The Journal of Systems and Software 208 (2024) 111878

Contents lists available at ScienceDirect

The Journal of Systems & Software

journal homepage: www.elsevier.com/locate/jss

An Empirical Investigation Into the Influence of Software Communities' Cultural and Geographical Dispersion on Productivity☆

Stefano Lambiase [a,*], Gemma Catolino [b], Fabiano Pecorelli [b], Damian A. Tamburri [b], Fabio Palomba [a], Willem-Jan van den Heuvel [b], Filomena Ferrucci [a]

[a] SeSa Lab - University of Salerno, Italy
[b] Jheronimus Academy of Data Science, The Netherlands

ARTICLE INFO

Keywords:
Global Software Engineering
Social aspects in software engineering
Productivity
Empirical software engineering

ABSTRACT

Estimating and understanding software development productivity represent crucial tasks for researchers and practitioners. Although different works focused on evaluating the impact of human factors on productivity, a few explored the influence of cultural/geographical diversity in software development communities. More particularly, all previous treatise addresses cultural aspects as abstract concepts without providing a quantitative representation. Improved knowledge of these matters might help project managers to assemble more productive teams and tool vendors to design software analytics toolkits that may better estimate productivity. This paper has the goal of enlarging the existing body of knowledge on the factors affecting productivity by focusing on *cultural and geographical dispersion* of a development community—namely, how diverse a community is in terms of cultural attitudes and geographical collocation of the members who belong to it. To reach this goal, we performed a mixed-method empirical study. First, we built a statistical model relating dispersion metrics with the productivity of 25 open-source communities on GitHub. Then, we performed a confirmatory survey with 140 practitioners. The key results of our study indicate that cultural and geographical dispersion considerably impact productivity, thus encouraging managers and practitioners to consider such aspects during all the phases of the software development lifecycle. We conclude our paper by elaborating on the main insights from our analyses and instilling implications that may drive further research.

1. Introduction

Software productivity refers to the efficiency and effectiveness of software development teams in creating, testing, and deploying software systems (Scacchi, 1995). High productivity levels typically lead to faster time-to-market, better quality software, and greater customer satisfaction (Behutiye et al., 2017; Cardozo et al., 2010). Nonetheless, assessing software productivity is challenging: while different productivity indicators have been proposed, *e.g.*, lines of code written or bugs fixed per time unit (Wagner and Ruhe, 2018; Oliveira et al., 2018; Petersen, 2011), the perception and feeling of productivity can vary among team members (Girardi et al., 2021), *e.g.*, a developer may feel highly productive when accomplishing difficult tasks rather than based on the time spent on coding.

Global Software Engineering (GSE) (Richardson et al., 2010; Cherry and Robillard, 2004)—*i.e.*, the set of practices and guidelines aimed at managing distributed teams—further challenges the estimation of development productivity. Indeed, in such a distributed scenario additional social and cultural aspects come into play (Herbsleb and Moitra, 2001; Mockus and Herbsleb, 2001). For instance, a sub-optimal management of the heterogeneity of development teams may cause heavier inter-team communication and misunderstanding among team members (Elbert, 2010; Cherry and Robillard, 2004; Casey and Richardson, 2008) that altogether negatively affect software communities' productivity (Murphy-Hill et al., 2019; Wagner and Ruhe, 2018; Mohagheghi and Conradi, 2007; Boehm et al., 2000; Graziotin et al., 2015; Vasilescu et al., 2015a).

The GSE research community has been conducting multiple studies targeting the relation between social and human factors on productivity (Machuca-Villegas et al., 2020, 2021, 2022; Gorla and Lam, 2004), other than the underlying reasons behind miscommunication, *e.g.*, community smells (Tamburri et al., 2019a; Palomba et al., 2021) and gender diversity (Catolino et al., 2019b,a). While these previous works have provided compelling evidence of how productivity can

☆ Editor: Raffaela Mirandola.

* Corresponding author.
E-mail address: slambiase@unisa.it (S. Lambiase).

https://doi.org/10.1016/j.jss.2023.111878
Received 8 March 2023; Received in revised form 5 October 2023; Accepted 12 October 2023
Available online 27 October 2023

be affected by social and human aspects, our research highlights a noticeable lack of knowledge of the role played by *cultural and geographical dispersion*, which are factors that can significantly impact how individuals perceive and measure productivity. Some cultures indeed emphasize teamwork, while others value individual effort (Hofstede et al., 2005; Hofstede, 2011): these differences may lead to variations in how team members perceive productivity and their overall effectiveness. Furthermore, geographically dispersed teams may experience concerns due to communication barriers and time zone differences, which might lead to inaccurate performance evaluations and hinder team effectiveness.

An improved understanding of the impact of cultural and geographical dispersion on productivity may lead to significant improvements in multiple project management knowledge areas and procedures (Project Management Institute, 2021). For example, knowing more about the dispersion of a development team since the initiation processes could lead to better risk management planning and a consequent more robust risk register, mitigation, and contingency strategies. Consequentially, better risk management knowledge positively impacts all the other knowledge areas—*e.g.*, cost, time, and communication—leading to a general improvement of the overall software development process.

In our previous empirical study (Lambiase et al., 2022a), we took a first step toward this direction by conducting a quantitative investigation on the relation between the cultural and geographical dispersion of a community—*i.e.*, the degree to which a community is formed by individuals growing up in and coming from different places globally—and its productivity, as measured by the number of commits in a specific time range (Lambiase et al., 2022a). We represented culture using the *Hofstede 6D Model*, namely, a framework that uses six measures to uniquely represent the culture of individuals based on their origin country, widely known and adopted for various work in the software engineering field (Borchers, 2003; Casey, 2011; Abufardeh and Magel, 2010; Venkateswaran and Ojha, 2019). We built a statistical regression model to assess the relationship between the two dispersion metrics and productivity, conducting the study on 25 open-source communities on GitHub. The results of our analysis showed that dispersion metrics influence developers' productivity both positively and negatively.

In this paper, we extend our previous work (Lambiase et al., 2022a) by complementing the statistical study with a *qualitative* empirical examination of the perceived value and impact of dispersion metrics over the productivity of software communities, hence assessing the overall goals of our study through *mixed-method research* (Johnson et al., 2007). More particularly, we complemented the statistical modeling findings with a survey study involving 140 participants with experience in distributed software development. By means of the survey study, we could first corroborate the statistical investigation with more qualitative insights aiming at understanding the value of cultural and geographical dispersion for the overall productivity of software communities and strengthen our findings. Secondly, we could also elaborate and further analyze a potential threat to the validity of the previous study, *i.e.*, we estimated productivity through the number of commits within a time range; however, developers might perceive productivity differently, hence influencing the conclusions drawn in our preliminary study.

In the first place, our findings confirm the results achieved in our preliminary investigation (Lambiase et al., 2022a): cultural and geographical dispersions sensibly impact a software development community's productivity, but not necessarily in a negative manner. Depending on the specific dispersion metrics considered in our study, we could draw implications for productivity. Our survey participants also confirmed the soundness of the design decisions taken in the statistical modeling exercise: the number of commits per time range is perceived as a valid proxy to estimate the actual community productivity. As a further result, both the survey and the statistical model confirm that socio-technical factors—*e.g.*, team size and truck factor—influence the productivity of a software development community.

To sum up, our work provides the following contributions:

1. Statistical insights into the role of dispersion metrics on software productivity;
2. A large-scale confirmatory survey study involving 140 developers to qualitatively validate the results of the statistical model and strengthen them;
3. An online replication package (Lambiase et al., 2023) publicly available to support replication and future work.

Our work finally offers and discusses a number of insights and implications to researchers, practitioners, and tool vendors, on how to effectively exploit cultural and geographical metrics to improve the overall productivity of software teams.

Structure of the paper. Section 2 describes the background of this work and the related work. In Section 3, we present and outline the design of our study, and Section 4 reports the study results. Section 5 discusses the insights of the paper, and Section 6 examines the threats to the validity and how we mitigated them. Finally, Section 7 concludes the paper and provides insights on our future research agenda.

2. Background and related work

This section describes the background and related work that is the foundation for our contributions.

2.1. Cultural aspects in software engineering

Nowadays, software development is often a geographically distributed effort involving stakeholders and practitioners collaborating from different places worldwide (Herbsleb and Moitra, 2001; Marinho et al., 2018; Mockus and Herbsleb, 2001). For such a reason, *Global Software Engineering* (GSE)—*i.e.*, the application of software engineering practices for managing and developing software distributed projects (Richardson et al., 2010; Marinho et al., 2018; Stray and Moe, 2020; Noll et al., 2011)—and the associated research field are becoming even more popular. Specifically, in the large set of topics the research community discusses, the social impact of such *"dispersion"* on product and project metrics is of particular interest. In particular, among the various problems, there is a growing interest in characterizing cultural aspects and their potentially catastrophic impact if not correctly managed (Marinho et al., 2018; Noll et al., 2011; Deshpande et al., 2010; Shah et al., 2012).

Culture has been defined by Kreitner et al. as a set of granted assumptions about how to act and think that characterize a community of individuals (Kreitner et al., 1999). Indeed, culture is a complex topic and is difficult to formalize and measure, so most of the work in literature treats it in an abstract manner (Deshpande et al., 2010). Nevertheless, having a tool to measure culture quantitatively is beneficial to conduct research (Furnham, 2012). For such a purpose, various frameworks and tools arise to allow researchers and practitioners in various fields to measure and represent the culture of individuals (Hofstede, 2011; Hall, 1989; Hampden-Turner et al., 2020). Most of them consisted of a representation based on a set of numerical values—called *dimensions*—whose combination uniquely characterizes the cultural behavior of an individual or a group of people. For example, Hampden-Turner et al. (Hampden-Turner et al., 2020) represented culture using three layers, *i.e.*, *explicit culture*, *norms and values*, and *assumption about existence*.

Of particular interest to our work is the framework proposed by Geert Hofstede, *i.e.*, the *Hofstede 6D Framework*, a set of six dimensions that assume values from zero to one hundred and which combination characterizes groups of individuals from a specific country globally (Deshpande et al., 2010; Shah et al., 2012; Richardson et al., 2010; Stray and Moe, 2020). The six Hofstede's dimensions are defined as follows:

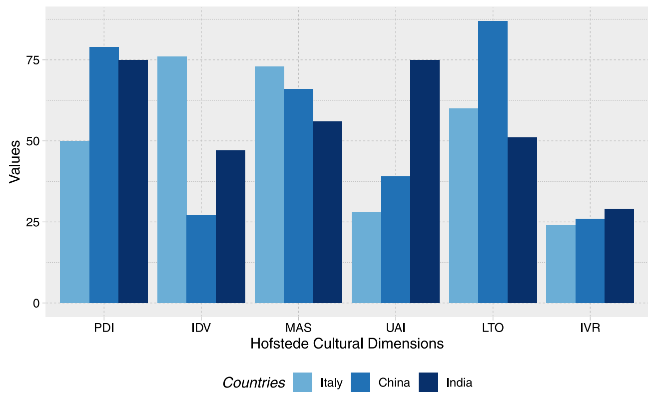

Fig. 1. Hofstede's cultural dimensions values computed for Italy, China, and India.

Power Distance Index (*PDI*). It refers to the degree of inequality that exists and is accepted between people with and without power in a community. A high score of *PDI* indicates that society accepts a hierarchical order in which everybody has a determined place. On the contrary, in societies with low *PDI*, people seek to equalize the power distribution between all community members.

Individualism vs. Collectivism (*IDV*). It represents the degree to which people in a society are integrated into groups and their perceived obligations and dependence on groups. A high level indicates a society where individuals are expected to care for only themselves and their immediate families. Conversely, a low level indicates a society in which people are supposed to be loyal to their group.

Masculinity vs. Femininity (*MAS*). It represents a contrast between the two preferences. The Masculinity side indicates a preference for achievement, heroism, assertiveness, and material rewards for success. In contrast, the Femininity side represents a preference for cooperation, caring for the weak, and quality of life.

Uncertainty Avoidance (*UAI*). It expresses the degree to which the members of a society feel uncomfortable with uncertainty and ambiguity. A high level of *UAI* indicates that people tend to (1) maintain rigid codes of belief and behavior and (2) are intolerant of unorthodox behavior and ideas. Conversely, a low level of *UAI* indicates societies that maintain a more relaxed attitude in which practice counts more than principles.

Long vs. Short Term Orientation (*LTO*). It measures how much people are oriented toward a long-term vision of life than a short-term one. A high score of *LTO* indicates that people emphasize persistence, perseverance, and long-term growth. On the contrary, a low score of *LTO* indicates that people emphasize quick results and respect for tradition.

Indulgence vs. Restraint (*IVR*). It refers to the extent and tendency of a society to fulfill its members' desires. A high level indicates that society allows relatively free gratification related to enjoying life and having fun. Conversely, a low level indicates a society that controls the gratification of needs and regulates it using strict social norms.

To better contextualize the Hofstede framework, Fig. 1 shows an example of the dimensions computed for three countries, *i.e.*, China, Italy, India.[1] For example, looking at the IDV score value of Italy, we can notice that the score is 75, representing a society in which individualism tend to be preferred to collaboration. On the contrary, China has a score of 28, meaning people act in the interests of the group and not necessarily of themselves. Moreover, analyzing the UAI scores, we can see that Italy and China have a value of 30 and 40, indicating that people in such countries tend to be more tolerant of risky situations and changes. In contrast, India has a value of 75, revealing a more traditional society.

Despite the diffuseness of this framework, the research community had mixed opinions on its validity (Roberts and Boyacigiller, 2012; Ailon, 2008; Baskerville, 2003) supporting in some cases its rejection (Brewer and Venaik, 2012, 2014; Sorge, 1983). For instance, Brewer and Venaik (2012, 2014) deemed unreliable the tool's ability to represent cultural profiles. Nonetheless, follow-up studies highlighted the potential of the framework; for example, Venkateswaran and Ojha (2019) showed that Hofstede's framework represents the most effective way to characterize the complex world of cultural aspects which has been already shown effective in several fields (Borchers, 2003; Abufardeh and Magel, 2010; Casey, 2011), *e.g.*, management, law, politics, ethics, architecture, medicine, and computer science (Hofstede, 2017).

Some works studied the use of Hofstede in the context of software engineering (Darwish and Henryson, 2019; Borchers, 2003). Recently, Darwish and Henryson (2019) published a thesis that studied how cultural background influences adopting a specific SE practice (*e.g.*, documentation design, refactoring, test driven development). They reported that developers with similar behaviors related to cultural dimensions (*i.e.*, PDI and UAI) tend to adopt similar practices (*e.g.*, making early design decisions and test driven development). In another work, Borchers (2003) conducted research on how cultural factors influence software engineering processes, such as code review. The study specifically examined three countries: Japan, India, and the United States. The findings revealed that various cultures approach software engineering processes differently. For instance, Japanese developers exhibit a significant level of UAI, which leads to slower decision-making. Borchers also emphasized that cultural differences within software teams can affect software architecture, suggesting further exploration of this topic in the field.

2.2. Productivity factors

Productivity is a complex concept to define and measure. Nevertheless, various metrics arise (Sadowski and Zimmermann, 2019; Mockus et al., 2002; Hernández-López et al., 2013; Boehm et al., 2000; Oliveira et al., 2020), and researchers agree that productivity measures should be expressed in terms of output produced in a given time given a specific input (Mockus et al., 2002; Oliveira et al., 2020). For example, some works define the productivity of a development community as the number of accomplished contributions by team members—*e.g.*, commits, push, or tasks completed in a given unit of time—for the entire project duration (Adams et al., 2009; Sornette et al., 2014; Oliveira et al., 2020).

Additionally, researchers investigated which factors influenced the productivity of a software development team (Murphy-Hill et al., 2019; Wagner and Ruhe, 2018; Mohagheghi and Conradi, 2007; Boehm et al., 2000; Graziotin et al., 2015; Vasilescu et al., 2015a; de Lemos Meira et al., 2010). From our review, we can divide these studies based on the metrics found relevant, *i.e.*, technical and social:

Technical Factors. Regarding technical factors, *i.e.*, product metrics and tools, Wagner and Ruhe (2018) conducted a systematic review, showing how metrics like code reuse, software size,

[1] Hofstede 6D Model website: https://www.hofstede-insights.com/country-comparison/china,india,italy/.

Table 1
Comparison with related work.

Related work	Main focus	Differences
Wagner and Ruhe (2018)	It is a systematic literature review showing how product metrics such as code reuse, software size, and programming language highly impact developers' productivity.	• No focus on social metrics has been put in the investigation. • No quantitative or qualitative methods have been used to validate the findings.
Graziotin et al. (2015)	They performed an observational study with eight participants to study the influence of affecting dimensions on self-assessed productivity.	• Paper did not investigate cultural and dispersions metrics. • Authors did not use quantitative methods.
Vasilescu et al. (2015a)	They performed a mixed-method approach study to investigate the influence of gender and tenure diversity on team productivity.	• Paper did not investigate cultural and dispersions metrics. • Authors did not focus on productivity measures.
Murphy-Hill et al. (2019)	They surveyed practitioners from three private companies to identify factors that can influence self-assessed productivity.	• Paper did not investigate cultural and dispersions metrics. • Authors did not focus on productivity measures. • Authors did not use quantitative methods.
Darwish and Henryson (2019)	They studied how cultural background influences adopting a specific SE practice aiming to provide insights that allow managers to take advantage of each culture's strengths.	• The context of the study was limited to Indonesian and Sweden developers. • The authors did not study the impact of culture on productivity.
Borchers (2003)	It is a report and analysis of the author's past experience in managing software teams composed of people from Japan, India, and the United States. The goal was to study how different cultures approach software development phases.	• The author used only three of the six Hofstede Dimensions to characterize culture. • The context of the study was limited to Japan, India, and the United States developers. • The authors did not study the impact of culture on productivity.

and programming language highly impact developers' productivity. Finally, Mohagheghi and Conradi (2007) investigated the relationship between productivity and software reuse, showing positive results.

Social Factors. Regarding social factors—mainly related to people and their relations—Murphy-Hill et al. (2019) surveyed practitioners, showing how social factors—*e.g.*, people's enthusiasm, peer support, and valuable feedback about job performance—strongly affect people's productivity. Moreover, Wagner and Ruhe (2018) demonstrated that social factors like corporate culture and working environment are essential to enhance software teams' productivity. Graziotin et al. (2015) showed how valence and dominance dimensions in developers affect self-assessed productivity. Finally, Vasilescu et al. (2015a) demonstrated how gender and tenure diversity are good predictors of productivity using a statistical model. Moreover, they adopted a mixed-method approach—*i.e.*, data mining and surveys—to strengthen their findings.

More recently, Murphy-Hill et al. (2019) studied the concept of *self-rated productivity* (Ramírez and Nembhard, 2004) and the factors able to influence it. Specifically, they conducted a survey with practitioners from private companies in IT fields—*i.e.*, Google, ABB, and National Instruments—to identify and study the aspects that practitioners could use to predict the productivity of a software development team. Their findings suggest that the factors most useful to predict productivity are not technical but social ones—*e.g.*, job enthusiasm, peer support for new ideas, and receiving helpful feedback about job performance.

With respect to the works discussed so far, we identified multiple differences and research gaps. Table 1 summarizes the major characteristics of the related work and points out the main research gaps that our study aims at filling.

Graziotin et al. (2015) and Vasilescu et al. (2015a) conducted investigations on social and human factors on productivity, not considering cultural and geographical indicators. Our study can be therefore seen as complementary, as it enlarges the body of knowledge with respect to these previous papers.

As for Murphy-Hill et al. (2019) and Wagner and Ruhe (2018), they conducted more general survey studies which did not have the explicit goal of analyzing the impact of culture. In other terms, the findings reported on the matter can be considered tangential and not comprehensive. Moreover, our mixed-method investigation allowed us to investigate cultural and geographical dispersion under different perspectives, hence strengthening the conclusion validity and generalizability of the findings.

Perhaps more importantly, the authors treat cultural aspects in an abstract and high-level manner without operationalizing them. On the contrary, we exploited Hofstede's 6-D framework (Hofstede et al., 2005), which was explicitly defined to represent cultural dimensions quantitatively.

Last but not least, Murphy-Hill et al. (2019) assessed *self-rated productivity* (Ramírez and Nembhard, 2004), while we experimented with the number of commits per time range. This metric was considered a valid proxy for productivity by the practitioners involved, hence potentially representing an additional insight provided by our work with respect to how productivity can be estimated.

3. Research study design

This section describes the research questions and the methods used to achieve the study's primary objective.

3.1. Research questions and goals

The *goal* of the study is to analyze the relationship between the *cultural and geographical dispersion* of a development community and its productivity, computed by counting the number of commits in a range of time (Mockus et al., 2002). The *purpose* is to increase awareness and allow practitioners to make more informed decisions based on their software development community. The *perspective* is of managers interested in effectively allocating resources, adhering to the project's requirements, or managing/monitoring complex organizational structures.

Although causation requires more than simple regression and correlation—the principal aim of this contribution—our work started from the hypothesis that the productivity of a development community may be influenced by its cultural and geographical dispersion. Fig. 2 depicts a potential cause–effect construct model on how dispersion metrics and productivity could be related. When a community is culturally diverse and spread across different geographical locations, it can harness a rich pool of perspectives and skills. However, this dispersion can

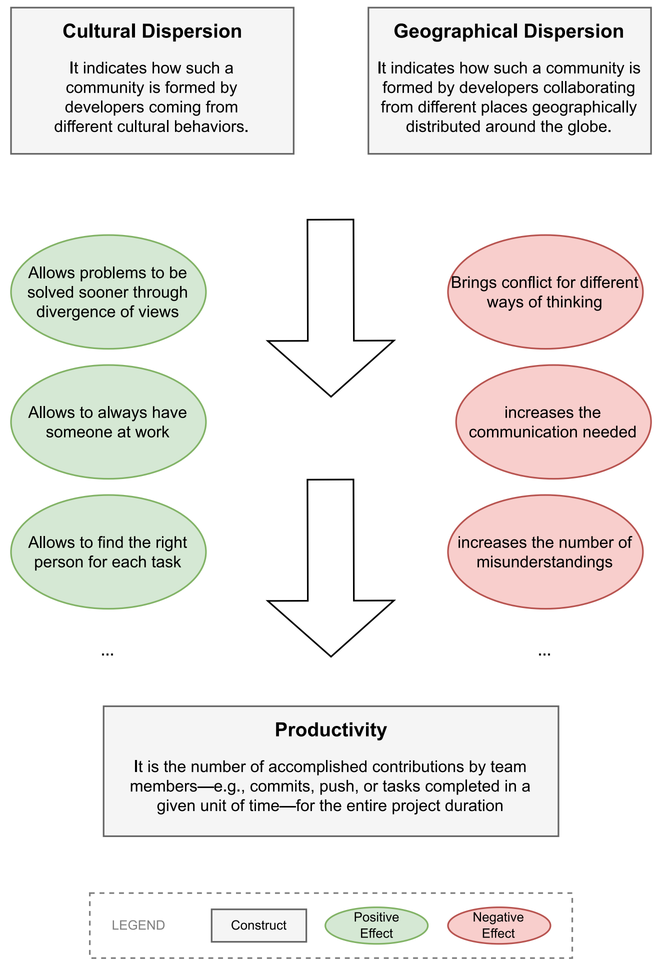

Fig. 2. Cause-effect construct example.

also pose challenges in terms of communication and collaboration; sub-optimal management of the heterogeneity of development teams may cause heavier inter-team communication and misunderstanding among team members that altogether negatively affect software communities' productivity.

The goals of the empirical study were mapped onto the following research question.

> **RQ—Dispersion Metrics Versus Productivity**
>
> *To what extent do cultural and geographical dispersion influence teams' productivity?*

Fig. 3 overviews the methods employed to address our research question. We adopted a mixed-method research approach (Johnson et al., 2007) in which both qualitative and quantitative studies are performed to reach theoretical saturation. Specifically, in our research, we performed the following studies:

Quantitative Investigation: we built a statistical model—a mixed-linear regression model (Lindstrom and Bates, 1988)—able to assess whether cultural and geographical dispersion relates to the productivity of open-source development communities.

Qualitative Investigation: we surveyed 140 practitioners with experience in distributed software development to gather their opinion on how different culturally originated behaviors in software teams can influence their productivity.

Table 2
Projects in the dataset.

Project	Progr. Language	# Windows
Akretion	Python	6
Bigcheese	C++	1
Burke	Go	5
Chapuni	C++	9
Cloudfoundry	Shell	7
CTSRD-CHERI	C++	2
Django	Python	23
Emberjs	Python	7
Fangism	C++	1
Genome	Perl	5
Holman	C	7
Jedi4ever	Shell	8
Jrk	C++	1
Liferay	Java	12
Loganchien	C++	1
Moodle	PHP	14
Mozilla - gecko-dev	C++	1
Mozilla - OpenBadger	Javascript	2
Mxcube	Python	2
Puppetlabs	Ruby	14
RobbyRussel	Python	15
Rspec	Ruby	13
Symfony	Python	13
Torvalds	C	17
Travis-ci	Javascript	10

In terms of reporting, we employed the guidelines by Wohlin et al. (2012), other than following the *ACM/SIGSOFT Empirical Standards*.[2] Moreover, a replication package of the study is available online (Lambiase et al., 2023).

3.2. Design of the quantitative study: Regression model

The first step of our study consisted of a statistical analysis analyzing the relationship between dispersion metrics—*i.e.*, cultural and geographical dispersion—and the productivity of software development communities. In the following, we provided information about our quantitative investigation.

3.2.1. Data collection

To conduct our study, we used the dataset from our previous study (Lambiase et al., 2022b) containing socio-technical metrics about 25 open-source software communities. Specifically, the dataset contains information for different time windows, made of 90 days. Therefore, we had information in various time slices for each software development community. Table 2 reports the list of the projects and the number of windows considered for each of them in our study.

Within the dataset, the most valuable metrics are represented by software communities' cultural and geographical dispersion indicators. As for cultural dispersion, the dataset contains six cultural metrics that can assume values from zero to fifty. Each metric corresponds to the standard deviation of the set containing the community members' value for one of the six dimensions of Hofstede (Hofstede et al., 2005). As for the geographical dispersion, the metric considered the standard deviation of the spherical distances (in miles) between each community member as a metric—computed using the GeoPy[3] library. To compute both metrics, we relied on the original country of the developers in the development communities, which was provided in the original dataset (Catolino et al., 2019b; Vasilescu et al., 2015b)—more details in the following section.

[2] Available at the following link: https://github.com/acmsigsoft/EmpiricalStandards. Given the nature of our study and the currently available standards, we followed the *"General Standard"*, *"Questionnaire Surveys"*, *"Mixed Methods"*, and *"Data Science"* definitions and guidelines.

[3] https://geopy.readthedocs.io/en/stable/index.html

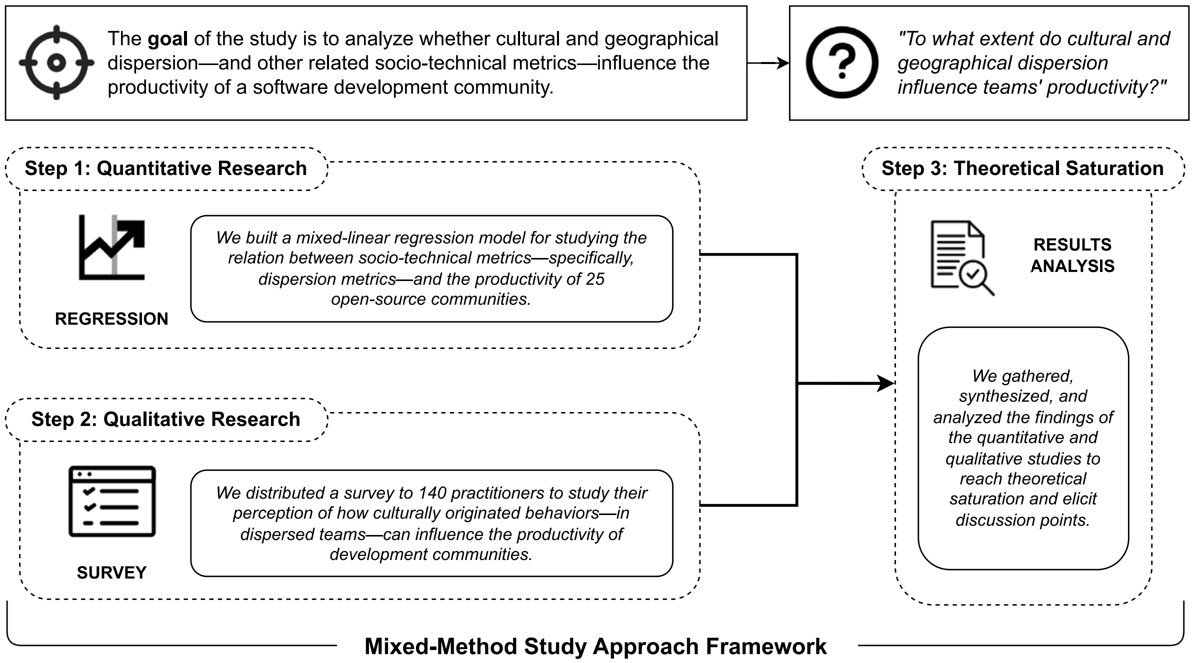

Fig. 3. Overview of the research methods used to address the research questions of the study.

3.2.2. Statistical model variables

To answer our research question, we built a statistical *linear mixed regression model* (Lindstrom and Bates, 1988) relating a development community's cultural and geographical dispersion to productivity—expressed in terms of the number of commits.

Independent variables. In the context of our study, we considered the following independent variables.

Cultural Dispersion. The *cultural dispersion* of a development community indicates how such a community is formed by developers coming from different cultural behaviors (Lambiase et al., 2022b; Tamburri et al., 2019b). Being a quantitative measure, it is necessary to represent the culture of individuals using a quantitative framework; hence, we used the six metrics in Hofstede's framework (Hofstede et al., 2005)—described in Section 2. By such a choice, we defined six cultural dispersion metrics—one for each Hofstede dimension—corresponding to the standard deviation of the set containing the Hofstede community members' values. The metrics are reported in the following:

- **PDID**: *Power Distance Index Dispersion* indicates how much community members tend to have a different idea of how power should be distributed between them.
- **IDVD**: *Individualism vs. Collectivism Dispersion* indicates how much community members tend to have different ideas regarding working in a group and sharing success or being individualistic.
- **MASD**: *Masculinity vs. Femininity Dispersion* indicates how much community members tend to have a different opinion about self-affirmation and help the weaker elements.
- **UAID**: *Uncertainty Avoidance Dispersion* indicates how much community members tend to have different ideas on taking risks and accepting new and controversial opinions.
- **LTOD**: *Long Term Orientation vs. Short Term Orientation Dispersion* indicates how much community members tend to have a different opinion about investing or not in the future and conserving old traditions and habits.
- **IVRD**: *Indulgence vs. Restraint Dispersion* indicates how much community members tend to have a different opinion about the rank in which the governing authority controls how people satisfy their needs and spend leisure.

Geographical Dispersion. The *geographical dispersion* of a development community indicates how such a community is formed by developers collaborating from different places geographically distributed around the globe. As shown in Eq. (1), we operationalized the geographical dispersion of a community as the standard deviation of the set containing the physical distances—expressed using spherical distance—between each community member (Tamburri et al., 2019b; Li et al., 2010).

$$GeographicalDispersion(X) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_X)^2} \tag{1}$$

where X is the set containing the physical distance between each pair of community members, N is the number of elements in X, X_i is the element of index i in X, and μ_X is the mean of X.

To provide an example for better explaining this metric, given a development community: (1) if all the members are working from different and distant locations, there is a high level of geographical dispersion; (2) if the members are working in clusters geographically distributed, there is a medium level of geographical dispersion; (3) if the members are all working in the same office, there is a zero level of geographical dispersion.

Geographical Dispersion Example

To provide an example of the adoption of the formula, let us consider the case in which we have a community of 15 practitioners. If they are separated into 3 clusters of the same size—i.e., Amsterdam, India, and Seattle—we have a set of 105 distances (the pairs of 15 items without repetition). Of these, 30 are 0 km value, 25 are 7849 km (distance between Seattle and Amsterdam), 25 are 12158 km (distance between India and Seattle), and 25 are 7146 km (distance between India and Amsterdam). The geographical dispersion is 4772,24.

We computed these metrics for each time window presented in our dataset. We make this choice because community members can change over time.

Dependent variable. Since our goal was to understand the impact of cultural and geographical dispersion on the productivity of a development community, we used the number of commits per time (Adams et al., 2009; Sornette et al., 2014; Oliveira et al., 2020) as a reference measurement. We relied on this metric since it has been widely used when dealing with productivity in past works (Adams et al., 2009; Sornette et al., 2014; Oliveira et al., 2020; Bao et al., 2022; Forsgren et al., 2021).

Control variables. When constructing a statistical model, it is essential to consider that beyond independent variables, several variables can affect the phenomenon analyzed, as demonstrated in the literature (Lambiase et al., 2022b; Catolino et al., 2019b, 2021; Palomba and Tamburri, 2021; Vasilescu et al., 2015a; Graziotin et al., 2015). For this reason, we considered the following variables:

- **Number of Committers**: It is defined as the number of people that have done at least one commit in a given project time window. Having more committers could imply high productivity in terms of the number of commits.
- **Team Size**: It represents the number of contributors per team in a given temporal window. The community's size can influence the number of commits done during the development of the project.
- **Turnover**: It concerns the fraction of the team in a given temporal slice that is different from the previous windows (*i.e.*, the *turnover ratio*). A high turnover means that team members change frequently. The constant introduction of new members might lead to the variability of productivity.
- **Project Age**: It represents the difference between the maximum index and the index of the 90-day temporal interval from the first commit. Older projects and their teams could have low productivity since their systems are running into a maintenance phase and not a developing phase that is generally more active.
- **Tenure diversity**: *Tenure measure* is defined as the experience of developers in various fields (MacCurtain et al., 2010), thus possibly affecting productivity (Vasilescu et al., 2015a). In our study, we considered two types of tenure: (1) *commit tenure* (that represents the coding experience of a contributor within all GitHub projects in which s/he contributed), and (2) *project tenure* (that represents the contributor's experience in the specific project considered).
- **Tenure median**: It represents the median project tenure and the commit median tenure and is used to complement *tenure diversity*.
- **Number of women in a team**: The number of women is computed as the difference between the total number of community members and the number of men belonging to the community.
- **Blau-Index**: Blau (Blau, 1977) defined *Blau diversity index* as $1 - \sum_{i=1}(\frac{n_i}{N})^2$ where n_i represents the number of individuals or entities in category i and N is the total number of individuals or entities in the entire population or group being analyzed. The values fluctuate between 0 and 0.5, at which there is the same percentage of male and female board members, and thus the diversity is maximized.
- **Socio-Technical Congruence**: STC (Valetto et al., 2007) represents "the state in which a software development organization harbors sufficient coordination capabilities to meet the coordination demands of the technical products under development.".
- **Truck Factor**: TF represents the minimum number of members of a team that have to quit before the project fails (Williams and Kessler, 2003; Avelino et al., 2016; Ferreira et al., 2017; Avelino et al., 2019).
- **Centrality**: It is defined as the strength of a community, and it is based on modularity measures (Hatala and George Lutta, 2009). A value over 0.3 means that the community is highly modular, thus clearly distinguishing the sub-communities in its development network. A value below 0.3 means that there are no sub-communities instead.

The above factors were considered as control variables in our statistical models, according to previous literature (Lambiase et al., 2022b; Catolino et al., 2019b, 2021; Palomba and Tamburri, 2021; Vasilescu et al., 2015a; Graziotin et al., 2015).

3.2.3. Statistical model construction

The dataset used (Lambiase et al., 2022b) consists of multiple temporal windows for each project analyzed. This means there are multiple snapshots of the community's situation for the same team, *i.e.*, our data expose a hierarchical structure (based on the team). For this reason, we constructed a *linear mixed model* able to capture measurements from within the same group (*i.e.*, within the same team) as a random effect (Lindstrom and Bates, 1988).

Linear mixed models (Lindstrom and Bates, 1988) extend simple linear models by incorporating both fixed and random effects. They are particularly employed when dealing with nonindependence in the data, often arising from hierarchical structures. Mixed models comprise fixed effects, which are parameters that remain constant, and random effects, which are parameters that are treated as random variables. Such a model has started to be adopted in software engineering research when analyzing the influence between variables in cases multiple snapshots for the same study item are provided (Catolino et al., 2019b; Lambiase et al., 2022b; Vasilescu et al., 2015a). In particular, we used the time window as a random effect and the rest of the above variables (in Section 3.2.2) as fixed effects. From an implementation perspective, we relied on the functions `lmer` and `lmer.test` available in the `R` package `lme4` (Bates et al., 2014).

It is important to note that we also faced the problem of multicollinearity (O'brien, 2007), which happens when an independent variable is highly correlated with one or more of the other independent variables, thus affecting the reliability of the results. For this reason, we used a stepwise variable removal process based on the *Companion Applied Regression* (`car`) `R` package,[4] using the `vif` function (O'brien, 2007).

To support our results, we computed the effect sizes of the coefficients using the well-known ANOVA statistical test (Cuevas et al., 2004). Variables are considered significant if they are statistically significant, *i.e.*, the p-value is less than 0.05. Finally, for the sake of results reliability, we built two baseline statistical models—the first one containing all the control variables and the random effect—comparing them through the *AIC* (*Akaike information criterion*) and *BIC* (*Bayesian information criterion*) (Burnham and Anderson, 2004; Akaike, 1998) estimators. Both are statistical measures used for model selection in the context of statistical modeling and hypothesis testing. Models with a low value of AIC and BIC is the one that better characterizes the sample analyzed. This comparison allows us to study whether adding the independent factors improves the model's capability to estimate software development productivity; the comparison with the second model is whether the obtained results reflected the random effect instead.

3.3. Design of the qualitative study: Survey design

The second step of our study consisted of a qualitative investigation to study how culturally originated behaviors, and geographical dispersion can influence the productivity of a software development community. Specifically, we surveyed 140 practitioners with experience in distributed software development and management. We adopted a survey as a qualitative method to study productivity based on previous evidence that showed how (i) they are straightforward and commonly used methods to measure self-assessed productivity (Meyer et al., 2017; Murphy-Hill et al., 2019) and (ii) they are flexible methods to collect general insight from industry (Kitchenham and Pfleeger, 2008). The following sections provide information about our qualitative research approach and dissemination actions.

[4] https://cran.r-project.org/web/packages/car/index.html

3.3.1. Survey structure

Our survey consisted of eight main sections described in the following paragraphs.

Survey initiation section. The first section was inspired by the work of Murphy-Hill et al. (2019) and aimed at providing a baseline for all respondents regarding the definition of productivity—to obtain more cohesive results. Moreover, we extracted insights from participants on how to measure software development team productivity.

Survey core section. The sections from the second to the seventh were mapped on the six Hofstede dimensions and aimed at extrapolating the perceived impact of cultural dispersion metrics on self-assessed productivity. We developed such sections—and associated questions—using a *vignette-based scenario approach* (Finch, 1987), mainly used in psychological and sociological experiments (Atzmüller and Steiner, 2010). Specifically, each scenario described the characteristics of one of the cultural dimensions contextualized in the software development environment.

For example, the box in the following shows the vignette used to describe the characteristics of the *Power Distance Index (PDI)* dimension:

Example Vignette Scenario

"Suppose your development team is working on the definition of an E-commerce application. During the development, you recognize the presence of people who (1) demand to equalize the distribution of the power between all team members, in contrast to others who (2) want to follow rigidly hierarchical organization."

In the above example and all the scenarios, we described two opposite behaviors, represented using the (1) and (2) items. We operationalized such a choice to allow participants to visualize better the situation we were interested in investigating (Atzmüller and Steiner, 2010). All the used scenarios are available in our online appendix (Lambiase et al., 2023).

In order to better explain this choice, it is essential to stress that each of Hofstede's dimensions describes two opposite behaviors. For example, regarding *Individualism vs. Collectivism*, a high level indicates a society where individuals are expected to take care of only themselves and their immediate families. In contrast, a low level indicates a society in which people are supposed to be loyal to their group. In the context of dispersion metrics, we are interested in how people characterized by two opposite behaviors coexist in the same community. For such a reason, we decided to use the strategy of the scenario described above. Moreover, to avoid bias in participants' answers, we described the two behaviors directly instead of referring to the culture of individuals—for this reason, in the following section, we introduce the concept of 'culturally originated behaviors'.

After each vignette, we asked questions aimed at understanding (1) whether the situation has ever happened to our participants in their past projects (Likert Scale from *Never* to *Always*); (2) to what extent the cultural background and physical distance influence the emergence of such behaviors (Likert Scale from *Not at all* to *To a Great Extent*); (3) how much the contrast between the behaviors impact productivity (Multiple choice grid with behaviors and Likert Scale from *Strongly Disagree* to *Strongly Agree*).

We decided to use Likert Scale with five values because they are largely suggested by Kitchenham and Pfleeger (Kitchenham and Pfleeger, 2008) for qualitative studies in software engineering. Such a scale allows us to increase the readability of our survey and be exhaustive without the risk of misconception by participants. Moreover, choosing five as the number of values allowed us to extract results without being too dispersed.

Although only the third question directly relates to our research question and objective, we included the other questions mainly for three reasons. First, to be sure that the participants were getting more into the context described by the scenario; second, to have additional data to assess the goodness and quality of the answers given; and third, to gather valuable cross-sectional information in the discussion phase of the results.

Control variables validation. The eighth section of the survey contained a single question asking participants their opinion on the influence of some socio-technical factors on the productivity of a software development community. Specifically, we aimed to evaluate practitioners' perceptions of the control variables used to build our statistical model. For such a purpose, we asked participants to express if a particular factor could impact productivity, using a Likert scale from *Definitely not* to *Definitely*. The control variables included were:

- Team Size: the dimension of the team in terms of people.
- Turnover: the changing of the workforce assigned to a production process.
- Project Age: the number of years from the start of the project—*i.e.*, the first commit.
- Tenure diversity: the difference of experience between the various team members.
- Blau-Index: the team's diversity in terms of biological gender—*i.e.*, male and female.

Survey demographic section. The last section of the survey was reserved for demographic information. Our target population was composed of practitioners. We included questions related to job positions (i.e, Project Manager, Product Owner, Software Architect, Software Engineer, and other), gender, programming/management experience, and size of the team they were considering when answering the survey questions. Furthermore, we also asked participants about their cultural backgrounds. Once we had developed the survey, and before releasing it, we sought and obtained approval from the Ethical Board Committee of the University of the second author.

Survey attention question section. Following the guidelines provided by Meade and Craig (2012), we included in our survey an *attention check question* (Meade and Craig, 2012)—*i.e.*, a question aimed at checking if the participant is reading the questions and is not answering randomly. Specifically, the attention question is the following "Respond with 'Never' to this question.". We discarded three participants' answers that did not correctly answer them.

3.3.2. Survey submission and participants characteristics

From a recruitment perspective, we carefully paid attention to the target population. Indeed, we used Prolific[5] to recruit experts. Prolific is a web-based platform to support researchers in finding participants for survey studies. The platform allows tuning the preference of surveyed, putting constraints, *i.e.*, people must be practitioners and experienced in distributed work. Prolific use an *opt-in* strategy (Hunt et al., 2013): this implies that participants get voluntarily involved, possibly leading to self-selection or voluntary response bias (Heckman, 1990; Sakshaug et al., 2016). We introduced an incentive of 3 dollars per valid respondent to mitigate this bias. Prolific automatically suggests this amount according to (i) the profiles selected and (ii) the survey time. Nonetheless, answers must be double-checked at the end of the survey before assigning the reward. The survey was made available from January 30 to February 22, 2023, and we surveyed 140 practitioners.

3.3.3. Survey design guidelines

During the design of the survey and its submission, we relied on various guidelines to improve our method and, consequentially, our findings. First, regarding the survey design, we relied on the guidelines provided by Kitchenham and Pfleeger (2008) and Andrews et al. (2007)—broadly adopted for software engineering qualitative

[5] Prolific website: https://www.prolific.co/.

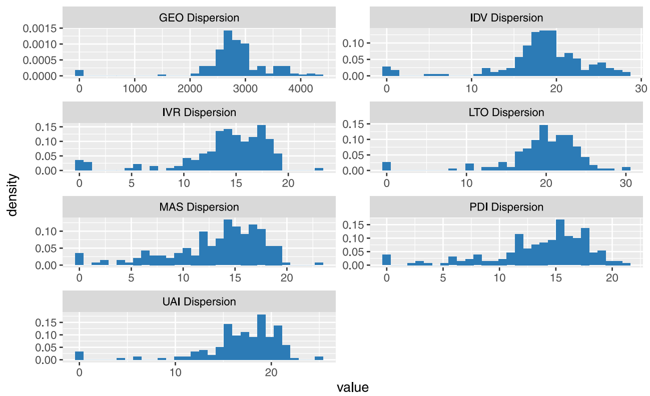

Fig. 4. Independent Variables Plots.

studies. Additionally, we considered the scale and guidelines provided by Kitchenham and Pfleeger (2008), thus allowing us to increase the readability of our survey. Finally, we took inspiration from previous qualitative investigations that focused on the productivity of software development teams (Murphy-Hill et al., 2019; Vasilescu et al., 2015a; Meyer et al., 2017).

Regarding the tool used for survey dissemination, *i.e.*, Prolific, we take inspiration from other work in literature (Reid et al., 2022; Ebert et al., 2022). Specifically, we adopted insights provided by Reid et al. (2022), which defined a series of recommendations to conduct surveys in the software engineering field using the platform.

Following the guidelines by Flanigan et al. (2008), we consciously kept the survey anonymous, preventing our influence on the answer. We created the survey as a Google form[6] and estimated a completion time of 10/15 min.

Regarding the analysis of the results, we used descriptive statistics/plots for the closed-ended questions of the surveys, while *content analysis* (Cavanagh, 1997)—*i.e.*, a research method where one or more inspectors go over the data of interest and attempt to deduct their meaning and/or the concepts they let emerge—for open-ended questions. The process was conducted by the first three authors of the paper, who jointly analyzed the individual responses to identify and label the main insights and comments left by participants.

4. Analysis of the results

This section illustrates the results of our study. For the sake of comprehensibility, we decided to divide the section into three subsections: the first one reporting the previous results from the quantitative analysis (Lambiase et al., 2022a); the second one reporting our findings from the analysis of the survey responses; and the third one, consisting of a sum-up and synthesizing of the results from both the studies. Such a report strategy is coherent with the mixed-method approach used for this study (Johnson et al., 2007). Due to space limitations and readability, detailed and raw results are available in the online appendix (Lambiase et al., 2023).

4.1. Quantitative study: Regression model

This section shows the results achieved when assessing the relationship between cultural and geographical dispersion and the productivity of a development community.

Table 3 and Fig. 5 report the details of the statistical models while Fig. 4 reports the frequency of all computed values for independent variables, including outliers. The first model (see the column "All

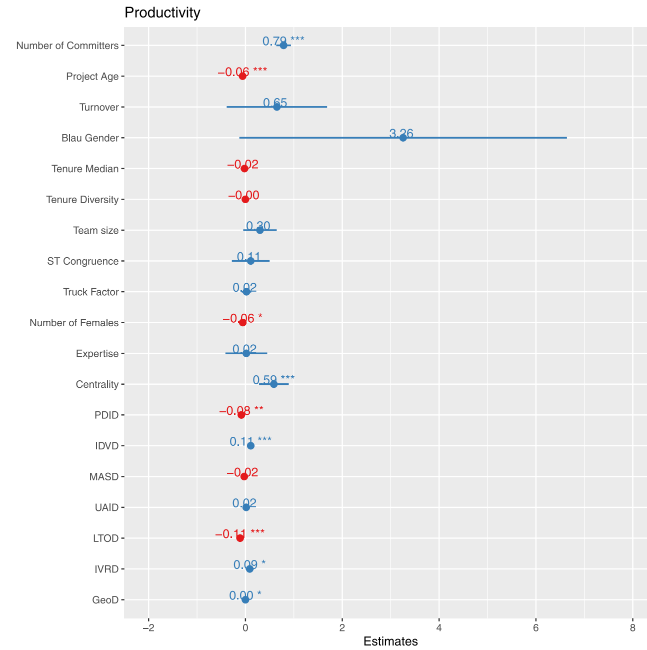

Fig. 5. Statistical Model Representation.

Variables") shows the results achieved considering both confounding factors and independent variables. In particular, the *number of committers* and the *centrality* seem to impact the productivity of a development community greatly. Moreover, the *project age* and the *number of females* in the team significantly impact the dependent variable too.

As for the cultural and geographical dispersion, we can notice that the most significant variables are *Individualism vs. Collectivism Dispersion* and *Long Term Orientation Dispersion*, followed by *Power Distance Index Dispersion*. *Indulgence vs Restraint Dispersion* and *Geographical Dispersion* seem significantly impact productivity, too. Therefore, we can claim that dispersion metrics influence the productivity of a software development community.

To provide a preliminary interpretation of our results—similar to our previous study (Lambiase et al., 2022b)—cultural and geographical dispersion influence the productivity of a development community both positively and negatively. As proof of this, *Individualism vs. Collectivism Dispersion*, *Indulgence vs. Restraint Dispersion*, and *Geographical Dispersion* impact the dependent variable positively, while *Power Distance Index Dispersion* and *Long Term Orientation Dispersion* negatively.

Table 4 shows the AIC and BIC value for the three models. The model with all variables had the lowest index value, *i.e.*, 502 and 572, compared to the one with only control and random variables. Thus, adding the independent variables contributes to explaining productivity better.

> **☞ Quantitative study: summary of the results.**
>
> Our study confirmed how socio-technical metrics could significantly impact the productivity of a development team. In addition, the study revealed that culture and geographical dispersion influence the productivity of a software development community.

4.2. Qualitative study: Survey

The following section reports (1) the background and demographic information of the study's participants and (2) the findings of our qualitative analysis.

[6] Google form website: https://www.google.com/forms/about/.

Table 3
Statistical model results.

Factor	All variables		Conf. variables		Random
	Estimate	*Sig.*	*Estimate*	*Sig.*	*Estimate*
(Intercept)	2.663		1.957		9.565
Number of Committers	0.788	***	0.899	***	
Project Age	−0.058	***	−0.051	**	
Turnover	0.649		0.092		
Blau Gender	3.256	·	5.289	**	
Tenure Median	−0.018		0.021		
Tenure Diversity	−0.001		−0.001		
Team Size	0.301	·	0.401	*	
Socio-Technical Congruence	0.109		0.171		
Truck Factor	0.021		−0.003		
Number of Females	−0.055	*	−0.055	*	
Expertise	0.018		0.059		
Centrality	0.587	***	0.563	**	
PDID	−0.084	**			
IDVD	0.109	***			
MASD	−0.024				
UAID	0.0158				
LTOD	−0.111	***			
IVRD	0.088	*			
GeoD	0.001	*			

*** $p < 0.001$.
** $p < 0.01$.
* $p < 0.05$.
· $p < 0.1$.

Table 4
AIC and BIC for the three models.

Metric	All variables	Conf. variables	Random
AIC	502	536	617
BIC	572	584	626

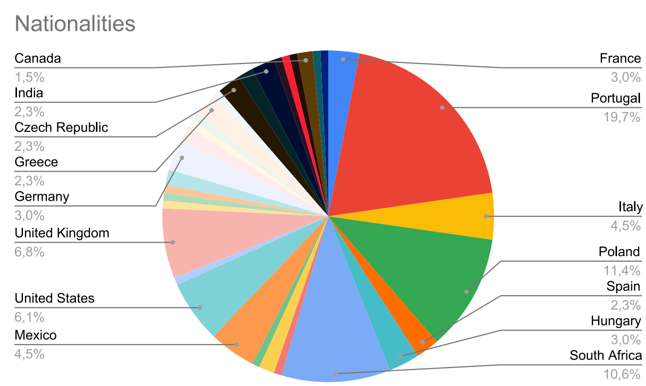

Fig. 6. Participants Nationalities.

4.2.1. Participants background information

Among the 140 participants, 70% of them worked as developers, 15% as Project Managers, and the remaining had other roles, *e.g.*, software architect and data scientist. Moreover, 50% of our sample was male, while 50% was female. Indeed, during the survey definition, we precisely decided to balance our sample based on gender in order to collect opinions from both of them. Furthermore, most of our participants self-assessed their management and programming skills as high or medium, 52% and 80%, respectively. In terms of cultural background, Fig. 6 reports the origin countries declared by our participants in the survey. Despite some countries presenting higher values (*e.g.*, Portugal and Poland), we can say that our sample is heterogeneous since the large number of countries of participants. Moreover, our final sample was composed of practitioners from all the cultural groups (according to the GLOBE framework), *i.e.*, clusters of countries exposing similar cultural behaviors and values (Javidan and Dastmalchian, 2009).[7] This further confirms that our set of participants is heterogeneous in terms of cultural background, ensuring the reliability of our results. From these basic descriptive statistics, we can claim that the answers collected provide reliable insights for validating the information gathered from the survey.

4.2.2. Perception of productivity measurement by developers

For our quantitative investigation, we selected the number of commits performed in a time range as a measure for representing the productivity of a software development community. Even though we selected such a metric motivated by previous usage in past literature (Hernández-López et al., 2013; Oliveira et al., 2020; Mockus et al., 2002)—with encouraging results—we were aware that it could be the origin of incorrect results; indeed, productivity is a complex concept to define and measure, and using a single representation could not be sufficient during a research analysis.

For the reason mentioned above, we decided to conduct a qualitative study, and we asked participants to provide us with some examples of productivity measures that they consider attentive. As shown in Fig. 7, participants consider the number of commits as a good metric to measure the productivity of a community. Therefore, we are confident that the quantitative and qualitative studies' results are reliable.

4.2.3. Culture and cultural behaviors

In the first question of each section, we asked participants to evaluate how many times they faced the described scenario in the past using a Likert scale from *Never* to *Always*. As shown in Fig. 8, most participants recognized the different cultural dimensions inside their team. In particular, *Individualism vs. Collectivism Index* and *Indulgence vs. Restraint Index* were the most frequent (respectively, 40% and 50% claimed that they often experienced it), while *Power Distance Index* was

[7] The Global Leadership and Organizational Behaviour Effectiveness (GLOBE) project is a multi-phase, multi-method project that examines the interrelationships between societal culture, organizational culture and leadership (Javidan and Dastmalchian, 2009). In its context, country groups were defined that exhibit different cultural behaviors when compared with each other.

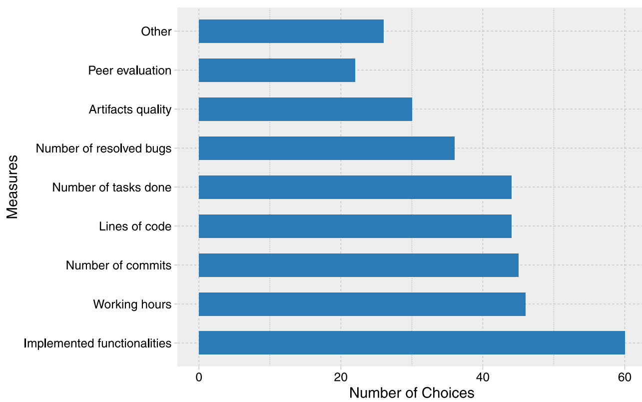

Fig. 7. Productivity measures.

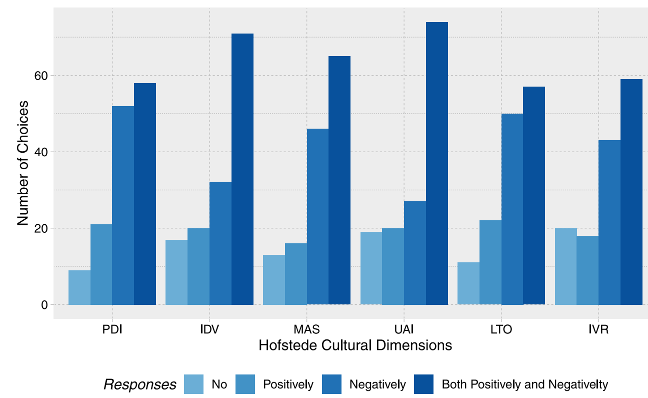

Fig. 10. Impact of cultural behaviors on productivity.

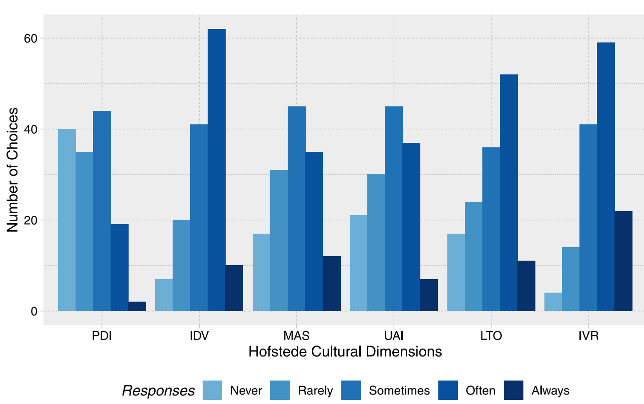

Fig. 8. Frequence of experienced cultural behaviors.

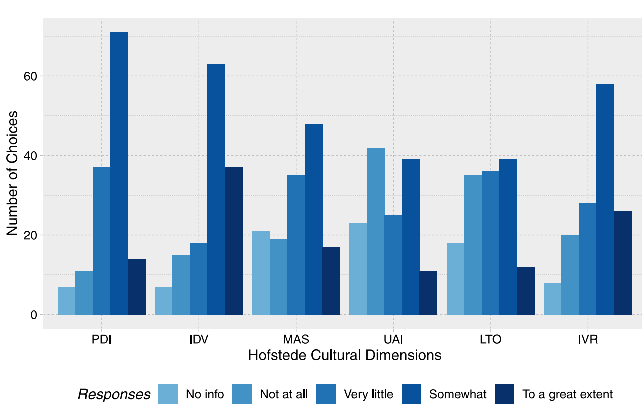

Fig. 9. Relation between culture and behaviors.

less frequent (30% of participants claimed that they *never* experienced it and 33% that they experienced it *sometimes*).

Besides studying the frequency of experienced behaviors, we were also interested in whether practitioners perceive culture as the principal root of such behaviors. Specifically, our goal was to verify if practitioners perceive a connection between these behaviors and the conjecture of Hofstede that relates them to cultural factors. For such a goal, we asked participants to what extent the individuals' culture could influence the emergence of specific behaviors. As shown in Fig. 9, most participants related the behaviors to cultural factors. In fact, in most cases, people do not answer *Not at all* or *Very little* but *Somewhat* and *To a great extent*. The only exceptions are for UAI, for which *Not at all* is the most selected (13%), and LTO, for which *Not at all* and *Very little* are the most selected ones (43% and 33%). Based on our result, we can conclude that the survey confirmed the relevance of cultural dimensions in software development communities and the relationship between belonging to different cultures and having certain behaviors.

4.2.4. Impact of dispersion metrics on productivity

As stated in Section 2.1, one cultural dimension is defined by two opposite behaviors, so in the context of the survey, we asked our participants whether the presence of both

1. does not affect productivity,
2. influences it positively (by increasing it),
3. influences it negatively (by decreasing it), or
4. influences it both positively and negatively.

In general, as reported in Fig. 10, practitioners perceive all the contrast between behaviors as impactful for productivity (as demonstrated by the low percentages of *No* answers). In fact, the higher value for *No* is nine responses—*i.e.*, 7% of respondents. Furthermore, they generally consider such an impact negatively or both negatively and positively. The only differences in this are for *Individualism vs. Collectivism* and *Uncertainty Avoidance Index* behaviors, in which we have a more significant number of responses for *Both negatively and positively* (respectively, 47% and 49%).

Regarding geographical dispersion, participants confirmed the statistical model results, assessing that physical distance is a crucial factor for the productivity of individuals. Indeed, one of the participants assessed that *"When people work remotely, and communication is not optimal, it is often necessary to repeat and meet several times before what needs to be done is done."*.

The survey confirmed the quantitative study results. Indeed *PDI*, *LTO*, and *IDV* are perceived as impactful. However, also the other behaviors are perceived as impactful by participants.

4.2.5. Impact of control variables

Regarding the perceived impact of control variables as shown in Fig. 11, most of them are perceived as impactful on software development productivity—*i.e.*, Team Size, Turnover, Project Age, Tenure diversity, and Blau-Index. The only exception to this is Blau gender, which is perceived as not impactful.

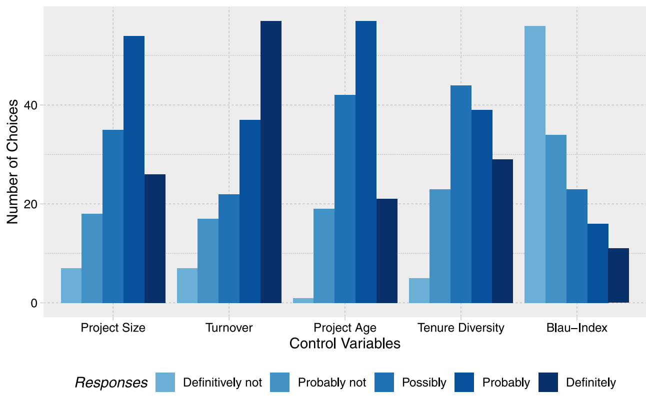

Fig. 11. Impact of control variables on productivity.

Qualitative study: summary of the results.

Practitioners perceive cultural and geographical dispersion as impactful for the productivity of a software team. Moreover, in most cases, such an impact is double-fashioned, *i.e.*, both negative and positive; this is because having a high level of cultural dispersion could—in general—lead to more extended discussions and lower team productivity. On the other side, having different points of view could help a team to identify problems sooner, thus focusing on important tasks rather than wasting time. Therefore, correctly managing cultural differences is essential to bring out the best in one's developers.

Theoretical Saturation.

Both the qualitative and the quantitative study converge to the same high-level result: dispersion metrics impact the productivity of a development community greatly. Moreover, the impact of cultural behaviors on the productivity of a software community can be categorized and analyzed using cultural dimensions, leading to the possibility for managers and team leaders to make more precise decisions. Indeed, cultural dispersion impacts productivity differently based on original Hofstede's dimension. Furthermore, the two studies reveal that the influence of dispersion metrics is positive and negative. The dispersion in the team led to extended discussion, resulting in (i) reducing productivity due to the time lost in communicating but (ii) increasing it because the discussion helped practitioners identify crucial tasks to perform. Using our findings, managers could find a converging point more easily and reduce productivity loss by knowing the contrast point—*i.e.*, the cultural dimension and corresponding behavior—enhancing the positive impact. For example, by understanding that the team is dispersed in terms of Power Distance Index Dispersion—*i.e.*, there are different attitudes in how relating to the distribution of the power—managers could implement ad hoc solutions—*e.g.*, changing the decision-making process or implementing a voting system—for reducing the negative impact of such dispersion and enhancing the positive aspects.

5. Discussion and implications

The insights of our study shed light on the role and the impact of cultural dimensions across software organizational structures revealing several lessons for software project management and community shepherding.

The role of culture as a factor influencing productivity in software development teams has yet to deepen. In this work, we tried to fill this gap by focusing on cultural and geographical dispersion. The following section discusses the final considerations based on the results achieved in Section 4.

Productivity metrics: Quantitative and qualitative. As already said, productivity is a complex concept to define and measure. Indeed, a plethora of work provided different interpretations and measurements (Hernández-López et al., 2013; Oliveira et al., 2020; Mockus et al., 2002). From such heterogeneity of materials on the matter, conducting a study that captures all the aspects of productivity comes to be a complex and tricky challenge. Nevertheless, managers can only manage what they can measure, so it is mandatory to study the phenomena and provide new insights. Aware of the complex side of productivity, we tried to address it by combining quantitative and qualitative representations in our studies. The former—*i.e.*, the number of commits—allowed us to be more "objective" in our findings; the latter—*i.e.*, self-assessed productivity—allowed us to capture more aspects and indicators of productivity. Ultimately, the two studies reported similar results, providing a potential precedent for ulteriorly enhanced findings on the matter using similar methods.

Individualism vs. Collectivism dispersion. Regarding the influence of *IDVD* on productivity, both the qualitative and the quantitative studies showed that such a dispersion metric could positively influence the productivity of a software development community. As proof, we informally discuss these results with some practitioners with experience as managers in open-source projects from a large company. They highlighted that the presence of both individualistic and collectivistic people in the same team could increase productivity. Specifically, one of them reports that *"In my attempt to try to integrate an individualistic team member, I found that it was dragging on performance for the other individuals and me."*. This assertion could also justify the second predominant result from the survey, *i.e.*, dispersion could also negatively affect the dependent variable—from a statistical point of view. Indeed, this situation can occur considering the example above, *i.e.*, trying to enforce a paradigm change in the communication and collaboration pattern of the individualistic person.

We could draw that supporting individualistic behavior leads to the emergence of "positive lone wolves", thus improving the community members' productivity.

Such a fact was also confirmed by various participants of the survey: for example, one of them—asked about the root causes of problems originated by such a dispersion—reported this: *"It is okay for people to prefer to work alone as long as they do the work and show up to all meetings to collaborate with the team and show the work they have done."*.

Long vs. Short term orientation and power distance index dispersion. Both *LTOD* and *PDID* negatively affect the productivity of a development community. These metrics reflect similar behaviors: indeed, *LTOD* indicates a contrast between people who tend to resist change—*e.g.*, try out new programming languages or technologies—in contrast to others who are more flexible; *PDID*, indicates a difference between individuals who demand to equalize the distribution of the power between all team members, in contrast to others who want to follow a rigidly hierarchical organization.

In software development, these metrics can lengthen the time required to make a decision, thus possibly decreasing productivity when developing software artifacts or making decisions.

As a confirmation of this, one of the participants in the survey reported that *"When the difference in behaviors is from the team managing the processes, this can lead to delays whilst decisions are made."*.

⚠ Moreover, an ulterior point could be that the resentment arising from the contrast between individuals could lead to the formation of *Organizational Silos—i.e.*, the situation in which different clusters of individuals in the same team arise and the communication between such "silos" is poor (Tamburri et al., 2016; Lambiase et al., 2022b) –, thus decreasing the communication and increasing the time to perform actions.

As a confirmation of this, one of the respondents to the questionnaire reports that *"Less collaboration occurs because people resent colleagues and do not want to share credit, so things get done slower."*.

Despite the negative influence that LTOD and PDID have on productivity, it is crucial to remember that low productivity does not necessarily mean low quality. In other words, more extended discussions—other than lowering developers' productivity—could lead to better identification of problems and their solutions, resulting in better quality—*e.g.*, a better understanding of requirements. The critical aspect here is to prevent discussions from escalating into unnecessarily lengthy communications or, at worst, toxic arguments that could negatively impact the community's productivity. Related to this, we have observed a strong correlation between cooperation and collaboration issues and the concept of community smells (Tamburri et al., 2016, 2013), which participants also mentioned, *i.e.*, organizational silos. In our previous research (Catolino et al., 2020), we have presented strategies to mitigate some of these smells. We believe that these strategies can be valuable in addressing the problems mentioned above and enhancing productivity.

One approach is facilitating mentoring activities to support team members in adapting to changes. This can involve organizing workshops, providing mentoring opportunities, or granting access to educational resources. By offering guidance and assistance, we can help individuals overcome their resistance to change and navigate new approaches more effectively. In addition, emphasizing clear communication is essential. It ensures the team comprehends the proposed changes' vision, goals, and rationale. When the reasoning is effectively communicated, it fosters understanding, reduces resistance, and encourages buy-in from the team members. Lastly, engaging in cohesion exercises with team leaders can play a vital role in promoting a positive attitude toward change and adaptation. Team leaders set an example for others to follow by demonstrating a willingness to embrace change. Being open to feedback and displaying a constructive mindset creates an environment where change is embraced as an opportunity for growth.

Control variables. Our study confirms previous findings (Wagner and Ruhe, 2018; de Lemos Meira et al., 2010): indeed, socio-technical metrics (Valetto et al., 2007; Cataldo et al., 2006), *e.g.*, turnover and tenure diversity, are strongly correlated with the productivity of a development community. For example, *centrality*—the degree to which a community is divided into sub-communities (Hatala and George Lutta, 2009)—positively influences the dependent variable. The reason could be that, with the increasing number of developers, modularizing the team could lead to better micro-management, improving the community's productivity.

In addition, as we might expect, the *number of committers* and the *age of the project* also affect productivity. Concerning the first variable, it probably depends on the way chosen to represent productivity in this study (the number of commits): more committers likely lead to more commits over time. Regarding the second variable, the project's age negatively impacts the community's productivity. This could be because some open-source communities tend to die over time, mainly if some core contributors migrate to other teams.

As a final note, the survey's participants found *Blau-Gender—i.e.*, the degree to which a community is different in terms of biological gender composition—as not relevant when analyzing the productivity of development communities. Both the analysis in this study confirmed these results. For example, Catolino et al. (2019a) found that gender diversity is perceived as being less important than experience or team size to mitigate the emergence of communication and collaboration issues—represented using community smells. Nevertheless, other quantitative investigations demonstrated that gender diversity could positively impact product and process metrics in software development (Lambiase et al., 2022b; Catolino et al., 2019b; Palomba et al., 2021). Undoubtedly, ulterior research is needed to assess the influence of gender diversity in software development communities, maybe introducing novel methodologies for conducting investigations.

Benefits for tool vendors and developers. Nowadays, managing software development teams and conducting software projects is an expensive effort—both in terms of money, time, and human resources. Moreover, such an effort ulteriorly arises in the context of distributed teams characterized by a plethora of factors, first of all, the extreme diversity of stakeholders in terms of behaviors and beliefs (Herbsleb and Moitra, 2001; Mockus and Herbsleb, 2001). Furthermore, developing guidelines, methods, and tools to help practitioners is complex due to the—at least apparently—abstractive nature of such social aspects. Nevertheless, our findings reveal that quantitative frameworks for measuring dispersion in development teams could be effective and can constitute a foundational step in developing recommendation systems for managers. Indeed, as a basilar example, using only the nationalities of team members—without any private information on the actual individuals—could allow managers to compute the dispersion rates and improve the estimation of productivity effort required for the project, constituting—if performed automatically—a worthwhile contribution.

⚠ Recommendation systems designers and developers should take the opportunity provided by quantitative representation of social aspects by developing tools to support practitioners in using and managing such factors. For example, computing a development team's dispersion metrics could contribute to the management process of software development projects.

We should care about software community health. These results shed light on an important aspect often underestimated by managers. Indeed, the health status of a community—its behaviors and diversity—needs to be carefully monitored when dealing with software development since it can affect the quantity (productivity) and quality of what it produces. Our conclusions can represent the first step toward better characterizing a software community in terms of "health".

6. Threats to validity

This section illustrates the threats to the validity of the study and how we mitigated them. Other than using the guidelines provided for qualitative studies (Kitchenham and Pfleeger, 2008), we identified and organized the threats using the well-known framework proposed by Wohlin et al. (2012, 2003).

6.1. Threats to construct validity

Threats in this category refer to the relationship between hypothesis and observations and are mainly due to imprecision in performed measurements (Wohlin et al., 2012).

The first threat concerns the usage of Hofstede (Hofstede et al., 2005) for characterizing cultural dimensions. Although a few researchers raised concerns about this framework (Roberts and Boyacigiller, 2012; Ailon, 2008; Baskerville, 2003), Venkateswaran and Ojha (2019) showed how the framework is the most efficient way to represent the complex world of cultural attitudes (Hofstede, 2017; Borchers, 2003; Casey, 2011; Abufardeh and Magel, 2010).

A further concern regards the usage of the metric chosen for measuring productivity. Specifically, we used the number of commits for the quantitative investigation, while for the qualitative one, we relied on self-assessed productivity. This difference could introduce some imprecision and threats to the theoretical saturation of the studies. To mitigate this, in the first section of the survey, we asked participants to provide us with some examples of productivity measures that they consider attentive. As shown in Fig. 7, participants consider the number of commits as a good metric to measure the productivity of a community.

As for the cultural and geographical dispersions metrics, we use the standard deviation since these metrics can be unreliable in the case of skewed measures. We applied the well-known Shapiro–Wilk test (Shapiro and Wilk, 1965) to verify the normality of the data. Moreover, these metrics have already been adopted in a few studies in the context of software development (Li et al., 2010; Lambiase et al., 2022b; Tamburri et al., 2019b). Nevertheless, we encourage replication of the study using different measures.

Another threat is related to the dataset chosen to conduct our study. To address such a threat, we relied on a dataset already used and tested in similar studies (Lambiase et al., 2022b; Vasilescu et al., 2015b; Catolino et al., 2019b).

Regarding the threat concerning the survey design, we followed the guidelines provided by Kitchenham and Pfleeger (2008) and Andrews et al. (2007) to define clear and explicit questions in the survey that could allow participants to get into the survey correctly. We also included open questions to let participants express themselves freely, without restriction. Finally, before sending the survey, we conducted a pilot study with four developers that reported possible biases and flaws we fixed before releasing the survey.

Our final concern pertains to the potential existence of sub-cultures within a larger cultural group. According to social science literature, individuals who share common backgrounds and beliefs might display minor variations in behavior, leading to the identification of sub-groups within the same cultural category [83, 84]. In our study, we utilized Hofstede's framework as a means to assess culture, which primarily focuses on defining culture at the country level, possibly overlooking subcultures. However, it should be acknowledged that Hofstede considered the presence of subcultures while designing the survey and assigning values to each country. Therefore, we are confident that the different sub-cultures behaviors are represented inside the used framework.

6.2. Threats to internal validity

Threats in this category are concerned with the possibility that the independent variable is affected by casualty factors without the researcher's knowledge (Wohlin et al., 2012).

The main threat is how we recruited our participants and their capacity to report about culturally dispersed teams. For the survey, we relied on voluntary participation through an online instrument like Prolific. In particular, this platform allows access to a pool of participants samples based on specific characteristics and backgrounds, *e.g.*, computer science, working as developers. In our case, we looked for people with (1) a background in computer science, (2) experience in information services and data processing, and (3) experience in distributed teams. Furthermore, we tried to balance the number of participants based on gender as possible.

The participant's capacity to report about culturally dispersed teams is a second critical threat. Prolific allows us to distribute the survey to globally distributed participants (more information on this is in the responses to the last section of the survey, in the online appendix (Lambiase et al., 2023)). For such a reason—combined with a large number of participants—we are confident that the target audience has been represented. Nevertheless, we encourage replications of our work to strengthen our findings and their generalizability.

6.3. Threats to conclusion validity

Threats in this category are concerned with the ability to draw correct conclusions about relations between treatments and outcomes of an experiment (Wohlin et al., 2012).

The first threat concerns the statistical model selected for our study. We used a *mixed-effect model* (Lindstrom and Bates, 1988; Bates et al., 2014) to manage the multiple time windows for each project, thus capturing information within the same group. Additionally, we used *vif* for dealing with multicollinearity (O'brien, 2007), and ANOVA test (Cuevas et al., 2004) for checking the significance of the results. Finally, to avoid omitting additional factors influencing a team's productivity, we included some socio-technical control factors identified by previous literature (Murphy-Hill et al., 2019; Wagner and Ruhe, 2018; de Lemos Meira et al., 2010; Hernández-López et al., 2013; Mohagheghi and Conradi, 2007), *e.g.*, socio-technical congruence.

A possible threat concerns the choice of Prolific Platform that involves the use of an incentive for people who do the survey, thus possibly affecting the results of our study. In our case, the participants involved in the survey obtained around 1$ for participating, thus allowing us to collect several answers. However, (1) the participation was voluntary, (2) we could collect various opinions from people with different experiences, thus increasing the reliability of our results, and (3) we asked to report their experience according to particular situations without influencing their answers. We released all the material used for this study to enable the verifiability of the conclusions described in our paper (Lambiase et al., 2023).

6.4. Threats to external validity

Threats in this category are concerned with the generalizability of the results (Wohlin et al., 2012).

The main threat relates to the generalizability of the results. We use a dataset (Lambiase et al., 2022b) containing information about big open-source projects on GitHub, with a large number of contributors. Nevertheless, we plan to extend the number of systems and perform some qualitative studies, *e.g.*, focus groups, surveys, and interviews, to strengthen the results as a future agenda. Moreover, we adopted a mixed-method approach, combining the results of a previously performed quantitative study with a large-scale survey. Although our numbers align with the SE research community's studies, we know how results strongly correlate with our sample, so replications are part of our agenda.

7. Conclusions

This study presents a qualitative extension of a quantitative empirical study (Lambiase et al., 2022a) that investigates the relationship between the cultural and geographical dispersion of a community—*i.e.*, the degree to which a community is formed by individuals growing up in and coming from different places globally—and its productivity, consider the number of commits in a specific range of time. The research question we wanted to answer was *"To what extent do cultural and geographical dispersion influence team's productivity?"*.

Our findings demonstrated that dispersion metrics impact productivity both positively and negatively. Indeed, how managers and leaders approach the differences between the cultural background of individuals is a crucial factor in determining their impact on the team. As an example of this, in the case of individualistic people collaborating with more collaboration-oriented ones, trying to integrate them into the team could lead to a useless and ineffective effort. Moreover, we found that the quantitative representations of social aspects are now mature to be exploited by practitioners. Developing tools and recommendation systems to track such metrics and visualize them to practitioners is an opportunity that should be considered.

As a future agenda, we plan to extend the generalizability of our results by adopting various productivity measurements. Moreover, we plan to develop instruments—*e.g.*, conversational agents—aiming to perform a technology transfer from the research field to the practitioners one, making our findings easily usable by practitioners.

CRediT authorship contribution statement

Stefano Lambiase: Formal analysis, Investigation, Data curation, Validation, Writing – original draft, Visualization. **Fabio Palomba:** Conceptualization, Methodology, Validation, Writing – review & editing. **Filomena Ferrucci:** Supervision, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data are publicly available on the web.

Acknowledgments

The authors would like to thank the participants of our study for the time and feedback provided. The authors would like to thank the two anonymous Reviewers for their insightful comments on our paper. Fabio is partially supported by the Swiss National Science Foundation - SNF Project No. PZ00P2_186090. Furthermore, this work has been partially supported by the EMELIOT national research project, which has been funded by the MUR under the PRIN 2020 program (Contract No. 2020W3A5FY).

References

Abufardeh, S., Magel, K., 2010. The impact of global software cultural and linguistic aspects on global software development process (GSD): Issues and challenges. In: 4th International Conference on New Trends in Information Science and Service Science. IEEE, pp. 133–138.
Adams, P.J., Capiluppi, A., Boldyreff, C., 2009. Coordination and productivity issues in free software: The role of brooks' law. In: 2009 IEEE International Conference on Software Maintenance. IEEE, pp. 319–328.
Ailon, G., 2008. Mirror, mirror on the wall: Culture's consequences in a value test of its own design. Acad. Manag. Rev. 33 (4), 885–904.
Akaike, H., 1998. Information theory and an extension of the maximum likelihood principle. In: Selected Papers of Hirotugu Akaike. Springer, pp. 199–213.
Andrews, D., Nonnecke, B., Preece, J., 2007. Conducting research on the internet:: Online survey design, development and implementation guidelines.
Atzmüller, C., Steiner, P.M., 2010. Experimental vignette studies in survey research. Methodology.
Avelino, G., Constantinou, E., Valente, M.T., Serebrenik, A., 2019. On the abandonment and survival of open source projects: An empirical investigation. In: 2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. ESEM 2019, Porto de Galinhas, Recife, Brazil, September 19–20, 2019, IEEE, pp. 1–12. http://dx.doi.org/10.1109/ESEM.2019.8870181.
Avelino, G., Passos, L., Hora, A., Valente, M.T., 2016. A novel approach for estimating truck factors. In: 2016 IEEE 24th International Conference on Program Comprehension. ICPC, IEEE, pp. 1–10.
Bao, L., Li, T., Xia, X., Zhu, K., Li, H., Yang, X., 2022. How does working from home affect developer productivity?—A case study of Baidu during the COVID-19 pandemic. Sci. China Inf. Sci. 65 (4), 142102.
Baskerville, R.F., 2003. Hofstede never studied culture. Account. Organ. Soc. 28 (1), 1–14.
Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.
Behutiye, W.N., Rodríguez, P., Oivo, M., Tosun, A., 2017. Analyzing the concept of technical debt in the context of agile software development: A systematic literature review. Inf. Softw. Technol. 82, 139–158.
Blau, P.M., 1977. Inequality and Heterogeneity: a Primitive Theory of Social Structure, Vol. 7. Free Press New York.
Boehm, B.W., Clark, Horowitz, Brown, Reifer, Chulani, Madachy, R., Steece, B., 2000. Software Cost Estimation with Cocomo II, first ed. Prentice Hall PTR, USA.
Borchers, G., 2003. The software engineering impacts of cultural factors on multicultural software development teams. In: 25th International Conference on Software Engineering, 2003. Proceedings. IEEE, pp. 540–545.
Brewer, P., Venaik, S., 2012. On the misuse of national culture dimensions. Int. Mark. Rev.
Brewer, P., Venaik, S., 2014. The ecological fallacy in national culture research. Organ. Stud. 35 (7), 1063–1086.
Burnham, K.P., Anderson, D.R., 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociol. Methods Res. 33 (2), 261–304.
Cardozo, E.S., Araújo Neto, J.B.F., Barza, A., França, A.C.C., da Silva, F.Q., 2010. SCRUM and productivity in software projects: a systematic literature review. In: 14th International Conference on Evaluation and Assessment in Software Engineering. EASE, pp. 1–4.
Casey, V., 2011. Imparting the importance of culture to global software development. ACM Inroads 1 (3), 51–57.
Casey, V., Richardson, I., 2008. A structured approach to global software development. In: European Systems and Software Process Improvement and Innovation, Dublin, Ireland.
Cataldo, M., Wagstrom, P.A., Herbsleb, J.D., Carley, K.M., 2006. Identification of coordination requirements: Implications for the design of collaboration and awareness tools. In: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work. pp. 353–362.
Catolino, G., Palomba, F., Tamburri, D.A., Serebrenik, A., 2021. Understanding community smells variability: A statistical approach. In: 2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Society. ICSE-SEIS, IEEE, pp. 77–86.
Catolino, G., Palomba, F., Tamburri, D.A., Serebrenik, A., Ferrucci, F., 2019a. Gender diversity and community smells: insights from the trenches. IEEE Softw. 37 (1), 10–16.
Catolino, G., Palomba, F., Tamburri, D.A., Serebrenik, A., Ferrucci, F., 2019b. Gender diversity and women in software teams: How do they affect community smells? In: 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society. ICSE-SEIS, IEEE, pp. 11–20.
Catolino, G., Palomba, F., Tamburri, D.A., Serebrenik, A., Ferrucci, F., 2020. Refactoring community smells in the wild: the practitioner's field manual. In: Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Software Engineering in Society. pp. 25–34.
Cavanagh, S., 1997. Content analysis: concepts, methods and applications. Nurse Res. 4 (3), 5–16.
Cherry, S., Robillard, P.N., 2004. Communication problems in global software development: Spotlight on a new field of investigation. In: International Workshop on Global Software Development, International Conference on Software Engineering. Edinburgh, Scotland, IET, pp. 48–52.
Cuevas, A., Febrero, M., Fraiman, R., 2004. An anova test for functional data. Comput. Statist. Data Anal. 47 (1), 111–122.
Darwish, A., Henryson, A., 2019. How do cultural characteristics and software engineering practices interplay? URL: https://gupea.ub.gu.se/handle/2077/62534.
de Lemos Meira, S.R., Barros, E.A., de Aquino, G.S., Silva, M.J.C., 2010. A review of productivity factors and strategies on software development. In: 2010 Fifth International Conference on Software Engineering Advances. IEEE, pp. 196–204.
Deshpande, S., Richardson, I., Casey, V., Beecham, S., 2010. Culture in global software development - a weakness or strength? In: 2010 5th IEEE International Conference on Global Software Engineering. pp. 67–76. http://dx.doi.org/10.1109/ICGSE.2010.16.
Ebert, F., Serebrenik, A., Treude, C., Novielli, N., Castor, F., 2022. On recruiting experienced GitHub contributors for interviews and surveys on prolific. In: International Workshop on Recruiting Participants for Empirical Software Engineering.
Elbert, C., 2010. Global software engineering: Distributed development, outsourcing, and supplier management. In: IEEE Computer Society Books.
Ferreira, M., Valente, M.T., Ferreira, K., 2017. A comparison of three algorithms for computing truck factors. In: 2017 IEEE/ACM 25th International Conference on Program Comprehension. ICPC, IEEE, pp. 207–217.
Finch, J., 1987. The vignette technique in survey research. Sociology 21 (1), 105–114.
Flanigan, T.S., McFarlane, E., Cook, S., 2008. Conducting survey research among physicians and other medical professionals: a review of current literature. In: Proceedings of the Survey Research Methods Section, American Statistical Association, Vol. 1. pp. 4136–4147.
Forsgren, N., Storey, M.-A., Maddila, C., Zimmermann, T., Houck, B., Butler, J., 2021. The SPACE of developer productivity: There's more to it than you think. Queue 19 (1), 20–48.
Furnham, A., 2012. The Psychology of Behaviour At Work: The Individual in the Organization. Psychology Press.
Girardi, D., Lanubile, F., Novielli, N., Serebrenik, A., 2021. Emotions and perceived productivity of software developers at the workplace. IEEE Trans. Softw. Eng. 48 (9), 3326–3341.
Gorla, N., Lam, Y.W., 2004. Who should work with whom? Building effective software project teams. Commun. ACM 47 (6), 79–82.

Graziotin, D., Wang, X., Abrahamsson, P., 2015. Do feelings matter? On the correlation of affects and the self-assessed productivity in software engineering. J. Softw. Evol. Process 27 (7), 467–487.

Hall, E.T., 1989. Beyond Culture. Anchor.

Hampden-Turner, C., Trompenaars, F., Hampden-Turner, C., 2020. Riding the Waves of Culture: Understanding Diversity in Global Business. Hachette UK.

Hatala, J.-P., George Lutta, J., 2009. Managing information sharing within an organizational setting: A social network perspective. Perform. Improv. Q. 21 (4), 5–33.

Heckman, J.J., 1990. Selection bias and self-selection. In: Econometrics. Springer, pp. 201–224.

Herbsleb, J.D., Moitra, D., 2001. Global software development. IEEE Softw. 18 (2), 16–20.

Hernández-López, A., Colomo-Palacios, R., García-Crespo, Á., 2013. Software engineering job productivity—a systematic review. Int. J. Softw. Eng. Knowl. Eng. 23 (03), 387–406.

Hofstede, G., 2011. Dimensionalizing cultures: The hofstede model in context. Online Read. Psychol. Cult. 2 (1), 2307–0919.

Hofstede, G., 2017. 50 Years memory lane – developing cultural dimensions from IBM data. In: Software of the Mind 2.0. IBM Netherlands.

Hofstede, G., Hofstede, G.J., Minkov, M., 2005. Cultures and Organizations: Software of the Mind, Vol. 2. Mcgraw-hill New York.

Hunt, K.J., Shlomo, N., Addington-Hall, J., 2013. Participant recruitment in sensitive surveys: a comparative trial of 'opt in'versus 'opt out'approaches. BMC Med. Res. Methodol. 13 (1), 1–8.

Javidan, M., Dastmalchian, A., 2009. Managerial implications of the GLOBE project: A study of 62 societies. Asia Pac. J. Hum. Resour. 47 (1), 41–58.

Johnson, R.B., Onwuegbuzie, A.J., Turner, L.A., 2007. Toward a definition of mixed methods research. J. Mix. Methods Res. 1 (2), 112–133.

Kitchenham, B.A., Pfleeger, S.L., 2008. Personal opinion surveys. In: Guide to Advanced Empirical Software Engineering. Springer, pp. 63–92.

Kreitner, R., Kinicki, A., Buelens, M., 1999. Organizational Behavior, First European Edition ed. London UK McGraw-Hill Publishing Company.

Lambiase, S., Catolino, G., Pecorelli, F., Tamburri, D.A., Palomba, F., Ferrucci, F., van den Heuvel, W.-J., 2023. An empirical investigation into the influence of software communities' cultural and geographical dispersion on productivity – online appendix. http://dx.doi.org/10.6084/m9.figshare.22147964.

Lambiase, S., Catolino, G., Pecorelli, F., Tamburri, D.A., Palomba, F., Van Den Heuvel, W.-J., Ferrucci, F., 2022a. "There and Back Again?" on the influence of software community dispersion over productivity. In: 2022 48th Euromicro Conference on Software Engineering and Advanced Applications. SEAA, IEEE, pp. 177–184.

Lambiase, S., Catolino, G., Tamburri, D.A., Serebrenik, A., Palomba, F., Ferrucci, F., 2022b. Good fences make good neighbours? On the impact of cultural and geographical dispersion on community smells. In: Proceedings of the 2022 ACM/IEEE 44th International Conference on Software Engineering: Software Engineering in Society. In: ICSE-SEIS '22, Association for Computing Machinery New York, NY, USA, pp. 67–78. http://dx.doi.org/10.1145/3510458.3513015, https://doi.org/10.1145/3510458.3513015.

Li, W., Yang, C., Yang, C., 2010. An active crawler for discovering geospatial web services and their distribution pattern–a case study of OGC web map service. Int. J. Geogr. Inf. Sci. 24 (8), 1127–1147.

Lindstrom, M.J., Bates, D.M., 1988. Newton—Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. J. Amer. Statist. Assoc. 83 (404), 1014–1022.

MacCurtain, S., Flood, P.C., Ramamoorthy, N., West, M.A., Dawson, J.F., 2010. The top management team, reflexivity, knowledge sharing and new product performance: A study of the Irish software industry. Creativity Innov. Manag. 19 (3), 219–232.

Machuca-Villegas, L., Gasca-Hurtado, G.P., Muñoz, M., 2021. Measures related to social and human factors that influence productivity in software development teams. Int. J. Inf. Syst. Project Manag. 9 (3), 43–67.

Machuca-Villegas, L., Gasca-Hurtado, G.P., Puente, S.M., Tamayo, L.M.R., 2022. Perceptions of the human and social factors that influence the productivity of software development teams in Colombia: A statistical analysis. J. Syst. Softw. 192, 111408.

Machuca-Villegas, L., Gasca-Hurtado, G.P., Restrepo Tamayo, L.M., Morillo Puente, S., 2020. Social and human factor classification of influence in productivity in software development teams. In: Systems, Software and Services Process Improvement: 27th European Conference, EuroSPI 2020, DÜSseldorf, Germany, September 9–11, 2020, Proceedings 27. Springer, pp. 717–729.

Marinho, M., Luna, A., Beecham, S., 2018. Global software development: practices for cultural differences. In: International Conference on Product-Focused Software Process Improvement. Springer, pp. 299–317.

Meade, A.W., Craig, S.B., 2012. Identifying careless responses in survey data. Psychol. Methods 17 (3), 437.

Meyer, A.N., Barton, L.E., Murphy, G.C., Zimmermann, T., Fritz, T., 2017. The work life of developers: Activities, switches and perceived productivity. IEEE Trans. Softw. Eng. 43 (12), 1178–1193.

Mockus, A., Fielding, R.T., Herbsleb, J.D., 2002. Two case studies of open source software development: Apache and mozilla. ACM Trans. Softw. Eng. Methodol. 11 (3), 309–346.

Mockus, A., Herbsleb, J., 2001. Challenges of global software development. In: Proceedings Seventh International Software Metrics Symposium. IEEE, pp. 182–184.

Mohagheghi, P., Conradi, R., 2007. Quality, productivity and economic benefits of software reuse: a review of industrial studies. Empir. Softw. Eng. 12 (5), 471–516.

Murphy-Hill, E., Jaspan, C., Sadowski, C., Shepherd, D., Phillips, M., Winter, C., Knight, A., Smith, E., Jorde, M., 2019. What predicts software developers' productivity? IEEE Trans. Softw. Eng. 47 (3), 582–594.

Noll, J., Beecham, S., Richardson, I., 2011. Global software development and collaboration: barriers and solutions. ACM Inroads 1 (3), 66–78.

O'brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. Qual. Quant. 41 (5), 673–690.

Oliveira, E., Conte, T., Cristo, M., Valentim, N., 2018. Influence factors in software productivity—a tertiary literature review. Int. J. Softw. Eng. Knowl. Eng. 28 (11n12), 1795–1810.

Oliveira, E., Fernandes, E., Steinmacher, I., Cristo, M., Conte, T., Garcia, A., 2020. Code and commit metrics of developer productivity: a study on team leaders perceptions. Empir. Softw. Eng. 25 (4), 2519–2549.

Palomba, F., Andrew Tamburri, D., Arcelli Fontana, F., Oliveto, R., Zaidman, A., Serebrenik, A., 2021. Beyond technical aspects: How do community smells influence the intensity of code smells? IEEE Trans. Softw. Eng. 47 (1), 108–129. http://dx.doi.org/10.1109/TSE.2018.2883603.

Palomba, F., Tamburri, D.A., 2021. Predicting the emergence of community smells using socio-technical metrics: a machine-learning approach. J. Syst. Softw. 171, 110847.

Petersen, K., 2011. Measuring and predicting software productivity: A systematic map and review. Inf. Softw. Technol. 53 (4), 317–343.

Project Management Institute, 2021. A Guide to the Project Management Body of Knowledge, seventh ed. p. 250.

Ramírez, Y.W., Nembhard, D.A., 2004. Measuring knowledge worker productivity: A taxonomy. J. Intellect. Cap.

Reid, B., Wagner, M., d'Amorim, M., Treude, C., 2022. Software engineering user study recruitment on prolific: An experience report. arXiv preprint arXiv:2201.05348.

Richardson, I., Casey, V., Burton, J., McCaffery, F., 2010. Global software engineering: A software process approach. In: Mistrík, I., Grundy, J., Hoek, A., Whitehead, J. (Eds.), Collaborative Software Engineering. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 35–56. http://dx.doi.org/10.1007/978-3-642-10294-3_2.

Roberts, K.H., Boyacigiller, N.A., 2012. 3. Cross-national organizational research: The grasp of the blind men. In: Societal Culture and Management. De Gruyter, pp. 51–69.

Sadowski, C., Zimmermann, T., 2019. Rethinking Productivity in Software Engineering. Springer Nature.

Sakshaug, J.W., Schmucker, A., Kreuter, F., Couper, M.P., Singer, E., 2016. Evaluating active (opt-in) and passive (opt-out) consent bias in the transfer of federal contact data to a third-party survey agency. J. Surv. Stat. Methodol. 4 (3), 382–416.

Scacchi, W., 1995. Understanding software productivity. In: Software Engineering and Knowledge Engineering: Trends for the Next Decade. World Scientific, pp. 273–316.

Shah, H., Nersessian, N.J., Harrold, M.J., Newstetter, W., 2012. Studying the influence of culture in global software engineering: Thinking in terms of cultural models. In: Proceedings of the 4th International Conference on Intercultural Collaboration. ICIC '12, Association for Computing Machinery, New York, NY, USA, pp. 77–86.

Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika 52 (3/4), 591–611.

Sorge, A., 1983. Review of culture's consequences: International differences in work-related values. Adm. Sci. Q. 28 (4), 625–629.

Sornette, D., Maillart, T., Ghezzi, G., 2014. How much is the whole really more than the sum of its parts? superlinear productivity in collective group actions. PLoS One 9 (8), e103023.

Stray, V., Moe, N.B., 2020. Understanding coordination in global software engineering: A mixed-methods study on the use of meetings and slack. J. Syst. Softw. 170, 110717.

Tamburri, D.A., Kazman, R., Fahimi, H., 2016. The architect's role in community shepherding. IEEE Softw. 33 (6), 70–79.

Tamburri, D.A., Lago, P., Vliet, H.v., 2013. Organizational social structures for software engineering. ACM Comput. Surv. 46 (1), 1–35.

Tamburri, D.A., Palomba, F., Kazman, R., 2019a. Exploring community smells in open-source: An automated approach. IEEE Trans. Softw. Eng. 47 (3), 630–652. http://dx.doi.org/10.1109/TSE.2019.2901490.

Tamburri, D.A., Palomba, F., Serebrenik, A., Zaidman, A., 2019b. Discovering community patterns in open-source: a systematic approach and its evaluation. Empir. Softw. Eng. 24 (3), 1369–1417.

Valetto, G., Helander, M., Ehrlich, K., Chulani, S., Wegman, M., Williams, C., 2007. Using software repositories to investigate socio-technical congruence in development projects. In: Fourth International Workshop on Mining Software Repositories. MSR'07: ICSE Workshops 2007, IEEE, p. 25.

Vasilescu, B., Posnett, D., Ray, B., van den Brand, M.G., Serebrenik, A., Devanbu, P., Filkov, V., 2015a. Gender and tenure diversity in GitHub teams. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 3789–3798.

Vasilescu, B., Serebrenik, A., Filkov, V., 2015b. A data set for social diversity studies of GitHub teams. In: 2015 IEEE/ACM 12th Working Conference on Mining Software Repositories. pp. 514–517. http://dx.doi.org/10.1109/MSR.2015.77.

Venkateswaran, R.T., Ojha, A.K., 2019. Abandon Hofstede-based research? Not yet! A perspective from the philosophy of the social sciences. Asia Pac. Bus. Rev. 25 (3), 413–434.

Wagner, S., Ruhe, M., 2018. A systematic review of productivity factors in software development. arXiv preprint arXiv:1801.06475.

Williams, L., Kessler, R.R., 2003. Pair Programming Illuminated. Addison-Wesley Professional.

Wohlin, C., Höst, M., Henningsson, K., 2003. Empirical research methods in software engineering. In: Empirical Methods and Studies in Software Engineering. Springer, pp. 7–23.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A., 2012. Experimentation in Software Engineering. Springer Science & Business Media.

Stefano Lambiase is a Ph.D. student in computer science at the University of Salerno. He is a member of the SeSa Lab (Software Engineering Salerno Lab), and he is advised by Professor. Filomena Ferrucci. He obtained the (magna cum laude) **Bachelor's Degree** defending a thesis on **Test Smell Detection and Suggestions** proposed by Prof. Andrea De Lucia. He got the (magna cum laude) **Master's Degree** on September 30th, 2021, defending a thesis regarding the human and social aspects of Software Development and Software Project Management.

Gemma Catolino is assistant professor at the Jheronimus Academy of Data Science, a collaboration between Tilburg University and Eindhoven Technical University. She is a member of the JADE Lab, supervised by Prof. Willem-Jan Van Den Heuvel. In 2020, she received the European Ph.D. Degree from the University of Salerno, advised by Prof. Filomena Ferrucci. She received (magna cum laude) the Master's Degree in Management and Information Technology from the University of Salerno (Italy) in 2016 defending a thesis on Software Quality Metrics, advised by Prof. Filomena Ferrucci. She got the Bachelor's Degree in Computer Science from the University of Molise in 2014 defending a thesis on Software Program Comprehension.

Fabiano Pecorelli is a Postdoctoral Researcher in the Jheronimus Academy of Data Engineering (JADE) lab at Jheronimus Academy of Data Science (JADS), Eindhoven University of Technology, The Netherlands. Before he has been a Postdoctoral Researcher in the Cloud Software Evolution and Assessment (CloudSEA) research group at Tampere University, Finland. In 2022, he received the European Ph.D. Degree, magna cum laude, in Computer Science from the University of Salerno, Italy, advised by Prof. Andrea De Lucia, working in the SeSa Lab (Software Engineering Salerno). In 2018, he received the Master's Degree, magna cum laude, in Computer Science from the University of Salerno, Italy, advised by Prof. Filomena Ferrucci. In 2016, he received the Bachelor's Degree, magna cum laude, in Computer Science from University of Salerno, Italy, advised by Prof. Andrea De Lucia.

Damian Andrew Tamburri is an Associate Professor at the Eindhoven University of Technology, the Jheronimus Academy of Data Science and Politecnico di Milano - DEIB, in s'Hertogenbosch, The Netherlands and Milan Italy, respectively. Damian completed his Ph.D. at VU University Amsterdam, The Netherlands in March 2014 "laude and mention" one year in advance of a standard Ph.D. His research interests lie mainly in Complex Software Architectures (with a focus on Big Data & Data-Intensive Architectures, Cloud & Microservices as well as Machine-Learning & Computational Intelligence Architectures), Complex Software Architecture Properties (with a focus on Privacy & Security and Social Sustainability), and Empirical Software Engineering (with a focus on Organisational, Social, and Societal aspects with Qualitative Methods, social-networks analysis as well as Machine-Learning). Damian has published over 100+ papers in either Journals such as the Transactions on Software Engineering (TSE) Journal, The ACM Computing Surveys (CSUR) Journal, the IEEE Software Magazine or top software engineering conferences (such as ICSE or FSE) and top software architecture conferences (such as ECSA or WICSA). Also, Damian has been an active contributor and lead research in many EU FP6, FP7, and H2020 projects, such as S-Cube, MODAClouds, SeaClouds, DICE, ANITA, DossierCLOUD, ProTECT, and more. In addition, Damian is IEEE Software and ACM TOSEM editorial board member, secretary of the TOSCA TC as well as secretary of the IFIP TC2, TC6, and TC8 WG on "Service-Oriented Computing".

Fabio Palomba is an Assistant Professor at the Software Engineering (SeSa) Lab (within the Department of Computer Science) of the University of Salerno. He received the European Ph.D. degree in Management & Information Technology in 2017. His Ph.D. Thesis was the recipient of the 2017 IEEE Computer Society Best Ph.D. Thesis Award.

His research interests include software maintenance and evolution, empirical software engineering, source code quality, and mining software repositories. He was the recipient of two ACM/SIGSOFT and one IEEE/TCSE Distinguished Paper Awards at the IEEE/ACM International Conference on Automated Software Engineering (ASE'13), the International Conference on Software Engineering (ICSE'15), and the IEEE International Conference on Software Maintenance and Evolution (ICSME'17), respectively, and Best Paper Awards at the ACM Computer Supported Cooperative Work (CSCW'18) and the IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER'18). In 2019 he was the recipient of an SNSF Ambizione grant, one of the most prestigious individual research grants in Europe, while in 2023 he has awarded with the prestigious IEEE Computer Society Technical Council of Software Engineering Rising Star Award, an early research career recognition assigned for the outstanding contribution to the field of code refactoring and code smells.

Prof. dr. **Willem-Jan van den Heuvel** is a full professor in Information Systems and managing director of the European Research Institute of Services Science (ERISS). His research interests are at the cross-junction of software service systems and business process management with an emphasis on (global) networked enterprises. In particular, his expertise revolves around the following major research themes: business process management, Big data analytics, service engineering (including service governance) and legacy system modernization.

Filomena Ferrucci has been a full professor of Computer Science since November 2011. She graduated (magna cum laude) in Information Science in 1990 from the University of Salerno. In 1995, she obtained a Doctorate in Applied Mathematics and Computer Science from the Federico II University of Naples. From 1995 to 2001, she was a researcher at the University of Salerno, where she held the position of associate professor from 2001 to 2011.

Her main research interests include software engineering (software environment development, cost/effort prediction, fault prediction, software metrics, web engineering, search-based software engineering, empirical software engineering, social and human aspects in software engineering with a particular focus on the contribution of women), human–computer interaction, software usability, visual languages, and computer science education. She has also worked on logic programming and e-learning. In recent years, she has focused on studying metrics to measure the functional size of software products and to estimate the effort required for their development. In her research, she has employed various techniques from machine learning, artificial intelligence, search-based methods, and evolutionary approaches.