Thwarting fake accounts by predicting their victims

Yazan Boshmaf
University of British Columbia

Dionysios Logothetis, Georgios Siganos
Telefonica Research

Matei Ripeanu, Konstantin Beznosov
University of British Columbia

AAAI 2014 Spring Symposium: Social Hacking and Cognitive Security on the Internet and New Media

a place of mind

TELEFÓNICA
INVESTIGACIÓN Y DESARROLLO

Improving fake account detection in OSNs by predicting potential victims

Why are users accounts valuable?

Yazan Boshmaf
University of British Columbia

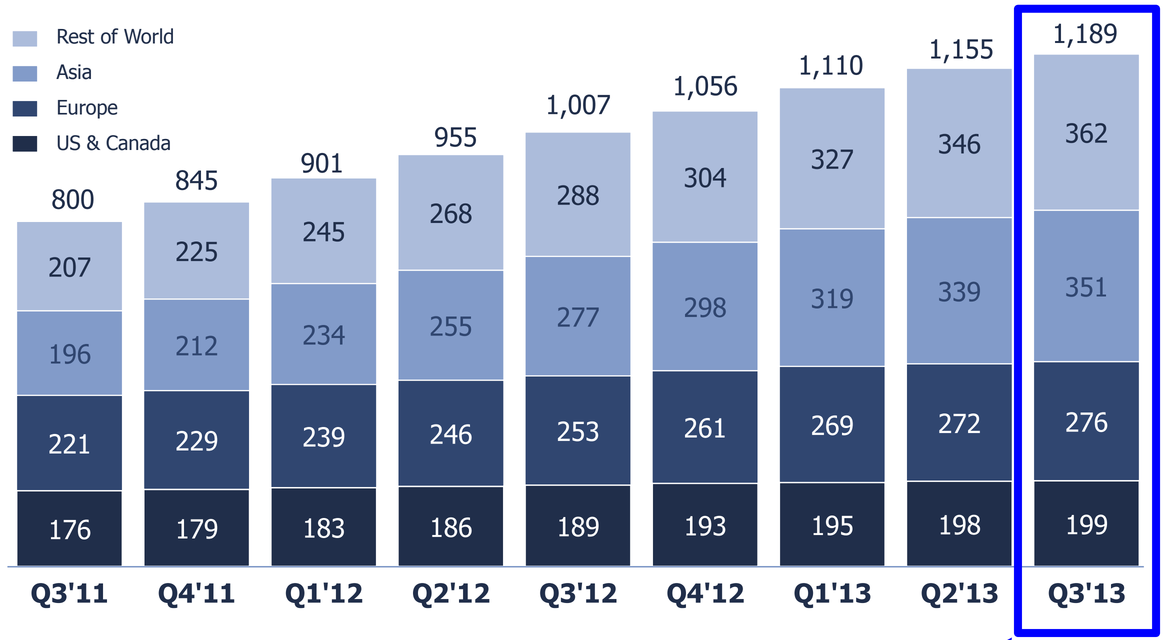
Dionysios Logothetis, Geogios Siganos
Telefonica Research

Matei Ripeanu, Konstantin Beznosov
University of British Columbia

Submitted to the 18th International Conference on Financial Cryptography and Data Security (FC'14)

User accounts are assets

Average MAU* in Facebook (Millions)**

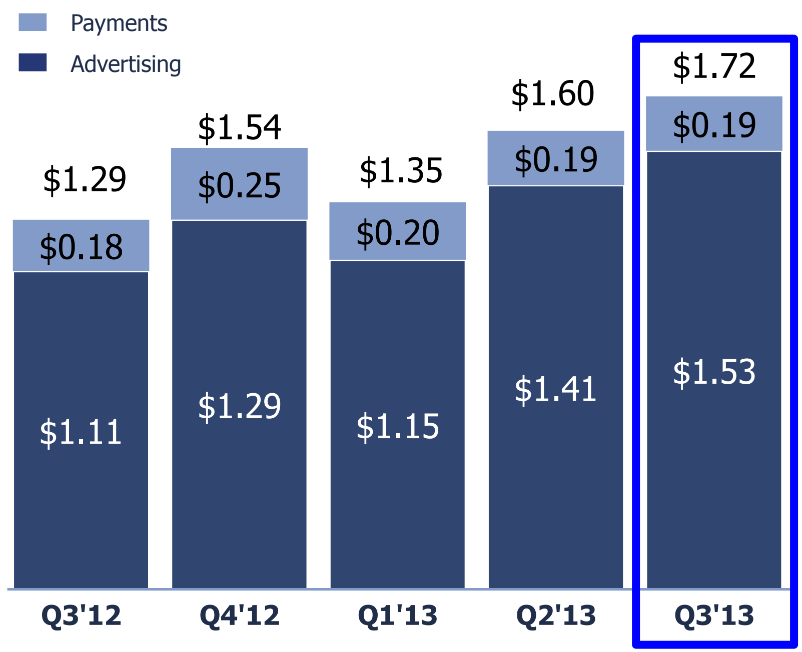


1.2 billion MAU in Q3'13

* Monthly active user (MAU): The basic user metric in Facebook
** Facebook Quarterly Reports, Facebook Investor Relations: http://investor.fb.com

User accounts generate revenue

Average revenue per Facebook user*



2 billion US dollars in Q3'13

* Facebook Quarterly Reports, Facebook Investor Relations: http://investor.fb.com

User accounts generate revenue

Average revenue per Facebook user*

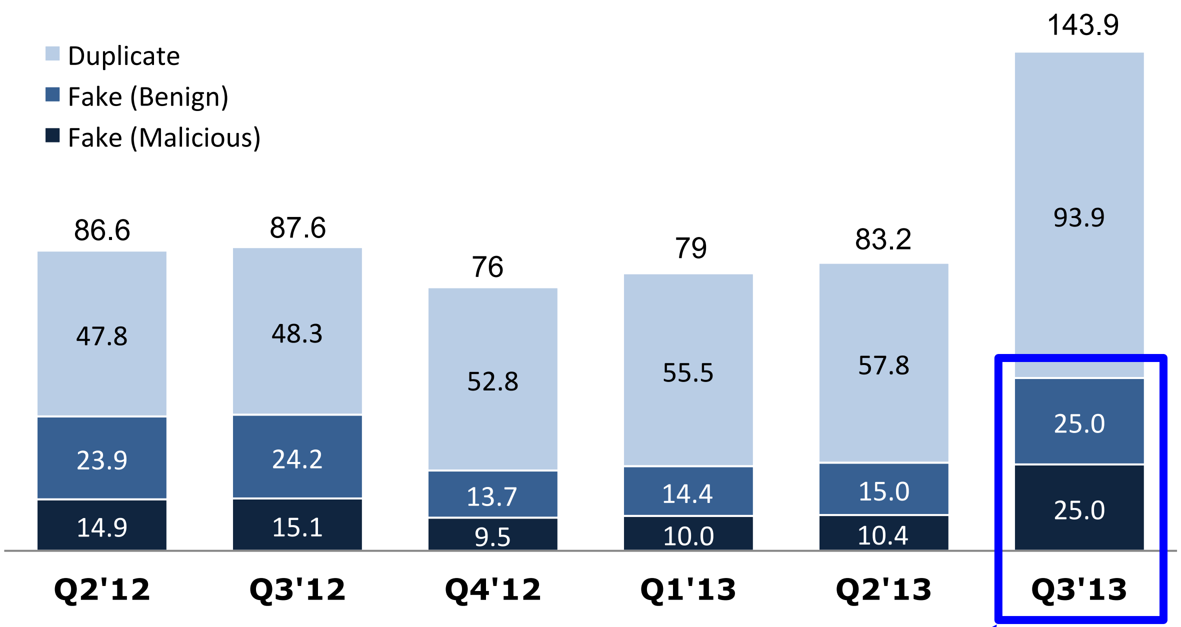
How many user accounts are fake?

2 billion US dollars in Q3'13

* Facebook Quarterly Reports, Facebook Investor Relations: http://investor.fb.com

Fake accounts are rising

Undesirable* accounts in Facebook (Millions)**

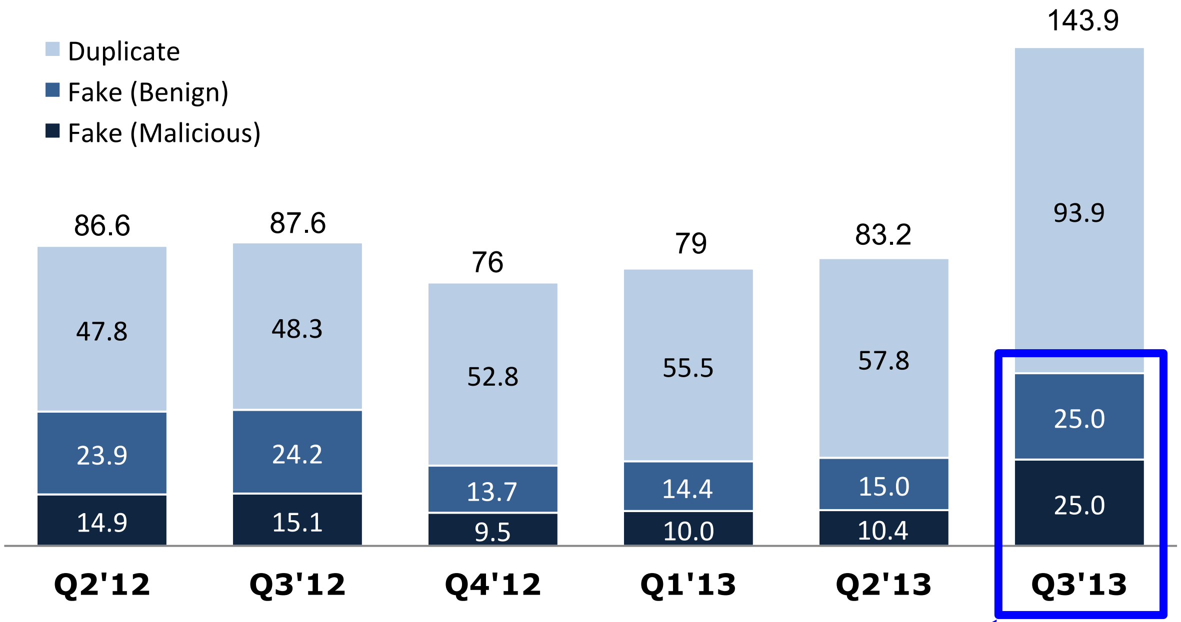


50 million fakes in Q3'13

* Undesirable Facebook accounts include both duplicates and fake accounts (worst case estimates)
** Facebook Quarterly Reports, Facebook Investor Relations: http://investor.fb.com

Fake accounts are rising

Undesirable* accounts in Facebook (Millions)**

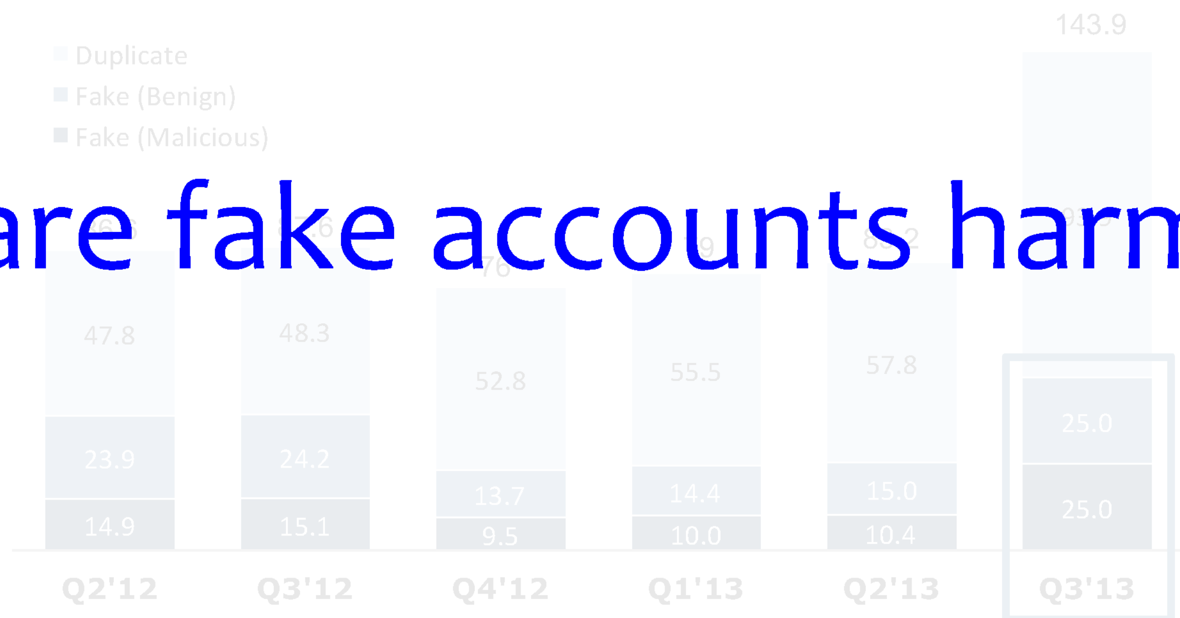


17.4 thousand fakes per hour on average

* Undesirable Facebook accounts include both duplicates and fake accounts (worst case estimates)
** Facebook Quarterly Reports, Facebook Investor Relations: http://investor.fb.com

Fake accounts are rising

Undesirable* accounts in Facebook (Millions)**



Why are fake accounts harmful?

17.4 thousand fakes removed per hour

* Undesirable Facebook accounts include both duplicates and fake accounts
** Facebook Quarterly Reports, Facebook Investor Relations: http://investor.fb.com

Fake accounts are bad for business

CBC

CBCnews | Technology & Science

Facebook shares drop on news of fake accounts

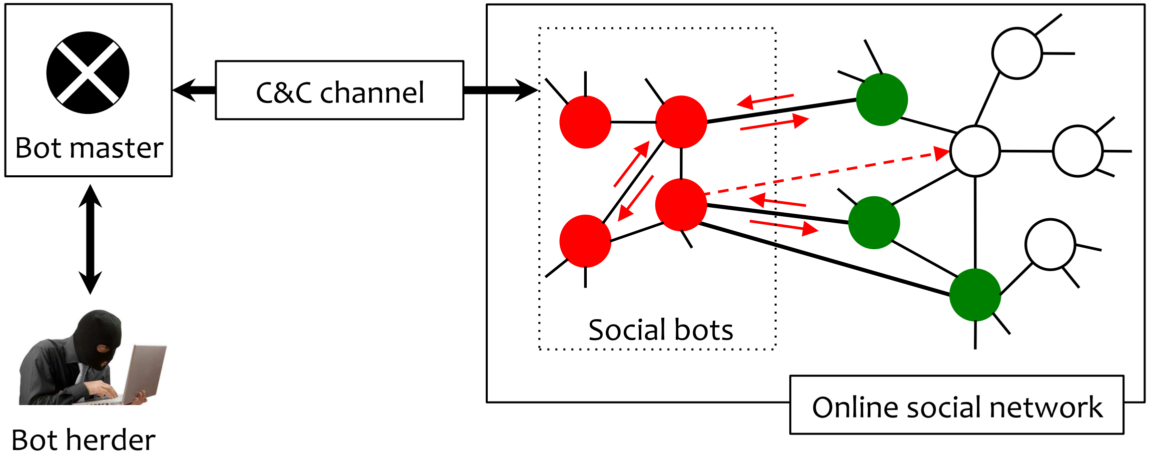
83 million accounts false or duplicates, company reveals

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 content*



Free infrastructure to steal data, spread malware & misinform

* Boshmaf et al. Design and analysis of a social botnet. Computer Networks, 2013.

Bad for users is bad for business

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

c|net

Koobface virus hits Facebook

An e-mail lure and a fake Adobe Flash update request could load a nasty virus on your PC.

by Robert Vamosi | December 4, 2008 4:36 PM PST

PCWorld

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

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

Fake accounts are bad for users

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

What's the role of fake accounts in today's underground economy?

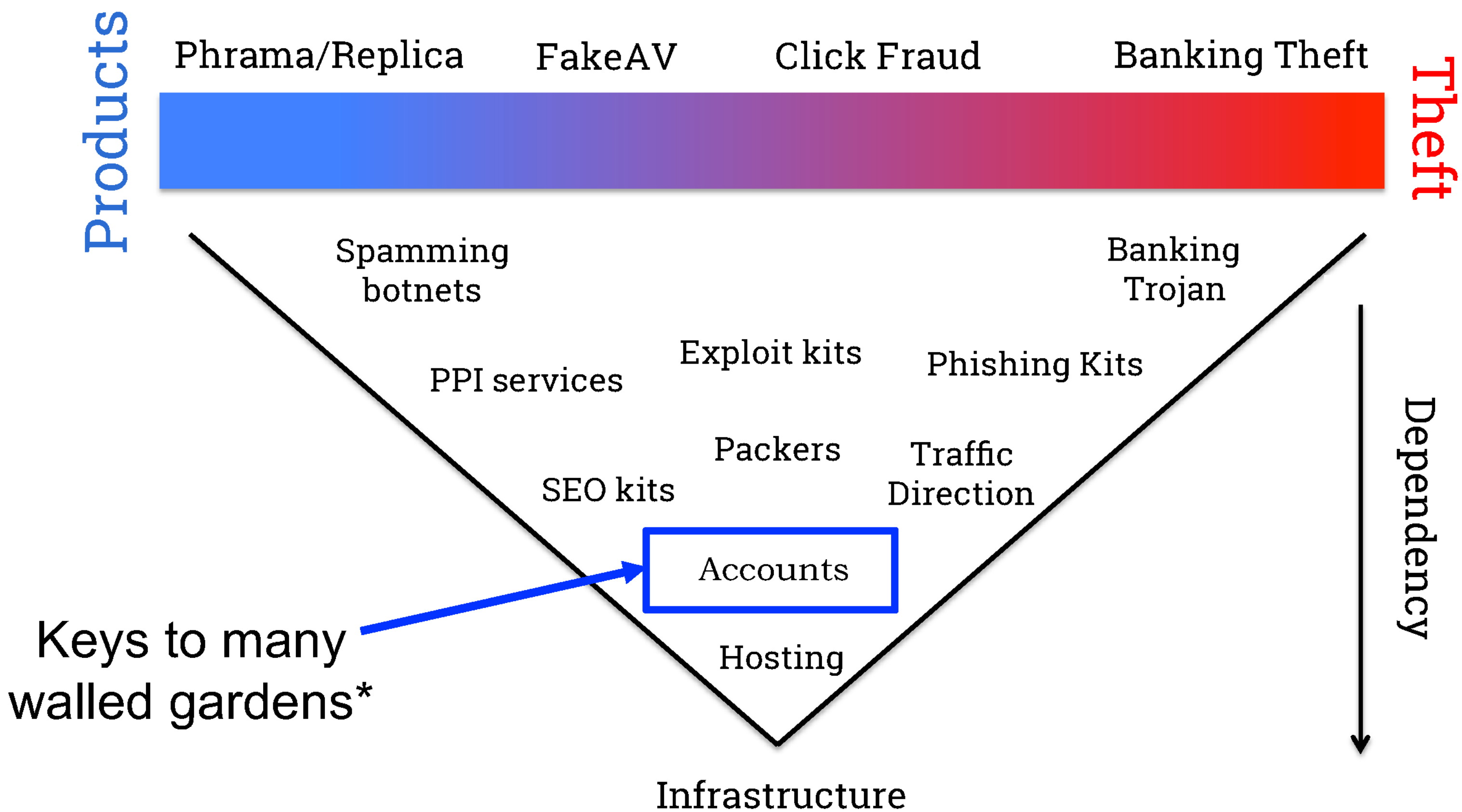
Koobface virus hits Facebook

by Robert Vamosi | December 4, 2008 4:36 PM PST

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

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

Fake accounts are market enablers



* Thomas et al. The role of the underground market in OSN spam and abuse. Usenix Security, 2013.

Fake account are profitable "commodity"

Prices from BuyAccs.com
СЕРВИС РЕГИСТРАЦИИ АККАУНТОВ

Web Service	Price per Thousand
Hotmail.com, resale*	$2.00
Hotmail.com	$4.00
Yahoo	$6.00
Twitter	$20.00
Google (PVA)**	$100.00
Facebook (PVA)**	$100.00

Already a multi-million dollar business

* Resale indicates account was previously used in another activity
** Phone Verified Accounts: A fake account verified by a text challenge-response using a cell phone

Fake account are profitable “commodity”

Prices from BuyAccs.com
СЕРВИС РЕГИСТРАЦИИ АККАУНТОВ

Web Service	Price per Thousand
Hotmail.com, resale*	$2.00
Hotmail.com	$4.00
Yahoo	$6.00
Twitter	$20.00
Google (PVA)**	$100.00
Facebook (PVA)**	$100.00

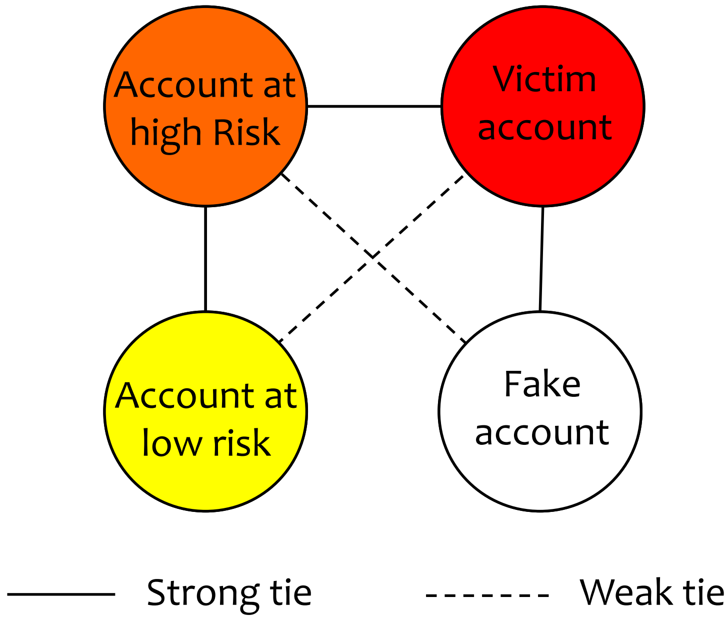
How do OSNs fight against fakes?

Already a multi-million dollar business

* Resale indicates account was previously used in another activity
** Phone Verified Accounts: A fake account verified by a text challenge-response using a cell phone

Threat model

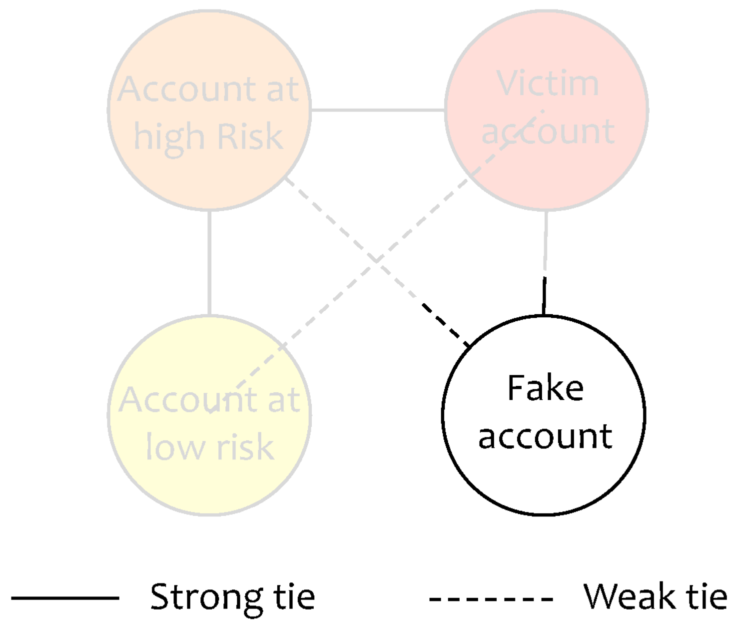
Attackers can create and control fakes in a botnet-like fashion



Attackers first infiltrate the OSN then mount subsequent attacks

Fake-centered security paradigm

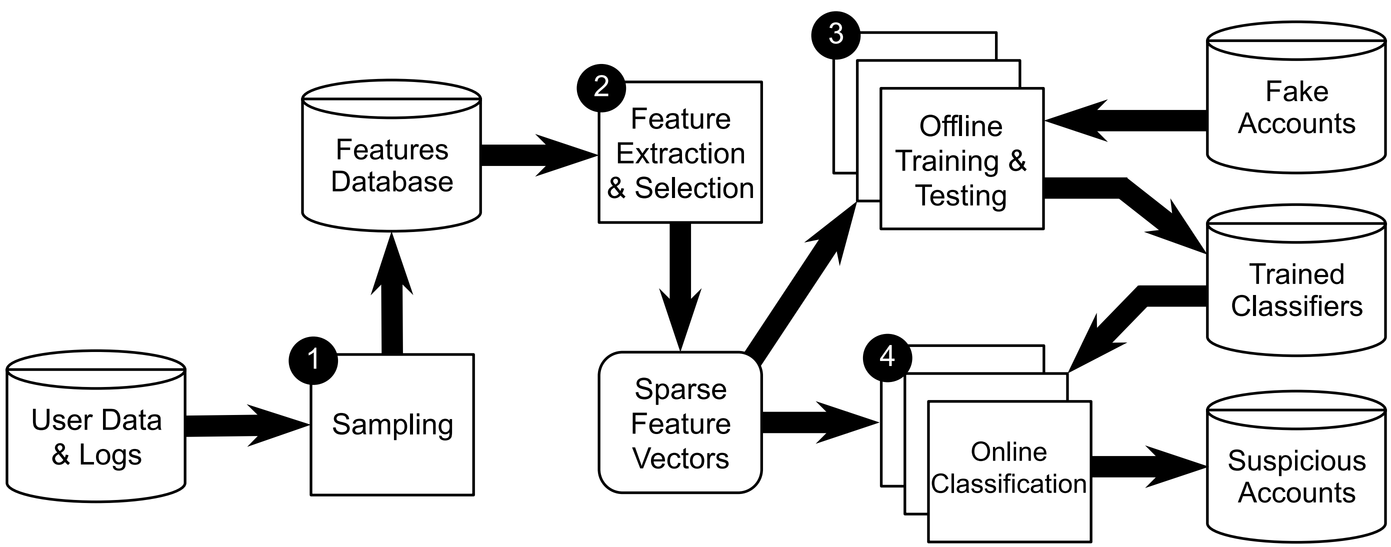
Detect fake account by identifying what "fakeness looks like"



Defined by anomalies in social content or structure

Feature-based detection

Identifies suspicious accounts using supervised machine learning



Relies on features extracted from real and fake accounts

* Stein et al. Facebook Immune System. EuroSys SNS, 2011

Which features to use?

Interactions?

Pictures?

Triadic closure?

Friends?

Ad clicks?

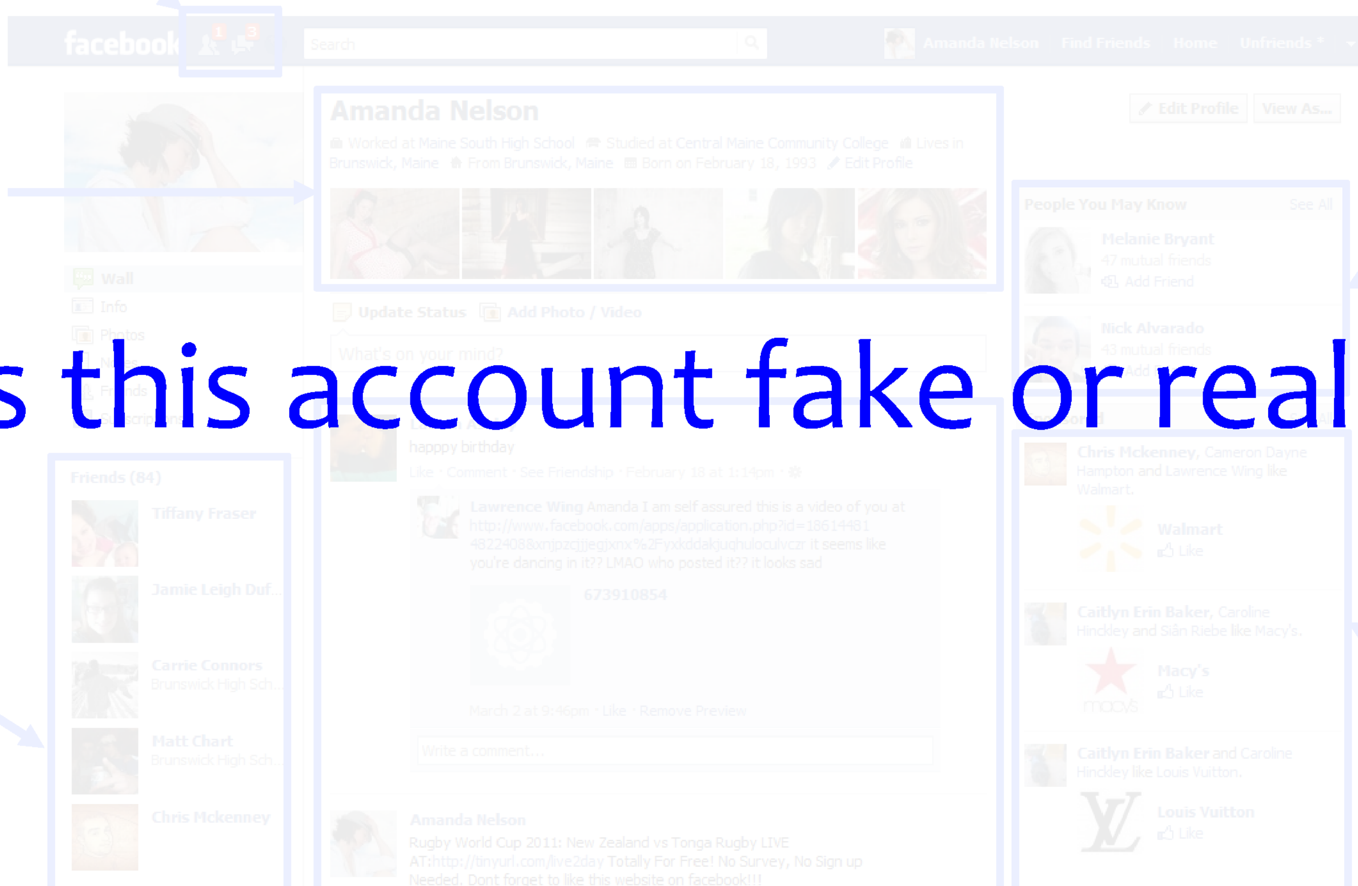
Posts?

Which features to use?

Interactions?

Pictures?

Triadic closure?



Friends?

Ad clicks?

Posts?

Is this account fake or real?

* Barreno et al. The security of machine learning. J. on Machine Learning, 2010

Fake accounts ≈ real accounts

Interactions?

Pictures?

Friends?

Triadic closure?

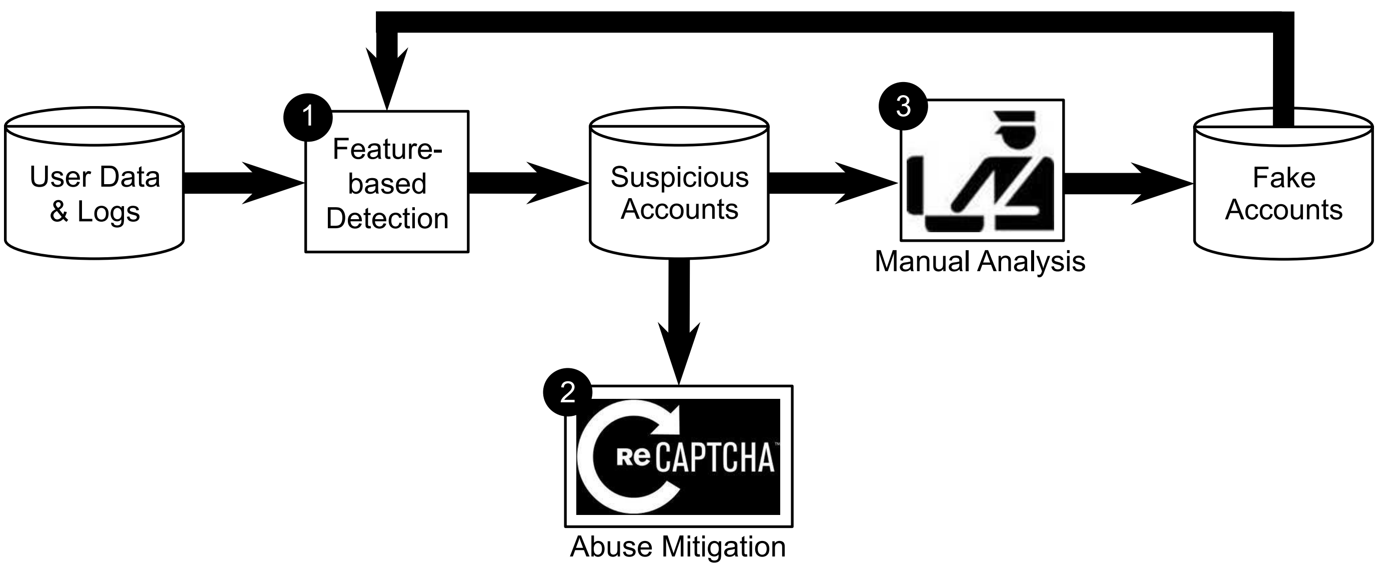
Ad clicks?

Posts?

* Barreno et al. The security of machine learning. J. on Machine Learning, 2010

How to build a ground-truth?

Analysts verify suspicious accounts and update ground-truth

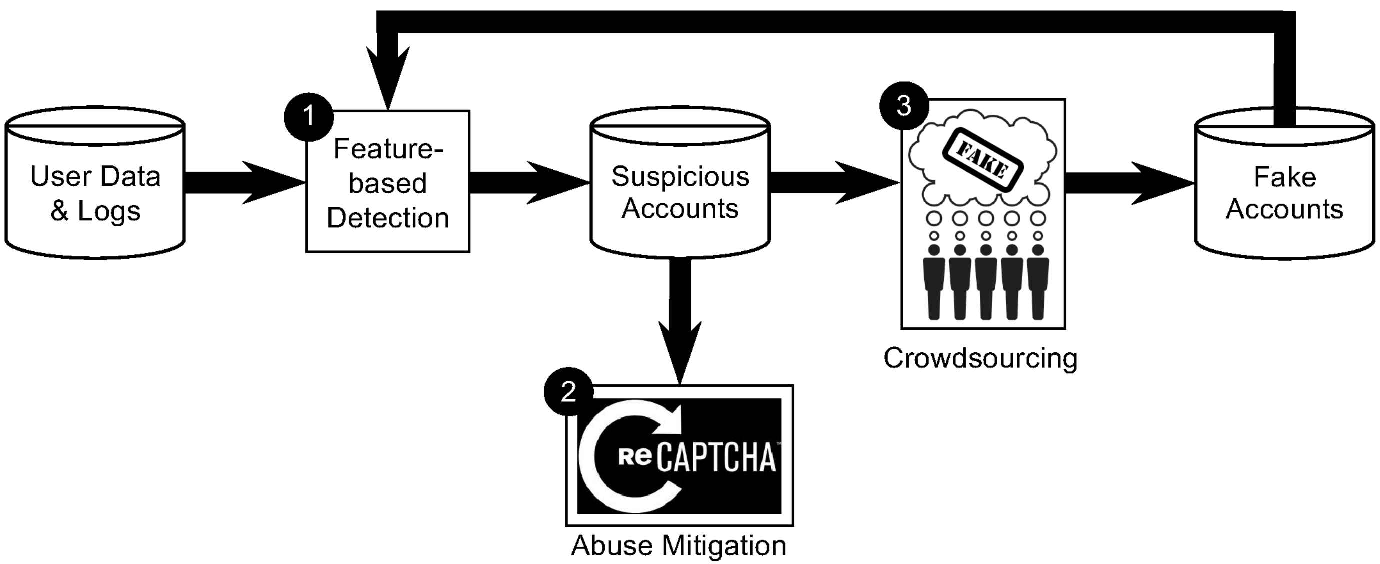


Roadblocks to “quarantine” highly-suspicious accounts

* Stein et al. Facebook Immune System. EuroSys SNS, 2011

How to build a ground-truth?

Internet crowds verify suspicious accounts to update ground-truth



Roadblocks to “quarantine” highly-suspicious accounts

* Wang et al. Social Turing Tests: Crowdsourcing Sybil Detection. NDSS, 2013

Fake accounts ≈ real accounts

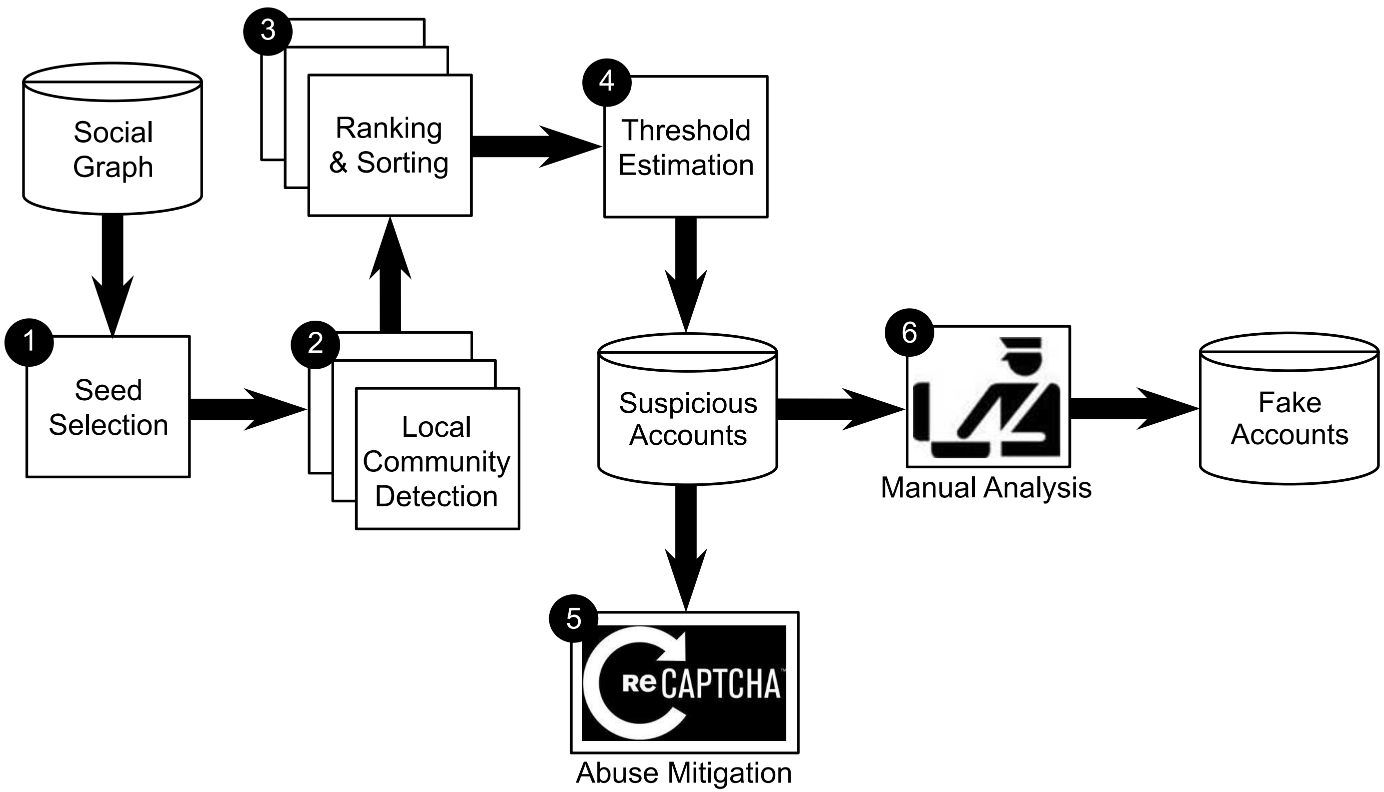
- ☐ Proactive protection
- ☑ Near real-time responses
- ☑ Scales to millions of users
- ☐ Hard to circumvent
- ☐ Accurate detection
- ☐ Provably secure

What else can we do?

* Stein et al. Facebook Immune System. EuroSys SNS, 2011

Graph-based detection

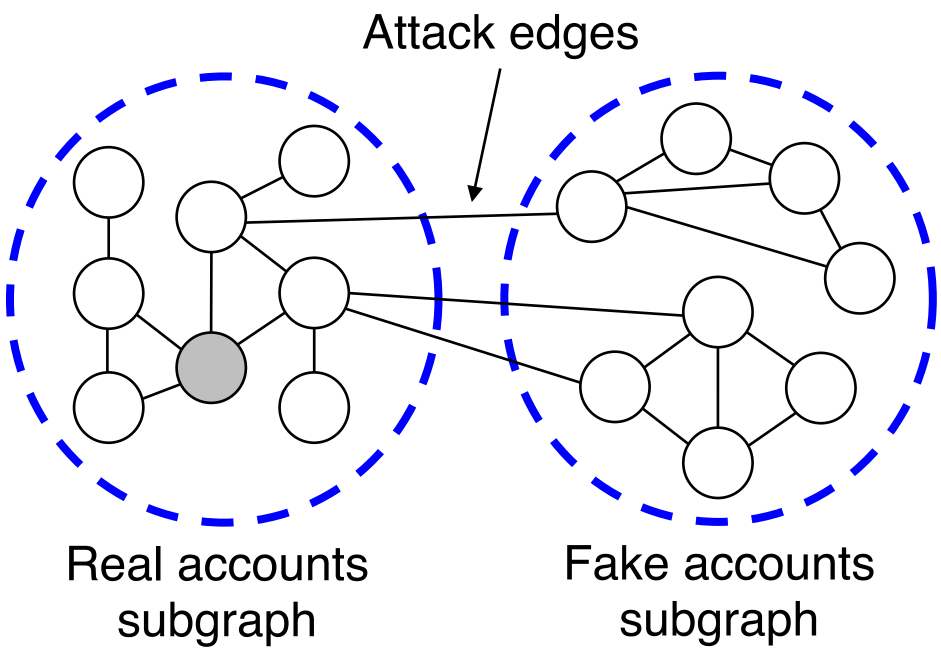
Identifies suspicious accounts using (network) graph analysis



Relies on the structural properties of real and fake accounts

* Boshmaf et al. Graph-based Sybil detection in social and information systems. ASONAM, 2013.

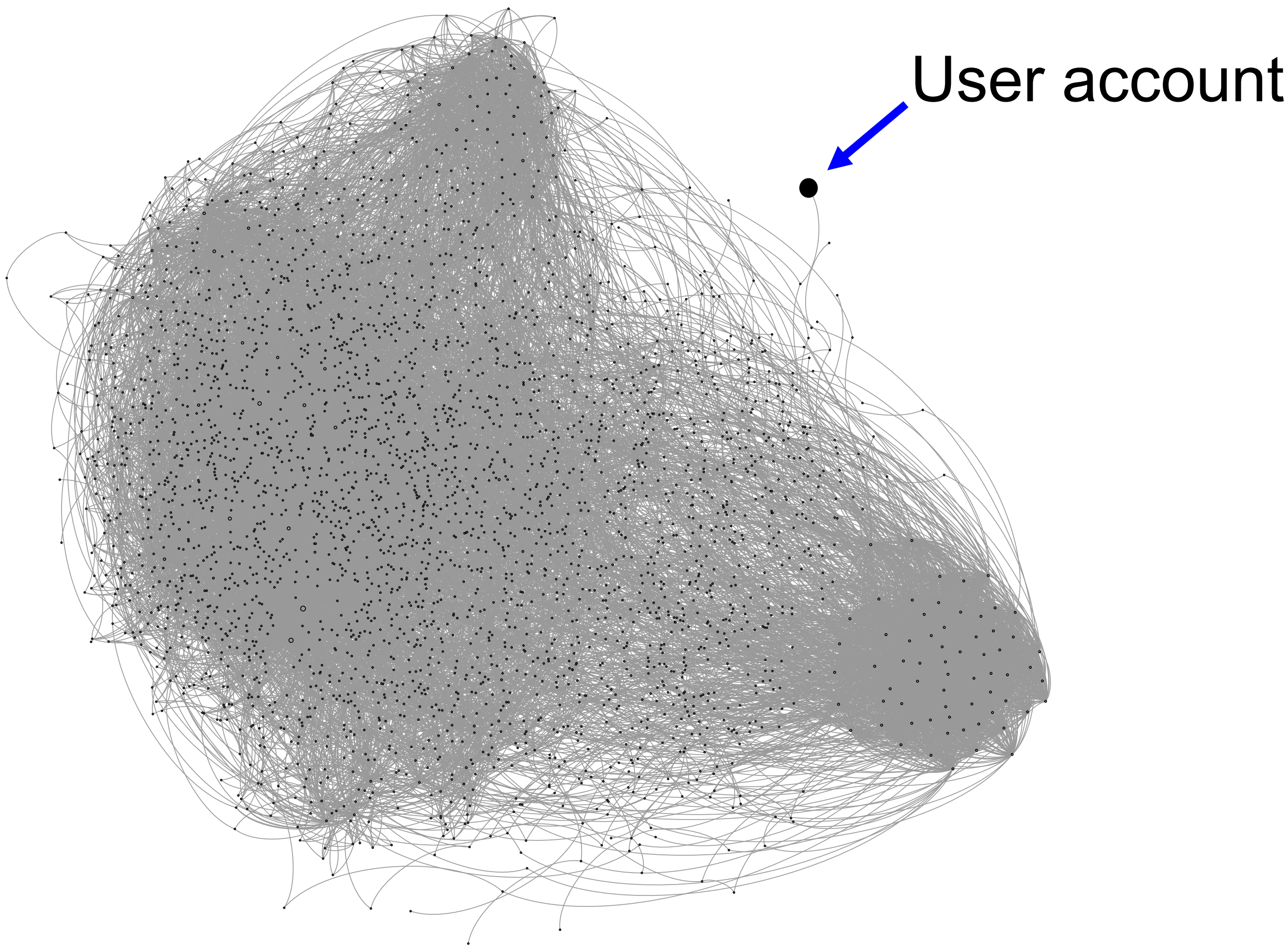
Which structural properties?



Find a (provably) sparse cut between the regions

* Spielman et al. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. ACM Theory of computing, 2004.

What about real-world graphs?



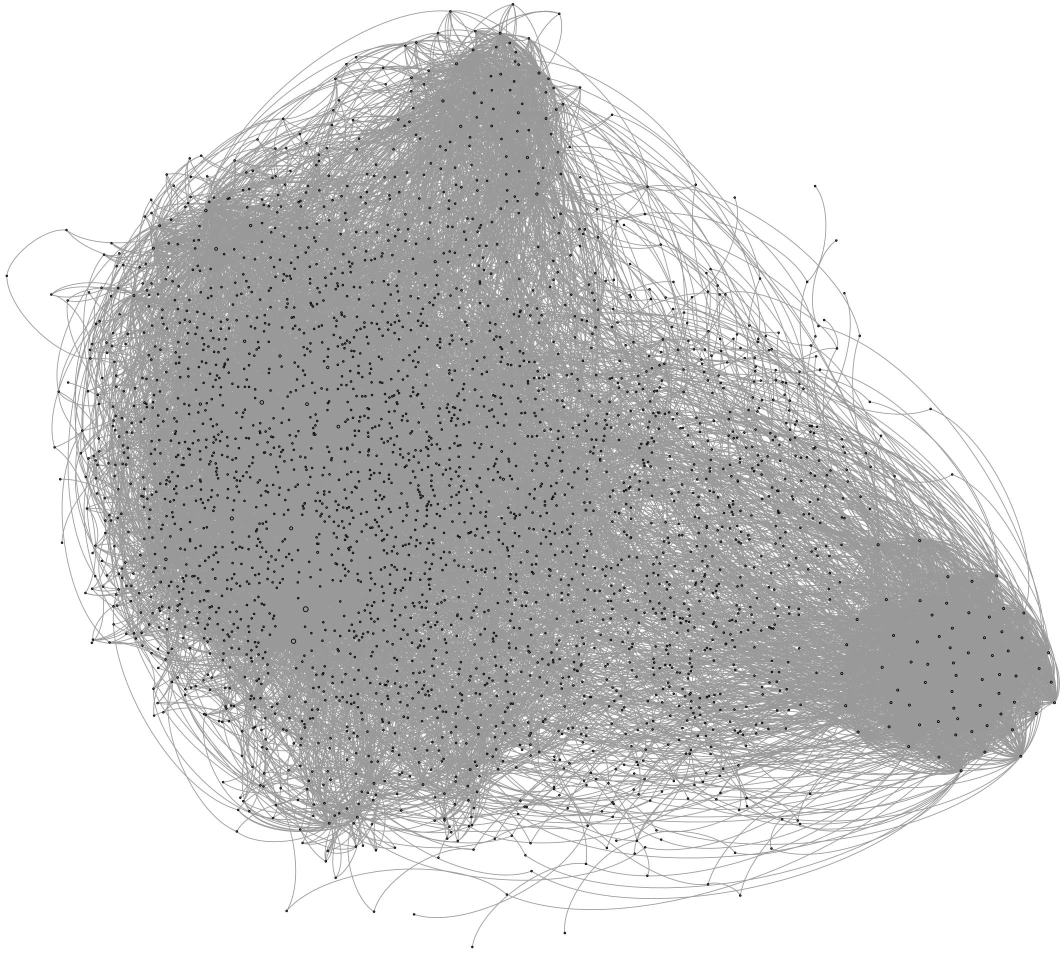
A Facebook community of 2,991 user accounts

What about real-world graphs?

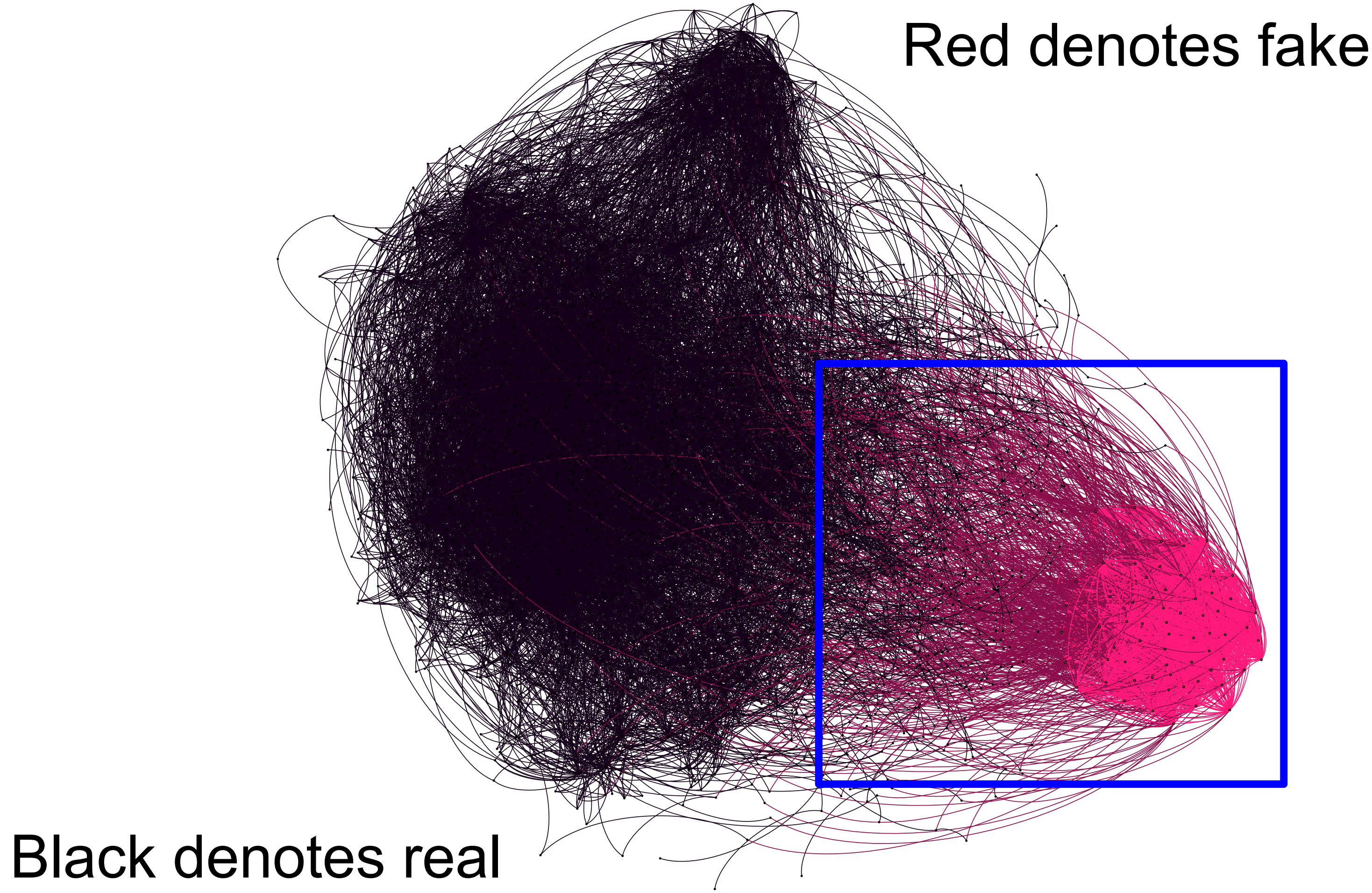
Is the community infiltrated?

A Facebook community of 2,991 user accounts

No sparse cut ≈ no fake accounts

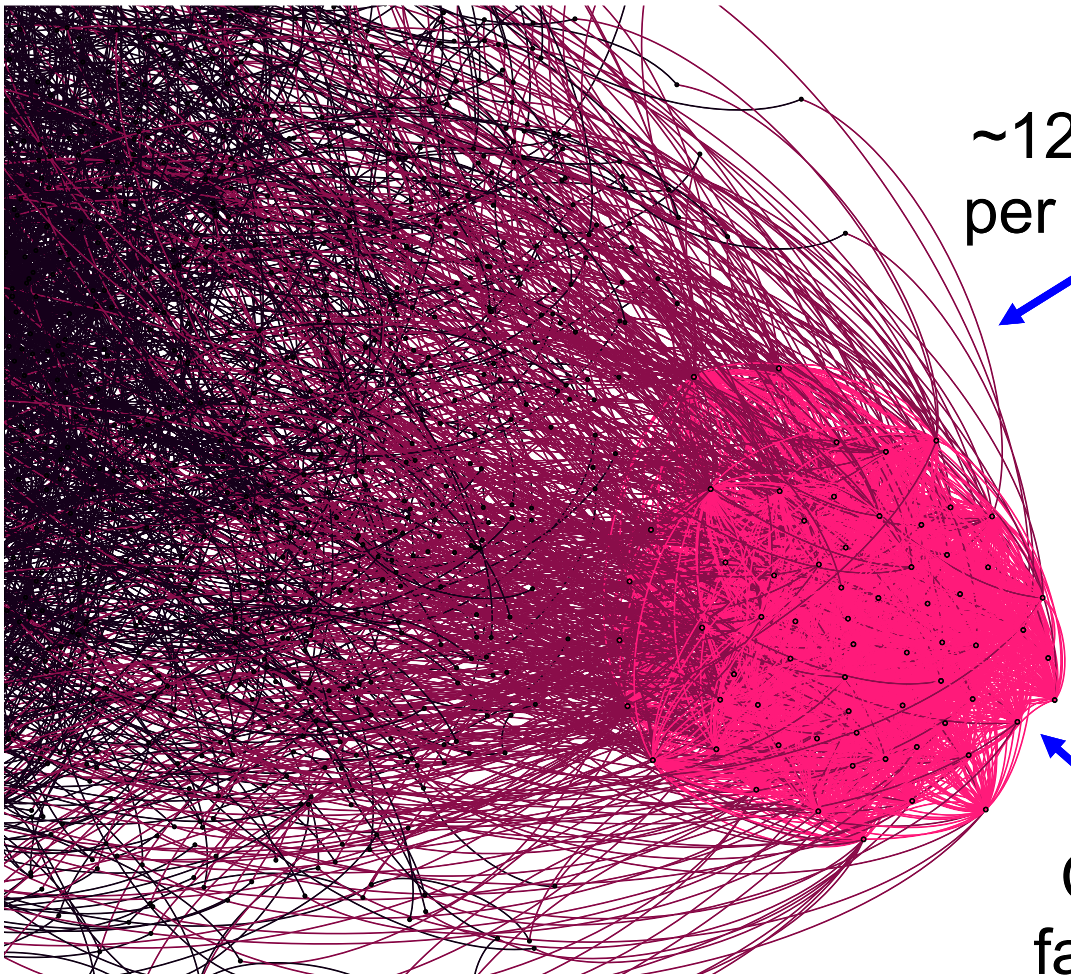


But users are easily deceived…



* Boshmaf et al. The socialbot network: When bots socialize for fame and money. ACSAC, 2011.

But users are easily deceived…



~12 attack edge per fake account

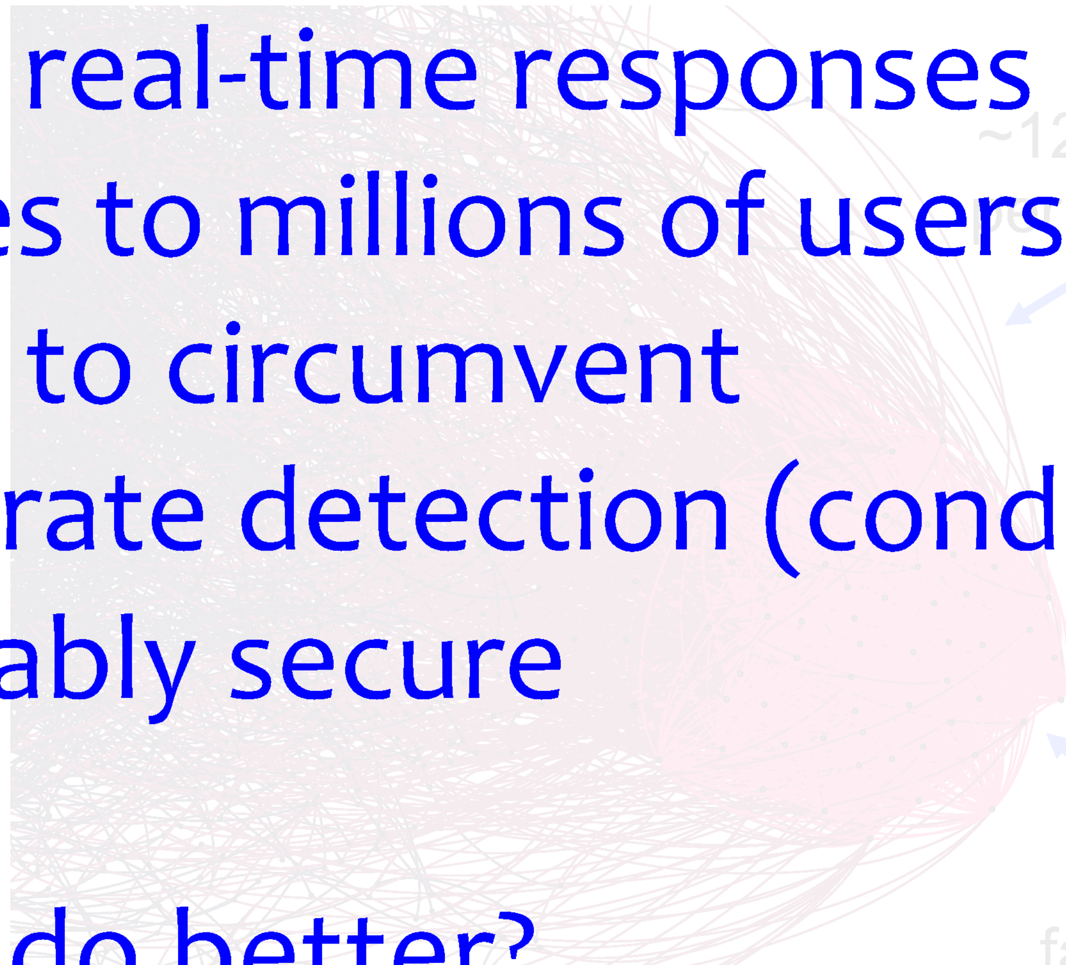
Clique of 65 fake accounts

* Boshmaf et al. The socialbot network: When bots socialize for fame and money. ACSAC, 2011.

Users are easily deceived

- ☐ Proactive protection
- ☐ Near real-time responses
- ☑ Scales to millions of users
- ☐ Hard to circumvent
- ☑ Accurate detection (conditional)
- ☑ Provably secure

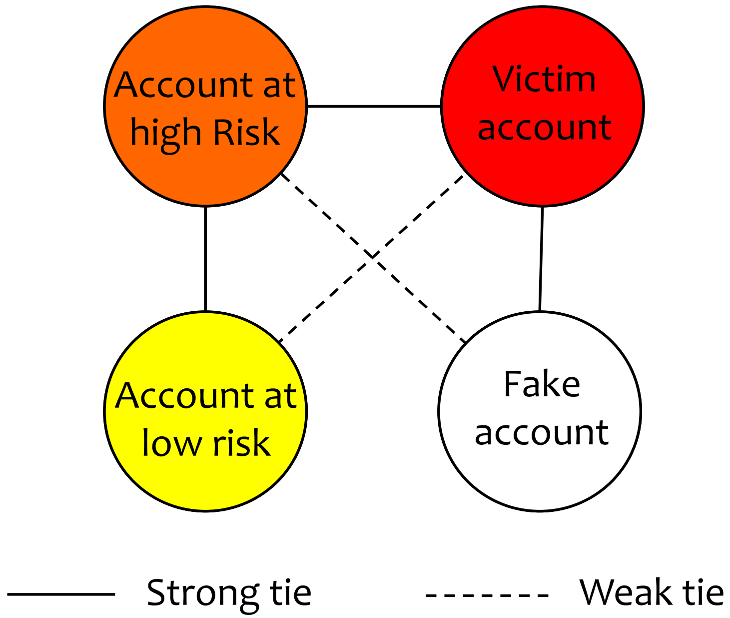
Can we do better?



* Boshmaf et al. The socialbot network: When bots socialize for fame and money. ACSAC, 2011.

Victim-centered security paradigm

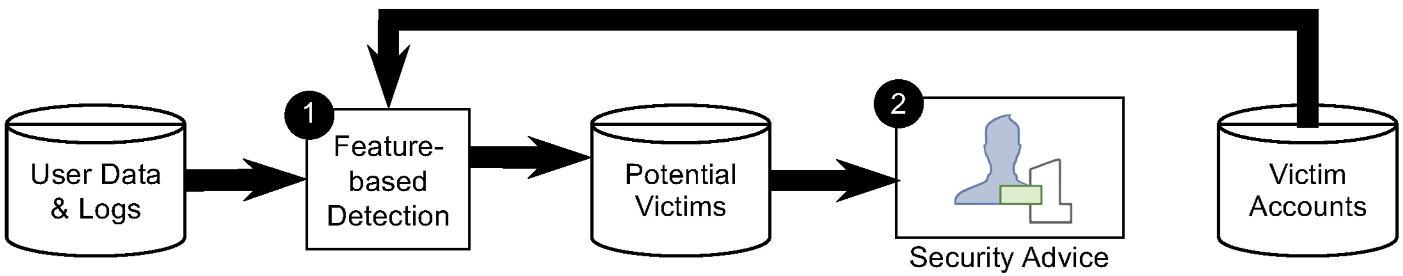
Detect fake accounts by first identifying their (potential) victims



This leads to a more resilient defense mechanism (epidemiology?)

SybilPredict in a nutshell

Predicts victims who (are likely to) have attack edges with fakes in O(n logn) time



Pros:

- Proactive protection
- Near real-time responses
- Scales to millions of users
- Hard to circumvent

Cons:

- Doesn't identify fakes
- Introduces usability problems
- Not provably secure

SybilPredict in a nutshell

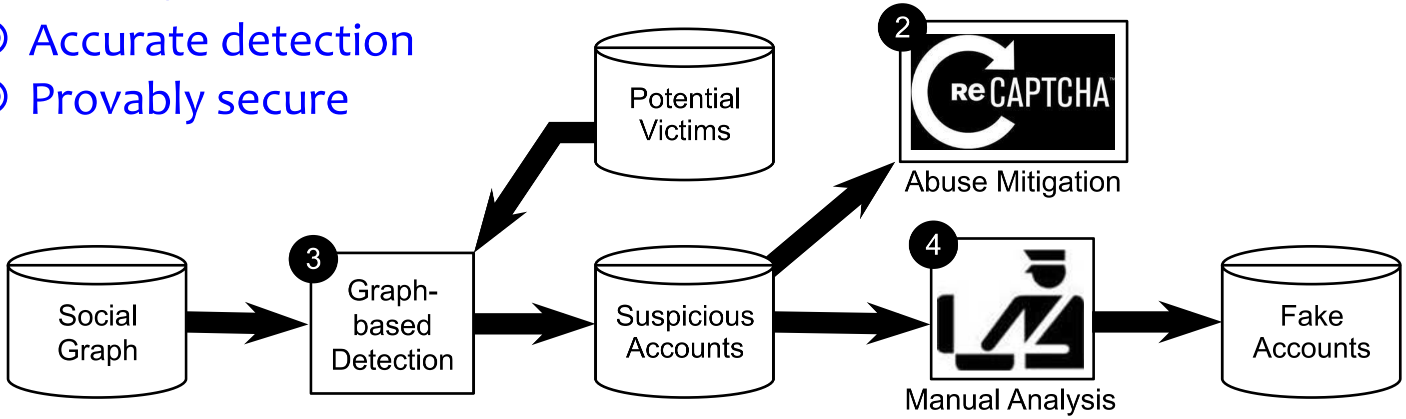
Embeds predictions into graph to identify suspicious accounts in $O(n \log n + m)$ time

Pros:

- Scales to millions of users
- Hard to circumvent
- Accurate detection
- Provably secure

Cons:

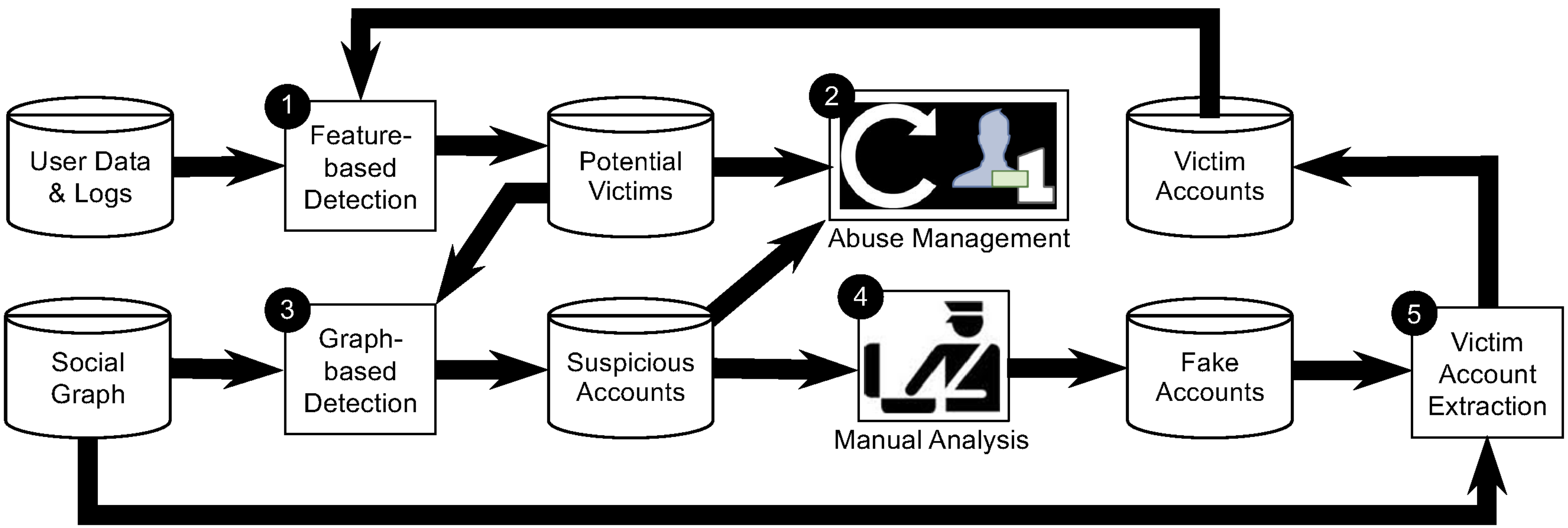
- Retroactive detection
- Batch processed



Uses short random walks biased against identified victims to rank users

SybilPredict in a nutshell

Uses distributed machine learning and graph processing infrastructure



Runs in O(n logn + m) time end-to-end

We claim SybilPredict is:

- ☑ Proactive protection
- ☑ Near real-time responses
- ☑ Scales to millions of users
- ☑ Hard to circumvent
- ☑ Accurate detection
- ☑ Provably secure

Challenges and research directions

* Boshmaf et al. Key challenges in defending against malicious socialbots. Usenix LEET, 2012

More info?

Fork or clone SybilPredict now: https://grafos.ml

On-going deployment at tuenti

For SybilPredict technical report, please email at boshmaf@ece.ubc.ca

More info?

Fork or clone SybilPredict now: https://grafos.ml

grafos

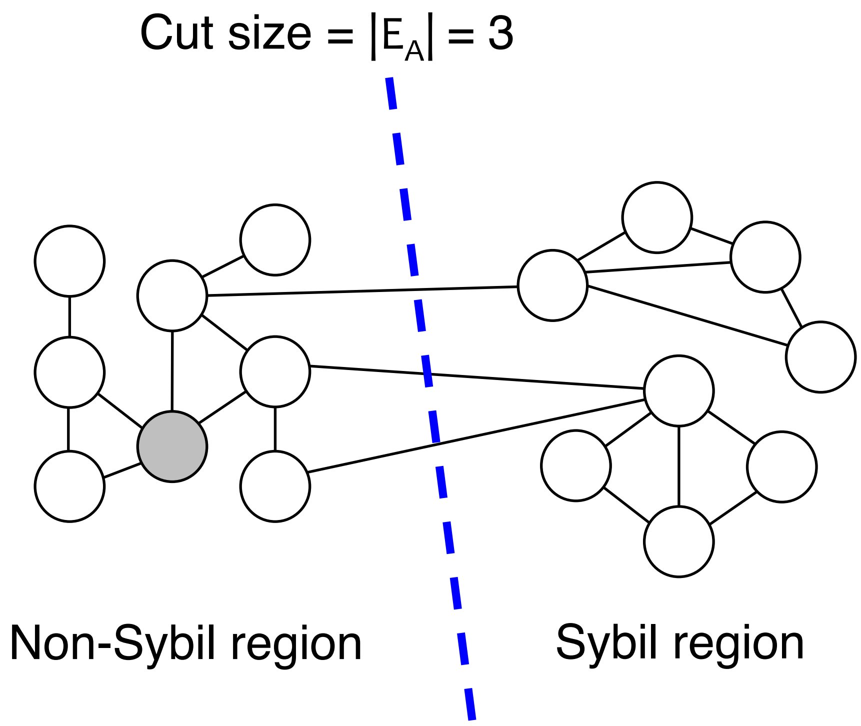
All you can Eat Giraph.

Details: Example & prelim results

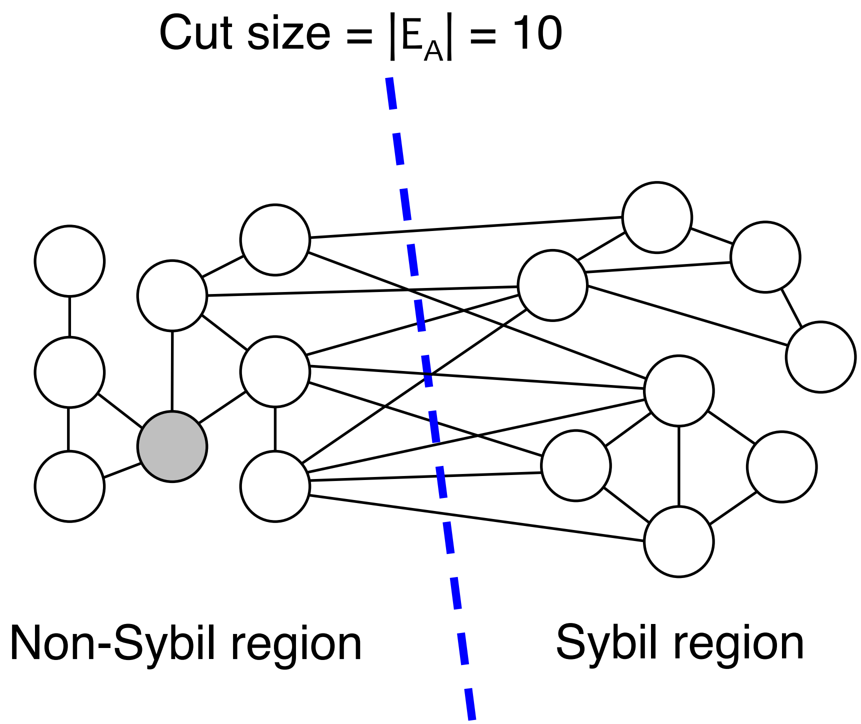
On-going deployment at tuenti

For latest technical report, please email at boshmaf@ece.ubc.ca

How does it work?



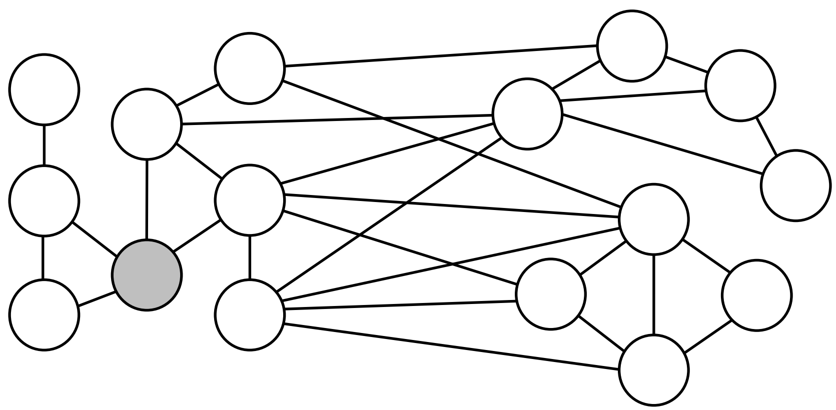
Graph-based detection fails now



Idea: Artificially prune attack edges based of victim prediction

Victim prediction

Trusted
Inspected
Victim
Not-victim

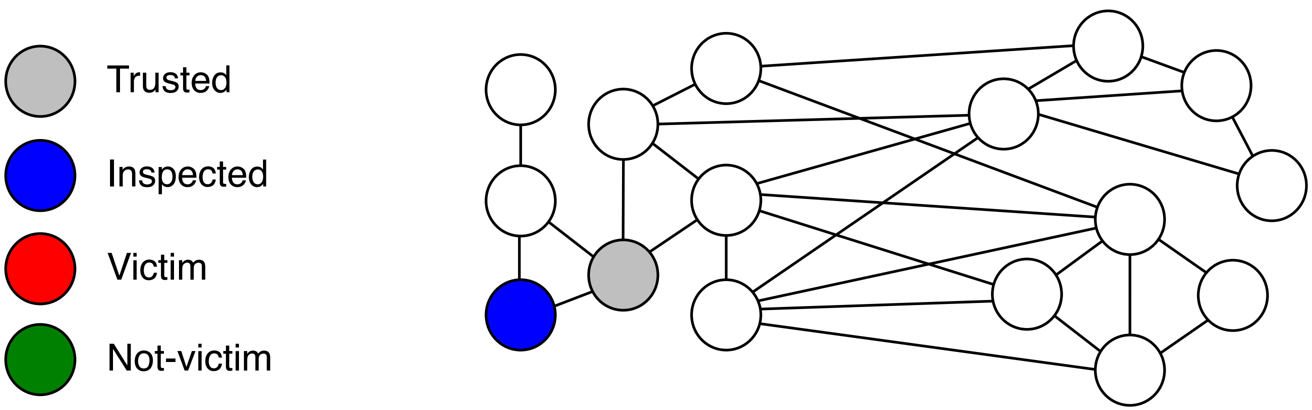


TP	TN	FP	FN
0	**0**	**0**	**0**

Victim prediction

#Friends	#Photos		#Posts	Victim?
1	54	…	22	0

Feature vector | Target variable



TP	TN	FP	FN
0	1	0	0

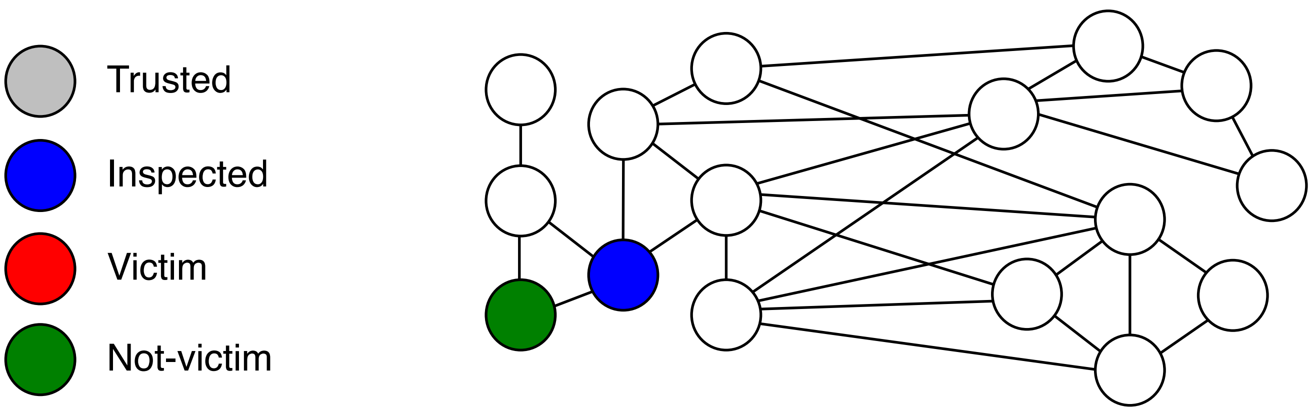
Victim prediction

#Friends	#Photos		#Posts
4	42	...	33

Feature vector

Victim?
0

Target variable



TP	TN	FP	FN
0	**2**	**0**	**0**

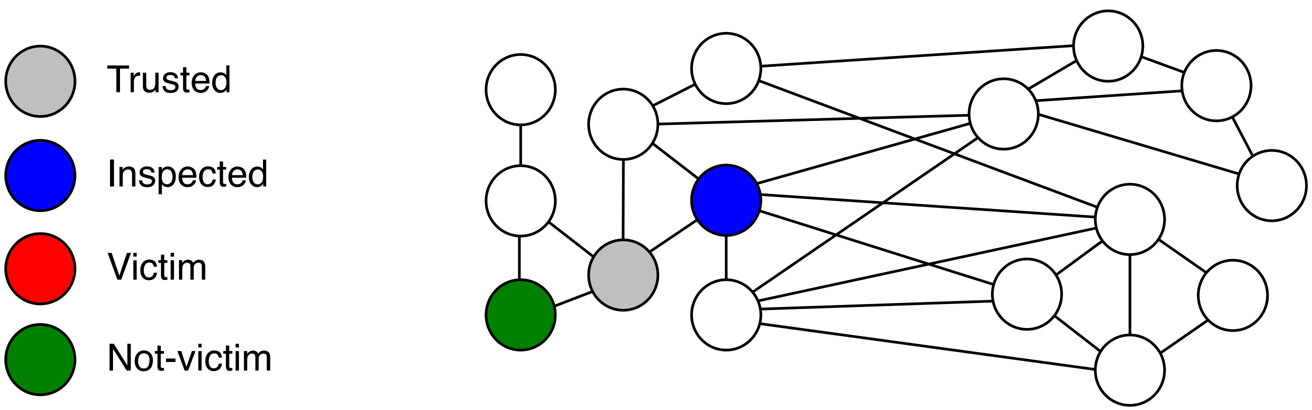
Victim prediction

#Friends	#Photos		#Posts
6	512	···	1024

Feature vector

Victim?
1

Target variable

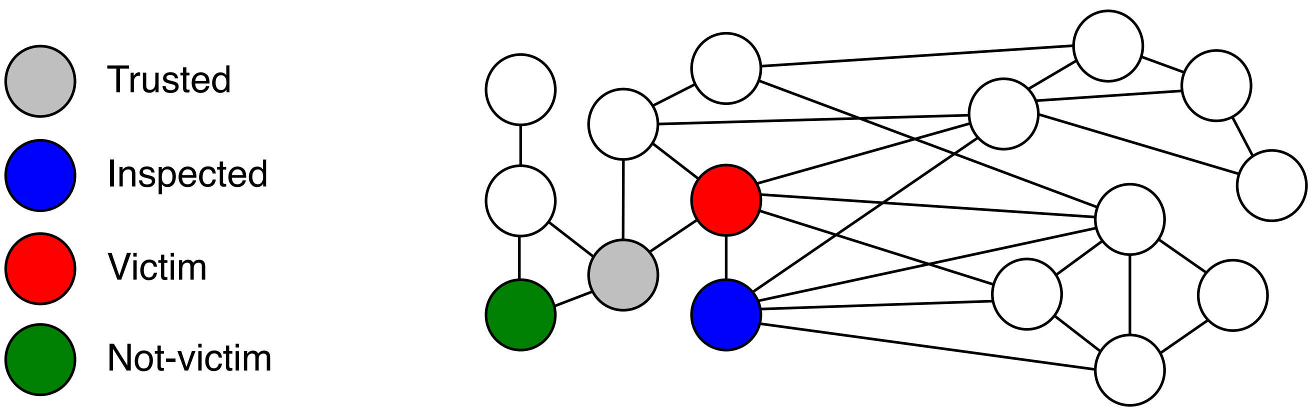


TP	TN	FP	FN
1	2	0	0

Victim prediction

#Friends	#Photos		#Posts	Victim?
5	476	...	688	**1**

Feature vector — Target variable



TP	TN	FP	FN
2	**2**	**0**	**0**

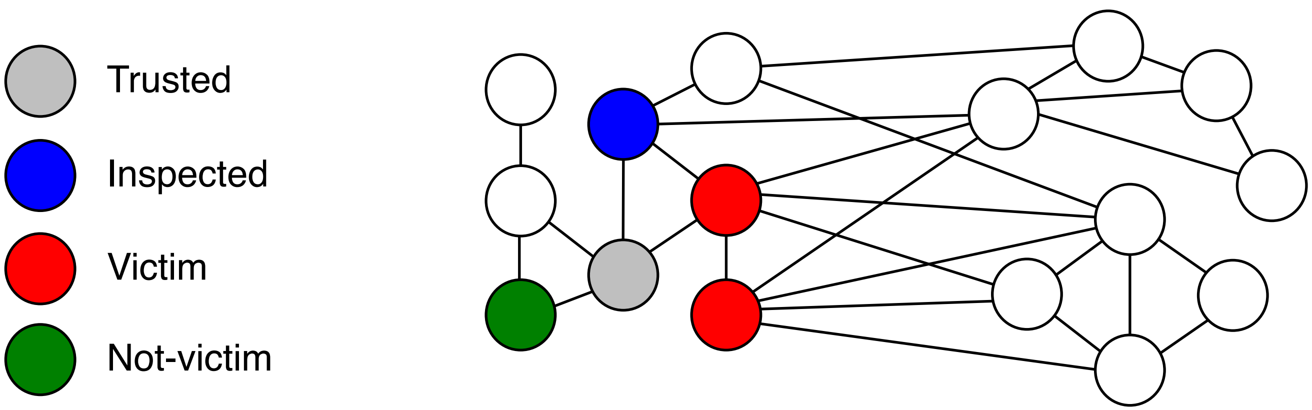
Victim prediction

#Friends	#Photos		#Posts
4	564	⋯	2033

Feature vector

Victim?
1

Target variable



TP	TN	FP	FN
3	2	0	0

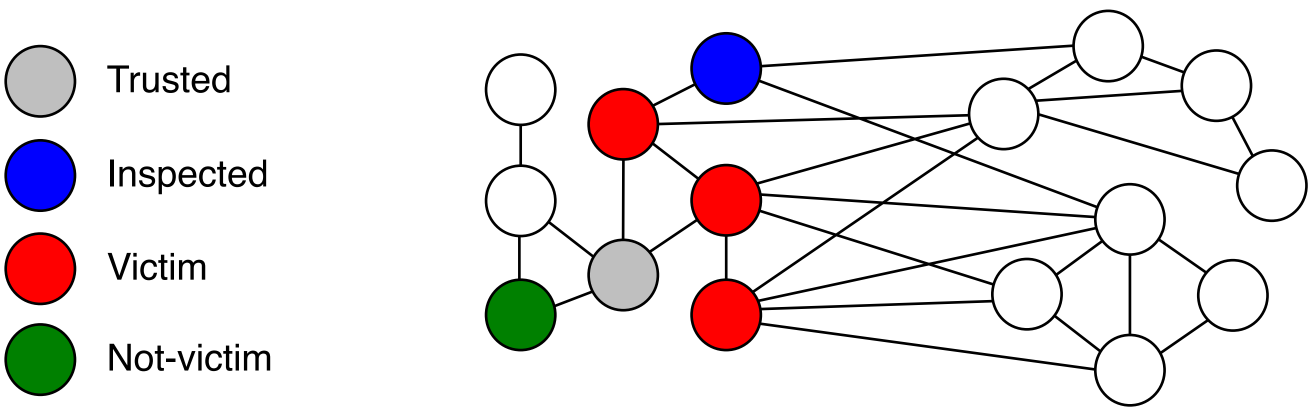
Victim prediction

#Friends	#Photos		#Posts
3	2066	⋯	766

Feature vector

Victim?
0

Target variable



TP	TN	FP	FN
3	**2**	**0**	**1**

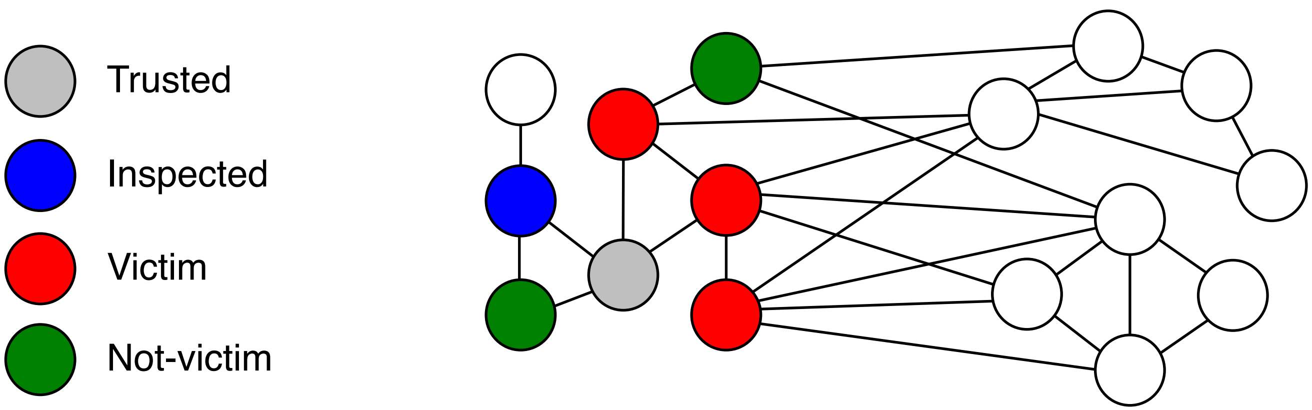
Victim prediction

#Friends	#Photos		#Posts
2	32	…	58

Feature vector

Victim?
0

Target variable



TP	TN	FP	FN
3	**3**	**0**	**1**

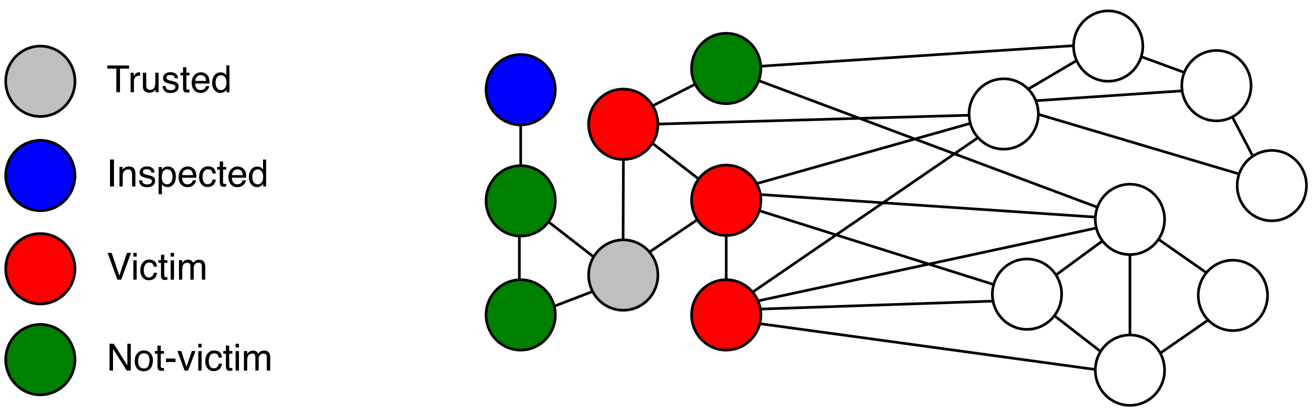
Victim prediction

#Friends	#Photos		#Posts
2	32	...	58

Feature vector

Victim?
0

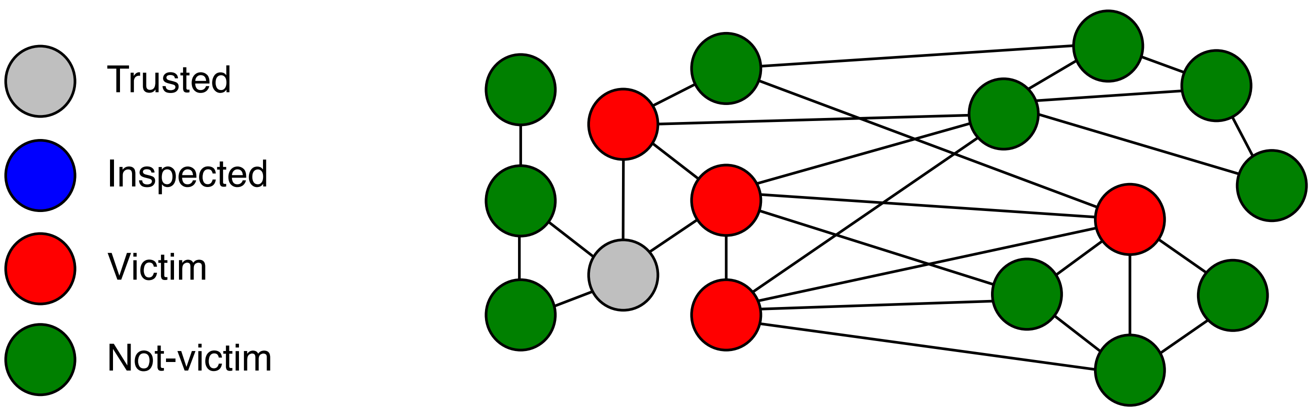
Target variable



TP	TN	FP	FN
3	**4**	**0**	**1**

Victim prediction

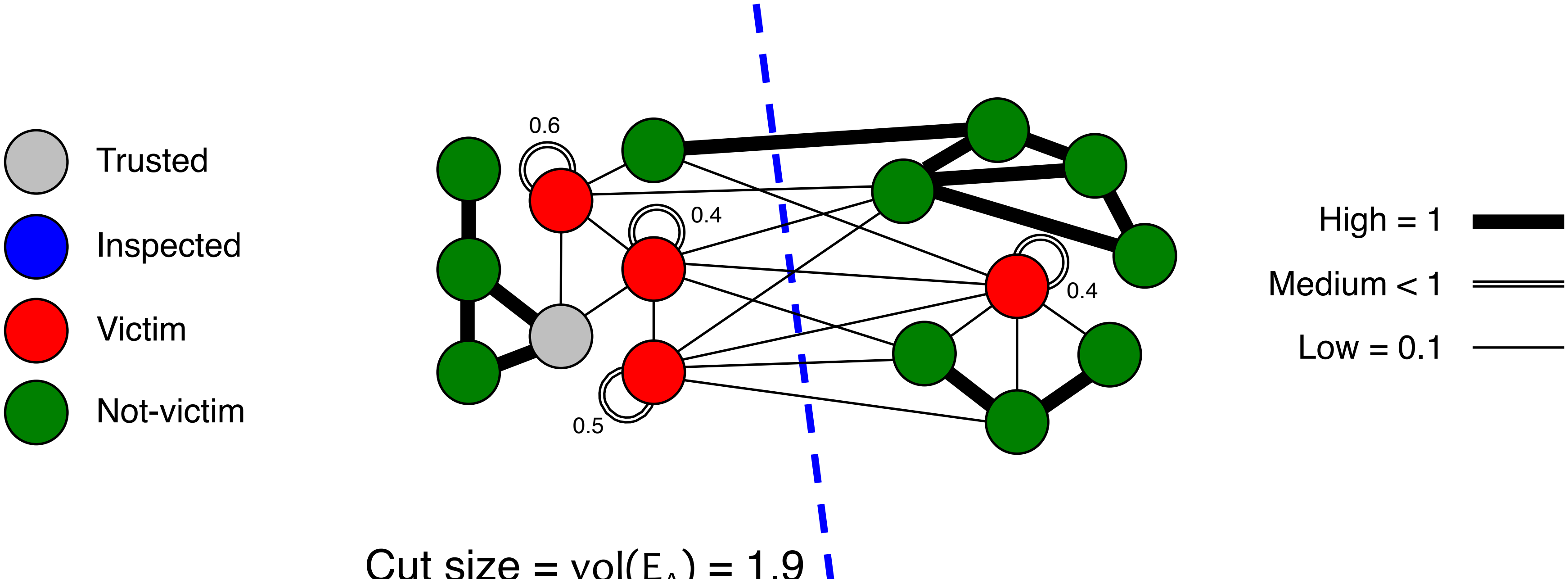
FPR = 8.3%, TPR = 75%



TP	TN	FP	FN
3	**11**	**1**	**1**

Malicious account detection

Assigns weights to edges based on victim predictions

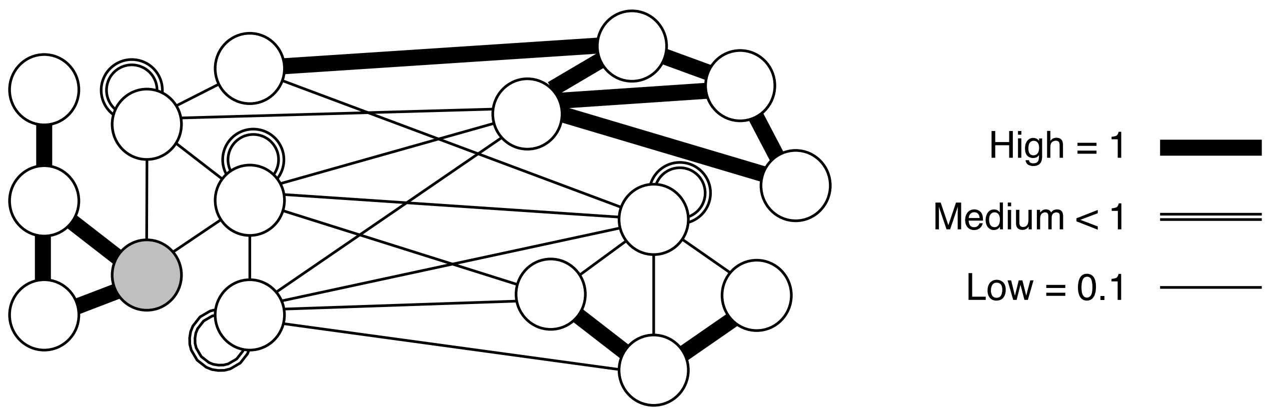


Cut size = vol(E_A) = 1.9

Penalizes relationships of identified victims (low edge weights)

Malicious account detection

Ranks accounts by degree-normalized landing probabilities of a short random walk



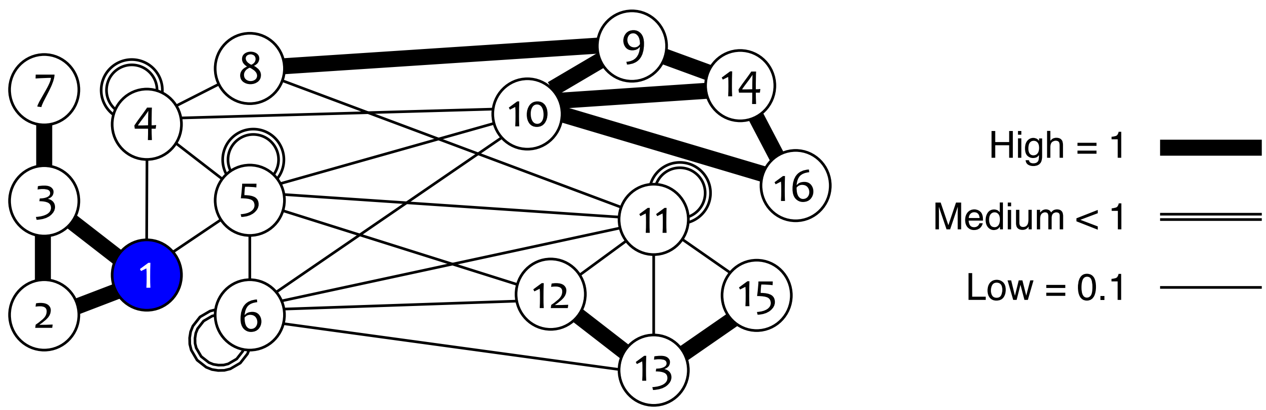
Real accounts ≈ similar ranks but malicious accounts ≈ significantly smaller ranks

Malicious account detection

	1	2	3	4	5	6	7	8	Total
Landing probability	1	0	0	0	0	0	0	0	1

Step

0



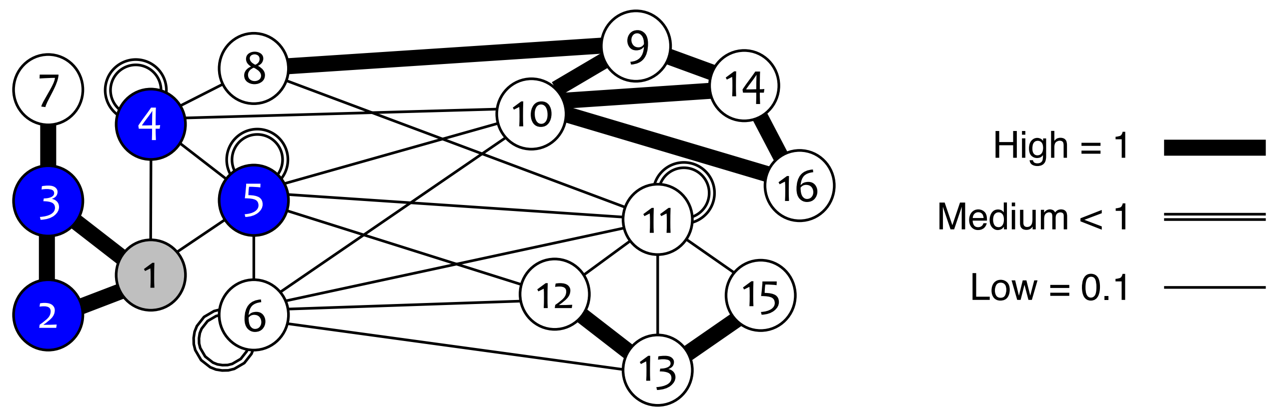
	9	10	11	12	13	14	15	16	Total
Landing probability	0	0	0	0	0	0	0	0	0

Malicious account detection

	1	2	3	4	5	6	7	8	Total
Landing probability	0	**0.4545**	**0.4545**	**0.0455**	**0.0455**	0	0	0	**1**

Step

1



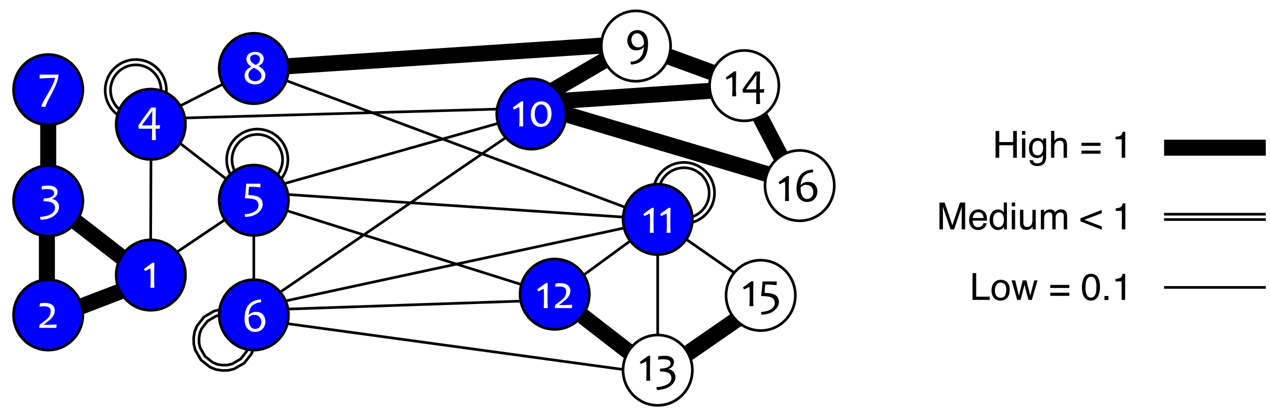
	9	10	11	12	13	14	15	16	Total
Landing probability	0	0	0	0	0	0	0	0	**0**

Malicious account detection

	1	2	3	4	5	6	7	8	Total
Landing probability	0.388	0.1515	0.2273	0.0319	0.0228	0.0046	0.1515	0.0046	~0.98

Step

2



	9	10	11	12	13	14	15	16	Total
Landing probability	0	0.0092	0.0046	0.0046	0	0	0	0	~0.02

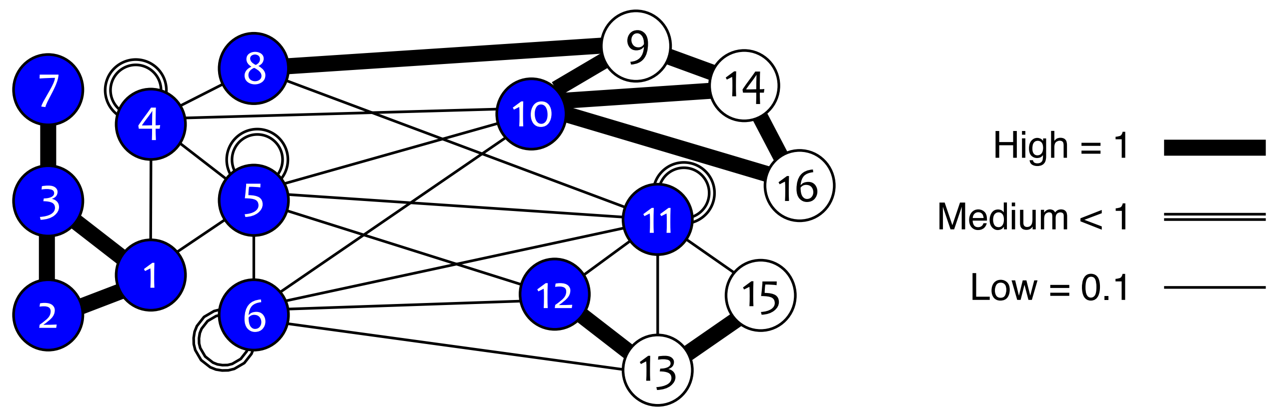
Early-terminates the random walk after O(logn) steps

Malicious account detection

	1	2	3	4	5	6	7	8
Normalized probability	0.1764	0.0758	0.0758	0.0319	0.0228	0.0046	0.1515	0.0038

Step

-

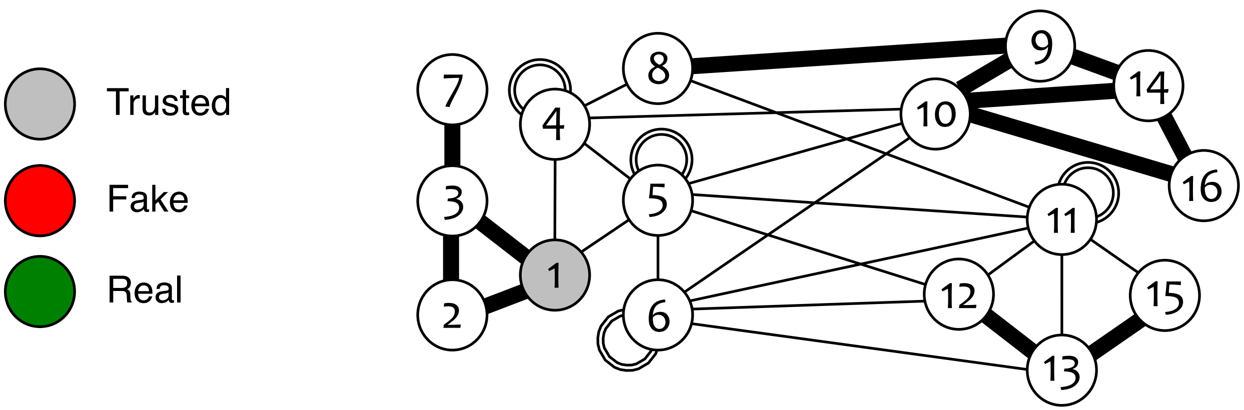


	9	10	11	12	13	14	15	16
Normalized probability	0	0.0028	0.0046	0.0035	0	0	0	0

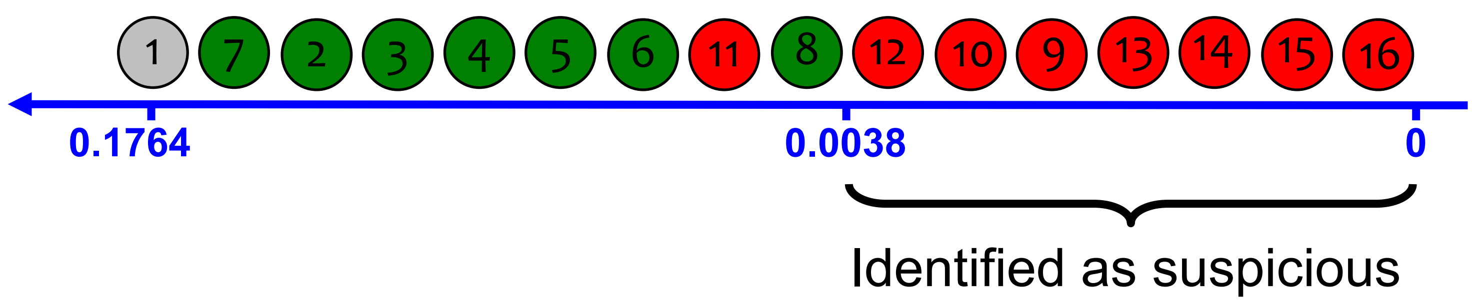
Ranks a node by its degree-normalized landing probability

Malicious account detection

Sorts accounts then estimates a threshold to identify suspicious ones

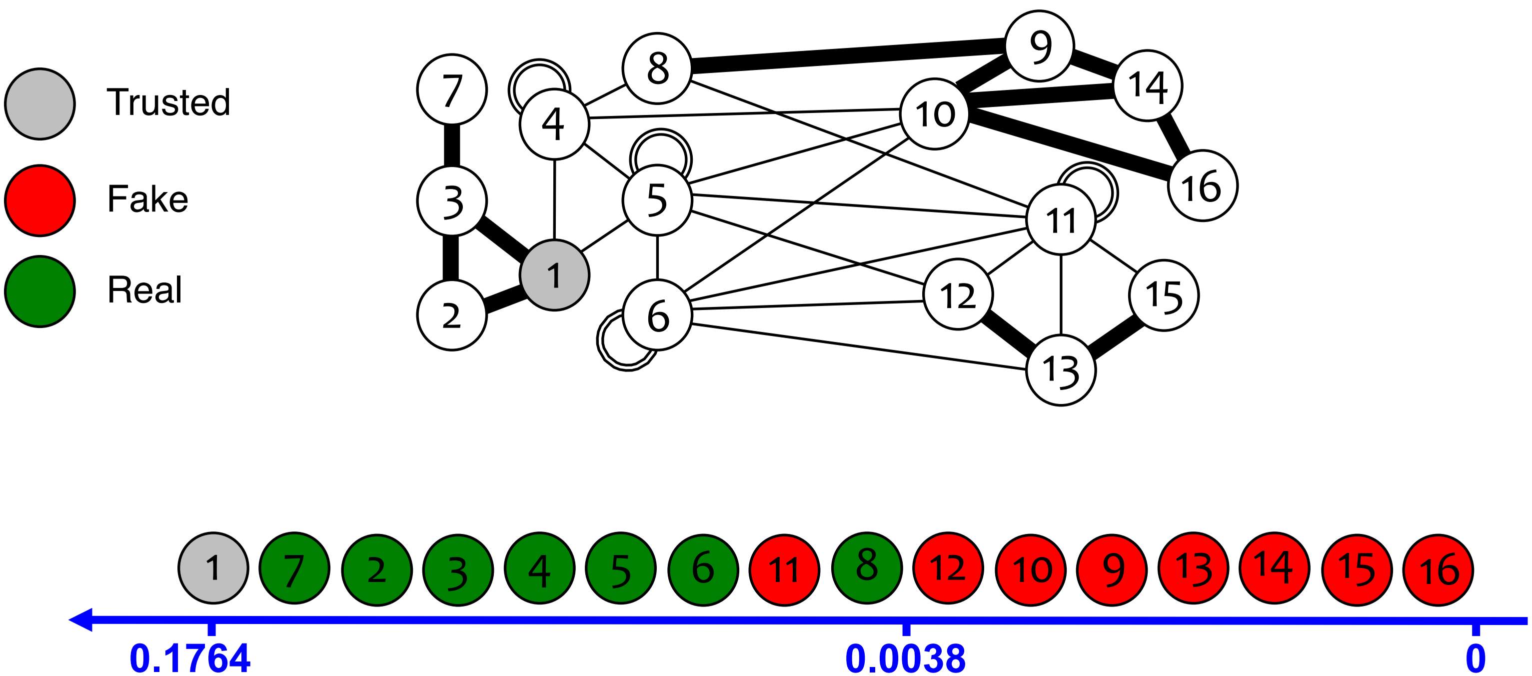


FPR = 0%
TPR = 88%



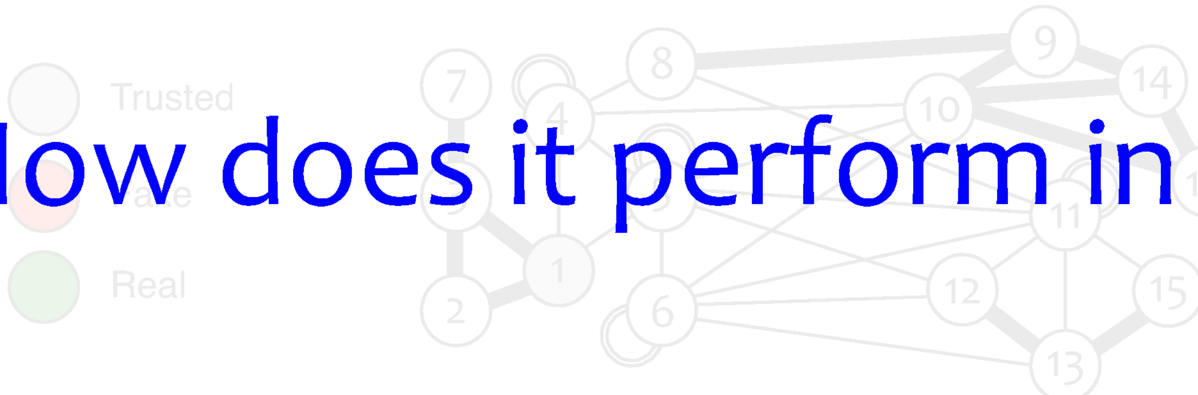
Malicious account detection

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



Malicious account detection

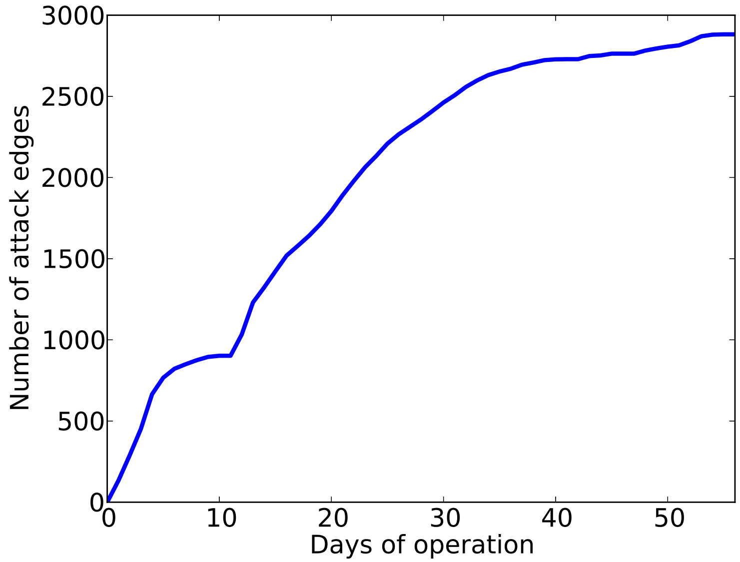
Theorem: Number of fake accounts that rank equal or higher than real accounts is O(vol(E_A) logn) where vol(E_A) ≤ $|A_E|$

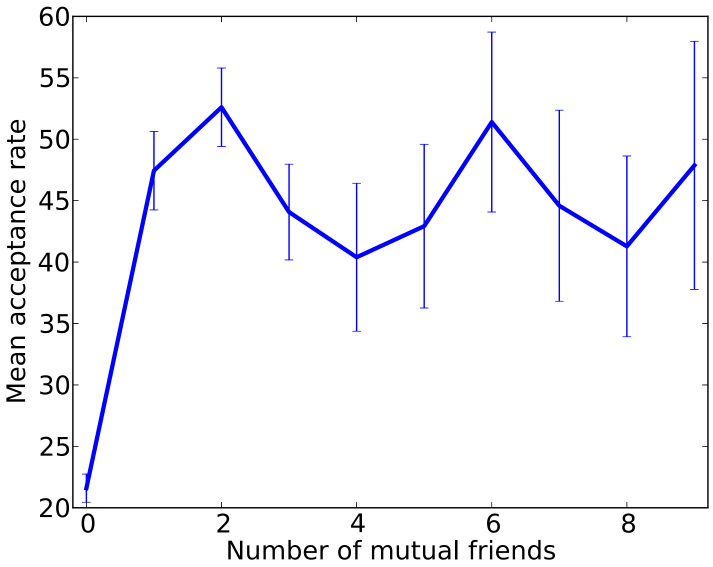


How does it perform in practice?

Real-world Sybil activity in Facebook

Data were collected in 2011 January 28 through March 23



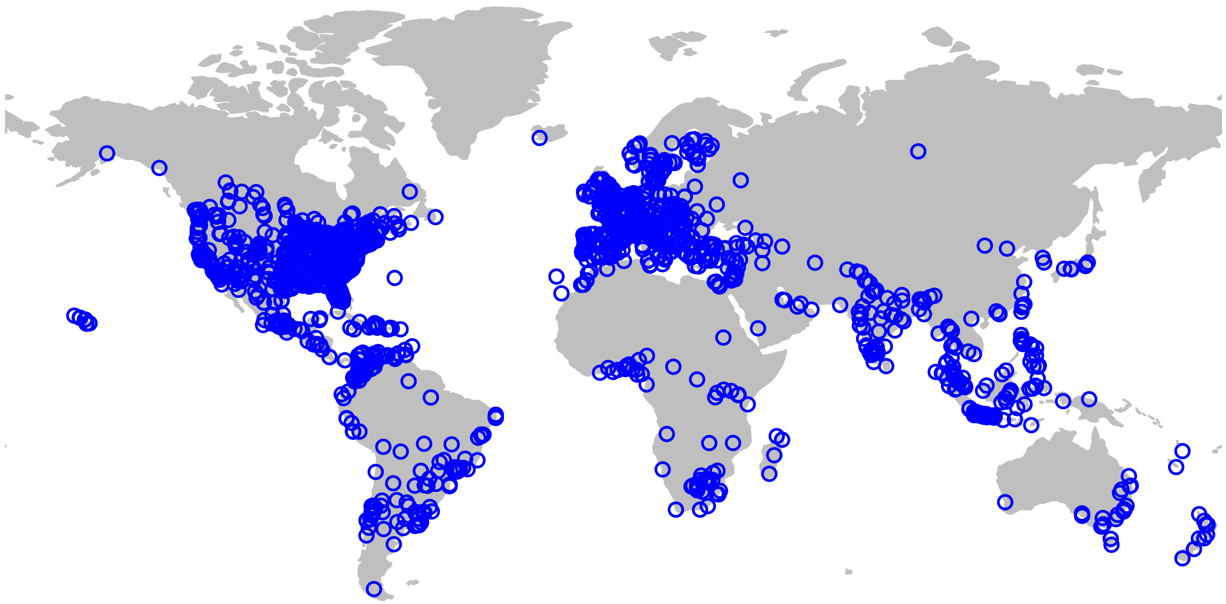


A total of 8.8K users received friend requests (32.4% victims)

More mutual friends more likely to accept a request sent by fakes

Real-world Sybil activity in Facebook

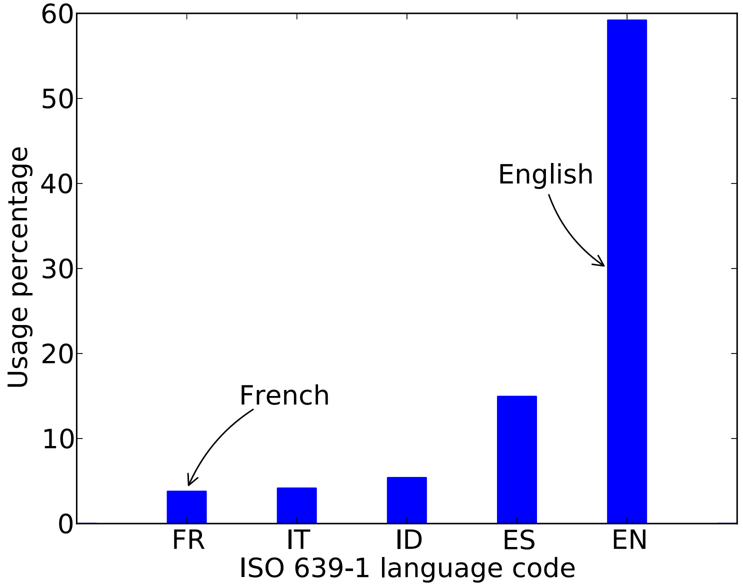
Data were collected in 2011 January 28 through March 23



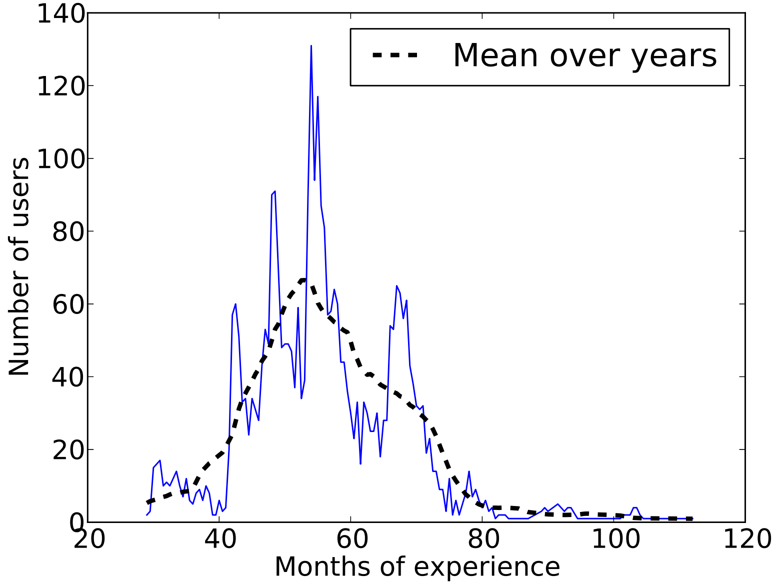
~2K different cities across 127 countries

Real-world Sybil activity in Facebook

Data were collected in 2011 January 28 through March 23



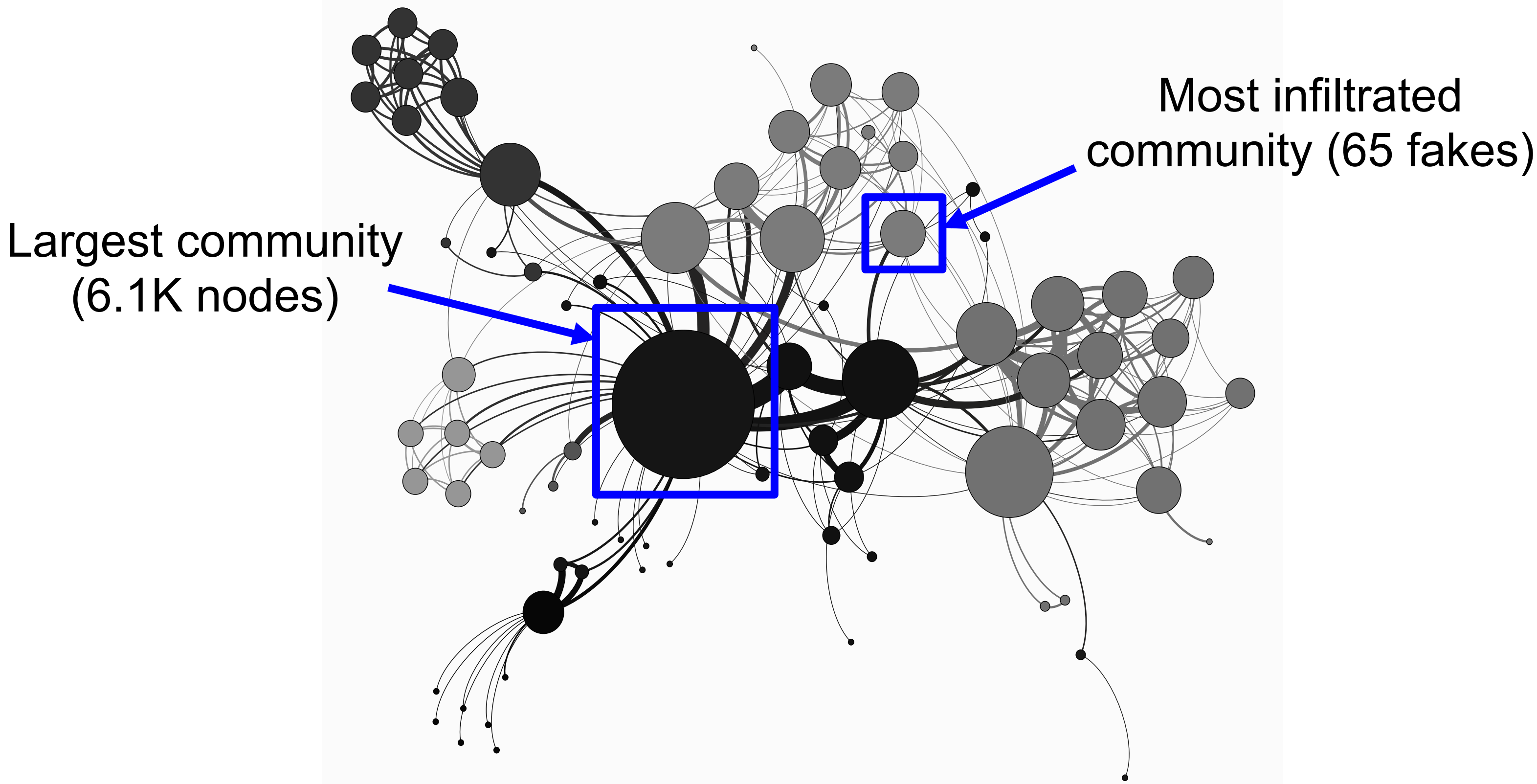
43 different languages



A mean of 5.4 years on Facebook

Real-world Sybil activity in Facebook

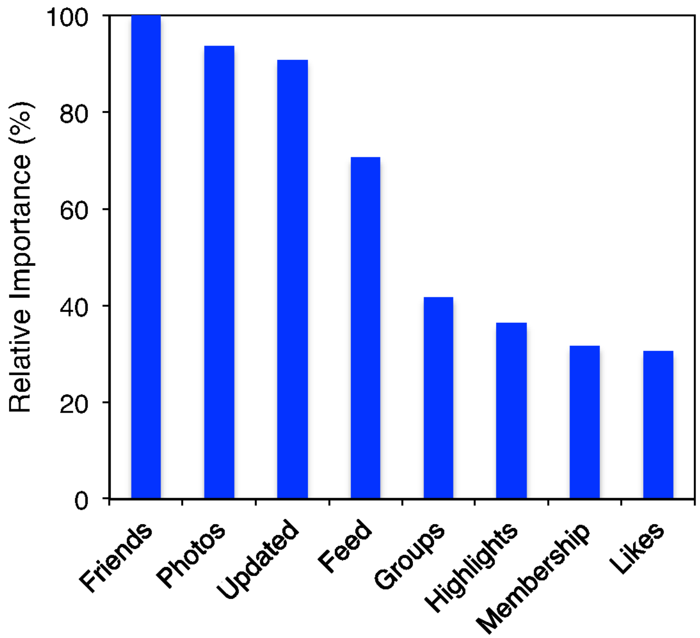
Data were collected in 2011 January 28 through March 23



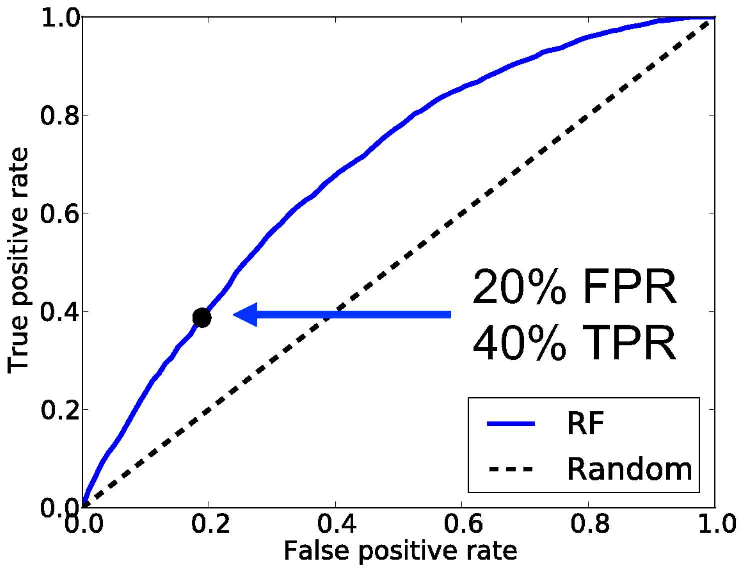
139K nodes, 660K edges, 74 communities, diameter of 9

Predicting victims

Random Forests (RF) is 40% better than random



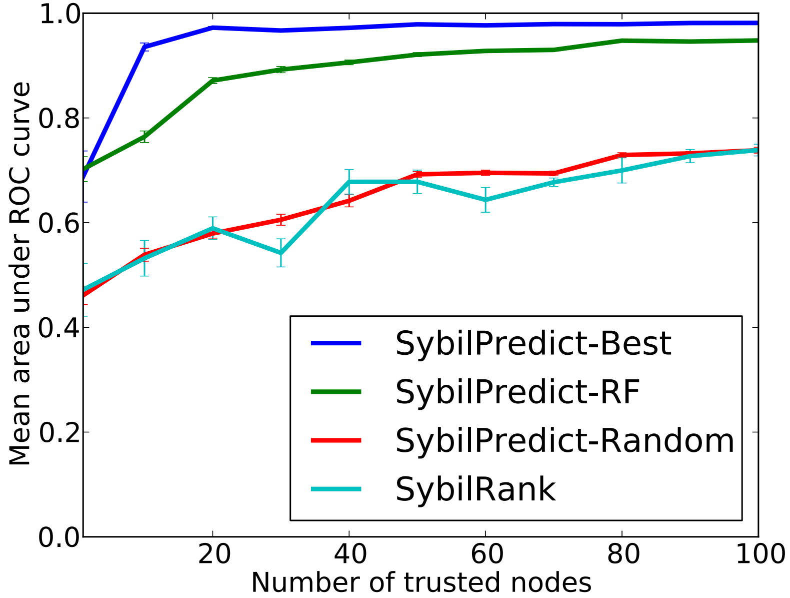
18 features from public profiles

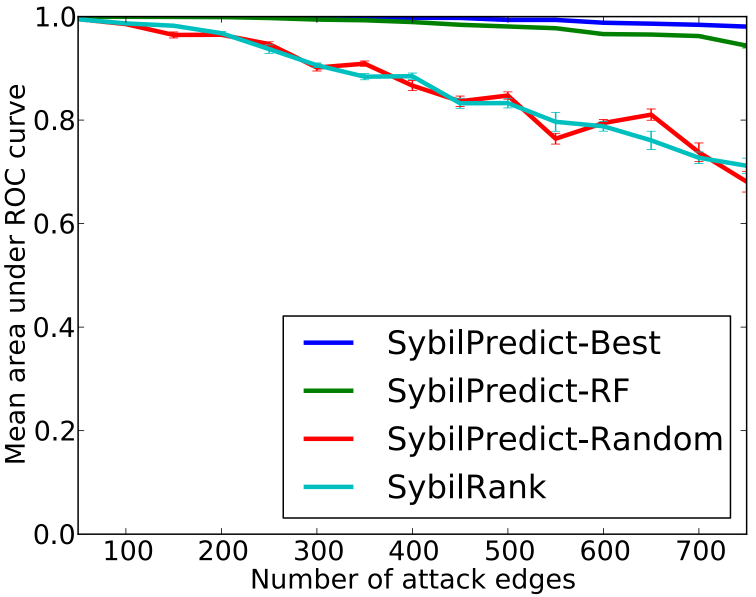


Random Forests classifier with AUC=0.7

Detecting malicious accounts

Trace-driven simulation on most infiltrated community (3K nodes)



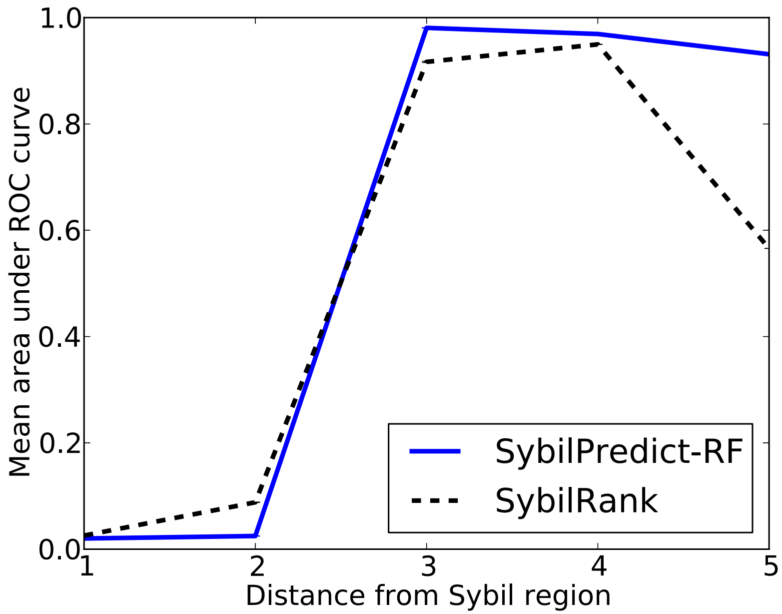


Few (random) seeds are enough

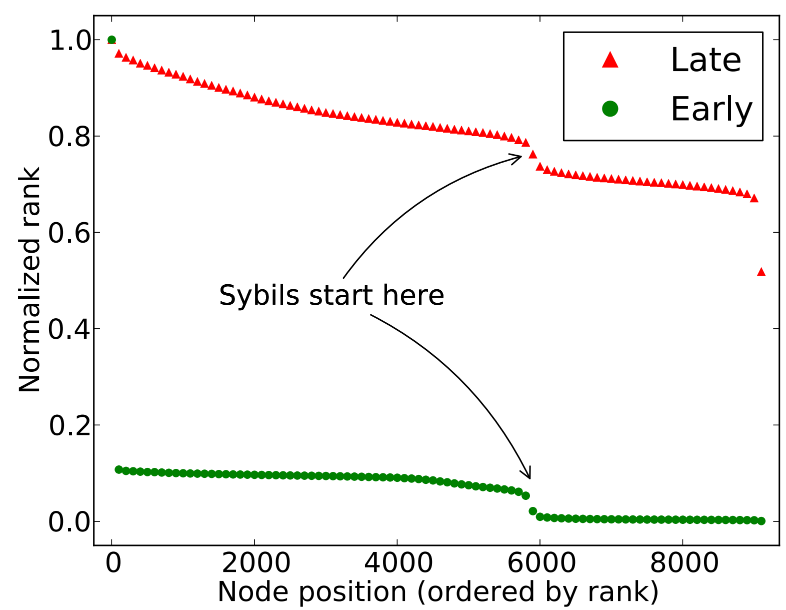
Near prefect ranking, up to 30% better

Detecting malicious accounts

Trace-driven simulation on most infiltrated community (3K nodes)



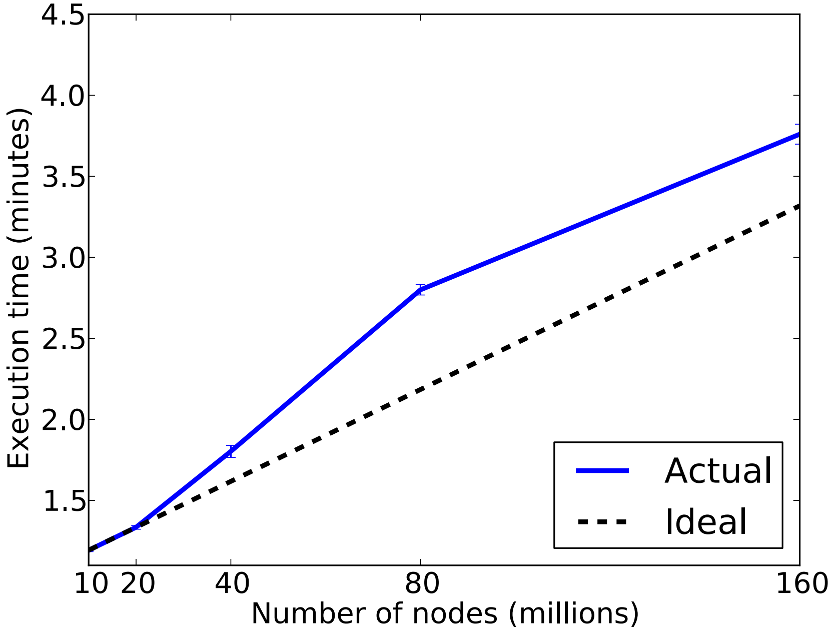
Seeds are sensitive to targeted attacks



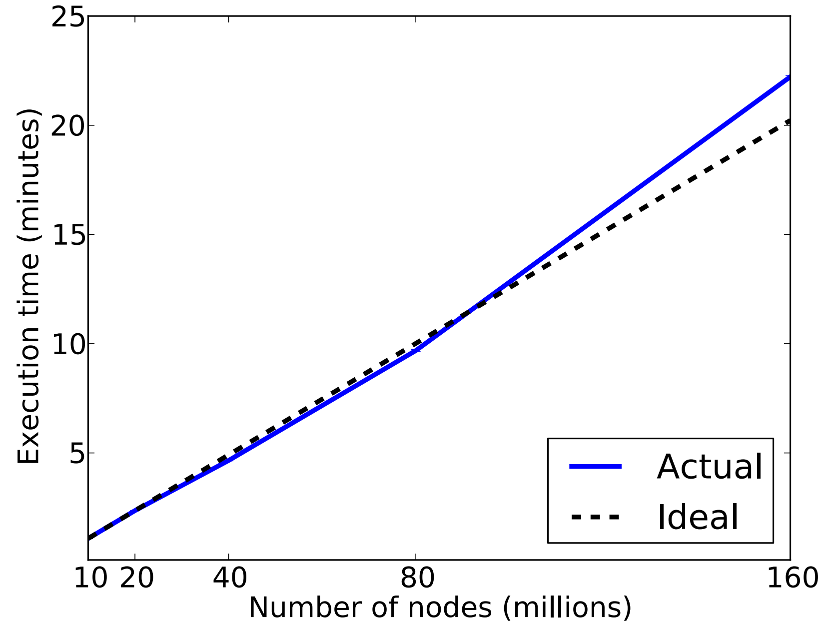
Clear cutoff threshold in rank distribution

Detecting malicious accounts

Near linear scalability with exponentially increasing order



RF is “embarrassingly parallel”



Ranking is “PageRank scalable”

In conclusion, SybilPredict is:

- ☑ Proactive protection
- ☑ Near real-time responses
- ☑ Scales to millions of users
- ☑ Hard to circumvent
- ☑ Accurate detection
- ☑ Provably secure

(patent disclosure submitted!)