Chair of Computer Science 5
RWTH Aachen University

Analysis of Overlapping Communities in Signed Complex Networks

Mohsen Shahriari, Ying Li, Ralf Klamma

Advanced Community Information Systems (ACIS)
RWTH Aachen University, Germany
shahriari@dbis.rwth-aachen.de

Mohsen Shahriari,
Ying Li,
Ralf Klamma

Agenda

- Introduction to OCD
- Related Work
- Motivation & Research Questions
- Overlapping Community Detection (OCD) Algorithms for Signed Networks
- Evaluation
- Results
- Conclusion and Outlook

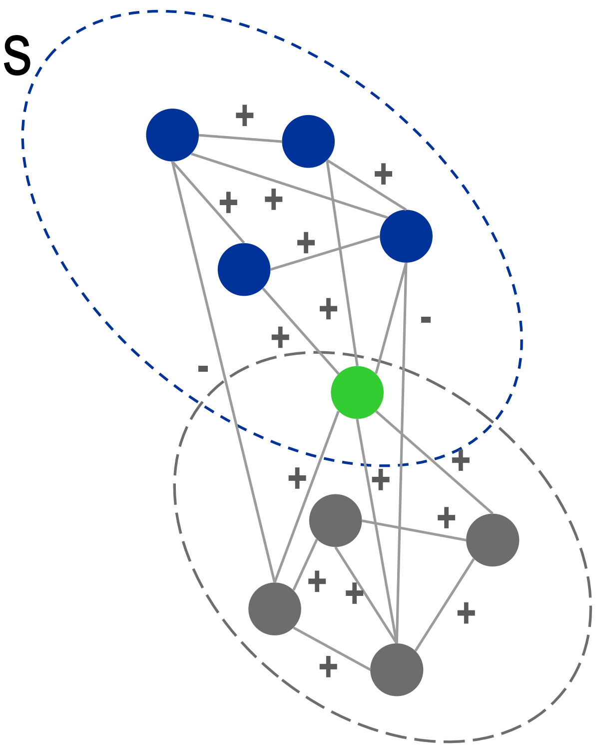
Mohsen Shahriari,
Ying Li,
Ralf Klamma

Introduction to OCD in Signed Networks

- Community detection as an important part of network analysis
- Two key characteristics of signed social networks
  - Nodes in the overlapping communities
  - Relations with signs
- Community structure

Inside Communities
- Dense
- Positive

Between Communities
- Negative
- Sparse



Learning Layers

Analysis of Overlapping Communities in Signed Complex Networks

Mohsen Shahriari,
Ying Li,
Ralf Klamma

Motivation

- Practical application of OCD in signed networks like
  - Informal learning networks
  - Review sites
  - Open source developer networks
- Contribute to the current research on OCD in signed networks with the following difficiencies
  - Few algorithms
  - No comparison between available algorithms

Mohsen Shahriari,
Ying Li,
Ralf Klamma

Related Work on Community Detection in Signed Graphs

- Non-overlapping community detection
  - Agent-based finding and extracting communities (FEC) [YaCL07]
  - Two-step approach by maximizing modularity and minimizing frustration [AnMa12]
  - Clustering re-clustering algorithm (CRA) [AmPi13]
- Overlapping community detection
  - Signed Disassortative Degree Mixing and Information Diffusion Algorithm (SDMID) [ShKl15]
  - Signed Probabilistic Mixture Model (SPM) [CWYT14]
  - Multi-objective Evolutionary Algorithm based on Similarity for Community Detection in Signed Networks (MEA_s-SN) [LiLJ14]

Mohsen Shahriari,
Ying Li,
Ralf Klamma

Research Questions

- How do Signed Disassortative degree Mixing and Information Diffusion (SDMID), Signed Probabilistic Mixture model (SPM) and Multi-objective Evolutionary Algorithm (MEA) perform in comparison with each other, in terms of knowledge-driven and statistical metrics?
- What are the structural properties of covers detected by SDMID, SPM and MEA and how do they differ?

Mohsen Shahriari,
Ying Li,
Ralf Klamma

Signed Disassortative Degree Mixing and Information Diffusion Algorithm: Phase 1

Identify leaders

- Calculate Local Leadership Value (LLD) using effective degree (ED) and normalized disassortativeness (DASS)

$$ED(i) = \frac{Max(|in^{+}(i)| - |in^{-}(i)|, 0)}{|in^{+}(i)| + |in^{-}(i)|} \qquad DASS(i) = \frac{\sum_{j \in Nei(i)} (\deg(i) - \deg(j))}{\sum_{j \in Nei(i)} (deg(i) + deg(j))}$$

$$LLD(i) = \alpha \times DASS(i) + (1 - \alpha) \times ED(i)$$

- Identify local leaders:

$$\forall j \in Nei(i), LLD(i) \geq LLD(j)$$

- Identify global leaders:

$$|FL(i)| > \frac{\sum_{j \in LL} |FL(j)|}{|LL|}$$

where FL: Follower Set, LL: Local Leader Set

Learning Layers

Analysis of Overlapping Communities in Signed Complex Networks

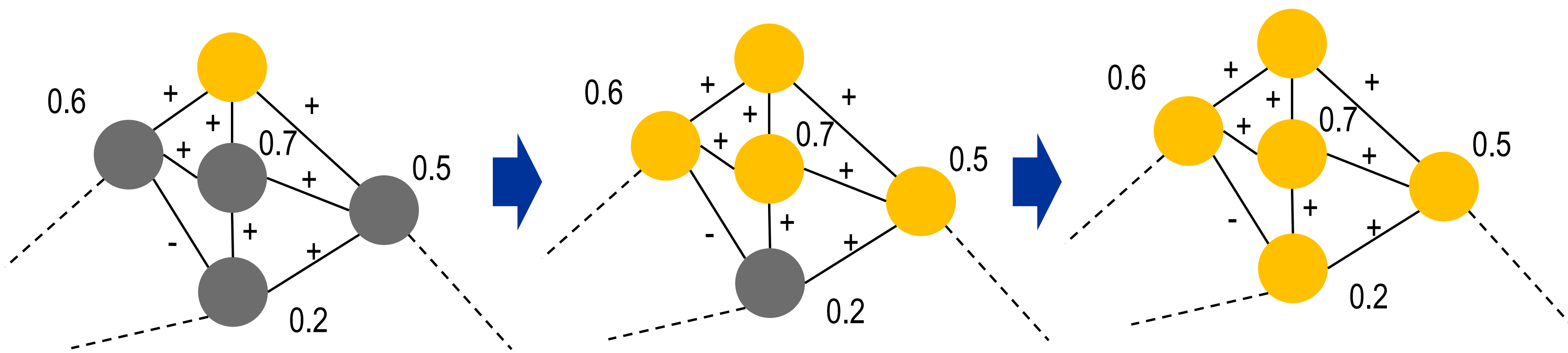
Mohsen Shahriari, Ying Li, Ralf Klamma

Signed Disassortative Degree Mixing and Information Diffusion Algorithm: Phase 2

Cascading (network coordination game)

- Assign a leader node *k* behavior *B* and all other nodes behavior *A*
- Node *i* with current behavior *A* will change its behavior to that (*B*) of its neighbors, if the potential payoff p_B(i) is above a predefined threshold, i.e. LLD:

$$p_B(i) = \frac{|\{u|u \in Nei^{+}(i) \text{ and } behavior(u) = B\}| - |\{v|v \in Nei^{+}(i)\ and\ behavior(v) = B\}|}{|\{u|u \in Nei^{+}(i)\ and\ behavior(u) = B\}| + |\{v|v \in Nei^{+}(i)\ and\ behavior(v) = B\}|}$$

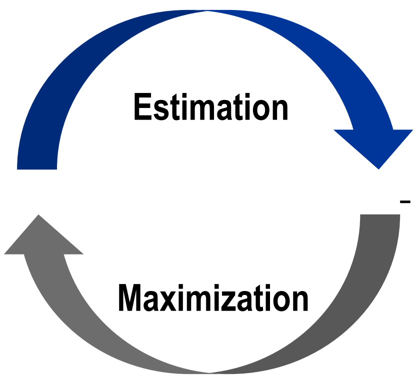


Mohsen Shahriari,
Ying Li,
Ralf Klamma

Signed Probabilistic Mixture Model

- Based on Expectation-Maximization (EM) method
- Maximize the log function of the marginal likelihood of the signed network:

$$P(E|\omega,\theta) = \prod_{e_{ij}\in E}\left(\sum_{rr}\omega_{rr}\,\theta_{ri}\theta_{rj}\right)^{A_{ij}^{+}}\left(\sum_{rs(r\neq s)}\omega_{rs}\,\theta_{ri}\theta_{sj}\right)^{A_{ij}^{-}}$$



Use ω,θ to compute

- ○ The probability of a positive edge from a community *r* : p_1
- ○ The probability of a negative edge from two communities *r* and *s*: p_2

Update ω,θ with p_1 and p_2 by maximizing $lnP(E|\omega,\theta)$

Mohsen Shahriari,
Ying Li,
Ralf Klamma

Multi-Objective Evolutionary Algorithm Based on Similarity for Community Detection in Signed Networks

- Based upon structural similarity between adjacent nodes

$$s(u,v) = \frac{\sum_{x \in B(u) \cap B(v)} \Psi(x)}{\sqrt{\sum_{x \in B(u)} w_{ux}^2} \cdot \sqrt{\sum_{x \in B(v)} w_{vx}^2}}$$

where $\Psi(x) = 0, \text{if } w_{ux} < 0 \text{ and } w_{vx} < 0;\ w_{ux}\, w_{vx}, otherwise$

- Objective functions
  - Maximize the sum of positive similarities within communities
  - Maximize the sum of negative similarities between communities

- Optimal solution is selected with MOEA/D (multiobjective evolutionary algorithm based on decomposition) [ZhLi07]
  - Decomposition into scalar optimization
  - Simultaneous optimization of these subproblems

Mohsen Shahriari,
Ying Li,
Ralf Klamma

Evaluation Metrics

- **Normalized mutual information**: regards M_{ik}, $M_{il'}$ as two random variables and determines the mutual information (M_i: membership vector, k: k-th community in detected cover, l': l'-th community in real cover)
- **Signed modularity**: measures the strength of a community partition by taking into account the degree distribution

$$Q_{SO} = \frac{1}{2(w^+)_e + 2|(w^-)_e|} \sum_{e_{ij}} \left[w_{ij} - \left(\frac{w_i^+ w_j^+}{2(w^+)_e} - \frac{w_i^- w_j^-}{2|(w^-)_e|} \right) \right] \delta(C_i, C_j),$$

where $\delta(C_i, C_j)$: No.of communities e_{ij} resides

- **Frustration**: normalized weighted weight sum of negative edges inside communities and positive edges between communities

$$Frustration = \frac{\alpha \times |(w_{intra}^-)_e| + (1-\alpha) \times |(w_{inter}^+)_e|}{(w^+)_e + |(w^-)_e|}$$

- **Execution time**

Mohsen Shahriari,
Ying Li,
Ralf Klamma

Synthetic Network Generator

- Comes from the idea of [LiLJ14] and is based on the Lancichinetti-Fortunato-Radicchi (LFR) model (directed and unweighted) and a model from [YaCL07]
- Parameters
  - From LFR: no. of nodes, average/max degree, minus exponents for the degree and community size distributions which are power laws, min/max community size, no. of overlapping nodes, no. of communities, fraction of edges that each node shares with other communities.
  - From [YaCL07]: proportion of negative edges inside communities P_- and proportion of positive edges between communities P_+

Learning Layers

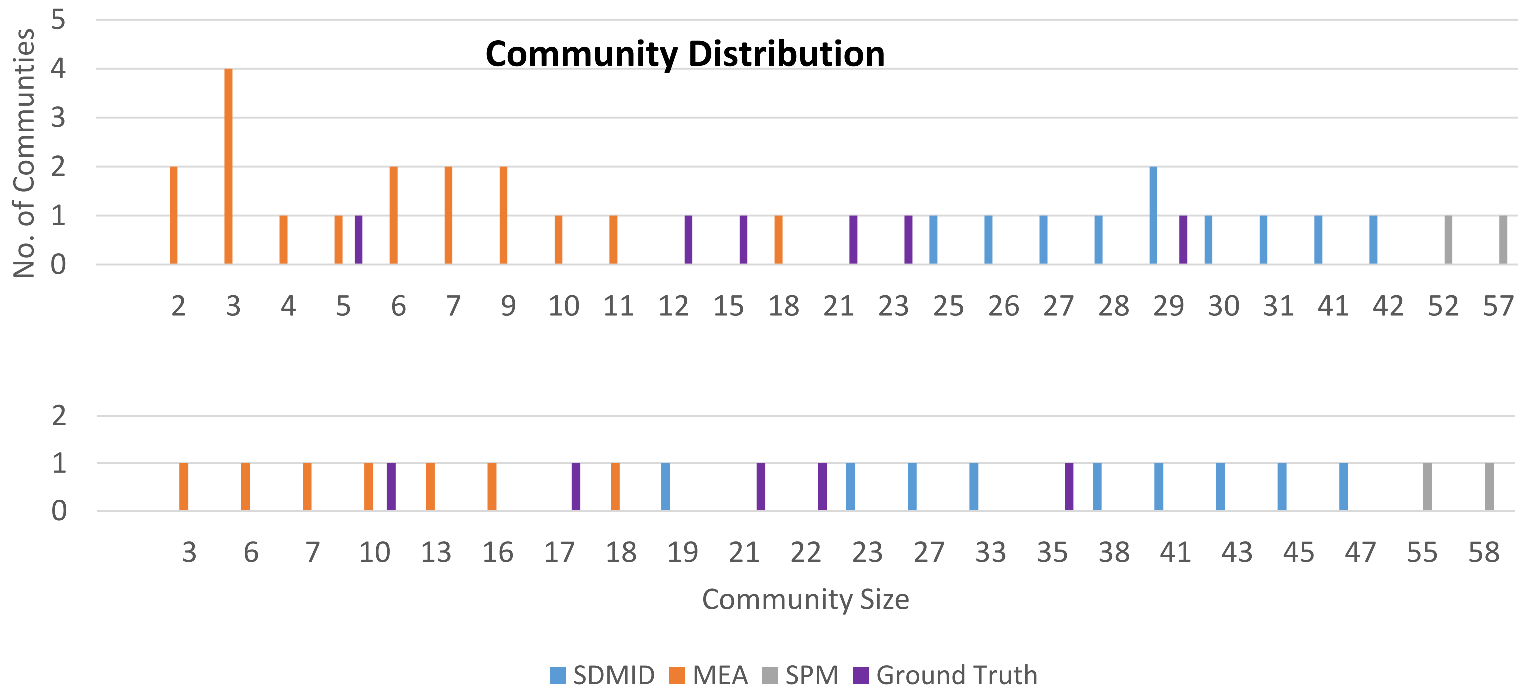
Mohsen Shahriari, Ying Li, Ralf Klamma

Experiments on Benchmark Networks: Community Structure (1)

- SDMID has a more similar community distribution in comparison to the ground truth
- SPM detects the biggest community sizes

Maxc=35

Maxc=40

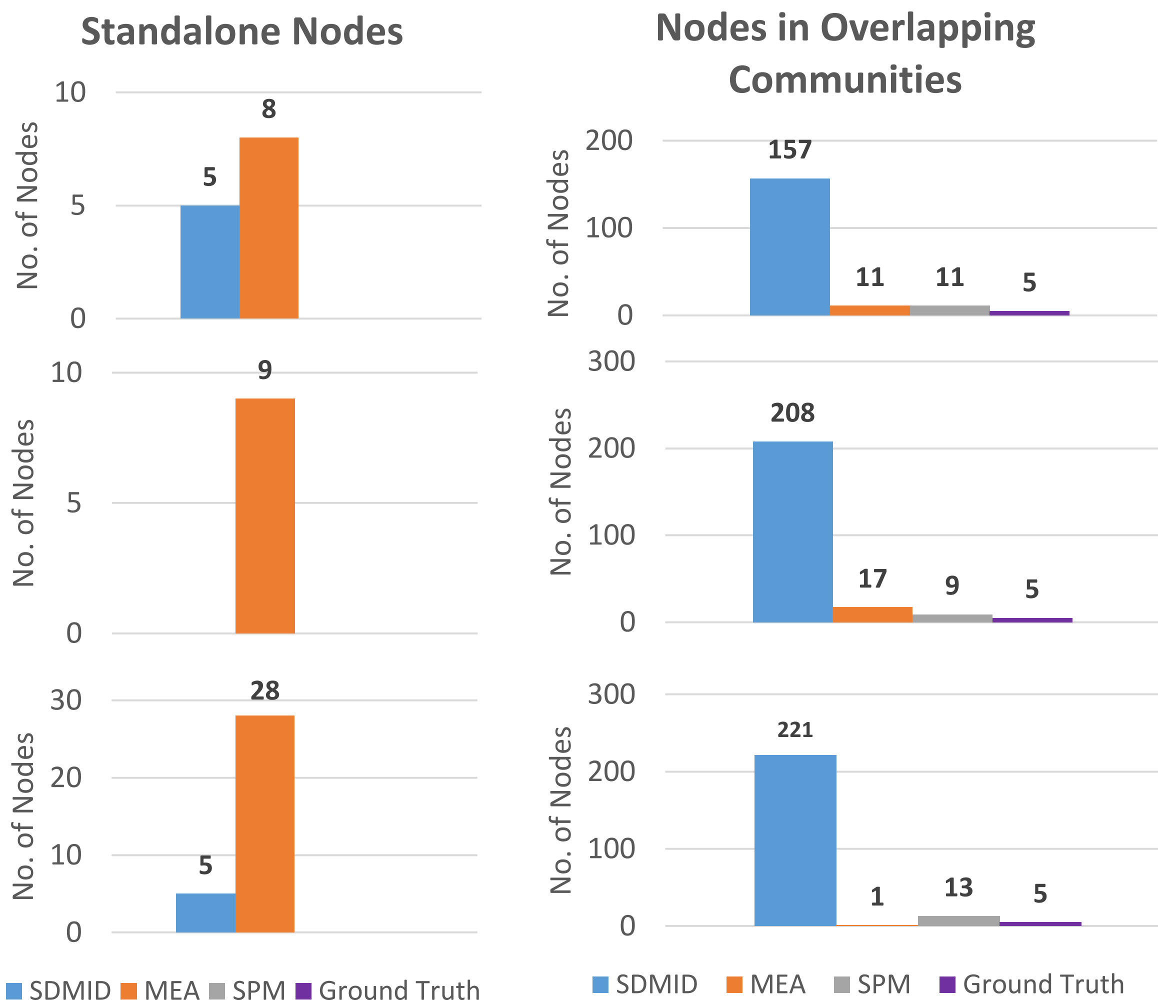


Parameters: n=100, k=3, maxk=6, μ=0.1, t_1=-2.0, t_2=-1.0, minc=5, on=5, om=2, P_-=0.01, P_+=0.01

Mohsen Shahriari,
Ying Li,
Ralf Klamma

Experiments on Benchmark Networks: Community Structure (2)

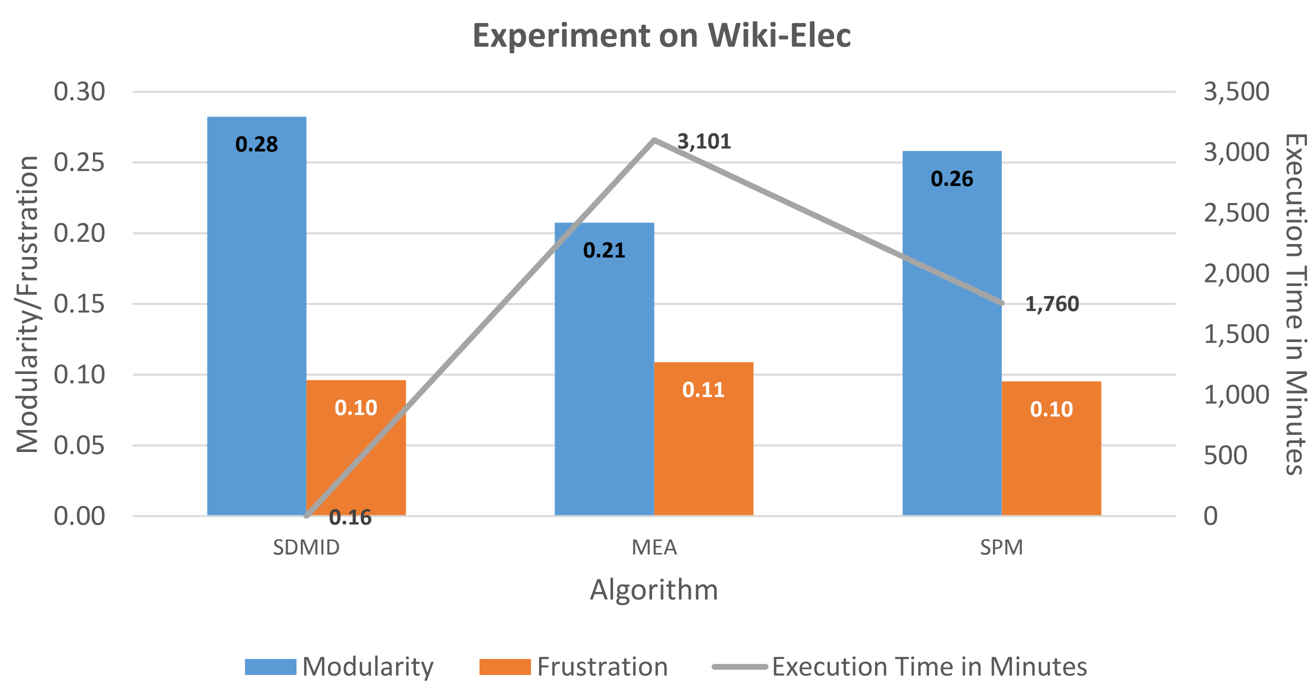
- MEA detects the highest number of standalone nodes
  - SDMID also identifies some of the nodes as standalone
- SDMID assigns most of the nodes as overlapping



Learning Layers

Experiment on Real World Network Wiki-Elec: Metric Values

- SDMID has the highest modularity value
- SDMID and SPM obtain the lowest frustration values
- SDMID is the best regarding the execution time



Analysis of Overlapping Communities in Signed Complex Networks

Mohsen Shahriari, Ying Li, Ralf Klamma

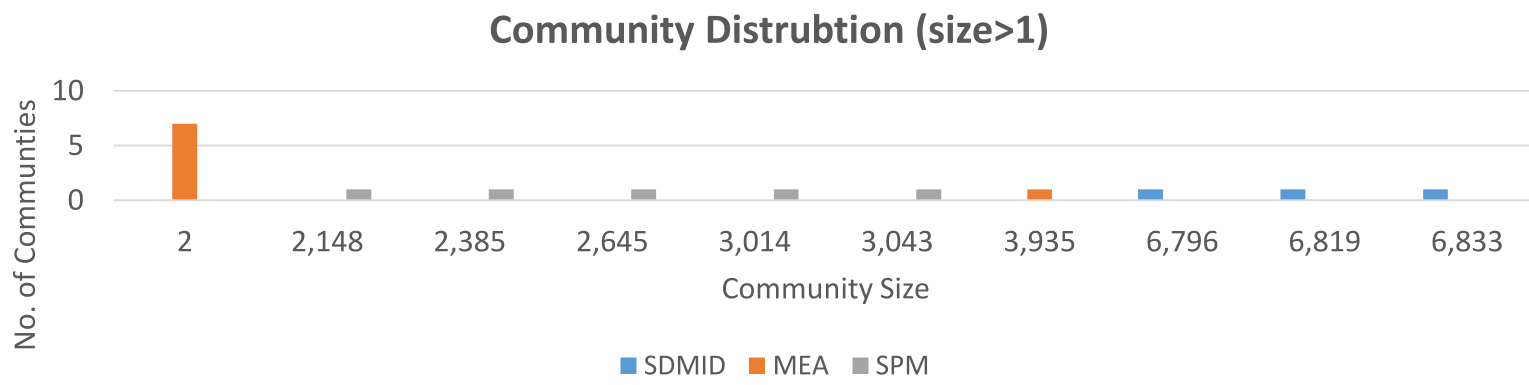
Lehrstuhl Informatik 5 (Information Systems) Prof. Dr. M. Jarke

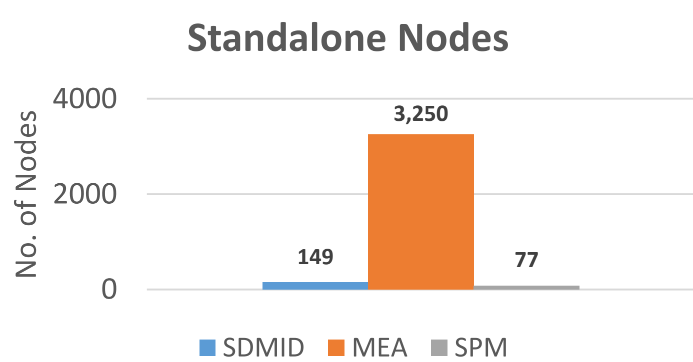
Learning Layers

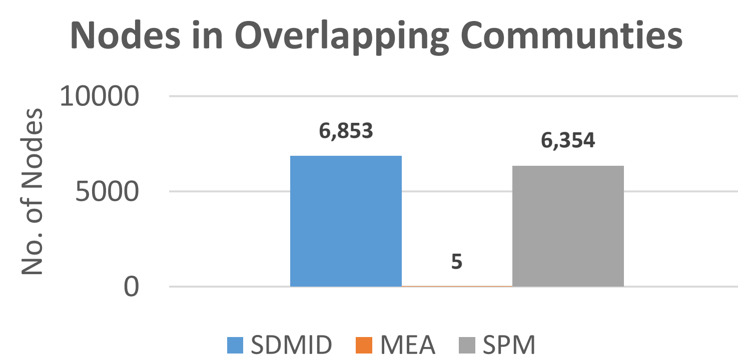
Mohsen Shahriari, Ying Li, Ralf Klamma

Experiments on Real World Network Wiki-Elec: Community Structure

- MEA detects most of the nodes as standalone and most of the nodes are in one community
- Fewest number of standalone nodes observed in SDMID and SPM
- SDMID and SPM approximately detect high number of overlapping ndoes







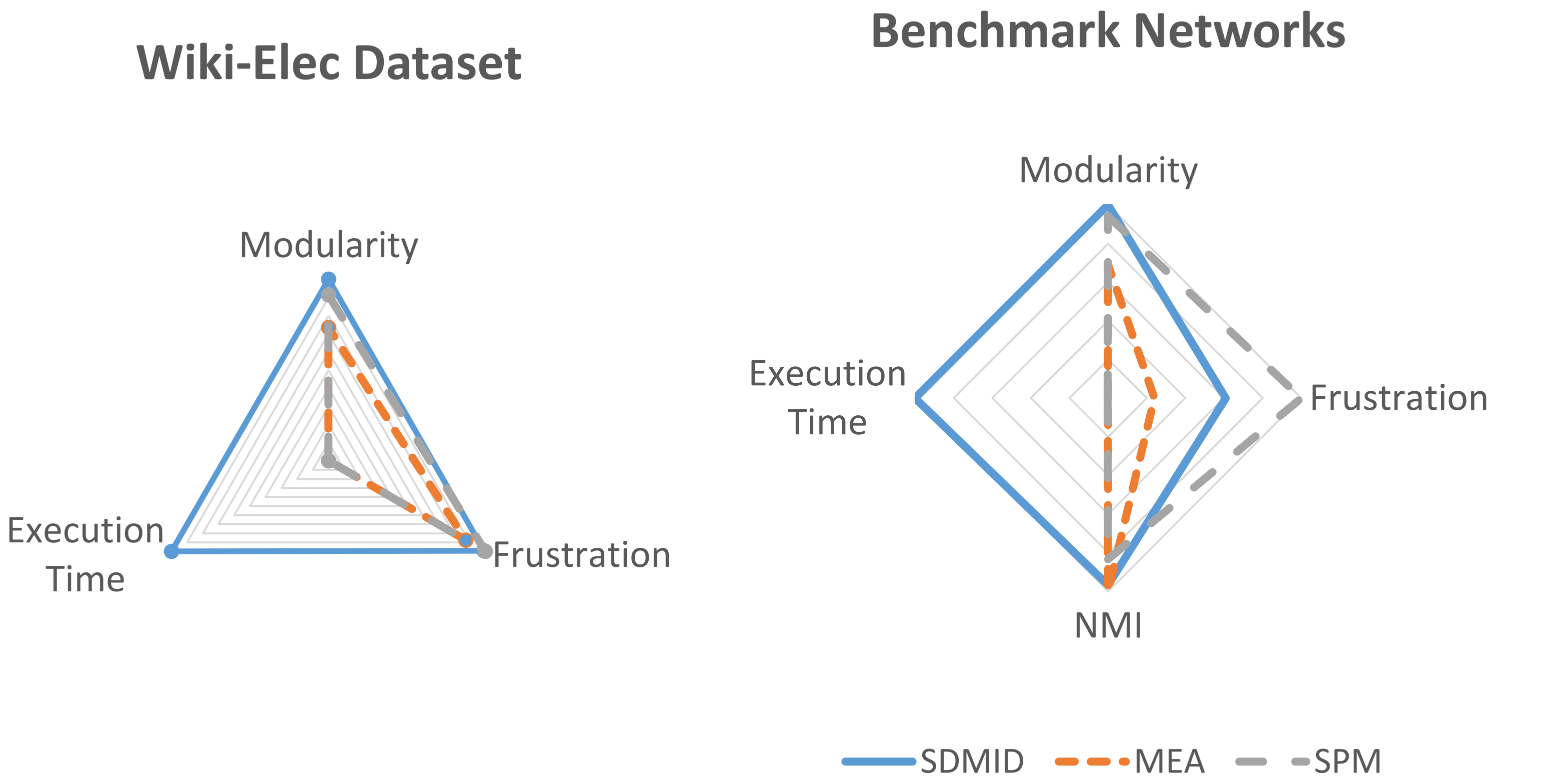
Learning Layers

Analysis of Overlapping Communities in Signed Complex Networks

Mohsen Shahriari, Ying Li, Ralf Klamma

Experiment Summary: Evaluation Radar

- In Wiki-Elec, SDMID has the best performance regarding modularity, execution time and frustration
- In Benchmark networks, SDMI has better performance regarding modularity, execution time and NMI
  - Performance of SPM is better regarding Frustration



Lehrstuhl Informatik 5 (Information Systems) Prof. Dr. M. Jarke

Mohsen Shahriari,
Ying Li,
Ralf Klamma

Experiment Summary: Community Structure

- SDMID
  - Big-sized communities
  - Large areas of overlapping
- MEA_s-SN
  - Small-sized communities
  - Few nodes in the overlapping area
  - Large number of stand-alone nodes
- SPM
  - Predefined number of communities k
  - Large areas of overlapping with a small k

Learning Layers

Analysis of Overlapping Communities in Signed Complex Networks

Mohsen Shahriari, Ying Li, Ralf Klamma

Conclusion & Message

- We compared SDMID, SPM and MEA OCD algorithms from different aspects
- There are few algorithms for overlapping community detection in signed networks
- Currently SDMID and SPM are the best options to be applied on datasets in signed networks
  - SDMID is the fastest and has the highest modularity
  - SDMID obtained the best performance on the real world network Wiki-Elec
- SDMID might be a better choice when diffusion of opinions is preferred across community borders

References

- [CWYT14] Yi Chen, Xiaolong Wang, Bo Yuan and Buzhou Tang. Overlapping Community Detection in Networks with Positive and Negative Links. In: Journal of Statistical Mechanics: Theory and Experiment 2014.3: P03021, 2014.
- [LiLJ14] Chenlong Liu, Jing Liu and Zhongzhou Jiang. A Multiobjective Evolutionary Algorithm Based on Similarity for Community Detection from Signed Social Networks. In:IEEE Transactions on Cybernetics 44.12: pp.2274-2286, 2014.
- [ShKl15] Mohsen Shahriari and Ralf Klamma. Signed Social Networks: Link Prediction and Overlapping Community Detection. In: Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2015.
- [YaCL07] Bo Yang, William K. Cheung, and Jiming Liu. Community Mining from Signed Social Networks. In: IEEE Transactions on Knowledge and Data Engineering 19.10: pp. 1333-1348, 2007.
- [ZhLi07] Qingfu Zhang and Hui Li. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. In:IEEE Transactions on Evolutionary Computation 11.6: pp. 712-731, 2007.

Mohsen Shahriari,
Ying Li,
Ralf Klamma

Thank you !