

Social Q&A

Rich Gazan

University of Hawaii, Department of Information & Computer Sciences, Library & Information Science Program,
1680 East-West Road, Honolulu, HI 96822. E-mail: gazan@hawaii.edu

This article presents a review and analysis of the research literature in social Q&A (SQA), a term describing systems where people ask, answer, and rate content while interacting around it. The growth of SQA is contextualized within the broader trend of user-generated content from Usenet to Web 2.0, and alternative definitions of SQA are reviewed. SQA sites have been conceptualized in the literature as simultaneous examples of tools, collections, communities, and complex sociotechnical systems. Major threads of SQA research include user-generated and algorithmic question categorization, answer classification and quality assessment, studies of user satisfaction, reward structures, and motivation for participation, and how trust and expertise are both operationalized by and emerge from SQA sites. Directions for future research are discussed, including more refined conceptions of SQA site participants and their roles, unpacking the processes by which social capital is achieved, managed, and wielded in SQA sites, refining question categorization, conducting research within and across a wider range of SQA sites, the application of economic and game-theoretic models, and the problematization of SQA itself.

Introduction

From the Cranfield experiments to the early days of the World Wide Web, information retrieval systems have operated on the same basic model: matching query representations with document representations. In social Q&A (SQA) sites, where users ask, answer, and rate content while interacting around it, queries are matched against an index of existing documents, and simultaneously become new documents themselves. Questions submitted to SQA sites undergo algorithmic, economic, and social processing strategies designed to entice individuals to append useful answers, ratings, and

comments, thus addressing the query and enriching the document with each contribution. With the onset of Web 2.0, an umbrella term referring to sites, systems, and applications that enable users to generate and share content on the Web, new systems and practices like SQA have made it possible for designers and users to fundamentally restructure traditional models of authorship, authority, access, retrieval, and relevance.

This review attempts to summarize and relate existing work in SQA, identify gaps in the literature, and suggest avenues for future research. It builds on an excellent review by Shah (2010), who focused on the broader realm of collaborative information seeking. One promising thread of research Shah identified was the need for a more holistic approach to the study of online information seeking, including awareness, user motivation, and information environments outside traditional single-user models. SQA sites, where the aggregate opinions of large numbers of individuals represent relative information quality, have been fertile grounds for this sort of holistic research, and what might appear to be a blizzard of informal information creation, provision, evaluation, and interaction on SQA sites is beginning to be better understood.

Terminology and Scope

The terms social Q&A and community Q&A appear with approximately equal frequency in the literature, but social Q&A is used throughout this review. Rosenbaum and Shachaf (2010) observe that the term “community” has been applied somewhat uncritically to Q&A sites as the literature has evolved, and they find conceptual merit for its application to Q&A sites in the large. However, in cases where participants in Q&A sites do not adopt and express an identity, engage in shared repertoire, or otherwise demonstrate formal indicators of community (Kling, Rosenbaum, & Sawyer, 2005), a broader term is needed. Hence, the term social Q&A, while it trades specificity for inclusiveness, was selected to scope this review.

Received February 2, 2011; revised March 24, 2011; accepted March 30, 2011

© 2011 ASIS&T • Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.21562

Shah, Oh, and Oh (2009) distinguish SQA from library-based digital reference services and ask-an-expert services on the grounds that in SQA, anyone can answer. Social reference, also called social information seeking, is a broader term, a conceptual superclass of the distributed information sharing activity that happens on SQA sites. Although an argument could be made that literature on digital reference services is an important precursor to SQA research, or that work on related Web 2.0 services such as social recommendation sites should be included, SQA sites are unique hybrids of social recommendation and question answering services, thus the focus of this review is limited to literature specific to SQA.

Forerunners

Before SQA sites existed, one thread of computer science research sought to improve on traditional document-based IR and focus on question-answering (QA) systems. Text REtrieval Conference (TREC) introduced a QA track in 1999, the goal of which was to conduct quantitative evaluations of systems designed to retrieve answers within documents, not just documents themselves, in response to test queries. Analyzing the syntax and structure of a query expressed in natural language to refine relevance judgments about potential answers promised an improvement over traditional keyword-based IR systems. The TREC QA track (Voorhees, 2003) has produced several papers about the superiority of a QA-based approach, though later researchers such as Lin (2007) found that traditional IR systems based on term frequency/inverse document frequency were “quite competitive” with QA-based systems. One of the primary emphases of the QA track has been to evaluate systems that can function in unrestricted domains, but TREC collections are generally limited to a small corpus of questions and focused on factoids, lists, and definitions as opposed to more social and conversational content.

Started in 1979, Usenet was a public bulletin board that comprised hierarchically organized newsgroups on a vast array of topics, where people posted and shared information in threaded conversations, usually moderated (Sudweeks, McLaughlin, & Rafaei, 1996; Fiore, LeeTiernan, & Smith, 2002; Viegas & Smith, 2004; Turner, Smith, Fisher, & Welser, 2005). Dedicated client software (newsreaders) did the work of sorting and displaying content, akin to Web browsers. Even in a low-bandwidth environment, individual newsgroups came to have diverse reputations and identities as places for support, measured discussion, or outright combat among participants. From Usenet were born now-familiar terms such as spam, trolls, and flaming, as well as emoticons. The overload of repetitive requests led to the creation of lists of frequently asked questions, or FAQs, which were collected by a moderator or other experienced participant in the newsgroup, stocked with canonical answers, or those hybridized from multiple past posters, a nascent form of SQA. However, once posted, there was no easy way for Usenet FAQs to be evaluated or amended by subsequent users as the content and communities evolved.

In an edited volume, Lueg and Fisher (2003) link the outgoing Usenet with incoming Web-based “social information spaces,” and paint a picture of one community, and then many subcommunities, each with self-generated membership and participation norms, created and maintained by their participants, surrounding topics of mutual interest. One of the primary differences in a Web 2.0 world is that online information can be rated, reposted, linked, commented upon, mashed up, and repurposed much more easily than was previously possible. Very few pieces of content are “safe” from the interpretations and opinions of subsequent readers. The popularization of peer-to-peer architecture in the original Napster file-sharing site exemplified the sort of disintermediation of traditional authority and expertise occurring across the Web, and it was in no small part these tools and this environment that allowed SQA to thrive.

Social Q&A

SQA sites exemplify the Web 2.0 model of user-generated and user-rated content. Anyone can ask and answer questions, rate content submitted by others, and view the community’s aggregate assessment of which questions, answers, and users are best. Sites featuring answers by experts are not strictly social Q&A, though many SQA sites offer a function to identify contributors as peer experts, given users’ assessments of their past contributions in particular topic areas. In their proposal for a social Q&A research agenda, Shah et al. (2009) propose a working definition of social Q&A as a site or service requiring:

- A method for users to present an information need in the form of a natural language question (as opposed to a keyword query)
- A forum for public response
- A community, based on participation level, in which the above transactions are embedded

They identify three primary research areas in social Q&A: user motivation and behavior, information quality assessment, and design and technological factors impacting participation.

In 2002, South Korean company NHN launched what is generally credited as the first SQA site, Knowledge-iN, as a component of its popular Naver search engine. The first SQA site in the United States was Answerbag, launched in April 2003, but when Yahoo! Answers launched in December 2005 (after a 6-month beta test), with its installed base of information-hungry searchers, SQA became popularized and institutionalized (Table 1). From its inception, Yahoo! Answers was by far the most widely used SQA site, with a reported peak of 62 million unique visitors per month in the United States alone in 2010. Only recently have competing SQA sites begun to challenge its popularity; after Answers.com merged several other properties into its overall database, its traffic approached Yahoo! Answers’ level at approximately 50 million unique visitors per month in the United States.

TABLE 1. Several of the more popular SQA sites.

SQA site	URL	Comments
Knowledge-iN	http://kin.naver.com	Also called Ji-Sik-In, first SQA site, Korean language, outranks Google in Korea
Yahoo! Answers	http://answers.yahoo.com	Most popular SQA site
Answers.com	http://www.answers.com	Composed of several sub-sites, e.g. WikiAnswers
Baidu Zhidao	http://zhidao.baidu.com	Also called Baidu Knows, Chinese language
Quora	http://www.quora.com	Hybrid SQA/wiki format
Answerbag	http://www.answerbag.com	First US SQA site
Askville	http://askville.amazon.com	Owned by Amazon.com
Blurtit	http://www.blurtit.com	Based in Norwich, Norfolk, UK
Fluther	http://www.fluther.com	Acquired by Twitter in 2010
Aardvark	http://vark.com	Real-time SQA
Yedda	http://aolanswers.com	Now AOL Answers
The AnswerBank	http://www.theanswerbank.co.uk	UK-centric SQA with quizzes, polls
Ask MetaFilter	http://ask.metafilter.com	SQA arm of Metafilter collaborative blog
Answerly	http://www.answerly.com	SQA metasearch site

Note. SQA = social Q&A.

Subject-specific, ask-an-expert, and pay-per-answer SQA sites are excluded.

The predominance of Yahoo! Answers as a data source in the literature is primarily because of the combination of its dominant market share and the ready availability of a subset of its data through a public application programming interface (API) (<http://developer.yahoo.com/answers/>). Other SQA sites are less forthcoming with their data, and researchers are limited to the data and tools available through the sites' public interfaces.

However, with the amount of data sharing and crossover traffic between sites, unique visitors per month is a less and less meaningful metric. SQA sites have created widgets or Facebook apps allowing users to access relevant topic-focused content from other sites, and there are differences in the rules and mechanics of site interaction that also have an effect on traffic. For example, Yahoo! Answers limits the length of time a question is open, from as little as four hours to a maximum of 8 days, and restricts the number of questions, answers or comments less experienced members can submit. Some systems allow askers to declare an answer the best, while others aggregate ratings and responses from other users indefinitely, and present the highest-rated answers first.

Conceptual Frameworks

The relative novelty of SQA sites has understandably yielded largely of descriptive, empirical research in the form of case studies and comparisons, but attempts to articulate theoretical frameworks surrounding SQA have begun to emerge. Understanding SQA requires considering broader questions about the value of socially produced information; from a pure supply-and-demand standpoint, the ease of creation and vast archive of existing information on the Web would seem to make SQA information essentially valueless. However, in a situation of oversupply, the same criteria people use to make relevance judgments and filter unwanted information are a starting point for analysis of the value created and derived from SQA information. Raban (2007)

summarized four overlapping frameworks for assessing the value of information: descriptive, rational, behavioral, and social. Following the work of Benkler (2006), Raban argues that the publishing and access of user-generated content has not fundamentally changed from traditional publication models, and that content vetting and selection have simply shifted to other actors and technologies.

Rosenbaum and Shachaf (2010) identify three points of contact between structuration theory (Giddens, 1979; Giddens, 1991) and communities of practice (Lave & Wenger, 1991; Wenger, 1998) germane to SQA: (a) the central role of social practices, (b) the importance of a fundamental duality (of participation and reification) through which the processes and structure of the community are maintained and changed over time, and (c) the extent to which identity is a critical outcome of the interactions among participants in the community. They cast SQA sites as instances of online communities of practice and call for research to encompass how communities are created and transformed as people interact both with and within them.

In a study of online travel communities, Lueg (2008) conceptualizes information seeking within online communities as not simple information exchange, but as a mediated interaction whereby the community, or individuals within it, help seekers understand and reconceptualize their information needs. While the day-to-day practices of asking, answering, and interacting imagined by the creators of SQA sites are where these structural negotiations play out, Rosenbaum and Shachaf (2010) also mention users engaging in "meta-conversations about the community and its rules" as more explicit evidence of user agency, echoing the literature on Usenet interaction patterns. Gazan (2009) identified 177 of these meta-conversations in the Answerbag SQA site, which included rituals of indoctrination and membership, debates about normative behavior, and the formation of subcommunities of like-minded users. These expressions of community self-awareness were associated with increased

levels of participation from high-ranking users, but also increased levels of conflict. The answers, comments, and conversations surrounding these meta-questions served as public spaces where competing ideas about appropriate content, rules, and behavior were debated.

The component of online identity formation suggested by Rosenbaum and Shachaf provides theoretical grounding for the idea that information exchange on SQA sites may not be motivated by classical notions of information retrieval and topical relevance. Sociologist Erving Goffman (1959) theorized that all social interaction has at its core a continuous impression management of the self one presents to others. Raban (2009) studied Google Answers and posited that the presentation of self is instrumental in SQA as well, and that answers alone are not enough to achieve social capital in the community. SQA sites operate on a model of aggregate peer authority (Gazan, 2008), a term which encompasses both the merged opinions of others about a single answer, but also the ability of those others to bestow site-specific expertise to a fellow user, based on the user's contributions as a whole.

Raban and Harper (2008) called for a conceptual framework for SQA encompassing both intrinsic and extrinsic motivations for participation. They suggest that an important component of why people participate in SQA sites is their perception that the overall system of SQA, where many minds merge to address common problems, has value and is worth perpetuating. To counter the argument that free-riding—deriving benefit from SQA content without contributing to it—is a systemic threat to SQA, Raban and Harper frame lurking as legitimate peripheral participation (Lave & Wenger, 1991), a necessary first phase of online community participation, where a prospective user observes the content and conduct of a community to better inform their decision about how and whether to ultimately participate. This conceptualization fits well with Rosenbaum and Shachaf's hybrid framework of structuration and communities of practice.

The meaning of reputation in a given community, and the processes by which it is achieved and leveraged to favor certain types and sources of content, must be understood in concert with a site's transaction log data and technological affordances, to have the most complete picture of the information being generated, evaluated, and exchanged. Chen, Ho, and Kim (2010), conceptualized Google Answers as a knowledge market, and while the price offered for answering a question was a significant factor in answer quality, they reported that the reputation of an answerer also helped lower barriers to Q&A transactions. Shachaf and Rosenbaum (2009) propose a Socio-Technical Interaction Network (STIN) framework approach as a conceptual lens for SQA. They describe STIN as a theoretical extension of social informatics, which accounts for the embedded co-evolution of social and technological systems (Kling, McKim, & King, 2003).

Shachaf (2010) suggests that social reference is characterized by collaborative work, and a unique, one-to-many structure of information seeking and exchange. Shachaf concludes that any theoretical framework that seeks to

understand, analyze, and evaluate social reference must account for social factors such as conflict management, trust, cohesiveness, motivation, coordination, and the maintenance of communication norms by the participants. Shachaf proposes an input–output–process model as a framework to identify and interrelate crucial variables in social reference environments, which can be summarized as follows:

Input variables

Task: question type, difficulty, clarity, topic

Users: participants, knowledge, skills, abilities, demographics, roles (askers, answerers, evaluators)

Context: information sources, site norms, policies, reward, and motivation

Process variables

Task: question negotiation, evaluation, categorization

Group: conflict management, motivation, trust, cohesiveness, team climate, norms

Output variables

Task: Answer quality (complete, accurate, verifiable, timely, sources cited)

Users: Satisfaction (askers, answerers, evaluators)

Context: Service (viability through participation, number of users, number of Q&A transactions, percentages of questions answered) and Technology (repository of past Q&A)

Although many SQA studies have focused on one or more of the variables proposed by Shachaf (2010), a holistic picture transcends any single study to date. Confounding a comprehensive approach is the fact that the most SQA sites make available to researchers only that data which is publicly viewable. A more significant barrier is that in the majority of cases, SQA site users cannot be contacted or studied directly, and reliable data about even basic demographic information is impossible to obtain in an online environment. Satisfaction measures and interaction patterns must be inferred from site-based metrics and quantitative analysis, and may not constitute an accurate picture of users' attitudes and behavior.

Question Classification and Retrieval

Some users seek information on SQA sites, while others seek conversation and contact. Early research in classifying questions submitted to SQA sites focused on factual content, and only after analysis did the need to include a class of social questions emerge. Jeon, Croft, and Lee (2005), using data from Naver, addressed the "lexical chasm problem" of similar questions being expressed with different words. They analyzed the answer text from pairs of similar questions in the Naver archive to create a probabilistic translation model and identify semantic similarities between words. They used this model to retrieve similar questions from the Q&A archive when a new question was submitted and found that their approach outperformed baseline retrieval models.

Ignatova, Toprak, Bernhard, and Gurevych (2009) used the Yahoo! Answers API to extract a sample of 755 questions related to data mining, natural language processing,

and e-learning, and they based their initial question type annotation framework on a pre-Web scheme developed by Graesser, McMahan and Johnson (1994) in the field of psycholinguistics:

- *Concept completion*: questions seeking to supplement known information
- *Definition*: of terms or acronyms
- *Procedural*: how-to
- *Comparison*: similarities and differences between two or more objects or concepts
- *Disjunctive*: objective or subjective opinion about relative merit of two or more objects or concepts
- *Verification*: confirmation of assumption included in the question
- *Quantification*: questions seeking a numerical answer
- *Causal*: seeking explanation for observed rules or phenomena
- *General information need*: including nonspecific requests

To account for ambiguous or cross-cutting questions, they allowed annotators to assign labels from two of the above categories if needed, and questions could also be flagged as seeking opinion. The researchers found that most questions in the subject area observed fell into the concept completion category, which they suggested might be usefully subdivided in future research. Questions seeking causal information were rare, suggesting that participants had a baseline level of knowledge, or possibly that they did not wish to risk asking questions that would give the impression they lacked basic knowledge. Close to one fifth of the questions were lexically, syntactically, or semantically ill-formed or otherwise not clearly expressed, and these questions yielded the highest rate of inter-annotator disagreement. The researchers also called for a more formal investigation of opinion-seeking questions in future work.

Building on the work of Pomerantz (2005) and Ignatova et al. (2009), Harper, Weinberg, Logie, and Konstan (2010) took a broader approach, and proposed a taxonomy of question types flexible enough to accommodate factual and opinion-based questions, as well as a universal range of question topics. They attempted to identify structural similarities in question types using rhetorical analysis and classified questions based on a framework derived from Aristotle's "species" of rhetoric:

- *Deliberative*, or future-focused questions, such as those seeking advice or contact with like-minded individuals
- *Epideictic*, or present-focused questions, such as those seeking opinions or consensus, whether subjective or objective
- *Forensic*, past-focused questions, such as how-to questions or seeking facts

A term frequency analysis suggested that questions including the words "I" or "me" tended to be seeking advice or information to support future action, and those including "you" or "your" sought identification with others in a more conversational sense.

Cao, Cong, Cui, Jensen, and Zhang (2009), preferring the term community QA, used the category structure of Yahoo! Answers in an attempt to improve the retrieval of relevant

existing questions from the archives of SQA sites. Their method uses hierarchical clustering and local smoothing methods to calculate a probability of a new question belonging to a set of existing categories, and it also accounts for improperly categorized questions within the corpus. Their results suggest a significant benefit to leveraging the category of past questions, though their framework does not consider answers and comments within questions as possible sources of relevant information, which they suggest as a direction for future work. Cao, Cong, Cui, and Jensen (2010) built on their prior work and proposed a more refined category-enhanced retrieval model that interpolates two relevance scores: a global score of a query and the category of a related question, and a local score of a query's relevance to the question without regard to its category.

Retrieval of SQA questions can also be improved by incorporating question paraphrases. For example, on WikiAnswers, a SQA subset of Answers.com, upon submitting a question, users are presented with possible matches from existing questions. Users have the ability to tag questions as reformulations, or alternate wordings of the same question. Bernhard and Gurevych (2008) used these reformulations as a "gold standard" for improving retrieval performance. They sampled 1,000 questions and 7,434 semantically similar paraphrases from the Education category of WikiAnswers, and they sought to match a target question in the smaller question set by retrieving relevant question paraphrases. Using a combination of query processing strategies and similarity metrics, they reported more than 80% question answering accuracy when presenting question paraphrases from the existing WikiAnswers collection.

Although a preponderance of research has identified semantic and structural differences between informational and conversational questions, these studies tend to be directed toward the construction of future algorithms that might process submissions and offer probabilistic evaluation of which sort of interaction the asker seeks. In Answerbag, users were provided with the ability to tag their own submissions to indicate whether they sought information or conversation, but fewer than 10% of askers used the flag (Gazan, 2009). One explanation may be related to the difficulties researchers have had creating strict classifications for questions and answers. The nature of SQA is such that regardless of the intent of the asker, questions and answers may be changed or co-opted by subsequent answerers and commenters, and it may be the case that an asker who explicitly limits desired responses might in that moment open themselves to targeting by defiant users.

Harper et al. (2010) enlisted two coders to categorize 300 questions drawn from Yahoo! Answers, Answerbag, and Ask Metafilter, and the coders reached agreement on a question's primary category 94.3% of the time. Cases of disagreement tended to be compound questions seeking both facts and interpersonal support. To control for variation between the different SQA sites from which the questions originated, the authors built a regression model predicting the quantitative outcome by site and question type. They conducted a Tukey's range test (Kramer, 1956) across the least squared

means of the question type to identify any differences. Because some SQA questions are personalized or immediate (e.g., Who do you think will win tonight's game?), while others are factual and not time-dependent (e.g., When did California become a state?), the researchers asked student coders to rate questions on their perceived "archival value," and found that contrary to previous research by Harper, Moy, and Konstan (2009), informational and conversational questions, particularly those seeking advice, were both rated highly in terms of their perceived lasting value.

Answer Classification and Quality Evaluation

One of the core assumptions of SQA is that at some point, the aggregated opinions of a large enough number of unknown answerers will equal or surpass the quality of those available through more traditional channels. Several studies, including Harper, Raban, Rafaeli, and Konstan (2008), Shachaf (2009) and Shachaf and Rosenbaum (2009) have compared answer quality on SQA sites to that of library virtual reference services and found that SQA sites often equal or exceed their professional counterparts.

In the comparative study by Harper et al. (2008), answer quality was typically higher in the fee-based Google Answers service than in the free SQA sites they studied, and answer quality rose with the price paid. However, they found evidence to support the SQA model of aggregate expertise: answers provided by nonexpert Yahoo! Answers users were judged to be as good or better than those provided by experts or professionals on other sites. Similarly, Chen, Ho, and Kim (2010) investigated the performance of price-based Q&A systems in their study of Google Answers. In contrast to Harper et al., they found that offering a higher price for an answer led to a significantly longer, but not better, answer, and that answerers with a higher reputation provide significantly better answers. Jeon, Kim, and Chen (2010) reconciled these apparently conflicting findings by reanalyzing both data sets and suggested that differences in handling unanswered questions was responsible for the divergent findings. They conclude that although price is a factor in whether a question receives an answer, it does not have an effect on the quality of the answer.

In a study of Answerbag, Gazan (2006) found that answers from "synthesists," those who provide links and supporting evidence in their answers but claim no expertise of their own, tend to be rated more highly than answers from "specialists," who claim expertise and provide no supporting evidence. In a comparative study of five SQA sites (not including Answerbag), Harper et al. (2009) found similar evidence and suggest that answers with citations and supporting evidence such as links to other sites tend to receive higher ratings than those without. Importantly, none of the SQA sites studied carries with it a rule or policy that users should prefer one sort of answer to another. These results suggest a social aspect to information quality assessment, that participants in SQA sites develop emergent standards about what constitutes a good answer.

Building on the work of Kim, Oh, and Oh (2007), Kim and Oh (2009) studied comments left by askers to determine the criteria by which they identified relevant answers. Their analytical framework included comments related to:

- *Content*
- *Cognition* (i.e., novelty, understandability)
- *Utility*
- *Source authority*
- *Extrinsic factors* (such as answer speed and available alternatives)
- *Socioemotional criteria* (including gratitude, sympathy, and humor)

They found that the socioemotional class of asker comments were the most frequently observed in their study, and comments related to content and utility were also common, though the distribution of comments varied across different topical categories. Kim (2010) adopted a conversational/informational analytical structure and interviewed 36 Yahoo! Answers questioners to determine the criteria by which they evaluated answer credibility, and found support for previous research suggesting three interrelated areas of credibility judgment:

- *Criteria related to the credibility of information itself* (e.g., quality, accuracy, clarity, currency, spelling and grammar, tone of writing, bias, usefulness)
- *Criteria related to the credibility of a source or sponsor of the website* (e.g., author's contact information, authority, expertise, affiliation, reputation, presence of the organization's address, type of source)
- *Criteria related to the credibility of a website as a whole* (e.g., design and look, reference by reputable sources, navigability, functionality, advertising, customer service, and website host)

In Kim's 2010 study, roughly 64% of the subjects asked informational questions, while the remaining 36% sought a conversational interaction, though research by Harper, Weinberg, Logie, and Konstan (2010) suggests this proportion can be reversed depending on the individual SQA site. Liu et al. (2008) observe that 78% of the "best answers" in Yahoo! Answers were found to be reusable when applied to future similar questions, but that only 48% of those answers were uniquely best. Their analysis sought to find alternatives to the best answer model via automated summarization and combination of multiple answers. Their results suggest that open and opinion questions tend to have multiple best answers, and that combining summarized answers improves on the quality of any single "best" answer.

Shah and Pomerantz (2010) sought to identify more refined metrics of answer quality assessment in Yahoo! Answers and predict which answer a user would choose as best. They operationalized an answer as best for a given question if it was identified as such by the asker, and if the asker awarded it three or more points. They used human coders to assess answer quality along 13 criteria and found strong agreement between the coders' opinions and those of the askers. Various iterations of their model achieved roughly 80%–85% best answer prediction accuracy, and they also concluded that one of the most significant factors in

assessment of answer quality is the profile of the answerer as measured by the points accumulated on the site.

Tutos and Mollá (2010) compared the performance of six search engines, databases, and question-answering systems (Google, Google on PubMed, PubMed, OneLook, MedQA, and Answers.com Brain Boost) on answer quality across a small sample of medical questions. Although Answers.com, the only SQA competitor, performed the worst of the six on questions requiring medical intervention, such as when to introduce solid foods to infants, it performed surprisingly well (third of six) with nonintervention questions such as, Is watermelon allergenic? These results support those of Shachaf (2009) and suggest that because certain types of questions tend to draw particularly good answers in SQA environments, SQA sites can be considered worthy complements to professional resources in terms of answer quality.

User Satisfaction

Overlapping answer quality assessment are other measures of user satisfaction with SQA sites. Agichtein, Liu, and Bian (2009) attempted to predict information seeker satisfaction from Yahoo! Answers data and proposed a system based on consideration of an answerer's history. They also suggest the importance of personalized modes of relevance assessment and satisfaction instead of a one-size-fits-all metric.

The speed with which answers are received is an important component of the popularity of SQA sites and user satisfaction. Shah (2011) studied an aggregate sample of over 3 million questions and 16 million answers from Yahoo! Answers, and found that 30% of questions received an answer within 5 minutes of submission and 92% of questions received an answer within an hour. However, the best answer, whether selected by the user or the community, tended to be submitted later. Shah's results showed a spike in the number of best answers bestowed to answers submitted sixth, which may be an artifact of subsequent users building on answers submitted earlier, though the second most common submission position for the best answer was the one submitted first.

Members of SQA sites interpret, adapt, and enforce site rules to fit the norms of the community, which has implications for user satisfaction. For example, while there is no official policy against the posting of homework questions on the Answerbag SQA site, Gazan (2007a) found that members were able to distinguish between homework questions submitted by "seekers," operationalized in the study as those who interact with the community and engage in conversation about their questions, and "sloths," those who post their homework questions apparently verbatim from the assignment and interact no further. Experienced users answered homework questions of seekers much more often than those of sloths, and flagged them much less often for removal from the site. The results suggested that SQA site members expressed in their answers, comments, ratings, and actions

normative standards for behavior and appropriate site use, in this case doing one's own homework. Though this strategy could result in a nonresponsive or even intentionally wrong answer that risks being downrated by the asker, the results suggest that the rest of the community supported these normative responses to homework questions and expressed their satisfaction by selectively uprating those expressions.

Motivation, Reputation, and Perceived Authority

Any recommender system improves as it accumulates data from which to make relevance judgments, and SQA sites thus have an incentive to reward frequent and long-term participation. SQA sites commonly attempt to both vet existing contributions and motivate future contributions by awarding points. A user who comes across an answer they deem useful can uprate it, which adds points to both the answer rated and to the answerer's overall score. Web search engines generally do not reveal the reason certain sites are presented at the top of the results list, but SQA sites sort answers according to either the best answer as chosen by the asker or other users, or the answer with the most aggregate rating points. SQA sites offer enticements such as level titles, badges, public acknowledgment on leaderboards, increased rating power, and other site privileges as a reward for the accumulation of positive rating points, and the social reward structures of Yahoo! Answers have been identified as critical to the site's success (Shah, Oh, & Oh, 2008).

From the definition in Shah et al. (2009), SQA sites do not enlist professional or expert answerers, though several SQA sites have allowed users to build a reputation within a particular question category and become known as an expert on the site. For example, Yahoo! Answers awards Top Contributor status to users who make strong contributions in up to three question categories, and Answerbag employed an algorithm to assign users Expert status in the categories in which their contributions were highly rated by other users. However, evaluation functions native to any site can be misleading or intentionally gamed and should be viewed with caution. People may read or respond to an answer without rating it, and some sites award points for actions having nothing to do with answer quality. Knowledge-iN awards users points simply for logging into the site each day (Nam, Ackerman, & Adamic, 2009). Answerbag gives users points for helping moderate the site, for example, by moving miscategorized questions and reporting spam (Gazan, 2006).

Although SQA may lend itself to "drive-by" visitors who enter the site to seek a single piece of information then never return, several studies have found that it is the long term participants who create value in SQA sites, personify site norms, earn social capital, and have a strong influence on answer quality assessment. In their study of Knowledge-iN, Nam et al. (2009) found that the best answers are correlated with consistent participation, and that social/affective factors such as altruism and a reward structure based on points motivate participation. For both informational and conversational questions in the study by Kim (2010), askers reported that

source credibility was an important factor in their assessment, as identified by the answerer's previous activity, and how their previous contributions had been rated by the community. These findings support those of Jeon, Croft, Lee, and Park (2006), who attempted to extract automated answer quality assessments using nontextual features such as ratings and recommendations, and found that apart from the length of the answer, the most important feature associated with quality assessment was the expertise of the answerer, as reflected by their aggregate ratings. Gazan (2010) studied SQA users in one community who defied system constraints to engage in brief, informal "microcollaborations" and found that when high-ranking users expressed interest by contributing content to a collaborative SQA thread, the number of subsequent page views increased by as much as a factor of two.

Rafaeli, Raban, and Ravid (2007) found that Google Answers researchers were more likely to answer questions that had drawn more comments than others. They concluded that public goods (comments) enhanced market activity around the private information goods (answers) and call for more research about the social factors involved in information value assessment and participation.

Adamic, Zhang, Bakshy, and Ackerman (2008) conducted a large-scale network analysis of Yahoo! Answers, and harvested over 8.4 million answers to over 1.1 million questions by approximately 700,000 users. They introduced the concept of user entropy to describe the breadth of participation patterns users demonstrate—for example, those who only ask, only answer, or who only participate in limited topical domains. They found that the lower a user's entropy, the higher their answer ratings tended to be, but that this effect was limited to factual categories. The results of their analysis were used to predict which answers would be judged best, but the most predictive attributes were answer length, the number of competing answers, and the track record of the answerer. These findings confirm those of Smith (2002), who studied social accounting metrics on Usenet and found that a mutual awareness of participants' contributions and relationships is critical to a cooperative outcome, and Fiore et al. (2002), who found that revealing author histories correlates with trust and a user's desire to read more content posted by those contributors. However, although using answer quality metrics native to a SQA community generally constitutes good research practice, Adamic et al. (2008) point out that the Yahoo! Answers "best answer" function must be viewed with caution, because only one answer can earn the distinction, though more than one answer may be equally correct. Also, the standards by which users judge answers best are frequently idiosyncratic as well.

Nam et al. (2009) suggested that SQA site design features include a point system to motivate participation, a filter to allow users to view and respond to unanswered questions, and point bonuses for quick answers to "high-urgency" questions. Their study encompassed 2.6 million questions and 4.6 million answers submitted to Knowledge-iN, and they found that answerers' motivations were a combination of altruism, learning, participating in a hobby that yielded

a sense of competence, business motives, and questing for points.

In a content analysis of the posts of experienced Microsoft Live QnA users, Welser, Gleave, Barash, Smith, and Meckes (2009) found relatively few contributors providing technical answers, and a majority of content related to opinion and discussion. They suggest that the tools and affordances of SQA sites can shape the roles and interactions of its participants. Rodrigues, Milic-Frayling, and Fortuna (2008) compared community generated question tags in the Microsoft Live QnA site with use of Yahoo! Answers topic categories and found evidence that user-generated tags were related to higher levels of social interaction, and that active users may establish social ties around specific tags.

Another cross-site comparison was carried out by Shah et al. (2008), who used a crawler to gather information from 55,005 Yahoo! Answers users and 83,846 Google Answers users, which the authors report was the entire user population of the latter site at that time. They found that users of Google Answers, which operated on a payment model, tended to post one or two questions and then never return to the site. Conversely, Yahoo! Answers is free to join and rewards participation with points bestowed by other users, and it cultivated an environment that attracted much more active and consistent use. They analyzed the distribution of Yahoo! Answers users based on their experience level and found that users at higher levels tended to be contributors more than consumers, and that they created a disproportionate share of the content and value of the site. However, their efforts did not go unrewarded by the community: Shah et al. (2008) observe that while the highest ranking users did not earn as many plaudits for submitting interesting questions as those at lower levels, they did earn more Best Answer designations, and only the latter carries with it point rewards leading to higher levels. High-ranking users may in fact be the most skilled answerers, or they may be more economically rational actors, and employ conscious strategies to maximize their potential reward.

Bian, Liu, Agichtein, and Zha (2008) studied some of the ways in which unscrupulous users can game the rating system in Yahoo! Answers data, and they proposed several unsupervised methods for detecting and minimizing the effect of "vote spam." They found that semantic and social features, such as the similarity between the query and question and the history of the answerer, are more significant factors in detecting probable vote spam than the ratings of other users. Yang, Adamic, and Ackerman (2008) studied winning strategies in Taskcn, a site where users submit competing answers for the chance to receive a monetary reward set by the asker. Analyzing over 3,000 tasks and 185,000 potential answerers on Taskcn, they found that the most significant variable influencing success is the number of competitors. The results suggested that the most successful participants learned to choose less crowded tasks and learned to submit later in the process where the number and quality of competitors can be best assessed. This result agrees with subsequent work by Shah (2011), who found

that best answers on Yahoo! Answers tended to be submitted later.

Although SQA sites generally place few limits on participation and point-giving, there is an implicit limited good of attention and awareness. Users are well aware of the exposure granted best-rated or top-rated answers, and they engage in activities both within and outside the rules of SQA sites to maximize their position. Gazan (2007b) conceptualized “rogue users” as high-ranking members of SQA sites who violate the rules or spirit of the community. Observed rogue behaviors included vindictive downrating of content to advance one’s own position, frivolously flagging content as inappropriate for the site, creating separate “sock puppet” accounts to downrate or attack competitors, and using the friending and answer notification functions of the site to marshal groups of users to swarm perceived offenders or enemies. Although only 46 individuals of the 40,000 studied met the rogue user criteria, six ranked among the top 50 contributors in terms of the quantity and rating of their submissions. High-ranking users tended to view and navigate the site through their friends lists and notification panels, not through browsing or searching from the main page; thus, users who could bestow the most points tended to selectively view content submitted by their similarly high-ranking friends, skewing the points structure to reflect the number and power of one’s friends network more than the content of one’s answers.

A subsequent redesign of Answerbag, which among other changes altered the points structure to lessen the advantage of highly-ranked, highly friended users, led to a user revolt and migration away from the community by 266 of the 519 most active users in the 3-month period prior to the redesign (Gazan, 2011). Rogue users actively exploited technical breakdowns connected with the transition to the new site, and they used the remaining site functionality to exhort people in their friends network to disrupt and abandon the site. Among the reasons identified were the lack of access to past questions that served as the archive of the community (Hansen, Ackerman, Resnick, & Munson, 2007), and a schism between groups of experienced users about how the site changes reflected the perceived level of commitment of the site administrators to what each group felt were the core values of the community.

Conclusion and Future Research Directions

The research reviewed here suggests that SQA sites are more than the sum of their questions and answers, and that people participate for reasons well beyond a quest for facts. These and other studies constitute an impressive body of research, especially considering how recent a phenomenon SQA is, and their results lead to several promising directions for future work.

Refine Question Categorization

The early success of various taxonomic structures and categorization strategies for SQA could be tested against users’

actual categorization behavior, with more analysis of the kinds of borderline questions that have already been identified in the literature. Also, an asker’s previous contributions may inform the most appropriate categorization of a subsequent question. If the goal of categorizing questions is to maximize the likelihood of a question receiving an answer, then query logs and traffic data could be analyzed to identify areas of the taxonomy that are more and less visited, and questions categorized accordingly. For example, while a question about the sleeping patterns of dogs might be appropriately categorized in a subcategory of science, it may be more apt to be answered on an SQA site if placed in a subcategory of pets. This work could also be extended to identify and match questions with likely answerers.

Propose Taxonomies of Askers, Answerers, and Roles

Many answer prediction studies have attempted to determine the likelihood that a generic user of an SQA site would determine an answer best, often based on the behavior of past participants. However, there is no such thing as a generic asker, answerer, or rater. Future research could build on the work of Agichtein et al. (2009) and explore more personalized forms of relevance and satisfaction measures. This mode of research would likely require a more ethnographic approach, and might focus on deep observation and description of user behavior, with the goal of creating systematic categorization structures for the various roles and motivations of participants in SQA sites. Do certain individuals adopt roles of information hubs and authorities in SQA sites? Does a person’s frequency of participation or length of time on the site have an effect on their rating? Do contributors in certain topic areas tend to value content differently than those having other conversations on SQA sites?

Unpack the Track Record of the Asker and Answerer

Many studies reviewed here identify an answerer’s track record as one of the primary predictors of answer quality in SQA sites, and assume no external influences. However, the means by which high-ranking individuals achieve their status in particular sites warrants more detailed study. SQA sites are, at their core, social: future research must account for the influence of self-presentation, online identity, and network effects. For example, is there a relationship between the number of friends one has and the community’s rating of their contributions? Are there external factors, such as relationships and communications across other sites, that influence participants’ behavior?

Conduct SQA Research in a Broader Range of Settings

Shah et al. (2008) also identified a challenge common to most comparative studies of SQA sites: extracting comparable data samples. Although Yahoo! Answers has been the dominant SQA site in terms of traffic, and should be commended for having made much of its data available to

researchers via an easily accessible API, its combination of parameters for content submission, answer evaluation, ratings, comments, categorization, moderation policies, and user tools are unique and thus difficult to compare across sites. Many smaller SQA sites have been introduced since 2006, which present the opportunity for research into the evolution of SQA communities and to test findings from large-scale sites in different SQA arenas.

Apply Economic and Game-Theoretic Models to SQA

Participation in SQA sites can be modeled as a game, where the goal might be to attract friends and followers, or to maximize one's content views, answer ratings, or aggregate points. Applying economic models to define and identify behavior patterns that maximize or minimize one's position relative to different goals in the SQA environment would extend existing research threads, inform the design of any site reliant on user-generated content, and increase understanding of participant motivation and interactions.

Account for Cross-Site Content Sharing and Other Technological Externalities

It is becoming less and less accurate to think of SQA sites in a vacuum. Social networking sites such as Facebook allow the sharing of questions users find interesting on any SQA site, and it is possible to rate and interact with the same SQA content across multiple platforms. In late 2010, Twitter acquired the Fluther SQA site, auguring an imminent realm of SQA where content is optimized for mobile devices and interactions are limited to 140 characters or less. Synchronous social Q&A services such as Aardvark, where questions are asked and answered via instant messaging or similar real-time services, and routed only to those most likely to help, are a fast-growing application area. Richardson and White (2011) describe the "IM-an-Expert" system piloted in an online community within Microsoft, which strives to balance asker satisfaction with answerer overload, and they found that a key factor was managing asker expectations of the speed and likelihood that their question would be answered based on past transactions. The influence of SQA literature can be seen in the authors' prescriptions for user history and profile information, reputation scores, archives of past Q&A, and predictions of answer accuracy and user satisfaction. However, they caution that providing an overly optimistic estimate of system performance may create a tragedy of the commons situation (Hardin 1968), where overuse of the system leads to degradation of its performance for all. This and other research seems to indicate that there may exist a theoretically ideal equilibrium between askers and answerers in SQA sites that could be explored in future research.

Cross-site content sharing yields more opportunities for participation, yet also more scatter and complexity. Relevance was born as a concept to capture topical relevance, either as an absolute measure or from the situational perspective of the asker, but in a broader sense it could always

be equated with the question: What's worth my time to view? Some of the research reviewed here suggests that the topical accuracy of SQA information may be less important to users than its aggregate rating, its subtextual statement about shared values, or its source, if posted by a friend. Relevance as a concept in SQA encompasses the point of view of both asker and answerers, and future research should consider these and other situational, multidirectional elements of relevance more directly. For example, the physical location of an answerer might be a natural element of a relevance decision for people asking where to eat in a given city. Instantaneous geo-location via mobile devices, perhaps even with a record of places one has visited previously, can easily be imagined as one component of a user's public SQA profile. Ethical questions of how SQA information might be spammed, corrupted, or otherwise gamed must be explored as well.

Problematize SQA Itself

In sum, what should also be considered is essentially the null hypothesis, that SQA sites are not a particularly significant advance in information retrieval. Although virtual reference services offered by libraries remain a popular option for many information seekers, in the short time since the 2005 Yahoo! Answers introduction, sites based on competing SQA models such as ask-an-expert and pay-per-answer have diminished or become defunct, and SQA based on aggregate peer authority may suffer the same fate. SQA may grow and thrive, it may become an erstwhile, little-remembered early model of democratic participation and knowledge production, or it may exemplify a larger movement away from topical relevance and toward the relevance of immediacy and familiarity, allowing critical evaluation to be offloaded to "friends," known or unknown.

As SQA evolves, and its systems, sources, and users become more diverse, novel avenues for research open up as well. And just as with SQA sites themselves, the journey begins with a question.

References

- Adamic, L., Zhang, J., Bakshy, E., & Ackerman, M. (2008). Knowledge sharing and Yahoo! Answers: Everyone knows something. Proceedings of the 17th ACM Conference on the World Wide Web (ACM WWW '08) (pp. 665–674). New York: ACM Press.
- Agichtein, E., Liu, Y., & Bian, J. (2009). Modeling information-seeker satisfaction in community question answering. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 3(2).
- Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.
- Bernhard, D., & Gurevych, I. (2008). Answering learners' questions by retrieving question paraphrases from social Q&A sites. In Proceedings of the Third ACL Workshop on Innovative Use of NLP for Building Educational Applications, (pp. 44–52). Stroudsburg, PA: Association for Computational Linguistics.
- Bian, J., Liu, Y., Agichtein, E., & Zha, H. (2008). A few bad votes too many? Towards robust ranking in social media. In Proceedings of the ACM Conference on Adversarial Information Retrieval (AIRWeb'08) (pp. 53–60). New York: ACM Press.

- Cao, X., Cong, G., Cui, B., & Jensen, C.S. (2010). A generalized framework of exploring category information for question retrieval in community question answer archives. In Proceedings of the ACM Conference on the World Wide Web (ACM WWW'10) (pp. 201–210). New York: ACM Press.
- Cao, X., Cong, G., Cui, B., Jensen, C.S., & Zhang, C. (2009). The use of categorization information in language models for question retrieval. In Proceedings of the ACM Conference on Information and Knowledge Management (ACM CIKM'09) (pp. 265–274). New York: ACM Press.
- Chen, Y., Ho, T.-H., & Kim, Y.-M. (2010). Knowledge market design: A field experiment at Google Answers. *Journal of Public Economic Theory*, 12(4), 641–664.
- Fiore, A.T., LeeTiernan, S., & Smith, M.A. (2002). Observed behavior and perceived value in Usenet newsgroups: Bridging the gap. In Proceedings of the ACM Conference on Human Factors in Computing (ACM CHI '02) (pp. 323–330). New York: ACM Press.
- Gazan, R. (2006). Specialists and synthesists in a question answering community. In Proceedings of the 69th Annual Meeting of the American Society for Information Science & Technology (ASIST'06). Medford, NJ: Information Today.
- Gazan, R. (2007a). Seekers, sloths and social reference: Homework questions submitted to a question-answering community. *New Review of Hypermedia & Multimedia*, 13(2), 239–248.
- Gazan, R. (2007b). Understanding the rogue user. In Diane Nahl & Dania Bilal, (Eds.). *Information & emotion: The emergent affective paradigm in information behavior research and theory*. Medford, NJ: Information Today, 177–185.
- Gazan, R. (2009). When online communities become self-aware. Proceedings of the 42nd Hawaii International Conference on System Sciences (HICSS-42). Washington, DC: IEEE Computer Society Press.
- Gazan, R. (2010). Microcollaborations in a social Q&A community. *Information Processing & Management*, 46(6), 693–702.
- Gazan, R. (2011). Redesign as an act of violence: Disrupted interaction patterns and the fragmenting of a social Q&A community. In Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (ACM CHI '11). New York: ACM Press.
- Giddens, A. (1979). *Central problems in social theory: Action, structure and contradiction in social analysis*. Berkeley, CA: University of California Press.
- Giddens, A. (1991). Structuration theory: Past, present and future. In Bryant, C.G.A. & Jary, D. (Eds.) *Giddens' theory of structuration: A critical appreciation* (pp. 201–221). New York: Routledge.
- Goffman, E. (1959). *The presentation of self in everyday life*. London: Allen Lane.
- Graesser, A., McMahan, C., & Johnson, B. (1994). Question asking and answering. In Gernsbacher, M. (Ed.) *Handbook of psycholinguistics* (pp. 517–538). San Diego, CA: Academic Press.
- Hansen, D.L., Ackerman, M.S., Resnick, P.J., & Munson, S. (2007). Virtual community maintenance with a collaborative repository. In Proceedings of the 70th Annual Meeting of the American Society for Information Science and Technology (ASIST 2007). Medford, NJ: Information Today.
- Hardin, G. (1968). The tragedy of the commons. *Science*, 162(3859) (pp. 1243–1248).
- Harper, F.M., Moy, D., & Konstan, J.A. (2009). Facts or friends?: Distinguishing informational and conversational questions in social Q&A sites. In Proceedings of the ACM Conference on Human Factors in Computing (ACM CHI '09) (pp. 759–768). New York: ACM press.
- Harper, F.M., Raban, D., Rafaei, S., & Konstan, J.A. (2008). Predictors of answer quality in online Q&A sites. In Proceedings of the ACM Conference on Human Factors in Computing (ACM CHI '08) (pp. 865–874). New York: ACM Press.
- Harper, F.M., Weinberg, J., Logie, J., & Konstan, J.A. (2010). Question types in social Q&A sites. *First Monday*, 15(7). Retrieved from <http://firstmonday.org/article/view/2913/2571>
- Ignatova, K., Toprak, C., Bernhard, D., & Gurevych, I. (2009). Annotating question types in social Q&A sites. *Tagungsband des GSCL Symposiums 'Sprachtechnologie und eHumanities'*, 44–49. Essen, Germany: Abteilung für Informatik und Angewandte Kognitionswissenschaft Fakultät für Ingenieurwissenschaften Universität Duisburg-Essen.
- Jeon, J., Croft, B., & Lee, J.-H. (2005). Finding similar questions in large question and answer archives. In Proceedings of the ACM Conference on Information and Knowledge Management (ACM CIKM'05) (pp. 84–90). New York: ACM Press.
- Jeon, J., Croft, B., Lee, J.-H., & Park, S. (2006). A framework to predict the quality of answers with non-textual features. In Proceedings of the 29th Annual International ACM SIGIR Conference on Research & Development in Information Retrieval (ACM SIGIR '06) (pp. 228–235). New York: ACM Press.
- Jeon, G.Y., Kim, Y.-M., & Chen, Y. (2010). Re-examining price as a predictor of answer quality in an online Q&A site. In Proceedings of the ACM Conference on Human Factors in Computing (ACM CHI '10). New York: ACM Press.
- Kim, S. (2010). Questioners' credibility judgments of answers in a social question and answer site. *Information Research*, 15(1). Retrieved from <http://informationr.net/ir/15-2/paper432.html>
- Kim, S., & Oh, S. (2009). Users' relevance criteria for evaluating answers in a social Q&A site. *Journal of the American Society for Information Science and Technology*, 60(4), 716–727.
- Kim, S., Oh, J., & Oh, S. (2007). Best-answer selection criteria in a social Q&A site from the user-centered relevance perspective. Proceedings of the 70th Annual Meeting of the American Society for Information Science and Technology (ASIST'07). Medford, NJ: Information Today.
- Kling, R., McKim, J., & King, A. (2003). A bit more to IT: Scholarly communication forums as socio-technical interaction networks. *Journal of the American Society of Information Science and Technology*, 55(2), 127–148.
- Kling, R., Rosenbaum, H., & Sawyer, S. (2005). Understanding and communicating social informatics: A framework for studying and teaching the human contexts of information and communication technologies. Medford, NJ: Information Today.
- Kramer, C. (1956). Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*, 12(3), 307–310.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Lin, J. (2007). Is question answering better than information retrieval? Towards a task-based evaluation framework for question series. In Proceedings of Human Language Technologies: The Conference of the North American Chapter of the Association for Computational Linguistics (pp. 212–219). Stroudsburg, PA: Association for Computational Linguistics.
- Liu, Y., Li, S., Cao, Y., Lin, C.-Y., Han, D., & Yu, Y. (2008). Understanding and summarizing answers in community-based question answering services. In Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008), (pp. 497–504). Stroudsburg, PA: Association for Computational Linguistics.
- Lueg, C.P. (2008). Querying information systems or interacting with intermediaries? Towards understanding the informational capacity of online communities. In Proceedings of the 71st American Society for Information Science and Technology Annual Meeting (ASIST'08). Medford, NJ: Information Today.
- Lueg, C.P., & Fisher, D. (Eds.) (2003). *From Usenet to CoWebs: Interacting with social information spaces*. London: Springer.
- Nam, K.K., Ackerman, M.S., & Adamic, L.A. (2009). Questions in, Knowledge-iN?: A study of Naver's question answering community. In Proceedings of the ACM Conference on Human Factors in Computing (ACM CHI '09) (pp. 779–788). New York: ACM Press.
- Pomerantz, J. (2005). Linguistic analysis of question taxonomies. *Journal of the American Society for Information Science and Technology*, 56(7), 715–728.
- Raban, D.R. (2007). User-centered evaluation of information: A research challenge. *Internet Research*, 17(3): 306–322.
- Raban, D.R. (2009). Self-presentation and the value of information on Q&A sites. *Journal of the American Society for Information Science and Technology*, 60(12), 2465–2473.

- Raban, D.R., & Harper, F.M. (2008). Motivations for answering questions online. In Caspi, D., & Azran, T. (Eds.) *New Media and Innovative Technologies*. Be'er Shiva, Israel: Burda Research Center and Ben Gurion University.
- Rafaeli, S., Raban, D., & Ravid, G. (2007). Social and economic incentives in Google Answers. *International Journal of Knowledge and Learning*, 3(1), 1–11.
- Richardson, M., & White, R.W. (2011). Supporting synchronous social Q&A throughout the question lifecycle. In *Proceedings of the ACM World Wide Web Conference (ACM WWW2011)*, New York: ACM Press.
- Rodrigues, E.M., Milic-Frayling, N., & Fortuna, B. (2008). Social tagging behaviour in community-driven question answering. In *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WIAT '08)* (pp. 112–119). New York: ACM Press.
- Rosenbaum, H., & Shachaf, P. (2010). A structuration approach to online communities of practice: The case of Q&A communities. *Journal of the American Society for Information Science and Technology*, 61(9), 1933–1944.
- Shachaf, P. (2009). The paradox of expertise: Is the Wikipedia reference desk as good as your library? *Journal of Documentation*, 65(6), 977–996.
- Shachaf, P. (2010). Social reference: A unifying theory. *Library & Information Science Research*, 32(1), 66–76.
- Shachaf, P., & Rosenbaum, H. (2009, February). Online social reference: A research agenda through a STIN framework. Paper presented at the iConference 2009, Chapel Hill, NC.
- Shah, C. (2010). Collaborative information seeking: A literature review. In Woodsworth, A. (Ed.) *Advances in Librarianship*, 32 (pp. 3–33). Bradford, UK: Emerald Group Publishing Limited.
- Shah, C. (2011). Effectiveness and user satisfaction in Yahoo! Answers. *First Monday* 16(2). Retrieved from <http://firstmonday.org/article/view/3092/2769>
- Shah, C., Oh, J.S., & Oh, S. (2008). Exploring characteristics and effects of user participation in online social Q&A sites. *First Monday*, 13(9). Retrieved from <http://firstmonday.org/article/view/2182/2028>
- Shah, C., Oh, J.S., & Oh, S. (2009). Research agenda for social Q&A. *Library & Information Science Research*, 31(4), 205–209.
- Shah, C., & Pomerantz, J. (2010). Evaluating and predicting answer quality in community QA. In *Proceedings of ACM Conference in Information Retrieval (ACM SIGIR '10)*. New York: ACM Press.
- Smith, M. (2002). Tools for navigating large social cyberspaces. *Communications of the ACM*, 45(4), 51–55.
- Sudweeks, F., McLaughlin, M., & Rafaeli, S. (1996). *Network and Net-play: Virtual Groups on the Internet*. Cambridge, MA: AAAI/MIT Press.
- Turner, T., Smith, M.A., Fisher, D., & Welser, H.T. (2005). Picturing Usenet: Mapping computer-mediated collective action. *Journal of Computer-Mediated Communication*, 10(4). Retrieved from <http://jcmc.indiana.edu/vol10/issue4/turner.html>
- Tutos, A., & Mollá, D. (2010). A study on the use of search engines for answering clinical questions. In Maeder, A. & Hansen, D. (Eds.) *Proceedings of the Fourth Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2010), Conferences in Research and Practice in Information Technology (CRPIT)*, Vol. 108, (pp. 61–68), Sydney, Australia: Australian Computer Society.
- Viegas, F.B., & Smith, M. (2004). Newsgroup crowds and AuthorLines: Visualizing the activity of individuals in conversational cyberspaces. In *Proceedings of the 37th Hawaii International Conference on System Sciences (HICSS-37)*. Washington, DC: IEEE Computer Society Press.
- Voorhees, E.M. (2003). Overview of the TREC 2003 question answering track. In *Proceedings of the 12th Text Retrieval Conference (TREC-2003)* (pp. 54–68). Washington, DC: NIST.
- Welser, H.T., Gleave, E., Barash, V., Smith, M., & Meckes, J. (2009). Whither the experts? Social affordances and the cultivation of experts in community Q&A systems. In *Proceedings of the IEEE International Conference on Computational Science and Engineering* (pp. 450–455). Washington, DC: IEEE Computer Society.
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, UK: Cambridge University Press.
- Yang, J., Adamic, L.A., & Ackerman, M.S. (2008). Crowdsourcing and knowledge sharing: Strategic user behavior on Taskcn. In *Proceedings of the ACM Conference on Electronic Commerce (ACM EC'05)* (pp. 246–255). New York: ACM Press.
- Zhang, J., Ackerman, M., & Adamic, L. (2007). Expertise networks in online communities: Structure and algorithms. In *Proceedings of the ACM World Wide Web Conference (ACM WWW'07)* (pp. 221–230). New York: ACM Press.