Finding low-tension communities

Esther Galbrun, Behzad Golshan, Aristides Gionis and Evimaria Terzi

esther.galbrun@inria.fr, Inria Nancy Grand Est, France
behzad@recruit.ai, Recruit Institute of Technology, US-CA
aristides.gionis@aalto.fi, Helsinki Institute for Information Technology, Aalto University, Finland
evimaria@cs.bu.edu, Department of Computer Science, Boston University, US-MA

A **community-search** problem with **opinion dynamics**:
find a subgraph that connects the seed nodes and has low social tension.

SOCIAL NETWORKS WITH PROFILES

We consider a social network $G = (V, E)$ where
nodes in V represent individuals, and
edges in E represent their interactions.

Each individual has his own preferences, habits, opinions, etc.
However, individuals may choose not to act in accordance with their true prefererences as they try to minimize peer pressure by conforming their preferences to those of their peers.

Each node i is associated to
a **latent profile**, x_i, representing the individual's true preferences, and
a **conformed profile**, f_i, representing his expressed preferences,
both take value in the interval $[0, 1]$.

PROBLEM STATEMENT

Given a network $G = (V, E)$, latent profiles $\mathbf{x}$ and a set of seed nodes $Q \subseteq V$,
find $V' \subseteq V$ such that $Q \subseteq V'$, the graph G' induced by V' on G is connected
and $T(G', \mathbf{x}, \mathbf{f})$ is minimized, where $\mathbf{f}$ is computed by the repeated averaging model on G'.

ALGORITHMS

Computing the conformed profiles is costly.
Instead, use proxy for contribution of each pair of neighboring nodes.
Assign weight $w_{ij} = |x_i - x_j|$ to each edge $(i, j) \in E$.

Spanning-tree approach

Build a spanning tree between the query nodes,
using the 2-approximation to Steiner tree problem.

Different ways to score the tree yield variants:

`CTree(e)` Number of edges involved
`CTree(s)` Sum of weights of the edges along the path

Top-down approach

Iteratively remove nodes until it is no longer possible without disconnecting the query nodes.

Different way to pick next node yield variants:

`CPeel(r)` Random
`CPeel(s)` Sum of adjacent edges weight
`CPeel(m)` Max of adjacent edges weight

MEASURING THE SOCIAL TENSION

Each node i bears
an **inner tension** caused by the difference between
its own latent and conformed profiles, and
a **cross tension** caused by the difference between
its own conformed profile and those of its neighbors:

$$T_i(G, \mathbf{x}, \mathbf{f}) = (x_i - f_i)^2 + \sum_{j \in N_G(i)} (f_i - f_j)^2.$$

The **social tension** of the network is the sum of the individual tensions:

$$\begin{aligned} T(G, \mathbf{x}, \mathbf{f}) &= \sum_{i \in V} T_i(G, \mathbf{x}, \mathbf{f}) \\ &= \sum_{i \in V} \left((x_i - f_i)^2 + \sum_{j \in N_G(i)} (f_i - f_j)^2\right) \\ &= \sum_{i \in V} (x_i - f_i)^2 + \sum_{(i,j) \in E} 2(f_i - f_j)^2. \end{aligned}$$

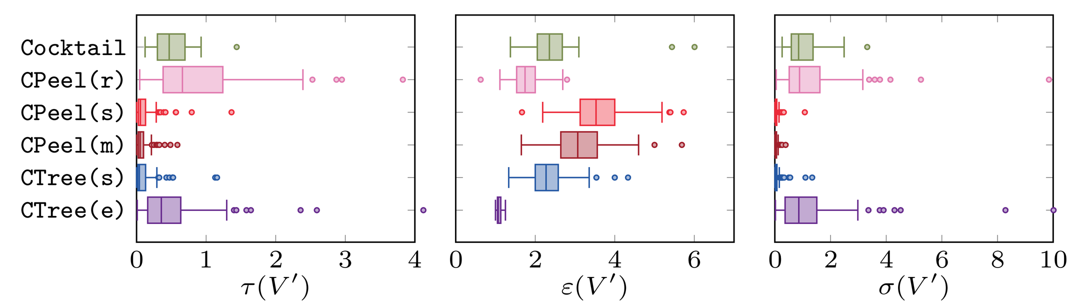
CONFORMATION PROCESS

Consider a **repeated averaging process** where at each step each node adjusts its conformed profile by setting it to the average of its latent profile and the conformed profile of its neighbors:

$$f_i(t+1) = \frac{x_i + \sum_{j \in N_G(i)} f_j(t)}{1 + |N_G(i)|}.$$

The repeated averaging model is equivalent to choosing f_i to minimize $T_i(G, \mathbf{x}, \mathbf{f})$.
It yields a Nash equilibrium for the tension, not a social optimum.

Results for the 2-hop ego-network of C.Papadimitriou with single-attribute latent profiles derived from conferences.



EXPERIMENTS

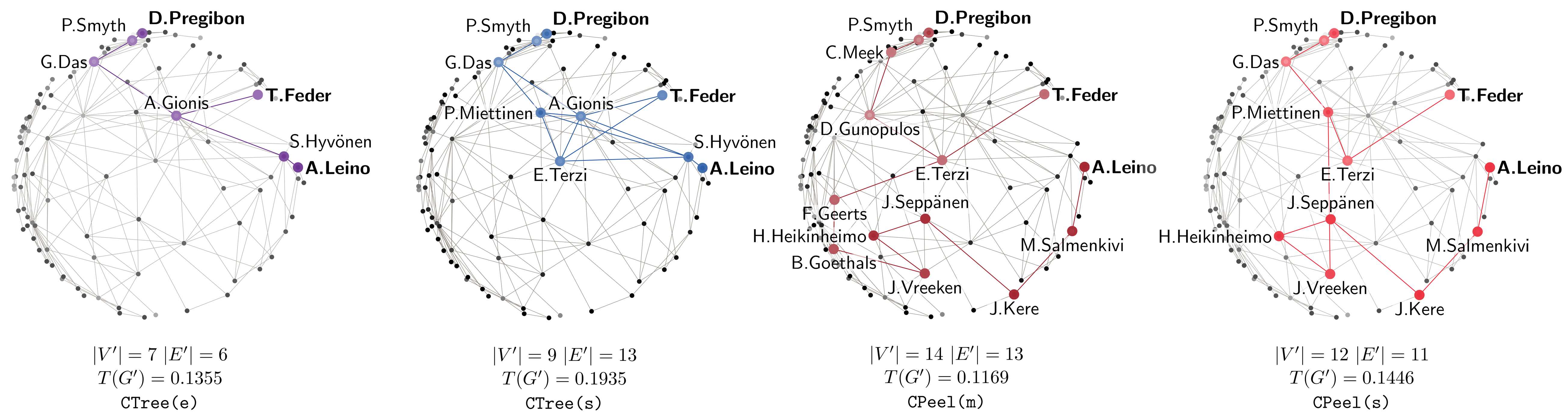
Comparing the algorithm variants and the `Cocktail` algorithm [3].

Evaluation measures:

standardized social tension (main measure) $\tau(V') = T(V')/(2e_b \cdot \overline{w^2}(V))$,
standardized solution size (auxiliary measure) $\varepsilon(V') = \left|E(V')\right| / e_b$, and
standardized average edge weight (auxiliary measure) $\sigma(V') = \overline{w^2}(V')/\overline{w^2}(V)$.

where e_b is the size of the minimum spanning tree connecting the query nodes,
and $\overline{w^2}(V)$ is the average squared edge weight in V.

Example solutions for connecting three seed nodes in the 1-hop ego-network of H.Mannila with single-attribute latent profiles derived from keywords.



REFERENCES

[1] Bindel, Kleinberg and Oren (2011) *How Bad is Forming Your Own Opinion?* FOCS
[2] Gionis, Terzi, and Tsaparas (2013) *Opinion Maximization in Social Networks.* SDM
[3] M. Sozio and A. Gionis (2010) *The community-search problem and how to plan a successful cocktail party.* KDD

BOSTON UNIVERSITY
HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY
Aalto University
informatics mathematics Inria