THE SNARE ARCHITECTURE OVERVIEW
Social Network Analysis and Reengineering Environment

Alexandre Barão and Alberto Rodrigues da Silva
Instituto Superior Técnico/INESC-ID, R.Alves Redol,9,1000-029 Lisboa, Portugal
alexandrebarao@yahoo.com, alberto.silva@acm.org

Keywords: Social network analysis, modeling, social network reengineering, transformations, ETL.

Abstract: Social network systems identify existing relations between social entities and provide a set of automatic inferences on these relations, promoting better interactions and collaborations between these entities. However, we find that most of existing organizational information systems do not provide in a native way, social network features, even though they have to manage somehow social entities. The focus on this paper starts from this fact, and proposes the SNARE system, short for "Social Network Analysis and Reengineering Environment". The SNARE's purpose is to promote social network capabilities in information systems not designed originally with that purpose. The paper overviews the architecture of SNARE system and discusses its applicability through different approaches.

1 INTRODUCTION

Social network analysis (SNA) is focused on describe and analyze relations between social entities, such as people or organizations (Wasserman and Faust, 1994). In SNA contexts, dynamics of groups are studied to identify relations and interactions among their members. Through SNA it's possible to uncover people interaction patterns (Freeman, 2007).

SNA presents the following benefits: (1) Improve information sharing by analyzing network relations; (2) Increase efficiency by identifying group or individual performances and bringing up new roles, or providing information to redistribute roles among groups or individuals; (3) Support diagnostic approaches by preventing bottlenecks in networks, e.g. evaluating periodically the weight of information flows; and (4) Supply measures to evaluate the impact of network changes. These benefits are crucial for decision-making and consequently can promote innovation and productivity in competitive organizations.

SNA traditional studies use much information residing in archives that were not created expressly for social research. Sometimes, such data provide measures of social ties and trace relationships of actors who are reluctant to interviews. Archival data are often inexpensive, especially when in electronic form if maintained over time, these archives can support SNA studies.

The validity of archival data rests on the correspondence between measured connections and the conceptual ties of research interest (Carrington, P. et al., 2005). Most social networks are domain specific, or, on the other hand, too much generic to hold relevant data. Thus, to solve specific problems, it is necessary to develop integration mechanisms to allow inferences on new social networks with different relations and topologies.

Our project, "Social Network Analysis and Reengineering Environment", SNARE system has the purpose to provide social network features in information systems not designed originally for the effect. Through the implementation of SNARE Social Network Metamodel, it is possible to do real time social analysis, i.e. analyze actors continuously and multiple relations among these actors.

Information does not flow unchanged through a human network. People add context, interpretation, and meaning as they receive information and pass it along (Cross and Parker, 2004). To minimize this fact, we assume that are actions associated to specific relations and SNARE provides an interface for real time actions registration. This fact enables automatic online social analysis between actors. We also assume that an information flow can be a special kind of action triggered by actors.

The SNARE metamodel allow researchers to dynamically construct real scenarios for SNA extraction and relational knowledge discovery.

We are currently considering applying SNARE into educational contexts. To introduce SNA features into these information systems, it is necessary to introduce ETL (acronym for Extract, Transform, and Load), a process that involves extracting data from outside sources, transforming it to fit business needs, and ultimately loading it into the data warehouse.

Social networks can be represented according different notations such as graphs, matrices, or algebraic notations (Wasserman and Faust, 1994). The data comprising social networks tend to be heterogeneous, multirelational, and semi-structured. Link mining is a confluence of research in social networks, link analysis, hypertext and Web mining, graph mining, relational learning, and inductive logic programming (Han and Kamber, 2006).

In this paper, we introduce the motivation and the context for social networks. Section 2 overviews the SNARE architecture. Section 3 discusses application scenarios of the SNARE framework. Finally, section 4 presents preliminary conclusions of the investigation.

2 SNARE ARCHITECTURE

The aim of the SNARE system is to extract and analyze social networks starting from information systems not designed originally with SNA features. This has several benefits, e.g. reduce survey dependent procedures, maximize reutilization of archival data, and provide a mechanism for ongoing social network analysis.

2.1 Architecture Approaches

SNARE provides mechanisms to extract data sets from other systems. SNARE is designed to extract data from a generic system by mapping data extraction components.

There are two approaches in SNARE architecture corresponding to two boundaries: (1) *Transparent,* to support SNA of an information system without the need to change it; and (2) *Intrusive,* to support SNA depending on full access to a generic information system source code with a need to change it.

As suggested in Figure 1, *Transparent* approach has two components: Information System (IS) and SNARE ETL Services. The IS denotes a generic information target system. SNARE ETL Services is an engine that provides ETL features, i.e. extract data from outside sources, transform it, and load it into the SNARE database.

On the other hand, *Intrusive* approach includes the Extended Information System (IS*) which is an IS refinement and the SNARE Services system, which provides a set of social network features through web services.

Besides these two architecture approaches, SNARE Services can be executed autonomously, i.e. through SNARE user interface applications, it is possible to perform social network analysis tasks without the need of external information systems integration.

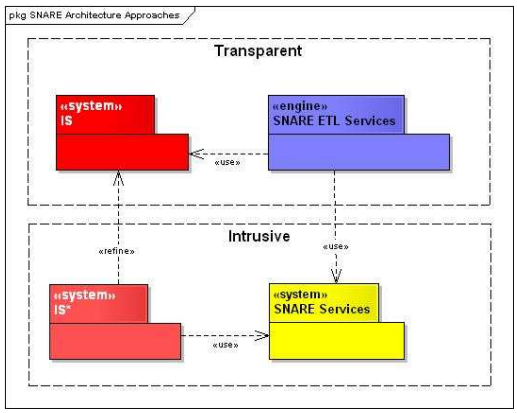

Figure 1: SNARE Architecture Approaches.

2.2 System Artefacts

SNARE is an application under development to be web accessed over a network such as the Internet or an intranet.

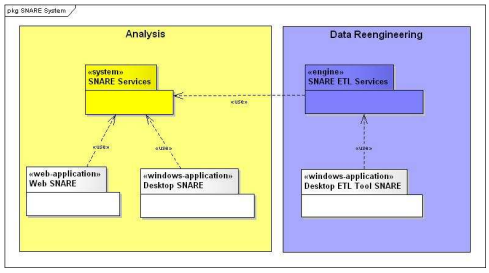

Figure 2: SNARE system packages.

As depicted in Figure 2, SNARE system involves a set of packages. Figure 2 also emphasizes the social network analysis context and the data reengineering context. Each package has a specific function, making possible to distinguish several stereotypes in SNARE components such as: web application, windows application, system, and

finally, an engine. Figure 3 shows the SNARE system components view.

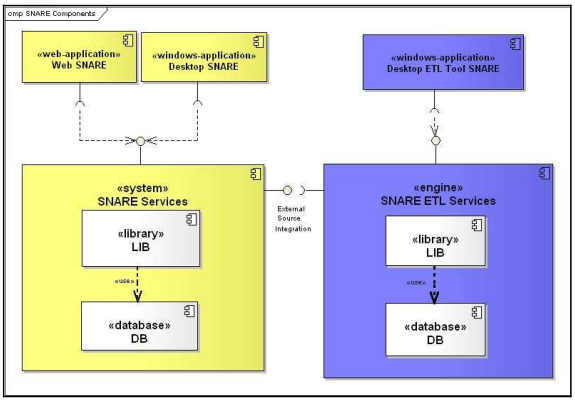

Figure 3: SNARE system components view.

2.2.1 Analysis Components

Based on social network models, SNARE Services provides methods to ensure the definition of a set of models to allow the representation of social networks.

Web SNARE is a web-based application while Desktop SNARE is the equivalent tool for a windows-based application. Both include four modules: Social Network Environment Settings; Social Network Definitions; Social Network Instantiation; and Social Network Explorer. Through Social Network Environment Settings, users can customize environment settings that automate data collection for SNA. Each user has an identity. Authorization and security policy is defined according to a flexible role-based mechanism. The Social Network Definitions Module is used to create social entities, roles, relations, actions, events, i.e. structural SNA data. The Social Network Instantiation Module make possible to instantiate relations, actions, events and surveys. Surveys allow investigators to decide on relationships to measure and on actors/objects to be approached for data and in the absence of archival records, surveys are often the most practical alternative (Carrington et al., 2005). SNARE survey instantiation interface mechanisms are useful for entering and visualizing personal or group social network data. Finally, the SNARE Social Network Module Explorer is the SNARE interface for social network data analysis requests and visualization.

SNARE Services include a library and a database as showed in Figure 3. SNA algorithms and model transformations are provided by SNARE Library, i.e. a set of transformations, allowing conversions between models and algorithms which may analyze the social networks to retrieve relevant information. Components interact with each other through provided and required interfaces. SNARE Services provide two main (classes set) of interfaces: (1) interfaces to support web and windows-based applications; and (2) interfaces to support external source integration. SNARE Services has a built-in XML parser for SNARE Schema, which is a schema to handle social networks relational data including relations, actions, events, social entities and roles, through provided interfaces (see section 3 for more details).

Developing models to represent social networks, allows SNARE Services implementation of specific graph algorithms which can extract valuable information from hidden social networks. Algorithms from Link Analysis, an area pertaining to the more general research field of data mining with the purpose of extracting new information on linked structures such as graphs, can be also applied on social networks . These include classification algorithms, which predict the value of nodes based on their links, and clustering algorithms, which attempt to group nodes of a same kind and are of special interest in finding communities on linked structures.

2.2.2 Data Reengineering Components

On the other hand, SNARE ETL Services provide a technical interface to Desktop ETL Tool, a desktop application to define and control ETL actions and has a required interface to execute SNARE Services methods. The aim of this component is to extract relevant social network data through ETL mechanisms. This tool allows users to specify transforms through a graphical user interface.

Data transformation can involve the following: (1) Smoothing, which works to remove noise from data; (2) Aggregation, where summary operations are applied to the data, typically used in constructing a data cube for analysis of the data at multiple granularities; (3) Generalization of the data, where low-level data are replaced by higher-level concepts through the use of concept hierarchies; (4) Normalization, where the attribute data are scaled so as to fall within a small specified range; and (5) Attribute construction, where new attributes are constructed and added from the given set of attributes to help the mining process (Han and Kamber, 2006).

Transparent approach use SNARE ETL Services. These services are encapsulated in the SNARE ETL library component. SNARE ETL

Services use the interface to external source integration provided by SNARE Services and SNARE metamodel is used to define this integration.

3 APPLICATION SCENARIOS

From our experience using and developing e-Learning based systems, we identify several system limitations, such as: difficulty to recognize explicit relations between learning objects, authors, students and teachers, or between teachers and students, or even between students interested in a common subject. The challenge of our research is to make explicit these relations, supported by a generic platform (such as SNARE). In order to enhance learning processes, it is possible to improve actor's relationships through social networks inferences on systems that were not design for that purpose.

For an e-learning scenario, we are considering Moodle platform ("Moodle", 2007), which is, under a constructivist perspective, an internet-based course management system designed using pedagogical principles, to help educators create online learning communities. Our project scenario project uses Moodle platform where users can freely sign-up and create educational contents using learning objects. In this scenario, we should apply the SNARE *Transparent* approach because we do not intend to change the Moodle source code.

To infer social networks on other systems, another scenario would be considered. This scenario is a Learning Management System to support school management activities, involving different actors, e.g. students, teachers, educators or parents. In this scenario we will apply the SNARE *Intrusive* approach.

4 CONCLUSIONS

This paper introduces the problems and motivation behind our research work and overviews the proposed SNARE system.

Throughout this work, we propose a set of components to analyze social networks from real application scenarios. The main purpose of SNARE is to analyze social networks on systems not previously designed for the effect.

Based on social network models, SNARE Services provides methods to ensure the definition of a set of models to allow the representation of social networks.

With SNARE ETL Services, the system provides ETL features in order to analyze databases with the specific purpose to extract, transform and load data to SNARE database. This fact allows the use of specific graph algorithms which can extract valuable information from hidden social networks.

SNARE metamodel ensures that relations, actions and events can have multiple extreme instances and the social network system keeps references to all previous metaclasses.

From the research preliminary discussed in this paper, we conclude that much work on the area of social network analysis is still open, and that this area has a growing potential that should be explored. As a consequence of this project, we hope to provide new approaches and technologies to improve the organizational environment and, in particular, to improve e-learning and scholar management systems user interactions, to maximize educational success.

Finally, it would be of interest the development of new systems, taking advantage of the proposed SNARE application in other organizational contexts.

REFERENCES

Carrington, P.J., J. Scott, and S. Wasserman, 2005. *Models and methods in social network analysis*. Structural analysis in the social sciences ; 27. Cambridge ; New York: Cambridge University Press. xiv, 328 p.

Cross, R.L. and A. Parker, 2004. *The hidden power of social networks : understanding how work really gets done in organizations*. Boston, Mass.: Harvard Business School Press. xiii, 213 p.

Freeman, Linton C., 2007. *What is Network Analysis.* International Network for Social Network Analysis. Accessed in November 2007. Available at: http://www.insna.org/INSNA/na_inf.html

Han, J. and M. Kamber, 2006. *Data mining : concepts and techniques*. 2nd ed. Morgan Kaufmann series in data management systems. Boston: Morgan Kaufmann. xxviii, 770 p.

Moodle. 2007. http://www.moodle.org

Wasserman Stanley and Faust, Katherine, 1994. *Social Network Analysis: methods and applications.* In: Structural analysis in social the social sciences series. Cambridge: Cambridge University Press, (1994) 2006. ISBN 0-521-38707-8.