Utilizing Group Affinity to Predict Community Formation in Social Networks

Joseph Leung

Department of Mathematics BYU

Student Research Conference

2019

Collaborators

Faculty	Students
Ben Webb	Julia Bohman
Emily Evans	Jason Kinghorn

Motivation

Motivation: Given an individual's previous group/community affiliations can we predict which group/community they will belong to in the future?

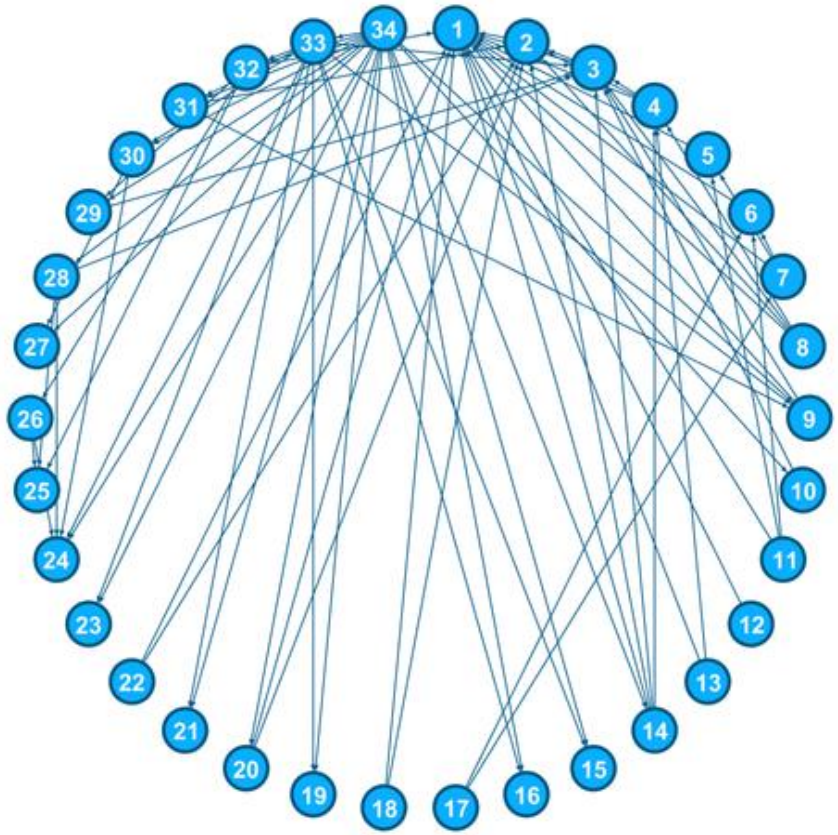
Motivation

Motivation: Given an individual's previous group/community affiliations can we predict which group/community they will belong to in the future?

Potential Applications:

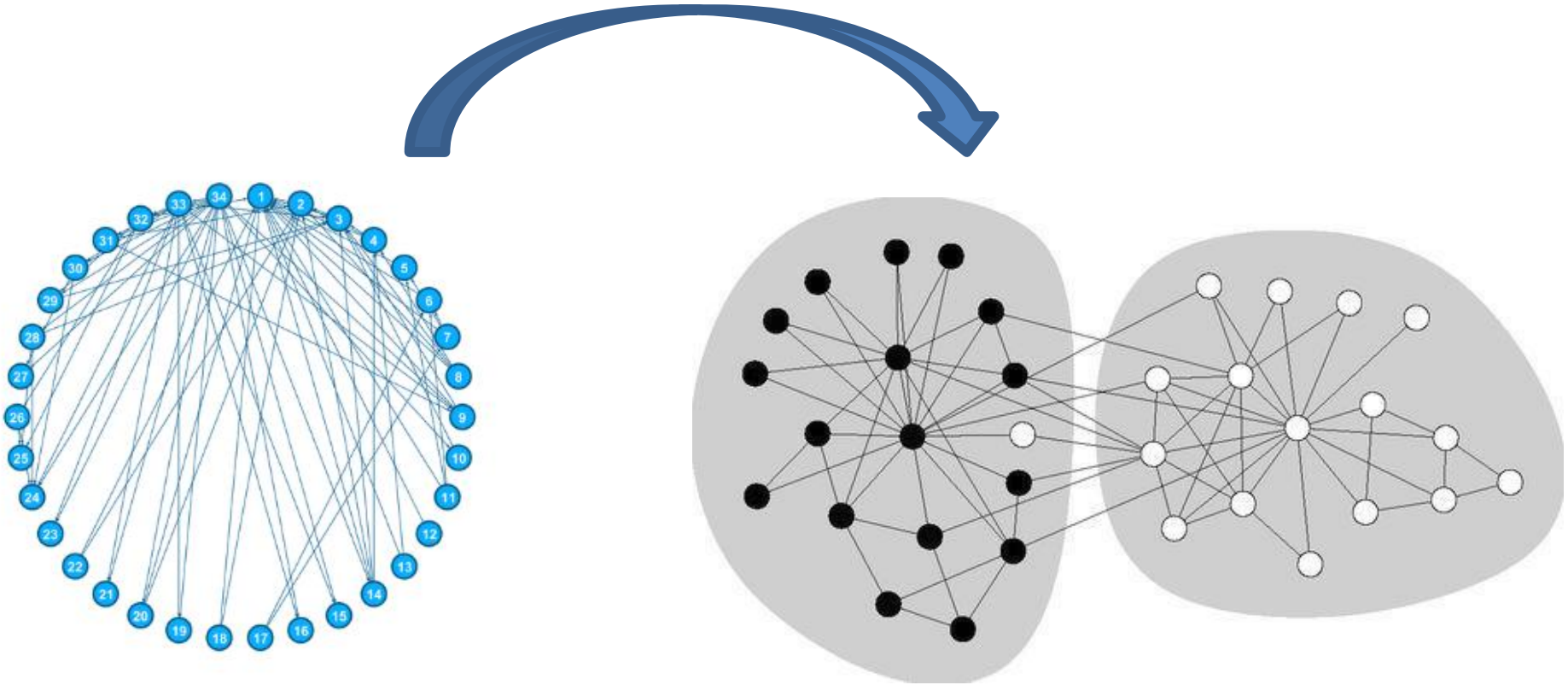
- Organizational structure
  - Terrorist Groups
  - Workplace Formation
  - School study groups

Zachary's Karate Club



https://en.wikipedia.org/wiki/Zachary%27s_karate_club

Zachary's Karate Club



- Community detection on Zachary's Karate Club with Simple Modularity Maximization

Networks: An Introduction(Newman)

Southern Women Dataset

Observed 18 women over the course of a year:

- Bi-Partite Graph
- Different communities depending on the algorithm
- Allows us to glimpse at the communities at periodic time periods

Southern Women Dataset

Observed 18 women over the course of a year:

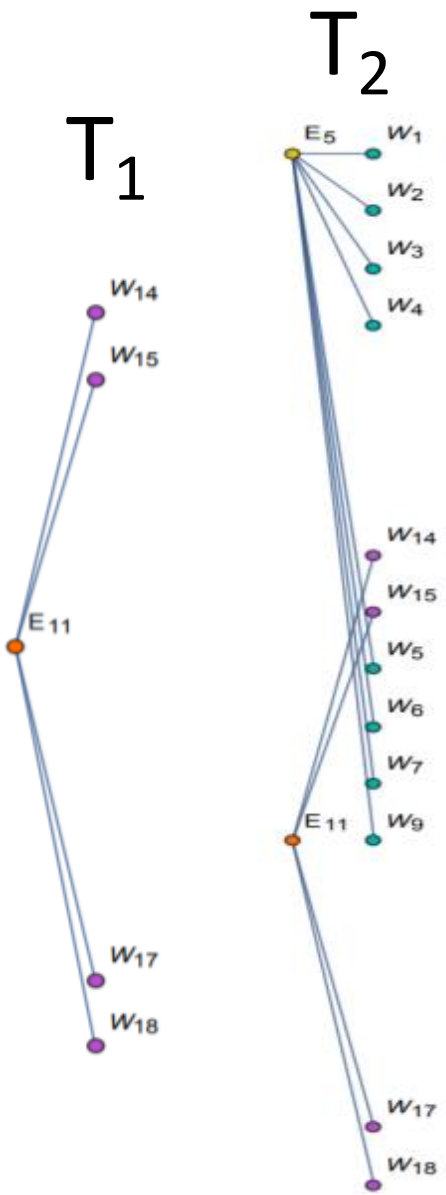
- Bi-Partite Graph
- Different communities depending on the algorithm
- Allows us to glimpse at the communities at periodic time periods

Gathered in 1941 by five sociologists who observed the women attend different events

Southern Women Dataset

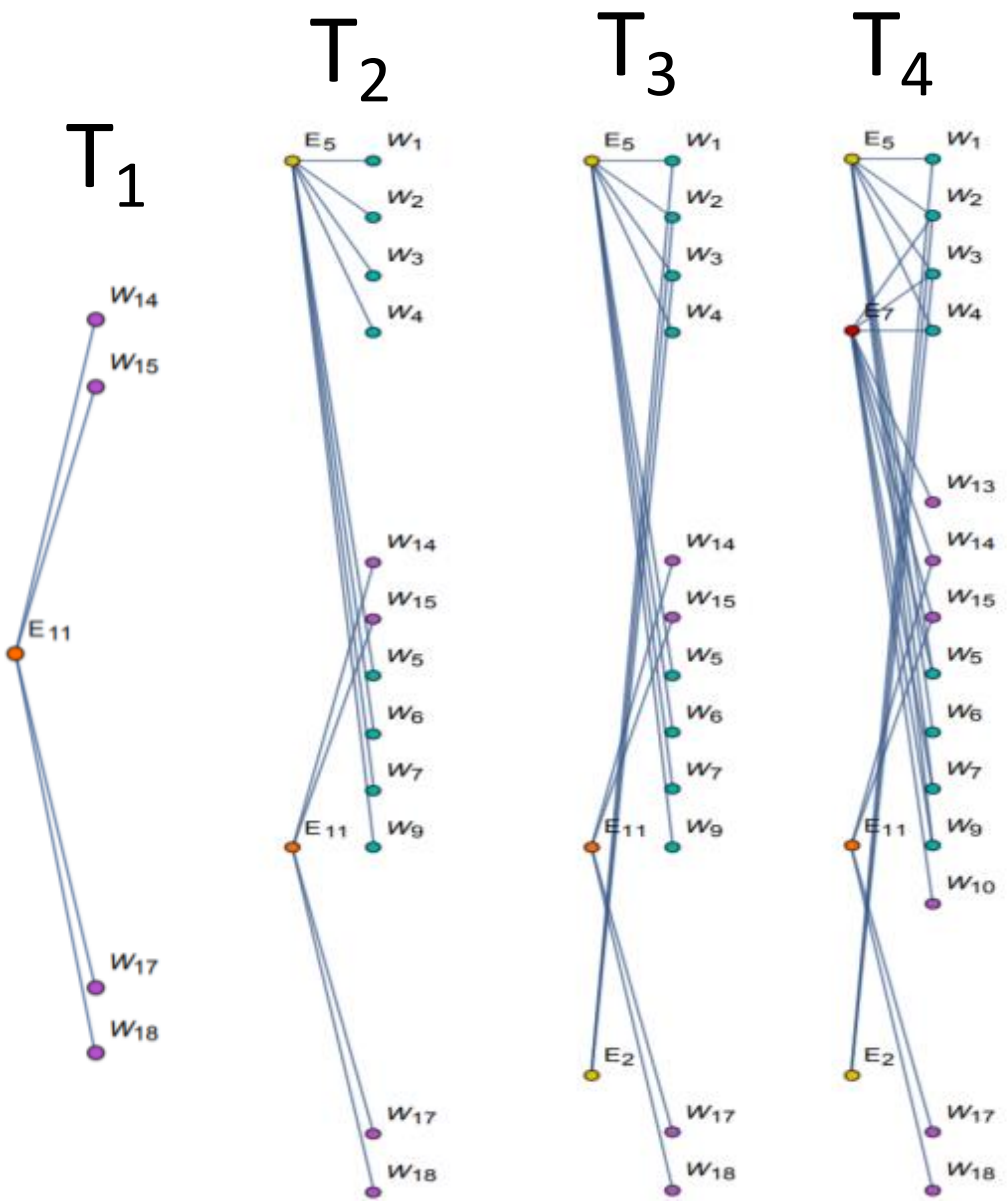
T_1

Southern Women Dataset

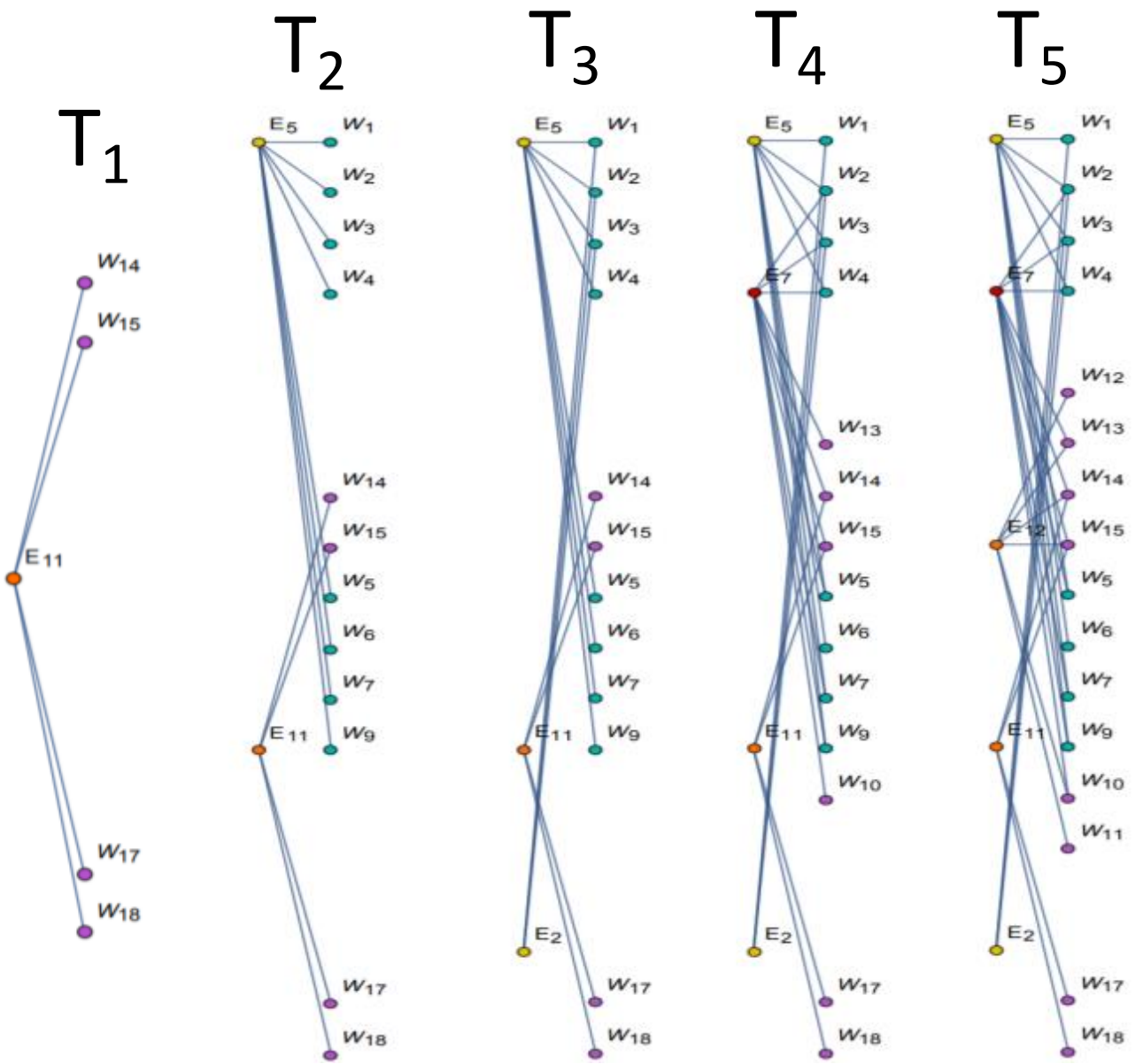


Southern Women Dataset

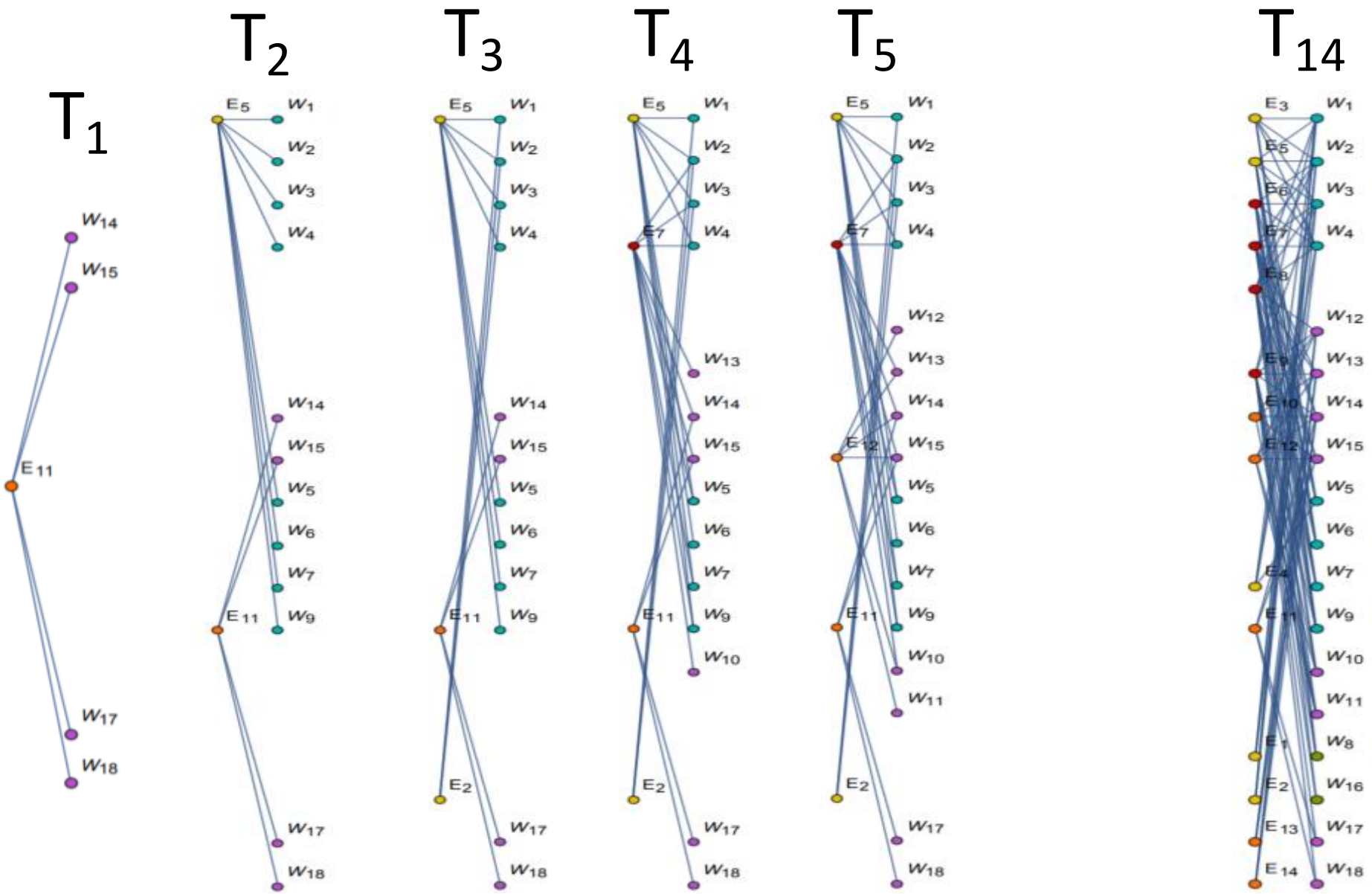
Southern Women Dataset



Southern Women Dataset



Southern Women Dataset



Community Detection Algorithms

There are many algorithms for community detection.

- **Spectral Partitioning**
- **Kernighan-Lin Algorithm**
- Simple Modularity Maximization
- Fluid Communities

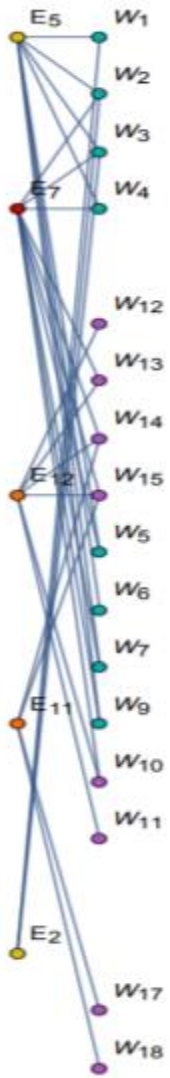
Southern Women Dataset – 2nd Event

Kernighan-Lin/
Spectral Partitioning
Division

Southern Women Dataset – 5th Event

Kernighan-Lin Division

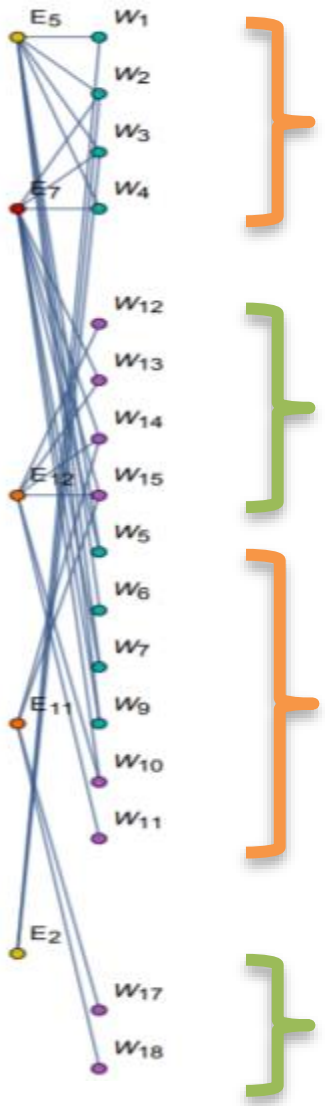
T_5



Southern Women Dataset – 5th Event

Kernighan-Lin Division

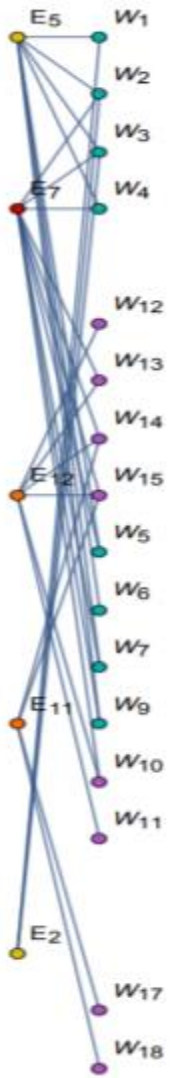
T_5



Southern Women Dataset – 5th Event

Spectral Partitioning Division

T_5



Southern Women Dataset – 5th Event

Spectral Partitioning Division

T_5

Affinity Score

Determining an Affinity Score:

A measurement that suggests how inclined an individual is to one group or the other

Affinity Score

Determining an Affinity Score:

- A measurement that suggests how inclined an individual is to one group or the other
  - Key information in predicting how the communities are ultimately determined

Affinity Score

Determining an Affinity Score:

- A measurement that suggests how inclined an individual is to one group or the other
  - Key information in predicting how the communities are ultimately determined

- No general rule for this in current literature

Affinity Score

Determining an Affinity Score:

- A measurement that suggests how inclined an individual is to one group or the other
  - Key information in predicting how the communities are ultimately determined
- Based on nature of the Community Detection Algorithm

Affinity Score

Determining an Affinity Score:

A measurement that suggests how inclined an individual is to one group or the other

Key information in predicting how the communities are ultimately determined

Based on nature of the Community Detection Algorithm

- Kernighan Lin Algorithm (modularity based algorithms)

Ran several times to find a probability of belonging to one group or the other

Affinity Score

Determining an Affinity Score:

A measurement that suggests how inclined an individual is to one group or the other

Key information in predicting how the communities are ultimately determined

Based on nature of the Community Detection Algorithm

- Kernighan Lin Algorithm (modularity based algorithms)

  Ran several times to find a probability of belonging to one group or the other

- Spectral Partitioning

  Used the eigenvalues of the Graph Laplacian as a score

Affinity Score

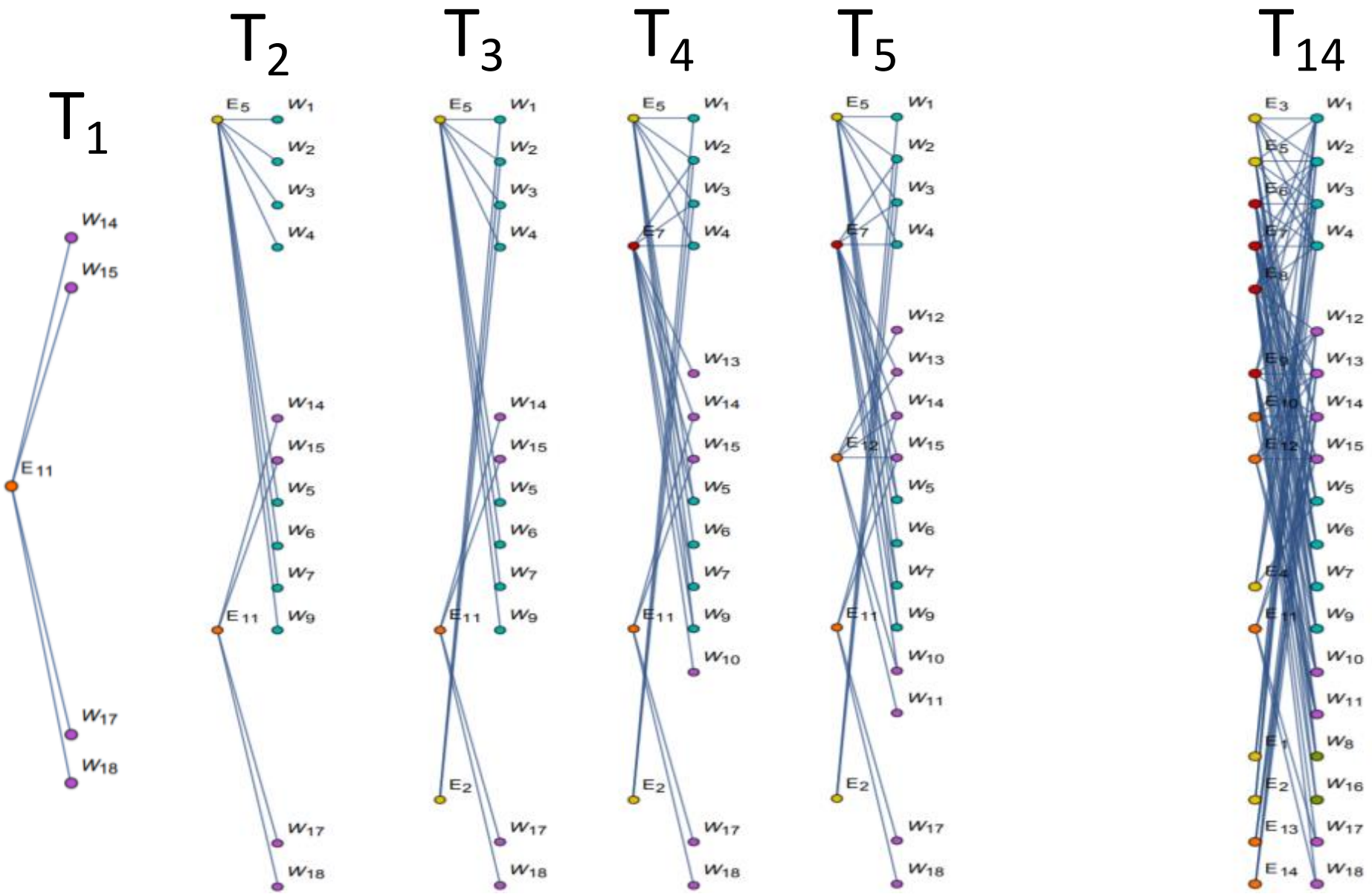
	Afg14	Afg15	Afg16	Afg17	Afg18	Afg19	Afg110	Afg111	Afg112	Afg113	Afg114
W14	3.63E-02	0.0280551	6.21E-02	6.21E-02	1.58E-17	1.58E-17	1.58E-17	1.58E-17	1.36E-17	1.36E-17	1.36E-17
W15	3.63E-02	0.0280551	2.51E-02	2.51E-02	-3.01E-02	-3.01E-02	-3.01E-02	-3.01E-02	-1.15E-16	-1.15E-16	-1.15E-16
W17	6.36E-01	0.6461203	2.17E-01	2.17E-01	-2.75E-01	-2.75E-01	-2.75E-01	-2.75E-01	4.47E-01	4.47E-01	4.47E-01
W18	6.36E-01	0.6461203	2.17E-01	2.17E-01	-2.75E-01	-2.75E-01	-2.75E-01	-2.75E-01	4.47E-01	4.47E-01	4.47E-01
W1	-1.84E-01	-0.1695486	-2.23E-03	-2.23E-03	3.06E-03	3.06E-03	3.06E-03	3.06E-03	-1.35E-15	-1.35E-15	-1.35E-15
W2	-1.26E-01	-0.1154864	-2.18E-01	-2.18E-01	2.45E-01	2.45E-01	2.45E-01	2.45E-01	-1.52E-01	-1.52E-01	-1.52E-01
W3	-1.26E-01	-0.1154864	-2.44E-17	-2.44E-17	9.14E-17	9.14E-17	9.14E-17	9.14E-17	7.21E-17	7.21E-17	7.21E-17
W4	-1.26E-01	-0.1154864	-2.18E-01	-2.18E-01	2.45E-01	2.45E-01	2.45E-01	2.45E-01	-1.52E-01	-1.52E-01	-1.52E-01
W5	-1.26E-01	-0.1154864	-2.18E-01	-2.18E-01	3.03E-01	3.03E-01	3.03E-01	3.03E-01	-6.84E-01	-6.84E-01	-6.84E-01
W6	-1.84E-01	-0.1695486	-6.99E-01	-6.99E-01	5.58E-01	5.58E-01	5.58E-01	5.58E-01	-1.52E-01	-1.52E-01	-1.52E-01
W7	-1.26E-01	-0.1154864	-2.18E-01	-2.18E-01	2.45E-01	2.45E-01	2.45E-01	2.45E-01	-1.52E-01	-1.52E-01	-1.52E-01
W9	-1.26E-01	-0.1154864	-1.28E-16	-1.28E-16	3.60E-16	3.60E-16	3.60E-16	3.60E-16	5.34E-17	5.34E-17	5.34E-17
W10	-1.12E-01	-0.0944937	6.21E-02	6.21E-02	-5.69E-02	-5.69E-02	-5.69E-02	-5.69E-02	1.22E-16	1.22E-16	1.22E-16
W13	-1.12E-01	-0.0944937	6.21E-02	6.21E-02	-5.69E-02	-5.69E-02	-5.69E-02	-5.69E-02	6.07E-17	6.07E-17	6.07E-17
W11		-0.0636739	2.17E-01	2.17E-01	-2.75E-01	-2.75E-01	-2.75E-01	-2.75E-01	9.93E-02	9.93E-02	9.93E-02
W12		-0.0636739	2.17E-01	2.17E-01	-2.75E-01	-2.75E-01	-2.75E-01	-2.75E-01	9.93E-02	9.93E-02	9.93E-02
W16			2.47E-01	2.47E-01	-3.26E-01	-3.26E-01	-3.26E-01	-3.26E-01	9.93E-02	9.93E-02	9.93E-02
W8			2.47E-01	2.47E-01	-3.09E-02	-3.09E-02	-3.09E-02	-3.09E-02	9.93E-02	9.93E-02	9.93E-02

Benchmark

Based off the most recent period

We see how close t_c matches t_f, and use that precision as a benchmark

Southern Women Dataset



Machine Learning Algorithms

Question: If we have the "community history" of a group can we correctly predict using some "affinity score" which communites individuals will end up in?

Machine Learning Algorithms

Question: If we have the "community history" of a group can we correctly predict using some "affinity score" which communites individuals will end up in?

Various Machine Learning Algorithms.

- Random Forrest
- XGBoost

Machine Learning Algorithms

Question: If we have the "community history" of a group can we correctly predict using some "affinity score" which communites individuals will end up in?

Various Machine Learning Algorithms.

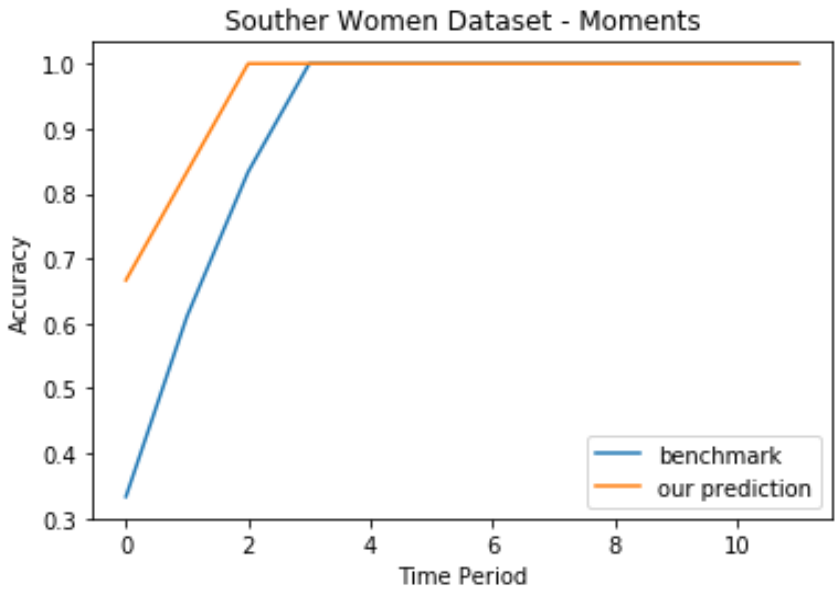
- Random Forrest
- XGBoost

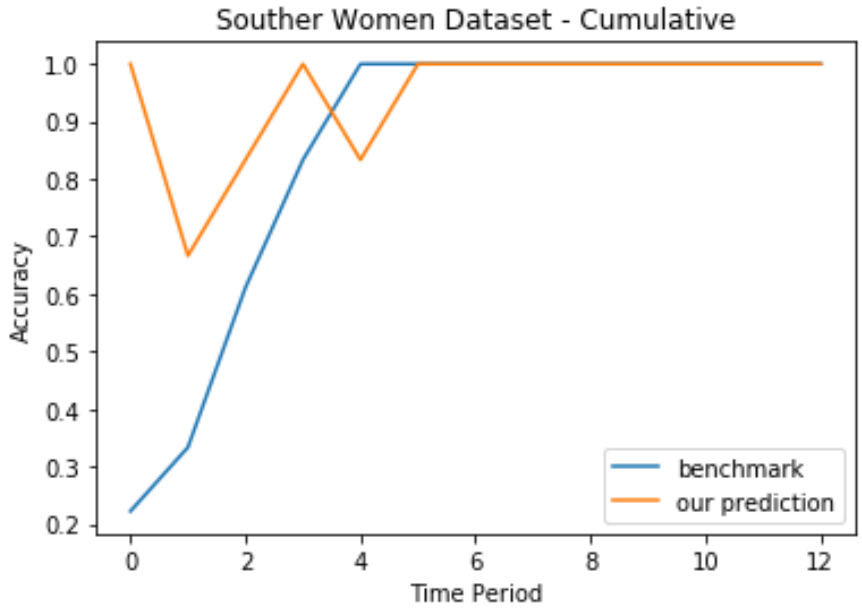
Existing Question: Is there a way to match the algorithm to the kind of network we observe?

Experimental Results

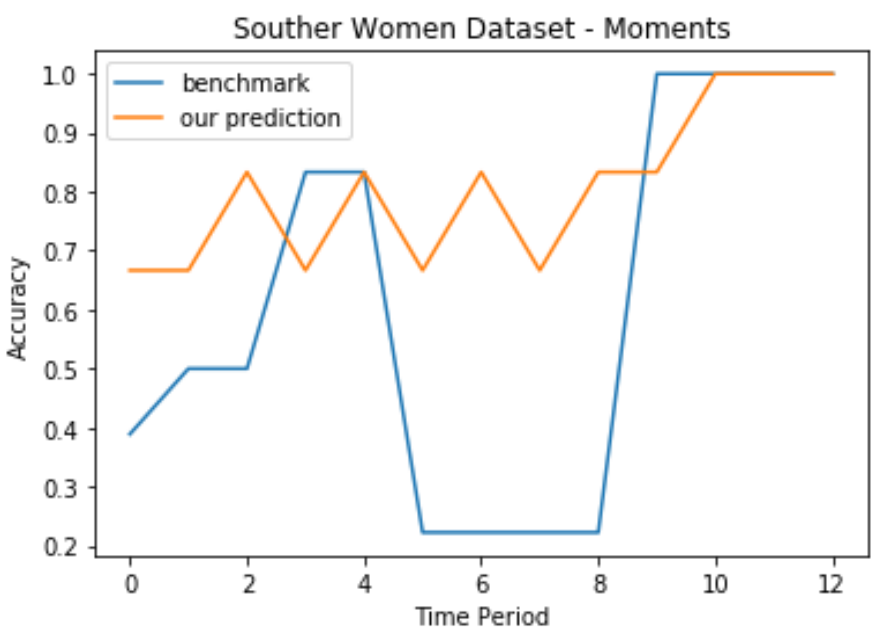
Results vary depending on community detection algorithm

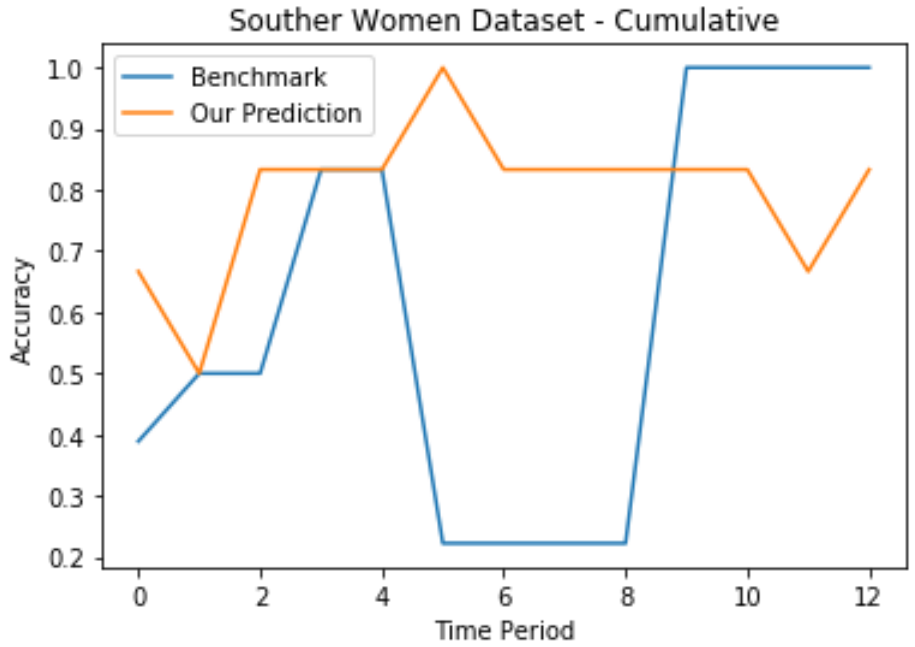
Experimental Results – Kernighan-Lin





Experimental Results – Spectral Partitioning





Other Networks

Advantages of using Southern Women Dataset

- Small => manageable
- Groups are intuitive

Disadvantages:

- Too small
- May not generalize

Other Networks

Other networks

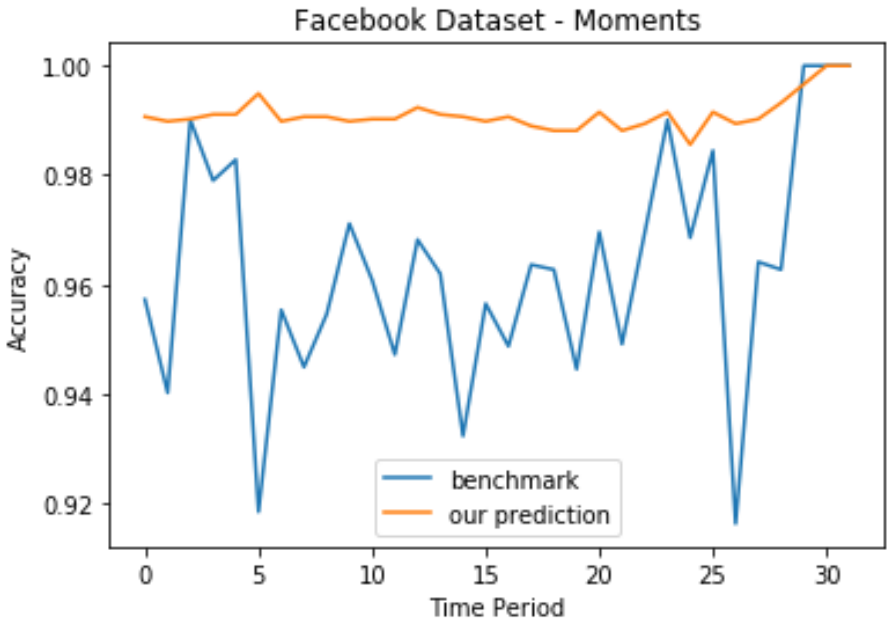
Facebook Dataset – Konect Database (2006-2009)

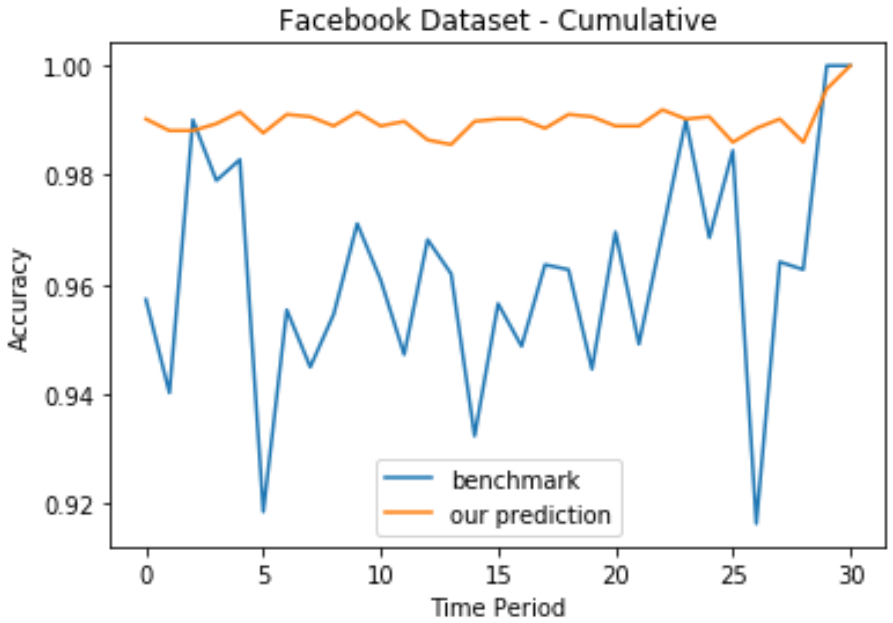
Subset of the Facebook social network

MIT Reality Mining Project – Konect Database

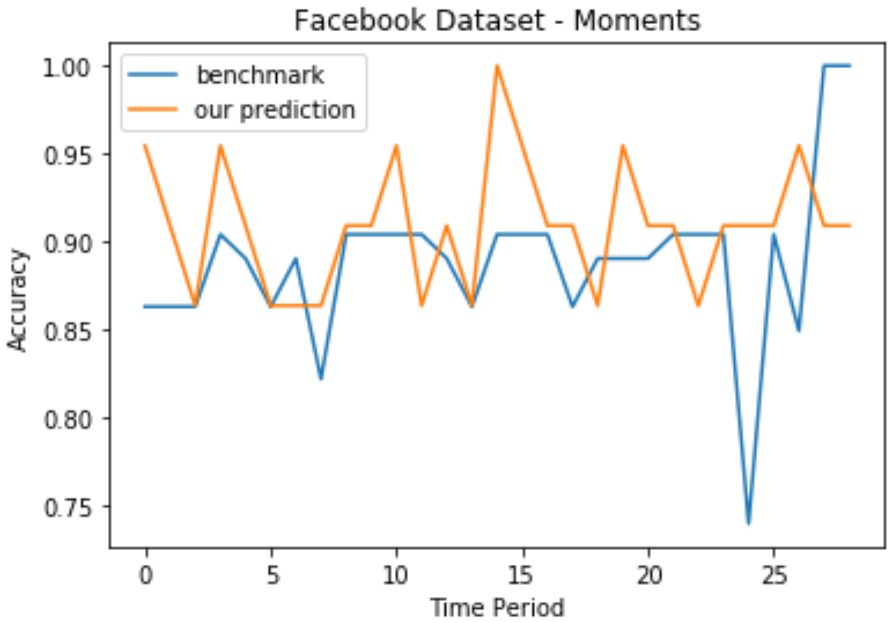
Contact between 100 individuals over time

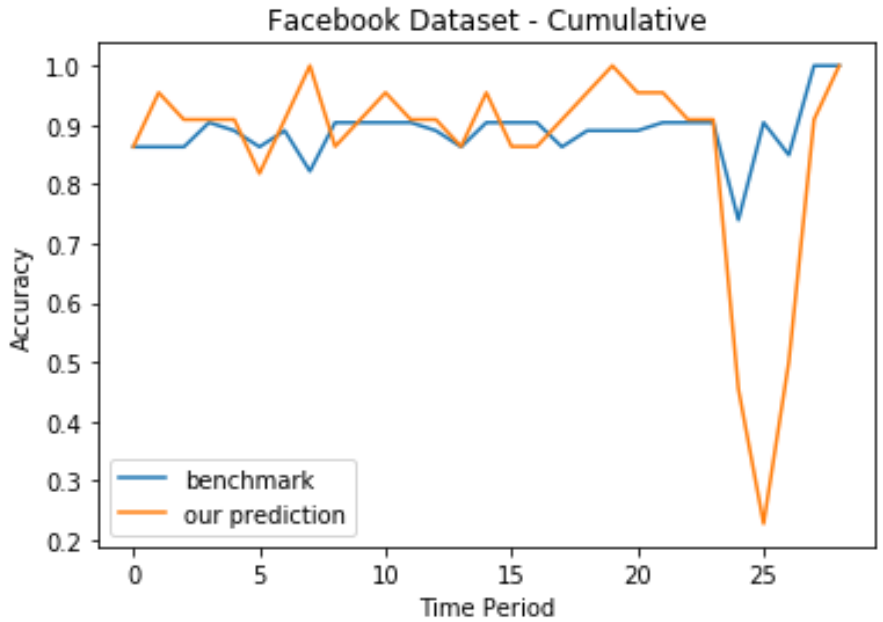
Experimental Results – Spectral Partitioning (2008-11)



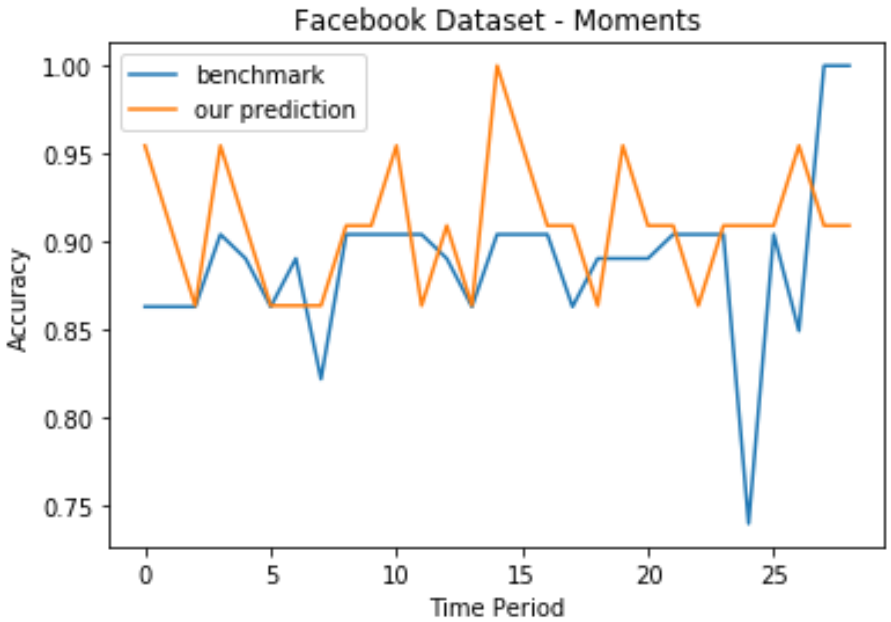


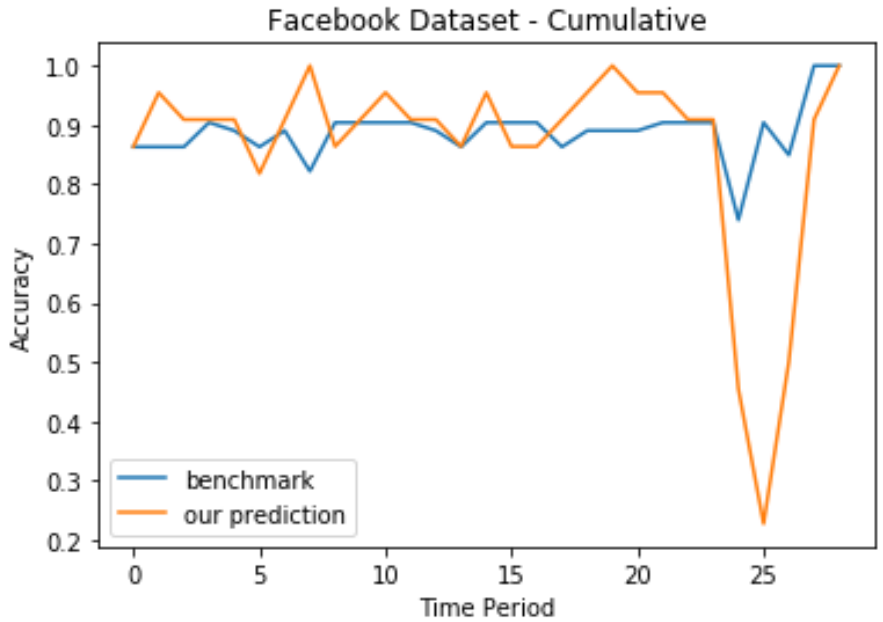
Experimental Results – (growth around a single node)





Experimental Results – (growth around a single node)

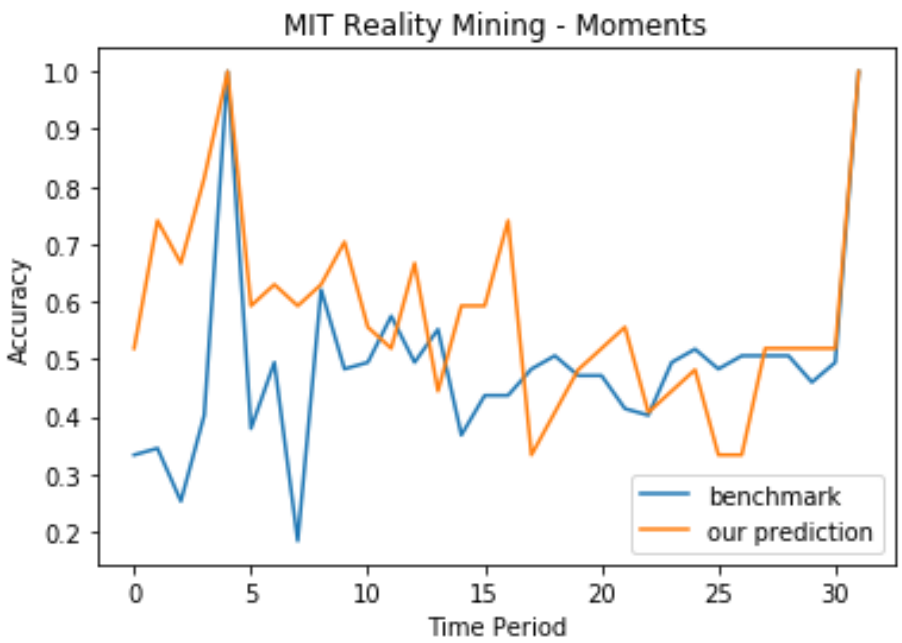


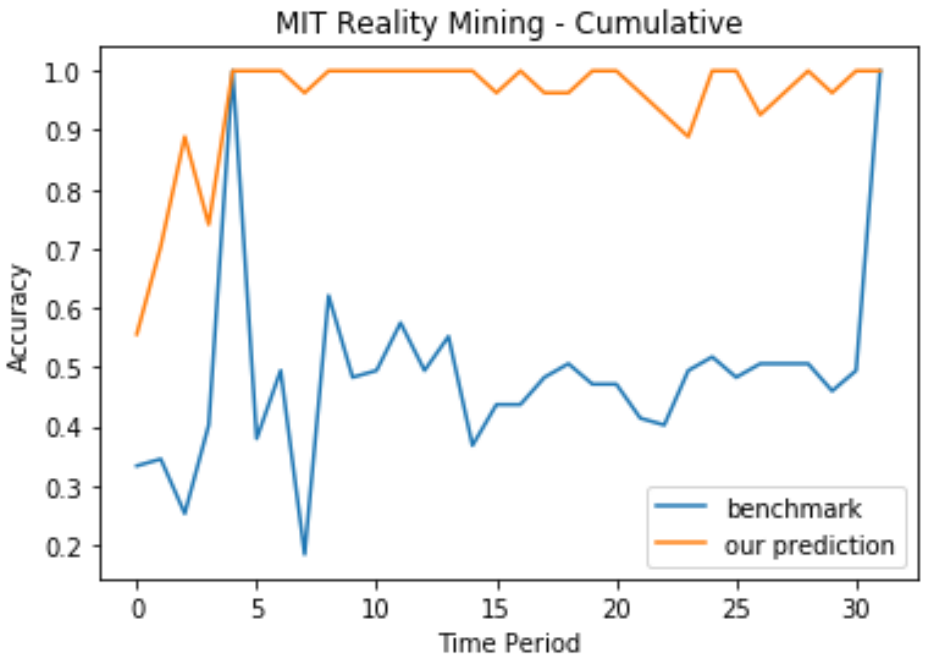


Result varies when observing the network around an individual or the entire network over a period of time

- Regardless, result is consistently better than the benchmark

Experimental Results – MIT





Summary

We can accurately predict group formation at a later point in time by observing the current structure of the communities based off of varying community detection algorithms

The End

Thank you!