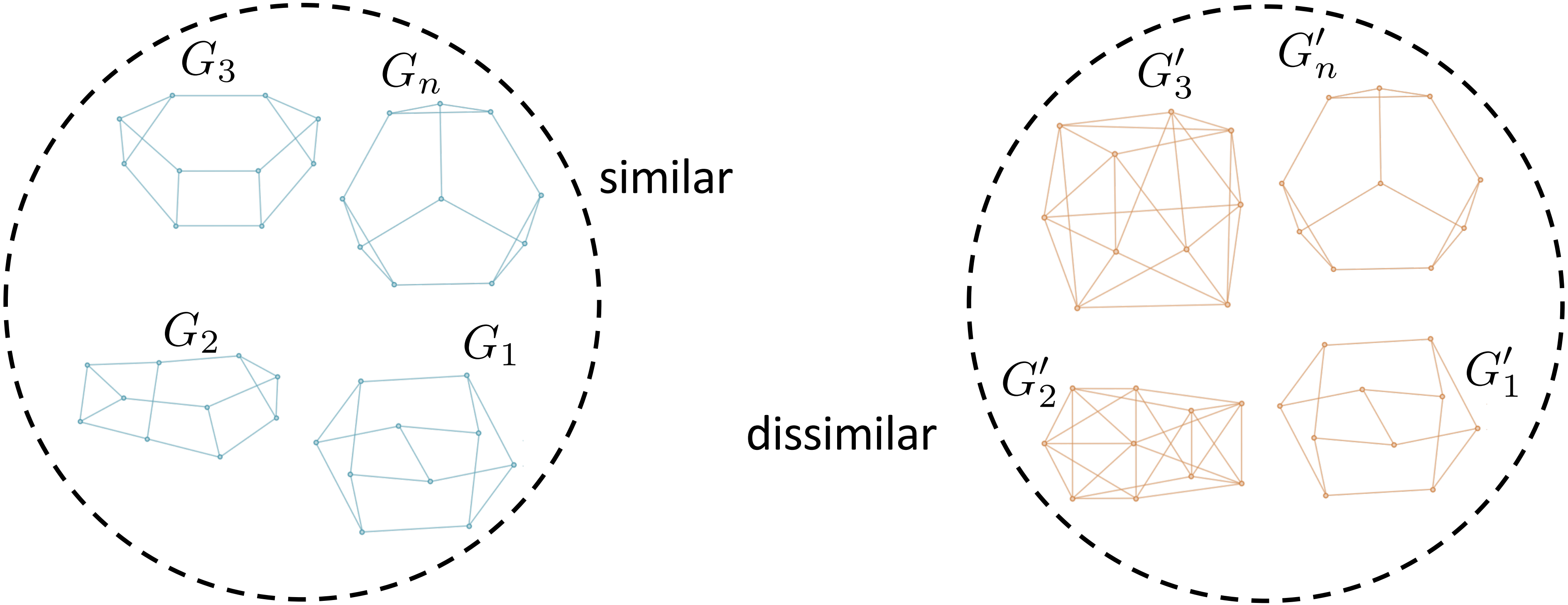
Problem

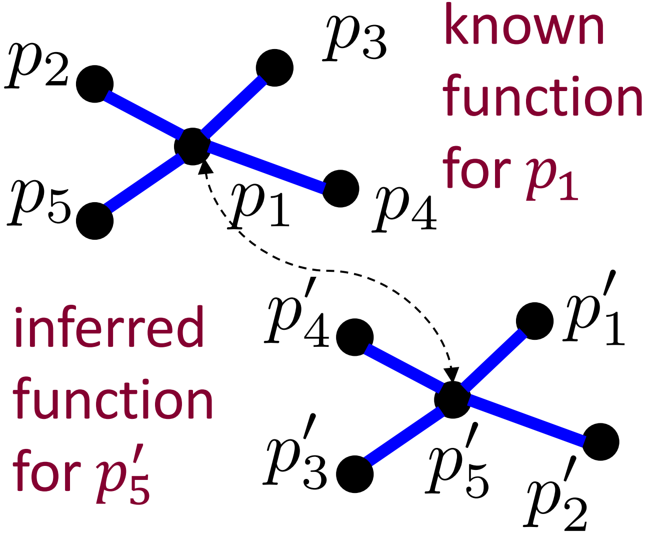


$$d(G_1, \ldots, G_n) \sim \text{small}$$

$$d(G'_1, \ldots, G'_n) \sim \text{big}$$

Additional goals

Find an association among graph nodes to show why graphs are similar. This allows, e.g., knowledge transfer.



The association among multiple graphs should be consistent.

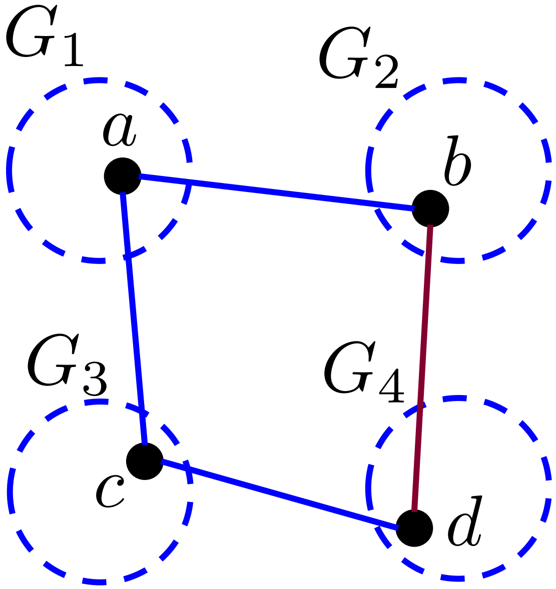
The distance function should satisfy intuitive properties of *metrics*.

$$d(G_1, G_3) \leq d(G_1, G_2) + d(G_2, G_3)$$

Additional goals

Find an association among graph nodes to show why graphs are similar. This allows, e.g., knowledge transfer.

The association among multiple graphs should be consistent.



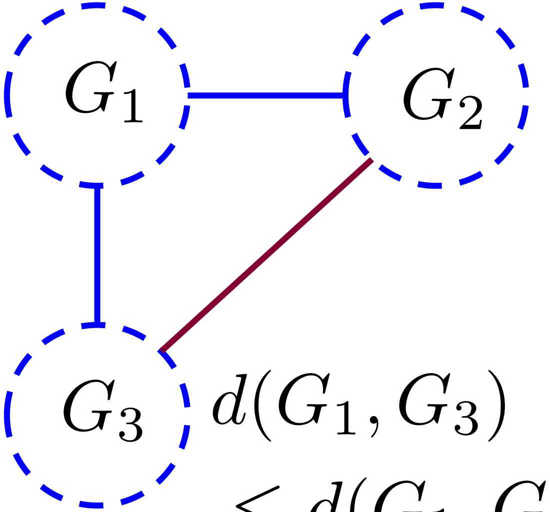
The distance function should satisfy intuitive properties of *metrics*.

Additional goals

Find an association among graph nodes to show why graphs are similar. This allows, e.g., knowledge transfer.

The association among multiple graphs should be consistent.

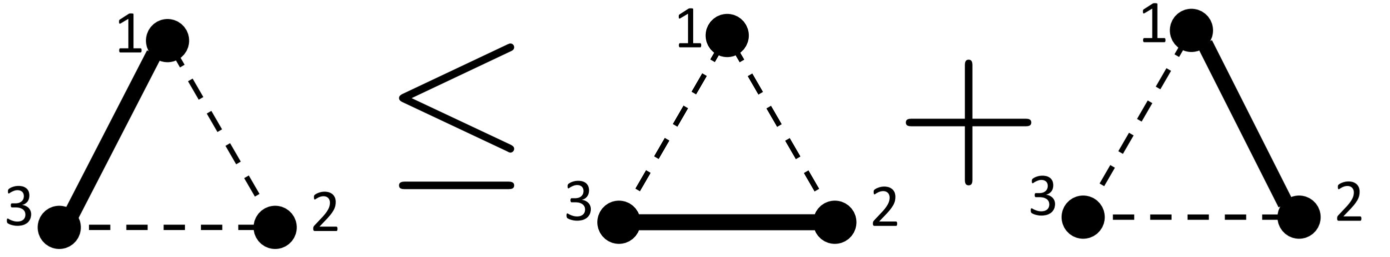
The distance function should satisfy intuitive properties of *metrics*.



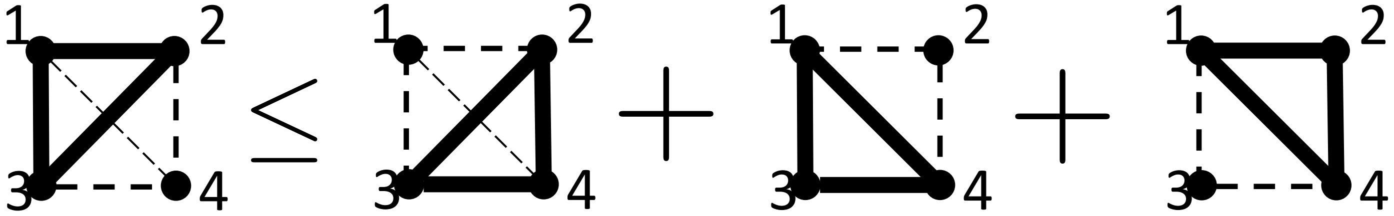
$$d(G_1, G_3) \leq d(G_1, G_2) + d(G_2, G_3)$$

n-metrics

$$d(G_1, G_3) \leq d(G_1, G_2) + d(G_2, G_3)$$



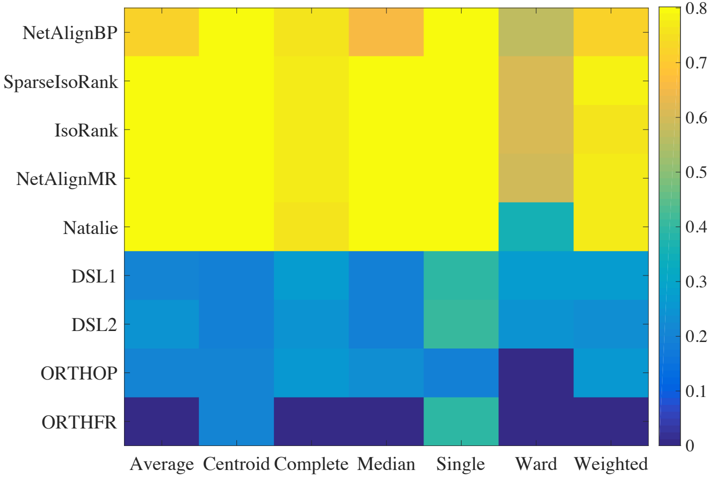
$$d(G_1, G_2, G_3) \leq d(G_2, G_3, G_4) + d(G_1, G_3, G_4) + d(G_1, G_2, G_4)$$



Why is this important?

1. Some algorithms can use the metric property to save computation time. E.g. a simple randomized algorithm can solve $\max_{G_1, G_2 \in S} d(G_1, G_2)$ in $\mathcal{O}(|S|)$ (1/2-approx. in expectation) v.s. $\mathcal{O}(|S|^2)$

2. Some algorithms show better accuracy when using metrics.



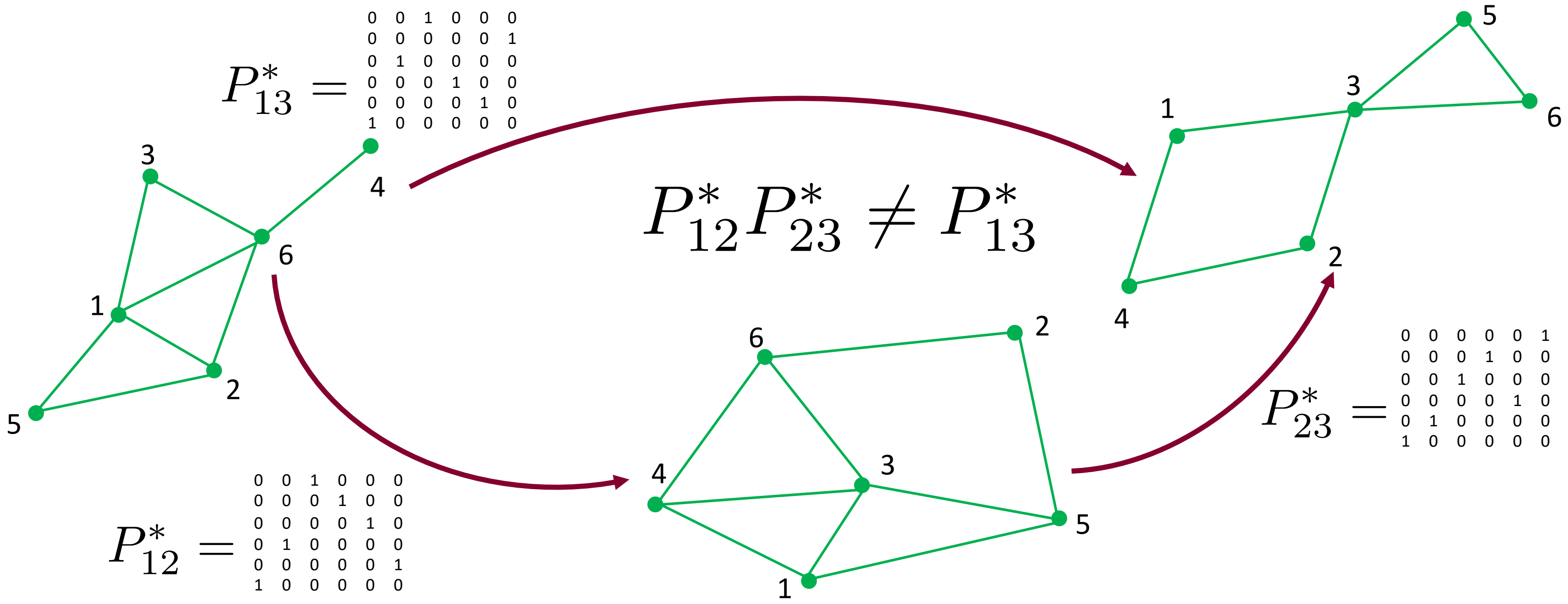
E.g. we can cluster graphs using distance-based clustering, and when we do so, we observe that using metrics results in better clustering performance than when using non-metrics.

Attempt 1

Given a metric $d(G_i, G_j)$, an easy way to obtain an n-metric is to define

$$d(G_1, \ldots, G_n) = \sum_{(i,j)} d(G_i, G_j) \quad ?$$

If $d(G_i, G_j)$ returns an assignment P_{ij}, we might not have consistency.



$$P^*_{13} = \begin{matrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

$$P^*_{12} P^*_{23} \neq P^*_{13}$$

$$P^*_{23} = \begin{matrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

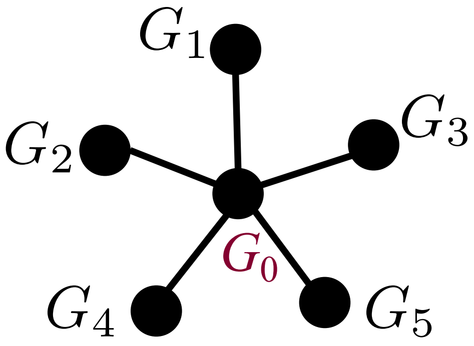
$$P^*_{12} = \begin{matrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

Attempt 2

Another easy way to obtain an n-metric is to define

$$d(G_1, \ldots, G_n) = \min_{G_0} \sum_{i=1}^{n} d(G_i, G_0)$$

This is called the Fermat distance associated with d.



We want G_0 to be close to all of the G_i 's. If we can find such a G_0, then the graphs are similar. If $d(G_i, G_0)$ returns an assignment P_{i0}, and we define $P_{ij} = P_{i0}(P_{j0})^{\mathrm{T}}$ then we have consistency.

However, the optimization over G_0 makes this definition hard to use.

Our definition: g-align

Under an appropriate choice of C, the following are n-metrics, reduce to solving a convex optimization problem, and satisfy a relaxed notion of alignment consistency.

$$d(G_1, ..., G_n) = \min_{\substack{P_{i,j} \in \mathcal{C} \\ P_{i,i} = I \\ \mathbf{P} \succeq 0}} \frac{1}{2} \sum_{i,j \in [n]} \|A_i P_{i,j} - P_{i,j} A_j\|$$

$$d(G_1, ..., G_n) = \min_{\substack{P_{i,j} \in \mathcal{C} \\ P_{i,i} = I \\ \|\mathbf{P}\|_* \leq mn}} \frac{1}{2} \sum_{i,j \in [n]} \|A_i P_{i,j} - P_{i,j} A_j\|$$

$\mathcal{C} =$ some convex set of matrices

For details, check Thursday's poster session

Pacific Ballroom #145