Computer Project II

Mining Chromium repository

Project details

- Writing scripts to mine Chromium's version control system
- Identifying top developers
- Identifying mostly edited file sets
- Building socio-technical network of Chrome project

- Team should consist of 3-4 people

Chromium project

- https://www.chromium.org/chromium-os/chromiumos-design-docs/source-code-management
- As version control system, they use Git.
- Check the chromium project in Github to learn more on where and how they store the source code.

Mining version control system

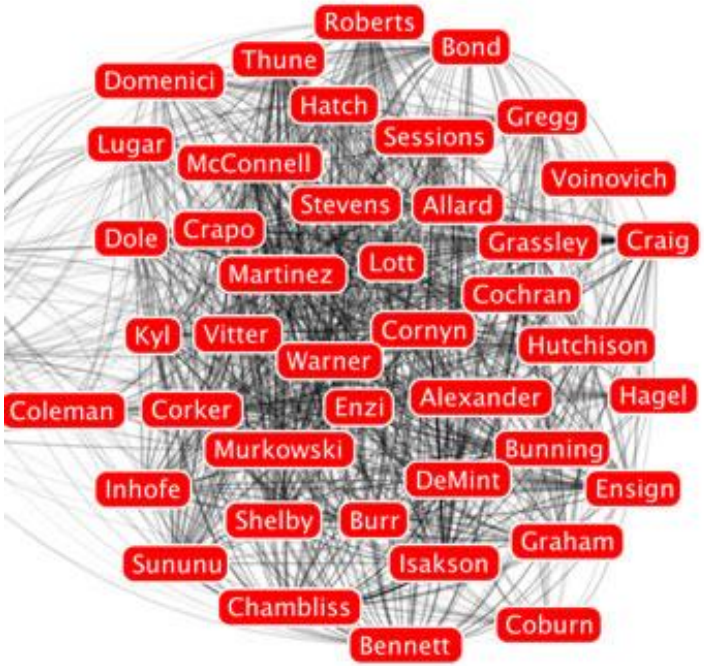
- Learn how to mine a project's development history from git
- Choose a particular time of development (e.g. last six months) OR a release OR a component/directory
- Get all commits done on that time/release/component/directory

Extracting useful data

- Learn to extract socio-technical information:
  1. Social: List of developers who contributed to the project
  2. Technical: Collaboration between developers in terms of contributed files

Visualizing social-technical network

- Use a graphic toolbox (e.g. SocialAction- http://www.cs.umd.edu/hcil/socialaction)
- Nodes will represent Developers
- Edges will represent Collaborations between Developers (in terms of commonly edited files)
- Nodes can grow to indicate top developers who has the most collaboration with the others
- Edges can be thicker based on the number of files contributed together



Graphic taken from http://www.cs.umd.edu/hcil/science20/

More on socio-technical dependencies in Chromium data

- Andrew Meneely, Alberto C. Rodriguez Tejeda, Brian Spates, Shannon Trudeau, Danielle Neuberger, Katherine Whitlock, Christopher Ketant, & Kayla Davis, An Empirical Investigation of Socio-technical Code Review Metrics and Security Vulnerabilities, Proceedings of the 6th International Workshop on Social Software Engineering pp. 37–44, 2014

- Felivel Camilo, Andrew Meneely, & Meiyappan Nagappan, Do Bugs Foreshadow Vulnerabilities? A Study of the Chromium Project, 2015 International Working Conference on Mining Software Repositories