COMS E6232 - Problem Set #3

Alex Wong (asw2181) - asw2181@columbia.edu

March 24, 2017

Problem 1

a. **Lemma 1a** There is always a directed cut that contains at least 1/4 of the edges.

Proof: For $i = 1, ..., |N|$ we assign node i to set S with probability 1/2 and to set $N - S$ with probability 1/2. For each directed edge (i, j), let $X_{(i,j)} = 1$ if $(i, j) \in$ directed cut from S to N-S, else 0 otherwise. Thus, we get the equation:

$$E[X_{(i,j)}] = P[(i, j) \in \text{directed cut from S to N-S}] = P[i \in S, j \in N - S] = 1/4$$

□

b.

c. The worst-case approximation ratio of the simple local search algorithm is unbounded. Take for example a node i and some number n of nodes where all edges are directed towards i. If we put node i in set S and the rest of the nodes in set $N - S$, then all the edges are directed from $N - S$ to S, giving a cut with 0 edges since only edges from S to $N - S$ count. Using only one local move will not improve the cut, regardless of which node we decide to move. If we move node i to $N - S$, then all nodes will be on that side. If we move any of the n nodes from $N - S$ to S, then the edge will reside within S, still not improving the cut. We can see that the maximum cut would have the n number of nodes in set S and node i in set $N - S$, which would give us a cut with n number of edges. Thus, we can see that the worst-case approximation ratio is unbounded using the simple local search algorithm since we can scale n to any arbitrarily large number.

d. We can improve the local search so that it guarantees ratio 1/4. The improvement of the local search algorithm is summarized below:

Given some directed graph $D = (N, E)$ and a random assignment of the nodes $i = 1, ..., |N|$ into either set S or $N - S$, while $\exists$ a node i that has more neighbors on the same side than the opposite side, move i to the opposite side. Based on slides 23 and 24 of the Lecture 5 slides, we will end up with $\geq 1/2|E|$ of edges that go between the sets S and $N - S$. From this, we find the value of the directed cut of these edges. Then, we move all the current nodes that are in S to $N - S$ and the current nodes of $N - S$ to S and find the value of the directed cut of these edges. We take the configuration of S and $N - S$ based on the larger of the two directed cuts that we found, which at least one of them $\geq 1/4|E|$. One of the cuts will be $\geq 1/4|E|$ because there are a total of $\geq 1/2|E|$ edges in the cut, so if $\leq 1/4|E|$ edges go from S to $N - S$ in one configuration, then there must be $\geq 1/4|E|$ edges going from $N - S$ to S, so we would move the nodes to orient the direction of the edges to go from S to $N - S$.

Problem 2

a. The quadratic program for this problem is given below:

$\max\frac{1}{2}\sum_{i<j} 1$

Problem 3

a. **Lemma 3a** If I is an independent set of G then I^k is an independent set of $G^{(k)}$.

Proof: For I to be an independent set of G, there must be no adjacent edges between any of the nodes of I. Given the definition of $E^{(k)}$ of $G^{(k)}$,

$$E^{(k)} = \{(u,v) \in N^{(k)} \times N^{(k)} \mid \exists i,j \in [k], (u_i, u_j) \in E \text{ or } (v_i, v_j) \in E \text{ or } (u_i, v_j) \in E\}$$

we can see that there are no edges between any pairs (u, v) of k-tuples of nodes that only contain nodes of I. When taking the Cartesian product of I, I^k, we create k-tuples using only those nodes from I so $I^k \subseteq N^{(k)}$ since $I \subseteq N$. Also, as explained earlier, the k-tuples of I^k do not have adjacent edges between them since they only contain nodes of I, thus showing that I^k is an independent set of $G^{(k)}$. $\square$

b. **Lemma 3b** If J is an independent set of $G^{(k)}$ then we can construct in polynomial time an independent set I of G of size at least $|J|^{1/k}$ and conclude that $\alpha(G^{(k)}) = (\alpha(G))^k$

Proof: From Lemma 3a, if I is an independent set of G, then I^k is an independent set of $G^{(k)}$. Since we have defined J to be an independent set of $G^{(k)}$, we can use Lemma 3a to say that $J = I^k$. The size of J is then equal to the size of I^k, which is $|I|^k$, thus $|J| = |I|^k \implies |I| = |J|^{1/k}$, showing that we can construct in polynomial time an independent set I of G of size at least $|J|^{1/k}$. Also, if given the maximum independent set of G, $\alpha(G)$, we can use our earlier conclusion and see that $(\alpha(G^{(k)}))^{1/k} = \alpha(G) \implies \alpha(G^{(k)}) = (\alpha(G))^k$. $\square$

c. **Lemma 3c** If the Maximum Independent Set problem can be approximated in polynomial time within some constant factor $c > 1$, then it has a PTAS.

Proof: We first define $\alpha(G)$ as the maximum independent set of G. From the PCP theorem, we can trivially say that the Maximum Independent Set problem has a 2-approximation algorithm. Let us define I as an independent set of G that agrees with the 2-approximation algorithm. This means that the size of I will be 1/2 the size of the maximum independent set of G, thus $\frac{|\alpha(G)|}{|I|} = 2$. Now, using Lemma 3b, if we take the k-th power of the graph G, we know that $\alpha(G^{(k)}) = (\alpha(G))^k$ and by Lemma 3a, I^k is an independent set of $G^{(k)}$. Thus, for some k-th power graph $G^{(k)}$, the approximation ratio is $\frac{|\alpha(G)|^k}{|I|^k} = (\frac{|\alpha(G)|}{|I|})^k = 2^k$ which shows that the approximation ratio grows as we apply the approximation algorithm to larger and larger k-th powers of graph G. Thus, the Maximum Independent Set problem can not be approximated in polynomial time within some constant factor, which means it does not have a PTAS unless $\mathbf{P} = \mathbf{NP}$. $\square$

Problem 4

a. **Lemma 4a** MDAS can be trivially approximated within a factor of 2.

Proof: Suppose we have an arbitrary ordering of nodes $v_1, ..., v_n$ and we have two subsets of edges A_1 and A_2 where

$$A_1 = \{(v_i, v_j) \mid (v_i, v_j) \in A, i < j\}$$

$$A_2 = \{(v_i, v_j) \mid (v_i, v_j) \in A, i > j\}$$

By separating the edges into these two subsets, the only way for A_1 to contain a cycle is if there is an edge where $i > j$, which would be contained in the A_2 subset, and the only way for A_2 to contain a cycle is if there is an edge where $i < j$, which would be contained in the A_1 subset. Thus, neither subset will contain a cycle and at least one of the two subsets will contain at least $|A|/2$ edges since $A = A_1 + A_2$. Thus, MDAS can be trivially approximated within a factor of 2 by taking the larger of the two subsets. $\square$

b. **Theorem 4.2** The Maximum Directed Acyclic Subgraph (MDAS) problem does not have a PTAS unless $\mathbf{P} = \mathbf{NP}$.

Proof: A known problem that does not have a PTAS is the Maximum Independent Set (MIS) problem. MIS does not have a PTAS for any graph with a maximum degree ≥ 3. We can do a linear reduction $MIS(3) \leq_L MDAS$: Given an undirected graph $G = (N, E)$ with maximum degree 3, we construct a directed graph $D = (V, A)$ where $V = \{(u_1, u_2) \mid u \in N\}$ and $A = \{(u_1, u_2) \mid u \in N\} \cup \{(u_2, v_1), (v_2, u_1) \mid (u, v) \in E\}$. We let $\alpha(G)$ denote the size of the maximum independent set of G and $\gamma(D)$ denote the number of edges of the maximum acyclic subgraph of D. From this, we have the following lemma:

Lemma 4.2.1 For all $\epsilon > 0$, if we are given an acyclic subgraph of D that has at least $(1 - (\epsilon/13))\gamma(D)$ edges, then we can compute in polynomial time an independent set G that has at least $(1 - \epsilon)\alpha(G)$ nodes.

Proof:
(4.2a) If I is an independent set of G, then $D' = (V, A')$ where $A' = \{(u_1, u_2) \mid u \in I\} \cup \{(u_2, v_1), (v_2, u_1) \mid (u, v) \in E\}$ is an acyclic subgraph of D. We can see that all edges $\{(u_2, v_1), (v_2, u_1) \mid (u, v) \in E\}$ do not create a cycle because as we have shown in Lemma 4a, a set of edges $\{(v_i, v_j) \mid (v_i, v_j) \in A, i > j\}$ do not create a cycle. We also know that the independent set I contains nodes that do not have any adjacent edges with each other. Thus, since A' contains all the linear reduction of edges of E, the only way to keep the edge set acyclic would be to only add the edges $\{(u_1, u_2) \mid u \in N\}$ where the nodes do not have any adjacent edges, which would be the independent set

I and gives us $\{(u_1, u_2) \mid u \in I\}$. Thus, if I is an independent set of G, then A' is an acyclic subgraph of D.

(4.2b) Now we show that if H is an acyclic subgraph of D with h edges, we can then derive efficiently from H an independent set of G with at least $h - 2|E|$ nodes. As we have shown in **4.2a**, acyclic subgraphs of D includes all edges $\{(u_2, v_1), (v_2, u_1) \mid (u, v) \in E\}$. We can see that each edge of E corresponds to two edges in A. If we took away all those edges from H, we would be left with edges $\{(u_1, u_2) \mid u \in N\}$ which we know are nodes that could create an independent set, as we have shown in **4.2a**. Thus, we can derive efficiently from H an independent set of G with at least $h - 2|E|$ nodes.

(4.2c) Lastly, since G has maximum degree 3, we can show that $\alpha(G) \geq |E|/6$. Without loss of generality, lets take a node $u \in N$ that has degree 3, which would imply 4 connected nodes. If we were to make all these nodes connected together, we would have a complete graph where each of the 4 nodes would have a maximum degree of 3. This complete graph has a 6 total edges. From this complete graph, we can only choose any one node as the maximum independent set $\alpha(G)$ as all nodes are adjacent to all other nodes. Thus, we can see that $\alpha(G) \geq |E|/6$.

Let us do the linear reduction from MIS to MDAS. From **4.2b** we know we can get an acyclic subgraph of D with $|I| + 2|E|$ edges. If $I = \alpha(G)$, then getting an acyclic subgraph D with $\alpha(G) + 2|E|$ would imply a maximum acyclic subgraph $\gamma(D)$. We can manipulate our conclusion from **4.2c** where $\alpha(G) \geq |E|/6 \implies |E| \leq 6 \cdot \alpha(G)$. Thus, we can say

$$\gamma(D) = \alpha(G) + 2|E| \leq \alpha(G) + 12 \cdot \alpha(G) = 13 \cdot \alpha(G)$$

Based on this L-reduction, it satisfies the first property that $OPT_{MDAS} \leq \alpha \cdot OPT_{MIS}$ where $\alpha = 13$. We can also see it satisfies the second property $|C_1 - OPT_{MIS}| = \beta \cdot |C_2 - OPT_{MDAS}|$ where $\beta = 1$. Combining both properties, we get the relative error:

$$\frac{|C_1 - OPT_{MIS}|}{OPT_{MIS}} \leq \frac{\alpha \cdot \beta \cdot |C_2 - OPT_{MDAS}|}{OPT_{MDAS}} \leq \alpha\beta\epsilon$$

Showing that the relative error of MIS is $13 \cdot \epsilon$ in relation to the relative error of MDAS being ϵ. Thus, if we wanted to express the relative error of MIS as just ϵ, that would mean that the relative error of MDAS would be $\epsilon/13$. Thus for all $\epsilon > 0$, if we are given an acyclic subgraph of D that has at least $(1 - (\epsilon/13))\gamma(D)$ edges, then we can compute in polynomial time an independent set G that has at least $(1 - \epsilon)\alpha(G)$ nodes, which proves the lemma. □

But because MIS does not have a PTAS as we have proven in problem 3, and since $MIS \leq_L MDAS$, that tells us that MDAS also does not have a PTAS unless **P** = **NP**, thus the theorem is proved. □