Modeling Similarity in Incentivized Interaction: A Longitudinal Case Study of StackOverFlow

Tanmay Sinha[1] **Wei Wei**[2] **Kathleen Carley**[2]

[1]**Language Technologies Institute**
Carnegie Mellon University
Pittsburgh, PA 15213
tanmays@cs.cmu.edu

[2]**Institute for Software Research**
Carnegie Mellon University
Pittsburgh, PA 15213
{weiwei, kathleen.carley}@cs.cmu.edu

Abstract

Multiple dimensions of Question Answering (Q&A) communities are indispensable to understanding its components, benchmarking success and catalyzing improvement. Gamification is one such dimension that has been shown to leverage user's natural desires for socializing, learning, achievement and altruism by providing opportunities for incentivized interaction. In this work we focus on dynamics of how much the community values a user, which is quantified via one salient instantiation of Q&A incentives - reputation. As research contributions, we first operationalize latent aspects of user participatory intent in a popular programming Q&A community (StackOverflow), and show that they significantly predict how much the community values the user. We then exploit underlying structure behind the online community to illuminate participatory patterns reflecting diverse user needs and preferred participation styles, and show how similarity in these patterns of interaction between users predict their reputation similarity over time. Our results, which differentiate underlying aspects of user interaction and their variations over time via a generic network analytic approach, have implications for community moderators in unmasking aggregated outcome measures such as reputation that look similar on a surface level. Modeling similarity in incentivized interaction is a step towards fine-grained tailoring of motivational affordances towards different participating classes of users in order to sustain their interest and engagement.

1 Introduction

In today's tech savvy world, increasingly many people have started leveraging the World Wide Web for various purposes such as communication, information seeking and education. With people exhibiting diverse ways of interacting with the technology, both online and offline, there arises the need for strong, active and lucrative support frameworks, which a)channel users toward finding relevant content, b)evoke long term contributions from experts on topics for which help is seeked, and c)encourage participation from newcomers [1]. Over the years, Q&A communities have grown to be excellent potential candidates for progressively building up this essential support framework on the web. Despite having made their presence felt in form of chatrooms and internet relay chats since decades, the advent of web 2.0 has led to proliferation of specialized Q&A communities catering to diverse user interests. Some of the core strengths of such Q&A communities lie in steering user participation and improving collaboration among them, while also reducing attrition by invoking game-like experiences through privilege escalation mechanisms such as badges, reputation points, bounties etc. At a broader level, by providing such opportunities for incentivized interaction, Q&A communities not only enhance their services, but also make the technology accessible and useful to all by ensuring that the online content is trustworthy and accurate. Prior work provides substantial evidence showing how these motivational affordances lead to psychological outcomes and further behavioral outcomes (for instance, motivation, engagement etc) [2].

In the current work, our objective is to study one among the many such incentives offered in Q&A communities, i.e, *reputation*, a score attributed to a participating user by virtue of his activities in the community that are made *technologically affordable*. Such activities may include providing answers, asking questions, casting votes as an expression of agreement or disagreement with others' opinions etc. The existing research on investigating reputation forums in online communities do not address specifically the questions of quantifying user's participatory intent and investigating its impact on how much the community values them. In addition, it is unclear whether similarity in user's participatory intent can help us apriori reason about similarity in their earned user reputation gain over time. Since a user's value to the community is dynamically influenced by how they naturally interact in the community over time, the summative construct of reputation is expected to exhibit a lot of longitudinal variance. Current published analyses of reputation forums in online communities have focused more at an individual level and have therefore not provided the needed visibility into how users tend to fall in similar reputation categories.

In this paper, we address these limitations in three ways by highlighting interaction mechanisms that differentiate online community users holding similar reputation points. First, in-part based on prior literature, we include in our analyses indicators of contribution to knowledge dissemination, discussion quality maintenance and passive information seeking derived from publicly available online community interaction logs, which potentially serve as proxies for participatory intent. Second, we leverage latent variable modeling in order to validate these latent aspects of user participatory intent and then use predictive modeling to demonstrate the generalizability of these aspects in predicting user reputation. Third, we employ a statistical model referred to as quadratic assignment procedure (QAP) in order to evaluate the factors that increase or decrease reputation similarity of users along the way, rather than focusing on the reputation of a single individual.

In the remainder of the paper, we first offer a more detailed review of important research thrusts related to the study of online communities in general. Next, we motivate the development of indicators of user reputation along the way, including measures of contribution to knowledge dissemination, discussion quality maintenance and passive information seeking. We then describe explanatory work on identifying how these measures associate to user reputation in our data via latent variable and predictive modeling approaches. We then present results from a quadratic assignment procedure model to investigate the factors that affect the likelihood of user reputation similarity in the community over time, while also suggesting a network analytic metric that can provide a finer-grained evaluation of specific user subgroups whose reputation similarity is more predictable than others. We conclude with discussion, limitations of our study and vision for future research, particularly focused on opportunities for development of a community progress index that can provides a holistic, objective, transparent and outcome-based measure of a community's growth.

2 Related Work

Q&A communities have been an active research area in the last couple of years. StackOverflow is a popular Q&A community in the domain of computer programming, relying heavily on crowd driven knowledge creation process [3]. Given this trend, there seems to be clear opportunity to add value for both community designers, producers and consumers. We briefly review related strands of work. The first strand of research deals with expertise detection. Existing methods have relied on quantitative properties of user activity in the community, partially accounting for effects that can be triggered by gamification incentives on these platforms [4, 5]. One interesting work tries to differentiate the behavioral patterns of highly active content producing users called sparrows and another group of expert users actually functional to knowledge creation called owls [6]. The second research thrust deals with analyzing various properties of Q&A in these communities, specifically focusing on a)what is being talked about, by leveraging topic modeling and qualitative coding approaches [7], b)predicting long lasting value of questions and whether they have been sufficiently answered [8], c)processes related to user activity around Q&A such as question response times [9] and effect of link sharing [10], d)predicting potential tags for posts made [11]. A substantial body of literature in this research foci has accounted for demographic factors such as age [12], gender [13], location [14] to explore representations and contributions to knowledge transfer.

The third research thrust most relevant to our research, focuses on steering user participation. First, this involves understanding personality traits of users that affects their participatory patterns [15].

Second, appropriate gamification incentives need to be optimally placed, in order to provide a lucrative environment for the users. Prior work along these lines has explored deployment of badges [16] and looked into reputation system in StackOverflow [17]. [18] have taken initial steps towards exploring how social interactions contribute to reputation building over time. In their analysis of StackOverflow, the authors considered adjacency matrices based on three types of interactions - between a user that asked a question and: a)any user who answered, b)the user who answered the accepted answer, if any, and c)any user who answered an answer that was upvoted. Finally, PageRank was used to identify important nodes and Singular Value Decomposition to identify anomalous questioner/answerer pairs. Although the work of [18] investigated what participation patterns distinguish high and low reputation users, they did not look specifically into the dynamics of reputation change over time. Moreover, they did not analyze the user interaction network from the perspective of how people tend to fall in similar reputation categories over time, as we demonstrate in our current analysis.

3 Reputation Building

The work of [8] describes how users on StackOverflow actually earn their reputation points, outlining different aspects that contribute to reputation change - questions and answers being upvoted (+5/+10) or downvoted (-2/-2), answer acceptance (+15), winning or offering bounties ($\pm$bounty amount), answers being marked spam (-100). As the authors mention, this system "is designed to incentivize users to produce high-quality content and to be generally engaged with the site". While votes on user contributions can be perceived to signal how much the online community values the user, we also believe that there are some other important latent aspects related to user's participatory intent that should be highly correlated with the aggregated reputation score. Since the primary purpose of Q&A communities is to facilitate knowledge flow by providing a healthy interaction environment where people can get quick, crisp and on-topic responses, the success of these crowdsourced systems depends a lot on people's intent to actively participate and help each other, in turn enabling formation of discussion networks in the community. In order to thus decouple the characteristics of highly reputed users, we started by operationalizing metrics of intent behind user interaction that could potentially affect reputation. For each of the identified factors below, we also describe explicit evidences of user interaction in the community that might reflect the characteristics of that factor.

First, we posited that highly reputed users should exhibit a positive intent by *contributing to knowledge dissemination (CKD)* in the community directly or indirectly. Traditionally, knowledge dissemination has been defined as "the transfer of knowledge within and across settings, with the expectation that the knowledge will be used conceptually or instrumentally" [19]. Because we expect such knowledge dissemination to be associated with positive outcomes such as increased awareness, ability to make informed choices among alternatives and the exchange of information, materials, or perspectives in the Q&A community, we operationalized CKD using the following factors: a)asking more questions and thereby providing an opportunity for more people to engage in discussions (*#questions asked*), b)providing more answers, indicating an increased propensity to participate in Q&A discussions and hence exhibiting a positive sign for knowledge sharing, healthy argumentation and increasing reciprocity (user's own question might be paid more attention in return) (*#answers made*), c)receiving more answers, indicating that the question has brought multiple and diverse problem solving perspectives together, thereby acting as a hotspot for influx of different ideas (*#answers received*), d)receiving more comments, indicating that the question has sparked off discussions in the community (*#comments received*).

Second, we posited that another important aspect of being a highly reputed user involves making contribution to *maintenance of discussion quality (DQM)* in the Q&A community. This serves to ensure that community standards are maintained in terms of specificity and appropriateness of questions and answers. Specifically, we inferred DQM from explicit user interactions such as a)upvotes casted, which acted as a proxy for appreciating good quality posting (*# upvotes casted*), b)downvotes casted, which acted as a proxy for regulating poor quality posting (*# downvotes casted*), c)marking questions as favorite, which acted as a measure for evaluating a post in terms of its overall quality and informativeness etc (*question popularity*). Affordances such as upvotes, downvotes and provision for star-marking favorite posts in the Q&A community neatly tie back to the criteria for discussion forum quality from prior literature [20] by allowing users to make decisions by providing proof so that the validity of information is assured (justification/judgment). Further, DQM serves to emphasize important issues on the topic of discussion, allowing participants to understand the focal issues related to any problem. We believe that highly reputed users would assist in prioritization of key knowledge in the community.

Finally, we posited that the third aspect of being a highly reputed user should take into account the fact that majority of people in online communities only view content. Such people would intervene or actively participate only when they want to seek help and hence would exhibit a bursty interaction pattern. This intuition also stems from the 90-9-1 rule in online communities, which states that 90% of the participants only view content, 9% of the participants edit content and 1% of the participants actively create new content. However, we believe that such *passive information seeking (PIS)* need not necessarily be a negative trait. Greater exposure might potentially lead to increased interest in participation and hence engagement in discussions. Specifically, we operationalized PIS by the page views made by a user (*#views*), which acted as a proxy for exposure to questions and ongoing discussions in the community.

4 Data and Explanatory Analysis

Our study context was a popular Stack Exchange Q&A community called StackOverflow, primarily focused on the computer programming domain. As of November 2014, this popular site had $3.7m$ registered users, $14m$ answers and $34m$ comments posted as responses to $8.5m$ questions categorized using $39k$ tags. In our work, we collected three roughly evenly spaced temporal data dumps of StackOverflow (licensed under a Creative Commons Attribute-ShareAlike license): a)October 2009 dump [1], b)August 2012 dump [2], c)September 2014 dump [3]. Each dump contained data collected since start of the Q&A community up to the specific month of the particular year. The dataset contained detailed information about a)participating users and their question-answering, voting, post editing activities that led to accumulation of badges, bounties and reputation points, b)posts and comments along with their edit/review history, attributes such as tags (that objectively characterize specific help seeking domains of the post) and flags (that objectively characterize post properties such as appropriateness).

In preparation for analyzing users with similar reputation, we first wanted to understand how the factors identified in section 3 affect reputation over time. At this stage, our work was exploratory rather than hypothesis driven. We began by informally selecting a data subset of $\approx$ 30000 active StackOverflow users (value of extracted interaction features > 0) from the start of the community till 2014, scaling their interaction variables and reputation using z-score, and finally dichotomizing user reputation along with all the interaction features into high and low categories based on equal frequency discretization criteria (which is invariant to the # users). Our goal was to investigate how the observed interaction features varied among low versus high reputation users, as well as what was the association of different feature combinations to reputation of users in the Q&A community.

Our qualitative visualizations (mosaic plot representations) revealed that the number of users belonging to low reputation class decreased (reflected by the change in class distribution) as user interaction features turned from low to high. Having laid the foundation for a conceptual model to understand the factors associated with user reputation in section 3, as a next step, we included estimates of Contribution to Knowledge Dissemination (CKD), Maintaining discussion quality (DQM) and Passive Information Seeking (PIS) in 3 latent variables, which intuitively relate to the propensity of users to participate in the community with a positive intent. These latent variables were further formalized by specifying associated set of observed interaction variables that we operationalized in section 3 (#answers made, #answers received, #questions asked, #comments received, #downvotes casted, #upvotes casted, question popularity, #views).

By leveraging the statistical technique of Structural Equation Modeling [21], we created three models for the years 2009, 2012 and 2014 respectively, evaluating each model using Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) [22]. Intuitively, for each variable in our latent factor set, we wanted to learn the associated weight parameter indicative of the contribution of each of the observed interaction variables. For the model fitting, we got a)RMSEA of 0.197 and SRMR of 0.122 with a CFI of 0.704 for 2009, b)RMSEA of 0.108 and SRMR of 0.072 with a CFI of 0.871 for 2012, c)RMSEA of 0.084 and SRMR of 0.053 with a CFI of 0.914 for 2014. Overall, the results suggested that

[1] http://tejp.de/files/so/dbdump/

[2] http://2013.msrconf.org/challenge.php

[3] https://archive.org/details/stackexchange

#answers made had the strongest influence on CKD latent factor, while #upvotes casted was most strongly associated with DQM latent factor. Interestingly, the link weights varied over time with, a)influence of #questions asked, #answers received, question popularity, #upvotes and #downvotes decreasing over time, b)influence of #answers made remaining nearly constant, while c)influence of #comments received increasing over time.

To further evaluate generality of the factors identified above, we created a predictive model to predict the two classes of user reputation (high/low) at the end of September 2014, using the CKD, DQM and PIS variables extracted from the 2014 data dump. Employing an L2 regularized Logistic Regression classifier trained using 10 fold cross validation, we varied the data sample size by randomly picking 10%, 30%, 50%, 70% and 90% of the data and observing variations in kappa values ((observed accuracy - expected accuracy)/(1 - expected accuracy)). For baselines, we began with the simplest models using CKD, DQM and PIS features alone, among which CKD features were the most influential predictors of user reputation class in 2014 (average kappa=0.76), while the DQM features were the least influential predictors of user reputation class in 2014 (average kappa=0.65). CKD+DQM+PIS, which was a combination of the latent aspects of contribution to knowledge dissemination, discussion quality maintenance and passive information seeking, further improved kappa by around 5% (average kappa=0.81). The combination of CKD+DQM+PIS+Reputation 2009+Reputation 2012 achieved the best performance in predicting user reputation in 2014 (average kappa=0.89), where Reputation 2009 and Reputation 2012 refer to the reputation score of a user till 2009 and till 2012.

These preliminary analyses helped us to validate and hone in on factors associated with highly reputed users, by showing that highly accurate predictions on user reputation category can be made leveraging these factors. In addition, they informed us (via an increasing/decreasing trend structural equation modeling link weights) that the intent behind user participation in Q&A communities changes with time, and hence provided a foundation for leveraging these factors in the rest of our work to model similarity in reputation. This also motivated our shift in focus from an individual to the dyadic (group) level, where our objective was to investigate how similarities and differences in interaction profiles among users might be reflected in their reputation similarities over time.

5 Method

Having established factors that contribute to reputation build-up in Q&A communities (StackOverflow) over time, we now turned toward investigating dynamics of the longitudinally earned reputation. More specifically, we wanted to analyze factors affecting reputation similarity through operationalizations of social positioning. For clarity, let's walk through this example - Consider user subset A in the online community having reputation greater than 2000 who have been mostly active answerers, while another user subset B also having reputation greater than 2000, but who have been mainly been active questioners and voters. The participatory patterns (potentially reflective of intent) of both these subset of users, which we will also refer to as interaction pathways henceforth, despite having similar reputation, are quite different. And, since these interaction pathways greatly reflect user needs and preferred style of participation, differentiating these underlying aspects of interaction and their variations over time is important. This will unmask aggregated outcome measures such as reputation that look similar on a surface level, in turn helping Q&A community moderators and designers to better direct their services toward different participating classes of users, for instance, by placing badges tailored to specific and more fine-grained ways of community involvement, rather than by just using the aggregated reputation score as a filter. To demonstrate our methodology in the following sub-sections, we randomly selected 10% ($\approx$ 3000) of active StackOverflow users from 2009 till 2014.

5.1 Attribute Similarity

Our *first research question* was the following: How was the increase or decrease in likelihood of similar participatory intent related to similarity in user reputation category in the Q&A community over time? We posited two supporting explanations. First, users might have a tendency to follow other users, and hence they might interact on the Q&A site in similar ways, in turn leading them to be connected to similar others who fall into the same reputation category as themselves (homophily hypothesis). Second, users could be influenced by other users in the same reputation category, in turn leading to their interaction mechanisms becoming similar over time (diffusion hypothesis). However, since reputation in Q&A communities is an outcome constructed as a result of user activities and interactions on the website, it is more likely that similar ways of participating in the community might lead users to fall in the same reputation category. This argument favored the homophily instead of the diffusion hypothesis.

In order to construct the dependent network, we first discretized the normalized reputation scores into 4 categories by equal frequency (low, medium, high, very high) for each data dump (2009, 2012 and 2014), and then formed a dyadic reputation similarity network, where 2 people had a link in the adjacency matrix if they fell into the same reputation category. However, we experimented with two ways for constructing independent networks: first, leveraging *exact attribute similarity*, where for each interaction feature (for e.g, #upvotes casted), users having *exactly* the same normalized value of the feature had a value 1 in their adjacency matrix (were connected); second, leveraging *approximate attribute similarity*, where for each interaction feature, we first discretized the normalized feature value into 4 categories by equal frequency (low, medium, high, very high) and then drew a binary link between users if they fell into the feature category.

Since standard regression methods can't be used to answer inferential questions in data that involves dependencies, we needed to first allow a representation that could focus on a dyadic level of interaction which was interdependent. Therefore, we utilized Quadratic Assignment Procedure (QAP) [23] to make stochastic inferences on the conceptualized network of reputation similarity and associated factors. Intuitively, the network analysis technique of QAP compares the degree of observed association between networks to that which would be expected to arise from a process in which individuals were randomly assigned to positions within the two networks, holding the structure constant. To calculate significance of the observed correlation, the QAP method compares the observed correlation to the correlations between large numbers of random sub-samples (matrices) that are just like the data matrices, but are known to be independent of each other (following a boot-strapping approach). The p value is then constructed by counting the proportion of these correlations among independent matrices that were as large as the observed correlation.

5.2 Subgroup Evaluation

Given that community users may be embedded in macro structures while participating in the Q&A community, and some of these macro structures could potentially be the interaction features that characterize various aspects of their exhibited behavior, we formulated our *second research question* as follows: Can we predict reputation similarity for *all* user sub-groups in the online community based on similarity in their participatory intent, or only for *specific* categories of user groups?

The E-I index measure [24] quantifies the extent to which macro-structures present in social networks cluster the connectivity patterns of individuals who fall within them. In our case, the E-I index can highlight what similarities and differences exist between users interaction patterns at a very fine-grained level. Concretely, given a partition of the dependent reputation similarity network into a number of mutually exclusive groups based on interaction features, E-I index would be defined as the number of ties external to the groups minus the number of ties that are internal to the group divided by the total number of ties. For our case, since we partitioned the reputation similarity network based on these interaction feature categories (group 1: low, group 2: medium, group 3: high, group 4: very high), internal links therefore corresponded to reputation similarity links in the same subunit and vice versa. Note than the E-I index ranges from 1 to -1 and can be seen as a measure of the extent a group chooses themselves. A value of -1 shows homophily and a value of +1 showing heterophily.

6 Results

In the *first stage*, by regressing interaction feature similarity networks (*exact similarity*) on reputation similarity networks using 50 permutations for QAP, and found the R^2 values to be low in general - 0.061 (2009), 0.055 (2012), 0.051 (2014). This meant that possessing information about the independent interaction feature similarity networks reduced uncertainty in predicting similarity in user reputation category by only 5-6%. In addition, the QAP intercept suggested that if users did not have exactly the same value on an interaction feature, their probability of falling into similar reputation category was 18-20% - 0.184 (2009), 0.202 (2012), 0.208 (2014). Interestingly, despite low R^2 values, we did have some significant predictors of reputation similarity ($p < 0.01$) that we depict in table 1. The unstandardized coefficient for these predictor networks indicate the increase in probability of falling into similar reputation category, if the network was known. For instance in table 1, if two users had exactly the same #upvotes casted up to 2009, this increased their probability of falling into similar reputation category by 0.268. We believe that one of the main reasons why the R^2 values following this approach are low is because the underlying predictor networks from which the QAP inference is made is fairly sparse. Using *exactly same* interaction feature value to form an interaction similarity link is a harsh criteria. And in fact on doing a sanity check, we found that the graph density of pairwise multiplex networks (formed using boolean combinations) [23] between the dependent reputation similarity network and each of the independent interaction feature similarity networks was very low too.

Thus, in the *second stage*, we regressed interaction feature similarity networks (*approximate similarity*) on reputation similarity networks using 50 permutations for QAP, and found the R^2 values to be comparatively much higher than the first stage - 0.207 (2009), 0.219 (2012), 0.225 (2014). These results suggest that possessing information about the independent interaction feature similarity networks now reduced uncertainty in predicting similarity in user reputation category by 20-22%, which is nearly 4 times the first approach. Significant predictors of reputation similarity ($p < 0.01$) are depicted in table 1.

Finally, we evaluated subgroups of similar reputation users via computation of E-I indices, by taking the significant interaction feature similarity networks discovered from approach 2 (*approximate attribute similarity*). Following are some of the important take-aways from results presented in table 2 - a)First, users falling into the highest and lowest category bins for the significant interaction feature similarity variables, have a tendency for group closure (negative E-I index), i.e, users of these sub-units are oriented inwardly rather than outwardly. Most of their reputation similarity links are within the same sub-unit, b)Second, the middle two bins are mostly heterophilous, meaning that interaction patterns in these bins are not very predictive of which reputation category users will fall into. There is a dominance of reputation similarity links outside the sub-unit. As a sanity check, we also discovered that partitioning the reputation similarity network based on non-significant interaction feature similarity variables discovered from QAP regression into the same four groups, led to partitions that were heterophilous (positive E-I indices).

Table 1: Regressing interaction feature similarity networks on reputation similarity networks using QAP. Only significant (**:p<0.01) unstandardized regression coefficients are shown. Approach 1: Exact Interaction Feature Similarity, Approach 2: Approximate Interaction Feature Similarity. Best scores (>0.150) are highlighted.

Approach	Independent Network used	Regression Coefficient (2009)	Regression Coefficient (2012)	Regression Coefficient (2014)
1	#questions asked	0.019	0.027	0.031
	#answers made	**0.160**	**0.236**	**0.225**
	#upvotes casted	**0.268**	**0.193**	**0.199**
	#downvotes casted	0.117	**0.205**	**0.220**
	#question popularity	0.029	0.051	0.063
	#views	0.095	0.094	**0.176**
	#comments received	-0.01	-	0.011
	#answers received	-	-	0.014
2	#answers made	**0.250**	**0.287**	**0.249**
	#upvotes casted	**0.161**	0.117	0.097
	#views	0.135	0.133	**0.204**
	#downvotes casted	-	-	0.135

Table 2: E-I indices for Reputation similarity network grouped by significant interaction features derived from leveraging approximate attribute similarity (A: #answers made, U: #upvotes casted, V: #views, D: #downvotes casted). Only users lying in groups 1 and 4 exhibit homophily tendency.

Group	2009			2012			2014			
	A	**U**	**V**	**A**	**U**	**V**	**A**	**U**	**V**	**D**
1 (Low)	-0.132	-0.147	0.040	-0.233	-0.077	-0.031	-0.276	-0.100	-0.181	-0.344
2 (Medium)	0.316	0.317	0.368	0.229	0.396	0.371	0.261	0.376	0.277	0.428
3 (High)	0.145	0.334	0.311	0.124	0.337	0.309	0.206	0.369	0.237	0.406
4 (Very High)	-0.431	-0.157	0.196	-0.398	-0.108	-0.186	-0.339	-0.031	-0.308	-0.002

7 Conclusion and Future Work

To summarize, in this work, we have shed light on the dynamics of reputation over time in the StackOverflow Q&A community. As a research contribution, our results effectively illustrate how the degree of association induced by the observed labeling of the reputation similarity and interaction similarity networks can be explained by their underlying structure. As a foundation for our study, we first validated and motivated the generic impact of the identified factors that affect user reputation, utilizing latent variable and predictive modeling. We then demonstrated the application of network analytic techniques such as QAP that can aid in analyzing and predicting how users in such Q&A communities tend to fall into similar reputation categories over time. Such a forecast can guide community designers and help them leverage this knowledge to optimally allocate resources and interventions, so that community participation (discussion) can be encouraged. Specifically, these data driven deployments would serve to improve user engagement and participation for factors along which less activity is expected in future, which might in turn lead to an overall decrease in user reputation score.

However, we believe there is a need for further modeling in order to fully understand the dynamics of the longitudinally earned user reputation. For instance, one limitation of our work is that we did not take network externalities into consideration. In the context of user reputation in Q&A communities, this effect refers to the additional value derived from being able to interact with other users in the community. Therefore, as a next step, we have begun to incorporate social interactions among users that are facilitated via Q&A postings into this framework. This will allow us to model the transitioning of users from one reputation category to another over time based on peer influence (for example, some users with a low initial reputation score might shift to a higher reputation category by participating in discussions with highly reputed users). In addition to monitoring changes in user interaction (for instance, answering significantly higher # questions than the before), the content and surrounding context of Q&A postings (for instance, similarity of topic-wise participation in different tag communities) also need to be taken into account. Second, interpretations drawn from our analysis must account for the noise inherently present in operationalizations of interaction feature categories. For instance, it can be argued that maintenance of discussion quality (DQM) and the passive information seeking (PIS) aspects used to describe user reputation could be manipulated to erroneously boost the reputation, and therefore an outlier pre-filtering layer before performing network analytics would be important. Third, CKD, DQM and PIS are by no means an exhaustive set of dimensions capturing user's participatory intent in Q&A communities and the current work is a step towards modeling and better utilizing these dimensions to make stochastic inferences. Fourth, since our analysis was solely done on the StackOverflow Q&A community in the computer programming domain, one should carefully consider how the present study would generalize to different Q&A community portals and domains. Thus, moving forward, we have also begun to extrapolate the concept of reputation as an index that aggregates diverse aspects of user interaction, toward developing a community progress index that can provides a holistic, objective, transparent and outcome-based measure of a community's growth. Specifically, we have collected longitudinal interaction data from 9 different Stack Exchange Q/A communities representing diverse application domains and operationalized measures of satisfaction and sustained engagement along multiple dimensions, which work in a feedback loop to help regulate interaction in Q&A communities. Preliminary results can distinguish between communities that follow a decreasing, increasing and relatively stable longitudinal progress trend. We are however, working further, to incorporate other potential indicators for community progress and conduct a more rigorous evaluation.

Acknowledgments
This work was supported in part by the Office of Naval Research (ONR) through a MURI N000140811186 on adversarial reasoning, DTRA HDTRA11010102, MINERVA N000141310835 and by Center for Computational Analysis of Social and Organization Systems (CASOS). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research, Defense Threat Reduction Agency (DTRA) or the U.S. government.

References

[1] Robert E Kraut, Paul Resnick, Sara Kiesler, Moira Burke, Yan Chen, Niki Kittur, Joseph Konstan, Yuqing Ren, and John Riedl. *Building successful online communities: Evidence-based social design*. Mit Press, 2012.

[2] Juho Hamari, Jonna Koivisto, and Harri Sarsa. Does gamification work?–a literature review of empirical studies on gamification. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, pages 3025–3034. IEEE, 2014.

[3] Lena Mamykina, Bella Manoim, Manas Mittal, George Hripcsak, and Björn Hartmann. Design lessons from the fastest q&a site in the west. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 2857–2866. ACM, 2011.

[4] Benjamin V Hanrahan, Gregorio Convertino, and Les Nelson. Modeling problem difficulty and expertise in stackoverflow. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion*, pages 91–94. ACM, 2012.

[5] Nidhi Raj, Lipika Dey, and Bhakti Gaonkar. Expertise prediction for social network platforms to encourage knowledge sharing. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference on*, volume 1, pages 380–383. IEEE, 2011.

[6] Jie Yang, Ke Tao, Alessandro Bozzon, and Geert-Jan Houben. Sparrows and owls: Characterisation of expert behaviour in stackoverflow. In *User Modeling, Adaptation, and Personalization*, pages 266–277. Springer, 2014.

[7] Anton Barua, Stephen W Thomas, and Ahmed E Hassan. What are developers talking about? an analysis of topics and trends in stack overflow. *Empirical Software Engineering*, 19(3):619–654, 2014.

[8] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Discovering value from community activity on focused question answering sites: a case study of stack overflow. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 850–858. ACM, 2012.

[9] Vasudev Bhat Adheesh Gokhale, Ravi Jadhav, and Jagat Pudipeddi Leman Akoglu. Min (e) d your tags: Analysis of question response time in stackoverflow.

[10] Carlos Gómez, Brendan Cleary, and Leif Singer. A study of innovation diffusion through link sharing on stack overflow. In *Mining Software Repositories (MSR), 2013 10th IEEE Working Conference on*, pages 81–84. IEEE, 2013.

[11] Avigit K Saha, Ripon K Saha, and Kevin A Schneider. A discriminative model approach for suggesting tags automatically for stack overflow questions. In *Proceedings of the 10th Working Conference on Mining Software Repositories*, pages 73–76. IEEE Press, 2013.

[12] Patrick Morrison and Emerson Murphy-Hill. Is programming knowledge related to age? an exploration of stack overflow. In *Mining Software Repositories (MSR), 2013 10th IEEE Working Conference on*, pages 69–72. IEEE, 2013.

[13] Bogdan Vasilescu, Andrea Capiluppi, and Alexander Serebrenik. Gender, representation and online participation: A quantitative study of stackoverflow. In *Social Informatics (SocialInformatics), 2012 International Conference on*, pages 332–338. IEEE, 2012.

[14] Dennis Schenk and Mircea Lungu. Geo-locating the knowledge transfer in stackoverflow. In *Proceedings of the 2013 International Workshop on Social Software Engineering*, pages 21–24. ACM, 2013.

[15] Blerina Bazelli, Abram Hindle, and Eleni Stroulia. On the personality traits of stackoverflow users. In *Software Maintenance (ICSM), 2013 29th IEEE International Conference on*, pages 460–463. IEEE, 2013.

[16] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Steering user behavior with badges. In *Proceedings of the 22nd international conference on World Wide Web*, pages 95–106, 2013.

[17] Amiangshu Bosu, Christopher S Corley, Dustin Heaton, Debarshi Chatterji, Jeffrey C Carver, and Nicholas A Kraft. Building reputation in stackoverflow: an empirical investigation. In *Proceedings of the 10th Working Conference on Mining Software Repositories*, pages 89–92. IEEE Press, 2013.

[18] Dana Movshovitz-Attias, Yair Movshovitz-Attias, Peter Steenkiste, and Christos Faloutsos. Analysis of the reputation system and user contributions on a question answering website: Stackoverflow. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 886–893. ACM, 2013.

[19] Janet R Hutchinson and Michael Huberman. Knowledge dissemination and use in science and mathematics education: A literature review. *Journal of Science Education and Technology*, 3(1):27–47, 1994.

[20] Dip Nandi, Shanton Chang, and Sandrine Balbo. A conceptual framework for assessing interaction quality in online discussion forums. *Same places, different spaces. Proceedings ascilite Auckland*, 2009.

[21] Kenneth A Bollen. Total, direct, and indirect effects in structural equation models. *Sociological methodology*, 17(1):37–69, 1987.

[22] Paul Barrett. Structural equation modelling: Adjudging model fit. *Personality and Individual differences*, 2007.

[23] Robert A Hanneman and Mark Riddle. Introduction to social network methods, 2005.

[24] David Krackhardt and Robert N Stern. Informal networks and organizational crises: An experimental simulation. *Social psychology quarterly*, pages 123–140, 1988.