Published in the Proceedings of
The 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI'06), Hong Kong, December 2006, pp. 113-116.

Relationship Algebra for Computing in Social Networks and Social Network Based Applications

Javed I. Khan & Sajid Shaikh
Internetworking and Media Communications Research Laboratories
Department of Computer Science
Kent State University
233 MSB, Kent, OH 44242, USA
Javed|sshaikh@kent.edu

Abstract

Online communities are the latest phenomena on the Internet. At the heart of each community lies a social network. In this paper, we show a generalized framework to understand and reason in social networks. Previously, researchers have attempted to use inference-specific type of relationships. In this paper, we propose a framework to represent and reason with general case of social relationship network in a formal way. We call it relationship algebra. This paper presents this algebra and then shows how this algebra can be used for various interesting computations on a social network weaved in the virtual communities.

1. Introduction

Online communities and social networking are one of the latest and fastest growing phenomena of the Internet. The websites providing social networking services are fast becoming an important cog in the borderless world of Internet [1]. Communities such as Orkut® [4], Yahoo360®, MySpace®, LinkedIn® are developing and offering various interesting social networking services. In the surface these networks are providing highly personalized services. For example, these provide tools so that people of similar interest or fields can meet, communicate and exchange ideas. One can also conveniently use it simply as a medium for personal and creative expression. However, at the core these are also collecting very interesting information about the complex web of relationship between individuals. These communities interweave these services within the context of relationship space of individuals. Figure-1 provides a snapshot portal from such two sites. These portals represent just one node in a vast network comprising of millions and millions of nodes. Even though the primary objective of many of these websites is to connect people over the Internet to their immediate friends, but the overall network potentially connects much larger people space. Consequently, these social networks are now capturing a whole new form of knowledge which was previously out side the ream of machine processing. For the first time, the community networks are now making wide variety of relationship information available in a digitally encoded format. A new generation of very powerful applications is now conceivable from the relationship information available in them. In this paper we present a generalized inference and computation framework on relationship information available in these networks and explain their application by way of two examples- conflict of interest assessment and immunization listing. The paper provides a serious glimpse at the techniques and a range of new applications now on the horizon.

2. Relationship Algebra

2.1 Representation

The world is comprised of unique entities that are each represented by an entity ID. Entities in this world are however, organized as members of various sets and these entities can be members of multiple sets. Members have also membership index in each set. The membership index of an entity does not have to be the same between its member sets. In a way, all objects in this world are members of a super set E and there exists relationship between various pairs of these smaller sets.

Let A and P be the subsets of the author and paper set respectively, then the cross product M^r=AXP is the *author-to-paper* relationship matrix. Each matrix element m_{ij} represents the relationship strength. A binary one denotes the presence of a relationship between a_i and p_j and a zero its absence.

Published in the Proceedings of
The 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI'06), Hong Kong, December 2006, pp. 113-116.

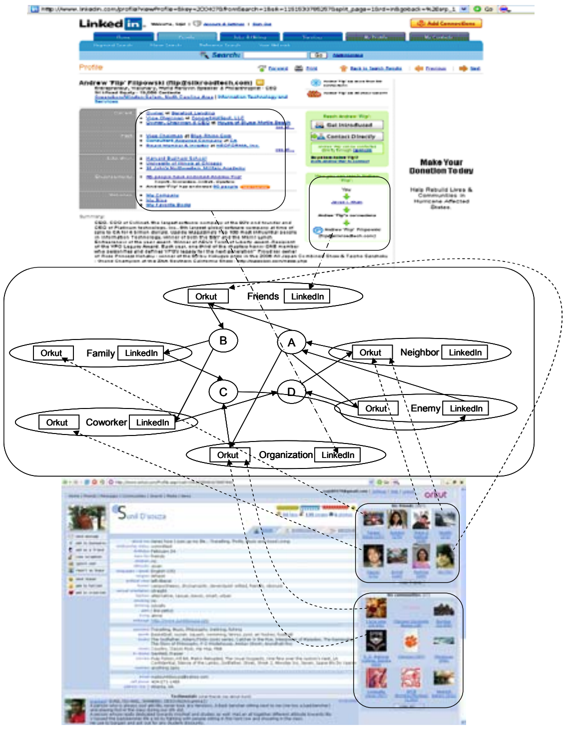

Figure 1: The top part shows a LinkIn® web page, the central part shows a social networks language graph and the bottom part of the figure shows an Orkut ® web page.

2.2 Relationship and Set Operators

This section defines the set operations and the semantic operators that are applied on the relation matrices. If A and B are relationship matrices then table 1 enumerates the various set operations and table 2 presents the semantic operators. 3. Examples of social networks

3.1. Publication Network's Language Graph

The major entities involved in the publication network are the Authors, the Organizations, the Paper, the Journal, the Reviewers, the Editors and the Topic Area. The topic area is the focal point of the network. Figure-2 shows a language schema to depict a publication network. Figure-3 shows an example of publication network defined using this language.

3.2. Application: Reviewer Selection

We would be using the publication network for illustrating how we can use the relationship algebra presented in section 2 for useful purposes such as *"Finding a set of reviewers for a particular paper"*. A set of constraints express the reviewer selection. Below is such an example set:

Reviewer Selection Constraints*: (i)The reviewer should not be a coauthor of the paper s/he is going to review, (ii) S/he should not be a coworker of the author for example the author and the reviewer should not be faculties in the same university. (iii) The reviewer should not have submitted a paper in the same journal or conference and (iv) finally s/he should be well acquainted with the subject area being discussed in the paper.*

Operation	Sym.	Explanation
Column Extraction	Ψ	$\Psi_i^X(M)$ where $X \text{ is } M_{ij} \geq \mu$ or $X \text{ is } M_{ij} < \mu$
Row Extraction	ρ	$\rho_j^X(M)$ where $X \text{ is } M_{ij} \geq \mu$ or $X \text{ is } M_{ij} < \mu$
Max Row	ξ	$\xi_j^X(M)$ where $X \text{ is } M_{ij} > \forall M_{kj}$
Max Column	ϕ	$\phi_i^X(M)$ where $X \text{ is } M_{ij} > \forall M_{ik}$
Zero Column	θ	$\theta^X(M)$ where $X \text{ is } M_{ij} = 0 \text{ and } 0 < i < n$

Table 1: Set Operators

Operation	Symbol	Explanation
Equivalence	A=B	$a_{ij} = b_{ij}$
Reflection	$R = A^T$	$r_{ij} = a_{ji}$
Synthesis	S = AxB	$s_{ij} = \sum_{r=1}^{n} a_{ir} b_{rj}$
Intersection	E=A⊗B	$e_{ij} = [a_{ij} \cap b_{ij}]$ where : $0 < i < n, 0 < j < m$
Union	U=A⊕B	$u_{ij} = [a_{ij} \cup b_{ij}]$ where : $0 < i < n, 0 < j < m$
Exclusion	X= AΘB	$x_{ij} = [a_{ij} - b_{ij}]$ where : $0 < i < n, 0 < j < m$
Dediagonalization	$(\hat{A})$	$A_{ij} = [0 \text{ if } i == j]$ where : $0 < i < n, 0 < j < m$
Quantization	$[A]^{\mu}$	$a_{ij} = \begin{bmatrix} 1 \text{ if } a_{ij} \geq \mu \\ 0 \text{ if } a_{ij} < \mu \end{bmatrix}$ where : $0 < i < n, 0 < j < m$

Table 2: Relationship Operators

Published in the Proceedings of
The 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI'06), Hong Kong, December 2006, pp. 113-116.

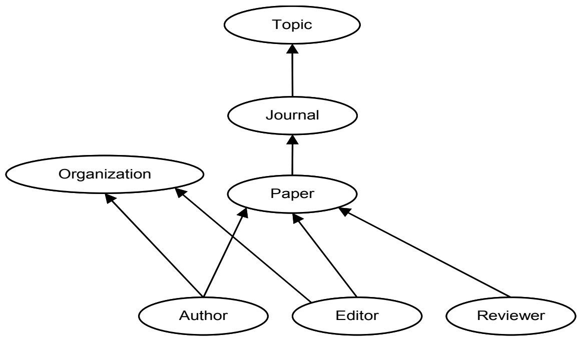

Figure 2 :Language Graph of Publication Network

Primary Relationship	Notation
Journal → Topic Area	$M_{J_i}^{J-T}$
Editor → Paper	$M_{E_i}^{E-P}$
Paper → Journal	$M_{P_i}^{P-J}$
Author → Paper	$M_{A_i}^{A-P}$
Reviewer → Paper	$M_{R_i}^{R-P}$
Author→ Organization	$M_{A_i}^{A-Org}$

Table 3: Publication Network's Primary Relationship

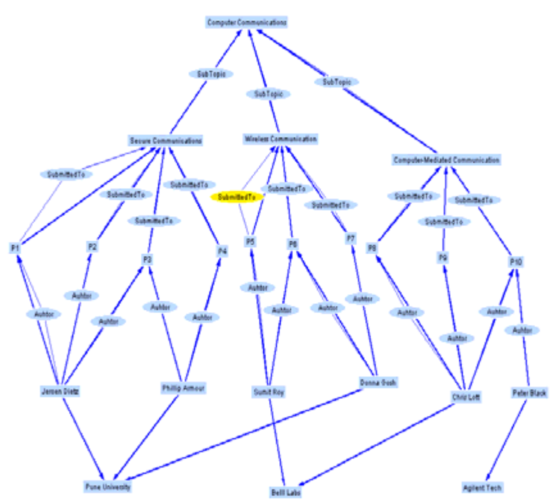

Figure 3: Instance Graph

The aim is to find a reviewer set for papers P_5 and P_6 from among the four authors available. The first step is to determine the authors and coauthors for P_5 and P_6. This is computed by multiplying the author-paper relationship matrix M_A^{A-P}, with its transpose matrix. The resultant matrix $M_{coAuthor}$ represents the co-author relationship between the various authors. The next step is to find out, which authors have submitted papers in the same journal. In order to determine this we first need to establish a relationship matrix depicting the relationship between the authors and the journals M_A^{A-J}. This matrix is obtained by multiplying the two matrices M_A^{A-P} and M_P^{P-J}. The resultant $M_{coJournal}$ is a product of M_A^{A-J} and its transpose. After this, we determine the coworker relationship between the authors. The coworker matrix $M_{coWorker}$ is the product of M_A^{A-Org} and its transpose. Finally, we compute the non-conflict of interest matrix $M_{nonConflict}$ by subtracting each of the *coAuthor, coJournal* and *coWorker* matrices from the matrix depicting the relationship between all the authors in our instance graph M_{all}. The reviewer set matrix $M_{reviewer}$ is calculated by multiplying the $M_{nonConflict}$ and the $\left(M_A^{A-P}\right)^T$ matrices. Applying the row extraction set operation ρ on the reviewer set matrix gives us the reviewer set for papers P_5 and P_6.

$$M_{reviewer} = M_A^{A-P}\left[M_{all} - M_{coAuthor} - M_{coJournal} - M_{coWorker}\right] \quad (1)$$

$$\text{ReviewerSet}(P_i) = \rho_i^{M_{ij}=1}\left(M_{reviewer}\right) \quad (2)$$

In the above discussion, we have mentioned the term *conflict of interest*, which can be defined as follows A **conflict of interest** consists of three entities, the source *"i"*, the sink *"j"* and the relationship between them *"R"*. It occurs if we have two distinct relationship trails R_1 and R_2 from i to j and their intersection set is nonempty.

$$S_i^j(R_1) \cap S_i^j(R_2) \neq \phi \quad (3)$$

3.3. Social Network's Language Graph

An individual's social network primarily consists of family, friends, neighbors, coworkers and the organizations with which he is affiliated. The circles denoted by A, B, C and D in figure 4 are individuals in a community. The ellipses denote the types of relationships we have considered for our example. The smaller rectangles within the ellipses denote the refined relationships for each class. Figure-5 shows an example of social network defined using this language. Table 4 enumerates the primary relationships of our example social network.

3.4. Application: Immunization

Suppose George is the virus carrier and our aim is to find out the people whom he might have infected and need vaccination. In order to achieve this we need to determine a *vaccination set* from George's social

Published in the Proceedings of
The 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI'06), Hong Kong, December 2006, pp. 113-116.

network. The people to be included in this set should satisfy certain conditions. For our particular example, the conditions are as follows:

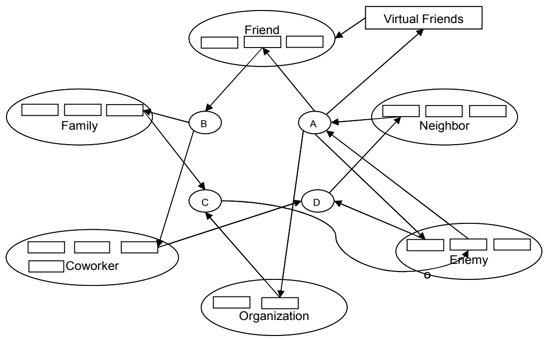

Figure 4: Language Graph of Social Network

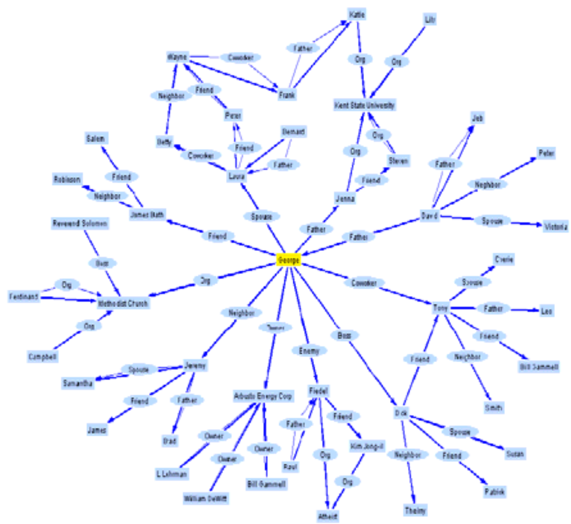

Figure 5: Social Network's Instance Graph

Primary Relationship	Notation
Individual → Friend	$M_{Ind}^{Ind-Ind(Friend)}$
Individual → Father	$M_{Ind}^{Ind-Ind(Father)}$
Individual → Neighbor	$M_{Ind}^{Ind-Ind(Neighbor)}$
Individual → Coworker	$M_{Ind}^{Ind-Ind(Cowoker)}$
Individual → Spouse	$M_{Ind}^{Ind-Ind(Spouse)}$

Table 4: Primary Relationship for The Social Network

Immunization Constraints: *(i) The first set of people comprised of the ones who are in close contact with George. We have to immunize them. (ii) The second set comprises of people who are likely to be infected belong to George's derived network i.e. his greater than 1-hop neighbors. The likelihood of them been infected depends upon their relationship strength with George's 1-hop neighbors. For our example, the threshold value is 0.6.*

Our first step is to derive the relationship strength between George and his greater than 1 hop neighbors. The matrices $M_{George}^{Spouse}, M_{George}^{Father}, M_{George}^{Neighbor}, M_{George}^{Coworker}, M_{George}^{Friend}$, help us in establishing these derived relationships. The next step is to apply column extraction on the first row of each of these matrices to get a set of people who are most likely to be infected.

$$set\left\{A\right\}^{Spouse} = \psi_{M_{ij}>0.6}^{A}\left(M_{A}^{Spouse}\right) \quad (4)$$

The final vaccination set is an union of the five smaller sets.

$$vaccinationset\{A\} = set\{A\}^{Spouse} \cup set\{A\}^{Father} \cup set\{A\}^{Friend} \cup set\{A\}^{Coworker} \cup set\{A\}^{Neighbor} \cup set\{A\}^{So} \quad (5)$$

4. Conclusions

Social networks are the latest wave of innovation on the Internet. Millions are participating. The variety and quantity of information being shared by individuals on these websites both are unprecedented. In this paper, we have demonstrated a few applications such as reviewer selection and immunization, which will be possible using the information collected by similar services. Trust propagation is another interesting area for the application of the algebra. The relationship algebra can be used to define various forms of trusts and determine various combinations and synthesis in a programmable way. Thus, the algebra helps in deriving complex and apparently hidden relationships in almost algebraic manner, which may not be obvious to any individual owner of information in the chain. The applications proposed here seems to be inevitable. In the light of such developments, it is important to be aware of the overwhelming implications in privacy.

5. References

[1] Scott, J., "*Social Network Analysis: A Handbook*", Sage Publications, London, 2nd ed. (2000).

[2] Wasserman, S. and Faust, K., "*Social Network Analysis*", Cambridge University Press, Cambridge (1994).

[3] Mizruchi, M. S., "*The American Corporate Network*, 1904–1974, Sage, Beverley Hills (1982).

[4] Meneses, J., "*The Orkut.com case: a reflection on the exploration of new ways to online sociability in the tradition of the study of virtual communities*". (2004)