}$ terms (for $k > 0$). Fortunately, we are interested in l_2. Letting $z = \exp(cx/k)$ we have

$$\begin{aligned} l_2(x) &= 1 - x + (-z)^{-k} \sum_{j=0}^{k-3} \binom{k}{j} \binom{k-j-1}{2} (-z)^j \\ &= -x \\ &\quad + z^{-k}(z-1)^{k-2} \left(z^2 + (k-2)z + \binom{k-1}{2} \right) \\ &> \left(1 - \frac{1}{z}\right)^k - x \ . \end{aligned}$$

□

Proof of Theorem 2.b Fix $\epsilon > 0$ and assume that $k \geq k(\epsilon)$ to be defined below. We will show that there exist $x_b, \zeta > 0$, both independent of k, such that for $x_b n \leq t \leq (x_b + \zeta)n$, $\mathrm{In}_1(t) > 1 + \delta$, for some $\delta > 0$. The theorem then follows from (c) of lemma 2. Recall that $c \geq (1+\epsilon)k \ln k$ for some arbitrary $\epsilon > 0$, so take $c = (1+\epsilon)k \ln k$. Let $\rho = 1+\epsilon$ and $f(x) = (1 - \exp(-cx/k))^k - x$. Note that for $x > 1/\rho$, $\lim_{k\to\infty} f(x) = 1 - x$. We will take

$$x_b = \frac{2+\rho}{3\rho} > \frac{1}{\rho} \ , \ \zeta = \frac{\rho - 1}{6\rho} > 0 \ , \text{ so that } x_b + \zeta < 1.$$

Since $1/\rho < x_b < x_b + \zeta$, using lemma 8, we can guarantee that for all $x \in [x_b, x_b + \zeta]$, for $k \geq k_1(\rho)$

$$l_2(x) > \frac{1}{2} \lim_{k\to\infty} f(x_b + \zeta) = \frac{\rho - 1}{4\rho} \ . \tag{8}$$

Since for all $x \in [x_2^e, 1]$, $u_2(x) \leq 0$, and for all $x \in [0, 1]$, by lemma 7, $u_2(x) \geq l_2(x)$, we get $x_b < x_b + \zeta < x_2^e$. Thus, for $x_b n \leq t \leq (x_b + \zeta)n$, we have $|U_2(t)| = u_2(t/n)n + \mathrm{o}(n) \geq l_2(t/n)n + \mathrm{o}(n)$, as desired. Let $\delta = \frac{\rho-1}{8\rho}$. If there is some t for which $|U_1(t)| \geq \delta n$ then k-GL a.s. fails, as we argued in the proof of (c) of lemma 2. Otherwise, for all $x_b n \leq t \leq (x_b + \zeta)n$

$$\begin{aligned} \mathrm{In}_1(t) &= \frac{2c}{kn} (|U_2(t)| - |U_1(t)|) \\ &> \frac{2c}{kn} (|U_2(t)| - \delta n) \\ &\geq 2\rho \ln k (l_2(t/n) - \delta) + \mathrm{o}(1) \ . \end{aligned} \tag{9}$$

By (8), (9) implies

$$\mathrm{In}_1(t) > 2\rho \ln k (\frac{\rho - 1}{4\rho} - \delta) = \frac{\rho - 1}{4} \ln k > 1 + \delta \ ,$$

for $k \geq k_2(\rho)$. Recall that $\rho = 1 + \epsilon$ and take $k(\epsilon) = \max\{k_1(\rho), k_2(\rho)\}$. □

Acknowledgments

We thank Alan Frieze for useful discussions.

References

[1] D. Achlioptas, E. Friedgut, *A threshold for random k-colorability*, manuscript in preparation.

[2] D. Achlioptas, L. Kirousis, E. Kranakis, D. Krizanc, *Rigorous results for* $(2+p)$*-SAT*, to appear in RALCOM 97. Available at URL http://www.cs.toronto.edu/~optas/papers/2psat.ps

[3] B. Bollobás, *Random Graphs*, Academic Press, London, (1985).

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

[5] B. Bollobás, A. Thomason, *Random graphs of small order*, Random Graphs '83. Annals of Discrete Mathematics, **28** (1985), 47–97.

[6] B. Bollobás, A. Thomason, *Threshold Functions*, Combinatorica, **7** (1986), 35–38.

[7] J. Bourgain, J. Kahn, G. Kalai, Y. Katznelson, N. Linial, *The influence of variables in product spaces*, Israel Journal of Mathematics, **77** (1992), 55–64.

[8] M.-T. Chao, J. Franco, *Probabilistic analysis of two heuristics for the 3-satisfiability problem*, SIAM Journal on Computing, **15** (1986), 1106–1118.

[9] M.-T. Chao, J. Franco, *Probabilistic analysis of a generalization of the unit-clause literal selection heuristic for the k-satisfiability problem*, Information Science, **51** (1990), 289–314.

[10] V. Chvátal, *Almost all graphs with 1.44n edges are 3-Colorable*, Random Structures and Algorithms, **2** (1991), 11–28.

[11] R.M. Corless, G.H. Gonnet, D.E. Hare, D.J. Jeffrey, D.E. Knuth, *On the Lambert W function*, Advances in Computational Mathematics, **5** (1996), 329–359.

[12] P. Erdös, A. Rényi, *On the evolution of random graphs*, Publication of the Mathematical Institute of the Hungarian Academy of Sciences, **5** (1960), 17–61.

[13] W. Fernandez de la Vega, *On the chromatic number of sparse random graphs*, Graph Theory and Combinatorics, B. Bollobás, Ed., Academic Press, New York, (1984), 321–328.

[14] E. Friedgut, *Necessary and sufficient conditions for sharp thresholds of graph properties and the k-SAT problem*, preprint, (1997).

[15] E. Friedgut, G. Kalai, *Every monotone graph property has a sharp threshold*, Proceedings of the American Mathematical Society, **124** (1996), 2993–3002.

[16] A. Frieze, C.J.H. McDiarmid, *Algorithmic theory of random graphs*, Random Structures and Algorithms, **10** (1997), 5–42.

[17] A. Frieze, S. Suen, *Analysis of two simple heuristics on a random instance of k-SAT*, Journal of Algorithms, **20** (1996), 312–355.

[18] G.R. Grimmett, C.J.H. McDiarmid, *On coloring random graphs*, Proceedings of the Cambridge Philosophical Society, **77** (1975) 313–324.

[19] S. Janson, D. Knuth, T. Łuczak B. Pittel, *The birth of the giant component*, Random Structures and Algorithms, **4** (1993) 231–358.

[20] J. Kahn, G. Kalai, N. Linial, *The influence of variables on Boolean functions*, Proceedings of the 29th Symposium on Foundations of Computer Science, (1988), 68–80.

[21] R. M. Karp, M. Sipser, *Maximum matchings in sparse random graphs*, Proceedings of the 22nd Symposium on Foundations of Computer Science, (1981) 364–375.

[22] L. M. Kirousis, E. Kranakis, D. Krizanc, *Approximating the unsatisfiability threshold of random formulas*, Proceedings of the 4th European Symposium on Algorithms, (1992) 27–38.

[23] D. E. Knuth, R. Motwani, B. Pittel, *Stable husbands*, Random Structures Algorithms, **1** (1990), 1–14.

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

[25] D.W. Matula, *Expose-and-merge exploration and the chromatic number of a random graph*, Combinatorica, **9** (1987), 275–284.

[26] M. Molloy, *The chromatic number of sparse random graphs*, Masters Thesis, University of Waterloo, (1992).

[27] M. Molloy, *A gap between the appearances of a k-core and a $(k+1)$-chromatic graph*, Random Structures and Algorithms, **8** (1996), 159–160.

[28] M. Molloy, B. Reed, *The dominating number of a random cubic graph*, Random Structures and Algorithms, **7** (1995), 209–221.

[29] B. Pittel, R.S. Weishar *Online coloring of sparse random graphs and random trees*, preprint, (1996)

[30] B. Pittel, J. Spencer, N. Wormald, *Sudden emergence of a giant k-core in a random graph*, Journal of Combinatorial Theory Series B, **67** (1996), 111–151

[31] L. Russo, *An approximate zero-one law*, Z. Wahrsch. Verw. Gebiete, **61** (1982), 129–139.

[32] E. Shamir, E. Upfal, *Sequential and distributed graph coloring algorithms with performance analysis in random graph spaces*, Journal of Algorithms, **5** (1984), 488-501.

[33] M. Talagrand, *On Russo's approximate zero-one law*, Annals of Probability, **22** (1994), 1576–1587.

[34] N. C. Wormald, *Differential equations for random processes and random graphs*, Annals of Applied Probability, **5** (1995), 1217–1235