Do I Belong? Modeling Sense of Virtual Community Among Linux Kernel Contributors

Bianca Trinkenreich
Northern Arizona University
Flagstaff, AZ, USA
bianca.trinkenreich@nau.edu

Klaas-Jan Stol
Lero, the SFI Research Centre for Software
University College Cork, Ireland
k.stol@ucc.ie

Anita Sarma
Oregon State University
Portland, OR, USA
anita.sarma@oregonstate.edu

Daniel M. German
University of Victoria
Victoria, Canada
dmg@uvic.ca

Marco A. Gerosa
Northern Arizona University
Flagstaff, AZ, USA
marco.gerosa@nau.edu

Igor Steinmacher
Northern Arizona University
Flagstaff, AZ, USA
igor.steinmacher@nau.edu

***Abstract*—The sense of belonging to a community is a basic human need that impacts an individual's behavior, long-term engagement, and job satisfaction, as revealed by research in disciplines such as psychology, healthcare, and education. Despite much research on how to retain developers in Open Source Software (OSS) projects and other virtual, peer-production communities, there is a paucity of research investigating what might contribute to a sense of belonging in these communities. To that end, we develop a theoretical model that seeks to understand the link between OSS developer motives and a Sense of Virtual Community (SVC). We test the model with a dataset collected in the Linux Kernel developer community (N=225), using structural equation modeling techniques. Our results for this case study show that intrinsic motivations (social or hedonic motives) are positively associated with a sense of virtual community, but living in an authoritative country and being paid to contribute can reduce the sense of virtual community. Based on these results, we offer suggestions for open source projects to foster a sense of virtual community, with a view to retaining contributors and improving projects' sustainability.**

***Index Terms*—sense of virtual community, belonging, open source, software developers, human factors, survey, PLS-SEM**

I. Introduction

The sustainability and long-term survival of Open Source Software (OSS) projects depend not only on attracting but, more crucially, retaining motivated developers [1]. The reasons behind a developer's decision to stay or leave an OSS project can depend on different intrinsic or extrinsic factors, including an individual's feelings of identity and belonging to the community [2]. Hagerty et al. defined a sense of belonging as *"the experience of personal involvement in a system or environment so that persons feel themselves to be an integral part of that system or environment"* [3]. The need to belong is a powerful, fundamental, and pervasive force that has multiple strong effects on emotional patterns and cognitive processes across all cultures and different types of people [4]. Maslow [5] positioned 'belonging' as a basic human need, and Hagerty et al. [6] posited that a sense of belonging represents a unique mental health concept. A sense of belonging is key to productivity, satisfaction, and engagement [4], and can help to avoid attrition [7]. In Science, Technology, Engineering, and Mathematics (STEM), a sense of belonging is strongly related to retention [8], especially for underrepresented groups [9].

The sense of belonging that members have towards others within a certain group is known as a *sense of community* [10]. The dimensions of a sense of community include feelings of membership and attachment to a group [11], a feeling that members matter to one another and to the group [12]. The concept of sense of virtual community (SVC) was developed by observing that virtual communities represent a new form of community, in which social relationships are predominantly forged in cyberspace [13].

Understanding SVC in OSS is relevant as it can influence the vitality and sustainability of a community [14], [15], and is linked to more satisfied, involved, and committed contributors [16]. Individuals who develop a psychological and relational contract with a community are supported by a state of being involved, rather than external factors such as earning something or climbing a career ladder and therefore tend to develop a deeper, reciprocal relationship with that community [10].

Since sustainability is a key concern for OSS projects, we must understand SVC in OSS communities. While several studies have investigated different motivations to contribute to OSS [1], [17]–[22], none have modeled how these factors can help or hinder in creating a sense of virtual community. Without a deeper understanding of how the different factors interplay to create a sense of community, strategies that aim to promote individual factors will likely be unsuccessful in creating a sustainable community. Understanding how different factors work together or against each other can help communities strategize how to retain their contributors. Therefore, in this paper, we ask the following research question:

Research Question: *How does a sense of virtual community develop in Open Source Software projects?*

We answer our research question by first developing a theoretical model of SVC grounded in prior literature (Sec. III).

We then evaluate our model through a sample (N=225) of Linux Kernel project contributors, using partial least squares structural equation modeling (PLS-SEM) (Sec. IV). The results of our analysis provide empirical support for part of our model, showing that *hedonism* (motivation that aims to maximize pleasure and fun and minimize pain [23]) and *social motives* (motivation that aims to maximize joint gains and others' gains [24]) have a positive association with a sense of virtual community, which can be weakened when contributors are *being paid* or are surrounded by an authoritative culture, i.e., national culture with a high index of power distance (Sec. V). We conclude the paper by discussing the implications of our findings, and threats to validity (Sec. VI).

II. Background

A. Sense of Virtual Community

While numerous definitions of the term 'community' exist, a common theme is that it involves human relationships based on some common characteristics [25]. The classical McMillan and Chavis [12] definition of 'Sense of Community' includes four characteristics: (1) feelings of membership (belonging to, and identifying with, the community), (2) feelings of influence (having an influence on, and being influenced by the community), (3) integration and fulfillment of needs (being supported by others in the community while also supporting them), and (4) shared emotional connection (relationships, shared history, and a 'spirit' of the community). *Virtual* communities typify a relatively new form of interaction whereby community members share information and knowledge in the virtual space for mutual learning, collaboration, or problem solving [13].

The development of OSS involves distributed problem solving within a virtual community [26]. Virtual communities are a particularly important type of virtual group because they are self-sustaining social systems in which members engage and connect with each other, developing a Sense of Virtual Community (SVC) [27]. The sense of community includes membership, identity, belonging, and attachment to a group that primarily interacts through electronic communication [11], [28], [29]. SVC has been tailored to virtual communities by deriving from McMillan's theory of sense of community [12]. The goal of measuring SVC is to assess the "community-ness" of a virtual community [11].

Community managers can assess and promote SVC to fulfill a core set of members' needs [30], so they feel they belong to a unique group. Such meaningful relationships are associated with increased satisfaction and communication with the virtual community, trust [31], and social capital in the project [32]. SVC has been shown to lead to an occupational commitment [33], and ultimately can help retain contributors and further attract potential newcomers [11], [34], who are critical to the sustainability of OSS projects [35].

B. Motivations to Contribute to Open Source Software

The software engineering literature suggests that, by managing developers' motivation and satisfaction, a software organization can achieve higher productivity levels and avoid turnover, budget overflows, and delivery delays [36]. Motivations for joining Open Source has been the topic of considerable research [1], [17]–[22]; motivations can be extrinsic or intrinsic. Extrinsic motivations are based on outside incentives that make people change their actions due to an external intervention [37]. As many companies, including Microsoft, Google, and IBM, hire or sponsor OSS contributors [38], career ambition and payment have become common extrinsic motivations [39]. However, intrinsic motivations also explain much of contributors' motivations [22], moving a person to act for fun or enjoy a challenge, kinship, altruistic reasons, or ideology, rather than in response to external pressures or rewards [40].

Previous research showed that several forces influence the decision of an OSS contributor to join, remain, or leave an OSS project [41]–[43]. Despite the extensive attention this topic has received, there are still no studies investigating how OSS contributors driven by different motivations develop a sense of virtual community. We argue that understanding how a sense of virtual community develops in OSS involves understanding the relationship between individual characteristics and motivations and the resulting community-related feelings.

III. Theory Development

Feelings of belonging in an online community can be influenced by several individual characteristics and factors of the surrounding environment [44]. In the education literature, researchers [45], [46] found associations between students' sense of belonging and a range of motivational variables. Motivational factors can be regarded as expectations related to the interaction with a virtual community (answering *why* users behave). Integration and fulfillment of needs refer to the idea that common needs, goals, and beliefs provide an integrative force for a cohesive community that can meet collective and individual needs. Thus, meeting members' needs is a primary function of a strong community [12].

Individuals who develop a psychological relationship contract with a community because it is focused on a state of being involved—rather than earning something or getting somewhere—tend to develop a sense of community [10]. Previous research on online communities also showed that individuals who are driven by *social motives* [47] tend to develop a sense of virtual community [28], [48]. Based on the Fundamental Social Motives Inventory, we included both kinship and altruism as social motives [47] and propose the following hypothesis:

Hypothesis 1 (H1). Open Source contributors motivated by social reasons have a higher sense of virtual community.

Most of the respondents in Gerosa et al.'s study (91%) agreed (or strongly agreed) that they contribute to OSS for entertainment (fun) [22]. Hedonic motivation is a type of motivation that aims to maximize pleasure and fun and minimize pain. It is an umbrella term that includes hedonic expectancy, perceived enjoyment, and playfulness [23]. Considering that expectations of enjoyable experiences, feelings of amusement, and being mentally or intellectually stimulated by interactions

are associated with a sense of virtual community [13], [15], and that changes in the perceived fulfillment of their entertainment needs can determine the change of their sense of virtual community [30], we propose the following hypothesis:

Hypothesis 2 (H2). Open Source contributors motivated by hedonic reasons have a higher sense of virtual community.

It is known that some open source contributors have a strong ideological basis for their actions [49], believing, for example, that source code should be freely available. Recently, Gerosa et al.'s study showed that, however, ideology is not a popular motivation—especially for young contributors [22].

Historically, the group-based morality of 'fighting' a shared dominant opponent incites the sense of virtual community among contributors [50]. This feeling was quite common in the 1990s, when big corporations characterized Open Source as 'communism' [51] and Linux as a 'cancer' [52]. Besides ideology, we include reciprocity in moral motives, as it represents the moral desire of contributors who aim for social justice by giving back to the community [53].

According to the Social Identity theory [54], sharing a moral vision is positively associated with feelings of belonging. Moreover, a homogeneous ideology throughout a religion was shown as being positively associated with a sense of virtual community [55]. Hence, we posit that:

Hypothesis 3 (H3). Open Source contributors motivated by moral reasons have a higher sense of virtual community.

Motivations may not always be strong enough to sustain an OSS contributor's participation [56]. Motivations may vary for different groups of people, depending on contextual factors. This implies the existence of moderating factors that change the relationship between motivations and a sense of virtual community. Cognitive Evaluation Theory suggests that feelings of autonomy are positively associated with intrinsic motivations and belonging, while tangible rewards negatively affect intrinsic motivating factors [57].

We evaluated the role of a feeling of autonomy using the variable of *power distance* from Hofstede's framework of Country Culture [58] as a proxy; a lower power distance would reflect in higher autonomy. We also evaluated the exposition to tangible rewards using the variable *is paid*.

People in societies exhibiting a large degree of power distance tend to accept a hierarchical order [59]. In high power distance cultures (where a high power differential between individuals is accepted and considered normal), information flows are usually constrained by hierarchy [58]. As an important cultural value describing the acquiescent acceptance of authority, power distance has received increasing attention in many domains [60], [61].

Prior research showed that, when surrounded by cultures with a high degree of power distance, students reported a lower sense of belonging to their school [62]. Therefore, in hierarchical cultures, leaders need control over the information flow, and the desire to restrict autonomy and access to critical information by lower-level team members could lead to significant organizational barriers to sharing knowledge and working in a community [63]. Thus, we define the following moderation hypotheses:

Hypothesis 4a (H4a). Power distance moderates the association between Open Source contributors' social motives and their sense of virtual community.

Hypothesis 4b (H4b). Power distance moderates the association between Open Source contributors' hedonic motives and their sense of virtual community.

Hypothesis 4c (H4c). Power distance moderates the association between Open Source contributors' moral motives and their sense of virtual community.

The traditional notion that OSS developers are all volunteers is now long outdated; many OSS contributors are currently paid, usually employed by a company, to contribute [39], [64], [65]. Indeed, many Linux Kernel contributors are paid to make their contributions, compensated by firms that have business models relying on the Linux Kernel [66]–[68].

In contrast to traditional paid software development work, and despite its benefits to OSS contributors, introducing financial incentives in OSS communities create complex feelings among OSS developers [69]. For example, developers on the Debian project expressed negative emotion because they felt payment went against the project's espoused values [70]. On the other side, not receiving pay for their work to support their livelihoods can frustrate OSS developers and affect their contributions [69].

Despite compensation, OSS contributors may be driven towards a project by both simultaneous feelings of belonging (intrinsic) and payment (extrinsic) [1], [39]. Nevertheless, there is no research examining the complex impact of receiving payment on intrinsic factors associated with SVC. As many OSS developers are currently paid, we would expect that the behavior of those who are paid and those who are not (volunteers) would diverge. Hence, we propose the following three moderating hypotheses:

Hypothesis 5a (H5a). Being paid moderates the association between Open Source contributors' social motives and their sense of virtual community.

Hypothesis 5b (H5b). Being paid moderates the association between Open Source contributors' hedonic motives and their sense of virtual community.

Hypothesis 5c (H5c). Being paid moderates the association between Open Source contributors' moral motives and their sense of virtual community.

IV. Research Design

The research design is summarized in Fig. 1. We conducted a survey among Linux Kernel contributors to evaluate our theoretical model. We studied one specific community to avoid confounding factors related to differences that each OSS community can pose. Introduced in 1991, the Linux Kernel represents one of the largest and most active OSS projects [71], boasting over ten million source lines of code and more

than 20,000 contributors from different countries and cultural backgrounds, including volunteers and paid developers from more than 200 companies [72], [73]. Linux Kernel's is impact is perceived in terms of processes and infrastructure tools that emerged from the community [73]. While the Linux Kernel Mailing List is known for its uncivil comments and toxic discussions that tend to discourage people from joining the community [74], community leaders aim to change the project's image and increase the sense of community among members. We closely collaborated with contributors and maintainers of the Linux Foundation involved with Linux Kernel, who had a crucial role in designing the data collection instrument and reaching out to potential participants. They engaged in several meetings with the team and reviewed the items of the questionnaires to provide their feedback, making sure that the instrument was appropriate for the study goals. They also distributed the survey to the Linux kernel community, playing an essential role in recruiting the participants for this study.

We used Partial Least Squares–Structural Equation Modeling (PLS-SEM) to analyze the relationships between motivations [75] and a sense of virtual community. SEM is a second-generation multivariate data analysis method; a recent survey (which also provides an introduction to the method) indicates that PLS-SEM has been used to study a variety of phenomena in software engineering [76]. SEM facilitates the simultaneous analysis of relationships among constructs, each measured by one or more indicator variables. The main advantage of SEM is being able to measure complex model relationships while accounting for measurement error when using latent variables (e.g., Sense of Virtual Community). PLS-SEM has previously been used in literature to evaluate factors that impact the sense of belonging in other contexts [77], [78].

In the following, we discuss the measurement model (i.e., operationalization of constructs), data collection, and analysis.

A. Measurement model

The theoretical model comprising the hypotheses is based on a number of theoretical concepts; some of the concepts may be directly observed (e.g., 'is paid'), but others cannot (e.g., sense of virtual community)—these concepts are represented as *latent* variables. A latent variable cannot be directly measured or observed but instead is measured through a set of indicators or manifest variables. For the latent variables in this study, we adapted existing measurement instruments.

Sense of Virtual Community: We adapted items about feelings of a virtual community from Blanchard's [11] instrument of sense of virtual community to better fit with the context of OSS contributions. In collaboration with a group of Linux Kernel community managers, we analyzed the items proposed by Blanchard et al. [33] and decided to use a subset of questions to compose a shorter version of the instrument to cover the dimensions of SVC. The subset was discussed synchronously by researchers and managers, and the items were considered appropriate and meaningful to represent SVC to the Linux Kernel contributors.

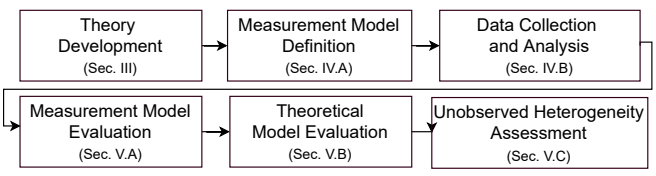

Fig. 1. Research Design and Phases for Results' Analysis

Intrinsic Motivations: We created items based on Gerosa et al. [22]'s instrument, which was built upon previous studies of motivations in OSS [18]–[20]. Following the community managers' request to make the questionnaire as short as possible, we grouped the intrinsic motivation factors from Gerosa et al.'s study [17] into three factors: 1. Social Motives (Kinship and Altruism) [47]; 2. Hedonic Motives (Joy and Fun) [23]; and 3. Moral motives (Ideology and Reciprocity) [53].

English Confidence: We reused four questions about the self-confidence of fluency levels during interactions by speaking and writing in technical and non-technical situations from a previous survey [79].

Power Distance: We asked in which country the respondent lived and used the value per country proposed by Hofstede's framework [58] as the Power Distance dimension in the model.

For the demographic questions, we adapted questions from surveys used in OSS communities to ask about tenure, self-identified gender, and compensation [80], [81]. Tenure was shown in 10-year slices in Table I for presentation purposes, but was included as a continuous variable in our analysis. We provided a dropdown list of years since 1991 (the year when the Linux Kernel was launched) for respondents to inform the year they started contributing to the project.

B. Data Collection and Analysis

We administered the online questionnaire using LimeSurvey, a leading Open Source survey software, to survey Linux Kernel contributors. We explored their motivations and their sense of virtual community. Our online appendix provides the instrument and replication package [82].

1) Designing the instrument: The questions were discussed during 12 online meetings between October 2020 and February 2021 in a group of five researchers experienced in OSS and survey studies and two members of the Linux community. The group discussed each of the questions until reaching a consensus. The questionnaire provides informed consent followed by closed questions about the importance of each motivation factor as a reason to contribute to the Linux Kernel and questions about their feelings about the Linux Kernel community. Finally, we added demographic questions aiming to segment analysis and understand the phenomenon considering the different dimensions of our participants, and an open question for additional comments. Investigating the forces that push people with different individual characteristics can help us better support a diverse community [22]. The questions included gender identity, financial compensation, starting the year at Linux Kernel, and country of residence.

2) Piloting the questionnaire: After designing the instrument, we piloted the questionnaire before distributing it to

the population of interest. In the first round, our collaborators from the Linux Foundation recruited a group of Linux Kernel maintainers, who answered the questionnaire and provided feedback. Although the feedback was overall positive, maintainers suggested reverse-coding some items for the SVC construct, i.e., items worded as negative statements (low score indicates agreement). Inverse, negative, or reverse-coded items can be defined as those having a directionality as opposed to the logic of the construct being measured [83]. We agreed with the suggestion and inverted two of the four items as this can help to mitigate acquiescence bias [84], which can occur when participants tend to agree with statements without regard for their actual content or due to laziness, indifference, or automatic accommodation to a response pattern [85]. The item *I feel at home in the group* was changed to *I don't feel at home in the group*. We inverted and adapted the question *I feel that my contribution is valued* to *I want to contribute more but I do not feel valued.*

After the first pilot, we ran two more pilot sessions with two researchers who are Open Source contributors. We used the think-aloud method [86] and recorded their suggestions while answering the questions. We made minor changes to the questionnaire and increased font size for better readability on different devices.

3) Recruiting participants: The Linux Foundation contributors who collaborated in this study took the lead in recruiting the participants from the Linux Kernel. They reached out to maintainers and contributors using mailing lists from the Linux Kernel community and interacted with maintainers to ask for engagement. Further, we presented the study motivation during the first day of the Linux Plumbers annual conference (https://lpc.events/event/11/), inviting participants to answer the questionnaire. The survey was available between August 12 and September 21, 2021.

4) Sample Analysis: We received 318 responses and carefully filtered the data to consider only valid responses. Respondents who did not complete the whole questionnaire were dropped (n=26). We also dropped the participants who answered "I'm not sure" to any of the items included with the five-point Likert scale for Motivations (n=16) and Sense of Virtual Community (n=51). In addition to the 5-point Likert scale, we included a 6th alternative ("I'm not sure") for participants who either preferred not to, or did not know how to, answer the question (to avoid forcing them), which is different from being neutral—based on the dissonance between ignorance and indifference [87], [88]. Therefore, we considered these responses as missing data. The efficacy of imputation methods has not yet been validated when using FIMIX-PLS; Sarstedt et al. [89] recommend removing all responses with missing values for any question before segmenting data into clusters.

After applying these filters, we retained 225 valid responses from residents of five different continents with a broad tenure distribution. The majority identified as men (84.4%), from Europe (52.9%), and paid to contribute (65.4%), matching previously reported distributions of OSS contributors [81].

TABLE I
DEMOGRAPHICS OF THE LINUX KERNEL RESPONDENTS (N=225)

Attribute	N	Percentage
Gender		
Man	190	84.4%
Woman	21	9.4%
Non-binary	5	2.2%
Prefer not to say	8	3.6%
Prefer to self describe	1	0.4%
Continent of Residence		
Europe	119	52.9%
North America	68	30.2%
Asia	32	14.2%
South America	6	2.7%
Starting year at the Linux Kernel		
2000 or earlier	28	12.4%
Between 2001 and 2010	77	34.2%
Between 2011 and 2021	120	53.4%
Current Compensation for the Linux Kernel contributions		
Paid	145	64.4%
Unpaid (volunteer)	80	35.6%

Table I presents a summary of the demographics.

To establish an appropriate sample size, we conducted a power analysis using the free G*Power tool [90] (see online appendix for details). The maximum number of predictors in our model is six (three motivations and three control variables to SVC). This calculation indicated a minimum sample size of 62 and our sample of 225 exceeded that number considerably.

We used the software package SmartPLS version 4 for the analyses. The analysis procedures for PLS-SEM comprise two main steps, with tests and procedures in each step. The first step is to evaluate the measurement model, which empirically assesses the relationships between the constructs and indicators (see Sec. V-A). The second step is to evaluate the theoretical (or structural) model that represents the hypotheses (see Sec. V-B).

V. ANALYSIS AND RESULTS

In this section, we describe our results, which include the evaluation of the measurement model (Sec. V-A), followed by the hypotheses evaluation in the structural model (Sec. V-B), both computed through our survey data. We assess the significance of our model by following the evaluation protocol proposed by previous research [76], [91] to make results consistent with our claims. The path weighting scheme was estimated using SmartPLS 4 [92].

Two tests are recommended to ensure that a dataset is suitable for factor analysis [93], [94]. We first conducted Bartlett's test of sphericity [93] on all constructs. We found a p-value $< .01$ (p values less than .05 indicate that factor analysis may be useful). Second, we calculated the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy. Our result (.81) is well above the recommended threshold of .60 [94].

A. Evaluation of the Measurement Model

Some of the constructs in the theoretical model (see Fig. 2) are modeled as latent variables, i.e., measured by more than

one observed variable (i.e., item/question on the survey). The first step in evaluating a structural equation model is to assess the soundness of the measurement of these latent variables—this is referred to as evaluating the 'measurement model' [91]. We present the assessment of several criteria.

1) Convergent Validity: First, we assess whether the questions (indicators) that represent each latent variable are understood by the respondents in the same way as they were intended by the designers of the questions [95], i.e., we assess the convergent validity of the measurement instrument. The assessment of convergent validity relates to the degree to which a measure correlates positively with alternative measures of the same construct. Our model contains two latent variables, both of which are reflective (not formative), as functions of the latent construct. Changes in the theoretical, latent construct are reflected in changes in the indicator variables [91].

We used two metrics to assess convergent validity: the Average Variance Extracted (AVE) and the loading of an indicator onto its construct (the outer loading).

The AVE is equivalent to a construct's communality [91], which is the proportion of variance that is shared across indicators. A reflective construct is assumed to reflect (or "cause") any change in its indicators. The AVE should be at least .50, indicating that it explains most of the variation (i.e. 50% or more) in its indicators [91]. This variance is indicated by taking the squared value of an indicator's loading. As Table II shows, all AVE values for both latent constructs in our model are above this threshold of .50.

TABLE II
INTERNAL CONSISTENCY RELIABILITY

	Cronbach's α	CR	AVE
English Confidence	.932	.951	.830
Sense of Virtual Community	.767	.839	.511

A latent variable is measured by two or more indicators; indicators with loading below .4 should be removed because this implies that a change in the latent construct that it purportedly represents (or 'reflects') does not get reflected in a sufficiently large change in the indicator [91]. Outer loading of .7 is widely considered sufficient, and .6 is considered sufficient for exploratory studies [91]. We followed an iterative process to evaluate the outer loading of the latent constructs; the indicators of the construct English Confidence all exceeded .7, but SVC had two indicators below .7. We removed the SVC indicator which had a loading below .4 (*svc6: a majority of the developers and I want the same thing*). After removing this indicator, the AVE value of SVC (now with five indicators) increased from .44 to .51 and all outer loadings were above .60.

2) Internal Consistency Reliability: Second, we verified how well the different indicators are consistent with one another and can reliably and consistently measure the constructs, i.e., we assess the internal consistency reliability. A high degree of consistency means that the indicators refer to the same construct. There are several tests to measure internal consistency reliability. We performed both the Cronbach's α and Composite Reliability tests; Cronbach's α frequently shows lower values, whereas the Composite Reliability (CR) is a more liberal test, which sometimes overestimates the values [91]. A desirable range of values for both Cronbach's α and CR is between .7 and .9 [91]. Values below .6 suggest a lack of internal consistency reliability, whereas values over .95 suggest that indicators are too similar and thus not desirable. The Cronbach α and CR values for both latent variables fell in the range .75–.95; only the CR for English Confidence was slightly over at .951. AVE values were both higher than .50.

3) Discriminant Validity: Third, we verified whether each construct represents characteristics not measured by other constructs, i.e., we assessed the discriminant validity of the instrument (indicating the distinctiveness of the constructs). Our model includes two latent variables (SVC and English Confidence). A primary means to assess discriminant validity is to investigate the Heterotrait-monotrait (HTMT) ratio of correlations, developed by Henseler et al. [96]. The discriminant validity could be considered problematic if the HTMT ratio exceeds .9 [96]; some scholars recommend a more conservative cut-off of .85 [91]. The HTMT ratio between the two latent constructs (SVC and English Confidence) was .24. We also assessed the cross-loadings of indicators and the Fornell-Larcker criterion. Items should only load onto their 'native' construct, the one they purportedly represent (Table III).

TABLE III
CROSS LOADINGS OF THE RETAINED INDICATORS ON THE CONSTRUCTS

		SVC	Engl. Conf.
svc1	I don't feel at home in the group	**.740**	.191
svc2	I feel that I belong to the group	**.822**	.231
svc3	If I have a problem, I know members in the group who I can ask for help	**.662**	.063
svc4	I want to contribute more but I do not feel valued	**.583**	.113
svc5	A majority of the developers in the group know me	**.722**	.242
eng1	Participating in a non-technical discussion on the email list	.180	**.881**
eng2	Performing Reviews	.216	**.925**
eng3	Speaking with others (face to face)	.226	**.890**
eng4	Participating in technical discussions on the email list	.278	**.939**

For the sake of completeness, we report the Fornell-Larcker procedure in the online appendix [82].

B. Evaluation of the Theoretical Model

We now evaluate and discuss the theoretical model, which includes the evaluation of the hypotheses.

1) Assessing Collinearity: Our theoretical model has three different exogenous variables of intrinsic motivations, the moderators 'Compensation' and 'Power Distance,' and the control variables 'English Confidence,' 'Gender,' and 'Tenure.' We hypothesized that the exogenous variables are associated with the endogenous variable Sense of Virtual Community. To ensure that the three exogenous constructs are independent, we calculate their collinearity using the Variance Inflation Factor

(VIF). A widely accepted cut-off value for the VIF is 5 [91], and in our model, all VIF values are below 5.

2) Path Coefficients and Significance: PLS does not make strong assumptions about the distribution (such as a Normal distribution) of the data, so parametric tests of significance should not be used. To evaluate whether path coefficients are statistically significant, PLS packages employ a bootstrapping procedure. This involves drawing a large number (usually five thousand) of random subsamples with replacement. The replacement is needed to guarantee that all subsamples have the same number of observations as the original data set. The PLS path model is estimated for each subsample. From the resulting bootstrap distribution, a standard error can be determined [91], which can subsequently be used to make statistical inferences. The mean path coefficient determined by bootstrapping can differ slightly from the path coefficient calculated directly from the sample; this variability is captured in the standard error of the sampling distribution of the mean.

Table IV shows the results for our hypotheses, including the mean of the bootstrap distribution (*B*), the standard deviation (*SD*), the 95% confidence interval, and the p-values.

The path coefficients in Fig. 2 and Table IV are interpreted as standardized regression coefficients, indicating the direct effect of a variable on another. Each hypothesis is represented by an arrow in the diagram in Fig. 2. For example, the arrow pointing from Hedonic Motives to SVC represents H2. Given its positive path coefficient (0.421), Hedonic Motives are positively associated with SVC. The path coefficient is 0.421; this means that when the score for "Hedonic" motives increases by one standard deviation unit, the score for "Sense of Virtual Community" increases by 0.421 standard deviation unit (the standard deviation is the amount of variation of a set of values).

Based on these results, we found support for Hypotheses H1 (p=.002), H2 (p=.000), H4a (p=.045), and H5b (p=.023). Hypothesis H3 was not supported, nor were H4b, H4c, H5a, or H5b (all p values $> .2$). The three control variables all have significant associations with SVC: English confidence, gender, and tenure ($p < .05$).

TABLE IV
STANDARIZED PATH COEFFICIENTS, STANDARD DEVIATIONS, CONFIDENCE INTERVALS, AND P VALUES

	B	SD	95% CI	*p*
H1 Social motives→SVC	.25	.11	(.04, .46)	**.002**
H2 Hedonic motives→SVC	.42	.11	(.19, .64)	**.000**
H3 Moral motives→SVC	−.14	.11	(-.36, .08)	.215
H4a Power distance × social motives → SVC	−.15	.08	(−.31,−.01)	**.045**
H4b Power distance × hedonic motives →SVC	−.05	.07	(−.18, .11)	.477
H4c Power distance × moral motives→SVC	.04	.07	(−.10, .17)	.539
H5a is Paid × social motives→SVC	−.05	.07	(−.32, .20)	.696
H5b is Paid × hedonic motives→SVC	−.33	.14	(−.62, .05)	**.023**
H5c is Paid × moral motives→SVC	.12	.14	(−.17, .36)	.404
Gender minorities→SVC	−.49	.17	(−.81,−.14)	**.004**
English confidence→SVC	.13	.01	(.01,.25)	**.025**

3) Coefficient of Determination: We assessed the relationship between constructs and the predictive capabilities of the model. The R^2 values of the endogenous variable in our model (SVC) was 0.4, which is considered weak-moderate [91], [97].

We also inspected Stone-Geisser's Q^2 [98] value, which is a measure of external validity, as an indicator of the model's predictive relevance [91], and can be obtained through a so-called blindfolding procedure (available within the SmartPLS software). Blindfolding is a resampling technique that omits certain data, predicts the omitted data points, then uses the prediction error to cross-validate the model estimates [99]. Q^2 values are calculated only for the SVC, the reflective endogenous construct of our model, with a value of .17. Values larger than 0 indicate the construct has predictive relevance, while negative values show the model does not perform better than the simple average of the endogenous variable would do.

The Standardized Root Mean Square Residual (SRMR) is a common fit measure that is appropriate to detect misspecification of PLS-SEM models [76]. SRMR is the square root of the sum of the squared differences between the model-implied

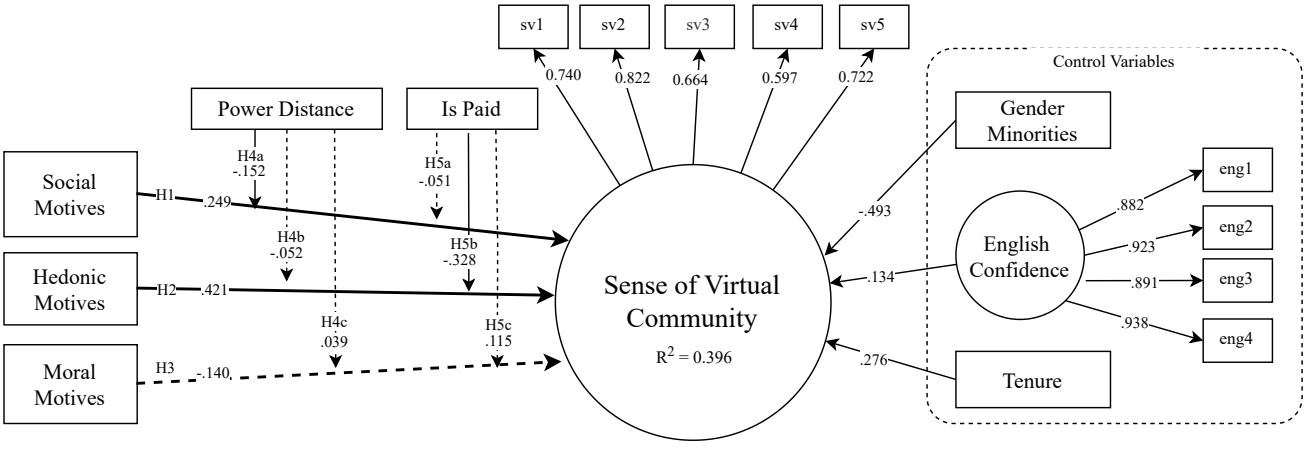

Fig. 2. Item loadings and path coefficients ($p < 0.05$ indicated by a full line). Non-significant links are indicated with a dashed line

and the empirical correlation matrix, or the Euclidean distance between the two matrices [100]. A value of 0 for SRMR would indicate a perfect fit, and a cut-off value of 0.08 is considered adequate [101]. Our results suggest a good fit of the empirical data in the theoretical model (SRMR = 0.06).

4) Moderating Factors: We examined our data to determine if the impact of each intrinsic motivation on a sense of virtual community would change when they are exposed to a high Power Distance culture or when they are financially compensated to contribute.

Only significant results at 0.05 are reported, with confidence intervals calculated through bootstrapping.

- Power Distance Country Culture: Being surrounded by a high power distance culture, in which leaders impose a high level of control and restrict the information flow [58], has been reported to negatively affect the sense of virtual community [63]. We did not find significant correlations between Power Distance and SVC for hedonic or moral motivations. Still, we found it for social motivations, which has a moderating effect on our model. Hence, we found support for H4a but do not support H4b and H4c.
- Compensation: Being paid to contribute reduces the sense of virtual community of contributors driven by hedonic motivations but not by social motivations and neither by moral motivations. Hence, we found support for H5b but rejected H5a and H5c.

Fig. 3 presents an interaction diagram showing the simple slopes for the relationship between the exogenous variable Social Motives and the endogenous variable SVC. All three slopes are positive, indicating a positive relationship; the top line (in green) is at +1 standard deviation of the moderator, Power distance; the bottom slope (in red) is at −1 standard deviation of the moderator. The middle slope (in blue) represents the relationship at the mean level of Power distance. The figure shows that given a higher level of Power distance, the relationship between social motives and SVC is *dampened* (flatter), whereas with lower levels of Power distance, the relationship is *strengthened* (steeper).

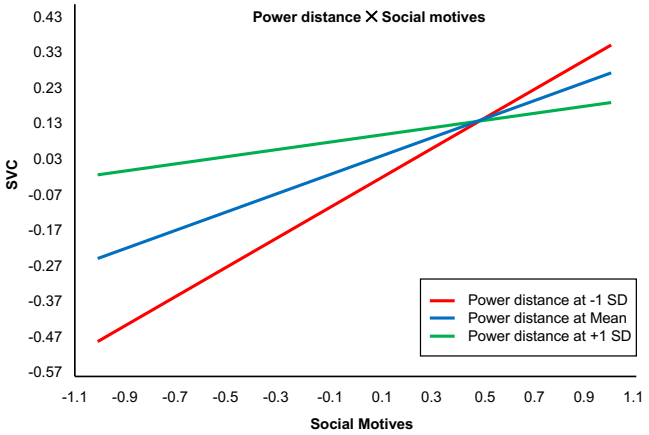

Fig. 3. Power distance as a moderator of Social motives → SVC

5) Control Variables: We also examined our data to determine if being part of gender minorities, tenure, or English Confidence could strengthen or weaken the sense of virtual community. We found that participants who identify as gender minorities tend to have a lower sense of virtual community, while participants with higher tenure and English Confidence reported a higher sense of virtual community.

C. Cluster Analysis: Detecting Unobserved Heterogeneity

While moderators and context factors capture *observed* heterogeneity (see Sec. V-B4), there may also be *unobserved heterogeneity*, or *latent classes* of respondents, the presence of which could threaten the validity of results and conclusions [89]. Latent classes of respondents refer to some groupings of respondents on one or more criteria that were not measured. The hypothesis results may differ for different groups.

We adopted Becker et al.'s approach [102], which jointly applies PLS-POS and FIMIX algorithms to identify latent classes. In Step 1, we used the minimum sample size for the maximum number of segments and ran FIMIX to find the optimal number of segments. In Step 2, we ran PLS-POS to compute the segmentation. In Step 3, we ran a multi-group analysis (PLS-MGA) and evaluated whether the segments were distinguishable. In Step 4, we checked if the resulting groups were plausible. We discuss the steps in more detail.

In Step 1, we assessed the maximum number of segments according to the minimum sample size (see Sec. IV-B4). Dividing the sample size (225) by the minimum sample size (62) yields a theoretical upper bound of three segments; each segment should satisfy the minimum sample size. We ran FIMIX for one (meaning, treating the original sample as a single segment), two, and three segments [89]. The results were compared using several different retention criteria (see Table V) [89]. For each criterion, the optimal solution is the number of segments with the lowest value (in *italics* in Table V), except in terms of criterion 'EN,' where higher values indicate a better separation of segments. Sarstedt et al. [103] argue that researchers should start the fit analysis by jointly considering the *combination* of modified Akaike's Information Criterion with factor 3 (AIC3) and Consistent AIC (CAIC) (Group 1 in Table V): when *both* criteria suggest the same number of segments, this result is likely to be most appropriate. As this is not the case here (AIC3 suggests 3 segments, CAIC suggests 1 segment), a second evaluation considers whether modified Akaike's Information Criterion with factor 4 (AIC4) and Bayesian Information Criterion (BIC) suggest the same number of segments (Group 2 in Table V). Again, this is not the case as AIC4 suggests 3 segments, and BIC suggests 1 segment). The third evaluation (Group 3) considers the joint analysis of Akaike's Information Criterion (AIC) and Minimum Description Length with factor 5 (MDL5); first, consider the number of segments indicated by the lowest values of AIC (3 segments) and MDL5 (1 segment). The appropriate number of segments should be lower than suggested by AIC (because it tends to overestimate) and higher than the number of segments suggested by MDL5 (because it tends

TABLE V
ESTABLISHING ADEQUATE NUMBER OF SEGMENTS

Group	Criterion	1-Segment	**2-Segment**	3-Segment
1	AIC3	574.153	540.241	*517.508*
	CAIC	*625.395*	646.141	678.065
2	AIC4	589.153	571.241	*564.508*
	BIC	*610.395*	615.141	631.065
3	**AIC**	559.153	509.241	***470.508***
	MDL5	***935.361***	1286.737	1649.292
4	EN	0	0.869	0.821

to underestimate). Hence, this combination suggests that a 2-segment solution is appropriate because 2 is lower than the 3 suggested by AIC and higher than the 1 suggested by MLD5. The value of EN is highest for the 2-segment solution.

In Step 2, we evaluated the segment sizes of the 2-segment solution and proportions of data to check whether groups were substantial or candidates for exclusion. A segment is not substantial if its size is considerably lower in proportion (e.g., a 2% segment size) or below the minimum sample size [102]. The 2-segment solution divided the dataset into groups with 158 (70.2%) and 67 (29.8%) observations; both considerable portions and larger than the minimum sample size [102].

In Step 3, we ran a multi-group analysis (PLS-MGA) with parametric tests to verify whether the segments were distinguishable [102], i.e., whether the results were different for the two segments. We found significant differences in hypotheses H4b-c, H5a-c, and on the control variables Tenure and English Confidence (see Table VI), thus, we conclude these two segments represent two different groups of respondents. Both groups presented R^2, goodness-of-fit (GoF), and SRMR [89] equal or more favorable than the original model. The values of the path coefficients and the explained variance of the endogenous variable SVC are shown in Table VI, which presents the results for the two segments, as well as the original estimates (see column *B* in Table IV).

TABLE VI
GROUP PATHS COEFFICIENTS: COEFICIENTS IN BOLD ARE SIGNIFICANT; LINES IN GRAY SHOW SIGNIFICANT DIFFERENCE BETWEEN SEGMENTS

	2-segment solution		Orig.
	Seg. 1 Hedonic	Seg. 2 Social	All
Sample size (N)	158	67	225
Coefficient of determination (R^2)	.57	.94	.40
H1 Social motives → SVC	−.04	**.22**	**.25**
H2 Hedonic motives→SVC	**.31**	.06	**.42**
H3 Moral motives→SVC	−.03	**−.23**	−.14
H4a Power distance × social mot.→ SVC	−.10	**−.24**	**−.15**
H4b Power distance × hedonic mot.→SVC	**−.07**	.14	−.05
H4c Power distance × moral motives→SVC	−.02	.22	.04
H5a is Paid × social motives→SVC	**.49**	**−.61**	−.05
H5b is Paid × hedonic motives→SVC	**−.50**	**.50**	**−.33**
H5c is Paid × moral motives→SVC	**−.15**	.32	.12
Gender minorities→SVC	**−.70**	**−.92**	**−.49**
English confidence→SVC	−.15	**.88**	**.13**
Tenure→SVC	**.43**	−.05	**.28**

In Step 4, we examined that groups were "plausible" [102] by explaining the different segments (highlighted in gray in Table VI) to label the segments. This labeling is somewhat speculative and not definitive, not dissimilar to the labeling of emergent factors in exploratory factor analysis. Given that for Segment 1 only Hedonic motives are significant, we posit that this segment represents *Hedonists* (*B*=.31); for Segment 2, we find that Social motives are significant (*B*=.22), thus we label Segment 2 as *Socially Motivated*. We note that moral motives were not significant in the original analysis (see column 'Orig.'), but this did become significant with a negative coefficient (*B*=−0.23) for Segment 2. For hedonists (Seg. 1), tenure (*B*=.43) is positively associated with SVC. When social motives are associated with SVC (Seg. 2), English Confidence positively affects SVC (*B*=.88). Both hedonists (*B*=−0.50) and socially motivated (*B*=−0.61) contributors have the association with SVC weakened when they are paid. Both groups showed that being part of a gender minority is associated with less SVC.

VI. DISCUSSION

In this study, we developed a theoretical model grounded in psychology literature to map the relationship between a Sense of Virtual Community and intrinsic motivations in OSS. The theoretical model includes a number of salient factors that have been shown to be important in belonging to an online community in general but not yet within the OSS domain. Over the past two decades, the nature of OSS communities (as a specific type of online community) has changed; traditionally, OSS was male-dominated and primarily volunteer-based, but being paid to contribute is now common, and increasingly we observe the participation of what we refer to as "minorities" in the broadest sense of the word, including women [104]. Our analysis highlights a number of key findings and implications; as we discuss these quantitative results, we bring exemplar quotes from the respondents' responses to the final open question of the survey to illustrate the discussion.

H1. Social motives → SVC: Social motives have a positive association with SVC. The intrinsic social motivations of kinship and altruism are positively associated with a sense of virtual community in OSS. This finding was corroborated by one of our respondents in the final open question, who associated SVC with social motivations: *"I did not fit in, in a big way. I was never able to create enough social capital to make networking effective, no matter who I tried to connect with."* Another respondent mentioned *"not being able to relate to colleagues* and named their perceived lack of SVC as *"a sense of otherness that never goes away."* However, the cluster analysis (Sec. V-C) indicated Segment 1 (which we labeled 'hedonic') to be non-significant, but Segment 2 (labeled 'social') is significant. We also found that for the 'socially motivated' English confidence is much more strongly related (*B*=.88 instead of .13) to SVC. This is intuitive because socially motivated people seek interaction, and English is the primary language within the Linux Kernel community.

H2. Hedonic Motives → SVC: Hedonic motives have a positive association with a Sense of Virtual Community. OSS

communities should seek to prevent toxic and other types of undesirable behavior that might reduce contributors' enjoyment; communities could also consider setting more clear community codes of conduct [74], [105], [106]. The cluster analysis showed that when only hedonism (not social motives) is associated with SVC (Seg. 1), Tenure is also associated with SVC. Hedonic-motivated contributors from our sample are also the ones who have longer tenure associated with SVC. Those contributors may have surpassed the initial barriers [107] and find enjoyment, or as mentioned by another respondent: *"It is therapeutic. When I feel bad about myself, [..] it calms me down emotionally to do Kernel development when I feel like that."*

H3. Moral Motives → SVC: The cluster analysis does not support H3. While social motives are positively associated with SVC (Seg. 2), moral motives are negatively associated with and reduce SVC. The first association is expected and not surprising [47]. People motivated by kinship or because they are happy to help others are keener to be part of the team and feel good in a community [28], [48]. Interestingly, the SVC presented a negative association with moral motivation. We argue that people motivated by ideological reasons may contribute regardless of how they feel about belonging there. They do it because they feel it is the right thing to do, either because it is the most ethical choice, as advocated by the Free Software Foundation (https://www.fsf.org/), or because they have a moral debt to the software project that they use, so they pay back [53]. Future research can investigate how strong the ties between these people and the community are and what roles they play in building SVC for the rest of the community.

H4a/b. Power Distance moderates the relationship between (a) Social and (b) Hedonic motives to SVC: Being surrounded by a culture with a high level of power distance weakens SVC for socially motivated contributors (when we consider all contributors). Still, if we consider Segment 1 (Hedonic) in the cluster analysis, we observe that power distance also weakens the SVC associated with hedonism. These results align with Cognitive Evaluation Theory [57]; contributors driven by hedonic (Seg. 1) or social motives (Seg. 2) need more autonomy (through less hierarchy—less Power Distance) to develop a Sense of Virtual Community. When not exerted in toxic and harsh ways to discipline community members, concerted control of communications can also ultimately play a pro-social role in increasing the SVC by increasing cohesiveness, commitment, and conformity [108].

H5a/b. Payment moderates the relationship between (a) Social and (b) Hedonic motives to SVC: Being paid to contribute weakens the association with SVC for hedonist contributors. The cluster analysis shows that being paid to contribute also weakens the SVC associated with social motives. Paid contributors, even those driven by hedonic or social motivations, showed a lower Sense of Virtual Community to the Linux Kernel. This result aligns with Cognitive Evaluation Theory [57] and might be explained by the conflicting identities and divided loyalties that paid contributors have to both their sponsoring firms and the Linux Kernel community [39]. We hypothesize that these contributors would leave the community if there were no payment to compensate for their participation.

Implications for OSS communities to retain contributors

SVC is associated with practices on *exchanging support* [15], [33], *creating identities and making identifications* [33], *producing mutual cognitive and affective trust* amongst members of a community [33], and establishing norms and a *"concertive (but not enforced) control"* [108], in which members of the community become responsible for directing their work and monitoring themselves. OSS communities can provide not only online interest groups for members, chat rooms, instant messaging, and discussion forums to encourage community involvement [109] but also online tools with shared spaces for contributors to work "together" on issues to be able to discuss and collaborate on similar interests. Better interactions can strengthen contributors' Sense of Virtual Community, especially those seeking social relationships. When the information being exchanged surpasses the technical content and includes socio-emotional support, it shows personal relationships among group members, and finally brings feelings of acceptance by members [33]. OSS communities should foster exchanging support among members to bring a positive impact on developing SVC [15]. The exchanging support includes technical and social support and happens through comments in pull requests and participation in mailing lists (by either reading or posting messages). Communities can manage pull requests and mailing lists to guarantee that members' posts are not being missed [74] and that the communication adheres to the code of conduct.

Implications for OSS communities to attract newcomers. Exchanging information and providing support to other community members are practices associated with positive feelings toward the community, and members' stronger attachment to the community [110]. Community members can encourage newcomers to become more active and move beyond the stage of 'lurker,' enticing them to participate in mailing lists [15] and to start making social connections to establish mutual trust, be known by other contributors, and facilitate the development of their Sense of Virtual Community. Conferences and meetups can help hedonic and socially motivated contributors have fun and increase their social capital.

Implications for Research. This study suggests a positive link between Social and Hedonic motivations and a Sense of Virtual Community. Further, the cluster analysis has detected unobserved heterogeneity within our sample, suggesting that there are different subgroups within the community for which different motivations play a more prominent role. Future work could explore how the challenges faced by contributors influence the development of a Sense of Virtual Community and how a Sense of Virtual Community influences the decisions to stay or leave a project. While we included three control variables, future work can consider additional variables, for example, demographic variables such as age.

Our study focuses on the Linux Kernel community, which is a limitation to the generalizability of this study; we suggest that our findings provide a useful starting point to conduct similar studies across other specific communities or across OSS developers regardless of which community they partake

in. When considering other projects, we also suggest that different project governance models might also play a role in SVC. This study has also demonstrated that payment plays a role in SVC and that minorities and marginalized individuals feel less part of the community.

Finally, our study has focused on the antecedents of a sense of virtual community in OSS but not on the consequences of it, and this could be a further area of focus in future work. Future work can investigate whether SVC is related to contributor satisfaction and whether a reduced SVC leads to contributors' attrition, thus jeopardizing a community's sustainability.

VII. Limitations and Threats to Validity

Construct Validity. We adopted and tailored existing measurement instruments for some constructs based on prior literature. Our analysis of the measurement model confirmed that our constructs were internally consistent and scored satisfactorily on convergent and discriminant validity tests. In this study we have used respondents' country of residence as a proxy for Power Distance as a dimension of culture as defined by Hofstede [59]. While also used in other studies [62], we acknowledge it is an approximation and not a perfect measure. One potential issue is that we do not know how *long* respondents have lived in their current country of residence. Another potential issue is that contributors' culture from where they grew up may differ from their current culture. This is why we report the metric as being surrounded by a specific culture instead of having a specific culture. Measuring culture in a more precise way is an important avenue for future work in general.

Internal Validity. Our hypotheses propose associations between different constructs rather than causal relationships, as the present study is a cross-sectional sample study [111]. We acknowledge the limitation that our respondents comprise contributors who are more likely to have a sense of virtual community as they dedicated their time to answering the questionnaire, suggesting a response bias. While it is clear that contributors motivated by some intrinsic-social reasons tend to experience a sense of virtual community and that power distance and financial compensation can influence those associations, a theoretical model such as ours cannot capture a complete and exhaustive list of factors. Other factors can play a role and our results represent a starting point for future studies.

External Validity. The Linux Kernel is a mature project that has attracted contributors for its value over the years, and while studied frequently and sometimes positioned as a 'quintessential' open source project, open source projects can vary in many ways. The specific context of the Linux Kernel project may therefore have impacted the results, which are therefore not necessarily generalizable to other OSS projects. Nevertheless, theory-building is a continuous and iterative process of proposing, testing, and modifying theory [112], in which a single case study is a first step towards constructing theories. In that sense, it is more valuable to interpret these results as a starting point and seek theoretical generalizability rather than statistical generalizability. Further, given the very important role that the Linux Kernel project plays in the software industry, this project we argue this is an appropriate starting point; further replication studies can validate and extend our theoretical model.

Our survey was conducted online and anonymously, but the numbers are aligned with the overall distribution of the Linux Kernel contributors. The Linux Kernel includes contributions of more than 20,000 developers [73], and are mostly paid to contribute [66], [68]. According to previous research, around 10% of contributors to Linux Kernel identify themselves as women [80], and the majority is from the USA, which is aligned with our sample. The responses were sufficiently consistent to find full or partial empirical support for four hypotheses.

VIII. Conclusion

A Sense of Virtual Community (SVC) helps individuals feel valued in their community, leading to more satisfied, involved, and committed contributors. While research has identified different motivations and challenges for contributors to OSS, it is unclear how a sense of community is created and what factors impact it.

In this paper, we close this gap by developing a theoretical model for sense of virtual community in OSS through a survey of Linux Kernel contributors. We found evidence that a subset of intrinsic motivations (social and hedonic motives) are positively associated with SVC; however, other extrinsic factors such as the country's culture and being paid to contribute can lessen SVC among contributors. Additionally, those with higher English confidence feel a higher sense of belonging in the community, and contributors who identify as part of a gender minority (non-men), tend to feel less of a sense of virtual community.

Our results also show heterogeneity in our respondents, suggesting that there are different subgroups within the community for whom different motivations play a more prominent role. This suggests that a "one size fits all" approach would not work when designing interventions to create an inclusive, welcoming community. Our SVC model can help researchers and community design interventions by highlighting the different factors that interplay in creating a sense of virtual community in OSS for different subgroups of contributors.

Acknowledgments

We are indebted to Kate Stewart, Vice President of Dependable Embedded Systems at the Linux Foundation, and Shuah Khan, maintainer and contributor of the Linux Kernel, for their invaluable support and assistance in the survey design and distribution. We much appreciate the extensive efforts and time they spent in meetings and reviews, and their thoughtful input to the survey design, and reaching out to the Linux Kernel community members. We are deeply grateful to all respondents of this survey. The National Science Foundation partially supports this work under Grant Numbers 1900903, 1901031, 2236198, 2235601, and Science Foundation Ireland grants no. 13/RC/2094-P2 to Lero, the SFI Research Centre for Software, and no. 15/SIRG/3293.

REFERENCES

[1] J. Roberts, I.-H. Hann, and S. Slaughter, "Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the apache projects," *Manage Sci*, vol. 52, 2006.

[2] A. Barcomb, K.-J. Stol, D. Riehle, and B. Fitzgerald, "Why do episodic volunteers stay in floss communities?" in *41st International Conference on Software Engineering (ICSE)*. IEEE, 2019, pp. 948–959.

[3] B. M. Hagerty, J. Lynch-Sauer, K. L. Patusky, M. Bouwsema, and P. Collier, "Sense of belonging: A vital mental health concept," *Archives of psychiatric nursing*, vol. 6, no. 3, pp. 172–177, 1992.

[4] R. F. Baumeister and M. R. Leary, "The need to belong: Desire for interpersonal attachments as a fundamental human motivation," in *Interpersonal development*, R. Zukauskiene, Ed. Routledge, 2017.

[5] A. Maslow, "A theory of human motivation." *Psychol Rev*, vol. 50, no. 4, 1943.

[6] B. M. Hagerty and K. Patusky, "Developing a measure of sense of belonging," *Nursing research*, vol. 44, no. 1, pp. 9–13, 1995.

[7] K.-A. Allen, "Making sense of belonging," *InPsych*, vol. 41, 2019.

[8] L. Espinosa, "Pipelines and pathways: Women of color in undergraduate stem majors and the college experiences that contribute to persistence," *Harvard Educational Review*, vol. 81, no. 2, pp. 209–241, 2011.

[9] L. Happe and B. Buhnova, "Frustrations steering women away from software engineering," *IEEE Software*, vol. 39, no. 4, 2021.

[10] S. M. Burroughs and L. T. Eby, "Psychological sense of community at work: A measurement system and explanatory framework," *Am J Community Psychol*, vol. 26, no. 6, pp. 509–532, 1998.

[11] A. L. Blanchard, "Developing a sense of virtual community measure," *CyberPsychology & Behavior*, vol. 10, no. 6, pp. 827–830, 2007.

[12] D. W. McMillan and D. M. Chavis, "Sense of community: A definition and theory," *J Community Psychol*, vol. 14, no. 1, pp. 6–23, 1986.

[13] J. Koh, Y.-G. Kim, and Y.-G. Kim, "Sense of virtual community: A conceptual framework and empirical validation," *International journal of electronic commerce*, vol. 8, no. 2, pp. 75–94, 2003.

[14] A. L. Blanchard, "Testing a model of sense of virtual community," *Computers in Human Behavior*, vol. 24, no. 5, pp. 2107–2123, 2008.

[15] L. Tonteri, M. Kosonen, H.-K. Ellonen, and A. Tarkiainen, "Antecedents of an experienced sense of virtual community," *Computers in Human Behavior*, vol. 27, no. 6, pp. 2215–2223, 2011.

[16] J. Kim, H.-M. Kim, and M. Kim, "The impact of a sense of virtual community on online community: does online privacy concern matter?" *Internet Research*, vol. 31, no. 2, pp. 519–539, 2020.

[17] G. Von Krogh, S. Haefliger, S. Spaeth, and M. W. Wallin, "Carrots and rainbows: Motivation and social practice in open source software development," *MISQ*, vol. 36, no. 2, pp. 649–676, 2012.

[18] A. Hars and S. Ou, "Working for free motivations of participating in open source software projects," in *Proc. HICSS*, 2004, pp. 25–31.

[19] R. Ghosh, R. Glott, B. Krieger, and G. Robles, "Free/libre and open source software: Survey and study," 2002.

[20] K. Lakhani and R. Wolf, "Why hackers do what they do: Understanding motivation and effort in free/open source software projects," *Open Source Software Projects*, 2003.

[21] G. Hertel, S. Niedner, and S. Herrmann, "Motivation of software developers in open source projects an internet-based survey of contributors to the linux kernel," *Res Policy*, vol. 32, no. 7, 2003.

[22] M. Gerosa, I. Wiese, B. Trinkenreich, G. Link, G. Robles, C. Treude, I. Steinmacher, and A. Sarma, "The shifting sands of motivation revisiting what drives contributors in open source," in *ICSE*, 2021.

[23] K. Tamilmani, N. Rana, N. Prakasam, and Y. Dwivedi, "The battle of brain vs. heart: A literature review and meta-analysis of "hedonic motivation" use in utaut2," *Int J Inf Manage*, vol. 46, 2019.

[24] C. G. McClintock, "Social motivation—a set of propositions," *Behav. Sci.*, vol. 17, no. 5, 1972.

[25] J. Gusfield, *Community: A critical response*. Harper & Row New York, 1975.

[26] M. R. Martínez-Torres and M. C. Díaz-Fernández, "Current issues and research trends on open-source software communities," *Technology Analysis & Strategic Management*, vol. 26, no. 1, pp. 55–68, 2014.

[27] H. Rheingold, *The virtual community, revised edition: Homesteading on the electronic frontier*. MIT press, 2000.

[28] H. H. Chang, P.-H. Hsieh, and C. S. Fu, "The mediating role of sense of virtual community," *Online Information Review*, 2016.

[29] R. Brown and S. Pehrson, *Group processes: Dynamics within and between groups*. John Wiley & Sons, 2019.

[30] J. Sutanto, A. Kankanhalli, and B. C. Y. Tan, "Eliciting a sense of virtual community among knowledge contributors," *ACM Trans Manag Inf Syst*, vol. 2, no. 3, pp. 1–17, 2011.

[31] A. L. Blanchard and M. L. Markus, "Sense of virtual community maintaining the experience of belonging," in *Hawaii International Conference on System Sciences*. IEEE, 2002, pp. 3566–3575.

[32] L. Zhao, Y. Lu, B. Wang, P. Y. Chau, and L. Zhang, "Cultivating the sense of belonging and motivating user participation in virtual communities: A social capital perspective," *International Journal of Information Management*, vol. 32, no. 6, pp. 574–588, 2012.

[33] A. Blanchard, D. A. Askay, and K. A. Frear, "Sense of community in professional virtual communities," in *Virtual communities: Concepts, methodologies, tools and applications*. IGI Global, 2011.

[34] G.-L. Chen, S.-C. Yang, and S.-M. Tang, "Sense of virtual community and knowledge contribution in a P3 virtual community: Motivation and experience," *Internet Research*, vol. 23, no. 1, 2013.

[35] I. Steinmacher, T. U. Conte, C. Treude, and M. A. Gerosa, "Overcoming open source project entry barriers with a portal for newcomers," in *Proceedings of the 38th International Conference on Software Engineering*, 2016, pp. 273–284.

[36] A. C. C. França, T. B. Gouveia, P. C. Santos, C. A. Santana, and F. Q. da Silva, "Motivation in software engineering: A systematic review update," in *EASE 2011*. IET, 2011, pp. 154–163.

[37] B. S. Frey, "On the relationship between intrinsic and extrinsic work motivation," *Int J Ind Organ*, vol. 4, no. 15, pp. 427–439, 1997.

[38] M. O'Neil, X. Cai, L. Muselli, F. Pailler, and S. Zacchiroli, *The coproduction of open source software by volunteers and big tech firms*. News and Media Research Centre (UC), 2021.

[39] M. Schaarschmidt and K.-J. Stol, "Company soldiers and gone-natives: role conflict and career ambition among firm-employed open source developers," in *Proceedings of the International Conference on Information Systems*. Association for Information Systems, 2018.

[40] R. Ryan and E. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being." *Am. Psychol.*, vol. 55, no. 1, p. 68, 2000.

[41] I. Steinmacher, M. A. Gerosa, and D. Redmiles, "Attracting, onboarding, and retaining newcomer developers in open source software projects," in *Workshop on GSD in a CSCW Perspective*, 2014.

[42] R. Kaur and K. Chahal, "Exploring factors affecting developer abandonment of open source software projects," *J. softw.: Evol. Process*, vol. 34, no. 9, 2022.

[43] F. Calefato, M. Gerosa, G. Iaffaldano, F. Lanubile, and I. Steinmacher, "Will you come back to contribute? investigating the inactivity of oss core developers in github," *Empir Softw Eng*, vol. 27, no. 3, 2022.

[44] K.-A. Allen, *The Psychology of Belonging*. Routledge, 2020.

[45] C. Goodenow and K. E. Grady, "The relationship of school belonging and friends' values to academic motivation among urban adolescent students," *The Journal of Experimental Education*, vol. 62, no. 1, 1993.

[46] D. Solomon, M. Watson, V. Battistich, E. Schaps, and K. Delucchi, "Creating classrooms that students experience as communities," *Am J Community Psychol*, vol. 24, no. 6, pp. 719–748, 1996.

[47] R. Neel, D. T. Kenrick, A. E. White, and S. L. Neuberg, "Individual differences in fundamental social motives," *J Pers Soc Psychol*, vol. 110, no. 6, 2016.

[48] E. Kim and M. Drumwright, "Engaging consumers and building relationships in social media: How social relatedness influences intrinsic vs. extrinsic consumer motivation," *Comput. Hum. Behav.*, vol. 63, 2016.

[49] K. Stewart and S. Gosain, "The impact of ideology on effectiveness in open source software development teams," *MISQ*, vol. 30, no. 2, 2006.

[50] D. McGowan, "Legal implications of open-source software," *U. Ill. L. Rev.*, 2001.

[51] G. Lea, "MS' Ballmer: Linux is a communism," 2000, the Register. [Online]. Available: https://www.theregister.com/2000/07/31/ms_ballmer_linux_is_communism/

[52] T. Greene, "Ballmer: 'Linux is a cancer'," 2001, the Register. [Online]. Available: https://www.theregister.com/2001/06/02/ballmer_linux_is_a_cancer/

[53] R. Janoff-Bulman and N. C. Carnes, "The model of moral motives," *The Guilford Press*, p. 223, 2018.

[54] H. Tajfel and J. C. Turner, "The social identity theory of intergroup behavior," in *Polit Psychol*. Psychology Press, 2004, pp. 276–293.

[55] R. Gan, L. S. Neo, J. Chin, and M. Khader, "Change is the only constant: The evolving role of women in the islamic state in iraq and syria (isis)," *Women & Criminal Justice*, vol. 29, no. 4-5, pp. 204–220, 2019.

[56] Y. Fang and D. Neufeld, "Understanding sustained participation in open source software projects," *J Manag Inf Syst*, vol. 25, no. 4, 2009.

[57] E. L. Deci and R. M. Ryan, "Cognitive evaluation theory," in *Intrinsic motivation and self-determination in human behavior*. Springer, 1985.

[58] G. Hofstede, *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage publications, 2001.

[59] ——, "Dimensionalizing cultures: The Hofstede model in context," *Online readings in psychology and culture*, vol. 2, no. 1, 2011.

[60] H. Fock, M. K. Hui, K. Au, and M. H. Bond, "Moderation effects of power distance on the relationship between types of empowerment and employee satisfaction," *J Cross Cult Psychol*, vol. 44, no. 2, 2013.

[61] S. Auh, B. Menguc, S. Spyropoulou, and F. Wang, "Service employee burnout and engagement: the moderating role of power distance orientation," *J. Acad. Mark. Sci.*, vol. 44, no. 6, 2016.

[62] K. Cortina, S. Arel, and J. Smith-Darden, "School belonging in different cultures: The effects of individualism and power distance," in *Frontiers in Education*, vol. 2, 2017.

[63] A. Ardichvili, "Learning and knowledge sharing in virtual communities of practice: Motivators, barriers, and enablers," *Adv Dev Hum Resour*, vol. 10, no. 4, 2008.

[64] J. Taylor and R. Dantu, "For love or money? examining reasons behind oss developers' contributions," *Inform Sys Manage*, vol. 39, no. 2, 2022.

[65] B. Trinkenreich, M. Guizani, I. Wiese, A. Sarma, and I. Steinmacher, "Hidden figures: Roles and pathways of successful oss contributors," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, 2020.

[66] J. Corbet, G. Kroah-Hartman, and A. McPherson, "Linux kernel development. how fast it is going, who is doing it, what they are doing, and who is sponsoring it," *The Linux Foundation*, 2012.

[67] D. Riehle, P. Riemer, C. Kolassa, and M. Schmidt, "Paid vs. volunteer work in open source," in *HICSS*, 2014, pp. 3286–3295.

[68] D. Homscheid, J. Kunegis, and M. Schaarschmidt, "Private-collective innovation and open source software: Longitudinal insights from linux kernel development," in *Conference on e-Business, e-Services and e-Society*. Springer, 2015.

[69] P. N. Sharma, S. L. Daniel, T. R. Chung, and V. Grover, "A motivation-hygiene model of open source software code contribution and growth," *J Assoc Inform Sys*, vol. 23, no. 1, pp. 165–195, 2022.

[70] J. Gerlach, C.-G. Wu, L. Cunningham, and C. Young, "An exploratory study of conflict over paying debian developers," *Int. J. Open Source Softw. Process.*, vol. 7, no. 3, 2016.

[71] D. Homscheid, "The Linux Kernel project," in *Firm-Sponsored Developers in Open Source Software Projects*. Springer, 2020, pp. 83–91.

[72] X. Tan, M. Zhou, and B. Fitzgerald, "Scaling open source communities: An empirical study of the linux kernel," in *International Conference on Software Engineering*. IEEE, 2020.

[73] K. Stewart, S. Khan, D. M. German, and et al., "2020 Linux Kernel History Report," https://www.linuxfoundation.org/resources/publications/linux-kernel-history-report-2020.

[74] C. Miller, S. Cohen, D. Klug, B. Vasilescu, and C. Kästner, "Did you miss my comment or what? understanding toxicity in open source discussions," in *Int'l Conf. Software Engineering*, 2022.

[75] C. Ringle, D. Silva, and D. Bido, "Structural equation modeling with the smartPLS," *Brazilian Journal Of Marketing*, vol. 13, no. 2, 2015.

[76] D. Russo and K.-J. Stol, "PLS-SEM for software engineering research: An introduction and survey," *ACM Comput Surv*, vol. 54, no. 4, 2021.

[77] C.-W. Chen and C.-S. Lin, "Building a sense of virtual community: The role of the features of social networking sites," *Cyberpsychology, Behavior, and Social Networking*, vol. 17, no. 7, pp. 460–465, 2014.

[78] H.-K. Ellonen, M. Kosonen, A. Tarkiainen, and L. Tonteri, "The positive outcomes of a sense of virtual community," *International Journal of Web Based Communities*, vol. 9, no. 4, pp. 465–482, 2013.

[79] I. Steinmacher, S. Balali, B. Trinkenreich, M. Guizani, D. Izquierdo-Cortazar, G. Cuevas, M. Gerosa, and A. Sarma, "Being a mentor in open source projects," *J Internet Serv Applic*, vol. 12, no. 1, 2021.

[80] Bitergia, "Gender-diversity analysis of the linux kernel technical contributions," https://blog.bitergia.com/2016/10/11/gender-diversity-analysis-of-the-linux-kernel-technical-contributions.

[81] F. Zlotnick, "Github survey," http://opensourcesurvey.org/2017/, 2017.

[82] "Replication package," https://figshare.com/s/5cedff387200ebc379cb.

[83] B. Weijters and H. Baumgartner, "Misresponse to reversed and negated items in surveys a review," *J Mark Res*, vol. 49, no. 5, 2012.

[84] H. Baumgartner and J.-B. Steenkamp, "Response styles in marketing research cross-national investigation," *J Mark Res*, vol. 38, no. 2, 2001.

[85] P. Podsakoff, S. MacKenzie, and Jeong-Yeon, "Common method biases in behavioral research: A critical review of the literature and recommended remedies," *J Appl Psychol*, vol. 88, no. 5, 2003.

[86] M. W. Van Someren, Y. F. Barnard, and J. A. Sandberg, "The think aloud method," *London: AcademicPress*, vol. 11, 1994.

[87] W. L. Grichting, "The meaning of "i don't know" in opinion surveys: Indifference versus ignorance," *Aust Psychol*, vol. 29, no. 1, 1994.

[88] P. Sturgis, C. Roberts, and P. Smith, "Middle alternatives revisited: How the neither/nor response acts as a way of saying "i don't know"?" *Sociological Methods & Research*, vol. 43, no. 1, pp. 15–38, 2014.

[89] M. Sarstedt, C. M. Ringle, and J. F. Hair, "Treating unobserved heterogeneity in PLS-SEM: A multi-method approach," in *Partial least squares path modeling*. Springer, 2017, pp. 197–217.

[90] F. Faul, E. Erdfelder, A. Buchner, and A.-G. Lang, "Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses," *Behav Res Methods*, vol. 41, no. 4, pp. 1149–1160, 2009.

[91] J. F. Hair, J. J. Risher, M. Sarstedt, and C. M. Ringle, "When to use and how to report the results of PLS-SEM," *Eur. Bus. Rev.*, 2019.

[92] M. Sarstedt and J.-H. Cheah, "Partial least squares structural equation modeling using smartPLS: a software review," 2019.

[93] M. Bartlett, "Tests of significance in factor analysis." *Br J Psychol*, 1950.

[94] J. F. Hair Jr, R. E. Anderson, R. L. Tatham, and W. C. Black, "Multivariate data analysis with readings," 1995.

[95] N. Kock, "Advanced mediating effects tests, multi-group analyses, and measurement model assessments in PLS-based SEM," *International Journal of e-Collaboration*, vol. 10, no. 1, 2014.

[96] J. Henseler, C. M. Ringle, and M. Sarstedt, "A new criterion for assessing discriminant validity in variance-based structural equation modeling," *J. Acad. Mark. Sci.*, vol. 43, no. 1, pp. 115–135, 2015.

[97] J. Henseler, C. M. Ringle, and R. R. Sinkovics, "The use of partial least squares path modeling in international marketing," in *New challenges to international marketing*. Emerald Group Publishing Limited, 2009.

[98] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *J R Stat Soc Series B Stat Methodol*, vol. 36, 1974.

[99] M. Tenenhaus, V. Vinzi, Y.-M. Chatelin, and C. Lauro, "Pls path modeling," *Computational statistics data analysis*, vol. 48, no. 1, 2005.

[100] J. Henseler, T. K. Dijkstra, M. Sarstedt, C. M. Ringle, A. Diamantopoulos, D. W. Straub, D. J. Ketchen Jr, J. F. Hair, G. T. M. Hult, and R. J. Calantone, "Common beliefs and reality about pls: Comments on rönkkö and evermann," *Organ Res Methods*, vol. 17, no. 2, 2014.

[101] J. Henseler, G. Hubona, and P. Ray, "Using pls path modeling in new technology research: updated guidelines," *Ind. Manag. Data Syst.*, 2016.

[102] J.-M. Becker, A. Rai, C. Ringle, and F. Völckner, "Discovering unobserved heterogeneity in structural equation models to avert validity threats," *MIS Quarterly*, vol. 37, no. 3, pp. 665–694, 2013.

[103] M. Sarstedt, J.-M. Becker, C. Ringle, and M. Schwaiger, "Uncovering and treating unobserved heterogeneity with fimix-pls: which model selection criterion provides an appropriate number of segments?" *Schmalenbach Business Review*, vol. 63, no. 1, 2011.

[104] B. Trinkenreich, I. Wiese, A. Sarma, M. Gerosa, and I. Steinmacher, "Women's participation in open source software: A survey of the literature," *ACM Trans Soft Eng Methodol*, vol. 31, no. 4, 2022.

[105] N. Raman, M. Cao, Y. Tsvetkov, C. Kästner, and B. Vasilescu, "Stress and burnout in open source: Toward finding, understanding, and mitigating unhealthy interactions," in *ICSE*, 2020.

[106] S. Cohen, "Contextualizing toxicity in open source: a qualitative study," in *Proc. 29th ESEC/FSE*. ACM, 2021, pp. 1669–1671.

[107] I. Steinmacher, M. A. G. Silva, and M. A. Gerosa, "Barriers faced by newcomers to open source projects: a systematic review," in *International Conference on Open Source Systems*. Springer, 2014.

[108] J. Gibbs, H. Kim, and S. Ki, "Investigating the role of control and support mechanisms in members' sense of virtual community," *Commun. Res.*, vol. 46, no. 1, 2019.

[109] B. Xu and D. Li, "An empirical study of the motivations for content contribution and community participation in wikipedia," *Information & management*, vol. 52, no. 3, pp. 275–286, 2015.

[110] A. Blanchard and M. Markus, "The experienced" sense" of a virtual community: Characteristics and processes," *ACM SIGMIS DATABASE*, vol. 35, no. 1, 2004.

[111] K.-J. Stol and B. Fitzgerald, "The ABC of software engineering research," *ACM TOSEM*, vol. 27, no. 3, 2018.

[112] F. Shull, J. Singer, and D. I. Sjøberg, "Guide to advanced empirical software engineering," 2007.