

*What's in a name?*  
*The Interplay between*  
*Titles, Content & Communities*  
*in Social Media*

*Himabindu Lakkaraju, Julian McAuley, Jure Leskovec*  
*Stanford University*

# Motivation

Content, Content Everywhere !!



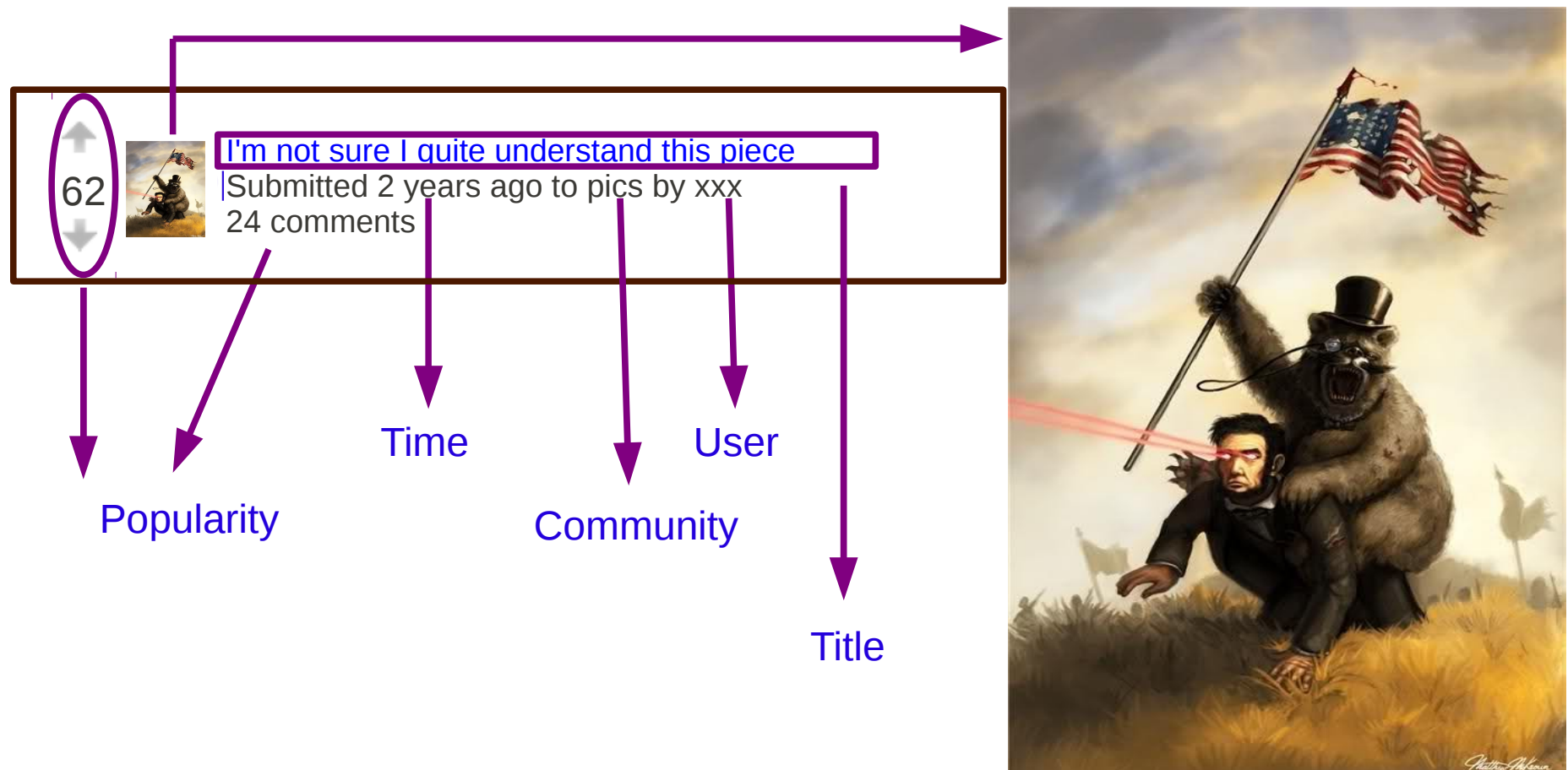
TechCrunch

How to get your content noticed amidst such information overload ?

# An Example

## Understanding a submission and its popularity

Content



# An Example

Understanding a submission and its popularity

Content

Is content the only factor  
in determining popularity ?

Popularity

Community

Title



# An Example

↑ [I'm not sure I quite understand this piece](#)

62 Submitted 2 years ago to pics by xxx

↓ 24 comments

↑ [How wars are won](#)

20 Submitted 18 months ago to WTF by xxx

↓ 1 comment

↑ [Murica!](#)

774 Submitted 1 year ago to funny by xxx

↓ 59 comments

↑ [Bring it on England, Bring it on !!](#)

10 Submitted 10 months ago to pics by xxx

↓ 4 comments

↑ [I believe this is quite relevant currently](#)

226 Submitted 7 months ago to funny by xxx

↓ 15 comments

↑ [God bless whoever makes these](#)

794 Submitted 1 month ago to funny by xxx

↓ 34 comments



# An Example

- ↑ I'm not sure I quite understand  
62 Submitted 2 years ago to funny by xxx  
↓ 24 comments
- ↑ How wars are won  
20 Submitted 18 months ago to funny by xxx  
↓ 1 comment
- ↑ Murica!  
774 Submitted 1 year ago to funny by xxx  
↓ 59 comments
- ↑ Bring it on England, Bring it on America  
10 Submitted 10 months ago to funny by xxx  
↓ 4 comments
- ↑ I believe this is quite relevant  
226 Submitted 7 months ago to funny by xxx  
↓ 15 comments
- ↑ God bless whoever makes these  
794 Submitted 1 month ago to funny by xxx  
↓ 34 comments



# An Example

- ↑ I'm not sure I c  
62 Submitted 2  
↓ 24 commen
- ↑ How wars a  
20 Submitted 1  
↓ 1 comment
- ↑ Murica!  
774 Submitted 1  
↓ 59 commen
- ↑ Bring it on B  
10 Submitted 1  
↓ 4 comments
- ↑ I believe this is quite relevant currently  
226 Submitted 7 months ago to funny by xxx  
↓ 15 comments
- ↑ God bless whoever makes these  
794 Submitted 1 month ago to funny by xxx  
↓ 34 comments

Content is not  
the only factor !!



- 
-



# An Example

↑ I'm not sure I quite understand this piece

62 Submitted

↓ 24 comments

↑ How was it?

20 Submitted

↓ 1 comment

↑ Murica!

774 Submitted

↓ 59 comments

↑ Bring it on

10 Submitted

↓ 4 comments

↑ I believe this

226 Submitted 7 months ago to funny by xxx

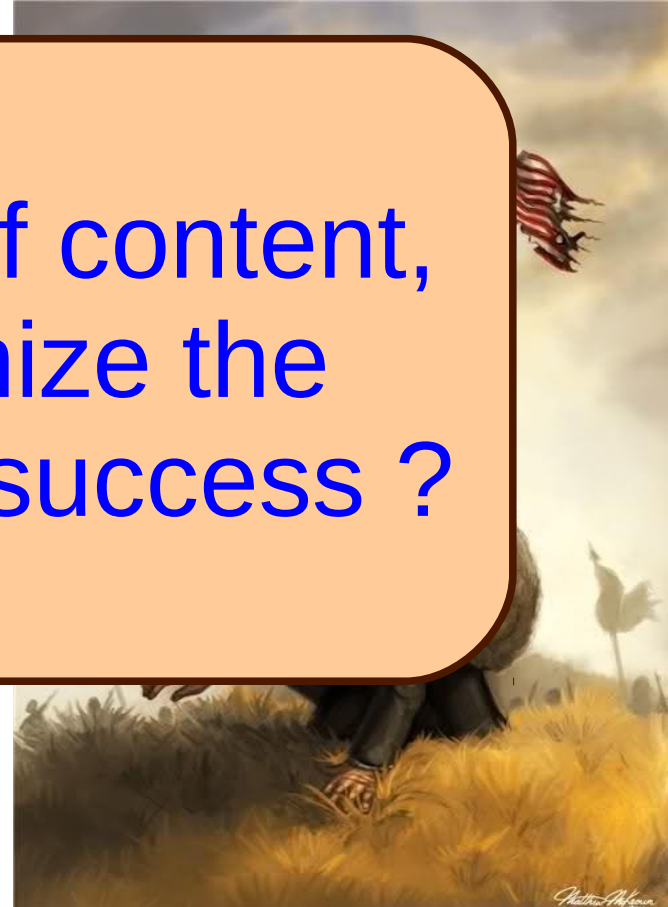
↓ 15 comments

↑ God bless whoever makes these

794 Submitted 1 month ago to funny by xxx

↓ 34 comments

Given a piece of content,  
can we maximize the  
probability of its success ?



- 
-



# Motivation

Factors influencing popularity

Community or Forum

Time of posting

Title of submission

Popularity of user

Previous submissions of same content

+

Content

and their confounding interplay!

# Motivation

Factors influencing popularity

Community or Forum

How do we  
tease apart  
these effects ?

Content

and their confounding interplay!

# Teasing apart..

How do we tease apart effects of various factors ?

Dataset which accomodates

Resubmissions of same content

Submissions across multiple communities

Communities with varying characteristics

Submissions by multiple users

# Teasing apart..

## Reddit to the rescue !

MY SUBREDDITS ▼ FRONT - ALL - RANDOM | PICS - FUNNY - POLITICS - GAMING - ASKREDDIT - WORLDNEWS - NEWS - VIDEOS - IAMA - TODAYILEARNED - WTF - AWW - ATHEISM - TECHNOLOGY - ADVICEANIMALS - SC

 **reddit** hot new rising controversial top wiki

- ↑ 2128 ↓  
**James Bamford: "The NSA has no constitutional right to secretly obtain the telephone records of every American citizen on a daily basis, subject them to sophisticated data mining and store them forever. It's time government officials are charged with criminal conduct, including lying to Congress"** (blog.sfgate.com)  
submitted 2 hours ago by trot-trot to politics  
146 comments share
- ↑ 1310 ↓  
 **Bajo and Hex, hosts of Australian TV shows Good Game and Good Game Spawn Point - AMA** (self.IAmA)  
submitted 2 hours ago\* by goodgameabctv to IAmA  
1192 comments share
- ↑ 2691 ↓  
**Majority of people worldwide believe corruption has worsened - governments less effective at curbing it since 2008 financial collapse** (nytimes.com)  
submitted 8 hours ago by oshunsmall to worldnews  
326 comments share

# Teasing apart..

## Our Dataset

- A novel dataset of 132K reddit submissions
- Every piece of content (image) submitted multiple times
- 16.7K original submissions
- Average of 7 resubmissions per image

Data available at <http://snap.stanford.edu/data>

# Our Goal

- To study the effect of the interplay between content, title, communities on a submission's popularity
- To understand how much of a submission's popularity is due to its
  - ◆ Inherent quality
  - ◆ Community choice
  - ◆ Time of posting
  - ◆ Characteristics of submission title

# Our Approach

- Model the popularity of a submission as a combination of various factors
- Evaluate the goodness of the model by predicting popularity
- How do we quantify popularity ?
  - Reddit score = # of upvotes - # of downvotes



# Our Contributions

Popularity = Community Model + Language Model

- **Community model:** choice of community + time of submission + previous submissions of same content
- **Language model:** linguistic features of submission title + language of community and,
- **a novel dataset** which allows the study of various factors

# Related Work

- Predicting the success of social media content
  - [Content based approaches](#) [Bandari et. al.] [Tsagkias et. al.]  
[Yano et. al.]
- Understanding the relationship between language and social engagement
  - [Analysis of lexical features](#) [Danescu-Niculescu-Mizil et. al.]  
[Hong et. al.] [Petrovic et. al.] [Suh et al.]

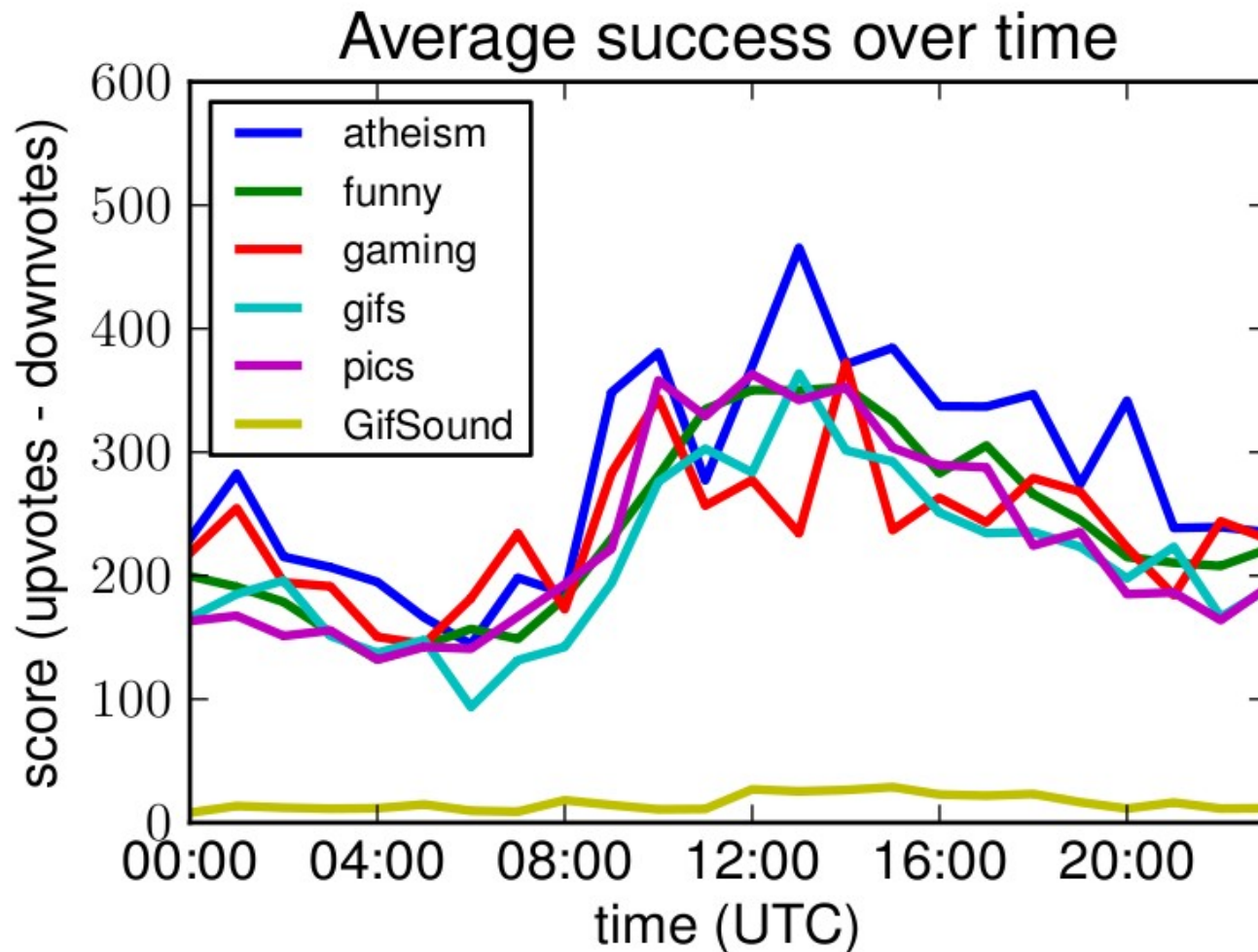
# Related Work

- Predicting the success of social media content
  - Content based approaches [Bandari et. al.] [Tsagkias et. al.] [Yano et. al.]
- Understanding the relationship between language and social engagement
  - Analysis of lexical features [Danescu-Niculescu-Mizil et. al.] [Hong et. al.] [Petrovic et. al.] [Suh et al.]

Our work focusses on the interplay between content, lexical features, communities and the resulting composite effect on popularity

# Insights

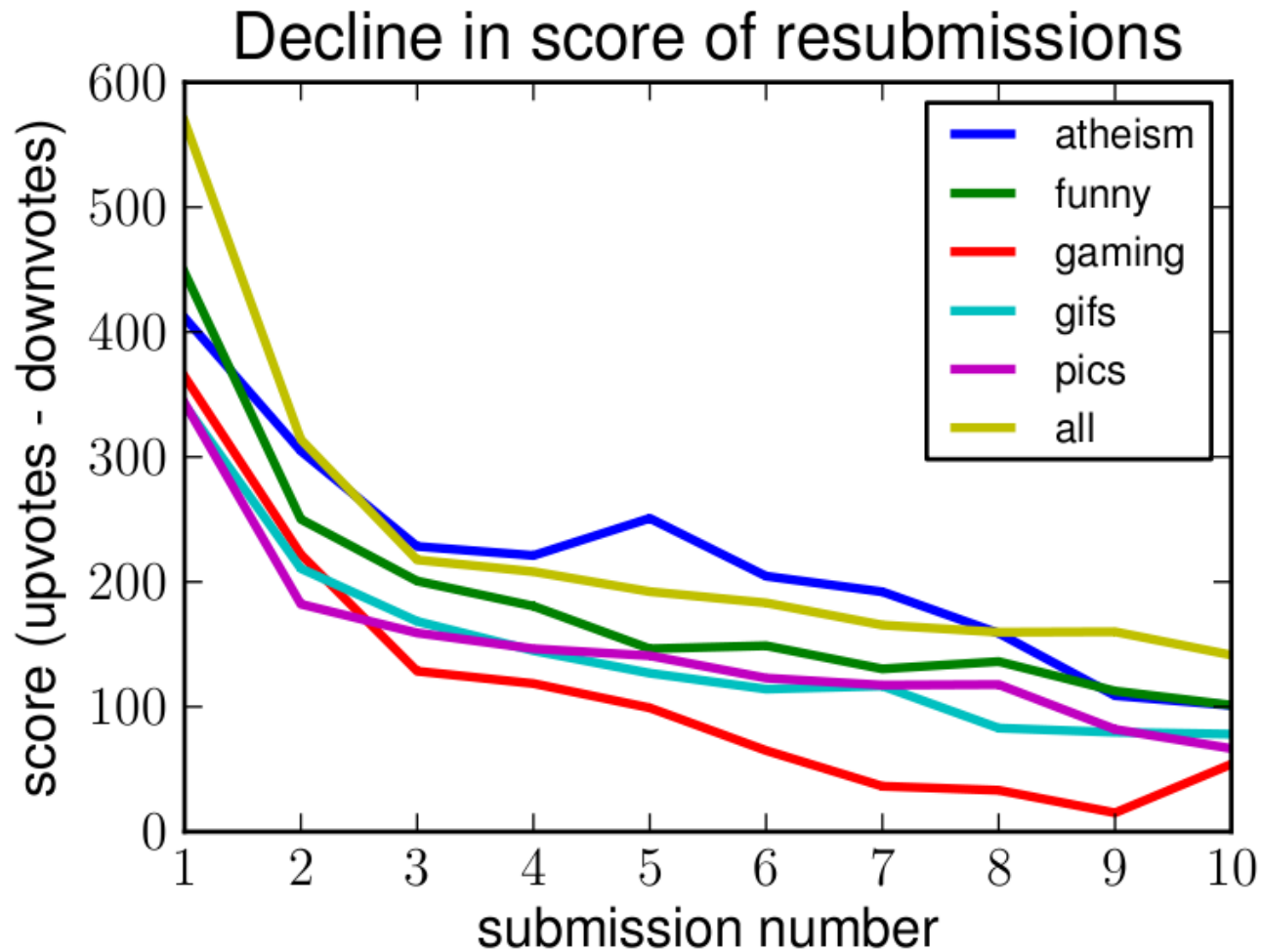
## Understanding community activity



Popularity varies with time of the day

# Insights

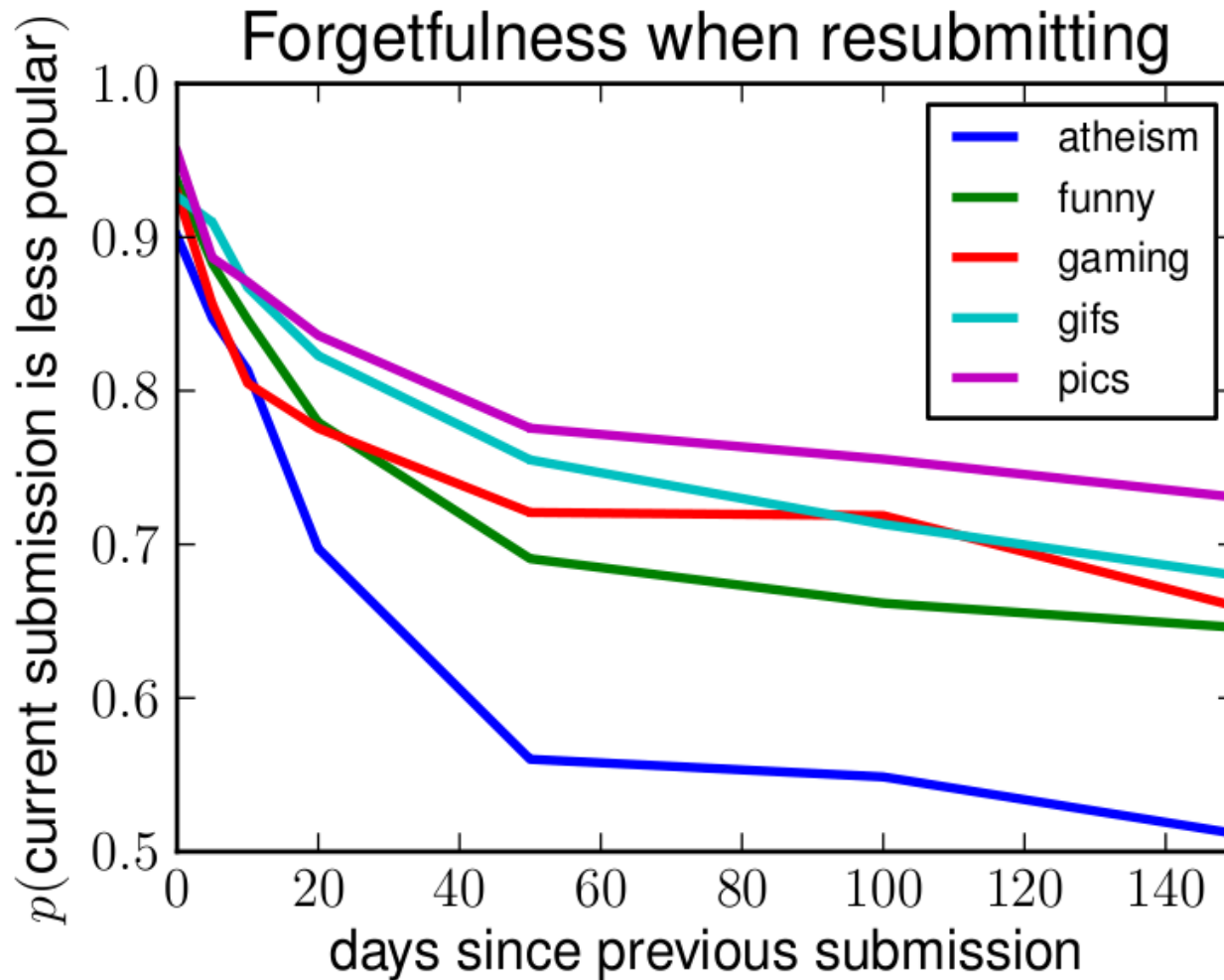
## Understanding community activity



Content is less popular with each resubmission

# Insights

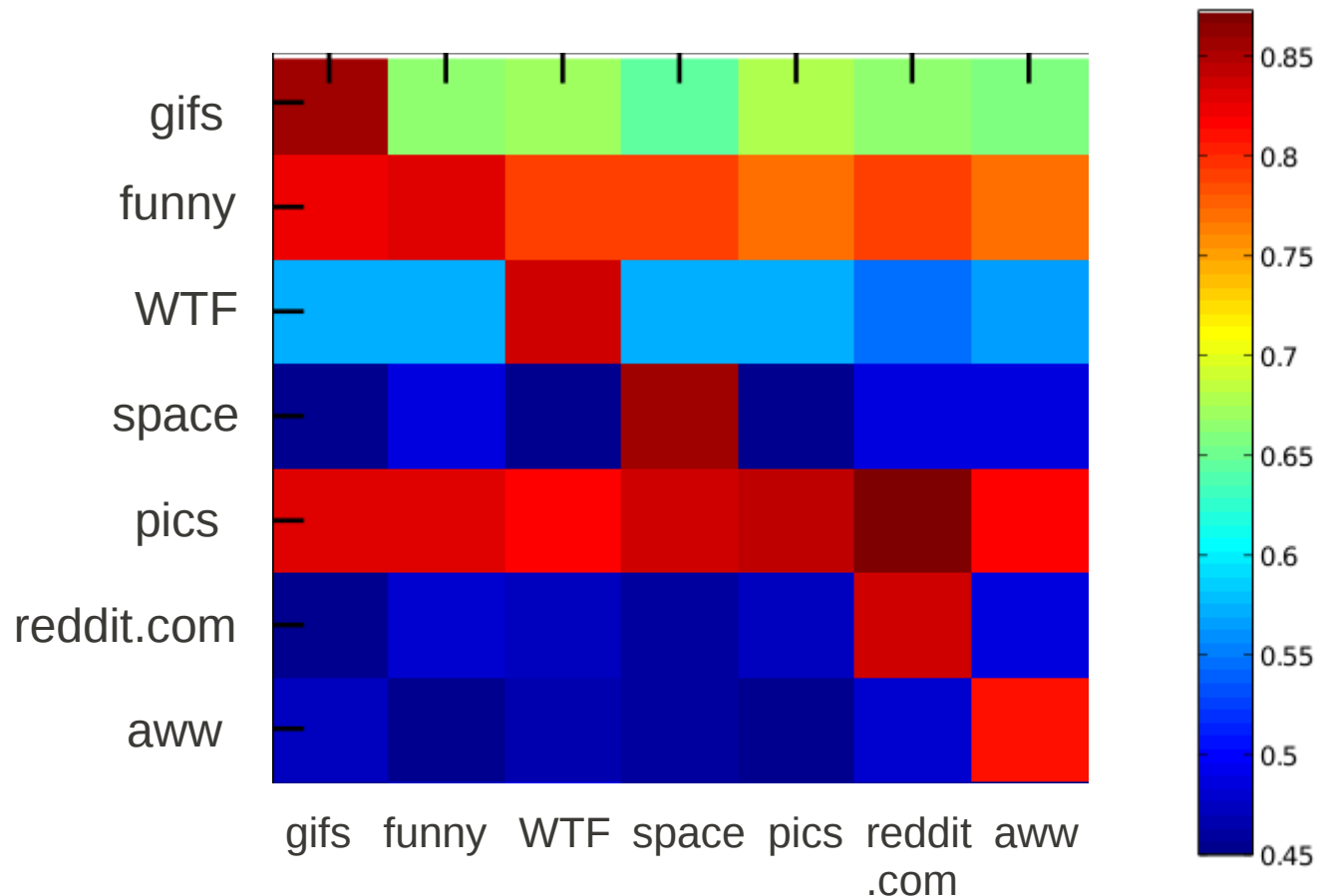
## Understanding community activity



Resubmissions are “forgiven” given enough time

# Insights

## Understanding inter-community effects



Don't resubmit to same community (diagonal)  
Don't resubmit highly visible content (rows)



# Our Approach

## Community Model

Input

Output

Inherent popularity

Resubmission decay

Forgetfulness

Inter-community effects



Popularity

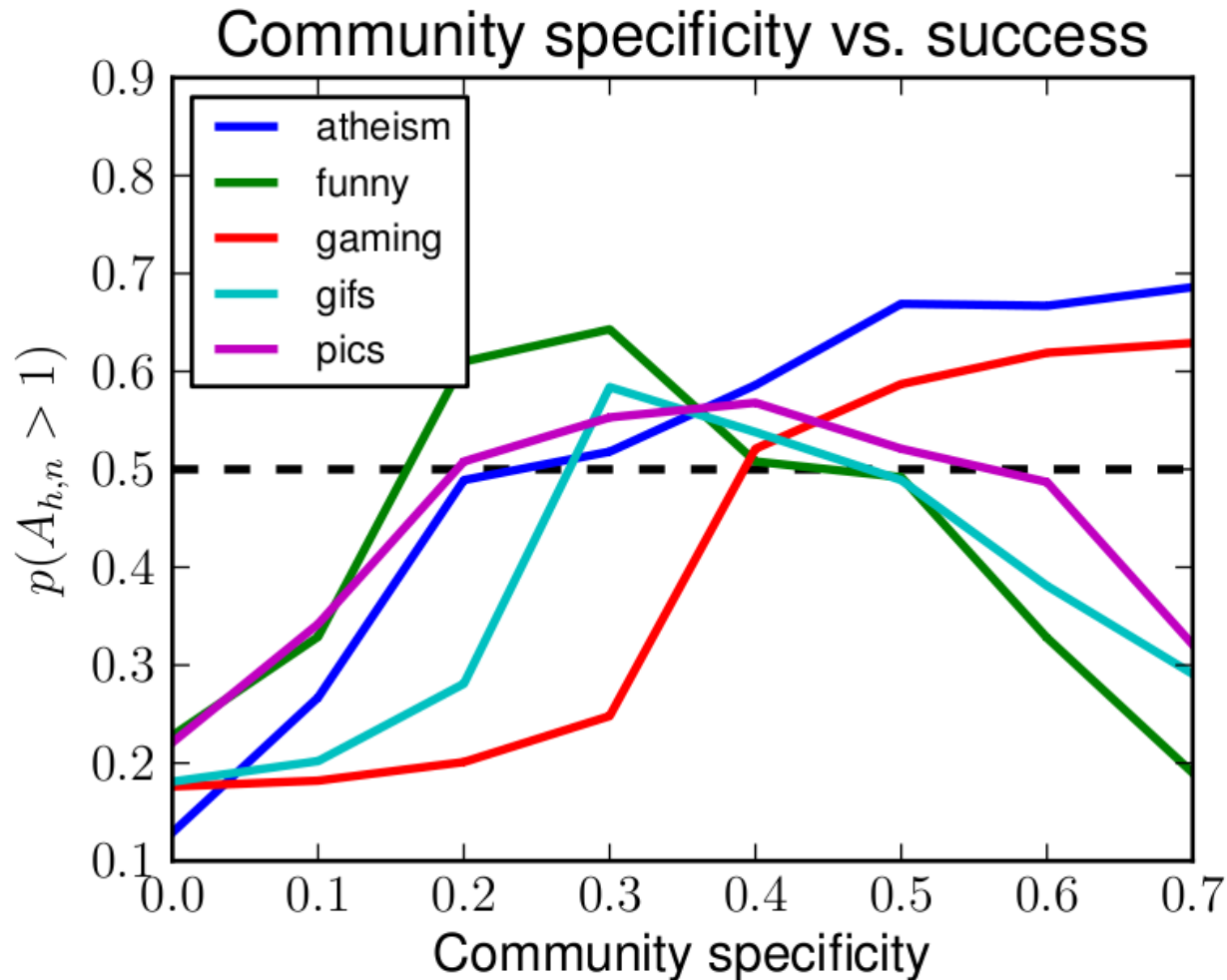
# Our Approach

## Language Model

- **Language of a Community** – Targeting title to a community
- **Content Specificity** – Title reflecting content
- **Title Originality** – Novelty of the title
- **Sentiment polarity, POS tags, # of words in title**

# Insights

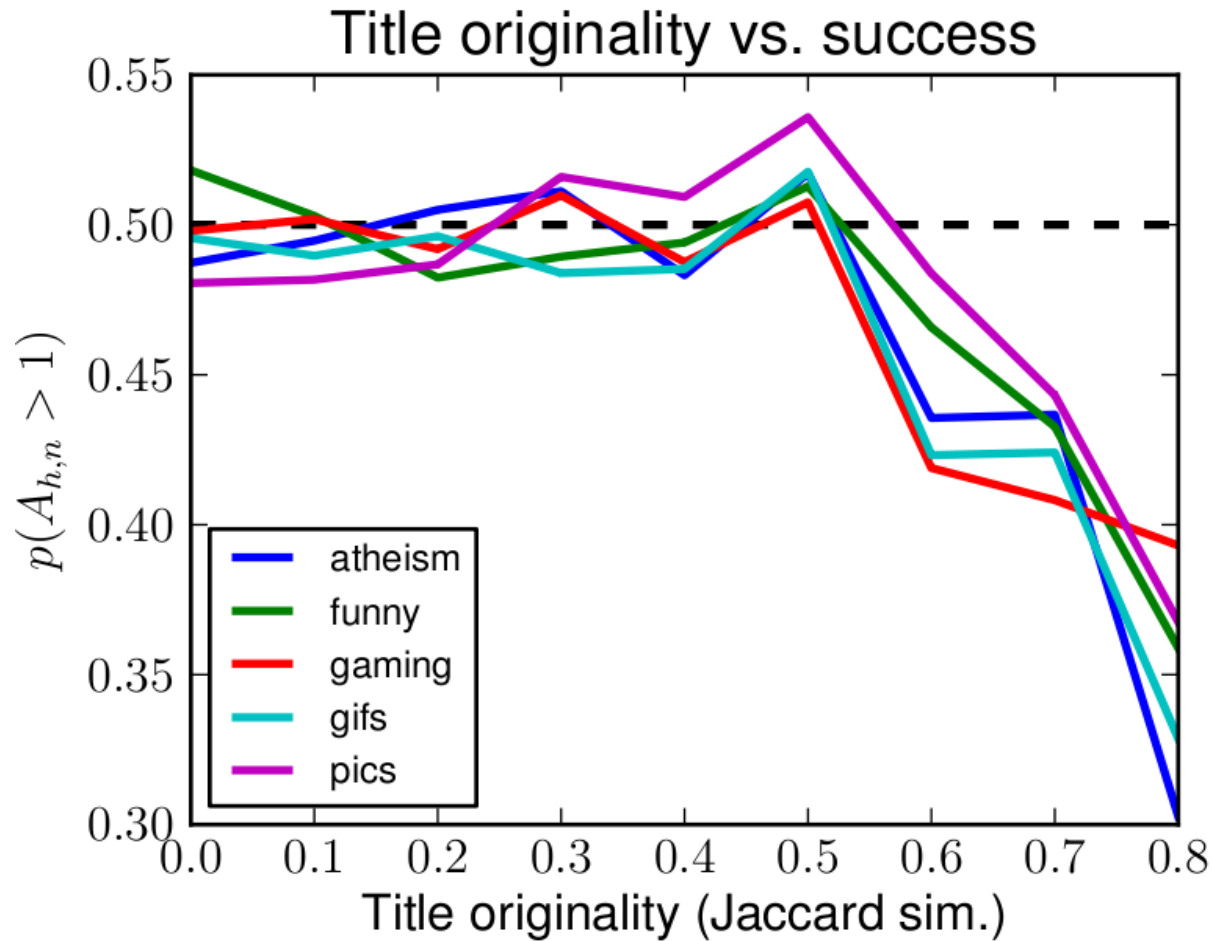
## Understanding language characteristics



Titles should balance novelty and familiarity

# Insights

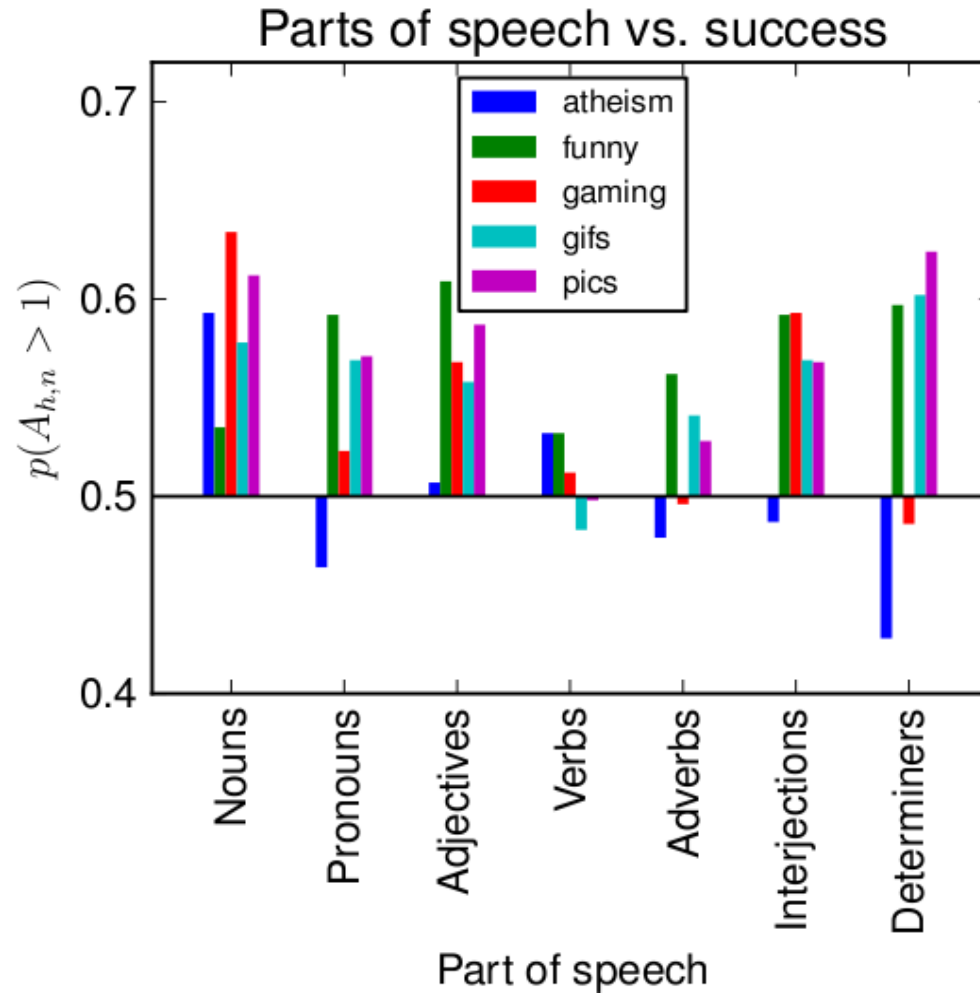
## Understanding language characteristics



Resubmissions benefit from novel titles

# Insights

## Understanding language characteristics



Various communities prefer different POS

# Quantitative Evaluation

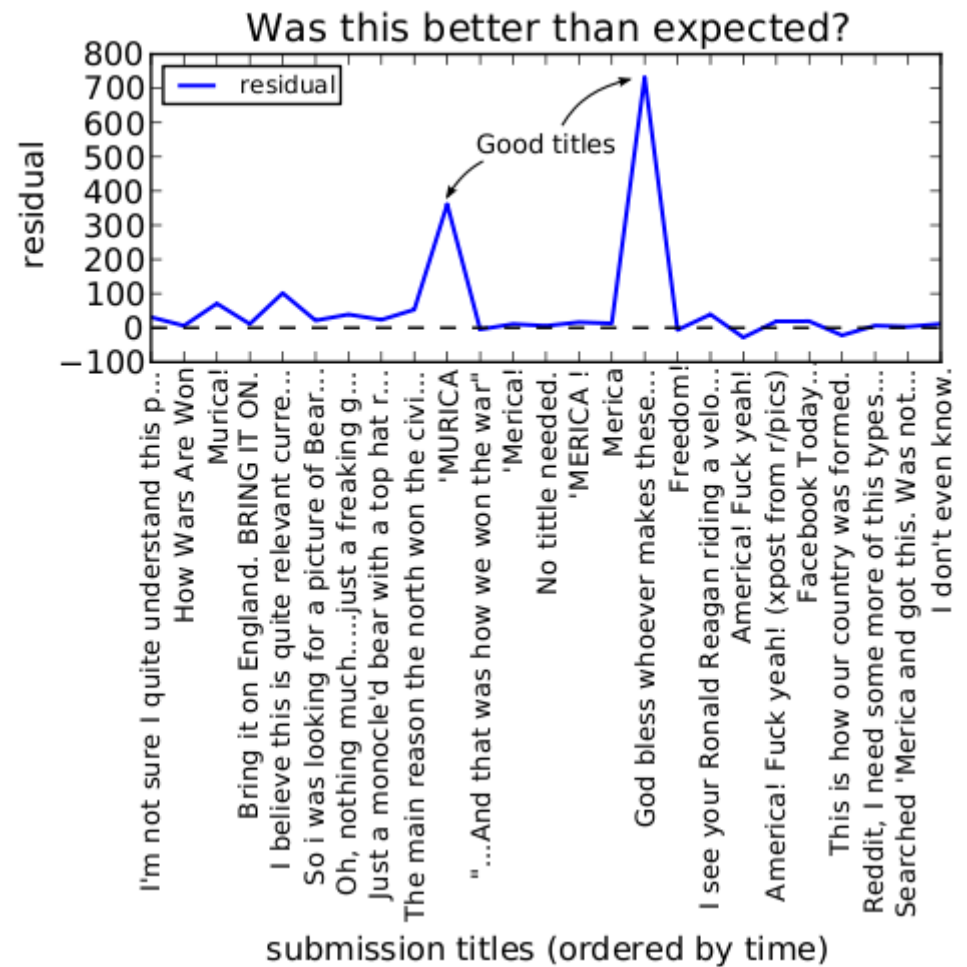
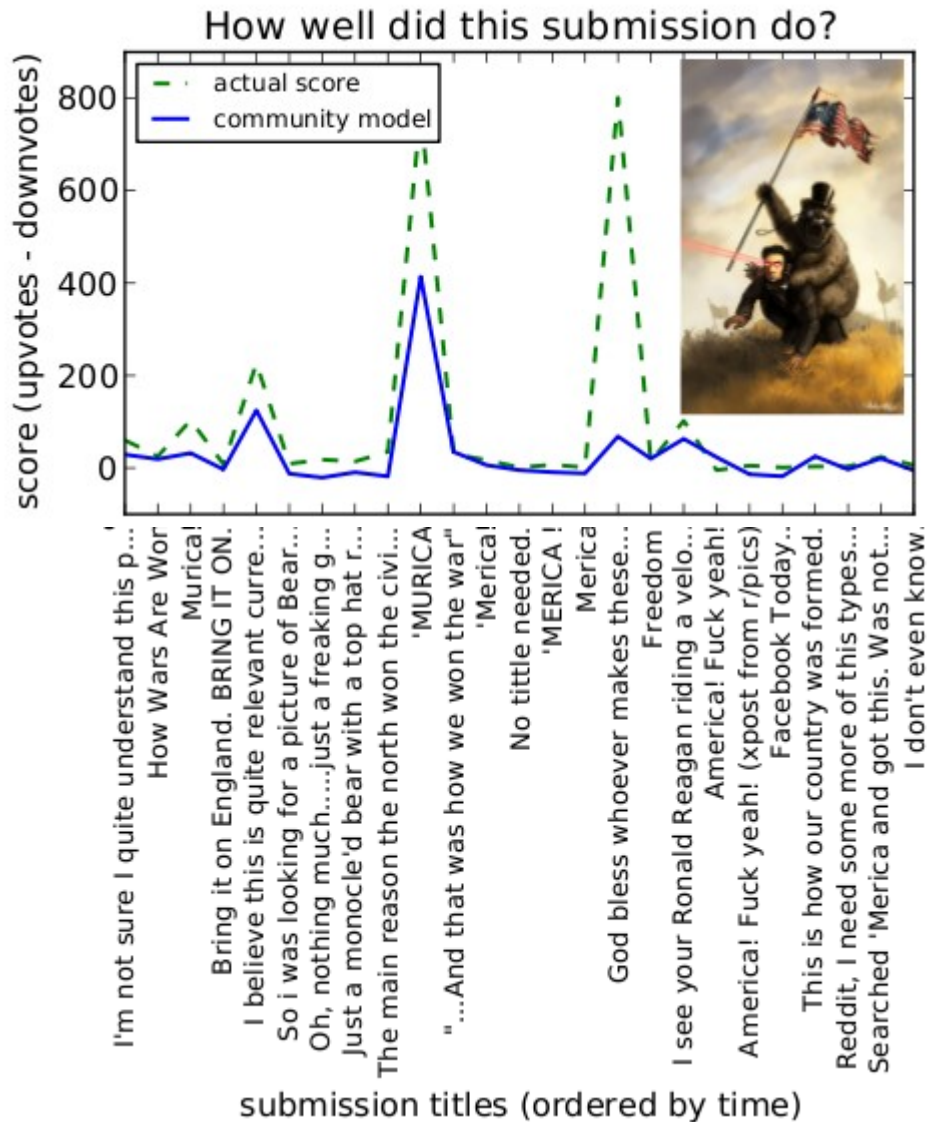
## Predicting reddit score

Evaluating predictive power on a held out test set of 25% of the data

Coefficient of determination –  $R^2$  statistic  
(value of 1.0 indicates perfect fit)

Model	$R^2$
Community Model	0.528
Language-only Model	0.081
Community + Language	<b>0.618</b>

# Qualitative Evaluation





# Qualitative Evaluation

Title	Community	SC	LC	O	POS
I'm not sure I quite understand this piece.	WTF	--	.	+	.
I believe this is quite relevant currently..	funny	-	++	.	.
The main reason the north won the civil war.	funny	.	++	-	+
'MURICA	funny	.	--	++	-
...And that was how we won the war	pics	+	.	.	+
No tittle needed.	pics	.	+	.	-
'MERICA !	pics	.	-	--	--
God bless whoever makes these...	funny	++	+	+	.
Freedom!	WTF	.	-	.	--
I see your Ronald Reagan riding a velociraptor, and raise you a bear riding Abe Lincoln.	funny	.	.	.	++

SC: words Specific to the Community  
LC: uses words Liked by the Community  
O: title is Original  
POS: Parts-of-Speech are appropriate

Top 10% (++) Top 25% (+)  
Bottom 25% (-) Bottom 10% (--)

# Qualitative Evaluation

Title	Community	SC	LC	O	POS
I'm not sure I quite understand this piece.	WTF	--	.	+	.
I believe this is quite relevant currently..	funny	-	++	.	.
The main reason the north won the civil war.	funny	.	++	-	+
'MURICA	funny	.	--	++	-
...And that was how we won the war	pics	+	.	.	+
No tittle needed.	pics	.	+	.	-
'MERICA !	pics	.	-	--	--
God bless whoever makes these...	funny	++	+	+	.
Freedom!	WTF	.	-	.	--
I see your Ronald Reagan riding a velociraptor, and raise you a bear riding Abe Lincoln.	funny	.	.	.	++

SC: words Specific to the Community  
 LC: uses words Liked by the Community  
 O: title is Original  
 POS: Parts-of-Speech are appropriate

Top 10% (++) Top 25% (+)  
 Bottom 25% (-) Bottom 10% (--)

# In Situ Evaluation

## Real time action on Reddit !

A sample of 85 images from our dataset  
Assigned a “good” and a “bad” title for each image

Total score of all “good” submissions is 3 times higher

2 of our “good” submissions hit Reddit front page

3 more featured on front pages of communities

# Conclusion

- Popularity is effected by the interplay of various content, language and community specific aspects
- We propose models which disentangle these effects
- Modeling these effects helps us understand what fraction of popularity can be attributed to each of these factors

# Thank you !!

- R. Bandari, S. Asur and B. Huberman. The pulse of news in social media: Forecasting popularity. In ICWSM 2012.
- M. Tsagkias, W. Weerkamp and M. Derijke. Predicting the volume of comments on online news stories. In CIKM 2009.
- Y. Yano and N. Smith. What's worthy of comment? content and comment volume in political blogs. In ICWSM 2010.
- C. Danescu-Niculescu-Mizil, M. Gamon and S. Dumais. Mark my words! Linguistic style accommodation in social media. In WWW 2011.
- L. Dang, O. Dan and B. Davison. Predicting popular messages in twitter. In WWW 2011.
- S. Petrovic, M. Osborne and V. Lavrenko. Rt to win! Predicting message propagation in twitter. In ICWSM 2011.
- B. Suh, L. Hong, P. Pirolli and E. Chi. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In SocialCom 2010.
- D. Blei and J. McAuliffe. Supervised topic models. In NIPS 2007.