ANALYSIS

5(15), September 1, 2013

Discovery

ISSN 2278–5469 EISSN 2278–5450

Link prediction in social network: a fuzzy approach

Thakur GS[1], Yachana Bhawsar[2☼]

1.Assistant Professor, Dept. of MCA, MANIT, Bhopal, MP, India
2.Research scholar, Dept. of MCA, MANIT, Bhopal, MP, India

☼Corresponding author: Research scholar, MANIT, Bhopal, MP, India, Mail: yachanabhawsar@gmail.com

Publication History
Received: 12 July 2013
Accepted: 19 August 2013
Published: 1 September 2013

Citation
Thakur GS, Yachana Bhawsar. Link prediction in social network: a fuzzy approach. *Discovery*, 2013, 5(15), 21-23

ABSTRACT

Link prediction in Social Network is an open and challenging problem which has been studied intensively in recent years. There are lots of algorithms available for link prediction. In this paper we applied fuzzy soft concept to predict link in a given social network. Effective methods for link prediction could be used to analyze such a social network to suggest promising interactions or collaborations that have not yet been identified within organization. Molodstov proposed soft set theory; it is a mathematical tool for dealing with uncertainty. We proposed a new method of link prediction using soft sets.

Keywords: Social network, link prediction, soft set.

1. INTRODUCTION

A computational analysis of social networks structure has great research scope. Data mining related tasks in social networks involve centrality analysis, community detection, classification, viral marketing and link prediction. A social network can be represented as a structure whose nodes represent people or other entities and whose edges represent interaction, collaboration, or influence between entities. Examples of social networks are the set of all scientists in a particular discipline, with edges joining pairs who have co-authored papers; the set of all employees in a large company, with edges joining pairs working on a common project; or a collection of business leaders, with edges joining pairs who have served together on a corporate board of directors. Link prediction problem can be defined as follows:

1.1. Definition 1

Given a snapshot of a social network at time t and aim to predict links that will emerge in the network between t and a later time t'. Identifying links using fuzzy soft set is a new concept. Up to the present, research on soft sets has been very active and many important results have been achieved in the theoretical aspect. We can understand this with the help of an example.

1.1.1. Example 1

As shown in the Figure 1, we have a social network with 12 nodes at time t. After some time the new links can be added in the network, which is (2,3), (5,7), (7,13), (4,12). These edges are predicted links.

1.2. Definition 2

Let U be an initial universe set and E be a set of parameters. Let P (U) denotes the power set of U and A ⊂ E. A pair (F, A) is called a soft set over U, where F is a mapping given by F: A→P (U).

1.2.1. Example 2

Let U be the set of four dresses, given by, U ={d1, d2, d3, d4} .Let E be the set of parameters, given by, E = {costly, cheap, comfortable, beautiful, gorgeous}.

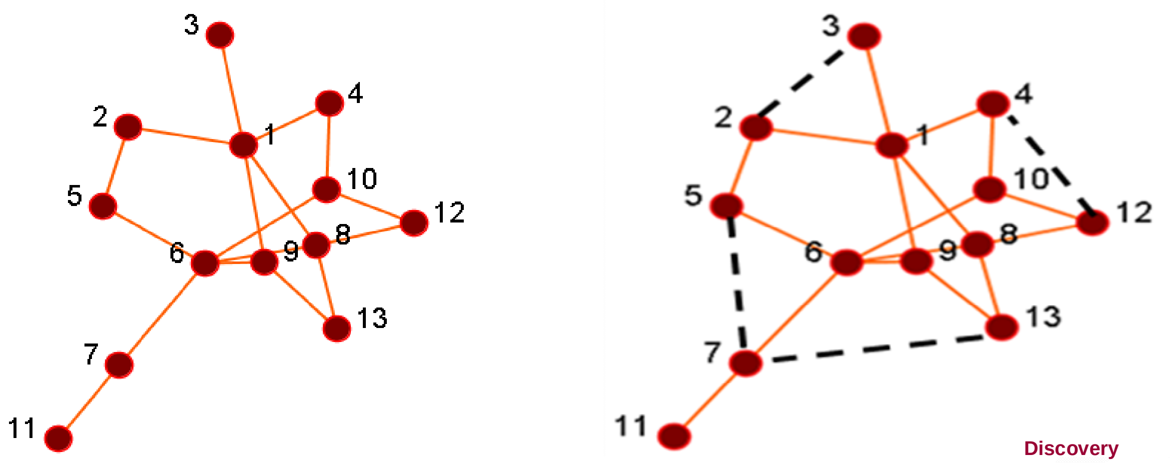

Figure 1
Social Network at time t and t'

Let A ⊂E, given by, A = *{costly*, cheap, comfortable, beautiful *}* = *{*e1, e2, e3, e4*}* where

e1 stands for the parameter 'costly',
e2 stands for the parameter 'cheap',
e3 stands for the parameter 'comfortable',
e4 stands for the parameter 'beautiful'.

Now suppose that, F is a mapping, defined as "dresses (·)" and given by, F(e1) = *{*d2, d4*}*, F(e2) = *{*d1, d3*}*, F(e3) = *{*d2, d3*}*, F(e4) = *{*d4*}*. Then the Soft Set (F, A) = *costly* dresses = *{*d2, d4*}*, cheap dresses = *{*d1, d3*}*, comfortable dresses = *{*d2, d3*}*, beautiful dresses = *{*d4*}}*.

The tabular representation of this soft set (F, A) is given in Table 1.

Table 1
Tabular representation of the soft set (F, A)

	e1	e2	e3	e4
d1	0	1	0	0
d2	1	0	1	0
d3	0	1	1	0
d4	1	0	0	1

Table 2
Tabular representation of the fuzzy soft set (F, A)

	e1	e2	e3	e4
C1	0.2	0	0.3	0.9
C2	0.9	1	0.4	0.1
C3	0.4	0.3	0.8	0.5
C4	0.6	0.4	0	0.3

1.3. Definition 3

Let U be an initial universe set and E be a set of parameters. Let P (U) denotes the set of all fuzzy sets of U and A ⊂ E. A pair (F, A) is called a fuzzy soft set over U, where F is a mapping given by F: A→P (U).

1.3.1. Example 3

Let U be the set of four cities, given by, U = *{*C1, C2, C3, C4*}* . Let E be the set of parameters (each parameter is a fuzzy word), given by, E = *{* highly, immensely, moderately, average, less*}* Let A ⊂E, given by, A = *{*highly, immensely, moderately, less*}* = *{*e1, e2, e3, e4*}*, where

e1 stands for the parameter 'highly',
e2 stands for the parameter 'immensely',
e3 stands for the parameter 'moderately',
e4 stands for the parameter 'less'.

Now suppose that,
F(e1) = {C1/.2, C2/.9, C3/.4, C4/.6},
F(e2) = {C2/1, C3/.3, C4/.4},
F(e3) = {C1/.3, C2/.4, C3/.8},
F(e4) = {C1/.9, C2/.1, C3/.5, C4/.3}

Then the fuzzy soft set is given by
(F, A) = {highly polluted city = {C1/.2, C2/.9, C3/.4, C4/.6},

immensely polluted city = {C2/1, C3/.3, C4/.4},
moderately polluted city = {C1/.3, C2/.4, C3/.8},
less polluted city = {C1/.9, C2/.1, C3/.5, C4/.3}}

The tabular representation of (F, A) is given in Table 2.

2. METHODS FOR LINK PREDICTION

Link prediction algorithms typically compute a probabilistic score for each candidate link and subsequently rank these scores and choose the largest ones. In the following we explained the existing methods of link prediction. We find that the number of proximity measures lead to predictions that outperform chance by factors of forty to fifty, indicating that the network topology does indeed contain the latent information from which to infer future interactions.

2.1. Common neighbors

The most direct implementation of this idea for link prediction is to define score (x,y) =| Γ(x) ∩ Γ(y)| , the number of neighbors that x and y have in common. Newman (2001) has computed this quantity in the context of collaboration networks, verifying a correlation between the number of common neighbors of x and y at time t and the probability that they will collaborate in the future.

2.2. Jaccard's coefficient and Adamic/Adar.

The Jaccard coefficient commonly used similarity metric in information retrieval (Gerard Salton et al. 1983) measures the probability that both x and y have a feature f, for a randomly selected feature f that either x or y has. If we take "features" here to be neighbors, this approach leads to the measure score(x , y) =| Γ(x)∩Γ (y)|/|Γ(x) U Γ (y)|.

2.3. Preferential attachment

It has received considerable attention as a model of the growth of networks. The basic premise is that the probability that a new edge has node x as an endpoint is proportional to =| Γ(x)|, the current number of neighbors of x. Newman (2001) and Barabasi et al. (2002) have further proposed, on the basis of empirical evidence, that the probability of co-authorship of x and y is correlated with the product of the number of collaborators of x and y. This proposal corresponds to the measure score(x; y) =| Γ(x)|.|Γ (y)|.

3. PROPOSED METHOD

We proposed a novel algorithm for link prediction based on fuzzy soft set, which is presented as follows:

Algorithm
Inputs: (a) Number of links ***n*** in social network.
(b) Generic parameters ***m***.
Output: Predicted link(s), (Table 3).
Step 1.Selection of a desired number of links (***n***).

Step 2.Selection of a desired number of parameters (***m***).
Step 3.Construction of a fuzzy soft set (F, E) and tabulation.
Step 4.Construction of Comparison Matrix C[i,j], (Table 4).

$$C[i,j] = \sum_{k=1}^{m} (f_{ik} - f_{jk})$$

Step 5.Construction of decision factor (Ri), (Table 5).
Step 6.The maximum value of Ri indicate that link Li is an predicated link.

Table 3
Tabular representation of links

U	E1	E2	E3	E4
L1	0.6	0.6	0.7	0.8
L2	0.5	0.7	0.7	0.8
L3	0.5	0.8	0.9	0.9
L4	0.8	0.5	0.5	0.6
L5	0.4	0.8	0.6	0.7
L6	0.5	0.8	0.9	1.0

Table 4
Comparison Matrix C[i,j]

U	L1	L2	L3	L4	L5	L6
L1	0.0	0.0	-0.4	0.3	0.2	-0.5
L2	0.0	0.0	-0.4	0.3	0.2	-0.5
L3	0.4	0.4	0.0	0.7	0.6	-0.1
L4	-0.3	-0.3	-0.7	0.0	-0.1	-0.8
L5	-0.2	-0.2	-0.6	0.1	0.0	-0.7
L6	0.5	0.5	0.1	0.8	0.7	0.0

Table 5
Decision factor R_i

Ri	values
R1	-0.4
R2	-0.4
R3	2.0
R4	-2.2
R5	-1.6
R6	2.6

$$R_i = \sum_{j=1}^{m} C[i,j]$$

4. EXPERIMANTAL SETUP AND RESULT

Let the initial universe U={L1,L2,L3,L4,L5 ,L6} be the six selected links of a Social Networks and E={E1,E2,E3,E4} be the generic parameters where E1,E2,E3,E4 are mutual friends, Like, Comments and Post respectively. Consider a Fuzzy Soft Set (F,E)={ {E1},{E2},{E3},{E4}} where

- F(E1)={L1/0.6,L2/0.5,L3/0.5,L4/0.8,L5/0.4, L6/0.5}
- F(E2)={L1/0.6,L2/0.7,L3/0.8,L4/0.5,L5/0.8, L6/0.8}
- F(E3)={L1/0.7,L2/0.7,L3/0.9,L4/0.5,L5/0.6, L6/0.9}
- F(E4)={L1/0.8,L2/0.8,L3/0.9,L4/0.6,L5/0.7, L6/1.0}

Decision factor Ri plays an important role in link prediction. Maximum values of Ri is used to select the corresponding link Li. The Maximum values of Decision factor Ri in Table are R6 =**2.6**, and R3 =2.0.Thus the values of R6, and R3 form table predicts that link L6 followed by link L3.

5. CONCLUSION

In this research work fuzzy soft set theory is used for link prediction of social network. The proposed fuzzy soft set based link prediction predicts the link(s). As per our knowledge this is the new work for predicting links in social network using fuzzy soft concept. Further research can be done if we have the same scores for two links, then it is difficult to predict the links. We are working in this direction also.

REFERENCE

1. Barabasi L, Jeong H, Neda Z, Ravasz E, Schubert A, Vicsek T. Evolution of the social network of scientific collaborations. *Physica A*, 2002, 311(3-4), 590-614
2. Gerard Salton, Michael J McGill. Introduction to Modern Information Retrieval. McGraw- Hill, 1983
3. Newman MEJ. Clustering and preferential attachment in growing networks. *Physical Review Letters E*, 2001, 64(025102)