# ANALYSIS OF THE GREEDY LIST COLORING ALGORITHM ON A RANDOM GRAPH

### MICHAEL KANE

As a special project for Spring 2007, a graph coloring algorithm proposed by Achlioptas and Molloy (1997) is analyzed.

#### 1. INTRODUCTION

#### 1.1. The Achlioptas Molloy Paper.

Achlioptas and Molloy (1997) introduce a greedy algorithm for k-coloring graphs. For a graph G = (V, E) with V vertices and E edges, denote  $L_v$  as the list of available colors for vertex v in V. The algorithm is given as follows:

- Initialize: U = V, and  $L_v = \{1, ..., k\}$  for every  $v \in V$ .
- While  $U \neq \emptyset$ :
  - (1) At time t select a vertex  $v_t$  uniformly at random from  $\{v_t \in U : |L_v| \text{ is smallest }\}$
  - (2) Assign to  $v_t$  a color  $\omega$  chosen uniformly at random from  $L_v$ , and set  $U = U \setminus \{v_t\}$
  - (3) For each  $u \in U$  that is a neighbor of  $v_t$  set  $L_u = L_u \setminus \{\omega\}$
  - (4) If  $L_v = \emptyset$  for any  $v \in V$  then the algorithm fails.

Achlioptas and Molloy analyzed the algorithm on random graphs to find the probability of its success or failure on Erdös-Rényi Graphs (Section 1.2). They claimed that for the three color problem an Erdös-Rényi Graph G(n, c/n) with 1.923n edges, the probability of colorability tends to 1 as nincreases. Furthermore, for the general k-coloring problem, they claimed that the analysis implies that for  $k \ge 3$ , if  $c \le k \log k - 3/2k$ , then the probability the succeeds goes to 1 as n increases. Finally, they claim that for any  $\varepsilon > 0$ , and k sufficiently large, if  $c \ge (1 + \varepsilon)k \log k$ , then the probability the algorithm fails goes to 1 as n increases.

1.2. The Erdös-Rényi Graph. This report is concerned with the probabilistic behavior of the given graph coloring algorithm when applied to an *Erdös-Rényi Graph* G(n, p) as n tends to infinite and p tends to zero at a rate of 1/n. The graph G(n, p) has n vertices with each of  $\binom{n}{2}$  possible

Date: June 21, 2007.

edges being considered independently for inclusion with probability p. Thus, the total number of edges is random with distribution

$$\operatorname{Bin}(m,p)$$
 where  $m = \binom{n}{2}$ 

It should be noted that in the analysis (as with this paper) it is assumed that the topology of the graph is not given entirely at the beginning of the algorithm. Instead, we learn about the graph as the algorithm proceeds. This is the principle of *late binding*, which has also been called the principle of *deferred decisions* (*le principe d'ajournement des décisions*) in Knuth (1976, Chapter 3) In the analysis of the algorithm, this method is employed so that at each step, the vertices that are exposed are those adjacent to colored vertices.

## 2. MODEL DESCRIPTION

For simplicity, I discuss the algorithm on Erdös-Rényi graphs only for the case k = 3. Let the colors being considered be denoted A, B, C. Let  $N_t$  denote the number of vertices with  $L_v = \{A, B, C\}$  at time t. Let  $X_1(t)$ ,  $X_2(t), X_3(t) Y_1(t), Y_2(t)$ , and  $Y_3(t)$  denote the number of vertices whose possible list colorings are AB, AC, BC, A, B, and C respectively at time t. Thus

$$S_t = X_1(t) + X_2(t) + X_3(t).$$

denotes the total number of vertices with 2 possible colorings at time t. Conditioning on all information up through time t, the expected value  $\mathbb{P}_t S_{t+1}$  depends only on its previous value  $S_t$ . Therefore the sequence of random variables,  $S_0, S_1, \dots, S_t$  is a Markov Chain.

For graph with n vertices, the algorithm starts with  $N_0 = n$ ,  $X_1(0) = X_2(0) = X_3(0) = Y_1(0) = Y_2(0) = Y_3(0) = 0$ , and V = U; At the first iteration:

- Select a  $v_1$  in V.
- Select its color. Say it is C.
- Say v₁ has b₁ neighbors. For each of these b₁ neighbors remove C from the list of possible colorings, leaving X₁(1) = b₁ and N₁ = n − b₁ − 1.
- Discard  $v_1$ , leaving uncolored vertices  $V = V \setminus \{v_1\}$ .

At the second iteration:

- Select  $v_2$  at random from  $b_1$  vertices with  $L_v = \{A, B\}$ .
- Select a color for  $v_2$  from  $\{A, B\}$ . Say it is B.

ANALYSIS OF THE GREEDY LIST COLORING ALGORITHM ON A RANDOM GRAPH 3



FIGURE 1. A Visualization for the List Counts: At time t, the algorithm is described by the counts  $X_1(t)$ ,  $X_2(t)$ ,  $X_3(t)$ ,  $Y_1(t)$ ,  $Y_2(t)$ ,  $Y_3(t)$ . If a vertex does not have possible coloring, the algorithm fails.

- Say  $v_2$  has  $b_2$  neighbors. For each of these  $b_2$  neighbors remove B from the list of possible colorings.
- Discard  $v_2$ .

At this stage  $X_2(2) = b_2$  and  $N_2 = n - S_2 - 2$ . There is also a small probability at this stage that a small number of vertices with color list  $\{A, B\}$  will have neighbor  $v_2$  in which case those vertices have color list  $\{A\}$  at the end of this stage.

A visualization of the vertex list counts is shown in Figure (1). At stages in the algorithm where there are 1-color vertices, the algorithm removes a 1-color vertex thereby reducing the 1-color count by 1. At the same time a number of vertices go from a 3-color to 2-color lists. Depending on the number of 2-color lists, some of the 2-color lists may become 1-color lists. The algorithm proceeds until either it fails or all vertices are colored.

For the algorithm to fail, there must be at least 2 vertices with the same 1-color list and these vertices must be adjacent to each other at some time t < n. At t + 1 a number of vertices have empty color lists, and this means the algorithm has failed to find a valid coloring.

# 3. The Simplified Model

It is not immediately clear how to analyze all list counts in the algorithm, and managing all of these variables may become unwieldy. Therefore, as a first-pass, I consider a simplified model with vertices only having color lists  $\{A, B, C\}$ ,  $\{A, B\}$ ,  $\{A, C\}$  and  $\{B, C\}$ . Also, under the simplified model, colored vertices will not be removed from the system.

In the first step of the simplified algorithm, t = 0,  $X_1(t)$ ,  $X_2(t)$ , and  $X_3(t)$  are chosen with equal probability. For t > 1, a vertex with color lists  $\{A, B\}, \{A, C\}, \text{ or } \{B, C\}$  is picked to be colored proportional to its respective list count. In other words, for vertices with color list  $\{A, B\}$  at time t, denoted  $V_{AB}(t)$ ,

$$\mathbb{P}\{\text{a vertex in } V_{AB}(t) \text{ is colored at } t+1|X_1(t), X_2(t), X_3(t)\} = \frac{X_1(t)}{S_t}.$$

Let  $b_{t+1}$  be the number of vertices that go from 3-color lists to 2-color lists. Then, the probability that the size of the 2-color list  $\{A, C\}$  goes from  $X_2(t)$  to  $X_2(t) + b_{t+1}$  is the probability that a vertex with color list  $\{A, B\}$  or  $\{B, C\}$  is picked to be colored times the probability that the color picked is B. Write  $\mathbb{P}_t$  for the conditional probability conditioned on information through time t, then we can write

$$\mathbb{P}_t \{ X_2(t+1) = X_2(t) + b_{t+1} \} = \frac{1}{2} \left( \frac{X_1(t) + X_3(t)}{S_t} \right)$$
$$= \frac{1}{2} \left( 1 - \frac{X_1(t)}{S_t} \right).$$

Let  $X_t$  be  $(X_1(t), X_2(t), X_3(t))$  and let  $\varepsilon_{t+1}$  indicate which stack get  $b_{t+1}$  items at t + 1. Then

$$\pi_t = \mathbb{P}_t \varepsilon_{t+1} = \frac{1}{2} \cdot (1 - X_t / S_t)$$

For the simplified model, the state of the system at time at time t being determined by the counts  $N_t$  and  $X_t = (X_1(t), X_2(t), X_3(t))$  is a Markov chain. Again, define  $S_t = \sum_i X_i(t)$ , then then transition to a state at time t + 1 is controlled by two conditionally independent quantities  $b_{t+1}$  with a  $\operatorname{Bin}(N_t, c/n)$  distribution and  $\varepsilon_{t+1}$  with a multinomial $(1, \pi_t)$  distribution.

Where

$$\pi_t = (\pi_1(t), \pi_2(t), \pi_3(t))$$
  
= 
$$\begin{cases} (1/3, 1/3, 1/3) \text{ if } t = 0\\ \frac{1}{2}(1 - X_t/S_t) \text{ if } t \ge 1. \end{cases}$$

It may be noted that this indicates that if at a given time t,  $X_1(t)$  is much larger than both  $X_2(t)$  and  $X_3(t)$ , then a vertex with color list  $\{A, B\}$  is more likely to be selected for coloring. Therefore, for i > 0,  $X_1(t + i)$  is unlikely to receive more items until its size is closer to  $X_2(t + i)$  or  $X_3(t + i)$ . A more formal justification of this balancing mechanism is provided in Section (3.2).

3.1. The Distribution of  $N_t$  and  $S_t$ . Letting q = 1 - c/n, the distribution of vertices with color list  $\{A, B, C\}$  is

$$N_t \sim \operatorname{Bin}(n, q^t).$$

For a vertex's color list to remain  $\{A, B, C\}$  at time t, it must survive t independent attempts at removing it. Each of these steps has a has probability q of failure.

In the simplified model,

$$S_t = n - N_t$$

and therefore, distribution of  $S_t$  is

$$S_t \sim \operatorname{Bin}(n, 1-q^t)$$

3.2.  $X_1(t)$ ,  $X_2(t)$ , and  $X_3(t)$  Remain Balanced. Define

$$\Delta_{t+1}X = X_{t+1} - X_t = b_{t+1}\varepsilon_{t+1}.$$

In order to show that  $X_1(t) \approx X_2(t) \approx X_3(t)$  with high probability, it is enough to consider each differences between 2-color stack sizes. The behavior of

$$Z_t = X_1(t) - X_3(t)$$

is typical.

Let

$$\gamma_{t+1} = \varepsilon_1(t+1) - \varepsilon_3 t + 1$$

and

$$\Delta_{t+1}Z = Z_{t+1} - Z_t.$$

References

The conditional probability of  $Z_t$  by  $\Delta_{t+1}Z$  is then

$$\begin{aligned} \mathbb{P}_t Z_t \Delta_{t+1} Z &= Z_t \mathbb{P}_t \Delta_{t+1} Z \\ &= Z_t \mathbb{P}_t b_{t+1} \mathbb{P}_t \gamma_{t+1} \\ &= Z_t N_t \frac{c}{n} \frac{1}{2} (1 - X_1(t) / S_t - 1 + X_3(t) / S_t) \\ &= -\frac{1}{2} \frac{c}{n} \frac{N_t}{S_t} Z_t^2 \\ &\leq 0 \end{aligned}$$

This implies that if, at some time, the difference  $|X_{AB}(t) - X_{BC}(t)|$  gets large, the stack selection process ensures that it will not remain large for long since subsequent iterations of the algorithm will tend to reduce this difference. Also, since the choice of stacks used to define  $Z_t$  was arbitrary, this negative feedback mechanism exists between all two-color each pair of elements in  $X_t$ . Now, if it can be shown that the expected conditional differences  $\mathbb{P}_t Z_{t+1}^2$  cannot grow too quickly, it can be concluded that the  $Z_t$ tends to be around zero.

3.3.  $\mathbb{P}_t Z_{t+1}^2$  is Bounded for by a Constant. We would like to show the expected conditional differences  $\mathbb{P}_t Z_{t+1}^2$  cann not grow too quickly in t

$$\begin{split} \mathbb{P}_t Z_{t+1}^2 &= \mathbb{P}_t (Z_t + \Delta_{t+1} Z)^2 \\ &= Z_t^2 + Z_t \mathbb{P}_t \Delta_{t+1} Z + \mathbb{P}_t \Delta_{t+1} Z^2 \\ &\leq Z_t^2 + \mathbb{P}_t b_{t+1}^2 \mathbb{P}_t \gamma_{t+1}^2 \\ &\leq Z_t^2 + \mathbb{P}_t b_{t+1}^2 \operatorname{since} \gamma_{t+1}^2 \leq 1 \\ &\leq Z_t^2 + (c+c^2) \operatorname{since} \mathbb{P}_t b_{t+1} \leq c+c^2. \end{split}$$

From this we can conclude

$$\mathbb{P}Z_{t+1}^2 \le \mathbb{P}Z_0^2 + t(c+c^2).$$

#### REFERENCES

- Achlioptas, D. and M. Molloy (1997). The analysis of a list-coloring algorithm on a random graph (extended abstract). *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science*, 204–212.
- Knuth, D. E. (1976). *Mariages Stables et Leurs Relations Avec D'autres Problemes Combinatoires*. Montreal: Les Presses De l'Université de Montréal.