Princeton University Spring '07
COS 594 : Limits on Approximation

Lecture 12: Directed EDP with Congestion

Lecturer: *Moses Charikar* Scribe: *Rajsekar Manokaran*

1 Overview of the lecture

In this lecture, we will study the directed edge-disjoint paths (EDP) problem with congestion. This problem, similar to the undirected EDP problem we studied earlier, asks for the maximum number of source-sink pairs that can be routed through a directed graph such that no edge is used by more c pairs. We will first show that the integrality gap of a natural LP relaxation is $\Omega\left(n^{\frac{1}{3c+11}}\right)$ for constant c. Then, we will extend the result to prove that the problem is hard to approximate within a factor of $\Omega\left(n^{\frac{1}{3c+19}}\right)$. A stronger version of this result was proven independently by Chuzhoy and Khanna[1] and Guruswami and Talwar[2]. The presentation here is from the paper by Chuzhoy and Khanna.

2 Linear Programming Relaxation

Before giving an LP relaxation, we state the problem formally.

Definition 1 (EDPwC) *Given a graph $G(V, E)$, a collection of source-sink pairs $(s_1, t_1), \dots, (s_k, t_k)$ and an integer c, route as many pairs as possible such that* less *than c paths go through each edge. The quantity c is called the congestion. The performance of algorithms for EDPwC is compared against the optimal solution with congestion 1.*

We have the following natural LP relaxation using multi-commodity flows:

$$\max \sum_i x_i \tag{1}$$

such that

$$x_i - \sum_{P \in P_i} f(P) = 0 \qquad 1 \le i \le k \tag{2}$$

$$\sum_{P : e \in P} f(P) \le 1 \qquad e \in E \tag{3}$$

$$x_i, f(P) \in [0, 1] \qquad 1 \le i \le k, P \in \mathcal{P} \tag{4}$$

Note that c does not appear in the above LP as we allow the fractional solution to have congestion at most 1. We will show the integrality gap by constructing a graph where the integral solution can route a small fraction of the vertices whereas the LP solution can route a $1/c$ fraction of every pair.

3 Integrality Gap

3.1 Gap Instance

We will now prove an integrality gap of $\Omega\left(n^{\frac{1}{3c+11}}\right)$ for the above LP relaxation. The gap instance we will construct will be a layered graph with the sources occupying the first layer and the last layer consisting of the sinks. Each intermediate layer is made up of a collection of blobs. Each blob has some "input" vertices to receive edges from the previous layer and "output" vertices which connect to the next layer. Edges in the graph are either within a blob or between layers.

More formally, we will have a source-sink pair for every pair (i, y) for $i \in [m]$ and $y \in Y$ for some set of labels Y and integer m which are yet to be chosen. Sources (and similarily sinks) of the form (i, y) will be called type-i sources (or sinks). As described above, the sources and the sinks occupy the first and the last layers. We will have Z intermediate layers each of which contains a blob for every label. Thus, we have $Z|Y|$ blobs in total (and $|Y|$ blobs per layer). Further, each blob has m input vertices and m output vertices.

3.2 Labeling Scheme

The set of labels Y will be $[2m^2Z] \otimes [2m^2Z]$ and hence a 2-dimensional vector space over $[2m^2Z]$. The component-wise addition modulo $2m^2Z$ gives a natural way to "add" two labels to obtain a new label. The collection of labels u_i, where $u_i = (i, i^2)$ for $1 \leq i \leq m$ will be called the increment vectors.

Let us denote the sources (i, y) (resp sinks) by $S(i, y)$ (resp $(T(i, y))$. Similarily, denote the blob in layer j and with label y by $B(j, y)$. The source $S(i, y)$ will be connected to the ith input vertex of $B(1, y + u_i)$ by an edge; the ith output of $B(1, y + u_i)$ will be connected to $B(2, y + 2u_i)$ the corresponding output being connected to $B(3, y + 3u_i)$ and so on till finally we connect the blob in the last layer to $T(i, y)$. More generally, if we (abuse notation and) denote the ith vertex of $B(j, y)$ by $B(j, y, i)$, then we have:

1. edges $B(j, y, i) \to B(j + 1, y + u_i, i)$ for all $1 \leq j \leq Z$, $y \in Y$ and $1 \leq i \leq m$
2. edges $S(i, y) \to B(1, y + u_i, i)$ for all $1 \leq i \leq m$ and $y \in Y$
3. edges $B(Z, y + Zu_i, i) \to T(i, y)$

3.3 Constructing the blobs

All that is left to complete the construction is to describe the edges within a blob. Note that each blob has m input vertices and m output vertices. We randomly group the m input vertices into m/c groups of size c each. For each group, introduce a new "special" edge and connect all the input vertices of the group to one side of the edge and the corresponding output vertices to the other side.

4 Proof of the Gap

4.1 Canonical Paths

From the construction, we immediately have the path $S(i,y) \to B(1,y{+}u_i) \ldots \to B(j,y{+}ju_i) \ldots \to T(i,y)$ called the canonical path $P(i,y)$. The following lemma proves that $P(i,y)$ is the only path that connects $S(i,y)$ to $T(i,y)$

LEMMA 1
For $k \leq Z+1$, let $u_{i_1}, u_{i_2} \ldots u_{i_k}$ and u_j be increment vectors such that $ku_j = u_{i_1} + u_{i_2} \ldots + u_{i_k}$. Then, $u_j = u_{i_1} = \ldots = u_{i_k}$.

PROOF: Since any component of an increment vector is at most m^2, the sum of $Z+1$ of them will strictly be less than $2m^2Z$. Thus, it is enough to prove the lemma for standard vector addition over the integers. Writing $u_j = \frac{u_{i_1}+u_{i_2}\ldots+u_{i_k}}{k}$, we see that u_j is a convex combination of the points $u_{i_1}, u_{i_2} \ldots u_{i_k}$. Since the curve (x, x^2) is strictly convex, the statement of the lemma immediately follows. $\square$

COROLLARY 2
The canonical path $P(i,y)$ is the only path from $S(i,y)$ to $T(i,y)$

REMARK 1 The construction and hence the hardness we will describe also applies to the All-or-Nothing flow problem. Here, the constraint that the pairs should be routed through a single path is relaxed to just requiring that a total flow of one unit between each pair. Due to the above corollary, this problem is no easier on the instance we consider.

4.2 Bounding Integral Solutions

The size of our construction is $O(mZ\,|Y|) = O(m^5Z^3)$. Further, the LP can route a $1/c$ fraction of flow through each canonical path hence attaining $m\,|Y|\,/c$ units of flow (fractionally) with congestion 1.

Suppose the integral solution routes more than $8c\,|Y|$ with congestion strictly less than c. This means that this many pairs are routed through each layer of blobs. By averaging, we want to lower bound the number of blobs (in each layer) that route many (say $4c$) pairs. Indeed, if x is the number of blobs that route $4c$ pairs, then $xm + |Y|\,(4c) \geq 8c\,|Y|$. Hence, at least $4c\,|Y|\,/m$ blobs in each layer route $4c$ pairs each. We call these blobs the "good" blobs.

Next, we want to upper bound the probability that a fixed good blob is not c-congested. Seeing the edges in the blobs as choosing a random c subset out of the m input vertices, the probability that the first edge choses all of them from the set (of size at least $4c$) of pairs routed through the blob is:

$$\frac{\binom{4c}{c}}{\binom{m}{c}} = \frac{(4c)(4c-1)\cdots(3c+1)}{m(m-1)\cdots(m-c+1)} \geq \left(\frac{3c}{m}\right)^c$$

Since we have Z layers and at least $4c\,|Y|\,/m$ good blobs in each layer, the probability that none of them is congested is

$$\left[1-\left(\frac{3c}{m}\right)^c\right]^{4c|Y|Z/m} \leq e^{-3^c 4|Y|Z(3c/m)^{c+1}} = e^{-O\left(\frac{|Y|Z}{m^{c+1}}\right)}$$

The number of such solutions is at most $2^{m|Y|} < e^{m|Y|}$. Hence, setting $Z = O(m^{c+2})$, we get (the existence of) a construction of size m^{3c+11} which achieves an integrality gap of $\frac{m|Y|}{c}/8c\,|Y| = O(m)$. This proves an integrality gap of $\Omega(n^{1/(3c+11)})$.

5 Hardness of approximating directed EDPwC

Now, we will extend the above result to obtain a hardness result for the approximation of the problem. We will perfrom a reduction from the independent set problem. Our starting point is the following result of Hastad[3].

THEOREM 3
For any $\epsilon > 0$, it is hard to distinguish between graphs with an independent set of size $n^{1-\epsilon}$ and graphs with no independent set of size n^{ϵ} unless **NP** $\subseteq$ **ZPP**.

5.1 Reduction from Independent Set Instance

Given an instance, G, of the independent set problem, we want to construct an instance of the directed EDP problem such that if G has a large independent set (and hence is an YES instance for the gap problem), then we can route at least an $\Omega(1/m^{\epsilon})$-fraction with congestion 1. On the other hand, if G is a NO instance, no more than $O(1/\sqrt{m})$-fraction of pairs can be routed.

The construction will be similar to the integrality gap construction. We will set m to be the number of vertices in G (the independent set instance) and we will chose Z later. The blobs will be constructed in a slightly different manner (from the graph G). If we see the type-i source-sinks as representing vertex i in G, then since each input to a blob is of a different type, the input vertices of any blob correspond to the vertex set of G. As before, we will group the m input vertices randomly into m/c groups of size c each. For each group of c vertices, if the corresponding vertices in G form a clique, we proceed in the usual way and connect all of them to a special edge inside the blob. On the other hand, if the group does not form a clique in the independent set instance, we connect them using c special edges (thus allowing them to be routed through the blob without any congestion).

5.2 Proof of Gap between YES and NO Instances

In the YES instance, we have an independent set S in the graph of size $m^{1-\epsilon}$. Then, we route the source-sink pair (v, y) for all $v \in S$ and $y \in Y$ concurrently through their respective canonical paths thus routing $m^{1-\epsilon}|Y|$ pairs.

We will show that in the NO instance, any subset of $4\,|Y|\,\sqrt{m}$ source-sink paths cause congestion c with high probability. To get an handle on the probability of congestion in a blob, we need to lower bound the number of c-cliques in a graph with small maximum independent set. For $c \geq 2$ and an integer s, $T(\alpha, c)$ denote the minimum number of c-cliques in a graph with α vertices and no independent set of size s. We will prove the following simple but arcane lemma.

Lemma 4
For $\alpha > (4s)^c$, $T(\alpha, c) \geq \frac{\alpha^c}{(2c)^c (4s)^{c^3}}$.

Proof: Base case: $c = 2$: For $\alpha > (4s)^2$, let the average degree be d. Then, by averaging argument, we have a set of size at least $\alpha/2$ with degree at most $2d$. Restricting our attention to those vertices, since each such vertex can connect to at most $2d$ more vertices in the group, we have an independent set of size at least $\alpha/2(2d+1)$. Thus, we have $d > \alpha/4s - 1/2$. The number of cliques of size 2 is $\alpha d/2 > \alpha^2/10s$

For the induction step, observe that at least $\alpha/2$ vertices must have degree at least $\alpha/2s - 1$ as otherwise we would have an independent set of size $\alpha/2/(\alpha/2s) = s$. Thus at least half the vertices have degree at least $\alpha/2s - 1 \geq \alpha/4s$. The neighbourhood of each such vertex consists of at least $\alpha/4s > (4s)^{c-1}$ vertices and hence contains at least $T(\alpha/4s, c-1)$ cliques of size $c-1$. Thus, summing over all such $\alpha/2$ vertices, and taking into account the fact that each c-clique may be counted at most c times, we get $T(\alpha, c) \geq \frac{\alpha}{2c} T(\alpha/4s, c-1)$. Iterating, we get

$$\begin{aligned}
T(\alpha, c) &\geq \frac{\alpha}{2c} \frac{\alpha/4s}{2(c-1)} \cdots \frac{\alpha/\left(4s\right)^{c-3}}{6} T(\alpha/\left(4s\right)^{c-2}, 2) \\
&\geq \frac{\alpha^{c-2}}{(2c)^{c-2}\left(4s\right)^{c^2/2}} T(\alpha/\left(4s\right)^{c-2}, 2) \\
&\geq \frac{\alpha^c}{(2c)^c \left(4s\right)^{c^3}}
\end{aligned}$$

□

For the NO instance, we will use the above lemma with $\alpha = \sqrt{m}$ and $s = m^\epsilon$. For some $\epsilon < 1/(3c^3)$, thus giving us the following corollary.

Corollary 5
Any graph H on $\alpha = \sqrt{m}$ vertices that does not contain an independent set of size $s = m^\epsilon < m^{1/(3c^3)}$ has at least $\Omega\left(m^{c/2-1/3}\right)$ distinct cliques of size c.

Now we prove that in the NO case, less than $4\left|Y\right|\sqrt{m}$ pairs can be routed. The proof is by contradiction and is similar in essence to the proof of the integrality gap. Assume that there is a solution that routes at least $4\left|Y\right|\sqrt{m}$ pairs. As in the proof of the integrality gap, we lower bound the "good" blobs that route many (here, $2\sqrt{m}$) pairs through them.

Claim 6
For each layer, the fraction of blobs which are good is at least $2/\sqrt{m}$.

Proof: Let x denote the fraction of good blobs. Then, we have $m.x|Y| + 2\sqrt{m}.|Y| \geq 4\left|Y\right|\sqrt{m}$ thus giving the required result. □

Next, we lower bound the probability that no edge in any good blob has congestion c.

Claim 7
The probability that no edge in any good blob has congestion c is at most exp $\left(-\Omega\left(\frac{1}{m^{\frac{c}{2}-\frac{1}{6}}}\right)\right)$

Proof: Let S be the set of vertices corresponding to the pairs routed through a good blob. Then $|S| \geq 2\sqrt{m}$. As before we see the random grouping as being done after the flows were routed. Let S_i denote the vertices in group i (for $1 \leq i \leq m/c$). For $i \leq \sqrt{m}/c$, $S - \cup_{j<i} S_j$ has atleast $\sqrt{m}$ vertices. Thus, the number of cliques that would cause a congestion of c is $\Omega\left(m^{c/2-1/3}\right)$. Thus, the probability that none of these $\sqrt{m}/c$ groups cause congestion c is:

$$\Pr[\text{no congestion in a good blob}] \leq \left[1 - \Omega\left(\frac{m^{c/2-1/3}}{m^c}\right)\right]^{\sqrt{m}/c} \leq e^{-\Omega\left(\frac{1}{m^{\frac{c}{2}-\frac{1}{6}}}\right)}$$

$\square$

Since we have a $2/\sqrt{m}$ fraction of good blobs in each layer and Z layers, the total probability is bounded by $\exp\left(-\Omega\left((|Y|\,Z)/m^{\frac{c}{2}+\frac{1}{3}}\right)\right)$. Since the number of possible solutions is at most $2^{|Y|m}$ setting $Z = O(m^{c/2+1/3+1})$, we have an instance of size $O(m^{9+3c/2})$ with a gap of $m^{1/2-\epsilon}$. Thus we have the following theorem.

Theorem 8
For any fixed positive constant $c \geq 1$, directed EDP with congestion c is hard to approximate within a factor of $\Omega(n^{\frac{1}{3c+18}})$ unless **NP** $\subseteq$ **ZPP**.

Remark 2 The above construction works for non-constant c upto $\log^{\lambda} N$ for some fixed λ although we will lose a little in the hardness. We can get a hardness of $\Omega\left(n^{\frac{1}{3c+20}}\right)$ for such c.

Remark 3 The papers that prove the above results also prove, using a different technique, a hardness of $n^{\Omega}(1/c)$ for c upto $\alpha \log n / \log\log n$ for some absolute constant α.

References

[1] Julia Chuzhoy and Sanjeev Khanna. Hardness of directed routing with congestion. `http://www.math.ias.edu/~cjulia/dir-multicut2.ps`.

[2] Venkatesan Guruswami and Kunal Talwar. Hardness of low congestion routing in directed graphs. `http://eccc.hpi-web.de/eccc-reports/2006/TR06-141/index.html`.

[3] J. Hastad. Clique is hard to approximate within $n^{1-\epsilon}$. In *FOCS '96: Proceedings of the 37th Annual Symposium on Foundations of Computer Science*, page 627, Washington, DC, USA, 1996. IEEE Computer Society.