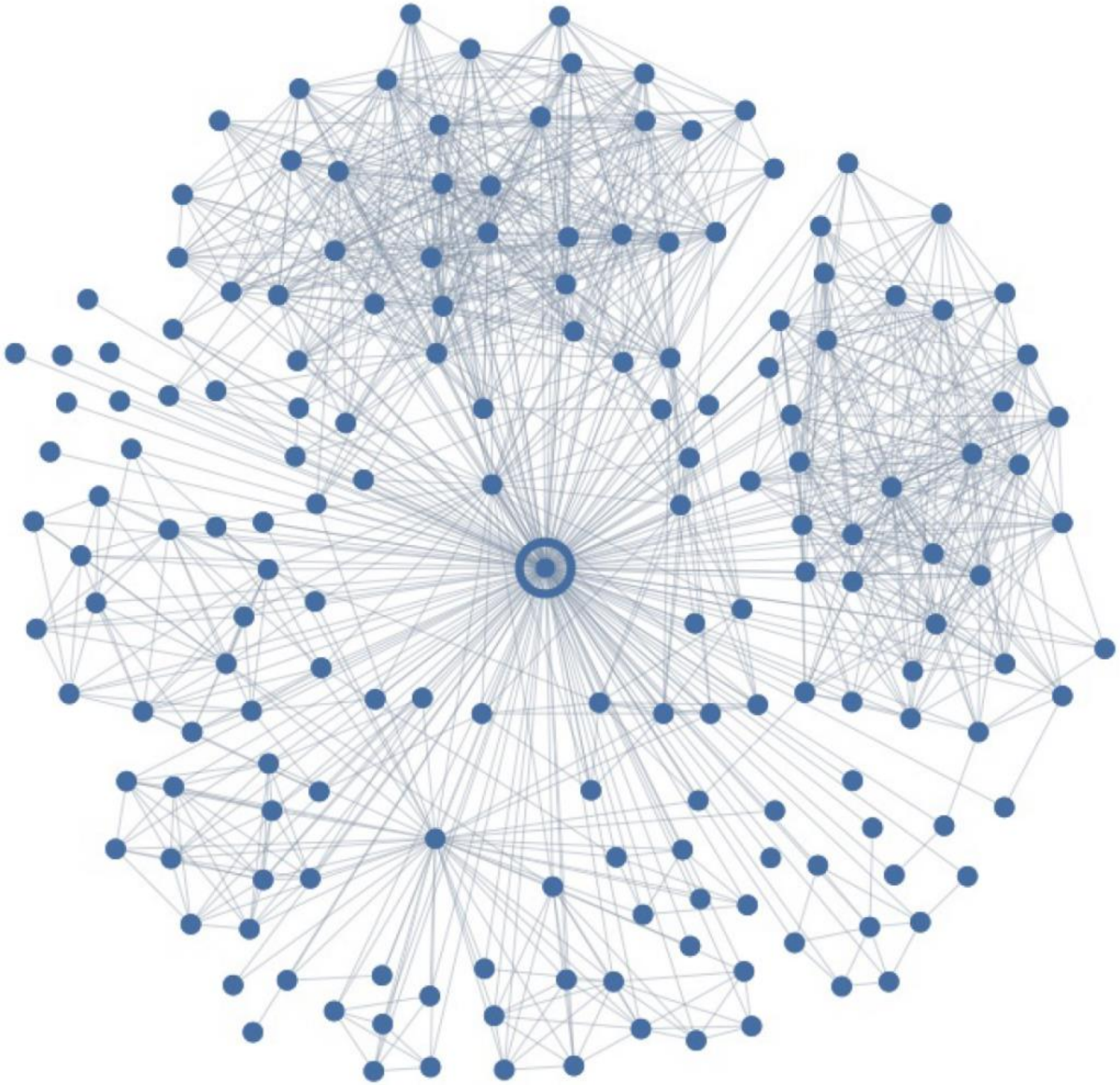
ETH zürich

Romantic Bundle

Roni Häcki
haeckir@student.ethz.ch

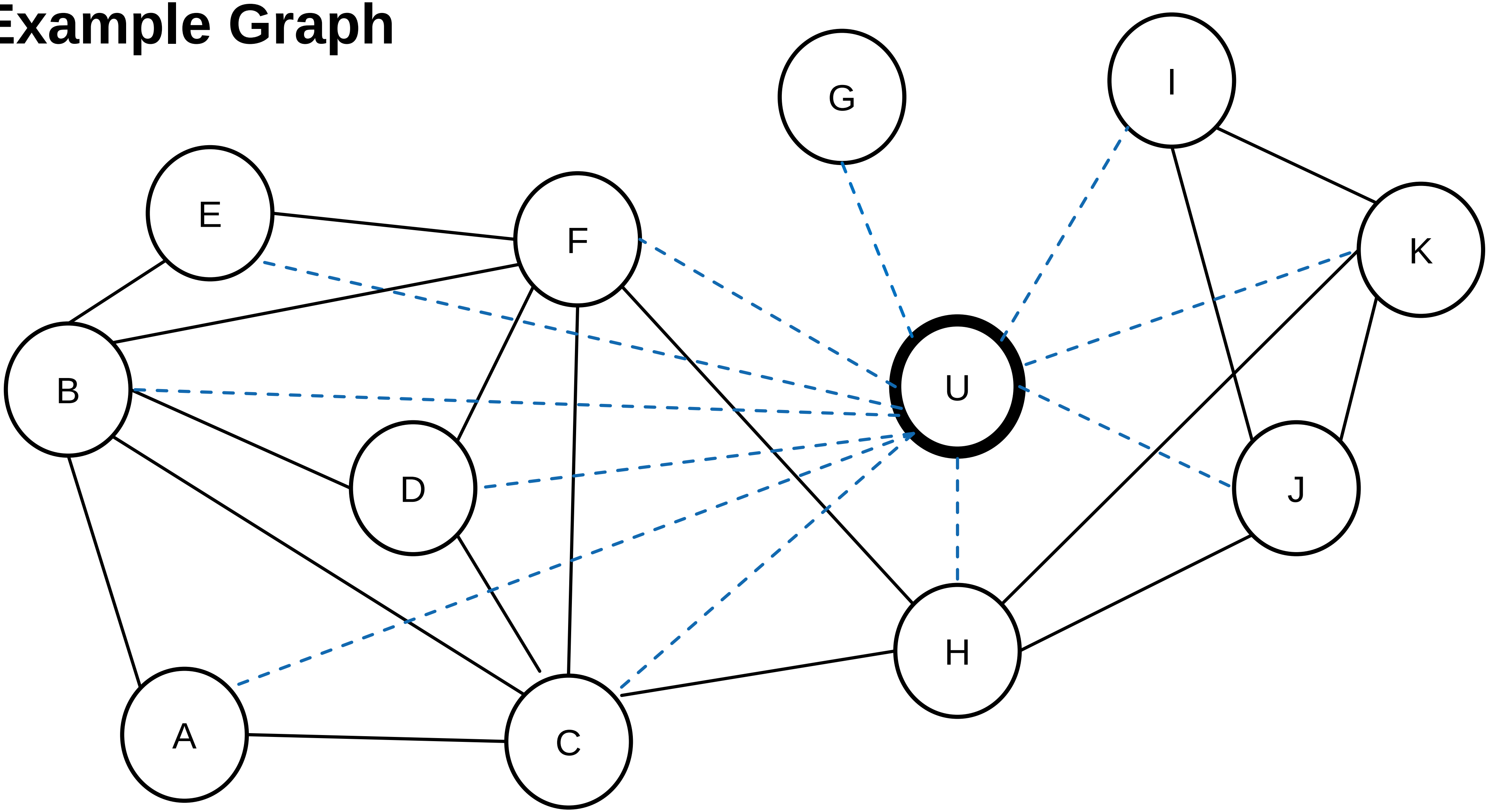
Motivation



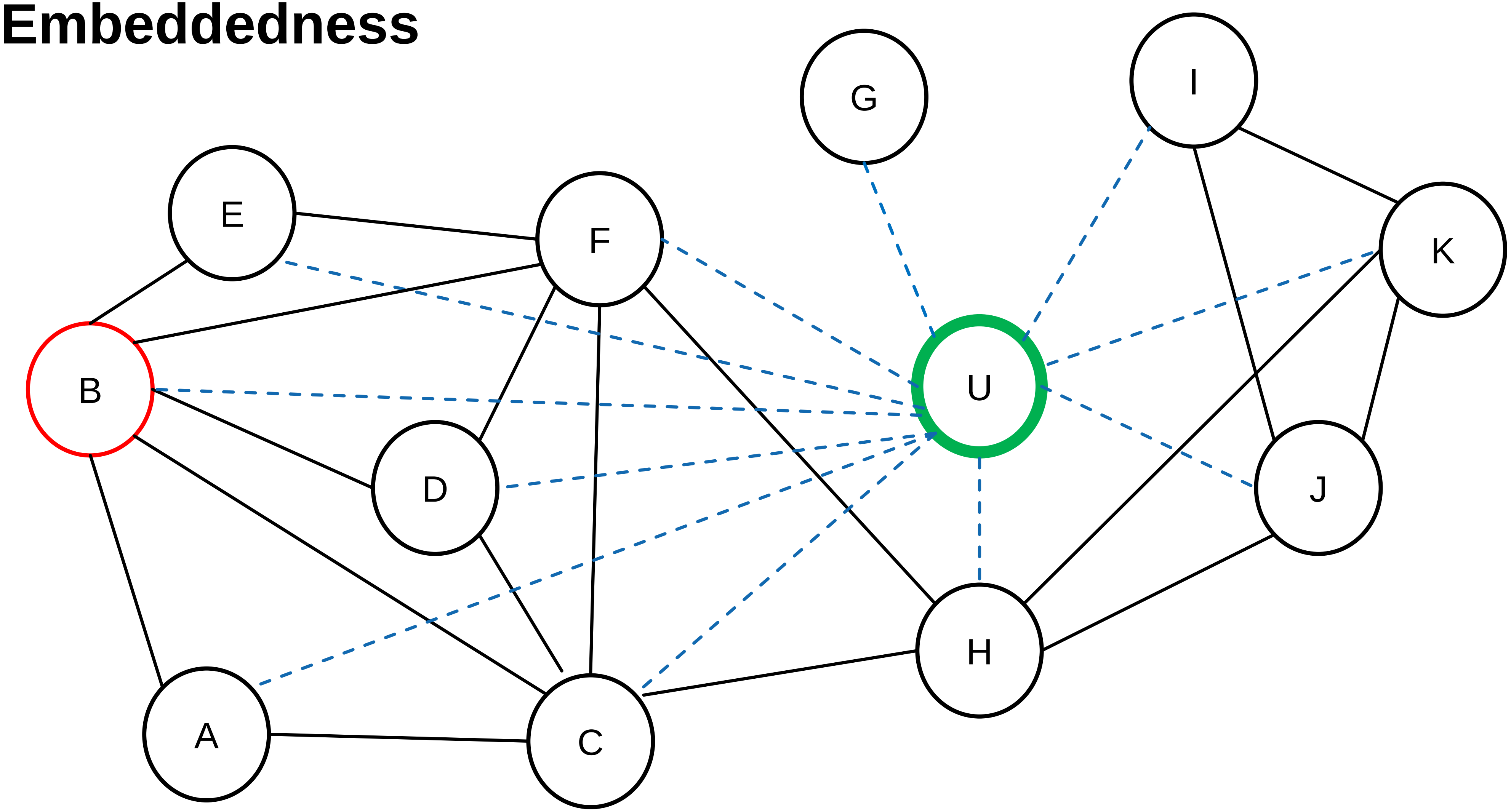
Embeddedness

- Embeddedness
  - The number of mutual friends two people share
  - Quantity that increases with tie strength

Example Graph

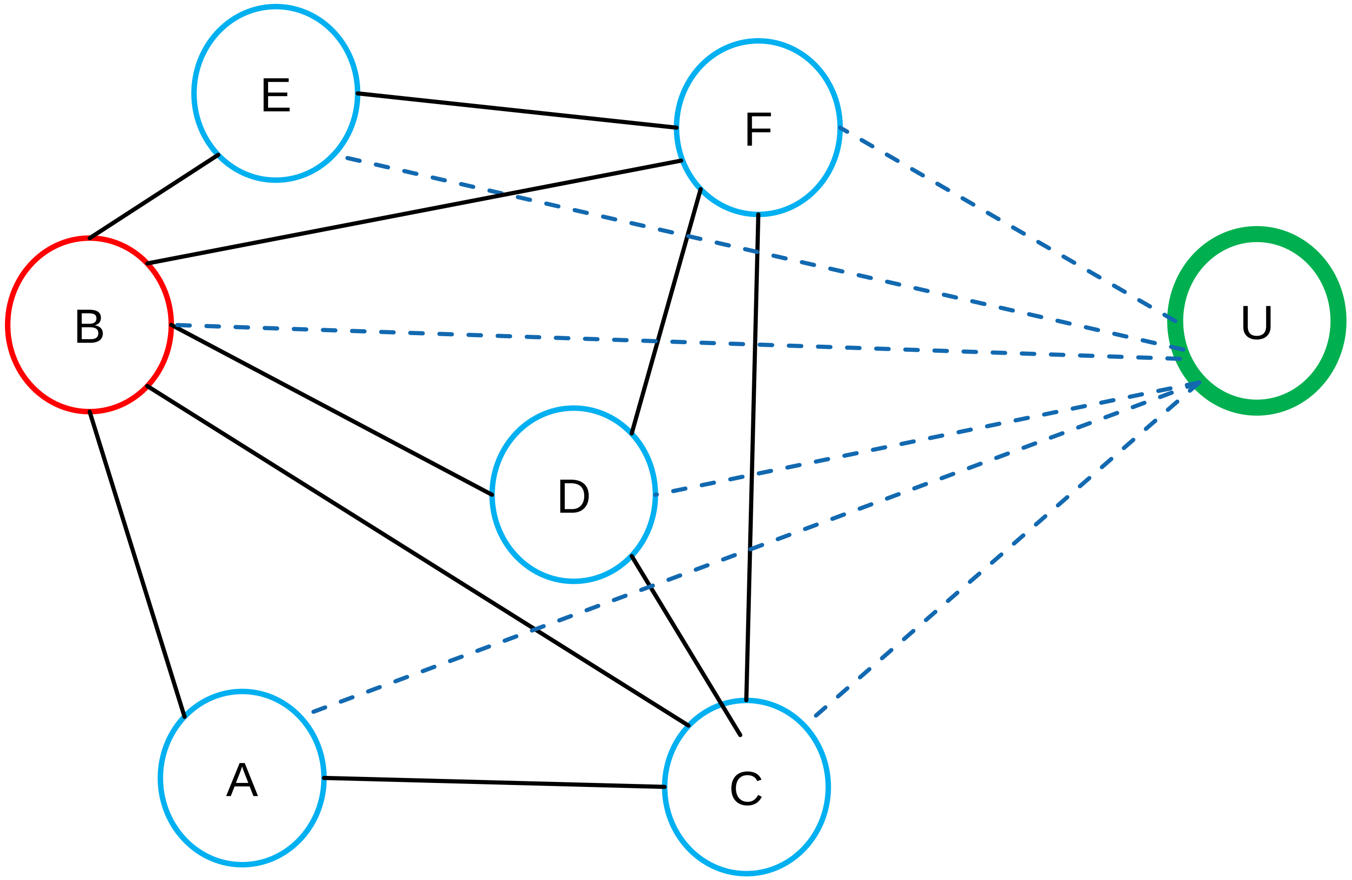


Embeddedness



ETH zürich

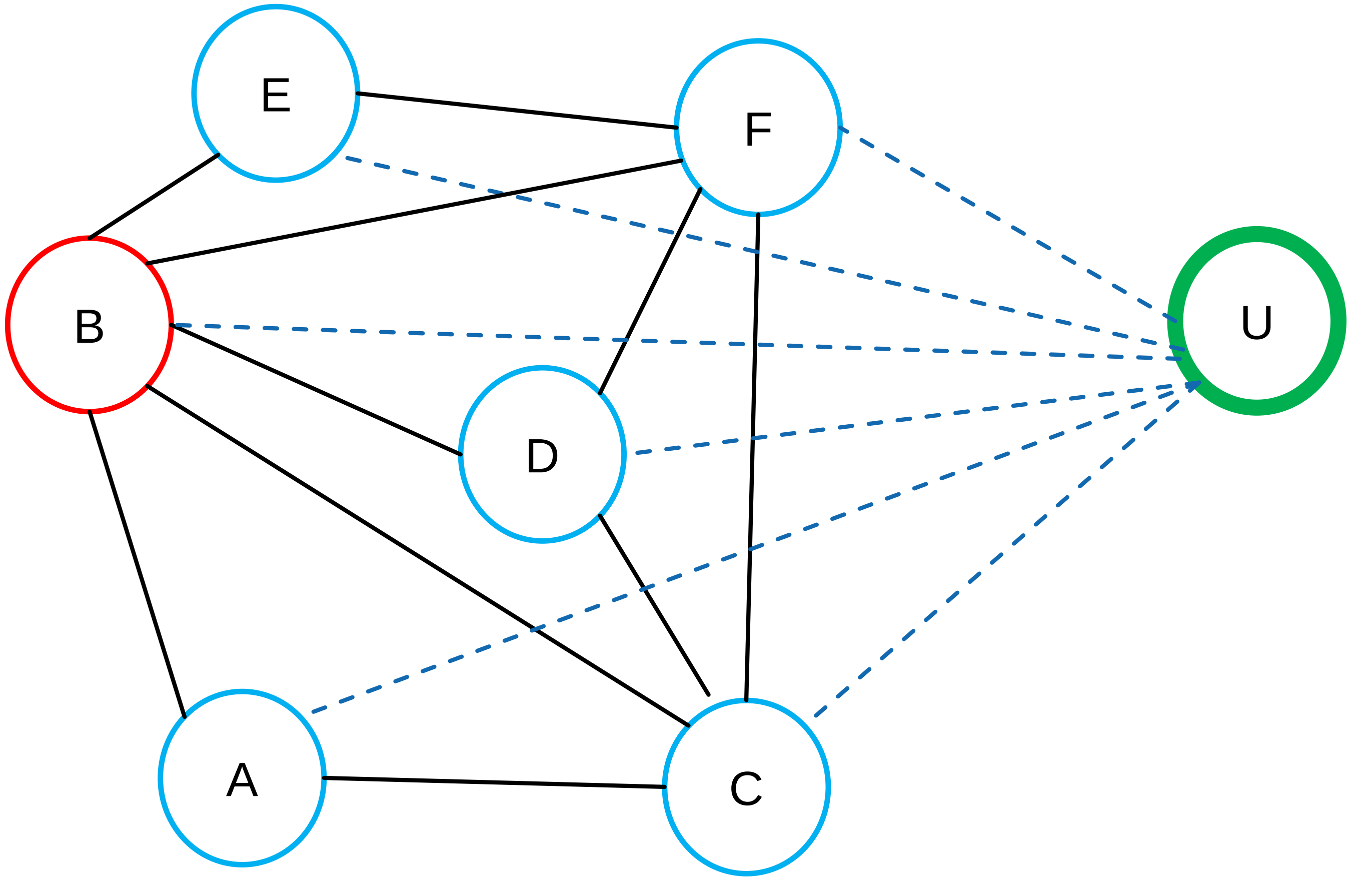
Embeddedness



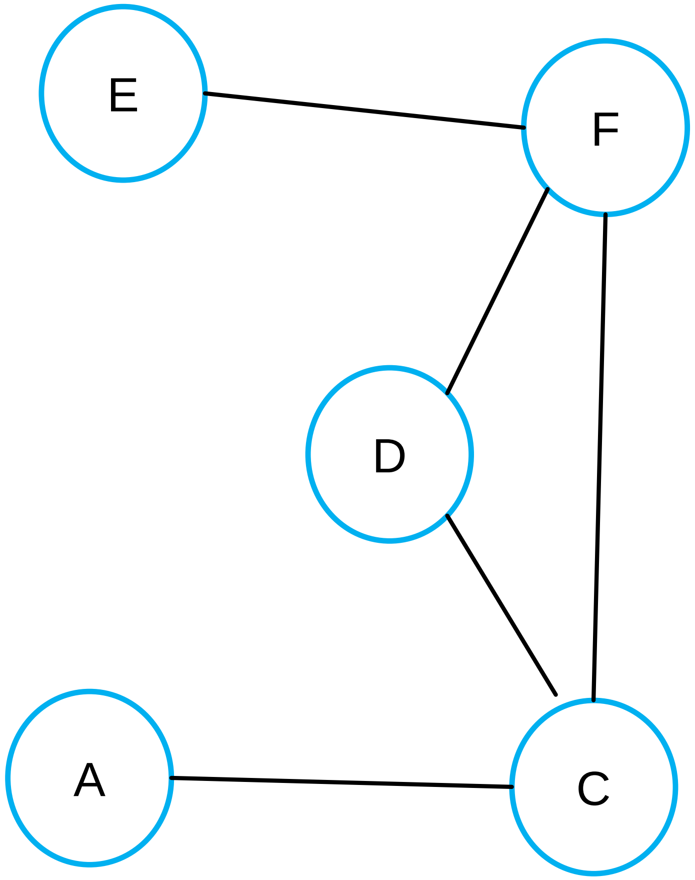
Dispersion

- Take into account how mutual friends are connected
- Partners introduce each other to different groups of friends
- Connection between different groups
- Not directly connected in graph of common friends
- No common neighbours

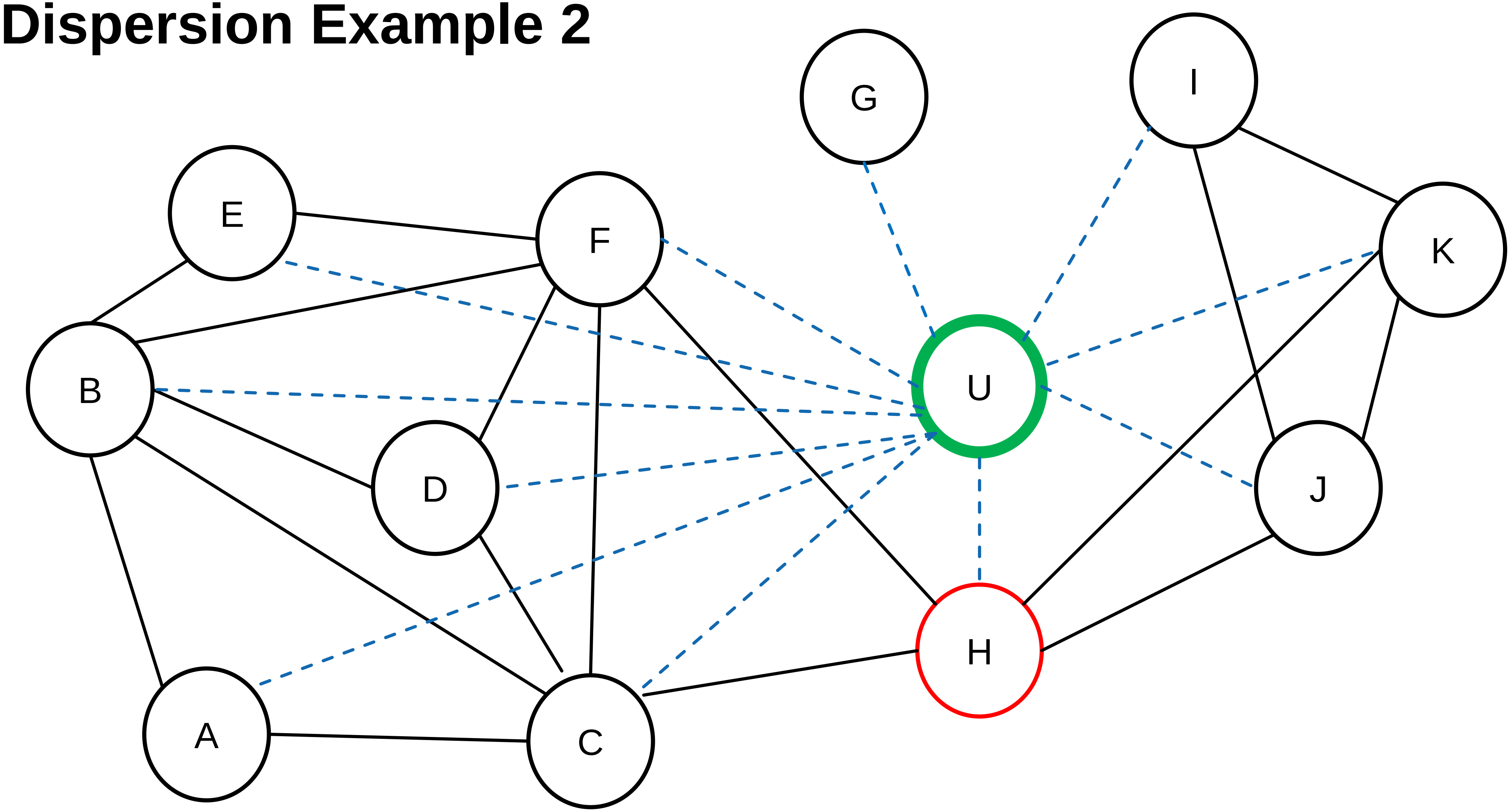
Dispersion Example 1



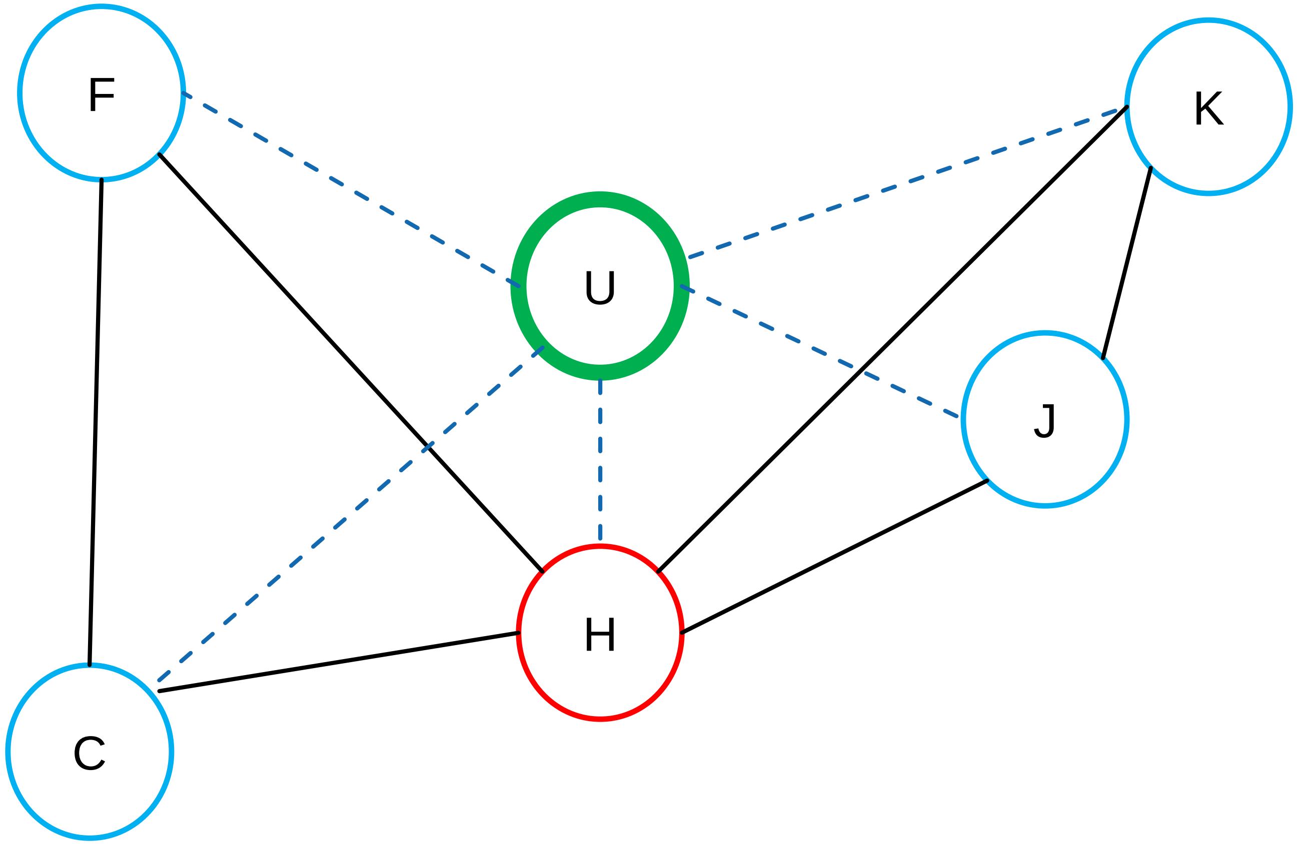
Dispersion Example 1



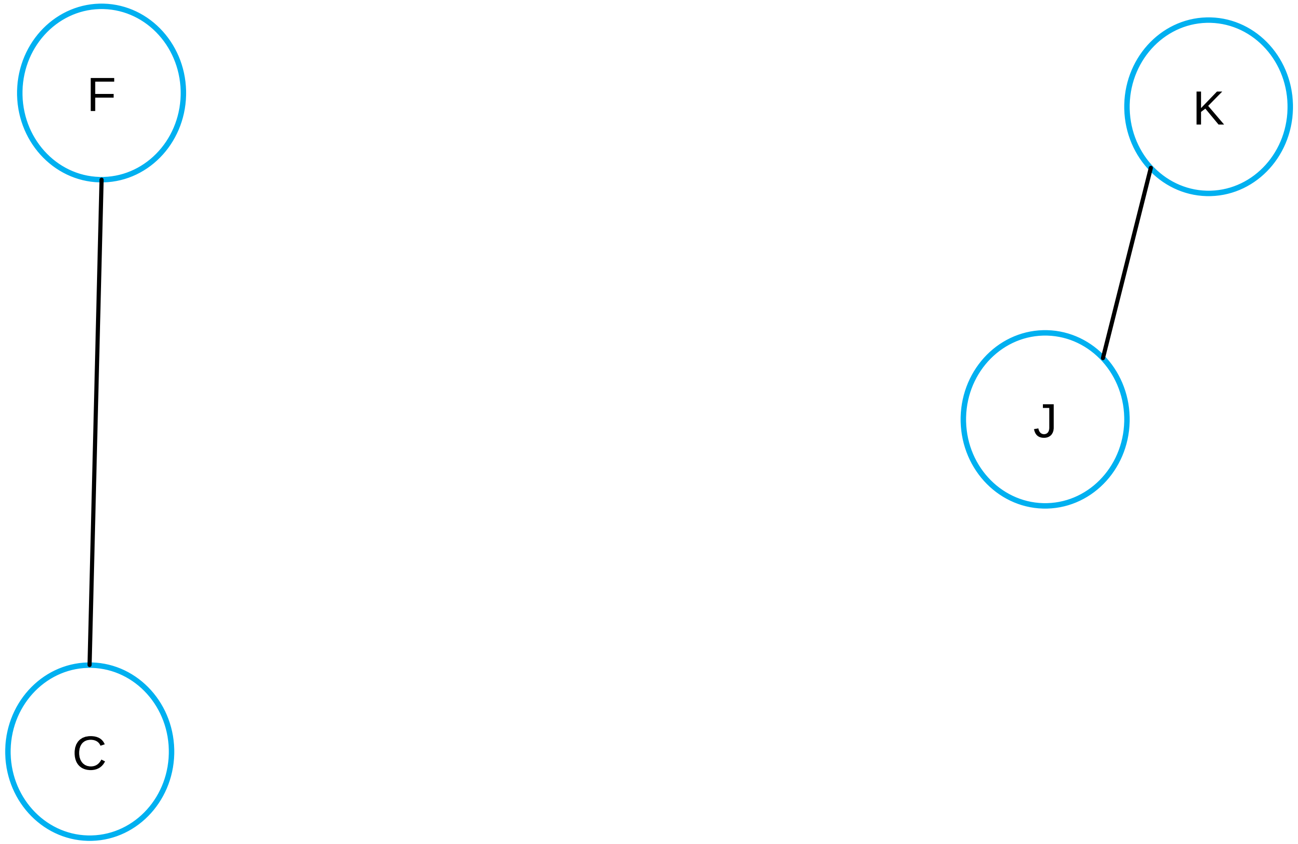
Dispersion Example 2



Dispersion Example 2



Dispersion Example 2



Dispersion

- More Mathematically
  - $disp(u, v) = \sum_{s,t \in C_{uv}} d_v(s, t)$
  - Where $d_v(s, t)$ = 1 only if s and t are not directly linked in G_{uv} and have no common neighbours in $G_u - \{u, v\}$
  - $d_v(s, t)$ = 0 otherwise

Normalized Dispersion

- For a fixed $disp(u,v)$ increasing $emb(u,v)$ decreases performance
- The normalized dispersion

$$norm(u,v) = \frac{disp(u,v)}{emb(u,v)}$$

$$norm(u,v) = \frac{(disp(u,v) + b)^{\alpha}}{(emb(u,v) + c)}$$

Recursive Dispersion

- Take into account dispersion of common neighbours who have a high dispersion on the link to u
- Initially $x_v = 1$ for all neighbours v of u, and iteratively update each x_v

$$x_v = \frac{\sum_{w \in C_{uv}} x_w^2 + 2 \sum_{s,t \in C_{uv}} d_v(s,t) x_s x_t}{emb(u,v)}$$

Data Set

- Randomly sampled Facebook users who declared a relationship in their profile
  - Married
  - Engaged
  - In a relationship
- «Extended Dataset»
  - 1.3 million Facebook users
  - Average 291 nodes and 6652 links
- «Primary dataset»
  - 73000 Neighbourhoods
  - At most 25000 links

Performance of Structural and Interaction Measures (1)

- Structural
  - Embeddedness
  - Recursive Dispersion
- Interaction
  - Rank neighbours based on how many times their profile was viewed by u
  - Rank neighbours based on number of photos they appear with u

Performance of Structural and Interaction Measures(2)

Type	Embed	Rec.disp.	Photo	Prof.view.
All	0.247	0.506	0.415	0.301
Married	0.321	0.607	0.449	0.210
Engaged	0.179	0.446	0.442	0.391
Relationship	0.132	0.344	0.347	0.441

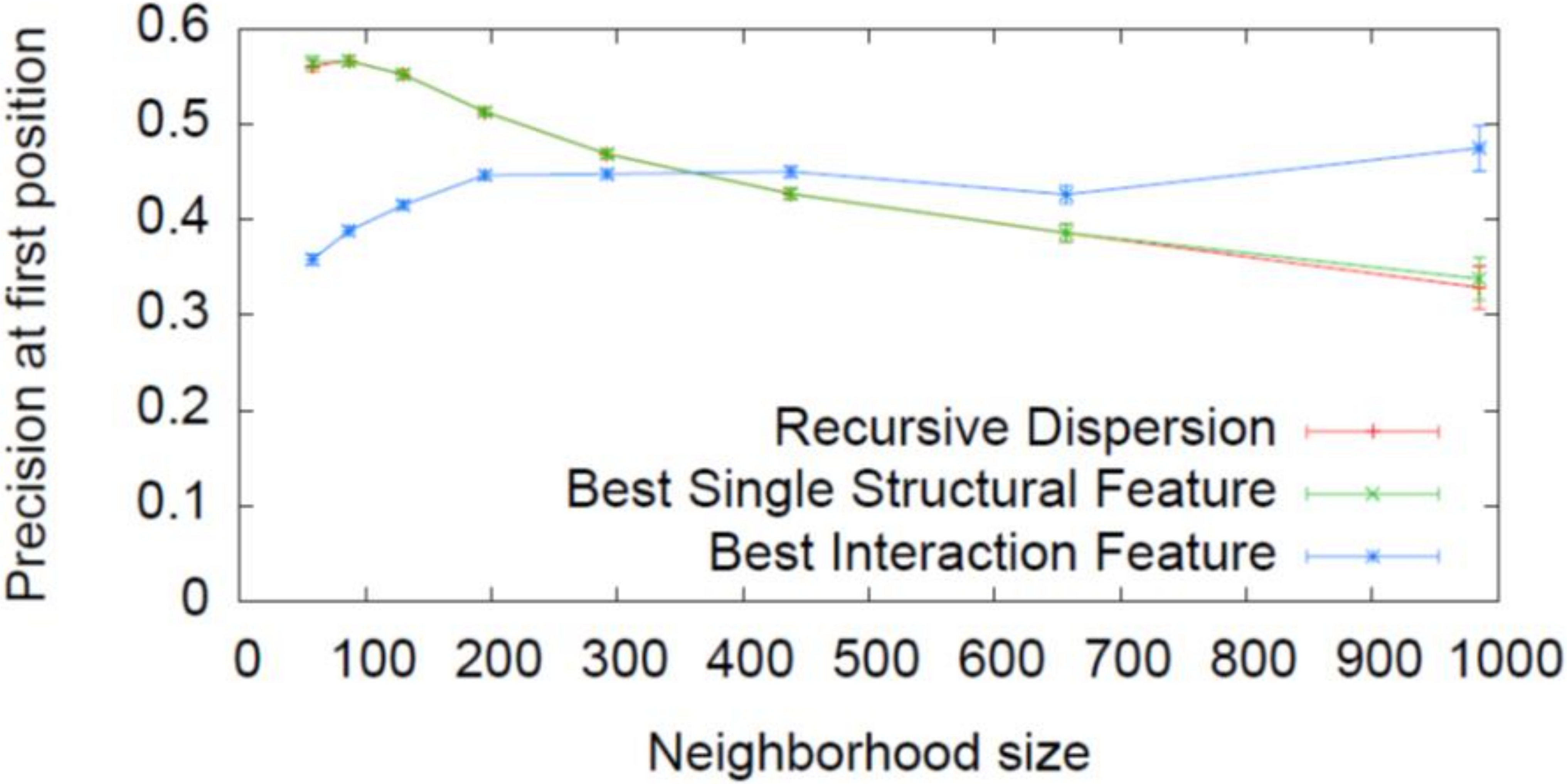
Performance of Structural and Interaction Measures(3)

Type	Embed	Rec.disp.	Photo	Prof.view.
Married (male)	0.347	0.667	0.511	0.220
Married (female)	0.296	0.551	0.391	0.202

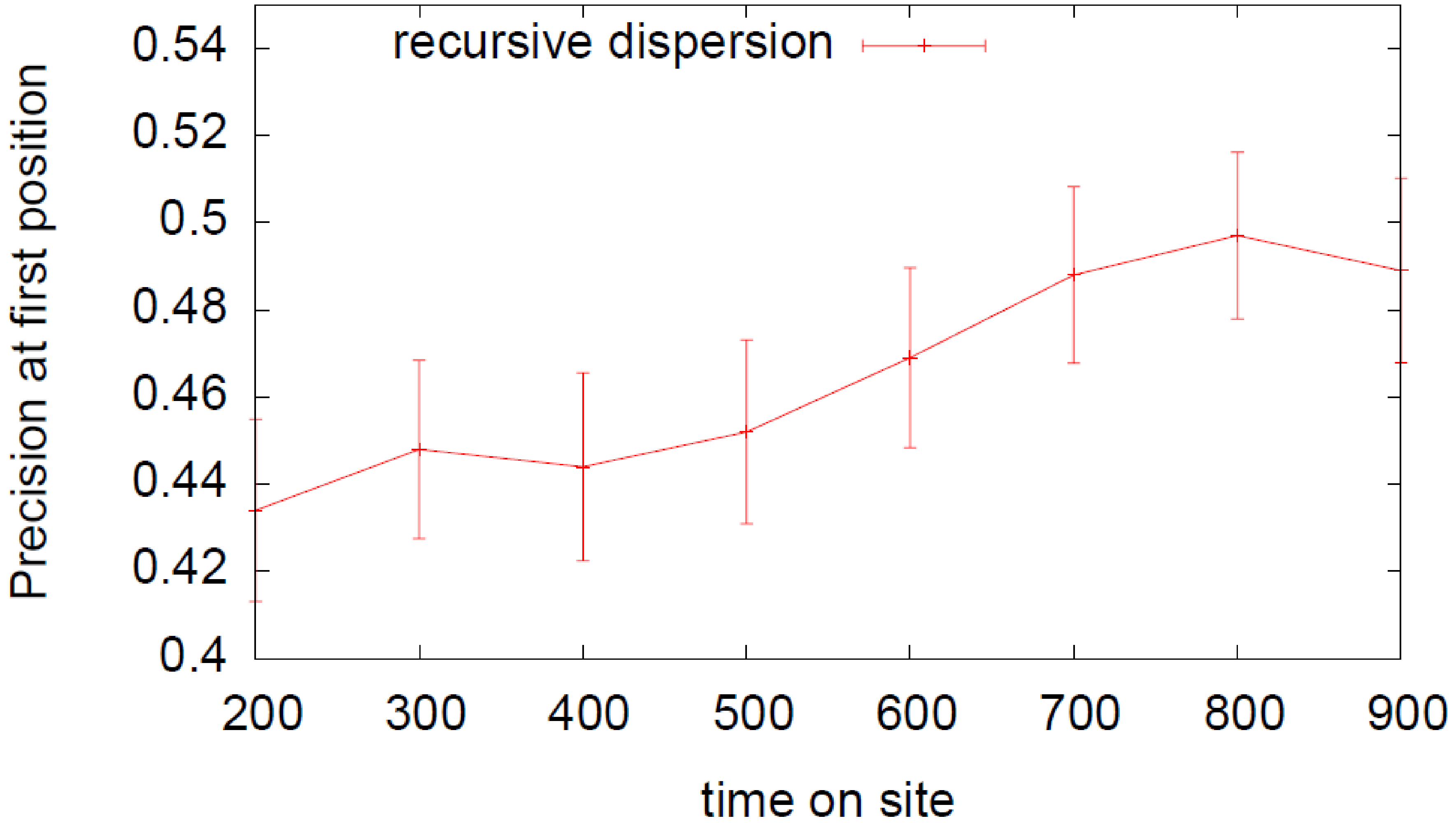
Influencing Factors

- Neighbourhood size
- Time on site
- Time since relationship reported

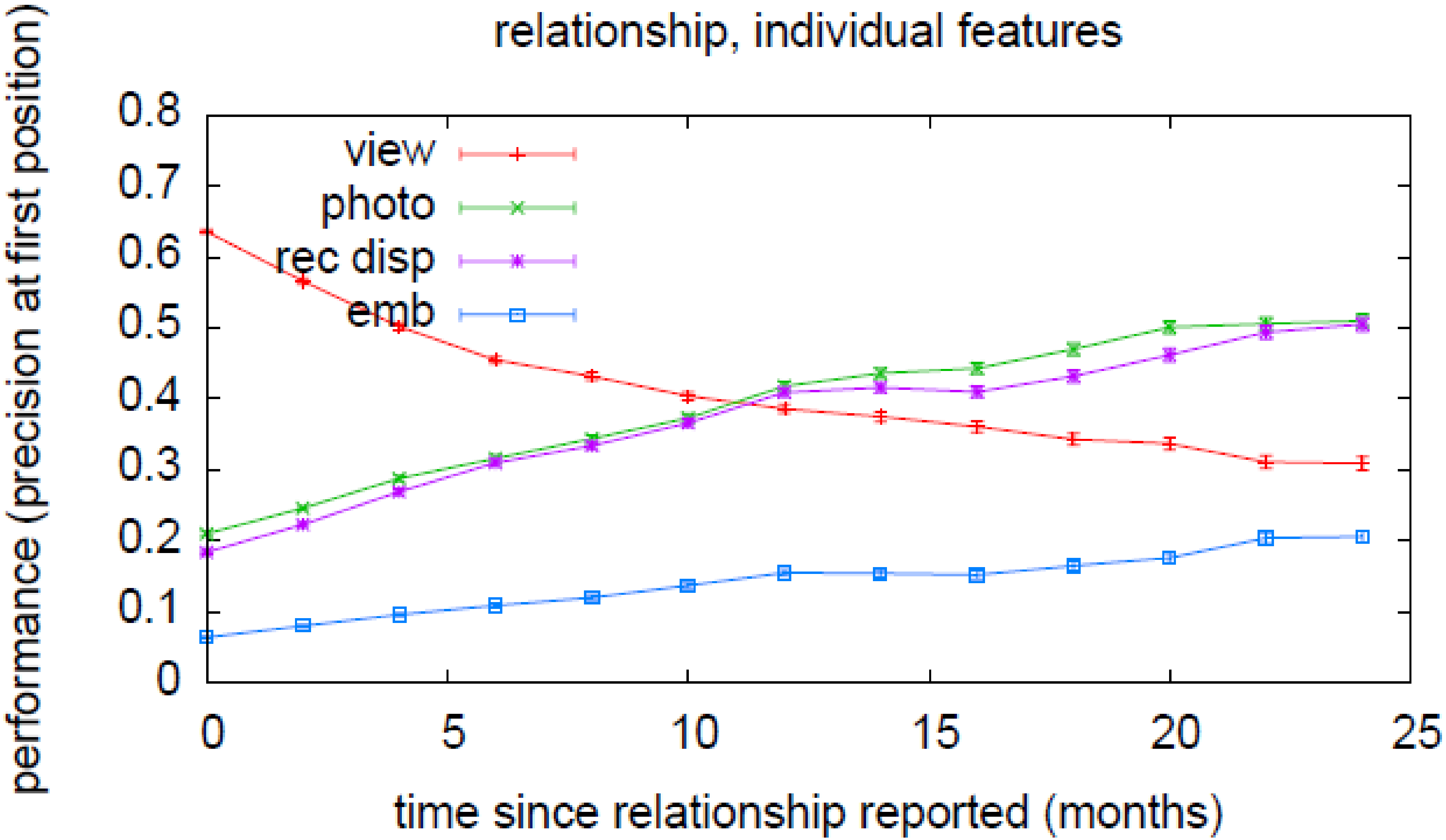
Neighbourhood Size



Time on Site


recursive dispersion
Precision at first position
0.54
0.52
0.5
0.48
0.46
0.44
0.42
0.4
200
300
400
500
600
700
800
900
time on site

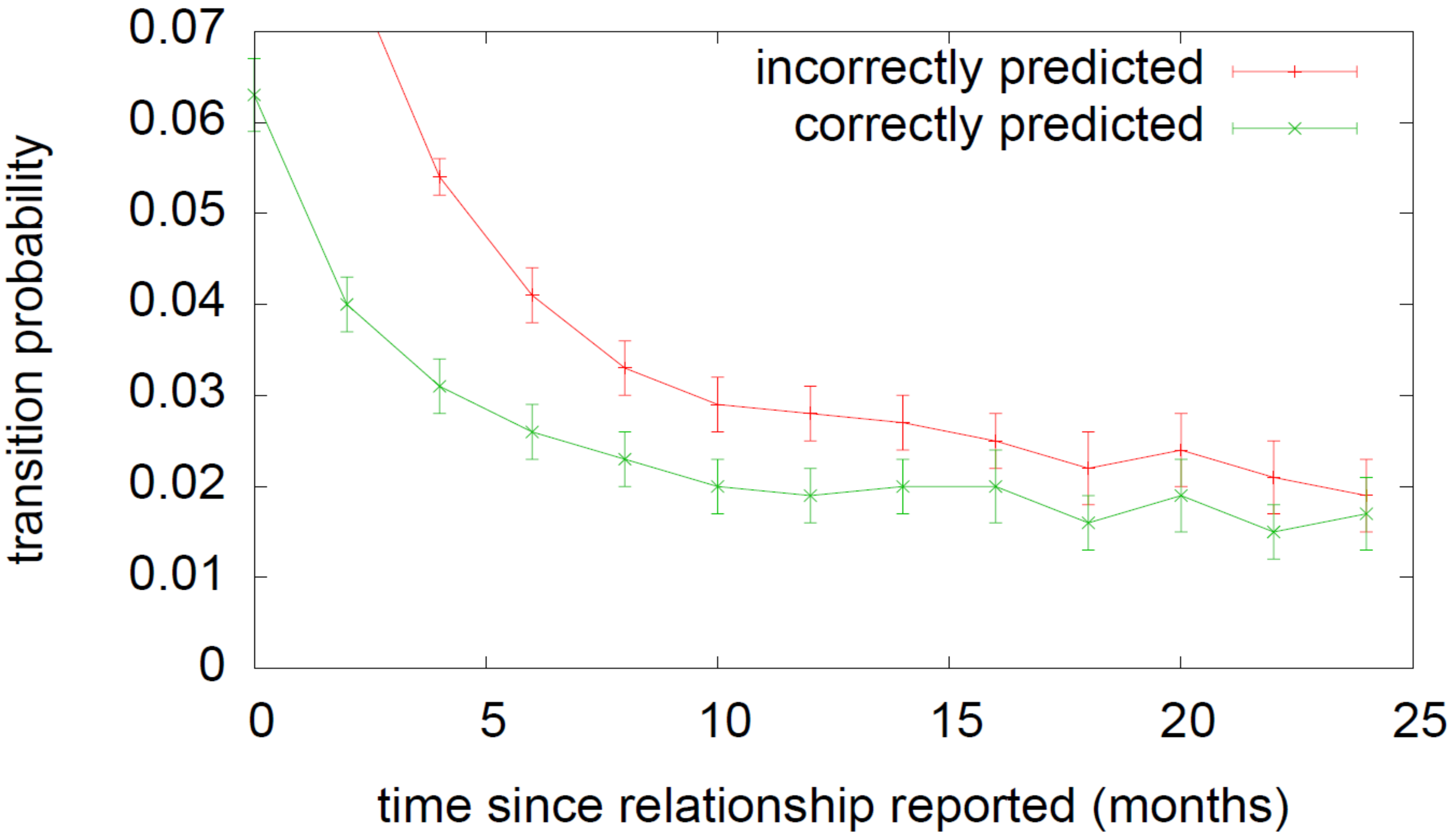
Time since Relationship Reported



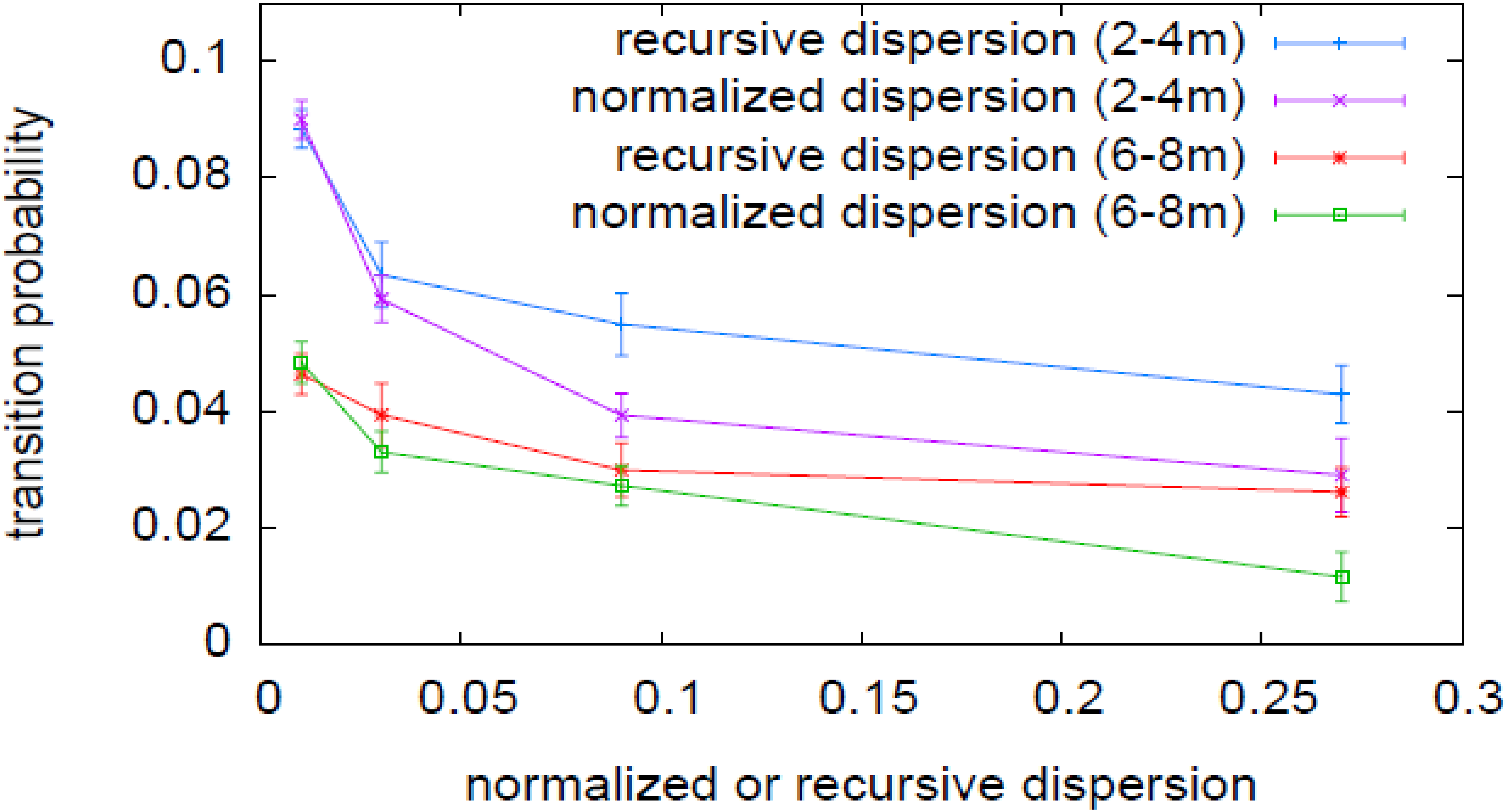
Other Findings

- Increased precision when relationship approaches the status «married»
- What if the partner was not predicted correctly?
  - Likely to be a family member
  - Increased chance of a transition from the status «in a relationship» to «single»
- Less likely to transition to «single» with high normalized or recursive dispersion

Transition Probability to Single (1)



Transition Probability to Single (2)



Machine Learning Based Approach

- 48 structural features
- 72 interaction features
- Try to find the romantic partner
- Try to predict relationship status

Machine Learning Performance Comparison

Type	Struct.	Inter.	Comb.
All	0.531	0.560	0.705
Married	0.624	0.526	0.716
Engaged	0.472	0.615	0.708
Relationship	0.377	0.605	0.682

Machine Learning Performance

- Based on 2 feature categories
  - Demographic (age, gender, country, time on site)
  - Structural features
  - Combination of these two
- Baseline just predicts the more common category

Type	Baseline	Demo.	Struct.	Both
Single vs. Any Rel.	0.598	0.679	0.616	0.683
Single vs. Married	0.566	0.78	0.661	0.79

Motivation (1)

Motivation (2)

- Classify relationships based on machine learning
  - Help organizing and prioritizing communication
  - Based on communication patterns and contact data
- Prevent self-disclosure and inappropriate social behaviour
  - Reminding people of the role they should enact
- Up to now only manually labelling groups
  - People are too lazy
- Groups created need to be periodically updated

Kind of Relationships

- Suggestion to classify users contact in 3 categories
  - Family
  - Work
  - Social

Data Collection

- Collected data from Android smartphones
  - Contact list
  - Call logs
  - SMS logs
  - Facebook friend list
- 40 Participants
  - Average age 28
  - At least 50 «friends» on Facebook
  - 55% students, 35% employed and 10% unemployed

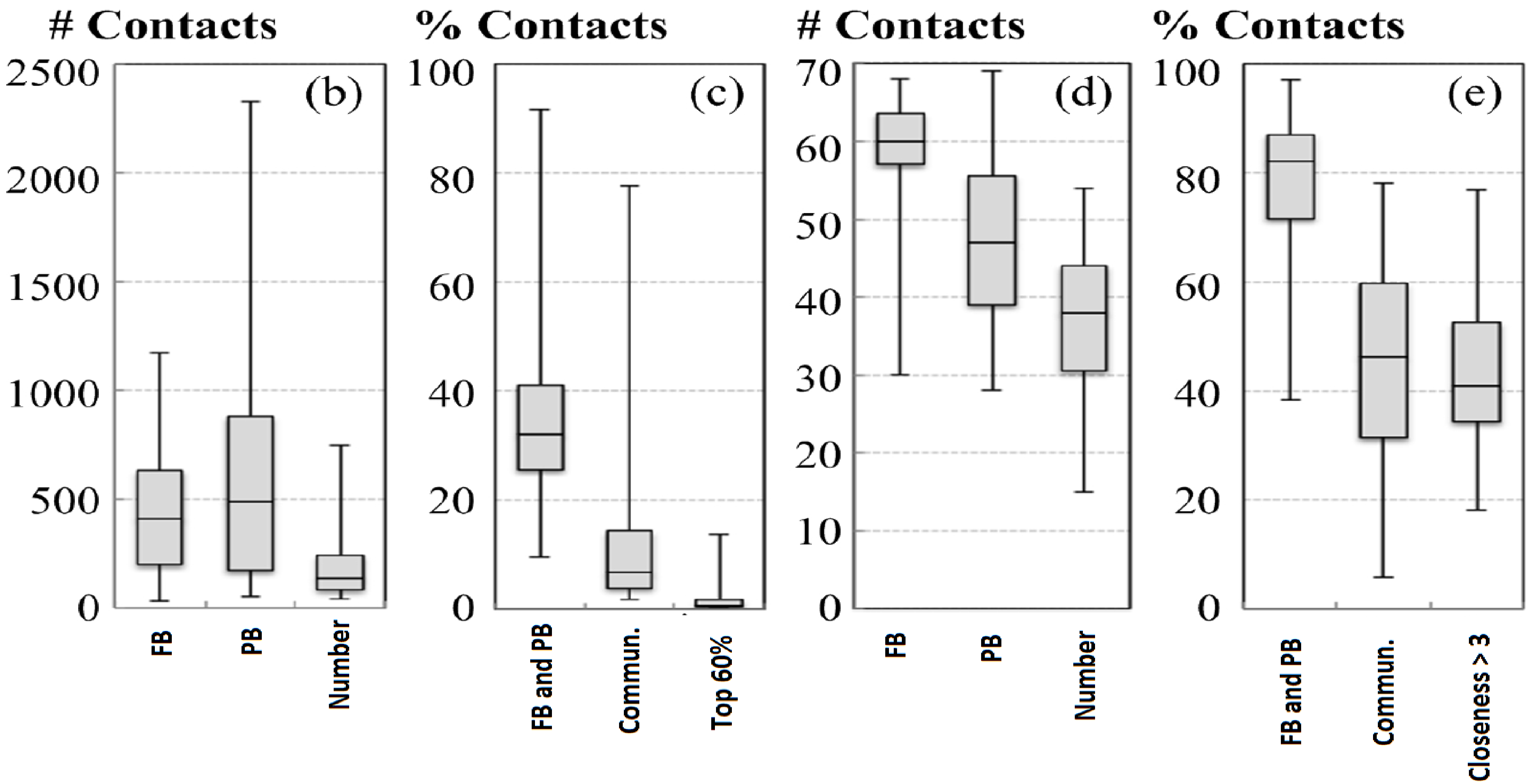
Ground Truth (1)

- Participants assign contacts to specific social groups
  - Family members (immediate, extended or significant other)
  - Work
  - Social (School, Hobby, …)
- Contacts may be added to multiple groups

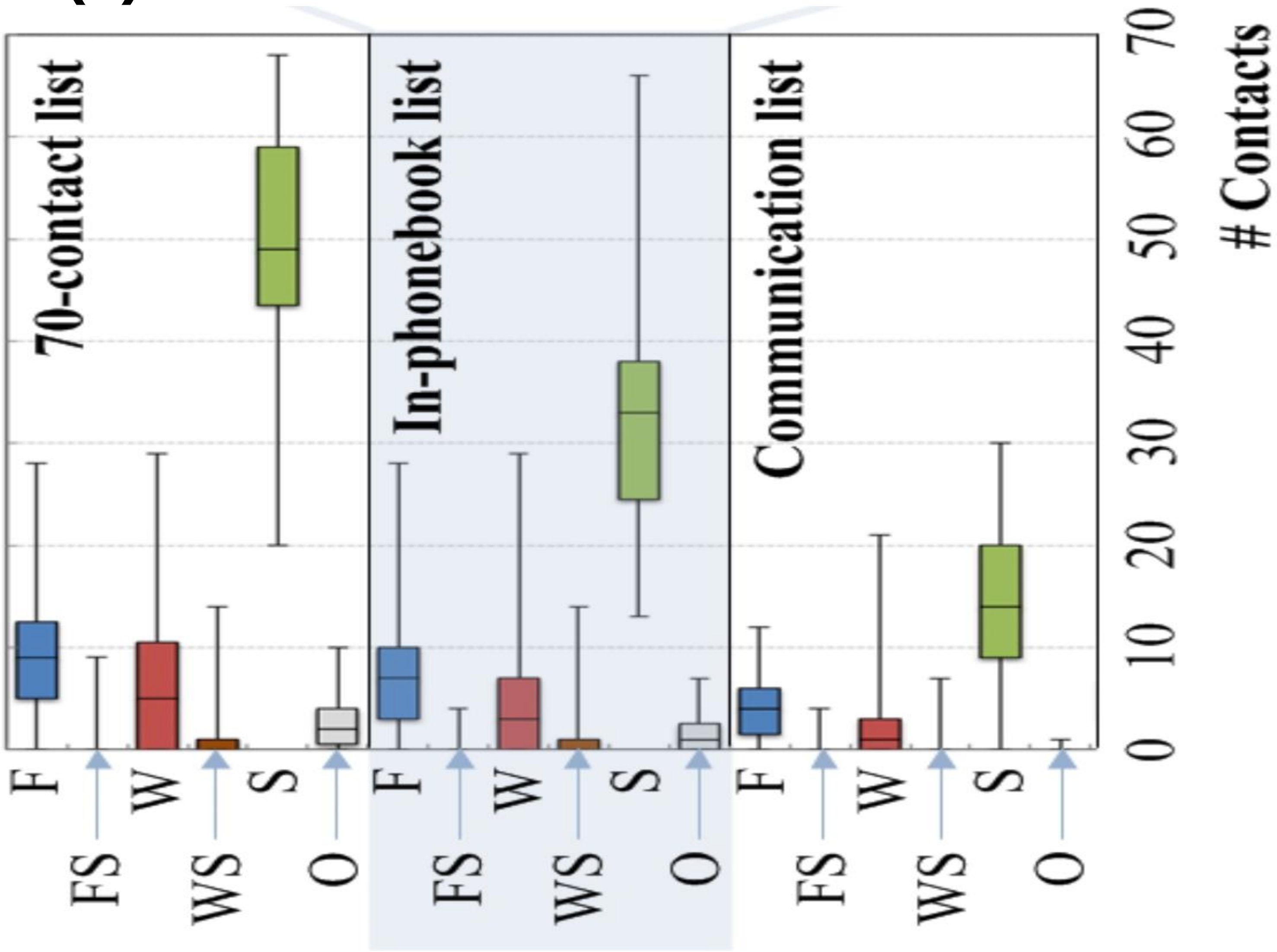
Ground Truth (2)

- «70-contact list»
  - 15 most frequently communicated with (Facebook, Phonebook, SMS)
  - Randomly added individuals from phone's contact list and Facebook friends
  - Removed duplicates
- Participants added relationship information to these contacts
  - Approximate age
  - Gender
  - 5-point scale rating perceived closeness
  - 7-point scale for frequency of interaction

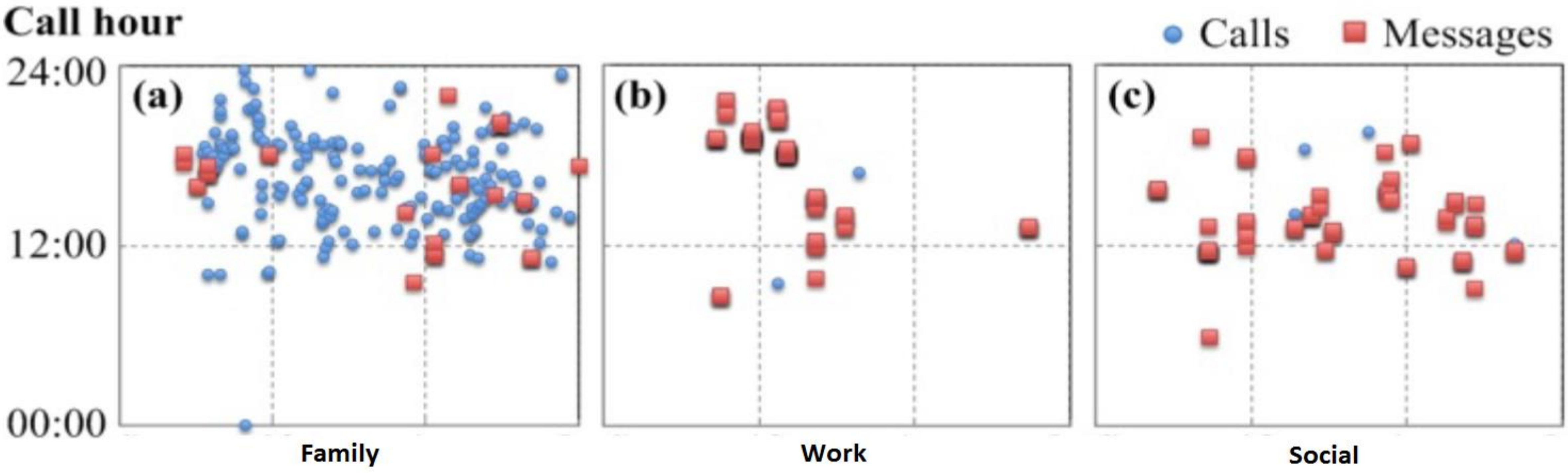
Data Distribution



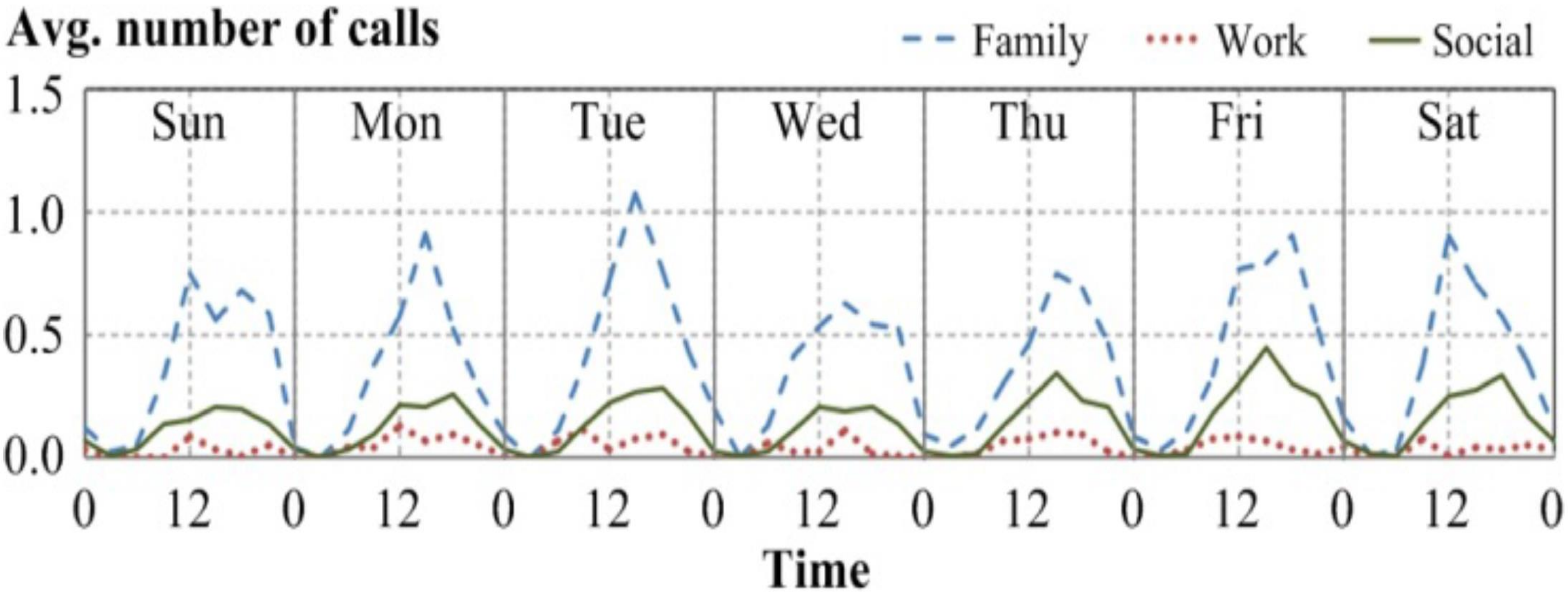
Data Analysis (1)



Data Analysis (2)



Data Analysis (3)



Applying Machine Learning (1)

- Profile Features
  - Basic contact information
- Survey Features
  - Same gender
  - Age difference
  - Is Facebook friend
  - Frequency seen
- Mobile-communication Features
  - Call logs (duration, incoming/outgoing, frequency, time of day)
  - MSG logs (length, incoming/outgoing, time of day)
  - COMM (Call & MSG logs)

Applying Machine Learning (2)

- Built model with SVMs (Support vector machines)
  - Binary classifier
  - Used 3 pairwise SVMs (family vs. work, family vs. social, work vs. social)
- Compared to two other models
  - WEKA: rule-based model (decision tree C4.5)
  - Probabalistic model (naïve Bayes)

Results

Data Set	Method	All Features
70-contact list	SVM	81.1%
	C4.5	74.2%
	NB	67.1%
In-phone book list	SVM	83.1%
	C4.5	76.1%
	NB	65.6%
Comm. list	SVM	90.5%
	C4.5	75.6%
	NB	76.6%

- Comm. list: Mobile communication features > Survey features > Profile features
- PB list: Survey features > Mobile communication features > Profile features

Correlation (1)

- Family
  - Call intensity
  - Event type (e.g. birthday)
  - Is starred
- Work
  - Call intensity (negatively)
  - Weekday communication
  - Frequency seen
- Social
  - Channel selection
  - Weekend communication
  - Age difference
  - Is starred

Correlation (2)

Profile	Family	Work	Social
Family	175	186	35
Work	68	247	71
Social	77	160	202

Correlation (2)

Mobile	Family	Work	Social
Family	288	31	76
Work	49	112	198
Social	15	23	430

Correlation (2)

Both	Family	Work	Social
Family	290	28	76
Work	46	124	198
Social	15	22	432

Correlation (2)

All	Family	Work	Social
Family	298	43	53
Work	27	305	44
Social	20	20	411

Limitations

- Android only stores 400 recent calls and 200 message per contact
- Limited to social media
- Small sample size (n=40)
- Only participants of the US (culture)
- Privacy concerns may lead to self-selection bias
- No incorporation of location data
- Only three groups

Summary (Romantic Partnership and the Dispersion of Social Ties: A Network Analysis of Relationship Status on Facebook)

- One basic question
- New measure dispersion
- Can identify romantic partners with high precision
  - Up to 60%
  - Better for male
- If prediction was wrong, it is most likely a family member
- Predictions about robustness of relationships
- Compared different features using machine learning

Summary (Mining Smartphone Data to Classify Life-Facets of Social Relationships)

- Classify contacts based on machine learning
- Three life facets (family, work, social)
- Three different feature categories (Profile, Survey, Mobile)
- If there are any communication logs the accuracy is high (90%)
- Correlations for different categories

Sources

- Romantic Partnership and the Dispersion of Social Ties: A Network Analysis of Relationship Status on Facebook
  - Lars Backstrom (Facebook Inc.), Jon Kleinberg (Cornell University)
  - CSCW' 14
- Mining Smartphone Data to Classify Life-Facets of Social Relationships
  - Jun-Ki Min, Jason Wiese, Jason I. Hong, John Zimmermann (Carnegie Mellon University)
  - CSCW' 13

Questions?