15-750: Graduate Algorithms	April 14, 2017

Lecture 32: Luby's Algorithm for Maximal Independent Set

Lecturer: Gary Miller *Scribe: Yanzhe Yang*

1 Overview

In this lecture, we:

1. Introduce MIS Problem.
2. Introduce parallel MIS algorithms, a simple version and a randomized version specifically.
3. Prove that the randomized Maximal Independent Set (MIS) algorithm terminates in $O(\log n)$ rounds with high probability.

2 Introduction to MIS Problem

Definition 2.1. For an undirected graph G = (V, E), $I \subseteq V$ is independent if no pair share an edge. Say, for all $(u, v) \in E$ either $u \notin I$ and/or $v \notin I$

The independent set I is maximal if $\forall w \in V - I, w \cup I$ is not independent. Another definition is maximum. The independent set I is a maximum independent set if for all independent sets I', $|I| \geq |I'|$. Note that finding a maximum independent set is NP-Hard.

We focus on finding a maximal independent set in this lecture. Denote maximal independent set of graph G as MIS. With a simple greedy algorithm, we can find a MIS in $O(m + n)$ time.

Here is the algorithm:

Algorithm 1 MIS(G)

```
I = ∅, V' = V
while V' ≠ ∅ do
    Choose v ∈ V' (in lexicographically order)
    I = I ∪ v
    V' = V' \ (v ∪ Neighbor(v))
end while
Return I
```

Note that computing the lexicographically first independent set is P-Complete.

3 Parallel MIS Algorithm

3.1 Parallel MIS algorithm

Finding maximal independent set is a classical problem as it plays a key role in parallel algorithms. Recall that algorithms for list ranking and parallel tree compression were of following form:

1. Find a large independent set of nodes
2. Process them (remove them), and recurse on remaining nodes

Our goal is to find a parallel MIS algorithm. Consider the algorithm of this form for $G = (V, E)$:

1. $I = \emptyset, G' = G$
2. While G' is not the empty graph
   (a) Choose a random set of vertices $S \subseteq V$ by selecting each vertex v independently with probability $Pr(v)$. Suppose $Pr(v) = \frac{1}{d_v}$, where $d_v \equiv$ degree of v.
   (b) For every edge $(u, v) \in E(G')$ if both endpoints are in S, then remove the vertex of lower degree from S (break ties). Denote the set after this step as S'.
   (c) Remove S' and Neighbor(S') and all adjacent edges from G'.
   (d) $I = I \bigcup S'$

What we would like this algorithm to do is that we took this big graph, run this algorithm, and hope it has the property that number of nodes in the graph, i.e. the graph size, is geometrically smaller. That is, after each round $|V'| \leq C|V|, C < 1$. (Note: V' is the vertices remained after each round.) Then after $O(\log n)$ rounds, we are done.

Of course that is not true, i.e. $|V'| \leq C|V|, C < 1$ is not always true. One example is a complete bipartite graph $K_{n,\sqrt{n}}$, like Figure 1. On the one side, A, we put down n nodes, and on the other side, B, we put down $\sqrt{n}$ nodes. Then we connect them together.

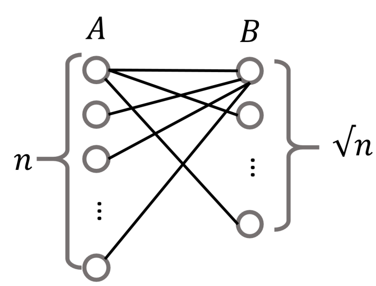

Figure 1: Bad Example

In this example, $\forall v \in A, d(v) = \sqrt{n}$ and $\forall u \in B, d(u) = n$. Then

$$\mathbb{E}(|I \cap A|) \approx n(\frac{1}{\sqrt{n}}) \approx \sqrt{n} \tag{1}$$

$$\mathbb{E}(|I \cap B|) \approx \sqrt{n}(\frac{1}{n}) = \frac{1}{\sqrt{n}} \approx 0 \tag{2}$$

From (2), we don't expect any nodes from B to join the independent set. Note that every node in B is adjacent to every node in A. So if any node in A is in the independent set, all nodes in B will be removed. As for edges, all edges adjacent to B or adjacent to I will be removed. Note that this is all edges in the graph.

$$\begin{aligned}\mathbb{E}(|V'|) &\approx \mathbb{E}(|V|) - \mathbb{E}(|I \cap A|) - |B| \\ &= n + \sqrt{n} - \sqrt{n} - \sqrt{n} \\ &= (1 - o(1))n\end{aligned}$$

So, after a round, $\mathbb{E}(|V'|) \approx |V|$ but $E' = \emptyset$. In this case, the number of vertices doesn't go down much after one round.

3.2 Randomized Parallel MIS

If v is a vertex and S is a set of vertices, define:

$$\begin{aligned}N(v) &= \{u | (u, v) \in E\} = \text{neighbors of } v \\ N(S) &= \bigcup_{u \in S} N(u) = \text{neighbors of } S \\ d(v) &= \text{degree of } v \\ \text{Edges}(S) &= \{(u, v) | u \in S \text{ or } v \in S\}\end{aligned}$$

Here is a randomized algorithm:

Algorithm 2 MIS(G=(V, E))

Add all singletons to I
$I \leftarrow \emptyset$
Each v picks a random value $P(v) \in [0, 1]$
if $P(v) < P(w)$, for all $w \in N(v)$ **then**
    Add v into I
end if
$V' = V \setminus (I \cup N(I))$
$E' = E \setminus \text{Edges}(I \cup N(I))$
Return $I \cup \text{MIS}(G' = (V', E'))$

What we should think about is that we choose a random permutation of vertices, and simply do the first step of simple greedy. We look at all vertices and its neighbor vertices. Then we put those satisfied nodes in the independent set and remove them. This can be regard as one step of parallel algorithm.

What is the probability for a vertex to join in this case? Note that a vertex joins the set when it is the smallest one in $v \cup N(v)$. So a vertex v joints with probability $\frac{1}{d_v+1}$.

3.3 Cost Analysis

In this section, we will prove see two theorem on analyzing this algorithm:

1. Main theorem. $\mathbb{E}(\#\text{edges removed}) \geq |E|/2$.
2. The algorithm terminates in $O(\log n)$ rounds with high probability.

The way to analyze is to define a bunch of variables, and prove things about these variables.

Consider random variables

$$Y = \#\text{ edges removed}$$

$$Y_{uv} = \begin{cases} 1, & \text{if } uv \in E \text{ is removed} \\ 0, & \text{otherwise} \end{cases}$$

Then $2Y = \sum_{uv \in E} Y_{uv}$

3.3.1 Main Theorem

Theorem 3.1. *(Main Theorem)* $2\mathbb{E}[Y] \geq |E|$, *i.e.* $\mathbb{E}(\#\textit{edges removed}) \geq |E|/2$.

Proof. The proof uses the notation in 2.2.

As for critical random variables, $uv \in E$,

$$X_{uv} = \begin{cases} 1, & \text{if } \forall w \in N(u) \cup N(v) \backslash u, P(u) \leq P(w) \\ 0, & \text{otherwise} \end{cases}$$

Note that $X_{uv} = 1 \Rightarrow u \in I$ and v is out.

Claim 3.2. $\mathbb{E}[X_{uv}] = Pr[X_{uv} = 1] \geq 1/(d_u + d_v)$

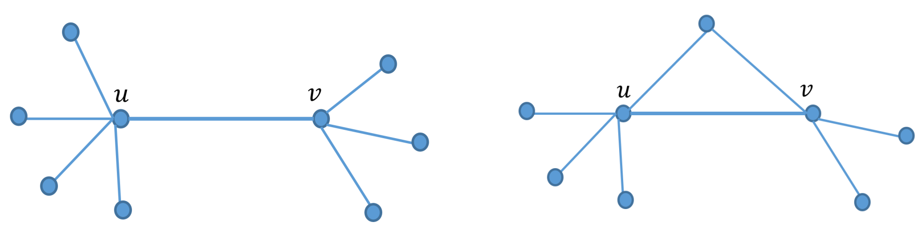

Figure 2: $u \in I$ and v is out.

We can see this as a dart game. The probability that u gets the minimal random nodes in $d(u)+d(v)$ values is $1/(d_u + d_v)$. Like what is shown in Figure 2:Left, $d(u) = 5, d(v) = 4$. $X_{uv=1}$ means u gets the minimal random value among all 9 blue nodes. The probability is 1/9.

That is an inequality because they may share vertices, like what is shown in Figure 2:Right. In that case, the probability that $x_{uv} = 1$ is 1/8.

Claim 3.3. $\sum_{uv \in E} d_v X_{uv} \leq 2Y$

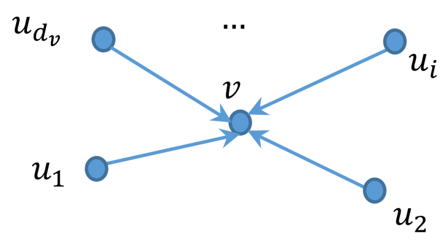

Figure 3: All edges into $v \in V$

Proof. Consider all edges into $v \in V$ (shown in figure 3). Note that at most one $X_{u_i v} = 1$. On the other hand, $\forall u_i \in N(v)$, if $X_{u_i v} = 1$, then $Y_{u_1 v} = Y_{u_2 v} = ... = Y_{u_{d_v} v} = 1$. That is, if any $X_{u_i v} = 1$, then every $Y_{u_j v} = 1, j \in [1, d_v]$.
Thus,

$$\sum_{u \in N(v)} d_v X_{uv} \leq \sum_{u \in N(v)} Y_{uv}$$

$$\text{and therefore} \quad \sum_{uv \in E} d_v X_{uv} \leq \sum_{u \in E} Y_{uv} = 2Y. \qquad \square$$

(back to proof of Main Theorem)
From Claim 3.3, we obtain that

$$\begin{aligned}
2\mathbb{E}[Y] &\geq \mathbb{E}[\sum_{uv \in E} d_v X_{uv} + d_u X_{vu}] && (3) \\
&= \sum_{uv \in E} d_v \mathbb{E}[X_{uv}] + d_u \mathbb{E}[X_{vu}] \\
&\geq \sum_{uv \in E} \frac{d_v}{d_u + d_v} + \frac{d_u}{d_u + d_v} && \text{(using Claim 3.2)} \\
&= \sum_{uv \in E} 1 \\
&= |E| && (4)
\end{aligned}$$

That is, $|E| \leq 2\mathbb{E}[Y]$ $\square$

3.3.2 How many rounds needed?

Theorem 3.4. *RandomParallelMIS terminates in $O(\log n)$ rounds with high probability.*

Proof. Consider random variables

$$Z_i = \begin{cases} 1, & \text{if round } i \text{ deletes at least } 1/3 \text{ remaining edges} \\ 0, & \text{otherwise} \end{cases}$$

Denote the number of remaining edges as m.

$$\begin{aligned}
Pr[Z_i = 0] &= Pr[Y \leq \frac{m}{3}] \\
&\leq \frac{\mathbb{E}[m - Y]}{m - \frac{m}{3}} && \text{(using Reverse Markovs Inequality)} \\
&= \frac{m - E[Y]}{\frac{2}{3}m} \\
&\leq \frac{\frac{1}{2}m}{\frac{2}{3}m} && \text{(using Theorem 3.1)} \\
&= \frac{3}{4}
\end{aligned} \tag{5}$$

So,

$$\begin{aligned}
Pr[Z_i = 1] &= 1 - Pr[Z_i = 0] \\
&\geq \frac{1}{4}
\end{aligned} \tag{6}$$

(Note: In Gary's lecture notes, the claim is $Pr[Z_i = 1] \geq \frac{2}{3}$ instead. But that doesn't affect the following proof as it is only a constant.)

Then after $4c \log n$ rounds,

$$\mathbb{E}[\#\text{rounds that satisfied } Z_i = 1] \geq c \log n \tag{7}$$

Because we choose different priorities uniformly at random for each round, the rounds are independent. Then we can apply Chernoff Bounds:

$$\begin{aligned}
Pr[\#\text{rounds that satisfied } Z_i = 1 \leq (1-\delta)c\log n] &\leq e^{\frac{\delta^2 c \log n}{3}} \\
&\leq^{-c' \log n} \\
&\leq \frac{1}{n^{c'}}
\end{aligned}$$

In conclusion, the randomized parallel MIS algorithm terminates in $O(\log n)$ rounds with high probability. □

References

[1] Lecture Notes for Randomized Algorithms
http://www.cc.gatech.edu/∼vigoda/RandAlgs/MIS.pdf

[2] Chernoff Bound
https://en.wikipedia.org/wiki/Chernoff_bound

[3] Luby's Algorithm
http://www.di.univaq.it/∼proietti/slide_algdist2015/Luby's Algorithm.pdf

[4] Luby, Michael. "A simple parallel algorithm for the maximal independent set problem." SIAM journal on computing 15.4 (1986): 1036-1053.