It is Hard to Know when Greedy is Good for Finding Independent Sets*

Hans L. Bodlaender, Dimitrios M. Thilikos, Koichi Yamazaki
Department of Computer Science, Utrecht University,
P.O. Box 80.089, 3508 TB Utrecht, the Netherlands
E-mail: {hansb,sedthilk,koichi}@cs.ruu.nl

Abstract

The classes $\mathcal{A}_r$ and $\mathcal{S}_r$ are defined as the classes of those graphs, where the minimum degree greedy algorithm always approximates the maximum independent set (MIS) problem within a factor of r, respectively, where this algorithm has a sequence of choices that yield an output that is at most a factor r from optimal, $r \geq 1$ a rational number. It is shown that deciding whether a given graph belongs to $\mathcal{A}_r$ is coNP-complete for any fixed $r \geq 1$, and deciding whether a given graph belongs to $\mathcal{S}_1$ is DP-hard, and belongs to $\Delta_2\mathbf{P}$. Also, the MIS problem remains NP-complete when restricted to $\mathcal{S}_r$.

Keywords: Analysis of algorithms, Combinatorial problems, Approximation algorithms

1 Introduction

A well known and well studied heuristic for the problem of computing a maximum independent set in a graph is the Minimum Degree Greedy algorithm (MDG). In this algorithm, one repeatedly selects a vertex of minimum degree from the graph, puts this vertex in the independent set, and removes the vertex and its neighbours from the graph, until an empty graph is left.

An interesting problem is when this MDG algorithm outputs a maximum independent set, or when its output differs a constant factor from a maximum independent set.

For several classes of graphs it is known that, if we require the input to belong to such a class, then MDG has a good approximation ratio; examples are the graphs of bounded degree or bounded average degree [6]. Also, MDG is known to output always a maximum independent set, when the input is a well-covered graph (a graph is *well-covered* if all its maximal independent sets are of the same cardinality – see [8]). Moreover, it is easy to verify that MDG outputs a maximum

*The second author was supported by the Training and Mobility of Researchers (TMR) Program, (EU contract no. ERBFMBICT950198). The third author was supported by Ministry of Education, Science, Sports and Culture of Japan as Overseas Research Scholar.

independent set when the input is a tree, split graph, complements of a k-tree, or a complete k-partite graph, for any k.

To consider the problem to determine when the MDG algorithm gives certain approximations of the maximum independent set, we introduce for each rational number r the graph class $\mathcal{A}_r$, consisting of those graphs where MDG always outputs an independent size such that the maximum independent set is at most r times as large. In other words, $\mathcal{A}_r$ is the class of graphs for which MDG is an approximation algorithm with performance ratio r.

Note that the MDG algorithm has a certain degree of non-determinism: when there are more vertices of minimum degree, the algorithm chooses one of them to remove. We define the graph class $\mathcal{S}_r$ (r rational number) as the set of graphs for which there exist some sequence of choices of minimum degree vertices for the MDG algorithm, such that the output is of size at least a constant fraction $1/r$ of the maximum independent set.

We prove that the hierarchies defined by classes $\mathcal{A}_r$ and $\mathcal{S}_r$ are proper i.e. for any $r_1 < r_2$ $\mathcal{A}_{r_1} \subset \mathcal{A}_{r_2}$. A consequence of this is (the non-surprising result) that for any function $f(n) = o(n)$ MDG is not a $f(n)$-approximation algorithm for the maximum independent set problem (n is the number of vertices in the input graph).

In this paper, we consider the complexity of the recognition problem for the classes $\mathcal{A}_r$ and $\mathcal{S}_r$ for rational r. We prove that for any r, the recognition problem of $\mathcal{A}_r$ is coNP-complete. Also, for any r, the recognition problem of $\mathcal{S}_r$ belongs to $\Delta_2\mathbf{P}$. We also prove that maximum independent set remains an NP-complete when restricted to graphs belonging to $\mathcal{S}_1$ and that the recognition of $\mathcal{S}_1$ is a **DP** hard problem.

Our results indicate that the problem of recognising the instances of the maximum independent set problem where the greedy algorithm has a nice approximation behaviour is a hard combinatorial problem. Clearly, the same results hold also for the maximum degree greedy algorithm for the clique problem (just take the complement of the graphs involved.)

2 Definitions and preliminaries

Throughout this paper all the graphs are considered to be without loops or multiple edges. Given a graph G we denote as $V(G)$ and $E(G)$ its vertex and edge set respectively. Given a set $S \subseteq V(G)$, we define the neighbourhood of S, denoted $N(S)$, to be the set of vertices not in S that are adjacent to vertices in S. Given a vertex $v \in V(G)$, we call the set $N(\{v\})$ the neighbourhood of v in G and we denote it as $N(v)$. Given some set $S \subseteq V(G)$ we denote as $G[S]$ the subgraph of G induced by S. A set $I \subseteq V(G)$ is an *independent set* if $E(G[I]) = \emptyset$. An independent set I is a *maximal* independent set when there is no independent set I with $I' \subset I$, $I' \neq I$. We call an independent set I *maximum*, when there is no independent set I' with $|I'| > |I|$. The *Maximum Independent Set* (MIS) problem, is the problem of finding a maximum independent set of a given graph. Finally, we denote the size of some maximum independent set in G as $\alpha(G)$. The decision version of the MIS problem asks, for given G,k, whether $\alpha(G) \geq k$.

One of the most simple and efficient algorithms to output a maximal independent set of a given graph is the one called *minimum-degree greedy* (MDG) algorithm.

Algorithm MDG
Input: A graph G
Output: A maximal independent set I of G.
1 **begin**
2 $I \leftarrow \emptyset$
3 Let $v \in V(G)$ be a vertex of minimum degree in G
4 $I \leftarrow I \cup \{v\}$
5 $G \leftarrow G[V(G) - \{v\} - N(v)]$
6 **if** $V(G) \neq \emptyset$ **then goto** 3
7 **end**

It is easy to see that line 3 of MDG algorithm introduces a certain degree of non-determinism, as there may be more than one minimum degree vertices to be chosen. To any graph G we associate the collection $\mathcal{I}_G$ of all possible maximum independent sets that MDG may output with input graph G, i.e., we look at all possible sequences of choices of vertices of minimum degree. We proceed with some definitions:

Definition 1 Let $r \geq 1$ be some rational number.
$\text{max-GR}(G) = max\{|I| : I \in \mathcal{I}_G\}$, $\text{min-GR}(G) = min\{|I| : I \in \mathcal{I}_G\}$,
$\mathcal{S}_r = \{G : \alpha(G)/r \leq \text{max-GR}(G)\}$, $\mathcal{A}_r = \{G : \alpha(G)/r \leq \text{min-GR}(G)\}$.

In other words, $\mathcal{A}_r$ is the class of graphs for which MDG is an approximation algorithm for MIS with performance ratio r. Also, $\mathcal{S}_r$ is the class of graphs for which there exist some sequence of minimum degree choices for the MDG algorithm such that the output has size at least a constant factor r of the MIS solution.

One can easyly verify that $\mathcal{A}_1$ contains all trees, cycles, split graphs, complete k-partite graphs and complements of k-trees. We also mention that $\mathcal{A}_1$ contains the class of well-covered graphs (the recognition problem of well-covered graphs has been proved to be a coNP-complete problem (see [4, 5])). Also, according to the results in [6], if $r \geq \frac{\Delta+2}{3}$, then $\mathcal{A}_r$ contains all the graphs with degree bounded by Δ.

Proposition 2.1 *For all rational numbers r_1, r_2 with $1 \leq r_1 < r_2$, $\mathcal{A}_{r_1}$ is a proper subset of $\mathcal{A}_{r_2}$, and $\mathcal{S}_{r_1}$ is a proper subset of $\mathcal{S}_{r_2}$.*

Proof. We look to the first part of the claim; the second part can be proved with the same construction. Note that it is sufficent to show that for any rational number $r \geq 1$, there exists a graph G with $\frac{\alpha(G)}{\text{min-GR}(G)} = r$.

Write $r = l/m$ with $l \geq m \geq 2$. We construct G in the following way: Take a vertex v_0 and a set $I = \{v_1, \ldots, v_l\}$ of l vertices adjacent to v_0. Let $\{I_1, \ldots, I_{m-1}\}$

be an arbitrary partition of I consisting of $m-1$ non-empty sets. Take additionally $m-1$ cliques $K_1, \ldots, K_{m-1}$, each consisting of $l+1$ vertices. The construction is completed by connecting each vertex in I_i with all vertices in K_i, for any $i = 1, \ldots, m-1$. We can easily verify that I is a maximum independent set in G. Also, MDG will always start choosing vertex v_0 and, because of this first choice, will finally output a maximum independent set consisting of v_0 and one vertex from each of the $m-1$ cliques $K_1, \ldots, K_{m-1}$ (an example for the case $r = \frac{5}{4}$ is shown in Figure 1).

Thus, $\alpha(G) = l$, but MDG outputs an independent set of size m: $G \in \mathcal{A}_r$ and $G \notin \mathcal{A}_{r'}, \forall r' < r$. □

The fact that for any $r \geq 1$ there are infinitely many graphs not in $\mathcal{A}_r$ shows that MDG is not an constant factor approximation algorithm. In fact, we can prove that MDG is not an aproximation algorithm for any aproximation factor of the form $f(n)$ where $\lim_{n\to\infty} \frac{n}{f(n)} = 0$ (n is the number of vertices of the input graph). For this, it is sufficient to see that if we apply the above construction for $l = l_0$ and $m = 2$ where $l_0 > 2f(2l_0+1)$, we obtain a graph G_{l_0} where $\alpha(G_{l_0}) = l_0$ and $\text{min-GR}(G_{l_0}) = 2$. As $|V(G_{l_0})| = 2l_0+1$, we have that $\frac{\alpha(G_{l_0})}{\text{min-GR}(G_{l_0})} = \frac{l_0}{2} > f(|V(G_{l_0})|)$, a contradiction to the existence of any $f(n)$-approximation algorithm.

We mention that MIS is not approximable within a factor of $n^{1/3-\epsilon}$ unless coRP=NP (see [1]).

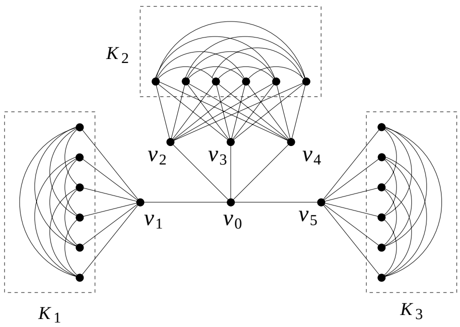

Figure 1: An example of a graph in $\mathcal{A}_{\frac{5}{4}}$ and/or $\mathcal{S}_{\frac{5}{4}}$

3 The complexity of recognizing $\mathcal{A}_r$

In this section we will prove that the recognition of those graphs where the MDG algorithm approximates the maximum independent set with approximation ratio any fixed rational number $r \geq 1$ is a coNP-complete problem.

Theorem 3.1 *For any fixed rational number $r \geq 1$, the problem to determine whether a given graph $G \in \mathcal{A}_r$ is coNP-complete.*

Proof. First, in order to show that the problem belongs to coNP, it is sufficient to observe that $G \notin \mathcal{A}_r$ if and only if there exist a set $I \subset V$, and a sequence of vertices $(v_1, \cdots, v_i)$, such that

- I is an independent set,
- $(v_1, \cdots, v_i)$ is an independent set which can be chosen by the MDG algorithm,
- $|I|/r > i$.

To prove hardness for coNP, we present a reduction from the problem, to determine whether for a given graph G and integer k, $\alpha(G) < k$, to the problem to determine whether for a given graph G' $G' \in \mathcal{A}_r$.

Let $G = (V, E)$ be a given graph, and k be a given positive integer. Write $r = l/m$, l, m integers ($l \geq m$). Construct G' as follows:

Take a clique A with $l \cdot |V(G)|$ vertices. Take a graph B consisting of l disjoint copies of G. Take a graph C consisting of $km - 2$ isolated vertices. Let G' be the graph such that $V(G') = V(A) \cup V(B) \cup V(C)$ and $E(G') = E(A) \cup E(B) \cup E(C) \cup \{\{u, v\} \mid u \in V(A) \cup V(C), v \in V(B)\}$, i.e., we make every vertex in B adjacent to all vertices in A and in C.

Since $G' \notin \mathcal{A}_r$ iff min-GR$(G') < \frac{m}{l}\alpha(G')$, it is sufficient to prove that $\alpha(G) \geq k$ iff min-GR$(G') < \frac{m}{l}\alpha(G')$. Notice that with G' as input, MDG algorithm always outputs a maximal independent set $V_C \cup \{p\}$, where p is a vertex in V_A and thus min-GR$(G') = km - 1$. Also, it is easy to see that $\alpha(G') = \max\{l\alpha(G), km - 1\}$:

Suppose that $\alpha(G) \geq k$. Then min-GR$(G') = km - 1 < m\alpha(G) = \frac{m}{l}\alpha(G')$.

Suppose that $\alpha(G) < k$. We distinguish two cases:

Case 1: $\frac{km-1}{l} \leq \alpha(G)$. We now have $\alpha(G') = l\alpha(G)$ and thus min-GR$(G) = km - 1 \geq m\alpha(G) = \frac{m}{l}\alpha(G')$.

Case 2: $\frac{km-1}{l} > \alpha(G)$. We now have $\alpha(G') = mk - 1$ and thus min-GR$(G') = km - 1 = \alpha(G') \geq \frac{m}{l}\alpha(G')$. □

It is easy to see that, using the same reduction with the one of Theorem 3.1, one can prove that the recognition problem for $\mathcal{S}_r$ is also a coNP-hard problem. In the next section we will prove a stronger result for $r = 1$.

It is a natural question to ask about the complexity of recognizing $\mathcal{A}_r$ (or $\mathcal{S}_r$) when r is considered to be an irrational number. One can actually prove that there are irrational numbers r, such that the recognition problem for $\mathcal{A}_r$, or $\mathcal{S}_r$ is undecidable. (Take any undecidable function $f : \mathbf{N} \to \{0, 1\}$, e.g., $f(n)$ tells whether the nth Turing machine in some recursive numbering halts on an empty input. Let $r = 1 + \sum_{i=1}^{\infty} 2^{-1} \cdot f(i)$. If testing membership in $\mathcal{A}_r$ or $\mathcal{S}_r$ is decidable, then one can compute the digits of r using graphs, as constructed in the proof of Proposition 2.1.)

4 Complexity results on $\mathcal{S}_r$

First, we show it does not help to know that a graph belongs to $\mathcal{S}_1$ (and hence, to any class $\mathcal{S}_r$ for $r \geq 1$) when we want to solve the maximum independent set problem.

Theorem 4.1 *The maximum independent set problem, restricted to $\mathcal{S}_1$ is NP-complete.*

Proof. We will give a reduction from the maximum independent set problem for arbitrary graphs. For a given (non-empty) graph G, we will construct a new graph $G' \in \mathcal{S}_1$ such that $\alpha(G') = \alpha(G) + |E(G)|$. G' is obtained from G by first replacing every edge in G by a path of length three (i.e., the edge is subdivided by putting two new vertices on it), and then taking two new adjacent vertices x, y and making these adjacent to all the original vertices in G. (See Figure 2 for an example.)

The original vertices from G are called the *real* vertices in G', the vertices introduced by the subdivisions are called the *dummy vertices*, and x and y are called the *additional* vertices.

We will now show that $\alpha(G') = \alpha(G) + |E(G)|$. Let I' be a maximum independent set of G'. Let $I = V(G) \cap I'$ be the set of real vertices in I'. Change I' in the following way: while there are vertices $v, w \in I$ that are adjacent in G, remove w from I' and instead add the dummy vertex neighbouring w on the path representing the edge $\{v, w\}$ to I'; update I accordingly. As a result, we obtain a maximum independent set I' such that $I = V(G) \cap I'$ is an independent set of G. Note that I' contains at most $|E(G)|$ dummy vertices. If $x \in I'$ or $y \in I'$, then $|I'| \leq |E(G)| + 1 \leq \alpha(G) + |E(G)|$, as no real vertex can belong to I'. Otherwise, also $|I'| \leq \alpha(G) + |E(G)|$. So we have $\alpha(G') \leq \alpha(G) + |E(G)|$.

Let now I be a maximum independent set of G. We take an independent set I' of G' in the following way: take all vertices in I, and for every edge $\{u, v\}$ in $E(G)$, we take on of the two dummy vertices corresponding to the edge: we can always take such a dummy vertex because either $v \notin I'$ or $w \notin I'$. So $\alpha(G') \geq |I'| = |I| + |E(G)| = \alpha(G) + |E(G)|$.

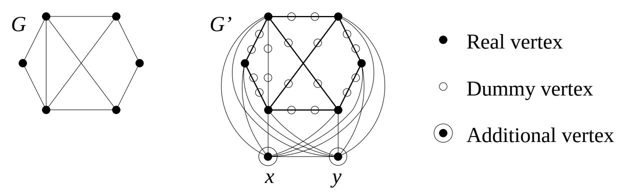

Figure 2: Graph G'

Also, we claim that $G' \in \mathcal{S}_1$. Let I be an independent set in G. We start by choosing $|E(G)|$ dummy vertices, not adjacent to vertices in I, as in the construction above: note that we can always do this, as all other vertices will have degree at least two (real vertices are adjacent to x and y, and x and y are adjacent to each other and at least one vertex in I; none of these is yet removed). At this moment, all vertices in I have degree two: they are only adjacent to x and y; all other vertices have degree at least two. Then, we can choose all vertices in I, and we end up with an independent set of size $|E(G)| + \alpha(G) = \alpha(G')$ (see also Figure 3).

Thus, the transformation, mapping (G, k) to $(G', k + |E(G)|)$ gives the required reduction, and the theorem follows. □

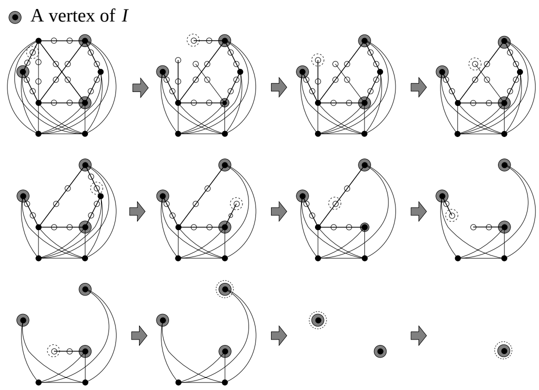

Figure 3: A sequence of steps for the MDG algorithm

As we have already mentioned, the recognition problem of S_r is a coNP-hard problem. In what follows, we will prove a stronger result for the recognition problem of S_1.

The complexity class **DP** is defined as the class of problems that can be expressed as a conjunction of two subproblems such that the one is in NP and the other in coNP (see [7]). An example of a **DP**-complete problem is Exact Vertex Cover, which asks, when given a graph G and a positive integer k, whether the size of the minimum vertex cover in G is exactly k. (See [3].) As the size of the vertex cover of a graph G equals to $|V(G)| - \alpha(G)$, it is clear that the following problem is also **DP**-complete.

Exact Independence Number
Instance: A graph G and a positive integer k.
Question: $\alpha(G) = k$?

Theorem 4.2 *The problem of determining whether a given graph G belongs to $\mathcal{S}_1$ is* **DP***-hard.*

Proof. We present a reduction from the Exact Independence Number. Given a graph G and a positive integer k, we will construct a graph G'' such that $G'' \in \mathcal{S}_2$ iff $\alpha(G) = k$.

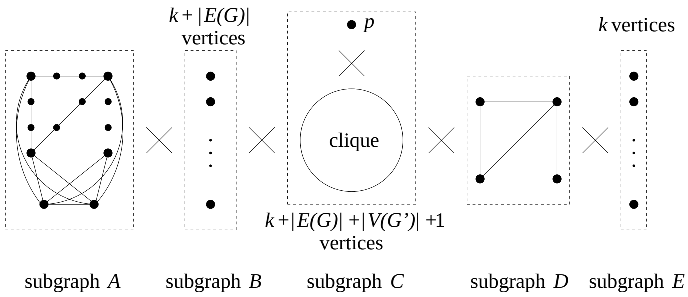

Figure 4: Graph G''.

The construction of G'' is as follows: First, let $G' \in \mathcal{S}_1$ be obtained from G, as in the proof of Theorem 4.1. Take a graph A, isomorphic to G'. Take a graph B consisting of $k + |E(G)|$ isolated vertices. Take a clique C with $k + |E(G)| + |V(G')| + 1$ vertices; distinguish an arbitrary vertex p from $V(C)$. Take a graph D, isomorphic to G. Take a graph E consisting of k isolated vertices. G'' is the graph with $V(G'') = V(A) \cup V(B) \cup V(C) \cup V(D) \cup V(E)$ and $E' = E(A) \cup E(C) \cup E(D) \cup \{\{u,v\} : u \in V(A), v \in V(B)\} \cup \{\{u,v\} : u \in V(B), v \in V(C) - \{p\}\} \cup \{\{u,v\} : u \in V(C) - \{p\}, v \in V(D)\} \cup \{\{u,v\} : u \in V(D), v \in V(E)\}$ (see Figure 4). (In other words, take the union of A, B, C, D, and E, and we add edges between vertices in A and vertices in B, between vertices in B and all vertices except p in C, between all vertices except p in C and vertices in D, and between vertices in D and vertices E.) It is easy to see that G'' can be constructed in polynomial time.

Now we show that $G'' \in \mathcal{S}_1$ iff $\alpha(G) = k$.

First, suppose $G'' \in \mathcal{S}_1$. Now, the MDG algorithm will start picking vertices in A and E, thus removing B and D. As $A \in \mathcal{S}_1$, $\alpha(A) = \alpha(G) + |E(G)|$ vertices in A will be chosen, and one vertex in C, and all k vertices from E. Thus, $\alpha(G'') = \alpha(G) + |E(G)| + k + 1$, using that $G'' \in \mathcal{S}_1$. As $V(B) \cup \{p\} \cup V(E)$ is an independent set of G'' with size $2k + |E(G)| + 1$, we have $\alpha(G'') \geq 2k + |E(G)| + 1$. It follows that $k \leq \alpha(G)$. Now suppose $k + 1 \leq \alpha(G)$. Then, consider an independent set consisting of $\alpha(G) + |E(G)|$ vertices in A, the vertex p and $\alpha(G)$ vertices in D. Thus, $\alpha(G'') \geq 2\alpha(G) + |E(G)| + 1 \geq (\alpha(G) + |E(G)| + k + 1) + 1$. This is a contradiction, since $\alpha(G'') = \alpha(G) + |E(G)| + k + 1$. So, $\alpha(G) = k$.

Now, suppose $\alpha(G) = k$. Any maximum independent set of G'' contains either $\alpha(A)$ vertices from A or all vertices from B, one vertex from C, and $\alpha(D)$ vertices from D or all vertices from E. Thus, $\alpha(G'') = 2k + |E(G)| + 1$. The MDG algorithm can output a set of this size: $k + |E(G)|$ vertices in A can be chosen (as in the proof of Theorem 4.1), p, and all vertices in E. Hence, $G'' \in \mathcal{S}_1$. □

We do not know whether the recognition problem for $\mathcal{S}_r$ is complete for **DP** for $r \geq 1$. Instead, we prove membership in the larger class $\Delta_2\mathbf{P}$. ($\Delta_2\mathbf{P}$ is the class of the problems that can be decided by a deterministic polynomial time oracle machine that uses an NP oracle). (See e.g. [3, 7].)

Lemma 4.3 *Let $r \geq 1$ be a rational number. The recognition problem for $\mathcal{S}_r$ belongs to $\Delta_2\mathbf{P}$.*

Proof. It is sufficient to see that for a given graph G, $G \notin S_r$ iff for some $k, 1 \leq k \leq n$: (i) $\alpha(G) \geq k$ and (ii) there is not any output of the MDG algorithm with at least k/r vertices. Finally, note that both (i) and (ii) can be answered by NP oracles. □

5 Open problems

We were unable to extend Theorem 4.2 to classes $\mathcal{S}_r$ for rational $r > 1$. Thus, it remains open to prove hardness for classes above NP for the recognition problems $\mathcal{S}_r$ with $r > 1$. Also, it is open whether the recognition problem of $\mathcal{S}_r$ is complete for **DP** or for some larger complexity class like $\Delta_2\mathbf{P}$, for all rational $r \geq 1$.

Acknowledgement

We thank Babette de Fluiter for discussions on this research.

References

[1] M. Bellare, Oded Goldreich, and M. Sudan, Free Bits, PCPs and Non-Approximability-Towards Tight Results, *Proc. 35th Annual IEEE Symp. on Computer Science* (1995), pp. 422-431.

[2] P. Berman and M. Fürer, Approximating Maximum Independent Set in Bounded Degree Graphs, in *Proc. Fifth ACM-SIAM Symp. on Discrete Algorithms* (1994), pp. 365-371.

[3] D. P. Bovet and P. Crescenzi, Introduction to the Theory of Complexity. Prentice Hall, 1993.

[4] V. Chvatal and P. J. Slater, A note on well-covered graphs, *Annals of Discrete Math.* 55 (1993), pp. 179-182.

[5] Y. Carol, A. Sebo, and M. Taris, Recognizing greedy structures, *J. Algorithms* 20 (1996), pp. 137-156.

[6] M. M. Halldorsson and J. Radhakrishnan, Greed is Good: Approximating Independent Sets in Sparse and Bounded-degree Graphs, in *Proc. 26nd Ann. ACM Symp. on Theory of Computing* (1994), pp. 439- 448.

[7] C.H. Papadimitriou. *Computational Complexity.* Addison - Wesley, 1994.

[8] D. Tankus and M. Tarsi, Well-Covered Claw-Free Graphs, *J. Combin. Theory Ser. B* 66 (1996), pp. 293-302.