Using Peer-Production to Foster Bias Awareness among Online Content Consumers

Andrew W. Vargo
Keio University
Yokohama, Japan
vargo@kmd.keio.ac.jp

Benjamin Tag
The University of Melbourne
Melbourne, Australia
benjamin.tag@unimelb.edu.au

ABSTRACT

Peer-production systems have been used to facilitate the creation of high-quality data products such as encyclopedias, repositories of technical information and software code, and creative efforts like cookbooks and literature. A proposed extension of the peer-production model is the creation and provision of bias analysis. An ideal application could provide users with a well-defined bias score of content or consumption. However, peer-production systems themselves can be vulnerable to bias and manipulation. Therefore, in this paper, we discuss recommendations for constructing a working peer-production model and provide an overview of the challenges and realities. In order to ground the discussion, we present the "SANCTUARY" framework, a distributed peer-production system for fostering bias awareness among online content consumers.

CCS CONCEPTS

• **Human-centered computing → Computer supported cooperative work**; *Collaborative content creation.*

KEYWORDS

Peer-Productions Systems, Bias Detection, CSCW

INTRODUCTION

The effectiveness of micro-targeting and the success of falsified information distributed through social media channels have led to increasing concerns about consumption of biased information. Algorithms today accurately learn from user behaviour and supply a virtually endless stream of additional content of interest to the user in order to increase engagment. An unintended side-effect is that these algorithms ultimately reinforce users' biases. Advances in artificial intelligence have led to a steady increase of learning algorithms that personalize information services, and ultimately foster the creation of social bubbles [18].

These technologies that enable targeted distribution of information are under scrutiny. To detect biases and raise awareness of biased information among users, different approaches have been deployed. These approaches often work on a **direct level** and/or on a **system level**. On a direct level, tests like Project Implicit[1] at Harvard University, aim at accurately presenting individuals their unconscious biases. On a system level, bias detection using Artifical Intelligence (AI) or machine learning (ML) relies on classifiers, e.g., Gao's system [12].

In all of these cases, a problem is the reliance on authority. This is especially important when considering that the people and organizations deciding what constitutes bias is vague [19]. Furthermore, the methodology for achieving the detection can be obtuse, and can be limited to specific areas and instances. For example, journalists could be cherry-picking edge cases which fail to convey the total truth and data scientists could be importing their own biases into their algorithms either unknowingly, or because they are using a data set with the faith that it is good. The data sets used in many studies [2, 9] either use or create data sets that rely on qualitative decisions. Since algorithms and in-depth online psychological profiling rely on mechanisms that work on levels beyond the understanding of laypersons, and/or may not work as popularly understood, as in the case of the Project Implicit [8, 13].

To solve these problems of trustworthiness of data sets, authority, and user agency, we propose the use of a peer-production based bias detection system, "SANCTUARY". Working in cooperative groups has been shown to result in well compiled accurate corpora of information and in short periods of time [3, 7]. Having a community come to a consensus on a data product based on qualitative determinations could lend credence to the corresponding data product. This also allows for a possible human relationship and affinity between the resulting product and the users, as opposed to the reliance on authority as mentioned above. In this position paper, we will describe the background of

[1] https://implicit.harvard.edu/implicit/

peer-production systems and the foundations necessary to develop a community, detail the application scenario, and conclude with a discussion of benefits and pitfalls of such peer-production systems.

BACKGROUND AND FOUNDATIONS

To foster the development of an effective peer-production system for bias-detection, we conducted an extensive investigation of peer-production systems. In the following we detail a summary of strengths, weaknesses, and crucial mechanics that inform the creation of SANCTUARY. Peer-production systems have a broad range of uses. Coined as ”Commons-based Peer Production” by Yochai Benkler [7], peer-production systems range from broad collaboration efforts to highly organized communities. In this section, we first clearly distinguish peer-production from crowdsourcing, before examining the spectrum of control and social ties that are present in peer-production systems.

Peer-Production and Crowdsourcing

A crucial point for this proposal is the utilization of ”Peer-Production” rather than that of ”Crowdsourcing”. While peer-production systems can be considered a subset of crowdsourcing, crowdsourcing carries a connotation of directed work that is non-community based [6]. That is, a requester asks a crowd or individuals from a crowd to fulfill an exact task. There is no community organization around this request, and freedom of choice in directing and accomplishing the task is severely limited. For example, a standard crowdsourcing task is to ask workers to label content for machine learning. The requester will provide the data set, the choice or range of labels, the time frame for task fulfillment, and the payment per label. There is no leverage for the crowd-member with the exception of accepting or rejecting the task. Just like this example, even though unpaid tasks do exist, many crowdsourcing tasks are motivated by monetary incentives.

In contrast to crowdsourcing, peer-production systems function on a community level, rather than depending on directed work. This means that motivation for which task is being complete is not driven by monetary incentives, but rather by a community that decides what to do and why. The production process is, therefore, under control of the participating users. This means that there is a level of freedom in how the community and individuals in the community approach and direct their own work. For example, a community iteratively developing its own taxonomy for tagging content. In summary, a fundamental aspect of peer-production is that there is no direct link between the output and monetization [5].

Models of Peer-Production Systems

Peer-production systems have many variations. In order to explain the general differences in peer-production systems, we compare two peer-production communities centered around computer programming.

Stack Exchange[2] and Quora[3] share a goal in that their target content is the asking of questions and their corresponding answers. However, the presentation and mechanics of the sites are noticeably different. This is because each community has a different drive towards achieving their goal. While any one question could be answered on any of the three sites, how it is answered is likely a function of the community.

Stack Exchange. Users compete in Question & Answer (QA) exchanges that are dictated by a reputation system. They vote on content and receive reputation points for contributions that are considered to be high quality. They can choose their own identity, resulting in a wide range of profiles, reaching from pseudo-anonymous to real-world anonymous [24]. Users who gain reputation points are awarded privileges that allow them to participate in collaborative moderation of the site, e.g., the right to vote to delete unfit content [10]. The overall goal of the site is to create a highly accurate and non-redundant corpora of information for each domain in the site network. Thus, content is often redacted or merged by the community, and content organization is strict [23]. The moderation effort is aided by moderators who are elected by popular vote in the community[4], and as of 2019, bots that can remove low-quality submissions and identify cheating.[5]

Quora. This site is a generalist QA community with a computer-programming subsection. Users do not compete for reputation points, but rather for exposure. Users typically expose their real-world identities which are linked to their areas of expertise. Users are essentially competing for views of their content and profiles. Quora has more lax rules for redundancy and content curation [26], and instead focuses on removing spam. As such, users can make requests to the moderators and bots to remove content.[6]

Both communities have differences in their scope and approach to QA as made clear by their differences in mechanics and site design. These differences both are a result of and describe an influence on the community and its users and their ties to the task.

The User Base

A key question for any peer-production system is how to grow and maintain a user base. Without a dedicated user base, it is impossible to have a stable site.

Achieving Critical Mass. Critical mass is the concept of having a number of users which allow a system to find equilibrium. Two common patterns for communities are for the number of users to accelerate before plateauing or to decelerate until extinction [20]. Achieving critical mass in a peer-production system is difficult as it is unclear what particular thing needs to hit critical mass [20]. However, it seems that building a large user base at the beginning of a community is more important to the success of a community than building up material, and contributions often lag behind membership [20]. This

[2] https://stackexchange.com/

[3] https://quora.com/

[4] https://stackoverflow.com/help/site-moderators

[5] https://stackoverflow.blog/2019/09/17/meet-the-bots-that-help-moderate-stack-overflow/

[6] https://qr.ae/T3ko5Y

is important because it may go against an intuitive decision to focus on the product, believing that users will come if the product is worthy. Instead, developers should at least concurrently discuss the development of the product with the prospective community.

Collaborative creation with the community. Continuing from the above point, it is possible to jump-start a community by directly talking to and involving the interested community [17]. An example would be Stack Overflow[7], the first domain site on Stack Exchange, where the owners made the creation of the site a collaboration with the community effort from the beginning. Having a relationship with an invested community makes it more likely that the system will benefit from a network effect [17, 21]. However, it can also create homophily and potential barriers to outside users [11].

Power-law. Power-law is a natural phenomenon for successful communities [20], describing the idea that a few users, relative to the many, contribute most of the ”valuable” content in a system [1]. For instance, in QA we might see that many users contribute a few questions, but a relatively few will contribute the best answers [3]. Wikipedia[8] is an example, where most successful projects have an equal distribution of edits between power-users (the most active) and rest of the community [20]. Despite this, attracting new users and maintaining casual users is also important to community growth [20, 27].

Culture and Community

Once a community has enough users to run itself, what can we expect from community behavior?

Communities are often not welcoming. An unfortunate feature of many peer-production sites is a calcification where new users are made to feel unwelcome by existing users [22]. Often, established users will criticize new users which has a direct negative impact on whether a new user stays and contributes value to the community. In addition, this sometimes even results in a community effectively cutting itself off from newer users, like Wikipedia [14]. Combating this entrenchment of established users is difficult, as appeals from moderators and administrators can go ignored [22].

Communities can become biased. A peer-production community can become biased to the point in which it no longer fulfills important aspects of its mission [16]. If leadership or power shifts towards a specific viewpoint, it can present the resulting data product as being a collective truth, rather than a representative truth. It, therefore, drives out opposing or non-complementary opinions and viewpoints. This describes the reinforcement of a social bubble, in which users will only seek to produce where they feel affinity with the group’s goals.

Cultural Differences . Peer-production is impossible to transfer without unexpected changes to the data product [4]. Review sites are notoriously culture sensitive, where there are different norms of

[7] https://stackoverflow.com

[8] https://wikipedia.com

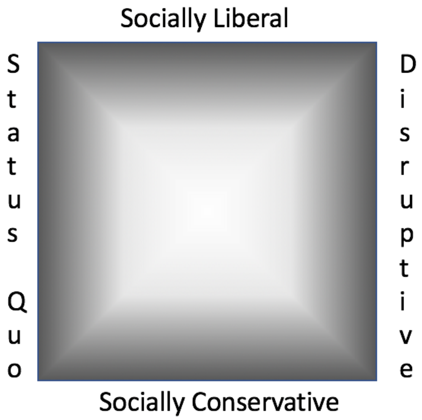

Figure 1: An Example of a Political Quadrant Ranking for SANCTUARY.

distribution [15]. Wikipedia and Stack Overflow are examples of where changing the language of a topic changes the content of the corresponding material, regardless of the technical nature of the topic [25].

OVERVIEW OF SANCTUARY

In order to have a grounded discussion of the merits of peer-production, we present the framework for SANCTUARY. SANCTUARY is an application to support individuals with monitoring their media and information intake. Previous research has shown that individuals tend to build social circles within which the majority of people are in agreement with each other on important topics, also known as “social bubbles” [18]. SANCTUARY aims to mitigate online bubbles and bias through fostering self-awareness of media consumption habits and a better understanding of crowd evaluation of content. In addition, this project will provide the public-at large with an important resource that is not based on data aggregated and controlled by large corporations or a government entities. This means that users can be more confident that their usage is not directly feeding entities that seek to control or extend their online media usage, and provides a check against automated services.

Peer-Production Mechanism

Users of the application would be both beneficiaries and contributors to the project. A user would provide evalutation on news content, identifying its political lean and controversy level, and would receive points for accurate evaluations that matches the crowd evaluation. At the same time, the user would receive peer-produced feedback on their consumption habits, showing them how they perceive information compared to their peers.

The proposed mechanics for the system is using an effective weighting system that leverages human-evaluation and scales for machine learning. The application will achieve its back-end through peer-produced ratings. The initial steps, therefore, requires a community of invested users to categorize material that can then be applied to the system. One of the essential parts of this project is to have an open platform that is independent of the overt influence of media producers and hosts, instead relying on its user base to power the application.

The base mechanism for this application is the rating of content from a variety of mainstream media sources. As shown in Figure 1., the proposed mechanic asks user to rate where the content places in the four political quadrants using a gradient scale. The evaluation of the content continues for each factor until the score stabilizes as visualized in Figure 2. Stabilization occurs when a maximal set of different score fails to move the aggregated score outside of the set threshold. The longer the stabilization takes, the higher the controversy coefficient.

Users are not allowed to see the ongoing score until stabilization occurs. The system mechanic gamifies the process by rewarding reputation points accurate evaluations and also rewarding earlier

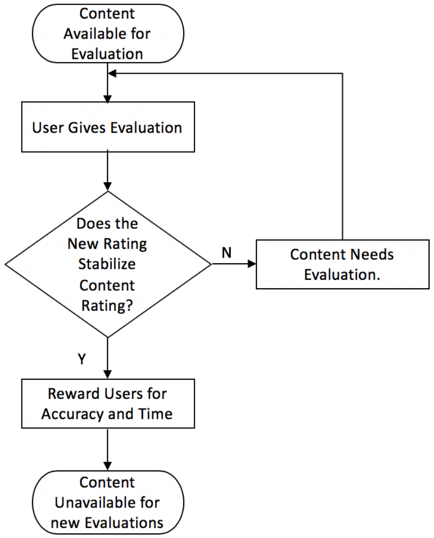
Content Available for Evaluation
User Gives Evaluation
Does the New Rating Stabilize Content Rating?
N
Content Needs Evaluation.
Y
Reward Users for Accuracy and Time
Content Unavailable for new Evaluations

Figure 2: Flowchart of Content Evaluation System for SANCTUARY.

accurate evaluations. The goal of this gamification is to encourage users to contribute evaluations which reflect an objective opinion rather than an emotional one.

Another aim of the project is to take the system from centralized control, to one that has distributed administration, and can start to handle content from a wider range of sources. In order to do this, users who earn reputation points will gain privileges to help with system upkeep and possible auxiliary features like discussion forums and moderation.

DESIGN RECOMMENDATIONS AND DISCUSSION

In order for SANCTUARY to be successful, it needs to engage with an enthusiastic community of core uses and read critical mass. This gives it the best chance of expanding a network of dedicated participants that can keep the system growing towards sustainability. However, it is important for the site designers to remember to pursue a purposefully diverse community. A risk for a peer-production system is that its community becomes an echo-chamber that produces agreement within itself but does not reflect a larger world. This may be the biggest challenge system designers, as it means they must continually invest in reinforcing the community message.

A peer-production system should not rely purely on altruism or self-improvement as motivations, but should encourage gamification and harness the power of competition. The benefits of competition seem to be an effective motivator for many users to contribute the best contributions possible, and at least serves as an important catalyst. However, developing the content rating systems and gamification is going to be an iterative process. It is vital for the developers to have a dialogue with the prospective community while developing the mechanisms. Engaging with a broad spectrum of potential users will help to ensure that the mechanism itself is not biased.

Determining the administrative level of control is going to be very important as the community ages. A loose administrative grip of the community would serve SANCTUARY as it seeks to scale upwards and have more distributed administration. But it also could lead to a situation in which the aim of the community is high-jacked by groups who hope to reduce system efficacy. A tight administrative grip, on the other hand, may fail to escape the biases of the system administrators and lose potential power-users. The designers will have to carefully negotiate the correct balance between direct and indirect, most likely with discussion and iterative attempts in the community building phase.

CONCLUSION

In this paper, we outlined the use of peer-production for the creation of a community-based system aimed at identifying and mitigating cognitive biases. A peer-production system has an important place in in this area, as it provides an alternative to authority driven solutions. Instead of relying on authority, the system relies on the aggregation of human judgement.

Implementation of a peer-production system is not easy and requires careful community cultivation and maintenance. A project that uses peer-production to create an easily consumed index of bias is unlikely to ever become a dominant world-wide factor in combating cognitive biases. The needs for user tie to the tasks and community organization, along with the march towards power-law means that in the best case scenario, a project like SANCTUARY could only ever hope to a small but active and sustainable community. Most people will never commit to such a project. However, if the goal is to provide a counterbalance to authority driven systems, then this method has potential to make a human-driven statement in the field of cognitive bias detection and mitigation.

ACKNOWLEDGEMENTS

This research is partially supported by a grant from the Keio University Fukuzawa Memorial Fund.

REFERENCES

[1] Lada A. Adamic, Jun Zhang, Eytan Bakshy, and Mark S. Ackerman. 2008. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In *Proceedings of the 17th International Conference on World Wide Web (WWW '08)*. ACM, New York, NY, USA, 665–674. https://doi.org/10.1145/1367497.1367587

[2] Desislava Aleksandrova, François Lareau, and Pierre André Ménard. 2019. Multilingual sentence-level bias detection in Wikipedia. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*. INCOMA Ltd., Varna, Bulgaria, 42–51. https://doi.org/10.26615/978-954-452-056-4_006

[3] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2012. Discovering Value from Community Activity on Focused Question Answering Sites: A Case Study of Stack Overflow. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '12)*. ACM, New York, NY, USA, 850–858. https://doi.org/10.1145/2339530.2339665

[4] Patti Bao, Brent Hecht, Samuel Carton, Mahmood Quaderi, Michael Horn, and Darren Gergle. 2012. Omnipedia: Bridging the Wikipedia Language Gap. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1075–1084. https://doi.org/10.1145/2207676.2208553

[5] Michel Bauwens. 2007. Why crowdsourcing isn't peer production. https://blog.p2pfoundation.net/why-crowdsourcing-is-peer-production/2007/03/08

[6] Yochai Benkler. 2016. Peer production and cooperation. *Handbook on the Economics of the Internet* (May 2016). https://www.elgaronline.com/view/edcoll/9780857939845/9780857939845.00012.xml

[7] Yochai Benkler and Helen Nissenbaum. 2006. Commons-based Peer Production and Virtue*. *Journal of Political Philosophy* 14, 4 (2006), 394–419. https://doi.org/10.1111/j.1467-9760.2006.00235.x

[8] Nicholas Buttrick, Jordan Axt, Charles R. Ebersole, and Jacalyn Huband. 2020. Re-assessing the incremental predictive validity of Implicit Association Tests. *Journal of Experimental Social Psychology* 88 (May 2020), 103941. https://doi.org/10.1016/j.jesp.2019.103941

[9] Soon Ae Chun, Richard Holowczak, Kannan Neten Dharan, Ruoyu Wang, Soumaydeep Basu, and James Geller. 2019. Detecting political bias trolls in Twitter data. In *WEBIST 2019 - Proceedings of the 15th International Conference on Web Information Systems and Technologies*. SciTePress, 334–342. https://www.researchwithnj.com/en/publications/detecting-political-bias-trolls-in-twitter-data

[10] Denzil Correa and Ashish Sureka. 2014. Chaff from the Wheat: Characterization and Modeling of Deleted Questions on Stack Overflow. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14)*. ACM, New York, NY, USA, 631–642. https://doi.org/10.1145/2566486.2568036

[11] Denae Ford, Justin Smith, Philip J. Guo, and Chris Parnin. 2016. Paradise unplugged: identifying barriers for female participation on stack overflow. In *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering (FSE 2016)*. Association for Computing Machinery, Seattle, WA, USA, 846–857. https://doi.org/10.1145/2950290.2950331

[12] Mingkun Gao, Hyo Jin Do, and Wai-Tat Fu. 2018. Burst Your Bubble! An Intelligent System for Improving Awareness of Diverse Social Opinions. In *23rd International Conference on Intelligent User Interfaces (IUI '18)*. Association for Computing Machinery, Tokyo, Japan, 371–383. https://doi.org/10.1145/3172944.3172970

[13] Anthony G. Greenwald, Mahzarin R. Banaji, and Brian A. Nosek. 2015. Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology* 108, 4 (2015), 553–561. https://doi.org/10.1037/pspa0000016

[14] Aaron Halfaker, R. Stuart Geiger, Jonathan T. Morgan, and John Riedl. 2013. The Rise and Decline of an Open Collaboration System How Wikipedia's Reaction to Popularity Is Causing Its Decline. *American Behavioral Scientist* 57, 5 (May 2013), 664–688. https://doi.org/10.1177/0002764212469365

[15] Noi Sian Koh, Nan Hu, and Eric K. Clemons. 2010. Do online reviews reflect a product's true perceived quality? An investigation of online movie reviews across cultures. *Electronic Commerce Research and Applications* 9, 5 (Sept. 2010), 374–385. https://doi.org/10.1016/j.elerap.2010.04.001

[16] Jaron Lanier. 2010. *You Are Not a Gadget*. Knopf Doubleday Publishing Group. Google-Books-ID: H76XlWv_FqQC.

[17] Lena Mamykina, Bella Manoim, Manas Mittal, George Hripcsak, and Björn Hartmann. 2011. Design Lessons from the Fastest Q&a Site in the West. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2857–2866. https://doi.org/10.1145/1978942.1979366

[18] Dimitar Nikolov, Diego F. M. Oliveira, Alessandro Flammini, and Filippo Menczer. 2015. Measuring online social bubbles. *PeerJ Computer Science* 1 (Dec. 2015), e38. https://doi.org/10.7717/peerj-cs.38

[19] Cathy O'Neil. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown. Google-Books-ID: NgEwCwAAQBAJ.

[20] Jacob Solomon and Rick Wash. 2014. Critical Mass of What? Exploring Community Growth in WikiProjects. In *Eighth International AAAI Conference on Weblogs and Social Media*. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8104

[21] Yla R. Tausczik and James W. Pennebaker. 2012. Participation in an Online Mathematics Community: Differentiating Motivations to Add. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 207–216. https://doi.org/10.1145/2145204.2145237

[22] Andrew W. Vargo and Shigeo Matsubara. 2016. Corrective or critical? Commenting on bad questions in Q&A. (March 2016). https://doi.org/10.9776/16199

[23] A. W. Vargo and S. Matsubara. 2016. Editing Unfit Questions in Q&A. In *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*. 107–112. https://doi.org/10.1109/IIAI-AAI.2016.83

[24] Andrew W. Vargo and Shigeo Matsubara. 2018. Identity and performance in technical Q&A. *Behaviour & Information Technology* 37, 7 (July 2018), 658–674. https://doi.org/10.1080/0144929X.2018.1474251

[25] A. W. Vargo, Benjamin Tag, Kai Kunze, and Shigeo Matsubara. 2018. Different Languages, Different Questions: Language Versioning in Q&A.

[26] Gang Wang, Konark Gill, Manish Mohanlal, Haitao Zheng, and Ben Y. Zhao. 2013. Wisdom in the Social Crowd: An Analysis of Quora. In *Proceedings of the 22Nd International Conference on World Wide Web (WWW '13)*. ACM, New York,

NY, USA, 1341–1352. https://doi.org/10.1145/2488388.2488506 event-place: Rio de Janeiro, Brazil.

[27] Haiyi Zhu, Amy Zhang, Jiping He, Robert E. Kraut, and Aniket Kittur. 2013. Effects of Peer Feedback on Contribution: A Field Experiment in Wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2253–2262. https://doi.org/10.1145/2470654.2481311