Estimating Relative User Expertise for Content Quality Prediction on Reddit

Wern Han Lim
Faculty of Information Technology
Monash University
lim.wern.han@monash.edu

Mark James Carman
Faculty of Information Technology
Monash University
mark.carman@monash.edu

Sze-Meng Jojo Wong
Faculty of Information Technology
Monash University
jojo.wong@monash.edu

ABSTRACT

Reddit as a social curation site relies on its users to curate content from the World Wide Web (WWW) for the consumption of other users. Content on the site is enriched through user comments, discussions and extensions. This additional content is of varying quality however – ranging from meaningful information to misleading content; depending on the reliability, expertise and intention of the authors. Reddit relies on the Wisdom of the Crowd (WotC) from its community as well as selected moderators to manage its content. We argue that this approach suffers from the cold start in collecting user votes and is at risk of user bias, particularly a group-think mentality. Besides that, managing the large collection of content on Reddit is expensive. In our study, we explore the estimation of relative user expertise through various content-agnostic approaches. We show that it is possible to infer information quality on Reddit using the expertise of the authors. This prediction of content quality could lead to an improved organisation of Reddit content (re-ranking) for user consumption and future information retrieval.

KEYWORDS

Reddit; User expertise; Information quality; Knowledge management; Information retrieval

1 INTRODUCTION

Content aggregation platforms are one-stop-fits-all destinations for information – where quality content from the World Wide Web (WWW) is identified and curated for the consumption of users. Such platforms have prospered since the beginning of the Internet (notably Slashdot[1], Digg[2] and Reddit[3]) commanding a large amount of internet traffic [30]. These platforms impact the WWW by:

- Identifying interesting content on the Web [30] that meets the information needs of the users. Popular and novel content [21, 33] often receives increased visibility [12].
- Directing a large volume of user traffic to the shared content. This is termed as the "Slashdot Effect" – a Flash Crowd occurrence [4, 7] where sites linked from Slashdot's stories receive a sudden swell in traffic.
- Playing a role in disseminating of information [8, 12], especially through influential users [20] to great effect [19]. Influence spreads beyond individual platforms and linked sites; shaping how stories can be viewed, enhancing content and creating new viral content (aka memes) that are picked up by mainstream media.
- From a sociological perspective, collaborative news aggregators provide an understanding in regard to the interest and importance that users place on topics – this characterises the Internet Zeitgeist [7].

These platforms have evolved over the years particularly in loosening the responsibility from hired editors to the user community of the platforms. This shift towards social curation discussed in Section 2 poses new challenges in the organization of large and diverse content with varied quality. One important content aggregation platform is Reddit which has gained popularity [29] to the point where it is sometimes referred to as the "Front Page of the Internet". We conducted a study of Reddit in Section 3.

This research proposed content-agnostic approaches towards the estimation of user expertise in Section 4. By estimating the user expertise, one can infer the quality of content – users of higher expertise have a higher probability to produce content of higher quality. The prediction of content information quality could then be applied to the improve management of content on social curation platforms; validated through an experimentation setup in Section 5. We analyse the findings in Section 6.

2 SOCIAL CURATION

Over the years, content aggregation platforms have shifted the responsibilities of content curation and creation from expert editors onto users acting collaboratively as a community. Slashdot was one of the earliest content curation platforms. Slashdot editors would select, edit, summarize and aggregate news or stories from the WWW for the platform.

Digg experimented with giving its users more responsibility as contributors on the platform, allowing its users to submit content. The submitted content were moderated through peer-moderation where users vote for content – actions coined as *digging* with a up-vote known as a *dig* and a down-vote known as a *bury*. These votes indicate how interesting the users find a piece of content [18]. When a user votes for an item of content, the content is shared to the voter's followers as recommendations [13] sorted by temporal order. Content on the highly visible front-page of Digg is still

[1] https://slashdot.org/
[2] http://digg.com/
[3] https://www.reddit.com/

HT '17, July 04–07, 2017, Prague, Czech Republic.

DOI: http://dx.doi.org/10.1145/3078714.3078720

however selected and managed by the editors of Digg; resulting in a competition between users for attention towards their contributed content.

Today, Reddit has built its reputation as one of the biggest websites with approximately 450 millions page views daily [30]. The growth in Reddit's adoption is through a large migration of users from Digg in 2010 following Digg $v4$, a phenomena known as the Great Digg Migration. The migration was largely motivated by the increased user empowerment offered by Reddit where users have the capability to create communities (*subreddits*), aggregate content and moderate the content with up-votes or down-votes. Users on Reddit are equals unlike the segregation of roles on Digg into editors, moderators, power users (who were perceived by Digg users as having too much power in $v4$ of the website) and anonymous users [20] – each with their own privileges including the number of votes possible and the magnitude of the votes. Without such roles, it is up to the user community to aggregate, curate and moderate the content including determining which content are pushed to the front-page.

2.1 Content on Social Curation Platforms

Often times, the lack of user contributions on social platforms can be of concern in order to achieve a good coverage of the Web. This is however not a challenge for many of the discussed social curation sites where we can observe a large amount of content which is very diverse in nature and topic domain [29]. Content threads are shared in the form of (1) a link to an external site such as news articles or image hosting services which define the platform role as a content aggregator; or (2) textual content written by the authors themselves as a curation platform like Reddit transitions towards a hybrid discussion board [29].

2.2 Organisation of Content

A very large amount of content exists on social curation sites. Over 60 million threads were submitted to Reddit between 2008 and 2012 [29] and we found that popular subreddits such as the *r/gaming* have an average of $1,257$ threads submitted daily[4]. The volume of data creates a need for the management and ordering of content on these platforms for user consumption and peer-moderation.

The ordering of content matters. Users were found to have a positional bias for links [6], where users were observed to display consumption patterns from the top of a webpage or a list [15] before proceeding down to the bottom [21] – top items having greater visibility for user consumption. Thus, the content on the users' stream should be sorted to better direct the users' attention towards content of good quality [21]. We note that a thorough study of how best to rank content on a user's stream is currently lacking for Reddit [30].

2.2.1 Peer Recommendation. Prior works on social curation platforms mainly look into the recommendation of content (particularly on Digg) and focused on analyzing the visibility and attraction of shared content. All the content competes for the general users' attention to be on the front-page of the platform. In Digg, recommendations are fed directly to a user's followers' content stream whenever the user votes for review of a content [13]. The items are sorted according to temporal order. The motivation for the feed is from the discovery that users do follow other users for content [19]. Researchers have looked to improve on the recommendation performance of Digg with user models [14] for identifying similar users. Users would then be pushed content that is highly rated by similar users as personalised recommendation [20], via a peer recommendation process [21] to better meet these information needs.

2.2.2 Content Popularity. A common approach in the ordering of content is to sort the content according to popularity. Studies have found that the popularity of items on Digg is largely influenced by quality [20], with content of high popularity found to be of higher quality than less popular ones [30]. It should however be noted that the novelty of the content degrades overtime [33] and that the popularity of Reddit content does not equate to quality. For example, studies have shown that popular image threads are often reposts of earlier threads with identical links often ignored by the community before achieving popularity [9].

Content popularity can be determined by looking at the number of votes garnered by the content [18]. It is however difficult to predict the future popularity of an item [31], particularly new content due to the cold start problem. The only available data initially comes from the content itself which are weak features for prediction [5].

2.3 Challenges in Social Curation

Social curation on Reddit faces multiple challenges which motivate our work in Section 4 for the estimation of user expertise.

2.3.1 Managing high amounts of unstructured content with varying quality. Users are of different expertise levels and produce or submit content based on their capabilities. This creates a diverse spread of content – some of which may not be suitable for consumption. If not properly managed, such misinformation could spiral out of control [8]. Thus, there is a need to infer the information quality of content for improved management of content.

2.3.2 Vulnerability to malicious users. As an open platform, users are welcomed to be a part of the communities – to contribute and help moderate the platform [21]. The existence of malicious users however can undo such efforts. For example, these users could game the system by creating proxy accounts for self promotion or coordinate attacks for propaganda [25]. By early identification of such users, social curation platforms could defend against such attacks.

2.3.3 Obtaining sufficient user interactions. Often, traditional social platforms rely on the user's personal motivation for contributing content [9] with the vast majority of users only consuming content without submitting any content on Reddit [26]. In our study of Reddit (Section 3), we see a high amount of user activities in creating threads and commenting on them. While large subreddits are immune to this, smaller subreddits and the community do suffer from the lack of user activities as the user interactions are centred around the larger subreddits [29]. We also note from the same study however that there is a lack of user votes for the moderation of content where the median score for threads and comments has the

[4]Figure from our dataset in Section 3.1

value 1 (Table 4 and Table 5). This deficiency in user review greatly affects the content moderation itself [18] on Reddit. Through the estimation of user expertise of content authors, it is possible to predict the information quality of the content despite the lack of user votes.

2.3.4 Vote bias. User bias does exist on social platforms, particularly in the judgment of content through their votes. Users tend to vote for the content which they have viewed early [32] and might not continue to browse the remaining content; resulting in the content not being judged well when the content is sorted by temporal order. Besides that, users can be biased by the "herding" effect – where the community's views on the content [28] such as the current votes gained by the content or an early swell in user votes that affect the decision of the users [27].

3 REDDIT

Reddit is a social curation platform that relies solely on User-Generated Content (UGC) which it provides for the consumption of its users – both the role as aggregators of content from the Web and also contributors of additional information to the discussions within threads. UGC has the information potential to enrich content especially on platforms like Reddit [30]. This research attempts to overcome the challenges presented in Section 2.3.

3.1 The Reddit Dataset

The study of Reddit is based on a publicly available dataset hosted on a Google BigData repository[5] from the *r/datasets* subreddit[6]. From this dataset, we made use of threads and comments for various subreddits between December 2015 and and July 2016; as downloaded on 27th September 2016[7].

3.2 Organization and Subreddits

Reddit introduces user-defined communities known as subreddits, $Z = \{z_1, z_2, ..., z_{|Z|}\}$. Users can create and manage the communities for content aggregation, sharing or consumption. Many of these subreddits are often topic-based. It is within the subreddit that threads are created for the sharing and discussion of content.

From the repository, this research investigated threads from the highly subscribed subreddits of rank 5th, 6th, 10th, 17th and 55th [8] – *r/science*, *r/worldnews*, *r/gaming*, *r/explainlikeimfive (elifive)* and *r/politics*. As we discuss below, user behaviour on these subreddits differ from one another; thus user expertise are estimated relative to that of the other users within the same subreddit.

The quantity of threads created daily differ between subreddits as shown in Figure 1. We note that the *r/science* subreddit recorded the lowest number of threads created daily and *r/gaming* the highest.

The number of threads created on these subreddits are relatively stable with the exception of the *r/politics* which fluctuates with regards to the political scene. Upon further inspection, we note that this subreddit despite its name is a subreddit only for the United States (US) political news. Thus, this contrasting behaviour would be an interesting exploration point against that of the global

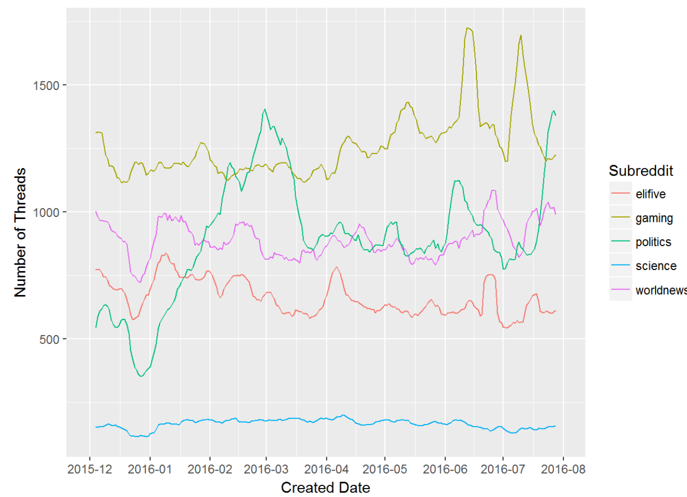

Figure 1: Moving Average (7 Days) of Reddit Threads Created

r/worldnews subreddit in future studies. The number of threads created is very stable for the *r/science* subreddit with little fluctuations when compared to the other subreddits.

From the plot, a consistent dip in the number of threads created towards the end of December 2015 and earlier January 2016 can be observed. Presumably, the holiday season has an effect on user activity – users are less active during such period with the exception of the *r/gaming* subreddit with holiday releases.

Such observations encourage the research to explore the possible performance difference between expertise estimation algorithms across these subreddits; each with varying amount of user activity and available information for estimation.

Table 1: Thread Density of Subreddits by Day

	Number of Threads by Day						
Subreddit	Minimum	1st Q	Median	Mean	3rd Q	Maximum	Total
elifive	446.0	593.0	648.0	663.2	717.8	1484.0	161827
gaming	973.0	1161.0	1227.0	1257.0	1328.0	2355.0	306735
politics	190.0	673.8	904.0	894.3	1061.8	1728.0	218210
science	65.0	118.8	174.0	164.8	199.0	263.0	40215
worldnews	503.0	774.8	922.0	892.8	1006.5	1567.0	217834

3.3 Reddit Threads

Reddit threads, $D = \{d_1, d_2, ..., d_{|D|}\}$ can be created by any users on any subreddit. These threads begin with a submission of an external link or self-written text [29]. Users including the thread starter can then interact with the thread by posting comments, critics and questions at any time point. Unlike many other UGC platforms, Reddit allows users to interact with other user comments in threads; creating a tree-like structure of various levels for the discussion. The structure of a Reddit thread is as shown in Figure 2.

The thread-based user interactions create a new challenge in estimating user expertise which differs from many other UGC platform such as community question-answering platforms [23]:

[5] Compiled by Reddit user *u/Stuck_In_the_Matrix*
[6] https://www.reddit.com/r/datasets/comments/3bxlg7
[7] Leaving a temporal gap for the threads and comments to stabilise
[8] As of 2nd September 2016 with healthy number of threads

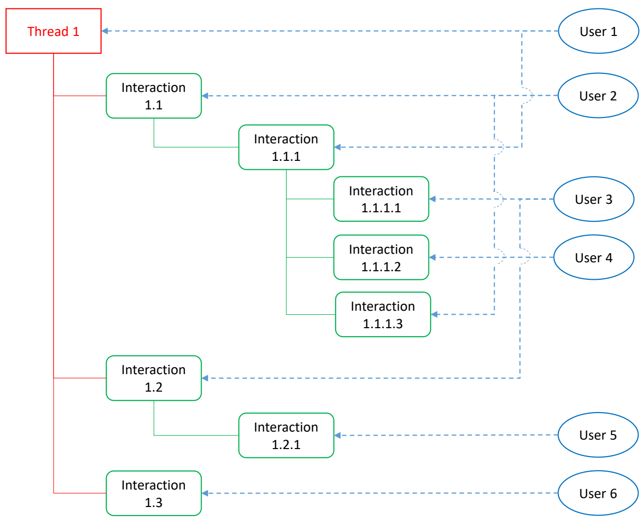

Figure 2: Structure of a Reddit Thread

- Multiple interactions from the same user.
- Not every comment in a thread directly respond to the content posted by the thread starter.
- Replies can both enhance, agree or disprove other comments.

3.4 Reddit Comments

Comments, $I = \{i_1, i_2, ..., i_{|I|}\}$ can provide additional meaningful information to the thread discussion. The distribution of user comments on threads do vary however and is very skewed to the right as we observed in Table 2. An interesting observation to be made here is that the threads in *r/explainlikeimfive* and *r/politics* would often receive at least 1 user comment; and on the other hand, at least half of the *r/gaming* threads have no user response.

Table 2: Comment Distribution of Reddit Threads

Subreddit	Number of Comments on Reddit threads					
	Minimum	1st Q	Median	Mean	3rd Q	Maximum
elifive	0.0	1.0	2.0	8.5	5.0	6105.0
gaming	0.0	0.0	0.0	9.9	2.0	9502.0
politics	0.0	1.8	3.0	43.8	15.0	47831.0
science	0.0	0.0	1.0	15.8	2.0	6370.0
worldnews	0.0	0.0	1.0	22.5	2.0	32901.0

Diving deeper by looking at the comment word count, it can be noted that the comments for *r/gaming* threads are usually short. The question-answering nature of threads in *r/explainlikeimfive* encourages more complete and elaborate responses which resulted in a higher word count for user comments.

3.4.1 Direct Comments. Direct comments are comments that respond directly to the thread. These comments could directly comment on, discuss or argue with the content shared by the thread starter. If the thread itself is a question, the comments would attempt to answer it. In Table 3, we observe that the comments in *r/politics* and *r/worldnews* are often responses to other comments in a discussion thread unlike the other subreddits here.

Table 3: Comment Types of Reddit Threads

Subreddit	Number of Comments		Correlation between Direct and All Comments in Threads	
	Direct	All	Pearson's	P-value
elifive	502010 (34.29%)	1463800	0.8957	$< 2.2 \times 10^{-16}$
gaming	524804 (36.23%)	1448582	0.9178	$< 2.2 \times 10^{-16}$
politics	1842373 (18.17%)	10137258	0.8371	$< 2.2 \times 10^{-16}$
science	210202 (30.78%)	682810	0.9267	$< 2.2 \times 10^{-16}$
worldnews	969053 (18.41%)	5263228	0.9413	$< 2.2 \times 10^{-16}$

There is a strong correlation in the number of direct comments with the number of comments in every Reddit thread. An interesting observation here is that despite the threads in *r/worldnews* having a much lower number of direct comments when compared to the other subreddits, it does record the highest correlation here telling us that it is very discussion-focused with users replying to each other a lot more.

3.5 Reddit Votes and Vote Difference

Reddit facilitates the peer moderation of content through user votes which are often regarded as the community's long-term judgement on the quality of content on many UGC platforms [1] – Up-votes (positive votes) to indicate content which they deem to be good, and down-votes (negative votes) to be poor. Often, negative votes are abused particularly for cyberbullying and thus are removed from many platforms [11]. Earlier studies have concluded however that the existence of down-votes does not negatively affect Reddit [25].

As a prevention of vote abuse by malicious users, Reddit is no longer displaying the number of up-votes and down-votes gained by a content (since June 2014) [16]. Instead, content votes are calculated internally in Reddit according to the up-votes and down-votes received through a function they called Vote Fuzzing (VF). The VF mechanism artificially inflates the number of up and down-votes for a resource, while maintaining the true difference between them which is then attached to the content itself. In this paper we extract user vote difference, Vote(I), as a content-agnostic indicator to estimate user expertise.

Starting with the threads themselves in Table 4, it can be noticed that there are no Reddit threads with negative vote difference. The vote difference distribution for Reddit threads are heavily skewed to the right particularly for the *r/explainlikeimfive* and *r/gaming*. On the other hand, the *r/politics* has almost half of its threads without any vote difference whereas the voted threads have a higher vote difference when compared to the other subreddits.

Unlike the threads, we observed Reddit comments with negative vote differences (shown in Table 5). These negative votes are however not dominant as they account for less than a quarter of the comments. We look to extend and adapt the algorithms from earlier work done on community question-answering data [23] to account for the negative vote difference. In general, most of the comments tend to have a relatively low vote difference value within the region of 1 to 3.

Table 4: Vote Difference Distribution of Reddit threads

	Scores of Reddit threads					
Subreddit	Minimum	1st Q	Median	Mean	3rd Q	Maximum
elifive	0.0	1.0	1.0	17.5	1.0	7074.0
gaming	0.0	1.0	1.0	57.7	1.0	9942.0
politics	0.0	0.0	1.0	111.2	12.0	9605.0
science	0.0	1.0	1.0	108.6	5.0	10924.0
worldnews	0.0	1.0	1.0	62.5	4.0	11850.0

Table 5: Vote Difference Distribution of Reddit comments

	Scores of Reddit comments					
Subreddit	Minimum	1st Q	Median	Mean	3rd Q	Maximum
elifive	-3276.0	1.0	1.0	6.8	2.0	8537.0
gaming	-865.0	1.0	1.0	9.5	3.0	6243.0
politics	-1763.0	1.0	1.0	5.2	3.0	7098.0
science	-206.0	1.0	1.0	6.9	3.0	5313.0
worldnews	-3032.0	1.0	1.0	8.0	3.0	6917.0

3.6 Reddit Users

Reddit users are integral to Reddit as a social curation platform – users share external content via thread links or produce their own textual content through self texts and comments. There are healthy number of users for each subreddit seen here in Table 6.

We however observed that user behaviour differs between subreddits which motivates this research to investigate the estimation of user expertise for these diverse subreddits. Unlike the other subreddits, the users of *r/explainlikeimfive* and *r/gaming* are active thread contributors, with almost 20% of the users of *r/gaming* only contribute to threads without any comments. Most users are active commentors on these subreddits particularly the users of *r/politics*, *r/science* and *r/worldnews*. Only a small portion of users in *r/science* and *r/worldnews* are active contributors of threads and comments. It is interesting that despite the question-answering nature of *r/explainlikeimfive*, 12.72% of its users are both thread starters and commentors; a phenomena explainable through the assumption that the thread starter themselves would post comments to further seek information, such as clarification from the other comments.

Table 6: Activities of Reddit Users

	Number of Users				
SubReddit	Threads	Comments	Threads only	Comments only	Total
elifive	81965 (27.55%)	253324 (85.16%)	44139 (14.84%)	215499 (72.45%)	297464
gaming	124630 (29.84%)	335491 (80.32%)	82221 (19.68%)	293083 (70.16%)	417713
politics	37914 (11.82%)	304619 (94.99%)	16067 (5.01%)	282773 (88.18%)	320687
science	13387 (9.47%)	131269 (92.84%)	10126 (7.16%)	128009 (90.53%)	141396
worldnews	48566 (9.56%)	476480 (93.81%)	31453 (6.19%)	459368 (90.44%)	507934

The distribution in the amount of items contributed by the users is heavily skewed to the right, as we observe at least 60% of the users of the explored subreddits created only 1 thread or made only 1 comment. At the 90th user percentile, they only created between 2-6 threads and 3-4 comments.

4 ESTIMATION OF USER EXPERTISE

This research attempts to overcome the challenges discussed in Section 2.3 with the estimation of user expertise through content-agnostic approaches and then apply it to predict the information quality of user-generated content. To the best of our knowledge, there exists no research that aims to predict the contribution quality of Reddit users through their estimated user expertise and instead all prior work focuses on the processing of the content itself [17, 30]. It should be noted that our work here can be used to supplement content-based approaches. The contribution from this research includes:

- Improved management of content according to the content's information quality. As a content-agnostic approach, the content quality could be predicted despite its unstructured nature.
- Reduces the platform's vulnerability towards malicious users. The estimated user expertise could be used to detect potential malicious or unreliable users earlier and reduce their influence from the get-go.
- Provides an early indicator of content quality which overcomes the cold start problem in garnering user votes for content moderation. The organisation of content can then be adjusted to improve peer-moderation.

Thus, the quality content can be promoted in the list for the consumption of users (since users do consume content from the highly visible top of a list to the bottom [15, 21]). This is the alternative to the temporal sorting of content that does not overcome the cold-start and is instead susceptible to temporal bias [9].

4.1 User Expertise on Reddit

Users of higher expertise tend to produce content of higher information quality [16]. This observation has encouraged works into identifying users of high expertise within domains – a process known as expert search [35] in order to improve content quality [22]. For this research on Reddit, the estimated user expertise is:

- **A relative measure.** The user expertise is estimated relative to other users. Comparing the expertise between two users enables us to infer the odds that one user would produce a better content than another user.
- **Domain sensitive.** A user cannot be an expert in every domain. Thus the user expertise is estimated within the domains where the user is active – i.e. we compute a different value for each subreddit.

This research looks at the content-agnostic approach towards user expertise estimation. In the following sections, we discuss the explored approaches which revolves around the central theme of user expertise according to their contribution significance.

4.2 Contribution Count (C-Count)

This is the simplest approach. It makes the assumption that expert users are those users who make a large number of significant contributions as judged by the vote difference. Thus, the approach counts the number of good contributions made by the users as their estimated user expertise.

4.2.1 Identifying Good and Bad Contributions. The vote difference gained is used to measure the contribution of each comment. The contribution significance of each comment, $\text{Sig}(I)$ can be judged within each thread according to:

- **Polarity.** A user comment is a "good contribution" if the vote difference gained by that comment is positive; and negative otherwise. This measure of polarity would however suffer from user voting bias where users are inclined to vote positively for good content but not negatively for poor content [28] – motivating us to explore the other variants.
- **Median.** The median vote difference over all comments in a thread can act as threshold to judge if a comment's contribution is significant. A contribution is good if it is "more significant" than half of the other contributions of the thread; consistent with the relative nature of our estimated user expertise. Moreover, this variant overcomes the imbalanced ratio of positive vs. negative vote difference in separating good and bad contributions.
- **Median Direct.** This variant is an extension of the median approach which uses the median vote difference of all direct comments instead.
- **Mean.** Similar to median variant, we adjust the threshold to consider the significance of the user comment. This approach could however punish the users when there is a non-uniform distribution over votes.
- **Mean Direct.** The threshold is adjusted according to the vote difference mean of direct user comments.

4.2.2 Punishing Bad Contributions? If a user produced a bad contribution $\text{Sig}(I) < 0$, the user could be punished. This research determines the effect from punishing users for their bad contributions by counting the following as the user's estimated expertise – (1) Only the number of good user contributions as in Function 1; or (2) The difference between the number of good vs. bad contributions as in Function 2.

$$\text{C-Count}(u) = |\text{Sig}(I_u) \geq 0| \tag{1}$$

$$\text{C-Count}(u) = |\text{Sig}(I_u) \geq 0| - |\text{Sig}(I_u) < 0| \tag{2}$$

4.2.3 Decay. User expertise changes overtime as users grow to become better experts in their area. Thus, the user's latest interactions would provide a better judgment for expertise estimation. We update a user's estimated expertise at temporal time t, $\text{C-Count}_t(u)$ by a power of the decay factor $\lambda = 1.0, 0.9, 0.5, 0.1$ on the earlier measure as shown in Function 3.

$$\text{C-Count}_{t+1}(u) = \text{C-Count}_t(u)^{\lambda} + |\text{Sig}_t(I_u) \geq 0| \tag{3}$$

4.2.4 Multiple Interactions. A Reddit user can contribute one or multiple comments in a thread. Here, this research takes into account all of these interactions. If a user is active, then it is up to the user to ensure that the comments made are "good contributions", or suffer the consequences otherwise. This research takes this stance as to not discourage user activity which is vital towards to generation of content on the platform. Besides that, this enabled the research to study the effect of user activity on user expertise estimation.

4.3 Contribution Z-Index (Z-Index)

The Contribution Z-Index is another baseline approach for this research as a possible improvement over the earlier Contribution Count approach. It is based on the Z-Index used in Community Question-Answering (CQA) platform [34] adapted for Reddit – it considers how many times more a user has made a good contribution rather than a bad one; based on the assumption that a random user is just as likely to make a good contribution as they are to make a negative one. Thus the count of positive vs negative contribution should follow a Binomial distribution. The Z-Index measure is used as the estimated user expertise. This approach also includes all the variants discussed earlier.

$$\text{Z-Index}(u) = \frac{|\text{Sig}(I_u) \geq 0| - |\text{Sig}(I_u) < 0|}{\sqrt{|\text{Sig}(I_u) \geq 0| + |\text{Sig}(I_u) < 0|}} \tag{4}$$

4.4 Contribution Scores (C-Score)

In this baseline, this research measures the significance of user contributions as a score. The vote difference of user contributions, $\text{Vote}(I)$ in the range of $(-\inf, \inf)$ are used to infer contribution [24] and experts are users with high collected scores from their comments. In this paper, we utilize this additional information as a measure of how good or bad user interactions are in contributing towards the estimation of user expertise.

4.4.1 Scores as Expertise. From the collected contribution scores, the user expertise can be estimated. Here, this research look at 2 variants – (1) sum of contribution scores as shown in Function 5; and (2) average of contribution scores shown in Function 6. In the sum variant, users are rewarded for being active. Similar to the Contribution Count and Contribution Z-Index; we decay the contribution scores of the earlier user comments before including the newest user comments for the estimation of user expertise.

$$\text{C-Score}(u) = \sum_{k=0}^{|I_u|} \text{Vote}(i_k) \tag{5}$$

$$\text{C-Score}(u) = \frac{\sum_{k=0}^{|I_u|} \text{Vote}(i_k)}{|I_u|} \tag{6}$$

4.4.2 Contribution Adjustments. The implementation of contribution scores can be tricky due to the varying number of user comments and the number of votes within a Reddit thread. Thus, there would be a need to adjust the contribution scores according to the thread where the interaction is. The variants explored here are:

- **Raw.** The vote difference is taken as the user's contribution score without any adjustments, enabling us to explore the effect of popular threads[9] on the contribution scores for user expertise estimation.
- **Mean and Mean Direct.** The mean variant attempts to normalize the contribution scores gained for each user interaction according to the mean vote difference of all comments or direct comments only.
- **Median and Median Direct.** Similar to the mean adjustment, here we normalise the contribution scores by the median instead.

[9] High number of comments and user votes

4.5 Contribution Rating (C-Rating)

The C-rating approach is a pairwise comparison approach for user interactions, inspired from its success in answer quality prediction on CQA platforms [23]. Here, we model each thread as a competition between the users who comment on it. The performance of the users are measured according to the significance of their contribution. The estimated expertise are the user ratings updated according to the users' performance in each thread, compared to that of other users in the thread. A user of higher rating is expected to be more likely to contribute significant contributions than users of lower rating.

4.5.1 Comparison Pairs. Comparisons pairs are built within each Reddit thread for the pairwise comparison between each comment in the thread; given that the authors of the comments are not the same. This Round Robin comparison with other direct comments instead of all comments reduces the runtime complexity of the approach[10].

4.5.2 Glicko-2 Rating. The user ratings are updated according to the Glicko-2 rating model [10]. The outcomes of each comparison pair are collected if they lie within the same rating period[11] t and are then used to update the user rating μ_u as shown in Equation 7. Here $w^t_{u_2}$ denotes the (aggregated) outcome of all pairwise comparisons between user u_1 and user u_2 in period t.

$$\mu^{t+1}_{u_1} = \mu^t_{u_1} + (\sigma^{t+1}_{u_1})^2 \cdot \sum_{i=2}^{k} g(\sigma^t_{u_i})(w^t_{u_i} - P(\mu^t_{u_1}, \mu^t_{u_i}, \sigma^t_{u_i})) \qquad (7)$$

The probability of user u_1 defeating opponent u_2 is modelled in Glicko using the logistic function of their score differences in Equation 8 where the uncertainty in the opponent's strength $\sigma^t_{u_2}$ is taken into accounted using the formula in 9.

$$P(\mu^t_{u_1}, \mu^t_{u_i}, \sigma^t_{u_2})) = \frac{1}{1 + exp(-g(\sigma^t_{u_2})(\mu^t_{u_1} - \mu^t_{u_2}))} \qquad (8)$$

$$g(\sigma) = \frac{1}{\sqrt{1 + 3\sigma^2/\pi^2}} \qquad (9)$$

The uncertainty in the estimated user ratings are also measured in Glicko-2 as the rating deviation σ_u based on δ^2 is the game outcome variance.. The third measure, the volatility ϕ is used during the rating updates to reduce the impact of a sudden user performance shift.

$$\sigma^{t+1} = \sqrt{\left(\frac{1}{(\sigma^t)^2 + (\phi^{t+1})^2} + \frac{1}{\delta^2}\right)^{-1}} \qquad (10)$$

In this paper, we measure a user expertise according to the user's rating deducted by its uncertainty, $\mu_u - \sigma_u$.

4.5.3 Win-Margin. Traditionally, competitions including real world sports are only concerned with the outcome of a match-up and the winner takes all. Thus, pairwise comparison approaches are often scoreless without looking at the margin of victory and often disregard draws [2, 24]. The win-margin variant is inspired by the Bradley-Terry model which correlates the win probability of a player in a game with the player's real rating μ_1. This win probability is $\frac{\mu_1}{\mu_1+\mu_2}$ when the player is against an opponent with a rating of μ_2 [3]. Our earlier work on CQA platform [23] found improvements towards the estimation of user expertise if we were to consider the win-margin between the winner and the loser of a pairwise comparison. Since Reddit is a different platform (with negative vote difference) we once again explore the scoreless and the win-margin variant. To account for the negative vote difference in comments,modifications to the win-margin are required:

- If the vote difference for both comments is positive, there is no change to the win-margin approach.
- If there is a negative vote difference, then it is a total victory (win-margin of 1) for the comment with the higher vote difference.

5 EXPERIMENT

The performance of the approaches discussed in Section 4 are evaluated for the estimation of user expertise on the collected Reddit dataset detailed in Section 3.1. The user expertise estimated from the approaches are used to predict the information quality of comments (future vote difference gained) contributed by the users.

The evaluation process investigates the user expertise estimation performance on 4 subreddits – *r/explainlikeimfive*, *r/gaming*, *r/science* and *r/worldnews*. The *r/politics* is not evaluated pending more data collection for future work due to the volatility surrounding the US 2016 Presidential Elections; unlike the global *r/worldnews*.

5.1 Training and Testing Cycles

Unlike traditional Web platforms with static content, a social curation platform like Reddit expand rapidly with the constant flow of user content. Beside building high accuracy models for the estimation of user expertise, the approaches explored are fast and efficient in taking in new content for the estimation process. Thus, the performance evaluation is modelled as a continuous cycle, beginning with a training cycle according to the thread timestamp.

5.1.1 Testing Cycle. In the testing cycle, the threads and comments of the threads are used for evaluation. Here, the estimated user expertise from the explored approaches are used to predict the information quality of comments according to the evaluation measures discussed in Section 5.2. Only comments of known users (with expertise estimated from their prior interactions) are evaluated.

5.1.2 Training Cycle. Once the threads and the comments have been used for evaluation, they are added to the training data. New users will have their expertise estimated for the first time and existing users have their expertise updated with the additional new data. The updated user expertise are then used for prediction in the following testing cycle.

5.2 Evaluation Measure

The proposed algorithms are evaluated according to their capabilities in estimating user expertise that can be used to predict information quality of comments.

5.2.1 Ground Truth. The actual vote difference gained is used as the ground truth judgment of information quality for thread

[10]This differs from our earlier work in building every comparison pairs[23].
[11]Daily on Reddit due to the high traffic.

comments. User comments within each Reddit thread are ranked according to this judgment. The explored approaches attempt to rank the same comments according to the estimated expertise of comments' author; to match the ground truth rank order to predict the future community assessment of content quality. Our experiment looked at predicting the ordering of – (1) direct comments; and (2) all comments in a thread. We then measure the Kendall's Tau correlation between the ground truth and the predicted ordering.

5.2.2 Kendall's Tau Rank Coefficient, Tau-B Measure. This research selects up to the 10 best comments for each Reddit thread[12] according to the ground truth. Joint observation pairs are then formulated for each of these comments in a Reddit thread. We then measure the number of concordant and discordant pairs for the observation pairs between the ground truth and the explored approaches. As it is possible for two comments to have the same vote difference or two users to have the same estimated expertise, we handle these ties with the Kendall's Tau-B, τ_B statistic.

6 RESULTS AND DISCUSSION

As detailed in Section 5, the evaluation process is based on an online training-test cycle. From a total of 724, 505 threads, 333, 175 threads have at least 1 direct comment created by a user with prior contribution (for the expertise to be estimated) and 344, 375 threads with at least 1 comment. Thus, 45.99% and 47.53% of the threads are suitable for direct comments and all comments evaluation respectively. We note that this value is much smaller for the *r/gaming* subreddit with only 23.03% and 25.70% respectively; possibly due to the low number of comments per thread and also almost 20% of its users are thread contributors only (as noted in Section 3). On the other hand, the *r/explainlikeimfive* subreddit recorded a higher number of suitable threads for evaluation with 78.78% and 79.51% of threads for direct comments and all comments respectively; possibly due to a higher density of threads in the subreddit with at least one interaction (see Table 2). The high number of threads and the comments in each thread ensures the significance of the evaluation measure.

6.1 Differentiating Users

The estimated user expertise is a relative measure which allow us to compare the odds of quality contributions between two users. For example in Reddit threads, it would be ideal for the expertise of the authors of the comments to not be the same unless they are all contributing comments of the same quality, otherwise there would not be a possible ordering in the prediction of information quality. Due to space limitations we present in Table 7, the evaluated approaches at their best variant by not having an order according to the estimated user expertise when there is an order for the ground truth. Here, we observe that the C-Count approach is unable to differentiate user expertise well within an evaluated thread when compared to the other approaches. Upon further inspection, we note that the C-Rating (scored variant) has the highest number of ties in order when the ground truth is a tie as well with only 0.24% of the ties going undetected.

[12]Computation are expensive for threads with very high number of comments

Table 7: Evaluated Approaches (Best Variant) with the Least Number of Orderless User Expertise in Reddit Thread Comments

	Percentage of Thread	
Approach	Direct Comments	All Comments
C-Count (mean threshold, no decay)	5.88%	5.12%
C-Score (averaged score, raw value, no decay)	5.31%	4.70%
Z-Index (mean threshold)	5.58%	4.88%
C-Rating (scoreless)	5.46%	4.77%

6.2 Prediction Performance

Table 8 summarises the comment quality prediction based on the estimated user expertise of the explored approaches at their best variant; measured with the Kendall's Tau-B Rank Coefficient against the ground truth. It can be observed that the best performing approach is the C-Rating approach followed by the C-Count, C-Score and finally the Z-Index.

Generally, the explored approaches perform well on all subreddits with the exception of the *r/gaming* subreddit where the threads lack user comments[13]. The C-Rating approach however performed very well in this subreddit as a robust approach.

6.3 Vote Difference as Contribution Measure

Similar to the earlier work on CQA platforms [23], the user vote information should not be directly used as an estimate of user expertise. The C-Score approach recorded a low performance measure with the average value variant outperforming the sum variant. Instead this information should be used as an indication of contribution significance – by counting the significance of a contribution using the C-Count or be used for pairwise comparison in the C-Rating. This is crucial for an environment with sparse user activity (a lower comment count in thread, a lower word count per comment) such as the *r/gaming* subreddit where we observe the largest performance gain.

6.3.1 Raw Values as Measures. Another interesting observation here is that the best performing variant for the C-Score uses the vote difference as a raw value; a value which we had feared would be inflated and thus should be normalised within the context of the thread itself. For example, popular threads tend to have higher number of user votes that inflate the vote difference of comments. What this finding suggests is that users who commented on such threads and recorded a high number of vote difference should be rewarded (for being able to make an impact on such a challenging thread in the first place). This is further supported by the C-Score having the average score value as its best variant, implying users should be providing consistently high quality content instead of more content of a lower quality.

6.3.2 Significance Threshold. The vote difference gained by a content is useful as an indicator of content contribution for a thread discussion. Often, the polarity of vote difference is used to differentiate contribution significance especially from the users' perspective. Our findings here showed that the best performing variant for both the C-Count and the Z-Index have the mean of vote difference for

[13]Discussed in Section 3

Table 8: **Kendall's Tau-B Rank Coefficient for Comments Ordering based on Quality Prediction with User Expertise Estimation with the Evaluated Approaches (Best Variant). Best performance in bold.**

	Direct Comments for Subreddit					All Comments for Subreddit				
Approach	elifive	gaming	science	worldnews	All	elifive	gaming	science	worldnews	All
C-Count (mean threshold, no decay)	0.9164	0.7923	0.9148	0.8890	0.8797	0.8284	0.6923	0.8610	0.8512	0.8084
C-Score (average score, raw value, no decay)	0.7816	0.7014	0.8873	0.7810	0.7720	0.6423	0.5702	0.8098	0.6941	0.6566
Z-Index (mean threshold)	0.7400	0.6928	0.8710	0.7088	0.7280	0.6258	0.5709	0.7975	0.6265	0.6256
C-Rating (scoreless)	**0.9796**	**0.9298**	**0.9760**	**0.9867**	**0.9716**	**0.9472**	**0.8786**	**0.9469**	**0.9777**	**0.9434**

all comments in a thread as a threshold for content significance – a comment is considered significant for a thread if the vote difference of that comment is above the mean.

6.4 Vote Difference as a Relative User Performance Measure

The high performance of the C-Rating approach is consistent with our earlier findings on CQA platforms [23] that it is possible to apply rating systems in user expertise estimation through pairwise comparisons, even when comparing against direct comments of threads only. User votes on content can be used as a judgment of user performance including the negative vote difference. The existence of negative vote difference however does not benefit the successful win-margin (scored) variant [23] as the scoreless variant recorded a slightly better performance with the exception on the *r/science* and *r/worldnews* subreddit.

6.5 Penalising Bad Content

This research explored the possibility to ignore bad contributions by users which maybe caused by user bias and instead only reward good contributions. The findings however suggest otherwise where all of such variants in the C-Count and C-Score were outperformed by variants which takes the bad contribution into account. Besides that, the C-Rating which naturally account for bad contributions as a losing performance measure recorded the highest evaluation measure.

6.6 Decaying User Expertise

Literature suggests that the user expertise improves overtime and earlier contributions should be discounted. A decay factor is introduced to reduce the impact of earlier contributions towards the estimation of user expertise. Our findings do not agree with this notion where the variants with decay for the C-Count and C-Score were outperformed by the non-decay variants; possibly be due to:

- There is a lack of consistent user contribution or activity on Reddit as discussed in Section 3.
- There should be an impact from the earlier user interaction, rewarding the good and penalising the bad. Thus, the users would need to be making significant contributions in the present and future to atone for the earlier bad interactions.

7 CONCLUSION

Social curation platforms have a large impact on the World Wide Web (WWW) today as a content-rich platform enriched by their users. The large amount of user-generated content (UGC) of varying information quality however creates a challenge for community-based content management. A detailed study was conducted using data from Reddit, one of the largest social curation platforms in use today. Here, we analysed 5 large subreddits which provided us with an understanding on Reddit, its content management and the behaviour of its users.

This research proposed the content-agnostic estimation of user expertise by extracting the vote difference from prior generated content; which is then used to predict the information quality of future content generated by the users themselves. Four main approaches were proposed based on previous work in expertise estimation; with multiple variants each for further exploration. These approaches were evaluated according to their prediction performance in the ranking of Reddit thread comments.

The findings suggest that it is possible to estimate user expertise for the prediction of content quality, with an average Kendall's Tau-B rank correlation of 0.9434 using the proposed C-Rating approach (which is based on competitive pairwise comparisons) to rank thread comments in comparison against the ground truth of unseen actual user moderation. This performance is consistently better than the other explored baseline approaches even in subreddits with sparse amount of user comment interactions such as the *r/gaming* subreddit; showcasing the robustness of the approach for the unpredictable nature of UGC platforms.

The vote difference is a good feature to indicate prior content quality for the explored approaches. This research discovers however that the vote difference should not be used directly as an indication of user expertise as the C-Score approach performs poorly. Instead, the vote difference can be used to identify the content of significance contribution in a thread discussion using the mean vote difference of comments in a thread in the C-Count approach; or as a relative measure for the pairwise comparison of user performance in the C-Rating approach. Both approaches were able to estimate user expertise for the prediction of content quality.

Findings from the research also suggest that the despite raw vote difference values are inflated within the popular threads, the values should not be normalised as users should be rewarded for interacting and making an impact on such competitive threads. Besides that, this research discovers that the polarity of vote difference gained by content is not a suitable threshold in judging contribution significance of content. Instead, a content is deemed to be significant if it registered a vote difference above the average of vote difference of other content within the same scope.

7.1 Future Works

Next, the research aims to further develop approaches to better estimate user expertise for content management on social platforms; by accounting for the comment chain in Reddit threads and learning user behaviours. A possible extension from this work is to explore applying user expertise to weight a user's votes.

REFERENCES

[1] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2012. Discovering Value from Community Activity on Focused Question Answering Sites: A Case Study of Stack Overflow. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '12)*. ACM, 850–858.

[2] Çiğdem Aslay, Neil O'Hare, Luca Maria Aiello, and Alejandro Jaimes. 2013. Competition-based Networks for Expert Finding. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '13)*. ACM, 1033–1036.

[3] Ralph A. Bradley and Milton E. Terry. 1952. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika* 39, 3/4 (1952).

[4] Xuan Chen and John Heidemann. 2005. Flash Crowd Mitigation via Adaptive Admission Control Based on Application-level Observations. *ACM Trans. Internet Technol.* 5, 3 (Aug. 2005), 532–569.

[5] Justin Cheng, Lada Adamic, P. Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. 2014. Can Cascades Be Predicted?. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14)*. ACM, New York, NY, USA, 925–936.

[6] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. 2008. An experimental comparison of click position-bias models. In *WSDM '08: Proceedings of the international conference on Web search and web data mining*. ACM, New York, NY, USA, 87fi?!94.

[7] Jeremy Elson and Jon Howell. 2008. Handling Flash Crowds from Your Garage. In *USENIX 2008 Annual Technical Conference (ATC'08)*. USENIX Association, Berkeley, CA, USA, 171–184.

[8] Adrien Friggeri, Lada Adamic, Dean Eckles, and Justin Cheng. 2014. Rumor Cascades. (2014).

[9] Eric Gilbert. 2013. Widespread Underprovision on Reddit. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 803–808.

[10] Mark Glickman. 2001. Dynamic paired comparison models with stochastic variances. *Journal of Applied Statistics* 28, 6 (2001).

[11] Joshua Guberman, Carol Schmitz, and Libby Hemphill. 2016. Quantifying Toxicity and Verbal Violence on Twitter. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion (CSCW '16 Companion)*. ACM, New York, NY, USA, 277–280.

[12] Nathan Oken Hodas and Kristina Lerman. 2012. How Visibility and Divided Attention Constrain Social Contagion. In *Proceedings of the 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust (SOCIALCOM-PASSAT '12)*. IEEE Computer Society, Washington, DC, USA, 249–257.

[13] Nathan Oken Hodas and Kristina Lerman. 2013. The Simple Rules of Social Contagion. *CoRR* abs/1308.5015 (2013).

[14] Tad Hogg, Kristina Lerman, and Laura M. Smith. 2013. Stochastic Models Predict User Behavior in Social Media. *CoRR* abs/1308.2705 (2013).

[15] Jeon-Hyung Kang and Kristina Lerman. 2015. VIP: Incorporating Human Cognitive Biases in a Probabilistic Model of Retweeting. In *Social Computing, Behavioral-Cultural Modeling, and Prediction*, Nitin Agarwal, Kevin Xu, and Nathaniel Osgood (Eds.). Lecture Notes in Computer Science, Vol. 9021. Springer International Publishing, 101–110.

[16] Simon Kassing, Jasper Oosterman, Alessandro Bozzon, and Geert-Jan Houben. 2015. Locating Domain-specific Contents and Experts on Social Bookmarking Communities. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing (SAC '15)*. ACM, New York, NY, USA, 747–752.

[17] Himabindu Lakkaraju, Julian McAuley, and Jure Leskovec. 2013. What's in a Name? Understanding the Interplay between Titles, Content, and Communities in Social Media. (2013).

[18] Kristina Lerman and Aram Galstyan. 2008. Analysis of Social Voting Patterns on Digg. In *Proceedings of the First Workshop on Online Social Networks (WOSN '08)*. ACM, New York, NY, USA, 7–12.

[19] Kristina Lerman and Tad Hogg. 2010. Using a Model of Social Dynamics to Predict Popularity of News. In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*. ACM, New York, NY, USA, 621–630.

[20] Kristina Lerman and Tad Hogg. 2012. Using Stochastic Models to Describe and Predict Social Dynamics of Web Users. *ACM Trans. Intell. Syst. Technol.* 3, 4, Article 62 (Sept. 2012), 33 pages.

[21] Kristina Lerman and Tad Hogg. 2014. Leveraging Position Bias to Improve Peer Recommendation. *PLoS ONE* 9, 6 (11 June 2014), e98914+.

[22] Baichuan Li and Irwin King. 2010. Routing Questions to Appropriate Answerers in Community Question Answering Services. In *Proceedings of the ACM Conference on Information & Knowledge Management (CIKM)*.

[23] Wern Han Lim, Mark James Carman, and Sze-Meng Jojo Wong. 2016. Estimating Domain-Specific User Expertise for Answer Retrieval in Community Question-Answering Platforms. In *Proceedings of the 21st Australasian Document Computing Symposium (ADCS '16)*. ACM, New York, NY, USA, 33–40.

[24] Jing Liu, Young-In Song, and Chin-Yew Lin. 2011. Competition-based User Expertise Score Estimation. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '11)*. ACM, 425–434.

[25] Richard Mills. 2011. Researching Social News -fi?! Is reddit.com a mouthpiece for the fiHive Mindfi, or a Collective Intelligence approach to Information Overload?. In *ETHICOMP 2011 Proceedings*. Sheffield Hallam University.

[26] Blair Nonnecke and Jenny Preece. 2000. Lurker Demographics: Counting the Silent. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*. ACM, New York, NY, USA, 73–80.

[27] Henrique Pinto, Jussara M. Almeida, and Marcos A. Gonçalves. 2013. Using Early View Patterns to Predict the Popularity of Youtube Videos. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining (WSDM '13)*. ACM, New York, NY, USA, 365–374.

[28] Maria Priestley and Alex Mesoudi. 2015. Do Online Voting Patterns Reflect Evolved Features of Human Cognition? An Exploratory Empirical Investigation. *PLoS ONE* 10, 6 (06 2015), e0129703.

[29] Philipp Singer, Fabian Flöck, Clemens Meinhart, Elias Zeitfogel, and Markus Strohmaier. 2014. Evolution of Reddit: From the Front Page of the Internet to a Self-referential Community?. In *Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion (WWW Companion '14)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 517–522.

[30] Greg Stoddard. 2015. Popularity and Quality in Social News Aggregators: A Study of Reddit and Hacker News. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 815–818.

[31] Gabor Szabo and Bernardo A. Huberman. 2010. Predicting the Popularity of Online Content. *Commun. ACM* 53, 8 (Aug. 2010), 80–88.

[32] Tim Weninger, Xihao Avi Zhu, and Jiawei Han. 2013. An Exploration of Discussion Threads in Social News Sites: A Case Study of the Reddit Community. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '13)*. ACM, New York, NY, USA, 579–583.

[33] Fang Wu and Bernardo A. Huberman. 2007. Novelty and collective attention. *Proceedings of the National Academy of Sciences* 104, 45 (2007), 17599–17601. arXiv:http://www.pnas.org/content/104/45/17599.full.pdf

[34] Jun Zhang, Mark S. Ackerman, and Lada Adamic. 2007. Expertise Networks in Online Communities: Structure and Algorithms. In *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*. ACM, 221–230.

[35] Zhi-Min Zhou, Man Lan, Zheng-Yu Niu, and Yue Lu. 2012. Exploiting User Profile Information for Answer Ranking in cQA. In *Proceedings of the Conference Companion on World Wide Web (WWW)*.