

Um mapeamento sistemático sobre a Análise de Redes Sociais aplicada à colaboração em projetos de software

Emanoel Carlos Gomes Ferraz Silva¹, José Claudemir Pacheco Júnior¹,
Vitor de Barros Costa¹, Alixandre Thiago Ferreira Santana²

¹ Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Caixa Postal 50.740–560 – Recife – PE – Brasil

² Universidade Federal Rural de Pernambuco – Unidade Acadêmica de Garanhuns
Caixa Postal 55.292–270 – Garanhuns – PE – Brasil

{ecgfs, jcpj, vbc}@cin.ufpe.br, alixandre@uag.ufrpe.br

Abstract. *Through a systematic mapping of the literature, this paper aims to map studies directly related to social network analysis in the context of collaboration in software development. We seek to list the key authors and data sources used by the primary studies. Also, identify the main metrics of social network analysis and research methods used.*

Resumo. *Através de um mapeamento sistemático da literatura, este trabalho pretende mapear estudos diretamente relacionados à análise de redes sociais no contexto da colaboração no desenvolvimento de software. Busca-se elencar os principais autores e fontes de dados utilizadas pelos estudos primários. Além disso, identificar as principais métricas de análise de redes sociais e métodos de pesquisa utilizados.*

1. Introdução

A indústria de tecnologia de informação se apoia cada vez mais nos trabalhadores do conhecimento. Pode-se afirmar que o contexto do desenvolvimento de software é uma atividade complexa e de natureza socio-técnica, uma vez que exige a participação de vários indivíduos nas mais diversas atividades do seu ciclo de desenvolvimento [FITSILIS et al. 2009, CHISAN and DAMIAN 2004].

Naturalmente, através do fluxo de informação são criadas conexões entre os integrantes de um projeto e uma estrutura social de interações implicitamente se forma. Recentes estudos [FITSILIS et al. 2009, SCHWIND and WEGMANN 2008] vem analisando tais estruturas procurando padrões, problemas e explicações utilizando conceitos de Análise de Redes Sociais [KADUSHIN 2012, WASSERMAN 1994].

Seja no desenvolvimento de sistemas de código aberto ou de sistemas proprietários, a dimensão da colaboração e a interação entre os agentes sociais envolvidos é um fator importante e que tem impacto na performance de projetos [ELLIS et al. 1991, FUKS et al. 2007]. Dessa forma, entender como a colaboração se comporta em projetos de software se apresenta como um tópico importante para a indústria de software.

Nesse sentido, através de um mapeamento sistemático, este trabalho tem como objetivo geral mapear e categorizar estudos diretamente relacionados à análise de redes sociais no contexto da colaboração no desenvolvimento de software, com o intuito de sistematizar o estado-da-arte.

2. Metodologia

A presente pesquisa se baseia na execução de um mapeamento sistemático da literatura, que segundo [KITCHENHAM 2001, p.3], “é um meio de identificar, mensurar e interpretar toda pesquisa relevante disponível para determinada questão de pesquisa, ou tópico, ou fenômeno de interesse”. A seguir será descrita a construção do protocolo de pesquisa.

2.1. Questões de Pesquisa

Neste trabalho, as pesquisas que aplicam ARS (Análise de Redes Sociais) em colaboração foram tratadas, analisadas e consolidadas com o objetivo de responder as seguintes questões de pesquisa: (i) Quais os principais métodos de pesquisa utilizados? (ii) Quais as principais métricas de análise de redes sociais utilizadas? (iii) Quais os objetivos de pesquisa mapeados nos estudos primários?

2.2. Estratégia de Busca

Este trabalho é parte de uma pesquisa mais abrangente sobre ARS em Projetos sob as seguintes dimensões: escopo, comunicação, conhecimento, liderança, estrutura time, stakeholder, inovação, complexidade, sucesso/fracasso, performance e colaboração. Aqui levou-se em consideração apenas os *projetos de software* classificados na dimensão de *colaboração*. Assim, a estratégia de busca utilizada foi construída com base nos termos *Análise de Redes Sociais e Projeto*, resultando na *string* de busca ilustrada pela Tabela 1.

Tabela 1. String de busca utilizada para o Mapeamento Sistemático

<i>String final de busca</i>
((project OR projects OR endeavor OR endeavors OR endeavour OR endeavours OR "Product Design"OR "Product Development") AND ("Social network"OR "Social networks"OR "Social Computing"OR "Social Networking"OR "social graph"OR "social graphs"OR sociometry OR sociometric OR sociometrics OR sociogram OR sociograms) AND ("metric"OR "metrics"OR measure OR measures OR "method"OR "methods"OR "analysis"OR "technique"OR "techniques"))

Considerando a relevância da quantidade e qualidade dos trabalhos publicados na área de software, a *string* de busca foi aplicada aos seguintes engines de busca: IEEE, Scopus, Science Direct e El compendex.

2.3. Critérios de inclusão e exclusão

Como critério de inclusão, foram selecionados apenas artigos que continham técnicas, métodos ou qualquer tipo de iniciativa para medir a aplicação de ARS tendo projetos como unidade de análise.

Por outro lado, foram excluídos aqueles que se enquadraram em pelo menos um dos seguintes critérios: (i) Publicado antes de janeiro de 2008; (ii) Artigos não escritos em inglês; (iii) Estudos não empíricos; (iv) Artigos com resultados duplicados, por exemplo, publicado em mais de um evento; (v) Resumo de *keynotes*, tutoriais, white papers e artigos incompletos, livros de capítulo, dissertações; (vi) Estudos secundários ou terciários; (vii) Artigos relacionados a redes de co-autoria;

2.4. Extração de dados

Para a execução das três fases, o grupo de pesquisadores envolvidos no mapeamento sistemático foi dividido aleatoriamente em três categorias: revisor, verificador e moderador. A extração de dados dos artigos foi dividida em três fases: análise do título e resumo, análise da introdução e conclusão e, por fim, extração de dados.

Tabela 2. Participação dos engenhos de busca

	Scopus	El Compendex	IEEE	Science Direct	Total
Fase 1	1019	609	217	101	1946
Fase 3	8	5	4	1	18

Tabela 3. Referência dos estudos primários selecionados

ID	Título	Ano	Fonte
23	Social Network Analysis of European Project Consortia to Reveal Impact of Technology-Enhanced Learning Projects	2011	[1]
120	Applying Centrality Measures to the Behavior Analysis of Developers in Open Source Software Community	2012	[1]
150	SVNNAT: Measuring Collaboration in Software Development Networks	2008	[1]
158	You scratch my back, we'll scratch yours: Exploring "open entrepreneurship" in a private-collective community	2012	[1]
237	Participation motifs and the emergence of organization in open productions	2013	[2]
449	The meso-level structure of F/OSS collaboration network: Local communities and their innovativeness	2010	[3]
453	Intelligent analysis of user interactions in a collaborative software engineering context	2012	[4]
459	Supporting situation awareness in FLOSS projects by semantical aggregation of tools feeds	2009	[4]
462	Evolution of developer collaboration on the Jazz platform: A study of a large scale agile project	2011	[4]
473	Bringing out collaboration in software development social networks	2011	[4]
506	Visualizing collaboration and influence in the open-source software community	2011	[4]
564	An exploratory study on collaboration understanding in software development social networks	2012	[3]
578	Improving developer activity metrics with issue tracking annotations	2010	[3]
612	What dynamic network metrics can tell us about developer roles	2008	[3]
657	Using social network analysis for mining collaboration data in a defect tracking system for risk and vulnerability analysis	2011	[3]
703	A quantitative study of social organisation in open source software communities	2012	[4]
822	Collaboration structure and performance in new software development: Findings from the study of open source projects	2010	[4]
971	Using social network analysis techniques to study collaboration between a FLOSS community and a company	2008	[4]

Sendo: [1] - IEEE, [2] - Science Direct, [3] - El Compendex, [4] - Scopus

3. Estatística dos meta-dados

A seguir serão apresentados os dados quantitativos resultantes da execução do protocolo de pesquisa.

3.1. Participação dos engenhos de busca

Na primeira fase um total de 1946 artigos foi retornado pelos engenhos de busca, incluindo duplicidades, sendo 1129 trabalhos únicos. Após o filtro da fase 1, aplicado pela leitura do título e *abstract*, a quantidade de estudos primários reduziu para 255.

Ao final do filtro da fase 2, um total de 152 artigos abordavam ARS aplicada a projetos. Desse total, 18 artigos abordavam colaboração em projetos sob a ótica da ARS (ver Tabela 3). A Tabela 2 mostra a distribuição dos estudos primários da primeira e terceira fase por cada engenho de busca.

3.2. Distribuição por país

Todos os artigos mapeados por este trabalho foram publicados por um total de 12 países, sendo que alguns artigos tiveram autoria de mais de um país. Foi o caso dos artigos [158], [237] e [449]. Para cada país a Figura 1 apresenta quantos artigos foram publicados, bem como o ano de publicação.

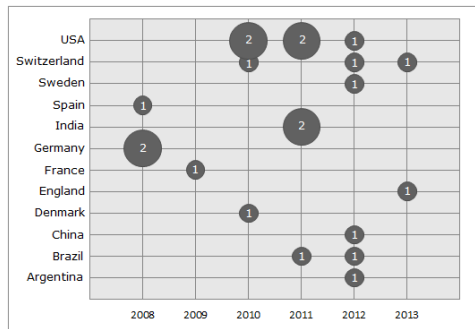


Figura 1. Participação dos países ao longo do tempo

Tabela 4. Autores dos estudos primários classificados pelo fator de impacto

Estudo primário	Autor	H_i scholar	Estudo primário	Autor	H_i scholar
23, 578	Laurie Williams	42	703	Ingo Scholtes	10
120	Yijian Huang	40	449	Francesco Rullani	10
703	Frank Schweitzer	33	120	Michael Schwind	10
506	Jeffrey Heer	30	578, 23	Andy Meneely	9
971	Gregorio Robles	29	657	Ashish Sureka	8
971	Jesus M. G. Barahona	28	564, 473	Andrea M. Magdaleno	7
612	Stephan Diehl	21	971	Juan Martinez-Romo	6
158	Robin Teigland	17	822	Jorge Colazo	5
703	Claudio J. Tessone	16	703	Marcelo Serrano Zanetti	4
453	Daniela Godoy	14	459	Olivier Berger	4
506	Brandon Heller	13	158	Paul M. Di Gangi	4
453	Silvia Schiaffino	13	657	Ayushi Rastogi	2
564, 473	Renata Mendes de Araujo	12	150	Peng He	2

3.3. Rede de Co-autoria

Considerando os 18 estudos primários mapeados por este trabalho, foram identificados 51 autores, implicando em uma média de 2,83 autores por artigo. A Tabela 4 mostra a lista dos autores que possuem fator de impacto H_i scholar.

A Figura 2 mostra a distribuição dos autores em função dos estudos primários, e a partir dela vemos a participação de alguns autores em mais de uma publicação. Foi o caso dos artigos [564] e [473], [449] e [237], e por fim, [23] e [578].

4. Resultados e Discussão

Nesta seção as perguntas de pesquisa definidas pelo protocolo serão respondidas e analisadas com base nas evidências apresentadas pelos estudos primários.

4.1. Quais os principais métodos de pesquisa utilizados?

Os seguintes métodos de pesquisa foram identificados nos artigos selecionados: Pesquisa Quantitativa e Estudo de Caso. Dos 18 artigos mapeados, 6 deles se enquadraram como pesquisa quantitativa, outros 12 se enquadraram como estudo de caso (ver Tabela 5).

Para este trabalho, foram considerados como estudo de caso aqueles artigos que tem como objeto de estudo um projeto em específico. Por outro lado, aqueles classificados

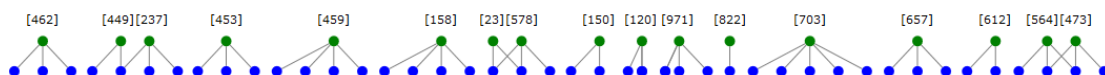


Figura 2. Distribuição dos autores (em azul) em função dos artigos (em verde)

Tabela 5. Métodos de pesquisa mapeados

Método de Pesquisa	Estudos Primários	Quantidade
Estudo de caso	[23], [158], [237], [453], [459], [462], [564], [578], [612], [657], [703], [971]	12 (66,7%)
Pesquisa Quantitativa	[120], [150], [449], [473], [506], [822]	6 (33,3%)

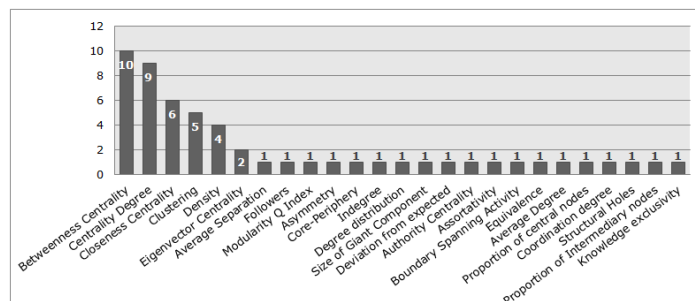


Figura 3. Frequência de uso das métricas de ARS

como pesquisa quantitativa não analisaram um projeto em específico, mas sim vários projetos, escolhidos aleatoriamente. [449], por exemplo, usou dados de 170.706 projetos hospedados no *SourceForge*.

4.2. Quais as principais métricas de análise de redes sociais utilizadas?

Dentre as métricas identificadas, as mais representativas foram: *Betweenness Centrality*, *Centrality Degree* e *Closeness Centrality*. A Figura 3 mostra a frequência de uso das métricas.

4.3. Quais os objetivos de pesquisa mapeados nos estudos primários?

Esta questão buscou categorizar os estudos primários de acordo com seus objetivos de pesquisa. As categorias foram criadas a partir da descrição dos objetivos de pesquisa dados pelos próprios autores em seus trabalhos. Ao todo foram identificadas oito categorias, cuja frequência está ilustrada pela Tabela 6 e descrição pela Tabela 7.

5. Conclusão

Este trabalho apresentou um mapeamento sistemático sobre ARS aplicada à colaboração em projetos de software. Através de quatro engenhos de busca, 1129 artigos únicos foram considerados pela fase 1 do mapeamento. Após aplicar critérios de inclusão/exclusão, 26 artigos foram identificados por aplicar ARS em colaboração de projetos (considerando projetos de qualquer natureza). Destes 26, um total de 18 era do contexto de projetos de software, e apenas estes foram considerados pela análise do mapeamento sistemático.

De acordo com os objetivos de pesquisa explicitados nos estudos primários, eles foram classificados em oito categorias, sendo as duas mais significativas: Acompanhar a dinâmica de colaboração e Identificar padrões de colaboração. De fato, espera-se que a ARS possa ajudar os líderes de projetos a entender como funciona a dinâmica de colaboração em um projeto de software.

5.1. Trabalhos Futuros

Espera-se que futuras pesquisas tenham como fonte de dados além da academia ou repositórios, uma presença significativa de dados da indústria. Para tanto, é necessário um maior envolvimento e a colaboração efetiva entre empresas e pesquisadores.

Tabela 6. Referência dos objetivos com os estudos primários

ID	Categoria	Estudos Primários	Quantidade
C01	Acompanhar a dinâmica de colaboração	[150], [453], [459], [462], [564], [971]	6 (33,33%)
C02	Identificar padrões de colaboração	[120], [473], [506], [612], [703]	5 (27,77%)
C03	Relacionar a topologia de redes colaborativas com resolução de problemas	[237], [578]	2 (11,11%)
C04	Relacionar a topologia da rede colaborativa com a performance	[822]	1 (5,55%)
C05	Identificar riscos, ameaças e vulnerabilidades na rede	[657]	1 (5,55%)
C06	Relacionar a topologia de redes colaborativas com inovação	[449]	1 (5,55%)
C07	Identificar a relevância de alguns indivíduos do grupo	[158]	1 (5,55%)

Tabela 7. Descrição das categorias

C01	Nesta categoria os artigos propuseram ferramentas para mensurar a colaboração na rede social formada pelos membros de um projeto de software. A ideia é fornecer informações úteis, sobretudo, aos gestores do projeto.
C02	Os artigos classificados nesta categoria analisaram um conjunto de projetos e buscaram notar padrões de colaboração, utilizando para tanto, métricas de ARS.
C03	Esta categoria buscou relações entre a topologia de uma rede social formada por desenvolvedores e a sua capacidade de resolução de problemas.
C04	O artigo desta categoria analisou 101 projetos de software de código aberto e buscou relação entre aspectos estruturais da rede social formada pelos desenvolvedores e performance, sobretudo em termos de produtividade e qualidade.
C05	O artigo desta categoria utilizou análise de redes sociais para procurar riscos, ameaças e vulnerabilidades na rede social formada pelo desenvolvimento de um projeto de software de código aberto. Por exemplo, pessoas que são críticas em um time de desenvolvimento, por serem únicas em dominar técnicas ou ferramentas.
C06	O artigo desta categoria analisou a estrutura social formada pelos integrantes de vários projetos de código aberto hospedados no <i>SourceForge</i> . O objetivo é relacionar a estrutura da rede social com potencial de inovação no projeto.
C07	O único artigo desta categoria buscou analisar a importância de indivíduos com comportamento empreendedor dentro da estrutura social de uma comunidade de software de código aberto.
C08	Esta categoria buscou validar algumas métricas de ARS. Seu objetivo foi confrontar os resultados obtidos pela aplicação das métricas contra a percepção das relações sociais notadas pelos integrantes de alguns projetos de código aberto.

Como continuação deste trabalho, serão reunidos esforços para entender o atual cenário do uso de ARS para analisar colaboração em projetos de software, bem como seus principais benefícios e limitações.

Referências

- CHISAN, J. and DAMIAN, D. (2004). Towards a model of awareness support of software development in gsd. page 28.
- ELLIS, C. A., GIBBS, S. J., and REIN, G. (1991). Groupware: some issues and experiences. 34(1):39–58.
- FITSILIS, P. et al. (2009). Using social network analysis for software project management. *Current Trends in Information Technology (CTIT)*, pages 1–6.
- FUKS, H. et al. (2007). The 3c collaboration model. pages 637–644.
- KADUSHIN, C. (2012). *Understanding social networks: Theories, concepts, and findings*. Oxford University Press.
- KITCHENHAM, B. (2001). Guidelines for performing systematic literature reviews in software engineering.
- SCHWIND, M. and WEGMANN, C. (2008). Svnmat: Measuring collaboration in software development networks. pages 97–104.
- WASSERMAN, S. (1994). *Social network analysis: Methods and applications*. Cambridge university press, New York.