Solution for CS5275 Problem Set 1

Prepared by TA Ivan Adrian Koswara

Contents

1 Problem 1

Let $T = (V, E)$ be any n-vertex tree.

(a) Prove that $h_{\text{out}}(T) = O(1/n)$.

(b) Prove that $h_{\text{out}}(T) = \Omega(1/n)$.

1.1 My solution

Recall the definition of h_{out}:

$$h_{\text{out}}(T) = \min_{0<|S|\leq n/2} \frac{|\partial_{\text{out}}(S)|}{|S|}$$

Here, $\partial_{\text{out}}(S)$ is the set of vertices in $V \setminus S$ that are adjacent to S.

We will prove part (b) first, then part (a).

Part (b). Take any S considered for the definition of h_{out}. Clearly $|S| \leq n/2$ by definition. On the other hand, in a connected graph, $|\partial_{\text{out}}(S)| \geq 1$ since S must be connected to something else in the graph. Therefore

$$\frac{|\partial_{\text{out}}(S)|}{|S|} \geq \frac{1}{n/2} = O(1/n).$$

This is true for all S, so the claim follows.

Part (a). In the following, the size of a connected component is the number of vertices in it. If v is a vertex, let $T \setminus v$ be the graph obtained by removing v (and all edges incident to it). Since T is a tree, $T \setminus v$ is a forest.

We first make the following observation.

Observation 1.1. *Fix any vertex v, and let C be any component of $T \setminus v$. Then there exists exactly one vertex $w \in C$ such that w is adjacent to v.*

Sketch: C must be connected to the rest of the graph. We know it's connected through v, and since T is a tree, it can't be connected any other way, otherwise C is not maximal and hence not a component.

We will make use of the following lemma.

Lemma 1.2. *There exists a vertex v in T – called a **centroid** – such that, if we remove v, each remaining component has size $\leq n/2$ vertices.*

(Note: The concept of centroid appears to be very well-known. I'm finding many articles discussing about the concept and its usefulness in competitive programming problems, but I can't find the original paper that introduces centroids in the first place.)

Proof of Lemma 1.2. For each vertex u, let $f(u)$ be the size of the largest component in $T \setminus u$. Let v be a vertex that minimizes f. We claim this v is what we're looking for.

Suppose otherwise, so $f(v) > n/2$. Note that only one component can have $> n/2$ vertices, let this component be C. By Observation 1.1, there exists a unique vertex $w \in C$ that is adjacent to v. We claim $f(w) < f(v)$, contradicting the minimality of v.

Indeed, look at $T \setminus w$. Now $V \setminus C$ forms one component. But since $|C| > n/2$, it follows $|V \setminus C| < n/2 < |C|$. Any other component is a proper subset of C, because it does not contain w, and so has size $< |C|$. This proves $f(w) < f(v)$.

Therefore, since such w cannot exist, the assumption $f(v) > n/2$ was wrong. So $f(v) \leq n/2$, proving the lemma. □

We will now prove the statement in the problem. It suffices to find one S such that $|\partial_{\text{out}}(S)| = 1$ and $n/4 \le |S| \le n/2$. Because, if we do so, then

$$h_{\text{out}}(T) \le \frac{|\partial_{\text{out}}(S)|}{|S|} \le \frac{4}{n} = O(1/n)$$

proving the claim.

To find this S, first find a centroid v as described in Lemma 1.2. Each component of $T \setminus v$ has size $\le n/2$. Moreover, if S is a union of some (≥ 1) components in $T \setminus v$, then $\partial_{\text{out}}(S) = \{v\}$. It suffices to find S with the desired cardinality.

If there exists a component of size $\ge n/4$, choose one such component as our S.

Otherwise, every component has size $< n/4$. Put in components into S in an arbitrary order, and stop once $|S| \ge n/4$. At this point, $|S| < n/2$, since each component added has size $< n/4$.

In both cases, we obtain our desired S.

1.2 Insight and discussion

Part (b). This problem is silly. I think, as long as you understand h_{out} (and its associated ∂_{out}), you should be able to arrive at this solution. It should be immediately obvious that $|S| = O(n)$, and it takes a bit more thought to realize $|\partial_{\text{out}}(S)| \ge 1$ but it shouldn't be difficult.

Part (a). After working out the above, you realize: to get $h_{\text{out}}(T) = O(1/n)$, you need some S with $|\partial_{\text{out}}(S)| = O(1)$ and $|S| = \Omega(n)$. A tree is barely connected, so it should be easy to find S with $|\partial_{\text{out}}(S)| = 1$; in other words, there is a bottleneck vertex v.

The problem is finding S of the appropriate size, because if v is something close to a leaf (or probably a leaf itself), the components of $T \setminus v$ are either tiny (and don't add up to $\Omega(n)$) or huge (and so $> n/2$ and cannot be considered for h_{out}). Once you want v to be roughly in the "middle" of the tree, you'll likely come up with the concept of centroid independently, if you didn't already know about it.

One way to come up with the centroid is by trying a "greedy algorithm". If we pick some v and there's still a large component, what happens if we move over into the large component? The large component will be split up. The other components will merge together, but if our large component is large enough, even the merged component will still be small. Formalizing this algorithm gives Lemma 1.2; the proof can either use infinite descent (basically the greedy algorithm), or a minimality argument (like in my official solution), or some other equivalent thing.

You don't have to come to the exact concept of centroids; you just need every component to have size $\le cn$ for some constant $c < 1$. But I believe you'll need to do a similar work either way to obtain any c, and while at that, might as well take $c = 1/2$, the best possible constant.

Either way, once you come up with said v, the rest of the solution is simply to make use of it. Every component is missing at least a constant proportion, so you won't excessively overshoot from $|S| = o(n)$ into $|S| = n - o(n)$, you'll always find some S in between. And that's enough.

Note that the idea of picking the S is very similar to "robustness against edge deletions" in lecture 1 slides 13–18. Also worth noting that the concept of centroids is also applicable for weighted trees, although the concept of ∂_{out} is probably only applicable for unweighted trees.

1.3 Grading criteria and common mistakes

1.3.1 Fatal error: Assuming a single T is enough

Unlike Problem 4, here the problem statement doesn't say you only need to prove the inequalities for a specific T, or even a specific T for each size n. You need to prove the inequalities for all (large enough) T.

I can see the problem doesn't exactly specify "for all T", so if you only give a specific T, I'm giving the benefit of the doubt and give a little bit of credit (0.5/2). But since the work doesn't do anything to the problem, I can't justify giving more.

1.3.2 Passable error: Treating T as a rooted tree

Quite a lot of the solutions treat T as a rooted tree, by starting with "let r be **the** root" (instead of "let r be an arbitrary vertex, root T at r"). I'm guessing this comes from a competitive programming background, where "tree" instinctively means rooted tree and all. But this is a mathematical course, tree just has the usual graph theoretical definition.

It usually doesn't change anything about the proof, it's just pretty funny to see.

1.3.3 (a) Fatal error: Wrong direction

To prove $h_{\text{out}}(T) = O(1/n)$, you need $|S| = \Omega(n)$. If you only say $|S| = O(n)$, that doesn't help at all; it gives the solution for part (b) instead. I can't give credit for this.

1.3.4 (a) Fatal error: Taking any $|S| = n/2$

Some of you simply take any connected S where $|S| = n/2$. This does not guarantee $T \setminus S$ is connected, because S might be "in the middle" of the tree. For that same reason, $|\partial_{\text{out}}(S)|$ also isn't bounded by a constant. I can't give credit for this.

1.3.5 (a) Passable error: Not citing centroid

To be honest, I've never heard of the concept of tree centroids before this problem (although I would be able to come up with the proof if I was told the definition). So if you call it a centroid, but don't give a proof of its existence, I'm wondering if a centroid will always exist.

That said, because searching online gives so many resources about tree centroids, I think I chalk it up to me being not familiar with the concept. Although, it's still a good idea to cite somewhere about where you read the concept of a centroid. (Even one of those competitive programming resources would be fine, you don't need a full academic paper.) In either case, I'm not deducting points for this.

1.3.6 (a) Passable error: Some imprecision with $n/3 \le |S| \le 2n/3$

When finding S of the appropriate size, some people use the approach in lecture 1 to come up with S such that $n/3 \le |S| \le 2n/3$. If $|S| > n/2$, then you need to invert S so that it can go into the definition of h_{out}.

The tricky thing is, if you invert S into $S' := V \setminus S$, you still need to take out v from S', so that you can have $\partial_{\text{out}}(S') = \{v\}$. Some of you didn't do so, which makes $|\partial_{\text{out}}(S')| \ge 1$, missing the bound on the numerator. And if you do remove v from S', then you can only guarantee $|S'| \ge n/3 - 1$.

To be fully correct, you'll need to adjust your proof slightly to handle these cases correctly. But they are extremely minor errors with easy fixes, so I don't deduct points for them.

2 Problem 2

Let $G = (V, E)$, and suppose it has a subgraph $G' = (V', E')$ such that

$$|E'| \geq (1 - \varepsilon) \cdot |E| \qquad \text{and} \qquad \Phi(G') \geq \phi > 0.$$

Consider the following procedure:

1. Initialize $W \leftarrow V$.
2. As long as there's some $S \subseteq W$ with $\mathrm{vol}_{G[W]}(S) \leq \mathrm{vol}_{G[W]}(W)/2$ such that $\Phi_{G[W]}(S) \leq \phi/2$:
   - Remove S from W, i.e. set $W \leftarrow W \setminus S$.
3. Return W.

Prove that there exist a threshold $\lambda > 0$ and a constant $c > 0$ – independent of G, n, ϕ – such that for each $\varepsilon \in (0, \lambda)$, the output W_{final} satisfies

$$\mathrm{vol}_G(W_{\text{final}}) \geq (1 - c \cdot \varepsilon) \cdot \mathrm{vol}_G(V).$$

(Note: You may assume $\mathrm{vol}_G(W_{\text{final}}) \geq \frac{1}{2} \cdot \mathrm{vol}_G(V)$. There is a solution without this assumption, but it's significantly more complicated.)

2.1 My solution

Note: This solution will prove the problem statement using the additional assumption. See Appendix 2.4 for how to handle the problem without the assumption.

Consider the procedure. Let m be the number of iterations it takes in step 2. On the k-th iteration, let the current W be W_k, and the trimmed set be S_k. Additionally, let $S = S_m$. Note that $S = V \setminus W_{\text{final}}$. The additional assumption means $\mathrm{vol}_G(S) \leq \mathrm{vol}_G(V)/2$, and we want to find a constant c such that $\mathrm{vol}_G(S) \leq c \cdot \varepsilon \cdot \mathrm{vol}_G(V)$ for all small enough ε.

We will put an upper bound and a lower bound on $|E(S, V \setminus S)|$. Then the upper bound is $\geq$ the lower bound, and from there, we will obtain our desired c.

Upper bound for $|E(S, V \setminus S)|$. To obtain an upper bound, consider the procedure. At the i-th step, we obtain S_i satisfying

$$\Phi_{G[W_i]}(S_i) = \frac{|E(S_i, W_i \setminus S_i)|}{\mathrm{vol}_{G[W_i]}(S_i)} \leq \frac{\phi}{2} \qquad \Longrightarrow \qquad |E(S_i, W_i \setminus S_i)| \leq \frac{\phi}{2} \cdot \mathrm{vol}_{G[W_i]}(S_i)$$

Now consider all S_i and add up the inequalities together. We obtain

$$\sum_{i=1}^{m} |E(S_i, W_i \setminus S_i)| \leq \frac{\phi}{2} \cdot \sum_{i=1}^{m} \mathrm{vol}_{G[W_i]}(S_i) \tag{1}$$

We will now weaken both sides by proving these two inequalities:

$$\text{(i)} \ \sum_{i=1}^{m} |E(S_i, W_i \setminus S_i)| \geq |E(S, V \setminus S)| \qquad \text{and} \qquad \text{(ii)} \ \sum_{i=1}^{m} \mathrm{vol}_{G[W_i]}(S_i) \leq \mathrm{vol}_G(S) \tag{2}$$

Inequality (i). Let uv be an edge where $u \in S$ and $v \notin S$. Then $u \in S_k$ for some k. Moreover, $v \in W_i$ for all i, in particular $v \in W_k$, and $v \notin S_k$. Therefore uv is in $E(S_k, W_k \setminus S_k)$. No other term in the summation counts uv: for $i < k$, neither u, v is in S_i; for $i > k$, u is not in either set. Therefore every edge in $E(S, V \setminus S)$ is counted exactly once.

Inequality (ii). First, $\mathrm{vol}_{G[W_i]}(S_i) \leq \mathrm{vol}_G(S_i)$, because $G[W_i]$ only removes edges from G (and doesn't add any). Second, all S_i's are disjoint, so the volume can simply be added together.

Combining (1) and the two inequalities in (2) gives

$$|E(S, V \setminus S)| \leq \frac{\phi}{2} \cdot \mathrm{vol}_G(S) \tag{3}$$

Lower bound for $E(S, V \setminus S)|$. First, let $S' = S \cap V'$ be the vertices of S that stay in G', and let $E^\times = E \setminus E'$ be the edges that are removed from G'. Note that $|E^\times| \leq \varepsilon \cdot |E|$ by definition of E'.

To get a lower bound, consider $E(S', V' \setminus S')$. Clearly

$$E(S, V \setminus S) \supseteq E(S', V' \setminus S') \qquad \Longrightarrow \qquad |E(S, V \setminus S)| \geq |E(S', V' \setminus S')|$$

Therefore it is enough to give a lower bound for $|E(S', V' \setminus S')|$.

Consider the conductance on G'. By definition of conductance, we have

$$\phi \leq \Phi(G') \leq \frac{|E(S', V' \setminus S')|}{\min\{\mathrm{vol}_{G'}(S'), \mathrm{vol}_{G'}(V' \setminus S')\}} \tag{4}$$

Now, note that

$$\mathrm{vol}_{G'}(S') \geq \mathrm{vol}_G(S) - 2 \cdot |E^\times| \geq \mathrm{vol}_G(S) - \varepsilon \cdot \mathrm{vol}_G(V) \tag{5}$$

The formalization is a bit messy, but the idea is, $\mathrm{vol}_{G'}(S')$ is exactly $\mathrm{vol}_G(S)$ without the edges removed in $E^\times$. But each edge contributes ≤ 2 into any volume. The second inequality comes from $|E^\times| \leq \varepsilon \cdot |E|$ and $\mathrm{vol}_G(V) = 2 \cdot |E|$.

The argument for (5) can also be applied to $V' \setminus S'$ in an identical manner:

$$\mathrm{vol}_{G'}(V' \setminus S') \geq \mathrm{vol}_G(V \setminus S) - \varepsilon \cdot \mathrm{vol}_G(V) \tag{6}$$

Therefore, (5) and (6) give

$$\min\{\mathrm{vol}_{G'}(S'), \mathrm{vol}_{G'}(V' \setminus S')\} \geq \min\{\mathrm{vol}_G(S), \mathrm{vol}_G(V \setminus S)\} - \varepsilon \cdot \mathrm{vol}_G(V) \tag{7}$$

By using the additional assumption, we know $\min\{\mathrm{vol}_G(S), \mathrm{vol}_G(V \setminus S)\} = \mathrm{vol}_G(S)$.

Now, combining (4) (after rearranging) and (7) gives

$$|E(S', V' \setminus S')| \geq \phi \cdot (\mathrm{vol}_G(S) - \varepsilon \cdot \mathrm{vol}_G(V)) \tag{8}$$

Combining upper and lower bounds. Finally, combining (3) and (8) gives

$$\phi \cdot (\mathrm{vol}_G(S) - \varepsilon \cdot \mathrm{vol}_G(V)) \leq |E(S', V' \setminus S')| \leq \frac{\phi}{2} \cdot \mathrm{vol}_G(S) \tag{9}$$

Taking only the two bounds and rearranging:

$$\begin{aligned} \phi \cdot (\mathrm{vol}_G(S) - \varepsilon \cdot \mathrm{vol}_G(V)) &\leq \frac{\phi}{2} \cdot \mathrm{vol}_G(S) \\ 2 \cdot (\mathrm{vol}_G(S) - \varepsilon \cdot \mathrm{vol}_G(V)) &\leq \mathrm{vol}_G(S) \\ \mathrm{vol}_G(S) &\leq 2\varepsilon \cdot \mathrm{vol}_G(V) \end{aligned}$$

Therefore $c = 2$ suffices.

2.2 Insight and discussion

Although the solution is quite messy, I think the idea is fairly straightforward.

First, you want to think about bounding $|E(S, V \setminus S)|$. This might come naturally along with the upper bound. The procedure removes low-conductance cuts, so the number of cut edges are small in comparison to the volume removed. Intuitively, that means the total number of cut edges over the entire procedure will also be small. This gives the upper bound in (3).

The lower bound is a bit more difficult, or perhaps unintuitive. G' has high conductance, so it is useful to restrict S into G' and see if we can get a bound. The conductance gives a lower bound for $|E(S', V' \setminus S')|$, but this lower bound is in $\min\{\mathrm{vol}_{G[V']}(S'), \mathrm{vol}_{G[V']}(V' \setminus S')\}$. It takes some algebraic trickery to manipulate this into a nicer form like in (5) into (7).

But once you get through the calculations, you end up with two very nice bounds on $|E(S, V \setminus S)|$. They are in $\mathrm{vol}_G(S)$, but the constants are the other way around: the constant $\phi \cdot \mathrm{vol}_G(S)$ for the lower bound is larger than the constant $\frac{\phi}{2} \cdot \mathrm{vol}_G(S)$ for the upper bound. This lets you bound $\mathrm{vol}_G(S)$ very cleanly.

Well, that's all nice and good... if you take the additional assumption for granted. Before this assumption was added, you would get a stumbling block: in (7), you could not say S is the smaller side, and you would need to handle the case if $V \setminus S$ is the smaller side. When I and Prof worked on this case, we realized it became much messier. See for yourself in Appendix 2.4.

We decided this distracts from the general idea we want, that you analyze an upper bound and a lower bound for $|E(S, V \setminus S)|$ based on the two forms of conductance given in the problem. That's why we added the additional assumption for you to use freely.

2.3 Grading criteria and common mistakes

In general, I'm pretty lenient at grading this specific problem. I excuse minor errors more than usual, especially if it's just small mathematical errors. I'm just looking that you make use of the two forms of conductance.

2.3.1 Not an error: Not handling the case $\mathrm{vol}(W_{\text{final}}) < \mathrm{vol}(V)/2$

As mentioned in the note for the problem, this is not an error. You will be given full credit even if you don't cover this case; and if you attempt to cover this case, you will still be given full credit if your solution for the main case (with the additional assumption) is correct, even if this case is wrong.

To my knowledge, there's only one student who tried to handle this case. (And they got it more or less correct.)

2.3.2 Passable error: Concluding S' is the smaller side than $V' \setminus S'$

In (4), the numerator is $\min\{\mathrm{vol}_{G'}(S'), \mathrm{vol}_{G'}(V' \setminus S')\}$. Although you have the additional assumption that S is the smaller side than $V \setminus S$, it doesn't follow that S' is the smaller side than $V' \setminus S'$. It's possible that G into G' loses exclusively stuff from $V \setminus S$, enough to turn $V' \setminus S'$ to be the smaller side.

This is a minor error, because $\mathrm{vol}_{G'}(S')$ can be written as $\mathrm{vol}_G(S)$ with a small piece taken out (like in (5)), and similarly with $\mathrm{vol}_{G'}(V' \setminus S')$. *Then* you can use the additional assumption. (Also, we find out S' is tiny anyway, so we can indeed assume S' is the smaller side, but that's a *post hoc* justification after the fact.)

I generally allow this error, since it's minor enough and doesn't significantly damage the proof.

2.3.3 Passable error: Using G instead of $G[W_i]$

When you take the conductance of the cut S_i, it is on $G[W_i]$. Therefore, the correct conductance is

$$\Phi_{G[W_i]}(S_i) = \frac{|E_{G[W_i]}(S_i, W_i \setminus S_i)|}{\mathrm{vol}_{G[W_i]}(S_i)} \leq \frac{\phi}{2}$$

There are three pretty common errors here.

- In the numerator, using edges in G instead of $G[W_i]$. By definition of Φ, this is an error, but it can be easily shown there's no additional or missing edge. So it's correct to simply say $|E(S_i, W_i \setminus S_i)|$ too.

- In the denominator, using G instead of $G[W_i]$ for the volume. Compared to the previous one, this is an actual error, because $\mathrm{vol}_G(S_i)$ is different from $\mathrm{vol}_{G[W_i]}(S_i)$. The good thing is, either one works, because $\mathrm{vol}_G(S_i) \geq \mathrm{vol}_{G[W_i]}(S_i)$. So the conductance is already $\leq \phi/2$ with $G[W_i]$; replacing the volume with G instead won't help making it larger.

- In the numerator, using $V \setminus S$ instead of $W_i \setminus S$. This one is an actual logical error, because it breaks the proof. $|E(S_i, V \setminus S_i)| \geq |E(S_i, W_i \setminus S_i)|$ in general, so the $\leq \phi/2$ bound no longer holds. But as I mentioned, I'm pretty lenient in grading this problem. So unless this descends into bizarre claims afterward, I'll just assume you made a typo here.

2.4 Appendix: Handling the additional assumption

It is possible to drop the additional assumption, but it requires a major rework on the solution.

First, let's investigate what the additional assumption really means in the first place. It says S is made of small cuts, so that even the union S is still small. Is it possible to have a large S, large enough to the point that $V \setminus S$ becomes the smaller side?

Yes. Even though $\mathrm{vol}_{G[W_i]}(S_i) \leq \mathrm{vol}_{G[W_i]}(W_i)/2$ individually, it's possible that some S_k has volume nearly half of its current W_k, so S jumps from being tiny to be around half of $\mathrm{vol}_G(V)$. This can be enough to tip it over to make $V \setminus S$ the smaller side. This is what the additional assumption prevents, and what we're trying to work around in this appendix.

Note that the additional assumption is *only* used in (7), when we take $\mathrm{vol}_G(S)$ as the minimum rather than $\mathrm{vol}_G(V \setminus S)$. But this is an important one, so that we can combine the two bounds in (9) to get an appropriate bound. If the minimum had been $\mathrm{vol}_G(V \setminus S) = \mathrm{vol}_G(V) - \mathrm{vol}_G(S)$, we instead end up with the following bizarre bound:

$$\mathrm{vol}_G(S) \geq \frac{2(1-\varepsilon)}{3} \cdot \mathrm{vol}_G(V) \tag{10}$$

This bound is interesting, because it says S must be *large* enough, specifically around $2/3$ of $\mathrm{vol}_G(V)$ when ε is small. But as we discussed earlier, intuitively S will jump from tiny to be around $1/2$ of $\mathrm{vol}_G(V)$. This suggests that this case cannot happen for small ε.

To formalize the claim, let's first slightly generalize the problem statement. Let $S_{\leq i} = S_1 \cup S_2 \cup \ldots \cup S_i$; note that $S_{\leq m} = S$. Instead of terminating the procedure when there is no more cut to be made, we allow the procedure to be terminated at any iteration. In other words, we claim the problem statement holds for any $S_{\leq i}$, instead of just for $S_{\leq m}$. Nowhere in our previous analysis did we rely on the terminating condition, so the revised problem statement is still true given the additional assumption.

Now suppose that there exists k such that $\mathrm{vol}_G(S_{\leq k}) > \mathrm{vol}_G(V)/2$, and let k be the smallest such index; in particular, $\mathrm{vol}_G(S_{\leq k-1}) \leq \mathrm{vol}_G(V)/2$. This is the step where S flips from being the smaller side to being the larger side. We analyze the equation

$$\mathrm{vol}_G(S_{\leq k}) = \mathrm{vol}_G(S_{\leq k-1}) + \mathrm{vol}_G(S_k). \tag{11}$$

First, we have

$$\mathrm{vol}_G(S_{\le k-1}) \le 2\varepsilon \cdot \mathrm{vol}_G(V). \tag{12}$$

This is just because the case up to $S_{\le k-1}$ is exactly what we have already proven; we do have the assumption $\mathrm{vol}_G(S_{\le k-1}) \le \mathrm{vol}_G(V)/2$ after all.

To bound $\mathrm{vol}_G(S_k)$, we first have the following inequality from the definition of S_k:

$$\mathrm{vol}_{G[W_k]}(S_k) \le \frac{\mathrm{vol}_{G[W_k]}(W_k)}{2} \le \frac{\mathrm{vol}_G(V)}{2} \tag{13}$$

Now we follow a similar idea as (5). Let $E^\times$ be the set of edges removed from G to $G[W_k]$. Note that

$$|E^\times| \le \mathrm{vol}_G(S_{\le k-1}) \le 2\varepsilon \cdot \mathrm{vol}_G(V) \tag{14}$$

The left inequality is because every edge in $E^\times$ is incident to something in $S_{\le k-1}$ – that's why the edge got removed – so each contributes at least 1 to the volume. The right inequality is from (12).

Then, using the exact same reasoning as in (5):

$$\mathrm{vol}_G(S_k) \le \mathrm{vol}_{G[W_k]}(S_k) + 2 \cdot |E^\times| \tag{15}$$

Now, take (15) and use the bounds from (13) and (14):

$$\mathrm{vol}_G(S_k) \le \frac{\mathrm{vol}_G(V)}{2} + 2 \cdot (2\varepsilon \cdot \mathrm{vol}_G(V)) = \left(\frac{1}{2} + 4\varepsilon\right) \cdot \mathrm{vol}_G(V) \tag{16}$$

Finally, adding up (12) and (16):

$$\begin{aligned} \mathrm{vol}_G(S_{\le k}) &= \mathrm{vol}_G(S_{\le k-1}) + \mathrm{vol}_G(S_k) \\ &\le 2\varepsilon \cdot \mathrm{vol}_G(V) + \left(\frac{1}{2} + 4\varepsilon\right) \cdot \mathrm{vol}_G(V) \\ &= \left(\frac{1}{2} + 6\varepsilon\right) \cdot \mathrm{vol}_G(V) \end{aligned} \tag{17}$$

Therefore, we now have an upper bound on $\mathrm{vol}_G(S_{\le k})$ in (17). We also have a lower bound for it: the bound in (10), obtained from our initial analysis except using $\mathrm{vol}_G(V \setminus S)$ as the minimum, also applies for $\mathrm{vol}_G(S_{\le k})$. (Here we simply stop the procedure at $S_{\le k}$.)

Combining (10) and (17):

$$\frac{2(1-\varepsilon)}{3} \cdot \mathrm{vol}_G(V) \le \mathrm{vol}_G(S_{\le k}) \le \left(\frac{1}{2} + 6\varepsilon\right) \cdot \mathrm{vol}_G(V).$$

Taking only the two bounds and rearranging:

$$\begin{aligned} \frac{2(1-\varepsilon)}{3} \cdot \mathrm{vol}_G(V) &\le \left(\frac{1}{2} + 6\varepsilon\right) \cdot \mathrm{vol}_G(V) \\ 2 - 2\varepsilon &\le \frac{3}{2} + 18\varepsilon \\ \frac{1}{40} &\le \varepsilon \end{aligned}$$

Therefore, $V \setminus S$ can only be the smaller side if $\varepsilon \ge 1/40$. So by taking $\lambda = 1/40$, we only focus our attention on when $\varepsilon < \lambda = 1/40$, when this case *cannot* happen. When $\varepsilon < 1/40$, we can prove the additional assumption (that S is the smaller side) always holds, so our original proof works to get $c = 2$.

(Alternatively, to avoid using λ, we weaken into $c = 40$ so that $c \cdot \varepsilon \ge 1$ when $\varepsilon \ge 1/40$, giving the trivial $\mathrm{vol}_G(S) \le \mathrm{vol}_G(V)$.)

(Note: It's likely possible to obtain a better (i.e. larger) λ with more careful analysis. But it's enough to show some λ exists.)

3 Problem 3

Let $G = (V, E)$ be an r-regular graph on n vertices, where $r = O(1)$ and $\Phi(G) = \Omega(1)$.

(a) Prove there are $\Omega(n^2)$ pairs of vertices with distance $\Omega(\log n)$.

(b) Let H be the clique K_n on the same vertex set V. Let $\mathrm{LR}(G, H)$ be the optimum value of the Leighton-Rao LP relaxation for this instance. Prove that

$$\mathrm{LR}(G, H) = O\left(\frac{\Phi(G, H)}{\log n}\right).$$

3.1 My solution

Because G is r-regular, it follows $\mathrm{vol}_G(S) = r|S|$. So vol_G and size (cardinality) are interchangeable.

Part (a). Fix a vertex v. For a natural number k, let $B(v, k)$ be the (closed) ball centered at v with radius k, i.e. the set of all vertices of distance $\leq k$ from v.

We have the following observation:

Observation 3.1. $|B(v, k+1)| \leq (r+1) \cdot |B(v, k)|$, *therefore* $|B(v, k)| \leq (r+1)^k$.

Sketch: Because the graph is r-regular, each vertex in $B(v, k)$ is adjacent to r vertices. So each $u \in B(v, k)$ supports $\leq r$ new vertices at distance $k+1$. Counting u itself, each $u \in B(v, k)$ supports $\leq r+1$ vertices in $B(v, k+1)$. The claim follows.

Now let $m = \lfloor \log_{r+1}(n/2) \rfloor$. This is $\Omega(\log n)$ because $r = O(1)$. By Observation 3.1, $|B(v, m)| \leq n/2$. Therefore there are $\geq n/2$ vertices of distance $> m$ from v.

Thus v contributes $\Omega(n)$ pairs of vertices with distance $\Omega(\log n)$. Summing over all n choices of v, we get the problem statement.

Part (b). Recall the definitions of $\Phi(G, H)$ and $\mathrm{LR}(G, H)$ (lecture 2 slide 29). Minimize

$$\frac{\sum_{u,v} w_G(u, v) \cdot d(u, v)}{\sum_{u,v} w_H(u, v) \cdot d(u, v)}$$

over all metric d satisfying: (i) d is a cut metric (for $\Phi(G, H)$), and (ii) d is any metric (for $\mathrm{LR}(G, H)$).

For simplicity, we will write

$$\Sigma(G, d) = \sum_{u,v} w_G(u, v) \cdot d(u, v)$$

and similarly for H. Therefore we want to minimize $\Sigma(G, d)/\Sigma(H, d)$ for both values $\Phi(G, H)$ and $\mathrm{LR}(G, H)$.

Furthermore, since G is unweighted, we have

$$w_G(u, v) = \begin{cases} 1 & \text{if } u, v \text{ are adjacent} \\ 0 & \text{if } u, v \text{ are not adjacent} \end{cases}$$

Therefore, $\Sigma(G, d)$ can be rewritten as

$$\Sigma(G, d) = \sum_{u,\, v \text{ adjacent}} d(u, v).$$

The same is true for H. Moreover, since $H = K_n$ in this case, we can simplify it even further: u, v are adjacent iff $u \neq v$.

We first give a lower bound for $\Phi(G, H)$. Then we construct d for $\mathrm{LR}(G, H)$, so we know $\mathrm{LR}(G, H) \leq$ the value for this d.

Lower bound for $\Phi(G, H)$. Let $S \subseteq V$ where $|S| = s$, and suppose $1 \leq s \leq n/2$. This gives rise to a cut metric d_S. All nontrivial cut metrics (i.e. not always zero) can be written in this way.

Now, $|E_H(S, V \setminus S)| = s(n-s)$, so $\Sigma(H, d_S) = 2s(n-s)$: each edge (u, v) in $E_H(S, V \setminus S)$ counts twice, as $d_S(u, v) = 1$ and as $d_S(v, u) = 1$.

We now consider the conductance $\Phi(S)$ of the set S. On one hand, $\Phi(S) \geq \Phi(G)$ by definition of conductance $\Phi(G)$ of the whole graph. On the other hand, recall $\mathrm{vol}_G(S) = rs$. Since $s \leq n/2$, S is the smaller side of the cut. Therefore

$$\begin{aligned}\Phi(G) &\leq \Phi(S) \\ &= \frac{|E_G(S, V \setminus S)|}{\min\{\mathrm{vol}_G(S), \mathrm{vol}_G(V \setminus S)\}} \\ &= \frac{|E_G(S, V \setminus S)|}{rs}\end{aligned}$$

Therefore, $|E_G(S, V \setminus S)| \geq rs \cdot \Phi(G)$.

Finally, consider $\Sigma(G, d_S)$. Note that $d_S(u, v) = 1$ if u, v is split by the set S, and 0 otherwise. Therefore $\Sigma(G, d)$ counts exactly the number of pairs u, v such that u, v are adjacent, and also split by S. This is exactly the definition of $|E_G(S, V \setminus S)|$.

Therefore

$$\frac{\Sigma(G, d_S)}{\Sigma(H, d_S)} = \frac{|E_G(S, V \setminus S)|}{2s(n-s)} \geq \frac{rs \cdot \Phi(G)}{2s(n-s)} = \frac{r \cdot \Phi(G)}{2(n-s)}$$

Finally, note that $r = \Omega(1)$ because it is a positive integer, $\Phi(G) = \Omega(1)$ is given, and $n - s = O(n)$. Therefore

$$\frac{\Sigma(G, d_S)}{\Sigma(H, d_S)} = \frac{r \cdot \Phi(G)}{2(n-s)} = \frac{\Omega(1)}{O(n)} = \Omega\left(\frac{1}{n}\right)$$

Since $\Phi(G, H)$ takes the minimum over all d_S, and all d_S satisfy the above, it follows $\Phi(G, H) = \Omega(1/n)$ too.

Upper bound for $\mathbf{LR}(G, H)$. Let $d = \mathrm{dist}_G$ be the shortest distance metric on G. We will compute $\Sigma(G, \mathrm{dist}_G)$ and $\Sigma(H, \mathrm{dist}_G)$.

$\Sigma(G, \mathrm{dist}_G)$ is easy to compute. Note that the summation is over pairs of adjacent u, v. But in this case, $\mathrm{dist}_G(u, v) = 1$. For each vertex u, it is adjacent to exactly r vertices, so in total $\Sigma(G, \mathrm{dist}_G)$ has nr terms, all of them 1. So $\Sigma(G, \mathrm{dist}_G) = nr$.

We now compute $\Sigma(H, \mathrm{dist}_G)$. By part (a), there are $\Omega(n^2)$ pairs of vertices with distance $\Omega(\log n)$. Therefore, for these pairs (u, v), we have $\mathrm{dist}_G(u, v) = \Omega(\log n)$. By definition of $\Sigma(H, d)$, we have

$$\Sigma(H, d) = \sum_{u \neq v} d(u, v) \geq \Omega(n^2) \cdot \Omega(\log n) = \Omega(n^2 \log n)$$

Therefore

$$\frac{\Sigma(G, d_S)}{\Sigma(H, d_S)} = \frac{nr}{\Omega(n^2 \log n)} = O\left(\frac{1}{n \log n}\right)$$

Since $\mathrm{LR}(G, H)$ takes the minimum over all d, which includes $d = \mathrm{dist}_G$, it follows $\mathrm{LR}(G, H) = O(1/n \log n)$.

Therefore, $\Phi(G, H) = \Omega(1/n)$ and $\mathrm{LR}(G, H) = O(1/n \log n)$. This proves the problem statement.

3.2 Insight and discussion

Part (a). The core idea is the ball-growing argument in lecture 1 slides 25–28, but we use it in the opposite direction. Basically, expanders make you visit a lot of vertices quickly, while low-degree graphs make you visit few vertices quickly. Once you get the idea, the calculation is straightforward and can be done in many ways.

Part (b). This problem is quite tricky, but it's made of a number of steps that individually make sense. The most tricky part is finding d for $\mathrm{LR}(G, H)$, but you're told what this d is with the hint. (I had trouble finding it without any hint, although on retrospect I think there's motivation to think that way; see below.) I think the main difficulty is the sea of symbols and definitions; this is your first practice to make use of $\Phi(G, H)$ and $\mathrm{LR}(G, H)$, so just take it slowly.

First, the definitions of $\Phi(G, H)$ and $\mathrm{LR}(G, H)$ are extremely scary with all the symbols. When you think about it, though, you realize: on an unweighted graph, $w_G(u, v)$ is just "1 if u, v are adjacent, 0 otherwise". So the summation is "sum of $d(u, v)$ over all pairs of adjacent vertices u, v". If we define $\Sigma(G, d)$ as sum of $d(u, v)$ over pairs of u, v adjacent in G, then for both Φ and LR, we want to minimize $\Sigma(G, d)/\Sigma(H, d)$.

First consider $\Phi(G, H)$. It considers only cut metrics d_S, which only takes values 1 and 0. So $\Sigma(G, d_S)$ is the count of adjacent u, v where $d_S(u, v) = 1$. Thanks to the definition of cut metric, this is exactly $|E_G(S, V \setminus S)|$. Now, this number also comes in the definition of $\Phi(G)$. So we make use of $\Phi(G) = \Omega(1)$ to give a lower bound for $|E_G(S, V \setminus S)|$, therefore a lower bound for $\Phi(G, H)$.

Now consider $\mathrm{LR}(G, H)$. Finding some d that works without hints is pretty difficult, but there's some motivation about it: you want $\Sigma(G, d)$ to be small, but $\Sigma(H, d)$ to be large. For instance, say $d(u, v) = O(1)$ for all adjacent pairs u, v, but there are a ton of large values of d that get counted for H but not for G. This may lead you to the shortest distance metric $d = \mathrm{dist}_G$, since $\mathrm{dist}_G(u, v) = 1$ for all adjacent u, v, but you can have large $\mathrm{dist}_G(u, v)$ elsewhere. There may be other d that works, but this might be the simplest one to think of.

(You can also arrive at this with some "meta"-motivation. Part (a) proved something about many pairs of far vertices, and since it's grouped with this problem, you can try to make use of that. Or, you know, just use the hint.)

In any case, once you think of $d = \mathrm{dist}_G$, it becomes fairly straightforward, using exactly the same train of thought described above.

3.3 Grading criteria and common mistakes

3.3.1 (a) Major error: Using Φ instead of r

Some of you use the ball-growing argument, but exactly as-is as given in the lecture, using the conductance Φ instead of the degree r. This gives the wrong direction, you end up with a lower bound on $|B(v, k)|$. And without any upper bound, you cannot claim $k = \Omega(\log n)$. For example, in a complete graph, $|B(v, 1)| = n$ already, so the smallest k such that $|B(v, k)| = \Omega(n)$ is $k = 1$, which is not $\Omega(\log n)$.

In general, this is a very damaging error since it does not lead to anything. However, I usually give some credit because you start by using balls $B(v, k)$, and the general idea of bounding $|B(v, k)|$ to get a bound on k is correct. Just the execution is very wrong.

3.3.2 (a) Passable error: $|B(v, k)| \le r^k$

When bounding $|B(v, k)|$, there are a few minor errors, such as saying $|B(v, k)| \le r^k$. This is incorrect for $k = 1$ (you have $|B(v, 1)| = r + 1$), and while correct for larger k, it is a bit more tricky to argue than the simpler $|B(v, k)| \le (r + 1)^k$. But it's such a tiny error that I don't usually comment anything about it, much less deduct any points.

3.3.3 (b) Minor error: Wrong direction for $\Phi(G, H)$

You know $\mathrm{LR}(G, H) = O(1/n \log n)$, so in order to show it is $O(\Phi(G, H)/\log n)$, you need to show $\Phi(G, H) = \Omega(1/n)$. Not $O(1/n)$. If you have $\Phi(G, H) = O(1/n)$, then both $\mathrm{LR}(G, H)$ and $\Phi(G, H)$ have upper bounds, and so they cannot be compared to each other.

This is a pretty important logical error, so I cannot give full credit to it. But it's a reasonably easy fix, so I only deduct minimal amount of points.

3.3.4 (b) Major error: Not computing $\Phi(G, H)$ correctly

There are a few errors when computing $\Phi(G, H)$. To get a proper lower bound, you must take an arbitrary S, then compute $\Sigma(G, d_S)$ and $\Sigma(H, d_S)$. In turn, computing $\Sigma(G, d_S)$ requires you to go into $\Phi(G)$.

A few students simply claim $\Phi(G, H) = \Omega(1/n)$ without any reasoning, which I can't give credit for.

There's one student who submitted $\Sigma(G, d_S) = \Theta(n)$ and $\Sigma(H, d_S) = \Theta(n^2)$. First, the $\Sigma(H, d_S)$ part is incorrect (for $|S| = o(n)$). I would accept $\Sigma(H, d_S) = \Theta(|S| \cdot n)$, but it is important to separate out $|S|$ to handle the numerator. Second, the $\Sigma(G, d_S)$ part is also incorrect if you don't look into $\Phi(G)$; after all, the cut might be sparse or even zero. It's from going into $\Phi(G)$ that you find out $\Sigma(G, d_S) = \Omega(|S|)$. (Not necessarily $\Theta(|S|)$, it might be denser than that, although I wouldn't make a fuss about this.) So this error is a bit more involved, but ultimately still damaging.

4 Problem 4

Let $G = (V, E)$ be connected unweighted graph, and let dist be the shortest-path metric on G. Let $p \in (0, 1)$ be a real number. Sample a random subset $A \subseteq V$ by including each vertex into A independently with probability p. Define the embedding to the line:

$$f_A(x) = \text{dist}(x, A) = \min_{a \in A} \text{dist}(x, a).$$

This gives the line metric $d_A(x, y) = |f_A(x) - f_A(y)|$. In the special case $A = \varnothing$, we say $d_A(x, y) = 0$.

The goal is to prove that the embedding has large distortion in expectation. For each n, construct a graph G with two distinguished vertices x, y, such that G has n vertices and

$$\mathbb{E}[d_A(x, y)] = O(n^{-0.01}) \cdot \text{dist}(x, y).$$

(Note: This can be written as

$$\frac{\text{dist}(x, y)}{\mathbb{E}[d_A(x, y)]} = \Omega(n^{0.01}).$$

This might explain better why we call this a "large" distortion: this fraction is not constant.)

(a) Construct such family of graphs when $p = 1/2$.

(b) Construct such family of graphs when $p = 1/n$.

4.1 My solution

Part (a). Let G be the path on n vertices, and x, y be its endpoints. We claim $\mathbb{E}[d_A(x, y)] = O(1)$. Since $\text{dist}(x, y) = n - 1$, this suffices, as we will have

$$\mathbb{E}[d_A(x, y)] = O(n^{-1}) \cdot \text{dist}(x, y).$$

First, use the triangle inequality to obtain

$$d_A(x, y) = |f_A(x) - f_A(y)| \le |f_A(x)| + |f_A(y)| = f_A(x) + f_A(y).$$

Therefore, by linearity of expectation and the fact that x, y are symmetric,

$$\mathbb{E}[d_A(x, y)] \le \mathbb{E}[f_A(x)] + \mathbb{E}[f_A(y)] = 2 \cdot \mathbb{E}[f_A(x)]$$

We now compute $\mathbb{E}[f_A(x)]$. By the tail-sum formula[1],

$$\mathbb{E}[f_A(x)] = \sum_{k=1}^{\infty} \Pr(f_A(x) \ge k).$$

We use the convention $f_A(x) = n$ if $A = \varnothing$. (This is fine, because if $A = \varnothing$, we're supposed to have $d_A(x, y) = 0$ anyway, while we're calling $d_A(x, y) \le f_A(x) + f_A(y) = 2n$. So this won't help $\mathbb{E}[d_A(x, y)]$ become smaller.) Then, note that $f_A(x) \le n$ for all A, so the sum actually only goes up to n.

Now, $\Pr(f_A(x) \ge k) = 2^{-k}$: this is just the probability that the k closest vertices to x, including x itself, are *not* in A. Each vertex is not in A with probability 2^{-1} and the vertices are independent, so the probabilities can just be multiplied together.

Therefore

$$\mathbb{E}[f_A(x)] = \sum_{k=1}^{n} 2^{-k} \le \sum_{k=1}^{\infty} 2^{-k} = 1$$

Thus $\mathbb{E}[d_A(x, y)] \le 2 \cdot 1 = 2 = O(1)$, proving the claim.

[1] I sometimes call it the "layer cake trick"

Part (b). Let G be a clique on n vertices, and let x, y be two of its vertices. We claim $\mathbb{E}[d_A(x,y)] = O(n^{-1})$. Since $\text{dist}(x,y) = 1$, this suffices, as we will have

$$\mathbb{E}[d_A(x,y)] = O(n^{-1}) \cdot \text{dist}(x,y).$$

Let E be the event $x, y \notin A$.

First, we claim that if event E happens, then $d_A(x,y) = 0$. This is pretty straightforward. If $A = \varnothing$, then $d_A(x,y) = 0$ by definition. Otherwise, there exists some $z \in A$. Since the graph is a clique, $f_A(x) = 1$ (adjacent to $z \in A$), and $f_A(y) = 1$ (same reason). Therefore

$$d_A(x,y) = |f_A(x) - f_A(y)| = 0.$$

Otherwise, we claim $d_A(x,y) \le 1$. This is also straightforward. $f_A(x) \le 1$ because every vertex is adjacent to x, and similarly $f_A(y) \le 1$. Therefore

$$d_A(x,y) = |f_A(x) - f_A(y)| \le \max\{f_A(x), f_A(y)\} \le 1.$$

Therefore, we can write the expectation:

$$\begin{aligned}\mathbb{E}[d_A(x,y)] &= \Pr(E) \cdot \mathbb{E}[d_A(x,y)|E] + \Pr(\overline{E}) \cdot \mathbb{E}[d_A(x,y)|\overline{E}] \\ &\le \Pr(E) \cdot 0 \qquad\qquad + \Pr(\overline{E}) \cdot 1\end{aligned}$$

So $\mathbb{E}[d_A(x,y)] \le \Pr(\overline{E})$.

Now, $\overline{E}$ is the event that at least one of x, y is in A. We have $\Pr(x \in A) = \Pr(y \in A) = p = 1/n$, so

$$\Pr(\overline{E}) \le \Pr(x \in A) + \Pr(y \in A) = \frac{2}{n} = O(n^{-1}).$$

Therefore $\mathbb{E}[d_A(x,y)] \le \Pr(\overline{E}) = O(n^{-1})$. This proves the claim.

4.2 Insight and discussion

Part (a). Since $p = \Omega(1)$, the closest vertex to x is very likely nearby; it gets exponentially less likely that it's far away. In fact, we can say that in expectation, the closest vertex is at constant distance away. In other words, $\mathbb{E}[f_A(x)] = O(1)$.

Now, a nifty trick on $d_A(x,y)$ is to weaken it using triangle inequality. It's a bit difficult to think of and justify, because it seems like it weakens things a lot. But it makes $d_A(x,y)$ much easier to manipulate, because the expectation $\mathbb{E}$ hates absolute values and loves linear combinations (i.e. additions). Once you break down $d_A(x,y)$ this way, you find out $\mathbb{E}[d_A(x,y)] = O(1)$ too, by using the previous observation.

Finally, since we have $\mathbb{E}[d_A(x,y)] = O(1)$ no matter the structure of the graph, just make x, y very far apart so $\text{dist}(x,y) = \Omega(n)$. The path is an obvious example.

Part (b). We cannot use the same idea as in part (a), because $\mathbb{E}[f_A(x)] = O(n)$ instead. We cannot have $\text{dist}(x,y) = \Omega(n^2)$. So we need a completely new trick. In particular, it's likely that we actually need to work with $d_A(x,y)$ as defined, instead of weakening it with triangle inequality.

A different idea is to simply make $d_A(x,y) = 0$. How can this happen? The straightforward execution is, x, y are both adjacent to a single vertex z. Then if $z \in A$, it follows $f_A(x) = f_A(y) = 1$ (as long as $x, y \notin A$, which is likely because p is small).

In fact, we can simply do this with *all* other vertices. Now, if neither of x, y is in A, then $d_A(x,y) = 0$. If either x, y is in A, then we also know $d_A(x,y) \le \text{dist}(x,y)$, which we can simply make constant $1 = O(1)$ by making x, y adjacent.

Because p is small, the probability either x, y is in A is also small, so this gives small $\mathbb{E}[d_A(x,y)]$. And that leads to the solution.

There are many other possible examples. See Appendix 4.4 for a list of them.

4.3 Grading criteria and common mistakes

4.3.1 (a) Passable error: Wrong $\mathbb{E}[f_A(x)]$

Using the path example, $\mathbb{E}[f_A(x)] \le 1$. Some of you write $\mathbb{E}[f_A(x)] = 1$ by taking the sum to infinity, but strictly speaking this is incorrect; the sum only goes up to $n-1$ because there are $n-1$ possible distances. But this is extremely minor, so I don't comment on it, much less deduct anything for it.

4.3.2 (a) Passable error: Using a cycle instead of a path

A construction I see is to use a cycle instead of a path, and taking x, y to be two opposite vertices. This can work, but $\mathbb{E}[f_A(x)]$ is a bit more difficult to calculate and it's easier to make an error. You should have $\Pr(f_A(x) > k) = 2^{-2k-1}$, so $\mathbb{E}[f_A(x)] \le 2/3$.

Of course, as long as you have $\mathbb{E}[d_A(x, y)] = O(1)$, the rest of the solution works. Just be careful about the calculation.

4.3.3 (b) Major error: Using a star and taking the center

Some of you use a star (see Appendix 4.4). If you use a star, the correct x, y is to take two of the leaves. If you take the center as one of them, your entire argument is wrong, because $d_A(x, y)$ will often be 1. (It is only 0 if x, y are both in A, with probability n^{-2}.)

Since the entire construction breaks down, I cannot accept this solution. But since you're thinking of a star, I'm willing to give a little bit of credit.

4.3.4 (b) Major error: Using the triangle inequality again

In (a), you analyze using the triangle inequality:

$$d_A(x, y) = |f_A(x) - f_A(y)| \le |f_A(x)| + |f_A(y)| = f_A(x) + f_A(y).$$

This approach doesn't work for (b). The main reason is, if you have a path of length k going out from x, then $\mathbb{E}[f_A(x)] = \Omega(k)$. To see this, recall that $(1 - 1/n)^n \to \exp(-1)$. Therefore, there exists some constant $\Omega(1)$ such that everything not on this path is not in A, and when this happens, we have:

$$\mathbb{E}[f_A(x) | A \text{ is entirely on the path}] = \sum_{i=1}^{k} \Pr(f_A(x) \ge i) = k \cdot \Omega(1) = \Omega(k).$$

(The probability is $\Omega(1)$ by using the same $(1 - 1/n)^n \to \exp(-1)$.)

Therefore $\mathbb{E}[f_A(x) + f_A(y)] = \Omega(k)$ too. But if x, y are at distance $\mathrm{dist}(x, y)$, then there is a path of length k going out from x by definition, it's the path to y. So $\mathbb{E}[f_A(x) + f_A(y)] = \Omega(\mathrm{dist}(x, y))$, so you don't get the $O(n^{-0.01})$ multiplier.

That's why you need to work with $d_A(x, y)$ directly, making use of the cases where $f_A(x) = f_A(y)$. Any argument using the triangle inequality will be doomed.

4.4 Appendix: Other examples

Here are various other examples I've seen, with a quick reasoning/proof of why they work.

4.4.1 (a) Thin bridge

Create two cliques (the "islands") of size $n/3$ and connect them with a path (the "thin bridge") of length $n/3$. Take x from one clique and y from the other.

Now $f_A(x) > 1$ only if everything in x's clique is not chosen. This happens with probability $2^{-n/3}$. So with probability $2 \cdot 2^{-n/3}$, the bad case happens, but at least $d_A(x, y) \leq n = o(2^{n/3})$. And with the rest of the probability, we have $f_A(x), f_A(y) \leq 1$ so $d_A(x, y) \leq 1$. This gives $\mathbb{E}[d_A(x, y)] = O(1)$.

You can change the sizes of the islands and the bridge as appropriate, and you can even leave out some of the vertices to form other parts of the graph. You just need the islands to have size $\Omega(\log n)$ with a large enough constant (≥ 0.02 or so, depending on the exponent in $O(n^{-0.01})$), and the bridge to be $\Omega(n)$ long. Any remaining vertices can do whatever they want, including being part of the islands: it will just push down the probability $f_A(x) > 1$ or $f_A(y) > 1$ even more.

4.4.2 (b) Star

Take the star graph ($K_{1,n-1}$) where x, y are two of its leaves.

It is still true that if $x, y \notin A$, then $d_A(x, y) = 0$. The proof is basically by cases. If the center is in A, then $f_A(x) = f_A(y) = 1$. If some leaf is in A instead, then $f_A(x) = f_A(y) = 2$.

Then the rest of the proof, the analysis of the probability and $\mathbb{E}[d_A(x, y)]$, is identical, just using $d_A(x, y) = 2$ instead of 1.

4.4.3 (b) Cloned vertex

We can take the clique (from the official solution) and the star (from above) and generalize it all the way to the following:

Let G be *any* connected graph, and take a vertex v. Now clone v into v', with exactly the same adjacencies. (vv' may be an edge but doesn't have to be.) Now $x = v$ and $y = v'$.

The analysis is once again identical. The main idea here is, for any vertex u, we have xu is an edge if and only if yu is an edge. So if x has a path leading to z, replacing the first vertex with y gives a path from y to z, and vice versa. So as long as $x, y \notin A$, the shortest distance to A is the same for both vertices.

4.5 Appendix: All the p, all the exponents

There are two quantities in the problem that can obviously be generalized. First, the exponent 0.01 can be generalized into an exponent $c \geq 0$ (although the case $c = 0$ is trivial since $d_A(x, y) \leq \text{dist}(x, y)$). Second, the probability p can be generalized into any function on n.

Intuitively, (a) covers the case of large p and (b) covers the case of small p. But is that enough? Do these two cases cover the entirety of p, or is there some value of p in between that is not covered by either case?

We can prove the following. Using the constructions we have:

- The thin bridge construction for (a) (see Appendix 4.4) works for all $c \leq 1$, as long as $p = \omega(n^{-1} \log n)$.
- The path construction for (a) (as in the main solution) works for $p = \Omega(n^{c-1})$.
- The cloned vertex construction for (b) (see Appendix 4.4; this includes the clique from the main solution) works for $p = O(n^{-c})$ and $1 - p = O(n^{-c})$, and only those.

Or, rephrasing in terms of c:

- If $c < 1$, then the thin bridge construction and the cloned vertex construction collectively cover all possible p for all large enough n. This is because the case $p = \Omega(n^{-c}) = \omega(n^{-1}\log n)$ is covered by the thin bridge construction, and the case $p = O(n^{-c})$ is covered by the cloned vertex construction.
- If $c = 1$, we cannot claim $\Omega(n^{-1}) = \omega(n^{-1}\log n)$, so there is a tiny gap between $p = \Omega(n^{-1})$ and $p = O(n^{-1}\log n)$ where those two constructions are not enough.
- If $c > 1$, then we have basically nothing for large p; neither the thin bridge construction nor the path construction work. (But also, see Section 4.5.5.)

It remains to give the analysis of the constructions. We will analyze them in reverse order of how they were mentioned above, since I think this is the most natural way to write it.

4.5.1 The clique construction

We follow the same analysis, just making it a bit more precise. If $x, y \notin A$, then $d_A(x, y) = 0$ indeed. If $x, y \in A$, then also $d_A(x, y) = 0$. The only case where $d_A(x, y) \neq 0$ is if one of x, y is in A, and the other isn't. This happens more precisely with probability $2p(1-p)$. (We earlier just gave the bound $\leq 2p$.)

Therefore,

$$\mathbb{E}[d_A(x, y)] = 2p(1-p).$$

Therefore, if $p = O(n^{-c})$, we have $\mathbb{E}[d_A(x, y)] = O(p) = O(n^{-c}) \cdot \operatorname{dist}(x, y)$. Similarly, if $1 - p = O(n^{-c})$, we also have $\mathbb{E}[d_A(x, y)] = O(1-p)$ with the same conclusion.

However, if $p, 1-p = \omega(n^{-c})$, note that at least one of them is $\Omega(1)$. So

$$\mathbb{E}[d_A(x, y)] = 2p(1-p) = 2 \cdot \omega(n^{-c}) \cdot \Omega(1) = \omega(n^{-c}) = \omega(n^{-c}) \cdot \operatorname{dist}(x, y)$$

so this construction will not work. So the *only* values of p that work with this clique construction are the ones where $p = O(n^{-c})$ or $1 - p = O(n^{-c})$.

4.5.2 The cloned vertex construction for (b)

The above analysis can actually be extended for the cloned vertex construction. It is still true that, if x, y are both in A or both not in A, then $d_A(x, y) = 0$. The only remaining case is if one is in A and the other isn't.

Now, if x, y are adjacent, then $\operatorname{dist}(x, y) = 1$ and $d_A(x, y) = 1$ too. Otherwise, because the graph is connected, x must be adjacent to some vertex z. The cloned vertex construction says y is also adjacent to z. Therefore $\operatorname{dist}(x, y) = 2$, and also $d_A(x, y) \in \{1, 2\}$ (either 1 if $z \in A$, or 2 if no such $z \in A$ and so y's nearest vertex is x). Therefore the equation just changes into

$$2p(1-p) \leq \mathbb{E}[d_A(x, y)] \leq 4p(1-p) \qquad \implies \qquad \mathbb{E}[d_A(x, y)] = \Theta(p(1-p))$$

and everything works identically as above.

4.5.3 The path construction

We again use the convention $f_A(x) = n$ if $A = \varnothing$. For the path construction, we have

$$\mathbb{E}[f_A(x)] = \sum_{k=1}^{\infty} \Pr(f_A(x) \geq k) = \sum_{k=1}^{n} (1-p)^k = \frac{1-p}{p} \cdot (1 - (1-p)^n)$$

We can also simplify this expression:

$$\mathbb{E}[f_A(x)] = \frac{1-p}{p} \cdot (1 - (1-p)^n) \leq \frac{1}{p}$$

In fact, if p is small, then $1-p$ and $1-(1-p)^n$ are both approximately 1 anyway, so $\mathbb{E}[f_A(x)] \approx 1/p$. (The error can be made precise, but I'll skip it to focus on the main idea.)

If we know $p = \Omega(n^{c-1})$, then $\mathbb{E}[f_A(x)] = O(n^{1-c})$, so

$$\mathbb{E}[d_A(x,y)] \le 2 \cdot \mathbb{E}[f_A(x)] = O(n^{1-c}) = O(n^{-c}) \cdot \mathrm{dist}(x,y).$$

On the other hand, if $p = o(n^{c-1})$, then this approach doesn't work, because

$$E[f_A(x)] = \omega(n^{1-c}) = \omega(n^{-c}) \cdot \mathrm{dist}(x,y)$$

so using $\mathbb{E}[d_A(x,y)] \le 2 \cdot \mathbb{E}[f_A(x)]$ will not give any appropriate upper bound.

Can we use a different analysis to get a better bound for this path construction? Possibly, but in that case, we have to work with $d_A(x,y) = |f_A(x) - f_A(y)|$ directly instead of using the triangle inequality. And it's just so horrible. You may attempt it if you want.

4.5.4 The thin bridge construction

To repeat the analysis for the thin bridge construction, we have two cases: $f_A(x) > 1$ and $f_A(x) \le 1$. We have good bounds on both cases. For the case $f_A(x) > 1$, this means nothing on x's island can be chosen, so this only happens with probability $\le (1-p)^{n/3}$. And note that we still have $f_A(x) \le n$. For the case $f_A(x) \le 1$, then well, $f_A(x) \le 1$.

Therefore

$$\begin{aligned}
\mathbb{E}[f_A(x)] &= \Pr(f_A(x) > 1) \cdot \mathbb{E}[f_A(x) | f_A(x) > 1] + \Pr(f_A(x) \le 1) \cdot \mathbb{E}[f_A(x) | f_A(x) \le 1] \\
&\le (1-p)^{n/3} \cdot n + (1 - (1-p)^{n/3}) \cdot 1 \\
&\le 1 + (n-1) \cdot (1-p)^{n/3} \\
&= (O(n^{-1}) + (1-p)^{n/3}) \cdot \Theta(n)
\end{aligned}$$

Note that $\mathrm{dist}(x,y) = \Theta(n)$, so we need $O(n^{-1}) + (1-p)^{n/3}$ to be $O(n^{-c})$. This means we need $c \le 1$. On the other hand, given that $c \le 1$, we just need $(1-p)^{n/3} = O(n^{-c})$; in other words, $\Theta(n^c) \cdot (1-p)^{n/3} = O(1)$.

First, suppose $p = \omega(n^{-1} \log n)$. (For instance, we might have $p = \Omega(n^{-k})$ for some $k < 1$.) By using $1 - x \le \exp(-x)$, we have

$$\begin{aligned}
\Theta(n^c) \cdot (1-p)^{n/3} &\le \exp(c \log n) \cdot \exp\left((-p) \cdot \frac{n}{3}\right) \\
&= \exp\left(c \log n - \omega(n^{-1} \log n) \cdot \frac{n}{3}\right) \\
&= \exp\left(\log n \cdot (c - \omega(1))\right) \\
&= n^{c - \omega(1)}
\end{aligned}$$

Now, $c - \omega(1) \to -\infty$, so $\Theta(n^c) \cdot (1-p)^{n/3} \to 0$, i.e. $o(1)$. Therefore this construction works for this p.

On the other hand, if $p = \Theta(n^{-1} \log n)$, then the analysis is similar except that the last line becomes $n^{c - \Theta(1)}$. So we need a very precise constant in order for this $\Theta(1)$ to be $\ge c$. The constant hidden in $p = \Theta(n^{-1} \log n)$ now very much matters.

Furthermore, if $p = o(n^{-1} \log n)$, for example $p = O(n^{-1})$, then the last line becomes $n^{c - o(1)}$. For $c > 0$, this is $\to \infty$, so it is not $O(1)$. So that p will not work. And if $c = 0$, the statement is trivial.

Therefore, this analysis works for $p = \omega(n^{-1} \log n)$, but does not work for smaller values of p.

Is it possible to get a better analysis by tightening the bounds? Once again, I'm not sure. The other island is at distance $n/3 = \Theta(n)$ and so already matches the bound we have. So any analysis we want to do will have to handle the thin bridge, which is a path. Tightening the bound means working with the path more closely, and this is something the previous analysis did. So if we don't yet have new progress for the path, it's likely the thin bridge here can't be improved either. But again, you're free to try.

4.5.5 The best c

We can ask from the other direction. Given p (as a function on n), what is the best possible c among *all* possible constructions? I don't have a complete answer, but there's some work that can be done.

Let's look into $\mathbb{E}[d_A(x,y)]$. This expectation contains at least the following one term: $x \in A$ and $y \notin A$. This happens with probability $p(1-p)$. When this happens, $d_A(x,y) \geq 1$. Therefore $\mathbb{E}[d_A(x,y)] \geq p(1-p)$.

Now, suppose $p, 1-p = \Omega(n^{-k})$. Note that at least one of them is $\Omega(1)$ anyway, so $p(1-p) = \Omega(n^{-k})$ too. Therefore $\mathbb{E}[d_A(x,y)] = \Omega(n^{-k})$.

Furthermore, note that $\text{dist}(x,y) = O(n)$. So this simple bound already gives

$$\mathbb{E}[d_A(x,y)] = \Omega(n^{-k-1}) \cdot \text{dist}(x,y).$$

Therefore, $c \leq k+1$.

Especially interesting is the case $k = 0$. This is when $p, 1-p = \Omega(1)$, so basically any constant p. This claims $c \leq 1$ is in fact the best possible c. We do have a construction for $c \leq 1$: either the path construction or the thin bridge construction suffices. So the problem is fully solved for constant p.

In the converse direction, if $c > 1$, then we need $k > 0$. That means for $c > 1$, many of the large p values will *never* work. We can only ask the question for $p = O(n^{1-c})$.

Finally, the clique construction works for $c \leq k$. There is still a gap in between, for $k < c \leq k+1$. We know the clique construction cannot work for this case, but is there another construction that make this work?

There are so many questions to answer, and there's not enough time. I will leave it to you in case you're interested in researching this problem further.