Dominating set

A distributed algorithm that outputs a dominating set of with approximation ration of at most $O(\log \Delta)$ in $O(\log n \log \Delta)$ rounds (assuming that all nodes know both n and Δ).

Special case of set cover problem: Given a ground set X and a collection of non-empty subsets of X with positive costs $c(S)$, the goal is to cover the ground set by a minimum cost subcollection of subsets.

Greedy Algorithm:

Select the largest degree vertex at each step in the residual graph.

Remove the vertex and its incident edges.

This gives a $O(\log \Delta)$ approximation.

Analysis of Greedy

Greedy chooses a set (vertex) at each step that realizes the minimum unit price, i.e., $p(e) = c(S)/|S|$.

To compute the cost of an algorithm we do the following accounting: when the algorithm picks a set, its cost is distributed equally to all elements it covers.

Each element is assigned a price only once, at the time it is covered by the algorithm.

For a subset $A \subseteq X$, let $g(A) = \sum_{e \in A} p(e)$.

Then $g(X)$, the sum of the unit prices, is the total cost incurred by greedy.

Theorem 1. *Approx ratio of greedy is* $O(\log \Delta)$.

Proof

For any set S, we can show that $g(S) \leq H_{|S|} c(S)$.

Sort the elements of S according to the time when they are covered by greedy breaking ties arbitrarily.

Let $e_1, \ldots, e_k$ be this numbering.

When greedy covers e_i, $p(e_i) \leq c(S)/(k - i)$.

We have, for any two sets A and B, $g(A \cup B) \leq g(A) + g(B)$.

Denoting with C^* an optimal cover, we have

$g(X) = g(\cup_{S \in C^*} S) \leq \sum_{S \in C^*} g(S)$

$\leq \sum_{S \in C^*} H_{|S|} c(S)$

$\leq max_S H_{|S|} \sum_{S \in C^*} c(S)$

$\leq max_S H_{|S|} OPT.$

Distributing Greedy

Greedy is sequential and a naive implementation can take $\Omega(\sqrt{n})$ time.

A candidate set is any set S such that $m \leq c(S)/|S| \leq 2m$, where m is the candidate set that gives the min cost at this step.

Adding any **one** candidate set still gives a $O(\log \Delta)$ approximation.

Adding all sets speeds up the algorithm, but destroys the approximation.

Idea: Come with a accounting scheme that distributed costs among the elements in a manner similar to (sequential) greedy.

For every set S selected by greedy, the cost $C(S)$ will be distributed among a subset $T \subseteq S$ of at least $|S|/4$ elements of T.

Elements of T will be charged only once.

Distributed Algorithm

Synchronous model.

Think of the process of choosing the candidate set as an election.

Elements vote for one of the candidate sets containing them.

Election rules:

1. A random permutation of the candidates is computed.

2. Among all the candidate sets that contain it, each voter votes for that set which has the lowest number in the permutation.

3. A candidate is elected if it obtains at least 1/4 of the votes of its electorate.

4. Elected candidates enter the dominating set.

The cost of the set can now be distributed equally among the elements that voted for it.

The algorithm is a sequence of $\log \Delta$ phases.

In phase $i = 1, 2, \ldots, \log \Delta$, the max degree of the graph is at most $\Delta/2^{i-1}$.

The candidates during phase i are all those vertices whose degree is in the interval $(\Delta/2^i, \Delta/2^{i-1}]$.

Initially all vertices are free. When a vertex becomes a dominator it does not participate in the algorithm any longer.

The voters are free nodes. Candidate can be free or dominated vertices.

Each phase consists of $O(\log n)$ elections whp.

Totally $O(\log n \log \Delta)$ rounds whp.

If Δ is not known then $O(\log^2 n)$ rounds.

Analysis

Bipartite graph: candidates on one side and its neighbors on the other.

Neighbors of a candidate c — electorate of c.

Neighbors of a voter v – pool of v.

We will show that the expected number of edges that are removed from the bipartite graph is a constant fraction in each election.

This will imply that $O(\log n)$ elections are enough to end a phase whp.

A voter is influential for a candidate c if at least $3/4$ of the voters in c's electorate have degree no greater than that of v. Let $d(v)$ denote the degree of v in the bipartite graph.

Lemma 1. *For any two voters v and w, $d(v) \geq d(w)$, in c's electorate, $\Pr(w \text{ votes } c | v \text{ votes } c) \geq 1/2$.*

Proof: N_b — the number of neighbors that v and w have in common.

N_v — number of neighbors of v that are not neighbors of w.

N_w — number of neighbors of w that are not neighbors of v.

$\Pr(w \text{ votes } c | v \text{ votes } c) = \frac{\Pr(w \text{ votes } c \text{ and } v \text{ votes } c)}{\Pr(v \text{ votes } c)} = \frac{1/(N_v + N_b + N_w)}{1/N_v + N_b} \geq 1/2$

Lemma 2. *Let v be an influential voter for c. Then* $\Pr(c$ *is elected* $|v$ *votes* $c) \geq 1/6$.

$X =$ number of votes for c.

$Y = c - X$. (c is the size of the electorate of c).

$E[X|v \text{ votes } c] \geq \sum_{w:d(w) \leq d(v)} \Pr(w \text{ votes } c|v \text{ votes } c)$

$\geq 3c/8$.

Hence,

$\Pr(c \text{ not elected } |v \text{ votes } c) = \Pr(X < c/4|v \text{ votes } c)$

$= \Pr(Y \geq 3c/4|v \text{ votes } c)$

$\leq E[Y|v \text{ votes } c]/(3c/4) = 4(c - E[X|v \text{ votes } c])/(3c)$

$\leq 5/6$.

Lemma 3. *In a phase, let m denote the total number of edges in the bipartite graph at any stage in this phase.*

Let X denote the number of edges removed from the bipartite graph after one election. Then $E[X] \geq m/24$.

Proof. An edge (v, c) is good is v is influential for c. At least $1/4$ edges are good.

$$E[X] \geq \sum_{(v,c)} \Pr(c \text{ is elected and } v \text{ votes } c) d(v)$$

$$\geq \sum_{(v,c) \text{ is good}} \Pr(c \text{ is elected and } v \text{ votes } c) d(v)$$

$$= \sum_{(v,c) \text{ is good}} \Pr(v \text{ votes } c) \Pr(c \text{ is elected } | v \text{ votes } c) d($$

$$= \sum_{(v,c) \text{ is good}} \Pr(c \text{ is elected } | v \text{ votes } c)$$

$$\geq m/24.$$