MAT 585: Max-Cut and Stochastic Block Model

Afonso S. Bandeira

April 2, 2015

Today we discuss two NP-hard problems, `Max-Cut` and minimum bisection. As these problems are NP-hard, unless the widely believed $P \neq NP$ conjecture is false, these problem cannot be solved in polynomial time for every instance. These leaves us with a couple of alternatives: either we look for algorithms that approximate the solution, or consider algorithms that work for "typical" instances, but not for all. We will study a particular algorithm of the first type for `Max-Cut` and an algorithm of the second type for minimum bisection.

1 Max-Cut

The objective of approximation algorithms is to efficiently compute approximate solutions to problems for which finding an exact solution is believed to be computationally hard. The `Max-Cut` problem is the following: Given a graph $G = (V, E)$ with non-negative weights w_{ij} on the edges, find a set $S \subset V$ for which $\text{cut}(S)$ is maximal. Goemans and Williamson [3] introduced an approximation algorithm that runs in polynomial time and has a randomized component to it, and is able to obtain a cut whose expected value is guaranteed to be no smaller than a particular constant α_{GW} times the optimum cut. The constant α_{GW} is referred to as the approximation ratio.

Let $V = \{1, \ldots, n\}$. One can restate `Max-Cut` as

$$\begin{array}{ll} \max & \frac{1}{2}\sum_{i<j} w_{ij}(1 - y_i y_j) \\ s.t. & |y_i| = 1 \end{array} \qquad (1)$$

The y_i's are binary variables that indicate set membership, i.e., $y_i = 1$ if $i \in S$ and $y_i = -1$ otherwise.

Consider the following relaxation of this problem:

$$\begin{array}{ll} \max & \frac{1}{2}\sum_{i<j} w_{ij}(1 - u_i^T u_j) \\ s.t. & u_i \in \mathbb{R}^n, \|u_i\| = 1. \end{array} \tag{2}$$

This is in fact a relaxation because if we restrict u_i to be a multiple of e_1, the first element of the canonical basis, then (2) is equivalent to (1). For this to be a useful approach, the following two properties should hold:

(a) Problem (2) needs to be easy to solve.

(b) The solution of (2) needs to be, in some way, related to the solution of (1).

We start with property (a). Set X to be the Gram matrix of $u_1, \ldots, u_n$, that is, $X = U^T U$ where the i'th column of U is u_i. We can rewrite the objective function of the relaxed problem as

$$\frac{1}{2}\sum_{i<j} w_{ij}(1 - X_{ij})$$

One can exploit the fact that X having a decomposition of the form $X = Y^T Y$ is equivalent to being positive semidefinite, denoted $X \succeq 0$. The set of PSD matrices is a convex set. Also, the constraint $\|u_i\| = 1$ can be expressed as $X_{ii} = 1$. This means that the relaxed problem is equivalent to the following semidefinite program (SDP):

$$\begin{array}{ll} \max & \frac{1}{2}\sum_{i<j} w_{ij}(1 - X_{ij}) \\ s.t. & X \succeq 0 \text{ and } X_{ii} = 1,\ i = 1, \ldots, n. \end{array} \tag{3}$$

This SDP can be solved (up to ϵ accuracy) in time polynomial on the input size and $\log(\epsilon^{-1})$.

There is an alternative way of viewing (3) as a relaxation of (1). By taking $X = yy^T$ one can formulate a problem equivalent to (1)

$$\begin{array}{ll} \max & \frac{1}{2}\sum_{i<j} w_{ij}(1 - X_{ij}) \\ s.t. & X \succeq 0\ ,\ X_{ii} = 1,\ i = 1, \ldots, n, \text{ and } \operatorname{rank}(X) = 1. \end{array} \tag{4}$$

The SDP (3) can be regarded as a relaxation of (4) obtained by removing the non-convex rank constraint. In fact, this is how we will later formulate a similar relaxation for the minimum bisection problem.

We now turn to property (b), and consider the problem of forming a solution to (1) from a solution to (3). From the solution $\{u_i\}_{i=1,\ldots,n}$ of the relaxed problem (3), we produce a cut subset S' by randomly picking a vector $r \in \mathbb{R}^n$ from the uniform distribution on the unit sphere and setting

$$S' = \{i | r^T u_i \geq 0\}$$

In other words, we separate the vectors $u_1, \ldots, u_n$ by a random hyperplane (perpendicular to r). We will show that the cut given by the set S' is comparable to the optimal one.

Let W be the value of the cut produced by the procedure described above. Note that W is a random variable, whose expectation is easily seen to be given by (**Draw!**)

$$\begin{aligned} \mathbb{E}[W] &= \sum_{i<j} w_{ij} \Pr\left\{\operatorname{sign}(r^T u_i) \neq \operatorname{sign}(r^T u_j)\right\} \\ &= \sum_{i<j} w_{ij} \frac{1}{\pi} \arccos(u_i^T u_j). \end{aligned}$$

If we define α_{GW} as

$$\alpha_{GW} = \min_{-1 \leq x \leq 1} \frac{\frac{1}{\pi} \arccos(x)}{\frac{1}{2}(1-x)},$$

it can be shown that $\alpha_{GW} > 0.87$.

It is then clear that

$$\mathbb{E}[W] = \sum_{i<j} w_{ij} \frac{1}{\pi} \arccos(u_i^T u_j) \geq \alpha_{GW} \frac{1}{2} \sum_{i<j} w_{ij}(1 - u_i^T u_j). \tag{5}$$

Let ρ_G be the maximum cut of G, meaning the maximum of the original problem (1). Naturally, the optimal value of (2) is larger or equal than ρ_G. Hence, an algorithm that solves (2) and uses the random rounding procedure described above produces a cut W satisfying

$$\mathbb{E}[W] \geq \alpha_{GW} \frac{1}{2} \sum_{i<j} w_{ij}(1 - u_i^T u_j) \geq \alpha_{GW} \rho_G, \tag{6}$$

thus having an approximation ratio (in expectation) of α_{GW}. Note that one can run the randomized rounding procedure several times and select the best cut.

2 The Stochastic Block Model

While approximation ratio guarantees are remarkable, they are worst-case guarantees and hence pessimistic in nature. In what follows we analyze the performance of relaxations on typical instances (where typical is defined through some natural distribution of the input).

We focus on the problem of minimum graph bisection. The objective is to partition a graph in two equal-sized disjoint sets (S, S^c) while minimizing $\text{cut}(S)$ (note that in the previous section we were maximizing it instead!).

We consider a random graph model that produces graphs that have a clustering structure. Let n be an even positive integer. Given two sets of $\frac{n}{2}$ nodes consider the following random graph G: For each pair (i, j) of nodes, (i, j) is an edge of G with probability p if i and j are in the same set, and with probability q if they are in different sets. Each edge is drawn independently and $p > q$. This is known as the Stochastic Block Model.

(Think of nodes as habitants of two different towns and edges representing friendships, in this model, people leaving in the same town are more likely to be friends)

The goal will be to recover the original partition. This problem is clearly easy if $p = 1$ and $q = 0$ and hopeless if $p = q$. The question we will try to answer is for which values of p and q is it possible to recover the partition (perhaps with high probability). As $p > q$, we will try to recover the original partition by attempting to find the minimum bisection of the graph.

2.1 What does the spike model suggest?

Motivated by what we saw in previous lectures, one approach could be to use a form of spectral clustering to attempt to partition the graph.

Let A be the adjacency matrix of G, meaning that

$$A_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E(G) \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Note that in our model, A is a random matrix. We would like to solve

$$\begin{aligned} \max & \sum_{i,j} A_{ij} x_i x_j \\ \text{s.t. } & x_i = \pm 1, \forall_i \\ & \sum_j x_j = 0, \end{aligned} \tag{8}$$

The intended solution x takes the value $+1$ in one cluster and -1 in the other.

The condition $x_i = \pm 1, \ \forall_i$ could be relaxed to $\|x\|_2^2 = n$ yielding a spectral method

$$\begin{aligned} \max \ & \sum_{i,j} A_{ij} x_i x_j \\ \text{s.t. } & \|x\|_2 = \sqrt{n} \\ & \mathbf{1}^T x = 0 \end{aligned} \tag{9}$$

The solution consists of taking the top eigenvector of A (orthogonal to the all-ones vector $\mathbf{1}$).

The matrix A is a random matrix whose expectation is given by

$$\mathbb{E}[A] = \begin{cases} p & \text{if } (i,j) \in E(G) \\ q & \text{otherwise.} \end{cases}$$

Let g denote a vector that is $+1$ in one of the clusters and -1 in the other (note that this is the vector we are trying to find!). Then we can write

$$\mathbb{E}[A] = \frac{p+q}{2}\mathbf{1}\mathbf{1}^T + \frac{p-q}{2} g g^T,$$

and

$$A = \left(A - \mathbb{E}[A]\right) + \frac{p+q}{2}\mathbf{1}\mathbf{1}^T + \frac{p-q}{2} g g^T.$$

In order to remove the term $\frac{p+q}{2}\mathbf{1}\mathbf{1}^T$ we consider the random matrix

$$\mathcal{A} = A - \frac{p+q}{2}\mathbf{1}\mathbf{1}^T.$$

It is easy to see that

$$\mathcal{A} = \left(\mathcal{A} - \mathbb{E}[\mathcal{A}]\right) + \frac{p-q}{2} g g^T.$$

This means that $\mathcal{A}$ is a superposition of a random matrix whose expected value is zero and a rank-1 matrix, i.e.

$$\mathcal{A} = W + \lambda v v^T$$

where $W = \left(\mathcal{A} - \mathbb{E}[\mathcal{A}]\right)$ and $\lambda v v^T = \frac{p-q}{2} n \left(\frac{g}{\sqrt{n}}\right)\left(\frac{g}{\sqrt{n}}\right)^T$. In previous lectures we saw that for large enough λ, the eigenvalue associated with λ pops outside the distribution of eigenvalues of W and whenever this happens, the leading eigenvector has a non-trivial correlation with g (the eigenvector associated with λ). We will soon return to such estimates.

Note that since to obtain $\mathcal{A}$ we simply subtracted a multiple of $\mathbf{1}\mathbf{1}^T$ from A, problem (9) is equivalent to

$$\begin{aligned} \max \ & \sum_{i,j} \mathcal{A}_{ij} x_i x_j \\ \text{s.t. } & \|x\|_2 = \sqrt{n} \\ & \mathbf{1}^T x = 0 \end{aligned} \tag{10}$$

Now that we removed a suitable multiple of $\mathbf{1}\mathbf{1}^T$ we will even drop the constraint $\mathbf{1}^T x = 0$, yielding

$$\begin{aligned} \max \ & \sum_{i,j} \mathcal{A}_{ij} x_i x_j \\ \text{s.t. } & \|x\|_2 = \sqrt{n}, \end{aligned} \tag{11}$$

whose solution is the top eigenvector of $\mathcal{A}$.

Recall that if $\mathcal{A} - \mathbb{E}[\mathcal{A}]$ was a Wigner matrix with i.i.d entries with zero mean and variance σ^2 then its empirical spectral density would follow the semicircle law and it will essentially be supported in $[-2\sigma\sqrt{n}, 2\sigma\sqrt{n}]$. From results derived in the homework problems, we would then expect the top eigenvector of $\mathcal{A}$ to correlate with g as long as

$$\frac{p-q}{2} n > \frac{2\sigma\sqrt{n}}{2}. \tag{12}$$

Unfortunately $\mathcal{A} - \mathbb{E}[\mathcal{A}]$ is not a Wigner matrix in general. In fact, half of its entries have variance $p(1-p)$ while the variance of the other half is $q(1-q)$.

If we were to take $\sigma^2 = \frac{p(1-p)+q(1-q)}{2}$ and use (12) it would suggest that the leading eigenvector of $\mathcal{A}$ correlates with the true partition vector g as long as

$$\frac{p-q}{2} > \frac{1}{\sqrt{n}} \sqrt{\frac{p(1-p)+q(1-q)}{2}}, \tag{13}$$

However, this argument is not necessarily valid because the matrix is not a Wigner matrix. For the special case $q = 1-p$, all entries of $\mathcal{A} - \mathbb{E}[\mathcal{A}]$ have the same variance and they can be made to be identically distributed by conjugating with gg^T. This is still an impressive result, it says that if $p = 1-q$ then $p-q$ needs only to be around $\frac{1}{\sqrt{n}}$ to be able to make an estimate that correlates with the original partitioning!

An interesting regime (motivated by friendship networks in social sciences) is when the average degree of each node is constant. This can be achieved by taking $p = \frac{a}{n}$ and $q = \frac{b}{n}$ for constants a and b. While the argument presented to justify condition (13) is not valid in this setting, it nevertheless suggests that the condition on a and b needed to be able to make an estimate that correlates with the original partition is

$$(a-b)^2 > 2(a-b). \tag{14}$$

Remarkably this was posed as conjecture by Decelle et al. [2] and proved in a series of works by Mossel et al. [5, 6] and Massoulie [4].

2.2 Exact recovery

We now turn our attention to the problem of recovering the cluster membership of every single node correctly, not simply having an estimate that correlates with the true labels. If the probability of intra-cluster edges is $p = \frac{a}{n}$ then it is not hard to show that each cluster will have isolated nodes making it impossible to recover the membership of every possible node correctly. In fact this is the case whenever $p \ll \frac{2\log n}{n}$. For that reason we focus on the regime

$$p = \frac{\alpha \log(n)}{n} \text{ and } q = \frac{\beta \log(n)}{n}, \tag{15}$$

for some constants $\alpha > \beta$.

Let $x \in \mathbb{R}^n$ with $x_i = \pm 1$ representing the partition (note there is an ambiguity in the sense that x and $-x$ represent the same partition). Then, if we did not worry about efficiency then our guess (which corresponds to the Maximum Likelihood Estimator) would be the solution of the minimum bissection problem (8).

In fact, one can show (but this will not be the focus of this lecture, see [1] for a proof) that if

$$\frac{\alpha+\beta}{2} - \sqrt{\alpha\beta} > 1, \tag{16}$$

then, with high probability, (8) recovers the true partition. Moreover, if

$$\frac{\alpha+\beta}{2} - \sqrt{\alpha\beta} < 1,$$

no algorithm (efficient or not) can, with high probability, recover the true partition.

In the next lecture we derive conditions that guarantee exact recovery of the true partition with high probability, using a semidefinite program.

References

[1] E. Abbe, A. S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *Available online at arXiv:1405.3267 [cs.SI]*, 2014.

[2] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E*, 84:066106, December 2011.

[3] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefine programming. *Journal of the Association for Computing Machinery*, 42, 1995.

[4] L. Massoulié. Community detection thresholds and the weak Ramanujan property. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, pages 694–703, New York, NY, USA, 2014. ACM.

[5] E. Mossel, J. Neeman, and A. Sly. Stochastic block models and reconstruction. *Available online at arXiv:1202.1499 [math.PR]*, 2012.

[6] E. Mossel, J. Neeman, and A. Sly. A proof of the block model threshold conjecture. *Available online at arXiv:1311.4115 [math.PR]*, January 2014.