# MLSM Project Proposal

## Predicting Pull Request Acceptance on GitHub from Social Factors

# Problem Statement

Research has previously been done to understand organizational learning and participation in communities of practice. With the rise of sites like GitHub, we can empirically identify what social factors contribute to acceptance in open source software developer networks. In particular, we are interested in predicting the likelihood that a first time contributor to a project will be accepted.

# Hypothesis

- Social factors can be used to accurately predict whether or not a pull request from a first time contributor will be accepted by project maintainers on GitHub.
  - Community members who have been active in previous discussions are more likely to have their code changes accepted.
  - Community members who have other popular projects are more likely to have their code changes accepted.
  - Community members who talk like project maintainers are more likely to have their code changes accepted.

# Data

- GitHub API is used to retrieve closed pull requests for five popular repositories.
- Data about the user who opened the pull request — their participation in previous pull request threads and the popularity (starred) repositories — will be used.
- The comments by the user who opened the pull request and the user who closed it will be compared.

# Methods

- A user’s contributions will be a count of pull request threads he participated in before submitting the current pull request.
- A user’s popularity will be a proportion of how many of his repositories have at least N stars.
- A linguistic similarity score will calculated based on comments by the contributor and the maintainer.
- These inputs variables will be used in a logistic regression model.

# Related Work

- Von Krogh, Georg, Sebastian Spaeth, and Karim R. Lakhani. "Community, joining, and specialization in open source software innovation: a case study." Research Policy 32.7 (2003): 1217-1241.
  - Mixed methods research developing a theory of open source participation.
  - Only looks at one project.
  - Propose idea of “joining script.”

# Related Work

- Bagozzi, Richard P., and Utpal M. Dholakia. "Open source software user communities: A study of participation in Linux user groups." Management science 52.7 (2006): 1099-1115.
  - Pulls from social psychology theories to describe participation in Linux user groups.
  - Tries to explain social identity of different types of members.

# Related Work

- Bryant, Susan L., Andrea Forte, and Amy Bruckman. "Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia." Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work. ACM, 2005.
    - Descriptive study of Wikipedia focusing on theoretical explanations of how users become contributors.

# Timeline

- 10/15: Data Collected
- 10/18: Data Formatted
- 11/05: Method implemented
- 11/12: Analysis of results
- 11/19: Presentations
- 11/26: Writeup of results