

# Pagamento Atrai Colaboradores de Longo Prazo ou Prestadores de Serviço? Uma Investigação Inicial sobre o Google Summer of Code 2014

**Jefferson O. Silva**  
Universidade de São Paulo  
São Paulo-SP-Brasil  
silvajo@ime.usp.br

**Igor Wiese, Igor Steinmacher**  
Universidade Tecnológica do Paraná  
Campo Mourão-PR-Brasil  
{igor, igorfs}@utfpr.edu.br

**Marco Aurelio Gerosa**  
Universidade de São Paulo  
São Paulo-SP-Brasil  
gerosa@ime.usp.br

## RESUMO

Comunidades online de colaboração aberta dependem da atração de novos colaboradores para manterem-se sustentáveis. Algumas comunidades promovem programas para atrair e reter novos colaboradores, muitas vezes utilizando incentivos financeiros. O Google Summer of Code (GSoC) é um programa global que possui como meta encorajar desenvolvedores a participarem em projetos de software livre por meio de recompensas financeiras. Esta pesquisa objetiva investigar se estudantes que participaram dos projetos do Apache Software Foundation no GSoC 2014 continuaram a contribuir após o término do programa. Nossas descobertas sugerem que os estudantes que participam tem perfil de prestadores de serviço, não voltando a contribuir após o término do programa, exceto para propor outros projetos em edições posteriores.

## ABSTRACT

Online open communities depend on attracting new contributors. In order to mitigate this problem, communities promote programs for luring and retaining new contributors, often using payments. Google Summer of Code (GSoC) is a global program aiming at inspiring developers to participate in open source projects using monetary rewards. This research aims at verifying whether GSoC 2014 students continued to contribute after the end of the program. Our findings suggest the students who participate do not continue to contribute to the projects as volunteers after the end of the program, but some come back to propose new projects in subsequent editions.

## Categories and Subject Descriptors

H.5.3 [Information Interfaces and Presentation (e.g., HCI)]: Group and Organization Interfaces – *computer supported cooperative work, organizational design.*

## Palavras-chave

Incentivos; Recompensas; Software Livre; Motivação; Google Summer of Code; Novatos; Hackathons

## 1. INTRODUÇÃO

Ambientes online de colaboração aberta, como a Wikipedia e comunidades de Software Livre (SL), promovem a produção coletiva por meio de colaboração mediada por

tecnologia com poucas barreiras para entrada e saída de participantes [11]. Muitas dessas comunidades são mantidas por voluntários e dependem da atração de novos colaboradores [11]. De acordo com Qureshi e Fang [12] é necessário motivar, engajar e reter novos colaboradores para promover uma comunidade sustentável. Kraut *et al.* [13] afirmam que novos colaboradores são fonte de informação, novas ideias e rotinas de trabalhos.

Comunidades de produção aberta promovem programas para atrair e reter novos colaboradores, algumas delas envolvendo custos financeiros. Como muitas comunidades abertas não geram receita e dependem de trabalho voluntário, devem escolher cuidadosamente como investir seus recursos, caso existentes, para atrair novos colaboradores. Para este artigo, focamos na análise da efetividade do uso de pagamento temporário para realização de uma tarefa no sentido de atrair e reter novos colaboradores de longo prazo.

Para isso, estudamos o programa Google Summer of Code (GSoC)<sup>1</sup>, que, dentre outras coisas, visa engajar novos desenvolvedores em projetos de software livre. O GSoC é um programa global, no qual os estudantes se inscrevem em projetos de software livre e, se aceitos, trabalham em parceria com mentores, que supostamente são desenvolvedores experientes provenientes dos projetos para os quais se inscreveram. Como prêmio, a Google oferece uma quantia de US\$ 5.500,00 ao estudante que conseguir finalizar com sucesso todas as três fases do programa. Dentre as metas definidas pelo GSoC, especificadas no site do programa<sup>2</sup>, destacamos: *inspirar jovens desenvolvedores a participarem do desenvolvimento de SL e auxiliar projetos de SL a identificar e trazer novos desenvolvedores.* Há outros programas similares, como o iRail/Open Summer of Code (OSoC)<sup>3</sup>, específico para estudantes belgas.

Dentro desse contexto, estabelecemos nossa questão de pesquisa: “*Em comunidades online de produção aberta cujo produto é SL, quão efetivos são os programas que encorajam a participação por meio de pagamento para*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SBSC10, November 4-6, 2015, Salvador, Bahia, Brazil.

Copyright 2010 ACM 1-58113-000-0/00/0010 ...\$15.00.

<sup>1</sup> <https://developers.google.com/open-source/soc/>

<sup>2</sup> [http://www.google-melange.com/gsoc/document/show/gsoc\\_program/google/gsoc2015/help\\_page#3](http://www.google-melange.com/gsoc/document/show/gsoc_program/google/gsoc2015/help_page#3). What are the goals of this program

<sup>3</sup> <https://hello.irail.be/irail-sumer-of-code/>

*realização de tarefas em termos de engajar colaboradores que continuem a contribuir após o término do programa?"*

Analisamos os dados dos estudantes que participaram de projetos de desenvolvimento da Fundação Apache na edição do GSoC de 2014. Os dados foram coletados das listas de arquivos de e-mails (*mailing lists archives*), do histórico do *issue tracker* e dos logs do Gerenciador de Configuração de Software (*Software Configuration Manager – SCM*) dos projetos nos quais os estudantes contribuíram. Em suma, esta pesquisa, de caráter exploratório, objetiva verificar se desenvolvedores participantes que contribuíram para os projetos da Apache Software Foundation na edição do GSoC 2014 continuaram a contribuir com os projetos após o término do programa.

Nossas descobertas sugerem que as metas do GSoC relacionadas à retenção de desenvolvedores não estão sendo alcançadas, o que significa que os estudantes investigados não colaboraram nos projetos após o término do GSoC.

Este trabalho está organizado da seguinte maneira: a Seção 2 apresenta uma revisão da literatura, necessária para um aprofundamento sobre o tema da pesquisa. A Seção 3 apresenta o método da pesquisa. Na Seção 4, são apresentados os resultados da pesquisa. A Seção 5 apresenta algumas limitações para a validade deste trabalho. Finalmente, a Seção 6 conclui esta pesquisa.

## 2. REVISÃO DA LITERATURA

Comunidades de produção aberta precisam continuamente encontrar maneiras de atrair novos membros para substituir aqueles que saem [18]. Por exemplo, a Wikipedia já expressou preocupação com o fato de que a taxa de entrada de novos colaboradores ser menor que a taxa de saída [23].

Determinar o que motiva os desenvolvedores a contribuírem voluntariamente com projetos de software livre (SL) também tem sido de grande interesse para os pesquisadores [2][22][18]. De forma geral, a desmotivação – ou ainda, a falta de motivação apropriada – dos desenvolvedores é identificada como uma causa comum de fracassos de projetos de software [7].

A decisão dos desenvolvedores de se engajar em um projeto de SL pode ser explicada como uma combinação de seus perfis [1], suas características pessoais [2], fatores contextuais [2], aspirações profissionais [3] e o tipo de mecanismo de incentivo ao qual estão submetidos [4].

É possível classificar os perfis dos desenvolvedores de SL, com relação ao incentivo financeiro, em quatro tipos [1]:

- **Ideologicamente motivados:** são os desenvolvedores que derivam seu senso de identidade do movimento de produção aberta, acreditando nos valores fundamentais de compartilhamento de conhecimento e na contribuição voluntária.
- **Intrinsicamente motivados:** são desenvolvedores com capacidade de tirar sua satisfação da própria

execução da tarefa, priorizando atividades que consideram prazerosas, desafiadoras e diferentes.

- **Extrinsicamente motivados:** são desenvolvedores principalmente motivados por fatores externos tais como salários, recompensas e/ou incentivos.
- **Motivados pelo senso de comunidade:** são desenvolvedores que derivam sua identidade das regras de coletividade, acreditando fortemente nos princípios da reciprocidade, colocando os interesses da comunidade acima de seus ganhos pessoais.

Desenvolvedores com diferentes perfis apresentam comportamento diferente com relação às recompensas financeiras [1][5]. Um exemplo é o caso da Mozilla, que oferece uma quantia para os desenvolvedores que encontram bugs em seu código. Uma reportagem relata<sup>4</sup> que cerca de 10% a 15% dos desenvolvedores recusam o recebimento da recompensa financeira. Krishnamurthy Shaosong, e Tripathi [1] concluíram que desenvolvedores *motivados pelo senso de comunidade* tendem a recusar recompensas financeiras quando são interpretadas como uma forma de ‘controle’. Quando o pagamento é interpretado como ‘suporte’ nenhum perfil demonstrou rejeição à recompensa.

O grau de motivação dos desenvolvedores é também diretamente relacionado às suas características pessoais [6]. A revisão sistemática de Beecham *et al.* [2] constata que as três características pessoais mais comuns aos desenvolvedores são: orientação ao crescimento, introversão e desejo por autonomia. Da mesma forma, a revisão sistemática lista *ambiente de trabalho precário* como fator mais desmotivador para um desenvolvedor.

Fatores contextuais, como por exemplo, o país onde o desenvolvedor reside ou para qual empresa trabalha, atuam como moderadores, isto é, determinam a intensidade com que as características pessoais ocorrem em cada um [2].

Horn Hann *et al.* [3] apresentam duas teorias diferentes com relação às aspirações profissionais dos desenvolvedores de SL: a teoria do capital humano e a teoria da sinalização. As explicações da teoria do capital humano são baseadas na aquisição de competências técnicas valorizadas pelo mercado. Nessa visão, a participação em um projeto de SL é vista como um investimento no qual o esforço aplicado e a experiência ganha no presente, serão transformados em maiores ganhos por meio de empregos de programação com salários maiores no futuro. A teoria da sinalização afirma que a participação dos desenvolvedores de SL serve como um sinal individual da capacidade produtiva do desenvolvedor para futuros empregadores.

Outro componente determinante na adesão dos projetos é relacionado ao mecanismo de incentivo usado pela organização. Scekic *et al.* [4] discutem o uso de

<sup>4</sup> <http://www.wincert.net/news/software/2209-more-than-1-in-10-mozilla-bug-finders-turn-down-cash-reward>.

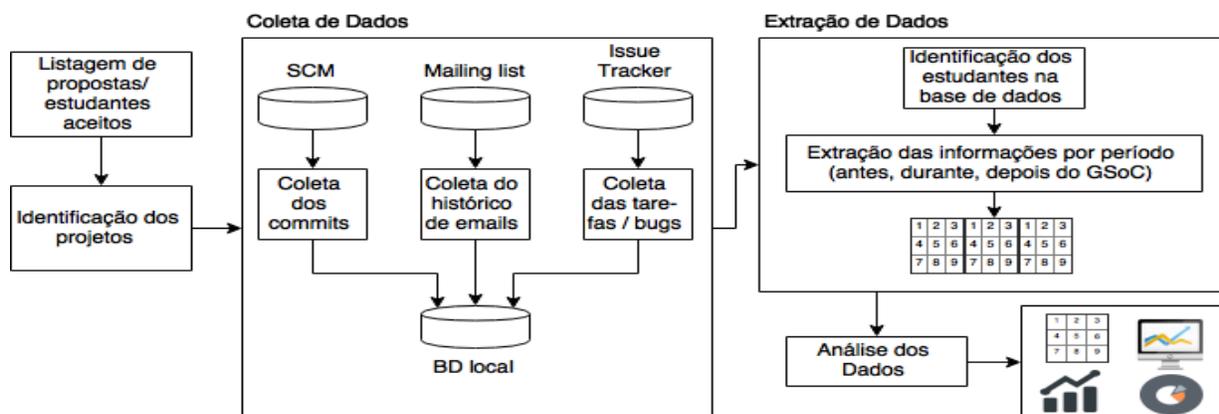


Figura 1. Método de pesquisa

diferentes mecanismos de incentivo. O mecanismo de incentivo do GSoC é uma avaliação realizada em três pontos pré-definidos no tempo, em que os participantes recebem as recompensas gradativamente. Este modelo é tipicamente utilizado para incentivar os participantes a trabalhar em tarefas mais longas, fazendo sentido ter uma avaliação objetiva do desempenho do participante em um momento particular. Os mecanismos de incentivos devem ser criteriosamente escolhidos, pois medidas de incentivos errôneas podem levar o participante a ter um comportamento indesejado [10].

Trainer *et al.* [18] apresentam um estudo qualitativo sobre a efetividade de dois tipos de programas que visam ao encorajamento de produção de código aberto: o GSoC e Hackathons. O estudo foca-se apenas em projetos de software científico. Com relação à efetividade da retenção dos estudantes, eles concluem que 18% dos estudantes, de sua amostra de 22 projetos, tornaram-se mentores.

Adicionalmente, alguns projetos do GSoC foram explorados como tema de pesquisa científica. Como exemplo, destacamos: o Biopython [15], que é um conjunto de bibliotecas para computação biológica; o Cytoscape [17], que é uma plataforma de software para visualizar redes de interações moleculares; o Wikipathways [16], que é um wiki para a contribuição e manutenção de conteúdo relacionado a caminhos biológicos.

### 3. MÉTODO DE PESQUISA

No intuito de responder nossa questão de pesquisa e conduzir nosso estudo exploratório, adotamos o método ilustrado na Figura 1.

O primeiro passo foi buscar os dados da edição do GSoC de 2014, que durou de maio a setembro desse ano. Os estudantes submeteram suas propostas até 21 de março. As propostas foram analisadas pelos membros do projeto e a lista de propostas selecionadas foi anunciada em 21 de abril de 2014. Nossa fase de coleta de dados para este estudo ocorreu entre 20 de abril de 2015 e 15 de maio de 2015.

Optamos por estudar a comunidade Apache pela diversidade de projetos que possui e por ser frequentemente

estudada empiricamente[19][20][21]. A comunidade está dispersa por vários países e conta com uma diversidade de perfis de desenvolvedores. Além disso, os projetos seguem algumas padronizações de processo e ferramentas, o que dá maior confiabilidade na interpretação e coleta de dados.

Cruzamos os nomes dos estudantes com projetos aceitos no site<sup>5</sup> do GSoC. Encontramos 33 estudantes aceitos para trabalhar nos projetos da Fundação Apache. Como não existe indicação explícita sobre para qual projeto Apache cada estudante foi selecionado para participar, realizamos uma busca manual pelos seus nomes usando o buscador da Google, o Apache Jira e os arquivos das listas de e-mails. Consideramos que identificamos o projeto quando encontramos evidências que ligavam os nomes dos estudantes ao projeto e quando havia menção ao GSoC 2014. Como resultado, identificamos 16 diferentes projetos da Fundação Apache que receberam estudantes.

O segundo passo consistiu da coleta de dados dos 16 projetos para os quais os estudantes contribuíram. Coletamos dados disponíveis publicamente de 3 diferentes fontes de cada projeto: *issue tracker*, logs do gerenciador de configuração de software e os arquivos de listas de e-mails.

Os dados foram coletados usando a suíte de ferramentas do MetricsGrimoire<sup>6</sup>. Estas ferramentas coletam os dados e os armazenam em bancos de dados relacionais, facilitando as buscas e inter-relacionamentos. As ferramentas que utilizamos foram:

- **Bicho**: utilizado para coletar informações do sistema de *issue tracking*; em nosso caso, Jira.
- **CVSAnalY**: utilizado para coletar informação dos sistemas de gerenciamento de código fonte. No contexto deste estudo, os sistemas de controle de versão usados foram o SVN e o GIT.
- **MailingListStats**: utilizado para extrair informação das listas de e-mails, arquivadas em formato mbox.

<sup>5</sup> <http://www.google-melange.com/gsoc/projects/list/google/gsoc2014>

<sup>6</sup> <http://metricsgrimoire.github.io/>

**Tabela 1. Número de estudantes que interagiram por ferramenta**

	Issue Tracker					SCM			Listas de emails	Todas ferramentas
	Issues Submetidas	Issues Atribuídas	Issues Comentadas	Anexos enviados	TOTAL	Commits	Commits (via outros)	TOTAL	Emails enviados	
Antes do GSoC	11	7	15	11	21	7	3	8	31	33
Durante o GSoC	14	10	17	13	24	12	6	17	27	33
Após o GSoC	2	5	2	0	7	5	3	7	13	18

Para identificar os estudantes em cada fonte, executamos múltiplas queries baseadas em seus nomes em uma tabela de banco de dados, que armazenou informação das pessoas usando os seguintes padrões (% é um curinga que corresponde a um número arbitrário de caracteres):

- <NomeCompleto>
- <Nome> <Sobrenome>
- %<Sobrenome>%
- %<Nome>%

Ao aplicar os padrões na ordem apresentada acima, fomos capazes de encontrar a maior parte dos estudantes. Encontramos 27 de 33 no *issue tracker* (81,8%), 32 (97,0%) nas listas de e-mails e 16 (48,5%) como committers no sistema de controle de versão. Com relação a esse sistema, executamos queries nas mensagens de commit, com os padrões apresentados anteriormente, visando encontrar commits feitos por outros colaboradores, e encontramos 6 estudantes, que não havíamos encontrado como colaboradores. Portanto, encontramos atividades de 22 estudantes no sistema de controle de versão (66,7%). Considerando todas as fontes, encontramos todos os estudantes interagindo com as comunidades.

O passo seguinte foi extrair informação de cada fonte, considerando três períodos distintos. Os períodos foram divididos para: ajudar a entender como os novatos agiam antes do programa; entender como eles se aproximavam e começavam a relação com a comunidade; entender seu comportamento durante o programa; para nos ajudar a verificar como eles interagiam com a comunidade enquanto desenvolviam seus projetos; e após o período do GSoC, para estudar a retenção dos novatos. Estes períodos foram definidos de acordo com o cronograma<sup>7</sup> do GSoC 2014, e são apresentados abaixo:

- **Antes do GSoC:** de 01-01-2013 até 21-04-2014 (publicação das propostas aceitas)
- **Durante o GSoC:** de 22-04-2014 (início do ‘período de entrosamento’) até 30-09-2014 (10 dias após o prazo para submissão do código)
- **Após o GSoC:** de 01-10-2014 até 15-05-2015 (final do período de coleta de dados)

Para cada período, coletamos as seguintes informações, de acordo com a fonte:

- Do *issue tracker*:
  - Issues submetidas pelo estudante
  - Comentários postados pelo estudante
  - Issues atribuídas ao estudante
  - Anexos (patches) submetidos pelo estudante.
- Da lista de e-mails:
  - E-mails enviados pelo estudante
- Do SCM:
  - Commits feitos pelo estudante
  - Commits feitos por outros colaboradores (identificados nas mensagens de commit)

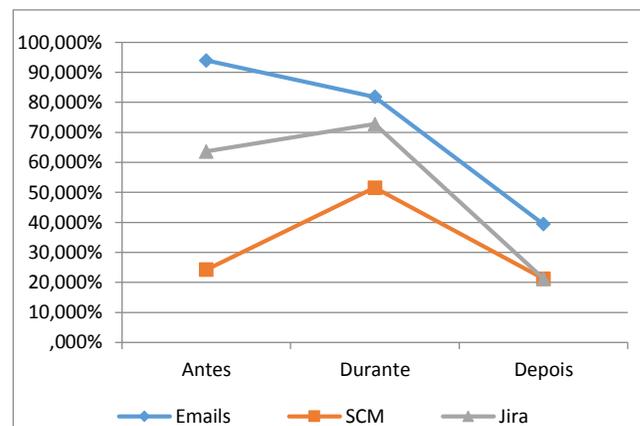
Coletamos estes números e analisamos o comportamento dos estudantes em cada período.

#### 4. RESULTADOS

Após a coleta, analisamos manualmente os dados para propiciar um melhor entendimento do processo de se colaborar em um projeto de software livre via GSoC e para responder nossa questão de pesquisa. Nas seguintes subseções, apresentamos análises diferentes feitas em nosso estudo exploratório.

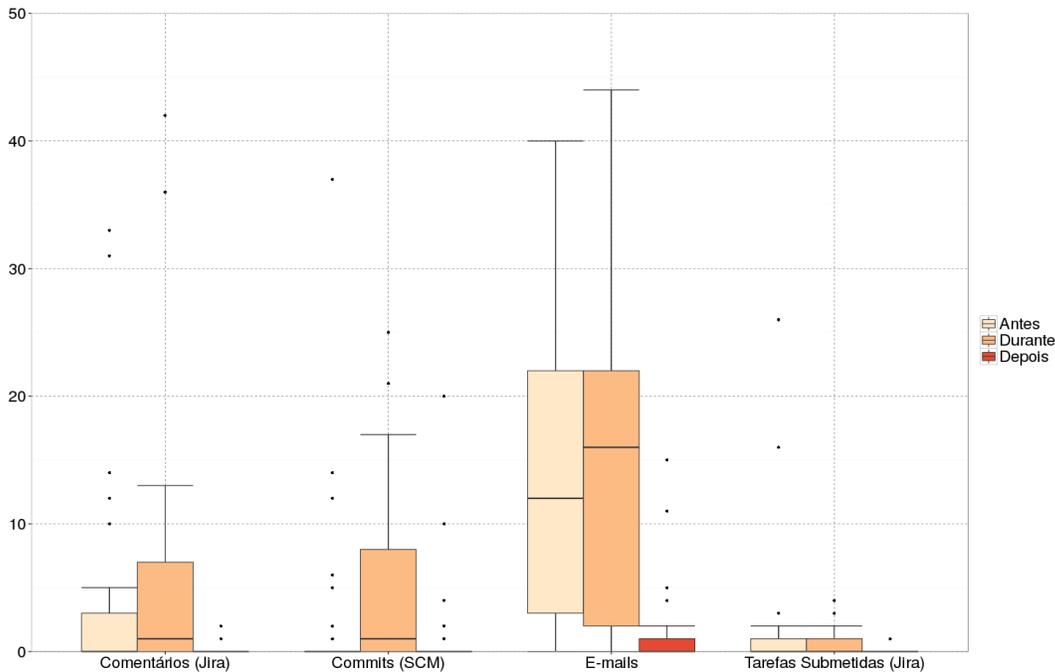
#### Análise da interação dos estudantes em diferentes ferramentas

Observando os números na Tabela 1 e na Figura 2, é claro que o número de estudantes que mantiveram interação com a comunidade após entregar o código do GSoC é menor que antes e durante o GSoC, independente da ferramenta que foi utilizada para interagir com a comunidade. Aproximadamente, apenas analisando a última coluna, é possível notar que 18 (54,5%) estudantes apareceram no projeto após o fim de sua tarefa.



**Figura 2. Quantidade de estudantes interagindo por ferramenta**

<sup>7</sup> <http://www.google-melange.com/gsoc/events/google/gsoc2014>



**Figura 3. Distribuição da interação por ferramenta**

Pode-se observar que este padrão é similar no Jira e no SCM, mas é ligeiramente diferente para os e-mails enviados na lista de e-mails. É possível notar que quase a totalidade dos estudantes interagiram com a comunidade via e-mail antes do GSoC, e então o número cai um pouco durante o período de contribuição. Isto acontece porque, normalmente, os estudantes precisam se apresentar e mandar suas propostas para a comunidade. Ao analisar manualmente suas mensagens, pudemos ver que isto acontece nas primeiras mensagens mandadas para a lista de e-mail do desenvolvedor

Na Figura 3, apresentamos a distribuição do número de interações de acordo com a ferramenta (e atividade no caso do *issue tracker*). É possível perceber que a lista de e-mails foi fortemente utilizada antes (mediana = 12 mensagens) e durante o GSoC (mediana = 16 mensagens). Também pode ser observado que o número de commits, e-mails e comentários no Jira aumentou durante o período de participação do estudante no GSoC. O aumento de participação era esperado, uma vez que os estudantes estavam trabalhando para entregar seus projetos.

Finalmente, percebemos que os estudantes que continuaram interagindo com os projetos após o GSoC apareceram como *outliers*, uma vez que grande parte deles não interagiu em qualquer ferramenta, levando as medianas a zero.

Para estender nossa avaliação do comportamento dos estudantes, também analisamos se eles interagiram usando apenas uma ou múltiplas ferramentas. Como resultado, apresentamos 3 diagramas de Venn, um por período, na Figura 4.

Antes do início do projeto, na Figura 4 (a), pudemos observar que 13 estudantes interagiram usando uma única ferramenta (11 usando listas de e-mails e 2 usando o *issue tracker*), mas a maioria interagiu com a comunidade por pelo menos 2 ferramentas.

Na Figura 4 (b), apresentamos o comportamento durante o período do GSoC. Esperávamos um alto percentual de estudantes aparecendo em todas as três ferramentas. Entretanto, encontramos apenas 10 (30,3%) com esse comportamento. Uma possível explicação é que em alguns projetos as discussões e feedbacks sobre os projetos ocorreram via *issue tracker*. Em outros casos, encontramos uma rica discussão e acompanhamento do andamento dos trabalhos ocorrendo via listas de e-mails. Se considerarmos a quantidade de estudantes que interagem com a comunidade via e-mail ou *issue tracker*, e que apareceram no sistema de controle de versão, encontramos 17 pessoas (51,5%). Em uma análise mais aprofundada, identificamos 6 estudantes que submeteram patches via *issue tracker* (anexo) e que não fizeram commits, e que não foram identificados nos logs de commits feitos por outros colaboradores. Portanto, identificamos patches para 23 de 33 participantes estudados.

Após o período do GSoC (Figura 4 (c)), encontramos alguns estudantes fazendo commits no SCM sem qualquer outro tipo de interação, e alguma interação residual com a lista de e-mails e o *issue tracker*. Exploramos mais detalhadamente o período após o GSoC nas subseções seguintes.

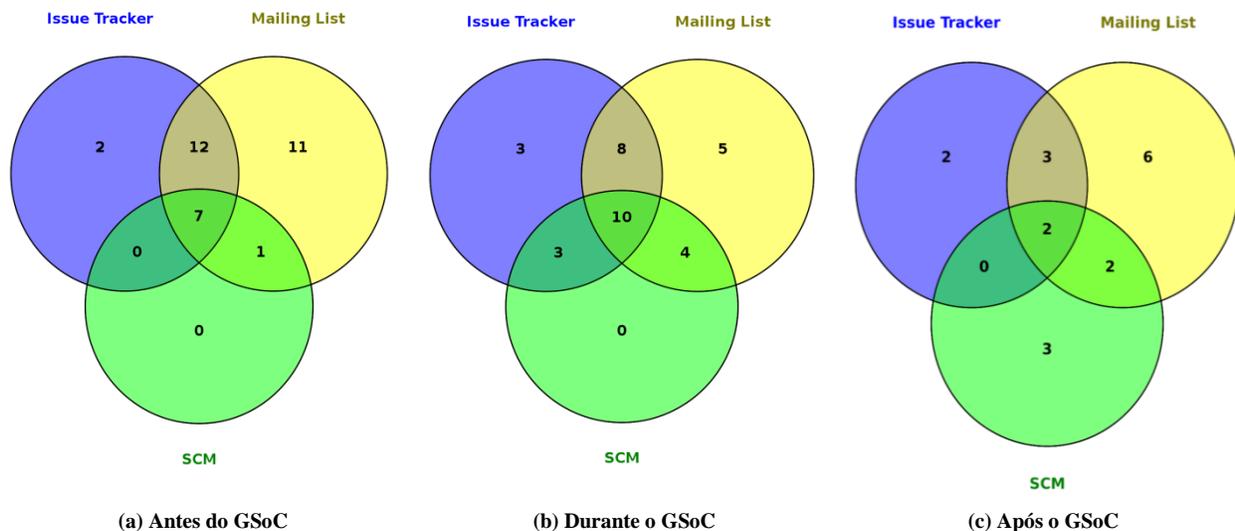


Figura 4. Diagrama de Venn representando as ferramentas usadas pelos estudantes para interagir com a comunidade

### Quão efetivo é o GSoC para reter estudantes?

Como nosso interesse era na retenção, exploramos estes dados um pouco mais. Encontramos que seis estudantes que interagiram via listas de e-mails apareceram somente em 2015, propondo um novo projeto para este ano. Outros três estudantes continuaram interagindo com as comunidades para concluir suas tarefas e para tirar dúvidas a respeito de seus projetos. Se excluirmos estes nove casos, somente quatro participantes continuaram interagindo via e-mail após o projeto.

Analisando os commits, encontramos que os commits de quatro estudantes, datados após o GSoC, tiveram relação com seus projetos do GSoC 2014, que não foram submetidos antes de setembro. Entre os outros três estudantes que submeteram código para o sistema de controle de versão após setembro de 2014, dois também continuaram interagindo por lista de e-mails. Analisando esses casos, verificamos que:

- **Estudante 1** é altamente ativo dentro da comunidade desde 2013. Participou no GSoC 2013, e se tornou membro do comitê gerenciador do projeto (PMC) no final de 2013. Após o GSoC 2014, manteve-se muito ativo com grande quantidade de e-mails enviados (80) e commits realizados (20), distribuídos de outubro de 2015 a abril de 2015.
- **Estudante 2** também é muito ativo dentro da comunidade desde 2013. Tornou-se *committer* (com direitos de enviar mudanças diretamente ao sistema de controle de versão) do projeto antes de começar o projeto do GSoC em 2014. Enviou e-mails esporadicamente e fez alguns commits (4 em 2015) após o GSoC.

Assim, após analisar os dados, podemos dizer que:

*Considerando os estudantes que participaram nos projetos Apache no GSoC 2014, o programa não foi efetivo no engajamento e na retenção de novos desenvolvedores como voluntários nas comunidades que eles participaram.*

De nossa amostra (33 estudantes), encontramos que somente 2 continuaram contribuindo após o término do programa, mas esses dois estudantes já eram colaboradores dos projetos antes do GSoC 2014 iniciar.

Acreditamos então que a participação no programa é motivada pela recompensa financeira, e, assim que a tarefa é concluída e o pagamento recebido, os estudantes param de interagir com o projeto. Em alguns casos, era claro observar esse comportamento orientado ao pagamento, porque os estudantes voltaram para a comunidade apenas para propor um novo projeto para a próxima edição do GSoC.

Neste ponto, é interessante apresentar outro resultado para nossa pesquisa exploratória nesta área, que é relacionada ao país de origem dos participantes do GSoC. Na Tabela 2, apresentamos o número de estudantes por país de origem. Identificamos seus países de origem procurando por seus perfis no GitHub, LinkedIn e páginas pessoais. Após encontrar nomes iguais aos dos estudantes nas listas, verificamos se eles mencionaram algo sobre o programa GSoC. Em todos os casos, os estudantes mencionaram o programa, especialmente em seus perfis do LinkedIn.

Como é possível notar, 20 estudantes são de 3 países. É interessante investigar as razões por detrás deste fenômeno. Seriam relacionadas ao dinheiro, uma vez que

o custo de vida (salários) nestes países é mais baixo? Isto explicaria o porquê os estudantes destes países procuram um lugar no GSoC, e poderia explicar as razões do porquê não continuam como voluntários, mas voltam a propor novos projetos em edições posteriores do programa.

**Tabela 2. Número de estudantes por país**

País	Participantes
Sri Lanka	13
China	5
Índia	3
EUA	2
Espanha	2
Irlanda	2
Reino Unido, França, Coréia do Sul, Portugal, Hungria, Estônia	1

## 5. AMEAÇAS À VALIDADE

Este trabalho de pesquisa contribui para a área não muito explorada da efetividade dos programas tais como o GSoC. Avança no entendimento das razões que levam os desenvolvedores a contribuir com projetos de SL, que é de interesse para a nossa sociedade. Entretanto, como acontece com qualquer pesquisa empírica, este trabalho sofre de algumas limitações.

Em primeiro lugar, selecionamos apenas estudantes de projetos da Apache Software Foundation do GSoC 2014. O GSoC tem sido realizado anualmente desde 2005, portanto nossos dados podem não ser representativos de toda a população de estudantes que já participaram. É possível que com uma seleção diferente de participantes, poderíamos chegar a outras conclusões.

Em segundo lugar, apenas analisamos os dados dos arquivos dos projetos e os disponíveis online. Uma pesquisa qualitativa com os participantes poderia revelar uma intenção positiva em colaborar com projetos de SL fora do GSoC, mas que não foi possível por razões que nos são desconhecidas.

## 6. CONCLUSÕES

Muitas comunidades de colaboração abertas são mantidas por voluntários, que podem facilmente parar de colaborar [8]. Por isso, é necessário um esforço para entender as reais motivações que os leva a colaborar [14].

Neste trabalho, apresentamos um estudo que investigou a efetividade do programa Google Summer of Code (GSoC), que tem entre seus objetivos recrutar novos colaboradores para atuar em projetos de software livre (SL). O GSoC se utiliza de um mecanismo de recompensas, em que o estudante gradualmente é premiado financeiramente à medida em que completa as fases do programa.

Um site de notícias reportou em 2005 que Chris Dibona, o gerente do programa de SL da Google, constatou que: 'cerca de 30% dos estudantes permanecem com seus

grupos após o GSoC<sup>8</sup>. Entretanto, em nosso estudo, analisando os 33 participantes do programa que selecionaram projetos da Fundação Apache na edição de 2014, não encontramos um único caso de novo desenvolvedor que tenha permanecido colaborando com o projeto após o término do programa. Constatamos que a maioria não voltou a colaborar após o término do programa, outros voltaram para propor projetos para a edição de 2015 e os únicos 2 que continuaram a contribuir para o projeto, já eram colaboradores ativos antes do programa. Portanto, não encontramos evidências que programas como GSoC sejam efetivos para encorajar a participação voluntária em projetos de software livre.

Para conduzir a pesquisa, buscamos identificar quais ferramentas de colaboração os participantes usam para interagir com a comunidade antes, durante e depois do programa. Para isso, mineramos dados de listas de e-mails, *issue trackers* e de sistemas de gerenciamento de configuração de software de projetos da Fundação Apache. Reconhecemos que nossa amostra não é ideal e fazemos nossa análise considerando os dados de poucos estudantes, e também considerando apenas a edição do GSoC 2014. No entanto, esta pesquisa ainda possui caráter exploratório, e nossos resultados levantam questões que podem ser investigadas em futuras pesquisas. Em particular, pode-se expandir esta pesquisa aumentando o número de projetos e edições e coletando dados complementares por meios de questionários e entrevistas.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. Krishnamurthy, S., Ou, S., & Tripathi, A. K. (2014). Acceptance of monetary rewards in open source software development. *Research Policy*, 43(4), 632–644. <http://doi.org/10.1016/j.respol.2013.10.007>
2. Beecham, S., Baddoo, N., Hall, T., Robinson, H., & Sharp, H. (2008). Motivation in Software Engineering: A systematic literature review. *Information and Software Technology*, 50(9-10), 860–878. <http://doi.org/10.1016/j.infsof.2007.09.004>
3. Hann, I.-H., Roberts, J., Slaughter, S., & Fielding, R. (2002). Economic Incentives for Participating Open Source Software Projects. Twenty-Third International Conference on Information Systems, ICIS 2002. Retrieved from <http://aisel.aisnet.org/icis2002/33>
4. Scekcic, O., Truong, H.-L., & Dustdar, S. (2013). Incentives and rewarding in social computing. *Communications of the ACM*, 56(6), 72. <http://doi.org/10.1145/2461256.2461275>

<sup>8</sup> 'something like 30 percent of the students stuck with their groups post SoC (Summer of Code) [sic]' <http://www.internetnews.com/dev-news/article.php/3607711>

5. Hsieh, G., Kraut, R. E., & Hudson, S. E. (2010). Why Pay?: Exploring How Financial Incentives are Used for Question & Answer. *Human-Computer Interaction*, 305–314. <http://doi.org/http://doi.acm.org/10.1145/1753326.1753373>
6. Misirli, A. T., Verner, J., Markkula, J., & Oivo, M. (2014). A Survey on Project Factors that Motivate Finnish Software Engineers.
7. Frangos, S. A. (1998). Motivated humans for reliable software products. *Microprocessors and Microsystems*, 21(10), 605–610. [http://doi.org/10.1016/S0141-9331\(98\)00062-3](http://doi.org/10.1016/S0141-9331(98)00062-3)
8. Sharp, H., & Hall, T. (2009). An initial investigation of software practitioners' motivation. *Proceedings of the 2009 ICSE Workshop on Cooperative and Human Aspects on Software Engineering, CHASE 2009*, 84–91. <http://doi.org/10.1109/CHASE.2009.5071418>
10. Hsieh, G., Hudson, S. E., & Kraut, R. E. (2011). Donate for Credibility: How Contribution Incentives Can Improve Credibility. *Group*, 3435–3438. <http://doi.org/10.1145/1978942.1979454>
11. Forte, A. and Lampe, C. 2013. Defining, Understanding, and Supporting Open Collaboration: Lessons From the Literature. *American Behavioral Scientist*. 57, 5 (2013), 535–547.
12. Qureshi, I. and Fang, Y. 2011. Socialization in Open Source Software Projects: A Growth Mixture Modeling Approach. *Org. Res. Methods*. 14, 1 (2011), 208–238.
13. Kraut, R.E., Burke, M., Riedl, J. and Resnick, P. 2012. *The Challenges of Dealing with Newcomers*. MIT Press. 179–230.
14. Wang, L.S., Chen, J., Ren, Y. and Riedl, J. 2012. Searching for the Goldilocks Zone: Trade-offs in Managing Online Volunteer Groups. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (Seattle, Washington, USA, 2012)*, 989–998.
15. Cock, P.J. a, Antao, T., Chang, J.T., Chapman, B. a, Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B. and de Hoon, M.J.L. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics (Oxford, England)*. 25, 11 (Jun. 2009), 1422–3
- 16 Pico, A.R., Kelder, T., van Iersel, M.P., Hanspers, K., Conklin, B.R. and Evelo, C. 2008. WikiPathways: Pathway Editing for the People. *PLoS Biology*. 6, 7 (2008), e184.
- 17 Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*. 13, 11 (Nov. 2003), 2498–504.
- 18 Erik H. Trainer, Chalalai Chaihirunkarn, Arun Kalyanasundaram, and James D. Herbsleb. 2014. Community Code Engagements: Summer of Code & Hackathons for Community Building in Scientific Software. In *Proceedings of the 18th International Conference on Supporting Group Work (GROUP '14)*. ACM, New York, NY, USA, 111-121. DOI=10.1145/2660398.2660420 <http://doi.acm.org/10.1145/2660398.2660420>
- 19 Bird, C. et al. 2006. Mining email social networks. *Proceedings of the 2006 international workshop on Mining software repositories (2006)*, 137–143.
- 20 Oliva, G.A. et al. 2012. Characterizing key developers: a case study with apache ant. *Collaboration and Technology*. Springer. 97–112.
- 21 Steinmacher, I. et al. 2012. Newcomers withdrawal in open source software projects: Analysis of Hadoop Common project. *Collaborative Systems (SBSC), 2012 Brazilian Symposium on (2012)*, 65–74
- 22 Burke, M., Marlow, C. and Lento, T. 2009. Feed Me: Motivating Newcomer Contribution in Social Network Sites. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Boston, MA, USA, 2009)*, 945–954.
- 23 Attracting and Retaining Participants: 2010. [https://strategy.wikimedia.org/wiki/Attracting\\_and\\_retaining\\_participants](https://strategy.wikimedia.org/wiki/Attracting_and_retaining_participants)