An Analysis of Open Source Software Development Using Social Network Theory and Agent-Based Modeling

Greg Madey
Computer Science & Engineering
University of Notre Dame
gmadey@nd.edu

Vince Freeh
Computer Science
North Carolina State University

Renee Tynan
Management
Mendoza College of Business
University of Notre Dame

Chris Hoffman
Mendoza College of Business
University of Notre Dame

Abstract

The open source software (OSS) development phenomenon appears to be a self-organizing process with emergent properties. Such processes are difficult to understand because emergent properties are by definition difficult to predict using traditional modeling and analytical techniques. An approach under evaluation is to use agent-based simulation techniques to study the OSS phenomenon. We are using the Swarm library and the Java programming language to model the self-organizing processes seen in the OSS phenomenon. A record of all events of interest is stored in a database during the simulation for post simulation analysis and comparison with other runs of the simulation. This permits analysis of process data, in addition to outcome data, generated by each simulation. Data mining techniques are applied to the process and outcome data across multiple simulations to identify self-organizing and emergent phenomenon.

We have collected data on OSS projects from several online OSS collaboratories. We define two software developers to be connected — part of a collaboration social network — if they are members of the same project, or are connected by a chain of connected developers. Project sizes, developer project participation, and clusters of connected developers are analyzed. We find evidence to support our hypothesis, primarily in the presence of power-law relationships on project sizes (number of developers per project), project membership (number of projects joined by a developer), and cluster sizes.

In our model open source software developers are agents. Each is an instance of a Java class with methods that encapsulate a real developer's possible daily interactions with the

development network. Developers can create, join, or abandon a project each day or continue their current collaborations. A separate Java method models each of the first three possibilities. A fourth method encapsulates a developer's selection of one of the three alternatives. Post simulation analysis and comparison with our empirically derived data on the OSS phenomenon is used to calibrate simulation parameters for model refinement.