Identifying the social signals that drive online discussions: A case study of Reddit communities

Benjamin D. Horne, Sibel Adalı, and Sujoy Sikdar
Rensselaer Polytechnic Institute
110 8th Street, Troy, New York, USA
{horneb, adalis, sikdas}@rpi.edu

Abstract—Increasingly people form opinions based on information they consume on online social media. As a result, it is crucial to understand what type of content attracts people's attention on social media and drive discussions. In this paper we focus on online discussions. Can we predict which comments and what content gets the highest attention in an online discussion? How does this content differ from community to community? To accomplish this, we undertake a unique study of Reddit involving a large sample comments from 11 popular subreddits with different properties. We introduce a large number of sentiment, relevance, content analysis features including some novel features customized to reddit. Through a comparative analysis of the chosen subreddits, we show that our models are correctly able to retrieve top replies under a post with great precision. In addition, we explain our findings with a detailed analysis of what distinguishes high scoring posts in different communities that differ along the dimensions of the specificity of topic and style, audience and level of moderation.

I. INTRODUCTION

In this paper, we study the following problem: What type of comments drive discussions on social media? First, we examine whether it is possible to predict which comments receive positive attention. In conjunction, we ask the following related questions: If prediction is possible, what features are useful in prediction? Secondly, what are the distinguishing features of comments that receive high attention in each community, and how do these differ from one community to another?

Increasingly, people form opinions based on information they consume on online social media, where massive amounts of information are filtered and prioritized through different communities. As a result, social media sites are often targets of campaigns for dissemination of information as well as misinformation [1] [2]. These campaigns can employ sophisticated techniques to hijack discussions by posting content with a specific point of view within posts and discussions in order to attract attention and steer the discussion or influence how users interpret the original content [3]. It is often observed that in our information saturated world, user attention is one of the most valuable commodities. Hence, it is crucial to understand which type of content receives high attention. We consider this as the first step in the study of information dissemination in online discussions and in building tools for information processing.

To address the central problem of this paper, we study a large dataset of comments from many different communities

on reddit. Reddit is one of the most popular platforms for news sharing and discussion, ranking #4th most visited site in US and #16 in the world. Reddit claims to be the front page of the internet, achieving its stated purpose by allowing users to *post* news, questions, and other information in the form of text, images and links to external websites. Users often engage with the posts by getting involved in or reading discussions consisting of comments made by other users in the community. Discussions are a vital and valuable feature of Reddit. Posts often generate lengthy and vibrant discussions, and comments that help users analyze and engage with the content, through the different perspectives and interpretations provided by members of the community.

Voting is the main mechanism reddit provides its users to affect the ranking and the visibility of posts and comments. Every post or comment on reddit is assigned a *score* based on the votes it receives. An *upvote* increases the score by one and a *downvote* decreases it. Posts and comments are sorted and presented to users (loosely) in order of the score they receive. While reddits' algorithm slightly obfuscates the ordering to prevent users from gaming the mechanism, the score is the primary and most significant factor in ordering posts and comments made within a small time period and is directly correlated with the votes. Voting allows users to steer the discussion and drive the most relevant, interesting, and insightful comments to prominence in the discussion.

It is undeniable that reddit has its own norms and culture, organized around its communities called subreddits. Subreddits differ from each other in many different ways, especially in four specific dimensions: topic, audience, moderation and style (see Figure 1). Some subreddits are topic specific (r/AskHistorians or r/Bitcoin), while others are from a general topic (r/AskReddit). Subreddits can differ in whether they target a specialized audience or not as in the case of r/Bitcoin for experts on this topic. Some subreddits have a very specific style for posting questions and comments such as in r/todayilearned that targets submissions that are verifiable facts. While all subreddits have rules regarding what types of content is allowable in that community (see Figure collage), the specificity of the rules and the level of moderation differ greatly from subreddit to subreddit. Even when the written standards are similar, reddit communities attract users with different interests and discussions of different nature. One expects that this results in other *unwritten* standards of quality that can only be inferred
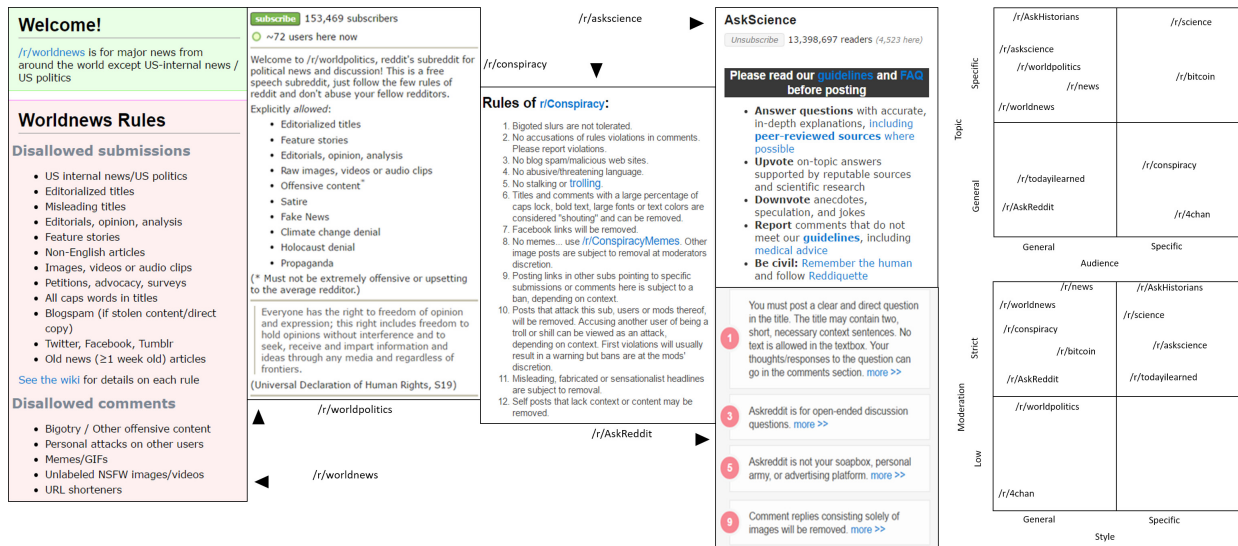
(a) Example subreddit rules   (b) Dimensions of subreddits

Fig. 1. Rules & dimensions: Wide variation in rules can be seen between r/worldnews & r/worldpolitics in part a and in the moderation dimension of part b.

from the readers' votes.

Given the very large user base of reddit and the diversity of subreddits along these four dimensions, it is not necessarily clear that high scoring content is predictable. In fact, a recent study reports fairly low accuracy results [4]. Previous work does not make it clear to which degree the prediction accuracy is impacted by the choice of features, communities or the learning method. We study this problem in detail and make the following contributions:

- We undertake a study featuring 11 popular subreddits that differ across the four dimensions discussed above. We sample a large number of posts analyze the content of comments under them and their relationship to scores.
- We analyse a comprehensive set of features, from previous work for predicting expertise, news engagement and readability as well as novel features geared towards the reddit culture, such as the self-referential nature of discussions.
- We train machine learning models using a combination of time, sentiment, relevance and content analysis features. Our models significantly outperform the state of the art and perform well across a range of subreddits, irrespective of topic, moderation, audience and style, consistently ranking the top comments by score with high precision. We find that sentiment based features are more useful than other categories.
- We perform a post-hoc analysis to find significant features that distinguish between high and low scoring comments. Many features are significant in many communities including our novel features. Some features are consistently positively or negatively correlated across communities, while others may flip sign between communities. Most notably the relationship to time of comments shows a more complex picture than previously reported in the literature.
- We include a detailed discussion of the similarities and

differences between communities based on this post-hoc analysis. We show that audience, specificity of topic and style matter greatly in understanding which features are prominent. Surprisingly, we find that subreddits with very different levels of moderation may show very similar behavior.

- We also study the impact of users' attributes and show that comments by high scoring and highly active users do not necessarily end up on top. However, comments by users with flairs often end up on top, even when these flairs are self-assigned. We speculate that flairs act as an easy to evaluate heuristic signaling expertise.

II. RELATED WORK

Most related to our work is work by Jaech et al [4], in which the authors explore language's impact on reddit discussions. The authors use a set of many complex natural language features to rank comment threads in 6 different subreddits using a SVM ranking algorithm. The ranking results achieve relatively low predictive power, only attaining an average of 26.6% precision in retrieving the top 1 comment correctly. However, these results show that feature importance can change across community types. Along with this, the authors study the relative impact of "high karma" users on discussions by computing the percentage of discussions where the top comment is made by the top h-index user, where h-index is the the number of comments in each user's history that have a score (karma) greater than or equal to k. This brief user analysis concluded that high h-index users have little impact on the popularity of a comment. For comparison, our work will have some overlaps with this work. Specifically, we will use 2 of the same communities, have some feature overlap, and use the h-index calculation in our user analysis. However, we incorporate many novel features, study a much larger data set with many more communities, achieve a far greater accuracy

and incorporate a unique post-hoc analysis showing striking differences across different communities.

In addition to Jaech et al [4], reddit has been well-studied from many different perspectives. Dansih et al. show that a comment's timing relative to the post matter in eliciting responses in the community `r/IAmA` [5]. Lakkaraju et al/ [6] and Tran et al. [7] show that reddit post titles and timing are important factors in the popularity of a post. However, this popularity can be delayed. In 2013, Gilbert determined that roughly 52% of popular reddit posts are unnoticed when first submitted. Further, the popularity and engagement of reddit posts can be reasonably determined by many factors [8]. Althoff et al. illustrate that temporal, social, and language features can play a role in successful requests in a study of altruistic requests in the reddit community `r/randomactsofpizza` [9]. More recently, Hessel, Lee, and Mimno show that visual and text features are important in image-based post popularity prediction. More over, Hessel, Lee, and Mimno show that user-based features do not predict popularity as well [10]. While all of these studies are exploring the posts on reddit rather than the comments as we do in this paper, they demonstrate the many perplexities in both the messages and messengers on reddit.

Also related to this problem is work on general information popularity, news engagement, expert finding, and information credibility. Sikdar et al. [11] develop models to predict credibility of messages on Twitter using several user and natural language features. This work shows that crowdsourced endorsements like upvoting contribute to predicting information credibility. Note that credibility of content does not necessarily imply its correctness, as one of the subreddits we choose explicitly allows conspiracy theories. A recent 2017 study explores the impact of reddit on news popularity in the community `r/worldnews` [12] and finds that well-known news popularity metrics are able to accurately predict the popularity of a news article on reddit. In essence, news behavior on this subreddit resembles news popularity in general. This work also shows that users tend to the change the news titles to be more positive and more analytical despite news being more negative overall [13]. When users change the titles of a news article, the article tends to become more popular. Similarly, another study predicts the popularity of news using sentiment and language features, showing that sentiment features are important in popularity prediction [14]. In 2016, Horne et al. [15] studied automatic discovery of experts on Twitter using simple language and meta heuristics. They found that experts tend to be more active on Twitter than their friends and that experts use simpler language than their friends, but more technical langauge than the users they mention. We will borrow features from many of these studies in our analysis of reddit comments as credibility and expertise are part of popularity of a comment.

III. FEATURES

To explore reddit discussions, we compute features on each comment in our data set. These features can be categorized into five categories: (1) sentiment (2) content (3) relevance (4) user,

and (5) time. A short description of our feature sets can be found in Table I.

Sentiment and Subjectivity: To compute sentiment features, we utilize the tool Vader-Sentiment [16] which is a lexicon and rule-based sentiment analysis tool proposed in 2014. We choose this tool as it is specifically built for "sentiment expressed on social media [16]" and has been shown to work well on reddit and news data [12] [14]. It provides 4 scores: negative, positive, neutral, and composite, where composite can be thought of as the overall sentiment in a text. We will include all 4 scores as features in our sentiment/subjectivity model.

Next, we utilize the Linguistic Inquiry and Word Count (LIWC) tool [17] for a mix of features for emotion and perceived objectivity of a comment. The emotion features include: positive and negative emotion, emotional tone, and affect. Other features from LIWC included in our sentiment/subjectivity model are: analytic, insight, authentic, clout, tentative, certainty, affiliation, present tense, future tense, and past tense. Despite LIWC computing these features using simple words counts, they have been shown to work well in a variety of settings [18].

Further, we include three features that directly measure the probability a comment is subjective or objective, computed by training a Niave Bayes classifier on 10K labeled subjective and objective sentences from Pang and Lee [19]. The classifier achieves a 92% 5-fold cross-validation accuracy and has been shown useful in predicting news popularity [12].

Content Structure: To analyse content structure, we take other word count features from LIWC such as parts of speech features (similar to what a POS tagger would provide), punctuation, and word counts for swear words and online slang. In addition, we capture the readability and clarity of a comment using three metrics: Gunning Fog, SMOG, and Flesch Kincaid, and the lexical diversity metric (Type-Token Ratio) [12].

Next, we compute "fluency" features based on the (log) frequency of words in a given corpus, capturing the relative rarity or commonality of a piece of text. These features can mean several things depending on the corpus used. Commonness of a word is in general is based on the Corpus of Contemporary American English (COCA) [20].It has been shown that humans tend to believe information that is more familar, even if that information is false [21]. To capture how well a comment fits into a given community style, we compute fluency on the corpus of each community. This localized fluency captures the well-known "self-referential" behavior of reddit and it's independent communities [22]. Some communities may have a very specific "insider" language, while others will not.

Relevance: To measure how much new information is added to a comment, we compute the similarity between the text of a comment and the post it is under as a notion of relevance of the comment to the post. To compute this feature, we first vectorize each word using word2vec [23] trained on COCA. Once all words are in vector format, we can compute centroids weighted by the inverse of the log word frequency in the text. We do this for the 5 rarest words in the post and the comment. Finally, we compute the cosine similarity between the post vector and the comment vector.

Abbr.	Description
h-index	number of comments in each users local history that have score $\geq$ h
activity	# of comments + # of posts made by user
flair	has flair or not, displayed with authors username, assigned by a moderator or self depending on subreddit

(a) User Features

Abbr.	Description
time_diff	difference between post and comment time

(b) Time Features

Abbr.	Description
vad_neg	negative sentiment score Vader-Sentiment
vad_pos	positive sentiment score Vader-Sentiment
vad_neu	neutral sentiment score Vader-Sentiment
vad_comp	composite sentiment score Vader-Sentiment
psubj	probability of subjectivity using a learned Naive Bayes classifier
pobj	probability of objectivity using a learned Naive Bayes classifier
subjcat	binary category of objective or subjective
posemo	number of positive emotion words
negemo	number of negative emotion words
tone	number of emotional tone words
affect	number of emotion words (anger, sad, etc.)
analytic	number of analytic words
insight	number of insightful words
authentic	number of authentic words
clout	number of clout words
tentative	number of tentative words
certain	number of certainty words
affil	number of affiliation words
focuspresent	number of present tense words
focusfuture	number of future tense words
focuspast	number of past tense words

(c) Sentiment Features (including subjectivity)

Abbr.	Description
relevance	cosine similarity between post vector and comment vector

(d) Relevance Features

Abbr.	Description
self_fluency	avg. frequency of most common words in subreddit corpus
coca_fluency	avg. frequency of least common 3 words using coca corpus
WC	word count
WPS	words per sentence
GI	Gunning Fog Grade Readability Index
SMOG	SMOG Readability Index
FKE	Flesh-Kincaid Grade Readability Index
ttr	Type-Token Ratio (lexical diversity)
conj	number of conjunctions
adverb	number of adverbs
auxverb	number of auxiliary verbs
pronoun	number of pronouns
ppron	number of personal pronouns
i	number of I pronouns
we	number of we pronouns
you	number of you pronouns
shehe	number of she or he pronouns
quant	number of quantifying words
swear	number of swear words
netspeak	number of online slang terms (lol, brb)
interrog	number of interrogatives (how, what, why)
per_stop	percent of stop words (the, is, on)
AllPunc	number of punctuation
quotes	number of quotes
function	number of function words
word_len	average word length

(e) Content Features

TABLE I
DIFFERENT FEATURES USED IN OUR STUDY

High similarity suggests little new information was added to the discussion, while low similarity suggests the opposite.

User: To study the influence of users on comment scores, we will use three simple features. Local h-index is the number of comments with score greater than or equal to h within a given community. This index is widely used to measure the scientific output of a researcher and reddit karma in [4]. This metric should capture a user's reputation in a community better than any central measure based on a user's historic comment scores. Local activity of a user is defined as the number of comments plus the number of posts a user makes within the community. Finally, flairs are visual badges displayed next to a user's screen name. Flairs are typically used to show a user's area of expertise and are given out by the moderators through a strict application process. However, some communities set these flairs up to be arbitrary user-selected tokens. These two types of flairs mean very different things as one is for expertise and the other simply for community involvement.

Time: To capture the timing of a comment, we will compute the difference between the post and comment submission times. Ranking using this time difference will be used as a baseline model.

IV. DATA SETS

To understand how discussion changes across communities, we gather comment threads from 11 different subreddits during a 6 month period in 2013. Once the comment threads are

TABLE II
SUBREDDITS USED IN STUDY (*Please note that some subreddits are NSFW and may contain offensive material. We highly recommend you use private browsing when visiting these subreddits.*)

subreddit	# posts	# cmnts	% users w/ flairs	# flairs
r/4chan	10225	89080	5.6	133
r/AskHistorians	10381	21926	3.2	304
r/AskReddit	10421	128487	0.0	0
r/askscience	10404	23302	3.9	637
r/Bitcoin	10468	52639	0.0	0
r/conspiracy	10471	45103	0.0	0
r/news	10583	46722	0.0	0
r/science	10533	59307	1.7	237
r/todayilearned	10453	102029	0.0	0
r/worldnews	10388	78120	0.0	0
r/worldpolitics	8057	24054	0.0	0
Total	**112K**	**582K**		**1.3K**

extracted, we randomly sample 10K comment threads from each subreddit. This data is extracted from Tan and Lees reddit post data set [24] and Hessel et al.s full comment tree extension to that reddit dataset [25], which contains 5692 subreddits, 88M posts, and 887.5M comments between 2006 and 2014. The statistics on our final extracted data sets can be found in Table II.

Communities: To understand the variation in noise and signal in online discussions, we collect 11 communities with respect to 4 dimensions: topic, audience, style, and modera-

tion. We explore communities that differ widely in moderation (r/worldnews and r/worldpolitics), communities based on expertise (r/science and r/askscience), communities based on news discussion (r/news and r/worldnews), communities that have large general audiences, (r/AskReddit and r/todayilearned), and communities that have smaller niche audiences (r/Bitcoin and r/conspiracy). In addition, we study r/4chan, a well-known "troll" and hate community that reaches a very specific audience using very little moderation.

Figure 1b shows where each community in our study falls with respect these four dimensions. In Figure 1a we provide example subreddit rules.

V. METHODOLOGY

To understand the voting behavior of different communities on reddit and to recover and uncover the communities' explicitly stated and hidden quality standards, we use the following methodology: (1) Learn a model to predict the score of a comment. (2) Evaluate learned models using learning to rank metrics. (3) Perform post-hoc analysis.

A. Learning to rank comments by score

We first describe the experimental setup. As described earlier, each subreddit consists of posts, under which users make comments. As a basic preprocessing step, we remove all posts that have fewer than 5 comments under them as the frequency distribution of the number of comments under a post is heavily skewed towards posts with just 1 or 2 comments. Including them in the dataset would heavily influence the learning to rank metrics such as the average precision and render them meaningless. Each dataset corresponds to a subreddit, and consists of comments, each described by a feature vector as well as information about the user who made the comment and the post under which the comment was made. In each subreddit, we pick 80% of the posts uniformly at random and use the comments under these posts to form the training set. The remaining comments form the test set.

Learning to predict score: We learn a regression model on the comments in the training set where each comment is described by a feature vector (see Table I) and the predicted variable is the score of the comment. In order to allow for easy introspection of the learned models, and in light of the non-linearity of some of our features, we chose to train simple linear models using the Python scikit-learn library [26]. We report results obtained from a model learned using ridge regression with regularization, where the regularization parameter is learned using 10-fold cross validation and the optimization objective is to minimize the L_2-norm between the predicted scores and the real scores from reddit data since it performed the best overall.

The learned model is used to predict the scores of the comments in the test set. We then rank the comments under each post in the test set according to their predicted score. We measure the performance of our models by comparing the predicted rankings versus the rankings according to their true scores on reddit.

Learning to rank metrics: We evaluate the performance on the test set by the following metrics from the learning to rank literature [27]:

1) Average Precision @ k: The percentage of the posts ranked among the top k as predicted by the learned model that are also among the top k posts by true scores, averaged over all posts.

2) Kendall-tau distance (KT-distance) @ k: Kendall-tau distance [28] between the relative ranking of the top k posts according to their true scores versus the relative ranking of the same k posts by their predicted scores.

We report the precision for $k = 1, 3, 5, 10$ and KT-distance for $k = 5, 10, 20$. KT-distance is a secondary feature, especially useful for posts that have a significantly large number of comments, giving us a complete picture together with precision. If we achieve high precision for posts with large number of comments and the Kendall-tau distance is low at some value of k, it means that: 1) comments predicted to be among the top k were truly among the top k by their true scores, and 2) the relative positions of the true top k comments are maintained in the predicted ranking. To summarize, a good model displays the following qualities:

- High average precision at low values of k.
- Low KT-distance for high values of k.
- KT-distance grows sub-linearly with k.
- At high values of k, high average precision and low KT-distance.

Good performance of learned models as measured by learning to rank metrics validates the predictive and descriptive power of the features. However, these models can still be hard to interpret. In order to gain greater insight into the voting behavior of reddits users in each community, we perform additional post-hoc analysis.

B. Post-hoc Feature Analysis

Our goal is to understand how each feature affects the score obtained by a comment. We start by dividing the data into two classes: low score comments (whose score are below the 50th percentile) and high score comments (whose scores are above the 90th percentile). How does the distribution of each feature affect whether a comment receives a low score or a high score? Since the features are not usually distributed normally, we use the Kolmogorov-Smirnov (KS) statistic as a robust measure of the effect size, which is independent of the distributions of the feature, and is sensitive to differences in the middle of the distribution which is of particular interest for this work. We then capture the top 15 features by effect size from each subreddit that were significant (with a p-value less than 0.05). We also capture the difference between the mean of the distributions of the feature values corresponding to the two classes to understand whether high scoring comments are attributed with higher or lower values of the feature.

VI. RESULTS AND DISCUSSION

In this section, we present results that answer the questions we set out to address in the introduction.

A. Yes, we can predict how comments are ranked.

The performance of our learned models are summarized in Table III which show that we can indeed predict ranking of comments with high precision. This is consistent across subreddits and the dimensions of style, moderation, subject and target audience. We achieve high average precision at all values of k including at $k = 1$. Moreover, the Kendall-Tau distance at k grows roughly linearly with the value of k (as opposed to growing exponentially). Our models significantly outperform the state of the art model [4]. Significantly, we achieved a significantly higher average precision at 1 result of 0.412 and 0.671 in the r/askscience and r/worldnews subreddits respectively (a 2 to 3 times improvement).

Since timeliness (represented by the feature time_dif) of a comment is widely cited as being a good predictor of the score of a comment (and in other literature of a post), we used a model trained using only timeliness as a feature as the baseline. Contrary to this widely held belief, we find that timeliness alone does not guarantee a high score. We also included time_dif as a feature in our sentiment, relevance and content models to measure the incremental improvement in performance by using these features. We found that relevance and sentiment alone are both highly predictive of the score of a comment. Sentiment held the highest predictive performance across subreddits. Unfortunately, but not surprisingly, the models performed poorly on the r/AskReddit dataset. This is likely because r/AskReddit is simply too diverse in terms of its topic and is aimed at a very general audience. It also somewhat loosely moderated. Surprisingly, prediction was possible for other loosely moderated subreddits such as r/worldpolitics and r/4chan.

B. There are both general and community specific factors that distinguish highly ranked comments.

To address our second main question, we perform a post-hoc analysis of the feature distribution for high and low score comments to determine which features distinguish high score comments, how this changes across communities, and how this corresponds to the explicit and implicit rules of the corresponding subreddits. The most important results can be found in Figure 2. The colors correspond to whether the feature is positively or negatively impacted the score of the comment on average while the intensity corresponds to the relative effect size normalized over the effect sizes of all features for each subreddit.

Timeliness is always important, but differently across communities: Saliently, we find that the importance of comment timing relative to the post is not consistent across communities. Specifically, we find that in the communities r/AskReddit, r/science, r/4chan, and r/news, comments that are made later in time tend to have higher scores, while the rest of the communities show the opposite effect. Previously, it has been shown that timing impacts the popularity of posts, in particular the time of the day or the week [10]. It has also been shown that comments which are submitted close the post submission time elicite more responses in the Multiple Inquirer Single Responder communtity r/IAmA [5]. Our result shows that the timing relative to the post is more dependent on the community than previously thought.

This result may appear for different reasons. For example, r/AskReddit often has posts reach and stay on the front page of reddit for a full day or more. This extended time may gain attention from multiple bursts of people, creating new sets of comments and new sets of votes later in time. While this bursty behavior inherently will not change the score of a popular post by much, it may change the number of comments and votes on comments by significant amounts. Similarly, timing of comments may be impacted by the average number of posts submitted to the community.

It is important to note that this does not nessarily negate the well-known rich-get-richer phenomenon on reddit [29] [10], but says that the rich-get-richer effect may not hold as strongly for comments as it does for posts.

Being relevant always matters: As expected, we find that comments that are more relevant to the post garner higher scores. This feature is the only feature that is globally consistent across all communities. Comment relevence was also shown to be important across several communities in [4].

New information over stale memes: Interestingly, we find that writing comments within community vocabulary (self_fluency) is not very important in comment popularity; in fact, it may hurt a comment's popularity. Specifically, we find that high score comments in expertise communities have a low self-fluency. This may mean the low score comments contain memes or jokes that have cycled in the community before or simply contain old information.

An alternate, and maybe more accurate, interpretation of this feature is how much new or rare information is in a comment relative to the community's history. Since this feature is the average frequency of highly frequent words in a corpus of community text, we are likely capturing how new or rare the information. This interpretation aligns well with how we expect expert communities to behave, as new information is more valuable information.

Moderation does not always impact behavior: Contrary to what we expected, moderation has much less of an impact on the normality of community behavior. In Figure 2, we can see significant similarity between r/worldnews and r/worldpolitics across many features. These common features include preferring more objective comments, less negative comments, more analytic comments, comments showing clout, and longer average word length. While the communities cover similar topics (i.e. international news), they are moderated in explicitly opposite ways. Figure 1a shows the completely opposing moderation structure of these two communities:r/worldnews provides well moderated, non-opinionated news stories, while r/worldpolitics has no restrictions on what news should be submitted, explicitly allowing propaganda, fake news, and offensive content. This notion is further supported by our prediction results, in which the strictness of moderation does not drastically impact our ability to rank comments.

Strikingly, these commonalities do not extend to the topic of news in general, as the community r/news (i.e. U.S. news) is exceedingly different from r/worldnews despite

Dataset	Model	Precision @ k				KT-distance @ k		
		$k=1$	$k=3$	$k=5$	$k=10$	$k=5$	$k=10$	$k=20$
r/4chan	Time	0.0	0.0	0.13	0.513	0.56	2.799	8.43
	Time+Sentiment	**0.682**	0.483	**0.585**	0.793	2.256	8.62	20.728
	Time+Relevance	0.544	0.451	0.576	0.806	2.681	9.294	21.128
	Time+Content	0.588	0.473	0.579	0.782	2.327	8.725	21.719
	All	**0.643**	0.483	0.588	0.795	2.183	8.519	20.721
r/AskHistorians	Time	0.0	0.0	0.382	0.84	0.896	2.549	4.417
	Time+Sentiment	**0.667**	0.514	**0.744**	**0.922**	2.306	5.396	10.396
	Time+Relevance	0.437	0.431	**0.688**	0.896	2.924	7.646	13.285
	Time+Content	0.563	0.468	0.696	0.936	2.514	7.09	10.812
	All	0.528	0.486	**0.708**	**0.917**	2.569	6.944	11.542
r/AskReddit	Time	0.0	0.0	0.181	0.678	2.258	8.553	16.249
	Time+Sentiment	0.254	0.285	0.457	0.796	1.197	6.072	12.931
	Time+Relevance	0.251	0.297	0.503	0.82	1.292	5.301	10.603
	Time+Content	0.19	0.287	0.507	0.828	1.296	5.292	11.023
	All	0.193	0.295	0.485	0.821	1.197	5.452	11.469
r/askscience	Time	0.0	0.0	0.379	0.813	1.462	4.643	9.72
	Time+Sentiment	**0.412**	0.396	**0.67**	**0.888**	2.099	4.758	8.176
	Time+Relevance	0.253	0.385	0.664	0.897	2.198	5.198	8.533
	Time+Content	0.368	0.368	0.63	0.897	1.94	5.115	8.593
	All	**0.456**	0.405	**0.664**	**0.909**	1.857	4.538	7.39
r/Bitcoin	Time	0.0	0.0	0.224	0.714	1.534	5.316	11.302
	Time+Sentiment	**0.465**	0.411	**0.624**	**0.868**	1.9	6.246	11.493
	Time+Relevance	0.426	39	0.583	0.864	1.839	6.617	12.444
	Time+Content	0.388	0.38	0.589	0.841	2.022	6.861	13.857
	All	0.407	0.387	0.595	0.842	1.971	6.652	13.111
r/conspiracy	Time	0.0	0.0	0.23	0.662	1.221	4.312	9.257
	Time+Sentiment	**0.643**	0.467	**0.638**	**0.858**	1.857	6.354	13.207
	Time+Relevance	0.51	0.444	0.642	0.851	2.232	6.787	14.156
	Time+Content	**0.618**	0.462	**0.635**	0.849	1.939	6.49	13.698
	All	0.563	0.457	0.602	0.823	1.92	6.559	14.994
r/news	Time	0.0	0.0	0.162	0.47	0.922	4.044	13.371
	Time+Sentiment	**0.679**	0.437	0.553	0.756	1.821	6.554	17.446
	Time+Relevance	0.524	0.392	0.54	0.741	2.068	7.615	19.966
	Time+Content	0.561	0.398	0.531	0.718	2.003	7.098	19.22
	All	0.541	0.401	0.514	0.703	1.993	7.233	21.405
r/science	Time	0.0	0.0	0.171	0.515	1.149	4.796	14.307
	Time+Sentiment	**0.502**	0.372	**0.542**	0.71	1.932	6.33	15.194
	Time+Relevance	**0.498**	0.388	**0.522**	0.693	1.9	6.476	16.388
	Time+Content	0.485	0.351	0.477	0.644	1.896	6.893	17.042
	All	**0.51**	0.384	**0.515**	0.69	1.919	6.497	14.506
r/todayilearned	Time	0.0	0.0	0.176	0.51	0.969	4.111	12.311
	Time+Sentiment	**0.714**	0.445	0.538	0.703	1.827	7.034	18.933
	Time+Relevance	**0.646**	0.437	0.55	0.734	2.127	7.14	18.014
	Time+Content	**0.647**	0.433	0.533	0.701	2.231	7.729	20.488
	All	0.627	0.436	0.54	0.717	2.033	7.466	20.202
r/worldnews	Time	0.0	0.0	0.153	0.462	1.05	4.045	13.259
	Time+Sentiment	**0.671**	0.451	**0.558**	0.705	1.912	7.183	19.847
	Time+Relevance	**0.575**	0.417	**0.526**	0.707	1.94	7.804	20.99
	Time+Content	**0.58**	0.429	**0.518**	0.707	2.06	8.173	21.387
	All	**0.601**	0.425	**0.534**	0.723	2.055	8.048	19.774
r/worldpolitics	Time	0.0	0.0	0.197	0.658	0.923	3.359	8.509
	Time+Sentiment	**0.722**	0.519	**0.653**	**0.844**	2.115	6.91	15.705
	Time+Relevance	0.491	0.427	0.626	0.835	2.637	8.218	17.197
	Time+Content	0.573	0.503	0.64	0.857	2.209	7.551	17.949
	All	0.615	0.489	0.625	0.842	2.06	7.594	17.487

very similar moderation restrictions. Particularly, we find that r/news prefers much less emotional comments, that are more authentic and less analytic than both r/worldnews and worldpolitics. Along with this, comments that are made long after the post submission in r/news may get higher score comments, while r/worldnews and worldpolitics have more short lived comment threads.

Audience can change behavior more than topic: The generality of a community's audience can also steer the community discussion. For example, we can see a distinct divide between the expert communities based on how niche the target audience is. The communities r/AskHistorians and r/askscience are built for explaining questions to a wide audience, while r/science and r/Bitcoin are build for discussions among a narrow audience (scientists and Bitcoin experts). Intuitively, r/AskHistorians and r/askscience prefer comments that are more lexically redundant, have less causation words (i.e think, know), and use more fluent or common terms. All of these features reflect the behavior of simply explaining the answer to a question. On
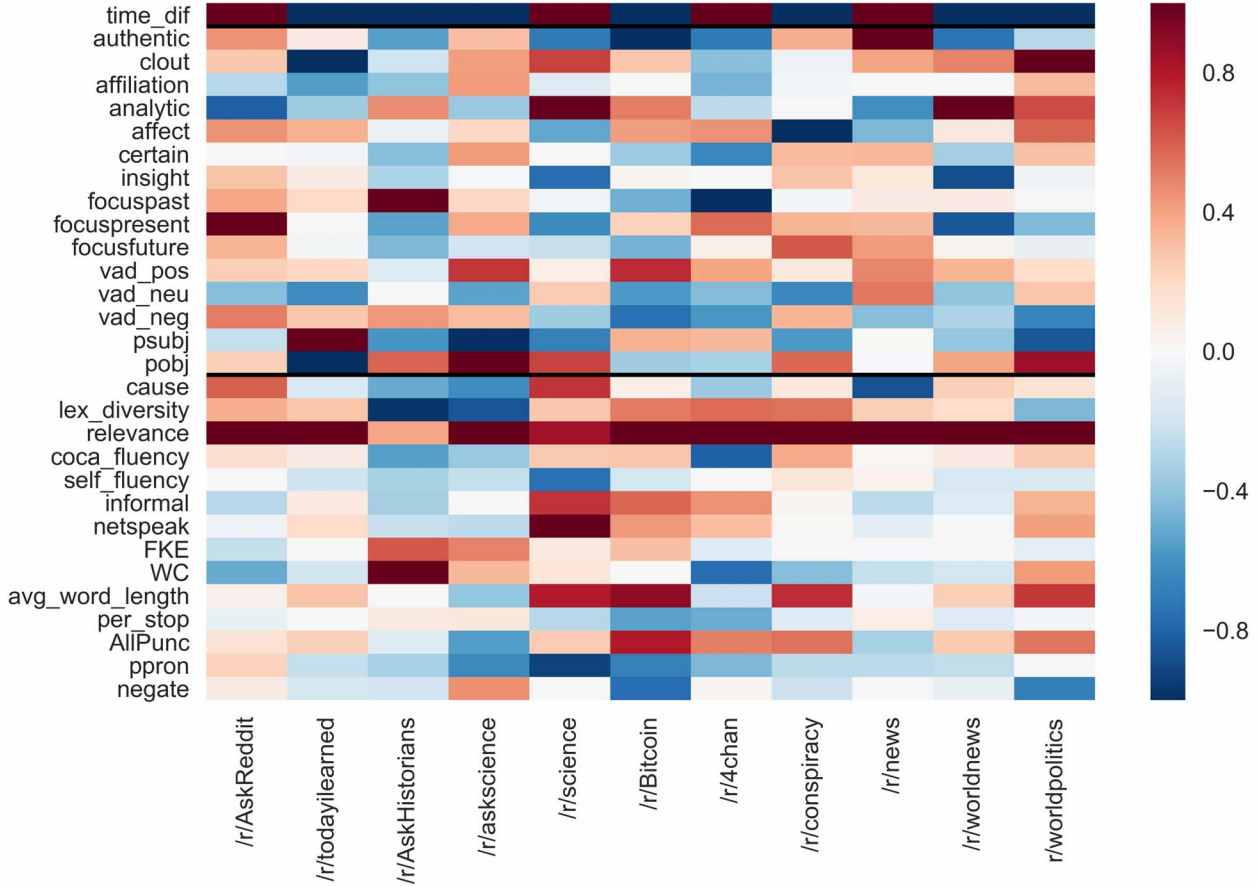
Fig. 2. A selection of the most important features according to the effect size using the KS statistic. For example, high scored comments in `r/AskHistorians`, `r/askscience` and `r/science` had lower subjectivity (psubj) given by blue and higher objectivity (pobj) given by red. The intensities signify their relative effect size among the sentiment features within each subreddit.
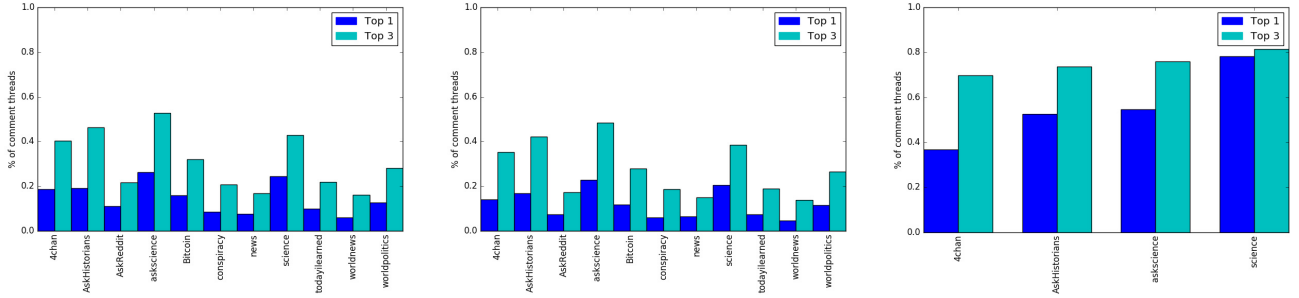

Fig. 3. Percent of comment threads with top comment by (**left**) the top h-index user (**center**) the top active user, and (**right**) by a user with a flair.

the other hand, we see `r/science` and `r/Bitcoin` prefer less lexically redundant comments that use more technical and analytic terms.

Even communities of the same type differ: It is clear from our results, and from the literature [4] that online communities differ vastly in discussion style and conduct. While we find many of the same features important in all communities, their direction of importance may change. For example, the authenticity of a comment is important across all subreddits; however, some prefer more authenticity, others significantly less. In general, emotion is important across

all subreddits, but some significantly dislike emotional comments, where as others significantly prefer emotional comments. Even very similar communities, like `r/worldnews` and `r/worldpolitics`, small differences can be found. `r/worldpolitics` uses significantly more netspeak and informal words, while `r/worldnews` uses more lexical diversity in discussion.

Highly active and high scoring users do not have a significant effect in the reception of comments, but existence of flairs do: We compute the percent of comment threads in which the top 1 and top 3 comments is by a user with the highest h-index,

the highest activity, or has a flair as shown in Figure 3. We find that both the local reputation and the local activity levels of a user have little impact on the score of the comment. This result is consistent with the literature [4]. More significantly, we find that users who have a flair have a high chance of being the highest score in a comment thread, especially if those flairs are expert flairs, despite only 1% to 5% of users having flairs (see Figure II). Further, we find that flaired users are more active overall and have a higher local h-index overall. In terms of expert flairs, this result is consistent with what we know about experts online [15]. We also test whether some flairs are more important than others, but find nothing significant.

VII. CONCLUSION AND FUTURE WORK

We confirm that it is possible to predict the score of comments, even when communities are unstructured, loosely moderated and noisy. Leveraging a carefully crafted set of features, our models outperform the state of the art model by a factor of 2, achieving high average precision and show that the relative rankings of comments remains close to the true ranking. Interestingly, the importance of features can vary vastly across communities. Despite this, we show some globally important features, such as relevance and emotion of comments and discuss potential reasons for the differences. Further, we show that user flairs can be an excellent predictor of highly popular comments, especially if those flairs are strictly controlled by moderators.

In the future, we would like to study what impact the users have on discussion, by conducting time-controlled experiments and collecting more user specific data such as years on reddit or if the user is a moderator of a community. Along with this, we would like to take a deeper look at the impact of different levels of moderation. While we show compelling results on the surprisingly small impact strict moderation has on discussion behavior, there may be many other behaviors impacted by moderation. We also would like to group different subreddits based on their similarities across our features. Another direction is to study how our findings change over time to see if changes to reddit scoring methods and other world events impact the findings.

VIII. ACKNOWLEDGMENTS

REFERENCES

[1] A. Bessi, F. Zollo, M. Del Vicario, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Trend of Narratives in the Age of Misinformation," *PLoS ONE*, vol. 10, no. 8, pp. e0134641–16, Aug. 2015.

[2] A. Bessi, F. Petroni, M. Del Vicario, F. Zollo, A. Anagnostopoulos, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Viral Misinformation - The Role of Homophily and Polarization." *WWW*, pp. 355–356, 2015.

[3] J. Maddock, K. Starbird, H. Al-Hassani, D. E. Sandoval, M. Orand, and R. M. Mason, "Characterizing online rumoring behavior using multi-dimensional signatures," in *ICWSM*, 2017.

[4] A. Jaech, V. Zayats, H. Fang, M. Ostendorf, and H. Hajishirzi, "Talking to the crowd: What do people react to in online discussions?" *arXiv preprint arXiv:1507.02205*, 2015.

[5] Y. Dahiya and P. Talukdar, "Discovering response-eliciting factors in social question answering: A reddit inspired study," *Director*, vol. 24196, no. 3295, pp. 13–61, 2016.

[6] H. Lakkaraju, J. J. McAuley, and J. Leskovec, "What's in a name? understanding the interplay between titles, content, and communities in social media." *ICWSM*, vol. 1, no. 2, p. 3, 2013.

[7] T. Tran and M. Ostendorf, "Characterizing the language of online communities and its relation to community reception," *arXiv preprint arXiv:1609.04779*, 2016.

[8] E. Gilbert, "Widespread underprovision on reddit," in *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 2013, pp. 803–808.

[9] T. Althoff, C. Danescu-Niculescu-Mizil, and D. Jurafsky, "How to ask for a favor: A case study on the success of altruistic requests," *arXiv preprint arXiv:1405.3282*, 2014.

[10] J. Hessel, L. Lee, and D. Mimno, "Cats and captions vs. creators and the clock: Comparing multimodal content to context in predicting relative popularity," *arXiv preprint arXiv:1703.01725*, 2017.

[11] S. Sikdar, B. Kang, J. O'Donovan, T. Hollerer, and S. Adah, "Understanding information credibility on twitter," in *Social Computing Conference (SocialCom)*. IEEE, 2013, pp. 19–24.

[12] B. D. Horne and S. Adali, "The impact of crowds on news engagement: A reddit case study," *arXiv preprint arXiv:1703.10570*, 2017.

[13] J. Reis, F. Benevenuto, P. O. de Melo, R. Prates, H. Kwak, and J. An, "Breaking the news: First impressions matter on online news," *arXiv preprint arXiv:1503.07921*, 2015.

[14] Y. Keneshloo, S. Wang, E.-H. Han, and N. Ramakrishnan, "Predicting the popularity of news articles," in *Proceedings of the 2016 SIAM International Conference on Data Mining*. SIAM, 2016, pp. 441–449.

[15] B. D. Horne, D. Nevo, J. Freitas, H. Ji, and S. Adali, "Expertise in social networks: How do experts differ from other users?" in *ICWSM*, 2016.

[16] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *ICWSM*, 2014.

[17] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: Liwc and computerized text analysis methods," *Journal of language and social psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[18] S. Wu, C. Tan, J. Kleinberg, and M. W. Macy, "Does bad news go away faster?" in *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[19] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *Association for Computational Linguistics (ACL)*, 2004, p. 271.

[20] M. Davies, "The corpus of contemporary american english: 520 million words, 1990-present." Available online at "http://corpus.byu.edu/coca/", 2008-, accessed: 7/21/2015.

[21] S. Lewandowsky, U. K. Ecker, C. M. Seifert, N. Schwarz, and J. Cook, "Misinformation and its correction continued influence and successful debiasing," *Psychological Science in the Public Interest*, vol. 13, no. 3, pp. 106–131, 2012.

[22] P. Singer, F. Flöck, C. Meinhart, E. Zeitfogel, and M. Strohmaier, "Evolution of reddit: from the front page of the internet to a self-referential community?" in *WWW*. ACM, 2014, pp. 517–522.

[23] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *ICLR Workshop*, 2013.

[24] C. Tan and L. Lee, "All who wander: On the prevalence and characteristics of multi-community engagement," in *WWW*. ACM, 2015, pp. 1056–1066.

[25] J. Hessel, C. Tan, and L. Lee, "Science, askscience, and badscience: On the coexistence of highly related communities," in *ICWSM*, 2016.

[26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[27] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. NY, NY, USA: Cambridge University Press, 2008.

[28] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1/2, pp. 81–93, 1938.

[29] M. J. Salganik, P. S. Dodds, and D. J. Watts, "Experimental study of inequality and unpredictability in an artificial cultural market," *Science*, vol. 311, no. 5762, pp. 854–856, 2006.