Simple and Local Independent Set Approximation

Ravi B. Boppana
Department of Mathematics
MIT, USA
rboppana@mit.edu

Magnús M. Halldórsson*
School of Computer Science
Reykjavik University, Iceland
mmh@ru.is

Dror Rawitz†
Faculty of Engineering
Bar-Ilan University, Israel
dror.rawitz@biu.ac.il

August 9, 2018

Abstract

We bound the performance guarantees that follow from Turán-like bounds for unweighted and weighted independent sets in bounded-degree graphs. In particular, a randomized approach of Boppana forms a simple 1-round distributed algorithm, as well as a streaming and preemptive online algorithm. We show it gives a tight $(\Delta + 1)/2$-approximation in unweighted graphs of maximum degree Δ, which is best possible for 1-round distributed algorithms. For weighted graphs, it gives only a Δ-approximation, but a simple modification results in an asymptotic expected 0.529Δ-approximation. This compares with a recent, more complex Δ-approximation [5], which holds deterministically.

1 Introduction

Independent sets are among the most fundamental graph structures. A classic result of Turán [20] says that every graph $G = (V, E)$ contains an independent set of size at least $\textsc{Turán}(G) \doteq n/(\overline{d}+1)$, where $n = |V|$ is the number of vertices and $\overline{d} = 2|E|/n$ is the average degree. Turán's bound is tight for regular graphs, but for non-regular graphs an improved bound was given independently by Caro [9] and Wei [21]:

$$\alpha(G) \geq \textsc{CaroWei}(G) \doteq \sum_{v \in V} \frac{1}{d(v) + 1} , \tag{1}$$

where $\alpha(G)$ is the cardinality of a maximum independent set in G and $d(v)$ is the degree of vertex $v \in V$.

*Supported by grants nos. 152679-05 and 174484-05 from the Icelandic Research Fund.
†Supported by the Israel Science Foundation (grant no. 497/14).

There are numerous proofs of the Caro-Wei bound, some involving simple greedy algorithms. Arguably the simplest argument known is a probabilistic one:

Uniformly randomly permute the vertices, and output the set of vertices that precede all their neighbors in the permutation. (2)

Each node v precedes its neighbors with probability $1/(d(v)+1)$, so by linearity of expectation the size of the set output matches exactly $\textsc{CaroWei}(G)$. This argument, which first appeared in the book of Alon and Spencer [2], is due to Boppana [8]. It clearly leads to a very simple local decision rule once the permutation is selected.

An alternative formulation of the algorithm is practical in certain contexts.

Each vertex v picks a random real number x_v from $[0,1]$. The vertex joins the independent set if its random number is larger than that of its neighbors. (3)

It suffices to select the numbers with precision $1/n^3$, for which collisions are very unlikely.

This leads to a fully *1-local* algorithm, in which each node decides whether to join the independent set after a single round of communication with its neighbors. The same $O(\log n)$ bits a node transmits go to all of its neighbors, which matches the Broadcast-CONGEST model of distributed algorithms. Furthermore, it is asynchronous. This is just about the simplest distributed algorithm one could hope for.

The simplicity of the approach also allows for other applications. The basic algorithm works well with edge streams, storing only the permutation and the current solution as a bit-vector. The storage can be reduced with an ϵ-min-wise permutation, at a small cost in performance. This can also be viewed as a preemptive online algorithm, where edges can cause nodes to be kicked out of the solution but never reenter.

Our contribution. The main purpose of this essay is to analyze the performance guarantees of Boppana's algorithm on graphs of maximum degree Δ. We show that it achieves a tight $(\Delta+1)/2$-approximation, which then also gives a bound on the fidelity of the Caro-Wei bound. In terms of the average degree $\overline{d}$, the performance is at most $(\overline{d}+2)/1.657$. We also show that the Turán bound has strictly worse performance than the Caro-Wei bound, but asymptotically the same for bounded-degree graphs or $(\Delta+1)/2+1/(8\Delta)$.

We then address the case of weighted graphs, and find that unchanged Boppana's algorithm gives only a $(\Delta+1)$-approximation. However, a slight modification yields an improved approximation which asymptotically approaches $0.529(\Delta+1)$.

1.1 Related work

Turán [20] showed that $\alpha(G) \geq \textsc{Turán}(G)$. Caro [9] and Wei [21] independently showed (in unpublished technical reports) that $\alpha(G) \geq \textsc{CaroWei}(G)$. The bound can also be seen to follow from an earlier work of Erdős [13], who showed that the bound is tight only for disjoint collections of cliques. Observe that $\textsc{CaroWei}(G) \geq \textsc{Turán}(G)$, for every graph G.

The min-degree greedy algorithm iteratively adds a minimum-degree node to the graph, removes it and its neighbors and repeats. It achieves the Caro-Wei bound [21] (see also [13]). Griggs [14] (see also Chvátal and McDiarmid [11]) showed that the max-degree greedy algorithm also attains the Caro-Wei bound, where the algorithm iteratively removes the vertex of maximum degree until the graph is an independent set. Sakai et al. [19] analyzed three greedy algorithms for weighted independent sets and showed them to achieve certain absolute bounds as well as a $(\Delta+1)$-approximation.

The best sequential approximation known is $\tilde{O}(\Delta/\log^2\Delta)$,[1] by Bansal et al. [4], which uses semi-definite programming. This matches the inapproximability result known, up to doubly-logarithmic factors, that holds assuming the Unique Games Conjecture [3]. The problem is known to be NP-hard to approximate within an $O(\Delta/\log^4\Delta)$ factor [10]. For small values of Δ, a $(\Delta+3)/5$-approximation [6] is achievable combinatorially, but requires extensive local search. As for simple greedy algorithms, it was shown in [16] that the performance guarantee of the min-degree greedy algorithm is $(\Delta+2)/3$, and also pointed out that the max-degree algorithm attains no better than a $(\Delta+1)/2$ ratio.

Most works on distributed algorithms have focused on finding maximal independent sets, rather than optimizing their size. Boppana's algorithm corresponds to the first of $O(\log n)$ rounds of Luby's maximal independent set algorithm (see also Alon et al. [?]). As for approximations, $n^{\Theta(1/k)}$-approximation is achievable and best possible for local algorithms running in k rounds [7], where the upper bound assumes both unlimited bandwidth and computation. Recently, Bar-Yehuda et al. [5] gave a Δ-approximation algorithm for weighted independent sets using the local ratio technique that runs in time $O(\mathrm{MIS}\cdot\log W)$ rounds, where MIS is the number of rounds needed to compute a maximal independent set and W is the ratio between the largest and smallest edge weight. We improve this approximation ratio by nearly a factor of 2 using only a single round, but at the price of obtaining a bound only on expected performance. Alon [1] gave nearly tight bounds for testing independence properties; his lower bound carries over to distributed algorithms, as we shall see in Sec. 2.4. For matchings, which correspond to independent sets in line graphs, Kuhn et al. [18] showed that achieving any constant factor approximation requires $\Omega(\max(\log\Delta/\log\log\Delta, \sqrt{\log n/\log\log n}))$ rounds.

Halldórsson and Konrad [?] examined how well the Caro-Wei bound performs in different subclasses of graphs. They also gave a randomized one-round distributed algorithm where nodes broadcast only a single bit that yields an independent set of expected size at least $0.24\cdot\textsc{CaroWei}(G)$ on every graph G. This is provably the least requirement for an effective distributed algorithm, as without degree information, the bounds are polynomially worse.

Streaming algorithms (including Boppana's) achieving Turán-like bounds in graphs and hypergraphs were considered in [15], and streaming algorithms for approximating $\textsc{CaroWei}(G)$ were given recently by Cormode et al. [?].

Motivated by a packet forwarding application, Emek et al. [12] considered the online set packing problem that corresponds to maintaining strong independent sets of large weight in hypergraphs under edge additions. We give a tight bound on their method for the special case of graphs.

2 Performance of Caro-Wei-Turán Bounds

We examine here how well the Caro-Wei and the Turán bounds perform on (unweighted) bounded-degree and sparse graphs.

Let $\textsc{opt}$ be an optimal independent set of size $\alpha=\alpha(G)$ and let $V'=V\setminus\textsc{opt}$. We say that a bound $B(G)$ has a performance ratio $f(\Delta)$ if, for all graphs G with $\Delta(G)=\Delta$ it holds that $\alpha(G)\geq B(G)\geq\alpha(G)/f(\Delta)$.

[1] $\tilde{O}(\cdot)$ suppresses $\log\log n$ factors.

2.1 Caro-Wei in Bounded-Degree Graphs

Theorem 1. CaroWei *has performance ratio* $(\Delta + 1)/2$.

Proof. Let G be a graph. Let O_i, for $i = 1, 2, \ldots, \Delta$, denote the number of vertices in OPT of degree i. Our approach is to separate the contributions of the different O_is to the Caro-Wei bound. The nodes of high degree have a smaller direct contribution, but also have an indirect contribution in forcing more nodes to be in V'.

Let m_{OPT} be the number of edges with an endpoint in OPT. Each such edge has the other endpoint in V', whereas nodes in V' are incident on at most Δ edges. Thus,

$$\sum_{i=1}^{\Delta} i \cdot O_i = m_{\text{OPT}} \leq \Delta |V'| \ . \tag{4}$$

We then obtain

$$\begin{aligned}
\textsc{CaroWei}(G) = \sum_{v \in V} \frac{1}{d(v)+1} &= \sum_{i=1}^{\Delta} O_i \cdot \frac{1}{i+1} + \sum_{v \in V'} \frac{1}{d(v)+1} \\
&\geq \sum_{i=1}^{\Delta} O_i \cdot \frac{1}{i+1} + |V'| \frac{1}{\Delta+1} \\
&\geq \frac{1}{\Delta+1} \sum_{i=1}^{\Delta} O_i \left(\frac{\Delta+1}{i+1} + \frac{i}{\Delta} \right) && \text{(Applying (4))} \\
&= \frac{1}{\Delta+1} \sum_{i=1}^{\Delta} O_i \left(2 + \frac{\Delta - i}{i+1} - \frac{\Delta - i}{\Delta} \right) \\
&\geq \frac{1}{\Delta+1} \sum_{i=1}^{\Delta} O_i \cdot 2 \\
&= \frac{2}{\Delta+1} \alpha(G) \ ,
\end{aligned}$$

obtaining the approximation upper bound claimed. Observe that the bound is tight only if all nodes in OPT are of degree Δ or $\Delta - 1$.

To see that the ratio attained is no better than $(\Delta + 1)/2$, observe that in any regular graph, the algorithm achieves a solution of exactly $n/(\Delta+1)$, while in bipartite regular graphs the optimal solution has size $n/2$. ☐

Remark. Selkow [?] generalized the Caro-Wei bound by extending Boppana's algorithm to two rounds. Namely, it adds also the nodes with no neighbor ordered earlier among those that did not get removed in the first round. For regular graphs, however, his bound reduces to the Caro-Wei bound, and thus does not attain a better performance ratio, given our lower bound construction.

2.2 Caro-Wei in Sparse Graphs

We now analyze the performance of the Caro-Wei bound in terms of the average degree $\overline{d} = 2|E|/n$. We shall use a certain application of the Cauchy-Schwarz inequality, which we state more generality in hindsight of its application in the following section.

Lemma 1. *If $x_1, x_2, \ldots, x_N$ and $w_1, w_2, \ldots, w_N$ are positive reals, then* $\displaystyle\sum_{i=1}^{N} \frac{w_i^2}{x_i} \geq \frac{\left(\sum_{i=1}^{N} w_i\right)^2}{\sum_{i=1}^{N} x_i}$.

Proof. The Cauchy-Schwarz inequality implies that for $u_1, u_2, \ldots, u_N$ and $v_1, v_2, \ldots, v_N$,

$$\left(\sum_{i=1}^{N} u_i v_i\right)^2 \leq \left(\sum_{i=1}^{N} u_i^2\right)\left(\sum_{i=1}^{N} v_i^2\right) .$$

The claim now follows using $u_i = \sqrt{x_i}$ and $v_i = w_i/\sqrt{x_i}$. □

Note that applying Lemma 1 with $w_v = 1$ and $x_v = d(v) + 1$ yields that

$$\textsc{CaroWei}(G) = \sum_{v \in V} \frac{1}{d(v)+1} \geq \frac{n^2}{\sum_v (d(v)+1)} = \frac{n}{\overline{d}+1} = \textsc{Turán}(G) .$$

Theorem 2. CaroWei *has performance ratio at most* $(\overline{d}+2)/1.657$.

Proof. Let OPT be an optimal independent set of size $\alpha = \alpha(G)$ and let $V' = V \setminus \text{OPT}$. Observe that when $|V'| = n - \alpha \geq \alpha$, the Turán bound gives $n/(\overline{d}+1) \geq \alpha \cdot 2/(\overline{d}+1)$, for a performance ratio of at most $(\overline{d}+1)/2$. We assume therefore that $\alpha \geq \frac{1}{2}n$.

Our approach is to first apply Lemma 1 separately on the parts of $\textsc{CaroWei}(G)$ corresponding to OPT and V'. We then show that the worst case occurs when all edges cross from OPT to V', indeed when the graph is bipartite with regular sides. Optimizing over the possible sizes of the sides then yields a tight upper and lower bounds.

Let m_{OPT} denote the number of edges with endpoint in OPT, $m_{V'}$ the number of edges with both endpoints in V' and $m = m_{\text{OPT}} + m_{V'}$ be the total number of edges. Observe that $\sum_{v \in \text{OPT}} d(v) = m_{\text{OPT}}$ while $\sum_{v \in V'} = m_{\text{OPT}} + 2m_{V'}$.

Lemma 1 (with $w_v = 1$ and $x_v = d(v)+1$) applied to OPT and V' separately yields that

$$\textsc{CaroWei}(G) = \sum_{v \in \text{OPT}} \frac{1}{d(v)+1} + \sum_{v \in V'} \frac{1}{d(v)+1} \geq \frac{\alpha^2}{m_{\text{OPT}} + \alpha} + \frac{(n-\alpha)^2}{m_{\text{OPT}} + 2m_{V'} + (n-\alpha)} ,$$

Denoting $t = m_{\text{OPT}}/m$, we get that

$$\textsc{CaroWei}(G) \geq \frac{\alpha^2}{t \cdot m + \alpha} + \frac{(n-\alpha)^2}{(2-t)m + n - \alpha} . \tag{5}$$

Considered as a function f of t, the r.h.s. of (5) has derivative

$$\frac{df}{dt} = -\frac{\alpha^2}{(tm+\alpha)^2} + \frac{(n-\alpha)^2}{((2-t)m + n - \alpha)^2} .$$

Since we assume $\alpha \geq n/2$, it holds that $\alpha^2(m+n-\alpha)^2 \geq (n-\alpha)^2(m+\alpha)^2$, and thus $df/dt \leq 0$ for all $t \in [0,1]$. Hence, denoting $\tau = \alpha/n$, we obtain that

$$\textsc{CaroWei}(G) \geq \frac{\alpha^2}{m+\alpha} + \frac{(n-\alpha)^2}{m+n-\alpha} = \alpha\left(\frac{\tau}{\overline{d}/2+\tau} + \frac{(1-\tau)^2/\tau}{\overline{d}/2+1-\tau}\right) . \tag{6}$$

The expression in the parenthesis then upper bounds the reciprocal of the performance guarantee of CaroWei.

To see that (6) is tightest possible, consider bipartite graphs G with regular sides. Let τ be such that τn is the size of the larger side and q is the degree of those vertices. Then the number of edges is $m = q \cdot \tau n$, average degree is $\overline{d} = 2m/n = 2q\tau$, and the degree of the nodes on the other side is $m/((1-\tau)n) = \overline{d}/(2(1-\tau))$. Clearly $\alpha(G) = \tau n$, while the Caro-Wei bound gives

$$\textsc{CaroWei}(G) = \frac{\tau n}{\overline{d}/(2\tau)+1} + \frac{(1-\tau)n}{\overline{d}/(2(1-\tau))+1} = \alpha(G)\left(\frac{1}{\overline{d}/(2\tau)+1} + \frac{(1-\tau)/\tau}{\overline{d}/(2(1-\tau))+1}\right) ,$$

which matches (6).

If we round up the lower order terms in the denominator of (6), we obtain a simpler expression for the asymptotic performance with $\overline{d}$:

$$\textsc{CaroWei}(G) \geq \alpha(G)\left(\frac{\tau + (1-\tau)^2/\tau}{\overline{d}/2+1}\right) ,$$

which is minimized when $\tau = 1/\sqrt{2}$, for a performance ratio at most $(\overline{d}+2)/(4(\sqrt{2}-1)) \leq (\overline{d}+2)/1.657$. $\square$

2.3 Turán Bound

Recall Turán's theorem that $\alpha(G) \geq \textsc{Turán}(G) = \frac{n}{\overline{d}+1} = \frac{n^2}{2|E|+n}$. We find that the guarantee of the Turán bound is strictly weaker than that of Caro-Wei, yet asymptotically equivalent.

Theorem 3. Turán *has performance ratio* $\dfrac{(2\Delta+1)^2}{8\Delta} = \dfrac{\Delta+1}{2} + \dfrac{1}{8\Delta}$.

Proof. Because OPT $= V \setminus V'$ is independent, each of the $|E|$ edges of G is incident to at least one vertex in V'. Conversely, each vertex in V' is incident to at most Δ edges. So by counting edges, we get

$$|E| \leq \Delta|V'| = \Delta(n-\alpha).$$

Therefore

$$2|E| + n \leq 2\Delta(n-\alpha) + n = (2\Delta+1)n - 2\Delta\alpha.$$

Multiplying by $8\Delta\alpha$ and using the inequality $4xy \leq (x+y)^2$ gives

$$8\Delta\alpha(2m+n) \leq 4(2\Delta\alpha)[(2\Delta+1)n - 2\Delta\alpha] \leq [(2\Delta+1)n]^2 .$$

Dividing both sides by $8\Delta(2m+n)$ gives

$$\alpha \leq \frac{(2\Delta+1)^2}{8\Delta} \cdot \frac{n^2}{2m+n} = \frac{(2\Delta+1)^2}{8\Delta}\textsc{Turán}(G) .$$

The argument above shows that the performance ratio of Turán's bound is at most $\frac{(2\Delta+1)^2}{8\Delta}$. This performance ratio is tight as a function of Δ. To see why, given $\Delta > 0$, let A, B, and C be disjoint sets of size $2\Delta-1$, $2\Delta-1$, and 2, respectively. Let G be any Δ-regular bipartite graph with parts A and B, together with two isolated vertices in C. We can check that $n = 4\Delta$, $|E| = (2\Delta-1)\Delta$, $\textsc{Turán}(G) = \frac{8\Delta}{2\Delta+1}$, and $\alpha(G) = 2\Delta+1$. So the performance ratio of Turán's bound on this graph is indeed $\frac{(2\Delta+1)^2}{8\Delta}$. $\square$

2.4 Limitations of Distributed Algorithms

We may assume that we are equipped with unique labels from a universe of N labels, where $N \geq \Delta \cdot n$. The nodes have knowledge of n, Δ and N, and have unlimited bandwidth and computational ability. The nodes have distinct ports for communication with their neighbors, but do not initially know there labels.

Our result for Boppana's algorithm is optimal for 1-round algorithms. Observe that the lower bounds below hold also for randomized algorithms.

Theorem 4. *Every 1-round distributed algorithm has performance ratio at least* $(\Delta+1)/2$*, even on unweighted regular graphs.*

Proof. In a single round, each node can only learn the labels of their neighbors and their random bits.

Consider the graph $G_1 = K_{\Delta+1}$, and G_2, which is any Δ-regular bipartite graph. Distributions over neighborhoods are identical. Hence, no 1-round algorithm can distinguish between these graphs.

All nodes will join the independent set with the same probability, averaged over all possible labelings, since they share the same views. This probability can be at most $1/(\Delta+1)$, as otherwise the algorithm would produce incorrect answers on $K_{\Delta+1}$. The size of the solution is then at most $n/(\Delta+1)$, while on every Δ-regular bipartite graphs, the optimal solution contains $n/2$ nodes. □

It is not clear if better results can be obtained when using more rounds. A weaker lower bound holds even for nearly logarithmic number of rounds.

Theorem 5. *There are positive constants* c_1 *and* c_2 *such that the following holds: Every* $c_1 \log_\Delta n$*-round distributed algorithm has performance ratio at least* $c_2\Delta/\log\Delta$*.*

Proof. Alon [1] constructs a Δ-regular graph G_1 of girth $\Omega(\log n/\log\Delta)$ with independence number $O(n/\Delta \cdot \log\Delta)$, and notes that it is well known that there exist a bipartite Δ-regular graph G_2 of girth $\Omega(\log n/\log\Delta)$. The distributions over the k-neighborhoods of these graphs are identical, for $k = O(\log n/\log\Delta)$. Hence, no k-round distributed algorithm can distinguish between the two. □

3 Approximations for Weighted Graphs

In the weighted setting, each node v is assigned a positive integral weight $w(v)$ and the objective is to find an independent set I maximizing the total weight $\sum_{v\in I} w(v)$. For a set $X \subseteq V$, denote $w(X) = \sum_{x\in X} w(x)$.

Boppana's algorithm can be applied unchanged to weighted graphs, producing a solution B of expected weight

$$\mathbb{E}[w(B)] = \sum_{v\in B} w(v) \cdot \frac{1}{d(v)+1} \ ,$$

by linearity of expectation. This immediately implies that $\mathbb{E}[w(B)] \geq w(V)/(\Delta+1)$, for a performance ratio at most $\Delta+1$. To see that this is also the best possible bound, consider the complete bipartite graphs $K_{N,N}$, where the nodes on one side have weight 1 and on the other side weight Q, for a parameter $Q \geq \Delta^2$. The expected weight of the algorithm solution is $(N+NQ)/(\Delta+1)$, while the optimal solution is of weight NQ. The performance ratio is then $(\Delta+1)/(1+1/Q)$, which goes to $\Delta+1$ as Q gets large.

We therefore turn our attention to modifications that take the weights into account.

3.1 Modified algorithm

We consider now a variation, MAX, previously considered in an online setting in [12].

> Each node v picks a random real number x_v uniformly from $[0,1]$. It broadcasts the values x_v and w_v to its neighbors, who compute from it $r_v = x_v^{1/w_v}$. As before, each node u joins the solution if its value r_u is the highest among its neighbors.

The only difference is the computation of r_v, which now depends on the weight w_v. Again the algorithm runs in a single round of Broadcast-CONGEST, with correctness following as before. The algorithm was previously shown in [12] to attain a Δ-approximation.

We obtain a tight bound, which does not have a nice closed expression.

Theorem 6. *The performance ratio $\rho(\Delta)$ of* MAX*, as a function of Δ, is given by*

$$\frac{1}{\rho} = \min_{x\leq 1}\left(\frac{x^2}{\Delta+x} + \frac{1}{x\Delta+1}\right) .$$

We prove Theorem 6 in the following subsection.

If we focus on the asymptotics as Δ gets large, we can ignore the additive terms in the denominators, obtaining that the performance ratio approaches

$$\rho(\Delta) \underset{\Delta\to\infty}{\rightarrow} (\Delta+1)\cdot\frac{1}{x^2+1/x} .$$

This is maximized when $x = 2^{-1/3}$ for a ratio of $2^{2/3}(\Delta+1)/3 \sim (\Delta+1)/1.89 \ \sim 0.529(\Delta+1)$.

Theorem 7. *The asymptotic performance ratio of* MAX *is $2^{2/3}(\Delta+1)/3 \sim 0.529(\Delta+1)$.*

Figure 1 shows $\rho(\Delta)/(\Delta+1)$ as a function of Δ. For $\Delta = 2$, we find that $1/\rho \sim 0.593$, or $\rho \sim 1.657 \sim 0.562(\Delta+1)$, which is about 6% larger than $0.529(\Delta+1)$, but 20% smaller than Δ. For $\Delta = 1$, the algorithm can made optimal by preferring nodes with higher weight than their sole neighbor.

3.2 Analysis

The key property of the MAX rule that leads to improved approximation is that the probability that a node is selected is now proportional to the fraction of its weight within its closed neighborhood (consisting of itself and its neighbors). We then obtain a bound in terms of weights of sets of nodes – the optimal solution and the remaining nodes – using the Cauchy-Schwarz inequality. We safely upper bound the degree of each node by Δ, but the main effort then is to show that the worst case occurs when the graph is bipartite with equal sides. This leads to matching upper and lower bounds.

Let $N(v)$ denote the set of neighbors of vertex v and $N[v] = \{v\}\cup N(v)$ its closed neighborhood. Let MAX also refer to the set of nodes selected by MAX.

The key property of the MAX rule is that the probability that a node is selected is now proportional to the fraction of its weight within its closed neighborhood. We provide a proof for the next lemma for completeness.

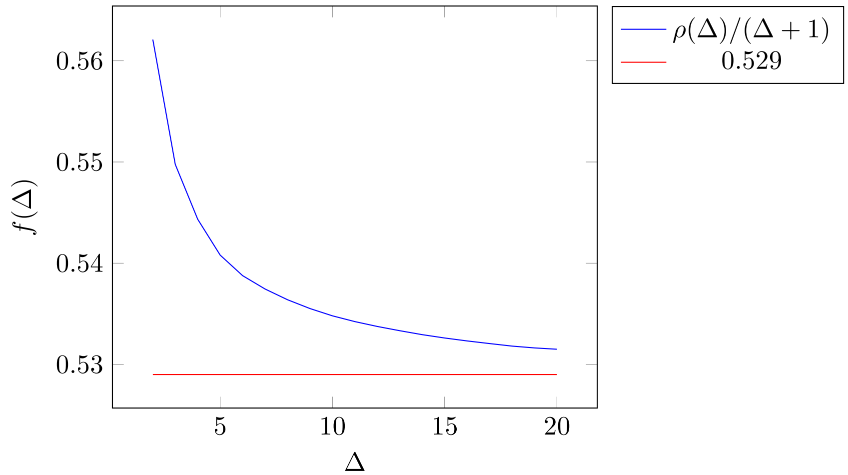

Figure 1: Bounds on performance ratio, for small values of Δ.

Lemma 2 ([12])**.** *For each vertex $v \in V$, we have that* $\mathbb{P}[v \in \textsc{Max}] = \dfrac{w(v)}{w(N[v])}$.

Proof. Let $r_{\max} = \max\{r_u : u \in N(v)\}$. By independence of the random choices we have, for $\alpha \in [0,1]$, that

$$\mathbb{P}[r_{\max} < \alpha] = \prod_{u \in N(v)} \mathbb{P}[r_u < \alpha] = \prod_{u \in N(v)} \mathbb{P}[x_u < \alpha^{w(u)}] = \alpha^{\sum_{u \in N(v)} w(u)} = \alpha^{w(N(v))} \ .$$

It follows that $r_{\max}$ has distribution $D_{w(N(v))}$, where the distribution D_z has density $f_z(\alpha) = z\alpha^{z-1}$, for $\alpha \in [0,1]$. Hence,

$$\mathbb{P}[r_v > r_{\max}] \ = \int_0^1 \mathbb{P}[r_{\max} < \alpha] \cdot f_{r_v}(\alpha) d\alpha \ = \int_0^1 \alpha^{w(N(v))} \cdot w(v) \alpha^{w(v)-1} d\alpha \ = \ \frac{w(v)}{w(N[v])} \ ,$$

as required. $\square$

Note that by Lemma 2 and linearity of expectation, we have that

$$\mathbb{E}[w(S \cap \textsc{Max})] = \sum_{v \in S} \mathbb{P}[v \in \textsc{Max}] \cdot w(v) = \sum_{v \in S} \frac{w(v)^2}{w(N[v])} \ , \tag{7}$$

for any subset $S \subseteq V$. Applying Lemma 1 (with $x_v = w(N[v])$) gives:

Lemma 3. *For any subset $S \subseteq V$ we have that* $\mathbb{E}[w(S \cap \textsc{Max})] \geq \dfrac{w(S)^2}{\sum_{v \in S} w(N[v])}$.

Applying Lemma 3 with $S = V$ gives an absolute lower bound on the solution size.

Lemma 4. $\mathbb{E}[w(\textsc{Max})] \geq \dfrac{w(V)^2}{\sum_{v \in V} w(N[v])} = \dfrac{w(V)^2}{\sum_{v \in V} (d(v)+1) w(v)} \geq \dfrac{w(V)}{\Delta + 1}$.

We need the following lemma when showing that worst case occurs for bipartite graphs.

Lemma 5. *Let $a > b > 0$ and let $Z - Y \geq X > 0$. Then*

$$\min_{t\in[0,1]} \left\{ \frac{a}{Y+tX} + \frac{b}{Z+(1-t)X} \right\} = \frac{a}{Y+X} + \frac{b}{Z} \ .$$

Proof. Let $f(t) = \frac{a}{Y+tX} + \frac{b}{Z+(1-t)X}$. We have that $\frac{df(t)}{dt} = -\frac{aX}{(Y+tX)^2} + \frac{bX}{(Z+(1-t)X)^2}$, which is negative for any $t \in [0,1]$, since $a > b$ and $Y + tX \leq Z + (1-t)X$. $\square$

Now we are ready to prove Theorem 6.

Proof of Theorem 6. Let $\textsc{opt}$ be an optimal solution, and define $V' \doteq V \setminus \textsc{opt}$, and $\beta \doteq w(V')/w(\textsc{opt})$. When $\beta \geq 1$, Lemma 4 implies that the performance ratio is at most $(\Delta+1)/2$. We therefore focus on the case where $\beta < 1$.

We first apply Lemma 3 separately on $\textsc{opt}$ and on V', obtaining:

$$w(\textsc{Max}) = w(\textsc{Max} \cap \textsc{opt})) + w(\textsc{Max} \cap V') \geq \frac{w(\textsc{opt})^2}{\sum_{v\in\textsc{opt}} w(N[v])} + \frac{w(V')^2}{\sum_{v\in V'} w(N[v])} \ . \tag{8}$$

Let $W = \sum_{v\in V'} w(v) \cdot |N(v) \cap \textsc{opt}| = \sum_{v\in\textsc{opt}} w(N(v))$ be the weighted degree of the nodes of V' into $\textsc{opt}$, which can be viewed as the total of the weights of neighborhoods of nodes in $\textsc{opt}$. Thus,

$$\sum_{v\in\textsc{opt}} w(N[v]) = w(\textsc{opt}) + \sum_{v\in V'} w(v)|N(v) \cap \textsc{opt}| = w(\textsc{opt}) + W \ . \tag{9}$$

and

$$\begin{aligned}
\sum_{v\in V'} w(N[v]) &= w(V') + \sum_{v\in V'} w(N(v)) \\
&= w(V') + \sum_{v\in\textsc{opt}} w(v) \cdot |N(v) \cap V'| + \sum_{v\in V'} w(v) \cdot |N(v) \cap V'| \\
&\leq w(V') + \Delta w(\textsc{opt}) + \sum_{v\in V'} w(v) \cdot (\Delta - |N(v) \cap \textsc{opt}|) \\
&= w(V') + \Delta w(\textsc{opt}) + \Delta w(V') - W \ .
\end{aligned} \tag{10}$$

Applying (9) and (10) to (8) gives

$$w(\textsc{Max}) \geq \frac{w(\textsc{opt})^2}{w(\textsc{opt}) + W} + \frac{w(V')^2}{w(V') + \Delta w(\textsc{opt}) + \Delta w(V') - W} \ .$$

Since $\beta < 1$ and $W \leq \Delta w(V')$ we can use Lemma 5 with $a = w(\textsc{opt})^2$, $b = w(V')^2$, $Y = w(\textsc{opt})$, $Z = w(V') + \Delta w(\textsc{opt})$, $X = \Delta w(V')$, and $t = W/X$. Hence,

$$w(\textsc{Max}) \geq \frac{w(\textsc{opt})^2}{w(\textsc{opt}) + \Delta w(V')} + \frac{w(V')^2}{w(V') + \Delta w(\textsc{opt})} = w(\textsc{opt}) \cdot \left(\frac{1}{1+\Delta\beta} + \frac{\beta^2}{\beta+\Delta} \right) \ . \tag{11}$$

The upper bound of the theorem therefore follows.

To see that bound (11) is tight, consider any Δ-regular bipartite graph $G = (V, E)$ with V partitioned into two sets L and R, where $|L| = |R|$. Set the weight of nodes in L and in R as 1 and β, respectively, for some $\beta \le 1$. Clearly, the weight of the optimal solution is $w(\textsc{opt}) = |L|$. Observe that

$$w(\textsc{Max}) = |L| \cdot \frac{1}{1+\Delta\beta} + |R|\beta \cdot \frac{\beta}{\beta+\Delta} = w(\textsc{opt}) \cdot \left(\frac{1}{1+\beta\Delta} + \frac{\beta^2}{\beta+\Delta} \right) ,$$

matching (11). □

Remark. Sakai et al. [19] considered the following greedy algorithm (named GWMIN2): add the vertex v maximizing $w(v)/w(N[v])$ to the solution, remove its closed neighborhood, and recurse on the remaining graph. They derived a $(\Delta + 1)$-approximation upper bound but not a matching lower bound. Since their algorithm attains the bound (7) (see [19]), our analysis implies that it also attains the bound of Theorem 6.

4 Conclusion

It's surprising that the best distributed approximations known of independent sets are obtained by the simplest algorithm. Repeating the algorithm on the remaining graph will certainly give a better solution – the challenge is to quantify the improvement.

References

[1] N. Alon. On constant time approximation of parameters of bounded degree graphs. In *Property Testing - Current Research and Surveys*, pages 234–239, 2010.

[2] N. Alon and J. H. Spencer. *The probabilistic method*. John Wiley & Sons, 2004.

[3] P. Austrin, S. Khot, and M. Safra. Inapproximability of vertex cover and independent set in bounded degree graphs. In *Computational Complexity, 2009. CCC'09. 24th Annual IEEE Conference on*, pages 74–80. IEEE, 2009.

[4] N. Bansal, A. Gupta, and G. Guruganesh. On the Lovász theta function for independent sets in sparse graphs. In *STOC*, pages 193–200, 2015.

[5] R. Bar-Yehuda, K. Censor-Hillel, M. Ghaffari, and G. Schwartzman. Distributed approximation of maximum independent set and maximum matching. In *PODC*, pages 165–174, 2017.

[6] P. Berman and T. Fujito. On approximation properties of the independent set problem for low degree graphs. *Theory Comput. Syst.*, 32(2):115–132, 1999.

[7] M. H. Bodlaender, M. M. Halldórsson, C. Konrad, and F. Kuhn. Brief announcement: Local independent set approximation. In *PODC*, New York, NY, USA, 2016. ACM.

[8] R. B. Boppana. Personal communication to Joel Spencer, 1987.

[9] Y. Caro. New results on the independence number. Technical report, Tel Aviv University, 1979.

[10] S. O. Chan. Approximation resistance from pairwise-independent subgroups. *Journal of the ACM (JACM)*, 63(3):27, 2016.

[11] V. Chvátal and C. McDiarmid. Small transversals in hypergraphs. *Combinatorica*, 12(1):19–26, 1992.

[12] Y. Emek, M. M. Halldórsson, Y. Mansour, B. Patt-Shamir, J. Radhakrishnan, and D. Rawitz. Online set packing. *SIAM Journal of Computing*, 41(4):728–746, 2012.

[13] P. Erdős. On the graph theorem of Turán (in Hungarian). *Mat. Lapok*, 21:249–251, 1970.

[14] J. R. Griggs. Lower bounds on the independence number in terms of the degrees. *J. Combin. Theory B*, 34:22–39, 1983.

[15] B. V. Halldórsson, M. M. Halldórsson, E. Losievskaja, and M. Szegedy. Streaming algorithms for independent sets in sparse hypergraphs. *Algorithmica*, 76:490–501, 2016.

[16] M. Halldórsson and J. Radhakrishnan. Greed is good: Approximating independent sets in sparse and bounded-degree graphs. *Algorithmica*, 18(1):145–163, 1997.

[17] M. M. Halldórsson and C. Konrad. Distributed large independent sets in a single round. *Distributed Computing*, 2017.

[18] F. Kuhn, T. Moscibroda, and R. Wattenhofer. Local computation: Lower and upper bounds. *J. ACM*, 63(2):17:1–17:44, Mar. 2016.

[19] S. Sakai, M. Togasaki, and K. Yamazaki. A note on greedy algorithms for the maximum weighted independent set problem. *Discrete Applied Mathematics*, 126(2-3):313–322, 2003.

[20] P. Turán. On an extremal problem in graph theory (in Hungarian). *Mat. Fiz. Lapok*, 48:436–452, 1941.

[21] V. Wei. A lower bound on the stability number of a simple graph. Technical report, Bell Laboratories, 1981.