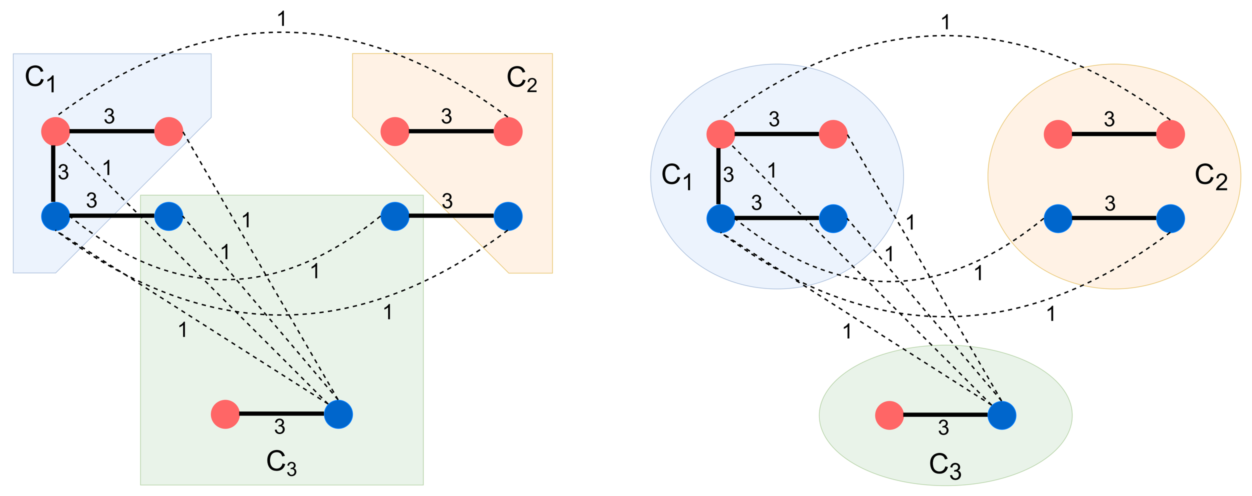
Scalable Spectral Clustering with Group Fairness Constraints

Ji Wang [1] Ding Lu [2] Ian Davidson [1] Zhaojun Bai [1]

[1]University of California, Davis [2]University of Kentucky

Group Fairness in Clustering

- **Clustering.** Given a graph $\mathcal{G}(V, W)$ with vertices $V = \{v_1, \dots, v_n\}$ and weighted adjacency matrix $W = (w_{ij}) \in \mathbb{R}^{n\times n}$, we want partition V into k disjoint clusters
$$V = C_1 \cup \dots \cup C_k,$$
such that the total weights within each subset are large and between two subsets are small.
- **Fair Clustering** (Chierichetti et al., 2017). For a given group partition
$$V = V_1 \cup \dots \cup V_h,$$
a clustering $V = C_1 \cup \dots \cup C_k$ is called **group fair (statistical parity)** if in each C_ℓ, the objects from each group are presented proportionately:
$$\frac{|V_s \cap C_\ell|}{|C_\ell|} = \frac{|V_s|}{|V|}, \qquad (1)$$
for $s \in [1 \cdots h]$ and $\ell \in [1 \cdots k]$.



(a) unfair clustering (b) fair clustering

- **Fair Clustering as Linear Constraints.** Let $H = \{0,1\}^{n\times k}$ be the *clustering indicator matrix*, and $G = \{0,1\}^{n\times h}$ be the *group indicator matrix*, then the group fairness constraints (1) is equivalent to
$$F^T H = 0, \qquad (2)$$
where $F \in \mathbb{R}^{n\times(h-1)}$ is the matrix by deleting last column of $F_0 := (I_n - \mathbf{1}_n\mathbf{1}_n^T/n)G$.

Spectral Clustering

- **Normalized cut (NCut)** (Shi and Malik, 2000).
$$\mathrm{NCut}(C_1, \cdots, C_k) := \sum_{\ell=1}^{k} \left(\frac{\sum_{v_i \in C_\ell, v_j \notin C_\ell} w_{ij}}{\sum_{v_i \in C_\ell} d_i} \right) = \mathrm{Tr}\,(H^T L H),$$
where $d_i = \sum_{j=1}^n w_{ij}$ is the *degree* of a vertex v_i,
$$L = D - W \quad \text{with } D = \mathrm{diag}(d_1, d_2, \dots, d_n)$$
is the graph *Laplacian*, and $H \in \mathbb{R}^{n\times k}$ is the *scaled clustering indicator matrix*.
- **Spectral Clustering** (Ng et al., 2001) solves the relaxed NCut minimization
$$\min_{H\in\mathbb{R}^{n\times k}} \mathrm{Tr}\,(H^T L H) \quad \text{s.t.} \quad H^T D H = I_k, \qquad (3)$$
which is equivalent to $LH = HD\Lambda_k$.
- **Algorithm 1: Spectral Clustering (SC)**

Input: Laplacian $L = D - W \in \mathbb{R}^{n\times n}$
Output: a clustering of indices $1:n$ into k clusters
1: compute the normalized Laplacian $L_n = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$;
2: compute the k smallest eigenvalues of L_n and the corresponding eigenvectors $X \in \mathbb{R}^{n\times k}$;
3: apply k-means clustering to the rows of $H = D^{-\frac{1}{2}} X$.

Fair Spectral Clustering and the Pitfalls

- **Fair Spectral Clustering.**
$$\min_{H\in\mathbb{R}^{n\times k}} \mathrm{Tr}\,(H^T L H) \ \text{ s.t. } \ H^T D H = I_k \text{ and } F^T H = 0. \qquad (4)$$
- **A Nullspace-based Algorithm** (Kleindessner et al., 2019).
The constraint $F^T H = 0$ implies
$$H = ZY \quad \text{for some } Y \in \mathbb{R}^{(n-h+1)\times k},$$
where $Z \in \mathbb{R}^{n\times(n-h+1)}$ is an orthogonal basis of the null(F^T). Therefore, (4) turns to
$$\min_{Y\in\mathbb{R}^{(n-h+1)\times k}} \mathrm{Tr}\,(Y^T [Z^T L Z] Y) \quad \text{s.t.} \quad Y^T [Z^T D Z] Y = I_k.$$
By a change of variables $Y = Q^{-1} X$ with $Q = (Z^T D Z)^{1/2}$, (4) is equivalent to the eigenvalue problem
$$MX = X\Lambda_k, \quad \text{with} \quad M = Q^{-1} Z^T L Z Q^{-1}. \qquad (5)$$
- **Algorithm 2: Fair Spectral Clustering (FairSC)**

Input: Laplacian $L = D - W \in \mathbb{R}^{n\times n}$, group indicator $F \in \mathbb{R}^{n\times(h-1)}$
Output: a clustering of indices $1:n$ into k clusters
1: compute an orthonormal basis Z of the nullspace of F^T; // $\mathcal{O}(n(h-1)^2)$
2: compute the matrix square root $Q = (Z^T D Z)^{1/2}$; // $\mathcal{O}((n-h+1)^3)$
3: compute $M = Q^{-1} Z^T L Z Q^{-1}$;
4: compute the k smallest eigenvalues of M and the corresponding eigenvectors $X \in \mathbb{R}^{n\times k}$;
5: apply k-means clustering to the rows of $H = ZQ^{-1}X$.

- **Pitfalls:** FairSC is not scalable due to its expensive kernels, such as computing nullspace and matrix square root.

A Novel Scalable FairSC algorithm

Our scalable fair spectral clustering algorithm is based on **nullspace projection** and **Hotelling's deflation** (Hotelling, 1943).

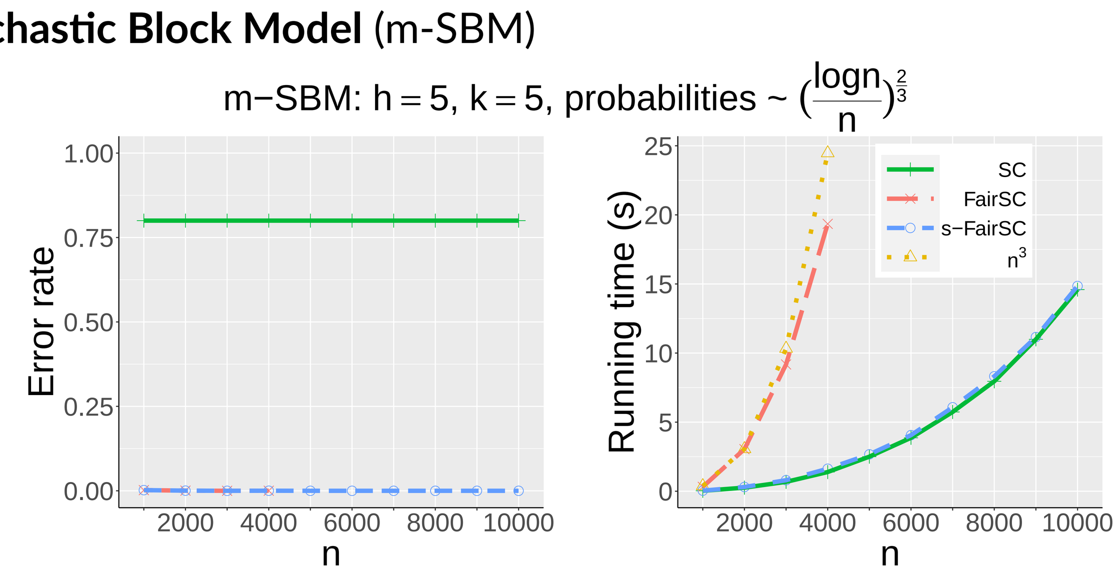
- **Constrained Eigenvalue Problem.** By change of variable $X = D^{\frac{1}{2}} H$, (4) gives
$$L_n X = X\Lambda_k \quad \text{s.t.} \quad C^T X = 0, \qquad (6)$$
where $L_n = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$ and $C = D^{-\frac{1}{2}} F$.
- **Nullspace Projection.** The constraint $C^T X = 0$ implies X lives in the nullspace of C^T, i.e.,
$$X \equiv PX \quad \text{where } P \text{ is an orthogonal projector onto null}(C^T).$$
This leads to the projected eigenvalue problem
$$(PL_nP)X = X\Lambda_k. \qquad (7)$$
Observation: The projected eigenvalue problem (7) shares the same solution as the constrained eigenvalue problem (6), except *it has k extra zero eigenvalues with an eigenspace* range(C) = range$(I - P)$.
- **Hotelling's deflation.** To avoid the unwanted zero eigenvalues, shift the matrix with a proper σ to
$$L_n^\sigma := PL_nP - \sigma(I - P) = P(L_n - \sigma I)P + \sigma I, \qquad (8)$$
and solve the eigenvalue problem $L_n^\sigma X = X\Lambda_k$.
Note: $\sigma = \|L_n\|_1$ for numerical stability.
- **Scalability:** The matrix-vector product $z = Pw$ for eigensolvers is done by solving sparse least squares
$$\min_z \|Cz - w\|_2$$
- **Algorithm 3: Scalable Fair Spectral Clustering (s-FairSC)**

Input: Laplacian $L = D - W \in \mathbb{R}^{n\times n}$, group indicator $F \in \mathbb{R}^{n\times(h-1)}$, shift σ
Output: a clustering of indices $1:n$ into k clusters
1: set $L_n = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$ and $C = D^{-\frac{1}{2}} F$;
2: compute the k smallest eigenvalues of L_n^σ and the corresponding eigenvectors $X \in \mathbb{R}^{n\times k}$;
3: apply k-means clustering to the rows of $H = D^{-\frac{1}{2}} X$.

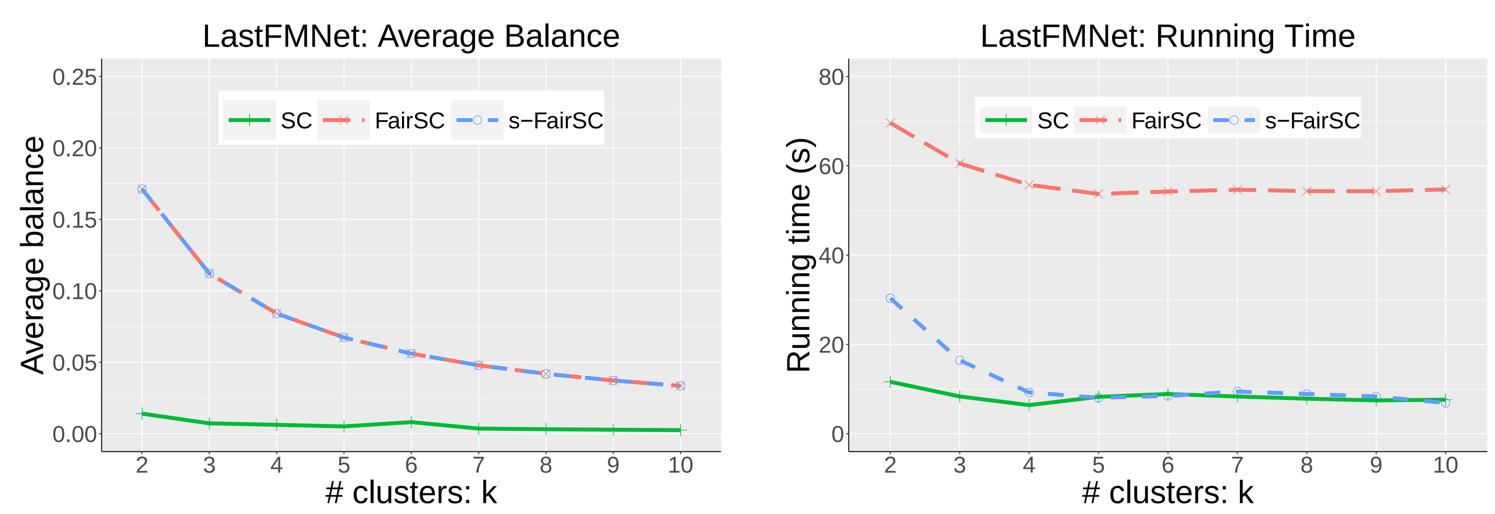
- **Time Complexity:** $\mathcal{O}(m + n(h^2 + k^2))$ where m is the number of non-zero elements of W.

Experiments

- **Modified Stochastic Block Model** (m-SBM)



- **LastFMNet** ($n = 5678$, 6 nationality groups)



Conclusions

- Our scalable fair spectral clustering algorithm (s-FairSC) combines nullspace projection and Hotelling's deflation, and it fully exploits the sparsity of the fair SC model.
- Experiments show s-FairSC improves fairness compared to SC, and is scalable in the sense that it only has a marginal increase in computational costs than SC.
- Possible extensions: (i) group overlapping; (ii) lenient group fairness; (iii) individual fairness.

References

F. Chierichetti, R. Kumar, S. Lattanzi, and S. Vassilvitskii. Fair clustering through fairlets. *Advances in Neural Information Processing Systems*, 30, 2017.

H. Hotelling. Some new methods in matrix calculation. *The Annals of Mathematical Statistics*, 14(1):1–34, 1943.

M. Kleindessner, S. Samadi, P. Awasthi, and J. Morgenstern. Guarantees for spectral clustering with fairness constraints. In *International Conference on Machine Learning*, pages 3458–3467. PMLR, 2019.

A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 14, 2001.

J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8): 888–905, 2000.

Code Repository