Lecture 9 – March, 2016

Prof. Ankur Moitra *Scribe: Trevor Henderson, David Huang*

1 Last Time: Capacity Scaling and Min-Cost Matching

Capacity Scaling gave us a mechanism to reduce the dependence from the max capacity u (pseudopolynomial) to $\log u$ (weakly polynomial). Capacity Scaling improved on Ford-Fulkerson by choosing a better than arbitrary augmenting path. Today, we will use the same approach to find a weakly polynomial algorithm for the min-cost flow problem.

2 Min-Cost Flow and the Goldberg-Tarjan Algorithm

2.1 Klein's Cycle-Canceling Algorithm

Recall that the algorithm is as follows:

Algorithm 1 Klein's Cycle-Canceling Algorithm

Find max s-t flow
while there exists a negative cost cycle $\mathcal{C}$ in the residual **do**
  Augment along $\mathcal{C}$ by bottleneck $\Delta_H(\mathcal{C})$
  Update the residual
end while

Last time with Min-Cost perfect matching we saw that a perfect matching M is a min-cost perfect matching if and only if its residual does not contain a negative cost cycle. For Min-Cost Flow, the analogue is the following:

Lemma 1. *A max s-t flow has a minimum cost (among all max s-t flows) if and only if there is no negative cost cycle in the residual*

With this in mind we need to answer the following questions:

1) How can we find a negative cost cycle if there is one?

   The answer is Bellman-Ford! (not Dijkstra's)

   Recall that Bellman-Form computes the cheapest path $s \to v$ for all vertices v, and works with negative cost edges. Moreover, if there is any negative cost cycle Bellman-Ford algorithm will find it. The running time is $O(mn)$ on a graph with m vertices and m edges.

2) How many iterations does Klein's algorithm take?

Assume all capacities and costs are bounded by U and C respectively. (Note that C is an upper bound on the absolute value of costs, since they can be negative).

At each step of Klein's algorithm the cost in the residual decreases by at least 1. The initial cost in the residual is 0 and the final cost is $\geq -mCU$, therefore Klein's algorithm takes at most mCU iterations.

Thus Klein's Cycle-Canceling algorithm has a running time of $O(m^2nCU)$ which is pseudopolynomial. With a small adjustment to the algorithm we will see that we can reduce this to be weakly polynomial.

2.2 Goldberg-Tarjan

The Goldberg-Tarjan algorithm modifies Klein's Cycle-Canceling algorithm by specifying which negative cost cycle should be used to augment the flow in each iteration. One might think to choose the cycle with the most negative cost, but finding that cycle is an **NP**-hard problem. Goldberg-Tarjan says to instead choose the cycle of minimum *mean* cost.

Algorithm 2 Goldberg-Tarjan algorithm

```
Find max s-t flow
while there exist negative cost cycles in the residual do
    Choose the negative cost cycle C that minimizes c_H(C)/|C|
    Augment along C by bottleneck Δ_H(C)
    Update the residual
end while
```

To analyze this algorithm we need a notion of *reduced cost*. For any potential function $p : V \to \mathbb{R}$, we define the reduced cost to be

$$c_p(u, v) = c(u, v) + p(u) - p(v)$$

Intuition: Let $\mathcal{C}$ be a cycle in the residual, then $\sum_{(u,v)\in\mathcal{C}} c(u, v) = \sum_{(u,v)\in\mathcal{C}} c_p(u, v)$. The reason is that the sum of the reduced costs along a cycle telescopes to be the sum of the original costs. Thus if there exists some potential function p where $c_p(u, v) \geq 0$ for all edges $(u, v) \in E$ then there is no negative cost cycle in the residual (with respect to the original cost function c).

We will prove this intuition as part of a much stronger statement. Let $\mu = \min\left(\frac{c_H(\mathcal{C})}{|\mathcal{C}|}\right)$ over all cycles $\mathcal{C}$ in the residual. Let $\epsilon = \max \gamma$ for which there exists some potential p such that $c_p(u, v) \geq \gamma$ for all edges $(u, v) \in E$.

Lemma 2. $\mu = \epsilon$.

Claim 1. $\mu \geq \epsilon$.

Proof. As we mentioned earlier, for any cycle $\mathcal{C}$, $c(\mathcal{C}) = c_p(\mathcal{C})$. By definition of ϵ, there exists a potential function p such that $c_p(u, v) \geq \epsilon$ for all edges $(u, v) \in A$. Therefore for all cycles $\mathcal{C}$,

$$\frac{c_H(\mathcal{C})}{|\mathcal{C}|} = \frac{c_p(\mathcal{C})}{|\mathcal{C}|} = \frac{1}{|\mathcal{C}|} \sum_{(u,v)\in\mathcal{C}} c_p(u, v) \geq \epsilon,$$

which implies that $\mu \geq \epsilon$. □

Claim 2. $\mu \leq \epsilon$.

Proof. Define a helper cost function c' as follows:

$$c'(u,v) = c(u,v) - \mu$$

By definition of μ, c' gives no negative cost cycles. Define the potential function $p(u)$ be the cost of cheapest path according to c' from s to v. By the definition of shortest path $p(v) \leq p(u) + c'(u,v)$. Rearranging:

$$\begin{aligned} 0 &\leq c'(u,v) + p(u) - p(v) \\ &= \underbrace{c(u,v) + p(u) - p(v)}_{c_p(u,v)} - \mu \end{aligned}$$

So $\mu \leq c_p(u,v)$, which concludes the proof. □

Next we show that ϵ (which is a negative number), is making progress towards 0, at which point there will be no minimum cost cycles in the residual. Let f be a flow and define $\epsilon(f)$ to be the value of ϵ on the residual graph of f.

Lemma 3. *Let f' be the flow obtained from f by canceling a minimum mean cost cycle in residual of f. Then $\epsilon(f') \geq \epsilon(f)$.*

Proof. Let p be a potential such that $c_p(u,v) \geq \epsilon$ for all edges (u,v). Recall that $\epsilon < 0$, since we only augment over negative cycles. Consider a minimum mean cost cycle in the residual graph of f. By Lemma 2, $\epsilon(f) = \mu$ and hence the reduced cost of all edges in the cycle is ϵ. After augmenting over the cycle, we will get some new backward edges. We show that using the same potential function p, $\epsilon(f') \geq \epsilon(f)$. To see this, the reduced cost of each newly added (backward) edge vu is $-c(u,v) + p(v) - p(u) = -c_p(u,v)$. Since ϵ is a negative value, all reduced costs in the residual of f' remain greater than or equal to ϵ with respect to p. Thus $\epsilon(f') \geq \epsilon(f)$. □

Let $\epsilon(f)$ denote the value of ϵ for the residual graph for flow f.

Lemma 4. *Let f be a maximum s-t flow and let f' be the flow obtained by m iterations of the Goldberg-Tarjan rule. Then, $\epsilon(f') \geq \left(1 - \frac{1}{n}\right)\epsilon(f)$.*

Proof. Let p be a potential such that $c_p(u,v) \geq \epsilon(f)$ for all edges (u,v) in the residual of f.

If we ever augment along a cycle with non-negative reduced cost, then the average cost of the cycle is at least $\epsilon\left(1 - \frac{1}{n}\right)$. Note that the average reduced cost a cycle is equal to the actual average cost of the cycle (and choice of p does not matter) and by Lemma 2, $\epsilon(f') \geq \epsilon\left(1 - \frac{1}{n}\right)$

Otherwise, in each iteration we replace at least one arc with negative reduced cost ϵ with one with positive cost in the opposite direction. Thus After at most m iterations we run out of arcs with negative reduced costs. □

Now we are ready to prove the main theorem.

Theorem 1. *Let $C = \max_{(u,v)\in E} |c(u,v)|$. The number of iterations in the Goldberg-Tarjan algorithm is at most $O(mn \log nC)$*

Proof. By setting $p = 0$, is straightforward to check that $\epsilon(f) \geq -C$. Applying Lemma 4,

$$\epsilon(f_{\text{last}}) \geq \epsilon(f_{\text{first}}) \left(1 - \frac{1}{n}\right)^{2n \ln(nC)} \geq -C \cdot e^{-2\ln(nC)} = \frac{-C}{n^2C^2} > -\frac{1}{n}.$$

Since all the original costs are integers, this implies there is no negative cost cycle. □

3 Application

Suppose we want to form a matching between sets of buyers B and houses H. It is only possible for a buyer to buy some subset of the houses due to price, location, etc. Additionally, each buyer b has a value associated with each of the houses h they can buy; $v(b,h)$. In many cases greedy buyers will buy houses that lead to a matching with sub-optimal total value and some buyers will be left without any houses to buy at all, stalling the market.

Is there some way for some higher power, like the government, to assign prices to the houses such that if the buyers are acting greedily they form a perfect matching that maximizes total value?

We can represent this problem as a Min-Cost Matching problem between the buyers and houses with directed edges from each buyer b to each of the houses they can buy h with cost $c(b,h) = -v(b,h)$. By termination condition of Goldberg-Tarjan algorithm, there is a minimum cost perfect matching M and a potential function $p : B \cup H \to \mathbb{R}$ such that:

$$\begin{aligned} &c_p(b,h) = c(b,h) + p(b) - p(h) = 0 && \forall (b,h) \in M \\ &c_p(b,h) > 0 && \forall (b,h) \in E \setminus M \end{aligned}$$

Set the price of house h to be $-p(h)$. A greedy buyer will buy house h if it maximizes his total value:

$$\underbrace{v(b,h')}_{\text{value}} + \underbrace{p(h')}_{\text{minus price}} < v(b,h) + p(h) \qquad \forall h' \neq h$$

By rearranging and substituting:

$$c(b,h) - p(h) < c(b,h') - p(h') \Rightarrow c_p(b,h) - p(b) < c_p(b,h') - p(b) \Rightarrow 0 = c_p(b,h) < c_p(b,h')$$

So for these prices everyone will agree on the maximum utility solution determined by M.

References

[1] Klein, M. A primal method for minimal cost flows with applications to the assignment and transportation problems. In *Management Science 14.3*. USA, 1967, pp. 205-220.

[2] Goldberg, A. V., and Tarjan, R. E. A new approach to the maximum flow problem. In *Proceedings of the 18th ACM Symposium on Theory of Computing*. AMC, New York, 1986, pp. 136-146.