An Efficient Local Search Approach for Polarized Community Discovery in Signed Networks

Linus Aronsson, Morteza Haghir Chehreghani

Chalmers University of Technology and University of Gothenburg

October 19, 2025

Problem and Motivation

- **Signed network**: $G = (V, E)$ with adjacency matrix A, where $A_{ij} \in \{-1, 0, +1\}$.
- **Goal**: *Polarized Community Discovery (PCD)*. Find k non-neutral groups $S_1, \ldots, S_k$ (with large similarity inside each cluster, and small similarity between each pair of clusters) and a neutral set S_0.
- **Use cases:** Analyzing social systems to detect and mitigate polarization, echo chambers and spread of misinformation.

PCD Objective

- **Polarity (common PCD objective)**:

$$\max_{S_0, S_1, \ldots, S_k} \frac{\sum_{m=1}^{k} \sum_{i,j \in S_m} A_{ij} - \frac{1}{k-1} \sum_{\substack{m,p=1 \\ m \neq p}}^{k} \sum_{i \in S_m} \sum_{j \in S_p} A_{ij}}{\sum_{m=1}^{k} |S_m|}.$$

- **Issue with polarity**: often leads to solutions with multiple empty clusters (large cluster size imbalance).

Contributions

- New PCD objective that encourages (reasonably) balanced non-neutral clusters.
- First scalable local search algorithm for PCD (motivated by success of local search algorithms for correlation clustering and other machine learning problems).
- Prove a linear convergence rate by connecting local search to block-coordinate Frank–Wolfe (FW) optimization (possible due to the structure of our proposed PCD objective).
- Our proposed method outperforms baselines on synthetic and real-world datasets.

Proposed PCD Objective

Problem (k-PCD)

$$\max_{S_0,S_1,\ldots,S_k} \sum_{m=1}^{k} \sum_{i,j\in S_m} A_{ij} - \alpha \sum_{\substack{m,p=1\\ m\neq p}}^{k} \sum_{i\in S_m} \sum_{j\in S_p} A_{ij} - \beta \sum_{m=1}^{k} |S_m|^2.$$

- $\alpha = \frac{1}{k-1}$ balances intra vs. inter terms.
- $\beta > 0$ regulates size and discourages imbalance.
- Theorem 1: k-PCD is NP hard.

Local Search and Block-Coordinate FW

Local Search (discrete)

Local Search for PCD

```
Initialize S_0, S_1, ..., S_k
repeat
    pick object i ∈ V randomly
    move i to cluster S_m, m ∈ {0, ..., k}, that maximally increases the objective
until no improving move
```

Initialize $S_0, S_1, \ldots, S_k$
repeat
  pick object $i \in V$ randomly
  move i to cluster S_m, $m \in \{0, \ldots, k\}$, that maximally increases the objective
until no improving move

Block-coordinate Frank–Wolfe (FW) (soft cluster membership)

Block-Coordinate FW

Soft memberships $x_i \in \Delta_{k+1}$ for all i
for $t = 0, 1, \ldots, T$ **do**
  pick block i
  $s_i \leftarrow \arg\max_{u \in \Delta_{k+1}} u^\top \nabla_i f(x)$
  $x_i \leftarrow (1 - \gamma)x_i + \gamma s_i$ $\quad \triangleright\ \gamma$ via line-search
end for

Equivalence and Convergence

Theorem 2 (informal)

The simple local search algorithm is equivalent to the block-coordinate FW algorithm.

Theorem 3 (informal)

The convergence rate of the local search algorithm is $O(1/t)$.

Note: For general non-concave FW the known rate is $O(1/\sqrt{t})$. Due to the structure of our objective, we prove the convergence rate $O(1/t)$, which is significantly better.

Local Search (efficient)

Local Search for PCD (efficient)

Initialize $S_0, S_1, \ldots, S_k$ and $X \in \{0,1\}^{n \times k}$ $\quad \triangleright X_{i,m} = 1$ iff $i \in S_m$
$M \leftarrow 2AX$
repeat
  pick object i and let $\hat{p}$ be its current cluster
  compute $G_{i,p} \leftarrow (1+\alpha)M_{i,p} - \alpha \sum_{q \in [k]} M_{i,q} - 2\beta|S_p| + 2\beta\mathbf{1}[i \in S_p] - \beta$ for all $p \in [k]$
  set $G_{i,0} \leftarrow 0$ $\quad \triangleright$ neutral set
  $p^\star \leftarrow \arg\max_{p \in \{0,\ldots,k\}} G_{i,p}$
  if $p^\star = \hat{p}$ **then** skip to next iteration
  Assign object i to cluster $S_{p^\star}$ and update X
  if $\hat{p} \in [k]$ **then** $M_{:,\hat{p}} \leftarrow M_{:,\hat{p}} - 2A_{:,i}$
  if $p^\star \neq 0$ **then** $M_{:,p^\star} \leftarrow M_{:,p^\star} + 2A_{:,i}$
until no improving move

Complexity: $O(kn^2 + T(n+k))$. Significant improvement over naive local search with $O(Tn^2k^2)$.

Results: Synthetic Datasets

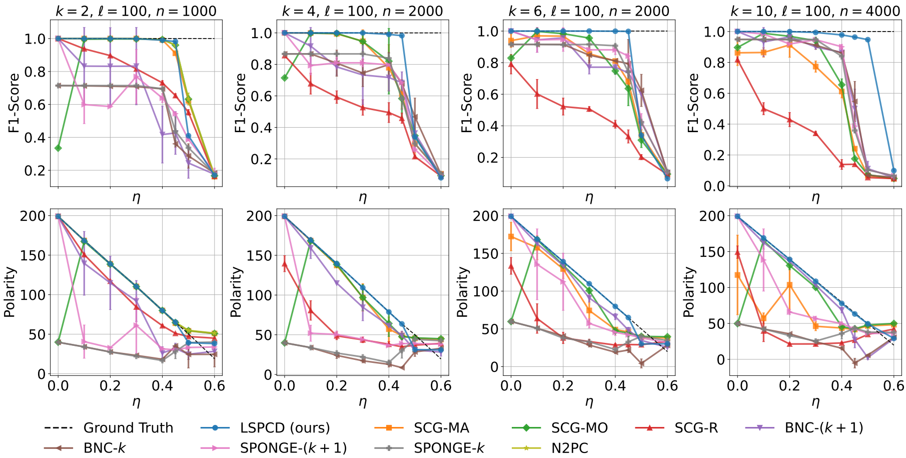

Figure: F1-score and polarity of different methods on synthetic graphs generated using the m-SSBM model, as the noise level η varies. See main text below for details.

Results: Real-World Datasets

Table: Polarity (POL) and imbalance factor (IF) for different methods and real-world datasets. $|E|$ denotes the number of edges with non-zero edge weight.

k		BTC POL	BTC IF	REF POL	REF IF	WikiC POL	WikiC IF	EP POL	EP IF	WikiP POL	WikiP IF
	$\|V\|$	6K		11K		116K		131K		138K	
	$\|E\|$	214K		251K		2M		711K		715K	
2	LSPCD (ours)	29.0	0.65	146.1	0.71	190.8	0.83	127.8	0.73	82.0	0.30
	SCG–MA	28.8	0.16	172.2	0.01	155.2	0.53	128.3	0.04	82.8	0.01
	SCG–MO	29.5	0.03	174.1	0.01	175.7	0.43	128.7	0.04	88.4	0.01
	N2PC ($\gamma = 1$)	29.6	0.02	173.6	0.01	172.8	0.46	169.7	0.00	87.5	0.00
	N2PC ($\gamma = 1.2$)	30.1	0.46	173.6	0.02	175.7	0.77	169.8	0.00	87.1	0.00
	N2PC ($\gamma = 1.5$)	24.4	1.00	130.3	0.94	158.2	0.99	169.9	0.00	86.6	0.02
	N2PC ($\gamma = 1.7$)	23.9	1.00	119.4	1.00	155.5	0.99	124.3	0.29	75.2	0.39
	N2PC ($\gamma = 2.0$)	24.1	1.00	118.1	1.00	142.0	1.00	76.7	0.99	48.3	0.96
4	LSPCD (ours)	23.3	0.47	139.2	0.41	113.6	0.56	111.5	0.58	71.6	0.27
	SCG–MA	25.1	0.22	94.5	0.68	104.9	0.06	127.4	0.30	56.5	0.52
	SCG–MO	25.3	0.22	82.1	0.70	117.9	0.24	129.0	0.34	39.7	0.30