Brief Announcement: Local Independent Set Approximation

Marijke H. L. Bodlaender* Magnús M. Halldórsson* Christian Konrad*
Fabian Kuhn †

Abstract

We show that the first phase of the Linial-Saks network decomposition algorithm gives a randomized distributed $O(n^{\epsilon})$-approximation algorithm for the maximum independent set problem that operates in $O(\frac{1}{\epsilon})$ rounds, and we give a matching lower bound that holds even for bipartite graphs.

1 Local Independent Set Computation

Local Algorithms. A central question in distributed computing is what can be computed locally, i.e., in O(1) communication rounds. Local algorithms have many desirable properties, e.g., they can recover quickly from failures, changes in the topology, or changes in the input [11].

Linial [7] showed that computing a maximal independent set on the ring requires $\Omega(\log^* n)$ communication rounds, where n is the number of nodes in the network, thus dashing the hope for local maximal independent set algorithms. In this work, we drop the maximality requirement of independent sets and aim at maximizing their sizes instead. Since the maximum independent set problem is NP-hard and is hard to approximate within a factor $n^{1-\epsilon}$, for any $\epsilon > 0$ [5], we cannot expect to obtain polynomial-time local algorithms.

Our Results. We give a $O(n^{\epsilon})$-approximation, $O(\frac{1}{\epsilon})$ round distributed algorithm for the maximum independent set problem in general graphs. In general graphs, exponential time computations are unavoidable due to the hardness of the problem, but in graph classes that allow the computation of a maximum independent set (or a good approximate thereof) efficiently such as bipartite graph, polynomial time suffices. Furthermore, we show that our algorithm is best possible, even for bipartite graphs.

Our algorithm employs the first phase of the Linial-Saks network decomposition algorithm [8]. A network decomposition is a partition of the vertex set V of a graph $G = (V, E)$ into connected subsets of vertices $(V_i)_i$, denoted clusters, each of bounded diameter, and a coloring of the clusters using a limited number of colors, such that adjacent clusters have different colors. Various network decomposition methods with different characteristics are known [1, 8, 10, 2, 3] and are employed as building blocks in a multitude of distributed algorithms.

Related Work. The work of Barenboim [2] and the follow-up work by Barenboim et al. [3] are closest to our work: They give local distributed algorithms for the minimum vertex coloring problem with approximation factors $\tilde{O}(\sqrt{n})$ (in [2]) and $\tilde{O}(n^{\epsilon})$ (in [3]), where nodes run exponential time algorithms. While our work makes use of the network decomposition of Linial and Saks [8], new network decomposition methods better suited to coloring problems are designed in [2] and [3].

Lower bounds for local algorithms are usually proved by pointing out pairs of graphs that are locally indistinguishable, but their global properties are quite different. In the context of independent sets/colorings, this technique has previously been applied for example for computing maximal independent sets [6] and coloring trees [7].

*ICE-TCS, School of Computer Science, Reykjavik University, Iceland. {marijke12, mmh, christiank}@ru.is
†Department of Computer Science, University of Freiburg, Germany. kuhn@cs.uni-freiburg.de

2 Algorithm

In each iteration of the Linial-Saks [8] network decomposition algorithm, a partial assignment of vertices into clusters is computed. The algorithm requires an integer parameter B and a probability p. Parameter B determines the maximum cluster diameter and thus the number of rounds required by the algorithm. Parameter p influences the average cluster diameter, which in turn influences the number of nodes that join a cluster (cf. Lemma 1). Each cluster consists of a cluster leader l and a subset of vertices $C_l \subseteq \Gamma_G^B[l]$, where $\Gamma_G^r[l]$ denotes the set of nodes in G at distance at most r from l including l (i.e., the inclusive r-neighborhood). Importantly, vertices of different clusters are not adjacent. Each iteration requires $2B$ rounds.

The key property that ensures the approximation factor of our algorithm is as follows:

Lemma 1 (Linial, Saks [8]). *For every v, the probability that v joins some cluster is at least $p(1-p^B)^n$.*

Our maximum independent set algorithm proceeds as follows: We set $B = \lceil \frac{1}{\epsilon} \rceil$ and $p = n^{-\frac{1}{B}}$, for the given $\epsilon > 0$, and run the first iteration of the Linial-Saks decomposition in $2B = 2\lceil \frac{1}{\epsilon} \rceil$ rounds, obtaining a set L of cluster leaders. At the same time, nodes collect the topologies of their B-neighborhoods. Each cluster leader $l \in L$ computes a maximum independent set I_l in the subgraph induced by C_l (in exponential time in general graphs). Finally, cluster leaders notify nodes of C_l in at most B rounds whether they are included in the independent set.

Theorem 1. *For every $\epsilon > 0$, there is a randomized distributed algorithm for the maximum independent set problem with expected approximation ratio $(2e \cdot n^{\frac{1}{\epsilon}})$ that runs in $3\lceil \frac{1}{\epsilon} \rceil$ rounds.*

Proof. First, recall that clusters are disconnected in the input graph, and thus the output $I = \cup_{l \in L} I_l$ is an independent set, establishing correctness of the algorithm.

Let $I^* \subseteq V$ denote a maximum independent set in the input graph, and let $C = \cup_{l \in L} C_l$ denote the set of nodes that are contained in some cluster. Then:

$$\begin{aligned} \mathbb{E}|I^* \cap C| &= \sum_{i \in I^*} \mathbb{P}\left[i \in C\right] \geq \sum_{i \in I^*} n^{-B^{-1}}(1 - n^{-B^{-1} \cdot B})^n \\ &\geq \sum_{i \in I^*} \frac{1}{n^\epsilon}(1 - \frac{1}{n})^n \geq |I^*| \frac{1}{n^\epsilon} \frac{1}{2e}, \end{aligned} \tag{1}$$

applying Lemma 1. Then, due to the local optimality of I_l,

$$|I| = \sum_{l \in L} |I_l| \geq \sum_{l \in L} |I^* \cap C_l| = |I^* \cap C|.$$

Taking expectations and applying Inequality 1 yields the approximation factor.

Running the first iteration of Linial-Saks and collecting the topology of local B-neighborhoods requires $2B$ rounds, and notifying nodes whether they are part of the independent set requires additional B rounds. □

Our algorithm can also give a $(1+\epsilon)$-approximation that runs in $O(\log(\frac{n}{\epsilon})/\epsilon)$ rounds by employing parameters $B = c\log(\frac{n}{\epsilon})/\epsilon$, for a small enough constant c, and $p = 1 - \frac{\epsilon}{2}$,

3 Lower Bound

Let $H_1 = (V_1, E_1), H_2 = (V_2, E_2)$ be d-regular graphs with girths at least g, and suppose that H_1 is bipartite. Let $n_i = |V_i|$, for $i \in \{1, 2\}$. Furthermore, for $i \in \{1, 2\}$, we label the vertices of H_i such that each vertex $v \in V_i$ receives a unique label (or ID) $\mathcal{L}_i(v)$, where the labeling function $\mathcal{L}_i$ is chosen uniformly at random from the set of injections from V_i to $\{1, \ldots, \max\{n_1, n_2\}\}$.

In the following, we denote the size of a maximum independent set in a graph F by $\alpha(F)$.

Lemma 2. *Let H_1 and H_2 be as defined above. Every possibly randomized $\lceil g/2 \rceil - 1$ rounds distributed algorithm for the bipartite maximum independent set problem computes an independent set of expected size at most $\alpha(H_2)\frac{n_1}{n_2}$ on H_1, where the expectation is taken over the labelings of H_1 and the random coin flips.*

Proof. Let $\mathcal{A}$ be a randomized distributed algorithm for the bipartite maximum independent set problem, and denote by I the independent set computed by $\mathcal{A}$ on graph H_1. For a vertex $v \in V_1$, let $p_v = \mathbb{P}[v \in I]$, where the probability is taken over the random coin flips of the algorithm and the labelings.

Let $g' = \lceil g/2 \rceil - 1$. Since algorithm $\mathcal{A}$ runs in g' rounds, the outcome of $\mathcal{A}$ when executed on vertex v depends on the structure and labels of the inclusive g'-neighborhood $\Gamma^{g'}_{H_1}[v]$, and on the random bits used by nodes of $\Gamma^{g'}_{H_1}[v]$. Since the girth of H_1 is g, each local neighborhood $\Gamma^{g'}_{H_1}[u]$ is isomorphic to a d-ary tree of depth g' rooted at u. Further, since labels are assigned uniformly at random, and random bits are uniform, the values p_u and p_v are identical, for every $u, v \in V_1$. Denote this value by p (i.e., $p = p_v$ for an arbitrary $v \in V_1$). Then, $\mathbb{E}|I| = \sum_{v \in V_1} p_v = n_1 p$.

We consider now the performance of $\mathcal{A}$ on graph H_2. Similar as in H_1, every local neighborhood $\Gamma^{g'}_{H_2}[u]$ is isomorphic to a d-ary tree of depth g' rooted at u. Furthermore, it can easily be seen that the output I' of $\mathcal{A}$ on H_2 constitutes an independent set. Similar considerations as before show that the expected size of I' computed by $\mathcal{A}$ on graph H_2 is $\mathbb{E}|I'| = n_2 p$, which is in turn is bounded from above by $\alpha(H_2)$. Thus, $p \leq \alpha(H_2)/n_2$, and hence $\mathbb{E}|I| \leq \alpha(H_2)\frac{n_1}{n_2}$. □

Next, in order to obtain our lower bound result, we employ the Ramanujan graphs of Lubotzky, Phillips, and Sarnak [9]. For p, q distinct primes congruent to 1 mod 4, there are $(p+1)$-regular graphs $X^{p,q}$ on n vertices with girth $\Omega(\log_p(q))$ that satisfy: If the Legendre symbol $\left(\frac{p}{q}\right) = -1$ then $X^{p,q}$ is bipartite and $n = q(q^2-1)$, while if $\left(\frac{p}{q}\right) = 1$ then $\alpha(X^{p,q}) = \mathrm{O}(\frac{n}{\sqrt{p}})$ and $n = q(q^2-1)/2$. Equipped with these graphs, we are ready to prove our lower bound result.

Theorem 2. *For every $\epsilon > 0$, there is an infinite family of bipartite graphs $\mathcal{G}$ such that every possibly randomized $\frac{1}{\epsilon}$-rounds distributed algorithm for the bipartite maximum independent set problem has approximation factor $n^{\Omega(\epsilon)}$ on every graph $G \in \mathcal{G}$, where n is the number of vertices of G.*

Proof. Let p, q_1, q_2 be distinct primes congruent to $1 \mod 4$ such that $\left(\frac{p}{q_1}\right) = -1$, $\left(\frac{p}{q_2}\right) = 1$, and $q_1, q_2 \in p^{\Theta(\frac{1}{\epsilon})}$. Let $H_1 = X^{p,q_1}$, $H_2 = X^{p,q_2}$, and for $i \in \{1,2\}$, let n_i be the number of vertices of graph H_i. Then, $n_1, n_2 \in p^{\Theta(\frac{1}{\epsilon})}$. H_1 and H_2 have girths $\Omega(\frac{1}{\epsilon})$ and are $(p+1)$-regular. Furthermore, H_1 is bipartite and $\alpha(H_2) = \mathrm{O}(n_2^{1-\Theta(\epsilon)})$.

By Lemma 2, for a small enough C, every possibly randomized $C\frac{1}{\epsilon}$-round distributed bipartite maximum independent set algorithm computes an independent set of size at most $\alpha(H_2)\frac{n_1}{n_2} = n_1/n_2^{\Theta(\epsilon)} = n_1^{1-\Theta(\epsilon)}$ on H_1. Since H_1 is bipartite, it contains an independent set of size at least $n_1/2$, implying approximation ratio of $\Omega(n_1^{\Theta(\epsilon)})$. □

4 Conclusion

Since our results provide tight bounds (even for bipartite graphs), an interesting question is to determine graph classes for which local algorithms with sub-polynomial approximations can be obtained. Progress has been made for example for *polynomially bounded-independence graphs* such as unit disc graphs, where poly-logarithmic approximation ratios can be achieved in a single communication round [4]. The same paper [4] implies a constant factor approximation for planar graphs, and, more generally, an $\mathrm{O}(d)$-approximation for graphs with average degree d.

5 Acknowledgements

Marijke H. L. Bodlaender, Magnús M. Halldórsson and Christian Konrad are supported by Icelandic Research Fund grants 120032011 and 152679-051. Fabian Kuhn is supported by ERC Grant No. 336495 (ACDC).

References

[1] Baruch Awerbuch, Andrew V. Goldberg, Michael Luby, and Serge A. Plotkin. Network decomposition and locality in distributed computation. In *30th Annual Symposium on Foundations of Computer Science, Research Triangle Park, North Carolina, USA, 30 October - 1 November 1989*, pages 364–369, 1989.

[2] L. Barenboim. On the locality of some NP-complete problems. In *Automata, Languages, and Programming - 39th International Colloquium, ICALP 2012, Warwick, UK, July 9-13, 2012, Proceedings, Part II*, pages 403–415, 2012.

[3] L. Barenboim, M. Elkin, and C. Gavoille. A fast network-decomposition algorithm and its applications to constant-time distributed computation - (extended abstract). In *Structural Information and Communication Complexity - 22nd International Colloquium, SIROCCO 2015, Montserrat, Spain, July 14-16, 2015, Post-Proceedings*, pages 209–223, 2015.

[4] Magnús M. Halldórsson and Christian Konrad. Distributed large independent sets in one round on bounded-independence graphs. In Yoram Moses, editor, *Distributed Computing*, volume 9363 of *Lecture Notes in Computer Science*, pages 559–572. Springer Berlin Heidelberg, 2015.

[5] Johan Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Mathematica*, 182:105–142, 1999.

[6] Fabian Kuhn, Thomas Moscibroda, and Roger Wattenhofer. What Cannot Be Computed Locally! In *23rd ACM Symposium on the Principles of Distributed Computing (PODC), St. Johns, Newfoundland, Canada*, July 2004.

[7] N. Linial. Locality in distributed graph algorithms. *SIAM J. Comput.*, 21(1):193–201, 1992.

[8] N. Linial and M. E. Saks. Low diameter graph decompositions. *Combinatorica*, 13(4):441–454, 1993.

[9] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, September 1988.

[10] Alessandro Panconesi and Aravind Srinivasan. On the complexity of distributed network decomposition. *J. Algorithms*, 20(2), March 1996.

[11] J. Suomela. Survey of local algorithms. *ACM Comput. Surv.*, 45(2):24:1–24:40, March 2013.