CS 598CSC: Approximation Algorithms	Lecture date: March 20th, 2009
Instructor: Chandra Chekuri	Scribe: Alina Ene

1 The Multicut Problem

In the MULTICUT problem, we are given a graph $G = (V, E)$, a capacity function that assigns a capacity c_e to each edge e, and a set of pairs $(s_1, t_1), ..., (s_k, t_k)$. The MULTICUT problem asks for a minimum capacity set of edges $F \subseteq E$ such that removing the edges in F disconnects s_i and t_i, for all i. Note that the MULTICUT problem generalizes the MULTIWAY CUT problem that we saw in the last two lectures. In this lecture, we will see a $\Theta(\log k)$ approximation algorithm for MULTICUT.

Let's start by giving an LP formulation for the problem. For each edge e, we have a variable d_e. We interpret each variable d_e as a distance label for the edge. Let $\mathcal{P}_{s_i,t_i}$ denote the set of all paths between s_i and t_i. We have the following LP for the problem:

$$\begin{aligned} \min \quad & \sum_{e \in E} c_e d_e \\ \text{s.t.} \quad & \\ & \sum_{e \in p} d_e \geq 1 \quad p \in \mathcal{P}_{s_i,t_i}, 1 \leq i \leq k \\ & d_e \geq 0 \quad\quad e \in E \end{aligned}$$

The LP tries to assign distance labels to edges so that, on each path p between s_i and t_i, the distance labels of the edges on p sum up to at least one. Note that, even though the LP can have exponentially many constraints, we can solve the LP in polynomial time using the ellipsoid method and the following separation oracle. Given distance labels d_e, we set the length of each edge to d_e and, for each pair (s_i, t_i), we compute the length of the shortest path between s_i and t_i and check whether it is at least one. If the shortest path between s_i and t_i has length smaller than one, we have a violated constraint. Conversely, if all shortest paths have length at least one, the distance labels define a feasible solution.

Let's also construct the dual of the previous LP. For each path p between any pair (s_i, t_i) we have a dual variable f_p. We interpret each variable f_p as the amount of flow between s_i and t_i that is routed along the path p. We have the following dual LP:

$$\begin{aligned} \max \quad & \sum_{i=1}^{k} \sum_{p \in \mathcal{P}_{s_i,t_i}} f_p \\ \text{s.t.} \quad & \\ & \sum_{p:\, e \in p} f_p \leq c_e \quad\quad e \in E(G) \\ & f_p \geq 0 \quad p \in \mathcal{P}_{s_1,t_1} \cup ... \cup \mathcal{P}_{s_k,t_k} \end{aligned}$$

The dual is an LP formulation for the Maximum Throughput Multicommodity Flow problem. In the Maximum Throughput Multicommodity Flow problem, we have k different commodities. For each i, we want to route commodity i from the source s_i to the destination t_i. Each commodity must satisfy flow conservation at each vertex other than its source and its destination. Additionally, the total flow routed on each edge must not exceed the capacity of the edge. The goal is to maximize the sum of the commodities routed.

The dual LP tries to assign an amount of flow f_p to each path p so that the total flow on each edge is at most the capacity of the edge (the flow conservation constraints are automatically satisfied). Note that the endpoints of the path p determine which kind of commodity is routed along the path.

We conclude this section with an exercise for the reader.

Exercise 1. *Write the* MULTICUT *LP and its dual in a compact form with polynomially many constraints.*

2 Upper Bound on the Integrality Gap

In this section, we will show that the integrality gap of the LP is $O(\log k)$ using a randomized rounding algorithm due to Calinescu, Karloff, and Rabani [1]. The first algorithm that achieved an $O(\log k)$-approximation for MULTICUT is presented in Chapter 20 of the textbook. This algorithm is due to Garg, Vazirani, and Yannakakis [2] and it is based on the region growing technique introduced by Leighton and Rao [3].

Let $B_d(v, r)$ denote the ball of radius r centered at the vertex v in the metric induced by the distance labels d_e.

Randomized Rounding:

Solve the LP to get the distance labels d_e
Pick θ uniformly at random from $[0, 1/2)$
Pick a random permutation σ on $\{1, 2, ..., k\}$
for $i = 1$ to k
  $V_{\sigma(i)} = B_d(s_{\sigma(i)}, \theta) \setminus \bigcup_{j<i} V_{\sigma(j)}$
Output $\bigcup_{i=1}^{k} \delta(V_i)$

Lemma 1. *The Randomized Rounding algorithm correctly outputs a multicut.*

Proof: Let F be the set of edges constructed by the algorithm. Suppose F is not a multicut. Then there exists a pair of vertices (s_i, t_i) such that there is a path between s_i and t_i in $G[E \setminus F]$ (recall that $G[E \setminus F]$ is the subgraph of G induced by the edges in $E \setminus F$). Therefore there exists a j such that V_j contains s_i and t_i. Since $V_j \subseteq B_d(s_j, \theta)$, both s_i and t_i are contained in the ball of radius θ centered at s_j. Consequently, the distance between s_j and s_i is at most θ and the distance between s_j and t_i is at most θ. By the triangle inequality, the distance between s_i and t_i is at most 2θ. Since θ is smaller than $1/2$, it follows that the distance between s_i and t_i is smaller than one. This contradicts the fact that the distance labels d_e are a feasible solution for the LP. Therefore F is a multicut, as desired. $\square$

Lemma 2. *The probability that an edge e is cut is at most $2H_k d_e$, where H_k is the k-th harmonic number and d_e is the distance label of the edge e.*

Proof: Let's fix an edge $e = (u, v)$. Let:

$$L_i = \min\{d(s_i, u), d(s_i, v)\}$$

$$R_i = \max\{d(s_i, u), d(s_i, v)\}$$

We may assume without loss of generality that $L_1 \leq L_2 \leq ... \leq L_k$.

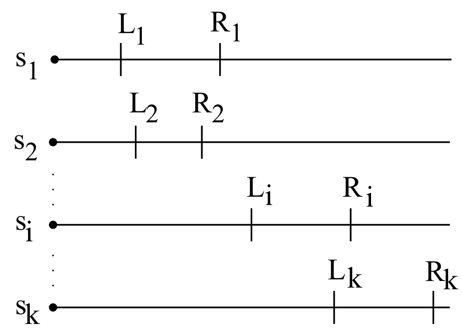

Figure 1.

Let A_i be the event that the edge e is cut first by s_i. More precisely, A_i is the event that $|V_i \cap \{u, v\}| = 1$ and $|V_j \cap \{u, v\}| = 0$ for all j such that $\sigma(j) < \sigma(i)$. Note that $|V_i \cap \{u, v\}| = 1$ simply says that s_i cuts the edge e. If s_i is the first to cut the edge e, for all j that come before i in the permutation, neither u nor v can be in V_j (if only one of u and v is in V_j, s_j cuts the edge e; if both u and v are in V_j, s_i cannot cut the edge e).

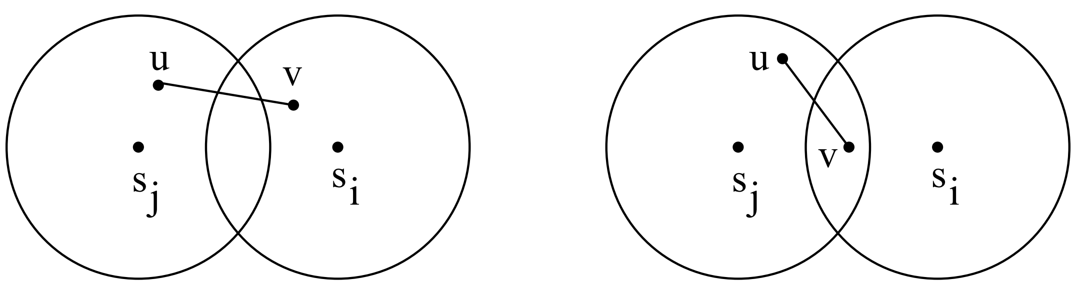

Figure 2: If $\sigma(j) < \sigma(i)$, s_i cannot be the first to cut the edge $e = (u, v)$ (on the left, s_j also cuts the edge; on the right, s_j captures the edge and therefore s_i cannot cut it).

Note that the event that the edge e is cut is the union of the disjoint events $A_1, ..., A_k$. Therefore we have:

$$\mathbb{P}(e \text{ is cut}) = \sum_i \mathbb{P}(A_i)$$

Let's fix $r \in [0, 1/2)$ and let's look at $\mathbb{P}(A_i|\theta = r)$. Note that s_i cuts the edge e only if one of u, v is inside the ball of radius r centered at s_i and the other is outside the ball. Differently said, s_i cuts the edge only if $r \in [L_i, R_i)$:

$$\mathbb{P}(A_i|\theta = r) = 0 \quad \text{ if } r \notin [L_i, R_i)$$

Now suppose that $r \in [L_i, R_i)$. Let's fix $j < i$ and suppose j comes before i in the permutation (that is, $\sigma(j) < \sigma(i)$). Recall that, since $j < i$, we have $L_j \leq L_i \leq r$. Therefore at least one of u, v is inside the ball of radius r centered at s_j. Consequently, s_i cannot be the first to cut the edge e. Therefore s_i is the first to cut the edge e only if $\sigma(i) < \sigma(j)$ for all $j < i$. Since σ is a random permutation, i appears before j for all $j < i$ with probability $1/i$. Therefore we have:

$$\mathbb{P}(A_i | \theta = r) \leq \frac{1}{i} \qquad \text{if } r \in [L_i, R_i)$$

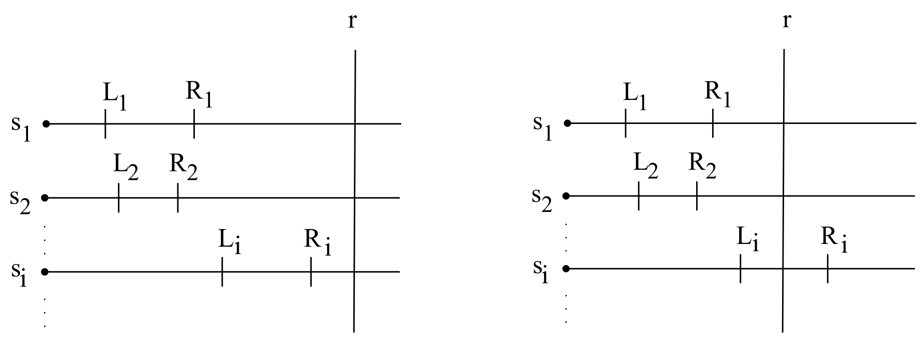

Figure 3.

Since θ was selected uniformly at random from the interval $[0, 1/2)$, we have:

$$\mathbb{P}(A_i) \leq \frac{1}{i}\, \mathbb{P}(\theta \in [L_i, R_i)) = \frac{2}{i}\, (R_i - L_i)$$

By the triangle inequality, $R_i \leq L_i + d_e$. Therefore:

$$\mathbb{P}(A_i) \leq \frac{2d_e}{i}$$

Consequently,

$$\mathbb{P}(e \text{ is cut}) = \sum_i \mathbb{P}(A_i) \leq 2H_k d_e$$

□

Corollary 3. *The integrality gap of the* Multicut *LP is* $O(\log k)$.

Proof: Let F be the set of edges outputted by the Randomized Rounding algorithm. For each edge e, let χ_e be an indicator random variable equal to 1 if and only if the edge e is in F. As we have already seen,

$$\mathbb{E}(\chi_e) = \mathbb{P}(\chi_e = 1) \leq 2H_k d_e$$

Let $c(F)$ be a random variable equal to the total capacity of the edges in F. We have:

$$\mathbb{E}(c(F)) = \mathbb{E}\left(\sum_e c_e \chi_e\right) = \sum_e c_e\, \mathbb{E}(\chi_e) \leq 2H_k \sum_e c_e d_e = 2H_k\, \mathrm{OPT}_{\mathrm{LP}}$$

Consequently, there exists a set of edges F such that the total capacity of the edges in F is at most $2H_k$ $\mathrm{OPT}_{\mathrm{LP}}$. Therefore $\mathrm{OPT} \leq 2H_k$ $\mathrm{OPT}_{\mathrm{LP}}$, as desired.

$\square$

Corollary 4. *We have:*

$$\max_{\text{m.c. flow } f} |f| \leq \min_{\text{multicut } C} |C| \leq O(\log k) \left(\max_{\text{m.c. flow } f} |f| \right)$$

where $|f|$ represents the value of the multicommodity flow f, and $|C|$ represents the capacity of the multicut C.

Proof: Let $\mathrm{OPT}_{\mathrm{LP}}$ denote the total capacity of an optimal (fractional) solution for the MULTICUT LP. Let OPT_{dual} denote the flow value of an optimal solution for the dual LP. Since $\mathrm{OPT}_{\mathrm{LP}}$ is a lower bound on the capacity of the minimum (integral) multicut, we have:

$$\max_{\text{m.c. flow } f} |f| = \mathrm{OPT}_{dual} = \mathrm{OPT}_{\mathrm{LP}} \leq \min_{\text{multicut } C} |C|$$

As we have already seen, we have:

$$\min_{\text{multicut } C} |C| \leq 2H_k \ \mathrm{OPT}_{\mathrm{LP}} = 2H_k \ \mathrm{OPT}_{dual} = 2H_k \left(\max_{\text{m.c. flow } f} |f| \right)$$

$\square$

Corollary 5. *The Randomized Rounding algorithm achieves an $O(\log k)$-approximation (in expectation) for the* MULTICUT *problem.*

Proof: As we have already seen, we have:

$$\mathbb{E}(c(F)) \leq 2H_k \ \mathrm{OPT}_{\mathrm{LP}}$$

where F is the set of edges outputted by Randomized Rounding algorithm and $c(F)$ is the total capacity of the edges in F. Since $\mathrm{OPT}_{\mathrm{LP}} \leq \mathrm{OPT}$,

$$\mathbb{E}(c(F)) \leq 2H_k \ \mathrm{OPT} = O(\log k)\mathrm{OPT}$$

$\square$

3 Lower Bound on the Integrality Gap

In this section, we will show that the integrality gap of the LP is $\Omega(\log k)$. That is, we will give a MULTICUT instance for which the LP gap is $\Omega(\log k)$. Let's start by looking at expander graphs and their properties.

3.1 Expander Graphs

Definition 1. *A graph $G = (V, E)$ is an α-edge-expander if, for any subset S of at most $|V|/2$ vertices, the number of edges crossing the cut $(S, V \backslash S)$ is at least $\alpha|S|$.*

Note that the complete graph K_n is a $(|V|/2)$-edge-expander. However, the more interesting expander graphs are also sparse. Cycles and grids are examples of graphs that are very poor expanders.

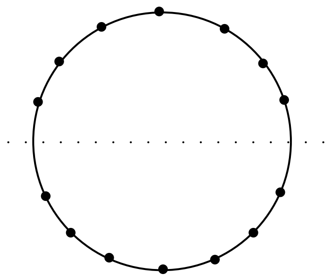
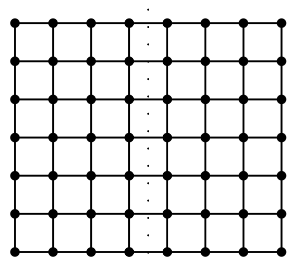

Figure 4. The top half of the cycle has $|V|/2$ vertices and only two edges crossing the cut. The left half of the grid has roughly $|V|/2$ vertices and only $\sqrt{|V|}$ edges crossing the cut.

Definition 2. *A graph G is d-regular if every vertex in G has degree d.*

Note that 2-regular graphs consist of a collection of edge disjoint cycles and therefore they have poor expansion. However, for any $d \geq 3$, there exist d-regular graphs that are very good expanders.

Theorem 6. *For every $d \geq 3$ there exists an infinite family of d-regular 1-edge-expanders.*

We will only need the following special case of the previous theorem.

Theorem 7. *There exists a constant $\alpha > 0$ and an integer n_0 such that, for all even integers $n \geq n_0$, there exists an n-vertex, 3-regular α-edge-expander.*

Proof Idea. The easiest way to prove this theorem is using the probabilistic method. The proof itself is beyond the scope of this lecture[1]. The proof idea is the following.

Let's fix an even integer n. We will generate a 3-regular random graph G by selecting three random perfect matchings on the vertex set $\{1, 2, ..., n\}$ (recall that a perfect matching is a set of edges such that every vertex is incident to exactly one of these edges). We select a random perfect matching as follows. We maintain a list of vertices that have not been matched so far. While there is at least one vertex that is not matched, we select a pair of distinct vertices u, v uniformly at random from all possible pairs of unmatched vertices. We add the edge (u, v) to our matching and we remove u and v from the list. We repeat this process three times (independently) to get three random matchings. The graph G will consist of the edges in these three matchings. Note that G is actually a 3-regular multigraph since it might have parallel edges (if the same edge is in at least two of the matchings). There are two properties of interest: (1) G is a simple graph and (2) G is an α-edge-expander for some constant $\alpha > 0$. If we can show that G has both properties with positive probability, it follows that there exists a 3-regular α-edge-expander (if no graph is a 3-regular α-edge-expander, the probability that our graph G has both properties is equal to 0).

[1] A more accurate statement is that the calculations are a bit involved and not terribly interesting for us.

It is not very hard to show that the probability that G does not have property (1) is small. To show that the probability that G does not have property (2) is small, for each set S with at most $n/2$ vertices, we estimate the expected number of edges that cross the cut $(S, V \backslash S)$ (e.g., we can easily show that $|\delta(S)| \geq |S|/2$). Using tail inequalities (e.g., Chernoff bounds), we can show that the probability that $|\delta(S)|$ differs significantly from its expectation is extremely small (i.e., small enough so that the sum – taken over all sets S – of these probabilities is also small) and we can use the union bound to get the desired result. □

Note that explicit constructions of d-regular expanders are also known. Margulis [4] gave an infinite family of 8-regular expanders. The vertex set of a graph G_n in Margulis' construction is $\mathbb{Z}_n \times \mathbb{Z}_n$, where $\mathbb{Z}_n$ is the set of all integers mod n. The neighbors of a vertex (x, y) in G_n are $(x+y, y)$, $(x-y, y)$, $(x, y+x)$, $(x, y-x)$, $(x+y+1, y)$, $(x-y+1, y)$, $(x, y+x+1)$, and $(x, y-x+1)$ (all operations are mod n). Another example is the following infinite family of 3-regular expanders. For each prime p, we have a 3-regular graph G_p. The vertex set of G_p is $\mathbb{Z}_p$. The neighbors of a vertex x in G_p are $x+1$, $x-1$, and x^{-1} (as before, all operations are mod p; x^{-1} is the inverse of x mod p, and we define the inverse of 0 to be 0)[2].

We conclude this section with the following observations (they will be very useful in showing the $\Omega(k)$ lower bound on the integrality gap of the LP).

Claim 8. *Let G be an n-vertex d-regular α-edge-expander, for some constants $d \geq 3$ and $\alpha > 0$. Then the diameter of G is $\Theta(\log n)$.*

Proof: For any two vertices u and v, let $dist(u, v)$ denote the length of a shortest path between u and v (the length of a path is the number of edges on the path). Let's fix a vertex s. Let L_i be the set of all vertices v such that $dist(s, v)$ is at most i. Now let's show that $(1+\alpha/d)|L_{i-1}| \leq |L_i| \leq d|L_{i-1}|$. Clearly, $|L_1| = d$ (since s has degree d). Therefore we may assume that $i > 1$. Every vertex in L_i is in L_{i-1} or it has a neighbor in L_{i-1}. Therefore it suffices to bound $|L_i \backslash L_{i-1}|$.

Note that any vertex in L_{i-1} has at least one neighbor in L_{i-1}. Therefore the vertices in L_{i-1} have at most $(d-1)|L_{i-1}|$ neighbors outside of L_{i-1}. Consequently, $|L_i| \leq d|L_{i-1}|$.

Now one of L_{i-1}, $V \backslash L_{i-1}$ has at most $|V|/2$ vertices. Let's assume without loss of generality that L_{i-1} has at most $|V|/2$ vertices (the other case is symmetric). Let $A = L_{i-1}$ and let B be the set of all vertices in $V \backslash L_{i-1}$ that have a neighbor in L_{i-1} (note that $|L_i| = |A| + |B|$). Let F be the set of all edges that cross the cut $(L_{i-1}, V \backslash L_{i-1})$. Now let's look at the bipartite graph $H = (A, B, F)$. Since G is an α-edge-expander, we have $|F| \geq \alpha|A|$. Moreover, $|F| = \sum_{v \in B} d_H(v)$, where $d_H(v)$ is the degree of v in H. Since $d_H(v)$ is at most d, we have $\alpha|A| \leq |F| \leq d|B|$. Therefore we have:

$$L_i = |A| + |B| \geq (1 + \alpha/d)|A| = (1 + \alpha/d)|L_{i-1}|$$

It follows by induction that $d(1 + \alpha/d)^{i-1} \leq |L_i| \leq d^i$. Therefore $dist(s, v)$ is $O(\log n)$ for all v and there exists a vertex v such that $dist(s, v)$ is $\Omega(\log n)$. Since this is true for any s, it follows that the diameter of G is $\Theta(\log n)$.

□

[2]Note that, unlike Margulis' construction, this construction is not very explicit since we don't know how to generate large primes deterministically.

Claim 9. *Let G be an n-vertex 3-regular α-edge-expander and let $B(v,i)$ be the set of all vertices u such that there is a path between u and v with at most i edges. For any vertex v, $|B(v,\log_3 n/2)| \leq \sqrt{n}$.*

Proof: Note that $B(v,\log_3 n/2)$ is the set of all vertices w such that $dist(v,w)$ is at most $\log_3 n/2$. As we have seen in the proof of the previous claim, we have $|B(v,\log_3 n/2)| \leq 3^{\log_3 n/2} = \sqrt{n}$. $\square$

3.2 The Multicut Instance

Let n_0, α be as in Theorem 7. Let $n \geq n_0$ and let G be an n-vertex 3-regular α-edge-expander. For each edge e in G, we set the capacity c_e to 1. Now let $X = \{(u,v) | u \notin B(v,\log_3 n/2)\}$. The pairs in X will be the pairs (s_i,t_i) that we want to disconnect. Let (G,X) be the resulting Multicut instance.

Claim 10. *There exists a feasible fractional solution for (G,X) of capacity $O(n/\log n)$.*

Proof: Let $d_e = 2/\log_3 n$, for all e. Note that, since G is 3-regular, G has $3n/2$ edges. Therefore the total capacity of the fractional solution is

$$\sum_e d_e = \frac{3n}{2} \cdot \frac{2}{\log_3 n} = \frac{3n}{\log_3 n}$$

Therefore we only need to show that the solution is feasible. Let (u,v) be a pair in X. Let's consider a path p between u and v. Since u is not in $B(v,\log_3 n/2)$, the path p has more than $\log_3 n/2$ edges (recall that $B(v,i)$ is the set of all vertices u such that there is a path between u and v with at most i edges). Consequently,

$$\sum_{e \in p} d_e > \frac{\log_3 n}{2} \cdot \frac{2}{\log_3 n} = 1$$

$\square$

Claim 11. *Any integral solution for (G,X) has capacity $\Omega(n)$.*

Proof: Let F be an integral solution for (G,X). Let $V_1, ..., V_h$ be the connected components of $G[E \backslash F]$. Let's fix an i and let v be a vertex in the connected component V_i. Note that, for any u in V_i, there is a path between v and u with at most $\log_3 n/2$ edges (if not, (u,v) is a pair in X which contradicts the fact that removing the edges in F disconnects every pair in X). Therefore V_i is contained in $B(v,\log_3 n/2)$. It follows from Claim 9 that $|V_i| \leq \sqrt{n}$. Since G is an α-edge-expander and $|V_i| \leq |V|/2$, we have $|\delta(V_i)| \geq \alpha |V_i|$, for all i. Consequently,

$$|F| = \frac{1}{2} \sum_{i=1}^{h} |\delta(V_i)| \geq \frac{\alpha}{2} \sum_{i=1}^{h} |V_i| = \frac{\alpha n}{2}$$

Therefore F has total capacity $\Omega(n)$ (recall that every edge has unit capacity). $\square$

Theorem 12. *The integrality gap of the* Multicut *LP is $\Omega(\log k)$.*

Proof: Note that $k = |X| = O(n^2)$. It follows from claims 10 and 11 that the LP has integrality gap $\Omega(\log n) = \Omega(\log k)$, as desired. $\square$

References

[1] G. Calinescu, H. Karloff, and Y. Rabani. Approximation algorithms for the 0-extension problem. In *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*, pages 8–16. Society for Industrial and Applied Mathematics Philadelphia, PA, USA, 2001.

[2] N. Garg, V.V. Vazirani, and M. Yannakakis. Approximate max-flow min-(multi) cut theorems and their applications. In *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, pages 698–707. ACM New York, NY, USA, 1993.

[3] T. Leighton and S. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM (JACM)*, 46(6):787–832, 1999.

[4] G.A. Margulis. Explicit constructions of expanders. *Problemy Peredaci Informacii*, 9(4):71–80, 1973.