

LEARNING IN CROWDSOURCED ENVIRONMENTS: WHERE ARE WE GOING AND HOW DO WE GET THERE?

Manasa Rath* & Chirag Shah

School of Communication and Information, Rutgers University, 4 Huntington Street, New Brunswick, New Jersey 08901, USA

* Address all correspondence to: Manasa Rath, School of Communication and Information, Rutgers University, 4 Huntington St., New Brunswick, NJ 08901, USA, E-mail: manasa.rath@rutgers.edu

Abstract

Internet users have many tools and options to fulfill their online information needs, for purposes such as routine decision-making and online learning. One such option is “learning from the crowd,” in a crowdsourced environment through the user-generated content found on various platforms, such as online question-answering sites. Studies have shown that these sites are used extensively. Amid growing usage of crowdsourced knowledge bases, research on crowdsourced learning has exposed multiple problems and hazards for users, such as content quality and reliability. In this article, the authors describe both the advantages and disadvantages of crowdsourced learning, and describe their current research to better understand and develop tools that assist users to evaluate the quality of user-generated content as a resource for learning.

KEY WORDS: crowdsourcing, online learning, question-answering sites

1. INTRODUCTION

Current Internet technologies permit individuals not only to communicate with one another but also to learn in the process. Lachman (1997) argued that learning can be understood as a change in an individual’s behavior due to experience. This definition helps us understand the dynamic of individuals using technology to learn. Another useful resource for understanding online learning is the body of work by scholars in the fields of cognitive and educational psychology that focuses on the role of communities in meaning making (Vygotsky, 1978). These perspectives are helpful as we observe and study patterns of behavior in an era of ubiquitous learning opportunities. Modern tools such as learning

management systems and initiatives such as the MIT Open Course Ware have revolutionized content access and delivery to users, and ushered in new forms of pedagogy. Moreover, these and other learning systems have revolutionized learning in both traditional and online environments. These advances paved the way for new modes of online learning, which has grown in recent years from formal (i.e., classroom) to informal (i.e., social media or self-regulated) conditions, and it has more recently been infused into the idea of self-regulated personal learning environments (Dabbagh and Kitsantas, 2012).

So far, crowdsourcing models have been developed and employed predominantly in educational settings (Skaržauskaitė, 2012). One common example of crowdsourcing is community participation in question-answering (Q&A) sites, where a member of the community asks a question that is then answered by the other members of that community. This participatory information exchange behavior is often associated with learning, and points to a possible future of crowdsourced scholarship (Gerber et al., 2001).

Scholars studying crowdsourcing in educational settings have investigated both the pros and cons of those environments. This paper acknowledges and summarizes those issues and examines focuses on ethical issues such as reliability, authenticity, and the quality of content. We conclude by proposing practical solutions through which crowdsourcing in educational contexts can scale to new heights.

2. RELEVANT PREVIOUS WORK AND THEORETICAL FRAMEWORKS

Scholars studying crowdsourced environments often have used socio-constructive, meta-theoretical approaches that assume human learning and development are socially situated and knowledge is created by interacting with others. Crowdsourced venues such as social question-answering sites are hubs for knowledge sharing (Shah et al., 2008), and often are viewed by scholars through the lens known as Social Exchange Theory (Emerson, 1976), where learning occurs by observing fellow members of the site in question. Understanding learning from this micro-perspective sees it as a form of guided behavior, i.e., learning with the help of social units (Bandura, 1986). Another framework for understanding learning is “situated learning.” This approach views learning as embedded within its other social surroundings such as activity, context, and culture (Anderson et al., 1996). For instance, a situated learning analysis of learning with web-enabled tools in physical classrooms might differ from a similar analysis of classroom-based teaching that does not make use of online technologies. Crowdsourcing in education also has been studied with approaches that focus on knowledge creation, such as the Socialization Externalization Combination Internalization (SECI) model that also is grounded on socio-constructive foundations. The SECI model has allowed previous innovative work studying

learning communities that are distributed in space and time in different ways (Chatti et al., 2007).

These and other theoretical approaches have yielded many useful studies of online learning, including analysis of ways to enhance and augment computational thinking among K-12 students (Cavanaugh et al., 2009), and how to produce better learning experiences in science education (Corneli and Mikroyannidis, 2012; Meyers et al., 2013; Hall, 2009; Porcello and Hsi, 2013). Significant work has been done in recent years on ways to usefully customize crowdsourced learning resources (Dontcheva et al., 2014; Hew and Brush, 2007; Picciano and Seaman, 2009) and producing high-quality instructional materials based on individual learning capabilities (Vovides et al., 2007). Additionally, important work has been done in the areas of understanding user motivations for participation in crowdsourced learning, and more generally related to personal learning environments (Attwell et al., 2007; Dabbagh and Kitsantas, 2012).

These are just some of the social learning theories and previous work to understand how individuals seek, learn, innovate, and make decisions in their everyday lives. All of the above-mentioned frameworks and theories can provide useful insights not just about formal learning, but also more generally about other activities involving fulfilling information needs in a participatory setting. Building upon this previous work, we examined crowdsourced question-answering sites as one form of social learning environments.

3. POTENTIAL BENEFITS OF CROWDSOURCED ONLINE LEARNING

3.1 Learning and Collaboration

Participatory sites such as question-answering platforms provide opportunities for learning and collaboration. Gazan (2010) found that Q&A sites have become popular venues for micro-collaborations around learning goals, through “brief, informal expressions of mutual interest and mutual effort toward seeking information on a given topic” (p. 693). That particular study demonstrated that instances of collaboration within Q&A sites were more evident when users were intended on helping fellow users or collaboratively fulfilling each other’s information needs (Shah et al., 2008). Q&A sites also have gained popularity due to the ever-increasing demand of free and easily accessible content.

3.2 Harnessing Collective Intelligence

Participatory sites are often considered as powerful venues of crowdsourcing or crowdfunded ideas (Brabham, 2008). For example, when we ask a question to the crowd and request an answer, we obtain a wide range of opinions about that question, which provides us with a wide perspective on that particular topic while also providing an external validation on the topic. Collective intelligence has been used on several occasions to solve

business problems (Alag, 2009). Users often turn to Q&A sites to obtain opinions and perspectives about particular tasks. For instance, a user might ask about resolving a societal problem (e.g., street parking issues) and receive helpful opinions and advice from fellow users. In some cases, as users help resolve significant problems, they are often being rewarded in different forms by the participatory site through items such as points or reputational rankings.

3.3 Rewarding Knowledge Sharing

Users who provide answers to questions on Q&A sites often can be upvoted or liked by fellow users, and often are thus rewarded in ways that encourage them toward further participation. Their contributions to Q&A sites are considered a part of knowledge sharing (i.e., participating with other users on the site to collaboratively provide answers) while they earn rewards. Scholars have studied this phenomenon of knowledge sharing on participatory sites and found that such interactions and rewards lead to further information sharing along with social networking among users (Vasilescu et al., 2014). Taking part in these practices, such as expressing information need, collaboratively working on fulfilling information needs, and in some cases, getting rewarded while fulfilling information needs, are just some of the potential benefits of knowledge sharing in emerging Web platforms.

3.4 Defining a Scholarly Identity on Social Spaces

Recently, some scholars have begun to disseminate their research and other works with fellow academics via social networking sites such as ResearchGate or Academia. This type of activity adds new potential dimensions to traditional forms of scholarly communication, including fruitful online collaborations and greater realization of the possibilities of online learning environments, as explained by Kaplan and Haenlein (2010). According to Van Noorden (2014), some scholars also now engage with their global research communities on social question-answering sites and social media such as Twitter, to disseminate their research to wider audiences.

4. POTENTIAL HAZARDS OF CROWDSOURCED ONLINE LEARNING

4.1 Content Quality

Online participatory sites essentially provide users with a platform to create and consume content. Because so many of these sites are essentially open to a global community of amateur participants and do not employ any effective screening for expertise except for socially generated feedback or rankings from other users, the quality of user-generated content is a major concern for educators and researchers. The content found on participatory sites often is indexed and optimized for easy Internet discovery and use in everyday decision-making. At a time of high public awareness of quality problems (e.g.,

fake news) and the complexity present in user-generated content (e.g., sarcastic content), we as a society still have not developed effective and generalizable systems for rating the quality and trustworthiness of user-generated content. This uncertainty for users significantly compromises the potential benefits of crowdsourced learning. Vint Cerf, a well-known computer scientist and acknowledged as one of the creators of the Internet, described assessing the quality of content on the Web as one of the herculean tasks of the 21st century (Cerf, 2018).

Moreover, unreliable user-generated content is caused not only by lack of user expertise. One of the other major concerns of crowdsourced sites is that they can contain intentionally false and malicious content (Krumm et al., 2008; Cerf, 2018). We inhabit an online environment where we are surrounded by dubious and half-true content. It is difficult to imagine scalable systems and methods for peer review and identification of unreliable and potentially harmful user-generated content. Other hazardous information such as defamatory content and hate speech is difficult to eliminate once posted, and can spread easily. Adding to the misery, studies have found that posts containing malicious content travels faster when compared to regular content. Again, we have no mechanisms or strict rules to cease the creation of such content (George and Scerri, 2007). To compound this issue, another major problem for users of participatory sites relates to ownership of the content they produce, which brings us to the following section on intellectual property.

4.2 Intellectual Property

Online participatory sites present numerous questions and issues related to intellectual property and ownership of content, ranging from brief answers and explanations, up to extensive creative works. Who owns the content presented on these types of Q&A sites? Who holds the copyright of the content created or shared on these sites? Claims of copyright infringement are common on many sites that allow users to create and share content. The participatory site YouTube has a policy in its terms and conditions stating that whoever creates the content on their platform holds its copyright and is its original creator of it (George and Scerri, 2007). However, some Internet users who are not familiar with copyright violation rules often get away with breaking the rules, thereby violating another's ownership by declaring it as their own (George and Scerri, 2007).

Such incidents fuel calls for stricter laws to curb copyright violations. Presently, however, misappropriation, stealing, and infringement of others' ownerships rights are too common. For instance, based on a study conducted by Turnitin, a plagiarism detection software company, it was found that an astonishing number of students reproduced the same content from different online venues without referencing the original source (Mergel, 2013).

Such incidents highlight the need for greater awareness among Internet users regarding online literacy and how to properly use online content.

4.3 Privacy

While posting content to online participatory sites, our identity is often revealed to other members of that site, which leads to privacy concerns. To understand the vulnerability of individuals online, one needs to look no further than the recent controversy and revelations about Facebook privacy violations in 2016, when data firm Cambridge Analytica (hired by the campaign of candidate Donald Trump) crawled for Facebook users' information in order to influence user behavior and political orientations (Persily, 2017). Such incidents demonstrate a need for stricter laws that could prevent some privacy invasions. However, a study conducted by Carnegie Mellon University revealed that although many users are aware of the potential harm in making personal information public, many are nonetheless comfortable providing personal and private information anyway (Govani and Pashley, 2005). The ongoing tensions and trade-offs between privacy and customized or targeted online services will not be resolved any time soon.

5. FURTHER COMPLICATIONS

Considering the potential benefits and hazards of using crowdsourced learning sites, we suggest that multiple problems must be addressed when building online learning environments. Studies have found users often rely on search engines over traditional knowledge keepers, such as librarians or teachers, in order to fulfill any information needs they may have. Moreover, many users prefer content found on participatory sites. Shah et al. (2008) demonstrated that large numbers of students utilize user-generated content on Q&A sites to complete their assignments. In a recent study conducted by the Stanford History Education Group (McGrew et al., 2018) found that university undergraduate students doing research online could be misled by manipulative features on some sites such as search result ranking, official-looking logos, and domain names intended to be similar to more reputable and trustworthy sites. This study shows a need for digital literacy programs to be embedded within the school curriculum.

6. EVALUATING THE QUALITY OF USER-GENERATED CONTENT

The issues mentioned above warrant a thorough rethinking of how to evaluate user-generated content. As part of an Institute of Museum and Library Services (IMLS) grant-funded research project ("Online Q&A in STEM Education: Curating the Wisdom of the Crowd") we have developed a model for evaluating the quality of user-generated information, based both on content as well as contextual factors. This model was developed by extensively reviewing the literature on the different information quality

frameworks, which was based on the user-generated content presented on sites such as Twitter and AnswerBag. Our work and previous research has yielded the following general evaluation criteria:

- Clear: content stated directly and is not confusing
- Complete: content is not missing any necessary or relevant information
- Correct: content is free from error and can be regarded as true
- Credible: content is provided by a genuine source and can be regarded as honest

By applying these criteria consistently, high-quality content can be identified more easily. High-quality content on a Q&A site attracts more users and enriches the knowledge-base of the site, which in turn improves the efficiency of interactions among the users at a macro level. On a micro level, high-quality content posted by any particular user cultivates a “micro collaboration” among users. The shutdown of Google Answers in 2011 for being unable to preserve the quality of content is a reminder of the continued study required in regards to Q&A sites (Rath et al., 2017). Furthermore, assessing the quality of the content on Q&A sites is important for preserving the sustainability of digital resources.

In order to advance our research, we plan to build a Web-based browser plugin that will automatically evaluate the quality of user-generated content. The envisioned tool will provide users with assessments of information quality using both the actual content and contextual, and will be able to incorporate objective evaluations provided by trained and trusted evaluators such as librarians and school media specialists. These quality evaluations will consider the criteria listed above: Correct, Clear, Credible, and Complete. Our tool will provide easily understandable visual cues about content users are viewing. Such assistance advances the possibilities for information fostering, (i.e., proactively providing suggestions to users while they fulfill their information-seeking tasks) as discussed by Shah (2018). The tool will help users understand both the relative quality of the content along with the social context within which the content was produced.

Our continuing work will not only develop the theoretical framework, but also associated metrics and tools to automatically evaluate the quality of user-generated content. The potential impact of our research is twofold, ranging from building rigorous evaluation infrastructure such as virtual tools to not just inform users about the quality of content, but also to better understand in the aggregate user sense-making behaviors when presented with indicators of content quality. Additionally, our work will aid us in developing the tools needed to help students attain their educational outcomes with the help of librarians and practitioners.

7. CONCLUSION

Crowdsourced learning through formal and informal settings has proven to be successful, although it has exposed certain issues and problems in recent years. In order to realize the true potential of learning in crowdsourced environments and minimize risk for users, we need better indicators of the quality of user-generated information. This effort will require scalable metrics, tools, workflows, and buy-in from stakeholders such as educators and users, and the operators of online crowd-sourced sites. Learning in crowdsourced environments is often enhanced with the help of fellow community members, requiring vigilant participation for the smooth functioning of such sites. This paper has reviewed some of the major potential benefits and hazards facing crowdsourced learning environments, and provided a summary of the work we are doing to help Q&A sites reach their fullest potential in support of learning.

REFERENCES

- Alag, S., *Collective Intelligence in Action*, Greenwich, CT: Manning, pp. 274–306, 2009.
- Anderson, J.R., Reder, L.M., and Simon, H.A., Situated Learning and Education, *Educ. Res.*, vol. **25**, no. 4, pp. 5–11, 1996.
- Attwell, G., Personal Learning Environments-The Future of eLearning?, *Elearning Papers*, vol. **2**, no. 1, pp. 1–8, 2007.
- Bandura, A., *Social Foundations of Thought and Action*, Englewood Cliffs, NJ: Prentice-Hall, 1986.
- Brabham, D.C., Crowdsourcing as a Model for Problem Solving: An Introduction and Cases, *Convergence*, vol. **14**, no. 1, pp. 75–90, 2008.
- Cavanaugh, C.S., Barbour, M.K., and Clark, T., Research and Practice in K-12 Online Learning: A Review of Open Access Literature, *Int. Rev. Res. Open Distributed Learning*, vol. **10**, no. 1, 2009. DOI: 10.19173/irrodl.v10i1.607
- Cerf, V.G., Unintended Consequences, *Commun. ACM*, vol. **61**, no. 3, p. 7, 2018.
- Chatti, M.A., Klamma, R., Jarke, M., and Naeve, A., The Web 2.0 Driven SECI Model Based Learning Process, in *Advanced Learning Technologies, 2007, ICALT 2007. Seventh IEEE Intl. Conf.*, Niigata, Japan, July 18–20, 2007, pp. 780–782, 2007.
- Corneli, J. and Mikroyannidis, A., Crowdsourcing Education on the Web: A Role-Based Analysis of Online Learning Communities, in *Collaborative learning 2.0: Open Educational Resources*, Hershey, PA: IGI Global, pp. 272–286, 2012.
- Dabbagh, N. and Kitsantas, A., Personal Learning Environments, Social Media, and Self-Regulated Learning: A Natural Formula for Connecting Formal and Informal Learning, *Internet Higher Ed.*, vol. **15**, no. 1, pp. 3–8, 2012.

Dontcheva, M., Morris, R.R., Brandt, J.R., and Gerber, E.M., Combining Crowdsourcing and Learning to Improve Engagement and Performance, in *Proc. SIGCHI Conf. Human Factors in Computing System*, Toronto, ON, Canada, April 26–May 1, 2014, pp. 3379–3388, 2014.

Emerson, R.M., Social Exchange Theory, *Ann. Rev. Sociol.*, vol. **2**, no. 1, pp. 335–362, 1976.

Gazan, R., Microcollaborations in a Social Q&A Community, *Inf. Process. Manage.*, vol. **46**, no. 6, pp. 693–702, 2010.

George, C.E. and Scerri, J., Web 2.0 and User-Generated Content: Legal Challenges in the New Frontier, *J. Inf., Law Technol.*, vol. **2007**, no. 2, 2007. Available from http://go.warwick.ac.uk/jilt/2007_2/george_scerri/.

Gerber, B.L., Cavallo, A.M., and Marek, E.A., Relationships among Informal Learning Environments, Teaching Procedures and Scientific Reasoning Ability, *Int. J. Sci. Educ.*, vol. **23**, no. 5, pp. 535–549, 2001.

Govani, T. and Pashley, H., Student Awareness of the Privacy Implications when Using Facebook, presented at the “Privacy Poster Fair” at the Carnegie Mellon University School of Library and Information Science, vol. **9**, pp. 1–17, 2005.

Hall, R., Towards a Fusion of Formal and Informal Learning Environments: The Impact of the Read/Write Web, *Electron. J. E-learning*, vol. **7**, no. 1, pp. 29–40, 2009.

Hew, K.F. and Brush, T., Integrating Technology into K-12 Teaching and Learning: Current Knowledge Gaps and Recommendations for Future Research, *Educ. Technol. Res. Dev.*, vol. **55**, no. 3, pp. 223–252, 2007.

Kaplan, A.M. and Haenlein, M., Users of the World, Unite! The Challenges and Opportunities of Social Media, *Bus. Horiz.*, vol. **53**, no. 1, pp. 59–68, 2010.

Krumm, J., Davies, N., and Narayanaswami, C., User-Generated Content, *IEEE Pervasive Comput.*, vol. **7**, no. 4, pp. 10–11, 2008.

Lachman, S.J., Learning is a Process: Toward an Improved Definition of Learning, *J. Psychol.*, vol. **131**, no. 5, pp. 477–480, 1997.

McGrew, S., Breakstone, J., Ortega, T., Smith, M., and Wineburg, S., Can Students Evaluate Online Sources? Learning from Assessments of Civic Online Reasoning, *Theory Res. Social Educ.*, vol. **46**, no. 2, pp. 165–193, 2018.

Mergel, I., Social Media Adoption and Resulting Tactics in the US Federal Government, *Gov. Inf. Q.*, vol. **30**, no. 2, pp. 123–130, 2013.

- Meyers, E.M., Erickson, I., and Small, R.V., Digital Literacy and Informal Learning Environments: An Introduction, *Learning, Media Technol.*, vol. **38**, no. 4, pp. 355–367, 2013.
- Persily, N., The 2016 US Election: Can Democracy Survive the Internet?, *J. Democracy*, vol. **28**, no. 2, pp. 63–76, 2017.
- Picciano, A.G. and Seaman, J., K-12 Online Learning: A 2008 Follow-Up of the Survey of US School District Administrators, Sloan Consortium, Newburyport, MA, 2009.
- Porcello, D. and Hsi, S., Crowdsourcing and Curating Online Education Resources, *Science*, vol. **341**, no. 6143, pp. 240–241, 2013.
- Rath, M., Shah, C., and Floegel, D., Identifying the Reasons Contributing to Question Deletion in Educational Q&A, *Proc. Assoc. Inf. Sci. Technol.*, vol. **54**, no. 1, pp. 327–336, 2017.
- Shah, C., Information Fostering-Being Proactive with Information Seeking and Retrieval: Perspective Paper, in *Proc. 2018 Conf. on Human Information Interaction & Retrieval*, New Brunswick, NJ, USA, March 11–15, 2018, ACM, pp. 62–71, 2018.
- Shah, C., Oh, J.S., and Oh, S., Exploring Characteristics and Effects of User Participation in Online Social Q&A Sites, *First Monday*, vol. **13**, no. 9, 2008. DOI: 10.5210/fm.v13i9.2182
- Skaržauskaitė, M., The Application of Crowd Sourcing in Educational Activities, *Soc. Technol.*, vol. **2**, no. 1, pp. 67–76, 2012.
- Van Noorden, R., Online Collaboration: Scientists and the Social Network, *Nat. News*, vol. **512**, no. 7513, pp. 126–129, 2014.
- Vasilescu, B., Serebrenik, A., Devanbu, P., and Filkov, V., How Social Q&A Sites are Changing Knowledge Sharing in Open Source Software Communities, in *Proc. 17th ACM Conf. on Computer Supported Cooperative Work & Social Computing*, Baltimore, Maryland, USA, February 15–19, 2014, pp. 342–354, 2014.
- Vovides, Y., Sanchez-Alonso, S., Mitropoulou, V., and Nickmans, G., The Use of e-Learning Course Management Systems to Support Learning Strategies and to Improve Self-Regulated Learning, *Educ. Res. Rev.*, vol. **2**, no. 1, pp. 64–74, 2007.
- Vygotsky, L., Interaction between Learning and Development, *Read. Dev. Children*, vol. **23**, no. 3, pp. 34–41, 1978.