Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE @ DIRECT®

Journal of Computer and System Sciences 67 (2003) 441–471

JOURNAL OF COMPUTER AND SYSTEM SCIENCES

http://www.elsevier.com/locate/jcss

Almost all graphs with average degree 4 are 3-colorable

Dimitris Achlioptas[a] and Cristopher Moore[b,c,*,1]

[a] Microsoft Research, Redmond, Washington, USA
[b] Computer Science Department, University of New Mexico, Farris Engineering Center, Albuquerque, NM 87131, USA
[c] Santa Fe Institute, Santa Fe, NM, USA

Received 14 July 2002; revised 28 December 2002

Abstract

We analyze a randomized version of the Brelaz heuristic on sparse random graphs. We prove that almost all graphs with average degree $d \leqslant 4.03$, i.e., $G(n, p = d/n)$, are 3-colorable and that a constant fraction of all 4-regular graphs are 3-colorable.

Keywords: Random graphs; Regular random graphs; Graph coloring; Threshold phenomena

1. Introduction

Let $G(n, p)$ be a random graph on n vertices where each edge appears independently of all others with probability p. The study of such graphs was pioneered in the seminal paper [10] of Erdős and Rényi where it was established that a number of monotone properties exhibit "sharp threshold" behavior. Let us say that a sequence of events $\mathscr{E}_n$ holds *with high probability* (w.h.p.) if $\lim_{n\to\infty} \Pr[\mathscr{E}_n] = 1$ and let us say that a property holds "for almost all graphs with average degree d" if it holds w.h.p. in $G(n, p = d/n)$.

Let $d_k = \sup\{d : G(n, d/n) \text{ is } k\text{-colorable w.h.p.}\}$. Determining d_3 was posed as an open problem by Erdős and Rényi in [10]. It remains open to date. Moreover, it is widely conjectured that for $d > d_k$, w.h.p. $G(n, d/n)$ is not k-colorable, i.e., that k-colorability has a sharp threshold for all $k \geqslant 3$. Locating the k-colorability threshold is a central open problem of random graph

*Corresponding author.
E-mail addresses: optas@microsoft.com (D. Achlioptas), moore@cs.unm.edu (C. Moore).
[1] Supported by NSF Grant PHY-0071139, the Sandia University Research Program, and Los Alamos National Laboratory.

0022-0000/$ - see front matter
doi:10.1016/S0022-0000(03)00120-X

theory and, in fact, this problem has attracted a fair amount of attention in other fields. For example, sophisticated computational experiments and non-rigorous arguments based on techniques from statistical physics [6,22] suggest $d_3 \approx 4.69$.

Upper bounds on d_k come from probabilistic counting arguments. Indeed, for $k = 3$ the best such bound is due to Kaporis et al. [11] who proved $d_3 \leqslant 4.99$. Lower bounds, on the other hand, have been established by constructive, i.e., algorithmic, arguments. In particular, the observation that a graph is k-colorable if it has no subgraph of minimum degree k, i.e., a k-*core*, allowed the derivation of all early lower bounds on d_3. Specifically, Łuczak [17] proved that $G(n, d/n)$ w.h.p. has no 3-core for $d = 1.0001$, i.e., after the giant component has emerged; shortly afterwards, Chvátal [9] improved the bound for the existence of a 3-core greatly to $d = 2.88$. Finally, in [23], Pittel, Spencer, and Wormald determined exactly the threshold for the emergence of a k-core for all $k \geqslant 3$, establishing $d_3 \geqslant 3.35\ldots$.

In [3], Achlioptas and Friedgut established that k-colorability has a sharp threshold, albeit a non-uniform one. That is, they proved that for all $k \geqslant 3$, there exists a function $d_k(n)$ such that for every $\varepsilon > 0$, if $np = (1 - \varepsilon)d_k(n)$ then w.h.p. $G(n, p)$ is k-colorable, but if $np = (1 + \varepsilon)d_k(n)$ then w.h.p. $G(n, p)$ is non-k-colorable. Let us say that a sequence of events $\mathscr{E}_n$ holds *with uniformly positive probability* (w.u.p.p.) if $\liminf_{n\to\infty} \Pr[\mathscr{E}_n] > 0$. This sharp threshold has the following, very useful, immediate corollary.

Corollary 1 (Achlioptas and Friedgut [3]). *If $G(n, d^*/n)$ is k-colorable w.u.p.p. then $d_k \geqslant d^*$.*

Achlioptas and Molloy [4] were the first to go beyond the 3-core lower bound for d_3 by analyzing a greedy list-coloring algorithm called 3-GL. Specifically, 3-GL maintains a list of available colors for each vertex. Initially, all vertices have the same 3 available colors. The algorithm attempts to color a graph by repeatedly (i) picking a random vertex v among those with fewest available colors, (ii) assigning v a random color c from its list, and (iii) removing color c from the lists of v's neighbors. In [4] it was shown that if $d < 3.847\ldots$, then 3-GL colors $G(n, d/n)$ w.u.p.p. implying $d_3 \geqslant 3.847$ via Corollary 1. That was the best known lower bound for d_3 prior to this work.

Here we consider a generalization of 3-GL in which ties between vertices with the same number of available colors are not broken uniformly at random, but rather by taking into account the number of remaining uncolored neighbors of each vertex. In particular, vertices with many uncolored neighbors are picked with higher probability, thus giving a probabilistic version of the Brelaz heuristic [8]. By analyzing our heuristic on $G(n, d/n)$ we prove the following.

Theorem 1. *$G(n, 4.03/n)$ is 3-colorable w.u.p.p.*

Invoking Corollary 1 we get

Corollary 2. *Almost all graphs with average degree* 4.03 *are* 3*-colorable.*

In proving Theorem 1 we actually derive a complete analysis of our heuristic on random graphs of bounded degree with an arbitrary degree sequence. This enables us to establish, for example, the first non-trivial result for the 3-colorability of random regular graphs. More precisely, for the set of all 4-regular graphs on n vertices with the uniform measure we prove

Theorem 2. *A random* 4*-regular graph is* 3*-colorable w.u.p.p.*

1.1. Organization of the paper

In the next section, we present our algorithm along with related work and motivation. In Section 3 we give an overview of the configuration model of random graphs and show how the analysis of our algorithm can be carried out in that model. In Section 4 we relate the configuration model to Theorems 1 and 2 and show how to separate out the tiny fraction of unbounded-degree vertices present in $G(n, d/n)$, leaving a list-coloring problem on a random graph of bounded degree. (We deal with these vertices separately in Section 12 as they are not central to our analysis.) In Section 5 we give a proof outline for the analysis of our algorithm on sparse random graphs with a fixed degree sequence and bounded degree. In Section 6 we introduce multitype branching processes and establish that they have certain crucial variational properties.

The main part of the analysis is given in Sections 7–9 where we use multitype branching process to apply the technique of differential equations. In particular, we derive a finite system of differential equations, parameterized by our preference function for higher degree vertices, that approximates the mean path of the random process corresponding to the uncolored vertices. In Section 10 we integrate these differential equations, with a strong preference for vertices of high degree, yielding the claimed results. Finally, in Section 11 we show how our system of differential equations parallels the one in [4] when the preference function is independent of the degrees of the vertices.

2. Our algorithm, related work, and motivation

Our results are based on analyzing the performance of the following algorithm, called `A`, which proceeds by maintaining a list of available colors $\ell(v)$ for each uncolored vertex v. In each step, some uncolored vertex w is chosen and permanently assigned a random color from $\ell(w)$; *there is no backtracking*. Initially, all lists contain the same set of three colors $\{R, G, B\}$. At any moment, for each uncolored vertex v, its list $\ell(v)$ consists of the colors originally available to v minus the colors assigned to its colored neighbors. Thus, the algorithm fails if we ever have $\ell(v) = \emptyset$ for some uncolored vertex. We will say that v is a "q-color vertex" if $|\ell(v)| = q$. Throughout the paper, unless specified otherwise, the degree of a vertex v, denoted by $\deg(v)$, will equal the number of its *uncolored* neighbors. The function h, determining the algorithm's preference for 2-color vertices with a given degree, will be specified later.

Remark. Note that in our implementation of `A`, if a 0-color vertex is ever generated, then `A` goes on forever (since such vertices remain uncolored). This impractical choice has the benefit of allowing us to analyze each iteration of the `while` loop without having to condition on `A` not having failed already. Note also that `A` colors only one 3-color vertex per connected component of the input graph (hence its lack of care in selecting such vertices).

Algorithm `A`

```
while there are uncolored vertices
   if there are 2-color vertices
      then
         pick a 2-color vertex v with probability proportional to h(deg(v));
         color(v)
      else
         pick a 3-color vertex v uniformly at random;
         color(v)
   while there are 1-color vertices
         pick a 1-color vertex v uniformly at random;
         color(v)
```

```
procedure color(v)
      pick c ∈ ℓ(v) uniformly at random;
      for all w adjacent to v set ℓ(w) ← ℓ(w) − c;
```

The list-coloring algorithm 3-GL considered in [4] is equivalent to `A` with h being the constant function. That is, when 2-color vertices exist 3-GL chooses from among such vertices uniformly at random. This method of choosing 2-color vertices has a property that greatly simplifies the analysis: if the original graph is distributed as $G(n,p)$, then after t vertices have been colored the graph induced by the uncolored vertices is distributed as $G(n-t,p)$. As a result, one can model the evolution of the graph induced by the uncolored vertices with a Markov process that has a very compact state representation; namely, the number of vertices having each of the 8 possible color lists. At the same time, it is intuitively clear that choosing 2-color vertices uniformly at random is not ideal. For example, it is very natural to give priority to the 2-color vertices that have high degree in the graph induced by the uncolored vertices. Such vertices, prima facie, are more constrained, so it makes sense to color them before we attempt to color less constrained vertices. Indeed, the Brelaz heuristic is exactly this idea taken to the extreme: $h(i) = 1$ if i is the maximum degree, and $h(i) = 0$ otherwise.

Our original motivation in this work was to improve upon the bound of [4] for the 3-colorability of $G(n, d/n)$ by analyzing algorithms that give priority to high-degree vertices. Clearly, for any such algorithm, the graph induced by the uncolored vertices will stop being $G(n,p)$ very quickly and our model for the graph induced by the uncolored vertices should, at a minimum, capture information about the different degrees. As we will see, `A` has the property that if the input graph is uniformly random *conditional on its degree sequence*, then the same is true for the graph induced by the uncolored vertices, for any choice of h. As a result, it turns out to be sufficient to refine our earlier notion of state so that it reflects how many vertices of degree i have each of the 8 possible color lists for every i. This blowup in the state representation is precisely what enables us to analyze a more sophisticated heuristic on $G(n,p)$. At the same time, since the input graph now need only be random conditional on its degree sequence, rather than be $G(n,p)$, we get a uniform analysis for a much larger class of graphs, e.g. random regular graphs. So, in the end we get to analyze a better heuristic in a more general setting.

To prove Theorems 1 and 2 there are two main technical challenges we need to overcome. The first challenge, as in [4], is establishing "list-stability", i.e., that each of the three different 2-color lists appears on a roughly equal number of 2-color vertices throughout the algorithm's execution. Since we now have to distinguish these vertices according to their degree, dealing with this issue in a probabilistic manner as in [1,4], or via the lazy-server lemma as in [2], is particularly cumbersome. Instead, here we first show how to model the evolution of the algorithm's state using multitype branching processes. Then we establish, algebraically, a certain variational stability for subcritical such processes. From that stability we will infer that perturbing any component of the state by $o(n)$ has a vanishing effect on the algorithm's probabilistic behavior. As a result, list-stability will follow readily from the symmetry of certain differential equations we will associate with the state evolution.

The second challenge has to do with the preference function h. Assume for a moment that we chose h to be as in the Brelaz heuristic, i.e., placing all the probability mass on maximum-degree 2-color vertices. While this certainly favors high-degree vertices, it results in an unwieldy probabilistic process. This is because the maximum degree among 2-color vertices is a very volatile random variable, potentially changing each time a 3-color vertex becomes a 2-color vertex. Maintaining this maximum degree information as part of the state would require an extremely "microscopic" representation of the random process and a correspondingly cumbersome analysis. In particular, with such a state representation it is not possible to apply the technique of differential equations for approximating the state evolution.

To overcome this difficulty, we observe that we can think of the Brelaz heuristic as setting $h(i) = i^\alpha$ where $\alpha \to \infty$. Therefore, taking $h(i) = i^\alpha$ where $\alpha \gg 0$ but finite, gives a "soft" version of the rule. The crucial point is that in this soft version, the probability of selecting a 2-color vertex of a given degree is a *smooth* function of the degree sequence. That is, if the number of vertices of each degree is perturbed by $o(n)$, this probability changes only by $o(1)$. This is crucial as it allows us to use differential equations to analyze A's performance. Moreover, considering larger and larger values of α allows us to get a better and better approximation of the Brelaz heuristic while maintaining a tractable process. In fact, we will see that taking $\alpha = 13$ appears to be enough to come very close to the limiting performance.

We want to point out that, up to now, relatively few algorithms have been analyzed in the degree sequence setting. While selecting vertices of a certain degree is often extremely useful, doing so can complicate the analysis greatly. The smoothing idea we introduce here provides the benefits of such selection, at least in a "soft" form, while maintaining a tractable analysis. Indeed, we consider this idea the main conceptual contribution of the paper.

3. The random configuration model

A random graph with a *fixed degree sequence* is a graph chosen uniformly at random among all graphs with that degree sequence. So, for example, a random graph on a degree sequence where all vertices have degree r is a uniformly random r-regular graph. It is not hard to see that a random graph $G(n,p)$ is also uniformly random conditional on its degree sequence. For this, first observe that a random graph $G(n,p)$ is uniformly random conditional on its number of edges and then observe that a uniformly random graph with a given number of edges is uniformly random conditional on its degree sequence. (We deal with the fact that the degree sequence of $G(n,p)$ is itself a random variable in the next section.)

As it turns out, studying random graphs with a fixed degree sequence directly is cumbersome. A very useful device in that vein is the *random configuration* model introduced by Bender and Canfield [5] and refined by Bollobás [7] and Wormald [24]. Suppose we are given a list V of vertices and their degrees, such that $\sum_v \deg(v)$ is even. In steps 1 and 2 below, we generate a random configuration (matching) with this degree sequence; in step 3 we use the configuration to form a random multigraph on the given vertices.

(1) Form a set V' consisting of $\deg(v)$ copies of each vertex $v \in V$.
(2) Pick a uniformly random perfect matching E' on V'.
(3) For each matching pair in E', add an edge between the corresponding vertices in V.

Clearly, a multigraph formed in this way may contain self-loops or multiple edges. If, though, it turns out to be simple, then it is a uniformly random graph on the given degree sequence. In the rest of the paper we will frequently use this equivalence of random multigraphs and random configurations, switching to the more suitable perspective for the statement at hand. In particular, our results will follow by running `A` on random configurations and proving that w.u.p.p. the corresponding multigraph is simple and 3-colorable. To do this, we first observe that step (2) above can be performed by sequentially selecting pairs of yet unmatched vertices and matching them. Then, we modify procedure `color`, so that the sequential matching in step (2) is performed by `A` along with the coloring of the vertices. Specifically, after assigning color c to vertex $v \in V$, `A` sequentially does the following for each copy $v' \in V'$ of v:

(a) Select a random copy $u' \in V' - v'$. Let u be the vertex of u'.
(b) Add the edge $\{u', v'\}$ to E'.
(c) Remove c from $\ell(u)$.
(d) Remove v' and u' from V'.

Alternatively, one can think of `A` as "discovering" (rather than generating) the random configuration as it colors it. In particular, we will sometimes refer to step (b) above as "exposing" copies u', v'. By the principle of deferred decisions [13], if the input is a uniformly random multigraph with a given degree sequence, then the uncolored vertices always induce a multigraph that is uniformly random conditional on its degree sequence. The degree of a vertex v, $\deg(v)$, is thus its number of unexposed copies. Note that if an edge e in the configuration corresponds to a multiple edge or self-loop, the algorithm still proceeds as desired, i.e., e has no effect on `A`'s coloring of the graph. Naturally, if there are no self-loops or multiple edges incident to v, $\deg(v)$ equals v's original degree minus the number of its colored neighbors.

We will refer to a step of the algorithm in which a 2- or 3-color vertex is colored as a *free step*, and a step in which a 1-color vertex is colored as a *forced step*. We will call a single iteration of `A`'s `while` loop a *round*. Thus, a round consists of a single free step and an ensuing sequence of forced steps, so that no 1-color vertices remain at the end of a round. Also, for the purposes of the analysis, it will be convenient to process 0-color vertices in the following manner (rather than leaving them uncolored): as soon as a 0-color vertex v is created, we add a random color $c \in \{R, G, B\}$ to $\ell(v)$ and label v `bad`. Clearly, the algorithm now fails if a `bad` vertex is ever created, but this trick ensures that at the beginning of every round only 2- and 3-color vertices are present.

4. Preliminaries and notation

Theorem 2, regarding random 4-regular graphs, is implied by following lemma, which we will establish by analyzing A on random 4-regular configurations.

Lemma 1. *W.u.p.p. a random* 4*-regular multigraph is simple and* 3*-colorable.*

To prove Theorem 1, regarding $G(n,p)$, we will rely on the fact that the degree sequence of $G(n,d/n)$ is very tightly concentrated around its expectation. In particular, it is random graph theory folklore that for any constant d, w.h.p. $G(n,d/n)$ has $(e^{-d}d^i/i!)\cdot n + o(n^{2/3})$ vertices of degree $i \leqslant 2\log n/\log\log n$ and no vertices of higher degree. Let $\mathscr{D}$ be any degree sequence such that w.h.p. $\mathscr{D}$ dominates the degree sequence of $G(n,d/n)$ and let π be any monotone graph property. It is easy to see that if a random multigraph on $\mathscr{D}$ is simple and has π with probability τ, then $G(n,d/n)$ must have π with probability at least $\tau - o(1)$. Thus, Theorem 1 follows from the following lemma regarding random configurations on degree sequences that w.h.p. dominate the degree sequence of $G(n,4.03/n)$.

Lemma 2. *Let $\mathscr{D}^*$ be any degree sequence with $(e^{-d}d^i/i!)\cdot n + o(n^{2/3})$ vertices of degree $i \leqslant 2\log n/\log\log n$ and no vertices of higher degree, where $d = 4.0309$. W.u.p.p. a random multigraph on $\mathscr{D}^*$ is simple and* 3*-colorable.*

To prove Lemma 2 we will separate the vertices of high degree and a few of their neighbors, handle them by a separate argument, and invoke A to color the remaining bulk of the graph.

Definition 3. Let $\Delta_{\max} = 30$. A vertex v has *high degree* if its initial degree is greater than $\Delta_{\max}$ and *low degree* otherwise. Let

$$\phi = \phi(d) = \sum_{i>\Delta_{\max}} ie^{-d}d^i/i!.$$

Remark. All subsequent references to $\Delta_{\max}$ and ϕ refer to Definition 3. Numerically, for $d = 4.0309$, we have $\phi = 4.475\ldots \times 10^{-16}$.

We divide the graph into a (mostly) high-degree part K, a low-degree part B and an interface set L as follows. Note that L is contained in both K and B.

Definition 4. For a random configuration $\mathscr{C}$ let E_H be the set of edges in $\mathscr{C}$ incident to high-degree vertices. Let H be the multigraph induced by E_H. Let Y be the set of low-degree vertices that lie in cyclic components of H. Let E_Y be the set of edges in $\mathscr{C}$ incident to vertices in Y. Let K be the multigraph induced by $E_H \cup E_Y$. Let B be the graph induced by $\mathscr{C} - \{E_H \cup E_Y\}$. Let L be the set of all low-degree vertices in K which are not in Y.

Lemma 2 follows from the following two lemmata.

Lemma 3. *Let $\mathscr{D}^*$ be any degree sequence as in Lemma 2 and let $\mathscr{C}$ be a random configuration on $\mathscr{D}^*$. Then*:

(1) *W.u.p.p. the multigraph K is simple and can be 3-colored so that all vertices in L have monochromatic neighborhoods.*
(2) *W.h.p. B has $b_i n + o(n^{2/3})$ vertices of each degree $0 \leqslant i \leqslant \Delta_{\max}$, where*

$$b_i = \sum_{j=i}^{\Delta_{\max}} \binom{j}{i} \left(\frac{\phi}{d}\right)^{j-i} \left(1 - \frac{\phi}{d}\right)^{i} \frac{e^{-d} d^j}{j!}.$$

(3) *W.h.p. $|L| < \phi n$.*

Lemma 4. *Let $\mathscr{B}$ be any degree sequence with $b_i n + o(n^{2/3})$ vertices of degree $0 \leqslant i \leqslant \Delta_{\max}$ with b_i as in part 2 of Lemma 3, and no vertices of higher degree. Assign lists to the vertices of $\mathscr{B}$ arbitrarily such that at most ϕn vertices have 2 available colors and all others have all 3 colors. W.u.p.p. a random multigraph on $\mathscr{B}$ is simple and list-colorable.*

Proof of Lemma 2. Let $\mathscr{D}^*$ be any degree sequence as in Lemma 2 and let $\mathscr{C}$ be a random configuration on $\mathscr{D}^*$. Let $\mathscr{E}_1$ be the event that K is simple and can be 3-colored so that all vertices in L have monochromatic neighborhoods. Let $\mathscr{E}_2$ be the analogous event for B. Clearly, if both $\mathscr{E}_1$ and $\mathscr{E}_2$ hold, then the multigraph induced by $\mathscr{C}$ is simple and 3-colorable (recall that the vertices in L belong both to the graph K and the graph B). Note that the events $\mathscr{E}_1, \mathscr{E}_2$ depend on disjoint sets of edges (but are not independent since the set L is not defined a priori) and Lemma 3 asserts that $\mathscr{E}_1$ holds w.u.p.p.

Whenever $\mathscr{E}_1$ holds we first 3-color K so that all vertices in L have monochromatic neighborhoods. We then uncolor every vertex in L and assign it the 2-color list avoiding the color assigned to its neighbors in K. We assign all 3 available colors to the other vertices in B. By parts 2 and 3 of Lemma 3 w.h.p. this yields a degree sequence and a color-list assignment that satisfies the conditions of Lemma 4. Therefore, w.u.p.p. the graph induced by B is list-colorable and simple. Since the list-colorability of B implies that each vertex in L has a monochromatic neighborhood, we get that $\mathscr{E}_1 \wedge \mathscr{E}_2$ holds w.u.p.p. □

The proof of Lemma 3 is postponed until Section 12, as it becomes much easier with multitype branching processes at our disposal (to bound the component-size distribution in K). The general idea is that if a graph has bounded average degree, then for every $\delta > 0$ there exists a constant $\Delta = \Delta(\delta)$ such that fewer than δn vertices have degree greater than Δ. In particular, if we take Δ to be large enough, then the subgraph induced by these high-degree vertices is sufficiently sparse so that, in a random graph, w.h.p. it consists of trees and a few unicyclic components. Thus, with some care, we can isolate this part of the graph from the remainder at the cost of removing at most one color from the lists of a few low-degree vertices. We note that in the special case of $G(n,p)$ we could have actually avoided dealing explicitly with the high-degree vertices as follows: if we set $h(i) = 0$ for all $i > \Delta_{\max}$, then the number of vertices of each degree $i > \Delta_{\max}$ follows a truncated Poisson distribution. Therefore, we could capture the behavior of the high-degree vertices by only adding a few extra variables to our state. However, the proof we give of Lemma 3 is more general, allowing one to analyze any other random graph model in which the degree distribution has a well-behaved tail.

5. Proof outline (and some more definitions)

We will say "at time t" to refer to the moment just before the tth round starts. We will divide the uncolored vertices present at the beginning of each round of A according to their (current) color list and degree. For each $i \geqslant 0$ and each color $C \in \{R, G, B\}$ we will denote by $C_i(t)$ the number of 2-color vertices which at time t have degree i and *do not* contain color C in their list. By $W_i(t)$ we will denote the number of 3-color vertices of degree i at time t and we will let $U_i(t) = R_i(t) + G_i(t) + B_i(t)$ be the total number of 2-color vertices of degree i. Finally, we let $E(t)$ denote the total number of unexposed copies, we let $U(t) = \sum_i iU_i(t)$ denote the number of unexposed copies belonging to 2-color vertices, and we let $H(t) = \sum_i h(i)U_i(t)$ denote the sum of the preference function over all 2-color vertices. We will use the term *list sequence* to refer to a degree sequence where each vertex v has been assigned a list $\ell(v) \subseteq \{R, G, B\}$. We will sometimes drop the reference to t in our random variables when that does not lead to confusion.

In the above terms, Lemma 1 asserts the colorability of a list sequence in which all vertices have 3 colors and degree 4, while Lemma 4 asserts the colorability of a list sequence in which most vertices have 3 colors, a small fraction have 2, and the degree distribution is described by part 2 of Lemma 3. To prove Lemmata 1 and 4 we will show that if we run A on each of these list sequences, w.u.p.p. no bad vertices are generated. In fact, in each case, it will be technically convenient to only run A for an a priori determined number of rounds (rather than until completion) and argue that the remaining graph is easy to color. This allows us to avoid the rather hairy analysis of A's last few rounds and use a much simpler (and more general) argument instead. Also, for technical reasons, it is easier to prove things about A's performance when initially there are $\varepsilon n + o(n)$ 2-color vertices of each color list and degree, for some small $\varepsilon > 0$. So, in particular, rather than applying A to the list sequences described by Lemmata 1 and 4, we will apply it to list sequences that result by stripping one color from the lists of some 3-color vertices. It is for these (slightly more constrained) list sequences that we prove that running A for a predetermined number of rounds leaves a residual list sequence that can be very easily colored. To make all this more precise we introduce the following definitions.

Definition 5. A list sequence on n vertices is (δ, ε)-*easy* if it has maximum degree $\Delta = O(1)$, $\sum_i i(i-2)(U_i + W_i) < -\varepsilon n$, all lists have at least 2 colors, and for every $i \geqslant 2$ there are distinct color lists ℓ_1, ℓ_2 such that at least δn vertices of degree i have list ℓ_1 and at least δn vertices of degree i have list ℓ_2.

The reason we call such list sequences easy is Lemma 5 below.

Lemma 5. *For all* $\delta, \varepsilon > 0$, *a* (δ, ε)-easy *list sequence is simple and* 3-*colorable w.u.p.p.*

Proof. Let $\mathscr{D}$ be any degree sequence with maximum degree $\Delta = O(1)$, where the number of vertices of degree $0 \leqslant i \leqslant \Delta$ is $\lambda_i n + o(n)$. It is a standard result in random graph theory [7] that a random multigraph on such a degree sequence is simple with probability at least $\theta = \theta(\Delta) > 0$. Moreover, in [21] it was shown that if there exists $\varepsilon > 0$ such that $\sum_i i(i-2)\lambda_i < -\varepsilon$ then w.h.p. a random multigraph on $\mathscr{D}$ has (i) no multicyclic component, and (ii) no more than $R \log n$ cycles, where $R = R(\Delta)$. In fact, implicit in the proof of [21] is the stronger fact that the expected number of cycles in such a multigraph is bounded by a constant $B = B(\Delta)$. Therefore, by Markov's inequality, for any constant $\zeta > 0$ there exists a constant $K = K(\zeta)$ such that a random multigraph on $\mathscr{D}$ has fewer than K cycles with probability at least $1 - \zeta$. Taking $\zeta < \theta$ we see that there exists some constant $L = L(\Delta)$ such that w.u.p.p. the multigraph induced by a (δ, ε)-easy degree sequence is simple, has at most L cycles, and has no multicyclic components.

Now, to prove that a given (δ, ε)-easy list sequence is colorable w.u.p.p. we consider the following experiment. Given any list sequence, we first generate a random multigraph with the same degree sequence and then we assign lists to its vertices randomly, subject to assigning each list to the correct number of vertices of each degree, per the given list sequence. Clearly, this experiment is the same as selecting a random multigraph on the vertices of the original list sequence. Now, for a (δ, ε)-easy list sequence, by our discussion above, we see that at the end of the first part of the experiment w.u.p.p. we have a simple graph that contains no multicyclic components and has at most L cycles. Moreover, since $\delta > 0$, in the second part of the experiment, with constant probability every cycle receives at least two different color lists. Therefore, w.u.p.p. at the end of the experiment we are left with a simple graph consisting of trees and unicyclic components where all vertices have at least two available colors and where there are no cycles of vertices with the same 2-color list. It is easy to see that such graphs are list-colorable. □

Definition 6. For an initial list sequence $\mathscr{L}$ and an integer t, let $G(t)$ be the random multigraph induced by edges incident to the colored vertices at time t. Let $\mathscr{L}(t)$ be the list sequence of the uncolored vertices at time t.

Lemmata 1 and 4 follow from combining Lemma 5 with Lemmata 6 and 7.

Lemma 6. *Let* $\mathscr{L}^4$ *be any list sequence where for all* $C \in \{R, G, B\}$, $C_4(0) = 10^{-3} \cdot n + o(n)$, *and* $W_4(0) = n - U_4(0)$. *There exist* $\delta, \varepsilon, T > 0$ *such that w.u.p.p.* (i) $G(T)$ *is simple and contains no* bad *vertices, and* (ii) $\mathscr{L}(T)$ *is* (δ, ε)-*easy*.

Lemma 7. *Let* $\mathscr{L}^*$ *be any list sequence where for all* $0 \leqslant i \leqslant \Delta_{\max}$ *and all* $C \in \{R, G, B\}$, $C_i(0) = \phi n + o(n)$, *and* $W_i(0) = (e^{-d} d^i / i!)n - U_i(0)$, *for* $d = 4.0309$. *There exist* $\delta, \varepsilon, T > 0$ *such that w.u.p.p.* (i) $G(T)$ *is simple and contains no* bad *vertices, and* (ii) $\mathscr{L}(T)$ *is* (δ, ε)-*easy*.

To get a rough idea of how the first parts of Lemmata 6 and 7 will be proved, observe the following: if a configuration has maximum degree Δ and δn copies and a given round has z forced steps, then the probability that a bad vertex is generated during that round is roughly proportional to z^2/n. To see this note that each time we expose a single copy in that round there is at most a $(z \times \Delta)/(\delta n)$ chance that its match lies among the yet unmatched copies of 1-color vertices waiting to be colored in that round. Clearly, if such a match never occurs, then no bad vertex is created. Thus if (i) $\Omega(n)$ copies remain unexposed at the beginning of a round and (ii) the conditional expected squared length of the round is bounded, then A has a $O(1/n)$ probability of failure in that round. Since there are at most n rounds, if we can show that w.h.p. (i) and (ii) hold at the beginning of every round, then Markov's inequality implies that A succeeds w.u.p.p.

To bound the second moment of a round's length it will be useful to think of each round as similar to a branching process where the progenitor is the vertex chosen on the free step and the progeny of each vertex v is the set of 1-color vertices created by coloring v. Clearly, not all vertices being colored are of equal "potency" as their degree and assigned color affects the distribution of their progeny. To capture this fact we use multitype branching processes, where a type amounts to a $\langle$color, degree$\rangle$ pair. Using this viewpoint, the heart of the matter becomes showing that w.h.p. every such branching process, throughout A's execution, is subcritical. This is because subcriticality implies a geometric tail for the round's length, which is more than enough to bound the length's second moment. As we will see, given a list sequence we can write a matrix M (corresponding to the expected progenies of a multitype branching process) such that subcriticality follows if M's largest eigenvalue is bounded below 1. So, with this in mind, our plan is to track the evolution of the list sequence of the uncolored vertices and prove that it is such that throughout A's execution the branching process corresponding to a single round is subcritical.

To perform this tracking observe that, as mentioned above, if the list sequence at the beginning of a round induces a subcritical branching process then the length of the round has a geometric tail. Therefore, the state of our process, namely the number of 2- and 3-color vertices of each degree, evolves in a very smooth manner, i.e., each round has an $O(1)$ expected effect on quantities of size $\Theta(n)$. This allows us to approximate the evolution of the list sequence using the technique of differential equations. In particular, we will model the evolution of the list sequence with a deterministic trajectory in $\mathbb{R}^{4\times(\Delta+1)}$ and argue that as long as this trajectory stays in the region corresponding to subcriticality, w.h.p. our random process stays very close to this trajectory.

6. Multitype branching processes

In the standard Galton–Watson (GW) branching process we have a progenitor vertex which gives rise to X children, where X is some non-negative integer-valued random variable. Each of those children then procreates independently, its offspring distribution being the same as that of the progenitor, and so on. The fundamental theorem of branching processes asserts that if $\mathbf{E}[X] < 1$, i.e., the branching process is subcritical, then extinction is certain. A natural generalization of the Galton–Watson process is one in which there are b vertex "types", the type of a vertex determining the probability distribution of its progeny. More precisely, for each type $1 \leqslant j \leqslant b$ there is a probability distribution $f_j : \mathbb{N}^b \to \mathbb{R}$, giving the probability that a vertex of type j will have progeny $(x_1, \ldots, x_b)$, i.e., x_i children of each type $1 \leqslant i \leqslant b$. The type of the progenitor is also governed by a probability distribution, i.e., the progenitor is of type j with probability p_j. The evolution is similar to the GW process: the progenitor is chosen according to p_j; from then on, for each $1 \leqslant i \leqslant b$, every vertex of type i procreates independently according to f_i.

Just as for the Galton–Watson branching process, in order to determine whether extinction is certain in a multitype branching process, it suffices to consider expectations. Rather than the scalar $\mathbf{E}[X]$, however, the key here is the matrix M where $M_{i,j}$ is the expected type-i progeny of a type-j vertex. The criterion for subcriticality is whether the maximal eigenvalue λ_1 of M is smaller than 1 (see the book by Mode [20] for a general treatment). In that case the expected total progeny is bounded and, moreover, it can be read off from M. The following lemma establishes exactly how the total progeny of each type relates to M and also gives two crucial variational properties of the total progeny.

Lemma 8. *Consider a multitype branching process with b types and let M be the $b \times b$ matrix where $M_{i,j}$ is the expected type-i progeny of a type-j vertex. Let the vector $p = (p_1, \ldots, p_b)$ give the probability that the progenitor is of type j and let $m = m(p) = (m_1, \ldots, m_b)$ be the total expected progeny of each type.*

If the maximal eigenvalue λ_1 of M satisfies $\lambda_1 < 1 - \delta$ for some $\delta > 0$ then

(1) *For all p, $m(p) = (I - M)^{-1} p$ where I is the identity matrix.*
(2) *If there exists some $\theta > 0$ such that for every type i*

$$\Pr[\textit{A vertex of type } i \textit{ has more than } t \textit{ children}] < (1 - \theta)^t,$$

then there exists $\rho = \rho(\theta, \delta) > 0$ such that

$$\Pr[\textit{Total progeny} > s] < (1 - \rho)^s.$$

(3) *Let n(q) be the total expected progeny of a multitype branching process with b types, matrix N, and progenitor distribution q. Let $\|\cdot\|$ denote the 2-norm. Suppose that every entry of M and N is at most B, the largest eigenvalue of N is less than $1 - \delta$, $\|M - N\| = \varepsilon$ and $\|p - q\| = \zeta$. Then there exists a constant $L = L(b, \delta, B)$ such that $\|m(p) - n(q)\| < L \times (\varepsilon + \zeta)$.*

Proof. (1) The linearity of expectation makes it trivial to see that the expected population of type j after z generations is $M^z p$. Therefore, the total expected progeny is given by the geometric sum

$$m(p) = \sum_{z=0}^{\infty} M^z p.$$

Note now that the sum $\sum_{z=0}^{\infty} M^z$ converges to $(I - M)^{-1}$ iff all eigenvalues of M have modulus less than one. Since the entries of M are real and nonnegative, its maximal eigenvalue is real and positive, so this amounts to $\lambda_1 < 1$.

(2) We iteratively reveal the progeny of the branching process as follows. At a given moment, vertices are designated "open", "closed", or "unexamined". Initially, we designate the progenitor

open and all other vertices unexamined (we assume that the type of the progenitor has already been chosen). In each step, we examine one open vertex, close it, and open all its children if there are any. The process ends when there are no open vertices left, i.e., when we have examined (opened and closed) all the progeny. At step t we have a random vector $g(t) = (g_1(t), \ldots, g_b(t))$ giving the population of open vertices of each type. Without loss of generality, we assume that the branching process is irreducible. Thus, the left eigenvector $s = (s_1, \ldots, s_b)$ corresponding to λ_1 is unique and strictly positive. Let v_t be the inner product $s^T \cdot g(t)$. We will focus on the random variable

$$X_t = \exp\left[\alpha\left(\beta \sum_{j \leqslant t} v_j + v_t\right)\right],$$

where $\alpha, \beta > 0$ will be chosen later. We will prove that for all $t \geqslant 0$, we have $\mathbf{E}[X_{t+1} \mid X_0, \ldots, X_t] < X_t$ for all t, implying $\mathbf{E}[X_t] < X_0$ for all $t > 0$. Since X_0 is bounded (for any progenitor type), we get the desired result by applying Markov's inequality to X_t and using the fact that s is strictly positive. Now,

$$\begin{aligned}\frac{\mathbf{E}[X_{t+1} \mid X_0, \ldots, X_t]}{X_t} &= \mathbf{E}[\exp(\alpha((1+\beta)v_{t+1} - v_t)) \mid v_0, \ldots, v_t] \\ &\equiv \mathbf{E}[\exp(\alpha Z_{t+1}(\beta))].\end{aligned}$$

We claim that (i) $Z_{t+1}(\beta)$ has an exponential tail for all $\beta > 0$ and that (ii) we can choose β such that $\mathbf{E}[Z_{t+1}(\beta)] < 0$. By standard arguments this implies that there exists some $\alpha > 0$ such that $\mathbf{E}[\exp(\alpha Z_{t+1}(\beta))] < 1$, which in turn implies that for such α, β we have $\mathbf{E}[X_{t+1} \mid X_0, \ldots, X_t] < X_t$. Claim (i) follows readily from the fact that the progeny of each vertex type has exponential tails. For claim (ii) we observe that if $\beta < \delta/(1-\delta)$, then

$$\begin{aligned}\mathbf{E}[Z_{t+1}(\beta)] &= (1+\beta)\mathbf{E}[v_{t+1} \mid v_0, \ldots, v_t] - v_t \\ &= (1+\beta)(s^T M \cdot v_t) - v_t \\ &= ((1+\beta)\lambda_1 - 1)\, v_t \\ &< 0.\end{aligned}$$

(3) First, note that since M and N are of fixed size b, their largest eigenvalues are bounded by $1-\delta$, and their entries are bounded by B, there exists a constant $Q = Q(b, \delta, B)$ such that $||(I-M)^{-1}|| < Q$ and $||(I-N)^{-1}|| < Q$. Then the triangle inequality and the facts $||Ax|| \leqslant ||A||||x||$ and $||p|| \leqslant 1$ imply

$$\begin{aligned}||m(p) - n(q)|| &= ||(I-M)^{-1}p - (I-N)^{-1}q|| \\ &= ||((I-M)^{-1} - (I-N)^{-1})p + (I-N)^{-1}(p-q)|| \\ &\leqslant ||((I-M)^{-1} - (I-N)^{-1})p|| + ||(I-N)^{-1}(p-q)|| \\ &\leqslant ||(I-M)^{-1} - (I-N)^{-1}|| + Q\zeta. \qquad (1)\end{aligned}$$

To bound the matrix norm in (1) we write

$$\begin{aligned}(I-M)^{-1}-(I-N)^{-1} &= (I-M)^{-1}((I-N)-(I-M))(I-N)^{-1}\\ &= (I-M)^{-1}(M-N)(I-N)^{-1}\end{aligned}$$

implying

$$||(I-M)^{-1}-(I-N)^{-1}|| \leqslant ||(I-M)^{-1}||\;||M-N||\;||(I-N)^{-1}|| \leqslant Q^2\varepsilon.$$

Then setting $L = Q^2$ completes the proof. □

7. A single round as a multitype branching process

We are now ready to analyze what happens in a single round of A. Recall that at the beginning of each round, only 2- and 3-color vertices remain. We focus on the case where there are $\Omega(n)$ unexposed copies belonging to 2-color vertices. This certainly holds initially for the list sequences in Lemmata 6 and 7 and we will see that in each case, by our respective choices of T, it will also hold w.h.p. for all T rounds that we will run A.

Each round starts with a free step in which we pick some 2-color vertex v and assign it a color c from its list. We then expose the partners of the unexposed copies of v. Some of these partners belong to 3-color vertices, while others belong to 2-color vertices. In either case, we update the lists of those vertices by removing c from their lists. This might lead to the creation of some new 1-color vertices which we process just as we did v, and the round proceeds with these forced steps until no 1-color vertices remain. As (the coloring of) each vertex gives rise to new vertices to be colored, we would like to map the set of colored vertices in each round to a multitype branching process, with the types corresponding to $\langle \text{assigned color}, \text{degree} \rangle$ pairs.

There are two main issues complicating such a mapping. The first one is that the graph induced by the edges incident to vertices colored in a given round might contain cycles, self-loops or multiple edges. In particular, if any such blemish occurs we cannot quite equate the number of 1-color vertices generated when a vertex is colored with its number of children in the branching process. (On the other hand, we are hoping that such blemishes are rare since any one of them destroys the graph's simplicity or might lead to a 0-color vertex.) The other issue is that the total number of vertices having each $\langle \text{color list}, \text{degree} \rangle$ pair shifts in the course of a round, as vertices are colored or lose colors and/or neighbors. Therefore, we do not quite have a fixed probability distribution governing the progeny of each type throughout the course of each round. To overcome these difficulties we will rely heavily on the fact that in the subcritical regime, with overwhelming probability, a round exposes no more than a polylogarithmic number of copies. As a result, with some work, we will be able to prove that in this regime the above two issues only affect lower-order terms. Below we show how to associate a matrix to the list sequence at time t by ignoring both issues and thinking of a round as corresponding exactly to a multitype branching process. Lemma 9, stated below and proved in the next section, asserts that this approximation is indeed exact up to $o(1)$ terms.

We will associate round t with a multitype branching process whose matrix $M = (M_{(x,i),(y,j)})$ gives the expected number of 1-color vertices with color list $\{x\}$ and degree i generated by assigning color y to a vertex v of degree j, for all $x, y \in \{R, G, B\}$ and $0 \leqslant i, j \leqslant \Delta$, where Δ is the maximum degree of the list sequence. Recall that whenever we color a vertex v of degree j, its j copies are matched with partners chosen uniformly among all other unexposed copies. Since $E(t)$ denotes the number of unexposed copies at time t, the expected number of v's copies that "hit" any given 2-color vertex of degree $i+1$ is then $j(i+1)/E(t) + o(1)$. As a result, M's entries are

$$M_{(x,i),(y,j)} = \begin{cases} \dfrac{j(i+1)C_{i+1}(t)}{E(t)} & \text{if } x \neq y, \\ 0 & \text{if } x = y, \end{cases} \tag{2}$$

where $C_{i+1}(t)$ is the number of 2-color vertices with $\ell = \{R, G, B\} - C = \{x, y\}$ and degree $i+1$ at time t. This gives us a square matrix M of size $3(\Delta + 1)$, corresponding to a branching process with one type for each $\langle$color, degree$\rangle$ pair. Parts 2 and 2b of Lemma 9 below assert that in the subcritical regime, this multitype branching process is an excellent approximation of the behavior of a single round. More precisely:

Definition 7. A list sequence $\mathscr{L}(t)$ is (α, β)-*subcritical* if $U(t) > \alpha n$ and the largest eigenvalue $\lambda_1 = \lambda_1(M)$ of the matrix M defined in Eq. (2) satisfies $\lambda_1 < 1 - \beta$. A list sequence is *subcritical* if it is (α, β)-subcritical for some $\alpha, \beta > 0$.

In Lemma 9 below, part 1 is unrelated to branching processes and comes from a straightforward calculation of the expected effect of the free step. In part 2a we use the branching process to determine the expected number of copies exposed while coloring vertices of each color in the course of a round. Note that this number is the expected total progeny of vertices of a given color, summed over all degrees. The expression in 2a is exactly what we would get from the corresponding branching process along with an $o(1)$ term that absorbs the effects of any potential cycles, self-loops or multiple edges. In part 2b we determine the expected change in our list sequence in the course of a single round by distributing the newly exposed copies to the vertices present at time t. Analogously to part 2a, the expression we get (up to the $o(1)$ term) represents the expectation of these changes if there was no shift in the list sequence in the course of a round. Finally, part 2c follows from the corresponding fact about branching processes, namely part 2 of Lemma 8.

Lemma 9. *Let $\mathscr{L}(t)$ be any (α, β)-subcritical list sequence.*

(1) *The probability $p_{(c,i)}$ that the vertex colored in the free step of round t receives color $c \in \{R, G, B\}$ and has degree i is*

$$p_{(c,i)} = \frac{1}{2} \frac{h(i)(X_i(t) + Y_i(t))}{H(t)}, \tag{3}$$

where $\{X, Y\} = \{R, G, B\} - c$.

(2) *Let $p \in \mathbb{R}^{3 \times (\Delta+1)}$ be the vector with entries $p_{(c,i)}$ for $c \in \{R, G, B\}$ and $0 \leqslant i \leqslant \Delta$, and let M be the matrix given by Eq.* (2). *If $\mathscr{L}(t)$ is subcritical then*:

(a) *The expected number of copies exposed while coloring vertices with color* $c \in \{R, G, B\}$ *in round* t *is* $k_c(M) + o(1)$, *where*

$$k_c = k_c(M) = \sum_{i=0}^{\Delta} i \times ((I - M)^{-1} p)_{(c,i)}. \tag{4}$$

(b) *Let* $k = k_R + k_G + k_B$. *For all* $C \in \{R, G, B\}$ *and for all* $0 \leqslant i \leqslant \Delta$,

$$\mathbf{E}[W_i(t+1) - W_i(t)] = -k \frac{iW_i(t)}{E(t)} + o(1), \tag{5}$$

$$\mathbf{E}[C_i(t+1) - C_i(t)] = k_C \frac{(i+1)(W_{i+1}(t) + C_{i+1}(t))}{E(t)} - k \frac{iC_i(t)}{E(t)} - \frac{h(i)C_i(t)}{H(t)} + o(1). \tag{6}$$

(c) *There exists* $\rho > 0$ *such that*

$$\Pr[\textit{More than } s \textit{ copies are exposed in round } t] < (1-\rho)^s + o(n^{-1}).$$

In Section 9 we will use Lemma 9 above to establish that we can model the evolution of $\mathscr{L}(t)$ by a system of differential equations. In particular, using that system, we will be able to establish that if we apply A to certain initial list sequences of interest for T rounds, then w.h.p. $\mathscr{L}(t)$ is subcritical for all $0 \leqslant t \leqslant T$. Along with the following lemma, this implies that in each such case, w.u.p.p. there are no bad vertices at the end of round T.

Lemma 10. *Assume that for a list sequence* $\mathscr{L}$ *and integer* T *the following holds: if we apply* A *to* $\mathscr{L}$ *for* T *rounds then w.h.p.* $\mathscr{L}(t)$ *is subcritical for all* $0 \leqslant t \leqslant T$. *Then, if we apply* A *to* $\mathscr{L}$ *for* T *rounds, w.u.p.p.* $G(T)$ *is simple and contains no* bad *vertices.*

In the next section we prove Lemmata 9 and 10 above. We observe that the analysis suggested by them is tight (up to the constant implicit in the w.u.p.p. statement). In other words, one can prove that w.u.p.p. A does indeed fail on the list sequences we consider.

8. A single round as a multitype branching process: proofs

Proof of Lemma 9. To prove part 1, note that since $\mathscr{L}(t)$ is subcritical there is at least one 2-color vertex right before round t begins. Thus, in the free step of round t some 2-color vertex is chosen. Since a given 2-color vertex with degree i is chosen with probability $h(i)/H(t)$, Eq. (3) follows.

For part 2, let $\tau_C(t)$ denote the total number of copies exposed while coloring vertices with color $C \in \{R, G, B\}$ in round t. Let $Q = \lfloor \log^2 n \rfloor$. We will prove below that if $\mathscr{L}(t)$ is subcritical, then

$$\Pr[\tau_R(t) + \tau_G(t) + \tau_B(t) \geqslant Q] = o(n^{-1}). \tag{7}$$

Let P_t be the random process corresponding to the tth round of A. With (7) in mind, let us define another random process S_t which is identical to P_t except that if S_t ever exposes Q copies in round t it stops, i.e., it does not expose any more copies (or color any other vertices) in round t. For each $C \in \{R, G, B\}$ let $v_C(t)$ be the analogue of $\tau_C(t)$ for S_t and observe that, by construction, $v(t) \equiv v_R(t) + v_G(t) + v_B(t) \leqslant Q$. We claim that

$$\mathbf{E}[W_i(t+1) - W_i(t)] = -\mathbf{E}[v(t)]\frac{iW_i(t)}{E(t)} + o(1), \tag{8}$$

$$\mathbf{E}[C_i(t+1) - C_i(t)] = \mathbf{E}[v_C(t)]\frac{(i+1)(W_{i+1}(t) + C_{i+1}(t))}{E(t)} - \mathbf{E}[v(t)]\frac{iC_i(t)}{E(t)} - \frac{h(i)C_i(t)}{H(t)} + o(1). \tag{9}$$

To prove our claim we first observe that, with the exception of the free step, the effect of P_t on the C_i and W_i is incurred by exposing the copies of the vertices colored in round t. Part 1 of the lemma asserts that the effect of the free step is given by the term $h(i)C_i(t)/H(t)$ in (9). We will prove that the effect of exposing copies in the course of S_t on W_i and C_i is given by the remaining terms in (8) and (9). Since at the beginning of every round there are $O(n)$ unexposed copies, (7) implies that the contribution of the cases where $S_t \neq P_t$ is $o(1)$.

To analyze the effect of exposing a single copy (in either P_t or S_t), it is convenient to think of the vertices of a given $\langle$color list, degree$\rangle$ pair as being grouped together in a "bucket". In particular, if a vertex v is assigned color c and one of its copies is matched with a copy of a neighbor w, this moves w from bucket $\langle \ell(w), \deg(w) \rangle$ to bucket $\langle \ell(w) - c, \deg(w) - 1 \rangle$. Specifically, any 3-color neighbor of degree i becomes a 2-color vertex of degree $i-1$, while a 2-color neighbor becomes a 1-color vertex if $c \in \ell(w)$, and stays a 2-color vertex if $c \notin \ell(w)$. In either case $\deg(w)$ becomes $\deg(w) - 1$. Thus, our claim is that (8) and (9) describe the expected "flows" into and out of these buckets.

As in P_t, every time a vertex is colored in S_t the partners of its copies are chosen uniformly among all unexposed copies at that time. Since S_t never exposes more than Q copies, this means that throughout its course, the number of unexposed copies E is at least $E(t) - Q$. Moreover, by the same token, throughout the course of S_t, we have $|W_i - W_i(t)| \leqslant Q$ and $|C_i - C_i(t)| \leqslant Q$ for every i and $C \in \{R, G, B\}$. Finally, recall that, since $\mathscr{L}(t)$ is subcritical, $E(t) = \Omega(n)$. Therefore, every time a copy is exposed in the course of S_t, the expected flow between any pair of buckets is within $O(\log^2 n)/n$ of what it was for the very first copy exposed in round t. Summing over all the copies exposed by S_t, by the linearity of expectation, we see that the total change in each of the W_i and C_i in S_t is given by (8) and (9) as claimed.

To conclude the proof of part 2 we will prove that for every $c \in \{R, G, B\}$,

$$\mathbf{E}[v_c] = \sum_{i=0}^{\Delta} i \times ((I - M)^{-1} p)_{(c,i)} + o(1), \tag{10}$$

where M is defined in (2) and p is specified by part 1 of the lemma.

We start by observing that every time a vertex v is colored in P_t, exposing its copies causes a number of vertices to become 1-color vertices and, thus, to be colored in subsequent steps. The same is true for S_t, with the possible exception that S_t stops because it has exposed Q copies (in which case some of the 1-color vertices generated might not get colored). Moreover, since the partner of each copy of v is picked uniformly among all unexposed copies, we can associate a probability distribution with the coloring of v, governing the number of 1-color vertices generated with each possible $\langle$color, degree$\rangle$ pair. Crucially, this distribution depends only on the list sequence just before v is colored. We claim that throughout the course of S_t, the expected number of 1-color vertices with color list $\{x\}$ and degree i generated by assigning color y to a vertex of degree j is

$$F_{(x,i),(y,j)} = \begin{cases} \dfrac{j(i+1)C_{i+1}(t)}{E(t)} + o(1) & \text{if } x \neq y, \\ 0 & \text{if } x = y, \end{cases}$$

where $C = \{R, G, B\} \backslash \{x, y\}$. To see this first observe that the very first copy exposed in round t "hits" a vertex with list $\{x, y\}$ and degree $i+1$ with probability exactly $(i+1)C_{i+1}(t)/E(t)$. Moreover, by arguing as we did earlier, since S_t never exposes more than Q copies this probability does not shift by more than $O(\log^2 n)/n$ throughout the course of S_t. Thus, our claim follows by summing the contribution over all degrees j and linearity of expectation.

Right before round t begins, let $f_{(c,i)} : \mathbb{N}^{3\times(\Delta+1)} \to \mathbb{R}$ be the probability distribution governing the number of 1-color vertices of each color and degree generated by assigning color c to a vertex of degree i, for $c \in \{R, G, B\}$ and $0 \leqslant i \leqslant \Delta$. Let us call each possible combination from $\mathbb{N}^{3\times(\Delta+1)}$ a "litter". From our discussion above regarding the list sequence shift during the course of S_t we see that we can readily construct a collection of probability distributions $f^{\mathrm{u}}_{(c,i)} : \mathbb{N}^{3\times\Delta} \to \mathbb{R}$ such that (i) each such distribution *dominates* the probability distribution governing the number of 1-color vertices generated when we color a vertex of degree i with color c in S_t, and (ii) each litter is assigned the same probability by $f^{\mathrm{u}}_{(c,i)}$ and $f_{(c,i)}$ within $o(1)$. Similarly, we can construct probability distributions $f^{\mathrm{l}}_{(c,i)} : \mathbb{N}^{3\times\Delta} \to \mathbb{R}$ such that (i) each such distribution *is dominated by* the probability distribution governing the number of 1-color vertices generated when we color a vertex of degree i with color c in S_t, and (ii) each litter is assigned the same probability by $f^{\mathrm{l}}_{(c,i)}$ and $f_{(c,i)}$ within $o(1)$.

Therefore, to bound the behavior of S_t we introduce two branching processes B^{u} and B^{l}, each with $3 \times (\Delta + 1)$ types, one for each (c, i) pair where $c \in \{R, G, B\}$ and $0 \leqslant i \leqslant \Delta$. In both, we let the probability that the progenitor is of a given type be the corresponding value for S_t. For each (c, i), the progeny of vertices of type (c, i) in B^{u} is determined by $f^{\mathrm{u}}_{(c,i)}$, while in B^{l} it is determined by $f^{\mathrm{l}}_{(c,i)}$. By construction, S_t can be coupled to B^{u} and B^{l} so that in every step, the number of 1-color vertices with each color and degree generated in S_t is dominated by the progeny of the corresponding vertex in B^{u}, while it dominates the progeny of the corresponding vertex in B^{l}. Moreover, observe that every entry of the matrix of expectations of B^{l} and B^{u} is within $o(1)$ of $F_{(x,i),(y,j)}$, which in turn is within $o(1)$ of the corresponding entry in the matrix M of Eq. (2). Recalling that $\lambda_1(M) < 1 - \delta$ and that M has finite size and bounded entries, by parts 1 and 3 of

Lemma 8, the expected number of vertices of degree i assigned color c in S_t is $((I - M)^{-1}p)_{(c,i)} + o(1)$. Summing over all degrees $0 \leqslant i \leqslant \Delta$ yields (10) as desired.

Finally, recall that we need to prove that the probability of S_t stopping because it exposes Q copies is $o(n^{-1})$. To prove this, we will establish that the probability the total progeny of B^{u} exceeds Q/Δ is $o(n^{-1})$. For that, observe that $\lambda_1(M) < 1 - \delta$ and that every entry in the matrix of expectations of B^{u} is within $o(1)$ of M. Therefore, the largest eigenvalue of B^{u} is also bounded below 1. As no type in B^{u} ever has more than Δ children, the desired claim follows from part 2 of Lemma 8 with room to spare. More generally, part 2 of Lemma 8 applied to B^{u} implies part 2c of Lemma 9. The $o(n^{-1})$ term covers the possibility that $P_t \neq S_t$. $\square$

Proof of Lemma 10. We remark that it is not obvious to us how to get a (stronger and more natural) lemma asserting that w.u.p.p. $G(T)$ is simple and contains no bad vertices whenever $\mathscr{L}(t)$ is subcritical for all $0 \leqslant t \leqslant T$. This is because the evolution of $\mathscr{L}(t)$ is almost, but not completely, independent of the generation of bad vertices. Insisting that $\mathscr{L}$ remains subcritical for all $0 \leqslant t \leqslant T$ w.h.p., i.e., in most runs, allows us to absorb such correlations.

We start by observing that every time we expose a copy in the course of a round, the probability that a bad event occurs, i.e., that we get a self-loop, a multiple edge, or a 0-color vertex, is bounded by the following ratio: the number of unexposed copies belonging to vertices that have had at least one copy exposed during the current round, divided by the total number of unexposed copies belonging to all other vertices. Moreover, note that this fact holds independently of the rest of the history of the process, i.e., these two quantities determine the probability of a bad event. At the same time, as we remark in the previous paragraph, the total number of copies exposed in a round is not independent from a bad event occurring in that round. Thus, to bound the probability of bad events we proceed as follows.

Throughout the algorithm's execution let us say that a copy is "dangerous" if its vertex has already had some copy exposed in the current round. Let $s = 0, 1, \ldots$ enumerate the copies exposed in the course of the algorithm and let $X(s)$ be the number of dangerous copies just before we expose the sth copy. Recall now that if $\mathscr{L}(t)$ is subcritical there are at least αn unexposed copies at time t. So, if $\mathscr{L}(t)$ is subcritical for all $0 \leqslant t \leqslant T$, there exists some $\gamma > 0$ such that when we expose the sth copy in the course of the T rounds, the probability of a bad event occurring is at most $X(s)/(\gamma n)$.

Imagine now that before the algorithm starts we perform Z Bernoulli trials, each one having probability of success $p = 1/(\gamma n)$, and "conceal" their outcomes (we will determine the value of Z later on). When we run the algorithm, when we expose the sth copy, we also expose the outcome of $2X(s)$ of the Bernoulli trials. We will say that this procedure fails if any of the following occurs: (i) some Bernoulli trial succeeds, (ii) $X(s) > \gamma n/2$, or (iii) we run out of Bernoulli trials. Recalling that $\Pr[\mathrm{Bin}(2n, p) > 0] \geqslant np$ for all $np < 1/2$, we see that the failure of this procedure dominates the occurrence of a bad event. Moreover, we can bound the probability that $\sum_s 2X(s) > Z$, i.e., that we run out of Bernoulli trials, as follows.

Observe that $\sum_s X(s)$ over the course of a single round cannot be greater than square of the number of copies exposed in that round, since $X(s)$ changes by at most 1 with each copy exposed. Moreover, note that if $\mathscr{L}(t)$ is subcritical for all $0 \leqslant t < T$, part 2c of Lemma 9 implies that there exists some $\rho > 0$ such that w.h.p. for all $0 \leqslant t < T$, the length of round t is dominated by a geometric random variable with parameter $1-\rho$. Thus, analogously to the Bernoulli trials above, we can construct a set of T independent, identically distributed geometric random variables $F_0, \dots, F_{T-1}$, so that F_t dominates the length of round t. Note that under this construction we also get that the event $X(s) > \gamma n/2$ is dominated by the event that at least one of these geometric random variables is greater than $\gamma n/2$. Now, by standard arguments, it is easy to show that if $T = \Theta(n)$ then w.h.p. $\sum_{t<T} F_t^2 < 2/(1-\rho) \times T$ and no $F_t > \gamma n/2$. Therefore, as long as $T = \Theta(n)$, we can take $Z = O(n)$ to guarantee that w.h.p. we do not run out of Bernoulli trials. Thus, the lemma follows by observing that $\mathscr{L}(t)$ is indeed subcritical w.h.p. for all $0 \leqslant t \leqslant T$ and that $\mathrm{Bin}(O(n), O(1/n))$ equals 0 w.u.p.p. □

9. The method of differential equations

The idea of using differential equations to approximate discrete random processes goes back at least to Kurtz [14,15]. It was first applied in the analysis of algorithms by Karp and Sipser [12] and has been greatly expanded since then by Mitzenmacher [16,18,19] and Wormald [25]. In this section we show how to employ the main theorem of [25] to track the evolution of $\mathscr{L}(t)$. That is, for $0 \leqslant i \leqslant \Delta$, where Δ is the maximum degree of the list sequence, we will track the random variables $R_i(t)$, $G_i(t)$, $B_i(t)$ and $W_i(t)$ for $0 \leqslant t \leqslant T$, where T is some a priori determined number of rounds. We start with the (rather technical) statement of the theorem and then discuss how it applies in our setting.

In the statement of Theorem 8 below, asymptotics denoted by o and O are for $n \to \infty$ but uniform over all other variables. In particular, "uniformly" refers to the convergence implicit in the $o()$ terms. For a random variable X, we say $X = o(f(n))$ *always* if $\max\{x \mid \Pr[X = x] \neq 0\} = o(f(n))$. We say that a function f satisfies a *Lipschitz condition* on $D \subseteq \mathbb{R}^j$ if there exists a constant $L > 0$ such that $|f(u_1, \dots, u_j) - f(v_1, \dots, v_j)| \leqslant L \sum_{i=1}^{j} |u_i - v_i|$ for all $(u_1, \dots, u_j)$ and $(v_1, \dots, v_j)$ in D.

Theorem 8 (Wormald [25]). *Let $Y_i(t)$ be a sequence of real-valued random variables, $1 \leqslant i \leqslant k$ for some fixed k, such that for all i, all t and all n, $|Y_i(t)| \leqslant Cn$ for some constant C. Let $\mathbf{H}(t)$ be the history of the sequence, i.e., the matrix $\langle \vec{Y}(0), \dots, \vec{Y}(t) \rangle$, where $\vec{Y}(t) = (Y_1(t), \dots, Y_k(t))$.*

Let $I = \{(y_1, \dots, y_k) : \Pr[\vec{Y}(0) = (y_1 n, \dots, y_k n)] \neq 0$ for some $n\}$. Let D be any bounded connected open set containing the intersection of $\{(s, y_1, \dots, y_k) : s \geqslant 0\}$ with a neighborhood of $\{(0, y_1, \dots, y_k) : (y_1, \dots, y_k) \in I\}$.[2] By "always", below, we mean for any value of $(t/n, Y_0(t)/n, \dots, Y_k(t)/n)$.

Let $f_i : \mathbb{R}^{k+1} \to \mathbb{R}$, $1 \leqslant i \leqslant k$, and suppose that for some $m = m(n)$,

(i) *for all i and uniformly over all $t < m$,*

$$\mathbf{E}[Y_i(t+1) - Y_i(t) | \mathbf{H}(t)] = f_i(t/n, Y_0(t)/n, \dots, Y_k(t)/n) + o(1), \text{ always;}$$

[2] That is, after taking a ball around the set I, we require D to contain the part of the ball in the halfspace corresponding to $s = t/n \geqslant 0$.

(ii) *for all i and uniformly over all* $t < m$,

$$\Pr[|Y_i(t+1) - Y_i(t)| > n^{1/5} | \mathbf{H}(t)] = o(n^{-3}), \ \textit{always};$$

(iii) *for all i*, f_i *is continuous and satisfies a Lipschitz condition on D.*

Then:

(a) *for* $(0, \hat{z}^{(0)}, \ldots, \hat{z}^{(k)}) \in D$ *the system of differential equations*

$$\frac{\mathrm{d}z_i}{\mathrm{d}s} = f_i(s, z_0, \ldots, z_k), \quad 1 \leqslant i \leqslant k$$

has a unique solution in D for $z_i : \mathbb{R} \to \mathbb{R}$ *passing through* $z_i(0) = \hat{z}^{(i)}, \ 1 \leqslant i \leqslant k$, *and which extends to points arbitrarily close to the boundary of D*;

(b) *with high probability*

$$Y_i(t) = z_i(t/n)n + o(n),$$

uniformly for $0 \leqslant t \leqslant \min\{\sigma n, m\}$ *and for each i, where* $z_i(s)$ *is the solution in* (a) *with* $\hat{z}^{(i)} = Y_i(0)/n$, *and* $\sigma = \sigma(n)$ *is the supremum of those s to which the solution can be extended.*

Remark. The theorem remains valid if the reference to "always" in (i) and (ii) is restricted to the event $(t/n, Y_0(t)/n, \ldots, Y_k(t)/n) \in D$.

Let us first discuss the conditions of this theorem less formally. Given a finite collection of random variables $Y_1, \ldots, Y_k$, Theorem 8 allows us to construct a set of deterministic real-valued functions $y_1, \ldots, y_k$ with the property that w.h.p. for all t considered, $Y_i(t) = y_i(t/n) \cdot n + o(n)$. This is achieved by taking a collection of equations describing the single-step conditional expected change of each random variable, such as Eqs. (5) and (6), and transforming them into a system of differential equations. (In our case, a single step corresponds to a round of A.) The functions y_i are the solutions to that system. Naturally, there are a number of issues that need to be addressed before this intuitive transformation can yield rigorous mathematical results. The statement of Theorem 8, while appearing rather technical, simply formalizes these issues. Specifically:

- We want the conditional expected change of each Y_i to always be bounded, i.e., for every possible history $\mathbf{H}$ of the process. Moreover, we want to be able to approximate that expectation within $o(1)$, even if we are only given the current value of each random variable within $o(n)$. We can view this as being able to approximate the conditional expected change even if we are only given a "macroscopic" view of the process, i.e., even if we only know the value of each Y_i within $o(n)$.
- For every possible history, we want the conditional change of each random variable in a round to have reasonable tail behavior, so that even over many rounds ($n^{2/3}$, say) the total change to $\vec{Y}(t)$ will be sharply concentrated around its expectation.
- The two conditions above already allow us to convert the probabilistic dynamics of the process into a deterministic, algebraic process. Moreover, we want this map to be a smooth function of the state, which translates to a certain stability of the underlying random process. In particular, we want there to be an absolute bound, i.e., a Lipschitz condition, on how much a small perturbation of the state can change the dynamics.

These are precisely conditions (i)–(iii) of Theorem 8. Note that in the above we demanded that the random process behaves "nicely" for any possible history. In many cases, this is too much to ask for. In particular, while the process might be expected to behave nicely in the vast majority of runs, one cannot definitively exclude the possibility that the process enters a bad regime in which "all bets are off." To deal with this issue, one can prespecify a set of "good" states such that for states in that set, all the desiderata are met. In particular, one can carve out a set in $\mathbb{R}^{k+1}$ (since t is also part of the state) and establish that as long as the (rescaled) state lies in that set, the conditions are met. This set is precisely the domain D in Theorem 8. As we remarked after stating the theorem, it suffices for the process to be "nice" for states inside this domain.

In our case, the set of good states amounts to the set of subcritical list sequences. To make this more precise let us first introduce the following notation.

Definition 9. For a sequence $\{s_i\}$, let $s^{(q)} = \sum_i i^q s_i$ denote its qth moment.

Let $z = (r_i, g_i, b_i, w_i)_{i=0}^{\Delta} \in \mathbb{R}^{4\times(\Delta+1)}$ and let $u^{(1)} = r^{(1)} + g^{(1)} + b^{(1)}$. Given such a point z, analogously to the matrix M of Eq. (2) for list sequences, we define a square matrix $A = A(z)$ of size $3(\Delta+1)$ as

$$A_{(x,i),(y,j)} = \begin{cases} \dfrac{j(i+1)c_{i+1}}{w^{(1)}+u^{(1)}} & \text{if } x \neq y, \\ 0 & \text{if } x = y, \end{cases} \tag{11}$$

where $c = \{r, g, b\} \setminus \{x, y\}$. We will say that z is (α, β)-*subcritical* if $u^{(1)} > \alpha$ and $\lambda_1(A) < 1 - \beta$. For fixed α, β we will let the domain $D = D(\alpha, \beta)$ consist of all (α, β)-subcritical points in $[0,1]^{4\times(\Delta+1)}$. With these definitions at hand, we are ready to show that we can apply Wormald's theorem to track the evolution of list sequences that correspond to points inside our domain.

(i) If $(R_i(t)/n, G_i(t)/n, B_i(t)/n, W_i(t)/n)_{i=0}^{\Delta}$ belongs in the domain, then the list sequence is (α, β)-subcritical by Definition 7 and we can apply Lemma 9 to get the conditional expected change in each of $R_i(t), G_i(t), B_i(t), W_i(t)$. Indeed, these changes are given by parts 1, 2a, and 2b of Lemma 9. Moreover, since the scaling by n factors out, we see that these expected changes are, within $o(1)$, functions of the rescaled parameters $(R_i(t)/n, G_i(t)/n, B_i(t)/n, W_i(t)/n)_{i=0}^{\Delta}$ implying that we satisfy condition (i) of Theorem 8.

(ii) If $(R_i(t)/n, G_i(t)/n, B_i(t)/n, W_i(t)/n)_{i=0}^{\Delta}$ belongs in the domain, part 2c of Lemma 9 establishes condition (ii) with room to spare.

Thus, if for $c \in \{r, g, b\}$ and $0 \leqslant i \leqslant \Delta$ we let

$$p_{(c,i)} = \tfrac{1}{2}\frac{h(i)(x_i + y_i)}{u^{(1)}}, \quad \text{where } \{x, y\} = \{r, g, b\} - c, \tag{12}$$

$$k_c = k_c(A) = \sum_{i=0}^{\Delta} i \times ((I - A)^{-1} p)_{(c,i)}, \tag{13}$$

$$k = k_r + k_g + k_b \tag{14}$$

and ignore the Lipschitz requirement of Theorem 8, we get the following system of differential equations when $h(i) = i^{\alpha}$:

$$\frac{\mathrm{d}w_i}{\mathrm{d}x} = -k\frac{iw_i}{w^{(1)} + u^{(1)}}, \tag{15}$$

$$\frac{\mathrm{d}c_i}{\mathrm{d}x} = k_c\frac{(i+1)(w_{i+1} + c_{i+1})}{w^{(1)} + u^{(1)}} - k\frac{ic_i}{w^{(1)} + u^{(1)}} - \frac{i^{\alpha}c_i}{u^{(\alpha)}}. \tag{16}$$

(Recall that $u^{(1)} = r^{(1)} + g^{(1)} + b^{(1)}$ so this is indeed a system of differential equations in r_i, g_i, b_i, w_i.)

(iii) To prove the Lipschitz condition for our system of differential equations we need to be more careful. In particular, it is here where the variational properties we proved for the total progeny of multitype branching processes will come into play. First, though, recall that if we are inside the domain then $u^{(1)} > \alpha$ implying that we readily get a Lipschitz constant $L = L(\alpha)$ for the term $1/(w^{(1)} + u^{(1)})$. Moreover, note that since Δ is fixed, if $u^{(1)}$ is bounded away from 0, then the same must hold for $u^{(\alpha)}$ for any $\alpha \geqslant 0$, implying a Lipschitz condition for $1/u^{(\alpha)}$.

For $k_c = k_c(A)$ we need to show that for all $c \in \{r, g, b\}$ and $0 \leqslant i \leqslant \Delta$, each of the coordinates of the vector $(I - A)^{-1}p$ satisfies a Lipschitz condition. For that we first observe that by (12) the coordinates of the vector p describing the probability distribution for the progenitor satisfy a Lipschitz condition. Moreover, the entries of the matrix A defined in Eq. (11) also satisfy a Lipschitz condition since $u^{(1)}$ is bounded away from 0 for (α, β)-subcritical list configurations. To conclude the argument we observe that for any two points $z, z' \in D$ the corresponding matrices $A(z)$ and $A(z')$ each have entries bounded by Δ, size $3(\Delta + 1)$, and largest eigenvalue bounded by $1 - \beta$, for some $\beta > 0$. The Lipschitz condition then follows by applying part 3 of Lemma 8.

Having satisfied the conditions of Theorem 8, we observe that for fixed x, y, i, each element $A_{(x,i),(y,j)}$ is proportional to j. Thus every column of A is a multiple of the $(r, 1)$ column, the $(g, 1)$ column, or the $(b, 1)$ column, so A has rank no more than 3. Further, observe that (16) is, in fact, symmetric with respect to the colors $c \in \{r, g, b\}$. So, under symmetric initial conditions, i.e., if $r_i(0) = g_i(0) = b_i(0)$ for all i, we see that $c_i = u_i/3$ and $k_c = k/3$ for each $c \in \{r, g, b\}$. In that case, the columns $(r, 1)$, $(g, 1)$ and $(b, 1)$ are identical, implying that A has rank 1. Therefore, for symmetric initial conditions we can rewrite system (15)–(16) as

$$\frac{\mathrm{d}w_i}{\mathrm{d}x} = -k\frac{iw_i}{w^{(1)} + u^{(1)}}, \tag{17}$$

$$\frac{\mathrm{d}u_i}{\mathrm{d}x} = k\frac{(i+1)(w_{i+1} + u_{i+1}/3) - iu_i}{w^{(1)} + u^{(1)}} - \frac{i^{\alpha}u_i}{u^{(\alpha)}} \tag{18}$$

and the largest (and only nonzero) eigenvalue of A is

$$\lambda_1 = \frac{2}{3}\frac{u^{(2)} - u^{(1)}}{w^{(1)} + u^{(1)}}. \tag{19}$$

Therefore, staying inside the domain amounts to maintaining $u^{(1)} > \alpha$ and

$$\frac{2}{3}\frac{u^{(2)} - u^{(1)}}{w^{(1)} + u^{(1)}} < 1. \tag{20}$$

It is instructive to compare (20) with the Molloy-Reed criterion for a degree sequence $\{a_i\}$ to be below the threshold for the emergence of a giant component in the configuration model [21]. In our notation, this criterion becomes

$$\lambda_{\text{MR}} = \frac{a^{(2)} - a^{(1)}}{a^{(1)}} < 1. \tag{21}$$

We can think of λ_{MR} as the expected progeny in the branching process corresponding to breadth-first search from a random copy, just as λ_1 is the expected progeny in the branching process corresponding to the forced coloring steps. Thus, we see that in both cases a vertex v is engaged by the process with probability proportional to its degree i, in which case its $i-1$ remaining copies become potential progeny (note that we are referring only to the 2-color vertices in the coloring case). This yields the numerator $u^{(2)} - u^{(1)}$ (resp. $a^{(2)} - a^{(1)}$) in each case. The denominators reflect the fact that the progeny of v is distributed among all unexposed copies, which in the case of coloring includes the copies of 3-color vertices as well as 2-color ones. Finally, in the case of breadth-first search all copies produce new progeny, while in the case of coloring with probability 1/3 the match of a 2-color copy does not become a 1-color vertex because its color list already lacked the assigned color of v.

Assume now that for a given initial list sequence $\mathscr{L}(0)$, the solution to the differential equations (17) and (18) remains strictly inside the domain, i.e., $\lambda_1 < 1 - \xi$ for some constant $\xi = \xi(\theta)$ and $u^{(1)}(t/n) > \alpha$ for all $0 \leqslant x \leqslant T/n$. Then Theorem 8 implies that w.h.p. for all $0 \leqslant t \leqslant T$ and all $C \in \{R, G, B, W\}$,

$$C_i(t) = c_i(t/n)n + o(n).$$

Thus, w.h.p. each entry of M is within $o(1)$ of the corresponding entry in A, implying that the largest eigenvalue of M is w.h.p. bounded below 1 for all $0 \leqslant t \leqslant T$. Along with the fact $u^{(1)}(t/n) > \alpha$, this implies that w.h.p. the list sequence $\mathscr{L}(t)$ is subcritical for all $0 \leqslant t \leqslant T$ and the conditions of Lemma 10 are fulfilled. Furthermore, if T is such that the list sequence at $x = T/n$ is (δ, ε)-easy for some $\delta, \varepsilon > 0$, we have proved Lemmata 6 and 7 as well. To determine the initial conditions and the T, δ and ε for which the above is true we need to solve the differential equations (17) and (18).

Eq. (20) allows us to calculate k explicitly from Eq. (4) and get

$$k = \frac{3(w^{(1)} + u^{(1)})}{3w^{(1)} + 5u^{(1)} - 2u^{(2)}} \frac{u^{(\alpha+1)}}{u^{(\alpha)}}.$$

It is also beneficial to rescale time as in [24] by taking

$$\mathrm{d}v = \frac{k}{w^{(1)} + u^{(1)}} \mathrm{d}x.$$

That is, we measure time not by how many rounds A has run, but by how many edges have been exposed, and scale that as well according to the number of remaining edges. This gives a new system of differential equations in terms of v:

$$\frac{\mathrm{d}w_i}{\mathrm{d}v} = -iw_i, \tag{22}$$

$$\frac{\mathrm{d}u_i}{\mathrm{d}v} = (i+1)(w_{i+1} + u_{i+1}/3) - \left(i + \frac{3w^{(1)} + 5u^{(1)} - 2u^{(2)}}{3u^{(\alpha+1)}} i^{\alpha} \right) u_i. \tag{23}$$

In Section 11 we will show how setting $\alpha = 0$ causes Eqs. (22) and (23) to collapse to a much simpler form, corresponding to the differential equations in the analysis of 3-GL [4], which can be solved analytically. For $\alpha > 0$, however, while some parts of this system are solvable analytically (for instance, $w_i(v) = w_i(0)e^{-iv}$), in general we have to resort to high-precision numerical integration. We note that perhaps some other form for the preference function may result in a system that has a closed-form solution, but we have not been able to find one.

Remark. Since, in the end, we are only interested in the case of symmetric initial conditions, our development of multitype branching processes might appear superfluous. This is far from true. Being able to consider the dynamics for the case where there is no 2-color symmetry (and prove that the process still behaves reasonably) is essential in allowing us to define a domain in which we can apply Theorem 8. Without this "wiggle room" around the trajectory corresponding to the symmetric case, that would be impossible. Indeed, multitype branching processes offer an algebraic way of establishing "list-stability", an approach which is much more robust and general than the probabilistic arguments used previously.

10. Integrating the differential equations

To prove Lemma 6, and complete the proof of Theorem 2, we integrate (22) and (23) with initial conditions $w_i = u_i = 0$ for $i \neq 4$, while $w_4 = 1 - 3 \times 10^{-3}$ and $u_4 = 3 \times 10^{-3}$. Integrating even with $\alpha = 0$, i.e., with no preference for high-degree vertices, we find that λ_1 is never more than 0.91289. We define the number of rounds T implicitly by running the differential equations until the rescaled time is $v = 2$. At this point the uncolored vertices are (δ, ε)-easy with $\delta = 0.00198$ and $\varepsilon = 0.00051$.

In fact, even if the initial degree distribution has κn vertices of degree 5 and $(1-\kappa)n$ of degree 4, for $\kappa = 0.219$ we have $\lambda_1 < 0.99973$ at all times for $\alpha = 0$, so we claim that these graphs are 3-colorable as well. Setting $\alpha = 20$ improves this to $\kappa = 0.3$, but even $\alpha = 50$ only increases this to 0.302—falling far short of 5-regular graphs.

Similarly, to prove Lemma 7, and complete the proof of Theorem 1, we start with initial conditions $w_i = e^{-d} d^i / i! - 3\phi$ and $u_i = 3\phi$ for all $0 \leqslant i \leqslant \Delta_{\max}$. With $d = 4.0309$ and $\alpha = 13$ we find that λ_1 is never more than 0.99909. At $v = 1.2$ the uncolored vertices are (δ, ε)-easy with $\delta = 10^{-41}$ and $\varepsilon = 0.010$ (δ is now tiny since originally only a 10^{-15} fraction of the vertices have degree $\Delta_{\max} = 30$ and these are the most likely vertices to be colored by the algorithm). By varying α and requiring that $\lambda_1 < 1$ at all times, we obtain the following series of lower bounds for d_3 (in all cases we integrated to 16 digits of precision and rounded $b(\alpha)$ down):

α	0	1	2	3	5	7	9	10	11	12	13
$b(\alpha)$	3.847	3.899	3.936	3.961	3.993	4.010	4.020	4.024	4.027	4.029	4.030

The first of these values is familiar from the original list-coloring algorithm of [4]. Indeed, in the next section we show that our differential equations reduce to those of [4] for the special case

$\alpha = 0$. As α increases, $b(\alpha)$ appears to converge somewhat slower than geometrically, but we conjecture that for $\alpha > 13$ the only improvement is in the third decimal digit. We also note that reducing the graph to its 3-core first (by repeatedly removing all vertices of degree smaller than 3) helps, but not very much. Using the results of [23] to get the degree sequence of the 3-core and plugging it in our differential equations with $\alpha = 13$ we get $d_3 \geqslant 4.04$, a slight improvement to Corollary 2.

11. Coloring with no degree preference

When $\alpha = 0$, our algorithm A becomes the list-coloring algorithm 3-GL of [4]. As mentioned in Section 2, under 3-GL the graph induced by the uncolored vertices is distributed at all times as $G(n', p = d/n)$, where n' is the number of uncolored vertices. As a result, the degree distribution of the uncolored vertices is asymptotically Poisson with a time-varying mean $\delta = pn'$. Moreover, the degree of each vertex is independent of its color list.

As a "sanity check" for our differential equations, we will show that for $\alpha = 0$, they lead to the differential equations in [4]. Recall that w_i and u_i are the degree distributions of the 3- and 2-color vertices, respectively, in our system. Using the information in the paragraph above, we "guess" that for $\alpha = 0$

$$w_i = \frac{\beta\delta^i}{i!} \quad \text{and} \quad u_i = \frac{\gamma\delta^i}{i!}, \tag{24}$$

for some time-varying parameters β, γ and δ. Moreover, if initially all vertices have 3 available colors and the graph is $G(n, p = d/n)$ then we must have

$$\beta(0) = e^{-d}, \quad \gamma(0) = 0, \quad \text{and} \quad \delta(0) = d. \tag{25}$$

From (24) we have

$$w^{(0)} = \beta e^{\delta}, \tag{26}$$

$$u^{(0)} = \gamma e^{\delta}, \tag{27}$$

$$w^{(1)} = \beta\delta e^{\delta}, \tag{28}$$

$$u^{(1)} = \gamma\delta e^{\delta}, \tag{29}$$

$$u^{(2)} = \gamma\delta(\delta + 1)e^{\delta}, \tag{30}$$

while the derivatives of w_i, u_i with respect to v (the time parameter) are

$$\frac{\mathrm{d}w_i}{\mathrm{d}v} = w_i\left(\frac{1}{\beta}\frac{\mathrm{d}\beta}{\mathrm{d}v} + i\frac{1}{\delta}\frac{\mathrm{d}\delta}{\mathrm{d}v}\right), \tag{31}$$

$$\frac{\mathrm{d}u_i}{\mathrm{d}v} = u_i\left(\frac{1}{\gamma}\frac{\mathrm{d}\gamma}{\mathrm{d}v} + i\frac{1}{\delta}\frac{\mathrm{d}\delta}{\mathrm{d}v}\right). \tag{32}$$

Now if we take Eqs. (22) and (23) with $\alpha = 0$, substitute (28)–(30), and match terms with (31) and (32), we find that (22) and (23) are satisfied for all i iff the following finite system of differential equations for the parameters β, γ and δ is satisfied:

$$\frac{d\beta}{dv} = 0, \quad \frac{d\gamma}{dv} = (\delta - 1)(\beta + \gamma), \quad \frac{d\delta}{dv} = -\delta. \tag{33}$$

With the initial conditions in (25), the solution to (33) is

$$\beta(v) = e^{-d}, \quad \gamma(v) = e^{-(v+de^{-v})} - e^{-d}, \quad \delta(v) = de^{-v}.$$

Using (26) and (27) and changing variables to $x = 1 - e^{-v}$ we get

$$w^{(0)}(x) = e^{-dx}, \quad u^{(0)}(x) = 1 - x - e^{-dx}$$

which is precisely the solution for 3-GL given in [4]. The reader can verify that Eq. (20) becomes

$$\lambda_1 = \frac{2\gamma\delta}{3(\beta + \gamma)} = \frac{2}{3} du^{(0)}.$$

This is maximized at $x = \ln d/d$, at which point $\lambda_1 = (2/3)(d - \ln d - 1)$. Setting this to 1 gives $d - \ln d = 5/2$, so $d = -W_{-1}(-e^{-5/2}) = 3.847\ldots$ where W_{-1} is the -1th branch of Lambert's W function, just as in [4].

We note that since Wormald's Theorem does not let us deal directly with infinite systems of differential equations, strictly speaking we cannot model A's progress with the infinite system (22) and (23) and then collapse it to the finite system of (33). Therefore, we cannot consider this as an alternate derivation of the results of [4]. We could make this approach rigorous either by truncating the degree sequence above some M and proving numerical lower bounds that approach the analytic bound of [4] as M increases, or by modelling the high-degree distribution as Poisson in the first place, e.g. as in [23].

12. Handling the high-degree vertices

Recall that our goal is to color the graph H of edges incident to high-degree vertices, so that we are left with a small number of 2-color low-degree vertices. If H were acyclic, i.e., if the high-degree vertices and their neighborhoods formed a forest, then we could accomplish this by simply 2-coloring H. In that coloring, every low-degree neighbor of a high-degree vertex trivially has a monochromatic neighborhood, so we could uncolor it and make it a 2-color vertex. Unfortunately, w.u.p.p. H cannot be colored so that all low-degree vertices have a monochromatic neighborhood, e.g. if two low-degree vertices occur in a 5-cycle. To remedy this we prove that after adding the neighbors of those low-degree vertices that appear in cyclic components of H, w.h.p. we can find a set of low-degree vertices that isolates the high-degree vertices from the rest of the graph.

Proof of Lemma 3. To prove part 1 of Lemma 3 we consider the following procedure, called High:

(1) Let E_H be the set of edges incident to high-degree vertices. Let H be the multigraph induced by E_H.
(2) Fail if any of the following is true:
    (a) H contains a component with more than one cycle.

   (b) H contains more than $\log n$ cyclic components.
   (c) H contains a component of size greater than $\log^2 n$.
(3) Let Y be the set of low-degree vertices in the cyclic components of H. Let E_Y be the set of edges incident to vertices in Y. Let K be the multigraph induced by $E_H \cup E_Y$.
(4) Fail if any of the following is true:
   (a) Some edge in E_Y connects two vertices in Y.
   (b) Some vertex of K that is not a vertex of H is contained in more than one edge of E_Y.
(5) Let L be the set of low-degree vertices in K that are not in Y. Find a 3-coloring η of K in which every vertex in L has a monochromatic neighborhood.

Lemma 11. *Step* 2 *of* `High` *succeeds w.h.p.*

Proof. Choose a random copy among all copies belonging to high-degree vertices and let v be the vertex to which it belongs. Let $C(v)$ be the number of copies belonging to vertices in v's connected component in the graph induced by E_H. We claim that there exists a constant $\rho > 0$ such that

$$\Pr[C(v) = s] < (1 - \rho)^s. \tag{34}$$

Using this claim and arguing as in Lemma 10 we get that the probability that a random copy lies in a cyclic component is $O(1/n)$ and the probability that it lies in a multicyclic component is $O(1/n^2)$. Therefore, the expected number of cyclic components in H is $O(1)$ while the expected number of multicyclic components is $o(1)$. Markov's inequality and the union bound, respectively, imply that w.h.p. H has no more than $\log n$ cyclic components and no multicyclic components. Moreover, since the number of vertices in $C(v)$ cannot be greater than its number of copies plus 1, (34) readily implies that w.h.p. H has no component of size $\log^2 n$.

To prove (34) we argue as in Lemma 9. That is, again, we consider a modified process S for exposing the component of v which stops if it ever exposes $\log^2 n$ copies, so that there is very little shift in the degree sequence during the course of the process. To bound the number of copies exposed by S we introduce a branching process B and show that the total progeny of B dominates the number of copies in S. The main difference is that this time, in order to account for the fact that low-degree vertices in H are never adjacent to other low-degree vertices, we distinguish the copies into two types, namely those belonging to low-degree vertices and those belonging to high-degree vertices. Specifically, a copy of a high-degree vertex gives birth to all other copies corresponding to the vertex of its partner, while a copy of a low-degree vertex does the same *only if its partner belongs to a high-degree vertex*. At the same time, the progenitor gives birth to all the copies of v other than itself. Thus we now have the 2×2 matrix,

$$M = \begin{pmatrix} p_{\text{high}}(d_{\text{high}} - 1) & p_{\text{high}}(d_{\text{low}} - 1) \\ p_{\text{low}}(d_{\text{high}} - 1) & 0 \end{pmatrix},$$

where p_{high} and p_{low} are the probabilities that a random copy in $\mathscr{D}^*$ is of high- or low-degree respectively, while d_{high} and d_{low} is the average degree of (the vertex corresponding to) a random high- or low-degree copy. Observe now that since S stops if it ever exposes $\log^2 n$ copies, adding a $o(1)$ term to each entry in M suffices to account for any shift in the degree sequence occurring during the course of S. The largest eigenvalue of M is

$$\lambda_1 = \frac{1}{2} p_{\text{high}}(d_{\text{high}} - 1)\left(1 + \sqrt{1 + 4\frac{p_{\text{low}}(d_{\text{low}} - 1)}{p_{\text{high}}(d_{\text{high}} - 1)}}\right).$$

For the degree sequence $\mathscr{D}^*$ with $\Delta_{\max} = 30$ a little arithmetic gives $\lambda_1 < 10^{-15}$. Since the degree sequence of $\mathscr{D}^*$ has an exponential tail, (34) follows from part 2 of Lemma 8. □

Lemma 12. *Step* 4 *of* `High` *succeeds w.h.p.*

Proof. Since we have passed conditions 2b and 2c, the total number of vertices in H lying on cyclic components is bounded by $\log^3 n$. Since the maximum degree of the graph is bounded by $2 \log n / \log \log n$, this implies $|E_Y| < \log^4 n$. Therefore, with certainty, there are $\Omega(n)$ unexposed copies in the rest of the graph, i.e., not in H, and those vertices have bounded degree. As a result, the probability of each of the events in step 4 is $O(|E_Y|^2/n) = O(\log^8 n)/n$. □

Lemma 13. *Step* 5 *of* `High` *succeeds w.h.p.*

To prove this we will use the following lemma.

Lemma 14. *Let K be a* 3*-colorable graph and let S be an independent set of K such that no cycle passes through any vertex in S. Then K can be* 3*-colored so that every vertex in S has a monochromatic neighborhood.*

Proof. Consider the connected components K_i of $K \backslash S$. Let us define a bipartite graph Γ on $S \cup \{K_i\}$ where $u \in S$ and K_i are connected if there is an edge between them in K. Clearly, Γ is acyclic since any cycle in Γ would induce a cycle in K passing through some vertex in S. Therefore, any two components K_i, K_j have at most one neighbor $u \in S$ in common. Moreover, each $u \in S$ is adjacent in K to at most one vertex $v_i(u)$ in each K_i.

To 3-color K it suffices to show that we can 3-color any connected component of Γ. Observe that since K is 3-colorable, so is each K_i. Moreover, for any $v \in K_i$ and $c \in \{R, G, B\}$, there is a 3-coloring η_i of K_i such that $\eta_i(v) = c$. So, for each connected component of Γ we start by coloring an arbitrary K_i in that component. Now, whenever we color some K_i with a coloring η_i, for every uncolored K_j that shares a neighbor $u \in S$ with K_i we do the following:

(1) Choose a color $C \neq \eta_i(v_i(u))$ and assign it to u.
(2) Color K_j with η_j such that $\eta_j(v_j(u)) = \eta_i(v_i(u))$.

Since Γ is acyclic, this procedure will never specify the color of more than one vertex of any K_i, so we will succeed in coloring every K_i. For the same reason, it never attempts to color any $u \in S$ more than once and, moreover, each $u \in S$ has a monochromatic neighborhood since when coloring it we require all its neighbors $v_j(u)$ to have the same color. □

It is a standard result in random graph theory [7] that a random multigraph induced by a degree sequence with bounded second moment is w.u.p.p. simple. Since we have already established that w.h.p. the graph K can be 3-colored so that all vertices in L have monochromatic neighborhoods, this concludes the proof of part 1 of Lemma 3.

To prove part 2 of Lemma 3, observe that a vertex in B has degree i if its original degree was $j \geqslant i$ and $j - i$ of its edges were incident to vertices in either H or Y. Moreover, observe that a given copy of a low-degree vertex is matched to a vertex in H with probability $\phi/d + o(n^{-1/3})$ and recall that w.h.p. $|E_Y| < \log^4 n$. Thus, an easy computation gives the expected number of vertices of degree i, while concentration follows from standard arguments.

Finally, part 3 of Lemma 3 follows by observing that the vertices in L are either neighbors of high-degree vertices or adjacent to some vertex in Y. Clearly, the number of vertices adjacent to high-degree vertices cannot exceed ϕn, while the number of vertices adjacent to Y cannot exceed $|E_Y|$. Moreover, it is clear that w.h.p. $\Omega(n)$ of the ϕn edges incident to high-degree vertices are between high-degree vertices. □

Acknowledgments

We thank Frank McSherry, Martin Nilsson, Yuval Peres, and Oded Schramm for helpful conversations. We also thank the referees for numerous useful comments.

References

[1] D. Achlioptas, Threshold phenomena in random graph coloring and satisfiability, Ph.D. Thesis, Department of Computer Science, University of Toronto, 1999.

[2] D. Achlioptas, Setting two variables at a time yields a new lower bound for random 3-SAT, in: 32nd Annual ACM Symposium on Theory of Computing (Portland, OR, 2000), ACM, New York, 2000, pp. 28–37.

[3] D. Achlioptas, E. Friedgut, A sharp threshold for k-colorability, Random Struct. Algorithms 14 (1) (1999) 63–70.

[4] D. Achlioptas, M. Molloy, The analysis of a list-coloring algorithm on a random graph, in: 38th Annual Symposium on Foundations of Computer Science (Miami, FL, 1997), IEEE Comput. Soc. Press, Los Alamitos, CA, 1997, pp. 204–212.

[5] E. Bender, E. Canfield, The asymptotic number of labelled graphs with given degree sequences, J. Combin. Theory Ser. A 24 (1978) 296–307.

[6] S. Boettcher, A. Percus, Optimization with extremal dynamics, Phys. Rev. Lett. 86 (2001) 5211–5214.

[7] B. Bollobás, Random Graphs, Academic Press, London, New York, 1985.

[8] D. Brelaz, New methods to color the vertices of a graph, Comm. ACM 22 (1979) 251–256.

[9] V. Chvátal, Almost all graphs with 1.44n edges are 3-colorable, Random Struct. Algorithms 2 (1) (1991) 11–28.

[10] P. Erdős, A. Rényi, On the evolution of random graphs, Magyar Tud. Akad. Mat. Kutató Int. Közl. 5 (1960) 17–61.

[11] A.C. Kaporis, L.M. Kirousis, Y.C. Stamatiou, A note on the non-colorability threshold of a random graph, Electron. J. Combin. 7 (2000) R29.

[12] R. Karp, M. Sipser, Maximum matchings in sparse random graphs, in: 22nd Annual Symposium on Foundations of Computer Science, IEEE Comput. Soc. Press, Los Alamitos, CA, 1981, pp. 364–375.

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

[14] T.G. Kurtz, Solutions of ordinary differential equations as limits of pure jump Markov processes, J. Appl. Probab. 7 (1970) 49–58.

[15] T.G. Kurtz, Approximation of Population Processes, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1981.

[16] M.G. Luby, M. Mitzenmacher, M.A. Shokrollahi, D.A. Spielman, V. Stemann, Practical loss-resilient codes, in: Proceedings of the 29th ACM Annual Symposium on the Theory of Computing, ACM, New York, 1997.
[17] T. Łuczak, Size and connectivity of the *k*-core of a random graph, Discrete Math. 91 (1) (1991) 61–68.
[18] M. Mitzenmacher, Analyses of load stealing models based on families of differential equations, Theory Comput. System 34 (1) (2001) 77–98.
[19] M. Mitzenmacher, Studying balanced allocations with differential equations, Combin. Probab. Comput. 8 (5) (1999) 473–482.
[20] C.J. Mode, Multitype Branching Processes. Theory and Applications, American Elsevier Publishing Co., Inc, New York, 1971.
[21] M. Molloy, B. Reed, A critical point for random graphs with a given degree sequence, Random Struct. Algorithms 6 (2–3) (1995) 161–179.
[22] R. Mulet, A. Pagnani, M. Weigt, R. Zecchina, Coloring random graphs, Phys. Rev. Lett. 89 (2002) 268701.
[23] B. Pittel, J.H. Spencer, N.C. Wormald, Sudden emergence of a giant *k*-core in a random graph, J. Combin. Theory Ser. B 67 (1) (1996) 111–151.
[24] N. Wormald, Some problems in the enumeration of labelled graphs, Ph.D. Thesis, Newcastle University, 1978.
[25] N.C. Wormald, Differential equations for random processes and random graphs, Ann. Appl. Probab. 5 (4) (1995) 1217–1235.