# RAIN: social Role-Aware INformation diffusion

Yang Yang*, Jie Tang*, Cane Wing-ki Leung#, Yizhou Sun+,
Qicong Chen*, Juanzi Li*, and Qiang Yang#

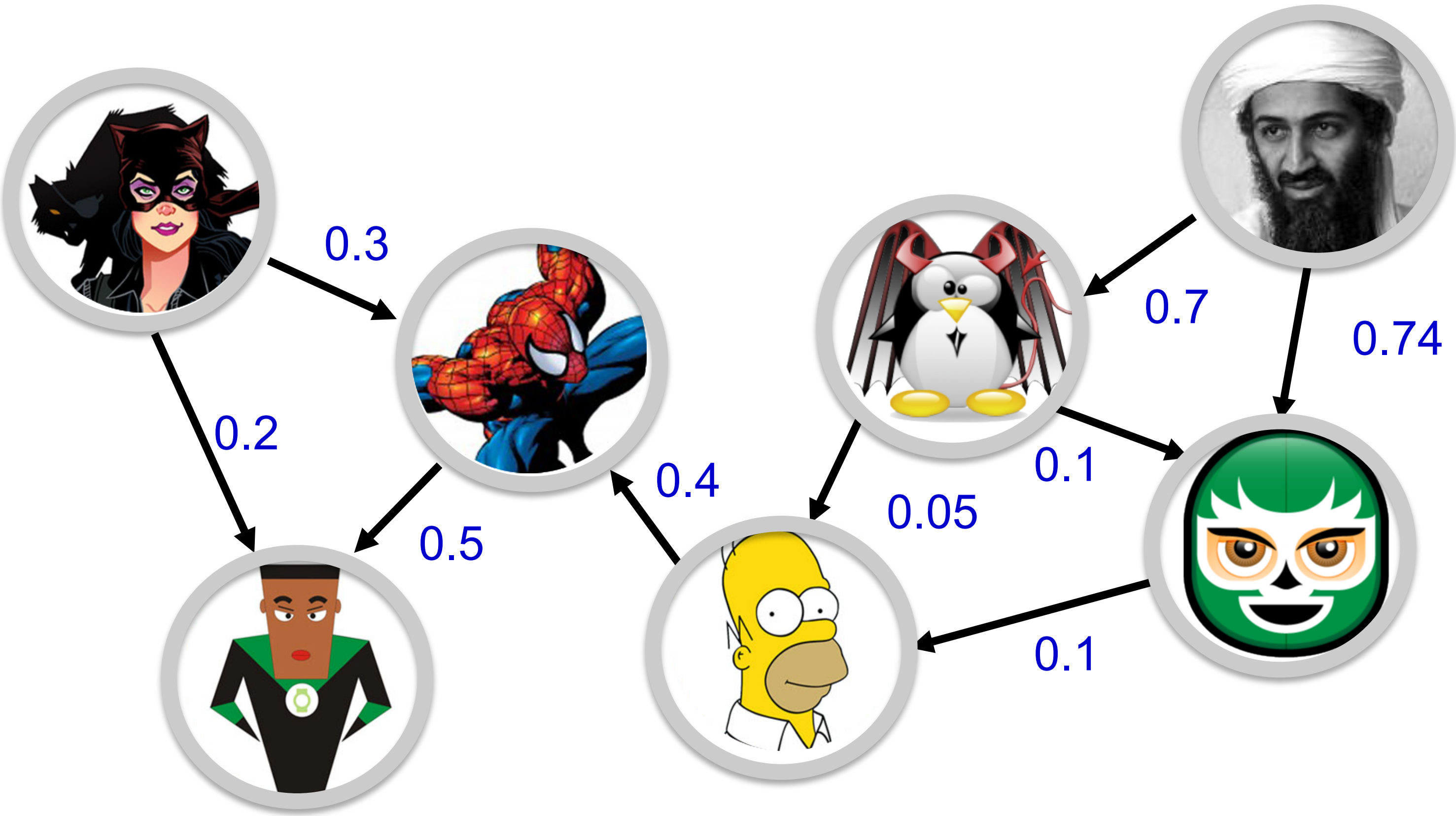
*Tsinghua University #Huawei Noah's Ark Lab +Northeastern University



Data&Code available at: http://aminer.org/rain

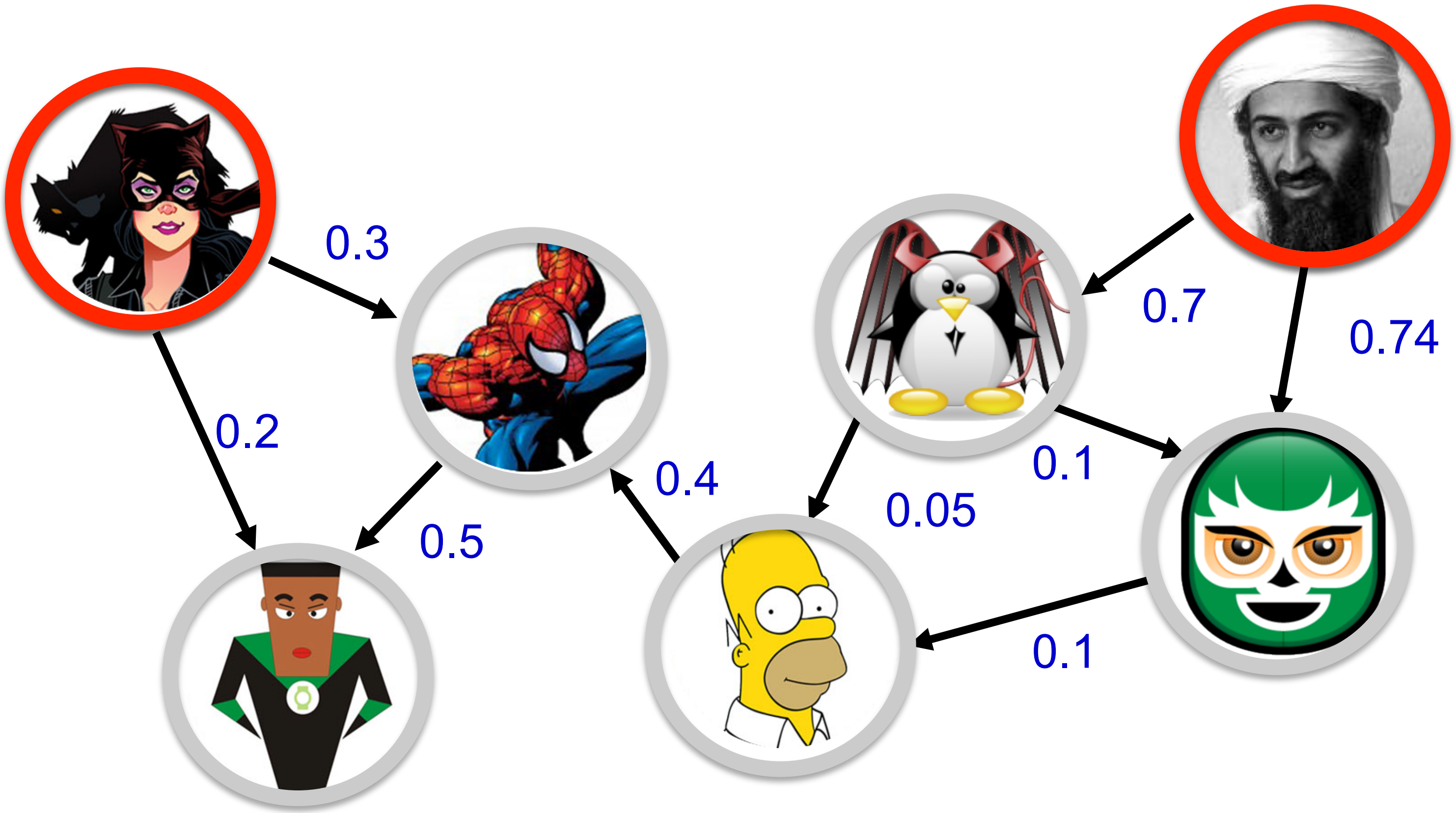
# Information Diffusion

- Users of a social network share information with neighbors



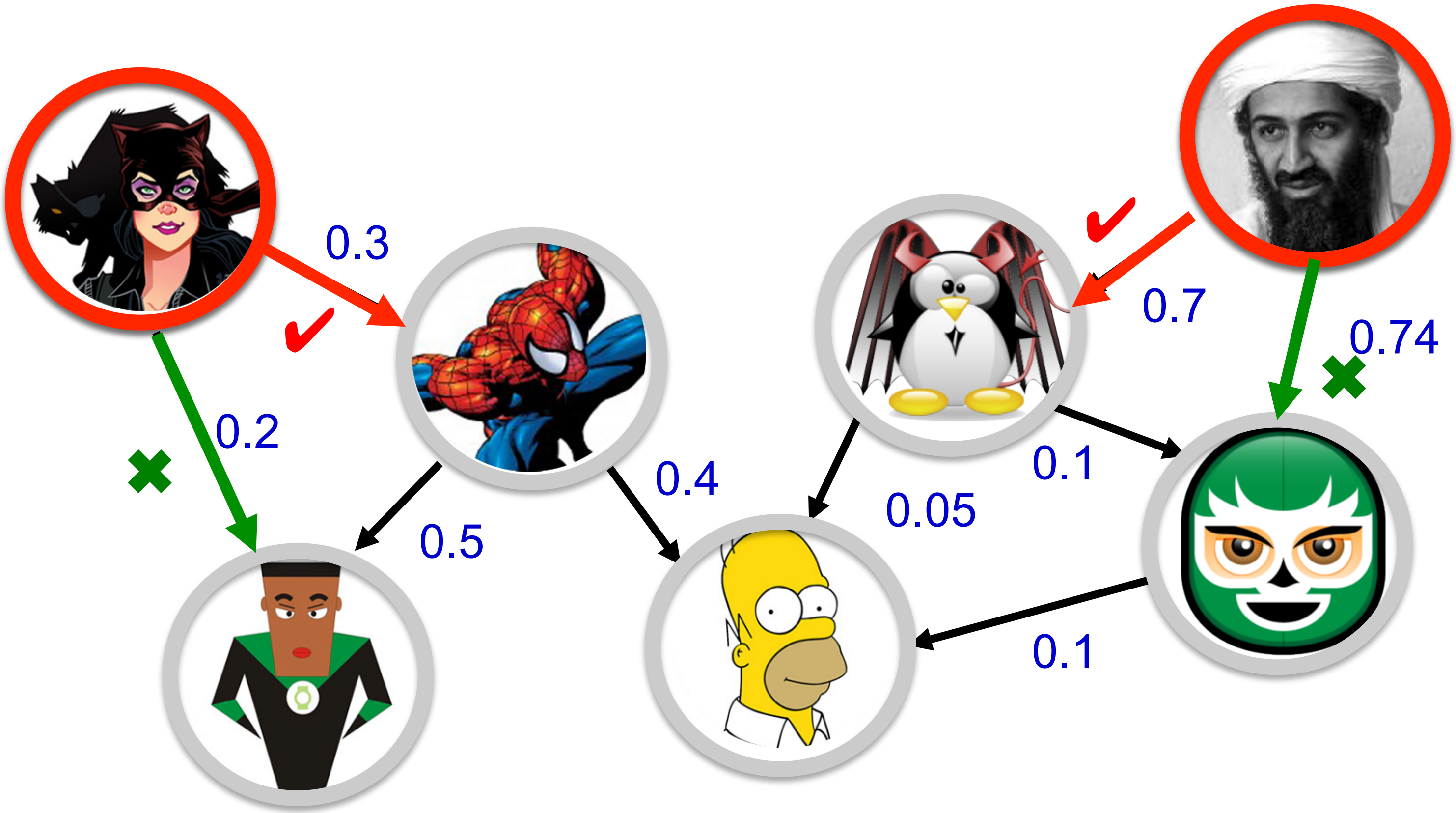
# Information Diffusion

- Users of a social network share information with neighbors



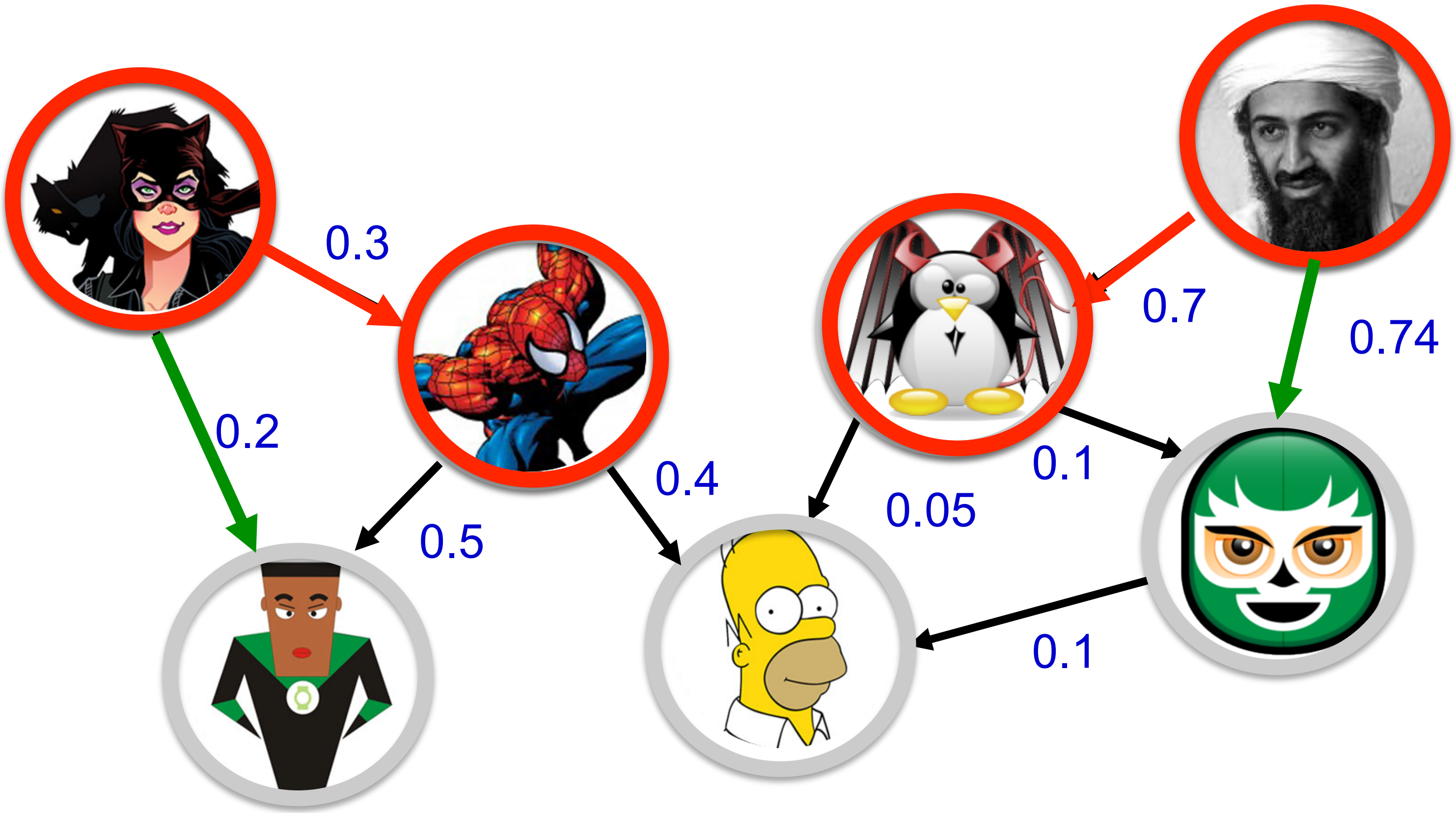
# Information Diffusion

- Users of a social network share information with neighbors



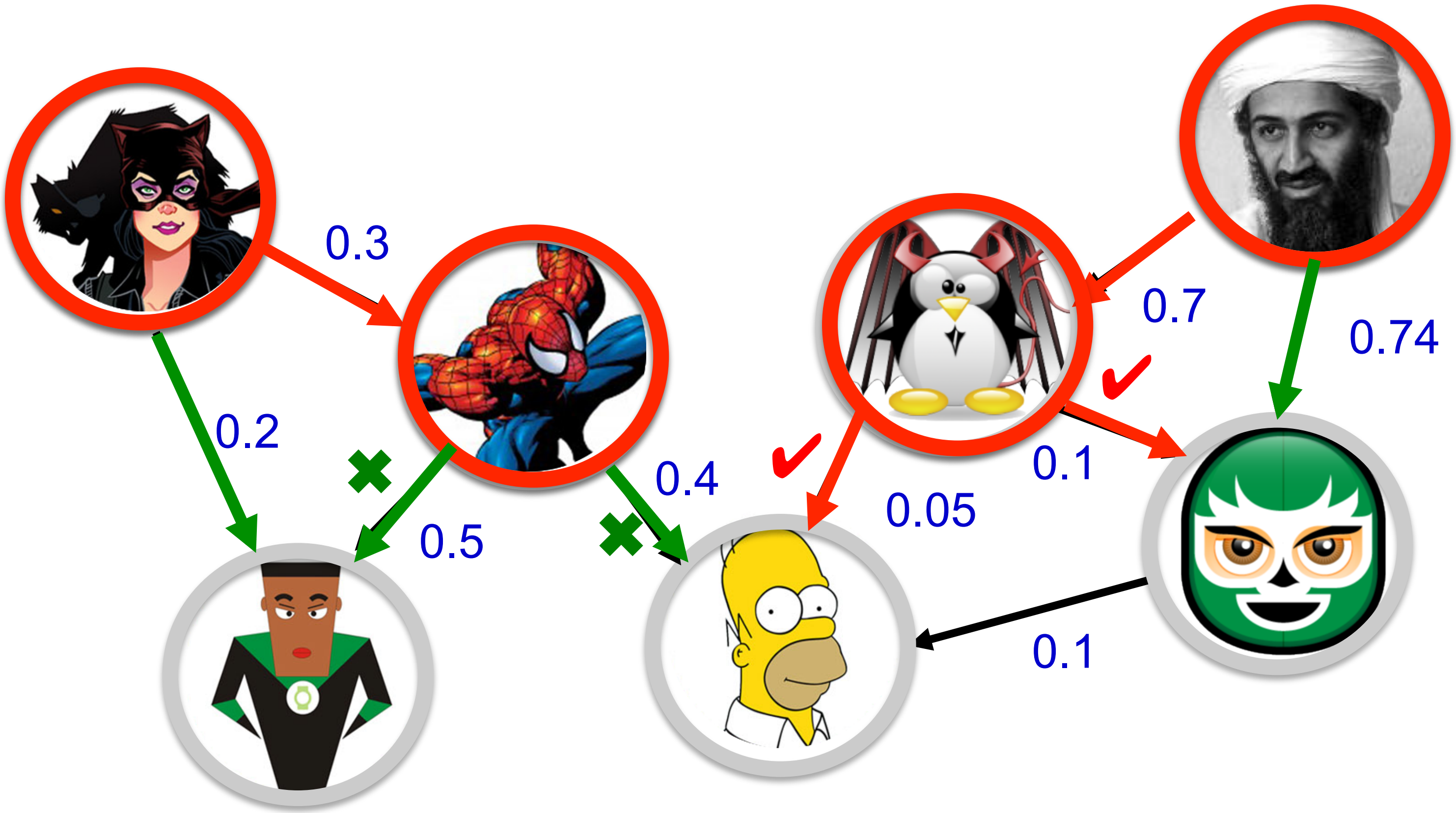
# Information Diffusion

- Users of a social network share information with neighbors



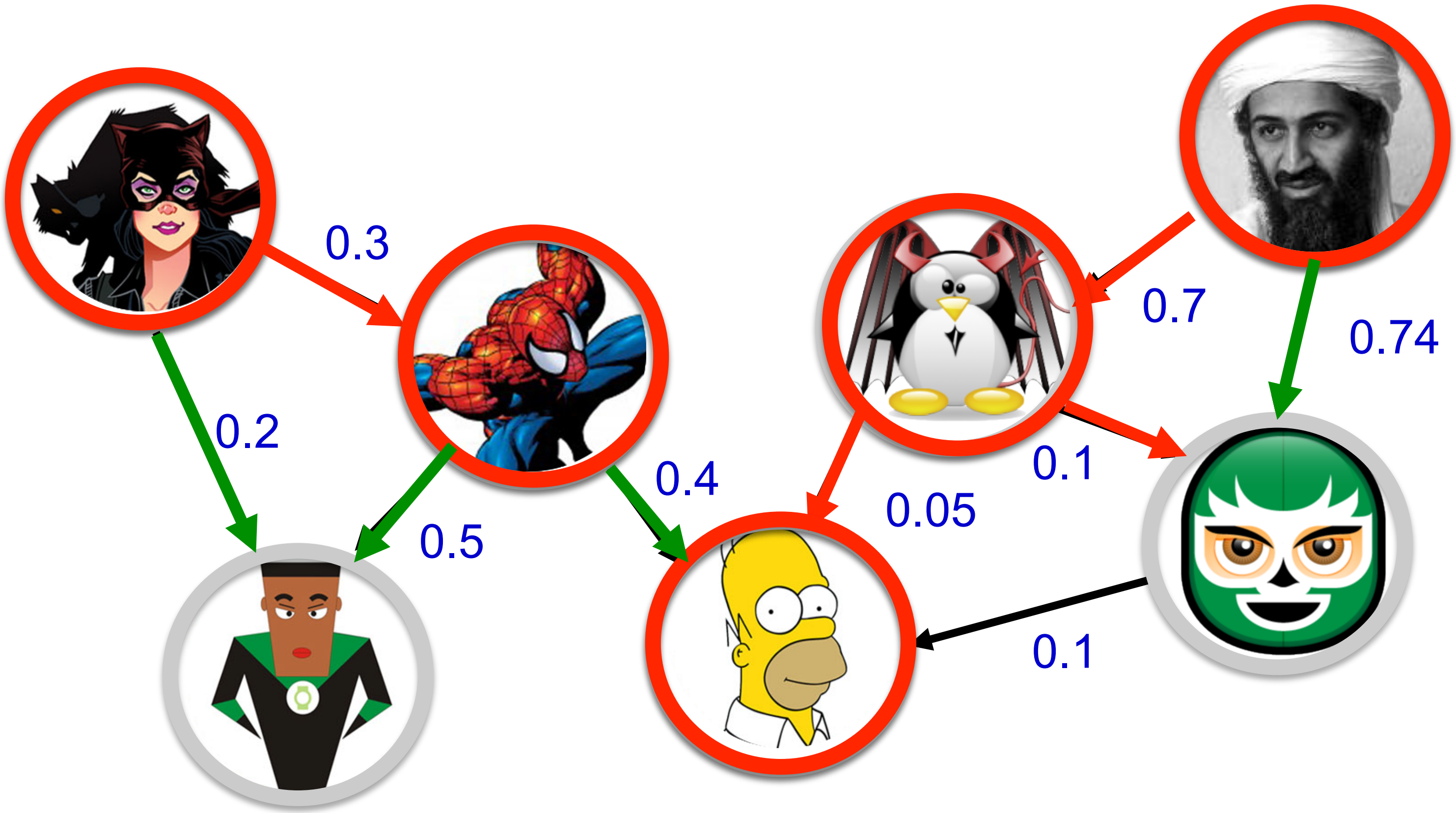
# Information Diffusion

- Users of a social network share information with neighbors



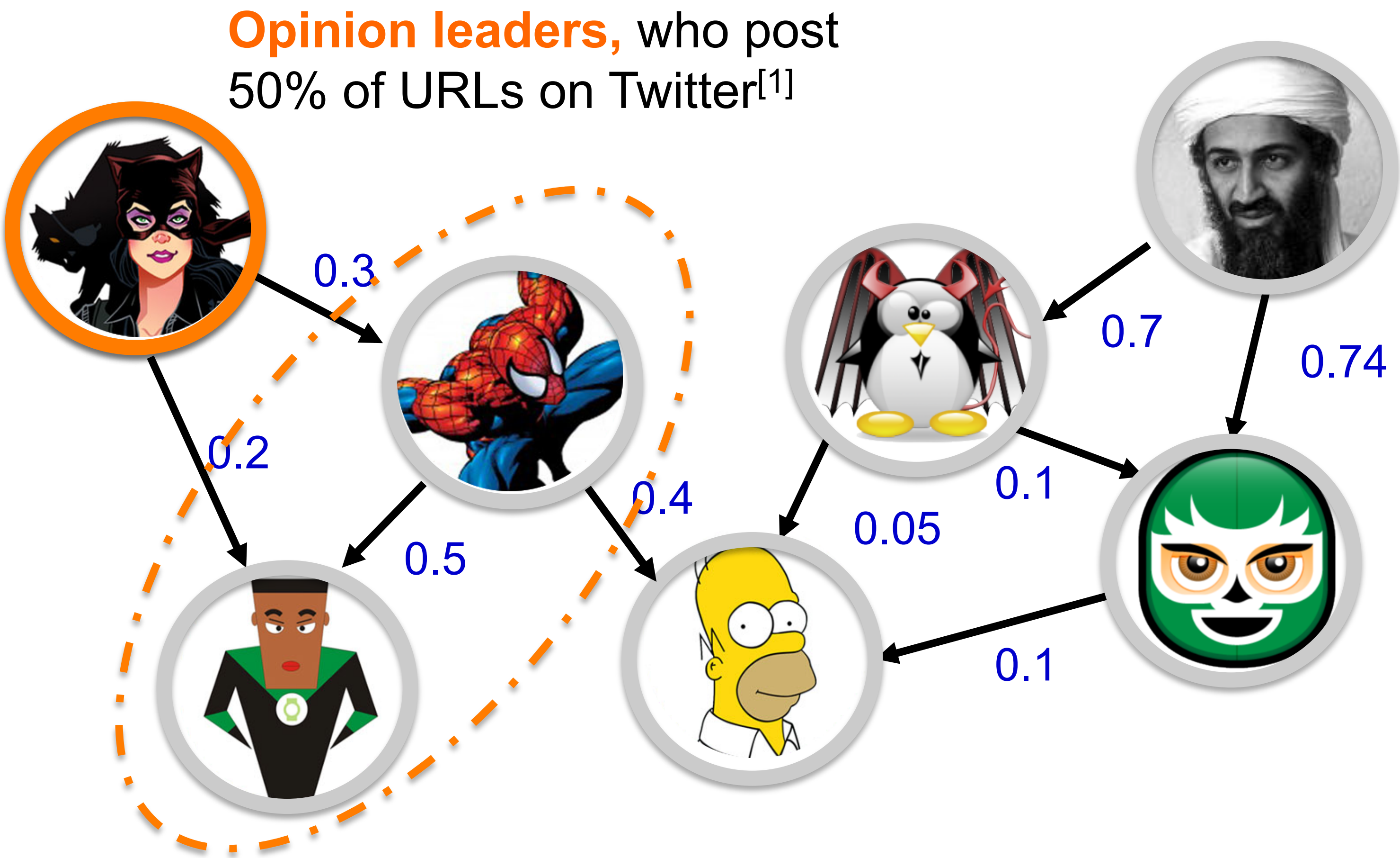
# Information Diffusion

- Users of a social network share information with neighbors



# Information Diffusion

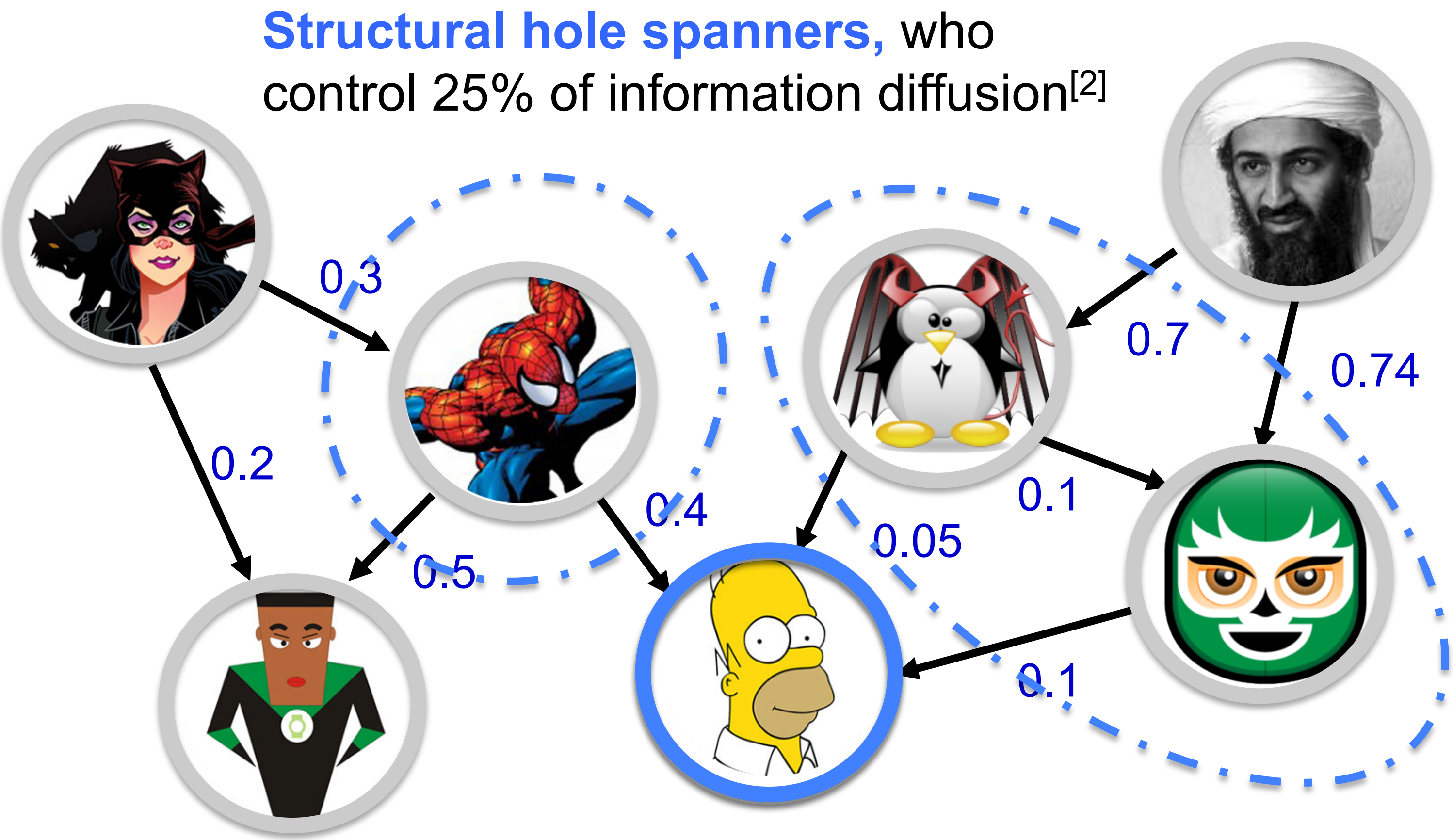
- Users of a social network share information with neighbors




[1] Wu, S.; Hofman, J. M.; Mason,W. A.; andWatts, D. J. 2011. Who says what to whom on twitter. In WWW'11 , 705–714.

# Information Diffusion

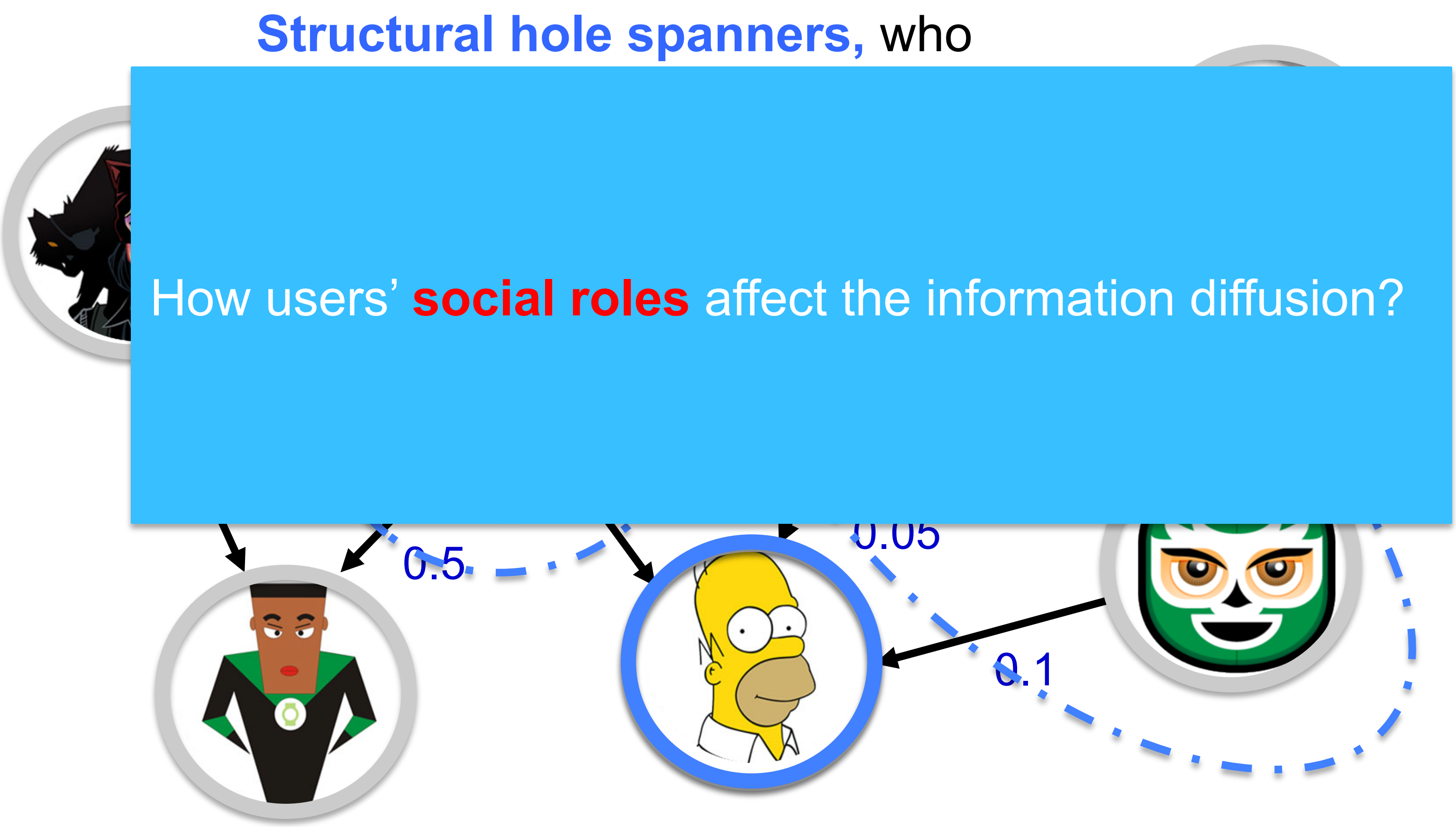
- Users of a social network share information with neighbors




[2] Lou, T., and Tang, J. 2013. Mining structural hole spanners through information diffusion in social networks. In WWW'13 , 825–836

# Information Diffusion

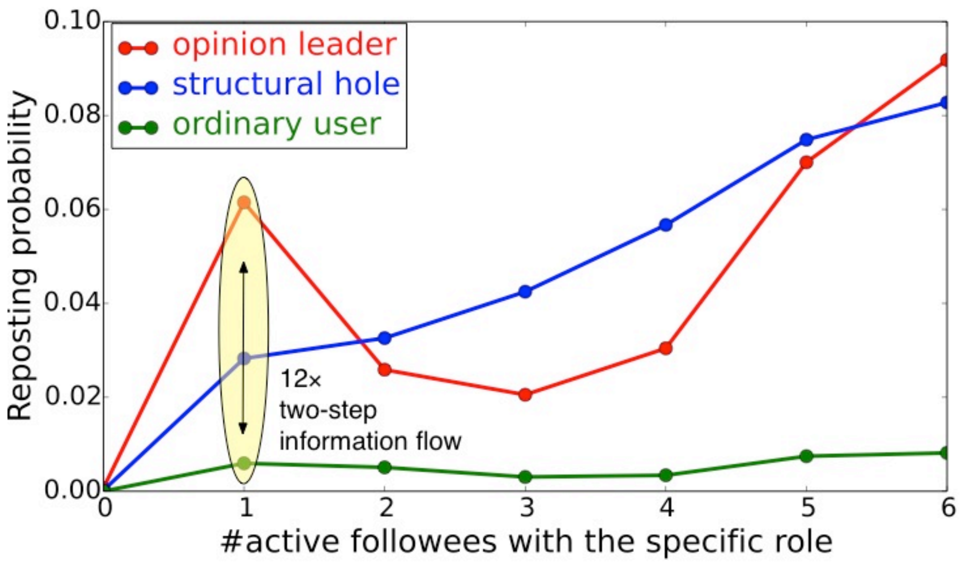
- Users of a social network share information with neighbors




[2] Lou, T., and Tang, J. 2013. Mining structural hole spanners through information diffusion in social networks. In WWW'13 , 825–836

# Exploratory Analysis

# Exploratory Analysis




**Study:** how users with different roles influence others

**Data**: a popular Chinese social media

>200 million users

>174 million posts

**Result**

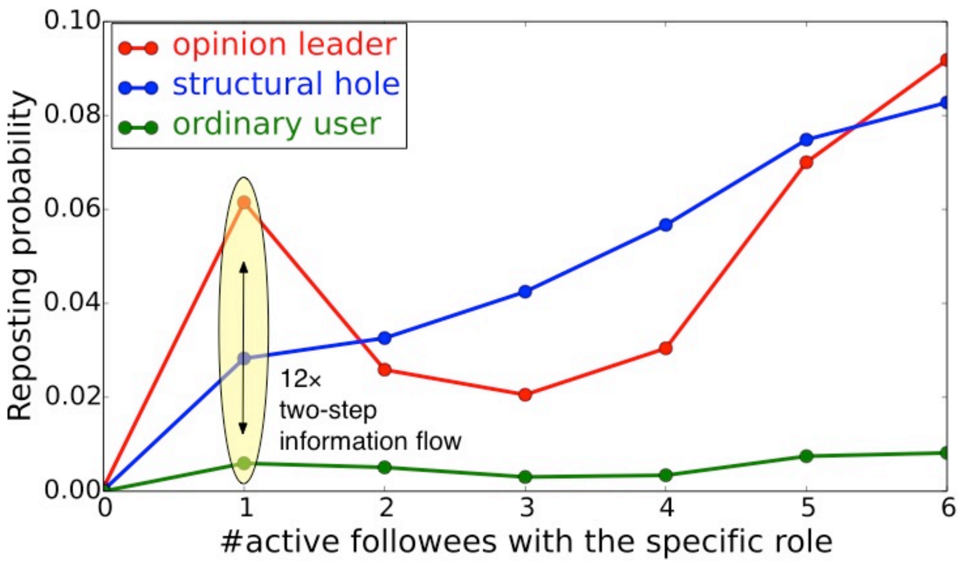
- X-axis: number of a user v's active followees with different social roles
- Y-axis: the probability of v bing activated

Opinion leader:

- Stage 1 - activation probability is12 times higher than ordinary user
- Stage 2 - information overload[1]: 2-3 opinion leaders are sufficient to spread a piece of information throughout a community.
- Stage 3 - information everywhere: spreading the information becomes a social norm to adopt.


[1] Lazarsfeld, P. F.; Berelson, B.; and Gaudet, H. 1944. The peoples choice: How the voter makes up his mind in a presidential election. New York: Duell, Sloan and Pearce .

# Exploratory Analysis




**Study:** how users with different roles influence others

**Data**: a popular Chinese social media

>200 million users

>174 million posts

**Result**

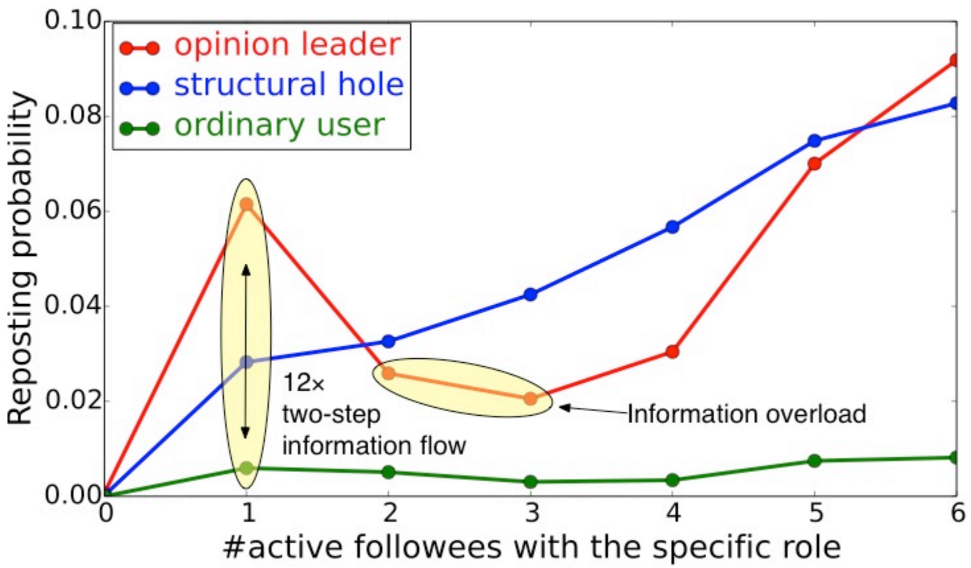
- X-axis: number of a user v's active followees with different social roles
- Y-axis: the probability of v bing activated

Opinion leader:

- Stage 1 - activation probability is12 times higher than ordinary user
- Stage 2 - information overload[1]: 2-3 opinion leaders are sufficient to spread a piece of information throughout a community.
- Stage 3 - information everywhere: spreading the information becomes a social norm to adopt.


[1] Lazarsfeld, P. F.; Berelson, B.; and Gaudet, H. 1944. The peoples choice: How the voter makes up his mind in a presidential election. New York: Duell, Sloan and Pearce .

# Exploratory Analysis




**Study:** how users with different roles influence others

**Data**: a popular Chinese social media

>200 million users

>174 million posts

**Result**

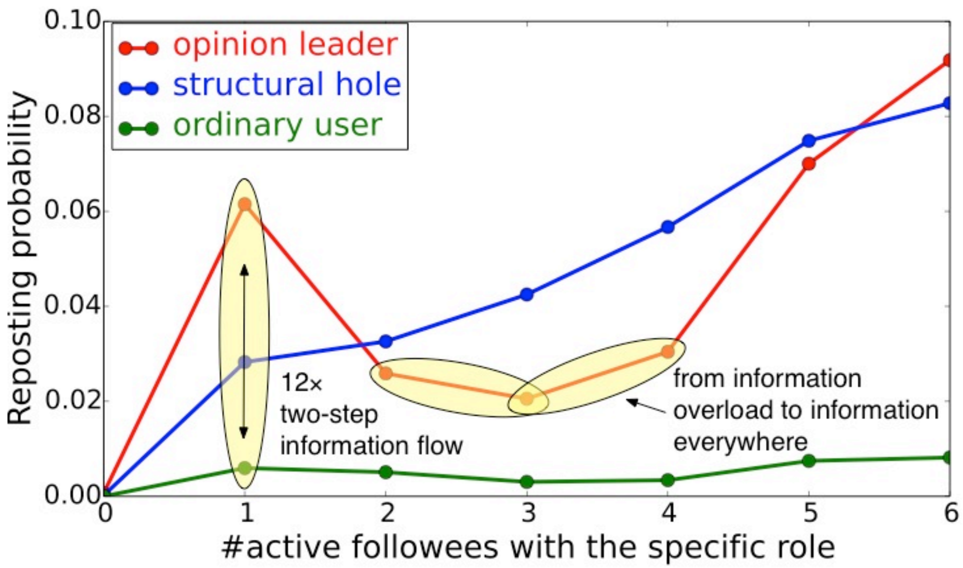
- X-axis: number of a user v's active followees with different social roles
- Y-axis: the probability of v bing activated

Opinion leader:

- Stage 1 - activation probability is12 times higher than ordinary user
- Stage 2 - information overload[1]: 2-3 opinion leaders are sufficient to spread a piece of information throughout a community.
- Stage 3 - information everywhere: spreading the information becomes a social norm to adopt.

[1] Lazarsfeld, P. F.; Berelson, B.; and Gaudet, H. 1944. The peoples choice: How the voter makes up his mind in a presidential election. New York: Duell, Sloan and Pearce .

# Exploratory Analysis




**Study:** how users with different roles influence others

**Data**: a popular Chinese social media

>200 million users

>174 million posts

**Result**

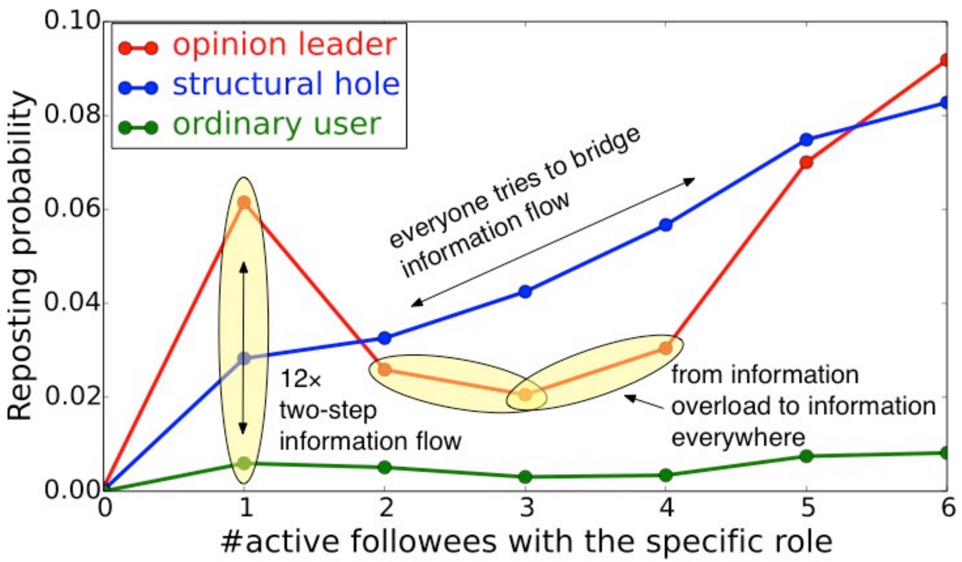
- X-axis: number of a user v's active followees with different social roles
- Y-axis: the probability of v bing activated

Opinion leader:

- Stage 1 - activation probability is12 times higher than ordinary user
- Stage 2 - information overload[1]: 2-3 opinion leaders are sufficient to spread a piece of information throughout a community.
- Stage 3 - information everywhere: spreading the information becomes a social norm to adopt.

[1] Lazarsfeld, P. F.; Berelson, B.; and Gaudet, H. 1944. The peoples choice: How the voter makes up his mind in a presidential election. New York: Duell, Sloan and Pearce .

# Exploratory Analysis




**Study:** how users with different roles influence others

**Data**: a popular Chinese social media

>200 million users

>174 million posts

**Result**

- X-axis: number of a user v's active followees with different social roles
- Y-axis: the probability of v bing activated

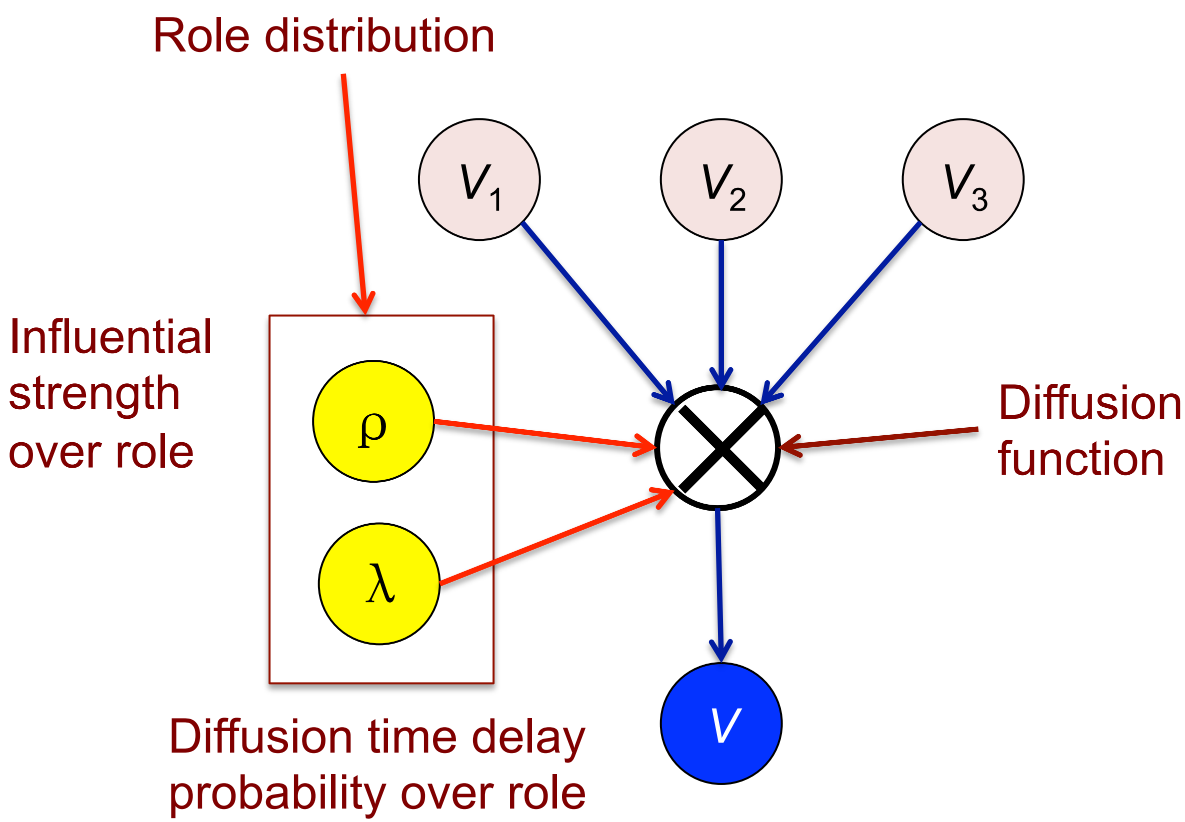
Structural hole spanners[2][3]:

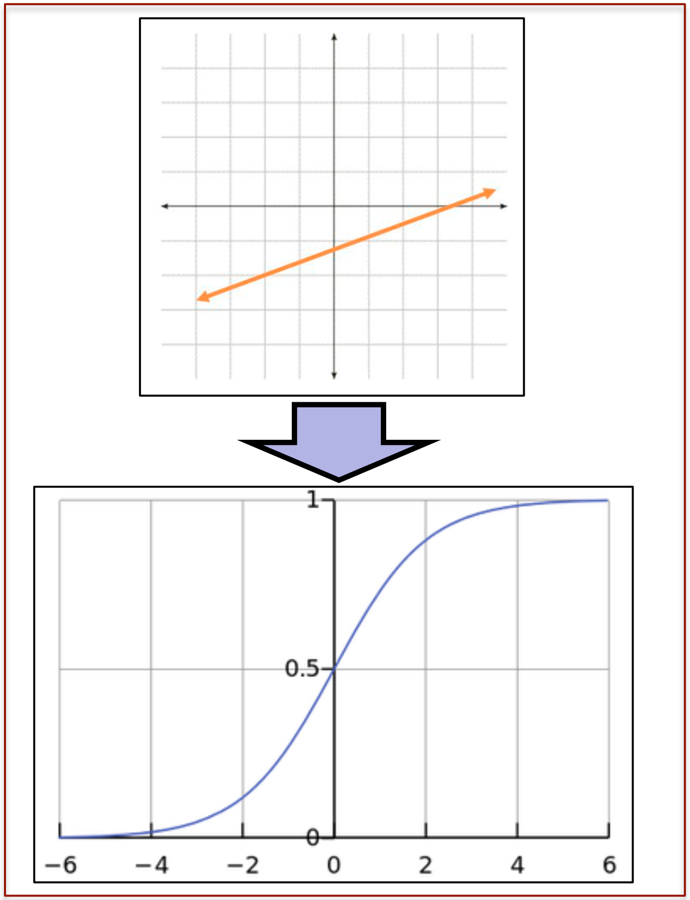
- SH tend to bring information that a certain community is rarely exposed to.
- Most users tries to bridge information flow between different groups.


[2] Burt, R. S. 2001. Structural holes versus network closure as social capital. Social capital: Theory and research 31–56.
[3] Burt, R. S. 2009. Structural holes: The social structure of competition . Harvard University Press.

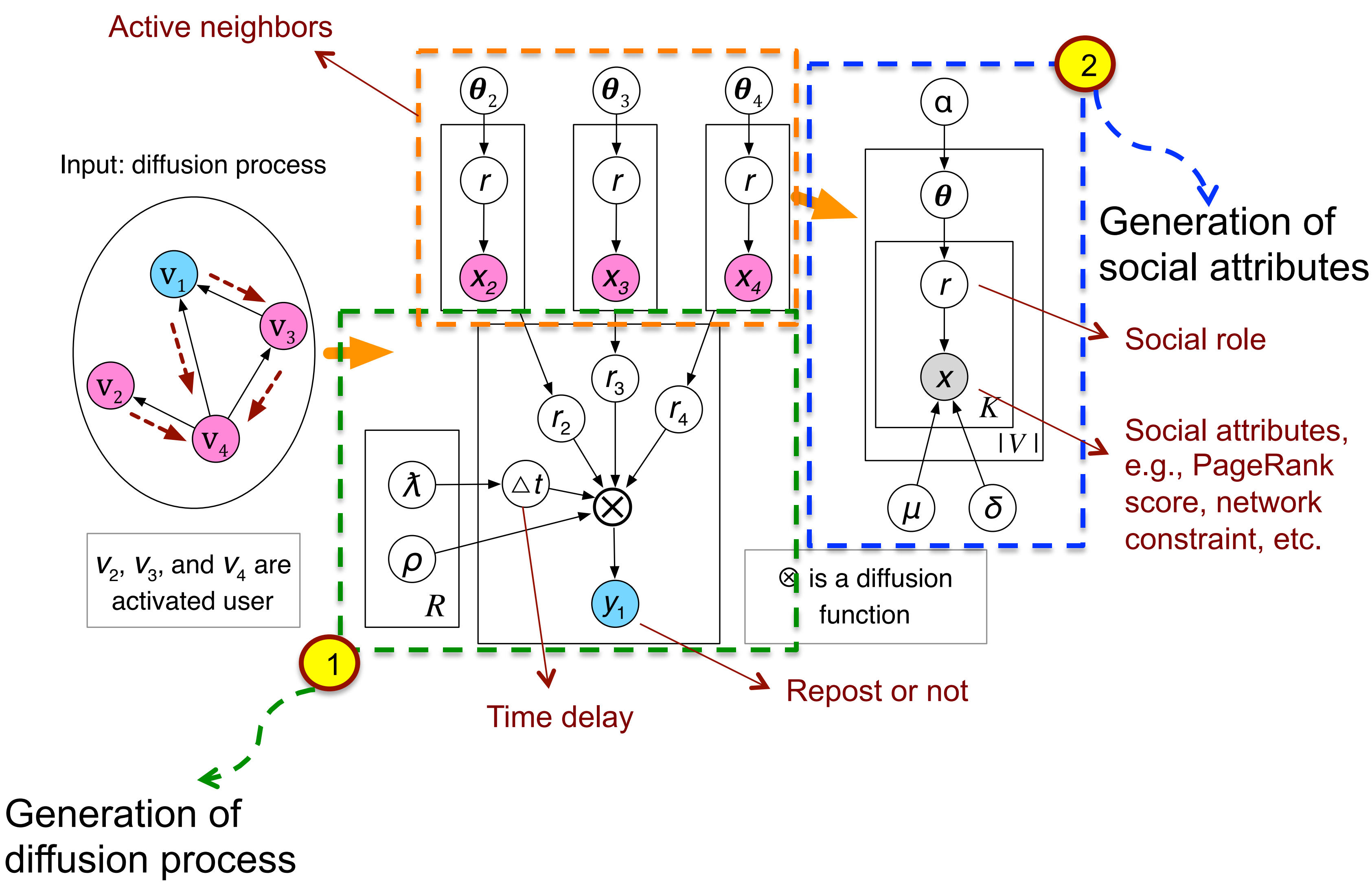
# Model

# Model: General Idea

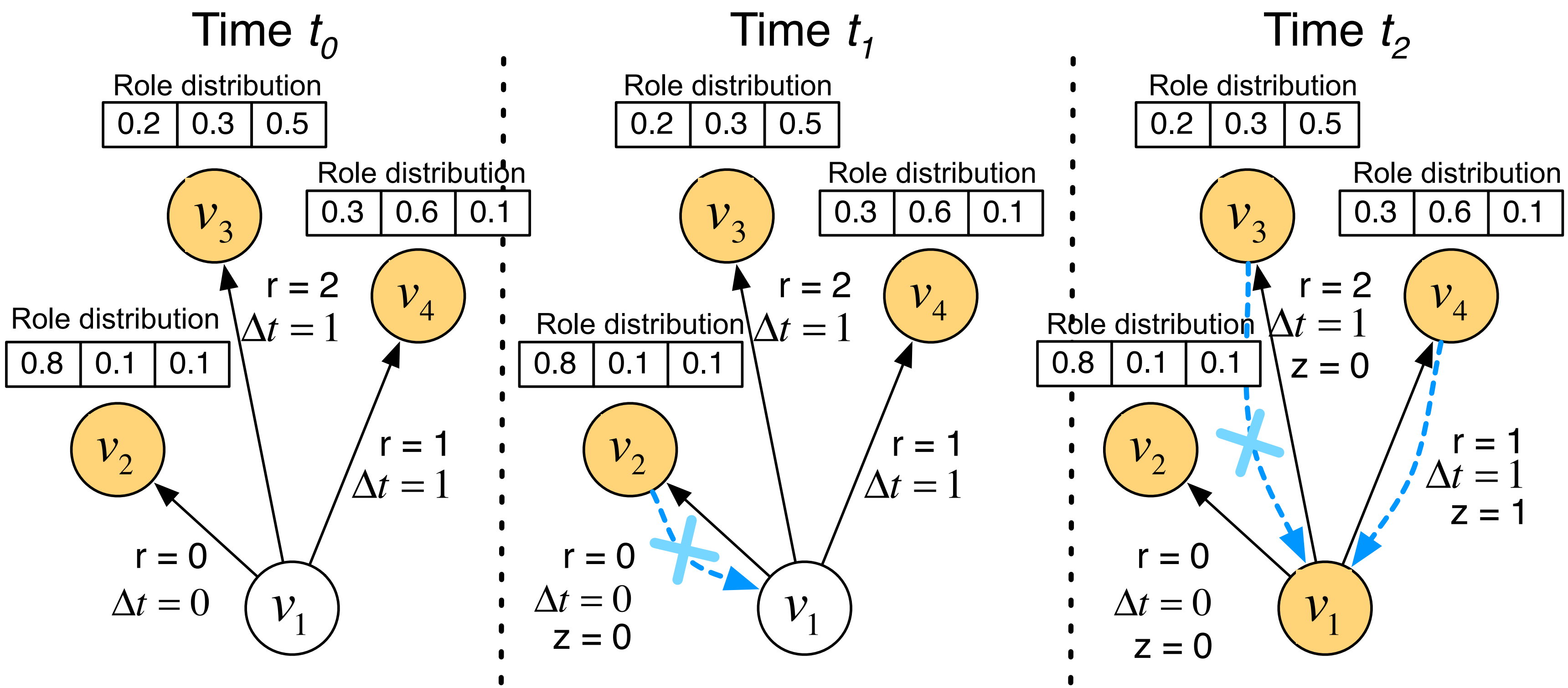


# RAIN (Role Aware INformation diffusion)



# An Example




$$P(v_1 \subseteq A_{t_1}) = \rho_0 \lambda_0 \theta_{v_2 0}$$

# RAIN: Objective Function

- Likelihood:

$$L = \prod_{i=1}^{I}\prod_{t=1}^{T}\prod_{v\in A_{it}} P(v \in A_{it}) \times \prod_{i=1}^{I}\prod_{v\notin D_{iT}} P(v \notin D_{iT})$$

$$\times \prod_{u\in V}\prod_{k=1}^{K} P(x_{uk}) \times \prod_{u\in V}\prod_{r=1}^{R} P(\theta_{ur}|\alpha)$$

$$\times \prod_{r=1}^{R}\{P(\rho_r|\beta) + P(\lambda_r|\gamma)\} \times \prod_{r=1}^{R}\prod_{k=1}^{K} P(\mu_{rk}, \delta_{rk}|\tau)$$

The probability of user *v* adopting the information *i* at time *t*

$$P(v \in A_{it}) = \boxed{\sum_{\mathbf{z}^t_{i*v}} P(\mathbf{z}_{i*v})} - \prod_{u\in B(v)\cap D_{it-1}} \boxed{P(z^t_{iuv} = 0)}$$

$$= \prod_{u\in B(v)\cap D_{it-1}} (\varphi^t_{iuv} + \varepsilon^t_{iuv}) - \prod_{u\in B(v)\cap D_{it-1}} \varepsilon^t_{iuv}.$$

All adoptions

Failed adoptions

The probability of user *v* never adopts the information *i*

$$P(v \notin D_{iT}) = \prod_{u\in B(v)\cap D_{iT}} \sum_r (1-\rho_r)\theta_{ur}.$$

Assumption here:
T >> the last observed timestamp

The probability of user *v* with the social attributes $x_{vk}$

$$P(x_{uk}) = \sum_r \sqrt{\frac{\delta_{rk}}{2\pi}} \exp\{-\frac{\delta_{rk}(x_{uk}-\mu_{rk})^2}{2}\}\theta_{ur}.$$

A mixture of Gaussian

Priors to model parameters

# Model Learning

- We utilize Gibbs sampling to estimate model parameters
- Sample latent role *r* for user u's each social attribute

$$P(r_{uk}|\mathbf{r}_{\neg uk}, \mathbf{x}) = \frac{P(\mathbf{x}, \mathbf{r})}{P(\mathbf{x}_{\neg uk}, \mathbf{r}_{\neg uk})} = \frac{n^{\neg uk}_{ur_{uk}} + \alpha}{\sum_r (n^{\neg uk}_{ur} + \alpha)} \boxed{\frac{\Gamma(\tau_2 + \frac{n_{r_{uk}k}}{2})}{\Gamma(\tau_2 + \frac{n^{\neg uk}_{r_{uk}k}}{2})}}$$
$$\times \frac{\sqrt{(\tau_1 + n^{\neg uk}_{r_{uk}k})}\,\eta(n^{\neg uk}_{r_{uk}k}, \bar{x}^{\neg uk}_{r_{uk}k}, s^{\neg uk}_{r_{uk}})}{\sqrt{(\tau_1 + n_{r_{uk}k})}\,\eta(n_{r_{uk}k}, \bar{x}_{r_{uk}k}, s_{r_{uk}})},$$

Using Stirling's formula to calculate the Gamma functions approximately

- Sample role *r*, activation delay *t*, and activation result *z* for each adoption

$$\begin{aligned}
&P(r_{iuv}, \Delta t_{iuv}, z_{iuv}|\mathbf{r}_{\neg iuv}, \Delta \mathbf{t}_{\neg iuv}, \mathbf{z}_{\neg iuv}, \mathbf{y}) \\
&= \frac{P(\mathbf{r}, \Delta \mathbf{t}, \mathbf{z}, \mathbf{y})}{P(\mathbf{r}_{\neg iuv}, \Delta \mathbf{t}_{\neg iuv}, \mathbf{z}_{\neg iuv}, \mathbf{y}_{\neg iuv})} \\
&= \frac{n^{\neg iuv}_{ur_{iuv}} + \alpha}{\sum_r (n^{\neg iuv}_{ur} + \alpha)} \times \frac{n^{\neg iuv}_{z_{iuv}r_{iuv}} + \beta_1^{z_{iuv}} \beta_0^{1 - z_{iuv}}}{n^{\neg iuv}_{1r_{iuv}} + \beta_1 + n^{\neg iuv}_{0r_{iuv}} + \beta_0} \\
&\times \frac{(n^{\neg iuv}_{r_{iuv}} + \gamma_1) \prod_{t=0}^{\Delta t - 2} (s^{\neg iuv}_{r_{iuv}} - n^{\neg iuv}_{r_{iuv}} + \gamma_0 + t)}{\prod_{t=0}^{\Delta t - 1} (\gamma_1 + s^{\neg iuv}_{r_{iuv}} + \gamma_0 + t)} \times \Phi,
\end{aligned}$$

- Update model parameters according to sampling results

# Experiments

**Micro-level:** predicting whether a user will repost a given message

**Macro-level:** predicting scale and duration of a diffusion process

# Micro-level Prediction

Table 2: Performance of repost prediction on several topics.

| Topic | Method | P@10 | P@50 | P@100 | MAP |
|---|---|---|---|---|---|
| Campus | C | | | | |
| | S | | | | |
| | IC | | | | |
| | R | | | | |
| Horoscope | C | | | | |
| | S | | | | |
| | IC | | | | |
| | R | | | | |
| Movie | C | | | | |
| | S | | | | |
| | IC | | | | |
| | R | | | | |
| History | C | | | | |
| | S | | | | |
| | IC | | | | |
| | R | | | | |
| Society | C | | | | |
| | S | | | | |
| | IC | | | | |
| | R | | | | |
| Health | C | | | | |
| | S | | | | |
| | IC | | | | |
| | R | | | | |
| Political | C | | | | |
| | S | | | | |
| | IC Model | [illegible] | [illegible] | [illegible] | [illegible] |
| | RAIN | **0.216** | **0.164** | **0.130** | **0.239** |
| Travel | Count | 0.142 | 0.056 | 0.031 | 0.103 |
| | SVM | 0.094 | 0.048 | 0.032 | 0.128 |
| | IC Model | **0.206** | 0.120 | 0.098 | 0.254 |
| | RAIN | 0.194 | **0.159** | **0.126** | **0.260** |

**Goal**: predict whether a user will repost a particular post

**Data**: complete Tencent Weibo data on Nov. 1-2, 2012

- >4.5 source posts
- >18k users
- Posts are categorized based on topics: *campus, constellation, movie, history, society, health, political, and travel*
- Posts on Nov.1 as train data, Nov. 2 as test data

# Micro-level Prediction

Table 2: Performance of repost prediction on several topics.

| Topic | Method | P@10 | P@50 | P@100 | MAP |
|---|---|---|---|---|---|
| Campus | Count | 0.028 | 0.010 | 0.006 | 0.068 |
| | SVM | | | | |
| | IC Model | | | | |
| | RAIN | | | | |
| Horoscope | Count | | | | |
| | SVM | | | | |
| | IC Model | | | | |
| | RAIN | | | | |
| Movie | Count | | | | |
| | SVM | | | | |
| | IC Model | | | | |
| | RAIN | | | | |
| History | Count | | | | |
| | SVM | | | | |
| | IC Model | | | | |
| | RAIN | | | | |
| Society | Count | | | | |
| | SVM | | | | |
| | IC Model | | | | |
| | RAIN | | | | |
| Health | Count | | | | |
| | SVM | | | | |
| | IC Model | | | | |
| | RAIN | | | | |
| Political | Count | | | | |
| | SVM | | | | |
| | IC Model | | | | |
| | RAIN | | | | |
| Travel | Count | | | | |
| | SVM | | | | |
| | IC Model | | | | |
| | RAIN | | | | |

## Baselines:

Count: ranks users by the number of active followees

SVM: Support Vector Machine, majorly considers features as

- *#active followers*
- *#active followees*
- *#whether the user have reposted similar messages*

IC Model: traditional IC model with fitted parameters[1]

**RAIN**: **R**ole **A**ware **IN**formation diffusion

## Evaluation Metrics:

Precision@K (K=10, 50, 100)

Mean Average Precision (MAP)

[1] Kimura, M.; Saito, K.; Ohara, K.; and Motoda, H. 2011. Learning information diffusion model in a social network for predicting influence of nodes. Intelligent Data Analysis 15(4):633–652.

# Micro-level Prediction

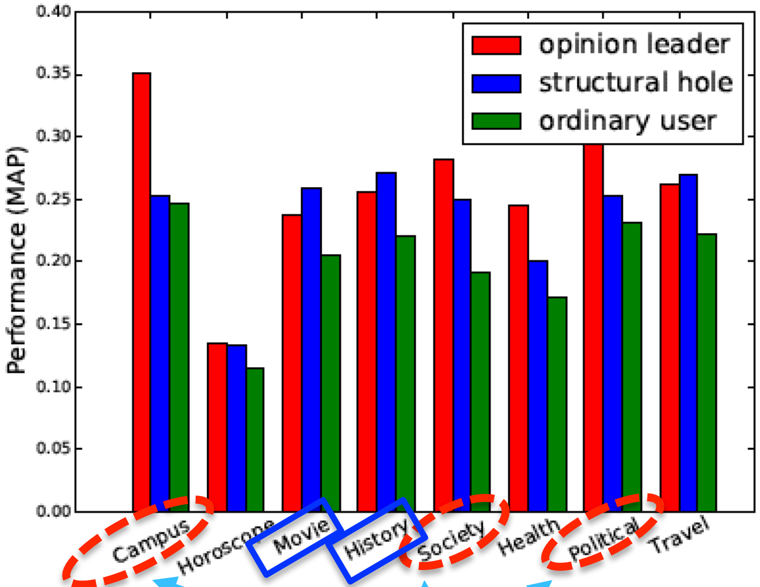
Table 2: Performance of repost prediction on several topics.

| Topic | Method | P@10 | P@50 | P@100 | MAP |
|---|---|---|---|---|---|
| Campus | Count | 0.028 | 0.010 | 0.006 | 0.068 |
| | SVM | 0.098 | 0.045 | 0.032 | 0.127 |
| | IC Model | **0.231** | 0.142 | 0.102 | 0.259 |
| | RAIN | 0.228 | **0.145** | **0.106** | **0.263** |
| Horoscope | Count | 0.019 | 0.010 | 0.006 | 0.005 |
| | SVM | 0.124 | **0.162** | 0.088 | **0.263** |
| | IC Model | 0.149 | 0.111 | 0.098 | 0.125 |
| | RAIN | **0.171** | 0.121 | **0.102** | 0.130 |
| Movie | Count | 0.015 | 0.007 | 0.004 | 0.009 |
| | SVM | 0.094 | 0.111 | 0.060 | 0.199 |
| | IC Model | 0.227 | 0.147 | **0.147** | 0.236 |
| | RAIN | **0.229** | **0.173** | 0.144 | **0.238** |
| History | Count | 0.191 | 0.056 | 0.033 | 0.096 |
| | SVM | 0.154 | 0.051 | 0.030 | 0.221 |
| | IC Model | 0.206 | 0.134 | **0.135** | 0.230 |
| | RAIN | **0.225** | **0.171** | 0.134 | **0.262** |
| Society | Count | 0.245 | 0.058 | 0.029 | 0.156 |
| | SVM | 0.100 | 0.023 | 0.012 | 0.122 |
| | IC Model | 0.171 | 0.131 | **0.109** | 0.198 |
| | RAIN | **0.176** | **0.140** | 0.106 | **0.204** |
| Health | Count | 0.041 | 0.008 | 0.005 | 0.035 |
| | SVM | 0.164 | 0.064 | 0.039 | **0.197** |
| | IC Model | 0.169 | 0.113 | 0.096 | 0.162 |
| | RAIN | **0.175** | **0.134** | **0.115** | 0.185 |
| Political | Count | 0.019 | 0.005 | 0.003 | 0.007 |
| | SVM | 0.104 | 0.077 | 0.039 | 0.176 |
| | IC Model | 0.209 | 0.132 | 0.102 | 0.224 |
| | RAIN | **0.216** | **0.164** | **0.130** | **0.239** |
| Travel | Count | 0.142 | 0.056 | 0.031 | 0.103 |
| | SVM | 0.094 | 0.048 | 0.032 | 0.128 |
| | IC Model | **0.206** | 0.120 | 0.098 | 0.254 |
| | RAIN | 0.194 | **0.159** | **0.126** | **0.260** |

**Comparison Results:**

- Count: performs worst due to the lack of supervised information.
- SVM: performs well on local topics but falls short on global topics.
- IC Model: suffers from model complexity.
- RAIN: improves the performance ***+32.6%*** in terms of MAP by reducing model complexity.

# Social Role Analysis




e.g., Campus, Society, Political

e.g., Movie, History

Opinion leaders can be better predicted on more regional and specialized topics.

Structural hole spanners can be better predicted on more general topics.

Ordinary users tend to behave more randomly and hard to be predicted.

# Macro-level Prediction

- We predict the ***scale*** of a diffusion process
  - X-axis: the number of reposts
  - Y-axis: the proportion of original posts with particular number of reposts

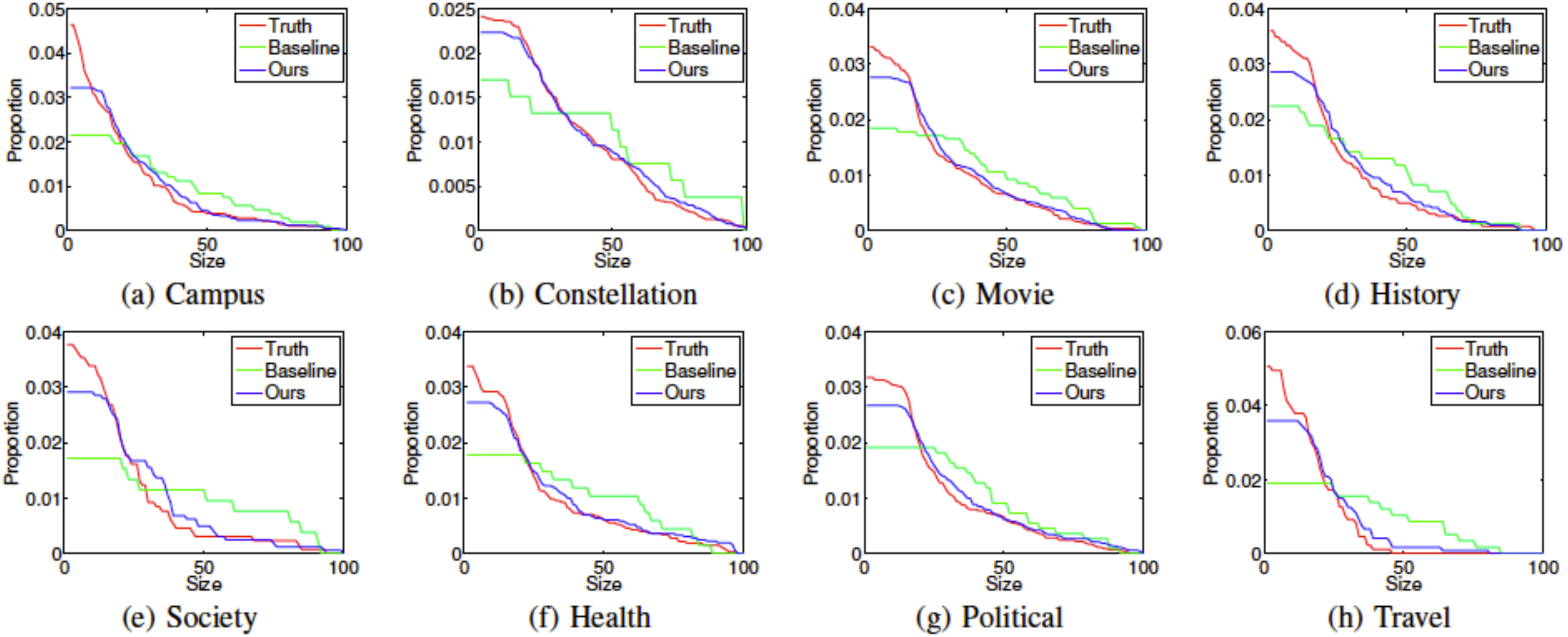



**Figure 8: Diffusion scale distributions of the different topics in the test set.**

# Macro-level Prediction

- We predict the ***duration*** of a diffusion process
  - X-axis: the time interval between the first and last posts
  - Y-axis: the proportion of original posts with particular time interval

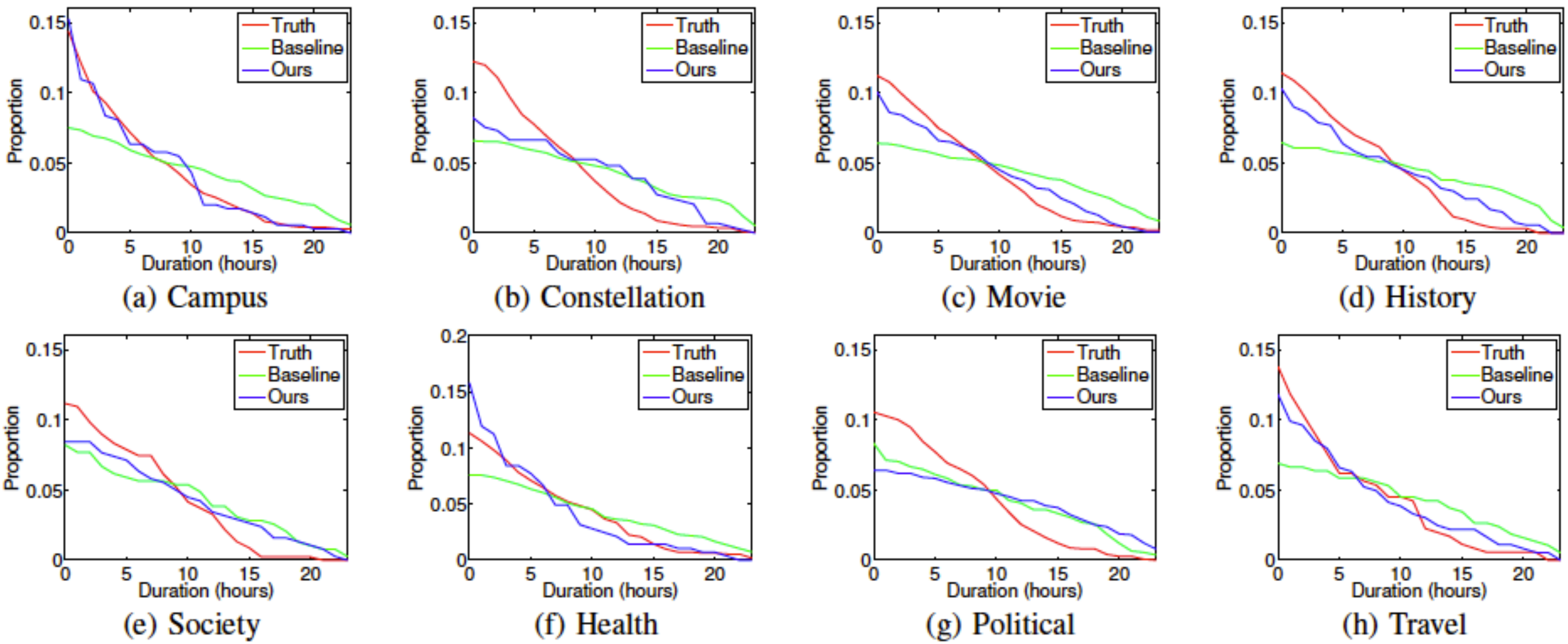



**Figure 9: Diffusion duration distributions of the different topics in the test set.**

# Conclusion

- We study the interplay between users' social roles and their influence to information diffusion.
- We propose a Role-Aware INformation diffusion (RAIN) model.
- We evaluate the proposed model on a real social media data set at both micro- and macro- levels.

# Thank You!

**RAIN: social Role-Aware INformation diffusion**

Yang Yang*, Jie Tang*, Cane Wing-ki Leung#, Yizhou Sun+,
Qicong Chen*, Juanzi Li*, and Qiang Yang#

*Tsinghua University #Huawei Noah's Ark Lab +Northeastern University

Tsinghua University

HUAWEI

Data&Code available at: http://aminer.org/rain