Coloring Random Graphs

A Short and Biased Survey

Lefteris M. Kirousis

RA Computer Technology Institute

and

Department of Computer Engineering and Informatics, University of Patras

The k-colorability problem (k-COL)

- Given a graph $G = (V, E)$ decide whether its vertices can be colored with at most k colors so that no adjacent vertices get the same color.

The List Coloring Algorithm

Input: A graph G together with a list of possible k colors for each of its vertices.

- At every step, choose a color from a list and assign it to its vertex.
- Delete this vertex and also delete the selected color from neighboring vertices.
- If the graph becomes empty, return “yes”; if a vertex with an empty list appears, return “no”.
- Vertices with one element in their list are given priority.

Improvements

- **The greedy list coloring algorithm:** Always choose a vertex with the least possible number of colors in its list. Ties are broken arbitrarily.
- **The Brelaz heuristic, 1979:** At ties, choose a vertex with the largest number of yet uncolored neighbors.

The Erdős–Rényi model

- $G_{n,p}$: Each edge is independently selected with probability p to be included in the graph (the number of edges is a random variable).
- $G_{n,m}$: Exactly m edges are uniformly and independently selected to be included in the graph.

The Erdős–Rényi model II

- We consider *sparse* graphs, i.e. graphs in $G(n, p)$ where $p = d/n$ for some constant d, or alternatively in $G(n, m)$, $m = dn/2$.
- Expected number of edges in $G(n, p = d/n)$ is $\sim dn/2$ and therefore expected "average" degree is d. The value $d/2$ is known as the *edge-density*.
- The two models although formally non-equivalent, they behave in a similar way.

Phase transition — non-rigorous results

[Mitchell et al. 1992] and other groups by *simulation experiments:*

General observation: for each fixed k (amenable to experimentation), there is a *threshold average degree* d_k^* such that

- If $d < d_k^*$, then a random graph with average degree d is a.a.s. k-colorable, while
- if $d > d_k^*$ then such a graph is a.a.s. non-k-colorable.

Note: “a.a.s.” means with probability approaching 1 asymptotically with the number of vertices.

Phase transition — continued

- Analytic (but non-rigorous) verification of the previous experimental results by methods of Statistical Physics.
- For $k = 3$, both experiments and analytic techniques suggest that $d_3^* \simeq 4.69$.

The Achlioptas–Friedgut theorem

Theorem *There is a sequence* $d^*(n)$ *such that* $\forall \epsilon :$

• *A random graph with average degree* $d^*(n) - \epsilon$ *is a.a.s.* 3*-colorable.*
• *A random graph with average degree* $d^*(n) + \epsilon$ *is a.a.s. non-*3*-colorable.*

- In other words, the transition interval can be made arbitrarily thin (sharp transition).

The Achlioptas–Friedgut theorem

Theorem *There is a sequence* $d^*(n)$ *such that* $\forall \epsilon :$

• *A random graph with average degree* $d^*(n) - \epsilon$ *is a.a.s.* 3*-colorable.*
• *A random graph with average degree* $d^*(n) + \epsilon$ *is a.a.s. non-*3*-colorable.*

- In other words, the transition interval can be made arbitrarily thin (sharp transition).

• **Still open question:** Does $d^*(n)$ *converge*? If yes, to what value?

Corollary *Given* d*, if for random graphs* G *of average degree* d*,* $\Pr[G$ *is* 3*-col.*$] > \epsilon$*, finally for all* n*, then for random graphs* G *of average degree any* $d' < d$*,* $\Pr[G$ *is* 3*-col.*$] \to 1$.

Upper bounds

The upper bound results are expressed in terms of the *edge-density* ($d/2$ rather than the average degree d). Because the $G(n,m)$ model works better in this case.

Reminder: Experimentally, the putative threshold occurs for edge-density 2.35 (average degree 4.69).

- First: 2.71 — observed by several researchers independently – *Markov's inequality*
- Current best: 2.427 Dubois and Mandler (2002) – *typical graphs plus the decimation technique*

Upper bounds

The upper bound results are expressed in terms of the *edge-density* ($d/2$ rather than the average degree d). Because the $G(n, m)$ model works better in this case.

Reminder: Experimentally, the putative threshold occurs for edge-density 2.35 (average degree 4.69).

- First: 2.71 — observed by several researchers independently – *Markov's inequality*
- Current best: 2.427 Dubois and Mandler (2002) – *typical graphs plus the decimation technique*

In between the above *less-than-three-tenths* improvement, several results that established intermediate values by Łuczak; Achlioptas and Molloy; Kaporis et al. and other groups

The basic upper bound technique

Let G be a random graph and $\mathcal{C}(G)$ the random class of its legal 3-colorings. Then

$$\Pr[G \text{ is } 3\text{-col.}] = \Pr[|\mathcal{C}(G)| \geq 1] \leq \mathbb{E}(|\mathcal{C}(G)|).$$

Since $\mathbb{E}(|\mathcal{C}(G)|)$ is easy to compute. So we find the values of edge-density for which it vanishes and thus we get a trivial upper bound, namely 2.71.

But why experiments suggest $d_3^* \simeq 2.35$?

A class of graphs with small probability but with *many* legal 3-colorings contributes *too much* to the expectation $\mathbb{E}(|\mathcal{C}(G)|)$ **(Lottery paradox).**

Improvements of the first moment

- Make $\mathcal{C}(G)$ (the set of all legal 3-colorings) "thinner", so that the "unrealistic" expectation of the cardinality of $\mathcal{C}(G)$, due to the Lottery paradox, gets smaller.
  - Consider *rigid* 3-colorings, i.e. colorings where any change of color to a higher one (in the RGB ordering) destroys the legality (Achlioptas and Molloy, further improvement by Kaporis, Kirousis et al.)
- Method first introduced for SAT by Kirousis et al., 1997.

Improvements of the first moment II

• Examine not the whole space of possible graphs, but a subset of it comprised of graphs that:

- are *typical* with respect to their degree sequence (Poisson) and
- *(the decimation technique)* have been repetitively depleted of vertices of degree 2 or less, as these vertices do not interfere with the colorability (Dubois and Mandler).

Algorithms for lower bounds

- Let d_k^- denote the lower bound for d_k^* that we try to compute.
- Consider list coloring or an improvement *without backtracking*, i.e. if failure is reported after the choice of a color, this failure is considered permanent).
- Prove that for $d < d_k^-$, the coloring algorithm a.a.s. succeeds.
- The more sophisticated the heuristic is, the more difficult or impossible its probabilistic analysis is.

Algorithmic lower bounds for 3-COL

- Achlioptas and Molloy (1997), and then Achlioptas and Moore (2004) analyzed the plain list coloring algorithm (the Brelaz heuristic respectively).

The best today lower bound for 3-COL: average degree 4.03 (Recall: experimental value of putative threshold: 4.69).

- Progress in the case of lower bounds is much slower and the techniques more involved (compared with upper bounds).
- Also, there is strong experimental evidence that the technique of analyzable local search algorithms cannot overcome a barrier smaller than the value of the experimental threshold (4.69). Why?

The case of general k-COL

By the first moment method: $d_k^* < 2k \ln k$.

By a result of Łuczak: $d_k^* > 2k(1-\epsilon)\ln k, \forall \epsilon > 0$ and for large enough k.

So $d_k^* \sim 2k \ln k$ (the asymptotic is w.r.t k).

Also, by another result of Łuczak: the chromatic number of graphs in $G(n, p = d/n)$ ranges a.a.s. within a window of only *two* possible integer values.

Alas, Łuczak gives no information on these two values, neither do the above asymptotics of the threshold d_k^*.

Main task: Make the asymptotics of d_k^* finer so that the two possible values of the chromatic number are found.

The geometry of k-colorings

For a random graph G with a given av. degree d, consider the space of all k-color assignments (legal or not) to the vertices of G. Then:

- For d below a certain value, all legal k-colorings form a unique cluster in this space (with respect to the Hamming distance).
- As the average degree increases, the clusters break down into exponentially many.
- Moreover, as d increases, exponentially many clusters correspond to color assignments that are illegal in a locally minimal way (i.e. any change in the colors of a few vertices gives rise to more illegally colored edges).

The geometry of colorings cont'ed

- Beyond the clustering point, sampling colorings becomes hard.
- Therefore no easily analyzable local search algorithms.
- Scant hopes to sufficiently improve the lower bound of d_k^* by algorithmic techniques.

Above results Krzącała et al. (2004) and Zdeborová and Krzącała (2007).

What is the way out?

Conclusion: Non-algorithmic approaches for lower bounds should be tried.

The second moment method

Let X be a non-negative variable (usually a *counting* variable) that depends on n.

- Lottery Paradox: As n grows large $\mathbb{E}(X)$ may also grow large, but yet $\Pr[X > 0]$ may approach zero.
- However if $\mathbb{E}(X^2)$ does not approach infinity too fast compared to $\mathbb{E}(X)$, then it may turn that $\Pr[X > 0]$ stays away from zero. Formally:

$$\Pr[X > 0] \geq \frac{(\mathbb{E}(X))^2}{\mathbb{E}(X^2)}.$$

The solution

- Achlioptas and Naor (2004). Let k_d be the smallest integer k such that $d < 2k \ln k$. Almost all $G_{n,p=d/n}$ random graphs have chromatic number either k_d or $k_d + 1$.
- Method of Proof: Second moment where X counts the number of balanced k-colorings of $G_{n,p=d/n}$.
- Balanced: each color is assigned to an equal number of vertices.
- Difficulty: The second moment of X turns out to be a sum of exponential terms. Locating the term with the largest base, which essentially gives the value of the sum, proved out to be a difficult task.

Random regular graphs

Different model from the Erdős–Rényi. Special case of the Newman model of random graphs with a preassigned degree sequence, intended to model large complex graphs.

Progress is much slower.

- Achlioptas and Moore (2004): The chromatic number of random regular graphs of degree d a.a.s. ranges in $\{k_d, k_d + 1, k_d + 2\}$, where k_d is the smallest integer k such that $d < 2k \ln k$.
- Shi and Wormald (2004): Algorithmic analogous results for values of d up to 10.
- Also, almost all 4-regular graphs have chromatic number 3, and
- almost all 6-regular graphs have chromatic number 4.

5-regular graphs

- Survey Propagation (Krząkała et al., 2004): almost all 5-regular graphs have chromatic number 3.
- The solution space of 3-colorings of 5-regular is on the edge of the clustering phase. Therefore, rigorously analyzable algorithmic techniques are expected not to work.
- Until recently, the only rigorous result for 5-regular graphs is that almost all of them have chromatic number 3 or 4.
- Second Moment: Fails when X counts 3-colorings, even if they are balanced.

Why the 2nd m/ent fails for reg. graphs?

By linearity of expectation and by summing over pairs of 3-colorings we have:

$$\mathbb{E}(X^2) = \sum_i E_i P_i$$

where E_i is the number of pairs of color assignments with a given *pattern* of color assignments (characterized by a parameter i) (entropy factor) and

P_i is the probability that a fixed pair of color assignments with pattern E_i is legal (energy factor).

Explanation of failure continued

- The term $E_i P_i$ that is equal (ignoring sub-exponential factors) to $(\mathbb{E}(X))^2$ is the *barycentric* term that corresponds to a completely symmetric pattern.
- But unfortunately unlike the case of $G(n, p)$ graphs, the barycentric term is not the prevalent one in the sum.
- Because there is a slight bias towards pairs of colorings that give the same color to a vertex.

How to eliminate this bias?

- Consider colorings where each vertex has neighbors with both the other two legal colors (rainbow or panchromatic colorings).
- Díaz, Kaporis, Kirousis, Kemkes, Pérez and Wormald (2009): 5-regular graphs are 3-colorable a.a.s.
- Method of proof: Apply second moment to the number of rainbow, balanced 3-colorings on 5-regular graphs.
- Result: A 5-regular graph is 3-colorable with positive probability independent of its size.

High probability

Pad up this probability to 1 (asymptotically).

Technique: Using the previous result that:

$$\Pr[X > 0] \geq \frac{(\mathbb{E}(X))^2}{\mathbb{E}(X^2)} \sim \text{constant}$$

show that:

$$\Pr_Y[X > 0] \geq \frac{(\mathbb{E}_Y(X))^2}{\mathbb{E}_Y(X^2)} \sim 1,$$

by conditioning over the number of small cycles of the graph.

Thank you