Performance and Creativity Indicators within Communication Patterns of the Open Source Community
Peter A. Gloor and Yared Kidane

Introduction
This research brief reports on the communication patterns of online communities of developers and users of the open source Eclipse Java development environment. The research involved measuring the productivity of each community and identifying correlations that exist between group communication characteristics and productivity attributes. The TeCFlow (Temporal Communication Flow) visualizer tool was used to create “movie maps” of the knowledge flow by analyzing the publicly accessible Eclipse developer mailing lists as an approximation of the social networks of developers and users. Thirty-three different Eclipse communities discussing development and use of components of Eclipse such as the Java Development Tools, the different platform components, the C/C++ Development Tools and the AspectJ extension were analyzed over a period of six months. The evolution of social network variables such as “betweenness centrality”, “density”, “contribution index”, and “degree” have been computed and plotted. Productivity of each development group is measured in terms of two indices – performance and creativity. The results indicate that there is a correlation between attributes of social networks such as density and betweenness centrality and group productivity measures in an open source development community.

Background
The Internet has enabled virtual teams to link across distance, time, culture, departments and organizations, thereby creating "anyone - anytime - anyplace" alternatives to the traditional same-time, same-place, functionally centered, in-house forms of organizational experience [Lueg, C. Fisher, D., 2003].

Differences in productivity observed in open source development teams have created increasing interest to study their pattern of communication and further uncover the existence of association between communication pattern characteristics (social network attributes) and teams’ productivity measures such as performance and creativity. In this paper a time series analysis of the communication patterns of thirty-three different component development groups of the Eclipse open source developers’ community [Eclipse project, 2004] was conducted. Assessment of their productivity was carried out in a bid to identify correlation between the two variables performance and communication patterns.

Communication Attributes and Productivity Measures
There are many attributes used to measure the characteristics of social networks. Two of the most prominent measures that are considered in this paper are betweenness centrality and density. Betweenness centrality measures how an actor controls information and relative access to network resources and information. It can also be interpreted as measuring the degree of independence from others in the network. The TeCFlow tool calculates it in a range from zero to one. The two possible extremes are a 100% centralization of the entire network, which is a case where one actor is bridge to all others, (star networks with betweenness centrality equal to one), and the other extreme is 0% centralization, which is a case where no actor is a bridge of another actor (circle network with betweenness centrality measure equal to zero).
Density on the other hand measures the readiness of the group to respond to changes, its complexity and solidarity. It is defined as the percentage of ties that exist in a network out of all possible ties. A density of 1 implies that every actor is connected to every other actor. A density of 0 implies that no actor knows any other actor. Productivity in software development is not only a measure of the time and/or cost required in delivering and maintaining a software system; but it is also a measure of the usefulness of the software system in satisfying the customers’ and users’ needs and expectation. Hence there are two major dimensions of software engineering productivity, the change in the quantity of software produced for a given period of time at a given cost (the process dimension) and the quality of the resultant software system (the software system dimension) [A.S. Duncan, 2003]. This paper limits itself to productivity measures from the software system perspective.

It considers two attributes, size and defect. “Size” measures how the software system evolves in time, enhancements and additional features incorporated. “Defect” is the amount of bugs made during a given period of time. These two attributes give rise to productivity indices used in this research i.e. creativity and performance where:

\[
\text{Creativity Index} = \frac{\# \text{ enhancements integrated in time period}}{\# \text{ bugs resolved in the same time period}}
\]

and

\[
\text{Performance Index} = \frac{\# \text{ bugs resolved in time period}}{\# \text{ bugs reported in the same time period}}
\]

Hypotheses
The following five hypotheses are defined:

**Hypothesis-1:** Decentralized software development groups promote performance and creativity by enabling members to share knowledge in a more efficient and effective manner than centralized ones. In centralized social networks, dissemination of knowledge takes longer as it needs to travel through the extended hierarchies. Therefore, the following hypothesis should hold “Group betweenness centrality is negatively correlated with group performance and creativity”.

**Hypothesis-2:** As the number of ties between actors in a social network grows; density, performance and creativity of a group increase. Alternative ways for knowledge to flow through the network will grow which facilitate collaboration. Therefore performance and creativity of software developers should improve. Therefore the following hypothesis follows “Group density is positively correlated with group performance and creativity”.

**Hypothesis-3:** The last hypothesis follows from the economic concept of “opportunity cost”: If a given developer group has to carry out competing tasks with the same resources (fixing bugs and plan, design and implement new features or enhance existing ones), there will be a tradeoff. As a result, an increase in one activity will entail a decrease in the other. Hence, hypothesis three follows “Group performance and creativity are negatively correlated to one another”.

**Hypothesis-4:** Teams that have a constant communication structure perform better. For high performance, it is therefore better for a group to have similar communication patterns over the lifetime of the group. It therefore follows that “Group performance is negatively correlated with the number of changes, i.e. fluctuations in betweenness centrality over time”.

**Hypothesis-5:** Teams that have changing communication structures are more creative. We speculate that alternations in communication structure between execution mode by hierarchical communication (high group betweenness centrality), and dense, balanced communications for brainstorming activities (low group betweenness centrality) are indicators of high creativity. It follows that “Group creativity is positively correlated with
the number of alternations in group betweenness centrality over time”.

**Research Approach**

To better understand the correlation between temporal communication patterns and performance of open source developer communities, mailing list archives of thirty-three Eclipse communities [Eclipse mailing lists, 2004] [Eclipse project, 2004] are analyzed. These communities are working on different modules of Eclipse such as the Java Development Tools, the different platform components, the C/C++ Development Tools and the AspectJ extension. Six months communication data was analyzed. Mailing lists are considered as an approximation of the social network for the fact that they are main means of communication by developers actually working on or otherwise contributing to day-today development to discuss and make decision on design and implementation issues.

Social network data was collected from the Eclipse component development groups’ online mailing lists by using the online process tool. Data on bugs and enhancements per each group was collected from the Eclipse bugzilla database by using built-in queries [Eclipse bugzilla, 2004]. The social network data was analyzed with the TeCFlow tool [Gloor, P. Zhao, Y., 2004]. The graph structure, group betweenness centrality, group density and similar attributes were calculated for each software development group. Based on the data found from the bugzilla bugs database performance and creativity indices were computed. Following these, statistical analysis, correlation and regression were used to test the hypothesis formulated above. Finally, on the basis of the results obtained from the analysis and hypothesis-testing phases a conclusion of the study is drawn.

**Analysis and Interpretation**

In the data collection phase communication data from mailing lists was parsed and imported into MySQL databases. Then, the TecFlow tool was used to merge datasets for the same “component” project groups in to one. Data cleansing was performed to avoid duplicate presence of the same entity in the database and mass and /or group mails. The TeCFlow tool was used to automatically calculate group betweenness centrality and density from the communication data stored in the database. There is neither manual step nor subjective interpretation involved in this process. The algorithms for calculating these social network measures are pre-built in the tool. By using the bugs data obtained from the bugzilla bugs database and formula described above creativity and performance indices were computed.

A correlation analysis between group betweenness centrality and creativity reveals an important relation between these two attributes: when the groups get more central the creativity drops, and vise versa. Groups, which tend to be decentralized, make proportionally more enhancements within the analyzed time period than centralized ones. On the other hand, group betweenness centrality shows an insignificant influence on the performance of a software development group (see table 1). A bi-variate correlation analysis supports hypothesis-2 that there is a strong positive correlation between density and performance of a group. The group with the highest density has resolved the highest number of bugs reported in the analyzed time period and the lowest–density group has the lowest performance index. Even though an inverse relation between density and creativity is observed, the relation is not statistically significant enough. Therefore, in this study, density was not an important social network attribute to affect creativity of a software development group.

<table>
<thead>
<tr>
<th>Performance Index</th>
<th>Creativity Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>P-Value</td>
</tr>
<tr>
<td>Group Betweenness</td>
<td>0.292</td>
</tr>
<tr>
<td>Group Density</td>
<td>0.369</td>
</tr>
</tbody>
</table>

Table 1. Correlation between Creativity - Performance and Density - Centrality
Our research indicates that performance and creativity are negatively correlated to one another with a correlation coefficient of $B=-0.224$, supporting Hypothesis-3. This illustrates that the economic law of “opportunity cost” prevails even in open source software development teams. Stated differently, adding more enhancements comes at the price of slower bug fixing. The relation discovered among group betweenness centrality, density, performance and creativity is illustrated by the social network graphs of the Eclipse open source developer teams in Figure 1 below.

Figure 1 illustrates our hypothesis; it displays snapshots of the network graphs of communication networks, temporal evolution of betweenness centrality, degree centrality, density, and creativity and performance numbers of two different Eclipse developer communities. As seen from the figure, the group with high communication density exhibit higher productivity index than the other one. Further, the lower creativity is observed in the group with relatively centralized communication pattern.

**Temporal Analysis**
We also aimed to better understand the impact of changes over times in social network structure on productivity and creativity of the software development groups. Towards that goal, data on the number of changes in the groups’ density, betweenness and degree centrality has been collected. We counted the number of peaks and troughs in the temporal density, betweenness, and degree centrality plots, i.e. the number of their local minima and maxima. Correlation analysis between these datasets and measures of performance and creativity has revealed interesting results as shown in Table 2 below.

<table>
<thead>
<tr>
<th></th>
<th>Performance Index</th>
<th>Creativity Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pearson Correlation</td>
<td>P-Value</td>
</tr>
<tr>
<td>Group Betweenness</td>
<td>-0.534</td>
<td>0.003</td>
</tr>
<tr>
<td>Group Density</td>
<td>-0.48</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Table 2: Correlation between creativity – performance and changes in group centrality and density

Statistically significant relationship between changes in groups' centrality, performance and creativity has been identified. In support of hypothesis-4, we found that groups, which are steady in their density over time, are more performing. On the other hand, groups, which exhibit more change in their centrality, are found to be more creative and less performing. This finding supports hypothesis-5, illustrating that oscillation between hierarchical and decentralized communication structure is a strong indicator for creativity. Similar analysis on density versus creativity has shown no statistically significant correlation between these two variables.

Figure 2 illustrates our findings graphically. The group at the top with a more fluctuating dynamic evolution of group betweenness centrality (top of Figure 2), and a changing social temporal surface [Gloor, P. 2005b] has higher creativity and lower performance than the group at the bottom of Figure 2.
Figure 2: Temporal social surfaces illustrating correlation between temporal evolution of group betweenness centrality and performance and creativity

Conclusion
The eclipse open source development groups, which tend to be central, are found to be less creative when compared to decentralized ones. Creativity in this sense is the amount of feature enhancement carried out by eclipse component development groups. The reason for this could be that as communication between development teams gets centralized, it prevents innovative ideas from coming up to the floor. On the other hand, it is found that centrality of a software development group is not a major factor affecting performance. This study illustrates that Eclipse development groups with high communication density seem to be better performers than those with low density. As development members get more connected they have better ways of dealing with bug resolution. More software bugs get fixed in an effective and efficient manner. Our continuing goal is to come up with recommendations for communication in virtual teams based on insights obtained through temporal social network analysis of open source developer teams. Whereas oscillations over time in group betweenness centrality and density are an indicator for high performance. This research presents preliminary results on the relationship between communication structure and productivity of open source developer teams. While conclusive evidence is not yet found that lower group betweenness centrality is correlated with higher creativity, the study has shown that higher density is an indicator of higher performance. Future work should focus on studying larger numbers of open source software development communities to come up with more insights that will help generalize the kind of correlation between group communication characteristics and productivity variables in open source development communities as a whole. It would also be useful to consider additional social network characteristic in the study. One should also consider how social network characteristics are related to the “process dimension” of software development groups’ productivity measures in order to see how cost of the development (in human resources, hardware resources, and calendar time), is affected by communication, collaboration and coordination variability.

References
Eclipse bugzilla bugs database. Retrieved on December 1, 2004 at URL: https://bugs.eclipse.org/bugs/reports.cgi
Eclipse mailing lists Retrieved on September 1, 2004 at URL http://www.eclipse.org/mail/index.html
Gloor, Peter A. 2005, “Swarm Creativity, Competitive advantage through collaborative innovation networks”, Oxford University Press
ABOUT THE MIT CENTER FOR DIGITAL BUSINESS
Founded in 1999, the Center for Digital Business is the largest research center in the history of the Sloan School. We are supported entirely by corporate sponsors whom we work with closely in directed research projects. The Center has funded more than 45 Faculty and performed more than 60 research projects. Our mission is to join leading companies, visionary educators, and some of the best students in the world together in inventing and understanding the business value made possible by digital technologies. Our interactions are a dynamic interchange of ideas, analysis, and reflection intended to solve real problems.

Examples of Current Focused Research Projects:
- Implications of e-Commerce for New Services and Structure of Logistics Systems
- How Do Intangible Assets Affect the Productivity of Computerization Efforts?
- Wireless and Mobile Commerce Opportunities for Payments Services
- Benchmarking Digital Organizations
- Security and the Extended Enterprise
- Pricing Products and Services in the High-Tech Industry

The Center for Digital Business is completing its Phase II, focusing more explicitly on business value, while at the same time including technologies beyond the Internet in its purview. Our goal, in part, is to reduce that timeline through basic and applied research, engagement with industry sponsors, and the sharing of best practice, and the MIT’s credo of combining rigor with relevance is well served.

We are organized into four areas of expertise – or Special Interest Groups:
1. Digital Productivity
2. Digital Marketing
3. Digital Services
4. Digital Health

Founding Sponsors
- BT
- Cisco Systems
- CSK Corporation
- France Telecom
- General Motors
- SAP
- Suruga Bank
- UPS

Research Sponsors
- Hitachi
- Liberty Mutual
- University of Lecce

Member Sponsors
- Intel
- Institute for Innovation and Information Productivity
- SAS

CONTACT INFORMATION
MIT Center for Digital Business
MIT Sloan School of Management
3 Cambridge Center, NE20-336
Cambridge, MA 02142
Telephone: (617) 253-7054
Facsimile: (617) 452-3231
http://digital.mit.edu/

David Verrill, Executive Director
Erik Brynjolfsson, Director
Glen L. Urban, Chairman
Carlene Doucette, Executive Assistant
Tammy Buzzell, Financial Assistant