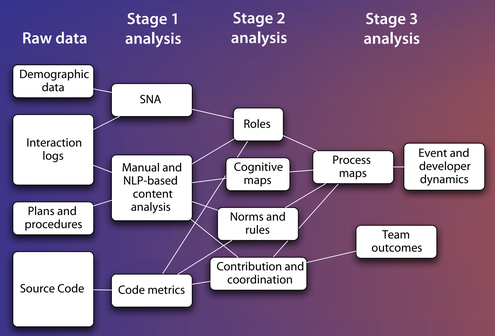
Investigating the Dynamics of FLOSS Development Teams

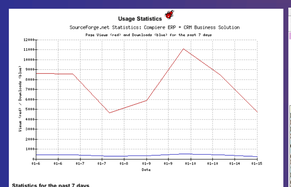
What are the dynamics through which self-organizing, distributed teams develop and work?

Research partially funded by NSF grant 05–27457, with prior support from 03–41475 & 04–14468

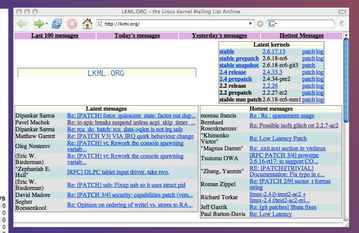
Analysis Plan

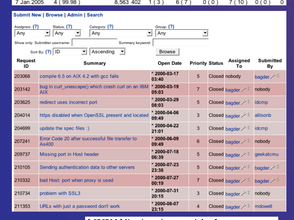


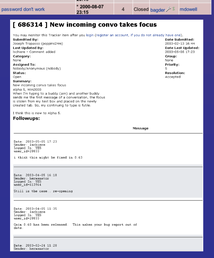
Raw Data



Date	Rank	Page Views	DI	Bugs	Support	Patches	All Trakr	Tasks CVS
15 Jan 2005	12 (99.93)	4,746	258	0 (0)	1 (0)	0 (0)	1 (0)	0 (0) 0
14 Jan 2005	11 (99.94)	5,469	414	1 (0)	6 (8)	0 (0)	7 (9)	0 (0) 0
10 Jan 2005	6 (99.97)	11,081	491	2 (1)	6 (6)	0 (0)	8 (7)	0 (0) 0
9 Jan 2005	4 (99.98)	5,893	340	0 (0)	0 (2)	0 (0)	0 (2)	0 (0) 0
8 Jan 2005	3 (99.99)	4,641	267	0 (0)	1 (1)	0 (0)	1 (1)	0 (0) 0
7 Jan 2005	4 (99.98)	8,563	402	1 (3)	6 (7)	0 (0)	7 (10)	0 (0) 0







Data Integration

People with their events

- Person ID: 48 (identified by sf_user_id: **993558**, real_name: **Jeff Johnson**, sf_username: **johnson1234**, email: **jgjohnso@wisc.edu**)

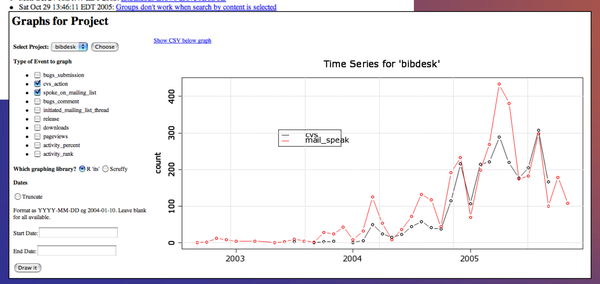
Participated in 34 events (of all kinds)

- bugs_submission: 22
- bugs_comment: 12
- release: 0
- cvs_action: 0
- initiated_mailing_list_thread: 0
- spoke_on_mailing_list: 0

Events in datetime order:

- Mon Mar 08 22:01:13 EST 2004: 1053361 on bugs 912358
- Tue Jun 21 10:33:12 EDT 2005: Accents in macros not displayed
- Tue Jun 21 10:40:03 EDT 2005: Crashed on changing Preview Pane preference
- Tue Jun 21 16:32:26 EDT 2005: 1682440 on bugs 1224896
- Thu Jul 07 21:40:07 EDT 2005: TypeInfo Editor doesn't quit edit mode
- Thu Jul 07 22:07:01 EDT 2005: Author sorting for two-word last names
- Sat Jul 09 16:03:43 EDT 2005: 1099934 on bugs 1234554
- Sat Jul 23 18:19:46 EDT 2005: Save changes dialog with no changes
- Wed Aug 31 20:39:21 EDT 2005: BibDesk 1.1.5 hangs on New Publication with Crossref
- Wed Aug 31 20:53:58 EDT 2005: BibDesk crashes when entering "/" character in raw BibTeX
- Tue Sep 06 11:31:35 EDT 2005: Gaps in list sorted by Booktitle
- Thu Sep 29 09:52:36 EDT 2005: Blank Container fields appear first in sorted list
- Wed Oct 05 19:32:13 EDT 2005: Preview pane stops updating after cancelled Close
- Sat Oct 08 20:16:37 EDT 2005: 1791318 on bugs 1308021
- Thu Oct 13 16:46:20 EDT 2005: Sorting not fully sticky
- Thu Oct 20 12:25:49 EDT 2005: Width of new Keywords/Group pane not sticky
- Mon Oct 24 16:51:47 EDT 2005: Extraneous arrows above scroll bar
- Sat Oct 29 13:46:11 EDT 2005: Groups don't work when search by content is selected

Graphs for Project



Research Plan

1. Manual examination of transcripts to locate evidence of concepts
2. Exploration of dynamic Social Network Analysis techniques
3. Specialization of NLP techniques for novel text types (e.g., chat logs)

Refining application of NLP techniques

Manual refining of identified features according to conceptual framework

Extraction of code and SNA metrics from growing sample

Utilizing NLP, SNA and Code metrics on large sample

Clustering of projects through optimal pattern matching of timelines

Comparison of dynamics across projects to provide generalizable knowledge

2006 2007 2008

Engagement with FLOSS community advisory board

Key Findings To Date

- FLOSS projects display highly distributed levels of centralization. Some projects that appear decentralized when examined at a snapshot in time are actually centralized, but with the individuals at the centre changing over time. However, the data suggest that the majority of projects analyzed maintain a single participant at the centre through their lifetimes.
- We identified different trajectories of decision making practices in two comparable projects, one successful and one not. In one, that trajectory connotes acceleration and energy (growth in number of participants, shorter decision time, more inclusive participation, and richer and more complex decision-making episodes), while the other shows signs of deceleration and entropy (shrinking participation, disappearing administrators, longer decision cycles, and increasingly simple and less complex decision-making episodes).

1 Manual examination of transcripts

Example: Coding team decision episodes

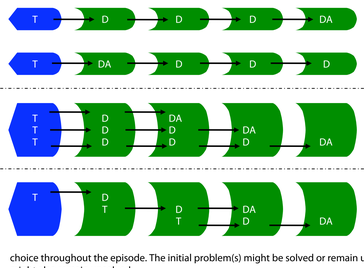
We defined a decision episode as a sequence of messages that begins with a message containing a trigger that presents an opportunity for choice (such as feature request, bug report or strategic problem). It includes discussion related to the issue, and a decision concerning the stated opportunity.

Decision Making Coding Scheme—Episode level

Item	Code	Description
Decision Type	Code	Central issues related to a change in the code/software, or an acceptance of a patch or lines of code that will become part of the code base
	Non-code	Central issues related to something other than code, e.g., legal issue, membership issue, funding, group maintenance etc
Trigger Type	Problem	Problem/crash without knowing exactly what causes it.
	Bug	Clear bug identification and bug report
	Patch	Patch or lines of code submission
	Feature	Clear identification of a desired functionality or change in code
	Releases	Release-related issue
	Mixed	Multiple, mixed issues listed together, but not related to a release
	Non-Code	Any problem / issue that isn't resolvable by a change in the code of the software being built, and complex or major issues that will eventually, but not immediately affect the code base, including major rewrites or redesign, voting, organizational changes, legal issues.

Preliminary results

Coded episodes from 6 project teams: 3 instant messenger clients and 3 enterprise software projects.
Selected 60 episodes from each project: 20 from start, 20 most recent and 20 surrounding a major release.



Type A (Single)

The episodes involve a single choice that responds to a single straightforward trigger

Type B (Multiple-simple)

The episodes respond in a linear fashion to a trigger that contains multiple opportunities for choice

Type C (Multiple-complex)

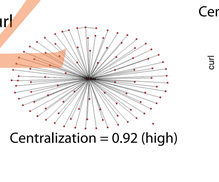
Regardless of the nature of the initial trigger, in these episode, new, sometimes unrelated triggers created additional opportunities for choice throughout the episode. The initial problem(s) might be solved or remain unsolved, and the new problems introduced might also remain unsolved.

T: Trigger; D: Discussion; DA: Decision Announcement

2 Exploration of dynamic social network analysis

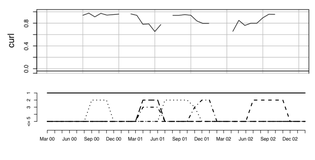
We analyzed the communications network in project bug tracker logs and found that projects exhibited a range of centralizations. We then examined the networks as they evolved over time and found evidence in some projects of changes in the central members.

Curl

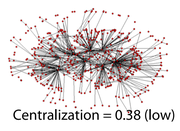


Centralization = 0.92 (high)

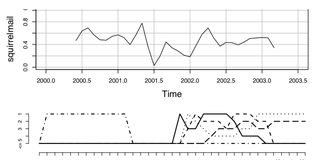
Centralization and most central individuals over time



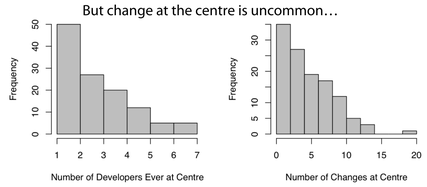
SquirrelMail



Centralization = 0.38 (low)



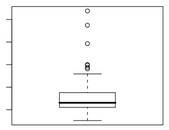
But change at the centre is uncommon…



…and few participants remain with projects

Most projects have participants active for only 10-20% of lifetime of the projects

Less than 4 have participation for > 33% of lifetime



3 Specialization of natural language processing techniques

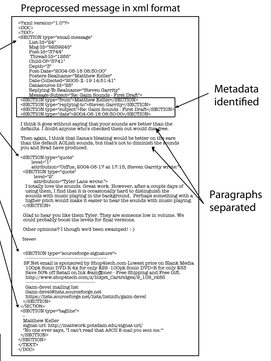
PreProcessing

The first step in creating structure from unstructured or semi-structured text is to prepare a text for further processing

- Identification of major sections of an email message, eg header, signature, tagline
- Removal of ">" symbols that interfere with sentence detection
- identification of quote levels embedded in messages and author of the quoted snippets

Original message (slightly edited)

Preprocessed message in xml format



Emoticons – finding them, categorizing them

Original Text: > Other opinions? I though we'd been swamped! :-)

Before – each element of punctuation is a separate token:

<sentence sid="S29"> I|PRP though|IN we|PRP have|VBD be|VBN swamp|VBN !|. </sentence>
<sentence sid="S30"> :|: -|:)|SYM >|SYM >|SYM </sentence>

After : Emoticon is pulled together as a single token. Each is categorized as neutral, positive, or negative. Part of speech is ignored for now.

<sentence sid="S43"> I|PRP though|IN we|PRP have|VBD be|VBN swamp|VBN !|. </sentence>
<sentence sid="S44"> <SYM cat="positive"> :-) |JJ </SYM> </sentence>

Automatic Coding – exploratory work

Group Maintenance/ Task Assignment

GAIM: doc55
frametype = entity
text = Task-Assignment
Ask-to-an-Unspecified-Person = Please submit any patch against cvs , it makes our life alot easy , some of what you describe makes it sound like you might be coding against 0.59.
senteceid = S27

Decision Announcements

GAIM: doc55
frametype = entity
text = Decision-Announcement
senteceid = S20
Decision-Announcement-Fix = This is fixed in cvs as far as I know .

Fire: doc1403
frametype = entity
text = Task-Assignment
senteceid = S88
Making-Results-Available = Here is a small patch for handling non-breaking spaces

Fire: doc1773
frametype = entity
text = Task-Assignment
senteceid = S25
Self-Assignment = I will take a look at fixing this

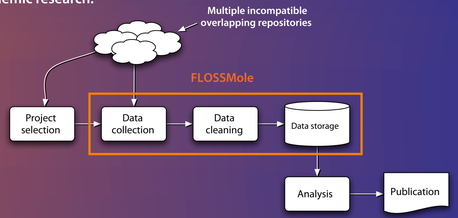
GAIM: docNT217
frametype = entity
text = Decision-Announcement
senteceid = S14
Decision-Announcement-Commit = I disagree (which is why I committed the change after KingAnt suggested it) .

People

not shown:
Keisuke Inoue
Sarah Harwell
Steven Rowe
Nancy McCracken

Products

- floss.syr.edu website with copies of papers and analyses
- FLOSSmole, a collaborative project to gather, share and store comparable data and analyses of free and open source software development for academic research.



Plans

- Continued qualitative analysis of email transcripts, including analysis of group maintenance, power and perhaps leadership
- Refining application of NLP techniques to automate aspects of qualitative analysis
- Development and extraction of software source code metrics
- Quantitative event sequence and time series analysis of processes
- Field study of shared mental models in FLOSS teams
- Literature review of empirical studies of Open Source Software development

Citations

Crowston, K., Wei, K., Li, Q., Eseryel, U. Y., & Howison, J. (In press). Self-organization of teams in free/libre open source software development. *Information and Software Technology Journal*, Special issue on Understanding the Social Side of Software Engineering, Qualitative Software Engineering Research, Accepted with major revisions.

Heckman, R., Crowston, K., Li, Q., Allen, E., Eseryel, Y., Howison, J. & Wei, K. (2006). Emergent decision-making practices in technology-supported self-organizing distributed teams. In *Proceedings of the International Conference on Information Systems (ICIS 2006)*. Milwaukee, WI, 10–13 Dec.

Howison, J., Inoue, K., & Crowston, K. (2006). Social dynamics of FLOSS team communications. In *Proceedings of The Second International Conference on Open Source Systems*, Como, Italy, 8–10 Jun. (First-runner up for best paper in conference.)

Li, Q., Crowston, K., Heckman, R., & Howison, J. (2006). Language and power in self-organizing distributed teams. Paper presented at the OCIS Division, Academy of Management Confer

Partners

- Barbara Scozzi, Politecnico di Bari, Italy
- Megan Conklin, Elon University, North Carolina