




Accelerating Cross-Project Knowledge Collaboration Using Collaborative Filtering and Social Networks

Masao Ohira Naoki Ohsugi Tetsuya Ohoka Ken-ichi Matsumoto
{masao, naoki-o, tetsuy-o, matumoto} @ is.naist.jp

Software Engineering Laboratory,
Graduate School of Information Science,
Nara Institute of Science and Technology, JAPAN





Background

- A large numbers of FOSS projects use hosting sites (e.g., SourceForge.net, Java.net)
- These sites provide FOSS projects with a variety of software repositories (e.g., CVS, Bugzilla, Mailman) for collaborative software development
- However, such the repositories are designed to support **collaborations in a single project**

Analysis of social interactions in hosting sites

- We found poor interactions in SF.net in Feb. 2005
 - 66.7% of projects had one developers.
 - 77.3% of developers belonged to one project.

Number of project with N developers

NUM. OF DEVELOPERS	NUM. OF PROJECTS	(%)
0	278	0.3
1	60665	66.7
2	14151	15.6
3	5854	6.4
4	3222	3.5
OVER 5	6732	7.4
TOTAL	90902	100

Number of developers with P projects

NUM. OF PARTICIPATING PROJECTS	NUM. OF DEVELOPERS	(%)
1	100408	77.3
2	18753	14.4
3	5980	4.6
4	2350	1.8
OVER 5	2406	1.8
TOTAL	129897	100



Need for cross-project knowledge collaboration

- The single-project support mechanism works well for large projects that have many developers and contributors, but does not work sufficiently for small projects
- Small projects need means to access information or knowledge holders of other projects related to problem solving



Goal and approach

- Goal: to provide FOSS projects with means to make cross-project collaborations easier

- Approach:
 - to use **social networks** for linking developers of small projects to developers of large projects
 - to use **collaborative filtering** for suggesting more relevant information and people of other projects

- Graphmania: a tool for visualizing relationships among developers and projects in SF.net

Graphmania: data collection

- Collecting data from project and developer information pages using an autopilot tool for SF

Project information

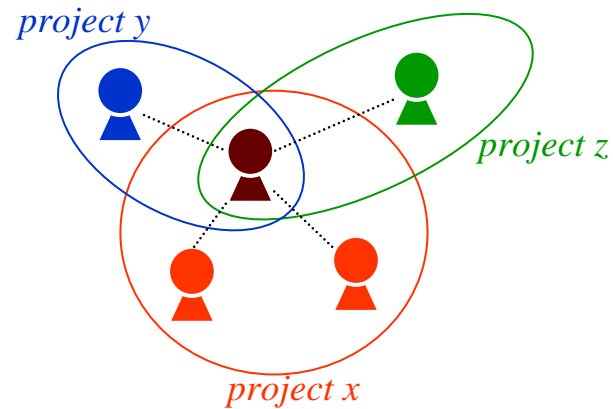
ATTRIBUTES	EXAMPLE
project name	phpMyAdmin
description	phpMyAdmin is a tool written in PHP intended to handle the administration of MySQL over the Web...
num. of developers	8 (garvinhicking, lem9, mkkech, ne0X, nijel, ...)
keywords	php, databases, ...
program lang.	PHP

Developer information

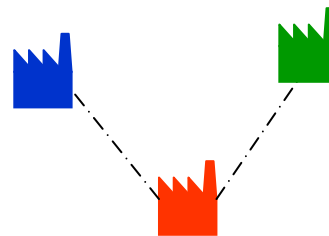
ATTRIBUTES	EXAMPLE
user name	naoki-o
public name	Naoki Ohsugi
email address	ohsugi at users.sourceforge.net
site member since	2002/6/10 22:16
group member of	NAIST Collaborative Filtering Engines, Game-R, Bullflog
skill inventory	C/C++: Competent: 5yr-10yr Perl: Competent: 5yr-10yr Java: Want to Learn: 2yr-5yr

Graphmania: social network analysis

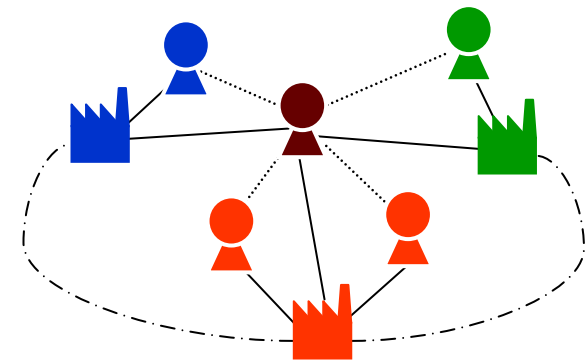
- Defining three kinds of social networks using the collected data



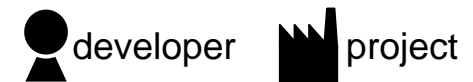
Developer network



Project network



Developer-project network



Graphmania: collaborative filtering

- Calculating cosine similarity based on
 - how often developers of a project (e.g **project X**) join other projects concurrently

	developer A	developer B	developer C	developer D	developer...	
project X	1 (member)	1 (member)	1 (member)	0 (not member)	1 (member)	
project a	1 (member)	1 (member)	1 (member)	0 (not member)	1 (member)	similarity: 1.0
project b	1 (member)	1 (member)	0 (not member)	0 (not member)	1 (member)	similarity: 0.8
project c	0 (not member)	1 (member)	0 (not member)	1 (member)	0 (not member)	similarity: 0.3

similarity between project X and other projects

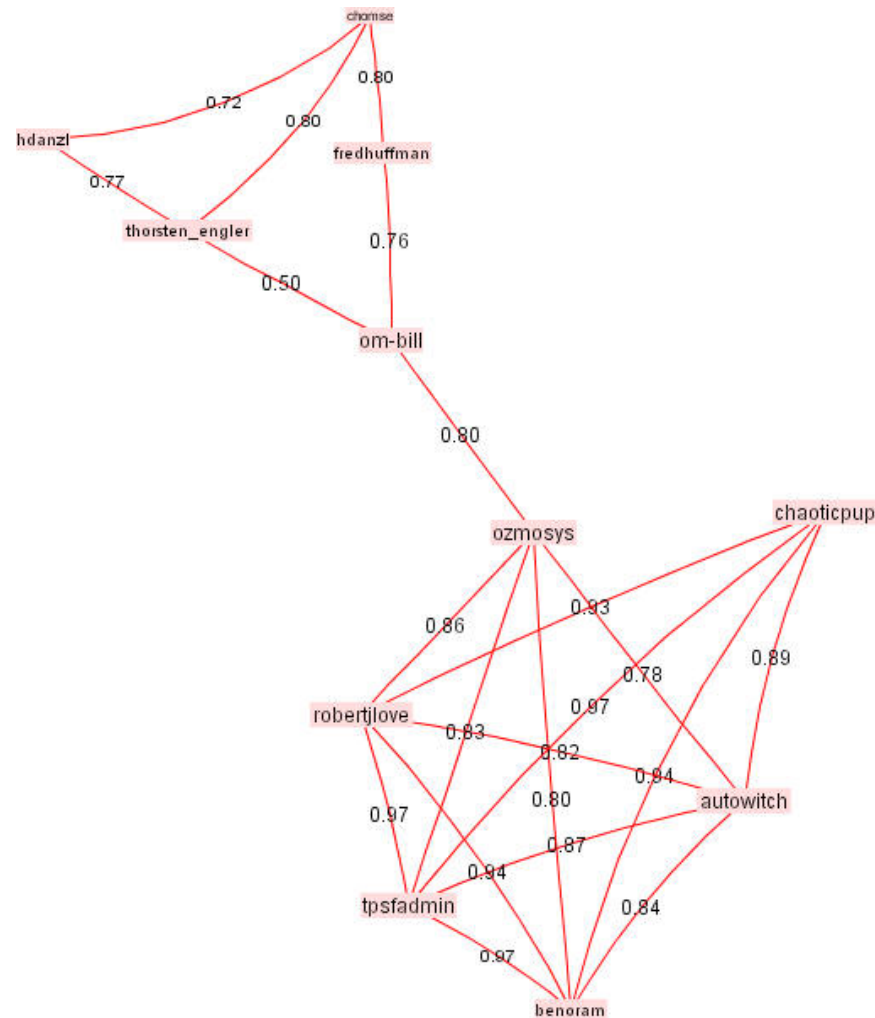
Graphmania: collaborative filtering

- Calculating cosine similarity based on
 - how often developers of a project (e.g. project X) join other projects concurrently
 - how often a developer (e.g. developer X) works with other developers

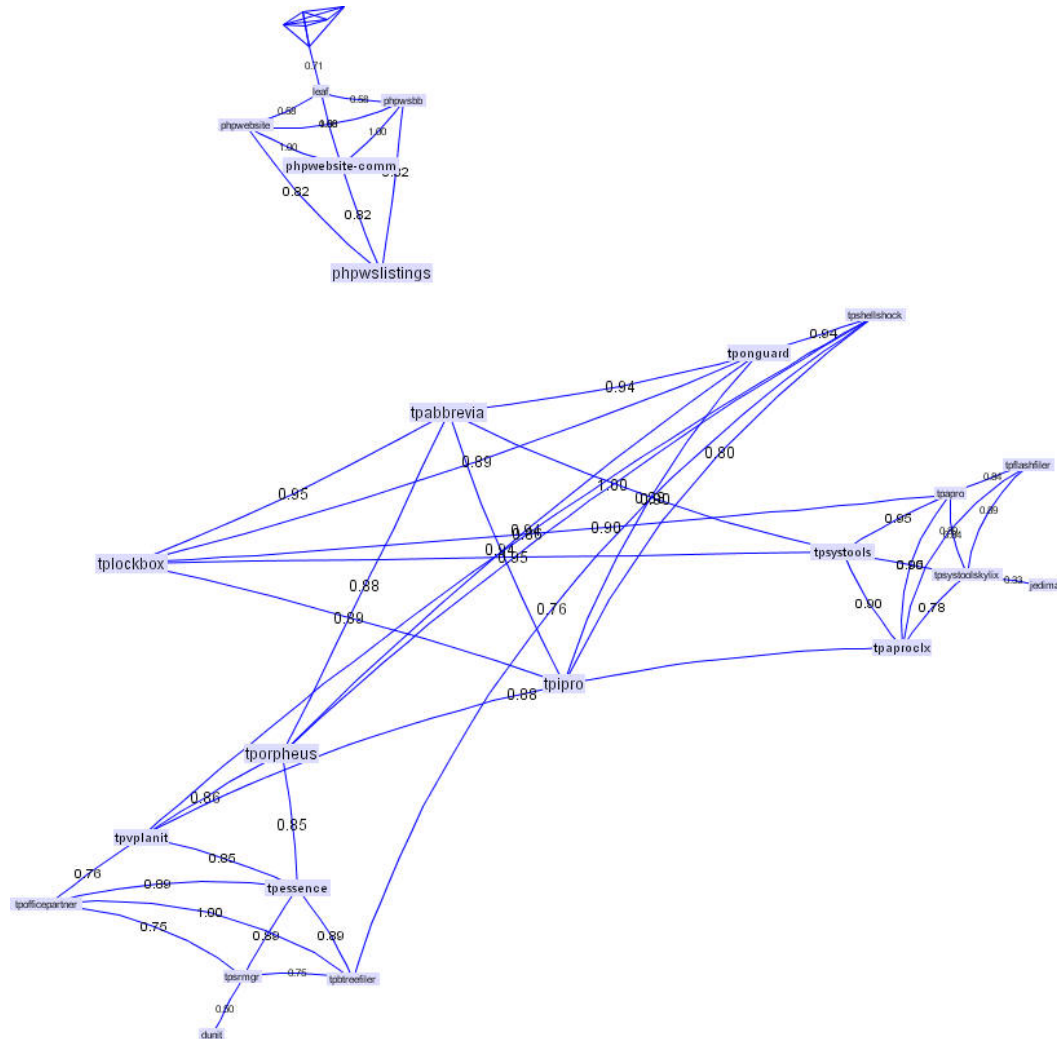
	project A	project B	project C	project D	project ...	
developer X	1 (member)	1 (member)	1 (member)	0 (not member)	1 (member)	
developer a	1 (member)	1 (member)	1 (member)	0 (not member)	1 (member)	similarity: 1.0
developer b	1 (member)	1 (member)	0 (not member)	0 (not member)	1 (member)	similarity: 0.8
developer c	0 (not member)	1 (member)	0 (not member)	1 (member)	0 (not member)	similarity: 0.3

similarity between developer X and other developers

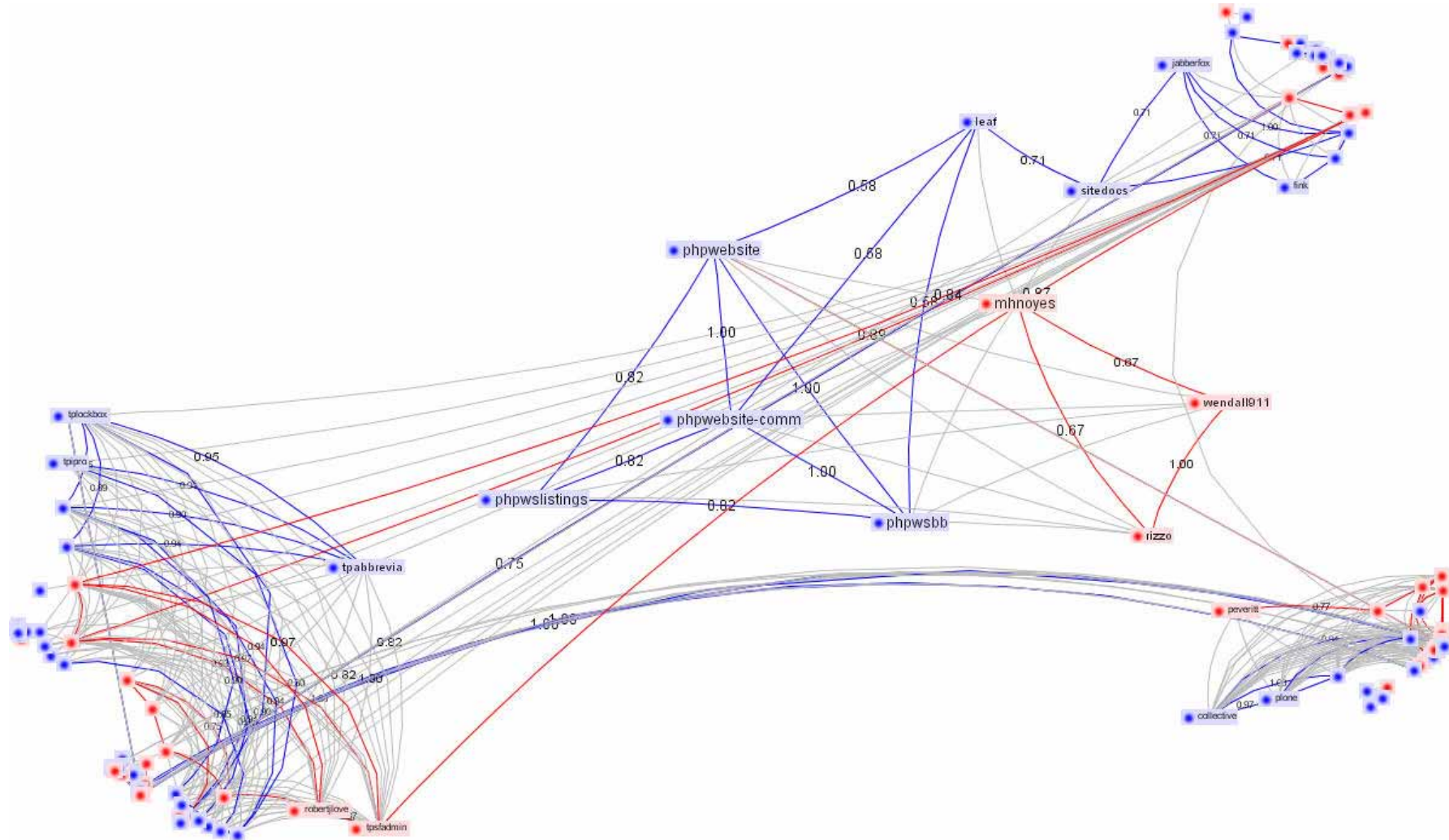
Graphmania: visualization (1) developer networks



Graphmania: visualization(2) project networks



Graphmania: Visualization (3) developer-project networks





Discussions

- What can users of Graphmania find out?
 - developer networks
 - linchpins who have many connections to others
 - the strength of relationships between developers
 - project networks
 - the relation between projects (or between own projects and other projects)
 - the tendency that similar projects are linked each other
 - developer-project networks
 - the relationships between developers and projects
 - complex visualization



Limitations and future work

- The results of visualizations do not suggest something useful to users currently
 - the data used for visualizations is “who is joining which projects” and “which projects have whom”
 - collaborative filtering are not used for recommending something useful such as Amazon.com but for calculating similarities
- Graphmania will be extended to be able to provide more effective visualizations and to recommend something useful for users