

Universidade Federal de Pernambuco
Centro de Informática

KARLA POLYANA SILVA FALCÃO
{kpsf@cin.ufpe.br}

Análise de Comunidades de Desenvolvedores Open Source

Ricardo Bastos Cavalcante Prudêncio
Orientador

CONTEXTO

O Github é a maior e mais popular plataforma de hospedagem, armazenamento e gerenciamento de código com o controle de versão distribuído Git (Git, 2008). Em 2018 o GitHub completa 10 anos e atualmente ela suporta aproximadamente 25 milhões de usuários, 3 bilhões de repositórios, dentre os quais apenas 25 milhões estão ativos e mais de bilhões de commits.

Além de hospedagem de código o github é caracterizado como uma rede social, por conta dos seus vários recursos sociais, como por exemplo, seguir outros desenvolvedores. Existem várias pesquisas envolvendo os dados do GitHub. Alguns resultados mostram que as atividades dos desenvolvedores servem como um portfólio profissional que gera uma impressão para os outros que então internalizam essas impressões para decidir quem e o que acompanhar ou onde contribuir em seguida [1,2,3]. Outros estudos mostram que os desenvolvedores veem os recursos sociais como um fator responsável pelo aumento da contribuição [4].

Algumas pesquisas envolvem análise de redes complexas. Thung et al. [5] construíram redes sociais de desenvolvedores envolvidos com 100.000 projetos do GitHub para demonstrar a estrutura social do ecossistema mundial do GitHub. Lima et al. [4] examinaram as localizações geográficas dos desenvolvedores do GitHub examinando informações de localização disponíveis nos perfis do GitHub.

OBJETIVO

Neste trabalho iremos analisar atividades dos usuários no GitHub e nos apoiaremos na teoria das redes complexas. Através da análise da estrutura do grafo social da comunidade de desenvolvedores do Github, desejamos encontrar grupos com interesses comuns, para assim ter um panorama do estado do desenvolvimento de software livre na cidade do Recife e fomentar idéias que apoiem o ecossistema de computação social.

METODOLOGIA

Primeiro, analisamos as informações dos dados calculando várias estatísticas,

então, construímos um grafo de relacionamentos desenvolvedor-desenvolvedor com a relação “seguir” e extraímos várias propriedades da rede complexa formada. Em seguida, identificamos desenvolvedores influentes nesta rede do GitHub usando algoritmos de influência e centralidade. Por fim, aplicamos alguns algoritmos renomados de clustering em redes para encontrar os grupos e os caracterizamos de acordo com o perfil dos desenvolvedores.

CRONOGRAMA

	Agosto	Setembro	Outubro	Novembro
Revisão bibliográfica	x			
Levantamento de dados	x	x		
Análises	x	x	x	
Escrita do Relatório		x	x	
Defesa				x

REFERÊNCIAS

[1] Kelly Blincoe, Jyoti Sheoran, Sean Goggins, Eva Petakovic, Daniela Damian. Understanding the popular users: Following, affiliation influence and leadership on GitHub. Information and Software Technology, ISSN: 0950-5849, Vol: 70, Page: 30-39. 2016

[2] F. Thung, T. Bissyande, D. Lo, and L. Jiang. Network structure of social coding in github. In 17th European Conference on Software Maintenance and Reengineering (CSMR), pages 323–326, 2013.

[3] Jason T. Tsay, Laura Dabbish, and James Herbsleb. Social media and success in open source projects. In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion (CSCW '12). ACM, New York, NY, USA, 223-226. 2012.

[4] Antonio Lima, Luca Rossi, Mirco Musolesi. Coding Together at Scale: GitHub as a Collaborative Social Network. Proceedings of 8th AAAI International Conference on Weblogs and Social Media, 2014.

[5] N. McDonald and S. Goggins. Performance and participation in open source software on github. In CHI'13 Extended Abstracts on Human Factors in Computing Systems, pages 139–

144. ACM, 2013.