Security Analysis of Malicious Socialbots on the Web

# Living in the (malicious) social web: Beyond friendships

Yazan Boshmaf

Yazan Boshmaf, Konstantin Beznosov, *Matei Ripeanu*,

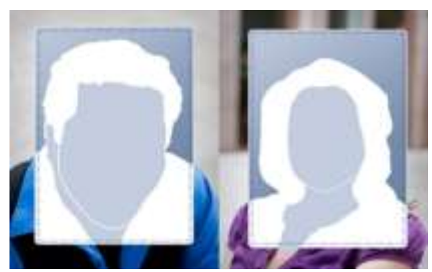
Dionysions Logothetis, Georgios Siganos, Jose Lorenzo

Dissertation presented in partial fulfillment of degree requirements of PhD in ECE, UBC

# Social bots

Automated fake accounts in online social networks (OSNs)


+



=

Designed to <u>deceive</u> and appear human

Hwang et al. Socialbots: Voices from the fronts. ACM Interactions 19, 2 (March 2012), 38-45.

# The threat of malicious social bots

Automated fake accounts in online social networks (OSNs)




+

=



**What is at stake?**

Designed to deceive and appear human

Hwang et al. Socialbots: Voices from the fronts. ACM Interactions 19, 2 (March 2012), 38-45.

# Fake accounts are bad for business

CBC

CBCnews | Technology & Science

**Facebook shares drop on news of fake accounts**

**83 million accounts false or duplicates, company reveals**

The Associated Press   Posted: Aug 03, 2012 10:47 AM ET | Last Updated: Aug 03, 2012 2:11 PM ET

"… 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…"

# Fake accounts are bad for users

OSNs are attractive medium for abusive users

Social Infiltration

Connecting with many benign users (friend request spam)

Bilge et al. All your contacts are belong to us: Automated identity theft attacks on social networks. Proc. of WWW, 2009

# Fake accounts are bad for users

OSNs are attractive medium for abusive users



Social Infiltration


Data collection

Online surveillance, profiling, and data commoditization

Nolan et al. Hacking human: Data-archaeology and surveillance in social networks. ACM SIGGROUP Bulletin 25.2, 2005

# Fake accounts are bad for users

OSNs are attractive medium for abusive users



Social Infiltration



Data collection

Misinformation

Influencing users, biasing public opinion, propaganda

Ratkiewicz et al. Detecting and tracking political abuse in social media. Proc. of ICWSM. 2011

# Fake accounts are bad for users

OSNs are attractive medium for abusive users



Social Infiltration




Data collection



Misinformation


Malware Infection

Infecting computers and use it for DDoS, spamming, and fraud

Thomas et al. The Koobface botnet and the rise of social malware. Proc. of MALWARE, 2010

# Our work


Threat characterization

Countermeasure design

Fake accounts are bad for users

OSNs are attractive medium for abusive content

Social Infiltration    Data collection    Misinformation    Malware Infection

Infecting computers and use it for DDoS, spamming, and fraud[1]

[1] Thomas et al. The Koobface botnet and the rise of social malware. Proc. of MALWARE, 2010.

# Questions

1 *How vulnerable are OSNs to social infiltration?*

- Vulnerability analysis
- Characterization of user behavior

2 What are the security and privacy implications of social infiltration?

- Quantification of privacy breaches
- Effectiveness of security defenses

3 What is the economic rationale behind infiltrating OSNs at scale?

- Scalability from economic context
- Profit-maximizing infiltration strategy

4 How can OSNs detect fakes or social bots that infiltrate on a large scale?

- Victim prediction for robust detection
- Framework for evaluation

# Questions

1 *How vulnerable are OSNs to social infiltration?*

- Vulnerability analysis
- Characterization of user behavior

2 *What are the security and privacy implications of social infiltration?*

- Quantifying privacy breaches
- Effectiveness of security defenses

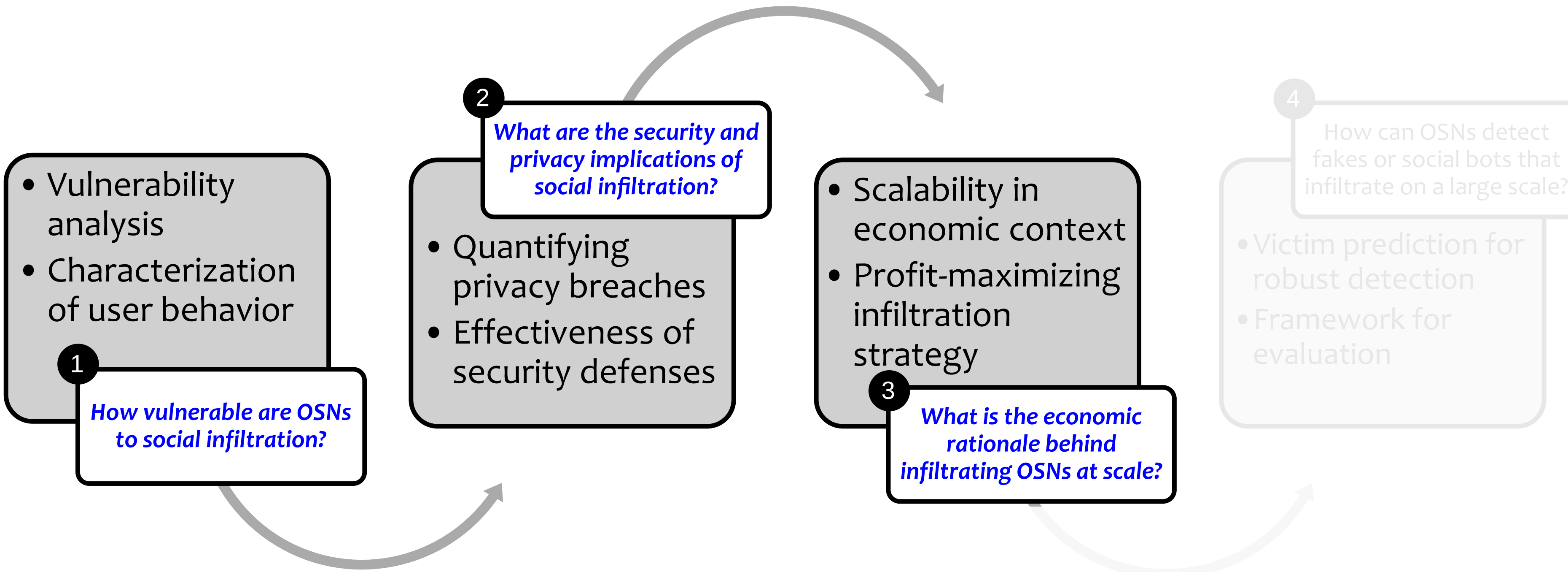
3 What is the economic rationale behind infiltrating OSNs at scale?

- Scalability from economic context
- Profit-maximizing infiltration strategy

4 How can OSNs detect fakes or social bots that infiltrate on a large scale?

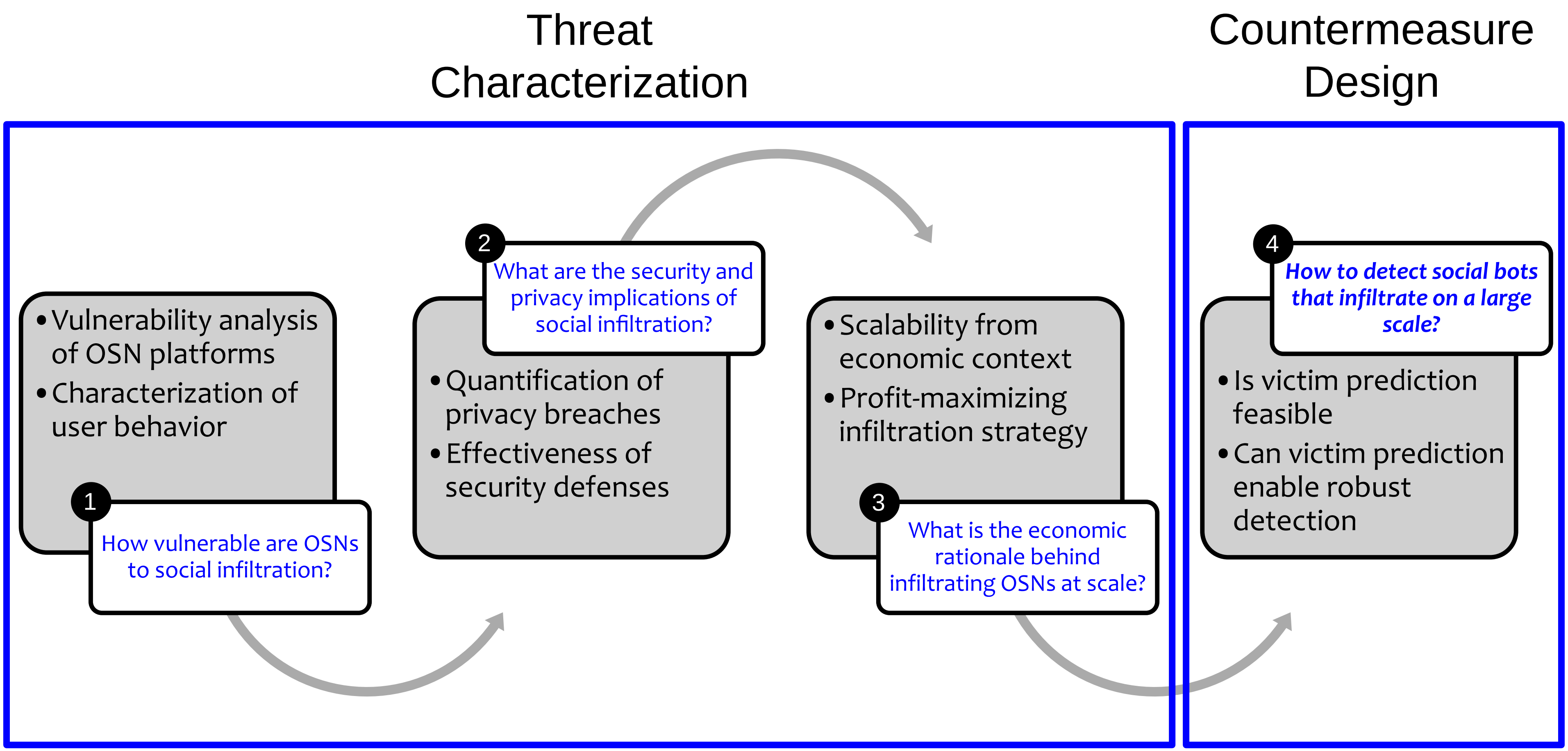
- Victim prediction for robust detection
- Framework for evaluation

# Questions



- Vulnerability analysis
- Characterization of user behavior
1
How vulnerable are OSNs to social infiltration?
2
What are the security and privacy implications of social infiltration?
- Quantifying privacy breaches
- Effectiveness of security defenses
- Scalability in economic context
- Profit-maximizing infiltration strategy
3
What is the economic rationale behind infiltrating OSNs at scale?
4
How can OSNs detect fakes or social bots that infiltrate on a large scale?
- Victim prediction for robust detection
- Framework for evaluation

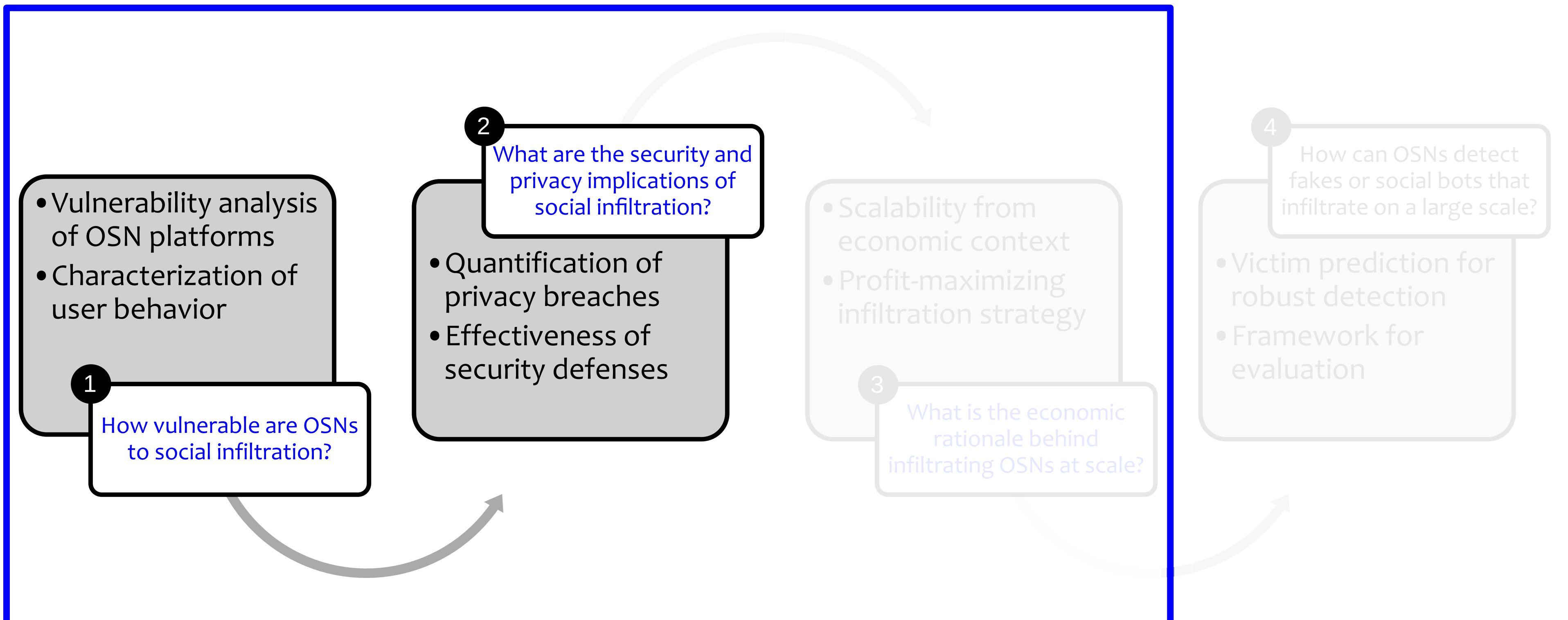
# Questions



Threat Characterization
Countermeasure Design
1
How vulnerable are OSNs to social infiltration?
• Vulnerability analysis of OSN platforms
• Characterization of user behavior
2
What are the security and privacy implications of social infiltration?
• Quantification of privacy breaches
• Effectiveness of security defenses
3
What is the economic rationale behind infiltrating OSNs at scale?
• Scalability from economic context
• Profit-maximizing infiltration strategy
4
How to detect social bots that infiltrate on a large scale?
• Is victim prediction feasible
• Can victim prediction enable robust detection

# Attack side: Social infiltration in OSNs

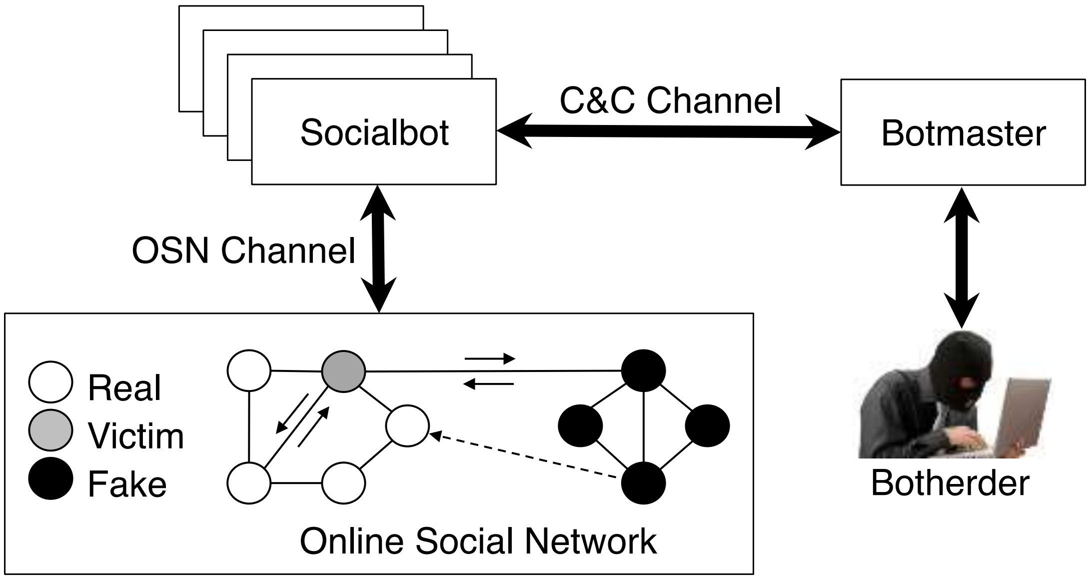
Threat Characterization




[1] *The socialbot network: When bots socialize for fame and money,* Boshmaf, Beznosov, Ripeanu, ACSAC, Dec 2011
[2] *Key challenges in defending against malicious socialbots*, Boshmaf, Beznosov, Ripeanu, USENIX LEET, April 2012
[3] *Design and analysis of a social botnet*, Boshmaf, Beznosov, Ripeanu, J. Comp. Net., 57(2), Feb 2013

# Social botnet: Experiment

Operated 100 socialbots on Facebook, single botmaster




Bots sent 9.6K friend requests send in 8 weeks,
35.7% requests from bots accepted (victims)

# Main findings

*(Platform-level vulnerability)*

It is feasible to automate social infiltration by exploiting platform and user vulnerabilities

Threat Characterization

# Main findings

*(Data breaches)*

Social infiltration results in serious privacy breaches, where personally identifiable information is compromised

# Victims are highly affected

| Profile Info | Direct (%) Before | Direct (%) After | Extended (%) Before | Extended (%) After |
|---|---|---|---|---|
| Birth Date | 3.5 | 72.4 | 4.5 | 53.8 |
| Email Address | 2.4 | 71.8 | 2.6 | 4.1 |
| Gender | 69.1 | 69.2 | 84.2 | 84.2 |
| Home City | 26.5 | 46.2 | 29.2 | 45.2 |
| Current City | 25.4 | 42.9 | 27.8 | 41.6 |
| Phone Number | 0.9 | 21.1 | 1.0 | 1.5 |
| School Name | 10.8 | 19.7 | 12.0 | 20.4 |
| Postal Address | 0.9 | 19.0 | 0.7 | 1.3 |
| IM Account ID | 0.6 | 10.9 | 0.5 | 0.8 |
| Married To | 2.9 | 6.4 | 3.9 | 4.9 |
| Worked At | 2.8 | 4.0 | 2.8 | 3.2 |
| Average | 13.3 | 34.9 | 15.4 | 23.7 |

2.62 times more private data collected after infiltration

# Friends of victims are affected too

| Profile Info | Direct (%) Before | Direct (%) After | Extended (%) Before | Extended (%) After |
|---|---|---|---|---|
| Birth Date | 3.5 | 72.4 | 4.5 | 53.8 |
| Email Address | 2.4 | 71.8 | 2.6 | 4.1 |
| Gender | 69.1 | 69.2 | 84.2 | 84.2 |
| Home City | 26.5 | 46.2 | 29.2 | 45.2 |
| Current City | 25.4 | 42.9 | 27.8 | 41.6 |
| Phone Number | 0.9 | 21.1 | 1.0 | 1.5 |
| School Name | 10.8 | 19.7 | 12.0 | 20.4 |
| Postal Address | 0.9 | 19.0 | 0.7 | 1.3 |
| IM Account ID | 0.6 | 10.9 | 0.5 | 0.8 |
| Married To | 2.9 | 6.4 | 3.9 | 4.9 |
| Worked At | 2.8 | 4.0 | 2.8 | 3.2 |
| Average | 13.3 | 34.9 | 15.4 | 23.7 |

1.54 times more, with more than
1 million affected users

# Friends of victims are affected too

| Profile Info | Direct (%) Before | Direct (%) After | Extended (%) Before | Extended (%) After |
|---|---|---|---|---|
| Birth Date | 3.5 | 72.4 | 4.5 | 53.8 |
| Email Address | 2.4 | 71.8 | 2.6 | 4.1 |
| Gender | 69.1 | 69.2 | 84.2 | 84.2 |
| Home City | 26.5 | 46.2 | 29.2 | 45.2 |
| Current City | 25.4 | 42.9 | 27.8 | 41.6 |
| Phone Number | 0.9 | 21.1 | 1.0 | 1.5 |
| School Name | 10.8 | 19.7 | 12.0 | 20.4 |
| Postal Address | 0.9 | 19.0 | 0.7 | 1.3 |
| IM Account ID | 0.6 | 10.9 | 0.5 | 0.8 |
| Married To | 2.9 | 6.4 | 3.9 | 4.9 |
| Worked At | 2.8 | 4.0 | 2.8 | 3.2 |
| Average | 13.3 | 34.9 | 15.4 | 23.7 |

From 49K birthdates to 584K

1.54 times more, with more than

1 million affected users



Acquisti et al. Predicting social security numbers from public data. Proc. Of Nat. Acad. of Sc. 106(27), 2009

# Vulnerabilities exploited to automate infiltration

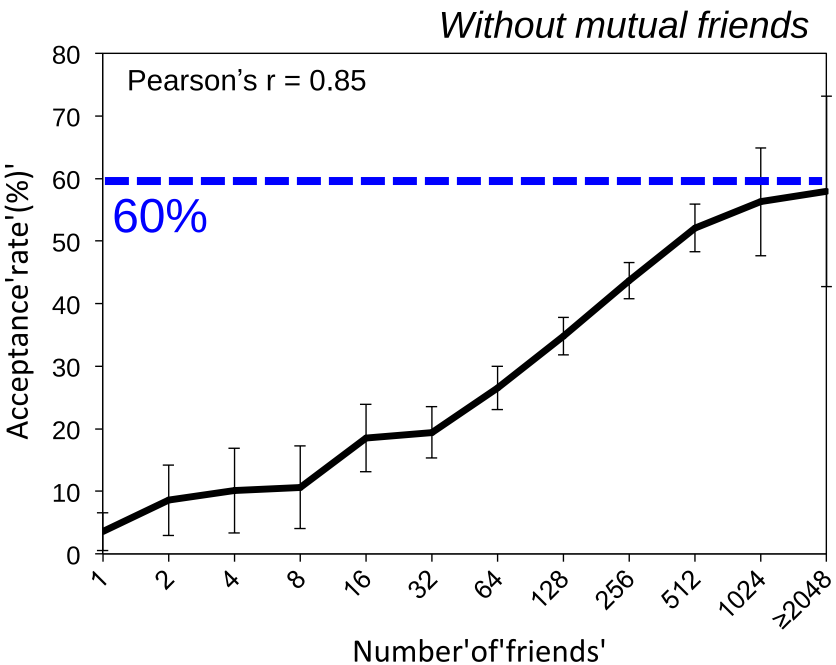
*(User behavior characterization)*


Some users are more susceptible to social infiltration, which partly depends on factors related to their social structure

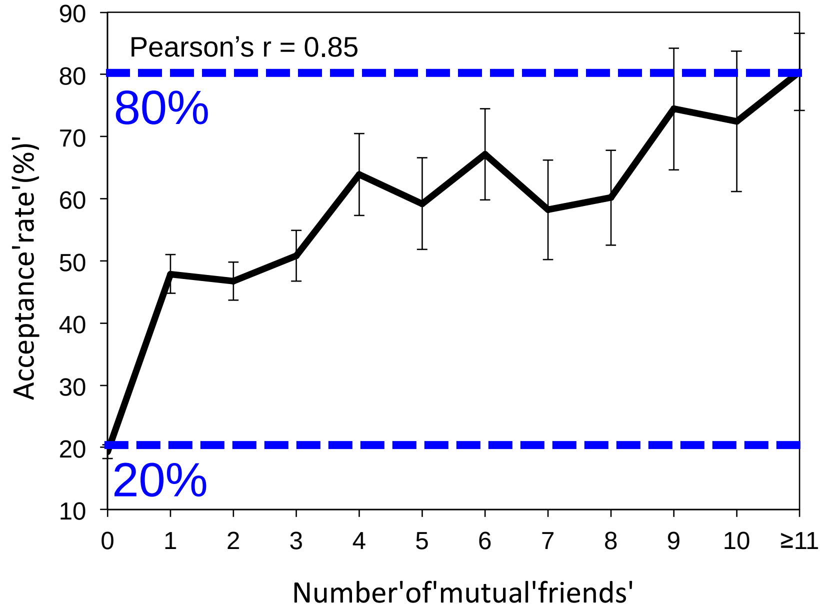
Large scale network crawls

Exploitable platforms and APIs

# User susceptibility to become a victim correlates with social structure




More friends, more susceptible to infiltration




More *mutual friends*, more susceptible to infiltration

# Fake accounts mimic real accounts

Only 20% of fakes were "detected"


All manually flagged by concerned users

# Friends of victims are affected too

*(Feature-based detection is ineffective)*

| Profile Info | Direct (%) Before | Direct (%) After | Extended (%) Before | Extended (%) After |
|---|---|---|---|---|
| Birth Date | [illegible] | [illegible] | [illegible] | 53.8 |
| Email Address | 2.4 | 71.8 | [illegible] | 4.1 |
| Gender | 69.1 | 69.2 | 84.2 | 84.2 |
| Home City | 26.5 | 46.2 | 29.2 | 45.2 |
| Current City | 25.4 | 42.9 | 27.8 | 41.6 |
| Phone Number | 0.9 | 21.1 | 1.0 | 1.5 |
| School Name | [illegible] | [illegible] | [illegible] | [illegible] |
| Postal Address | 0.9 | 19.0 | 0.7 | 1.3 |
| IM Account ID | 0.6 | 10.9 | 0.5 | 0.8 |
| Married To | [illegible] | [illegible] | [illegible] | [illegible] |
| Worked At | 2.8 | 4.0 | 2.8 | 3.2 |
| Average | 13.3 | 34.9 | 15.4 | 23.7 |

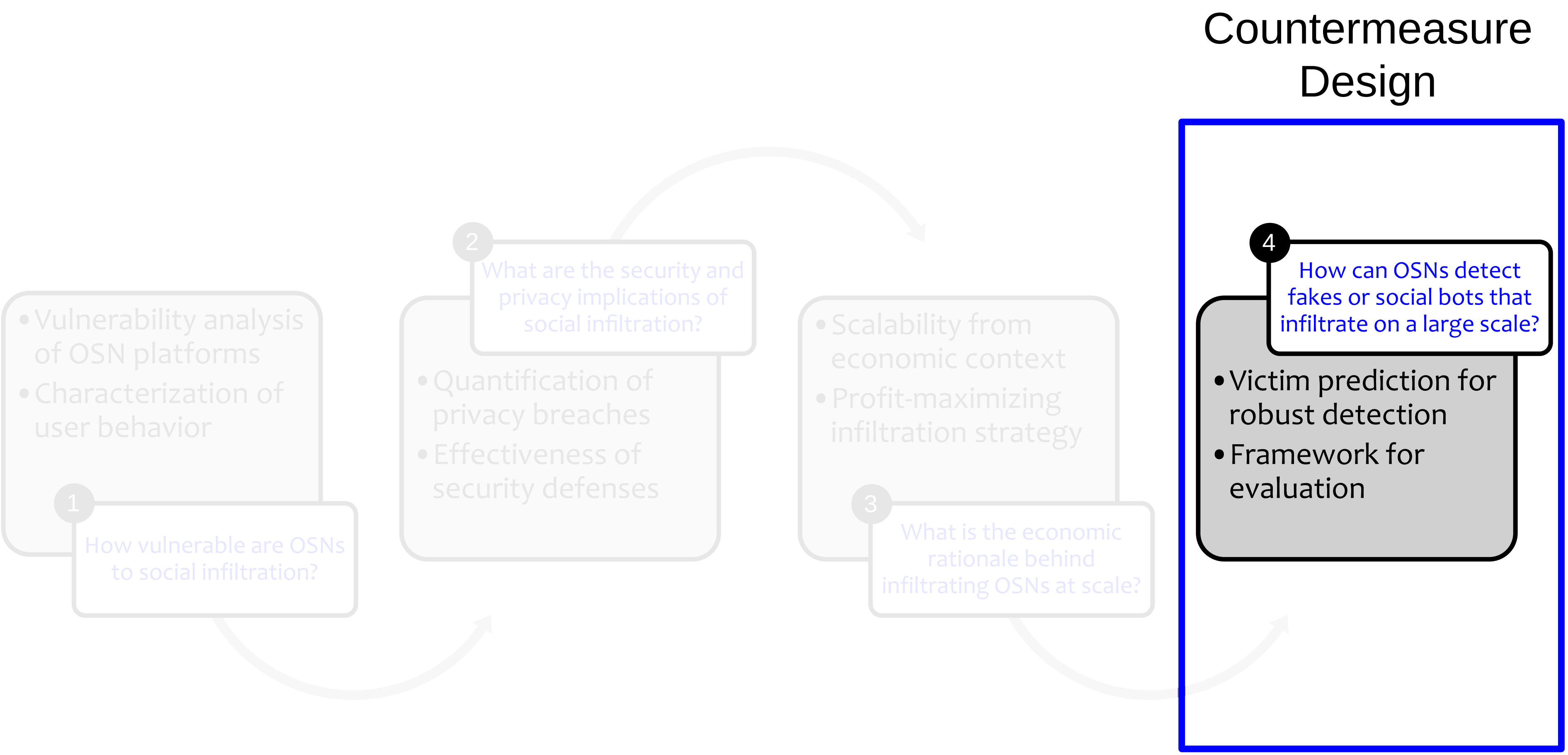
From 49K birthdates to 584K

Socialbots leads to arms race and render feature-based fake account detection ineffective

1.54 times more, with more than 1 million affected users

Acquisti et al. Predicting social security numbers from public data. Proc. Of Nat. Acad. of Sc. 106(27), 2009

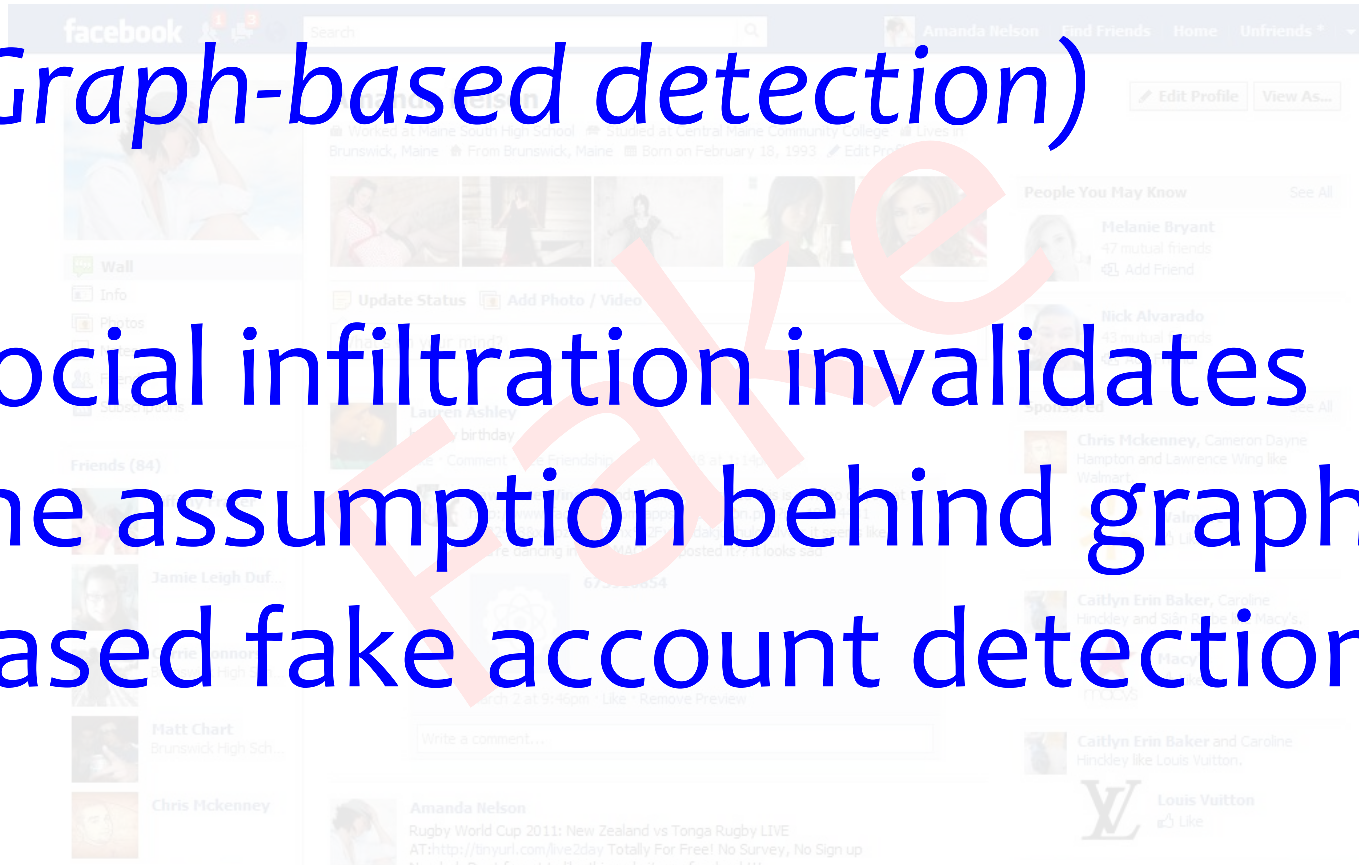
# Defense side: Infiltration-resilient fake account detection




[1] *Graph-based Sybil detection in social and information systems*. In Proc. of ASONAM, Aug 2013
[2] *Integro: Leveraging victim prediction for robust fake account detection in OSNs.* NDSS, Feb 2015
[3] *Thwarting fake accounts by predicting their victims.* Submitted to TISSEC, Feb 2015

# Feature-based detection is ineffective

Only 20% of fakes were “detected”



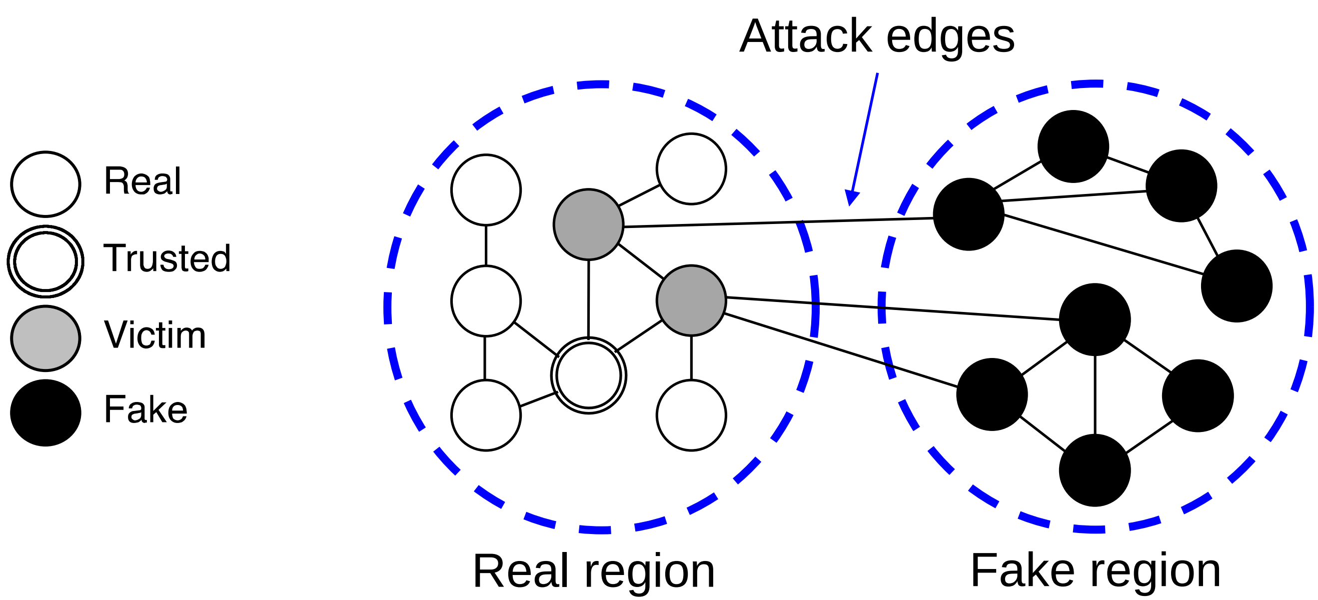

*(Graph-based detection)*

Social infiltration invalidates the assumption behind graph-based fake account detection

All manually flagged by concerned users

# Graph-based detection

Assumes social infiltration on a large scale is infeasible

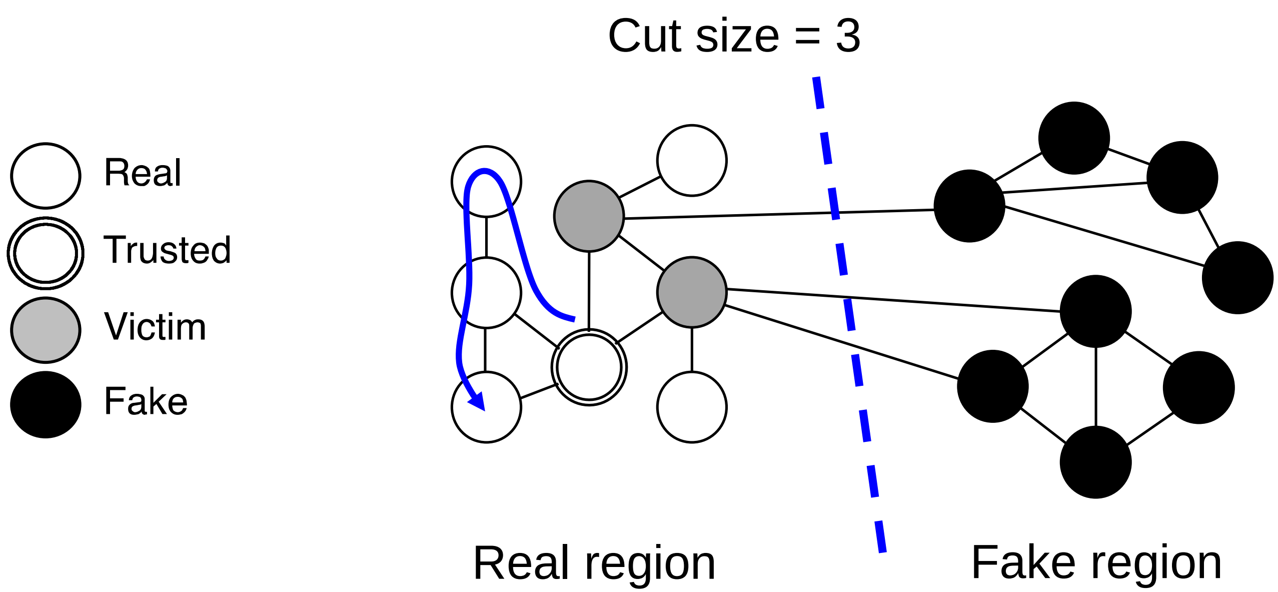



Finds a (provably) sparse cut between the regions by ranking

Alvisi et al. The evolution of Sybil defense via social networks. IEEE Security and Privacy, 2013.

# Graph-based detection

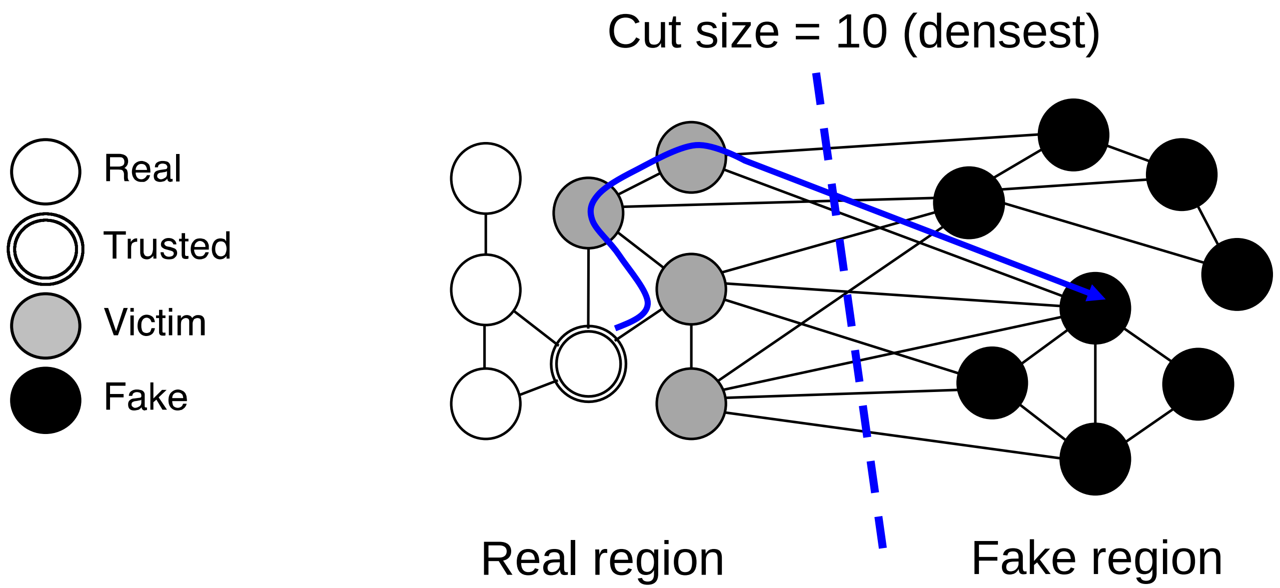
Ranks computed from landing probability of a <u>short</u> random walk




Most real accounts rank higher than fakes

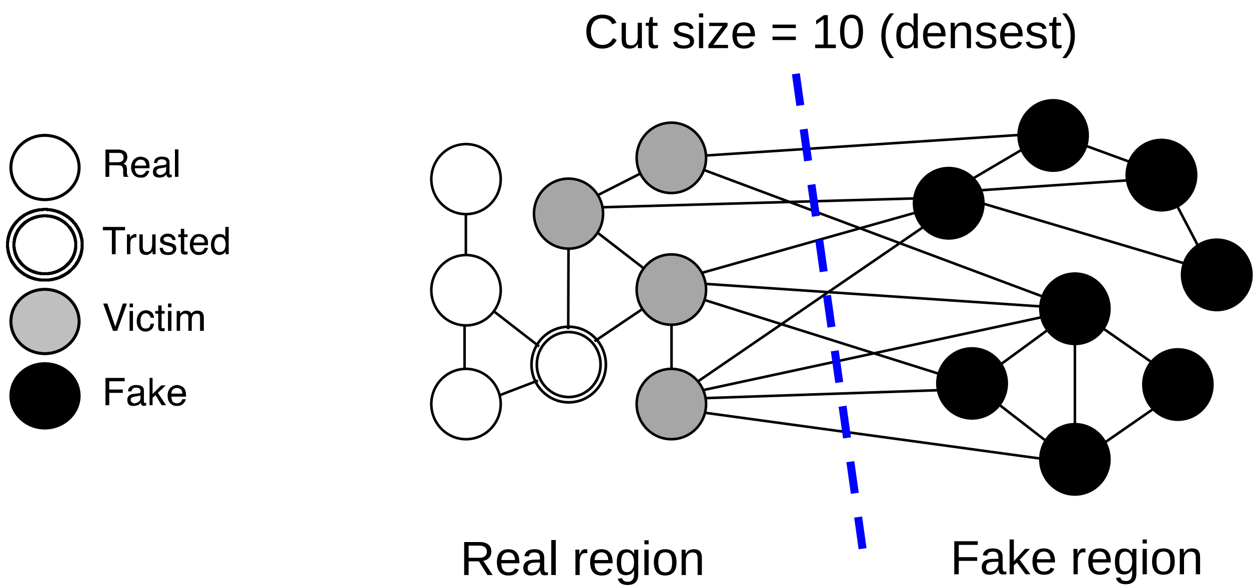
Alvisi et al. The evolution of Sybil defense via social networks. IEEE Security and Privacy, 2013.

# Graph-based detection is not resilient to social infiltration

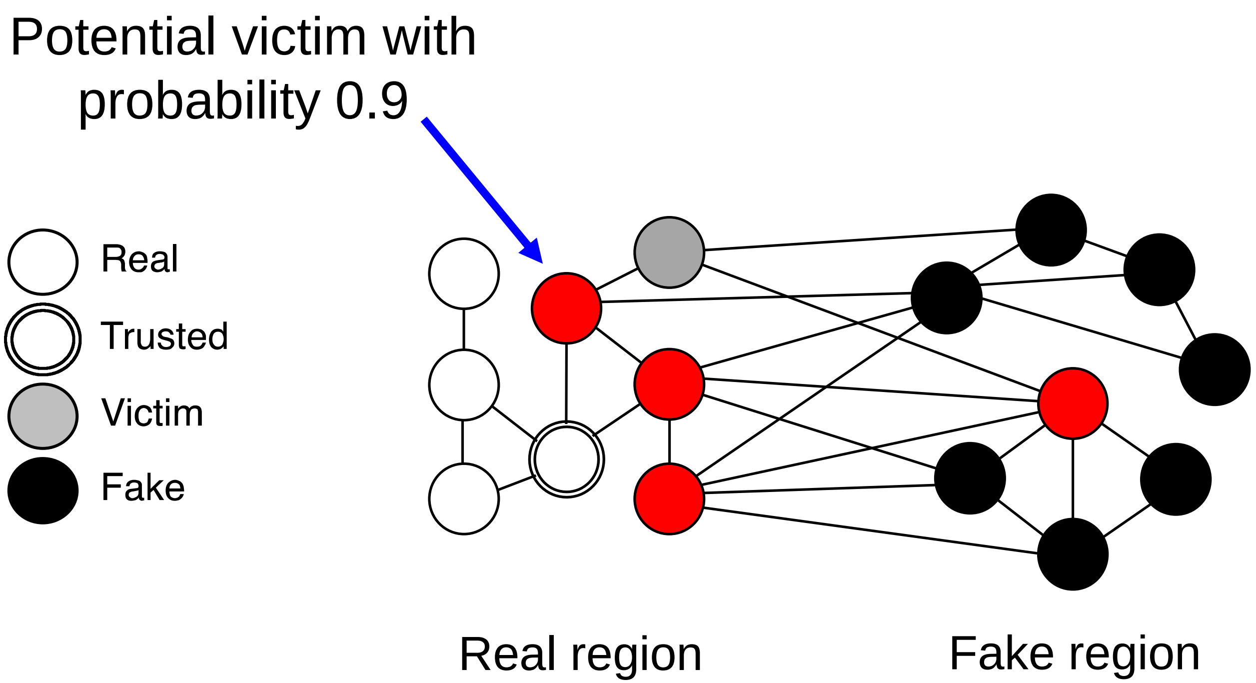



50% of bots had more than 35 attack edges

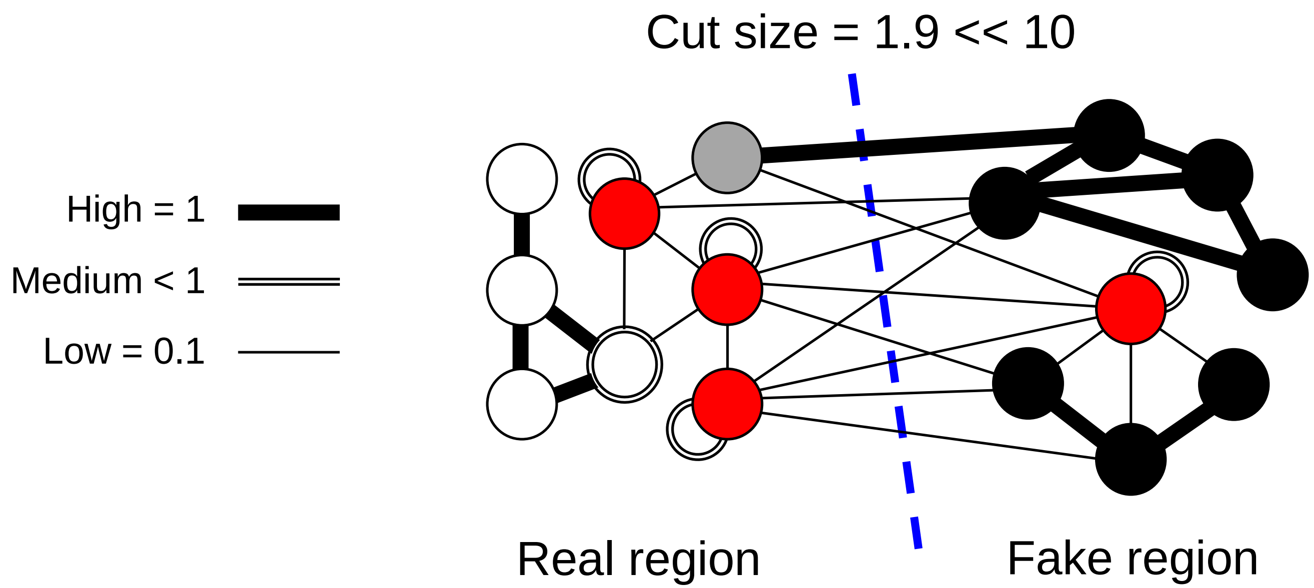
# Premise: Regions can be tightly connected



# **Key idea:** Identify potential victims with some probability



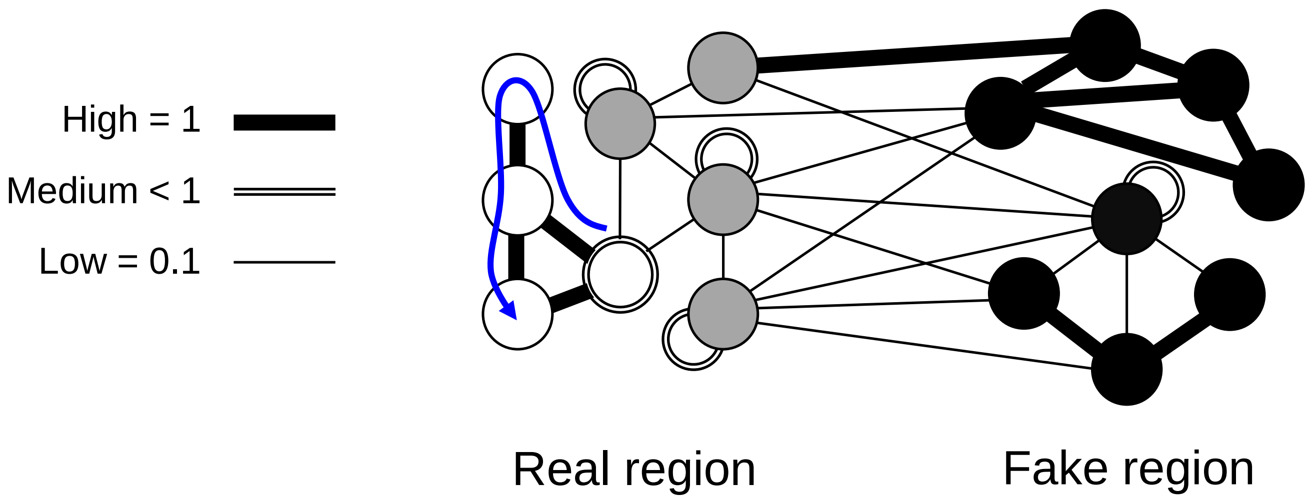
# **Key idea:** Leverage victim prediction to reduce cut size




Assign lower weight to edges incident to potential victims

# Delimit the real region by ranking accounts

Ranks computed from landing probability of a <u>short</u> random walk




Most real accounts are ranked higher than fake accounts

# Delimit the real region by ranking accounts

Ranks computed from landing probability of a short random walk

## *Result 1: Bound on ranking quality*

Number of fake accounts that rank equal to or higher than real accounts is $O(vol(E_A) \log n)$ where $vol(E_A) \leq |E_A|$

Medium < 1

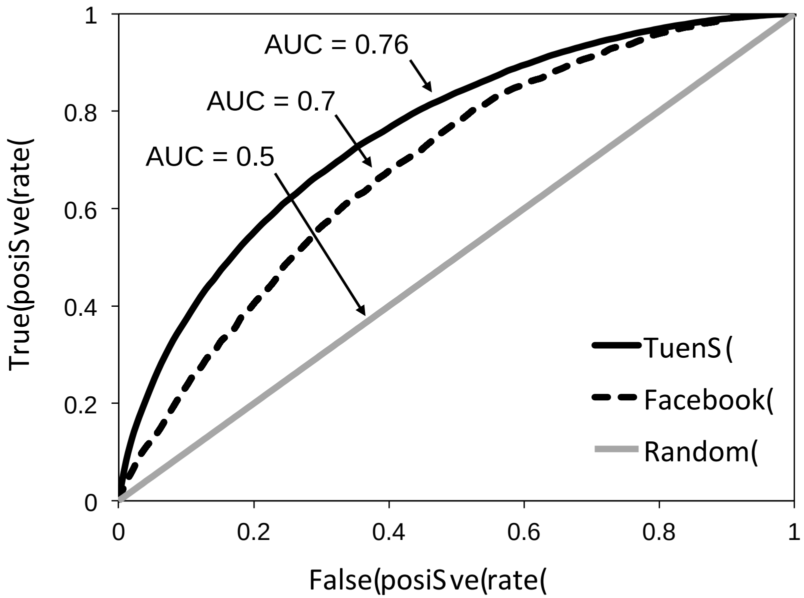
Real region

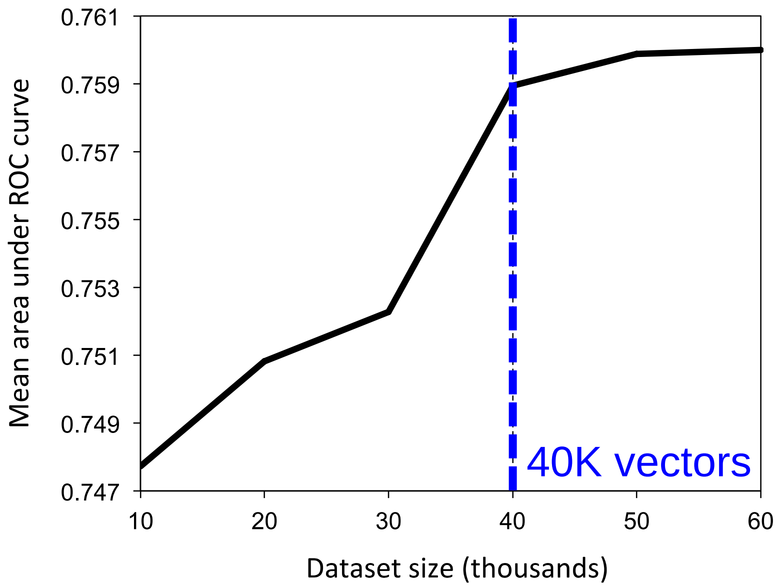
Fake region

Most real accounts are ranked higher than fake accounts

Assuming a fast mixing real region and an attacker who establishes attack edges at random

# Result 2: Victim classification is feasible (even using low-cost features)



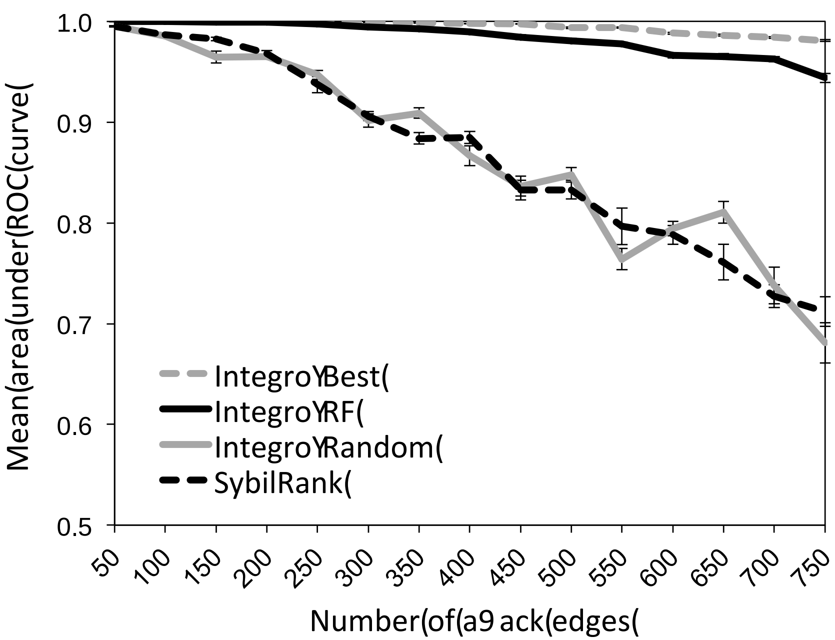

Random Forests (RF) achieves up to 52% better than random

No need to train on more than 40K feature vectors on Tuenti

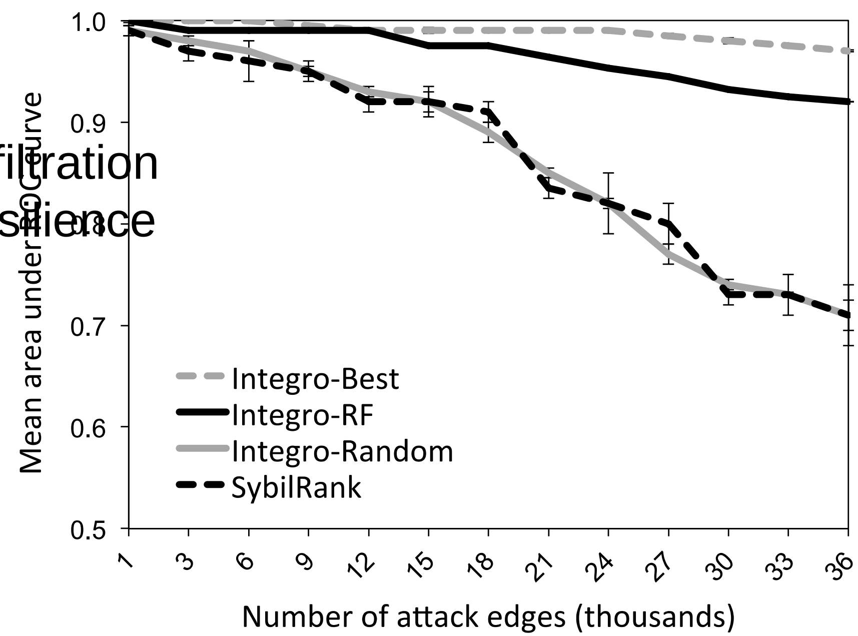
*Integro: Leveraging victim prediction for robust fake account detection in OSNs*. NDSS, Feb 2015
*Thwarting fake accounts by predicting their victims*. Submitted to TISSEC, Feb 2015.

# Result 3: Ranking is resilient to infiltration

Integro delivers up to 30% higher AUC, and AUC is always > 0.92




Infiltration resilience




Targeted-victim attack

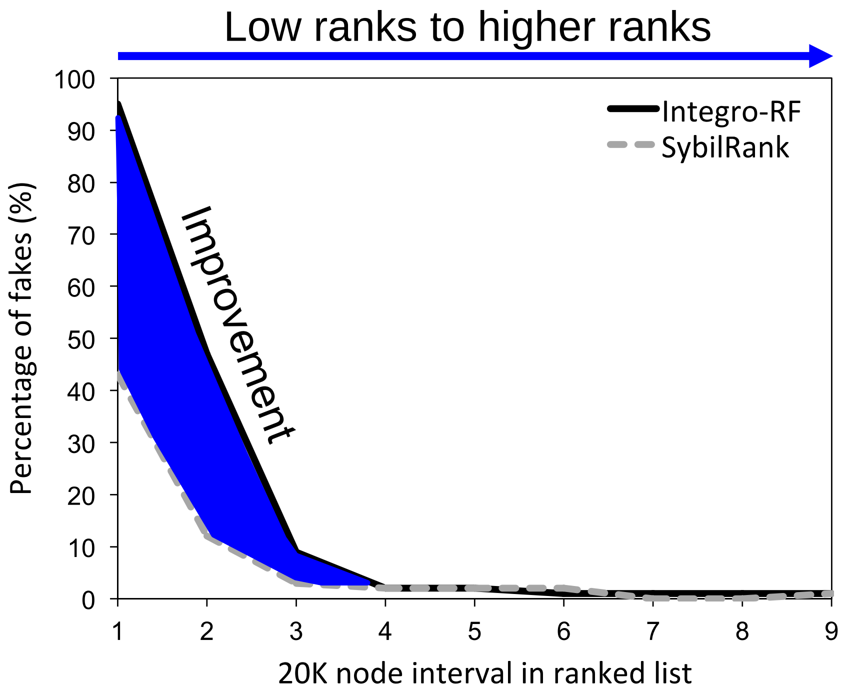
Random-victim attack



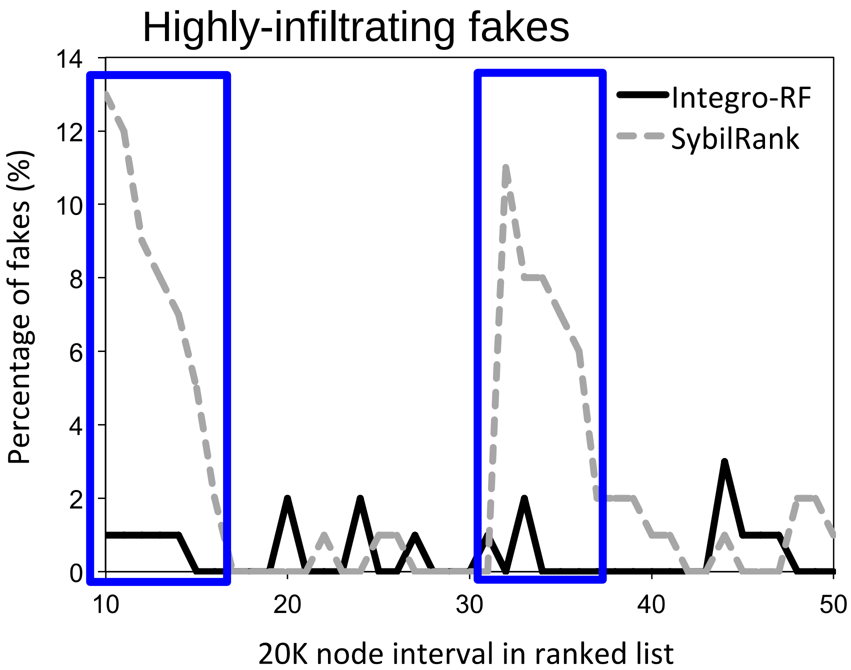
Cao et al. Aiding the Detection of Fake Accounts in Large Scale Social Online Services, NSDI'12

# Deployment at Tuenti confirms results

Integro delivers up to an order or magnitude better <u>precision</u>

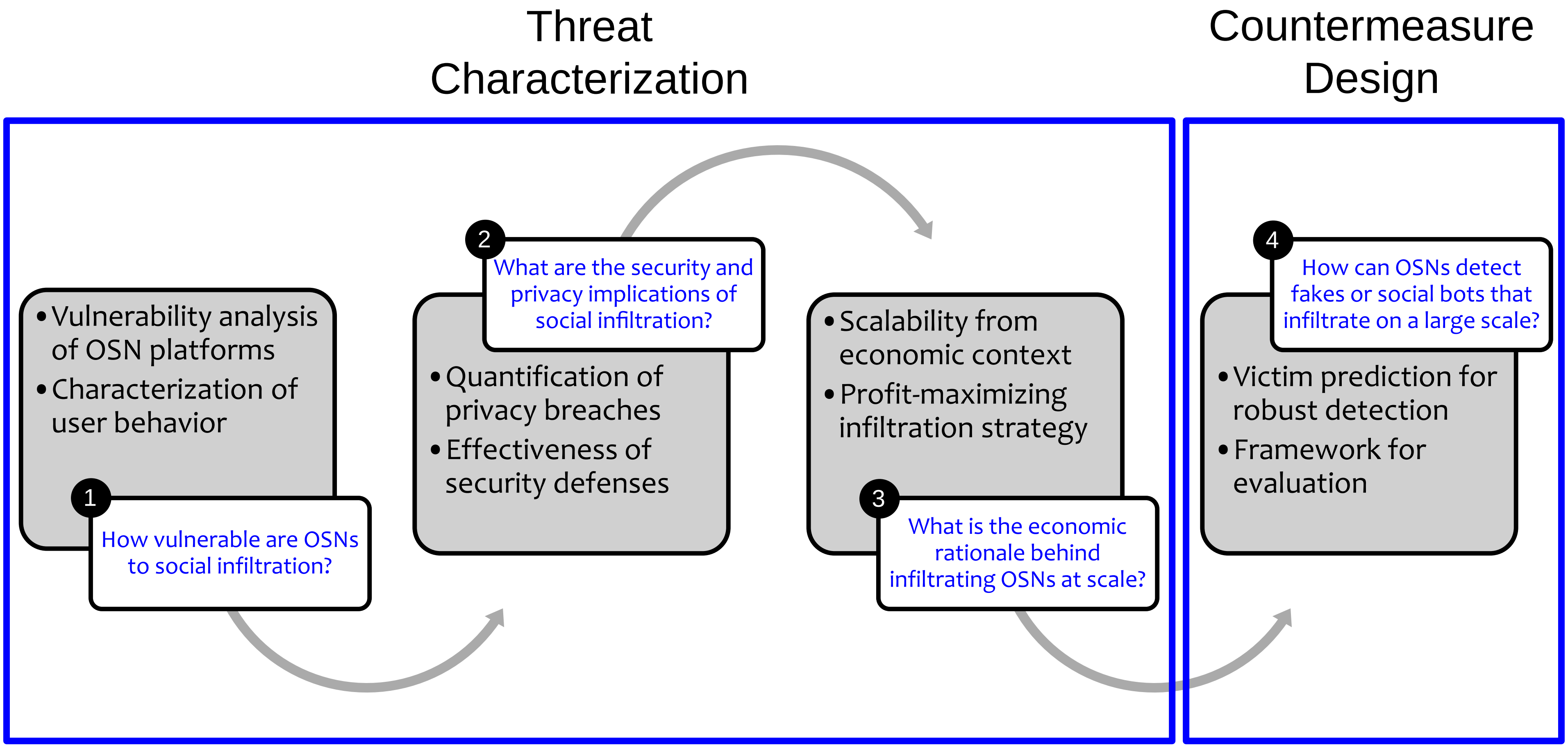



Precision at lower intervals




Precision at higher intervals

# Research Questions and Contributions



# Impact

## Public education & further studies

PCWorld

'Socialbots' Invade Facebook: Cull 250GB of Private Data

By John P. Mello Jr., PCWorld Nov 2, 2011 2:20 PM

BBC NEWS

Socialbots used by researchers to 'steal' Facebook data

Researchers have demonstrated a new technique capable of stealing personal information from Facebook.

2 November 2011 Last updated at 08:00 ET

IW InfoWorld

APRIL 08, 2013

Your Facebook friends may be evil bots

Computer scientists have unleashed hordes of humanlike social bots to infiltrate Facebook -- and they're awfully effective

By Eagle Gamma | InfoWorld

CBC news

Facebook easily infiltrated, mined for personal info

Socialbot network could mine 175 chunks of personal data per bot per day

## Production-class deployment

tuenti

## Open-source, public release

grafos

All you can Eat Giraph.

# Publications

**Primary:**

1. Boshmaf et al. *The socialbot network: When bots socialize for fame and money.* Proc. of ACSAC, Dec 2011 (20% acceptance rate, **best paper award**)

1. Boshmaf et al. *Key challenges in defending against malicious socialbots*. In Proc. of USENIX LEET, April 2012 (18% acceptance rate)

1. Boshmaf et al. *Design and analysis of a social botnet*. J. Comp. Net., 57(2), Feb 2013 (1.9 impact factor)

1. Boshmaf et al. *Graph-based Sybil detection in social and information systems*. In Proc. of ASONAM, Aug 2013 (13% acceptance rate, **best paper award**)

**Related:**

1. Boshmaf et al. *The socialbot network: are social botnets possible?* ACM Interactions, March-April, 2012

1. Sun et al. *A billion keys, but few locks: The crisis of web single sign-on*. In Proc. of NSPW, Sept 2010

1. Rashtian et al. *To befriend or not? A model for friend request acceptance on Facebook*. In Proc. of SOUPS, July 2014