CS 270 - Homework 4

Walid Krichene (23265217)

(1.a) Consider the linear program for maximum weight bipartite matching. Show that the optimum may not be integer for non bipartite graphs.

answer Let $G = (V, E)$ be the graph. Let $n = |V|$ and $m = |E|$, and $w \in \mathbb{R}_+^m$ be the vector of edge weights ($w_e \geq 0$ is the weight of edge e). Consider the matrix $A \in \{0,1\}^{n\times m}$ defined by

$$A_{v,e} = \begin{cases} 1 & \text{if edge } e \text{ is incident to vertex } v \\ 0 & \text{otherwise} \end{cases}$$

then the maximum weighted matching LP is given by

$$\begin{aligned} \text{maximize}_{x\in\mathbb{R}_+^m} \quad & \sum_{e\in E} w_e x_e \\ \text{subject to} \quad & Ax \leq 1 \end{aligned}$$

the inequality constraint reads: for all vertex v, $\sum_{e\in E: v\in e} x_e \leq 1$, i.e. at most one edge should be selected.

If the graph is not bipartite, the LP may not have an integer solution. Indeed, consider the graph $V = \{1, 2, 3\}$, such that the edges $\{(1, 2), (2, 3), (3, 1)\}$ have weight 1, and all other edges have weight 0. The LP is

$$\begin{aligned} \text{maximize}_{x\in\mathbb{R}_+^3} \quad & x_1 + x_2 + x_3 \\ \text{subject to} \quad & x_1 + x_2 \leq 1 \\ & x_2 + x_3 \leq 1 \\ & x_3 + x_1 \leq 1 \end{aligned}$$

the value of the LP is 3/2. This cannot be achieved by any integer solution (by simple enumeration of the feasible integer points).

(1.b) Edmonds showed that if one adds the condition where the total value of x_e for edges across every odd cut $(S, E \setminus S)$ is at least 1, then there is an integer optimum. Add a set of constraints that correspond to this condition (exponential number of constraints)

answer First, to ensure that there can be a perfect matching, if $n = |V|$ is odd, we add a virtual vertex (with zero weight edges). Therefore we assume without loss of generality that n is even (this is important, otherwise the condition that the cut is odd is vacuous, since all cuts will be odd). The constraints are: for each cut $(S, V \setminus S)$, we add a constraint $\sum_{e\in C(S,V\setminus S)} \geq 1$, for all cuts $(S, \bar{S})$ with $|S|$ odd. Here $C(S, \bar{S})$ are the edges across the cut, i.e. $C(S, \bar{S}) = S \times \bar{S} = S \times (V \setminus S)$. The additional inequality constraint is

$$\sum_{e\in C(S,\bar{S})} x_e \geq 1 \ \forall S \subset V \text{ such that } |S| \text{ is odd}$$

if we enumerate the odd cuts from S_1 to S_K (K is exponential in the size m), then let $B \in \{0,1\}^{k\times m}$ defined by

$$B_{k,e} = \begin{cases} 1 & \text{if } e \in C(S_k, \bar{S}_k) \\ 0 & \text{otherwise} \end{cases}$$

then the additional constraints can be rewritten $-Bx \leq -1$, and the LP becomes

$$\begin{array}{ll} \text{maximize}_{x \in \mathbb{R}^m_+} & \sum_{e \in E} w_e x_e \\ \text{subject to} & Ax \leq 1 \\ & -Bx \leq -1 \end{array}$$

(1.c) Take the dual of the linear program, and give an interpretation of it. Use the dual to show that the example in 1.a has no integer solution.

answer Forming the Lagrangian, with variable $\lambda \in \mathbb{R}^n_+$ associated to the constraint $Ax \leq 1$ and variable $\mu \in \mathbb{R}^K_+$ associated to the constraint $-Bx \leq -1$, we have

$$\begin{aligned} L(x, \lambda, \mu) &= w^T x + \lambda^T (1 - Ax) + \mu^T (Bx - 1) \\ &= (w - A^T \lambda + B^T \mu)^T x + \lambda^T 1 - \mu^T 1 \end{aligned}$$

the dual function is given by minimizing the Lagrangian over $x \geq 0$, and is equal to

$$\max_{x \geq 0} L(x, \lambda, \mu) = \begin{cases} \lambda^T 1 - \mu^T 1 & \text{if } w - A^T \lambda + B^T \mu \leq 0 \\ +\infty & \text{otherwise} \end{cases}$$

finally, the dual LP is given by

$$\begin{array}{ll} \text{minimize} & \sum_v \lambda_v - \sum_k \mu_k \\ \text{subject to} & w - A^T \lambda + B^T \mu \leq 0 \end{array}$$

the inequality constraint can be written: $\forall e = (u, v)$, $w_e - \lambda_u - \lambda_v + \sum_{k : v \in C(S_k, \bar{S}_k)} \mu_k \leq 0$, or

$$\lambda_u + \lambda_v - \sum_{k : v \in C(S_k, \bar{S}_k)} \mu_k \geq w_e$$

one interpretation is: the w_e's are the values of the edges, the λ_v's are prices on the vertices, and μ_k's are discounts (one per cut). The discounts affect all edges in the cut. To buy an edge, the sum of its vertex prices minus the total discounts $(\lambda_u + \lambda_v - \sum_{k : v \in C(S_k, \bar{S}_k)})$ have to be greater than the value of the edge w_e. This defines the constraints. The objective of the optimizer is to buy all edges, while minimizing the total amount spent (sum of prices minus sum of discounts).

The dual of the LP in (1.a) is given by

$$\begin{array}{ll} \text{minimize} & \lambda_1 + \lambda_2 + \lambda_3 \\ \text{subject to} & \lambda_1 + \lambda_2 \geq 1 \\ & \lambda_2 + \lambda_3 \geq 1 \\ & \lambda_3 + \lambda_1 \geq 1 \end{array}$$

Therefore at optimum, at most one λ_i is zero (otherwise a constraint would be violated). Thus for a primal x to be optimal, we must have at least two tight primal constraints (by complementary slackness). But an integer solution cannot have two tight primal constraints (since this would mean having two edges in the matching, which is not possible by inspection). Therefore any integer solution is not optimal.

(2.a) Briefly argue that two convex bodies with empty intersection can be separated with a hyperplane.

answer Let A, B be two convex sets. Assume they are closed. Then there exist $a \in A$ and $b \in B$ such that $|a-b| = \min_{x \in A, y \in B} |x-y|$. Consider the hyperplane $H = \{x | w^T(x-c) = 0\}$ where $w = b - a$ and $c = \frac{1}{2}(a+b)$ (this does define a hyperplane since $a \neq b$, since the sets are disjoint and closed).

Then H (strictly) separates A and B, i.e. $\forall x \in A$, $w^T(x-c) < 0$, and $\forall x \in B$, $w^T(x-c) > 0$. Indeed, assume by contradiction that this is not the case. Without loss of generality, assume that the condition is violated for A (the problem is symmetric in a, b). Then

$$\exists x_1 \in A \text{ such that } w^T(x_1 - c) \geq 0$$

Since A is convex and contains both a and x, it contains any convex combination of a and x. Let $\lambda \in [0,1]$, and consider the point $x_\lambda = a + \lambda(x_1 - a)$. The squared distance of x_λ to b is

$$\begin{aligned} d(\lambda) &= |a + \lambda(x_1 - a) - b|^2 \\ &= |a-b|^2 + \lambda^2|x_1 - a|^2 + 2\lambda(a-b)^T(x_1 - a) \end{aligned}$$

we have

$$\begin{aligned} d'(\lambda) &= 2\lambda|x_1 - a|^2 + 2(a-b)^T(x_1 - a) \\ &= 2\lambda|x_1 - a|^2 - 2w^T(x_1 - c + c - a) \\ &\leq 2\lambda|x_1 - a|^2 - 2w^T(c - a) \end{aligned}$$

since $w^T(x_1 - c) \geq 0$ by assumption. Now $2w^T(c-a) = 2(b-a)^T(\frac{a+b}{2} - a) = (b-a)^T(b-a) = |b-a|^2$, thus

$$d'(\lambda) \leq 2\lambda|x_1 - a|^2 - 2|b-a|^2$$

$d'(\lambda) < 0$ for all $\lambda \in [0, \lambda_0]$ where $\lambda_0 = \frac{|b-a|^2}{|x-a|^2}$. Thus there exists a point x_λ with distance to b strictly smaller than $|a-b|$. This contradicts the definition of a.

(2.b) Prove Farkas version B.

Lemma 1 *Farkas B: Exactly one of the following holds*

1. $\exists x$ *such that* $Ax \leq b$
2. $\exists y \geq 0$ *such that* $y^T A = 0$ *and* $y^T b < 0$

proof we seek to show that $1 \Rightarrow \neg 2$ and $\neg 1 \Rightarrow 2$.

- $1 \Rightarrow \neg 2$: if there exists x such that $Ax \leq b$, then for all $y \geq 0$, $y^T Ax \leq y^T b$. Therefore 2 cannot hold, for if $y^T A = 0$, then $y^T Ax \leq y^T b$ becomes $y^T b \leq 0$.
- $\neg 1 \Rightarrow 2$: assume that for all $x \geq 0$, $Ax > b$. In other words, the convex sets $C = \{Ax, x \in \mathbb{R}^n\}$ and $D = \{b + s, s \leq 0\}$ are disjoint, indeed

  $$\begin{aligned} C \cap D \neq \emptyset &\Leftrightarrow \exists x \in \mathbb{R}^n \text{ and } s \leq 0 \text{ such that } Ax = b + s \\ &\Leftrightarrow \exists x \in \mathbb{R}^n \text{ such that } Ax \leq b \end{aligned}$$

  by the convex separation theorem, the sets C and D can be (strictly) separated by a hyperplane, i.e. there exists a vector y and a scalar α such that

  $$\begin{aligned} y^T p &\geq \alpha & \forall p \in C \\ y^T q &\geq \alpha & \forall q \in D \end{aligned}$$

in other words,

$$\begin{aligned} y^T A x &> \alpha & \qquad \forall x \in \mathbb{R}^n \\ y^T(b+s) &< \alpha & \qquad \forall s \leq 0 \end{aligned}$$

In particular for $x = 0$, we have $0 > \alpha$, then in particular for $s = 0$, we have $y^T b < \alpha < 0$. This proves the first inequality for 2. To prove the equality, for all i, take x to be $x = te_i$, where $t \in \mathbb{R}$ and e_i is the i-th canonical vector. Then we have $\forall t$, $y^T A(te_i) < \alpha$, i.e.

$$\forall t > 0,\ y^T A_i < \alpha/t$$
$$\forall t < 0,\ y^T A_i > \alpha/t$$

where A_i is the i-th column of A. Taking the limit when $t \to \pm\infty$ gives $y^T A_i \geq 0$ and $y^T A_i \leq 0$, which proves $y^T A_i = 0$ for all i.

Finally, to show that $y \geq 0$, take $s = te_i$, where $t < 0$. Then $y^T(b + te_i) < \alpha$, i.e.

$$\forall t < 0,\ ty_i < \alpha - y^T b$$

or

$$\forall t < 0,\ y_i > \frac{1}{t}(\alpha - y^T b)$$

taking the limit as $t \to -\infty$, we have $y_i \geq 0$.

(3.a) Notice that the algorithm for rounding the LP relaxation solution had expected facility cost equal to expected LP facility cost, but had a possibly larger connection cost (2 times the LP connection opt plus the LP connection cost).

Let F and C be the optimal facility and connection cost respectively. Given a solution with connection cost C', show that there is a facility i with facility cost f_i whose opening reduces the connection cost by Δ_i, where $\frac{\Delta_i}{f_i} \geq \frac{C'-C}{F}$

answer Observe that opening a facility fractionally with fraction y_i reduces the connection cost by $y_i\Delta_i$. Now if we open all facilities fractionally with fractions y_i, then the resulting connection cost is $C' - \sum_i y_i\Delta_i$, and it is a lower bound on the OPT connection cost C (since we have opened all facilities with fraction y_i). Thus

$$C' - \sum_i y_i \Delta_i \leq C$$

Now assume by contradiction, that for all i, $\frac{\Delta_i}{f_i} < \frac{C'-C}{F}$. Summing over i, we have

$$\sum_i y_i \Delta_i < \frac{C'-C}{F} \sum_i y_i f_i$$

but $\sum_i y_i f_i = F$, thus

$$\sum_i y_i \Delta_i < C' - C$$

which is a contradiction. qed.

(3.b) Argue that you can find a facility location solution with cost at most $F + C + (1 + \ln(\frac{2F+2C}{F}))F$

answer Let I_0 be the set of facilities that are open by the rounding algorithm. Cal the connection cost and facility cost C_0 and F_0 respectively. We have

$$\begin{aligned} C_0 &\leq 2F + 3C \\ F_0 &\leq F \end{aligned}$$

Then at each iteration k, find a facility i_k such that opening i_k reduces connection cost by $\Delta_{i_k} = C_{k-1} - C_k$ that satisfies $\frac{\Delta_{i_k}}{f_{i_k}} > \frac{C_{k-1}-C}{F}$. Stop when $\frac{C_k - C}{F} \leq 1$. Let K be the last iteration.

Then we have at each iteration

$$\begin{aligned} f_{i_k} &\leq F \frac{C_{k-1} - C_k}{C_{k-1} - C} \\ &= F\left(1 - \frac{C_k - C}{C_{k-1} - C}\right) \\ &\leq -F \ln\left(\frac{C_k - C}{C_{k-1} - C}\right) && \text{using } \ln x \leq x - 1 \end{aligned}$$

summing over iterations, we have

$$\begin{aligned} \sum_{k=1}^{K} f_{i_k} &\leq -F \ln \prod_{k=1}^{K} \frac{C_k - C}{C_{k-1} - C} \\ &= F \ln \frac{C_0 - C}{C_K - C} \end{aligned}$$

therefore adding $F_0 + C_K$ to both sides, we obtain a bound on the final cost

$$C_K + F_K \leq F \ln \frac{C_0 - C}{C_K - C} + F_0 + C_K$$

using the bounds $C_0 \leq 2F + 3C$ and $F_0 \leq F$, we have

$$C_K + F_K \leq F \left(1 + \ln \frac{2F + 2C}{C_K - C}\right) + C_K$$

our stopping criterion, we have $C_K \leq C + F$. Observing that the bound is increasing in C_k (indeed, if $f(x) = F + F \ln \frac{2F+2C}{x-C} + x$, then $f'(x) = -\frac{F}{x-C} + 1 > 0$ for $x < C + F$), the bound is also valid replacing C_K by $C + F$. This gives the desired

$$C_K + F_K \leq F \left(1 + \ln \frac{2F + 2C}{F}\right) + C + F$$

(3.c) Consider an instance of the facility location problem where the distance between clients and facilities are either 1 or 3. Give a randomized rounding method that produces a solution with expected connection cost of $(1 + 2/e)C$ and expected facility cost of F (where F and C are the facility and connection cost in a linear program relaxation)

answer Consider the (relaxed) LP

$$\begin{aligned}
&\text{minimize}_{x \geq 0, y \geq 0} && \sum_{i,j} x_{i,j} d_{i,j} + \sum_i y_i f_i && \\
&\text{subject to} && x_{i,j} \leq y_i && \forall i, j \\
& && \sum_i x_{i,j} \geq 1 && \forall j
\end{aligned}$$

where i is in the set of facilities, and j is in the set of clients. Let (x, y) be a solution to the LP, and let F and C be the corresponding facility and connection cost, respectively.

$$F = \sum_i y_i f_i$$
$$C = \sum_{i,j} x_{i,j} d_{i,j}$$

Consider the following randomized algorithm:

1. for all i, open facility i with probability y_i (note that the solution satisfies $y_i \in [0, 1]$).
2. for each client j, assign j to the closest open facility.

Analysis:

- Expected facility cost: let $Y_i = 1$ if facility i is open, and 0 otherwise. Then the opening cost is $\sum_i Y_i f_i$, and taking the expectation

$$\mathrm{E}[\sum_i Y_i f_i] = \sum_i \mathrm{E}[Y_i] f_i = \sum_i y_i f_i = F$$

- Expected connection cost: consider a client j, and let $\bar{C}_j$ be its connection cost. Let $C_j = \sum_i x_{i,j} d_{i,j}$ be the part of the LP connection cost associated to j. Let $G_j = \{i : d_{i,j} = 1\}$ be the set of good facilities for j. Then we have

$$\bar{C}_j = \begin{cases} 1 & \text{if } Y_i = 1 \text{ for some } i \in G_j \\ 3 & \text{if } Y_i = 0 \text{ for all } i \in G_j \end{cases}$$

therefore

$$\begin{aligned}
\mathrm{E}[\bar{C}_j] &= 3 \prod_{i \in G_j} (1 - y_i) + 1. \left(1 - \prod_{i \in G_j} (1 - y_i)\right) \\
&= 1 + 2 \prod_{i \in G_j} (1 - y_i) \\
&\leq 1 + 2 \prod_{i \in G_j} (1 - x_{i,j}) && \text{using the fact that } y_i \geq x_{i,j} \\
&\leq 1 + 2 \prod_{i \in G_j} e^{-x_{i,j}} && \text{using } 1 - x \leq e^{-x} \\
&= 1 + 2e^{-\sum_{i \in G_j} x_{i,j}}
\end{aligned}$$

on the other hand, we have

$$\begin{aligned}
C_j &= \sum_i x_{i,j} d_{i,j} \\
&= 1. \sum_{i \in G_j} x_{i,j} + 3 \left(1 - \sum_{i \in G_j} x_{i,j}\right) \\
&= 3 - 2 \sum_{i \in G_j} x_{i,j}
\end{aligned}$$

to show that $\mathrm{E}[\bar{C}_j] \leq (1 + 2/e) C_j$, it suffices to show that $1 + 2e^{-u} \leq (3 - 2u)(1 + 2/e)$ for $u \in [0, 1]$, and then apply the inequality with $u = \sum_{i \in G_j} x_{i,j}$.

The desired inequality follows from convexity of e^{-x} between the points 0 and 1

$$e^{-u} \leq (1 - u)e^{-0} + ue^{-1}$$

i.e. $e^{-u} \leq 1 - u + u/e$, thus $1 + 2e^{-u} \leq 3 - 2u + 2u/e$, which is clearly $\leq (3 - 2u)(1 + 2/e)$.

This proves that $\mathrm{E}[\bar{C}_j] \leq (1 + 2/e) C_j$, and summing over all the clients j we have $\mathrm{E}[\bar{C}] \leq (1 + 2/e) C$

Therefore the total expected cost is $\leq F + (1 + 2/e) C$.

(4) One of the basic problems in relational databases is computing the size of the join of two relations. Recall that for two relations (tables in a database) $r(A, B)$ and $s(A, C)$, with a common attribute A, we define the join $r \times s$ to be a relation consisting of all tuples (a, b, c) such that $(a, b) \in r$ and $(a, c) \in s$. Therefore if $f_r(a)$ is the number of occurrences of a in r, then the size of the join is $\sum_{a \in A} f_r(a) f_s(a)$. Give an efficient streaming algorithm for estimating the size of the join: scan the items of the two relations in a single pass in streaming fashion to estimate the size of the join (the error can be in terms of the variances $\sum_a f_r(a)^2$, $\sum_a f_s(a)^2$, and the size of the join itself $\sum_a f_r(a) f_s(a)$)

answer Call the r stream (x_i^r, y_i^r), $i \in [m]$, and the s stream (x_j^s, y_j^s), $j \in [m]$, and assume that the sizes of the sets A, B, C are $O(n)$. The number of occurrences of a in the two streams are

$$\begin{aligned} f_r(a) &= |\{i | x_i^r = a\}| \\ f_s(a) &= |\{j | x_j^s = a\}| \end{aligned}$$

Let $h : [n] \to \{-1, +1\}$ be a random hash function drawn from a 4-wise independent family $\mathcal{H}$. Consider the estimator

$$Z = \left(\sum_i h(x_i^r) \right) \left(\sum_j h(x_j^s) \right)$$

grouping the terms that appear in each of the sums, we have

$$\begin{aligned} Z &= \left(\sum_{a \in A} f_r(a) h(a) \right) \left(\sum_{a \in A} f_s(a) h(a) \right) \\ &= \sum_{a \in A} f_r(a) f_s(a) h^2(a) + 2 \sum_{a < a'} f_r(a) f_s(a') h(a) h(a') \end{aligned}$$

- Expectation: taking the expectation over the hash functions (the streams are deterministic and fixed), we have

  $$\mathop{\mathrm{E}}_{h \in \mathcal{H}}[Z] = \sum_a f_r(a) f_s(a) + 2 \sum_{a < a'} f_r(a) f_s(a') \, \mathrm{E}[h(a) h(a')]$$

  by linearity, and the fact that for all a, $h^2(a) = 1$. Then for each fixed pair $a \neq a'$, $h(a)$ and $h(a')$ are independent, therefore $\mathrm{E}[h(a)h(a')] = \mathrm{E}[h(a)] \, \mathrm{E}[h(a')] = 0$. This proves

  $$\mathop{\mathrm{E}}_{h \in \mathcal{H}}[Z] = \sum_a f_r(a) f_s(a) = \mu$$

- Variance: we seek to bound the variance in terms of the size of the join μ, and the variances (or self joins) v_s and v_r defined by

  $$\begin{aligned} \mu &= \sum_{a \in A} f_r(a) f_s(a) \\ v_r &= \sum_{a \in A} f_r(a)^2 \\ v_s &= \sum_{a \in A} f_s(a)^2 \end{aligned}$$

  We have

  $$\begin{aligned} \mathrm{E}[Z^2] &= \mathrm{E}\left[\sum_{a_1, a_2, a_3, a_4} f_r(a_1) f_r(a_2) f_s(a_3) f_s(a_4) h(a_1) h(a_2) h(a_3) h(a_4) \right] \\ &= \sum_{a_1, a_2, a_3, a_4} f_r(a_1) f_r(a_2) f_s(a_3) f_s(a_4) \, \mathrm{E}[h(a_1) h(a_2) h(a_3) h(a_4)] \end{aligned}$$

by linearity of the expectation. Then using the fact that if $a \neq a'$, $\mathrm{E}[h(a)h(a')] = 0$, we observe that the terms $\mathrm{E}[h(a_1)h(a_2)h(a_3)h(a_4)]$ are non-zero only if each a_i appears an even number of times, in which case $\mathrm{E}[h(a_1)h(a_2)h(a_3)h(a_4)] = 1$. Therefore the sum can be reduced to

$$\begin{aligned}
\mathrm{E}[Z^2] &= \sum_{a_1=a_2=a_3=a_4} f_r(a_1)^2 f_s(a_1)^2 + \sum_{a_1=a_2, a_3=a_4} f_r(a_1)^2 f_s(a_3)^2 + \sum_{\substack{a_1=a_3, a_2=a_4 \\ or \\ a_1=a_4, a_2=a_3}} f_r(a_1) f_s(a_1) f_r(a_2) f_s(a_2) \\
&= \sum_a f_r(a)^2 f_s(a)^2 + 2 \sum_{a<a'} f_r(a)^2 f_s(a')^2 + 4 \sum_{a<a'} f_r(a) f_s(a) f_r(a') f_s(a')
\end{aligned}$$

on the other hand

$$\begin{aligned}
\mathrm{E}[Z]^2 &= (\sum_a f_r(a) f_s(a))^2 \\
&= \sum_{a,a'} f_r(a) f_s(a) f_r(a') f_s(a') \\
&= \sum_a f_r(a)^2 f_s(a)^2 + 2 \sum_{a<a'} f_r(a) f_s(a) f_r(a') f_s(a')
\end{aligned}$$

therefore the variance is

$$\begin{aligned}
\mathrm{E}[Z^2] - \mathrm{E}[Z]^2 &= 2 \sum_{a<a'} f_r(a)^2 f_s(a')^2 + 2 \sum_{a<a'} f_r(a) f_s(a) f_r(a') f_s(a') \\
&\leq \left(\sum_a f_r(a)^2 \right) \left(\sum_{a'} f_s(a')^2 \right) + \left(\sum_a f_r(a) f_s(a) \right)^2 \\
&= v_r v_s + \mu^2
\end{aligned}$$

In order to decrease the variance (to obtain a better bound on the probability that $|Z - \mu|$ is big), we run the core algorithm K times in parallel, by drawing a random hash function from $\mathcal{H}$ for each run. We obtain estimators $Z_1, \ldots, Z_K$, which we combine by taking the average

$$Y = \frac{1}{K} \sum_{k=1}^{K} Z_k$$

then we still have the correct expectation by linearity

$$\mathrm{E}[Y] = \mathrm{E}[Z] = \mu$$

and by independence of the terms, the variance is divided by K

$$\mathrm{var}[Y] = \frac{1}{K^2} \sum_k \mathrm{var}[Z_k] = \frac{1}{K} \mathrm{var}[Z] = \frac{1}{K} (v_r v_s + \mu^2)$$

Finally, by the Chebyshev inequality, we have

$$\begin{aligned}
\Pr[|Y - \mu| > \epsilon(\sqrt{v_r v_s} + \mu)] &\leq \frac{1}{\epsilon^2 (\sqrt{v_r v_s} + \mu)^2} \mathrm{var}[Y] \\
&= \frac{1}{K\epsilon^2} \frac{v_r v_s + \mu^2}{(\sqrt{v_r v_s} + \mu)^2} \\
&\leq \frac{1}{K\epsilon^2}
\end{aligned}$$

therefore for a fixed small δ and small ϵ, if we run the algorithm $K = \frac{1}{\delta\epsilon^2}$ times, we have

$$\Pr[|Y - \mu| > \epsilon(\sqrt{v_r v_s} + \mu)] \leq \delta$$

the space complexity is then $O(\frac{1}{k\epsilon^2} \log n \log m)$, since for each run, we need to specify the hash function $h : [n] \to \{-1, +1\}$, which requires $O(\log n)$ space (pick random vectors of size $O(n)$), and each counter requires $O(\log m)$ space.

(5) Project: I would like to work on the experts algorithm, applied to routing. I am interested in particular in what happens when the cost function for routing is not linear. The problem we discussed in class was

$$\min_f \max_p c(p, f)$$

where we minimize over path flows f (vector of size the number of routes), the maximum congestion over edge distributions p (vector of size the number of edges). Here the congestion function $c(p, f)$ is linear and can be written as $p^T A f$ for an appropriate matrix A ($A_{e,r} = 1$ if edge e is on route r), making this a zero sum 2 player game. We applied the experts algorithm to this LP (in fact to its dual) to find approximate solutions. When $c(p, f)$ is not linear the analysis might not work so nicely. I will look at the convex case first.

Another version of the problem is

$$\min_f \sum_e f(e) c(e, f)$$

$f(e)$ is the flow on edge e. We want to minimize the average cost experienced by everyone. This is easy to solve if we control all the flow. But if some portion of the flow is selfish (i.e. chooses best available route) the problem (Stackelberg routing game) becomes hard (Roughgarden has some results on this). I want to see whether we could apply experts to find an approximate solution.

I have some other ideas involving experts. In particular, I believe there might be a connection with the EM algorithm for learning graphical models (both problems can be written as a solution to an optimization problem, where we optimize over distributions, regularized by the entropy of the distribution). I would like to see if there is an interpretation of the EM algorithm in terms of experts. A quick search for "expectation maximization experts" does not seem to return anything interesting.