Social Roles, Interactions and Community Sustainability in Social Q&A Sites: A Resource-based Perspective

Yuyang Liang
Michigan State University
liangyuy@msu.edu

Josh Introne
Michigan State University
jintrone@msu.edu

Abstract

Online tech support communities have become valuable channels for users to seek and provide solutions to specific problems. From the resource exchange perspective, the sustainability of a social system is contingent upon the size of its members as well as their communication activities. To further extend the resource-based model, the current research identifies a variety of social roles in a large tech support Q&A forum and examines longitudinal changes in the community's structure based on the identification. Moreover, this study also investigates the relationship between the community's functionality and its traffic. Results suggest that the proportion of unsolved questions negatively impacts the number of future incoming questions and the outcome of a given question is not only dependent on users' interactions within the discussion, but also on the community activities preceding the question. These observations can help community managers to improve system design and task allocation.

1. Introduction

Social Question and Answer (Q&A) sites provide information seekers spaces and opportunities to ask questions and look for solutions. For question askers, answers on Q&A platforms clearly fulfill a need. At the same time though, answerers must also derive benefits through the act of providing answers. In this sense, questions provided by askers are a *resource* that allows answerers to fulfill a need; similarly, their answers become resources for the askers. It is thought that one way a socio-technical system can become *sustainable* is through such a balanced exchange of resources [4, 29]. A sustainable social platform is one that roughly maintains (or even increases) its rate of user contributions over time without requiring infusions of external resources (e.g., paid contributions).

Recent literature has focused on different, discrete aspects of Q&A sites. Chief among these is the content and the quality of information provided [24], which includes question topics [20], question quality [16, 30] and answer quality [9, 12, 25]. Another focal area is classifying and modeling users' behaviors and expertise [8, 22, 31], which sheds light on the structures and the dynamics of various types of users in Q&A communities.

On the other hand, the amount of research regarding the underlying knowledge sharing process and the longitudinal evolution of the social system that supports it is relatively small [24]. Some studies have applied social network analysis to understand the global communication patterns in Q&A communities and their growth [1, 23, 26], and a few studies have examined the knowledge sharing process at the thread level [15, 27].

From a resource exchange perspective though, information quality, social structures, and site activity are interwoven and jointly determine whether sufficient resources are generated to meet the aggregate needs of a population of users. Some early work applied this perspective to examine Q&A interactions in Usenet discussion forums. In seminal work, Welser et al. [29] visualized the structural signature of various social roles, and argued that the balanced interactions among these roles (primarily askers and answerers) sustained continuing levels of site activity.

Although Welser et al.'s [29] analysis was compelling in part because of its parsimony, a more granular, predictive model would be of great value for designers and platform administrators. Modern Q&A platforms also offer a variety of new affordances and signifiers that may influence the resource exchange process, and ultimately, the sustainability of a modern Q&A site.

In this paper, we adopt a resource-based perspective to develop such a model. Our analysis focuses on a large online Q&A forum hosted on Reddit.com, `/r/excel`[1]. We chose Reddit as the site of our analysis for two reasons. First, the Reddit Q&A forum is a stable and successful community, and appears (on the surface) to

[1] www.reddit.com/r/excel

URI: https://hdl.handle.net/10125/59717
ISBN: 978-0-9981331-2-6

be a sustainable system, and we are interested in those factors contributing to sustainability. Second, unlike more carefully designed sites (e.g., StackOverflow), the socio-technical organization (including social structures as well as design features) of the Excel Q&A forum has evolved organically over time from a general-purpose discussion forum. As a basis for future work, we are interested in how a site comes to organize itself without the guidance of a designer's hand.

The remainder of this paper starts with a detailed discussion of the resource-based model of social structures, social roles in online contexts, and a brief overview of prior work on Q&A sites. In our analysis, we first seek to identify social roles in the Excel Q&A forum, and examine their interplay over time. We find that the proportion of unsolved questions is predictive of overall site traffic. This motivates our final analysis, which focuses on identifying those factors that weigh heavily in whether or not questions are answered successfully. These analyses allow us to articulate an overall model of sustainability for the forum, which we present in the discussion section of the paper.

2. Related Work

2.1. A Resource-based Model of Sustainability

Social structures are sustainable when the provided benefits outweigh the cost of participation [14]. In Butler's [4] resource-based model of sustainable social structures, site activity is sustained via a feedback loop of benefit provision. Current members of a social platform are key providers of resources. Their communication activities create a range of benefits for a heterogeneous population, enabling the community to develop social structures that attract new and retain existing members. The receipt and provision of benefits increase engagement and commitment among members, enabling the site to sustain (or increase) levels of activity.

At the core of this model are the communication activities of users, which transform available resources into valued benefits [4]. In online Q&A communities, the central communication activity is question posting and answering. Both askers and answerers are resource providers, in the sense that askers produce questions so that answerers are able to generate replies and display expertise while answerers provide solutions to satisfy askers' information needs. The exchange of resources among individuals creates dynamics (temporal variance in forum activities) for the system as a whole, and so the dynamics of a forum are connected to aspects of the resource exchange process in interesting ways. For instance, Anderson et al. [2] found that questions that elicit high volumes of communication reflect the community's general interest in the question, and generate higher reputation scores for answerers. By also treating members as resources, Dev et al. [7] examined the interdependence between questions and answers and showed that an increase of the inputs leads to a constant increase of the outputs in the content creation process.

This investigation extends the previous research on social Q&A communities by using the resource-based framework to understand how such dynamics are connected within the context of the overall socio-technical system. Prior work focuses mainly on answers, or individual question-answer pairs, but does not consider how these interactions contribute to the sustainability of the community. By considering different kinds of communication activities as an exchange of resources among users that derive benefits from them, our study seeks to develop an explanation for how Q&A communities can be sustainable.

One challenge for our work is that an individual's needs and the benefits they obtain are not visible in trace data. However, regularities in the behavioral patterns of site visitors provide a strong signal about the sorts of activities that satisfy their needs. These regularities have been described as *social roles*, and they can be an important tool in a resource-based analysis.

2.2. Social Roles in Online Communities

Welser et al. [29] built on Butler's [4] resource-based model by illustrating how online interactions between individuals in different social roles produce sustainability due to the balance between those resources sought and obtained. Simply stated, askers seek answers and provide questions, while answerers seek questions and provide answers.

In the symbolic interactionist tradition of social role theory [5], social roles are defined as cultural objects that are "recognized, accepted, and used to accomplish pragmatic interactive goals in a community". Studies have sought to identify roles using various methods [e.g., 6, 17, 28], and Gleave et al. [10] sought to standardize the usage of social roles for online media research. They argued that social roles have two key dimensions: structure, which refers to the patterns embedded in relationships and resources in a population; and culture, which means social roles are contingent on the social context of a group.

Practically speaking, one way to use trace data to characterize roles in online communities is by analyzing the behavioral metrics and relationships that emerge during participation in focal activities. This can be done in a data-driven approach. For example, Furtado et al. [8] mined and clustered behavioral patterns in multiple

Stack Exchange sites to identify ten different types of roles. We follow a similar approach here.

Roles are important for the resource exchange framework because individuals who adopt different social roles have different needs, and generate social structures that produce different benefits [3, 10]. Thus, whereas needs and benefits cannot be observed directly in trace data, social roles may be, and can be used as an observable proxy for pools of potential needs and benefits. For instance, in social Q&A communities, some individuals might provide the role of 'expert answerers,' who provide solutions for some thorny problems, filling a small but important niche in the *role ecology* [10] of the platform.

Prior work on social Q&A communities has focused significant attention on social roles [1, 6, 19]. The most salient roles in these spaces include question people, answer people and discussion people. In our study, we follow a data-driven approach to provide a more elaborate analysis of the roles that are important from a resource exchange perspective.

3. Research Questions

Butler proposed that the size of the membership base was critical to site sustainability [4], and Welser et al. [29] extended this analysis to show how different sub-populations can play a distinct role in a balanced resource exchange process. We continue this line of work to provide a more granular analysis of the `/r/excel` community, and further to provide a predictive model that helps isolate the critical factors underlying the community's sustainability. We frame our research around three research questions:

RQ1: What social roles can be identified based on community members' behaviors and their relational networks?

RQ2: How do interactions among individuals in these roles influence overall site activity (rate and types of contributions)?

RQ3: Which key factors appear to drive the sustainability of the system?

4. Dataset Description

We analyzed data from a large online Q&A forum hosted on Reddit.com, `/r/excel`. This forum is launched in 2009, and currently has more than 74,000 subscribers. Most of the posts in the community are questions about Microsoft Excel but there are also threads concerning general discussions and tips for the software. The community had a major design and management change in mid-2014, and an automated moderator was introduced to manage the status of the questions. Users can ask questions by starting new posts and later replies are organized as grouped messages, known as discussion threads. In addition to plain text, both questioners and answerers can use code and formula formatting or insert HTML links to facilitate the process.

The forum offers several socio-technical features that played a role in our analysis. First, and as will other Reddit forums, questions (and comments) receive a score that depends on how many people vote a question up or down. Another feature that is a key differentiator between it and other forums hosted on Reddit is the ability of members to tag a question as "solved." The original poster must perform this action, but there are several socio-technical factors that motivate this activity. First, the community guidelines explicitly state that question posters must mark a post as solved. Second, upon doing so, they will receive "ClippyPoints" which are public indicators of good community behavior. Finally, an automated bot will notify the original poster if they have ignored a question for a long period of time. For this reason, the "solved" status of a question is a reliable indicator that a question was indeed solved.

The dataset used in the study contains a trace of 29 months of activities in the community, starting from January 1, 2015, with 32,733 questions and 193,769 replies in total. To examine the longitudinal changes in the community, the data are discretized into 29 monthly time windows according to the creation time of the questions (thus corresponding replies belong to the same time window). The size of the time window is chosen to reduce the fluctuation in users' activities due to events like holidays and to ensure there are enough data points in each window so that reliable estimates can be obtained for further analysis.

5. RQ1: Role identification

5.1. Metrics Used

As discussed, following prior research on Q&A communities, we used three groups of metrics to identify social roles: *network relations*, *question posting behaviors* and *replying behaviors*.

To obtain the metrics of network relations, we transformed users' activities in the community into weighted directed networks, where each user is represented as a vertex, and the weight of each edge reflects the number of messages exchanged between the users (i.e. forum posts that reply to a previous poster). Directed networks are critical for two reasons. First, we are interested in the social role that individuals play in relation to their activities, rather than the strength of their relationships, and so the direction of messages is

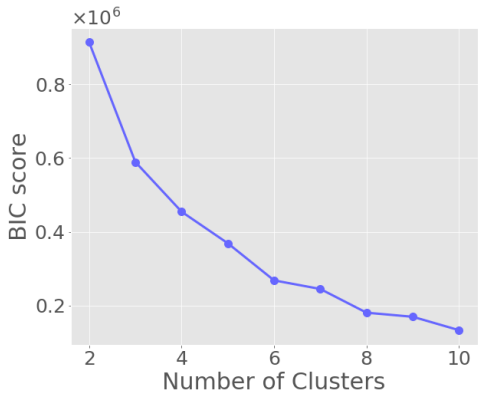

Figure 1 BIC score of each cluster solution

important. Second, directed edges indicate of how resources (carried by communication) flow among social roles.

We derive two user-specific metrics from this network. Outdegree (*Out*) is the total number of messages a user sends out, and the difference between indegree and outdegree (*Diff_In_Out*) is the number of messages a user receives minus the number of messages that user sends, which helps to capture the relative imbalance of a user's contributions.

Metrics of question posting behaviors capture the frequency of posting as well as the sophistication and utility of questions. For each user, we define: number of questions posted (*Num_Q*); percentage of questions that contains code/formula formatting and/or URLs (*Pct_Q_Special*); average length of the questions (*Avg_Q_Length*); and average score of the questions evaluated by other users (*Avg_Q_Score*).

Metrics of reply behaviors measure the responsiveness, sophistication, and utility of the replies. We include: number of the direct replies to the initial posts/questions (*Num_R_Direct*); average maximum depth of the replies in the discussion threads (*Avg_R_Depth*); average time ranking of the replies (*Avg_R_Timerank*), where all the replies in the same thread are ranked in ascending order based on its creation time (initial posts always have the highest rank); percentage of replies that contains code/formula formatting and/or URLs (*Pct_R_Special*); average length of the replies (*Avg_R_Length*); average standardized score (Z-score) of the replies (*Avg_R_Score*), where the score of each reply is evaluated by other users and standardized in relation to other replies in the same thread.

5.2. Clustering Algorithm and Results

To cluster our population, we follow Pal et al.'s [22] approach for identifying experts in a Q&A community. Pal et al. [22] used Gaussian Mixture Models (GMM) to identify clusters in the dataset. GMM can flexibly approximate the underlying density function of the data by using a combination of a finite number of Gaussian distributions, and can be considered as a generalization of the K-Means clustering method. One of the benefits of GMM is that the algorithm does not assume the independence of the data and can incorporate information about the covariance structure. Moreover, Bayesian Information Criterion (BIC) can be used to select the number of clusters (denoted by K) in an efficient way. Before the application of the algorithm, all data are standardized (Z-score) with respect to their own time window.

According to BIC, given a finite set of models, the model with a lower BIC value is preferred. However, since the value is likely to gradually decrease as the number of clusters increases. As in Pal et al. [22], we used visual inspection of the data to estimate an optimal cutoff, such that adding additional clusters did not provide much improvement in model fit Based on the results shown in Figure 1, the reduction of BIC value starts to level off at $K = 6$ and therefore the number of clusters is selected as 6.

Once the number of clusters is determined, the center, or the mean, of each Gaussian component is estimated and evaluated. Figure 2 presents the centers the clusters and based on the patterns, we developed a set of labels that we felt captured the characteristics of each cluster:

- *Frequent Questioner* (FQ), users who frequently post questions and the positions of their replies are deep in discussion threads;
- *Occasional Questioner* (OQ), users who infrequently post questions and their questions tend to be short and simple;
- *Occasional Answerer* (OA), users who infrequently post replies and send out more messages than they receive; their messages are short and simple;
- *Community Activist* (CA), users who send out a large number of messages and direct replies; they tend to be quick repliers;
- *Elaborative Questioner* (EQ), users who tend to post long questions with sophisticated formatting;
- *Experienced Answerer* (EA), users who usually post long and sophisticated replies and receive higher scores for their contributions.

The clustering results illustrate the diversity of users in the Q&A community. The FQs actively posts questions and are highly involved in the discussions while the EQs tend to be less active but more advanced questioners. Meanwhile, the CAs are extremely active posting a large volume of replies, and occasionally submitting questions. In comparison, the EAs are more marked by their ability than their activity levels. Occasional users (the OQs and the OAs) are less

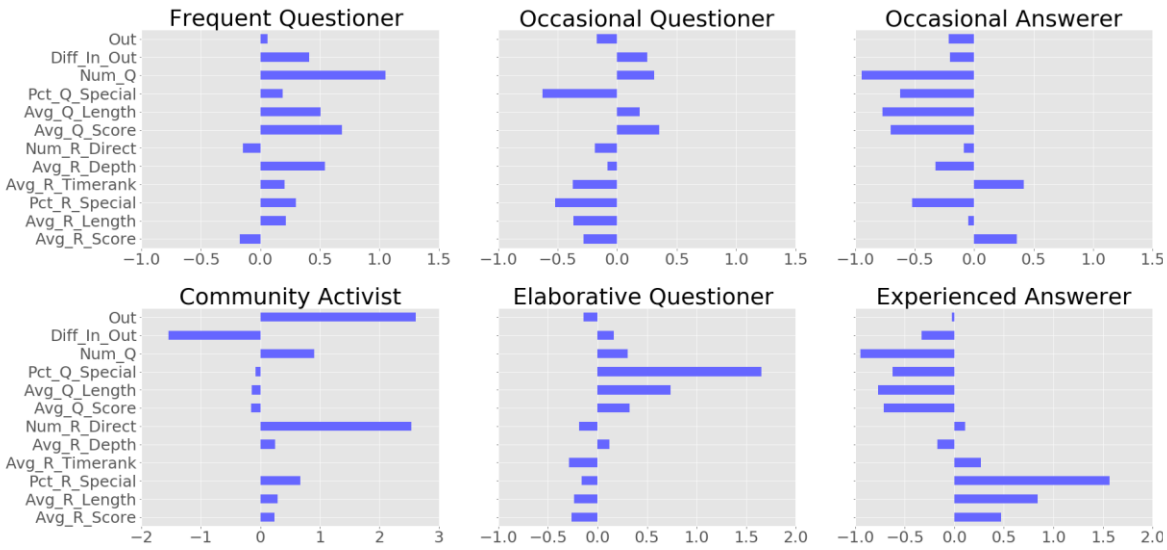

Figure 2 Centers of each cluster. Note that the horizontal axis is the standardized score and some of the horizontal scales vary.

engaged in the community as they are less active and their contributions tend to be simple.

6. RQ2: Role Dynamics and Community Traffic

6.1. Role Dynamics and Community Structure

As discussed, we use social roles as proxies for sets of needs and resources in the community. To understand how resources are exchanged across these roles, we constructed a directed network by aggregating the edges from members of each cluster. Figure 3 illustrates the proportion of the messages exchanged between roles (from row to column), averaged over all time windows.

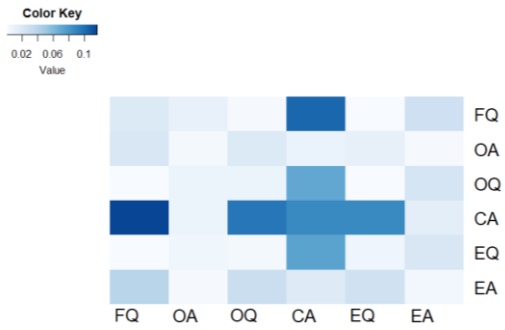

Figure 3 The average proportion of messages sent from one role (row) to another role (column).

In general, messages involving the CAs account for larger proportions of communication traffic; specifically, the exchange between the FQs and the CAs has the largest volume. In fact, the FQs and the CAs are the most active contributors in the community in terms of posting questions and replies, respectively. The distributions of messages in the OQ and the EQ group are similar, mainly concentrating on the interactions with the CAs, followed by the EAs and the OAs. The OAs' messages generally have the lowest volume.

Figure 4 presents the proportion of membership size of each role as well as the percentages of questions and replies contributed by each role. Occasional users are the largest groups, while the CAs make up the smallest. All of the questioner roles produce a similar proportion of questions, whereas the CAs produce a relatively small

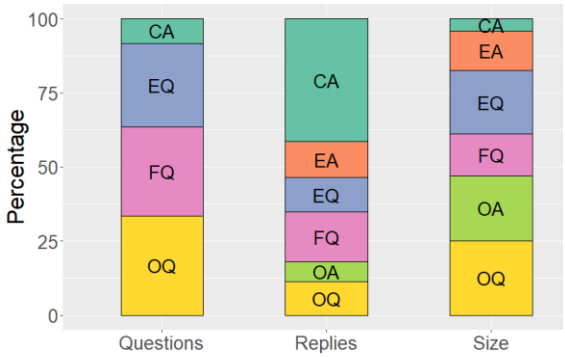

Figure 4 The proportion of each role and the percentages of questions and replies produced by each role.

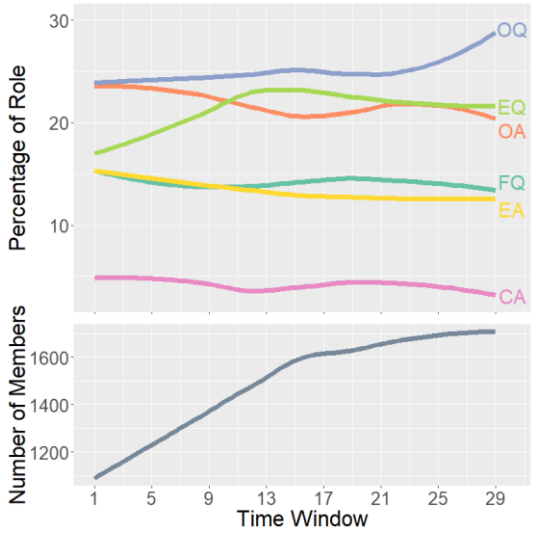

Figure 5 Longitudinal changes (with Loess smoothing) in the proportion of each roles (top) and in the membership size (bottom).

proportion. However, in line with the findings of other studies [8, 13, 18], the CAs contribute the most replies, despite being the smallest group. It is also notable that the FQs do engage in the discussion of their own questions, so the proportion of their replies is larger than the other two questioner groups. Figure 5 shows the longitudinal changes in the proportional makeup of the population by role (top) and the total number of members (bottom). The graphs reflect the active monthly population (the number of unique individuals who post in each monthly time window), and the plots are smoothed to highlight trends. Over the course of 29 months examined, the total active monthly population grows by nearly 30%. Note that the proportions of the OQs and the EQs increase while the OAs exhibits a downward trend. The other groups are relatively stable over time. Therefore, as the membership size expands, the proportion of questioners, especially the occasional ones, also increases. In the meantime, such expansion of the membership size is accompanied by a proportional increase in the size of the most active users (the CAs).

The analysis reveals disproportionate balance between questioners and answerers, and between different types of answerers in the community. 60.4% of the community members are the questioners while the most active answerers take less than 5%. In the resource-based model, the expansion of the resource base, measured by membership size, depends on attracting new and retaining existing members [4]. From the quantity perspective, in this technical support Q&A community, as the membership size and the proportion of questioners grow, more questions are produced, thus supporting the answerers' behaviors; meanwhile, the large volume of replies contributed by the active

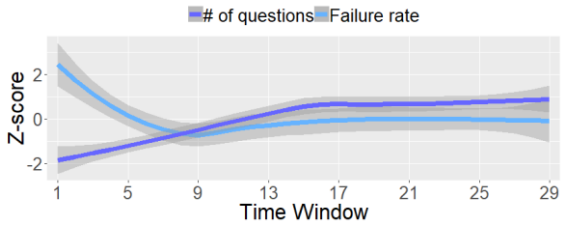

Figure 6 Changes of the number of questions and the failure rate over time, with Loess smoothing.

answerers may signal that the resources are readily available for those who are looking for solutions to their problems, thus attracting more new members to post questions.

The distribution of the resources, reflected by the composition of users' roles, is also important. A large proportion of questions comes from the FQs, and the interactions between the FQs and the CAs are more frequent than with other pairs of roles. This suggests that the FQs stimulate and sustain the CAs' behaviors. On the other hand, the stable proportions of CAs and EAs form the basis of a stable resource pool for information seekers, increasing the likelihood that they will obtain a solution. In the following, we build additional empirical support for these inferences.

6.2. Community Traffic, Sustainability and Functionality

To understand how the resource exchange process influences sustainability we seek to identify correlations among different factors and levels of communication traffic (i.e. posting rate). Although all proffered answers may be considered resources, those that are successful are particularly important. If a tech support community fails to provide useful solutions, users will be likely to cease to ask questions there and turn to other channels. We therefore focused our analysis on how the number of questions in each time window correlated with the proportion of questions that are marked as 'unsolved' (denoted as *failure rate*), in previous time windows. We use the number of questions rather than overall traffic because questions are a key external driver of activity on the site. Notably, the number of questions is strongly correlated with the number of replies generated in the same time window ($\rho = 0.93$, $p < 0.001$).

The changes of the number of questions and the failure rate over time are presented in Figure 6. Both time series are scaled (Z-score) and smoothed. During the data collection period, the number of questions is gradually growing whereas the failure rate drops. The cross-correlation between these two time series is estimated as -0.47 (see Table 1), with the number of

Table 1 Lag -1 cross-correlation between the failure rate and the number of questions (by role and in total). Values in bold are significant at the 0.05 level.

	FQ	OQ	CA	EQ	All
Cross-Correlation	**-.48**	**-.39**	-.35	**-.44**	**-.47**

questions lagging one time window behind the failure rate. The result indicates that the failure rate negatively predicts the number of questions in the community; therefore, a higher failure rate can lead to the decrease of community traffic in the future. In addition, the lag 1 autocorrelation of the failure rate is estimated to be 0.49, and the value drops below the significant level as the lag increases, suggesting that the failure rate is strongly correlated with the failure rate in the previous month.

The cross-correlations by role, shown in Table 1, further enrich our understanding of the forum. The FQs' questions have the strongest negative correlation with the failure rate, followed by the EQs' and the OQs', while the correlation between the number of the CAs' questions and the failure rate is not significant. Therefore, as the failure rate increases, the FQs and the EQs are less likely to post questions in the future. This suggests that the FQs and the EQs derive the most direct benefit from incoming questions.

One plausible explanation for these results is that an increased failure rate leads users to become reluctant to post new questions because they think their information needs cannot be adequately satisfied in the community, thus reducing the overall posting traffic in the forum. The impact of increased failure rate is the greatest of the FQs and the EQs, who are responsible for a large number of questions. The functionality of the community, largely maintained by the CAs and the EAs, thus plays a crucial role in attracting and retaining these more engaged questioners. This finding provides a basis for a predictive model, presented in the next section.

7. RQ3: Predicting Question Outcome

The previous section examined the connection between the ability of a community to successfully answer questions (henceforth referred to as its *functionality*) and posting traffic. Our findings suggest that the functionality of a forum is a critical driver of posting traffic. We now seek to connect this finding back to the activities of other roles, to develop a predictive model of the forum's functionality, and hence its sustainability.

To do this, we used a random forest classifier to determine which factors predict the outcome of a question. The outputs of a random forest classifier indicate the relative importance of a set of features, where importance is an indicator that may be intuitively interpreted as how much that feature contributes to the variation in a response variable (reported as *Mean Decrease in Impurity (MDI)*). In this task, the response variable is a dichotomous indicator of whether the question is solved or not. The sample consists of 32,590 threads, of which 63.2% are marked as 'solved'. The features used in the task are drawn from users' roles and activities in the community as well as the structural aspects of problem-solving conversations. Specifically, three classes of features are included in the prediction task:

- *Role configuration* (10 features): the role of the questioner, and the proportion of replies from each role (excluding the questioner) in the thread;
- *Community activities* (12 features): features describing the activities in the community happening 24, 72 and 120 hours before a question is posted. We included the proportion of unsolved questions, the number of repliers and the proportion of questions from each role, and the average number of comments in each question;
- *Thread structure* (4 features): the total number of comments, the thread's maximum depth, the number of unique branches (i.e., direct reply to the initial post), and the h-index (i.e., the deepest discussion tree level h which has at least h replies and is used as an indicator for controversy; see [11] for further details).

In sum, we developed an initial set of 26 features. We evaluated the classifier via 5-fold cross validation, and report accuracy and the area under the ROC curve (AUC). The community activity features with the three time-frames produce similar results, with the 24-hour timeframe performing slightly better. Here, we report results based on this timeframe.

On average, the classifier achieved an accuracy of 74.5%, an approximately 11% improvement over a random classifier, with an AUC score of 0.812. Table 2 gives the importance of the features whose importance scores are greater than 0.01 (the sum of all scores is 1.0). The proportion of the CAs' replies has the greatest impact on predicting the outcome of the question, followed closely by the number of comments. Overall, community activities in the day before a question made accounted for roughly 55% of the total MDI, role-based features for 22%, and thread-based features for 23%. Notably, the role of the questioner has small predictive power (less than 0.01) of their questions' outcomes, suggesting that the outcome of a question is more likely to depend on the interactions between users engaged within the question and the larger community than with the questioners themselves.

Table 2 Relative importance (RI) of features for predicting the outcome of questions. Note that only features with the score greater than 0.01 are included.

Community	Avg # of Previous Comments	.064
	Pct. of Unsolved Questions	.061
	Pct. of OQ's Questions	.059
	Pct. of FQ's Questions	.059
	Pct. of EQ's Questions	.059
	Pct. of CA's Questions	.056
	# of OA Users	.047
	# of CA Users	.047
	# of EA Users	.046
	# of FQ Users	.024
	# of EQ Users	.014
	# of OQ Users	.013
Thread	# of Comments	.090
	Max. Depth	.053
	# of Branches	.047
Role	Reply from CA	.092
	Reply from EA	.054
	Reply from OA	.040

Our results help illustrate that the responsiveness of the CAs is a central factor in successful questions, in two ways. Not only is the proportion of replies by the CAs an important on its own, but because the CAs tend to reply more quickly than other members (the EAs and the OAs), they also may provide questioners with important signals about the availability of resources. This is especially important for the FQs, who are likely to cease participating in the discussion if they do not receive responses in a short amount of time. This finding has been previously reported in Anderson et al.'s analysis of Stack Overflow [2].

However, the CAs are not solely responsible for developing the lengthy discussions that are also important in successful questions. We examined the relationship between the number of comments and the role composition in a thread, using a multiple linear regression model with the number of comments (log scaled) as the dependent variable. We found that the number of comments is significantly positively predicted by the proportion of the EAs' ($\beta_1 = 0.15$, $p < 0.01$) and the OAs' replies ($\beta_2 = 0.25$, $p < 0.01$), while the effect of the proportion of the CAs' replies is small and not significant ($\beta_3 = -0.01$, $p = 0.65$). Hence, the inputs from different roles may benefit the discussion by offering more information to the questioner.

More generally, the strong predictive power of the community activity—including the type of community members who are asking questions, who the repliers are, and how all members contribute—in the day leading up to a question may reflect the resources the site can muster to respond to a question. This interpretation is

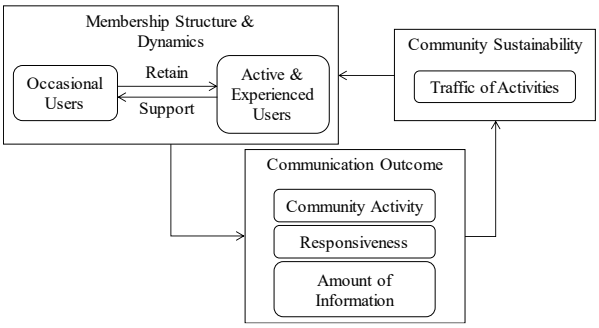

Figure 7 Schematic of the inferred resource exchange process in /r/excel

consistent with Butler's observations [4], about member size and resource availability, but also highlights the sensitivity of the platform to short-term fluctuations.

8. Discussion

From a resource-based perspective, community members are providers of different kinds of resources, and have different needs, but these cannot be directly observed in historical trace data. However, social roles correlate with "bundles" of resources and needs, and roles are revealed through individuals' behavioral regularities and network signatures. Thus, by identifying interaction patterns among different social roles, the analytical procedure we have followed helps elucidate the resource exchange process.

Earlier work with the resource-based model focused on membership size without delineating the more granular patterns in the resource exchange process sustaining a community. Our analysis helps extend the resource-based model by illustrating some of the complexity underlying the resource exchange process in a Q&A community.

In contrast with Welser et al.'s [29] analysis we find that the sustainability of the community is more complex than the balanced exchange of questions and answers. In Figure 7, we offer a schematic depiction of the resource-exchange process we infer from our findings. Notably, time becomes a much more important factor in our view of the resource exchange process. The responsiveness of the CAs, who appear to devote a significant amount of time to monitoring and engaging with the platform, is a key driver of site activity. Without it, questioners are likely to disengage from the platform, and seek other venues.

However, the continued attention of the broader population is a stronger indicator of whether or not the platform will be able to answer a question successfully. This echoes Page's [21] theory that diverse populations

are better at solving difficult problems than more homogeneous, expert populations.

In summary, we observe several nested patterns. The CAs are continually engaged on the site, responding to one another and also incoming questions quite rapidly. This initial responsiveness is important for questioners, and it may provide questioners with an early indicator that an answer will be forthcoming. This initial activity may serve to bridge the gap between the time when a question is posted, and when the community can actually produce an answer. At the same time, a steady stream of questions will help to keep this larger group of diverse but somewhat less responsive users engaged. The continued engagement of this diverse population is important in the functionality of the forum.

It is premature for us to draw general conclusions from these findings. Ours is a single case study, and is limited in several ways. Without qualitative data, we cannot know what anyone actually 'needs' and what benefits they derive from the platform. In particular, we have little insight into why the CAs are so active on the platform. In light of other research on social platforms [e.g., 13, 17], our finding the CAs often reply to other CAs might indicate that the Reddit Excel community is an important virtual space for socialization for these members. This is an important avenue for future work.

Nonetheless, we can extract a range of insights that are of value for the `/r/excel`, and might be useful for designers of other platforms as well. First, because responsiveness is important, system designers and moderators may want to optimize the real-time display of the system status so that the active members can be more efficiently directed to the threads that need attention. For instance, designers might offer support for push notifications and "dashboard" interfaces that allow active members to quickly assess the status of a forum.

We also note that a continuing stream of questions helps to keep the broader population of answerers engaged on the platform. A gap in the stream of questions could have cascading effects that lead to further reductions in the stream of questions. Moderators might use competitions or actively recruit questioners to fill such gaps. At the same time though, unsolved questions may dissuade future questions, so moderators should strive to keep them from piling up. Affordances for moving unsolved questions to a less visible archive might reduce their potentially deleterious effect.

Finally, maintaining the diversity of the population is important for platform functionality. To help maintain this diversity, moderators and designers might seek ways to invite contributions from less frequent users. One possibility might be to provide a range of incentives that might appeal to different classes of users, and selectively reward initial contributions more heavily.

9. Conclusion

In this paper we have extended work on the resource exchange model of online communities, providing a granular analysis of the communication patterns amongst distinct social roles on a social Q&A site. Through our analysis, we are able to identify several features that we believe are essential both to continuing traffic on the site, and its ability to function effectively as a technical Q&A support platform. Although our findings are likely to be specific to the platform we have analyzed, our methods can be easily replicated on other social platforms. Our findings illustrate the power of this systems approach for analyzing online communities, and we believe that following this approach will enable us to design more effective, sustainable socio-technical platforms in the future.

10. References

[1] Adamic, L.A., J. Zhang, E. Bakshy, and M.S. Ackerman, "Knowledge sharing and yahoo answers: everyone knows something", *WWW*, ACM Press (2008).

[2] Anderson, A., D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Discovering Value from Community Activity on Focused Question Answering Sites: A Case Study of Stack Overflow", *SIGKDD*, ACM Press (2012), 850–858.

[3] Baker, W.E., and R.R. Faulkner, "Role as Resource in the Hollywood Film Industry", *American Journal of Sociology 97*(2), 1991, pp. 279–309.

[4] Butler, B.S., "Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures", *Information Systems Research 12*(4), 2001, pp. 346–362.

[5] Callero, P.L., "From Role-Playing to Role-Using: Understanding Role as Resource", *Social Psychology Quarterly 57*(3), 1994, pp. 228–243.

[6] Danyel Fisher, H.W., Marc A. Smith, "You Are Who You Talk To: Detecting Roles in Usenet Newsgroups", *HICSS*, IEEE (2006).

[7] Dev, H., C. Geigle, Q. Hu, J. Zheng, and H. Sundaram, "The Size Conundrum: Why Online Knowledge Markets Can Fail at Scale", *WWW*, ACM Press (2018), 65–75.

[8] Furtado, A., N. Andrade, N. Oliveira, and F. Brasileiro, "Contributor profiles, their dynamics, and their importance in five q&a sites", *CSCW*, ACM Press (2013), 1237.

[9] Gkotsis, G., K. Stepanyan, C. Pedrinaci, J. Domingue, and M. Liakata, "It's all in the content: state of the art best answer prediction based on discretisation of shallow linguistic features", *WebSci*, ACM Press (2014), 202–210.

[10] Gleave, E., H.T. Welser, T.M. Lento, and M.A. Smith, "A Conceptual and Operational Definition of 'Social Role' in Online Community", *HICSS*, IEEE (2009), 1–11.

[11] Gómez, V., A. Kaltenbrunner, and V. López, "Statistical analysis of the social network and discussion threads in slashdot", *WWW*, ACM Press (2008), 645.

[12] Harper, F.M., D. Raban, S. Rafaeli, and J.A. Konstan, "Predictors of answer quality in online Q&A sites", *CHI*, ACM Press (2008), 865.

[13] Introne, J., B. Semaan, and S. Goggins, "A Sociotechnical Mechanism for Online Support Provision", *CHI*, ACM Press (2016), 3559–3571.

[14] Levine, J.M., and R.L. Moreland, "Group Socialization: Theory and Research", *European Review of Social Psychology 5*(1), 1994, pp. 305–336.

[15] Liang, Y., "Knowledge Sharing in Online Discussion Threads: What Predicts the Ratings?", *CSCW*, ACM Press (2017).

[16] Liu, Z., and B.J. Jansen, "Factors influencing the response rate in social question and answering behavior", *CSCW*, ACM Press (2013), 1263.

[17] Maloney-Krichmar, D., and J. Preece, "A Multilevel Analysis of Sociability, Usability, and Community Dynamics in an Online Health Community", *ACM Trans. Comput.-Hum. Interact. 12*(2), 2005, pp. 201–232.

[18] Mamykina, L., B. Manoim, M. Mittal, G. Hripcsak, and B. Hartmann, "Design lessons from the fastest q&a site in the west", *CHI*, ACM Press (2011), 2857.

[19] Nam, K.K., M.S. Ackerman, and L.A. Adamic, "Questions in, knowledge in?: a study of naver's question answering community", *CHI*, ACM Press (2009), 779.

[20] Nie, L., Y.-L. Zhao, X. Wang, J. Shen, and T.-S. Chua, "Learning to Recommend Descriptive Tags for Questions in Social Forums", *ACM Transactions on Information Systems 32*(1), 2014, pp. 1–23.

[21] Page, S.E., *The difference: How the power of diversity creates better groups, firms, schools, and societies*, Princeton University Press, 2008.

[22] Pal, A., S. Chang, and J. Konstan, "Evolution of Experts in Question Answering Communities", *ICWSM*, AAAI Press (2012), 274–281.

[23] Rechavi, A., and S. Rafaeli, "Knowledge and Social Networks in Yahoo! Answers", *HICSS*, IEEE (2012), 781–789.

[24] Srba, I., and M. Bielikova, "A Comprehensive Survey and Classification of Approaches for Community Question Answering", *ACM Transactions on the Web 10*(3), 2016, pp. 1–63.

[25] Toba, H., Z.-Y. Ming, M. Adriani, and T.-S. Chua, "Discovering high quality answers in community question answering archives using a hierarchy of classifiers", *Information Sciences 261*, 2014, pp. 101–115.

[26] Wang, G., K. Gill, M. Mohanlal, H. Zheng, and B.Y. Zhao, "Wisdom in the social crowd: an analysis of quora", *WWW*, ACM Press (2013), 1341–1352.

[27] Wang, G.A., H.J. Wang, J. Li, A.S. Abrahams, and W. Fan, "An Analytical Framework for Understanding Knowledge-Sharing Processes in Online Q&A Communities", *ACM Transactions on Management Information Systems 5*(4), 2014, pp. 1–31.

[28] Wang, X., K. Zhao, and N. Street, "Social Support and User Engagement in Online Health Communities", In X. Zheng, D. Zeng, H. Chen, Y. Zhang, C. Xing and D.B. Neill, eds., *Smart Health*. Springer International Publishing, 2014, 97–110.

[29] Welser, H.T., E. Gleave, D. Fisher, and M. Smith, "Visualizing the Signatures of Social Roles in Online Discussion Groups", *The Journal of Social Structure 8*(2), 2007.

[30] Yao, Y., H. Tong, T. Xie, L. Akoglu, F. Xu, and J. Lu, "Detecting high-quality posts in community question answering sites", *Information Sciences 302*, 2015, pp. 70–82.

[31] Zhang, J., M.S. Ackerman, and L. Adamic, "Expertise networks in online communities: structure and algorithms", *WWW*, ACM Press (2007), 221.