

# Toward Generalizing Comment Quality Prediction in Online Communities

Sadia Afroz and Rachel Greenstadt  
Drexel University  
sadia.afroz@drexel.edu, greenie@cs.drexel.edu

## ABSTRACT

With the rise of the social web, community discussion has become an integral part of the Internet. Many machine learning based approaches have been proposed to augment human-driven community filtering system. However, there is no research showing how approaches used in one commenting system can be applied on a different community with different demographics and commenting structure. The goal of this work is to generalize machine-driven community filtering systems. In this work, we analyzed comments and user activities of an online community, Jezebel.com. We demonstrated the underlying similarities in online communities by comparing our approach with other approaches that studied commenting systems on sites such as Slashdot and Digg. The novelty of our work is that we show how features used to predict comment quality in one community can be translated into predicting comment quality in another community, despite differences in community structure, commenter demographics and commenting practice.

## INTRODUCTION

Online communities are rapidly becoming the modern public square. Community filtering has the potential to make the space vibrant and useful and/or degenerate into a form of censorship. Content filtering mechanisms are necessary, without them, information overload reduces utility as users are overwhelmed by noise. How we filter, however, is not value-neutral and shapes the nature of online spaces and the opportunities they afford online speakers. The nature of filtering mechanisms determines which voices reach a large audience and which are silenced by obscurity.

Despite the best efforts of sites, both automated and crowd-sourced commenting systems are widely regarded as of poor quality<sup>1</sup>. Best practices in producing discussion worth reading still requires the tireless efforts of human moderators. However, this approach is expensive and does not scale well to large sites.

<sup>1</sup><http://www.theatlantic.com/technology/archive/2011/12/more-than-a-decade-in-and-internet-comments-continue-to-be-terrible/249379/>

In this work, we study community discussions performed at Jezebel.com using statistical machine learning. Jezebel.com<sup>2</sup> is a Gawker media site that publishes gossip, culture, and fashion news and allows user discussions. It is the most commented site among the Gawker Media sites<sup>3</sup>. Crucially, Jezebel depends on the work of trusted paid and unpaid moderators to promote (and ban) users and comments rather than a crowd-sourced approach.

Several studies have been done on the Slashdot and Digg community filtering systems [3, 8, 2, 7, 6, 10, 9]. We chose to study Jezebel.com for three reasons. First, Jezebel's commenting system is different from other online communities. Jezebel has dedicated moderators who check and approve comments before the comments are made public. It supports limited crowd-based moderation where only the trusted commenters can 'promote' or 'demote' a comment. However, unlike Slashdot or Digg, Jezebel does not aggregate crowd ratings. Comments from Jezebel are examples of human-annotated comment dataset, where a 'trusted' human user manually verifies the quality of each comment. Second, while Slashdot and Digg are predominantly male communities<sup>4,5</sup>, Jezebel represents one of the biggest feminist communities on the web with 2.1 million monthly readers. 95% of its users are female and 82% are of 18-34 age range<sup>6</sup>. The goal of this study is to understand how commenting practice varies from community to community depending on user demographics and whether or not similar features that were found useful in one community are salient in another community. Third, Jezebel is one of the many Gawker media sites with similar commenting system. By analyzing Jezebel, we can understand commenting system of many other Gawker sites. Recently, New York Times has also adopted a similar commenting system<sup>7</sup> by introducing 'trusted' commenters. Our analysis on Jezebel would be helpful in analyzing any online communities.

This work has three major novel contributions. (1) We demonstrate the underlying similarity in different online communities. We show that features similar to those used in crowd-sourced sites such as Digg and Slashdot also work well on a site with a different commenting system and different demographics, strengthening the argument for the generality of

<sup>2</sup><http://jezebel.com>

<sup>3</sup><http://jezebel.com/5310875/fasten-your-seatbeltsits-gonna-be-a-bumpy-sight>

<sup>4</sup><http://www.alexa.com/siteinfo/digg.com>

<sup>5</sup><http://www.alexa.com/siteinfo/slashdot.org>

<sup>6</sup><http://advertising.gawker.com/jezebel/>

<sup>7</sup><http://www.nytimes.com/2011/12/01/business/media/a-note-to-our-readers-about-comments.html?hp>

this approach. (2) We show how agents based on learning can potentially ease the load on human moderators, even in a largely human-curated site such as Jezebel. (3) Lastly, we suggest that using a human-curated data set such as Jezebel comments provides a useful and large, pre-annotated dataset to study social media and text quality.

The rest of the paper is organized as follows. We discuss related works in online community content analysis in section 2. We describe the Jezebel community and its commenting practices in section 3. We compare Jezebel with two other online community, Slashdot and Digg, in section 4. Our comment analysis approach and the features used for the analysis are described in section 5. Section 6 provides our evaluation which includes data collection, methodology, and results. Lastly in section 7 we discuss the result which is followed by conclusion and future works.

## RELATED WORK

The discourse quality of an online community has an important impact on the quality and topics of the future content of the site [4]. However, it is hard to create a community that self regulates. Content moderation and filtering in an online community helps to sustain enthusiastic user participation and a high-level of user satisfaction [13]. The popular news and opinion site HuffingtonPost.com receives approximately three million comments every month [4]. With the explosion of user-generated content on the web, some kind of automation is necessary to help readers make sense of complex online environments.

Several approaches have been studied to choose high-quality comments on Slashdot and Digg. One approach is to predict the community ratings of the Slashdot comments using a wide variety of features from the Slashdot metadata and posts' contents [3]. Another approach is to predict the Slashdot comment filtering threshold of a new reader by analyzing filtering choices of the existing readers [8].

Similar studies performed on Digg. Hsu et al. predicted the ranking of comments using comment visibility, user reputation and content specific features [6]. This work showed that features related to users' past history have the highest impact on comment ratings. Lerman studied Digg's article rating system to show the role of social networks in promoting content and predicted story ratings based on how the story was shared in the social network [9, 10]. Another approach is to collect discussion comments from a large number of people in the community and distill the information using text summarization [1].

Our work is most similar to that of Hsu et al. [6] and Brennan et al. [3], however, it differs in that we study a site with a different demographic base and commenting system. We study the Jezebel community to identify features that are unique to this community and use those to analyze the community practice. We perform a feature-based comparison of our work with the Slashdot and Digg studies and show that similar features are useful in predicting community ratings in all three of these communities.

## THE JEZEBEL COMMUNITY

Jezebel: "Celebrity, Sex, Fashion for Women. Without Airbrushing" is a women-focused online community launched on May, 2007 under the Gawker Media umbrella. Jezebel has a group of authors, and editors who publish news and blog posts. Jezebel allows community discussions. Each blog post has its own comment series. Anyone can comment on the site. If the commenter is new, her comment has to be 'approved' by the editors before it appears on the site. Jezebel also has a dedicated group of comment moderators who can moderate any discussion at any time. Due to the large number of comments, Jezebel implemented a 'star' system for users. A commenter who maintains a history of posting high quality comments is awarded a 'star.' A starred commenter's comment is not moderated in advance and is considered as a part of the featured discussion for a post. A 'good' comment is 'promoted' by another starred user or Jezebel editors. A comment can also be tagged as trollpatrol (offensive comments), cotd (comment of the day), complaints and bodysnark (making offensive remark about the bodies of others). In this paper we only considered promoted and regular comments. Comments that are not promoted or posted by unstarred commenters are generally hidden. The featured discussion of a post is loaded by default with the post. A reader can choose to read all comments by explicitly choosing the 'all' option at the end of a post.

A starred commenter can both approve new users' comments and promote other commenters' comments. The star privilege can be revoked if a starred commenter posts too many off-topic comments or approves comments of others that are considered inappropriate by the editors. An editor can also demote and delete comments and ban any commenters for posting low quality comments.

According to Jezebel commenting guideline<sup>8</sup>, an insightful comment that adds new information to a post or inspires an interesting discussion is considered a 'good' comment. Comments containing personal attacks on other commenters or Jezebel editors, vulgarity, self-promotion, banality, complaining (for example "I don't want to read about this, can't we see pictures of puppies?") are considered bad comments. Other bad commenting practices are "blog pimping" which means advertising someone's own blog, "body snarking" which means criticize female figures, usage of all caps, consecutive punctuation (for example, !!!??), excessive sarcasm, and overheated rhetoric<sup>9</sup>.

Transgressions of the commenting guidelines might result in 'disemvoweling,' which means removing the vowels from a comment, or deletion of the particular comment. If this behavior continues the commenter might get banned from using the site.

The Jezebel commenting system is different from that of Slashdot, Digg and Reddit in several ways. First, a comment does not have any explicit numeric quality rating. Second,

<sup>8</sup><http://jezebel.com/5053058/the-girls-guide-to-commenting-on-jezebel-version-12>

<sup>9</sup><http://jezebel.com/339424/this-year-lets-call-it-quits-on-the-nasty-nit-picking>

Jezebel does not aggregate community choices. Whether a comment should be promoted or not depends on the particular ‘star’ commenter who promoted it. We chose Jezebel to discover how salient features from crowd-rating based communities like Slashdot transfer to Jezebel.

A sample blog post<sup>10</sup> and promoted comment<sup>11</sup> based on it is given below:

*Excerpt from a sample post*

*Title: Tiger Mom Under Attack by Proponent of Controversial “Loving Your Kids” Approach*

Another Chinese-American mom—this one also a psych researcher—has come forward to challenge Amy Chua’s famous/infamous suggestion that “tiger” parenting is the best way to raise successful kids. Shockingly, she says moderation is actually best. According to ScienceDaily, Desiree Baolian Qin’s research on Chinese immigrant parents and their families have found that such parents often do adopt “tiger mom” tactics, pressuring their kids to excel and comparing them unfavorably to less-successful siblings. In another study, she found that Chinese-American students at an East Coast high school had higher levels of depression and anxiety—and lower self-esteem—than their white classmates.

*Sample promoted comment*

*plzpretypuss on 12 Jan 2012 11:02 AM*

Playdates? We do playing by appointment now? My mom used to take me to the playground and just set me free, let me roam like the buffalo. I’m not opposed to purposeful socialization, but I always got the impressions regimented or overly-scheduled activities weren’t good for children either.

I will always try to ascribe to George Carlin’s view on children - “If you wanna help your children, leave them the fuck alone!”

*\*promoted by SorciaMacnasty*

**COMPARISON AMONG JEZEBEL, SLASHDOT, AND DIGG**

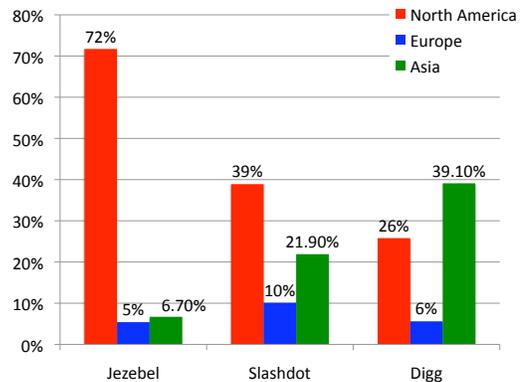
We compare Jezebel.com with two previously studied online communities: Slashdot.org and Digg, in terms of commenters demographics, commenting system and community structure. These factors drive the news and articles presented on the site and community discussions. Our results show that despite differences in these three communities, similar features are important in comment quality assessment in all three sites.

*User demographics*

Jezebel has a female dominated community with 95% female users. Slashdot and Digg are more male dominated (shown in Figure 2). Jezebel community is mainly North America based (71.7% are from North America) whereas Slashdot and Digg community are more diverse (shown in 1).

<sup>10</sup><http://jezebel.com/5875099/researcher-criticizes-tiger-moms-advocates-controversial-loving-your-kids-approach>

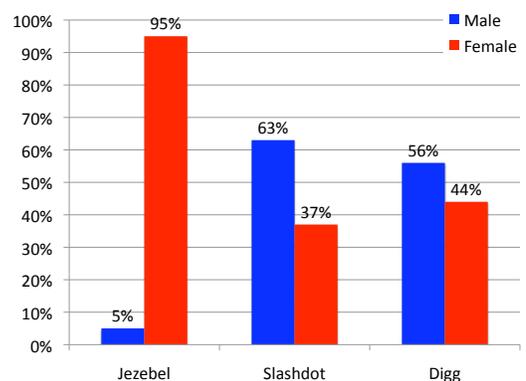
<sup>11</sup><http://jezebel.com/5875099/tiger-mom-under-attack-by-proponent-of-controversial-loving-your-kids-approach?comment=45935417#comments>



**Figure 1. Ethnic diversity of the commenters in Jezebel, Slashdot, and Digg**

*Commenting system*

Jezebel does not aggregate crowd ratings. Good comments are rewarded by a promotion. Only a few users with good reputation are given the power, a star, to promote a comment. A comment is considered ‘promoted’ if a single reputed commenter promotes it. Bad comments are not promoted and might get deleted or disemvoweled depending on how ‘bad’ it is. Slashdot and Digg aggregate crowd ratings. On Slashdot,



**Figure 2. Gender diversity of the commenters in Jezebel, Slashdot, and Digg**

each user can rate a comment using a scale from -1 to 5, with 5 signifying the comments most worth reading. Total rating of a comment is its final score. Comments with higher scores are considered better than the comments with lower scores. Comments that receive a very low score are typically hidden and comments with a higher score are highlighted.

On Digg user can rate each comment using a scale from -1 (thumbs-up) to +1 (thumbs-down). Total rating of a comment is its final score.

*Community structure*

Each community employs some mechanisms to reward good activity in the community. Jezebel awards stars to a selected

number of commenters based on their past activities and reputation. Starred commenters can promote/demote other comments and their comments are designated as “featured” comments.

On Slashdot, each commenter has “karma,” which is the total ratings of all of his past comments. Commenters with high karma are eligible to become moderators in the system and their comments are assigned higher ratings (+2) than default (+1). Digg does not have any explicit user ranking. A user’s activity rate, profile views, and joining date could be used as reputation features, though it is not clear whether or not Digg internally uses that.

## APPROACH

We represent each document as  $(\vec{x}, y)$  where  $\vec{x} \in \mathbb{R}^n$  is a vector of  $n$  features,  $n = 91$ , and  $y \in \{Promoted, Regular\}$  is the type of the document. For classification, we used Support Vector Machine (SVM) with Sequential Minimal Optimization (SMO) [11] implemented in the WEKA tool [5] with a linear kernel. We tested our dataset with other classifiers in the WEKA tool such as k-Nearest Neighbor, Naive Bayes, J48 Decision Tree, Logistic Regression and SVM with RBF kernel. We chose to focus on the SMO SVM as it outperformed other classifiers in most of the test cases.

## Features

We used three kinds of features to predict comment quality. The features are contextual, commenter reputation, and linguistic. The full list of features is shown in Table 1. We used similar features that were used to analyze Slashdot [3] and Digg [6] as our goal is to generalize comment prediction.

### Contextual features

We chose contextual features to understand the context in which the comment was made. In general, relevancy of the comment with the original post is the most important contextual factor. Irrelevant comments are not promoted. To see whether a comment is on-topic we calculated percentage of words in the original post, title of the post and tags of the post that also appear in the comment. Time difference between post and comment is another important feature as comments that are made long after the original post are less likely to be seen and get promoted. Rating of a comment also depends on whether or not the comment was made as a reply to another comment. Comments that are made as reply to another promoted comment or part of a thread with other promoted comments have a higher chance of being seen and getting promoted. From our manual analysis of Jezebel comments, we found that comments that contain funny graphics are often promoted. The same is true for comments that refer to other relevant news or posts. To check this we search for image html tag (img src) and URLs in a comment.

### Reputation and Activity features

Commenter reputation and activity features are chosen to understand commenter’s interest level for the community and her status in the community. We chose two reputation features to measure commenter popularity in the community: commenter type and number of promoted comments. In commenter type, we check whether or not the commenter is

starred. A starred commenter’s comment is by default considered as a featured comment, thus it has more visibility than a non-starred commenter’s comment. Number of promoted comments of a commenter represents how valued the commenter’s comment is in the Jezebel community. If a commenter has a history to posting high-quality comments, she has a high chance of getting a ‘star’ and her future comments have a high chance of getting promoted.

Commenter activity features are: commenter id, total comments made by the commenter, percentage of comments that are posted in reply to other comments, percentage of comments initiated new thread, average comments per post and average of the linguistic features of the comments.

### Linguistic features

Linguistic features measure the linguistic quality of a comment. Some of the features used are: readability index that measures the comprehensiveness of a text, lexical density that measures the ratio of content words to grammatical words in any given text (spoken or written), percentage of different parts of speech, average word length and total characters, words, and sentences.

## EVALUATION AND RESULT

In this study we conducted three experiments. First, we predicted comment quality. We distinguished promoted and regular comments. Our second experiment was to predict commenter’s reputation from her past activities. We predicted whether a commenter is starred or non-starred. The third experiment was to identify salient features in comments posted by commenters with similar reputation. In this experiment we analyzed comments that were posted by only non-starred and only starred commenters to find important features.

### Data Collection

We collected over 2 million comments from Jezebel.com which were posted from Dec 31, 2010 to Nov 1, 2011. These comments were made on 14193 posts by 22922 commenters. 61.53% of the comments were promoted comments and the rest were regular comments. 11.75% of the commenters are starred commenters and 86.65% are regular commenters. Other commenters were Jezebel editors, post authors and comment admins. 68.12% of the promoted comments were posted by starred commenters. Details of data statistics are shown in Table 2 and Table 3.

Table 2. Jezebel Data Statistics: Comments

Comment Type	Percent (Total)
Promoted Comments	61.53% (1239295)
Regular Comments	38.47% (775051)
Total Comments	2014167

For our experiments, we used a sample of 100k comments, half of the comments were promoted and the rest were regular comments. We split the dataset into ten subsets of 10k comments in each. We trained the classifier on one set of 10k comments and tested on another set of 10k comments. This process was repeated 10 times and the average was taken. We also trained and tested the classifier with lower number

Table 1. Feature set

FeatureType	Number	Description
Commenter reputation	17	Commenter id; commenter type (star, regular, author, editor or removed user); commenter activity such as total comments, total promoted comments, percentage of comments made in reply to other comment, average comments per posts, average of linguistic features of the comments, average comment time difference with original post
Contextual	9	Comment id, number of words also appear in the original post title, post body, and post tags; time difference with the post, reply or thread starter, contains image, url.
Linguistic	75	Total characters (with and without space), readability index, percentage of punctuation, consecutive punctuation marks, unique words, parts of speech, uppercase letters, all caps, self references ( I, me, my), percentage of words from different word groups such as assent words (agree, OK, yes), swear words (ass, damn, hell), tentative words (guess, maybe, perhaps), etc.

Table 3. Jezebel Data Statistics: Commenters

Commenter Type	Percent (Total)
Starred Commenter	11.75% (2695)
Regular Commenter	86.65 % (19862)
Removed User	1.08% (315)
Editors	less than 1% (16)
Authors	less than 1% (18)
Comment Admins	less than 1%(16)
Total Commenter	22922

of comments, but did not notice any significant performance improvement after 3k samples of the data.

### Experimental Results

#### Baseline

In comment quality prediction, we considered two baseline cases. The first is prediction using only commenter type (starred/non-starred) and the second is prediction using time difference between the post and comment. On Jezebel starred commenters’ comments are by default considered as part of the featured discussion, which is the default comment viewing choice. Thus, comments made by starred commenters have more visibility than the comment made by regular commenters. Time difference is chosen as a baseline as comments that are made long after the original post are less likely to be seen and get promoted. Our analysis shows that on average comments made after 120 minutes of a post are less likely to get promoted (figure 3).

In our dataset, only 81% of the comments can be correctly predicted using commenter type (shown in Table 5). Another important result is the effect of time difference between comment and post is not at all salient, as it performs less than random chance.

#### Predicting comment quality

In our dataset, our classifier had 84.8% accuracy in predicting promoted comments and 80.5% accuracy in predicting regular comments using all contextual, reputation and linguistic features, as shown in Table 4. This result is significantly better than the baseline cases (Table 5). According to our results,

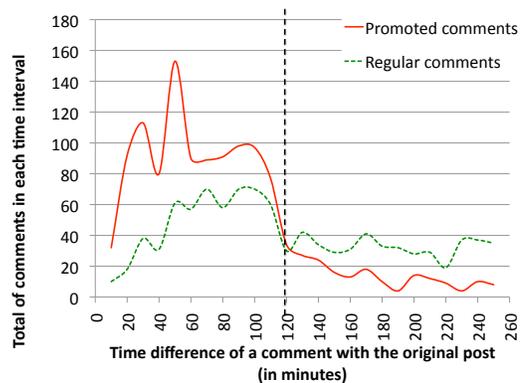


Figure 3. Time difference of a comment with the original post. On average, comments made after two hours (120 minutes) of a post are less likely to get promoted.

reputation and activity based features are the most salient features in predicting comment quality. These features can detect comment type with 83.8% accuracy (shown in Table 5). Linguistic features performed less than random chance.

Table 4. Comment prediction result (P = precision, R= recall, and F1= F-measure)

Comment Type	P	R	F1
Promoted	93.9%	77.3%	84.8%
Regular	75.5%	93.3%	83.5%

Table 6. Comment prediction result: only starred users (P = precision, R= recall, and F1= F-measure)

Comment Type	P	R	F1
Promoted	95.8%	99.6%	97.7%
Regular	88.9%	40.7%	55.8

We also performed comment prediction experiment with comment posted by only starred and only non-starred commenters. For comments posted by non-starred commenters, our classifier can predict promoted comments with 72.4% f-measure and not-promoted comments with 85.8% f-measure

**Table 5. Effect of features on comment prediction (P = precision, R= recall, and F1= F-measure)**

Feature Type	P	R	F1
All features	Promoted: 93.9% Regular: 75.5%	Promoted: 77.3% Regular: 93.3%	Promoted: 84.8% Regular: 83.5%
Reputation	Promoted: 92.5% Regular: 77.9%	Promoted: 78.8% Regular: 91.6%	Promoted: 85.1% Regular: 83.6%
Contextual	Promoted: 56.5% Regular: 0	Promoted: 100% Regular: 0	Promoted: 72.2% Regular: 0
Reputation and Contextual	Promoted: 92.3% Regular: 77%	Promoted: 79% Regular: 91.4%	Promoted: 85.1% Regular: 83.6%
Commenter type	Promoted: 84.6% Regular: 74.7	Promoted: 77.6 Regular: 82.3	Promoted: 81% Regular: 78.3%
Time difference	Promoted: 55.5% Regular: 0	Promoted: 99.9% Regular: 0	Promoted: 71.4% Regular: 0
Linguistic	Promoted: 56.5% Regular: 42.9%	Promoted: 96.8% Regular: 3.1%	Promoted: 71.4% Regular: 5.8%

**Table 7. Comment prediction result: only non-starred users (P = precision, R= recall, and F1= F-measure)**

Comment Type	P	R	F1
Promoted	82%	64.7%	72.4%
Regular	80.9%	91.3%	85.8%

(Table 7). For comments posted by starred commenters, majority (97.7%) of the comments were considered as promoted (Table 6).

#### Predicting commenter reputation

To predict commenter reputation we randomly chose 4000 unique commenters, 2000 of them were starred and 2000 non-starred. Using only the commenter activity features, an SVM classifier can predict non-starred commenters with 95.7% f-measure and star commenters with 87.5% f-measure.

#### Feature study

To understand the effect of different features in comment quality prediction, we ranked the features based on their Information Gain Ratio (IGR) [12]. IGR of a feature  $F_i$  is defined as,

$$IGR(F_i) = (H(D) - H(D|F_i))/H(F_i),$$

where  $D$  is comment class and  $H$  is entropy. We used WEKA to calculate IGR. Table 8 shows the top 10 features based on information gain ratio. Commenter activity and reputation features are found to be the most useful features in comment quality prediction. Of these, the previous history of promoted comments were found to be the most salient feature. We also performed a feature study with comments posted by only non-starred and only starred commenters, the top 5 features are shown in Table 9. Percentage of promoted comments is the most important feature for non-starred commenters, whereas comment id is the most salient feature for starred commenters.

#### Misclassification

Our classifier failed to predict comment quality when two comments with exactly same text posted by commenters

**Table 8. Top 10 features for predicting comment quality**

Feature Name	Gain Ratio
Starred commenters	0.275
Percentage promoted comments	0.208
Non-starred commenters	0.2
Commenter id	0.082
Authors	0.078
Total promoted comments	0.069
Commenter's total comments	0.044
Total character	0.037
Average comments per posts	0.026
Percentage of comment made as replies	0.025

**Table 9. Top 5 features for predicting comment quality posted by only starred and non-starred commenters**

Starred commenter	
Feature	Gain ratio
Comment Id	0.571
Percentage of promoted comments	0.326
Commenter Id	0.253
Average comment length	0.045
Percentage of comment posted as replies	0.003
Non-starred commenter	
Feature	Gain ratio
Percentage of promoted comments	0.178
Comment Id	0.15
Total promoted comments	0.083
Total comments	0.056
Percentage of thread-starter	0.053

with similar reputation were considered promoted and regular based on the context in which the comments were posted. For example, two non-starred commenters both posted ‘So cute!’ in a same post but in different discussion threads, one was promoted but another was not. Context and relevance of a comment are hard to measure. Sometimes a starred user promotes a comment only because she agrees with it or because it was post by someone who is a friend. The classifier also failed to predict quality of funny and sarcastic comments and comments with funny pictures.

#### Similarity among online communities

We compared our result with Slashdot [3] and Digg [6]. We chose these two studies because both of them performed similar studies on online communities using similar features we used. Interestingly, similar features were found to have similar effects in all communities, as shown in Figure 4. In all the studies commenter reputation features were found as the most salient features in predicting discourse quality and linguistics features were found as the least salient features.

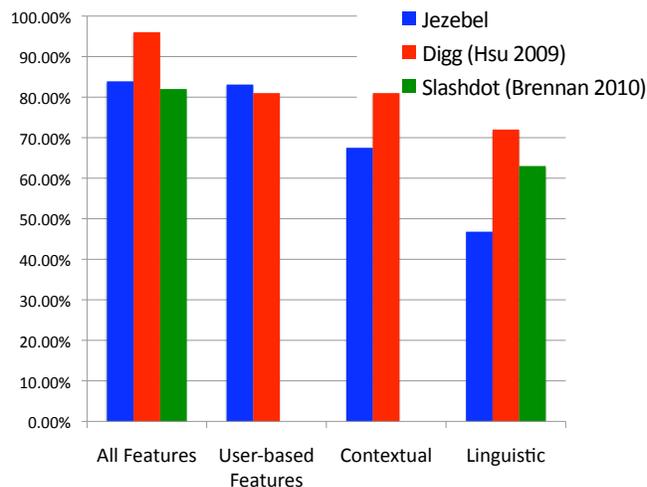


Figure 4. Effect of different features on Jezebel, Slashdot, and Digg. Similar features have similar effects on all online communities in predicting quality discourse.

However, not all features used in one community could be used in other communities. For example, on Slashdot, comments with more first person pronoun were rated higher and comments with swear words were rated lower. Such effects were not seen in the Jezebel community.

#### DISCUSSION

In this study we used machine learning methods to replicate community ratings automatically. Ratings of a comment in an online community depend more on the commenter’s reputation and past history and less on the content itself. One interesting future research direction is to understand how reputations develop over time and the way positive and negative reinforcement affect participation rates. By plotting the participation and reputation histories of a group of users over time,

we can identify good commenters and also provide quantitative data to augment the qualitative analysis.

During our analysis we noticed that a comment that disagrees with the opinion presented in the original post or in the popular discussion thread is hardly considered as high quality, despite of its high quality content. Automated comment prediction approach can be used to augment these shortcomings of current community filtering system by highlighting both machine-rated and user-rated high quality content.

The problems posed by socially intelligent community filtering systems require both insight into human behavior and social interaction as well as a keen awareness of how the technology can be changed. This work contributes to understanding of how values are embedded in technologies, how communities develop reputations and norms, and how socio-technical communities can combine human and machine computation.

#### CONCLUSION

In this work, we analyzed comments and user activity of an online community, Jezebel.com, to predict discourse quality and commenter reputation. Despite the fact that our community differed in comment regulation approach and demographic profile from communities that are traditionally studied, we showed that similar feature-types work across these communities. Our approach used statistical machine learning as a way to objectively gain insight into the workings of community filtering mechanisms. By extracting features that replicate their workings, we can better understand community filtering, improve the way the community uses the ratings of their members, and design agents that augment community decision-making. By analyzing a community such as Jezebel we showed the underlying similarity among online communities and demonstrated that it is possible to generalize machine learning driven community filtering approaches.

#### ACKNOWLEDGMENT

We want to thank Jezebel.com for providing us easy access to the public user comments. This work is funded by NSF Eager grant 1048515.

#### REFERENCES

1. Ackerman, M., Swenson, A., Cotterill, S., and DeMaagd, K. I-diag: from community discussion to knowledge distillation. *Communities and Technologies*, M. Huysman, E. Wenger, and V. Wulf, Eds. Kluwer BV, Deventer, The Netherlands (2003), 307–325.
2. Arnt, A., and Zilberstein, S. Learning to perform moderation in online forums. In *Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on*, IEEE (2003), 637–641.
3. Brennan, M., Wrazien, S., and Greenstadt, R. Learning to extract quality discourse in online communities. In *Workshops at the Twenty-Fourth AAAI Conference on Artificial Intelligence* (2010).
4. Diakopoulos, N., and Naaman, M. Towards quality discourse in online news comments. In *Proceedings of*

- the ACM 2011 conference on Computer supported cooperative work*, ACM (2011), 133–142.
5. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter 11*, 1 (2009), 10–18.
  6. Hsu, C., Khabiri, E., and Caverlee, J. Ranking comments on the social web. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, vol. 4, IEEE (2009), 90–97.
  7. Kaltenbrunner, A., Gomez, V., and Lopez, V. Description and prediction of slashdot activity. In *Web Conference, 2007. LA-WEB 2007. Latin American*, IEEE (2007), 57–66.
  8. Lampe, C., Johnston, E., and Resnick, P. Follow the reader: filtering comments on slashdot. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2007), 1253–1262.
  9. Lerman, K. Social networks and social information filtering on digg. *Arxiv preprint cs/0612046* (2006).
  10. Lerman, K., and Galstyan, A. Analysis of social voting patterns on digg. In *Proceedings of the first workshop on Online social networks*, ACM (2008), 7–12.
  11. Platt, J. Sequential minimal optimization: A fast algorithm for training support vector machines. *Advances in Kernel MethodsSupport Vector Learning 208*, MSR-TR-98-14 (1998), 1–21.
  12. Quinlan, J. Induction of decision trees. *Machine learning 1*, 1 (1986), 81–106.
  13. Wise, K., Hamman, B., and Thorson, K. Moderation, response rate, and message interactivity: Features of online communities and their effects on intent to participate. *Journal of Computer-Mediated Communication 12*, 1 (2006), 24–41.