Lecture Notes CS:5350 Luby's Randomised & Distributed Algorithm for MIS: Analysis
Lecture 12: Oct 3, 2019
Scribe: Krishnamoorthy V. Iyer

In this lecture, we focus on a simplified analysis of Luby's algorithm that shows that an expected running time of $O(\log \Delta \cdot \log n)$ suffices to obtain an MIS, where $\Delta \stackrel{\text{def}}{=} \max_{v \in V} deg(v)$ and $n = |V|$ (standard notation).

1 Overall Strategy of the (Simplified) Analysis

The heart of the simplified analysis is the following lemma. (For the proof, skip ahead to section 2).

Lemma: Let v be a vertex s.t. $deg(v) \in [\frac{\Delta}{2}, \Delta]$. Then the probability that v is deactivated at the end of the first iteration is $\geq \frac{1}{2}\left(1 - \frac{1}{e^{\frac{1}{4}}}\right)$, a constant.

- Throughout the course of the proof of the lemma, we will refer to nodes satisfying the above condition as "high" degree (for that round) nodes.

- Since the probability that a "high"-degree node is deactivated at the end of one iteration is a constant, if we repeat the process for $\log n$ iterations, the probability that the node will be deactivated is of the form $1 - \frac{1}{n^c}$, where $c \geq 1$ is a constant. If n is "large", this means that **after $\log n$ iterations, with high probability (w.h.p. henceforth), there will be no nodes with "high" degree i.e., degree in the range** $[\frac{\Delta}{2}, \Delta]$. Call this the first "round" (i.e., consisting of $\log n$ iterations).

- We can now focus on the remaining vertices, all of whom will, w.h.p., have degrees $\leq \frac{\Delta}{2}$. We repeat the process above, treating vertices with degrees in the range $[\frac{\Delta}{4}, \frac{\Delta}{2}]$ as "high"-degree vertices. So after a further $\log n$ iterations (equivalently, at the end of 3 rounds), the maximum degree will (w.h.p) be $\leq \frac{\Delta}{4} = \frac{\Delta}{2^2}$.

- At the end of k rounds – equivalently $k \log n$ iterations – the maximum degree will have dropped (w.h.p) to $\frac{\Delta}{2^k}$.

- After $\log \Delta \cdot \log n$ iterations, the maximum degree will have dropped to less than 1. But since the degree must be an integer, this means that the degrees of all nodes will be, w.h.p., 0. This gives us the required MIS.

2 Proof of the Lemma

Preliminaries: Crucial to understanding the proof are the following basic rules of probability. In the following, the symbols $X, Y, \overline{X}, X_1, X_2, \ldots, X_k$ represent events:

- Rule 0: Probability of (an event) X = 1 - Probability of the complementary event of X:

$$P(X) = 1 - P(\overline{X})$$

  where $\overline{X}$ represents the non-occurrence of X i.e., the complementary event.

- Rule 1: Chain rule for probability :

$$P(X \wedge Y) = P(X) \cdot P(Y|X)$$

- Rule 2: The probability of the simultaneous occurrence of a collection of *independent* events $X_1, X_2, \ldots, X_k$ is the product of their probabilities:

$$P(X_1 \wedge X_2 \wedge \cdots \wedge X_k) = \prod_{i=1}^{k} P(X_i)$$

  The LHS above is the probability of the conjunction of the independent events $X_1, X_2, \ldots, X_k$.

- Rule 3: The *union* bound, which upper bounds the probability of the disjunction of events $X_1, \ldots, X_k$:

$$P(X_1 \vee X_2 \vee \cdots \vee X_k) \leq \sum_{i=1}^{k} P(X_i)$$

  The LHS of the above is often written as: $P(\bigcup_{i=1}^{k} X_i)$ where the '$\bigcup$' stands for 'union', and hence this is known as the 'Union Bound'.

2.1 Proof of the Lemma

Now we turn to the proof of the lemma. (Finally!)

A vertex v can be deactivated – call this event $\mathcal{E}$ – if either of the following events occurs:

- Either v joins the MIS. Call this event $\mathcal{E}_1$.
- Some neighbour of v joins the MIS. Call this event $\mathcal{E}_2$.

Now, we have:

$$P(\mathcal{E}) = P(\mathcal{E}_1 \cup \mathcal{E}_2) \geq P(\mathcal{E}_2)$$

Our proof strategy for the lemma consists of obtaining a lower bound on $P(\mathcal{E}_2)$. This will automatically give us the desired lower bound on $P(\mathcal{E})$.

For the event $\mathcal{E}_2$ i.e., "some neighbor of v joins the MIS" to occur, *both* of the following events must occur, to wit:

- Some neighbor of v is marked. Call this event $\mathcal{A}$.
- Some *marked* neighbor of v survives the tie-breaking competition with *its marked neighbors*. Call this event $\mathcal{B}$.

We have:

$$P(\mathcal{E}_2) \overset{\text{def of } \mathcal{E}_2}{=} P(\mathcal{A} \wedge \mathcal{B}) \overset{\text{Rule 1}}{=} P(\mathcal{A}) \cdot P(\mathcal{B}|\mathcal{A})$$

To obtain a lower bound on $P(\mathcal{E}_2)$, we will lower bound each of $P(\mathcal{A})$ and $P(\mathcal{B}|\mathcal{A})$.

2.1.1 Lower Bounding $P(\mathcal{A})$:

$$\begin{aligned} P(\mathcal{A}) &= P(\text{some neighbor of } v \text{ is marked}) \\ &\overset{\text{Rule 0}}{=} 1 - P(\text{no neighbor of } v \text{ is marked}) \\ &= 1 - P(\overline{\mathcal{A}}) \end{aligned}$$

Now:

$$P(\overline{\mathcal{A}}) = P(\text{no neighbor of } v \text{ is marked}) = P(\bigwedge_{\omega \in N(v)} (\omega \text{ not marked }))$$

where $N(v)$ is the subset of vertices that are neighbors of v i.e., are adjacent to v.

But whether a vertex is marked or not is independent of whether or not other vertices are marked. In other words, *the events "being marked" – or not – are independent.* Hence we can apply Rule 2. Thus, we have:

$$P(\bigwedge_{\omega \in N(v)} (\omega \text{ not marked })) = \prod_{\omega \in N(v)} P(\omega \text{ not marked})$$

So:

$$\begin{aligned} P(\overline{\mathcal{A}}) &= \prod_{\omega \in N(v)} P(\omega \text{ not marked}) \\ &\overset{(a)}{=} \prod_{\omega \in N(v)} [1 - P(\omega \text{ marked})] \\ &\overset{(b)}{=} \prod_{\omega \in N(v)} [1 - \frac{1}{2 \cdot deg(\omega)}] \\ &\overset{(c)}{\leq} \prod_{\omega \in N(v)} [1 - \frac{1}{2 \cdot \Delta}] \\ &\overset{(d)}{=} [1 - \frac{1}{2 \cdot \Delta}]^{deg(v)} \\ &\overset{(e)}{\leq} [1 - \frac{1}{2 \cdot \Delta}]^{\frac{\Delta}{2}}. \end{aligned}$$

in (a), we have applied Rule 0 to each individual factor in the product,
in (b), we have used probability of ω being marked,
in (the inequality) (c), we have used $deg(\omega) \leq \Delta$. [Check this!],
in (d), we use that all product terms are identical, and there are $deg(v)$ factors,
and the inequality (e) follows from the observation that $deg(v) \geq \Delta/2$ (by choice) and if $\alpha \in [0, 1]$, then $\alpha^p \leq \alpha^q$ if $p \geq q$. (Whew!)

We now use $1 + x \leq e^x\ \forall x$ with $x \stackrel{\text{set}}{=} -\frac{1}{2\cdot\Delta}$ to obtain:

$$P(\overline{\mathcal{A}}) \leq \left[e^{-1/2\cdot\Delta}\right]^{\Delta/2} = e^{-\frac{1}{4}}.$$

Finally, we get the lower bound desired:

$$P(\mathcal{A}) = 1 - P(\overline{\mathcal{A}}) \geq 1 - \frac{1}{e^{1/4}}.$$

Note: One question that was raised in class was "Why do we attempt to bound $P(\mathcal{A})$ for high-degree nodes only? Why do we ignore low-degree nodes?". The answer is that we cannot obtain a constant lower bounding value for $P(\mathcal{A})$ for low-degree nodes. This can be seen easily in the extreme case where we consider a node of degree = 1. Then the product will have only one factor, and bounding is not possible. Another response to the query is to study [9, MIS II, p.75] or [3, Sec 12.3, p. 341]

2.1.2 Lower Bounding $P(\mathcal{B}|\mathcal{A})$:

Consider $P(\mathcal{B}|\mathcal{A})$:

$$= P(\text{some marked nghbr of } v \text{ survives tie-breaking}|\text{some nghbr of } v \text{ is marked})$$

Consider the marked neighbor of v with highest degree. Call it w.[1] Subtlety: How do we know such a marked neighor even exists? There must exist such a neighbor i.e., the event is well-defined because we have conditioned on there being neighbors of v that have been marked. Now:

$$\begin{aligned}
&P(\text{some mrkd nghbr of } v \text{ survives tie-breaking}|\text{some nghbr of } v \text{ mrkd}) \\
&= P(\bigcup_u \text{mrkd nghbr } u \text{ survives tie-breaking}|\text{some nghbr of } v \text{ mrkd}) \\
&\geq P(\text{mrkd nghbr } w \text{ of highest deg survives tie-breaking}|\text{some nghbr of } v \text{ mrkd})
\end{aligned}$$

To compute a lower bound on the quantity on the right of the inequality above, we proceed as follows. We partition the neighborhood of w as: $N(w) = [N(w) \cap N(v)] \sqcup [N(w) \setminus N(v)]$ where the symbol '$\sqcup$' means 'disjoint union'.

Only those neighbors of w that are not neighbors of v i.e. $N(w) \setminus N(v)$ can defeat w in the tie-break. Why? Because w, by virtue of being the highest degree node among v's neighbors, will win the

[1] Initially, ignore the possibility that there may be more than one highest degree neighbor – ties can be handled as in the algorithm by means of IDs, for example, by giving priority to the node with higher ID.

tie-break among its marked neighbors who are also neighbors of v i.e., in $N(w) \cap N(v)$. Thus, we need only focus our attention on $N(w) \setminus N(v)$.

Only a neighbor of w in $N(w) \setminus N(v)$ that is marked and has higher degree will defeat w in the tie-break. Conversely, for w to survive the tie-break, none of its marked neighbors in $N(w) \setminus N(v)$ must have higher degree than w. So we now have:

$$\begin{aligned} P(\mathcal{B}|\mathcal{A}) &\geq P(\text{no nghbr of } w \in N(w) \setminus N(v) \text{ of higher degree is marked}) \\ &= 1 - P(\text{some nghbr of } w \in N(w) \setminus N(v) \text{ of higher degree is marked}) \\ &= 1 - P(\bigcup_{u \in N(w) \setminus N(v): deg(u) > deg(w)} u \text{ is marked}) \end{aligned}$$

We use the union bound (Rule 3) to upper bound:

$$\begin{aligned} P(\bigcup_{u \in N(w) \setminus N(v): deg(u) > deg(w)} u \text{ is marked}) &\leq \sum_{u \in N(w) \setminus N(v): deg(u) > deg(w)} P(\{u \text{ is marked}\}) \\ &= \sum_{u \in N(w) \setminus N(v): deg(u) > deg(w)} \frac{1}{2 \cdot deg(u)} \\ &\leq \sum_{u \in N(w) \setminus N(v): deg(u) > deg(w)} \frac{1}{2 \cdot deg(w)} \\ &\leq deg(w) \cdot \frac{1}{2deg(w)} \\ &= \frac{1}{2} \end{aligned}$$

where the second-to-last inequality follows because $deg(u) > deg(w)$ and the last inequality holds because the number of terms in the sum is $\leq deg(w)$.

Substituting, we obtain:

$$P(\mathcal{B}|\mathcal{A}) \geq 1 - \frac{1}{2} = \frac{1}{2}$$

2.1.3 Putting all the pieces together:

The product of the lower bounds on $P(\mathcal{A})$ and $P(\mathcal{B}|\mathcal{A})$ together gives $\frac{1}{2} \cdot \left(1 - e^{-1/4}\right)$.

Finally, noting that:

$$P(\mathcal{E}) \geq P(\mathcal{E}_2) = P(\mathcal{A}) \cdot P(\mathcal{B}|\mathcal{A}) \geq \frac{1}{2} \cdot \left(1 - e^{-1/4}\right)$$

gives the required lower bound.

References

[1] https://en.wikipedia.org/wiki/Maximal_independent_set

[2] Luby, "A Simple Parallel Algorithm for the Maximal Independent Set Problem"

[3] Motwani, Raghavan, "Randomized Algorithms"

[4] Barenboim, Elkin, Pettie, Schneider, "The Locality of Distributed Symmetry Breaking"

[5] Ghaffari, "An Improved Distributed Algorithm for Maximal Independent Set"

[6] Daum, Ghaffari, Gilbert, Kuhn, Newport, "Maximal Independent Sets in Multichannel Radio Networks" Available at: http://tr.informatik.uni-freiburg.de/reports/report275/report00275.pdf

[7] Basagni, "Finding a Maximal Weighted Independent Set in Wireless Networks"

[8] Hsin-Hao Su, "Algorithms for Fundamental Problems in Computer Networks"

[9] Gopal Pandurangan "Distributed Network Algorithms"