SOCIAL COMPUTING: AN INTELLIGENT AND RESPONSIVE SYSTEM

Dev Rishi Tekriwal, Undergraduate Student
Arkaprava De, Undergraduate Student
Ezhilmaran D, Assistant Professor
School of Computing Science and Engineering,
Vellore Institute of Technology, Vellore, India

Abstract

       This paper deals with the advantages of a graphical database over conventional database in terms of real world applications. The unique property exhibited by the graphical databases which is, that the edges inside the graph which is the relationship between the nodes can be dynamically changed in real time without any computational overrides thereby providing a more optimized structure to compute and manipulate, is given as a hypothesis in this paper. It also throws light on how the data mining techniques can be implemented in the social networking, thus making it more intelligent and responsive systém. It also propose a methodology to generate a social graph of user's action and predict future social activities using graph mining. Through this model, we believe that it becomes clearer that data from different contexts can be related such that new solutions can be explored and thus, it may provide illumination for the aforementioned problems and stimulate new research.

Keywords: Graph mining, coloring, influence factor

Introduction

       Social networking has become an important area of study because of the sprawling and rapidly evolvement over the last decade with the advance of Internet and Web Technologies. Computing is shifting to the edges of network, and individual users are empowered with technology to engage in Social interaction, sharing content and distributing information. As a result the data is increasing day by day at a rate which is difficult to control and process. Hence a model is needed which can handle the discrepancies in the system connecting the people of similar interest which consists of following some pattern which could be extremely useful to obtain information that could be of interest to different people of different areas. Hence we would be

using a graphical model to represent social network data, describing and analyzing patterns of social relations. A novel approach for this would be visualizing datasets into dynamic, directed, and weighted graphs. The datasets will be obtained from the online social networking site and the graph would be constructed using the same dataset. The graph depicts the user's status in real time scenario using links and the weights on the links. The association mining rules would be generated based on the graph and these rules would be used for doing various predictions. One such example of Graph Database System is the Facebook new Architecture which is also known as the Graph Search. Facebook Graph Search is based on Unicorn which is a graph traversing tool. Using traditional information-retrieval systems to mix keyword and structured queries is fairly well understood. But Facebook required the structure also to find responses more than a single connection away, such as "hotels liked by my friends from India." Unicorn allows you to write a simple s-expression query to get a list of all Facebook Ids which are your Friends of friends. Using s-expression you can traverse the graph in multiple depths in just one go. Facebook stores two types of data: objects, and the relationships between those objects. And with the use of Open Graph, the model has been extended to third-party applications and websites.

Visualization

In general, the way graph databases work is that you use some index to find a starting point, called a node. Once you have a starting point, you traverse the graph node by node by traveling across relationships. Designing for a graph database is all about turning your data into nodes and relationships. For Facebook Graph that starting point is often your account, every item in the Graph is a node, and your actions (like, share, etc.) are relationships between you and those nodes.

We know that graph is a set of vertices and connecting edges between them, where each vertex or node represents users/ organization etc. and the edges representing the relationship between them. In the case of social networking, the nodes represents people, organization while the edge represents the relationship like "friends of", "married to", "like", etc. thus we can represent the social structure in form of a social graph where the behavior of the people could be leant. If an edge exists {a, b} then we can say that nodes a and b are related to each other. The edges themselves can be unordered pairs of nodes or in a directed graph (digraph), ordered pairs of nodes where each edge has a direction, sometimes called an arc. Graphs are (generally) no reflexive; nodes are not related to themselves. Order is # of nodes, size is # of edges. (Sun, Tang, Wang & Yang 2009)

Most of the social networks are not static and hence it is useful to think about how a network evolves over time i.e. how nodes arrive and

depart and how the edges form and vanish. If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends too. This could be understood by the Triadic closure-a principle Triadic closure-a principle. The formation of the edge between Band C illustrates triadic closure. B and C have an incentive to become friends (form a BC edge). (Cameron, Leung & Tanbeer 2011)

(a) Before B-C edge forms.   (b) After B-C edge forms.
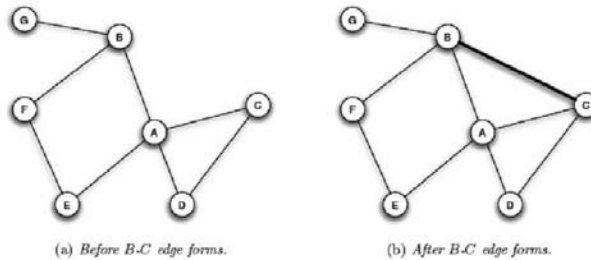
Hence we are proposing data mining technique which could be used to form a social graph in the real scenario of dynamic nature. The data obtained, contains information of diverse people from different areas of world and these data are stored in a database. This database can be mined using mining techniques for social marketing professionals, advertisement, analysts and also as find risky group of people for security purpose. Now these predictions can only be done using proper association rules techniques or algorithms that scans multiple databases for association rules. But techniques like apriori , fptree algorithms produces more complexity in terms of time and space as databases are traversed multiple times thus reducing efficiency. (Bhatia & Kadge 2012)

Thus we approach the problem with social graph technique where the database is broken down and visualized as a directed, weighted graph that dynamically increases its size. First the static data is taken from user during registration of the user and a graph is formed which dynamically increases based on usage or behavior of the user within the social networking site. The links within the graph represents association or relationship and the weights represent the strength of the relationship within the links. And so we can use association techniques and suggest friends to the user based on his/hers mutual friends, or become a part of an organization by understanding the likes and dislikes of the user. And this is performed where the database of the user acts as training datasets for social graph technique.(Ilkan, Safaei & Sahan 2009)

Visualizing the proposed concept

Why we need the proposed technique – To increase user friendliness of the social networking site and make it more intelligent towards its users, and also to increase social awareness and improvements

How it is done ?-   The dataset containing user information is converted to a directed, weighted social graph that dynamically increases its size based on user interaction and behavior and trains itself to intelligently form social relationships between other users, thus creating a social web of information

Rules in the social graph formation

1. Each edge represents the relationship between the nodes, which is the activeness of the user and it dynamically increases on association techniques
2. Each edge consists of weights that represent how strong the relationship is.
3. Each edges or link are when linked between two users and unidirectional if link is between community and user (Bhatia & Kadge 2012)

Example

Two users are represented using nodes:
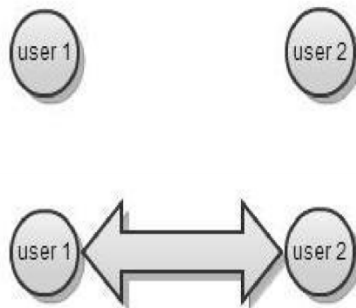
If two they are friends then they are represented with nodes connected by bidirectional edge.

And depending on the closeness of friendship between them, weights are assigned like: just friends, best friends, medium friends, etc.

Thus a social graph is formed related to all the relationship between users and also between users and community and they are related using edges with weights.
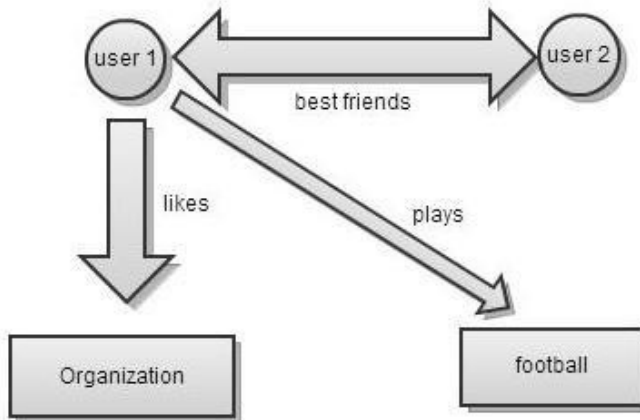
Now take this concept to calculate the influence factor among the nodes in our graph. To get the picture, consider the following scenario. Even the news feed in Facebook is an efficient engine in finding the influence factor between people i.e. the important people whose posts are shown.

In a social network, people who are more influenced are considered to me more important to us. Even Facebook uses the same technique to classify the important news feeds from the other by the method of influence function which is determined by the technique below.

We can measure the influence between users by means of connectivity and strength of relationship between them. We know that the degree of a node is the number of nodes connected to it. And from this connectedness we can find the influence of each node over the other by the weights between the connectivity i.e. the strength of the relationship. Let's represent influence by the function I where , (Borgatti & Everett 1996)

I = degree (p) = number of friends of the person p

The main advantage of this is to find the influence. The graph is formed using the adjacency matrix calculated, then the influence of a node is the sum of the row in the corresponding row and the operation is less complex in term of performance, but considered to be the naïve approach. Let us consider this graph:

Single person with high Degree
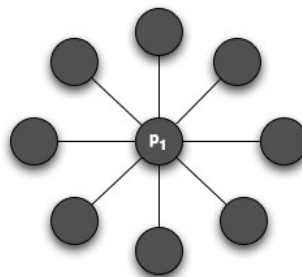
Single person low degree but high connectivity
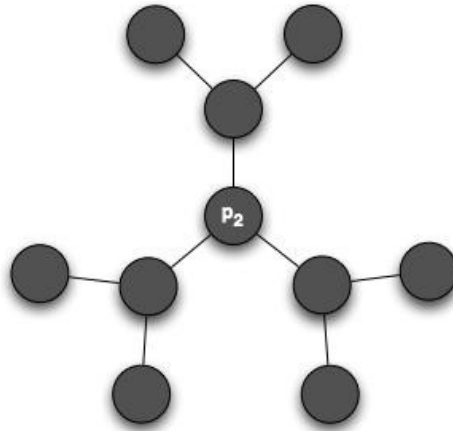
I_d denotes measure of influence and hence the first person, P1, has a greater ration of influence because it is connected directly to 8 nodes. The second person, P_2, however can influence up to 9 nodes (people).The same is the case in real world also. This sort of indirect influence can be measured by using a measurement called eigenvalue centrality. The idea is that a person's influence is comparative to the total influence of the people to whomever he/she is linked. So let's this influence measure be I_e. (Borgatti & Everett 1996)

Let's consider someone is the President at A Corp. The Corp has four VPs and each of them has influence factor of five and these are the persons to whom the President is directly connected to. Then this measurement says that there is some λ (number) such that

I_e (President) = 1/ λ [I_e (VP$_1$)] + [I_e (VP$_2$)] + [I_e (VP$_3$)] + [I_e (VP$_4$)]

λ is the key factor determining the level of influence which each of the person share through the connectivity. Small value of λ tells that the president has much influence, if it is big then he has slight. How λ is to be calculated?

Let G be a social graph, where vertices represents persons and edges represent social relationship, and let A be the adjacency matrix of graph G. If there are N people in the given social network, labeled P$_1$, P$_2$, the above could be generalized and we can conclude:

$$I_e(p_i) = \frac{1}{\lambda} \sum_{j=1}^{N} A_{i,j} I_e(p_j)$$

Reminisce that $A_{i,j}$ is 1 if p_i and p_j are linked through an edge and 0 otherwise. It's also significant to notice that λ is a function of the *graph* and not of any individual node (person). If we say $x_i = I_e(p_i)$ then a vector x can

be formed whose ith coordinate is the influence of the ith person. The above equation could be rewritten using matrices and vectors :

$X = 1/ \lambda \ (A/X)$

Roles/Positions

We take a digraph represented by D(V,E), where the in-neighbor of node v is denoted by Ni(v) which is the set of vertices that sends arcs towards v, which is $N_i(v) = \{u: (u,v) \square E\}$. We denote the out neighbors of a node v by $N_o(v)$ which is the set of vertices that get arcs from *v*. and is denoted by, $N_o(v) = \{u: (v,u) \square E\}$.

A coloration *C* is considered as an assignment of colors to the vertices *V* of the digraph. *C(v)* denotes the vertex v color and the set of diverse colors which is assigned to the nodes in a set *S* is represented *C(S)* and is labeled the spectrum of *S*. In the Figure 1, a coloration of the nodes has been showed by labeling the nodes with single letters such as 'y' for yellow and 'r' for red. If two nodes are of same color, then they are said equivalent. (Albert & Barabasi 2002)
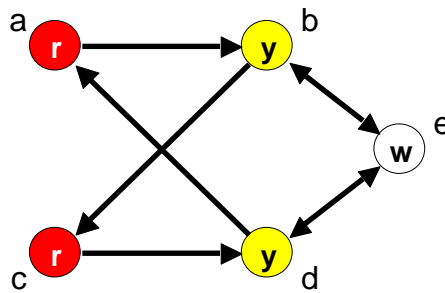

Figure 1 .

A coloration is said to be *strong coloration* if the nodes are allotted with the similar color iff they have similar in-and-out neighborhoods. Which means, for all u,v $\square$V, C(u) = C(v) if and only if Ni(u) = Ni(v) and No(u) = No(v). Therefore the coloration in Figure 1 is a strong coloration. This could be checked by taking node pairs and verifying if they are of the same color, then they have identical neighborhoods, and if they have different color, then they have dissimilar neighborhoods. For example, *b* and *d* are of same color, and both of their neighborhoods entail {a,c,e}.

A coloration C is said to be *regular* if $C(u) = C(v)$ implying that $C(Ni(u)) = C(Ni(v))$ and $C(No(u)) = C(No(v))$ for all *u, v* $\square$V. To put it in different words, in regular colorations, every pair of nodes having identical color must receive arcs from nodes encompassing the same set of colors and must direct arcs to nodes containing the same set of colors. Also every structural coloration is also a regular coloration.
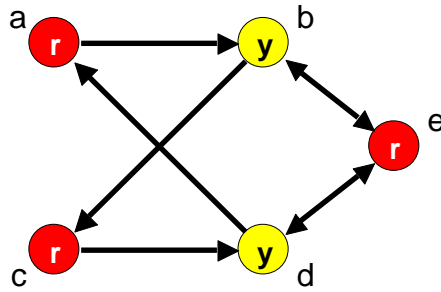
Figure 2 . Regular coloration.

The coloration in Figure 2 is regular, but not strongly structural. This can be checked by considering, every red node which has an out-neighborhood comprising only yellow nodes and an in-neighborhood comprising only yellow nodes, whereas every yellow node has an out neighborhood comprising only red nodes and an in-neighborhood comprising only red nodes. An another regular coloration has been depicted in figure 3. Node *g* could not be colored the same as f or i, because it has an out neighborhood containing a white node, while f and i have no out neighborhood at all. This also gives an implication that g cannot be colored the same as f and i because it has a tie from a node of a other color. (Albert & Barabasi 2002)
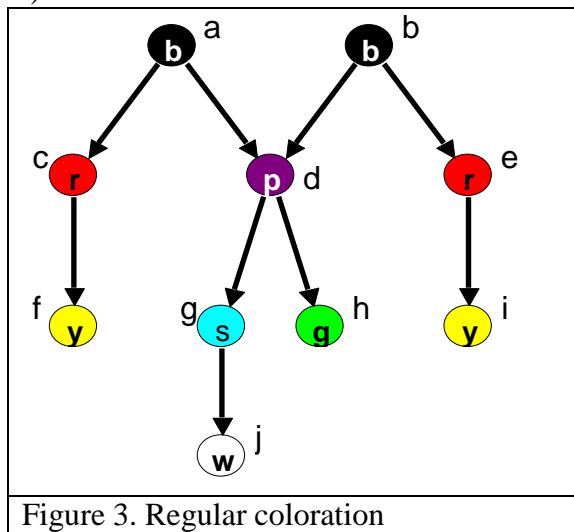

Figure 3. Regular coloration

If we represent social network in form of graphs, the colors could be thought as defining types or classes of people such that if one member of some class (blue) has outgoing links to members of exactly two other classes (yellow and green), then all further members of that (blue) class have outgoing links to members of those similar two classes (yellow and green). Therefore according to the concept of regular colorations, members of social network are classified according to their pattern of relations of others, and

two persons are classified in the same class if there is some interaction in same ways with the same kind of others.

Conclusion

Thus we have proposed a mixed approach towards the formation of a social graph. This mixed approach is a combination of mining technique, graph coloration technique and influence factor to form a more stable, dynamic structure of the social web. This system also predicts the activities that could be done by the user. Based on this prediction technique, we could also predicts the activity, membership of the user in different community or between two communities. We have also visualized, how these techniques are easier and has less time complexity and better performance than other association rules like apriori algorithm or fptree formation. Thus we conclude the paper by proposing a better approach or improvement in social networking area.

References:

Juan J. Cameron ,Carson Kai-Sang Leung, Syed K. Tanbeer, "Finding Strong Groups of Friends among Friends in Social Networks", Ninth IEEE International Conference on Dependable, Autonomic and Secure Computing, 2011.

Marjaneh Safaei, Merve Sahan, Mustafa Ilkan, "Social Graph Generation & Forecasting using Social Network Mining", 33rd Annual IEEE International Computer Software and Applications Conference, 2009.

Sanam Kadge, Gresha Bhatia, "Graph Based Forecasting for Social Networking Site", 2012 International Conference on Communication, Information & Computing Technology (ICCICT) , Oct 19-20, Mumbai, India

Martin G. Everett , Stephen P. Borgatti, "Exact colorations of graphs and digraphs", School of Computing and Mathematical Sciences, University of Greenwich, Wellington Street, London, SE18 6PF, UK

Jie Tang, Jimeng Sun, Chi Wang and Zi Yang, "Social Influence Analysis in Large-scale Networks"

R. Albert and A. L. Barabasi. Statistical mechanics of complex networks. Reviews of Modern Physics, 74(1), 2002