Integro: Leveraging Victim Prediction for Robust Fake Account Detection in OSNs

Yazan Boshmaf, Matei Ripeanu, Konstantin Beznosov
University of British Columbia

Dionysios Logothetis, Georgios Siganos
Telefonica Research

Jorge Laria, Jose Lorenzo
University of British Columbia

Presented at NDSS'15, San Diego, Feb 2015

Why is it important to detect fakes?

Integro: Leveraging Victim Prediction for Robust Fake Account Detection in OSNs

Yazan Boshmaf, Matei Ripeanu, Konstantin Beznosov
University of British Columbia

Dionysios Logothetis, Georgios Siganos
Telefonica Research

Jorge Laria, Jose Lorenzo
University of British Columbia

Presented at NDSS'15, San Diego, Feb 2015

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

Social Infiltration

Connecting with many benign users (friend request spam)

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

Fake accounts are bad for users

OSNs are attractive medium for abusive content

Social Infiltration

Data collection

Online surveillance, profiling, and data commoditization

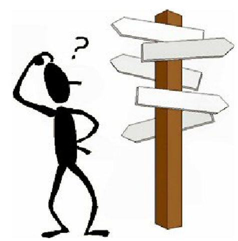
Nolan et al. Hacking human: Data-archaeology and surveillance in social networks. ACM SIGGROUP Bulletin 25.2, 2005

Fake accounts are bad for users

OSNs are attractive medium for abusive content

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

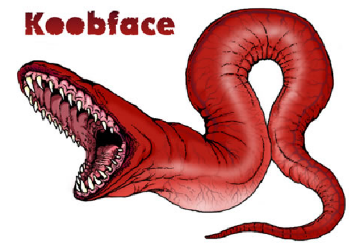
Fake accounts are bad for users

OSNs are attractive medium for abusive content

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

Fake accounts are bad for users

OSNs are attractive medium for abusive content

How do OSNs detect fakes today?

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

Feature-based detection

Interactions

Pictures

Triadic closure

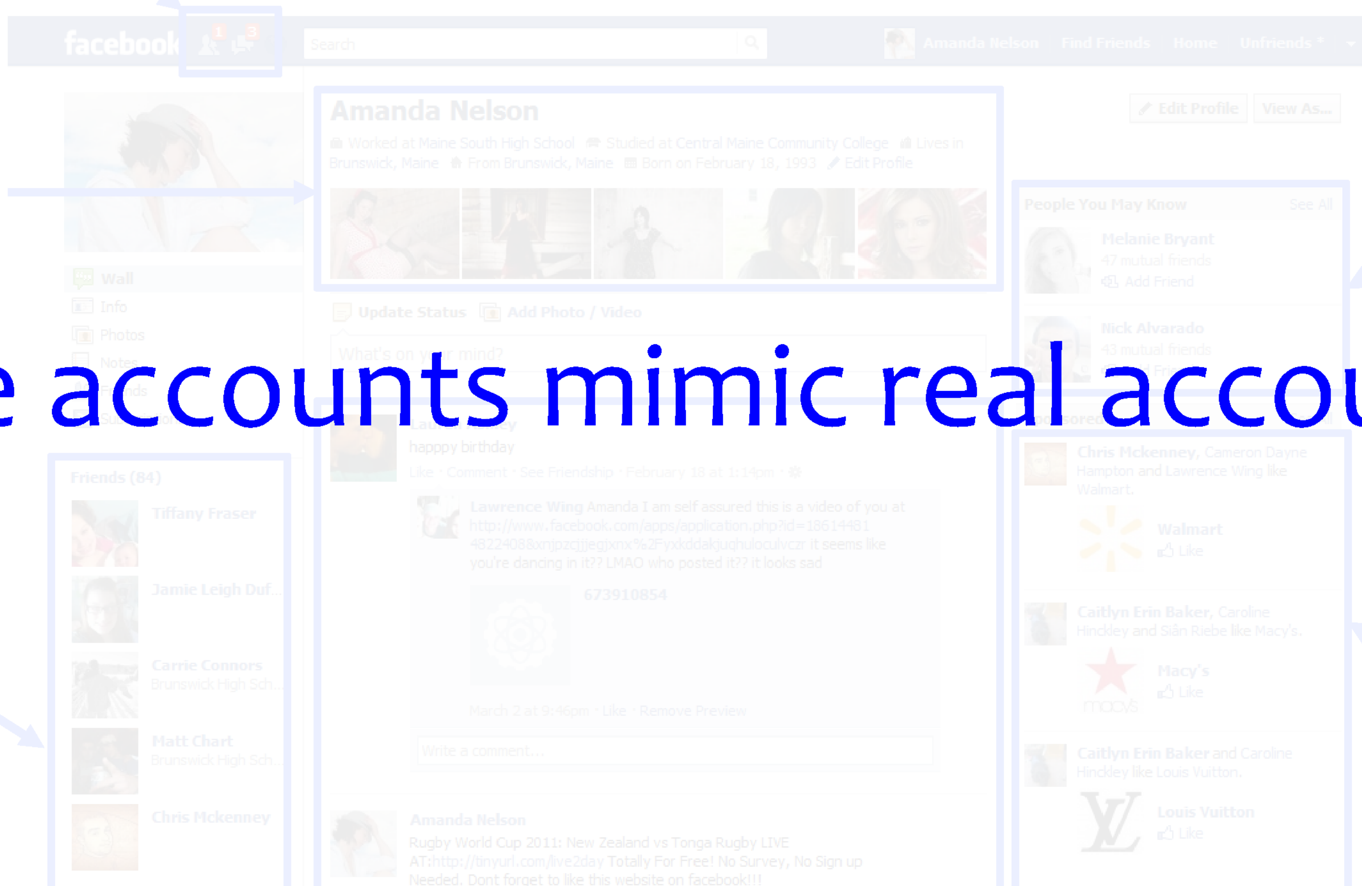
Friends

Ad clicks

Posts

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

Feature-based detection



Fake accounts mimic real accounts

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

Feature-based detection is ineffective

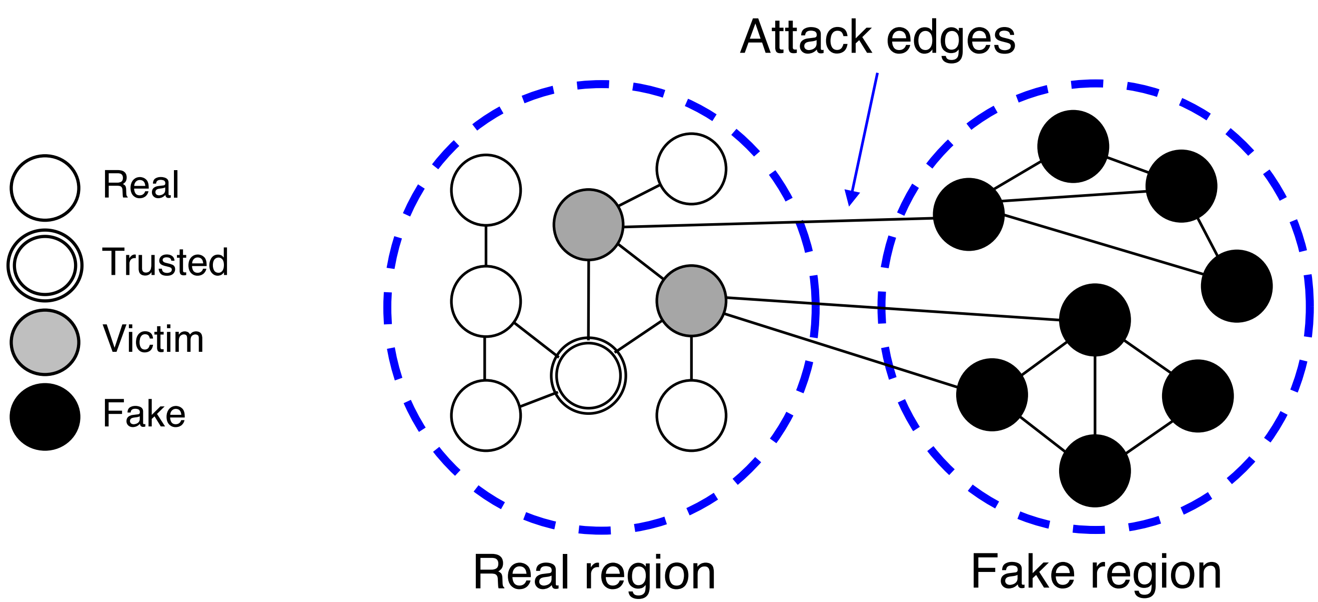
Only 20% of fakes were detected

All manually flagged by concerned users

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

Graph-based detection

Assumes social infiltration on a large scale is infeasible

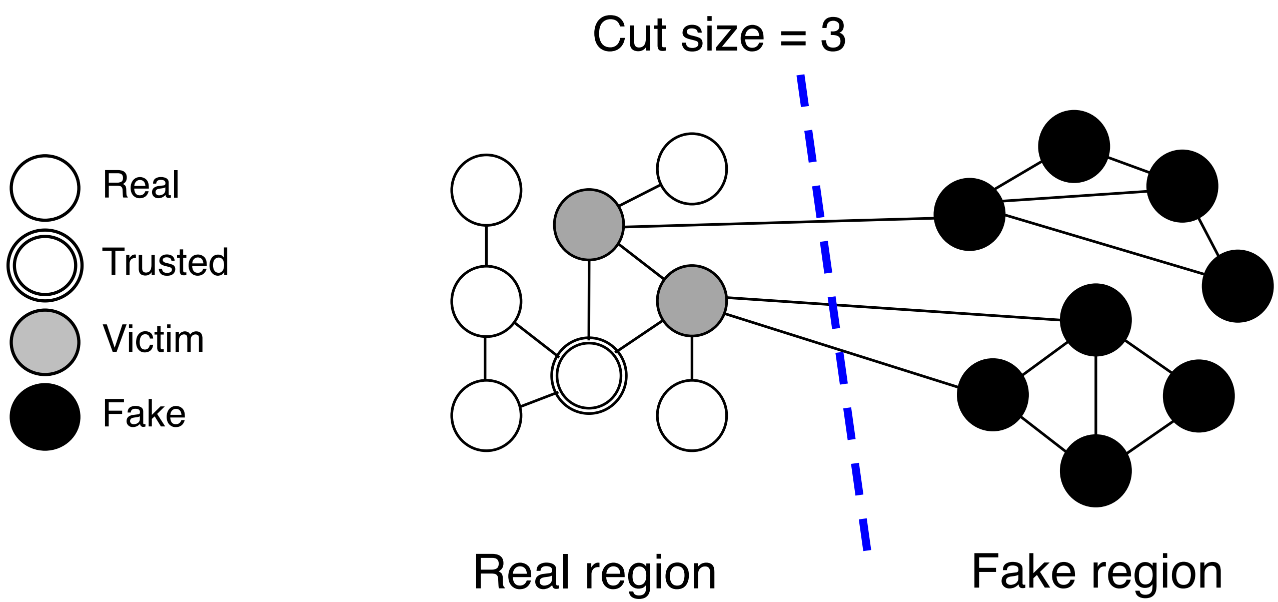


Finds a (provably) sparse cut between the regions by ranking

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

Graph-based detection

Assumes social infiltration on a large scale is infeasible

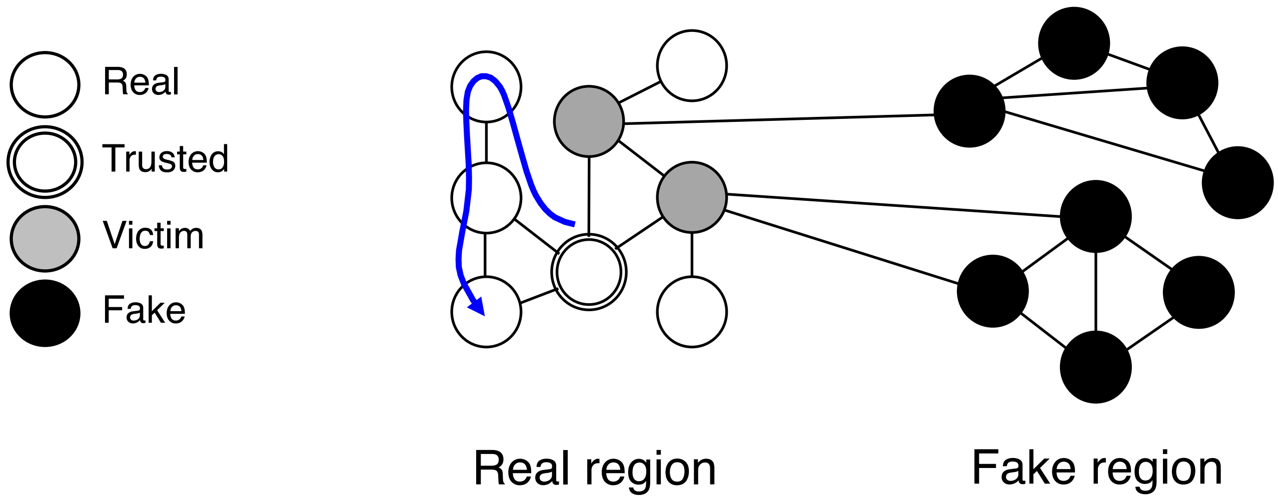


Finds a (provably) sparse cut between the regions by ranking

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

Graph-based detection

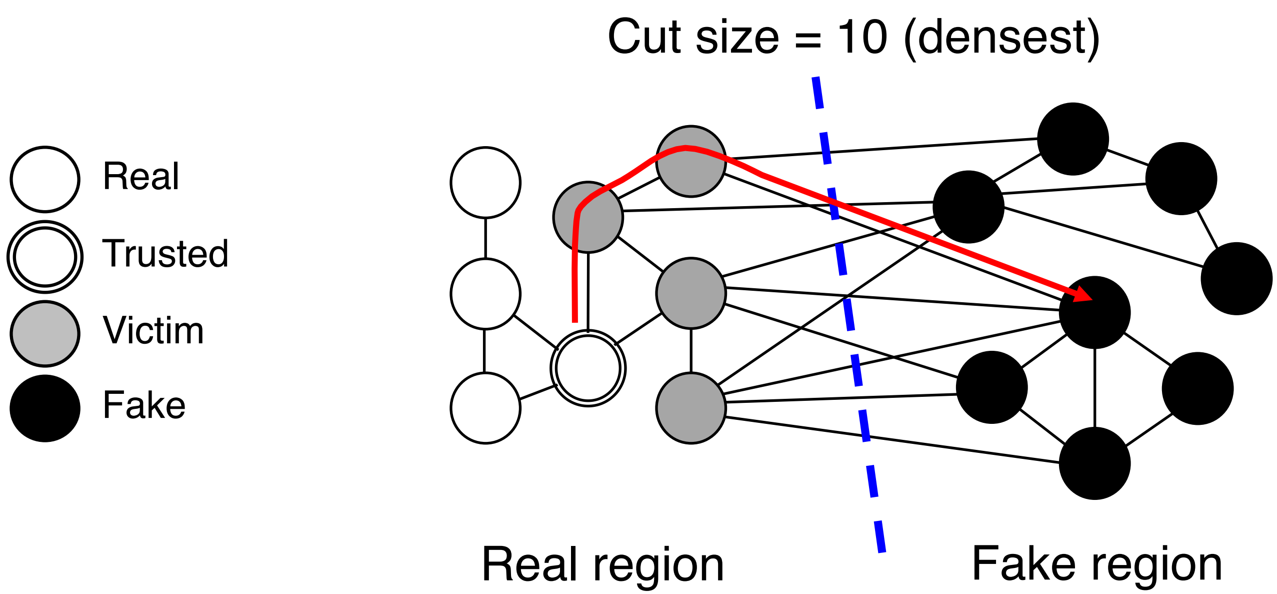
Ranks computed from landing probability of a short random walk



Most real accounts rank higher than fakes

Cao et al. Aiding the detection of fake accounts in large scale social online services, In proc. of NSDI, 2012

Graph-based detection
is not resilient to social infiltration



50% of fakes had more than 35 attack edges

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

Graph-based detection <u>is not</u> resilient to social infiltration

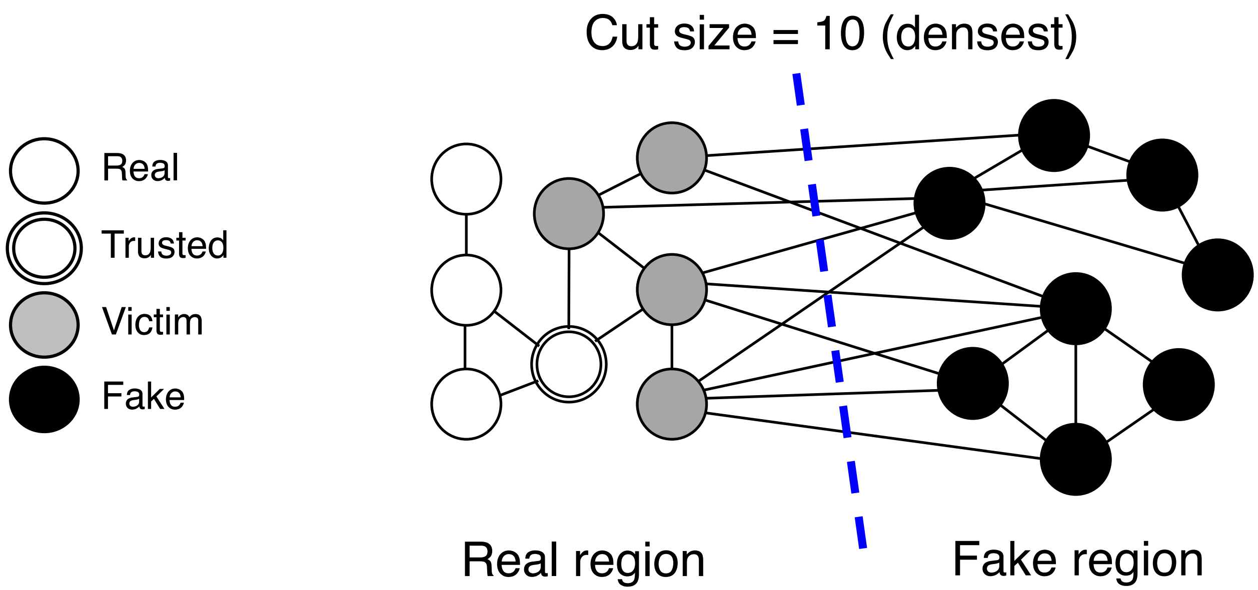
Can we do better?

Hint: What if we <u>integrate</u> both?

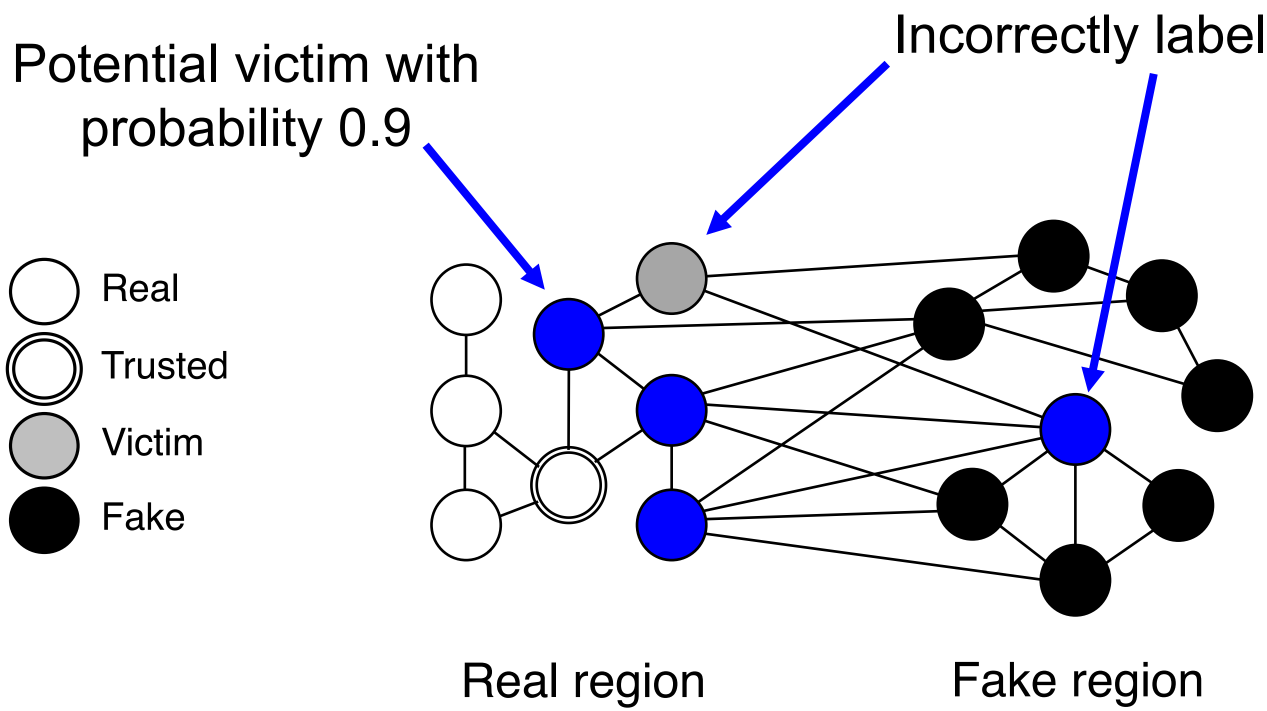
50% of bots had <u>more than</u> 35 attack edges

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

Premise: Regions can be tightly connected

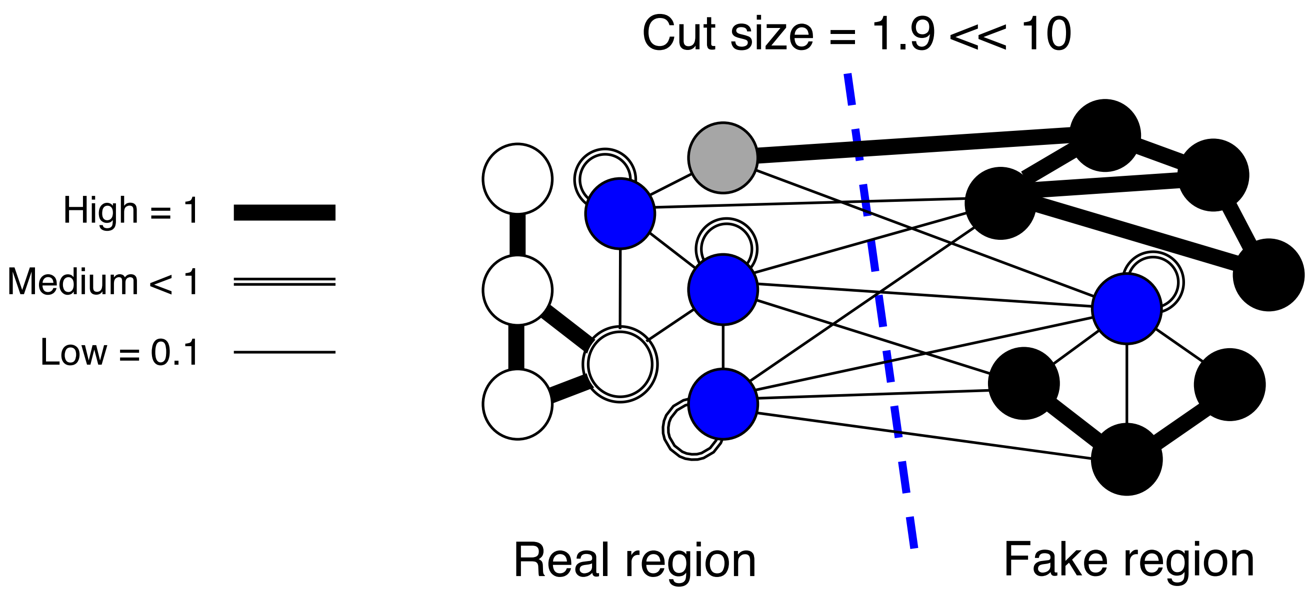


Identify potential victims with some probability



Potential victims are real accounts that are likely to be victims

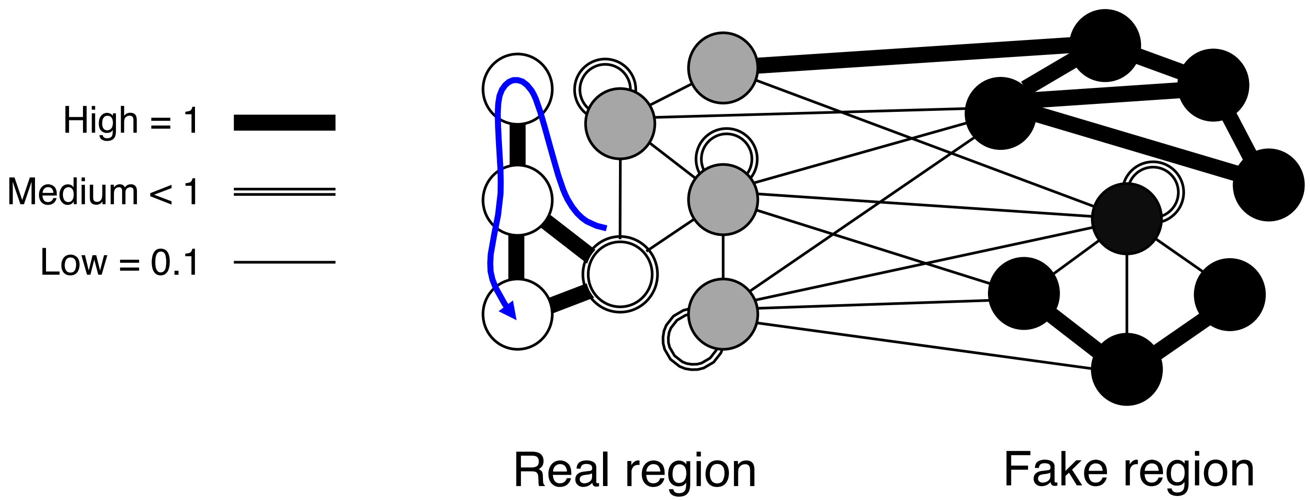
Leverage victim prediction to reduce cut size



Assign lower weight to edges incident to potential victims

Delimit the real region by ranking accounts

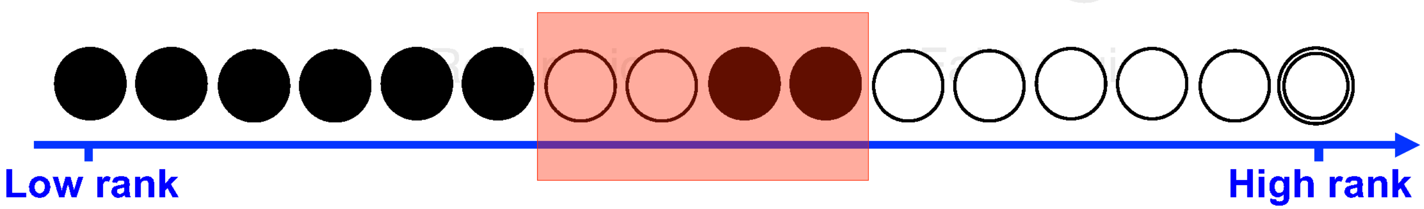
Ranks computed from landing probability of a short random walk



Most real accounts are ranked higher than fake accounts

(Bound on ranking quality)

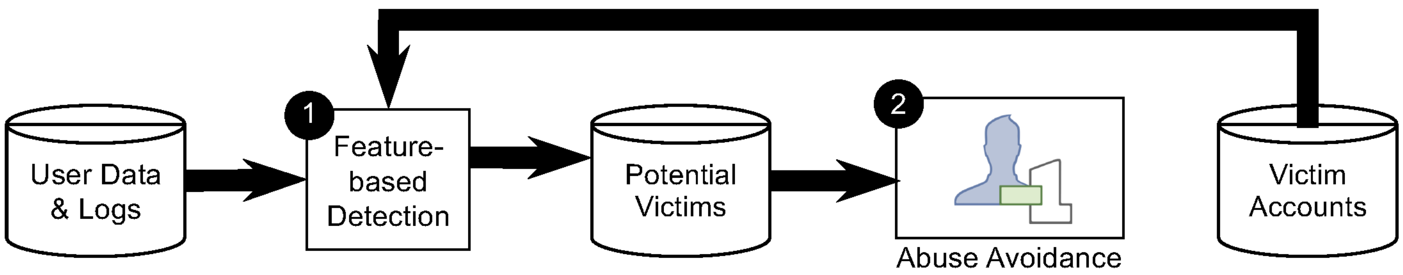
Number of fake accounts that rank equal to or higher than real accounts is $O(vol(E_A) \log n)$ where $vol(E_A) \leq |E_A|$



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

Integro: Victim classification

Identifies potential victims in O(n logn) time



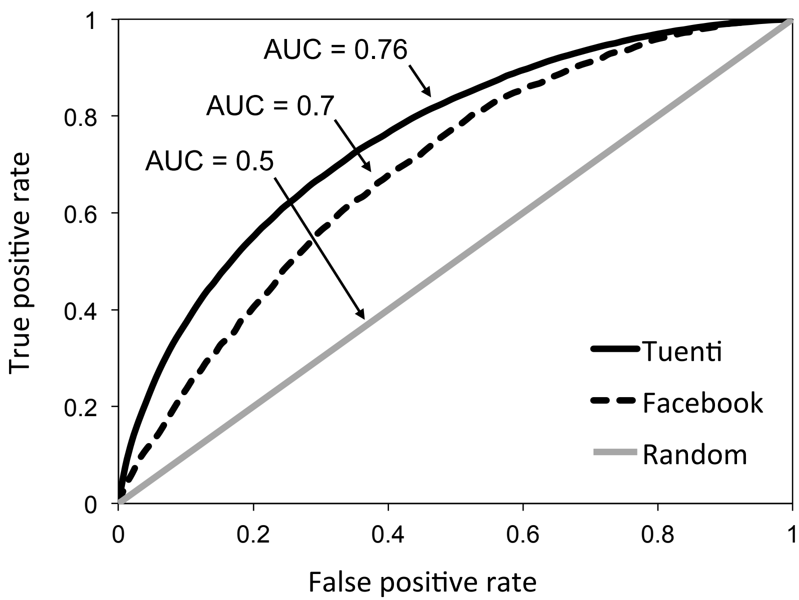
Pros:

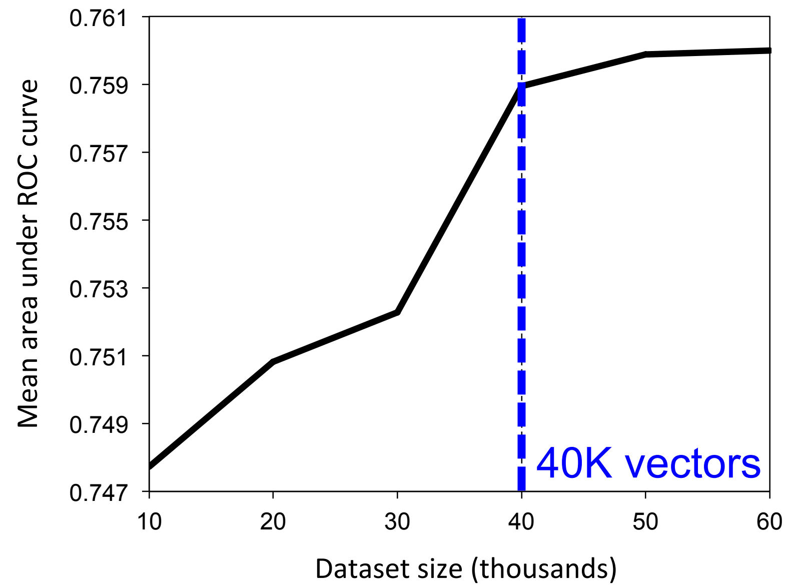
- Proactive protection
- Near real-time responses
- Scales to millions of users
- Hard to circumvent

Cons:

- Doesn't identify fakes
- May introduce usability issues
- Not provably secure

Victim classification is feasible 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: User account ranking

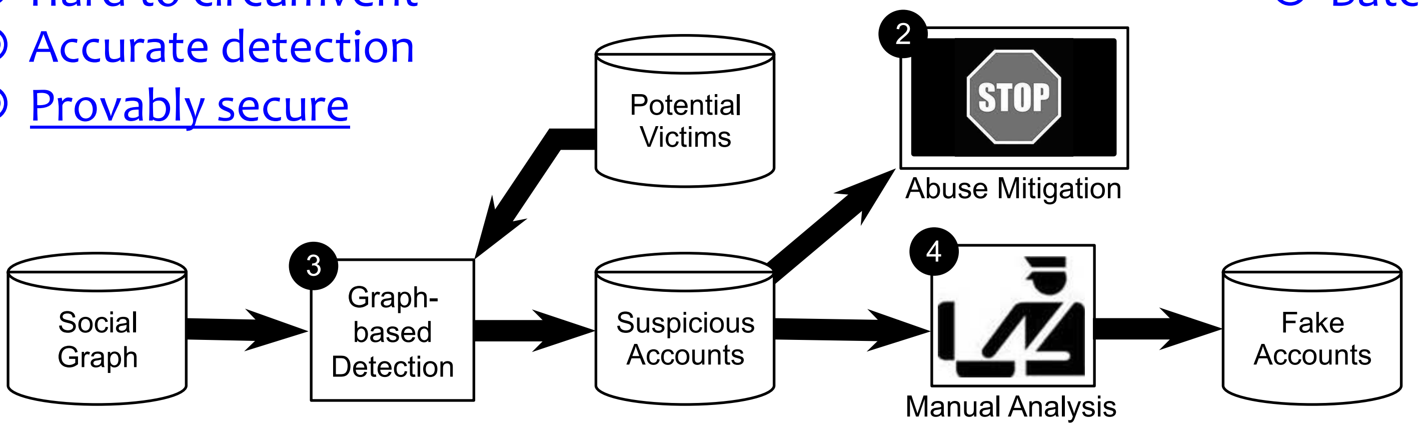
Integrates victim classification (labels + probabilities) into graph as edge weights

Pros:

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

Cons:

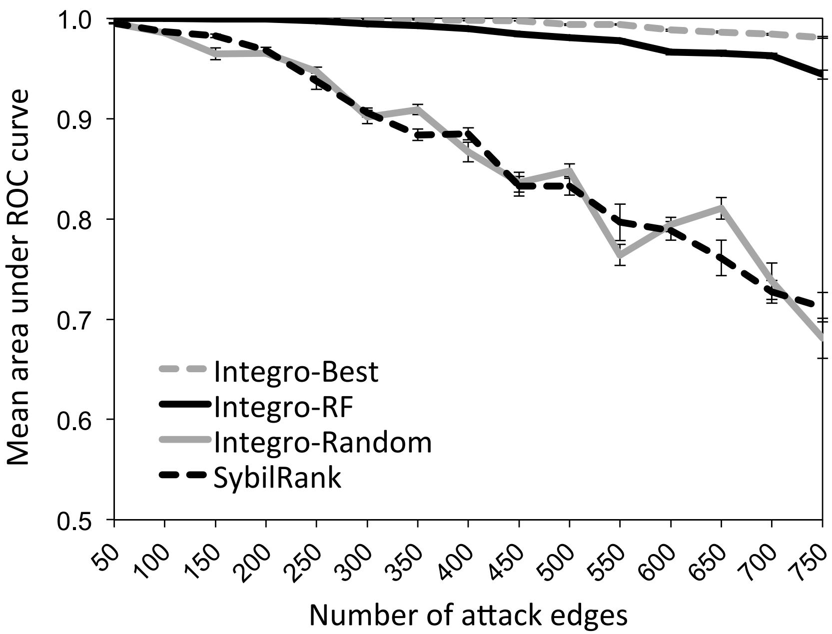
- Reactive protection
- Batch processed



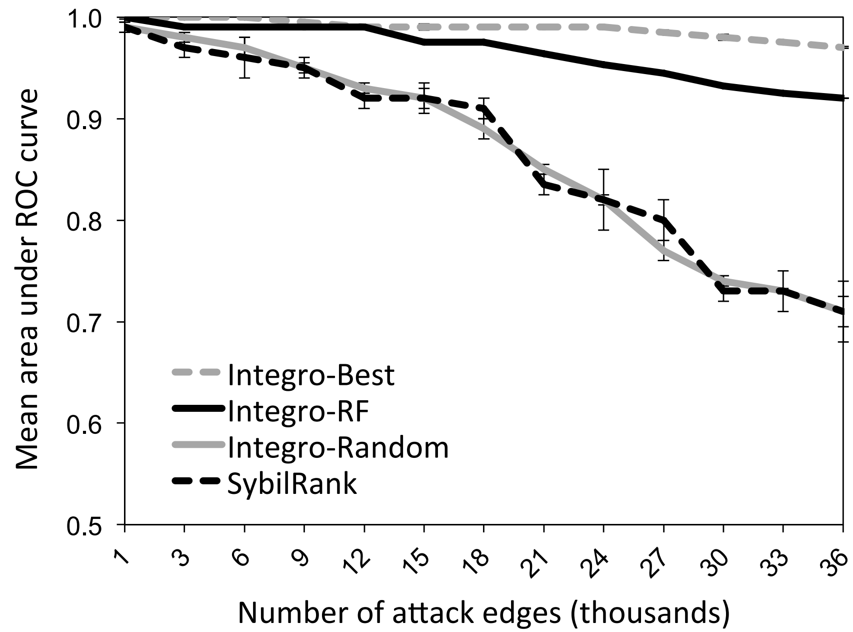
Ranks accounts based on a *short* random walk in O(n logn + m) time

Ranking is resilient to infiltration

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



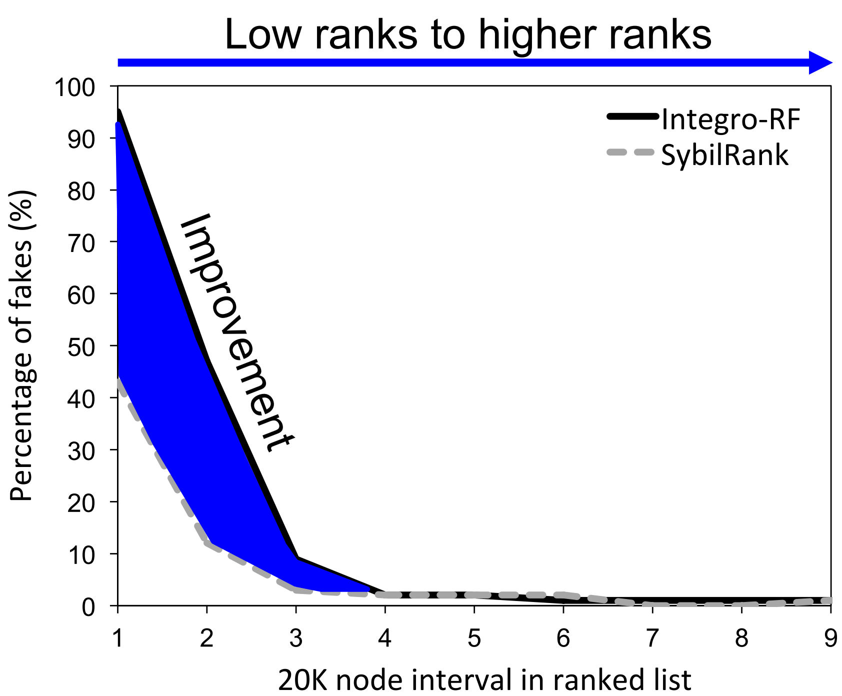
Targeted-victim attack



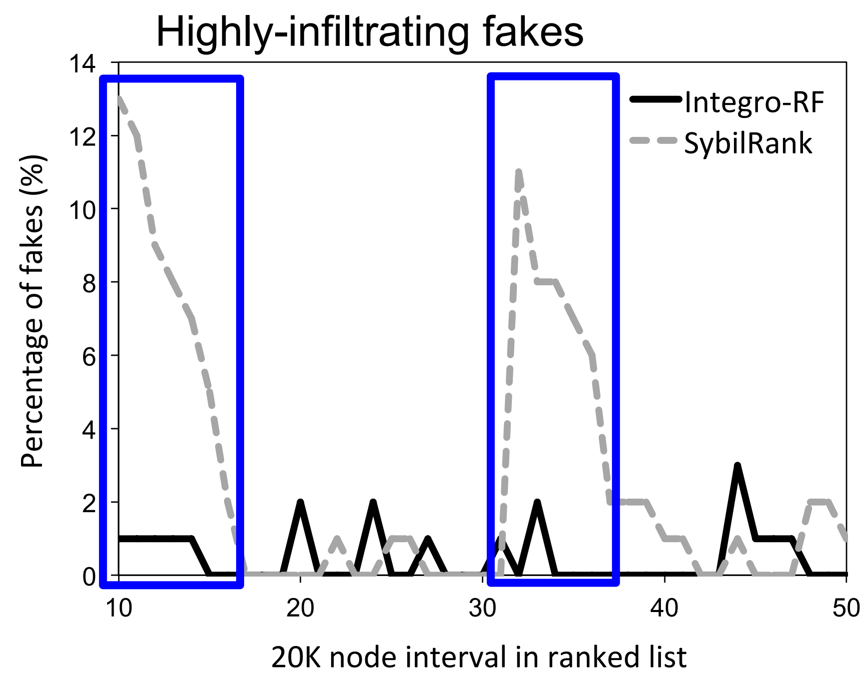
Random-victim attack

Deployment at Tuenti confirms results

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



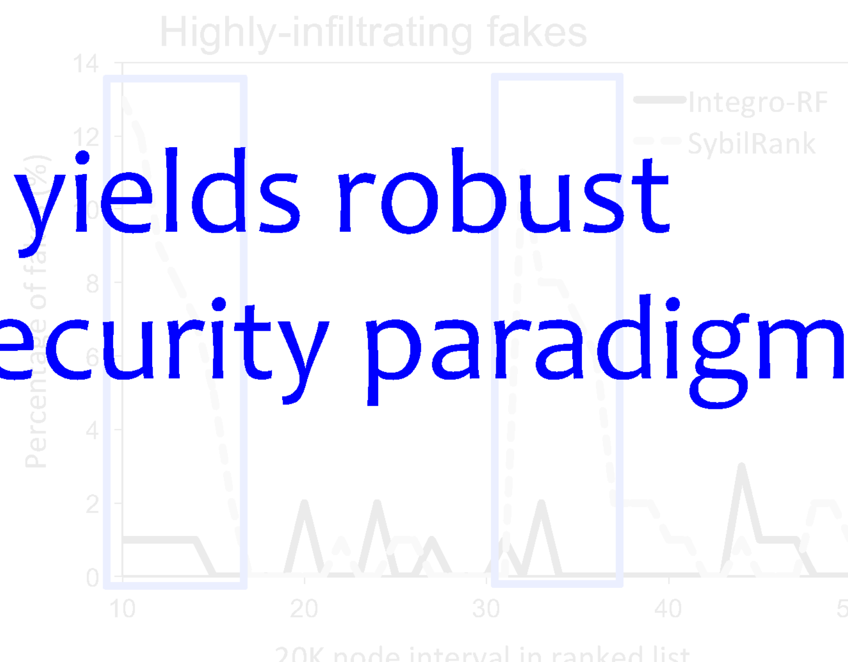
Precision at lower intervals



Precision at higher intervals

Deployment at Tuenti confirms results

Integro delivers up to an order or magnitude better precision



Victim prediction yields robust detection (new security paradigm)

Precision at lower intervals

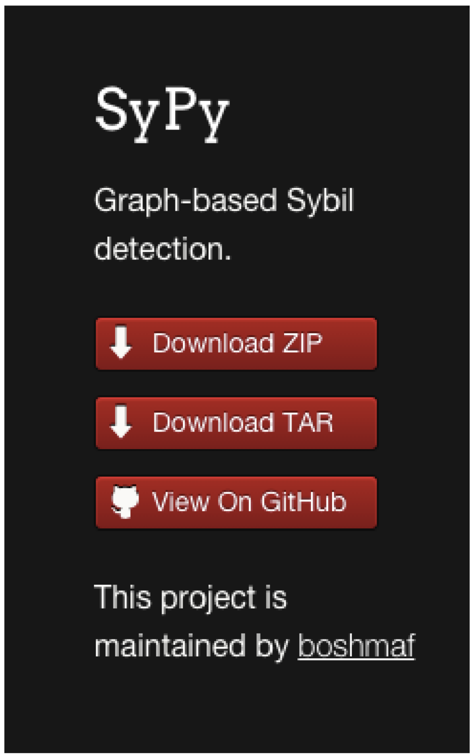
Precision at higher intervals

In conclusion, Integro achieves:

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

Fork or clone Integro now!

SyPy and Integro are publicly released



http://boshmaf.github.io/sypy

https://grafos.ml

Fork or clone Integro now!

SyPy and Integro are publicly released

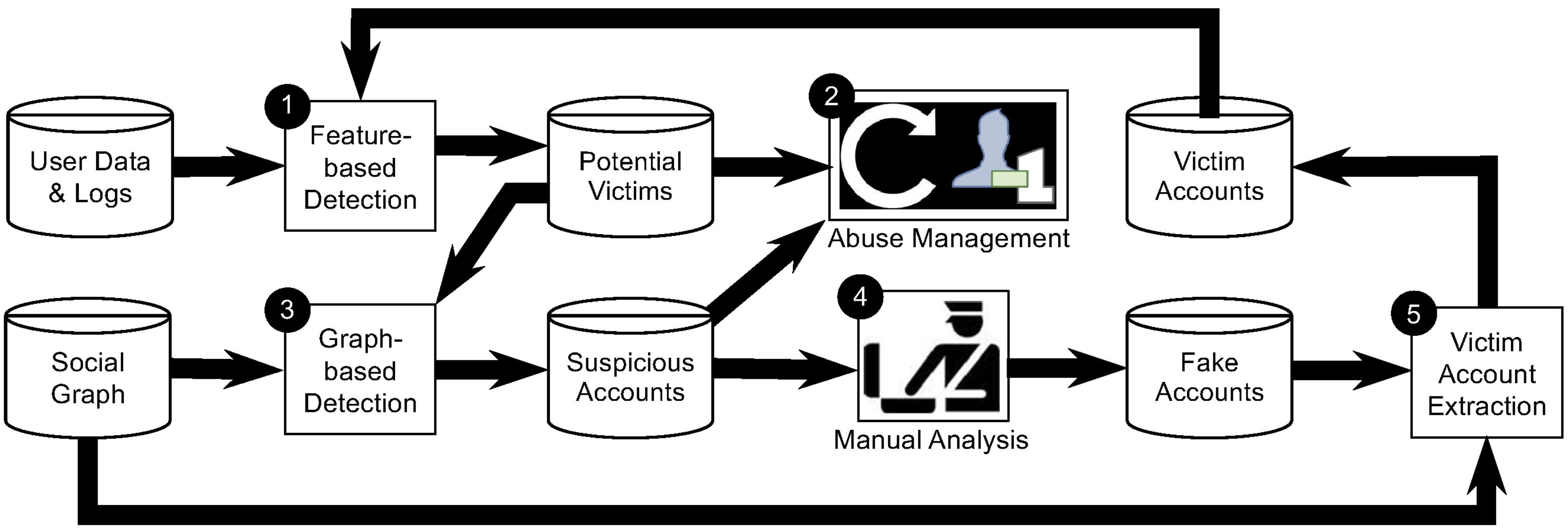
Backup

http://boshmaf.github.io/sypy

https://grafos.ml

Integro in a nutshell

Uses distributed machine learning and graph processing infrastructure



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

Datasets

- Labeled feature vectors
  - 8.8K public Facebook profiles (32% victims)
  - 60K full Tuenti profiles (50% victims)

- Graph samples
  - Time stamped infiltration targeting 2.9K real accounts, with 65 fakes and 748 attack edges
  - 6.1K real accounts

Feature engineering

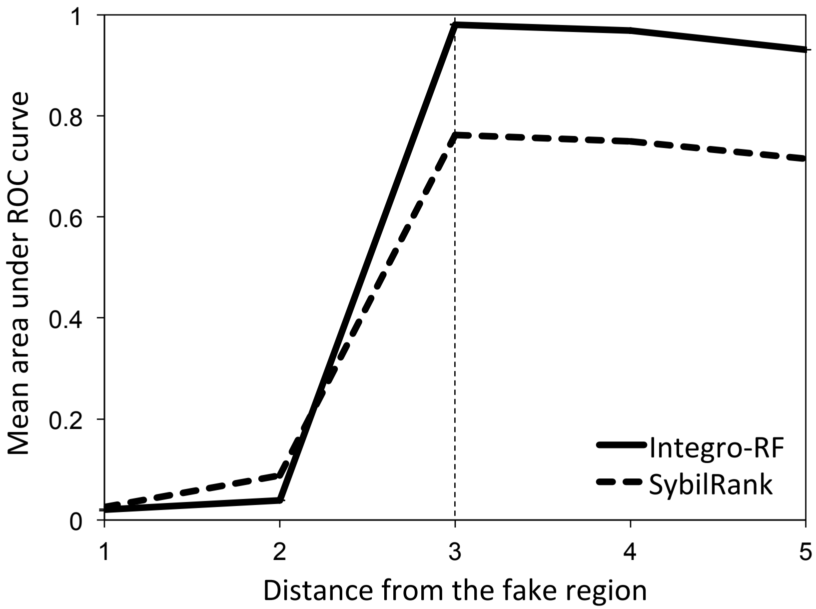
Most important features

Feature	Brief description	Type	RI Score (%)	
			Facebook	Tuenti
User activity:				
Friends	Number of friends the user had	Numeric	100.0	84.5
Photos	Number of photos the user shared	Numeric	93.7	57.4
Feed	Number of news feed items the user had	Numeric	70.6	60.8
Groups	Number of groups the user was member of	Numeric	41.8	N/A
Likes	Number of likes the users made	Numeric	30.6	N/A
Games	Number of games the user played	Numeric	20.1	N/A
Movies	Number of movies the user watched	Numeric	16.2	N/A
Music	Number of albums or songs the user listened to	Numeric	15.5	N/A
TV	Number of TV shows the user watched	Numeric	14.2	N/A
Books	Number of books the user read	Numeric	7.5	N/A
Personal messaging:				
Sent	Number of messages sent by the user	Numeric	N/A	53.3
Inbox	Number of messages in the user's inbox	Numeric	N/A	52.9
Privacy	Privacy level for receiving messages	5-Categorical	N/A	9.6
Blocking actions:				
Users	Number of users blocked by the user	Numeric	N/A	23.9
Graphics	Number of graphics (photos) blocked by the user	Numeric	N/A	19.7
Account information:				
Last updated	Number of days since the user updated the profile	Numeric	90.77	32.5
Highlights	Number of years highlighted in the user's time-line	Numeric	36.3	N/A
Membership	Number of days since the user joined the OSN	Numeric	31.7	100
Gender	User is male or female	2-Categorical	13.8	7.9
Cover picture	User has a cover picture	2-Categorical	10.5	< 0.1
Profile picture	User has a profile picture	2-Categorical	4.3	< 0.1
Pre-highlights	Number of years highlighted before 2004	Numeric	3.9	N/A
Platform	User disabled third-party API integration	2-Categorical	1.6	< 0.1

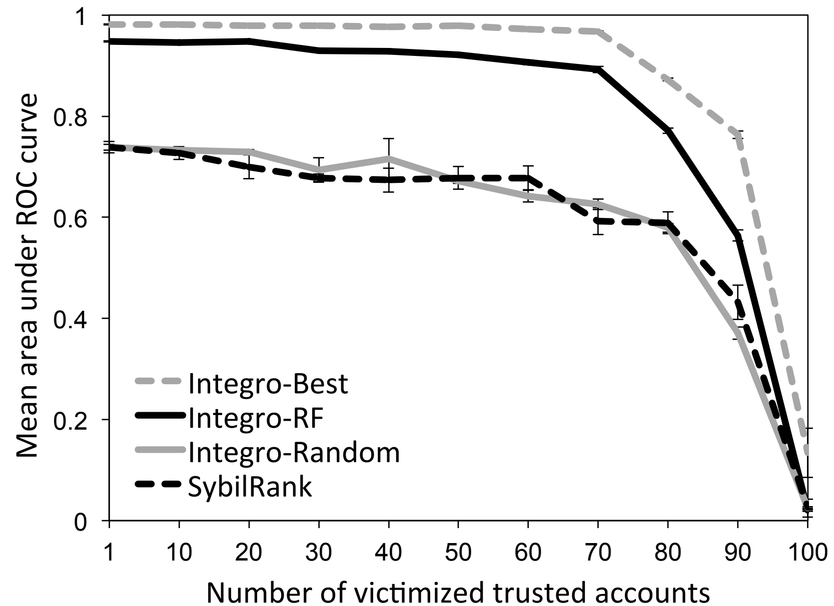
18 features(Facebook), 14 features (Tuenti)

Sensitivity to seed-targeting

Both systems are sensitive to seed-targeting attack, follow seed selection strategy



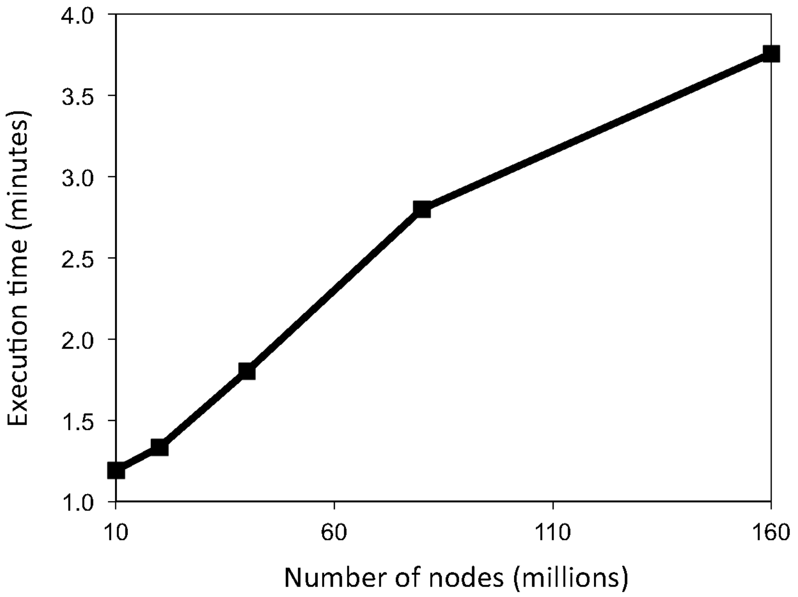
Distant-seed attack



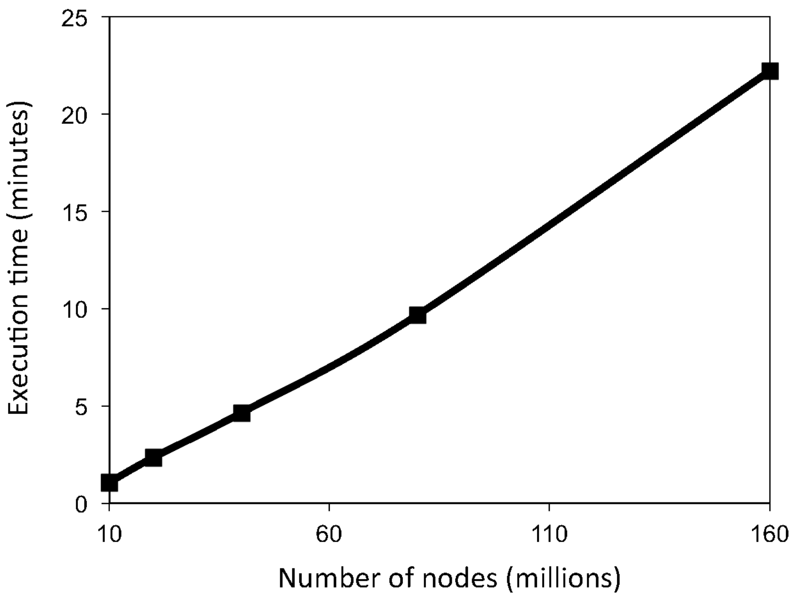
Random-seed attack

Scalability

Near linear scalability with number of accounts



RF is “embarrassingly parallel”



Ranking is “PageRank scalable”