



University of Milan - Bicocca
Master Degree in Data Science
Academic Year 2022/2023

r/seduction: community and content understanding of the subreddit



Alessio Pasinato - 887000
Julia Bui Xuan - 882385
Michele Salvaterra - 891109

Introduction



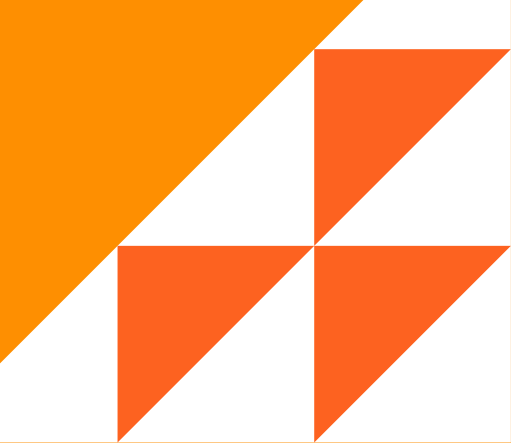
reddit

The main objective of this project consists in the analysis the **r/seduction subreddit**, a forum for discussing and sharing tips on how to attract and pick up romantic partners.

The subreddit has faced controversy in the past for:

- promoting **toxic** and **manipulative behaviour**
- objectifying and **devaluing women**
- promoting **false** and **harmful stereotypes** about women and relationships





- Analysis of the subreddit's contents, discussed topics and their relative sentiment
- Identification of most influential users
- Analysis of the subreddit structure and users interactions
- Identification of communities and their relative characteristics

Research Questions

Dataset

Dataset: **Posts**

Data collected: 1349

Period of time: 1st october - 31st October 2022

Language: English

Attribute **description**:

- **title**: title of the post
- **text**: body of the post
- **author**: author of the post
- **id**: post identifier
- **created_utc**: time the submission was created, represented in Unix Time
- **URL**: post's URL
- **upvote_ratio**: the percentage of upvotes from all votes on the submission
- **num_awards**: number of awards the post has received

Dataset: **Comments**

Data collected: 14089

Period of time: 1st october - 31st October 2022

Language: English

Attribute **description**:

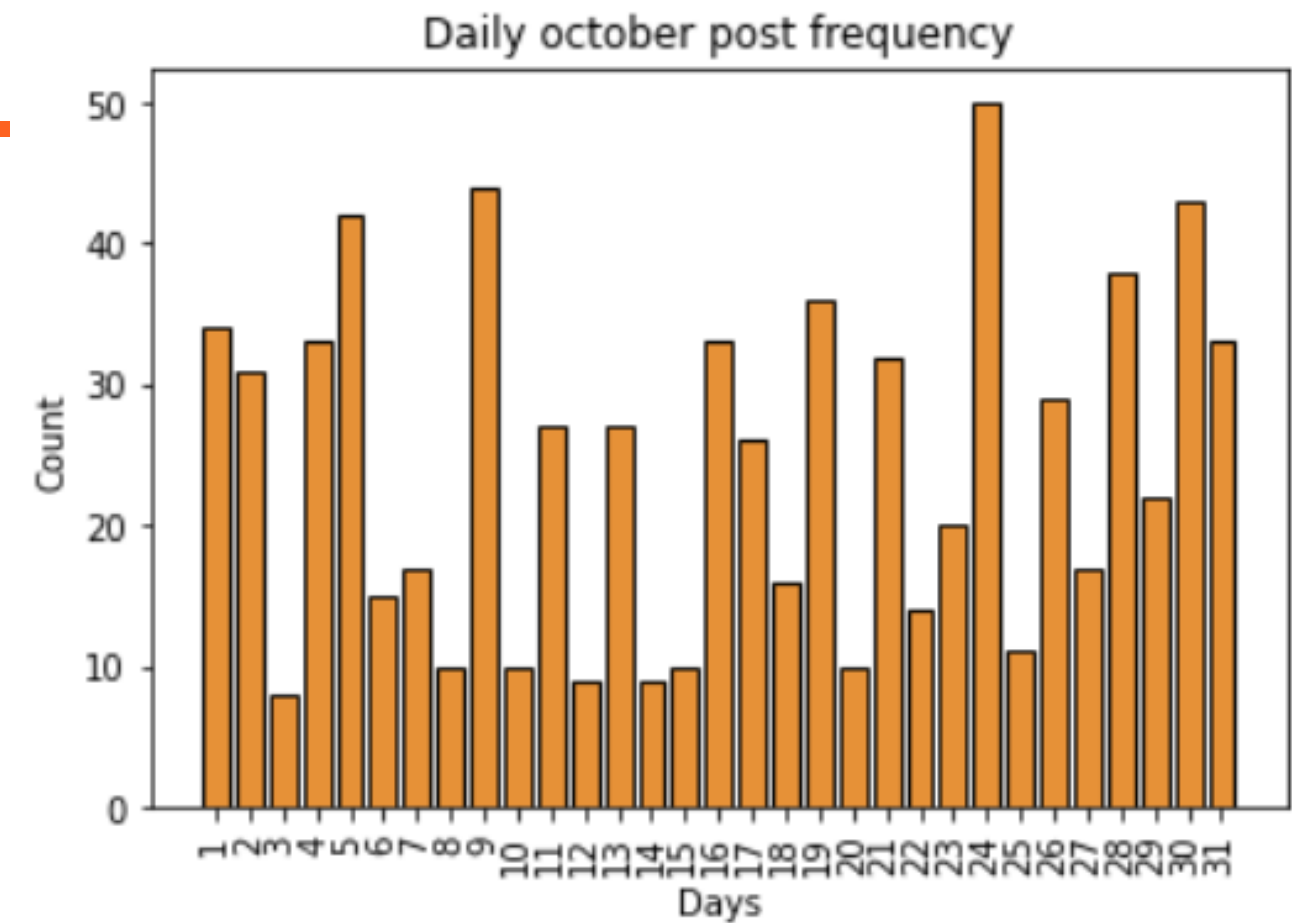
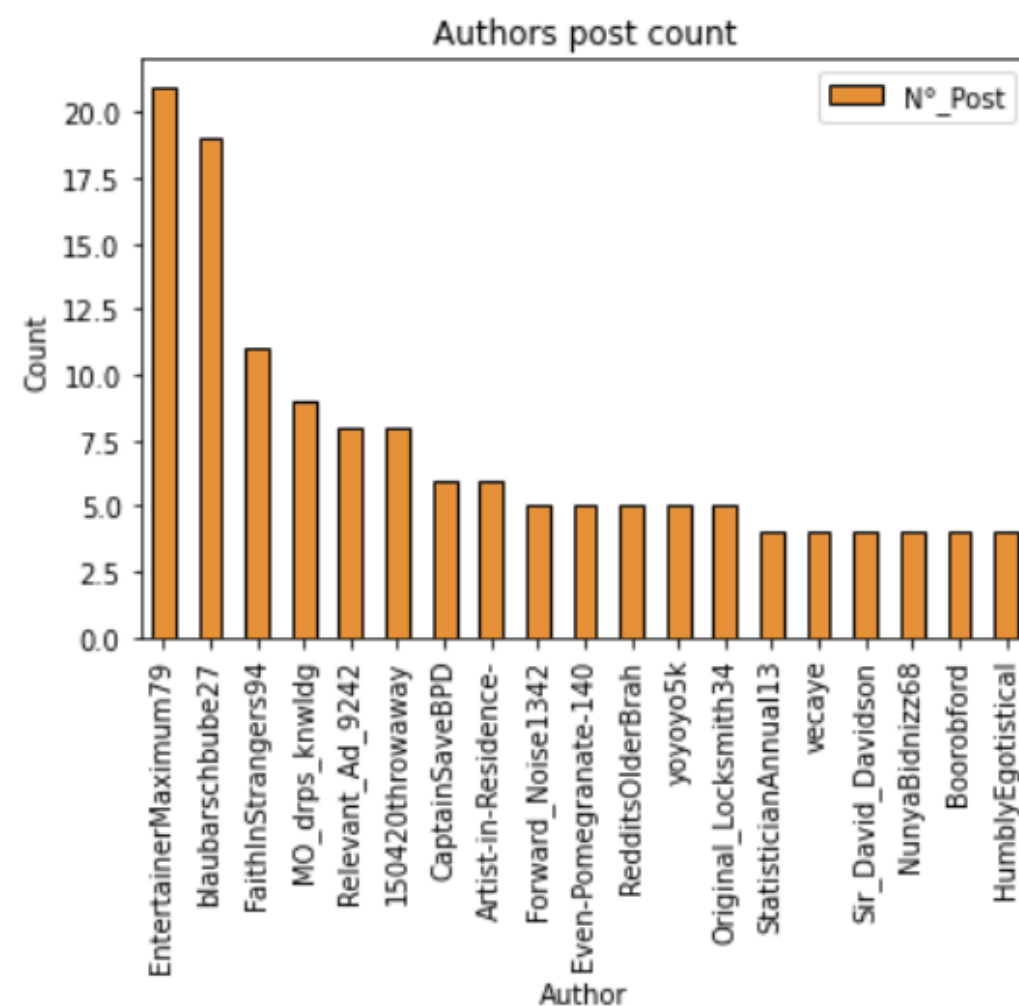
- **author**: author of the comment
- **body**: comment's content
- **comment id**: id of the comment
- **created_utc**: time the comment was created, represented in Unix Time
- **URL**: comment's URL
- **n upvotes**: upvotes of the comment
- **parent id**: the id of the parent comment

Data exploration

Minimum: 8 post, 3 october

Maximum: 50 post, 23 october

Average: 24 post



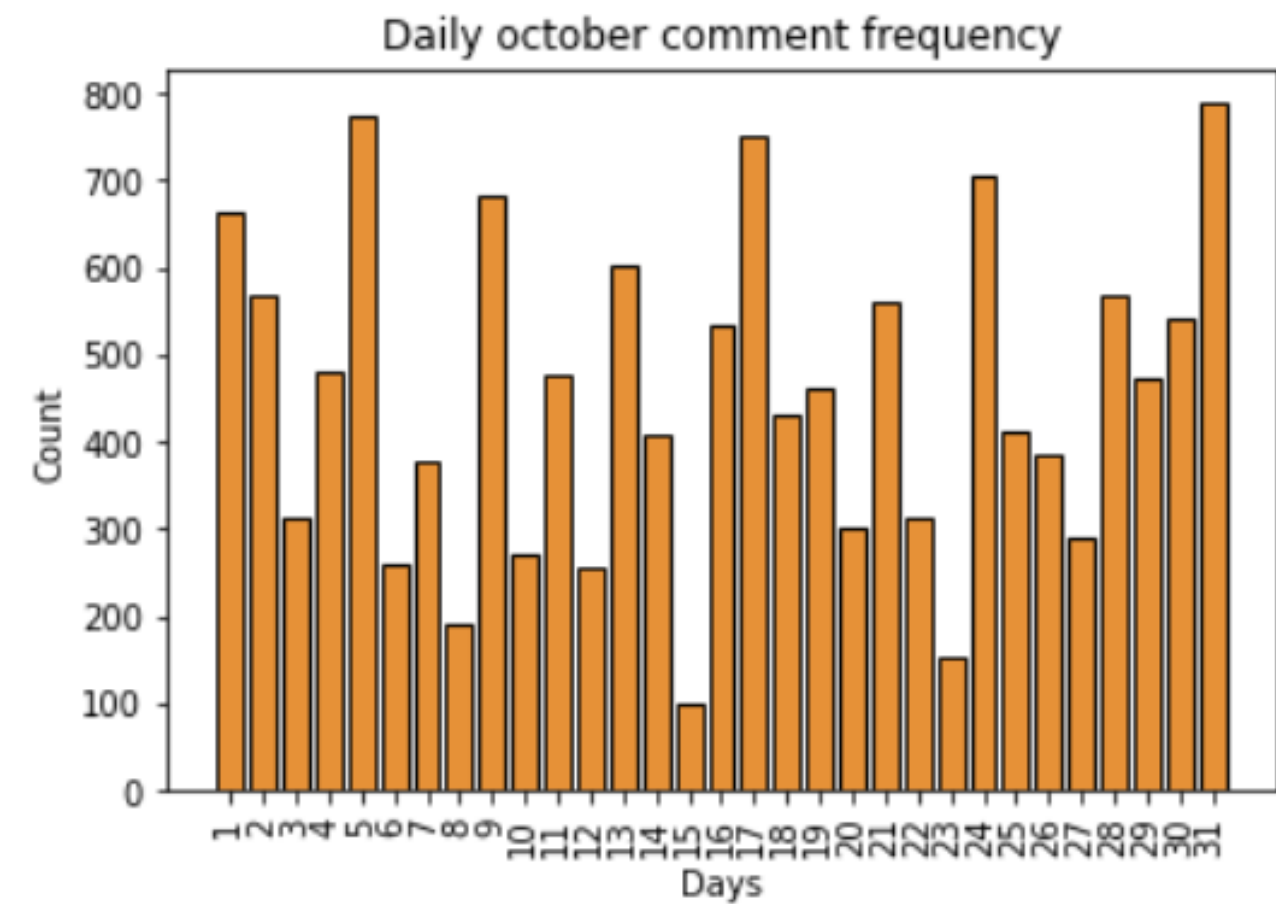
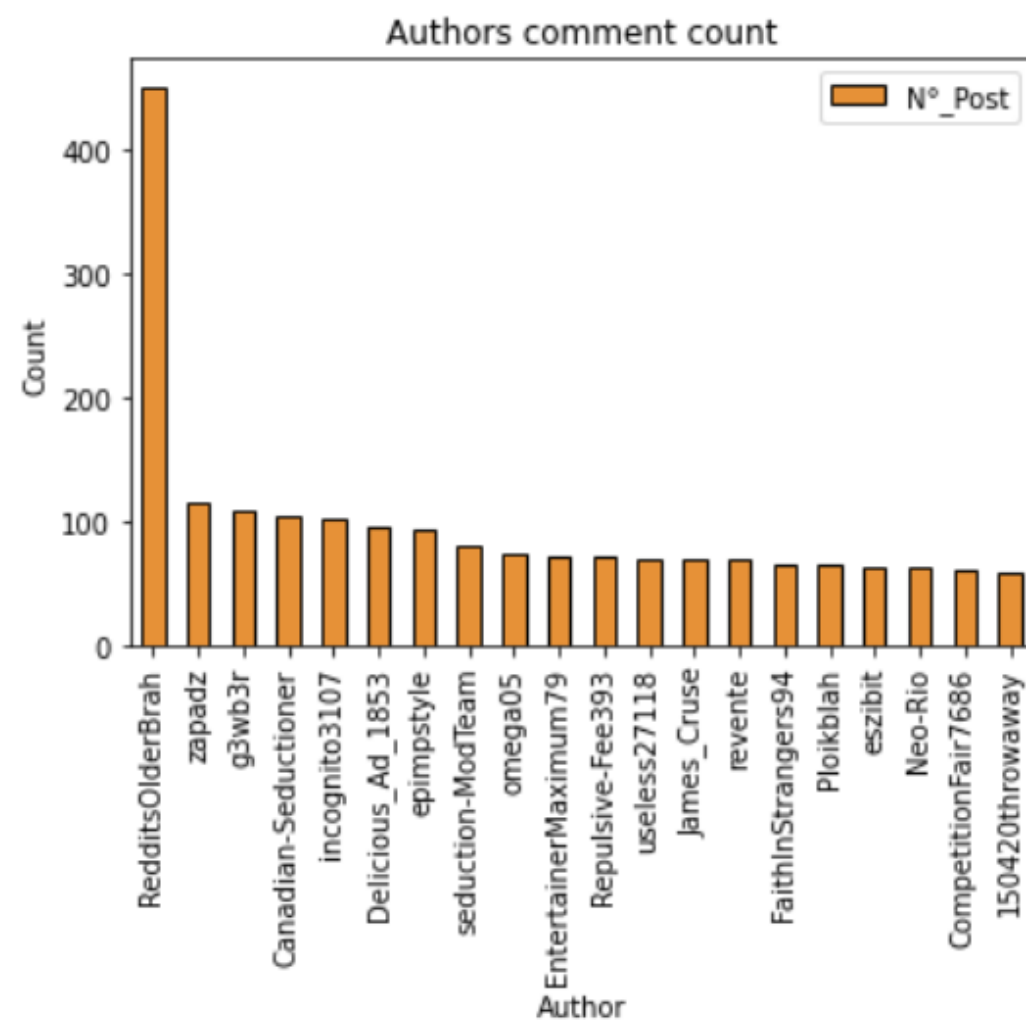
The author with gratest number of posts is "*EntertainerMaximum79*" , with 21 posts created.

Data exploration

Minimum: 99 comments, 15 October

Maximum: 789 comments, 31 October

Average: 450 comments



The author with the greatest number of comments is "*RedditsOlderBrah*", with 451 comments created.



Data exploration

Best and worst comment



Author *"thatstotallyracist"*

"Are you telling me that there are hot, single women waiting to meet me in my area?"

Vote 552

Author *"Environmental_Cress2"*

"prolly. I prefer nerdy girls. But I'm mostly around party tiktok type of girls. So that affects my perception"

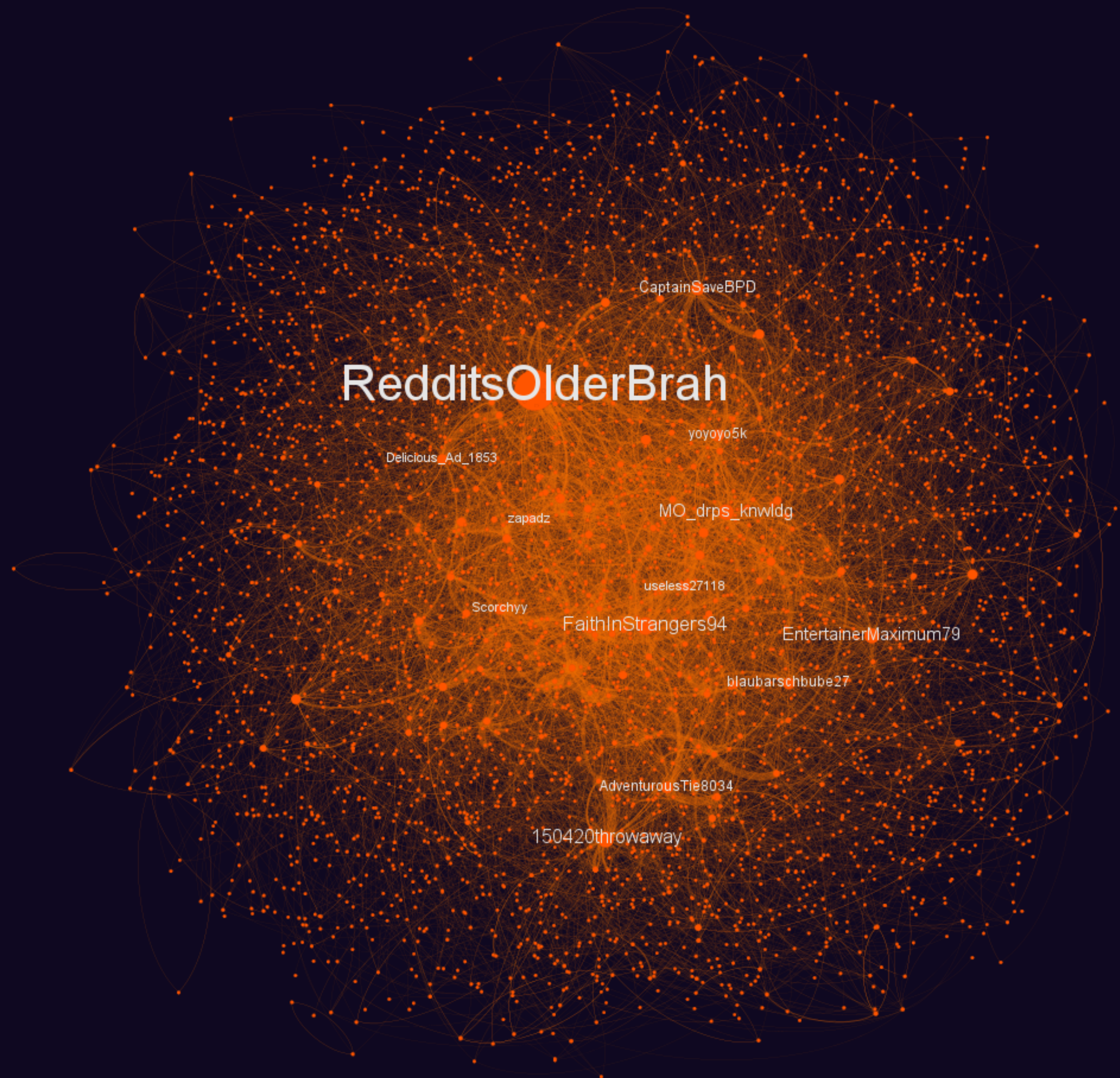
Vote -63





- 4145 nodes
- 13690 edges
- Directed
- Weighted
- Multigraph

Graph



Graph metrics

In-degree centrality

User	In-degree centrality
RedditsOlderBrah	0.084
FaithInStrangers94	0.053
MO_drps_knwldg	0.048
150420throwaway	0.044
EntertainerMaximum79	0.039

Table 1. Top 5 highest in-degree centralities

Betweenness centrality

User	Betweenness centrality
RedditsOlderBrah	0.106
FaithInStrangers94	0.034
MO_drps_knwldg	0.026
150420throwaway	0.020
EntertainerMaximum79	0.020

Table 3. Top 5 highest betweenness centralities

Out-degree centrality

User	Out-degree centrality
RedditsOlderBrah	0.098
zapadz	0.027
g3wb3r	0.024
incognito3107	0.024
Canadian-Seductioner	0.023

Table 2. Top 5 highest out-degree centralities

Closeness centrality

User	Closeness centrality
RedditsOlderBrah	0.303
FaithInStrangers94	0.287
MO_drps_knwldg	0.278
EntertainerMaximum79	0.273
blaubarschbube27	0.272

Table 4. Top 5 highest closeness centralities

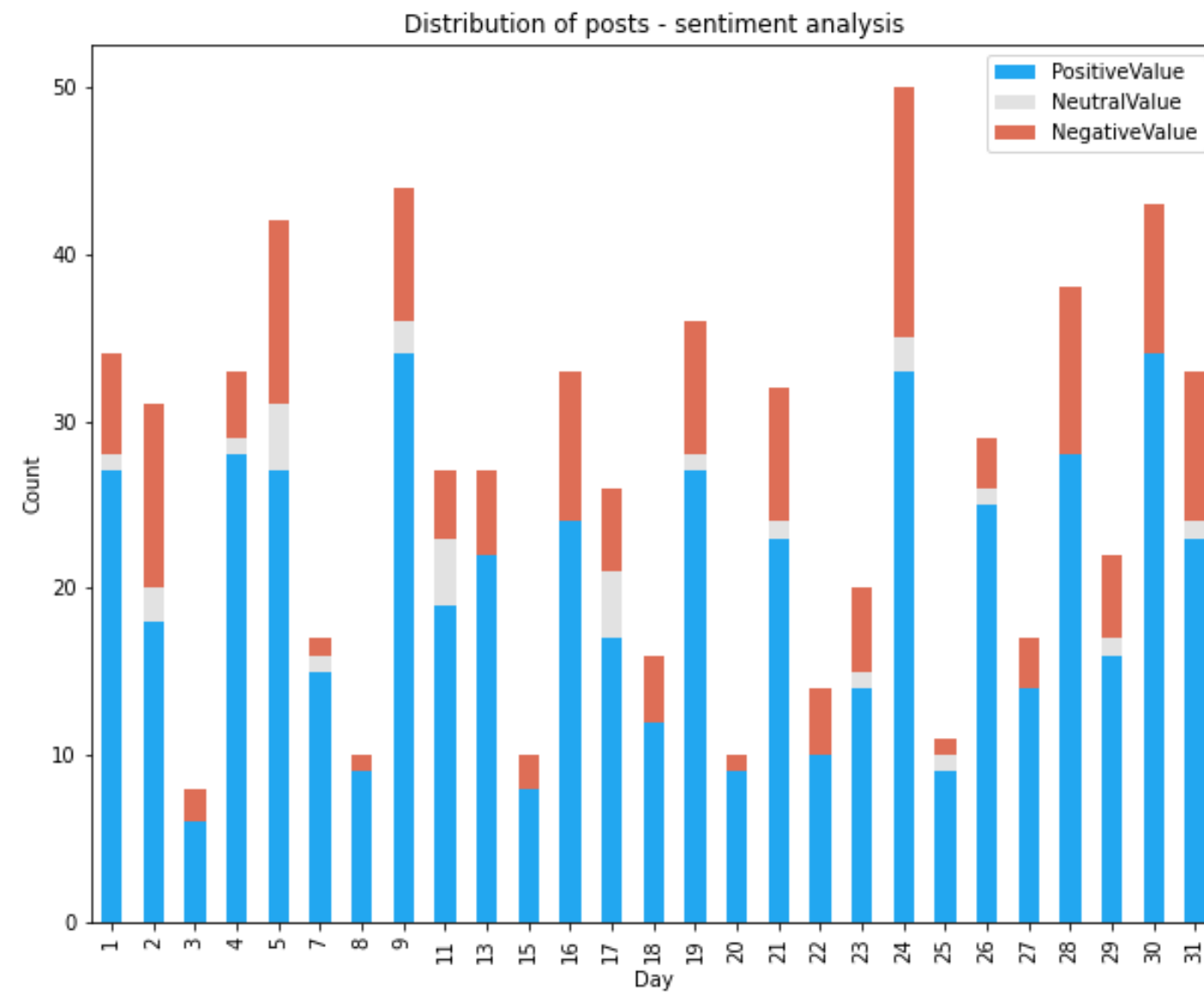


Sentiment analysis

- Model used: **VADER** (Valence Aware Dictionary for Sentiment Reasoning)
- **Polarity classification** (positive, negative, neutral)
- Sentiment of a text: summing up the valence scores of each word, adjusted by the **context** in which the word is used.
- Ability to handle: use of negations, sarcasm, emoticons, capitalization, punctuations, punctuation emphasis, degree modifiers

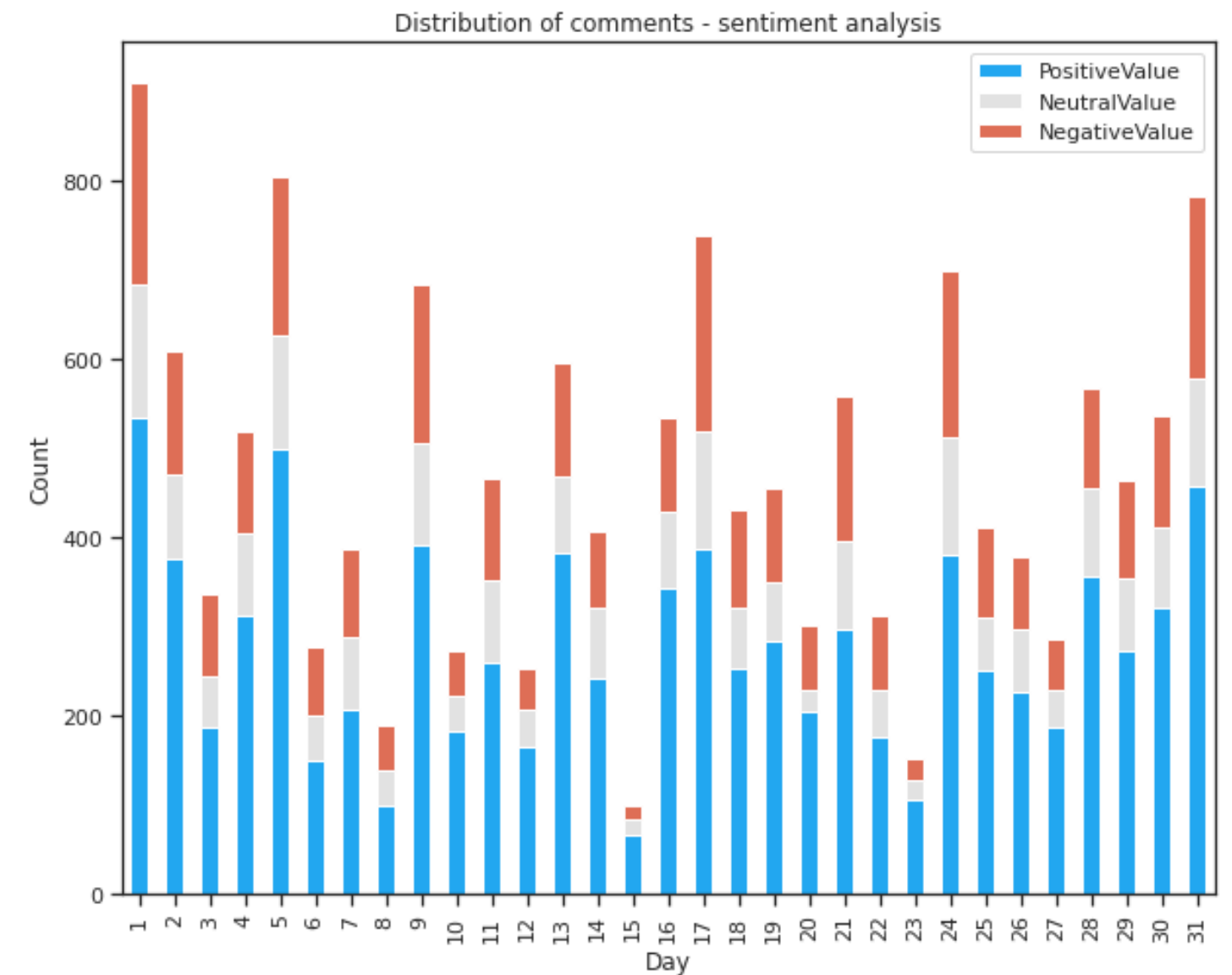


Post dataset



Positive: 72.6%%
Neutral: 4.2%
Negative: 20.2%

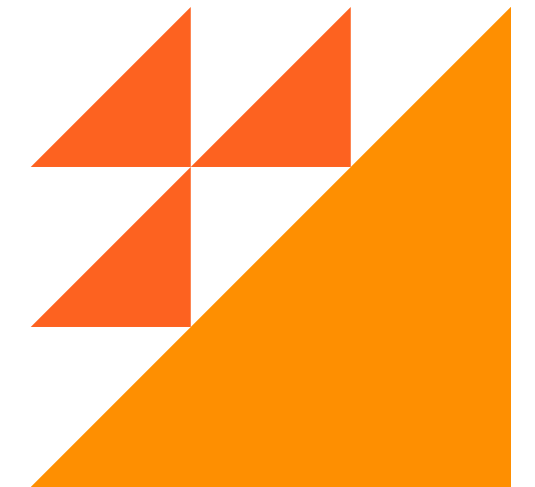
Comments dataset



Positive: 59.4%
Neutral: 16.6%
Negative: 23.9%

Topic modelling

- Approaching girls
- Ways to improve appearance
- Approaches and rejections
- First dates
- Relationships
- Sexual relations
- Body language



Community detection and analysis

Community	Positive	Neutral	Negative
1	58.1%	14.4%	27.4%
2	58.9%	16.7%	25.4%
3	58.7%	17.8%	23.5%
4	54.7%	20.9%	24.5%

Table 5.

Four biggest communities' sentiment

Table 6.

Most positive and most negative communities' sentiment

Communities		
	Most Positive	Most Negative
Positive	69.3%	52.0%
Neutral	14.5%	13.3%
Negative	16.2%	34.7 %

Conclusions



reddit

- Most **influential** and **central** user: "RedditsOlderBrah"
- Prevalent **sentiment**: Positive
- Most discussed **topics**: ways to improve appearance, approaches to girls, rejections, first dates, relationships, sexual relations and body language
- **Communities** detected: 26, similar sentiment analysis
- Most negative community: 34.7% of negative comments
- Most positive community: 69.3% of positive comments

