ANALYSIS OF THE GREEDY LIST COLORING ALGORITHM ON A RANDOM GRAPH: PART 2

MICHAEL KANE

1. An Upper Bound on the Maximum $S_2(t)$

Define b_t as before. Using the fact that the number of vertices going from the 3-stack to a 2-stack is stochastically dominated by b_t,

$$\max_{1 \leq t \leq n\varepsilon} S_2(t) \leq \sum_{t=1}^{n\varepsilon} b_t$$

and therefore to get an upper bound on the biggest $S_2(t)$ it is sufficient to get a bound on the size of the sum of b_t. This can be done by centering each of the b_t values and then applying the Bennett Inequality (Pollard 2001, Chapter 11)

$$\begin{aligned} \mathbb{P}\left\{\sum_{t=1}^{n\varepsilon} b_t \geq \nu\right\} &= \mathbb{P}\left\{\sum_{t=1}^{n\varepsilon} b_t - n\varepsilon c \geq \nu - n\varepsilon c\right\} \\ &\leq \exp\left(-\frac{(\nu - n\varepsilon c)^2}{2c(1-c/n)}\psi\left(\frac{n(\nu - n\varepsilon c)}{c(1-c/n)}\right)\right) \end{aligned}$$

where

$$\psi(x) := \begin{cases} 2\left((1+x)\log(1+x) - x\right)/x^2 & \text{for } x \geq -1 \text{ and } x \neq= 0 \\ 1 & \text{for } x = 0. \end{cases}$$

As ν gets larger than the expected value of the sum of b_t's, that is $\nu > n\varepsilon c$, there is a sharp decrease in the probability of getting an $S_2(t)$ larger than ν. Therefore, we expect that $S_2(t)$ will not be much bigger than this value.

2. The Restricted 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, consider a simplified Markov chain similar to that in Figure **??** consisting of only $S_3(t)$, $X_\alpha(t)$ for $\alpha \in \{A, B, C\}$, and Y_β for $\beta \in \{\{A, B\}, \{A, C\}, \{B, C\}\}$. This means that a given vertex may go from the 3-color stack to a 2-color stack, the 3-color stack to the fixed color

Date: August 31, 2007.

stack, or a 2-color stack to the fixed color stack. This will be referred to as the restricted model. To begin with, the added simplification of no fixed-color stack will be imposed. In this case a "colored" vertex will retain its color list and the algorithm will proceed. It will not be added to the fixed-color stack. After the simplified model is well understood, the fixed-color stack will be introduced and analyzed.

In the first step of the algorithm, $t = 1$ in the restricted, no fixed-color stack, model, one of the n vertices is chosen uniformly at random and assigned a color which is chosen uniformly at random from 1 of its 3 colors. This color is removed from the color-lists of each of its adjacent vertices. For $t > 1$, a vertex with color lists $\{A, B\}$, $\{A, C\}$, 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 $V_{AB}(t)$,
(2.1)

$$\mathbb{P}\{\text{a vertex in } V_{AB}(t) \text{ is colored at } t{+}1 | X_{AB}(t), X_{AC}(t), X_{BC}(t)\} = \frac{X_{AB}(t)}{S_2(t)}.$$

Let b_{t+1} be the number of vertices that go from 3-color lists to 2-color lists at time $t + 1$. The probability that $X_{AC}(t + 1)$ is selected to receive b_{t+1} vertices from $S_3(t)$ 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

$$\begin{aligned} \mathbb{P}_t\{X_{AC}(t + 1) = X_{AC}(t) + b_{t+1}\} &= \frac{1}{2}\left(\frac{X_{AB}(t) + X_{BC}(t)}{S_2(t)}\right) \\ &= \frac{1}{2}\left(1 - \frac{X_{AB}(t)}{S_2(t)}\right). \end{aligned}$$

Let X_t be $(X_{AB}(t), X_{AC}(t), X_{BC}(t))$ and let ε_{t+1} indicate which stack get b_{t+1} items at $t + 1$. Then

$$\pi_t = \mathbb{P}_t \varepsilon_{t+1} = \frac{1}{2}. (1 - X_t/S_2(t))$$

For the restricted, no fixed-color stack, model the transition to a state at time $t + 1$ is controlled by two conditionally independent quantities b_{t+1}

and ε_{t+1} Where

$$\begin{aligned}\pi_t &= (\pi_{AB}(t), \pi_{AC}(t), \pi_{BC}(t)) \\ &= \begin{cases} (1/3, 1/3, 1/3) \text{ if } t = 0 \\ \frac{1}{2}\left(1 - X_t/S_2(t)\right) \text{ if } t \geq 1. \end{cases}\end{aligned}$$

If at a given time t, $X_{AB}(t)$ is much larger than both $X_{AC}(t)$ and $X_{BC}(t)$, then a vertex with color list $\{A, B\}$ is more likely to be selected for coloring. Therefore, for some $i > 0$, $X_{AB}(t+i)$ is unlikely to receive more items until its size is closer to $X_{AC}(t+i)$ or $X_{BC}(t+i)$. After the distributions of $S_3(t)$, $S_2(t)$, and b_t are established in the next subsection a more formal justification of this balancing mechanism is provided in Section (2.2).

2.1. **The Distribution of $S_3(t)$, $S_2(t)$, and b_t.** To find the distribution of $S_3(t)$, the 3-stack size, start by looking at the probability a vertex remains in the 3-stack through time t. For this to happen it must survive t independent attempts at removing it. For each of these attempts a 3-stack vertex has probability $(1 - c/n)$ of of remaining in the 3-stack. Therefore the size of the 3-stack is

$$S_3(t+1) \sim \text{Bin}(n, (1 - c/n)^t).$$

In the simplified, no fixed-color state, model

$$S_2(t+1) = n - S_3(t+1)$$

and therefore, distribution of $S_2(t+1)$ is

$$S_2(t+1) \sim \text{Bin}(n, 1 - (1 - c/n)^t)$$

Until now the number of vertices moving from the 3-stack to the 2-stack, b_{t+1} has been approximated as $\text{Bin}(n, c/n)$ since at each step of the algorithm, there is a c/n chance that a vertex in a 3-stack will move to a 2-stack, and there are at least n 3-stack vertices. A more accurate estimate is reached by realizing that at time t there are $S_3(t)$ possible 3-stack vertices. Therefore, a more accurate estimate of b_{t+1} is given by

$$b_{t+1} \sim \text{Bin}(S_3(t), c/n).$$

This approximation wil be used for the remainder of this paper.

2.2. $X_1(t)$, $X_2(t)$, and $X_3(t)$ Remain Balanced when Colors are Not Fixed.

Define

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

In order to show that $X_{AB}(t) \approx X_{AC}(t) \approx X_{BC}(t)$ with high probability, consider the differences between 2-color stack sizes. The behavior of

$$Z_t = X_{AB}(t) - X_{BC}(t)$$

is typical.

Let

$$\gamma_{t+1} = \varepsilon_{AB}(t+1) - \varepsilon_{BC}(t+1)$$

and

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

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

$$\begin{aligned}
\mathbb{P}_t Z_t \Delta_{t+1}Z &= Z_t \mathbb{P}_t \Delta_{t+1}Z \\
&= Z_t \mathbb{P}_t b_{t+1} \mathbb{P}_t \gamma_{t+1} \\
&= Z_t S_3(t) \frac{c}{n}\frac{1}{2}(1 - X_1(t)/S_2(t) - 1 + X_3(t)/S_2(t)) \\
&= -\frac{1}{2}\frac{c}{n}\frac{S_3(t)}{S_2(t)} Z_t^2 \\
&\leq 0
\end{aligned}$$

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.

2.3. $\mathbb{P}_t Z_{t+1}^2$ is Bounded for by a Constant When Colors are Not Fixed.

We would like to show the expected conditional differences $\mathbb{P}_t Z_{t+1}^2$ cannot grow too quickly in t

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

From this we can conclude

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

2.4. **Balance in the Fixed Color Case.** The previous 2 sections show that a balance mechanism exists in the 2-stacks which keeps their sizes comparable when colored vertices are not removed from the system. We would like to extend this to the case where colored vertices are moved to a fixed color state and removed from the 2 and 3-stack counts.

From Equation 2.1 we know that a vertex in a 2-stack is colored is proportional to its size. A new $\Delta_{t+1}X$ can then be defined as

$$\Delta_{t+1}X = X_{t+1} - X_t = b_{t+1}\varepsilon_{t+1} - X_t/S_2(t)$$

Defining γ_{t+1} as before

$$\begin{aligned}\mathbb{P}_t Z_t \Delta_{t+1} Z &= Z_t(\mathbb{P}_t b_{t+1}\gamma_{t+1} - Z_t/S_2(t)) \\ &= \frac{1}{2}\frac{c}{n}\frac{S_3(t)}{S_2(t)}Z_t^2 - \frac{1}{S_2(t)}Z_t^2\end{aligned}$$

The first term is the same as when the the fixed-color stack was not included and along with this negative term there is another negative term. This implies that the feedback is stronger in the model including the fixed-color stack.

Unfortunately, we cannot recursively bound the second moment of Z_{t+1} when the fixed-color stack is introduced. The problem is that the conditional second moment of Z_{t+1} is a function of $Z_t/S_2(t)$ and since the 2-stack size of zero may occur with positive probability we cannot give an upper bound for the $1/S_2(t)$ term.

References

Pollard, D. (2001). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.