Local Correlation Clustering with Asymmetric Classification Errors

Jafar Jafarov
University of Chicago

Joint work with

Sanchit Kalhan, Northwestern U.
Konstantin Makarychev, Northwestern U.
Yury Makarychev, TTIC

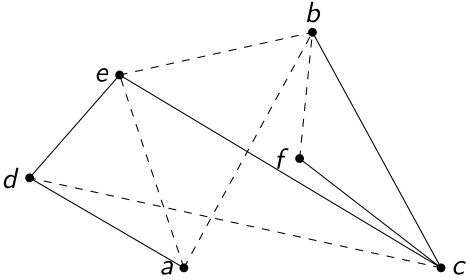
ICML 2021

Correlation Clustering

- Introduced by Bansal, Blum, and Chawla [2004]
- Many applications in Machine Learning
  - Image Segmentation (Wirth [2010])
  - Spam Detection (Bonchi et al. [2014], Ramachandran et al. [2007])
  - Coreference Resolution (Cohen and Richman [2001, 2002])
  - Multi-Person Tracking (Tang et al. [2016, 2017])
  - Data Mining (Filkov and Skiena [2003])
  - ...

Problem Definition

- Input: $G = (V, E)$, $weight : E \to \mathbb{R}_{\geq 0}$, $label : E \to \{+, -\}$
- Output: Clustering $\mathcal{C}$ of the vertex set V

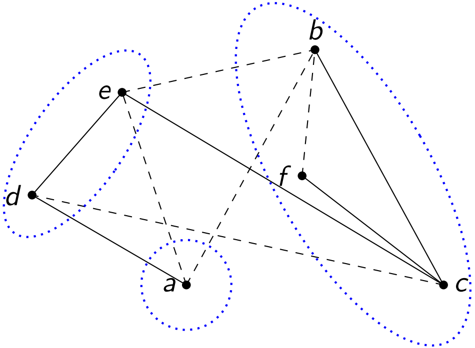


$-$: dashed line

$+$: solid line

Problem Definition

- Input: $G = (V, E)$, *weight* : $E \to \mathbb{R}_{\geq 0}$, *label* : $E \to \{+, -\}$
- Output: Clustering $\mathcal{C}$ of the vertex set V



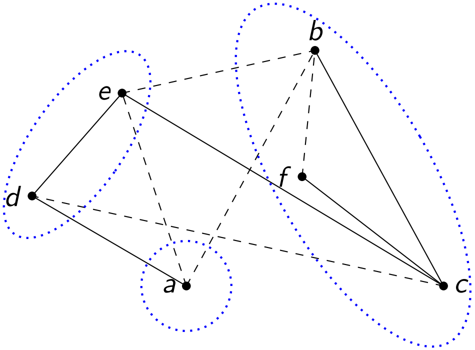
$-$: dashed line

$+$: solid line

$\mathcal{C} = \{\{e, d\}, \{a\}, \{b, f, c\}\}$

Problem Definition

- $(u, v) \in E^+$ is in disagreement with $\mathcal{C}$ if $\mathcal{C}(u) \neq \mathcal{C}(v)$.
-



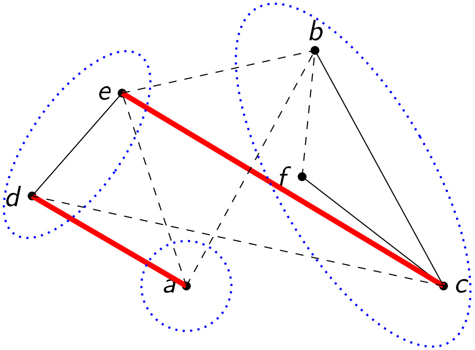
$-$: dashed line

$+$: solid line

$$\mathcal{C} = \{\{e, d\}, \{a\}, \{b, f, c\}\}$$

Problem Definition

- $(u, v) \in E^+$ is in disagreement with $\mathcal{C}$ if $\mathcal{C}(u) \neq \mathcal{C}(v)$.
-



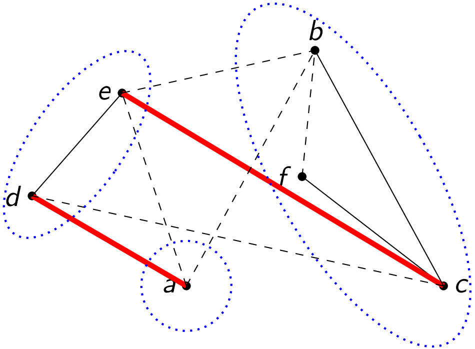
$-$: dashed line

$+$: solid line

$$\mathcal{C} = \{\{e, d\}, \{a\}, \{b, f, c\}\}$$

Problem Definition

- $(u, v) \in E^+$ is in disagreement with $\mathcal{C}$ if $\mathcal{C}(u) \neq \mathcal{C}(v)$.
- $(u, v) \in E^-$ is in disagreement with $\mathcal{C}$ if $\mathcal{C}(u) = \mathcal{C}(v)$.



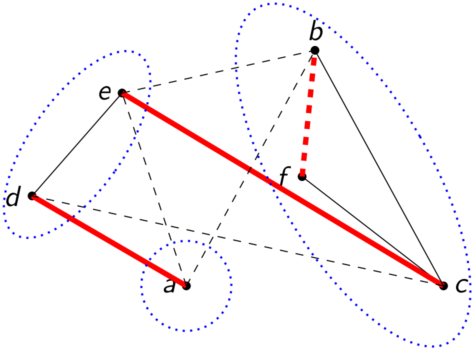
$-$: dashed line

$+$: solid line

$$\mathcal{C} = \{\{e, d\}, \{a\}, \{b, f, c\}\}$$

Problem Definition

- $(u, v) \in E^+$ is in disagreement with $\mathcal{C}$ if $\mathcal{C}(u) \neq \mathcal{C}(v)$.
- $(u, v) \in E^-$ is in disagreement with $\mathcal{C}$ if $\mathcal{C}(u) = \mathcal{C}(v)$.



$-$: dashed line

$+$: solid line

$$\mathcal{C} = \{\{e, d\}, \{a\}, \{b, f, c\}\}$$

Local Objectives in Correlation Clustering

- Let the disagreements vector be a vector indexed by the vertices of G.

Local Objectives in Correlation Clustering

- Let the disagreements vector be a vector indexed by the vertices of G.
- Given a clustering $\mathcal{P}$ for each vertex $u \in V$,

$$\mathrm{dis}_u(\mathcal{P}) = \sum_{(u,v)\in E} w_{uv} \cdot 1\{(u,v) \text{ is in disagreement with } \mathcal{P}\}.$$

Local Objectives in Correlation Clustering

- Let the disagreements vector be a vector indexed by the vertices of G.
- Given a clustering $\mathcal{P}$ for each vertex $u \in V$,

$$\mathrm{dis}_u(\mathcal{P}) = \sum_{(u,v) \in E} w_{uv} \cdot 1\{(u, v) \text{ is in disagreement with } \mathcal{P}\}.$$

- ℓ_p objective is to find a clustering $\mathcal{P}$ that minimizes the ℓ_p-norm of the disagreements vector:

$$\min \left(\sum_{u \in V} |\mathrm{dis}_u(\mathcal{P})|^p \right)^{\frac{1}{p}}.$$

Local Objectives in Correlation Clustering II

- ℓ_1 objective is equivalent to minimizing the total weight of disagreements

Local Objectives in Correlation Clustering II

- ℓ_1 objective is equivalent to minimizing the total weight of disagreements, the **MinDisagree** objective.

Local Objectives in Correlation Clustering II

- ℓ_1 objective is equivalent to minimizing the total weight of disagreements, the **MinDisagree** objective.

- ℓ_∞ objective is equivalent to minimizing the weight of disagreements at the vertex that is worst off

Local Objectives in Correlation Clustering II

- ℓ_1 objective is equivalent to minimizing the total weight of disagreements, the **MinDisagree** objective.
- ℓ_∞ objective is equivalent to minimizing the weight of disagreements at the vertex that is worst off, **Min Max Correlation Clustering.**

Local Objectives in Correlation Clustering II

- ℓ_1 objective is equivalent to minimizing the total weight of disagreements, the **MinDisagree** objective.
- ℓ_∞ objective is equivalent to minimizing the weight of disagreements at the vertex that is worst off, **Min Max Correlation Clustering.**
- ℓ_1 objective is a global objective.

Local Objectives in Correlation Clustering II

- ℓ_1 objective is equivalent to minimizing the total weight of disagreements, the **MinDisagree** objective.
- ℓ_∞ objective is equivalent to minimizing the weight of disagreements at the vertex that is worst off, **Min Max Correlation Clustering.**
- ℓ_1 objective is a global objective.
- For higher values of p, ℓ_p objective becomes a local objective.

Known Results: Complete Unweighted Graph

ℓ_1 objective

Approximation Ratio	
≈ 20000	Bansal, Blum, and Chawla [2004]
4	Charikar, Guruswami, and Wirth [2003]
3 and 2.5	Ailon, Charikar, and Newman [2008]
2.06	Chawla, Makarychev, Schramm, and Yaroslavtsev [2015]
Integrality Gap	
2	Charikar, Guruswami, and Wirth [2003]

Known Results: Complete Unweighted Graph

ℓ_p objective

Approximation Ratio	
48	Puleo and Milenkovic [2018]
7	Charikar, Gupta, and Schwartz [2017]
5	Kalhan, Makarychev, and Zhou [2019]

Known Results: Arbitrary Weighted Graph

ℓ_1 objective

Approximation Ratio	
$O(\log n)$	Charikar, Guruswami, and Wirth [2003]; Demaine, Emanuel, Fiat, and Immorlica [2006]
Integrality Gap	
$O(\log n)$	Charikar, Guruswami, and Wirth [2003]; Demaine, Emanuel, Fiat, and Immorlica [2006]

Known Results: Arbitrary Weighted Graph

ℓ_p objective

Approximation Ratio	
$O(\sqrt{n})$ (for $p = \infty$)	Charikar, Gupta, and Schwartz [2017]
$O\left(n^{\frac{1}{2}-\frac{1}{2p}} \cdot (\log n)^{\frac{1}{2}+\frac{1}{2p}}\right)$	Kalhan, Makarychev, and Zhou [2019]
Integrality Gap	
$\Omega\left(n^{\frac{1}{2}-\frac{1}{2p}}\right)$	Kalhan, Makarychev, and Zhou [2019]

Correlation Clustering with Asymmetric Classification Errors

Jafarov, Kalhan, Makarychev, and Makarychev [2020]

- Let G be a complete graph, $\alpha \in (0, 1]$ and $\boldsymbol{w} > 0$ a scaling parameter.
- For every positive edge $e \in E^+$ we have $\boldsymbol{w_e} \in [\alpha \boldsymbol{w}, \boldsymbol{w}]$
- For every negative edge $e \in E^-$ we have $\boldsymbol{w_e} \in [\alpha \boldsymbol{w}, \infty)$

Correlation Clustering with Asymmetric Classification Errors

Jafarov, Kalhan, Makarychev, and Makarychev [2020]

- Let G be a complete graph, $\alpha \in (0,1]$ and $\boldsymbol{w} > 0$ a scaling parameter.
- For every positive edge $e \in E^+$ we have $\boldsymbol{w_e} \in [\alpha \boldsymbol{w}, \boldsymbol{w}]$
- For every negative edge $e \in E^-$ we have $\boldsymbol{w_e} \in [\alpha \boldsymbol{w}, \infty)$
- $3 + 2 \ln \frac{1}{\alpha}$ approximation for the ℓ_1 objective (Jafarov et al. [2020])

Main Result

Main Theorem

There exists a polynomial-time $O\left((\frac{1}{\alpha})^{\frac{1}{2}-\frac{1}{2p}} \cdot \log \frac{1}{\alpha}\right)$*-approximation algorithm for minimizing the* ℓ_p *objective in the Correlation Clustering with Asymmetric Classification Errors model.*

Main Result

Main Theorem

There exists a polynomial-time $O\left((\frac{1}{\alpha})^{\frac{1}{2}-\frac{1}{2p}} \cdot \log \frac{1}{\alpha}\right)$*-approximation algorithm for minimizing the* ℓ_p *objective in the Correlation Clustering with Asymmetric Classification Errors model.*

- $p = 1$: We get $O(\log \frac{1}{\alpha})$ approximation

Main Result

Main Theorem

There exists a polynomial-time $O\left((\frac{1}{\alpha})^{\frac{1}{2}-\frac{1}{2p}} \cdot \log \frac{1}{\alpha}\right)$*-approximation algorithm for minimizing the* ℓ_p *objective in the Correlation Clustering with Asymmetric Classification Errors model.*

- $p = 1$: We get $O(\log \frac{1}{\alpha})$ approximation
- $p = 2$: We get $\tilde{O}\left((1/\alpha)^{1/4}\right)$ approximation

Main Result

Main Theorem

There exists a polynomial-time $O\left((\frac{1}{\alpha})^{\frac{1}{2}-\frac{1}{2p}} \cdot \log\frac{1}{\alpha}\right)$*-approximation algorithm for minimizing the* ℓ_p *objective in the Correlation Clustering with Asymmetric Classification Errors model.*

- $p = 1$: We get $O(\log\frac{1}{\alpha})$ approximation
- $p = 2$: We get $\tilde{O}\left((1/\alpha)^{1/4}\right)$ approximation
- $p = \infty$: We get $\tilde{O}\left(\sqrt{1/\alpha}\right)$ approximation

Main Result

Main Theorem

There exists a polynomial-time $O\left(\left(\frac{1}{\alpha}\right)^{\frac{1}{2}-\frac{1}{2p}} \cdot \log \frac{1}{\alpha}\right)$*-approximation algorithm for minimizing the* ℓ_p *objective in the Correlation Clustering with Asymmetric Classification Errors model.*

- $p = 1$: We get $O(\log \frac{1}{\alpha})$ approximation
- $p = 2$: We get $\tilde{O}\left((1/\alpha)^{1/4}\right)$ approximation $\ll \tilde{O}\left(n^{1/4}\right)$ when $1/\alpha \ll n$.
- $p = \infty$: We get $\tilde{O}\left(\sqrt{1/\alpha}\right)$ approximation $\ll O\left(\sqrt{n}\right)$ when $1/\alpha \ll n$.

Main Result

Main Theorem

There exists a polynomial-time $O\left(\left(\frac{1}{\alpha}\right)^{\frac{1}{2}-\frac{1}{2p}} \cdot \log\frac{1}{\alpha}\right)$*-approximation algorithm for minimizing the* ℓ_p *objective in the Correlation Clustering with Asymmetric Classification Errors model.*

- $p = 1$: We get $O(\log\frac{1}{\alpha})$ approximation
- $p = 2$: We get $\tilde{O}\left((1/\alpha)^{1/4}\right)$ approximation $\ll \tilde{O}\left(n^{1/4}\right)$ when $1/\alpha \ll n$.
- $p = \infty$: We get $\tilde{O}\left(\sqrt{1/\alpha}\right)$ approximation $\ll O\left(\sqrt{n}\right)$ when $1/\alpha \ll n$.

Thank you!

Nir Ailon, Moses Charikar, and Alantha Newman. Aggregating inconsistent information: ranking and clustering. Journal of the ACM (JACM), 55(5):23, 2008.

Nikhil Bansal, Avrim Blum, and Shuchi Chawla. Correlation clustering. Machine learning, 56(1-3):89–113, 2004.

Francesco Bonchi, David García-Soriano, and Edo Liberty. Correlation clustering: from theory to practice. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, page 1972, 2014.

Moses Charikar, Venkatesan Guruswami, and Anthony Wirth. Clustering with qualitative information. In IEEE Symposium on Foundations of Computer Science. Citeseer, 2003.

Moses Charikar, Neha Gupta, and Roy Schwartz. Local guarantees in graph cuts and clustering. In Proceedings of the Conference on Integer Programming and Combinatorial Optimization, pages 136–147, 2017.

Shuchi Chawla, Konstantin Makarychev, Tselil Schramm, and Grigory Yaroslavtsev. Near optimal LP rounding algorithm for correlation clustering on complete and complete k-partite graphs. In Proceedings of the Symposium on Theory of Computing, pages 219–228, 2015.

William Cohen and Jacob Richman. Learning to match and cluster entity names. In Proceedings of the ACM SIGIR-2001 Workshop on Mathematical/Formal Methods in Information Retrieval, 2001.

William W Cohen and Jacob Richman. Learning to match and cluster large high-dimensional data sets for data integration. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 475–480, 2002.

Erik D Demaine, Dotan Emanuel, Amos Fiat, and Nicole Immorlica. Correlation clustering in general weighted graphs. Theoretical Computer Science, 361(2-3):172–187, 2006.

V. Filkov and S. Skiena. Integrating microarray data by consensus clustering. In Proceedings. 15th IEEE International Conference on Tools with Artificial Intelligence, pages 418–426, 2003. doi: 10.1109/TAI.2003.1250220.

J. Jafarov, Sanchit Kalhan, Konstantin Makarychev, and Yury

Makarychev. Correlation clustering with asymmetric classification errors. In Submission, 2020.

Sanchit Kalhan, Konstantin Makarychev, and Timothy Zhou. Improved algorithms for correlation clustering with local objectives. CoRR, abs/1902.10829, 2019. URL http://arxiv.org/abs/1902.10829.

G. J. Puleo and O. Milenkovic. Correlation clustering and biclustering with locally bounded errors. IEEE Transactions on Information Theory, 64 (6):4105–4119, 2018.

Anirudh Ramachandran, Nick Feamster, and Santosh Vempala. Filtering spam with behavioral blacklisting. In Proceedings of the Conference on Computer and Communications Security, pages 342–351, 2007.

Siyu Tang, Bjoern Andres, Mykhaylo Andriluka, and Bernt Schiele. Multi-person tracking by multicut and deep matching. In European Conference on Computer Vision, pages 100–111, 2016.

Siyu Tang, Mykhaylo Andriluka, Bjoern Andres, and Bernt Schiele. Multiple people tracking by lifted multicut and person re-identification. In Proceedings of the Conference on Computer Vision and Pattern Recognition, pages 3539–3548, 2017.

Anthony Wirth. Correlation clustering. In Encyclopedia of Machine Learning, pages 227–231. Springer, 2010.