Advances in the SourceForge Research Data Archive (SRDA)

Matt Van Antwerp
Greg Madey
University of Notre Dame

September 10, 2008
20th World Computer Congress
OSS 2008 / WoPDAcSD 2008
Milan, Italy

Supported in part by the National Science Foundation's CISE IIS-Digital Society & Technology program under Grant ISS-0222829 and by the National Science Foundation's CISE Computing Research Infrastructure program under Grant CNS-0751120
Advances in the SRDA - Outline

- Monthly Data Dumps
- Data Access
- Documentation
- Additional Data
- Administration
- Community Building
- Features in Development
'Big Data' Curation

- Monthly Data Dumps – SourceForge provides us with monthly snapshot of their database
- Each dump is about 7GB, compressed
- Over 600GB total
- Over 40 schemas (monthly dumps)
- PostgreSQL database
- 'Wrapped' in a Wiki
SourceForge Research Data Archive (SRDA): A Repository of FLOSS Research Data

- This Wiki contains information about using the SourceForge Research Data Archive. Links to query the SourceForge Research data warehouse can be found to the right. Your same userID and password that provides access to the query form, also can be used to login to the wiki. After logging in, all content can be edited, including creating new pages. Data access is given to academic and scholarly researchers; additional information is here.

Recent News

- The archive is being rebuilt. We are hoping for as little downtime as possible.
- May dump is now loaded.
- April dump is now loaded.
- We have posted a few ER diagrams.
- The links in the "Top Links" template have been updated to point to the new pages. The old ones are still active in case you have problems with the new ones. They are here (query form) and here (schema browser).
- Try out the new query page.
- Web Services are deployed. They have been tested but there are bound to be problems. Please let us know if you have trouble.
- Looking for particular data? Start here
- See the ND group page or check out info pages about researchers at ND: Greg Maday, Matthew Van Antwerp, Scott Christley, Jin Xu, and Yongqin Gao.
- We are working to enhance the community aspect of the website. We encourage all members to create a user or research profile and participate in discussions.
Data Access

- Data Access – provided through web front end only, initially

- Must use SQL, but recent additions include:
  - Added: drop-down query builder
  - Added: user query history
  - Added: schema browser
  - Added: web service (SOAP) access
    - Programmatic access allows queries to be scripted or batched
    - Still has issues that need to be addressed, e.g., no WSDL
SourceForge.net Research Archive Query Form

Autocomplete Implemented

Begin typing and a drop down menu will appear with a list of options beginning with the text you have already entered. Autocompletes is "tokenized" meaning it will start a new match after a space a period or a comma. It is not goof-proof since it does not check syntax on the fly and the options are all possible schemas, tables, or fields, or some combination thereof depending on the clause (select, from, or where). However, it should help you cut down on typing and typos.

Example Queries

SELECT * FROM sf0306.users WHERE user_id < 100

SELECT user_name FROM sf1104.users a, sf1104.artifact b WHERE a.user_id = b.submitted_by AND b.artifact_id = 304727

Quick Lookup

Examples:
sf0306
sf0107.users
users

News

- We are trying to make this page more interactive and implement helpful features
- Be sure to check out the many recent additions to the wiki
- If you have trouble with this page, here is the old one
- Your most recent successful query is filled in by default. Hit Clear to remove it. Reset will replace all fields with your most recent query.
- May dump (sf0508) is now loaded
- April dump (sf0408) is now loaded
- If you notice any errors, please let us know (oss at nd dot edu)
| 1. Wed Jun 18 11:53:26 2008 -- SELECT \texttt{max(date)} FROM \texttt{sf0508.forum} |
| 2. Fri Jun 13 14:29:38 2008 -- SELECT \texttt{max(date)} FROM \texttt{sf0308.forum} |
| 3. Fri Jun 13 14:29:00 2008 -- SELECT \texttt{max(date)} FROM \texttt{sf0408.forum} |
| 4. Fri Jun 13 14:26:48 2008 -- SELECT \texttt{date} FROM \texttt{sf0408.forum} WHERE \texttt{true order by date desc limit 1} |
| 5. Fri Jun 13 14:23:15 2008 -- SELECT \texttt{count(*)} FROM \texttt{sf0508.forum} |
| 6. Fri Jun 13 12:00:07 2008 -- SELECT \texttt{date} FROM \texttt{sf0508.forum} WHERE \texttt{true order by date desc limit 1} |
| 7. Fri Jun 13 10:55:01 2008 -- SELECT \texttt{date} FROM \texttt{sf0508.forum} WHERE \texttt{true order by date limit 1} |
| 8. Fri Jun 13 10:52:10 2008 -- SELECT \texttt{count(*)} FROM \texttt{sf0508.users} |
| 9. Tue May 20 15:21:23 2008 -- SELECT \texttt{count(*)} FROM \texttt{sf0408.users} |
| 10. Wed May 14 12:34:34 2008 -- SELECT \texttt{count(*)} FROM \texttt{sf0408.users} |
| 11. Wed May 14 12:33:31 2008 -- SELECT \texttt{*} FROM \texttt{sf0408.frs_release} WHERE \texttt{true limit 5} |
| 12. Tue May 6 09:58:17 2008 -- SELECT \texttt{*} FROM \texttt{sf0308.frs_release} WHERE \texttt{true limit 5} |
| 13. Thu May 1 19:05:49 2008 -- SELECT \texttt{*} FROM \texttt{sf0308.trove_agg} WHERE \texttt{group_name='dreichl'} |
| 14. Thu May 1 18:57:04 2008 -- SELECT \texttt{*} FROM \texttt{sf0308.users} WHERE \texttt{user_id=853915} |
| 15. Thu May 1 18:56:41 2008 -- SELECT \texttt{*} FROM \texttt{sf0308.users} WHERE \texttt{true limit 15} |
| 16. Fri Apr 25 18:45:27 2008 -- SELECT \texttt{a.group_id, b.*} FROM \texttt{sf0208.trove_group_link a, sf0208.trove_cat b WHERE a.group_id IN (235,237) and a.trove_cat_id=b.trove_cat_id and b.root_parent IN (199,160,6) ORDER BY group_id} |
| 17. Fri Apr 25 16:43:24 2008 -- SELECT \texttt{a.group_id, b.*} FROM \texttt{sf0208.trove_group_link a, sf0208.trove_cat b WHERE a.group_id IN (235,237) and a.trove_cat_id=b.trove_cat_id and b.root_parent=198 ORDER BY group_id} |
Usage: Queries & Downloads

Queries and Downloads by Month (2006-2008)

- Number of Queries
- Gigabytes Downloaded
Web Service Usage

Queries Per Day Since Inception April 2008

Number of Queries

queries

Apr 22  May 06  May 20  Jun 03  Jun 17  Jul 01  Jul 15
Documentation

- Documentation (meta-data) added
- User and administrator-provided content is located in the wiki
  - Each table has its own page
  - Additional resources like FAQ, E-R diagrams, and general instructions also on wiki
- Schema browser is a dynamic tool for displaying schema and table descriptions
### SourceForge.net Research Archive Schema Browser

**Instructions:**

Each link gives the month and year for the schema data followed by the scheme name. Click on a schema name to view its tables. From there, you can click on a table name to view its full description.

**Quick Lookup**

Examples:
- sf0208
- sf0107.users
- users

<table>
<thead>
<tr>
<th>2003</th>
<th>2004</th>
<th>2007</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 2003</td>
<td>sf0103</td>
<td>January 2007</td>
<td>sf0107</td>
</tr>
<tr>
<td>November 2004</td>
<td>sf1104</td>
<td>February 2007</td>
<td>sf0207</td>
</tr>
<tr>
<td>December 2004</td>
<td>sf1204</td>
<td>March 2007</td>
<td>sf0307</td>
</tr>
<tr>
<td></td>
<td></td>
<td>April 2007</td>
<td>sf0407</td>
</tr>
<tr>
<td></td>
<td></td>
<td>May 2007</td>
<td>sf0507</td>
</tr>
<tr>
<td></td>
<td>users</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

You can query the database here.
'Good Netizen' gentle data extraction to minimize load on SourceForge CVS/SVN servers

CVS/Subversion metadata for over 100,000 SourceForge projects is in a beta stage of development

Similar metadata available for BerliOS and GNU Savannah

New web frontend being tested – more interactive
Simple Query Interface

Users

Groups
gcc

Submit Query
Users who have contributed to 3dl1df:

gmannds
lfinsto1

Distinct Commit Times:

Time - cumulative net line changes

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fri Jul 9</td>
<td>11:09:50</td>
<td>0</td>
</tr>
<tr>
<td>Fri Jul 9</td>
<td>12:06:43</td>
<td>2</td>
</tr>
<tr>
<td>Fri Jul 9</td>
<td>12:23:03</td>
<td>2</td>
</tr>
<tr>
<td>Fri Jul 9</td>
<td>12:43:01</td>
<td>2</td>
</tr>
<tr>
<td>Sun Jul 18</td>
<td>16:48:41</td>
<td>2</td>
</tr>
<tr>
<td>Sun Jul 18</td>
<td>17:39:32</td>
<td>2</td>
</tr>
<tr>
<td>Mon Jul 19</td>
<td>11:08:30</td>
<td>17850</td>
</tr>
<tr>
<td>Wed Jul 21</td>
<td>10:35:04</td>
<td>17850</td>
</tr>
<tr>
<td>Wed Jul 21</td>
<td>12:01:52</td>
<td>17850</td>
</tr>
<tr>
<td>Wed Jul 21</td>
<td>16:20:54</td>
<td>17860</td>
</tr>
<tr>
<td>Thu Jul 22</td>
<td>14:27:51</td>
<td>17910</td>
</tr>
<tr>
<td>Fri Jul 23</td>
<td>12:51:00</td>
<td>17910</td>
</tr>
<tr>
<td>Fri Jul 23</td>
<td>13:07:48</td>
<td>17917</td>
</tr>
<tr>
<td>Tue Jul 27</td>
<td>14:42:46</td>
<td>17940</td>
</tr>
<tr>
<td>Wed Jul 28</td>
<td>15:15:05</td>
<td>17957</td>
</tr>
<tr>
<td>Sat Aug 7</td>
<td>03:45:43</td>
<td>17957</td>
</tr>
<tr>
<td>Sun Aug 8</td>
<td>17:19:29</td>
<td>17957</td>
</tr>
<tr>
<td>Thu Aug 12</td>
<td>16:25:18</td>
<td>17957</td>
</tr>
<tr>
<td>Thu Aug 12</td>
<td>16:35:32</td>
<td>17963</td>
</tr>
<tr>
<td>Fri Aug 13</td>
<td>11:32:45</td>
<td>17968</td>
</tr>
<tr>
<td>Mon Aug 16</td>
<td>12:36:15</td>
<td>17968</td>
</tr>
<tr>
<td>Tue Aug 17</td>
<td>15:36:33</td>
<td>17980</td>
</tr>
<tr>
<td>Wed Aug 18</td>
<td>15:03:41</td>
<td>17990</td>
</tr>
<tr>
<td>Wed Aug 18</td>
<td>16:43:47</td>
<td>17999</td>
</tr>
<tr>
<td>Fri Aug 20</td>
<td>12:16:21</td>
<td>18011</td>
</tr>
<tr>
<td>Fri Aug 20</td>
<td>15:42:48</td>
<td>18018</td>
</tr>
<tr>
<td>Mon Aug 23</td>
<td>15:11:54</td>
<td>18028</td>
</tr>
<tr>
<td>Mon Aug 23</td>
<td>17:22:45</td>
<td>18035</td>
</tr>
<tr>
<td>Tue Aug 24</td>
<td>16:41:50</td>
<td>18046</td>
</tr>
<tr>
<td>Thu Aug 26</td>
<td>15:28:29</td>
<td>18073</td>
</tr>
<tr>
<td>Thu Aug 26</td>
<td>16:02:57</td>
<td>18075</td>
</tr>
<tr>
<td>Fri Aug 27</td>
<td>14:22:58</td>
<td>18088</td>
</tr>
</tbody>
</table>
Project 3dldf has been alive for 46 months from **Fri Jul 9 11:09:50 2004** to **Wed Feb 13 15:54:36 2008**
There have been 237 total file commits in that time span.
File list:

/source/3dlc3dlc/Attic/eqdexpr.wv
/source/3dlc3dlc/Attic/pprdepxpr.wv
/source/3dlc3dlc/CWEB/Group/Attic/main.web.v
/source/3dlc3dlc/Group/00NEWS.v
/source/3dlc3dlc/Group/00README.v
/source/3dlc3dlc/Group/ANIMATNS/animctrl_web.v
/source/3dlc3dlc/Group/ANIMATNS/Attic/conveys_web.v
/source/3dlc3dlc/Group/ANIMATNS/ChangeLog.v
/source/3dlc3dlc/Group/ANIMATNS/loader_web.v
/source/3dlc3dlc/Group/ANIMATNS/LOGS/animctrl_web.log.v
/source/3dlc3dlc/Group/ANIMATNS/LOGS/conveys_web.log.v
/source/3dlc3dlc/Group/ANIMATNS/LOGS/loader_web.log.v
/source/3dlc3dlc/Group/ANIMATNS/LOGS/Makefile.am.log.v
/source/3dlc3dlc/Group/ANIMATNS/LOGS/parser_web.log.v
### Revisions by user lfinsto1 for project 3dldf

<table>
<thead>
<tr>
<th>file name</th>
<th>revision</th>
<th>time</th>
<th>user_name</th>
<th>state</th>
<th>lines plus</th>
<th>lines minus</th>
<th>branches</th>
<th>message</th>
</tr>
</thead>
<tbody>
<tr>
<td>sources/3dldf/3dldf/Attic/pd:expr.pw.y</td>
<td>1.2</td>
<td>Wed Nov 8 16:35:11 2006</td>
<td>lfinsto1</td>
<td>dead</td>
<td>0</td>
<td>0</td>
<td></td>
<td>Removed. This file shouldn't have been here in the first place.</td>
</tr>
<tr>
<td>sources/3dldf/3dldf/Attic/pd:expr.pw.y</td>
<td>1.1</td>
<td>Fri Jan 20 20:51:58 2006</td>
<td>lfinsto1</td>
<td>Exp</td>
<td></td>
<td></td>
<td></td>
<td>Added.</td>
</tr>
<tr>
<td>sources/3dldf/3dldf/Attic/pd:expr.pw.y</td>
<td>1.2</td>
<td>Wed Nov 8 16:35:11 2006</td>
<td>lfinsto1</td>
<td>dead</td>
<td>0</td>
<td>0</td>
<td></td>
<td>Removed. This file shouldn't have been here in the first place.</td>
</tr>
<tr>
<td>sources/3dldf/3dldf/Attic/pd:expr.pw.y</td>
<td>1.1</td>
<td>Fri Jan 20 20:48:30 2006</td>
<td>lfinsto1</td>
<td>Exp</td>
<td></td>
<td></td>
<td></td>
<td>Added.</td>
</tr>
<tr>
<td>sources/3dldf/3dldf/CWEP/Group/Attic/main.web.y</td>
<td>1.2</td>
<td>Fri Jul 9 12:06:43 2004</td>
<td>lfinsto1</td>
<td>dead</td>
<td>1</td>
<td>1</td>
<td></td>
<td>*** empty log message ***</td>
</tr>
<tr>
<td>sources/3dldf/3dldf/CWEP/Group/Attic/main.web.y</td>
<td>1.1</td>
<td>Fri Jul 9 11:06:50 2004</td>
<td>lfinsto1</td>
<td>Exp</td>
<td></td>
<td></td>
<td></td>
<td>Initial version on CVS (Group)</td>
</tr>
<tr>
<td>sources/3dldf/3dldf/GN/Group/GNNEWS.y</td>
<td>1.18</td>
<td>Sun Dec 23 15:20:28 2007</td>
<td>lfinsto1</td>
<td>Exp; lines: +2 -2; committed: SNp91gOnAklzFsKs</td>
<td></td>
<td></td>
<td></td>
<td>Working on matrix multiplication. Added &quot;2008&quot; to copyright notices.</td>
</tr>
<tr>
<td>sources/3dldf/3dldf/GN/Group/GNNEWS.y</td>
<td>1.18</td>
<td>Wed Sep 19 11:47:00 2007</td>
<td>lfinsto1</td>
<td>Exp; lines: +3 -3; committed: HyEeqvHe87c8ys</td>
<td></td>
<td></td>
<td></td>
<td>Changed the version number from &quot;1.2.0.0&quot; to &quot;1.2.0&quot;. Both numbers were used in different places before. Working on [Focuses]</td>
</tr>
</tbody>
</table>
Users who have contributed to `3dldf`:

- gmanndsu
- lfinst01

File Revision History: `/sources/3dldf/3dldf/Group/00README.v`

<table>
<thead>
<tr>
<th>revision</th>
<th>time</th>
<th>user_name</th>
<th>state</th>
<th>lines plus</th>
<th>lines minus</th>
<th>branches</th>
<th>message</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.16</td>
<td>Sun Dec 23 15:20:29 2007</td>
<td>lfinst01</td>
<td>Exp</td>
<td>+2 -2; commitid: 3Np91gOnAkLfyzs</td>
<td></td>
<td></td>
<td>Working on matrix multiplication. Added &quot;2008&quot; to copyright notices.</td>
</tr>
<tr>
<td>1.17</td>
<td>Wed Sep 19 11:47:00 2007</td>
<td>lfinst01</td>
<td>Exp</td>
<td>+3 -3; commitid: lHyEqaHcS07cS8ys</td>
<td></td>
<td></td>
<td>Changed the version number from &quot;1.2.0&quot; to &quot;1.2.0&quot;. Both numbers were used in different places before, Working on [Focuses].</td>
</tr>
<tr>
<td>1.16</td>
<td>Mon Jul 9 14:23:25 2007</td>
<td>lfinst01</td>
<td>Exp</td>
<td>+17 -17; commitid: 4VZ0h7VSEcY16ps</td>
<td></td>
<td></td>
<td>Working on functions for hyperbolae. Updated copyright and license information in many files.</td>
</tr>
<tr>
<td>1.15</td>
<td>Sat Feb 3 12:41:16 2007</td>
<td>lfinst01</td>
<td>Exp</td>
<td>4</td>
<td>3</td>
<td></td>
<td>Updated copyright notice and address of the Free Software Foundation.</td>
</tr>
<tr>
<td>1.14</td>
<td>Thu Apr 14 14:18:41 2005</td>
<td>lfinst01</td>
<td>Exp</td>
<td>7</td>
<td>33</td>
<td></td>
<td>Corrected errors in <code>CWWEB/Makefile.am</code>. <code>3dldf</code> should now be built correctly using <code>ctools3DL</code>, which is generated by the shell script <code>gctools3DL</code>. Modified Files: Group/00README Group/ChangeLog Group/CWWEB/Group/CWWEB/Makefile.am CWWEB/parser.c++ Group/CWWEB/parser.h Group/CWWEB/parser.h++ Group/CWWEB/parser/output Group/CWWEB/parser.y++ Group/CWWEB/LOGS/Makefile.am Group/LOGS/00README.log Group/SNAPSHOTS/3DLDSnap.tar.gz Group/SNAPSHOTS/3DLDSnap.tar.gz</td>
</tr>
</tbody>
</table>
### Users who have contributed to 3dldf:

- gmarndseu
- lfinstol

### Revisions occurring at Wed Sep 19 11:47:00 2007 for project 3dlfd

<table>
<thead>
<tr>
<th>file name</th>
<th>revision</th>
<th>time</th>
<th>user_name</th>
<th>state</th>
<th>lines plus</th>
<th>lines minus</th>
<th>branches</th>
<th>message</th>
</tr>
</thead>
<tbody>
<tr>
<td>/sources/3dlfd/3dlfd</td>
<td>1.16</td>
<td>Wed Sep 19 11:47:00 2007</td>
<td>lfinstol</td>
<td>Exp; lines: +3 -3; commitid: lHyEeqflE87c6lys</td>
<td></td>
<td></td>
<td></td>
<td>Changed the version number from &quot;1.2.0.0&quot; to &quot;1.2.0&quot;. Both numbers were used in different places before. Working on [Focuses].</td>
</tr>
<tr>
<td>/sources/3dlfd/3dlfd</td>
<td>1.17</td>
<td>Wed Sep 19 11:47:00 2007</td>
<td>lfinstol</td>
<td>Exp; lines: +3 -3; commitid: lHyEeqflE87c6lys</td>
<td></td>
<td></td>
<td></td>
<td>Changed the version number from &quot;1.2.0.0&quot; to &quot;1.2.0&quot;. Both numbers were used in different places before. Working on [Focuses].</td>
</tr>
<tr>
<td>/sources/3dlfd/3dlfd</td>
<td>1.45</td>
<td>Wed Sep 19 11:47:00 2007</td>
<td>lfinstol</td>
<td>Exp; lines: +5 -2; commitid: lHyEeqflE87c6lys</td>
<td></td>
<td></td>
<td></td>
<td>Changed the version number from &quot;1.2.0.0&quot; to &quot;1.2.0&quot;. Both numbers were used in different places before. Working on [Focuses].</td>
</tr>
<tr>
<td>/sources/3dlfd/3dlfd</td>
<td>1.4</td>
<td>Wed Sep 19 11:47:00 2007</td>
<td>lfinstol</td>
<td>Exp; lines: +14 -14; commitid: lHyEeqflE87c6lys</td>
<td></td>
<td></td>
<td></td>
<td>Changed the version number from &quot;1.2.0.0&quot; to &quot;1.2.0&quot;. Both numbers were used in different places before. Working on [Focuses].</td>
</tr>
</tbody>
</table>
CVS Data Schema

SQL queries: `unix_group_name` and `user_name` can be foreign keys into the rest of SRDA
SVN Data Schema

SQL queries: `unix_group_name` and `user_name` can be foreign keys into the rest of SRDA
Administration

- Administration (account creation and the majority of wiki updates) are semi-automated
- Authentication is done through the wiki user table – scripts do all other housekeeping
- For each new dump, wiki pages are updated with new info regarding the dump
  - Done using a wiki versioning tool called MVS

Community Building

- Research Project Pages
- Rather than YAML (Yet Another Mail List), users are encouraged to use the FLOSSmole mail list
Community Participation

Category: Research Projects

Research Projects Using the SourceForge Research Archive

Articles in category "Research Projects"

There are 104 articles in this category.

A
- Abed Noor
- Adamic
- Ahuja
- Alexander
- Au

H cont.
- Hillison
- Hinds
- Hinz
- Hoffmann
- Howison
- Hutcheson

R cont.
- Ranka
- Ratan
- Robledo
- Roy

S
- Santos
Howison

Project Title

The Development and Practices of Effective Distributed Work Teams

Researchers

James Howison, Syracuse University

Kevin Crowston, Syracuse University

Project Abstract

In the current study, we will examine how distributed teams develop shared mental models to guide members' behavior, roles to mediate access to resources, and norms and rules to shape action, as well as the dynamics by which independent, geographically-dispersed individuals are socialized into team roles and behaviors.
Adamic

Project Title

- An Exploratory Study on the Relationship Between OSS Project Popularity and Network Characteristics
- Using Networks to Visualize and Understand Participation on SourceForge.net

Researchers

Lada Adamic, School on Information, University of Michigan
Min-seok Pang, University of Michigan
Nathan Oostendorp, University of Michigan and SourceForge.net

Project Abstract

Two projects from a course offered by Professor Lada Adamic, SI 708/ CSCS 535 Networks: Theory and Application, Winter 2007

References


Link: http://www-personal.umich.edu/~ladamic/courses/si708w07/index.html
Hinds

Project Title

Relationship Between Open Source Software Project Community Structure and Project Success

Researchers

David Hinds, Florida International University
Ronald M. Lee, Florida International University

Project Abstract

In recent years, a surprising new phenomenon has emerged in which globally distributed online communities collaborate to create useful and sophisticated computer software. These open source software groups are comprised of generally unaffiliated individuals and organizations who work in a seemingly chaotic fashion and who participate on a voluntary basis without direct financial incentive.

The purpose of this research is to investigate the relationship between the social network structure of these intriguing groups and their level of output and activity, where social network structure is defined as 1) closure or connectedness within the group, 2) bridging ties which extend outside of the group, and 3) leader centrality within the group. Based on well-tested theories of social capital and centrality in teams, propositions were formulated which suggest that social network structures associated with successful open source software project communities will exhibit high levels of bridging and moderate levels of closure and leader centrality.

The research setting was the SourceForge hosting organization and a study population of 143 project communities was identified. Independent variables included measures of closure and leader centrality defined over conversational ties, along with measures of bridging defined over membership ties. Dependent variables included source code commits and software releases for community output, and software downloads and project site page views for community activity. A cross-sectional study design was used and archival data were extracted and aggregated for the two-year period following the first release of project software. The resulting compiled variables were analyzed using multiple linear and quadratic regressions, controlling for group size and conversational volume.

References

Work in Process

- Refinements
  - Stored queries
  - Automated graph production
  - Web service improvement

- New Data
  - mbox archives?
  - Continual updates to CVS/SVN from SourceForge
Questions?

SourceForge Research Data Archive (SRDA)

http://zerlot.cse.nd.edu/

Thank you!

Supported in part by the National Science Foundation's CISE IIS-Digital Society & Technology program under Grant ISS-0222829 and by the National Science Foundation's CISE Computing Research Infrastructure program under Grant CNS-0751120