doi:10.1093/iwcomp/xxxxxx

Gender, Representation and Online Participation: A Quantitative Study

Bogdan Vasilescu,[1] Andrea Capiluppi,[2] and Alexander Serebrenik[1]

[1] *Eindhoven University of Technology, Eindhoven, The Netherlands,* {b.n.vasilescu/a.serebrenik}@tue.nl
[2] *Brunel University, London, United Kingdom,* andrea.capiluppi@brunel.ac.uk

Online communities are flourishing as social meeting web-spaces for users and peer community members. Different online communities require different levels of competence for participants to join, and scattered evidence suggests that female and minorities as participants can be under-represented. Additional anecdotal evidence suggests that women withdraw from unfriendly online communities.

Due to the limited amount of empirical evidence on the matter, this paper provides a quantitative study of the phenomenon, in order to assess the representation and social impact of gender in online communities. This study positions itself within recent and focused international initiatives, launched by the European Commission in order to encourage women in the field of sciences and technology.

Focusing on technical support networks around web content management tools (e.g., Drupal and WordPress) and on questions&answers websites (e.g., StackOverflow), this paper unearths a spectrum of online communities, in which women participate to various degrees.

Categories and subject descriptors: .....

Keywords: online community, gender, StackOverflow, empirical study

Responsible Editorial Board Member: Editorial Board member name

1. INTRODUCTION

Online communities and social sites represent an extension of the very well known "open source" phenomenon: individuals use their own free time to gather around a common web space, to discuss, socialize or request support from other contributors. Current online communities target a wide spectrum of diverse users, ranging from the general audience (e.g., Facebook), professionals (e.g., LinkedIn), or IT experts specifically (e.g., StackOverflow). The purposes of these sites can also be very diverse: some provide the ability to share content such as source code fragments (e.g., snipplr), entire projects (e.g., GitHub, bitbucket) or images (e.g., Flickr). Others support knowledge sharing by means of questions and answers (e.g., StackOverflow) or news postings (e.g., reddit).

Software developers (e.g., professionals working in a company, or voluntaries active in open-source) create and maintain software by standing on the shoulders of others [Storey et al., 2010]: they reuse components and libraries, and go *foraging* on the Web for information that will help them in their tasks [Brandt et al., 2009]. When foraging online, developers often turn to mailing lists, or programming question and answer (Q&A) communities such as StackOverflow[1] (SO). StackOverflow is becoming one of the most visible venues for sharing knowledge about software development [Mamykina et al., 2011].

[1] http://stackoverflow.com

Participation in such communities is a two-way process: by *asking* (e.g., sending emails to a mailing list, or posing questions on a Q&A platform), developers can seek help and advice from their peers about undocumented technology features [Parnin et al., 2012], report bugs, or discuss future changes to the software; by *answering* (e.g., Q&A questions posed by others), developers can share their knowledge and expertise, and educate their peers. Therefore, such dedicated online communities are an integral part of the working lives of software developers, exponents of Science, Technology, Engineering, and Mathematics (STEM) professionals.

The gender representation in STEM-related subjects raises significant attention of researchers and academics [Miliszewska et al., 2006, Blickenstaff, 2005, Fisher et al., 1997, Roberts et al., 2002, Hill et al., 2010], as well as of policy-makers [Initiative, 2012], all noting a significant under-representation of women. The main reasons behind such under-representation have been studied mostly qualitatively, to delineate the issue, formulate the main reasons of the imbalance between the number of female and male participants, and propose initiatives to attract more women to sciences, in terms of both majoring in STEM-related studies as well as choosing STEM-related career paths. In particular, encouraging more women to participate in Information Technology, Computer Science, and Computer Engineering is seen as having potential benefits not only to women, but also to society [Whitecraft and Williams, 2010, p.235].

In addition to gender and STEM-related studies and careers in general, there is also the issue of representation of women in the use of technology and online communities. The use of Internet technologies is not as unbalanced as the access to careers and vocational studies [Bimber, 2000]. Still, software development remains a predominantly male activity, especially for Open Source: all surveys reviewed by David and Shapiro [2008] agree that only 1-5% of the open source developers are women. This is in sharp contrast with the 28% female employees with computer and mathematical occupations reported by the National Science Foundation [2004].

The focus on gender under-representation in online communities is further motivated by anecdotal observations: it has been suggested that the Q&A website StackOverflow strongly promotes one-upmanship; fosters flame-wars and the down-voting of individuals; and it is based on earning prizes, reputation and badges, that allow participants to access new features and gain more control on others' postings [VA, 2011a,b]. Experience suggests that this results in a lesser participation by female users, who do not engage with this sort of communities. In a similar vein Turkle [2005, p.194] observed that the focus of hackers in winning and achieving recognition "*makes their world peculiarly male in spirit, peculiarly unfriendly to women*". Moreover, some female StackOverflow users prefer to use gender-neutral names to be accepted by the mostly-male audiences: similar feeling of a strong pressure to make themselves invisible as women has been reported by Nafus et al. [2006]. Male StackOverflow users also masquerade as females sometimes, with the belief that other (male) users would be less aggressive towards them and their questions. Similar "gender swapping" has been observed (and for similar purposes) in an online poker community [Wood and Griffiths, 2008].

This paper is an attempt to quantitatively evaluate the presence of women in specific software-development-related online communities, and to compare their levels and duration of engagement (as compared to the male counterparts). The main rationale for this study stems from the scarcity of empirical studies so far on how gender plays a role in highly-skilled software-development-related online communities, while most of the evidence remains at the anecdotal level. This paper analyses a sample of female and male users from the Q&A website StackOverflow, and compares its gender engagement with the interaction on the Drupal and WordPress mailing lists, whose topics (web Content Management Systems) are associated with skills subject to gendered evolution [Kotamraju, 2003].

This paper is organised as follows: Section 2 reports on relevant insights from the existing literature on gender and technology as well as on publications related to the kind of data we analyse and analysis techniques we apply. Section 3 deals with the research design of this study following the *Goal-Question-Metric* approach [Basili and Weiss, 1984]. Section 4 discusses gender and self-representation in StackOverflow, Drupal and WordPress; Section 5 summarises the issues with collecting and analysing the necessary data, and includes our gender resolution approach; Section 6 presents the results of a pilot survey conducted to obtain insights in the demographics of StackOverflow, and to test the accuracy of the gender resolution algorithm; Section 7 shows and compares the results of the experiments on the analysed communities; Section 8 identifies the threats to validity. Finally, Section 9 summarises the paper and concludes that two types of communities emerged: while women are still a minority in both, in the first type of community women and men show broadly similar contribution patterns; while in the second type of community, women's levels of participation are significantly lower than men's.

2. RELATED WORK

This article is an extended and revised version of our previous conference paper [Vasilescu et al., 2012] that focussed on StackOverflow (SO). There we have observed that: (i) gender of many SO users could not be

identified; (ii) the percentage of users engaged in SO and presenting themselves as feminine was greatly imbalanced in comparison to the percentage of users presenting themselves as masculine; (iii) not only do male-perceived users represent the vast majority of contributors, but they also participate more, earn more reputation points, and engage in the "game" imposed by SO more than female-perceived users do. However, the question remained open on *whether a lesser participation of the latter users in SO can be attributed to specifics of SO (such as prizes, reputation and badges), or it is a common phenomenon in online developer communities*. Therefore, in the current work we compare SO with two additional online communities, Drupal and Wordpress.

As the question we address is at the intersection of technology studies, studies of online communities and gender studies, in Subsection 2.1 we discuss relevant insights from the literature on gender and technology (technology in general, and gender and information technology/online communities specifically), while in Subsection 2.2 we focus on publications related to the kind of data we analyse and gender resolution technique we apply.

2.1. Gender and technology

Our work builds upon the existing studies of the relation between gender and technology in general [Lerman et al., 2003] and specifically, gender and information technology [tra, 2006].

2.1.1. Feminist scholarship

Several qualitative studies have focused on the reasons why women are not willing to embark in STEM-related subjects and careers [Fisher et al., 1997, Roberts et al., 2002, Hill et al., 2010]. Various reasons have been given to the under-representation of women in STEM subjects: a general lack of interest in STEM subjects [Hill et al., 2010], stereotyped thinking by family and teachers [Gras-Velazquez et al., 2009], lack of role models [Margolis and Fisher, 2003], and most often a combination of various causes together [Trauth et al., 2004]. These approaches frequently adhere to the liberal feminism vision [Lohan and Faulkner, 2004]: technology is inherently gender-neutral but controlled by men and misrepresented in the classroom [Kelly, 1987a,b]. The gap between women and technology should, hence, be addressed by adapting aspirations, values and behaviour of women [Wajcman, 2010]. Alternative approaches to conceptualization of the link between gender and technology have been provided by ecofeminist thinking and socialist feminism [Grint and Gill, 1995, Hodgkinson, 2000]. Ecofeminists associate women and nature, on the one hand, and men and mechanisms, on the other hand. They reject the reductionist approach characteristic for modern science, as subjugating both nature and women, in favour of holism [Merchant, 1990, pp.290–295]. Socialist feminism, also known as "technology as masculine culture", shares with radical feminism the idea of gendered technology but rejects the essentialism of ecofeminism [Wajcman, 2010, Grint and Gill, 1995, p.8]. This rejection of essentialism by "technology as masculine culture" is close to the concept of "gender as performance" advocated by Butler [1999, p.33]. Technofeminism [Wajcman, 2010] asserts that gender is integral to the sociotechnical process: technology affords or inhibits the doing of particular gender power relations. This also means that technofeminist approaches can expose concrete technological practices leading to exclusion of women.

We subscribe to the latter stance. The ultimate goal of our research consists, therefore, in understanding communication practices in software engineering leading to exclusion of women. Our previous study [Vasilescu et al., 2012] has shown that communication in SO leads to lesser participation by women. Understanding whether this observation is SO-specific, however, calls for a comparative study, which is the subject of this article. In our paper, we also follow the liberal-feminist approach: we inherently assume that communities and their activity are gender neutral, and we test whether this assumption is correct or not. The StackOverflow community, as well as the Drupal and Wordpress ones, are technical in nature, so devoted to proposing solutions to technical problems. Gender should not play a role in such communities: an answer to a question is not "more correct" based on the gender of the contributor who solves the problem.

2.1.2. Gender and information technology

Different paths of feminist thinking have been also applied to information technology [Van Zoonen, 1992, Turkle, 2005, pp.111–114]. For instance, it has been pointed out that the use of Internet technologies is not as unbalanced as the access to careers and vocational studies [Bimber, 2000]. Moreover, the "hacker" culture tends to be male-dominated [Taylor, 1999] and unfriendly to women [Turkle, 2005, p.194], although the number of female hackers is not easy to evaluate [Adam, 2003] and in some countries, such as Malaysia, women account for about half of all students in higher computer science education [Lagesen, 2008]. Conversely, a major advocate of the open source phenomenon posits that the hacker culture does not favor a specific gender, but rather is asexual when dealing with technology-oriented problems [Raymond, 1999]. This has been questioned by recent results showing an "active discrimination" towards women [Nafus et al., 2006].

The alienation of women and, in general, the unwillingness observed in groups of minorities to

participate in online communities have created the so-called "impostor syndrome" amongst women: in spite of having good knowledge and being professionally well-settled, women believe they are disqualified or are doing a fraud by fooling others [Clance, 1985]. Sullivan [2012] recommends "Do not feed the Trolls" that can become discouragement amongst the users or members. Apart from the "Science: It's a Girl Thing" initiative launched in June 2012 [Initiative, 2012] by the EU commission, there are also some other gender specific online communities who encourage girls to engage in STEM vocational fields [Schimke et al., 2007] and support women in computing and also provide them with a private space to take advice from other members in the same field[2].

Relating to this second sector of related work, our approach is to produce a quantitative experiment of how gender engages with computing technology, without assuming that the "hacker" culture is more or less biased towards one or the other gender, or that there are relevant differences in the behaviour of male and female users, when dealing with technical content, and the production of user-generated content.

The ultimate application of our research consists in creating gender-neutral communities supporting software development. However, before this application can be realised, the gender implications of current communication practices in software engineering should be understood.

2.1.3. Gender and online communities

A number of studies targeted a broader question of differences in the online behaviour of men and women [Parks, 1996, Herring, 1996]. Specifically, impact of gender on participation in online communities has been studied, e.g., in communities targeting cancer [Ginossar, 2008] and travel [Wang and Fesenmaier, 2004]. Women have been more actively involved in cancer communities than men [Ginossar, 2008], despite the common observation that computer-mediated communication, in general, is a male-dominant technology which privileges men [Shaw and Gant, 2002]. In the online travel community [Wang and Fesenmaier, 2004] it has been found that men, holding age and educational level constant, have been community members for a longer period of time.

The ways participants of online communities chose to perform their gender through the choice of names, locations and avatars (discussed in detail in Section 4) are related to the issues of self-representation and identity display in online communities [Grasmuck et al., 2009, Van House, 2011]. Indeed, while the online communities we study cannot be considered social networking sites as defined in [Boyd and Ellison, 2007], both these communities and social networking sites allow the user to construct a public or a semi-public profile. Flexibility of self-presentation in online communities means that by constructing such a profile the users can practice forms of gender vagueness and impersonation [Armentor-Cota, 2011, Marshall, 2006] such as "gender swapping", i.e., self presentation as the opposite gender [D. and R., 1999, Turkle, 2011]. "Gender swapping" is well-studied in the context of multiplayer real-time virtual worlds (known as MUDs [Turkle, 2011]), and has been furthermore observed both in SO and in poker communities [Wood and Griffiths, 2008]. "Gender swapping" can be seen as an attempt to *queer SO*, i.e., cross the prescribed norms, or go in an adverse or opposite direction [Light, 2011].

Our work on participation of women and men in online software engineering communities is at the intersection of gender studies in information technology and studies of computer-mediated communication. Given the male dominance in both domains, we expect to observe male dominance in online software engineering communities as well. However, it is not clear how big the disparity between men and women would be in these communities, and especially whether this disparity would be also visible at the level of engagement.

2.2. Methodology

Our empirical analysis is based on analysing data from mail archives and StackOverflow, and the core step of the empirical analysis is gender resolution.

2.2.1. Studies of mail archives and StackOverflow

Our analysis is based on the data from mail archives of Open Source projects and StackOverflow (SO). Mail archives have been extensively studied in software engineering research (e.g., cf. [Squire, 2012]), while Q&A websites, and specifically SO, are gaining more and more interest from the research community: since 2010, more than forty research papers were based on the StackOverflow data [Vasilescu, 2013], e.g., [Mamykina et al., 2011, Treude et al., 2011]. However, these studies do not consider gender issues, with the aforementioned analysis of gender in mail archives [Kuechler et al., 2012] being the only notable exception.

2.2.2. Gender resolution

The core step of our empirical analysis is gender resolution, i.e., inferring the gender (of the participants active in the chosen online communities), otherwise undisclosed, from other pieces of information which are publicly available (such as their names or profile pictures). As opposed to using interviews in the sociological studies

[2] E.g., the Anita Borg Institute for Women and Technology, http://anitaborg.org, The Women's Foundation of Colorado, https://www.wfco.org/ but also LinuxChix, GrrrTalk, Ada Initiative, among others; cf. [Lin, 2006]

of genders [Boivie, 2010], researchers in information science and software engineering propose automatic or semi-automatic gender resolution approaches. By using quantitative methods one can consider a much broader group of participants. Name-based gender resolution has been attempted before: Herdağdelen and Baroni [2011] and Kuechler et al. [2012] base their decisions on the data from the U.S. Census.

Alternatively, genders can be recognised based on the style of writing [Argamon et al., 2003]. Style-based gender resolution involves counting so called markers that are more frequently used by writers of a certain gender, e.g., pronouns "I", "you" and "she" are significantly more often used by females, while "of"-phrases ("garden of roses") are more typical for male writers [Argamon et al., 2003]. The authors report accuracy of gender resolution as high as 80% [Koppel et al., 2002]. An obvious advantage of the style-based gender-resolution is its robustness against "gender swapping" or individuals presenting themselves in gender-neutral or gender-ambiguous ways. Unfortunately, style-based gender resolution is challenged by a statistical relationship between the gender of the author and the absence of gendered stylistic language features, as described in the study of blogs by Herring and Paolillo [2006]. Moreover, style-based gender-resolution is likely to be affected by the writing style, i.e., the existing style-based gender resolution approaches may not be applicable to SO questions and answers, as SO questions and answers are neither similar to fiction nor to non-fiction documents (i.e., scientific papers) considered by Argamon et al. [2003], nor to blogs considered by Herring and Paolillo [2006]. Finally, gender-resolution accuracy will be affected by errors made by non-native speakers.

Complementary approaches to gender resolution have been proposed by the image processing community [Lu et al., 2005]. Ideally, these approaches could have simplified or even replaced the manual avatar analysis. Unfortunately, many avatars cannot be regarded as facial images (symbols, cartoons, body parts). Moreover, application of image processing approaches such as [Lu et al., 2005] would require a manual preprocessing step involving cropping, resizing and rotation of images. Moreover, in online communication users can guess each other's gender by discussing the kinds of things an off-line person of that gender could be expected to know, e.g., related to period, ring sizes and pantyhoses [Suler, 2004]. This approach, however, cannot be applied to communication in online software development communities. First of all, it does not transfer well across different cultures [Marshall, 2006], hence it is not applicable in a multicultural context, such as the one present in these communities. Second, communication in online developer communities is topic-restricted and does not cover more personal topics as mentioned above.

Based on the discussion above, similarly to Herdağdelen and Baroni [2011], Kuechler et al. [2012], *in this article* we opt for a name-based gender resolution approach, augmented with manual analysis. However, while Herdağdelen and Baroni [2011] and Kuechler et al. [2012] report using name lists for USA only, we employ a much broader search across 30 countries, and use additional heuristics going beyond name frequency (Section 5.2).

3. RESEARCH DESIGN

This section presents the research design of this study following the *Goal-Question-Metric* (GQM) approach [Basili and Weiss, 1984].

3.1. Goal

The aim of this work is to produce a comprehensive study on how and when women engage in software-development-related online communities (hereafter referred to as "online communities"), or whether specific communities perform better at attracting and retaining women.

Rationale—What is currently known about this "disengagement" phenomenon is still at the anecdotal level, and qualitative and quantitative studies are needed for three reasons: first, to produce a solid and reproducible understanding of the main cause of this disengagement; second, to evaluate the possible long-term implications of such online behaviour; third, to focus on gender-balanced online communities, in order to understand the reasons (if any) of gender-balanced communities and to feed them back to less welcoming communities.

3.2. Questions

This paper addresses the following research questions:

> RQ_1: **What are the challenges with identifying gender in online communities?**

Rationale — The identification of gender in online activities is complicated by several factors: some communities do not record the gender of their participants; users often choose gender-neutral names, or opposite-sex names to cope with a male-dominated environment; in specific countries, certain names are "unisex", therefore the names–to–gender translation has to be country-specific.

> RQ_2: **What is the rate of participation by women in online communities?**

Rationale — While the sharp decline of women in STEM-related subjects is well known, a quantitative study of how many women participate in online communities has not been achieved yet. Before trying to understand the possible reasons behind a possible under-representation of women, it is necessary to first delineate the issue.

RQ$_3$: **What is the level of engagement of women in online communities?**

Rationale — Even in case of extreme imbalance in the representation of gender in online communities, it is important to define whether women and men follow similar patterns of contribution and engagement. Showing that women engage less than men in these communities, but achieve similar levels of contribution, would produce a picture of a (relatively) "healthy" community. On the other hand, a community with an imbalanced representation of gender, where the levels of participation vary substantially with gender, would suggest a gender-specific community.

3.3. Metrics

The representation of gender and their levels of engagement are measured using various attributes:

- The *number of women and men* participating in online activities. In the case of mailing lists, the participation event occurs when users post messages to the appropriate list; in the case of SO, when a user proposes a new question, or attempts to answer an existing one[3].
- The *number of questions* posed by an individual to the online community. In the case of mailing lists, these are the messages with "no precedence", i.e. just posted, and not written "in reply to" any other messages. In the case of SO, these are the questions that are posted to the community, and that can be voted upon by members, edited, or even closed and deleted.
- The *number of answers* given by an individual. In the mailing lists, this is the number of messages (per individual) "in reply-to" other messages; in SO, it is the number of answers (that can be voted, down-voted, or edited/deleted) given by a user to pending questions.
- The *length of engagement* in an online community. Regarding the mailing lists, this was counted, per individual, as the number of days between the first and the latest message to the mailing list. For SO participants, the number of days between the first question or answer, and the latest question or answer given by a user.

Based on the questions and metrics formulated above, we posit a number of null hypotheses (reported in Table 1), to be tested via statistical testing (the exact machinery used is described in Section 5.4). The alternative hypotheses test whether the communities under study are gender-specific and biased towards men, and they are drawn from the collected anecdotal evidence.

4. GENDER AND SELF REPRESENTATION IN STACKOVERFLOW, DRUPAL AND WORDPRESS

In this section we discuss how the StackOverflow, Drupal and WordPress participants present themselves, in general, and their gender identity in particular.[4] Individuals construct their gender, e.g., as feminine or masculine in various ways, and they might present it, consciously or not, in a fashion that is contrary to one's sex. Moreover, individuals might be less eager to disclose their gender or reject the traditional gender binary.

4.1. StackOverflow

StackOverflow (SO) is a programming questions and answers (Q&A) website collaboratively built and maintained by programmers, and owned by Stack Exchange, Inc. Created in 2008, SO is the first and largest Q&A website in the Stack Exchange network, and was followed by more than 100 "siblings" on different topics, such as English language, video gaming, photography, or parenting. Moreover, SO has been followed by a number of international clones such as a German-language CodeKicker (started in 2009)[5] and a Russian-language HashCode (started in 2010)[6].

The SO user profile page (Figure 1) combines the representation of oneself (self-determined) with activity-related information (automatically provided by SO). Self-representation is expressed through the choice of a user name, the user's real name, personal web-site, email, location and age, as well as a free text self description. A user's gender is not explicitly recorded, but can be constructed here through name, location and avatar choice.

[3]Other events are possible in SO, such as commenting/editing posts. We only analyze participation related to posing/answering questions, considered the core activities for a Q&A website.

[4]Following Harding [1986, pp. 52–57] we distinguish between symbolic, structural and individual gender. In this and the forthcoming sections we focus on the individual gender, i.e., what counts as masculine and feminine identity and behaviour.

[5]http://codekicker.de/

[6]http://hashcode.ru/

Table 1. Null hypotheses to be tested, and their relation to the research questions

Null (H_0)	H_1	RQ
$H_{1,0}$: women and men are similarly represented	$H_{1,1}$: the number of women in online communities is under-represented, and it reflects the current trend of women enrolling in STEM-related subjects	RQ_1, RQ_2
$H_{2,0}$: women engage for a length of time statistically similar to men's	$H_{2,1}$: men engage for longer	RQ_2, RQ_3
$H_{3,0}$: women formulate a number of questions statistically similar to men's	$H_{3,1}$: men formulate more questions	RQ_2, RQ_3
$H_{4,0}$: women provide a number of answers statistically similar to men's	$H_{4,1}$: men provide more answers	RQ_2, RQ_3
$H_{5,0}$: women and men achieve similar levels of reputation	$H_{5,1}$: men achieve higher reputation levels	RQ_3

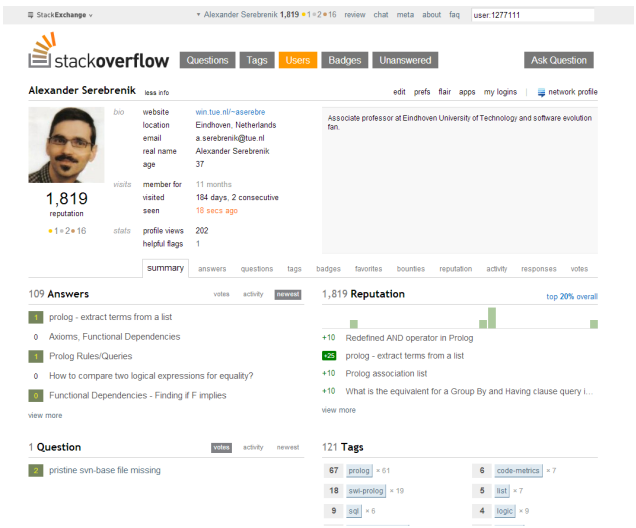

Figure 1. An SO user page combines the representation of oneself with the activity-related information.

Activity-based information (in the lower part of Figure 1) includes the reputation score, badges, questions asked and answers given, as well as tags, i.e., topics, frequent in the user's questions or answers. SO uses *gamification* and an activity-based reputation system: users receive "badges" for different actions performed (e.g., resurrecting and editing posts that were inactive for long periods, up-voting competing answers, or sharing links to questions in order to attract more viewers); similarly, users earn "reputation" points by posting interesting questions and answers (as reflected by the up-votes received from the SO community). The higher the reputation and the more badges one has, the more control she has over SO and other members' postings (e.g., users having earned certain badges can be elected to help moderate the site).

Registration is not obligatory in order to participate in SO: as indicated by a female programmer, referred to as B^{***}, the users might even prefer to select a different name for each question or answer [B^{***}, 2012]. The SO user page for unregistered users contains the same information as for regular users, with the additional indication *Unregistered* accompanying the user name.

4.2. Drupal and WordPress

Drupal[7] and WordPress[8] are free and open source content management platforms used as back-end systems for many websites, ranging from personal blogs to the Eclipse marketplace[9] or the Ubuntu homepage[10]. Similarly to SO, the Drupal user page combines self-representation with activity indication. Gender is recorded explicitly although this field is not mandatory: should the user be interested in indicating her gender, she can choose between *male*, *female*, *transgender* and *other*. Moreover, gender is shown through the choice of names (full name, first name, last name and user name) and location but, interestingly, not an avatar.

In the WordPress community, gender is not explicitly recorded, but it is shown by the members through the choice of the username and the avatar. While WordPress records the users' real name, the real name is not published. In both the WordPress and Drupal mailing lists, produced by the respective communities, full names and emails are available for inspection.

[7] http://drupal.org
[8] http://wordpress.org
[9] http://marketplace.eclipse.org
[10] http://ubuntu.com

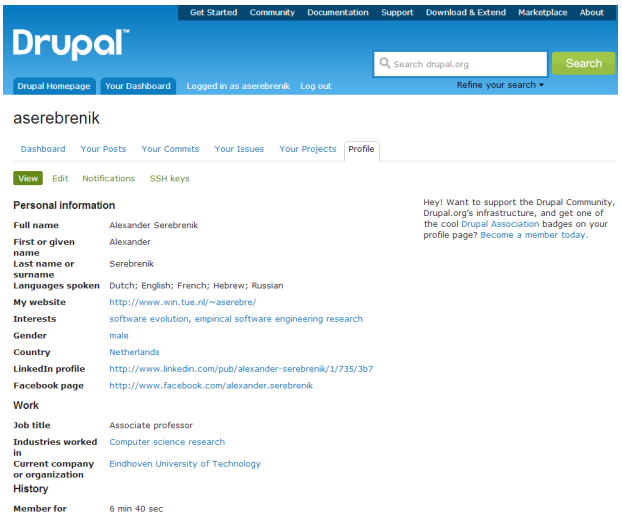

Figure 2. A Drupal user page records gender-related information (for the community members who chose to do so).

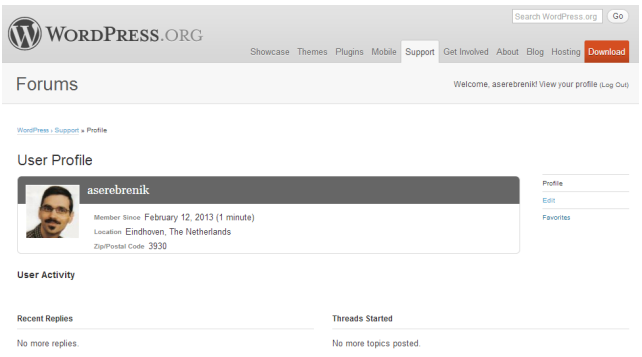

Figure 3. A WordPress user page does not reveal the real name of the community member.

5. EMPIRICAL APPROACH

As discussed in Section 4, gender of participants in online communities is not always recorded. The StackOverflow and WordPress communities do not record gender at all, while Drupal records gender-related information only for the community members who decide to do so. However, as detailed below, it was found that only a fraction of "registered users" are also "participants" in the Drupal mailing lists. Therefore, one needs to *infer* the gender of participants (as opposed to "registered users") in all three online communities.

The empirical approach followed involves *automatic* and *manual* steps. The *automatic* process comprises inferring gender *based on a person's name* and, if available, their location. Despite the cases of "gender swapping" discussed in Section 2 and anecdotal evidence that "gender swapping" occurs in StackOverflow, Armentor-Cota [2006] indicated that, e.g., on chat forums there were many more performances of dominant gender identities than there were transgressive acts. Moreover, given the masculine culture of technology [Wajcman, 2010] and our focus on online communities focusing on technology, we expect male names to predominantly correspond to masculine identities. We can be even more confident in genders reconstructed for Drupal and WordPress mailing lists: Marshall [2006] conjectures impersonation to be less frequent when people are identified by e-mail addresses rather than usernames.

The *manual* process is based on inferring gender *based on a person's avatar picture* (only available for StackOverflow participants), or *based on additional data sources*, such as GitHub, Twitter, Flickr, Facebook, GMail or LinkedIn. The manual process was performed only for those participants for which the automatic process did not result in a gender inference. All results were manually reviewed. The details about data extraction, as well as specific challenges pertaining to the datasets are described below. The accuracy of the gender resolution process is discussed in Section 6.

5.1. Obtaining the data

5.1.1. *StackOverflow*

All public data in SO (including the list of members and data about their activity) can be downloaded as part of the Stack Exchange data dump[11]. In this article we explore the data dump dated April 2012, containing information about 1,078,708 registered users. The SO dataset starts in August 2008, which corresponds to the creation date of the website. Since our gender-resolution process is only partly automatic, we decided to sample a smaller set: to obtain a 2% margin error and 99% confidence, a random sample of 4,144 SO users was extracted. Hereafter, whenever we discuss SO results, they have been obtained on this sample.

5.1.2. *Drupal and WordPress Mailing Lists*

The mailing lists analysed for this study are the public lists appearing in the Drupal [2012] and WordPress [2012] websites. The data from the mailing lists and their participants was analysed using the *mlstats* tool [Robles

[11] From http://www.clearbits.net/torrents/2017-apr-2012

Table 2. Characteristics of the studied mailing lists. The value in the "overall" row is not the sum of the preceding rows, since the same person may have participated in multiple lists.

Drupal	first	last	messages	participants
consulting	12/2005	05/2012	5,790	746
development	04/2005	05/2012	38,702	1,541
support	04/2005	05/2012	22,308	1,627
themes	12/2005	05/2012	1,279	309
translations	12/2005	05/2012	920	154
overall				3,342
WordPress	first	last	messages	participants
wp-accessibility	08/2009	08/2011	29	18
wp-docs	02/2005	05/2012	3,133	281
wp-edu	11/2008	05/2012	578	127
wp-hackers	02/2005	05/2012	43,744	2,128
wp-testers	03/2005	05/2012	14,504	1,451
wp-ui	01/2010	05/2012	279	105
wp-xmlrpc	06/2007	05/2012	633	80
overall				3,611

et al., 2009]. Data on messages, dates, senders and receivers are recorded and stored automatically. Table 2 presents basic facts of the studied mailing lists. Note that the mailing lists are generally older than StackOverflow (which started in 2008). Note also that the owners of the Drupal and WordPress mailing lists further partitioned these lists into specific sublists, to keep topics clearly separated. For example, WordPress separates discussions between members of the community who want to contribute to the documentation (hosted in the "wp-docs" list), from discussions between people interested in extending WordPress through plugins or improvements to the core code (hosted in the "wp-hackers" list). Such separation is often also indicative of the differences in expected vocabulary (or skills) of the participants: for example, "wp-docs" states that "no coding skills [are] required, just a lot of patience", while "wp-hackers" assumes "a certain level of working knowledge of WordPress and PHP" from its participants [WordPress, 2012].

Two additional public sublists for WordPress have been excluded from the analysis, since the vast majority of their traffic is automatically generated: "wp-trac" mirrors changes to the issue tracking system of WordPress, while "wp-svn" mirrors updates to the WordPress SVN repository. While in both cases these changes have likely been originally performed by humans, their mirrored versions, automatically generated, lose this authorship information. No such lists exist for Drupal.

Since the two communities are relatively small (3,342 participants for Drupal, and 3,611 for WordPress, respectively), no sampling was required to perform the analysis.

5.2. Automatic gender resolution

A person's name is often indicative of their gender. For example, John is a common male English first name, while Claire is a common female English first name. Corroborated with location information, even more accurate inferences can be made about one's gender based on their name. For example, Andrea is a common male first name in Italy, but a common female one in Germany. To implement this approach, we iteratively built lookup tables with first names for countries where this information was accessible online. Whenever available (e.g., when the data came from national statistics institutes), we also record the name usage frequency per gender[12]. Both our gender inference tool as well as the data on which it is based are available on GitHub.[13]

5.2.1. Preprocessing

The goal of the preprocessing step consists in obtaining the *(name, country)* tuples whenever possible. The first activity to perform is preprocessing the names, and purge the accents, special characters and symbols. We also converted the names from Leet [Ross, 2006] to Latin: for example, w35l3y is converted to Wesley.

Next, we tried to identify the real names of those participants who choose not to disclose them, but instead use *nicknames* (e.g., Carrotman, CoffeeCode) or, in case of SO, standard SO-assigned usernames (e.g., user4106). To determine the real names of such SO users we crawled and parsed their personal webpages, if linked from their SO profile pages. Moreover, if one person has multiple SO accounts, information obtained from one of the accounts can be used to infer gender for another one. To identify accounts belonging to the same person, we made use of email hashes[14]: accounts associated with the same hash have the same email address, hence belong to the same person. For example, user357032 shares the same email hash with Thomas Johnson, so gender can be inferred for the former using the latter's name.

Location information is only available for the StackOverflow users that choose to describe it in their SO profiles, and is not available for mailing list participants. Even for SO, only a fraction of the users in our sample

[12] We have compiled lists for Albania, Australia*, Belgium*, Brazil, Canada*, Czech Republic, Finland, France, Greece, Hungary, India, Iran, Ireland*, Israel, Italy*, Latvia, Norway*, Poland, Romania, Russia, Slovenia*, Somalia, Spain*, Sweden*, The Netherlands, Turkey, UK*, Ukraine, USA*, and Vietnam. The asterisk denotes countries with frequency information.

[13] https://github.com/tue-mdse/genderComputer.

[14] The actual email addresses are not publicly available, for privacy reasons; the MD5 hashes, however, are.

specify location (821 out of 4,144, or 19.8%), and not all user-specified addresses refer to geographic locations (e.g., The Matrix). Whenever available, the locations were fed into the Google Maps geocoding service[15], and the relative country (if available) was recorded.

5.2.2. Gender resolution process

In order to resolve the gender of a participant, we developed a Python tool that resolves a name using the lookup tables discussed above, and a number of heuristics. The tool, available on GitHub at `https://github.com/tue-mdse/genderComputer`, takes a *(name, country)* tuple as input, and returns one of "feminine", "masculine", or "x" (i.e., no gender can be inferred). The resolution algorithm starts with the identification of the first and the last name, and continues with gender detection based on gender-specific last name forms (e.g., *-ova* in Russian), country-specific lookup tables, cross-country lookup and diminutive resolution.

For example, given *(Anna Akhmatova, Russia)* the tool infers "feminine" due to a gender-specific last name form. For *(Jean Delville, Belgium)* the tool chooses "masculine" since Jean is much more frequent as a male name in Belgium (90,211 occurrences as a male name, as opposed to only 98 as a female name). In general we require the name to be at least twice more frequently used for men as for women to infer "masculine", and to be at least twice more frequently used for women as for men to infer "feminine". For *(Bogdan Lalić, Croatia)* the tool again chooses "masculine" despite the fact that we do not have data for Croatia: Bogdan is recorded only as male in all lookup tables that include this name. Observe, however, that we cannot (reliably) infer gender for *(Andrea Demirović, Montenegro)*, since we do not have data for Montenegro and different countries list Andrea both as a male and a female name.

If none of the above results in a resolved gender, and the name contains a single name part (i.e., it resembles a username), we assume it is formatted according to common naming conventions for usernames [Bird et al., 2006] (e.g., johns for John Smith), and restart the process (e.g., with john derived from johns). We evaluate the accuracy of our automatic gender resolution algorithm in Section 6.3, by comparing against the results of a small survey conducted among the SO-users, in which the respondents indicated gender.

5.3. Manual gender resolution

Considering the SO names, not all participants choose names that the automatic name-based gender resolution algorithm can successfully parse (e.g., because they prefer nicknames such as CoffeeCode). To improve the accuracy of the automatic gender inference process, whenever it does not result in an inferred gender (i.e., one of either masculine or feminine), we manually inspect additional sources of information: avatar pictures (available for some of the SO users), and websites such as GitHub, Twitter, Flickr, or LinkedIn, which may help reveal a person's real name.

SO users have the option of displaying an avatar picture on their profile pages. We have manually inspected the avatar pictures of the SO users in our sample, and tried to infer gender. However, not all users upload pictures of themselves (e.g., some use default geometric patterns, celebrity stock photos, cartoons or symbols). We do not infer gender from geometric patterns or symbols. Moreover, some of the SO users choose to display pictures of their babies or children, and some display photos of animals, places, or artificial symbols. We chose not to infer gender from such avatars. Ultimately we rely on heuristics to infer gender from the gender of the person or character depicted in the photo. For example, we infer "masculine" from a picture of Kenny McCormick, the South Park character, and "feminine" from a picture of Angelina Jolie. Clearly, the manual gender identification is a subjective process that can be put in jeopardy by "gender swapping" discussed in Subsection 2.1.3; this and additional threats to the validity of our results are discussed in Section 8.

We have also observed that people often use the same way to identify themselves (e.g., the same avatar picture, or the same nickname) in other online communities where they are participating (e.g., GitHub, Twitter, Flickr, or LinkedIn). However, the level of personal information available for a given person in each of these communities may differ. For example, a person's Twitter account may also display her full name, or a person's avatar picture may also be used when she comments on blog posts, where she signs with her full name. Whenever we could not directly infer gender using the approaches above, we manually investigated the information available from one's participation in other online communities, and tried to infer gender from full names, as discussed above.

As mentioned above, we also evaluated the accuracy of the combination of the automatic and manual gender resolution on the pilot survey data (for which the respondents themselves have indicated gender), described in Section 6.3.

5.4. Statistical analysis

Testing for significance the hypotheses reported in Table 1 can be viewed as a comparison of multiple distributions: for example, given the total number of questions asked on a mailing list, we had to compare three distributions, two

[15] `https://developers.google.com/maps/documentation/geocoding/`

generated by the masculine and feminine gender, and one by the unknowns, i.e., those users for which we could not infer gender.

Traditionally, comparison of multiple groups follows a two-step approach: first, a global null hypothesis is tested, and then multiple comparisons are used to test sub-hypotheses pertaining to each pair of groups. The first step is commonly carried out by means of ANOVA or its non-parametric counterpart, the Kruskal-Wallis one-way analysis of variance by ranks [Holander and Wolfe, 1973]. The second step uses the t-test or the rank-based Wilcoxon-Mann-Whitney test [Wilcoxon, 1945], with Bonferroni correction [Dunn, 1961, Sheskin, 2007]. Unfortunately, the global test null hypothesis may be rejected while none of the sub-hypotheses are rejected, or vice versa [Gabriel, 1969].

Therefore, one-step approaches are preferred: these should produce confidence intervals which always lead to the same test decisions as the multiple comparisons. To this end, we employ the recently-proposed multiple contrast test procedure $\widetilde{\mathbf{T}}$ [Konietschke et al., 2012] using the traditional 5% family-wise error rate (equivalent to a level of significance of 95%). $\widetilde{\mathbf{T}}$ is robust against unequal population variances and is applicable to different types of contrasts, including comparisons of all pairs of distributions (using the so-called *Tukey-type contrast* [Tukey, 1951]).

However, not all all combinations of mailing list and metric lead to comparisons of three groups (e.g., there are no unknowns in the "wp-accessibility" sublist of WordPress). In such cases we fall back on the non-parametric Wilcoxon-Mann-Whitney test [Wilcoxon, 1945] for assessing whether one of two samples of independent observations (i.e., "masculine" or "feminine") tends to have larger values than the other for a given metric (e.g., number of questions asked). We selected the Mann-Whitney test rather than the t-test since we observed that the distribution of values in each sample is highly skewed (i.e., not normal), while the t-test requires the distributions of values to be normal. Indeed, few participants contribute a lot (e.g., in terms of number of messages or answers), and many others participate very little (e.g., sending very few messages). The Wilcoxon-Mann-Whitney test consists of calculating a test statistic U and comparing its distribution with a distribution which is known under the null hypothesis. The result of this comparison is a p-value. If the p-value is lower than a predefined threshold (we use the traditional threshold of 0.05, equivalent to a level of significance of 95%), then we can reject the null hypothesis and accept the alternative hypothesis.

6. ACCURACY OF GENDER RESOLUTION

To obtain insights in the demographics of SO, and to test the accuracy of the gender resolution algorithm (discussed in Sections 5.2 and 5.3), we conducted a pilot survey. The survey was not used to evaluate the gender of all the participants of the SO network; rather, it served as a proof of concept of the assumptions used in the semi-automated approach.

6.1. Design and participation

We asked the respondents to indicate their SO userid, gender, age, country of birth, country of residence, highest education level obtained and years of professional experience, as well as involvement in open-source and proprietary software development. We obtained 141 responses, including 127 valid ones (e.g., a unique SO userid mapped to an individual). Since the responses were obtained voluntarily, composition of the sample is likely to be affected by a selection bias. However, this data was only used to derive qualitative conclusions.

6.2. Results

Our first observation is that the majority of respondents are men: only 12 respondents from 127 have identified themselves as women. Using the 'score' method of Wilson [1927] as described by Newcombe [1998] and implemented by Lowry[16], with a 95% confidence interval we estimated the proportion of women in SO in the $[5.5\%, 15.8\%]$ interval.

We also noted that the respondents are predominantly involved either exclusively in proprietary software (47 respondents) or both in proprietary and open source software (47), while the number of exclusively open source developers is lower (17)[17]. This means that *a priori* one could have expected the share of women on StackOverflow to be between 1-5% reported for open source projects [David and Shapiro, 2008] and 28% reported for proprietary software [National Science Foundation, 2004].

The 95% confidence interval calculated based on the survey data fits within the expected range. Still, it could be the case that women were less inclined to participate in the survey as the information about age, country of birth or country of residence can be considered private, and they might prefer not to disclose it. Hence, we verified this expectation on a larger scale in Section 7.

[16] http://www.vassarstats.net/prop1.html

[17] The remaining respondents are either not involved in software development at all or they indicated a more elaborate answer than "yes"/"no".

Table 3. Accuracy of the automatic gender resolution approach described in Section 5.2 on the pilot survey data.

Gender as described by 117 respondents	**Gender as inferred by automatic approach (name-based)**			
	masculine	feminine	unknown	total
masculine	60	3	43	106
feminine	2	5	4	11
Precision $\simeq 0.93$	Recall $\simeq 0.58$		Accuracy $\simeq 0.55$	

Table 4. Accuracy of enhanced gender resolution approach, combining the automatic steps described in Section 5.2 with the manual steps described in Section 5.3, on the pilot survey data.

Gender as described by 117 respondents	**Gender as inferred by the enhanced approach**			
	masculine	feminine	unknown	total
masculine	90	3	13	106
feminine	2	9	0	11
Precision $\simeq 0.95$	Recall $\simeq 0.88$		Accuracy $\simeq 0.85$	

Finally, we observed that a significant group of respondents (25 out of 127) no longer resides in the countries of their birth due to personal, professional or educational reasons.

6.3. Accuracy in the sampled users

Since gender is one of the entries the participants in the pilot survey filled in, we can use this data to evaluate the accuracy of the gender resolution approaches. However, 10 out of 127 valid responses (1 woman, 9 men) correspond to users who joined SO after the data dump was released, so we cannot verify whether their "inferred" gender matches their "described" gender. Therefore, we restrict our accuracy analysis to the remaining 117 SO users.

Tables 3 and 4 display the number of participants with correctly- and incorrectly-inferred gender, the number of persons with unresolved gender, as well as precision, recall, and accuracy for the automatic and combined automatic+manual gender resolution approaches. Recall from the previous section that the manual approaches (based on avatar pictures or information from other online communities) were only used as an enhancement to the automatic name-based approach, whenever it could not infer a gender. We observe that although the automatic approach has high precision (91%), it has only average recall (58%), due to many SO users preferring usernames or nicknames over real names.

Furthermore, 5 users are misclassified (3 users described themselves as male while the automatic approach inferred "feminine", and 2 described themselves as female and the approach inferred "masculine") due to (parts of) their usernames corresponding to either male or female first names in some countries.

On the other hand, the enhanced approach, combining the automatic steps with the analysis of avatar pictures and additional data sources, correctly identified gender of 34 further individuals labeled as "unknown" by the automatic analysis (cf. Table 4). In this way, the enhanced gender resolution approach achieves higher recall (88%) with high precision (94%), although gender for 13 out of 117 users (or 11%) still cannot be inferred. Failure to infer any gender is due to absence of avatar pictures and use of common words as usernames (e.g., Ubiquité, phant0m, or bitmask), which cannot be confidently assumed to belong to the same individual in other online communities.

Given the sample size, and with a 95% confidence interval, we estimate the precision and recall within the $[84\%, 97\%]$ and $[49\%, 67\%]$ ranges (respectively) for the automatic approach; similarly for the enhanced approach, the precision and recall are in the $[89\%, 98\%]$ and $[81\%, 93\%]$ ranges, respectively.

(RQ_1) Gender of SO participants is typically not recorded. To infer gender one can use both automatic and manual techniques, with different accuracies. Give a person's *name* and, if available, their *location*, gender can be *automatically* inferred with high precision ($\simeq$ 95%). However, the method has relatively low recall ($\simeq$ 60%), largely due to the fact that not all participants in online communities disclose their real names, and even fewer reveal their locations. However, recall can be further improved (up to $\simeq$ 90%) by *manually* inspecting other data sources, such as *avatar pictures*, or profile pages of the same users in *other online communities* (e.g., GitHub, Twitter, Flickr, or LinkedIn). Such sources may provide additional clues as to a person's real name and location, if not directly their gender.

7. RESULTS

This section is articulated in two parts: we start with the mailing list results in 7.1, then discuss the StackOverflow results in 7.2.

7.1. Mailing Lists: Drupal and Wordpress

We explore the representation (*what is the fraction of mailing list participants in each gender?*) and engagement of gender (*how does the length of engagement with the community and the amount of activity vary with gender?*) along two dimensions: (i) granularity level (i.e., considering all the sublists together and analysing the mailing lists at large, or at the level of each sublist); and (ii) type of engagement (i.e., distinguishing between *asking* and *answering*, since not all participants are necessarily involved in both). Moreover, we test the robustness of the analysis against the "unknowns" (those participants for which we could not infer gender).

7.1.1. Hypothesis testing – Gender Representation

When analysing the Drupal mailing lists at large (considering all the sublists together), we inferred "masculine" as gender for 2,879 (3,043 in Wordpress) participants, "feminine" for 328 (282) participants, and "unknown" for 135 (286) participants[18]. These numbers show a first result that is shared for both Drupal and WordPress: some 8-10% of the overall participants construct their gender as feminine, while the vast majority of participants construct their gender as masculine (some 85%). The fraction of participants labelled "unknown" is approximately 4% in Drupal and 8% in WordPress.

Repeating the analysis at the level of individual sublists (i), on the one hand, and only for the populations of askers and answerers (ii), on the other hand, yields similar results: as an example, Figure 4 shows the gender ratios of participants in the individual sub-lists of Drupal (top), and of participants answering questions (bottom) in the same community. Bar chart labels correspond to the names of the Drupal mailing lists (cf. Table 2). Similar proportions are found in the number of participants posing questions in Drupal (not shown), and in all the sublists of the Wordpress community (also not shown).

Variation across sublists

The ratio of women participating in these sublists appears quite stable. However, small differences can be observed between different lists or within certain populations (e.g., from Figure 4 it seems that there are fewer women in the "development" sublist than on other sublists, both when considering the entire population, as well as only the population of participants answering questions). To check whether these differences are statistically significant, we applied the $\widetilde{\mathbf{T}}$-procedure. For participants labelled "unknown", we considered three different scenarios:

[A] excluding unknowns from the sample,

[18] For most of these unknowns, we observed a very poor engagement to the lists, and often a spamming activity to the same lists.

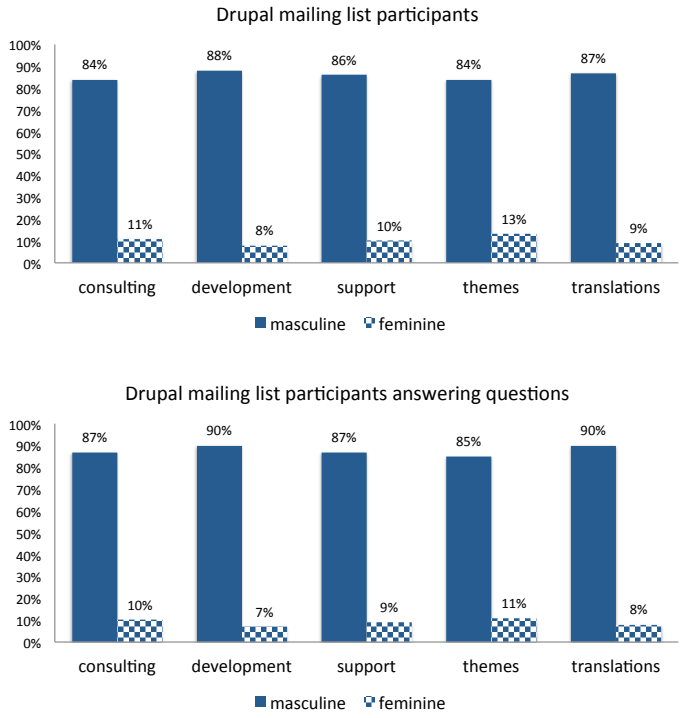

Figure 4. Gender of participants in the Drupal mailing lists. Overall (top) and participants answering questions only (bottom). "Unknowns" (not shown) account for the remainder up to 100%.

[B] assuming all unknowns to be "men", and
[C] assuming all unknowns to be "women".

For the Drupal community, we found that the only significant differences were in the the population of *askers* in the "development" and "consulting" mailing lists. If the individuals labelled "unknown" are excluded (scenario [A]), or all of them are assumed to be men (scenario [B]), the number of women asking questions on the "development" mailing list is lower than on the "consulting" mailing list. In all the other scenarios, and mailing lists, no statistically significant differences exist. In the WordPress dataset, statistical significance of the comparisons is more sensitive to the scenario (likely due to a higher proportion of "unknowns" than in Drupal).

However, what is common to both Drupal and WordPress is that the differences in gender participation occur mostly between mailing lists focussing on *designing* technology ("development", "wp-hackers" and "wp-xmlrc") and *using* technology ("consulting", "wp-docs" and "wp-edu"). This observation confirms the anecdotal evidence that "the more technical the role, the less likely a woman will be involved" [Lyman, 2005]. Moreover, it concurs with the old adage "boys invent things and girls use things boys

invent" [Darrow, 1970] which, as observed by Margolis and Fisher [2003, p.2], seems to remain "uncomfortably true today".

In conclusion, and in terms of a formal test to ascertain whether gender is under-represented, the null hypothesis $H_{1,0}$ states that "women and men are similarly represented" in online communities. Given the disparity in numbers in both Drupal and WordPress, we had to reject the null hypothesis $H_{1,0}$ regarding the gender representation in the mailing lists, while not being to able to completely reject the alternative hypothesis $H_{1,1}$ with an adequate significance.

Active versus Passive Users

The results on gender representation can be complemented with additional observations, regarding active and passive users of the WordPress and Drupal communities:

- A larger layer of "users" of these communities is composed of visitors, potentially not engaged in the discussions, nor registered as users. The collected data for the US audience provided by the QuantCast company[19] for the two communities[20], and related to the gender of the visitors, shows even less divide between the genders: 44% of the traffic on the Drupal website comes from female users (and, similarly, 39% for WordPress).
- A different ratio is also observed in the number of "interested" users: when screen-scraping the users registered on the Drupal.org website (i.e., the "interested" users), and classified as male or female members[21], it is found that some 12,000 users are registered as female, while 63,000 are registered as male, hence reflecting a 15/85 divide between genders (of registered users). Only 43 users indicated their gender as "transgender" and 209 as "other".
- The registered users do not always match the participants in the Drupal mailing lists: many of those users do not participate in the discussions, and various mailing list participants are not registered as Drupal users.

(RQ_2) The numbers of active participants in the Drupal and WordPress mailing lists (at large, and in the sublists) show that indeed there is a disparity in the gender representation in those two online communities (women account for approximately 10% of the participants). Such disparity is even more acute in the presence of more focused and technical aspects of developing and programming these software systems. On the contrary, this imbalance is not visible in the number of interested (or "passive") users: the male÷female ratio of content users, or readers, shows a much more balanced divide between gender.

7.1.2. Hypothesis testing – Gender Engagement

The previous section shows clearly that the gender representation in the Drupal and WordPress communities is highly imbalanced. The interested audience (i.e., the "readers"), the registered users and the mailing list participants get in fact more and more biased towards a male audience, as long as more active involvement is needed. However, these facts alone are not enough to quantify the level of engagement, and the differences between gender in these communities: given for granted that a larger number of masculine participants exist, is it true that their activity and length of engagement with the community are significantly different from the feminine ones?

To answer this question, for each participant and each mailing list we recorded (i) their gender, (ii) the number of questions that they posed to the mailing list audiences, (iii) the number of answers that they gave to any of the questions posed[22], and (iv) the number of days between the first and the last posted messages.

In order to discount for the experience of a user, we normalised the numbers of questions and answers by the number of days in which she was active in the mailing list (likely, a user being around for longer asked and answered more questions than a novice user).

We separated the analysis into two parts: first, we analysed the engagement within the community, and we tested the following null hypothesis:

$H_{2,0}$: women engage for a length of time statistically similar to men's.

This hypothesis corresponds to the observation of Kuechler et al. [2012] based on six different open source projects: no statistically significant differences were observed between men and women.

Second, we compared activity (number of questions asked, and number of answers given), and we tested the following null hypotheses:

$H_{3,0}$: women formulate a number of questions statistically similar to men's.

$H_{4,0}$: women provide a number of answers statistically similar to men's.

[19] http://www.quantcast.com

[20] http://www.quantcast.com/drupal.org#!demo&anchor=panel-GENDER, http://www.quantcast.com/wordpress.org#!demo&anchor=panel-GENDER

[21] http://drupal.org/profile/profile_gender/male and http://drupal.org/profile/profile_gender/female

[22] The *mlstats* tool, used for the analysis of the mailing lists, can tag a message as an "answer" if it was sent as a response to an earlier question.

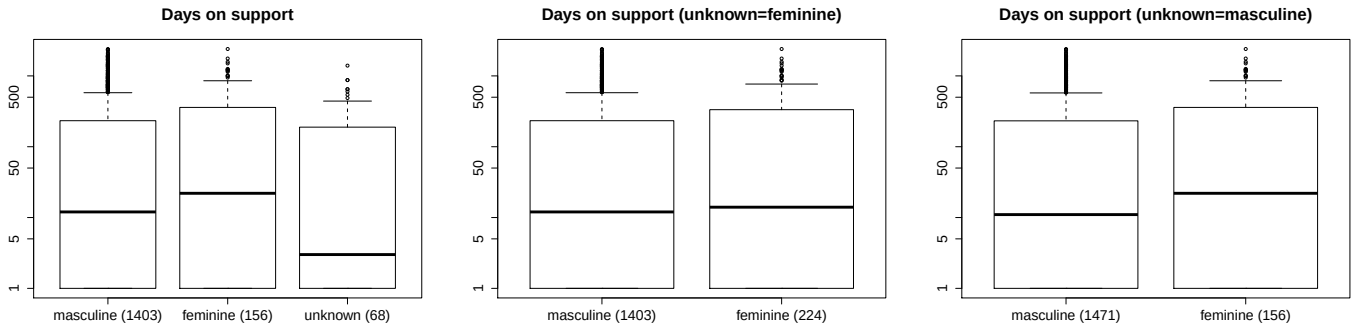

Figure 5. Left: length of engagement on the "support" list in Drupal, by gender; middle and right: hypothetical "what if" scenarios where the unknown gender is considered feminine and masculine, respectively. With each gender we display the number of users between parentheses.

Length of engagement with the community

To illustrate the analysis, consider the example of the Drupal "support" sublist depicted in Figure 5. The boxplots capture the distributions of the length of engagement on this sublist (measured in days), for each gender and for the unknowns (those users for which our gender inference process did not produce a result). In the middle and right subfigures we depict hypothetical "what if" scenarios, wherein the unknown gender is considered feminine and masculine, respectively, to assess the sensitivity of the analysis to the gender inference process. By applying the $\widetilde{\mathbf{T}}$-procedure to the three groups (masculine, feminine and unknown; left subfigure), we do not find sufficient evidence to reject the null hypothesis

> $H_{2,0}$: women engage for a length of time statistically similar to men's.

at a 95% confidence level. Then, by applying the Wilcoxon-Mann-Whitney test to the two groups (masculine and feminine; middle and right subfigures) in the "what if" scenarios, we again did not find sufficient evidence to reject the $H_{2,0}$ with a 95% confidence level ($p = 0.89$ and $p = 0.45$, respectively).

Table 5 summarises the results of the statistical testing for the other sublists. For each sublist (row), we report the number of masculine, feminine and unknown-gender participants. The ticks show where the relative null hypotheses could not be rejected with a 95% confidence level. The "what if" column describes the three scenarios of treating unknowns. As an example, the line

> **development**; 1,362; 120; 59; ✓; $H_{2,0}$ is rejected at 95% confidence level in favour of $H_{2,1}$ (men are engaged for longer) if all the unknowns are feminine.

from Table 5 shows that on the "development" sublist of Drupal we have identified 1,362 users as masculine and 120 as feminine, we could not infer gender for 59 users, and we cannot reject the null hypothesis $H_{2,0}$: "women engage for a length of time statistically similar to men's". However, if all the unknowns were feminine, $H_{2,0}$ would be rejected in favour of $H_{2,1}$: "men engage for longer".

Note that for both Drupal and WordPress, the null hypothesis $H_{2,0}$ cannot be rejected for any of the sublists: there is no significant difference in the length of engagement with the community between feminine and masculine participants.

> (RQ_3) Both Drupal and WordPress have a larger number of masculine participants than feminine participants in their mailing lists. However, this imbalance does not describe the length of gender engagement with the community: both genders engage for lengths of time which are statistically similar.

Amount of activity

To compare the amount of activity on the mailing lists across genders, we repeat the procedure described above. The results of the statistical testing for the number of questions asked, and for the number of answers given on each list are presented in Table 6. The values have been normalised to account for the potentially confounding effect of the length of engagement. Note that the number of masculine, feminine and unknown-gender participants in Table 6 differs from the corresponding values in Table 5: indeed, as mentioned earlier, not all users engage in both asking *and* answering questions.

The null hypotheses ($H_{3,0}$: "women formulate a number of questions statistically similar to men's" and $H_{4,0}$: "women provide a number of answers statistically similar to men's") cannot be rejected for any of the sublists.

Table 5. Results of the statistical analysis for length of engagement with the community. For each sublist we display the number of users tagged masculine, feminine and unknown. ✓ - hypothesis cannot be rejected at 95% confidence, X - hypothesis is rejected

DAYS - DRUPAL MAILING LISTS					
List	masculine	feminine	unknown	$H_{2,0}$	"What if" scenarios
consulting	630	85	31	✓	$H_{2,0}$ is rejected at 95% confidence level in favour of $H_{2,1}$ (men are engaged for longer) if all the unknowns are feminine, or if all the unknowns are masculine.
development	1,362	120	59	✓	$H_{2,0}$ is rejected at 95% confidence level in favour of $H_{2,1}$ (men are engaged for longer) if all the unknowns are feminine.
support	1,403	156	68	✓	no change
themes	259	39	11	✓	no change
translations	134	14	6	✓	$H_{2,0}$ is rejected at 95% confidence level in favour of $H_{2,1}$ (men are engaged for longer) if all the unknowns are feminine.
DAYS - WORDPRESS MAILING LISTS					
List	masculine	feminine	unknown	$H_{2,0}$	"What if" scenarios
wp-accessibility	12	5	1	✓	no change
wp-docs	223	38	20	✓	no change
wp-edu	103	20	4	✓	no change
wp-hackers	1,870	130	128	✓	$H_{2,0}$ is rejected at 95% confidence level in favour of $H_{2,1}$ (men are engaged for longer) if all the unknowns are feminine, or if all the unknowns are masculine.
wp-testers	1,184	116	151	✓	$H_{2,0}$ is rejected at 95% confidence level in favour of $H_{2,1}$ (men are engaged for longer) if all the unknowns are feminine, or if all the unknowns are masculine.
wp-ui	105	11	7	✓	no change
wp-xmlrpc	73	1	6	✓	$H_{2,0}$ is rejected at 95% confidence level in favour of $H_{2,1}$ (men are engaged for longer) if all the unknowns are feminine.

Moreover, the results are remarkably stable with respect to the unknown-gender.

(RQ_3) Both Drupal and WordPress come across as "healthy" communities, in which no discrimination is visible in terms of contributed activity. Although women are outnumbered, this imbalance does not describe the level of gender engagement: the contributing activities of men and women are in fact comparable, in number of questions asked, answers given and length of engagement.

7.2. StackOverflow Q&A

StackOverflow is similar to mailing lists as the core activities are asking and answering questions. Software developers can forage or actively engage in SO or the mailing lists to seek solutions to technical challenges or help others by sharing their knowledge and expertise. However, the StackOverflow platform leaves room for additional actions besides Q&A (e.g., moderating the website, or editing existing posts); in addition, its gamified environment has the potential to alter the behaviour of its users, who compete to achieve higher reputation and status by being more active.

In particular, the fact that there can be registered users who neither asked nor answered any questions requires a more careful analysis than with the mailing lists, where the "visible" participants (i.e., those encountered by mining the list archives) have either asked or answered at least one question (in other words, have sent at least one email message, otherwise they wouldn't leave any "trace" in the archives).

7.2.1. Hypothesis testing – Gender Representation

Overall, we observed 2,296 users with masculine identities, 291 users with feminine identities, and 1,557 users for which a gender could not be identified, and are not considered as either. The high percentage of unknown-gender users (37.5%) contrasts sharply with the previous situation of the mailing lists, where we could not infer gender for at most 5-8% of the users. For 616 out of these 1,557 (or 40%) it is impossible to infer gender, since they have standard SO-assigned user-names (e.g., user1234), no avatar pictures, and no MD5 email hashes in common with other SO users. This suggests that as opposed to participating in the mailing lists, where not disclosing one's name and email address is more difficult

Table 6. Results of the statistical analysis for the number of questions asked and number of answers given. For each sublist we display the number of users tagged masculine, feminine and unknown. ✓ - hypothesis cannot be rejected at 95% confidence, X - hypothesis is rejected

NUMBER OF QUESTIONS - DRUPAL MAILING LISTS					
List	masculine	feminine	unknown	$H_{3,0}$	"What if" scenarios
consulting	299	51	16	✓	no change
development	884	81	42	✓	no change
support	1,054	129	55	✓	no change
themes	134	25	5	✓	no change
translations	94	10	4	✓	no change
NUMBER OF QUESTIONS - WORDPRESS MAILING LISTS					
List	masculine	feminine	unknown	$H_{3,0}$	"What if" scenarios
wp-accessibility	5	4	0	✓	no change
wp-docs	144	23	14	✓	no change
wp-edu	52	10	2	✓	no change
wp-hackers	1,384	102	98	✓	no change
wp-testers	760	77	101	✓	no change
wp-ui	39	7	5	✓	no change
wp-xmlrpc	56	1	4	✓	no change
NUMBER OF ANSWERS - DRUPAL MAILING LISTS					
List	masculine	feminine	unknown	$H_{4,0}$	"What if" scenarios
consulting	522	58	20	✓	no change
development	1,120	83	38	✓	no change
support	1,042	108	45	✓	$H_{4,0}$ is rejected at 95% confidence level in favour of $H_{4,1}$ (men ask more questions) if all the unknowns are feminine, or if all the unknowns are masculine.
themes	206	27	9	✓	no change
translations	103	9	3	✓	no change
NUMBER OF ANSWERS - WORDPRESS MAILING LISTS					
List	masculine	feminine	unknown	$H_{4,0}$	"What if" scenarios
wp-accessibility	9	4	1	✓	no change
wp-docs	174	27	12	✓	no change
wp-edu	75	15	3	✓	no change
wp-hackers	1,485	86	96	✓	no change
wp-testers	896	82	95	✓	no change
wp-ui	80	8	3	✓	no change
wp-xmlrpc	54	0	3	✓	no change

(e.g., it would require using "fake" email accounts created specifically for this purpose), SO users prefer to make use of the possibility offered by the platform to remain anonymous (since filling in personal information in SO profiles is not mandatory).

Another observation is that 34% of the users (or 1,416) in our SO sample have neither asked nor answered any question during the analysed period (Figure 6). A closer inspection reveals that only three of these have been engaged in other activities, such as editing posts, or commenting on existing ones. This suggests that virtually all of these inactive users create SO accounts but never use them. Note that SO does not require lurkers to create accounts in order to access existing posts, since all access is public. Moreover, we note that it is more likely for unknown-gender users in the inactive category to be "unresolvable" (i.e., impossible to infer gender for)—51%, than for those in the active category—29%. However, even for the active category, the unknown-gender users generally have a very low reputation on SO, denoting very little activity.

Out of the 2,587 users for which we could infer gender, for 18.9% of them we had to resort to additional sources of information besides their name (in other words, 81.1% of the masculine and feminine users have been resolved automatically using the name-based technique, while the rest have been resolved manually using the enhanced

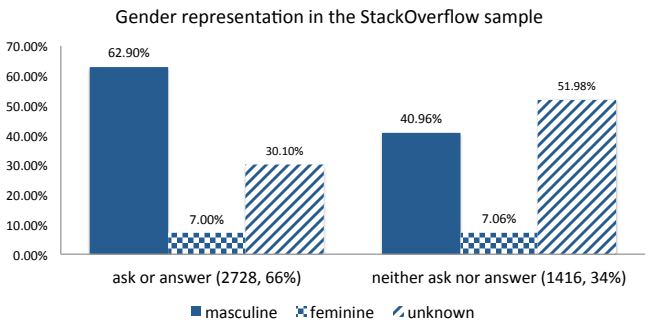

Figure 6. Gender representation in StackOverflow, across two different groups of users: those who ask or answer at least one question (left), versus those who neither asked nor answered any question during the analysed period (right).

technique): for example, we relied on avatar pictures for 181 (or 37%) out of these manually resolved users.

The numbers in both groups (active and inactive users) show an overall representation of women at around 7% of the participants[23], and a vast majority of male users, again rejecting the null hypothesis $H_{1,0}$ regarding the balanced gender representation in SO, and accepting the alternative hypothesis $H_{1,1}$. The overrepresentation of males is less visible for the group of inactive users (it seems it is more likely for low-activity users to remain anonymous than for high-activity ones, based on the observation above).

On the other hand, Quantcast reports that the audience of the SO website (i.e., people potentially not engaged in the discussions, nor registered as users) is more gender-balanced than the user community itself: 27% of the SO traffic comes from females[24].

(RQ_2) There is disparity in the gender representation of StackOverflow participants in our sample: approximately 55% present themselves as masculine and 7% as feminine, while gender could not be inferred for the remaining 38%. Inactive users are more likely to remain anonymous.

7.2.2. Hypothesis testing – Gender Engagement

The analysis of activity of StackOverflow users follows a similar two-step approach as in the case of the mailing lists. First, we assert whether men and women engage for lengths of time which are statistically similar. Second, we compare their activity in terms of the number of questions asked and the number of answers given. These values are

[23] Using the same methodology as in Section 6.2 we estimate the 95% confidence interval as [6.2%, 7.8%].

[24] http://www.quantcast.com/stackoverflow.com

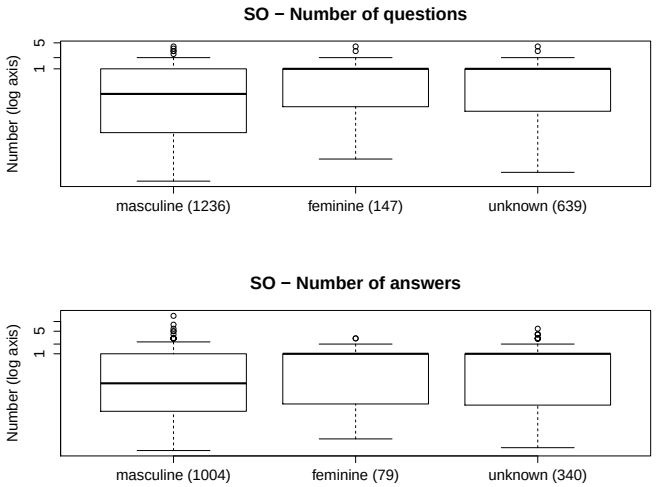

Figure 7. Gender engagement in StackOverflow: number of questions asked (top), number of answers given (bottom). The number of users in each group is displayed in between parentheses. Only users that ask, respectively answer questions have been compared, similar to the mailing lists.

again normalised, to account for the confounding effect of the length of engagement (see the discussion above). Moreover, computations are performed only on the group of active users (those asking or answering at least one question), to enable the comparison with the mailing lists (length of engagement cannot be determined for users that neither asked nor answered any questions).

By applying the $\widetilde{\mathbf{T}}$-procedure to the three groups (masculine, feminine and unknown), we reject the null hypothesis $H_{2,0}$–"women engage for a length of time statistically similar to men's" at 95% confidence level ($p = 5.41 \times 10^{-6}$), thus accepting the alternative hypothesis $H_{2,1}$–"men engage for longer". A "what if" analysis reveals no changes if the unknown-gender users were all masculine or feminine, respectively.

In terms of their asking activity (Figure 7 top), the $\widetilde{\mathbf{T}}$-procedure reveals statistically significant differences between the feminine and masculine users, thus rejecting the null hypothesis $H_{3,0}$–"women formulate a number of questions statistically similar to men's" ($p = 1.4 \times 10^{-4}$). However, the alternative hypothesis is different than in the case of the mailing lists: after controlling for length of engagement, women formulate more questions than men. The result does not change when the unknown-gender users are considered all masculine or all feminine, respectively.

In terms of their answering activity (Figure 7 bottom), the $\widetilde{\mathbf{T}}$-procedure does not reveal statistically significant differences between the feminine and masculine users: women provide a number of answers statistically similar to men's (thus $H_{4,0}$ cannot be rejected at 95% confidence

level). The result would only change if all unknowns were feminine, in which case an alternative hypothesis "women answer more questions than men" cannot be rejected at 95% confidence level.

(RQ_3) On StackOverflow it is clear that men engage for longer periods of time than women. Relative to the duration of their engagement in SO, however, women formulate more questions than men, while providing a similar number of answers. This results in a relatively "unhealthy" community where women disengage sooner, although their activity levels are comparable to men's.

7.3. Discussion and Implications

This paper aims at quantifying the gender representation in online communities, by focusing on very different environments. While it is *a priori* clear that men will constitute the majority of participants in the online developer communities, in the same way they constitute the majority of computer professionals [David and Shapiro, 2008, National Science Foundation, 2004], it was not *a priori* clear how big the disparity between men and women would be in these communities, and whether this disparity would be also visible at the level of engagement.

While studying the related work, we assumed a gender-neutral approach to the quantitative evaluation of the participation in the studied online communities, therefore we were not expecting any difference of participation and engagement between men and women in StackOverflow, Drupal or WordPress.

We have observed that Drupal, WordPress and StackOverflow are close in terms of gender representation: 7-10% of the participants in these communities construct their gender as feminine. This is in line with what has been discussed in the related literature, showing how gender is indeed under-represented in various parts of society: our finding is important to define what proportion of women *actively* engage with the studied communities, while it is still unclear whether a similar proportion could be found in a "passive" use of the online content.

In terms of gender engagement, two different kinds of communities seem to emerge. On the one hand, Drupal and WordPress come across as "healthy" communities, in which no discrimination is visible in terms of contributed activity. Although women are outnumbered, this imbalance does not describe the level of gender engagement: the contributing activities of men and women are in fact comparable, in number of questions asked, answers given and length of engagement. StackOverflow, on the other hand, appears to be a relatively "unhealthy" community where women disengage sooner, although their activity levels are comparable to men's. From the liberal-feminist approach, discussed in the related work, the Drupal and WordPress communities appear as gender-neutral in terms of engagement, but not on the level of participation. StackOverflow, on the other hand, seems to be better explained using a socialist-feminist framework, where technology is indeed linked to the masculine culture.

While a formal analysis of the reasons for the differences observed between Drupal and WordPress, on the one hand, and SO, on the other hand, goes beyond the scope of this paper, we would like to make a couple of remarks on this point. As also noticed by the participants of StackOveflow [VA, 2011a,b], the community around this Q&A site seems to create and maintain higher barriers to entry for women. It does so by designing an approach to collaboration based on earning prizes, achieving higher status and promoting and fostering extremely fast responses by the participants [Mamykina et al., 2011], in turn producing more reputation and status.

As opposed to StackOverflow, communication on Drupal and WordPress mailing lists focusses on the essence, asking and answering questions, and is not being publicly quantified by means of prizes, statuses or reputation points, and as such might be more attractive for women.

Reflecting on the differences between performance of women and men in a more (StackOverflow) and less (Drupal, WordPress) competitive environments, we conjecture that *the competitive nature of an online community not only creates higher entry barriers for women, but is also detrimental to their effectiveness*. This conjecture concurs with a series of laboratory experiments on gender differences in competitive environments. Gneezy et al. [2003] have observed women to be less effective than men in competitive environments, even if they are able to perform similarly in non-competitive environments. Moreover, the observed effect has been stronger when women have to compete against men than in single-sex competitive environments. In a different experiment, Niederle and Vesterlund [2007] have observed that women shy away from competition and men embrace it. We believe that quantitative verification of the conjecture above constitutes an important direction for future work.

Going beyond the differences we have observed between different online communities, we stress a number of unsolved challenges pertaining to gender-specific reluctancy in equally participating to specific online communities: from encouraging women to enter the field of technology; to their engagement in online communities

for expert help and advice; to their sharing of knowledge with other members.

8. THREATS TO VALIDITY

The validity of this study is subject to several threats. In the following, threats to *internal validity* (whether confounding factors can influence the findings), *external validity* (whether results can be generalized), and *construct validity* (relationship between theory and observation) are illustrated.

Internal validity – The following threats to internal validity have been individuated:

(i) The observation (pilot survey) that a significant number of StackOverflow users no longer reside in their birth country can affect the internal validity of the name-based gender resolution. Indeed, names associated with one gender in the birth country may be associated with a different gender in the residence country. Since StackOverflow users indicate the country of residence (if anything) as their location, this means that the gender-resolution heuristics will make a wrong choice: e.g., a male Andrea born in Italy but living in Germany will be identified as female.

(ii) Furthermore, information on gender has been constructed through name, avatar, and writing style as "masculine", "feminine" or something in between, and may be presented, consciously or not, in a fashion that is contrary to one's sex. This aspect is very relevant when considering the separation between male/female and masculine/feminine identities. We chose to treat names as dimensions of masculine and feminine gender performance.

External validity – The presented results are only valid for very specific communities: we suspect that other online communities act with similar gender barriers (e.g., gaming communities), while others are more gender- and minorities-friendly (e.g., those related to web technologies).

Construct validity – Below we report several points to be considered against the construct validity of this study:

(i) Limitations of the automatic approach: as illustrated in the description of the automatic gender resolution tool in subsection 5.2.2, such algorithm cannot be expected to map *all* usernames to genders, nor can it be expected to *always* map names to genders correctly.

Failure to map a username to gender can be related to inherent limitations of the resolution technique. For instance, we cannot resolve *(Norbertas, Lithuania)* since we do not have a name list for Lithuania, and none of the name lists we consider contains the name Norbertas. Furthermore, as explained above, we cannot resolve the gender for *(Andrea Demirović, Montenegro)* since we do not have data for Montenegro and different countries list Andrea both as a male and a female name. Moreover, some individuals might prefer not to disclose personal information and use standard SO-assigned usernames (e.g., *user1234*). Finally, location information, which is used to improve the accuracy of the technique, is not available for mailing list participants, hence could not be used when inferring gender for these users.

Gender resolution might produce incorrect answers, e.g., if (parts of) the usernames correspond to either male or female first names in some countries. To illustrate this point consider the following artificial example: *(lajos, Belgium)* is recognised as "masculine" since Lajos is known as a male name in Belgium. However, this username could have also been used by a woman called *La*ura *Jos*ten. Moreover, we have chosen to infer "masculine" ("feminine") if the name is known to be at least twice more frequently used for men (women) as for women (men). While predominant popularity of a name can be seen as a strong indication of the gender, it might still produce wrong results for an individual user: some women might have first names more commonly associated with men and some men might have first names more commonly associated with women: e.g., Coffey and McLaughlin [2009] report 102 women called "John" (vs. 38,730 men) and 25 men called "Carolyn" (vs. 8,615 women).

(ii) Limitations of the manual approach: as any manual process, avatar-based gender recognition could have been affected by the evaluator's fatigue or failure to recognise an image, e.g., a cartoon character, as being male or female. To address this threat, we have reviewed the genders inferred and corrected the few instances of misclassified genders.

(iii) Another threat to validity is formed by StackOverflow users purposefully using as avatars images of the opposite gender, e.g., male users with erotic stock photos of female models. While we have done our best to identify these cases and resolve them during the review, we cannot guarantee absolute correctness of the manual resolution approach.

(iv) Furthermore, we note potential human error when inferring gender from an avatar picture, or when deciding whether a certain profile in another

data source (e.g., Twitter) belongs to the same StackOverflow or mailing list participant.

(v) Another threat to the construct validity is formed by StackOverflow users purposely using as avatars images of the opposite gender, e.g., male users with erotic stock photos of female models. These cases have been identified and resolved during the manual review.

9. CONCLUSION AND FUTURE WORK

The issue of gender and STEM-related subjects has been studied for several years, and mostly from the point of view of "why" women do not engage with scientific studies or careers. Lesser attention has so far been given to quantify the phenomenon and representation of women in online communities (as technology-"users"), what are their levels of participation, and whether differences can be detected at the gender level. Only anecdotal evidence has been gathered on how specific communities actively encourage (or discourage) women from participating.

This study quantitatively investigated two different types of communities, the one gathered around the StackOverflow Q&A website, and the mailing lists of two web Content Management Systems (Drupal and WordPress). The number of male and female participants, and the type and length of participation were measured to compare the two communities. The main objective of the study was to add facts to current anecdotal evidence, which suggests that StackOverflow actively discourages the participation of women, and that web technologies tend to attract more female users.

Special tools were developed to infer gender based on name and nationality; however, since the gender inference was partly manual, the StackOverflow data was sampled. In the analysis and attribution of gender to participants of online communities, it was found that a large proportion of users (StackOverflow users in particular) are not identifiable. The large number of untrackable identities goes against the benefits that this community could produce: instead of uniquely identifying themselves, a large proportion of contributors prefer to remain anonymous, which could reflect an unfriendly community, or a way to abstract the solutions to problems from the people providing them.

It was found that the percentage of women engaged in the studied communities is greatly imbalanced, and men represent the vast majority of users and contributors. This finding is in line with the recent down-fall in number of graduates in STEM (and computing in particular) subjects. On the other hand, two types of communities emerged: in the first, the genders are not balanced, but women and men show broadly similar contribution patterns in terms of questions and answers, and length of engagement, when considering "traditional" mailing lists. Differences in participation were only detected in specific WordPress mailing lists, associated with "hacking" and more technical content.

In the second type of community, women are still a minority, but their levels of participation are significantly different from men's: when considering the specific case of StackOverflow, men engage for longer periods of time than women. However, relative to the duration of their engagement in the community, women are even more active than men when asking questions, while providing a comparable number of answers. This result suggests a relatively "unhealthy" community, in which women disengage sooner, although their activity levels are comparable to men's.

This second set of findings, together with the tendency to remain anonymous, show that the StackOverflow community in particular, beside producing excellent technical content, sets up higher barriers to entry to its participation, particularly to women. This further exacerbates the division in gender participation and engagement to online communities, which produce quality content at the expense of a generalised participation.

Future work should expand on the current notion of gender as a binary phenomenon (male/female), an approach that has been already criticised by some of the gender-technology students [Boivie, 2010, Van Lenning, 2004]. Indeed, the conflation of gender and heterosexuality has been observed to complicate social relations in male-dominated domains like computing [Stepulevage, 2001], and lesbian women may feel attracted to software development for the same reasons that heterosexual women may feel disinterested in this field [Landström, 2007]. Therefore, as a possible direction for future work we consider going beyond the gender binary and investigating how sexual orientation affects individual involvement in online software developer communities.

As stated in Section 7.3, a crucial direction for further investigation consists in verifying quantitatively whether the competitive nature of online communities creates higher entry barriers for women and is detrimental for their effectiveness (cf. laboratory experiments in [Gneezy et al., 2003, Niederle and Vesterlund, 2007]).

In addition, based on the observations from different online communities, we would like to design a next generation online platform for software developers, a kind of StackOverflow++. Similarly to the study of spreadsheet software [Burnett et al., 2011] the platform should take gender differences into account, without penalising either gender.

Ultimately, the data extrapolated in this study is purely quantitative. However, we acknowledge the value of qualitative research, and we plan to triangulate the

reported results by doing qualitative analysis. Specific female-only, online communities (LinuxChix, GrrrTalk, Ada Initiative, among others; cf. [Lin, 2006]), devoted to technical and computer science aspects, have been contacted already for interest in the research and potential future interviews and questionnaires.

ACKNOWLEDGMENTS

We are grateful to a number of people for helping with name-based gender resolution of contributors from their countries, as well as avatar-based gender resolution in general: Ali Afroozeh, Amrita Mishra, Benjamin Le Cadre, Florian Grolleau, Josselin Tillay-Doledec, Jyoti Kalra, Kévin Morin, Roberto Lie, Stephen Flenet, Tanir Ozcelebi, Ulyana Tikhonova, Yanja Dajsuren, Yann Cadalen, Yaping Luo. We are also grateful to the participants in the StackOverflow pilot survey, as well as to the StackOverflow community for comments and discussion around the pilot survey [Anonymous, 2012].

We are also grateful to gender students Elena Gapova, Tanya Krzywinska and Adi Kuntsman for pointing out relevant sociological studies.

The first author gratefully acknowledges the support of the Dutch Science Foundation, through the project "Multi-Language Systems: Analysis and Visualization of Evolution—Analysis" (612.001.020). Part of this research has been carried out during the third author's stay at the Université de Mons, Belgium, supported by the F.R.S-F.N.R.S. under the grant BSS-2012/V 6/5/015.

REFERENCES

Encyclopedia of gender and information technology, 2006. Idea Group Reference.

A.E. Adam.
Hacking into hacking: Gender and the hacker phenomenon.
ACM SIGCAS Computers and Society, 33(4):3, 2003.

Anonymous.
Gender, representativeness and reputation in Stack Overflow.
http://meta.stackoverflow.com/q/139901/182512, 2012.

Shlomo Argamon, Moshe Koppel, Jonathan Fine, and Anat Shimoni.
Gender, genre, and writing style in formal written texts.
Text, pages 321–346, 8 2003.

Janet Armentor-Cota.
Gender and chat rooms.
In *Encyclopedia of gender and information technology*, pages 361–364, 2006.

Janet Armentor-Cota.
Multiple perspectives on the influence of gender in online interactions.
Sociology Compass, 5(1):23–36, 2011.

B***.
email about stack overflow.
http://pastebin.com/14TeVv79, November 2012.

Victor R. Basili and David M. Weiss.
A methodology for collecting valid software engineering data.
Software Engineering, IEEE Transactions on, SE-10(6):728–738, 1984.

B. Bimber.
Measuring the gender gap on the internet.
Social Science Quarterly, 81(3):868–876, 2000.

C. Bird, A. Gourley, P.T. Devanbu, M. Gertz, and A. Swaminathan.
Mining email social networks.
In *MSR*, pages 137–143. ACM, 2006.

J.C. Blickenstaff.
Women and science careers: leaky pipeline or gender filter?
Gender and Education, 17(4):369–386, 2005.

Inger Boivie.
Women, men and programming : Knowledge, metaphors and masculinity.
In *Gender Issues in Learning and Working with Information Technology: Social Constructs and Cultural Contexts*, pages 1–24. IGI Global, 2010.

Danah M. Boyd and Nicole B. Ellison.
Social network sites: Definition, history, and scholarship.
J. Computer-Mediated Communication, 13(1):210–230, 2007.

Joel Brandt, Philip J. Guo, Joel Lewenstein, Mira Dontcheva, and Scott R. Klemmer.
Two studies of opportunistic programming: interleaving web foraging, learning, and writing code.
In *CHI*, pages 1589–1598. ACM, 2009.

Margaret M. Burnett, Laura Beckwith, Susan Wiedenbeck, Scott D. Fleming, Jill Cao, Thomas H. Park, Valentina Grigoreanu, and Kyle Rector.
Gender pluralism in problem-solving software.
Interacting with Computers, 23(5):450–460, 2011.
Feminism and HCI: New Perspectives.

Judith Butler.
Gender Trouble: Feminism and the Subversion of Identity.
Routledge Classics. Routledge, 1999.

P.R. Clance.
The impostor phenomenon: overcoming the fear that haunts your success.
Peachtree Publishers, 1985.

Bentley Coffey and Patrick A. McLaughlin.
Do masculine names help female lawyers become judges? evidence from south carolina.
American Law and Economics Review, 11(1):112–133, 2009.

Roberts Lynne D. and Parks Malcolm R.
The social geography of gender-switching in virtual environments on the internet.
Information, Communication and Society, 2(4):521–540, 1999.

Whitney Darrow.
I'm Glad I'm a Boy! I'm Glad I'm a Girl!

Windmill Books, 1970.

Paul A. David and Joseph S. Shapiro.
Community-based production of open-source software: What do we know about the developers who participate?
Information Economics and Policy, 20(4):364–398, 2008.

Drupal.
Drupal mailing lists.
http://drupal.org/mailing-lists/, 2012.

O. J. Dunn.
Multiple comparisons among means.
J AM STAT ASSOC, 56:52–64, 1961.

Allan Fisher, Jane Margolis, and F. Miller.
Undergraduate women in computer science: experience, motivation and culture.
ACM SIGCSE Bulletin, 29(1):106–110, 1997.

K. R. Gabriel.
Simultaneous test procedures—some theory of multiple comparisons.
ANN MATH STAT, 40(1):224–250, 1969.

Tamar Ginossar.
Online participation: a content analysis of differences in utilization of two online cancer communities by men and women, patients and family members.
Health Communication, 23(1):1–12, 2008.

Uri Gneezy, Muriel Niederle, and Aldo Rustichini.
Performance in competitive environments: Gender differences.
The Quarterly Journal of Economics, 118(3):1049–1074, 2003.

A. Gras-Velazquez, A. Joyce, and M. Debry.
Women and ict.
Why are girls still not attracted to ICT studies and careers?, 2009.

Sherri Grasmuck, Jason Martin, and Shanyang Zhao.
Ethno-racial identity displays on facebook.
Journal of Computer-Mediated Communication, 15(1):158–188, 2009.

K. Grint and R. Gill.
The Gender-Technology Relation: Contemporary Theory and Research: An Introduction.
Gender & society: feminist perspectives on the past & present series. Taylor & Francis Group, 1995.

Sandra G. Harding.
The Science Question in Feminism.
Cornell University Press, 1986.

A. Herdağdelen and M. Baroni.
Stereotypical gender actions can be extracted from web text.
J. Am. Soc. Inf. Sci. Technol., 62(9):1741–1749, September 2011.

Susan C. Herring.
Gender and democracy in computer-mediated communication.
In *Computerization and Controversy: Value Conflicts and Social Choices*, pages 476–489. Academic Press, San Diego, CA, USA, 2nd edition, 1996.

Susan C. Herring and John C. Paolillo.
Gender and genre variation in weblogs.
Journal of SocioLinguistics, 10(4):439–459, 2006.

C. Hill, C. Corbett, and A. St Rose.
Why So Few? Women in Science, Technology, Engineering, and Mathematics.
ERIC, 2010.

L. Hodgkinson.
Is technology masculine? Theorising the absence of women.
In *International Symposium on Technology and Society.*, pages 121–126. IEEE, 2000.

M. Holander and D. A. Wolfe.
Nonparametric statistical methods.
Wiley, 1973.

European Union Initiative.
Science: It's a girl thing!
http://science-girl-thing.eu/, 2012.
Accessed: 30/07/2012.

Alison Kelly.
The construction of masculine science.
In *Science for Girls*, pages 66–77. Open University Press, 1987a.

Alison Kelly.
Why girls don't do science.
In *Science for Girls*, pages 12–17. Open University Press, 1987b.

Frank Konietschke, Ludwig A. Hothorn, and Edgar Brunner.
Rank-based multiple test procedures and simultaneous confidence intervals.
Electronic Journal of Statistics, 6:738–759, 2012.
ISSN 1935-7524.

Moshe Koppel, Shlomo Argamon, and Anat Shimoni.
Automatically categorizing written texts by author gender.
Literary and Linguistic Computing, 17(4):401–412, 2002.

Nalini P. Kotamraju.
Art versus code: The gendered evolution of web design skills.
In *Society Online: The Internet in Context*, pages 189–200. Sage Publications, 2003.

Victor Kuechler, Claire Gilbertson, and Carlos Jensen.
Gender differences in early free and open source software joining process.
In *Open Source Systems: Long-Term Sustainability*, volume 378 of *IFIP Advances in Information and Communication Technology*, pages 78–93. Springer, 2012.

Vivian A. Lagesen.
A cyberfeminist utopia?: Perceptions of gender and computer science among malaysian women computer science students and faculty.
Science Technology Human Values, 33(1):5–27, January 2008.

Catharina Landström.
Queering feminist technology studies.
Feminist Theory, 8(1):7–26, 2007.

Nina Lerman, Ruth Oldenziel, and Arwen P. Mohun.
Gender and Technology: A Reader.
Gender and Technology. Johns Hopkins University Press, 2003.

Ann Light.
HCI as heterodoxy: Technologies of identity and the queering of interaction with computers.
Interacting with Computers, 23(5):430–438, 2011.

Yuwei Lin.
A techno-feminist view on the open source software development.
In *Encyclopedia of gender and information technology*, pages 1148–1153, 2006.

Maria Lohan and Wendy Faulkner.
Masculinities and technologies: Some introductory remarks.
Men and Masculinities, 6(4):319–329, 2004.

Xiaoguang Lu, Hong Chen, and Anil Jain.
Multimodal facial gender and ethnicity identification.
In *Advances in Biometrics*, volume 3832 of *Lecture Notes in Computer Science*, pages 554–561. Springer, 2005.

Jay Lyman.
Getting in touch with the feminine side of open source.
Linux Today, August 2005.

Lena Mamykina, Bella Manoim, Manas Mittal, George Hripcsak, and Björn Hartmann.
Design lessons from the fastest Q&A site in the west.
In *CHI*, pages 2857–2866. ACM, 2011.

Jane Margolis and Allan Fisher.
Unlocking the clubhouse: Women in computing.
The MIT Press, 2003.

Jonathan Marshall.
Online life and gender vagueness and impersonation.
In *Encyclopedia of gender and information technology*, pages 932–937, 2006.

Carolyn Merchant.
The Death of Nature: Women, Ecology, and the Scientific Revolution.
ISSR library. HarperCollins, 1990.

I. Miliszewska, G. Barker, F. Henderson, and E. Sztendur.
The issue of gender equity in computer science- what students say.
Journal of Information Technology Education, 5(1):107–120, 2006.

D. Nafus, J. Leach, and B. Krieger.
FLOSSPOLS Deliverable D 16 Gender: Integrated Report of Findings.
http://www.flosspols.org/deliverables/D16HTML/FLOSSPOLS-D16-Gender_Integrated_Report_of_Findings.htm, 2006.

National Science Foundation.
Women, Minorities, and Persons with Disabilities in Science and Engineering.
http://www.nsf.gov/statistics/wmpd/pdf/nsf04317.pdf, 2004.
NSF 04-317.

Robert G. Newcombe.
Two-sided confidence intervals for the single proportion: comparison of seven methods.
Statistics in Medicine, 17(8):857–872, 1998.

Muriel Niederle and Lise Vesterlund.
Do women shy away from competition? do men compete too much?
The Quarterly Journal of Economics, 122(3):1067–1101, 2007.

Malcolm R. Parks.
Making friends in cyberspace.
J. Computer-Mediated Communication, 1(4), 1996.

Chris Parnin, Christoph Treude, Lars Grammel, and Margaret-Anne Storey.
Crowd documentation: Exploring the coverage and the dynamics of API discussions on Stack Overflow.
Technical report, Georgia Institute of Technology, 2012.

E.S. Raymond.
The cathedral and the bazaar: Musings on linux and open source by accidental revolutionary revised edition.
Cambridge, UK, O'Reily, 1999.

E.S. Roberts, M. Kassianidou, and L. Irani.
Encouraging women in computer science.
ACM SIGCSE Bulletin, 34(2):84–88, 2002.

Gregorio Robles, Jesus M. Gonzalez-Barahona, Daniel Izquierdo-Cortazar, and Israel Herraiz.
Tools for the study of the usual data sources found in libre software projects.
International Journal of Open Source Software and Processes, 1(1):24–45, Jan-March 2009.

Nigel Ross.
Writing in the information age.
English Today, 22:39–45, 6 2006.

Diana Schimke, Heidrun Stoeger, and Albert Ziegler.
The relationship between social presence and group identification within online communities and its impact on the success of online communities.
In *Proceedings of the 2nd international conference on Online communities and social computing*, pages 160–168. Springer, 2007.

Lindsay H. Shaw and Larry M. Gant.
Users divided? exploring the gender gap in internet use.
CyberPsychology & Behavior, 5(6):517–527, 2002.

David J. Sheskin.
Handbook of Parametric and Nonparametric Statistical Procedures.
Chapman & Hall, 4 edition, 2007.

Megan Squire.
How the FLOSS research community uses email archives.
International Journal of Open Source Software and Processes, 4(1):37–59, 2012.

Linda Stepulevage.
Gender/Technology relations: complicating the gender binary.
Gender and Education, 13(3):325–338, 2001.

Margaret-Anne D. Storey, Christoph Treude, Arie van Deursen, and Li-Te Cheng.
The impact of social media on software engineering practices and tools.

In *FoSER*, pages 359–364. ACM, 2010.

John Suler.
Do boys (and girls) just wanna have fun? gender switching in cyberspace.
In *Gender communication*, pages 149–152. Kendall/Hunt, 2004.

Nicole Sullivan.
Don't feed the trolls.
http://www.youtube.com/watch?v=ulNSlES1Fds, 2012.

P.A. Taylor.
Hackers: crime in the digital sublime.
Psychology Press, 1999.

E.M. Trauth, J.L. Quesenberry, and A.J. Morgan.
Understanding the under representation of women in it: Toward a theory of individual differences.
In *Proceedings of the 2004 SIGMIS conference on Computer personnel research: Careers, culture, and ethics in a networked environment*, pages 114–119. ACM, 2004.

Christoph Treude, Ohad Barzilay, and Margaret-Anne D. Storey.
How do programmers ask and answer questions on the web?
In *ICSE*, pages 804–807. ACM, 2011.

John W. Tukey.
Quick and dirty methods in statistics, part II, Simple analysis for standard designs.
In *American Society for Quality Control*, pages 189–197, 1951.

Sherry Turkle.
The Second Self: Computers and the Human Spirit.
The MIT Press, 2005.

Sherry Turkle.
Life on the Screen.
Simon & Schuster, 2011.

VA.
What can Stack Overflow do to persuade female programmers to participate more?
http://meta.stackoverflow.com/q/30411/185480, 2011a.

VA.
Never see any women answering questions?
http://meta.stackoverflow.com/q/34070/185480, 2011b.

Nancy A. Van House.
Feminist HCI meets facebook: Performativity and social networking sites.
Interacting with Computers, 23(5):422–429, 2011.

Alkeline Van Lenning.
The body as crowbar: Transcending or stretching sex?
Feminist Theory, 5(1):25–47, 2004.

Liesbet Van Zoonen.
Feminist theory and information technology.
Media, Culture and Society, 14(1):12–35, 1992.

Bogdan Vasilescu.
Academic papers using Stack Overflow data.
http://meta.stackoverflow.com/q/134495/185480, 2013.

Bogdan Vasilescu, Andrea Capiluppi, and Alexander Serebrenik.
Gender, representation and online participation: A quantitative study of StackOverflow.
2012 International Conference on Social Informatics, pages –, 2012.

Judy Wajcman.
Feminist theories of technology.
Cambridge Journal of Economics, 34(1):143–152, January 2010.

Youcheng Wang and Daniel R. Fesenmaier.
Modeling participation in an online travel community.
Journal of Travel Research, 42(3):261–270, 2004.

Michele A. Whitecraft and Wendy M. Williams.
Why aren't more women in computer science?
In *Making Software: What Really Works, and Why We Believe It*, pages 221–238. O'Reilly Media, Inc., 2010.

F. Wilcoxon.
Individual comparisons by ranking methods.
Biometrics Bulletin, 1(6):80–83, 1945.

Edwin B. Wilson.
Probable inference, the law of succession, and statistical inference.
Journal of the American Statistical Association, 22(158):209–212, 1927.

Richard T. A. Wood and Mark D. Griffiths.
Why swedish people play online poker and factors that can increase or decrease trust in poker web sites: A qualitative investigation.
Journal of Gambling Issues, 21:80–97, 2008.

WordPress.
Wordpress mailing lists.
http://codex.wordpress.org/Mailing_Lists, 2012.