MODU

Mesurer les émOtions des DiscUssions dans les collaborations en ligne

Measuring emotions of discussions in online collaboration

A post-doc project proposal for eNSEMBLE's call for post-doc proposals #2025 round 1. This proposal targets CONGRATS as it aligns with Task 4.2.2 Community health, affect and cohesiveness.

The MODU project aims to measure emotions in online collaborative discussions, focusing on how emotional context influences knowledge production in virtual teams. By analyzing textual data from platforms like Wikipedia and Debian, the project seeks to understand the impact of emotions on cooperation and productivity. It addresses key challenges such as limited contextual information (interactions through text production only), precise emotion measurement in that context, and linking emotional states and expression to knowledge/information creation events. The findings will provide insights for managers to foster positive collaboration and prevent toxic behaviors in online communities as well as fostering the comprehension of Online Open Collaborative Communities (OOCCs). The project combines two broad streams of studies. On the one hand, we rely on managerial and communication studies to theorize overall organizational behaviour in such a realm. On the other hand, we adopt natural language processing approaches as well as mixed methods to gather and analyze data, but also to deliver actionable recommendations for participants for supervising and enhancing virtual team dynamics.

Context, Objectives, Challenges, Approach, Originality, Methods, Anticipated result

Since the crisis of COVID-19, virtual teamwork has become increasingly important. Organizations are leveraging digital infrastructures to facilitate collaboration and knowledge creation. This form of organization is not inherently new; OOCCs have existed on the internet since the 1990s, with notable examples such as Debian and Wikipedia. While these projects have democratized participation in the production of commons, they have also faced challenges such as disinformation, dehumanization, and lack of data privacy. Psychological safety within working teams is crucial for effective collaboration, as emotions significantly affect cooperative efforts (Edmondson, 1999). The research project aims to explore how emotions are experienced and expressed in virtual environments and how they impact team performance and knowledge production.

The primary focus of the project is to understand how emotions influence cooperative efforts and how they are shaped by environmental factors such as social cues and the emotions expressed by others. This aspect is grounded in communication theories, including the General Appraisal Theory (Scherer, 2021), which suggests that expressed emotions result from an individual's evaluation of an external stimulus. The project also considers the Communication Accommodation Theory (CAT) and emotional mimicry theories (Hess & Fischer, 2022), which explore how individuals align their communication styles during interactions.

One of the key challenges the research project aims to address is the adaptation of emotional psychology concepts to the specific setting of online collaborations. Online interactions often lack face-to-face cues, making it difficult to reconstruct the social identity and emotions of participants. The project aims to develop tools to analyze emotions in computer mediated communication (CMC), using only the information available to participants through the corpus of text present in their online environment. This involves evaluating visible text and understanding how past interactions influence the expression and experience of emotions.

Another aspect the project seeks to address is the spectrum of emotions considered in similar studies. While past research has focused on valence (pleasantness) and arousal (excitement) in emotions (Gallus and Bhatia, 2020), this project aims to expand this spectrum by including dimensions such as fear, anger, and happiness. Furthermore, the project will consider codes, registers, and styles typical of CMC, such as emoji or Internet slang, and aims to identify predetermined social behaviors like hate speech through text analysis.

Ultimately, the project seeks to understand how expressed emotions impact the progress of collaborative tasks, such as editing a Wikipedia article or developing software in Debian. This involves creating a dataset with timestamps and developing tools to analyze how emotional "temperature" and social behaviors affect project evolution and quality. In the case of Wikipedia, for instance, we would like to see how the longevity of changes on an Article is linked with the emotions expressed within the annexed discussion page. By linking the online environment, individual emotions, and collective outcomes, the project aims to enhance our understanding of the dynamics behind the collective action problem that characterizes OOCCs and social media platforms. The project also aims to provide organizational and societal benefits by helping online communities better understand and manage discussions. The societal benefits are also represented by the potential opening of the discussion on the actual impartiality that collective negotiation processes, such as those on Wikipedia, may or may not guarantee. This makes the discussion on the reliability of what is produced and readable online increasingly concrete.

The main deliverables include datasets on discussion activities, algorithms for evaluating emotional content, and insights into the relationship between cooperation, emotion, and knowledge production. The results will be published openly, with data and algorithms licensed under open data principles, in line with ANR recommendations and eNSEMBLE data management plan.

Project organisation (feasibility, duration, milestones, partnerships, etc.)

The project is divided into several key tasks, including literature review, technical training, data collection, and analysis. The timeline is structured to facilitate iterative feedback between data and results, with a Gantt chart (see Figure 1 in the Appendix) outlining the major steps. The project is divided into three main parts:

The first part focuses on acquiring and refining the technical and theoretical tools necessary for the project. The intermediary part involves refining research activities and developing initial drafts for the final deliverables. The final part concerns completing the deliverables and disseminating the findings and overall research. This dissemination ideally involves developing collaborations and workshops with practitioners to discuss and refine the understanding of how the findings and managerial tools identified can be effectively used and what impacts they have on day-to-day activities.

The project is not starting from scratch, as an initial paper written with Nicolas Jullien (IMT-Atlantique) has laid the foundations for this type of research. Specifically, the article was written during visiting

periods at the Brest Campus at various times in 2023 and 2024. During this period, contacts and discussions were also initiated with researchers from TAL.

The project has already received support from Région Bretagne (18 months are funded), and this request focuses on securing funding for an additional year to cover environmental costs (computer, travel). The total duration of the post-doc is expected to be 30 months.

At the same time, it is worth mentioning that the project is not standalone but complementary and closely related to other research involving IMT Atlantique and the Venice School of Management (VSM) of Ca' Foscari University of Venice. Another paper is being written with Francesco Rullani (Ca' Foscari) and Nicolas Jullien, focusing on the interactions between Debian and Ubuntu developers. The aim is to understand the impact of the business model on how developers and organizations interact with each other. Such proximity of research interests can, on the one hand, benefit from economies of scope due to overlapping areas in terms of theory and methodology and, on the other hand, lead to closer collaboration between the universities and research centers. Considering the ongoing research in the already existing network, this project can easily develop additional partnerships with partners of PEPR eNSEMBLE such as Nadia Gauducheau, Michel Marcoccia. At the same time, potential partnerships can involve fellow researchers of VSM of Ca' Foscari as well as the Center for the futures of work, information and technology of Syracuse University.

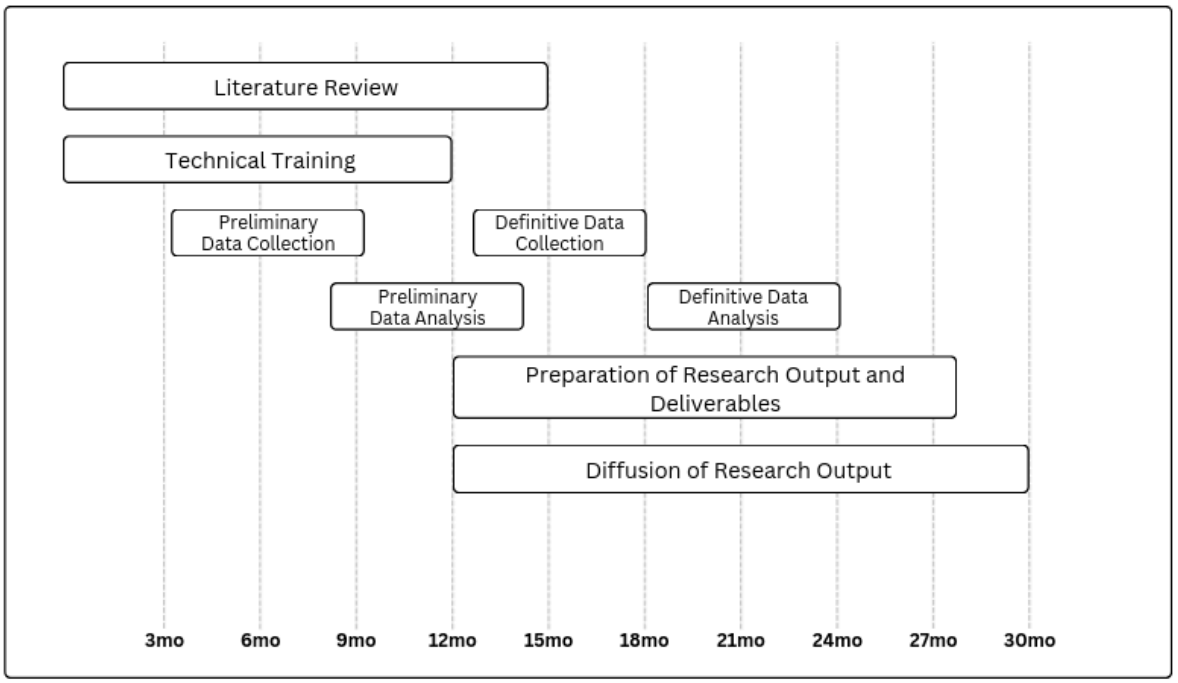
Regarding methods and tools, the project's intention is to deeply connect the socio-psychological mechanisms and dynamics at play in the observed phenomenon. To tell a "thick tale," we intend to adopt a mixed-methods approach. On one hand, the use of ad-hoc software built with Python allows for the application of quantitative methods that are perfectly adaptable to the research project without the need for paid proprietary software. In particular, in the mentioned ongoing research, we utilized Word Frequency Averaging and Word Embeddedness to estimate the emotion expressed in text. Future approaches could broaden the models utilized for such estimates with the aim of achieving a more context-aware and multidimensional understanding of emotion in text. This project will complement the work already conducted by members of the CONGRATS projects at Université Grenoble Alpes (UGA). In particular, it will build upon the sentiment analysis research on Reddit, conducted as part of Ribhu Misra's PhD. Although the goals of the projects are similar, the fields of study differ, and the sentiments analyzed may be more extreme in the Reddit context. But the possibility to exchange techniques and results will help both research projects. The second project looks at emotion detection using facial recognition. . While this project attempts to detect emotions using "rich" sources (the complexity of facial expressions), and thus faces the difficulty of analyzing complex and (too) abundant information, this one aims to detect emotions and social behavior using "poor" and limited sources (though rich in terms of quantity). Understanding if the same kinds of emotions is broadcasted in these different online environments of cooperation will help understand what are the emotional information that are fundamental to ensure cooperation. Secondly, CONGRATS will benefit from sharing the interviews of participants in these different projects, or from mixing the participants of the different platforms in focus groups to better understand the reason why emotions are important to cooperate, and for which tasks. Such a qualitative approach can be developed in a symbiotic relationship with the final phase of the project in which we intend to involve practitioners and individuals that have or had to deal with cooperation within online environments and communities.

Adequacy with the objectives of PEPR eNSEMBLE and PC 4.

This project is directly connected with Task 4.2.2 Community health, affect and cohesiveness of CONGRATS, and will contribute to the second deliverable of this task “Methodological framework for measuring emotions and affect in online community collaboration”.
It is worth stressing that the methodology proposed is compatible with the overarching goal of CONGRATS, that is providing “actionable tools for communities”, in the sense of tools that are built on what participants see and evaluate when they are involved in such projects.
As already mentioned, other research is conducted to assess emotions, something that echoes other researchers’ interest in CONGRATS but also in PILOT. This post-doc also aims at facilitating the coordination of such initiatives, but also at developing collaborations with other researchers from the laboratories participating in eNSEMBLE. We have already identified such possible researchers in the LIST3N at UTT.

Appendix

Figure 1


Literature Review
Technical Training
Preliminary Data Collection
Definitive Data Collection
Preliminary Data Analysis
Definitive Data Analysis
Preparation of Research Output and Deliverables
Diffusion of Research Output
3mo
6mo
9mo
12mo
15mo
18mo
21mo
24mo
27mo
30mo