Chemnitz University of Technology

Master Thesis

Efficient Approximation of Independent Sets in Graphs

Matthias Baumgart

Content

Introduction

- Boppana and Halldórsson developed an approximation algorithm for maximum independent set based on the Ramsey Theory.
- Idea: A graph on $R(k,l)$ vertices contains a clique of size k or an independent set of size l (or both).
- If the algorithm finds a large clique (and only a small independent set), then remove this clique and search on the remaining graph.

- The same holds for the maximum clique problem because in this case we have to remove independent sets.

Idea of U. Feige: Remove not only independent sets, but also sparse subgraphs.

Definition 1 *Let* $G = (V, E)$ *be a graph with a clique* C *of size* $|C| \geq n/k$. *A vertex induced subgraph* S *is called poor if it does not contain a clique* C_S *of size* $|C_S| \geq |S|/(2k)$.

Theorem 2 *Let* $G = (V, E)$ *be a graph with a clique* C *of size* $|C| \geq n/k$. *Let* $S_1, \ldots, S_l$ *be arbitrary disjoint poor subgraphs of* G. *Let* $G' = (V', E')$ *be the vertex induced subgraph of* G *that remains after removing the poor subgraphs. Then* $|V'| \geq n/(2k)$ *and* G' *contains a clique* C' *of size* $|C'| \geq |V'|/k$.

Proof:

∘ union $S = \bigcup_{i=1}^{l} S_i$ is a poor subgraph on at most n vertices and contains therefore no clique of size $n/(2k)$

∘ precondition: G contains a clique C of size $|C| \geq n/k$

$\Longrightarrow$ at least $n/(2k)$ vertices of C must be in G'

Assume: G' contains only cliques C' with $|C'| < |V'|/k$

$\Longrightarrow$ in this case there are at least

$$\frac{|V|}{k} - \frac{|V'|}{k} = \frac{|S|}{k}$$

vertices of C in the set S

$\Longrightarrow$ contradiction because S is poor $\square$

The Algorithm of Feige

The algorithm of Feige finds a clique of size

$$|C| \geq t \cdot \log_{3k}(n/t - 3)$$

whenever a graph $G = (V, E)$ has a clique of size n/k.

- ◦ It is divided into phases and iterations.
- ◦ Each phase consists of several iterations.

- The input to a phase is a graph $G' = (V', E')$ with a clique of size $|V'|/k$.
- After several iterations a phase ends and one of the following two conditions holds:

1. A clique C of size
   $$|C| \geq t \cdot \log_{3k} (|V'|/(6kt))$$
   is found.
2. A poor subgraph is found.

Each iteration works on a graph $G'' = (V'', E'')$ and processes the following steps:

1. If $|V''| < 6kt$, end the phase and output C.
2. Partition V'' into disjoint parts P_i, each with $2kt$ vertices.
3. In each part P_i consider all possible subsets S of vertices of cardinality t.
4. Let $N(S)$ be the set of vertices in $V'' \setminus S$ that are neighbors in G'' to every vertex in S. Call S *good* if the subgraph of G'' induced on S is a clique and $|N(S)| \geq |V''|/(2k) - t$.
5. If some set S is good then $C = C \cup S$ and go to the next iteration with the subgraph induced on $N(S)$.
6. Otherwise declare V'' as *poor*, and end the phase.

The Algorithm of Feige – Correctness

Theorem 3 *If a phase declares a set V'' poor, then indeed the subgraph of G induced on V'' does not contain a clique C of size*

$$|C| \geq \frac{|V''|}{2k} \; .$$

Proof: *Assume* $V'' = P_1 \cup \ldots \cup P_l$ contains a clique C with

$$|C| \geq \frac{|V''|}{2k} = t \cdot l$$

$\Longrightarrow$ by the pigeon-hole principle, at least one P_i will contain at least t vertices from this clique

$\Longrightarrow$ there is at least one S with the property *good* □

Theorem 4 *If a phase ends by outputting the set* C*, then this set contains at least*

$$|C| \geq t \cdot \log_{3k} \frac{|V'|}{6kt}$$

vertices, and these vertices form a clique in $G' = (V', E')$.

Proof:

- ○ obviously the set C is a clique
- ○ each – except the last – iteration adds t vertices to C
- ○ we have to lower bound the number of iterations

- the first iteration starts with $|V'|$ vertices
- a new iteration starts with at least $|V''|/(2k) - t$ vertices
- because of $|V''| \geq 6kt$ we have $t \leq |V''|/(6k)$
- the number of vertices of a new iteration is at least

$$\frac{|V''|}{2k} - \frac{|V''|}{6k} = \frac{|V''|}{3k}$$

- the number of vertices of iteration $i + 1$ is at least

$$\frac{|V'|}{(3k)^i}$$

- How many iterations are needed to reduce the number of vertices to $6kt$?

$$\begin{aligned} & \frac{|V'|}{(3k)^x} &<& \quad 6kt \\ \iff & \frac{|V'|}{6kt} &<& \quad (3k)^x \\ \iff & \log_{3k}\frac{|V'|}{6kt} &<& \quad x \end{aligned}$$

- There are at least $\log_{3k}(|V'|/(6kt))$ iterations and each of them adds t vertices to C, thus:

$$|C| \geq t \cdot \log_{3k}\frac{|V'|}{6kt} \qquad \square$$

The Algorithm of Feige – Running Time

- running time is polynomial bounded in n if $\binom{2kt}{t}$ is polynomial bounded in n
- choice of t affects the size of the clique C as well as the running time of the algorithm
- to maximize the size of the clique C and to achieve a polynomial running time we set

$$t = \Theta\left(\frac{\log n}{\log \log n}\right)$$

- in this case the clique C is of size

$$|C| = \Omega\left(\left(\frac{\log n}{\log \log n}\right)^2\right)$$

An Approximation Ratio of $O(n(\log\log n)^2/(\log n)^3)$

Given a graph $G = (V, E)$ with a clique C of size

$$|C| \geq \frac{n}{k}$$

case 1: $k \geq (\log n)^3$

- ◦ simply output a single vertex $v \in V$ to achieve an approximation ratio of $O(n/(\log n)^3)$

case 2: $k \leq \log n/(2\log\log n)$

- ◦ algorithm *IndependentSetRemoval* of Boppana and Halldórsson

- ◦ approximation ratio of $O(n\log\log n/(\log n)^3)$

case 3: $\log n/(2\log\log n) < k < (\log n)^3$

- the algorithm of Feige achieves only an approximation ratio of $O(n(\log\log n)^3/(\log n)^3)$
- if $k > \log n$ then the ratio is already $O(n(\log\log n)^2/(\log n)^3)$
- we have to save a factor of $\Omega(\log\log n)$

$\Longrightarrow$ modify the algorithm of Feige

$\Longrightarrow$ change the definition of a *good* subgraph *modification:*

- ○ call a subgraph S *good* if S is a clique and $|N(S)| > n_{test} - t$ holds, where n_{test} is the largest value still satisfying

$$n_{test} \le \frac{\log n_{test}}{2\log\log n_{test}} \cdot \frac{|V''|}{2k}$$

- ○ if $|V''|/(2k) - t \le |N(S)| \le n_{test} - t$ holds then apply the algorithm *IndependentSetRemoval* on $G[S \cup N(S)]$

  $\Longrightarrow$ if it finds a clique of size $(\log n_{test})^3/(6\log\log n_{test})$ join this clique to C and end the algorithm

  $\Longrightarrow$ otherwise do not consider S to be good

$\Longrightarrow$ approximation ratio of $O(n(\log\log n)^2/(\log n)^3)$

References

[1] R. Boppana und M.M. Halldórsson. Approximating Maximum Independent Sets by Excluding Subgraphs. *BIT*, 32(2):180–196, 1992.

[2] U. Feige. *Approximating Clique by Removing Subgraphs*. Manuscript, appears in SIAM Journal on Discrete Mathematics, 2002.

[3] M.M. Halldórsson und J. Radhakrishnan. A Still Better Performance Guarantee for Approximate Graph Coloring. *Information Processing Letters*, 45:19–23, 1993.