

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Ciência da Computação

Diferenças Comportamentais Entre Gêneros em Comunidades de Perguntas e Respostas

Milena Sales Araujo

Dissertação submetida à Coordenação do Curso de Pós-Graduação em
Ciência da Computação da Universidade Federal de Campina Grande -
Campus I como parte dos requisitos necessários para obtenção do grau
de Mestre em Ciência da Computação.

Área de Concentração: Ciência da Computação

Linha de Pesquisa: Sistema de Computação

Nazareno Ferreira de Andrade

(Orientador)

Campina Grande, Paraíba, Brasil

©Milena Sales Araujo, 15/07/2015

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

A663d Araujo, Milena Sales .
Diferenças comportamentais entre gêneros em comunidades de perguntas e respostas / Milena Sales Araujo. – Campina Grande, 2015.
83 f. : color.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2015.

"Orientação: Prof. Nazareno Ferreira de Andrade".
Referências.

1. Estudo de Gênero. 2. Sistemas Colaborativos. I. Andrade, Nazareno Ferreira de. II. Título.

CDU 004.773.7(043)

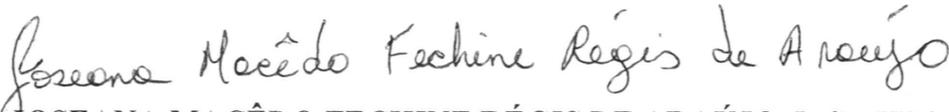
**"DIFERENÇAS COMPORTAMENTAIS ENTRE GÊNEROS EM COMUNIDADES DE
PERGUNTAS E RESPOSTAS"**

MILENA SALES ARAÚJO

DISSERTAÇÃO APROVADA EM 15/07/2015


NAZARENO FERREIRA DE ANDRADE, D.Sc, UFCG
Orientador(a)


TAYANA UCHÔA CONTE, Dr.^a, UFAM
Examinador(a)


JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO, D.Sc, UFCG
Examinador(a)

CAMPINA GRANDE - PB

Resumo

Sites de perguntas e respostas são conhecidos por serem uma fonte de conhecimento simples e confiável. Os sites do *StackExchange* se destacam por estarem dentre as maiores comunidades de perguntas e respostas online, principalmente se considerarmos aquelas relacionadas à *STEM*. Assim como em outras comunidades online, mulheres são uma minoria nos sites do *Stackexchange*, fazendo com que estes sites não se beneficiem de pontos de vista diversificados na criação de seu conteúdo. Esta pesquisa investiga se mulheres que se dedicam aos sites do *StackExchange* contribuem menos e/ou deixam a comunidade precocemente. Para isso, examinamos diferenças entre os gêneros quanto ao número de contribuições feitas aos sites, o tempo que passaram contribuindo e quanto a diferença de qualidade das suas contribuições segundo a comunidade. Nossos resultados mostram que, na maioria dos sites, mulheres contribuem tanto quanto homens, sendo suas contribuições com qualidade e por intervalos de tempo similares aos homens. Quando percebe-se diferença no número de contribuições, são as mulheres que tendem a contribuir mais. Além disso, encontramos que a proporção de contribuições provenientes de mulheres nos sites relacionados a *STEM* do *StackExchange* está aumentando na maioria das comunidades onde encontramos diferença estatisticamente significativa. Estes resultados ajudam a melhorar a visão que temos das contribuições feitas por mulheres neste tipo de site, gerando repercussão para pesquisas futuras, assim como discussão de design para os administradores.

Abstract

Social Question and answer (Q&A) sites are known as simple and reliable knowledge databases. Sites from the StackExchange platform presently stand out as some of the largest and most social Q&A spaces, particularly the sites related to STEM. As in some other collaborative online communities, women are underrepresented in StackExchange platform, hindering its sites from contemplating more diverse views in their content creation process. This work investigates whether women who are engaged with sites from StackExchange are influenced to contribute less and to leave the community. For that, we examined gender differences in the number of contributions made to the sites, the time spent contributing, and on the evaluation of quality that the community provides for posted content. Our results point that, in the greater part of sites, women contributors contribute as much as men, with similar quality and for a similar period. When difference do occur, it happens most often that women tend to contribute more than men. Nevertheless, we found that the proportion of contributions coming from women in the STEM sites of StackExchange is increasing in several sites. These results point to a richer picture of female contribution in those sites, and raise several implications for future research, as well as to site administrators.

Agradecimentos

Gostaria de agradecer aos meus amigos e familiares pelo apoio emocional, à CAPES pelo apoio financeiro. Ao meu orientador pela paciência e dedicação. Ao grupo de estudos de sistemas colaborativos, pela amizade e valiosas discussões.

Por fim, agradeço à comunidade StackExchange por ser incrível e pelos dados fornecidos para esta pesquisa. Em especial, agradeço à comunidade *Cross Validated* e ao usuário *Glen_b* que me deram suporte técnico e estatístico para variadas questões que surgiram durante esta pesquisa.

Conteúdo

1	Introdução	1
1.1	Motivação	1
1.2	Objetivos	3
1.3	Abordagem de pesquisa	3
1.4	Resultados	4
1.5	Estrutura do texto	5
2	Contexto	7
2.1	Comunidades de Perguntas e Respostas	7
2.2	<i>StackExchange</i>	8
2.2.1	História	8
2.2.2	Modo de funcionamento	10
3	Literatura	12
3.1	Diferenças de gênero em comunidades online	12
3.2	<i>Feedback</i> e competitividade	14
3.3	Mulheres, <i>STEM</i> e autoconfiança	15
3.4	Nossa contribuição	16
4	Metodologia	18
4.1	Dados utilizados	18
4.2	Deduzindo o gênero dos usuários	20
4.3	Manuseio dos dados	21
4.3.1	Classificando os usuários quanto ao gênero	22
4.3.2	Cálculos estatísticos	22

4.4	Medindo contribuições e engajamento	23
4.4.1	Número de contribuições	23
4.4.2	Qualidade	23
4.4.3	Engajamento	24
4.4.4	Contribuições e novos registros ao longo do tempo	25
4.5	Distribuições e testes	26
5	Resultados	27
5.1	Número de contribuições	27
5.2	Qualidade	29
5.3	Dedicação	31
5.4	Contribuições e novos registros ao longo do tempo	31
5.5	Focando nos maiores sites	32
6	Discussão	36
6.1	As semelhanças	36
6.2	As diferenças	37
7	Conclusões	41
7.1	Limitações	42
A	Sumário dos usuários identificados	50
B	Cinco maiores comunidades	54
C	Estatísticas da primeira pergunta de pesquisa	60
D	Estatísticas da segunda pergunta de pesquisa	67
E	Estatísticas da terceira pergunta de pesquisa	73
F	Estatísticas da quarta pergunta de pesquisa	77

Lista de Símbolos

STEM - *Science, Technology, Engineering and Mathematics*

IRC - *Internet Relay Chat*

Lista de Figuras

2.1	Exemplo de uma página de pergunta no <i>StackOverflow</i>	9
2.2	Perfil de um usuário cadastrado	10
2.3	Perfil de um usuário não cadastrado	11
5.1	Comparação do número de perguntas	28
5.2	Comparação do número de respostas em comunidades onde mulheres respondem mais.	30
5.3	Comparação do número de respostas em comunidades onde homens respondem mais.	32
5.4	Comparação do número de contribuições na 5 maiores comunidades.	35
6.1	Distribuição das variáveis para o site <i>User Experience</i>	39
B.1	Distribuição das variáveis para o site <i>StackOverflow</i>	55
B.2	Distribuição das variáveis para o site <i>SuperUser</i>	56
B.3	Distribuição das variáveis para o site <i>ServerFault</i>	57
B.4	Distribuição das variáveis para o site <i>Mathematics</i>	58
B.5	Distribuição das variáveis para o site <i>Programmers</i>	59

Lista de Tabelas

2.1	Resumo das categorias do <i>StackExchange</i>	8
4.1	Lista de comunidades estudadas	19
5.1	Diferença de tempo de vida entre homens e mulheres	33
A.1	Sumário dos usuários identificados	53
C.1	Resultados estatísticos para o número de contribuições	66
D.1	Resultados estatísticos para a qualidade das contribuições	72
E.1	Resultados estatísticos para o tempo de dedicação à comunidade	76
F.1	Resultados estatísticos para as contribuições ao longo do tempo	80
F.2	Resultados estatísticos para os registros ao longo do tempo	83

Capítulo 1

Introdução

Desde o seu surgimento na década de 1990 [2], fontes de informação colaborativa online têm adquirido a sua importância no meio web por seu conteúdo confiável e de fácil acesso. Tais fontes nada mais são do que sites ou aplicativos que podem ser acessados via internet e todo o seu conteúdo provém dos seus usuários. Podemos destacar, dentre tantas outras, a *Wikipedia*, *Yahoo Answers* e os sites da plataforma *StackExchange*. Estes dois últimos pertencem a uma categoria específica de fonte de informação colaborativa – são comunidades de perguntas e respostas online. Estas comunidades são fontes de informação colaborativa online onde a toda a dinâmica de construção da informação se dá a partir de perguntas e respostas. A plataforma *StackExchange*, foco do nosso estudo, apesar das suas mais de 100 comunidades de perguntas e respostas, destaca-se no campo de *STEM* (*Science, Technology, Engineering and Mathematics*), com comunidades com mais de 4 milhões de usuários [1].

Sabendo que o conteúdo destas comunidades de perguntas e respostas é criado por voluntários, estes sites são extremamente dependentes da adoção e contribuição de seus usuários. Assim como a variedade no conteúdo, a diversidade de visões de um certo assunto no conteúdo da comunidade também depende da diversidade dos usuários voluntários.

1.1 Motivação

A motivação inicial para este estudo foi um receio por parte dos pesquisadores quanto à diversidade de gêneros ou *gender gap* em comunidades online, principalmente àquelas relacionadas a *STEM*. Já era de se esperar que comunidades voltadas à *STEM* não apresentassem

diversidade de gêneros, levando em conta os estereótipos, mas relatos recentes apontam que não são apenas estas comunidades que apresentam um déficit de mulheres.

Com a divulgação dos relatórios de diversidade por parte de grandes companhias como *Google* [5] e *Linkedin* [6] mostrando que a proporção de mulheres em cargos de liderança é baixo, associamos este resultado às comunidades *Open Source* que apresentam números parecidos [38]. É possível que estes dados estejam relacionados, visto que tanto as companhias quanto as comunidades *Open Source* focam no desenvolvimento de software.

Quanto às comunidades de perguntas e respostas, como o *StackExchange* onde resultados recentes mostram que mulheres são uma minoria extrema dentre os contribuidores do *StackOverflow* [48], a maior comunidade dentro da plataforma *StackExchange*, chegando a representar apenas 7% dos usuários que contribuem para o site. Ademais, nota-se que pessoas que questionam a razão do *gender gap* no *StackOverflow* podem ser mal entendidas e até distratadas¹, tendo os outros membros as acusando de preconceito. Isto é preocupante, pois existe um problema de balanço de diversidade, não se sabe a razão. A própria comunidade não parece estar disposta a examinar causas e consequências.

Por fim, notamos que, em quatro das maiores comunidades do *StackExchange* relacionadas à *STEM* (*StackOverflow*, *SuperUser*, *Mathematics* e *Programmers*) não foi possível identificar nenhuma mulher no ranking dos 100 maiores contribuidores destas comunidades, o que exaltou mais a nossa suspeita da falta de participação, por parte das mulheres, nas comunidades que pretendíamos estudar.

Tais fatos coincidem com estereótipos comuns de mulheres em campos de *STEM* [42] assim como a baixa influência que mulheres possuem em grupos nos quais há homens e mulheres [27]. Contudo, se mulheres realmente são influenciadas a contribuir menos que homens nestas comunidades, os sites do *StackExchange* podem estar perdendo não só contribuidores de qualidade, mas também a oportunidade de produzir conteúdo que representa as visões de ambos os gêneros. Além disso estes sites apresentam melhor funcionamento, sendo uma comunidade com grupos de gêneros distintos [32].

¹<http://is.gd/eAnI8R>

1.2 Objetivos

Apesar de estudos anteriores apontarem que a fração de contribuidoras é bem menor que a de contribuidores, não está claro se mulheres tendem a, individualmente, contribuir menos, por menos tempo ou têm suas contribuições vistas como de menor qualidade pela comunidade. Estas observações podem nos orientar se estas comunidades são menos acolhedoras para contribuidoras e em que aspecto. Outrossim, se mulheres não contribuem menos, por menos tempo nem tem seu conteúdo avaliado como de baixa qualidade, podemos deduzir que uma campanha de ampliação da diversidade nestas comunidades deve ter seu foco diferente de apenas convocar mulheres.

Verificamos que, na literatura, não há um estudo que observe os usuários das comunidades do *StackExchange* com relação às características citadas acima e ainda faça um comparativo com estudos de gênero, não só no contexto de comunidades online, mas também em ambientes offline, relacionando as diferenças e semelhanças encontrados nos demais estudos com relação à interação e à competitividade das pessoas de diferentes gêneros e aos comportamentos observados nos sites estudados.

Portanto, a pesquisa aqui descrita tem como objetivo estudar as consequências no comportamento de um usuário que identifica seu gênero em comunidades do *StackExchange*. Para atingir o nosso objetivo, estudamos usuários que participaram da comunidade, dos quais pudemos identificar seu gênero facilmente. Destes, os seguintes aspectos foram analisados: participação, qualidade das contribuições, tempo ativo na comunidade e contribuições ao longo do tempo. Mais ainda, estudamos como se comporta a proporção de novos registros por cada gênero, em cada comunidade, ao longo do tempo.

1.3 Abordagem de pesquisa

No contexto dos sites da plataforma *StackExchange*, quisemos entender melhor como se dá a diferença de comportamento entre homens e mulheres que participam da plataforma. Para isto, a primeira medida que tomamos foi a de identificar automaticamente o gênero dos usuários que contribuíram de alguma forma para qualquer um dos sites. Utilizamos um classificador para que obtivéssemos o maior número de usuários identificados de uma

maneira rápida e confiável, na medida do possível para uma comunidade onde a maioria dos usuários é anônima.

Após esta primeira etapa, procuramos entender a diferença de comportamento respondendo a seguinte pergunta de pesquisa: *Como se dá (caso existente) a diferença de comportamento entre homens e mulheres nos sites do StackExchange?* E, para respondê-la, a dividimos em quatro subperguntas de pesquisa para estudar aspectos distintos do comportamento dos usuários.

A primeira questão que nos vem a mente quando comparamos comportamento de contribuições é quanto ao número de contribuições realizada por cada usuário. A primeira subpergunta é *O número de contribuições feitas por homens e por mulheres difere significativamente nas comunidades observadas?* Foi possível analisar e comparar a quantidade de cada tipo de contribuição (perguntas, respostas e comentários) realizadas por cada usuário de cada gênero.

Um aspecto importante a ser estudado é como os usuários percebem a qualidade uns dos outros, para isto temos a segunda subpergunta: *As contribuições feitas por diferentes gêneros são vistas, pela comunidade, com níveis de qualidade diferentes?* E, para estudar o engajamento dos usuários, analisamos o tempo de vida na comunidade e frequência de postagem, respondendo à terceira subpergunta: *O engajamento de homens e mulheres difere nas comunidades estudadas?*

Por fim, para estudar como as contribuições e novos registros realizados por usuários de cada gênero têm se dado ao longo do tempo, fizemos a quarta e última pergunta de pesquisa: *A proporção de contribuições e novos registros feitos por usuários de cada gênero tem aumentado ou reduzido ao longo do tempo?*

1.4 Resultados

Nosso resultados mostram que, na maioria das comunidades estudadas, mulheres que expõem seu gênero no site, não só contribuem tanto quanto os homens, mas também têm suas contribuições vistas com o mesmo nível de qualidade e se dedicam ao site pelo mesmo período de tempo que os homens. Nas comunidades onde podemos verificar uma diferença significativa entre as contribuições advindas dos dois gêneros, os padrões detectados vão de en-

contro aos estereótipos: na maioria das vezes, mulheres tendem a contribuir mais do que homens. Isto acontece até nas comunidades relacionadas à *STEM*, inclusive na *StackOverflow*, apesar de que, em nenhuma destas comunidades mulheres contribuem com mais respostas do que homens. Também conseguimos notar um crescimento na proporção de novos registros e contribuições provenientes de mulheres em várias comunidades relacionadas a *STEM*.

Estes resultados questionam o senso comum de que mulheres são menos presentes em sites como os estudados porque suas contribuições podem não ser tão bem vindas quanto as dos homens. Em vez disso, os vários indicadores de contribuição que utilizamos mostram que mulheres que manifestam seu gênero, em geral, contribuem de maneira similar aos homens com a mesma característica. Estes resultados mostram comunidades de perguntas e respostas, e, em particular, aquelas com assuntos relacionados à *STEM*, não são comunidades exclusivas para homens e necessitam de motivos alternativos para explicar o *gender gap* nestes sites. A razão pode estar relacionada à falta de auto confiança no campo por parte das mulheres, como pode ser observado em outros sistemas [7]. Mais ainda, saber que mulheres que se engajam nestes sites são tão ativas quanto o que observamos sugere que o foco na campanha a favor da diversidade de gênero nestas comunidades deve ser voltado ao recrutamento e ao engajamento inicial.

1.5 Estrutura do texto

De forma breve, descrevemos aqui a estrutura do restante deste documento. No Capítulo 2 apresentamos comunidades de perguntas e respostas, o contexto da pesquisa, apresentando sua definição e objetivos, focando principalmente no nosso objeto de pesquisa que é a plataforma *StackExchange*. No Capítulo 3 fazemos uma visão geral do que já foi estudado sobre gênero, comunidades online e a interseção entre estes estudos.

Uma descrição dos dados e metodologia utilizados no estudo podem ser encontrados no Capítulo 4. No Capítulo 5 apresentamos, de forma detalhada, os resultados obtidos utilizando a metodologia do Capítulo anterior e discutimos os resultados e suas implicações no Capítulo 6.

Por fim, no Capítulo 7.1 apresentamos as limitações encontradas durante esta pesquisa e fechamos a descrição do estudo com o Capítulo 7, que descreve as conclusões que chegamos

e aponta caminhos para estudos futuros. Os Apêndices A à F resumizam e descrevem o objeto de estudo e as variáveis estudadas.

Capítulo 2

Contexto

Neste capítulo descrevemos qual a finalidade e como funcionam comunidades de perguntas e respostas. Ademais, focando na plataforma *StackExchange*, descrevemos sua história, estrutura, além de como são dispostos os perfis dos usuários.

2.1 Comunidades de Perguntas e Respostas

Dentre os vários sites e comunidades que se propuseram a formar um banco de informações online e disponível à toda internet, comunidades de perguntas e respostas têm se destacado por serem um meio social e colaborativo para formar tal depósito de conhecimento.

Sites como *Quora*, *Yahoo! Answers* e as comunidades do *StackExchange* são exemplos de sucesso de tais comunidades. E, apesar do design e das abordagens diferentes, todas têm o mecanismo de funcionamento similar: usuários postam perguntas que são expostas publicamente e divididas por tópicos ou *tags*. Outros usuários, então, podem responder tais perguntas e suas respostas recebem votos pelo questionador e demais membros da comunidade, como um controle de qualidade.

Tais sites atraem milhões de usuários passivos (conhecidos como *lurkers*) e ativos que, sem recompensa financeira alguma, são responsáveis por todo o conteúdo gerado por estes sites e que também é disponibilizado à comunidade sem custo algum. Tal conteúdo é considerado como confiável e com um nível de qualidade não esperado para um site cujo conteúdo é produzido por qualquer pessoa que se voluntarie, fazendo com que essas comunidades e seu conteúdo sejam consideradas um forte atrativo para pesquisadores.

2.2 StackExchange

Com mais de 100 sites sobre assuntos diversos, o *StackExchange* não se trata apenas de uma comunidade de perguntas e respostas, mas de uma plataforma. Apesar de se destacar nas comunidades relacionadas à *STEM*, o *StackExchange* possui sites nos mais diversos assuntos, desde viagens, passando por culinária e chegando até a comunidades dedicadas a línguas e religiões como alemão e islamismo. A Tabela 2.1 mostra um resumo com as categorias de comunidades presentes na plataforma *StackExchange*, a quantidade de sites em cada categoria e a idade média dos sites em cada categoria.

<i>Categoria</i>	<i>Comunidades</i>	<i>Idade (meses)</i>
business	3	43
culture-recreation	21	39
life-arts	15	44
professional	2	25
science	12	43
technology	32	47

Tabela 2.1: Esta tabela apresenta o número de comunidades por categoria e a média das idades das comunidades de cada categoria, em meses.

2.2.1 História

Tudo começou com o site *StackOverflow*, que foi criado em 2008 por dois engenheiros de software chamados Joel Spolsky e Jeff Atwood, com o intuito de ser uma alternativa a fóruns voltados para programadores, que costumavam ser populares na época. Com seu design diferenciado, o *StackOverflow* foi cativando a comunidade, atraindo centenas de usuários e, no ano seguinte, mais duas comunidades foram criadas pelos mesmos criadores do *StackOverflow*, para suprir a demanda por sites similares focando em assuntos não relacionados à programação de computadores: *ServerFault* e *SuperUser*.

A plataforma StackExchange foi lançada publicamente no início de 2011, já contando

com 33 sites sobre diversos temas. Hoje, a plataforma conta com 144 comunidades, divididas em seis categorias: Tecnologia, Cultura, Artes, Ciência, Negócios e Profissional. Com mais de cinco milhões de usuários cadastrados e centenas de milhares de visitas por dia, o *StackExchange* cresceu muito desde 2008, mas sua comunidade mais popular continua sendo a *StackOverflow* [3; 4].

Dates difference in days as a fraction

▲ 1 ▼ ★

I have two dates (datetime objects) which I want to subtract from each other. The result I am expecting is the number of days between the two. This is how I am doing it:

```
def get_days_between(datePast, dateFuture):
    difference = dateFuture - datePast
    return difference.days
```

The problem is, I need the number of days as a fraction. Example: If the past and future dates are only 12 hours apart, I am expecting the result to be 0.5 day.

How can I achieve that ?

python date python-2.7 datetime

share improve this question

asked Aug 19 '14 at 14:57

 Milena Araujo
129 ● 2 ● 17

2 Divide the number of hours by 24? $12 / 24 = 0.5$ – iwin Aug 19 '14 at 14:59

@iwin The timedelta object has no "hours" attribute. It was my first shot :-/ – Milena Araujo Aug 19 '14 at 16:31

add a comment

1 Answer

active oldest votes

▲ 6 ▼ ✓

```
from datetime import timedelta

def get_days_between(datePast, dateFuture):
    difference = dateFuture - datePast
    return difference.total_seconds() / timedelta(days=1).total_seconds()
```

share improve this answer

answered Aug 19 '14 at 15:01

 wim
48.3k ● 19 ● 89 ● 171

Figura 2.1: Exemplo de uma página de pergunta no *StackOverflow*.

2.2.2 Modo de funcionamento

Como dito anteriormente, os sites do StackExchange funcionam de maneira idêntica. Um usuário pode contribuir com perguntas e os outros usuários podem respondê-la. Tanto perguntas quanto resposta podem receber comentários e votos. Na Figura 2.1 é apresentado um exemplo dos elementos supracitados.

É importante notar que, durante a maior parte da produção de informação nos sites do *StackExchange*, os usuários só têm acesso ao nome, foto e reputação dos usuários que estão contribuindo concomitantemente. Para ter acesso aos demais dados de um colega, caso estejam disponíveis, é necessário que o usuário clique no link para o perfil dele ou dela. Na Figura 2.2 é mostrado um perfil com todos os campos de informação pessoal propostos pela plataforma preenchidos.

Os votos são o meio que o *StackExchange* usa para permitir que a própria comunidade mantenha a qualidade das contribuições. Para isso, eles são restritos aos usuários registrados. Os que possuem uma reputação de 15 pontos ou mais podem votar afirmando que uma contribuição é boa. Apenas aqueles usuários com 125 pontos ou mais podem votar afirmando que uma contribuição é ruim. Restringindo assim aos usuários que já participaram do site ou da plataforma o controle de qualidade do conteúdo do site.

Milena Araujo

:D

129 REPUTATION

2 answers, 8 questions, ~10k people reached

Brazil
mih
mil3na
milena.com
Member for 4 years, 2 months
523 profile views
Last seen 1 min ago

Stack Overflow Careers profile

Communities (11)

Stack Overflow	129
Cross Validated	91
Area 51	51
Server Fault	1
Home Improvement	1

Top Tags (24)

python	SCORE 3	POSTS 8	POSTS % 80
django	SCORE 3	POSTS 4	rest
tastypie	SCORE 3	POSTS 2	primefaces
primefaces	SCORE 0	POSTS 2	lazy-loading
lazy-loading	SCORE 0	POSTS 2	

Figura 2.2: Perfil de um usuário cadastrado.

Uma particularidade da plataforma StackExchange é que, em seus sites, o anonimato do usuário é respeitado e preservado. Primeiramente, apesar da plataforma sugerir que o usuário preencha uma vasta quantidade de informações pessoais ao se cadastrar, apenas o seu *email* e nome de usuário (que pode ser gerado automaticamente) são obrigatórios. Por decisão interna, o gênero do usuário não é incluso dentre as informações pessoais que um usuário pode inserir em seu perfil.

Ademais, um usuário não precisa estar cadastrado para participar dos sites. Qualquer pessoa pode responder a qualquer pergunta aberta (que ainda possa receber respostas). Um perfil temporário é criado para este usuário e não pode mais ser acessado pelo mesmo após sua sessão expirar. Tal perfil é marcado como "não registrado". Um exemplo pode ser encontrado na Figura 2.3.

Apesar da vantagem da anonimidade, estes usuários que optaram por não se cadastrar não poderão usufruir das regalias que os usuários cadastrados que participam da comunidade possuem. Poder contribuir com perguntas e comentários, ter uma reputação associada ao seu nome e poder, caso tenha reputação suficiente, participar do processo de avaliação de qualidade do conteúdo do site, são alguns exemplos de habilidades restritas aos usuários cadastrados.

The screenshot shows the profile of an unregistered user named 'msPark'. The profile includes a purple geometric avatar, a reputation score of 219, and a 'top 94% this year' badge. The user's statistics are: 1 answer, 0 questions, and ~77k people reached. The user is a member for 3 years and 1 month, with 0 profile views and last seen on Apr 26 '12 at 5:51. The profile also shows 'Communities (1)' with Stack Overflow at 219 reputation, and 'Top Tags (4)' including windows (SCORE 21, POSTS 1, POSTS % 100), windows-7 (SCORE 21, POSTS 1), startup (SCORE 21, POSTS 1), and uac (SCORE 21, POSTS 1).

Figura 2.3: Perfil de um usuário não cadastrado. É criado um perfil temporário que não pode mais ser acessado pelo usuário após o término da sessão.

Capítulo 3

Literatura

Baseado na psicologia dos gêneros, encontramos duas teorias principais para justificar a ausência de mulheres em campos relacionados à tecnologia: a Essencialista, que vê as causas na biologia e o Constructo Social, que vê as causas no meio, o qual é considerado *male-friendly* (propício aos homens) mas não *female-friendly* (propício às mulheres). Contudo, ainda é importante levar em consideração as diferenças individuais destas pessoas ao estudar esse *gender gap* [47].

Na pesquisa aqui descrita estudamos gênero como fator social, como as pessoas se identificam, e não como fator biológico. Portanto usaremos o termo "gênero" ao invés de "sexo". Também generalizamos e estudamos apenas os dois extremos do espectro: aqueles usuários que se identificam como homens e os que se identificam como mulheres.

3.1 Diferenças de gênero em comunidades online

O comportamento de homens e mulheres em comunidades online nem sempre é o mesmo. Podemos observar diferenças desde quanto ao uso da própria internet já que mulheres julgam sua eficiência online como sendo baixa e isto influencia as tarefas escolhidas para serem efetuadas online [23]. Outro fator que podemos notar diferença é quanto à linguagem usada online. Esta última pode influenciar na visão que cada gênero tem de quão relevante e útil é o conteúdo da comunidade. Em outras palavras, a percepção de qualidade da comunidade [20].

Em ambientes de aprendizado online, homens interagem mais com o ambiente e enviam mensagens mais longas, com um contexto mais social, que as mulheres. Estas tendem a ter

uma participação mais pontual e relacionada ao tema das aulas, mostrando que as mulheres tendem a ser mais cuidadosas com o conteúdo que postam online, prezando por qualidade e pontualidade [9]. Na *Wikipedia*, apesar de homens e mulheres terem comportamentos de contribuição similares, as mulheres são as que produzem contribuições mais longas [21].

O *IRC (Internet Relay Chat)* é um meio de comunicação online que foi muito popular nos anos 2000. Seu funcionamento é simples: usuários escolhem um *nick*, se conectam ao servidor de sua preferência e podem conversar com grupos de pessoas de uma só vez nos *canais* ou com apenas uma no *privativo*. Hoje, apesar de contarmos com servidores e canais sobre os mais diversos assuntos, a maior parte deles é dedicada às comunidades *Open Source* de desenvolvimento de software. Quanto à questão de gênero no *IRC*, foi encontrado que, em canais de *IRC* sobre diversos assuntos, usuários que apresentam um pseudônimo feminino tendem a receber 25 vezes mais mensagens maliciosas que usuários com pseudônimo masculinos e 4 vezes mais que usuários com pseudônimos que não se identificava o gênero [33]. Isto pode ser um dos motivos pelo qual mulheres não participam tanto quanto homens deste tipo de comunidade.

Interessante comparar este resultado encontrado no *IRC*, que, teoricamente, não foi projetado especificamente para homens utilizarem com comunidades *MMORPG (Massively multiplayer online role-playing game)* – comunidades de jogos online – onde tais jogos foram explicitamente criadas por homens e para homens. Observou-se que estas comunidades que são conhecidas por não serem *female-friendly* e serem cenário de inúmeros casos de perseguição às mulheres, registraram um crescimento na proporção de mulheres participantes, este número chegando até a ser maior que o de homens [45]. Isso mostrando que, quando há o interesse, mulheres conseguem sim *sobreviver* em comunidades que apresentam um ambiente hostil a elas.

Contudo, a situação torna-se preocupante quando observamos resultados de pesquisas em comunidades online profissionais, já que estas podem refletir o que acontece fora do mundo virtual. Em comunidades *Free Open Source (FOSS)* – comunidades de desenvolvimento de software livre e gratuito – a proporção de mulheres contribuintes não passa de 5%, sendo a média 2%. Ademais, membros destas comunidades tendem a associar a culpa pela baixa proporção de mulheres à falta de interesse por parte delas, esquecendo os tantos casos de preconceito implícito e explícito divulgados em *blogs* e pela própria imprensa [38]. Tais pre-

conceitos podem aparecer de diversas formas: desde de mensagens entre os contribuidores até palestras que incluem mensagens que menosprezam mulheres. Não é para menos que, assim como a proporção de mulheres é baixa, seu tempo de vida nestas comunidades não é alto.

Por fim, encontramos apenas um estudo dedicado a gênero e comunidades do *StackExchange*. Vasilescu *et al* [48] estudaram a maior e mais antiga comunidade da plataforma, a *StackOverflow*, quanto a alguns aspectos da participação de homens e mulheres no site. Os resultados não são animadores: foi observado que mulheres são uma pequena minoria dentre os usuários e homens dedicam mais tempo à comunidade que as mulheres. Resultados similares aos encontrados em comunidades *FOSS*[38; 37]. Pode-se afirmar que estes resultados já eram esperados, visto os relatos espalhados por toda a internet, mas colocando em números, o problema parece ser mais grave e, aparentemente, sem uma explicação exata.

Em resumo, não podemos generalizar o comportamento de mulheres para todas as comunidades online, mesmo que elas tenham designs similares e propósitos iguais. Como todas as comunidades podem se beneficiar de usuários mais diversos, é importante estudar o comportamento de cada uma para ser possível sugerir soluções para o problema do *gender gap*.

3.2 Feedback e competitividade

Feedback mostra-se como uma ferramenta que influi bastante no comportamento dos usuários em comunidades online. Sabe-se que *feedback* negativo pode trazer um efeito indesejado à comunidade, fazendo com os que o recebem mantenham-se ativos e produzindo conteúdo de baixa qualidade [13]. O estudo mostra também que o efeito é viral: pessoas que recebem *feedback* negativo tendem a dar *feedback* negativo para os demais usuários indiscriminadamente. Em contrapartida, na *Wikipedia*, acredita-se que *feedbacks* positivos são um fator importante na motivação de mulheres a participar da comunidade, visto que este aumenta a auto-estima delas quanto ao seu conhecimento no assunto da contribuição [14].

Com relação à competitividade, estudos em laboratório mostram que mulheres não são tão eficientes quanto homens em ambientes competitivos, principalmente quando é neces-

sário competir com o gênero oposto [22]. As comunidades do *StackExchange* podem ser vistas como ambientes competitivos, visto que as respostas passam por um processo seletivo para cada pergunta. É possível que um baixo desempenho das mulheres nestas comunidades possa ser frustrante e levá-las a sair cedo da comunidade. Ademais, nota-se que, enquanto a maioria das mulheres as evitam, homens sentem-se confortáveis em competições [35; 15].

3.3 Mulheres, *STEM* e autoconfiança

Sabe-se que a disciplina matemática é base para muitos dos campos em *STEM*. Por isso, vários estudiosos quando tentam entender as diferenças entre gêneros em *STEM*, recorrem à matemática para realizar experimentos. Tais estudos mostram que mulheres têm performance tão boa quanto à de homens em atividades de matemática [25; 12], apesar da baixa autoconfiança que mulheres têm na sua capacidade cognitiva em matemática e da forte convicção, principalmente por parte dos homens, de que matemática é uma disciplina masculina. Tais estigmas e estereótipos se tornam impraticáveis já que outros estudos mostram que mulheres costumam ter notas melhores em disciplinas relacionadas à *STEM* [43] e que já se sabe que a performance das mulheres neste campo não depende de igualdade política, econômica ou social [43].

Contudo, o estereótipo de que mulheres são menos competentes em campos relacionados a *STEM* é bem visível e documentado. Por exemplo, Moss *et al* [34], ao estudar o processo de seleção de contratação de pesquisadores para universidades, observou que mesmo obtendo currículos idêntico aos dos seus colegas do sexo masculino, mulheres não eram escolhidas por serem consideradas menos competentes. Ademais, aquelas mulheres que conseguiam a vaga, muitas vezes tinham um salário inicial menor que o dos seus colegas com habilidades similares [34].

Apesar do seu potencial, mulheres ainda são minoria em campos relacionados à *STEM* que acredita-se que é necessário um talento inato para poder alcançar o sucesso na área [30]. Este é o caso de Ciência da Computação e Física, por exemplo. O motivo para que isto aconteça pode ser a crença por parte de ambos os gêneros de que mulheres não possuem tal talento [46; 28]. Porém, não foi provado ainda que esse "talento inato" é exclusivo de

homens ou mulheres [24].

Tais estereótipos de que mulheres são menos habilidosas ou que não possuem um talento inato para disciplinas e atividades relacionadas à *STEM* é extremamente danoso, não só para a carreira das mulheres, mas para a comunidade acadêmica e profissional no geral que pode sofrer com a falta de diversidade dentro do ambiente de estudo / trabalho. Mais ainda, estudos mostram que se mulheres absorverem este estereótipo, elas podem achar que estes campos não são para elas [49; 40], o que não é verdade, visto os resultados de estudos que compararam o desempenho de pessoas de ambos os gêneros em atividades de matemática e *STEM* [43; 25; 12].

O problema parece estar na falta de autoconfiança por parte das mulheres que tendem a duvidar que possuem tal talento e sentem-se desencorajadas a tentar campos como matemática e física [30]. Já foi observado que mulheres realmente possuem baixa autoconfiança em tarefas de ciências, se comparadas aos homens [18] e que resultados em tarefas de matemática afetam mais mulheres do que homens com relação a sua auto confiança em sua performance [12]. Esta falta de confiança em seu potencial já chegou até a afetar comunidades como a *Wikipedia*, onde a maior parte das contribuições vêm de uma minoria de homens [8; 29].

Ademais, o número baixo de mulheres em *STEM* parece influenciar outras mulheres a não se envolver na área: o fato de outras mulheres não escolherem um determinado campo pode ser um sinal de que ele não seja bom para mulheres no geral [40]. Contudo, mulheres tendem a não levar em consideração seu gênero quando decidem se querem seguir carreira em Ciência da Computação. Ao invés, condições culturais e o ambiente onde irão trabalhar são os fatores mais influentes em sua escolha [10]. Mais ainda, estudos apontam que mulheres que têm contato com modelos em ciência podem mudar a sua postura com relação à disciplina de uma maneira positiva [41], e isto pode ter um impacto positivo em seus interesses educacionais [36].

3.4 Nossa contribuição

Considerando a literatura apresentada neste capítulo, vemos que os estudos apontam que não devemos esperar que mulheres sejam menos habilidosas que homens em tarefas relaci-

onadas à *STEM*. O ambiente em que elas trabalham ou estudam juntamente com o *feedback* recebido para suas atividades mostram-se como verdadeiros fatores que influenciam o relacionamento de mulheres com campos relacionados a *STEM*. Contudo, não nos restou claro até que ponto o ambiente e *feedback* recebido pelas mulheres nas comunidades da plataforma *StackExchange* tem influenciado na participação e performance de mulheres na construção da base de conhecimento disponibilizado por estes sites, principalmente na área de *STEM*. Portanto, este estudo tem como o intuito preencher essa lacuna de conhecimento, estudando as diferenças e similaridades em comportamento entre homens e mulheres nas comunidades dedicadas a diversos assuntos pertencentes à plataforma *StackExchange*.

Capítulo 4

Metodologia

Este capítulo detalha os dados e métodos utilizadas para medir quantidade, qualidade e frequência das contribuições dos usuários, assim como o compromisso dos usuários com a comunidade. Aqui também explicamos como identificamos os gêneros dos usuários. Todo o código utilizado neste experimento pode ser encontrado em um repositório no *Github*¹.

4.1 Dados utilizados

Nosso estudo usa dados de todas as comunidades do *StackExchange* em seu último *dump* de Setembro de 2014. Contudo, para que fosse possível realizar a análise, tivemos que remover algumas comunidades que não possuíam atividades suficientes, por parte das mulheres, para serem analisadas. Como comunidades muito novas não apresentavam dados suficientes para serem estudados, removemos da nossa pesquisa todas as comunidades do *StackExchange*, presentes no *dump* de Setembro de 2014, que tenham menos de 18 meses de idade, além de comunidades que não tinham um número razoável de mulheres. Estas medidas nos garante que as comunidades estudadas possuem um número estatisticamente relevante para as duas amostras estudadas (homens e mulheres) em cada comunidade, evitando um viés ou erro estatístico por amostras de tamanho insuficiente. A lista das 85 comunidades estudadas pode ser encontrada na Tabela 4.1. Mais ainda, apenas os usuários que tenham feito algum tipo de contribuição e têm, no mínimo, 50 pontos de reputação foram estudados. Esta limitação é para garantir que estudamos usuários que estiveram ativos em algum momento na

¹<https://github.com/mil3na/gender-study-final>

comunidade estudada.

Comunidades Estudadas					
academia	cooking	french	magento	productivity	spanish
android	crypto	gamedev	math	programmers	sqa
anime	cs	gaming	mathematica	quant	stackapps
apple	cstheory	gardening	mathoverflow	raspberrypi	stackoverflow
askubuntu	dba	genealogy	mechanics	rpg	stats
bicycles	diy	german	money	russian	superuser
biology	drupal	gis	movies	salesforce	tex
bitcoin	dsp	graphicdesign	music	scicomp	travel
chemistry	electronics	hermeneutics	outdoors	scifi	unix
chinese	ell	history	parenting	security	ux
christianity	english	islam	philosophy	serverfault	webapps
codegolf	expressionengine	japanese	photo	sharepoint	webmasters
codereview	fitness	judaism	physics	skeptics	wordpress
cogsci	freelancing	linguistics	pm	sound	workplace
writers					

Tabela 4.1: Lista das comunidades do *StackExchange* estudadas nesta pesquisa.

As informações utilizadas neste estudo foram obtidas no *dump* trimestral² oferecido pelo *StackExchange*. O grupo costuma publicar regularmente todos os dados de todos os seus sites que não contenham informações privadas do usuário (e.g. *email*). Tais dados incluem perguntas, respostas e comentários postados pelos usuários, assim como os votos (positivos e negativos) recebidos por cada tipo de post. O *dump* também contém a reputação de cada usuário em cada comunidade. As comunidades presentes nestes dados são classificadas pelo *StackExchange* dentre estas seis categorias: *Technology*, *Culture/Recreation*, *Life/Arts*, *Science*, *Business* e *Professional*.

²<https://archive.org/details/StackExchange>

4.2 Deduzindo o gênero dos usuários

Os perfis do *StackExchange* não possuem explicitamente o gênero de cada usuário, já que a plataforma não exige do usuário esta informação no seu perfil. Sendo assim impossível identificar o gênero de cada usuário automaticamente, com precisão. Já que nosso objetivo neste estudo é apenas investigar o efeito de um usuário, explicitamente, manifestar seu gênero, decidimos focar apenas naqueles usuários que têm a intenção de identificar seu gênero nestas comunidades.

As comunidades do *StackExchange* possuem três tipos de atividades principais: perguntar, responder e comentar. Durante a participação de um usuário nas atividades principais, só é possível identificar o gênero do seu colega por dois meios: sua foto ou nome de usuário, como pode ser visto na Figura 2.2. Utilizamos o nome de usuário para identificar o seu gênero. Sabe-se que a maioria dos usuários do *StackExchange* vêm de países ocidentais, provavelmente tendo nomes típicos ocidentais [39], e pesquisas anteriores mostram que, para pessoas com nomes ocidentais, inferir o gênero de alguém a partir de seu primeiro nome é um método acurado tanto para redes sociais que requerem o nome real do usuário [44], quanto para as que não o requerem [11; 31].

Nosso método de inferência de gênero é similar ao encontrado em Liu *et al.* [31] e Cunha *et al.* [16]. Nós utilizamos o *Global Name Data* [26], um banco de dados de nomes baseado em registros de nascimentos nos Estados Unidos e Reino Unido, que contém mais de 100,000 nomes únicos e a frequência com que cada nome é associado a um menino ou menina. Com este banco de dados, fizemos um classificador que categoriza os usuários em três classes: Feminino, Masculino e Desconhecido, baseado na contagem de frequência probabilística de cada nome. Um usuário é classificado como Masculino se existir uma proporção significativamente maior de meninos registrados com esse nome, segundo o *Global Name Data*. O gênero de um nome é considerado Desconhecido caso não haja diferença estatística entre a proporção de homens e mulheres catalogados com este nome. Nossa análise só leva em consideração aqueles usuários que possuem nomes os quais é possível inferir com confiança o gênero, de acordo com este método. Com este método, conseguimos identificar, em média, 37% dos usuários (mínimo: 27%, máximo: 52%) das comunidades estudadas. Sendo sua grande maioria homens: em média, 93% são homens (mínimo: 74%, máximo:

98%). O Apêndice A apresenta o sumário dos usuários identificados por gênero, em cada comunidade.

4.3 Manuseio dos dados

Os dados concedidos pelo *StackExchange* são disponibilizados no formato *XML*, o qual dificulta a manipulação desses. Portanto, a primeira etapa foi a inserção dos dados em um banco de dados. Como pretendíamos realizar várias associações entre as comunidades e não tínhamos um *schema* fixo, optamos por um banco de dados flexível, orientado a documentos, o *MongoDB*. Por praticidade, utilizamos a linguagem de programação *Python* para filtrar de dentro dos arquivos *XML* os dados nos quais estávamos interessados, transformar no formato adequado, criar um sub banco de dados no *MongoDB* para cada comunidade e inserir seus respectivos dados nessa comunidade.

A próxima etapa foi filtrar os usuários que não nos interessam. Esta pesquisa foca em comportamento de usuários em um site, então os usuários que não participam da comunidade não são relevantes para o nosso estudo. Um script em *Python* foi criado para remover do nosso banco de dados todos aqueles usuários que realizaram nenhuma contribuição ou que tenham uma reputação muito baixa. Consideramos uma reputação baixa aquela que não permite que um usuário possa realizar qualquer uma das principais contribuições possíveis na plataforma. Na época do estudo, o usuário precisaria de 50 pontos de reputação para poder perguntar, responder ou comentar no site.

Para cada usuário foi criado um sumário da sua participação em cada comunidade e inserido em sua entrada no banco de dados. Mais uma vez um script escrito em *Python* foi utilizado para realizar esta tarefa. O sumário de cada usuário conta com o número de identificação do usuário no site, sua reputação, a data de ingresso, a média da pontuação obtida nas perguntas, respostas e comentários feitos, o número total de perguntas, respostas e comentários feitos, a taxa de aceitação e utilidade média das respostas, as datas de todas as contribuições feitas e o tipo da primeira contribuição.

4.3.1 Classificando os usuários quanto ao gênero

Para classificar os usuários quanto ao seu gênero, criamos um classificador de gênero dos usuários baseado nos nomes que eles forneciam em seu perfil. Para isso fizemos o *download* do banco de dados oferecido pelo *Global Name Data* por meio do site *Github*³ e o inserimos em nosso banco de dados *MongoDB*. O classificador foi escrito em *Python* utilizamos os dados já filtrados que se encontram em nosso banco de dados. Para cada usuário de cada site, selecionamos seu nome e verificamos se ele se encontra no banco de dados de nomes. Caso seja encontrado, inserimos a classe a qual ele pertence (Feminino, Masculino ou Desconhecido) ao seu sumário, de acordo com a classificação fornecida pelo *Global Name Data* para cada nome através de um estimador binomial. Se não, este usuário será classificado como tendo o gênero desconhecido. Apenas os usuários que foram classificados como Feminino ou Masculino são utilizados neste estudo.

4.3.2 Cálculos estatísticos

Visto que a manipulação dos dados feitas até este ponto do experimento foram realizadas utilizando scripts em *Python*⁴ e a não compatibilidade do banco de dados *MongoDB*⁵ com a linguagem *R*⁶, na época em que o estudo foi realizado, optamos por utilizar as bibliotecas *pandas*⁷, *numpy*⁸, *scipy*⁹ e *statsmodels*¹⁰ que provêm uma funcionalidade similar a encontrada na linguagem *R*. Ademais, utilizamos a ferramenta *ipython notebook*¹¹ que nos permite dispor código e resultados (sejam estes numéricos ou gráficos) de forma simples e clara, facilitando o entendimento e reprodução do experimento. Estes *notebooks* assim como todo o código utilizado para esta pesquisa podem ser encontrados no repositório *Github*¹² da pesquisa.

³<https://github.com/OpenGenderTracking/globalnamedata/tree/master/assets>

⁴<https://www.python.org/>

⁵<https://www.mongodb.org/>

⁶<https://www.r-project.org/>

⁷<http://pandas.pydata.org/>

⁸<http://www.numpy.org/>

⁹<http://www.scipy.org/>

¹⁰<http://statsmodels.sourceforge.net/>

¹¹<http://ipython.org/notebook.html>

¹²<https://github.com/mil3na/gender-study-final>

4.4 Medindo contribuições e engajamento

A questão principal do nosso estudo é esclarecer o quão diferente é a participação de homens e mulheres em comunidades de perguntas e respostas, comparando aquelas relacionadas à *STEM* (pertencentes às categorias *Technology* e *Science*) com as demais. Para responder esta questão, nós a dividimos em quatro perguntas mais específicas, que serão relatadas a seguir.

4.4.1 Número de contribuições

Nossa primeira pergunta foi: *O número de contribuições feitas por homens e por mulheres difere significativamente nas comunidades observadas?* Esta pergunta é relacionada aos três tipos principais de contribuição na plataforma *StackExchange*: perguntas, respostas e comentários, além da soma de todas. Portanto, aqui analisamos o número total de perguntas, respostas, comentários e contribuições totais feitas por cada usuário e comparamos se existe uma diferença estatisticamente relevante para o número de contribuições feitas por homens e mulheres.

Com a proporção de mulheres sendo menor que a de homens na maioria dos sites das plataformas e, em alguns sites, esta proporção atingindo menos de um terço dos usuários identificados, esperávamos que isto se refletiria nas contribuições provenientes das mulheres. Em outras palavras, acreditávamos que a resposta para esta primeira pergunta seria que, independente do tipo de contribuição, mulheres contribuiriam menos do que os homens.

4.4.2 Qualidade

O controle de qualidade das perguntas e respostas na plataforma *StackExchange* é feito pela própria comunidade por meio do sistema de votos. Usuários com uma determinada reputação podem votar "para cima" e "para baixo", caso considerem o post como uma boa ou má contribuição, respectivamente. Esses votos geram pontos, positivos ou negativos, que, dentre outros fatores, formam a reputação de um usuário.

Tendo em vista o processo de avaliação de qualidade descrito acima, a nossa segunda pergunta foi: *As contribuições feitas por diferentes gêneros são vistas, pela comunidade, com níveis de qualidade diferentes?* Caso exista uma avaliação negativa indiscriminada das contribuições provenientes de mulheres, elas podem estar contribuindo menos porque suas

contribuições não são bem provenientes, fazendo com que elas sintam que falharam, as desencorajando de continuar participando.

Utilizamos a pontuação criada pelo sistema de votos para medir a qualidade das perguntas e respostas de cada usuário. A qualidade das perguntas de um usuário é definida como a média da pontuação de todas as perguntas feitas pelo usuário, ou zero, caso o usuário tenha contribuído com nenhuma pergunta. Para respostas, utilizamos duas métricas: a primeira é a taxa de aceitação que mede a proporção de respostas do usuário que foram aceitas como melhor resposta pelo interrogador. A segunda é a utilidade média das respostas, proposta por Furtado *et al.* [19]. Para calcular esta métrica, primeiramente faz-se uma normalização do saldo de votos de todas as respostas à uma pergunta. Esta métrica é a média dos saldos de votos das respostas de um usuário comparado às outras respostas de uma mesma pergunta. Para ambas as métricas que medem a qualidade das respostas, apenas usuários com pelo menos uma resposta são levados em consideração.

4.4.3 Engajamento

Neste estudo, usamos a definição de engajamento, a qual o descreve como o esforço de um usuário ao longo do tempo. Nós medimos engajamento tanto em relação ao tempo total que o usuário permanece na comunidade (seu tempo de vida) quanto em relação a sua frequência de contribuição durante seu tempo de atividade.

Levando em consideração estas definições, nossa terceira pergunta foi: *O engajamento de homens e mulheres difere nas comunidades estudadas?* É importante notar que, caso haja diferença no hábito de contribuição entre homens e mulheres, podemos verificar se usuários de algum dos gêneros deixam a comunidade mais cedo ou têm uma taxa de contribuição menor, nos ajudando a entender melhor as diferenças (ou similaridades) entre homens e mulheres com relação ao seus hábitos de contribuição.

O tempo de vida de um usuário em uma comunidade do *StackExchange* não pode ser considerado apenas como a diferença entre a data de sua primeira contribuição e a última. Já que uma pergunta, e suas respectivas respostas, podem ser transferidas entre comunidades¹³. Desse modo, definimos a seguir o tempo de vida de um usuário em uma determinada comunidade.

¹³<http://meta.stackExchange.com/questions/2683/move-questions-between-stack-exchange-sites>

Antes de tudo, definimos o primeiro dia de atividade do usuário como a data mais recente entre o dia de cadastro do usuário e a data da sua primeira contribuição na comunidade. Esta medida evita que levemos em consideração um período que o usuário não esteve ativo na comunidade em questão, mas esteve em outra comunidade e sua pergunta foi transferida. Mais ainda, definimos que um usuário está inativo quando ele não produz nenhuma contribuição por um período de tempo maior que o intervalo de morte da comunidade. Este intervalo, medido em dias, é calculado da seguinte maneira: para cada usuário obtemos o maior intervalo entre duas de suas contribuições consecutivas. O intervalo de morte é a média dos maiores intervalos de todos os usuários. Por fim, o tempo de vida de um usuário é a diferença entre a data da sua última contribuição e a data da sua primeira atividade (ou cadastro), medida em dias.

A frequência de participação, segunda métrica usada nesta pergunta, é definida como o total de contribuições feita por um usuário (perguntas, respostas ou comentários) dividido pelo número de dias nos quais o usuário estava ativo. Ao contrário do tempo de vida, aqui nós consideramos apenas os dias em que o usuário contribuiu para a comunidade.

A nossa expectativa para a resposta desta pergunta era que mulheres se engajam menos, principalmente às comunidades que abordam assuntos relacionados à *STEM*. Existe um estigma de que estas comunidades não são atrativas às mulheres e que também são excludentes, fazendo com que mulheres não sejam bem-vindas. Por esses motivos, as poucas mulheres que se cadastrassem em qualquer site o deixaria em pouco tempo.

4.4.4 Contribuições e novos registros ao longo do tempo

Para esta pergunta, verificamos se a falta de contribuição por parte das mulheres é algo novo nestas comunidades, ou ainda se existem padrões de contribuição ao longo do tempo, seja de aumento ou redução. Investigamos também se existe alguma tendência no número de registros feito por cada gênero. Em outras palavras, buscamos responder à seguinte pergunta: *A proporção de contribuições e novos registros feitos por usuários de cada gênero tem aumentado ou reduzido ao longo do tempo?*.

Se mulheres contribuem pouco e por pouco tempo é lógico esperar que a proporção de contribuições por parte das mulheres tenha aumentado por um tempo, mas estagnado depois e, provavelmente até diminuído nos últimos meses. O mesmo é o que esperamos para a

proporção de novos registros, até porque estas comunidades são conhecidas por não serem *female-friendly*.

4.5 Distribuições e testes

A maioria das variáveis possui uma distribuição bastante enviesada e com certeza não normal, como pode ser visto no Apêndice B. Portanto, utilizamos testes não paramétricos. Para comparar o número de contribuições (e de cada tipo de contribuição), a qualidade das contribuições e a frequência de contribuição de cada gênero, utilizamos o teste um teste de dominância estocástica, o *Mann-Whitney-U* não pareado. Primeiramente, a versão bilateral do teste para verificar se há diferença estatística e, em seguida, o mesmo teste é aplicado aos grupos que apresentaram diferença estatística mas, desta vez, utilizando sua versão unilateral para verificar se o grupo de mulheres era o que apresentava dominância estocástica. Por fim, utilizamos o estimador de *Hodges-Lehmann* para averiguar o tamanho da diferença entre cada amostra, para cada variável onde utilizamos o teste estatístico de *Mann-Whitney*

Para testar se homens e mulheres possuem um tempo de vida divergente na comunidade utilizamos a *Survival Analysis* junto com um teste *LogRank*. Para a quarta pergunta tomamos a proporção de contribuição e novos registros feitos por cada grupo de gênero, a cada seis meses, e verificamos se esta proporção aumenta ou não, com a ajuda de uma regressão utilizando o método dos mínimos quadrados.

Sumário das variáveis estudadas, para cada pergunta de pesquisa, podem ser encontrados nos Apêndices C, D, E e F.

Capítulo 5

Resultados

Este capítulo apresenta os resultados para cada uma das quatro perguntas de pesquisa. Quando descrevemos as diferenças para uma determinada variável de um conjunto de sites, nós utilizamos da notação $M/S/H$, em que M é o número de sites onde mulheres obtiveram valores para esta variável significativamente maiores, S o número de sites onde não obteve-se diferença significativa e, por fim, H é o número de sites onde homens obtiveram valores significativamente maiores. Mais ainda, nas figuras aqui apresentadas, as caixas representam os quartis das distribuições enquanto os traços mostram o primeiro e nonagésimo nono percentis. Para visualizar a distribuição em escala de log, acrescentamos 1 ao valor original do número de contribuições realizadas por cada usuário. Por fim, lembramos que o teste MWW pode achar diferença estatística na soma dos rankings mesmo que os quartis sejam similares nas duas amostras.

5.1 Número de contribuições

Para todos os tipos de contribuições considerados, um padrão em comum surgiu em nossos resultados: a diferença da quantidade de contribuições provenientes de homens e mulheres não é significativa na maioria dos sites (teste *Mann-Whitney U*, $\alpha = 0.05$). E, mesmo para os sites onde pudemos observar diferença estatística, esta não foi grande. O estimador de Hodges-Lehmann nos indica que, para qualquer tipo de contribuição, não há diferença maior do que uma unidade de contribuição entre homens e mulheres.

Dos sites em que pudemos observar uma diferença estatística significativa, existe um

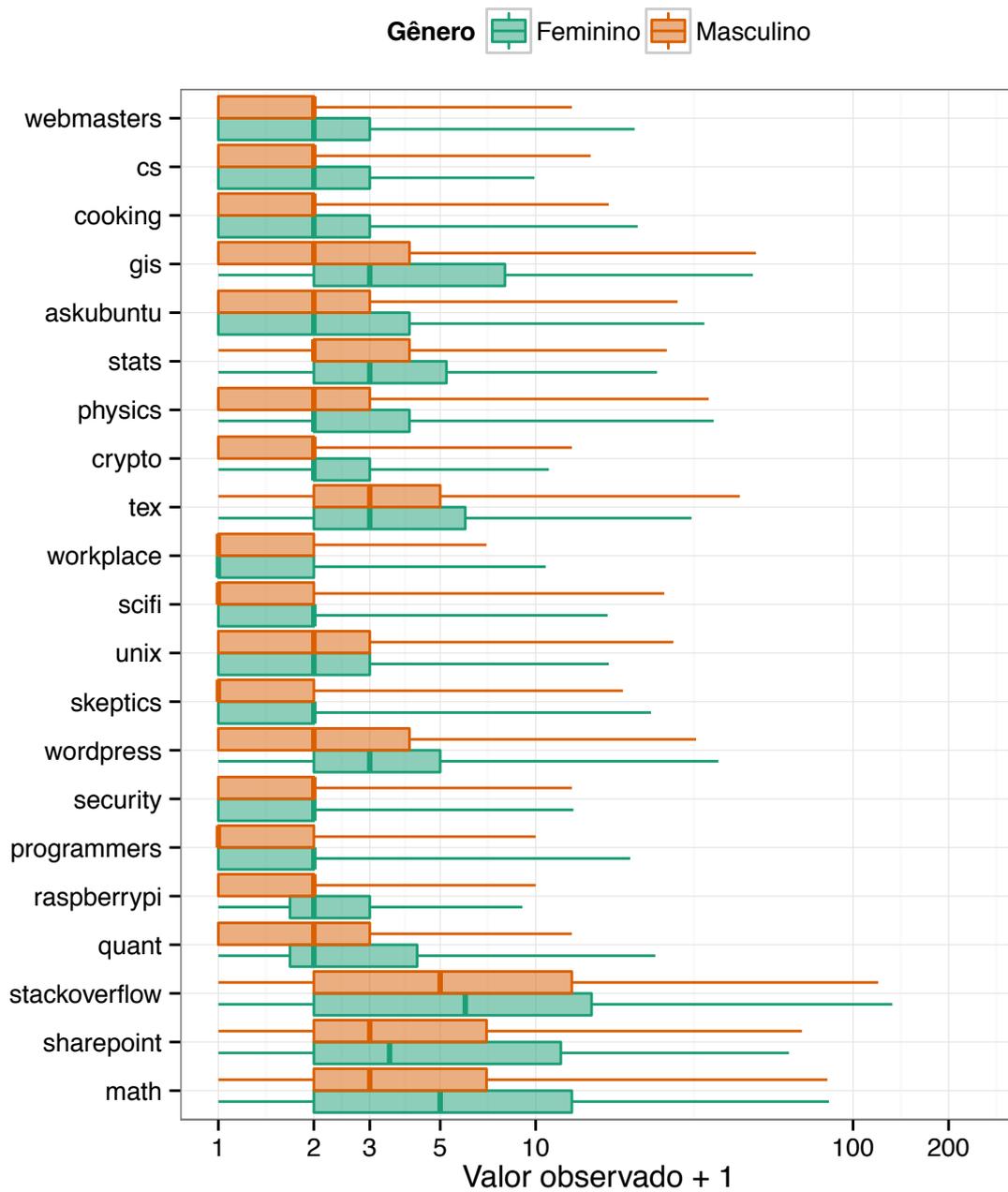


Figura 5.1: Comparação do número de perguntas provenientes de contribuidores de diferentes gêneros nos sites onde pode-se verificar uma diferença estatisticamente significativa e mulheres contribuem mais do que homens.

número maior de comunidades nas quais podemos observar que mulheres são mais ativas que homens do que o contrário. Esta diferença é maior quando observamos o número de perguntas (21/62/2), seguido pelo número de respostas (18/56/11) e, finalmente, o número total de contribuições (10/73/2).

Interessante notar que existe uma diferença no padrão de categorias nas quais podemos notar diferenças no número de respostas entre homens e mulheres, comparando aos outros tipos de contribuição. Para perguntas, comentários e a soma de todas as contribuições, as comunidades onde mulheres são geralmente mais ativas pertencem a categorias relacionadas à *STEM*. Para perguntas a distribuição de comunidades relacionadas à *STEM* é 16/28/0, como pode ser observado na Figura 5.1. Contudo, quando olhamos para respostas, apenas três das dezoito comunidades onde mulheres respondem mais do que homens são relacionadas a *STEM* (como pode ser visto na Figura 5.2). É mais comum encontrar comunidades das categorias *Life/Arts* ou *Professional* dentre os sites que pode-se observar que mulheres respondem significativamente mais do que homens. Mais ainda, duas das três comunidades relacionadas à *STEM*, onde mulheres respondem mais que homens abordam assuntos que são estereotipados como mais femininos: *UX* e *Software Quality Assurance*. As comunidades onde homens respondem significativamente mais do que mulheres podem ser vistas na Figura 5.3.

Estes resultados nos mostram que mulheres que identificam seu gênero e que se engajam à comunidade, na maioria dos sites, não contribuem menos do que os homens. Ao contrário: quando podemos observar diferenças, essas frequentemente apontam que mulheres contribuem mais do que homens.

5.2 Qualidade

Poucos foram os sites que apresentaram uma diferença significativa em como a comunidade avalia a qualidade das contribuições de homens e mulheres. Esta diferença ocorre menos frequentemente do que as diferenças em número de contribuições. Para a taxa de aceitação (7/76/2) e utilidade média das respostas (9/74/2) dos usuários, um número pequeno de comunidades tende a atribuir às respostas das mulheres uma qualidade maior do que às dos homens. Para a média dos votos de uma pergunta, podemos observar um equilíbrio nos sites

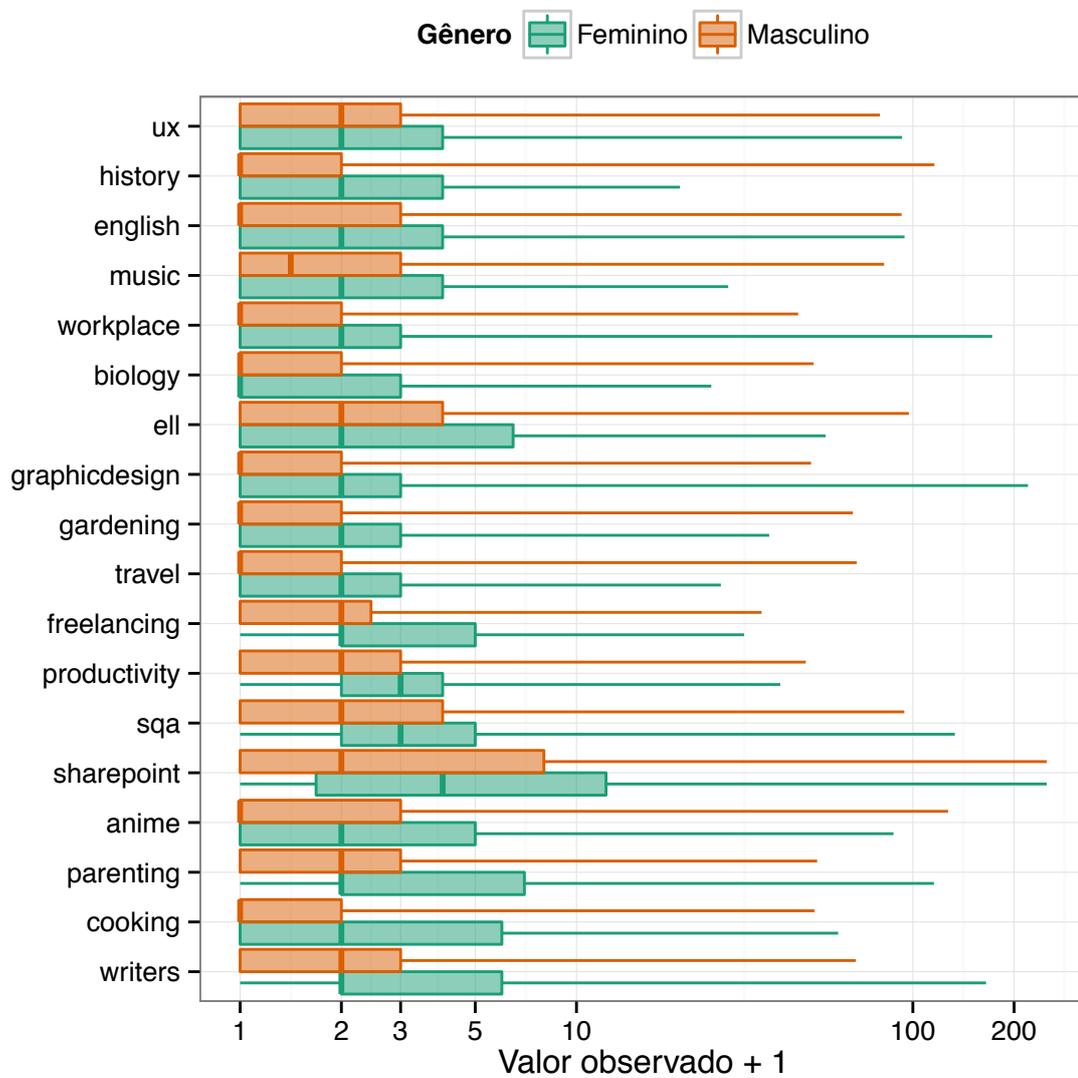


Figura 5.2: Comparação do número de respostas provenientes de contribuidores de diferentes gêneros nos sites onde pode-se verificar uma diferença estatisticamente significativa e mulheres contribuem mais do que homens.

que estudamos (8/67/10). No encontramos um padro relacionado s categorias a partir dos quais podemos encontrar diferena significativa na avaliao de qualidade. E, assim como para o nmero de contribuices, o estimador de Hodges-Lehmann nos indica um valor infimo na diferena de qualidade entre homens e mulheres, quando o teste estatístico indica que h uma diferena entre as amostras.

5.3 Dedicaco

Para as mtricas relacionadas ao tempo dedicado  comunidade, as diferenas observadas possuem um padro bem claro. Primeiramente, para a maioria dos sites, no h diferena estatística quanto ao tempo de vida ou frequncia de atividade de homens e mulheres. Contudo, quando constatamos diferena no tempo de vida, estas acontecem porque homens passam mais tempo na comunidade do que mulheres (1/70/14). E o contrrio acontece com relao  frequncia de contribuico em dias ativos: mulheres contribuem mais frequentemente do que homens (10/74/1).

Ns tambm notamos que as maiores diferenas entre os gneros quanto ao tempo de vida na comunidade ocorre no StackOverflow, o mais popular e tambm o mais antigo site do *StackExchange*. Os prximos trs sites em grandeza de diferena de tempo de vida tambm so antigos e relacionados a *STEM*: *MathOverflow*, *SuperUser* e *ServerFault*. Estas diferenas podem ser observadas na Tabela 5.1.

Com relao  frequncia, um padro diferente pode ser observado. Em todos os sete sites relacionados a *STEM* que apresentaram diferena estatística com relao  frequncia de atividade, mulheres contribuem mais frequentemente do que homens. Assimilando isto com o resultado relacionado ao tempo de vida, podemos inferir que, apesar de as mulheres deixarem a comunidade antes dos homens, elas contribuem mais durante o tempo que ainda esto associadas ao site.

5.4 Contribuices e novos registros ao longo do tempo

Observando as contribuices de homens e mulheres ao longo do tempo, notamos que em, aproximadamente, um quarto dos sites estudados a proporo das contribuices provenientes

de mulheres está aumentando (20/60/5). Apesar de não tão evidente, das comunidades que apresentaram diferença estatística quanto aos registros feitos por homens e mulheres ao longo do tempo, é possível ver um padrão: os registros de mulheres têm aumentado (8/76/1). Ambos resultados são mais visíveis se considerarmos apenas os sites relacionados à *STEM*: 16/25/3 para contribuições ao longo do tempo e 6/38/0 para a proporção de novos registros.

5.5 Focando nos maiores sites

Os três maiores sites do *StackExchange* – *StackOverflow*, *ServerFault* e *SuperUser* – são muito maiores e mais antigos que os demais sites disponíveis na nossa amostra. Juntamente com os sites *Mathematics* e *Programmers* eles possuem uma proporção significativa dos usuários de toda a plataforma *StackExchange*, sendo os cinco maiores sites, em número

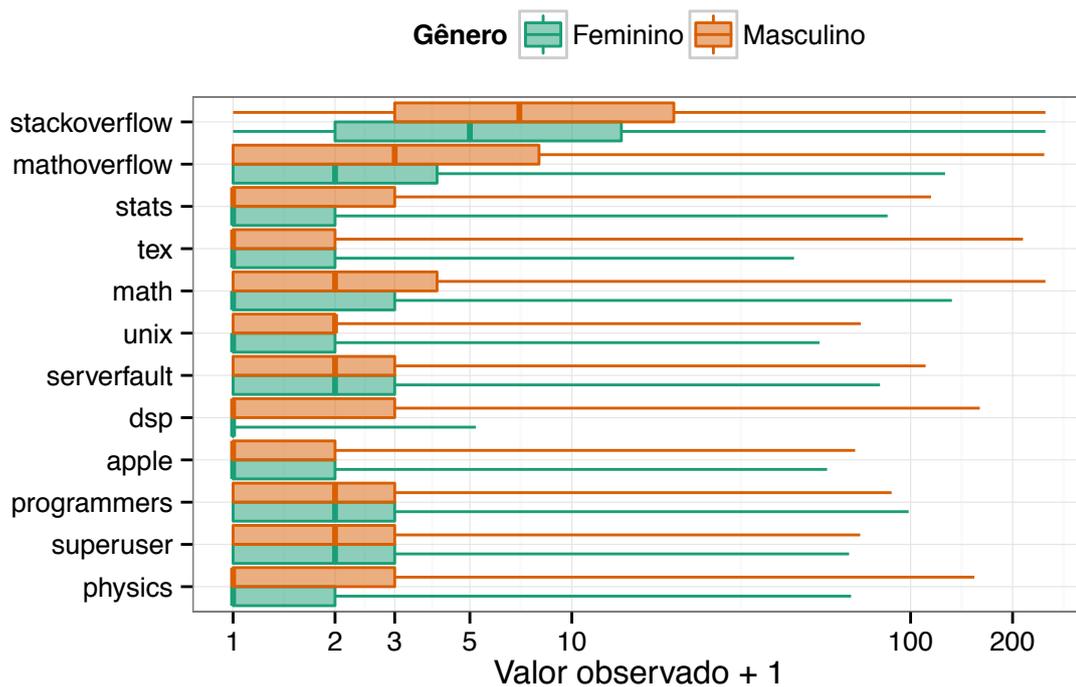


Figura 5.3: Comparação do número de respostas provenientes de contribuidores de diferentes gêneros nos sites onde pode-se verificar uma diferença estatisticamente significativa e homens contribuem mais do que mulheres.

de usuários, da nossa amostra, achamos por bem evidenciar o comportamento dos usuários nestas comunidades.

No *StackOverflow*, primeira comunidade da plataforma, mulheres perguntam mais e homens costumam responder mais, apesar de não encontrarmos diferença estatística no total

<i>Dif. Mediana</i>	<i>Categoria</i>	<i>Comunidade</i>	<i>Idade (meses)</i>
-346.87	technology	stackoverflow	76
-327.76	science	mathoverflow	62
-70.93	technology	superuser	64
-35.29	technology	serverfault	67
-17.15	technology	dsp	39
-14.50	technology	apple	51
-12.62	life-arts	diy	52
-9.05	culture-recr.	english	51
-8.80	life-arts	scifi	46
-7.26	technology	networkengineering	18
-6.42	culture-recre.	mechanics	44
-4.98	life-arts	photo	52
-1.92	culture-recre.	travel	41
-1.66	science	philosophy	41
74.72	technology	wordpress	51

Tabela 5.1: Sites onde podemos observar uma diferença estatística significativa entre homens e mulheres quanto ao seu tempo de vida na comunidade. Apresentamos a diferença entre a mediana de dias ativos, a idade da comunidade em meses e sua categoria.

de contribuições feita por ambos os gêneros. Perguntas e respostas vistas como mais qualificadas pelos usuários advêm de homens. Ademais, o padrão de homens possuírem maior tempo de vida na comunidade, enquanto mulheres possuem maior frequência de contribuição também pode ser visto neste site. E, ainda, a proporção de mulheres se cadastrando e contribuindo para a comunidade tem aumentado nos últimos anos.

Os três maiores sites do *StackExchange* – *StackOverflow*, *ServerFault* e *SuperUser* – são muito maiores e mais antigos que os demais sites disponíveis na nossa amostra. Juntamente com os sites *Mathematics* e *Programmers* eles possuem uma proporção significativa dos usuários de toda a plataforma *StackExchange*, sendo os cinco maiores sites, em número de usuários, da nossa amostra, achamos por bem evidenciar o comportamento dos usuários nestas comunidades.

Em todas as demais comunidades dentre as cinco maiores comunidades da nossa amostra, homens respondem mais que mulheres. Tal fato já era esperado, já que todas tratam de assuntos relacionados à *STEM*. Mulheres perguntam mais nos sites *Programmers* e *Mathematics*. E, nesse último, mulheres comentam e têm um número total de contribuições maior que o dos homens também. Quanto à qualidade, apenas a *ServerFault* apresentou diferença estatística e em apenas uma métrica (média de votos das perguntas), favorável aos homens. Nos sites *ServerFault*, *SuperUser*, homens apresentam tempo de vida na comunidade maior do que as mulheres, enquanto apenas na *Mathematics* mulheres possuem uma frequência de postagem maior que a dos homens. As demais não apresentam diferença estatística. Ao longo do tempo, mulheres têm contribuído mais do que homens nos sites *Programmers*, *Mathematics* e *ServerFault*.

Por fim, uma comparação do número de contribuições nos cinco maiores sites do *StackExchange* relacionados a *STEM* pode ser encontrada na Figura 5.4.

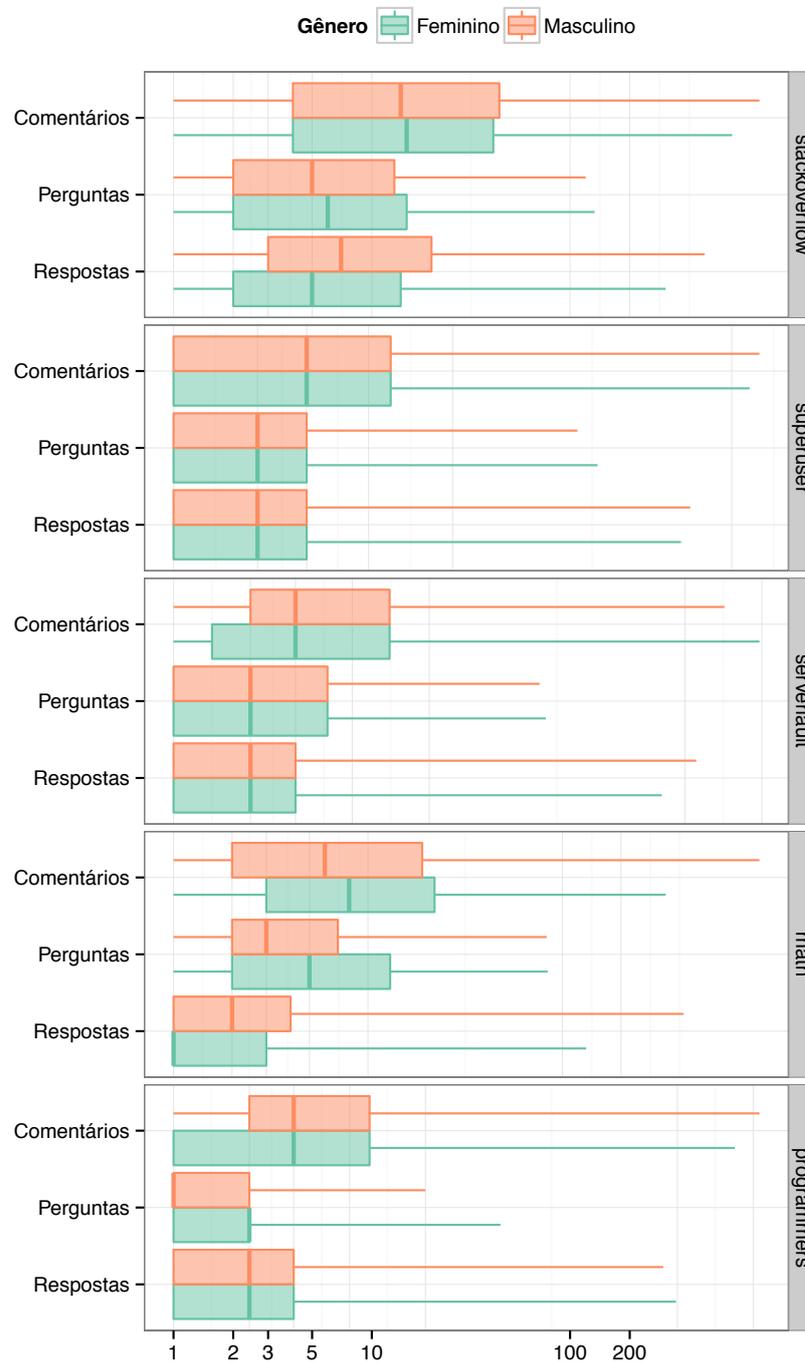


Figura 5.4: Comparação do número de contribuições provenientes de usuários de diferentes gêneros nos cinco maiores sites (em quantidade de usuários) relacionados a *STEM* do *StackExchange*.

Capítulo 6

Discussão

Neste capítulo iremos comentar os resultados obtidos no nosso experimento e associá-los à literatura relacionada ao estudo aqui descrito. Dividimos nossos resultados entre as semelhanças e diferenças encontradas e, apesar da importância do primeiro tópico, focamos nas diferenças por estas serem mais controversas.

6.1 As semelhanças

Nossa expectativa ao observar contribuições individuais de usuários de diferentes gêneros era que mulheres, por serem uma minoria, iriam participar menos destes sites, assim como na *Wikipedia* [8] e comunidades *FOSS* [38]. Contudo, o que encontramos não foram diferenças, mas uma grande equivalência entre o comportamento de homens e mulheres nos sites estudados.

No geral, saber que mulheres contribuem de maneira similar aos homens na maioria dos sites aponta que uma mulher que se cadastrar em um site aleatório do *StackExchange* provavelmente irá encontrar um ambiente que não a estimule a contribuir menos ou menos frequentemente que os homens. Tal resultado é animador visto que, apesar da pequena proporção de mulheres nestas comunidades, o comportamento das mesmas não segue a linha das pesquisas anteriores onde mulheres contribuem menos que homens quando estão em minoria.

6.2 As diferenças

Tendo em mente todo o estereótipo em *STEM* que mulheres são enquadradas, esperávamos que sua baixa representatividade também representaria baixa contribuição. Ademais, esperávamos que o estereótipo de mulheres em tecnologia [25] e aquele em que mulheres não possuem um talento inato para ciência [30] fosse influenciar na opinião que os outros usuários teriam das suas colegas, fazendo com que mulheres recebessem *feedback* negativo nas suas contribuições, indiscriminadamente, levando o conteúdo das mulheres a serem classificados com menor qualidade pela comunidade. Sem esquecer que outros estudos já tinham identificado o conteúdo vindo de mulheres como de qualidade inferior comparado ao de homens, tanto em tarefas de matemática [12] quanto em comunidades online [14].

Contudo, não é isso que acontece. Na maioria dos sites, mulheres produzem conteúdo de tão boa qualidade quanto homens, segundo a visão da comunidade. Vemos que estas mulheres contribuem bastante com perguntas e comentários, tendo uma frequência de contribuições muitas vezes maior que a de homens.

Por outro lado, apesar de mulheres responderem mais do que homens em várias comunidades, quando observamos apenas comunidades relacionadas à *STEM* isto não acontece. Mulheres contribuem com menos respostas do que homens em todos os quatro maiores e mais antigos sites da plataforma: *StackOverflow*, *SuperUser*, *ServerFault* e *Mathematics*. Se estes sites influenciam mulheres a participar menos no processo de resposta, a grandeza deles pode implicar que um grande número de mulheres está contribuindo com menos respostas do que poderia estar. Sabendo que existem sites em outras categorias nas quais mulheres comuns respondem mais do que homens, nossos resultados sugerem que há um fator característico em alguns sites – e provavelmente à comunidade – atuando no comportamento das mulheres quanto às suas contribuições. A ampla diferença no tempo de vida entre homens e mulheres nestas comunidades complementa este pensamento (ver Tabela 5.1).

Uma característica comum a algumas comunidades relacionadas à *STEM* do *StackExchange* é que elas são ramos do *StackOverflow* que foi criada por um grupo de homens e se manteve assim por bastante tempo, até se tornar popular. É provável que este grupo de homens tenha sua própria cultura dentro do *StackOverflow*, com seus valores e práticas que não sejam amigáveis às mulheres, e esta cultura tenha passado para os demais sites,

fazendo com que estes sejam sites com características sociais que não atraiam as mulheres.

Nosso resultado se alinha com o de Yang *et al.* [50], que apesar de não estudar a mesma comunidade de perguntas e respostas, defendeu que usuários que têm o tempo de vida maior na comunidade preferem responder à perguntar. No nosso estudo, verificamos que estes usuários são geralmente homens, assim como entenderam Vasilescu *et al.* [48]. Vasilescu *et al* consideram este resultado insalubre para a comunidade, visto que bons contribuidores estão deixando a comunidade. Nós acreditamos que exista uma barreira para que mulheres continuem contribuindo, podendo ser desde a aquisição, por parte delas, do pensamento de que não são capazes de contribuir, até por perseguições/assédios dentro da própria comunidade por parte dos outros usuários. Estas hipóteses, contudo, pedem uma investigação mais detalhada, fora do escopo desta pesquisa.

Apesar de não ser relacionada com sua real performance, a percepção que uma pessoa tem de sua capacidade cognitiva em um campo influencia como esta pessoa acha que se sairá em uma determinada tarefa. No mais, achar que não vai se dar bem em uma tarefa pode fazer com que, caso seja dada a opção, a pessoa não chegue a tentar realizar tal atividade. Mulheres acham que não são boas em tarefas sobre ciência e isso as influencia a não participar do campo [17]. Isso pode explicar porque a proporção de mulheres é tão baixa nas comunidades sobre *STEM*, mas elas se dão tão bem quanto homens na maioria delas.

Se olharmos para comunidades de áreas de *STEM* que são conhecidas por terem mais mulheres, observamos que o resultado é similar aos de comunidades não relacionadas à *STEM*. Tomando como exemplo a comunidade *UX* (*user experience*), cujas variáveis podem ser observadas na Figura 6.1, mulheres respondem mais nesta comunidade, o que pode ser uma demonstração de auto confiança no seu conhecimento no assunto. Mais ainda, *UX* não é considerada uma disciplina que necessita de um conhecimento inato para ter sucesso na área, sendo este mais um motivo para mulheres não se inibirem em participar desta comunidade de todas as formas possíveis.

Ver que mulheres têm aumentado sua proporção de contribuições ao longo do tempo, principalmente em comunidades sobre assuntos relacionados à *STEM*, nos dá uma esperança de que a barreira que existe para que mulheres entrem e contribuam para estas comunidades esteja diminuindo. Já que outras comunidades, como as comunidades *FOSS* [37], mulheres que já estão envolvidas tendem a continuar fazendo parte dela. Uma maneira de aumentar a

UX

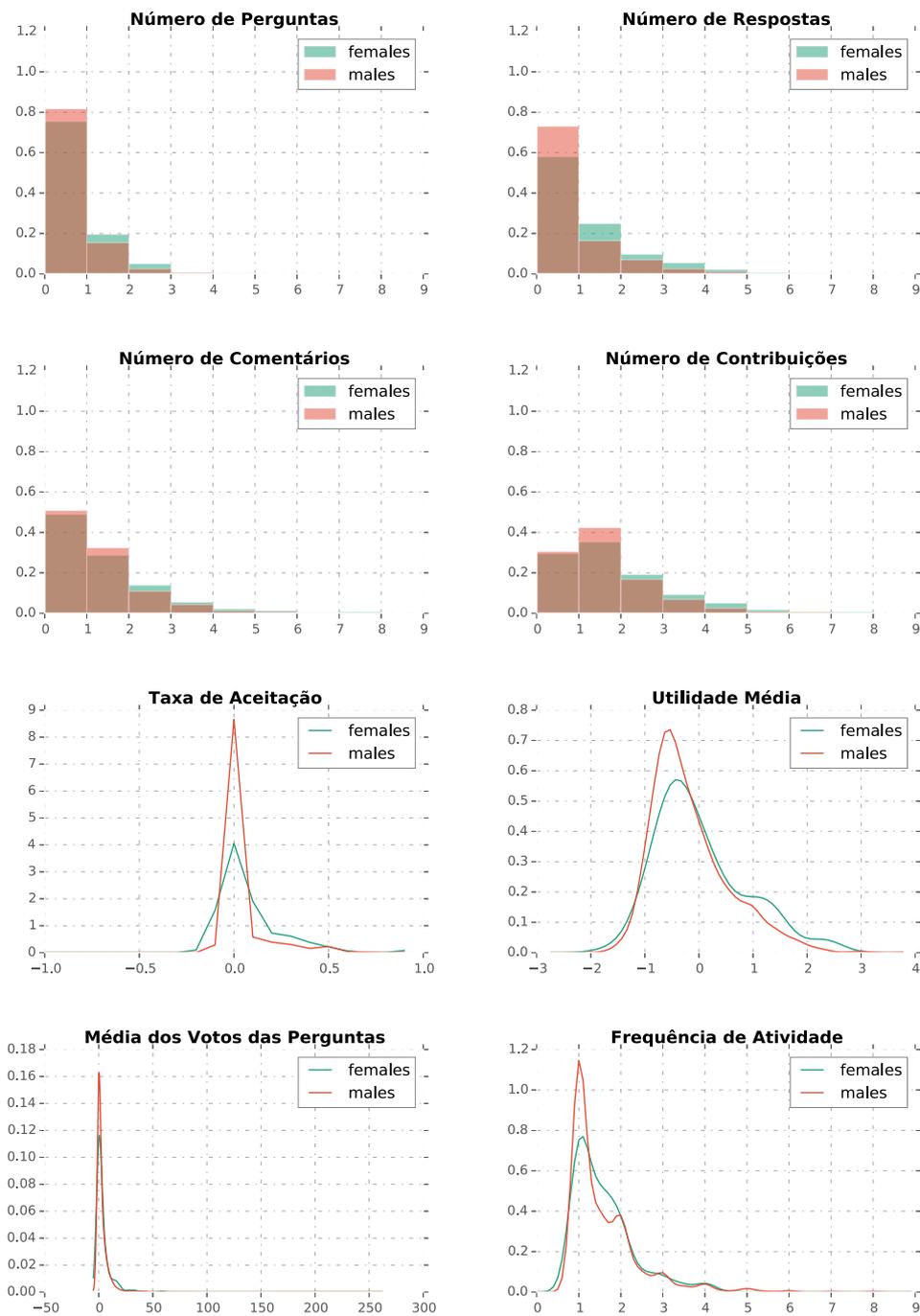


Figura 6.1: Distribuição das variáveis para o site *User Experience*. Mostramos um histograma do log do valor da variável mais um para as variáveis discretas e a gráfico da função de densidade, para as variáveis contínuas.

autoconfiança destas mulheres quanto ao seu conhecimento também é essencial para aumentar estas proporções, já que elas mesmas consideram ter autoconfiança algo essencial para se manter em comunidades nas quais elas são maioria [37].

No geral, nossos resultados mostram que a maioria dos sites do *StackExchange* possui comunidades com poucas diferenças de comportamento relacionadas ao gênero. Porém, conseguimos enxergar claramente duas barreiras para a participação de mulheres nestes sites: uma na entrada, considerando a baixa proporção de contribuidoras e outra relacionada à partida antecipada de mulheres engajadas. Ambas as barreiras são mais expressivas nas maiores comunidades relacionadas à *STEM*, sugerindo que nestas comunidades, mulheres sentem ainda mais dificuldade em participar. Todos esses pensamentos, juntamente com o fato de que mulheres tendem a contribuir com mais perguntas do que homens nestas comunidades, requerem uma investigação qualitativa aprofundada de modo a entender por que este padrão existe.

Capítulo 7

Conclusões

Enquanto estudos de gênero em comunidades de perguntas e respostas costumam focar em número de usuários de cada gênero, esta pesquisa tomou uma abordagem diferente, estudando a participação, qualidade das contribuições e engajamento de usuários de ambos os gêneros. No geral, nossos resultados nos mostram que, individualmente, mulheres que identificam seu gênero, se comprometem aos sites tanto quanto homens. Em vários sites, mulheres contribuem até mais do que homens.

Semelhante à *Wikipedia* [29], mulheres contribuem com conteúdo relevante para sites de perguntas e respostas, mas a grande maioria das contribuições ainda vêm de homens. Nosso estudo nos revela que, depois de passar um tempo contribuindo para os sites que estudamos, mulheres parecem encontrar uma barreira para continuar produzindo conteúdo de valia e continuar engajadas tanto quanto os homens. Não é objetivo do nosso estudo investigar porque há tão poucas mulheres dentro destas comunidades, mas saber que ambientes diversos beneficiam o funcionamento e qualidade de uma comunidade, nos leva a concluir que não incluir ou não motivar mulheres a entrar nestas comunidades pode ser um grande desperdício de contribuidores que podem manter, ou até aumentar, a qualidade dessas comunidades.

Um aspecto, que gostaríamos de destacar, destes resultados é a percepção de quão ativas mulheres podem ser nestas comunidades. Este fato pode ser aproveitado para incentivar a participação de mulheres nestes sites. Apresentar as contribuidoras que encontramos nestas comunidades como modelo de comportamento para outras mulheres pode ser de grande ajuda para aumentar a diversidade de gênero nestas comunidades, visto que mulheres tendem a ter uma boa resposta a este tipo de modelo [41; 36]. Mais ainda, saber que mulheres

conseguem progredir e ter sucesso nestas comunidades pode fazer com que mulheres que apenas observam os sites se identifiquem com as outras e mudem sua percepção de o quão bem podem se adaptar, fazendo com que mudem sua opinião com relação a participar da comunidade [17].

Em resumo, nosso estudo contradiz várias crenças comuns, além algumas pesquisas anteriores relacionadas a gêneros, *STEM* e sites de perguntas e respostas. Isso nos soa como um reforço pertinente que o problema da disparidade de gênero pode aparecer de várias formas e não existe uma solução única para todas as versões do problema. Em nosso caso, observamos que mulheres precisam enfrentar uma barreira para começar a contribuir ou para se identificar como mulher nestas comunidades. Às vezes, uma barreira até em continuar contribuindo uma vez engajada ao site. Contudo, menos esforço parece ser necessário para manter a contribuição proveniente de mulheres já engajadas.

Por fim, é importante notar que o conhecimento preciso das diferenças comportamentais entre homens e mulheres com relação às contribuições para sites de perguntas e respostas não é só relevante para reduzir a disparidade na proporção de mulheres nestas comunidades. Esperamos que este estudo também sirva como arcabouço para iniciar e melhorar estudos de diferenças entre outras minorias e assim aumentar a diversidade de forma geral em comunidades de perguntas e respostas.

7.1 Limitações

Apesar de todas as resoluções as quais chegamos nesta pesquisa, ela não é completa. Uma pesquisa qualitativa soa como um bom começo para complementar as análises que começamos aqui. Outrossim, adicionar mais fatores relacionados à participação pode ajudar a ampliar a visão que temos da situação.

Quanto à inferência de gênero, limitamos nossa amostra àqueles usuários que identificaram a si mesmo com nomes contidos no *Global Name Data*. Esta abordagem pode ter descartado usuários de outros países cujos nomes não são comuns nos registros de nascimento dos Estados Unidos e Reino Unido. Uma pesquisa futura poderia realizar as mesmas comparações que fizemos, mas utilizando um outro método de identificação de gênero, como a foto do perfil do usuário.

Embora nossa amostra contenha sites com uma diversidade significativamente de temas, todos eles pertencem a uma mesma plataforma: o *StackExchange*. Comparar nossos resultados com outros sites que tenham um propósito similar, mas um design diferente pode ajudar a entender melhor a relação entre o design do site, a diversidade de gêneros da comunidade e o comportamento de seus usuários.

Finalmente, nossa abordagem quantitativa e conjunto de dados usado não nos permite investigar profundamente quanto às razões de várias diferenças e similaridades encontradas. Acreditamos ser necessário e oportuno realizar um estudo qualitativo que complementasse nossa iniciativa.

Bibliografia

- [1] About stackexchange. Website. Retrieved August 13, 2015 from <http://stackexchange.com/about>.
- [2] History of wikipedia. Website. Retrieved August 13, 2015 from https://en.wikipedia.org/wiki/History_of_Wikipedia#Historical_overview.
- [3] Wikipedia - stackexchange. Website. Retrieved August 13, 2015 from https://en.wikipedia.org/wiki/Stack_Exchange.
- [4] Wikipedia - stackoverflow. Website. Retrieved August 13, 2015 from https://en.wikipedia.org/wiki/Stack_Overflow.
- [5] Google's workforce diversity report. Website, January 2014. Retrieved April 26, 2015 from <http://www.google.com/diversity/at-google.html>.
- [6] LinkedIn's workforce diversity report. Website, June 2014. Retrieved April 26, 2015 from <http://blog.linkedin.com/2014/06/12/linkedins-workforce-diversity>.
- [7] Stem confidence gap. Website, January 2015. Retrieved May 3, 2015 from <http://blog.piazza.com/stem-confidence-gap/>.
- [8] Judd Antin, Raymond Yee, Coye Cheshire, and Oded Nov. Gender differences in wikipedia editing. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, pages 11–14. ACM, 2011.
- [9] Elizabeth Barrett and Vic Lally. Gender differences in an on-line learning environment. *Journal of computer assisted learning*, 15(1):48–60, 1999.

-
- [10] Lenore Blum, Carol Frieze, Orit Hazzan, and M Bernardine Dias. A cultural perspective on gender diversity in computing. *Reconfiguring the firewall: recruiting women to information technology across cultures and continents*. AK Peters, Ltd, Wellesley, MA, 2007.
- [11] John D Burger, John Henderson, George Kim, and Guido Zarrella. Discriminating gender on twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1301–1309. Association for Computational Linguistics, 2011.
- [12] Nancy K Campbell and Gail Hackett. The effects of mathematics task performance on math self-efficacy and task interest. *Journal of Vocational Behavior*, 28(2):149–162, 1986.
- [13] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. How community feedback shapes user behavior. *arXiv preprint arXiv:1405.1429*, 2014.
- [14] Benjamin Collier and Julia Bear. Conflict, criticism, or confidence: an empirical examination of the gender gap in wikipedia contributions. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 383–392. ACM, 2012.
- [15] Rachel Croson and Uri Gneezy. Gender differences in preferences. *Journal of Economic literature*, pages 448–474, 2009.
- [16] Evandro Cunha, Gabriel Magno, Marcos André Gonçalves, César Cambraia, and Virgilio Almeida. He votes or she votes? female and male discursive strategies in twitter political hashtags. *PloS one*, 9(1):e87041, 2014.
- [17] Joyce Ehrlinger and David Dunning. How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of personality and social psychology*, 84(1):5, 2003.
- [18] Mary F Fox and Glenn Firebaugh. Confidence in science: The gender gap. *Social Science Quarterly*, 1992.

- [19] Adabriand Furtado, Nazareno Andrade, Nigini Oliveira, and Francisco Brasileiro. Contributor profiles, their dynamics, and their importance in five q&a sites. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 1237–1252. ACM, 2013.
- [20] David Gefen and Catherine M. Ridings. If you spoke as she does, sir, instead of the way you do: A sociolinguistics perspective of gender differences in virtual communities. *SIGMIS Database*, 36(2):78–92, June 2005.
- [21] Ruediger Glott, Philipp Schmidt, and Rishab Ghosh. Wikipedia survey–overview of results. *United Nations University: Collaborative Creativity Group*, 2010.
- [22] Uri Gneezy, Muriel Niederle, and Aldo Rustichini. Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3):1049–1074, 2003.
- [23] Eszter Hargittai and Steven Shafer. Differences in actual and perceived online skills: the role of gender*. *Social Science Quarterly*, 87(2):432–448, 2006.
- [24] Janet Shibley Hyde. The gender similarities hypothesis. *American psychologist*, 60(6):581, 2005.
- [25] Janet Shibley Hyde, Elizabeth Fennema, Marilyn Ryan, Laurie A Frost, and Carolyn Hopp. Gender comparisons of mathematics attitudes and affect. *Psychology of women Quarterly*, 14(3):299–324, 1990.
- [26] Adam Hyland. Global name data. <http://bocoup.com/weblog/global-name-data/>, jun 2013. Accessed: 2014-05-28.
- [27] Christopher F Karpowitz, Tali Mendelberg, and Lee Shaker. Gender inequality in deliberative participation. *American Political Science Review*, 106(03):533–547, 2012.
- [28] Bruce Kirkcaldy, Peter Noack, Adrian Furnham, and Georg Siefen. Parental estimates of their own and their children’s intelligence. *European Psychologist*, 12(3):173, 2007.
- [29] Shyong Tony K Lam, Anuradha Uduwage, Zhenhua Dong, Shilad Sen, David R Muscant, Loren Terveen, and John Riedl. Wp: clubhouse?: an exploration of wikipedia’s

- gender imbalance. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, pages 1–10. ACM, 2011.
- [30] Sarah-Jane Leslie, Andrei Cimpian, Meredith Meyer, and Edward Freeland. Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, 347(6219):262–265, 2015.
- [31] Wendy Liu and Derek Ruths. What’s in a name? using first names as features for gender inference in twitter. In *AAAI Spring Symposium: Analyzing Microtext*, 2013.
- [32] Joan E Marshall and Richard Heslin. Boys and girls together: Sexual composition and the effect of density and group size on cohesiveness. *Journal of Personality and Social Psychology*, 31(5):952, 1975.
- [33] Robert Meyer and Michel Cukier. Assessing the attack threat due to irc channels. In *Dependable Systems and Networks, 2006. DSN 2006. International Conference on*, pages 467–472. IEEE, 2006.
- [34] Corinne A Moss-Racusin, John F Dovidio, Victoria L Brescoll, Mark J Graham, and Jo Handelsman. Science faculty’s subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, 109(41):16474–16479, 2012.
- [35] Muriel Niederle and Lise Vesterlund. Do women shy away from competition? do men compete too much? Technical report, National Bureau of Economic Research, 2005.
- [36] Lucia A Nixon and Michael D Robinson. The educational attainment of young women: Role model effects of female high school faculty. *Demography*, 36(2):185–194, 1999.
- [37] WHITNEY E Powell, D SCOTT Hunsinger, and B DAWN Medlin. Gender differences within the open source community: An exploratory study. *Journal of Information Technology*, 21(4):29–37, 2010.
- [38] Karen Rustad. "suck it up, princess": Outreach and diversity in foss communities. 2011.

- [39] Dennis Schenk and Mircea Lungu. Geo-locating the knowledge transfer in stackoverflow. In *Proceedings of the 2013 International Workshop on Social Software Engineering*, pages 21–24. ACM, 2013.
- [40] Casey A Shapiro and Linda J Sax. Major selection and persistence for women in stem. *New Directions for Institutional Research*, 2011(152):5–18, 2011.
- [41] Walter S Smith and Thomas Owen Erb. Effect of women science career role models on early adolescents’ attitudes toward scientists and women in science. *Journal of Research in Science Teaching*, 23(8):667–676, 1986.
- [42] Steven J Spencer, Claude M Steele, and Diane M Quinn. Stereotype threat and women’s math performance. *Journal of experimental social psychology*, 35(1):4–28, 1999.
- [43] Gijsbert Stoet and David C Geary. Sex differences in academic achievement are not related to political, economic, or social equality. *Intelligence*, 48:137–151, 2015.
- [44] Cong Tang, Keith Ross, Nitesh Saxena, and Ruichuan Chen. What’s in a name: A study of names, gender inference, and gender behavior in facebook. In *Database Systems for Adanced Applications*, pages 344–356. Springer, 2011.
- [45] TL Taylor. Multiple pleasures women and online gaming. *Convergence: The International Journal of Research into New Media Technologies*, 9(1):21–46, 2003.
- [46] Joachim Tiedemann. Gender-related beliefs of teachers in elementary school mathematics. *Educational Studies in Mathematics*, 41(2):191–207, 2000.
- [47] Eileen M Trauth, Jeria L Quesenberry, and Allison 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.
- [48] Bogdan Vasilescu, Andrea Capiluppi, and Alexander Serebrenik. Gender, representation and online participation: A quantitative study. *Interacting with Computers*, 2013.
- [49] Allan Wigfield and Jacquelynne S Eccles. Expectancy–value theory of achievement motivation. *Contemporary educational psychology*, 25(1):68–81, 2000.

-
- [50] Jiang Yang, Xiao Wei, Mark S Ackerman, and Lada A Adamic. Activity lifespan: An analysis of user survival patterns in online knowledge sharing communities. In *ICWSM*, 2010.

Apêndice A

Sumário dos usuários identificados

<i>Comunidade</i>	<i>Categoria</i>	<i>Idade</i> <i>(meses)</i>	<i>Número</i> <i>Homens</i>	<i>Número</i> <i>Mulheres</i>	<i>Proporção</i> <i>Identif.</i>
academia	life-arts	33	1124	99	34.41%
android	technology	50	2828	166	35.05%
anime	culture-recreation	23	193	28	27.80%
apple	technology	51	5321	210	38.20%
askubuntu	technology	52	9502	430	33.75%
bicycles	culture-recreation	51	907	38	40.18%
biology	science	35	617	70	36.50%
bitcoin	business	39	815	29	35.63%
chemistry	science	31	400	36	33.72%
chinese	culture-recreation	35	245	22	29.70%
christianity	culture-recreation	39	632	41	40.49%
codegolf	technology	46	1637	63	34.16%
codereview	technology	46	4046	155	35.96%
cogsci	science	34	312	32	38.10%
cooking	life-arts	52	1602	202	39.05%
crypto	technology	40	653	29	35.65%
cs	science	32	969	53	35.96%
cstheory	science	51	1107	51	44.11%
dba	technology	46	3268	127	38.32%

<i>Comunidade</i>	<i>Categoria</i>	<i>Idade</i>	<i>Número</i>	<i>Número</i>	<i>Proporção</i>
		<i>(meses)</i>	<i>Homens</i>	<i>Mulheres</i>	<i>Identif.</i>
diy	life-arts	52	1613	70	41.31%
drupal	technology	44	1372	108	33.19%
dsp	technology	39	460	32	32.50%
electronics	technology	50	2505	92	34.65%
ell	culture-recreation	22	704	67	34.04%
english	culture-recreation	51	5348	441	37.66%
expressionengine	technology	24	358	27	51.96%
fitness	life-arts	44	734	63	37.23%
freelancing	professional	18	159	21	39.22%
french	culture-recreation	39	321	33	34.10%
gamedev	technology	52	2803	128	34.43%
gaming	culture-recreation	52	3671	201	32.28%
gardening	life-arts	41	433	61	40.86%
genealogy	life-arts	25	108	39	52.50%
german	culture-recreation	42	499	44	34.13%
gis	technology	52	1757	159	36.33%
graphicdesign	life-arts	46	1497	110	36.37%
hermeneutics	culture-recreation	37	211	20	43.02%
history	culture-recreation	37	495	41	38.31%
islam	culture-recreation	29	231	21	35.80%
japanese	culture-recreation	42	271	32	30.27%
judaism	culture-recreation	42	374	38	36.69%
linguistics	science	38	327	25	37.13%
magento	technology	22	488	37	39.89%
math	science	52	9568	808	34.57%
mathematica	technology	34	857	63	36.05%
mathoverflow	science	62	3965	283	52.69%
mechanics	culture-recreation	44	575	25	37.90%
money	life-arts	51	1356	81	38.79%
movies	life-arts	36	712	52	34.82%

<i>Comunidade</i>	<i>Categoria</i>	<i>Idade</i>	<i>Número</i>	<i>Número</i>	<i>Proporção</i>
		<i>(meses)</i>	<i>Homens</i>	<i>Mulheres</i>	<i>Identif.</i>
music	life-arts	43	788	52	36.05%
outdoors	culture-recreation	34	269	24	37.76%
parenting	life-arts	44	702	114	41.30%
philosophy	science	41	623	32	37.88%
photo	life-arts	52	1729	99	39.64%
physics	science	48	2992	165	35.79%
pm	business	45	609	40	41.93%
productivity	life-arts	41	619	58	35.90%
programmers	technology	50	9812	415	39.20%
quant	business	46	402	24	35.32%
raspberrypi	technology	29	858	28	35.44%
rpg	culture-recreation	51	889	67	32.92%
russian	culture-recreation	29	168	20	34.75%
salesforce	technology	28	574	41	41.41%
scicomp	science	36	453	23	38.76%
scifi	life-arts	46	2058	216	37.03%
security	technology	48	3065	127	36.14%
serverfault	technology	67	16627	571	37.64%
sharepoint	technology	43	1448	92	42.44%
skeptics	culture-recreation	45	1238	80	38.41%
sound	technology	48	381	21	43.55%
spanish	culture-recreation	36	267	22	37.39%
sqa	technology	42	423	35	44.81%
stackapps	technology	56	509	24	40.63%
stackoverflow	technology	76	149340	9806	35.50%
stats	science	52	2512	200	36.01%
superuser	technology	64	21331	931	36.21%
tex	technology	52	4321	340	36.94%
travel	culture-recreation	41	1240	114	37.64%
unix	technology	51	5086	218	33.42%

<i>Comunidade</i>	<i>Categoria</i>	<i>Idade</i> <i>(meses)</i>	<i>Número</i> <i>Homens</i>	<i>Número</i> <i>Mulheres</i>	<i>Proporção</i> <i>Identif.</i>
ux	technology	51	3324	242	40.09%
webapps	technology	53	2990	170	37.72%
webmasters	technology	52	3042	175	38.88%
wordpress	technology	51	2700	184	38.39%
workplace	professional	31	1441	140	36.72%
writers	life-arts	48	620	70	40.66%

Tabela A.1: Sumário mostrando, para cada comunidade, sua categoria, sua idade, em meses, o número total de homens e mulheres identificados e a proporção de usuários que conseguimos identificar.

Apêndice B

Cinco maiores comunidades

stackoverflow

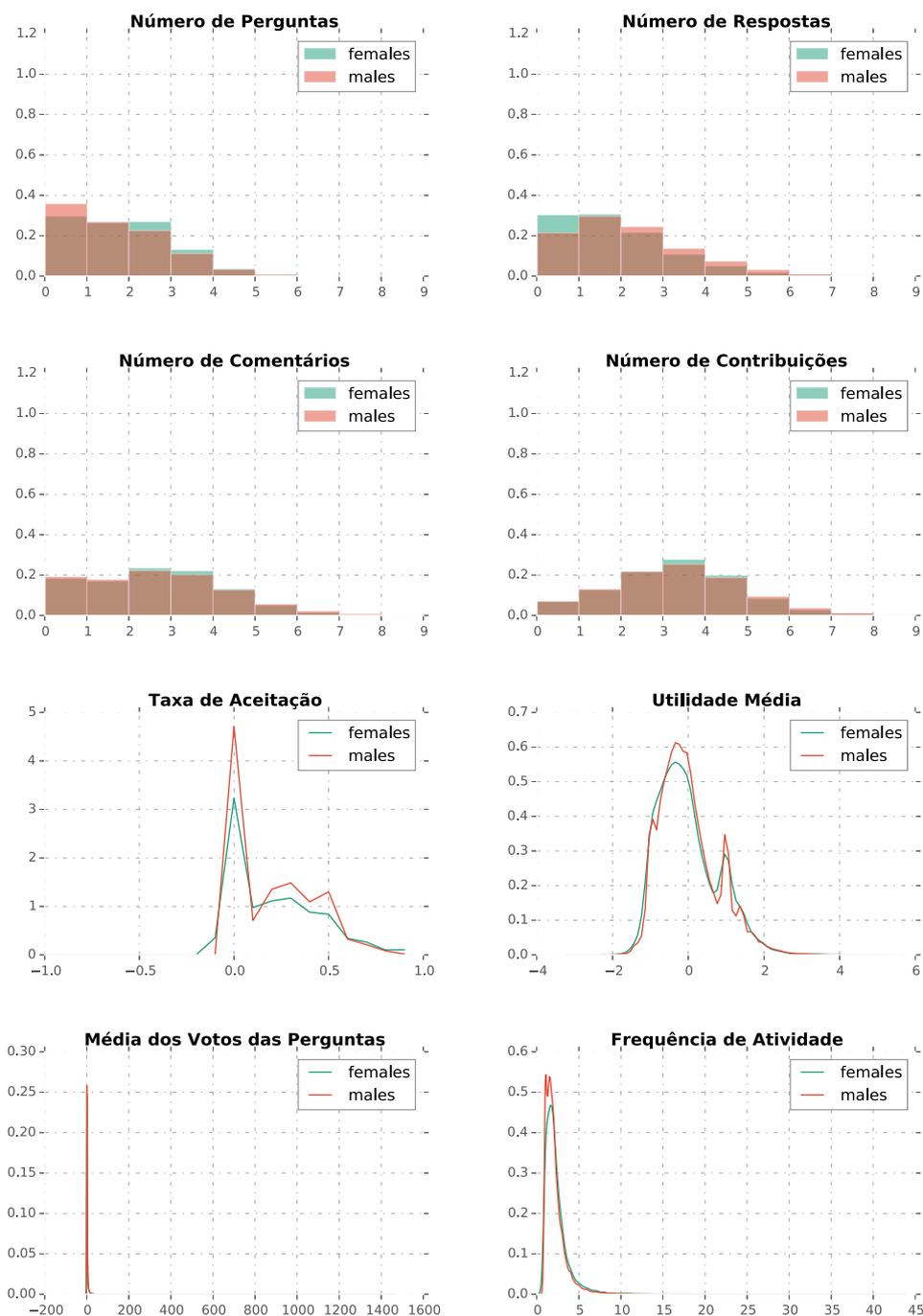


Figura B.1: Distribuição das variáveis onde foi aplicado o teste *Mann-Whitney-U* para o site *StackOverflow*. Mostramos um histograma do log do valor da variável mais um para as variáveis discretas e a gráfico da função de densidade, para as variáveis contínuas.

superuser

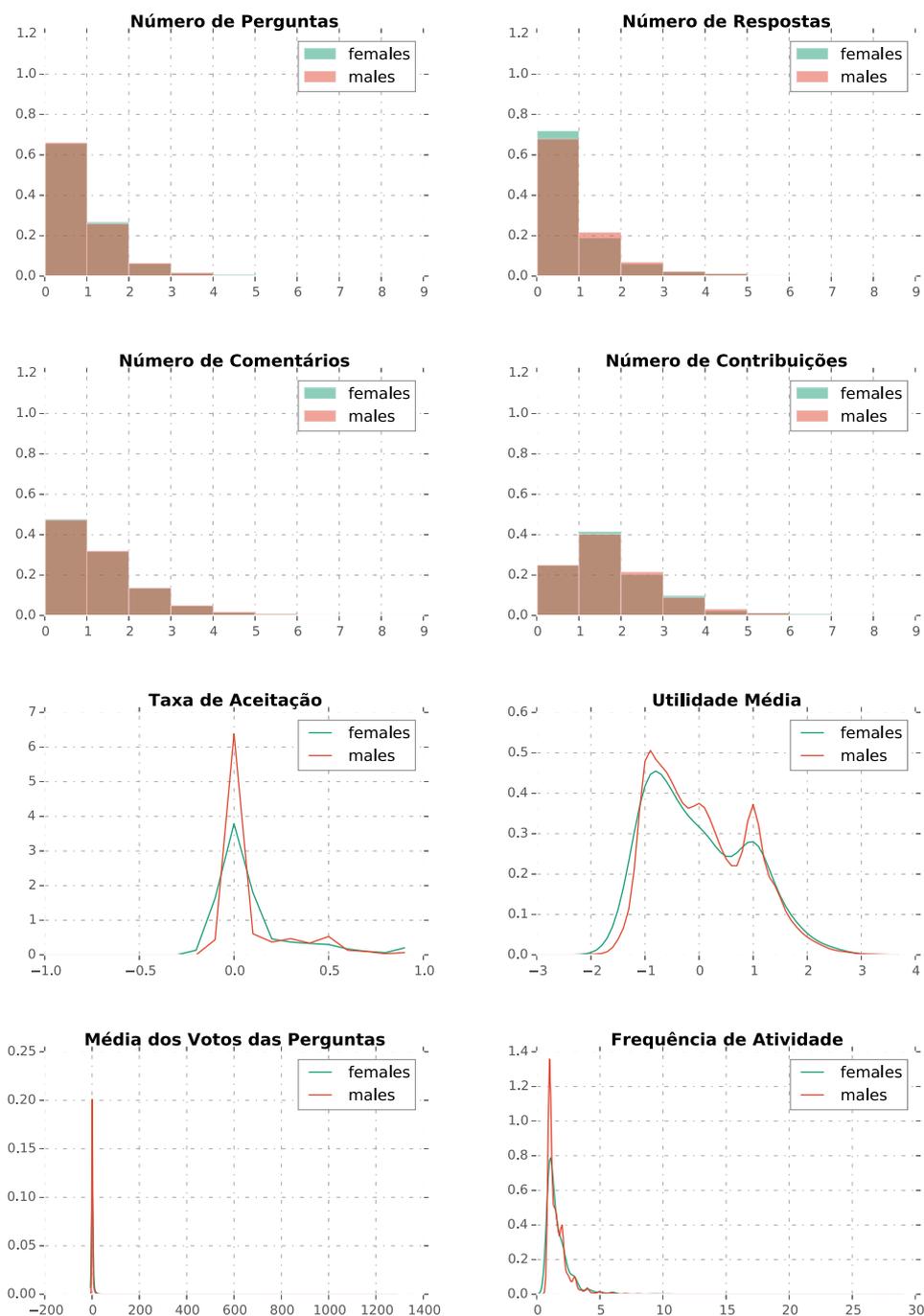


Figura B.2: Distribuição das variáveis onde foi aplicado o teste *Mann-Whitney-U* para o site *SuperUser*. Mostramos um histograma do log do valor da variável mais um para as variáveis discretas e a gráfico da função de densidade, para as variáveis contínuas.

serverfault

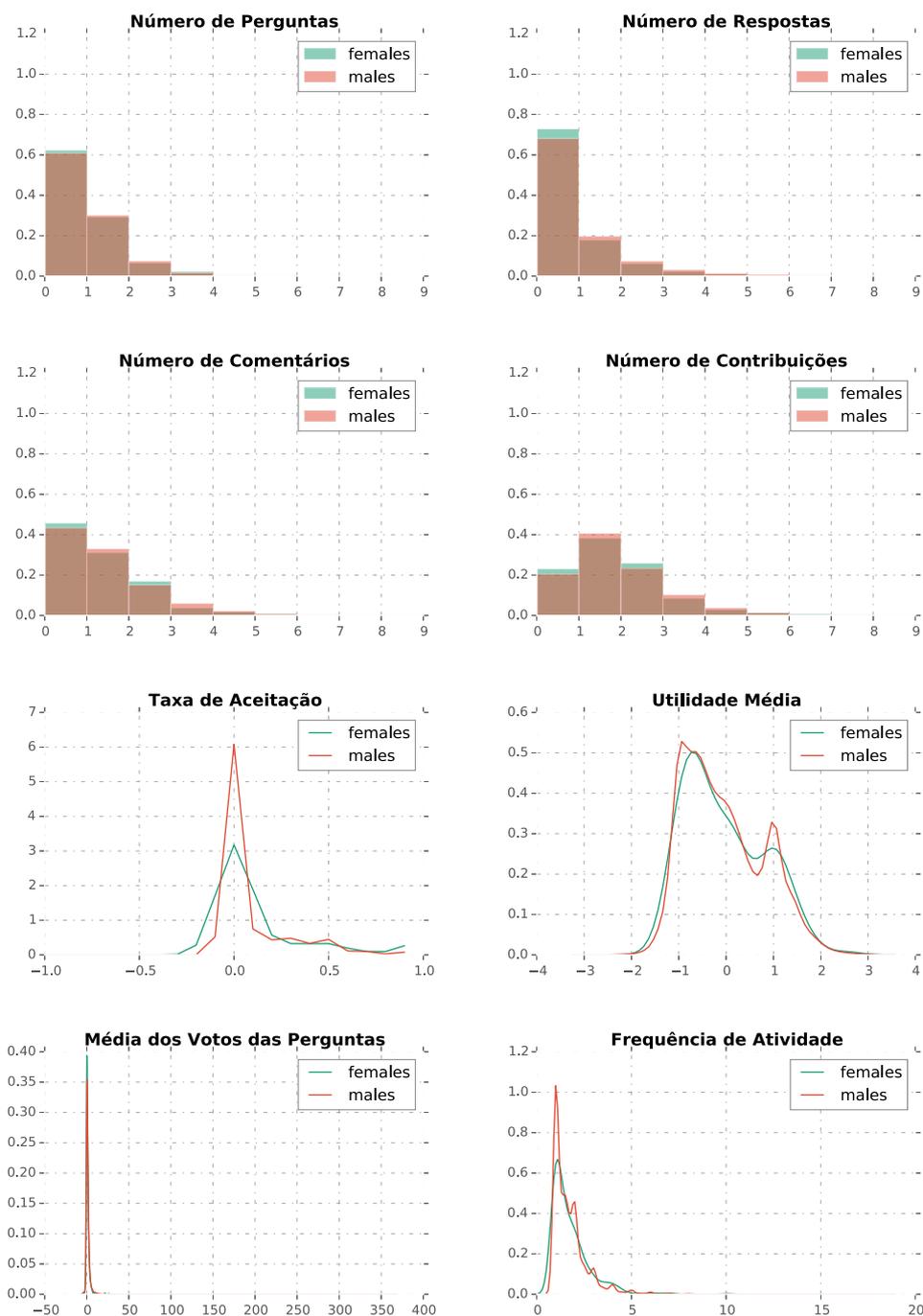


Figura B.3: Distribuição das variáveis onde foi aplicado o teste *Mann-Whitney-U* para o site *ServerFault*. Mostramos um histograma do log do valor da variável mais um para as variáveis discretas e a gráfico da função de densidade, para as variáveis contínuas.

math

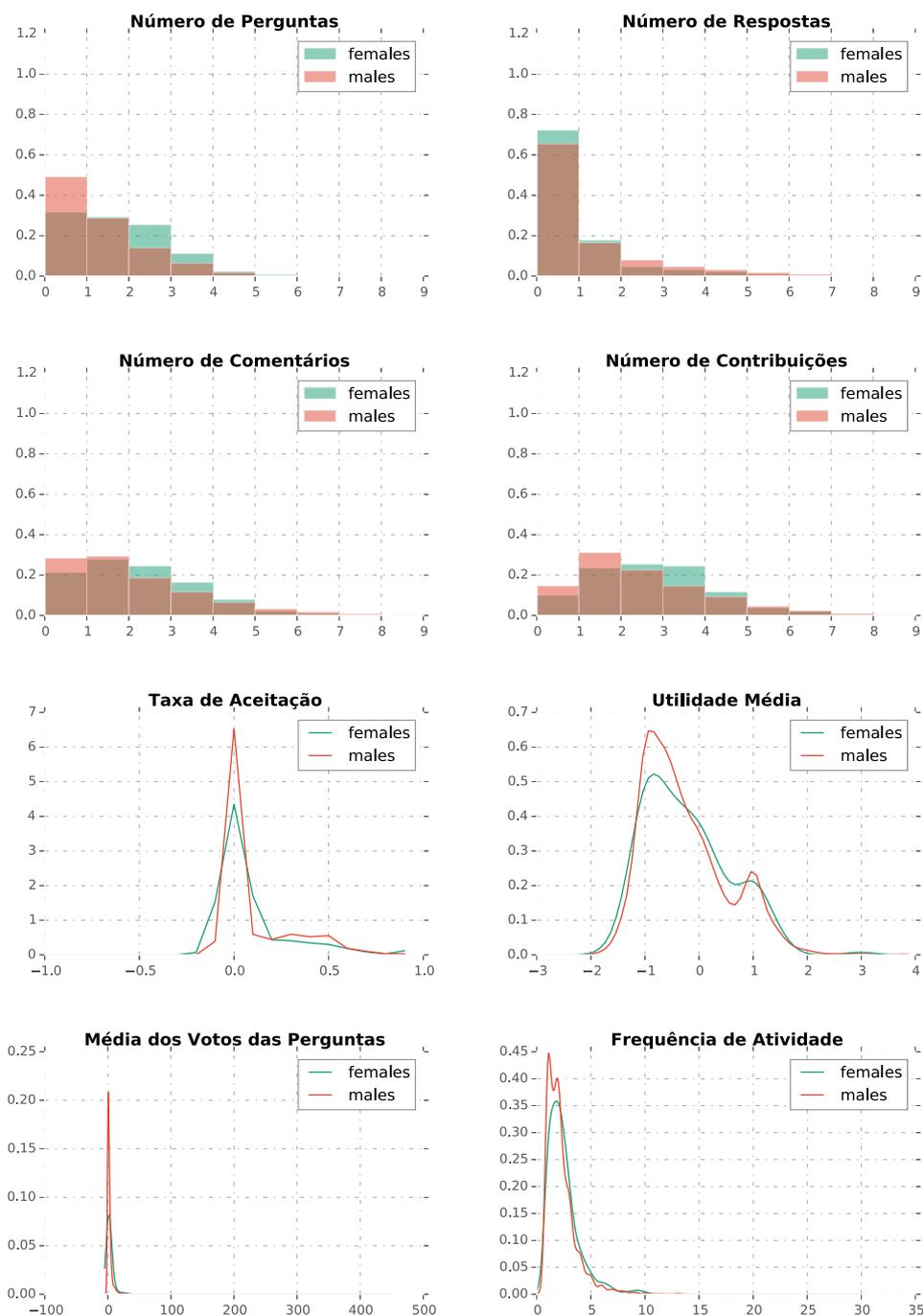


Figura B.4: Distribuição das variáveis onde foi aplicado o teste *Mann-Whitney-U* para o site *Mathematics*. Mostramos um histograma do log do valor da variável mais um para as variáveis discretas e a gráfico da função de densidade, para as variáveis contínuas.

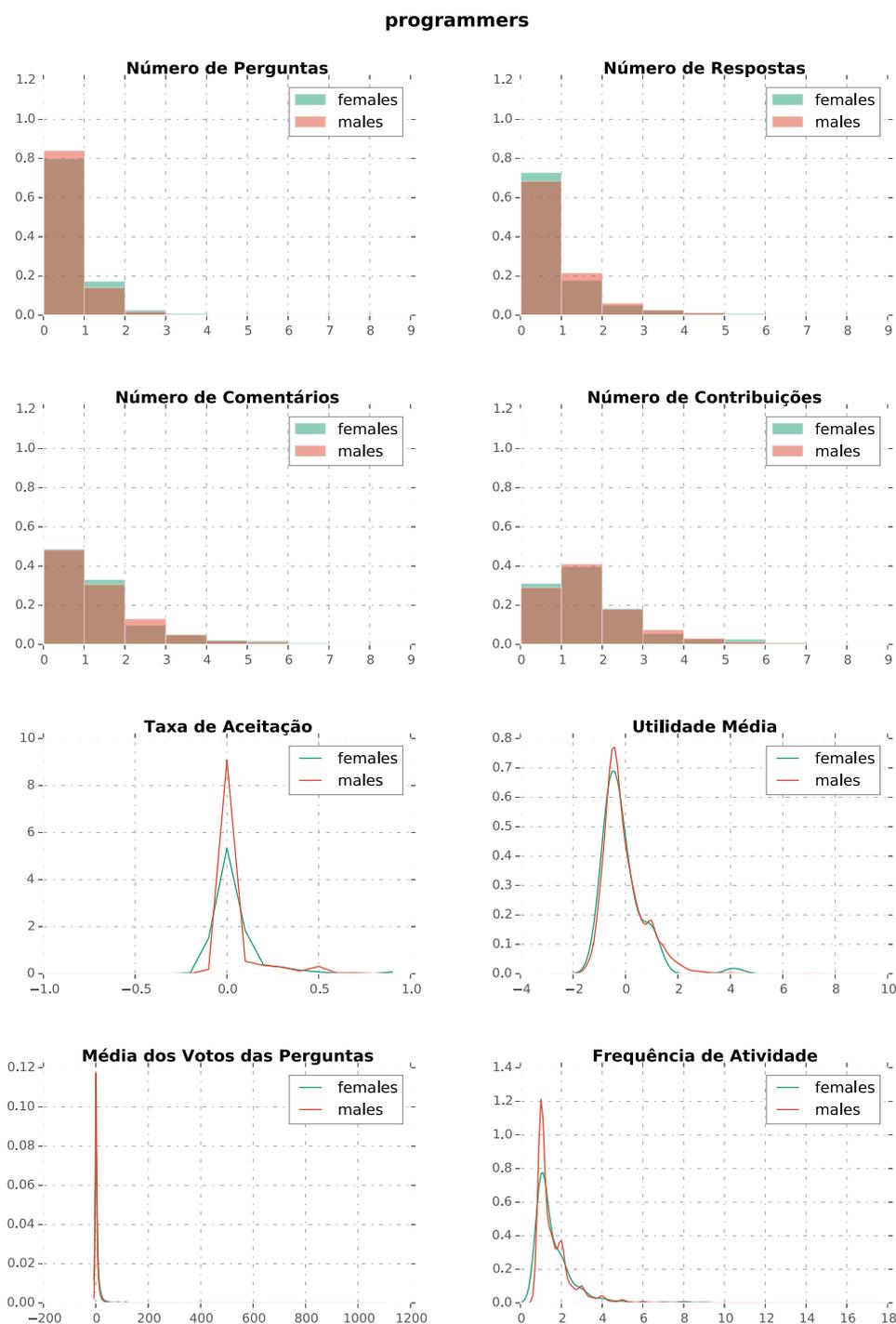


Figura B.5: Distribuição das variáveis onde foi aplicado o teste *Mann-Whitney-U* para o site *Programmers*. Mostramos um histograma do log do valor da variável mais um para as variáveis discretas e a gráfico da função de densidade, para as variáveis contínuas.

Apêndice C

Estatísticas da primeira pergunta de pesquisa

* = Diferença entre a média das mulheres e a média dos homens.

** = Diferença entre a mediana das mulheres e a mediana dos homens.

*** = *P-valor* do teste de hipótese utilizado.

<i>Comunidade</i>	<i>Categoria</i>	<i>Res- pos- tas</i>	<i>Res- pos- tas</i>	<i>Res- postas</i>	<i>Co- mentá- rios</i>	<i>Co- mentá- rios</i>	<i>Co- mentá- rios</i>	<i>Contri- buições</i>	<i>Contri- buições</i>	<i>Contri- buições</i>	<i>Per- guntas</i>	<i>Per- guntas</i>	<i>Per- gun- tas</i>
		*	**	***	*	**	***	*	**	***	*	**	***
academia	life-arts	-1.95	0.00	0.82	-6.58	0.00	0.80	-8.26	0.00	0.55	0.28	1.00	0.13
android	technology	-1.54	0.00	0.80	-3.55	0.00	0.82	-5.18	-1.00	0.66	-0.09	0.00	0.68

<i>Comunidade</i>	<i>Categoria</i>	<i>Res- pos- tas</i>	<i>Res- pos- tas</i>	<i>Res- postas</i>	<i>Co- mentá- rios</i>	<i>Co- mentá- rios</i>	<i>Co- mentá- rios</i>	<i>Contri- buições</i>	<i>Contri- buições</i>	<i>Contri- buições</i>	<i>Per- guntas</i>	<i>Per- guntas</i>	<i>Per- guntas</i>
		*	**	***	*	**	***	*	**	***	*	**	***
anime	culture-recreation	3.50	1.00	0.03	13.34	1.00	0.45	17.34	1.00	0.30	0.50	-1.00	0.56
apple	technology	-1.22	0.00	0.04	-2.75	0.00	0.07	-4.44	-0.50	0.06	-0.47	0.00	0.41
askubuntu	technology	-0.34	0.00	0.61	-0.88	0.00	0.83	-0.60	1.00	0.20	0.63	0.00	0.00
bicycles	culture-recreation	-3.73	0.00	0.10	-10.44	1.00	0.80	-14.63	0.00	0.76	-0.47	0.00	0.67
biology	science	-0.39	0.00	0.03	-1.76	0.00	0.11	-2.46	-0.50	0.69	-0.31	0.00	0.44
bitcoin	business	-2.70	1.00	0.50	-4.15	0.00	0.12	-7.16	-1.00	0.64	-0.31	0.00	0.36
chemistry	science	0.26	0.00	0.36	-2.87	1.00	0.29	-1.69	1.00	0.18	0.91	0.00	0.32
chinese	culture-recreation	1.72	0.00	0.25	-0.35	0.50	0.42	0.98	0.00	0.35	-0.39	0.00	0.42
christianity	culture-recreation	-6.64	0.00	0.52	-16.49	-1.00	0.13	-24.27	-1.00	0.45	-1.14	0.00	0.21
codegolf	technology	-1.69	0.00	0.35	-5.28	1.00	0.12	-6.93	1.00	0.18	0.04	0.00	0.10
codereview	technology	-0.03	0.00	0.18	-0.81	-1.00	0.95	-0.82	0.00	0.66	0.02	1.00	0.19
cogsci	science	-0.73	1.00	0.19	-4.19	0.00	0.44	-5.84	0.00	0.88	-0.92	0.00	0.67
cooking	life-arts	2.75	1.00	0.00	4.09	0.50	0.24	7.77	1.50	0.00	0.93	0.00	0.04
crypto	technology	-2.72	0.00	0.95	-8.80	0.00	0.70	-10.60	0.00	0.78	0.93	0.00	0.02
cs	science	-3.13	0.00	0.07	-10.94	-1.00	0.68	-13.78	1.00	0.68	0.29	0.00	0.01
cstheory	science	-2.46	-1.00	0.09	-12.62	-1.00	0.13	-15.67	-1.00	0.22	-0.59	0.00	0.87

<i>Comunidade</i>	<i>Categoria</i>	<i>Res- pos- tas</i>	<i>Res- pos- tas</i>	<i>Res- postas</i>	<i>Co- mentá- rios</i>	<i>Co- mentá- rios</i>	<i>Co- mentá- rios</i>	<i>Contri- buições</i>	<i>Contri- buições</i>	<i>Contri- buições</i>	<i>Per- guntas</i>	<i>Per- guntas</i>	<i>Per- guntas</i>
		*	**	***	*	**	***	*	**	***	*	**	***
dba	technology	0.01	0.00	0.49	-1.29	0.00	0.54	-1.21	1.00	0.31	0.07	0.00	0.53
diy	life-arts	-3.38	0.00	0.51	-6.07	-1.00	0.11	-10.04	0.00	0.10	-0.60	0.00	0.13
drupal	technology	-2.24	0.00	0.63	-3.89	0.00	0.89	-6.20	2.50	0.55	-0.07	1.00	0.77
dsp	technology	-5.11	0.00	0.00	-11.05	0.00	0.52	-16.14	-1.00	0.27	0.03	0.00	0.11
electronics	technology	-7.01	0.00	0.74	-14.41	-1.00	0.35	-20.59	-1.00	0.66	0.83	0.00	0.17
ell	culture-recreation	-0.78	0.00	0.04	-2.33	1.00	0.34	-1.45	2.00	0.04	1.66	0.00	0.81
english	culture-recreation	0.06	1.00	0.00	-5.47	-1.00	0.04	-5.78	1.00	0.23	-0.37	0.00	0.23
expressionengine	technology	-1.44	1.00	0.63	10.57	2.00	0.14	11.45	8.00	0.11	2.32	2.00	0.06
fitness	life-arts	1.17	1.00	0.56	0.35	0.00	0.55	2.03	0.00	0.68	0.51	0.00	1.00
freelancing	professional	1.09	0.00	0.02	-0.58	0.00	0.76	0.09	0.00	0.36	-0.41	0.00	0.20
french	culture-recreation	6.89	0.00	0.08	14.84	0.00	0.99	21.27	2.00	0.33	-0.46	-1.00	0.86
gamedev	technology	-0.62	0.00	0.28	2.17	0.00	0.82	1.82	0.00	0.48	0.27	0.00	0.15
gaming	culture-recreation	8.70	0.00	0.19	22.65	0.00	0.24	35.30	0.00	0.95	3.95	0.00	0.89
gardening	life-arts	0.38	1.00	0.00	-2.05	0.00	0.30	-1.75	-1.00	0.83	-0.08	0.00	0.00
genealogy	life-arts	-0.97	0.00	0.21	-4.50	-1.00	0.36	-6.32	-1.00	0.89	-0.86	-1.00	0.16
german	culture-recreation	-2.33	0.00	0.46	-11.48	-1.00	0.12	-14.95	0.00	0.92	-1.13	0.00	0.04

<i>Comunidade</i>	<i>Categoria</i>	<i>Res- pos- tas</i>	<i>Res- pos- tas</i>	<i>Res- postas</i>	<i>Co- mentá- rios</i>	<i>Co- mentá- rios</i>	<i>Co- mentá- rios</i>	<i>Contri- buições</i>	<i>Contri- buições</i>	<i>Contri- buições</i>	<i>Per- guntas</i>	<i>Per- guntas</i>	<i>Per- guntas</i>
		*	**	***	*	**	***	*	**	***	*	**	***
gis	technology	-5.23	0.00	0.20	-4.85	1.00	0.02	-8.78	2.00	0.04	1.30	1.00	0.00
graphicdesign	life-arts	4.17	1.00	0.00	8.16	0.00	0.63	12.69	1.00	0.28	0.37	0.00	0.84
hermeneutics	culture-recreation	-5.68	0.00	0.41	-2.58	0.50	0.88	-7.88	1.50	0.90	0.62	-1.00	0.81
history	culture-recreation	-1.44	1.00	0.04	-12.69	0.00	0.99	-13.79	1.00	0.58	0.33	-1.00	0.86
islam	culture-recreation	-0.58	0.00	0.65	-6.51	-1.00	0.63	-6.68	1.00	0.99	0.41	0.00	0.18
japanese	culture-recreation	-5.10	-1.00	0.15	-21.60	-1.00	0.03	-28.73	-1.50	0.05	-2.04	0.00	0.51
judaism	culture-recreation	-	0.00	0.17	-23.30	0.00	0.43	-40.44	-0.50	0.52	-1.28	0.00	0.74
		15.86											
linguistics	science	-2.55	0.00	0.17	-11.25	-1.00	0.31	-14.16	-1.00	0.12	-0.37	0.00	0.69
magento	technology	-4.10	1.00	0.34	-5.48	0.00	0.56	-7.47	1.00	0.54	2.11	0.00	0.29
math	science	-	-1.00	0.00	-34.13	2.00	0.00	-45.93	6.00	0.00	3.29	2.00	0.00
		15.08											
mathematica	technology	-2.61	0.00	0.40	-14.34	2.00	0.36	-16.77	1.00	0.36	0.19	0.00	0.13
mathoverflow	science	-7.99	-1.00	0.00	-26.43	-1.00	0.00	-35.75	-3.00	0.00	-1.33	0.00	0.14
mechanics	culture-recreation	-3.85	0.00	0.06	-5.90	-1.00	0.11	-10.25	-1.00	0.03	-0.50	0.00	0.13
money	life-arts	-0.97	1.00	0.32	-2.97	0.00	0.08	-4.42	-1.00	0.26	-0.48	0.00	0.07
movies	life-arts	-2.92	0.50	0.91	-8.97	0.00	0.06	-12.67	0.00	0.45	-0.78	0.00	0.32

<i>Comunidade</i>	<i>Categoria</i>	<i>Res- pos- tas</i>	<i>Res- pos- tas</i>	<i>Res- postas ***</i>	<i>Co- mentá- rios</i>	<i>Co- mentá- rios</i>	<i>Co- mentá- rios</i>	<i>Contri- buições</i>	<i>Contri- buições</i>	<i>Contri- buições</i>	<i>Per- guntas</i>	<i>Per- guntas</i>	<i>Per- guntas</i>	
		*	**		*	**	***	*	**	***	*	**	***	
music	life-arts	-0.76	0.50	0.01	-2.50	0.50	0.13	-3.19	1.50	0.04	0.07	0.00	0.35	
outdoors	culture-recreation	-0.24	0.00	0.31	-4.60	0.00	0.08	-5.74	-1.00	0.48	-0.91	0.00	0.35	
parenting	life-arts	6.56	0.00	0.00	5.10	0.00	0.57	11.87	0.00	0.03	0.21	0.00	0.12	
philosophy	science	-3.13	-1.00	0.49	-12.62	-1.00	0.14	-16.20	-0.50	0.38	-0.45	0.50	0.99	
photo	life-arts	-4.85	-1.00	0.29	-9.64	0.00	0.10	-14.61	0.00	0.21	-0.13	0.00	0.76	
physics	science	-4.39	0.00	0.05	-14.62	0.00	0.18	-18.38	1.00	0.91	0.63	0.00	0.00	
pm	business	-2.73	0.00	0.91	-2.89	0.50	0.40	-5.53	2.00	0.19	0.09	0.50	0.52	
productivity	life-arts	1.84	1.00	0.01	0.78	0.00	0.50	2.40	1.00	0.05	-0.22	0.00	0.41	
programmers	technology	-0.37	0.00	0.02	0.52	0.00	0.26	0.54	0.00	0.23	0.40	1.00	0.00	
quant	business	-2.37	-1.00	0.19	-2.68	0.50	0.36	-3.27	2.00	0.26	1.77	0.00	0.04	
raspberrypi	technology	0.67	0.00	0.22	1.20	1.00	0.87	2.50	1.00	0.32	0.64	0.00	0.02	
rpg	culture-recreation	-2.67	1.00	0.53	-7.90	-1.00	0.24	-11.27	0.00	0.84	-0.71	0.00	0.63	
russian	culture-recreation	1.30	0.00	0.90	12.02	-0.50	0.54	15.29	-1.00	0.86	1.97	0.00	0.06	
salesforce	technology	-	0.00	0.73	-16.68	3.00	0.84	-24.71	3.50	0.83	2.19	0.00	0.66	
scicomp	science	10.22	-6.15	-1.00	0.08	-13.05	-1.00	0.13	-18.62	-1.00	0.14	0.58	0.00	0.29

<i>Comunidade</i>	<i>Categoria</i>	<i>Res- pos- tas</i>	<i>Res- pos- tas</i>	<i>Res- postas</i>	<i>Co- mentá- rios</i>	<i>Co- mentá- rios</i>	<i>Co- mentá- rios</i>	<i>Contri- buições</i>	<i>Contri- buições</i>	<i>Contri- buições</i>	<i>Per- guntas</i>	<i>Per- guntas</i>	<i>Per- guntas</i>
		*	**	***	*	**	***	*	**	***	*	**	***
scifi	life-arts	-1.38	0.00	0.07	-4.07	0.00	0.00	-4.41	0.00	0.06	1.03	1.00	0.00
security	technology	-3.19	0.00	0.27	-3.90	1.00	0.77	-6.71	0.00	0.96	0.38	0.00	0.05
serverfault	technology	-0.76	0.00	0.04	0.31	0.00	0.27	-0.08	0.00	0.17	0.37	0.00	0.60
sharepoint	technology	5.99	2.00	0.03	3.71	4.50	0.01	10.95	7.00	0.00	1.26	0.50	0.04
skeptics	culture-recreation	0.53	0.00	0.06	0.65	-1.00	0.55	2.16	0.00	0.50	0.98	1.00	0.01
sound	technology	-1.31	-3.00	0.31	46.14	-2.00	0.25	55.65	-5.00	0.25	10.81	0.00	0.94
spanish	culture-recreation	0.42	0.00	0.79	1.42	-1.00	0.69	0.90	-1.50	0.50	-0.95	-1.00	0.12
sqa	technology	6.60	1.00	0.01	8.79	1.00	0.46	15.39	1.00	0.19	0.00	0.00	0.40
stackapps	technology	-1.40	0.00	0.59	-4.55	0.00	0.39	-6.29	0.00	0.34	-0.34	0.00	0.71
stackoverflow	technology	-	-2.00	0.00	-15.73	1.00	0.52	-26.13	0.00	0.15	2.13	1.00	0.00
stats	science	12.52	-3.00	0.00	-3.14	1.00	0.02	-5.36	2.50	0.01	0.77	1.00	0.00
superuser	technology	-1.10	0.00	0.00	-1.97	0.00	0.20	-2.91	0.00	0.48	0.16	0.00	0.20
tex	technology	-6.38	0.00	0.01	-14.78	1.00	0.39	-21.03	0.50	0.13	0.13	0.00	0.00
travel	culture-recreation	-0.09	1.00	0.00	-4.85	0.00	0.00	-5.73	0.00	0.77	-0.78	0.00	0.15
unix	technology	-2.34	-1.00	0.04	-5.19	-0.50	0.32	-7.78	0.00	0.76	-0.25	0.00	0.02

<i>Comunidade</i>	<i>Categoria</i>	<i>Res- pos- tas</i>	<i>Res- pos- tas</i>	<i>Res- postas</i>	<i>Co- mentá- rios</i>	<i>Co- mentá- rios</i>	<i>Co- mentá- rios</i>	<i>Contri- buições</i>	<i>Contri- buições</i>	<i>Contri- buições</i>	<i>Per- guntas</i>	<i>Per- guntas</i>	<i>Per- guntas</i>
		*	**	***	*	**	***	*	**	***	*	**	***
ux	technology	2.18	0.00	0.00	5.40	1.00	0.70	7.89	1.00	0.02	0.32	1.00	0.11
webapps	technology	-0.17	0.00	0.73	-1.27	0.00	0.03	-1.63	0.00	0.63	-0.19	0.00	0.56
webmasters	technology	-0.92	0.00	0.40	-0.77	0.00	0.73	-1.03	0.00	0.27	0.66	0.00	0.01
wordpress	technology	-0.97	0.00	0.13	1.45	1.00	0.01	1.60	3.00	0.01	1.12	1.00	0.01
workplace	professional	4.31	1.00	0.01	8.46	0.00	0.69	13.18	1.00	0.06	0.41	0.00	0.01
writers	life-arts	11.21	0.00	0.00	20.89	0.00	0.86	31.73	0.00	0.19	-0.37	0.00	0.52

Tabela C.1: Para cada comunidade apresentamos nesta tabela sua categoria, a diferença entre a média de mulheres e homens quanto ao número de respostas / comentários / contribuições / perguntas realizadas por usuários de cada gênero; a diferença entre a mediana de mulheres e homens quanto ao número de respostas / comentários / contribuições / perguntas realizadas por usuários de cada gênero; e o *p-valor* obtido utilizando o Mann-Whitney-U test, para cada tipo de contribuição, para cada gênero.

Apêndice D

Estatísticas da segunda pergunta de pesquisa

67

* = Diferença entre a média das mulheres e a média dos homens.

** = Diferença entre a mediana das mulheres e a mediana dos homens.

*** = *P-valor* do teste de hipótese utilizado.

<i>Comunidade</i>	<i>Categoria</i>	<i>Taxa</i>	<i>Taxa</i>	<i>Taxa</i>	<i>Utilidade</i>	<i>Utilidade</i>	<i>Utilidade</i>	<i>MVP *</i>	<i>MVP **</i>	<i>MVP</i>
		<i>Aceit. *</i>	<i>Aceit. **</i>	<i>Aceit. ***</i>	<i>Média *</i>	<i>Média **</i>	<i>Média ***</i>			<i>***</i>
academia	life-arts	0.03	0.00	0.43	0.11	0.12	0.37	0.52	-0.33	0.85
android	technology	0.01	0.00	0.81	-0.06	-0.35	0.51	0.30	0.00	0.96
anime	culture-recreation	0.21	0.50	0.02	0.29	0.16	0.32	1.46	0.63	0.52
apple	technology	-0.07	0.00	0.08	0.00	-0.04	0.96	-0.29	-0.42	0.33

<i>Comunidade</i>	<i>Categoria</i>	<i>Taxa</i>	<i>Taxa</i>	<i>Taxa</i>	<i>Utilidade</i>	<i>Utilidade</i>	<i>Utilidade</i>	<i>MVP *</i>	<i>MVP **</i>	<i>MVP</i>
		<i>Aceit. *</i>	<i>Aceit. **</i>	<i>Aceit. ***</i>	<i>Média *</i>	<i>Média **</i>	<i>Média ***</i>			<i>***</i>
askubuntu	technology	-0.01	0.05	0.98	0.07	0.07	0.22	0.02	-0.17	0.46
bicycles	culture-recreation	-0.05	0.00	0.07	-0.14	-0.28	0.24	3.27	2.75	0.01
biology	science	0.06	0.08	0.35	0.13	0.34	0.49	0.72	0.17	0.38
bitcoin	business	0.14	0.33	0.12	0.02	0.17	0.94	-2.97	-1.00	0.00
chemistry	science	0.03	-0.01	0.71	-0.19	-0.12	0.43	0.23	-0.15	0.46
chinese	culture-recreation	0.02	0.00	0.81	0.17	0.13	0.52	-0.65	0.00	0.50
christianity	culture-recreation	-0.04	0.00	0.12	0.06	0.16	0.64	2.18	-0.61	0.82
codegolf	technology	0.01	0.00	0.69	-0.07	0.00	0.79	-5.68	-3.50	0.08
codereview	technology	0.02	0.00	0.52	-0.10	-0.23	0.57	0.62	1.00	0.21
cogsci	science	-0.05	-0.14	0.67	0.28	0.54	0.22	3.26	1.98	0.02
cooking	life-arts	0.03	0.00	0.02	0.14	0.20	0.03	-0.02	0.33	0.75
crypto	technology	0.02	0.00	0.84	-0.10	-0.21	0.70	-1.19	-0.89	0.11
cs	science	0.10	0.25	0.32	-0.09	-0.33	0.68	-1.41	0.00	0.08
cstheory	science	0.10	0.25	0.26	-0.02	0.06	0.98	-1.10	-0.34	0.36
dba	technology	-0.09	-0.04	0.30	-0.32	-0.36	0.04	-0.59	-0.20	0.23
diy	life-arts	-0.03	0.00	0.19	0.14	0.04	0.59	0.49	0.00	0.35
drupal	technology	-0.04	-0.04	0.40	0.04	0.09	0.83	0.09	0.00	0.61
dsp	technology	-0.05	-0.20	0.56	0.21	0.08	0.54	-0.91	-0.75	0.42

<i>Comunidade</i>	<i>Categoria</i>	<i>Taxa</i>	<i>Taxa</i>	<i>Taxa</i>	<i>Utilidade</i>	<i>Utilidade</i>	<i>Utilidade</i>	<i>MVP *</i>	<i>MVP **</i>	<i>MVP</i>
		<i>Aceit. *</i>	<i>Aceit. **</i>	<i>Aceit. ***</i>	<i>Média *</i>	<i>Média **</i>	<i>Média ***</i>			<i>***</i>
electronics	technology	-0.10	-0.10	0.05	0.06	0.06	0.86	0.17	0.00	0.75
ell	culture-recreation	0.04	0.11	0.30	-0.15	0.00	0.90	0.69	0.72	0.29
english	culture-recreation	-0.01	0.00	0.39	0.11	0.12	0.04	1.05	0.00	0.55
expressionengine	technology	0.00	-0.09	0.64	-0.31	-0.18	0.15	-0.55	-0.07	0.40
fitness	life-arts	0.02	0.00	0.58	0.24	0.49	0.16	1.68	1.67	0.03
freelancing	professional	0.09	0.00	0.07	0.05	-0.03	0.64	-2.25	-2.00	0.17
french	culture-recreation	0.11	0.25	0.09	0.29	0.44	0.11	0.64	1.50	0.45
gamedev	technology	0.03	0.00	0.68	0.14	0.37	0.15	-0.93	-0.75	0.17
gaming	culture-recreation	0.00	0.02	0.84	0.17	0.29	0.04	1.20	1.00	0.00
gardening	life-arts	-0.03	0.00	0.68	0.05	0.01	0.71	0.32	-1.00	0.65
genealogy	life-arts	0.02	0.00	0.71	0.35	0.30	0.09	-0.35	0.00	0.86
german	culture-recreation	-0.09	0.00	0.41	-0.07	-0.11	0.63	-0.80	-1.00	0.38
gis	technology	0.00	-0.01	0.87	-0.03	-0.03	0.95	-0.04	-0.36	0.12
graphicdesign	life-arts	-0.02	0.00	0.95	0.01	0.18	0.70	1.58	0.00	0.12
hermeneutics	culture-recreation	0.04	0.00	0.99	-0.09	0.06	0.69	0.60	1.02	0.46
history	culture-recreation	0.05	0.17	0.22	0.24	0.46	0.12	1.31	2.26	0.04
islam	culture-recreation	0.03	0.14	0.34	-0.37	-0.42	0.15	0.87	-0.23	0.48
japanese	culture-recreation	0.10	0.01	0.56	0.31	0.69	0.47	-1.12	-1.00	0.23

<i>Comunidade</i>	<i>Categoria</i>	<i>Taxa</i>	<i>Taxa</i>	<i>Taxa</i>	<i>Utilidade</i>	<i>Utilidade</i>	<i>Utilidade</i>	<i>MVP *</i>	<i>MVP **</i>	<i>MVP</i>
		<i>Aceit. *</i>	<i>Aceit. **</i>	<i>Aceit. ***</i>	<i>Média *</i>	<i>Média **</i>	<i>Média ***</i>			<i>***</i>
judiasm	culture-recreation	0.03	-0.01	0.88	0.52	0.62	0.02	0.81	-0.42	0.60
linguistics	science	-0.09	0.00	0.19	0.10	0.15	0.70	-0.94	-0.44	0.70
magento	technology	0.11	0.12	0.09	-0.16	-0.05	0.66	0.12	0.05	0.89
math	science	0.00	-0.05	0.46	0.04	0.11	0.53	0.47	0.00	0.45
mathematica	technology	-0.01	0.08	0.95	-0.26	-0.30	0.38	0.24	-0.13	0.41
mathoverflow	science	0.02	-0.17	0.39	0.15	0.05	0.10	0.33	0.42	0.31
mechanics	culture-recreation	-0.27	-0.22	0.13	-0.39	-0.22	0.56	1.62	0.33	0.99
money	life-arts	0.08	0.00	0.14	0.11	0.18	0.58	1.77	0.96	0.00
movies	life-arts	0.09	0.00	0.57	0.64	1.02	0.00	1.70	0.63	0.40
music	life-arts	0.01	0.00	0.46	0.11	-0.13	0.59	0.22	0.71	0.44
outdoors	culture-recreation	0.10	0.00	0.39	0.16	0.10	0.48	2.36	3.67	0.22
parenting	life-arts	0.03	0.00	0.04	0.26	0.33	0.00	0.98	1.74	0.08
philosophy	science	0.00	0.00	0.93	0.10	-0.12	0.70	4.55	-1.00	0.33
photo	life-arts	0.07	0.00	0.26	-0.09	-0.08	0.30	0.96	1.00	0.03
physics	science	-0.01	-0.06	0.70	-0.22	-0.30	0.09	-0.14	-0.38	0.35
pm	business	0.05	0.00	0.85	0.16	0.29	0.27	1.01	2.00	0.12
productivity	life-arts	-0.05	0.00	0.41	-0.13	-0.10	0.35	0.27	2.00	0.65
programmers	technology	-0.01	0.00	1.00	-0.07	-0.04	0.13	1.29	1.00	0.34

<i>Comunidade</i>	<i>Categoria</i>	<i>Taxa</i>	<i>Taxa</i>	<i>Taxa</i>	<i>Utilidade</i>	<i>Utilidade</i>	<i>Utilidade</i>	<i>MVP *</i>	<i>MVP **</i>	<i>MVP</i>
		<i>Aceit. *</i>	<i>Aceit. **</i>	<i>Aceit. ***</i>	<i>Média *</i>	<i>Média **</i>	<i>Média ***</i>			<i>***</i>
quant	business	0.15	0.25	0.16	-0.48	-0.55	0.12	-1.07	-0.68	0.31
raspberrypi	technology	0.06	0.33	0.38	-0.02	0.03	0.92	-1.04	0.00	0.49
rpg	culture-recreation	0.00	0.00	0.42	-0.11	-0.16	0.22	1.99	1.00	0.20
russian	culture-recreation	0.04	0.00	0.72	-0.13	-0.51	0.27	-1.63	-1.00	0.43
salesforce	technology	-0.02	-0.01	0.73	0.09	0.00	0.78	-0.77	-0.83	0.01
scicomp	science	0.07	0.00	0.88	0.38	0.79	0.29	-0.14	1.50	0.94
scifi	life-arts	-0.01	0.00	0.66	0.11	0.09	0.13	-0.04	0.45	0.89
security	technology	0.06	0.00	0.35	0.12	0.09	0.41	0.44	-0.25	0.48
serverfault	technology	0.02	-0.02	0.82	0.02	0.06	0.85	-0.55	0.00	0.00
sharepoint	technology	0.01	-0.01	0.77	0.11	0.16	0.31	-0.11	0.00	0.93
skeptics	culture-recreation	0.23	0.50	0.00	0.41	0.97	0.02	-1.57	0.33	0.94
sound	technology	-0.11	-0.03	0.01	-0.06	-0.03	0.70	-0.82	-1.00	0.09
spanish	culture-recreation	0.08	0.14	0.32	0.43	0.20	0.10	1.31	0.00	0.30
sqa	technology	-0.01	0.00	0.99	0.37	0.14	0.15	-0.79	-1.00	0.46
stackapps	technology	-0.06	0.00	0.97	0.46	0.39	0.08	0.43	1.00	0.97
stackoverflow	technology	0.01	-0.02	0.38	-0.01	-0.02	0.04	-0.45	-0.27	0.00
stats	science	-0.04	-0.14	0.28	0.00	0.10	0.85	0.60	0.17	0.09
superuser	technology	0.00	-0.03	0.77	-0.05	-0.08	0.18	0.66	0.00	0.29

<i>Comunidade</i>	<i>Categoria</i>	<i>Taxa</i>	<i>Taxa</i>	<i>Taxa</i>	<i>Utilidade</i>	<i>Utilidade</i>	<i>Utilidade</i>	<i>MVP *</i>	<i>MVP **</i>	<i>MVP</i>
		<i>Aceit. *</i>	<i>Aceit. **</i>	<i>Aceit. ***</i>	<i>Média *</i>	<i>Média **</i>	<i>Média ***</i>			<i>***</i>
tex	technology	-0.03	0.00	0.29	-0.01	-0.01	0.94	-0.77	-0.35	0.05
travel	culture-recreation	-0.01	0.00	0.88	0.04	0.16	0.49	1.29	0.88	0.05
unix	technology	0.01	0.00	0.77	0.03	-0.03	0.78	-0.92	-0.43	0.00
ux	technology	0.04	0.00	0.01	0.18	0.11	0.02	0.80	0.33	0.33
webapps	technology	0.07	0.04	0.07	0.16	0.31	0.14	1.57	0.25	0.56
webmasters	technology	0.02	0.00	0.94	-0.02	0.20	0.99	-0.02	0.16	0.40
wordpress	technology	0.08	0.12	0.01	-0.08	-0.02	0.41	0.20	0.43	0.02
workplace	professional	0.02	0.00	0.30	0.12	0.26	0.06	2.55	2.00	0.08
writers	life-arts	0.12	0.13	0.00	0.32	0.45	0.01	0.17	1.00	0.33

Tabela D.1: Para cada comunidade apresentamos nesta tabela sua categoria, a diferença entre a média de mulheres e homens quanto a taxa de aceitação das respostas / utilidade média das respostas / média dos votos das perguntas realizadas por usuários de cada gênero; a diferença entre a mediana de mulheres e homens quanto a taxa de aceitação das respostas / utilidade média das respostas / média dos votos das perguntas realizadas por usuários de cada gênero; e o *p-valor* obtido utilizando o Mann-Whitney-U test, para cada métrica de qualidade, para cada gênero.

Apêndice E

Estatísticas da terceira pergunta de pesquisa

* = Diferença entre a média das mulheres e a média dos homens.

** = Diferença entre a mediana das mulheres e a mediana dos homens.

*** = *P-valor* do teste de hipótese utilizado.

<i>Comunidade</i>	<i>Categoria</i>	<i>Frequência</i>	<i>Frequência</i>	<i>Frequência</i>	<i>Tempo de</i>	<i>Tempo de</i>
		*	**	***	<i>Vida</i> **	<i>Vida</i> ***
academia	life-arts	0.20	0.20	0.23	-2.12	0.30
android	technology	0.06	0.19	0.16	0.00	0.20
anime	culture-recreation	0.38	0.46	0.06	2.07	0.93
apple	technology	0.04	0.04	0.52	-14.50	0.00
askubuntu	technology	0.11	0.23	0.00	-29.00	0.96
bicycles	culture-recreation	0.05	0.05	0.58	-7.87	0.12
biology	science	-0.14	-0.25	0.16	9.05	0.30
bitcoin	business	-0.07	0.20	0.99	-1.43	0.32
chemistry	science	0.05	0.09	0.32	9.37	0.66
chinese	culture-recreation	0.64	0.34	0.54	-6.69	0.46
christianity	culture-recreation	-0.15	0.00	0.53	-6.88	0.60
codegolf	technology	0.16	0.38	0.14	4.54	0.79
codereview	technology	0.02	0.00	0.55	0.23	0.48
cogsci	science	0.10	-0.01	0.76	-0.83	0.47

<i>Comunidade</i>	<i>Categoria</i>	<i>Frequência</i>	<i>Frequência</i>	<i>Frequência</i>	<i>Tempo de</i>	<i>Tempo de</i>
		*	**	***	<i>Vida</i> **	<i>Vida</i> ***
cooking	life-arts	0.11	0.30	0.00	15.97	0.27
crypto	technology	-0.11	0.07	0.62	14.33	0.71
cs	science	-0.04	-0.08	0.68	8.51	0.50
cstheory	science	-0.09	-0.02	0.68	-20.26	0.09
dba	technology	-0.03	0.00	0.92	1.91	0.79
diy	life-arts	0.01	-0.01	0.95	-12.62	0.01
drupal	technology	-0.04	-0.01	0.79	101.76	0.77
dsp	technology	0.17	0.00	0.72	-17.15	0.00
electronics	technology	0.21	0.15	0.86	-10.53	0.21
ell	culture-recreation	0.14	0.08	0.30	4.34	0.98
english	culture-recreation	0.29	0.04	0.01	-9.05	0.00
expressionengine	technology	0.15	-0.02	0.56	inf	0.21
fitness	life-arts	0.13	-0.33	0.46	-1.53	0.24
freelancing	professional	-0.29	-0.33	0.20	1.64	0.13
french	culture-recreation	-0.08	-0.08	0.60	2.45	0.16
gamedev	technology	0.19	0.26	0.06	-10.72	0.75
gaming	culture-recreation	0.11	0.04	0.39	-4.63	0.33
gardening	life-arts	0.03	-0.14	0.62	-0.50	0.41
genealogy	life-arts	0.14	0.50	0.17	-1.77	0.42
german	culture-recreation	-0.07	0.16	0.45	-7.07	0.25
gis	technology	0.10	0.17	0.05	23.48	0.45
graphicdesign	life-arts	-0.07	-0.06	0.67	2.29	0.35
hermeneutics	culture-recreation	-0.01	0.02	0.65	13.52	0.81
history	culture-recreation	0.02	0.00	0.97	2.40	0.20
islam	culture-recreation	-0.20	-0.25	0.57	-6.67	0.42
japanese	culture-recreation	-0.09	-0.31	0.28	-7.08	0.16
judaism	culture-recreation	-0.22	-0.10	0.32	-52.95	0.37
linguistics	science	-0.35	-0.43	0.15	-5.65	0.20
magento	technology	0.00	0.04	0.38	6.29	0.60
math	science	0.19	0.10	0.00	-16.23	0.05

<i>Comunidade</i>	<i>Categoria</i>	<i>Frequência</i>	<i>Frequência</i>	<i>Frequência</i>	<i>Tempo de</i>	<i>Tempo de</i>
		*	**	***	<i>Vida</i> **	<i>Vida</i> ***
mathematica	technology	0.19	-0.13	0.51	-49.06	0.32
mathoverflow	science	-0.01	-0.01	0.59	-327.76	0.00
mechanics	culture-recreation	-0.17	-0.30	0.18	-6.42	0.02
money	life-arts	0.01	-0.08	0.47	-2.31	0.14
movies	life-arts	-0.27	-0.08	0.03	-0.05	0.67
music	life-arts	0.06	0.17	0.16	26.31	0.44
outdoors	culture-recreation	0.13	0.11	0.29	-12.11	0.08
parenting	life-arts	0.28	0.35	0.01	-1.28	0.26
philosophy	science	-0.10	-0.24	0.39	-1.66	0.02
photo	life-arts	0.04	-0.17	0.54	-4.98	0.00
physics	science	-0.09	0.00	0.71	0.06	0.14
pm	business	0.08	0.47	0.17	5.03	0.31
productivity	life-arts	0.09	0.17	0.16	6.25	0.35
programmers	technology	-0.02	-0.11	0.36	-15.12	0.08
quant	business	0.31	0.03	0.23	57.63	0.72
raspberrypi	technology	-0.19	-0.17	0.37	13.43	0.33
rpg	culture-recreation	-0.03	0.00	0.85	-24.27	0.05
russian	culture-recreation	-0.10	-0.33	0.30	20.18	0.77
salesforce	technology	-0.12	-0.08	0.51	-13.21	0.93
scicomp	science	-0.09	0.00	0.76	-35.94	0.07
scifi	life-arts	0.08	-0.11	0.79	-8.80	0.00
security	technology	0.01	0.08	0.51	-1.86	0.45
serverfault	technology	0.02	0.00	0.87	-35.29	0.04
sharepoint	technology	0.31	0.17	0.01	145.00	0.39
skeptics	culture-recreation	-0.03	0.00	0.77	6.53	0.49
sound	technology	0.16	0.12	0.97	-92.24	0.41
spanish	culture-recreation	0.02	0.22	0.70	-17.34	0.13
sqa	technology	0.04	0.29	0.24	17.91	0.31
stackapps	technology	0.16	0.00	0.41	-0.70	0.15
stackoverflow	technology	0.14	0.15	0.00	-346.87	0.00

<i>Comunidade</i>	<i>Categoria</i>	<i>Frequência</i>	<i>Frequência</i>	<i>Frequência</i>	<i>Tempo de</i>	<i>Tempo de</i>
		*	**	***	<i>Vida</i> **	<i>Vida</i> ***
stats	science	0.15	0.17	0.02	37.66	0.91
superuser	technology	0.07	0.00	0.17	-70.93	0.00
tex	technology	0.11	0.15	0.05	11.03	0.08
travel	culture-recreation	0.03	0.18	0.45	-1.92	0.01
unix	technology	0.02	0.04	0.74	-4.30	0.13
ux	technology	0.01	0.25	0.12	13.68	0.49
webapps	technology	-0.04	0.00	0.44	-0.54	0.06
webmasters	technology	0.02	-0.20	0.77	3.59	0.81
wordpress	technology	0.13	0.23	0.02	74.72	0.02
workplace	professional	0.00	0.25	0.21	2.89	0.24
writers	life-arts	-0.05	0.00	0.91	1.12	0.53

Tabela E.1: Para cada comunidade apresentamos nesta tabela sua categoria, a diferença entre a média de mulheres e homens quanto a frequência de contribuição feita por usuários de cada gênero; a diferença entre a mediana de mulheres e homens quanto a sua frequência de contribuição e tempo de vida dos usuários de cada gênero; e o *p-valor* obtido utilizando o Mann-Whitney-U test, para cada tipo de contribuição, para cada gênero.

Apêndice F

Estatísticas da quarta pergunta de pesquisa

<i>Comunidade</i>	<i>Categoria</i>	<i>Coeficiente</i>	<i>p-value</i>	<i>R²</i>
academia	life-arts	0.00	0.52	0.11
android	technology	0.00	0.00	0.80
anime	culture-recreation	-0.07	0.02	0.97
apple	technology	0.00	0.11	0.23
askubuntu	technology	0.01	0.00	0.71
bicycles	culture-recreation	0.00	0.29	0.16
biology	science	0.00	0.94	0.00
bitcoin	business	0.01	0.13	0.40
chemistry	science	0.00	0.98	0.00
chinese	culture-recreation	-0.01	0.72	0.03
christianity	culture-recreation	0.00	0.52	0.09
codegolf	technology	-0.03	0.19	0.18
codereview	technology	0.00	0.00	0.86
cogsci	science	0.00	0.99	0.00
cooking	life-arts	0.00	0.62	0.04
crypto	technology	0.01	0.03	0.59
cs	science	0.00	0.19	0.27

<i>Comunidade</i>	<i>Categoria</i>	<i>Coeficiente</i>	<i>p-value</i>	<i>R²</i>
cstheory	science	0.00	0.77	0.01
dba	technology	0.00	0.47	0.05
diy	life-arts	0.00	0.21	0.21
drupal	technology	0.01	0.01	0.54
dsp	technology	0.00	0.05	0.44
electronics	technology	0.00	0.33	0.11
ell	culture-recreation	0.01	0.68	0.10
english	culture-recreation	0.00	0.08	0.34
expressionengine	technology	-0.02	0.11	0.79
fitness	life-arts	0.01	0.45	0.10
freelancing	professional	0.02	0.36	0.72
french	culture-recreation	0.08	0.02	0.72
gamedev	technology	0.01	0.01	0.55
gaming	culture-recreation	-0.01	0.24	0.15
gardening	life-arts	-0.01	0.41	0.10
genealogy	life-arts	-0.02	0.13	0.59
german	culture-recreation	0.00	0.51	0.09
gis	technology	0.01	0.00	0.77
graphicdesign	life-arts	0.00	0.73	0.02
hermeneutics	culture-recreation	0.03	0.01	0.88
history	culture-recreation	-0.01	0.10	0.46
islam	culture-recreation	0.02	0.24	0.42
japanese	culture-recreation	0.00	0.99	0.00
judaism	culture-recreation	0.01	0.00	0.70
linguistics	science	0.00	0.18	0.21
magento	technology	0.01	0.26	0.54
math	science	0.00	0.00	0.76
mathematica	technology	0.00	0.95	0.00
mathoverflow	science	0.00	0.01	0.63
mechanics	culture-recreation	0.00	0.97	0.00

<i>Comunidade</i>	<i>Categoria</i>	<i>Coeficiente</i>	<i>p-value</i>	<i>R²</i>
money	life-arts	0.00	0.18	0.21
movies	life-arts	-0.01	0.10	0.53
music	life-arts	-0.01	0.11	0.36
outdoors	culture-recreation	0.00	0.86	0.01
parenting	life-arts	0.01	0.55	0.06
philosophy	science	0.00	0.17	0.34
photo	life-arts	0.00	0.73	0.02
physics	science	0.01	0.02	0.57
pm	business	-0.01	0.04	0.47
productivity	life-arts	0.02	0.01	0.76
programmers	technology	0.00	0.04	0.32
quant	business	0.01	0.10	0.39
raspberrypi	technology	0.02	0.08	0.70
rpg	culture-recreation	0.00	0.45	0.08
russian	culture-recreation	-0.05	0.11	0.63
salesforce	technology	0.01	0.18	0.50
scicomp	science	0.00	0.55	0.10
scifi	life-arts	-0.01	0.11	0.36
security	technology	0.00	0.52	0.06
serverfault	technology	0.00	0.00	0.81
sharepoint	technology	0.00	0.66	0.02
skeptics	culture-recreation	0.00	0.63	0.04
sound	technology	-0.02	0.00	0.86
spanish	culture-recreation	-0.01	0.50	0.12
sqa	technology	0.01	0.48	0.11
stackapps	technology	0.00	0.95	0.00
stackoverflow	technology	0.00	0.00	0.83
stats	science	0.00	0.12	0.25
superuser	technology	0.00	0.10	0.22
tex	technology	0.00	0.01	0.46

<i>Comunidade</i>	<i>Categoria</i>	<i>Coefficiente</i>	<i>p-value</i>	<i>R²</i>
travel	culture-recreation	-0.01	0.07	0.51
unix	technology	0.00	0.01	0.49
ux	technology	0.01	0.00	0.67
webapps	technology	0.00	0.05	0.37
webmasters	technology	0.01	0.00	0.67
wordpress	technology	0.00	0.21	0.21
workplace	professional	0.01	0.55	0.13
writers	life-arts	0.02	0.15	0.31

Tabela F.1: Para cada comunidade apresentamos nesta tabela sua categoria e os resultados da regressão simples onde verificamos se a proporção de contribuições provenientes de mulheres tem aumentado ao longo do tempo. Especificamente, são mostrados o coeficiente do modelo, o p-valor do coeficiente e o R^2 da regressão.

<i>Comunidade</i>	<i>Categoria</i>	<i>Coefficiente</i>	<i>p-value</i>	<i>R²</i>
academia	life-arts	0.01	0.01	0.82
android	technology	0.00	0.60	0.05
anime	culture-recreation	-0.01	0.84	0.03
apple	technology	0.00	0.94	0.00
askubuntu	technology	0.00	0.16	0.26
bicycles	culture-recreation	0.01	0.15	0.27
biology	science	-0.01	0.36	0.21
bitcoin	business	0.00	0.75	0.02
chemistry	science	-0.01	0.51	0.16
chinese	culture-recreation	-0.02	0.02	0.78
christianity	culture-recreation	-0.01	0.24	0.26
codegolf	technology	0.00	0.79	0.01
codereview	technology	0.00	0.69	0.03
cogsci	science	-0.01	0.48	0.13

<i>Comunidade</i>	<i>Categoria</i>	<i>Coeficiente</i>	<i>p-value</i>	<i>R²</i>
cooking	life-arts	-0.01	0.45	0.08
crypto	technology	0.02	0.01	0.80
cs	science	-0.01	0.29	0.27
cstheory	science	0.03	0.15	0.27
dba	technology	0.00	0.15	0.31
diy	life-arts	0.00	0.06	0.43
drupal	technology	0.00	0.33	0.16
dsp	technology	0.03	0.12	0.41
electronics	technology	0.00	0.14	0.25
ell	culture-recreation	-0.01	0.37	0.39
english	culture-recreation	0.01	0.01	0.64
expressionengine	technology	0.00	0.88	0.01
fitness	life-arts	-0.01	0.07	0.46
freelancing	professional	0.02	0.75	0.14
french	culture-recreation	0.03	0.06	0.54
gamedev	technology	0.00	0.21	0.21
gaming	culture-recreation	0.00	0.33	0.14
gardening	life-arts	-0.01	0.22	0.28
genealogy	life-arts	-0.04	0.65	0.12
german	culture-recreation	0.01	0.35	0.18
gis	technology	0.00	0.92	0.00
graphicdesign	life-arts	-0.01	0.06	0.48
hermeneutics	culture-recreation	0.01	0.22	0.35
history	culture-recreation	0.00	0.84	0.01
islam	culture-recreation	0.01	0.51	0.16
japanese	culture-recreation	0.01	0.24	0.26
judaism	culture-recreation	0.00	0.73	0.02
linguistics	science	-0.01	0.51	0.12
magento	technology	0.00	0.96	0.00
math	science	0.00	0.06	0.43

<i>Comunidade</i>	<i>Categoria</i>	<i>Coeficiente</i>	<i>p-value</i>	<i>R²</i>
mathematica	technology	0.00	0.48	0.13
mathoverflow	science	0.00	0.02	0.53
mechanics	culture-recreation	0.00	0.83	0.01
money	life-arts	-0.01	0.10	0.30
movies	life-arts	-0.01	0.27	0.29
music	life-arts	0.00	0.71	0.03
outdoors	culture-recreation	0.00	0.72	0.04
parenting	life-arts	-0.01	0.24	0.26
philosophy	science	0.00	0.66	0.04
photo	life-arts	0.00	0.59	0.04
physics	science	0.00	0.95	0.00
pm	business	0.01	0.13	0.34
productivity	life-arts	0.00	0.51	0.09
programmers	technology	0.00	0.13	0.30
quant	business	0.01	0.12	0.36
raspberrypi	technology	0.00	0.73	0.05
rpg	culture-recreation	0.00	0.18	0.24
russian	culture-recreation	0.00	0.95	0.00
salesforce	technology	0.00	0.83	0.02
scicomp	science	0.01	0.28	0.28
scifi	life-arts	0.00	0.67	0.03
security	technology	0.00	0.22	0.23
serverfault	technology	0.00	0.06	0.34
sharepoint	technology	0.00	0.54	0.05
skeptics	culture-recreation	0.00	0.81	0.01
sound	technology	0.01	0.25	0.16
spanish	culture-recreation	0.00	0.32	0.24
sqa	technology	-0.01	0.10	0.45
stackapps	technology	0.00	0.95	0.00
stackoverflow	technology	0.01	0.00	0.96

<i>Comunidade</i>	<i>Categoria</i>	<i>Coefficiente</i>	<i>p-value</i>	<i>R²</i>
stats	science	0.01	0.15	0.28
superuser	technology	0.00	0.01	0.53
tex	technology	0.00	0.05	0.44
travel	culture-recreation	0.00	0.68	0.04
unix	technology	0.00	0.71	0.02
ux	technology	0.01	0.18	0.24
webapps	technology	0.00	0.46	0.08
webmasters	technology	0.01	0.04	0.49
wordpress	technology	0.00	0.04	0.47
workplace	professional	0.01	0.17	0.51
writers	life-arts	-0.01	0.08	0.43

Tabela F.2: Para cada comunidade apresentamos nesta tabela sua categoria e os resultados da regressão simples onde verificamos se a proporção de registros provenientes de mulheres tem aumentado ao longo do tempo. Especificamente, são mostrados o coeficiente do modelo, o p-valor do coeficiente e o R^2 da regressão.