DSP 2014 August 22, 2014

Defense Against Sybil Attacks in Directed Social Networks

Pengfei Liu, Xiaohan Wang, Xiangqian Che, Zhaoqun Chen, and Yuantao Gu

Department of Electronic Engineering,
Tsinghua University

Outline

- Introduction
- Model
- Proposed method
- Experiment results
- Conclusion

Spamming in microblogging services

- Microblogging services become popular, such as Twitter and Sina Weibo
  - Relations are directed
  - Directed social networks
- Spammers in microblogging services
  - Phishing
  - Advertising on counterfeit products
  - Propagating illegal messages
  - Faking trends
  - Misleading public opinion

Defense against spammers: methods

- User-profile based
  - e.g. Profile Integrity; Photo; Number of Follower; Number of Following; Follower/Following Ratio
- Microblog based
  - e.g. URL count (or %); @ count (or %); microblog frequency; avg. time between microblogs; number of #

Defense against spammers: methods

- Social relation based
  - Existing work for spammer detection methods in undirected social networks, they use e.g. modularity, random walks
  - We haven't seen any work for directed social network including microblogging services

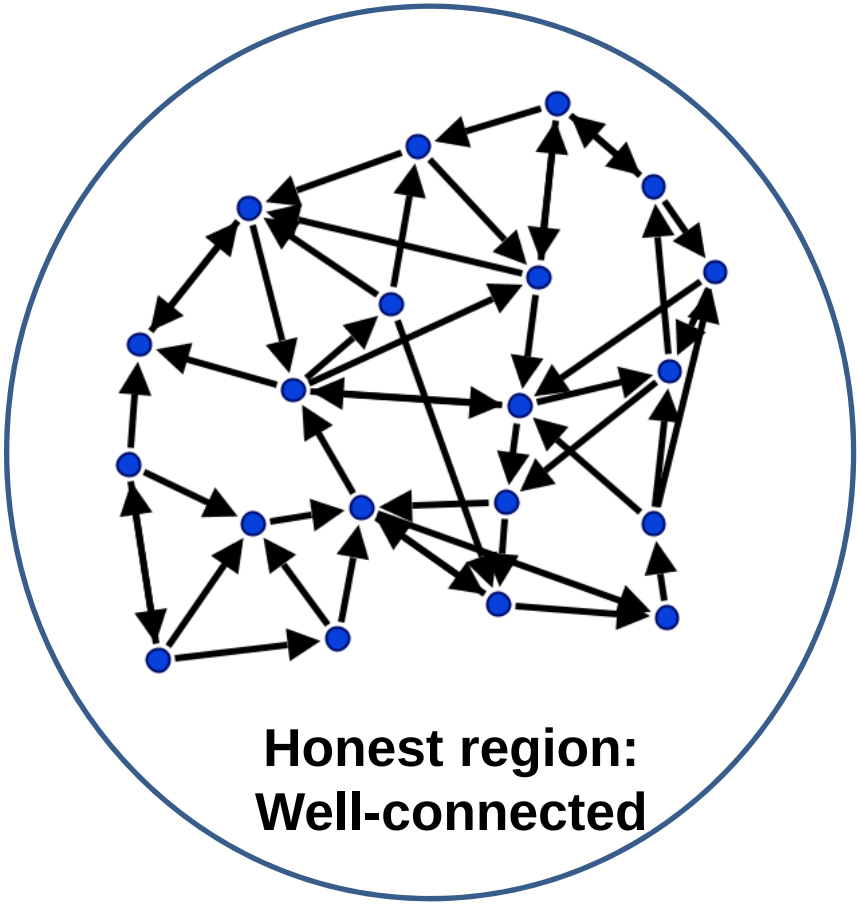
Outline

- Introduction
- Model
- Proposed method
- Experiment results
- Conclusion

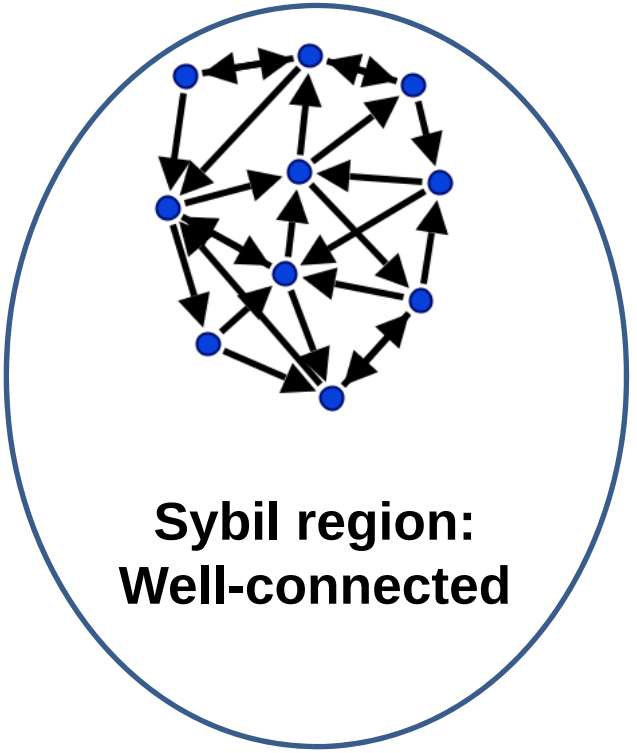
Attack model

- Sybil Attack
  - Comes from a novel, where the heroine “Sybil” has multiple personalities
  - One adversary has multiple false identities (microblog accounts) in the attack

Attack model



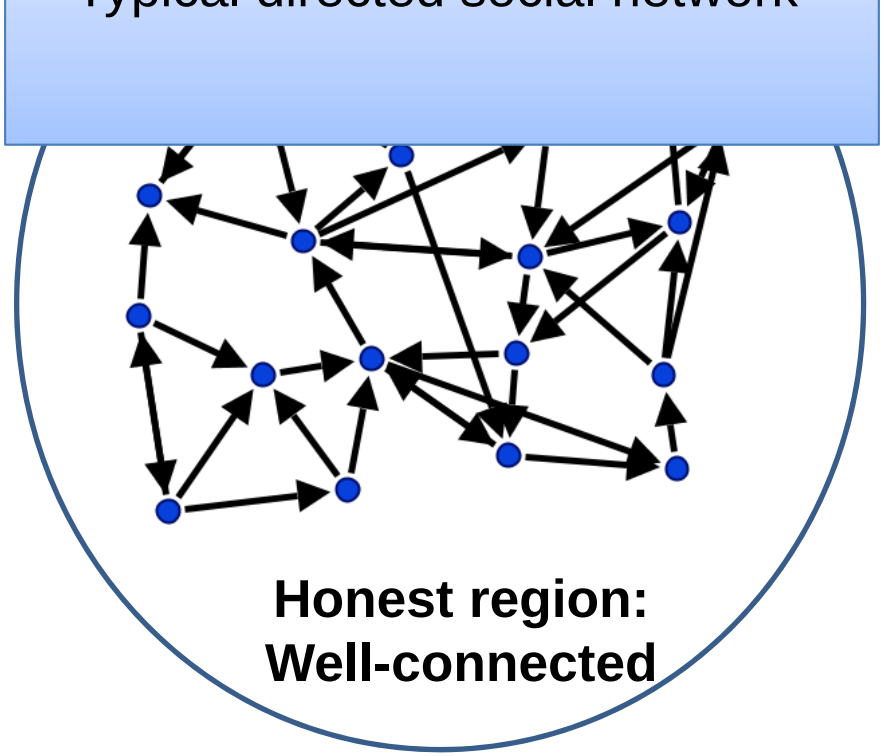
Honest region:
Well-connected

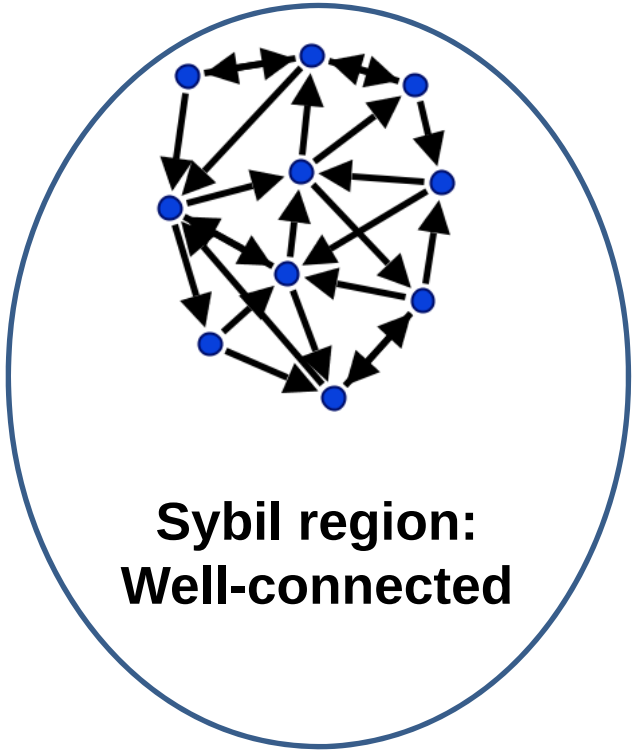


Sybil region:
Well-connected

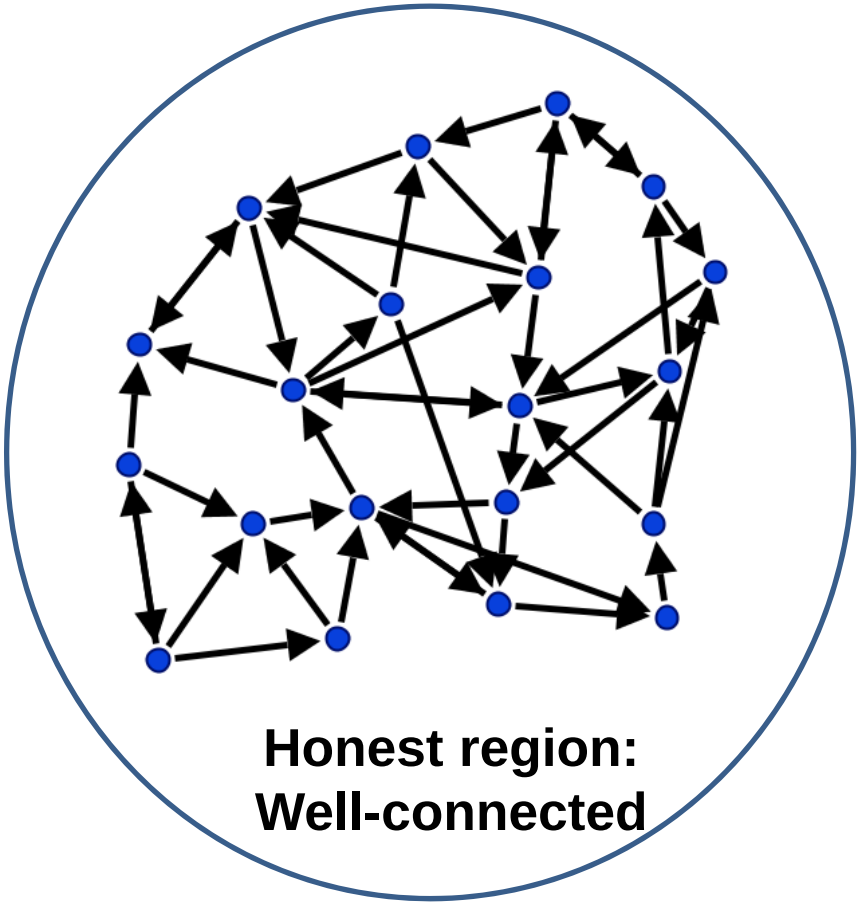
Attack model

Typical directed social network



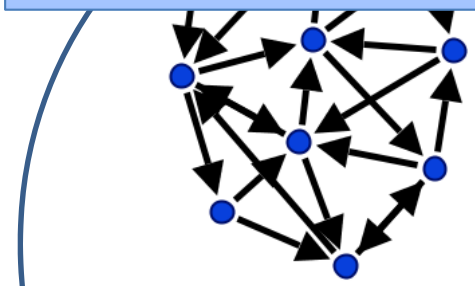


Attack model



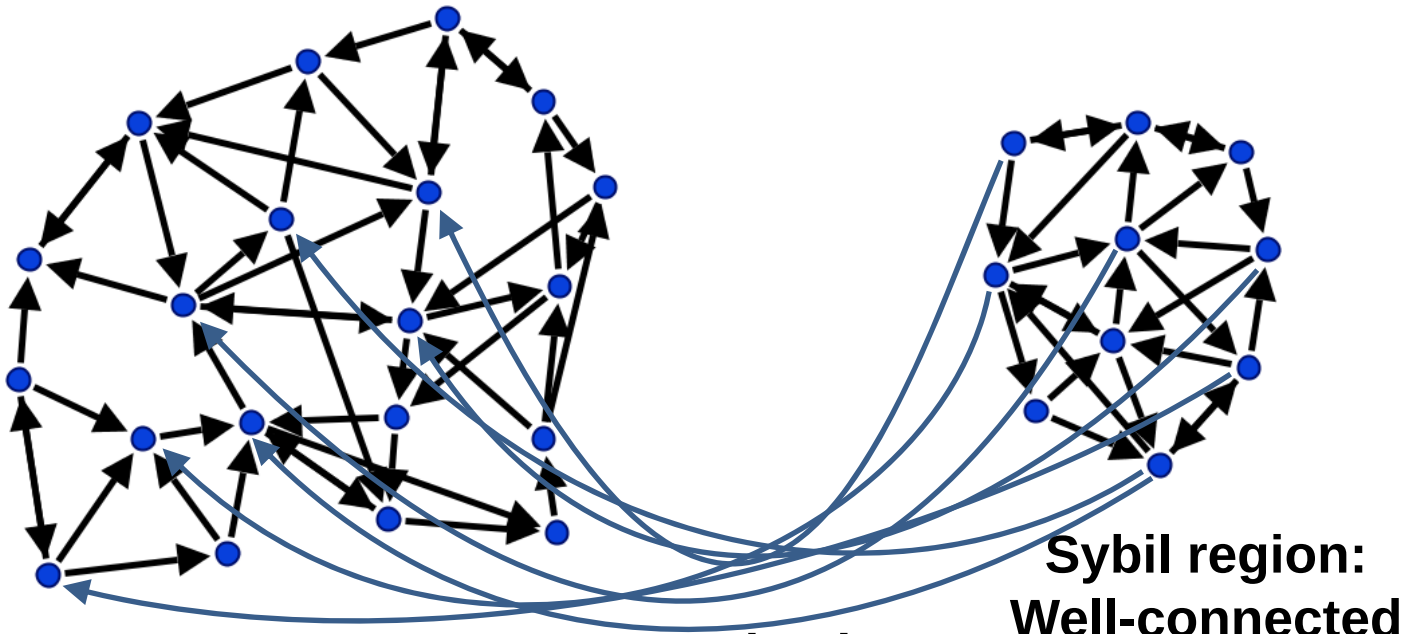
Honest region:
Well-connected

In order to improve their rankings



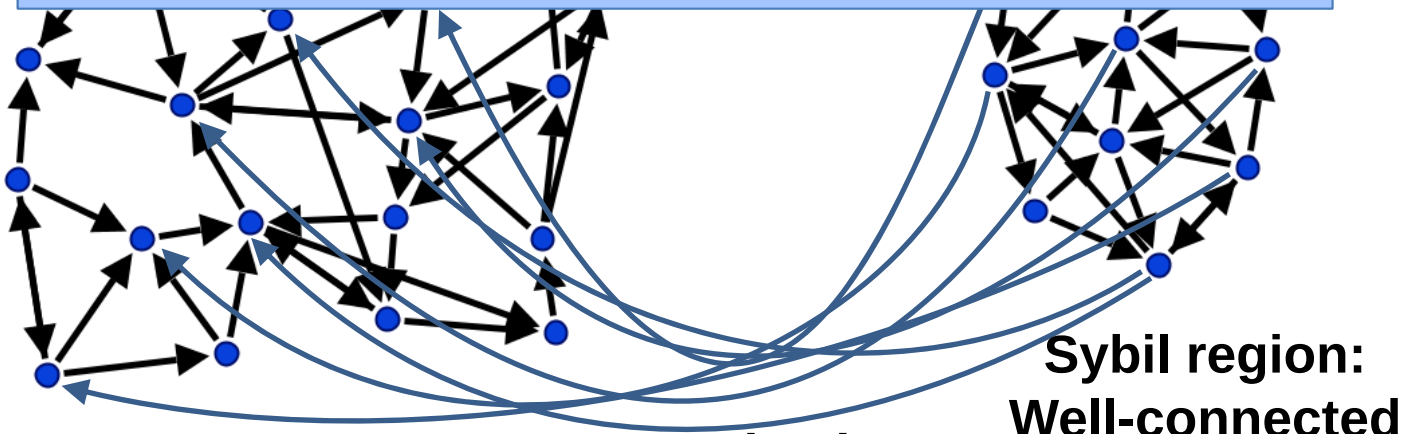
Sybil region:
Well-connected

Attack model



Attack model

Attack edges are arbitrarily selected by attackers

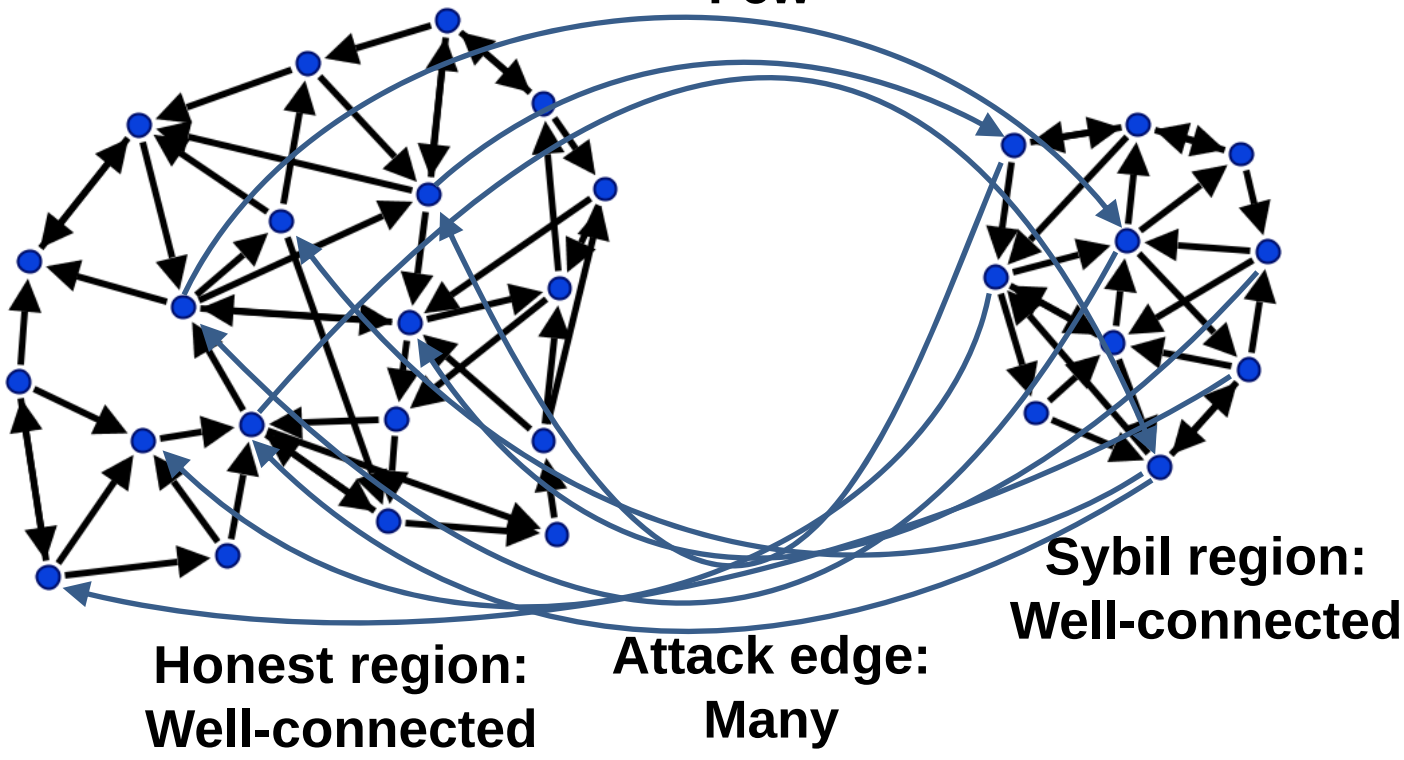


Honest region: Well-connected

Attack edge: Many

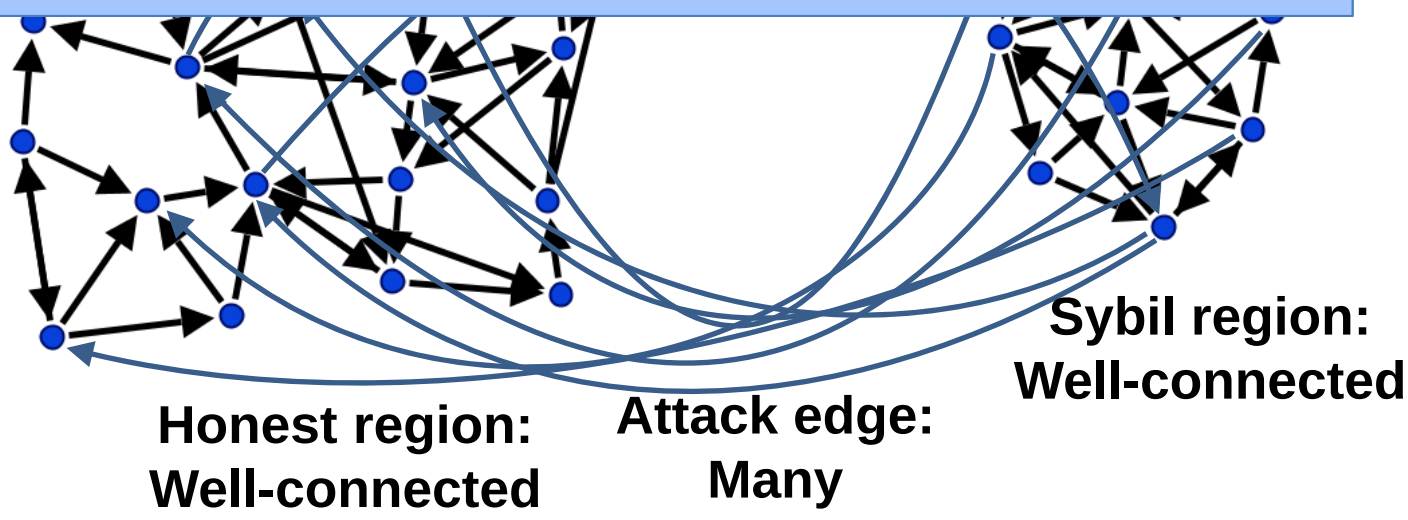
Sybil region: Well-connected

Attack model



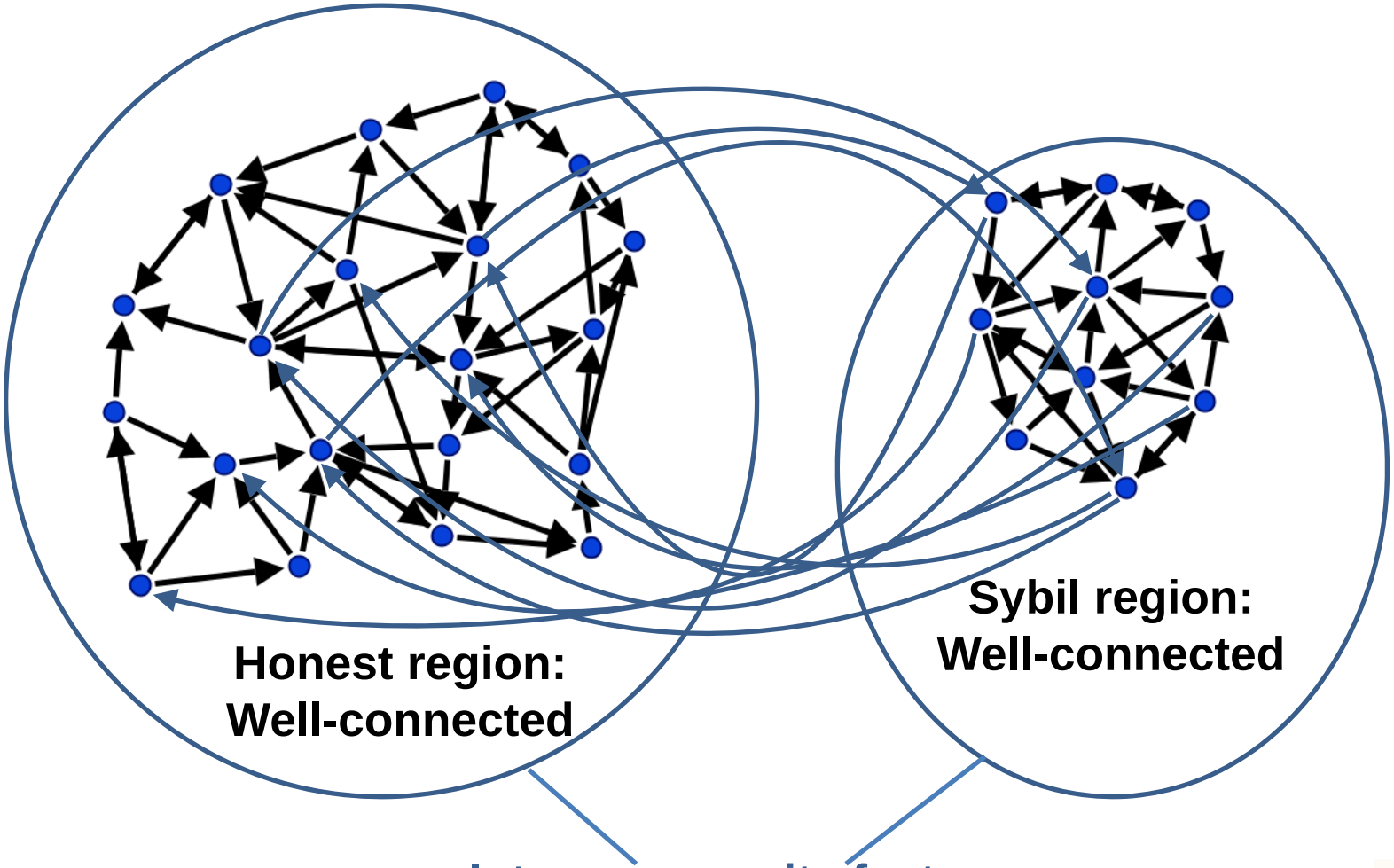
Attack model

A study of spammers on Twitter* shows that compromised Edge/Attack Edge ratio is about 10%.

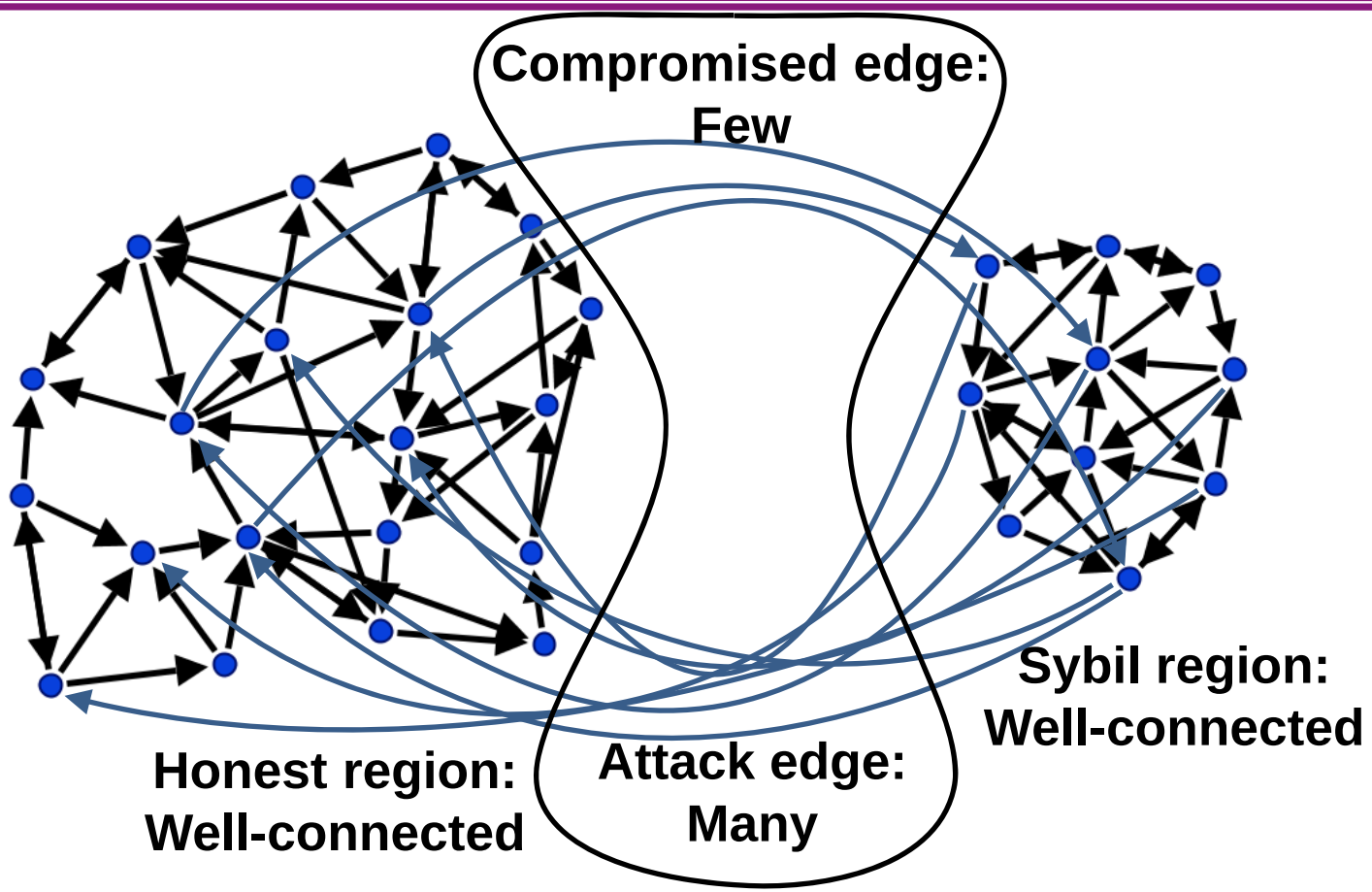


*C. Yang, R. Harkreader, et.al, “Analyzing Spammers Social Networks for Fun and Profit: A Case Study of Cyber Criminal Ecosystem on Twitter,” in *WWW*, 2012, pp. 71-80.

Attack model: analysis

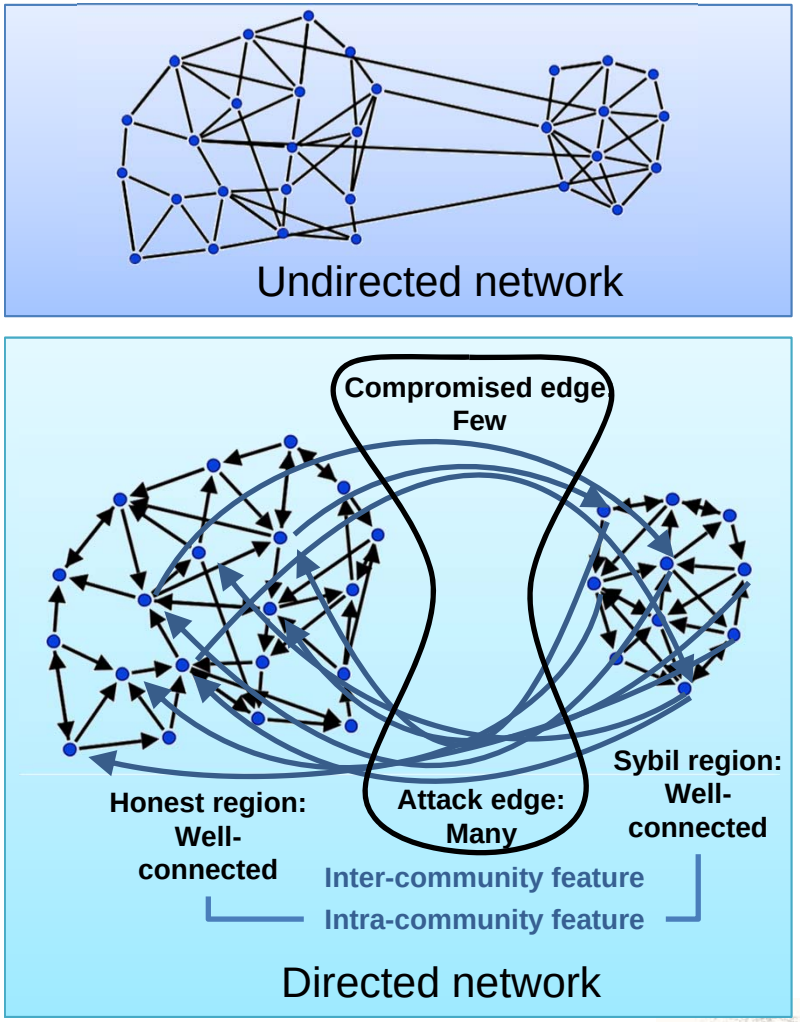


Attack model: analysis



Attack model: analysis

- Intra-community feature
  - Similar for directed and undirected networks
- Inter-community feature
  - Different
  - The key for the detection of Sybil nodes

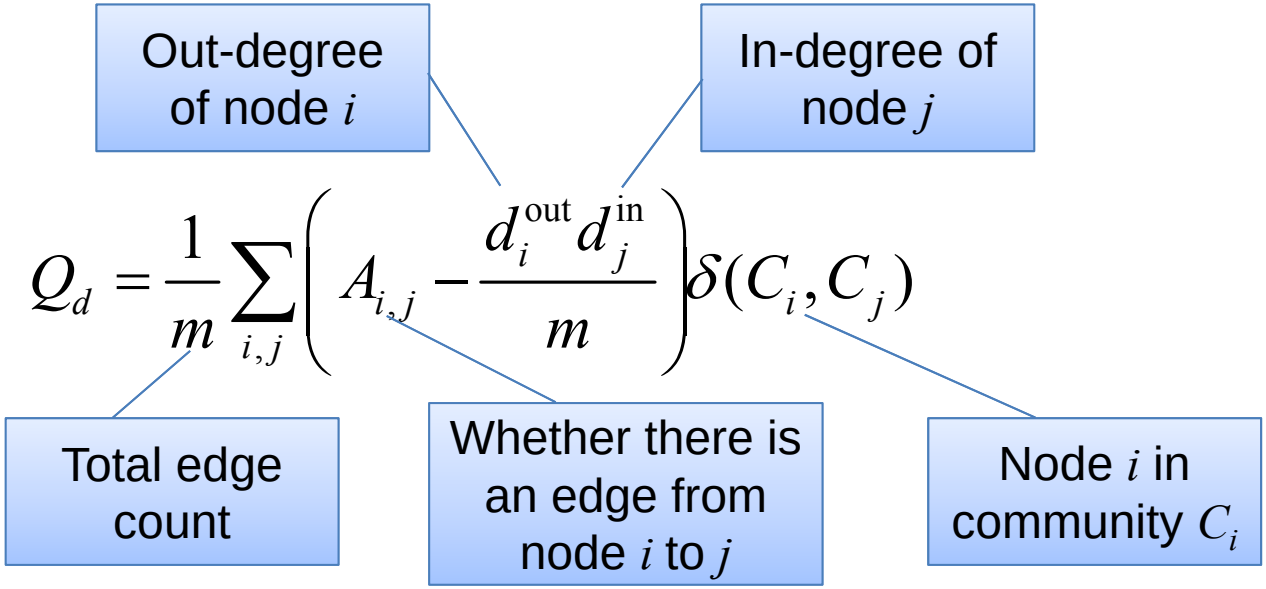


Outline

- Introduction
- Model
- Proposed method
- Experiment results
- Conclusion

How to evaluate a network partition?

- Modularity is widely used, but not enough for solving the problem
  - Modularity for directed networks is defined as*,



*E. A. Leicht, and M. E. J. Newman, "Community structure in directed networks," *Physical Review Letters*, vol. 100, no. 11, 2008.

How to evaluate a network partition?

- Proposed set of measures for directed networks partition evaluation

$$Q = \sum_{\substack{i,j \\ \text{node } i,j \text{ in} \\ \text{the same community}}} Q_{ij} - \sum_{\substack{i,j \\ \text{node } i,j \text{ in} \\ \text{different communities}}} Q_{ij},$$

Q_{ij} measures the confidence that node i and j are in the same community

How to evaluate a network partition?

- Proposed set of measures for directed networks partition evaluation

$$Q = \sum_{\substack{i,j \\ \text{node } i,j \text{ in} \\ \text{the same community}}} Q_{ij} - \sum_{\substack{i,j \\ \text{node } i,j \text{ in} \\ \text{different communities}}} Q_{ij},$$

Those with high confidence should be in the same community

Those with low confidence should be in different communities

How to evaluate a network partition?

- Proposed set of measures for directed networks partition evaluation

$$Q = \sum_{\substack{i,j \\ \text{node } i,j \text{ in} \\ \text{the same community}}} Q_{ij} - \sum_{\substack{i,j \\ \text{node } i,j \text{ in} \\ \text{different communities}}} Q_{ij},$$

where

$$Q_{ij} = \begin{cases} F_{ij}, & i, j \text{ connected;} \\ G_{ij}, & i, j \text{ not connected,} \end{cases}$$

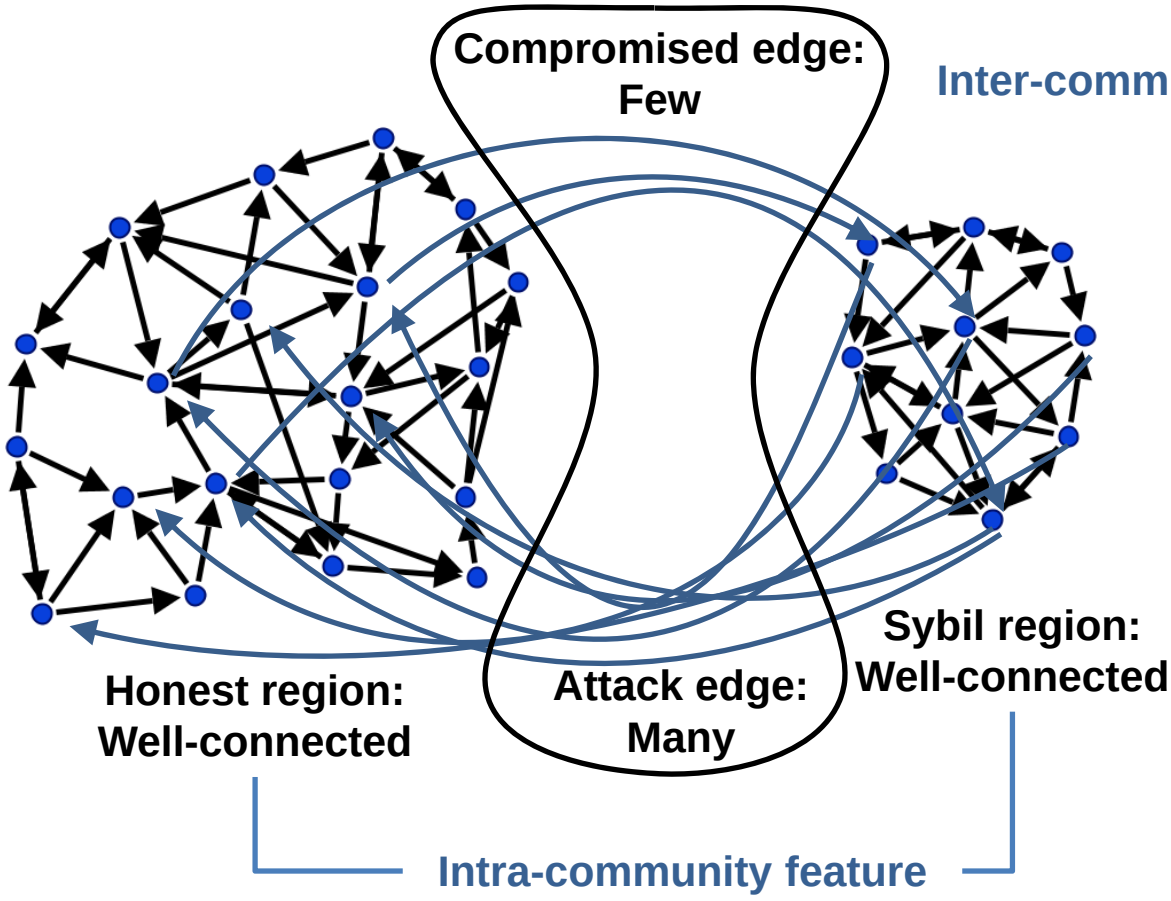
- With arbitrary selection of F_{ij} and G_{ij}, various properties of the network can be measured
- Modularity is a special case of the set of measures

How to optimize Q ?

- Take 2-community partition as example
- Select an initial state
- Iteration
  - For each node, calculate the increment of Q when moving the node to the other community
  - Get the max gain for the previous step, and move the corresponding node to the other community
- Stop when the max gain is no larger than 0

How to choose measure functions?

- Recall: intra-community feature and inter-community feature



How to choose measure functions?

- Intra-community feature
  - Modularity
- Inter-community feature
  - Edge balance ratio*
- Overall
  - Selecting

$$F_{ij} = 1 - \frac{d_i^{\text{out}} d_j^{\text{in}}}{m} + \lambda \log \frac{d_j^{\text{in}}}{d_i^{\text{out}}}$$

  - and

$$G_{ij} = -\frac{d_i^{\text{out}} d_j^{\text{in}}}{m}$$

*X. Wang, Z. Chen, P. Liu, and Y. Gu, “Edge balance ratio: Power law from vertices to edges in directed complex network”, *IEEE Journal of Selected Topics in Signal Processing*, vol.7, no.2, pp. 184-194, 2013.

How to choose measure functions?

- Intra-community feature
  - Modularity
- Inter-community feature
  - Edge balance ratio
- Overall
  - Selecting
  - and

$$F_{ij} = 1 - \frac{d_i^{\text{out}} d_j^{\text{in}}}{m} + \lambda \log \frac{d_j^{\text{in}}}{d_i^{\text{out}}}$$

$$G_{ij} = -\frac{d_i^{\text{out}} d_j^{\text{in}}}{m}$$

In-degree of node j

Out-degree of node i

Total edge count

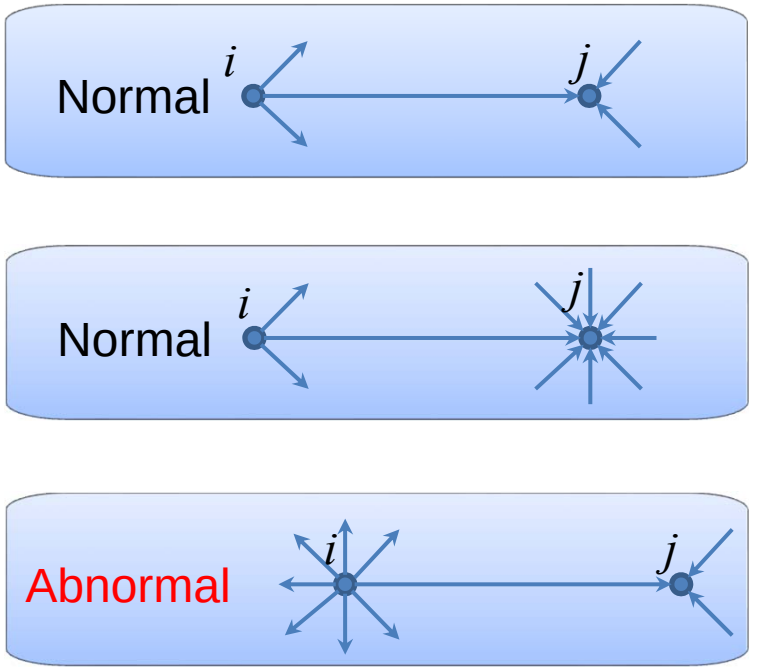
How to choose measure functions?

- Intra-community feature
  - Modularity
- Inter-community feature
  - Edge balance ratio
- Overall
  - Selecting

$$F_{ij} = 1 - \frac{d_i^{\text{out}} d_j^{\text{in}}}{m} + \lambda \log \frac{d_j^{\text{in}}}{d_i^{\text{out}}}$$

  - and

$$G_{ij} = -\frac{d_i^{\text{out}} d_j^{\text{in}}}{m}$$



Might be
Sybil node

*X. Wang, Z. Chen, P. Liu, and Y. Gu, “Edge balance ratio: Power law from vertices to edges in directed complex network”, *IEEE Journal of Selected Topics in Signal Processing*, vol.7, no.2, pp. 184-194, 2013.

How to choose measure functions?

- Intra-community feature
  - Modularity
- Inter-community feature
  - Edge balance ratio
- Overall
  - Selecting

$$F_{ij} = 1 - \frac{d_i^{\text{out}} d_j^{\text{in}}}{m} + \lambda \log \frac{d_j^{\text{in}}}{d_i^{\text{out}}}$$

  - and

$$G_{ij} = -\frac{d_i^{\text{out}} d_j^{\text{in}}}{m}$$

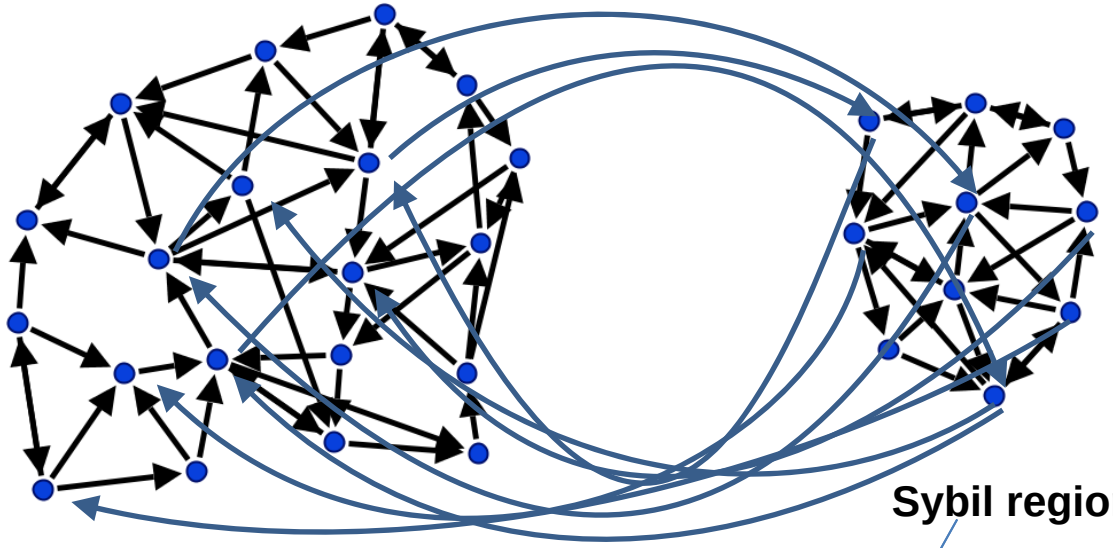
regularization factor

Outline

- Introduction
- Model
- Proposed method
- Experiment results*
- Conclusion

*Codes can be found at
http://gu.ee.tsinghua.edu.cn/publications#sybil

Experiment setup

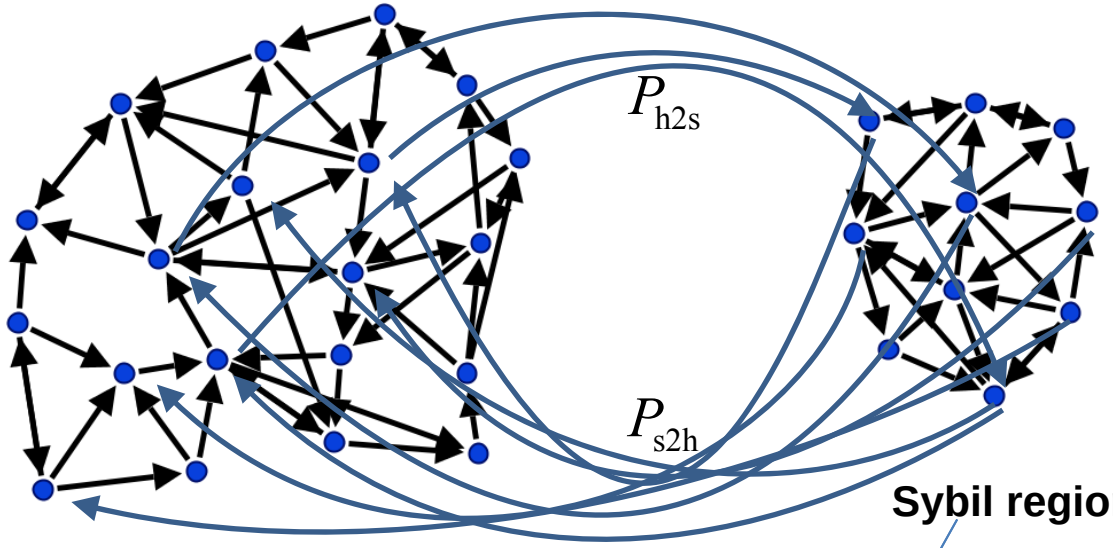


10,000 top users of Sina Weibo and 699,236 edges among them

Constructed using ER model with 1,000 nodes and 10,000 edges

- ***To get ground truth***

Experiment setup



10,000 top users of Sina Weibo and 699,236 edges among them

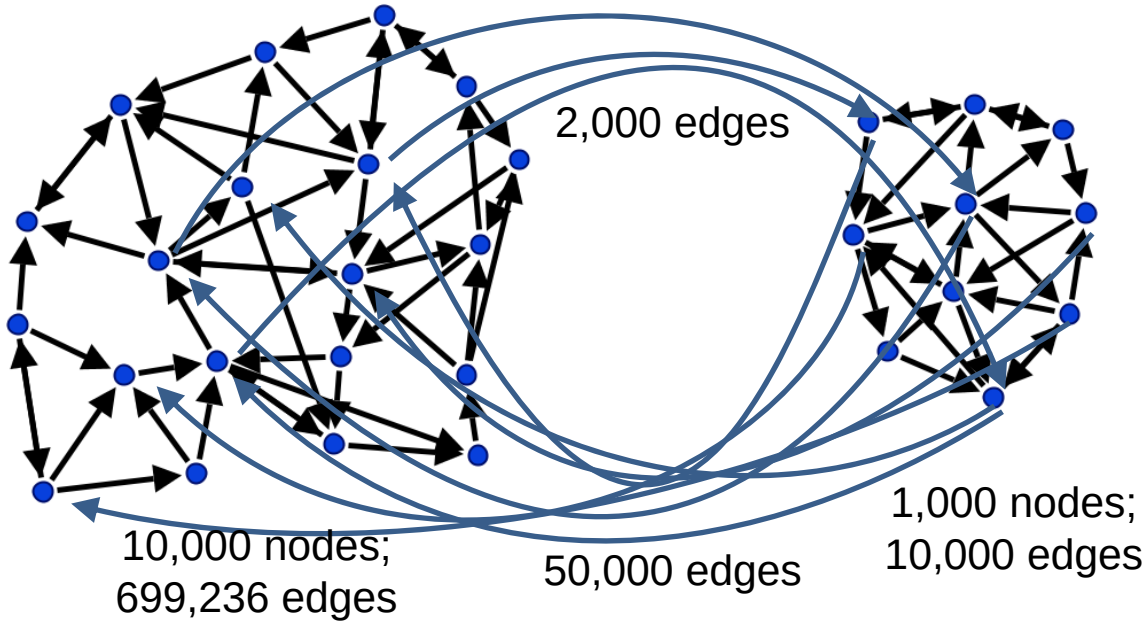
Constructed using ER model with 1,000 nodes and 10,000 edges

- ***To get ground truth***

Basic results

- $P_{\text{h2s}} = 2\times10^{-4}$, $P_{\text{s2h}} = 5\times10^{-3}$

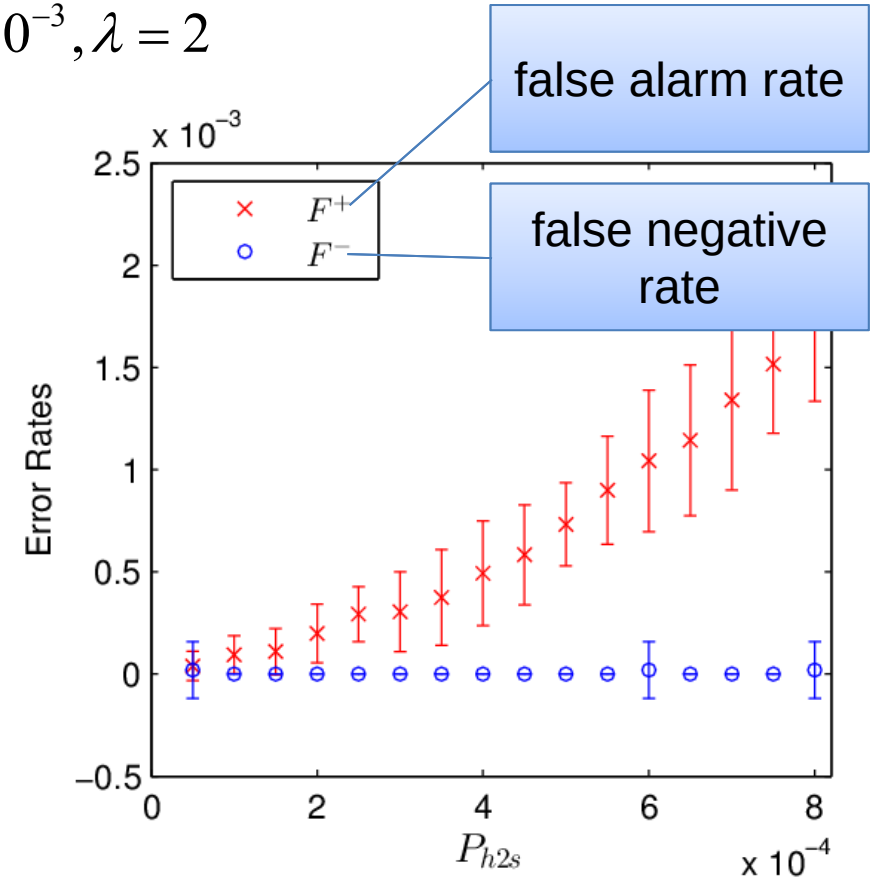
  $\lambda = 2$



- false alarm rate: 0.017%±0.014%
- false negative rate: 0

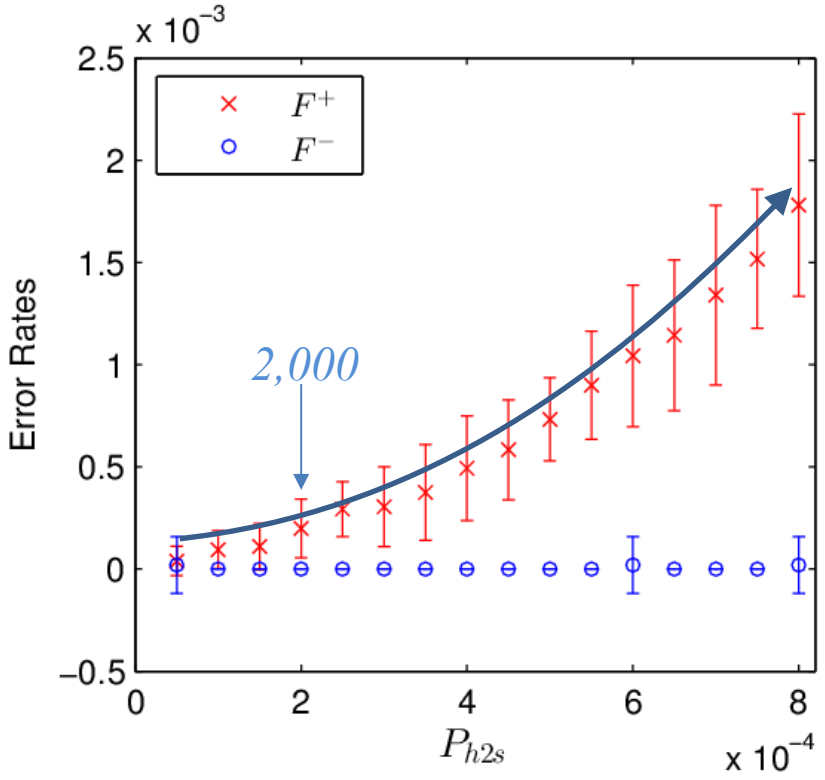
Variation of construction parameters

- Vary P_{h2s}
- $P_{\text{s2h}} = 5 \times 10^{-3}, \lambda = 2$



Variation of construction parameters

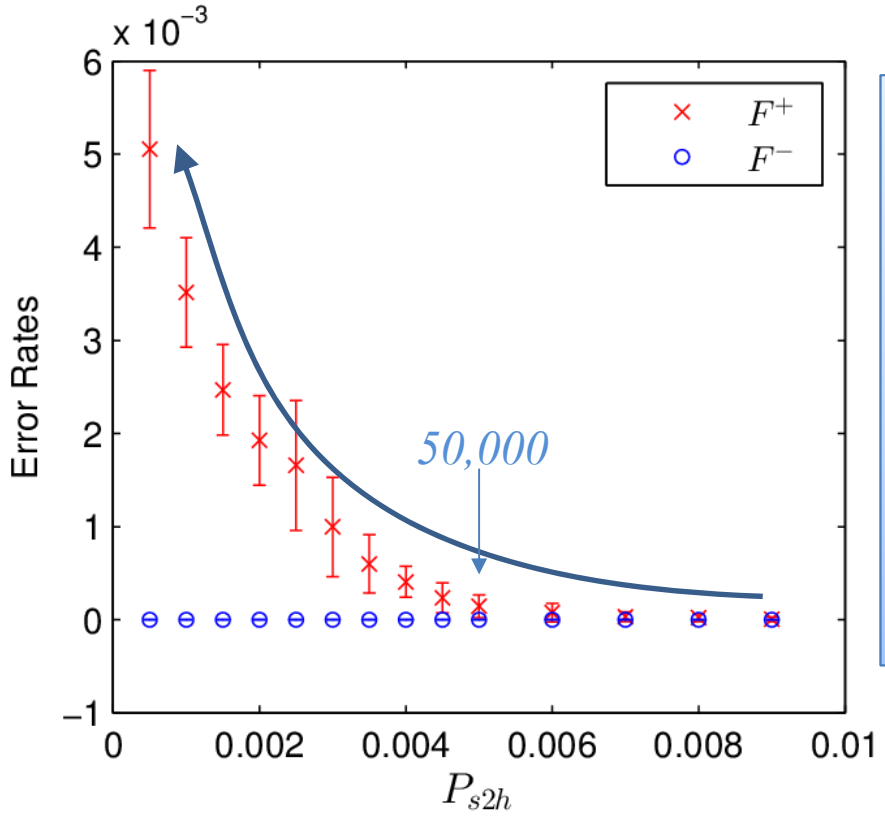
- Vary P_{h2s}
- $P_{\text{s2h}} = 5 \times 10^{-3}, \lambda = 2$ *50,000 attack edges*



More compromised edges, more compromised nodes falsely identified as Sybil nodes

Variation of construction parameters

- Vary P_{s2h}
- $P_{\text{h2s}} = 2\times10^{-4}, \lambda = 2$ *2,000 compromised edges*

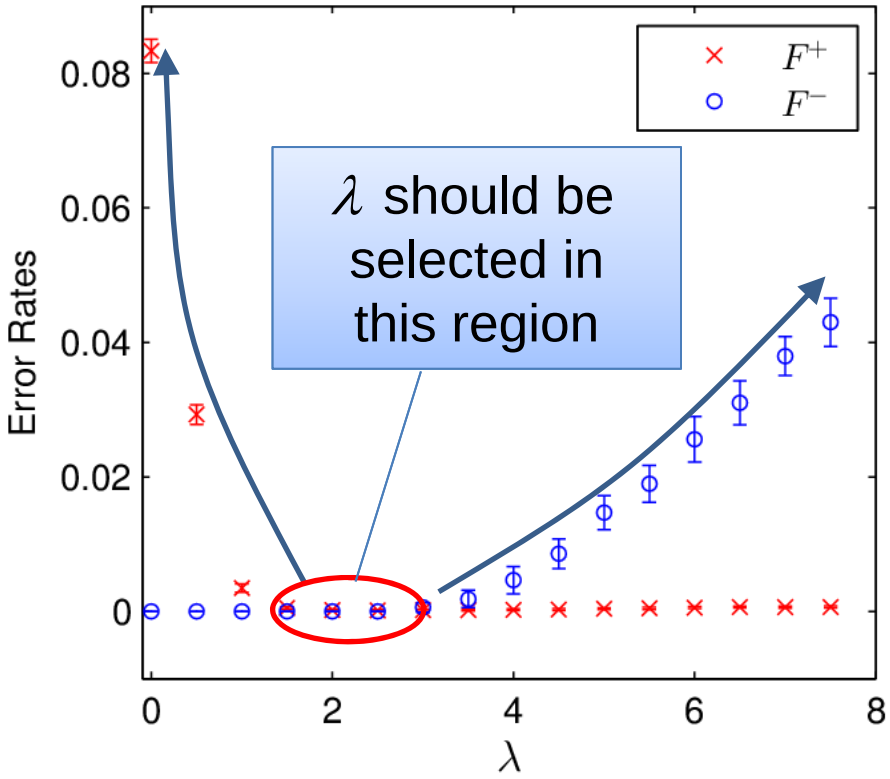


Less attack edges, more compromised nodes falsely identified as Sybil nodes

Variation of regulation parameter

- Vary λ
- $P_{\text{h2s}} = 2\times10^{-4}$, $P_{\text{s2h}} = 5\times10^{-3}$

Intra-community dominating



Inter-community dominating

Comparison with SybilDefender

- No existing schemes specialized in the Sybil detection problem in *directed* social networks
- SybilDefender* is one of the most effective Sybil-defending schemes in *undirected* social networks

P_{s2h}	10^{-5}		$2x10^{-5}$	
Error type	false alarm	false negative	false alarm	false negative
Proposed	0	0	0.002%	0
SybilDefender	2.42%	0	2.70%	8.2%

*W. Wei, X. Fengyuan, C. C. Tan, and Q. Li, “SybilDefender: Defend against sybil attacks in large social networks,” in *INFOCOM*, 2012, pp. 1951-1959.

Outline

- Introduction
- Model
- Proposed method
- Experiment results
- Conclusion

Conclusion

- A set of measures for the evaluation of network partitions of directed networks
- A social relation based method for the defense against Sybil attacks in directed social networks
- Promising results and outperforms the reference algorithm
- Future work
  - Adaptation of regularization factor
  - Mixed method that utilize profile-based, tweet-based and graph-based methods for spammer detection