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

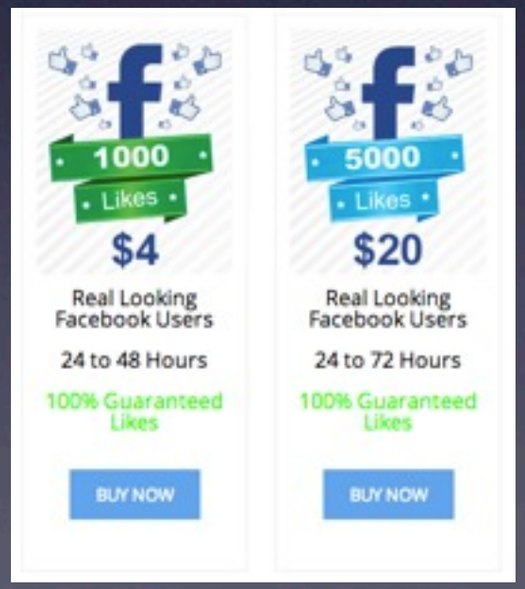
Service abuse in social networks today

Several black-market services are available today to

- Manipulate content ratings
- Manipulate influence/popularity of a user

One can buy likes for their Facebook page on the black-market

Quality of traffic on social ad platforms is also questionable

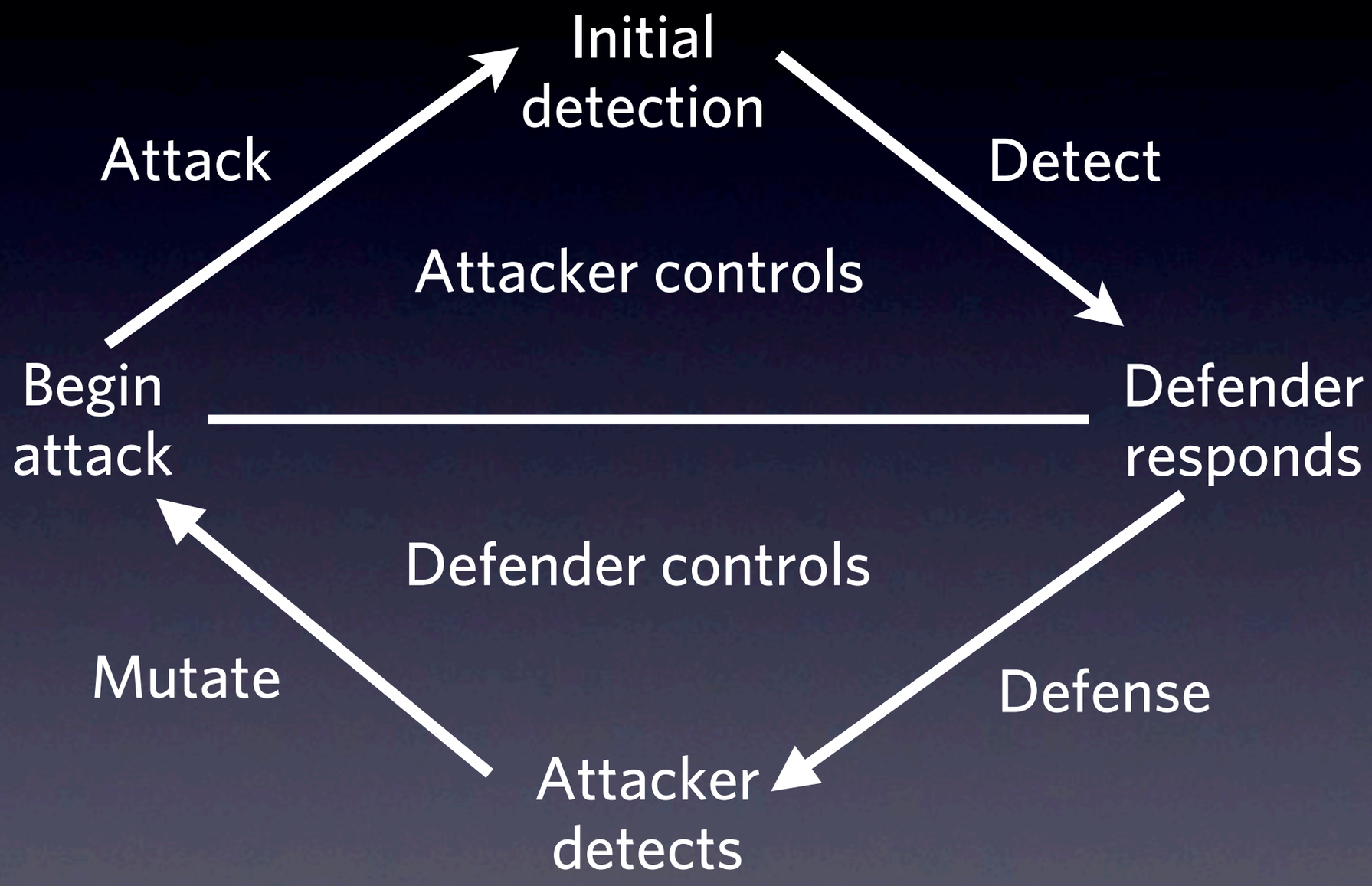


Our goal

Detect misbehaving identities in the service

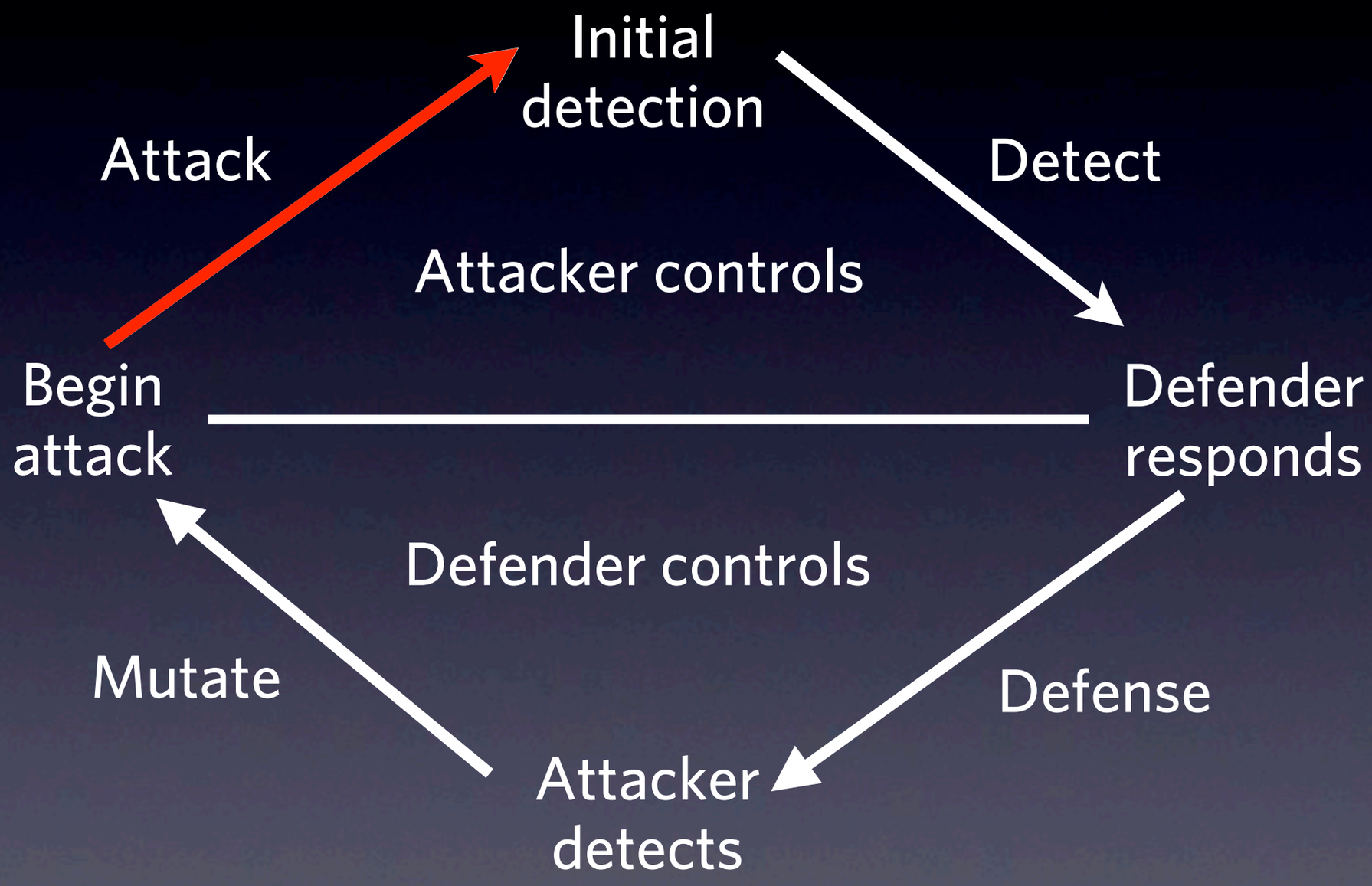
- Suspend the misbehaving user or nullify their actions

Existing approaches for defense



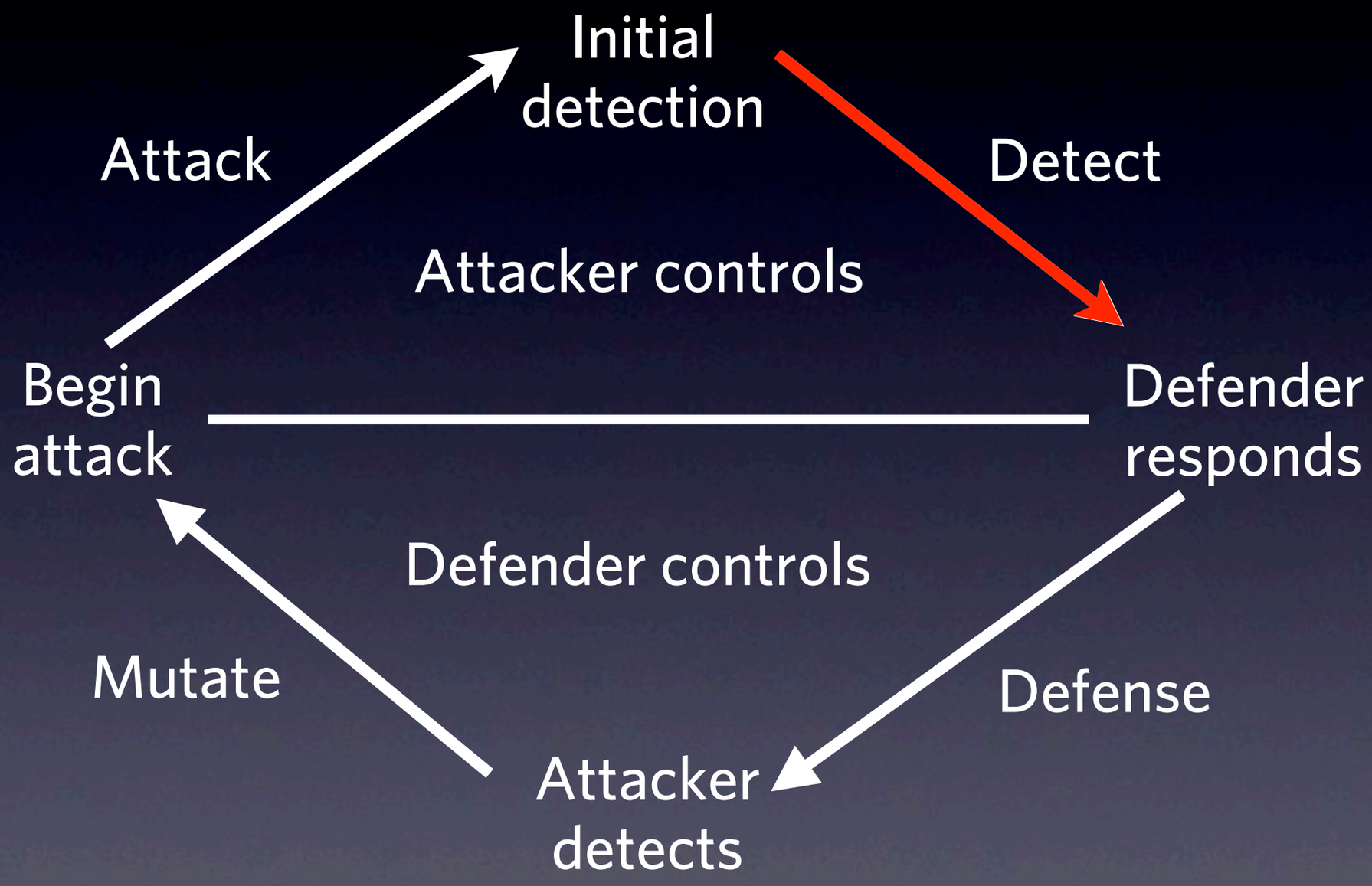
"Facebook Immune System", SNS'10

Existing approaches for defense



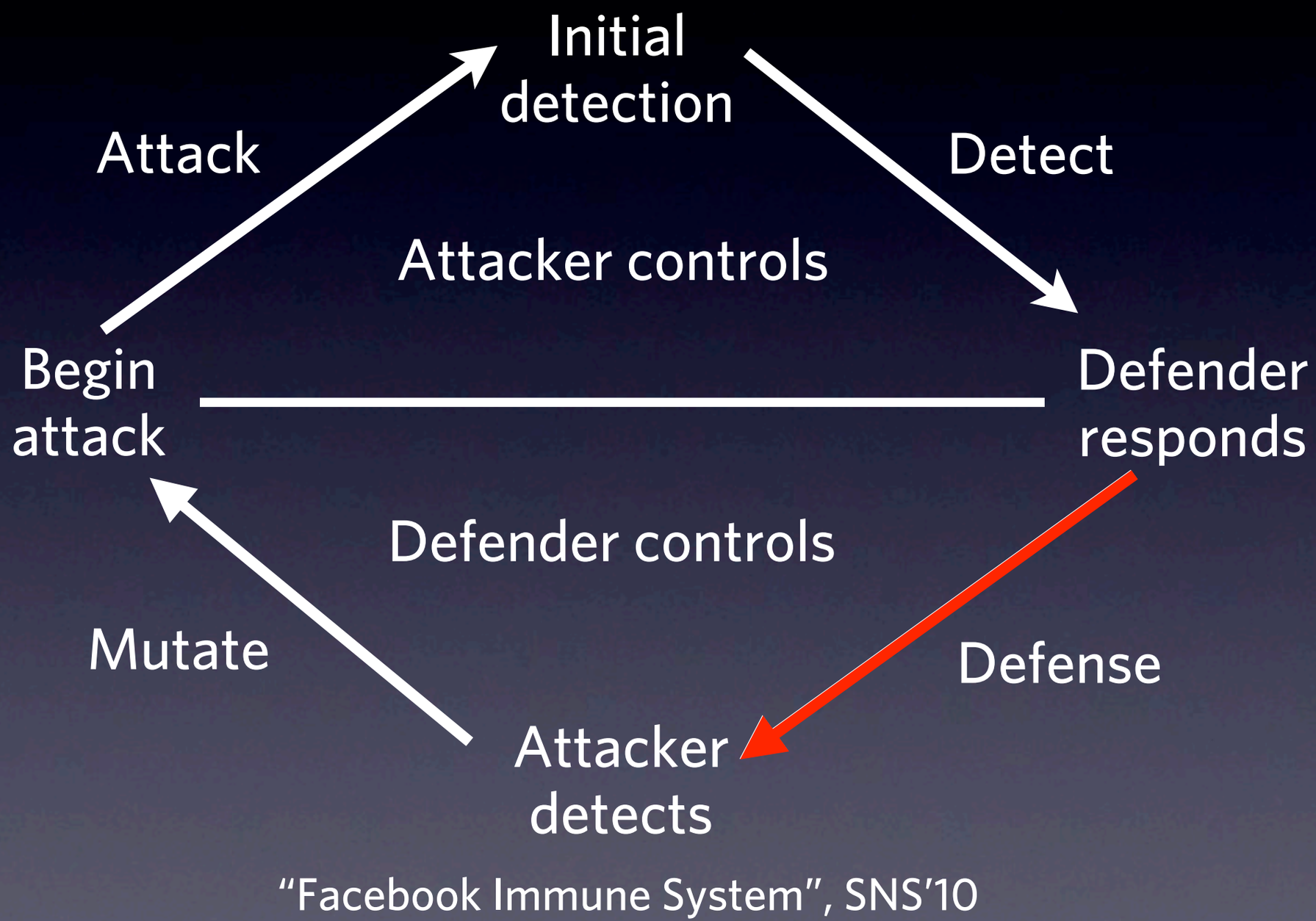
"Facebook Immune System", SNS'10

Existing approaches for defense



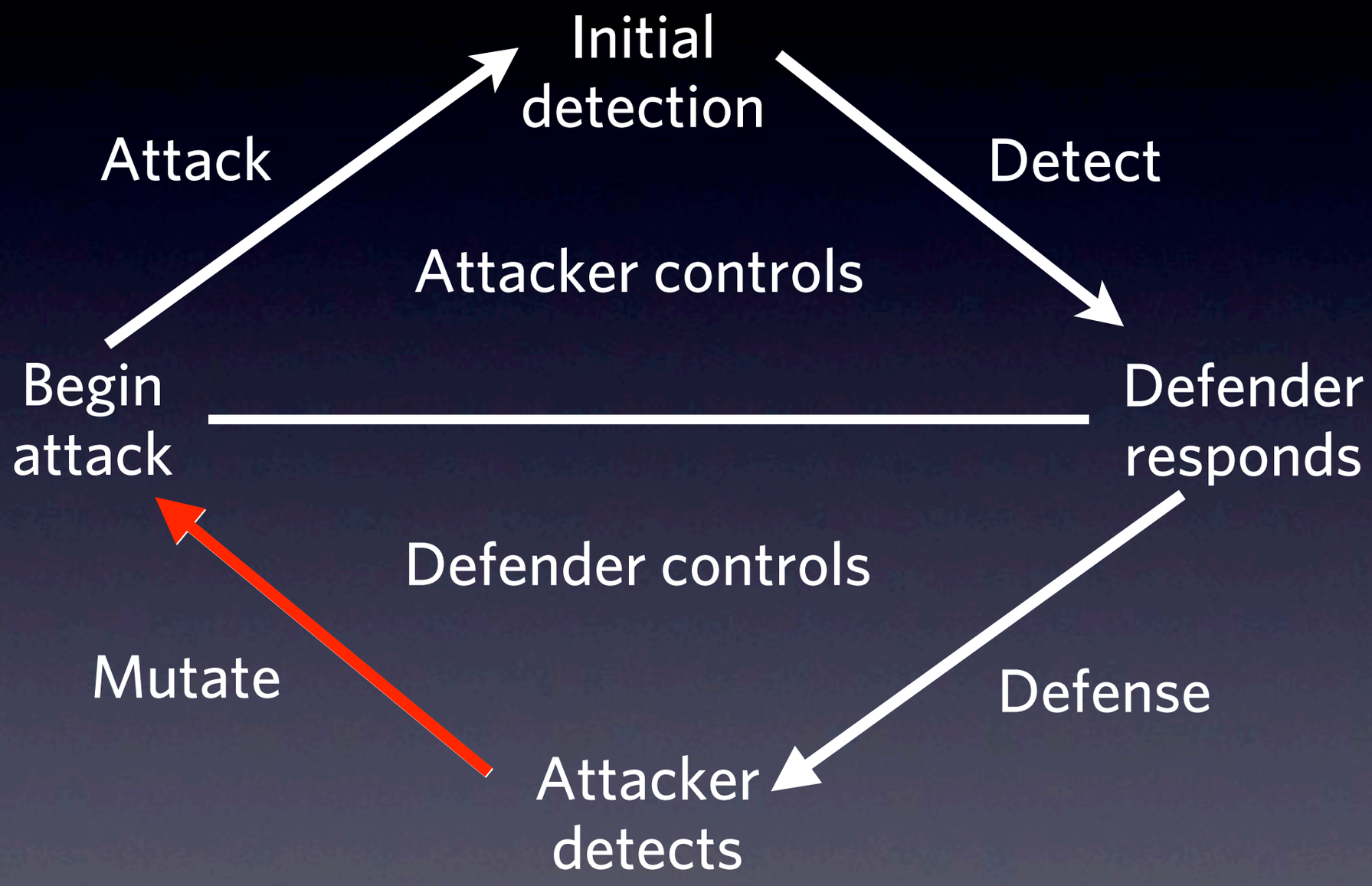
"Facebook Immune System", SNS'10

Existing approaches for defense



"Facebook Immune System", SNS'10

Existing approaches for defense



"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 unsupervised anomaly detection

Our approach at a high level

We use an unsupervised anomaly detection technique

We build an Anomaly classifier

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

For learning phase:

- Input only includes behavior of unlabeled random sample of users

This approach has the potential to catch diverse attack strategies

Key contributions

An approach to identify anomalous user behavior

Detect Like spammers on Facebook

- Our approach detects diverse attack 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

Methodology

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

Spatial feature: Distribution of #page categories liked

Football	2
Cricket	3
Photography	10

Normal user

Body building	30
Dolls	32
Rock climbing	41
Beauty care	29
Medicine	30
Motorcycle	35
Cartoons	51

Anomalous user

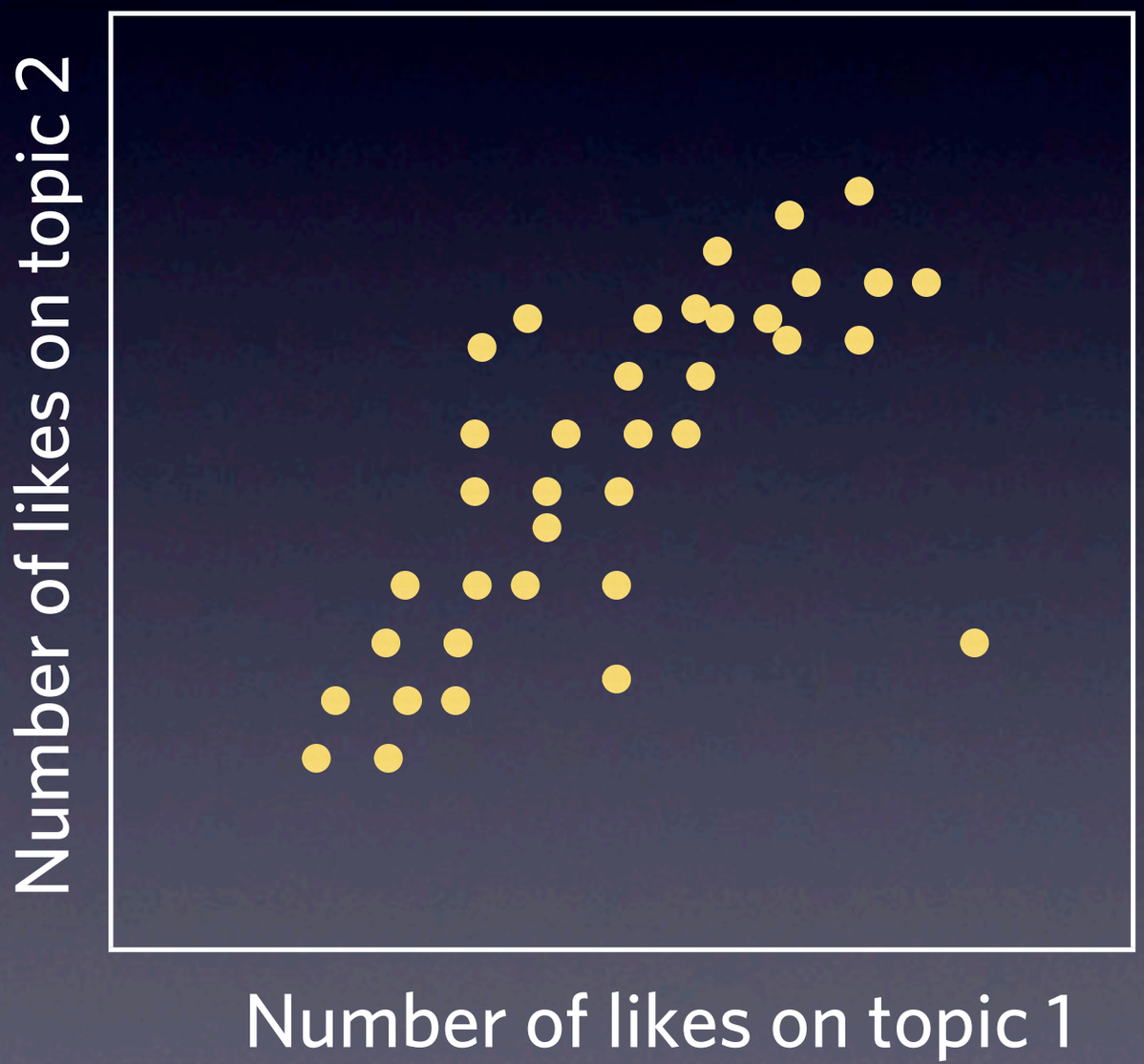
Challenges in modeling behavior

How do you model complex user behavior in social networks?

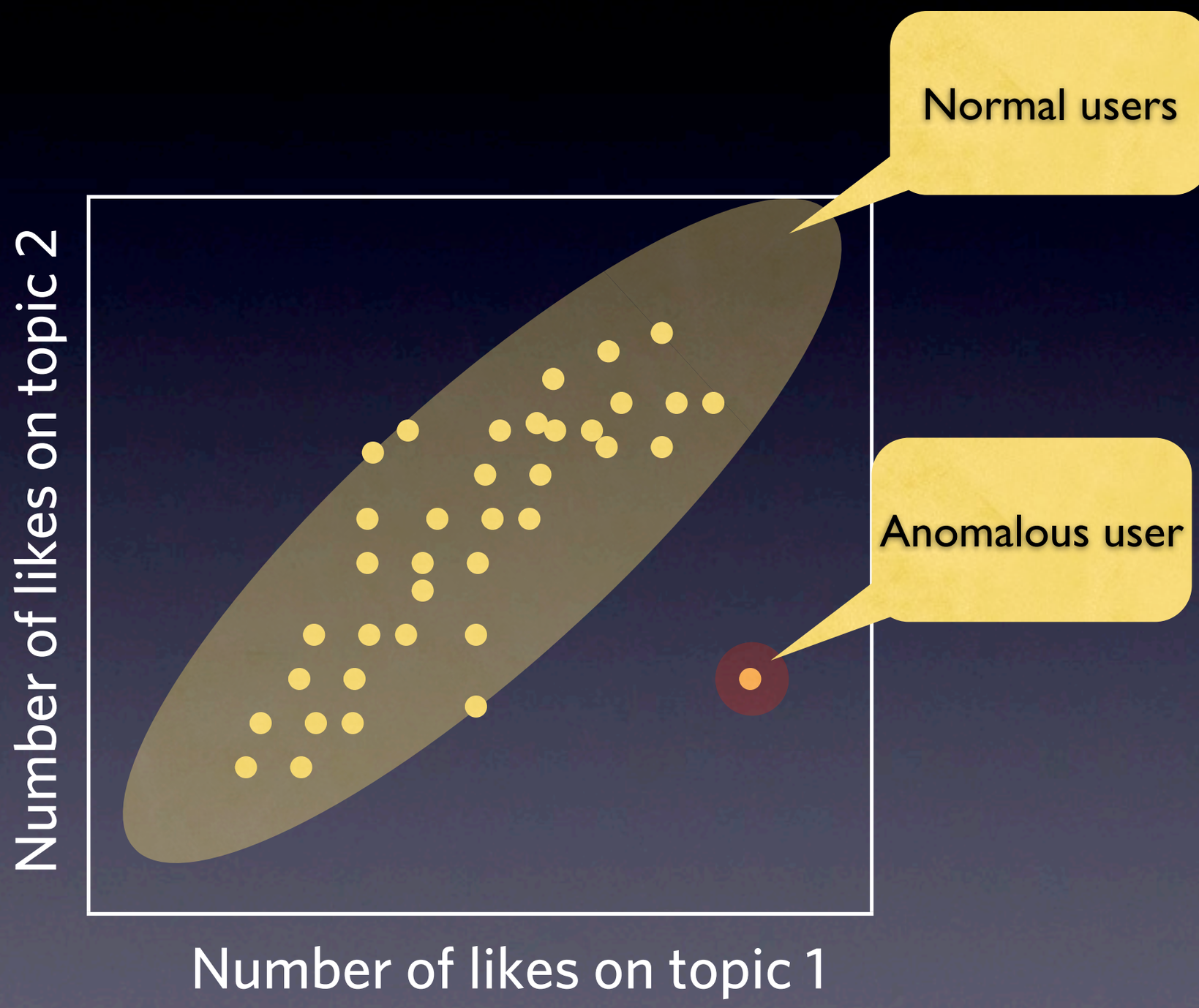
- User behavior is high dimensional
  - Spatial feature: Behavior defined as distribution of topic categories
  - Temporal feature: Time-series of number of likes per day
- User behavior can change over time
- User behavior can be noisy

Anomaly detection using PCA

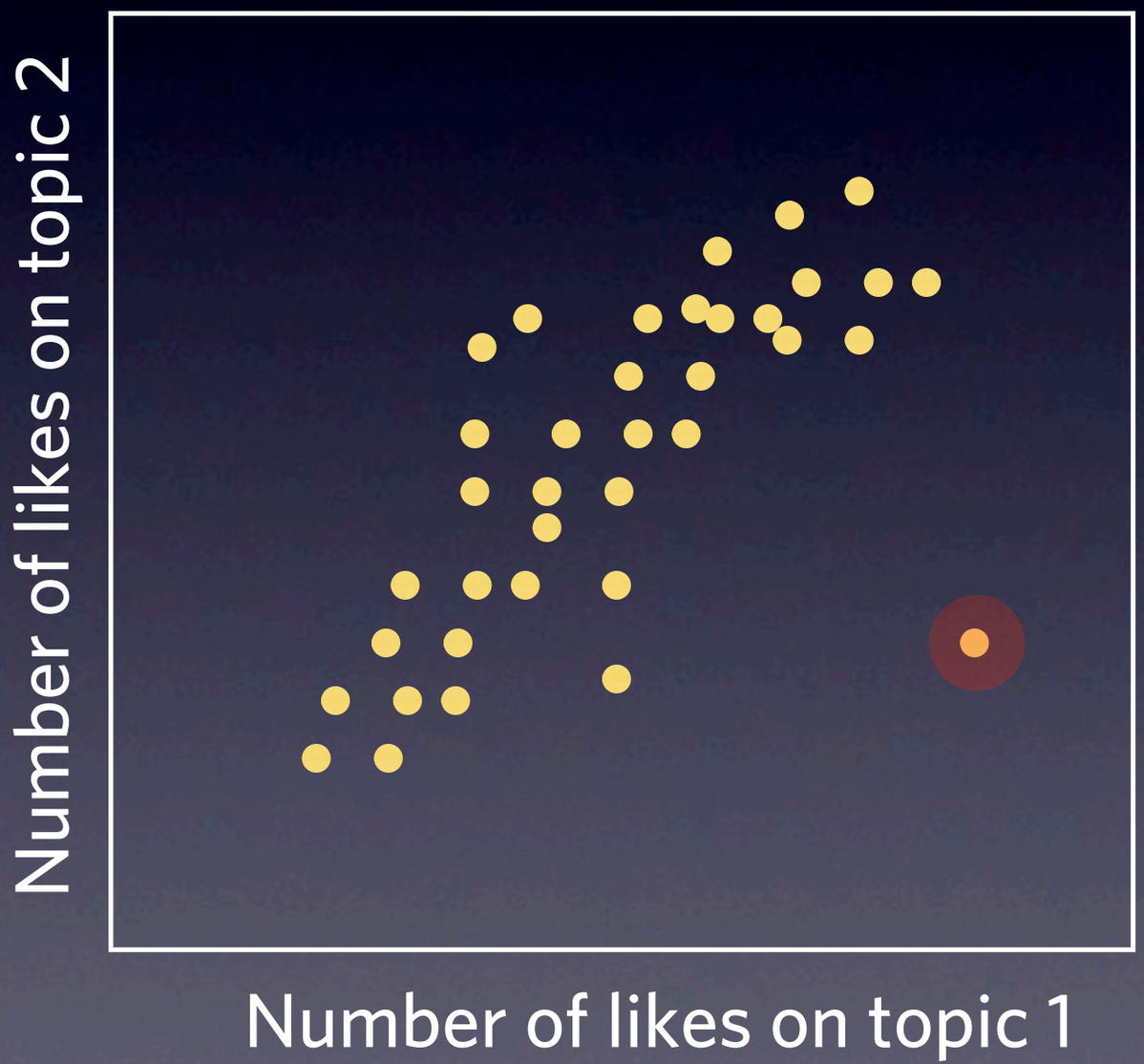
Anomaly detection using PCA



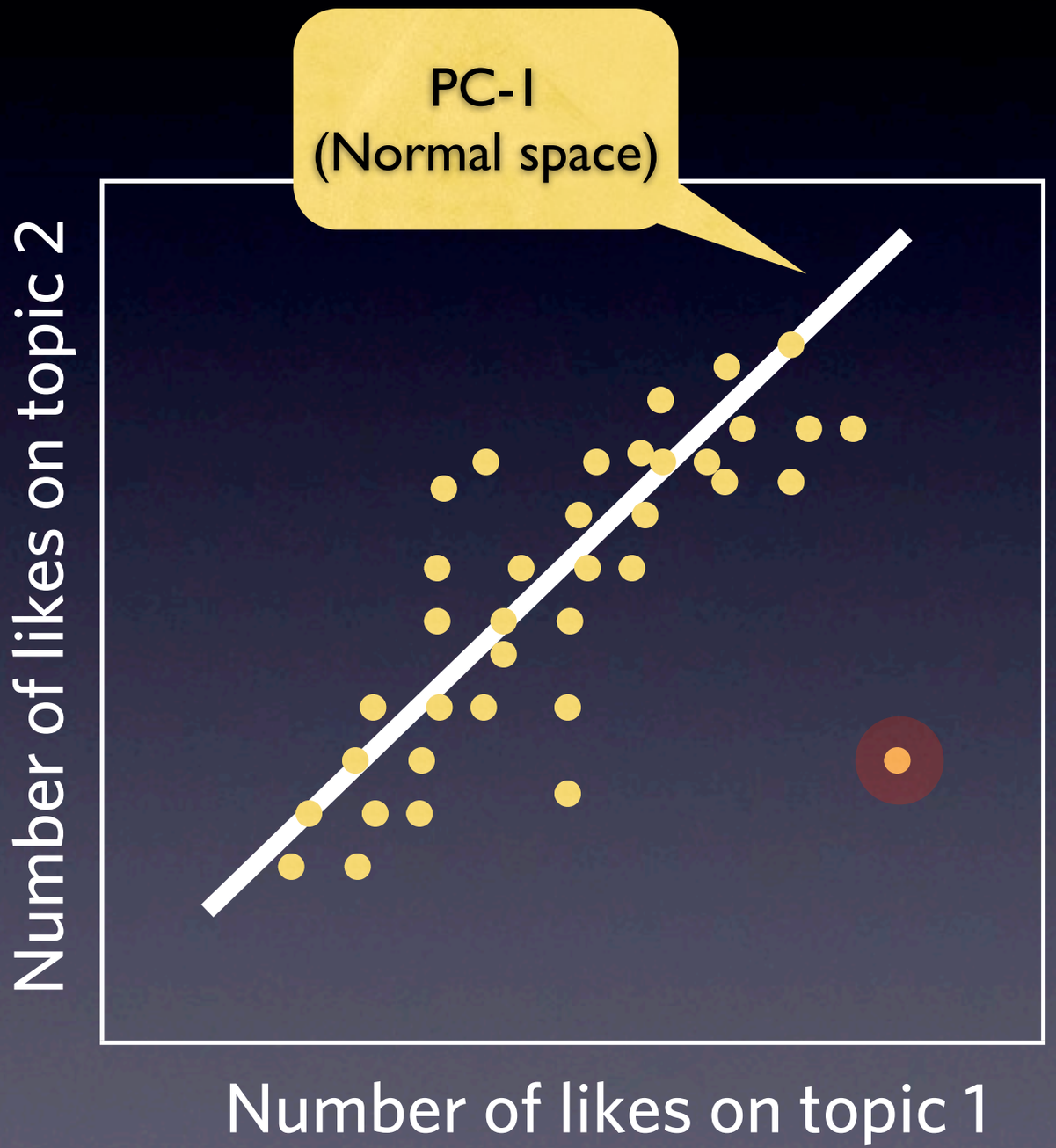
Anomaly detection using PCA



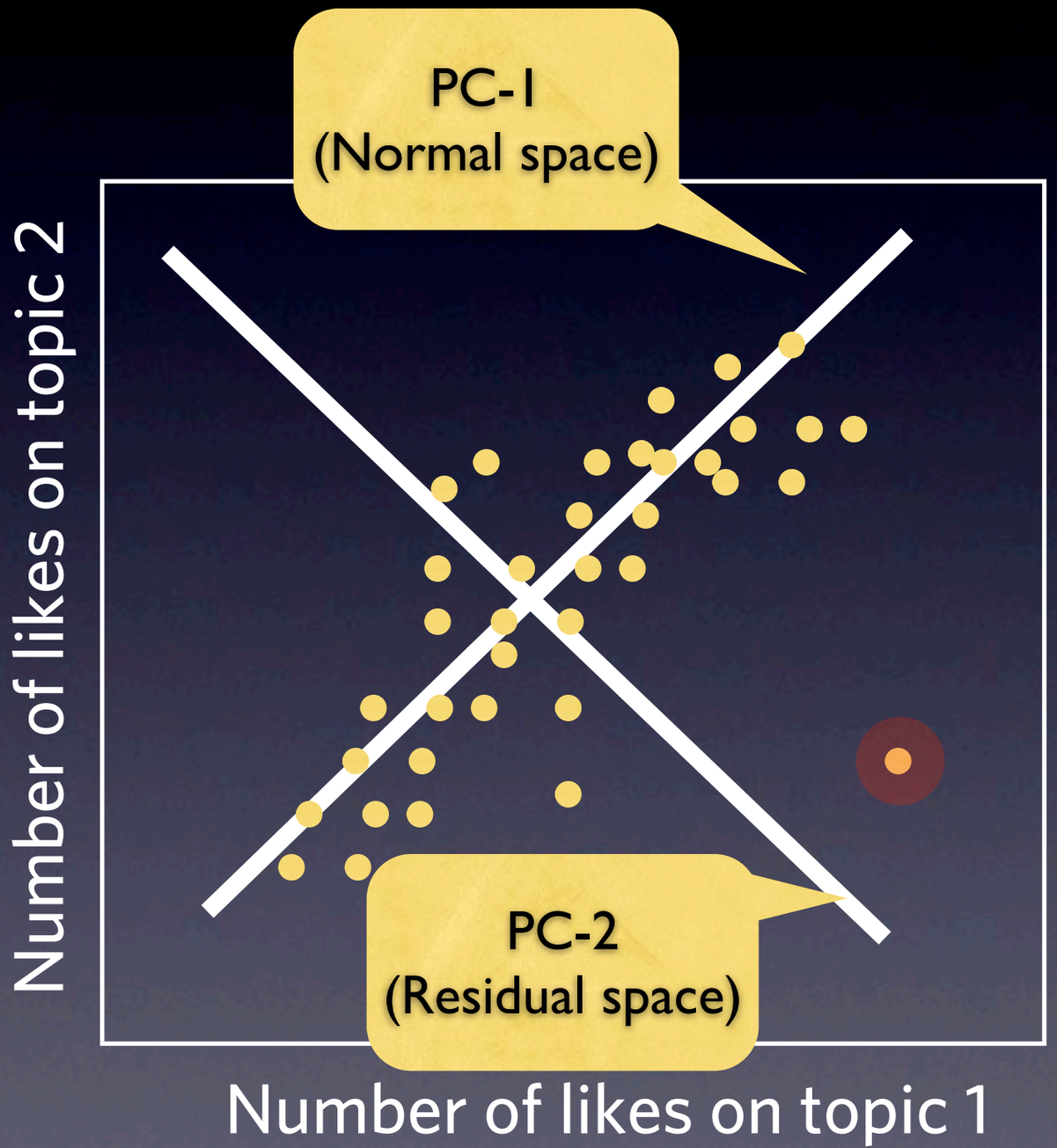
Anomaly detection using PCA



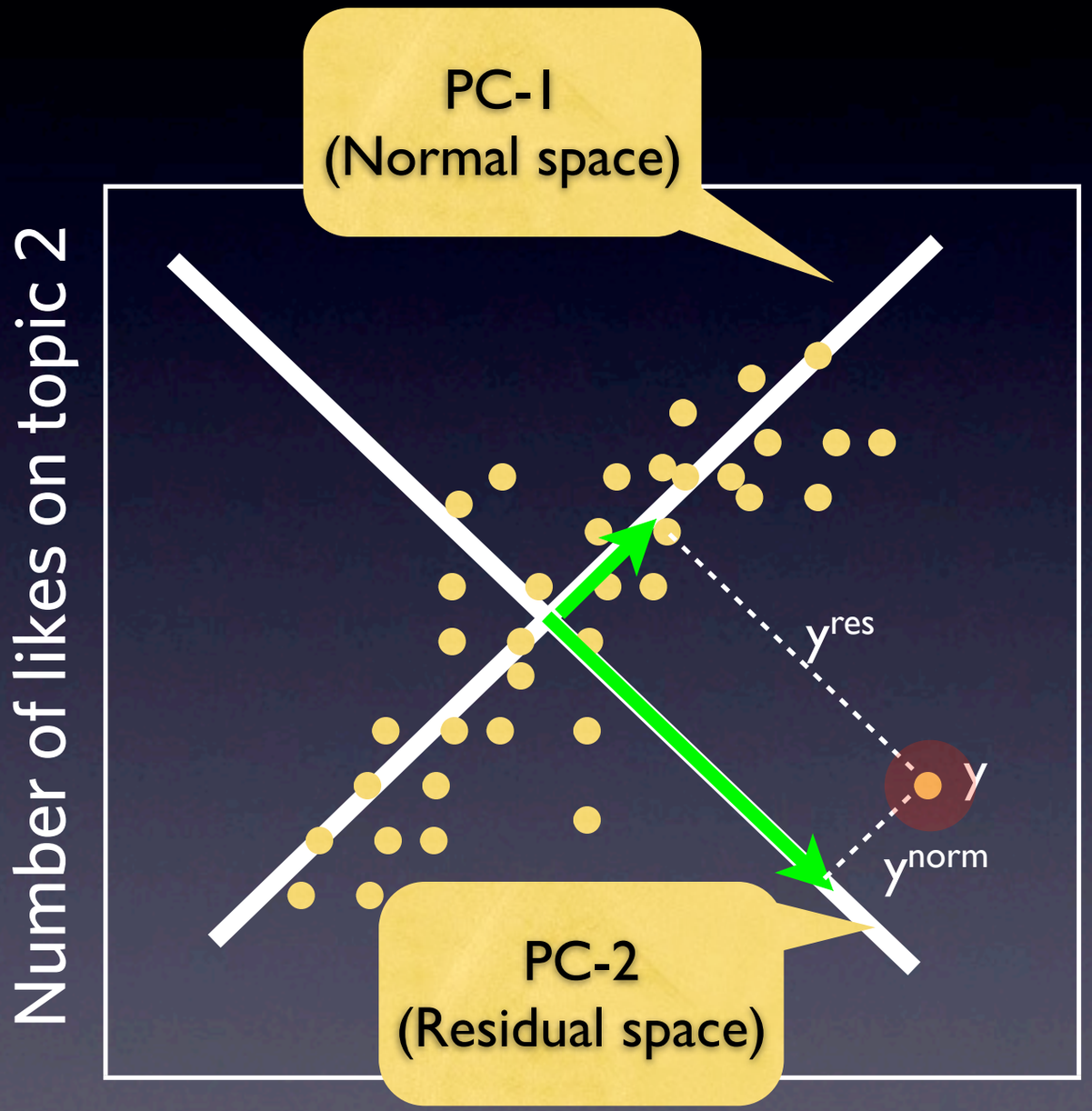
Anomaly detection using PCA



Anomaly detection using PCA



Anomaly detection using PCA



If y^{res} > Threshold, 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

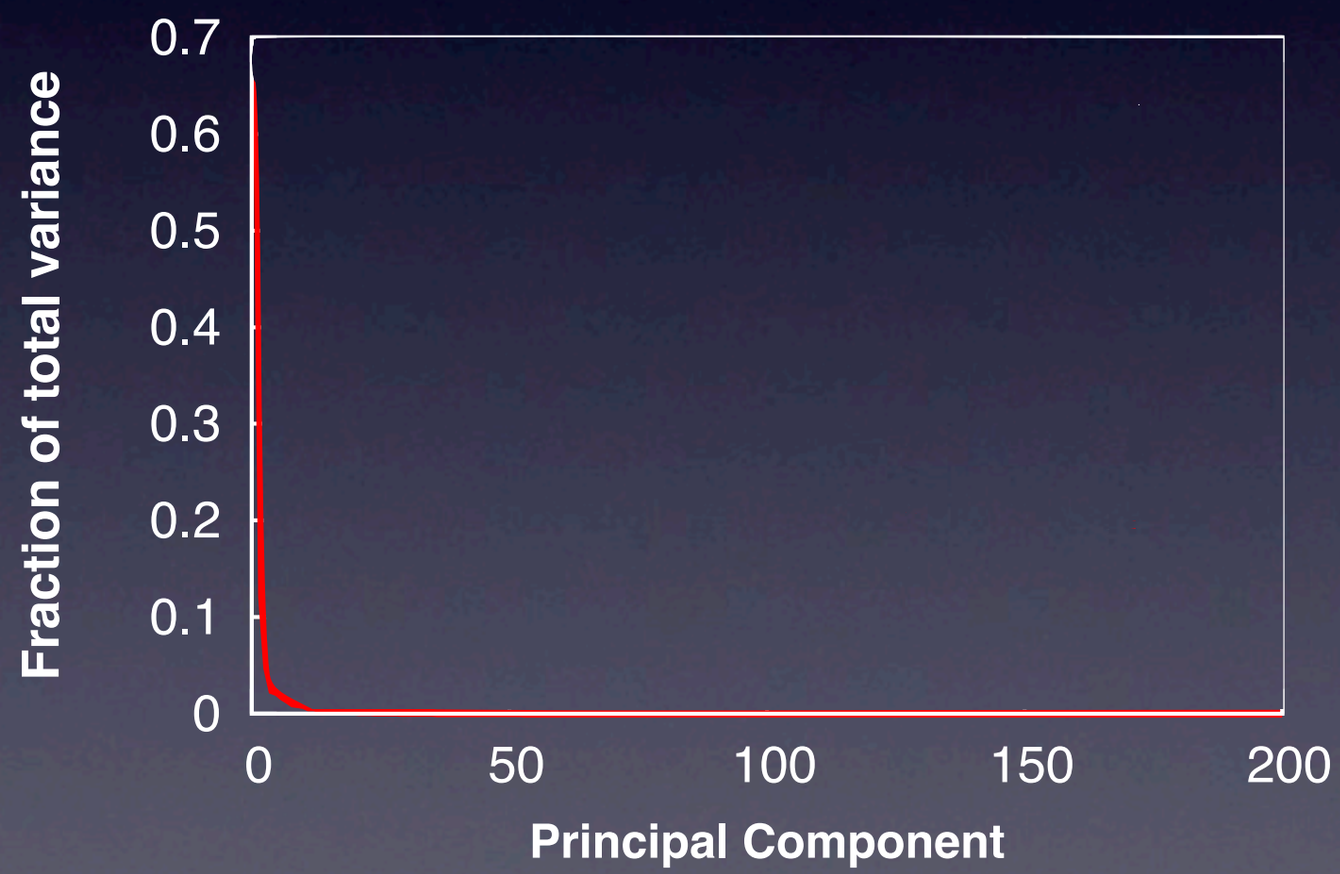
We apply PCA to user behavior defined over 224 page topics

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

We apply PCA to user behavior defined over 224 page topics

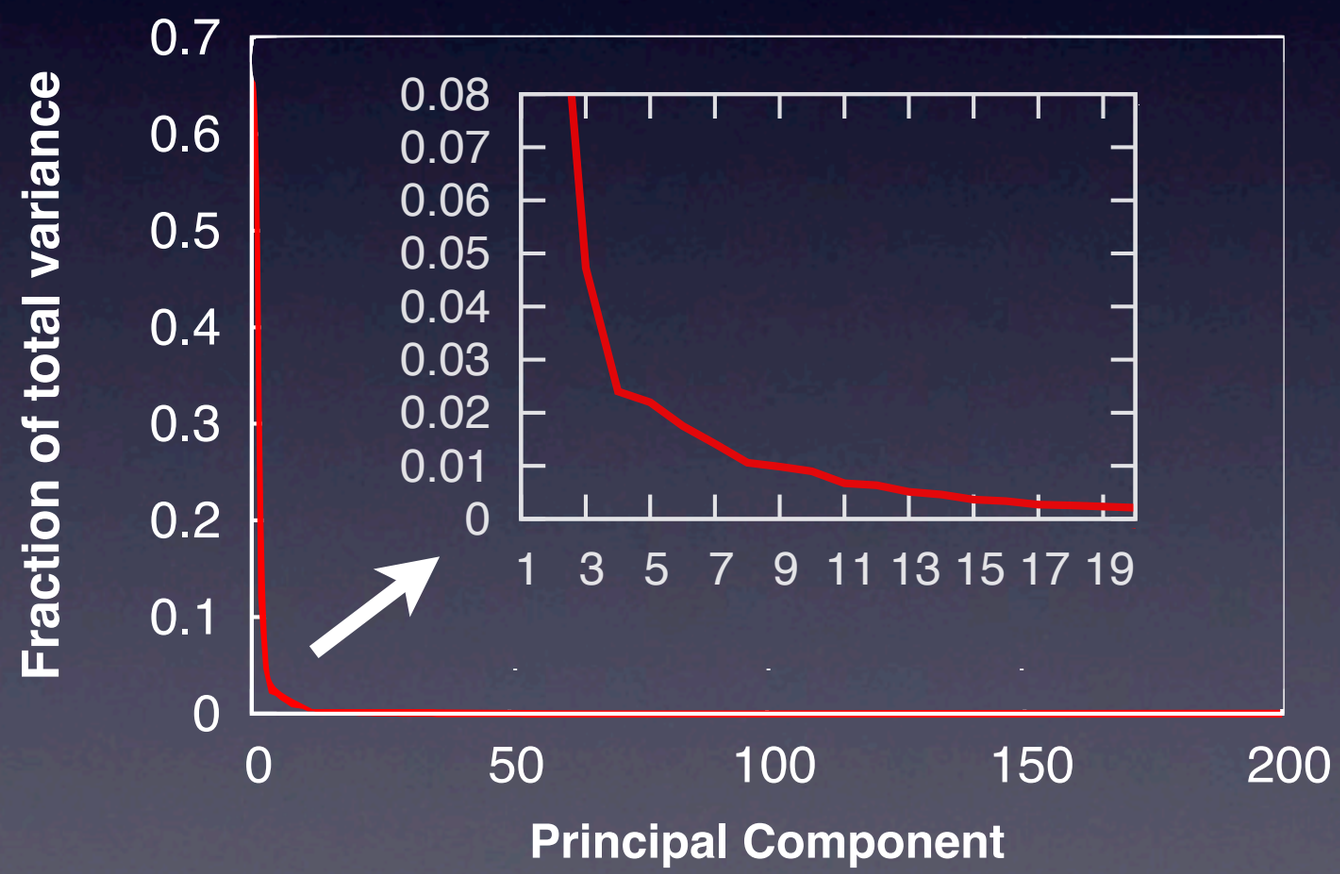


Capturing normal behavior patterns

Are there a few patterns of behavior that are dominant?

Can be answered by looking at variance captured by each PC

We apply PCA to user behavior defined over 224 page topics

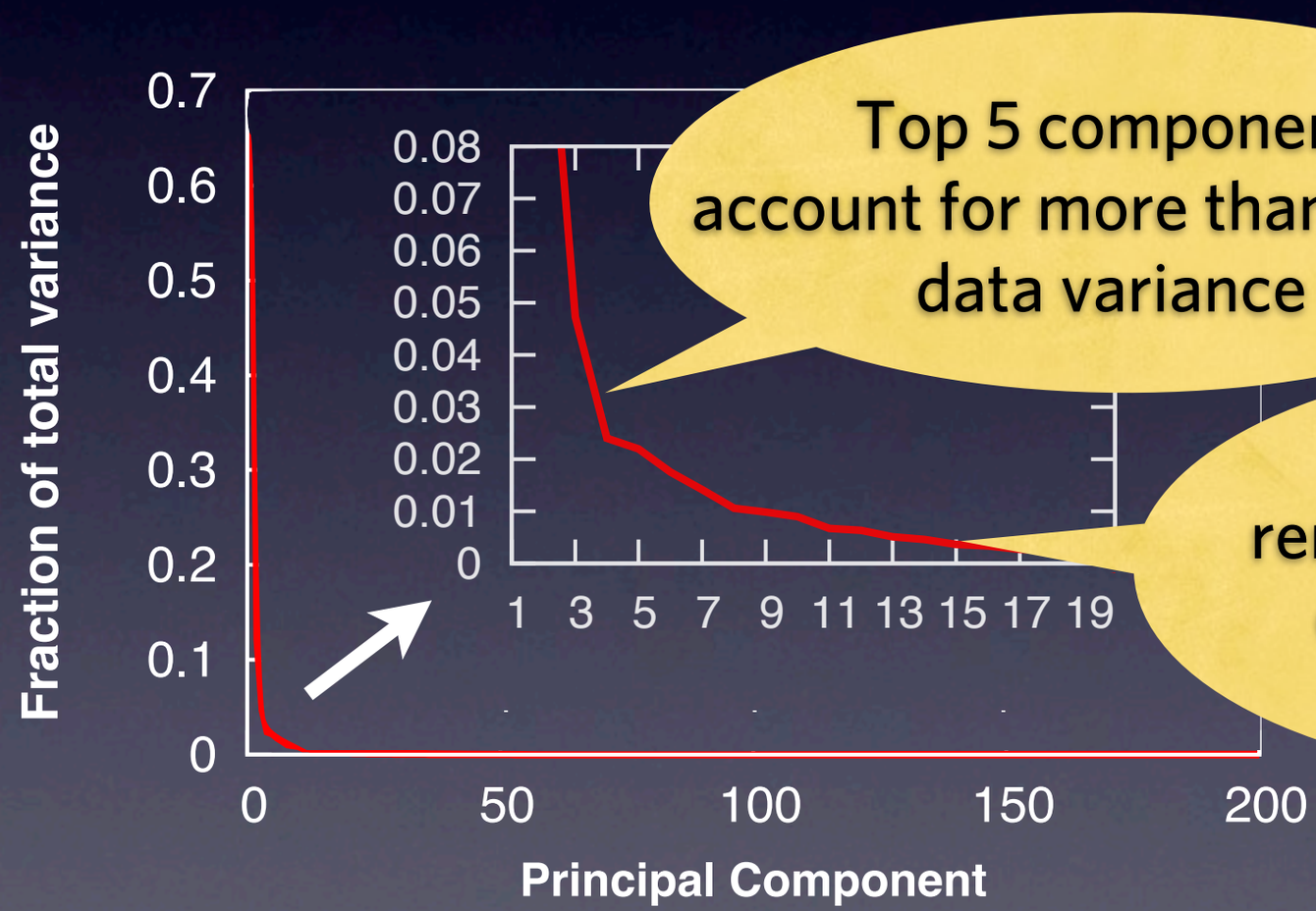


Capturing normal behavior patterns

Are there a few patterns of behavior that are dominant?

Can be answered by looking at variance captured by each PC

We apply PCA to user behavior defined over 224 page topics

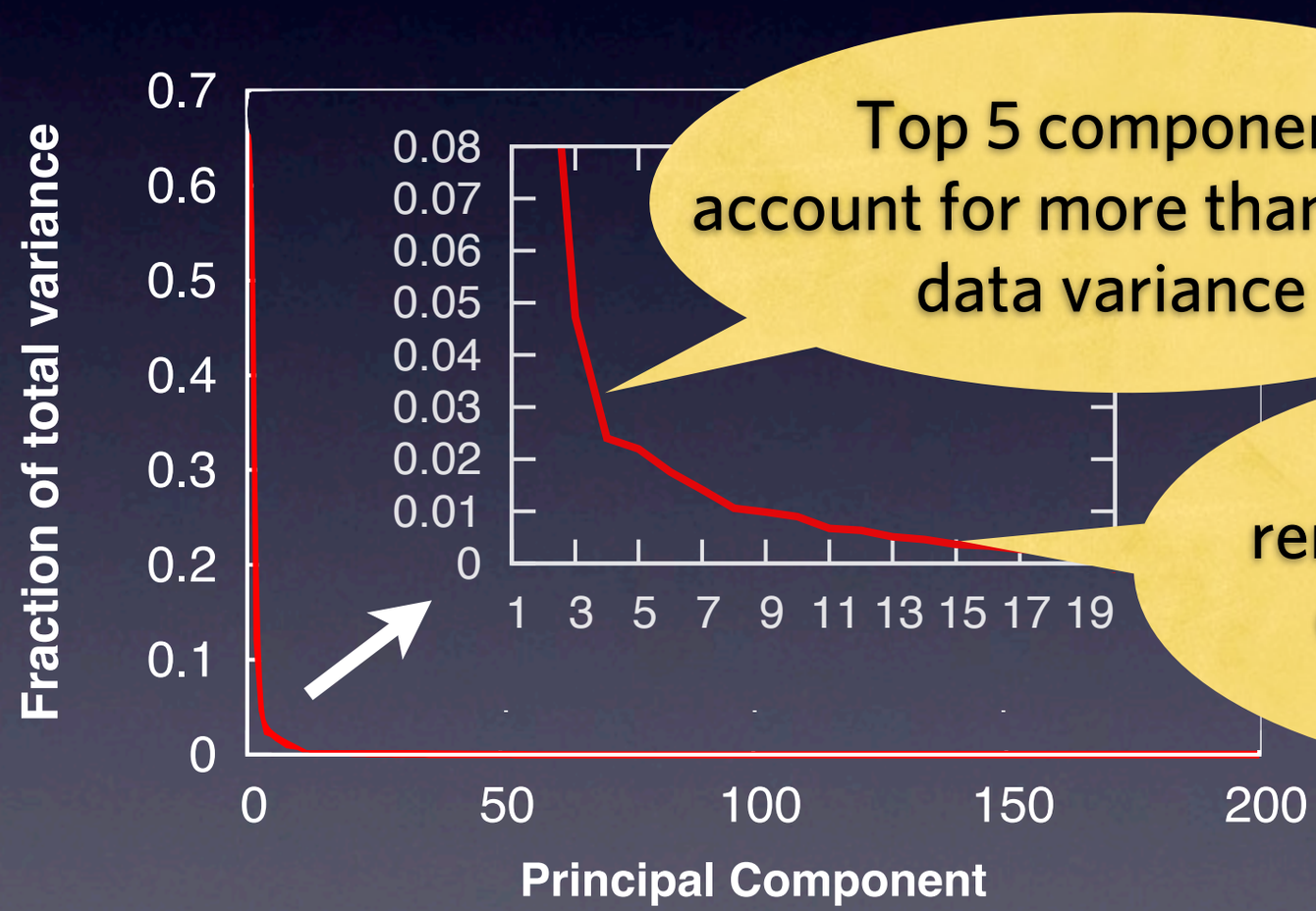


Capturing normal behavior patterns

Are there a few patterns of behavior that are dominant?

Can be answered by looking at variance captured by each PC

We apply PCA to user behavior defined over 224 page topics



We observe such patterns in other social networks too

Evaluation: Detecting Like spammers on 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

Estimating threshold for anomalous behavior

- Find threshold such that 3% of random users are flagged
  - Facebook reported in 2013 that 3% of all users are suspicious

We observe a false positive rate of 3.3%

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

Evaluation: Detecting click-spam on Facebook ads

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 bluff ad and a real ad targeting users in USA
- Heavily instrumented the landing page to capture user activity

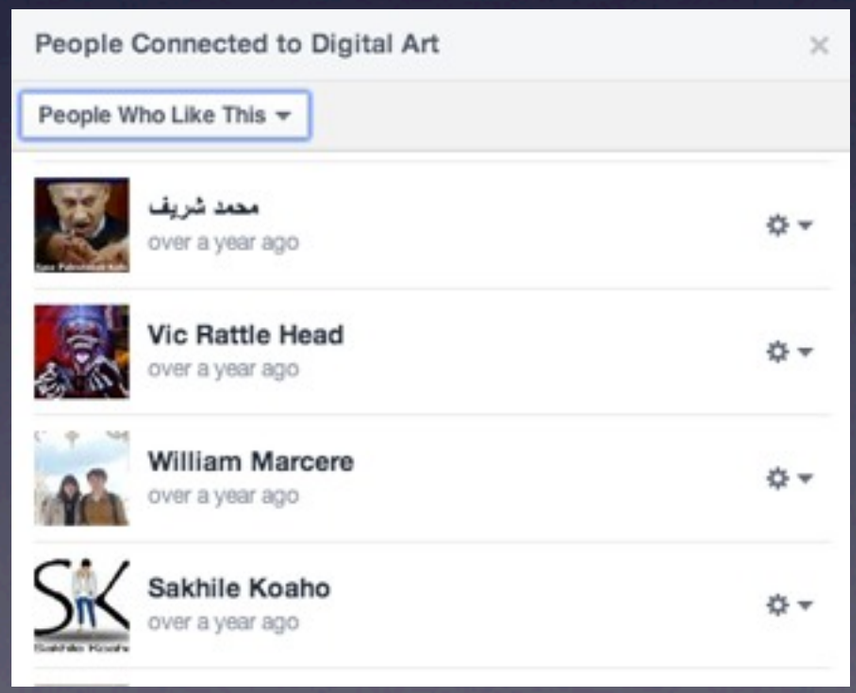
Both bluff and real ad performed nearly identically

- e.g., similar number of clicks and similar levels of activity on landing page

Experiment to catch anomalous clicks

Set up ad to get likes to our page

Find identity of user who liked our page



Click-spam identified

We set up 10 ad campaigns targeting 7 countries

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

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

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:

- More than 85% of all likes by ad users were removed after 4 months
  - Confirms our findings of click-spam
- But a lot of likes by known misbehaving users still exist
  - 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