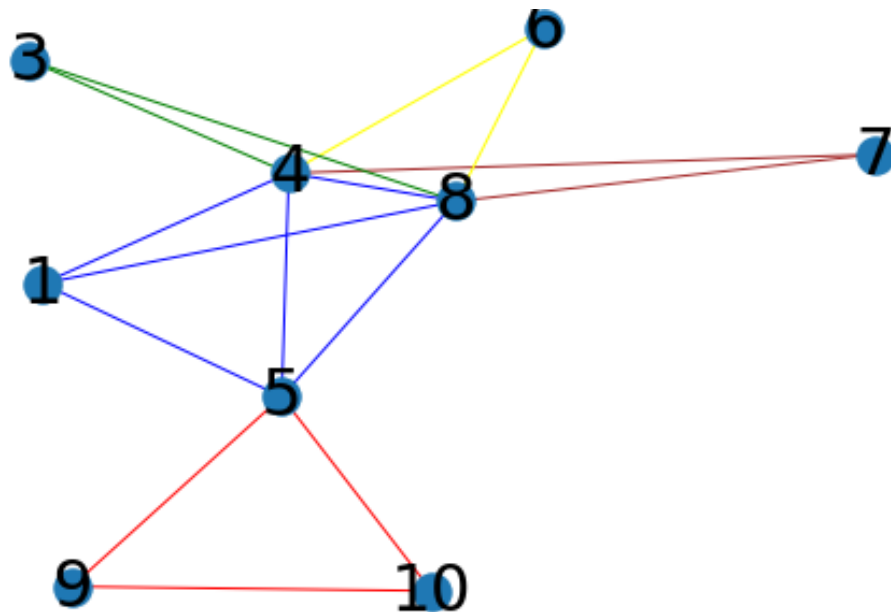


# Detecting Opinion Spammer Groups

*Conference of KPIS Seminar,  
Hamburg University,  
11th of February 2021,  
Department of Informatics,  
Knowledge Technology*



# Agenda

1. Importance of the field
2. What are opinion spamming groups and why do we search for them?
3. Yelps dataset and review system
4. Group finding approach and features
5. Discussion and results

# Importance of the field

- Online reviews influence customers decision making
- Second most trusted source are online product reviews
- ~20% of online reviews are fake
- Payed fake online reviews more and more common

## Goal:

- Restoring trust in online reviews
- Prevent market manipulation

# Opinion spamming groups

- Spammer often act in organised groups
- An agency gets booked and promotes one specific product

## Goal by identifying groups:

- Applying the Clique Percolation Method (CPM)
- Identify groups of users acting suspiciously similar
- Examine the behavior of these groups
- By finding one spammer we can identify a whole group

# Yelps dataset and review system

- Recommendation platform based on user-generated user reviews
- Publically available reviews about businesses (e.g. Restaurants, Hotels, etc.)

## **Dataset:**

- Freely available
- > 8 million reviews and > 200.000 businesses
  - Our focus on reviews in the area of Charlotte, USA in 2016

## 4: Group finding approach (CPM) and features

### Build Search Space

- $G = (U, E)$
- $U$  = Users,  $E$  = Relationships
- Take 2 reviewers ( $u_j, u_k$ ) and add to  $G$ 
  - Same business
  - Same rating
  - Within 6 days

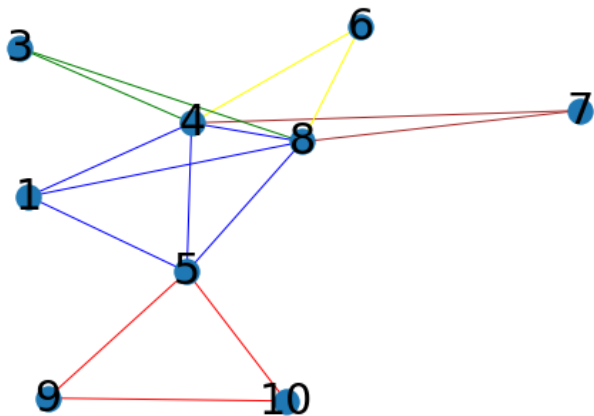
### Find Groups

- $C \subseteq U$
- Where  $C$  is a subgraph of  $G$  where reviewers in  $C$  (nodes) do have an undirected relationship  $e$  to each other (node).

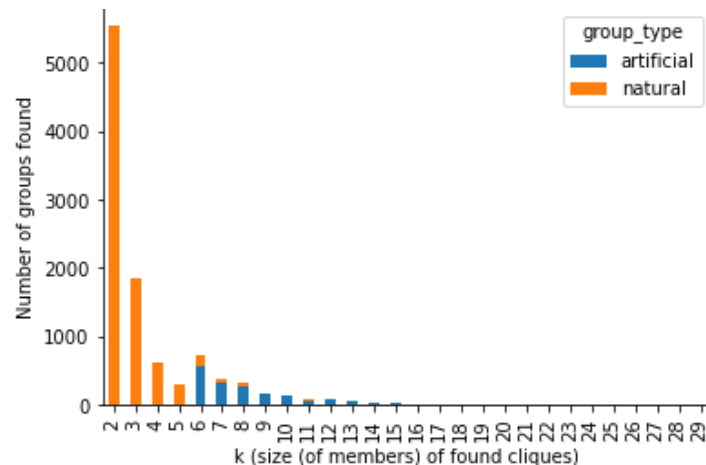
### Rank Groups

- Cosine Similarity (CS)
- Extreme Rating (EXT)
- Burstiness (BST)
- Combined to Suspicious Score (Sus)

## 4: Group finding approach and features



Clique examples out of the suspicious user graph  $G$   
(own figure)



Natural and artificial group-size distribution in our dataset (own figure)

## 4: Group finding approach and features

### Burstiness (BST):

- The smaller the distance between the reviews given by a user (in days), the greater is the BST of that user.

### Extreme Rating (EXT)

- When a user has given only 1 or 5-star reviews an EXT of 1 is given to that user, 0 otherwise.

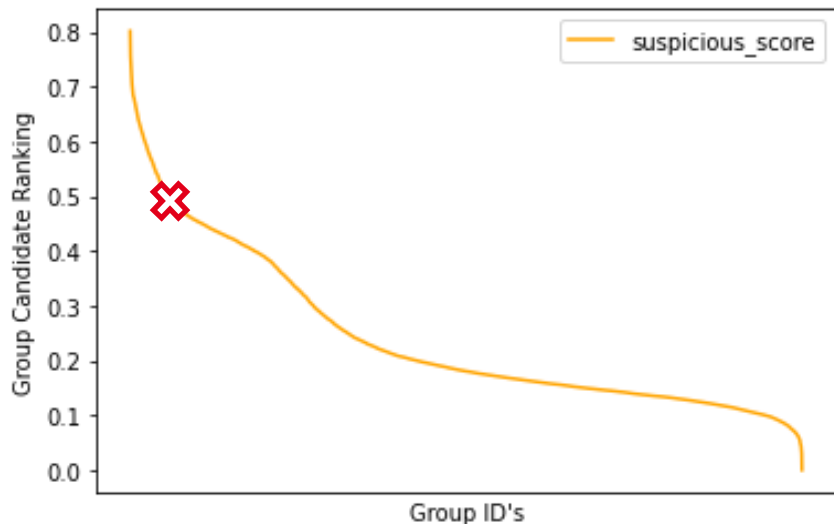
### Cosine Similarity(CS)

- On all pairs (permutations) of reviews in a group, the cosine function is applied. After that, the mean value of all calculated CS this way is aggregated to the group.



## 5: Discussion and results

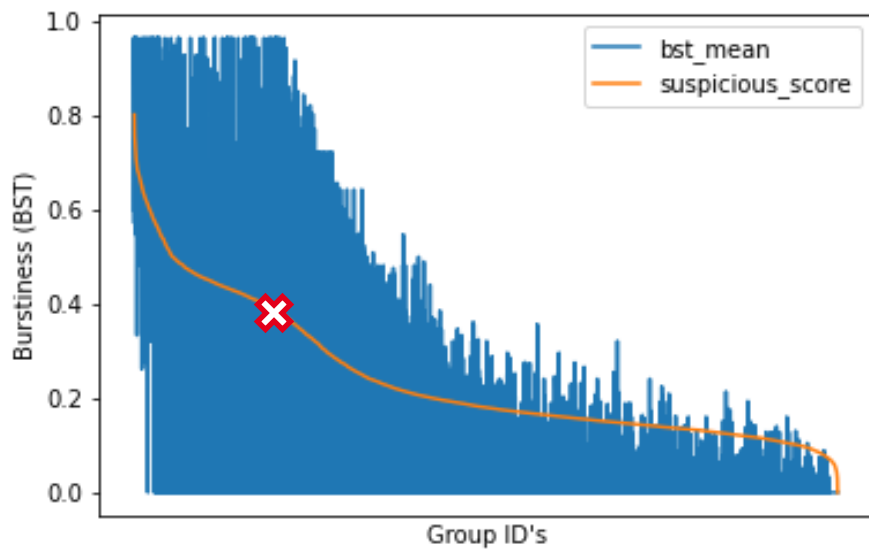
Group Candidate Ranking (own figure)



- 4 % of all groups do have a Sus of  $> 0.5$  (~500)
- 94 % of them are natural groups
- Max Sus value is 0.8

## 5: Results: BST

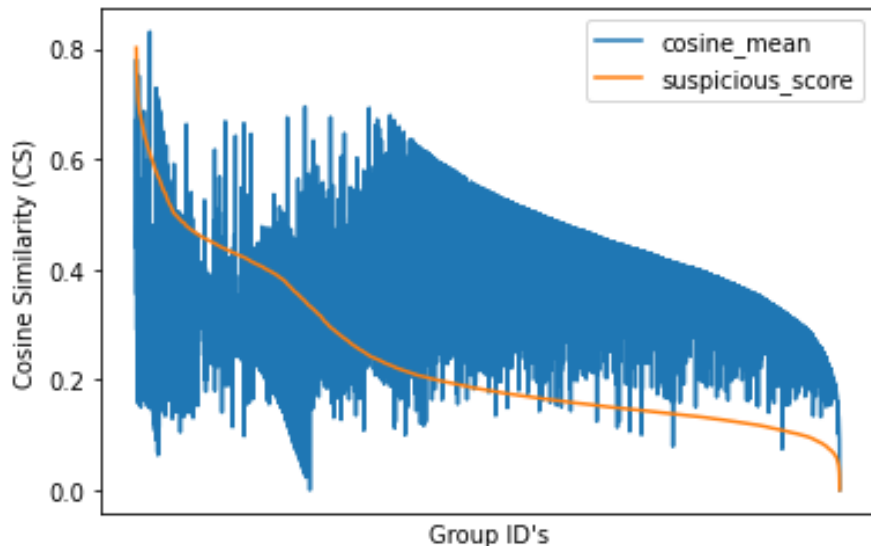
Burstiness (BST) own figure



- Groups are sorted by Sus
- At  $Sus \geq 0.4$  BST is almost 1
- Is indicating that the top 18 % of our members in our groups are only active within a few days
- But it varies strongly group by group

## 5: Results: CS

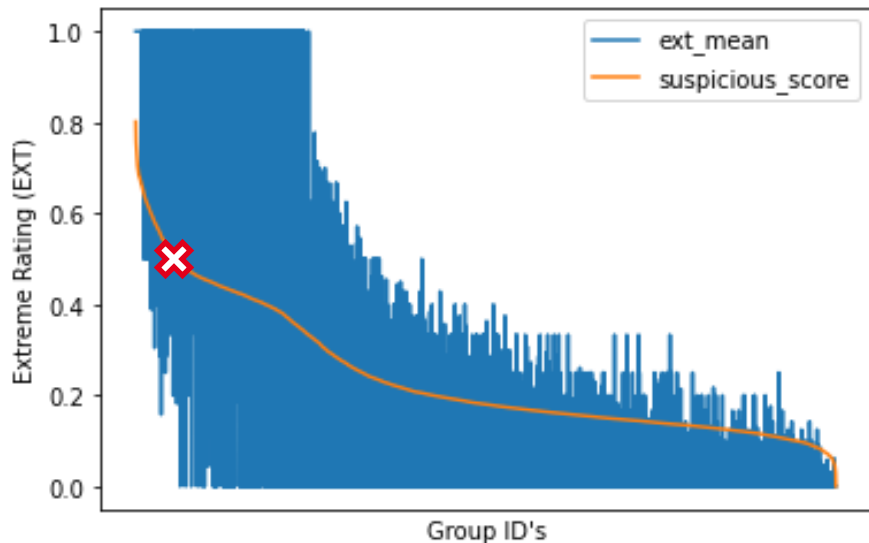
Cosine Similarity (CS) (own figure)



- CS strongly varies
- Assumingly spammers try to avoid copy-paste in order to not get detected
- Maybe Yelp filters out obvious duplicates

## 5: Results: EXT

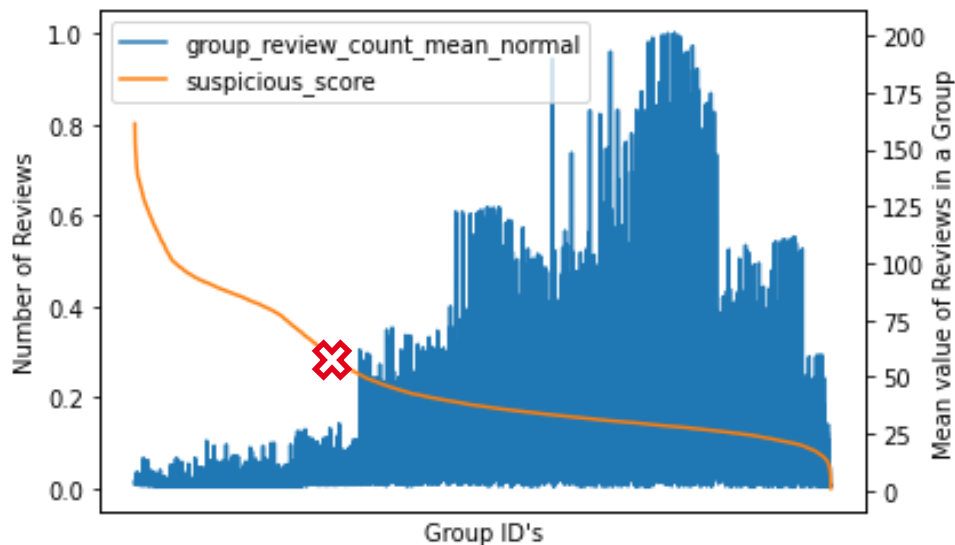
Extreme Rating (EXT) (own figure)



- Constantly increasing
- After a Sus of 0.5 the EXT is always above 0, indicates that those groups are suspicious to demote/promote target products together

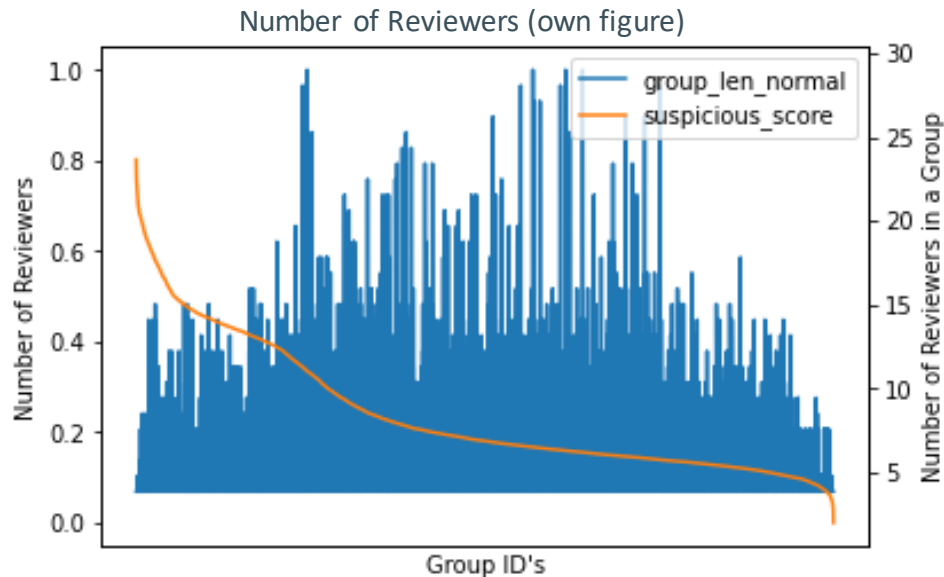
## 5: Results: Number of Reviews

Number of Reviews (own figure)



- The number of reviews of each user in a group decreases as the Sus value goes up
- Where  $Sus > 0.3$  the average member of a group writes 3 reviews as a  $Sus < 0.3$  writes 27
- Indicates spammers only write a few reviews
  - To reduce the damaged duo to the deletion
  - Allows to write multiple reviews for one target product

## 5: Results: Number of Reviewers



- Strongly varies
- Highly suspicious groups are mostly small groups and therefore natural ones

## 5: Discussion and results

- The CPM-Method is able to identify groups with highly suspicious behaviour
- Due to the lack of implemented features the approach failed to deliver better results
- Hyperparameter optimization and different datasets could have improved the results
- We could not prove that large organized (artificial) groups exist

# Thanks for your attention!