A Social Information Flow Graph: Design and Prototypical Implementation

Thomas Reschenhofer, Patrick Bürgin, and Florian Matthes

Technical University of Munich, Munich, Germany,
reschenh@in.tum.de, patrick.buergin@tum.de, matthes@in.tum.de

Abstract. Companies are increasingly turning to Enterprise 2.0 technologies to harness collective intelligence for information management and analysis. In this context, social network analysis (SNA) provides insights into which knowledge workers interact with which information assets and thus improves cooperation among them. However, current approaches to SNA disregard semantic relationships between the information assets or the relevance of the applicability of SNA by end-users. In the paper at hand we address this issue by proposing the social information flow graph (SIFG) as a tool for end-user-oriented SNA. In this sense, the SIFG provides a holistic and social perspective on relationships between individuals and information assets within an Enterprise 2.0 environment. We showcase its technical applicability by a prototypical implementation. By conducting case studies in three different application domains, we show that the SIFG and its explorability enable novel opportunities and use cases for end-user-oriented SNA.

Keywords: Information Flow Graph, Social Network Analysis, Enterprise 2.0, Information Visualization

1 Introduction

Today's enterprises are facing not only the challenge of managing an increasing amount of digital information, but also to do this in an efficient way [5]. To tackle this challenge, enterprises apply collaborative technology to foster the contribution of multiple knowledge workers [3] in order to harness collective intelligence. By the collaborative creation and management of information and services, knowledge workers implicitly form a social network. Analyzing those interactions can reveal interesting knowledge (e.g., communication patterns) about an organization's social structure and dynamics [6]. However, making this knowledge accessible to the organization's users and thus to enable end-user-oriented social network analysis (SNA) is a big challenge [8].

On another note, data has to be processed and presented according to the specific needs of knowledge workers in order to support decision making based on it. Knowledge workers often have different requirements regarding the representation of the same information, e.g., they might be interested in different levels of detail. End-user-oriented business intelligence (BI) tools enable them to process and represent a given information according to their needs without having to rely on IT specialists. Thereby, knowledge workers are able to define custom data transformations and tailored data visualizations. However, if multiple knowledge workers define data transformations and visualizations based on the same information, it is challenging to keep track of which information is transformed and visualized in which way.

To address this issue, information flow diagrams make the information flow from its source through its transformations to visualizations transparent. While in a non-collaborative setting, individual knowledge workers define how information is processed and represented based on their individual needs and independently from co-workers, in a collaborative environment users not only cooperate in managing of information, but also in its processing and presentation, e.g., by sharing transformations or visualizations, or by defining them jointly. This inevitably leads to a more complex information flow structure, which makes it more difficult for users to understand connections between information assets.

In the present paper, we propose a concept for the utilization of information flow graphs for end-user-oriented SNA. In this sense, a social information flow graph (SIFG) not only visualizes how the information flows through a software system, but also which users are interacting with it. Therefore, it provides a social and holistic perspective on information assets and their relations. We also address the challenge of making the SIFG explorable by end-users to ensure its practical applicability. Finally, we identify concrete concerns in different application domains which can be addressed by a SIFG.

2 Conceptual Design of a Social Information Flow Graph

As described by Chi and Riedl's [2] *Information Visualization Data State Reference Model*, data is transformed to analytical abstractions, and subsequently bound to visualization abstractions. These visualization abstractions are the basis for the final view which is consumable by users. In order to keep the SIFG model simple and applicable, we subsume Chi and Riedl's concepts of *Data* and *Analytical Abstraction* under *Data*, and their notion of *Visualization Abstraction* and *View* under *View*. Furthermore, we explicitly model Chi and Riedl's *Data Transformation* and *Visualization Transformation*, while we name the latter one *Data Binding* in accordance to the model-based approach of Hauder et al. [7]. A *Data Transformation* derives data from potentially multiple data elements, while a *Data Binding* connects potentially multiple data elements to visualizations.

The concepts *Data*, *View*, and *Transformation* are considered as *InformationAsset*s as defined by Khatri and Brown [12], i.e., they are documented facts with a potential value for a stakeholder. Therefore, we introduce a corresponding class as super type for *Data*, *View*, and *Transformation*. Each *InformationAsset* has *MetaData* representing descriptive information about its structure, context, or semantics [4]. This also includes different kinds of relations between *InformationAsset*s and *User*s, e.g., who is responsible for a certain *InformationAsset* [12],

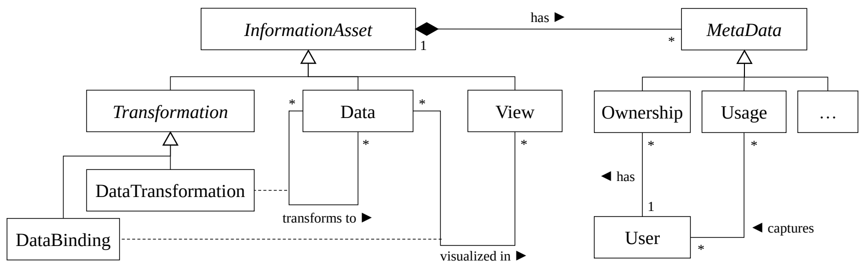

Fig. 1. The conceptual model for a holistic social information flow graph integrating concepts of visualization models [2, 7] and meta data management [4, 11, 12, 16].

who uses it [16], and who creates it [11]. The concrete meta data types depend on the actual use case, and can be added to the model as indicated in Figure 1.

The SIFG enables the derivation of new social relationships by considering interactions with different information assets which are connected to each other through a chain of transformations. For example, the owner of a particular data object can be associated with a user consuming the same data object through a corresponding visualization. As a consequence, the proposed model opens new opportunities for improving organizational social network analysis [13].

3 A Prototypical Implementation

The prototype is built upon a model-based Enterprise 2.0 platform called Hybrid Wikis [14]. In this platform, the initially unstructured data entities can be structured incrementally and collaboratively by enabling end-users to dynamically define a data model consisting of entity types, attributes, and relationships. On the top of this data modeling concept, the Hybrid Wiki platform integrates the model-based expression language (MxL) for the user-driven definition of queries [15] as well as configurable dashboards consisting of potentially multiple visualizations. MxL's static type-safety enables the automatic detection ofo semantic dependencies between expressions and elements they are referring to. Those dependencies are the input for the automatic generation of an information flow graph. For example, a query defined by an MxL expression and referring to a data entity implies a dependency from the query to the data entity, while the information flows into the opposite direction.

Regarding the conceptual model as described in the previous section, the Hybrid Wiki entities as *Data*, dashboards and visualizations as *View*s, and MxL expressions as both *DataTransformation*s and *DataBinding*s form the system's *InformationAsset*s. Each of those information assets is definable and editable by users, and possesses a specific set of meta data attributes. For example, each information asset has a version history capturing its temporal evolution, or authorization rules controlling the access to it in a collaborative environment. These meta data attributes—and particularly the user-related meta information—can

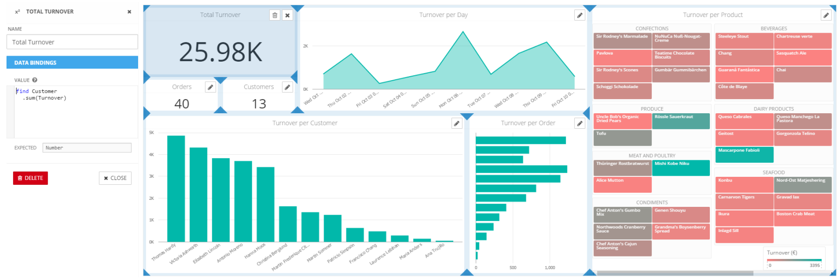

Fig. 2. An exemplary dashboard consisting of different configurable visualizations.

be used by the prototype to enrich the aforementioned information flow graph and to generate the SIFG. Therefore, the prototype represents a full implementation of the conceptual model as presented in Section 2.

In the following, we illustrate the prototype by implementing a simple sales scenario which involves customers, orders, products, and categories. A customer's turnover is defined as derived attribute, i.e., the value is calculated automatically according to a given computation prescription. Derived attributes represent data transformations as defined in Section 2. In the prototype, a customer's turnover can be defined by the MxL expression *Orders.sum(Product.Price)*. Based on this, a query computing the total turnover of all customers can be defined as *find Customer.sum(Turnover)*. The *find* construct retrieves all customers. Based on this collection as well as a lambda expression referring to the turnover of each customer, the *sum* function computes the total turnover. In this example, there is a semantic dependency from the query to the derived attribute turnover, which in turn depends on the order's product and the respective product's price. Therefore, there is an information flow from the product's price through the derived attribute turnover to the total turnover query.

Figure 2 shows a dashboard illustrating, e.g., the temporal evolution of the daily turnover within a line chart, or a cluster map organizing the products within their categories. Each visualization is configured separately, i.e., for each visualization users have to define a corresponding data binding by using MxL expressions. Based on the data model, data transformations (e.g., derived attributes), as well as visualizations and their data bindings, the prototype automatically generates the SIFG in real-time. As shown in Figure 3, the information flow's starting points are the Hybrid Wiki system's entity types (e.g., *Order* and *Customer*) which are organized in so-called workspaces (the dark-blue boxes on the left, e.g., *Northwind*). The internal structure of the types (e.g., attributes and derived attribute) is not shown, since this would add to much complexity to the view. Visualizations are illustrated by respective icons and composed to dashboards (the red boxes on the right, e.g., *Northwind*). In between, there are functions representing generic data transformations, e.g., *turnoverPerDay*.

The prototype provides UI features inspired by Heer and Boyd [8] to improve the usability of the SIFG for end-users. When clicking on a node, the prototype

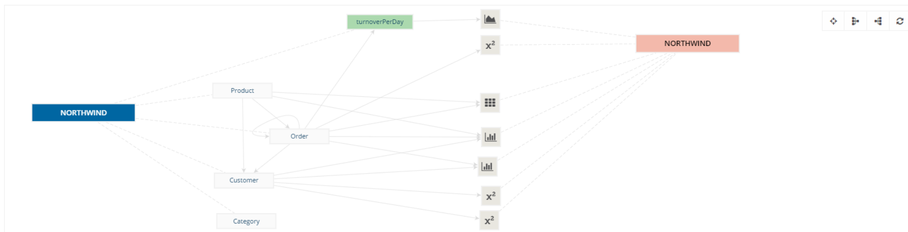

Fig. 3. The social information flow graph (SIFG) for the dashboard in Figure 2.

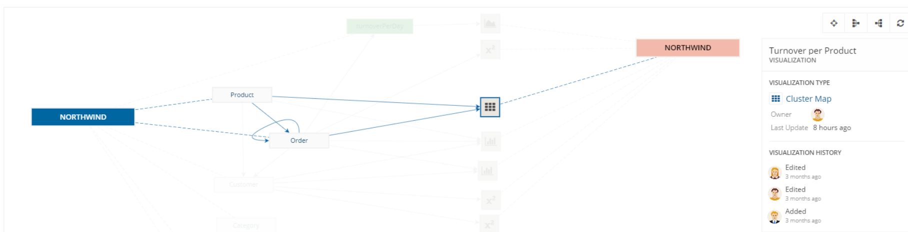

Fig. 4. Selecting the nodes of the SIFG allows users to explore their meta information.

shows the meta information for the respective information asset on the right-hand side of the graph view. For example, in Figure 4 the meta information for a visualization is shown, including user-related meta information like its ownership and version history. Depending on the actual use case, the set of displayed meta information can be extended accordingly, e.g., by information on the usage of the selected information asset. Furthermore, clicking on a particular node highlights all (transitive) predecessors as well as (transitive) successors. In this way, a user can identify all information sources which the selected node depends on, and all visualizations for which the selected node is an input. This enables users, e.g., to find the contact person for the input of a particular visualization.

To abstract and hide certain parts of the graph view, users can collapse workspaces or dashboards. This can reduce the complexity of the graph view and increase its understandability. As shown in Figure 5, collapsing a dashboard hides all its visualizations. All information flows which initially targeted those visualizations are now heading to the dashboard instead.

4 Case Study

We conducted a case study including three cases in different application domains and different German companies. In each of those cases, knowledge workers already use a BI tool for analyzing and visualizing domain-specific data in an end-user-driven way. The study shows concrete concerns those knowledge workers can address when using the SIFG prototype. To this end, we asked them for their opinion on the SIFG and for related concerns which they face in their

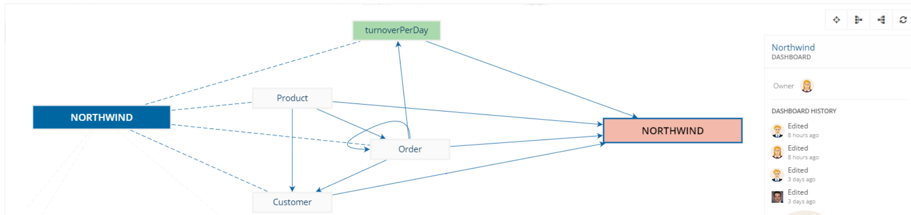

Fig. 5. Nodes can be collapsed to hide certain parts of the information flow network.

daily work, and which could be resolved by our approach. The first interviewee is an enterprise architect who currently manages the internal architecture of a German IT services provider (5,000 - 10,000 employees), while the second one is an IT infrastructure manager of a subsidiary of a German investment company (10,001+ employees). The third interviewee is a data quality manager of a German company (10,001+ employees). For the sake of discussion, we exemplarily implemented a representative dashboard for each of the observed cases.

The interviews revealed the following concerns which are addressable by the demonstrated prototype in particular, and the SIFG in general:

Usage Analysis: As stated by the interviewees, "the collection of data induces the highest cost in our daily business". By using the SIFG, data model maintenance process can be optimized by stopping to maintain unused data.

Stakeholder Identification: The interviewees stated that the SIFG could be used to identify stakeholders which are related to certain information assets, enabling them to proactively contact the users which express their interest in a given dashboard.

Impact Analysis: According to the interviewees, users tend to avoid changes of certain information assets, since they are afraid to "break something". Faced with the SIFG, the interviewees expressed that "this would certainly help to make the models and relations more transparent".

Support for Data Provenance: As stated by the interviewee, sharing data among colleagues is "part of the daily business". Thus it is "interesting to know what the story behind a given information asset is", raising questions such as "Where does the data come from? Who worked on it? When was the last change and by whom?". Providing an environment to visually explore the SIFG may reduce the "significant operative cost" induced by struggles when identifying co-workers that faced similar problems in the past.

Addressing Compliance Demands: Driven by compliance demands, certain organizations (particularly in the financial services domain) have to be able to reveal the full calculation process behind a given measure or visualization. The SIFG may help to meet these demands.

Support for Data Consolidation: The interviewees stated that the SIFG is helpful to initiate consolidations of dashboards and views based on overlapping information, i.e., information which is visualized in different dashboards.

5 Related Work

Hüner et al. [11] discuss requirements of a collaborative MDM repository, and highlight the relevance of collaboration for effective MDM. However, they do not elaborate on how the analysis of a information flow graph or social networks could improve the repository's support for MDM processes. For example, they propose a process for the maintenance of business meta data including activities like "Identify responsibility", which can be supported by the SIFG.

On another note, Dinter et al. [4] investigate the opportunities of MDM for different kinds of stakeholders. Thereby, they identify benefits like "Impact analysis" and "Data lineage/provenance", which match with the identified concerns as described in Section 4. Their study shows that the potential benefit of particularly those three mentioned benefits is exceptionally high—especially for BI users—while the level of actual implementation is still low.

Hermans et al. [9] discuss data flow diagrams for spreadsheets. They present a tool for automatically creating them based on cell references. By evaluating the use of data flow diagrams for the analysis of spreadsheets by conducting a survey, they found out that a holistic perspective on the information flow within a spreadsheet is helpful for 80 % of survey participants. However, they also state that the level of detail and understandability of data flow diagrams has to be well-balanced in order to provide helpful insights into the spreadsheet dynamics.

In addition to researchers, software vendors also try to utilize the dependencies and relationships between both information assets and users in order to expose new knowledge [10]. For example, Microsoft's Office Delve allows users to analyze information and activities within the Office 365 Suite.

6 Conclusion

This work presents a concept and a prototypical implementation for a SIFG, which not only makes semantic dependencies between information assets visible and explorable for end-users, but also integrates relations to users of the system. In this way, the SIFG supports the identification of new social relationships between users, and thus fosters collaboration among knowledge workers. Its conceptual model as described in Section 2, the UI features described in Section 3, and the interviews as described in Section 4 represent the answers to the research questions raised in the present paper's introduction.

With respect to the validity of the exploratory case study, three cases form a small foundation for both an evaluation of the SIFG as well as the identification of concrete concerns addressable by it. However, the interviews represent a preliminary study for a more extensive study on concerns in one particular domain. While the interviewees agreed on the high potential of the SIFG, they also outlined ethical and privacy issues [1]. All interviewees mentioned that their company's work councils "would have worries if users can see all the activities of their co-workers". However, the interviewees also stated that a proper balance between transparency and privacy would ensure a reasonable practicability while also taking into account ethical and legal concerns.

Based on the results of the paper in hand, our future research activities will focus on the improvement of the SIFG with respect to its application in a specific domain, namely EAM. Therefore, we aim for an extension of this concept and the prototype in order to address the specific concerns as identified in the interview with an enterprise architect. Furthermore, we plan to do an extensive evaluation within a research community including more than 20 enterprises.

References

1. Borgatti, S.P., Molina, J.L.: Ethical and Strategic Issues in Organizational Social Network Analysis. Journal of Applied Behavioral Science 39(3), 337–349 (2003)
2. Chi, E.H.h., Riedl, J.: An Operator Interaction Framework for Visualization Systems. Proceedings of the IEEE Symposium on Information Visualization (1998)
3. Davenport, T.H.: Thinking for a Living: How to Get Better Performances and Results from Knowledge Workers. Harvard Business Press (2013)
4. Dinter, B., Gluchowski, P., Schieder, C.: A Stakeholder Lens on Metadata Management in Business Intelligence and Big Data–Results of an Empirical Investigation. Proceedings of the Americas Conference on Information Systems (2015)
5. Edmunds, A., Morris, A.: The Problem of Information Overload in Business Organisations: A Review of the Literature. International journal of information management 20(1), 17–28 (2000)
6. Freeman, L.: The Development of Social Network Analysis. Createspace, Vancouver (2004)
7. Hauder, M., Matthes, F., Roth, S., Schulz, C.: Generating Dynamic Cross–Organizational Process Visualizations through Abstract View Model Pattern Matching. Proceedings of the International Conference on Interoperability for Enterprises Systems and Applications pp. 95–101 (2012)
8. Heer, J., Boyd, D.: Vizster: Visualizing Online Social Networks. IEEE Symposium on Information Visualization (INFOVIS) (2005)
9. Hermans, F., Pinzger, M., van Deursen, A.: Supporting Professional Spreadsheet Users by Generating Leveled Dataflow Diagrams. Proceedings of the International Conference on Software Engineering 2011 (2011)
10. Hoffmann, L.: Mine Your Business. Communications of the ACM 53(6), 18–19 (2010)
11. Hüner, K.M., Otto, B., Österle, H.: Collaborative Management of Business Metadata. International journal of information management 31(4), 366–373 (2011)
12. Khatri, V., Brown, C.V.: Designing Data Governance. Communications of the ACM 53(1), 148–152 (2010)
13. Kilduff, M., Brass, D.J.: Organizational Social Network Research: Core Ideas and Key Debates. The Academy of Management Annals 4(1), 317–357 (2010)
14. Matthes, F., Neubert, C., Steinhoff, A.: Hybrid Wikis: Empowering Users to Collaboratively Structure Information. Proceedings of the International Conference on Software and Data Technologies pp. 250–259 (2011)
15. Reschenhofer, T., Monahov, I., Matthes, F.: Type-Safety in EA Model Analysis. Proceedings of the International Enterprise Distributed Object Computing Conference Workshops and Demonstrations pp. 87–94 (2014)
16. Vaduva, A., Vetterli, T.: Metadata Management for Data Warehousing: An Overview. Journal of Cooperative Information Systems 10(3), 273–298 (2001)