

Applying Social Network Analysis and Centrality Measures to Improve Information Flow Analysis

Stephan Kiesling¹(✉), Jil Klünder¹, Diana Fischer², Kurt Schneider¹,
and Kai Fischbach²

¹ Software Engineering Group, Leibniz Universität Hannover, Hannover, Germany

{stephan.kiesling, jil.kluender,
kurt.schneider}@inf.uni-hannover.de

² University of Bamberg, Bamberg, Germany

{diana.fischer, kai.fischbach}@uni-bamberg.de

Abstract. In software development projects, documents are very important for sharing requirements and other information among employees. However, information can be transported in different ways. Conversations, meetings, workshops and emails convey and impart information as well. Especially large companies struggle in dealing with unclear and incorrect information flows. These information flows can be improved by means of information flow analysis and flow patterns. One technique to analyze information flows is the FLOW method. It supports visualization and analysis of information flows to detect lacks and anomalies and thereby improves information flows. An analyst gathers information transported in the company. Afterwards, information flows are visualized and analyzed based on patterns and personal experience. Nevertheless, analysis based on individual knowledge is error-prone. Hence, we improve the FLOW method with the help of social network analysis applying centrality measures to the FLOW method and to support the FLOW analyst.

1 Introduction

Documents play a major role in process-driven software development companies. They share requirements and other information among team members. However, information is also transferred through other channels like conversations, meetings or workshops. Agile methods, becoming more and more established these days, prefer direct information communication channels and use less documents. Especially big companies have to cope with unclear and wrong information flows in and between teams. Companies struggle to localize these communication problems. To find the problems, information flows must be analyzed.

Information flow analysis can uncover disruptions in information flows in companies, so that they can be corrected. The first step of the analysis is to determine the flow of information in the company. After that, the flow needs to be visualized and analyzed. The analysis may uncover findings that can be used for information flow improvement.

There are various possibilities to conduct an information flow analysis. One technique is the FLOW method, which is in focus in this paper [1].

FLOW analysis is a systematic method to visualize, analyze and improve information flow [2]. The FLOW method provides a graphical notation (Fig. 1) as well as an approach to identify and gather information flows. The graphical notation provides two different types of information flows. Non-documented, verbal or informal flows of information are denoted as fluid, whereas solid information flows are always documented. By using this graphical notation, so-called FLOW diagrams can be created. These diagrams represent all information flows and all information stores for a task or process. On the one hand, FLOW diagrams help the FLOW analyst to analyze gathered information flows. On the other hand, they can be used after analysis to present findings to the company the analysis was conducted for.

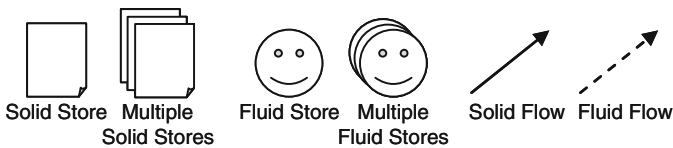


Fig. 1. The basic FLOW notation

The FLOW method consists mainly of three phases: (1) Information flows need to be collected by interviewing appropriate persons. (2) After the elicitation, FLOW diagrams are created by combining partial diagrams from each interviewee into an overall one. (3) Afterwards, the analysis phase starts. The analyst regards the FLOW diagram's structure and looks for bad patterns to decide on improvement recommendations. Moreover, the analyst also compares the depicted information flow with additional information elicited in phase (1) to look for possible contradictions or conflicts.

Mainly the last phase entails several problems. The analysis can be very subjective due to varying personal experiences. The analyst can overlook or misinterpret aspects. He may look into some details more closely and dismiss others only based on a gut feeling.

To reduce the influence of personal experience during analysis, we extend FLOW with methods from social network analysis. We choose centrality measures that are suited for FLOW and use them to compute key indicators for FLOW diagrams. An analyst can use these key indicators as support for his calculations and indicators that point to possible information flow issues. These indicators are specifically intended to support analysis of complex and large FLOW diagrams.

The remaining work is structured as follows. Section 2 presents related work. Section 3 introduces the concepts of the new method. Finally, Sects. 4 and 5 conclude this paper with a discussion and outlook.

2 Related Work

In the following section, we provide an overview of existing research on the relevant topics in this paper.

Modelling and analyzing **information flow** in organizations is motivated by the need to improve the information processes, e.g. by eliminating redundant processes, minimizing the duplication of information, or managing and sharing intra- and inter-organizational information flow [3]. The FLOW method is a method for diagrammatic modelling information flow in teams with the aim to identify critical points which may cause loss of information.

Stapel et al. [2] considered the information flow within different kinds of project teams to detect critical points like lacks of information, a wrong amount of information sharing or gatekeepers who are very central within the network.

Stapel [1] extends this approach by presenting possibilities to analyze the FLOW diagrams which visualize the information flows within the team or business.

In the organizational context, **social network analysis** can help understand how inter- and intra-organizational networks are linked to outcomes and processes [4], such as job exit, team performance, innovation and individual satisfaction (for an overview see, for example, [5]). In addition, social network analysis has been applied to explore the information flow in and between organizations. For example, Braha and Bar-Yam [6] analyzed the information flow structure of intra- and inter-organizational networks in large-scale product development organizations. Friedkin [7] found that the strength of relationships between employees has an impact on intra-organizational information flow. Ryyänänen et al. [8] observed the internal information flow network in the project sales process to understand and improve the information flow between the employees.

Information flow networks resulting from the FLOW method are directed networks. Analyzing directed networks requires a distinction between measures for incoming ties and measures for outgoing ties [9]. This is important for the measures, which will be applied in the network analysis in this contribution.

Social network analysis provides a wide range of measures to analyze a network (e.g. [9]). In order to understand and analyze the roles of actors in networks, many studies in social science have relied on centrality measures [6]. Centrality measures help to understand the individual actor's prominence according to the actor's position in the network [10]. We argue that centrality measures are suitable means to analyze the directed flow networks.

3 Improving the FLOW Method

This section gives an overview of our concept to enrich FLOW diagrams with social network analysis by calculating key indicators.

3.1 Selection of Appropriate Centrality Measures

This approach concentrates on the use of centrality measures which help to analyze FLOW networks to simplify the process. Nevertheless, the FLOW analysis cannot be fully automated, since a FLOW network will not include all information – in particular, interpersonal aspects will be hard to visualize. However, there are some centrality measures which help to identify critical points and support the individual analysis. We will focus on embedding five centrality measures in FLOW analysis.

Degree centrality is a local measure counting the number of edges of a node, i.e. the sum of sources (in-degree) and receivers (out-degree) of information. In directed networks, in- and out-degree can be considered as two different values to differentiate between the quantities of incoming and outgoing information [11]. Applied to FLOW analysis, degree centrality measures the frequency and the amount of incoming and outgoing information. Degree centrality is a first indicator for a person being important for information sharing. It identifies central persons who need to share or receive much information. A drop out of one of these central persons can threaten project success since all information passing this person would either be lost or delivered in a longer time.

Closeness centrality measures the average distance of a node to each other node within the network [11]. It can be seen as a measure for the well-positioning of a person within a network [10]. This measure assumes that information originates from all other persons with equal probability and that all information flows along the shortest paths [10]. A low raw closeness score is an indicator for a person being well-positioned to obtain novel information early [10]. Nodes with a short distance to other nodes tend to receive flows sooner, assuming comparable times of information transport between all nodes in the network. Thus, closeness centrality is normally interpreted as the time until arrival of something flowing through a network [12]. Persons with high closeness scores have short distances and hence a high collaboration [13] with other persons and will receive information sooner. In FLOW analysis, these persons are very important for information sharing. Identifying these persons is one of the aims of the FLOW method, since they obtain novel information early. In addition, they bundle much information that can be shared within the network. In some cases, there are tasks referencing to several processes. Furthermore, they bundle information of many processes and hence can transport them to responsive persons. Replacing such persons by farseeing persons with less closeness centrality would go along with a drop in performance.

Betweenness centrality measures how often a node is located on the shortest (i.e. geodesic) paths between two other nodes. Persons are central if they have the potential to mediate the flow of resources or information between other actors [14]. Thus, persons who can mediate the flow of information between other actors are considered as central. This measure identifies persons who are indispensable for information sharing between other persons. Furthermore, regular exits due to holidays, retirement, or termination, can cause problems since it is difficult to replace the respective persons adequately. Therefore, knowing these persons is desirable.

Flow betweenness extends betweenness centrality in two ways: It considers all paths between nodes and not only geodesics, and it is appropriate for graphs and weighted graphs in which larger weights indicate stronger ties between actors. Edge weights are

taken to represent the potential for the flow of information between nodes assuming that the constraints for a metric in a mathematical view (except for the symmetry in directed graphs) hold [15]. Applied to FLOW analysis, flow betweenness describes the amount of information, forwarded by a single person by considering all information transported via this specific person. These persons are interesting because they coordinate the process-independent information flow.

As an extension of degree centrality, **eigenvector centrality** defines actors to be central if they have ties to other actors who, in turn, are central themselves [14]. In FLOW analysis, eigenvector centrality identifies persons who are important for information flow since they probably get many information in a short time. Furthermore, they can share their information with other important persons with little effort. These persons themselves can share the information so that the information can spread in a short time.

Each of these centrality measures is an indicator for problems concerning the information flow within a team. Hence, we use them to help the FLOW analyst identify critical points in FLOW networks. Having determined these measures, the FLOW analyst can interpret the network and help the team ensure the right amount of information flow and unburden persons who must not fail.

3.2 Transforming Flow Diagrams to Networks

In order to apply centrality measures to a FLOW diagram, the diagram has to be transformed into a valid network first. The ongoing of this transformation is not trivial, since FLOW diagrams highly resemble networks but differ in some significant aspects. To clarify this, we provide a short introduction to structure of FLOW diagrams and their elements below.

A FLOW diagram consists of information stores, information flows, and FLOW activities.¹ Having a nested structure, a FLOW activity can contain information stores, information flows, and other FLOW activities. To convert the FLOW diagram into a network, information stores and information flows can be mapped directly to nodes and edges of these networks but nested structures need to be dissolved.

As illustrated in Fig. 2, there are three different ways of transforming a FLOW activity. The left side of the figure shows an example of a FLOW diagram containing an activity with a nested structure to be transformed. On the right side of the arrows are all possible alternatives to transform the nested structure of the activity:

1. Directly connecting all incoming and outgoing stores of the activity.
2. Representing the activity through a distinct node connected to all incoming and outgoing stores.
3. Specifying the underlying information flow structure by deciding for each incoming and contained store to which contained or outgoing store it has to be connected. In this case, “A” is defined as an incoming store, “X” and “Y” are defined as contained stores and “B” is defined as an outgoing store.

¹ Refer to [16] for a detailed description of the FLOW syntax.

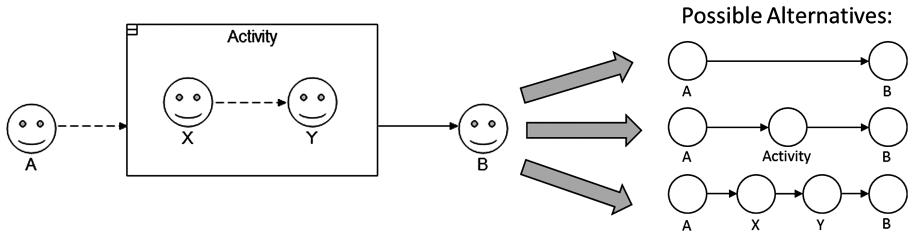


Fig. 2. Three alternatives to represent the same FLOW activity (left) in a network (right)

To transform FLOW activities properly, we have to clarify the purpose of the actual activity first. In the first alternative, store “A” provides information for the activity, which might or might not be processed and changed in the activity. It is obtained by neither “X” nor “Y”. “B” receives the possibly changed information. This signifies that an internal way of processing this information is not considered to be relevant for the overall process.

In the second alternative, the activity represents an abstract task or sub-process, which can neither be transformed nor removed. This may occur, if a FLOW diagram contains uncertain information, which could not be elicited during FLOW interviews. For example, “A” and “B” do not know how information is being passed in the activity and “X” and “Y” could not be identified and are thereby obscured for the FLOW analyst.

The third alternative represents a fully unrestricted way of interconnecting incoming and outgoing stores. This case should be used in all cases except alternative 1 and 2. This case can occur if the activity contains a sub-process of the overall modelled process. In this example, “A” provides information for “X”, while “Y” provides information for “B”.

Only the analyst can decide which alternative of these three above he should apply. He has to decide for each FLOW activity which case is appropriate and which incoming and outgoing information stores have to be selected.

3.3 Calculating Key Indicators

We developed a converter to transform FLOW diagrams from xml files created with our FLOW editor ProFLOW into networks. The transformed network now has to be analyzed by applying centrality measures. For this task, we use UCINET² to calculate key indicators based on our centrality measures depicted above. The converted FLOW networks can be imported to UCINET, which calculates the key indicators. This process cannot be done automatically yet because user interaction with UCINET is necessary. After calculation, the determined key indicators are exported by the analyst for further analysis of the FLOW network.

² UCINET is a common used tool for social network analysis.

4 Discussion

In this section we point out the limitations of our method followed by an interpretation of our findings.

4.1 Limitations and Interpretation

Applying information flow analysis in a software company does not imply to remedy all company's problems, but can improve information flows. Furthermore, findings uncovered through information flow analysis must be accepted by the company and implemented by the employees.

The benefit of information flow analysis depends on the chosen flow analysis method. In our case, we have chosen FLOW to analyze information flows which suited well for the analyses we conducted so far. Applying the FLOW method takes a lot of time and is highly based on personal experience. The approach requires a considerable number of manual steps. One reason is that these steps require a decision from the FLOW analyst that cannot be automated (see Fig. 2). Another reason is the interface between UCINET and our ProFLOW editor. UCINET does not allow automated processes. It is likely that the analysis will become more time-consuming for the analyst but he will gain more insights.

A way to decrease manual steps would be to integrate the converter into the used FLOW editor and relieve the user to switch between applications and to choose the FLOW diagram to convert.

In this contribution we applied centrality measures which can enrich the original FLOW method. Considering the added information (as key indicators) can lead to results that have been overlooked before.

5 Conclusion

This paper describes concepts, methods and tools to improve the analysis of information flows by improving the FLOW method with social network analysis. The approach is to convert FLOW diagrams into networks and to calculate centrality measures for the converted diagrams. These centrality measures support the analysis of structure and meaning of the original FLOW diagram and enables an easier detection of inconsistencies and anomalies in information flows. This improves processes and procedures in the project as well as in the entire software company.

We accomplished our goal to reduce the influence of personal experience on the FLOW method as well as our purpose to uncover more findings in information flows in software companies.

Acknowledgements. This work was supported by the German Federal Ministry of Education and Research under grant number K3: FKZ 13N13548 (2015-2018) and by the German Research Foundation (DFG) under grant number 263807701 (Project TeamFLOW, 2015-2017).

References

1. Stapel, K.: *Informationsflusstheorie der Softwareentwicklung*, 1st edn., Univ, Hannover, Dr. Hut, München (2012)
2. Stapel, K., Knauss, E., Schneider, K.: Using FLOW to improve communication of requirements in globally distributed software projects. In: *Workshop on Collaboration and Intercultural Issues on Requirements: Communication, Understanding and Softskills (CIRCUS 2009)*, pp. 5–14 (2009)
3. Durugbo, C., Tiwari, A., Alcock, J.R.: Modelling information flow for organisations: a review of approaches and future challenges. *Int. J. Inf. Manag.* **33**(3), 597–610 (2013)
4. Merrill, J., et al.: Findings from an organizational network analysis to support local public health management. *J. Urban Health Bull. N.Y. Acad. Med.* **85**(4), 572–584 (2008)
5. Borgatti, S.P., Foster, P.C.: The network paradigm in organizational research: a review and typology. *J. Manag.* **29**(6), 991–1013 (2003)
6. Braha, D., Bar-Yam, Y.: Information flow structure in large-scale product development organizational networks. In: *Vervest, P., Heck, E., Pau, L.-F., Preiss, K. (eds.) Smart Business Networks*, 1st edn., pp. 105–125. Springer, Heidelberg (2005)
7. Friedkin, N.E.: Information flow through strong and weak ties in intraorganizational social networks. *Soc. Netw.* **3**(4), 273–285 (1982)
8. Ryyänänen, H., Jalkala, A., Salminen, R.T.: Supplier’s internal communication network during the project sales process. *Proj. Manag. J.* **44**(3), 5–20 (2013)
9. Wasserman, S., Faust, K.: *Social Network Analysis: Methods and Applications*, 1st edn. Cambridge University Press, Cambridge (1994)
10. Borgatti, S.P.: Centrality and network flow. *Soc. Netw.* **27**(1), 55–71 (2005)
11. Freeman, L.C., Roeder, D., Mulholland, R.R.: Centrality in social networks: II. Experimental results. *Soc. Netw.* **2**(2), 119–141 (1979)
12. Borgatti, S.P.: Centrality and AIDS. *Connections* **18**(1), 112–114 (1995)
13. Schneider, K., Liskin, O.: Exploring FLOW distance in project communication. In: *2015 IEEE/ACM 8th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)*, pp. 117–118 (2015)
14. Faust, K.: Centrality in affiliation networks. *Soc. Netw.* **19**(2), 157–191 (1997)
15. Newman, M.J.: A measure of betweenness centrality based on random walks. *Soc. Netw.* **27**(1), 39–54 (2005)
16. Stapel, K., Schneider, K., Lübke, D., Flohr, T.: Improving an industrial reference process by information flow analysis: a case study. In: *Münc, J., Abrahamsson, P. (eds.) PROFES 2007. LNCS*, vol. 4589, pp. 147–159. Springer, Heidelberg (2007). doi:[10.1007/978-3-540-73460-4_15](https://doi.org/10.1007/978-3-540-73460-4_15)