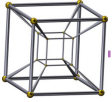




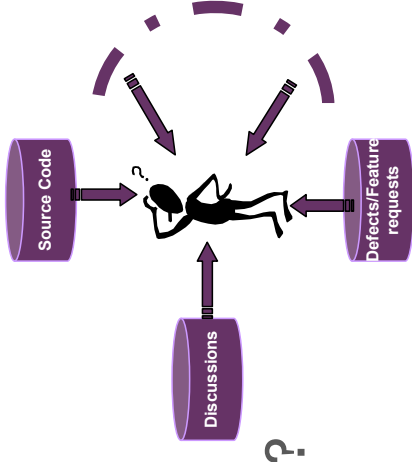
Tesseract: Interactive Visual Exploration of Socio-Technical Relationships in Software Development

Anita Sarma, Larry Maccherone, Patrick Wagstrom, and James Herbsleb
Institute for Software Research, School of Computer Science
Carnegie Mellon University

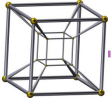


Some Questions in Software Development

- Who do I go to for help?
- Which other artifacts are affected by my changes?
- Which developer is affected by my changes?
- Whose changes are affecting my changes?
- Who should be assigned to this task?
- Which tasks need to be completed before the other?
- Which artifacts are brittle or buggy?
- ...



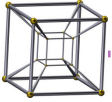
Answering these questions is non-trivial



Problem

- Need for coordination arises because of complex relationships among project elements
- Data is siloed
- These relationships change over time
- Social and technical relationships are tightly coupled
- Lack of interactive exploratory environment for software projects

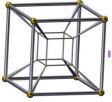




Objective

- Provide an interactive project exploration environment
 - aggregate data across data different sources
 - cross-link and visualize relationships
 - present changes in relationships over time
 - treat social and technical relationships as first order objects
- Allow the 'lay user' to explore relationships and project dynamics

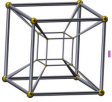




Research Challenges

- Where to get the information?
- How to deal with scale?
- How to meaningfully extract and relate linkages?
- Allow investigating a particular problem
- Which information should be displayed?
- Help find interesting patterns

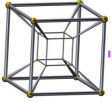




Tesseract

- Interactive and explorative environment to correlate and understand the complex relationships among:
 - code
 - developers
 - communication records
 - tasks (issues/ features)
 - time





Tesseract

Project
gnome/rhythmbox

Search 8685

DoubleClick node to drill down upon all selected.

Show Settings

Files

File name
data/ui/.cvsignore
data/views/.cvsignore
doc/.cvsignore
help/C/.cvsignore
rhythmdb/rhythmdb-property-moc
rhythmdb/rhythmdb-property-moc
rhythmdb/rhythmdb-query-model
rhythmdb/rhythmdb-query-model
rhythmdb/rhythmdb-tree.c
rhythmdb/rhythmdb-tree.h
rhythmdb/rhythmdb.c
rhythmdb/rhythmdb.h
rhythmdb/rhythmdb-entry-model.

Developers

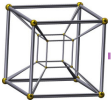
Name
christian neuma
william jon mcci
benjamin otte
carlos perello m
yann bizeul
bastien nocera
colin walters
abel cheung
james willcox
christophe ferg

Issues

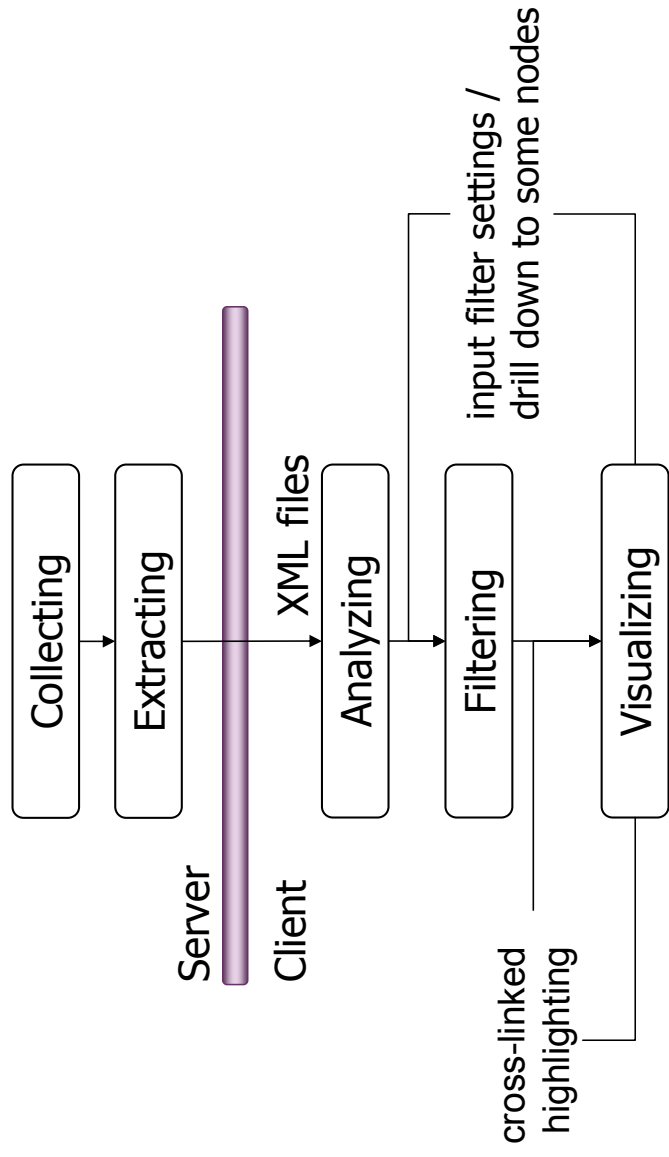
Severity: Enhancement Trivial Minor Normal Major Critical Blocker

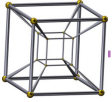
Files

Bug ID	Start	End	Severity	Priority	Status	Resolution	Description
7360	2004-04-01	2004-12-21	critical	Normal	RESOLVED	FIXED	player crashes starting s
7361	2004-04-01	2005-01-13	critical	Normal	CLOSED	INCOMPLETE	Crashes when using mu

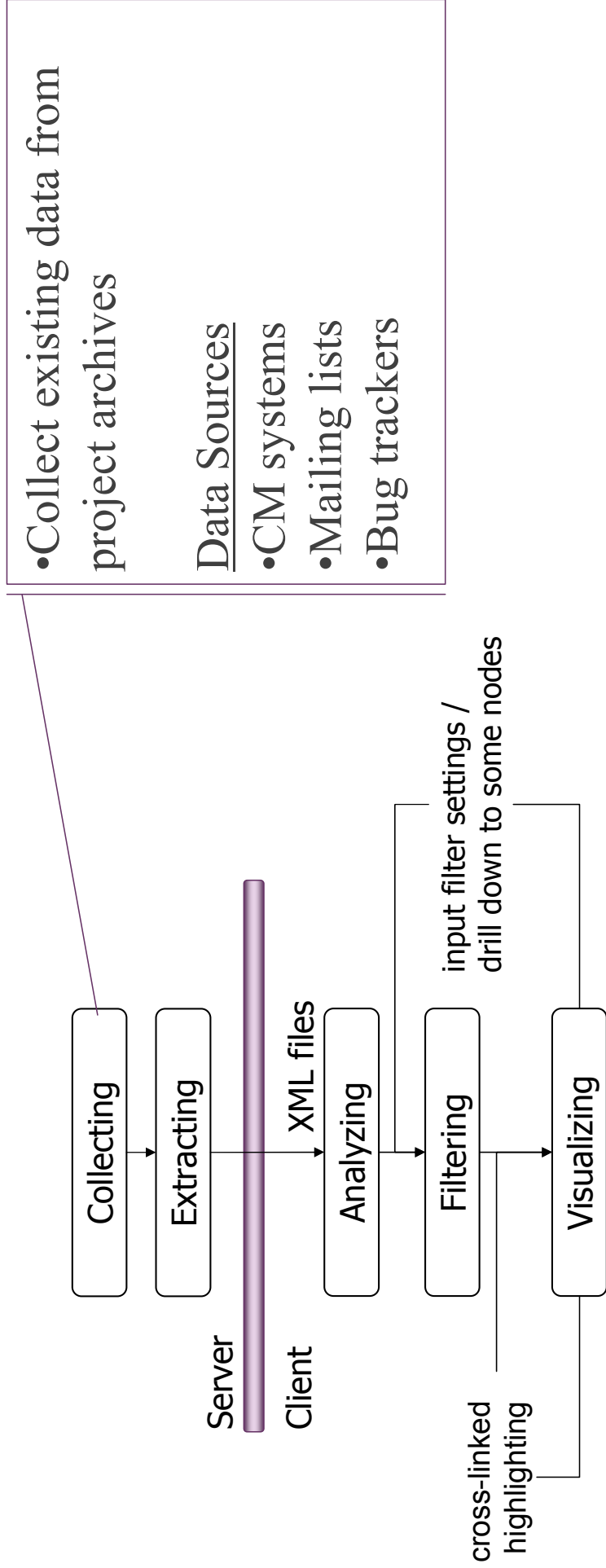


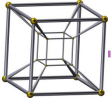
Information Flow



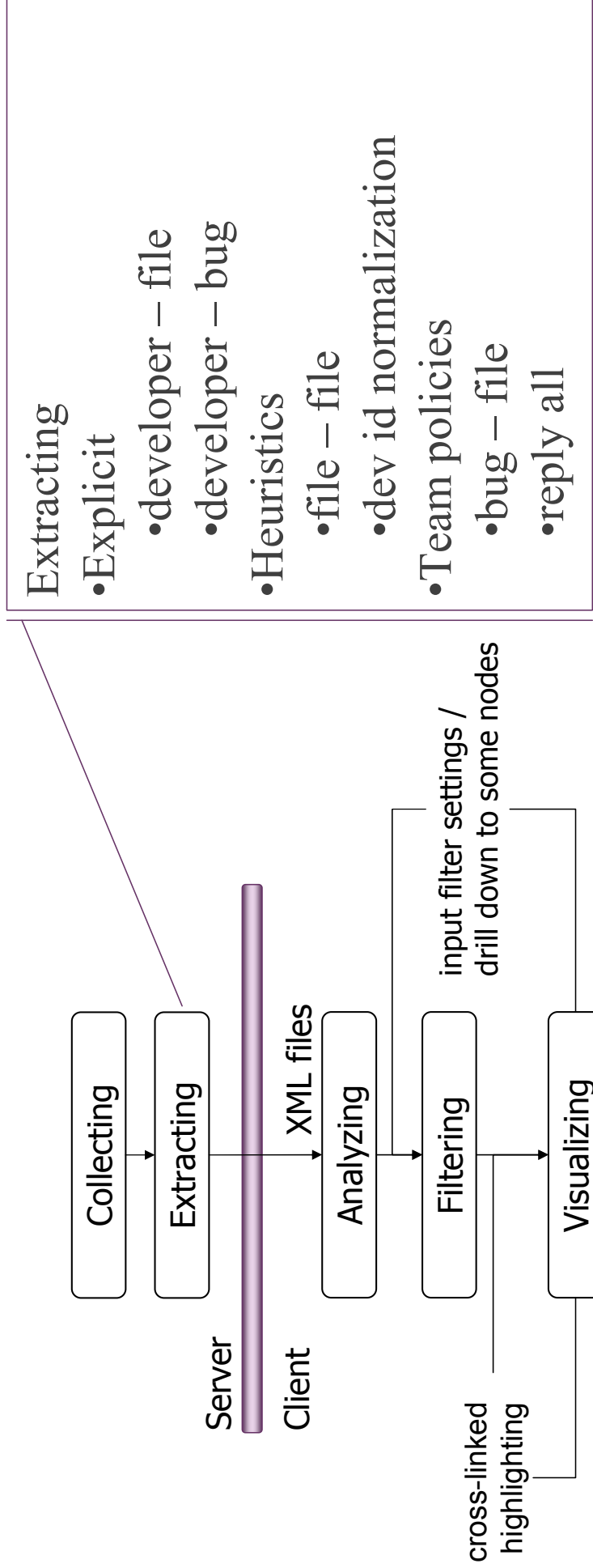


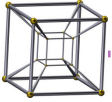
Information Flow



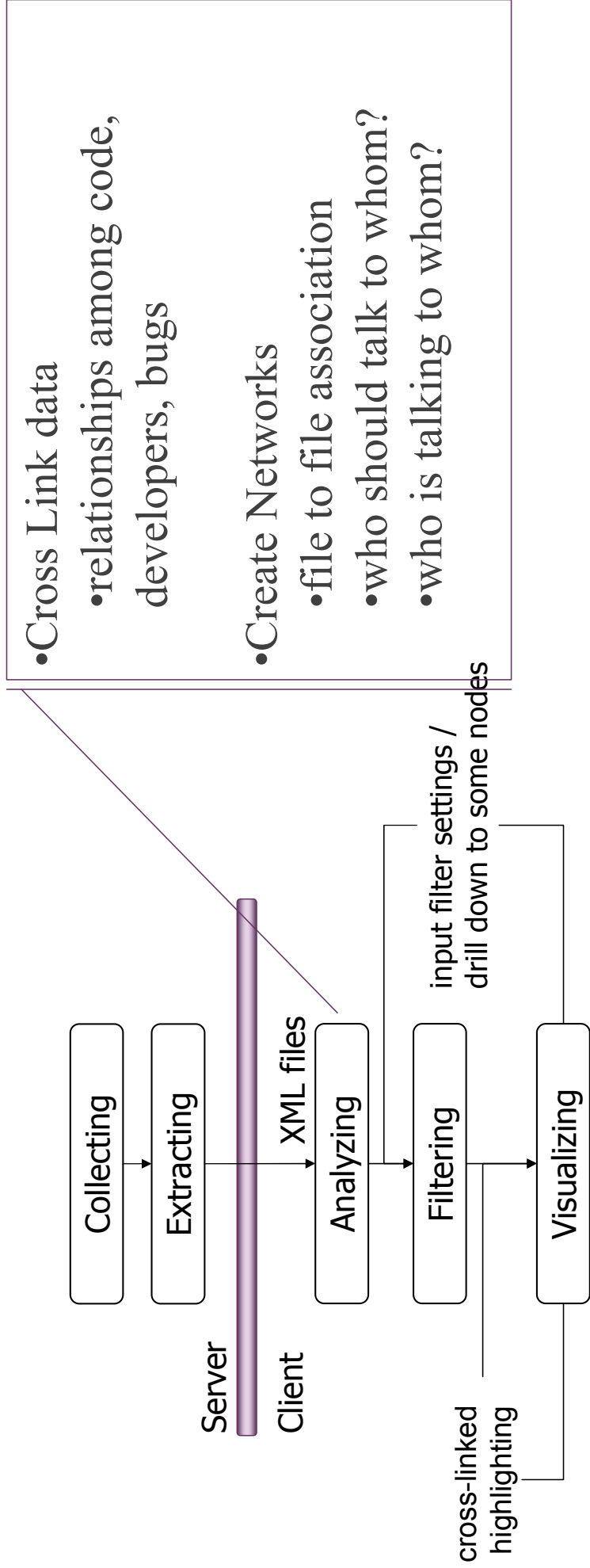


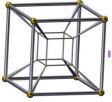
Information Flow





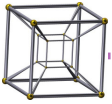
Information Flow





File Association

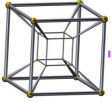
- Logical Coupling
- Frequently co-committed files are logically associated with each other (Gall, Hajek, Jazayerri 1988)
- Especially useful in situations where
 - code base contains different programming language files
 - call site separated from target (e.g., network connection or even transmitted by event bus)
- Better measure of dependency for our purposes (Cataldo et al. 2006, 2008, 2009)



Developer Testimony

“The implicit dependency stuff, that, I think could be really useful in and of itself. So things that which end up being changed together but don't necessarily have an inheritance relationship, or compositional -- knowing that, I've changed this thing it looks like something in isolation, but in reality whenever someone changes something here, these thirty other things change because of **some ripple effect, that would be useful...”**



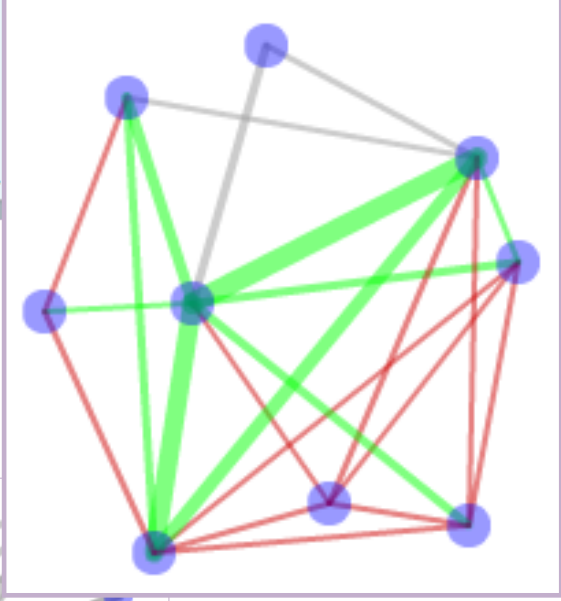
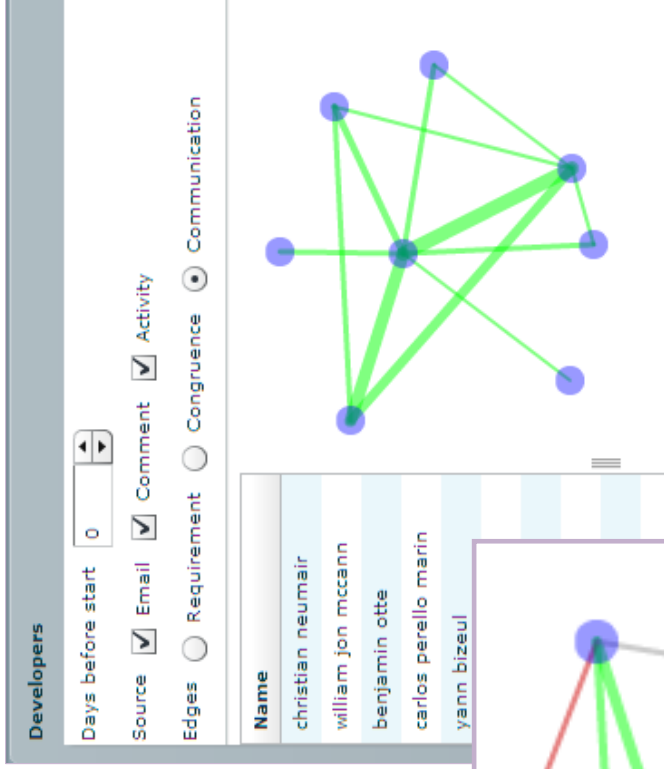
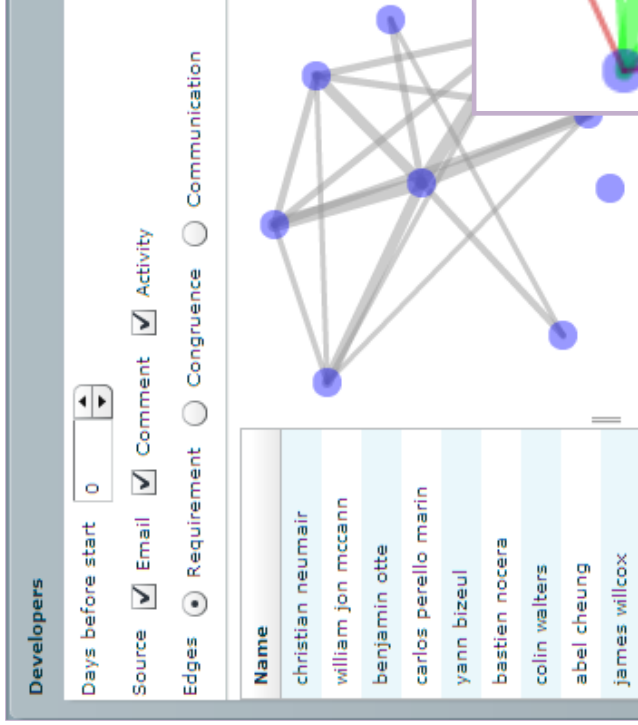


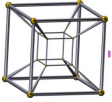
Developer Network



Coordination Requirements

Communication Pattern

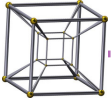




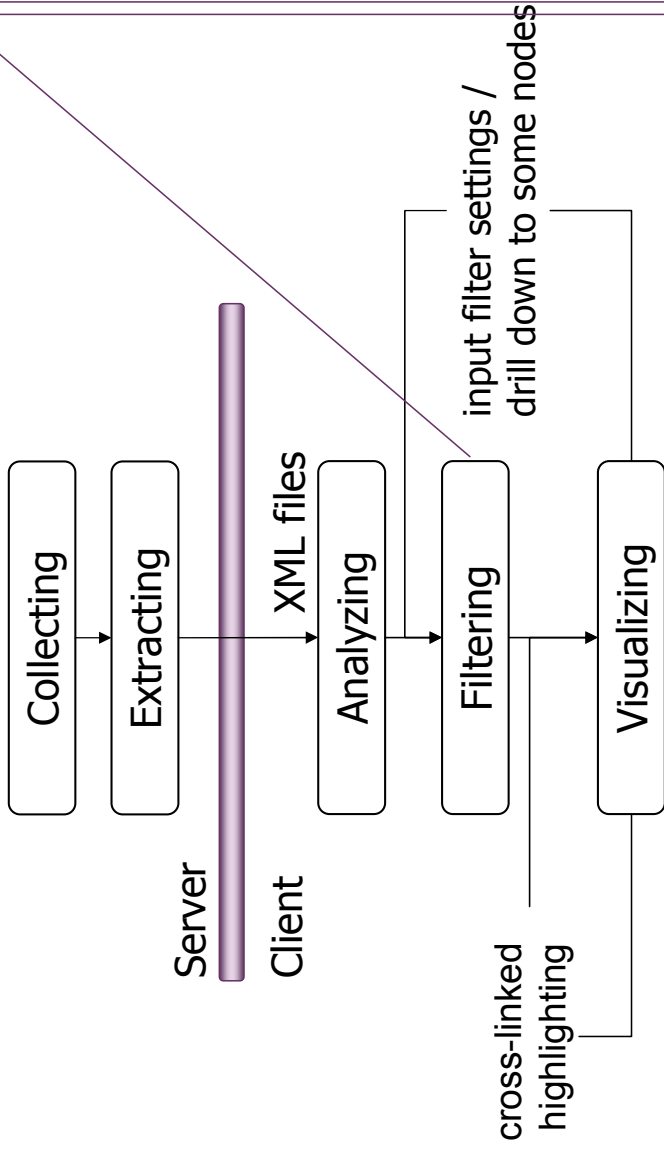
Developer Testimony



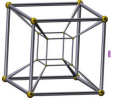
“this [developer pane] is a project manager view.
What I know is, I am this person, three people have
red flag and one person has green flag. **My
dashboard says you need to talk to
[developer] because he made these
changes...**”



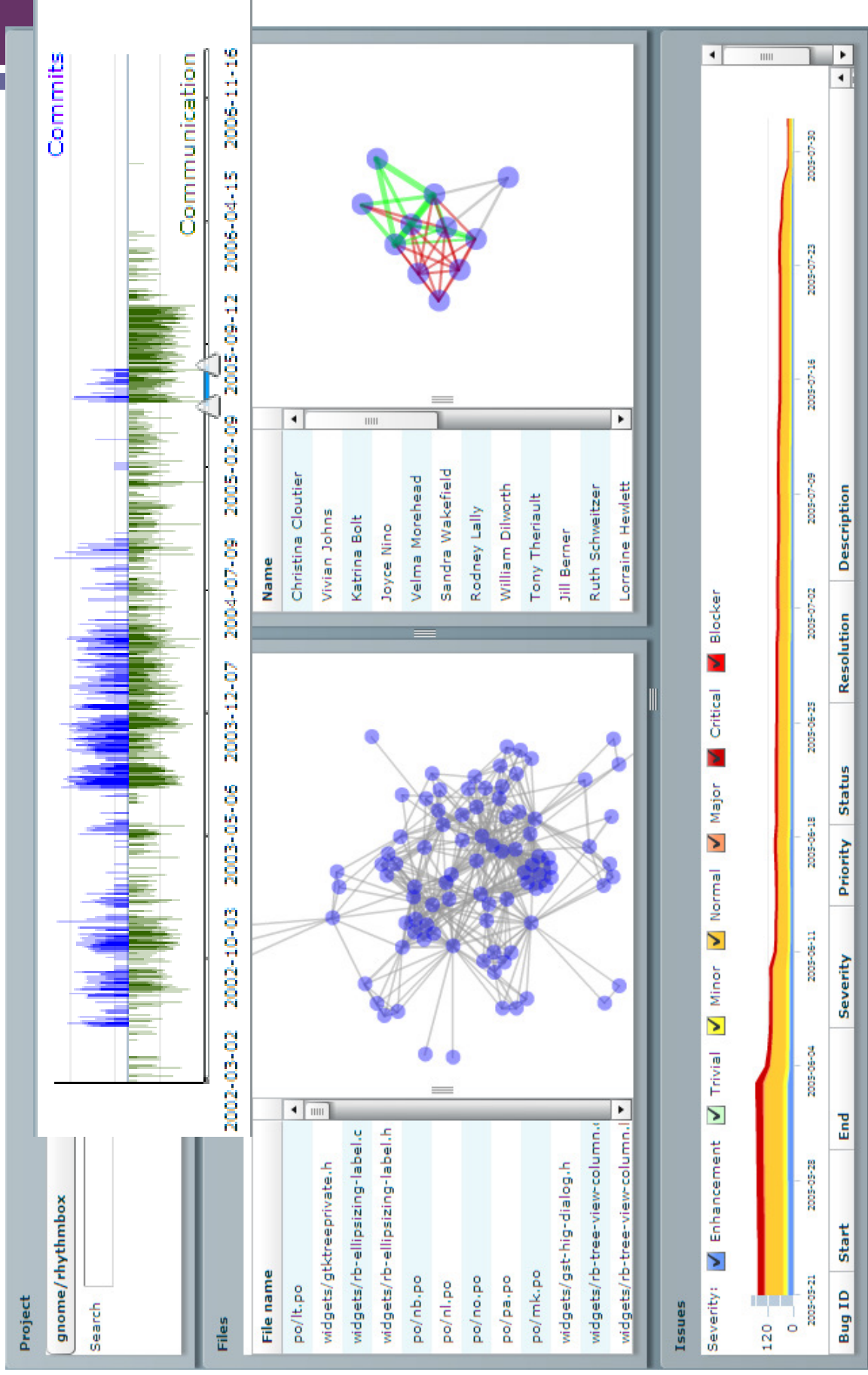
Information Flow

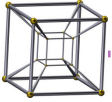


- Project activity view
- Drill down
- Thresholding
 - density of file, developer association
 - total commits in a set
 - file types
 - communication type
- Text search



Tesseract Visualization

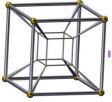




Tesseract Visualization

The screenshot displays the Tesseract project visualization interface, which is divided into several sections:

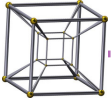
- Project:** Shows the project name "gnome/rhythmbox" and a search bar. A tooltip indicates: "DoubleClick node to drill down upon all selected."
- Commits:** A bar chart showing commit activity over time from 2002-03-02 to 2008-11-16. The chart has two series, one in blue and one in green.
- Files:** A list of files committed together, with a threshold of 20. The list includes:
 - po/lt.po
 - widgets/gtktreeprivate.h
 - widgets/rb-ellipsizing-label.c
 - widgets/rb-ellipsizing-label.h
 - po/nb.po
 - po/nl.po
 - po/no.po
 - po/pa.po
 - po/mk.po
 - widgets/gst-hig-dialog.h
- Developers:** A list of developers:
 - Christina Cloutier
 - Vivian Johns
 - Katrina Bolt
 - Joyce Nino
 - Velma Morehead
 - Sandra Wakefield
 - Rodney Lally
 - William Dilworth
 - Tony Theriault
 - Jill Berner
 - Ruth Schweitzer
 - Lorraine Hewlett
- Issues:** A bar chart showing issue activity over time from 2005-05-21 to 2008-07-30. The chart has a color gradient from blue to red. A legend indicates severity levels: Trivial, Enhancement, Minor, Normal, Major, Critical, and Blocker.



Tesseract Visualization

The screenshot displays the Tesseract visualization tool interface, which is divided into several sections:

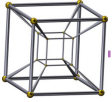
- Project:** Shows the current project as 'gnome/rhythmbox'. It includes a search bar and a 'DoubleClick node to drill down upon all selected.' instruction. Below this is a 'Commits' chart showing activity from 2002-03-02 to 2007-01-25, with a secondary 'Communication' chart below it.
- Files:** A list of files with a search bar. The list includes files like 'po/lt.po', 'widgets/gtkprivate.h', and 'data/ui/rhythmbox-ui.xml'. A 'Show Settings' button is located to the right.
- Developers:** A list of developer names, including Christina Cloutier, Vivian Johns, Katrina Bolt, and others. A 'Show Settings' button is also present.
- Network Graphs:** Two network graphs are shown. The left graph highlights a path of nodes, with a tooltip for 'shell/rb-shell.c'. The right graph shows a more complex network of connections between developers.
- Issues:** A section at the bottom right showing a severity distribution chart from 2005-05-21 to 2005-08-01. The chart shows a peak in 'Critical' issues around 2005-07-20. A legend indicates severity levels: Enhancement, Trivial, Minor, Normal, Major, Critical, and Blocker.



Tesseract Visualization

The screenshot displays the Tesseract visualization tool interface, which is divided into several sections:

- Project:** Shows the current project as 'gnome/rhythmbox' with a search bar below it. A tooltip indicates: 'DoubleClick node to drill down upon all selected.'
- Commits:** A horizontal bar chart at the top showing commit activity over time from 2002-03-02 to 2007-01-25. The chart has two series, one in blue and one in green.
- Files:** A list of files with 'shell/rb-shell.c' highlighted in orange. The list includes: 'rhythmdb/rhythmdb-query-mc', 'rhythmdb/rhythmdb-query-mc', 'rhythmdb/rhythmdb-tree.c', 'rhythmdb/rhythmdb.c', 'rhythmdb/rhythmdb.h', 'library/.cvsignore', 'shell/Makefile.am', 'shell/main.c', 'shell/rb-shell.c', 'shell/rb-remote.c', 'radio/Makefile.am', 'radio/rb-station-properties-di', 'shell/rb-playlist-manager.c', 'shell/rb-shell-player.c', and 'shell/rb-shell-player.h'.
- Developers:** A list of developers with 'Thelma Gillman' highlighted in yellow. The list includes: 'Jill Berner', 'Ruth Schweitzer', 'Lorraine Hewlett', 'Tina Raab', 'Alicia Dimaggio', 'Stephen Walther', 'Charles Kautz', 'Margie Carruth', 'Norman Buffington', 'Thelma Gillman', 'Jane Pagan', 'Sonya Pinckney', 'Douglas Parry', 'Joseph Truesdale', and 'Rodney Hanwood'.
- Issues:** A horizontal bar chart at the bottom showing issue severity over time from 2005-05-21 to 2005-08-01. The chart has a color scale from red (low severity) to yellow (high severity). A legend indicates severity levels: Enhancement (checked), Trivial (checked), Minor (checked), Normal (checked), Major (checked), Critical (checked), and Blocker (checked).



Tesseract Visualization

Project

gnome/rhythmbox

Search

DoubleClick node to drill down upon all selected.

Commits

Communication

Show Settings

Files

File name
metadata/Makefile.am
metadata/.cvsignore
shell/rb-source-header.c
metadata/rb-metadata.h
metadata/rb-metadata-xine.c
shell/rb-tray-icon.c
shell/rb-tray-icon.h
metadata/rb-metadata-gst.c
.cvsignore
lib/rb-file-helpers.c
shell/rb-statusbar.h
shell/rb-shell-preferences.c
sources/Makefile.am

Show Settings

Developers

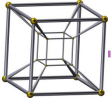
Name
Tony Theriault
Jill Berner
Ruth Schweitzer
Lorraine Hewlett
Tina Raab
Alicia Dimaggio
Stephen Walther
Charles Kautz
Margie Carruth
Norman Buffington
Thelma Gillman
Jane Pagan
Sonya Pinckney

Show Settings

Issues

Severity: Enhancement Trivial Minor Normal Major Critical Blocker

Bug ID	Start	End	Severity	Priority	Status	Resolution	Description
9020	2004-11-28	2005-06-04	normal	Normal	NEW	0	rhythmbox cant open urls when the url is a commandline argument
9030	2004-08-26	2005-06-06	critical	High	RESOLVED	OBSOLETE	Crash when resuming play
9048	2004-12-07	2005-06-10	normal	Normal	RESOLVED	OBSOLETE	import folder window titled "choose files or directory"
9057	2004-12-01	2005-07-29	critical	High	RESOLVED	FIXED	app crash when editing radio category



Developer Testimony



Without Tesseract

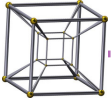
“It’s usually just talking to people about what happened, going back to the CVS and trying to see what happened with the file changes [is] **kinda fruitless.**”

With Tesseract

“...from a grunt developer standpoint, the file listing and cross reference of who has worked before – **that would be very, very nice.**”

But probably not much use for experienced developers

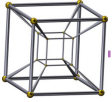
“..This stuff most useful for the initial developers...they err... for getting to know the code base. **I have been for eight years...have this stuff in mind.**...If someone new, help in how to find stuff...”



Formative Evaluation

- Instrumented with GNOME data
 - 10 years data
 - 1,000 developers, 48,000 commits
 - 200,000 bugs
- Usability studies
 - five tasks to evaluate the understanding of cross-panel referencing
 - five participants
- Open source developer feedback
 - interview to verify the need and usage scenario
 - five open source developers from different projects

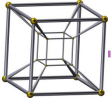




Conclusions

- Allow Interactive explorations of project relationships
 - cross-linked across different data sources
 - over time
- Treat both social and technical relationships as first order elements
- Use logical coupling for file associations
- Determine fit between communication needs and behavior
- Formative evaluations that demonstrate the need for such an approach

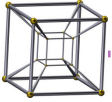




Future Work

- Summative user evaluation
- Visualization Enhancements
 - hierarchical grouping of nodes (e.g., packages, directories)
 - clustering algorithm (Newman grouping)
 - sticky layout of networks
 - integration with source repositories
- Analysis
 - other analyses (temporal consideration for congruence, SNA metrics)
 - difference in networks between two time slices





Questions!



Demo presentation on Friday 11am, Salon C

Authors gratefully acknowledge support from the following agencies:

- NSF IIS-0414698, IIS-0534656
- IBM Jazz Innovation and Faculty grants
- Alfred P. Sloan Foundation (2008-09)