Nuclear Norm minimization for the planted clique and biclique problems

Brendan Ames[1] Stephen Vavasis[1]

[1]Department of Combinatorics & Optimization
University of Waterloo

2009-Aug-26 / ISMP

Outline

Maximum clique and biclique problems

- Clique: Given an undirected graph (V, E), find k vertices mutually interconnected such that k is maximized
- Biclique: Given a bipartite graph (U, V, E), find a subgraph (U^*, V^*, E^*) containing all possible $|U^*| \cdot |V^*|$ edges such that $|U^*| \cdot |V^*|$ is maximized
- Max Clique and Max Biclique are both NP-hard
- They are simple models of information retrieval problems

Biclique reformulation as rank minimization

- *Rank minimization* refers to the problem of minimizing the rank of X subject to convex constraints on X.
- Existence of an mn biclique as rank minimization:

$$
\begin{array}{lll}
\min & \mathrm{rank}(X) & \\
\text{s.t.} & X(i,j) \in [0,1] & \forall (i,j) \in U \times V \\
& X(i,j) = 0 & \forall (i,j) \in (U \times V) - E \\
& \sum_{(i,j)} X(i,j) \geq mn &
\end{array}
$$

- Similar formulation exists for clique

Nuclear norm relaxation

- *Nuclear norm* of a matrix X, written $\|X\|_*$, is defined as the sum of X's singular values.
- Recent work by Recht, Fazel, Parrilo and others suggests that nuclear norm is a good relaxation for rank minimization.
- Nuclear norm relaxation of biclique:

$$(NNR)\quad \begin{array}{lll} \min & \|X\|_* & \\ \text{s.t.} & X(i,j) \geq 0 & \forall (i,j) \in U \times V, \\ & X(i,j) = 0 & \forall (i,j) \in (U \times V) - E, \\ & \sum_{(i,j)} X(i,j) \geq mn. & \end{array}$$

- This relaxation is convex.

Our results for clique

- Consider an N-node graph G consisting of an n-node clique K_n plus diversionary edges:
  - Up to $O(n^2)$ deterministically-placed diversionary edges; at most $O(n)$ K_n-vertices adjacent to any non-K_n-vertex, or,
  - All nonclique edges inserted independently at random with probability p, and $N = O(n^2)$.
- Then the nuclear norm relaxation finds the maximum clique and a certificate that it is the maximum.
- Similar results for biclique.
- Previous results by Feige and Krauthgamer; Alon, Krivelevich and Sudakov.

Relationship to compressive sensing

- Recht, Fazel and Parrilo: main results of compressive sensing extend to rank minimization.

$$\|\mathbf{x}\|_0 \Longrightarrow \mathrm{rank}(X)$$
$$\|\mathbf{x}\|_1 \Longrightarrow \|X\|_*$$

- Our result is similar to theirs in spirit: an NP-hard rank minimization problem is solvable in polynomial time by convex relaxation if low-rank solution is known to exist and if data includes randomized constraints.
- We use the KKT conditions in the same way that RFP use them.

Subgradient of nuclear norm

- Proof technique: show that the maximum clique is optimal for (NNR) by showing that KKT conditions are satisfied. Furthermore show that optimal solution is unique.
- Suppose $A \in \mathbf{R}^{m \times n}$ has rank r and SVD $A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$. Then $\phi \in \partial \|A\|_*$ iff $\phi = \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \mathbf{u}_r \mathbf{v}_r^T + W$ s.t. $\|W\| \leq 1$, $\mathrm{span}(W) \perp \mathrm{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$, $\mathrm{span}(W^T) \perp \mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$.

KKT conditions (biclique case)

Theorem. Suppose X is a feasible rank-one matrix $X = \bar{\mathbf{u}}\bar{\mathbf{v}}^T$, where $\bar{\mathbf{u}}, \bar{\mathbf{v}}$ are the characteristic vectors of $U^* \subset U$, $V^* \subset V$ resp, $|U^*| = m$, $|V^*| = n$. Then X is optimal for (NNR) iff $\exists W \in \mathbf{R}^{M\times N}, \lambda \in \mathbf{R}^{M\times N}, \mu \in \mathbf{R}$ s.t.

$$\frac{\bar{\mathbf{u}}\bar{\mathbf{v}}^T}{\sqrt{mn}} + W = \mu\mathbf{e}\mathbf{e}^T + \sum_{(i,j)\in(U\times V)-E} \lambda_{ij}\mathbf{e}_i\mathbf{e}_j^T$$

with $\|W\| \leq 1$, $W^T\bar{\mathbf{u}} = \mathbf{0}$, $W\bar{\mathbf{v}} = \mathbf{0}$, $\mu \geq 0$. In this case, U^*, V^* is an optimal solution for the max biclique problem.

KKT conditions (biclique case)

Theorem. Suppose X is a feasible rank-one matrix $X = \bar{\mathbf{u}}\bar{\mathbf{v}}^T$, where $\bar{\mathbf{u}}, \bar{\mathbf{v}}$ are the characteristic vectors of $U^* \subset U$, $V^* \subset V$ resp, $|U^*| = m$, $|V^*| = n$. Then X is optimal for (NNR) iff $\exists W \in \mathbf{R}^{M\times N}, \lambda \in \mathbf{R}^{M \times N}, \mu \in \mathbf{R}$ s.t.

$$\frac{\bar{\mathbf{u}}\bar{\mathbf{v}}^T}{\sqrt{mn}} + W = \mu \mathbf{e}\mathbf{e}^T + \sum_{(i,j)\in(U\times V)-E} \lambda_{ij}\mathbf{e}_i\mathbf{e}_j^T$$

with $\|W\| \le 1$, $W^T\bar{\mathbf{u}} = \mathbf{0}$, $W\bar{\mathbf{v}} = \mathbf{0}$, $\mu \ge 0$. In this case, U^*, V^* is an optimal solution for the max biclique problem. If, in addition, $\mu > 0$ and $\|W\| < 1$, X is the unique optimizer.

Finding W, λ, μ

Thus, showing that (NNR) finds the optimal biclique reduces to constructing W, λ, μ. Define $\mu = 1/\sqrt{mn}$. Define W as follows

$$W = \frac{1}{\sqrt{mn}} \cdot \begin{array}{c} \\ U^* \\ \\ U-U^* \end{array} \begin{array}{c} \begin{array}{cc} V^* & \qquad\qquad V-V^* \end{array} \\ \left[\begin{array}{cc} 0 & \begin{cases} 1 \text{ in } E \\ \frac{-q_j}{m-q_j} \text{ in } \bar{E} \end{cases} \\ \begin{cases} 1 \text{ in } E \\ \frac{-p_i}{n-p_i} \text{ in } \bar{E} \end{cases} & \begin{cases} 1 \text{ in } E \\ -\gamma \text{ in } \bar{E} \end{cases} \end{array}\right] \end{array}$$

$p_i = \#$ of V^*-nodes adjacent to $i \in U - U^*$
$q_j = \#$ of U^*-nodes adjacent to $j \in V - V^*$

Finding W, λ, μ

Thus, showing that (NNR) finds the optimal biclique reduces to constructing W, λ, μ. Define $\mu = 1/\sqrt{mn}$. Define W as follows

$$W = \frac{1}{\sqrt{mn}} \cdot \begin{array}{c} \\ U^* \\ \\ U-U^* \end{array} \begin{array}{c} V^* \qquad\qquad\qquad V-V^* \\ \left[\begin{array}{cc} 0 & \begin{cases} 1 \text{ in } E \\ \frac{-q_j}{m-q_j} \text{ in } \bar{E} \end{cases} \\ \begin{cases} 1 \text{ in } E \\ \frac{-p_i}{n-p_i} \text{ in } \bar{E} \end{cases} & \begin{cases} 1 \text{ in } E \\ -\gamma \text{ in } \bar{E} \end{cases} \end{array}\right] \end{array} \quad \begin{array}{l} W\bar{\mathbf{v}} = \mathbf{0} \\ W^T\bar{\mathbf{u}} = \mathbf{0} \end{array}$$

$$p_i = \# \text{ of } V^*\text{-nodes adjacent to } i \in U - U^*$$
$$q_j = \# \text{ of } U^*\text{-nodes adjacent to } j \in V - V^*$$

Finding W, λ, μ

Thus, showing that (NNR) finds the optimal biclique reduces to constructing W, λ, μ. Define $\mu = 1/\sqrt{mn}$. Define W as follows

$$W = \frac{1}{\sqrt{mn}} \cdot \begin{array}{c} \\ U^* \\ \\ U-U^* \end{array} \overset{\begin{array}{cc} V^* \qquad\qquad & V-V^* \end{array}}{\begin{bmatrix} 0 & \begin{cases} 1 \text{ in } E \\ \frac{-q_j}{m-q_j} \text{ in } \bar{E} \end{cases} \\ \begin{cases} 1 \text{ in } E \\ \frac{-p_i}{n-p_i} \text{ in } \bar{E} \end{cases} & \begin{cases} 1 \text{ in } E \\ -\gamma \text{ in } \bar{E} \end{cases} \end{bmatrix}}$$

$W\bar{\mathbf{v}} = \mathbf{0}$
$W^T\bar{\mathbf{u}} = \mathbf{0}$
$\|W\| < 1$?

$p_i = \#$ of V^*-nodes adjacent to $i \in U - U^*$
$q_j = \#$ of U^*-nodes adjacent to $j \in V - V^*$

Norm of W: deterministic case

- In the case of deterministic (adversary-chosen) diversionary edges, use fairly weak bound that $\|W\| \leq \|W\|_F$. Fractions $p_i/(n - p_i)$ and $q_j/(m - q_j)$ are bounded by imposing assumption that $p_i \leq \alpha n$, $q_j \leq \beta n$. Remaining entries bounded by assuming $|E - (U^* \times V^*)| \leq O(mn)$.
- With these assumption, $\|W\|_F \leq 1$.
- The bounds in these assumptions are the best possible (up to constant factors).

Norm of W: randomized case

Assume non-clique edges are included in G independently at random with probability p.

$$W = \frac{1}{\sqrt{mn}} \cdot \begin{array}{c} \\ U^* \\ \\ U-U^* \end{array} \overset{\begin{array}{cc} V^* & \qquad\qquad V-V^* \end{array}}{\begin{bmatrix} 0 & \begin{cases} 1 \text{ in } E \\ \frac{-q_j}{m-q_j} \text{ in } \bar{E} \end{cases} \\ \begin{cases} 1 \text{ in } E \\ \frac{-p_i}{n-p_i} \text{ in } \bar{E} \end{cases} & \begin{cases} 1 \text{ in } E \\ -\gamma \text{ in } \bar{E} \end{cases} \end{bmatrix}}$$

Take $\gamma = -p/(1-p)$; then W is fairly close to a random matrix with independent entries of mean 0.

Norm of a random matrix

- Theorem (Geman, 1980). Suppose $\hat{W}$ is an $M \times N$ random matrix with $M \sim N$ and with entries chosen independently from a fixed distribution whose mean is 0 (plus a few other assumptions). Then with probability exponentially close to 1, $\|\hat{W}\| \leq O(\sqrt{N})$.
- Much better than Frobenius norm estimate.
- Note: $W \approx \hat{W}/\sqrt{mn}$.
- Implies that we can take M, N as large as m^2, n^2 and still obtain $\|W\| \leq 1$.

Analysis of the perturbation

- Must derive bound on $\|W - \hat{W}/\sqrt{mn}\|$.
- Must bound the difference between entries of the form $-p_i/(n - p_i)$ and $-p/(1 - p)$ in $(U - U^*) \times V^*$ block of W.
- Our analysis uses Bernstein's technique described in Hoeffding (1962).

Analysis of randomization in clique case

- Follows the same lines, except in place of Geman's theorem we require Füredi and Komlós's (1981) analysis of the norm of a random symmetric matrix.
- Similar result obtained: our algorithm can find a "planted" clique of with n nodes, $n(n-1)/2$ edges, in a random graph with $O(n^2)$ vertices (and hence $O(n^4)$ edges).

Algorithms

- Nuclear norm minimization can be reduced to SDP, hence (NNR) is solvable with an SDP package.
- We have a preliminary implementation of an (NNR) solver that uses Boyd's cvx on top of SDPT3.
- Toh and Yun's (2009) work on nuclear norm minimization seems promising but applies only to equality constraints.

Conclusions and open questions

- Convex relaxation can find a clique or biclique in a graph that contains the clique and biclique plus many diversionary edges.
- If the diversionary edges are placed at random, then the algorithm can tolerate many more of them.
- Would be interesting to extend the technique to other information retrieval problems.