Report Notes: LP and Rounding for MinDisagree Correlation Clustering on Complete Graphs

Problem Formulation and Algorithm

1. Problem Definition (Weighted MinDisagree)

Given a complete graph $G = (V, E)$ with positive and negative edges, each edge (i, j) has weight w_{ij}.
The objective is to minimize disagreement:

- If (i, j) is positive and separated: cost w_{ij}
- If (i, j) is negative and clustered together: cost w_{ij}

2. Integer Program (IP)

$$\min \sum_{(i,j)\in +} w_{ij}x_{ij} + \sum_{(i,j)\in -} w_{ij}(1 - x_{ij})$$

Subject to triangle inequalities:

$$x_{ik} \leq x_{ij} + x_{jk} \quad \forall i, j, k$$

$$x_{ij} \in \{0, 1\}$$

Interpretation:

$$x_{ij} = \begin{cases} 0 & \text{if } i, j \text{ in same cluster} \\ 1 & \text{if } i, j \text{ separated} \end{cases}$$

Relaxing to $x_{ij} \in [0, 1]$ gives the LP.

3. Rounding Algorithm (1/2-ball Clustering)

1. Initialize $S = V$
2. Pick $u \in S$
3. Define $T = B_x(u, 1/2) = \{i \mid x_{ui} \leq \frac{1}{2}\}$
4. If $\frac{1}{|T|} \sum_{i\in T} x_{ui} \geq \frac{1}{4}$, then output singleton cluster $\{u\}$, otherwise output cluster $\{u\} \cup T$
5. Remove the cluster and repeat

Approximation Analysis (Unweighted Complete Graph)

The goal is to show that the LP pays for the algorithmic cost.

Case 1: Output Singleton $\{u\}$

We know that $\sum_{i \in T} x_{ui} \geq \frac{|T|}{4}$, so LP pays at least $\sum_{i \in T} x_{ui} \geq \frac{|T|}{4}$ and the algorithm cuts at most $|T|$ edges incident to u. Thus $ALG \leq 4 \cdot LP$.

Case 2: Output Cluster $\{u\} \cup T$

For an edge (i,j), assume (w.l.o.g.) that $x_{ui} \leq x_{uj}$, i.e., j is farther from u than i in the LP metric. For negative edges, fix a vertex $j \in T$; for positive edges, fix a vertex $j \notin T$. Let p_j (resp. n_j) be the number of positive (resp. negative) edges adjacent to j. Let LP_j denote the contribution charged to j in the LP objective. Using triangle inequality $x_{ij} \leq x_{ui} + x_{uj}$, We can separately obtain the following two bounds:

$$LP_j \geq \min\{\frac{p_j}{8} + \frac{3n_j}{8}, \ \frac{p_j}{4} + \frac{n_j}{4}\}$$

$$LP_j \geq \min\{\frac{p_j}{2}, \ \frac{p_j}{4} + \frac{n_j}{4}\}$$

The algorithmic cost at j is at most $p_j + n_j$, hence the LP lower bound absorbs the algorithmic cost, yielding a 4-approximation.

Conclusion

For unweighted complete graphs:

- LP relaxation with triangle inequalities
- 1/2-ball rounding
- Vertex-wise LP lower bounds (charging argument)

Together imply a 4-approximation.

Possible Improvements

(1) Pivot Selection Order

In the current algorithm, the pivot u is chosen arbitrarily. A possible refinement is to adjust the selection order, for example, prioritize outputting a singleton or forming a non-singleton cluster whenever possible.
Different pivot orders may improve constants.

(2) Stability of the Threshold

The rounding uses threshold 1/2 to define the ball $B_x(u, 1/2)$.
It is natural to ask whether small perturbations of this threshold (e.g. $1/2 \pm \varepsilon$) significantly change the rounding outcome.
Studying this stability may help understand the robustness of the rounding scheme and its impact on the approximation ratio.