Discovering Opinion Intervals from Conflicts in Signed Graphs

Peter Blohm*, Florian Chen*,
Aristides Gionis, Stefan Neumann




Aalto University

UNIVERSITY OF OXFORD

KTH VETENSKAP OCH KONST

TU WIEN

NeurIPS 2025

European Research Council

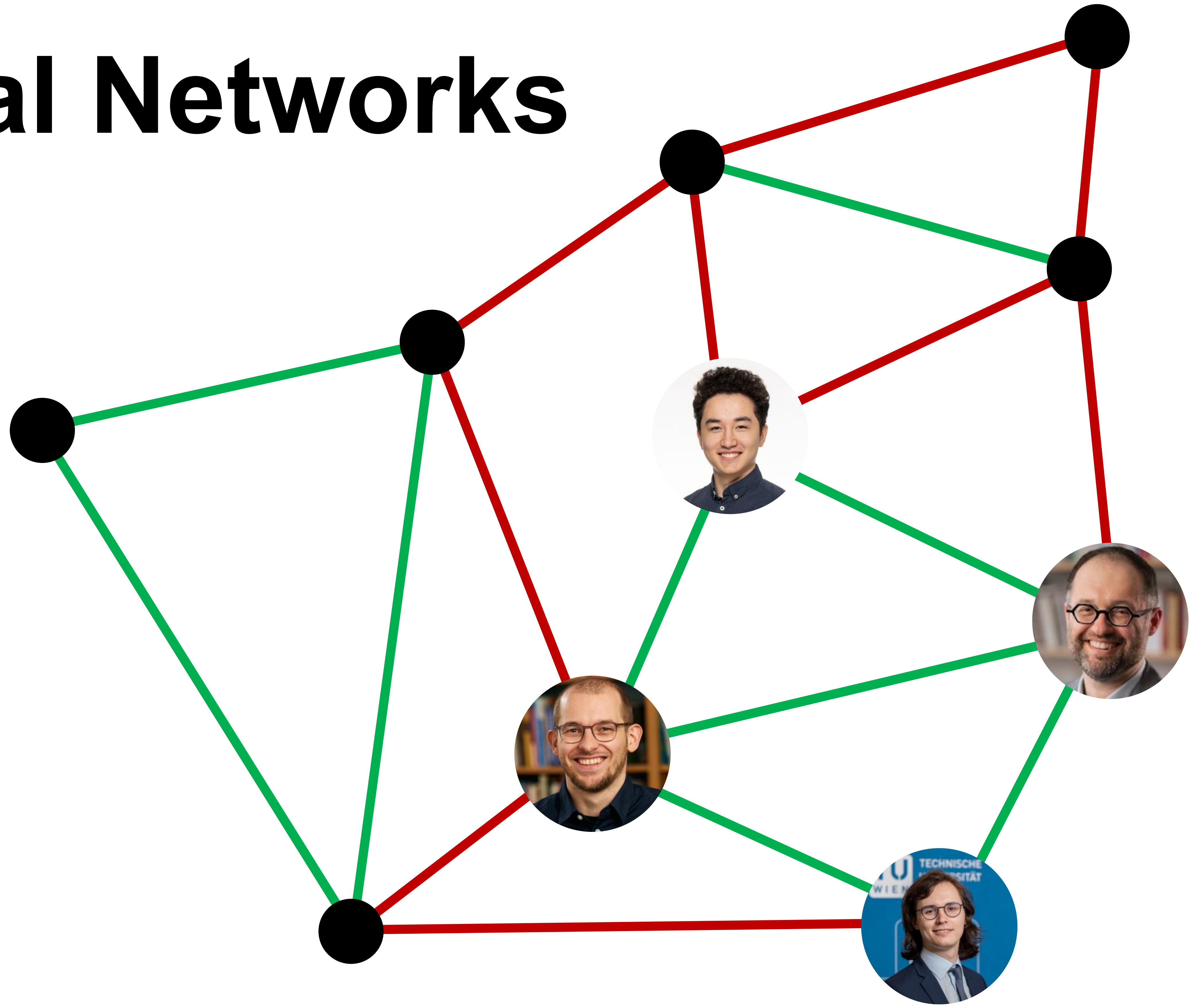
WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

W|W|T|F

Interactions in Social Networks

San Diego Sightseeing Preferences

- We observe a social network modelled as a **signed** graph

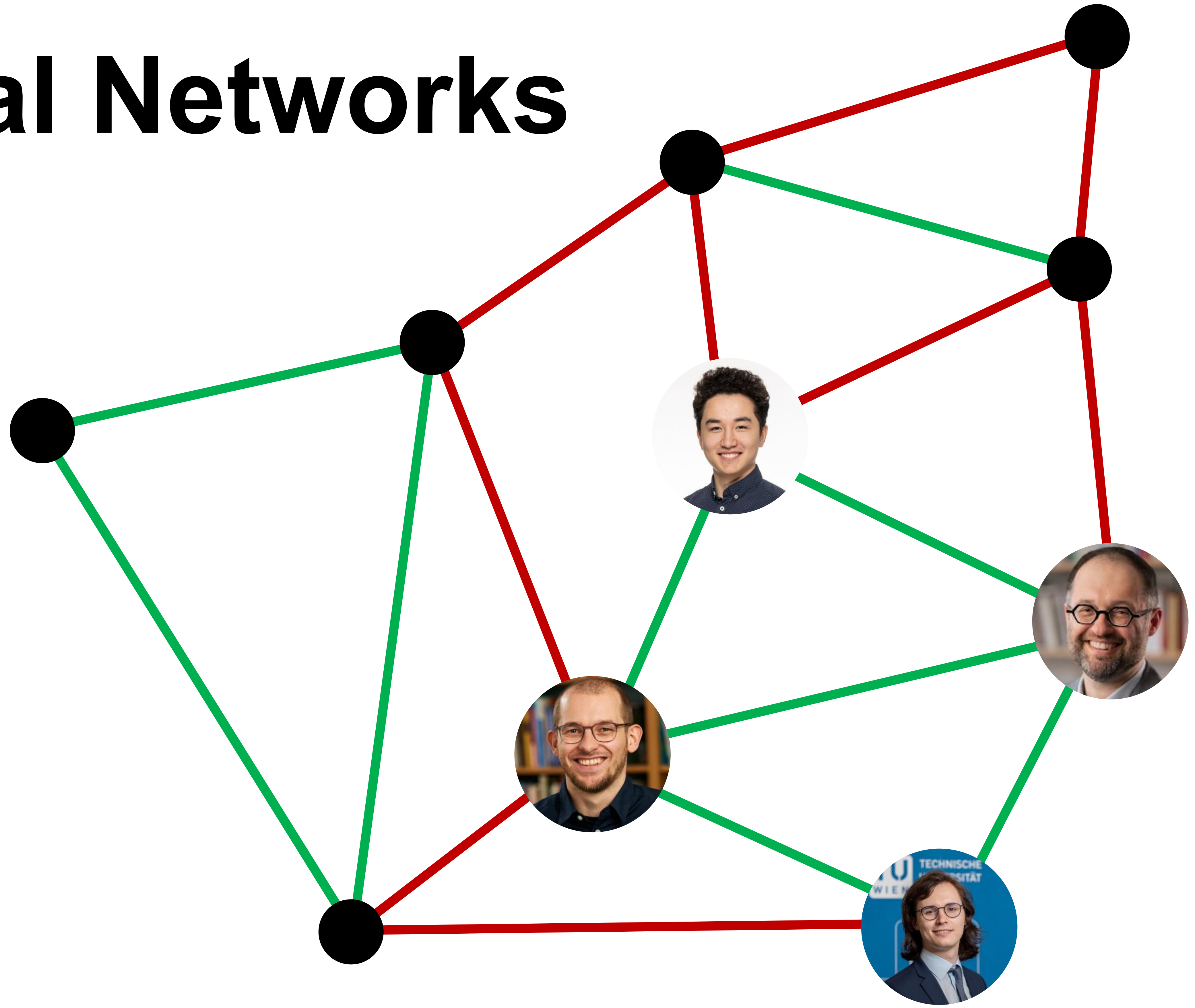


Interactions in Social Networks

San Diego Sightseeing Preferences

- We observe a social network modelled as a **signed** graph

- Each edge has a **sign** + or – for
  - Positive + interactions or
  - Negative – interactions



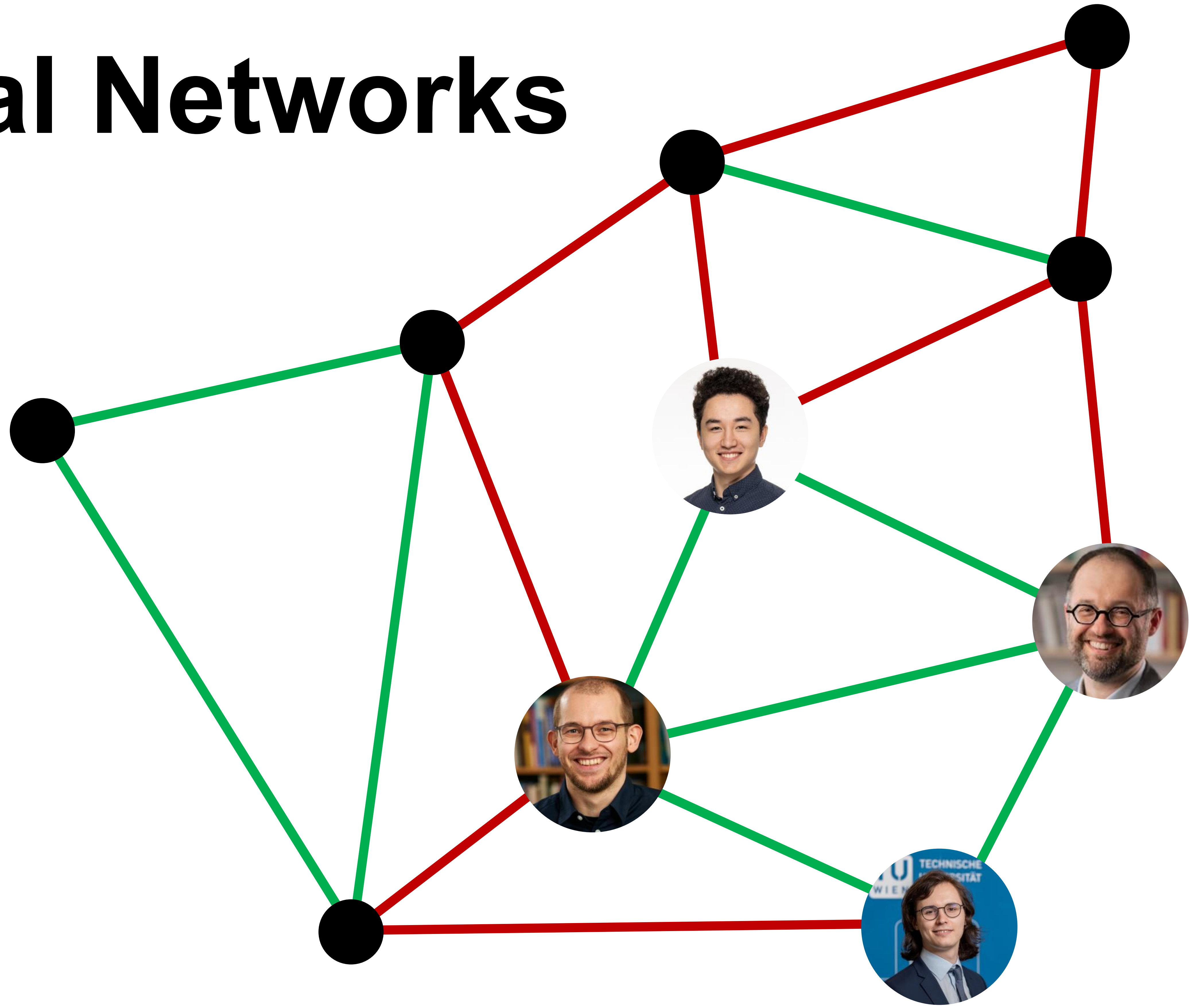
Interactions in Social Networks

San Diego Sightseeing Preferences

- We observe a social network modelled as a **signed** graph

- Each edge has a **sign** + or – for
  - Positive + interactions or
  - Negative – interactions

- **Goal:** Discover individuals' opinions



Interactions in Social Networks

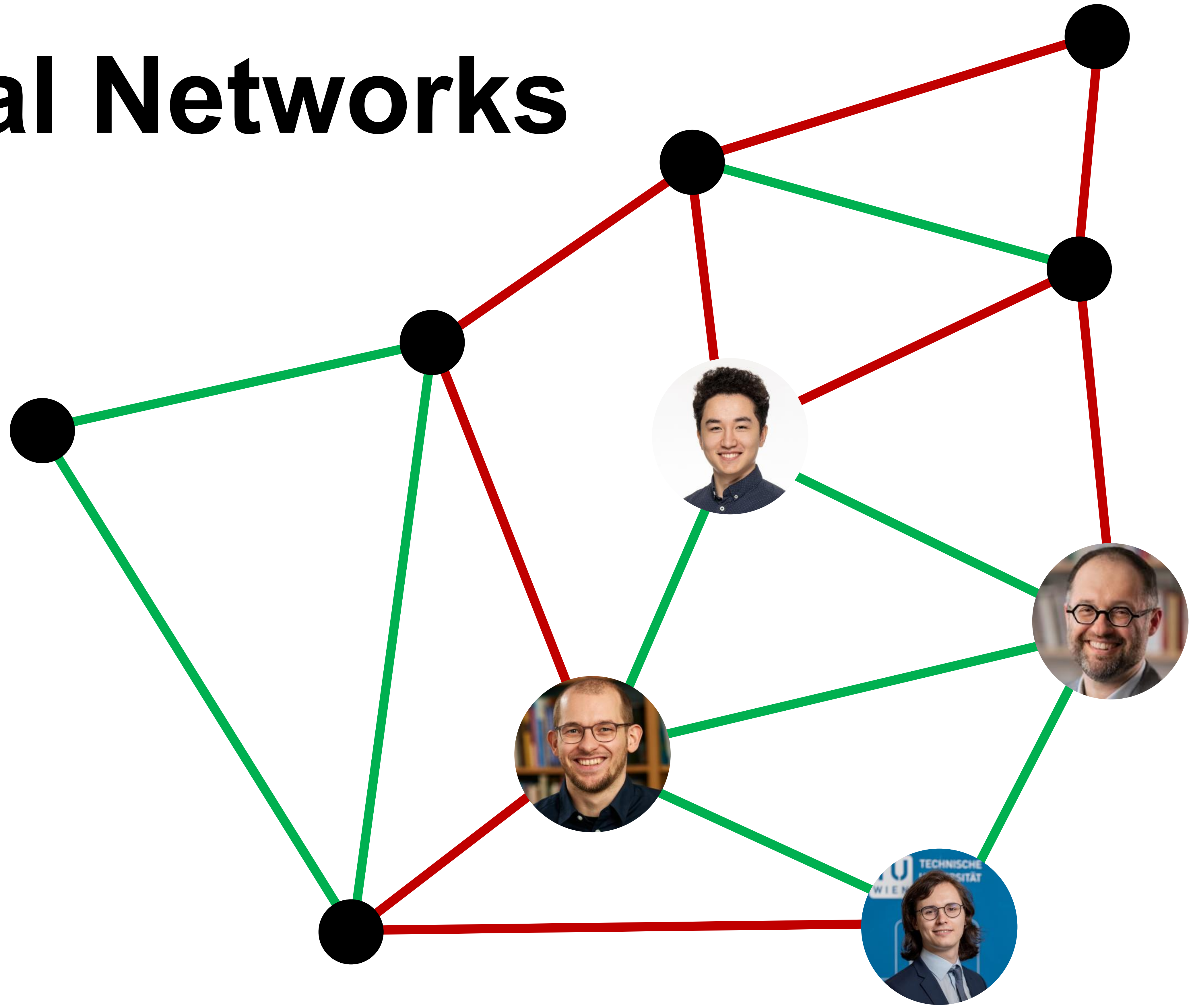
San Diego Sightseeing Preferences

CORRELATION CLUSTERING:

Assign each vertex $v \in V$ a cluster label $\ell_v \in \mathbb{N}$ to maximize the number of

(1) $\{u, v\} \in E^+$ for which $\ell_u = \ell_v$

(2) $\{u, v\} \in E^-$ for which $\ell_u \neq \ell_v$



Interactions in Social Networks

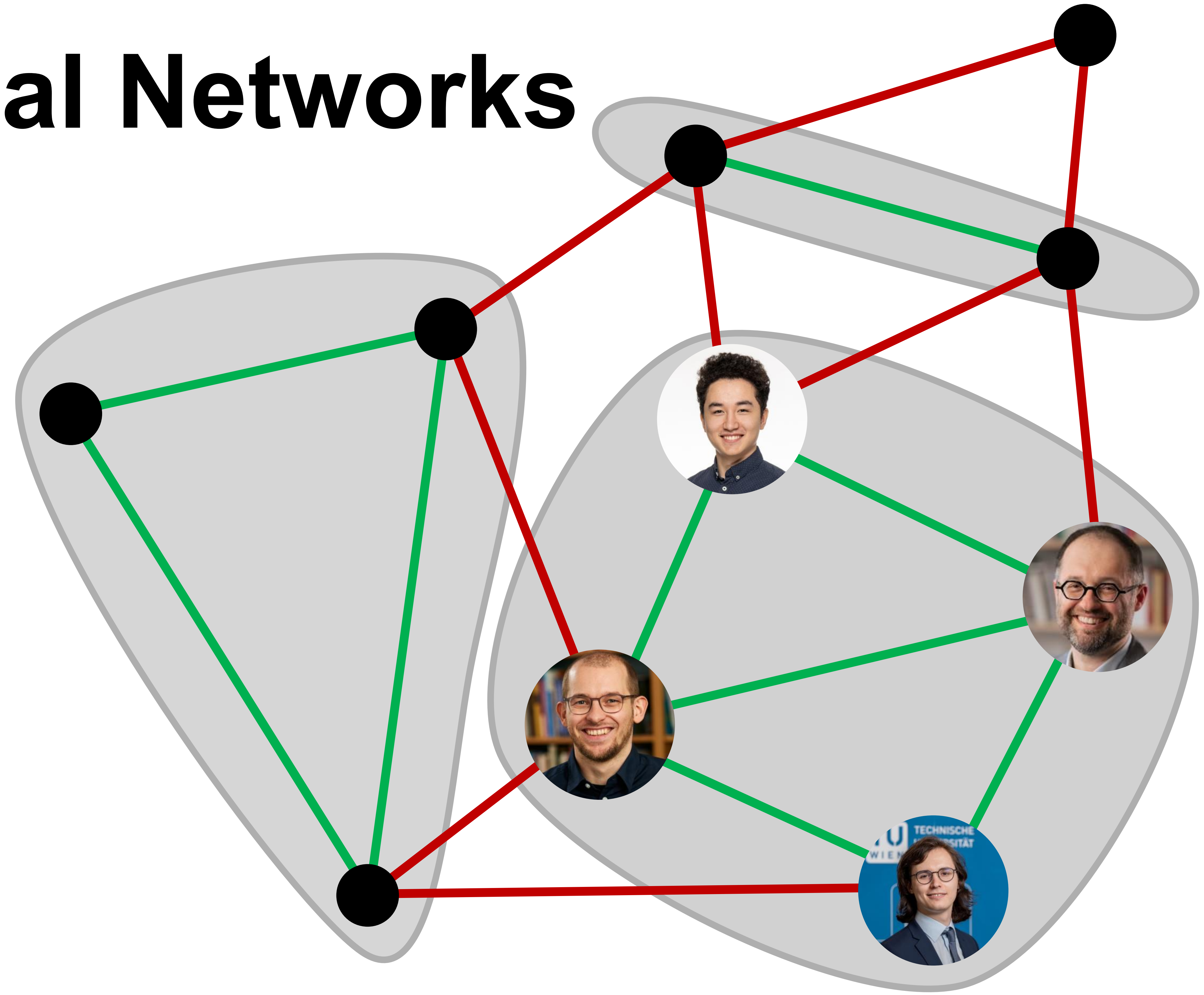
San Diego Sightseeing Preferences

CORRELATION CLUSTERING:

Assign each vertex $v \in V$ a cluster label $\ell_v \in \mathbb{N}$ to maximize the number of

(1) $\{u, v\} \in E^+$ for which $\ell_u = \ell_v$

(2) $\{u, v\} \in E^-$ for which $\ell_u \neq \ell_v$



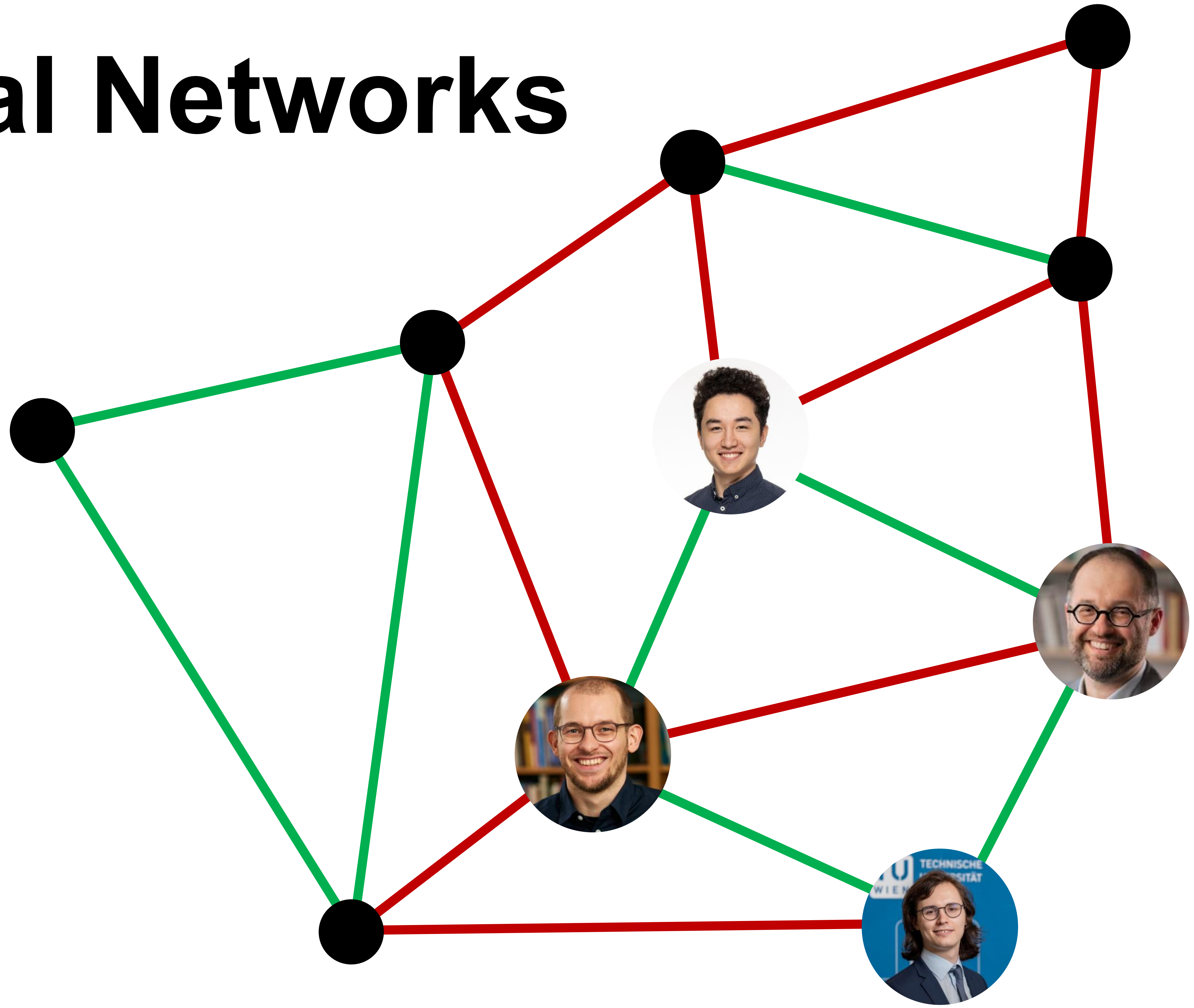
Interactions in Social Networks

San Diego Food Preferences

CORRELATION CLUSTERING:
Assign each vertex $v \in V$ a cluster label $\ell_v \in \mathbb{N}$ to maximize the number of

(1) $\{u, v\} \in E^+$ for which $\ell_u = \ell_v$

(2) $\{u, v\} \in E^-$ for which $\ell_u \neq \ell_v$



Interactions in Social Networks

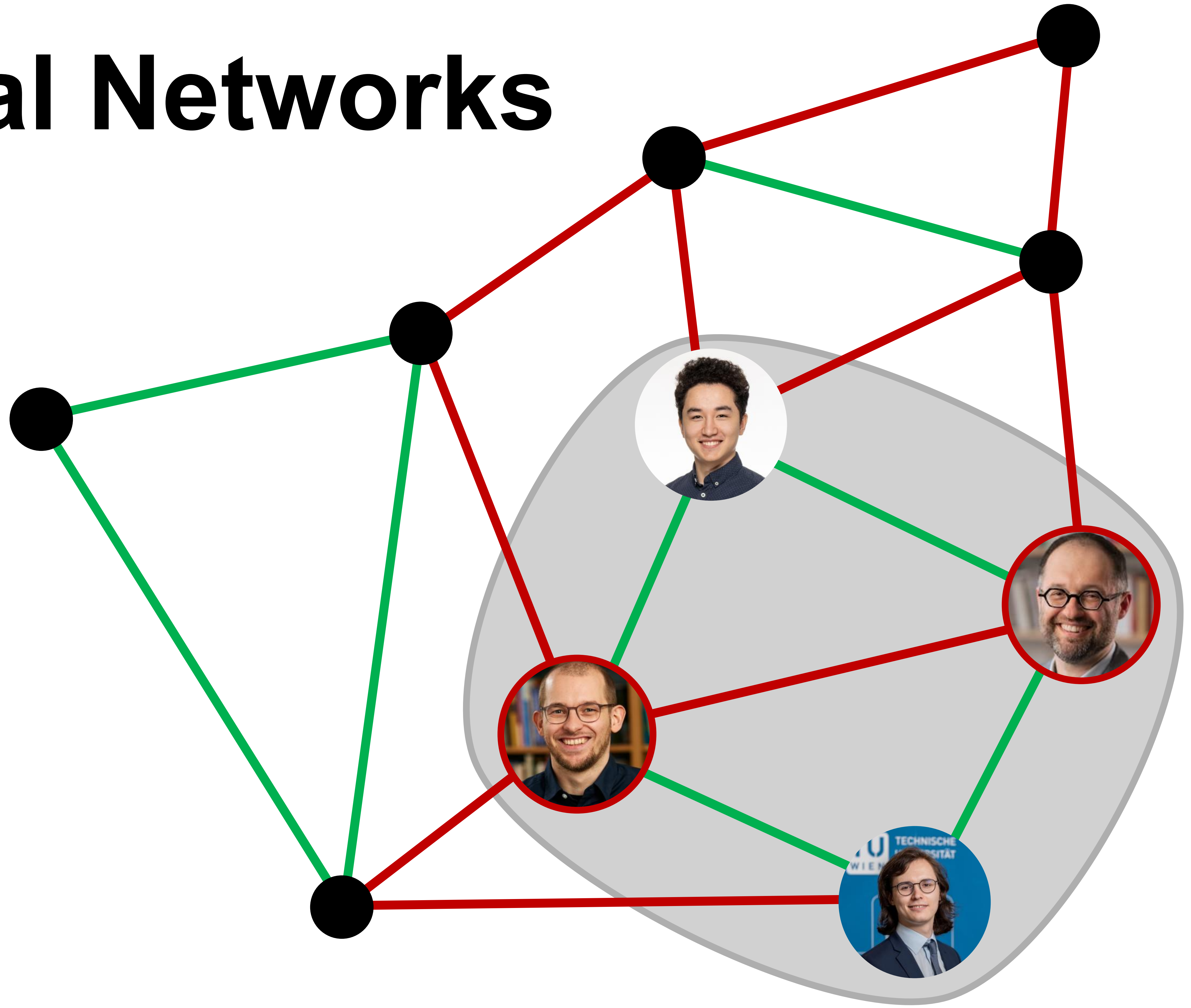
San Diego Food Preferences

Correlation Clustering:
Assign each vertex $v \in V$ a cluster label $\ell_v \in \mathbb{N}$ to maximize the number of

(1) $\{u, v\} \in E^+$ for which $\ell_u = \ell_v$

(2) $\{u, v\} \in E^-$ for which $\ell_u \neq \ell_v$

Problem: Disjoint clusters cannot model complex node interactions



Interactions in Social Networks

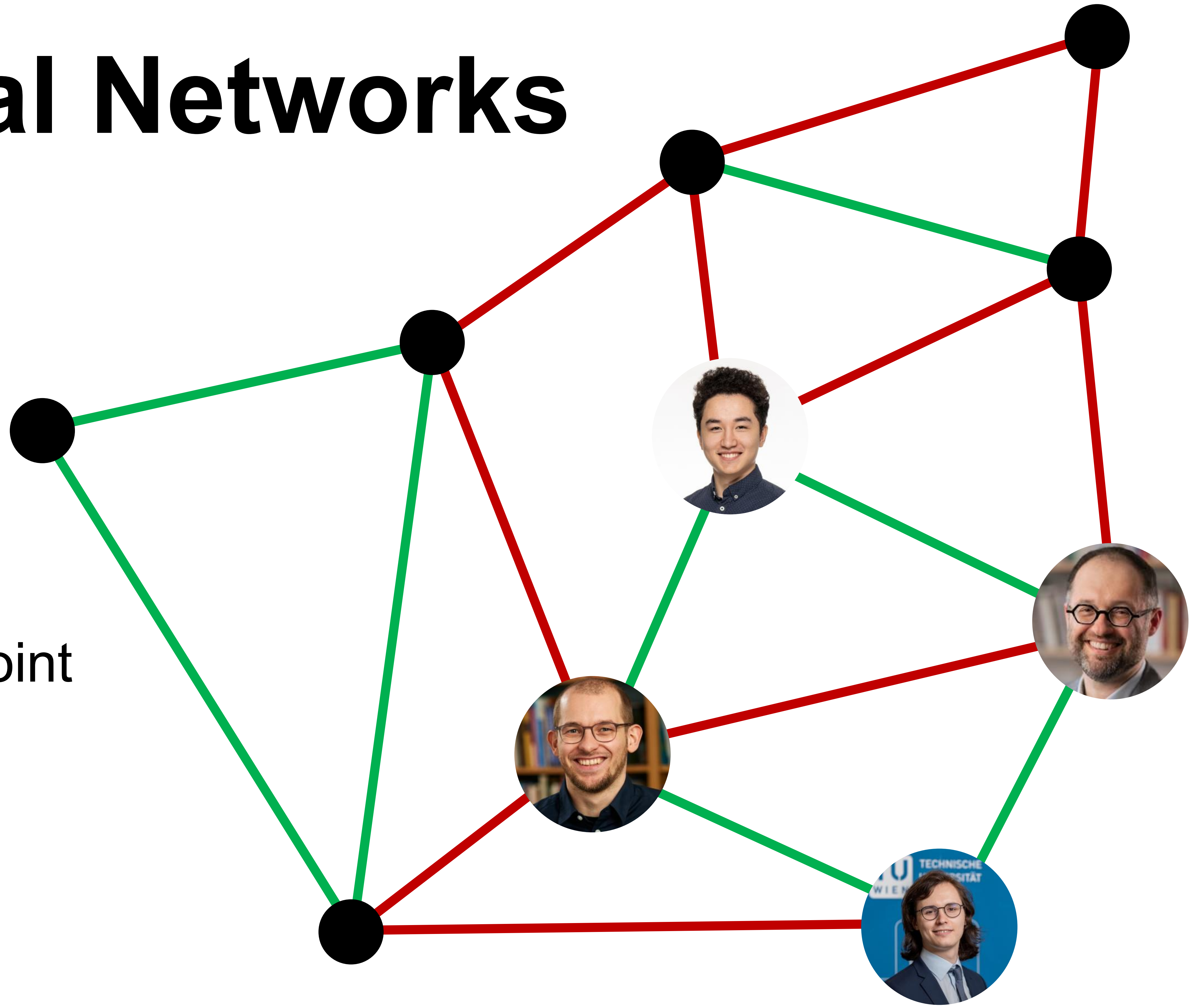
San Diego Food Preferences

Best Interval Approximation:

Assign each vertex an interval $\{I_v : v \in V\}$ with $I_v \subset \mathbb{R}$ to maximize the number of

(1) $\{u, v\} \in E^+$ for which I_u and I_v overlap

(2) $\{u, v\} \in E^-$ for which I_u and I_v are disjoint



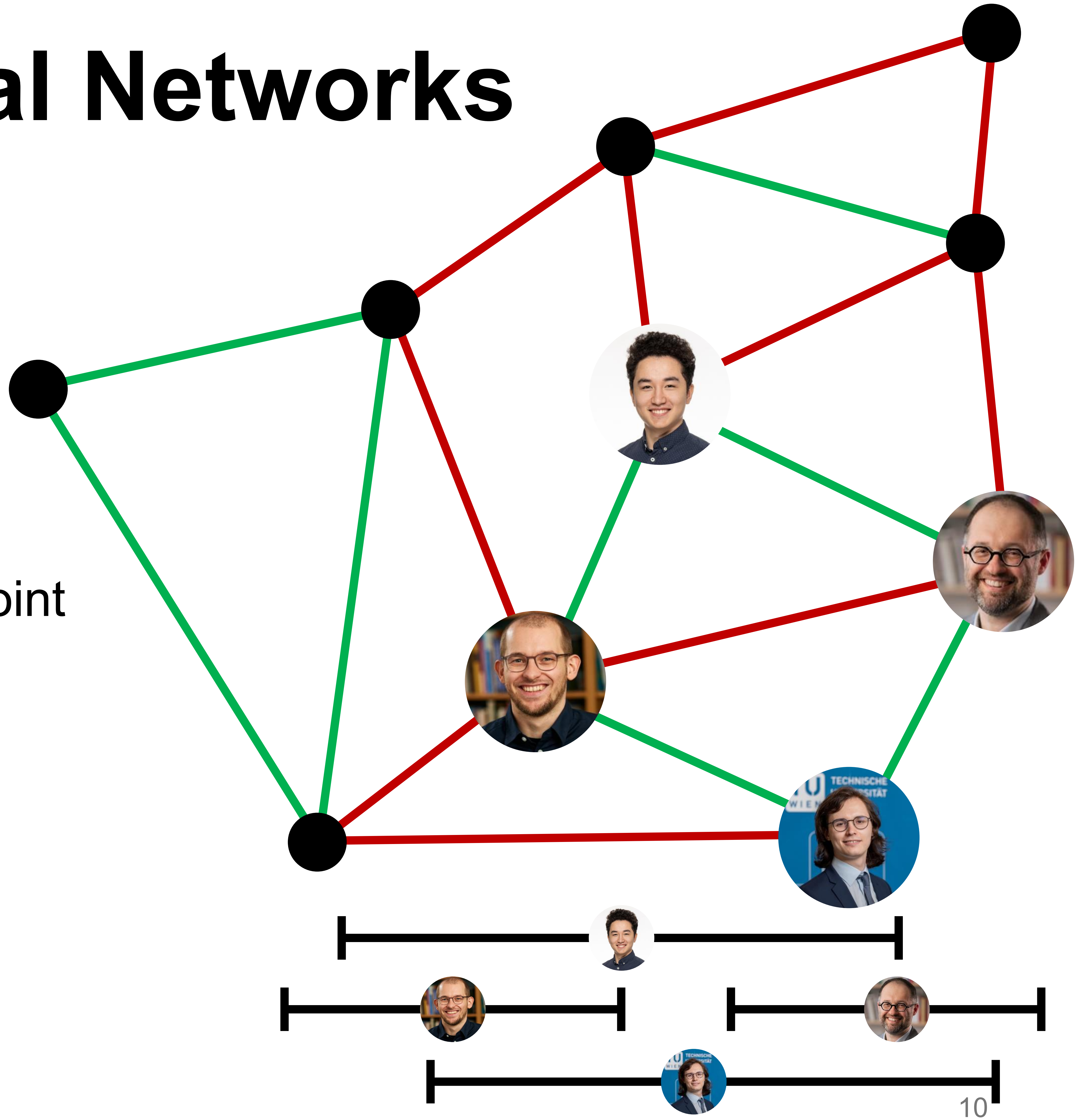
Interactions in Social Networks

San Diego Food Preferences

BEST INTERVAL APPROXIMATION:
Assign each vertex an interval $\{I_v : v \in V\}$ with $I_v \subset \mathbb{R}$ to maximize the number of

(1) $\{u, v\} \in E^+$ for which I_u and I_v overlap

(2) $\{u, v\} \in E^-$ for which I_u and I_v are disjoint



Interactions in Social Networks

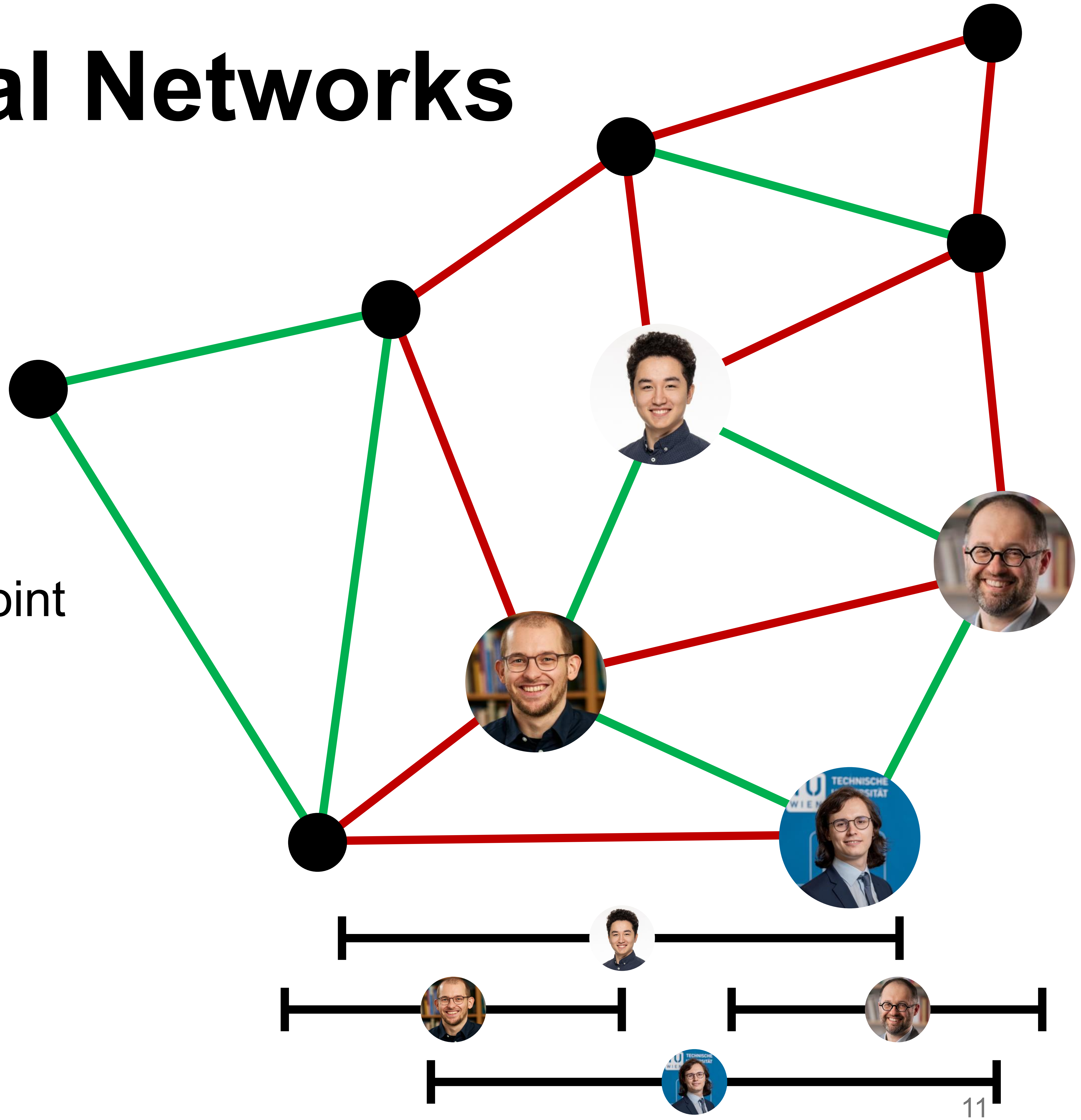
San Diego Food Preferences

BEST INTERVAL APPROXIMATION:
Assign each vertex an interval $\{I_v : v \in V\}$ with $I_v \subset \mathbb{R}$ to maximize the number of

(1) $\{u, v\} \in E^+$ for which I_u and I_v overlap

(2) $\{u, v\} \in E^-$ for which I_u and I_v are disjoint

- This is more expressive than CORRELATION CLUSTERING



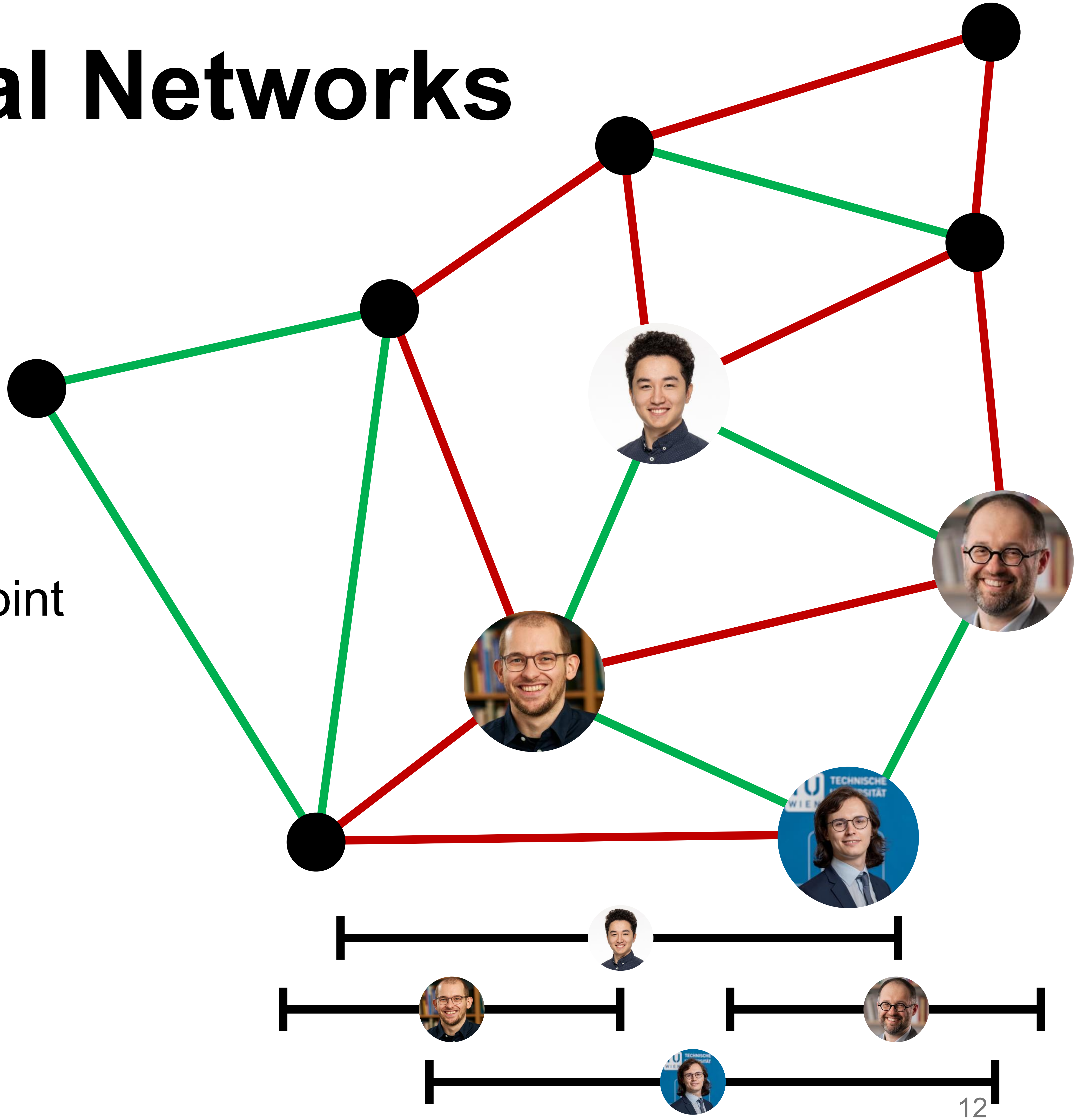
Interactions in Social Networks

San Diego Food Preferences

BEST INTERVAL APPROXIMATION:
Assign each vertex an interval $\{I_v : v \in V\}$ with $I_v \subset \mathbb{R}$ to maximize the number of

(1) $\{u, v\} \in E^+$ for which I_u and I_v overlap

(2) $\{u, v\} \in E^-$ for which I_u and I_v are disjoint



- This is more expressive than CORRELATION CLUSTERING
- The model captures the **tolerance** an individual has for other opinions

Problem Analysis

NP-Hardness

Theorem

BEST INTERVAL APPROXIMATION is NP-hard. This follows via a reduction from ACYCLIC DIGRAPH PARTITION.

NP-Hardness

Theorem

BEST INTERVAL APPROXIMATION is NP-hard. This follows via a reduction from ACYCLIC DIGRAPH PARTITION.

- One cannot get a constant factor approximation on the number of mistakes in OPT

NP-Hardness

Theorem

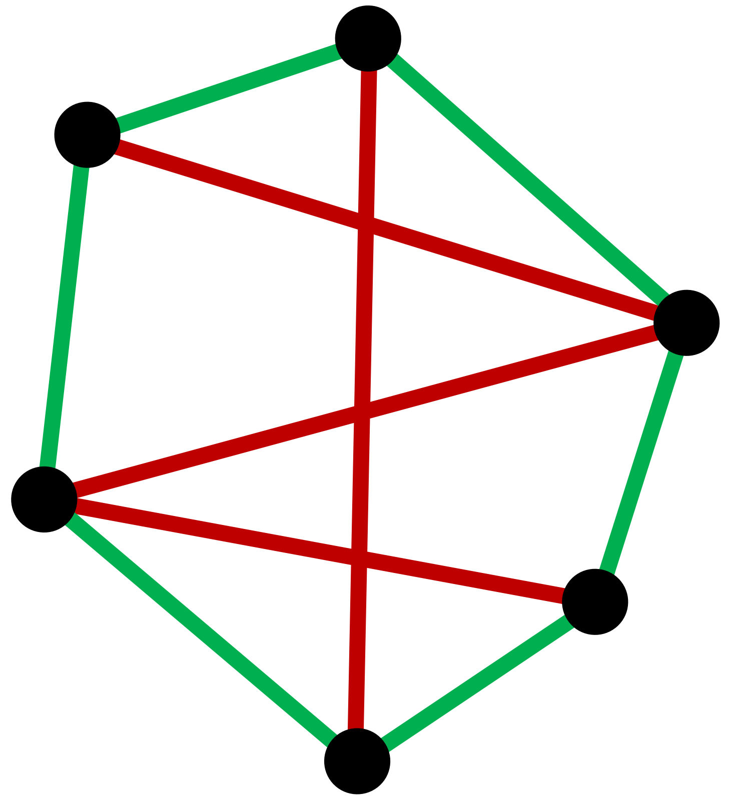
BEST INTERVAL APPROXIMATION is NP-hard. This follows via a reduction from ACYCLIC DIGRAPH PARTITION.

- One cannot get a constant factor approximation on the number of mistakes in OPT
- This result holds even if $G^+ = (V, E^+)$ is a chordless cycle

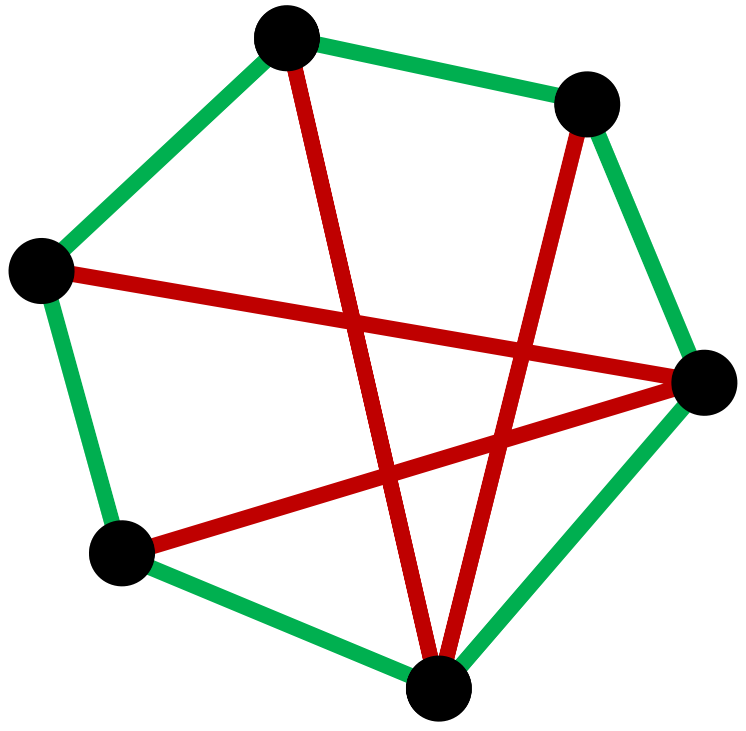
NP-Hardness

Theorem

BEST INTERVAL APPROXIMATION is NP-hard. This follows via a reduction from ACYCLIC DIGRAPH PARTITION.



- One cannot get a constant factor approximation on the number of mistakes in OPT
- This result holds even if $G^+ = (V, E^+)$ is a chordless cycle



Algorithms

PTAS in restricted setting

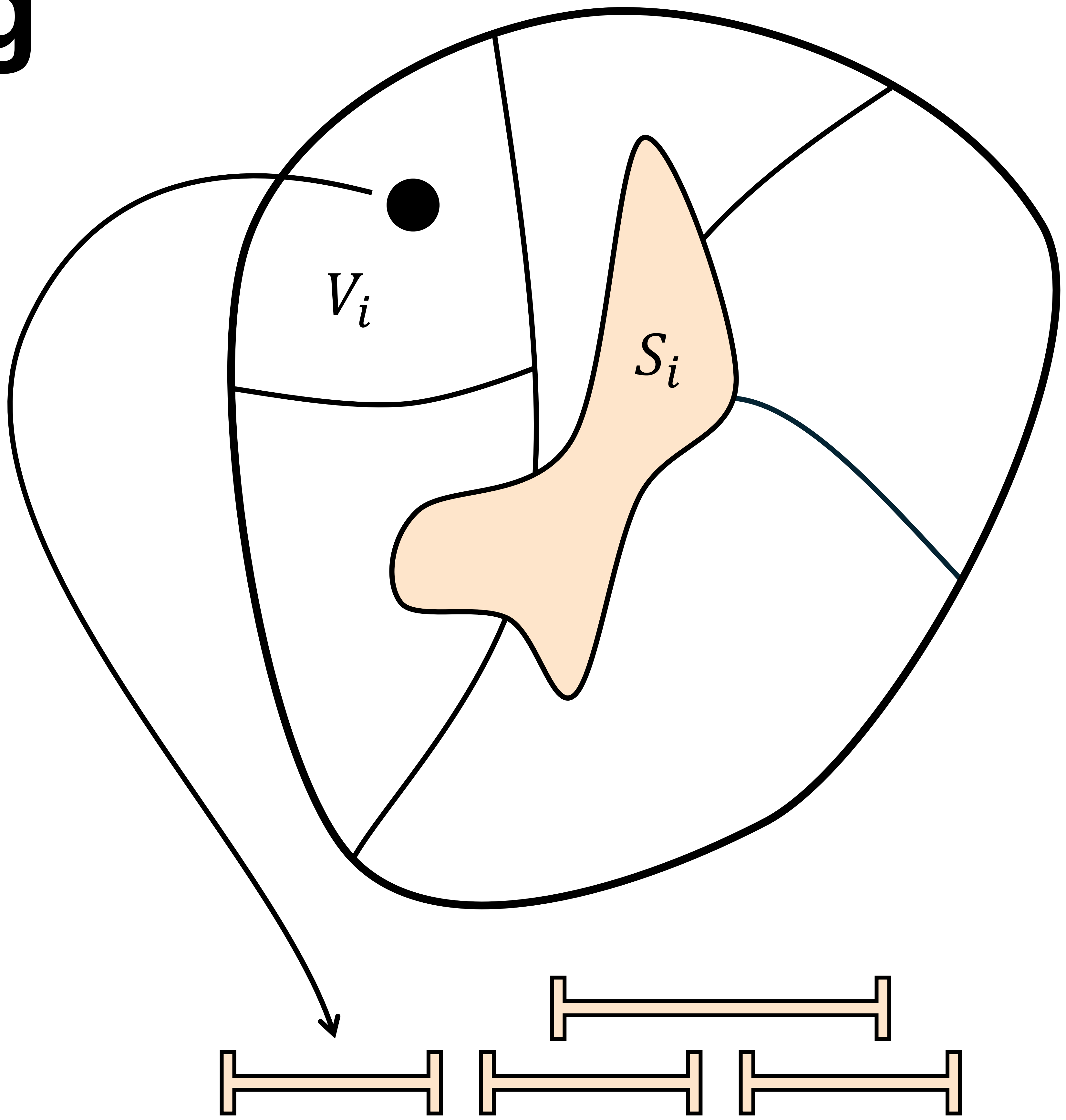
For complete signed graphs and fixed interval configurations find a $(1+\epsilon)$-approximation:

1. Partition V into sets $V_1, \dots, V_m$
2. For each subset V_i:

   Sample S_i with replacement from $V \setminus V_i$

   Solve S_i optimally

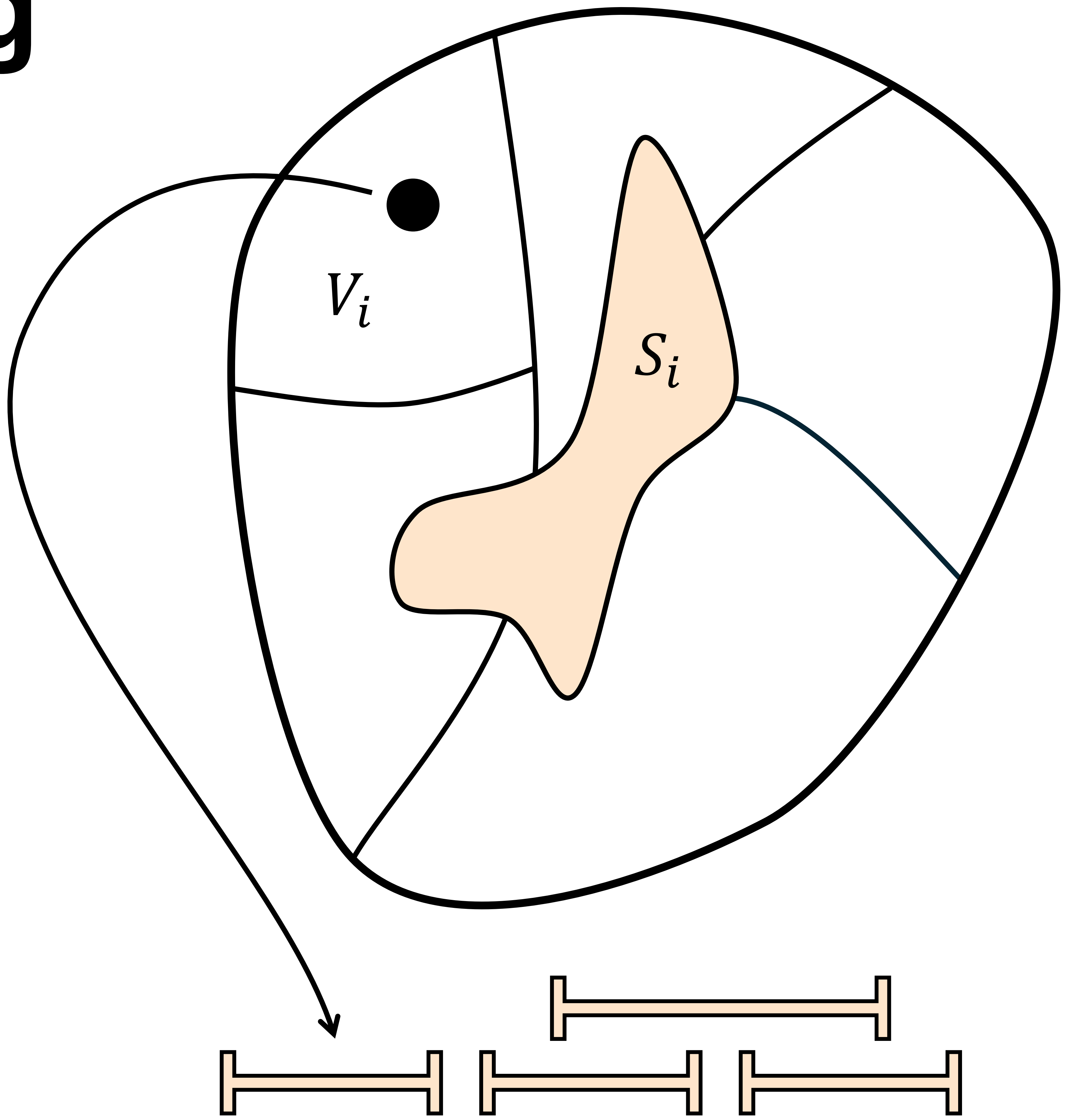
   Assign V_i greedily
3. Combine solutions across all V_i



PTAS in restricted setting

For complete signed graphs and fixed interval configurations find a $(1+\epsilon)$-approximation:

1. Partition V into sets $V_1, \ldots, V_m$
2. For each subset V_i:
   Sample S_i with replacement from $V \setminus V_i$
   Solve S_i optimally
   Assign V_i greedily
3. Combine solutions across all V_i



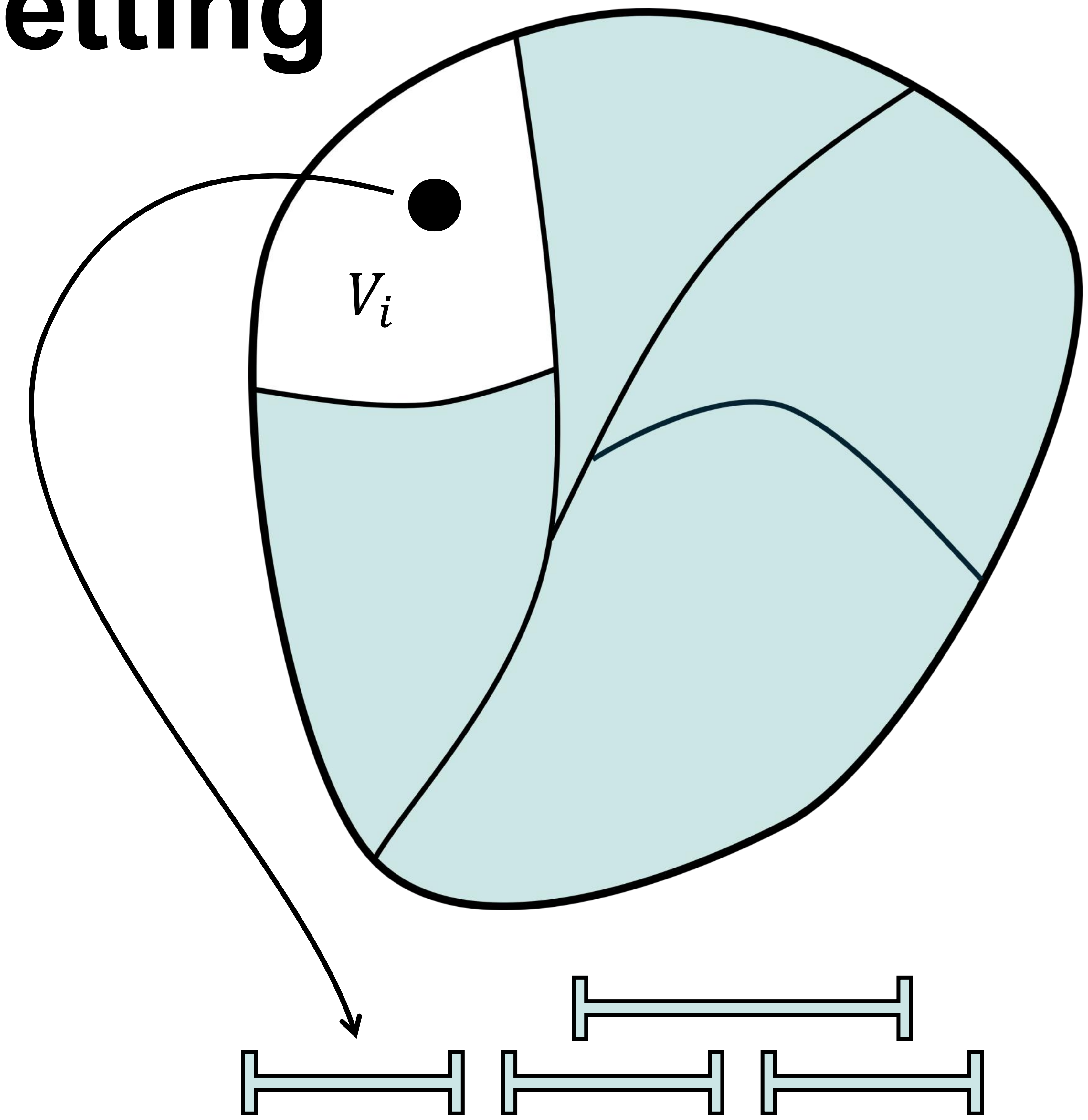
Heuristics in restricted setting

For ~~complete~~ signed graphs and fixed interval configurations find a heuristic solution:

1. Partition V into sets $V_1, \dots, V_m$
2. For each subset V_i:

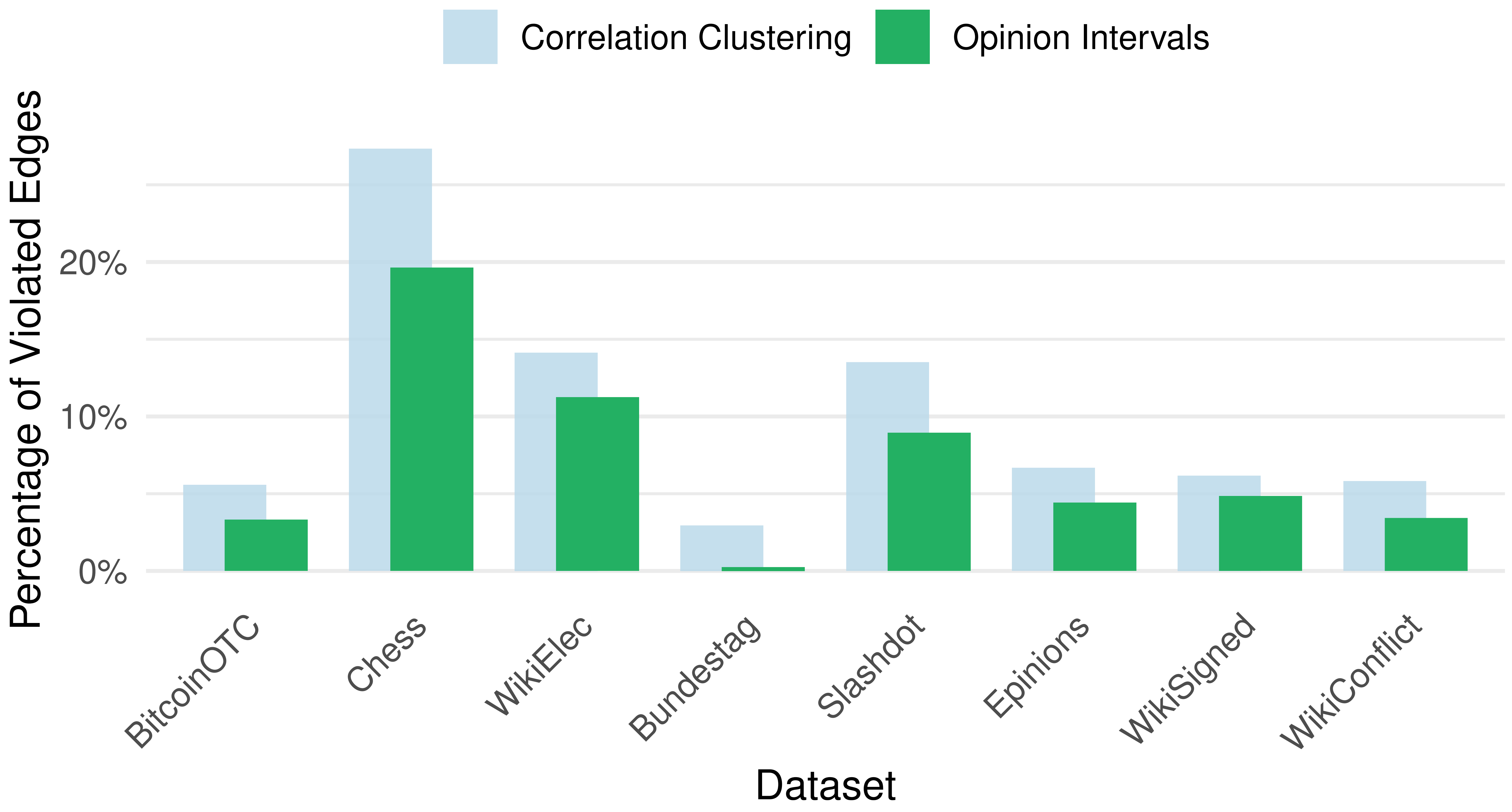
   Reuse solution on $V \setminus V_i$

   Assign V_i greedily
3. Combine solutions across all V_i

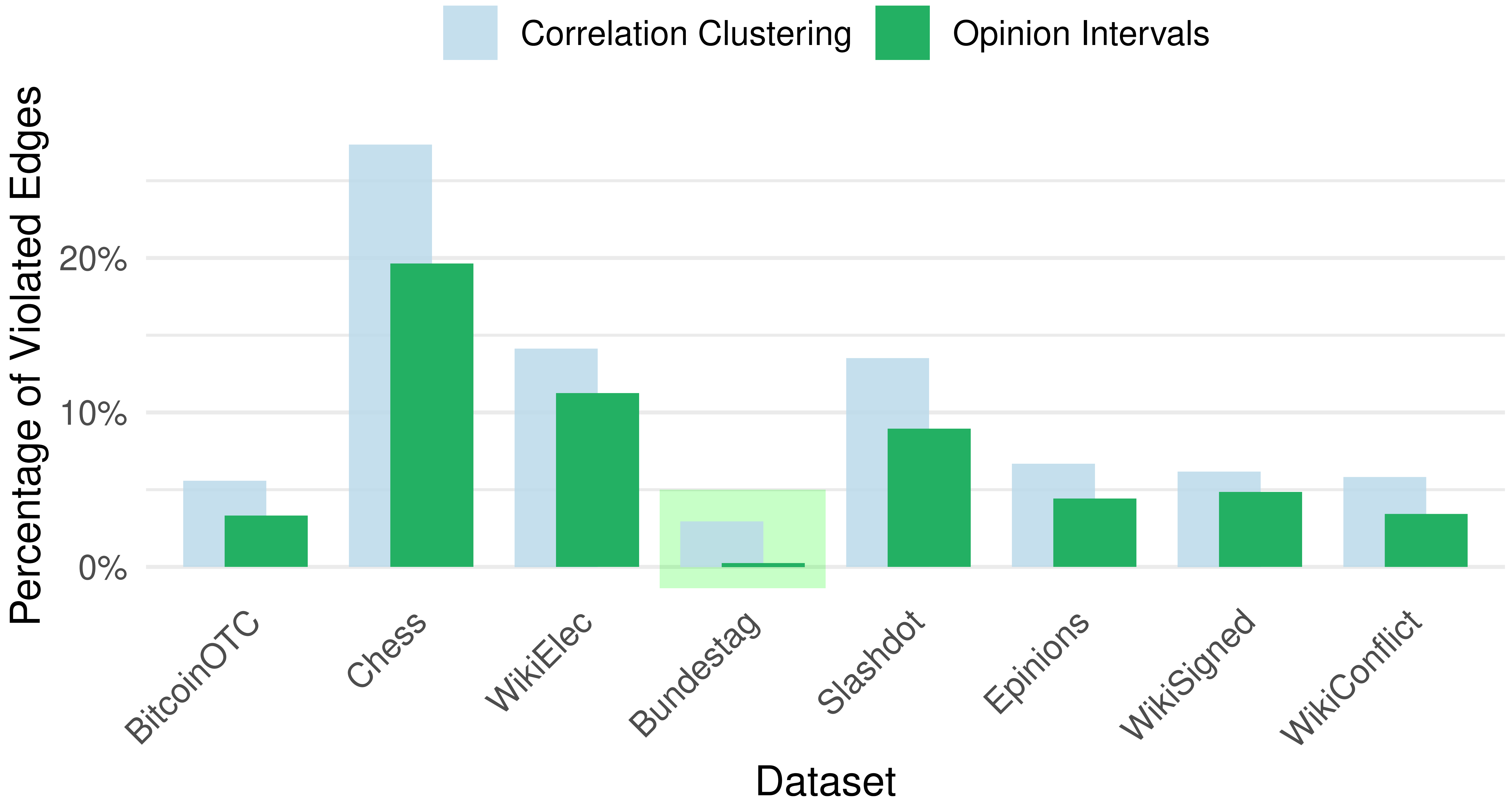


Experiments

38% fewer disagreements than Correlation Clustering

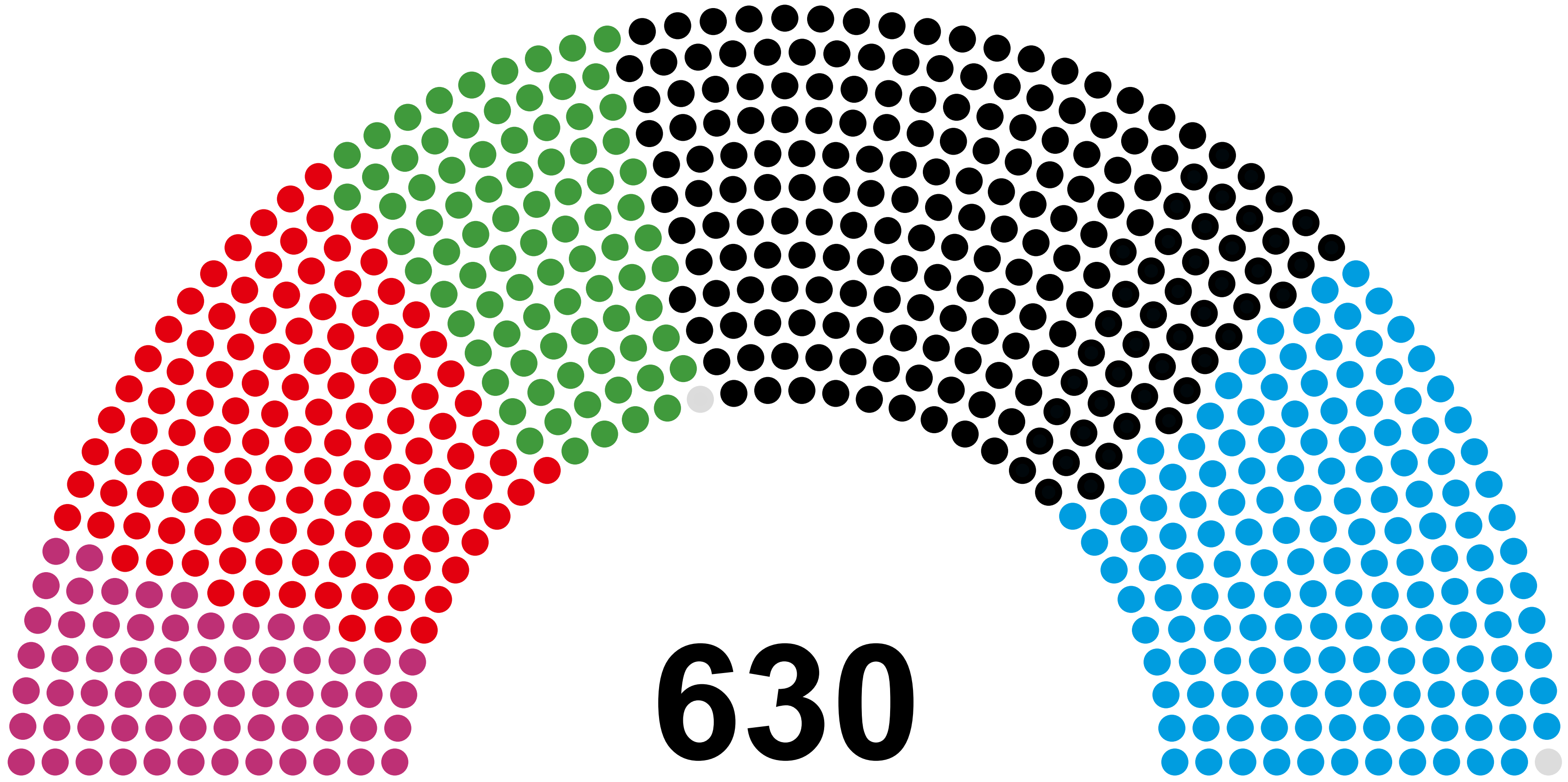


38% fewer disagreements than CORRELATION CLUSTERING

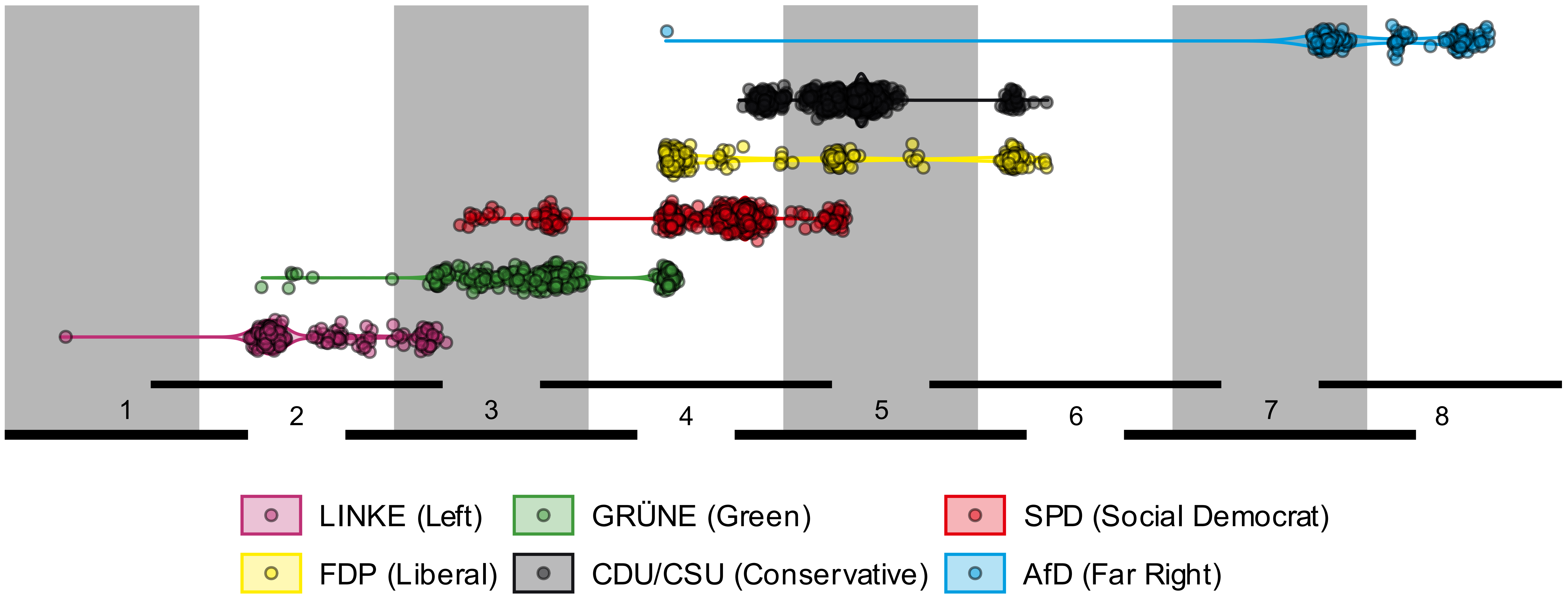


German Bundestag

- Multi-party political system
- Coalitions are formed to achieve majority
- We create a signed graph based on historic co-voting behaviour



Reconstructing the German political spectrum



Conclusion

We introduced the problem **Best Interval Approximation**

We provided a **PTAS** for complete graphs and a fixed interval configuration

We provided scalable and effective **heuristic algorithms**

Open Questions

Can we find better **approximation guarantees** in the unrestricted setting? (>75%?)

How can we efficiently find interval structures for **Best Interval Approximation?**

Can this approach be extended to **higher dimensions?**