Aiding the Detection of Fake Accounts in Large Scale Social Online Services

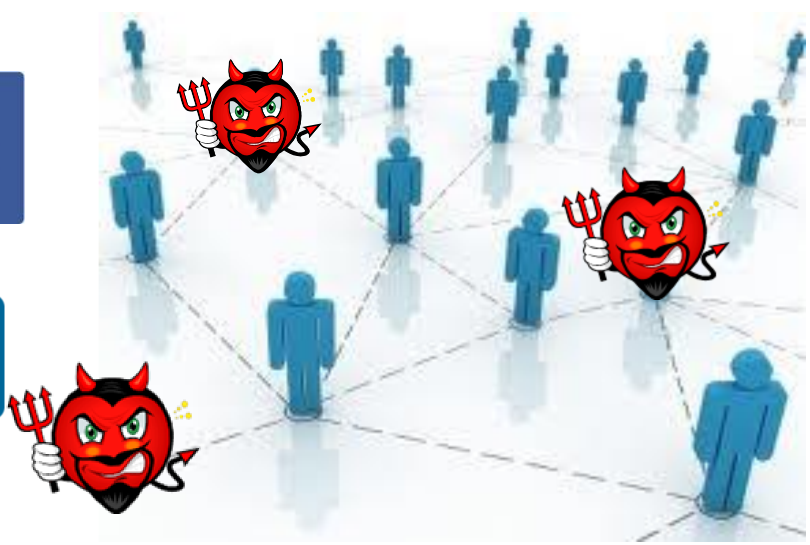
Qiang Cao
Duke University

Michael Sirivianos
Cyprus Univ. of Technology
Telefonica Research

Xiaowei Yang
Duke University

Tiago Pregueiro
Tuenti, Telefonica Digital

Fake accounts (Sybils) in OSNs



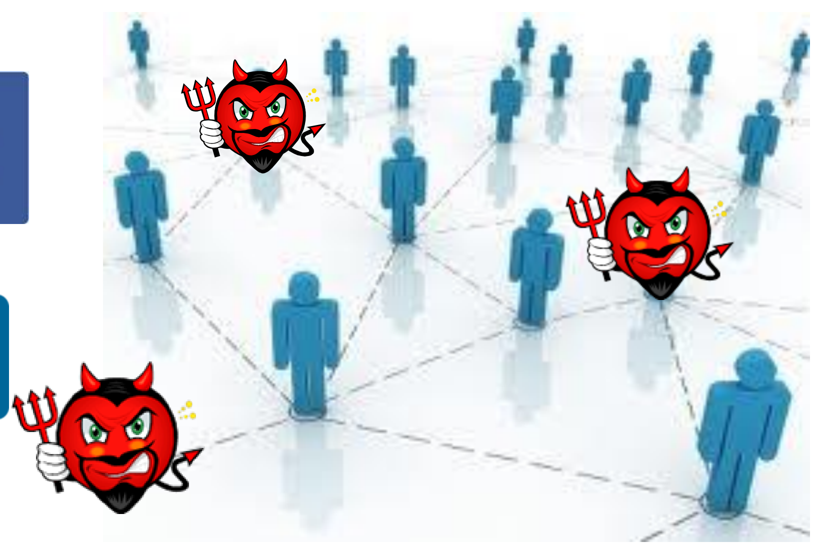
Fake accounts (Sybils) in OSNs

Home / News & Blogs / Friending Facebook

Facebook: 5-6% of accounts are fake

By Emil Protalinski | March 8, 2012, 8:17am PST

Summary: *Facebook estimates somewhere between 42.25 million and 50.70 million Facebook accounts are either false or duplicate. This is the first time the social networking giant has revealed such numbers.*



Fake accounts for sale

HOME PAGE | TODAY'S PAPER | VIDEO | MOST POPULAR | TIMES TOPICS

The New York Times

Internet

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

VeriSign

2010

During several weeks in February, iDefense tracked an effort to sell log-in data for 1.5 million Facebook accounts on several online criminal marketplaces, including one called Carder.su.

freelancer.com

BlackHat World
BlackHat Seo Forum

WICKEDFIRE

Why are fakes harmful?

- **Fake (Sybil) accounts in OSNs can be used to:**
  - **Send spam [IMC'10]**
  - **Manipulate online rating [NSDI'09]**
  - **Access personal user info [S&P'11]**
  - **…**

Why are fakes harmful?

ZDNet

Facebook: 5-6% of accounts are fake

By Emil Protalinski | March 8, 2012, 8:17am PST

“the geographic location of our users is estimated based on a number of factors, such as IP address, which may not always accurately reflect the user's actual location. If advertisers, developers, or investors do not perceive our user metrics to be accurate representations of our user base, or if we discover material inaccuracies in our user metrics, our reputation may be harmed and advertisers and developers may be less willing to allocate their budgets or resources to Facebook, which could negatively affect our business and financial results.”

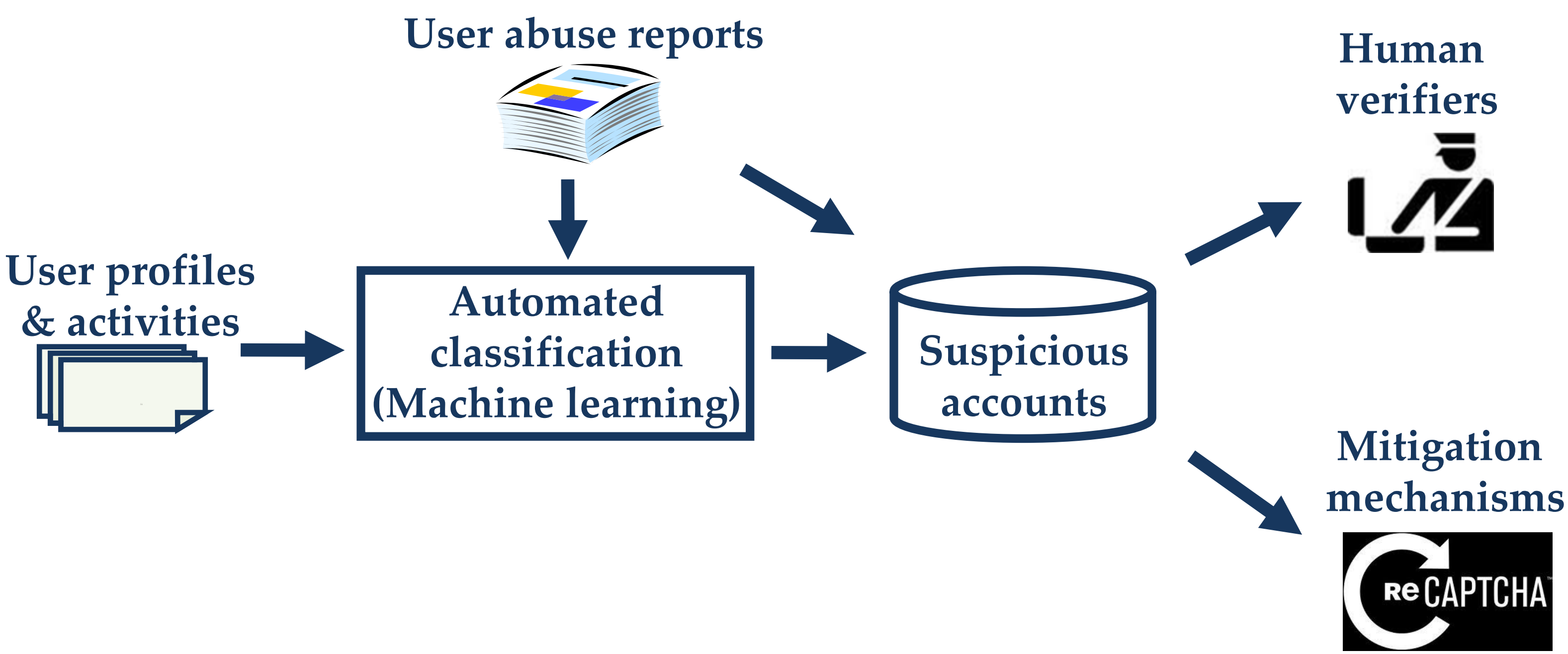
Detecting Sybils is challenging

- **Sybils may resemble real users**

Difficult to automatically detect using profile and activity features

Current practice

- Employs many counter-measures
- False positives are detrimental to user experience
  - Real users respond very negatively



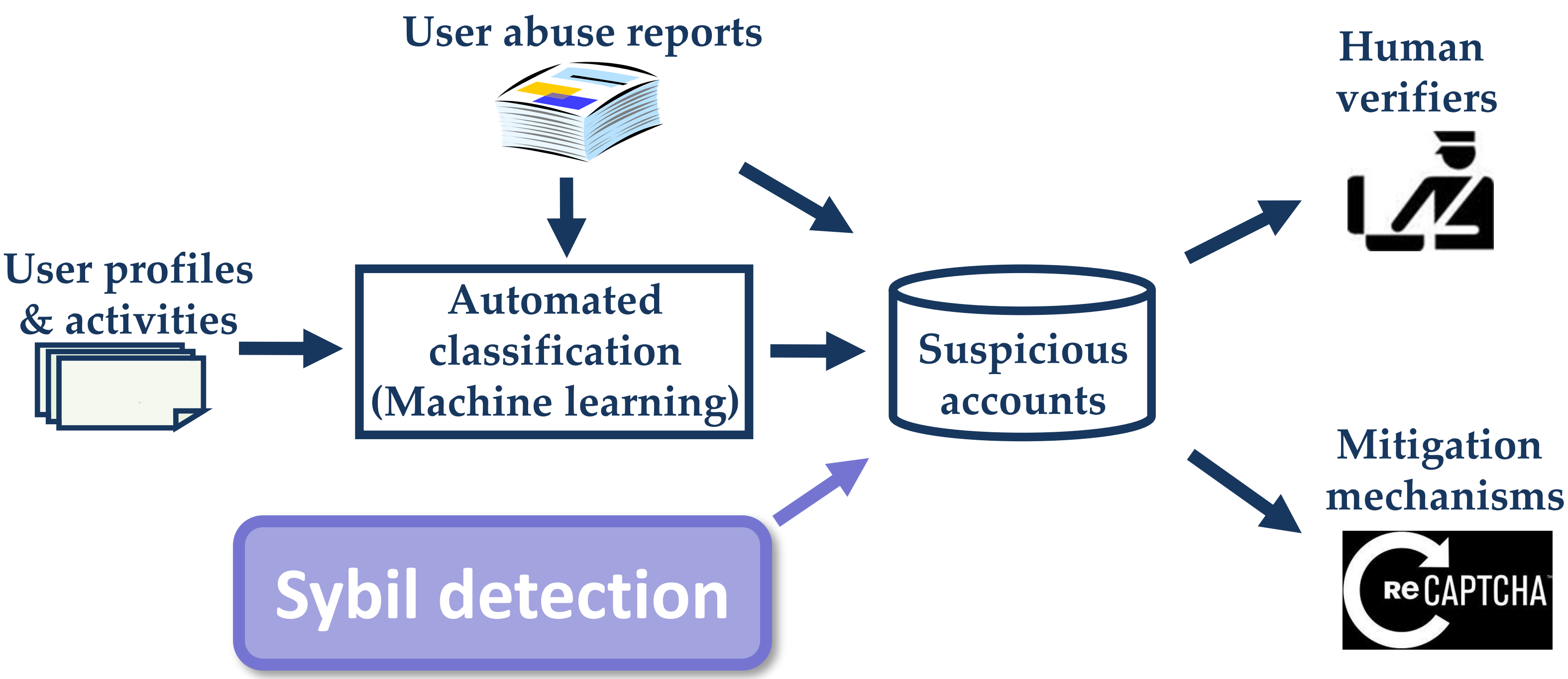
Current practice

- Employs many counter-measures
- False positives are detrimental to user experience
  - Real users respond very negatively
- **Inefficient use of human labor!**

- Tuenti's user inspection team
  - Reviews ~12, 000 abusive profile reports per day
  - An employee reviews ~300 reports per hour
  - Deletes ~100 fake accounts per day

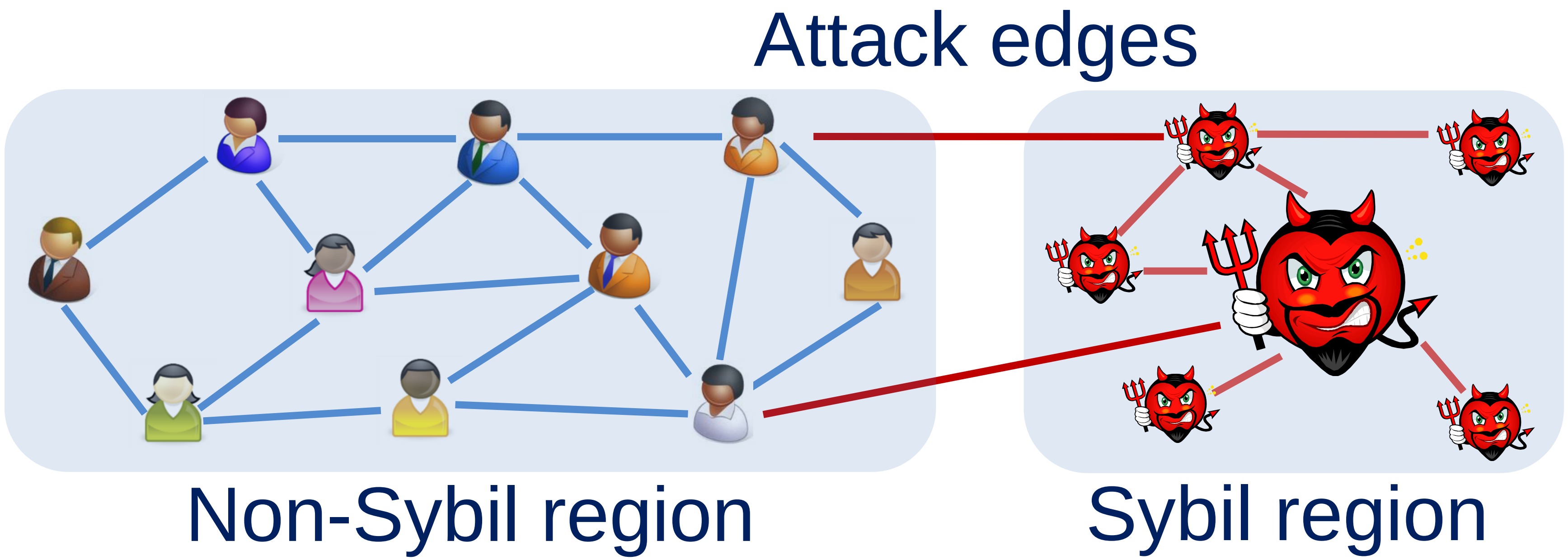
Mitigation mechanisms

Can we improve the workflow?



Leveraging the social relationship

- **The foundation of social-graph-based schemes**
  - **Sybils have limited social links to real users**
- **Can complement current OSN counter-measures**

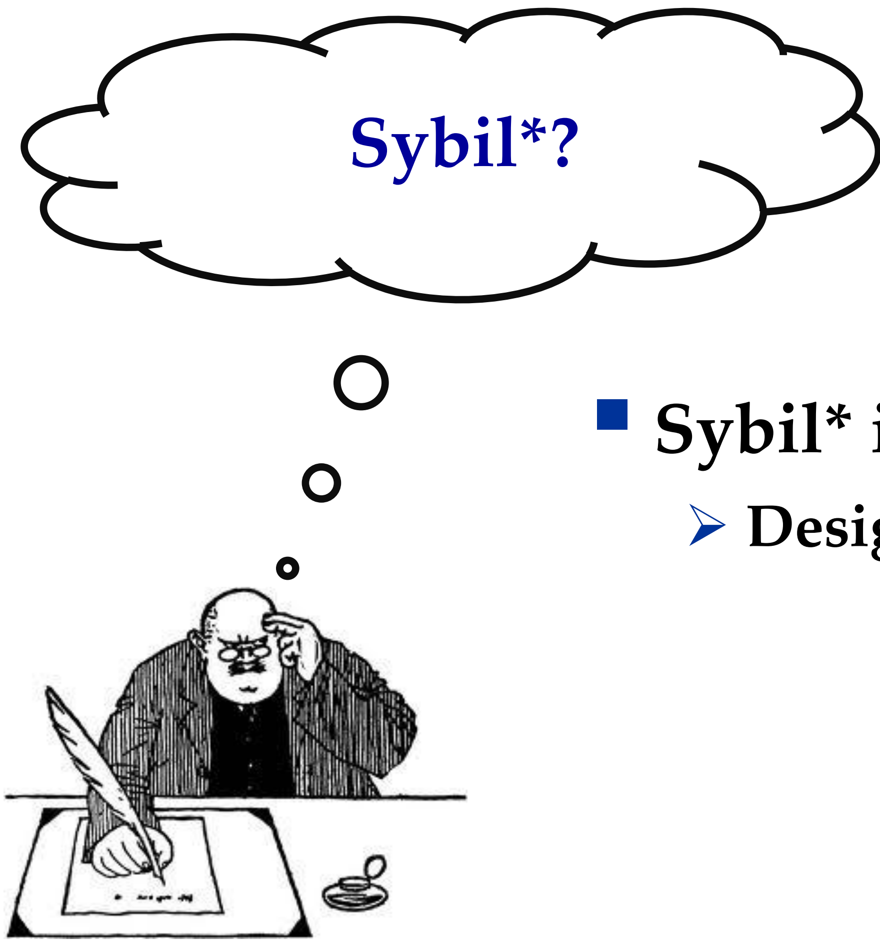


Goals of a practical social-graph-based Sybil defense

- **Effective**
  - **Uncovers fake accounts with high accuracy**

- **Efficient**
  - **Able to process huge online social networks**

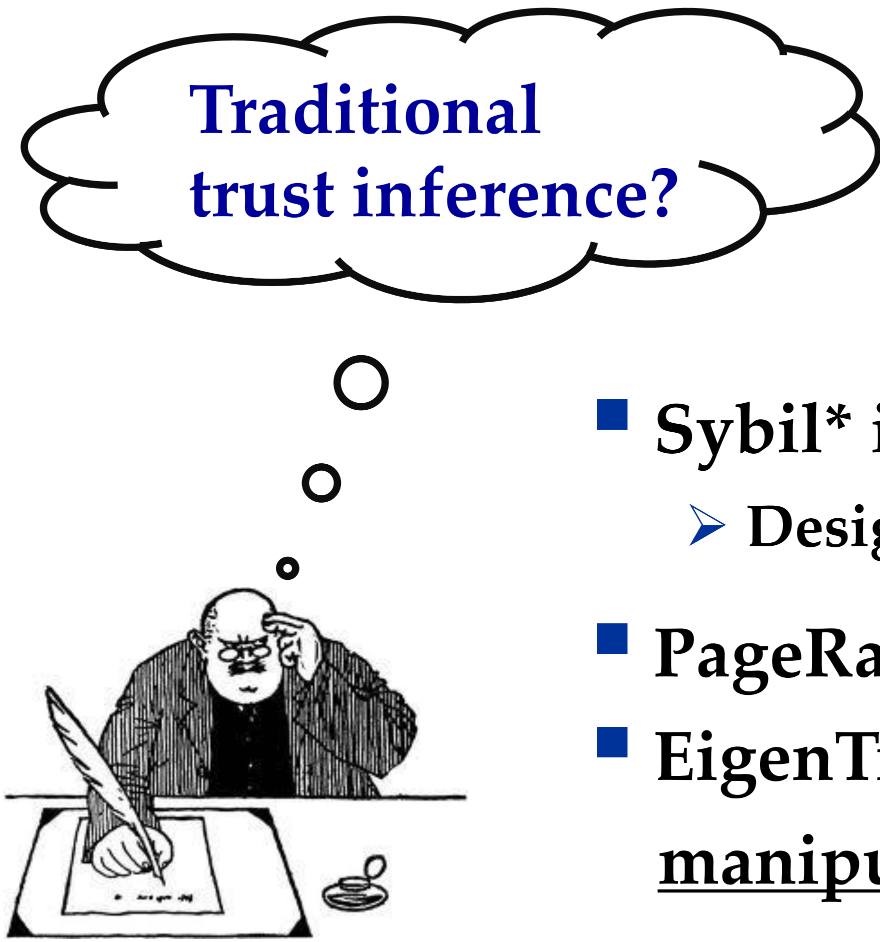
How to build a practical social-graph-based Sybil defense?



SybilGuard [SIGCOMM'06]
SybilLimit [S&P'08]
SybilInfer [NDSS'09]

- **Sybil* is <u>too expensive</u> in OSNs**
  - **Designed for decentralized settings**

How to build a practical social-graph-based Sybil defense?



PageRank [Page et al. 99]
EigenTrust [WWW'03]

- **Sybil* is too expensive in OSNs**
  - **Designed for decentralized settings**
- **PageRank is not Sybil-resilient**
- **EigenTrust is substantially manipulable [NetEcon'06]**

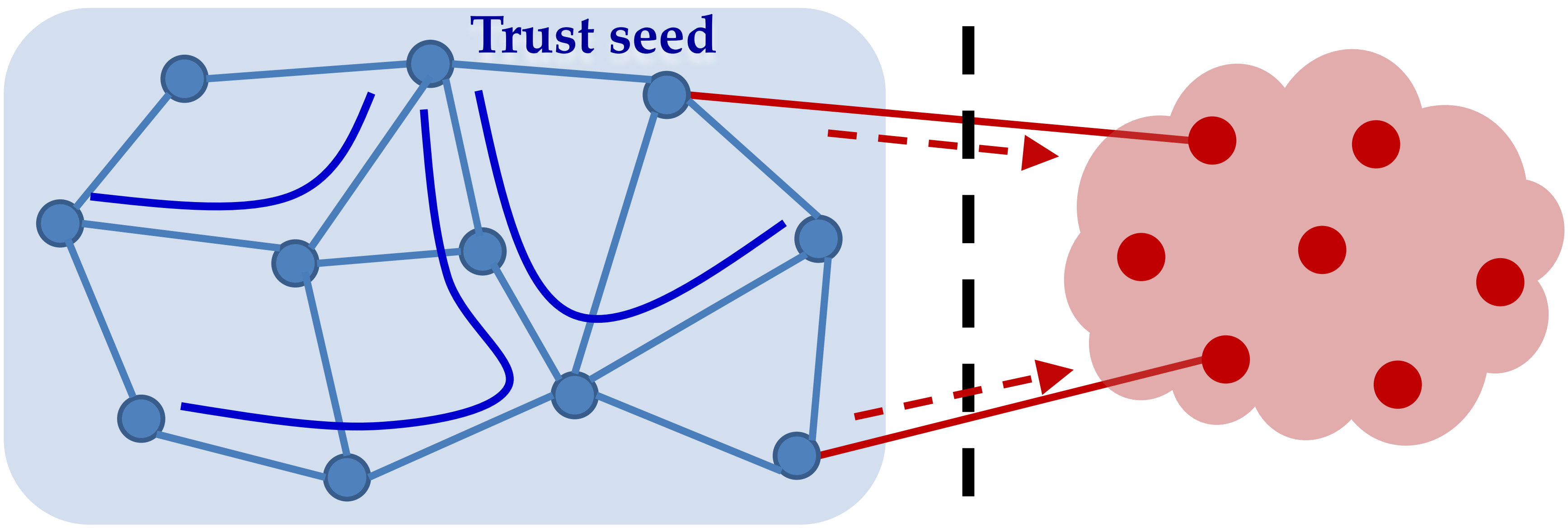
SybilRank in a nutshell

- **Uncovers Sybils by ranking OSN users**
  - **Sybils are ranked towards the bottom**
  - **Based on short random walks**
  - **Uses parallel computing framework**

- **Practical Sybil defense: efficient and effective**
  - **Low computational cost: O(n log n)**
  - **≥20% more accurate than the 2nd best scheme**
  - **Real-world deployment in Tuenti**

Primer on short random walks

- **Short random walks**



Limited probability of escaping to the Sybil region

SybilRank's key insights

- **Main idea**
  - **Ranks by the landing probability of short random walks**

- **Uses power iteration to compute the landing probability**
  - **Iterative matrix multiplication (used by PageRank)**
  - **Much more efficient than random walk sampling (Sybil*)**
    - **O(n log n) computational cost**
  - **As scalable as PageRank**

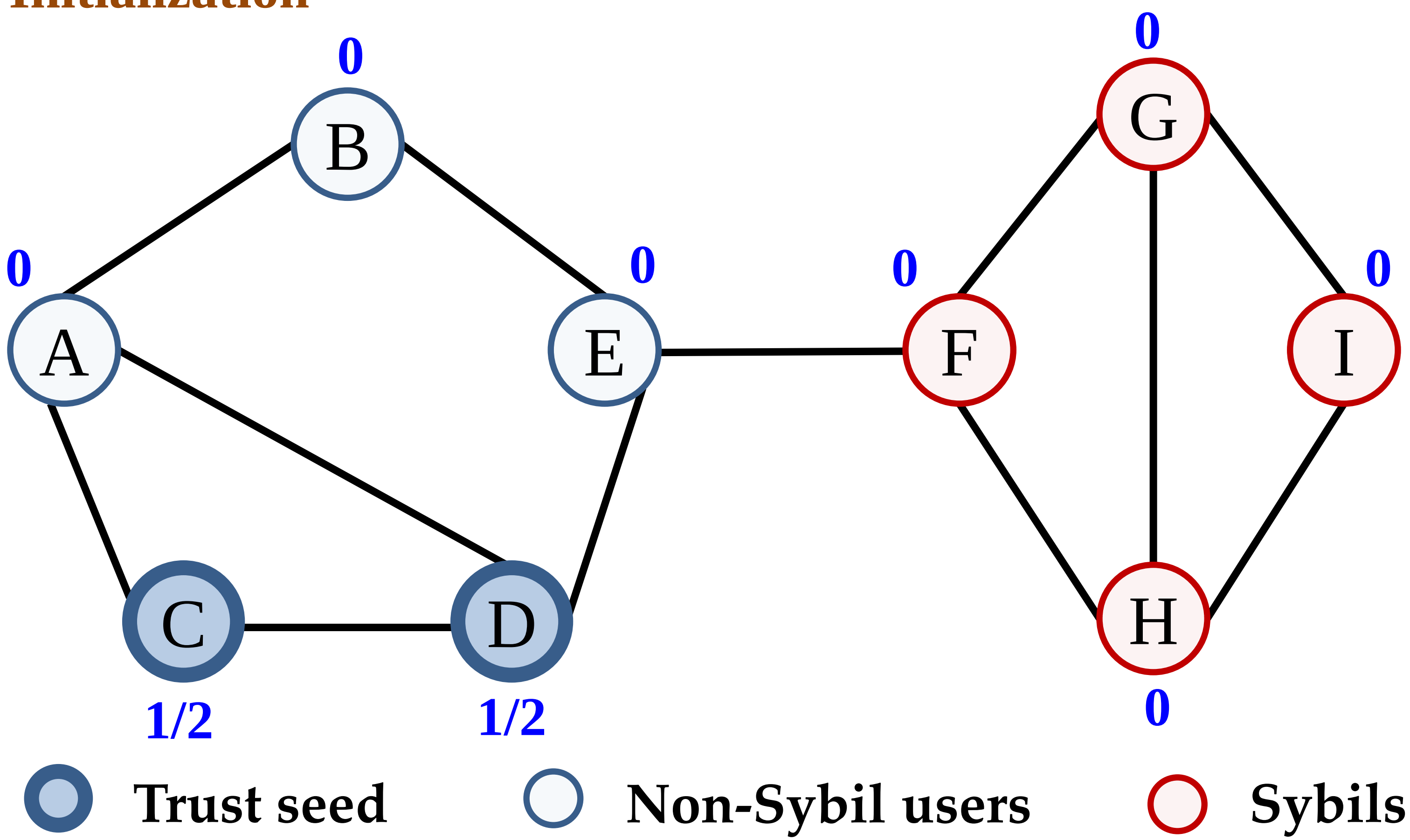
Google

PageRank™

An example

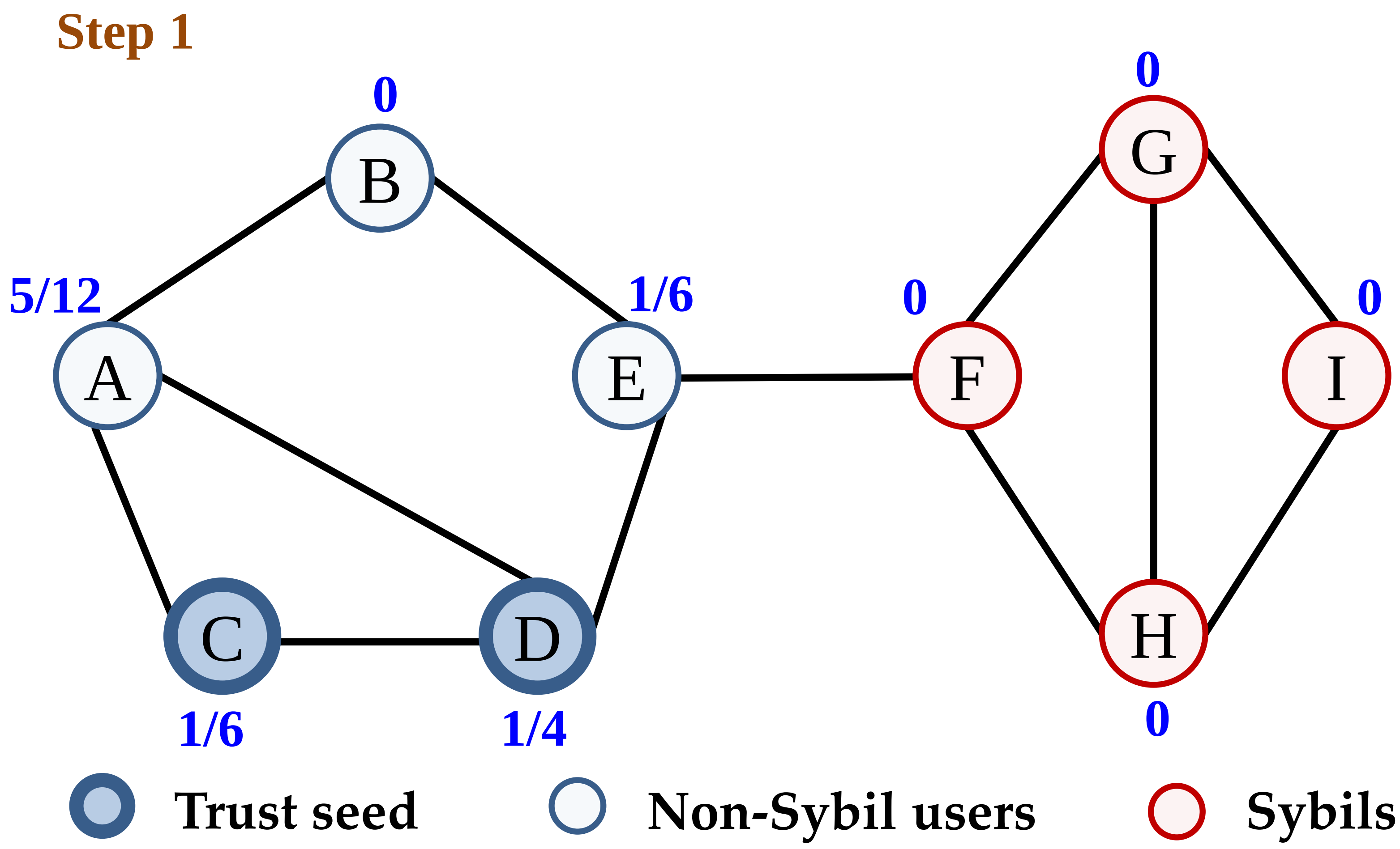
- **Landing probability of short random walks**

Initialization



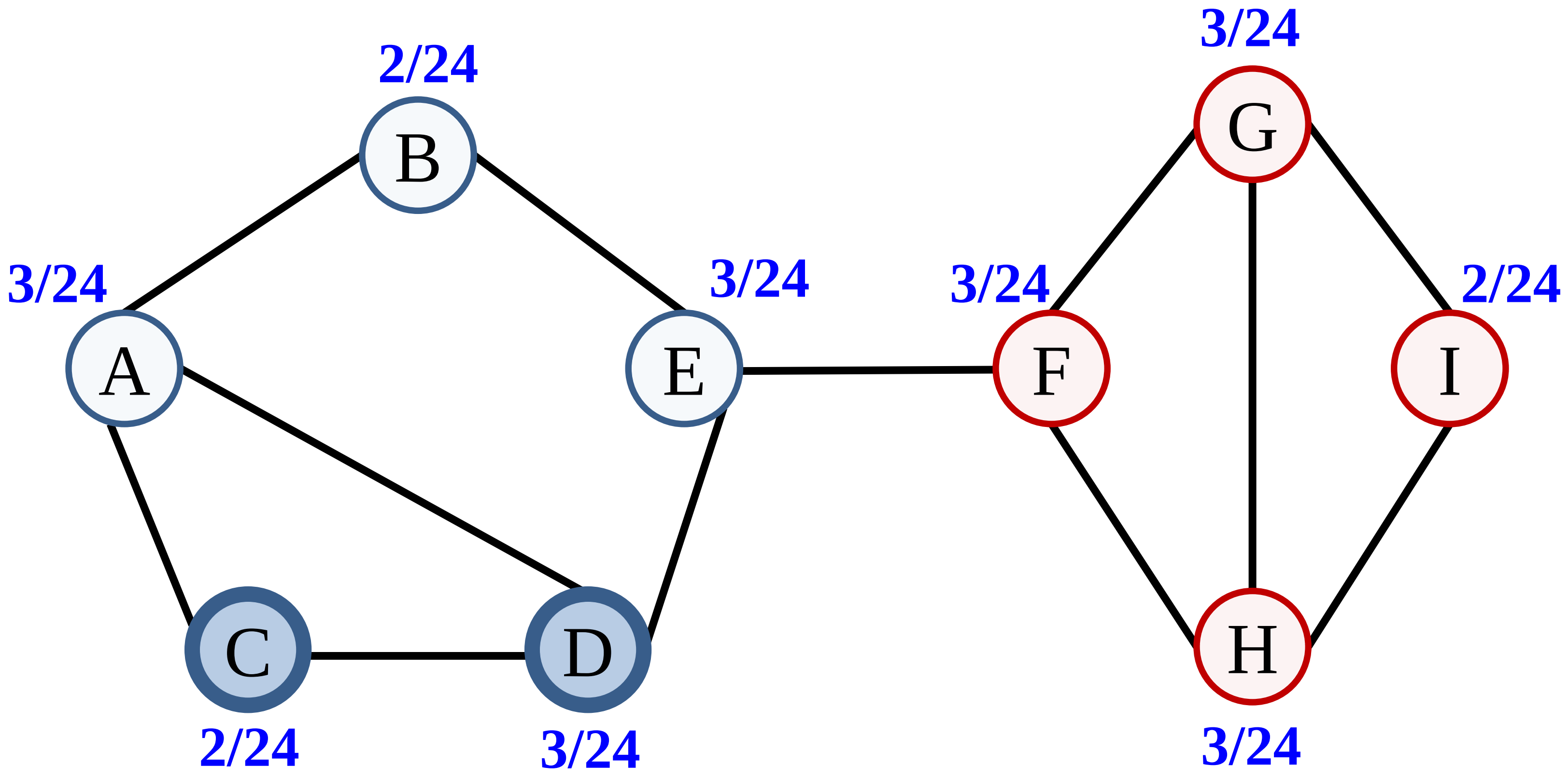
An example

- **Landing probability of short random walks**



An example

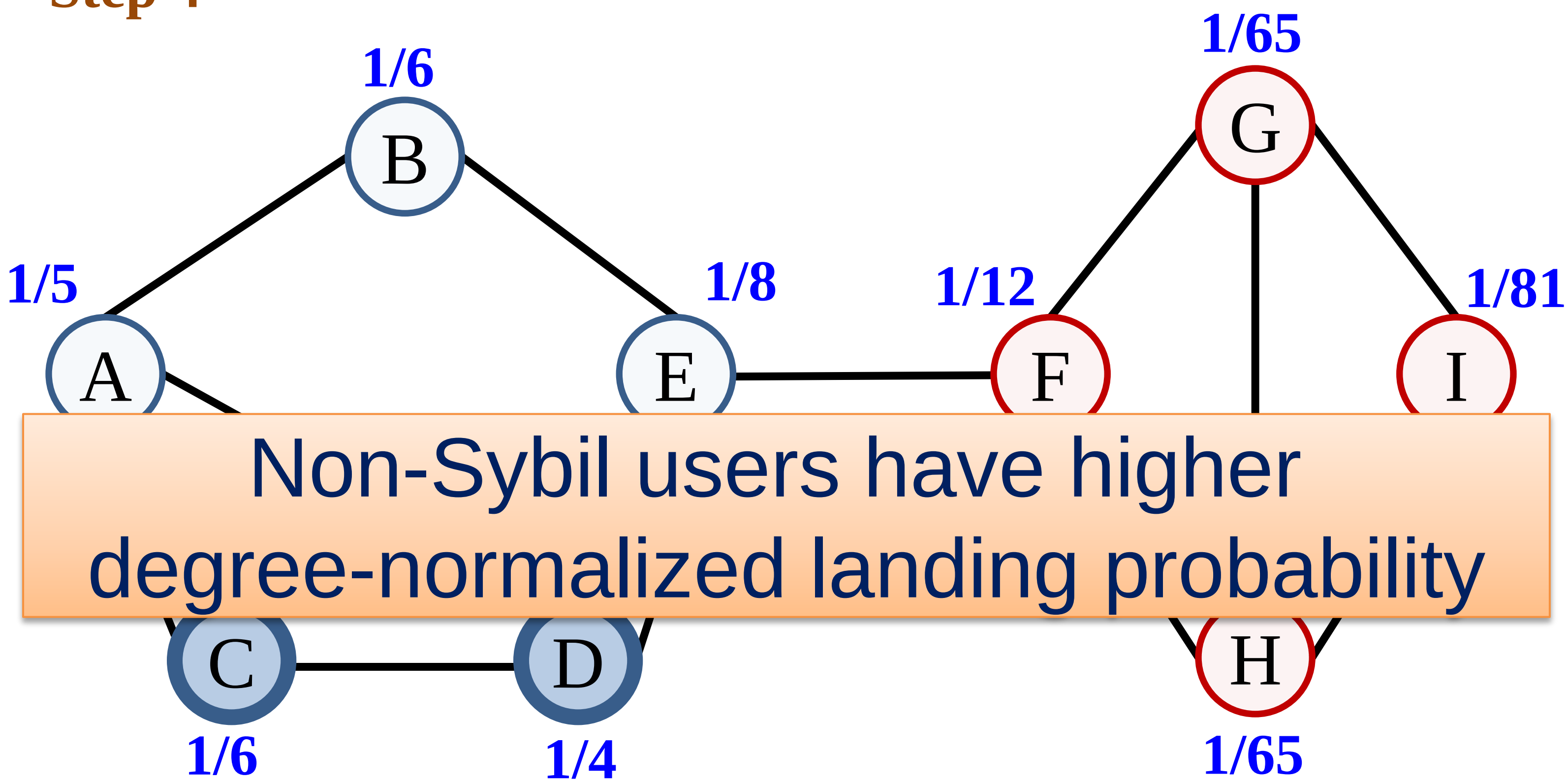
- **Stationary distribution**
  - **Identical degree-normalized landing probability: 1/24**



An example

- ~~Stationary distribution~~ **Early Termination**

Step 4

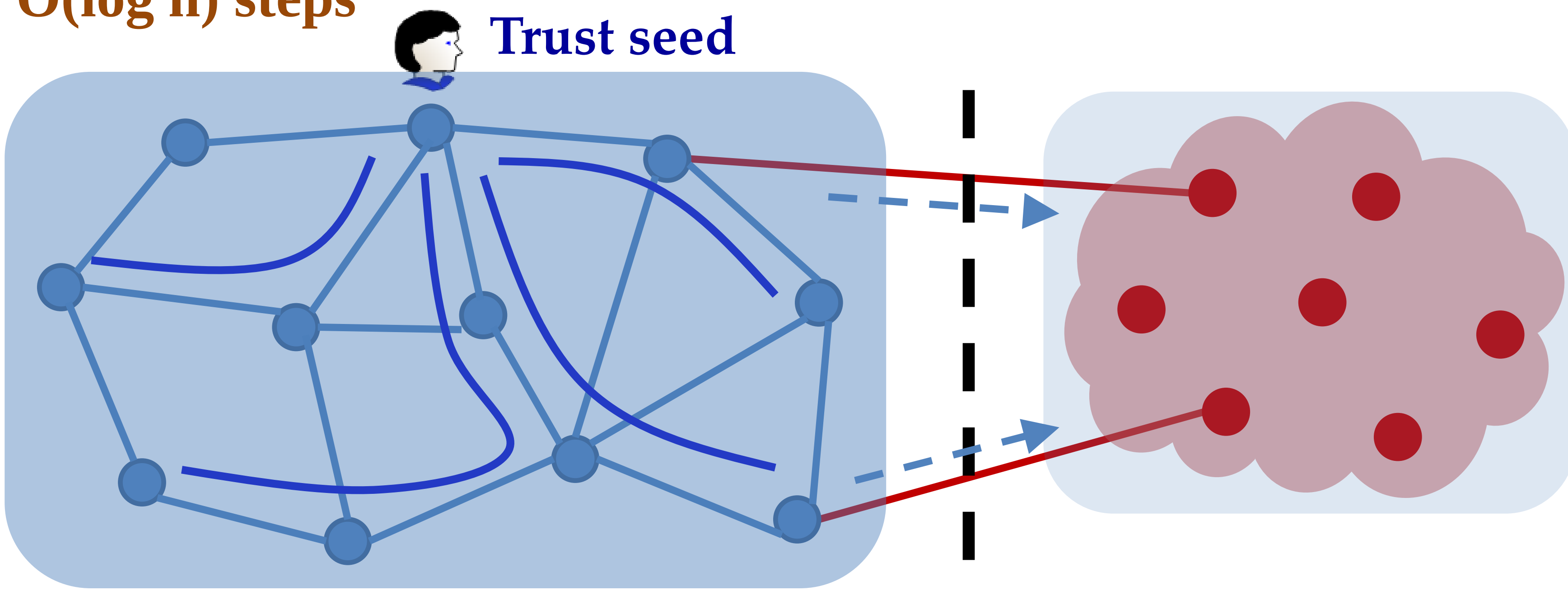


Non-Sybil users have higher degree-normalized landing probability

Rankings B C D A E F I G H

How many steps?

- **O(log n) steps to cover the non-Sybil region**
  - **The non-Sybil region is fast-mixing (well-connected) [S&P'08]**



Stationary distribution approximation

Overview

- **Problem and Motivation**
- **Challenges**
- **Key Insights**
- **Design Details**
- **Evaluation**

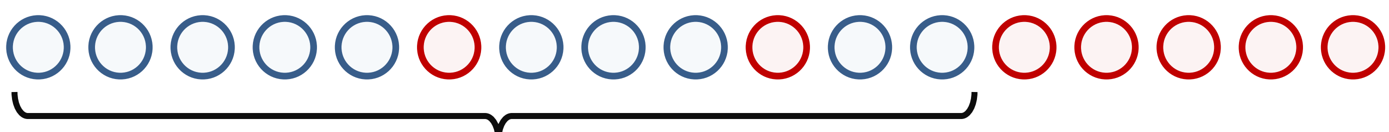
We divide the landing probability by the node degree

- **Eliminates the node degree bias**
  - **False positives: low-degree non-Sybil users**
  - **False negatives: high-degree Sybils**

- **Security guarantee**
  - **Accept O(log n) Sybils per attack edge**

Theorem: When an attacker randomly establishes *g* attack edges in a fast mixing social network, the total number of Sybils that rank higher than non-Sybils is O(*g* log *n*).

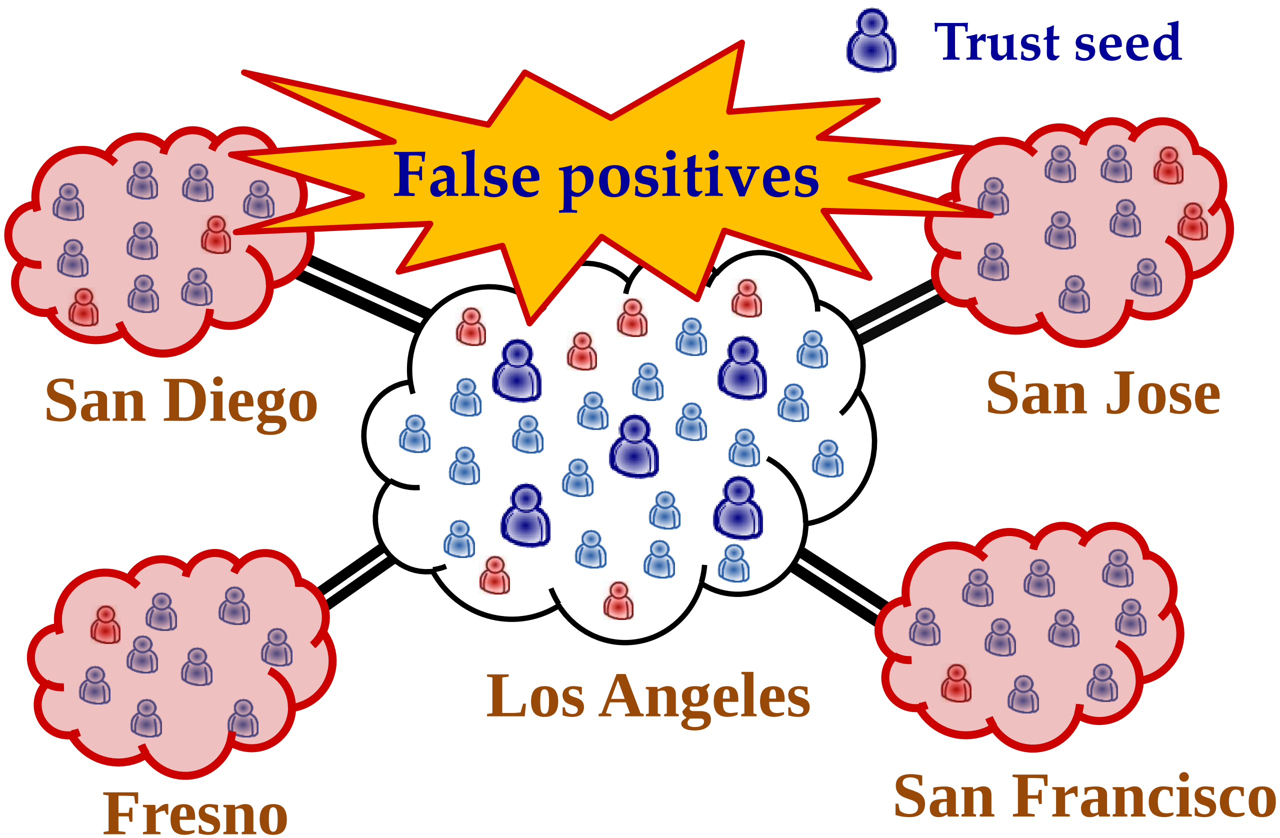
Rankings



Only O(g log n)

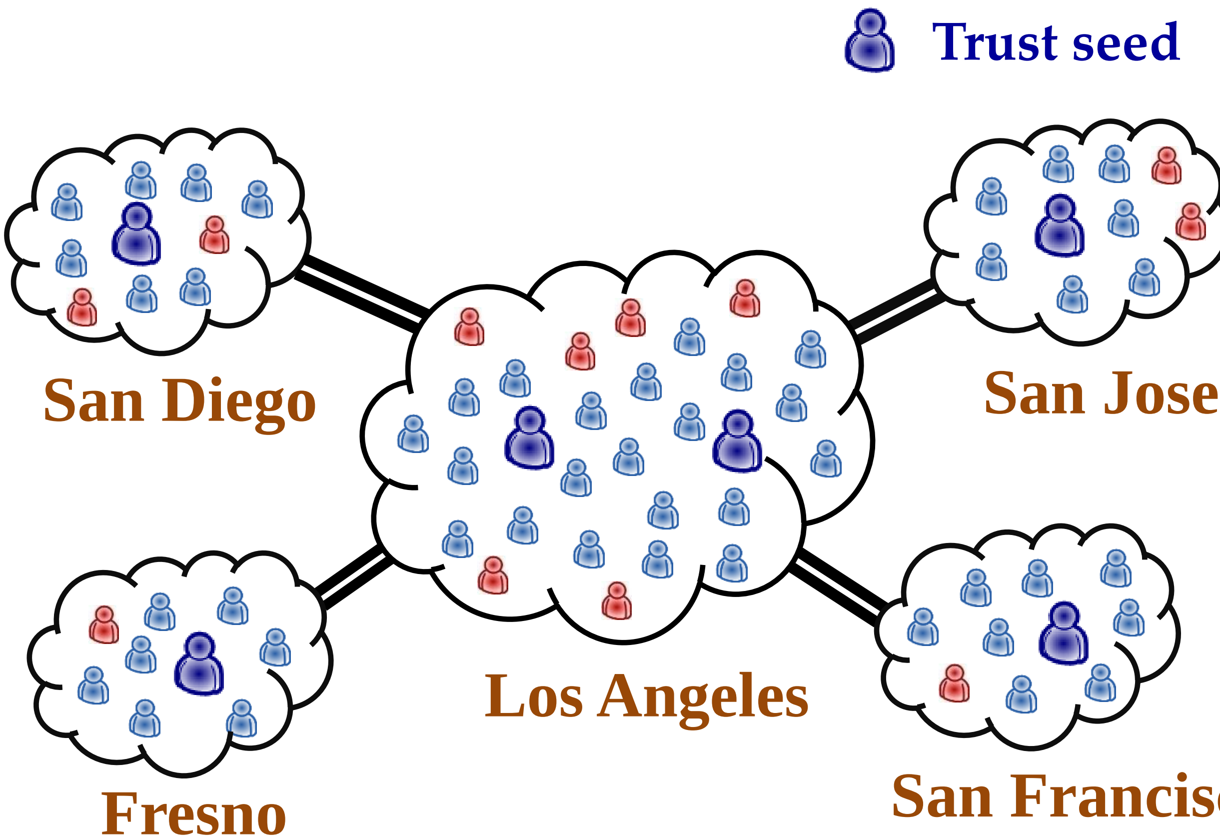
Coping with the multi-community structure

- **A weakness of social-graph-based schemes [SIGCOMM'10]**



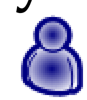
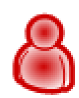
Coping with the multi-community structure

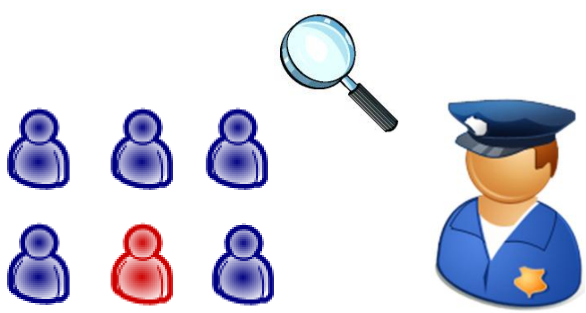
- **Solution: leverage the support for multiple seeds**
  - **Distribute seeds into communities**



How to distribute seeds?

- **Estimate communities**
  - **The Louvain method**
    [Blondel et al., J. of Statistical Mechanics'08]

- **Distribute non-Sybil seeds in communities**
  - **Manually inspect a set of nodes in each community**
  - **Use the nodes that passed the inspection as seeds** 
  - **Sybils cannot be seeds** 

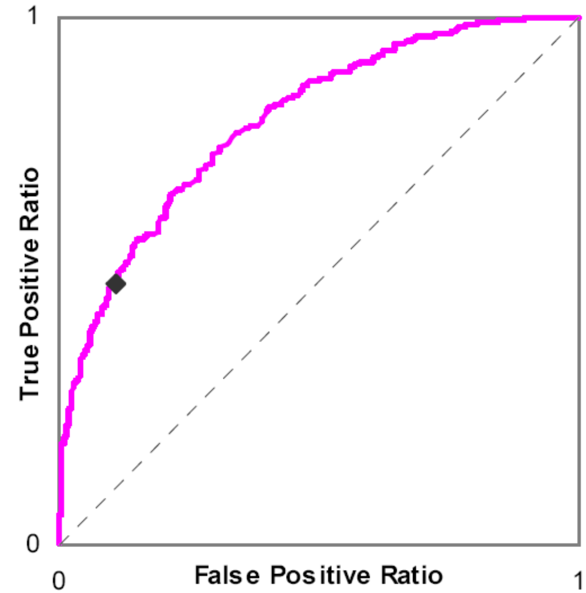


Evaluation

- Comparative evaluation
- Real-world deployment in Tuenti

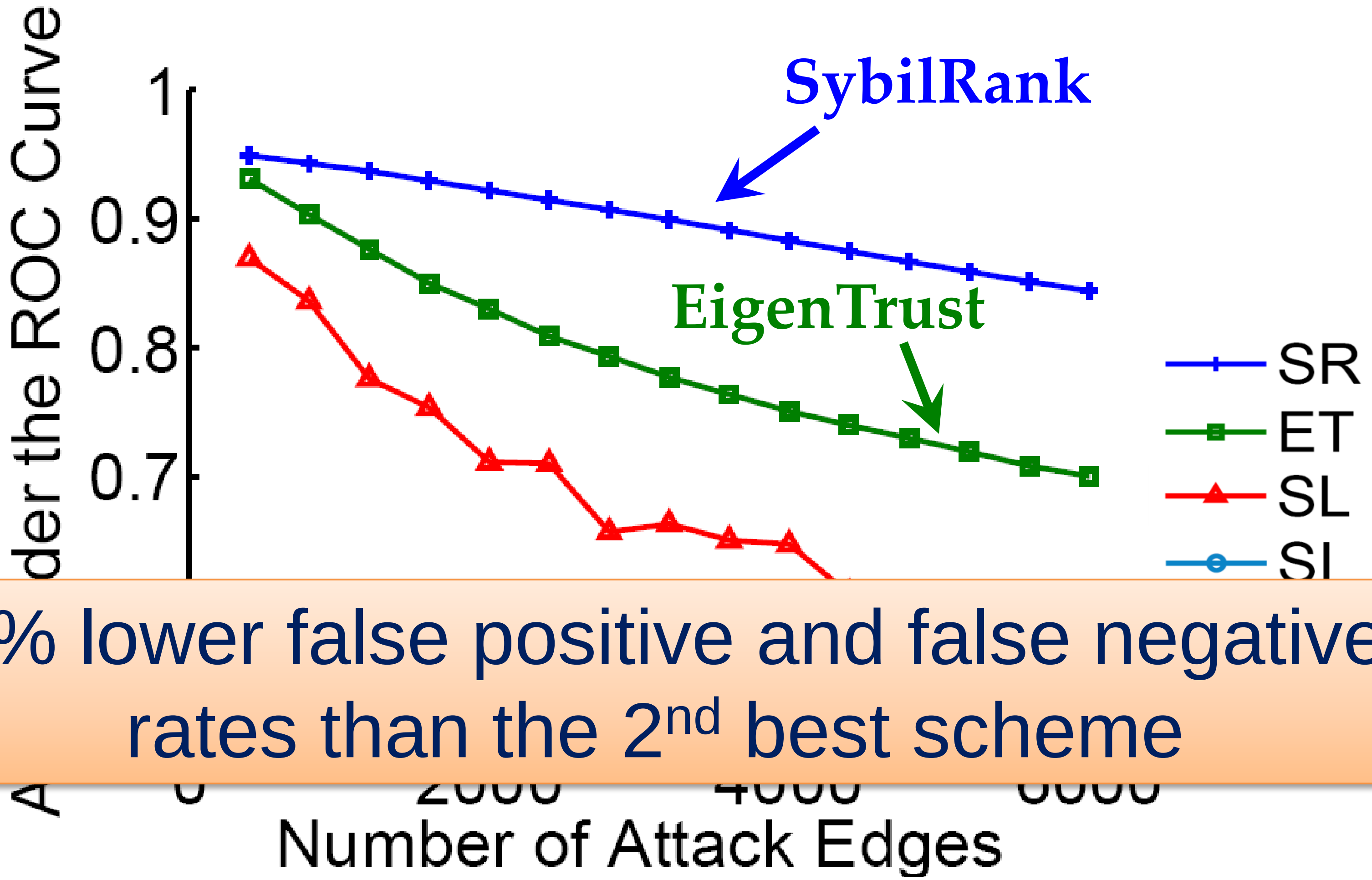
Comparative evaluation

- Stanford large network dataset collection
- Ranking quality
  - Area under the Receiver Operating Characteristics (ROC) curve [Viswanath et al., SIGCOMM'10]
- Compared approaches
  - SybilLimit (SL)
  - SybilInfer (SI)
  - EigenTrust (ET)
  - GateKeeper [INFOCOM'11]
  - Community detection [SIGCOMM'10]



[Fogarty et al., GI'05]

SybilRank has the lowest false rates

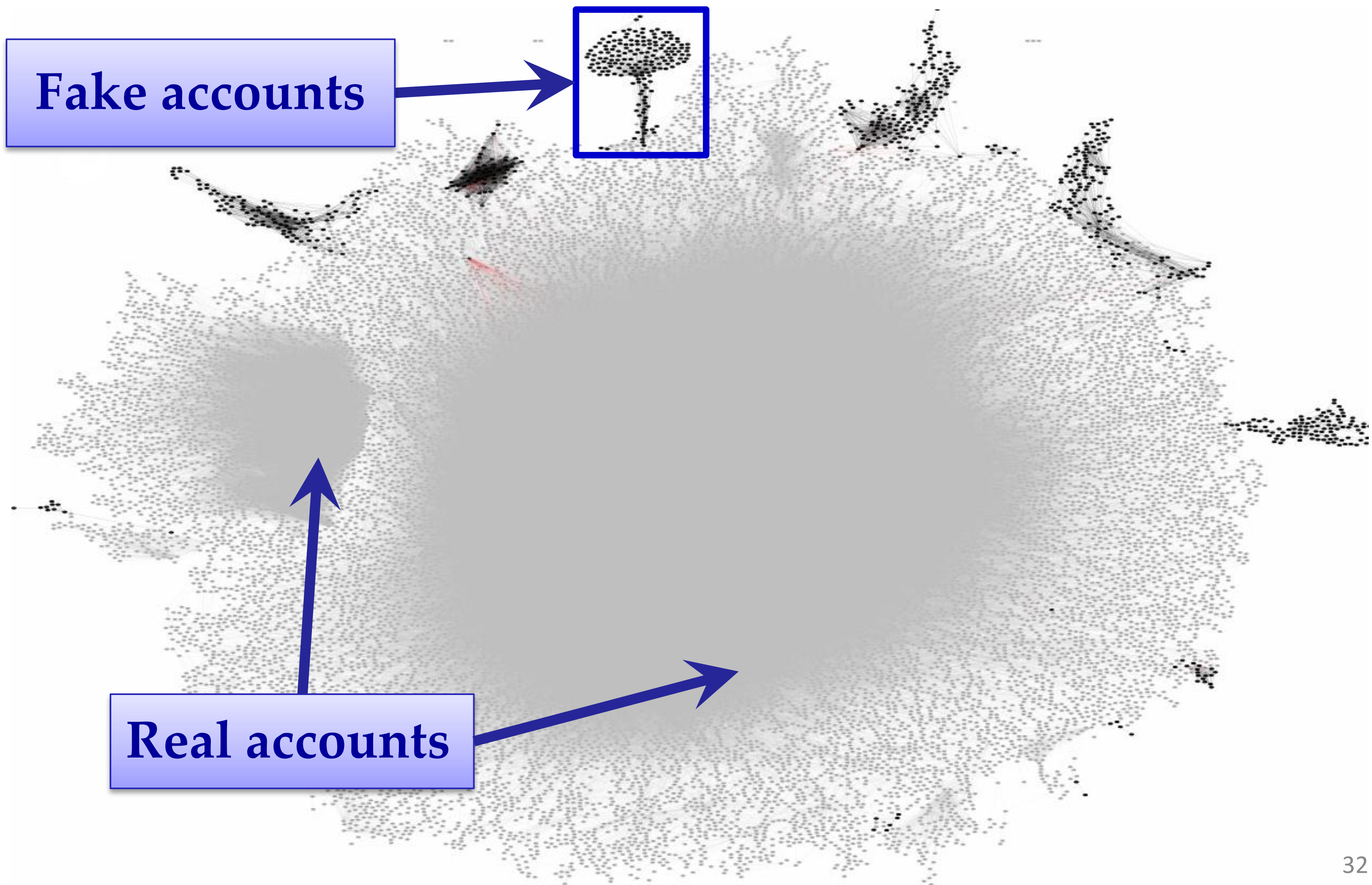


20% lower false positive and false negative rates than the 2nd best scheme

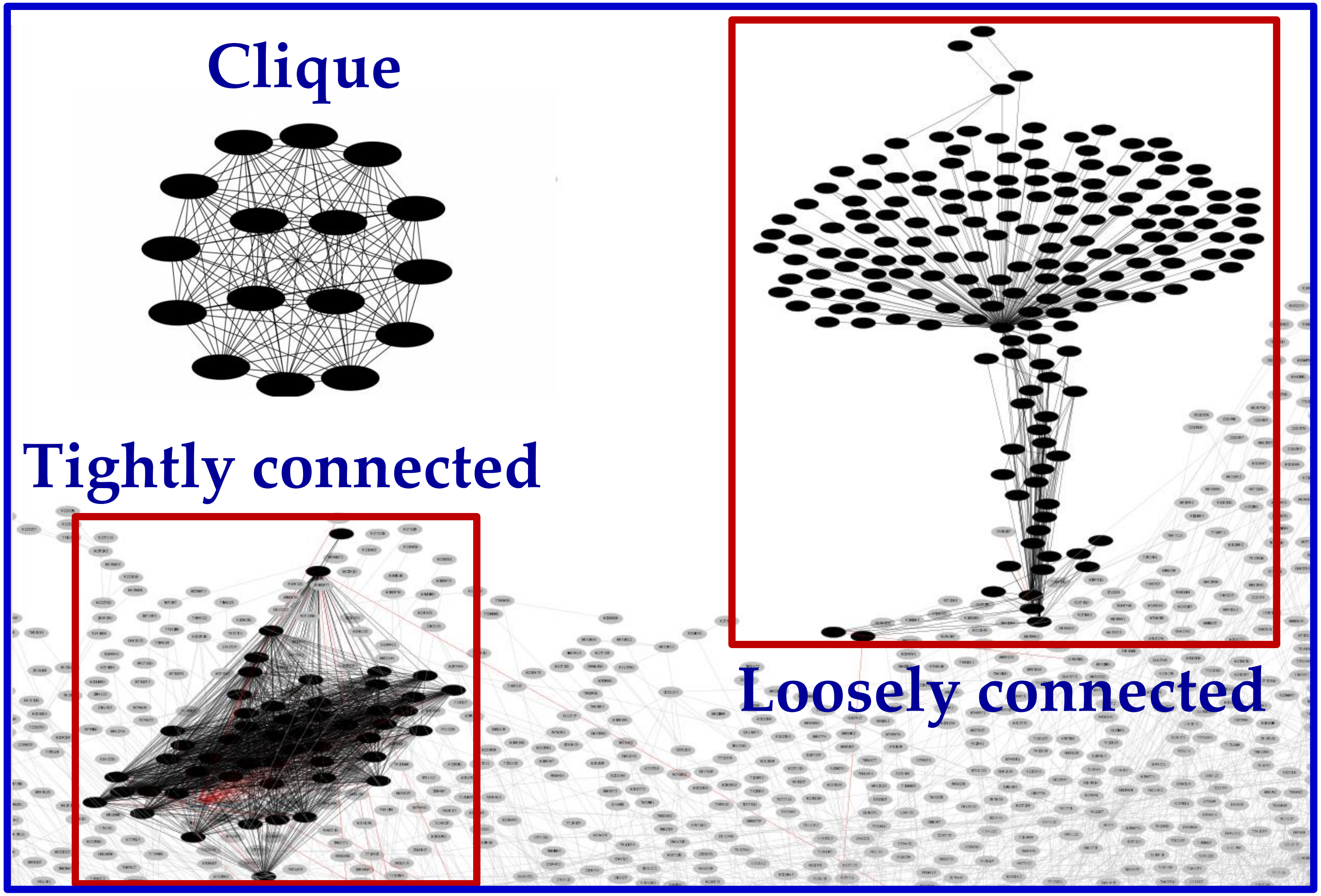
Real-world deployment

- **Used the anonymized Tuenti social graph**
  - **11 million users**
  - **1.4 billion social links**
  - **25 large communities with >100K nodes in each**

A 20K-user Tuenti community



Various connection patterns among suspected fakes

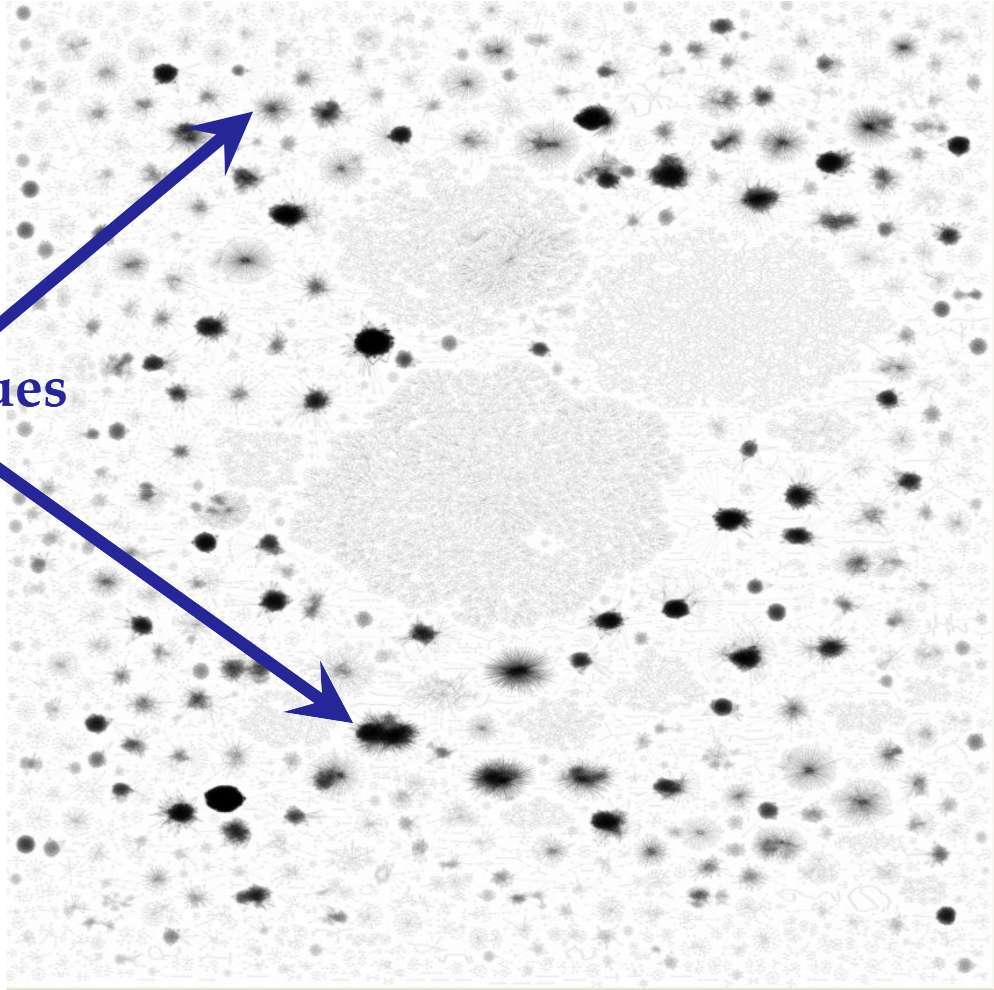


A global view of suspected fakes' connections

50K suspected accounts

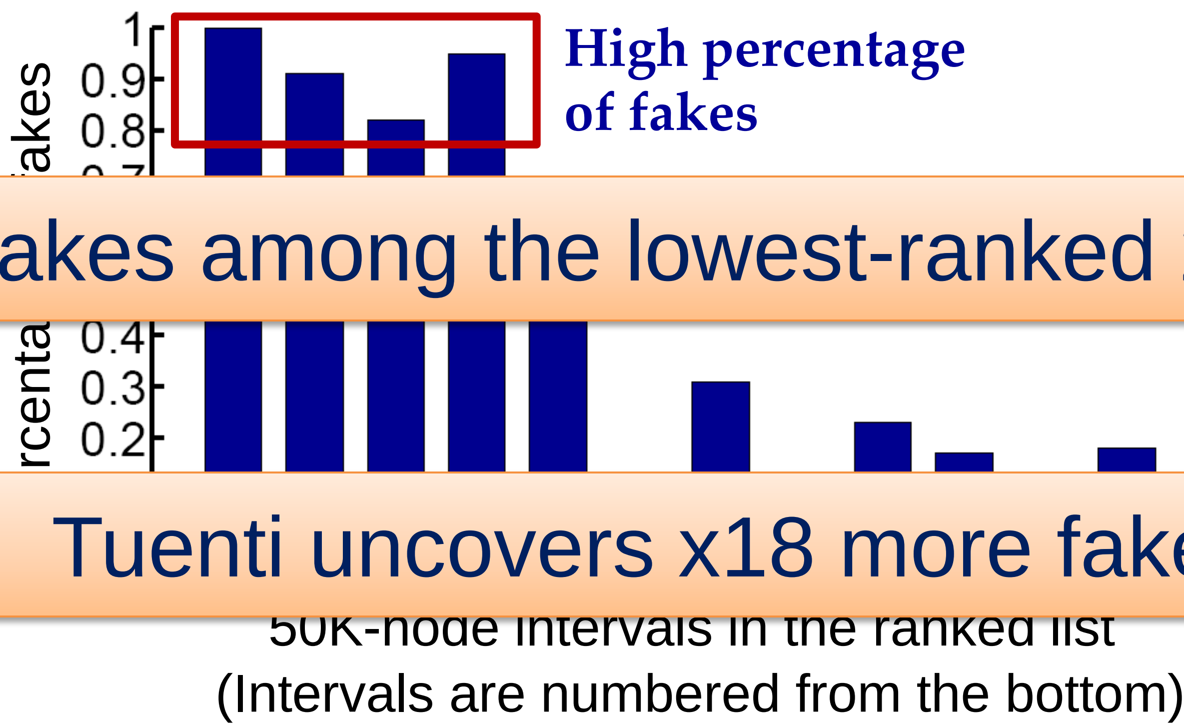
Small clusters/cliques

Controlled by many distinct attackers



SybilRank is effective

- **Percentage of fakes in each 50K-node interval**
  - **Estimated by random sampling**
  - **Fakes are confirmed by Tuenti's inspection team**



High percentage of fakes

~180K fakes among the lowest-ranked 200K users

Tuenti uncovers x18 more fakes

Conclusion: a practical Sybil defense

- **SybilRank: ranks users according to the landing probability of short random walks**
  - **Computational cost O(n log n)**
  - **Provable security guarantee**

- **Deployment in Tuenti**
  - **~200K lowest ranked users are mostly Sybils**

- **Enhances Tuenti's previous Sybil defense workflow**

Thank You!

qiangcao@cs.duke.edu
michael.sirivianos@cut.ac.cy
xwy@cs.duke.edu
tiago@tuenti.com

Questions?