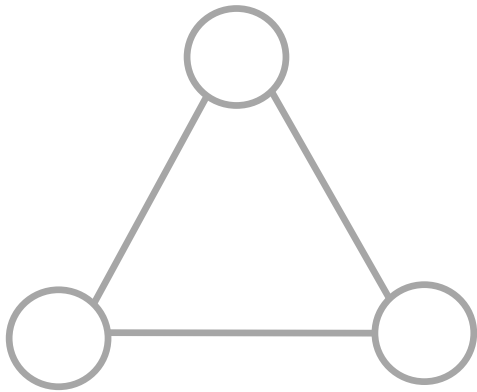


Mining Triadic Closure Patterns in Social Networks



Xiaoming Fu, University of Göttingen

Joint work with:

Hong Huang, University of Göttingen

Jie Tang, JarDer Luo, Tsinghua University

Sen Wu, Stanford University

Lu Liu, Northwestern University

Networked World

facebook

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

twitter



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

amazon.com

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



Alibaba Group
阿里巴巴集团

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



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

新浪微博
weibo.com

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



- **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

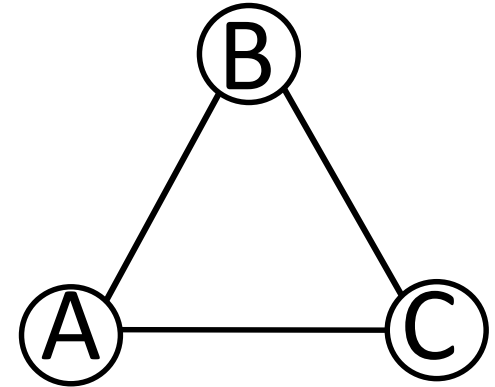
<http://techcrunch.com/2010/08/07/why-online2offline-commerce-is-a-trillion-dollar-opportunity/>

Simmelian tie

- Three (a triad) or more of reciprocal strong ties in a group
- Even stronger than a regular strong tie

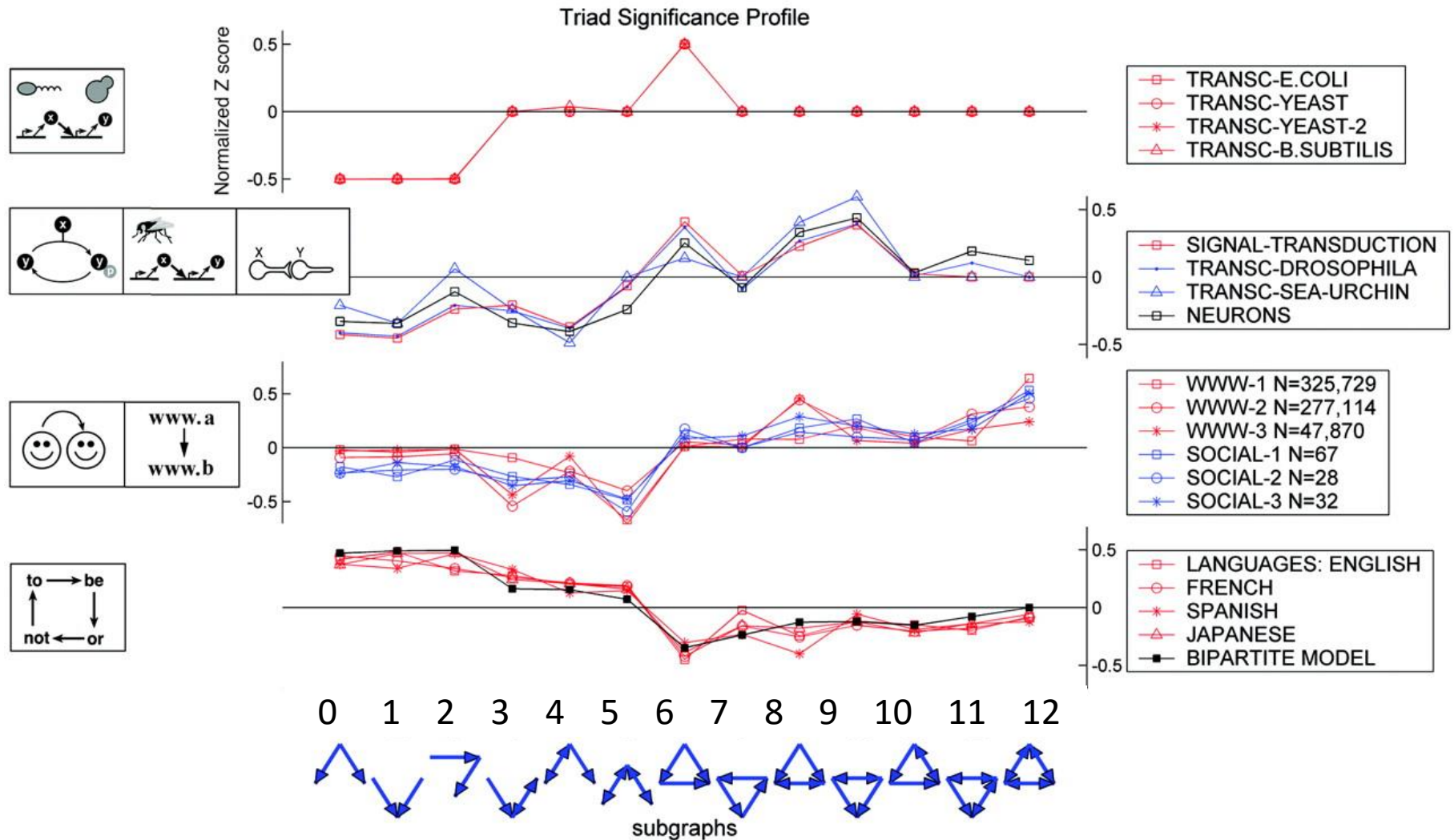
“Triangle Laws”

- Real social networks have a lot of triangles
 - Friends of friends are friends[Wasserman Faust '94]
- Any patterns?
 - 2X the friends, 2X the triangles?



How many different structured triads can we have?

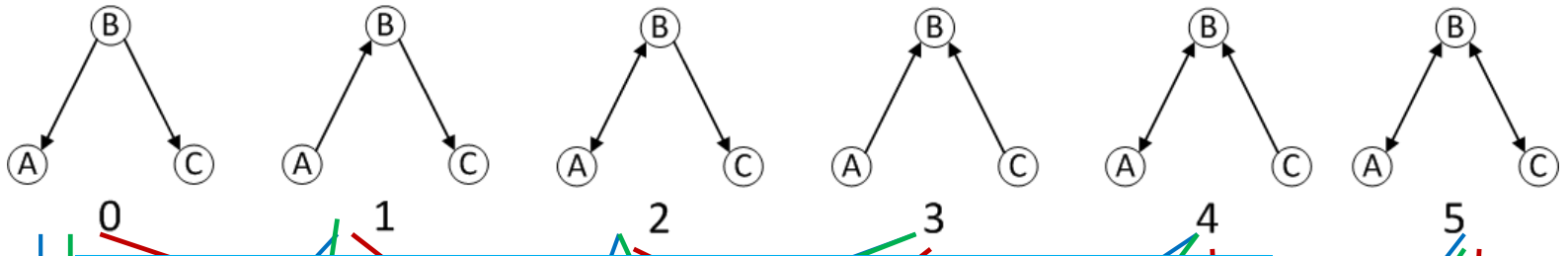
Triads in networks



Open Triad to Triadic Closure

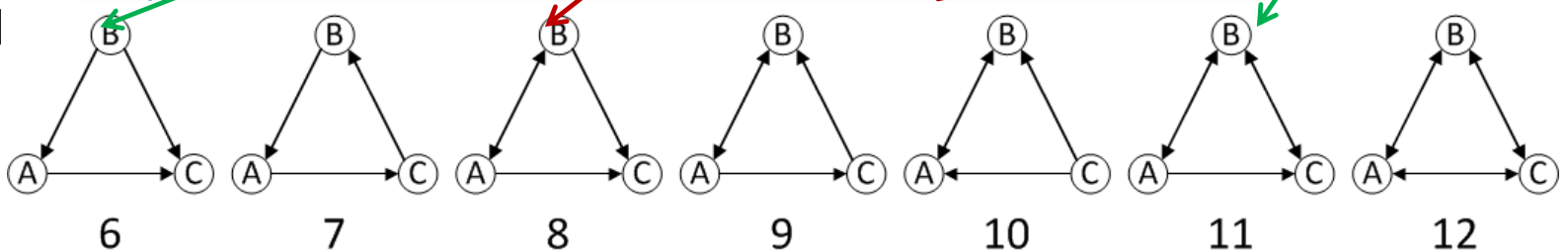


Open Triad

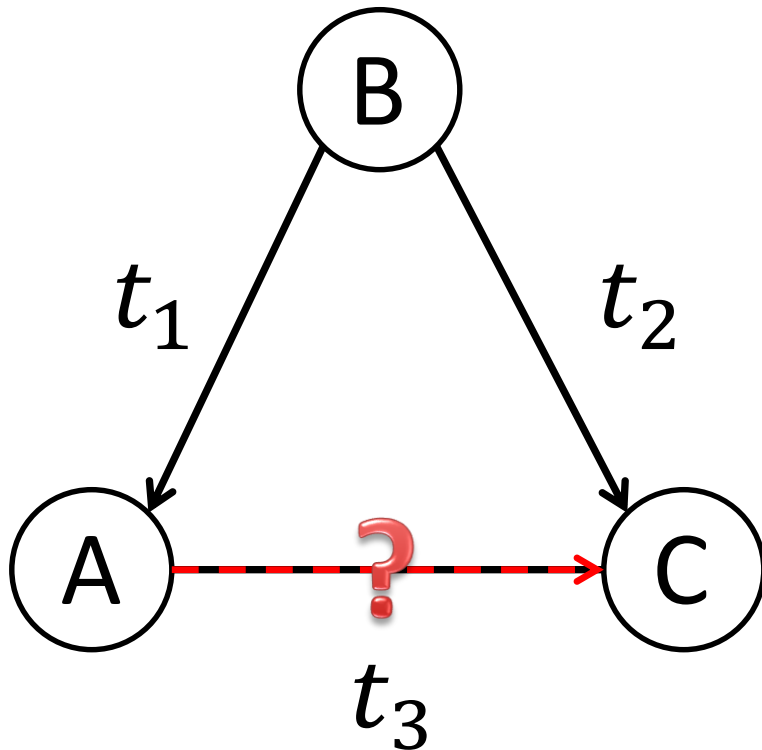


However, the formation mechanism is not clear...

Closed Triad



Problem Formalization



$$t_3 > t_2 > t_1$$

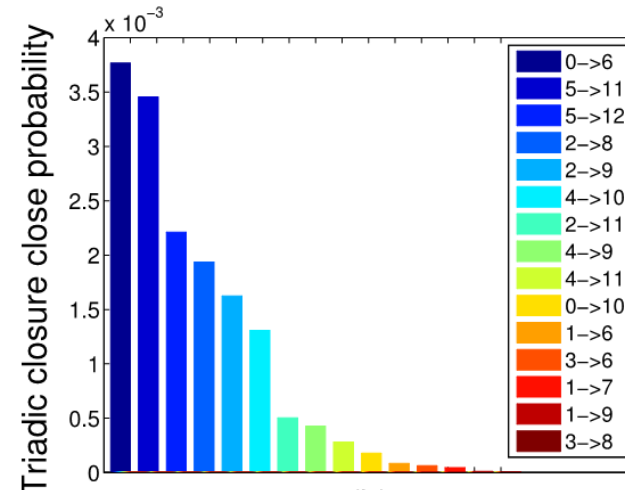
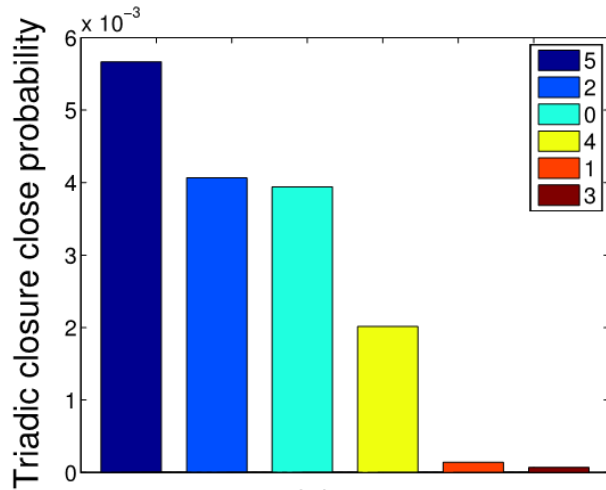
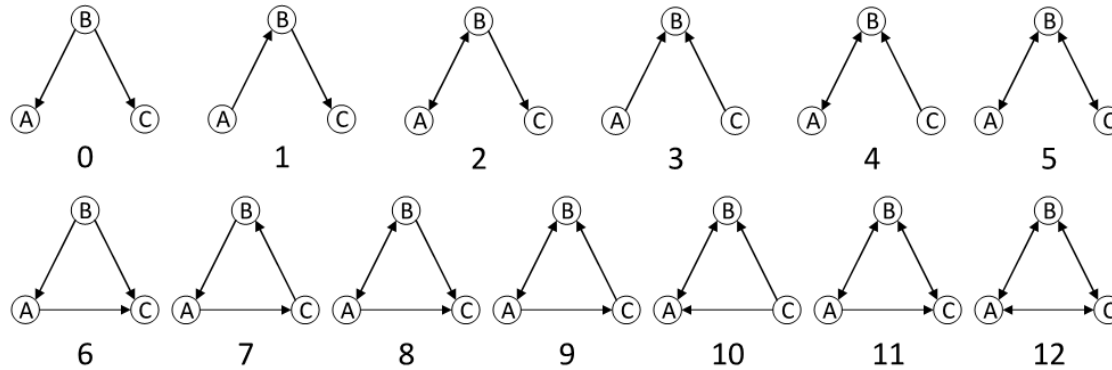
- Given network $G^t = (V, E)$, Y^T are candidate open triad:
- Goal: Predict the formation of triadic closure

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

Dataset

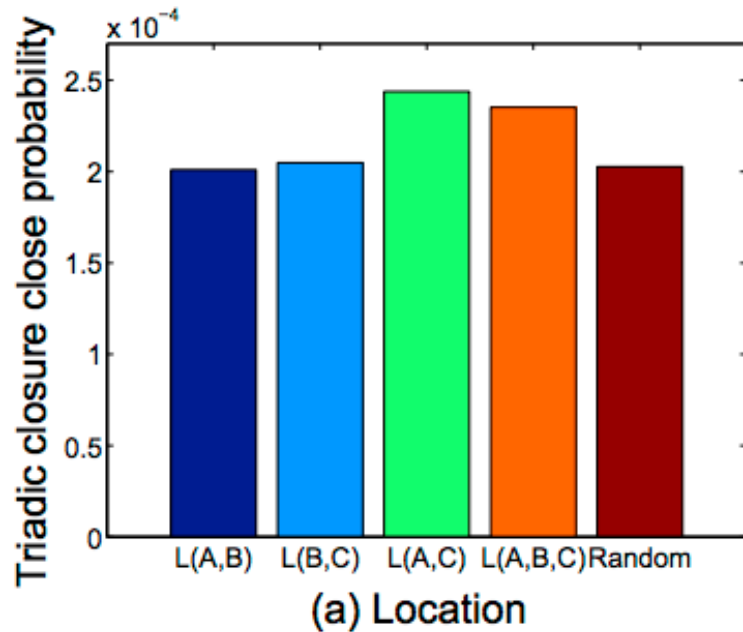


Observation - Network Topology

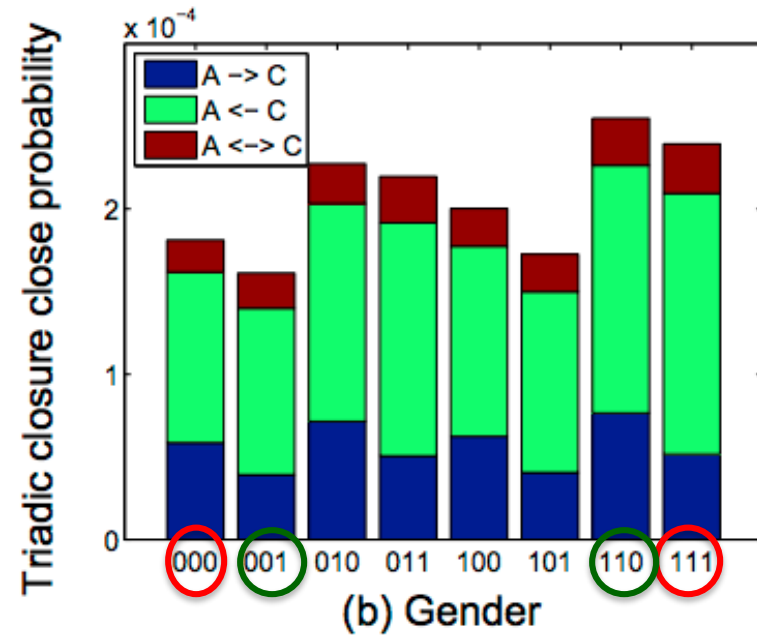


Y-axis: probability that each open triad forms triadic closures

Observation - Demography



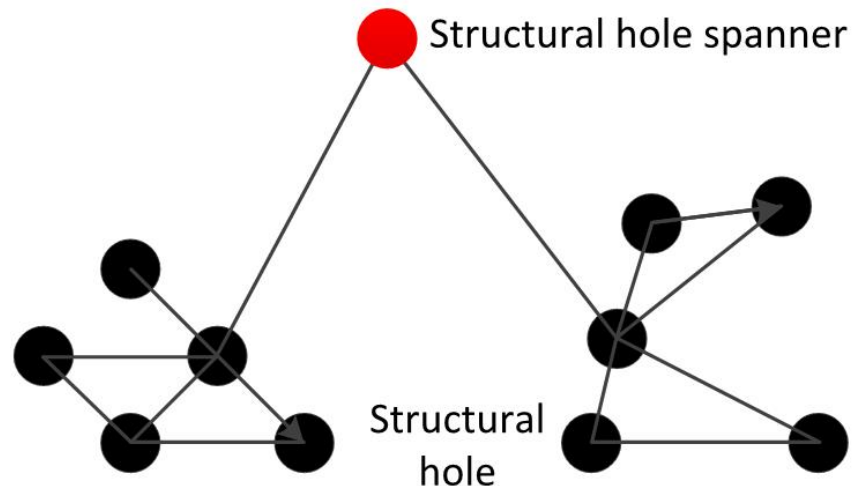
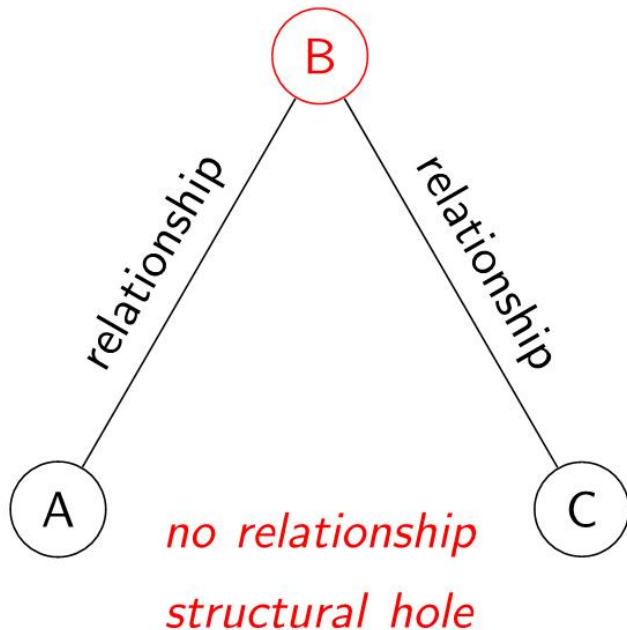
L(A, B) means A and B are from the same city



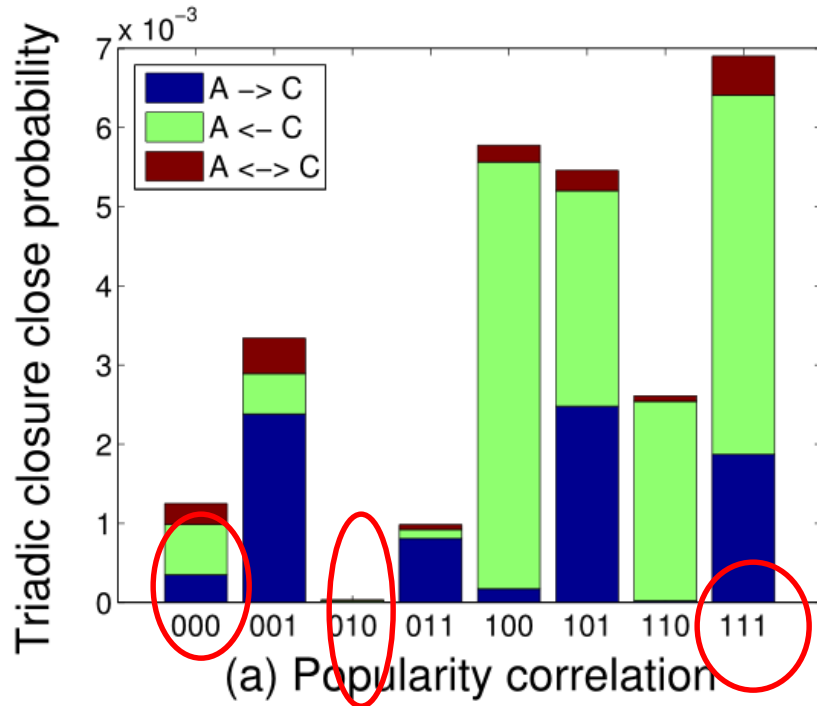
0—female; 1—male
e.g., 001 means A and B are female while C is male.

Structural hole

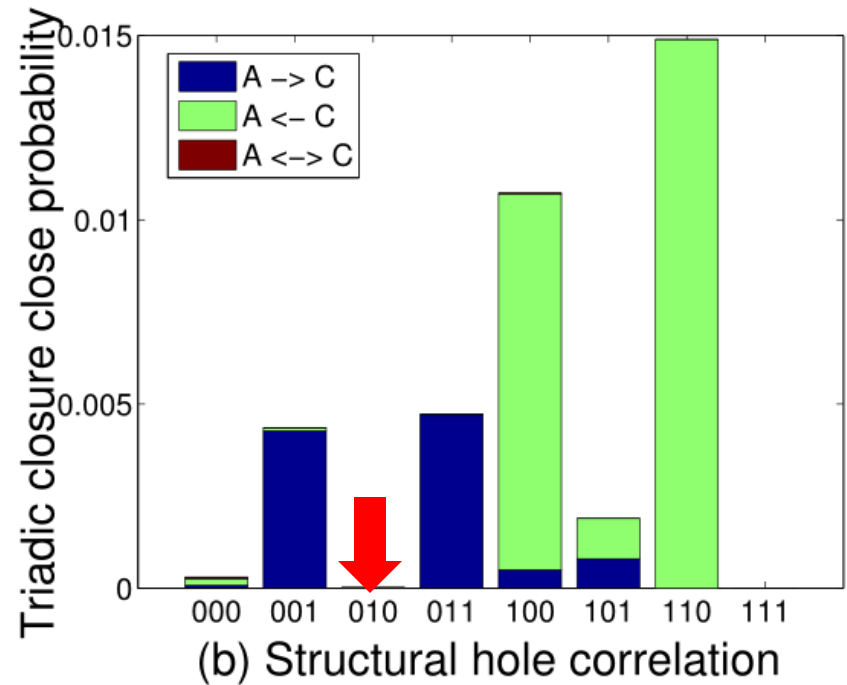
- When two separate clusters possess non-redundant information, there is said to be a structural hole between them



Observation - Social Role



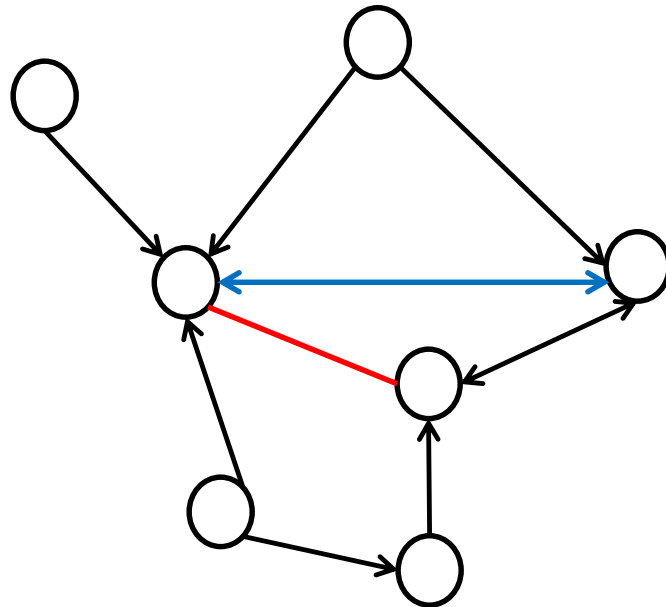
0—ordinary user;
 1—opinion leader
 e.g., 001 means A and B are ordinary user while C is opinion leader.



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

Summary

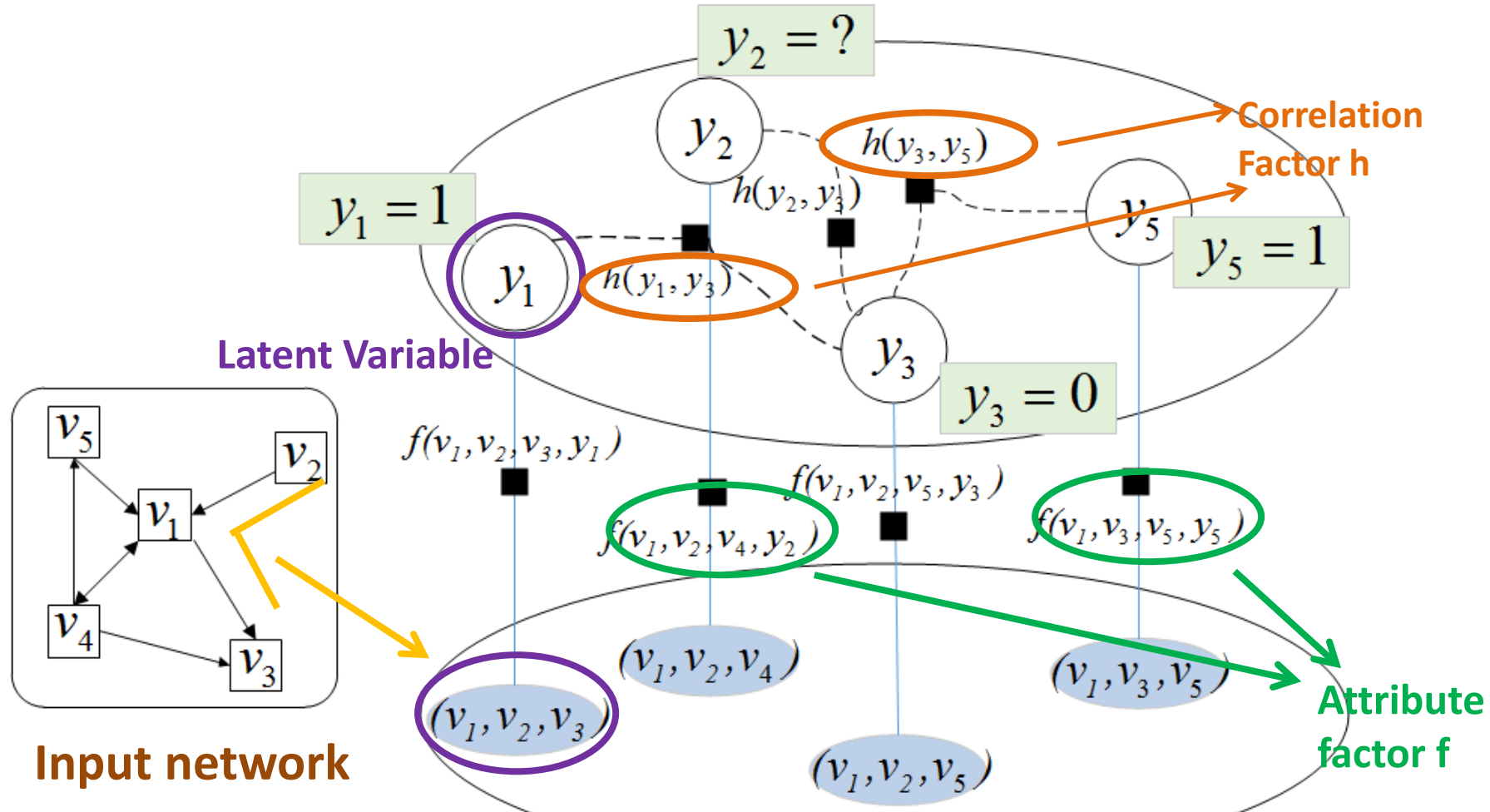
- Intuitions:
 - Men are more inclined to form triadic closure
 - Triads of opinion leaders themselves are more likely to be closed.
 - ...
- Correlation



Considered the intuitions and correlations...

THE PROPOSED MODEL AND RESULTS

Triad Factor Graph (TriadFG) Model



Map candidate open t

Example:

Whether three users come from the same place?

Solution

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

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

$$\cdot \frac{1}{Z_2} \exp\left\{ \sum_c \sum_k \mu_k h_k(Y_{Tr_c}) \right\}$$

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

Correlation
factor h

Learning Algorithm

Input: network G^r , 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*;

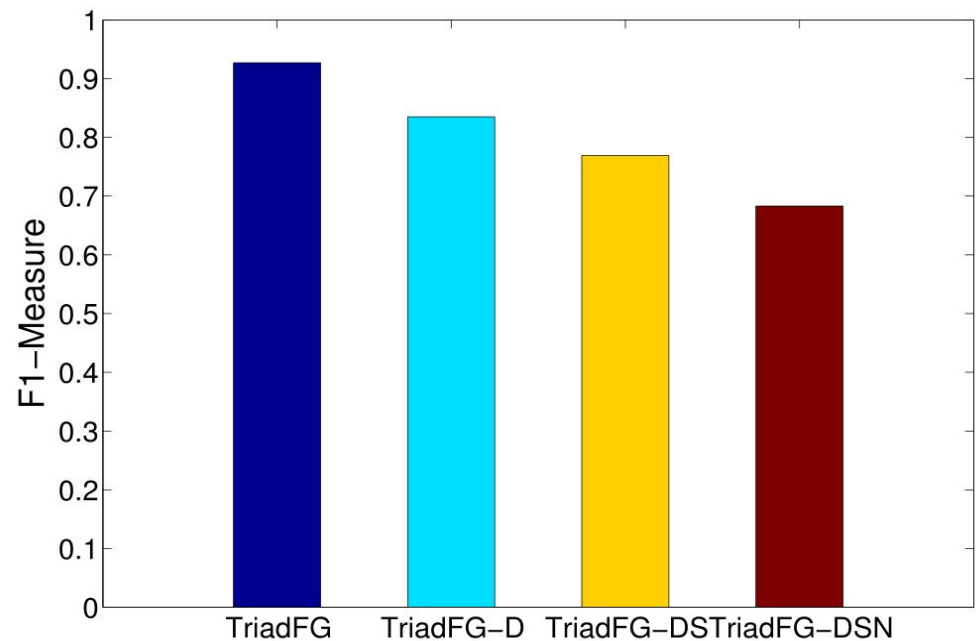
Results on the Weibo data

- Baselines: SVM, Logistic

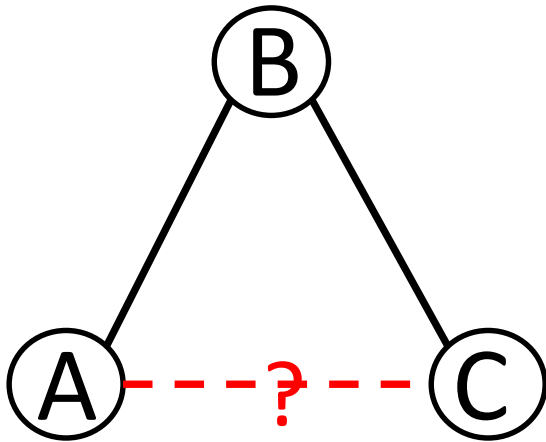
Algorithm	Precision	Recall	F1	Accuracy
SVM	0.890	0.844	0.866	0.882
Logistic	0.882	0.913	0.897	0.885
TriadFG	0.901	0.953	0.926	0.931

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**