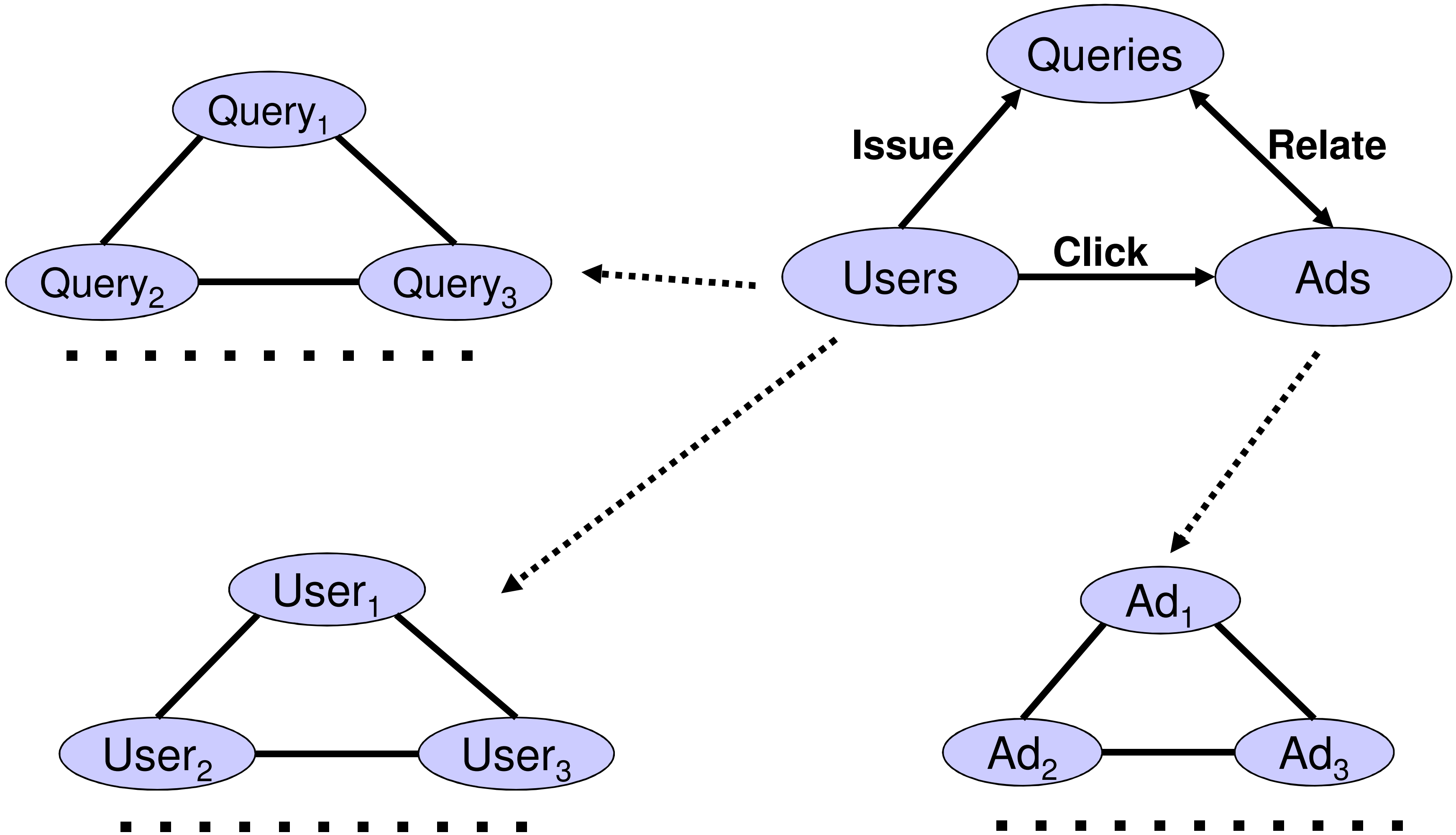
Community Evolution in Dynamic Multi-Mode Networks

Lei Tang, Huan Liu, Jianping Zhang, Zohreh Nazeri

ASU ARIZONA STATE UNIVERSITY

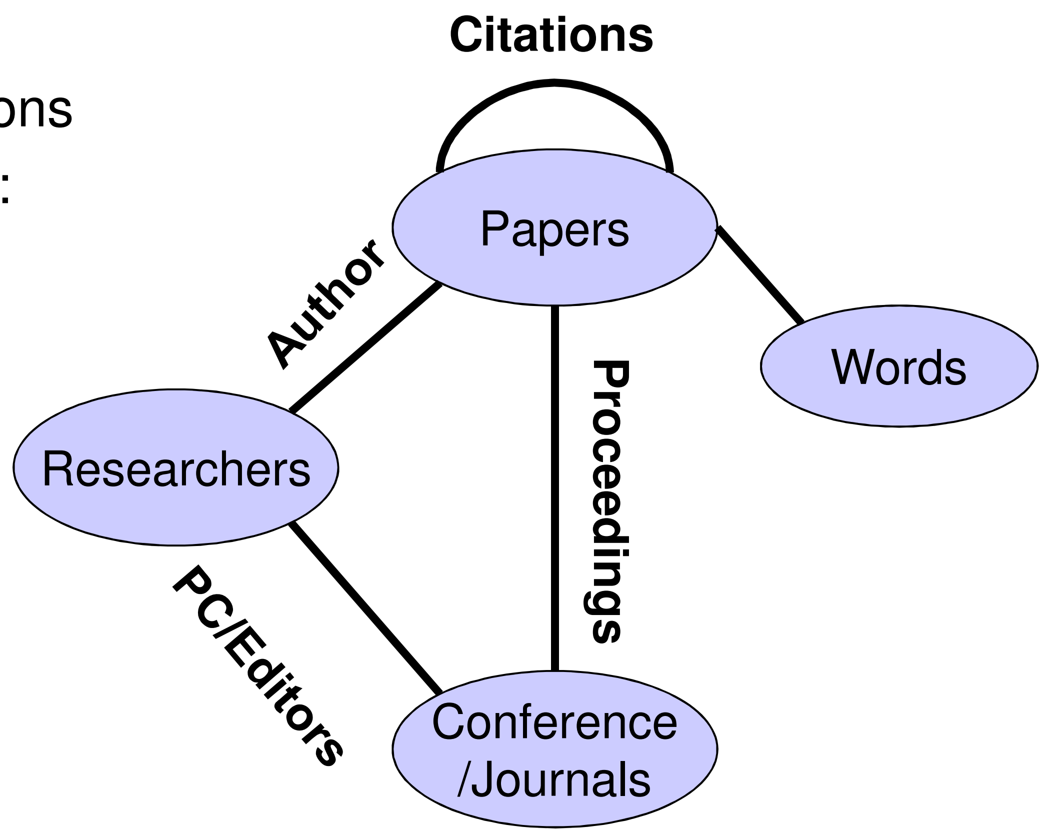
MITRE

One-Mode Network



Multi-Mode Network

- Multi-Mode Network
  - multiple mode of actors
  - Heterogeneous interactions
  - More complicated cases:
    - Actor attributes
    - Self interaction
    - Directed Interaction
- Applications
  - Targeting
    - users, queries, ads
  - Collaborative filtering
    - users, objects



Community Evolution

- Actors in a network tends to form groups/communities.
- Communities of different modes are correlated.
  - Researchers working on data mining attending conferences with similar topics : ICML, KDD, ICDM
- Community membership evolves gradually.
  - A researcher could divert his research interest
  - Hot topics change gradually
- Different modes present different evolution pattern.
  - The venue community is much more stable
- Needs to identify community evolution in dynamic multi-mode networks.

Discovery Community Evolution

- Given:
  - Multiple consecutive snapshots of the multi-mode network
- Output:
  - Identify community membership evolution

- Possible Applications:
  - Detect user interests shift leading to more effective targeting
  - Browse history of networks by showing the long-term trend
  - Anomaly / Buzz detection
  - ……

Spectral Approach

Interaction Approximation

$$R^t_{i,j} \approx C^{(i,t)} A^t_{i,j} (C^{(j,t)})^T$$

Spectral Relaxation $\longrightarrow$

$$\min \quad ||R^t_{i,j} - C^{(i,t)} A^t_{i,j} (C^{(j,t)})^T||^2_F$$

s.t. $\quad (C^{(i,t)})^T C^{(i,t)} = I_{k_i}$

+

Temporal Smoothness

$$\cancel{||C^{(i,t)} - C^{(i,t-1)}||^2_F}$$

$$||C^{(i,t)} (C^{(i,t)})^T - C^{(i,t-1)} (C^{(i,t-1)})^T||^2_F$$

$$\mathbf{F_2}: \quad \min \sum_{t=1}^{\mathrm{T}} \sum_{i<j} w_a^{(i,j)} ||R^t_{i,j} - C^{(i,t)} A^t_{i,j} (C^{(j,t)})^T||^2_F$$

$$+ \frac{1}{2} \sum_{i=1}^{m} w_b^{(i)} \sum_{t=2}^{\mathrm{T}} ||C^{(i,t)} (C^{(i,t)})^T - C^{(i,t-1)} (C^{(i,t-1)})^T||^2_F$$

$$s.t. \quad (C^{(i,t)})^T C^{(i,t)} = I_{k_i}$$

$$with \quad i = 1, \cdots, m, \quad t = 1, \cdots \mathbb{T}$$

Solution

- Difficult to find global solution
- Much easier when performing block alternating optimization
  - Optimal A_{ij} can be solved given C_i and C_j at timestamp t

$$A^t_{i,j} = (C^{(i,t)})^T R^t_{i,j} C^{(j,t)}$$

  - Optimal C_i at time t can be computed giv…

THEOREM 3. *Given* $C^{(j,t)}$ *and* $C^{(i,t\pm1)}$ … *puted as the top left singular vectors of th*… *catenated by the following matrices in column*…*ise.*

$$\left[\left\{\sqrt{w_a^{(i,j)}}R^t_{i,j}C^{(j,t)}\right\}_{i<j},\left\{\sqrt{w_a^{(k,i)}}(R^t_{k,i})^T C^{(k,t)}\right\}_{k<i},\sqrt{w_b^{(i)}}C^{(i,t\pm1)}\right]$$

Other cluster indicators are just attributes

Extensions to realistic cases

- Online Clustering

$$\left[\left\{\sqrt{w_a^{(i,j)}}R_{i,j}^{t}C^{(j,t)}\right\}_{i<j},\left\{\sqrt{w_a^{(k,i)}}(R_{k,i}^{t})^{T}C^{(k,t)}\right\}_{k<i},\sqrt{w_b^{(i)}}C^{(i,t\pm1)}\right]$$

- Inactive Actors
  - Delete the corresponding entry in C_i
- Emerging Actors
  - Add an entry in C_i with default value 0
- Actor Attributes
  - $$\left[\left\{\sqrt{w_a^{(i,j)}}R_{i,j}^{t}C^{(j,t)}\right\}_{j\neq i},\sqrt{w_b^{(i)}}C^{(i,t\pm1)},\sqrt{w_c^{(i)}}F_i^{t}\right]$$
- Within-mode Interaction
  - Add to the similarity matrix calculated via “attributes”

Algorithm

Algorithm: Evolutionary Multi-mode Clustering

Input: R, k_i, $w_a^{(i,j)}$, $w_b^{(i)}$
Output: $idx^{(i,t)}$, $C^{(i,t)}$, $A_{i,j}^t$.

1. Generate initial cluster indicator matrix $C^{(i,t)}$.
2. **Repeat**
3. **For** $t = 1 : \mathbb{T}$, $i = 1 : m$
4. shrink / expand $C^{(i,t\pm 1)}$ if necessary;
5. calculate P_i^t (or M_i^t) as in Theorem 3;
6. calculate SVD of P_i^t (or eigen vectors of M_i^t);
7. update $C^{(i,t)}$ as top left singular (eigen) vectors;
8. **Until** the relative change of the objective (**F**3) $\leq \epsilon$.
9. calculate $A_{i,j}^t$ as in Theorem 2
10. calculate the cluster $idx^{(i,t)}$ with k-means on $C^{(i,t)}$

Experiments

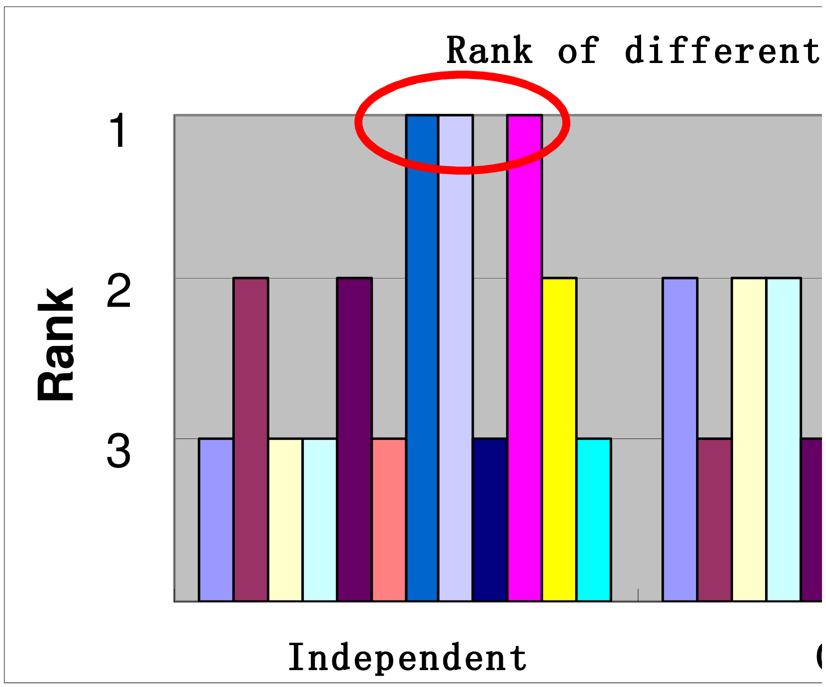
- Two publically available real-world data
  - Enron data (Apr. 2001 – Mar.2002)
    - 3 modes: people, email, words
  - DBLP data (1980 – 2004)
    - 4 modes: authors(347013) , papers(491726), venues(2826), terms(9523)

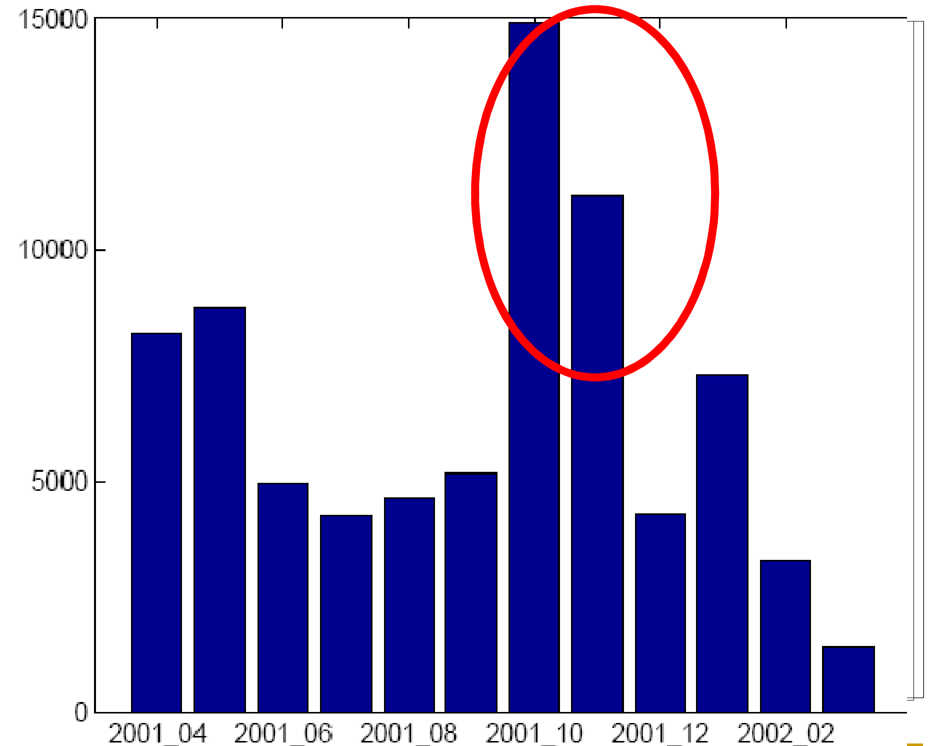
- Methods:
    - Independent clustering (without temporal information)
    - Online Clustering (only consider temporal information in the past)
    - Evolutionary Clustering

- Evaluation
  - No ground truth
  - Adopt "cross-validation" strategy
  - Relative measure to compare different methods:
    - constant - approximation error (the larger, the better)

Performance on Enron

- Evolutionary Clustering consistently approximates the Interaction better most of the time.
- Independent Clustering outperforms others when there's enough interaction traffic.





Performance on DBLP

- 11 out of 15 years, evolutionary outperforms other clustering approaches.
- The other 4 winners are online clustering.
- Example:
  - NIPS:
    - Aligned with Neural Network conferences in 1995
    - More with machine learning in 2004

NIPS, COLT, ECML, ICML, ICANN,
IJCIA, IWANN, Machine Learning,
Inter. C. on Algorithmic Learning Theory,
Int. J. Neural Syst., Neural Computation,
Int. J. Computational Intelligence and Applications.

Computation Time

- Algorithm typically converges in few iterations
- All three methods demonstrate computation time of the same order
- The majority of the computation time is actually spent on K-means rather than SVD

Table 4: Computation Time

Method	Independent	Online	Evolutionary
Enron	5.0699×10^3	8.4974×10^3	1.1076×10^4
DBLP	2.1033×10^3	2.6945×10^3	5.0491×10^3

Conclusions

- A spectral approach to address community evolutionary clustering in dynamic multi-mode networks.
- Easy to extend to handle hibernating/emerging actors, actor attributes, within-mode interaction.
- Empirically find more accurate community structure.

- In this framework, we only capture the membership change. Currently trying to develop new algorithm to simultaneously detect
  - Micro-evolution (membership change)
  - Macro-evolution (group interaction change)