Knowledge fixation and accretion: Longitudinal analysis of a social question-answering site

Abstract

Purpose – The aim of this work was to investigate longitudinal features of an established social question-answering site to study how question-answer resources and other community features change over time.

Design/methodology/approach – Statistical analysis and visualisation was performed on the full data dump from the StackOverflow social question-answering site for programmers.

Findings – The timing of answers is as strong a predictor of acceptance - a proxy for user satisfaction - as the structural features of provided answers sometimes associated with quality. While many questions and answer exchanges are short-lived, there is a small yet interesting subset of questions where new answers receive community approval and which may end up being ranked more highly than early answers.

Research limitations/implications – As a large-scale data oriented research study, this work says little about user motivations to find and contribute new knowledge to old questions or about the impact of the resource on the consumer. This will require complementary studies using qualitative and evaluative methods.

Practical implications – While content contribution to social question-asking is largely undertaken within a very short time frame, content consumption is usually over far longer periods. Methods and incentives by which content can be updated and maintained need to be considered. This work should be of interest to knowledge exchange community designers and managers.

Originality/value – Few studies have looked at temporal patterns in social question-answering and how time and the moderation and voting systems employed may shape resource quality

Keywords: Online communities; knowledge exchange; social QA; question answering; social epistemology

Article Type: Research paper

Introduction

Social question-answering (Q & A) sites provide a platform for users to post questions on general or more specialist topics and receive answers from the community of users. Curated question-answer thread pages then become a resource of value more widely to web users looking for the same, or similar solutions to problems. Aspects of gamification and collaborative filtering - introducing points and rankings for asking and answering queries, and allowing users to vote on others' contributions - are common in these sites, and act as an incentive for members to provide answers. At the same time, they shape the way the information resource evolves and appears to the web user on subsequent visits. While the time to "populate" a question-answer resource may be quite short (fixation) a question will remain open for subsequent interaction, to be further enhanced, rated or edited by the community (accretion).

Information resources and architectures may be understood from a "pace layering" perspective (Morville , 2005). A concept originally applied to physical buildings, pace layering describes the differential rate of change between layers - such as structure/services/space plan in a building context - with a more permanent outer shell giving way to more rapidly changing inner layers. Over time, inner changes from new or fast-moving activities may be absorbed into the more permanent core. In the information context, this may be the process by which changes in content can be recognised and supported by the interface, navigation and social organisation of the system (Campbell and Fast, 2006). If an information resource is to remain pertinent and timely, then it needs to have this ability to absorb and adapt to new information.

This article will examine knowledge fixation and accretion in the technology - oriented social Q & A community "Stack Overflow", a resource widely used by computer programmers to provide programming language- or technology-specific answers to their questions. Before looking at the mechanics, successes and controversies within the Stack Overflow community and analysing longitudinal effects and patterns, I will discuss philosophical ideals for knowledge exchange and psychological aspects of knowledge seeking and contribution. Longitudinal aspects of other collaborative information environments will be discussed. I will then focus on social Q & A and their emergent dynamics.

Desirable norms and architecture for a knowledge commons

Online, social Q&A lowers the barriers of entry to a community of knowledge to a simple interest and willingness to contribute. Social Q & A provides an online analogue of a social epistemological system, with community members providing "testimony" as to the best answer to a question. The foremost theorists in social epistemology regard their field as normative, in providing a framework for an epistemological system to converge on the true knowledge (Goldman , 1999). Conditions of such systems include *inclusivity*, *diversity* and *authority*. A system should be inclusive, in allowing anyone to contribute testimony. It should be diverse, in encouraging a range of opinion on which the information seeker can draw. It should be authoritative, in that users should be able to gauge the authority of the testifier and their warrant for knowledge (Matthews and Simon, 2012).

Epistemic systems also need to accommodate different types of question, not only the simplified philosophical proposition of "whether p", but how-tos, whichs, whys and more general advice and recommendation seeking. Many of these wider question types do not have anything like a "correct" answer but are open to a diversity of views among which the receiver of testimony needs to choose. In doing so, we may formulate a tentative hypothesis until evidence seems sufficient to reject it (Goldman , 1999).

To these above conditions I would add that the system should be *dynamic*, flexible enough to accommodate new conditions that require knowledge to be updated, or new opinions or evidence to be included for consideration.

Social Q & A and user motivations

Sites such as Stack Overflow attract two overlapping yet partially distinct user groups - those seeking answers and those providing them (Mamykina, Manoim et al., 2011). Some mutual attraction is apparent, with the availability of answerers make the site attractive to questioners (Kumar, Liftshits et al., 2010). To understand the relative motivations for the two activities, we can look to well-established social psychology, information behaviour and communication theories.

Askers may often have a "high need for cognitive closure" (Kruglanski , 1990). That is, they are highly motivated to receive knowledge which may be specific (e.g. confirming a hypothesis) or non-specific (a "best answer" to an open question) in a short time. When in receipt of candidate answers, questioners may be prone to "satisfice" or accept good-enough answers that enable them to progress in their current task (Prabha, Lynn et al., 2007).

For answerers, many motivators for knowledge contribution in online communities have been identified, among them self-efficacy, sense of community and the principles of social exchange, reciprocity and individual recognition (Matthews and Simon, 2012, Kosonen , 2009). In Stack Overflow, answer contributors similarly show both extrinsic and intrinsic motivation, to share knowledge and help the community while gaining externally recognisable kudos through the reputation gained. (Mamykina, Manoim et al., 2011)

Some interesting and important research has been reported by Dearman and Truong into reasons that potential answerers *do not* contribute (2010). This questionnaire-based research revealed that the third most common reason - after ignoring dubious questions and questions for which they did not know an answer - was that a question had already received answers. Within this group, some mentioned that a satisfactory answer was given which they felt they could not improve on. Importantly though, some answerers reported not answering for fear that their contribution would be lost or ignored in amongst the existing posts. Here we see that users may feel discouraged from adding knowledge based on an assessment of its (low) potential impact.

Quality in social Q & A

Quality assessment of a Q&A resource might consider question and answer quality individually or look instead at the quality of the emergent question and answer resource.

Stack Overflow has a range of guidelines and suggestions to users on how to formulate a question as well as detailed information on the type of questions that should be asked. The site FAQ encourages answerable, practical questions specific to the domain – open-ended, chatty questions are explicitly discouraged[i]. Question guidelines request that answers first be properly researched, that users are specific as possible, that the question is on-topic and relevant to others[ii]

Quality of question and answers may be modelled using recoverable features and training against user-provided quality assessments, either those implicit in the system or those collected separately for the purposes of the research evaluation. Agichtein et al. (2008) , for instance, show that for Yahoo Answers, a combination of link analysis features, textual features, and user interaction features, can provide a model with up to 0.761 accuracy (area under the ROC curve) for predicting question quality and 0.878 for predicting answer quality against a cross validation set. The most important features for making the prediction for questions were usage statistics (such as click-throughs), user assigned scores (such as stars given to a question asker), and textual features (such as the amount of punctuation in the question text). For answers, answer length was significant, as were the frequency of words appearing in the overall corpus and the relative up vote proportion.

In terms of answers, one further route may be to assess answers based on which are "accepted" as a best answer by the questioner. Adamic et al (2008) used a logistic regression on data from the social question answering site Yahoo Answers, to predict whether an answer would be accepted at the "best" answer. They selected data from a set containing 1,178,983 questions and 8,452,337 answers, balancing for accepted or not accepted answers. They claimed a 62% prediction accuracy - achieved using cross-validation - for accepted answers, based on reply length, the length of the thread (negative), the number of best answers given by the answerer and the number of replies (negative). There are, however, as discussed below, some temporal bias issues with treating accepted answers (or even those with the most upvotes) as proxies for the highest quality and we should be cautious here.

At a smaller scale and taking a mixed methods approach, Blooma et al (2008) analysed 300 question and answer pairs from Yahoo Answers, relating textual and non-textual features to answer quality. A multiple regression model was used, taking an overall quality rating of the answers by human volunteers as the dependant variable. Model variables were the reputation and authority of asker and answerer and the category (non-textual variables) and accuracy, completeness, language, reasonableness and answer length (textual features, again as ratings from the volunteers). The textual features of accuracy, completeness and length were the most significant regressors (Blooma, Chua et al., 2008).

   While such structural features presented above may provide a reasonable predictor of quality, there remains quite a large amount of variability unexplained by quantitative measures that may be better accessed by qualitative approaches (Shah and Kitzie, 2012) In a study of acceptance criteria in Yahoo Answers, for instance, Kim & Oh showed that socio-emotional criteria may be cited in addition to those that are content-related (Kim and Oh, 2009).

It is notable also that the above studies have largely neglected temporal features of the interactions, or only including them implicitly (e.g. by counting the number of answers received or the overall click-through count). One exception is a recent study by Shah (Shah , 2011) on Yahoo Answers. He found a that within 15 minutes a majority of questions had been answered to the user's apparent satisfaction, and that first answers were the most likely to be chosen as best among the first five received. Notably, however, a third of questions where not satisfactorily answered until over an hour after the question and the answers ranked sixth in time were the most likely of all to be accepted. This appears to support the observation that the most significant quality features that has been identified: answer length, correctness of grammar etc, are probably better associated with questions and answers that more time to prepare and present.

Fewer studies have targeted the quality of the overall question-answer thread as a resource, which may consist of the question plus a number of different - though equally relevant - answers. In a way, well

phrased questions with several high quality candidate answers may be the best overall quality resources and equivalent to Wikipedia's featured articles. Some support for this comes from Fichman (Fichman , 2011) for thread completeness and verifiability but interestingly not for overall accuracy (due to the presence of contradictory information in the threads studied)

In a learning environment, multiple wiki texts - even when each text is itself collaboratively authored - may be more helpful to students than a single text, with individual preferences differing as to which may be most helpful (Fountain , 2007). Such intertextuality is often associated with the building of a more sophisticated understanding of the problem area and, perhaps counter intuitively, prior naivety towards a topic may actually help in building a nuanced understanding across multiple texts (Bråten, Strømsø et al., 2008).

In a game-theoretical analysis of social Q&A, Jain et al (2009) usefully distinguish between new answers that *substitute* for existing ones (providing an overall better answer alternative which renders existing answers redundant) and those that *complement* them (adding to and enriching the information already provided). Through a simulation exercise, they show that the existing voting system for Yahoo Answers is satisfactory in the case of substitution answers but less so for complementary answers. In the latter case, the ability for the question poser to award votes proportionally to a number of answers was found to be optimal. This analysis points to the possibility that current collaborative ranking systems may introduce bias to the detriment of overall resource quality.

Sources of bias through moderation, community size and the use of gamified interfaces

A number of forms of bias - both epistemically positive and negative - may be found within socio-technical systems. Firstly, there is bias introduced by the technology platform itself, through the way in which information is selected and presented to the user. Secondly, there is bias relating to the user community (norm-oriented bias) and how it may self-moderate or police. A third type of bias emerges as a result of user interaction with the technology, and may become manifest as the filtering effected on an information resource, or the order of presentation such that the user is more likely to read a particular section of the information.

In social Q & A and the Stack Overflow interface, the "fastest gun" issue has been identified and criticised both from within and outwith the community. Here, users compete to answer questions as quickly as possible, in order to gain the associated reputation points if their answer is accepted by the questioner (Mamykina, Manoim et al., 2011). As one community user puts it, after admitting to watching the very newest questions being posted to the site in order to select which to answer : "You can usually catch questions first if you watch this page. I think it is the easiest way to score rep."[iii]. While this has the desirable effect of the asker receiving an answer quickly, but may sacrifice quality given that the answerer is in a rush to provide information. The extent of the problem was highlighted, with the median answer time calculated at 11 minutes after the question was posted.

In epistemic online communities that reach a certain size, the sheer volume of information becomes very difficult to manage. The addition of new posts at a certain rate means that new questions may slip quickly from view and not be detected by potential responders. A number of corrective measures may be taken to refloat unanswered questions and to rank them by votes or activity, and Stack Overflow makes use of both

of these approaches. These measures are only partially successful, however, and the activity window on new questions remains by and large fairly brief.

Collaborative filtering, in the form of aggregated user votes, provides an important mechanism to rank answers provided to users' questions. Again, "fastest gun" effects my lead to bias here as early answers start to gain votes before - possibly higher quality - answers accumulate. As we shall see below, this can be verified in Stack Overflow by looking at Answer scores in relation to the time they were contributed.

Knowledge accretion in collaborative authoring environments

Perhaps the most widely studied collaborative knowledge technology is the wiki, and the specific instance most widely studied is Wikipedia. Here, approaches have looked at the dynamics of article authoring and editing for different types of topic, and have reached a quite good understanding of how knowledge builds and changes over time.

Overall, although a large number of users may work on any particular article, additions and minor style edits are far more common than deletions or more thorough reorganisations (Jones , 2008).

Quality Wikipedia articles are those which over time achieve a balanced objectivity and a relative stabilisation of information (Sanger , 2009). In cases of contested or controversial topics, the overall direction of the article may be determined by the more persistent - as opposed to expert - editors.

The "Featured article" status of the very best pages are often taken as a proxy for quality, and these may have more policy edits than non - featured articles (Jones , 2008).

The more prolific Wikipedia editors tend to have a better grasp of Wikipedia norms (e.g. a neutral point of view, and will cite these in edit page discussions about addition or change to content (Panciera, Halfaker et al., 2009). Notably, however, "generalist" regular contributors may not provide as high a quality of contributions as expert, irregular contributors (Luyt, Aaron et al., 2008)

A "first mover advantage" in Wikipedia articles has been identified. Content contributed early in a page's life tends to persist, and that may include any errors introduced early on (Luyt, Aaron et al., 2008, Ehmann, Large et al., 2008). This was first identified and visualised in the "History Flow" project undertaken by IBM research (Viégas, Wattenberg et al., 2004). It is not clear what the underlying explanation for the first mover advantage is, though it is possible that subsequent edits look at style and formatting issues and neglect the factual content of the articles.

While the knowledge produced by a collaborative Q & A resource may be viewed as relatively ephemeral in comparison to a Wikipedia, user studies have highlighted that even conversational and advice-seeking related questions have a high perceived archival value (Gazan , 2011).

Methods and Approach

This work used the Creative Commons data dump of the Stack Exchange web sites, a periodically published dataset containing partial information for a range of social Q&A sites, with Stack Overflow the largest. Most of the work was conducted using the May 2011 data, but where indicated the December 2011 data dump was used. The May 2011 dataset contains some 6,479,788 posts - incorporating both questions and answers - from amongst 756,695 users. By December 2011, the totals had grown to 7,397,507 posts and 887,372 users.

Preprocessed data and R code used in the analysis is available in the following repository:
https://github.com/paulusm/InfoAccretion

Findings

Temporal bias in answer acceptance

An analysis of temporal factors in relation to answer acceptance confirmed the tendency for the first few answers received to be accepted in preference to later ones. Figure 1 shows the histogram of 1000 randomly selected answers plotted by temporal rank. We can see that the first answers to be received are accepted at a rate of 0.65, but by the fourth answer the acceptance rate has dropped to 0.2.

Figure 1: Accepted and not-accepted answers by temporal rank
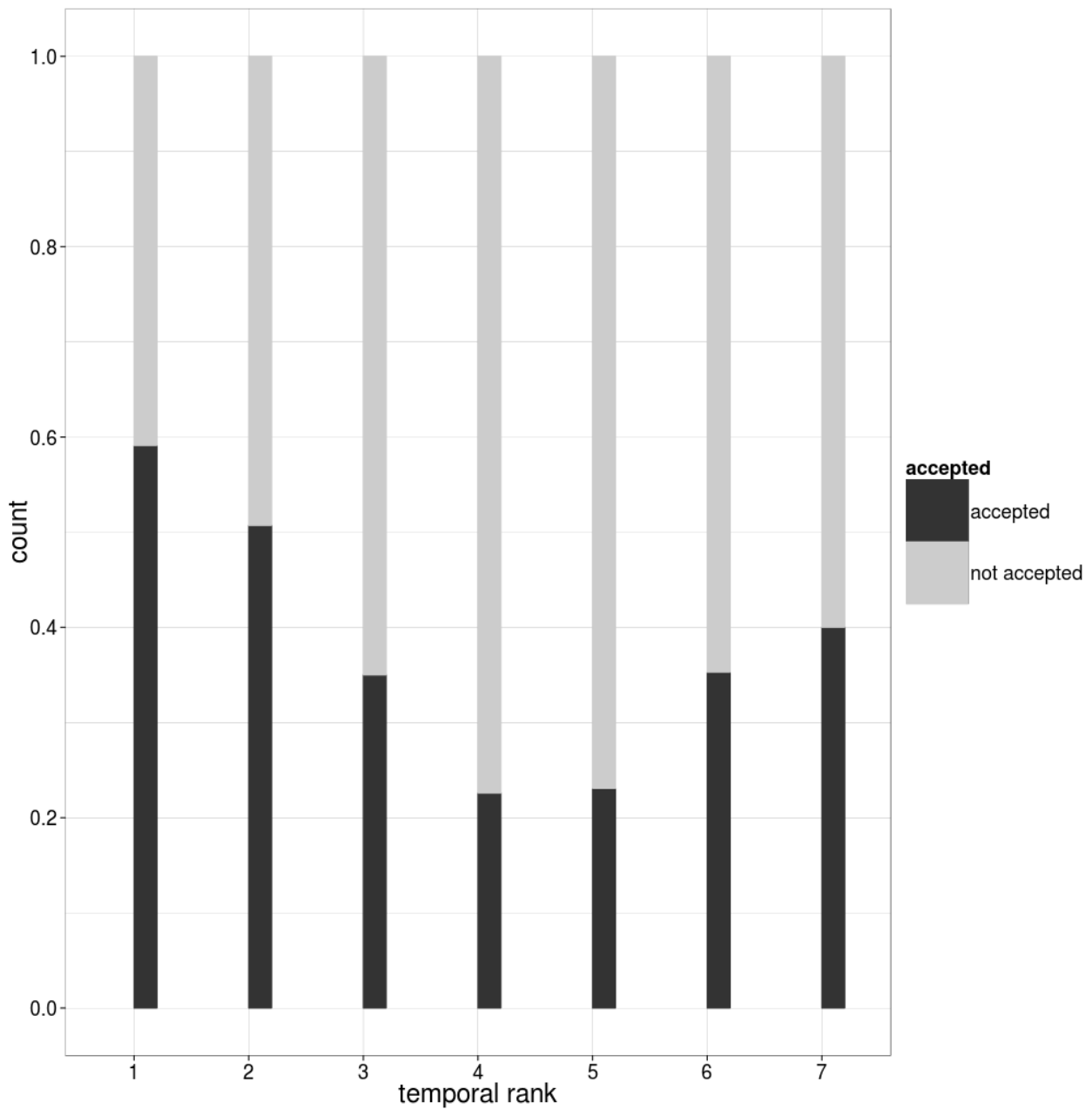
In order to compare temporal factors with other structural features previously identified as important to acceptance or answer quality, multiple logistic models were constructed to provide predictions of answer acceptance. 5000 randomly selected answers were used to construct the models, with the features of temporal rank, the time from question to answer, the post length in characters and whether the post contained a code example. The non-temporal features had been identified in earlier studies as important features in acceptance decisions in Q&A generally or, for code examples, the Stack Overflow community (Adamic, Zhang et al., 2008, Treude, Barzilay et al., 2011).

Models were tested using a further separate sample of 5000 answers. The prediction accuracy in terms of a 0.5 prediction threshold and the area under the ROC curve are shown in Table 1, below.

Table 1: Predicting acceptance from temporal and textual features

	Model	Prediction Accuracy(%)	Area Under ROC Curve
1	temporal rank	0.603	0.644
2	quadratic temporal rank	0.603	0.644
3	log of temporal rank	0.616	0.644
4	time after question posted	0.518	0.500
5	answer length (chars)	0.561	0.586
6	Contains code	0.561	0.586
7	temporal rank + time after question	0.605	0.632
8	temporal rank + answer length	0.635	0.686
9	temporal rank + answer length + contains code	0.640	0.695
10	log of temporal rank + answer length + contains code	0.636	0.693

We see that temporal rank provides the best single predictor of acceptance, with the timing of the response and the structural features relatively less reliable. Combining features only provides piecemeal improvements, with the best accuracy of 0.644 achieved with a combination of temporal rank, answer length and the "contains code" flag.

Question longevity and "breaking" answers

As noted in previous studies, the activity window around new questions tends to be very brief, with most answers contributed on the same day, usually within hours of the question being added. As with many online activities, however, there is a long tail of activity and some questions are revisited months later.

Table 2 shows a summary of questions and the time at which answers were contributed. The majority of question activity ceased after 30 days. In some cases, activity continued after 30 days but those questions contributed fewer user votes. Only in around 2% of cases did new questions garner the majority of votes.

*Table 2: Question longevity in terms of vote distribution*

	Description	count(*)
1	Answers above 30 days old scoring more highly	24286
2	Answers less than 30 days old scoring more highly	160620

	Description	count(*)
3	No answers after 30 days	1620674

Figure 2 shows a typical question timeline, with the majority of answers, and up-votes to these - signalling approval from other community members - received in the few days after the question was asked. The upper part of the graph presents the answers, each in different colour, with their cumulative up-votes over time. The lower part shows voting activity for the answers, with the size of the marker proportional to the number of upvotes received at that time.

The typical voting patterns are shown in figure 2, and tend to show the most activity on the day of question posting, with voting tailing off thereafter, but the majority of votes going to answers that were in place early in the timeline[iv].

Figure 2: Typical question timeline

When to rewrite a code base from scratch
<web-development><architecture><tdd><testing> , 1386 views
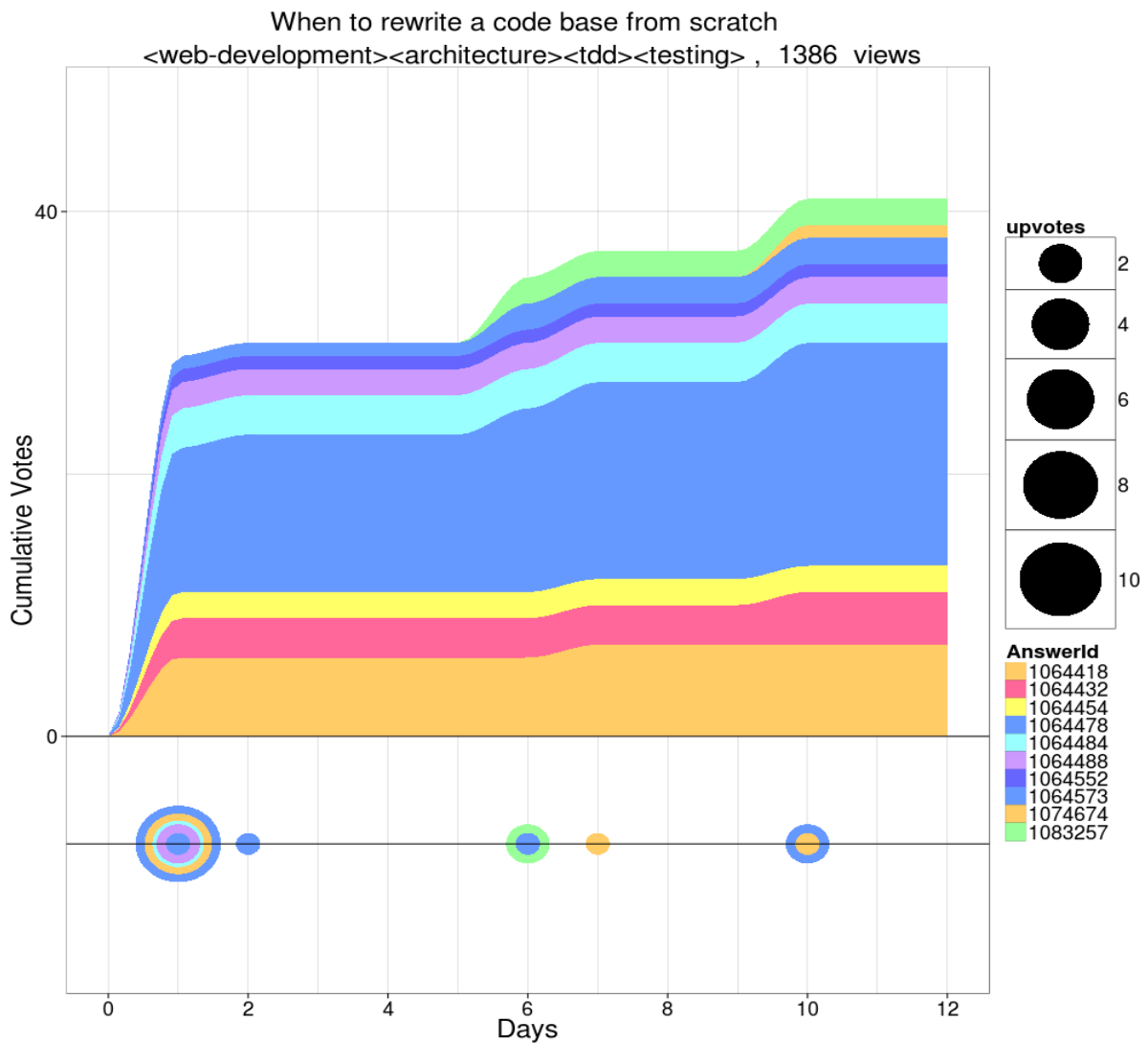
In a "poll" type question, voting is more evenly spread out, with new answers appearing at intervals[v]. This type of question, while attracting community participation and attention, is also a type disfavoured by moderators and liable to closure for being too subjective (see evolution of moderation, below).
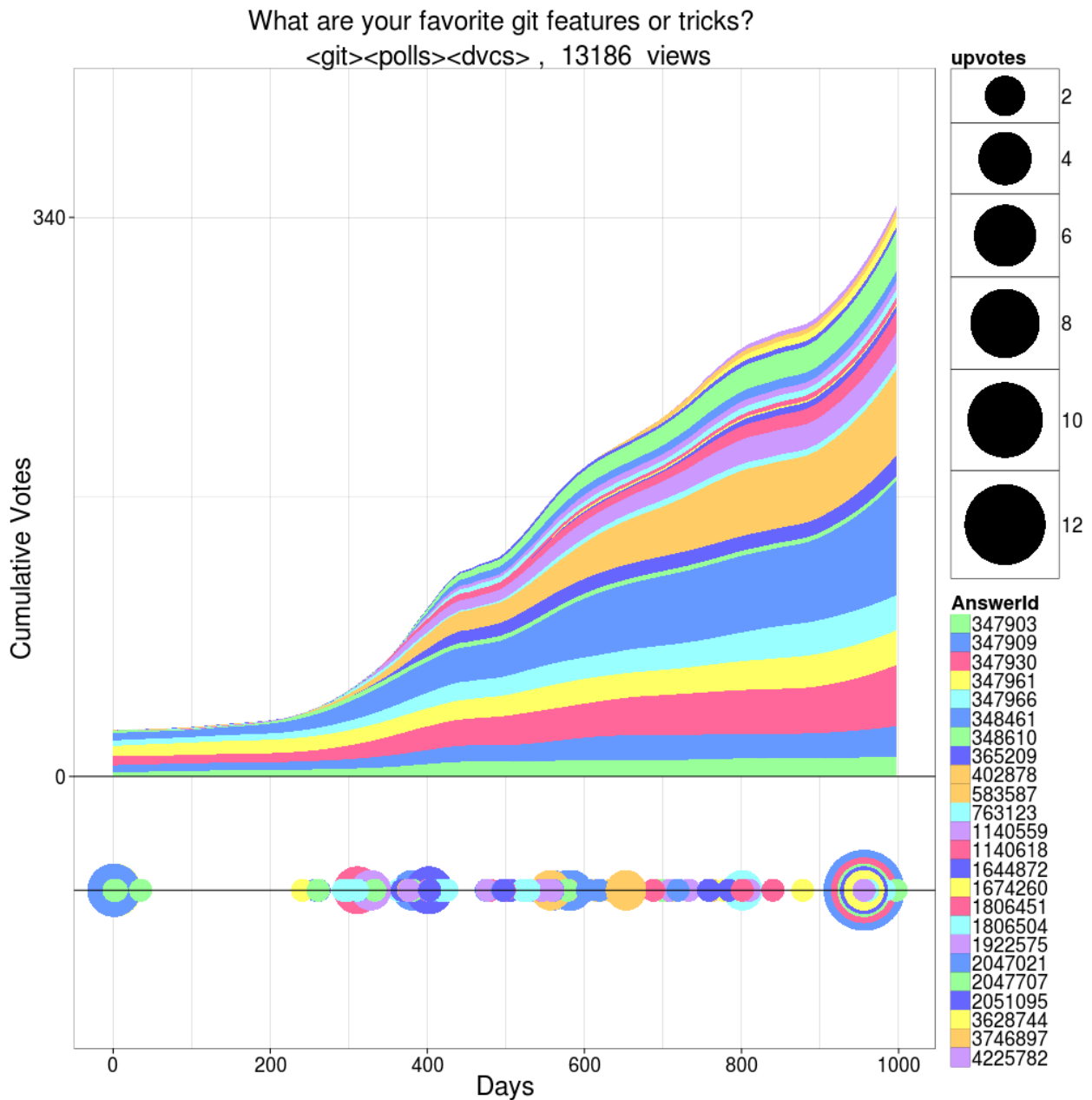

*Figure 3: Poll type question timeline (answers receiving at least 5 upvotes)*

In the minority of cases where a new, high quality candidate answer is received that manages to attract community attention, we see a timeline as portrayed in figure 4 below. Here, the question is quite specifically related to the "Project Euler" mathematical computing challenge and as such it might be argued that there would be a very focused, expert sub-community interested in the question[vi].

Figure 4: New preferred answer timeline (answers receiving at least 6 votes)
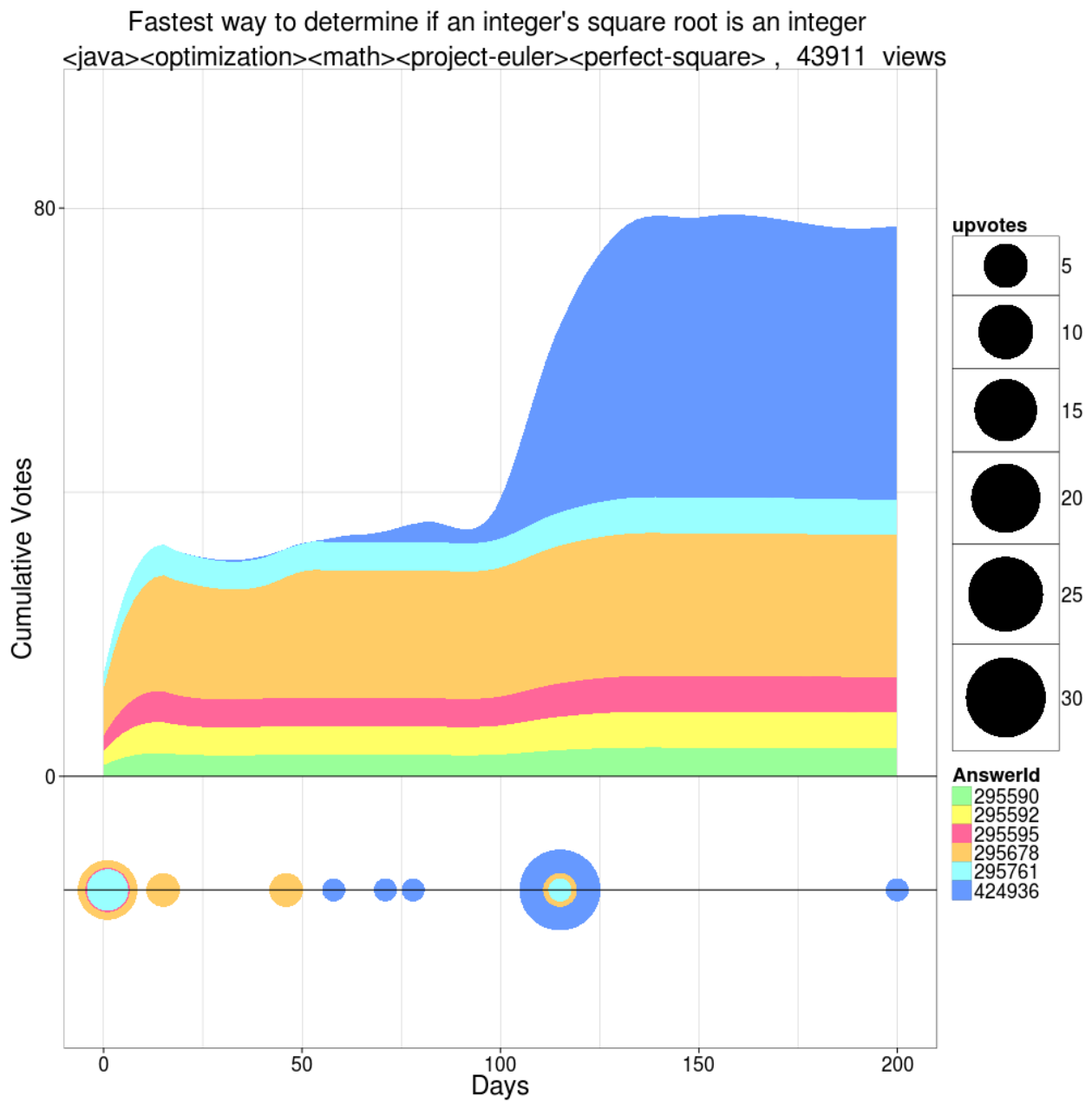
We see from subsequent voting on this question that the new answer received gradual upvotes over time, promoting it up the ranking of answers according to vote tally.
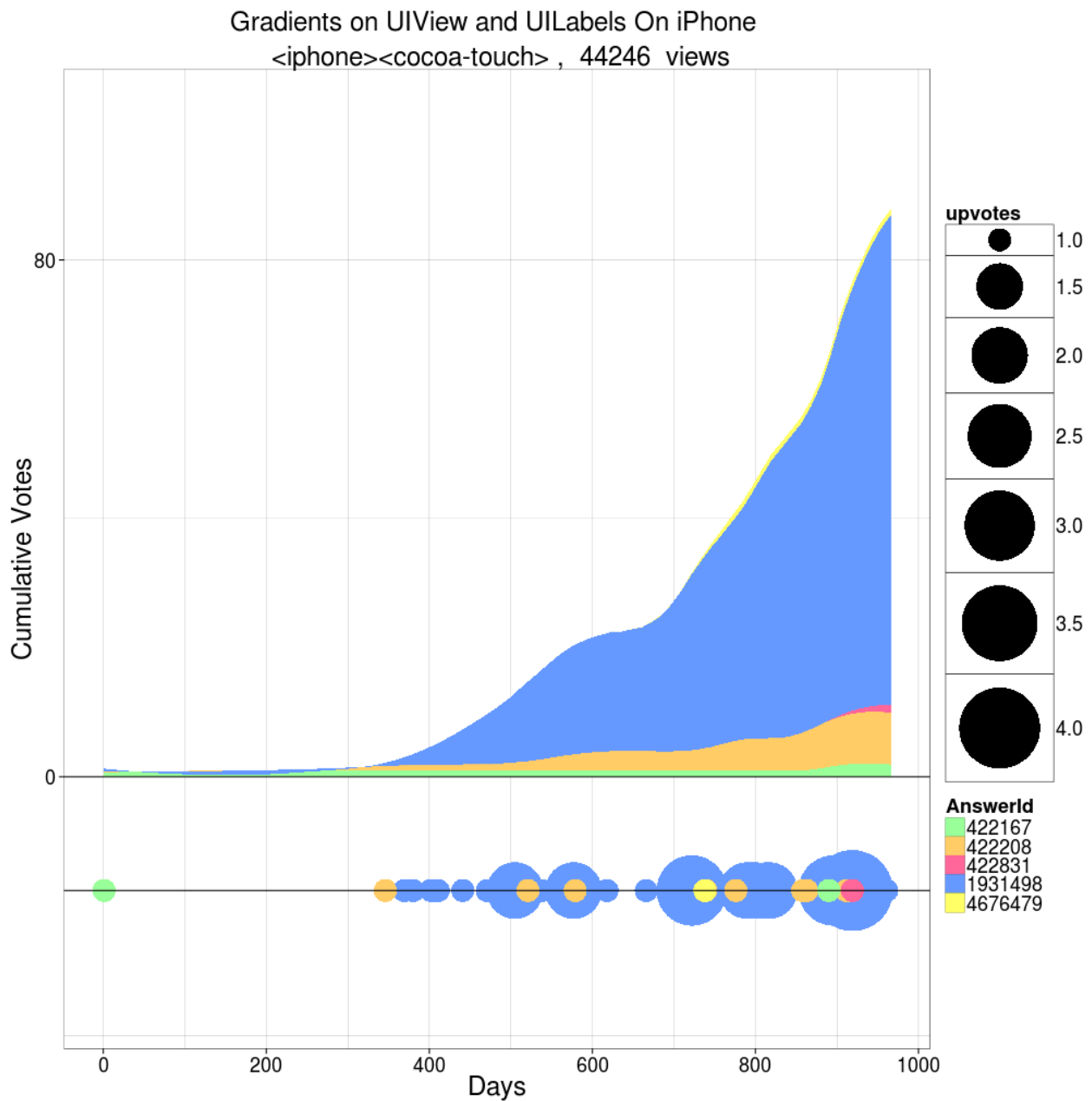
Figure 5: "Breaking answer" timeline
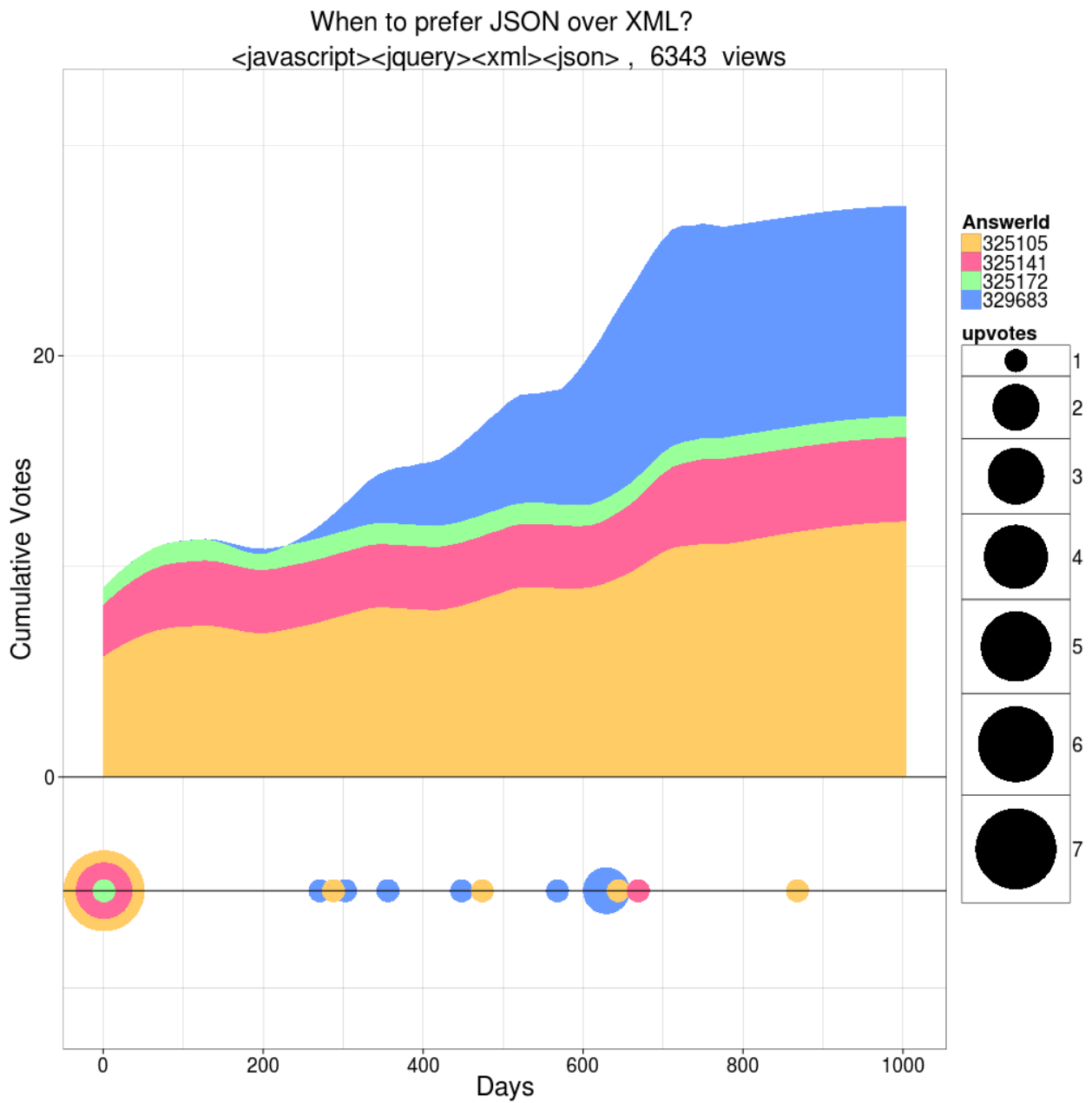
In other cases we see questions where answers are mostly temporally close to the question, but the first mover advantage puts the earliest ahead initially. In time, however, a slightly later question (in this case just a few days) gains over subsequent months. Here, the newer question author has provided a more factual, broader summary of the pros and cons of web data formats than the originally upvoted answers[vii] (Figure 5)

In some cases a new technical approach emerges in time which supercedes previously provided answers. We also see here a rise in question popularity, probably coupled with increased general activity in the area - in that case of iOS development (iPhone apps)[viii].

Figure 6: New technology question timeline

Gradients on UIView and UILabels On iPhone
<iphone><cocoa-touch> , 44246 views

upvotes

•	1.0
●	1.5
●	2.0
●	2.5
●	3.0
●	3.5
●	4.0

AnswerId
- 422167
- 422208
- 422831
- 1931498
- 4676479

In this case, unlike the accepted answer, the preferred new answer provided almost a year after the question supplies a fuller answer, including a code example. The new answer is fourth if the questions are ranked temporally, but two of those ahead - including the accepted answer - are very brief and one only attracted a single upvote. This suggests that the new answer was visible with only minimum vertical scrolling on the part of the user and rose above two of the older answers quite quickly. The utility of the new answer is clear by the quantity of upvotes it attracted - but relied on the answer provider actively seeking out an older question and choosing to update the content with new information.

Question and answer edits over time

Above certain reputation thresholds, users are able to edit the question and answers of other users. In order to investigate the extent of editing, a "diff" was calculated using the Python difflib library[ix]. The library takes a "Gestalt" approach to text comparison, calculating the percentage match between two texts by working outward from the largest sequence matched. A diff ratio can then be calculated:

> "If T is the total number of elements in both sequences, and M is the number of matches, this is 2.0*M / T. Note that this is 1.0 if the sequences are identical, and 0.0 if they have nothing in common." - difflib documentation

For the full corpus of questions in the September 2011 dataset, the mean diff overall was 0.92. Questions had received an average of 1.54 body edits with 35% of all questions receiving at least one edit. For answers, the mean diff was 0.94 and the average number 1.38 body edits - 20% of all answers received at least one edit. So more questions than answers receive an edit, and of those that are edited, questions are slightly more likely to be edited repeatedly.

To examine any relationship between the diffs of questions and answers to associated aspects of the posts, Spearman's correlation matrices were created and are shown in Tables 3 and 4, below. The "minsbetween" variable was the time in minutes between the first and last edits. Score is the sum of upvotes minus downvotes, and ViewCount the number of web views the post has received.

*Table 3: Correlation matrix of question edit features*

	score	viewcount	diff	minsbetween	bodyedits	titleedits
score	1.00	0.37	0.01	0.12	0.04	0.04
viewcount	0.37	1.00	-0.05	0.18	0.06	0.04
diff	0.01	-0.05	1.00	-0.24	-0.36	-0.09
minsbetween	0.12	0.18	-0.24	1.00	0.27	0.07
bodyedits	0.04	0.06	-0.36	0.27	1.00	0.12
titleedits	0.04	0.04	-0.09	0.07	0.12	1.00

Table 4:Correlation matrix of answer edit features

	score	diff	minsbetween	bodyedits
score	1.0000	-0.0218	0.0532	0.1467
diff	-0.0218	1.0000	-0.0230	-0.3399
minsbetween	0.0532	-0.0230	1.0000	0.2782
bodyedits	0.1467	-0.3399	0.2782	1.0000

While we see no significant relationship between the extent of edits and quality ratings (score and viewcount), there are indications that the time between edits is related to score and viewcount for questions, and that the number of body edits may be important for answers. In the case of questions, this may be because edits over time keep the thread active and therefore more visible through the interface. In the case of answers, it is indicative of possible quality improvements introduced through iterative edits (even if more major changes are not associated with more views).

Figure 7: Question(solid line) and answer (dashed line) diffs for posts with 2 or more edits
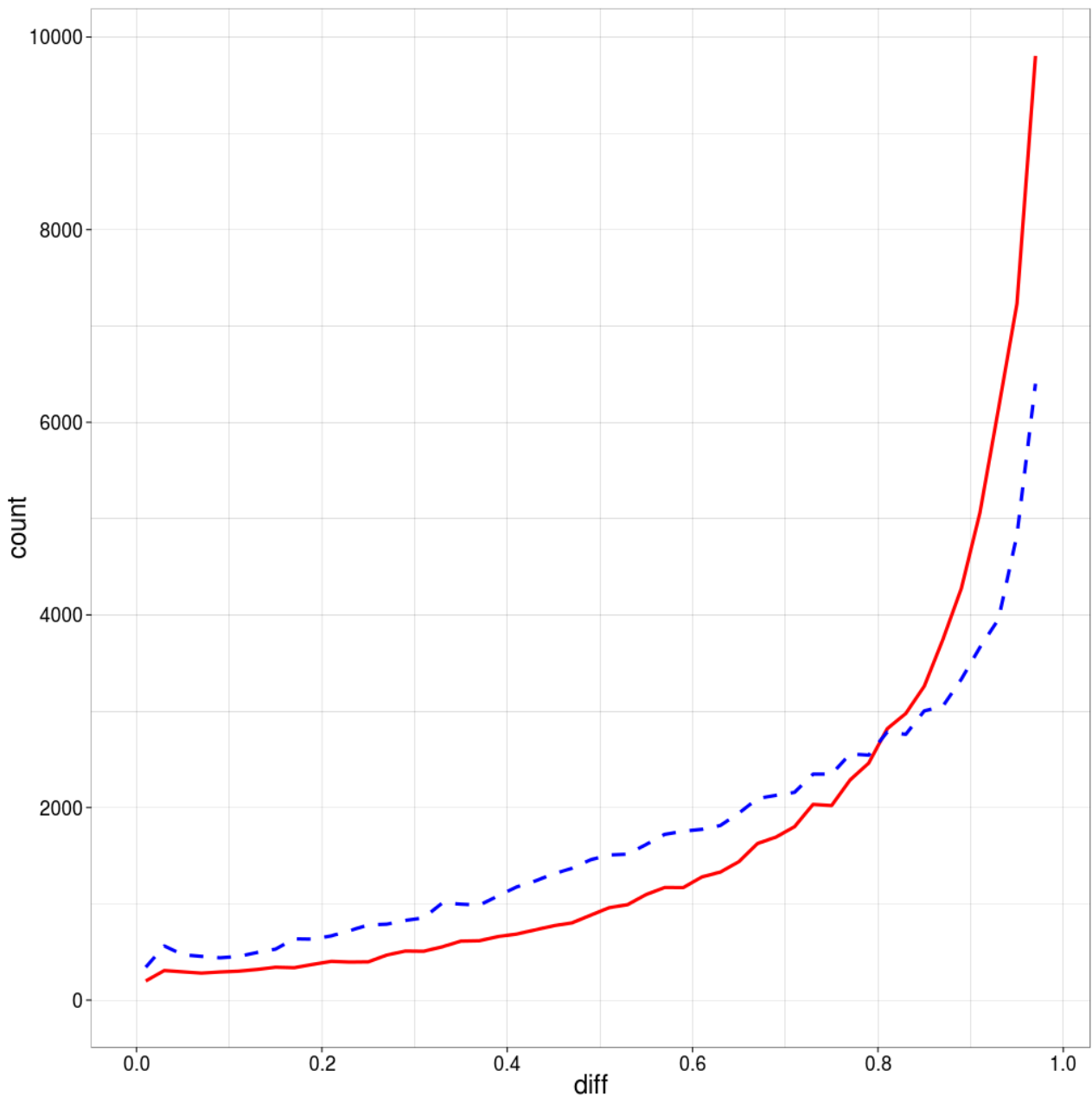
To further investigate this difference between questions and answers, the diff distribution was plotted for a random sample of 10,000 of each and is shown in Figure 7. This reveals that questions tend to be edited more shallowly - and accords with user testimony that much of question editing is to do with more minor formatting[x]

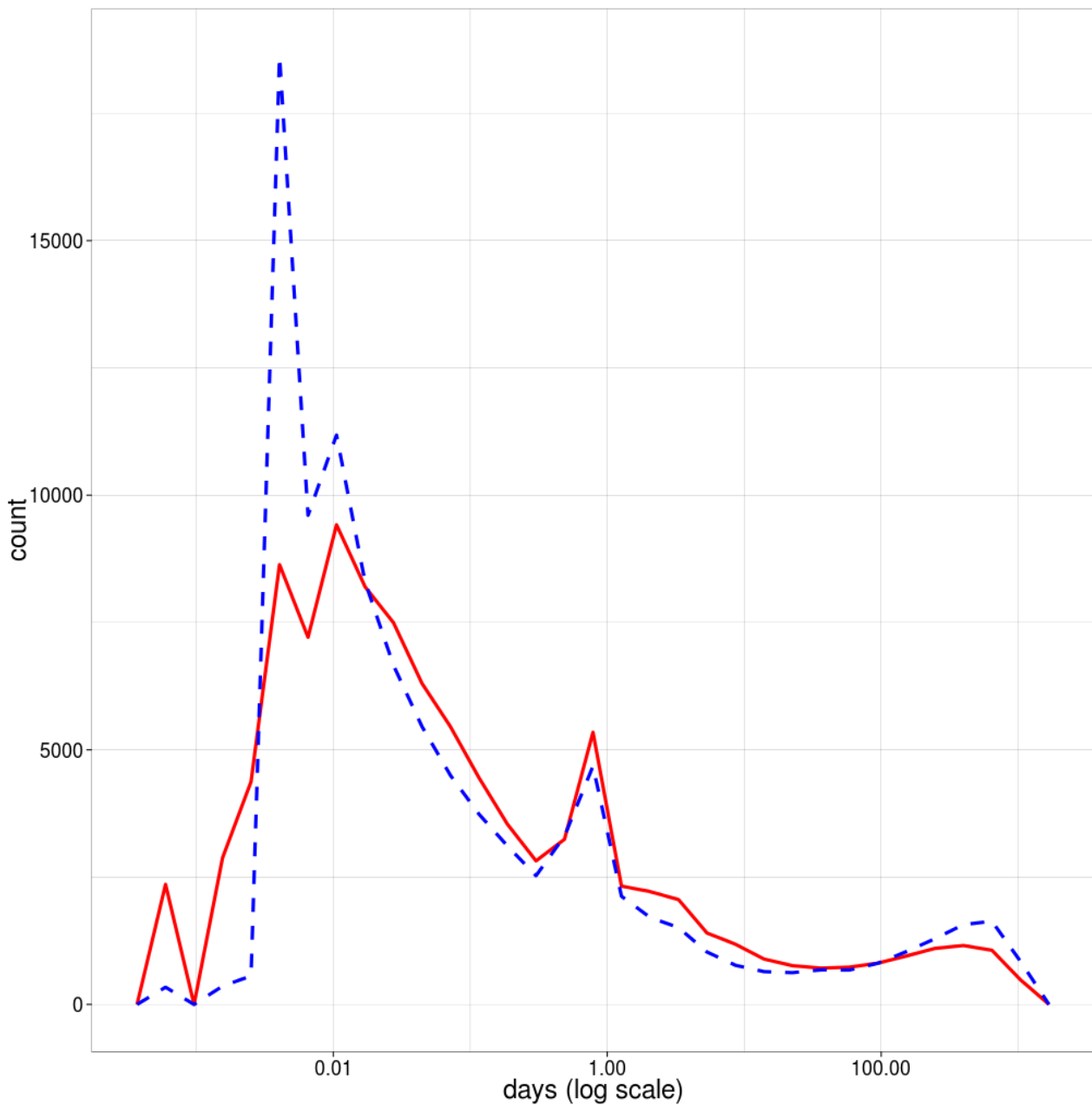*Figure 8: Question(solid line) and answer (dashed line) editing period for posts with 2 or more edits*

Figure 8 illustrates the temporal distribution of edits. In the majority , edits are made quickly but there is long tail of edits over time. The peaks in the editing curves may partly due to behavioural aspects, such as users revisiting their posts at the same time the following day.

Evolution of moderation

In the Stack Overflow community, question duplicates are an acknowledged problem. The objective is a single, canonical answer to a specific question and the emergence of duplicates threatens this. With the success and expansion of the community this has been a growing problem. Users with a certain reputation level are able to flag questions as duplicates and vote for them to be closed, at which point the question is merged into the original.

We see from a comparison of duplicate closure votes over time, that this is in fact increasing with the accumulation of questions, indicating that policies to encourage the identification of duplicates are meeting some success (Figure 9).

Figure 9: Questions closed as duplicates as a percentage of total questions (with loess smoothed mean line and 0.95 confidence limit in shading)
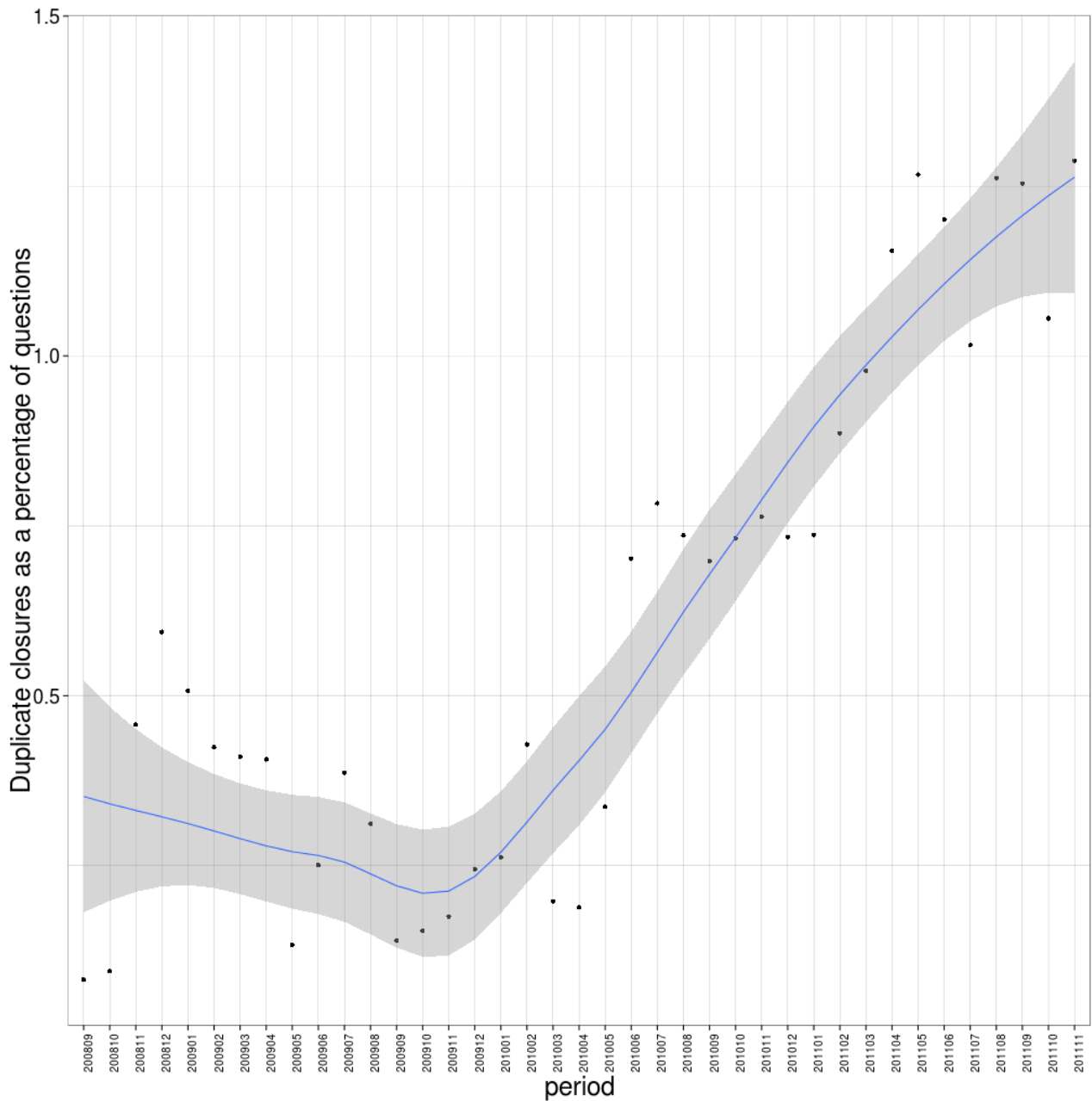
A related issue is the closure of questions deemed to be off-topic and not suited to the remit of the community. Here again, there is a trend towards increasing numbers of questions being closed as off-topic (Figure 10). This is something that is itself a bone of contention within the StackOverflow community, as "off-topic" may be to some extent a matter of opinion.

Figure 10: Questions closed as off topic as a percentage of total questions (with loess smoothed mean line and 0.95 confidence limit in shading)
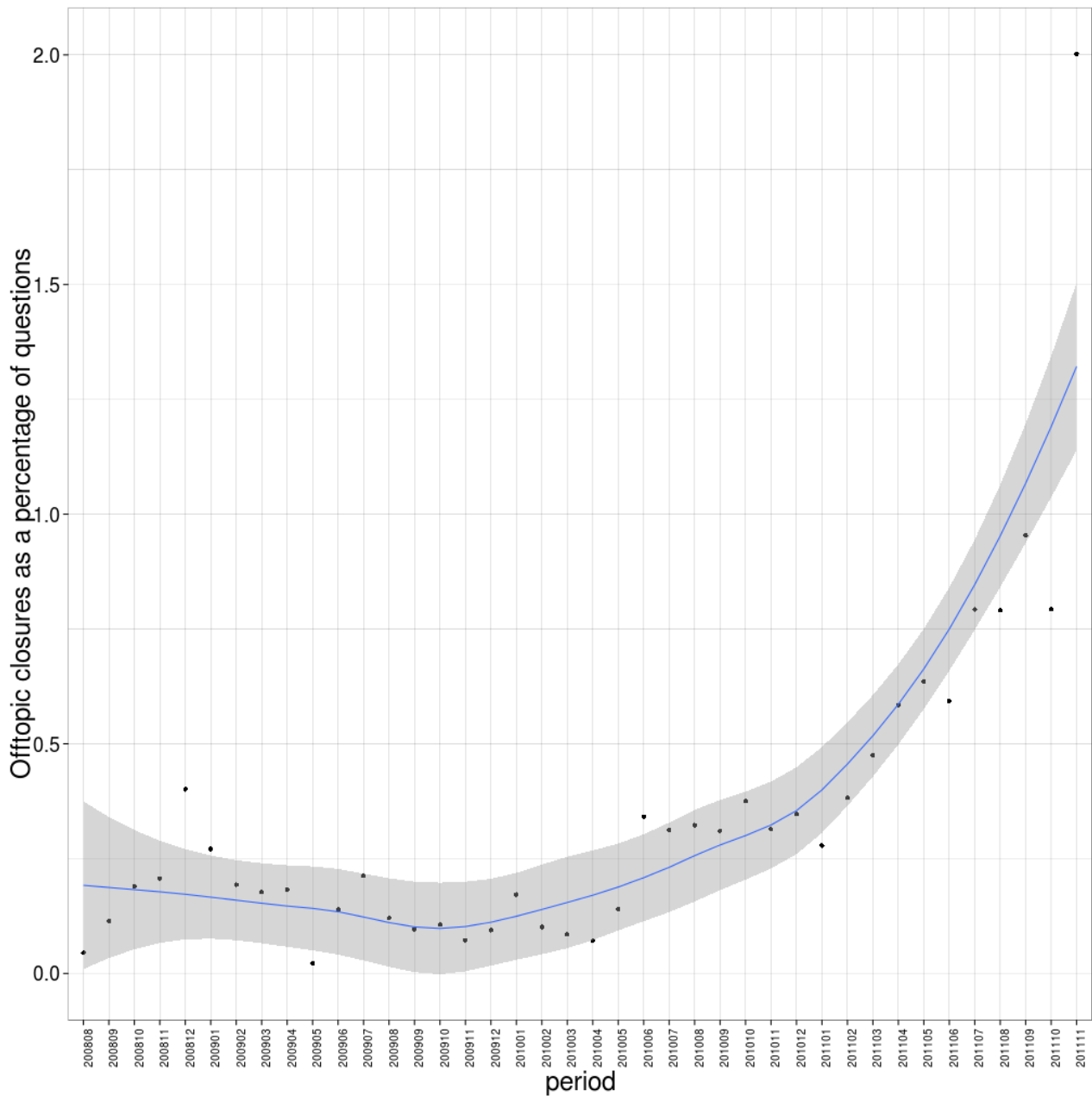
Conclusion and Discussion

The analysis largely confirms the temporally biased nature of this type of social Q&A community. Askers require quick answers, and will choose speed over quality if they allow them to achieve some form of cognitive/epistemic closure. The community at large will also mostly limit voting on a question-answer thread while it is active, most commonly in the few days following creation.

We do see, however, that a minority of questions receive high quality answers much later, improving the quality of the resource for the community and web users more generally. Understanding how this happens - and encouraging the drivers thereof - could make for more dynamic and flexible resources.

Old questions, new knowledge

It seems that the altruistic motivations of social QA members are influenced by an awareness of building a generally useful public resource by updating old questions over time. The following exchange, from the "meta" discussion pages of diy.stackexchange.com, a sister site to Stack Overflow, illustrates this well:

> "The fact that an old question can reappear and get new answers is by design. Things change, what was the best answer a year or two ago might not be any more. - ChrisF
>
> I agree: it is a good thing when an old question is rekindled and receives new answers (typically from someone who was not here when the question first was posted). Remember that we're building a resource for posterity, not just helping the OP."– Vebjorn Ljosa
>
> It is a very negative experience to spend time working on an answer, only to realize that the asker will never see it. – dbracey
>
> @dbracey - don't focus on the asker. Focus on all the other people who'll find the question because they have the same problem. – ChrisF"[xi]

There has indeed been discussion on how to reward new answers to old questions[xii],[xiii]. Some have felt that existing mechanics are sufficient, in that questions are "bumped" back to active and that the original asker will receive a notification that a new answer is available. Others have suggested that the system needs to more actively promote new answers, or "decay" votes on older answers that were contributed earlier. It is certainly possible to conceive of algorithmic adjustments that detect potentially "fast rising" answers and provide additional weighting for them, in addition to mechanisms for rewarding users reopening and updating older threads. Indeed, Stack Overflow has implemented the "Archaeologist" badge specifically to encourage this[xiv].

Tag Watching

The true power of a technical specialist community such as Stack Overflow is that it is really a community of communities, with a range of specialist users focussing on particular languages, toolkits or frameworks. Here, the ability to watch a tag and hence view new questions specifically for that topic adds to the length of the visibility time window and opportunities for questions to receive inputs. The online community "warren and plaza" metaphor (Zhang , 2010) here favours the warren - a personalised information space only visible to a subset of users - for enriching specialist knowledge. This in turn accords with one of Goldman's (1999) criteria for a social epistemology, that personal interest is what adds value to the knowledge-seeker's belief states.

Related to tag watching is answer provider notification, and recent work has pointed to ways to recommend questions to individuals who may be in a position to provide answers (Zheng, Hu et al., 2012).

Redressing Voting Bias

Analyses of information aggregation have long recognised the need for certain conditions to apply in order for combined knowledge to outperform individual opinion. When voting is anonymous and the pool of

information providers have a better than even chance of being right then conditions are ideal. When, however, social deliberation is involved, then reputational and information cascades - where individuals withhold information or defer to more opinionated people - can adversely affect the outcome (Sunstein , 2006). These bias cascading effects are clearly in operation in online collaboratively filtered knowledge exchanges, with recent studies highlighting the "rich get richer" phenomenon. Answers with more votes in turn tend to attract more votes. (Paul, Hong et al., 2012). In addition to the possibility of distributing allocated upvotes amongst complementary answers (Jain, Chen et al., 2009) it has been suggested that voting on randomly presented samples of content might also alleviate this problem (Stray , 2009), though such an approach has yet to be field - tested.

Shallow Editing

The diffs analysis presented here suggests that the extension of editing of questions and answers over longer time periods also appears related to the overall value of a question thread to web users. It is only in the long tail, however, that we see posts with edits that represent significant revisions of the original text, and many may only be cosmetic. This highlights the difference between the social Q&A resource and a community edited wiki such as Wikipedia. In the former, knowledge fixation is driven by the demand-supply dynamic, whereas in the latter there appears to be more of a commitment on the part of editors to incrementally improve resource quality.

Policy evolution

Finally, in the large community under study, we see self-moderation behaviour geared toward improving the overall resource by removing repetition and content considered out of scope. Whilst clearly both of these aims are laudable and are succeeding to some extent, it is not without the risk of alienating users and leading to the criticism of over-policing. Indeed, application of such policies appears to be as much an art as a science, with for instance some debate over what constitutes duplicate information.

Limitations and Further Work

A purely data-driven analysis of social Q&A dynamics may be informative, yielding numerous insights into community dynamics and question and answer characteristics. To really understand user motivations and the impact of the emergent resource on the knowledge seeker , however, we need to meld these with qualitative and ethnographic studies that provide deeper insight into psychological and epistemological dynamics.

We know, for instance, that profile information (especially user reputation) and answer scores may be used as heuristics by the information seeker in selecting which posts to prioritise. What is not clear, however, is the extent to which the user understands how reputation and scoring really works and how influential they may be. In terms of social epistemological value, these mechanisms provide potential benefits but may also threaten inclusivity, genuine authority and the flexibility of the resulting resource.

In terms of the temporal aspects discussed here, it seems very important to continue to improve our understanding of the conditions under which new, better quality answers can rise to the top – both from the point of view of user motivation to supply these answers and from the interplay of system dynamics and user attention that allow their promotion.

References

Adamic, L., Zhang, J., et al. (2008) Knowledge Sharing and Yahoo Answers: Everyone Knows Something. [Online]. *In:* Anon. *17th International World Wide Web Conference (WWW2008), Beijing, China 2008.* ACM, Available from: http://www2008.org/papers/fp840.html [Accessed June 2009].

Agichtein, E., Castillo, C., et al. (2008) Finding High-Quality Content in Social Media. [Online]. *In:* Anon. *Proceedings of the international conference on Web search and web data mining, Palo Alto, California, USA* New York, NY, USA: ACM, pp 183-194. Available from: http://doi.acm.org/10.1145/1341531.1341557 .

Blooma, M.J., Chua, A.Y.K., et al. (2008) A Predictive Framework for Retrieving the Best Answer. [Online]. *In:* Anon. *Proceedings of the 2008 ACM symposium on Applied computing, Fortaleza, Ceara, Brazil* New York, NY, USA: ACM, pp 1107-1111. Available from: http://doi.acm.org/10.1145/1363686.1363944 .

Bråten, I., Strømsø, H.I., et al. (2008) Are sophisticated students always better? The role of topic-specific personal epistemology in the understanding of multiple expository texts. *Contemporary educational psycholog*y 33(4), 814-840.

Campbell, D.G., and Fast, K.V. (2006) From Pace Layering to Resilience Theory: The Complex Implications of Tagging for Information Architecture. *Paper presented at Information Architecture Summit, 200*6.

Dearman, D. and Truong, K.N. (2010) Why Users of Yahoo!: Answers do Not Answer Questions. [Online]. *In:* Anon. *Proceedings of the 28th international conference on Human factors in computing systems, Atlanta, Georgia, USA* New York, NY, USA: ACM, pp 329-332. Available from: http://doi.acm.org/10.1145/1753326.1753376 .

Ehmann, K., Large, A., et al. (2008) Collaboration in context: Comparing article evolution among subject disciplines in Wikipedia. *First Monda*y [online]. 13(10), 6.

Fichman, P. (2011) A comparative assessment of answer quality on four question answering sites. *Journal of Information Scienc*e [online]. 37(5), 476-486.

Fountain, R. (2007) Conceptual Enhancement Via Textual Plurality:: A Pedagogical Wiki Bow Towards Collaborative Structuration. [Online]. *In:* Anon. *Proceedings of the 2007 international symposium on Wikis,* ACM, pp 43-46.

Gazan, R. (2011) Social Q&A. *Journal of the American Society for Information Science and Technolog*y [online]. 62(12), 2301-2312.

Goldman, A.I. (1999) *Knowledge in a Social World.* Oxford.

Jain, S., Chen, Y., et al. (2009) Designing Incentives for Online Question and Answer Forums. [Online]. *In:* Anon. *Proceedings of the 10th ACM conference on Electronic commerce, Stanford, California, USA* New York, NY, USA: ACM, pp 129-138. Available from: http://doi.acm.org/10.1145/1566374.1566393 .

Jones, J. (2008) Patterns of Revision in Online Writing. *Written Communication*} [online]. 25(2), 262-289.

Kim, S., and Oh, S. (2009) Users' relevance criteria for evaluating answers in a social Q&A site. *Journal of the American Society for Information Science and Technolog*y 60(4), 716-727.

Kosonen, M. (2009) Knowledge sharing in virtual communities a review of the empirical research. *International Journal of Web Based Communitie*s 5144-163.

Kruglanski, A.W. (1990) Lay Epistemic Theory in Social-Cognitive Psychology. *Psychological Inquir*y 1(3), 181-197.

Kumar, R., Liftshits, Y., et al. (2010) Evolution of Two-Sided Markets. [Online]. *In:* Anon. *Third ACM International Conference on Web Search and Data Mining, New York February 3-6 2010.* Available from: http://www.wsdm-conference.org/2010/proceedings/docs/p311.pdf .

Luyt, B., Aaron, T.C.H., et al. (2008) Improving Wikipedia's accuracy: Is edit age a solution? *Journal of the American Society for Information Science and Technolog*y [online]. 59(2), 318-330.

Mamykina, L., Manoim, B., et al. (2011) Design Lessons from the Fastest q\&\#38;a Site in the West. [Online]. *In:* Anon. *Proceedings of the 2011 annual conference on Human factors in computing systems, Vancouver, BC, Canada* New York, NY, USA: ACM, pp 2857-2866. Available from: http://doi.acm.org/10.1145/1978942.1979366 .

Matthews, P. and Simon, J. (2012) Evaluating and Enriching Online Knowledge Exchange: A Socio-epistemological Perspective. *In:* A.A. Lazakidou, R. Sharda and S. Voay, eds. *Virtual Communities, Social Networks and Collaboration.* [Online]. Springer New York.

Morville, P. (2005) *Ambient Findability.* O'Reilly.

Panciera, K., Halfaker, A., et al. (2009) Wikipedians are Born, Not made: A Study of Power Editors on Wikipedia. [Online]. *In:* Anon. *Proceedings of the ACM 2009 international conference on Supporting group work, Sanibel Island, Florida, USA* New York, NY, USA: ACM, pp 51-60. Available from: http://doi.acm.org/10.1145/1531674.1531682 .

Paul, S.A., Hong, L., et al. (2012) Who is Authoritative? Understanding Reputation Mechanisms in Quora. [online]. Available from: http://arxiv.org/abs/1204.3724v1 http://arxiv.org/abs/1204.3724v1.

Prabha, C., Lynn, S.C., et al. (2007) What is enough? Satisficing information needs. *Journal of Documentatio*n [online]. 63(1), 74-89.

Sanger, L.M. (2009) The Fate of Expertise after WIKIPEDIA. *Episteme* 6(1), 52-73.

Shah, C. (2011) Measuring effectiveness and user satisfaction in Yahoo! Answers. *First Monda*y6(2), .

Shah, C., and Kitzie, V. (2012) Social Q&A and virtual reference : comparing apples and oranges with the help of experts and users. *Journal of the American Society for Information Science and Technolog*y [online]. n/a-n/a.

Stray, J. (2009) *Rating Items by Number of Votes: Ur Doin It Rong* [online] Available from: http://jonathanstray.com/rating-items-by-number-of-votes[Accessed 10/22/2012 2012].

Sunstein, C.R. (2006) *Infotopia: How many minds produce knowledge.* New York: Oxford.

Treude, C., Barzilay, O., et al. (2011) How do Programmers Ask and Answer Questions on the Web?: NIER Track. [Online]. *In:* Anon. *Software Engineering (ICSE), 2011 33rd International Conference on,* pp 804-807.

Viégas, F.B., Wattenberg, M., et al (2004) *Studying Cooperation and Conflict between Authors with history flow Visualizations* [online] Available from: http://alumni.media.mit.edu/~fviegas/papers/history_flow.pdf[Accessed 4/16/2012 2012].

Zhang, X. (2010) *Social Software Sundays #2 – The Evaporative Cooling Effect* [online] Available from: http://blog.bumblebeelabs.com/social-software-sundays-2-the-evaporative-cooling-effect/[Accessed 10/12/2012 2012].

Zheng, X., Hu, Z., et al. (2012) Algorithm for recommending answer providers in community-based question answering. *Journal of Information Scienc*e [online]. 38(1), 3-14.

i   http://stackoverflow.com/faq

ii   http://stackoverflow.com/questions/how-to-ask

iii   http://meta.stackoverflow.com/questions/8137/do-you-watch-the-unanswered-page-home-page-or-another-page

iv   http://stackoverflow.com/questions/1064403/when-to-rewrite-a-code-base-from-scratch

v   http://stackoverflow.com/questions/347901/what-are-your-favorite-git-features-or-tricks

vi   http://stackoverflow.com/questions/295579/fastest-way-to-determine-if-an-integers-square-root-is-an-integer

vii   http://stackoverflow.com/questions/325085/when-to-prefer-json-over-xml

viii   http://stackoverflow.com/questions/422066/gradients-on-uiview-and-uilabels-on-iphone

ix   http://docs.python.org/library/difflib.html

x   http://meta.stackoverflow.com/users/165455/aaron-bertrand

xi   http://meta.diy.stackexchange.com/questions/668/how-can-we-keep-old-threads-from-coming-back

xii   http://meta.stackoverflow.com/questions/23996/does-it-make-sense-to-answer-old-questions

xiii   http://meta.stackoverflow.com/questions/15805/how-can-we-make-good-answers-to-old-questions-float-to-the-top

xiv   http://stackoverflow.com/badges/1286/archaeologist