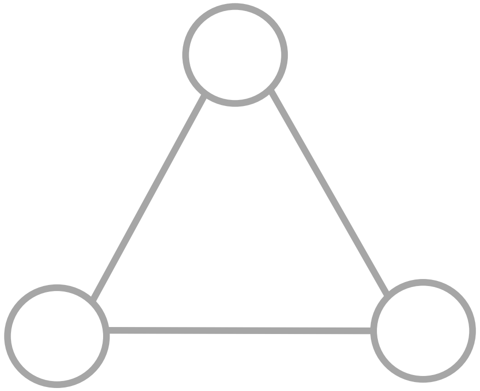
Seoul, Korea
WWW 2014
The 23rd International World Wide Web Conference
April 7-11, 2014 COEX

Mining Triadic Closure Patterns in Social Networks



Hong Huang, University of Goettingen
Jie Tang, Tsinghua University
Sen Wu, Stanford University
Lu Liu, Northwestern University
Xiaoming Fu, University of Goettingen

Networked World

- **1.26 billion** users
- **700 billion** minutes/month

- **555 million** users
- **.5 billion** tweets/day

- **79 million** users per month
- **9.65 billion** items/year

- **280 m**
- **80% o**

- **560 million** users
- **influencing** our daily life

- **500 million** users
- **35 billion** on 11/11

- **800 million** users
- **~50% revenue** from network life

A Trillion Dollar Opportunity

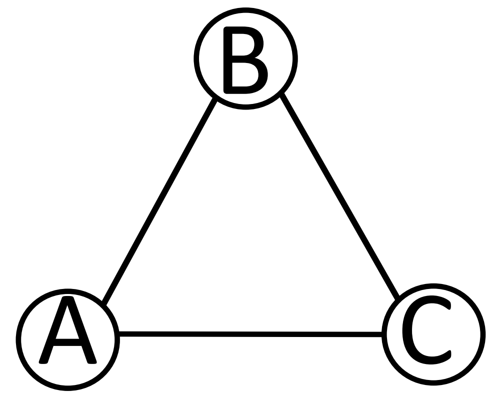
Social networks already become a bridge to connect our daily **physical** life and the **virtual** web space

On2Off **[1]**

[1] Online to Offline is trillion dollar business
http://techcrunch.com/2010/08/07/why-online2offline-commerce-is-a-trillion-dollar-opportunity/

"Triangle Laws"

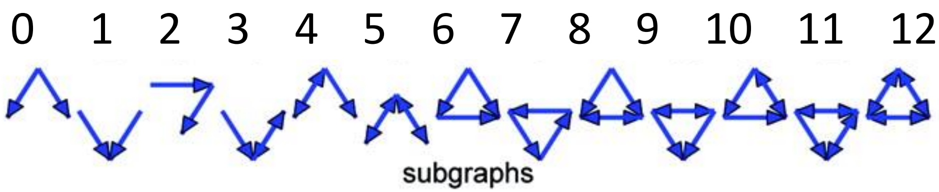
- Real social networks have a lot of triangles
  - Friends of friends are friends
- Any patterns?
  - 2X the friends, 2X the triangles?



How many different structured triads can we have?

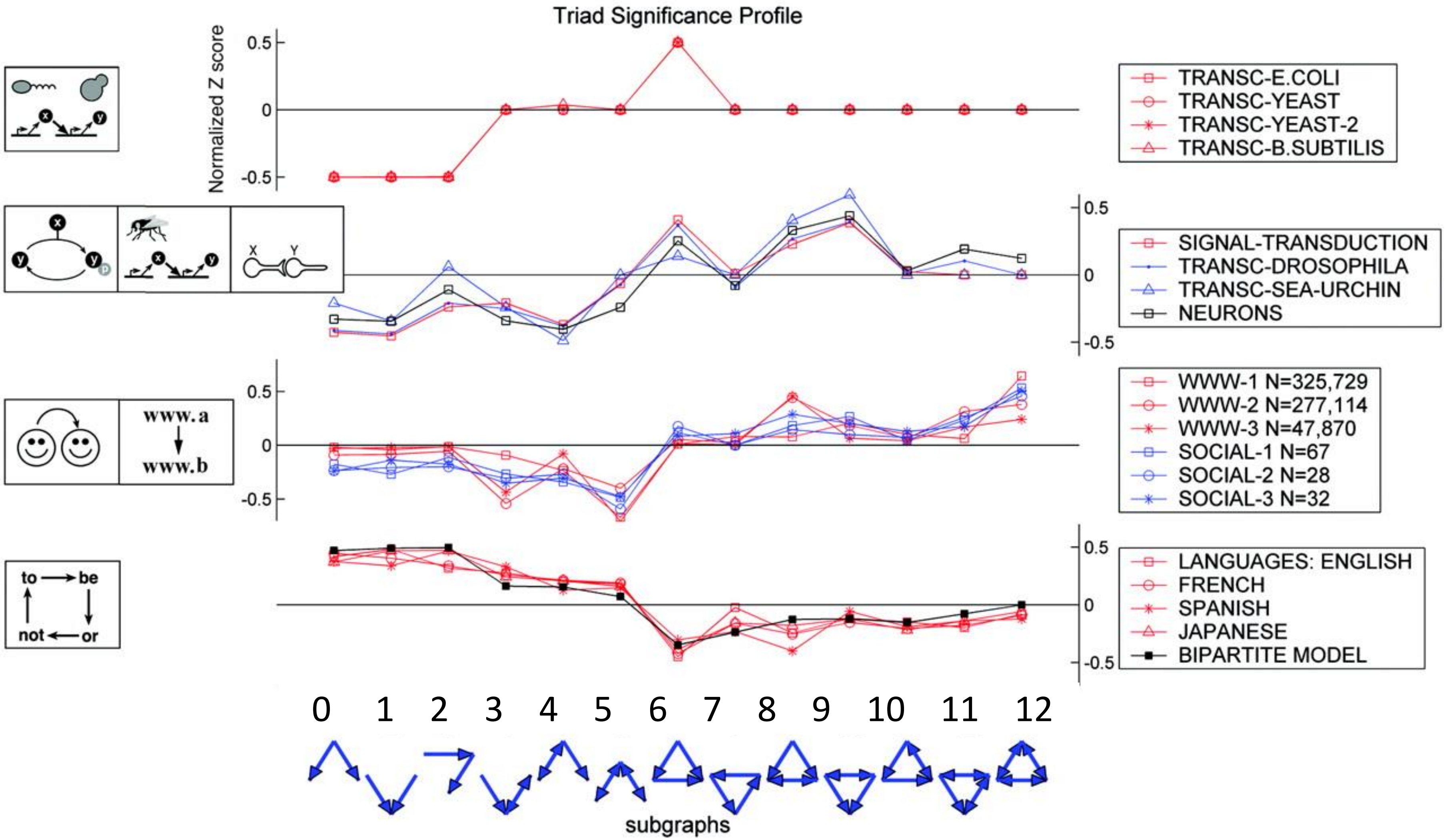
Christos Faloutsos's keynote speech on Apr.9

Triads in networks



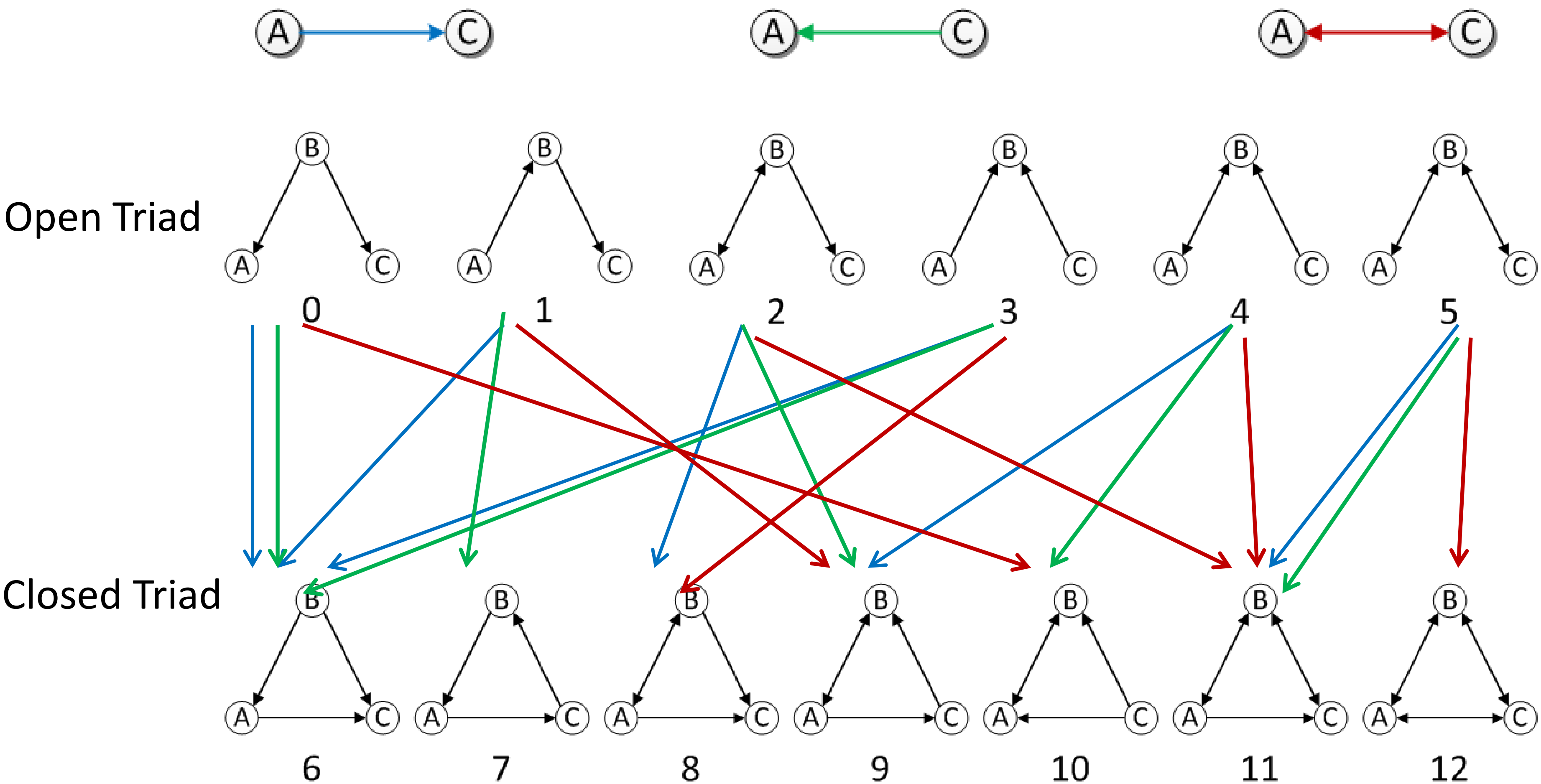
Milo R, Itzkovitz S, Kashtan N, et al.. Superfamilies of evolved and designed networks. Science, 2004

Triads in networks

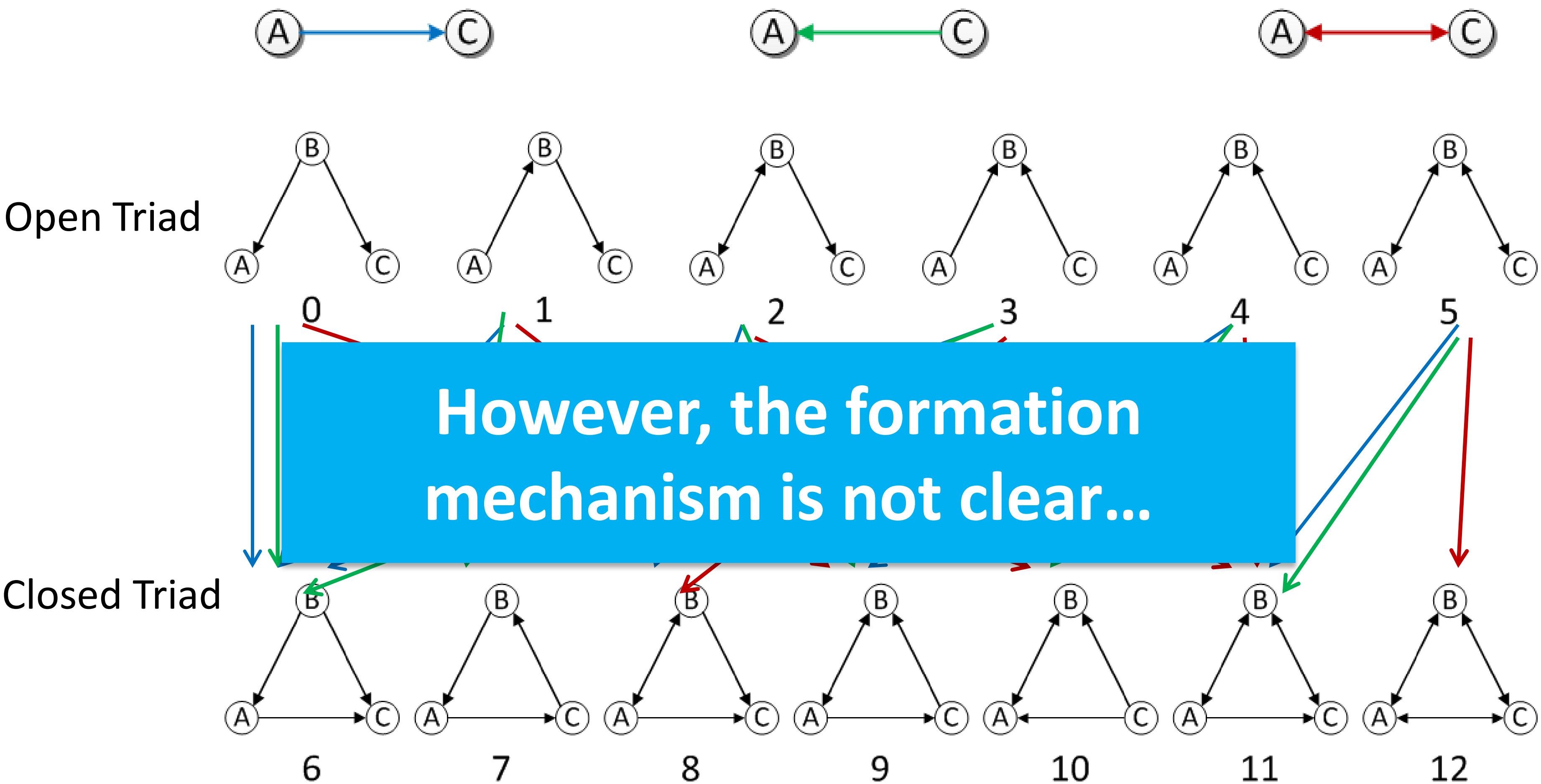


Milo R, Itzkovitz S, Kashtan N, et al.. Superfamilies of evolved and designed networks. Science, 2004

Open Triad to Triadic Closure

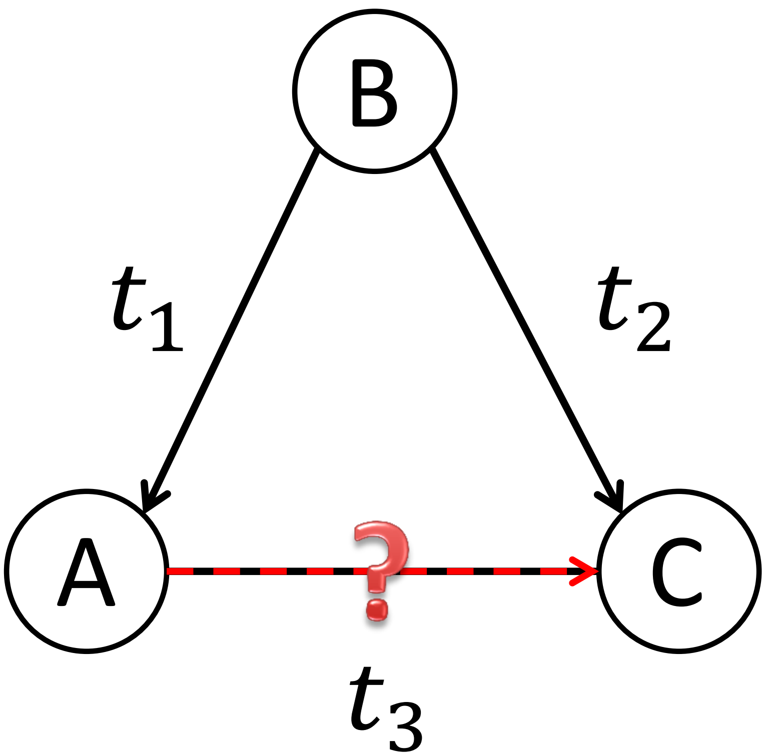


Open Triad to Triadic Closure



However, the formation mechanism is not clear…

Problem Formalization



$t_3 > t_2 > t_1$

- Given network $G^t = (V, E)$, Y^T are candidate open triad:
- Goal: Predict the formation of triadic closure

$$f: (\{G^t, Y^t, X^t\}_{t=1,\dots T}) \rightarrow Y^{T+1}$$

Dataset

44 thousand newly formed closed triads per day

Time span: Aug 29th, 2012 – Sep 29th, 2012

Weibo

新浪微博
weibo.com

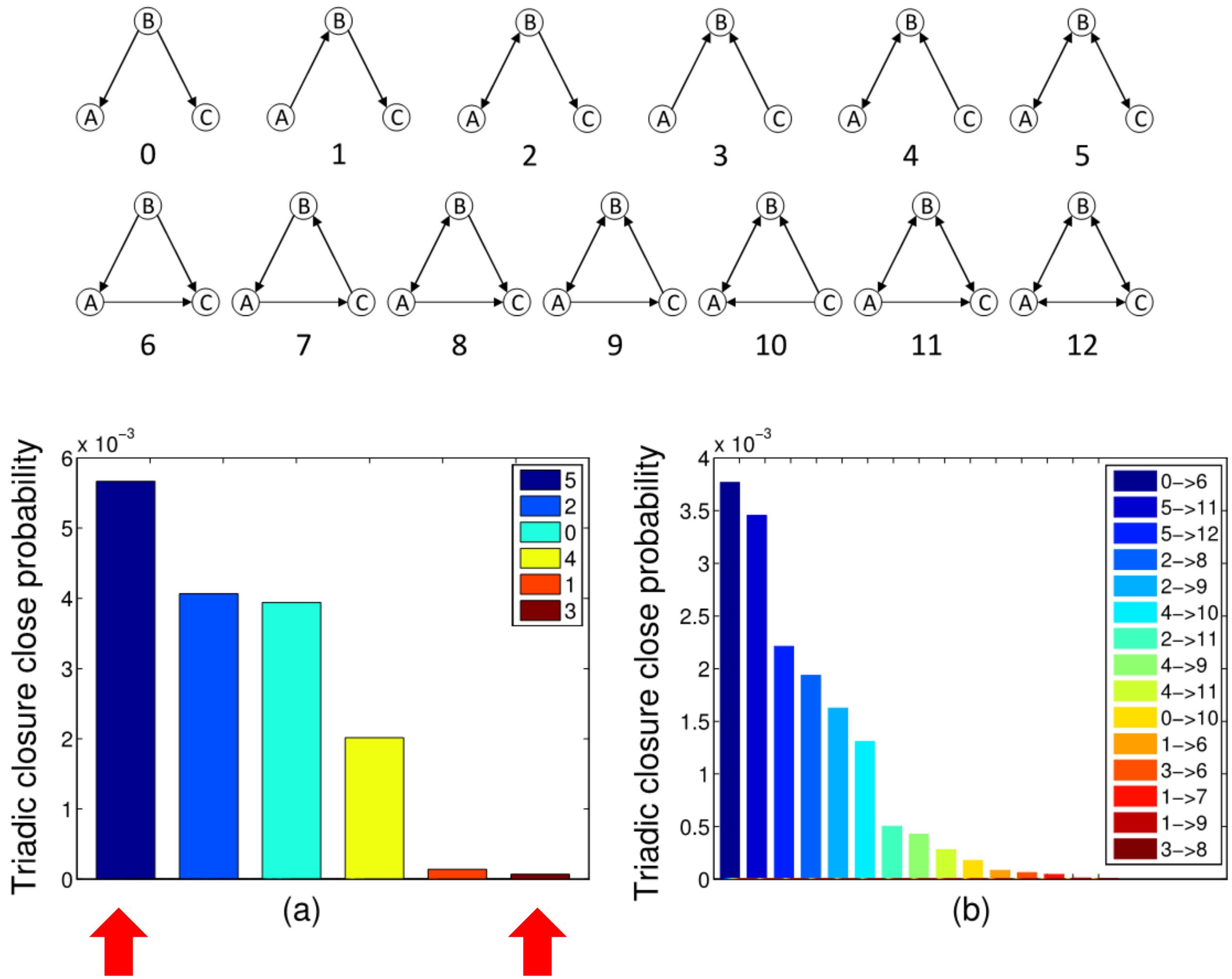
700 thousand nodes

360 thousand new links

200 out-degree per user

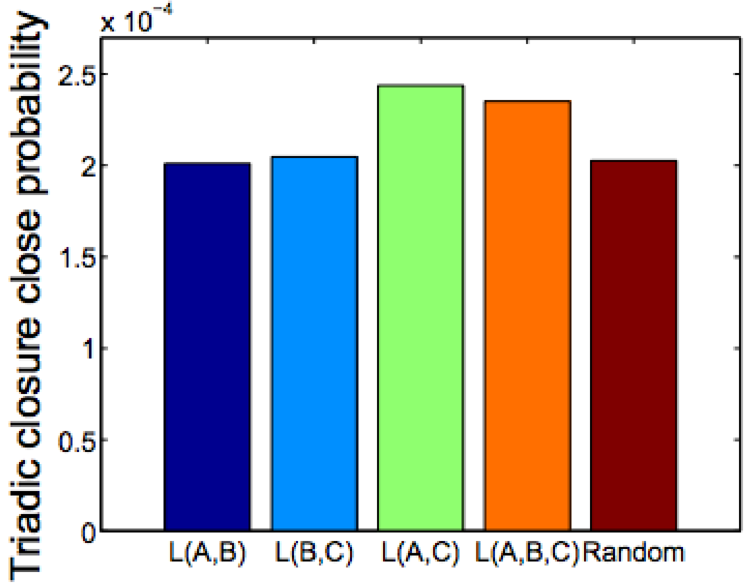
400 million following links

Observation - Network Topology

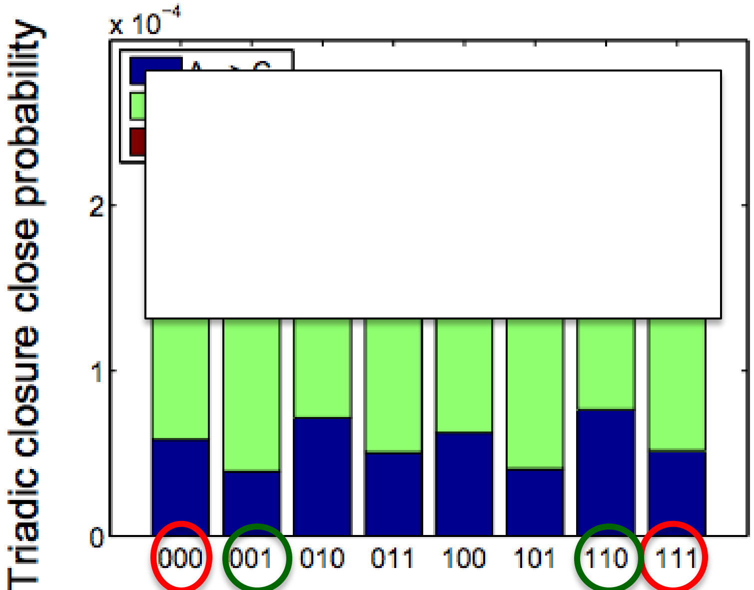


Y-axis: probability that each open triad forms triadic closures

Observation - Demography



(a) Location

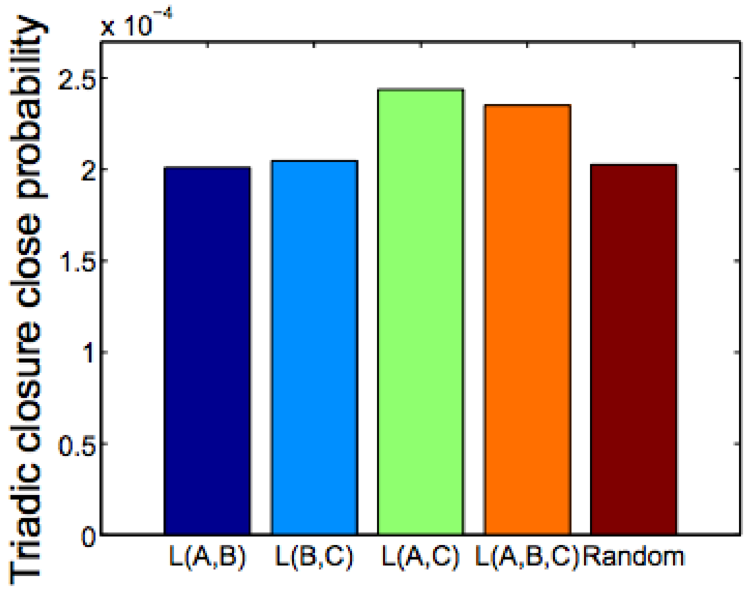


(b) Gender

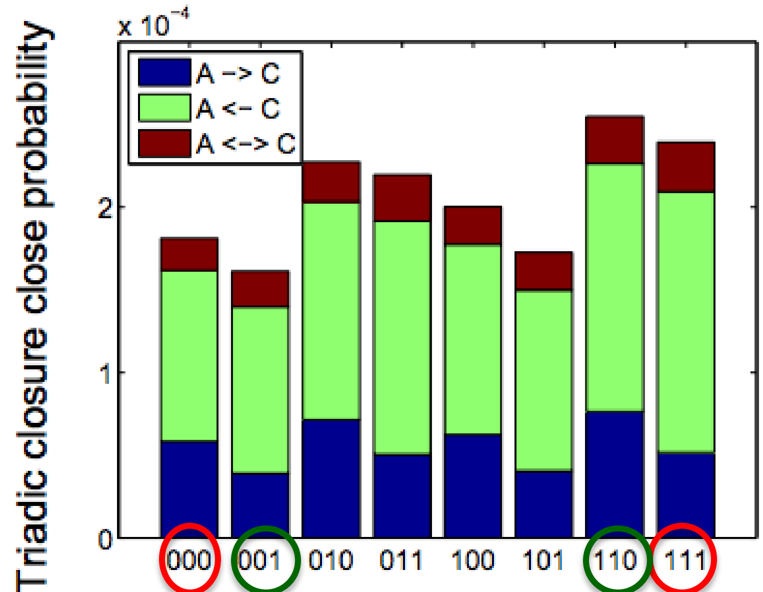
L(A, B) means A and B are from the same city

0—female; 1—male
e.g., 001 means A and B are female while C is male.

Observation - Demography



(a) Location

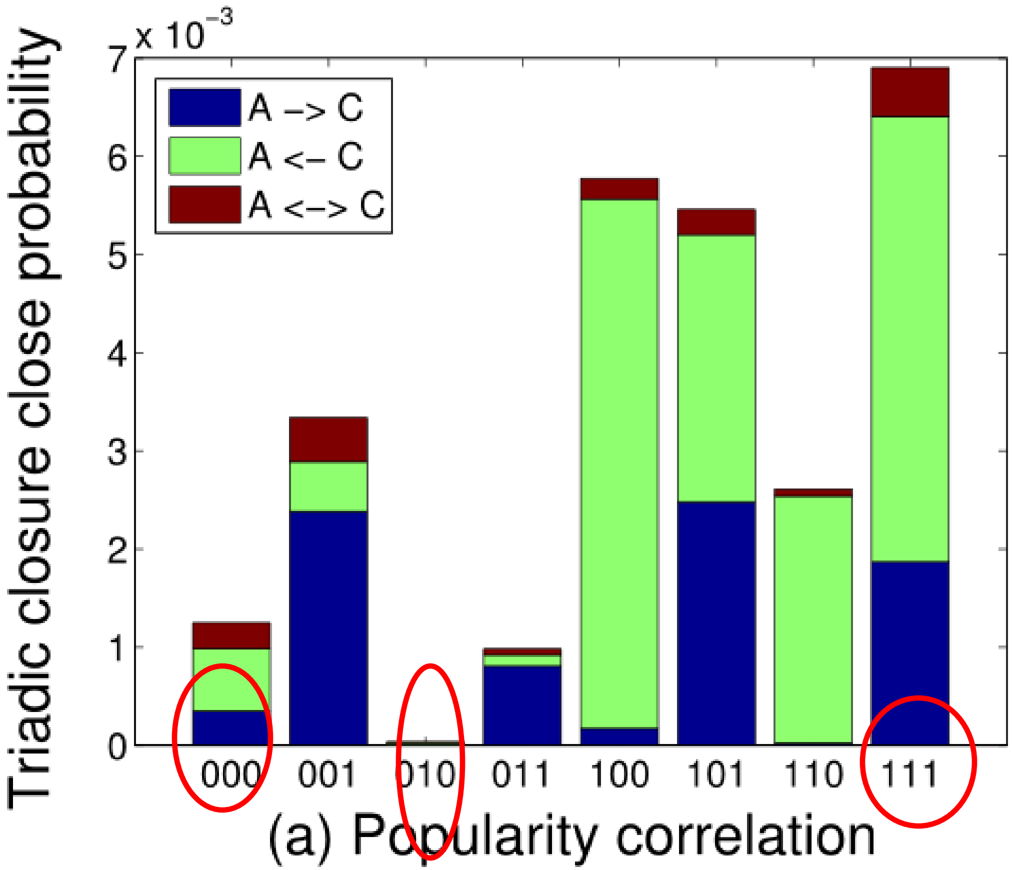


(b) Gender

L(A, B) means A and B are from the same city

0—female; 1—male
e.g., 001 means A and B are female while C is male.

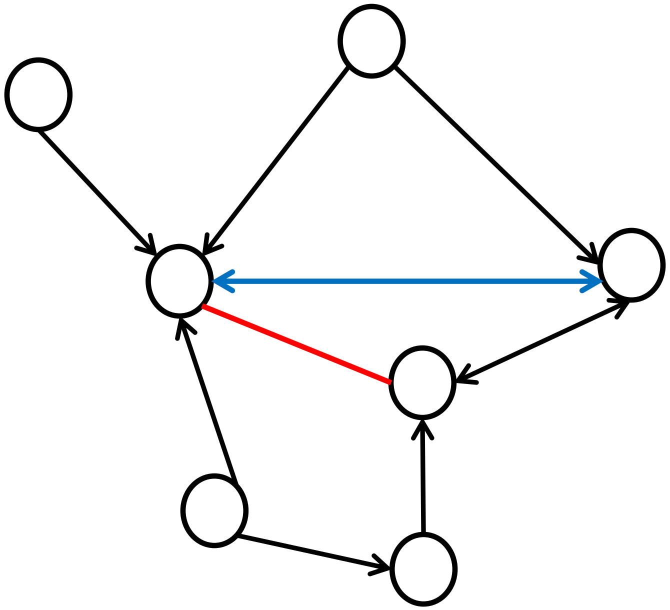
Observation - Social Role



0—ordinary user
1—opinion leader (top 1% PageRank)
e.g., 001 means A and B are ordinary user while C is opinion leader.

Summary

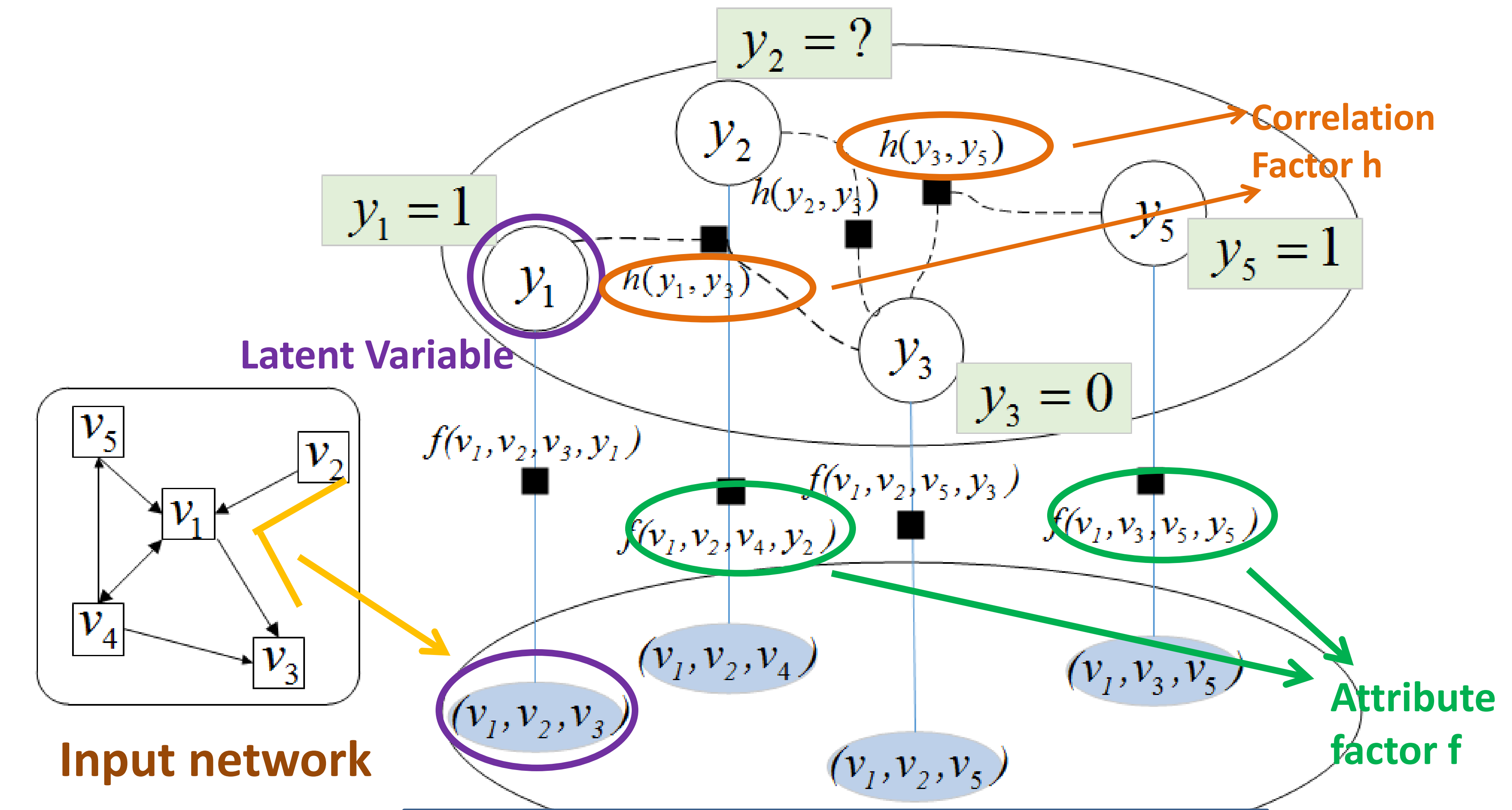
- Intuitions:
  - Men are more inclined to form triadic closure
  - Triads of opinion leaders themselves are more likely to be closed.
  - …
- Correlation



Considered the intuitions and correlations…

THE PROPOSED MODEL AND RESULTS

Triad Factor Graph (TriadFG) Model



Example:
Whether three users come from the same place?

Solution

- Given a network $G = \{V, E, X, Y\}$
- Objective function: $\varphi_\theta = logP_\theta(Y|X, G)$
- $P(Y|X, G) \propto P(X|Y) \cdot P(Y|G)$

attribute factor f

$$= \frac{1}{Z_1} \exp\{\sum_{i=1}^{|Tr|} \sum_{j=1}^{d} \alpha_j f_j(x_{ij}, y_i)\}$$

$$\cdot \frac{1}{Z_2} \exp\{\sum_{c} \sum_{k} \mu_k h_k(Y_{Tr_c})\}$$

- $\theta = (\{\alpha_j\}, \{\mu_k\})$

Correlation factor h

Learning Algorithm

Input: network G^t, learning rate η
Output: estimated parameters θ

Initialize $\theta \leftarrow 0$;
repeat

Perform LBP to calculate marginal distribution of unknown variables $P(y_i|x_i, G)$;
Perform LBP to calculate the marginal distribution of triad c, i.e., $P(y_c|\mathbf{X}_c, G)$;
Calculate the gradient of μ_k according to Eq. 7 (for α_j with a similar formula):

$$\frac{\mathcal{O}(\theta)}{\mu_k} = \mathbb{E}[h_k(Y_c)] - \mathbb{E}_{P_{\mu_k}(Y_c|\mathbf{X},G)}[h_k(Y_c)]$$

Update parameter θ with the learning rate η:

$$\theta_{\text{new}} = \theta_{\text{old}} + \eta \cdot \frac{\mathcal{O}(\theta)}{\theta}$$

until *Convergence*;

Lou T, Tang J, Hopcroft J, et al. Learning to predict reciprocity and triadic closure in social networks[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2013, 7(2): 5.

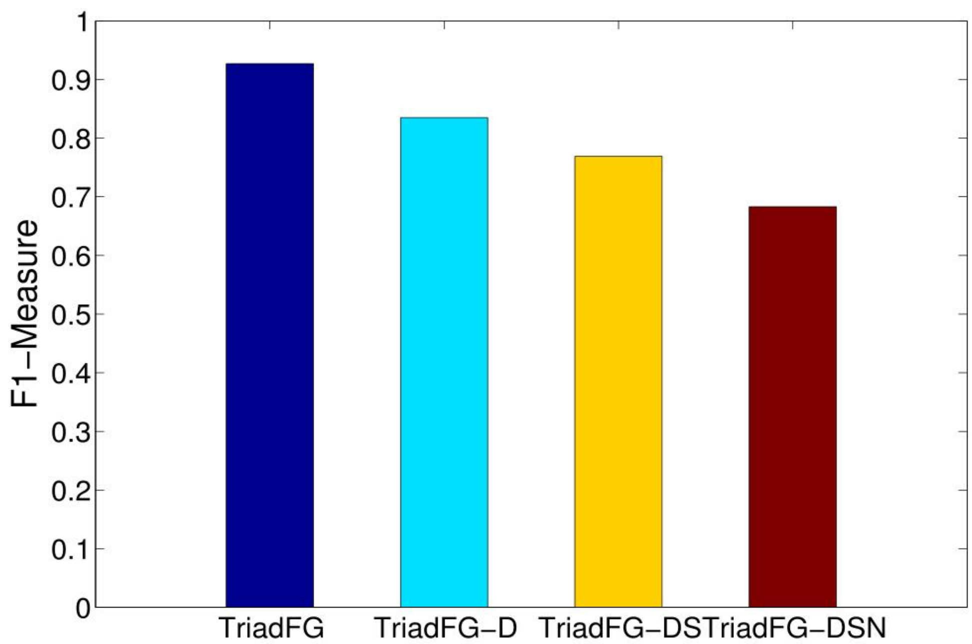
Results on the Weibo data

- Baselines: SVM, Logistic

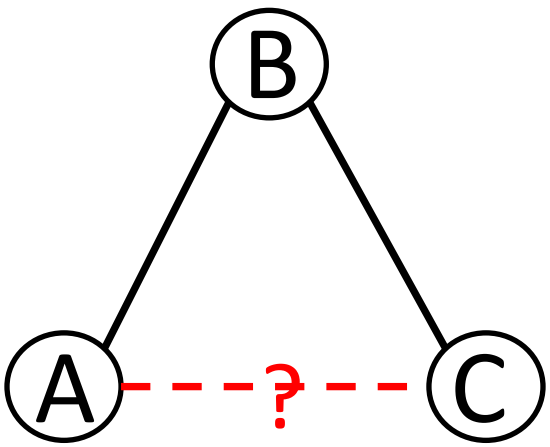
Algorithm	Precision	Recall	F1	Accuracy
SVM	0.890	0.844	0.866	0.882
Logistic	0.882	0.913	0.897	0.885
TriadFG	**0.901**	**0.953**	**0.926**	**0.931**

Factor Contribution Analysis

- Demography(D)
- Popularity(S)
- Network Topology(N)
- Structural hole (H)



Conclusion



- **Problem**: Triadic closure formation prediction
- **Observations**
  – Network Topology
  – Demography
  – Social Role
- **Solution**: TriadFG model
- **Future work**

Thanks
Jing Zhang in Tsinghua Uni.
for sharing her Weibo data!

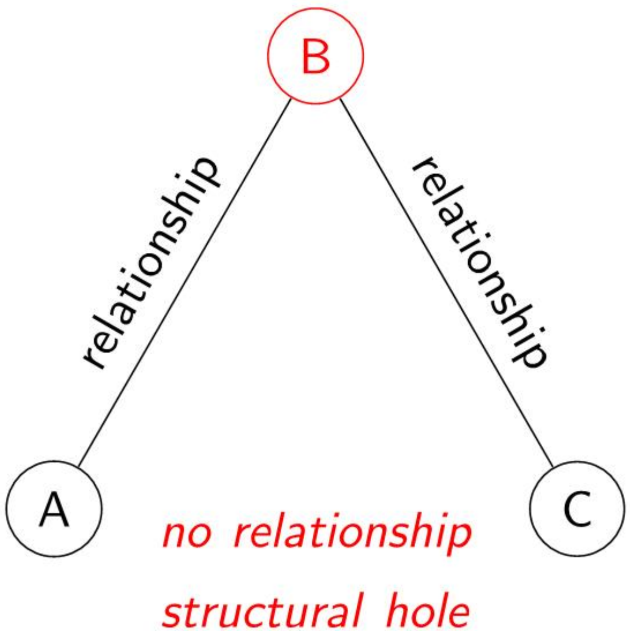
THANK YOU!

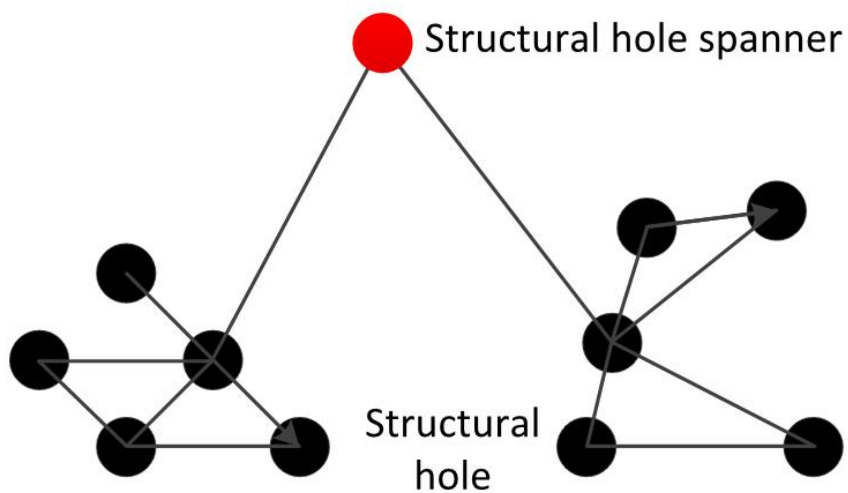
Attribute factor Definition

	Feature	Function
Network topology	Is open triad 5/2/0/4/1/3	1/1/1/1/0/0
Demography	A,B,C from the same place	1
	A,C from the same place	1
	C is female	1
	B is female	1
Social role	A/B/C is popular user	1/0/1
	A,B,C are all popular user	1
	Two users are popular	1
	One user is popular	1
	A/B/C is structural hole spanner	1/0/1
	Two users are structural hole spanner	1
	One user is structural hole spanner	1

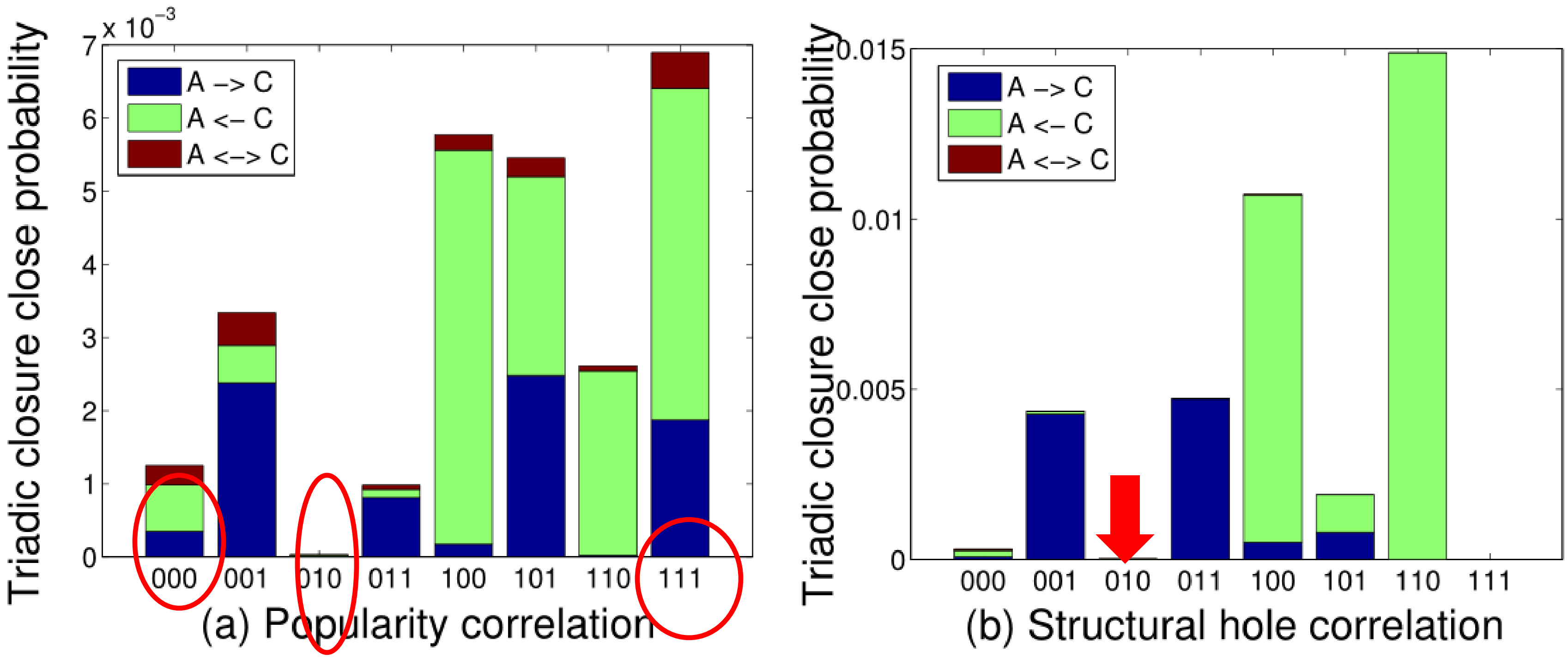
Structural hole

- When two separate clusters possess non-redundant information, there is said to be a structural hole between them





Observation - Social Role



0—ordinary user; 1—opinion leader
e.g., 001 means A and B are ordinary user while C is opinion leader.

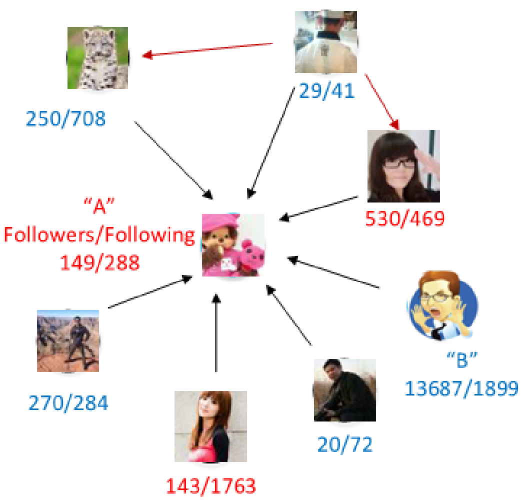
0—ordinary user; 1—structural hole spanner
e.g., 001 means A and B are ordinary user while C is structural hole spanner .

Lou T, Tang J. Mining structural hole spanners through information diffusion in social networks, www2013

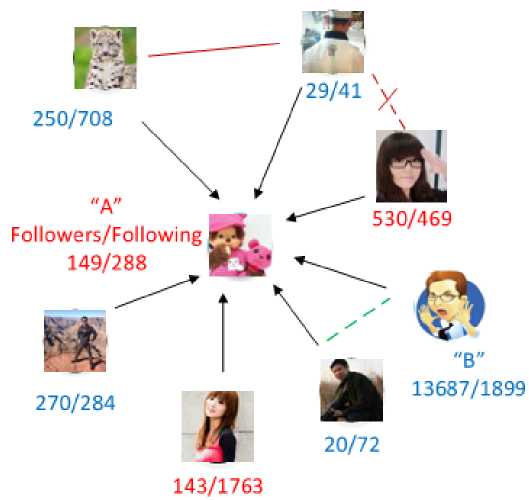
Popular users in Weibo vs. Twitter

- The rich get richer (Both)
  - $P(1XX) > P(0XX)$, validates preferential attachment
- In twitter, popular users functions in triadic closure formation, while in Weibo reverse
  - In Twitter, $P(X1X) > P(X0X)$
  - In Weibo, ordinary users have more chances to connect other users.
- Popular users in China are more close

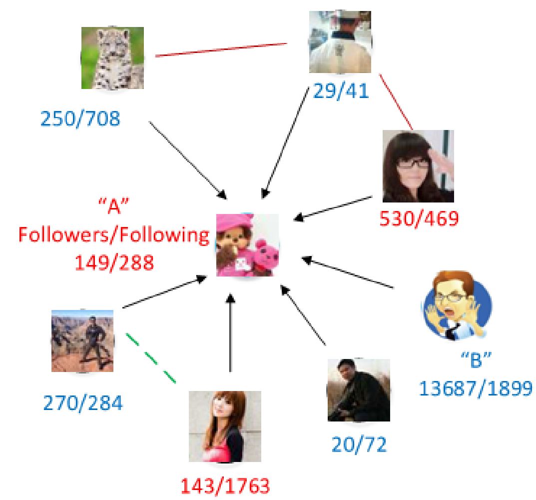
Qualitative Case Study



(e) Ground Truth



(f) SVM



(g) Our approach