Current Flows in Electrical Networks for Fuzzy Social Network Analysis (FSNA)

Alejandro Gutiérrez-Muñoz and Abraham Kandel (IEEE Fellow)
{agutierrez,kandel}@cse.usf.edu

Department of Computer Science and Engineering, University of South Florida
4202 E. Fowler Ave. Tampa, FL 33620-9951, USA

***Abstract*—Connections among different entities of a social network can be extracted based on social interaction activity records. Activity records present an inherent level of uncertainty when ranking methods are introduced to rate the importance of each record based on a predetermined scale. With time, activity records may become stale and cluttered with non-relevant information. The extraction of key actors on a social network, and the ability of identifying primary interaction routes between entities of a network is of paramount importance in social network analysis (SNA). In this paper, we introduce a methodology that incorporates into the social interaction activity records the uncertainty and time sensitiveness of the events through Fuzzy Social Networks Analysis (FSNA). Also, we investigate an approach based on the analysis of current flows in electrical networks for the extraction of primary routes of interaction among key actors in a social network.**

***Index Terms*—fuzzy graphs, social network analysis, current flows, FSNA, fuzzy sets.**

I. Introduction

The use of social networks for the representation of human social interaction has become one of the most important approaches for the analysis of human activities in a network that models interactions among its entities. Social networks are usually represented as graphs, where actors are nodes, and the relationships between entities are edges of the graph. Several methods for the construction of social network graphs have been proposed over the last two decades [1]. In [2], a Fuzzy Social Network Graph (FSNG) representation based on a consolidation operation that combines activity records between entities has been proposed. In this paper, we investigate a methodology for the construction of a fuzzy social network which incorporates the uncertainty introduced to the network by a raking system of the activity records. Also, we introduce a time sensitive function that allows for the evaluation of activity records based on relevancy (how current the records are relative to the evaluation date) and the length of time that a piece of information is considered relevant before it becomes stale.

The analysis and comparison of graphs representing social networks, when the size of the graph is too large, can prove to be computationally complex. Several error-tolerant graph matching techniques for very large graphs have been proposed [3], [4], [5], [6], [7], [8]. Error-tolerant graph matching techniques offer suboptimal solutions to the subgraph isomorphism problem, and some of them are computational feasible for very large graph structures. Here, we investigate an approach that uses current flow analysis in electrical networks to identify key actors on a social network, and primary interaction routes between the entities.

II. Related Work

FSNGs provide the necessary tools to incorporate the uncertainty in the relationships between the entities of a social network. In [2], Nair and Sarasamma propose an approach to model fuzzy social network graphs. Their approach is based on the reduction of the size of a FSNG accomplished by combining connections between actors that represent activity records between the same pair of entities. A new fuzzy operation has been proposed: the consolidation operation, where the fuzzy graph is a fuzzy graph of type V [9], in this case the set of vertices and edges are crisp, but the edge weights are fuzzy. The consolidation operation is defined as follows: *If,* $\alpha, \beta \in [0,1]$*, then* $\otimes : [0,1] \times [0,1] \rightarrow [0,1]$ *is defined by* $\alpha \otimes \beta = \alpha + \beta - \alpha\beta$. Multiple edge weights between the same pair of entities are then consolidated as a single edge using the consolidation operation.

Current flow analysis has been discussed in [10], [11] for the extraction of dense subgraphs and is related to electrical currents in a network of resistors. This approach deals with the problem of finding the connection subgraph that can deliver as many units of electrical current as possible. For this purpose, a graph $\mathcal{G} = (V, E)$ is treated as an electrical network, where edge weights represent conductance: $C(u,v)$, and vertices represent the nodes of the electrical circuit. The voltages: $V(u)$, at each node of the circuit are calculated by combining Ohm's law and Kirchhoff's current law.

$$\forall u, v : I(u,v) = C(u,v)(V(u) - V(v)) \tag{1}$$

$$\forall v \neq s, t : \sum_u I(u,v) = 0 \tag{2}$$

Having *s* as the source node, and *t* as the target node, equations (1) and (2) determine the voltages and currents as the solution to the following linear system:

$$V(u) = \sum_v \frac{V(v)C(u,v)}{C(u)} \quad \forall u \neq s, t \tag{3}$$

$$V(s) = 1, V(t) = 0 \tag{4}$$

$$C(u) = \sum_{v} C(u,v) \tag{5}$$

Solving (3) with boundary conditions (4) determines the voltages at each node. $C(u)$ represents the total conductance of node u, this is, the sum of all edge weights adjacent to u. Once the values of the current, $I(u,v)$ are available, the current along a particular path: $\hat{I}(P)$ $P = \{s,\ldots,t\}$ is defined as the pro-rated current along that path from source to target.

$$\hat{I}(s,u) = I(s,u) \tag{6}$$

$$\hat{I}(s = u_1,\ldots,u_i) = \hat{I}(s = u_1,\ldots,u_{i-1})\frac{I(u_{i-1},u_i)}{I_{out}(u_{i-1})} \tag{7}$$

where $I_{out}(u) = \sum_{\{v|u\to v\}} I(u,v)$. $I_{out}(u)$ is the total current leaving a node, which is equal to the sum of all currents leaving the node in a downhill stream, where a downhill stream from node *u* to node *v* means that voltage at node *u* is higher than voltage at node *v*, $V(u) > V(v)$. Since the idea of the approach presented in [10], [11] is to find the best connection subgraph, the concept of capture flow $CF(\mathcal{H})$ is introduced. $CF(\mathcal{H})$ of a subgraph $\mathcal{H}$ of $\mathcal{G}$ is the total delivered current, summed over all paths from source *s* to target *t* that belong to $\mathcal{H}$.

III. FSNG Based on Activity Records

The construction of a FSNG based on activity records is done by using the information obtained through several sources and collected in an activity record journal, where interactions among entities is captured as well as rated for future reference. An activity record is defined as a tuple: $\langle date, description, importance\rangle$. *Date* represents the date of the event. *Description* represents the description of the event, including parties involved, places, etc. *Importance* represents a ranking of the event compared to others. An example of an activity records looks like this:

$\langle$*10/03/07, Phone call between entities A and B, High*$\rangle$.

We can clearly use a fuzzy linguistic variable [12] when describing the *importance* of the event. Different experts could rate similar events in different ways, at different levels of importance. We define *importance* as a fuzzy linguistic variable represented by the following fuzzy sets:

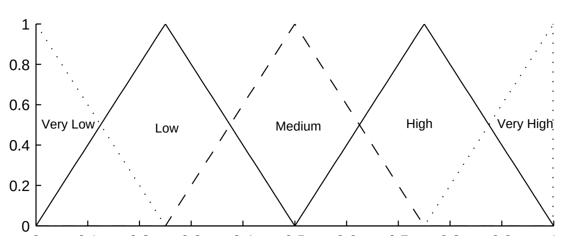

Figure 1. Importance variable

From Fig. 1, we observe that five fuzzy sets have been defined: very low, low, medium, high, and very high. Each of these fuzzy sets can be represented using a triangular fuzzy number. Table I shows the value of each of the fuzzy sets representing the *importance* variable. We use the L-R fuzzy number notation, where a fuzzy number is defined as a tuple $\langle m,\alpha,\beta\rangle$, where *m* is the mean value, α is the left spread, and β is the right spread [13].

Table I
L-R fuzzy numbers for *importance* ***variable***.

Fuzzy set	L-R fuzzy number
very low	(0.00,0.00,0.25)
low	(0.25,0.25,0.25)
medium	(0.50,0.25,0.25)
high	(0.75,0.25,0.25)
very high	(1.00,0.25,0.00)

Now that we have defined a fuzzy number for each of the possible values assigned to the *importance* variable, we can redefine our activity records as follows:

$\langle$*date, entities involved, importance*$\rangle \rightarrow$
$\langle$*10/03/07, [A,B] , (0.75,0.25,0.25)*$\rangle$

so far we have introduced an approach to include the inherent uncertainty of activity reports within our activity record format. But there is still a problem with the activity record definition. We are not taking into account the validity of the information based on how old the information is. Activity records of more than 5, or 10 years could be meaningless for our purposes. Therefore, we need a way to cope with the problem of information aging. Thus we propose a function that will modify the *importance* variable over time, in the form of

$$m = x_m \times e^{-\frac{days}{365\times Y}} \tag{8}$$

$$\alpha = \begin{cases} m & m - 0.25 < 0 \\ 0.25 & m - 0.25 \geq 0 \end{cases} \tag{9}$$

$$\beta = \begin{cases} 1-m & m + 0.25 > 1 \\ 0.25 & m + 0.25 \leq 1 \end{cases} \tag{10}$$

where *x* is a L-R fuzzy number, *m* is the mean value of the *x*, and the variable *days* represents the number of days that have passed since the event date. *Y* is a constant that represents the number of years during which the event will have some relevance. For example, assume the following activity records:

P:$\langle$*08/15/05, [A,B] , (0.75,0.25,0.25)*$\rangle$
Q:$\langle$*10/03/07, [A,B] , (0.25,0.25,0.25)*$\rangle$

Event *P* occurred back on 2005, and event *Q* is a more recent event. Event *P* has a *high* importance, while event *Q* has a *low* importance. We evaluate both events as of 11/01/2007, with a *Y* constant factor of 5 years. After applying equations (8), (9), and (10) to calculate the importance value as of 11/01/2007 we get:

P:$\langle$*08/15/05, [A,B] , (0.4817,0.25,0.25)*$\rangle$
Q:$\langle$*10/03/07, [A,B] , (0.2461,0.2461,0.25)*$\rangle$

As we can observe, event *P* after 808 days has an importance close to *medium*. While event *Q*, which is very recent, keeps its *low* importance.

Based on activity records, one can construct a graph representing the relationships between the entities involved in joined activities. The methodology we propose is to transform the original activity records, replacing the *importance* rating given by the expert, to a fuzzy linguistic variable as defined in Table I. Once all records are represented within the fuzzy rating framework, the *importance* assessment based on evaluation time, as defined by (8), (9), and (10) must be applied to each record on the activity log. With the new activity log reflecting the uncertainty and time sensitivity of the records, we can proceed to construct the FSNG.

The set of nodes of the FSNG is defined as the set of actors in the activity records. The set of edges of the graph represents how strong is the relation between actors based on the activity records. In order to calculate the weight of the edges, the *importance* factor, after it has been transformed, is recomputed for each record relating the same two actors, via equations (11) and (12) below:

$$Imp(a,b) = \sum_{\forall\, activity\; a,b} importance(a,b) \qquad (11)$$

$$\begin{gathered} \mathcal{G} = \langle V, E, w \rangle \\ V = \{x : x\, is\, an\, actor\} \\ E_{x,y} = \{\forall x \in V, y \in V:\; Imp(x,y) > 0\} \\ w(E_{x,y}) = EV(Imp(x,y)) \end{gathered} \qquad (12)$$

The edge weights as defined by (11) and (12) are defined as the expected value (EV) [12] of the sum of all the activity records between actors a and b. The expected value of a triangular fuzzy number $x = (a,b,c)$ is defined as $EV(x) = \frac{a+b+c}{3}$. Since we are using the *L-R* fuzzy number representation, the expected value (EV) of fuzzy number $x = (m, \alpha, \beta)$ is defined as:

$$EV(x) = m + \frac{\beta - \alpha}{3} \qquad (13)$$

We now present an example based on the activity log described in Table II. Table IV represents the connections (edge weights) between all of the actors based on the transformations to the *importance* factor represented in Table III. Figure 2 shows the FSNG extracted from Table IV.

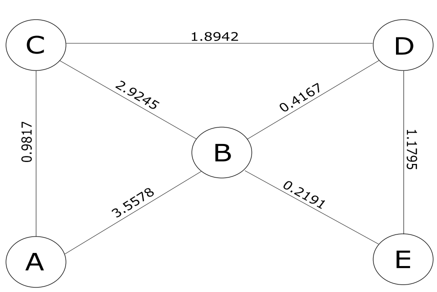

Figure 2. FSNG based on Activity Log

Table II
ACTIVITY LOG

Date	Actors	Event Description	Rating
01/05/05	A,B	Phone Call from A to B	High
02/10/05	B,C	$10,000 deposit from B to C	High
02/12/05	B,D	Phone call from B to D	Low
02/12/05	C,D	Fax from C to D	Very High
03/01/05	A,C	Phone Call from C to A	Very Low
09/10/05	A,B	$1,200 deposit from A to B	Medium
09/25/05	B,C	$500 deposit from B to C	Low
09/26/05	B,C	$200 deposit from B to C	Very Low
09/30/05	C,D	$800 deposit from C to D	Low
10/02/05	B,C	Phone Call from C to B	Very High
10/28/05	A,B	Phone Call from B to A	High
11/02/05	C,D	Phone Call from D to C	Medium
12/01/05	D,E	Phone Call from D to E	High
08/01/06	A,B	Phone Call from B to A	High
09/20/06	C,D	Fax from C to D	Very Low
10/23/06	B,E	Phone Call from B to E	Low
01/03/07	A,B,C	Meeting Between A,B, and C	Very High
05/04/07	C,D	Phone Call from D to C	Medium
06/01/07	B,D	Phone Call from D to B	Low
08/17/07	A,C	Phone Call from C to A	Very Low
09/30/07	A,B	$30,000 deposit from A to B	Very High
10/05/07	B,C	$10,000 deposit from B to C	High
10/20/07	C,D	Fax from C to D	Low
10/25/07	D,E	Phone Call from E to D	Medium

Table III
MODIFIED ACTIVITY LOG AS OF 11/01/2007

Days	Actors	Rating	Importance
1030	A,B	(0.75,0.25,0.25)	(0.4265,0.25,0.25)
994	B,C	(0.75,0.25,0.25)	(0.4350,0.25,0.25)
992	B,D	(0.25,0.25,0.25)	(0.1452,0.1452,0.25)
992	C,D	(1.00,0.25,0.00)	(0.5807,0.25,0.25)
975	A,C	(0.00,0.00,0.25)	(0.00,0.00,0.25)
782	A,B	(0.50,0.25,0.25)	(0.3257,0.25,0.25)
767	B,C	(0.25,0.25,0.25)	(0.1642,0.16421,0.25)
766	B,C	(0.00,0.00,0.25)	(0.00,0.00,0.25)
762	C,D	(0.25,0.25,0.25)	(0.1647,0.1647,0.25)
760	B,C	(1.00,0.25,0.00)	(0.6594,0.25,0.25)
734	A,B	(0.75,0.25,0.25)	(0.5016,0.25,0.25)
729	C,D	(0.50,0.25,0.25)	(0.3353,0.25,0.25)
700	D,E	(1.00,0.25,0.00)	(0.6814,0.25,0.25)
457	A,B	(0.75,0.25,0.25)	(0.58389,0.25,0.25)
407	C,D	(0.00,0.00,0.25)	(0.00,0.00,0.25)
374	B,E	(0.25,0.25,0.25)	(0.2037,0.2037,0.25)
302	A,B,C	(1.00,0.25,0.00)	(0.8475,0.25,0.1525)
181	C,D	(0.50,0.25,0.25)	(0.4528,0.25,0.25)
153	B,D	(0.25,0.25,0.25)	(0.2299,0.2299,0.25)
76	A,C	(0.00,0.00,0.25)	(0.00,0.00,0.25)
32	A,B	(1.00,0.25,0.00)	(0.9826,0.25,0.0174)
27	B,C	(0.75,0.25,0.25)	(0.7390,0.25,0.25)
12	C,D	(0.25,0.25,0.25)	(0.2484,0.2484,0.25)
7	D,E	(0.50,0.25,0.25)	(0.4981,0.25,0.25)

Table IV
EDGE WEIGHTS BASED ON TABLE 3

Actors		$Imp(a,b)$	$EV(Imp(a,b))$
A	B	(3.6679,1.5,1.1699)	3.5578
A	C	(0.8475,0.25,0.6525)	0.9817
B	C	(2.8451,1.1642,1.4025)	2.9245
B	D	(0.3751,0.3751,0.5)	0.4167
B	E	(0.2037,0.2037,0.25)	0.2191
C	D	(1.7818,1.163,1.5)	1.8942
D	E	(1.1795,0.5,0.5)	1.1795

Once the FSNG has been constructed based on the activity log, we can proceed with the calculation of the current flows. In the next section, we focus on the current flows calculation along two actors of the graph. The current flows between two actors, representing the flow of information, or resources, between these two actors based on the connections with other entities in the network.

IV. Current flow along two actors of the social network.

Now that we have investigated how to generate the FSNG, we can proceed with the current flow calculation along any pair of entities. Calculating currents is performed using nodal analysis as described in section II. Each relationship can be described as a path from source to target, $P = (s, \ldots, t)$. As noted by equation (4) voltages in source and target are given by $V(s) = 1, V(t) = 0$. In this scenario, the source will act as the voltage source, and the target as the ground. In the case that the graph has some nodes of degree 1 that are *not* the source nor the target, these nodes are considered to be grounds as well, hence the voltage at these nodes is $V(u) = 0$.

Once the voltage for source node and ground nodes have been specified, we can proceed to solve a system of linear equations with *n* variables, where *n* is equal to the total number of vertices of the graph, minus source and grounds nodes. This system can be reduced to solving an eigenvector calculation of the form

$$\begin{bmatrix} A & B \\ 0 & I \end{bmatrix} V = V \tag{14}$$

$$A = \left\{ a_{ij} = \frac{w_{ij}}{w_j} \right\} \tag{15}$$

where matrix *A* represents the relationship between the nodes based on their connection weights, *B* represents the boundary conditions for *s, t,* and other ground nodes, and *I* is the identity matrix. The solution to the system of equations represents the voltages at each of the *n* nodes. With the voltages for each node we can now calculate the current for each edge using equation (1). Once we have calculated the current along each edge, we need to calculate the current flow along each possible path between the two selected actors. This is done using equations (6) and (7). The current flow is a pro-rated amount from source to target based on the total current along each node in the path and the total current leaving each of the nodes along the path.

Based on the FSNG shown in Figure 2, and taking actors *A* and *E* as source and target, we have the following system of equations:

$$\begin{bmatrix} -7.1181 & 2.9245 & 0.4167 \\ 2.9245 & -5.8004 & 1.8942 \\ 0.4167 & 1.8942 & -3.4904 \end{bmatrix} V = \begin{bmatrix} -3.5578 \\ -0.9817 \\ 0.00 \end{bmatrix} \tag{16}$$

$$V(A) = 1 \quad V(E) = 0 \tag{17}$$

The above system of equations produces the following voltages:

$$V(B) = 0.84 \;\; V(C) = 0.76 \;\; V(D) = 0.51 \tag{18}$$

Calculating the currents along each edge using equation (1), the FSNG, represented by its voltages and currents is represented in Figure 3.

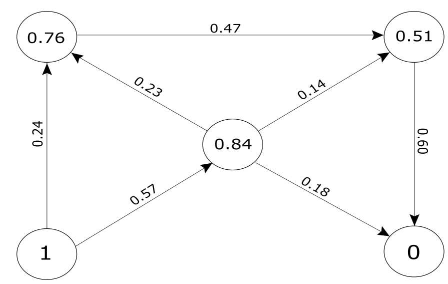

Figure 3. Voltages and Currents

Now we proceed to calculate the current along each path between *A* and *E*. It should be noted that based on the current flow only downhill paths can be calculated. A downhill path from node *u* to node *v* means that voltage at node *u* is higher than voltage at node *v*, $V(u) > V(v)$:

$A \to B \to E = 0.19$
$A \to C \to D \to E = 0.24$
$A \to B \to D \to E = 0.14$
$A \to B \to C \to D \to E = 0.24$

In [10], [11], the concept of captured flow $CF(\mathcal{H})$ is introduced. $CF(\mathcal{H})$ of a subgraph $\mathcal{H}$ of $\mathcal{G}$ is the total delivered current, summed over all paths from source *s* to target *t* that belong to $\mathcal{H}$. Here, we select the path that delivers the most current flow. In the case of several paths that deliver the same current flow we prefer those where the ratio between the current flow (the flow of information) between source and target, and the number of nodes along the path is lower. This is, given two paths with the same current flow value, the path that provides a stronger relation between source and target is the path with a higher number of nodes in it. In our example, we prefer path $A \to B \to C \to D \to E$ over path $A \to C \to D \to E$. This is due to the fact that the same amount of information (or resources) is flowing between the two paths, taking into account the fact that the longest path has more chances to diverge this information to other nodes. This does not happen, so we consider this an indication that the flow of information along this path is stronger than along the shortest path with the same delivered current.

Based on our results and on the activity log defined in Table II we observe that the flow of information usually goes from A to B, then B contacts C, C contacts D, and D contacts E. Other approaches that do not take into account the uncertainty and time sensitiveness of the events may conclude that the most likely path from A to E is from A to B to E. Analyzing the relationships between the actors we observe that there is only one interaction between actors B and E and is almost one year old based on the evaluation date. Also, this interaction was rated as *low* importance. This is the reason why our approach did not prefer the shortest path in lieu of the longest path that has stronger relationships among the actors.

V. Conclusions

FSNGs and current flows analysis are being investigated as an alternative approach to the extraction of information based on activity records. A methodology for analyzing activity records integrating the uncertainty and time sensitiveness of the information is introduced. Also, a method for the creation of a fuzzy social network graph is proposed. Finally, an analysis of the information encoded on the graph is discussed based on the analysis of current flows in electrical networks. Fuzzy set theory allow us to define these characteristics based on fuzzy linguistic variables. The evaluation of the relevancy of the information based on a time frame is accomplished by the means of fuzzy algebra. Current flows analysis emerges as an alternative approach for the analysis of information/resources flow along a network of entities. The ability to capture the influence of all nodes involved in a network over a particular path represents a promising avenue for the extraction of characteristics of the social network assuming that uncertainty and time sensitiveness are parameters of the information stored on activity logs that cannot be ignored and must be accounted for.

References

[1] S. Wasserman and K. Faust, *Social network analysis*. Cambridge: Cambridge University Press, 1994.

[2] P. S. Nair and S. T. Sarasamma, "Data mining through fuzzy social network analysis," in *NAFIPS '07: Annual Meeting of the North American Fuzzy Information Processing Society*. San Diego, CA, USA: IEEE Computer Society, 2007, pp. 251–255.

[3] B. T. Messmer and H. Bunke, "A new algorithm for error-tolerant subgraph isomorphism detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 5, pp. 493–504, 1998.

[4] M. Neuhaus and H. Bunke, "Edit distance-based kernel functions for structural pattern classification," *Pattern Recogn.*, vol. 39, no. 10, pp. 1852–1863, 2006.

[5] H. Bunke and K. Shearer, "A graph distance metric based on the maximal common subgraph," *Pattern Recogn. Lett.*, vol. 19, no. 3-4, pp. 255–259, 1998.

[6] D. Justice, "A binary linear programming formulation of the graph edit distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 8, pp. 1200–1214, 2006, fellow-Alfred Hero.

[7] M.-L. Fernández and G. Valiente, "A graph distance metric combining maximum common subgraph and minimum common supergraph," *Pattern Recogn. Lett.*, vol. 22, no. 6-7, pp. 753–758, 2001.

[8] W. D. Wallis, P. Shoubridge, M. Kraetz, and D. Ray, "Graph distances using graph union," *Pattern Recogn. Lett.*, vol. 22, no. 6-7, pp. 701–704, 2001.

[9] M. Blue, B. Bush, and J. Puckett, "Unified approach fuzzy graph problems," *Fuzzy Sets Syst.*, vol. 125, no. 3, pp. 355–368, 2002.

[10] C. Faloutsos, K. S. McCurley, and A. Tomkins, "Fast discovery of connection subgraphs," in *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM Press, 2004, pp. 118–127.

[11] D. J. Cook and L. B. Holder, *Mining Graph Data*. John Wiley & Sons, 2006.

[12] G. J. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1995.

[13] K. Tanaka, *An Introduction to Fuzzy Logic for Practical Applications*. Springer, 1997.