1 Announcements

- HW4 is out, due next Wednesday.
- Solutions for HW2 are posted (or will be posted very soon).

2 Lecture Recap and Questions?

Any questions from the mini-lectures or pre-class-quiz? (Metric Embeddings; Bourgain’s Embedding)

3 Warm-Up

**Group Work**

Let $G = (V, E)$ be a weighted, undirected graph, on $n$ vertices with edge weights $w_{uv}$ on the edge $\{u, v\} \in E$. Let $d : V \times V \to \mathbb{R}$ be the associated graph metric.

Explain how to efficiently find and apply a map $f : V \to \mathbb{R}^k$, for $k = O(\log^2 n)$, so that

$$\frac{\sum_{\{u,v\} \in E} \|f(u) - f(v)\|_1}{\sum_{\{u,v\} \in \binom{V}{2}} \|f(u) - f(v)\|_1} \leq O(\log n) \frac{\sum_{\{u,v\} \in E} d(u,v)}{\sum_{\{u,v\} \in \binom{V}{2}} d(u,v)}$$

holds with high probability. Above, $\binom{V}{2}$ refers to the set of all unordered pairs $\{u, v\}$ for $u, v \in V$ and $u \neq v$.

**Group Work: Solutions**

Let $f : V \to \mathbb{R}^k$ be the map given by Bourgain’s embedding. Then for all $u, v$, we have (for some constant $b$)

$$\frac{k}{b \log n} d(u, v) \leq \|f(u) - f(v)\|_1 \leq k d(u, v),$$

and so

$$\frac{\sum_{\{u,v\} \in E} d(u,v)}{\sum_{\{u,v\} \in \binom{V}{2}} d(u,v)} \geq \frac{\sum_{\{u,v\} \in E} \frac{1}{k} \|f(u) - f(v)\|_1}{\sum_{\{u,v\} \in \binom{V}{2}} \frac{b \log n}{k} \|f(u) - f(v)\|_1} = \frac{1}{b \log n} \frac{\sum_{\{u,v\} \in E} \|f(u) - f(v)\|_1}{\sum_{\{u,v\} \in \binom{V}{2}} \|f(u) - f(v)\|_1}.$$
Multiplying both sides by $b \log n$ establishes the statement.

4 Minimum Cuts

[Will present on the whiteboards, summary is below]

For a graph $G = (V, E)$, define

$$\phi(G, S) = \frac{|E(S, \bar{S})|}{|S||\bar{S}|},$$

and

$$\phi(G) = \min_{S \subset V, S \neq \emptyset, V} \phi(G, S),$$

where above $\bar{S} := V \setminus S$ denotes the complement of $S$, and $E(S, \bar{S})$ denotes the set of edges that have one endpoint in $S$ and one endpoint in $\bar{S}$.

Intuitively, if $\phi(G, S)$ is small, then $S$ is pretty “disconnected” from $\bar{S}$. Notice that the denominator, $|S||\bar{S}|$, is the number of edges that would be between $S$ and $\bar{S}$ in the complete graph, so $\phi(G, S)$ is the fraction of possible edges between $S$ and $\bar{S}$ that actually exist in $G$.

Finding $S$ to minimize $\phi(G, S)$ is useful, for example, in clustering applications. However, it’s also NP-hard. Today we’ll see a randomized algorithm to find an $S$ so that $\phi(G, S)$ is approximately minimized. More precisely, we’ll find $S$ so that $\phi(S, G) \leq O(\log n)\phi(G)$.

Question: How is this definition of $\phi(G)$ different/better than simply asking for the sparsest cut? (Recall we saw a randomized algorithm for the sparsest cut back in Week 1...)

4.1 Connection to metrics

**Group Work**

In this group work, you will show that

$$\phi(G) = \min_{f} \frac{\sum_{\{u,v\} \in E} \|f(u) - f(v)\|_1}{\sum_{\{u,v\} \in (V)^2} \|f(u) - f(v)\|_1},$$

where the minimum is over all functions $f : V \to \mathbb{R}^k$ for some $k$, so that $f$ takes on at least two distinct values. (This last bit is needed so that the denominator doesn’t vanish).

1. Show that

$$\phi(G) = \min_{f : V \to \{0,1\}} \frac{\sum_{\{u,v\} \in E} |f(u) - f(v)|}{\sum_{\{u,v\} \in (V)^2} |f(u) - f(v)|},$$

where the minimum is over all functions $f : V \to \{0,1\}$ so that $f$ takes on both values 0 and 1. (The difference between this and the expression above is that $f$
maps to \( \{0, 1\} \) instead of \( \mathbb{R}^k \) for some \( k \).

**Hint:** Consider mapping functions \( f \) to sets \( S \) by the relationship \( S = \{ u : f(u) = 1 \} \).

2. Think about why the above extends to show that

\[
\phi(G) = \inf_{f : V \to \mathbb{R}} \frac{\sum_{\{u,v\} \in E} |f(u) - f(v)|}{\sum_{\{u,v\} \in \binom{V}{2}} |f(u) - f(v)|},
\]

where now the minimum is over \( f : V \to \mathbb{R} \) instead of \( f : V \to \{0,1\} \).

(Don’t worry about a formal proof here, just kind of convince yourself intuitively that this is true).

**Hint:** Using part (a), it suffices to show that the infimum over all \( f : V \to \mathbb{R} \) is actually attained by some \( f \) that maps vertices in \( V \) to \( \{0,1\} \). To see this, consider the following steps:

- **Suppose that** \( f : V \to \mathbb{R} \) **takes on three distinct values,** \( a < b < c \). Consider a new function \( f_x : V \to \mathbb{R} \), so that \( f_x(u) = x \) if \( f(u) = b \), and \( f_x(u) = f(u) \) otherwise. That is, \( f_x(u) \) just replaces the value \( b \) with \( x \). Show that either \( R(f_x) \leq R(f) \) or \( R(f_c) \leq R(f) \), where

\[
R(f) = \frac{\sum_{\{u,v\} \in E} |f(u) - f(v)|}{\sum_{\{u,v\} \in \binom{V}{2}} |f(u) - f(v)|}.
\]

(That is, by sliding the middle value \( b \) towards either \( a \) or \( c \), you can decrease this quantity.)

**Sub-hint:** when you vary \( x \in [a,c] \), you can get rid of the absolute values in \( R(f_x) \). Looking at a small example might be helpful.

- **Argue that** the above logic implies that there’s an \( f \) that attains the infimum that takes on only two values.

- **Argue that** those two values may as well be 0 and 1.

3. Think about why the above extends to show that

\[
\phi(G) = \min_{f : V \to \mathbb{R}^k} \frac{\sum_{\{u,v\} \in E} \|f(u) - f(v)\|_1}{\sum_{\{u,v\} \in \binom{V}{2}} \|f(u) - f(v)\|_1},
\]

where the minimum is over all functions \( f : V \to \mathbb{R}^k \) for any \( k \).

**Hint:** You may want to use the inequality that \( \sum a_i \leq \min_i \frac{a_i}{b_i} \) for \( a_i, b_i > 0 \).
Group Work: Solutions

1. Using the connection in the hint, the numerator is exactly $|E(S, \bar{S})|$, and the denominator is the number of edges between $S$ and $\bar{S}$ in the complete graph, which is $|S||\bar{S}|$.

2. Note: this proof is a bit involved; there is an easier proof, but this one involves the least machinery and also is somewhat algorithmic, which will be useful later. I didn’t expect students to get all of the details of this proof in group work, I only wanted you to get some basic intuition. For convenience, let $$R(f) = \frac{\sum_{\{u,v\} \in E} |f(u) - f(v)|}{\sum_{\{u,v\} \in (V/2)} |f(u) - f(v)|}.$$ Notice that both the numerator and the denominator of $R(f_{b'})$ are linear in $b'$, for $b' \in [a, c]$. This is because if both $f(u), f(v) = b$, then $|f_{b'}(u) - f_{b'}(v)| = |f(u) - f(v)| = 0$; if neither are equal to $b$, then the expression does not change; and if only one is equal to $b$ (say WLOG that $f(u) = b$), then the other one is either $\leq a$ or $\geq c$ (say WLOG $\leq a$), meaning that $|f_{b'}(u) - f_{b'}(v)| = |b' - f(v)| = b' - f(v)$ is linear in $b'$.

Further, the denominator of $R(f_{b'})$ doesn’t vanish, since there’s at least one nonzero term in it (e.g., the term $|c - a|$). But then $R(f_{b'})$ is the ratio of linear functions in $b'$, and the denominator never vanishes. It’s not too hard to see (e.g., with some calculus) that $R(f_{b'})$ is thus is either increasing or decreasing (or constant), and in particular it attains a minimum at one of the endpoints $a$ or $c$ of the relevant interval.

We could have done this for any $f$ so that there are $\geq 3$ distinct values in the range. By doing this repeatedly, we see that for any $f$ with $\geq 3$ distinct values, there is some $f^*$ with only two values (say, $a$ and $b$) so that $R(f^*) \leq R(f)$. But notice that $R(f^*)$ doesn’t change if we change the values of $a$ and $b$ to 0 and 1 respectively. (That is, replace $f^*(x)$ with $\frac{f^*(x) - a}{b - a}$).

This implies that $\inf_{f:V \to \{0,1\}} R(f) \leq \inf_{f:V \to \mathbb{R}} R(f)$, and since there are only a finite number of functions $f : V \to \{0,1\}$, the infimum is actually a minimum.

3. We have shown that $\phi(G) = \min_{f:V \to \mathbb{R}} R(f)$. We clearly have

$$\phi(G) = \min_{f:V \to \mathbb{R}} R(f) \geq \min_{f:V \to \mathbb{R}^k} R(f),$$

since the set we are minimizing over on the right. On the other hand, for any
\( f : V \to \mathbb{R}^k \), we can write \( f = (f_1, \ldots, f_k) \), and so

\[
R(f) = \frac{\sum_{\{u,v\} \in E} \sum_i |f_i(u) - f_i(v)|}{\sum_{\{u,v\} \in (V_2)} \sum_i |f_i(u) - f_i(v)|}
\]

\[
= \sum_i \frac{\sum_{\{u,v\} \in E} |f_i(u) - f_i(v)|}{\sum_{\{u,v\} \in (V_2)} |f_i(u) - f_i(v)|}
\]

\[
\geq \min_i \sum_{\{u,v\} \in E} |f_i(u) - f_i(v)|
\]

\[
= \min_i R(f_i)
\]

\[
\geq \min_{g : V \to \mathbb{R}} R(g)
\]

\[
= \phi(G).
\]

Since the above reasoning held for any \( f : V \to \mathbb{R}^k \), we conclude

\[
\min_{f : V \to \mathbb{R}^k} R(f) \geq \phi(G).
\]

### 4.2 A randomized algorithm

#### Group Work

1. First, all quietly read the following: Based on the result that we got in the first group work, we might think of the following approach:

   Find \( f : V \to \mathbb{R}^k \) to minimize

   \[
   R(f) := \frac{\sum_{\{u,v\} \in E} \|f(u) - f(v)\|_1}{\sum_{\{u,v\} \in (V_2)} \|f(u) - f(v)\|_1}
   \]

   Unfortunately, this doesn’t turn out to be an easy optimization problem to solve. Instead, we’ll consider the optimization problem:

   Find values \( d_{u,v} \in \mathbb{R} \) for all \( u \neq v \in V \) to minimize

   \[
   Q(d) := \sum_{\{u,v\} \in E} d_{u,v}
   \]

   subject to:

   - \( d_{u,v} = d_{v,u} \geq 0 \) for all \( u, v \)
   - \( d_{u,v} + d_{v,w} \geq d_{u,w} \) for all \( u, v, w \)
\[
\sum_{(u,v) \in \binom{V}{2}} d_{u,v} = 1
\]

It turns out that this problem can be solved efficiently, using linear programming. (If you don’t know what that is, it’s okay, all that matters now is that we can find \( \ddot{d} \) to minimize this efficiently).

2. Suppose that \( d^* \) is the minimizer of the problem above. Explain why \( Q(d^*) \leq \phi(G) \).

3. Find a randomized algorithm to approximate \( \phi(G) \). More precisely, give a randomized algorithm that finds \( f : V \to \mathbb{R}^k \) so that, with high probability,
\[
\frac{\sum_{(u,v) \in E} \| f(u) - f(v) \|_1}{\sum_{(u,v) \in \binom{V}{2}} \| f(u) - f(v) \|_1} \leq O(\log n) \phi(G).
\]

**Hint:** Your warm-up exercise might be relevant.

**Hint:** If it comes up, you may assume that Bourgain’s embedding works just fine on pseudo-metrics, which are functions \( d(u,v) \) that obey all of the axioms of metrics except that maybe \( d(u,v) = 0 \) for \( u \neq v \).

4. Given \( f \) as in the previous part, explain how to efficiently find a set \( S \subset V \) so that \( \phi(G,S) \leq O(\log n) \phi(G) \).

**Hint:** Our proof in the first group-work was somewhat algorithmic...

---

**Group Work: Solutions**

1. Notice that because of the final constraint, and the fact that the \( \ell_1 \) norm satisfies
\[
\| c(f(u) - f(v)) \|_1 = c\| f(u) - f(v) \|_1,
\]
\[
R(f) = Q(d_f),
\]
where
\[
d_f(u,v) = \frac{\| f(u) - f(v) \|_1}{\sum_{u',v' \in \binom{V}{2}} \| f(u') - f(v') \|_1}.
\]
But \( Q(d^*) \) is the minimum over all (pseudo-)metrics (aka, distances \( d \) that satisfy \( d(u,v) = d(v,u) \geq 0 \) and also satisfy the triangle inequality), so in particular \( d_f \) is in the domain that we are minimizing over. Thus, \( Q(d^*) \leq Q(d_f) = R(f) \).
Since this holds for any \( f \),
\[
Q(d^*) \leq \min_f R(f) = \phi(G)
\]
using the previous group work.

2. Apply Bourgain’s embedding to the metric $d^*$ to get some embedding $f$. The warm-up exercise exactly implies that

$$\frac{\sum_{(u,v) \in E} \|f(u) - f(v)\|_1}{\sum_{(u,v) \in \binom{V}{2}} \|f(u) - f(v)\|_1} \leq O(\log n) Q(d^*) \leq O(\log n) \phi(G).$$

3. Given $f : V \rightarrow \mathbb{R}^k$, we saw that we can just find the coordinate $f_i$ of $f$ with the smallest $R(f_i)$ value and that will have $R(f_i) \leq R(f)$. From there, if $f$ takes on more than two values, we can “push” any intermediate value to one of its two neighbors. Repeating this leaves us with $f$ taking on only two values, and then we can renormalize $f$ to take on values that are only 0 and 1. Then we let $S \leftarrow \text{Supp}(f)$. 
