

Chapter 11

Social Network Database Querying Based on Computing with Words

Ronald R. Yager

Abstract Fuzzy relationships and their role in modeling weighted social relational networks are discussed. We describe how the idea of computing with words can provide a bridge between a network analyst's linguistic description of social network concepts and the formal model of the network. We then turn to some examples of taking an analyst's network concepts and formally representing them in terms of network properties. We first do this for the concept of clique and then for the idea of node importance. Finally we introduce the idea of vector-valued nodes and begin developing a technology of social network database theory.

1 Introduction

Social relational networks have rapidly become an important technology in our digital based information intense world [1–8]. Among the notable examples of social network sites are Facebook and LinkedIn. In addition to providing an ability for enabling people from all over the world to connect with each other each they provide a vast source of information about the individual participants in the network, the so called nodes. Each of these participants can be viewed as a kind of database containing information about themselves. We see here that a social network can be viewed as kind of network of databases. This leads us to understand that the development of appropriate technologies to manage social networks requires a combination of the use of network technologies, graph theory and database technologies [9]. Furthermore the development of intelligent social network management requires the extension of these two technologies by the introduction of ideas from soft and granular computing, computational intelligence [8]. In addition to the intelligent extension of these two technologies with soft computing an important task on the road map

R. R. Yager(✉)

Machine Intelligence Institute, Iona College, New Rochelle, NY 10801, USA

e-mail: yager@panix.com

of developing intelligent social networks is the combining of network theory with database theory. Since many advances have been made toward the development of intelligent database technologies, especially by contributors to this volume, here we shall take some steps in the other two tasks, the intelligent extension of social network theory and the combining of social networks with databases. We note that one area that has received a considerable amount of research is the mining of social networks [4, 5]. We emphasize that our interest here is not on this problem but we shall be more interested in the issue of querying the social network with its nodes seen as individual databases. We shall refer to this as SONDAB-Q as an acronym for **SO**cial **Ne**twork **DA**tabase **Q**uerying.

The current social network technology can be extended and enriched to help in modeling these newly emerging applications by introducing ideas from fuzzy sets and related granular computing technologies [10–16]. We can provide this extension and enrichment in two ways. The first is with the introduction of fuzzy graphs representing the networks [17–19]. This allows a generalization of the types of connection between nodes in a network from simply connected or not to weighted or fuzzy connections. Here the idea of strength of connection becomes important. The second and perhaps more interesting extension is the use of Zadeh’s fuzzy set based paradigm of computing with words [10–12] to provide a bridge between a human network analyst’s linguistic description of social network concepts and the formal model of the network. Fundamental to this capability is the realization that both formal network models and the paradigm of computing with words are built upon set based technologies. More specifically, the formal representation of a social network is in terms of a mathematical set object called a relationship [1] and computing with words uses a set object, fuzzy subsets, to formally represent the semantics of linguistic terms. This common underlying set based technology allows us to take human concepts and formally represent them in terms of network properties. This in turn allows an analyst to estimate the truth or falsity of observations about a network as well helps in the mining of social relation networks. In an attempt to help the reader get an understanding of the technology useful in this approach we provide few examples of how we would model some social network concepts.

Another useful idea we discuss is vector-valued nodes. Here we associate with each node a vector whose components are the attribute values of the node. Combining this with the machinery of Zadeh’s computing with words we are then able to intelligently query the network with questions that involve both attributes and connections. We see this as the basis of an emerging discipline of social network database querying, SONDAB-Q

2 Fuzzy Graphs

Since a social network can be formally viewed as graph whose nodes represent the members of the social network we begin by introducing some ideas from fuzzy graphs. Here we first describe the idea of a fuzzy relationship. The concept of a

fuzzy relationship plays a fundamental role in modeling a type of weighted graph called a fuzzy graph [17–19]. Let X be a set of elements. A fuzzy relationship on X is a mapping $R : X \times X \rightarrow [0, 1]$ where $R(x, y)$ indicates the degree of relationship between x and y . We note that we can view a fuzzy relationship as a fuzzy subset on $X \times X$. This allows us to use much of the formalism of fuzzy sets. For example we can say that $R_1 \subseteq R_2$ if $R_1(x, y) \leq R_2(x, y)$ for all (x, y) . We note that $R_1 \subseteq R_2$ means that R_1 is a subset of R_2 . We shall also use the terminology $R_1 \leq R_2$.

Some notable properties that can be associated with fuzzy relationships are

- (1) **Reflexivity:** $R(x, x) = 1$ for all x
- (2) **Symmetry:** $R(x, y) = R(y, x)$
- (3) **Transitivity:** $R(x, z) \geq \text{Max}_y [R(x, y) \wedge R(y, z)]$

An important operation on fuzzy relations is composition. Assume R_1 and R_2 are two relations on X . The composition $R = R_1 \blacklozenge R_2$ is also a relationship on X such that

$$R(x, z) = \text{Max}_y [R_1(x, y) \wedge R_2(y, z)]$$

The composition operation can be shown to be associative

$$(R_1 \blacklozenge R_2) \blacklozenge R_3 = R_1 \blacklozenge (R_2 \blacklozenge R_3)$$

The associativity property allows us to use the notation $R^k = R \blacklozenge R \blacklozenge \dots \blacklozenge R$ for the composition of R with itself k times. In addition we shall define R^0 to be such that $R^0(x, y) = 0$ for all x and y .

If R is reflexive then $R^{k_2} \supseteq R^{k_1}$ for $k_2 > k_1$. On the other hand if R is transitive, it can be shown that $R^{k_2} \subseteq R^{k_1}$ if $k_2 > k_1$. From this we see that if R is reflexive and transitive then $R^{k_2} = R^{k_1}$ for all k_1 and $k_2 \neq 0$.

We shall now define various types of fuzzy graphs. Let X be a set of elements, which we shall refer to using graph terminology as nodes or vertices. We shall further assume R is a reflexive fuzzy relationship on X . The pair $\langle X, R \rangle$ can be seen as defining a fuzzy or weighted graph in which $R(x, y)$ is the weight associated with the arc $x \rightarrow y$, (x, y) . More generally if F is a fuzzy subset of X we can also define a fuzzy graph as $\langle X, F, R \rangle$. Here in addition to having degrees of connection we also have a degree to which each of the nodes belongs to the network. In this case we let $R.F$ be a relationship on X defined such that $R.F(x, y) = R(x, y) \wedge F(x) \wedge F(y)$ and say that $R.F$ is the relationship R on F . We note here that $R.F(x, x) = F(x)$. If $F = X$ then $R.F(x, y) = R(x, y)$. It can be shown here that $R.F^{k_1} \subseteq R.F^{k_2}$ if $k_2 > k_1$. We note that if $F = X$ then $\langle X, F, R \rangle = \langle X, R \rangle$.

If R is symmetric we shall say $\langle X, F, R \rangle$ is an undirected fuzzy graph. We note that if R is symmetric then $R.F$ is also symmetric. If R is not symmetric we shall refer to $\langle X, F, R \rangle$ as a directed graph and we refer to a pair (x, y) as an arc. Here the weight on the arc (x, y) and arc (y, x) may be different. In the case of where R symmetric we shall refer to the pair (x, y) as an edge. Since we shall primarily be concerned with undirected graphs, we shall simply use the unmodified term graph or network to refer to this case where R is symmetric.

At times, especially when working with undirected graphs, we shall find it convenient to consider the space U which consists of all unordered pairs of distinct elements which we denote as $\{x, y\}$. In this case we can refer to R_U as the reflection of R on U . In this $R_U(\{x, y\}) = R(x, y) = R(y, x)$. We note that at a formal level we can also view U as consisting of all subsets of X consisting of two elements.

Assume $G = \langle X, F, R \rangle$ is a fuzzy graph. A path ρ in G is a sequence of distinct nodes $x_0 x_1 \cdots x_n$. The number of links in the path is n . The strength of the path is defined as

$$ST(\rho) = \mathbf{Min}_{i=1 \text{ to } n} [R.F(x_{i-1}, x_i)].$$

If $F = X$ then $ST(\rho) = \mathbf{Min}_{i=1 \text{ to } n} [R(x_{i-1}, x_i)]$.

Two nodes for which there is a path ρ with $ST(\rho) > 0$ between them are called connected. We call ρ a cycle if $n \geq 2$ and $x_0 = x_n$

Consider the graph $G = \langle X, R \rangle$ let us now consider R^k . We can show that $R^k(x, y)$ is the strength of the strongest path from x to y containing at most k links. We see that if X has n nodes then R^{n-1} provides the strongest connection between two nodes using any number of links. If $R^k(x, y) \neq 0$ we can say x and y are connected at least of order k .

We note that if $G = \langle X, F, R \rangle$ we can make statements similar to the above about $R.F$.

Assume $G = \langle X, R \rangle$ is a fuzzy graph. Let $\rho = x_0 x_1 \cdots x_n$ be a path in X . A concept introduced by Rosenfeld [17] is the length of the path. He defined

$$L(\rho) = \sum_{i=1}^n \frac{1}{R(x_{i-1}, x_i)}$$

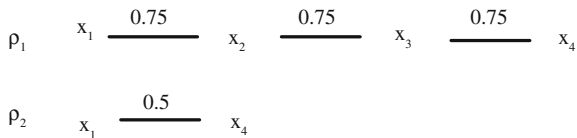
Clearly $L(\rho) \geq n$. We note that if there exists one $R(x_{i-1}, x_i) = 0$ then $L(\rho) = \infty$ and $St(\rho) = 0$. We note that if R is crisp and $St(\rho) \neq 0$ then $L(\rho) = n$. Using this idea we can define the distance between two nodes x and y in the graph G as

$$\delta(x, y) = \mathbf{Min}_{\text{all paths } x \text{ to } y} [L(\rho)]$$

It is the length of the shortest path from x to y . It can be shown that δ is a metric [17], $\delta(x, x) = 0$, $\delta(x, y) = \delta(y, x)$ and $\delta(x, z) \leq \delta(x, y) + \delta(y, z)$. Using the terminology common in social network analysis [1, 2] we can refer to the path ρ such that $L(\rho) = \delta(x, y)$ as the geodesic between x and y and refer to $\delta(x, y)$ as the geodesic distance.

While there appears to be some inverse connection between strength of a path and its length as for example in the case where $ST(\rho) = 0$ implies $L(\rho) = \infty$ this is not a strict correlation. Consider for example the two paths ρ_1 and ρ_2 shown in Fig. 1. We see that $ST(\rho_1) = 0.75 > Str(\rho_2) = 0.5$. On the other hand $L(\rho_1) = \frac{4}{3} + \frac{4}{3} + \frac{4}{3} = 4 \geq L(\rho_2) = 1$

Fig. 1 Two paths



Let $\langle X, R \rangle$ be a weighted graph. At times it can be useful to view this from the level set point of view [20]. This will allow us to make use of the representation theorem [20] to extend operations to these fuzzy relationships. We recall that if R is a fuzzy relationship on $X \times X$ then $R_\alpha = \{(x, y)/R(x, y) \geq \alpha\}$ is called the α level set of R . We see that each R_α is a crisp relationship on X .

We note that if R^k is the k composition of R then $R^k_\alpha = \{(x, y)/R^k(x, y) \geq \alpha\}$. It can also be shown that [21]

$$R^k_\alpha = R_\alpha \blacklozenge R_\alpha \blacklozenge \dots \blacklozenge R_\alpha,$$

the k composition of R_α is also R^k_α .

The representation theorem allows us to represent fuzzy relationship in terms of the collection of its level sets. This can be used to extend operations that are well defined on crisp sets to be defined on fuzzy sets. Using this we can express

$R^k = \bigcup_{\alpha=0}^1 \{R^k_\alpha\}$, here we are using the standard fuzzy set notation where $\frac{\alpha}{R^k_\alpha}$ indicates that α is the membership grade of R^k_α .

Undirected fuzzy graphs, which are also transitive, provide a very interesting class of graph. In these graphs if x is related to y and y is related to z then x is related to z . In social networks transitivity captures the property “friend of a friend is a friend”.

Many of the concepts introduced in the preceding are valid for both directed and undirected graphs. A fundamental difference is the following. Assume $x_0x_1 \dots x_n$ is a sequence of points that constitute a path. In an undirected graph its transpose $x_nx_{n-1} \dots x_0$ is a path having the same strength and length. In a directed graph we can't say anything about the transpose sequence. Whenever possible in an undirected graph we shall refer a path as between x_0 and x_n while in a directed graph we shall refer to the path from x_0 to x_n .

3 Computing with Words

Our goal here is to extend our capabilities for analyzing social relational networks by associating with these networks concepts with which human beings view social network relationships in such a way that they are comprehensible to both humans and machines. On one hand human beings predominantly use linguistic terms in which to communicate, reason and understand the world. Machines on the other hand require much more formal symbols. One of the most useful approaches to providing a

bridge between man and machine comprehension is the general framework provided by granular computing [8] and more specifically Zadeh’s fuzzy set based paradigm of computing with words. This technology allows for a high level of man-machine cooperation by providing a framework in which concepts can be modeled in a manner amenable to both. The potential for applying fuzzy set based technologies within the domain of social network analysis is particularly promising given that the computer modeling of these networks is in terms of mathematical relationships which as we already noted are equivalent to fuzzy sets.

In the following we introduce some ideas from the fuzzy set based approach to computing with words. Let U be some attribute that takes its value in the space Y . An example of such an attribute is age, which, for human beings takes its value in the set $Y = \{0, \dots, 100\}$. A fundamental concept in computing with words is the idea of linguistic value [22]. A linguistic value is some word used to express the value of U . In the case of age, some examples of linguistic values are “old”, “young”, “about 30”. A linguistic value can be seen as a granule, it is a collection of values from the space Y . As we noted it is with the aid of linguistic values that human beings best understand and reason about their environment.

By a vocabulary we shall mean a collection of commonly understood words that are used to express the linguistic values associated with an attribute. These are the words that people use to communicate with each other. They are also the words people use to reason with about the attribute. Fuzzy sets provide a useful tool for formalizing the idea of a vocabulary in a way that allows machine computation and understanding. If W is a word in the vocabulary associated with the variable U we can express W as a fuzzy subset W of the domain of U . Here then for any element $y \in Y$ the membership grade of y in W , $W(y)$, indicates the compatibility of the value y with the linguistic value W . Thus the fuzzy subset W provides a machine comprehensible interpretation of the word W .

We are now in a position to bridge the gap in man-machine communication with respect to the analysis of social relational network by allowing the human to build a vocabulary of linguistic terms associated with an attribute and then provide a representation for these terms by fuzzy sets (see Fig. 2). Thus here now we have a communal vocabulary coherent to both the human and the machine.

What must be emphasized here is that the choice of the vocabulary as well as the associated meaning of the words in terms of fuzzy sets is completely in the hands of the human partner. This greatly simplifies this task. The vocabulary that will be used is imposed, we are giving the computer meaning. Particularly noteworthy is the fact that learning algorithms need not necessarily be required, the computer is told these are the terms I will be using and this is what they mean in your language, fuzzy

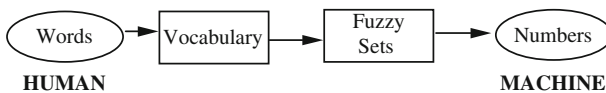
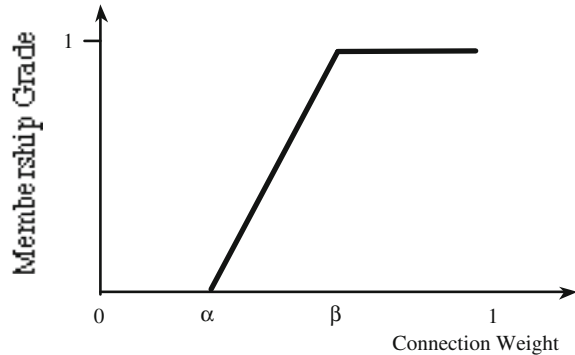


Fig. 2 Paradigm of man-machine understanding

Fig. 3 Fuzzy subset of term *strong*



sets. This is not to say that the construction of the communal vocabulary is not an important and complex task to which future research must be dedicated so that it is thoughtfully and appropriately done but only to reflect the fact that it is within our power to impose what we have decided. This situation allows us in the following to assume the availability of a communal vocabulary associated with attributes of interest.

In analyzing weighted relational networks there are a number of attributes about which it will be useful to have a vocabulary of commonly accepted terms. One such attribute is strength of connection. This is an attribute whose domain is the unit interval, $I = [0, 1]$. Terms like *strong*, *weak*, *none* would be part of a vocabulary associated with this attribute. In this case we would define the word *strong* as a fuzzy subset S of $[0, 1]$ such that for any $y \in [0, 1]$ the value $S(y)$ would indicate the degree to which y satisfies the working definition of the concept *strong* connection. We would assume that S would be such that $S(0) = 0$, $S(1) = 1$ and S is monotonic, $S(y_1) \geq S(y_2)$ if $y_1 \geq y_2$. A prototypical example of the definition of the term *strong* would be the piecewise linear fuzzy subset shown in Fig. 3.

Another attribute for which it will be useful to have a communal vocabulary is the number of links in a path, path length. Some words in such a vocabulary would be *short* and *long*. In the case of this attribute we would provide a semantics for the words of the vocabulary, in terms of fuzzy subsets of the domain $H = \{0, 1, \dots, n\}$ when n is the number of vertices in the network.

Concepts, in addition to actual physical objects, can provide attributes of interest. One such concept that we shall find useful are proportions. Here U is an attribute that takes its value in the set $I = [0, 1]$ where $r \in [0, 1]$ is a proportion. Examples of linguistic values that could be part of a communal vocabulary associated with this attribute are *many*, *most* and *about half*. As noted by Zadeh these terms provide a generalization of the quantifiers “all” and “none” that are used in logic. We can refer to these as linguistic quantifiers.

4 Clusters and Cliques

An important idea in classical graph theory is the concept of a cluster; here we want to extend this to weighted graphs. Let us first review the ideas from crisp graph theory. Let $\langle X, R \rangle$ be a graph where R is a crisp relationship. We are implicitly assuming our graph is undirected so R is symmetric. One approach is to call a subset C of X a cluster of order k if

- (a) For all node pairs x and y in C we have $R^k(x, y) = 1$
- (b) For all nodes $z \notin C$ there is some $x \in C$ such that $R^k(x, z) = 0$.

Note: When $k = 1$ we call C a clique.

Note: The order k of the cluster is reflected in the term R^k .

In [17] Rosenfeld suggested extending these ideas to fuzzy graphs. Here we let $\langle X, R \rangle$ be a fuzzy graph where R is a symmetric fuzzy relationship. A crisp subset $C \subset X$ is called a fuzzy cluster of order k if

$$\text{Min}_{x,y \in C} [R^k(x, y)] > \text{Sup}_{z \notin C} [\text{Inf}_{w \in C} R^k(w, z)].$$

In the following we suggest an alternative, more human meaningful, definition of a clique and then using the paradigm of computing with words we can provide a procedure for evaluating how well a subset of nodes satisfies our definition.

Let A be a subset of elements from X . We shall define A to be a clique if

- C1: All elements in A are connected by a **short strong** path.
- C2: No element not in A is connected to an element in A by a **strong** path.

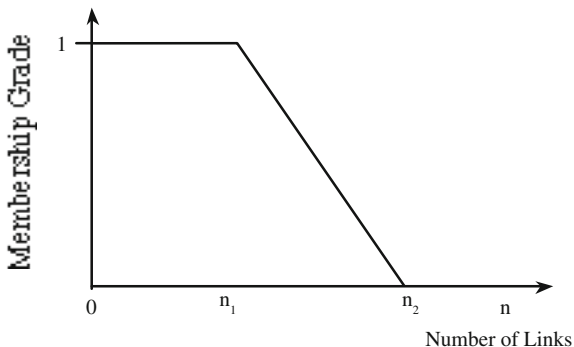
Here then we have two criteria that need to be satisfied for a subset $A \subset X$ to be considered as a clique, C_1 and C_2 . If we let $C_1(A) \in [0, 1]$ be the degree to which it is true that A satisfies C_1 and $C_2(A) \in [0, 1]$ be the degree to which it is true that A satisfies C_2 then $C(A) = \text{Min}[C_1(A), C_2(A)]$ is the degree to which it is true that A is a clique.

We must now formulate the two criteria in terms of features from the network $\langle X, R \rangle$. Here we shall make use of the communal vocabularies for the attributes strength of connection and path length. We shall assume the availability of the word **strong** for strength of connection and **short** for length of path in the vocabulary, that is we have expressions for the meaning, semantics, of these words as fuzzy subsets.

We first focus on C_1 . Here we will make use of the term “strong connection” which we will represent as a fuzzy subset S of the unit interval I . In addition we need use the linguistic term “short path” which we represent as a fuzzy subset SH of the space N .

Assume x_i and x_j are two nodes in the proposed clique A . Here $R^k(x_i, x_j)$ indicates the strength of connection of x_i and x_j for a path of at most k links. For any $R^k(x_i, x_j)$ the value $S(R^k(x_i, x_j))$ is the degree to which $R^k(x_i, x_j)$ is a strong connection. We recall that $R^k(x_i, x_j)$ is monotonic in k , that is if $k_2 > k_1$ then $R^{k_2}(x_i, x_j) \geq R^{k_1}(x_i, x_j)$. In addition as we noted earlier S , strong connection, is also monotonic

Fig. 4 Fuzzy set representation of short path



in strength, $S(a) \geq S(b)$ if $a > b$. From this we can conclude that for $k_2 > k_1$ then $S(R^{k_2}(x_i, x_j)) \geq S(R^{k_1}(x_i, x_j))$.

The concept short path with respect to the number of links it contains can be defined as a fuzzy subset SH of $N = \{1, 2, \dots, n\}$. Here we can observe that generically SH should have the following properties; $SH(1) = 1$, $SH(n) = 0$ and $SH(k_1) \geq SH(k_2)$ if $k_1 < k_2$. Here it is monotonic decreasing in k , increasing k leads to decrease the satisfaction. A prototypical example of this concept is the piecewise linear fuzzy subset as shown in Fig. 4.

Using these ideas we can determine the degree to which there exists a short-strong connection between x_i and x_j . We define this using a form of the Sugeno integral [23] as

$$C_1(x_i, x_j) = \text{Max}_{k=1 \text{ to } n} [SH(k) \wedge S(R^k(x_i, x_j))]$$

In the above \wedge is the Min operator. We see that as k increases $SH(k)$ tends to get smaller while $S(R^k(x_i, x_j))$ tends to get bigger.

We now can use this to determine the degree to which all elements in A are connected by a short-strong path,

$$C_1(A) = \text{Min}_{\substack{x_i, x_j \in A \\ x_i \neq x_j}} [C_1(x_i, x_j)].$$

It is very important to observe the marriage of different types of set objects used in making up the definition of $C_1(x_i, x_j)$. We first see that we have used $R^k(x_i, x_j)$, which is essentially the set-based definition of the network. In addition we have used the fuzzy sets SH and S, which are the fuzzy set definitions of the words short and strong. Here we have taken advantage of the fact that both the basic representation of a graph (network) and the meaning of words can be formulated using a set-based formulation to mix language and network representation. In [8] we referred to this approach as PISNA, the **Paradigm for Intelligent Social Network Analysis**.

We now consider the second criteria, C_2 , no element not in A has a strong connection with an element in A . In the following we shall let $x \in A$ and $z \notin A$. For

these elements $R^n(x, z)$ is the strength of the strongest path between x and z of any length. Here then $S(R^n(x, z))$ is the degree to which there exists a strong path between x and z . If we calculate $\text{Max}_{x \in A, z \notin A} [S(R^n(x, z))]$ we obtain the degree to which there exists a strong path between an element in A and an element not in A . Using this can calculate $C_2(A)$ as

$$C_2(A) = 1 - \text{Max}_{x \in A, z \notin A} [S(R^n(x, z))]$$

We now can use these formulations to determine the degree to which a subset A is a clique.

In the following we will make some observations regarding the preceding approach to defining the concept of clique.

In the preceding we defined the concept of short path in an absolute way as a fuzzy subset SH of the set of number of elements in a graph. It is possible to express short in a more universal way as a subset of the proportion of the number of elements in the graph. Thus here we can define “short path” as a fuzzy subset SH_p defined on the unit interval where for any $r \in [0, 1]$ the value $SH_p(r)$ indicates the degree to which the proportion r of elements in the network constitute a “short path.” Thus here we are defining “short path” as a proportion of number of vertices in the network. This allows us to have a universal definition of the concept of “short path” independent of the number of nodes in the network. In this case where we have SH_p we calculate

$$C_1(x_i, x_j) = \text{Max}_{k=1 \text{ to } n} \left[SH_p \left(\frac{k}{n} \right) \wedge S(R^k(x_i, x_j)) \right]$$

We note that SH_p will have the same form as SH , $SH_p(r)$ decreases as r increases.

We also note that it is possible to somewhat relax the second criteria. Instead of having

C_2 = no element, not in A , is connected to an element in A by a strong path.

We can say that

C_2 = no element not in A is connected to an element in A by a strong path that is not long.

Here we need obtain the word **Long** from our communal vocabulary. A typical example of a fuzzy subset representing such a definition is shown in Fig. 5.

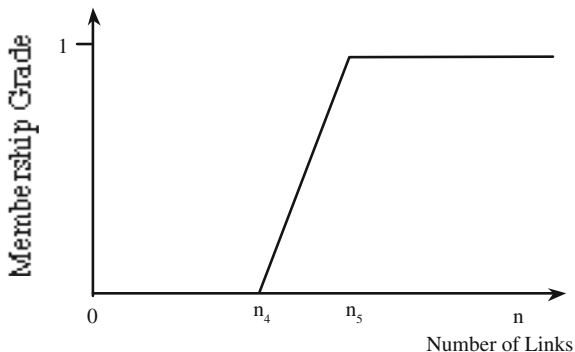
If L is the fuzzy subset defined on N corresponding to Long then not Long is a fuzzy subset $N.L$ defined such that $N.L(k) = 1 - L(k)$. Using this we obtain

$$\text{Max}_{k=1 \text{ to } n} [N.L(k) \wedge S(R^k(x, z))]$$

as the degree to which there is a strong and not long path between x and z . Finally from this we obtain

$$C_2(A) = 1 - \text{Max}_{x \in A, z \notin A} \left(\text{Max}_{k=1 \text{ to } n} [N.L(k) \wedge S(R^k(x, z))] \right)$$

Fig. 5 Fuzzy set representing term Long



We now consider the situation in what we allow the set A to be a fuzzy set. Thus we want to determine if A is a fuzzy clique. We shall consider the same two original conditions C_1 and C_2 as defining a clique. We shall use \tilde{A} to indicate our fuzzy Clique.

Here we first look at C_1 . In this case we still get for any two nodes x and $y \in X$.

$$C_1(x, y) = \text{Max}_{k=1 \text{ to } n} [SH(k) \wedge S(R^k(x, y))]$$

as the degree to which there is a short and strong path between the nodes x and y . Using this we obtain

$$C_1(\tilde{A}) = \text{Min}_{x \neq y \in X} [(1 - \tilde{A}(x)) \vee (1 - \tilde{A}(y)) \vee C_1(x, y)]$$

We note that in the case where \tilde{A} is crisp this reduces to our original definition of C_1 . For if x or y are in \tilde{A} then the disjunction reduces to for $C_1(x, y)$ while if either x or y is not in \tilde{A} then the argument becomes 1.

In this case of C_2 , for a fuzzy clique we get

$$C_2(\tilde{A}) = 1 - \text{Max}_{\substack{x, z \in X \\ x \neq z}} [\tilde{A}(x) \wedge (1 - \tilde{A}(z)) \wedge S(R^n(x, z))]$$

We note that if \tilde{A} is crisp this reduces to our previous definition. For if $x \in \tilde{A}$ and if $z \notin \tilde{A}$ then it becomes the original format. If either $x \notin \tilde{A}$ or $z \in \tilde{A}$ then $\tilde{A}(x) \wedge (1 - \tilde{A}(z)) \wedge S(R^n(x, z)) = 0$ and it doesn't effect the calculation of $C_2(\tilde{A})$.

5 Centrality

An important concept in social network analysis is centrality [1, 2]. The centrality of a node is closely related to its importance in the network. Assume $\langle X, R \rangle$ is a relational network where R is a crisp relation. The measure of centrality of node x_i is the number of nodes connected to it by at most k links. In this case

$$C^k(x_i) = \sum_{\substack{j=1 \\ j \neq i}}^n R^k(x_i, x_j)$$

is the measure of the centrality of node x_i .

If we have a network where R is a weighted graph a straightforward way to extend this is to calculate $C^k(x_i)$ as in the above. One problem that can arise here is that a large number of weak connections, small values of $R^k(x_i, x_j)$, can add up to appear as a strong connection. Here we can suggest some other alternative methods for obtaining this measure of centrality.

One method is to use the level set representation and obtain a fuzzy set representation for the centrality. Here we can express

$$C_\alpha^k(x_i) = \sum_{\substack{j=1 \\ j \neq i}}^n R_\alpha^k(x_i, x_j)$$

Thus here $C_\alpha^k(x_i)$ is the number of nodes connected to x_i with strength of at least α using at most k links. Using this we can define

$$\tilde{C}^k(x_i) = \bigcup_{\alpha \in [0,1]} \left\{ \frac{\alpha}{C_\alpha^k(x_i)} \right\},$$

here we are using the standard fuzzy notation indicating that α is the membership grade of $C_\alpha^k(x_i)$.

In this case $\tilde{C}^k(x_i)$ is a fuzzy number. We should note that since for $\alpha_2 > \alpha_1$ we have $R_{\alpha_2}^k \subseteq R_{\alpha_1}^k$ then $C_{\alpha_1}^k \supseteq C_{\alpha_2}^k$.

Let us consider another way to extend the concept of centrality to the case of a weighted graph. An alternative and perhaps more appropriate definition of centrality would be the “number of **strong** connections using at most k links.” Here we shall define the concept “strong” as a fuzzy subset of unit interval. For example see Fig. 3. Using this definition we can obtain

$$\tilde{C}^k(x_i) = \sum_{\substack{j=1 \\ j \neq i}}^n \text{Strong } R^k(x_i, x_j)$$

Thus here we transform the scores via the concept strong.

6 Social Network Databases

In the following we shall consider a weighted network $\langle X, R \rangle$ where each of the nodes has an associated vector of attribute (feature) values. In these types of networks each of the node objects have various attributes, properties and features. This structure can be viewed as the combination of a network and database.

In these networks we have a collection of q attributes U_1, \dots, U_q . In the case where the nodes are people, examples of attributes could be nationality, age or income. Each of the attributes takes a value from a domain Y_i . In this situation, each node has an associated q vector whose i th component is the value of the i th attribute for that node.

We shall use the notation $U_i(x_j)$ to indicate the variable corresponding to the attribute U_i in the case of node x_j . For example with U_i being the attribute age then $U_i(x_j)$ would indicate the variable age of x_j . If U_i is the attribute country of birth then Y_i would be a list of countries and $U_i(x_j)$ would be the variable corresponding the country of birth of x_j . We shall let v_{ij} indicate the value of the variable $U_i(x_j)$, thus $U_i(x_j) = v_{ij}$. Thus in this case any node in our network has an associated vector V_j whose i component v_{ij} corresponds to the value of attribute U_i for node x_j . We should observe the above network could in some ways be viewed as a kind of database.

In the following we shall begin to describe techniques that can be used to analyze, investigate and question networks with vector-valued nodes. Here we shall be using flexible/fuzzy-querying techniques [9, 24].

In the following we shall assume that country of residence is one of the attributes, we shall denote this as U_1 . Thus $U_1(x_j)$ is the variable denoting the country of residence of x_j . In this case the domain Y_1 associated with U_1 is the set of countries. In addition, a communal vocabulary associated with this attribute would consist of terms such as “Middle East”, “North America”, “South America” and “Southeast Asia”. Other terms such as “mountainous country”, “Spanish speaking”, and “Oil producing” can also be part of the vocabulary. Each of the terms in the vocabulary would be defined in terms of subsets of Y_1 . Some of these terms can be defined using crisp subsets while others will require fuzzy subsets.

In addition we shall assume age is another of the attributes associated with the network nodes. We shall denote this as U_2 with its domain Y_2 being the set of non-negative integers. Here we shall also assume the availability of mutually understandable vocabulary of commonly used linguistic terms to describe a person’s age. These terms will be defined in terms of subsets of Y_2 .

Assume x_j is some node in our network. We can ask “To what extent is x_j strongly connected to a person residing in South America?” In the following we shall let SA indicate the subset of Y_1 corresponding to South America. Using this we can obtain as the answer to our question

$$\text{Max}_{i, i \neq j} [\text{SA}(U_1(X_i)) \wedge R^n(x_i, x_j)]$$

More specifically we can ask “To what extent is x_j strongly connected to a young person residing in South America?” In this case with U_2 being the attribute age and Young being the subset corresponding concept young person we get as the answer

$$\text{Max}_{i,i \neq j} [\text{SA}(U_1(X_i)) \wedge \text{Young}(U_2(X_j)) \wedge R^n(x_i, x_j)]$$

We note this value is also the truth of the question “Does x_j have a connection to a Young South American”.

We note that if we want to find out “does x_j have a *strong* connection to a Young South American” then we obtain the truth of this as

$$\text{Max}_{i,i \neq j} [\text{SA}(U_1(X_i)) \wedge \text{Young}(U_2(X_j)) \wedge \text{Strong}(R^n(x_i, x_j))]$$

Here we have replaced the predicate $R^n(x_i, x_j)$ with Strong ($R^n(x_i, x_j)$)

A related question is the following. Let B be some crisp subset of X . We now ask what is the strongest connection between an element in B and a Young South American not in B . The answer is then obtained from the following

$$\text{Max}_{X_i \in B} \left[\text{Max}_{x_j} [\text{SA}(U_1(X_i)) \wedge \text{Young}(U_2(X_j)) \wedge R^n(x_i, x_j) \wedge \bar{B}(X_j)] \right]$$

If in the above we are interested in only direct connections rather than any connection we replace R^n by R .

We now consider the question: Do all people residing in South America have a strong connection with each other? We shall denote the truth of this question $\text{Tr}[Q]$ We calculate this truth-value as

$$\text{Tr}(Q) = \text{Min}_{x_i, x_j \in X} [(1 - \text{SA}(U_1(x_i))) \vee (1 - \text{SA}(U_1(x_j))) \vee \text{Strong}(R^n(x_i, x_j))]$$

Let us look at this for the special case where SA is a crisp set. We first see that in the case of a pair (x_i, x_j) in which at least one of the elements do not reside in South America then either $\text{SA}(U_1(x_i)) = 0$ or $\text{SA}(U_1(x_j)) = 0$ and therefore

$$(1 - \text{SA}(U_1(x_i))) \vee (1 - \text{SA}(U_1(x_j))) \vee \text{Strong}(R^n(x_i, x_j)) = 1.$$

This case will not be the min. For the case in which both x_i and x_j reside in South America then $\text{SA}(U_1(x_i)) = \text{SA}(U_1(x_j)) = 1$ and hence

$$(1 - \text{SA}(U_1(x_i))) \vee (1 - \text{SA}(U_1(x_j))) \vee \text{Strong}(R^n(x_i, x_j)) = \text{Strong}(R^n(x_i, x_j))$$

From this we get as expected $\text{Tr}(Q) = \text{Min}_{x_i, x_j \in \text{SA}} [\text{Strong}(R^n(x_i, x_j))]$

Now we shall consider the slightly more complicated question of whether “*most* of the people residing in *western* countries have *strong* connections with each other?”

We shall here assume the term *most* is available in our common vocabulary where it is defined as a fuzzy subset over the unit interval. In addition we shall assume that the concept western country, is a concept that is defined by the fuzzy subset W over the domain Y_1 . In this case for each $x_i \in X$ we have $W(U_1(x_i))$ indicates the degree to which it is true that x_i is from a western country. In the following we shall set P to be the set of all unordered pairs of distinct elements from X . P is the set of all the subsets of X consisting of two elements. We see that the number of elements in P , $n_p = \frac{(n)(n-1)}{2}$ where n is the number of elements in X . We shall denote an element $\{x_i, x_j\}$ in P as p_k . Here then k goes from 1 to n_p .

For each pair $p_k = \{x_i, x_j\}$ we obtain two values. The first $V_k = \text{Min}(W(U_1(x_i)), W(U_1(x_j)))$ indicates the degree to which p_k consists of pair of elements both from a western country. The second value is $S_k = \text{Strong}(R^n(x_i, x_j))$, indicates the degree to which there is a strong connection between the pair $\{x_i, x_j\}$. We shall use the technology of OWA operators to answer our question [25]. We proceed to obtain the answer as follows:

- (1) Order the S_k and let $\text{ind}(j)$ be the index of the j th largest of the S_k . Thus here $S_{\text{ind}(j)}$ is the j th largest of the S_k and $V_{\text{ind}(j)}$ is its associated V value.
- (2) We next calculate

$$R = \sum_{j=1}^{n_p} V_{\text{ind}(j)}$$

- (3) We next obtain a set of weight w_j for $j = 1$ to n_p where

$$w_j = \text{Most}\left(\frac{R_j}{R}\right) - \text{Most}\left(\frac{R_j - 1}{R}\right)$$

here $R_j = \sum_{i=1}^j V_{\text{ind}(i)}$

- (4) We finally calculate the truth of the question as

$$\text{Tr}(q) = \sum_{j=1}^{n_p} w_j S_{\text{ind}(j)}$$

An interesting special case of the preceding occurs if the subset W , Western Country, is a crisp set. In this case $V_k = \text{Min}(W(U_1(x_i)), W(U_1(x_j)))$ is a binary value, either one or zero. It is one if both x_i and x_j are from western countries and zero if either is not from a western country.

Another question we can ask is whether the young people form a clique. Since the young people provide a fuzzy subset over the space X and we have previously indicated a process for determining whether a fuzzy set is a clique we can answer this question.

7 Conclusion

We discussed the idea of fuzzy relationships and their role in modeling weighted social relational networks. The paradigm of computing with words was introduced and the role that fuzzy sets play in representing linguistic concepts was described. We discussed how these technologies can provide a bridge between a network analyst's linguistic description of social network concepts and the formal model of the network. Some examples of taking an analyst's network concepts and formally representing them in terms of network properties were provided. We applied this to the concept of clique and then to the idea of node centrality. Finally we introduced the idea of vector-valued nodes and began developing a technology of social network database theory. Clearly this newly introduced idea of social network database theory will provide many applications and will need a more formal mathematical framework, a task for future research.

References

1. Wasserman, S., Faust, K.: *Social Network Analysis: Methods and Applications*. Cambridge University Press, New York (1994)
2. Scott, J.: *Social Network Analysis*. SAGE Publishers, Los Angeles (2000)
3. Newman, M.: *Networks: An Introduction*. Oxford University Press, New York (2010)
4. Russell, M.A.: *Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites*. O'Reilly Media, Schasptopol, CA (2011)
5. Aggarwal, C.C.: *Social Network Data Analytics*. Springer, New York (2011)
6. Prell, C.: *Social Network Analysis: History, Theory and Methodology*. Sage Publisher, London (2012)
7. Kadushin, C.: *Understanding Social Networks: Theories, Concepts and Findings*. Oxford University, New York (2012)
8. Yager, R.R.: Intelligent social network analysis using granular computing. *Int. J. Intell. Syst.* **23**, 1196–1219 (2008)
9. Pivert, O., Bosc, P.: *Fuzzy Preference Queries to Relational Databases*. World Scientific, Singapore (2012)
10. Zadeh, L.A.: Fuzzy logic = computing with words. *IEEE Trans. Fuzzy Syst.* **4**, 103–111 (1996)
11. Zadeh, L.A.: From computing with numbers to computing with words—from manipulation of measurements to manipulations of perceptions. *IEEE Trans. Circuits Syst.* **45**, 105–119 (1999)
12. Zadeh, L.A.: Outline of a computational theory of perceptions based on computing with words. In: Sinha, N.K., Gupta, M.M. (eds.) *Soft Computing and Intelligent Systems*, pp. 3–22. Academic Press, Boston (1999)
13. Zadeh, L.A.: Generalized theory of uncertainty (GTU)-principal concepts and ideas. *Comput. Stat. Data Anal.* **51**, 15–46 (2006)
14. Zadeh, L.A.: Fuzzy logic. In: Meyers, A.R. (ed.) *Encyclopedia of Complexity and Systems Science*. Springer, Heidelberg (2009)
15. Bargiela, A., Pedrycz, W.: *Granular Computing: An Introduction*. Kluwer Academic Publishers, Amsterdam (2003)
16. Yager, R.R.: Human behavioral modeling using fuzzy and Dempster-Shafer theory. In: Liu, H., Salerno, J.J., Young, M.J. (eds.) *Social Computing, Behavioral Modeling and Prediction*, pp. 89–99. Springer, Berlin (2008)

17. Rosenfeld, A.: Fuzzy graphs. In: Zadeh, L.A., Fu, K.S., Tanaka, K., Shimura, M. (eds.) *Fuzzy Sets and their Applications to Cognitive and Decision Processes*, pp. 77–97. Academic Press, New York (1975)
18. Delgado, M., Verdegay, J.L., Vila, M.A.: On the valuation and optimization problems in fuzzy graphs (a general approach and some particular cases). *ORSA J. Comput.* **2**, 74–83 (1990)
19. Koczy, L.T.: Fuzzy graph in the evaluation and optimization of networks. *Fuzzy Sets Syst.* **46**, 307–319 (1992)
20. Yager, R.R.: Level sets and the extension principle for interval valued fuzzy sets and its application to uncertainty. *Inf. Sci.* **178**, 3565–3576 (2008)
21. Zadeh, L.A.: Similarity relations and fuzzy orderings. *Inf. Sci.* **3**, 177–200 (1971)
22. Zadeh, L.: The concept of a linguistic variable and its application to approximate reasoning: Part 1. *Inf. Sci.* **8**, 199–249 (1975)
23. Murofushi, T., Sugeno, M.: Fuzzy measures and fuzzy integrals. In: Grabisch, M., Murofushi, T., Sugeno, M. (eds.) *Fuzzy Measures and Integrals*, pp. 3–41. Physica-Verlag, Heidelberg (2000)
24. Zadrozny, S., de Tré, G., de Caluwe, R., Kacprzyk, J.: An overview of fuzzy approaches to database querying. In: Galindo, J. (ed.) *Handbook of Research on Fuzzy Information Processing in Databases*, vol. 1, pp. 34–54. Information Science Reference, Hershey, PA (2008)
25. Yager, R.R.: Quantifier guided aggregation using OWA operators. *Int. J. Intell. Syst.* **11**, 49–73 (1996)