Network Analysis:

The Hidden Structures behind the Webs We Weave

17-338 / 17-668

Graph Theoretic Signatures of Social Processes

Thursday, September 12, 2024

Patrick Park & Bogdan Vasilescu

Carnegie Mellon University
School of Computer Science

S3D
Software and Societal Systems Department

2-min Quiz, on Canvas

Quick Recap – Last Thursday's Lecture

Structural Balance: triads of friends and enemies

But, most real world social networks are not perfectly balanced

Many different triadic relationships exist

Triadic closure – two nodes that are connected to the same set of other nodes have a higher probability of forming an edge

Q: Why do social networks exhibit triadic closure?

Local clustering coefficient (probability that two neighbors of a node are connected) measures the extent of triadic closure in a network

Today

Continue to explore how social context relates to graph structure

Three example signatures:

- Graph-level: spanning tree
- Dyad-level: joint-bridging (or “network dispersion”)
- Node-level: distribution of interactions (or the “social signature”)

Case Study: Graph-Level Signature

Graph-Level Signature of Romantic Relationships

Romantic and sexual networks directly influence the contagion dynamics of STD

Accurately describing the network structure helps us understand contagion dynamics

Network structure emerges from the aggregate of individual partner choices

Identifying the reasons for those individual choices is important for public health policy (e.g., incentives to suppress emergence of detrimental network structures in terms of contagion)

Graph-Level Signature of Romantic Relationships

Bearman, Moody, and Stovel 2004 “Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks” *American Journal of Sociology*

Motivation of the study

- Romantic and sexual networks directly influence the contagion dynamics of STD
- Accurate description of network structure helps us understand contagion dynamics
- Network structure emerges from the aggregate of individual partner choices
- Identifying the reasons for those individual choices is important for public health policy (e.g., incentives to suppress emergence of detrimental network structures in terms of contagion)

Graph-Level Signature of Romantic Relationships

Bearman, Moody, and Stovel 2004 “Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks” *American Journal of Sociology*

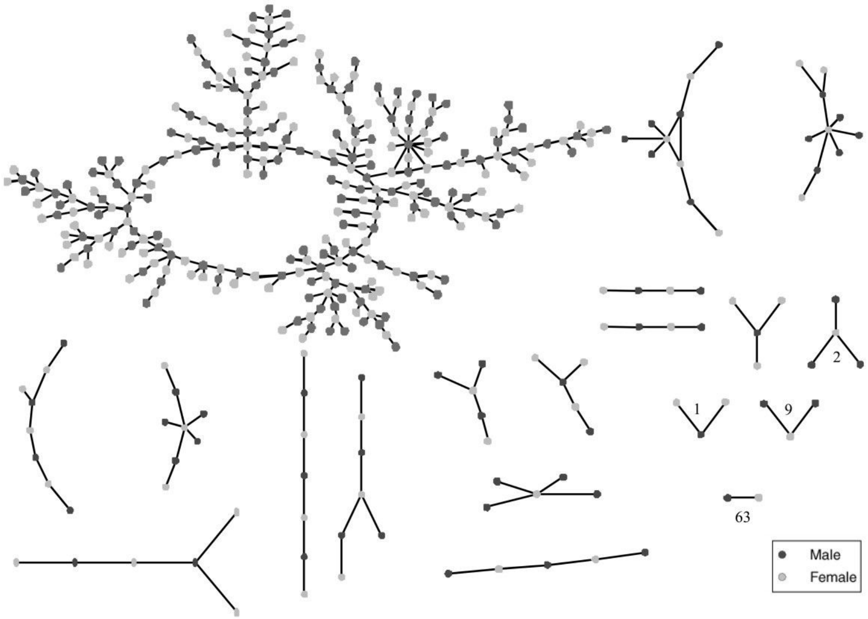
Analytic Strategy

- Describe the observed network features that affect contagion
    - against random network baselines
- Explore social factors of network structure
    - salient factors related to partner choice (homophily)
    - Incorporate social factors in constructing the random network baseline
- Explore salient graph features and deduce social factors
    - Theorize what norms / preferences generate those graph features
    - Incorporate those features into the random network baseline

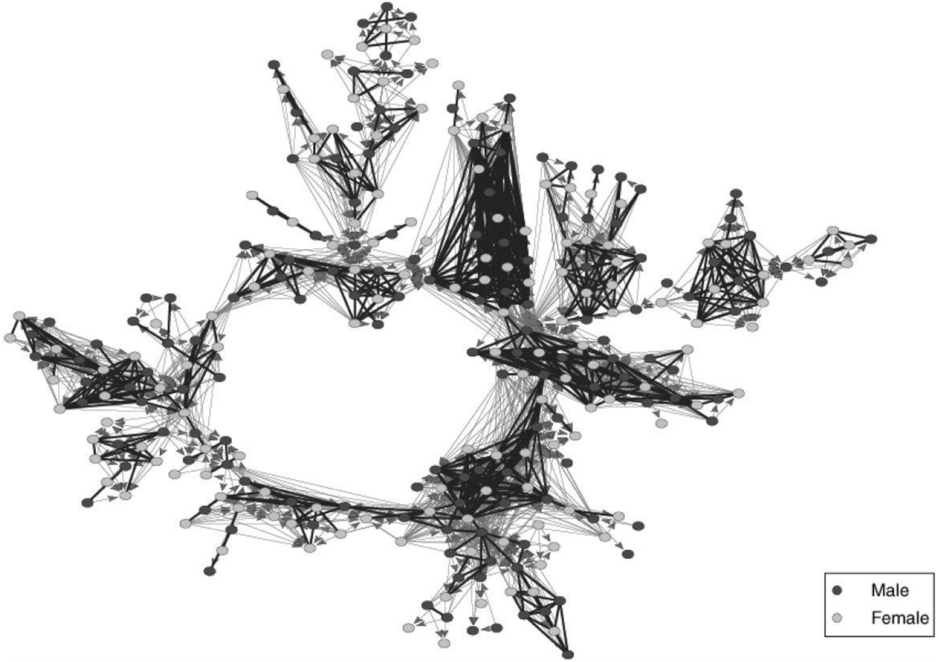
Description of Observed Network

Temporal ordering of dating ties make it possib
broad contagion of STD across the network
component

Spanning tree structure at Jefferson High

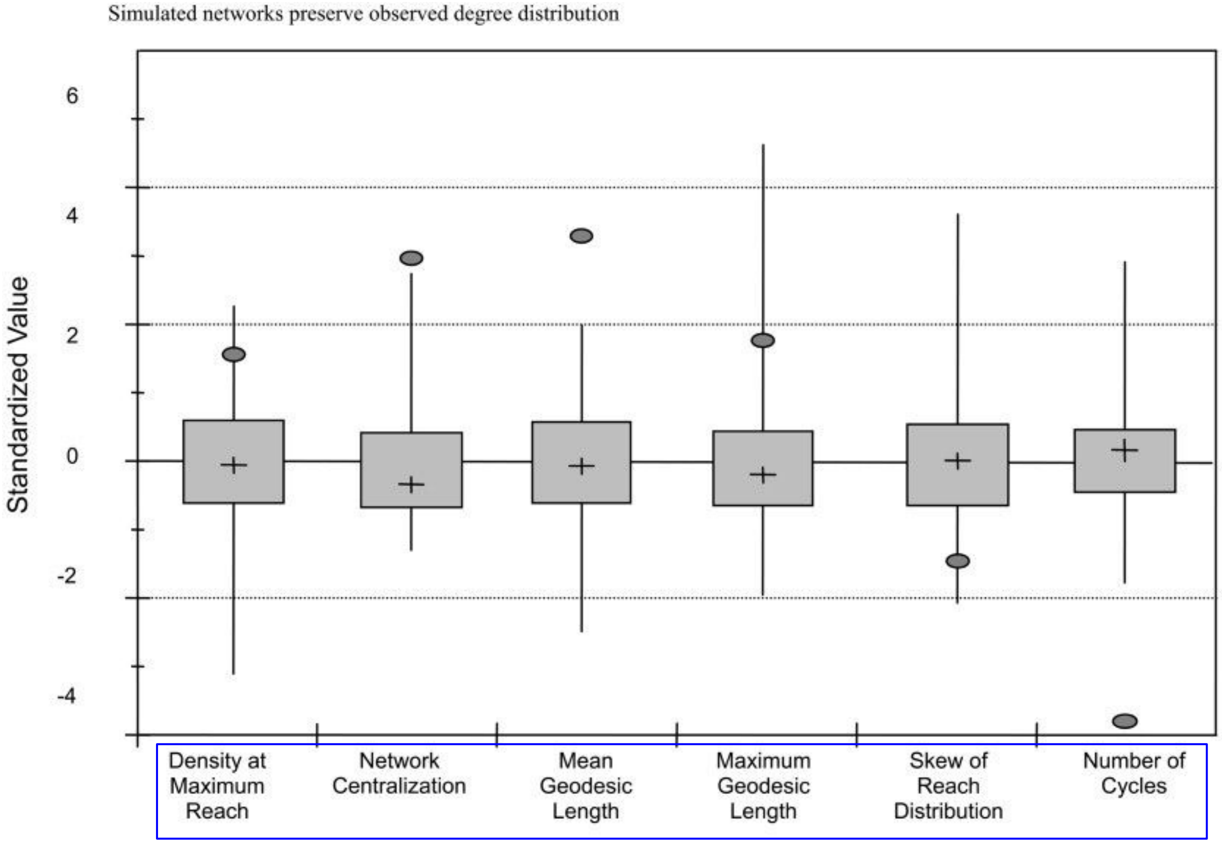


Dating Ties ignoring temporality



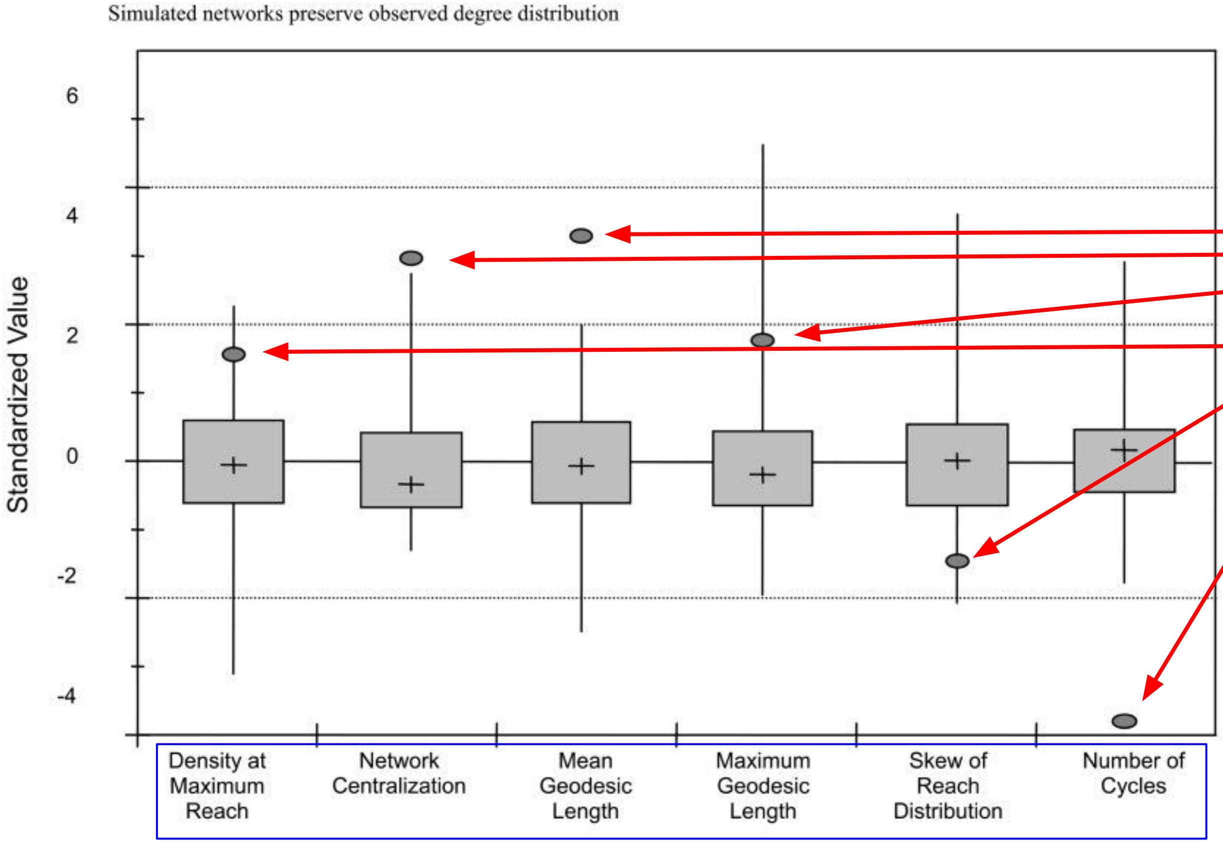
Temporally ordered Ties

Description of Observed Network against Random Graphs



Network features important for contagion

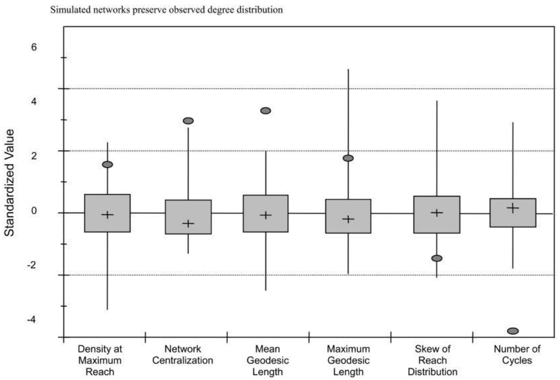
Description of Observed Network against Random Graphs



Observed network features deviate from random distribution

Network features important for contagion

Important: How to Study Social Mechanisms of Networks



Study the salient features, preferences, and norms in partner choice

Then, **translate** these social features into graph characteristics

Incorporate those graph characteristics as **constraints** that the random graph generator should respect

If the resulting constrained random graph becomes similar to the observed graph, you **conjecture** that those social features generated the observed graph structure

Important: How to Study Social Mechanisms of Networks

Translation is hard

Requires creativity

Study the salient features, preferences, and norms in partner choice

Then, **translate** these social features into graph characteristics

Incorporate those graph characteristics as **constraints** that the random graph generator should respect

If the resulting constrained random graph becomes similar to the observed graph, you **conjecture** that those social features generated the observed graph structure

Factors Related to Partner Choice

Partners shared these features (positive coefficients)

- SES
- Grade
- GPA
- Gets drunk
- Vocabulary

Table 2 Homophily in Student Pairs

Variable	QAP Mean Difference[a]	
	Full Network	Cross-Sex Only
Family SES...	.299***	.295***
Grade...	.331***	.367***
GPA...	.096**	.102***
Expect to graduate college...	.202***	.222***
School attachment...	.118***	.132***
Trouble in school...	.029	.019
Gets drunk...	.180***	.195***
Delinquency[b]...	−.058	−.070
Hours watching TV...	−.149	−.027
Religiosity (praying)...	−.006	−.012
Popularity (in-degree)...	−.377*	−.211
Self-esteem...	.004	.008
Autonomy...	.008	.002
Expect to get HIV...	.003	−.007
Expect to marry by 25...	.025	.020
Attractiveness...	.013	.047
Vocabulary (AH_PVT)...	1.508***	1.671***
Religion...	−.034*	−.043*
Sexually active...	−.100***	−.124***
Smoking...	−.087***	−.110***
School suspension...	−.028	−.066**
Tattoo...	−.003	−.016

Translating Social Preference to Graph Featur

The social preference:

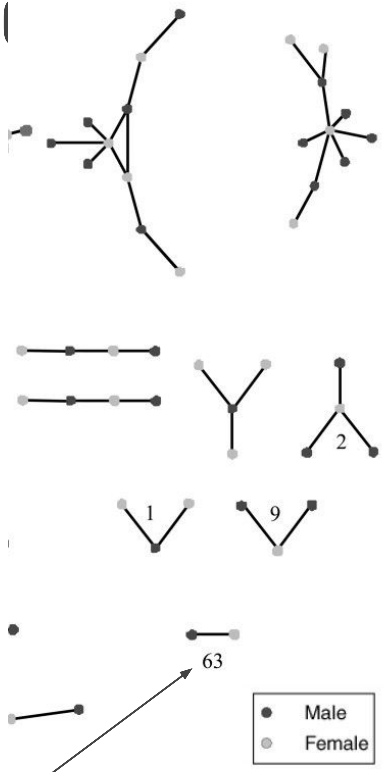
- People prefer partners with similar levels of dating experience

Corresponding graph feature:

- Isolated dyad: partners i and j did not have past partners

Incorporate graph feature into random graph:

- Force the random graph generation algorithm to create the same number of isolated edges

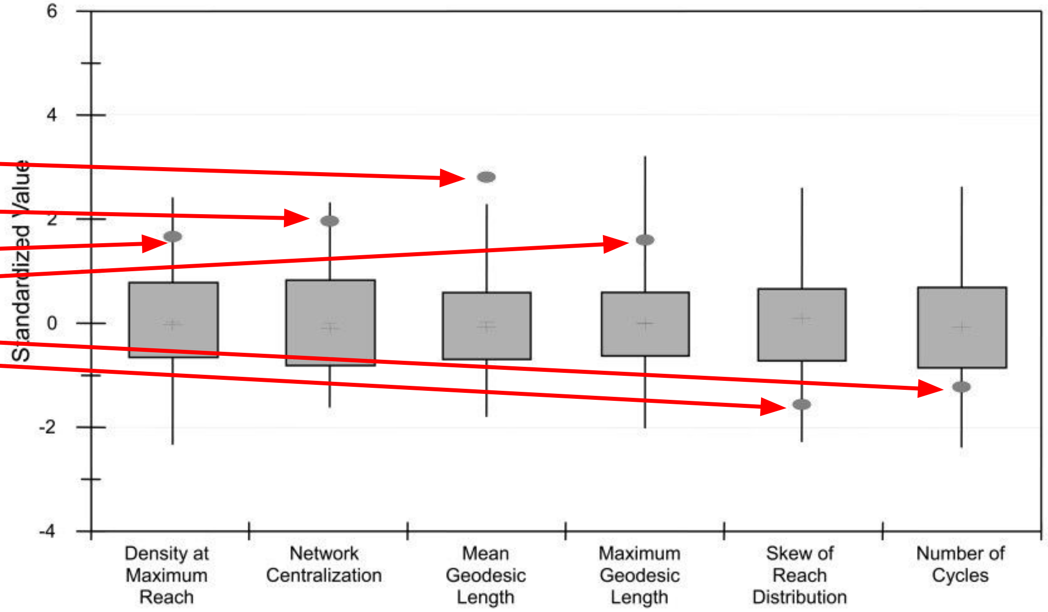


Translating Social Preference to Graph Features

Incorporate graph feature into random graph:

- Force the random graph generation algorithm to create the **same number of isolated edges**

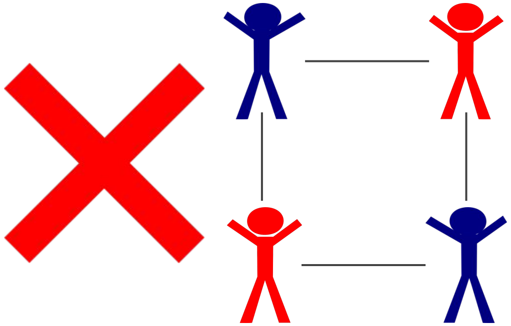
Now, observed network deviates less from these constrained random graphs



Reverse-Translating Graph Features to Social Preferences

Observed graph feature:

- The absence of four-cycles



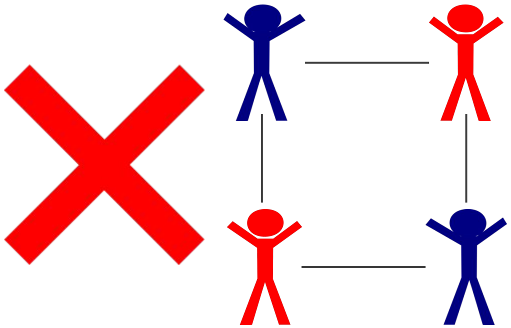
Corresponding social preferences / norms:

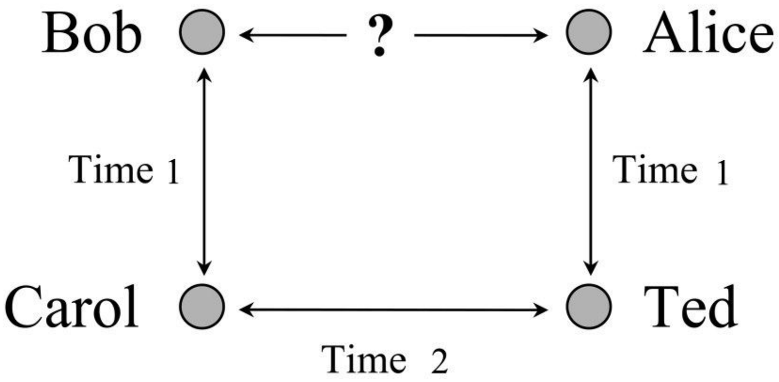
- Can you guess?

Reverse-Translating Graph Features to Social Preferences

Observed graph feature:

- The absence of four-cycles





Corresponding social preferences / norms:

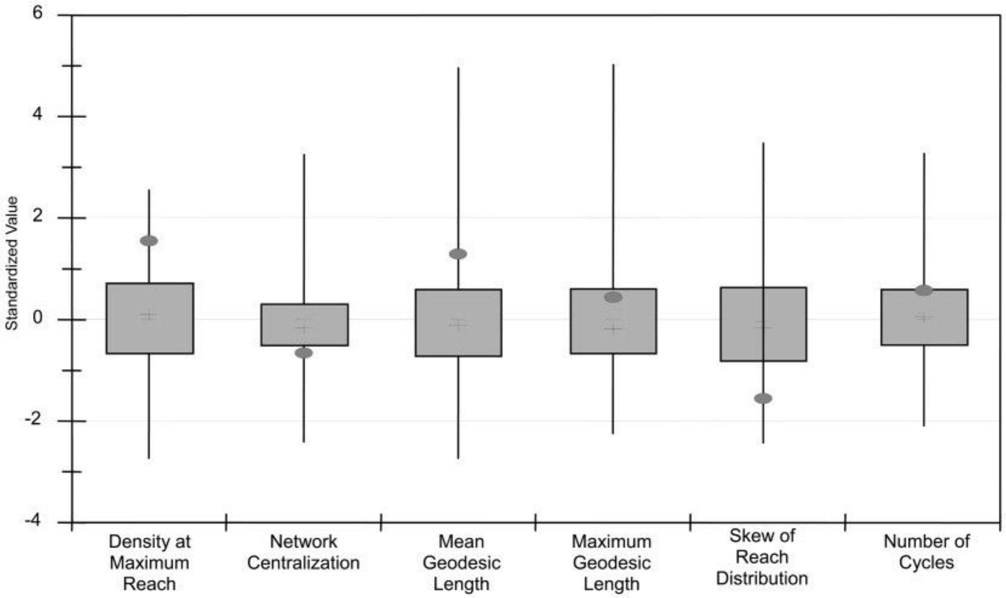
- **Avoidance of losing status**
- Hidden norm: Don’t Date your ex’s current partner’s ex

Reverse-Translating Graph Features to Social Preferences

Incorporate graph feature into random graph:

- Force the random graph generation algorithm to **suppress four-cycles**

Now, observed network does not deviate much from the random graphs that constrain four-cycles



Graph-Level Signature of Romantic Relationships

Q: Alternative explanations for absence of four-cycles?

Graph-Level Signature of Romantic Relationships

Q: Alternative explanations for absence of four-cycles?

Q: Is the lack of four-cycles a general signature in romantic networks beyond the high school context?

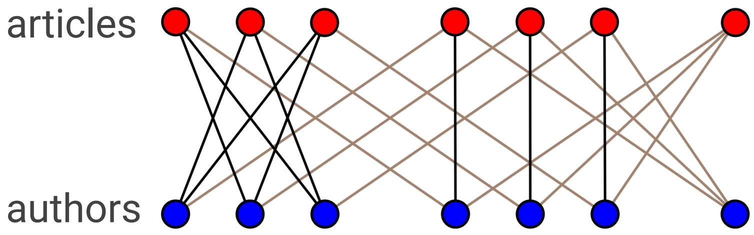
Graph-Level Signature of Romantic Relationships

Q: Alternative explanations for absence of four-cycles?

Q: Is the lack of four-cycles a general signature in romantic networks beyond the high school context?

The Jefferson High dating network was largely heterosexual: bipartite graph

Q: Do you think the bipartite graph of authors and articles lack four-cycles? Why?

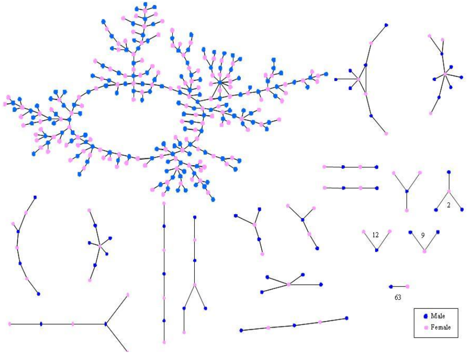


Case Study: Edge-Level Signature

Edge-Level Signature of Romantic Ties

As the high school romantic relationship network example demonstrates, sometimes certain relationship types in specific social contexts (e.g., school) leave a visible structural marker

in high school context

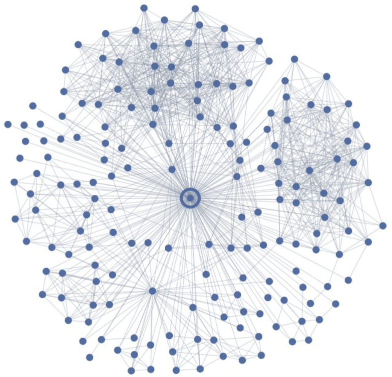


Graph Signature of Social Ties

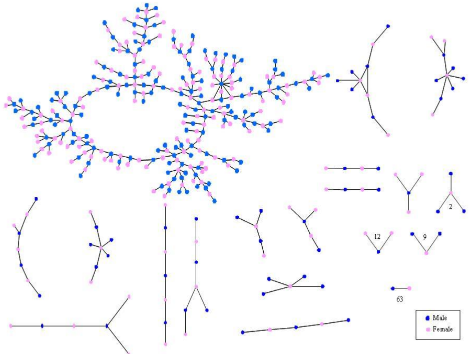
As the high school romantic relationship network example demonstrates, sometimes certain relationship types in specific social contexts (e.g., school) leave a visible structural marker

The same type of relationship can leave different structural markers in different social contexts

in Facebook



in high school context

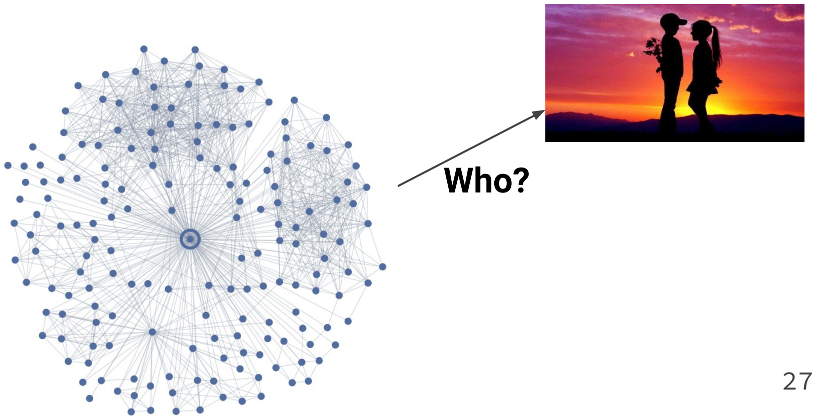


Graph Signature of Social Ties

An Illustrative Problem:

Predict the significant other (romantic partner / spouse) of a Facebook user solely from the user’s friendship graph

Q: Can you think of a graph characteristic that can hint at romantic partners or spouses?



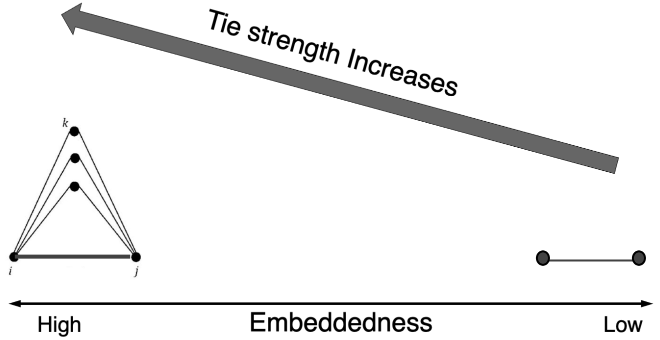
Graph Signature of Social Ties

An Illustrative Problem:

Predict the significant other (romantic partner / spouse) of a Facebook user solely from the user's friendship graph

A network analyst who learned about strong ties and triadic closure may reason:

- A social tie that is highly embedded tends to be strong



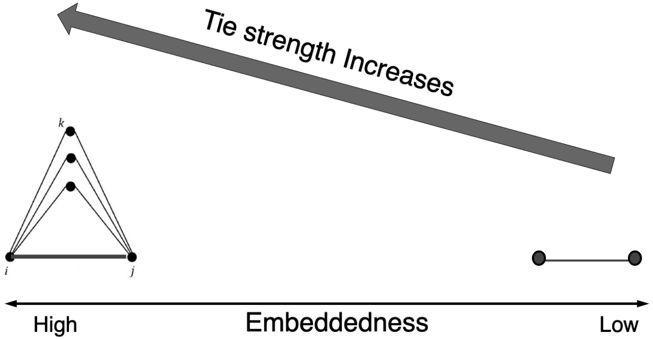
Graph Signature of Social Ties

The Problem:

Predict the significant other (romantic partner / spouse) of a Facebook user solely from the user's friendship graph

A network analyst who learned about strong ties and triadic closure may reason:

- A social tie that is highly embedded tends to be strong
- A partner is one of the strongest ties with many friends in common



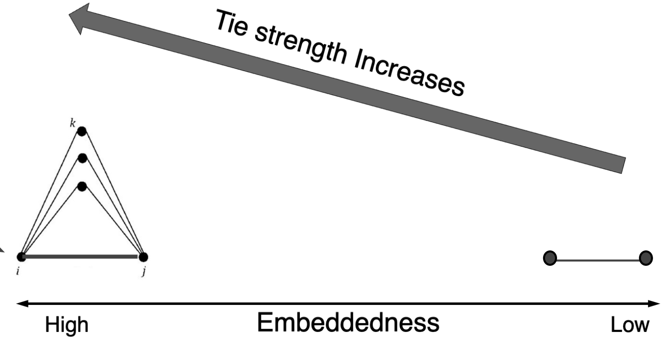
Graph Signature of Social Ties

The Problem:

Predict the significant other (romantic partner / spouse) of a Facebook user solely from the user's friendship graph

A network analyst who learned about strong ties and triadic closure may reason:

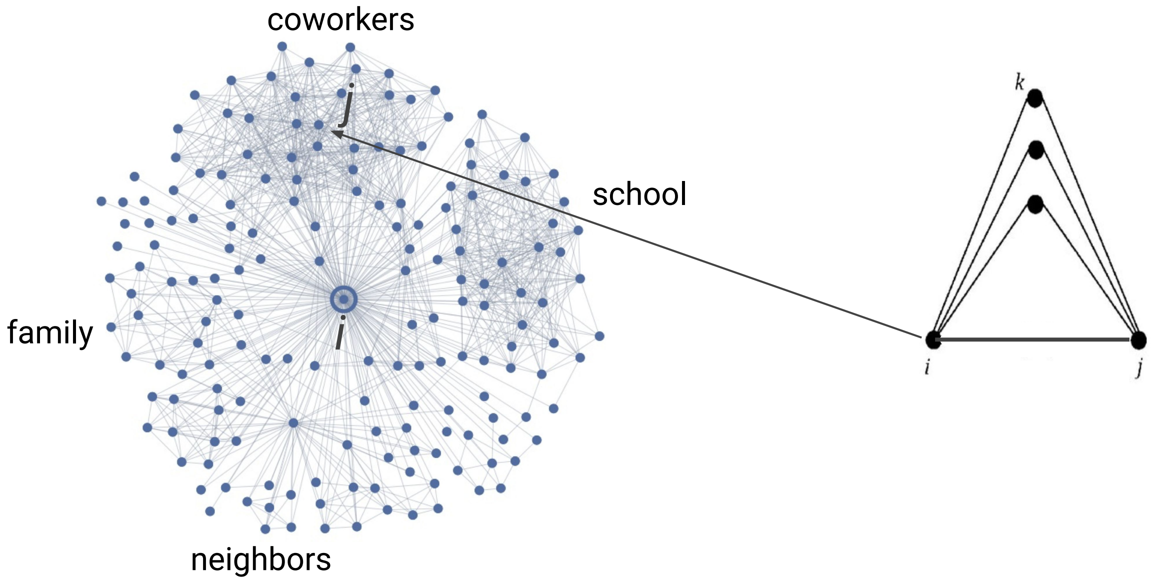
- A social tie that is highly embedded tends to be strong
- A partner is one of the strongest ties with many friends in common
- Therefore, the node with **highest embeddedness** is likely to be the partner



Graph Signature of a Significant Other

In practice, the friend with highest embeddedness is someone who is highly connected in the largest cluster

- Example: coworker, college friend, often not the significant other



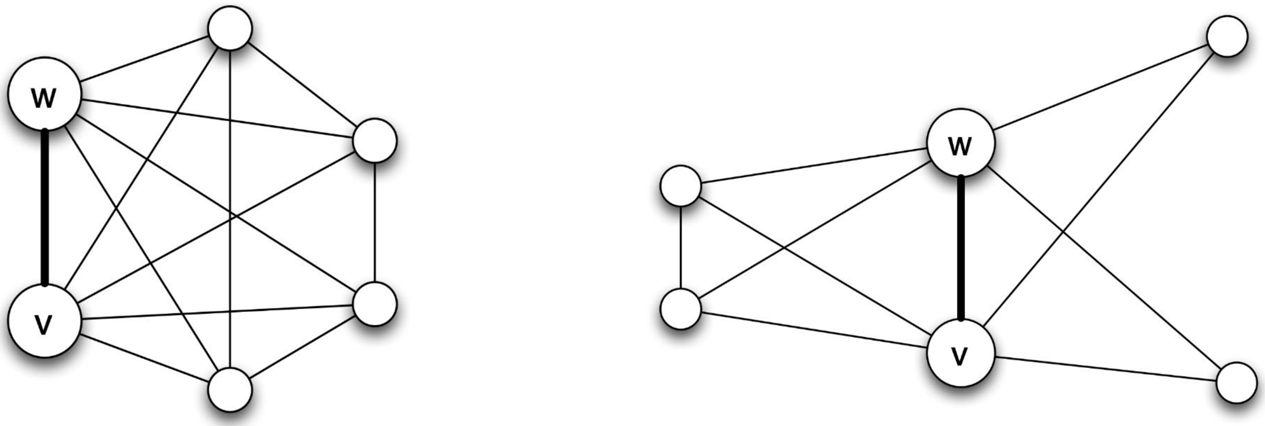
Graph Signature of a Significant Other

Backstrom and Kleinberg draw insight from the psychology literature on the characteristics of intimate ties

- a sense of intimacy, voluntary investment in the companionship
- an interest in **being together** as much as possible through interactions **in multiple social contexts** over a long period
- a sense of **mutuality** and support for partner’s needs

They focus on the fact that many couples are together in multiple social contexts

Graph Signature of a Significant Other

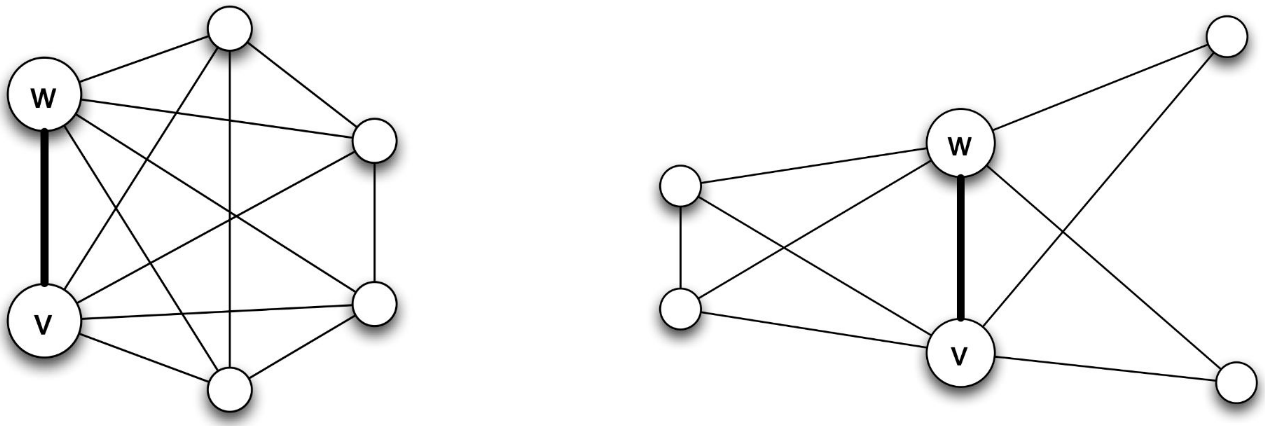


Instead of just counting mutual friends, look at their structure.

- How well connected are the common endpoints of edge e?
- If not well connected, suggests something about v-w relationship.
- v-w cannot be easily "explained" by any one social focus.

Source: [Jon Kleinberg's slide presentation]

Graph Signature of a Significant Other

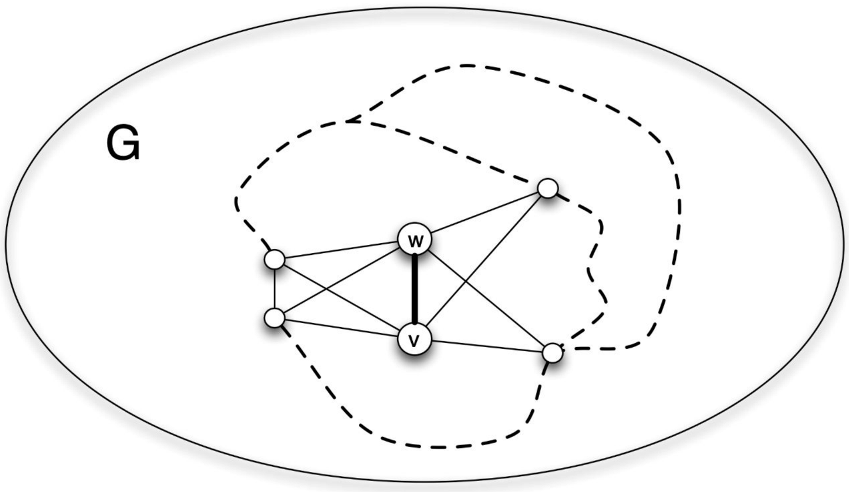


w-v tie on the left is highly embedded, but in a single social context

w-v tie on the right participates in three different social contexts

Together, they constitute a local bridge connecting these different contexts

Intuitively, the tie on the right is more likely to be partners



C_{vw} = common neighbors of v and w.

Sum of distances between pairs in C_{vw}, after deleting v and w:

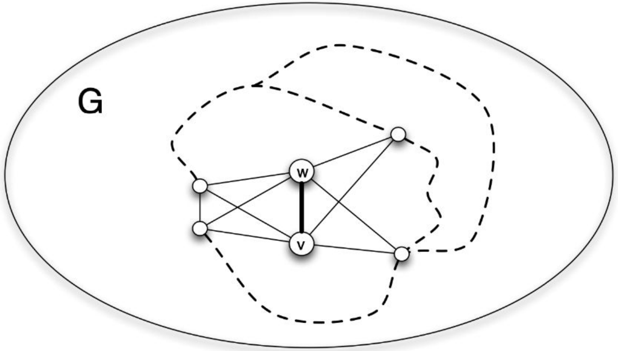
$$\sum_{s,t \in C_{vw}} d_{G\text{ - }\{v,w\}}(s, t).$$

The dispersion of edge (v, w) with respect to distance function d.

- Should use 0-1-valued metric; normalize by $|C_{vw}|$.

Source: Jon Kleinberg's slide presentation

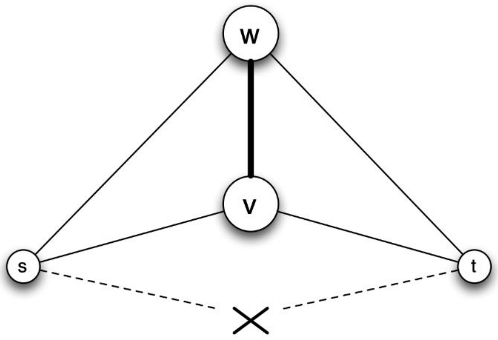
Can use many possible functions d.
$disp(v, w) = \sum_{s,t \in C_{vw}} d_{G - \{v,w\}}(s, t)$.



- $d(s,t) = \begin{cases} 0 \text{ if } (s,t) \text{ is an edge} \\ 1 \text{ otherwise} \end{cases}$
- $d(s,t) = \begin{cases} 0 \text{ if shortest } s\text{-}t \text{ path avoiding } v, w \text{ has} \leq k \text{ edges} \\ 1 \text{ otherwise} \end{cases}$

Can also normalize the dispersion: $\dfrac{disp(v,w)}{|C_{vw}|^{\alpha}}$.

- Analogue of clustering coefficient [Watts-Strogatz 98] is $k = 1$ and $\alpha = 2$.
- Searching over choices of k, α shows $k = 2$ and $\alpha = 1$ nearly optimal.



Source: [Jon Kleinberg's slide presentation]

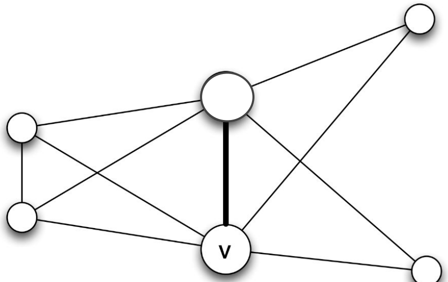
Evaluating the Methods

For evaluation, use 1.3 million Facebook users who:

- Declare a relationship partner in their profile (symmetric).
- Have between 50 and 2000 friends.
- Are at least 20 years old.

For each user v, rank all friends w by competing metrics:

- Embeddedness of v-w edge.
- Dispersion of v-w edge.
- Number of photos in which v and w are both tagged.
- Number of times v viewed w's profile in last 90 days.



For what fraction of all users v is the top-ranked w the relationship partner?

Source: Jon Kleinberg's slide presentation

A random guess for a user with 100 friends

= 1% accuracy

Highest dispersion

= 50.6% accuracy

type	embed	dispersion	photo	profile view
all	0.247	0.506	0.415	0.301
married	0.321	0.607	0.449	0.210
married (female)	0.296	0.551	0.391	0.202
married (male)	0.347	0.667	0.511	0.220
relationship	0.132	0.344	0.347	0.441
relationship (female)	0.139	0.316	0.290	0.467
relationship (male)	0.125	0.369	0.399	0.418

Notes:

- Embeddedness vs. dispersion
- Structural vs. activity-based
- Married vs. in a relationship
- Female vs. male
- Combining all via machine learning: 0.716 married, 0.682 relationship
- Approx 34-38% of dispersion's incorrect guesses are family members.

Source: [Jon Kleinberg's slide presentation]

Prediction performance much higher for married couples, compared to unmarried relationships

Why?

type	embed	dispersion	photo	profile view
all	0.247	0.506	0.415	0.301
married	0.321	0.607	0.449	0.210
married (female)	0.296	0.551	0.391	0.202
married (male)	0.347	0.667	0.511	0.220
relationship	0.132	0.344	0.347	0.441
relationship (female)	0.139	0.316	0.290	0.467
relationship (male)	0.125	0.369	0.399	0.418

Notes:

Embeddedness vs. dispersion

Structural vs. activity-based

Married vs. in a relationship

Female vs. male

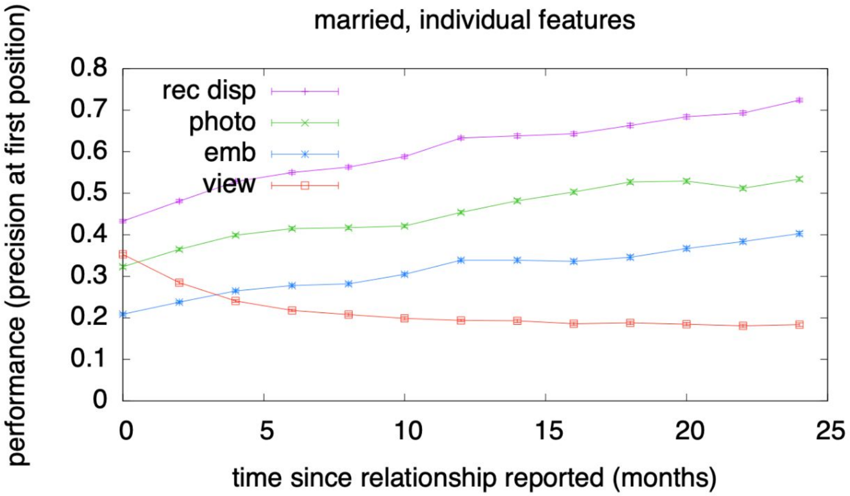
Combining all via machine learning: 0.716 married, 0.682 relationship

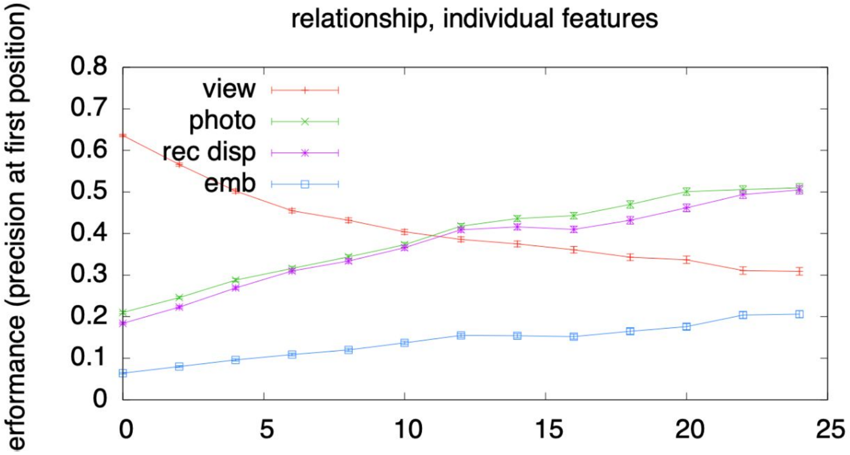
Approx 34-38% of dispersion's incorrect guesses are family members.

Source: Jon Kleinberg's slide presentation

Because it takes time for a couple to share multiple social contexts

Recall, intimate ties have an interest in being together as much as possible through interactions in multiple social contexts **over a long period**





Source: [Jon Kleinberg's slide presentation]

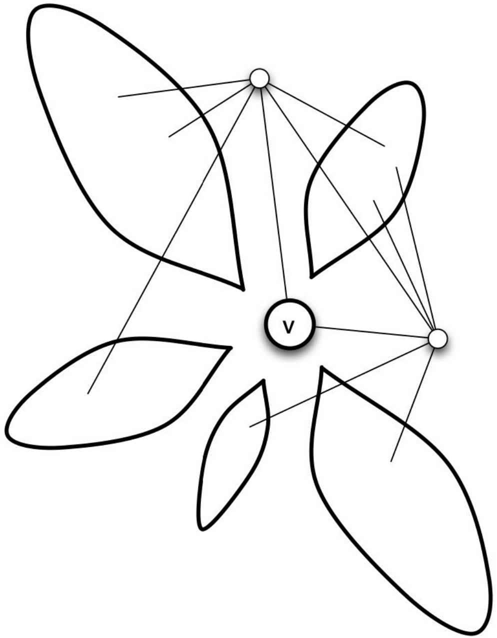
Graph Signature of a Significant Other

So, a significant other is a person who navigates the social world with you as a single unit, a companion

Lesson 1: Seek insights from the social and try to map them on to quantitative features in the graph

Example: Being together in multiple contexts→ network dispersion

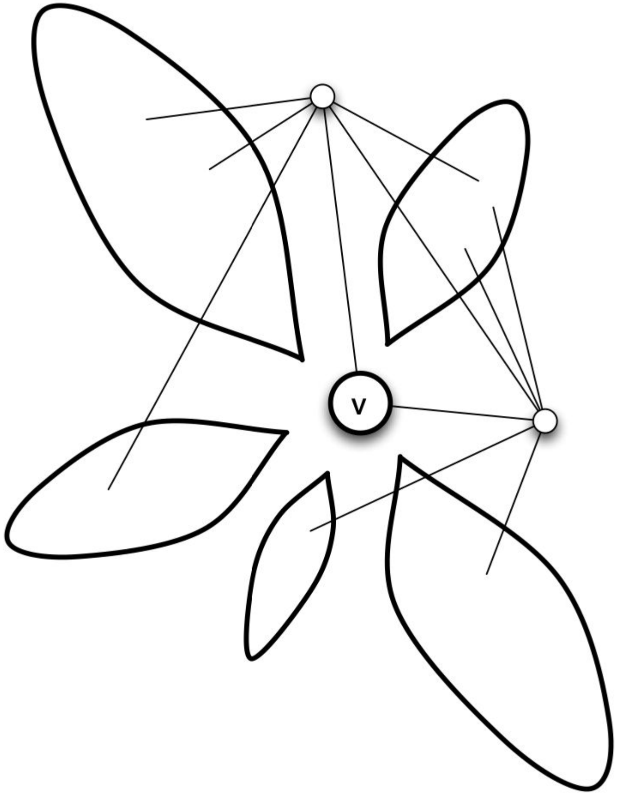
Lesson 2: Analyze those graph features and circle back to evaluate how well they capture the relationships within a social context



Graph Signature of a Significant Other

Q: Suppose *i* and *j* are partners in real life

If *j* gets the highest dispersion score from *i*'s network, but *i* does not get the highest dispersion score in *j*'s network, what do you think this mismatch suggests of their romantic relationship?



Case Study: Node-Level Signature in Communication

People Allocate Communication Volume Differently

Q: Do people maintain the same distribution of interaction volume across friends?

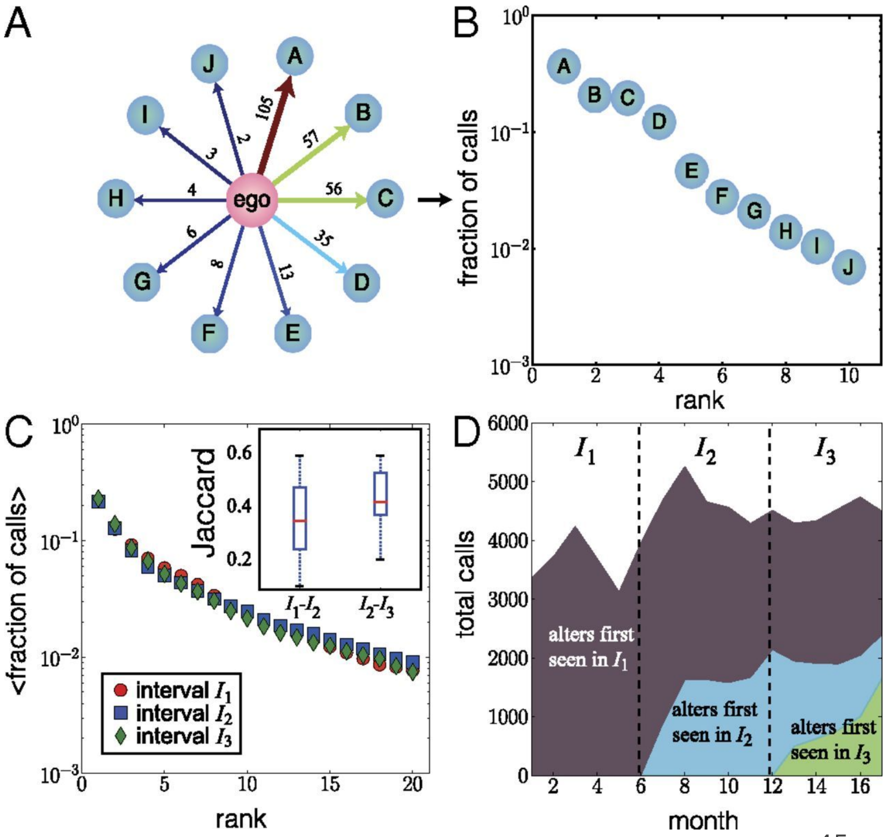
People Allocate Communication Volume Differently

Do people maintain the same distribution of interaction volume across friends?

- Apparently, they do

Each individual has a unique distribution of communication across network neighbors

- This distribution is temporally stable
- Despite network churn
- The distribution is a “social signature”



Saramaki et al. 2014

Summary

Network signatures

- Graph level: high school romantic network
- Edge level: network dispersion
- Node level: Communication distribution
- Translating the social features to graph characteristics (and vice versa)

Where We Are in the Course

Basic building blocks of networks: nodes, links, dyads, triads

Basic tools for analyzing networks: graph theory, BFS, random graph model

Universal properties (natural sciences) vs context and nuance (social sciences)

Fundamental properties of social networks (e.g., small worlds, reciprocity, triadic closure)

- Short paths connecting nodes
    - Random Wikipedia articles https://www.thewikigame.com
    - Co-authorship distance https://www.csauthors.net/distance
- Triangles formed by common neighbors
- Similarity between neighbors ← more next time

Graph signatures