THE SNARE LANGUAGE OVERVIEW

Alexandre Barão and Alberto Rodrigues da Silva
IST/INESC-ID, R. Alves Redol, 9, 1000-029 Lisboa, Portugal

Keywords: Social network, Design, Modeling.

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, from scratch, 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 Language as the conceptual framework for 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 for the effect. Visual models are needed to infer and represent new or established patterns of relations. This paper overviews the SNARE language and shows its applicability through several models regarding the application of the SNARE to the LinkedIn real scenario.

1 INTRODUCTION

A *social network* consists of a finite set of actors and the relations defined among them (Wasserman and Faust, 1994). *Actors* are discrete individuals, corporate or collective social units, and are linked to one another by social *ties* (Wasserman and Faust, 1994). A *dyad* is a linkage or relation between two actors. *Triads* are triples of actors and associated ties. To a large extent, the power of network analysis lies in the ability to model the relations among systems of actors. A *subgroup* of actors is any subset of actors and all ties among them. A *group* is the collection of all actors on which ties are to be measured. The collection of ties of a specific kind of members of a group is called a *relation* (Wasserman and Faust, 1994). Actors may be referred as *social entities*.

An entity is social if involves a network of relations with other social entities (Masolo et al., 2004). A social entity play several roles in the same network. A role is a combination of particular sets of behavioral, meaningful, and structural attributes (Welser et al., 2007). The nature of roles and the way of representing them have been discussed in different fields, e.g. knowledge representation, knowledge engineering, object-oriented and conceptual modeling, multi-agent systems, linguistics, and cognitive semantics (Masolo et al., 2004). Four common features about social roles can be found: (1) roles are properties, e.g. different entities can play the same role; (2) roles are anti-rigid and they have dynamic properties, e.g. an entity can play different roles simultaneously, an entity can change role, an entity can play the same role several times, simultaneously, a role can be played by different entities simultaneously or at different times, the sequence in which roles may be acquired and relinquished can be subject to restrictions; (3) roles have a relational nature, i.e. roles imply patterns of relations; and (4) roles are linked to contexts, i.e. a contextual approach refer to a variety of factors, including relations, events, organizations and behaviors. The term "context" can have different interpretations, e.g. metaphysical context, cognitive context; and linguistic context. See (Masolo et al., 2004) for a further review.

There are different types of social networks. *One-mode networks* involve just a single set of social entities. *Two-mode networks* involve two sets of actors, or one set of actors and one set of events (Wasserman and Faust, 1994). *Events* have a time associated with them and it is possible for relations, positions and roles to change over time. In spite, events can occur at different times, the organizers of events change over time, and a different set of actors might participate in each event (Licamele et al., 2005). *Dyadic networks* and *affiliation networks* are

Barão A. and Rodrigues da Silva A. (2010).
THE SNARE LANGUAGE OVERVIEW.
In *Proceedings of the 12th International Conference on Enterprise Information Systems - Information Systems Analysis and Specification*, pages 344-349
DOI: 10.5220/0002975203440349

particular cases of two-mode networks. Another kind of network is the *ego-centered network* where a focal actor (termed “ego”) has a set of alters who have ties to ego, and measurements on the ties among these alters. It is possible to consider three or more mode networks, but rarely have social network methods been designed for such data structures (Wasserman and Faust, 1994).

In Social Network Analysis (SNA) scope, dynamics of groups are studied to identify relations and interactions among their members. Starting from these interactions it is possible to identify social patterns (Haythornthwaite, 2005) and it is possible to detect or propose social or organizational changes that reveal how networks grow or should change. Also, it is possible to find potential causes and consequences of a network change, previewing and controlling networks evolution (Churchill and Halverson, 2005). These features are dependent of metrics to allow group properties identification or to characterize individual influence on a specific group. Typically scenarios are strategic alliances and collaborations, flows of information (communication), affect (friendship), goods and services (workflow), and influence (advice) (Brass et al., 2004). Network research represents a different paradigm of research which requires new concepts and methods (Borgatti, 2003).

Traditional SNA 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 relations of social entities who are reluctant to interviews. Archival data are often inexpensive, especially when in electronic form. The validity of archival data rests on the correspondence between measured connections and the conceptual ties of research interest (Carrington et al., 2005). The data comprising social networks tend to be heterogeneous, multirelational, and semi-structured. Link mining is a relevant example showing 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).

New visual models are needed to infer and represent patterns of relations, and this paper proposes the SNARE language as the conceptual framework for SNARE system. The SNARE system purpose is to promote social network capabilities in information systems not designed originally for the efect.

In Section 1, we introduce social network concepts. Section 2 overviews social networks modeling techniques and the motivation for a social network language. Section 3 purposes the SNARE language. Finally, Section 4 presents preliminary conclusions of the investigation.

2 MODELING SOCIAL NETWORKS

Social network models allow researchers to conceptualize social structures as patterns of relations, and understand how an individual is influenced by a social structural environment (Wasserman and Faust, 1994). The aim for using formal methods to show social networks such as mathematical and graphical techniques is to represent the descriptions of networks compactly and systematically. In the analysis of complete networks, three strategies for modelling social networks can be found: (1) descriptive methods, also through graphical representations; (2) mathematical analysis procedures, often based on a decomposition of the adjacent matrix; and (3) statistical models based on probability distributions (Jamali and Hassan, 2006).

First, graphical representations: graph theory provides a vocabulary which can be used to label social structural properties. Also, gives a representation of a social network as a model consisting of a set of actors and the ties between them. When a graph is used as a model of a social network, points or nodes are used to represent social entities, and lines, connecting the points, are used to represent the ties between them. Figure 1 (adapted) is a graph that describes the structure of relations between the entities A, B, C, D, E, F and G (Churchill and Halverson, 2005). In the figure, the circles are nodes and lines between them are links (Churchill and Halverson, 2005). The links corresponds to the sending messages act between entities. Entity A is linked to two subgroups and also to an isolated entity G. The arrows shows which are the directed connections (e.g. A sends mail to E) or undirected (e.g. A sends mail to F and F sends email to A) (Churchill and Halverson, 2005). Node *A* can be characterized as a boundary entity between two subgroups, and potentially a point of connection between them.

The visual representation of data that a graph offers allows researchers to uncover patterns that might otherwise go undetected. Graphs have been widely used in SNA as a mean of formally represent social relations and quantify social structural

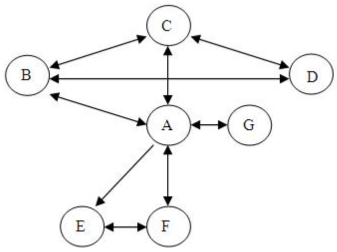

Figure 1: “Send Mail To” Graph.

wide review on this approach (Iacobucci, 1994). Still, inspired by social tagging mechanisms, Peter Mika has formulated a generic model of semantic-social networks in the form of a graph of person, concept and instance associations, extending the traditional concept of ontologies (concepts and instances) with the social dimension. His work showed how community-based semantics emerges from this model through a process of graph transformations (Mika, 2005).

Second, mathematical analysis procedures: matrices are another way to represent networks. A matrix contains the same information as a graph, but is more useful for computer analysis, because matrix operations are widely used for definition and calculation in SNA (Wasserman and Faust, 1994). The adjacency matrix is the primary matrix used in SNA, usually referred as a sociomatrix (Iacobucci, 1994). The entries in the matrix indicate whether two nodes are adjacent or not. The incidence matrix, records which lines are incident with which nodes (Iacobucci, 1994). Figure 2 (adapted) shows a matrix representing the Figure 1 connections (Churchill and Halverson, 2005). In the matrix, value 1 indicates the presence of a connection and value 0 the absence. The absence of a link between *A* and *D* is represented by the zero value in both cells of the matrix. Entity *A* is related to *E* via a directed link and *E* is not directed to *A* (Churchill and Halverson, 2005).

And third, statistical models: earlier statistical methods for SNA were introduced by Wasserman and Faust (Wasserman and Faust, 1994), but in recent years there has been a growing interest in exponential random graph models (ERGMs) called the *p** class of models. The exponential random graph models describe a general probability distribution of graphs on *n* nodes. The possible ties among nodes of a network are regarded as random variables, and assumptions about dependencies among these random tie variables determine the general form of the ERGM for the network. The Markov random graphs are one particular class of

	A	B	C	D	E	F	G
A	1	1	1	0	1	1	1
B	1	1	1	1	0	0	0
C	1	1	1	1	0	0	0
D	0	1	1	1	0	0	0
E	0	0	0	0	1	1	0
F	1	0	0	0	1	1	0
G	1	0	0	0	0	0	1

Figure 2: “Send Mail To” Matrix.

ERGMs (see (Robins et al., 2007) for a summary to the formulation and application of ERGMs for social networks). Mathematical and graphical SNA techniques allow to represent the descriptions of networks compactly and systematically. For small populations of actors (e.g. the people in a neighbourhood, or the business firms in an industry) it is possible to describe the pattern of social relations that connect the actors using words. However, to list all logically possible pairs of actors, and describe each kind of possible relations, if the number of actors and number of relation types is large, formal representations ensure that all the necessary information is systematically represented, and provides rules for doing so in ways that are much more efficient than lists (Hanneman, 2010). Robert Hanneman considers that social network analysis is more a branch of "mathematical sociology” than "statistical or quantitative analysis" though networkers most certainly practice both approaches (Hanneman, 2010). He advocates that the distinction between the two approaches is not clear. In his words: “Mathematical approaches to network analysis tend to treat the data as deterministic. That is, they tend to regard the measured relationships and relationship strengths as accurately reflecting the real or final or equilibrium status of the network. Mathematical types also tend to assume that the observations are not a sample of some larger population of possible observations; rather, the observations are usually regarded as the population of interest.

Statistical analysts tend to regard the particular scores on relationship strengths as stochastic or probabilistic realizations of an underlying true tendency or probability distribution of relationship strengths” (Hanneman, 2010). In reviewing the main results of the analysis and modelling of networks, Watts describes the main network modelling approaches, regarding structure, connectivity, searchability, and degree distributions. He concludes that the current generation of network-related research is a rapidly emerging, and a highly interdisciplinary synthesis occurs, with new analytical techniques with greater computing power, and an unprecedented volume of data (Watts, 2004).

The scope and use of statistical approaches have been extended in recent years, through methods for SNA focusing on longitudinal network data, which is understood as two or more repeated observations of a graph on a given node set. Longitudinal network data is the most frequently network data format in social sciences. Several lectures discuss models designed to analyze such data, as proposed in (Snijders, 2001) and (Snijders et al., 2007). There is little difference between conventional statistical approaches and SNA approaches. Regarding to (Hanneman, 2010), univariate, bi-variate, and even many multivariate descriptive statistical tools are commonly used to describe, explore, and model social network data. Social network data is easily represented as arrays of numbers. For Hanneman, algorithms from statistics are commonly used to describe characteristics of individual observations and the network as a whole, and concludes that statistical algorithms are very heavy used in assessing the degree of similarity among social entities, and if finding patterns in network data (e.g. factor analysis, cluster analysis, multi-dimensional scaling). Even the tools of predictive modeling are commonly applied to network data (e.g. correlation and regression).The most common emphasis in the application of inferential statistics to social science data is to answer questions about the stability, reproducibility, or generalizability of results observed in a single sample (Hanneman, 2010).

Using mathematical and graphical SNA techniques to describe a network, consider a scenario with multiple relations, i.e. where there may be more than one relation in a social network data set (e.g. more than one relation defined on pairs of actors from group $\mathcal{N}$), R represents the number of relations. Each relation can be represented as a graph and has a set of arcs $\mathcal{L}_{\mathrm{r}}$ containing L_{r} ordered pairs of actors with $1 \leq r \leq R$. Each set R defines a directed graph with nodes in $\mathcal{N}$. These graphs can be seen in one or more pictures. Each relation is defined on the same set of nodes, and each has a different set of arcs. A relation r is given by $(\mathcal{N}, \mathcal{L}_{\mathrm{r}})$ with $r = 1, 2, \ldots, R$. Consider Figure 3 which presents a scenario for three possible relations.

We conclude that the methods outlined above are essential to analyze existing social networks but we believe that new visual models are needed to infer and represent new or established patterns of relations.

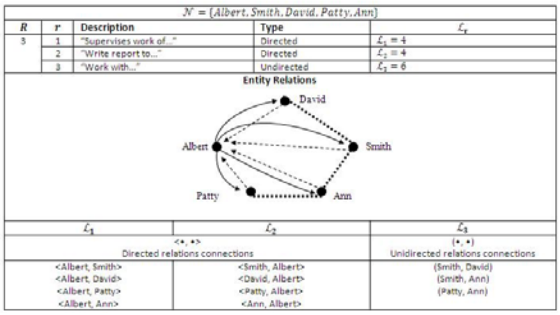

Figure 3: Multiple Relations Scenario.

3 SNARE LANGUAGE

The SNARE language proposed in this paper provides a representation of an abstract social network structure using UML (www.uml.org) as a formal descriptive method. SNARE is acronym for "Social Network Analysis and Reengineering Environment". It is an engineering artifact to represent social networks and allow researchers to design and build real scenarios for social networks extraction and relational knowledge discovery. As mentioned before, this language is the conceptual root for SNARE system, which has the main purpose to support social network analysis in information systems not designed originally for the effect. Through the instantiation of SNARE language, it is possible to analyze social entities and multiple relations among them. Based on dynamic and multiple aggregations, this language supports *N-mode* networks. SNARE language main concepts are: *Social Entity*, *Relation*, *Role*, *Action and Event* as depicted in Figure 4. Three additional concepts to support multiplicity and give flexibility were engineered: *RelationExtreme*, *ActionExtreme* and *EventExtreme*.

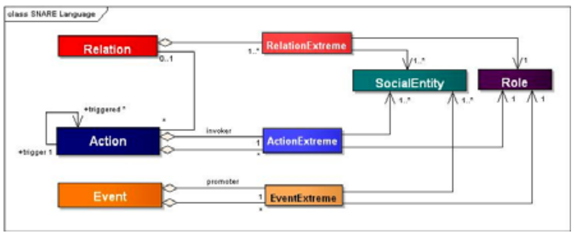

Figure 4: SNARE Language Key Concepts.

The *SocialEntity* represents an entity, typically a person, an organization, a department or a group in general terms.

The *Relation* represents a kind of connection between two or more social entities and can be expressed in different ways depending on the scope. In a family context we consider relations such as *IsFatherOf*, *IsSonOf*, *IsBrotherOf*, etc. In an enterprise context other relations emerge, for example, *IsColleagueOf*, *IsBossOf*, *IsCustomerOf* or *IsSupplierOf*. In an academic context, we find relations such as *IsTeacherOf*, *IsMentorOf*, *IsStudentOf* or *IsResearcherWith*. The *Relation* encapsulates much of the semantic that characterizes the connection between social entities. However, is not always a simple task to find the correct semantic to describe a relation. If we consider the relation *IsColleagueOf*, the relation's semantic definition is facilitated since the relation is bidirectional. That is, if the person A is a colleague of the person B, then B is also colleague of A. On the other hand, considering the family context again, the *father-son* relation, we can not concentrate all the semantics that characterizes the relation in a single class *Relation*. If A is the father of B, we can not say that B is the father of A. I.e. the relation is not bidirectional. However, this language supports the representation of any real-world connection due to the *RelationExtreme* concept, that gives flexibility and characterizes any type of relation. Each entity has a role. So, the *RelationExtreme* maintains the consistency of the connection as it allows differentiate roles in the same relation.

The *Role* support semantic roles in a given context, such as teacher, student, father, child, administrator or executive director. Returning to the family context model, the case *father-son*, probably the best solution is to define a relation *IsMemberOfFamily* and use the roles of each social entity to differentiate the semantic aspects of *father* and *son*. The SNARE language supports also this flexibility.

In social relations, it occurs sometimes different types of actions. The *Action* concept captures these flows between entities. The SNARE language makes it possible to keep track actions performed by social entities. For example, in an academic context, considering *ResearchesWith* relation, *WriteAnArticle* can be defined as an *Action*. Thus, the SNARE language allows keeping track of all articles written by participants in the relation: *ResearchesWith*.

Finally, events are part of people's lives. Participants of events do not always have met before. However, when participating in an event, probably a social entity will get involved in new relations. In order to accomplish this fact, we decide to include in our language the concept *Event*.

The SNARE language ensures that relations, actions and events can have multiple extreme instances, this is a flexibility requirement.

4 DISCUSSION

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

Social network analysis is an emergent technique to identify and understand relations and interactions among social entities, related patterns and meanings, to support social or organizational changes that reveal how networks grow, and find potential causes and consequences of a network change, previewing and controlling networks evolution.

In the analysis of complete networks, we found three strategies for modelling social networks: (1) descriptive methods, also through graphical representations; (2) mathematical analysis procedures, often based on a decomposition of the adjacent matrix; and (3) statistical models based on probability distributions. Graphical representations such as graphs tend to show connections between social entities ignoring valuable semantic aspects of the relations. Mathematical and statistical methods are focused on achieving results which are translated into algebric expressions, numerical matrices or coefficients to analyze. We consider that mathematical analysis procedures and statistical models complement our work, but our approach is a different way for graphic representation of social networks and semantic descriptions. Common techniques shows social networks as maps of entities and connections between them. These concepts are often displayed in a diagram, where nodes are the points and ties are the lines. There can be many kinds of ties between the nodes. I.e. there can be many relation types between social entities, and the resulting diagrams are often very complex to uncover related semantic concepts. In order to understand how an individual is influenced by a social structural environment, it is also necessary to identify the semantic of relations in a given social network. This process helps researchers to conceptualize and identify social structures as patterns of relations.

To conceptualize social structures in a network using our language to model social networks, the process requires the instantiation of social entities, roles, relations, actions and events. To do this instantiation, a set of stereotypes can be used. The richness of this language to model social networks comes from the flexibility to combine these stereotypes. The flexibility is expressed by all possible links that may exist on a network without adding redundant instantiations. E.g. a social entity can play several roles in the same relation, and this concept should be achieved through the instantiation of a factorized *Role* stereotype. Regarding to the connection patterns we studied, SNARE language captures all the possible social network relations. Also, it is possible to introduce new stereotypes or adapt existing ones. SNARE language ensures that relations, actions and events can have multiple extreme instances and the social network system keeps references to all previous concepts. After applying SNARE language to several scenarios, we conclude that it is flexible to fit the needs of modeling social networks. Considering organizational consulting processes, instead of statistical or mathematical representations, the notation we use leads to a significant easing of communication, visualization and discussion. When comparing with other presented social networks representation techniques (Figures 1, 2 and 3), SNARE language includes a new collection of diagrammatic model elements. These elements are more expressive to capture social network semantic concepts. Also, they are unambiguous and supported by UML tools. The SNARE language notation is a *well-known* standard derived, which can grows as the requirements for modeling grow. If the basic functionality of SNARE language is not sufficient, it is possible to extend it through the use of stereotypes.

From the research 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 social network analysis for organizational environments. In the future, our goal is to provide a tool for social networks patterns design and analysis.

REFERENCES

Borgatti, S. P., *The State of Organizational Social Network Research Today*, Boston University, 2003.

Brass, D. J. G., Joseph; Greve, Henrich R.; Tsai, Wenpin, *Taking Stock of Networks and organizations: a multilevel perspective*, in *Academy of Management Journal, 47(6)*, 2004.

Carrington, P. J., J. Scott, and S. Wasserman, *Models and methods in social network analysis*. In *Structural analysis in the social sciences*, 27, Cambridge University Press, 2005.

Churchill, E. and C. Halverson, *Social Networks and Social Networking.* IEEE Internet Computing, 9 (5), 2005.

Han, J. and M. Kamber, *Data mining : concepts and techniques*. Morgan Kaufmann, 2006.

Haythornthwaite, C., *Social Network Methods and Measures for Examining E-Learning. Graduate School of Library and Information Science*, University of Illinois at Urbana-Champaign. 2005. Available at: http://classweb.lis.uiuc.edu/~haythorn/

Iacobucci, D., *Graphs and Matrices*, in *Social Network Analysis: methods and applications. Structural analysis in social the social sciences series*, C.U. Press, 1994.

Jamali, M. A. and A., Hassan, *Different Aspects of Social Network Analysis*, in *Proceedings of the 2006 IEEE/WIC/ACM*, 2006.

Licamele, L. B., Mustafa; Getoor, Lise; Roussopoulos, Nick, *Capital and Benefit in Social Networks*, in *Proceedings of the LinkKDD*, 2005.

Masolo, C. V., Laure; Bottazzi, Emanuele; Catenacci, Carola; Ferrario, Roberta; Gangemi, Aldo; Guarino, Nicola. *Social Roles and their Descriptions*. in *Proceedings of the 9th ICPKRR, AAAI Press*. 2004.

Mika, P., *Ontologies are us: A unified model of social networks and semantics*, in *Proceedings of the 4th ISWC*, LNCS 3729, Springer-Verlag. 2005.

Robins, G. P., Pip; Kalish, Yuval; Lusher, Dean, *An Introduction to exponential random graph (p*) models for social networks*. In Social Networks, 29, Elsevier, 2007.

Wasserman, S. and K. Faust, *Social network analysis: methods and applications*. In *Structural analysis in the social sciences*, 8, Cambridge University Press. 1994.

Watts, D., *The "New Science of Networks"*, in *Annual Rev. Sociol. 30*, 2004.

Welser, H. T. G., Eric; Fisher, Danyel; Smith, Marc, *Visualizing the Signatures of Social Roles in Online Discussion Groups*, in *The Journal of Social Structure*. 8 (2), 2007.

Snijders, T. A. B., *The statistical evaluation of social network dynamics*. In M. E. Sobel and M.P. Becker (Eds.), *Sociological methodology*. Basil Blackwell, 2001.

Snijders, T. A. B. et al., *Modeling the co-evolution of networks and behavior*. In K. van Montfort, H. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences*, Lawrence Erlbaum, 2007.

Hanneman, Robert A., *Introduction to Social Network Methods*, Riverside, 2010. Available at: www.analytictech.com/networks.pdf

Object Management Group, *UML Website* (accessed in May 2010): www.uml.org.