Joseph Feller, Jonas Gamalielsson, Benjamin Mako Hill and Gregorio Robles
(Eds.)

Proceedings of the Doctoral Consortium at the 15th International Symposium on Open Collaboration

Skövde, Sweden, 19 August 2019

UNIVERSITY OF SKÖVDE

Skövde University Studies in Informatics 2019:1

Feller, J., Gamalielsson, J., Mako Hill, B., and Robles, G. (Eds.) (2019) Proceedings of the Doctoral Consortium at the 15th International Symposium on Open Collaboration, Skövde University Studies in Informatics 2019:1, ISSN 1653-2325, ISBN 978-91-983667-4-7, University of Skövde, Skövde, Sweden.

Skövde University Studies in Informatics 2019:1
ISSN 1653-2325
ISBN 978-91-983667-4-7

www.his.se

UNIVERSITY
OF SKÖVDE

Proceedings of the Doctoral Consortium at the 15th International Symposium on Open Collaboration, 2019

Edited by:

Joseph Feller
University College Cork, Ireland

Jonas Gamalielsson
University of Skövde, Sweden

Benjamin Mako Hill
University of Washington, USA

Gregorio Robles
Universidad Rey Juan Carlos, Spain

Preface

The field of open collaboration brings together different strands of open collaboration research and practice, seeking to create synergies and inspire new collaborations between people from computer science, information science, social science, humanities, and everyone interested in understanding open collaboration and how it is changing the world. The OpenSym conference is a premier publication venue in the field and has reached its fifteenth edition this year.

To facilitate new researchers with an arena to present and receive feedback on their research, the OpenSym conference has had a doctoral consortium for several years. The principle objective of the consortium is to provide doctoral students the opportunity to present their research at various stages of production – from early drafts of their research design to near completion of their dissertation – in a forum where they can receive constructive feedback from a community of interested scholars and other students as they work to finish their degree. This volume contains the three papers, each of which was reviewed by members of the program committee. After the doctoral consortium, authors were given the opportunity to revise their papers based on the input they received from the reviewers and participants who provided feedback during the event.

This volume contains the revised versions of the papers, which were presented and discussed at the Doctoral Consortium at the 15th International Symposium on Open Collaboration, in Skövde, Sweden on 19 August 2019.

We wish to thank the reviewers and members of the program committee of the doctoral consortium who have provided valuable feedback on the papers. We also thank all Ph.D. students for their participation. Finally, we are grateful for the financial support from the TOTO project (University College Cork, Ireland) which is funded by the Lewis Charitable Foundation (USA).

Joseph Feller
Jonas Gamalielsson
Benjamin Mako Hill
Gregorio Robles

Program Committee

Joseph Feller	University College Cork, Ireland
Benjamin Mako Hill	University of Washington, USA
Gregorio Robles	Universidad Rey Juan Carlos, Spain

Table of Contents

Open science hardware: towards more democratic science and technology in Latin America? 1
Author & presented by: Julieta C. Arancio

Bots and Their Impact on Knowledge Diversity in Wikidata 6
Authors: Mariam Farda-Sarbas and Claudia Müller-Birn
Presented by: Mariam Farda-Sarbas

Ecology of Online Organizations 10
Author & presented by: Nathan TeBlunthuis

Open science hardware: towards more democratic science and technology in Latin America?

Julieta C. Arancio
CENIT-UNSAM; Universidad Nacional de Quilmes
Buenos Aires, Argentina
jarancio@unsam.edu.ar

ABSTRACT

Open Science Hardware (OScH) refers to the practice of sharing the designs of tools used for scientific research so anyone can obtain, assemble, use, study, modify, distribute and sell them[12]. OScH growth is evidenced by the upsurge of initiatives, communities, specialized journals, publications and platforms worldwide during the last five years[20][31]. Activists claim OScH democratizes science and technology especially in Global South contexts; however, how OScH practice enables this transition still remains mainly unexplored. This thesis, in data collection stage, aims to analyze democratization in OScH initiatives in Latin America by combining a socio-technical transitions framework with categories from social movements, feminist studies, user innovation literature and the capabilities approach. We expect to gain understanding on how OScH strategies can foster more and more diverse participation and empowerment around science and technology in a Global South context, and propose an analytic framework for studying transitions towards democratization in socio-technical systems.

ACM Classification Keywords

B.m. Hardware: Miscellaneous; K.4.m. Computers and Society: Miscellaneous

Author Keywords

Open hardware; open science; democratization; Latin America.

INTRODUCTION

Open Science Hardware (OScH) refers to the practice of sharing the designs of tools used for scientific research, so anyone can obtain, assemble, use, study, modify, distribute and sell them[12]. OScH practitioners present a diversity of mindsets, including free software philosophy, do-it-yourself culture, hacker culture and Appropriate Technology among others[24]. Similar to other domains such as Open Education, Open Access to scientific research or Open Pharma, OScH can be considered part of the open and collaborative production phenomena, enabled by massification of internet access[6].

OScH expansion in the last five years is evidenced by an exponential growth in number of publications, projects and diversity of domains[20], emergence of global and local communities[12] and creation of specialized infrastructure such as journals[16][11] and documentation platforms.

Drivers behind this trend include lower fabrication costs due to access to digital fabrication tools like 3D printing, access to specialized Free Open Source Software for hardware design and cheaper electronic components, combined with lower knowledge barriers due to design and expertise sharing, facilitated by internet access and open licensing[17].

Open Science Hardware uses

One of the domains where OScH has flourished is education, greatly influenced by the irruption of the Arduino community. Originally created for teaching interactive design, Arduino boards are currently used in robotics, labware, STEM education[27] among many others. Heradio et al.[20] identify a growing trend of publications about OScH in education since 2007, in an early stage oriented towards engaging students and facilitating women inclusion in STEM subjects; from 2010 on, moving towards enhancing university education and lowering research costs using OScH.

Other domain where OScH is gaining popularity is academic research. Since 2012, when Joshua Pearce published an early paper[29] on how to build scientific equipment with open source hardware, an exponential number of projects have appeared in the most diverse domains: microscopy[4], analytical chemistry[15], microfluidics[2], nanotechnology[14], agriculture[8], medical appliances[26].

A paradigmatic case within academia, the European Organization for Nuclear Research (CERN) has been developing and implementing open source hardware in the Large Hadron Collider control systems since 2009[40]. It has also released the CERN Open Hardware License in 2011[3] and maintains the Open Source Hardware Repository[10], which hosts both CERN and non-CERN projects.

Open science hardware is also used in extra-academic contexts, blurring the limits of expertise[33]. Examples include community-based environmental monitoring[7], community

OpenSym Doctoral Consortium, Aug 20–22, 2019, Skovde, Sweden

networks for development of synthetic biology projects[13] and projects combining art, science and activism[39].

Challenges and opportunities

Challenges for OScH scaling include limited funding due to dominance of the IP model among private funders and public agencies, slow dissemination due to OScH's distributed nature and resistance to adoption at the institutional level[31]. On top of these, other difficulties identified by the community include the need of better calibration methods and quality standards, development of successful business models, how to increase diversity in the community and methods to lower even more the current access barriers for users[12].

In terms of benefits, advocates claim OScH makes research processes more efficient: it lowers costs, enables repairing and customization of tools, minimizes dependence from suppliers, increases reproducibility and fosters user innovation[25]. Besides these efficiency-related benefits practitioners also highlight the potential of OScH as a democratization force of science and technology, in some cases linking it explicitly to the concept of appropriate technology[30]. Especially in Global South countries, arguments are OScH enables co-creation processes and capacity building, facilitates research in resource-constrained contexts and provides tools for new voices from community science to emerge[23].

RESEARCH PROBLEM

As mentioned in the introduction, OScH advocates claim it presents multiple advantages in diverse areas. Those associated to efficiency have been widely explored[29][4][25]. However, how OScH practice can enable a transition to more democratic knowledge production processes still remains mainly unexplored.

The main goal of this thesis is to determine if OScH initiatives in Latin America constitute a feasible niche for science and technology democratization in the region. We decompose this goal in four research questions:

1. RQ1: which are the framings, contexts and visions in OScH initiatives in Latin America that define them as an alternative niche of science and technology production?
2. RQ2: which are the resources mobilized and strategies developed by OScH initiatives in Latin America?
3. RQ3: who participates and who doesn't in OScH initiatives in Latin America, and how is agency distributed between participants?
4. RQ4: how do learning processes in OScH initiatives in Latin America translate into capacity building and empowerment of participants?

RELATED RESEARCH

OScH as a field of study is recent, but references to participation, democratization and collaboration can be found in related fields such as open innovation, open design or Do-It-Yourself/maker studies.

The following research is related to the topic of OScH and democratization of technology:

Tanenbaum et al.[37] describe how initiatives from maker culture present elements of 'democratized technological practices' which challenge traditional conceptions of technology user, and begin to form a politics of appropriation. Authors categorize these democratic elements in terms of i) playfulness, ii) decisions around tool use, iii) leveraging of industrial infrastructures around materials and standards, iv) learning and knowledge sharing.

Kera[23] studies how OScH global platforms can be considered examples of 'geek diplomacy'. The author defines it as grassroots involvement in science which bridges various knowledge and infrastructural divides, to create a more inclusive R&D response to challenging global multidimensional issues. According to the author, democratic features of OScH are increased participation, co-creation and collaboration, decentralization of research infrastructure and empowerment in terms of assessing risks, making decisions and formulating new context-relevant research questions. In this way OScH builds a public sphere for free deliberation and enables technological empowerment which is material, discursive and social.

Bonvoisin et al.[9] analyze documentation of open hardware projects online in order to determine how participative they are in terms of transparency, workload distribution, number of contributors and (de)centralization. Results of repository mining show distributed development in two thirds of projects analyzed and different forms of organization similar to those studied in Free Open Source Software projects.

ANALYTIC FRAMEWORK

OScH proposes a distributed, open and collaborative paradigm of science hardware production in opposition to the current closed and centralized approach, tied to Intellectual Property (IP). Through the lens of socio-technical transitions theory[32][22] and its multi-level perspective categories -niche, regime, landscape-[19] we aim to analyze if OScH in Latin America constitutes a viable *niche* for change from the current IP *regime* to a new configuration of actors, practices, regulations, artifacts and their links; and which dynamics that change may take according to modifications in the wider political, cultural and economic background -the *landscape* level-.

The innovations associated with OScH practice are more complex than mere technological enhancements: the changes proposed are value-driven, powered by a common vision resulting from the interaction and negotiation between multiple heterogeneous mindsets. In order to study this complexity we use categories from social movements literature[38][5], in particular from grassroots innovation movements[36][21]. These allow us to analyze the *visions* and motivations within *framings* in the OScH community, the contexts in which these framings arise, the *strategies, resources and spaces* operated by the community to achieve its vision and how it can contribute to the creation of *alternative pathways* that otherwise would not exist.

Advocates claim OScH democratizes science: it "allows a diversity of values and voices to ask research questions and to make technology", "more people and more types of people can take part in and benefit from science", it "empowers

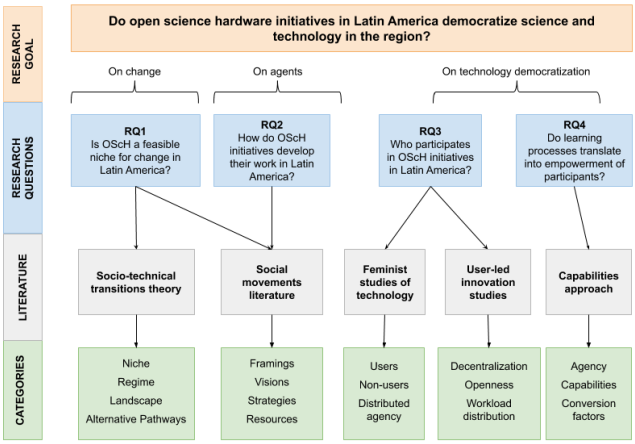

Figure 1. Scheme of analytic framework showing main goal, research questions, main literature and categories used

people to achieve their ideas at low cost". In order to turn the idea of democratization of technology[18][35] into operative categories for analysis, we take concepts from feminist studies[1][28][41], user-led innovation literature[17] and the capabilities approach[34], all of which are anchored in the users' perspective.

First we ask about diversity and power: *users* who participate in OScH projects and *non-users*; *agency distribution* among them and how they interact with technology through the artifacts' *scripts*. Second, we want to understand participation: we look at models of interaction between users and between projects, degrees of *openness* of product and process, and degrees of *(de)centralization*. Last, we aim to analyze, at the user level, if learning processes within OScH projects lead to *capacity building*, how *conversion factors* affect this process, and if it translates into empowerment through *agency*.

The analytic framework of this thesis, as shown in Figure 1, is therefore a combination of three main themes: socio-technical transitions theory, social movements literature and democratization of technology, the latest made operative through three user-centered approaches: feminist studies, user innovation literature and capabilities approach.

METHODOLOGY

Methodology of this project includes a combination of qualitative and quantitative approaches, using primary sources of information such as interviews, online surveys and field notes together with valuable metadata about OScH projects available in online repositories.

Initial activities included a comprehensive literature review on open science hardware, its main uses, opportunities, challenges and significant features; definition of the analytic framework combining categories from different bodies of literature to understand viability of OScH as a niche towards democratization of socio-technical systems, and methodological design of the project.

Empiric analysis activities included:

1. We built a database of OScH projects in Latin America from public information on attendance to the Global Open Science Hardware (GOSH) community gatherings combined with data from online forums, journals and repositories.

2. We carried on exploratory interviews to participants of GOSH 2017 gathering in Santiago, Chile.

3. We interviewed 14 members of the Global Open Science Hardware community in 2018, with the aim of reconstructing its origins and understand better the global actors' framings, motivations and strategies.

4. Insights from exploratory interviews were used as input for the design of an online survey, sent to all identified OScH projects in Latin America (still open).

5. Results of this survey were used to enrich the original database, detect a preliminary typology and select five cases to be studied more deeply through semi-structured interviews, starting May 2019.

Each case will contain a qualitative analysis structured according to the categories of the analytic framework complemented with a section for descriptive statistics and social network analysis. To complement the regional analysis with a global perspective we will include for each case an analysis of a similar global initiative, and in the macro level a chapter on the history of the Global Open Science Hardware community, enriched with data analysis from online sources.

As per commitment to open science values and methodologies, this project constitutes in itself an effort for developing the whole research process as openly as possible. For that purpose only free open source software tools are used for reference management, interviews, data collection and analysis. Evolution of the writing process can be followed at `https://github.com/thessaly/phd` together with preliminary versions of databases and code for analysis, taking into account consent and privacy according to GDPR standards. A reflection on outcomes, opportunities and barriers encountered during this process will be also included in the final text.

EXPECTED CONTRIBUTION

We expect to provide a comprehensive open survey of OScH initiatives in Latin America, which is currently not available, insights on the mechanisms and strategies put into play, identification of barriers and opportunities, and use it as a baseline for making policy recommendations aimed to support OScH initiatives in Global South contexts.

We expect to contribute to the Global Open Science Hardware community with a document containing information on its origins and an open and collaborative database powered by WikiData showing actors, resources, strategies and networks, that can be used for education, research and dissemination purposes. A very basic example can be found in Figure 2.

We also expect the reflections on the process of writing a Social Studies PhD in an open and collaborative way can provide useful insights and motivation for future researchers in the field.

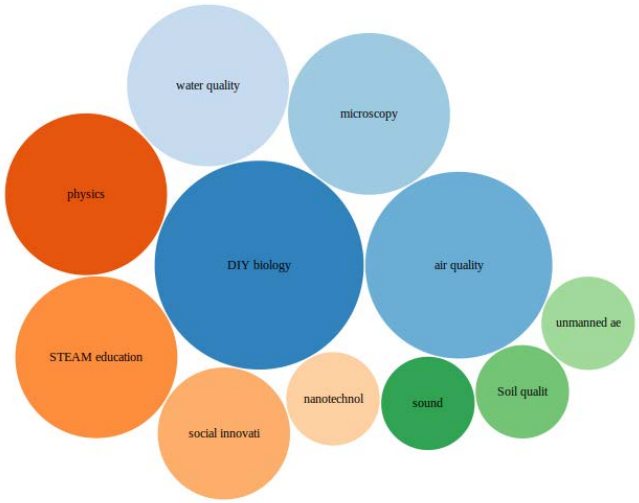

Figure 2. A preliminary bubble graph showing frequency of OScH projects in Latin America by field of work, using Wikidata Query Service (https://w.wiki/5K7).

Finally, we propose a framework we consider valuable for studying transitions towards democratization of socio-technical systems, a phenomena that has gained relevance in the last decades and researchers[18][35] consider will continue growing in the years to come.

REFERENCES

1. Madeleine Akrich. 1992. The De-Scription of Technical Objects. In *Shaping Technology/Building Society. Studies in Sociotechnical Change*. MIT Press, Cambridge, MA.
2. Justyna Ausareny, Denisa Kera, Stefania Druga, and Yair Reshef. 2014. Open Source Hardware (OSHW) Supporting Interaction between Traditional Crafts and Emergent Science: Wayang Kulit over Microfluidic Interfaces. In *SIGGRAPH Asia 2014 Designing Tools For Crafting Interactive Artifacts on - SIGGRAPH ASIA '14*. ACM Press, Shenzhen, China, 1–4. DOI: http://dx.doi.org/10.1145/2668947.2668955
3. Myriam Ayass and Javier Serrano. 2012. The CERN Open Hardware License. *IFOSS L. Rev.* 4 (2012), 71.
4. Tom Baden, Andre Maia Chagas, Greg Gage, Timothy Marzullo, Lucia L. Prieto-Godino, and Thomas Euler. 2015. Open Labware: 3-D Printing Your Own Lab Equipment. *PLoS Biology* 13, 3 (2015), 1–12. DOI: http://dx.doi.org/10.1371/journal.pbio.1002086
5. Robert D. Benford and David A. Snow. 2000. Framing Processes and Social Movements: An Overview and Assessment. *Annual Review of Sociology* 26, 1 (Aug. 2000), 611–639. DOI: http://dx.doi.org/10.1146/annurev.soc.26.1.611
6. Yochai Benkler. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Vol. 7. Yale University Press, New Haven [Conn.], USA. DOI:http://dx.doi.org/10.1177/0894439307301373
7. Paz Bernaldo and Gustavo Pereyra Irujo. 2018. Proyecto "Vuela". *Liinc em Revista* 14, 1 (2018), 11. DOI:http://dx.doi.org/https://doi.org/10.18617/liinc.v14i1.4237
8. Giovanni Bitella, Roberta Rossi, Rocco Bochicchio, Michele Perniola, and Mariana Amato. 2014. A Novel Low-Cost Open-Hardware Platform for Monitoring Soil Water Content and Multiple Soil-Air-Vegetation Parameters. *Sensors* 14, 10 (Oct. 2014), 19639–19659. DOI:http://dx.doi.org/10.3390/s141019639
9. Jérémy Bonvoisin, Tom Buchert, Maurice Preidel, and Rainer G. Stark. 2018. How Participative Is Open Source Hardware? Insights from Online Repository Mining. *Design Science* 4, E19 (2018). DOI: http://dx.doi.org/doi:10.1017/dsj.2018.15
10. CERN. 2009. Open Hardware Repository. (2009). Retrieved 2019-04-22 06:42:49 from https://localhost/project/ohr-support/wikis/welcome/raw
11. GOSH community. 2017. Journal of Open Hardware. (2017). Retrieved 2019-04-22 06:42:49 from http://openhardware.metajnl.com/
12. Global Open Science Hardware (GOSH) Community. 2018. GOSH Roadmap. (2018). Retrieved 2019-04-22 06:42:49 from http://openhardware.science/wp-content/uploads/2017/12/GOSH-roadmap-smll.pdf
13. Hackteria community. 2009. Hackteria. (2009). Retrieved 2019-04-22 06:42:49 from https://www.hackteria.org/
14. Tulsi R. Damase, Daniel Stephens, Adam Spencer, and Peter B. Allen. 2015. Open Source and DIY Hardware for DNA Nanotechnology Labs. *Journal of Biological Methods* 2, 3 (Aug. 2015), 24. DOI: http://dx.doi.org/10.14440/jbm.2015.72
15. Michael D. M. Dryden, Ryan Fobel, Christian Fobel, and Aaron R. Wheeler. 2017. Upon the Shoulders of Giants: Open-Source Hardware and Software in Analytical Chemistry. *Analytical Chemistry* 89, 8 (April 2017), 4330–4338. DOI: http://dx.doi.org/10.1021/acs.analchem.7b00485
16. Elsevier. 2016. HardwareX. (2016). Retrieved 2019-04-22 06:42:49 from https://www.journals.elsevier.com/hardwarex/
17. Eric von Hippel. 2005. *Democratizing Innovation*. The MIT press.
18. Andrew Feenberg. 2002. *Transforming Technology: A Critical Theory Revisited*. Oxford University Press, New York, N.Y.
19. Frank W Geels. 2002. Technological Transitions as Evolutionary Reconfiguration Processes: A Multi-Level Perspective and a Case-Study. *Research Policy* 31 (2002), 1257–1274. DOI: http://dx.doi.org/10.1016/S0048-7333(02)00062-8

20. Ruben Heradio, Jesus Chacon, Hector Vargas, Daniel Galan, Jacobo Saenz, Luis De La Torre, and Sebastian Dormido. 2018. Open-Source Hardware in Education: A Systematic Mapping Study. *IEEE Access* 6 (2018), 72094–72103. DOI: http://dx.doi.org/10.1109/ACCESS.2018.2881929

21. David Hess, Steve Breyman, Nancy Campbell, and Brian Martin. 2007. Science, Technology, and Social Movements. In *The Handbook of Science and Technology Studies*. MIT Press, 26.

22. René Kemp, Johan Schot, and Remco Hoogma. 1998. Regime Shifts to Sustainability through Processes of Niche Formation: The Approach of Strategic Niche Management. *Technology Analysis & Strategic Management* 10, 2 (1998), 175–198. DOI: http://dx.doi.org/10.1080/09537329808524310

23. Denisa Kera. 2015. Open Source Hardware (OSHW) for Open Science in the Global South: Geek Diplomacy? In *Open Science, Open Issues*. IBICT, Brasilia, 292.

24. Denisa Kera. 2017. Science Artisans and Open Science Hardware. *Bulletin of Science, Technology & Society* 37, 2 (June 2017), 97–111. DOI: http://dx.doi.org/10.1177/0270467618774978

25. André Maia Chagas. 2018. Haves and Have Nots Must Find a Better Way: The Case for Open Scientific Hardware. *PLOS Biology* 16, 9 (Sept. 2018), e3000014. DOI:http://dx.doi.org/10.1371/journal.pbio.3000014

26. Gerrit Niezen, Parisa Eslambolchilar, and Harold Thimbleby. 2016. Open-Source Hardware for Medical Devices. *BMJ Innovations* 2, 2 (April 2016), 78–83. DOI: http://dx.doi.org/10.1136/bmjinnov-2015-000080

27. Mehmet Akif Ocak. 2017. Where does ArduinoâĂŹs power come from?: An extended literature review. *Journal of Learning and Teaching in Digital Age (JOLTIDA)* 3, 1 (2017), 21–34.

28. Nelly Oudshoorn, Margo Brouns, and Ellen van Oost. 2005. Diversity and Distributed Agency in the Design and Use of Medical Video-Communication Technologies. In *Inside the Politics of Technology*, Hans Harbers (Ed.). Amsterdam University Press, Amsterdam, 85–106. DOI: http://dx.doi.org/10.1515/9789048503841-005

29. Joshua M Pearce. 2012a. Building research equipment with free, open-source hardware. *Science* 337, 6100 (2012), 1303–1304.

30. Joshua M. Pearce. 2012b. The case for open source appropriate technology. *Environment, Development and Sustainability* 14, 3 (01 Jun 2012), 425–431. DOI: http://dx.doi.org/10.1007/s10668-012-9337-9

31. Joshua M. Pearce. 2017. Emerging Business Models for Open Source Hardware. *Journal of Open Hardware* 1, 1 (March 2017), 2. DOI:http://dx.doi.org/10.5334/joh.4

32. A Rip and R Kemp. 1998. *Technological Change*. Vol. Volume 2. Battelle Press, Columbus, Ohio.

33. S., Jalbert Wylie K., Dosemagen S., and Ratto M. 2014. Institutions for Civic Technoscience: How Critical Making Is Transforming Environmental Research. *The Information Society* 30, 2 (2014), 116–126. DOI: http://dx.doi.org/10.1080/01972243.2014.875783

34. Amartya Sen. 1999. *Development as Freedom*. Alfred A. Knopf, New York.

35. Adrian Smith and Andrew Stirling. 2018. Innovation, Sustainability and Democracy: An Analysis of Grassroots Contributions. *Journal of Self-Governance and Management Economics* 6, 1 (2018), 64–97. DOI: http://dx.doi.org/10.22381/JSME6120183

36. Adrian Smith et al. 2017. *Grassroots Innovation Movements*. Routledge, Abingdon, Oxon ; New York, NY.

37. Joshua G Tanenbaum and Simon Fraser. 2013. Democratizing Technology: Pleasure, Utility and Expressiveness in DIY and Maker Practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 2603–2612.

38. Charles Tilly. 2008. *Contentious Performances*. Cambridge University Press, New York, NY, USA.

39. Mary Tsang. 2017. *From Biomolecules to Biopolitics... Hormones with Institutional Biopower!* Doctoral Dissertation. Massachusetts Institute of Technology. Retrieved 2019-04-22 06:42:49 from https://dspace.mit.edu/handle/1721.1/112560

40. E van der Bij, J Serrano, T Wlostowski, M Cattin, E Gousiou, P Alvarez Sanchez, A Boccardi, N Voumard, and G Penacoba. 2012. Open Hardware for CERN's Accelerator Control Systems. *Journal of Instrumentation* 7, 01 (2012), C01032–C01032. DOI: http://dx.doi.org/10.1088/1748-0221/7/01/c01032

41. Sally Wyatt. 2005. Non-Users Also Matter: The Construction of Users and Non-Users of the Internet. In *Now Users Matter: The Co-Construction of Users and Technology*. MIT Press, 67–79.

Bots and Their Impact on Knowledge Diversity in Wikidata

Mariam Farda-Sarbas
Freie Universität Berlin
Berlin, Germany
mariam.fs@fu-berlin.de

Claudia Müller-Birn
Freie Universität Berlin
Berlin, Germany
clmb@inf.fu-berlin.de

ABSTRACT

Wikidata is one of the most edited knowledge bases which was launched in 2012. The contributing community, as shown by research, consists of humans and bots, with bots being the most active editors of Wikidata. Bots can edit larger amounts of data and at a higher pace than humans, thus, their edits can have considerable impact on data quality in Wikidata. Data quality as 'fitness for use' in the Wikidata context is the data that can be accessed by anyone anywhere in the world, and is partially implemented by allowing multiple values for an entity in the form of aliases, multilingual labels, multiple statements, claims and references (sources). This schema design for coexistence of diverse or contradictory statements is called plurality which is a design principle of Wikidata, and refers to the concept of knowledge diversity. Knowledge diversity supports Wikidata's overarching goal of becoming a global knowledge base through storage of knowledge without a judgment of good or bad. This research aims to study how data quality in Wikidata can be improved by improving knowledge diversity using bots.

Author Keywords
Knowledge diversity; Wikidata; Bot; Data quality.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous; See http://acm.org/about/class/1998 for the full list of ACM classifiers. This section is required.

INTRODUCTION

Wikidata, the sister project of Wikipedia, is a collaborative knowledge base, which is freely accessible and contains human-readable, as well as, machine-readable data. The main goal behind Wikidata development was to provide a centralized structured data source for Wikimedia projects, and to overcome the data inconsistencies of Wikipedia. The keywords defining Wikidata are "free, collaborative, multilingual, secondary database, collecting structured data, support for Wikimedia Wikis and to anyone in the world[1]" which refers to the capability of representing diverse knowledge open for anyone anywhere in the world. Thus, data quality as `fitness for use' in this context can be to serve anyone anywhere in the world with data which is understandable (language), relevant (relates to geographical, values and religious context), usable (friendly UI) and available. Efforts have been made to improve multilingualism and make data understandable, however, results show that Wikidata is currently reflecting western knowledge and a majority of languages are not yet well covered[1][2][3]. In comparison to Wikipedia, which also allows knowledge diversity in form of different language versions of Wikipedia, Wikidata has all the information in a unified from[4]. In Wikipedia every language's editors are free to add knowledge which is in the context of their own values (e.g. culture, language, religious beliefs and values). In Wikidata, data is stored in a language independent form, using ids where the letter Q succeeded by numbers (e.g., Q64 for Berlin) is used for items and the letter P succeeded by numbers (e.g., P1082 for Population) is used for properties. Then, labels are used to provide information in a certain language. This uniform representation of diverse data raises the question, if all human knowledge without judgment of good or bad, can be represented in Wikidata. There are currently more than 54 million entities[2] contributed by both, humans and bots. While, humans have an active community of around 20,000 active users, the majority of contributions come from bots which are much less in number (270[3]). This indicates a considerable impact of bots' edits on the diversity and quality of data in Wikidata. Thus, further related issues for investigation are the current status of diversity in the existing knowledge which has mostly come from active contribution of bots, the impact of diversity on

[3] https://www.wikidata.org/wiki/Wikidata:Bots

trustworthiness and quality of data, and improvement of data quality in Wikidata through knowledge diversity.

LITERATURE REVIEW

Wikidata has been studied from different angles including data quality and bots, however, knowledge diversity in Wikidata is not yet given much attention and only multilingualism or language diversity dimension of Wikidata is explored so far. In the following comes existing related literature on bots, data quality and knowledge diversity in Wikidata to present state-of-the-art in the field.

Bots

Bots are most commonly defined as software programs that automate tasks [5], usually repetitive [6] or routine tasks [7] which humans consider time consuming and tedious.

Due to this general definition, it is a challenge to differentiate between bots and their similar counterparts like, scripts, programs, software agents, and robots. Wikidata as the sister project of Wikipedia, besides other things from Wikipedia, has inherited the usage of bots as well.

Steiner [8] aims to understand editing distribution of user on Wikidata and Wikipedia. His developed application observes real-time edit activity on Wikidata and tracks bots, logged-in and anonymous users. The study shows that most of Wikidata, i.e. 88\%, is edited by bots. Müller-Birn et. al [5] study the community editing patterns of Wikidata through a cluster analysis of contributors' editing activities and identify six editing patterns of the participating community. Hall et. al [9] develops a machine learning model to detect bots through informal editing characteristics and implicit behaviors. The importance of identifying bots is claimed by the fact that 2-3\% of the contributions (more than 1 million edits), considered as human contributions are actually coming from unidentified bots. This could be very damaging in case of vandalism. There are also researches on bots behaviors which study how bots use SPARQL query service and what are the bots editing patterns. Bielefeldt et. al [10] analyze the Wikidata SPARQL Query Service access logs and distinguish them between organic (i.e. human) and robotic traffic. The study shows that organic queries are more complex and diverse, while, robotic queries are simple and make the largest portion of the queries.

Research shows that Wikidata has most of the edits done by bots, thus, it is important to monitor their tasks and analyze their impact.

Data Quality

Data quality is intended to be fitness for use, and in Wikidata it is mostly studied from a completeness perspective, as Prasojo et al. [11] develops "COOL-WD", a tool for supporting the completeness lifecycle of Wikidata. Similarly, Ahmeti et al. [12] and Balaraman et al. [13] propose and develop Recoin, a relative completeness tool for evaluating completeness of entities in Wikidata. Brasileiro et al. [14] discuss the quality of taxonomic hierarchies in Wikidata to have a consistent data model and representation schema. Piscopo et al. [15][16] analyze Wikidata quality from the provenance perspective, the relevance and authoritativeness of Wikidata external references. Razniewski et al. [17] introduce the problems and limitations of properties in Wikidata and propose entity-specific property ranking for Wikidata.

Knowledge diversity

Knowledge diversity in the context of knowledge bases needs further efforts and research. Diversity is defined by Giunchiglia et al. as "the co-existence of contradictory opinions and/or statements (some typically non-factual or referring to opposing beliefs/opinions)"[18]. A number of the related diversity dimensions mentioned in [18] are diversity of: sources, resources, language, view point, geographical and temporal diversity. Among the mentioned dimensions, multilingualism or diversity of language have been studied so far.

Although, Wikidata supports language diversity by nature, research shows that not all languages have the same data representation state and only some languages like, English, Dutch, French, Deutsch, Spanish, Italian, and Russian cover most of the content in Wikidata, while, other languages are not well represented [1]. In another study [19], authors propose a mechanism to enrich Wikidata multilingual content by retrieving semantic relations based on alignment between info-box properties and Wikidata properties in various languages, and "the outcomes mainly contribute these semantic relations back to Wikidata and Wikipedias, especially ones are based on the Latin alphabet".

HYPOTHESIS

Data quality in Wikidata can be affected by bots, as they are doing most of the edits, thus, the assumption here is: bots edits have impact on knowledge diversity, and by improving knowledge diversity, data quality in Wikidata can be improved.

The hypothesis contains three main issues (i.e., bots, knowledge diversity, data quality) in the context of Wikidata, and to address this hypothesis investigation is needed on how these issues relate to each other, how to measure diversity and quality of bot edits, how to improve knowledge diversity and, as a result, data quality.

PROPOSED SOLUTION

To improve knowledge diversity in Wikidata, the proposed solution is intended to, first, investigate for a measurement approach to assess knowledge diversity in Wikidata, second, develop a recommender system which recommends different data diversity dimensions for editors to improve data diversity, and third, measure diversity aspect of data after application of recommender system, and compare the results of first and third steps to judge the improvements in knowledge diversity.

METHODOLOGY

The approach considered for this research is mainly divided in the following three perspectives:

1. Theoretical Perspective: This stage includes survey of related literature on Wikidata, bots, data quality and knowledge diversity, to provide state-of-the-art in each area and define the hypothesis more precisely.
2. Engineering Perspective: Based on the insights from the theoretical perspective, approaches are investigated to measure the existing levels of diversity, and based on the measurements, model the current knowledge diversity aspect in Wikidata. Then, analyze the current model to identify diversity gaps in Wikidata and propose approaches to improve the diversity model and refine the current model accordingly.
3. Empirical Perspective: Based on the proposed model of diversity in Wikidata, a software (i.e., recommender system) is developed and tested to measure how diversity levels have changed in comparison to the results prior to software implementation. Then, analyze the final results with relation to the stated hypothesis.

CURRENT STATE OF RESEARCH:

I have so far performed a review of research on Wikidata and studied bots in Wikidata.

REFERENCES

[1] L.-A. Kaffee, A. Piscopo, P. Vougiouklis, E. Simperl, L. Carr, and L. Pintscher, "A Glimpse into Babel: An Analysis of Multilinguality in Wikidata," in *Proceedings of the 13th International Symposium on Open Collaboration*, New York, NY, USA, 2017, pp. 14:1–14:5.

[2] L.-A. Kaffee *et al.*, "Mind the (Language) Gap: Generation of Multilingual Wikipedia Summaries from Wikidata for ArticlePlaceholders," in *The Semantic Web*, 2018, pp. 319–334.

[3] L.-A. Kaffee *et al.*, "Learning to Generate Wikipedia Summaries for Underserved Languages from Wikidata," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 2018, vol. 2, pp. 640–645.

[4] D. Vrandečić and M. Krötzsch, "Wikidata: a free collaborative knowledgebase," *Commun. ACM*, vol. 57, no. 10, pp. 78–85, 2014.

[5] C. Müller-Birn, B. Karran, J. Lehmann, and M. Luczak-Rösch, "Peer-production system or collaborative ontology engineering effort: What is Wikidata?," in *Proceedings of the 11th International Symposium on Open Collaboration*, 2015, p. 20.

[6] M.-A. Storey and A. Zagalsky, "Disrupting Developer Productivity One Bot at a Time," in *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering*, New York, NY, USA, 2016, pp. 928–931.

[7] R. S. Geiger, "Bots are Users, Too! Rethinking the Roles of Software Agents in HCI," *Tiny Trans. Comput. Sci.*, vol. 1, p. 1, 2012.

[8] T. Steiner, "Bots vs. Wikipedians, Anons vs. Logged-Ins (Redux): A Global Study of Edit Activity on Wikipedia and Wikidata," in *Proceedings of The International Symposium on Open Collaboration*, New York, NY, USA, 2014, pp. 25:1–25:7.

[9] A. Hall, L. Terveen, and A. Halfaker, "Bot Detection in Wikidata Using Behavioral and Other Informal Cues," *Proc. ACM Hum.-Comput. Interact.*, vol. 2, no. CSCW, pp. 1–18, Nov. 2018.

[10] A. Bielefeldt, J. Gonsior, and M. Krötzsch, "Practical Linked Data Access via SPARQL: The Case of Wikidata," in *Linked Data on the Web (LDOW2018)*, Lyon, France, 2018, p. 10.

[11] R. E. Prasojo, F. Darari, S. Razniewski, and W. Nutt, "Managing and Consuming Completeness Information for Wikidata Using COOL-WD," in *Proceedings of the 7th International Workshop on Consuming Linked Data co-located with 15th International Semantic Web Conference (ISWC 2015)*, Kobe, Japan, 2016, vol. Vol-1666.

[12] A. Ahmeti, S. Razniewski, and A. Polleres, "Assessing the Completeness of Entities in Knowledge Bases," in *The Semantic Web: ESWC 2017 Satellite Events*, 2017, pp. 7–11.

[13] V. Balaraman, S. Razniewski, and W. Nutt, "Recoin: Relative Completeness in Wikidata," in *Companion Proceedings of the The Web Conference 2018*, Republic and Canton of Geneva, Switzerland, 2018, pp. 1787–1792.

[14] F. Brasileiro, J. P. A. Almeida, V. A. Carvalho, and G. Guizzardi, "Applying a multi-level modeling theory to

assess taxonomic hierarchies in Wikidata," in *Proceedings of the 25th International Conference Companion on World Wide Web*, 2016, pp. 975–980.
[15] A. Piscopo, L.-A. Kaffee, C. Phethean, and E. Simperl, "Provenance Information in a Collaborative Knowledge Graph: An Evaluation of Wikidata External References," in *The Semantic Web – ISWC 2017*, 2017, pp. 542–558.
[16] A. Piscopo, C. Phethean, and E. Simperl, "What Makes a Good Collaborative Knowledge Graph: Group Composition and Quality in Wikidata," in *Social Informatics*, 2017, pp. 305–322.
[17] S. Razniewski, V. Balaraman, and W. Nutt, "Doctoral Advisor or Medical Condition: Towards Entity-Specific Rankings of Knowledge Base Properties," in *Advanced Data Mining and Applications*, 2017, pp. 526–540.
[18] F. Giunchiglia *et al.*, "Foundations for the representation of diversity, evolution, opinion and bias," p. 94, Aug. 2018.
[19] T. H. Ta and C. Anutariya, "A Model for Enriching Multilingual Wikipedias Using Infobox and Wikidata Property Alignment," in *Semantic Technology*, Springer, Cham, 2014, pp. 335–350.

Ecology of Online Organizations

Nathan TeBlunthuis
Department of Communication
Community Data Science Collective
Wikimedia Foundation
Seattle, WA, USA
nathante@uw.edu

ABSTRACT

Would Wikipedia be one of the most visited websites in the world if other online collaborative encyclopedia projects had been more established when it was founded? Or was Wikipedia helped by the fact that its predecessors had engaged and trained hundreds of its future contributors? Do new discussion communities on Reddit compete with one another over content or contributors? Is the evolving world of online organizations better understood as a competitive struggle for resources or as symbiotic relationships that support a web of interdependent communities?

Established approaches to the comparative study of online organization success have almost exclusively looked inside organizations [e.g., 7, 11, 15, 16, 23]. But answering the questions above requires an *ecological understanding* of online organizations that accounts for the complex dynamic interactions between communities and their environments.

Analyses of ecological factors in the life sciences enable effective wildlife management, pest control, and sustainable utilization of renewable resources. In sociology and management, organizational ecology provides compelling explanations for the life-cycles of industries, organizational specialization, and patterns of collaborative partnerships. Similarly, ecology may offer designers and managers of online organizations new conceptual and methodological tools for sustaining platform ecosystems.

ACM Reference Format:
Nathan TeBlunthuis. 2019. Ecology of Online Organizations. In *Proceedings of Open Sym 2019 (OpenSym).* ACM, New York, NY, USA, 6 pages. https://doi.org/10.475/123_4

1 STAGE OF PH.D

I am a 4th year Ph.D student, I have completed my exams. I will have defended my proposal by the time of the conference. I will graduate in 2020 or 2021.

OpenSym, August 2019, Skövde, Sweden

2 PROBLEM DOMAIN

Online organizations are a dynamic, growing, and increasingly important form of organization.[1] Through peer production, the Wikipedia has produced the largest collaborative effort and most important reference work in human history. Free/libre open source software (FLOSS) project have produced tens of billions of dollars worth of software made freely available online [3]. Other online groups like Reddit sometimes engage in peer production, and often provide information, social support, and entertainment to millions of people. Online platforms support millions of attempts to organize projects or communities but only a tiny percentage manage to mobilize participants and to sustain collaboration [9, 10, 15, 18]. The significance of these accomplishments should not be understated. Provisioning new categories of global public goods like free software and encyclopedias is a rare achievement for an organizational form that places peer production in an elite class shared with governments, markets, social movements, and universities.

Despite the success of peer production, the decline in active contributors to Wikipedia motivates renewed attention to the challenges of maintaining a pool of active contributors to an online organization [19]. Most prior studies of the growth, survival, and success of online organizations have focused almost exclusively on their internal features with an emphasis on attracting and retaining participants [10, 11]. Consider Kraut et al.'s claim that people will contribute to online groups when the benefits to them exceed the costs [11]. Benefits of participation, such as intrinsic or extrinsic motivations for contributing, and identity and bonds-based commitments of participants to the group, in turn depend on maintaining a *critical mass* of active contributors [11].

On the other hand, the size of online groups may also be limited by the social structures and technical tools used to manage quality and regulate behavior [7, 22]. An increasing group size leads to increasing costs to participation and so communities will grow to an equilibrium size at their *carrying capacity*. Butler considers costs of communication in

[1] Online organizations are also commonly called "online communities." Because I am considering an organizational theory known as "community ecology", I choose the terms "online organization" and "online group" to avoid confusion.

larger groups [4] and the decline of Wikipedia suggests that that systems for maintaining the established order and content quality impose high costs on inexperienced and underrepresented participants [7].

Yet evidence supporting such lifecycle-based accounts of online organization success may be confounded by external factors that can drive participation in online organizations. Because this approach does not account for the fact that contributor time and energy are finite, it cannot rule out alternative ecological explanations including the presence of competitors, an exhausted pool of potential participants, and so on [19]. For example, the contemporaneous rise of Facebook and other social networking sites remains a possible alternative explanation for Wikipedia's transition from rise to decline. As a result, we still do not understand how the growth of online organizations is limited by these types of external resources.

3 KEY IDEA AND RESEARCH QUESTIONS

This project proposes to develop an Ecological approach to understanding online organization success that builds on established approaches to the ecological study of biological populations [24, 27] and organizations [8, 13]. No competent wildlife biologist predicting the survival of an animal population would only consider the species' internals and physiology in isolation. Instead, they would consider the availability of necessary resources like food and shelter, the presence or absence of other organisms, and competitive dynamics that might give the species an advantage. This project's overarching goal is to transform current understandings of the conditions for successful online organization through an ecological analysis of the dynamic interactions between organizations and their environments.

I will conduct three interlinked projects answering fundamental ecological questions: (1) How does the growth and survival of online organizations depend on their ecological communities? (2) How do ecosystems of online groups themselves develop over time? (3) How do the choices of individual participants in online organizations lead to ecological dynamics? These three questions relate three different levels of analysis: ecological communities (networks of interdependent groups), the interdependent online groups that comprise them, and the individuals who contribute to the groups.

4 RELATED WORK

This approach to online organizations is not altogether new. Indeed a handful studies published in HCI have already taken up the framework [20, 21, 26, 28, 29]. Most notable are a series of of three related studies, [26, 28] and [29], that apply density dependence theory from organizational ecology in studies of three different community ecosystems. One of the most striking findings from this work is that newly formed Wikia wikis have greater longevity when they have many early contributors who also participate in more established wikis [29].

Prior applications of organizational ecology to online communities that only translate theories from organizational contexts to online communities have encountered surprising results that suggest some specific open opportunities to study how ecological dynamics depend on distinctive features of the online environment. In two similar studies, [26] observe evidence for competition between Usenet groups while [29] observe evidence for mutualism between Wikia wikis. These findings usefully illustrate how resource overlaps do not necessarily lead to a given type of ecological relationship [24], but importantly they also suggest that ecological dynamics in online ecosystems may depend on factors specific to the context and nature of online communities: the topic or identity associated with a community, the distinctive motivations of online community organizers, characteristics of platforms hosting communities, and the stage of the development of ecological communities [26].

These ecological studies of online organizations in HCI draw from population ecology, one of several strands of ecological theory in organization science. Population ecology focuses on the growth and survival of a single population of organizations (such as firms with the same business model). Therefore it assumes that the organizations under consideration are interchangeable. But I think that online communities are generally not interchangeable. Indeed it seems to me that no two communities are alike. They each have their own ecological niche. In this way, each community is more akin to an ecological population than to an individual organism.

In contrast to population ecology, which "emphasizes forces that make organizations more uniform rather than more diverse," community ecology focuses on the diverse roles of each organization in the ecosystem [1, p. 224]. I draw most directly on McPherson et al.'s application of the community ecology approach to study interdependent voluntary organizations like churches, unions, and clubs to contribute a new focus on the relationships (such as competitive or mutualistic relationships) between them [12]. Indeed prior studies have not attempted to model online environments in terms of such dyadic interactions—a first step in community ecology analysis [24].

Many other studies in approach relationships between online groups, but without adopting an explicitly ecological approach. Tan explores reproduction and inheritance in online organizations and references ecology as an inspiration, but doesn't synthesize its empirical contributions into a general theory of community growth or survival [20]. In

general, such contextually focused empirical analyses of interdependence [e.g. 5, 6, 25] are valuable contributions making informative empirical contributions and demonstrate the potential relevance of ecological modeling to online organizations, but more can be done to marshal the full power of ecology to explain why some online groups grow and survive while others do not.

These studies provide evidence suggesting that ecological dynamics are important drivers of online community outcomes, and that neglecting them is one reason that researchers have had difficulty in understanding or predicting outcomes. Explicating a theory of ecological dynamics in the online environment requires going beyond straightforward translation of theory from one domain into another. It requires new theoretical development with careful attention to how ecological dynamics that drive online organizations' successes may diverge from those found in other domains. Ecology has potential to provide new conceptual, analytical, and computational tools to inform community leaders and platform designers. More theoretically and empirically rigorous steps toward an ecology of online organizations are overdue.

5 CONTRIBUTION TO HCI, SOCIAL COMPUTING, AND PEER PRODUCTION

Conventional approaches to explaining online community growth and survival in HCI are typically based in social psychology and engineering. However, they account for only a small amount of variation in communities' growth, longevity, and performance. This dissertation will contribute by drawing ecological approaches from organization science as an alternative approach.

In biology and organization studies, ecological approaches have shown that success is largely —and sometimes overwhelmingly—a function of what others in an individual's environment are doing [8]. From an ecological view, the individual organism or online community cannot be fully understood except through in its interdependence relations with others; its function and role in the broader ecosystem [27].

Ecology also promises implications for the design and management of online organizations. Analyses of ecological factors in the life sciences enable effective wildlife management, pest control, and sustainable utilization of renewable resources. In sociology, organizational ecology provides compelling explanations for the life-cycles of industries, organizational specialization, and patterns of collaborative partnerships.

Similarly, an ecological understanding of online organizations can identify environmental conditions favoring projects in given niches to suggest the creation of new communities to fill them; predict how regulatory decisions that platforms

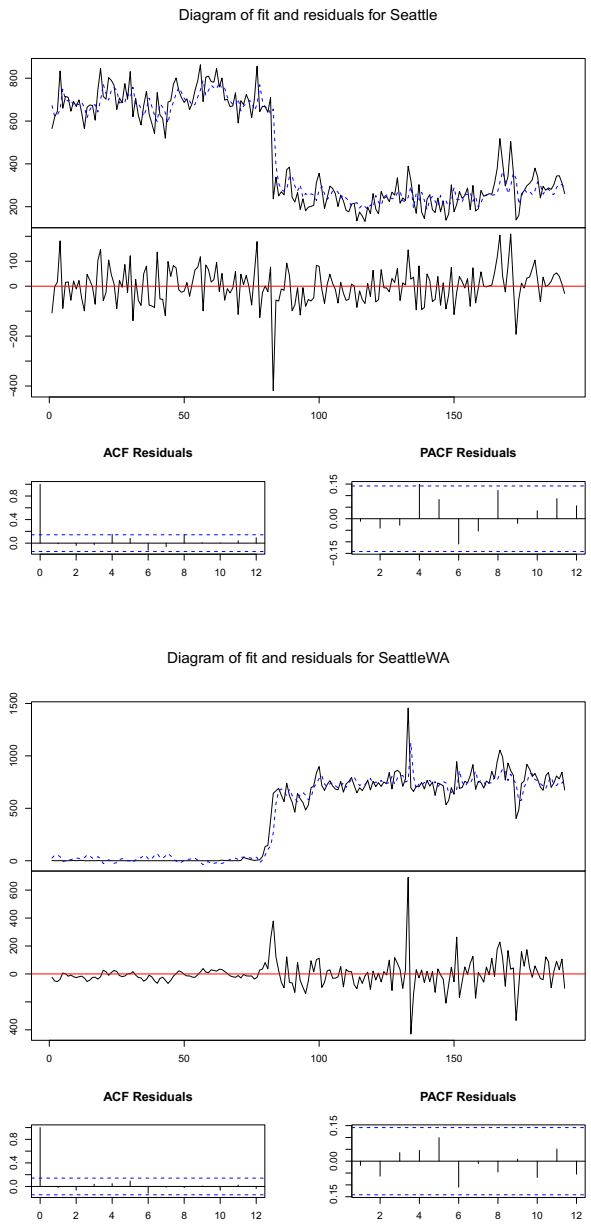

Figure 1: Fit of VAR model to data from /r/SeattleWA and /r/Seattle. The top plots show the data in black and the model fit in blue. The plot below that shows the residuals. The small plots at the bottom are diagnostic plots showing that the residuals are not serially correlated as consistent with model assumptions.

make through policy or code will affect the health of particular individual communities; simply by better explaining the success and failure of communities in terms of forces beyond those communities' direct control, the ecological approach can help us know whether we should attribute the success or failure of given communities to their own choices or to external forces which may drive their fates. In statistical terms, ecological forces can confound observational studies of online organizations. Finally, shifting our view of success to focus on ecological dynamics can help the open collaboration movement consider how to act collectively as a body of disparate, but interrelated organizations.

In addition, this project will transfer advanced statistical approaches to time-series analysis (briefly described below) from ecology and social science to HCI.

6 EXECUTION PLAN

To answer my research question (1), I propose an analysis in terms of the ecological relationships between communities (i.e. competition and mutualism) and of their overlapping human and topical resources. I will use a hierarchical vector auto-regression model (hVAR), a cutting-edge statistical approach from ecology, finance, political science, and economics to identify ecological relationships and their associations with user and content overlaps [2]. To answer research question (2), I will extend the hVAR model to analyze how these relationships change over time and test hypotheses from ecology and organizational sociology which propose that ecological communities become more stable over time.

I have already experimented with these kinds of models in some preparatory work. I'll briefly illustrate by this approach using the two different subreddits about Seattle: /r/Seattle and /r/SeattleWA.

The story of /r/SeattleWA and /r/Seattle is one of ecological competition between communities with (at this point I assume) a high degree of topical and user overlap. As a small-time participant in these communities I have first-hand knowledge of the rise of /r/SeattleWA in 2016.[2] The short story is that the /r/Seattle community rebelled against a heavy-handed moderator and most members of /r/SeattleWA defected to organize a new community at /r/SeattleWA. Thereafter the two communities co-exist, with /r/Seattle having around 800 active weekly contributors and /r/Seattle around 300. Figure 1 shows the fit of a VAR model to the time series of active commentators to /r/Seattle and /r/SeattleWA.

As shown in Figure 2 The model predicts that an exogenous shock to /r/Seattle causing a cumulative increase in 400 participants over 10 weeks is expected to cause a decrease of around 200 participants in /r/SeattleWA. Similarly,

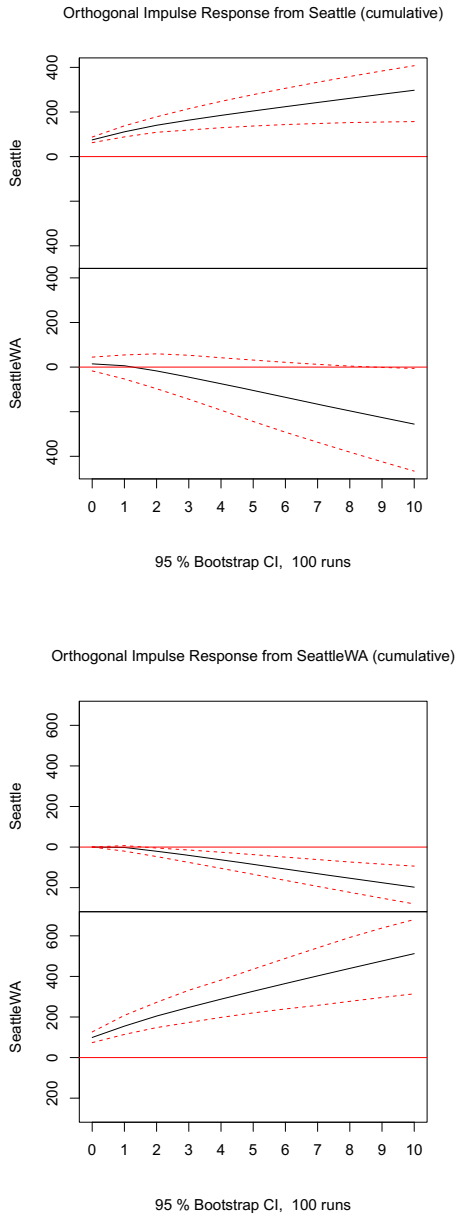

Figure 2: The left plot shows how the model predicts the two communities would be effected by an exogenous increase in participation in /r/Seattle and the right shows the response from a comparable shock to /r/SeattleWA. Qualitatively, the plot shows evidence of competition between the two communities.

[2] Also see https://www.seattleweekly.com/news/seattles-reddit-community-is-big-active-and-at-war-with-itself/ for media coverage.

an exogenous shock to /r/SeattleWA causing an increase of around 500 participants over 10 weeks will cause a decrease of around 200 participants to /r/Seattle. This illustrates how I will identify competitive or mutualistic relationships, but I will extend this approach to jointly identify the relationships and learn whether they are more or less likely to occur under membership and content overlaps.

For the research question (3) I will use an agent-based model (ABM) to demonstrate how a parsimonious description of an individual decision making process can lead to empirically observed macro level patterns of community growth and survival and of competitive and mutualistic relationships between communities. Similar processes have successfully been studied in this way. Perhaps most famously, Shelling [14] developed a model to show that even a mild preference for racial homophily, if sufficiently widespread, can lead to stable patterns of racial segregation. ABMs have also been used in ecological studies in the tradition of McPherson to model how organizations that isolate their members to avoid competition diminish their chances for recruitment and that to survive organizations should adopt strategies with a level of turnover appropriate to local environments [17].

I will evaluate my ABM in terms of its ability to produce empirical patterns at high levels of analysis such as the participation levels, community longevity, and the number of communities in which individuals participate.

REFERENCES

[1] W. Graham Astley. 1985. The Two Ecologies: Population and Community Perspectives on Organizational Evolution. *Administrative Science Quarterly* 30, 2 (1985), 224–241. https://doi.org/10.2307/2393106 00620.

[2] Miguel A. G. Belmonte, Gary Koop, and Dimitris Korobilis. 2014. Hierarchical Shrinkage in Time-Varying Parameter Models. *Journal of Forecasting* 33, 1 (2014), 80–94. https://doi.org/10.1002/for.2276

[3] Yochai Benkler, Aaron Shaw, and Benjamin Mako Hill. 2015. Peer Production: A Form of Collective Intelligence. In *Handbook of Collective Intelligence*, Thomas W. Malone and Michael S. Bernstein (Eds.). MIT Press, Cambridge, MA, 175–204.

[4] Brian S. Butler. 2001. Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures. *Information Systems Research* 12, 4 (Dec. 2001), 346–362. https://doi.org/10.1287/isre.12.4.346.9703

[5] Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. You Can'T Stay Here: The Efficacy of Reddit's 2015 Ban Examined Through Hate Speech. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW (Dec. 2017), 31:1–31:22. https://doi.org/10.1145/3134666

[6] Srayan Datta, Chanda Phelan, and Eytan Adar. 2017. Identifying Misaligned Inter-Group Links and Communities. *Proceedings of the ACM on Human-Computer Interaction, CSCW* 1 (Dec. 2017), 1–23. https://doi.org/10.1145/3134672

[7] 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

[8] Michael T. Hannan and John Freeman. 1989. *Organizational Ecology* (1 ed.). Harvard University Press, Cambridge, MA.

[9] Kieran Healy and Alan Schussman. 2003. The Ecology of Open-Source Software Development. (2003). http://kieranhealy.org/files/drafts/oss-activity.pdf

[10] Benjamin Mako Hill and Aaron D. Shaw. 2019. Studying Populations of Online Communities. In *The Handbook of Networked Communication*, Wells B Foucault and Sandra González-Bailón (Eds.). Oxford University Press, New York, New York, 25.

[11] Robert E. Kraut, Paul Resnick, and Sara Kiesler. 2012. *Building Successful Online Communities: Evidence-Based Social Design.* MIT Press, Cambridge, MA.

[12] J. Miller McPherson. 1983. An Ecology of Affiliation. *American Sociological Review* 48, 4 (1983), 519–532. https://doi.org/10.2307/2117719

[13] Martin Ruef. 2000. The Emergence of Organizational Forms: A Community Ecology Approach. *Amer. J. Sociology* 106, 3 (Nov. 2000), 658–714. https://doi.org/10.1086/318963

[14] Thomas C. Schelling. 1971. Dynamic Models of Segregation. *The Journal of Mathematical Sociology* 1, 2 (July 1971), 143–186. https://doi.org/10.1080/0022250X.1971.9989794

[15] Charles M. Schweik and Robert C. English. 2012. *Internet Success: A Study of Open-Source Software Commons.* MIT Press, Cambridge, MA.

[16] Aaron Shaw and Benjamin Mako Hill. 2014. Laboratories of Oligarchy? How the Iron Law Extends to Peer Production. *Journal of Communication* 64, 2 (2014), 215–238. https://doi.org/10.1111/jcom.12082

[17] Yongren Shi, Fedor A. Dokshin, Michael Genkin, and Matthew E. Brashears. 2017. A Member Saved Is a Member Earned? The Recruitment-Retention Trade-Off and Organizational Strategies for Membership Growth. *American Sociological Review* 82, 2 (April 2017), 407–434. https://doi.org/10.1177/0003122417693616

[18] Clay. Shirky. 2008. *Here Comes Everybody : The Power of Organizing without Organizations.* Penguin Press, New York, NY.

[19] Bongwon Suh, Gregorio Convertino, Ed H. Chi, and Peter Pirolli. 2009. The Singularity Is Not near: Slowing Growth of Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration (WikiSym '09).* ACM, New York, NY, 1–10. https://doi.org/10.1145/1641309.1641322

[20] Chenhao Tan. 2018. Tracing Community Genealogy: How New Communities Emerge from the Old. In *Proceedings of the Twelfth International Conference on Web and Social Media (ICWSM '18).* AAAI, Palo Alto, California, 395–404. https://aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17811

[21] Chenhao Tan and Lillian Lee. 2015. All Who Wander: On the Prevalence and Characteristics of Multi-Community Engagement. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15).* International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 1056–1066. https://doi.org/10.1145/2736277.2741661

[22] Nathan TeBlunthuis, Aaron Shaw, and Benjamin Mako Hill. 2017. Density Dependence without Resource Partitioning: Population Ecology on Change.Org. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17 Companion).* ACM, New York, NY, USA, 323–326. https://doi.org/10.1145/3022198.3026358

[23] Nathan TeBlunthuis, Aaron Shaw, and Benjamin Mako Hill. 2018. Revisiting "The Rise and Decline" in a Population of Peer Production Projects. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18).* ACM, New York, NY, 355:1–355:7. https://doi.org/10.1145/3173574.3173929

[24] Herman A Verhoef and Peter J Morin. 2010. *Community Ecology: Processes, Models, and Applications.* Oxford University Press, Oxford. OCLC: 876676566.

[25] Nicholas Vincent, Isaac Johnson, and Brent Hecht. 2018. Examining Wikipedia with a Broader Lens: Quantifying the Value of Wikipedia's Relationships with Other Large-Scale Online Communities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, 566:1–566:13. https://doi.org/10.1145/3173574.3174140

[26] Xiaoqing Wang, Brian S. Butler, and Yuqing Ren. 2013. The Impact of Membership Overlap on Growth: An Ecological Competition View of Online Groups. *Organization Science* 24, 2 (2013), 414–431. https://doi.org/10.1287/orsc.1120.0756

[27] Donald Worster. 1994. *Nature's Economy: A History of Ecological Ideas*. Cambridge University Press, Cambridge; New York, NY, USA. http://site.ebrary.com/id/10740550 OCLC: 855524849.

[28] Haiyi Zhu, Jilin Chen, Tara Matthews, Aditya Pal, Hernan Badenes, and Robert E. Kraut. 2014. Selecting an Effective Niche: An Ecological View of the Success of Online Communities. ACM Press, 301–310. https://doi.org/10.1145/2556288.2557348 00001.

[29] Haiyi Zhu, Robert E. Kraut, and Aniket Kittur. 2014. The Impact of Membership Overlap on the Survival of Online Communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, 281–290. https://doi.org/10.1145/2556288.2557213

Feller, J., Gamalielsson, J., Mako Hill, B., and Robles, G. (Eds.) (2019) Proceedings of the Doctoral Consortium at the 15th International Symposium on Open Collaboration, Skövde University Studies in Informatics 2019:1, ISSN 1653-2325, ISBN 978-91-983667-4-7, University of Skövde, Skövde, Sweden.

Skövde University Studies in Informatics 2019:1
ISSN 1653-2325
ISBN 978-91-983667-4-7

www.his.se

UNIVERSITY
OF SKÖVDE