Community Discovery in Dynamic Rich-context Social Networks

Yu-Ru Lin, Ph.D. Candidate **Advisor: Hari Sundaram**
Computer Science + Arts Media Engineering, Arizona State University

Motivation and Context

Contemporary *social media* culture

Grass-root network | Social movement | Decision making

Data characteristics

- Voluminous
- Dynamic
- Context-rich

New challenges

- Lack of well-defined attributes
- Limitation in network analysis
- Scalability requirement

What is a community?

A *cluster* of people with *sustained relationship* among them

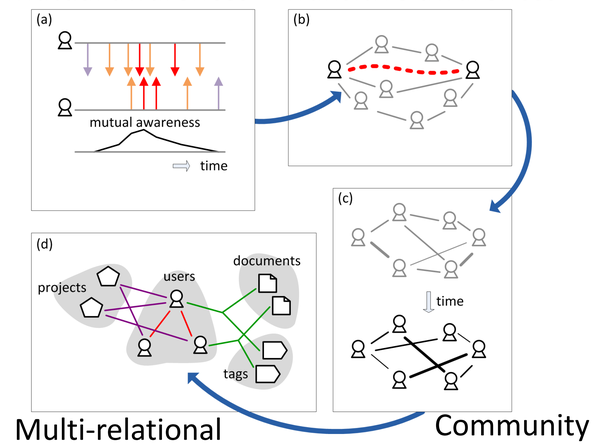
Summary

Goal

Large scale community discovery in the context of everyday social media use

Current work

Mutual awareness | Transitive awareness

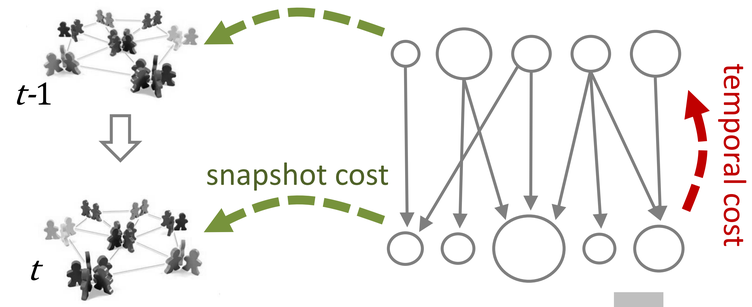


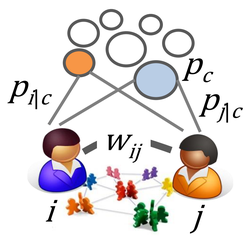
Contribution

Leveraging interdisciplinary observations into developing new mining framework

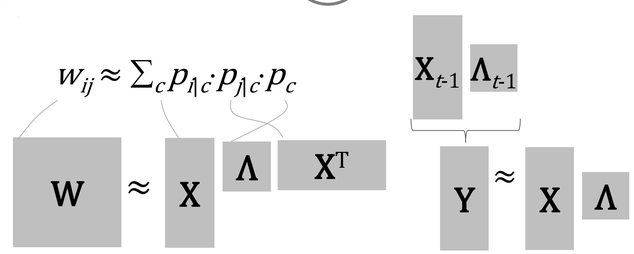
Community Evolution

Extracting *sustained evolving* communities





$w_{ij} \approx \sum_c p_{i|c} \cdot p_{j|c} \cdot p_c$

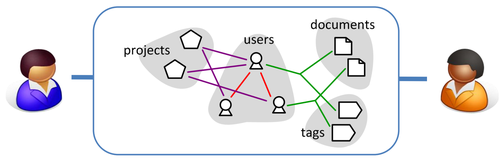


objective function minimize $\alpha\, D(\mathbf{W} \,||\, \mathbf{X}\Lambda\mathbf{X}^T) + (1-\alpha)\, D(\mathbf{Y} \,||\, \mathbf{X}\Lambda)$

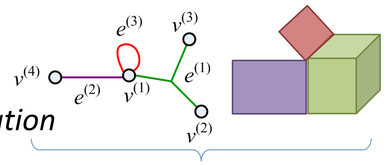
KL divergence

Multi-relational Communities

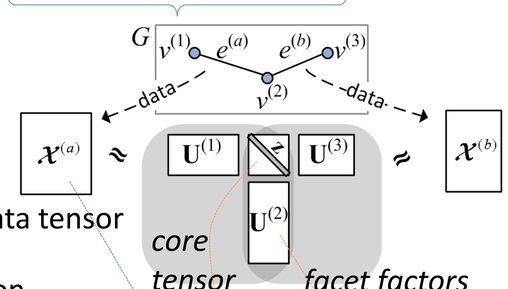
Tracking communities from *rich context*



matagraph
node: *facet*
hyperedge: *relation*



matafac
multi-tensor factorization on a metagraph



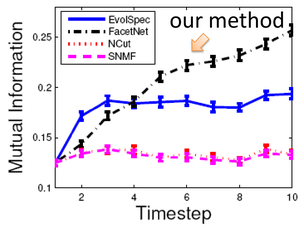
objective function
minimize $\text{cost}(G) = \sum_{r \in E} D(\mathcal{X}^{(r)} \,||\, [\![\mathbf{z}]\!] \prod_{m: v^{(m)} \sim e^{(r)}} \times_m \mathbf{U}^{(m)})$

KL divergence

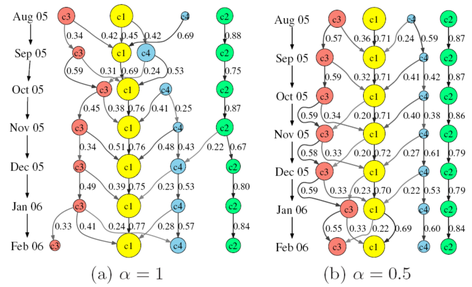
Results

Clustering accuracy

on synthetic evolving networks

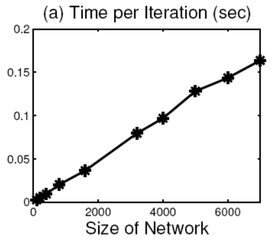


Community evolution in blog networks



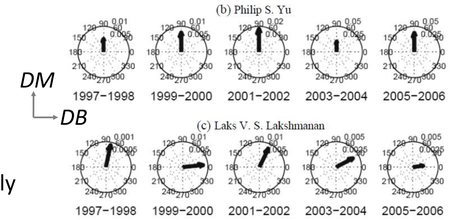
use historic information to prevent drastic change

Scalability



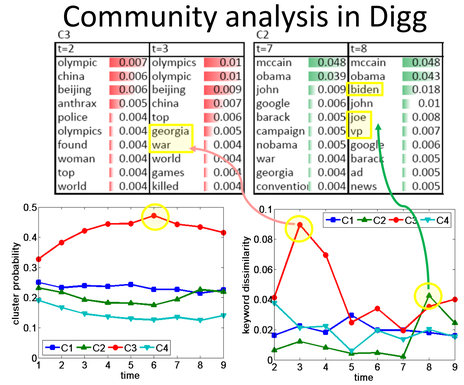
running time scales linearly with network size

Time-dependent ranking in DBLP

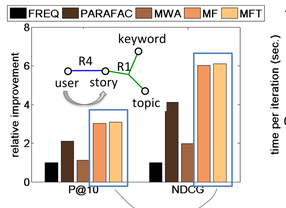


Results

Community analysis in Digg

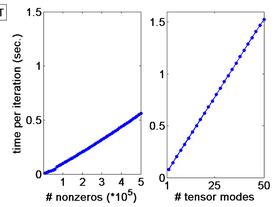


Digg prediction



our method

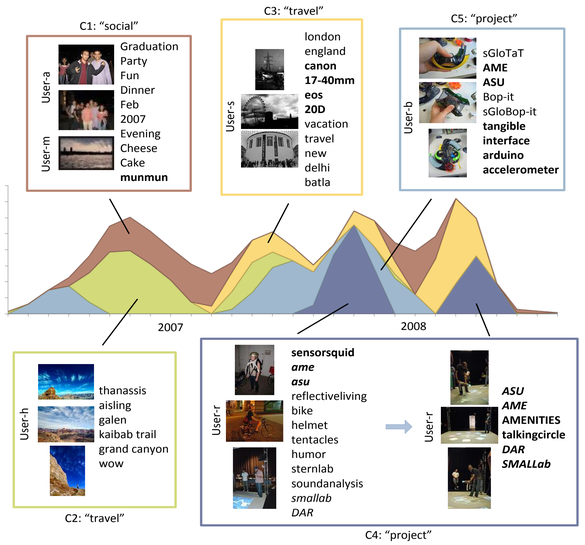
Scalability



z, {**U**} can be solved with linear time complexity

Social media artifact organization

user study using Flickr group data



Contact: yu-ru.lin@asu.edu