Existing BigBang notebooks:

(for MailingLists)

* analyze senders --> distribution of top-senders, including smart ways to deal with name inconsistency
* assortativity study --> computes degree assortativity on email reply network (whether users tend to interact with likely central users)
* auditing fernando --> follow a single user’s contribution over time
* cohort analysis --> trend over time of new users joining the discussion
* collaboration robustness --> various analysis on how relationships between users evolve
* correlation between betwenness centrality and community membership --> whether the importance of a user in brokering information flow is related to the user belonging to a definable type of community
* list correlation --> [not really clear yet] does something with correlation between time series and community detection
* participants over time --> time series on email senders
* plot activity --> various examples of how to explore a mailing list archive: load full data; plot evolution of email activity per day; plot top-senders; merge very similar senders’ names
* show interaction graph --> graph of interactions in the mailing lists
* single word trend --> trend of a specified word in a single mailing list
* special words analysis --> various interesting ways to identify special words in the mail texts: unique / common words between mailing lists; words introduced by different people in different lists; words introduced by same people in different lists; words that quickly flow from a user to others
* testing power law response time hypothesis --> check if the distribution of response time in their mailing list fits a power law [spoiler: no]
* threads / threads-research-in-progress --> some analyses about threads in a mailing list: number of threads; distribution of messages per thread; duration of a thread; properties of a single thread
* Word type classifier --> [not clear yet] perform some analyses on words shared by two distinct mailing lists

(for GitHub)

* committer dominance --> identify key-contributors to projects
* git commit network --> create graph where nodes are commits and edges connect parent and child commits
* git diffs --> create bipartite graph where nodes are committers and files, and edges are files edited by committers
* git interaction graph --> create graph of collaborations between contributors of a repository
* git loading tutorials --> tutorial on how to interact with GitRepo object
* MultiGitRepos --> allows to collect data from multiple repositories, and performs: bi-partite graph with nodes as committers and files and as committers and repositories; projected graph of repositories sharing committers

(combining Mailing Lists and GitHub)

* git collection --> compare trends in commits and emails; identifies top-committers; distribution of contributions by cohorts of committers
* walkers and talkers --> [really funny one!] compares the activity of contributors to a project in terms of commits and in terms of emails: tests whether those who talk a lot are also those who work or not