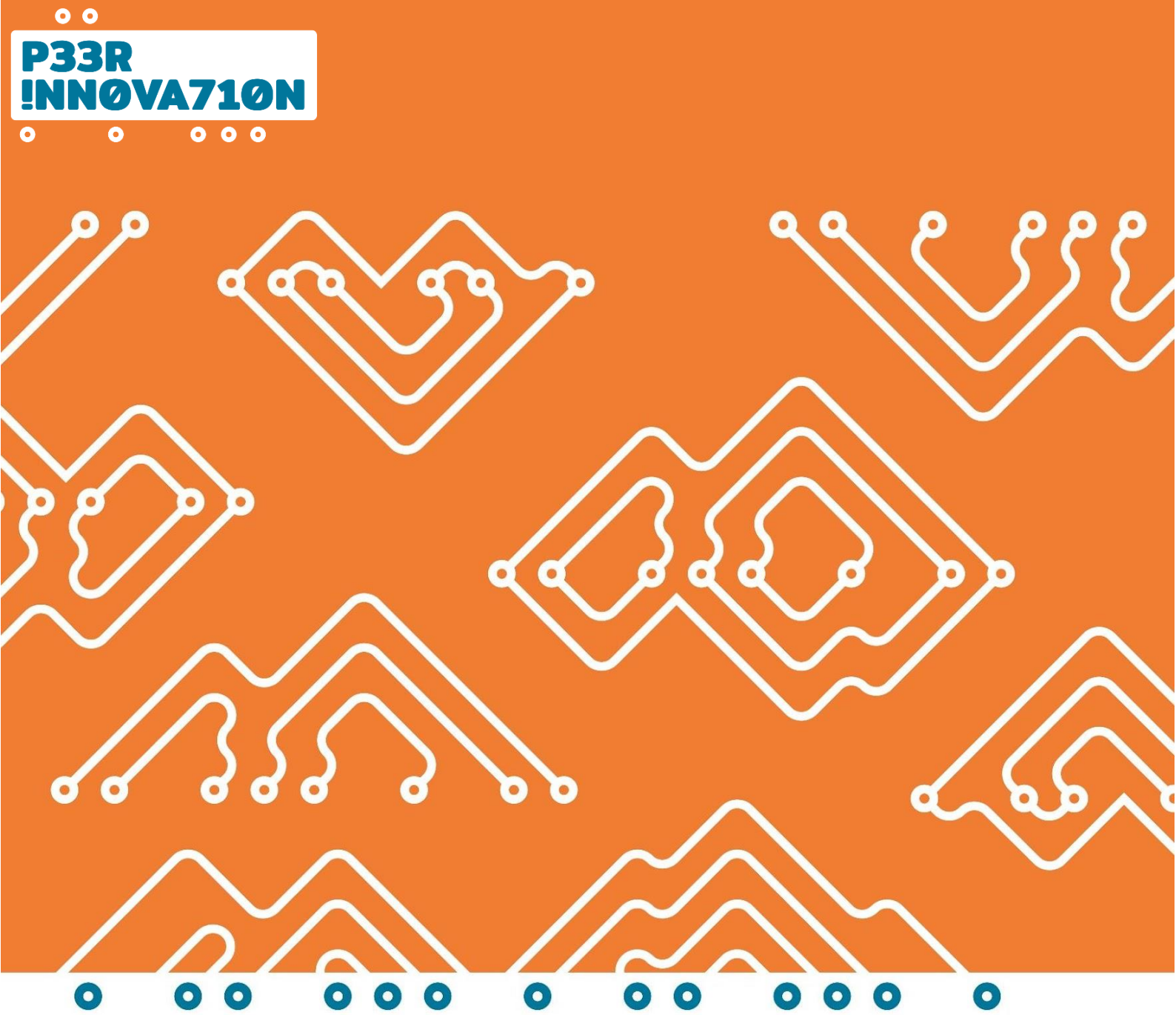


Measuring peer innovation in online communities

Discussion paper

Measuring peer innovation in online communities

Discussion paper for the expert workshop "How innovative are peer communities? Approaches to the empirical assessment of peer innovation" on 14 January 2022

Philipp Heß (TU Berlin), Anna Gleu (TU Berlin)

philipp.hess@tu-berlin.de

Abstract: This discussion paper describes the why and how of our current approach to measuring innovation activities in online communities. A special focus lies on the description of how we operationalized the concept of "peer innovation" to manually label interactions in online forums. We label whole forum threads and threads' initial posts to be able to train, evaluate, and select indicators for peer innovation that are based on properties of contributions, contributors, and community extracted from online forum data. Our current dataset, for which we present first labelling results, comprises the online forums of the Precious Plastic, OpenEnergyMonitor, and OpenStreetMap communities.

Introduction

Household sector innovation (HHI) is a growing and already far-reaching phenomenon (Franke et al., 2016; Gambardella et al., 2016; von Hippel et al., 2012). It receives increasing recognition in the innovation community, most prominently visible in the extended definition of innovation in the latest version of the Oslo Manual, which acknowledges the role of non-firm actors both on the supply and demand side of innovation (OECD/Eurostat, 2019, p. 61). Yet, policymakers and industry systematically underestimate the relevance of HHI (Bradonjic et al., 2019; Gault, 2018; Nielsen, 2020). This is also due to a lack of appropriate innovation indicators. Established indicators mainly focus on the activity of market actors (Gault, 2013), which leads to an under-representation of HHI in official statistics (Franke et al., 2016).

A major part of HHI takes place in web-based peer communities (Baldwin & von Hippel, 2011; Dahlander & Frederiksen, 2012). Prior work has highlighted how peer communities contribute to innovation and sustainability (Peuckert, Färber, et al., 2020). More formally, such "peer innovation" can be defined as "*a specific form of innovation in the household sector in which private individuals voluntarily and self determinedly collaborate to develop novel technical solutions and openly share them with a peer community via online platforms*" (Peuckert, Färber, et al., 2020, p. 17).

Our project addresses this gap between the relevance of peer innovation and a lacking recognition that is associated with missing means of quantification. We aim to generate insights into peer innovation processes with the goal of

generating an analytical framework that involves the establishment of generalizable and automatable peer innovation indicators. A first and decisive step for the achievement of this goal is the operationalization of the concept of peer innovation, which lays the foundation for any following quantitative analysis. This paper consequently focuses on our approach to creating "ground truth" by manually labelling exchanges in peer innovation communities according to the incidence of innovation (activities).

The following sections describe the overall measurement concept, our labelling rules, and the labelling process. We further shortly describe employed data and first labelling outcomes before closing with a discussion of limitations and future considerations, which we would like to further develop together with the workshop participants.

Methodology

Our approach to creating an analytical framework for measuring peer innovation builds on the subjective assessment of (the incidence of) peer innovation in the online forums of exemplary communities. We focus on labelling innovation *activities* instead of a "final innovation outcome," arguing that peer innovation is a continuous process of knowledge creation with no natural end state and is rather characterised by a sequence of interrelated actions by different actors. Our level of observation is the forum thread, which is a sequence of contributions by different actors from the community. Following the model of an innovation lifecycle (Korbel & Grosse, 2020; Gambardella et al., 2016), we assess whether exchanges in a thread involve the innovation activities *presentation of idea, evaluation, implementation, modification,* or *improvement.* The assessments are made based on rules that are iteratively updated with growing knowledge about the phenomenon. While our main focus lies on the evaluation of innovation activities, we decided to further subjectively assess the potential for innovations to emerge from the exchanges in a thread (3-item scale: "no potential", "some potential", "potential"), and to label the threads' initial posts according to Grosse et al. (2018), as this equips us with extended means of evaluating our own assessments.

The evaluation process can be seen as a netnographic study that gives the involved researchers a deeper understanding of the communities and equips them with more domain knowledge. The labelling outcome, a set of labelled threads, serves as input to a comparison of potential innovation indicators with the researchers' understanding of peer innovation (activity). The generated training data set can furthermore serve as input for additional analyses of peer innovation based on (semi-) supervised machine learning.

For the following evaluation of indicators for peer innovation, the project builds on prior research on innovation indicators in the context of online communities, user innovation, and open innovation. A literature analysis conducted in a previous work package yielded a selection of potential indicators that satisfied four criteria (Pohlisch & Peuckert, 2021): Providing an indication of innovation activities of non-commercial actors in online communities (1) based on publicly available data (2), while solely relying on quantitative measurement (no qualitative interpretation) (3), and being rooted in innovation theory (4). The identified indicators represented a wide range of approaches focused on different units of observation, features, and different data types. They typically addressed either contributions made in online forums (both single posts or whole threads), the contributors (users), or, to a lesser extent, the whole community. Features that were found to be related to innovation ranged from contribution uniqueness and elaboration, over contributor diversity, to the sentiments expressed in a discussion or the position of the contributors within the community network.

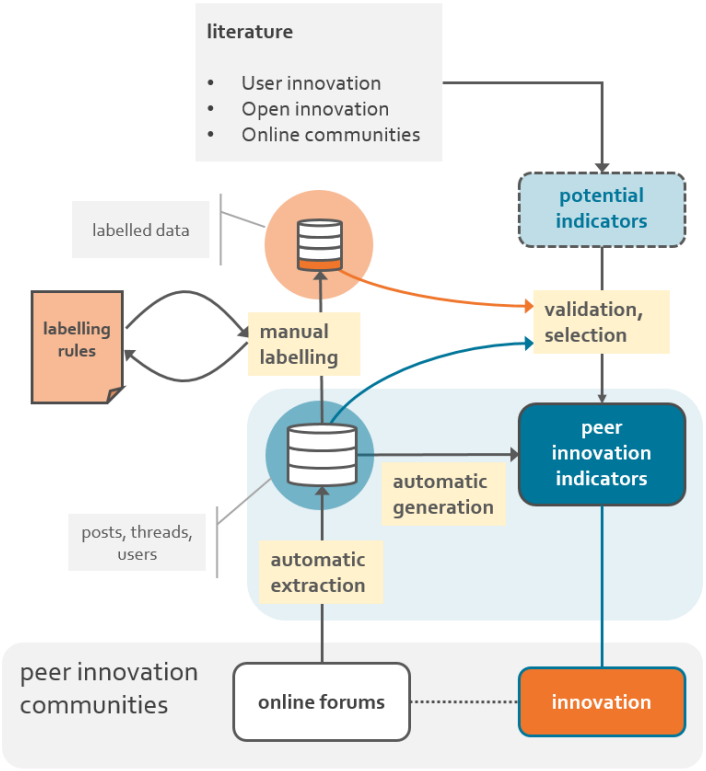


Figure 1: Measuring approach

Labelling rules

After multiple iterations of reading through forum threads, applying labels, accessing inter-rater agreement, discussing labelling decisions, and updating label specifications (see results section), we arrived at the following definitions and examples for the five types of innovation activity. Threads were labelled according to the incidence of different innovation activities and the assessed innovation potential. Initial posts are labelled with the likely intention(s) that motivated the contributor. We developed our own labelling categories for innovation activities and potential. For labelling the initial posts, we use the definitions by Grosse et al. (2018). Table 1 gives an overview of all labels. The exact definitions are stated in the following sub-sections.

Table 1: Labels

Level of observation	Category	Label	Exclusivity
Thread	Innovation activity	Presentation of idea	Non-exclusive
		Evaluation	
		Implementation	
		Modification	
		Improvement	
	Innovation potential	Has no potential (0)	Exclusive
		Has some potential (1)	

		Has potential (2)	
Initial post	Intention	Needing help	Exclusive
		Requesting feedback	
		Providing feedback	
		Disseminating information	
		Sharing developments	
		Calling for action	

Innovation activity

Presentation of idea:

- An idea with added value or novelty is presented in a post within the thread.
- Irrespective of the associated development status or utilization
- Examples:
    - A thread with an initial post asking "How about creating an iPhone app for OSM?" is not labelled "presentation of idea" because the reaction in the thread shows that such apps already exist (novelty criterion).
    - A thread with an initial post that presents pictures of a finalized machine is labeled "presentation of idea" because building the machine is based on an initial idea.

Evaluation:

- The thread contains an evaluation of a product, process, or idea that originated within the community.
- A typical evaluation assesses the idea's value, potential, feasibility, or relevance or expresses the evaluator's general interest.
- Reactions with an organizational or administrative character, or with the intent of pointing the contributor of the idea to additional material, are not counted as evaluation, e.g., "this is the wrong forum to post in", "you should take a look at this website: XY".

Implementation:

- One or multiple posts in the same thread indicate that an idea that originated from the community has been implemented by another or the same community member.
- An implementation can be anything that transports the idea/contribution from the interaction in the online forum to the realm of another community member with the aim of diffusing, testing, modifying, or improving it by means of installation, reproduction, or imitation.
- Examples are: building a machine based on blueprints that were made available to the community, installing new software that was introduced in the thread, running a posted query on a database, applying a newly proposed set of rules to a process.

Modification:

- A product, process, or idea that originated in the community has been modified based on a novel idea by the same person or another member of the community.

- The modification of an idea or process does not require an implementation.
- A modification is not necessarily an improvement.
- Examples for modifications are: building a machine from available blueprints but with different materials or following a different process, implementing a proposed software with a modified algorithm, proposing an altered set of rules from those currently active in the community

Improvement:

- A product, process, or idea previously discussed in the community has been improved by the same or another member of the community by either a) modification of the original or b) by a novel extension ("add on") or replacement.
- The improvement of a process or idea does not require an implementation.
- Example: Correcting an error in a previously presented idea or product is not (necessarily) an improvement. E.g., finding and correcting an error in a posted software code is not an improvement (in the sense of an innovation activity). However, adding a (potentially missing) new feature is.

Additionally, we set the following *general rules* for labelling innovation activities. First, the assessment and subsequent labelling must take into account all posts in the thread. This means, for example, that the presentation of an idea does not necessarily occur in the initial post and that an otherwise "non-innovative" thread can contain "spin-off" sections that do contain innovation activities. Second, information from external sources outside the forum (e.g., linked documents) must not be used to assess the incidence of innovation activities. While external data could enhance the understanding of the "quality" of an idea or its following success, its integration into the automatic calculation of innovation indicators could prove very difficult. However, incorporating data into the labelling process, but not into the labelled training data, would create biases when validating indicators or training ML models.

Initial post

Grosse et al. (2018) define the following six different triggers of initial posts. Their label definitions are exclusive, meaning that a labeller can assign at most one label per post.

- **Needing help:** "Threads started by a user asking for help with a specific problem"
- **Requesting feedback:** "when a user shared a partially developed product or idea with the intention of receiving constructive feedback from the community"
- **Providing feedback:** "posts in which users supplied feedback to a product or feature developed by the community"
- **Sharing developments:** sharing "a product development simply to inform potentially interested community members without formulating a specific request for any community response"
- **Disseminating information:** "If a user opened a thread to disseminate information they thought was interesting without formulating an expectation of a community reaction"
- **Calling for action:** "Threads that were started by users with a specific idea or problem in mind, prompting the community to develop a solution"

We relaxed some requirements to extend the applicability of the labels that Grosse et al. (2018) specifically defined for the OEM community. We decided that the labels "sharing developments" and "disseminating information" can also apply when, instead of "without formulating an expectation [or request]", the intention to disseminate information clearly outweighs the intention to receive a reaction. Furthermore, the label "calling for action" does

also apply if the intention is not to prompt the community to develop a solution, but also if it is more generally prompted to perform other types of activities, such as testing software, building a machine, or performing real-life tasks (collecting and sorting plastic at the beach).

First results & discussion

We collected data on interactions in the peer-innovation communities Precious Plastic (PP), OpenEnergyMonitor (OEM), and OpenStreetMap (OSM) by scraping all posts and author profiles that were publicly available in their online forums[1]. We restricted our analysis to posts that were published in the time period of 01/2017 to 12/2019, in order to exclude potentially biasing founding and shut-down phases and increase comparability (OEM forum started in 2016, PP "Dave Hakkens forum" shut down in 2020; see figure 1). In total, the resulting dataset contained more than 200,000 posts (20,000 threads) by 12,000 users. Our labelling focus primarily lay on technology-oriented sub-forums such as "Machine development" (PP), "EmonCMS" (OEM), "Development" (OSM), or "Garmin Maps" (OSM). We particularly excluded sub-forums that dealt with non-technical (meta-) topics, such as organizing the forum itself.

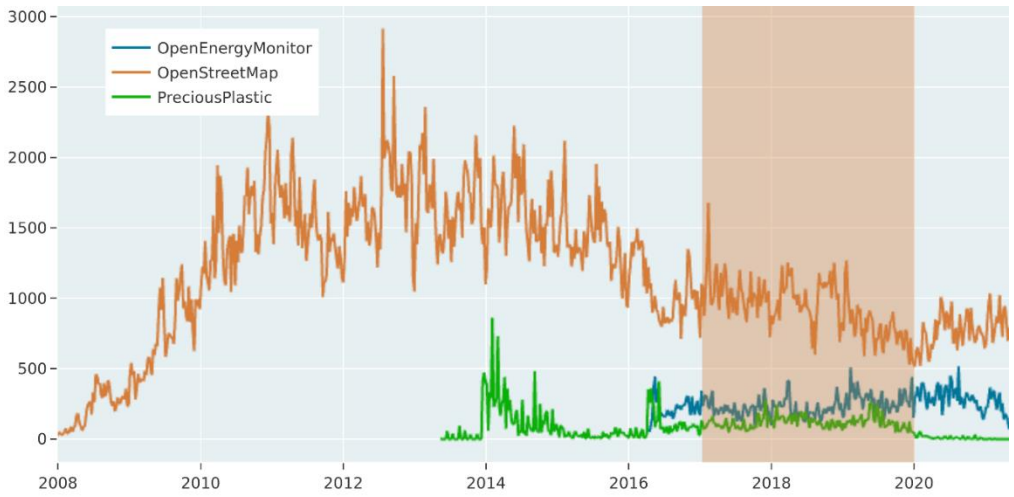


Figure 2: Number of posts per week in the three examined online forums. Orange background: Time period considered in subsequent analyses (01/2017 – 12/2019).

Four innovation researchers (the authors of this paper, none domain experts on matters discussed in the forums) individually labelled a random selection of almost 100 threads from the four communities. As described in the previous section, this was an iterative process, which involved re-defining and concretizing the labelling rules. After a few iterations, we assessed the inter-rater reliability using suitable metrics such as Cronbach's kappa or Krippendorff's alpha (e.g., Zapf et al., 2016). We found that inter-rater agreement between the four raters was mostly "ok" (rule of thumb, α around 0.66) for most labels. The exact agreement per thread (considering all labels) lay between 68% to 84% per labeller pair. Most disagreements stemmed from the application of the label *implementation*, which we found had a rather ambiguous definition and was interpreted differently by the labellers.

[1] https://davehakkens.nl/community/forums/, https://community.openenergymonitor.org/, https://forum.openstreetmap.org/

We were surprised by the high agreement on "innovation potential" (α = 0.75), considering the need for rather subjective evaluations.

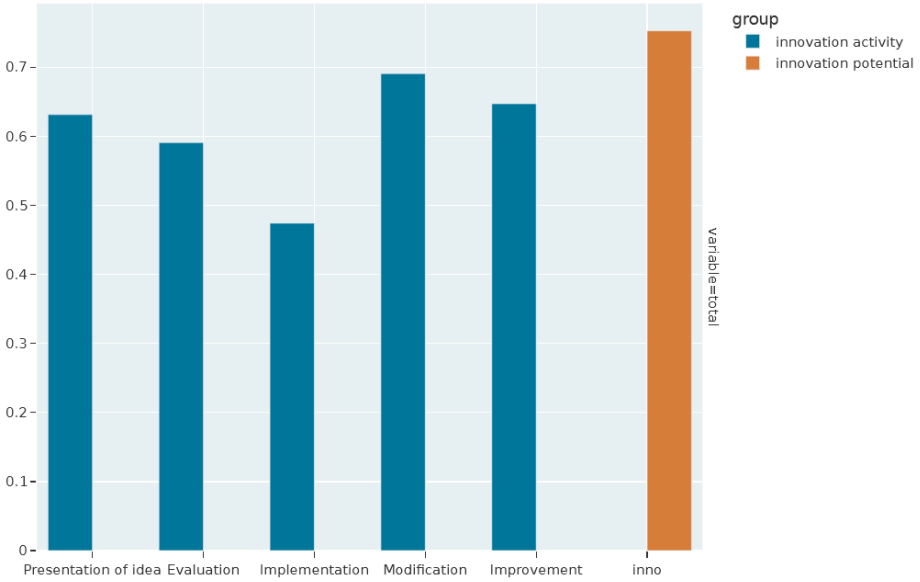


Figure 3: Krippendorff's alpha (inter-rater agreement on labels for innovation activity and innovation potential)

As we considered the disagreement still too high for our purposes, two of the four labellers re-examined all threads with deviating labels. They were able to resolve all disagreements and updated the rules once more to remove ambiguities and inconsistencies.

Overall, we found that labelling innovation activity appeared to be an actionable method and that the definition of labels did not lead to empty or redundant categories. We had the impression that a certain typology of innovation activities emerged, and that there was a correspondence between initial posts and (subsequent) innovation activity, which would be in line with findings from Grosse et al. (2018). We plan to further analyse these aspects after having labelled more data. Our experiences while analysing forum threads furthermore confirmed our decision to label innovation *activities* rather than *innovations*, as we found that in many cases, innovation activities had *not yet* resulted in an innovation outcome, or that outcomes were hard to assess due to diffusion or "materialization" outside the forum.

Labelling initial posts was perceived as much simpler. Initial confusions about the difference between needing help and requesting feedback were resolved by emphasizing that needing help refers to a specific problem (e.g., I have this error, can anybody help me with that?) while requesting feedback can refer to a general topic (e.g., I'm interested in XY, does anybody have experience with that?). The responses to "needing help" contain a solution to a problem, while the responses to a request for feedback contain evaluation, opinions, or experiences.

Our labelling activity was primarily limited by two factors: very long threads and very high knowledge complexity. Especially discussions in the OEM forum were sometimes "too technical" to be completely understood by the labellers, which made it hard to assign labels or led them to completely abandon labelling such cases. We furthermore skipped several threads that we felt were too long to analyse (in some cases 100+ posts). Further analysis should address whether this exclusion leads to any specific biases.

Conclusion

This paper gives a short overview of our approach to measuring peer innovation in online communities, particularly focusing on our rules for manual labelling. We use labelled data to select, validate and potentially train innovation indicators that were previously derived from adjacent innovation literature. After collecting data from online forums of three peer innovation communities, we performed first labelling cycles. An interim analysis of inter-rater agreement and the detailed resolving of disagreements resulted in refined labelling rules.

We plan to continue labelling with a reduced number of labellers to arrive at a set of around 500 to 1,000 labelled threads in total. Our hope (and current impression) is that the labelling process will proceed with increased speed and quality due to the updated rules and better-trained labellers.

We would like to use the expert workshop to discuss our overall approach to measuring peer innovation, to gather feedback on our labelling rules and process, and to gain insights into what experts and members of the peer innovation communities understand by peer innovation.

References

Baldwin, C., & von Hippel, E. (2011). Modeling a Paradigm Shift: From Producer Innovation to User and Open Collabo-rative Innovation. Organization Science, 22(6), 1399–1417. JSTOR.

Bradonjic, P., Franke, N., & Lüthje, C. (2019). Decision-makers' underestimation of user innovation. Research Policy, 48(6), 1354–1361. https://doi.org/10.1016/j.respol.2019.01.020

Dahlander, L., & Frederiksen, L. (2012). The Core and Cosmopolitans: A Relational View of Innovation in User Commu-nities. Organization Science, 23(4), 988–1007. https://doi.org/10.1287/orsc.1110.0673

Franke, N., Schirg, F., & Reinsberger, K. (2016). The frequency of end-user innovation: A re-estimation of extant find-ings. Research Policy, 45(8), 1684–1689. https://doi.org/10.1016/j.respol.2016.04.012

Gambardella, A., Raasch, C., & von Hippel, E. (2016). The User Innovation Paradigm: Impacts on Markets and Welfare. Management Science, 63(5), 1450–1468. https://doi.org/10.1287/mnsc.2015.2393

Gault, F. (2013). Innovation indicators and measurement: An overview. Handbook of Innovation Indicators and Measurement. https://www.elgaronline.com/view/edcoll/9780857933645/9780857933645.00008.xml

Gault, F. (2018). Defining and measuring innovation in all sectors of the economy. Research Policy, 47(3), 617–622. https://doi.org/10.1016/j.respol.2018.01.007

Grosse, M., Pohlisch, J., & Korbel, J. J. (2018). Triggers of collaborative innovation in online user communities. Journal of Open Innovation: Technology, Market, and Complexity, 4(4), 59.

Korbel, J. J., & Grosse, M. (2020). Identification and classification of household sector innovation: the parallel product market and the importance of context and complexity. Innovation, 22(2), 193-215.

Nielsen, K. R. (2020). Policymakers' views on sustainable end-user innovation: Implications for sustainable innovation. Journal of Cleaner Production, 254, 120030. https://doi.org/10.1016/j.jclepro.2020.120030

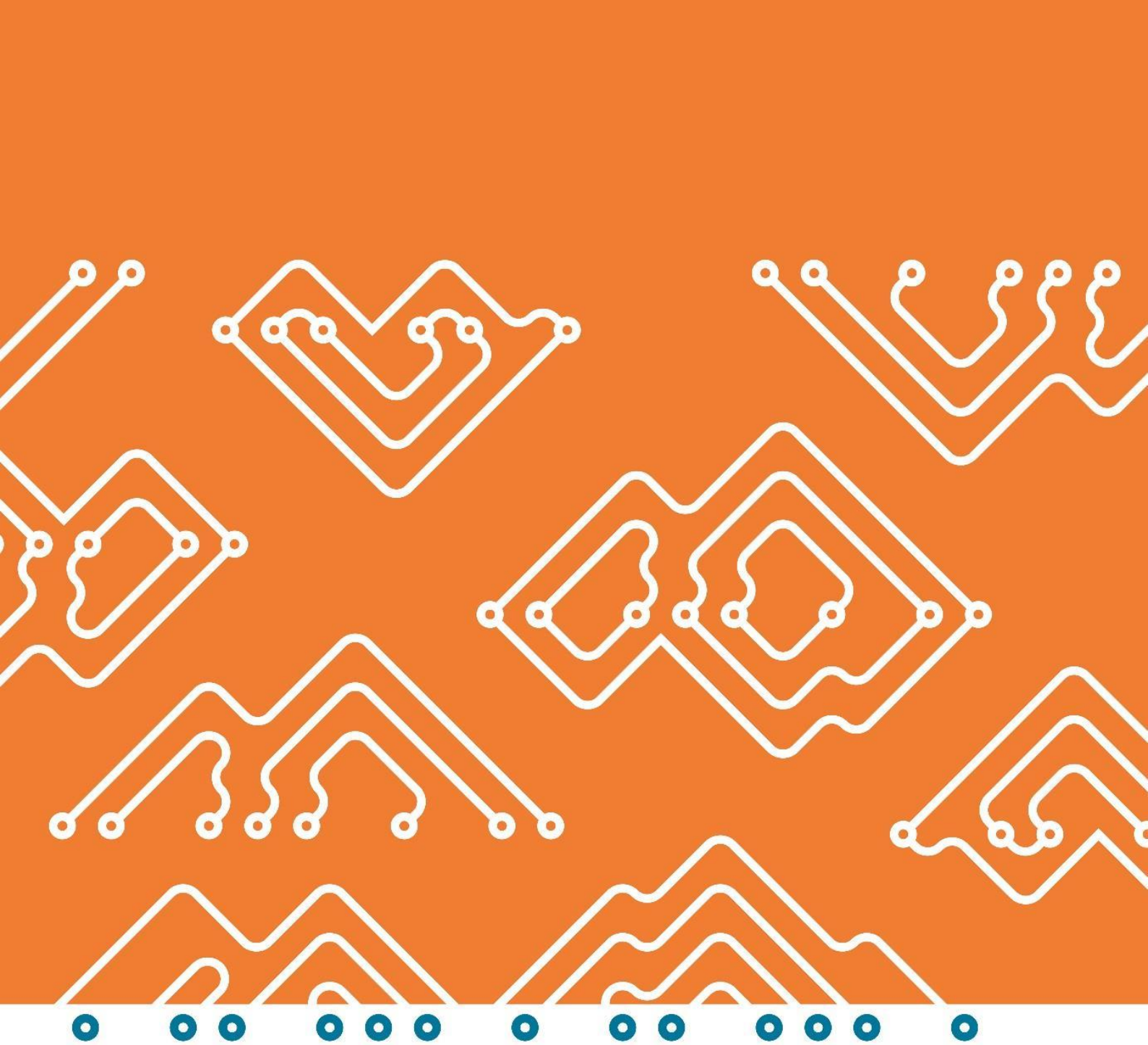
OECD/Eurostat. (2019). Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation. OECD Publishing. https://www.oecd-ilibrary.org/content/publication/9789264304604-en

Peuckert, J., Färber, L., Pohlisch, J., & Kern, F. (2020). Peer Innovation – Beiträge zu Innovation und Nachhaltigkeit (Arbeitsbericht Nr. 1; PeerInnovation). Institut für ökologische Wirtschaftsforschung. https://www.peer-innovation.de/app/download/9037004814/PeerInnovation_Arbeitsbericht_1.pdf?t=1614942321

Pohlisch, J., Peuckert, J. (2021): Peer Innovation – Indikatoren für Innovation in Peer-Communities (Arbeitsbericht Nr. 1; PeerInnovation). Institut für ökologische Wirtschaftsforschung. https://www.peer-innovation.de/app/download/9063905214/PeerInnovation+Arbeitsbericht+2.pdf?t=1623754128

von Hippel, E., de Jong, J. P. J., & Flowers, S. (2012). Comparing Business and Household Sector Innovation in Consumer Products: Findings from a Representative Study in the United Kingdom. Management Science, 58(9), 1669–1681. JSTOR.

Zapf, A., Castell, S., Morawietz, L., & Karch, A. (2016). Measuring inter-rater reliability for nominal data–which coefficients and confidence intervals are appropriate? *BMC medical research methodology*, *16*(1), 1-10.



www.peer-innovation.de