An Investigation into the Free/Open Source Software Phenomenon using Data Mining, Social Network Theory, and Agent-Based Simulation

Greg Madey
Associate Professor
Computer Science & Engineering
College of Engineering
University of Notre Dame

UIUC - NSF Workshop on Continuous (Re)Design of
Open Source Software
University of Illinois, Urbana-Champaign
October 8-9, 2003

Abstract

We report on results from an ongoing study of the Free/Open Source Software (F/OSS) phenomenon at the community level. Publicly available data about F/OSS projects, developers, processes, and their relationships have been collected from SourceForge.net. Numerous statistical results, including the existence of many power-law relationships, are presented. The F/OSS community is modeled as a collection of ad hoc, social networks consisting of heterogeneous agents, self-organizing into projects and clusters of projects. A computer simulation of the F/OSS community model is developed using SWARM, an agent-based simulation toolkit. Empirical data, obtained using web-mining and data-mining techniques, are used to parameterize the simulation, which in turn is used to investigate hypotheses about processes and mechanisms leading to F/OSS community formation. The quantitative data, the model, and the simulation may offer insight into F/OSS project coordination and communication.