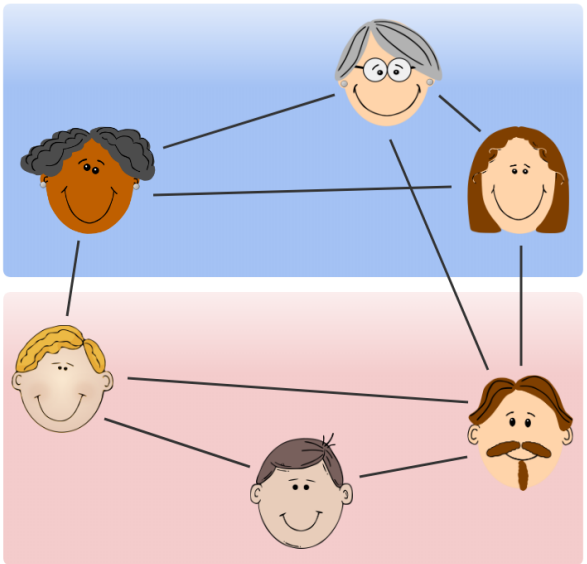
Guarantees for Spectral Clustering with Fairness Constraints

Matthäus Kleindessner, **Samira Samadi**, Pranjal Awasthi & Jamie Morgenstern

RUTGERS Georgia Tech

Spectral Clustering (SC) and Fairness

SC is the method of choice for clustering the nodes of a graph.



Friendship network: SC can result in highly unfair clustering with respect to the two demographic groups.

Fair clustering (Chierichetti et al. 2017): in every cluster, each group V_s should be represented with (approximately) the same fraction as in the whole data set V.

Goal: Study spectral clustering with fairness constraints.

Spectral Clustering

Goal: Partition V into k clusters with min *RatioCut objective* value.

◇ Encode a clustering $V = C_1 \dot{\cup} \ldots \dot{\cup} C_k$ by $H \in \mathbb{R}^{n \times k}$ with

$$H_{il} = \begin{cases} 1/\sqrt{|C_l|}, & i \in C_l \\ 0, & i \notin C_l \end{cases} \tag{1}$$

$\text{RatioCut}(C_1, \ldots, C_k) = \text{Tr}(H^T L H)$. L is the *graph Laplacian matrix*.

◇ The exact problem:

$$\min_{H \in \mathbb{R}^{n \times k}} \text{Tr}(H^T L H) \text{ subject to } H \text{ is of form (1)}$$

◇ Solve the relaxed version:

$$\min_{H \in \mathbb{R}^{n \times k}} \text{Tr}(H^T L H) \text{ subject to } H^T H = I_k.$$

◇ Apply k-means clustering to the rows of H.

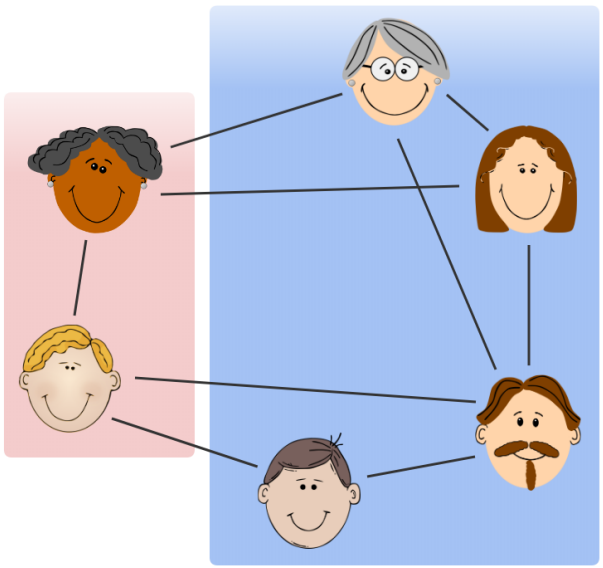
Spectral Clustering with Fairness Constraints

Approach: Incorporate fairness as a linear constraint

$$\min_{H \in \mathbb{R}^{n \times k}} \operatorname{Tr}(H^T L H) \quad \text{subject to } H^T H = I_k \,\&\, F^T H = 0.$$

Convert the program to the standard form and solve.

$\rightsquigarrow$ Our approach is analogous to existing versions of constrained SC that try to incorporate must-link constraints (e.g. Yu and Shi '04)



Friendship network: Our algorithm finds a fair clustering with respect to the two demographic groups.

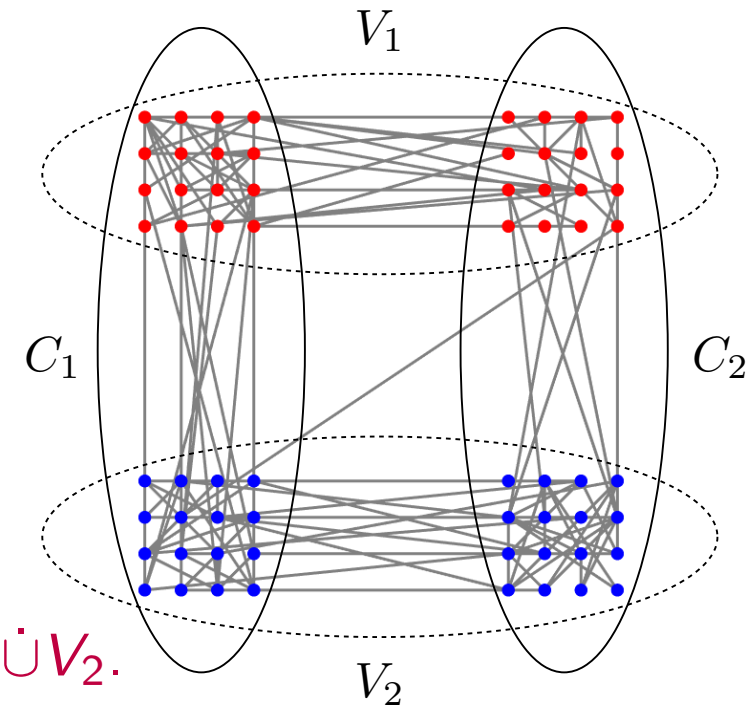
Analysis on Variant of Stochastic Block Model

Given V with a fair ground-truth clustering e.g., $V = C_1 \dot{\cup} C_2$

$$\Pr(i,j) = \begin{cases} a, & i \text{ and } j \text{ in same group and in same cluster} \\ b, & i \text{ and } j \text{ in same group, but in different clusters} \\ c, & i \text{ and } j \text{ in different groups, but in same cluster} \\ d, & i \text{ and } j \text{ in different groups, and in different clusters} \end{cases}$$

for some $a > b > c > d$.

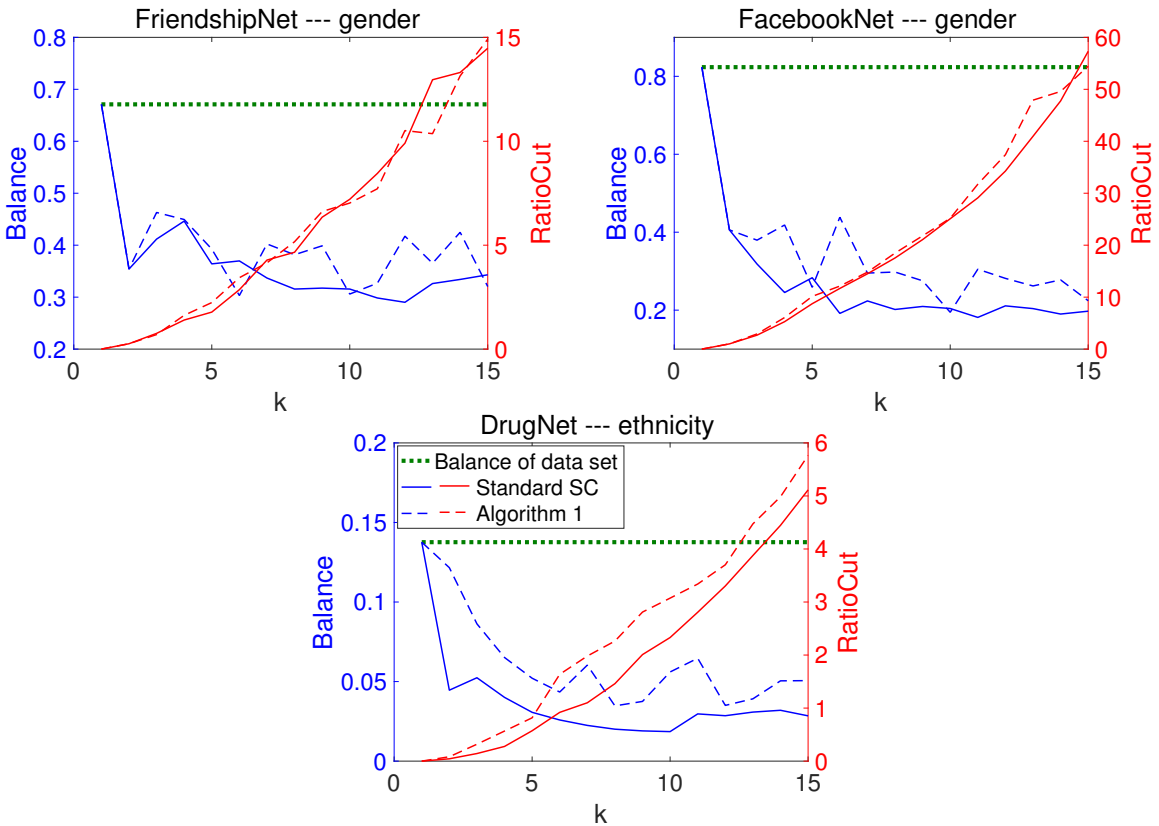
Theorem *(informal): Fair SC recovers the ground-truth clustering $C_1 \dot{\cup} C_2$ with high probability.*



Standard SC is likely to return $V_1 \dot{\cup} V_2$.

Experiments on Real Networks

FriendshipNet, FacebookNet, DrugNet



Average balance of clusters and RatioCut value as a function of number of clusters.

Thank you!

Poster #195