

Análise do papel das GitHub Discussions no ambiente de desenvolvimento de Software Livre

Ana Claudia Maciel¹, Igor Wiese², Igor Steinmacher²

¹Universidade Estadual de Maringá (UEM), Maringá - PR

²Universidade Tecnológica Federal do Paraná (UTFPR), Campo Mourão - PR

anacm.maciel@gmail.com, {igor, igorfs}@utfpr.edu.br,

Resumo. *GitHub Discussions é uma nova funcionalidade do GitHub que possibilita a abertura de discussões, e perguntas e respostas, fora do ambiente das pull requests e issues. Dada a novidade desta funcionalidade, é importante entender o papel que ela desempenha na entrada de novos e sustentabilidade de projetos no GitHub. Assim, o objetivo deste trabalho é analisar a relação das GitHub Discussions com as issues e pull requests e investigar como as discussões podem impulsionar a entrada de novos contribuidores. Em análises preliminares com o projeto Vercel/next.js, nota-se que um grande número de usuários teve sua primeira interação no projeto por meio das GitHub Discussions e que uma pequena parcela se envolveu em issues e pull requests. Como consolidação deste trabalho, novos projetos serão analisados e novas análises serão realizadas a fim de ter uma melhor compreensão do papel das GitHub Discussions no ecossistema do GitHub.*

1. Introdução

Canais de comunicação mostraram potencial para alavancar um grande número de contribuidores para criar bases de conhecimento ricas e eficientes em comunidades abertas [Nam et al. 2009]. Neste sentido, as discussões entre os desenvolvedores de software livre são necessárias para garantir a alta qualidade do código [Gousios et al. 2015]. Para apoiar as necessidades de colaboração dos desenvolvedores, as ferramentas de desenvolvimento modernas são frequentemente integradas ou complementadas por canais de comunicação e mídia social, como por exemplo, e-mail, fóruns, chat, serviços de microblog, etc. [Storey et al. 2017]. Entretanto, após o grande *boom* do GitHub, isso mudou, já que a plataforma centraliza todo o processo de colaboração.

O GitHub é uma plataforma de hospedagem de código-fonte e arquivos com controle de versão, além de gerenciar o relacionamento entre desenvolvedores projetos. Ele possibilita que desenvolvedores —ou quaisquer usuário cadastrado na plataforma— contribuam em projetos privados ou públicos de qualquer lugar do mundo. Em um ambiente diverso e distribuído como este, tópicos de discussão específicos de projeto de software podem ser uma opção para estabelecer uma ponte entre a comunicação entre os membros [Borges et al. 2016]. Pensando nisso, o GitHub oferece agora uma nova funcionalidade, as *GitHub Discussions* [Ushev 2020].

GitHub Discussions é uma funcionalidade do GitHub que possibilita a abertura de tópicos de discussões, e perguntas e respostas, fora do ambiente das *pull requests*—que são solicitações para que alterações dos artefatos sejam incorporadas ao projeto—e

das *issues*—que são as tarefas dos projetos [Hata et al. 2022]. Tal funcionalidade tem atraído uma grande quantidade de pessoas aos projetos, o que pode indicar o potencial de atração. Assim, dada a novidade desta funcionalidade, é importante entender o papel que ela desempenha no contexto do GitHub.

Nesse contexto, o objetivo deste trabalho é analisar a relação das GitHub *Discussions* com as *issues* e *pull requests* e investigar o papel das discussões na entrada de novos desenvolvedores em projetos de software livre. Para guiar o trabalho, foram definidas as seguintes questões de pesquisa:

Q1. Quais os papéis dos usuários envolvidos nas GitHub *Discussions*?

Essa questão de pesquisa visa agrupar os membros dos projetos de acordo com as ações que eles desempenham na comunidade. Além disso, pretende-se explorar o relacionamento dos membros do projeto com outras ferramentas do GitHub.

Q2. Como a inclusão das GitHub *Discussions* mudaram a entrada dos novatos do projeto?

Essa questão de pesquisa tem o objetivo de identificar se os novatos estão utilizando as GitHub *Discussions* como meio de entrada nos projetos.

Q3. As GitHub *Discussions* são mais atrativas para novatos do que os outros meios de interação (*Issues* e *Pull Requests*)?

Essa questão de pesquisa visa identificar se as GitHub *Discussions* estão atraindo mais novatos do que os outros meios mais técnicos e além disso se elas estão retendo esses membros que entraram pelas discussões.

Considerando que as Github *Discussions* ainda são uma novidade, é importante analisar o papel que ela está desempenhando nos projetos.

2. Trabalhos Relacionados

Este capítulo descreve estudos diretamente relacionados à realização desse trabalho. Na literatura são encontrados vários trabalhos que estudam comunicação no ambiente de software livre, alguns serão apresentados neste capítulo. Este projeto vai estudar o papel das GitHub *Discussions* no desenvolvimento de software livre e até o momento nenhum trabalho foi realizado nesse sentido, apenas Hata et al. [2021] que apresentaram essa nova funcionalidade.

Grande parte dos trabalhos acadêmicos [Guzzi et al. 2013, Storey et al. 2017, Kafer et al. 2018, Bi et al. 2021] analisaram listas de email como o principal mecanismo de comunicação em projetos de software livre. As listas de email continuam aparecendo na literatura nos trabalhos atuais, apesar de vários novos meios de comunicação estarem disponíveis para as comunidades de software livre, como por exemplo, as redes sociais (Twitter, Facebook) [Storey et al. 2017]. Várias discussões relacionadas a aspectos de desenvolvimento estão mudando recentemente para as ferramentas de *issue tracker* e sistemas de mensagem instantânea como Slack e Gitter [Kafer et al. 2018].

Embora diferentes conhecimentos em relação a diferentes canais de comunicação sejam abordados na literatura, o novo recurso das discussões do GitHub é uma nova base de conhecimento na comunidade de software livre e foi apresentado por Hata e colegas [2021], que realizaram um estudo para entender como os desenvolvedores usam esse

novo recurso e como isso impacta os processos de desenvolvimento. Foi conduzido um estudo de métodos mistos com base nos primeiros usuários das discussões do GitHub de janeiro até julho de 2020. Hata et al. [2021] chegaram a conclusão que há uma relação positiva em relação ao envolvimento dos membros do projeto e a frequência das discussões. Os desenvolvedores consideraram as discussões do GitHub úteis, mas enfrentam um problema em relação a duplicação entre discussões e questões. Por fim, concluem que as discussões desempenham um papel crucial no avanço no desenvolvimento de projetos.

Neste trabalho visamos estender o que foi analisado por Hata, objetivando entender melhor o papel das GitHub *Discussions* frente ao que já existe no GitHub, bem como da perspectiva da atratividade de novos contribuidores.

3. Método de Pesquisa

Este trabalho visa avançar o conhecimento atual sobre a nova funcionalidade do GitHub, as GitHub *Discussions*. Primeiro, foi feita uma coleta de dados das *issues*, *pull requests*, e das discussões de projetos do GitHub. Posteriormente serão aplicadas técnicas variadas de análise dos dados coletados.

A coleta das *issues* e *pull requests* foi realizada por meio da GitHub REST API¹. No entanto, a API ainda não suportava o novo recurso das discussões, portanto para coletar os dados das discussões foi utilizada GraphQL² que é uma linguagem de consulta para consumir a API.

Foram definidas três questões de pesquisa para a realização deste trabalho, as quais cada uma necessita de uma ou mais análises diferentes que serão descritas abaixo.

Q1. Quais os papéis dos usuários envolvidos nas GitHub *Discussions*?

Serão utilizadas análises de clusterização e análises de rede social para responder esta questão. A análise de rede social utiliza estruturas sociais por meio do uso de redes e da teoria dos grafos. Os métodos de clusterização visam encontrar uma solução que maximize simultaneamente a homogeneidade dentro do grupo e a heterogeneidade entre os grupos.

Técnicas de clusterização já são utilizadas em outros estudos. Ogawa et al. [2007] analisaram, por meio de clusters as listas de emails de projetos já consolidados, como o Apache webserver e mostrou ser uma boa técnica. Neste trabalho, portanto, também aplicaremos técnicas de clusterização.

Técnicas de análise de rede social já foram encontradas na literatura como em Faqir et al. [2019] que utilizaram, por exemplo, métricas como grau (número de conexões na rede), *betweenness* (média da menor distância de um nós em relação aos outros nós) e *closeness* (caminho mais curto entre os outros nós) para caracterizar os membros da rede social criada. Neste trabalho de doutorado serão utilizadas análises similares a essa.

Q2. Como a inclusão das GitHub *Discussions* mudaram a entrada dos novatos do projeto?

Para essa questão de pesquisa será feita uma análise quantitativa em relação as *issues*, *pull requests* e discussões. Será observado qual é a data da primeira interação no projeto

¹<https://docs.github.com/en/rest>

²<https://graphql.org/>

e qual foi o meio utilizado, diante dessa análise teremos todos os membros que utilizaram as GitHub *Discussions* como meio de entrada no projeto.

Izquierdo e Cabot [2022] apresentaram uma análise quantitativa que pode servir de referência para esta proposta. Mediante essa análise, os autores classificaram todas as ações e membros do projeto com base em uma definição precisa de possíveis papéis de acordo com a forma de contribuição no GitHub. Também, calcularam várias métricas relacionadas à composição, diversidade e evolução de papéis.

Q3. As GitHub *Discussions* são mais atrativas para novatos do que os outros meios de interação (*Issues* e *Pull Requests*)?

Para verificar se as GitHub *Discussions* são mais atrativas do que os outros meios mais técnicos, iremos utilizar a análise feita na Q2. E além disso, verificar se os novatos que entraram no projeto pelas GitHub *Discussions* permanecem no projeto, ou seja, eles precisam ter pelo menos mais uma interação para que seja contabilizado nessa questão. E diante disso, concluir se as GitHub *Discussions* retêm mais membros do que os outros meios do GitHub. Pretende-se também, comparar estatisticamente o número de novatos que entraram pelas discussões e aqueles que entraram por outros meios e verificar se as discussões impactam o desenvolvimento do projeto.

4. Resultados Preliminares

A proposta deste trabalho de doutorado tem por objetivo avaliar o papel das GitHub *Discussions* no como mecanismo de entrada em projetos de Software Livre. Este trabalho está em fase inicial, porém foi feito um estudo preliminar para conhecimento do ambiente das *Discussions*.

As análises e estudos preliminares foram realizadas no projeto Vercel/next.js, primeiro projeto que teve as GitHub *Discussions* liberadas. Neste projeto, os usuários que utilizaram as discussões como meio de entrada no projeto foram selecionados, examinando fatores como: quão ativos esses usuários são, considerando a quantidade de *issues* e *pull requests*. Destes, selecionamos os mais ativos, médios e menos ativos, e idade do usuário no GitHub.

Essas análises preliminares, mostraram que temos cerca de 28% dos usuários (4000 usuários, aproximadamente) os quais utilizaram as GitHub *Discussions* como meio de entrada no projeto. Destes, 97 usuários criaram pelo menos uma *pull request* e uma *issue*. Além disso, 10 usuários criaram pelo menos uma *pull request* e uma *issue* e fizeram comentários nas *issues* e nas *pull requests*, ou seja, interagiram em todos os meios. O conjunto desses 10 usuários contém usuários novos e antigos e usuários mais e menos ativos no projeto.

Diante dessa amostra, há evidências de que os mantenedores do projeto abraçaram a iniciativa e estão entre os mais ativos nas GitHub *Discussions* e outros meios mais técnicos. Também, as GitHub *Discussions* estão atraindo novos usuários que podem tornar-se ativos por meio dessa nova funcionalidade.

No entanto, isso não pode ser generalizado, pois pode ser que em outros projetos, os membros comportem-se de maneira diferente. Outros projetos serão estudados e outras análises serão aplicadas para consolidação desse trabalho.

Agradecimentos

Agradecemos o apoio financeiro do CNPq (#313067/2020-1) e da UTFPR-Campus Campo Mourão.

Referências

- Bi, T.; Ding, W.; Liang, P.; Tang, A. (2021). Architecture information communication in two OSS projects: The why, who, when, and what. *Journal of Systems and Software*, 181:111035.
- Borges, H.; Hora, A.; Valente, M. T. (2016). Predicting the Popularity of Github Repositories. In *Proceedings of the The 12th International Conference on Predictive Models and Data Analytics in Software Engineering*, New York, NY, USA. Association for Computing Machinery.
- Faqir, Y. E.; Arroyo, J.; Serrano, A. (2019). Visualization of the Evolution of Collaboration and Communication Networks in Wikis. In *Proceedings of the 15th International Symposium on Open Collaboration*, New York, NY, USA.
- Gousios, G.; Zaidman, A.; Storey, M.-A.; van Deursen, A. (2015). Work Practices and Challenges in Pull-Based Development: The Integrator’s Perspective. In *Proceedings of the 37th International Conference on Software Engineering - Volume 1*, page 358–368, Florence, Italy. IEEE Press.
- Guzzi, A.; Bacchelli, A.; Lanza, M.; Pinzger, M.; Deursen, A. v. (2013). Communication in Open Source Software Development Mailing Lists. In *Proceedings of the 10th Working Conference on Mining Software Repositories*, page 277–286.
- Hata, H.; Novielli, N.; Baltés, S.; Kula, R. G.; Treude, C. (2022). Github Discussions: An exploratory study of early adoption. *Empirical Software Engineering*, 27(1):1–32.
- Izquierdo, J.; Cabot, J. (2022). On the analysis of non-coding roles in open source development: An empirical study of npm package projects. *Empirical Software Engineering*, 27.
- Kafer, V.; Graziotin, D.; Bogicevic, I.; Wagner, S.; Ramadani, J. (2018). Communication in Open-Source Projects-End of the e-Mail Era? In *Proceedings of the 40th International Conference on Software Engineering: Companion Proceedings*, page 242–243, New York, NY, USA. Association for Computing Machinery.
- Nam, K. K.; Ackerman, M. S.; Adamic, L. A. (2009). Questions in, Knowledge in? A Study of Naver’s Question Answering Community. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, page 779–788, New York, NY, USA. Association for Computing Machinery.
- Ogawa, M.; Ma, K.-L.; Bird, C.; Devanbu, P.; Gourley, A. (2007). Visualizing social interaction in open source software projects. In *2007 6th International Asia-Pacific Symposium on Visualization*, pages 25–32, Sydney, NSW, Australia. IEEE.
- Storey, M.-A.; Zagalsky, A.; Filho, F. F.; Singer, L.; German, D. M. (2017). How Social and Communication Channels Shape and Challenge a Participatory Culture in Software Development. *IEEE Transactions on Software Engineering*, 43(2):185–204.
- Ushev, D. (2020). My thoughts on github discussions. <https://dzhavat.github.io/2020/04/04/my-thoughts-on-github-discussions.html>.