NSEC lab
Network Security and Privacy Protection

Smoke Screener or Straight Shooter: Detecting Elite Sybil Attacks in User-Review Social Networks

Haizhong Zheng[1] Minhui Xue[2] Hao Lu[1] Shuang Hao[3]

Haojin Zhu[1] Xiaohui Liang[4] Keith Ross[2]

[1]Shanghai Jiao Tong University
[2]New York University Shanghai
[3]University of Texas at Dallas
[4]University of Massachusetts Boston

February, 2018

User-Review Social Networks

大众点评 dianping.com

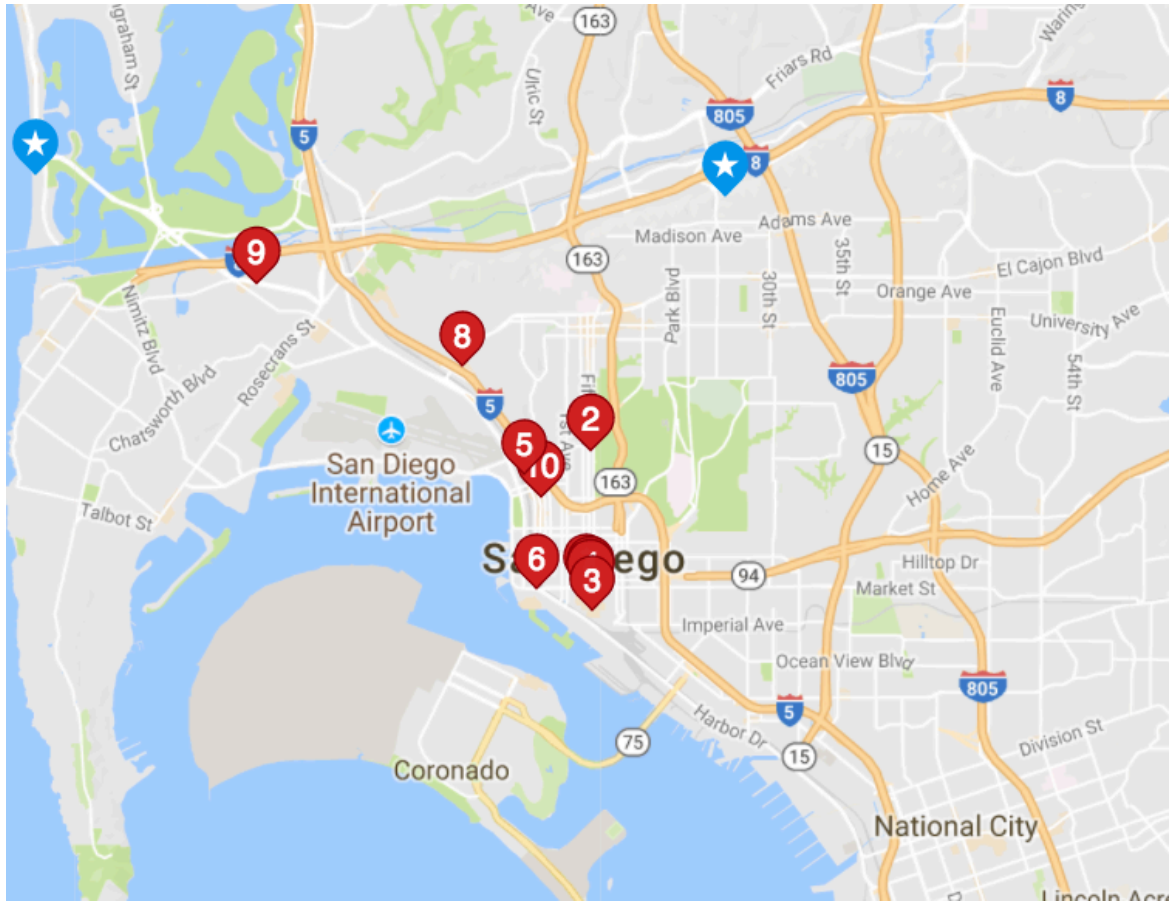
美团网 meituan.com

tripadvisor

Google My Business

yelp

Yelp: Search restaurants in San Diego



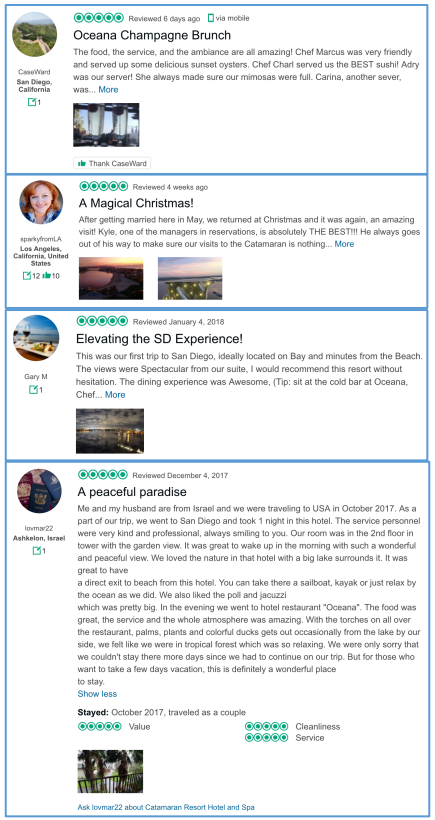
Reviews in User-Review Social Networks (URSNs)

大众点评 dianping.com

yelp

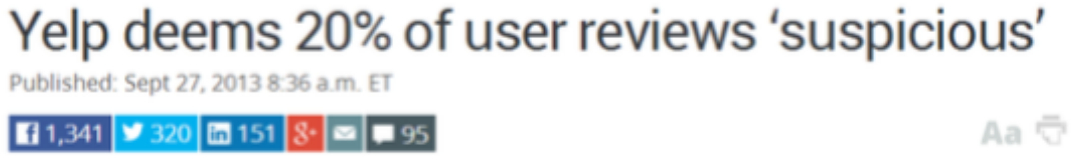
tripadvisor

Google My Business



Sybil Attacks in URSNs

MarketWatch:


Yelp deems 20% of user reviews 'suspicious'

Published: Sept 27, 2013 8:36 a.m. ET

1,341 320 151 95 Aa

Financial Times:

TripAdvisor fined in Italy over false reviews

Online travel website to challenge €500,000 penalty

Phoenix Finance: Dianping deleted 6 million fake reviews in half a year.

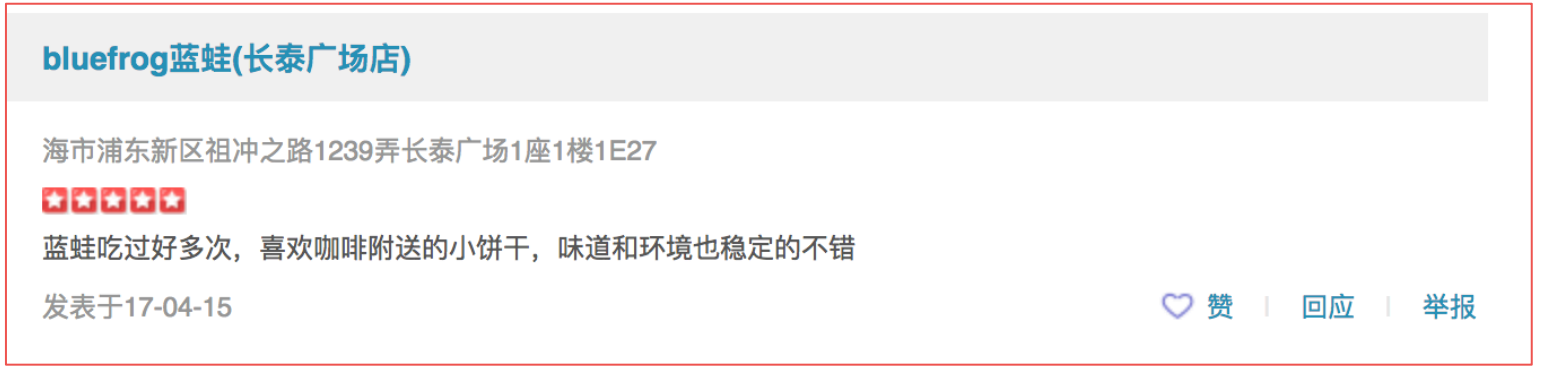
大众点评 dianping.com

Examples of fake reviews

Which one is the fake review?

Review 1

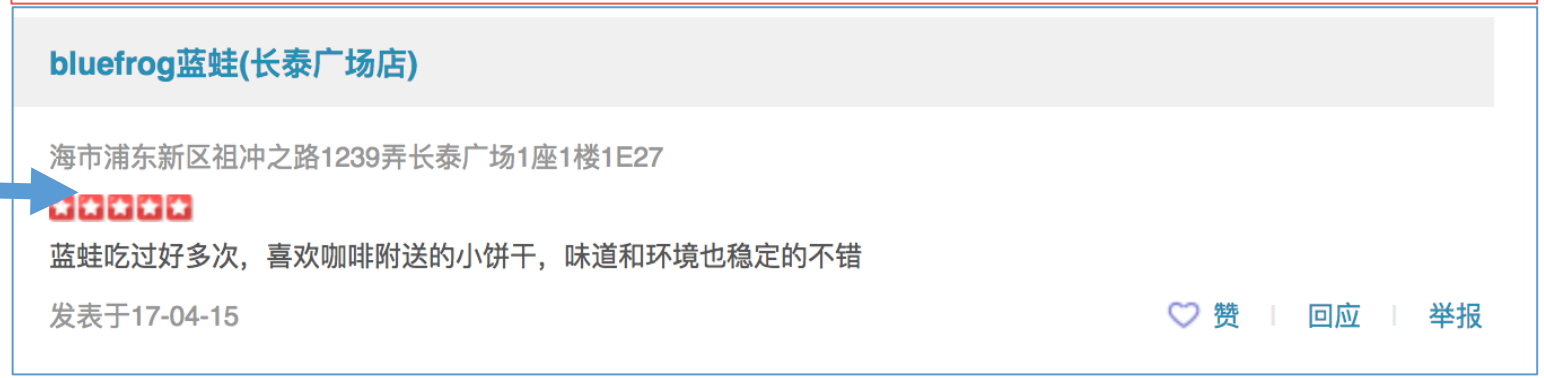
Review 2



Examples of fake reviews

A fake review posted by the volunteer

A real review posted by my friend



A New Type of Sybils

Elite Sybil users consist of two kinds of users.

1. Sybil accounts creating reviews not belonging to Sybil tasks to mimic benign users.
2. Accounts owned by benign users, but occasionally fulfilling a Sybil task to get the payments.

Elite Sybil User

Regular Sybil User

Outlines

01 Elite Sybil Attacks

02 ELISEDET

03 Evaluation & Measurement

04 Conclusion

01

ELITE SYBIL ATTACKS

Sybil Organization Architecture

01

Customers (or Overhyped stores)

Stores want to boost their scores rapidly.

02

Agents

Agents are responsible for accepting tasks from customers and launching campaigns.

03

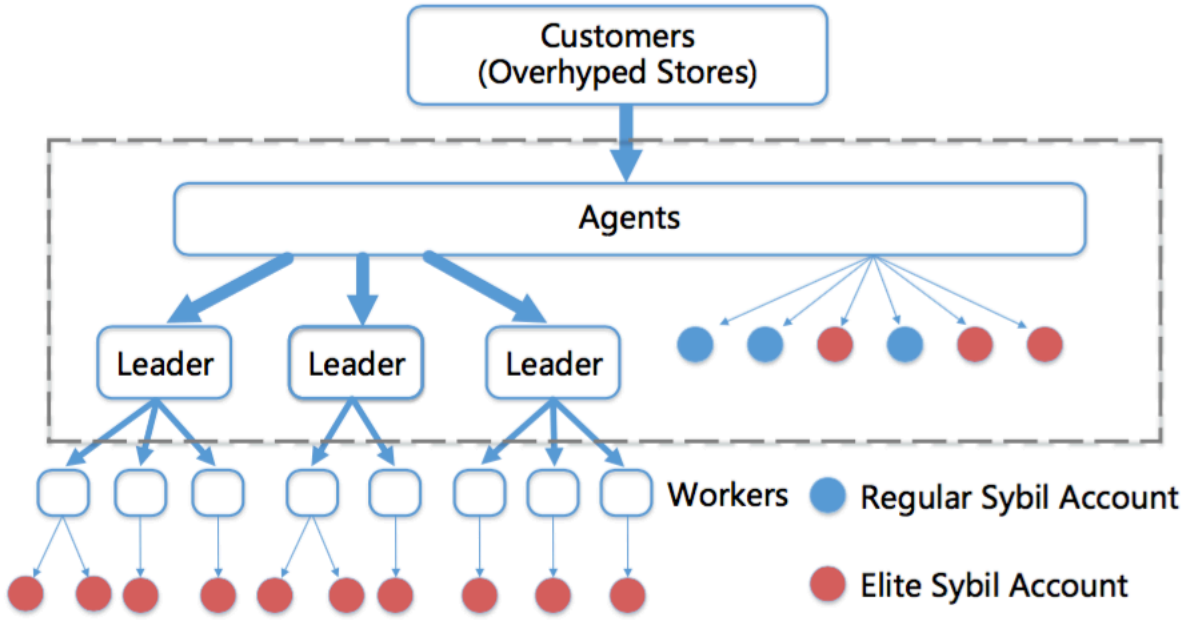
Leaders

Leaders take charge of recruiting workers and distributing tasks to workers.

04

Elite Sybil workers

Elite Sybil workers are internet users post fake reviews for profit.



01 ELITE SYBIL ATTACKS

Sybil Organization Architecture

01

Customers (or Overhyped stores)

Stores want to boost their scores rapidly.

02

Agents

Agents are responsible for accepting tasks from customers and launching campaigns.

03

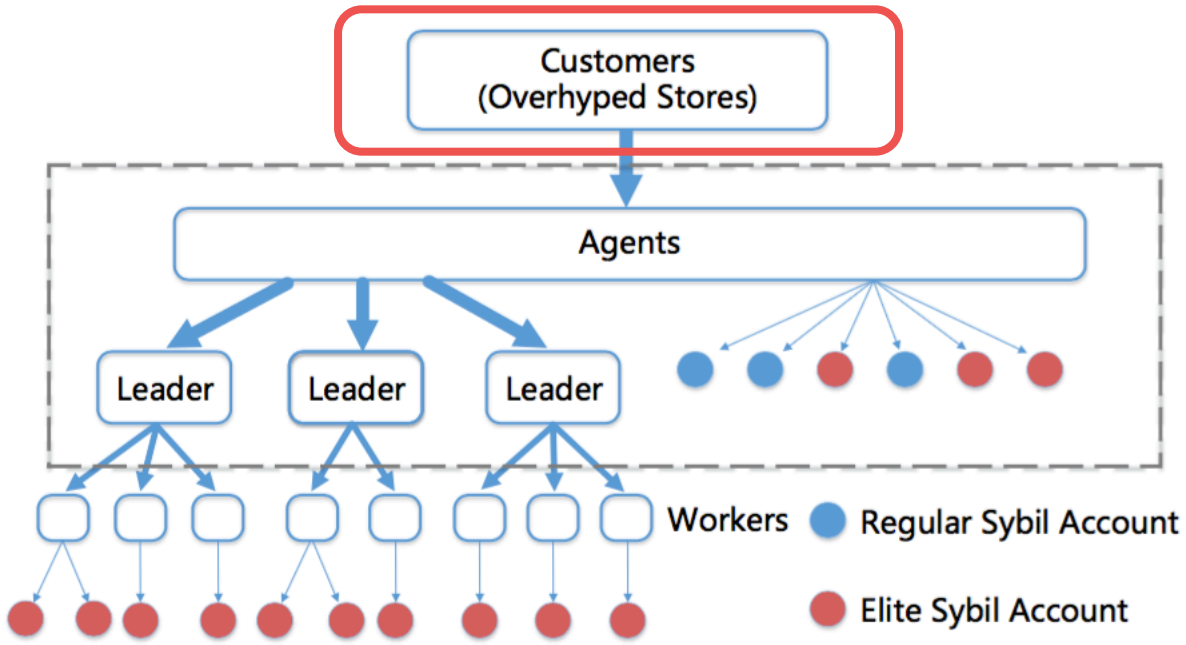
Leaders

Leaders take charge of recruiting workers and distributing tasks to workers.

04

Elite Sybil workers

Elite Sybil workers are internet users post fake reviews for profit.



01

ELITE SYBIL ATTACKS

Sybil Organization Architecture

01

Customers (or Overhyped stores)

Stores want to boost their scores rapidly.

02

Agents

Agents are responsible for accepting tasks from customers and launching campaigns.

03

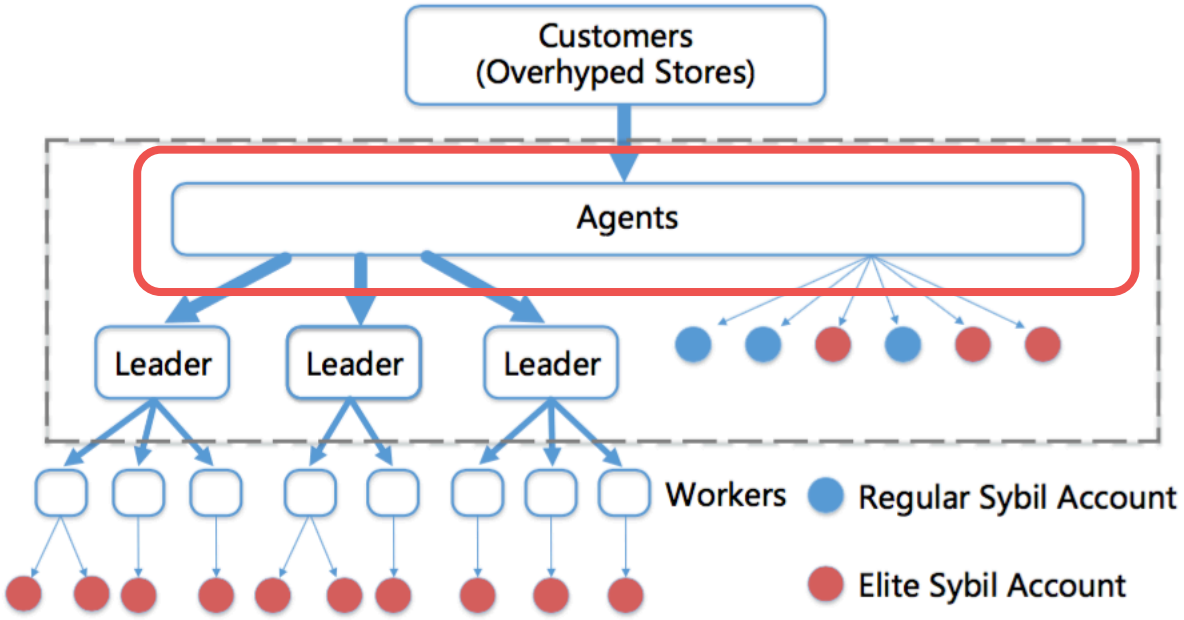
Leaders

Leaders take charge of recruiting workers and distributing tasks to workers.

04

Elite Sybil workers

Elite Sybil workers are internet users post fake reviews for profit.



01 ELITE SYBIL ATTACKS

Sybil Organization Architecture

01 Customers (or Overhyped stores)

Stores want to boost their scores rapidly.

02 Agents

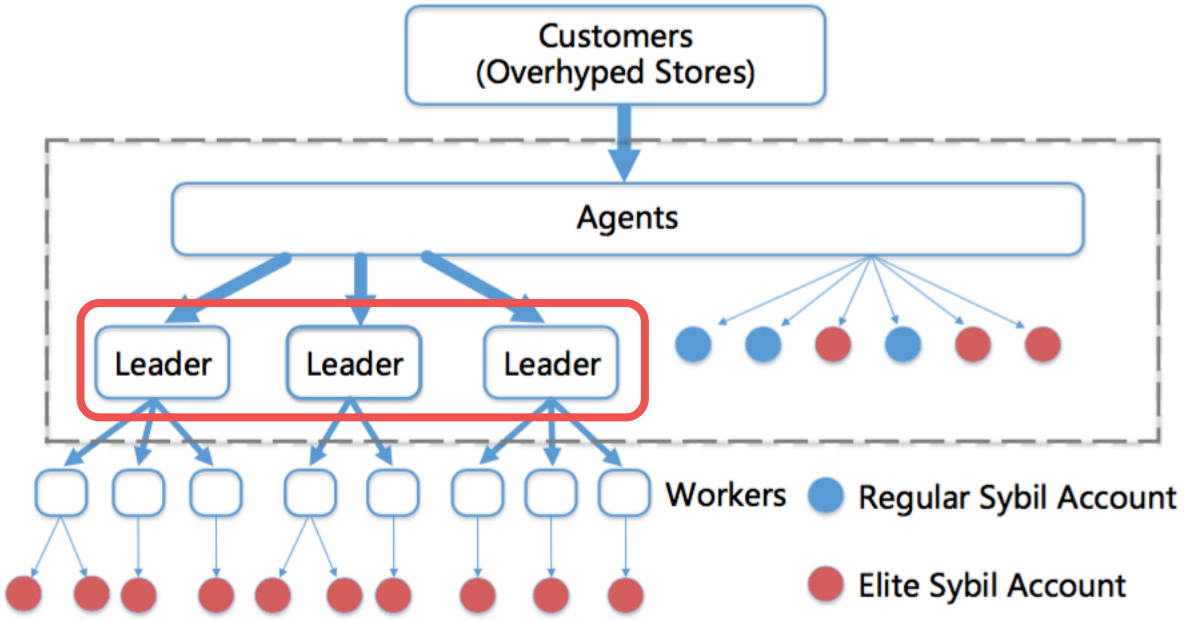
Agents are responsible for accepting tasks from customers and launching campaigns.

03 Leaders

Leaders take charge of recruiting workers and distributing tasks to workers.

04 Elite Sybil workers

Elite Sybil workers are internet users post fake reviews for profit.



01 ELITE SYBIL ATTACKS

Sybil Organization Architecture

01

Customers (or Overhyped stores)

Stores want to boost their scores rapidly.

02

Agents

Agents are responsible for accepting tasks from customers and launching campaigns.

03

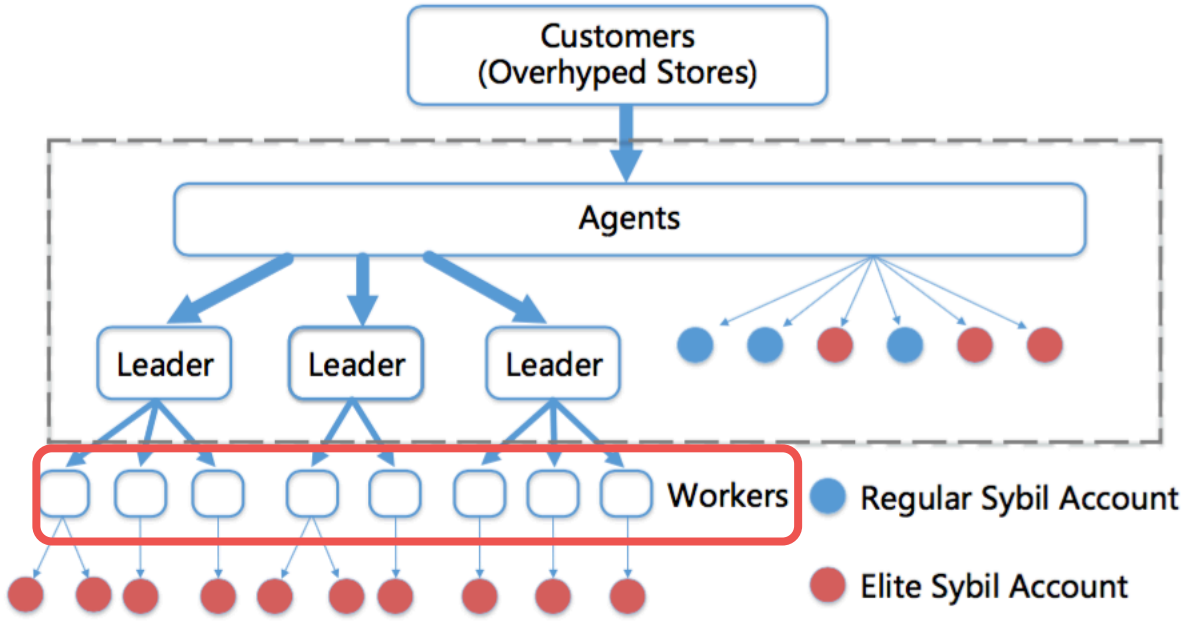
Leaders

Leaders take charge of recruiting workers and distributing tasks to workers.

04

Elite Sybil workers

Elite Sybil workers are internet users post fake reviews for profit.



01 ELITE SYBIL ATTACKS

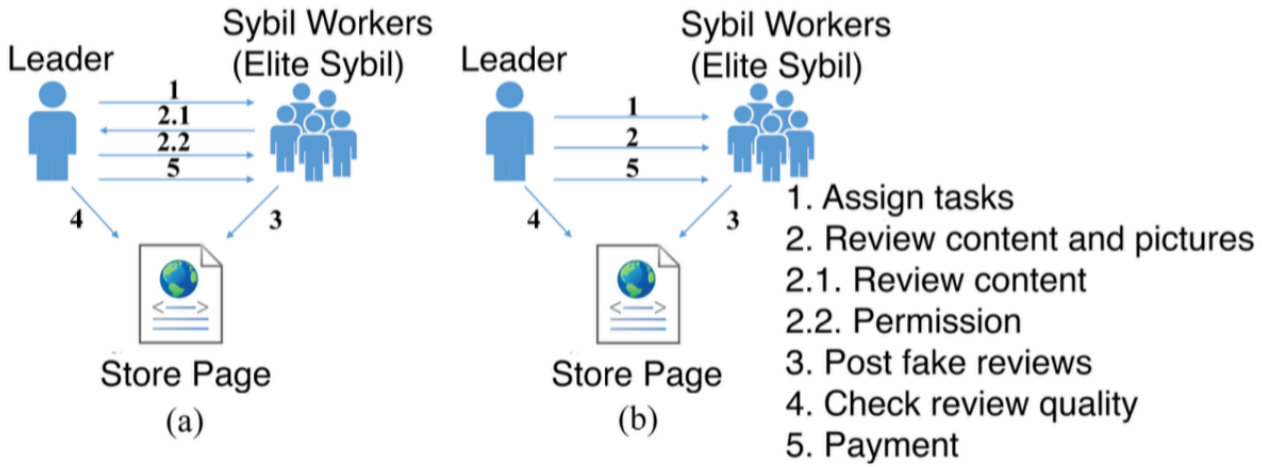
Typical Workflow

a **Leader-supervised model**

The reviews generated by elite Sybil workers

b **Leader hands-on model**

The content of reviews is provided by leaders.



The quality of fake reviews is strictly controlled!!!

Why do Sybil organizations put so much effort to control the quality of reviews?

Economic Factors

The payments vary with ratings of accounts:

TABLE I. HIERARCHICAL REWARDS FOR (ELITE) SYBIL WORKERS

Ratings of Accounts	Rewards per Submission
0-star, 1-star	$0.30
2-star	$0.75
3-star	$1.50
4-star	$3.74
5-star, 6-star	**$5.98**

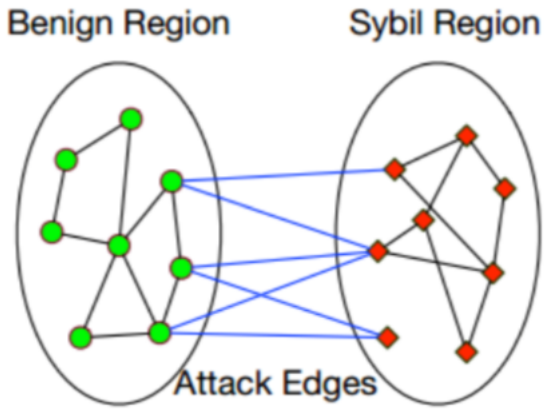
01

ELITE SYBIL ATTACKS

Challenges

Weak connectivity of User-Review Social Networks (URSNs)

Graph-based detection approaches may not work.



Highly adaptive and professional

Imitate real users in user profiles and review content.
Behavior-based detection approaches may not work.

More active out of Sybil campaigns

Lower percentage of fake reviews in their posts.
Clustering approaches may not work.

Outlines

01 Elite Sybil Attacks

02 ELISEDET

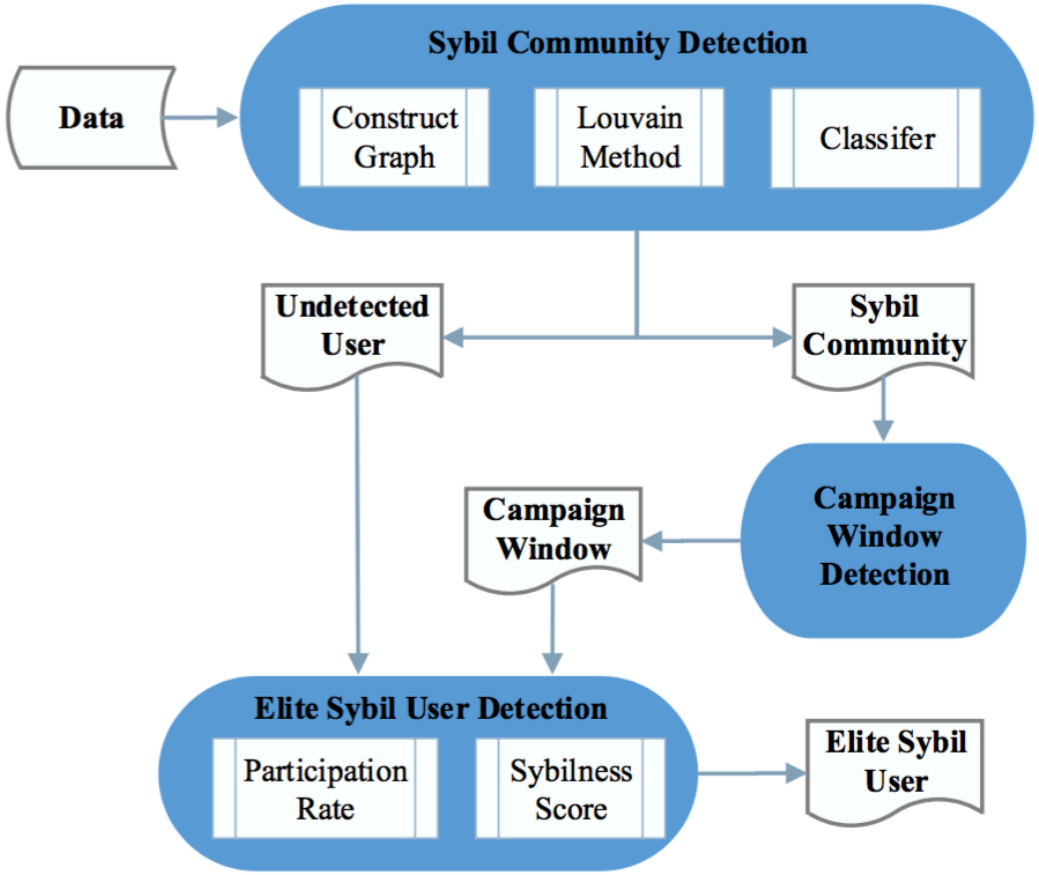
03 Evaluation & Measurement

04 Conclusion

02 ELISEDET

Overview

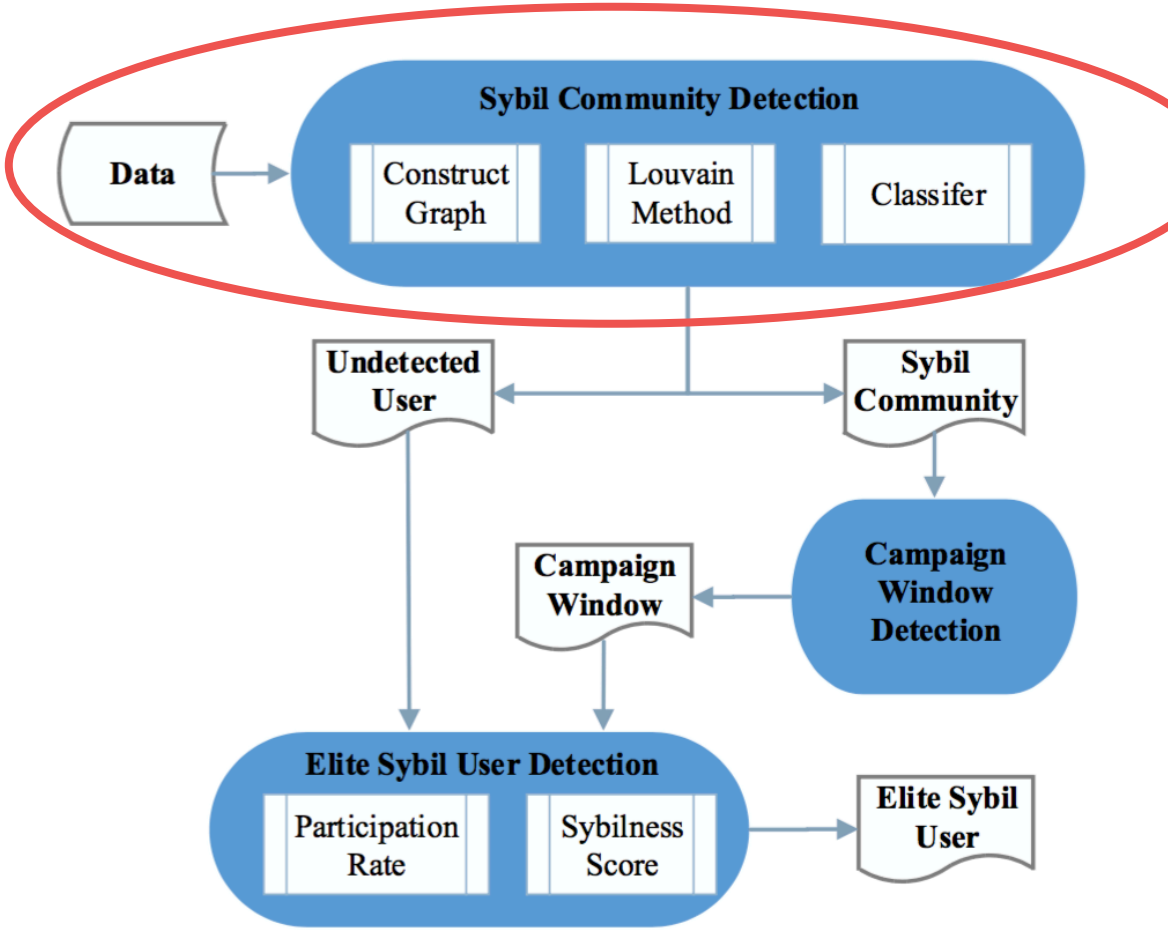
- Sybil Community Detection
- Campaign Window Detection
- Elite Sybil User Detection



02 ELISEDET

Overview

- Sybil Community Detection
- Campaign Window Detection
- Elite Sybil User Detection



02 ELISEDET

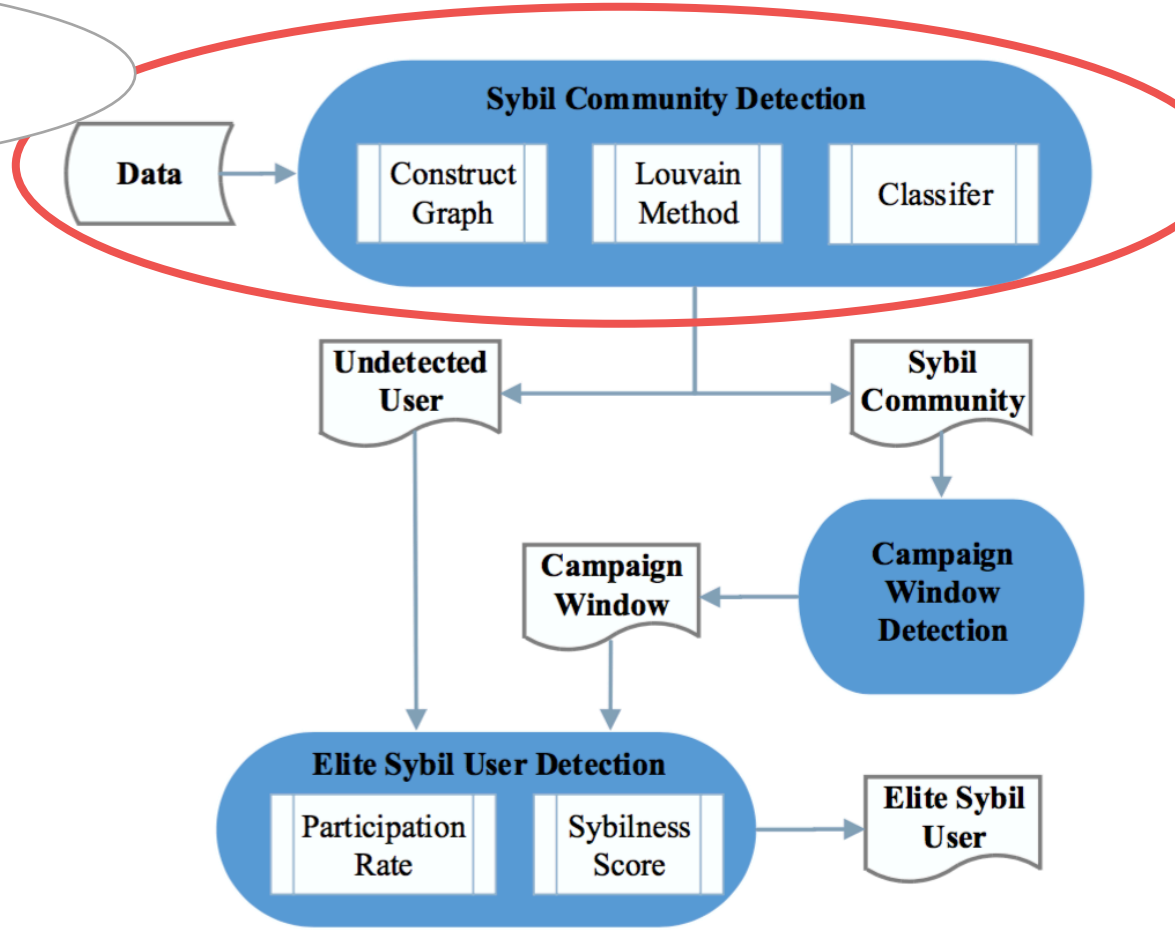
Overview

Challenge 1: Weak connectivity
Challenge 2: Professional fake reviews

Sybil Community Detection

Campaign Window Detection

Elite Sybil User Detection



Challenge 1: Build the relation between users via collusive reviews.

02 ELISEDET

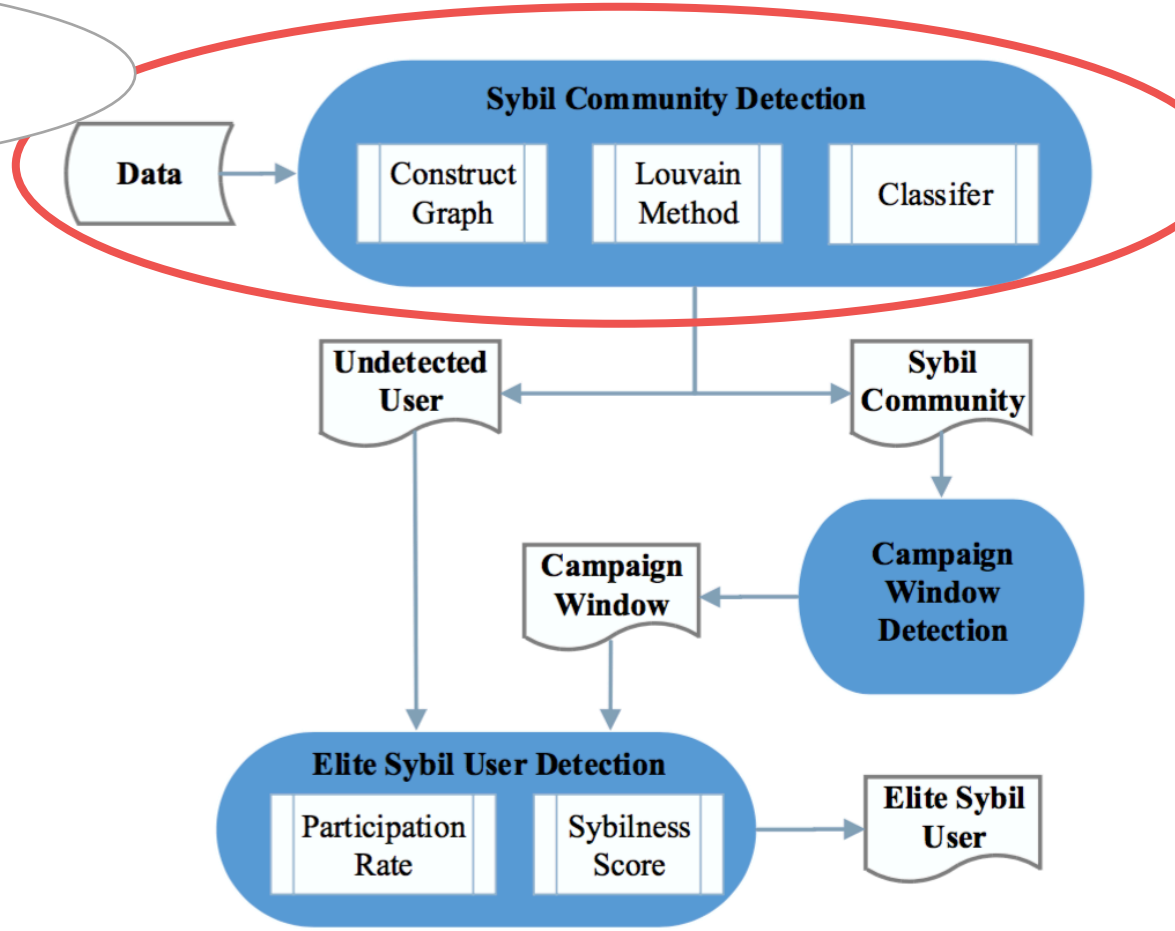
Overview

Challenge 1: Weak connectivity
Challenge 2: Professional fake reviews

Sybil Community Detection

Campaign Window Detection

Elite Sybil User Detection

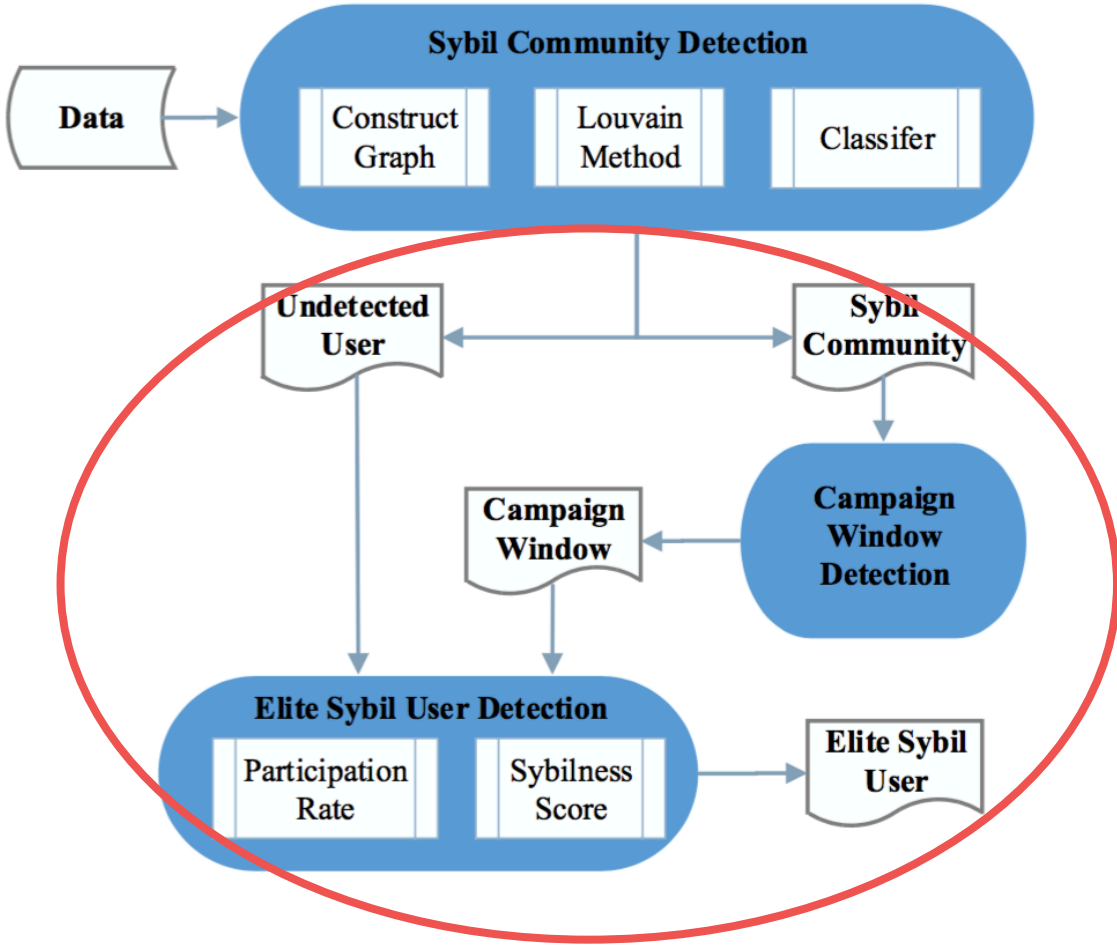


Challenge 2: Use features irrelevant to user profiles and review content.
(community-based features, network features, content-free user features)

02 ELISEDET

Overview

- Sybil Community Detection
- Campaign Window Detection
- Elite Sybil User Detection



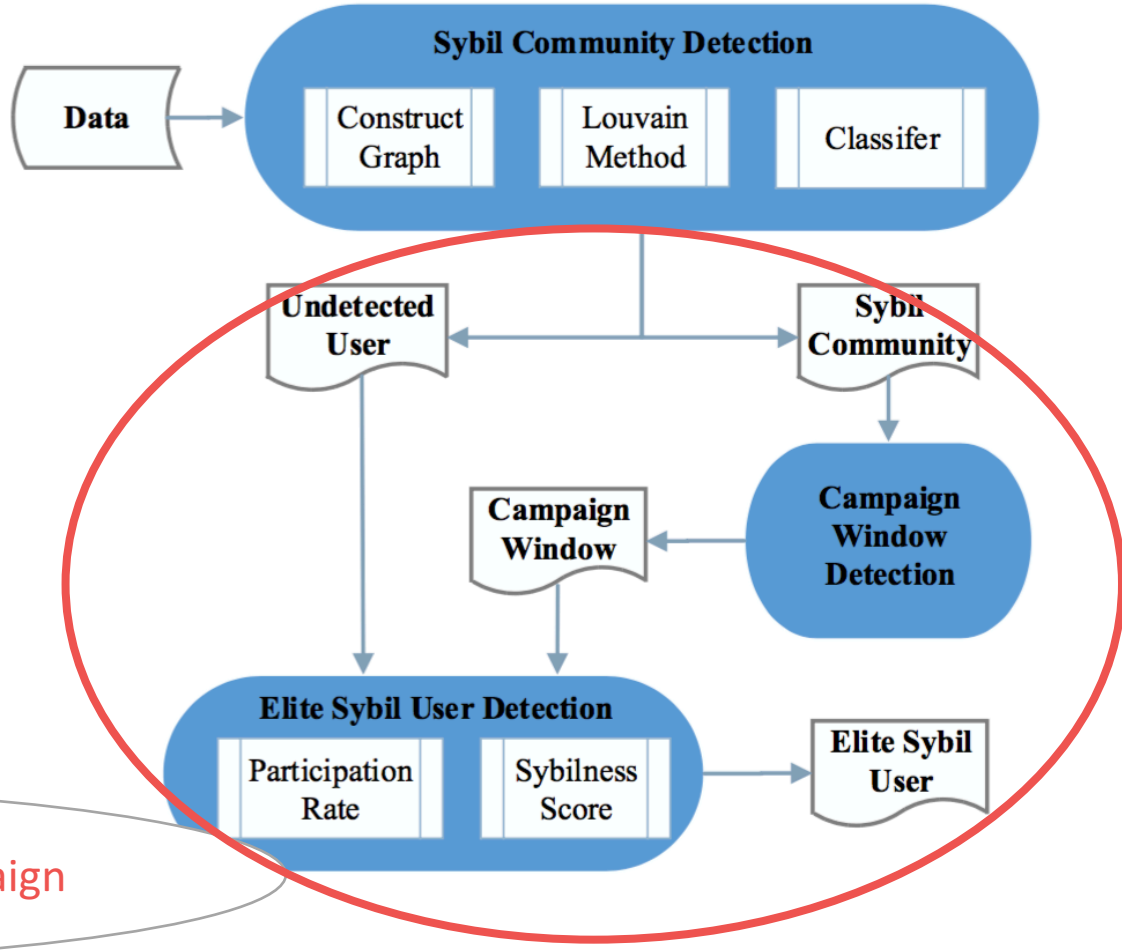
02 ELISEDET

Overview

Sybil Community Detection

Campaign Window Detection

Elite Sybil User Detection



Challenge 3: More active out of Sybil campaign

Challenge 3: Use campaign windows to filter out the reviews irrelevant to Sybil Campaigns

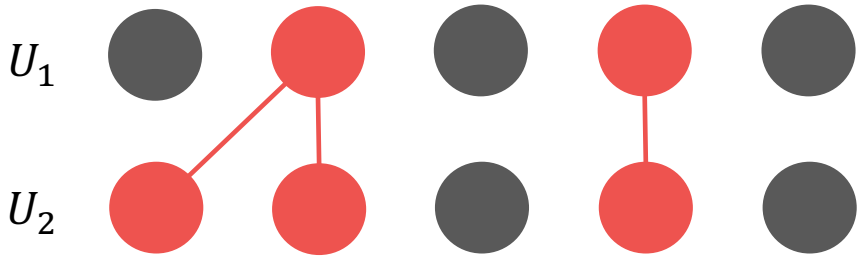
Build Relations between Users

01 Collusive reviews

Posted by two users in a same store in a short time period Δ T with the same rating (same goal).

02 Collusive relation between users

$$R(U_1, U_2) = \frac{|\text{Collusive Reviews}|}{|\text{Reviews of } U_1| + |\text{Reviews of } U_2|}$$



$$R(U_1, U_2) = \frac{5}{5+5} = 0.5$$

Sybil Community Detection

Detect Sybil communities in three steps:

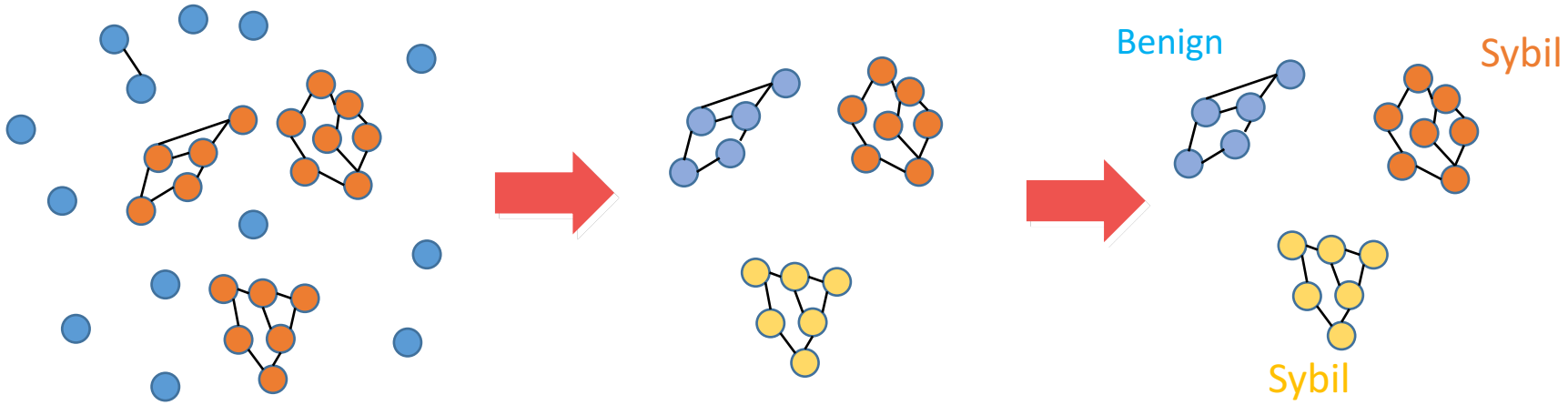
Build Relations between Users → Louvain Community Detection → Classification

- Community-based Features
- Network Features
- Content-free User Features

Build collusive relations and construct graphs

Detect communities

Classify click farming communities



02 ELISEDET

Campaign Window Detection

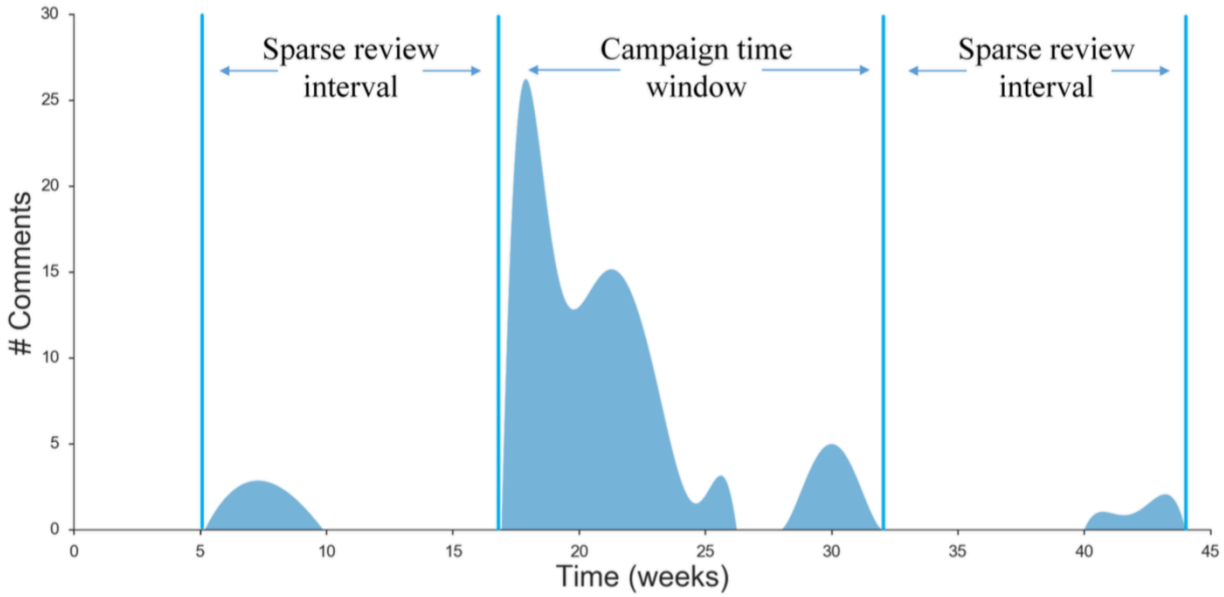
01 Sparse interval:

The number of weeks with at least one review is less than the number of weeks without any reviews.

02 Iteratively find and delete sparse review intervals

ALGORITHM 1: Detecting Campaign Time Windows

Input: A list L_{review} whose item $L_{review}[i]$ denotes the number of reviews posted in the ith week.
Output: The start point l and end point r of the campaign time window.
Initial:
$l \leftarrow 0$;
$r \leftarrow length(L_{review}) - 1$;
while (true) do
  $I_{l,l'} \leftarrow find(left, l)$; {Find the first sparse interval $I_{l,l'}$ from left.}
  $I_{r',r} \leftarrow find(right, r)$; {Find the first sparse interval $I_{r',r}$ from right.}
  if ($l' = r$ **and** $r' = l$) **then** {There is no sparse interval.}
    break;
  end if
  if ($|I_{l,l'}| \leq |I_{r',r}|$) **then** {Choose the interval with fewer reviews.}
    $l \leftarrow l' + 1$;
  else
    $r \leftarrow r' - 1$;
  end if
end while
return l, r;



02 ELISEDET Elite Sybil User Detection

01 Participation rate between users and communities

Just consider the reviews posted in Sybil campaigns.

$N_C(k)$: the accumulated number of reviews posted C within the kth time window of community C.

N_C^{max}: the maximum number of reviews posted within all time windows of community C.

$$P_C(k) = \frac{N_C(k)}{N_C^{max}}$$ Indicate the importance of kth time interval

$$N_{u \in C} = \sum_k P_C(k) \cdot N_{u \in C}(k)$$ Indicate the importance of reviews u posted for community C

Participation rate: $$\rho_{u \in C} = \frac{1}{1 + \exp^{-\frac{N_{u \in C} - \mu_C}{\sigma_C}}}, \quad \text{for any } u \in C,$$ Standardize with sigmoid function

02 Sybilness

$$f(u) = \sum_C \rho_{u \in C} \cdot N_{u \in C}$$ **Perceived likelihood indicating if a user is an elite Sybil user**

Outlines

01 Elite Sybil Attacks

02 ELISEDET

03 Evaluation & Measurement

04 Conclusion

03 EVALUATION

Detection Results

Dataset

32,933 stores, 3,555,154 users and **10,541,931 reviews**

Community

566 Sybil communities with 22,324 users and 144 benign communities with 5,222 users

Elite Sybil User

1. User ***u*** does not belong to any Sybil communities.
2. There is a community ***C*** and the user participation rate $\boldsymbol{\rho_{u \in C}}$ is larger than 0.5 (average participation rate in a community).
3. **12,292 elite Sybil users.**

03 EVALUATION

Detection Results

Dataset

32,933 stores, 3,555,154 users and **10,541,931 reviews**

大众点评 dianping.com

Community

566 Sybil communities with 22,324 users and 144 benign communities with 5,222 users

Elite Sybil User

1. User ***u*** does not belong to any Sybil communities.
2. There is a community ***C*** and the user participation rate $\rho_{u \in C}$ is larger than 0.5 (average participation rate in a community).
3. **12,292 elite Sybil users.**

TABLE III. BREAKDOWNS OF STORES

Type	# Stores	# Overhyped Stores	Percentage of Overhyped Stores
Cinema	235	71	**30.21%**
Hotel	1,738	134	7.71%
Restaurant	22,474	1,244	5.54%
Entertainment	1,384	73	5.27%
Wedding Service	320	8	2.50%
Beauty Store	1,460	35	2.40%

03 EVALUATION

Performance

Sybil Community

Classifier	Precision	Recall	F1	AUC
Decision tree	93.80 %	92.90 %	93.60 %	92.83 %
SVM	96.74 %	96.47 %	**96.45%**	**99.42%**
GNB	94.21 %	93.44 %	93.57 %	97.64 %
KNN	96.75 %	96.47 %	96.50 %	97.45 %
Ada boost	93.84 %	93.54 %	93.60 %	97.92 %
Random forest	93.16 %	94.01 %	92.99 %	97.42 %

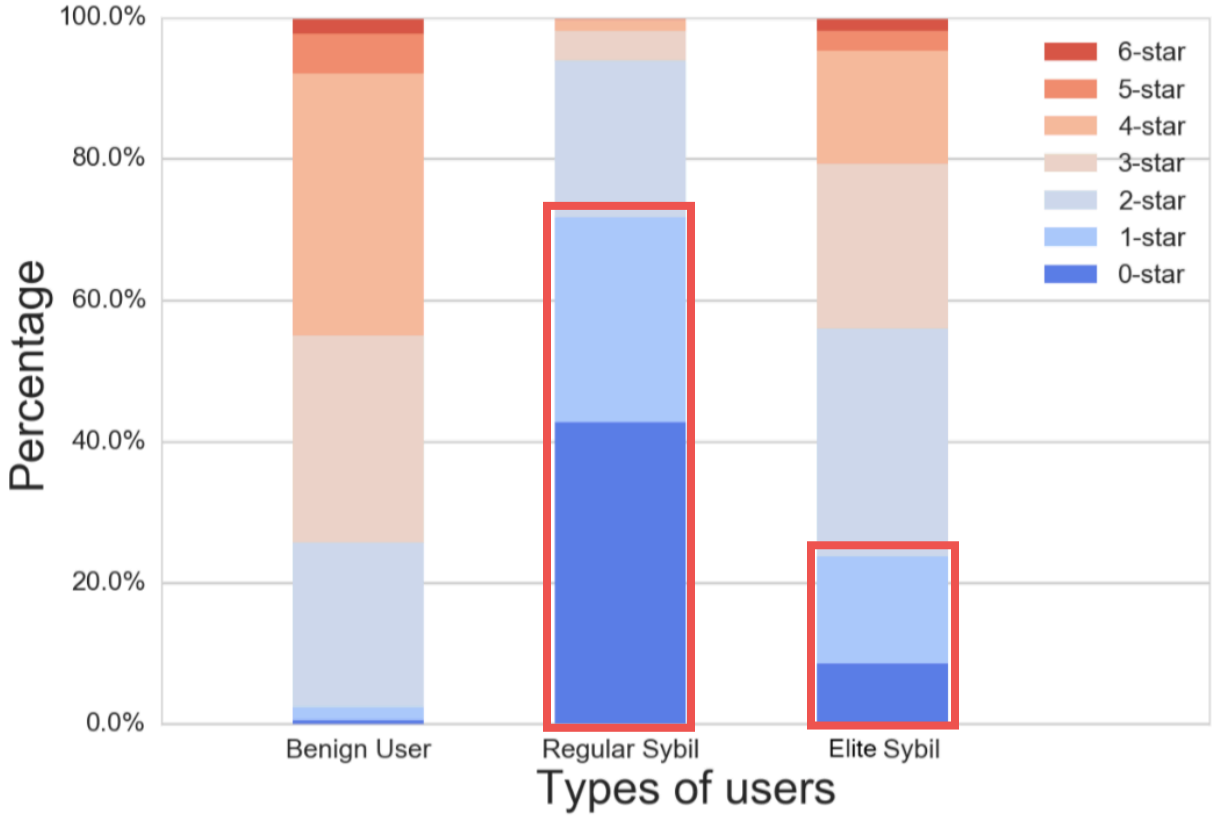
Elite Sybil User

Type of users	Precision
Sybilness top 1,000 users	93.80%
Random 1,000 users	90.70%

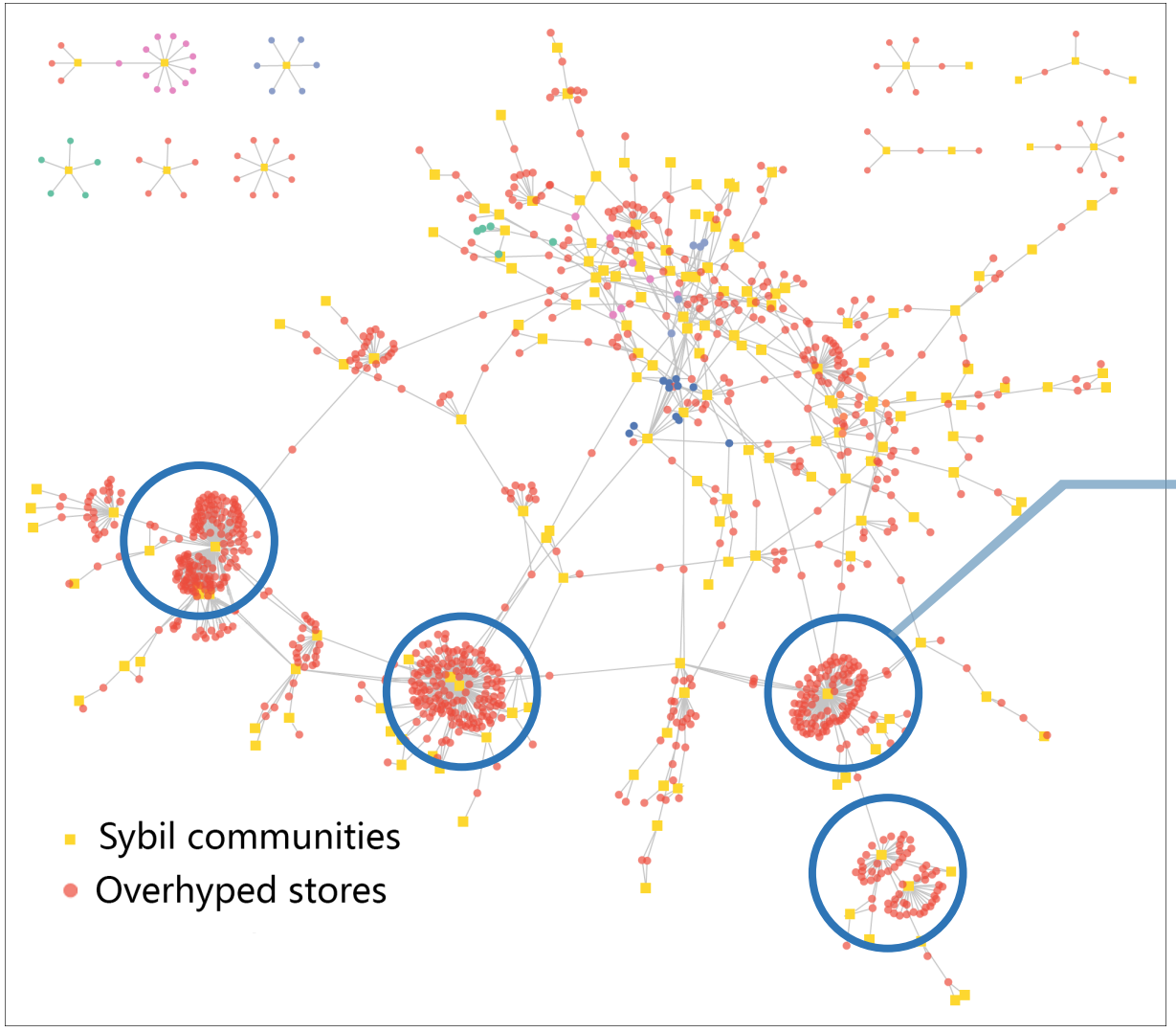
03 MEASUREMENT Comparison with Regular Sybil User

User-level star ratings

Compared to regular Sybil users,
elite Sybil users have higher user levels.



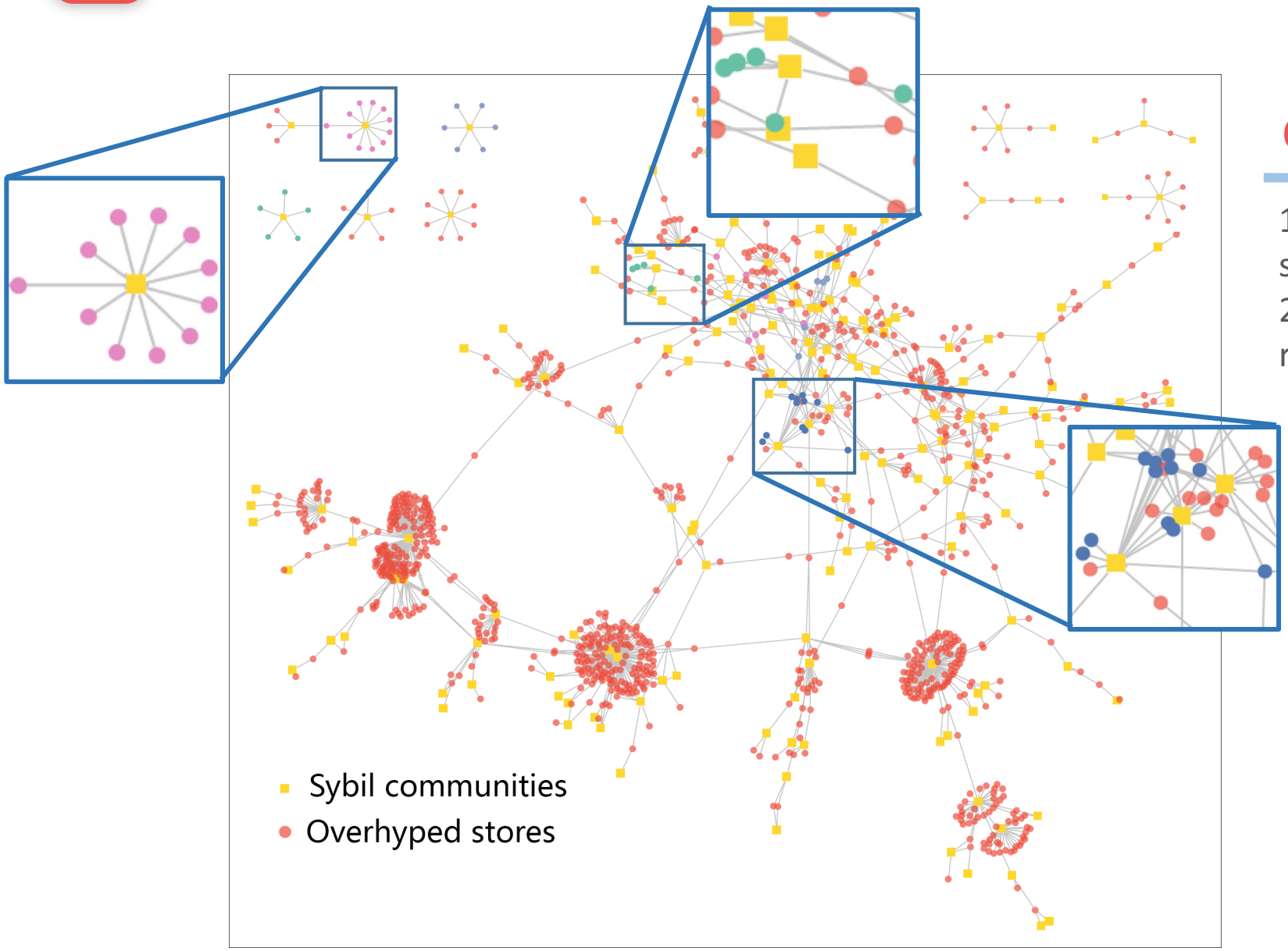
Relation between Communities and Stores



Sybil Communities

Some Sybil communities provide services for a huge number of overhyped stores.

Relation between Communities and Stores



Chain store

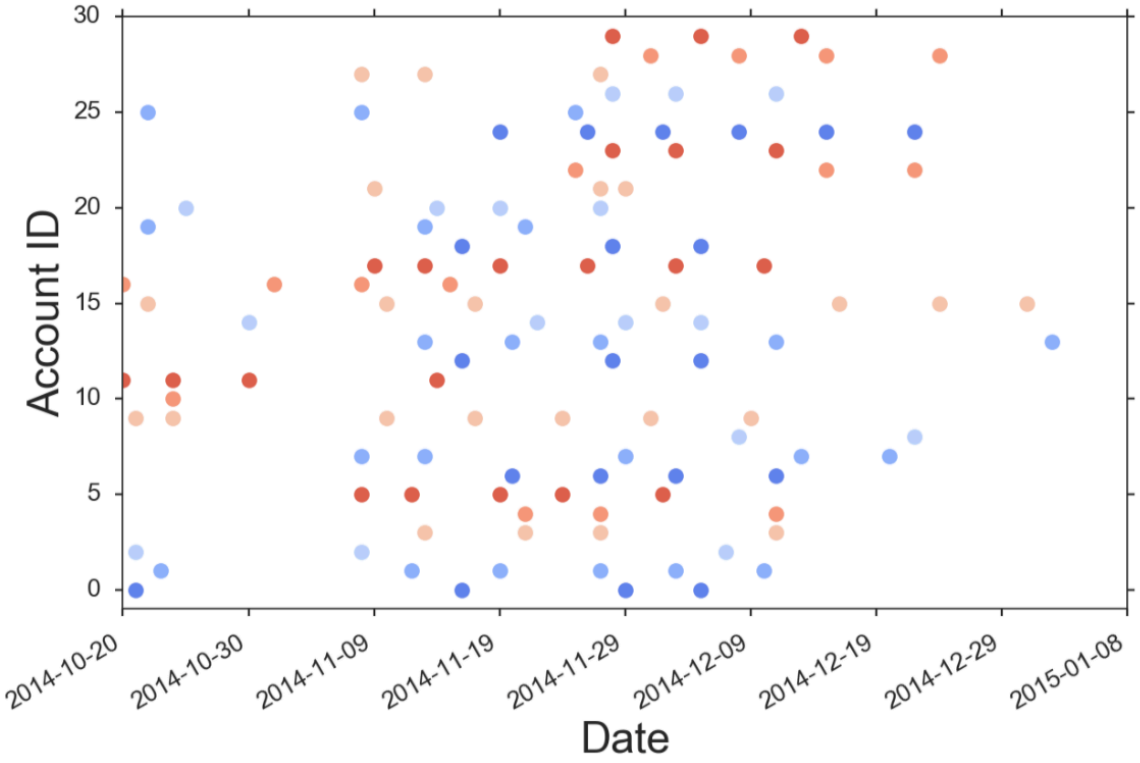
1. Except for red dots, other dots with same colors are chain stores.
2. 12.37% of Sybil communities post fake reviews for chain stores.

03 MEASUREMENT

Case Study

User posting period

Elite Sybil users post fake reviews periodically.
Evading detection method basing on posting frequency.



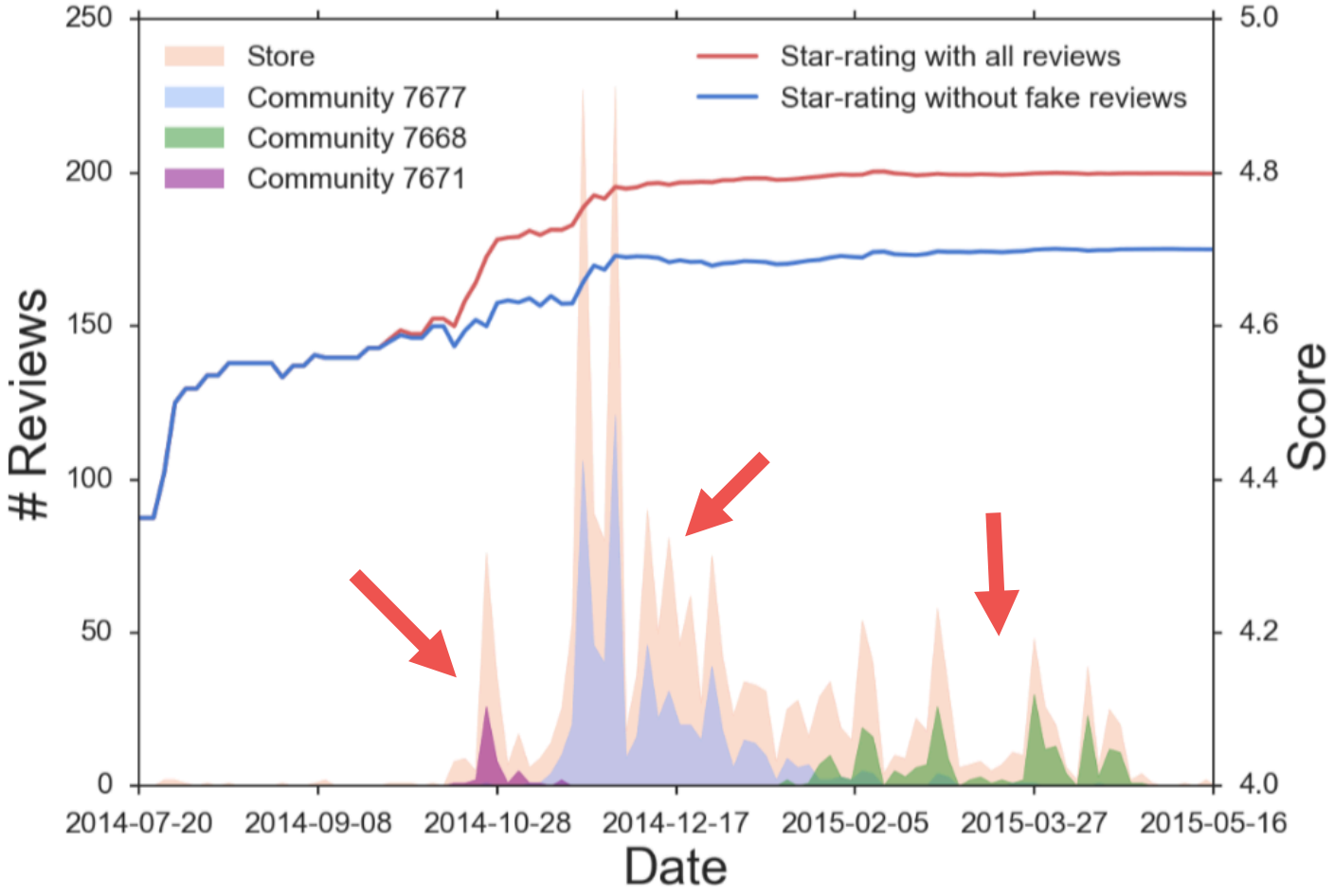
Reviews posted by Community 4559 in Store 4112200

03 MEASUREMENT

Case Study

The review spike

Three different Sybil communities posted fake reviews in a store and caused the review spikes.
Increase in the number of reviews.



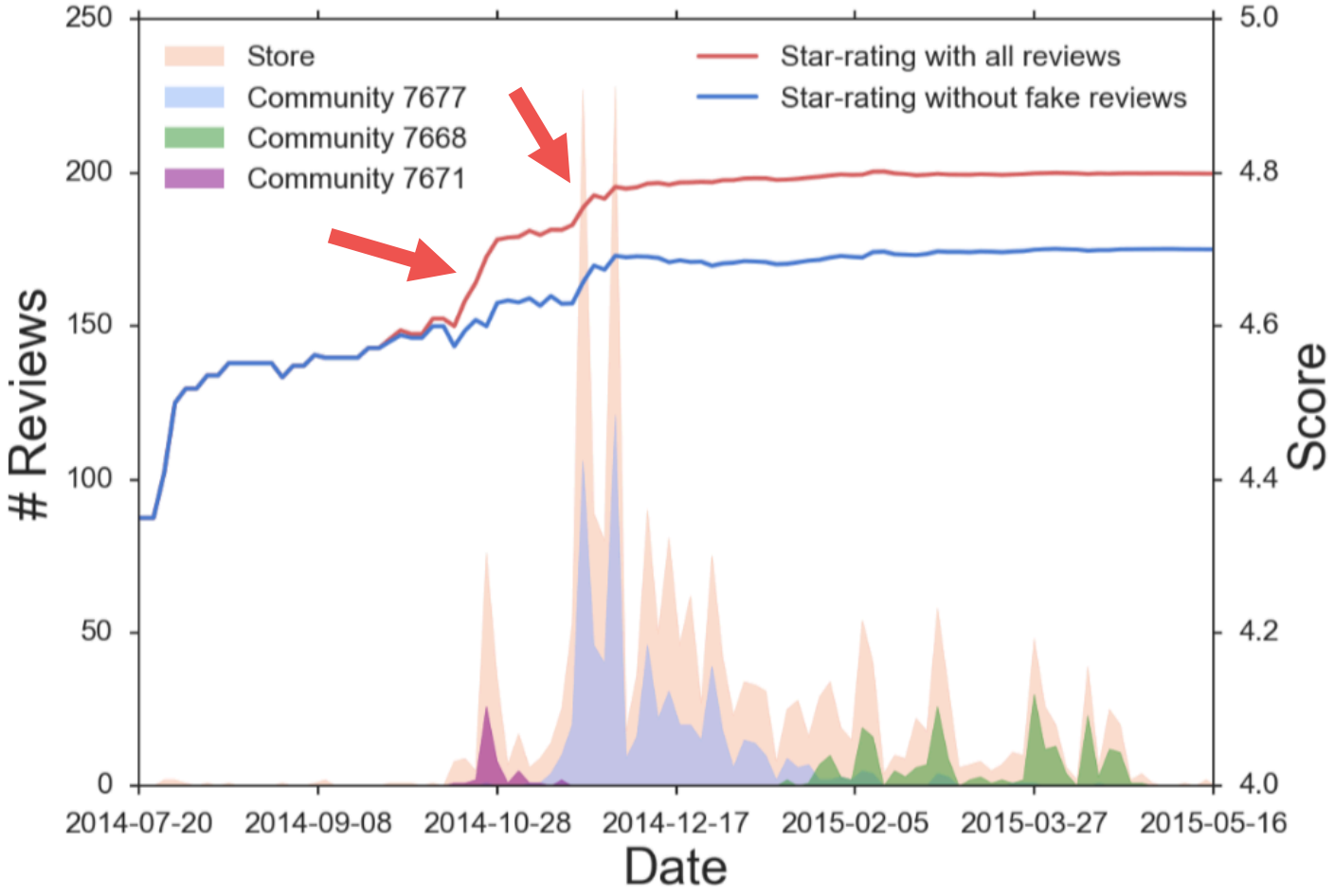
Variation of star ratings and the number of reviews of a hotel

03 MEASUREMENT

Case Study

Increase in rating

The reviews spikes lead to increases in the rating of the store.



Variation of star ratings and the number of reviews of a hotel

Outlines

01 Elite Sybil Attacks

02 ELISEDET

03 Evaluation & Measurement

04 Conclusion

04 CONCLUSION

Novel attack: Elite Sybil attack

Elite Sybil users are more spread out temporally, craft better edited contents, but have fewer reviews filtered

Elite Sybil Detection System: ELISEDET

Both highly effective and scalable as a standalone system

Measurement about elite Sybil users

Elite Sybil have higher user-levels and deliberately control the posting time

Thanks for listening!

Questions?

sjtu.zhenghaizhong@gmail.com