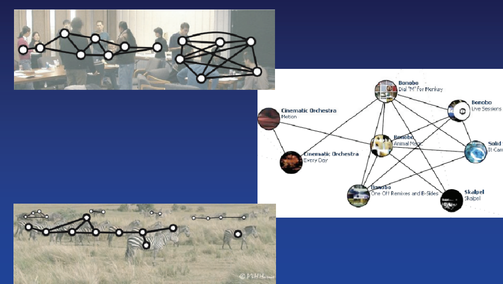
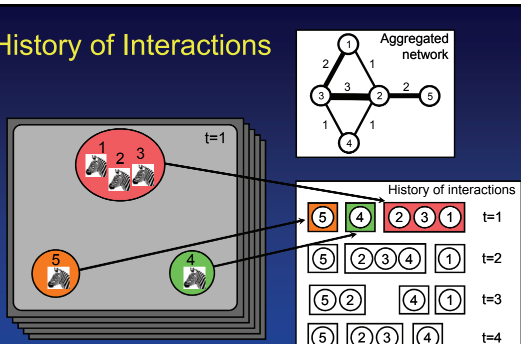
A Framework for Finding Communities in Dynamic Social Networks

Slides from Chayant Tantipathananandh, Tanya Berger-Wolf
University of Illinois at Chicago
David Kempe
University of Southern California

Social Networks



History of Interactions



Assume discrete time and interactions in form of complete subgraphs.

Community Identification

What is community?

"Cohesive subgroups are subsets of actors among whom there are relatively strong, direct, intense, frequent, or positive ties." [Wasserman & Faust '97]

Notions of communities:

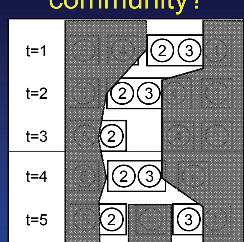
Static

- Centrality and betweenness [Girvan & Newman '01]
- Correlation clustering [Basal et al. '02]
- Overlapping cliques [Palla et al. '05]

Dynamic

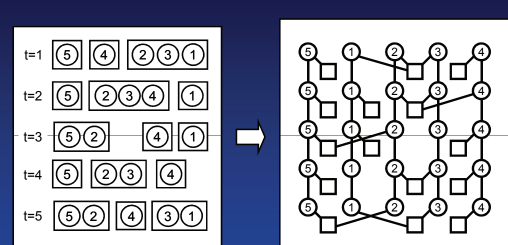
- Metagroups [Berger-Wolf & Saia '06]

The Question: What is dynamic community?



- A dynamic community is a subset of individuals that stick together over time.
- NOTE: Communities ≠ Groups

Approach: Graph Model



Approach: Assumptions

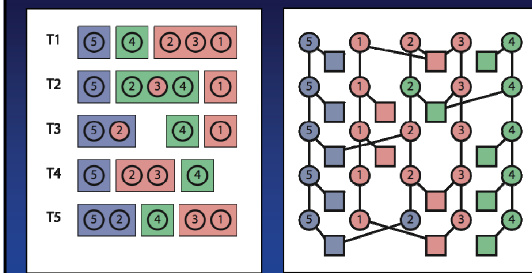
Required

- Individuals and groups represent exactly **one** community at a time.
- Concurrent groups represent **distinct** communities.

Desired

- **Conservatism:** community affiliation changes are rare.
- **Group Loyalty:** individuals observed in a group belong to the same community.
- **Parsimony:** few affiliations overall for each individual.

Approach: Color = Community



Valid coloring: distinct color of groups in each time step

Approach: Assumptions

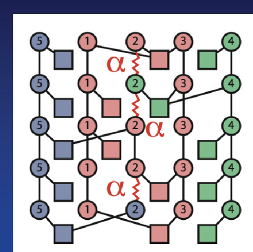
Required

- Individuals and groups represent exactly **one** community at a time.
- Concurrent groups represent **distinct** communities.

Desired

- **Conservatism:** community affiliation changes are rare.
- **Group Loyalty:** individuals observed in a group belong to the same community.
- **Parsimony:** few affiliations overall for each individual.

Costs



- Conservatism: switching cost (α)
- Group loyalty:
  - Being absent (β1)
  - Being different (β2)
- Parsimony: number of colors (γ)

Approach: Assumptions

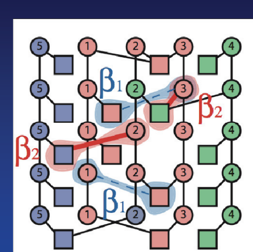
Required

- Individuals and groups represent exactly **one** community at a time.
- Concurrent groups represent **distinct** communities.

Desired

- **Conservatism:** community affiliation changes are rare.
- **Group Loyalty:** individuals observed in a group belong to the same community.
- **Parsimony:** few affiliations overall for each individual.

Costs



- Conservatism: switching cost (α)
- Group loyalty:
  - Being absent (β1)
  - Being different (β2)
- Parsimony: number of colors (γ)

Approach: Assumptions

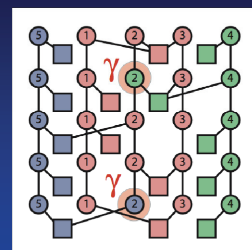
Required

- Individuals and groups represent exactly **one** community at a time.
- Concurrent groups represent **distinct** communities.

Desired

- **Conservatism:** community affiliation changes are rare.
- **Group Loyalty:** individuals observed in a group belong o the same community.
- **Parsimony:** few affiliations overall for each individual.

Costs



- Conservatism: switching cost (α)
- Group loyalty:
  - Being absent (β1)
  - Being different (β2)
- Parsimony: number of colors (γ)

Problem Definition

- Minimum Community Interpretation
  For a given cost setting, (α,β1,β2,γ), find vertex coloring that minimizes total cost.
  - Color of group vertices = Community structure
  - Color of individual vertices = Affiliation sequences
- Problem is NP-Complete and APX-Hard
  - Reduction from Minimum Multiway Cut problem

Model Validation and Algorithms

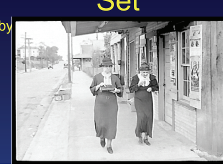
- Model validation: exhaustive search for an exact minimum-cost coloring
  - Fix group coloring and find optimal coloring for individuals using DP
- Heuristic algorithms evaluation: compare heuristic results to OPT.
  - Bipartite Matching Heuristic
  - Greedy Heuristics

Results

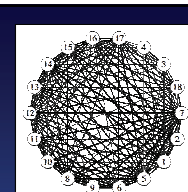
- Synthetic Data sets
  - Assembly Line
  - Dutiful Children
- Real-World Data sets
  - Southern Women
  - Grevy's Zebra Data set

Southern Women Data Set

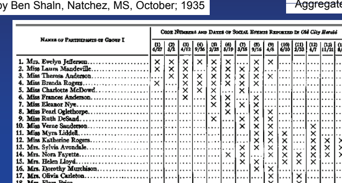
by 1



Photograph by Ben Shaln, Natchez, MS, October; 1935



Aggregated network

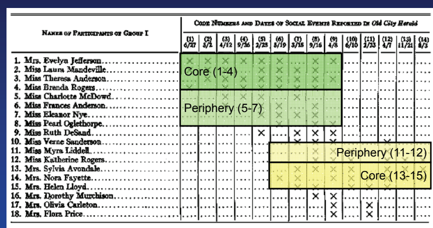


NAMES OF PARTICIPANTS OF GROUP I	CODE NUMBERS AND DATES OF SOCIAL EVENTS REPORTED IN *Old City Herald*													
	(1) 6/27	(2) 3/2	(3) 4/12	(4) 9/26	(5) 2/25	(6) 5/19	(7) 3/15	(8) 9/16	(9) 4/8	(10) 6/10	(11) 2/23	(12) 4/7	(13) 11/21	(14) 8/3
1. Mrs. Evelyn Jefferson	X	X	X	X	X	X		X	X					
2. Miss Laura Mandeville	X	X	X		X	X	X	X						
3. Miss Theresa Anderson		X	X	X	X	X	X	X	X					
4. Miss Brenda Rogers	X		X	X	X	X	X	X						
5. Miss Charlotte McDowd			X	X	X		X							
6. Miss Frances Anderson			X		X	X		X						
7. Miss Eleanor Nye					X	X	X	X						
8. Miss Pearl Oglethorpe						X		X	X					
9. Miss Ruth DeSand					X		X	X	X					
10. Miss Verne Sanderson							X	X	X			X		
11. Miss Myra Liddell								X	X	X		X		
12. Miss Katherine Rogers								X	X	X		X	X	X
13. Mrs. Sylvia Avondale							X	X	X	X		X	X	X
14. Mrs. Nora Fayette						X	X		X	X	X	X	X	X
15. Mrs. Helen Lloyd							X	X		X	X	X		
16. Mrs. Dorothy Murchison								X	X					
17. Mrs. Olivia Carleton									X		X			
18. Mrs. Flora Price									X		X			

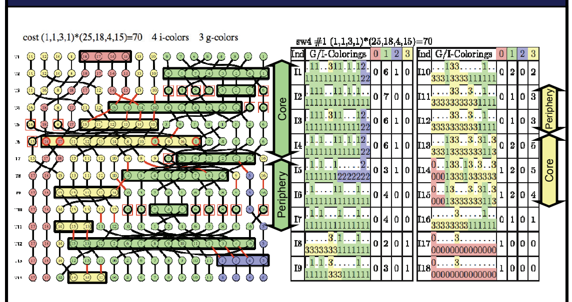
Event participation

Ethnography

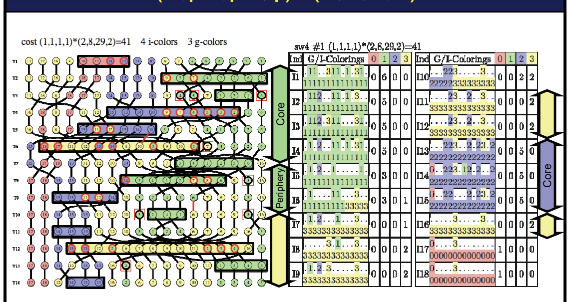
by Davis, Gardner, and Gardner, 1941



An Optimal Coloring: (α,β1,β2,γ)=(1,1,3,1)



An Optimal Coloring: (α,β1,β2,γ)=(1,1,1,1)



Conclusions

- An optimization-based framework for finding communities in dynamic social networks.
- Finding an optimal solution is NP-Complete and APX-Hard.
- Model evaluation by exhaustive search.
- Heuristic algorithms for larger data sets. Heuristic results comparable to optimal.

Discussion

- Scalability – datasets considered are small. What about large datasets? E.g., Facebook
- Assumption are too tight: Relaxation of their constraints
  - Individuals and groups represent exactly **one** community at a time
  - How does it effect scalability?
- What next after finding communities? Is their work motivating?

Thank You