Identifying Collaboration Patterns in Software Development Social Networks

Taisa Alves dos Santos[1,2], Renata Mendes de Araujo[1,2], Andrea Magalhães Magdaleno[2,3]

[1]Graduate Program in Informatics
[2]Research and Practice Group in Technology
Federal University of the State of Rio de Janeiro (UNIRIO)

[3]Systems and Computing Engineering Program
Federal University of Rio de Janeiro (UFRJ)

Abstract. *Software development is a collaborative activity which involves the effective coordination of groups displaying variations in their skills and responsibilities. This paper argues that, by understanding the way collaboration is performed, participants and managers can better understand the development process in order to conduct their activities. This paper proposes an approach based on social networks analysis to identify collaboration patterns in software development process instances which can be used as a resource for collaboration awareness and understanding.*

Keywords: Collaboration, software process, awareness, social networks.

1 Introduction

Software development is characterized as a collaborative activity [9]. One of the main challenges of coping with collaboration both in distributed and co-located settings is how to make the work visible to all participants, making them aware of what is happening in the development process [1]. To face this challenge, proposals for collaborative support through computational tools have been suggested [4] [16] wherein collaborative supporting aspects are provided, such as coordination, communication, group memory, and awareness [1] [6] [8].

This work suggests that the social network [20] achieved as a result of software development interactions can provide information about the collaboration existing therein. However, only the view of a social network topology using visualization tools [3] [19] may not be enough to help participants and project managers to understand and analyze the collaboration level of the team. This work proposes the possibility of identifying collaboration patterns through the analysis of social networks properties. According to the collaboration patterns and with the help of social network visualization tools, developers and project managers will be able to interfere, change, redistribute or reflect about the process and work being conducted.

This paper is organized as follows: Section 2 reports the research work on how to provide awareness in software development processes; Section 3 summarizes social network properties and the tools that can be used to identify collaborative patterns; Section 4 discusses CollabMM as a reference for identifying collaboration levels in business processes; Section 5 presents preliminary essays in identifying collaboration patterns using social network properties based on CollabMM. Section 6 concludes the paper and outlines future work.

2 Awareness in software development processes

In collaborative support, awareness can be defined as being conscious of the presence of other users and of their actions while interacting through applications [6][14][17]. Awareness aims to reproduce or even increase, in a virtual environment, the elements of a real, face-to-face interaction. To achieve this objective, awareness mechanisms can be used to represent, for instance, the presence of a group member, the position of each participant in the shared workspace, or even to distinguish each participant by using different colors [7]. These mechanisms are used to extend user awareness about information they cannot notice alone or information that they would possibly not consider as relevant for the work [15].

In this work, we consider three types of awareness for software development processes [1]: social, process and collaboration. Social awareness allows users to recognize the group in which they are included for a possible interaction. Process awareness involves acknowledging the current process enactment state, the activities complete, the activities being performed, which activities are waiting to be performed by an individual and which should be performed by the entire group. Collaboration awareness focuses collaboration among group members, contributes to the understanding of their interactions inside the group and fosters future improvements in process interactions. All these kinds of awareness have been studied by different researchers and were implemented in collaborative tools to support group work in software development. The following paragraphs present examples of tool proposals for each type of awareness.

The OpenMessenger tool [5] represents the social awareness by "tickets" considered a user's photo avatar. Users can rotate their avatar to indicate how busy they are. An avatar in full view indicates that the user is available, and the more the picture is turned away, the busier the user is (Figure 1).

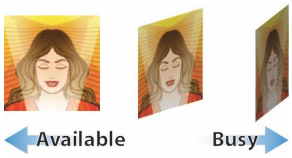

Figure 1 – Avatars and rotation in OpenMessenger [4]

The PIEnvironment [1] explores the possibility of extracting information about participants' interactions from the software process models defined to be enacted in a workflow system. In this case, process awareness can be understood through the sequence of activities performed by the group; collaboration awareness is presented by modeling user interactions extracted from process enactment (Figure 2).

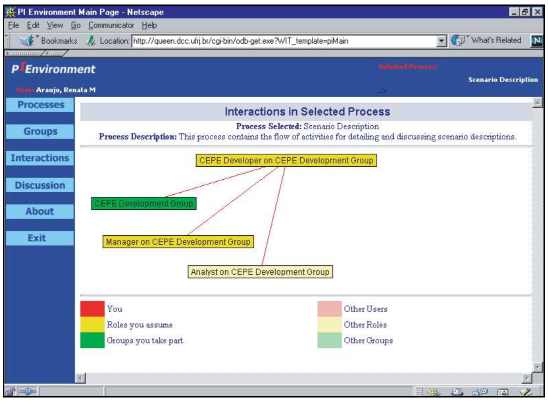

Figure 2 – Interaction graph (Collaboration awareness) [1]

Processes mining approaches [18] also strive to obtain, from the process execution data, the identification of interactions which may occur within process participants. According to an event log (Table 1), obtained from workflow tools, the social interactions which occurred in the work environment of a particular team can be understood and visualized. Figure 3 shows a social network mined from the event log in Table 1. The first graph (Figure 3a) shows the control-flow structure expressed in terms of a Petri net. The second (Figure 3b) is the organizational structure expressed in terms of an activity-role-performer diagram, and the last one (Figure 3c) is a sociogram based on transfer of work done.

The approaches presented above rely on the possibility of collecting data to be presented as awareness information for process participants interacting through computational tools. They focus on how to provide development teams with the resources for being aware of the process they execute. These awareness resources play a fundamental role in helping people recognize and learn the way they actually work, as well as recognize problems and improvement possibilities.

Case	Activity	ID	Timestamp
Case 1	Activity A	John	9-3-2004:15.01
Case 2	Activity A	John	9-3-2004:15.12
Case 3	Activity A	Sue	9-3-2004:16.03
Case 3	Activity B	Carol	9-3-2004:16.07
Case 1	Activity B	Mike	9-3-2004:18.25
Case 1	Activity C	John	10-3-2004:9.23

Table 1 – Event log [18]

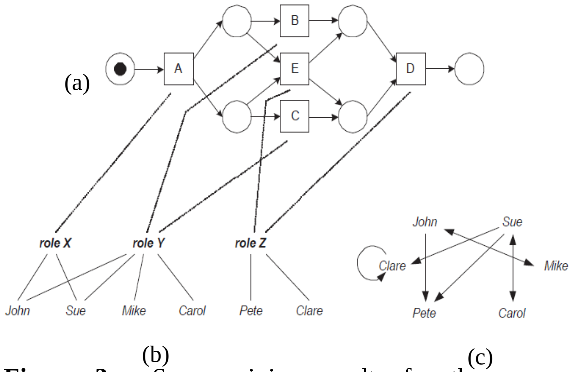

Figure 3 - Some mining results for the process perspective (a) and organizational (b and c) perspective based on the event log shown in Table 1 [18].

In this work, we discuss how this information can be used for understanding the levels of collaboration being achieved by a team. We claim that it is possible to identify collaboration patterns from the analysis of process interactions. These collaboration patterns can help participants and project managers to understand the different levels of collaboration existing and make decisions about changes on improving the process.

3 Social networks

The concept of 'network' is as simple as: a set of links among nodes. A social network means the set of links among people [20], where a node represents an actor and links among actors[1] represent possible relationships among them. The semantics of a link depends on which analysis we wish to conduct. This can be communication, relationship, friendship and so on. Social network analysis is a way to understand the interaction and social organization within a group [3]. In software development, we aim at understanding coordination and communication relations among process participants.

3.1 Social networks properties

Social networks can be examined through the analysis of its properties [20]. For the purpose of this work we have selected an initial set of properties which we believe have potential to provide information about collaboration patterns.

---

[1] For the purpose of this work, the terms *actor* and *node* will be used as synonyms.

Properties related to actor centrality are based on the links one actor bears with other actors [20]. Therefore, each actor has a value within the network which can be considered when comparing it to the other nodes. These properties render the node more visible to other actors. There are three types of actor centrality [20]:

Degree centrality: the degree centrality of the actor is measured by the inputs and outputs of the node, i.e. it sums the number of its relationships [7]. The actor with high degree centrality will be in direct contact with more actors, occupying a central role in the network. The node which has the greatest value is called a central node. Central nodes in a network are called *hubs* [3] [20]. The Figure 4 is an example of a social network with four actors. Node 3 is the central node because its degree centrality is equal to 3.

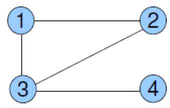

Figure 4 – Degree centrality

Betweness centrality: measured by the number of times a node appears in the path of other nodes [7]. Actors which are between two nodes which are not neighbors have control over the link between them. To have high betweness centrality, an actor must be in the path of different actors. In the Figure 5 the node 2 is the actor with the higher betweness centrality because it is on the way of 1, 3 and 4 actors.

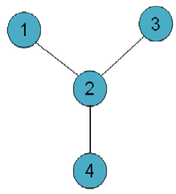

Figure 5 – Betweness centrality

Closeness centrality: calculated by the inverse of the sum of distance between one source node to different destination nodes [7]. This property is based on distance and represents how close or far an actor is to other nodes. A central node, for instance, can interact quickly with other nodes and can be highly productive in information sharing with the overall group, as they have a fast communication path with other nodes. In the star network, presented in Figure 6, the node 2 is adjacent to all others. Therefore is has maximum closeness centrality – starting from node 2, any other node can be reached following just one link.

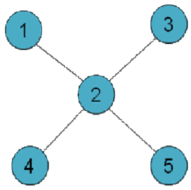

Figure 6 – Closeness centrality

3.2 Social network tools

The properties presented in the last session can be used as a basis for social network mining and visualization tools [2][13][19]. The SVNNAT tool gives evidence of collaboration awareness to software development group over the analysis of data extracted from Subversion (SVN) configuration system [13]. The Figure 7 shows an example of the social network exported and the data mined.

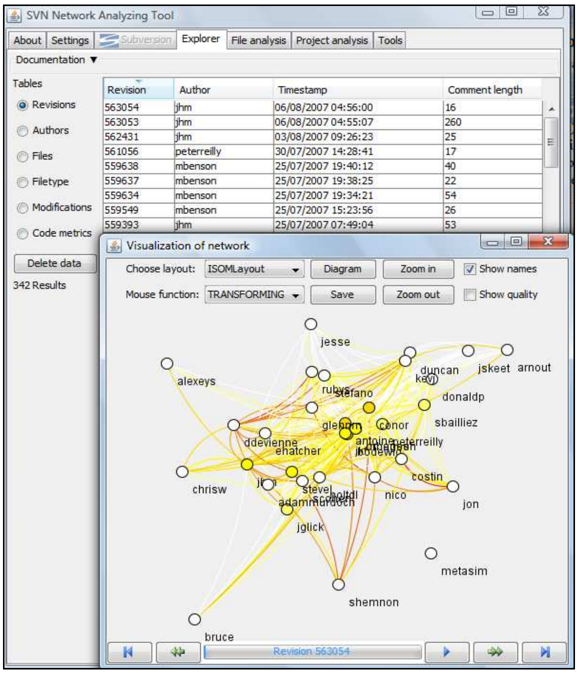

Figure 7 – SVNNAT tool [13]

The OSSNetwork tool [2] extracts from open source development communities the interactions which occur among group members and the source code, mail lists and forums. Figure 8 shows the social network exported and the properties for this network provided by the tool.

The MiSoN tool (Figure 9), which is part of the ProM framework [19], is used to mine social networks extracted from workflow event logs. The event log (Table 1) is the input data used to generate the social network. This tool allows the analysis of mined networks using the above mentioned properties.

The visualization of a social network topology and the availability of its properties provided by these tools are relevant information to allow for understanding relationships in a work group setting. However, they are not enough to permit understanding the level of collaboration therein.

In this work, we argue that social network topology and properties can be associated to different levels of group collaboration maturity. In order to evaluate that, a collaboration maturity framework – CollabMM – was used and will be detailed in the next section.

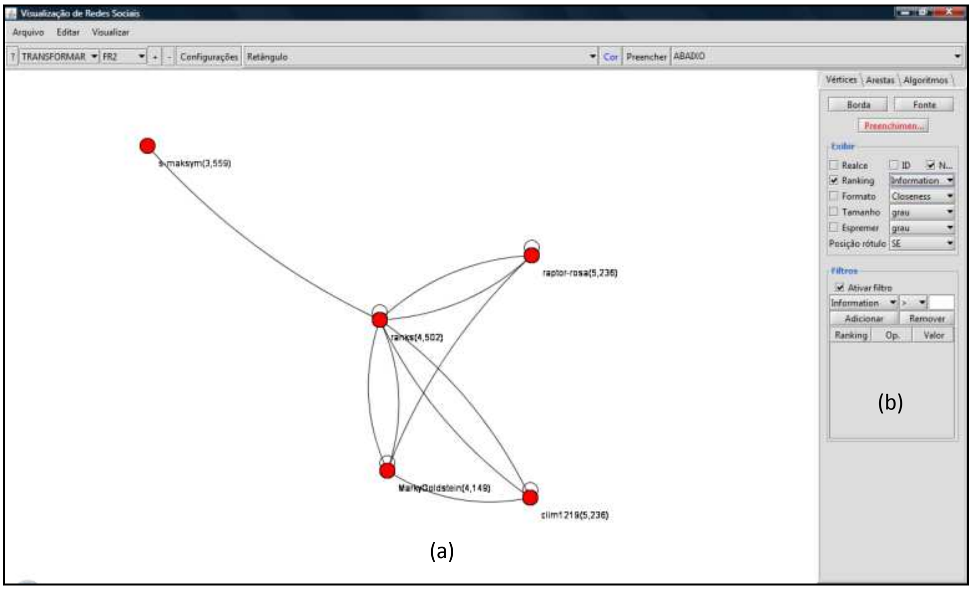

Figure 8 – OSSNetwork tool [2] (a) Social network mined, and (b) properties to analyze social networks.

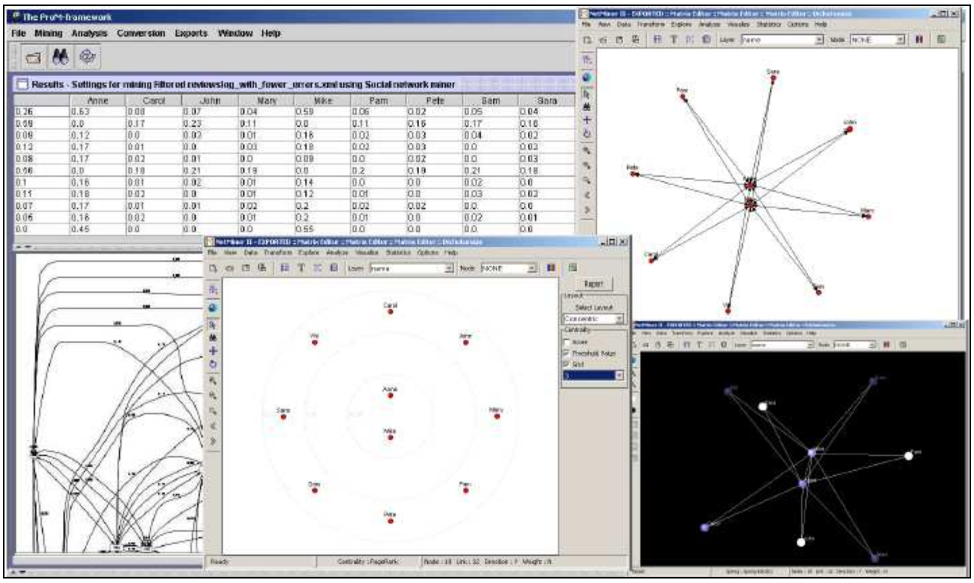

Figure 9 – MiSoN tool [19]

4 The CollabMM model

Magdaleno et al. [10] proposed a collaboration maturity model for business processes – CollabMM - that aims to organize a set of practices which can enhance collaboration in business processes. CollabMM describes an evolutionary path in which processes can progressively achieve higher capability on collaboration, organized in four maturity levels: Ad-hoc, Planned, Aware and Reflexive, as shown in Figure 10. Levels are a way of prioritizing practices for improving collaboration in a process, according to the collaboration support aspects (communication, coordination, group memory and awareness). A specific level comprises a group of related activities which can be executed together, aiming at improving process collaborative capability (Figure 10). The CollabMM collaborative levels can be summarized as follows:

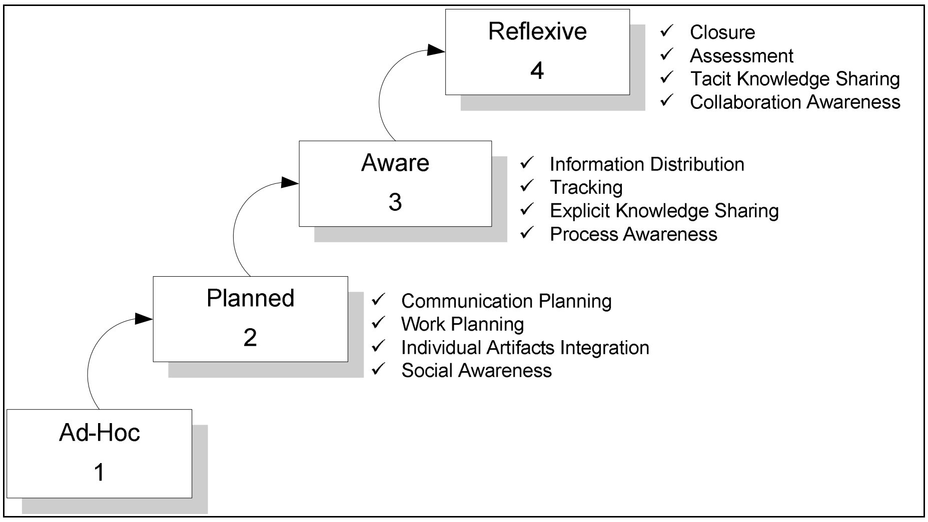

Figure 10 – CollabMM model [10]

- **Ad-hoc level**: In this level, collaboration is not explicitly represented in a process. However, processes in this first level cannot be featured as with total absence of collaboration. Collaboration may happen, but it is still dependent on individual initiative and skills, and its success depends on the relationship and/or affinity among people. The aspects of communication, coordination, group memory and awareness are present, but they occur in an ad-hoc manner. Figure 11 presents a metaphor of individual effort where people do not act like a group.

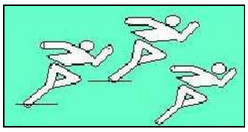

Figure 11 - Metaphor for Level 1 – Ad-hoc [10]

- **Planned level**: In this level, business processes start to be modified aiming at including basic collaboration activities. The coordination is a strong aspect in this level because groups need leadership and management in order to work well. Work groups – created to execute a project, process or a specific activity – are formally established. Figure 13 shows a metaphor for this level.

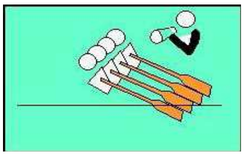

Figure 12 - Metaphor for Level 2 – Planned [10]

- **Aware Level**: In this level, the process includes activities for monitoring and controlling how collaboration occurs. Centralized coordination is not highly relevant, since group members are aware of their tasks and responsibilities and are committed towards them. Group members understand the process in which they are engaged and, its main objectives, as well as their roles and responsibilities and how their activities are related with others to perform these objectives. Additionally, processes at this level are characterized for shared knowledge, mainly through the artifacts produced by the group. Figure 14 shows a metaphor for this level.

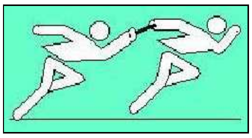

Figure 13 - Metaphor for Level 3- Aware [10]

- **Reflexive level**: In the reflexive level, processes are designed to provide self-understanding, identifying the relevance of the results which had been produced and sharing this knowledge inside the organization, this can be represented by metaphor of collective disseminated effort in Figure 14. Considering communication, processes must be formally concluded and their results communicated. Lessons learned can be captured; strengths and weaknesses are analyzed; successes and challenges are shared; ideas for future improvements are collected; and workgroup results are published and celebrated. Group members are aware of the way in which the group collaborates during process execution, while process tacit knowledge is shared through ideas, opinions and experiences, thereby enhancing group memory.

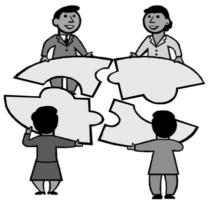

Figure 14 - Metaphor for Level 4 – Reflexive [10]

CollabMM has been used to assist organizations in introducing different levels of collaboration in their business process models [11]. It also has been discussed as a framework for assessing collaboration levels in a business process [10].

Our aim in this work is to use CollabMM as a guide, based on the properties of the social network produced in a development process, for identifying collaboration patterns or levels, as discussed in the next section.

5 Identifying collaboration patterns

The purpose of this section is to discuss our hypothesis on how collaboration patterns can be identified from social networks, reviewing and detailing previous ideas presented in [12]. The main idea is that social network properties can be associated to the characteristics of the different collaboration levels suggested by CollabMM.

All social networks exemplified in this paper were mined since September 2009 from Sourceforge.net *(http://sourceforge.net/)*. The software development projects selected should meet the following criteria: more than 5 years of community activity; and an expressive number of downloads, characterizing their stability as development communities.

From these social networks, developer's interactions can be perceived using information obtained from the online source code repository history. In these networks, the nodes represent the developers and the relationships are established among the individuals who work in the same part of the code, modifying it. The SVNNAT tool [13] was used to mine and visualize these social networks that represent the intrinsic collaboration.

Our hypothesis is that the degree, betweness and closeness centralities are the social network properties which emphasize the coordination aspect. The CollabMM levels explore the coordination as a strong aspect of collaboration and the collaboration patterns will be divided according to these levels.

The social networks classified as planned level will be fitted in the collaboration pattern that seeks for centralized coordination. In other words, in this type of network there is the presence of hubs, which predominate in the three centralities properties (degree, betweness and closeness), and give evidence of the "winner takes all" pattern [3].. This pattern describes the idea that a single node becomes so strong that it may dominate the network

In the aware level, social networks will show a decentralized coordination. This type of network will have more than one hub, so new central nodes will appear according to degree, betweness and closeness centralities. For example, in open source projects, the core development team shares the coordination of the project.

In the reflexive level, the coordination tends to be distributed and the figure of the central node disappears. The degree, betwenness and closeness centralities values are too close, where the existence of hubs is not clear.

5.1 Ad-hoc level

In this level of collaboration, social networks may not show specific collaboration patterns. The relationships among the participants of this network vary extensively, with instability and, possibly, lack of patterns for analysis.

5.2 Planned level

As described in the CollabMM model, the coordination is an important aspect of the planned level. Coordination at this level is characterized by strong leadership and management in order to guide all the work. The collaboration pattern of this level is characterized from the degree, betweness and closeness centrality properties. The existence of a strong central node or hub, may characterize the network at this level as a centralized social network [4].

The social network obtained for the devkitPro project is an example. Figure 15 shows the devkitPro social network and the Table 2 details the data analyzed for each of nodes that represent the devkitPro social network. As noticed by the properties in Table 3, *wntrmute* is a central node that stands out among the rest. This probably characterizes this developer as a key node for project and work coordination.

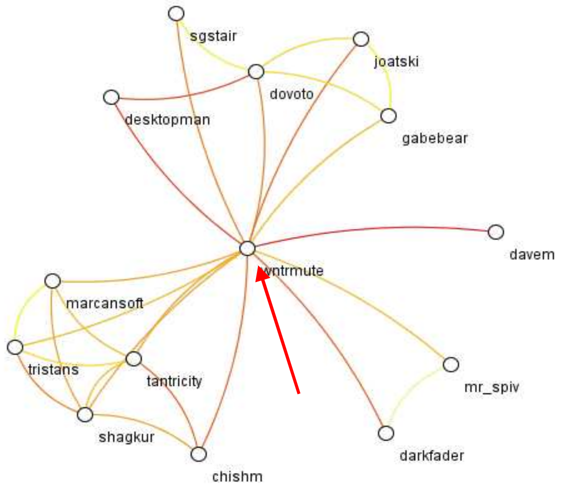

Figure 15 – DevKitPro Social Network

Actor	Degree	Betweness	Closeness
wntrmute	**828.00**	**60.00**	**100.00**
shagkur	272.00	0.66	61.90
dovoto	259.00	2.50	61.90
tantricity	256.00	0.66	61.90

Table 2 – DevKitPro project data

5.3 Aware level

At this level, the group members are aware of their tasks and responsibilities and can act more autonomously. So, the main characteristic of the reflexive level is the existence of more than one hub, differently previous level, a few number of nodes distinguish according to betweness and closeness centrality. The coordination becomes decentralized, since the existence of more than one actor representing the central node [4].

Figure 16 illustrates this collaboration pattern and the software project WinMerge data is detailed below in the Table 3. The nodes *kimmov* and *gerund* are hubs according to degree, betweness and closeness centrality. This emphasizes the collaboration pattern that represents the aware level.

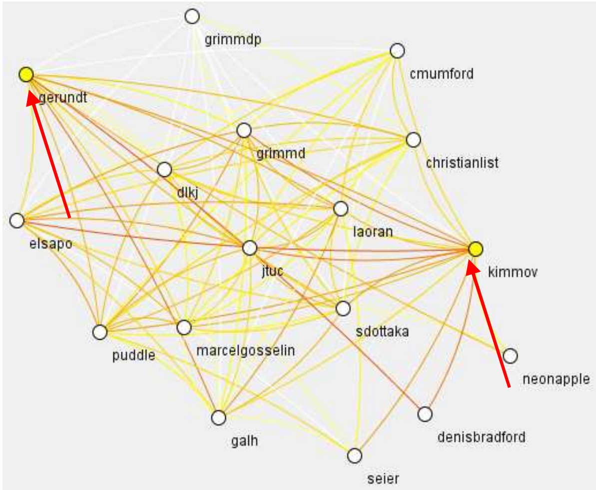

Figure 16 – WinMerge Social Network

Actor	Degree	Betweness	Closeness
kimmov	**732.00**	**12.92**	**100.00**
gerundt	567.00	**12.92**	**100.00**
puddle	547.00	1.75	88.88
elsapo	537.00	1.75	88.88
laoran	535.00	1.75	88.88

Table 3 –WinMerge project data

5.4 Reflexive level

In the reflexive level, the main characteristic is knowledge exchange and the self-understanding about the group work. The collaboration pattern that represents this level can be perceived by the absence of hubs. Different nodes have very close degree, betweness and closeness centralities values.

Figure 17 shows the NHibernate project social network that may represent a network at the reflexive level and Table 4 details the values of its nodes properties and demonstrates the reflexive social network level has a distributed coordination between nodes [4].

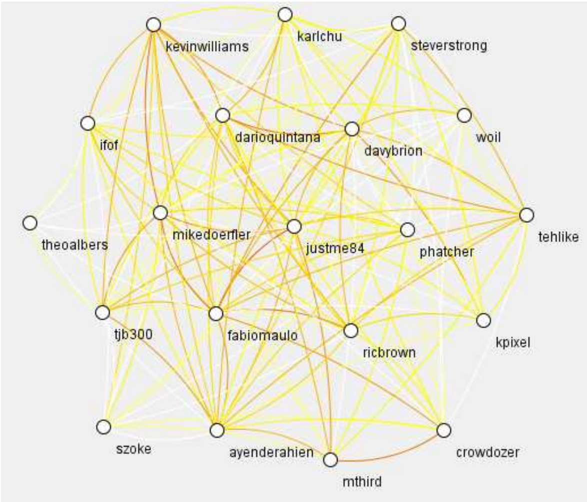

Figure 17 – NHibernate Social Network

Actor	Degree	Betweness	Closeness
fabiomaulo	687,00	4,86	100,00
justme84	644,00	4,86	100,00
ayenderahien	596,00	4,86	100,00
darioquintana	564,00	4,86	100,00
kevinwilliams	561,00	1,16	79,16
fabiomaulo	687,00	4,86	100,00

Table 4 – NHibernate project data

Table 5 summarizes our hypothesis of the relationship between social network properties and CollabMM maturity levels. Each of the social networks properties emphasize the aspect of coordination. The various levels of the CollabMM model attend the coordination aspect in different ways which is also the goal of our proposal.

Level	Degree Centrality	Betweness Centrality	Closeness Centrality
Planned	Single central node	Single central node	Single central node
Aware	Few central nodes	Few central nodes	Few central nodes
Reflexive	No central nodes	No central nodes	No central nodes

Table 5 – Collaboration patterns as CollabMM levels

6 Conclusion

In this paper, we discuss the potential of social networks analysis in the identification of collaboration patterns in software development processes. Our aim is to contribute to research related to the understanding of collaboration in different development models – in house/distributed, disciplined/agile/open source - arguing that the understanding of collaboration can be a way to promote balance between these different approaches, as well as, a tool for management purposes.

As future work, it will be necessary to evaluate and detail our hypothesis by conducting different analysis over different development process settings. Further, the information about collaboration patterns derived from this analysis can be used as input for development process enactment or management tools in order to help managers and participants to be aware of collaboration they participate in.

Acknowledgments

This work is partially funded by CNPq under processes nº. 142006/2008-4 and 557.128/2009-9 and by FAPERJ (Carlos Chagas Filho Foundation for Research of the State of Rio de Janeiro) under process nº. 101.252/2008 and E-26/170028/2008 (Programa INC&T - Projeto: Instituto Brasileiro de Pesquisa em Ciência da Web).

References

[1] ARAUJO, R. M., BORGES, M. R. S., 2007, "The role of collaborative support to promote participation and commitment in software development teams". Software Process: Improvement and Practice, v. 12, n. 3, pp. 229-246.

[2] BALIEIRO, M. A., SOUSA JÚNIOR, S. F. DE, DE SOUZA, C. R. B., 2008, “Facilitating Social Network Studies of FLOSS using the OSSNetwork”, In: Open Source Development, Communities and Quality: International Conference on Open Source Systems, Milan, Italy. Proceedings of International Conference on Open Source Systems. Spring Street, NY: IFIP International Federation for Information Processing, a Springer Series in Computer Science, 2008. p. 343-350.

[3] BARABASI, A. L., 2002, Linked: The New Science of Networks. Cambridge, MA, Perseus Publishing.

[4] BARAN, P., 1964, “On distributed communications network”, IEEE Transactions on Communications, Vol. 12, No. 1, March, pp. 1-9.

[5] BIRNHOLTZ, J. P., GUTWIN, C., RAMOS, G., WATSON, M., 2008, “OpenMessenger: gradual initiation of interaction for distributed workgroups”, In: Proceedings of CHI’2008, pp. 1661-1664.

[6] DOURISH, P., BELLOTTI, V., 1992, "Awareness and coordination in shared workspaces". In: Proceedings of the 1992 ACM conference on computer supported cooperative work, pp. 107-114, Toronto, Ontario, Canada, November.

[7] FREEMAN, L. C., 1979, "Centrality in Social Networks: I. Conceptual Clarification." Social Networks, 1, 215-239.

[8] GUTWIN, C., PENNER, R., SCHNEIDER, K, 2004, “Group awareness in distributed software development”. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW, p 72-81, Computer Supported Cooperative Work - Conference Proceedings, CSCW 2004.

[9] KRISHNAMURTHY, B., NARAYANASWAMY, K., 1994, “CSCW 94 Workshop to Explore Relationships between Research in Computer Supported Cooperative Work & Software Process – Workshop Report”, Software Engineering Notes, v. 20, n. 2 (Abr.), pp.34-35.

[10] MAGDALENO, A. M., ARAUJO, R. M., BORGES, M. R. S., 2009, “A maturity model to promote collaboration in business processes”. In: International Journal of Business Process Integration and Management, v. 4, p. 111-123

[11] MAGDALENO, A. M., CAPPELLI, C., BAIAO, F. A., SANTORO, F. M., ARAUJO, R. M., 2008, “Towards Collaboration Maturity in Business Processes: An Exploratory Study in Oil Production Processes”. In: Information Systems Management, v. 25, p. 302-318.

[12] SANTOS, T. A. L., ARAUJO, R. M., MAGDALENO, A. M. “Padrões para Percepção da Colaboração em Redes Sociais de Desenvolvimento de Software”. In: Workshop de Desenvolvimento Distribuído de Software (WDDS), 2009, Fortaleza, Brasil (in portuguese).

[13] SCHWIND, M., WEGMANN, C., 2008, “SVNNAT: Measuring collaboration in software development networks”, In: Proceedings 10th IEEE Joint Conference on E-Commerce Technology and the 5th Enterprise Computing, E-Commerce and E-Services, CEC 2008 and EEE 2008, p. 97-104.

[14] SOHLENKAMP, M., 1998, “Supporting group awareness in multi-user environment through perceptualization”. Paderborn: Fachbereich, Mathematik-Informatik der Universität - Gesamthochschule.

[15] STOREY, M-A. D., CUBRANIC, D., GERMAN, D. M., 2005, “On the use of visualization to support awareness of human activities in software development: A survey and a framework.” In Proceedings of 2nd ACM Symposium on Software Visualization, 2005, St. Loius, Missouri, USA.

[16] TEE, K., GREENBERG, S., GUTWIN, C., 2006, “Providing artifact awareness to a distributed group through screen sharing”, In: Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW, p 99-108, Proceedings of the 20th Anniversary ACM Conference on Computer Supported Cooperative Work, CSCW 2006.

[17] TOLLMAR, K., SUNDBLAD, Y., 1995, “The Design and Building of the Graphic User Interface for the Collaborative Desktop”, In: Computers & Graphics. 19(2). 179-188.

[18] VAN DER AALST, W., REIJERS, H., WEIJTERS, A., DONGEN, B., MEDEIROS, A., SONG, M., VERBEEK, H., 2007, “Business Process Mining: An Industrial Application”, Information Systems, v. 32, n. 5, pp. 713-732.

[19] VERBEEK, H.M.W., DONGEN, B.F. van, MENDLING, J., AALST, W.M.P. van der, 2006, “Interoperability in the ProM framework”. In: Proceedings of the EMOI-INTEROP Workshop at

the 18th International Conference on Advanced Information Systems Engineering (CAiSE'06), pp. 619-630, Namur University Press.

[20] WASSERMAN, S. and FAUST, K., 1994, Social Network Analysis, Cambridge University Press, New York, NY.

[21] WWF, 2003, “Redes: uma introdução às dinâmicas da conectividade e da auto-organização”, Relatório Técnico, WWF - Brasil. Home Page: http://www.wwf.org (in portuguese).