Approximation Algorithms Workshop June 13-17, 2011, Princeton

Open Problems Session

Scribe: Rajesh Chitnis

Please send any corrections/comments to rajeshchitnis@gmail.com

1 Almost Cliques : Does more information really help ?

Proposed by Sanjeev Arora

If we know that $\omega(G) = k$ then can we find an almost $k-$clique, i.e., can we find in polynomial time a subgraph on k vertices having $(1-\epsilon)$ fraction of the edges of a $k-$clique.

Suppose we have the following stronger assumption : Our graph G consist of 1000 disjoint cliques of size $\frac{n}{1000}$ each. Can we then find in polynomial time a subgraph on $\frac{n}{1000}$ vertices having $(1-\epsilon)$ fraction of edges of an $(\frac{n}{1000})-$clique ?

2 Directed k-Spanner Problem

Proposed by Grigory Yaroslavtsev, for more details see slides here and a paper here

Problem statement Let $G(V, E)$ be a weighted *directed* graph. A k-spanner is a subset of edges of G, which preserves distances in the original graph up to a factor k. Formally, a k-spanner is a graph $G_H(V, E_H)$, where $E_H \subseteq E$ and $\forall\ (u, v) \in E$ we have $d_{G_H}(u, v) \leq k \cdot d_G(u, v)$.

We want to find a directed $k-$spanner which *minimizes* the number of edges $|E_H|$. What is the best approximation factor that we can get?

Most recent previous work

- Dinitz and Krauthgamer (STOC 2011) gave a $\tilde{O}(n^{\frac{2}{3}})$ approximation and showed an integrality gap of $\Omega(n^{\frac{1}{3}-\epsilon})$.
- This was improved to $\tilde{O}(\sqrt{n})$ by Berman, Bhattacharya, Makarychev, Raskhodnikova and Yaroslavtsev (ICALP 2011).
- Hardness: Elkin and Peleg (STOC 2000) show that it is quasi-NP-hard to approximate with ratio better than $2^{log^{1-\epsilon} n}$.
- Integrality gap: $\Omega(n^{1/3-\epsilon})$ (for constant k) by Dinitz and Krauthgamer.

Questions :

- Can we beat this ratio?
- Current method is randomized rounding of an LP relaxation, combined with sampling. What other techniques can we use?
- What is a natural online setting for this problem?

Some comments by the listeners

- **Q:** How about undirected spanners?

  A: They are very different, because girth arguments work there.

- **Q:** How about directed spanners of minimum cost?

  A: The best result is by Dodis and Khanna (STOC 1999), who give $\tilde{O}(n)$-approximation.

- **Q:**Is there an example, when there is a sparse directed spanner of a dense graph?

  A: For every k it is easy to construct a graph with $\Omega(n^2)$

- In above problem we had fixed k and wanted to minimize $|E_M|$. We can also consider the problem where we have a bound on $|E_M|$ and then want to minimize k.
- How well can we approximate the size of the sparsest $2k$-spanner, relative to the size of the sparsest k-spanner?

3 Framework of Minimizing Movement

Proposed by MohammadTaghi Hajiaghayi

Demaine et al. (SODA 2007) considered the following problem : Given an (un)directed graph $G = (V, E)$ where $|V| = n$. We have a set of m pebbles and a property P on configurations where a configuration is a function mapping pebbles to vertices in V. We are given an initial configuration of the pebbles and we want to bring out a motion of the pebbles so that the property P is achieved in the graph. Note that the pebbles can move only along edges of the graph.

We have two particular vertices s and t. Pebbles are placed on some of the other vertices in the graph. There are 3 general problems we can consider in this framework :

1. Minimize the number of pebbles moved in achieving property P
2. Minimize the maximum movement of a pebble in achieving property P
3. Minimize the total movement of all pebbles to achieving property P

Demaine et al. (SODA 2007) gave approximation algorithms for various special cases of the property P such as connectivity which says that the graph induced on pebbles must be connected. Among other results, for the special case where P is having a pebble path between 2 fixed vertices s and t they give $O(\sqrt{\frac{m}{OPT}})$-approximation. Friggstad and Salvatipour (FOCS 2008) consider a generalization of the problem where we have client nodes and facility nodes and we want to move these nodes such that finally each client is at a node which is destination of some facility. Via an approximation preserving reduction they show that minimizing the total movement of clients and facilities in this framework generalizes the classical $k-$median problem and give an 8-approximation. One important open problem is when the facilities have additional capacity constraints.

This framework has not been well-studied yet and many variants are still open.

4 Decomposing graphs into two balanced parts such that each edge belongs to at least one of the parts

Proposed by Guy Kortsarz

Remark: This is only the simplest question of a subject called *channel allocation*. See a paper by Rajiv Gandhi, Khuller, Srinivasan and Wang. Approximation algorithms for channel allocation problems in broadcast networks, Networks, 47(4):225-236, March 2006.

We want two balanced subsets $S_1, S_2 \subseteq V(G)$ such that each edge in G belongs to at least one of $G[S_1]$ or $G[S_2]$ and $\max\{|S_1|, |S_2|\}$ is minimized. A naive solution of $S_1 = V$ and $S_2 = \emptyset$ gives a 2 approximation. Can we do better ?

Let A and B be the optimum and let $C = A \cap B$. Consider $A - C$ and $B - C$. There could not be edges between $A - C$ and $B - C$ as these edges do not belong to S_1 nor to S_2. Thus C is a graph separator and should be balanced if we want to get good approximation. The optimum may find a bisection with $C = o(n)$ and get $|S_1|$ and $|S_2|$ of size about $n/2$.

There is an $O(\sqrt{\log n})$ approximation of a variant of vertex separators due to Feige, Hajiaghayi and Lee (STOC 2005).

I can give a series of question of the same flavor : Can we beat the approximation ratio when OPT is large? For example, in VC if we know that there is a vertex cover of size $n/1.5$ then can we find a VC of size $0.9999n$? This question therefore can be translated to a collection of questions.

5 Probabilistic s-t (Dis)Connectivity in Directed Graphs

Proposed by Rajmohan Rajaraman

This problem is a probabilistic version of s-t connectivity. Given a directed graph G with source s and sink t, suppose we want to remove nodes (or edges) from G so as to disconnect s from t. This can be solved easily by the standard max-flow min-cut theorem.

Suppose now that when we select any nodes for removal, they are actually removed only with a probability p, say $1/2$. Then, we can ask the following question: Given a directed graph G with source s and sink t, what is the set S of nodes that maximizes the probability that s is disconnected from t, when each vertex in S is removed independently with probability p?

One motivation for this problem arises from considering epidemics in networks. We may want to protect node t from an epidemic originating at s. We may have a budget to install some "protection" at k nodes (for example, by vaccination), but this protection may only succeed with a certain probability p. Then, what is the set of nodes to target such that t is disconnected from s with maximum probability?

We could also consider the following variant: Find a set S of k nodes (if it exists) whose removal – in the probabilistic sense as above – disconnects s from t with probability at least 1/2.

One challenge in attacking the above problem is that even the following computation may not be tractable: Given a set S, estimate the probability that s is disconnected from t if each node in S is removed with probability 1/2. It is known that computation of such probabilities is #P-hard. One can, however, find good approximations using Monte Carlo simulations. By just sampling a polynomial number of instances of the above and counting the number of times that s is connected to t, we can get an high-probability approximation that is within an additive $1/poly(n)$ value of the true answer. There is relevant work by Marco Laumanns and Rico Zenklusen (European Physics Journal B, 2009) on the complexity of estimating s-t connectivity probability and of estimating the epidemic size. For certain other related problems such as the all-terminal network reliability (e.g., work by Karger STOC 1995), fully polynomial-time randomized approximation schemes are known.

One can also define several variants of the above connectivity problem; e.g., having multiple sources and sinks, and probabilistic variants of standard cut problems.

As far as I know, there are no results known on the above optimization problem.

6 Beating $O(\sqrt{log\ n})$ for Sparsest Cut

Proposed by David Steurer

The best known approximation for Sparsest Cut is $\sqrt{log\ n}$ due to Arora, Rao and Vazirani (STOC 2004). Can we get better approximation via SDP hierarchies (lift and project approach)?

The following Structure theorem is known : If $v_1, \ldots, v_n$ are unit vectors which are *well-spread* and obey squared triangle inequality, then if we can partition $[n]$ into S, T such that $|S|, |T| \geq \frac{n}{3}$ and $\langle v_i, v_j \rangle < 1 - \epsilon$ for every $i \in S, j \in T$ then we have a $\frac{1}{\epsilon}$-approximation for Sparsest Cut. Arora, Rao and Vazirani show this for $\epsilon = \sqrt{\frac{1}{log\ n}}$. We cannot get it for smaller ϵ because the example when the vertices $v_1, v_2, \ldots, v_n$ are vertices of a $(log\ n)-$dimensional hypercube is a tight example. The question now is whether the tight example for the Structure theorem is inherently low-dimensional. if so then this could lead to better approximation via SDP hierarchies.

Let $d(\epsilon)$ be the smallest dimension such that for unit vectors $v_1, v_2, \ldots, v_n$ satisfying $\mathbb{E}_{i,j}\langle v_i, v_j \rangle^2 < \frac{1}{d(\epsilon)}$ there is a partition S, T of $[n]$ such that $\langle v_i, v_j \rangle < 1-\epsilon$ for every $i \in S, j \in T$. Global Correlation due to Barak, Raghavendra and Steurer (2010) shows that we can get $\frac{1}{\epsilon}$-approximation for Sparsest Cut in $2^{d(\epsilon)}\text{poly}(n)$ time. A conjecture is that $d(\epsilon) = n^{\epsilon}$

7 Computing Operator Norm of Matrices

Proposed by Aravindan Vijayaraghavan

For a matrix A the operator norm of A is defined by $||A||_{q \to p} = \max_{x \neq 0} \dfrac{||Ax||_p}{||x||_q}$. For different values of p, q, we get very different problem flavors. For $p = 2 = q$ this is the maximum singular value of the matrix. For $p = 1$ and $q = \infty$ this is the Grothendick problem.

We have reasonable understanding of the problem when $p \leq q$. In the case when $2 < p \leq q$ or $p \leq q < 2$, the problem is very hard to approximate (even upto LabelCover like inapproximability factors) in general . However in this range, when entries of A are non-negative (say, adjacency matrices of graphs) then there is a PTAS. Also, when $p > 2, q < 2$ or $p < 2, q > 2$, then a constant factor approximation (roughly factor 2) holds due to Nesterov.

The problem however is completely open for the range $q < p \leq 2$ and $2 \leq q < p$ (a trivial $O(n^{1/4})$ approximation is known by Holder's inequality). These norms (called hypercontractive norms) seem to have interesting connections with well-studied problems. For instance, Ailon and Liberty showed that matrices with small $||A||_{2 \to 4}$ are useful for fast Johnson-Lindenstrauss dimension reductions. Further, graphs whose adjacency matrices have bounded $||A||_{(2-\epsilon) \to 2}$ (called hypercontractive graphs) are integrality gap instances for SDP relaxations of unique games and related problems. Hence, it would be useful to have an algorithm which certifies that the matrix has small (hypercontractive) norm.

Further, computing the $||A||_{(2-\epsilon) \to 2}$ norm seems to be closely related to algorithms for the densest k-subgraph (DkS) problem. The recent algorithm of Bhaskara, Charikar, Chlamtac, Feige and Vijayaraghavan (STOC 2010) when given a graph G of degree n^δ containing a k-subgraph of higher log-density i.e. induced average degree $k^{\delta+\gamma}$ with $\gamma > 0$ can roughly recover a k-subgraph of degree k^γ. A good approximation to computing $||A||_{(2-\epsilon) \to 2}$, would yield better approximations to DkS, when the optimal k-subgraph has higher log-density than that of G (however, this would not yield an better approximation algorithm overall, since it would yield much when the log-density of the optimal k-subgraph is smaller).