

The Influence of the Cooperation of Developers on OSS Project Success from the Perspective of Complex Network

Peng Liu, Jiawei Zheng*, Nianxin Wang

School of Economics and Management, Jiangsu University of Science and Technology, Zhenjiang 212003, China

*Corresponding author. Email: jiawei8683@163.com

ABSTRACT

The success of OSS projects is closely related to the collaborative network formed by the spontaneous interaction of members within the community. Although some studies have discussed about that, they only focus on the complex network characteristics presented by the developer collaboration network, so the research in this area still needs to be further deepened. In this research, we construct the developer collaboration network by collecting 418 projects data from GitHub community. After considering the influence of the characteristics of the project itself. Our results suggest that the clear degree of technical division of labor among developers, The strength of collaboration between each other and The proportion of developers involved in the collaboration are positively correlated with the project's success. while the degree of differences in the status about division of members negatively correlated with the success of the project. This discovery helps us to better improve the probability of the success of open source projects.

Keywords: Complex network, Developer collaboration, Open source software, Project success.

复杂网络视角下成员分工协作对OSS项目成功的影响研究

刘鹏, 郑佳伟*, 王念新

江苏科技大学经济管理学院, 镇江, 江苏, 中国.

*通讯作者. 邮箱: jiawei8683@163.com

中文摘要

开源软件的成功与社区内成员自发交互形成的协作网络是密不可分的.已有研究虽然对这方面进行了探讨,但只是针对开发者协作网络呈现出的某种复杂网络特性,因此这方面研究还有待进一步深入.本文通过收集 Github 社区中 418 个项目数据并构建开发者协作网络,在考虑了项目自身特征因素影响之后,我们的结果表明成员技术分工的明确程度、彼此间的协作强度以及参与到协作中的人员比例均正向影响项目成功,而成员地位分工的差异程度负向影响项目成功.这一发现有助于我们可以更好的提高开源项目成功的可能性.

关键词: 复杂网络, 分工协作, 开源软件, 项目成功

1. 引言

近十多年来,开源软件得到广泛应用.为了探索开源软件成功的奥秘,学者们尝试着从诸多角度进行研究,比如项目特征^[1]、许可范围^[2]、意识形态^[3]等等,但这些研究隐含的依赖于开源团队非社会化观点,即很大程度上脱离了它的社会背景.由于需要依靠成员自发的分工协作来完成,但这种具有内在自治性与复杂性的分工协作是如何影响到项目的成功目前还没有确切的定论.

Crowston 等人^[4]发现核心成员和外围成员的沟通量有助于项目成功;Rullani 等人^[5]则发现核心成员和外围成员的不同分工对项目的成功都很重要,但这些研究仅探讨了网络中特定位置的成员分工协作,因此存在一定的不足.随着复杂网络理论的兴起,有学者通过分析成员之间的交互结构来解决这一问题,例如 Singh^[6]通过收集 Sourceforge 上 15 个开源软件社区的日志文件和邮件通讯数据来构建开发者协作网络,研究结果表明协作网络不同程度的小世界特性与项目成功具有相关性,但仅关注协作网络的某种复杂网络特性,并未深入探讨成员分工协作是如何影响到项目的成功.

基于此,本文通过收集 Github 代码托管平台的项目和代码提交数据,利用代码协作关系构建开发者协作网络,并对其网络结构属性特征展开分析,并进一步探讨成员分工协作是如何影响到项目成功,本文力图从复杂网络角度深入探讨影响开源项目成功的关键因素.

2. 研究假设

2.1. 成员分工对项目成功的影响

在实际开发过程中,学者研究发现,开发人员在工作上的层级结构和专业分工显得较为明显^[7].即在同一项目中,不同的个体常常围绕不同的模块进行代码提交完成协作工作,这种因技术水平而形成的分工使得成员逐渐清楚自己的任务并形成较为稳定的社群,团队内部成员都能发挥自己的专长.同时由于成员在资源控制方面的差异性使得成员在组织内部产生了隐性的“管理结构”^[8],这种因资源分配引起的地位分工的差异性易造成成员之间因资源分配不合理使得成员在进行活动时受到很大限制,长期下去,那些处于边缘的开发者可能会产生反感情绪,从而减少贡献甚至退出团队,基于此,本文提出以下假设:

H1a: 技术分工的明确度正向影响项目成功

H1b: 地位分工的差异度负向影响项目成功

2.2. 成员协作对项目成功的影响

Georg 等人^[9]表示开源项目的 Bug 修复和新功能的推出非常依赖于开发者之间的协作沟通,良性的协作

沟通有利于高质量信息的出现和传播.我们认为成员间的协作可以促进成员之间的学习交流^[10].然而如果参与协作中的成员过少,并且彼此之间的协作不足则会阻碍有效知识的产出,从而影响成员学习.基于此,本文提出以下假设:

H2a: 成员协作强度正向影响项目成功

H2b: 参与到协作中的成员比例正向影响项目成功

3. 数据获取与处理

3.1. 数据来源

我们利用 Github 代码托管平台的 418 个项目的 git 数据作为构建开发者协作网络的基础,每条提交数据包含了提交者的邮箱地址、提交时间、代码修改情况及所涉及的文件等信息,本文主要通过社区开发者之间的代码协作关系构建开发者协作网络.

3.2. 变量定义

因变量:我们使用 CVS 提交的数量作为项目成功的度量,并已被多数学者用于相关的研究^[2,6].

自变量:技术分工的明确度和地位分工的差异度在复杂网络视角下表现为网络整体的模块化(MD)程度和中心势(DC)高低^[7],本文参考 Georg 等人^[9]计算网络的模块度,并参考 Wasserman^[11]等人的方式计算网络点度中心势.成员协作强度与参与到协作中的人员比例在复杂网络视角下表现为网络节点平均度(AD)和最大连通子图大小占比(LOC)^[9].点度表示与目标节点直接相连的节点数量,连通子图表示网络中一部分直接或间接相连节点的子图,其中规模最大的连通子图称为最大连通子图.

控制变量:为了保证我们结果的可靠性.我们控制了项目规模(Size),先前的研究表明不同规模的项目在吸引贡献者与协调新任务等方面有些差异^[1];我们也控制项目是否使用了 GnuGPL (Gnu Public License),因为先前的研究表明,许可证的选择会影响用户的兴趣^[2];我们还控制项目在 Github 上发布的时间(Time),以季度(90天周期)来衡量时间,从项目创建之时,即第0季度开始,直到2019年底前的最后一个完整季度;除此之外我们还控制了项目的语言数量(Language),虽然涉及多种语言的项目可能表现出更强的适应性,但是协调不同语言对于成员来说也有很大挑战^[12];我们还控制了项目核心开发者的数量(NoOfCorD),我们将核心开发者定义为在项目期间代码贡献量高于平均代码贡献量的开发者,有研究表明核心开发者数量也会影响到项目的成功^[9];最后为了使得研究更为严谨,本文还控制了协作网络的小世界特征强度(Q),已有研究表明开发者协作网络的小世界特性有助于促进成员进行沟通进而做出贡献^[6],对于小世界特征强度计算采用了 Singh 等人的界定方法^[6].

4. 研究结果

经检验,自变量间的相关系数远小于 0.700,变量间相关性不显著.进一步计算方差膨胀因子和容忍度指标,结果显示方差膨胀因子小于上限阈值 10,容忍度大于下限阈值 0.100,表明研究模型不会出现多重共线性问题.

表 1 层次线性模型预测项目成功

Variables	Ln(commit)		
	Model1	Model2	Model3
Intercept	-0.069	-0.072	-0.064
Time	0.255**	0.094*	0.191**
Size		0.241**	0.208**
NoOfCorD		0.035	0.055
Language		0.05	0.015
Gnu Public License		0.149***	0.119*
Q		0.163**	0.124*
MD			0.127*
DC			-0.099*
LOC			0.211**
AD			0.157**
R ²	0.109	0.374	0.459
Adjust R ²	0.105	0.359	0.433
ΔR ²		0.265	0.084

Significance levels: * p≤0.05, ** p≤0.01

由表 1 数据可知,多数控制变量均对项目成功产生显著影响,这与现有的文献结果基本吻合.在成员分工方面,从模型 3 可以看出技术分工的明确度正向影响项目成功($\beta=0.127, p<0.05$),这是因为专业化分工使得成员更加清楚自己的角色定位从而表现出更加稳定的贡献行为进而促进项目成功,故 H1a 成立.同时成员地位分工的差异程度越大,项目越不容易取得成功($\beta=-0.099, p<0.05$),即较高的地位差异度不利于成员做出贡献,故 H1b 成立,由于扁平化的组织结构更易于信息流通,从而可以提高成员的学习效率和处理问题的速度,进而有利于项目的成功;在成员协作方面,从模型 3 可以看出成员协作强度和协作参与度均正向影响项目成功($\beta=0.211, p<0.01; \beta=0.157, p<0.01$),故假设 H2a, H2b 均成立,协作强度与协作人数的增多容易刺激成员想法的诞生和技术的贡献,成员之间可以快速的寻求新的问题解决方案,这些开发人员的贡献主要体现在具体的代码开发上--包括项目原有代码的审核、修订,以及新代码的提供.

5. 意义与局限性

理论上,本文基于复杂网络视角考察了开源软件项目成功的影响因素,既是对已有研究的补充:先前针对网络结构对项目成功的影响主要是基于开发者分工协作网络呈现出的某种复杂网络特性.其次,本文对不同领域的开源项目成员分工协作对项目成功的影响展开实证分析,分析结果有助于探究成功的项目是否

存在共性的分工协作模式,并以此来指导项目管理者可以更加科学的引导项目走向成功.实践上,对于开源软件开发本身来说,了解其开发模式成功的影响因素有利于社区管理者在项目的发展阶段对其施加干预,改进开源软件项目的建设,引导开源软件项目走向成功.

本文尚存在一些局限,值得进一步深入研究.第一,本文主要侧重从基本的网络结构特征出发来考察对项目成功的影响研究,忽略了结构以外的因素对项目成功的影响,这可能使得研究结果存在一定的缺陷.第二,本文使用的是 Github 社区中的项目为例进行实证研究,由于衡量开源项目成功的指标存在差异,因此可能结论和发现不具备普适性.

致谢

基金项目: 国家自然科学基金面上项目(71871108)资助

REFERENCES

- [1] Subramaniam C, Sen R, Nelson M L. Determinants of open source software project success: A longitudinal study[J]. Decision Support Systems, 2009, 46(2): 576-585.
- [2] Singh P V, Phelps C. Networks, Social Influence, and the Choice Among Competing Innovations: Insights from Open Source Software Licenses[J]. Information Systems Research, 2013, 24.
- [3] Daniel S L, Maruping L M, Cataldo M, et al. The Impact of Ideology Misfit on Open Source Software Communities and Companies[J]. MIS quarterly, 2018, 42(4):1069-1096.
- [4] Crowston K, Shamshurin I. Core-Periphery Communication and the Success of Free/Libre Open Source Software Projects[J]. Journal of Internet Services and Applications, 2017, 8(1):10.
- [5] Rullani F, Haefliger S. The periphery on stage: The intra-organizational dynamics in online communities of creation[J]. Research Policy, 2013, 42(4):941-953.
- [6] Singh P V. The small-world effect: The influence of macro-level properties of developer collaboration networks on open-source project success[J]. ACM Transactions on Software Engineering & Methodology, 2010, 20(2): 1-27.
- [7] Aoxiang Xia, Xiao Zhang, Xingzhou Zhang. Study on collaborative network of OpenStack OSS developers [J]. System Engineering Theory and Practice, 2017, 37(5): 1373-1382.

- [8] Seidel M D L, Stewart K J. An Initial Description of the C-Form[J]. *Research in the Sociology of Organizations*, 2011, 33:37-72.
- [9] Georg, von, Krogh, et al. Community, joining, and specialization in open source software innovation: a case study[J]. *Research Policy*, 2003.
- [10] Dahlander L, O'Mahony S. Progressing to the Center: Coordinating Project Work[M]. *INFORMS*, 2011.
- [11] Wasserman S, Faust K. Social network analysis theory and applications[J].
- [12] Medappa P K, Srivastava S C. Does Superposition Influence the Success of FLOSS Projects? An Examination of Open-Source Software Development by Organizations and Individuals[J]. *Information Systems Research*, 2019, 30(3):764-786.