Social Measures and Flexible Navigation on Online Contact Networks

Nikolaos Korfiatis and Miguel–Angel Sicilia

***Abstract*— This paper discusses ways of navigating online contact networks - networks of social connections defined under a relational context - on a way that can provide much more meaningful information to those that use them. In particular we use the concept of a weak and strong relational tie to address the different levels of social distance and express them using a trapezoid function of fuzzy numbers. We provide a numerical example using a fuzzy graph where the strength or weakness of the connections is defined as a fuzzy number and address some of the advantages of this approach to provide better information for the members of this contact network.**

I. Introduction

Sociological explorations of formal representations about groups of interconnected entities such as people or institutions have been on the focus of graph/network theoretic community for several decades [1]. In principle the paradigm of networks has been explored in several applications from organizational science to applications in diverse fields such as biology and recently on the internet [2], putting lately a general focus on social science and organizational research [3]. With the development of the internet the concept of online social networking applications has contributed new dynamics into the way social networks are formed and used. Due to the fact that online communication such as email has made geospatial and time distance not more an issue, the expansion of the network and the acquisition of new members is something that can be done very easily through the use of online communication. The later has lead to a huge ocean of social connections that can be exploited for the benefit of an individual actor or a group of actors (e.g. connection with old friends, colleagues etc). One particular type of this online social networks are contact networks. Contact networks represent social ties for some specific purpose. In the context of professional practice, one of its principal applications is job search through personal contact or recommendations. Web-based systems such as LinkedIn provide the means to represent incrementally contact networks by a simple model in which individuals ask others (usually by email) to connect to them in the on-line system and provide some explanation of the history of its social tie.

However a major drawback of these online contact systems is the expressiveness of the social connections that can be defined in these contact networks and furthermore the strength and the importance of these ties to those that use the network. A particular example could be for instance the exploitation of the contact network for a purpose (e.g. find a good job). Although the online contact systems provide the ability to reach several degrees of relations there is an issue on finding those contacts or "bridges" that will be the most reliable/useful. This happens mainly due to the difficulty of capturing the quality of this online social connection [4]. There can be cases that a connection is established typically usually after a short meeting in a conference or is established as an old co-worker or classmate. In both cases the relation is defined as an existent/not-existent which doesnt give any clue of each quality. Therefore the goal of this paper is to describe this problem using imprecise models of social connectedness over social networks and discuss their possible applications into the effective design of online contact systems.

Previous work in the application of fuzzy set theory to modeling social nets include group models in information retrieval [5] the use of fuzzy models combined with query resolution strategies [6] and the adaptation of well-known Web page ranking algorithms to a social context [7].

The rest of this paper is structured as follows. Section 2 discusses sociometric foundations of contact/social network navigation and their shortcomings. Section 3 discusses the application of fuzzy set theory and fuzzy discrete structures, in that case fuzzy graphs, to the modeling of social ties and ranking on fuzzy based contact networks. We present a numerical example as a case study on section 4 as well as conclusions and discussion for future research.

II. Representing and Navigating Contact Networks

Navigation in contact networks entails traversing the graph of ties in some way. Following the notation of *social relational systems* in [8], we have a single set of g actors $\mathcal{N}$ and several relations defined on that set $\mathcal{L}_1, \ldots, \mathcal{L}_k$. Each of these relations define a directed graph $\mathcal{G}_i =< \mathcal{L}_i, \mathcal{N} >$, and individual ties are denoted as $n_i \rightarrow n_j$.

In a contact network the social distance between all actors is defined as the *characteristic path lentgh* $(D, \mathcal{G})$, that is the typical distance $d(i, j)$ between every actor (ego) and every other actor(alter). Let us consider a pair of two actors : i and j. In that case $d(i, j)$ the social distance is defined as the *shortest path length* between the ego i and the alter j.

Navigation in contact networks starts usually from a single node n_i (e.g. the user logged in the Web contact system), and proceeds by levels. This level-based navigation has several major drawbacks:

Nikolaos Korfiatis is with the Department of Informatics, Copenhagen Business School (CBS),Howitzvej 60, DK-2000 Frederiksberg (Copenhagen) ,Denmark (email: nk.inf@cbs.dk).

Miguel–Angel Sicilia is with the Department of Computer Science, University of Alcalá, Ctra. Barcelona, km. 33.6 28871, Alcala de Henares (Madrid), Spain (email: msicilia@uah.es).

TABLE I

FORMAL AND LINGUISTIC REPRESENTATION OF THE CHARACTERISTIC PATH LENGTH AND THE SOCIAL DISTANCE

Characteristic Path Length	Expresions of Social Distance
$D = 1$	Very Close
$D = 2$	Close
$D = 3$	Quite Close
$D > 3$	Distantly Close

i Every tie to other actors in the network is not considered, i.e the *strength* of the ties is not considered.

ii The position of the actors tied to oneself is also not considered, i.e. there is not a consideration of “prestige”.

iii There is not a consideration of the formation of “clusters” in the overall network.

A straightforward representation of a fuzzy tie can be introduced in current contact models to obtain a model that considers [i]. Then, the (crisp) level–based navigation in the network can be substituted by a *fuzzy level navigation*. The strength of ties in a contact network can only be assessed by the actors themselves. This fits in the current functionality of contact networks in which it is the user that “invite” others to link with him/her. In consequence, the valence of a concrete tie $n_i \rightarrow n_j$ is considered to be provided by n_i. Classical fuzzy graph network models [9] provide several options to express the “length” of arcs. In the domain under consideration, it seems reasonable to let the users assess the strength of the tie by a fuzzy number. It should be noted that from a pragmatic viewpoint, the representation of those fuzzy numbers could be hidden to the actors by providing, for example, a single linguistic classification scale.

Fig. 1. Contact Network

III. FUZZY MODELS FOR SOCIAL DISTANCE

A. The model

We continue our discussion by introducing a model of a contact network using fuzzy set theory and discussing the related representations and concepts. In particular we lconsider a fuzzy directed graph $\widetilde{\mathcal{G}}_i =< \mathcal{V}, \mathcal{E}, \widetilde{\mathcal{C}} >$, where $\mathcal{V}$ is a set of contacts in the contact network $\mathcal{G}$ and $\mathcal{E}$ is the set of a binary relation of their social connection such as : $e_k = (u_i, u_j) \in VxV$ is part of the adjacency matrix. The fuzzy set $\widetilde{\mathcal{C}}$ contains theindividual ties of an imprecise strength are which denoted as $t_k = n_i \widetilde{\rightarrow} n_j$. Each tie has an associated fuzzy number denoted as $\mu(n_i \widetilde{\rightarrow} n_j)$, so that $\mu : t_k \rightarrow \mathcal{F}([0,1])$, i.e. the tie membership function maps ties to fuzzy numbers defined on the [0, 1] segment of the real domain. The upper and lower bound denotes special cases of social connectedness. In particular the case $t_k \rightarrow 0$ denotes that the social entities (in that case the individuals) are unknown to each other and there exists no path of undirected connection that they can reach each other. Figure shows an example network generated using the JUNG libraries and assigning a userdatum key with a trapezoid fuzzy number

B. Finding the network at a fuzzy distance

We consider the fuzzy distance of a contact network represented as in A as an individual tie of an imprecise strength such as: $D_f = e_{ij} * t_k$. To illustrate our case we use a simple graph labeling algorithm based on the BFS (Breadth First Search). The reason for using the BFS algorithm as the base for our variant is due to the fact that the contact network is non-labeled which means that every edge has the same length with another adcacent pair. The process starts by considering two inputs: the set $\mathcal{V}$ of the contacts in the contact network and the fuzzy set $\widetilde{\mathcal{C}}$ that contains the expressions of the connectedness. The process outputs the contact network labeled with a fuzzy distance metric which we can use later for ranking and importance.

```
LABELFUZZYNET(𝒱, 𝒞̃)
    ▷ 𝒱 is the set of the actors of the contact network
    ▷ 𝒮 is the people graph
𝒮 ← COMPUTESOCIALRELEVANCE(𝒮)
for each v ∈ VERTEX(𝒟)
      do
         v.source.relevance ←
         NODES(𝒮)[v.source].relevance

𝒟 ← WEIGHTEDPAGERANK(𝒟)
```

The complexity of the algorithm follows the general complexity of the BFS algorithm and is $O(|\mathcal{V}| + |\mathcal{E}|)$ where $|\mathcal{V}|$ and $|\mathcal{E}|$ is the number of actors in the contact network and the number of of their adjacent connections respectively.

C. Introducing fuzzy measures of prestige

In network navigation one important issue is to find the prominent nodes, those that have the higher status. Such representation in social network analysis is the prestige. A prestige measure is a direct representation of the social status of an entity which depending on the type of the network employs the reciprocal/non-reciprocal connections (social choices) provided to that entity along with the influence that this entity might provide to the neighboring entities. A simple measure of prestige is the in-degree of the entity on the network. On our model this can be formalized as follows:

$$k_i^{in} = \sum e_{i,j}, j > 0$$

However the inner degree index of a vertex makes sense only in cases where a directional relationship is available (the connection is non-reciprocal) and this cannot be applied in the study of non directed networks such as an online contact network. In that case the prestige is computed by taking into account the influence domain of the vertex. For a non directed graph the influence domain of a vertex is the number or proportion of all other vertices which are connected by a path to that particular vertex.

$$d_i = \frac{1}{N-1} \sum_{j=1} e_{j,i}$$

On the above measure represents the set of paths between the vertices and and is the number of all available nodes in the graph (The total number of nodes minus the node that is subject to the metric). The combination of the above two metrics of prestige results to the “Proximity Prestige” of a vertex. This encompasses the normalization of the inner degree of the vertex by its degree of influence such as:

$$PP_i = \frac{k_i^{in}}{d_i}$$

IV. Case study

We used the JUNG libraries for the modeling and presentation of the numerical example based on the graph shown in figure XXX . We used a set of java classes of fuzzy numbers and in particular the trapezoid number function as well as some aggregation operators developed in the Information Engineering Research Unit in the University of Alcala. We first

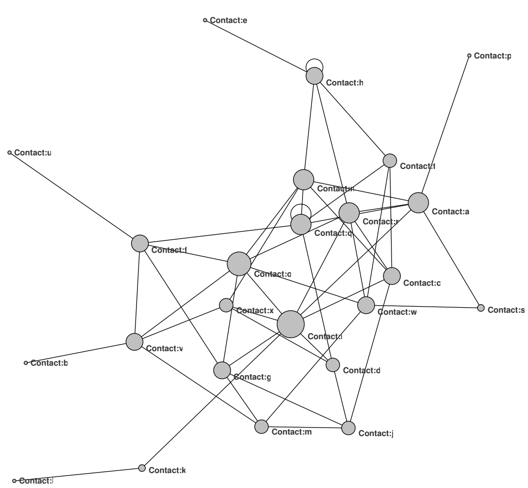

Fig. 2. The Contact Network used in our case study. The size of the vertices the prestige of each vertex based on the indegree.

In our example we extend the default graph labeling algorithm so it can assign a fuzzy number as an extra label (userdatum key) on each arc. As can be seen in the figure we select two nodes and present the distance on three respective layers similar to the three degrees of distance that is a common concept on contact networks.

TABLE II
Social and Fuzzy Based Distance in the example Contact Network

Characteristic Path Length	Expressions of Social Distance
$D = 1$	Very Close
$D = 2$	Close
$D = 3$	Quite Close
$D > 3$	Distantly Close

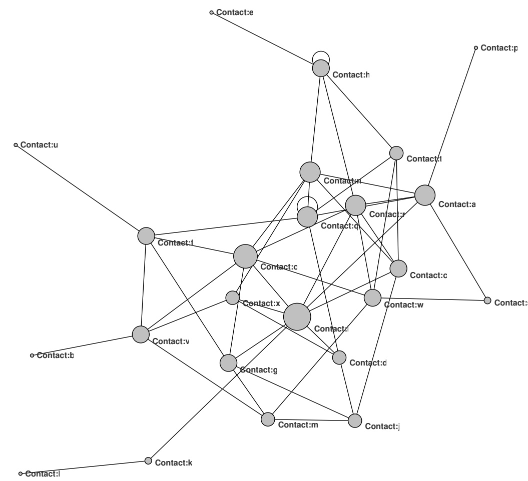

Fig. 3. The Contact Network used in our case study. Vertex Size represents the prestige of each vertex based on the indegree.

V. Conclusions and Future Research Directions

We have presented an example using fuzzy set theory and fuzzy discrete structures of the application of the above concepts to social information handling. In particular we consider the application of fuzzy set theory to the definition and manipulation of social ties in other environments, for instance collaboration networks such as wikipedia [10]. We consider the fuzzy set theory as an ideal framework for describing social connections where from their definition the connection is imprecisely perceived.

Nikolaos Korfiatis acknowledges support from the DREAMS Project at Copenhagen Business School. DREAMS is partially supported by the Danish Research Agency, grant number 2106-04-0007.

References

[1] F. Harary, R. Norman, and D. Cartwright, *Structural Models: An Introduction to the Theory of Directed Graphs.* Wiley, 1965.
[2] A.-L. Barabasi, *Linked: The New Science of Networks.* Perseus Publishing, 2002.
[3] S. Borgatti and P. Foster, “The Network Paradigm in Organizational Research: A Review and Typology,” *Journal of Management*, vol. 29, no. 6, p. 991, 2003.
[4] J. N. Cummings, B. Butler, and R. Kraut, “The quality of online social relationships,” *Communications of the ACM*, vol. 45, pp. 103–108, July 2002.

[5] M. Sicilia and E. García, “Fuzzy Group Models for Adaptation in Cooperative Information Retrieval Contexts,” *Lecture Notes in Computer Science*, vol. 2932, pp. 324–334, 2004.

[6] F. O. Cirit, M. Nikravesh, and S. E. Alptekin, “Consumer profiling using fuzzy query and social network techniques,” in *Proceedings of the BISC FLINT-CIBI International joint workshop on Soft Computing for Internet and Bioinformatics*, 2003.

[7] E. García and M. A. Sicilia, “Filtering Information with Imprecise Social Criteria: A FOAF-based backlink model,” in *Proceedings of the Fourth Conference of the European Society for Fuzzy Logic and Technology*, 2005.

[8] S. Wasserman and K. Faust, *Social Network Analysis: methods and applications*. Cambridge University Press, 1994.

[9] D. Malik and J. Mordeson, *Fuzzy Discrete Structures*. Physica Verlag, 2000.

[10] N. Korfiatis, M. Poulos, and G. Bokos, “Evaluating Authoritative Sources using Social Networks: an Insight from Wikipedia,” *Online Information Review*, vol. 30, no. 3, pp. 252–262, 2006.