This dissertation contains their essays on the study of open source software collaboration. The first essay, titled “Overview of Open Source Software Development”, highlights the link between developer incentives under the non-monetary setting and the sustainability issue in the open source community. Motivated by the well-known fatal flaw Heartbleed in the widely used open source network security framework OpenSSL, Chapter 1 discusses the lack of manpower in the open source community from the perspective of developer incentives. I introduce the empirical setting of this project, GitHub, one the most popular open source development platforms. Both the past literature as well as the social network features on GitHub suggest that the motivations to contribute to open source projects could include (but not limited to) ego stratification, reputation, and career concern.

The second Chapter, “An Empirical Case Study of Open Source Software Community”, provides descriptive statistics of the contribution history data for Python projects from GitHub, and establishes several empirical patterns of open source contribution to showcase the lack of manpower in the community. I find that the distribution of contribution of a project follows power law, in the sense that most projects have very few contributors and number of commits while very few projects have a large number of contributors and commits. In addition, during the eight years from 2010 to 2018, the growing rate of new projects appears to be faster than new contributors.

In the last Chapter, titled “A Structural Model of Decentralized Open Source Software Development”, I build a dynamic discrete choice model to characterize individual developer preferences when choosing an open source project to contribute to. Using finite-dependence (Arcidiacono and Miller, 2011, 2019b,a), I identify and estimate the parameters of an individual developer’s utility function. The estimation results show that developers prefer to contribute to popular projects. At the same time, they prefer their own contribution to not be “diluted” by their peers. This is the first study, to the extent of my knowledge, to build and estimate a dynamic discrete choice model to characterize open source software developer preferences. Lastly, given the estimated structural parameters, I conduct counterfactual analysis by increasing the expected popularity of projects except for new projects. The results show that the number of contributors choosing not to commit would decrease under the counterfactual regime.