# CS680: Distributed Software Development

# Project Report - Spring Term 2010

**Team Members:**

1. First Name: Jordan Last Name: Osecki E-mail: [Jordan.Osecki@gmail.com](mailto:Jordan.Osecki@gmail.com)

2. First Name: Sammie Last Name: Stahlback E-mail: [Sws28@drexel.edu](mailto:Sws28@drexel.edu)

3. First Name: Michael Last Name: Kim E-mail: [Mikim303@yahoo.com](mailto:Mikim303@yahoo.com)

**Project Name:**

* Analysis Team – Core/Periphery Analysis of MySQL from

Data Mining Open Source Project

# Abstract

About 300 words, summarizing your problem and results, as they will be discussed in the rest of the document.

# Overview (1 page max – excluding figures)

State the DSD issue that you have embraced with your project and the basic idea of what the project is about.

What objective you were supposed to achieve.

Project goals:  
1) Recreate community / social network for each release (or something more finer grained)  
     -Core/Periphery Structure using metrics from other papers  
     -Create a social network structure with nodes being the alises and the edges being messages and weighted.  
     -Look for tools that will calculate more metrics for us based on our own initial metrics.  
2) Statistics  
     -Volume of traffic within community for each release (or something more finer grained)  
         -The final product may simply be a chart showing this

Project goals:

1. Recreate community for each release

Core

Periphery

1. Statistics

Volume of traffic within community

How you have decided to pursue that objective.

# Background (1 page max – excluding figures)

Provide any background information that is useful to better understand the context in which you developed your project.

May have to do with the description and characteristics of the project you have worked with.

The techniques that were used.

The environments and tools that you employed.

If there is any related work that you think is relevant to the work you have done and should be referenced, discuss it here.

# Approach (2 pages max – excluding figures)

Describe in detail the work you have done, as a team or individuals.

The methods you have employed to achieve the intended results.

Detail whatever you have built and how it works

Discuss here also any issue of process and collaboration if relevant. – MINING ISSUES

High level overview of project tasks:

1. Collect data from (<http://lists.mysql.com>) NEWS GROUPS, CHANGE LOG
2. Create Postgres/PostgreSQL database using schema from sql.tar.gz
3. Process data collected from (http://lists.mysql.com)
4. Populate Postgres/PostgreSQL database - SCHEMA
5. Analyze communication (reference previous papers discussed in class)
6. Analyze one particular MySql Database Server:  
       - Mine archived developer's mailing list for last 10 years on lists.mysql.com
7. (One list from Server links, have to determine mainone)  
   - An assumption will be in the tree of messages, that a reply is  
   meant for the author of the previous message in tree.  
     
   High level overview of project tasks:  
   1) Decide on discussion/blog area  
     a) Private group in bb vista  
   2) Collect data from mailing list archives (<http://lists.mysql.com>)  
   3) Create Postgres/PostgreSQL database using schema from sql.tar.gz  
   (Use initMails.sql and person and alias table from initBase.sql)  
   4) Process data collected from (<http://lists.mysql.com>)  
   5) Populate Postgres/PostgreSQL database with the data  
   6) Analyze communication (reference previous papers discussed in  
   class, see Project Goals above)  
   7) Produce a final report analyzing our findings  
     
   Immediate TODO's:  
   -Get pointers to the relevant papers  
   -Get pointers to tools that will help us calculate c/p  
   -Determine if mySQL has its mailing list archived  
   -Determine correct mailing list to use  
   -Get a database set up on tux  
   Timeline:  
   -Data Collection:  Approximately 2 Weeks  
   -Data Analysis:  Approximately 4 Weeks

# Evaluation (max 2 pages – excluding figures)

Here is where you present all of your results and discuss them as in depth as you can.

Analyze and discuss what those results mean with respect to the issue chosen for the project.

Release information

mysql-3.20.32a.tar.gz@  
mysql-3.21.33b.tar.gz@ 6/1998  
mysql-3.22.32.tar.gz@ 2/2000  
mysql-3.23.57.tar.gz@ 6/2003  
mysql-3.23.58.tar.gz@ 9/2003  
mysql-4.0.26.tar.gz@ 9/2005  
mysql-4.0.27.tar.gz@ 5/2006  
mysql-4.1.21.tar.gz@ 7/2006  
mysql-4.1.22.tar.gz@ 11/2006  
mysql-5.0.45.tar.gz@ 7/2007  
mysql-5.1.22-rc.tar.gz@ 9/2007  
mysql-5.1.23-rc.tar.gz@ 1/2008  
mysql-5.2.0-falcon-alpha.tar.gz@ 1/2007  
mysql-5.2.3-falcon-alpha.tar.gz@ 2/2007  
mysql-6.0.0-alpha.tar.gz@ 4/2007  
mysql-6.0.2-alpha.tar.gz@ 8/2007  
mysql-6.0.3-alpha.tar.gz@ 11/2007  
mysql-6.0.4-alpha.tar.gz@ 2/2008

What I'd like you guys to do is to figure out a date for each of the   
above releases, which will be the moment we'll use to take snapshots of the Social Network mined from the ML and compute its Core/Periphery metrics.

If you look at the references of last week's Crowston's paper, you'll find a paper by Borgatti + 1 co-author which described the C/P metric. About the MLs, I was wrong: the relevant ones (developers' communications) are "internal" ones: specifically Internals and  possibly also Bugs. Let's start with the former.

Cover both strengths and shortcoming of your approach.

# Conclusions (max 1 page)

Summarize the lessons you have learned from the project.

What is the most significant contribution?

How would you do it anew, if you had the hindsight you have now?

Is there any promise?

How can this work be further extended?

# References

Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

Crowston/Borgatti Paper