

# Meme Propagation Through Reddit

James Chen  
Hoik Jang  
Steven Oliver  
Adrian Perez

November 27, 2018



## 1 Background

- Reddit
- Summary Statistics

## 2 Objectives

- Motivation
- Objectives

## 3 Results

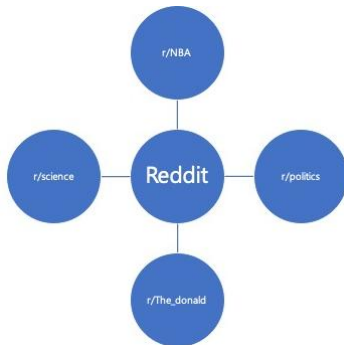
- Comment Activity
- Subscriber and Comment Counts
- Danielle Meme
- The Killer Plot

## 4 Conclusion



# Reddit: "The Front Page of the Internet"

Reddit operates as a collection of over a million communities known as *subreddits*. These are message boards oriented around a topic of interest such as:



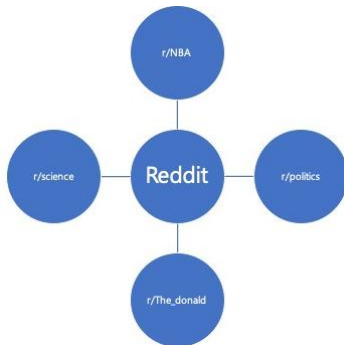
Users interact within the subreddits by:

- posting content (images, gifs, or text discussion)
- commenting on posts
- voting



# Reddit: "The Front Page of the Internet"

Reddit operates as a collection of over a million communities known as *subreddits*. These are message boards oriented around a topic of interest such as:



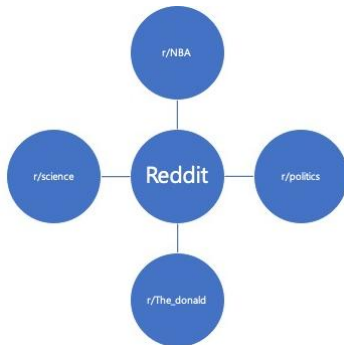
Users interact within the subreddits by:

- posting content (images, gifs, or text discussion)
- commenting on posts
- voting



# Reddit: "The Front Page of the Internet"

Reddit operates as a collection of over a million communities known as *subreddits*. These are message boards oriented around a topic of interest such as:



Users interact within the subreddits by:

- posting content (images, gifs, or text discussion)
- commenting on posts
- voting



# Summary Statistics

- Data size: 6 months of data. On average, each month contains 70 Million Rows with 12 columns.
- File type: JSON
- Process: Convert JSON in to SQL Database, then create subtables tables
- Established server



# Summary Statistics

- Data size: 6 months of data. On average, each month contains 70 Million Rows with 12 columns.
- File type: JSON
- Process: Convert JSON in to SQL Database, then create subtables tables
- Established server



# Summary Statistics

- Data size: 6 months of data. On average, each month contains 70 Million Rows with 12 columns.
- File type: JSON
- Process: Convert JSON in to SQL Database, then create subtables tables
- Established server





# Summary Statistics

- Data size: 6 months of data. On average, each month contains 70 Million Rows with 12 columns.
- File type: JSON
- Process: Convert JSON in to SQL Database, then create subtables tables
- Established server



Reddit is one of the largest communities on the internet for new, revolutionary, and influential purposes. User interaction within popular subreddits can have a great influence across the communities. Due to the richness and size of the dataset, we have proposed the following objectives:



# Objectives

## Objective 1

Investigate relationships of the top 100 most active subreddits by tracking the spread of memes across those communities.

## Objective 2

Analyze patterns of user comment activity in a multi-month span on Reddit.

## Objective 3

Analyze the fluctuations of user activity across subreddits by time of day.



# Objectives

## Objective 1

Investigate relationships of the top 100 most active subreddits by tracking the spread of memes across those communities.

## Objective 2

Analyze patterns of user comment activity in a multi-month span on Reddit.

## Objective 3

Analyze the fluctuations of user activity across subreddits by time of day.



# Objectives

## Objective 1

Investigate relationships of the top 100 most active subreddits by tracking the spread of memes across those communities.

## Objective 2

Analyze patterns of user comment activity in a multi-month span on Reddit.

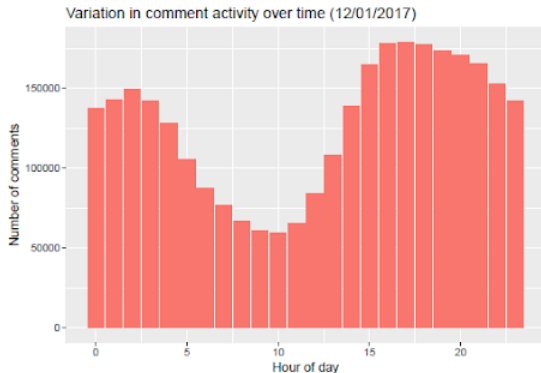
## Objective 3

Analyze the fluctuations of user activity across subreddits by time of day.



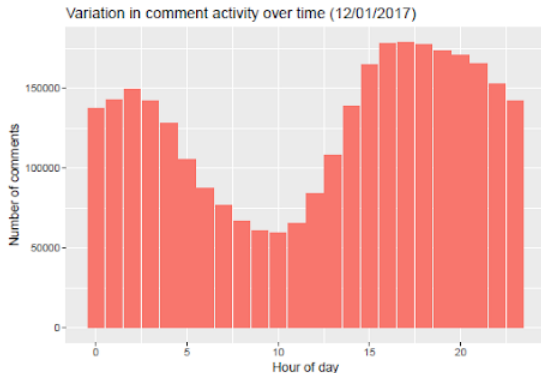
# Comment Activity

- Time is in UTC (+6:00 CST)
- Users are from the US
- Peak activity during Lunch and 9pm
- Expected more night activity, can be due to young use age



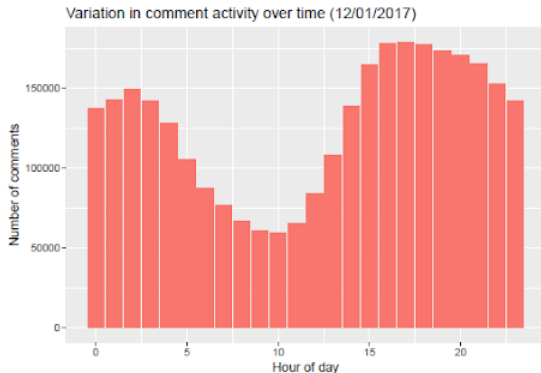
# Comment Activity

- Time is in UTC (+6:00 CST)
- Users are from the US
- Peak activity during Lunch and 9pm
- Expected more night activity, can be due to young use age



# Comment Activity

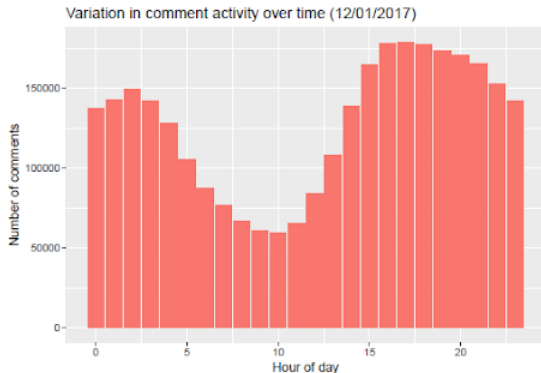
- Time is in UTC (+6:00 CST)
- Users are from the US
- Peak activity during Lunch and 9pm
- Expected more night activity, can be due to young use age





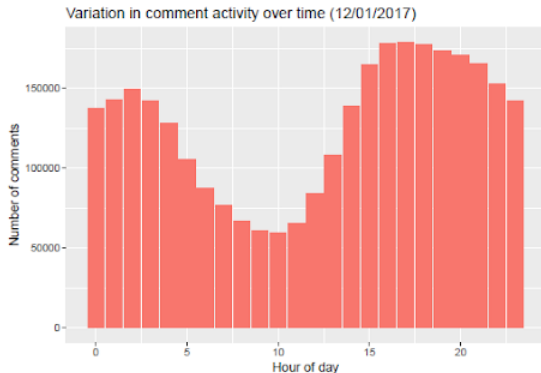
# Comment Activity

- Time is in UTC (+6:00 CST)
- Users are from the US
- Peak activity during Lunch and 9pm
- Expected more night activity, can be due to young use age



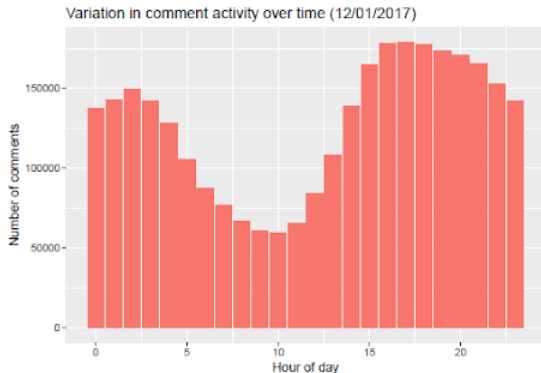
# Comment Activity

- Time is in UTC (+6:00 CST)
- Users are from the US
- Peak activity during Lunch and 9pm
- Expected more night activity, can be due to young use age



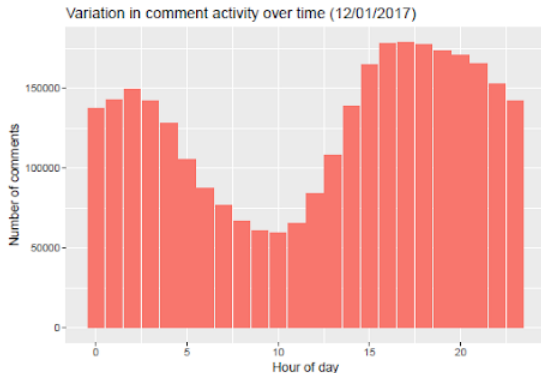
# Comment Activity

- Time is in UTC (+6:00 CST)
- Users are from the US
- Peak activity during Lunch and 9pm
- Expected more night activity, can be due to young use age



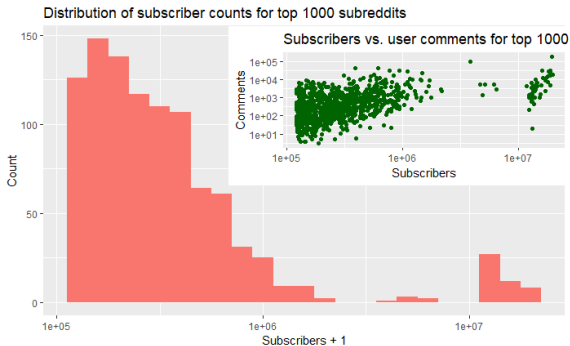
# Comment Activity

- Time is in UTC (+6:00 CST)
- Users are from the US
- Peak activity during Lunch and 9pm
- Expected more night activity, can be due to young use age



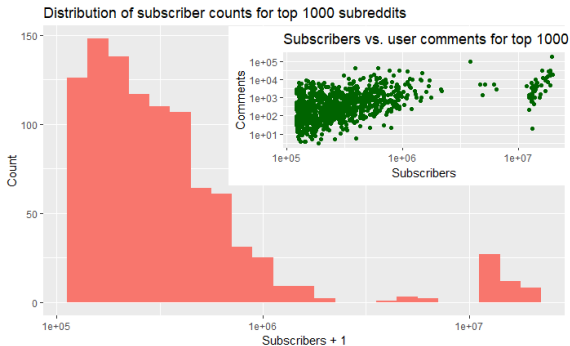
# Subscriber and Comment Counts

- Expected decay with top subreddits
- Anomaly is explained by auto subscriptions
- Breaks linear regression pattern as well



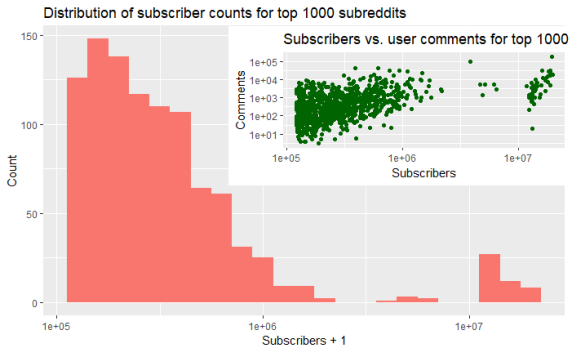
# Subscriber and Comment Counts

- Expected decay with top subreddits
- Anomaly is explained by auto subscriptions
- Breaks linear regression pattern as well



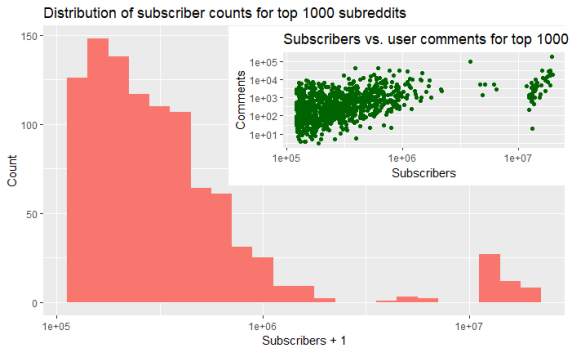
# Subscriber and Comment Counts

- Expected decay with top subreddits
- Anomaly is explained by auto subscriptions
- Breaks linear regression pattern as well



# Subscriber and Comment Counts

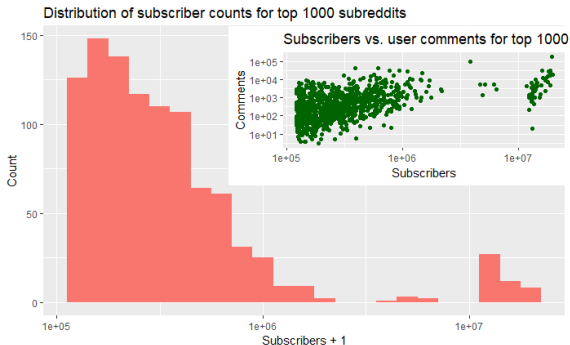
- Expected decay with top subreddits
- Anomaly is explained by auto subscriptions
- Breaks linear regression pattern as well





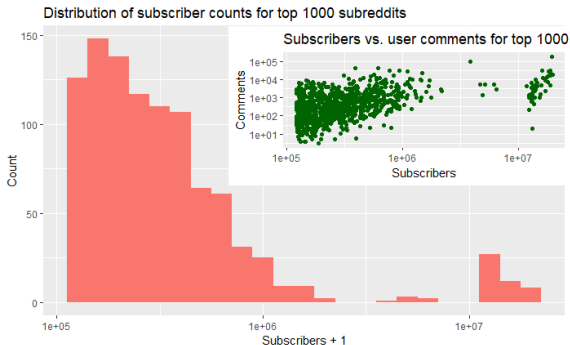
# Subscriber and Comment Counts

- Expected decay with top subreddits
- Anomaly is explained by auto subscriptions
- Breaks linear regression pattern as well



# Subscriber and Comment Counts

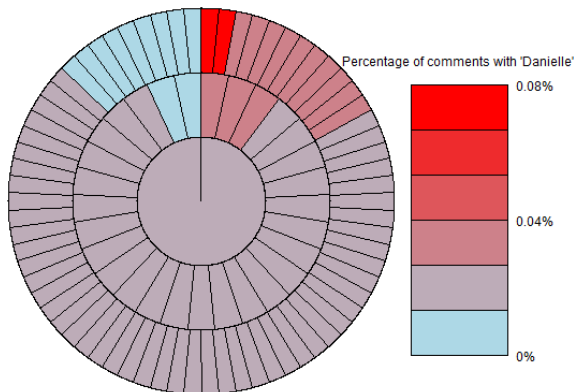
- Expected decay with top subreddits
- Anomaly is explained by auto subscriptions
- Breaks linear regression pattern as well





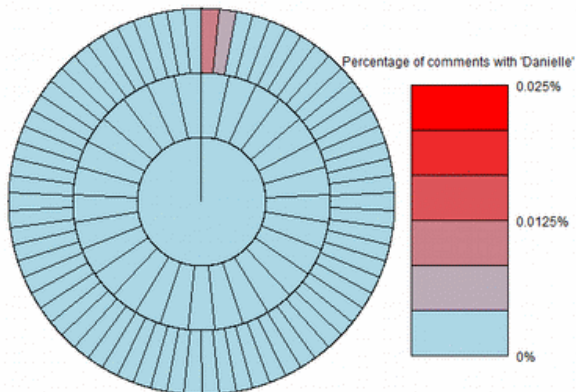
# The Killer Plot

Plot of frequency of 'Danielle' for each subreddit  
January 2017 (stratified by closeness to r/AskReddit)



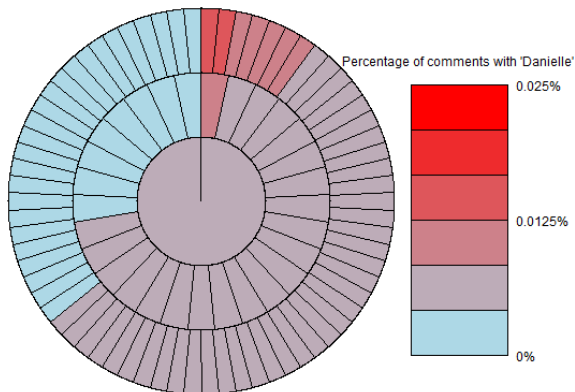
# The Killer Plot

Plot of frequency of 'Danielle' for each subreddit  
January 2017 Week 1 (stratified by closeness to r/AskReddit)



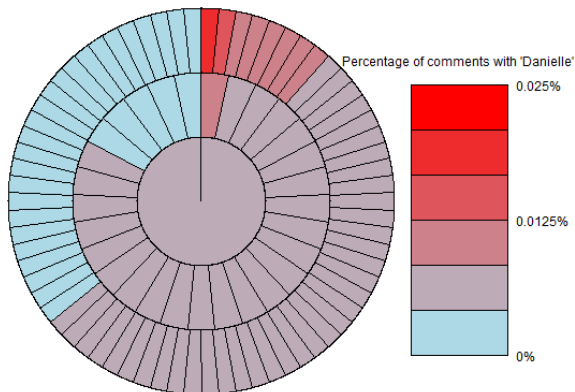
# The Killer Plot

Plot of frequency of 'Danielle' for each subreddit  
January 2017 Week 2 (stratified by closeness to r/AskReddit)



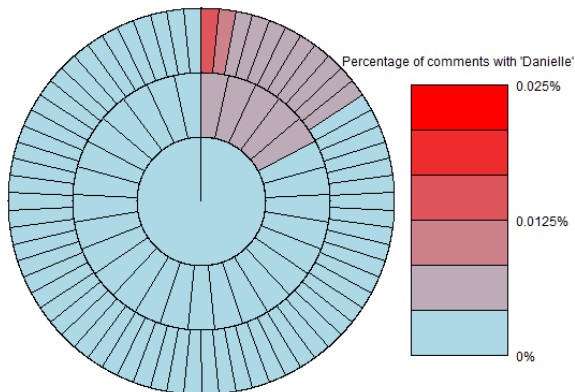
# The Killer Plot

Plot of frequency of 'Danielle' for each subreddit  
January 2017 Week 3 (stratified by closeness to r/AskReddit)



# The Killer Plot

Plot of frequency of 'Danielle' for each subreddit  
January 2017 Week 4 (stratified by closeness to r/AskReddit)





# The Killer Plot

- Center is r/AskReddit, largest subreddit, and one of the first that Danielle meme appeared in
- Inner ring is 29 subreddits most closely related to r/AskReddit, based on cross-commenting activity
- Outer ring is 70 other subreddits less closely related to r/AskReddit
- Color scale is saturation percent of comments (percentage of comments containing the meme)



# The Killer Plot

- Center is r/AskReddit, largest subreddit, and one of the first that Danielle meme appeared in
- Inner ring is 29 subreddits most closely related to r/AskReddit, based on cross-commenting activity
- Outer ring is 70 other subreddits less closely related to r/AskReddit
- Color scale is saturation percent of comments (percentage of comments containing the meme)



# The Killer Plot

- Center is r/AskReddit, largest subreddit, and one of the first that Danielle meme appeared in
- Inner ring is 29 subreddits most closely related to r/AskReddit, based on cross-commenting activity
- Outer ring is 70 other subreddits less closely related to r/AskReddit
- Color scale is saturation percent of comments (percentage of comments containing the meme)



# The Killer Plot

- Center is r/AskReddit, largest subreddit, and one of the first that Danielle meme appeared in
- Inner ring is 29 subreddits most closely related to r/AskReddit, based on cross-commenting activity
- Outer ring is 70 other subreddits less closely related to r/AskReddit
- Color scale is saturation percent of comments (percentage of comments containing the meme)



## Takeaways:

- Meme started in week 1, grew in weeks 2 and 3. and started dying off in week 4
- Two subreddits with most saturation of meme where in least correlation subreddits
  - r/MMA
  - r/hiphopheads
- r/MMA might be the real center of the meme's spread

## Future directions:

- Should check with r/MMA but SQL queries take a while to run!
- Try this with more meme in other months (convfefe, Nothing but Respect for MY President).



## Takeaways:

- Meme started in week 1, grew in weeks 2 and 3. and started dying off in week 4
- Two subreddits with most saturation of meme where in least correlation subreddits
  - r/MMA
  - r/hiphopheads
- r/MMA might be the real center of the meme's spread

## Future directions:

- Should check with r/MMA but SQL queries take a while to run!
- Try this with more meme in other months (convfefe, Nothing but Respect for MY President).



## Takeaways:

- Meme started in week 1, grew in weeks 2 and 3. and started dying off in week 4
- Two subreddits with most saturation of meme where in least correlation subreddits
  - r/MMA
  - r/hiphopheads
- r/MMA might be the real center of the meme's spread

## Future directions:

- Should check with r/MMA but SQL queries take a while to run!
- Try this with more meme in other months (convfefe, Nothing but Respect for MY President).



## Takeaways:

- Meme started in week 1, grew in weeks 2 and 3. and started dying off in week 4
- Two subreddits with most saturation of meme where in least correlation subreddits
  - r/MMA
  - r/hiphopheads
- r/MMA might be the real center of the meme's spread

## Future directions:

- Should check with r/MMA but SQL queries take a while to run!
- Try this with more meme in other months (convfefe, Nothing but Respect for MY President).





## Takeaways:

- Meme started in week 1, grew in weeks 2 and 3. and started dying off in week 4
- Two subreddits with most saturation of meme where in least correlation subreddits
  - r/MMA
  - r/hiphopheads
- r/MMA might be the real center of the meme's spread

## Future directions:

- Should check with r/MMA but SQL queries take a while to run!
- Try this with more meme in other months (convfefe, Nothing but Respect for MY President).



## Takeaways:

- Meme started in week 1, grew in weeks 2 and 3. and started dying off in week 4
- Two subreddits with most saturation of meme where in least correlation subreddits
  - r/MMA
  - r/hiphopheads
- r/MMA might be the real center of the meme's spread

## Future directions:

- Should check with r/MMA but SQL queries take a while to run!
- Try this with more meme in other months (convfefe, Nothing but Respect for MY President).



## Takeaways:

- Meme started in week 1, grew in weeks 2 and 3. and started dying off in week 4
- Two subreddits with most saturation of meme where in least correlation subreddits
  - r/MMA
  - r/hiphopheads
- r/MMA might be the real center of the meme's spread

## Future directions:

- Should check with r/MMA but SQL queries take a while to run!
- Try this with more meme in other months (convfefe, Nothing but Respect for MY President).

