Today, teamwork has become increasingly virtual: people collaborate over digital infrastructures without meeting face-to-face. Improving virtual team efforts for knowledge production has become strategic for organizations. However, **organizations face challenges in implementing** such **virtual teams** (Edison et al. 2020) due to difficulties in capturing and evaluating the work of team members, signaling **which competencies are needed for which teams and organizational tasks, and identifying and teaming the "best" for the task**.

In voluntary online projects, team members self-select by choosing the task(s) in which they want to participate. Previous research (Barcomb et al. 2019) showed that these team members and project leaders do this by evaluating the other team members based on the observable elements they have access to, mostly previous collaboration. But there is still a lack of methodology that can capture these complex interrelationships of virtual teamwork. This is partly because it requires a pluridisciplinary approach to discuss concepts of management and CSCW (what is a good team, what is a good production, what are the criteria to define these concepts?) with data science (how to mine/represent the collaborators & the collaboration to inform the defined criteria).

This phd proposal, positioned **in the field of computational social science**, aims to address this challenge by providing a framework and the associated measures to evaluate what is a good team and what is good teaming, based only on the information that is accessible in a virtual community.

From a **management science perspective**, the impact of team member diversity on performance and viability is important for groups that collaborate virtually to produce knowledge. “Good” characteristics of the involved team members in terms of individual experiences, skills, and previous collaborative connections matter (Morrison-Smith & Ruiz, 2020). Other team members rely on this information to choose their involvement (Ren et al. 2020), but it is not always present (CVs may not be part of the platform infrastructure, or may not be filled in). So the used information to study teams and teaming usually comes from the previous interactions (available data) and then allow network-based analysis to study team organization. Thus, in terms of team structure, there is an extensive literature describing digital platforms that create a core-periphery structure and their management (Safadi et al. 2021). This structure is characterized by a dense, cohesive core and a sparse, disconnected periphery. This suggests that the core members strongly determine the viability and quality of the knowledge produced, while the peripheral members add new perspectives and pieces of knowledge that are integrated by the core members (Rhyn et al. 2017). However, **the right balance between core and peripheral members to support the viability of a team** to produce a good piece of knowledge is still a subject of research. Finally, on the output side, if research has progressed to define what a good output is and how to measure it, from the knowledge production part (see the recent review by Moás & Lopes, 2023), it still needs to be linked to the characteristics of the good team to do so. Beyond this micro-level of knowledge production, much research focuses on the creation of a team for a specific task, and less on the stability of a team and its capacity to handle several tasks consecutively (what we coin as team health), when Guimera et al. (2005) showed that team stability is key to achieving efficiency and success, with a certain amount of refreshment. But what is the reality of stability? Does the whole team have to stay together, or is it more about the stability of the core members? Are there phases according to the tasks, or envies of its members?

From a **computational science perspective**, it means to understand and describe the dynamics of these communities. As part of the complex networks research domain, many methods have been proposed to detect communities (Rossetti & Cazabet, 2018; Hartmann et al., 2016), while some of them address the problem of tracking their evolutionary behavior (Mohammadmosaferi, K.K. and Naderi, H., 2020). Predicting the future changes and life cycle of communities for time-evolving systems is one of the challenges. The first results in this direction were related to predicting whether a community will grow and/or survive or instead disappear (Kairam et al. 2012; Patil et al. 2013). However, a closely related problem has been very little studied: the analysis of the internal evolution of communities, as has been done for static communities (Dao et al. 2020). For example, structural (e.g., density, cohesion) and temporal features (e.g., join node ratio and left node ratio, activity over time) are used as variables to predict life cycle events such as continuation, shrinkage, growth, and dissolution (Pavlopoulou et al. (2017), Liu et al. (2019)). However, the internal characterization of communities is not studied per se.

To do this, we will rely on Wikipedia, which proposes the largest dataset regarding the task (writing articles) and the team (people who join to participate in this writing). These are organized at the "project" level (science, medicine, etc.), which allows us to discuss the issue of assigning people to concurrent projects. The Wikipedia case study will provide both examples and real data sets to help design concepts and algorithms, but will also bring new insights to management research, particularly with regard to teamwork in online volunteer teams aimed at producing knowledge.

PhD thesis objectives

In this PhD thesis, we aim to explore community dynamics at the micro level, i.e., patterns of relationships within communities. How does internal organization evolve? Are there core-periphery structural patterns? What are their dynamics? How do they aggregate, expand? Are they supported by periodic evolutionary models? **The different stages of the work will be:**

- To provide a state of the art of internal organizational dynamics of communities both in management (criterion to describe a virtual team, its constitution and its outcomes), and complex networks domains;
- To collect and model interactions (co-working) in Wikipedia is the first technical challenge of the work. The project proposes to focus on a Wikipedia "project" (such as economics, still unstudied, or US military history, the most active and best structured);
- To model interactions. This is a non-trivial task in temporal networks. Probably the simplest approach is to define snapshots, corresponding to static graphs, representing the state of the evolving network at a given time. Another way is to treat the dynamics in a stream fashion with link streams;
- To develop new measures of core/periphery structures and of team members and roles in dynamic groups/graphs;
- To evaluate these measures for their ability to predict the ability of teams and teaming to produce good outcomes in the context of Wikipedia.

Indicative Bibliography (in bold research conducted by the supervisors)

Barco**mb, A., Jullien, N., Meyer, P., Olteanu, A. L. (2019). Integrating managerial preferences into the qualitative multi-criteria evaluation of team members. In Multiple criteria decision making and aiding (pp. 95-143).**

Borgatti, S. P., Everett, M. G. (2000). Models of core/periphery structures. Social networks, 21(4), 375-395.

Dao, V.L., Bothorel, C., Lenca, P. Community structure: A comparative evaluation of community detection methods. Network Science, 2020, 8 (1), pp.1-41.

Guimera, R., Uzzi, B., Spiro, J., Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. Science, 308(5722), 697-702.

Edison, H., Carroll, N., Morgan, L., & Conboy, K. (2020). Inner source software development: Current thinking and an agenda for future research. Journal of Systems and Software, 163, 110520.

Hartmann, T., Kappes, A., & Wagner, D. (2016). Clustering evolving networks. Algorithm engineering: Selected results and surveys, 280-329.

Kairam, S.R., Wang, D.J., Leskovec, J. (2012). The life and death of online groups: Predicting group growth and longevity. Proceedings of the Fifth ACM International Conference on Web Search and Data Mining. pp 673–682.

Liu, W., Saganowski, S., Kazienko, P., Cheong, S.A. (2019). Predicting the Evolution of Physics Research from a Complex Network Perspective. Entropy. 2019; 21(12):1152.

Moás, P. M., Lopes, C. T. (2023). Automatic Quality Assessment of Wikipedia Articles-A Systematic Literature Review. ACM Computing Surveys.

Mohammadmosaferi, K.K. and Naderi, H. (2020). Evolution of communities in dynamic social networks: An efficient map-based approach. Expert Systems with Applications 147.

Morrison-Smith, Sarah, Jaime Ruiz (2020). Challenges and barriers in virtual teams: a literature review. SN Applied Sciences 2: 1-33.

Patil, A., Liu, J., Gao, J. (2013). Predicting group stability in online social networks. In: Proceedings of the 22nd International Conference on World Wide Web (WWW'13). pp 1021–1030.

Pavlopoulou, M. E. G., Tzortzis, G., Vogiatzis, D., & Paliouras, G. (2017). Predicting the evolution of communities in social networks using structural and temporal features. In 2017 12th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP) (pp. 40-45). IEEE.

Rhyn, Marcel, Ivo Blohm, and Jan Marco Leimeister. "Understanding the Emergence and Recombination of Distant Knowledge on Crowdsourcing Platforms." ICIS. 2017.

Rossetti, G., Cazabet, R. (2018). Community discovery in dynamic networks: A survey. ACM Comp. Surv., 51(2): 1–37.

Safadi, H., Johnson, S. L., Faraj, S. (2021). Who contributes knowledge? Core-periphery tension in online innovation communities. Organization Science, 32(3), 752-775.

Ren, R., Yan, B., Jian, L. (2020). Show me your expertise before teaming up: sharing online profiles predicts success in open innovation. Internet Research, 30(3), 845-868.