Our research projects aims to determine how the structure of the social network and which characteristics of developers involved in the creation of Open Source software favor creation of innovation in the Open Source community. By *innovations*, we mean new software and improvements which are created through cooperation among developers.

Our analyses will be based on the unique database, obtained by *web scrapping* GitHub, which is currently one of the largest repository services used for Open Source software development. The database contains publicly available information about the developers registered on GitHub. Developers are related to each other in various ways (e.g., they express interest in new activity by other developers, new activity about specific projects, or develop software cooperatively), hence they can be represented as nodes of a social network.

Utilizing the data above we will check whether the developer's reputation (measured e.g. as the number of people who have asked to be informed about the developer's activity, or the number of distinctions awarded to the projects) increases the probability of generating innovations by cooperating with the developer. We will also study to what extent the phenomenon of *homophily* affects the creating of innovations, i.e., whether programmers who have similar characteristics are more likely to cooperate.

The first task is to organize the raw data obtained from GitHub and publicly accessible databases: GHTorrent and GithubArchive. Information about developers is publicly available, but it is distributed among various sites, and may contain errors. It needs to be aggregated and cleaned up.

The second task is to prepare our data set for the needs of the analytic models estimated in our research. We have to harmonize and aggregate the data. Furthermore, we will create a hyperbolic map of the social networks. Application of hyperbolic geometry to represent social networks is a relatively new approach, which allows to describe relationships in the social networks which are created based on a combination of popularity and similarity between nodes.

In the third task, we will study how the probability of creation of innovation between two developers depends on the structure of the social network. We will also study the characteristics of programmers which favor collaboration with them.

The four task is to validate the results obtained in the previous tasks. We will collect a second snapshot of the social network, and try to predict whether an innovation between two programmers has been created, based on the previously obtained models. We will compare the predicted graph with the real one. This step will allow us to verify the extent of possible generalization of our results.

Already existing studies of the Open Source community were usually based on samples of 100-300 developers, working on a specific project. There were almost no studies which applied econometrics and the theory of social networks to analyze the community. The presented project aims to fill these gaps. Our project is interdisciplinary: it combines computer science, economy, and other social sciences, and is based on a unique database representative for the Open Source community.