18.409 The Behavior of Algorithms in Practice March 14, 2002

Lecture 10

Lecturer: Dan Spielman *Scribe: Fernando Ordóñez*

1 Smoothed model for bandwidth minimization

In the last lecture we presented Turner's analysis of the Cuthill-Mckee heuristic for bandwidth minimization. The problem of bandwidth minimization of a graph is to find an ordering of the nodes of the graph which has the minimum possible bandwidth, where the bandwidth is defined as the largest difference in ordering among nodes that are linked.

This problem corresponds to finding the order of rows and columns that minimizes the bandwidth for the vertex adjacency matrix. This is a square, symmetric, sparse matrix. The problem is studied for a graph $G_b(n,p)$ on n nodes, such that neighbors up to b nodes away are connected with probability p. This graph can be represented by the matrix $M \in \{0,1\}^{n\times n}$ such that

$$M_{i,j} = \begin{cases} 1 & |i-j| \leq b, \text{ with probability } p \\ 0 & |i-j| \leq b, \text{ with probability } 1-p \\ 0 & |i-j| > b \ . \end{cases}$$

In the next section, we will see that Turner's analysis can be directly translated to smooth analysis.

Let $G_b(n)$ be the family of graphs on n nodes such that if $|i-j| > b$ then $(i,j) \notin E$. That is the family of graphs which have bandwidth at most b. Let A be an algorithm, such that $A(G)$ is the quality of the output of algorithm A.

1.1 Blum and Spencer, 1995

Blum and Spencer suggested to analyze an algorithm in a semi-random model, where the inputs distribution is partially controlled by an adversary. This model is represented by:

$$\max_{G\in G_b(n)} E_{P\in G_b(n,p)} \left[A(G \oplus P)\right] \ ,$$

an adversary chooses $G \in G_b(n)$, in other words, the adversary chooses for each pair of vertices that are potential neighbors ($|i - j| \leq b$), whether he joins them or not. This choice can be flipped with probability p. This is the addition of the random perturbation $P \in G_b(n,p)$, where the addition allows to add or remove edges. This is a natural smoothed model for the problem of bandwidth minimization.

Turner's theorem aplies with the same proof to this model, although this does not become obvious until we strengthen the model.

Blum and Spencer suggested a stronger model, known as the monotone adversary model, in which the graph is drawn ramdomly, $P \in G_b(n,p)$ and then an adversary is allowed alter the random decisions in a monotonic manner, that is only allowed to add edges.

How this is related to the Blum and Spencer model? We first note that the semi-random model is weaker than a semi random model that only allowes random addition of edges, i.e.

$$\max_{G \in G_b(n)} E_{P \in G_b(n,p)} [A(G \oplus P)] \leq \max_{G \in G_b(n)} E_{P \in G_b(n,p)} [A(G \cup P)] \ .$$

This is true because the random deletion of edges, can be done by the adversary in selecting the graph G.

We now note that the monotone adversary model can be obtained by from the semi-random model in which edges are only added by reversing the orders of the quantifiers. That is,

$$\max_{G \in G_b(n)} E_{P \in G_b(n,p)} [A(G \cup P)] \leq E_{P \in G_b(n,p)} \max_{G \in G_b(n)} [A(G \cup P)] \ ,$$

where the left-hand term gives the performance in the monotone adversary model. Feige and Krauthgamer point out that Turner's analysis applies directly to the monotone adversary model. Thus, Turner's analysis provides lower bounds on the resulting bandwidth of the graph and performace guarantees for level algorithms to find the bandwidth in the monotone adversary model. By the arguments above, this analysis can be applied to the semi-random adversary or smoothed model as well.

2 Graph Bisection

The problem is: for a given graph G, to split the nodes in two groups evenly (in reality don't need to cut evenly), cutting as few edges as possible

Definition 1. (A first algorithm)

1. *Start with any bisection.*
2. *Greedy: find nodes in each side with the highest degree to the other side and swap.*
3. *Repeat 2.*

This First algorithm works well but tends to get stuck at local minimum. The following algorithm incorporates a method that allows to escape a local minimum.

Definition 2. (Kernighan-Lin algorithm)

1. *Begin with any bisection*
2. *Do Greedy $n/2$ times*
   - *never reusing nodes*
   - *do even if hurts*

   Note that after doing all $n/2$ swaps, all nodes have changed side.
3. *Recall best configuration*
4. *Repeat 2.*

This algorithm, which is of the class of "tabu search" methods (we keep a list of tabu nodes, which can not be used), works very well in practice. But there are no theoretical results on this.

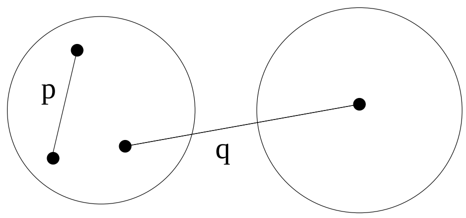
A way to check the improvement of these algorithms, is to construct lower bounds to the optimal number of bisection, by for example a Linear Programming relaxation formulation.

Definition 3. (Simulated Annealing algorithm, Jerrum - Sorkin made the analysis)

- *Begin with a bisection*
- *Pick random pairs*
- *Flip is improvement is above a threshold*

Now we turn to a special type of graph, which allows for a nicer analysis of some bisection algorithms. We will consider graphs made from the Planted Bisection Model.

Definition 4. Planted Bisection Model



Is a graph constructed by separating the nodes in two sets, defining edges within a group independently with probability p, and between groups independently with probability q. For $q < p$.

It can be shown that if q is sufficiently $< p$, then the original is the only optimal bisection.

There is a paper by R. B. Boppana, which analyzes this problem with the average case analysis, when $p - q \sim \frac{1}{\sqrt{n}}$.

Condon and Karp, 2000 provide a linear-time algorithm, that if $p - q \geq \frac{1}{n^{.5-\epsilon}}$, for some $\epsilon > 0$, then the algorithm finds the optimal partition with high probability.

McSherry considers the matrix of probabilities of the edges

$$A = \begin{pmatrix} p & \cdots & p & q & \cdots & q \\ \vdots & & \vdots & \vdots & & \vdots \\ p & \cdots & p & q & \cdots & q \\ q & \cdots & q & p & \cdots & p \\ \vdots & & \vdots & \vdots & & \vdots \\ q & \cdots & q & p & \cdots & p \end{pmatrix}$$

Let B be the adjecency matrix of the graph obtained by the planted partition. Then think of B as a rounding of A. Therefore the eigenvalues of these matrices are related.

Matrix A has rank 2, and by inspection we note that the largest eigenvalue is $\frac{n}{2}(p+q)$ with eigenvector the vector of all ones, that is

$$w_1^t = (1, \cdots, 1) \ ,$$

and the second largest eigenvalue is $\frac{n}{2}(p-q)$ with eigenvector

$$w_2^t = (1, \cdots, 1, -1, \cdots, -1) \ .$$

All other eigenvalues of matrix A are equal to 0.

Definition 5. (McSherry's algorithm)

- *Take 2 eigenvectors of B with the largest eigenvalues.*
- *Use second eigenvector to partition the graph.*

For the analysis of McSherry's algorithm, we need the following two theorems

Theorem 6. *For A is a symmetric matrix and $E = A - B$. If $\lambda_1, \lambda_2, \cdots, \lambda_n$ are eigneval-ues of A, with eigenvectors $v_1, v_2, \cdots, v_n$ and $w_1, w_2, \cdots, w_n$ eigenvectors of B, with $\theta_i = \mathrm{angle}(v_i, w_i)$. Then*

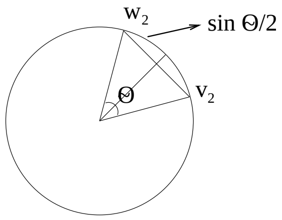
$$\theta_i \sim \frac{1}{2}\sin(2\theta_i) \leq \frac{\|E\|}{\mathrm{gap}(i, A)} \; ,$$

where $\mathrm{gap}(i, A) = \min_{j \neq i} |\lambda_i - \lambda_j|$.

Theorem 7. *(Alon-Krivelevich-Vu)*

In this case $\|E\| \leq 4p\sqrt{n}$ with probability $1 - \frac{1}{\exp(n)}$.

In the analysis of McSherry's algorithm, note that $\mathrm{gap}(2, A) = \frac{n}{2}(p - q)$ which implies that $\theta_2 \leq \frac{8p}{\sqrt{n}(p-q)}$. Note that $v_2^t = \frac{1}{\sqrt{n}}(1, \cdots, 1, -1, \cdots, -1)$ and that w_2^t is such that $(w_2)_i > 0$ for one group and $(w_2)_i < 0$ for the other group. If we missclassify k entries, then $\|v_2 - w_2\| \geq \sqrt{\frac{k}{n}}$, since for each missclassified entries there is least a contrbution of $\frac{1}{\sqrt{n}}$ in $v_2 - w_2$.



By the geometric argument above we note that $\|v_2 - w_2\| \leq 2\sin(\theta_2/2) \sim \frac{1}{2}\sin(2\theta_2)$. From the previous bounds,

$$\sqrt{\frac{k}{n}} \leq \|v_2 - w_2\| \leq \frac{1}{2}\sin(2\theta_2) \leq \frac{8p}{\sqrt{n}(p-q)} \; ,$$

this implies that

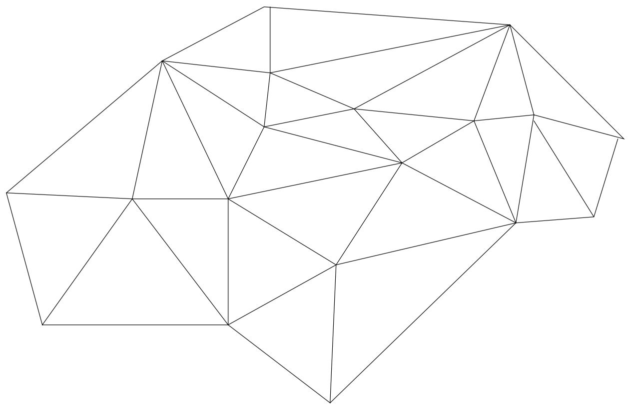
$$k \leq \left(\frac{8p}{p-q}\right)^2 \; .$$

This is a bound, independent of n, of the number of missclassified entries.

Ideas on Random Graphs

Graphs in practice are not planted bisection. What can be said for other types of graphs?

For example Fully triangulated planar graphs.



Consider perturbations where:
- delete a random edge
- add the other diagonal in the square formed

What can be proven with this perturbation model?

Property-Preserving Perturbations

Define a function f, from the space of graphs considered, that for a graph G, returns in $f(G)$ a vector where each coordinate is a property of interest (max cut, coloring number, graph is planar or not, etc).

Some coordinates can be in the reals (the max flow of the graph), while other coordinates can be binary, the graph has a property or does not.

Can define different perturbation models. For

$$P \leftarrow G(n,p) \quad \text{analyze} \quad f(G+P) \ ,$$

or

$$P \leftarrow G(n,p) \quad \text{consider perturbations that} \quad f(G+P) = f(G) \ ,$$

or

$$P \leftarrow G(n,p) \quad \text{consider perturbations that} \quad \|f(G+P) - f(G)\| \leq \epsilon \ .$$

paper on the subject distributed in the next class.