

# Are All Successful Communities Alike?

## Characterizing and Predicting the Success of Online Communities



Tiago Cunha  
tiolivei@umich.edu  
@tocunha



David Jurgens  
jurgens@umich.edu  
@david\_\_jurgens



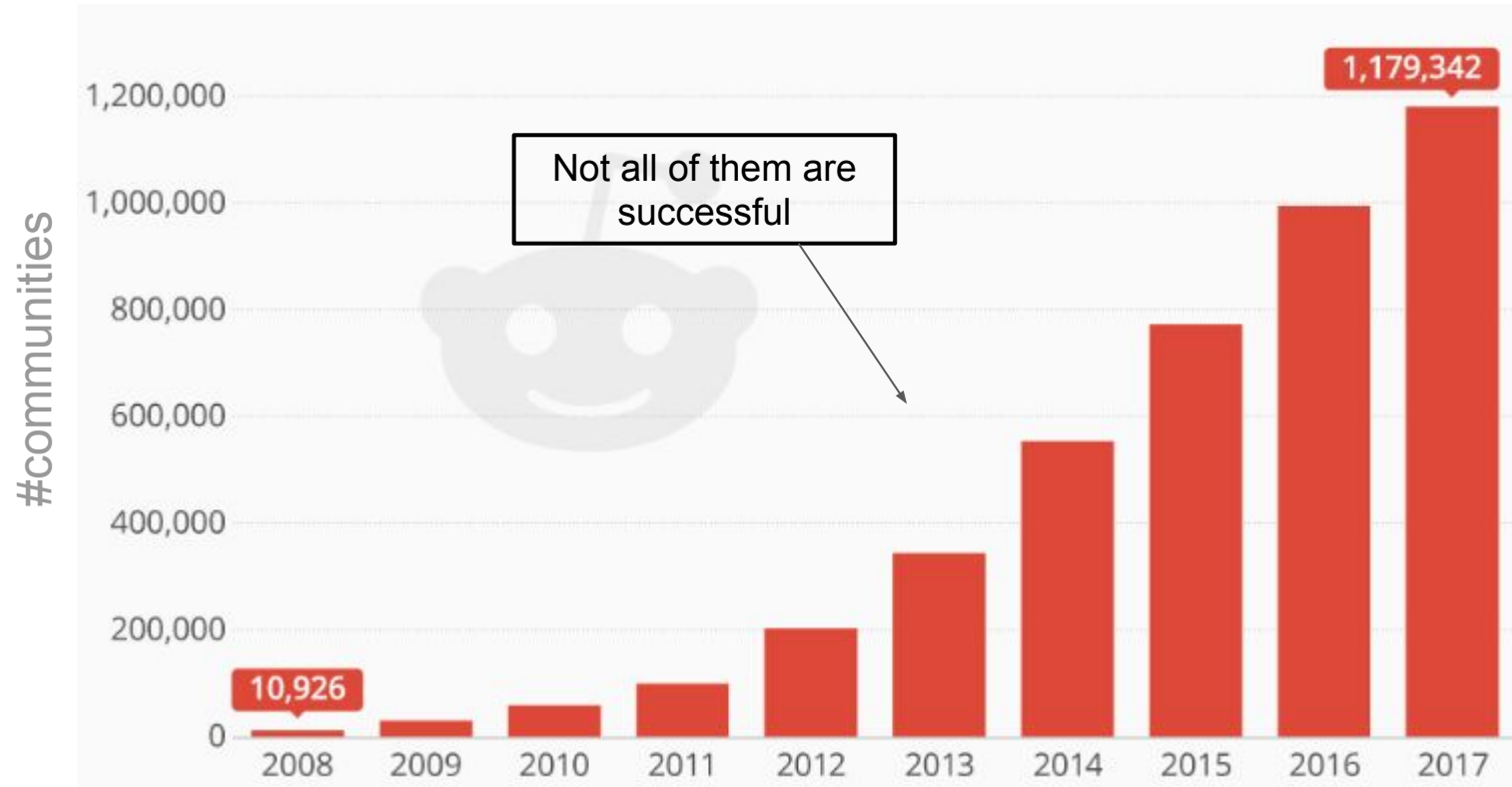
Chenhao Tan  
chenhao.tan@colorado.edu  
@ChenhaoTan



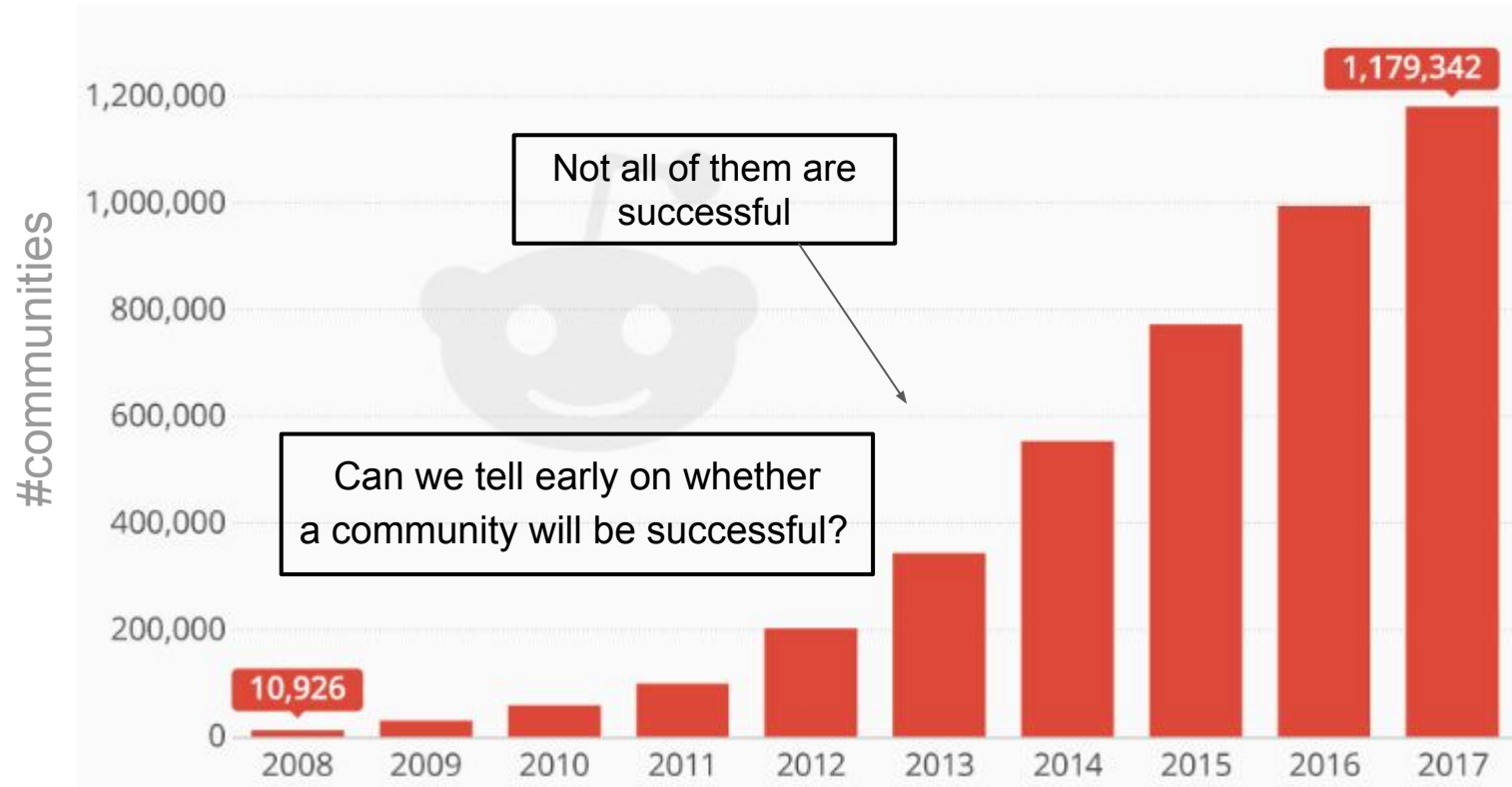
Daniel Romero  
drom@umich.edu  
@DanielMRomero



# The explosion of the number of online communities



# The explosion of the number of online communities



# What is community success?

- Existing research has typically considered a single measure of success.
- Growth in the number of members the community.
- Considering a single measure of success has limited our understanding on what makes communities successful.

# How can we measure success?

Grad students



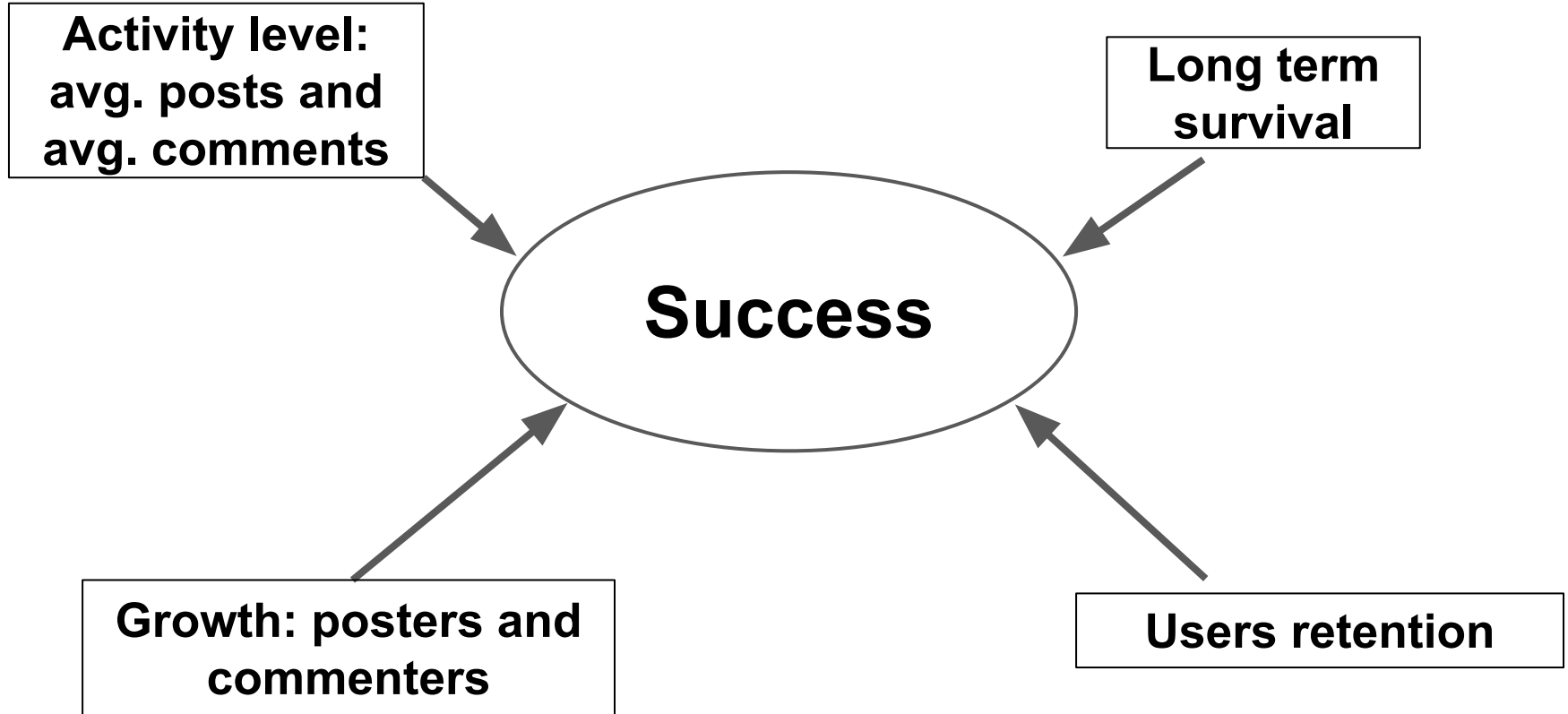
X

Global warming awareness



Our work aims to fill this gap by exploring a variety of success measures of newly created communities on Reddit.

# Defining Community Success



# Defining Community Success

- Growth: number of members after one year.
- Users retention: average monthly retention rate in the first year.
- Survival: if the community is still active after two years.
- Volume of activities: average number of posts and comments after one year.

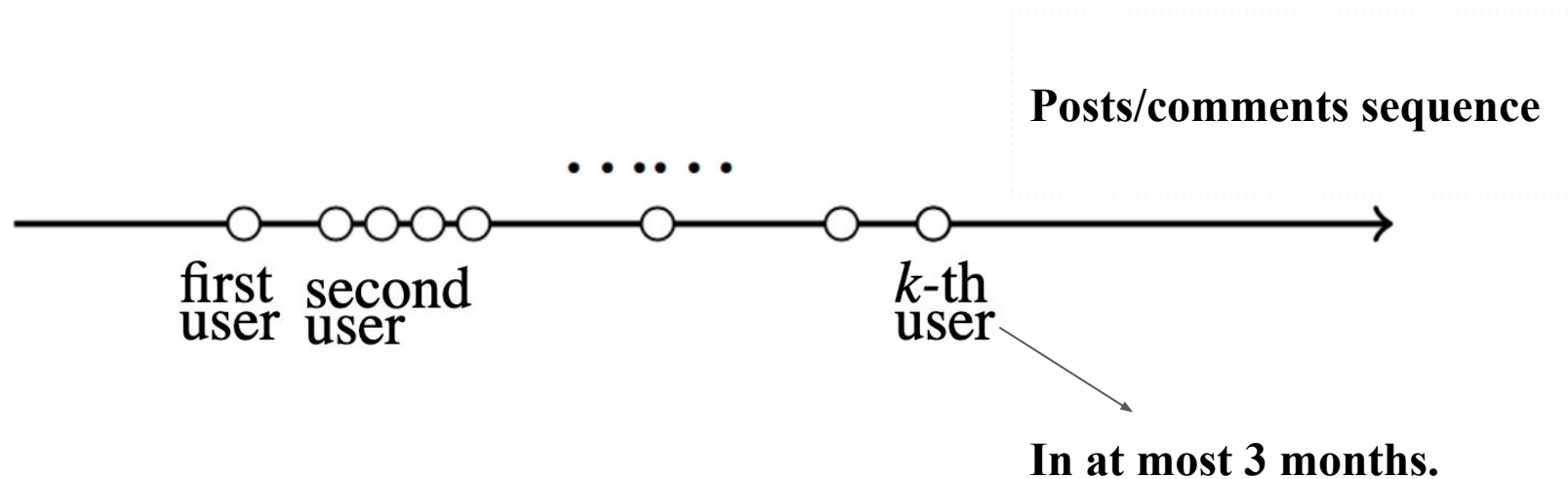


# Building dataset



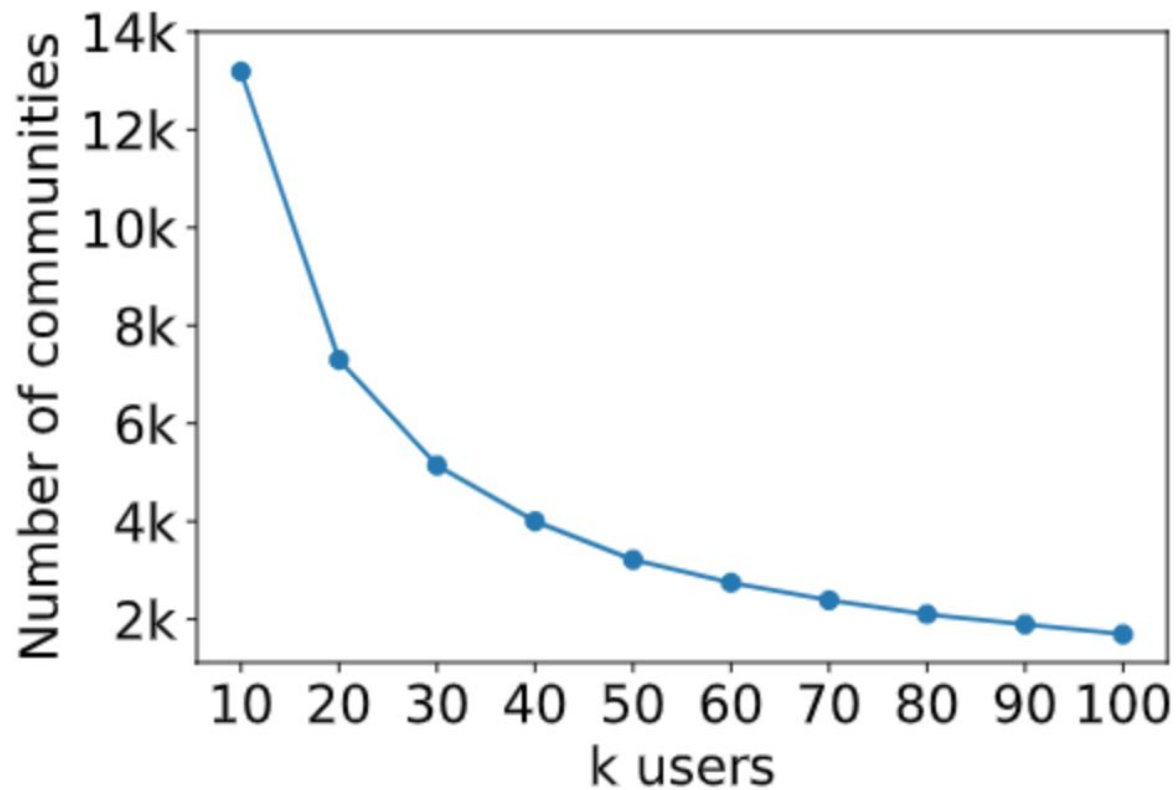
- We obtained all posts and comments from subreddits created in 2014.
- Focus on early members of the communities.

# Focus on the early members of a community



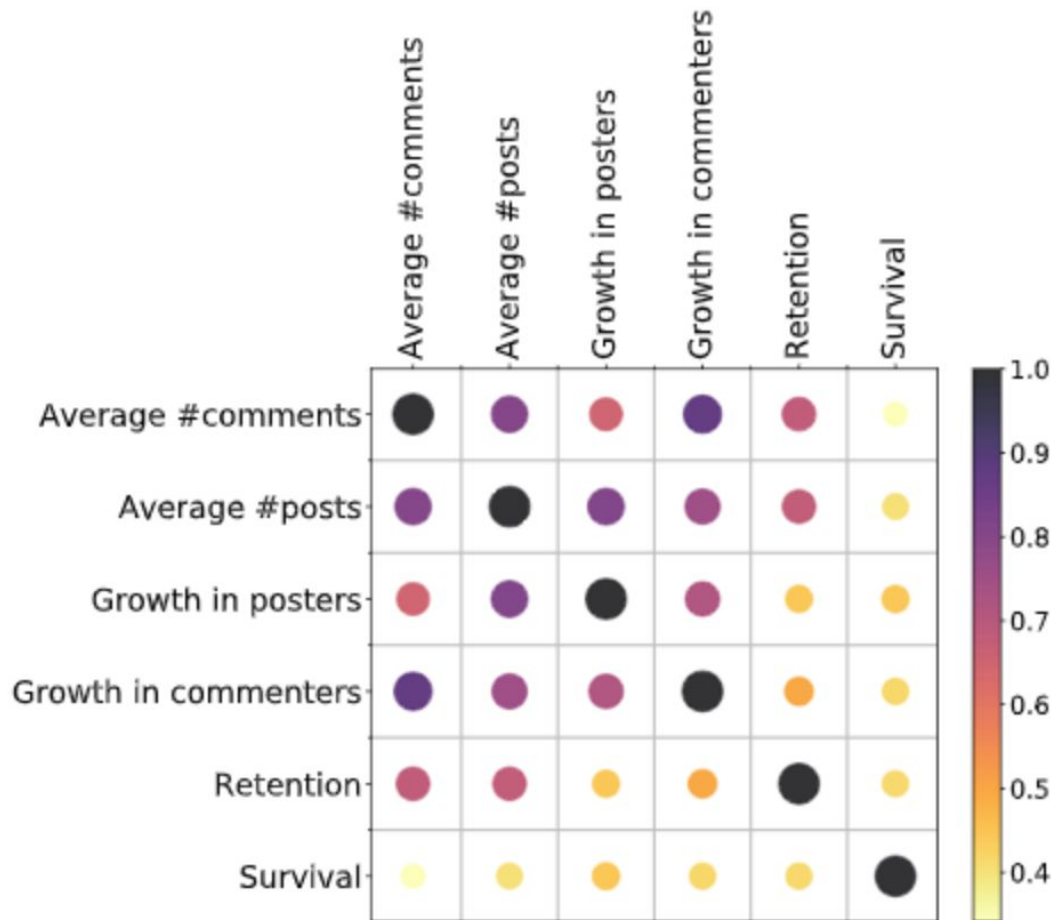
**K varying from 10 to 100.**

# Dataset

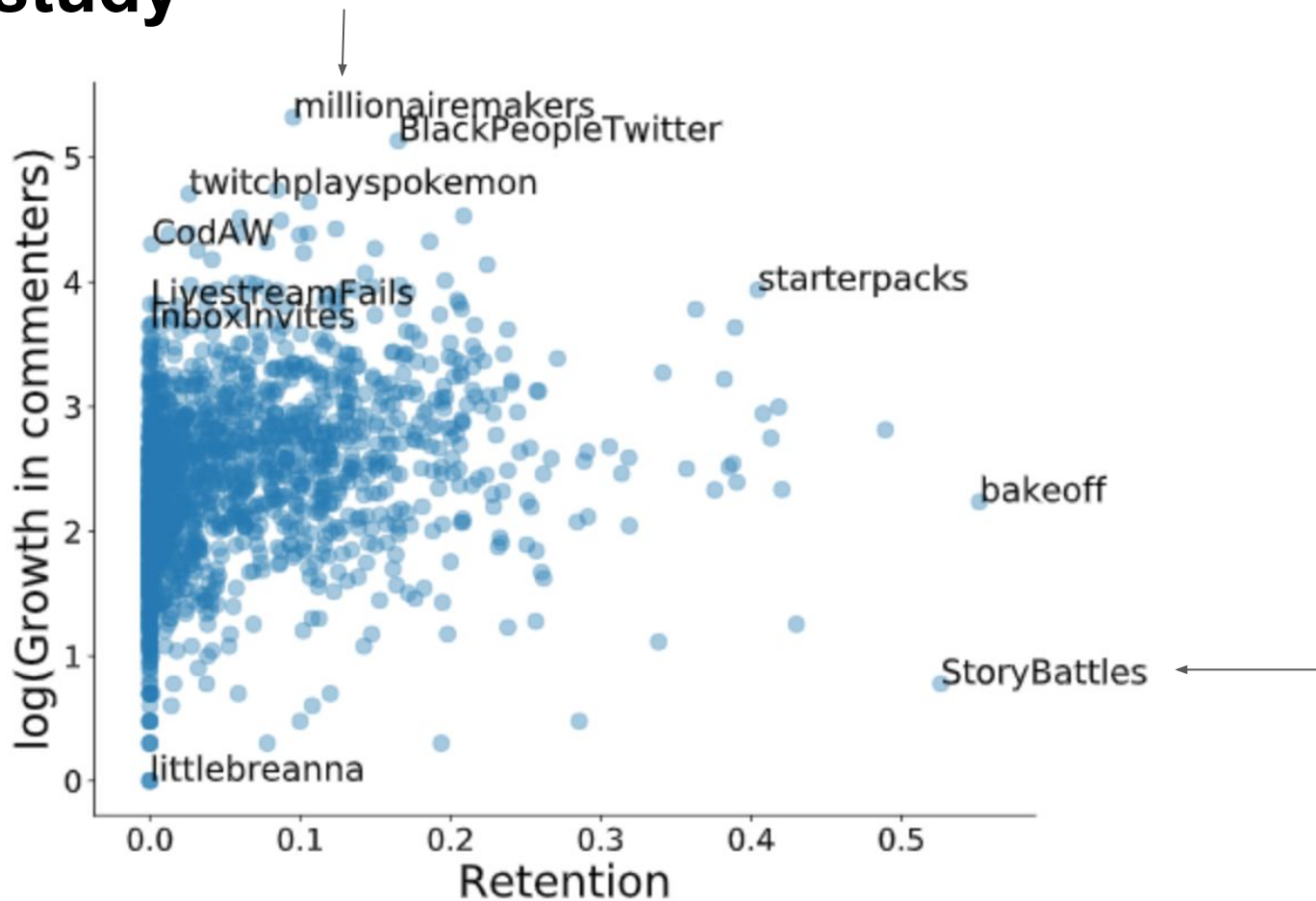


**What is the relationship between the different success measures?**

# Correlation between Success Metrics



# Case study



# **Can Success Be Predicted?**

# Predictive features of success

- Inspired by the literature we defined 6 sets:
- Volume and speed of activities.
- Distribution of activities
- User composition
- Linguistic style
- Social networks
- Parents communities



# Volume and speed of activities

- Number of posters
- Number of commenters
- The date the community was created
- Number of posts
- ...

# Distribution of activities

- How the content produced in the community is spread among users and over time.
- We capture this spread through the Gini coefficient.

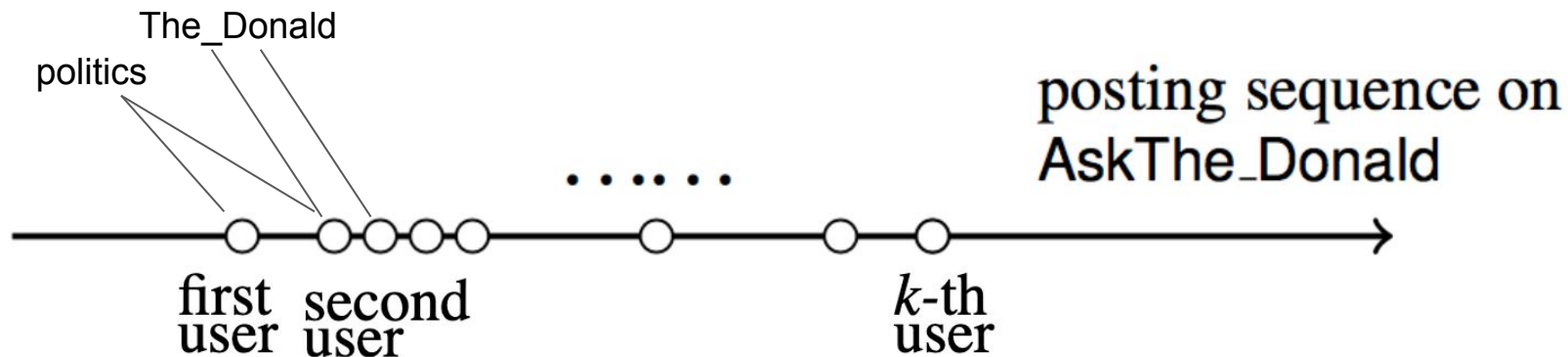
# Linguistic style

- Capture how the language in the content created in the communities can help understand the desirable characteristics of a community.

# Social networks

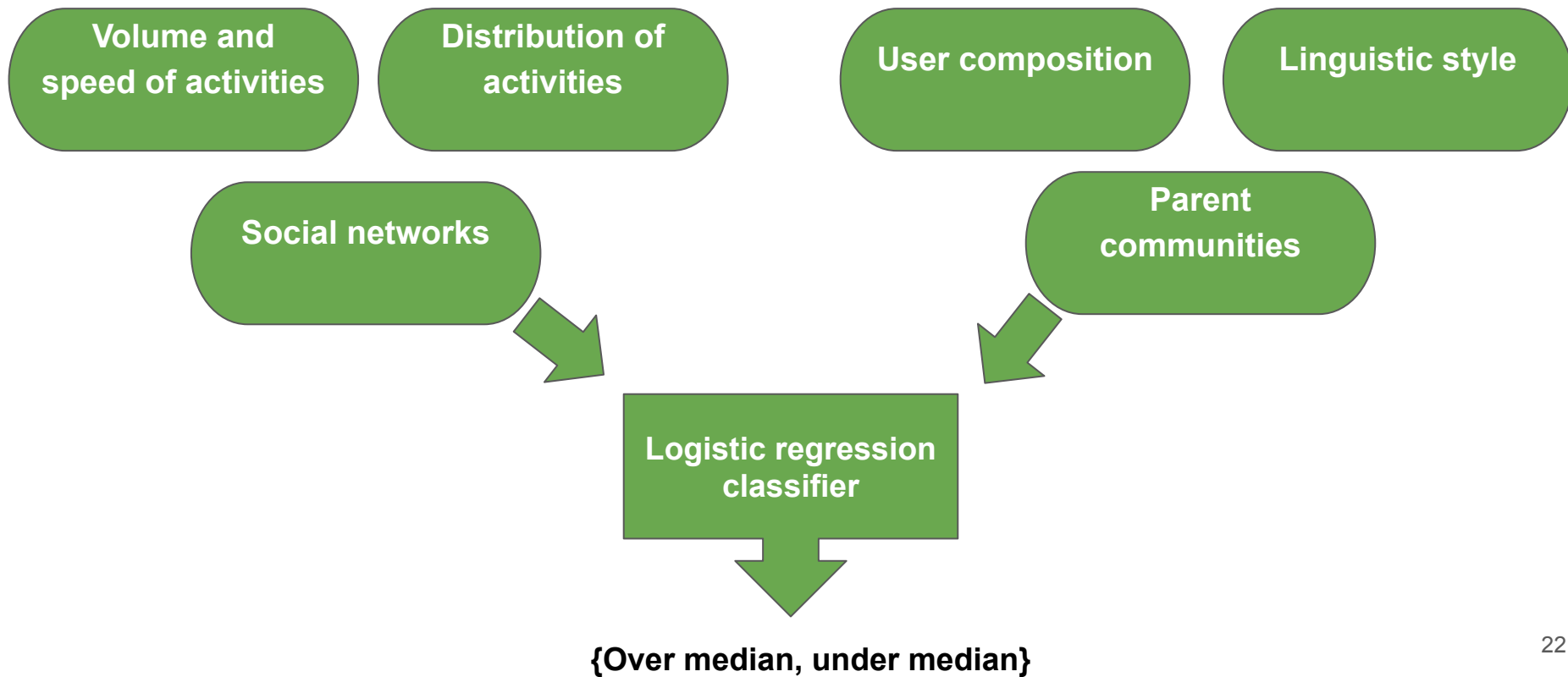
- We construct a communication network among early members and extract a variety of features that describe its structure.
- Transitivity
- Average clustering coefficient
- density
- ...

# Parent communities

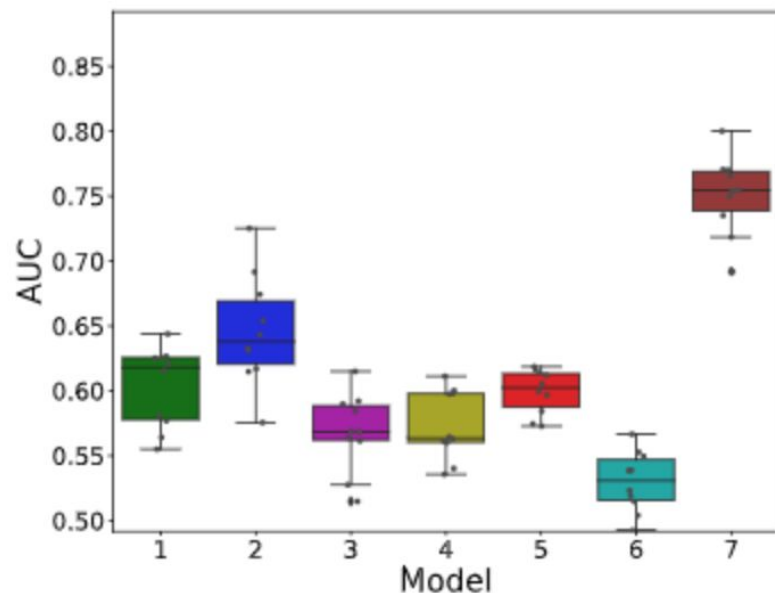


Subreddits where users  
were recently active in

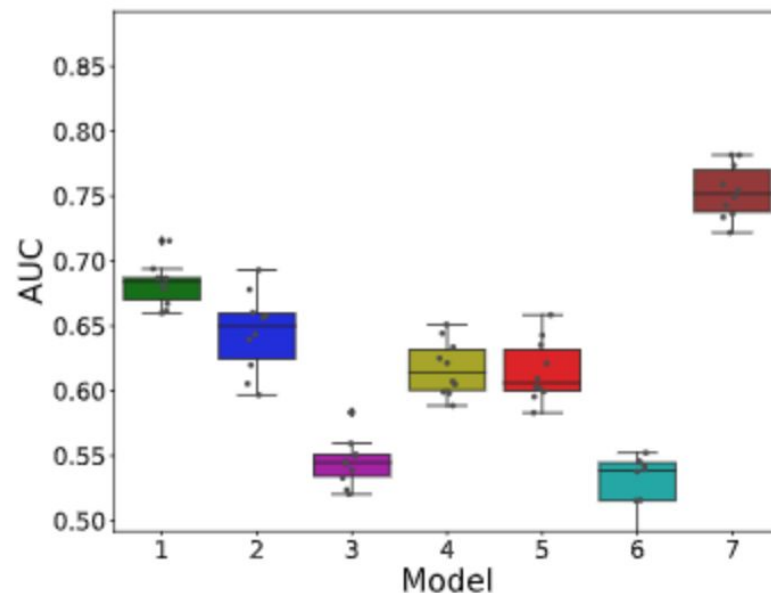
# Building a classifier



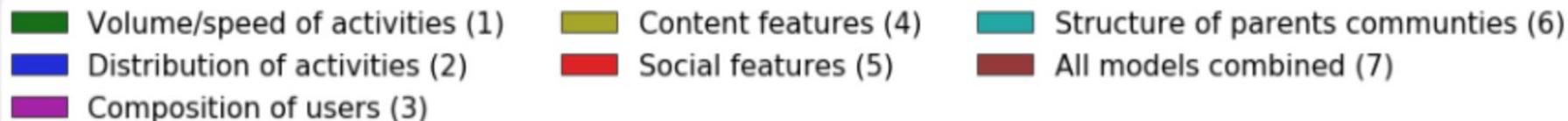
# Classifier performance



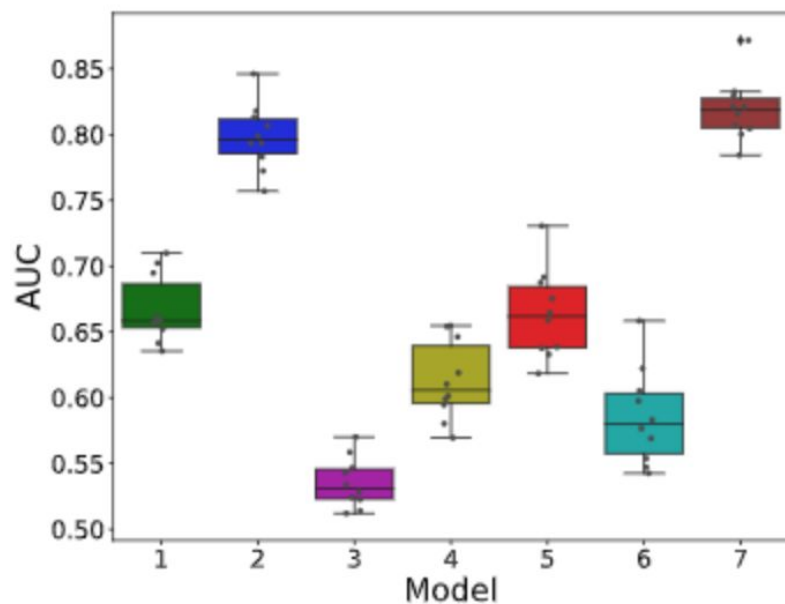
(a) Growth in commenters



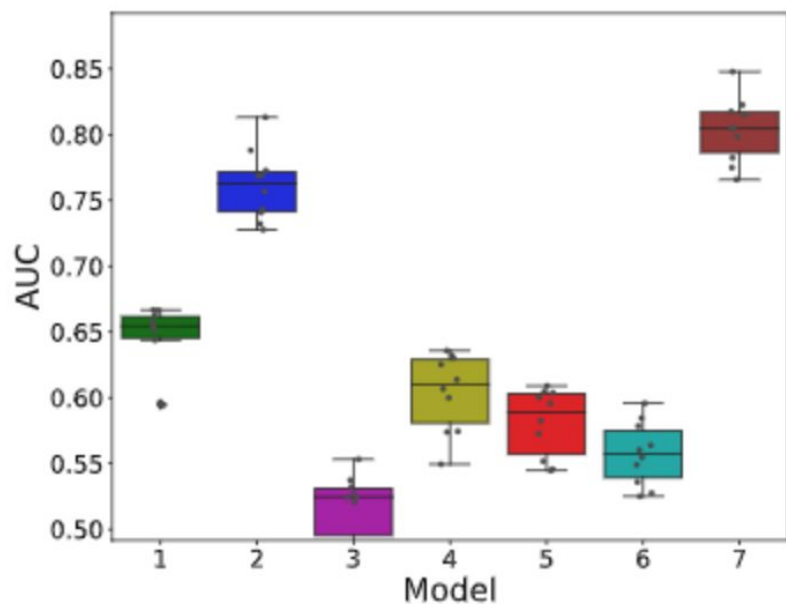
(b) Growth in posters



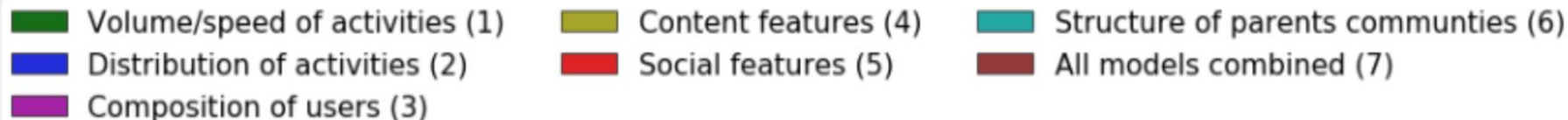
# Classifier performance



**(e) Average #comments**

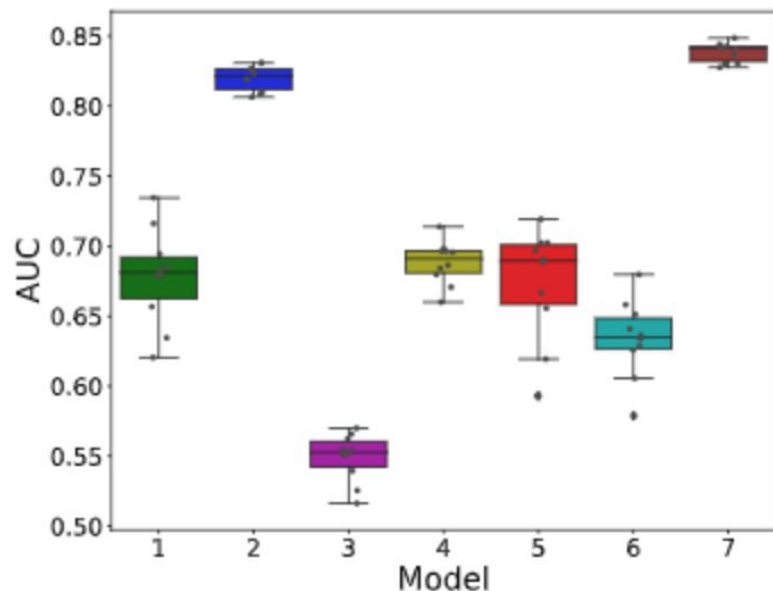


**(f) Average #posts**

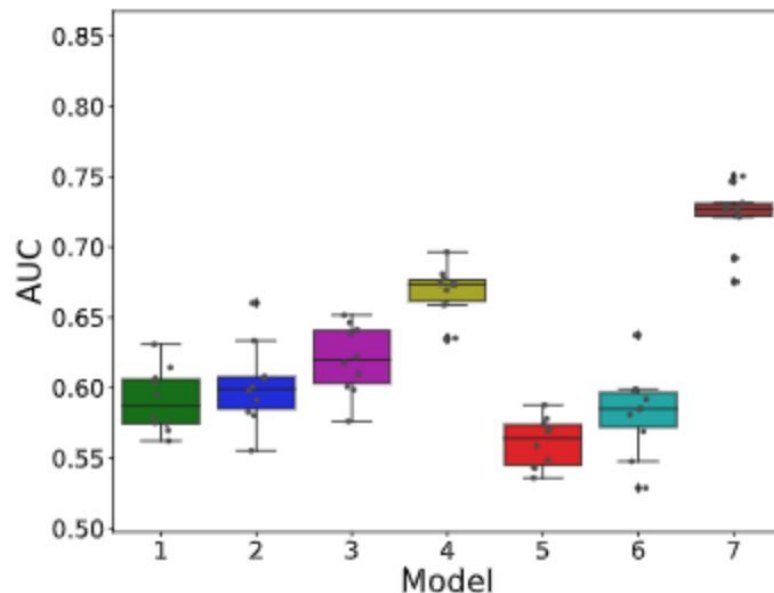




# Classifier performance



(c) Retention



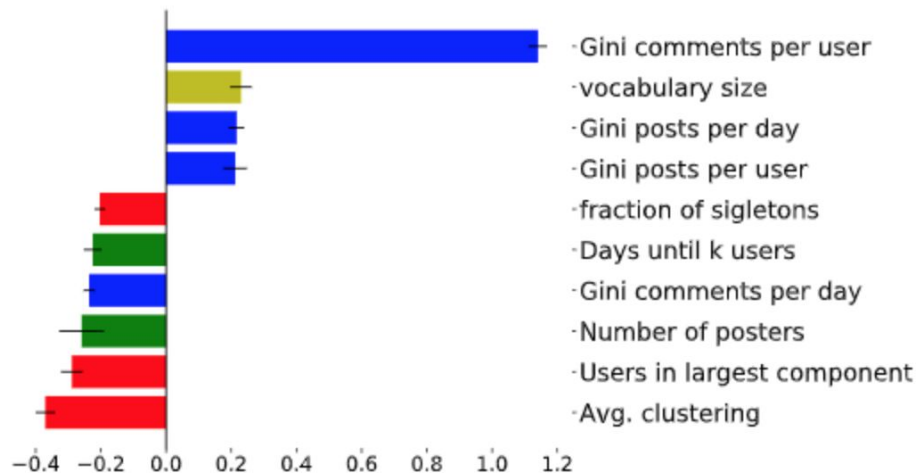
(d) Survival

Volume/speed of activities (1)  
Distribution of activities (2)  
Composition of users (3)

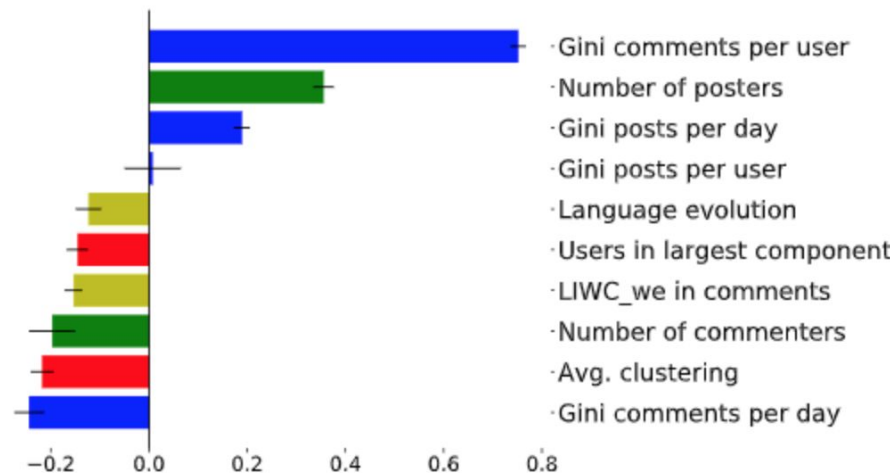
Content features (4)  
Social features (5)

Structure of parents communities (6)  
All models combined (7)

# Which features predict success?

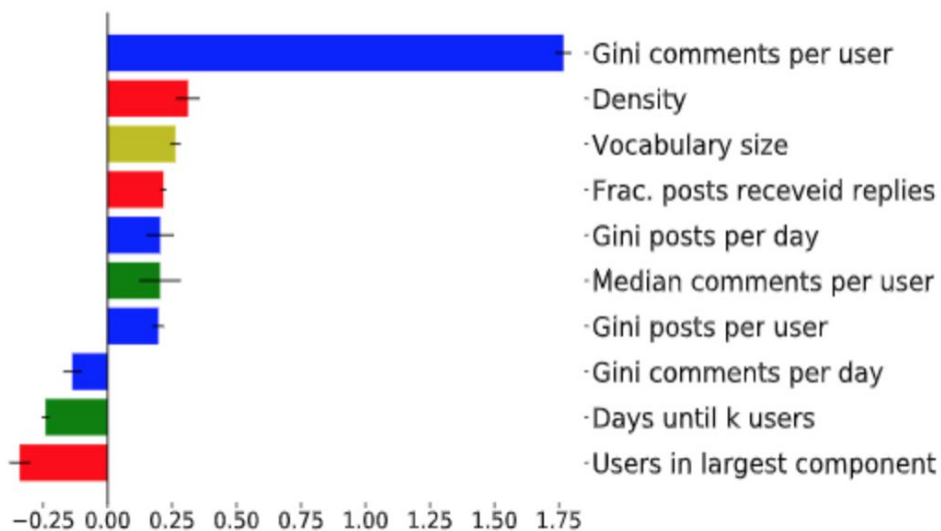


Top 10 features for growth in commenters.

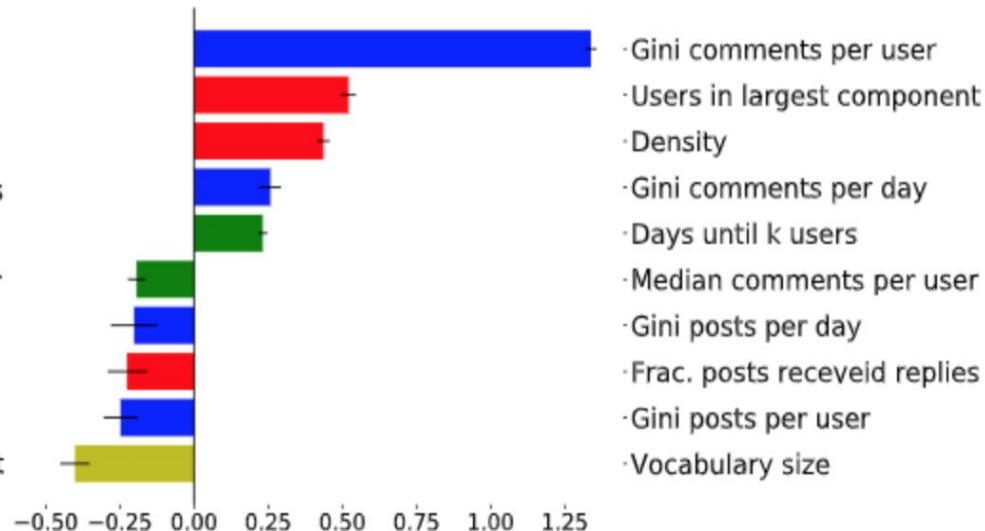


Top 10 features for growth in posters.

# Which features predict success?

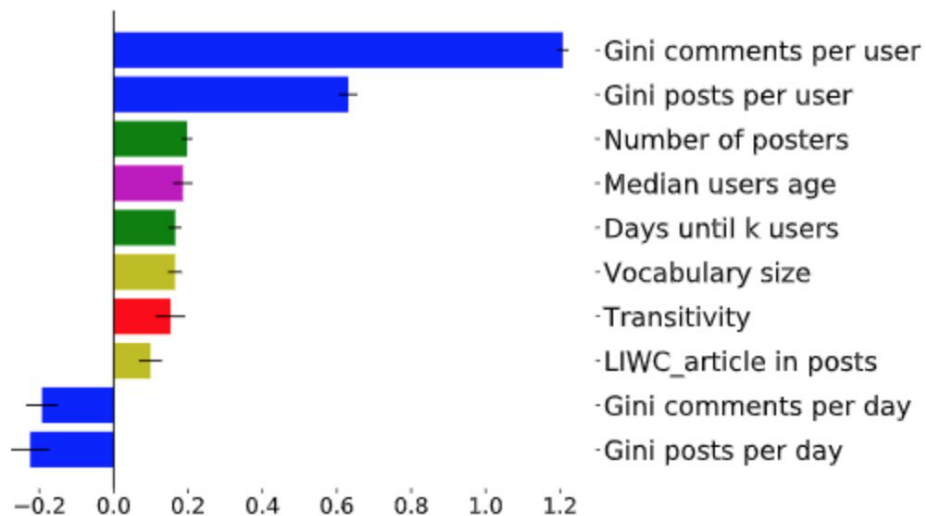


Top 10 features for average #comments.

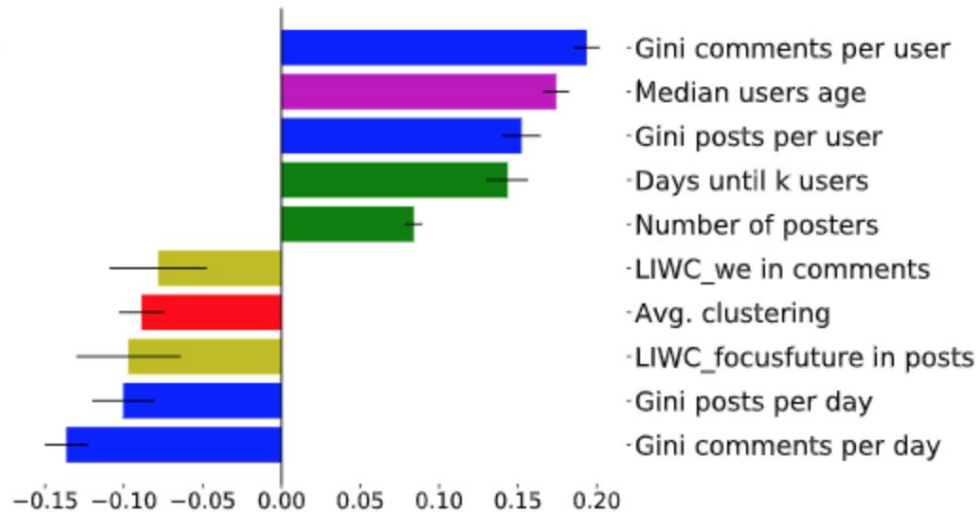


Top 10 features for average #posts.

# Which features predict success?



Top 10 features for users retention.



Top features for community survival.

# How early can we predict success?

Success measure	Std. AUC
Growth in commenters	0.030
Growth in posters	0.020
Retention	0.006
Survival	0.022
Average #comments	0.023
Average #posts	0.024

# Recommendations [Our results are observational!!]

- Having a small group of highly committed groups of participants early on is key to future success.
- Taking longer to attract  $k$  members is positively associated with users retention and average #posts.
- Average clustering is predictive of retention, suggesting that having a close-knit set of participants who bond well with each other is important for retention but not growth,

# Limitations

- Not controlling for effect of a community's topic
- Experiments focus on subreddits created in 2014
- Causality
- Our analysis only uses posts and comments

# Summary of findings

- Groups succeed in their own way, in part due to the diversity of why a group forms
- Future success of a group can be predicted along each of the four measures
- No single behavior drives a group to be successful in each dimension.



