Towards Detecting Anomalous User Behavior in Online Social Networks

Bimal Viswanath

M. Ahmad Bashir, Mark Crovella, Saikat Guha, Krishna Gummadi, Balachander Krishnamurthy, Alan Mislove

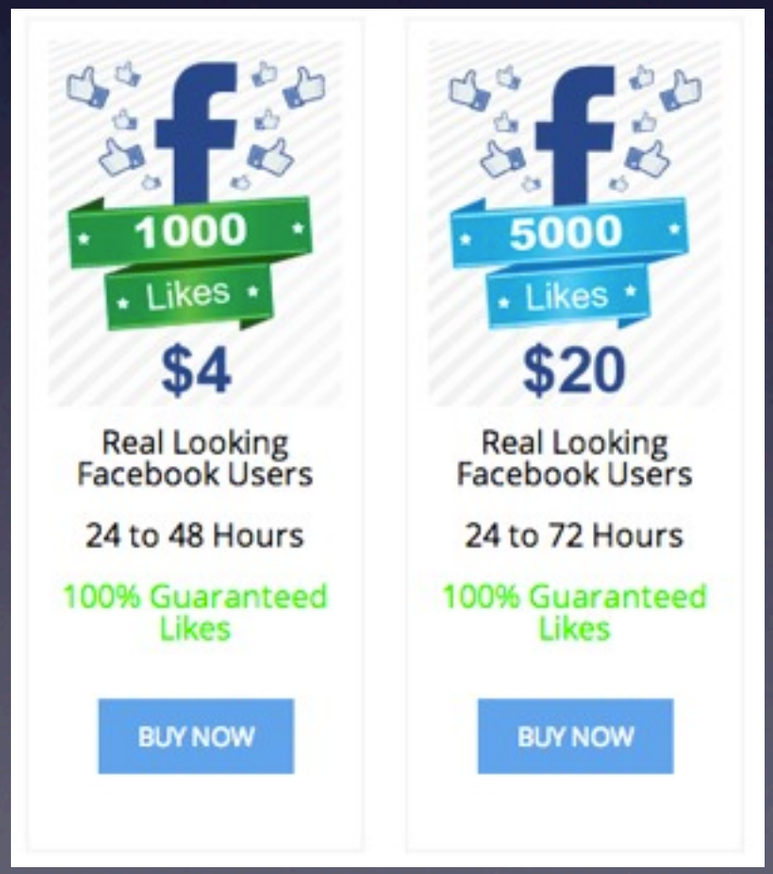
Usenix Security'14

Service abuse in social networks today

Several black-market services are available today to

- Manipulate content ratings
- Manipulate popularity of a user

Like spammers try to boost popularity of Facebook pages

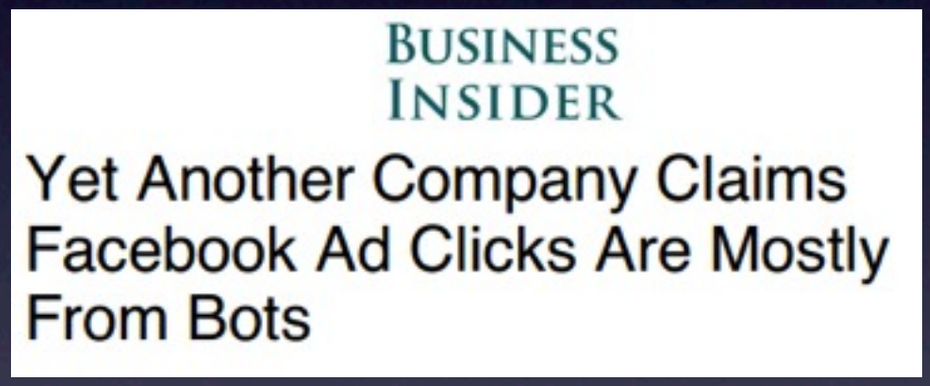


Service abuse in social networks today

Service abuse can have significant economic consequences

Social advertising services also seem to be targeted by attackers





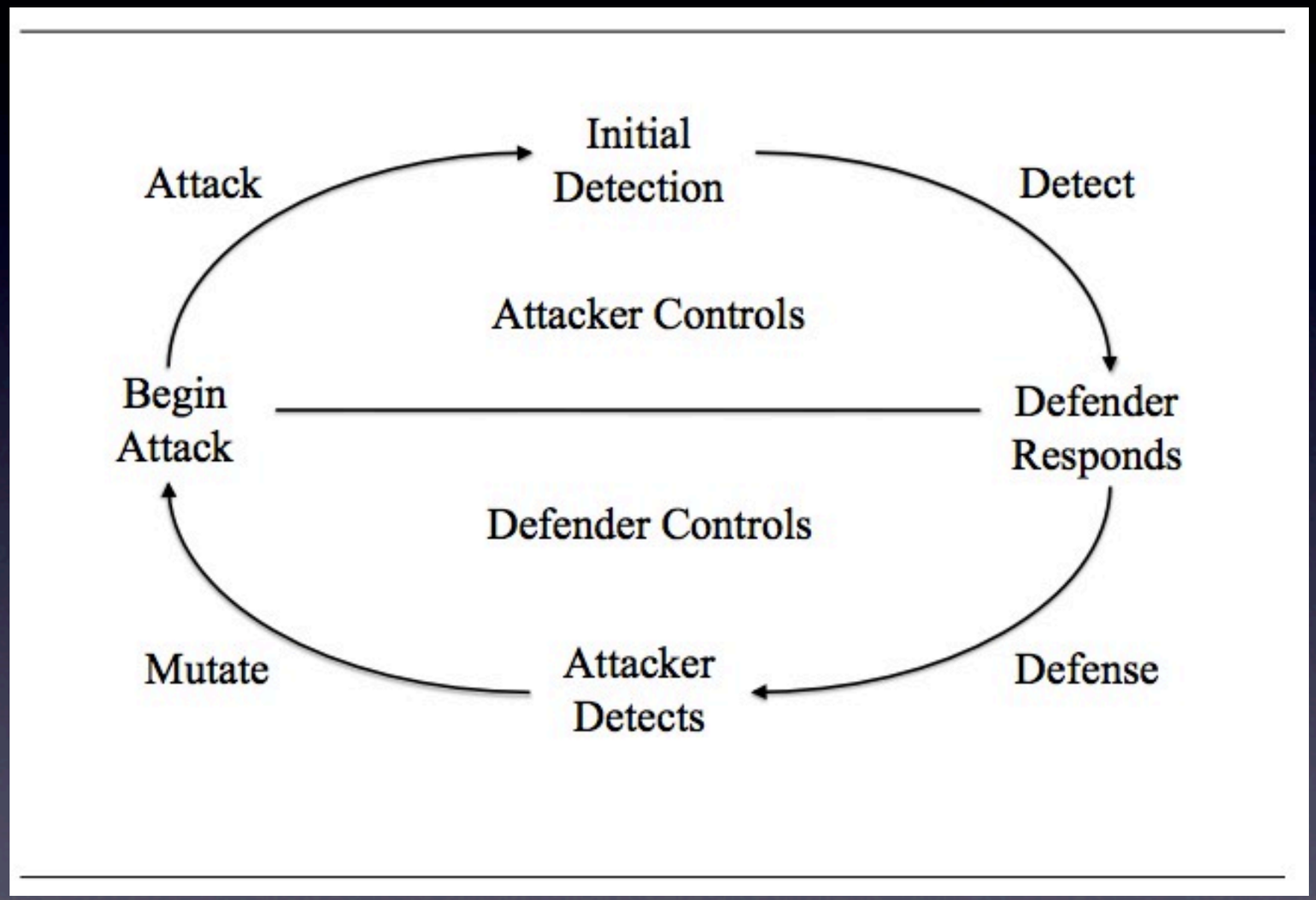


Our goal

Detect misbehaving identities in a social networking service

- Suspend the misbehaving user or nullify their actions

Adversarial cycle today



"Facebook Immune System", SNS'10

Limitations of existing approaches

Relies on detecting specific known patterns of misbehavior

Attackers mutate and use diverse strategies today:

- Fake accounts are created for Sybil attacks
- Some real users tend to collude to boost each other's popularity
- Real user accounts are compromised for better social reach

Existing approaches are vulnerable against an adaptive attacker

Idea: Use anomaly detection on user behavior

Our approach at a high level

We build an Anomaly classifier

- That learns normal patterns of user behavior
- Any behavior that deviates significantly from normal is anomalous

Our technique is unsupervised

- Learning only requires behavior of unlabeled random sample of users

This approach has the potential to catch diverse attacker strategies

- Because we do not require any a priori knowledge of attacker strategy

Contributions

An approach to identify anomalous user behavior
- Without requiring any a priori knowledge of attacker strategy

Detect like spammers on Facebook who use diverse strategies:
- Using Sybil accounts
- Compromised accounts
- Colluding accounts

Detect fraudulent clicks in the Facebook social ad platform
- Observe that a significant fraction of clicks look anomalous

Rest of the talk

1. Methodology

2. Detecting like spammers on Facebook

3. Detecting click-spam on Facebook ads

4. Corroboration by Facebook

Learning normal patterns of behavior

For our approach to work:

We have to learn normal patterns of user behavior

If user behavior is too noisy - i.e., everyone behaves very differently

Attacker can potentially hide in the noise and evade detection

We want to see if there are a few patterns of behavior that are dominant among normal users

Why would this work against attackers?

To evade detection, attacker would have to behave normally

- Will have to limit himself to the few patterns of normal behavior
- This constrains the attacker and bounds the scale of the attack

Why would this work against attackers?

To evade detection, attacker would have to behave normally

- Will have to limit himself to the few patterns of normal behavior
- This constrains the attacker and bounds the scale of the attack

User's like behavior: Distribution of #page categories liked

Why would this work against attackers?

To evade detection, attacker would have to behave normally

- Will have to limit himself to the few patterns of normal behavior
- This constrains the attacker and bounds the scale of the attack

User's like behavior: Distribution of #page categories liked

Category	#likes
Football	2
Cricket	3
Photography	10

Normal user

Why would this work against attackers?

To evade detection, attacker would have to behave normally

- Will have to limit himself to the few patterns of normal behavior
- This constrains the attacker and bounds the scale of the attack

User's like behavior: Distribution of #page categories liked

Category	#likes
Football	2
Cricket	3
Photography	10

Normal user

Category	#likes
Body building	30
Dolls	32
Rock climbing	41
Beauty care	29
Medicine	30
Motorcycle	35
…	…

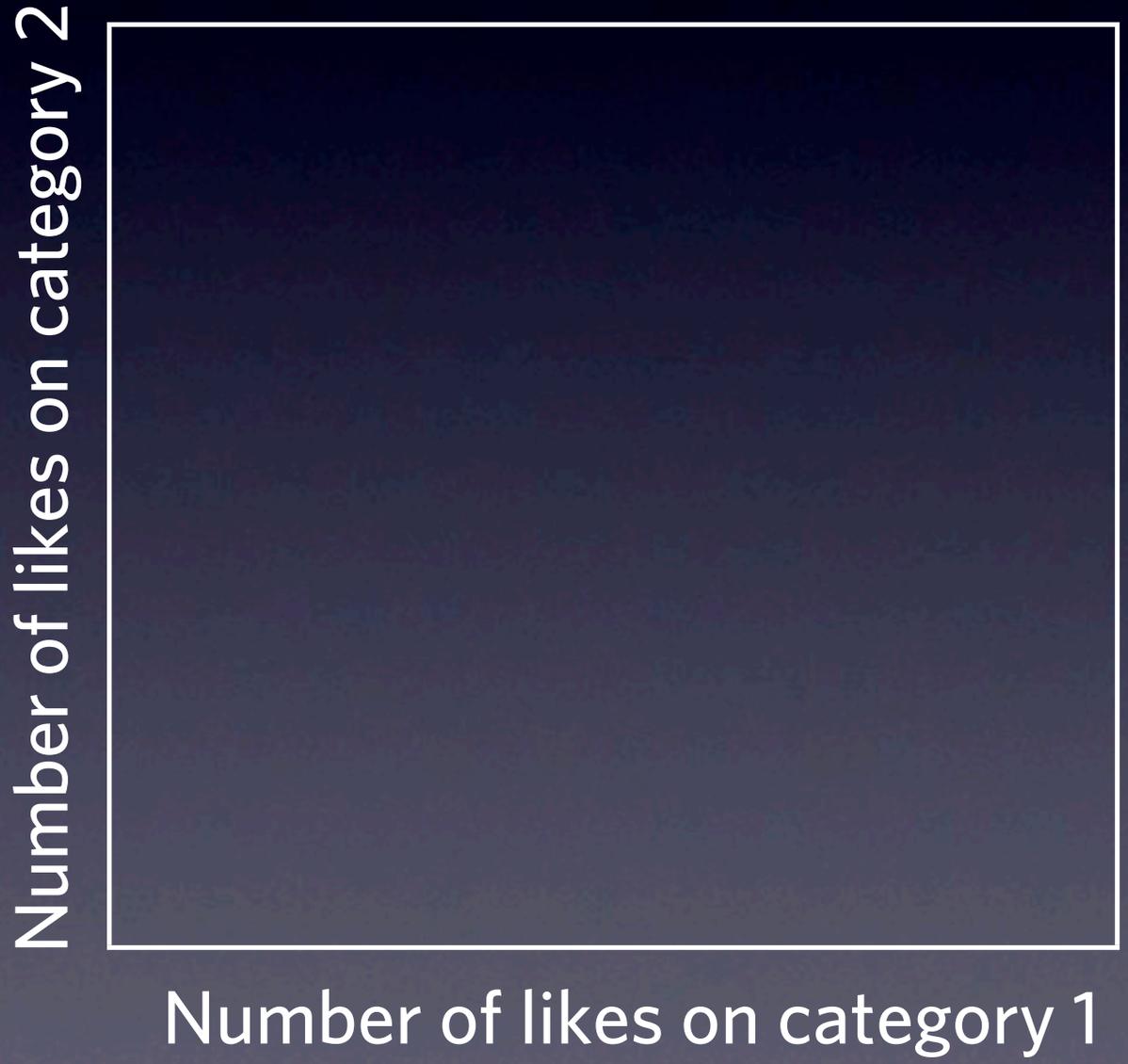
Anomalous user

Challenges in modeling behavior

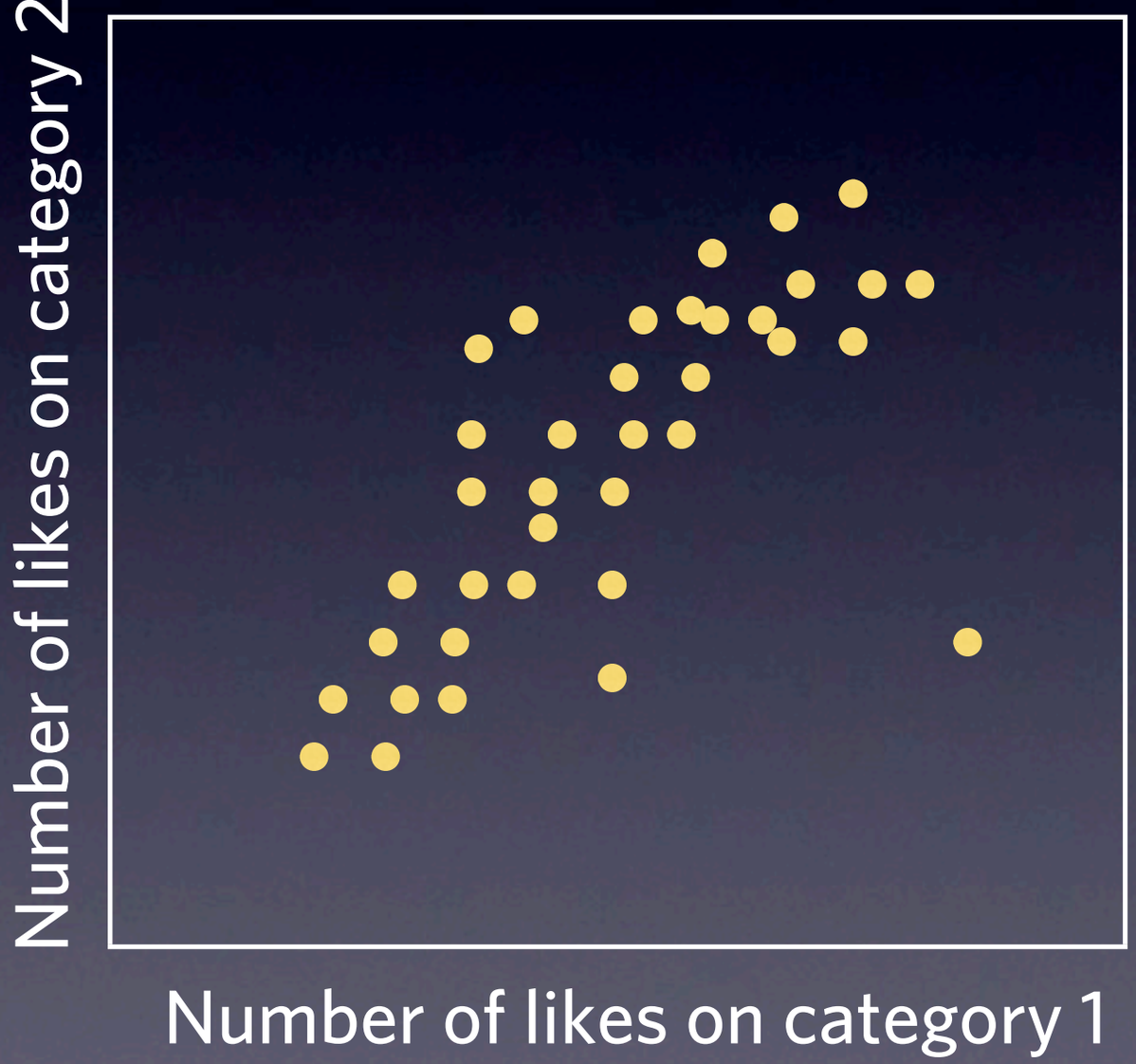
How do you model complex user behavior in social networks?

- User behavior is high dimensional
  - Number of likes for different page categories
  - Time-series of number of likes per day
- User behavior can change over time
- User behavior can be noisy

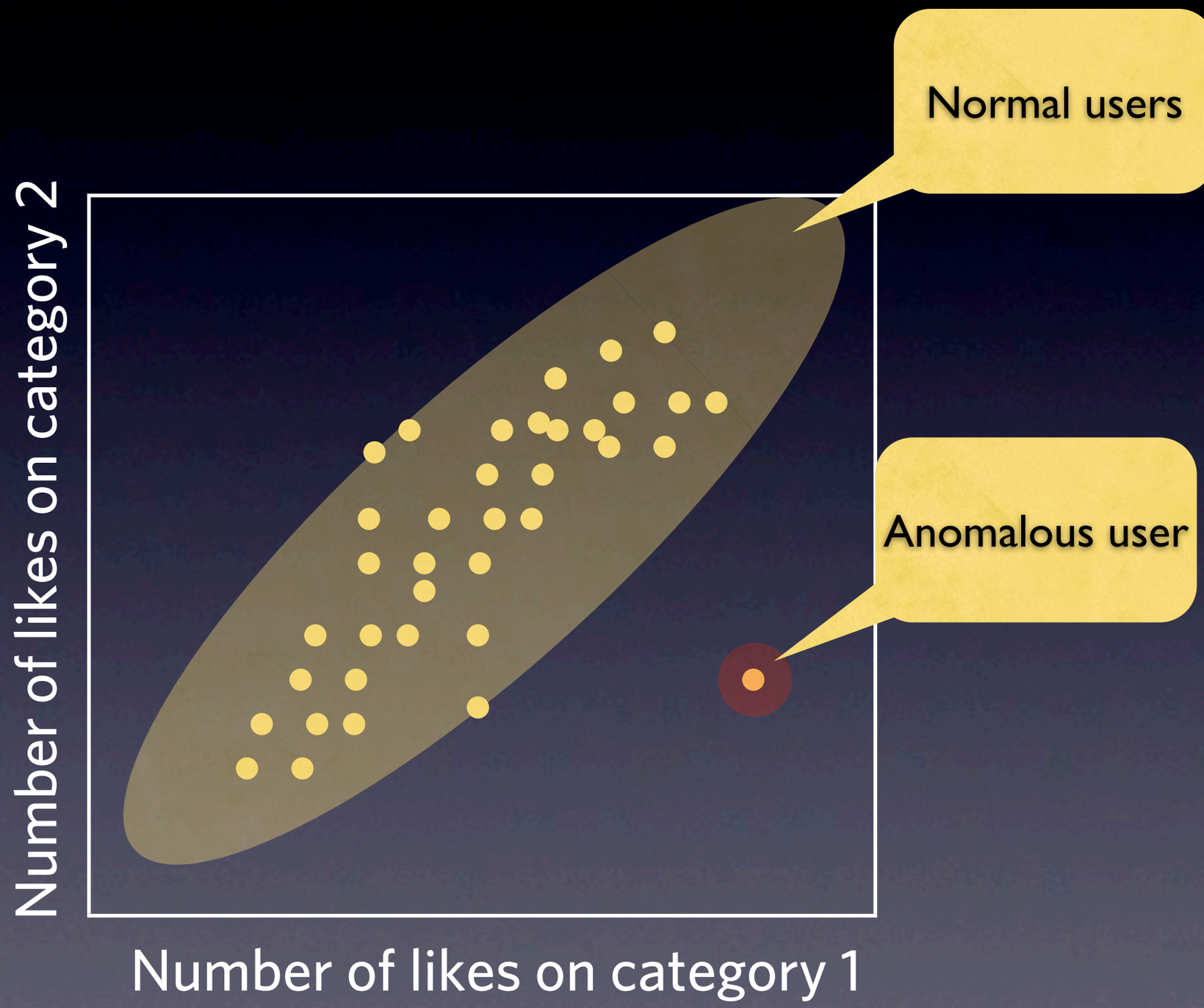
Anomaly detection using PCA



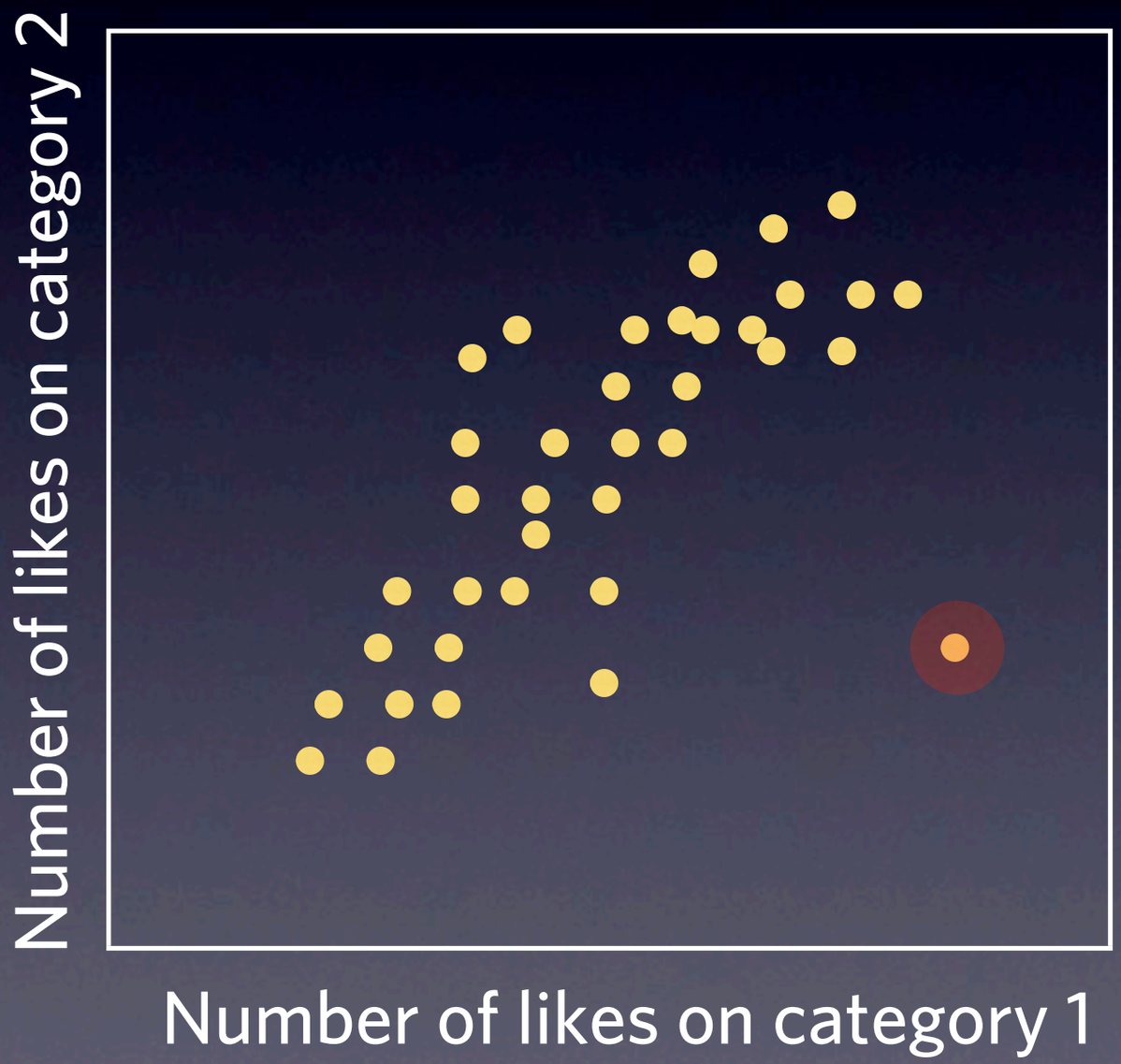
Anomaly detection using PCA



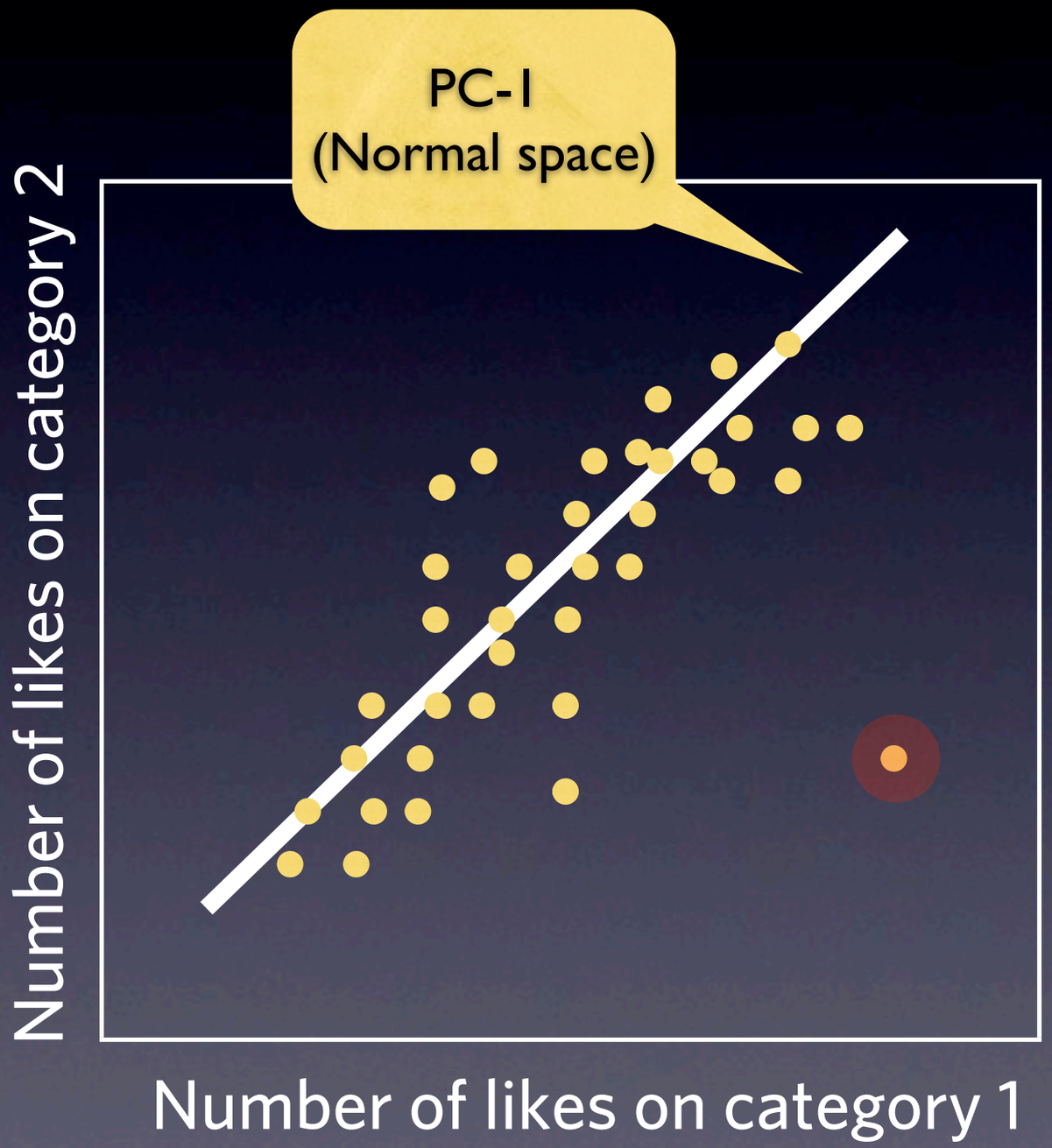
Anomaly detection using PCA



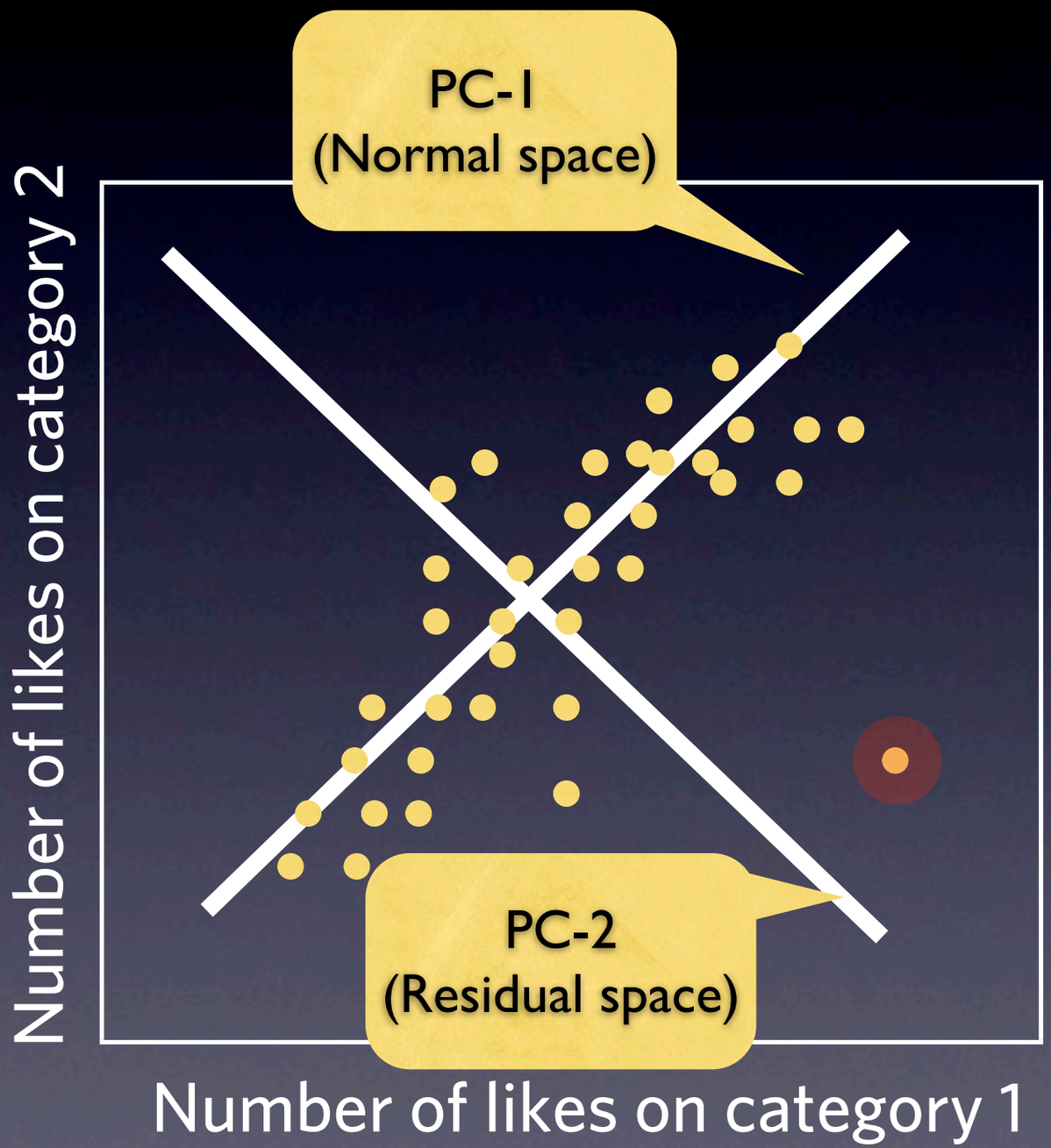
Anomaly detection using PCA



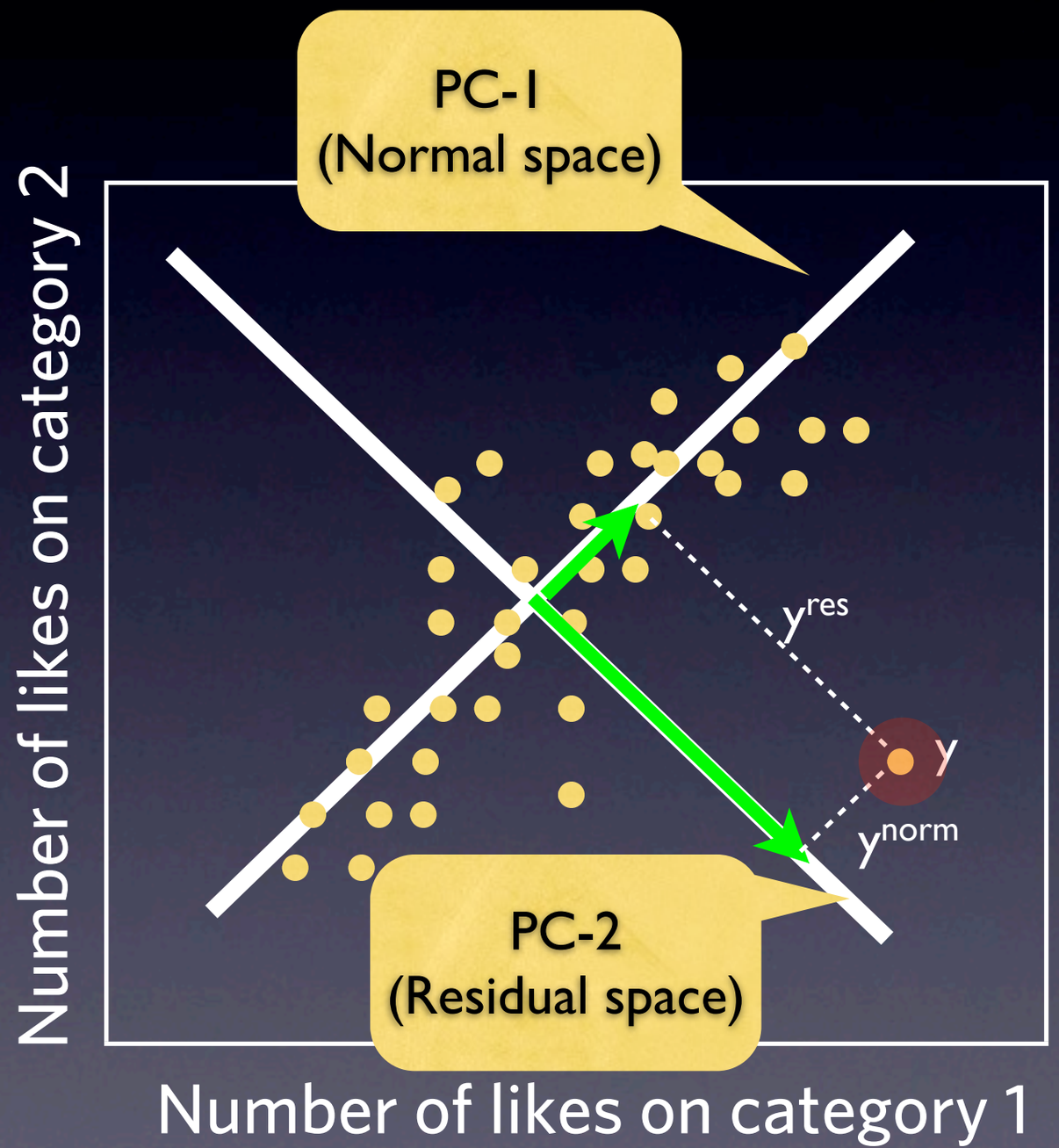
Anomaly detection using PCA



Anomaly detection using PCA



Anomaly detection using PCA



If y^{res} is unusually high, user is anomalous

Capturing normal behavior patterns

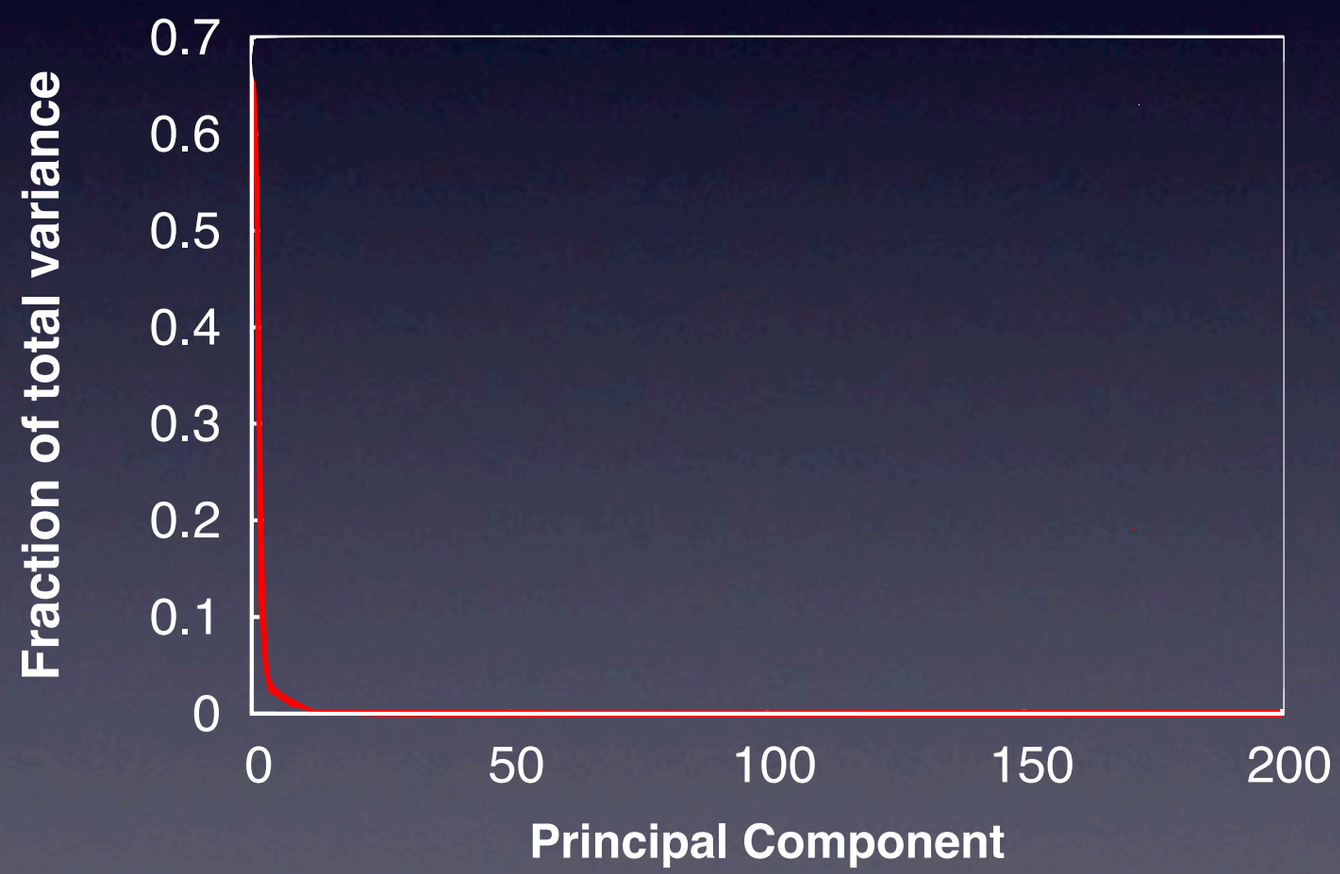
Are there a few patterns of behavior that are dominant?

Can be answered by looking at variance captured by each PC

Capturing normal behavior patterns

Are there a few patterns of behavior that are dominant?

Can be answered by looking at variance captured by each PC

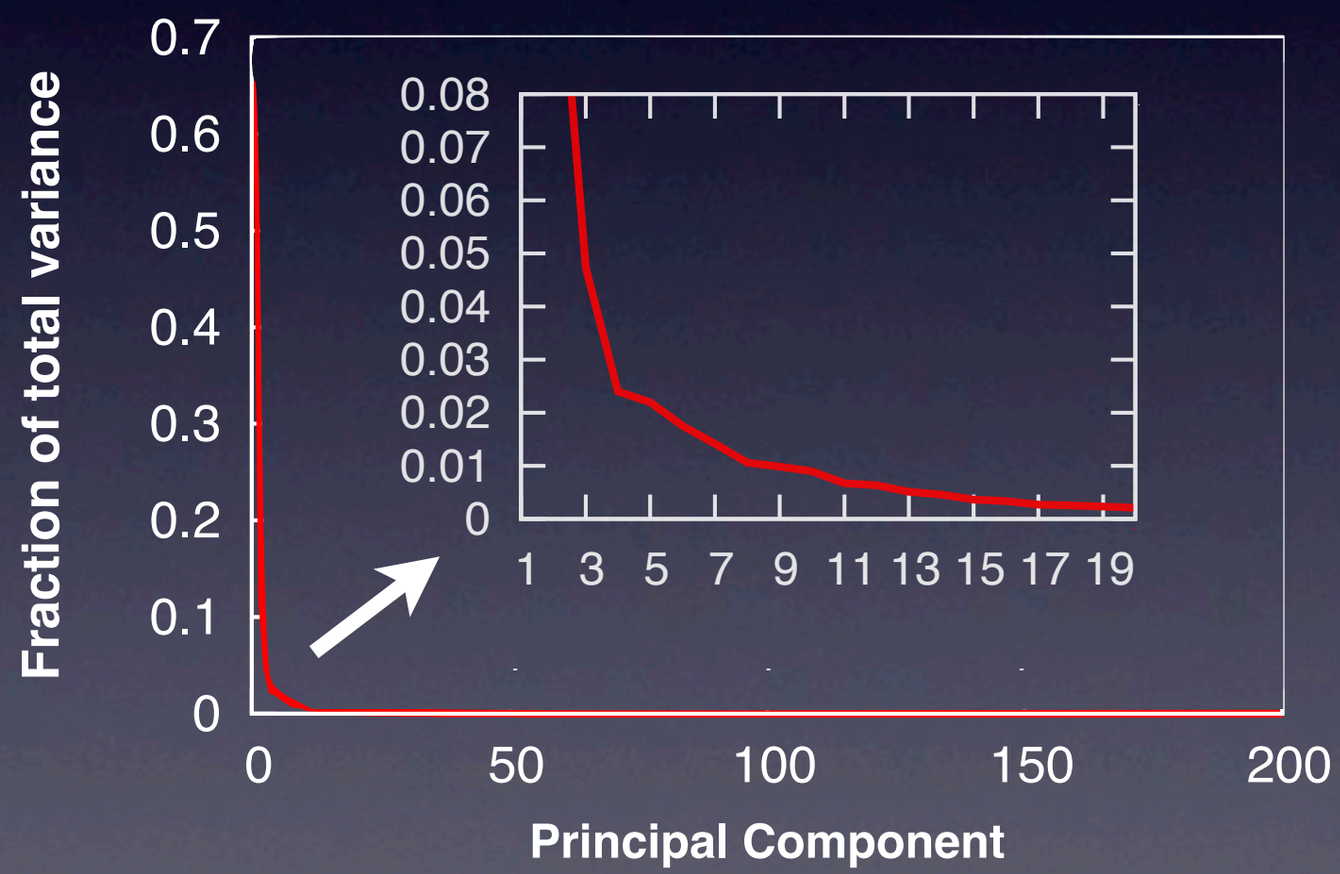


Facebook like behavior defined over 224 page categories

Capturing normal behavior patterns

Are there a few patterns of behavior that are dominant?

Can be answered by looking at variance captured by each PC

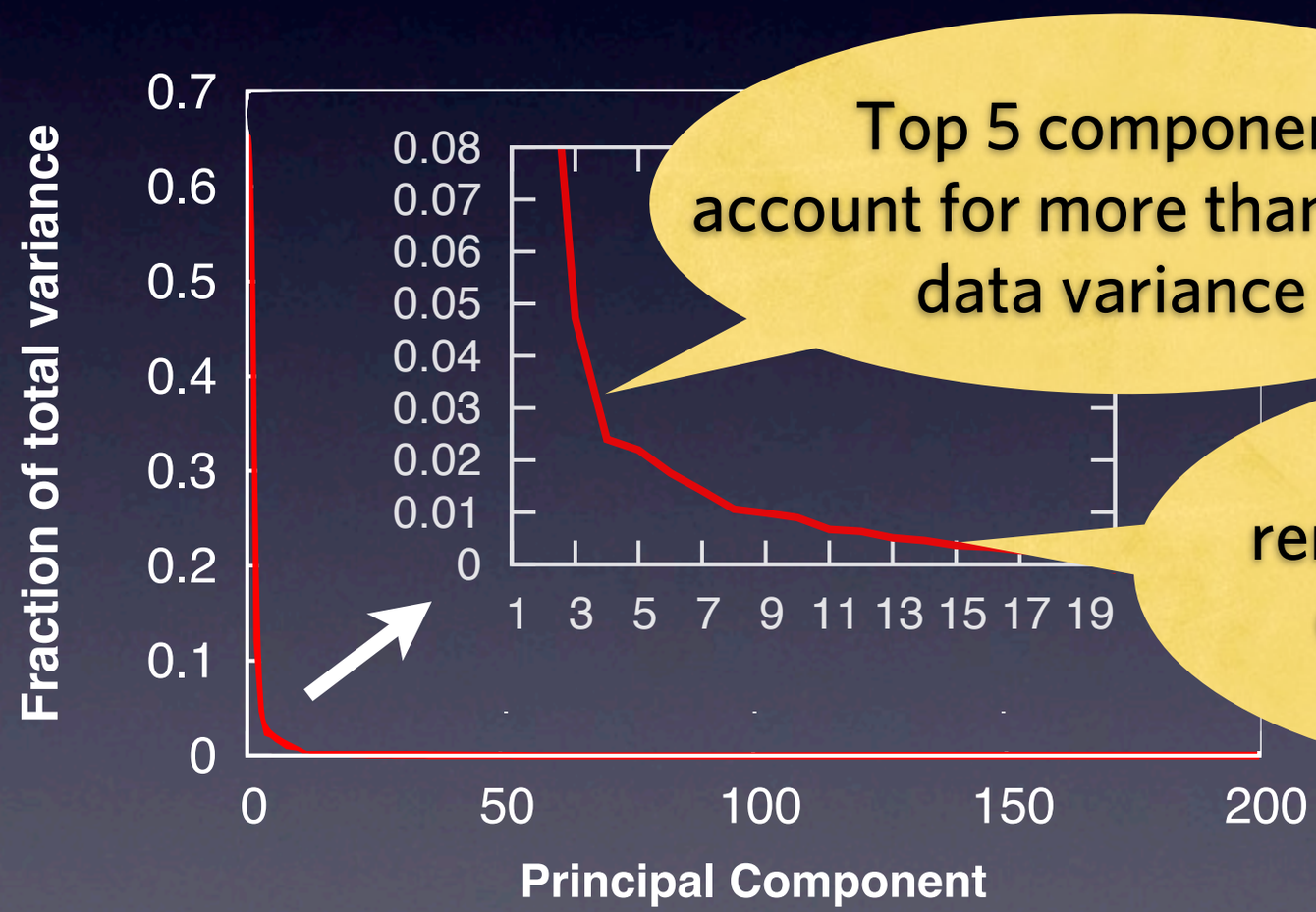


Facebook like behavior defined over 224 page categories

Capturing normal behavior patterns

Are there a few patterns of behavior that are dominant?

Can be answered by looking at variance captured by each PC



Top 5 components account for more than 85% of data variance

book like

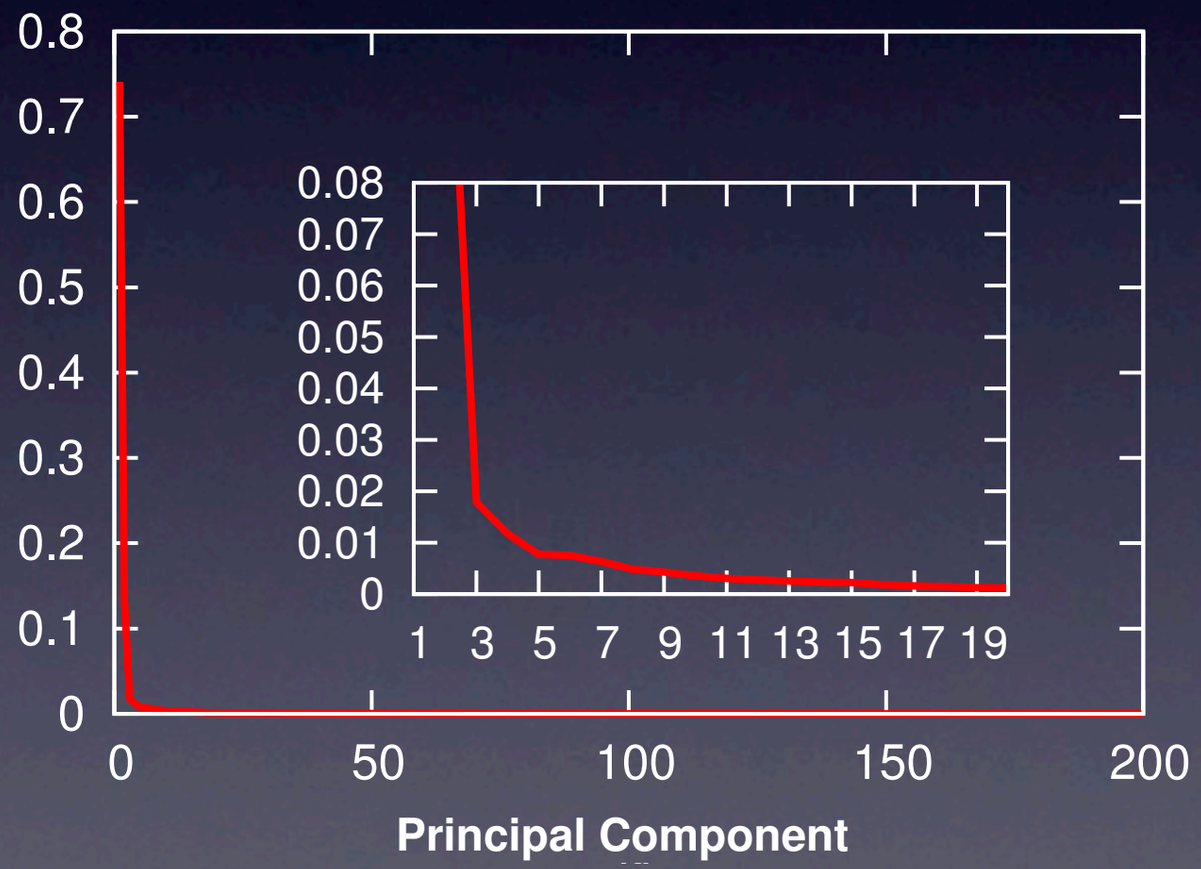
behavior defined

Each of the remaining components capture very small variance

Capturing normal behavior patterns

Are there a few patterns of behavior that are dominant?

Can be answered by looking at variance captured by each PC

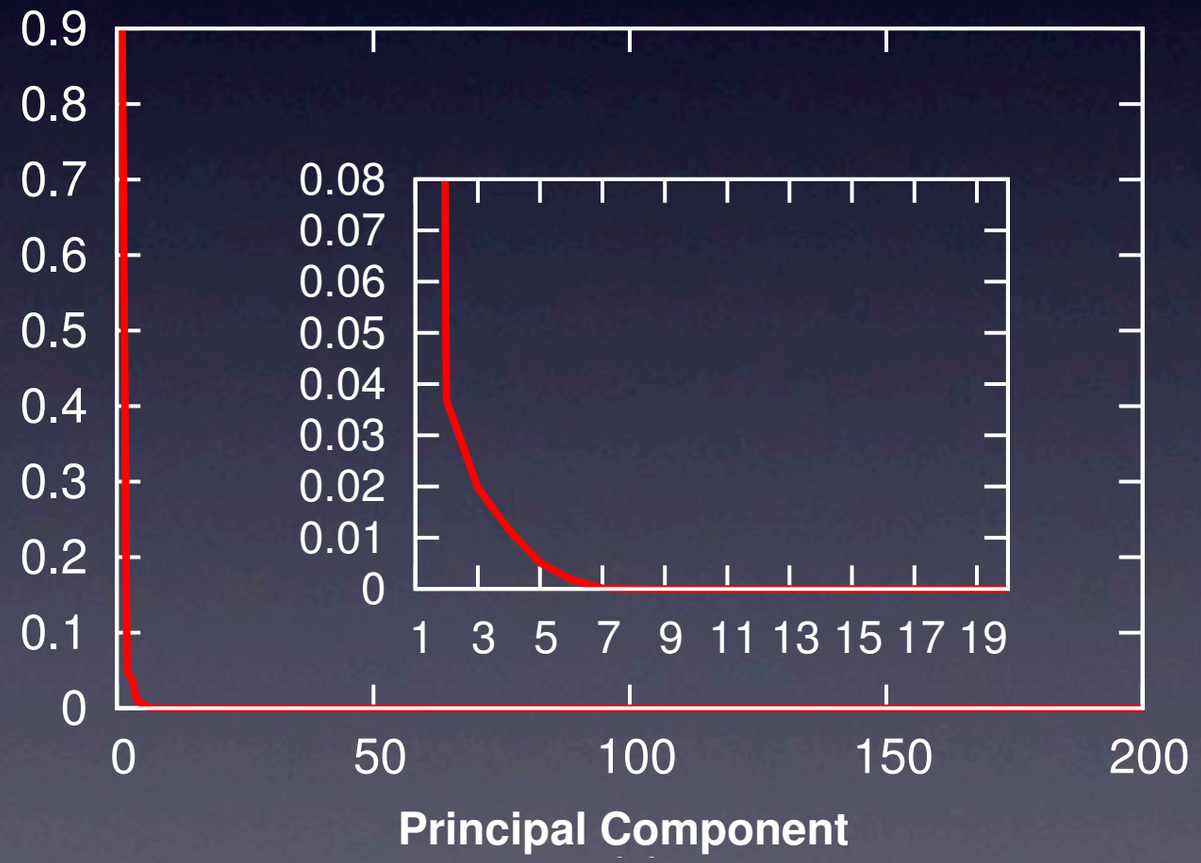


Twitter user behavior

Capturing normal behavior patterns

Are there a few patterns of behavior that are dominant?

Can be answered by looking at variance captured by each PC



Yelp user behavior

Rest of the talk

1. Methodology

2. Detecting like spammers on Facebook

3. Detecting click-spam on Facebook ads

4. Corroboration by Facebook

Data collected

Training data:

Random users: 12k random users sampled from Facebook

Testing data:

Identity type	#Users
Black-market	3.2k
Compromised	1k
Colluding	900
Normal	1.2k

Detected anomalous behavior

Identity type	Likes flagged
Black-market	99%
Compromised	64%
Colluding	92%

When tested on normal users, we flag 3.3% of them (false positives)

Rest of the talk

1. Methodology

2. Detecting like spammers on Facebook

3. Detecting click-spam on Facebook ads

4. Corroboration by Facebook

Click-spam on Facebook

Advertisers lose money on spam clicks

- They might lose confidence in the advertising platform
- Affects the sustainability of the social networking service

Preliminary experiment to understand click-spam in Facebook ads

- Set up a real ad and a bluff ad targeting users in USA

Click-spam on Facebook

Advertisers lose money on spam clicks

- They might lose confidence in the advertising platform
- Affects the sustainability of the social networking service

Preliminary experiment to understand click-spam in Facebook ads

- Set up a real ad and a bluff ad targeting users in USA



Real ad

Click-spam on Facebook

Advertisers lose money on spam clicks

- They might lose confidence in the advertising platform
- Affects the sustainability of the social networking service

Preliminary experiment to understand click-spam in Facebook ads

Set up a real ad and a bluff ad targeting users in USA



Real ad



Bluff ad

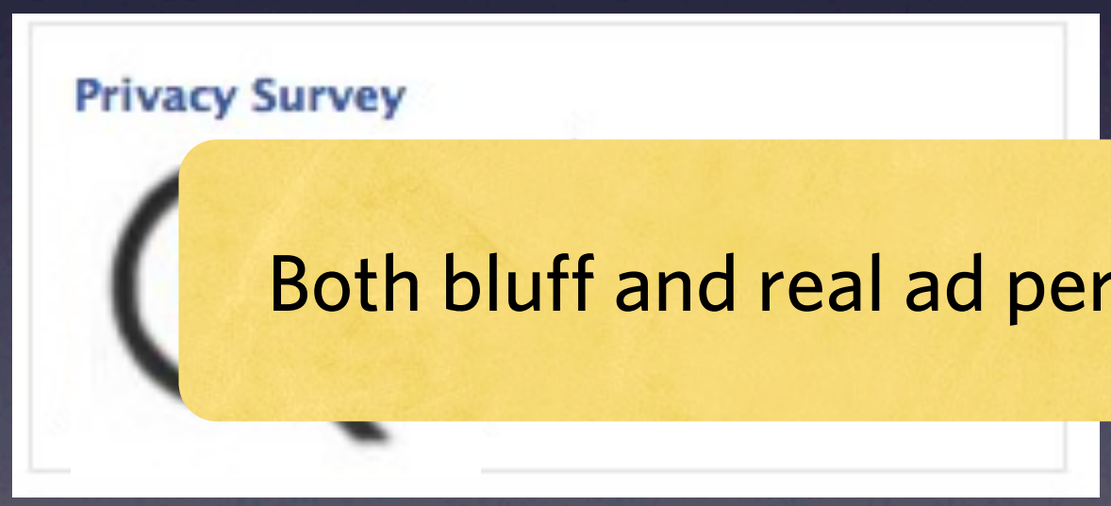
Click-spam on Facebook

Advertisers lose money on spam clicks

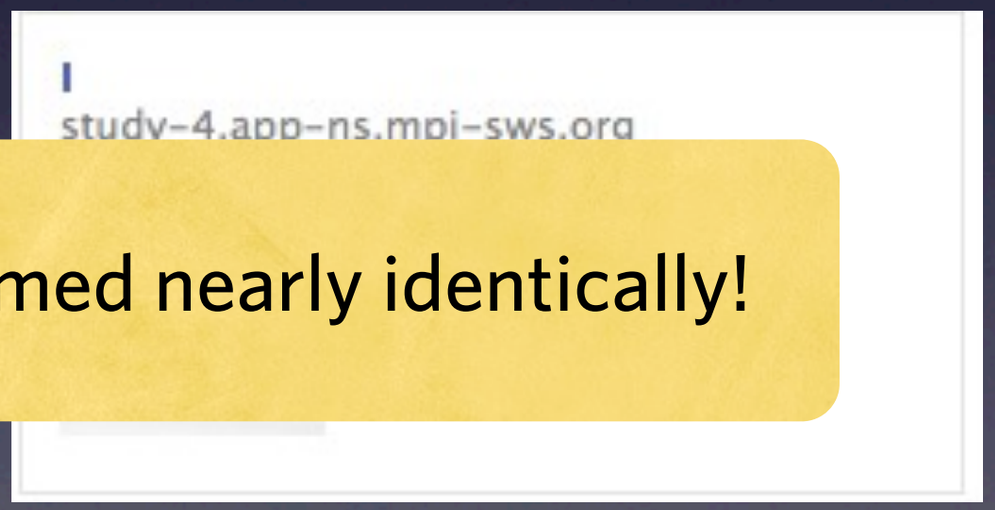
- They might lose confidence in the advertising platform
- Affects the sustainability of the social networking service

Preliminary experiment to understand click-spam in Facebook ads

- Set up a real ad and a bluff ad targeting users in USA



Real ad



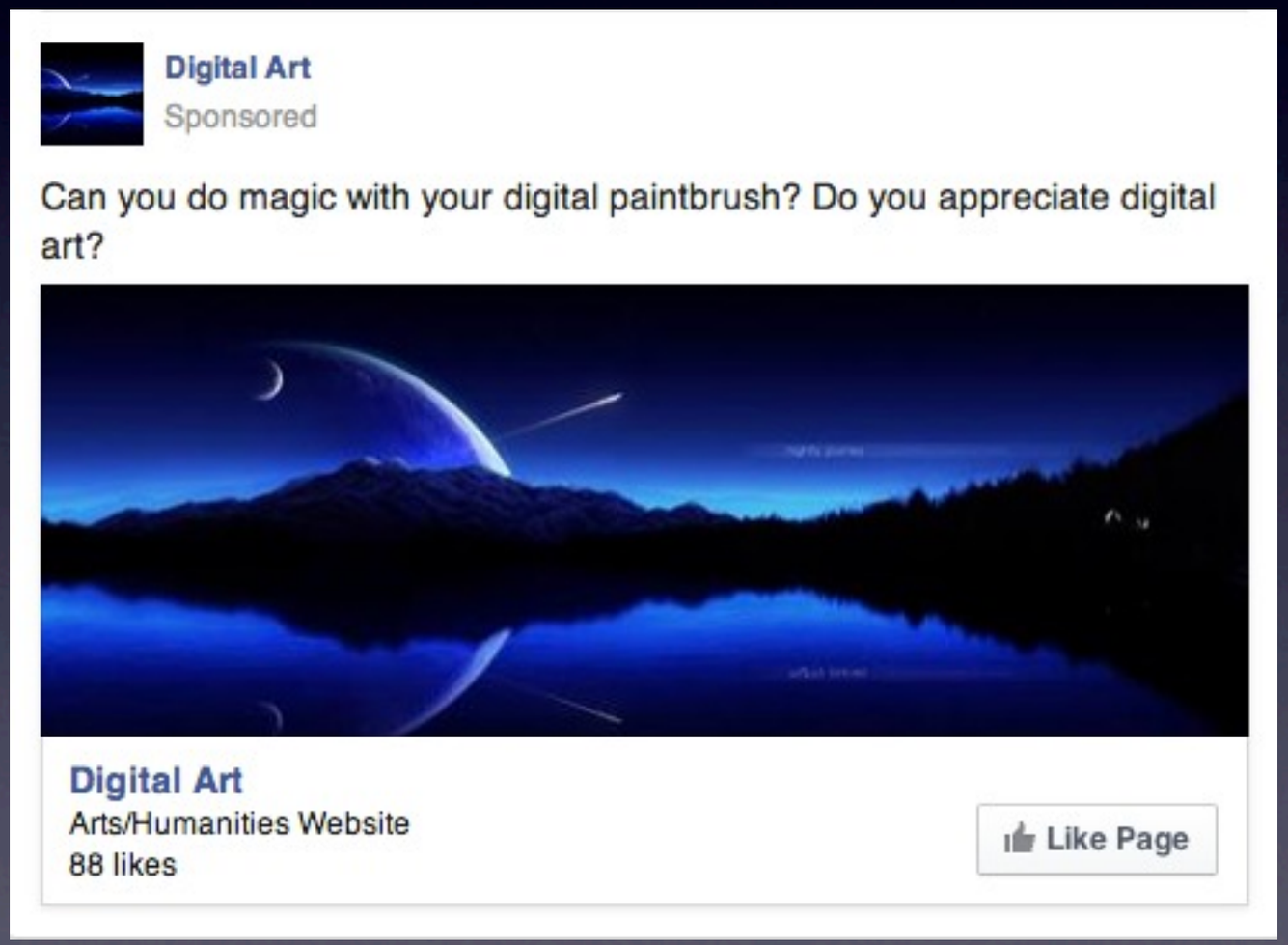
Bluff ad

Both bluff and real ad performed nearly identically!

Experiment to detect click-spam

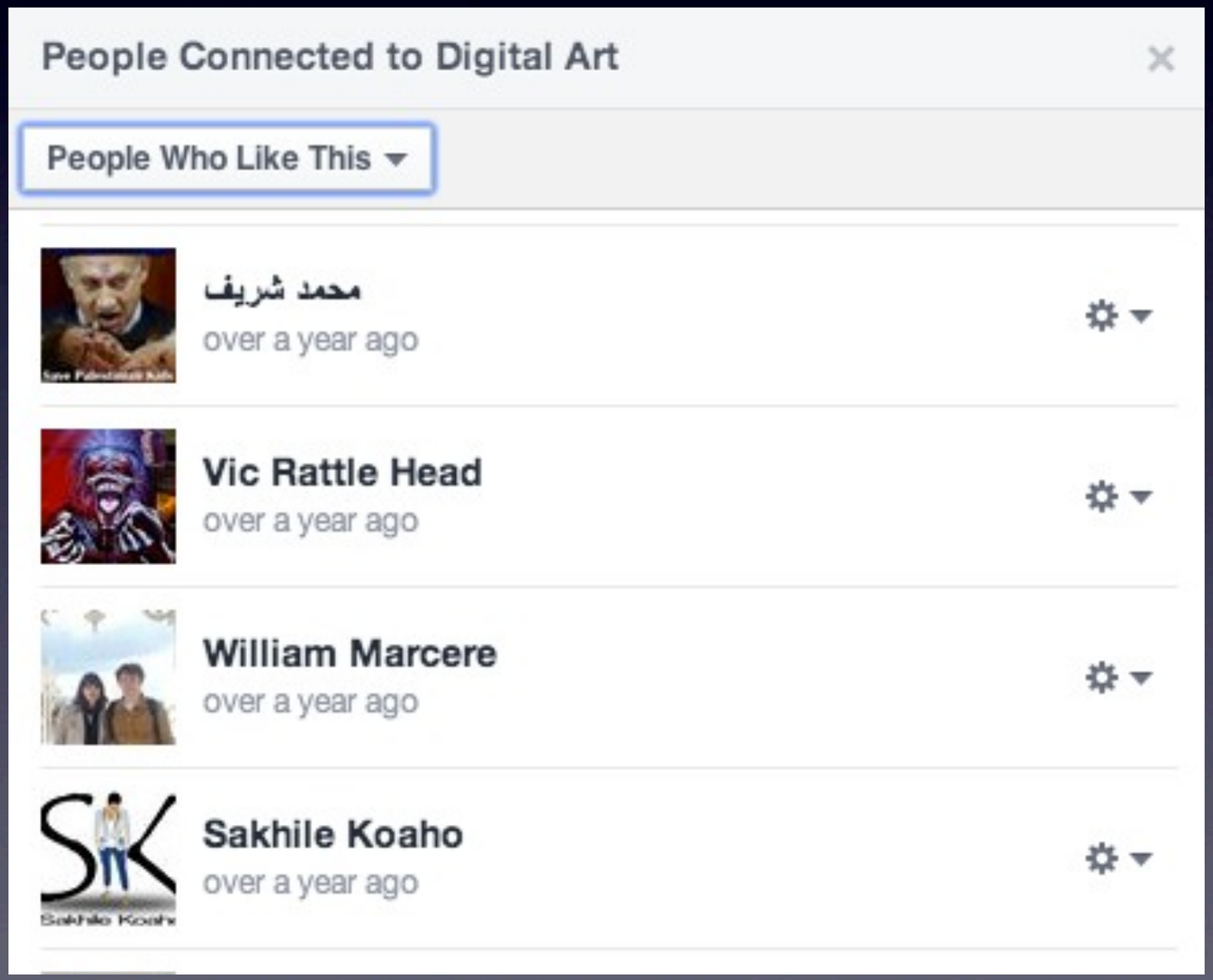
Step 1. Create ad to get likes to our Facebook page

Facebook then targets users who are more likely to like the ad/page



Experiment to detect click-spam

Step 2. Apply anomaly classifier to users who clicked (liked) on the ad



Experiment to detect click-spam

We set up 10 such ad campaigns targeting 7 countries

USA, UK, Australia, Egypt, Philippines, Malaysia, India

Click-spam identified

1,867/2,767 (67%) users who click on ads look anomalous

8 out of 10 campaigns have a majority of clicks that look anomalous

US,UK campaigns have more than 39% anomalous clicks

Rest of the talk

1. Methodology

2. Detecting like spammers on Facebook

3. Detecting click-spam on Facebook ads

4. Corroboration by Facebook

Corroboration by Facebook

We analyzed the state of flagged users and their likes in June 2014

Users:

- Most of the flagged users still exist
- 92% of black-market and 93% of ad users are still alive

Likes:

- Confirms click-spam findings (where there was no ground-truth)
- More than 85% of all likes by ad users were removed after 4 months

- But Facebook's system is still behind on removing a lot of misbehavior
- Over 48% of likes by black-market users still exist after 10 months

Conclusion

Service abuse is a huge problem in social networks today

- Attackers use diverse strategies and also tend to adapt

We propose an unsupervised anomaly detection scheme

- PCA serves as a nice tool to model behavior and detect anomalous ones

We evaluate our technique on extensive ground-truth data of anomalous behavior

We apply our approach to detect click-spam in a social ad platform