THE WEB CONFERENCE ACM

APRIL 30-MAY, 4, 2023

Austin, TX

Fairness-Aware Clique-Preserving Spectral Clustering of Temporal Graphs






Dongqi Fu (UIUC)

Dawei Zhou (Virginia Tech)

Ross Maciejewski (ASU)

Arie Croitoru (GMU)

Marcus Boyd (UMD)

Jingrui He (UIUC)

Presenter: Dongqi Fu (dongqif2@illinois.edu)

GitHub: https://github.com/DongqiFu/F-SEGA

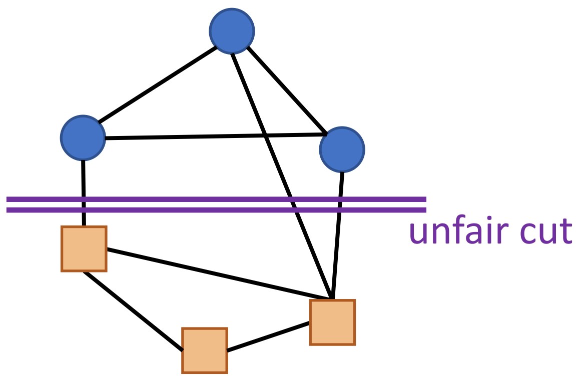
NSF ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN VT ASU UNIVERSITY OF MARYLAND 1856

Roadmap

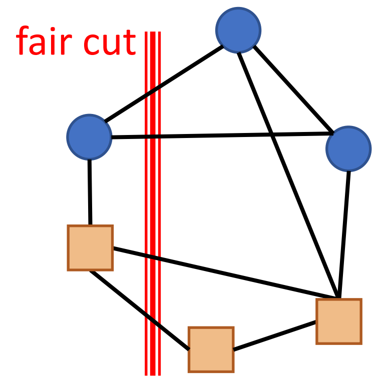
- Motivation
- Proposed F-SEGA Method
- Experiments
- Conclusion

Fairness-Aware & Clique-Preserving

- **Fairness-Aware** Clustering on Graphs [1]

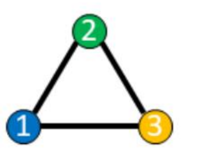


Dense but unfair
(w.r.t. node type distribution)

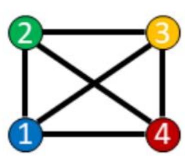


Comparable-dense but fair
(w.r.t. node type distribution)

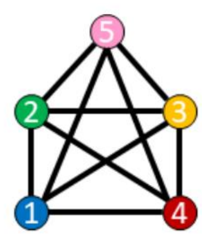
- **Clique-Preserving** Clustering on Graphs [2]



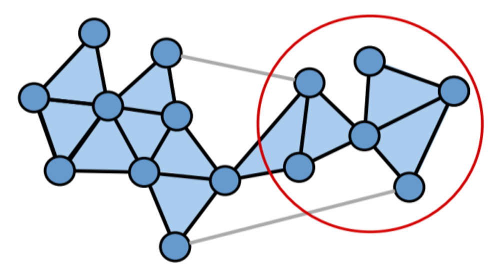
3-clique



4-clique



5-clique



Triangle-Preserving Clustering

[1] Matthäus Kleindessner, Samira Samadi, Pranjal Awasthi, Jamie Morgenstern: Guarantees for Spectral Clustering with Fairness Constraints. ICML 2019
[2] Hao Yin, Austin R. Benson, Jure Leskovec, David F. Gleich: Local Higher-Order Graph Clustering. KDD 2017

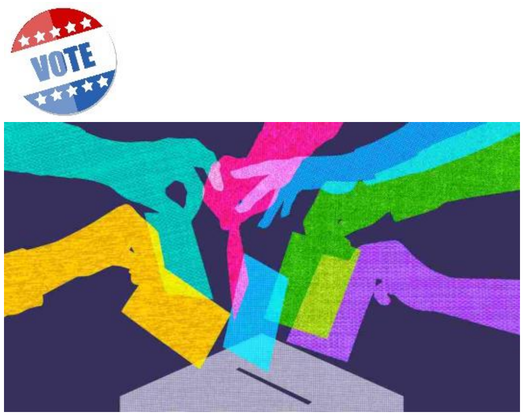
Why these two need to be co-optimized?

- k-clique community (clustering) is densely connected
- When k increases, the meaning of communities is more specialized [1], then those communities can be used for
  - Recommendation and voting (based on similar interests) [2]
  - Collaboration (based on similar expertise) [3]

[1] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. Nature, 2005
[2] Amanda L. Traud, Peter J. Mucha, and Mason A. Porter. Social structure of facebook networks. CoRR, 2011
[3] Liangyue Li and Hanghang Tong. Computational approaches to the network science of teams. 2020

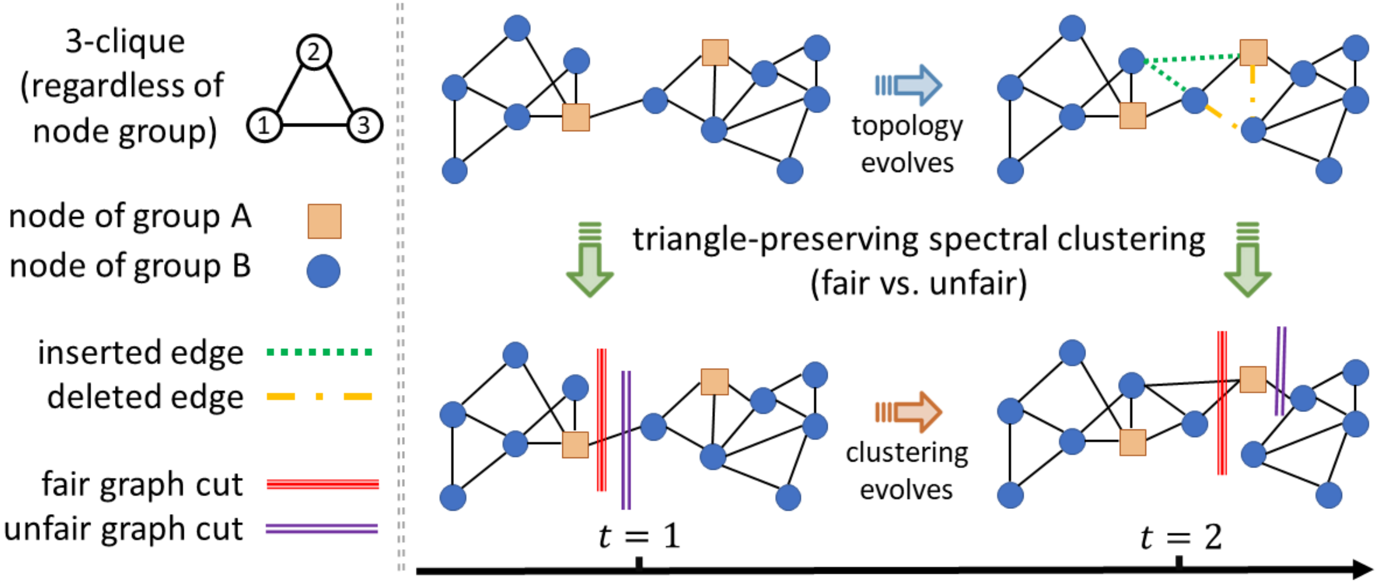
Why these two need to be co-optimized?

- However, without proportional demographics in communities
  - The voice of different groups, especially minority groups, can barely be heard when voting.
  - The team could not handle interdisciplinary tasks requiring diverse backgrounds.



When graph topology starts to evolve …

- Suppose we already have a static solution for fairness-aware and clique-preserving graph clustering algorithm.
- Will the previous fairness and high-order density be broken, when graph structure evolves?



Roadmap

- Motivation
- **Proposed F-SEGA Method**
- Experiments
- Conclusion

Problem Setting

- **Input**: Given a temporal graph $\mathcal{G} = \{G^{(1)}, G^{(2)}, \dots, G^{(T)}\}$, a number of desired clusters q, and a target k-clique
- **Output**: F-SEGA aims to identify clusters $\left\{C_1^{(t)}, C_2^{(t)}, \dots, C_q^{(t)}\right\}$ for $t \in \{1, 2, \dots, T\}$ satisfying

$$\boxed{\min_{C_i^{(t)}} \sum_{t=1}^{T} CPNcut(C_1^{(t)}, \dots, C_q^{(t)}, \mathbb{N})}$$

/* Clique-density Constraint */

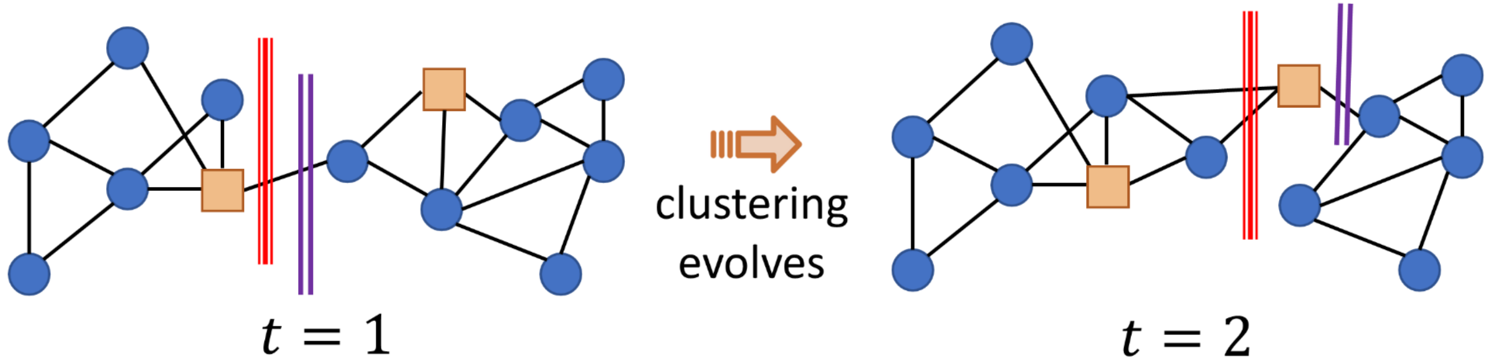
$$CPNcut(C_1, \dots, C_q, \mathbb{N}) = \sum_{i=1}^{q} \frac{cut(C_i, V \setminus C_i, \mathbb{N})}{\mu(C_i, \mathbb{N})}$$

$$\boxed{\forall s \in \{1, \dots, h\} : \frac{|V_s \cap C_i^{(t)}|}{|C_i^{(t)}|} = \frac{|V_s|}{|V|}, t \in \{1, \dots, T\}}$$

/* Demographical Fairness Constraint */

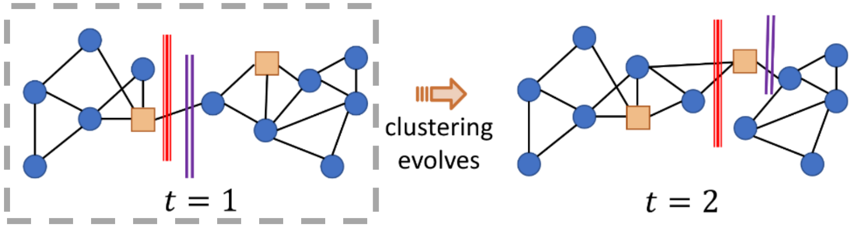
Theoretical Contribution of F-SEGA

- First, we propose **a static solution** for the fairness-aware clique-preserving spectral clustering on graphs.



- Then, we adapt this solution **to dynamic setting**, through
  - Laplacian Update via Edge Filtering and Searching
  - Eigen-Pairs Update with Singularity Avoided

Static Solution



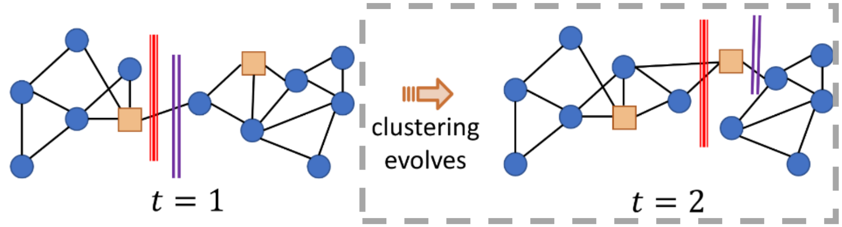
- Core Idea
  - Spectral Clustering

- Detail
  - $\mathcal{M}$ is fairness-constrained clique-weighted Laplacian matrix
  - $\mathcal{M} = \underbrace{Q^{-1}Z^{T}}_{\text{encode the demographical distribution of the entire graph}} LZQ^{-1} \in \mathbb{R}^{(n-h+1)\times(n-h+1)}$

    encodes the clique distribution of the entire graph
  - Eigen-decompose $\mathcal{M}$, get the low-rank matrix and apply the k-means [1,2] for obtaining the clustering

[1] Matthäus Kleindessner, Samira Samadi, Pranjal Awasthi, Jamie Morgenstern: Guarantees for Spectral Clustering with Fairness Constraints. ICML 2019
[2] Ulrike von Luxburg. A tutorial on spectral clustering. Stat. Comput., 2007

Laplacian Update via Edge Filtering and Searching



$$\mathcal{G} = \{G^{(1)}, G^{(2)}, \dots, G^{(T)}\}$$

$$\mathcal{M} = Q^{-1} Z^T L Z Q^{-1} \in \mathbb{R}^{(n-h+1)\times(n-h+1)}$$

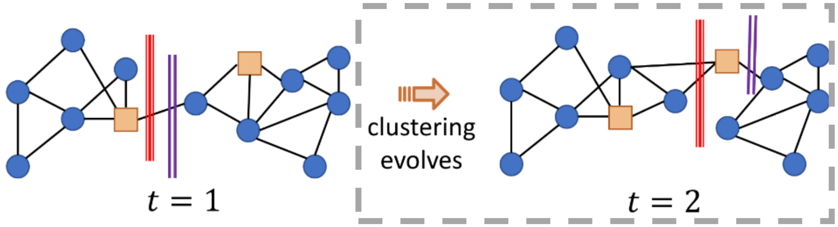
- Core Idea
  - Update $\mathcal{M}^{(t)}$

fairness-constrained clique-weighted Laplacian matrix

- Detail
  - Identify insensitive updated edges those will not change the last time clustering and ignore them
- Time Complexity
  - Given k-clique, $k \geq 2$, updating the clique-weighted adjacency matrix costs $O(k\alpha^{k-2} m^{(t)})$
    - α is the arboricity of snapshot $G^{(t)}$
    - $m^{(t)}$ is the number of edges in $G^{(t)}$

arboricity: minimum number of spanning forests needed to cover all the edges of the graph

Eigen-Pairs Update with Singularity Avoided



- Core Idea
  - Track eigen-pairs of $\mathcal{M}^{(t)}$ instead of solving it from scratch

- Detail
  - Approximate eigen-pair (λ, u) of Laplacian matrix perturbation $\Delta\mathcal{M} = \mathcal{M}^{(t+1)} - \mathcal{M}^{(t)}$

$$\lambda_i^{(t+1)} = \lambda_i^{(t)} + \Delta\lambda_i, \quad \text{s.t. } \Delta\lambda_i = \boldsymbol{u}_i^{(t)\top} \Delta\boldsymbol{M}\boldsymbol{u}_i^{(t)}$$ /* Eigenvalue update*/

$$\boldsymbol{u}_i^{(t+1)} = \boldsymbol{u}_i^{(t)} + \Delta\boldsymbol{u}_i, \quad \text{s.t. } \Delta\boldsymbol{u}_i = \sum_{j=1}^{q} \frac{\boldsymbol{u}_j^{(t)\top} \Delta\boldsymbol{M}\boldsymbol{u}_i^{(t)}}{\lambda_i^{(t)} - \lambda_j^{(t)}} \boldsymbol{u}_j^{(t)}$$ /* Eigenvector update*/

- Time Complexity
  - Given ΔM from $M^{(t)}$ to $M^{(t+1)}$, getting new eigen-pair costs $O(q^4 + nq^2)$, where q is num. of clusters and n is num. of nodes

Roadmap

- **Motivation**
- **Proposed F-SEGA Method**
- **Experiments**
- **Conclusion**

Real-World Datasets

- Highschool-2011
  - 126 nodes (male and female students)
  - 28,561 temporal edges in 4 days
- Highschool-2013
  - 327 nodes (male and female students)
  - 188,509 temporal edges in 5 days
- Hospital
  - 75 nodes (of patients, nurses, medical doctors, and administrative staff)
  - 32,424 temporal edges
- PrimarySchool
  - 232 nodes (male and female students)
  - 125,773 temporal edges
- ASA
  - 5,767 nodes (male and female employees)
  - 873,716 temporal edges in 10 years

Performance over Real-World Graphs

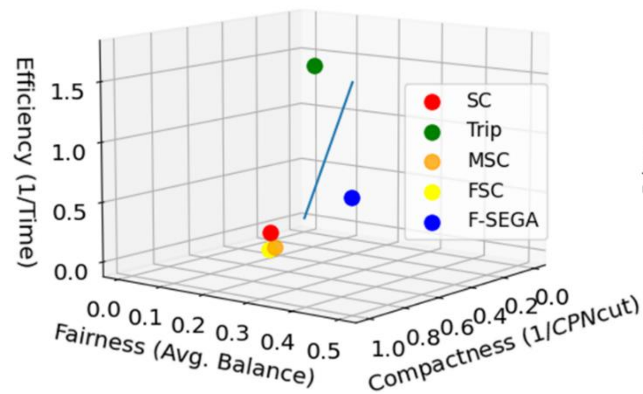
- When the distribution of input graph is not demographically fair
- When the distribution of input graph is already demographically fair

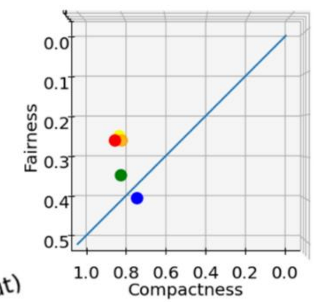
Data	HighSchool-2011 (Small Number of Clusters)			
Method \ Metric	$Ncut$ ↓	$CPNcut$ ↓	Avg. Balance ↑	Time (cs) ↓
SC	3.1389 ± 0.8599	3.0331 ± 0.9046	0.4596 ± 0.0454	9.5270 ± 2.4491
TripSC	3.9756 ± 1.0791	3.9507 ± 1.1274	0.4519 ± 0.0669	4.7160 ± 0.2390
MSC	3.1443 ± 0.8973	2.9554 ± 0.8900	0.3888 ± 0.0850	17.0819 ± 2.1950
FSC	3.4110 ± 0.7931	3.3047 ± 0.8312	0.4457 ± 0.0185	23.8289 ± 2.3470
F-SEGA	4.4525 ± 0.9885	4.4435 ± 0.9947	0.6281 ± 0.0851	15.4022 ± 0.9090

Data	HighSchool-2013 (Small Number of Clusters)			
Method \ Metric	$Ncut$ ↓	$CPNcut$ ↓	Avg. Balance ↑	Time (cs) ↓
SC	1.4866 ± 0.4334	0.6458 ± 0.2347	0.4708 ± 0.0135	33.2589 ± 2.3160
TripSC	1.7915 ± 0.2823	1.0755 ± 0.5641	0.4531 ± 0.0182	27.5309 ± 0.4920
MSC	1.4664 ± 0.4205	0.6483 ± 0.1829	0.4695 ± 0.0139	63.0641 ± 2.2970
FSC	1.5203 ± 0.4895	0.6620 ± 0.2860	0.5160 ± 0.0466	52.8459 ± 2.5430
F-SEGA	1.5296 ± 0.3493	0.6728 ± 0.1800	0.4620 ± 0.0058	23.1415 ± 0.1730

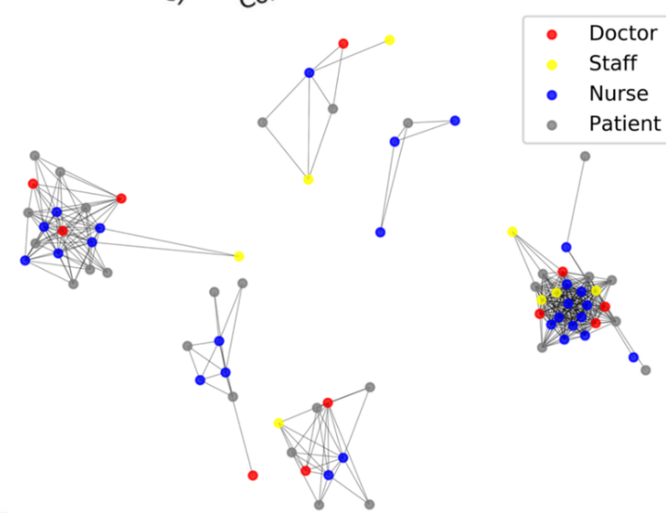
Visualization

- Comprehensiveness, i.e., trade-off among
  - Fairness
  - Density
  - Efficiency





- Case Study
  - Proportional human resource allocation in the hospital graph



Roadmap

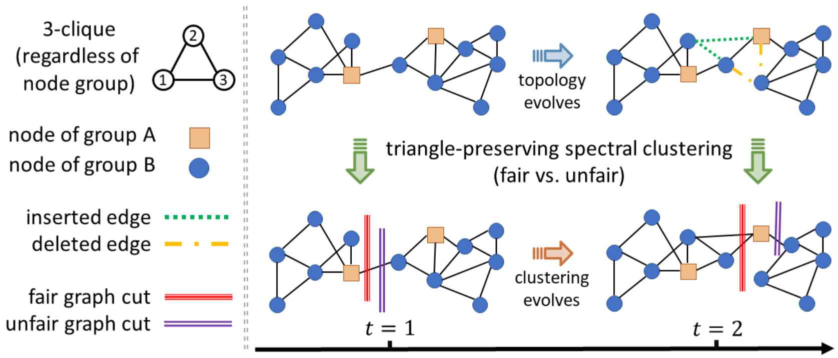
- Motivation
- Proposed F-SEGA Method
- Experiments
- Conclusion

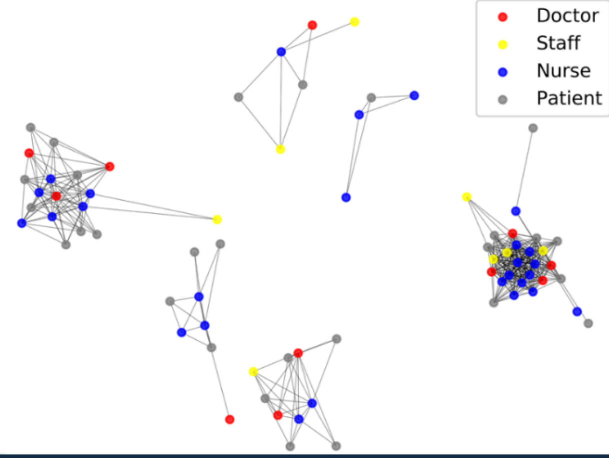
Conclusion

- Problem: Fairness-Aware Clique-Preserving Spectral Clustering of Temporal Graphs

- Algorithm: F-SEGA
  - Static Solution + Dynamic Update
  - Bounded time complexity
  - Easy to code

- Evaluation
  - Effectiveness
  - Efficiency and Robustness
  - Case Study
  - Ablation Studies





THE WEB CONFERENCE ACM

APRIL 30-MAY, 4, 2023

Austin, TX

Thanks!

					
Dongqi Fu (UIUC)	**Dawei Zhou (Virginia Tech)**	**Ross Maciejewski (ASU)**	**Arie Croitoru (GMU)**	**Marcus Boyd (UMD)**	**Jingrui He (UIUC)**

Presenter: Dongqi Fu (dongqif2@illinois.edu)

GitHub: https://github.com/DongqiFu/F-SEGA