Densest Subgraph Problem

Anil Maheshwari

anil@scs.carleton.ca
School of Computer Science
Carleton University
Canada

Outline

Problem Statement

Greedy Peeling Algorithm

Problem Statement

Problem Statement

Input: A simple undirected graph $G = (V, E)$.

Aim: Output a subset of vertices $S \subseteq V$ that maximizes $\frac{|E(S)|}{|S|}$, where $E(S)$ is the set of edges in the induced subgraph on S.

Main Source: Chekuri, Quanrud and Torres, SODA 2022.

Note: We follow the description from the survey paper of Lanciano et al. from arXiv 2024.

Notation

- Graph $G = (V, E)$, where $V = \{v_1, \ldots, v_n\}$.
- For a subset $S \subseteq V$, let $G(S) = (S, E(S))$ be the graph induced by vertices in S.
- $E(S)$ is the subset of edges of E that are among the vertices in S. I.e., the edges in $G(S)$.
- Let $e(S) = |E(S)|$.
- Let $deg_S(v)$ is the degree of vertex v in $G(S)$.
- Density of the subgraph $G(S)$: $d(S) = \frac{e(S)}{|S|}$.
- $S^* \subseteq V$ has the maximum density. I.e., $d(S^*) \geq d(S)$ for any $S \subseteq V$.

Greedy Peeling Algorithm

Greedy Peeling Algorithm

Step 1: $S_n := V$ and $i := n$.

Step 2: While $i > 1$ do

1. $v_{\min} = \underset{v \in S_i}{\text{argmin}}\{deg_{S_i}(v)\}$.
   I.e., pick the vertex of minimum degree in S_i.
2. $S_{i-1} = S_i \setminus \{v_{\min}\}$.

Step 3: $S_{\max} = argmax\{d(S) | S \in \{S_1, \ldots, S_n\}\}$.

Step 4: Return $S_{\max}$

Main Claim

The density of $S_{\max}$ is at least $1/2$ of the density of an optimal subset of V. I.e. $d(S_{\max}) \geq \frac{1}{2} d(S^*)$.

Lower bound on degree of vertices in S^*

Claim 1

For any vertex $v \in S^*$, $deg_{S^*}(v) \geq d(S^*)$.

Proof:

- By definition $d(S^*) = \frac{e(S^*)}{|S^*|} \geq \frac{e(S^* \setminus \{v\})}{|S^*|-1}$.

- $e(S^* \setminus \{v\}) = e(S^*) - deg_{S^*}(v)$.

- We have $\frac{e(S^*)}{|S^*|} - \frac{e(S^* \setminus \{v\})}{|S^*|-1} \geq 0$.

$\Leftrightarrow \frac{e(S^*)}{|S^*|} - \frac{e(S^*) - deg_{S^*}(v)}{|S^*|-1} \geq 0$.

$\Leftrightarrow \frac{deg_{S^*}(v)}{|S^*|-1} - \frac{e(S^*)}{|S^*|(|S^*|-1)} \geq 0$.

$\Leftrightarrow \frac{deg_{S^*}(v)}{|S^*|-1} - \frac{d(S^*)}{|S^*|-1} \geq 0$.

$\Leftrightarrow deg_{S^*}(v) \geq d(S^*)$ $\square$

Competitive Ratio of Greedy Peeling

Main Claim

The density of $S_{\max}$ is at least $1/2$ of the density of an optimal subset of V. I.e. $d(S_{\max}) \geq \frac{1}{2}d(S^*)$.

Proof: GP removes vertices, and at some point it will remove a vertex from S^* for the first time.
- Let v^* be the first vertex in S^* that is removed by GP, and just before its removal let the set of remaining vertices be S'.
Observe

1. $S^* \subseteq S'$
2. For any vertex $v \in S'$, $deg_{S'}(v) \geq deg_{S'}(v^*) \geq deg_{S^*}(v^*)$.
   1st inequality follows from greedy choice.
   2nd inequality follows from the fact that $S^* \subset S'$ and v^* can have potentially more neighbors in S' compared to S^*.

Competitive Ratio of Greedy Peeling (contd.)

$$
\begin{aligned}
d(S') &= \frac{\frac{1}{2}\sum\limits_{v \in S'} deg_{S'}(v)}{|S'|} \\
&\geq \frac{\frac{1}{2}\sum\limits_{v \in S'} deg_{S'}(v^*)}{|S'|} \\
&= \frac{\frac{1}{2}|S'|deg_{S'}(v^*)}{|S'|} \\
&= \frac{1}{2}deg_{S'}(v^*) \\
&\geq \frac{1}{2}deg_{S^*}(v^*) \\
&\geq \frac{1}{2}d(S^*) \qquad \text{(By Claim 1)} \qquad \square
\end{aligned}
$$

Summary of Greedy Peeling

Theorem

Greedy Peeling runs in linear time and it outputs a subset whose density is at least $1/2$ of the density of an optimal densest subset.

Remarks:

1. There are examples where GP doesn't do better than $1/2$.
   Consider a graph that is a disjoint union of a bipartite graph $K_{d,D}$ and several disjoint K_{d+2}'s, where $D >> d$. An optimal solution consists of $K_{d,D}$ and its density is $\approx d$. In GP, first most of the vertices of $K_{d,D}$ will be removed followed by vertices in cliques. Competitive ratio of GP $\approx \frac{d}{2}$
2. An optimal densest subset can be found in polynomial time using network flow.
3. The problem of finding densest subset that has exactly k vertices, called the densest k-subset problem, is $\mathbf{NP}$-Hard and can't be approximated within a constant factor.

References

1. T. Lanciano, A. Fazzone, A. Miyauchi, and F. Bonchi, A Survey on the Densest Subgraph Problem and Its Variants, arXiv:2303.14467v2 [cs.DS] 18 Apr 2024.
2. C. Chekuri, K. Quanrud, and M. Torres, Densest subgraph: Supermodularity, iterative peeling, 2022 ACM-SIAM Symposium on Discrete Algorithms.
3. A. Goldberg, Finding a maximum density subgraph, Technical Report, University of California at Berkeley, 1984.