

Identifying Most Relevant Node Path To Increase Connection Probability In Graph Network

Abhiram Gandhe

Computer Science and Engineering
Visvesvaraya National Institute of Technology
Nagpur, India

abhiram.gandhe@gmail.com

Parag Deshpande

Computer Science and Engineering
Visvesvaraya National Institute of Technology
Nagpur, India

psdeshpande@cse.vnit.ac.in

Abstract

In social networks, one of the most challenging problems is to find the best way to establish a relationship between two nodes. Different attributes (Topological, Non-Topological) can be used to define friendship score between two nodes which indicates the strength of a relationship. Non-Topological attributes can be used to define the strength of a relationship even if two nodes are not connected. The concept of friendship score to define the strength of a relationship between two nodes transforms social network into a complete graph where each node is connected to every other node and where friendship score is used as link attribute. The information on already existing connections in social media network and graph which is formed based on friendship score can be used to find out best way of connecting two different nodes even if no path is in existence in social media network between these nodes.

In this paper, we propose a novel way of estimating friendship score using non-topological attributes based on available information in social media network and algorithm to find out best way of connecting two nodes in the form of chain of reference. The chain of reference between node X_1 and X_n is a path $X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_{n-1} \rightarrow X_n$ where each link $X_i \rightarrow X_j$ is having high friendship score. The chain of reference indicates how X_1 can be connected to X_n even if no path exists between X_1 and X_n in social media network.

Keywords: Friendship Link, Online Social Network, Graph Network, Node Path, Reference Chain.

1. INTRODUCTION

Online Social Networks (OSN) like Facebook, Google+, Twitter, LinkedIn, etc. are very dynamic. With the growing influence of the social network and their far-fetched reach, every individual is trying to get the maximum out of his social network. OSN have million to billions of users communicating and connecting with each other on a daily basis. Every individual is on a lookout for good information of his social network and more and more connect on the go.

While the connections in the network grow and more relevant and interesting individuals get added to the network, it is always a demand for tools to look for the best and most relevant ways of connecting with the people of interest. The friend recommendations in today's OSN are restrictive in the way they suggest connects.

In most of the OSNs, connects are suggested by:

- a) Identifying two unconnected individuals with maximum mutual friends
- b) Mutual friends are asked to suggest connects between unconnected friends

- c) Unconnected individuals having multiple short length paths in the graph are suggested a connect
- d) Unconnected individuals commenting on the same conversation, multiple times are suggested a connect

There is no way an individual to find out the best way or a recommended path to a distant individual who is not in his network or in FOAF (Friend of a Friend) network [5]. LinkedIn has a function to “get introduced”, but that is only limited to FOAF network. Anything above the second level contact doesn’t have a way to get introduced. If person X who is a Software Developer in Bangalore is looking forward to connecting with Y, who is a software programmer in Brazil there is no way to find out a path of reference. In this paper we propose an algorithm, which with the help of Friendship Score (fScore), a score obtained to identify the probability of friendship between to unconnected individuals, will provide the most relevant Friendship Link, on which if chain of reference is invoked, can provide the best possibility of attaining a connect.

If a node needs to connect to a distant unconnected node, the best way to do it is to get introduced to the node of interest through the already existing linked node path from source node to destination node. If a source node is connected to destination node by multiple unknown paths then it is always a best way to get connected to the destination node by invoking introduction reference along the most relevant path from source node to destination node. To find such a node path the most important steps would be:

- a) Finding multiple paths existing between source node and destination node
- b) Most relevant path of all the existing paths between source node and destination node

The Importance of the problem, as stated in the paper “The small world problem” [6], brings under discussion the network of acquaintances and friends in the social structure. The two schools of thought about the small world problem are:

- i. Two individuals how much ever remotely situated they are, they are linked to each other by a finite number of links between acquaintances and friends and the number of intermediate links between them are relatively small than expected.
- ii. There are unbridgeable gaps between various groups and, therefore, the two distant individuals will never link up. A message from one circle will never be able to jump to the other circle.

In this paper, we do not take sides of any of the schools of thought stated above. Here, we propose to find out the most relevant path between two individuals which with recommendation and connects followed, will lead to a friendship between the two individuals or else report that no relevant or strong path exists between the two individuals.

2. RELATED WORK

There is a variety of local similarity measures [5] (i.e. FOAF algorithm, Adamic/Adar index, Jaccard Coefficient, etc.) for analyzing the “proximity” of nodes in a network and suggesting links and friendship. FOAF [2] is adopted by many popular OSNs, such as facebook.com and hi5.com for the friend recommendation task. FOAF is based on the common sense that two nodes x, y are more likely to form a link in the future if they have many common neighbors. In addition to FOAF algorithm, there are also other local-based measures such as Jaccard Coefficient [5] and Adamic/Adar index [1]. Adamic and Adar have proposed a distance measure to decide when two personal home pages are strongly “related”. There is a variety of global approaches in paper[5]. As comparison partners of global approaches, we consider Katz status index [4], Random Walk with Restart algorithm [9] and RWR algorithm [7]. Katz defines a measure that directly sums over all paths between any pair of nodes in graph G , exponentially damped by length to count short paths more heavily. RWR considers a random walker that starts from node x , and chooses

randomly among the available edges every time, except that, before he makes a choice, with probability α , he goes back to node x (restart).

3. PROBLEM STATEMENT

A person 'A' in social network wishes to connect unknown person 'B' because of some professional interest, what will be the best path to connect person Y?

In most of the OSNs today the request won't be sent by the OSN functionality as "A" doesn't directly know "B". If "A" sends a friend request to "B" then even if the request lands with "B", "B" won't accept the friendship request as "B" doesn't know "A". The only way by which "B" will consider "A"'s request is when "A" is recommended or introduced to "B" by some of "B"'s acquaintance or friend.

In this paper, we propose a novel method to identify the most relevant path between "A" and "B" when the recommendations/introductions can be forwarded by the nodes on the path, with least cost and better accuracy, to form the most reliable friendship path between "A" and "B". If A and B exist in disjoint graphs or have weak linkage path, then the method proposed will specify that "No reliable and strong reference path exists between A and B at this moments snapshot of the network".

4. METHODOLOGY

4.1 Friendship Score (fScore)

Non-Topological Attribute values of two nodes help determine Friendship Score (fScore) [12]. If two individuals have Same Gender then a specific weighted score is added to the fScore. In the paper titled "Use of Non-Topological Node Attribute values for Probabilistic Determination of Link Formation", we have detailed the mathematical models to calculate fScore between two individuals.

If a Link is present between two nodes, a and b, then:

$$\mathbf{FScore[a, b] = 1 + FNT(a, b)}$$

Where FNT(a,b) is a function non-topological attribute of a and b. If a link is not formed between two nodes, a and b, then friendship score needs to be calculated using non-topological attribute data. The fScore is calculated as:

$$\mathbf{FScore[a, b] = FNT(a, b)}$$

FNT(a,b) is a weighted sum of non-topological attributes. The weights are calculated by collecting sample data from the Facebook. The sample data for calculating weights in function consists of 7637 user data and 101904 existing friend links. The mathematical formula, to calculate the function value

$$FNT [a, b] = \frac{\sum_{i=1}^n \sum_{j=1}^m W_{ij} \times V_{ij}}{n}$$

Where:

i: Attributes/Features e.g. Gender (if there is no value associated with the attributes in any/both of the nodes a or/and b then that attribute will not be taken up for the calculation of fScore)

j: Different Attribute Values possible e.g. Same Gender, Different Gender (No gender available is also a valid possible value, but it will already be excluded from the equation because of the elimination of attributes while collating the final set of i's)

Attribute Value Weights calculated using Conditional Probability Bayes rule on the existing data set from Facebook:

W (Same School)	0.0235561	W (Same Gender)	0.0047786
W (Same Location)	0.0106998	W (Different Gender)	0.0038902
W (Different Favorite Athlete)	0.0064703	W (Different Location)	0.0037144
W (Same Favorite Athlete)	0.0055419	W (Different School)	0.0032468

So, formula for finding fScore for two nodes, having values for gender, location, School and favorite athlete attributes, is as follows:

$$FNT(SG, SL, SS, SFA) = \frac{WSG \times VSG + WDG \times VDG + WSL \times VSL + WDL \times VDL + WSS \times VSS + WDS \times VDS + WSFA \times VSFA + W DFA \times VDFA}{4}$$

Topological attributes can also be used to effectively to calculate friendship score [13]. We have used Non Topological attribute Value between two nodes to calculate friendship score [12]. Method to use Topological Attribute to calculate friendship measure [13] combined with the method suggested above to calculate Friendship score using Non Topological attribute values can result in a more inclusive friendship score calculation. Calculation of Friendship Score (fScore) is not within the purview of this paper. We will be going ahead with the Non Topological Function FNT equation stated above.

5. PROPOSED MODEL

We propose a novel approach, which uses fScore, as the linkage probability predictor, will provide an existing path in the network, from source to destination, which can be termed as the most relevant path for invoking chain of reference.

5.1 Detailing Steps In The Proposed Approach

Identify most relevant path from Source to destination.

Source	Destination
A	B

Step 1:

Are A and B Friends (fScore [A, B] >= 1)?

- i. Yes: A -> B has an existing friendship. No Need to proceed further.
- ii. No: Process to Step 2.

Step 2: Initialization

Configure Algorithm Parameters: Max Hops, Min fScore Threshold, Qualifying percentage

Initialize Source Set with Source Node and Destination Set with Destination node.

Level = 0

Source Set	Destination Set
A	B

Step 3:

- a. Identify Neighbors for nodes in Destination Set at this level
NB={NB1, NB2, ---, NB_i}
i connected neighbors of B
- b. Get fScore of all neighbors with all nodes in Source Set
- c. Sort the resultant list with the decreasing order of fScore

- d. Trim the sorted list by taking only top Qualifying percentage of the list with maximum values of fScore selected.
- e. Remove all the items in the trimmed list that have fScore less than Min fScore Threshold.
- f. Check for duplicates in the list and remove the duplicates, retaining maximum fScore in the trimmed list
- g. Add the trimmed list to the destination set.
TNB = Trimmed NB

Source Set	Destination Set
A	B + TNB

Step 4:

- a. Check in the destination set if there is any node with fScore more than or equal to 1
 - a. Yes: Path is identified from Source to Destination which has highest relevance w.r.t. fScore
 - i. Process the Source Set and Destination Set to print relevant path
 - ii. Exit
 - b. No: Continue with next steps

Step 5:

- a. Identify Neighbors for nodes in Source Set at this level
NA={NA1, NA2, ---, NAj}
j connected neighbors of A
- b. Get fScore of all neighbors with all nodes in Destination Set
- c. Sort the resultant list with the decreasing order of fScore
- d. Trim the sorted list by taking the only top Qualifying percentage of the list with maximum values of fScore selected.
- e. Remove all the items in the trimmed list that have fScore less than Min fScore Threshold.
- f. Check for duplicates in the list and remove the duplicates, retaining maximum fScore in the trimmed list
- g. Add the trimmed list to source set.
TNA = Trimmed NA

Source Set	Destination Set
A +TNA	B + TNB

Step 6:

- a. Check in the source set if there is any node with fScore more than or equal to 1
 - a. Yes: Path is identified from Source to Destination which has highest relevance w.r.t. fScore
 - i. Process the Source Set and Destination Set to print relevant path
 - ii. Exit
 - b. No: Continue with next steps

Step 7:

If no path is identified, till this step:

- a. Level ++;
- b. Process Step 3,4,5,6 for the next hop with set of neighbors for this level.

Step 8:

Is the path identified?"

- a. Yes: We have the most relevant path

No: We can process again by changing the configuration algorithm parameter (Max Hops, Min fScore Threshold, Qualifying percentage) values

6. ALGORITHMIC REALIZATION

The proposed approach is realized in the form of an algorithm. The algorithm, Relevant Path algorithm, will propose the best path, which when traversed from source to destination, will provide the most relevant path. The Path given as the output of the algorithm is most relevant as fScore is the parameter used to weigh the relevance of the node considered for finding the linkage path.

We have put below the components, which will give the Relevant Path algorithm outline with all the other components needed to execute Relevant Path algorithm.

- fScore function calculates the Friendship Score between two nodes as stated above in this paper
- getRelevantNeighbors function provides the list of relevant neighbors as detailed in step 2 and 4 above
- getRelevantConnectPath function using the getRelevantneighbor and fScore function finds out the most relevant path, inside the network, from source to destination
- printRelevantPath just helps to print the path identified by getRelevantConnectPath

Relevant Path algorithm when processed on a dataset obtained from Facebook® using graph query, the same data set which was used to obtain the weights stated above. After processing the above algorithm on the network to find out relevant path for few set of nodes, the results are as follows.

6.1 Configuration Parameter Values

max_hops = 6
 minfScore = 0.001
 slab_percentatge = 10

Query 1:

getRelevantConnectPath('651556926', '540415531')

Output:

Source Set

Level 0: fScore=>0.004552525,

Node ID =>651556926

Destination Set

Level 1: fScore =>1.0039132666667

Node ID =>619017668

Level 0: fScore => 0.004552525

Node ID =>540415531

Query 2:

getRelevantConnectPath('100002435033939', '100001284823018')

Output:

Source Set

Level 3: fScore =>1.00940775

Node ID =>100000352452195

Level 2: fScore => 0.004330425

Node ID =>651556926

Level 1: fScore => 0.0036171333333333

Node ID =>520669641

Level 0: fScore] => 0

Node ID =>100002435033939

Destination Set

Level 0: fScore] => 0

Node ID =>100001284823018

Query 3:

getRelevantConnectPath('1641471744', '895005331');

Output:

Source Set

Level 2: fScore => 1.0040127	Node ID => 651556926
Level 1: fScore => 0.0077392	Node ID => 100000352452195
Level 0: fScore => 0.003913266666667	Node ID => 1641471744

Destination Set

Level 2: fScore => 0.01416735	Node ID => 100001740004686
Level 1: fScore => 0.0040127	Node ID => 664919666
Level 0: fScore => 0.003913266666667	Node ID => 895005331

7. CONCLUSION AND FUTURE SCOPE

Relevant Path Algorithm proposed here, using fScore between nodes as a linkage probability predictor, identifies a network path, from the source node to the destination node, as the most relevant path to invoke the chain of reference. This relevant path identified using Relevant Path algorithm will yield the best result on invocation of chain of reference on the path than any other network path between source node and destination node.

The other algorithms to predict friendship using a network path used in graph networks is very restrictive. The example of the same is FOAF (Friend of a Friend) algorithm. It restricts the traversal of the network only to the depth of 2 from the source network. Katz Algorithm defines a measure that directly sums over all paths between any pair of nodes in the graph and is exponentially damped by length to count short paths more heavily. Other algorithms for shortest path identification using edge weights, in our case fScore, like Bellman-Ford Algorithm need to calculate the path from each node to every other node. With continuous changes in the weights due to change in any Node attribute value makes the maintenance of the resultant matrix in this algorithm continuously changing due to which it has huge computational needs.

The proposed Relevant Path Algorithm is a real-time path calculation algorithm and is capable of accommodating the continuous changes in the network topology and Node attribute values. It is very much possible that a different path can be proposed by relevant path algorithm between the same set of source node and destination node at two different time instances t1 and t2.

In this proposed method, Relevant Path Algorithm, specifies the strength of a friendship link using Non Topological attributes. The proposed method, bases the relevant path algorithm on strength of the link. We have proposed an algorithm to find the most relevant path algorithm which uses fScore between nodes as a linkage probability predictor. The algorithm proposed can be strengthened more by enriching the formula for fScore calculation.

In the future scope, the strength of the link can also incorporate topological attributes along with the non-topological attribute. The integrated model, thus designed, will strengthen the relevance of the path, by taking topological along with Non Topological attributes.

8. REFERENCES

- [1] L. Adamic and E. Adar. How to search a social network. *Social Networks*, 2005, pp. 187–203.
- [2] J. Chen, W. Geyer, C. Dugan, M. Muller, and I. Guy. Make new friends, but keep the old: recommending people on social networking sites. *Proceedings of the 27th international conference on Human factors in computing systems*, 2009, pp. 201–210.
- [3] K. C. Foster, S. Q. Muth, J. J. Potterat, and R. B. Rothenberg. A faster katz status score algorithm. *Compute Math Organ Theory*, Dec 2001, pp. 275–285,

- [4] L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 1953, pp. 39-43.
- [5] D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. *Proceedings of the 12th International Conference on Information and Knowledge Management (CIKM)*, 2003
- [6] S. Milgram. The small world problem. *PsychologyToday*, 1967, pp. 61–67.
- [7] J. Pan, H. Yang, C. Faloutsos, and P. Duygulu. Automatic multimedia cross-modal correlation discovery. *Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 653– 658.
- [8] F. Rubin. Enumerating all simple paths in a graph. *IEEE Transactions on Circuits and Systems*, 1978, pp. 641–642.
- [9] H. Tong, C. Faloutsos, and J. Pan. Fast random walk with restart and its applications. In *ICDM '06: Proceedings of the 6th International Conference on Data Mining*, 2006, pp. 613–622.
- [10] S.Wasserman and K. Faust. *Social network analysis: Methods and applications*. 1994.
- [11] A. Papadimitriou, P.Symeonidis and Y. Manolopoulos. Friendlink: Link Prediction in Social Networks via Bounded Local Path Traversal. *International Conference on Computational Aspects of Social Networks (CASoN)*, 2011.
- [12] A. Gandhe, P. Deshpande. Use of Non-Topological Node Attribute values for Probabilistic Determination of Link Formation. *International Journal of Advanced Computer Science and Applications(IJACSA)*, Volume 6 Issue 2, 2015, PP. 186-191.
- [13] M. Fire, L. Tenenboim, O. Lesser, R. Puzis, L. Rokach, Y. Elovici. Link Prediction in Social Networks using Computationally Efficient Topological Features. *IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing*, 2011.