

User Documentation

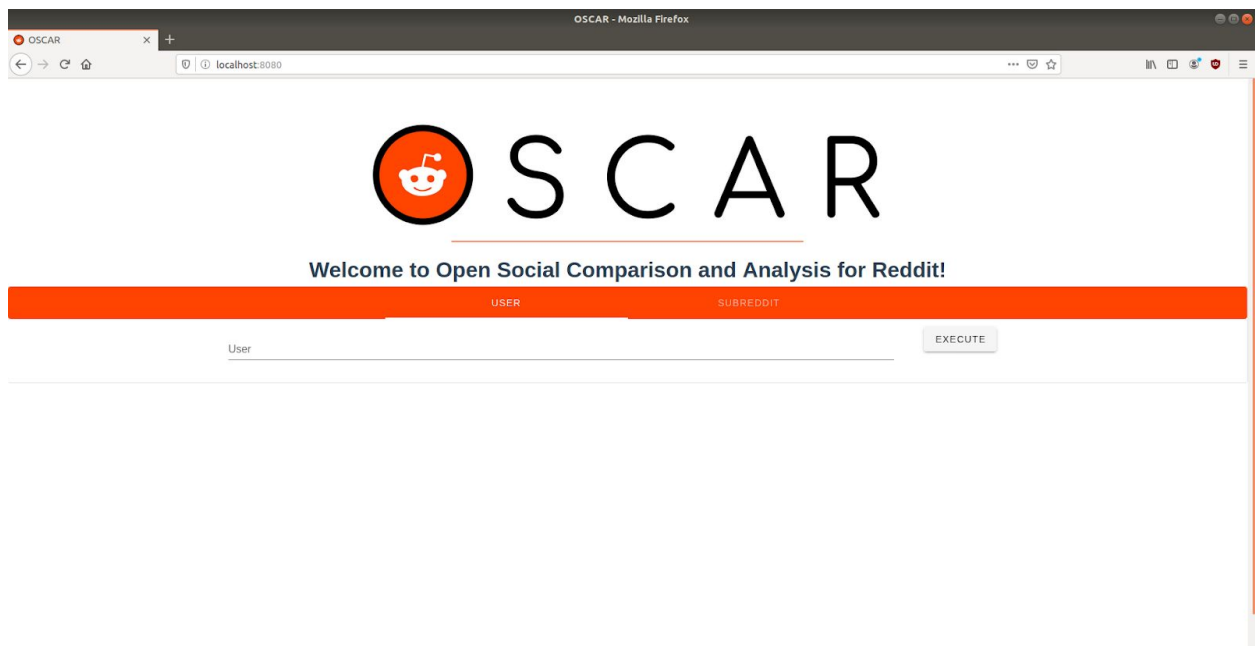
Getting Started

About

OSCAR, or Open Source Comparative Analysis for Reddit, is a tool that allows people to easily view summarized analysis of subreddits and users they are interacting with on Reddit. An OSCAR user can submit a subreddit or user they are interested in learning more about for processing. Once processing is complete, the user can interact with the results of the analysis. The aim of OSCAR is to allow Reddit users to become more educated about the people and topics they are interacting with online and potentially detect suspicious users or biased information more easily.

Getting Started

To get OSCAR up and running, use the installation instructions located in the User Manual. Once the OSCAR user interface is running, you will see this screen.



The home screen for OSCAR is very simple; it simply includes two tabs, "USER" and "SUBREDDIT", which are indicated in the image below. Click on these labels to switch between the two tabs.



Welcome to Open Social Comparison and Analysis for Reddit!

USER

SUBREDDIT

User

EXECUTE

The User tab is used to submit a user for analysis. The Subreddit tab is used to submit a subreddit for analysis.

Using the Interface

Submitting a User

Ensure that you are on the User tab. Use the text box, indicated by the black box in the image below, to type in the username of the Reddit user that you want to analyze. Click the Execute button, indicated by the blue box in the image below, to start the analysis.



Welcome to Open Social Comparison and Analysis for Reddit!

USER

SUBREDDIT

User

EXECUTE

Once the analysis is complete, the screen should be populated with visualizations, similar to the screenshots below.



Welcome to Open Social Comparison and Analysis for Reddit!

USER

SUBREDDIT

spacks

EXECUTE

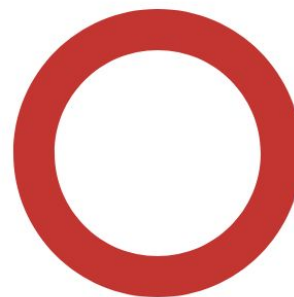
Most Popular Words for spacks

- thread
- posts
- please
- see
- town
- megathread
- get
- moving
- post
- reddit
- people
- covid19
- chili
- new
- live
- got



Most Popular Links for spacks

- shelterforce.org

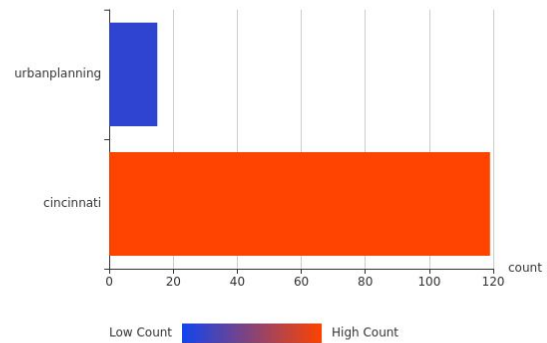


Proportion of Controversial Posts for spacks

- non-controversial
- controversial

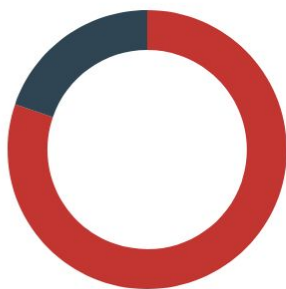


SubReddits that spacks Participates in



spacks's Karma

- Comment Karma
- Link Karma



Evaluation Results

Delay Percentile	Normal
Action Count	Abnormal: Action Count Percentile: 1
Relationship with Closest Neighbors	Normal

Interpreting User Analysis

The output of user analysis includes several visualizations that give information about the submitted user. A brief explanation of each visualization is below.

Most Popular Words

This visualization is a pie chart indicating the words most frequently used in submissions and comments made by a user. It does not include stopwords like “the” and “that.” The sections of the pie chart can be hovered over in order to see the exact percentage and count per word.

Most Popular Links

This visualization is a pie chart indicating the websites most frequently linked to in submissions made by a user. It does not include websites that would convey little information about the user, like Youtube and Twitter. The sections of the pie chart can be hovered over in order to see the exact percentage and count per website.

Proportion of Controversial Posts

This visualization is a pie chart indicating the number of controversial posts vs non-controversial posts made by a user. A controversial post is a post where the number of dislikes is roughly equal to the number of likes, indicating that a significant portion of people that viewed the post did not like its content. This can often indicate that content is offensive or incorrect. The sections of the pie chart can be hovered over in order to see the exact percentage and count per each post type.

Subreddits Participated In

This visualization is a bar chart indicating subreddits that the user frequently posts or comments in. The count is the number of submissions or comments the user has made on that subreddit.

Karma

This visualization is a pie chart indicating a user’s comment karma vs their link karma. Reddit karma is accumulated by a user when they receive likes on things they post. High comment karma means the user frequently receives likes on their comments (and likewise for link karma). This can also indicate if a user is a frequent commenter or link poster. The sections of the pie chart can be hovered over in order to see the exact percentage and count per each karma type.

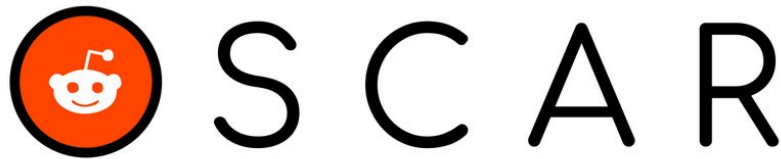
Evaluation Results

This section displays the results of the SocksCatch algorithm if they are available for the user. An “Abnormal” rating in more than one category could be an indicator that this user is a member of a sock puppet organization. Hovering over each category will pull up a tooltip that explains the meaning of the category and what threshold indicates abnormal activity.

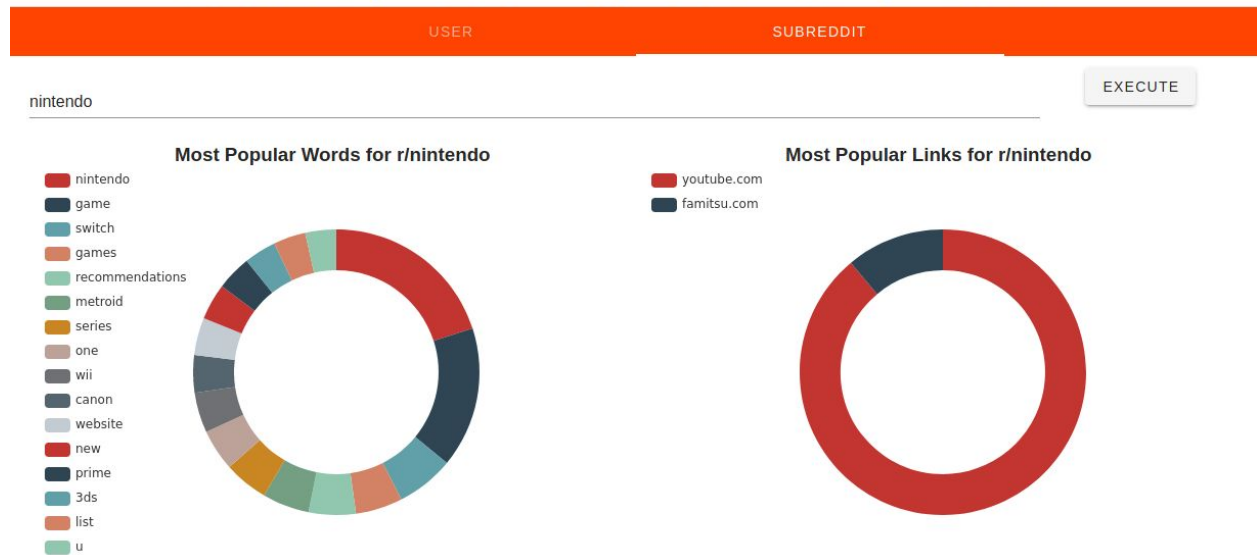
Submitting a Subreddit

Ensure that you are on the Subreddit tab. Use the text box to type in the name of the subreddit that you want to analyze, the same as you would to analyze a user. Click the Execute button to start the analysis.

Once the analysis is complete, the screen should be populated with visualizations, similar to the screenshots below.

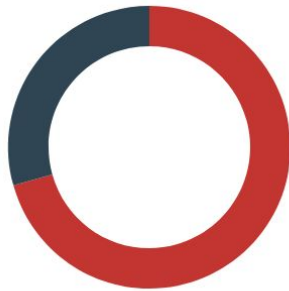


Welcome to Open Social Comparison and Analysis for Reddit!

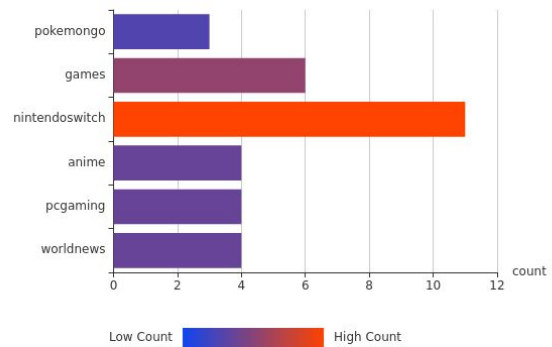


Proportion of Controversial Posts on r/nintendo

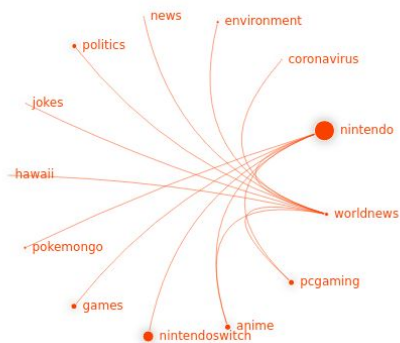
■ non-controversial
■ controversial



SubReddits that Users of r/nintendo Also Post on



Popular SubReddits of r/nintendo Expanded Based on User's Top SubReddits to Post on



Inpreting Subreddit Analysis

The output of subreddit analysis includes several visualizations that give information about the submitted subreddit. A brief explanation of each visualization is below.

Most Popular Words

This visualization is a pie chart indicating the words most frequently used in the top submissions on this subreddit. It does not include stopwords like “the” and “that.” The sections of the pie chart can be hovered over in order to see the exact percentage and count per word.

Most Popular Links

This visualization is a pie chart indicating the websites most frequently linked to in top submissions on this subreddit. It does not include websites that would convey little information about the subreddit, like Youtube and Twitter. The sections of the pie chart can be hovered over in order to see the exact percentage and count per website.

Proportion of Controversial Posts

This visualization is a pie chart indicating the number of controversial posts vs non-controversial posts in top submissions on this subreddit. A controversial post is a post where the number of dislikes is roughly equal to the number of likes, indicating that a significant portion of people that

viewed the post did not like its content. This can often indicate that content is offensive or incorrect. It is important to note that since this analysis is done on top (popular) submissions, the number of controversial posts is often low. The sections of the pie chart can be hovered over in order to see the exact percentage and count per each post type.

Subreddits that Users of this Subreddit also Post on

This visualization is a bar graph indicating the most popular subreddits that top posters on the analyzed subreddit also post on. This indicates that these subreddits share a significant portion of their user base.

Popular Subreddits, Expanded

This visualization is a graph demonstrating the relationships between the analyzed subreddit and other subreddits frequented by users of the analyzed subreddit. This is an expanded view of the previous visualization and includes several more layers of relationships between subreddits. The graph can be zoomed in on and dragged around to get a closer look at its content.

Frequently Asked Questions (FAQ)

Why doesn't the user I submitted for analysis have evaluation results?

Due to the limited amount of computational power available to us when completing this project, we could not scrape Reddit for data on every single layer. We started with a small subset of users to demonstrate the SocksCatch algorithm in action. We built the tools such that over time we could scrape more Reddit data and expand this pool of users until we would eventually support all users. This pool could also expand if our computational capabilities increased.

The subreddit I'm submitting is being rejected for analysis. What's going on?

A subreddit can only be analyzed if it exists on Reddit and is not quarantined. Make sure the spelling of the subreddit name is exactly the same as it appears on Reddit. Additionally, make sure you are not attempting to analyze a subreddit that is quarantined or has other restrictions on its access.

What are quarantined subreddits?

Per Reddit's Help subreddit, a quarantined subreddit is defined as the following:

A community will be Quarantined on Reddit when we deem its content to be extremely offensive or upsetting to the average redditor. The purpose of quarantining a community is to prevent its content from being accidentally viewed by those who do not wish to do so. Restrictions on a quarantined community include:

- *Requiring an account with a verified email address*

- *Requiring an explicit opt-in*
- *No custom images*
- *Will generate no revenue, including ads or Reddit Gold*

Why aren't quarantined subreddits supported for analysis?

We chose not to support the analysis of quarantined subreddits due to them containing potentially offensive or upsetting content. We did not want to appear to promote this content or the communities that create it. Additionally, we did not want users to unexpectedly view offensive or upsetting content when using our product.

Why does this operation take so long?

Many of our processes rely on Reddit's API. In order to maintain compliance with it, our rate is limited to 100 items per ~2 seconds. We are also limited in our processing power due to this being a free product that runs on open source software.