Inferring Geographic Coincidence in Ephemeral Social Networks

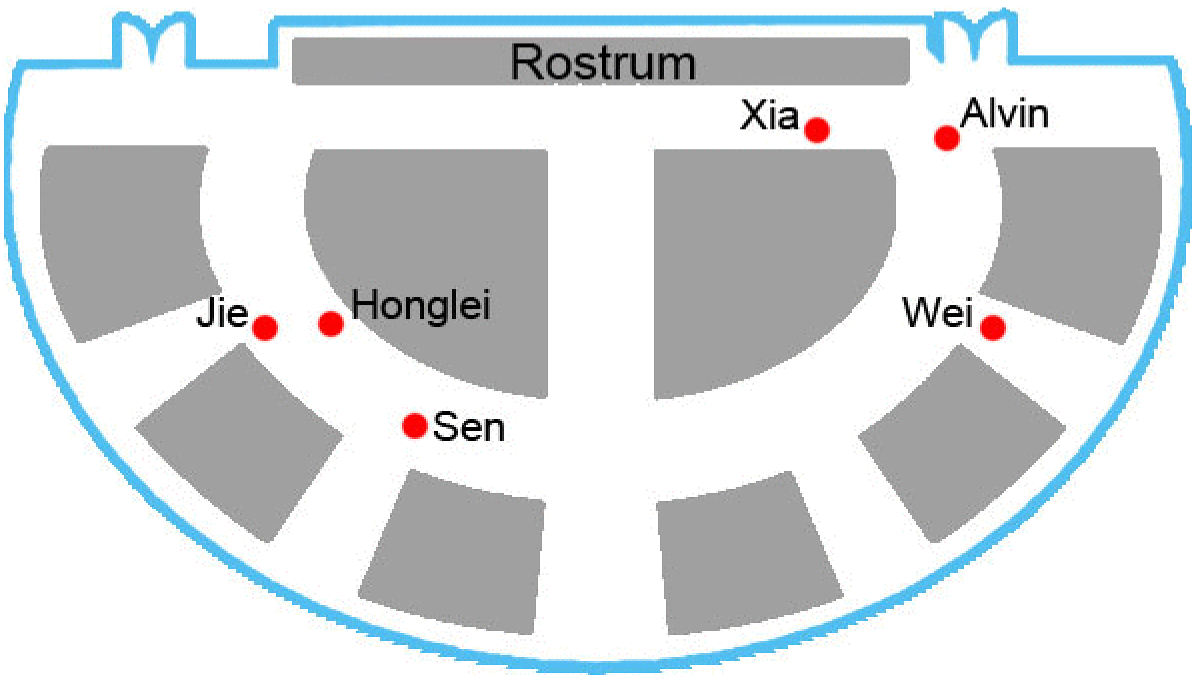
Honglei Zhuang[1], Alvin Chin[2], Sen Wu[1],
Wei Wang[2], Xia Wang[2], Jie Tang[1]

[1]Tsinghua University
[2]Nokia Research Center Beijing

2012/9/26

What is an Ephemeral Social Network?

- A social network temporarily created during an event such as a conference, game or banquet.



Questions: W^3

- *Who do you want to meet at a conference?*
- *Who did you finally chat with?*
- *What made a conversation last up to 2 hours?*

Research Problem

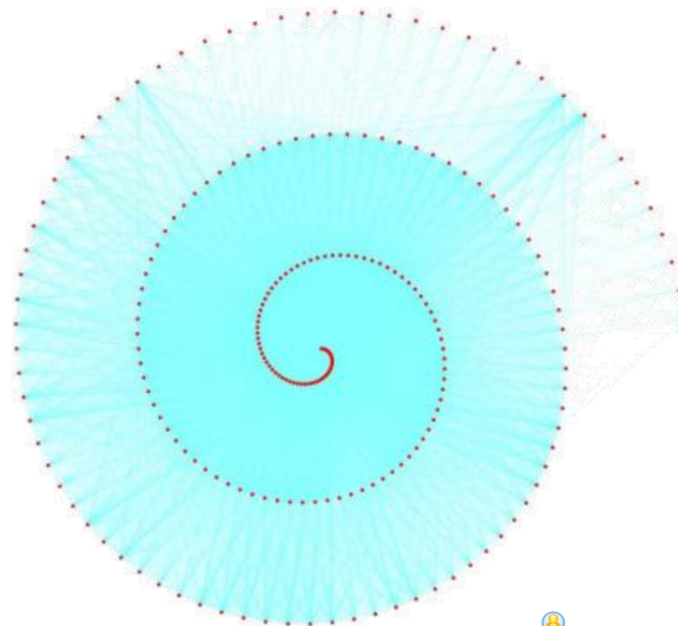
- An interesting and basic question to explore in an ephemeral social network is, can we recommend and predict whom a user will meet in the future?

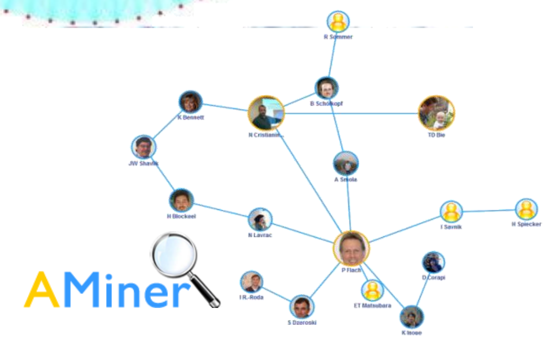
Nokia Find & Connect System

- Designed for events including Ubicomp 2011.
- Allows users to locate friends, check users nearby, and share personal information.

Data Sets

- **UbiComp**
  - Location logs are collected by Find & Connect system during *UbiComp*.
  - Author information, including publication list, coauthor relationship are provided by *ArnetMiner*.

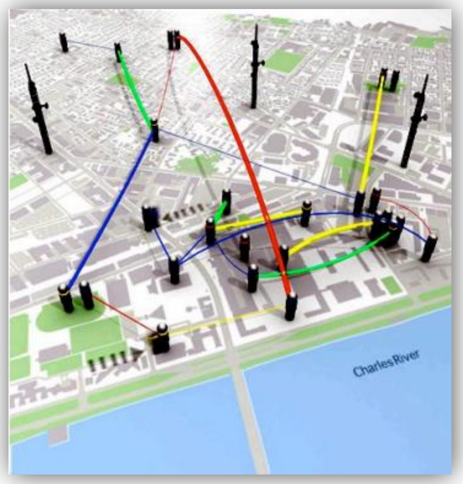




Ubicomp2012: http://www.ubicomp.org/ubicomp2011/
ArnetMiner: http://aminer.org

Data sets (cont')

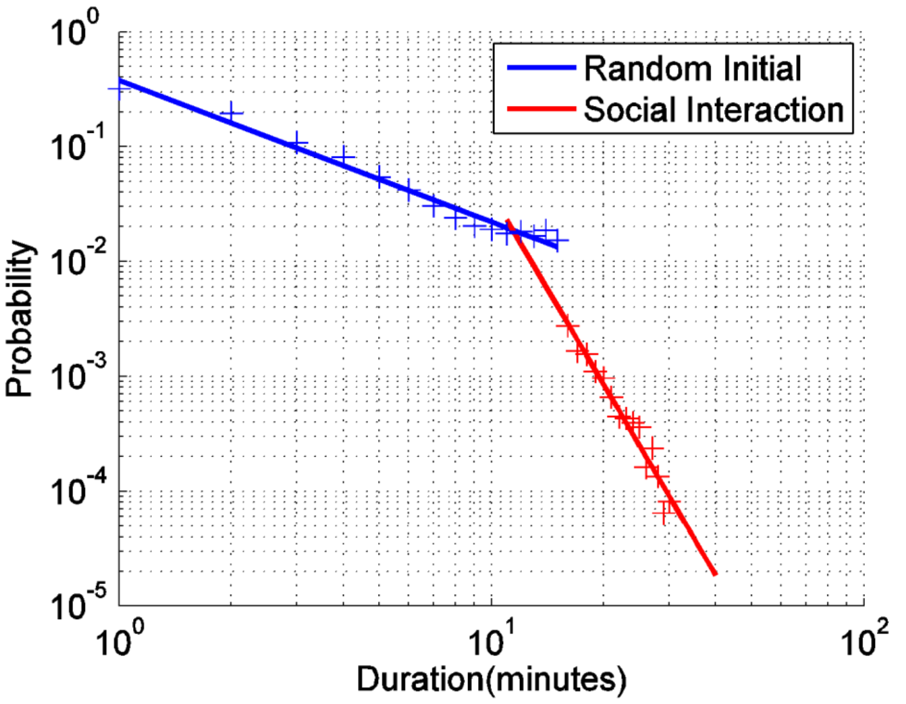
- **Reality** [Eagle et al. 2009]
  - Collected by MIT Media Lab from 2004-2005.
  - Physical proximity logs are recognized by Bluetooth device on mobile phones of 106 users.
  - A survey for all subjects provides the friendship network.



Data sets	UbiComp	Reality
User number	234	106
Duration	3 days	10 months
Logs number	69,844	~ 4,000,000
Normal network	Coauthor	Friendship

Observation 1: 10-minute Phenomenon

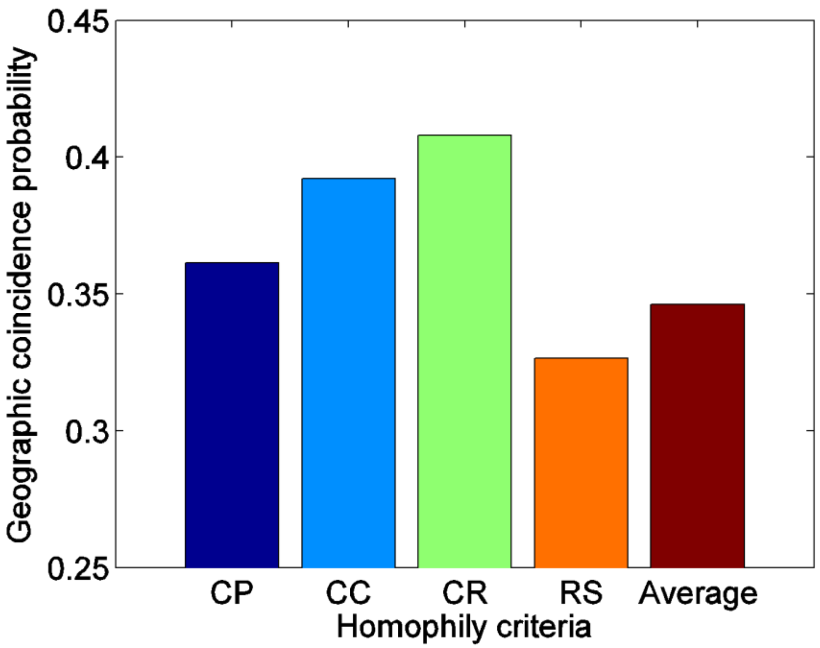
- Probability of a geographic coincidence vs. its duration.
- A two-stage power-law distribution where 10-minute is the inflexion point.
- Implication: geographic coincidences with real social interaction behind would usually last for longer than 10 minutes.



Observation 2: Old Folks or New Ideas?

- Does homophily play an important role in geographic coincidences between people?
- Probability of geographic coincidences for user pairs with the most common paper, common conference, and common research interests.

- Users with high research similarity are less likely to encounter.
- Users with high common papers/conference/coauthors are still more likely to encounter.

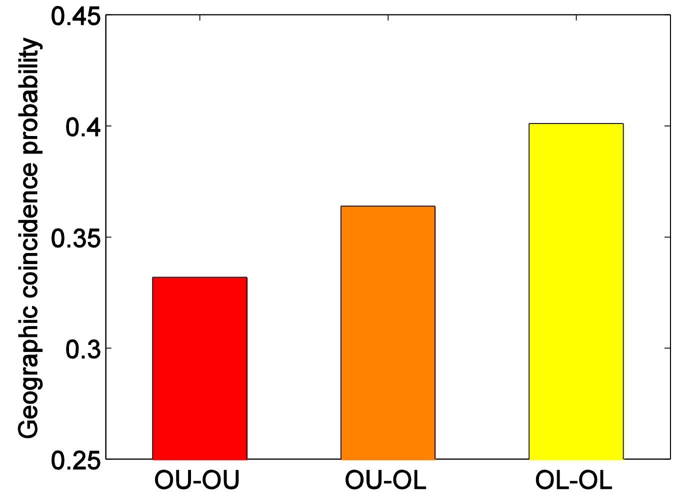


*CP for coauthored papers, CC for common coauthors, CR for common conference, RS for similarity of research interests.

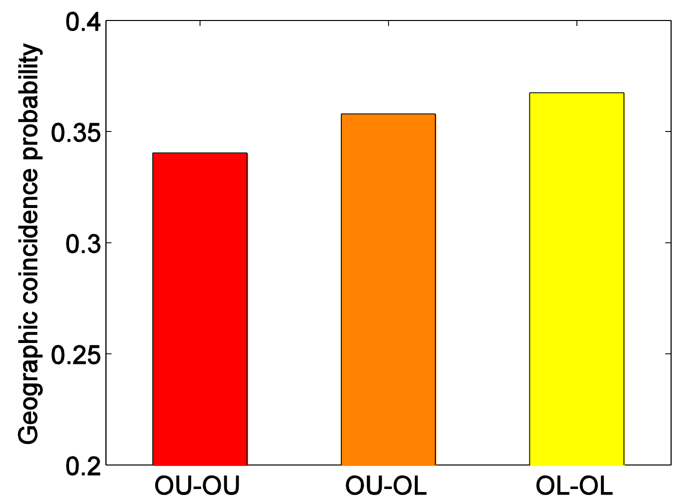
Observation 3: Opinion Leaders

- Opinion leaders are more likely to have geographic coincidences with other people.
- A somewhat surprising observation is, two opinion leaders are even more likely to have geographic coincidences than the pair of an opinion leader and another ordinary user.

*OU for ordinary users, OL for opinion leaders

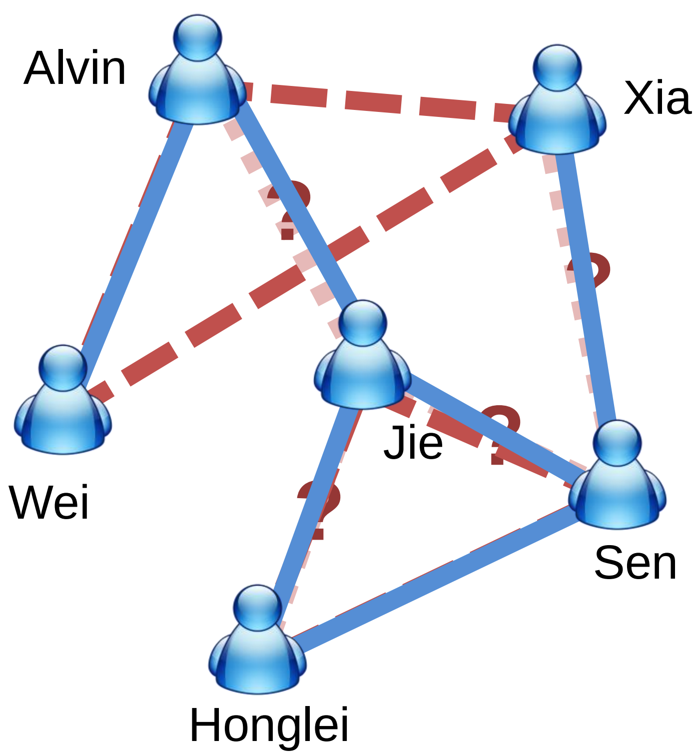


(a) Publication count



(b) H-index

Problem Formulation

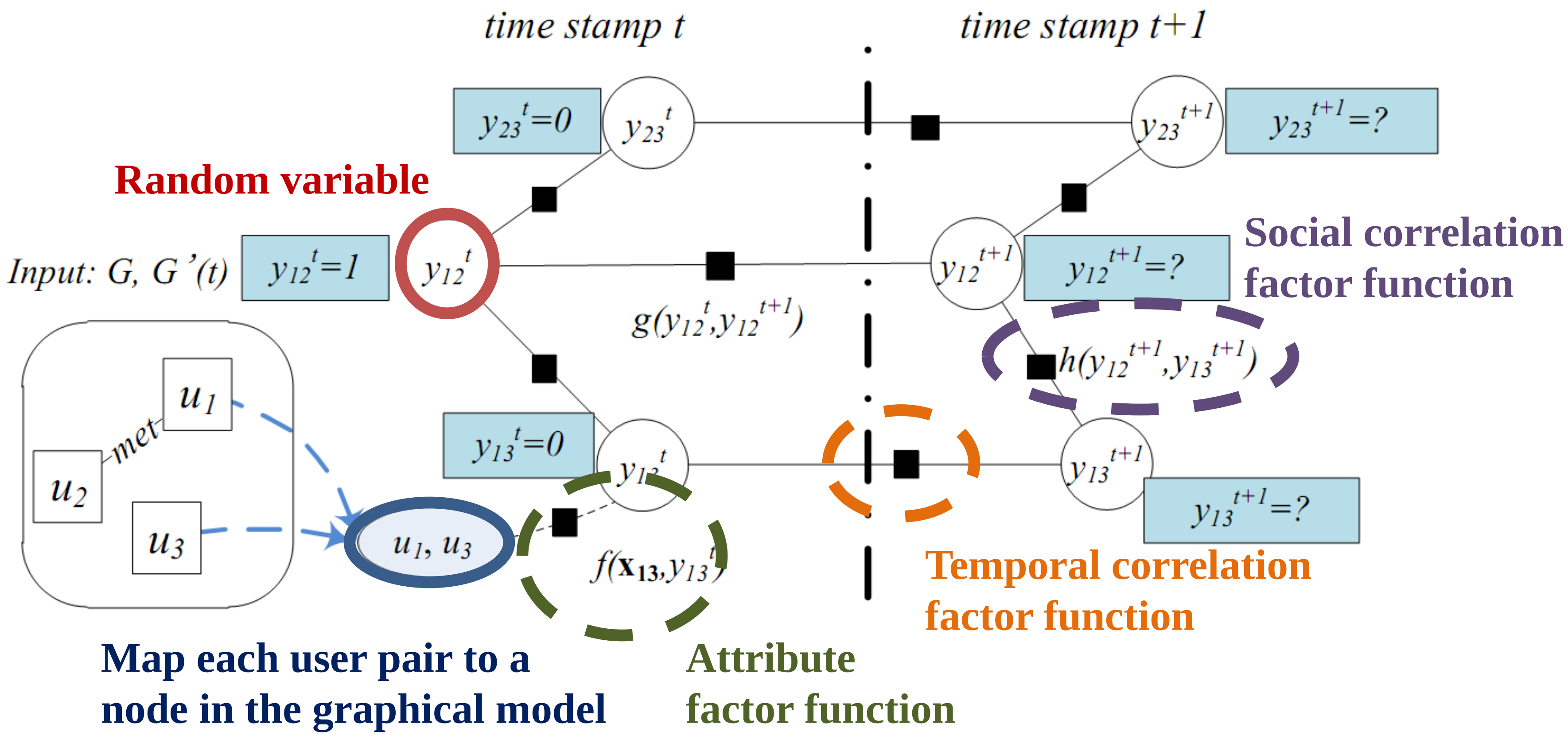


Input

Output

$(U^t, \mathbf{X}^t, Y^t)$

Factor Graph Model



Factor Graph Model (cont')

- The objective function can be written as

$$\mathcal{O}(\theta) = \log \sum_{Y|Y_L} \exp \theta^T \mathbf{S} - \log \sum_{Y} \exp \theta^T \mathbf{S}$$

where $$\theta = [\alpha^T, \beta^T, \lambda^T]^T$$

$$\mathbf{S} = [\sum_t \sum_{y_{ij}^t} \mathbf{\Phi}(\mathbf{x}_{ij}, y_{ij}^t), \sum_{i,j} \sum_t \mathbf{g}(y_{ij}^t, y_{ij}^{t+1}), \sum_t \sum_{Y_c^t} \mathbf{h}(Y_c^t)]^T$$

- **Model Learning**
  – To find a parameter configuration s.t.

$$\theta^* = \mathrm{argmax}_\theta P(Y_L|G, G')$$

- **Predicting geographic coincidences**
  – Find a Y_U to optimize the objective function

$$Y_U^* = \mathrm{argmax}_{Y_U} P(Y|G, G')$$

Experimental Results

- **Performance comparison**

Table 2. Prediction performance comparison(%)

Date set	Method	Precision	Recall	F1-score
UbiComp	SVM	**34.5**	20.4	25.6
	CRF	33.2	39.4	36.0
	FGM	34.0	**65.4**	**44.7**
Reality	SVM	84.1	64.4	72.9
	CRF	73.6	**85.8**	79.2
	FGM	**85.1**	81.0	**83.0**

- **Baselines**
  - SVM – only uses attributes of users.
  - CRF – further consider the time correlation.

Experimental Results (cont')

- **Social Correlation Factors Analysis**
  - Performance achieved by only considering one particular social correlation.

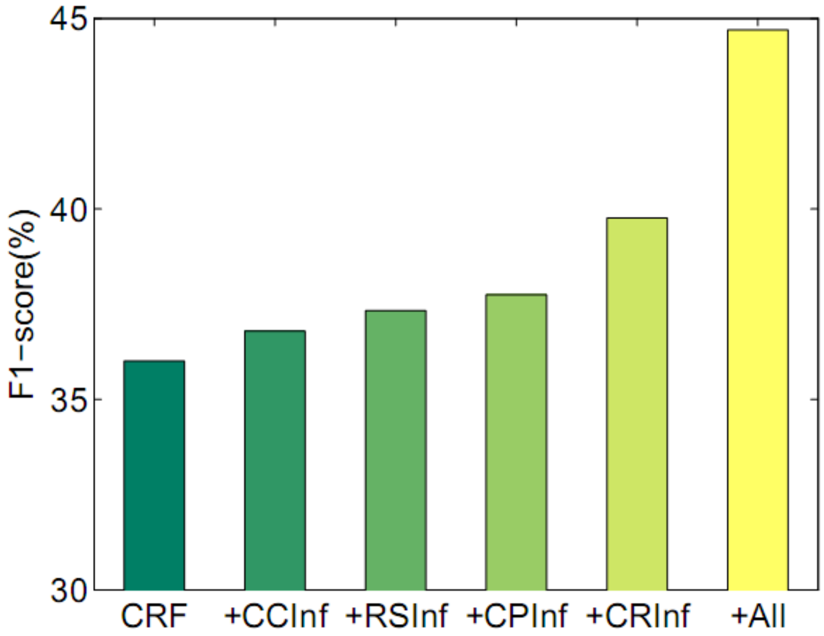
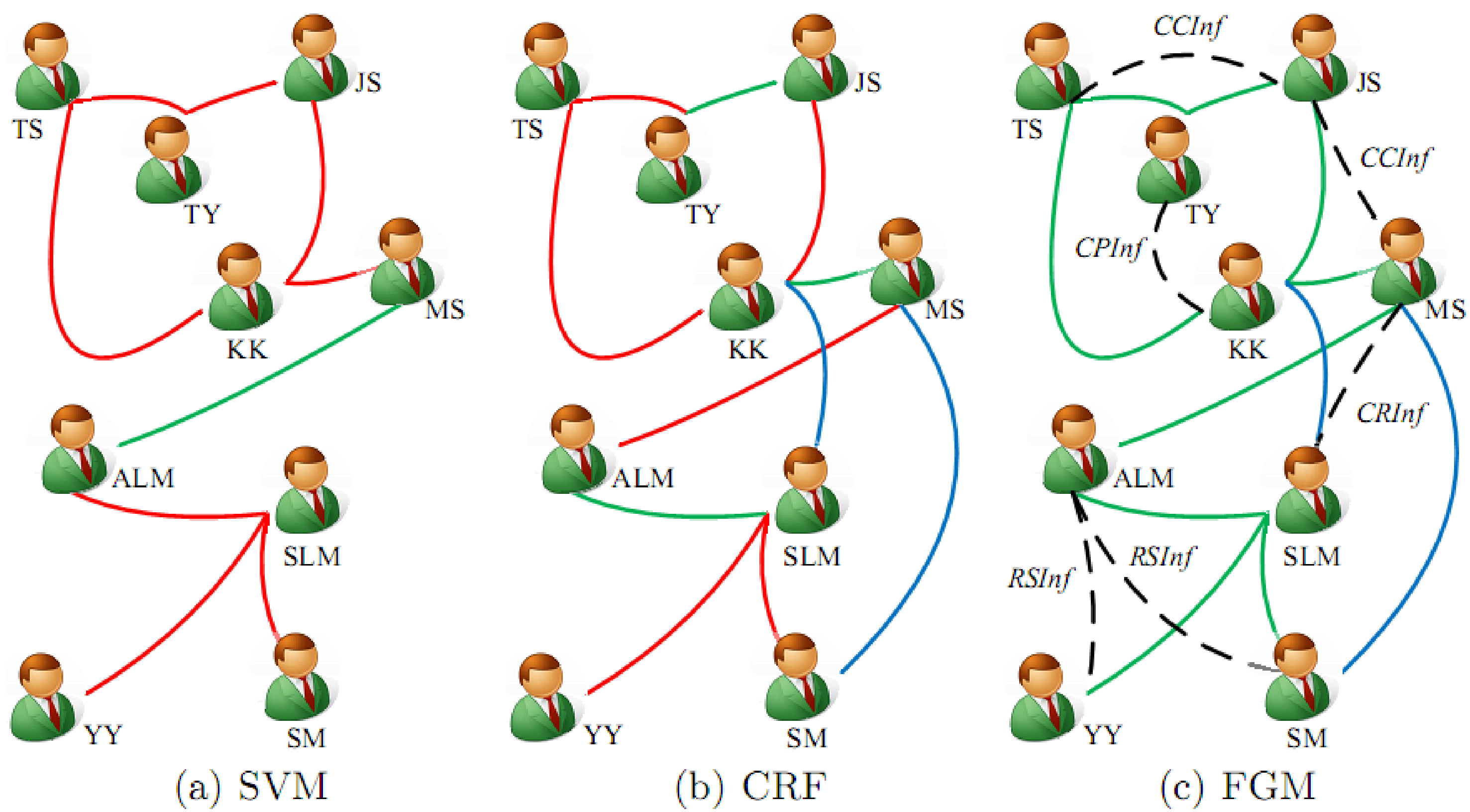


*Figure for UbiComp data set. CPInf for correlation factor defined by coauthor paper count, CCInf for common coauthor count, CRInf for common conference ratio, RSInf for research similarity.

Case Study



*Green for true positive, red for false negative, blue for false positive.
**Dash lines for social correlation.

Summary

- A formal definition of inferring geographic coincidences on ephemeral social networks.
- A series of observations has been conducted on the UbiComp data set.
  - Two-stage power-law distribution for duration where 10-minute becomes the watershed of random encounters and planned talks.
  - Both similar research experiences and different research interests motivate people to meet others.
- A Factor Graph Model has been proposed and tested on two data sets.
  - 3.8~8.7% improvement of F1-score over baselines.
  - Common conference correlation factors can improve the performance the most (~3%)

Future Work

- Studies on ephemeral social network will create a new research direction. It is expected to provide insight on the evolution and formation of traditional social networks.

- By combining information from large-scale online social network (e.g. Facebook), can we further predict when and where a geographic coincidence will occur?

- By employing information from ephemeral social networks, can we predict a new link in online social network?

THANK YOU

2012/9/26