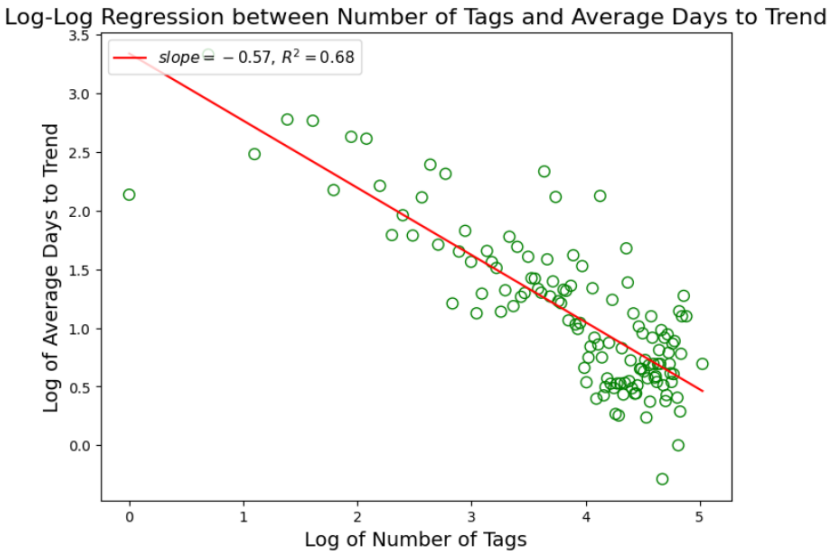
Analyzing Trending YouTube Video's Tags

Dataset: Trending YouTube Video Statistics
https://www.kaggle.com/datasets/datasnaek/youtube-new

1. Scatter Plot

The Impact of Tag Quantity on Trending Speed of Youtube Videos



Legend Explanation

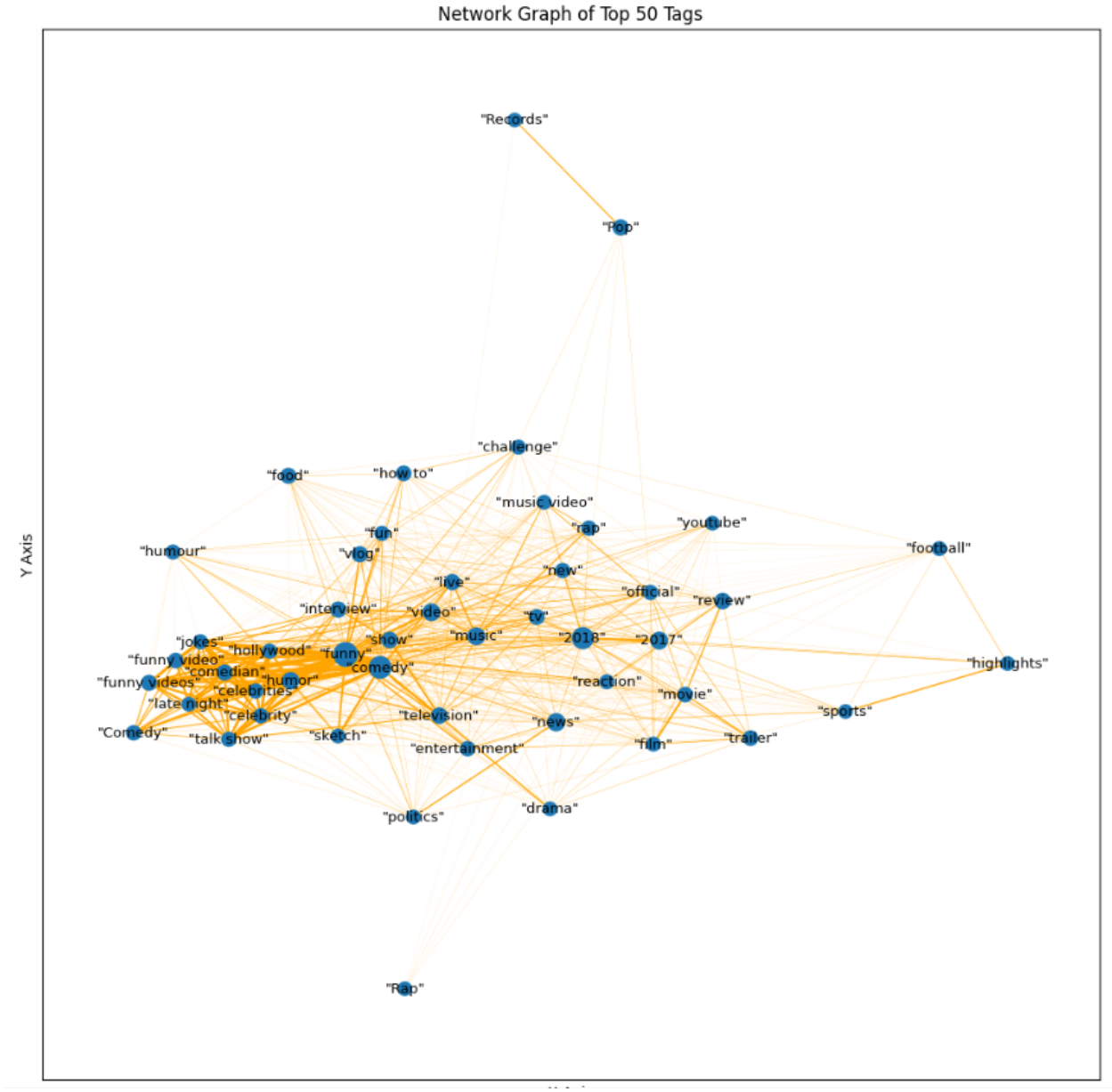
- Green Circles: each shows a video's data point, with tag count and trending days both log-scaled.
- Red Line: the result of a linear regression analysis on the log-transformed data. Its slope (given as -0.57 in the legend) represents the percentage decrease in the average days to trend for every 1% increase in the number of tags.

Highlights

- Negative Correlation: The figure displays a noticeable downward trend, suggesting a significant negative correlation between the number of tags and the average days to trend when both are log-transformed.
- Slope Interpretation: A slope of -0.57 indicates that a 1% increase in the number of tags is associated with an approximate 0.57% decrease in the average days to trend.
- Goodness of Fit: value of 0.68 indicates that approximately 68% of the variability in the log-transformed average days to trend can be explained by changes in the log-transformed number of tags.
- This finding suggests that to make a video popular more quickly, it's best to add several relevant tags.

2. Network Figure

Top 50 Tags Network



Legend Explanation

- Points (Nodes): represent the top 50 tags from the dataset. Their sizes vary based on the frequency of each tag, with larger nodes indicating more frequently used tags.
- Lines (Edges): represent the relationships between the tags. If two tags often appear together in the dataset (more than 30 times), there will be an edge connecting them. The width of the line indicates the strength of the relationship: wider lines mean that the tags are more frequently paired together.

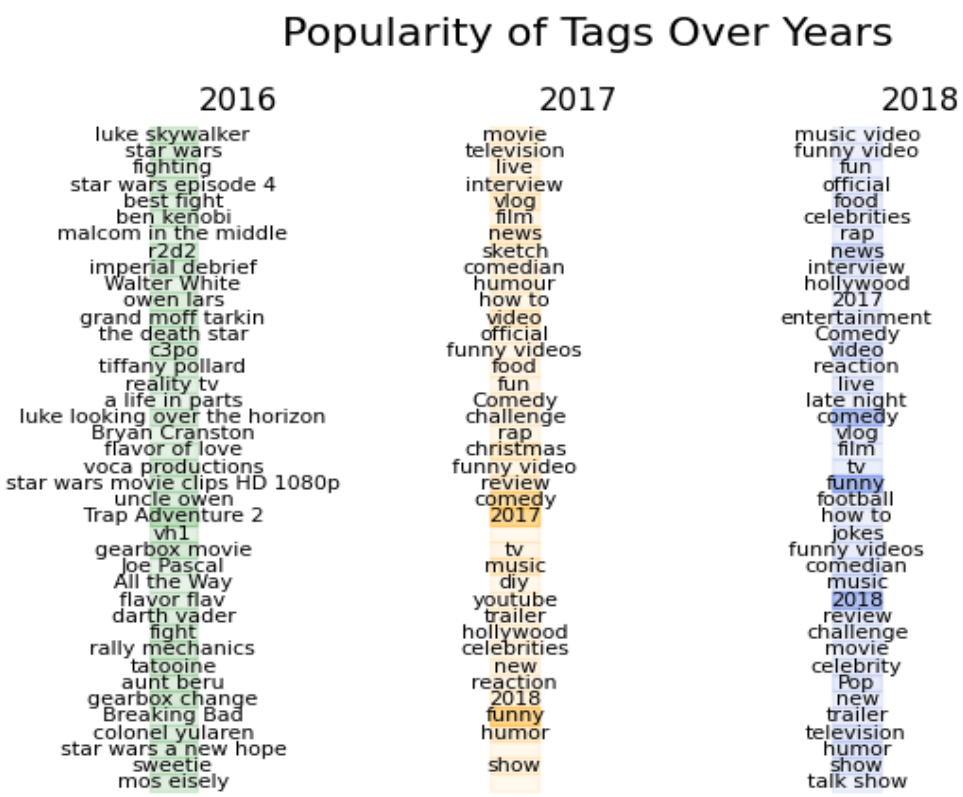
Highlights

- High Connectivity: tags such as "youtube", "music video", "how-to", and "interview" appear to be highly connected, suggesting they are commonly used with many other tags.
- Tag Clusters: There are certain clusters of tags that frequently appear together, which indicates common topics. For instance, the cluster involving "comedy", "funny", "talk show", and "celebrity" indicates content that is humorous and entertainment-focused.

Significance Statement

- To identify the most common-used tags within the dataset.
- Find tag combinations for better tag recommendation.

3. Time-series Word Cloud



Legend Explanation

- Block: each small block represents a tag.
- Bar: each represents a year's popular tags.
- Opacity reflect tag frequency: high frequency tags have deeper color.

Highlights

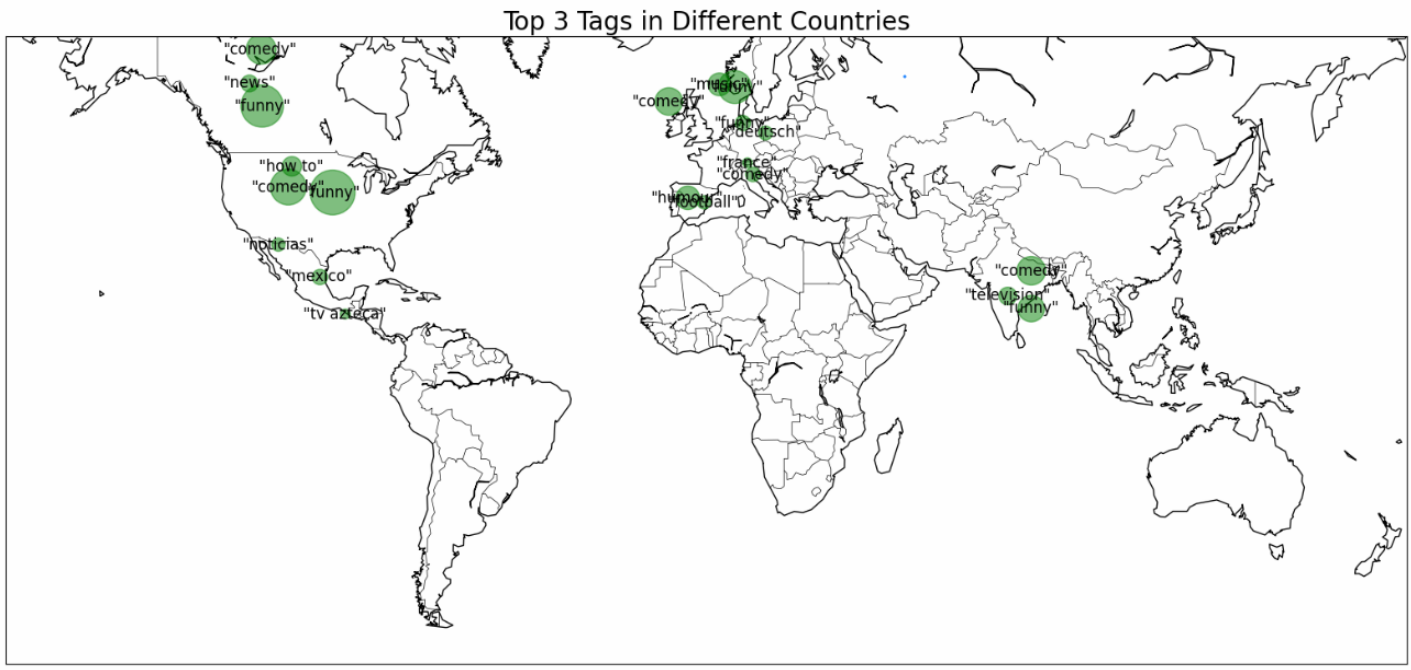
- Tags from 2016 are longer and more dispersed, while the tags from 2017 and 2018 are more general and more concentrated.
- The topics of these 3 year varies: in 2016, tags were predominantly related to movies, especially the "Star Wars" series. The year 2017 saw a diversification of tags, with terms like "movie," "interview," and "vlog" emerging. By 2018, tags related to music and entertainment, such as "music video," "funny," and "football," gained popularity.

Significance Statement

- This figure highlights the shifting trends in social media interests over time, which is crucial for content creators, marketers, and social media strategists. It reveals the movement in public interests, aiding these professionals in better understanding their audience and tailoring their strategies accordingly.

4. Map

Top 3 Tags in Different Countries



Legend Explanation

- Green circles: Size corresponds to the frequency of the top tag within the country.
- Text labels: Represent the top 3 social media tags in each country.

Findings:

- "comedy" and "funny" are prevalent tags in multiple regions, indicating a universal appeal for humor.
- Specific regional interests are evident, such as "news" in North America.
- Europe shows a diverse range of top tags, reflecting a mix of entertainment and informational content preferences. Also, "football" is among the top 3 tags of Europe.

Data and Method

- Data: used a map<region, map<tag, count>> to store tags and their frequency from different regions.
- Method: used mpl_toolkits.basemap to draw the world map. Manually set the cordinates of different regions, and then get the top 3 tags from the processed data. (It's easy to get more tags to make this figure more informative, however, due to limited size of this figure, I chose top 3.) Then, I used randomly slightly changed cordinates to display these 3 tags.

Significance Statement

Understanding the popularity of social media tags across different countries offers valuable insights into cultural trends, interests, and the global digital landscape. This figure highlights the commonalities and differences in social media usage around the world, which is essential for content creators, marketers, and researchers studying global digital communication patterns.