Paper ID : 091

KDD2015

Probabilistic Community and Role Model for Social Networks

Yu Han[1] and Jie Tang[1,2,3]

[1]Department of Computer Science and Technology, Tsinghua University

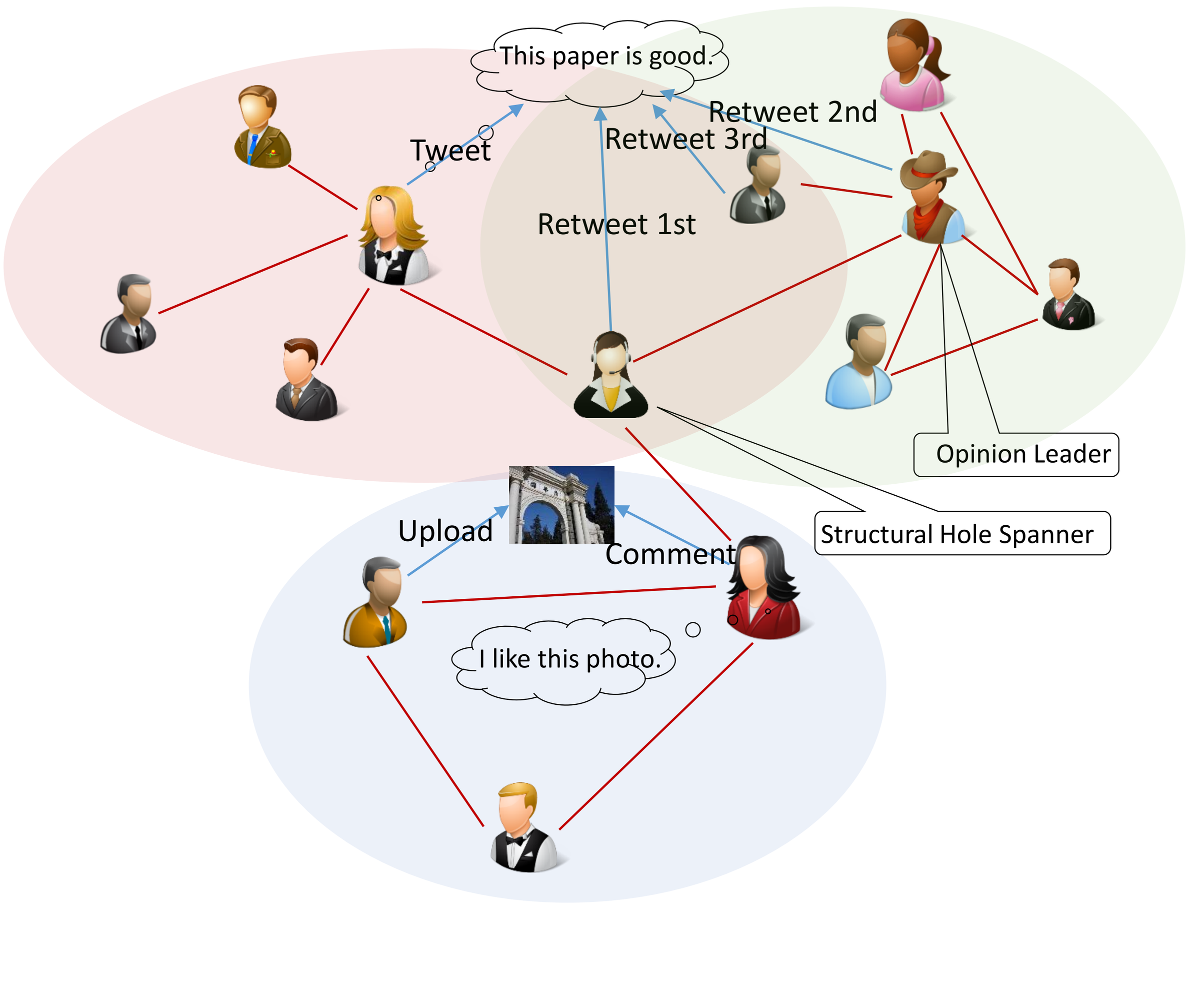
[2]Tsinghua National Laboratory for Information Science and Technology (TNList)

[3]Jiangsu Collaborative Innovation Center for Language Ability, Jiangsu Normal University, China

yuhanthu@126.com, jietang@tsinghua.edu.cn

Motivation

Example



- ☺ There are visible(users, links, actions) and invisible(communities, roles) elements in social networks.
- ☺ Visible and invisible elements interact and affect each other.

Challenges

- How should we model a complex social network so that the model can capture the intrinsic relations between all these elements, such as conformity influence, individual attributes, and actions?
- How do we use a social network model to handle issues such as community detection and behavior prediction without changing model itself?

Intuitions

Links:

- Locally inhomogeneous.
- Each node may belong to several communities.

Attributions:

- Each node has many attributes.
- Based on these attributes, we can classify the nodes into clusters.
- Each cluster can be regarded as a role that nodes play.

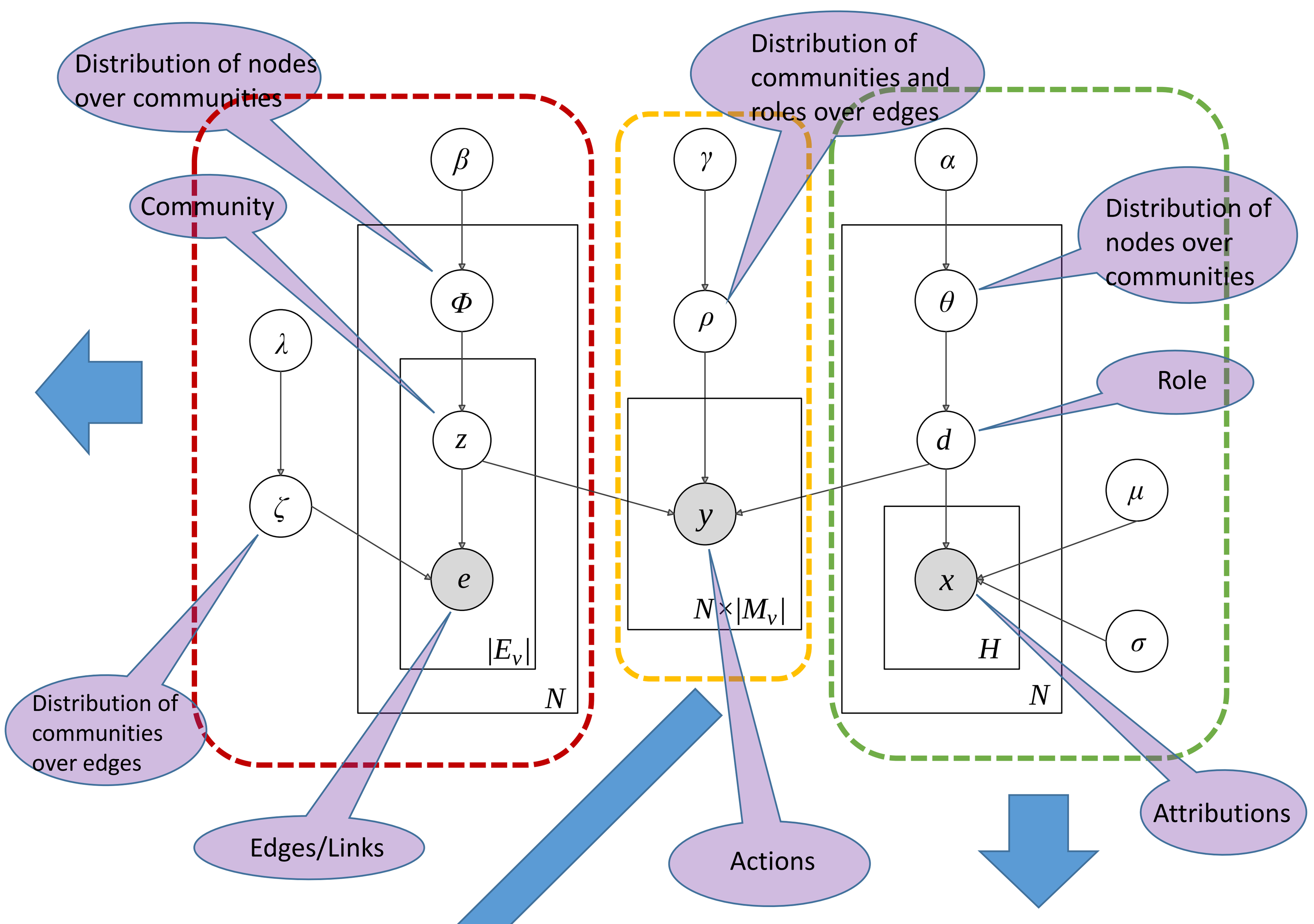
Actions:

- Whether a node takes a specific action partly depends on the community it belongs to.
- Whether a node takes an action may also depend on the role it plays.

Assumptions

1. Each node has a distribution over the communities.
2. Each community has a distribution over the links.
3. The attributes of each role satisfy a specific distribution—such as a Gaussian distribution.
4. Each node has a distribution over roles according to its attributes.
5. Community and role have a distribution over actions.

Our approach : CRM model



The Process to Generate Edges

For each node v in the graph:

1. Draw ζ from $Dirichlet(\lambda)$;
2. Draw a φ_v from $Dirichlet(\beta)$ prior;
3. For each edge $e_{v,i}$:
   - Draw a community $z_{v,i} = c$ from multinomial distribution φ_v ;
   - Draw an edge $e_{v,i}$ from a multinomial $\zeta^{(c)}$ specific to community c.

Inference and Parameter Estimation

$$p(z_{v,i} = c|z_{-v,-i}, E) \propto \frac{n^{(v)}_{-v,-i,c} + \beta}{|E_v| + |C| \cdot \beta} \frac{n^{(e)}_{-v,-i,c} + \lambda}{n^{(e)}_{-v,-i,\cdot} + |E| \cdot \lambda}$$

$$\varphi_{v,c} = \frac{n_{v,c} + \beta}{|E_v| + |C| \cdot \beta} \qquad \zeta_{c,e} = \frac{n_{c,e} + \lambda}{n_c + |E| \cdot \lambda}$$

The Process to Generate Attributes

For each node v in the graph:

1. Draw a θ_v from $Dirichlet(\alpha)$ prior;
2. Draw a role $d_v = r$ from multinomial distribution θ_v;
3. For each attribute of v, draw a value $x_h^{(r)} \sim N(\mu_{r,h}, \sigma_{r,h}^2)$.

Inference and Parameter Estimation

E-step:

$$\theta_{v,r} = \frac{\prod_h (2\pi)^{-\frac{1}{2}} \sigma_{r,h}^{-1} e^{\frac{(x_{v,h} - \mu_{r,h})^2}{2\sigma_{r,h}^2}}}{\sum_{d_v} \prod_h (2\pi)^{-\frac{1}{2}} \sigma_{r,h}^{-1} e^{\frac{(x_{v,h} - \mu_{r,h})^2}{2\sigma_{r,h}^2}}}$$

M-step:

$$\mu_{r,h} = \frac{\sum_v \theta_{v,r} x_{v,h}}{\sum_v \theta_{v,r}} \qquad \sigma_{r,h} = \sqrt{\frac{\sum_v \theta_{v,r} (x_{v,h} - \mu_{r,h})^2}{\sum_v \theta_{v,r}}}$$

The Process to Generate Actions

For each action y_m:

1. Draw ρ from $Dirichlet(\gamma)$ prior;
2. Draw a community c_v for v from φ_v;
3. Draw a community c_u for u, which is the target of the action, from φ_u;
4. Draw a role r from θ_v;
5. Draw $y_m \sim Multinomial(\rho^{\tau,r})$.

Inference and Parameter Estimation

$$p(a_v = \tau, d_v = r|a_{-v}, r_{-v}, y) \propto (\varphi_v \varphi_v^T)\theta_v \frac{n_{-v,-m,\tau,r} + \gamma}{|M| + 2|H|\gamma}$$

$$\rho = \frac{n_{v,m,\tau,r} + \gamma}{|M| + 2|H|\gamma}$$

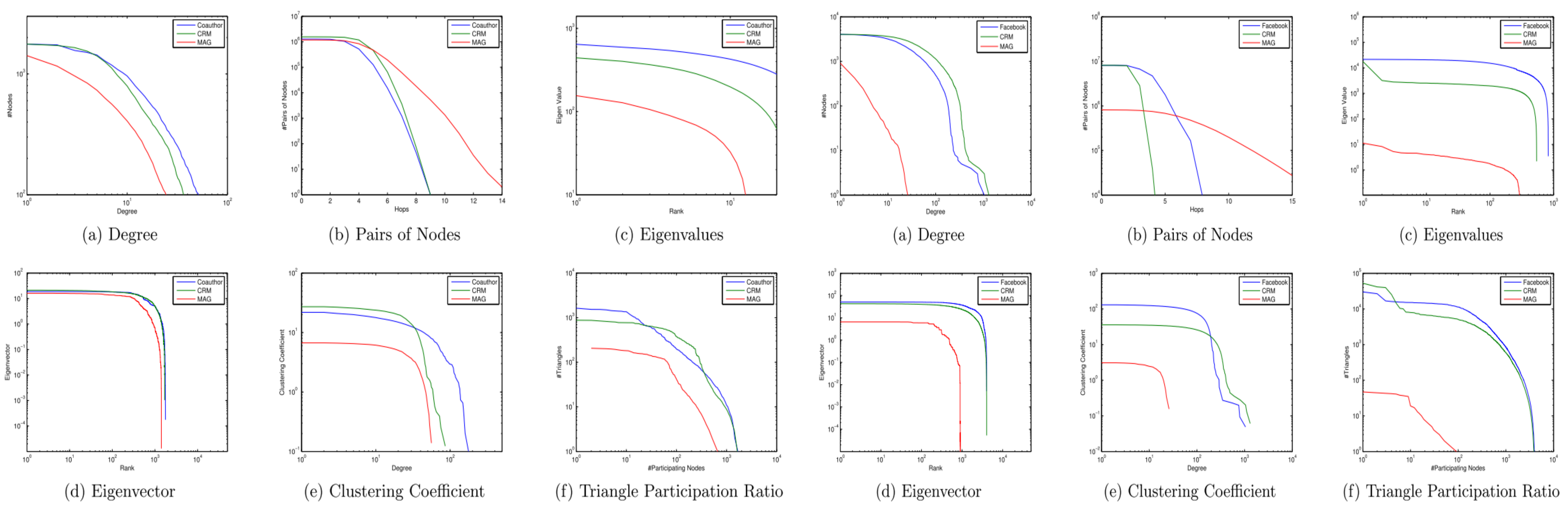
Experiments

We first use a real dataset to learn the parameters of CRM. Then we use the parameters to generate a synthetic social network. Then we evaluate CRM by three tasks:

- Structural Recovery. Baseline: MAG(UAI'11)
- Behavior Pediction. By parameter ρ.
- Community Detection. By parameter ζ.

The datasets we used include Coauthor(1,765 nodes, 13,415 links), Facebook(4,039 nodes, 88,234 links), Weibo(1,776,950 nodes, 308,489,739 links).

Structural Recovery



Coauthor

Facebook

Behavior Prediction

Date set	Method	Precision	Recall	F1-measure	AUC
Coauthor	SVM	**0.8838(0.1725)**	0.5562(0.3183)	0.6827(0.2054)	0.7360(0.1111)
	SMO	0.8647(0.1218)	0.8142(0.1260)	0.8387(0.1138)	0.9218(0.0366)
	LR	0.8668(0.1242)	0.8292(0.1022)	0.8476(0.1016)	0.9642(0.0196)
	NB	0.8183(0.1830)	0.8115(0.1444)	0.8149(0.1549)	0.9417(0.0335)
	RBF	0.8552(0.1058)	0.8353(0.1165)	0.8451(0.1081)	0.9477(0.0271)
	C4.5	0.8328(0.0518)	0.8015(0.1286)	0.8169(0.1478)	0.9065(0.1165)
	CRM	0.8562(0.1490)	**0.8630(0.0598)**	**0.8596(0.1013)**	**0.9800(0.0199)**
Weibo	SVM	0.5067(0.1405)	0.5027(0.1185)	0.5047(0.1150)	0.6068(0.1113)
	SMO	0.5074(0.1464)	0.5209(0.1099)	0.5141(0.1271)	0.6145(0.0363)
	LR	0.5199(0.1306)	0.5469(0.1073)	0.5331(0.1157)	0.6330(0.0377)
	NB	0.5112(0.1245)	0.5692(0.1083)	0.5386(0.1172)	0.6397(0.0394)
	RBF	0.5225(0.1361)	0.4679(0.1117)	0.4937(0.1217)	0.5945(0.0085)
	C4.5	0.5237(0.1367)	0.5322(0.1114)	0.5279(0.1211)	0.6271(0.1083)
	CRM	**0.7017(0.1300)**	**0.7305(0.1079)**	**0.7158(0.1149)**	**0.8174(0.0233)**

Data Sets	Precision	Recall	F1-measure	AUC
Coauthor	0.37%	13.76%	$\boxed{7.04\%}$	9.45%
Weibo	36.22%	40.14%	$\boxed{38.14\%}$	32.08%

Community Detection

We use a case study on Coauthor dataset to demonstrate its effectiveness in detecting communities qualitatively. According to ζ after training, we can obtain the representative researchers with the highest probabilities in each community.