

A Cognitive-based Scheme for User Reliability and Expertise Assessment in Q&A Social Networks

Konstantinos Pelechrisin^{*}, Vladimir Zadorozhny^{*}, Vladimir Oleshchuk[‡]

^{*}University of Pittsburgh

{kpele, vladimir}@pitt.edu

[‡]University of Agder, NO

vladimir.oleshchuk@uia.no

Abstract—Q&A social media has gained a great deal of attention during recent years. People rely on these sites to obtain information due to the number of advantages they offer as compared to conventional sources of knowledge (e.g., asynchronous and convenient access). However, for the same question one may find highly contradictory answers, causing ambiguity with respect to the correct information. This can be attributed to the presence of unreliable and/or non-expert users. In this work, we propose a novel approach for estimating the reliability and expertise of a user based on human cognitive traits. Every user can individually estimate these values based on local pairwise interactions. We examine the convergence performance of our algorithm and we find that it can accurately assess the reliability and the expertise of a user and can successfully react to the latter’s behavior change.

I. INTRODUCTION

During the last decade, advancements in computing and networking have drastically changed the way people acquire information. For example, printed sources of information and knowledge (e.g., scientific magazines, books etc.) are being supplanted by digital media, while functions of traditional libraries are being taken over by online digital libraries and search engines. In OSNs, users might seek help from their peers for specific topics. As an example, members of the Yahoo! Answers network can post a specific question, and the rest of the users are free to provide answers. The same is possible via the most popular OSN to date, Facebook, which has introduced a new feature called “Questions“. For quick answers, such online forums, Q&A SNs, online tutoring, etc., have the advantage of being asynchronous, often without requiring face-to-face communications, and in general being more convenient.

Common to all of these situations is the lack of vetting of these modern sources of information for their quality, correctness and accuracy, among other characteristics. For instance, in the physical world, an oculist is an eponymous source who has been recognized as an *authority* on eye diseases. The same holds true for a book used in a reputable medical school; its usage in the medical school automatically imbues it with the status of infallibility. On the contrary, it is clear that for information provided by an online source, the same property does not hold. In social psychology studies, people have been found to place a higher trust on information provided from sources classified as authorities [1], even though the classification (e.g., book used in university) itself is subjective. In [2], a study of how a diverse set of human participants searches for and appraises medical information, it was found that a “professional

look” of a web site made it appear to be more authoritative. Inappropriate banner ads affected the credibility of the site. Nevertheless, an unverified source can still be preferable to humans if it is easy to access and convenient. Studies have shown that individuals may rely on less trustworthy but more accessible sources to obtain the information they need even though the accuracy of the information itself is in doubt [3]. This however, increases the possibilities that their search is inadequate or less reflective, and the obtained information will be flawed.

The reputation¹ and the expertise of the answer *provider* has a direct impact on the quality of the information obtained. In this paper, we consider a novel approach to assess user reliability and expertise by utilizing human behavioral patterns. The main *fact* our scheme is based on is the **inability of a person to know everything about all things**. In other words, expertise is context dependent; Bob is a highly-reliable person, an excellent Java programmer and (with high probability) can answer any question with regards to this topic. However, he will not be able to answer questions about heart diseases. Since every question posted is related to a specific topic (e.g., “Java programming“, “Soccer“, etc.), we keep track of a user’s (say Jack) activity per category with the help of the *response matrix* (to be defined in the following). We define statistical metrics (estimated on the response matrix) that capture the compliance/deviation of Jack’s behavior with the expected profile and update our belief on his expertise and reliability. The main advantages of our assessment system are its lightweight nature and the fact that it can be applied *locally* from every user individually.

We would like to state upfront that the current work studies the characteristics of the assessment algorithm under the assumption that the cognitive traits considered are true. Challenging these traits is left for future work.

The rest of the paper is organized as follows. Section II provides a simple example illustrating the basic idea of our approach. Section III briefly discusses prior related studies. Our cognitive-based assessment scheme is presented in Section IV. Section V presents our simulation results, while Section VI forms our conclusions and discusses the scope of our work.

II. SYSTEM MODEL

Consider a simple scenario with two users, Bob and Jack, replying to each others’ questions about various topics. For

¹We use the terms user reputation and user reliability interchangeably.

our example, we consider three topics of interest: “Football”, “Medicine” and “Programming”. Each user should be able to judge the *quality* of the information obtained from the other user. Assume that Bob received some information from Jack related to “Medicine”. Intuitively, the quality of this information should be assessed using (1) Jack’s knowledge about “Medicine”, and (2) reputation of Jack. However, it would be unrealistic to assume that there is a globally consistent and adequate way to estimate both (1) and (2) for just any user. Achieving global consensus in such judgments is problematic even in relatively small user communities, and it is practically impossible in large scale social networks. Instead, we propose to estimate (1) Bob’s subjective opinion about Jack’s knowledge of “Medicine” and (2) Bob’s subjective opinion about Jack’s reputation.

In this paper, we introduce a scalable and automatic way to assess individual opinions as well as to further aggregate them. We utilize cognitive principles of human reactions to requests of information. If a user tends to respond consistently to questions related to a particular topic, we consider her knowledgeable in that area. Meanwhile, if the user is willing to reply to many remotely related topics, it would be safer to assume that this person is an amateur in each of those areas and her replies should be treated as less than reliable. We formally capture these behavioral patterns by maintaining pairwise user views of each other in the form of **response matrices (RM)**. Columns in a response matrix correspond to topics of interests, while rows reflect history of user responses. Figure 1 shows an example of two response matrices reflecting Bob’s views of Jack and vice versa. Note that Bob has a high opinion about Jack’s knowledge in “Programming” since Jack’s replies are consistently focused on this topic; Bob’s opinion about Jack’s reliability is also high, since Jack’s responses are not spread over various remote topics. Meanwhile Jack has a low opinion about Bob’s knowledge in “Medicine”, as well as Bob’s reliability.

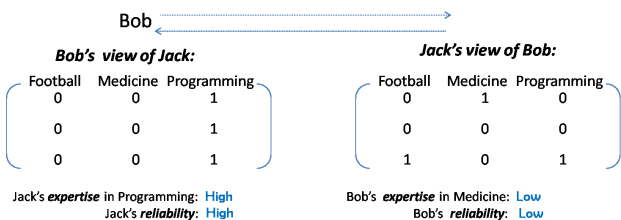


Fig. 1. Example of Response Matrices reflecting high and low opinions

To sum up, a user’s overall reliability is reflected through the spread of 1s over rows of the RM, while a user’s expertise in particular topics is represented as the density of 1s in the corresponding columns. In Section IV, we further describe our approach building on this example.

We would like to emphasize on the fact that opinions are generated and propagated automatically without explicit involvement of users. For this purpose we do not require users to evaluate the quality of responses from their peers.

III. RELATED WORKS

In this section, we will briefly discuss existing work on reputation systems and expertise inference.

Reputation systems: Reputation models have been primarily considered in the context of online electronic markets. Users of each specific market rate each other, and a centralized authority computes the trust value (reliability) on every single entity [4]. These computations are mainly based on simple statistics acquired from users’ feedback (e.g., positive and negative feedback). Sabater *et al.* [5] designed the regret system. They describe their scheme using an example borrowed from an online marketplace and they show how their system exploits the social relations among the different users. In brief, the reliability that a user (say Bob) has in any of his peers (say Jack) is based on their direct interactions as well as the interactions of *witnesses* (say Alice) with Jack and their social relation with him. Huynh *et al.* [6] introduce the notion of certified reputation. If Bob has no interaction with Jack and he cannot find any witnesses to report reputation information for Jack, Jack can present certified information about his past performance. These are essentially references from other agents who have interacted with Jack. Certified reputation is very useful for open multi-agent systems, where users can leave and join the system arbitrarily in time.

Expertise inference: There exists some related work in the literature that tries to assess the expertise metric. Zhang *et al.* [7] identify *expert* users in an online Java forum using a centralized approach that leverages social network analysis tools considering the network graph structure. Similar work can be found in [8], [9], and [10]. The well-known PageRank algorithm [11] ranks web pages based on their *popularity* on specific topics as seen from Web users; it can potentially form the basis for expertise inference [12]. A large portion of the existing work tries to locate expert users within a large network. For instance, Contact Finder [13] does not directly reply to users’ queries but identifies a set of peers that can provide a “good” reply to the specific question. Other similar approaches can be found in [14] [15] and [16].

There exist literature that deals with closely related and interesting issues from the perspective of cognitive sciences. For instance, [17] examines the way a user builds expertise. However, to the best of our knowledge, *to date there exists no work in the literature that tries to exploit cognitive and behavioral characteristics of humans to reach the joint estimation of reliability and expertise.*

IV. ASSESSMENT SCHEME

In this section we will present our scheme which estimates for user i (say Jack) his reliability r_i and his expertise $e_{i,q}$ on queries of type q (say “Football”), building on the example of Section II.

Response matrix (RM): The Q&A SN’s participating entities can be both consumers of information, as well as providers. When a consumer Bob asks a query he obtains responses directly from multiple providers (e.g., Jack). The goal of the

SN is to assess the quantities r_{Jack} and $e_{Jack,q} \forall q \in Q$, where Q is the set of different topics (in our case $Q = \{\text{"Football"}, \text{"Medicine"}, \text{"Programming"}\}$). Bob can obtain locally a *subjective* opinion about Jack's (i) reliability and (ii) expertise in q .

The first step is for Bob to derive the RM for Jack, $M_{Jack}^{Bob} \in \Pi^{w \times n}$; $\Pi^{w \times n}$ is the set of $w \times n$ matrices, w is the number of questions considered (e.g., posted from Bob) and n is the number of different topics. In our example we have $w = n = 3$. Note that there is no actual correspondence between the actual time and the rows except that the queries were made within the time interval T_{RM} corresponding to the RM. Thus, multiple "ones" in a row simply imply responses obtained to multiple queries on different topics within T_{RM} . A single RM can be thought as a single snapshot of the network (with respect to Jack's activity in Bob's view).

Assessment of $e_{Jack, \text{"Football"}}^{Bob}$: The expertise of Jack is tightly related with a *specialization*. An expert on one topic is expected to be rather engaged on the related questions. Thus, being *consistently* active is a sign of expertise in the corresponding category². For this task, Bob will use the column of M_{Jack}^{Bob} that corresponds to "Football" (let it be column j). Column j is a vector, denoted by $\vec{\lambda}_{Jack}^{Bob,j}(t) \in \mathbb{R}^{w \times 1}$, of "0"s and "1"s. $\vec{\lambda}_{Jack}^{Bob,j}(t)$ can be thought of as an observation vector. Its h^{th} element, denoted by $[\lambda_h(t)]_{Jack}^{Bob,j}$, is equal to 1 if Jack responded to the h^{th} "Football" question in the snapshot t ; otherwise, it is 0. We *measure* the interest of Jack in "Football" through his active participation in the corresponding discussions; this can roughly capture his *tendency* to be an expert in the field.

Each one of the questions in a snapshot can be thought as a Bernoulli trial X . The trial is successful if Jack responds. Thus, the probability of success p of X is equal to Jack's expertise on "Football", which we assume to be constant throughout the snapshot. In random variables terminology, the outcome of the h^{th} trial $[\lambda_h(t)]_{Jack}^{Bob,j}$, is 0 if Jack did not respond to the h^{th} "Football" question, and 1 otherwise. Hence, the probability density function (pdf) of X is:

$$f_h(X = \lambda_h) = p^{\lambda_h} \cdot (1-p)^{1-\lambda_h} \quad (1)$$

By replacing p with $e_{Jack, \text{"Football"}}^{Bob}$, the pdf described by Equation 1 can be thought as the formal definition of Jack's expertise. Given the expertise sample set we have collected, we use the Maximum Likelihood Estimation (MLE) [18] framework to obtain an estimate on parameter p . In particular, this estimate corresponds to the solution of the following optimization problem:

$$\max_p \frac{1}{w} \cdot \sum_{i=1}^w \log(f_i(\lambda_i|p)) \quad \text{subject to } p \in [0, 1] \quad (2)$$

Considering one snapshot/RM of the network at time t provides Bob with a single sample set of observations. Thus, by solving the MLE problem, he acquires a single point estimate

²Later, in Section V, we will describe scenarios where expertise is falsely inferred and how we can mitigate these occurrences.

$\tilde{p}(t)$. In order to compute the uncertainty on the expertise value with respect to Jack, we propose the use of m snapshots in time, which will provide m sample sets. Using the estimates computed from MLE for each of the above sets, Bob can compute the average estimator $\bar{\tilde{p}}$ and its standard deviation \tilde{p}_{sd} . In turn, this provides a method to obtain an expertise interval E of width \tilde{p}_{sd} , centered at $\bar{\tilde{p}}$.

Assessment of r_{Jack}^{Bob} : Reliability is a personality trait, related to the "good will" of an entity. Given its highly subjective nature, there are no clear metrics of Jack's reliability. However, as mentioned before, a reliable person can be *roughly profiled* as follows:

- 1) Given that Jack cannot be an expert in a large variety of topics, he is expected to reply to only a few topics. This translates to the matrix $M_{Jack}^{Bob}(t)$ of a reliable person being dominated by "0"s.
- 2) Reliable Jack is expected to consistently reply to the topics within his interest/expertise. This translates to the matrix $M_{Jack}^{Bob}(t)$ having a *minimum* number of "1"s which are *clustered* to a few columns.

Using the profile above we can formally define the r_{Jack}^{Bob} . Let R_1 be the number of "1" entries in $M_{Jack}^{Bob}(t)$. With δ_{xy} being the Kronecker's delta, $R_1 = \sum_{i=1}^w \sum_{j=1}^n \delta_{[m_{ij}]_{Jack}^{Bob}, 1}$. Furthermore, let vector $\vec{\pi}_{Jack}^{Bob} = [\pi_j]_{Jack}^{Bob} = [\sum_{i=1}^w \delta_{[m_{ij}]_{Jack}^{Bob}, 1}]_{Jack}^{Bob}$. Each element of $\vec{\pi}_{Jack}^{Bob}$ is the number of Jack's replies in each query category. Finally, let R_2 be the number of modes in the sample set $\vec{\pi}_{Jack}^{Bob}$ (see Appendix for details in the mode calculation). Then, Jack is considered to be *reliable*, that is $r_{Jack}^{Bob} = 1$, iff:

$$\alpha \leq R_1 \leq \beta \quad \wedge \quad R_2 \leq \gamma \quad (3)$$

Here α , β and γ are functions of the dimensions of $M_{Jack}^{Bob}(w, n)$. When the first part of (3) does not hold, we need to penalize Jack. For example, if $R_1 < \alpha$, Jack can be thought as acting *selfishly*; not providing any answers at all (even at the topics of his expertise)³. In this case, Bob penalizes Jack based on (i) the deviation of R_1 from its lower bound, that is $d_1 = \alpha - R_1$, as well as (ii) the deviation of R_2 from γ ($d_2 = \gamma - R_2$):

$$\begin{aligned} r_{Jack}^{Bob} &= y_1 \cdot (1 - \frac{1}{\alpha} \cdot (\alpha - R_1)) + y_2 \cdot (1 - \frac{1}{\gamma} \cdot (\gamma - R_2)) \\ &= y_1 \cdot \frac{R_1}{\alpha} + y_2 \cdot \frac{\gamma}{R_2}, \quad y_1 + y_2 = 1 \end{aligned} \quad (4)$$

If $R_1 > \beta$, Jack is *unreliable*. He is *talkative* and simply provides answers in many areas where he has no background. This can lead to the diffusion of low quality information. In this case, Bob penalizes Jack based on (i) the deviation of R_1 from its upper bound, that is $d_3 = R_2 - \beta$, as well as (ii) the deviation

³Of course, Jack might have no expertise at all and therefore, if reliable, he will exhibit extremely low activity. As discussed in Section V in the current work we are not interested into distinguishing between a selfish user and a non-expert.

of R_2 from γ ($d_2 = \gamma - R_2$):

$$\begin{aligned} r_{Jack}^{Bob} &= x_1 \cdot \frac{((w \cdot n - \beta) - (R_1 - \beta))}{(w \cdot n - \beta)} + x_2 \cdot \left(1 - \frac{1}{\gamma} \cdot (\gamma - R_2)\right) \\ &= x_1 \cdot \frac{(w \cdot n - R_1)}{(w \cdot n - \beta)} + x_2 \cdot \frac{\gamma}{R_2}, \quad x_1 + x_2 = 1 \end{aligned} \quad (5)$$

Note here that, the coefficients y_1 , y_2 , x_1 and x_2 , can also be functions of R_1 and/or R_2 . For instance, when $R_1 < \alpha$, it might be the case that the number of modes present (i.e., R_2) is within the limit of γ . In this case, we should not use d_2 (which is negative!) to penalize Jack, since he adheres to the expected behavior. Therefore,

$$y_1 = \begin{cases} \rho_y & \text{if } R_2 > \gamma \\ 1 & \text{otherwise} \end{cases} \quad (6)$$

$$y_2 = \begin{cases} 1 - \rho_y & \text{if } R_2 > \gamma \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Similar definitions can be given for x_1 and x_2 , controlled by a different parameter ρ_x .

However, even if the R_1 is kept below β it might be the case that this happens not because Jack focuses on his topics of expertise but because he is very little engaged to replying (spreading his low activity across a number of topics). Thus the right part of (3) needs to hold as well. In this case, Bob reduces the reliability of Jack based on the number of excessive modes present ($d_2 = \gamma - R_2$):

$$r_{Jack}^{Bob} = \begin{cases} 0 & \text{if } R_2 = w \\ \frac{\gamma}{R_2} & \text{otherwise} \end{cases} \quad (8)$$

Mapping intervals to opinions: For more convenient representation of our results, we will not directly make use of the intervals obtained from the assessment scheme, but we will represent them using the notion of *opinion* borrowed from subjective logic [19]. This will also enable the use of subjective logical consensus operators for combining opinions from multiple users in our future work.

Let t , d and u be non-negative values such that $t + d + u = 1, \{t, d, u\} \in [0, 1]^3$. Then, the triple $\omega = \{t, d, u\}$ is called an *opinion*, where components t , d and u represent levels of trust, distrust and uncertainty. For example, high distrust with some uncertainty (0.1) could be expressed as an opinion $\omega_1 = \{0.0, 0.9, 0.1\}$, while high trust with a minor uncertainty of 0.04 could be expressed as opinion $\omega_2 = \{0.96, 0.00, 0.04\}$.

Assuming that $r_{Jack}^{Bob} = [a, b]$, we generate the subjective logic opinions using the following equation (likewise, a mapping can be designed for the expertise opinion triplet $\omega_{Jack, "Football"}^{Bob}$):

$$\omega_{Jack}^{Bob} = \left\{ \frac{a+b}{2}, 1 - \frac{a+b}{2} - \frac{b-a}{2}, \frac{b-a}{2} \right\} \quad (9)$$

V. EVALUATIONS

In this section, we present our simulation set up and the evaluations of our assessment scheme.

A. Experimental Setup

In order to obtain the RMs, we emulate the behavior of an information provider. In our study, we are primarily interested in identifying four categories of users; “Reliable expert”, “Talkative expert”, “Reliable amateur” and “Talkative amateur”. The names are self explanatory but to give an example, a “Talkative expert” is someone who is a *real* expert on a few topics (as expected), but she is also replying to questions outside her specialization. On the contrary, a provider can be classified as “Reliable amateur” if she does not have any expertise (something which can also be common) and is sincere enough not to provide any uncertain answers to any category. Simply put, a “Reliable amateur” is aware of her non-expertise and provides only a few answers (e.g., for the questions she is certain about). We would like to emphasize the fact that we make the implicit assumption that providers are not *selfish*; thus, a real expert will always reply to questions that fall into her area of specialization. Otherwise it will be extremely hard, if possible at all, to distinguish between a “Selfish expert” and a “Reliable amateur”.

Every user in the network has an *a priori* fixed expertise in each topic (expertise vector) and a reliability value. By sampling two uniform distributions and comparing the samples with the real expertise/reliability, we can obtain the RMs based on the diagram depicted in Figure 2. Furthermore, unless otherwise stated, the values of the simulation parameters used are shown in the table in the same figure. Finally, in our experiments that involve dynamic behaviors, the notion of time is not tightly related with the absolute time (e.g., seconds). A *jiffy/time ticks* is equal to a full RM snapshot. In other words, time $t = x$, means that there exists x snapshots (i.e., $x \cdot n$ questions in total) since the time we started observing the network.

We would like to reiterate that currently we are only considering the existence of a reply or not, assuming that users strictly adhere to the cognitive profiles presented. Nevertheless, in reality the quality of an answer is not binary. In the future, we seek to use data from real networks in order to examine (i) the compliance of real users with the traits considered, (ii) the effectiveness of our inference schemes and (iii) the improvements possible (if any) by incorporating expert knowledge information.

B. Experimental Results

Recovering the real expertise/reliability: In our first set of experiments, we opt to examine the accuracy of the individual assessment scheme. We consider a set of 10 users to monitor. After obtaining the RMs, we apply our framework and obtain the corresponding opinions. We first examine the columns of the RMs in order to obtain an estimation for the expertise of the user with regards to each topic of interest. We then examine the structure of the whole matrix in order to assess its reliability. Denoting the real value of the attribute (topic expertise/reliability) with r , if $r \in [t-u, t+u] \vee |t-r| \leq p \cdot r$, $p \in [0, 1]$, we have a correct inference. The value of p dictates the strictness

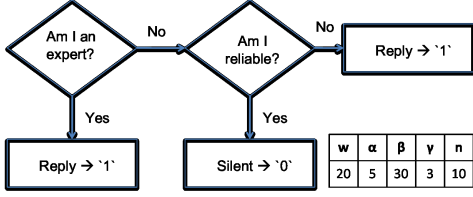


Fig. 2. Flow diagram of our user model and our simulation parameters.

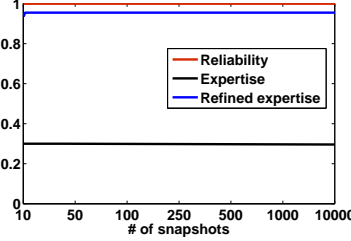


Fig. 3. Inference accuracy of our scheme.

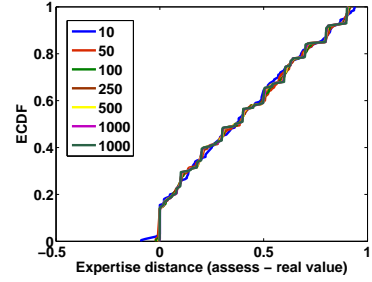


Fig. 4. Overestimating expertise (no refinement phase).

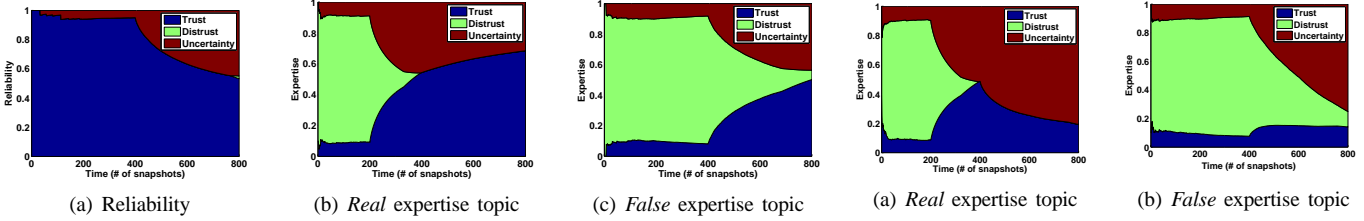


Fig. 5. Dynamics with no refinement phase.

Fig. 6. Dynamics with refinement phase.

of the convergence. Smaller values correspond to more strict convergence. In our experiments, we have set $p = 0.15$. Our results are depicted in Figure 3 where accuracy is shown for the different number of snapshots used. Despite the fact that we were able to recover the reliability for all the users, the accuracy with regards to the expertise is relatively low ($\sim 30\%$).

The reason for this performance can be attributed to the fact that when applying MLE on each column of the RM, the correctness of the answer is not considered. As a result, the presence of multiple “1”s in a column is considered as a sign of expertise. Nevertheless, a “Talkative” user will exhibit this pattern over more columns than just the ones of his actual expertise (if any). Thus, there will be an overestimation of user expertise which results in the low accuracy. Figure 4 depicts the CDF of the difference between the trust for the expert opinion and the real expertise value for different number of snapshot used for the estimation. As we can see with high probability, the inferred value is much larger than the actual one. For instance, with a probability greater than 40% this difference is greater than 0.5. In order to overcome this problem, we include a *refinement phase*. After estimating the reliability of a user (which is extremely accurate), we scale down the initial estimation of the expertise opinion (trust value) using the assessed reputation. Since it must hold $t + d + u = 1$, the expertise distrust and uncertainty both increased proportionally to their initial inferred values. As observed in Figure 3, the refined estimation is very accurate ($\sim 95\%$).

Response to dynamic behavior: The above results correspond to static scenarios; the (real) expertise and reputation values do not change during the network evolution. However, in reality a user might change her behavior over time for

various reasons. For instance, Alice is an expert in “Medicine“, but her account got compromised by Eve who is a computer scientist and knows nothing about medical questions. In this set of experiments, we seek to examine the effect of similar dynamic behaviors on the assessed quantities. We simulate 800 network snapshots with a behavior change taking place every 200 snapshots. The cycle followed is: “Reliable amateur“ \rightarrow “Reliable expert“ \rightarrow “Talkative expert“ \rightarrow “Talkative amateur“.

Figure 5 shows the reliability of a user (say Alice) along with her expertise (no refinement phase) with respect to two different topics. The real expertise topic corresponds to a subject in which Alice indeed has expertise (i.e., “Medicine“), while the false expertise topic corresponds to a category in which she is not knowledgeable at all⁴. As one can observe, Alice’s reliability follows the behavioral cycle we simulated. For the first 400 snapshots, her reliability is high, while for the rest of the simulation period her reputation degrades. The real reputation reduces to 0.1, however the degradation in the assessed value is much less steep due to the accumulated nature of the estimation (i.e., no RMs are ignored in the current scheme even if they correspond to old snapshots that might have become stale). The expertise assessment is more challenging as alluded to above. As can be clearly seen from Figures 5(b) and 5(c), when Alice becomes *talkative*, her assessed expertise is boosted in both types of topics. In the case of a non expertise topic for the period between 400-800 snapshots, Alice’s expertise is falsely increased. The same holds true for

⁴Note here that, even for the expertise topic, there can be periods for which Alice is an amateur and has no knowledge for this topic. This can correspond to periods where she is building knowledge, her account is compromised as mentioned above, etc.

the period between 600-800 snapshots in Figure 5(b), during which Alice is an amateur (e.g., due to her account being misused). However, if we examine the reliability and expertise assessments in combination, we can identify the periods of false expertise assessment, due to the low reliability of Alice. This falls back to the refinement phase we introduced in the static experiments. Simulating the same scenario using the refinement engine, we obtain Figures 6(a) and 6(b). As it is evident the non-expertise topic no longer exhibits false assessment. Furthermore, there is a degradation of the trust in expertise for the real topic of specialization, when Alice morphs from “Talkative expert” to “Talkative amateur” as should be the case. Nevertheless, there is a degradation of her expertise during the “Talkative expert” period. This is an expected outcome of the refinement performed: the trust in a user’s expertise degrades with the reduction of the user reliability. The fact that Alice is unreliable should affect our general trust in her replies. In our example, the rate of degradation is relatively small.

To summarize, under the assumption that the cognitive traits considered are true, our scheme provides high accuracy. Furthermore, it can react to users’ behavioral changes fairly fast.

VI. CONCLUSIONS & SCOPE OF OUR WORK

To date very little attention has been paid to the quality of the information delivered from a network. Assessing the expertise and reliability of an information provider is the first step towards quantifying this quality. Even though we focus on Q&A SNs, we believe that similar approaches can be taken for other kinds of data communication networks. For example, in a sensor infrastructure even if the reporting device is reliable (i.e., it has not been compromised by a malicious entity), its report might not be very accurate due to its physical distance from the event’s location. In this scenario, this distance defines the context related to the expertise attribute. Identifying the expected traits for an expert and reliable user of the underlying network is the only requirement to apply our framework in different scenarios.

REFERENCES

[1] John P. Kotter. *Power and Influence: Beyond Formal Authority*. Free Press, ISBN 0-02-918330-8, 1985.
 [2] G. Eysenbach and C. Kohler. How do consumers search for and appraise health information on the world wide web? qualitative study using focus groups, usability tests, and in-depth interviews. *British Medical Journal*, 324:573 – 577, 2002.

[3] B. Means, Y. Toyama, R. Murphy, M. Bakia, and K. Jones. Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies. In *U.S. Department of Education Office of Planning, Evaluation, and Policy Development Policy and Program Studies Service*, 2009.
 [4] e-bay. <http://www.auctionbytes.com/cab/abn/y06/m08/i01/s04>.
 [5] Jordi Sabater and Carles Sierra. Reputation and social network analysis in multi-agent systems. In *AAMAS*, 2002.
 [6] Trung Dong Huynh, Nicholas R. Jennings, and Nigel R. Shadbolt. An integrated trust and reputation model for open multi-agent systems. In *Journal of Autonomous Agents and MultiAgent Systems*, pages 119–154, 2006.
 [7] J. Zhang, M. A. Ackerman, and L. Adamic. Expertise networks in online communities: Structure and algorithms. In *WWW*, 2007.
 [8] L. Streeter and K. Lochbaum. Who knows: A system based on automatic representation of semantic structure. In *RIAO*, 1988.
 [9] M.S. Ackerman and D.W. McDonald. Answer garden 2: merging organizational memory with collaborative help. In *CSCW*, 1996.
 [10] B. Dom, I. Eiron, A. Cozzi, and Y. Zhang. Graph-based ranking algorithms for e-mail expertise analysis. In *DMKD*, 2003.
 [11] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. In *Stanford Digital Libraries Technologies Project*, 1998.
 [12] A. John and D. Seligmann. Collaborative tagging and expertise in the enterprise. In *WWW*, 2006.
 [13] B. Krulwich and C. Burkey. The contactfinder agent: Answering bulleting board questions with referrals. In *National Conference of Artificial Intelligence*, 1996.
 [14] H. Kautz, A. Milewski, and B. Selman. Agent amplified communication. In *National Conference of Artificial Intelligence*, 1996.
 [15] L.N. Foner. Yenta: A multi-agent, referral-based matchmaking system. In *Agents*, 1997.
 [16] H. Kautz, B. Selman, and M. Shah. Referralweb: Combining social networks and collaborative filtering. In *ACM Communications*, vol. 40, no. 3, 1997.
 [17] M. T. H. Chi, R. Glaser, and M. Farr. The nature of expertise. *Occasional Paper No. 107. Ohio State University, Columbus. National Center for Research in Vocational Education*, 1988.
 [18] Steven M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, ISBN 0-13-345711-7, 1993.
 [19] A. Josang. Artificial reasoning with subjective logic. In *Second Australian Workshop on Commonsense Reasoning*, 1997.

APPENDIX

Given a data set, the mode is the value that occurs more frequently. In our case the sample set $\vec{\Pi}_{Jack}$ is a vector whose i^{th} element π_i , is the number of responses from Jack with respect to category i . For a topic of expertise j we expect to have $\pi_j = w$, which will be the mode of $\vec{\Pi}_{Jack}$ (since this is the maximum possible value). By defining the set S as follows:

$$S = \{i | \pi_i \geq z \cdot \max_{k \in \{1, 2, \dots, n\}} \{\pi_k\}\} \quad (10)$$

we have R_2 to be equal to the cardinality of S , that is, $R_2 = |S|$. In our set of experiments we have set $z = 0.8$.