Journal of High Technology Management Research 20 (2009) 52–61

Contents lists available at ScienceDirect

Journal of High Technology Management Research

Social networks and coordination performance of distributed software development teams

Liaquat Hossain [a,*], David Zhu [b]

[a] Project Management Graduate Programme, The University of Sydney, NSW 2006, Australia
[b] Accenture, Sydney, Australia

ARTICLE INFO

Available online 16 April 2009

Keywords:
Coordination
Social networks
Performance
Software development
Defect management
OSS (Open Source Software)

ABSTRACT

In this study, we explore the coordination performance of the geographically distributed software development teams by exploring OSS (Open Source Software) development dataset available through SourceForge.com. OSS team structures have traditionally been geographically dispersed and therefore, the coordination of post release activities such as testing efforts have been carried out by means of communication via electronic forms, such as email or message boards and forums. In our current communication-enriched environment, best practices for coordination are adopted by all software projects yet some still fail to achieve their target performance. Does team structure have any bearing on the performance outcome of the project? How does the communication between teams and their external parties affect ultimate success or failure of projects? We seek to answer above questions by applying existing theories and analytical methods from social networks for exploring the coordination performance of defect management activities found in OSS projects. We propose social networks based theoretical model for exploring distributed coordination structure and apply that for the case of OSS defect management process for exploring the structural properties, which induce the greatest coordination performance. The outcome of our suggest that there is correlation between certain network measures such as density, centrality and betweenness and coordination performance measures of defect management systems such as quality and timeliness.

1. Introduction

The Open Source approach to software development involves a group of loosely knitted volunteers to collaborate over a public medium of communication, most popularly the Internet, to create software (de Souza, Froehlich, & Dourish, 2005). The source code is open to public access and it is this readily available, which result in faster and more responsive development cycles, thus producing more robust and secure software. These inherent characteristic has some interesting connotations to OSS project teams and how the software is tested. What is phenomenal about an OSS projects, is that its participants tend to form a community that is "bounded by their shared interest in using/developing the system" (Ye & Kishida, 2003).

In this study, the particular coordination activities we are interested are related to bug reporting, fixing and knowledge sharing post release date of the OSS. The general focus of existing studies have been coordination between developers during pre-release phase of the project with minimal attempts to include OSS project community's involvement. We suggest that the evolution of any OSS project originates from the input or identification of defects from the OSS project's community. We explore questions such as: Does the degree of centrality, betweenness and density of the network have any bearing on the number of defect fixed per software promotion? Does the degree of centrality, betweenness and density of the network have any bearing on the number of defects reported at different severity levels? Does the degree of centrality, betweenness and density of the network have any bearing on

* Tel.: +61 2 9306 9110; fax: +61 2 9351 8642.
E-mail address: l.hossain@usyd.edu.au (L. Hossain).

1047-8310/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.hitech.2009.02.007

the average number of days for a defect to be fixed for each project team? Does the degree of centrality, betweenness and density of the network have any bearing on the average number of days for responses by developers on defects?

2. Towards a social network based coordination approach

The conceptual origin of social network analysis can be traced as far back as the 1930's when leading 'Gestalt' theorists Jacob Moreno devised the 'sociogram' to represent the formal properties of social configuration (Scott, 2000). Moreno's (1930) sociogram (as cited in Scott, 2000) depicted individuals using points and their social relationship with others by lines, he argued using this type of representation researchers could identify leaders(stars) and isolated individuals(isolates), to uncover asymmetry and reciprocity and to map chains of connection. In modern studies, researchers has taken used Moreno's (1930) suggestions and analysed networks based on either team structure. (Pearce & David, 1983) concluded there to be a definite relationship between 'organisation design variables and group performance' from their study of two types of organizational design models each with their own setup of stars, isolates and liaisons and levels of reciprocity.

The framework of social network analysis was finally crystallized in the 1950's through the works of fieldworkers associated with the Department of Social Anthropology at Manchester University. Clyde Michell's can be attributed to turning the existing studies of graph theory used for sociometric analysis and formulating it to fit a specific sociological framework (Scott, 2000). John Barnes conducted intensive studies on the part kinship, friendship and neighbouring relationships played in the formation of communities. Barnes claimed that 'the whole of social life' could be set as 'a set of points some which are joined by lines' to form a 'total network of relations'. The informal sphere of interpersonal relations can be seen as a part of a 'partial network', of this total network (Barnes, 1954). The final breakthrough that solidified the global properties of SNA (social Networks Analysis) in all fields of social life was made at Harvard by Harrison White and his associates. The first part was the development of algebraic models of groups using set theory to model relationships and the second was the development of multidimensional scaling, which was a technique for translating relationships into social distances and for mapping social space, which was the key idea behind Lewins work (Scott, 2000).

The earliest example of combining theories of networks with coordination can be found in research by Alex Bavelas in the early 1950's. The experiment was devised so that individuals had to find other person who possesses the card with the same symbol as they did. The way of communication was through written messages and the group was setup in different formations of networks. The result of their research was that groups in the star or Y networks solved the problem the quickest, with fewest errors and least mistakes and those at the center of the star or Y network recorded the highest level of satisfaction (Kosfeld, 2003). de Souza et al. (2005) conducted on social network based investigation into distributed software development with its focus using artifacts and activities to determine the social structure of the software project (de Souza et al., 2005). With regards to examining distributed coordination, Sandusky and Gasser (2005) observed the impact of 'negotiation' as a mechanism of coordination in distributed software problem management systems (SWPM). The research was centered on publicly accessible bug repositories for OSS and negotiation in the SWPM process (Sandusky & Gasser, 2005). It is important to note that from the research it was found that OSS community bug repositories represent the major coordination mechanism in OSS projects. Finally, Gutwin, Penner et al. (2004) studied the coordination of distributed software development focusing on group awareness with the aim of unearthing how such geographically dispersed group stay aware of each other work and coordinates the development of large complex software projects. The result of the study outlined the importance of organizational culture and text based communication tools to maintain awareness (Gutwin, Penner et al., 2004). From the literature reviewed, it is evident that focus of the work has been conducted in the field of distributed software development of OSS has primarily been on pre-release coordination activities. For example, the focus of bug repositories study by Sandusky and Gasser (2005) was negotiation and not coordination measures. This leaves scope for investigating social network properties and their implications on coordination measures of OSS teams.

2.1. Structural variables of social networks for understanding distributed coordination of OSS

There are certain variables of social networks that will be the focus of our analysis for the coordination in OSS projects. There are some general concepts and definitions that will be extensively used in the analysis sections that need to be firstly introduced. These definitions will help form the foundation of our social network based analysis of the OSS projects and their wider community. Social networks analysis is deeply rooted in graph theory, however it is important to distinguish that a graph in this context is 'simply a set of lines connection points' and graph theory consists of a 'body of mathematical axioms and formulae that describe the properties of the patterns formed by the lines' (Scott, 2000).

- Two points are said to be adjacent if a line connects them
- Those points which are adjacent to a particular point is said to be the neighbourhood
- The total number of points within a point's neighbourhood is said to be its degree of connection
- Points can be connected directly by a single line or indirectly connected by a sequence of lines. The sequences of lines is said to be a walk
- A walk in which each line and point is distinct is called a path
- The length of a path is said to be the number of lines that make up the path
- The distance between two points is the length of the shortest path between them.