EPiC Series in Computing
Volume 69, 2020, Pages 242–255
Proceedings of 35th International Conference on Computers and Their Applications

An Integrated Framework for Friend Recommender System Using Graph Theoretic Approach

Runa Ganguli[1], Akash Mehta[2], Narayan C. Debnath[3], Sultan Aljahdali[4], Soumya Sen[1]
[1]A.K. Choudhury School of I.T., University of Calcutta
[2]Department of Computer Science, The Bhawanipur Education Society College
[3]Department of Software Engineering, Eastern International University, Vietnam
[4]Computer Science Department, Taif University, Saudi Arabia
runa.ganguli@gmail.com, akash93mehta@gmail.com, narayan.debnath@eiu.edu.vn, aljahdali@tu.edu.sa, iamsoumyasen@gmail.com

Abstract

Study of relationships established in social media is an emerging area of research. Online Social Network (OSN) is a collection of social entities carrying a lot of information that enriches the network. A structured modelling of the OSN dataset is required for informative knowledge mining and efficient Social Network Analysis (SNA). Graphical representation of data helps in analyzing the structural properties, study of dense substructure, cluster formation and identifying the numerous types of entities exhibiting associations based on different activity fields. This paper discusses about various ways of graph theoretic representations of OSN including structure-based and content or interaction-based approaches. An integrated framework is proposed in this paper that learns from various user attributes and its associated interactions, network structure, timeline history, etc. from a polarized OSN Graph for generating an efficient Friend Suggestion Recommender System.

1 Introduction

A Social Networking site is an online platform which allows people to build social networks or social relationship with others. The users share their personal and career interests, and snapshots from their day-to-day activities. The social network is distributed across various computer networks which is used to link people and organizations throughout the world. Social networking sites like Facebook, LinkedIn, Instagram, Twitter, etc vary in format and the number of features they offer. Social media analysis (SNA) (Bonchi, Castillo, Gionis, & Jaimes, 2011) involves the congregation and investigation of data obtained from social networking sites. One application of SNA is the tracking of online conversations about products and companies. This enables market researchers to obtain the pulse of an entire community about their products. Social media analytics is defined as, “the art and science of extracting valuable hidden insights from vast amounts of semi-structured and unstructured social media data to enable informed and insightful decision making” (Bonchi et al., 2011). In order to perform different analysis on Social networking commonly graph based models are used for representation.

A Graph G is defined as a two-tuple (V, E) where V is the set of vertices and E is the set of edges. Edge is as an ordered pair of vertices (v_i, v_j), where v_i, v_j ε V. In a paper by Riaz and Ali (Riaz & Ali, 2011) it is shown that the concepts of Graph Theory are widely used to study and model various applications in different fields. They include, study of molecules, atoms and construction of bonds in chemistry. Graph Theory also finds its application in the field of Biology, Sociology, Combinatorial Operational Research, Game Theory, modelling of transport network, etc. There are many algorithms based on Graph Theory like Shortest Path algorithm, minimum spanning tree detection, finding graph planarity, finding adjacency matrices, finding connectedness, finding the cycles in a graph, graph search algorithms (BFS, DFS), etc. These algorithms are utilized in solving not only computer science application problems, but also many real-world problems. The inherent structure of a network can be best modelled using graph which utilizes vertices and edges. Graph Coloring (Runa Ganguli & Roy, 2017) is an important concept that has multiple applications in solving various scheduling problems like time-table scheduling (Runa Ganguli & Roy, 2017)**,** resource allocation problems, optimization problems and so on. The vertex cover algorithm of Graph Theory is used in network security. In a paper [2] authors have discussed role of graph theory related to the issues in Mobile ADHOC Networks (MANET) (Kurkowski, Camp, & Colagrosso, 2005). A graph anonymization framework is presented in (R. Ganguli & Roy, 2018) to combat against IP Spoofing in Ethernet LAN. In addition to these, graph theory is also used in service modelling (Bhattacharya, Sen, Sarkar, & Debnath, 2016)**,** data warehouse applications (Roy, Sen, Sarkar, Chaki, & Debnath, 2013), visualization of substructures using Graph Database (Shrivastava & Pal, 2009), multi-source graph data mining (Koutra, 2017), bioinformatics mining (Tang & Tan, 2011) and so on**.** However, not only graphical modelling of a problem and application of graph algorithms on it, the study of the structural properties of the representation is a broad area of research. Hence it can be said that a graph is inherently the most appropriate data structure to model any social network.

In Graph theoretical approach, SNA is the study of a network where the vertices of the graph are social entities, often termed as actors and the focus is primarily on the structure of the ties (or edges connecting the vertices). Additional information of social entities like status, age, sex, location, etc. are carried out by node properties, whereas edge attributes focus mostly on relationship type, intensity, strength, distance, etc. Facebook is the largest online social network that exists today. Users send friend request to each other and if the request is accepted, the users get linked with each other. This allows the users to directly communicate through wall-posting, photo-tagging, commenting on posts and so on. The inherent graphical nature of Facebook makes it desirable for researchers to pursue a graph-based analysis of the social network.

This paper provides a critical analysis of the various ways of social network data representation using Graph Theoretic Approach. Section 2 briefly discusses the existing methodologies for structure-based and content-based analysis of social network data. An integrated framework for a recommender system that gives friend suggestions after extensive graph mining is proposed in Section 3. Finally, the paper is concluded by discussing the future applications of our proposed framework.

2 Related Work

So far, we have learnt that social media data can be well represented using graph and this graph can be analysed to extract meaningful information. Online social networks provide unparalleled challenges and opportunities for knowledge discovery and data mining. Social network analysis can be done in two different approaches (Aggarwal, 2011)**:**

i. **Linkage-Based and Structural Analysis** – where behaviour of the network is analysed to determine popular or important nodes, communities or strongly connected clusters, establishment of ties or links and growing regions of the underlying network.

ii. **Content-Based Analysis** – all online social network platforms carry huge amount of content-based information in the form of text and images. By mining them efficiently we can obtain additional information about the OSN which would not only help us to quantize the edge weights appropriately but also help us in making informed decisions.

Another important approach of analysis, takes time as an important parameter. These temporal analysis-based algorithms can be categorised into - *dynamic analysis and static analysis* (Aggarwal, 2011). In static analysis, the social network is assumed to change slowly over time, and the whole network in batch mode is analysed over particular snapshots. On the other hand, dynamic analysis is applicable to many networks such as instant messaging networks, where interactions occur continuously at high rate and the scenario is constantly evolving. Online relationships, like their real-world counterparts, tend to evolve with time – either the relationship grows stronger and deeper, or they tend to fade out. Another research work (Viswanath, Mislove, Cha, & Gummadi, 2009) considers the edge weight to be a function of user activity, user pairwise interaction and time – in other words, the edge weights are dynamic in nature. It studies the interaction between users and its effect on the overall network structure. This paper (Viswanath et al., 2009) classifies the users as either low-interaction users or high-interaction users. The relationships involving the low-interaction users did not show much change, whereas those involving the high-interaction users tend to deteriorate. The findings indicate that the number of links deteriorate, showing that users tend to minimise their relationships with others over time, whereas the structural properties of the graph, like average node degree, clustering coefficient, and average path length remain stable.

However, for efficient and effective results, a study combining all approaches should be used. Interaction based approach can be thought of as the combined approach. Contents like user attributes, node types, link types, comments, image contents, etc. are all read along with the topological structure of the network and interaction history over time. A unified graph mining framework (Shrivastava & Pal, 2009) is presented that depicts entities, relationship between entities, dense sub-structure or cluster detection, discovery of frequent substructures, informative and interactive visualization of the mined knowledge from the graph representation. The proposed framework in (Shrivastava & Pal, 2009) is broadly classified into five modules based on its functionalities.

- **Graph Preprocessing**- Integrates data from heterogeneous sources like XML, relational database, web crawl and transforms them to an appropriate graph format suitable for subsequent analysis. If required, some cleaning is also done on the graphs before storing in the graph database (McKnight, 2014).
- **Graph Database**- Data is represented as undirected labeled graph where each vertex corresponds to an entity and each edge corresponds to the relation between two entities. In the proposed framework, the labels associated with vertices or edges may not be unique.
- **Frequent Substructure Discovery**-It targets at the detection of such substructures where the number of times a sub-graph exists in the entire graph is above a specified threshold. FSG, gSpan, SPIN is some of the frequent substructure discovery algorithms.

- **Dense Substructure Extraction**-Dense substructure like maximal clique, bipartite graph generally represents clusters within a graph. Here, mention of few dense substructure extraction algorithms are found like trawling, shingling, connection sub-graphs and maximal clique detection.
- **Graph Visualization-** There are open source graph visualization tools that take combinatorial descriptions of graph as input and gives an interactive visual exploration of the graph as output that facilitates discovery of clusters and associations.

The challenges that this model (Shrivastava & Pal, 2009) has taken is to make the graph mining tool scalable, efficient and robust for mining. Also, the visualization tool should be interactive enough for easy readability. The paper wishes to extend the framework to graph classification techniques and synthetic graph generation using graph generator model. Designing a graph mining tool that helps in discovery of cluster association is important for further research. The previous research work has been extended in (Humski, Pintar, & Vranic, 2019) that dealt with the generation of synthetic datasets. Facebook data isn't readily available because of privacy concerns, so it becomes important to expand upon a small data set and generate a large one. Apart from considering the existence of connection between people, this paper also considers the type and intensity of interaction between various users. This paper, instead of considering binary social graphs, works with Expanded Social Graphs (ESG). In an ESG, multiple edges exist between people, with each edge describing a type/factor of relationship. The edge weights represent the strength of the relationship based on that factor. Such a graph gives a much better representation of human relationships than a binary graph. The rest of the paper deals with synthesizing such graphs. The concept of multiple edges is very important that needs special analysis as the data represented through multiple edges gives granular level information.

Many important concepts of Graph Theory which are relevant to Social Network Analysis have been discussed in (De Nooy, 2012)**.** Among the various behavioral hypothesis discussed in this article, one important analysis states that social entities adjust their behavior, beliefs, attitude and interests according to the other members of the social system they are part of. In other words, the *ego-network* (a network containing the vertex, its neighbors and all edges among the selected vertices) of a vertex affects its behavior to a certain extent. When it comes to study of social relationships, the real scenario of social inequalities is well reflected by the asymmetric ties in the network. A popular node or in other words active node can be determined by *degree centrality*. Popularity or attractiveness of actors or the fact of being chosen can be depicted by the number of incoming ties, i.e. the in-degree of a vertex. Such networks produce informal social hierarchies. A trend is observed here (De Nooy, 2012) that actors tend to associate with other actors of the same rank or prestige by means of positive ties and tend to have negative ties with actors of lower status groups. Strong ties are generally hypothesized to contribute to subgroup formation while weak ties are meant for linking remote parts. The main idea behind improving the efficiency of a network is to strengthen the weak ties, i.e. ensuring information propagation throughout the network. To improve access to information and number of go-betweens, the entire network should remain connected with minimum cut vertices and short-length paths. SNA is concerned with many subjects including study of cohesive sub-group formation tendency (often termed as *homophily*) which can be determined by *transitive triads*, affiliation to common attributes or characteristics. These cohesive subgroups are maximal complete subgraphs called cliques. However, presence of *structural holes*, i.e. absence of a tie between two neighbors of a social entity, can be thought of as incomplete triads which have the possibility of brokerage. This paper (De Nooy, 2012) primarily discusses many studies and theoretical concepts around the local structure of the network, rather than the overall network structure. But the overall structure is only a collection or overlapping of local structures. In practical social ties, status of social entities or actors keep changing with time, thereby impacting the overall network structure. A methodology for ranking users based on parameters such as likes, shares and user comments is presented by Chaudhary and Kumar (Satapathy, Joshi, Modi, & Pathak, 2016). This paper (Satapathy et al., 2016) proposes two ranking techniques – one based on cosine similarity, and the other based on features, like user comments. This helps to determine the reputation of the user, detect the highly reputed users and users with the most negative behaviour.

A generalized Markov Graph model is introduced by Wang et al. in (Wang, Krim, & Viniotis, 2013) and its application in social network classification and social network synthesis are studied. The main findings from this paper state that the degree distribution, the clustering coefficient distribution, and the crowding coefficients are three fundamental statistics for characterizing generic social networks. This paper finds that the clustering coefficient is the result of the dependence between higher order structures, namely the triads, in social networks.

To study the structure of a network in order to determine clusters or highly connected sub-networks is the primary focus in (Krid, Grati, & Robbana, 2016). The study or analysis of real-world networks actually helps in better understanding of the particular properties that are common in OSNs. This analysis leads to the study of Community, which is a set of nodes having strong connectivity among themselves and weak connectivity with nodes outside. The nodes or profiles are in a community if they tend to have same type of interests (film, sports, hobbies, work, etc.). Several studies outline that in OSNs, there are few nodes that are highly connected and remaining are moderately connected. This paper (Krid et al., 2016) proposed a model having three types of node- *private profiles, public pages and groups*. The links that connect these types of nodes are of *friendship, followership and membership*. Friendship links are bi-directional, whereas the latter two are unidirectional. Few configurable parameters of the model include Number of nodes of the network, Proportion of private profiles, public pages in the network, groups in the network, Min- Max degree of isolated private profiles, Min-Max number of members of every group, Min-Max number of followers of every public page, etc. The authors have implemented Facebook data using OSNSim, a new tool that allows users to generate instances of OSNs according to the proposed model (Krid et al., 2016)**.** Parameters are chosen such that the generated instance has the structure of Facebook (highly connected communities, a good number of private profiles, etc.). To evaluate the results of the model, software SocNetV (Social Network Visualizer) which is a cross-platform tool for the analysis and visualization of Social Networks in the form of mathematical graphs has been used. However, this paper leaves scope for future study of information propagation and its privacy protection in OSNs.

One of the most important problems in SNA is classifying users as trustworthy or deceitful. SWTrust Framework is used to generate trusted graphs for the evaluation of trustworthiness in an OSN (Jiang & Wang, 2011).The framework (Jiang & Wang, 2011) comprises of two stages:

- Pre-processing stage which involves the classification of a user's neighbours by their social distance and calculation of the neighbours' priority, based on their topic and target-related degree;
- The Development of a distributed algorithm which leads to the generation of the trusted graph.

A semi-supervised approach of Machine Learning for detecting anomalies and outliers in an OSN is shown by researchers (Reza Hassanzadeh, 2013). The OSN is modelled as a graph and its properties are extracted and studied for detecting outliers in a community. A clustering algorithm is developed in order to group the users present in a network, and then fuzzy logic is applied to calculate the membership score of each user. The user with minimum membership score is then labelled as an outlier. In a study, Catanese et al. (Catanese, De Meo, Ferrara, & Fiumara, 2010) also developed an agent for visiting the friend list of a Facebook user for extracting relevant information regarding the user and his/her friend circle. This phase generates an undirected graph. The structural properties of the graph are then studied on the basis of a few network metrics, like Maximum geodesic Distance, number of connected components and the maximum number of vertices and edges in a connected component. This helps to visualize the data in an efficient way which allows the development of a more robust data mining algorithm for mining graph-based data. A text mining approach to generate an enriched social graph, modelling cross-thread community interactions and interests of Web forum users is proposed in a research work (Anwar & Abulaish, 2012). Message-similarity relationship technique is used to keep track of all similar posts resulting out of deviated discussions in different threads. The paper has also mentioned techniques, that not only consider the existence of relationship between users belonging to separate social communities, but also proceeds to quantify them based on their posts.

The principle of reciprocity is an important factor in determining the strength of the relationship between two social entities. The theory states that the behaviour of any person in social world is determined by an exchange process i.e. humans tend to reciprocate positively to people who have helped them in the past whereas they tend to ignore or respond negatively to those who have rebuked them. The validity of this principle in an online social network is verified in (Surma, 2016). An observation is made in the paper is that greater the number of messages broadcasted by a user, greater is the number of responses received. This paper also hypothesizes that greater the degree of reciprocity shown by a user, greater is the degree of reciprocity received. A user assigns scores and meanings to various interactions, and this score plays an important role in deciding whether future communications will take place between certain users or not. Facebook is studied here (Surma, 2016) as the online social platform and a strong positive correlation between *Reciprocity_Likes_Sent* (likes sent in response to likes received) and *Reciprocity_Likes_Recieved* (likes received in response to likes sent) is found. However, it is also pointed out that social exchanges tend to decrease as people grow old. The paper provides a strong experimental evidence for a reciprocity phenomenon on Facebook.

Estimating friendship or relationship strength between two users cannot be limited to single interaction field. Two users can have connections between them on multiple grounds of activity or interest. The interaction level for different activity fields vary between same user pairs. The weight or the impact factors for different activity fields, play a vital role in computation of the overall relationship strength. The strength for the individual activity fields contribute to the overall friendship score. The overall friendship score is obtained by weighted average of the individual scores. In a study, Zhao et al. (Zhao, Yuan, Li, Chen, & Li, 2012) uses a mathematical model to compute a full spectrum of friendship strength and not mere binary status, i.e. strong or weak tie. For measuring the relationship strength, this paper uses two typical information sources - *User's Profile information* and *Various Interaction Activities.* This paper has worked on Facebook data by crawling through some selected User profiles and has collected their information regarding key attributes. Also, the interaction activities associated with these profiles are well organized in the form of Interaction Activity Document. The model uses LDA (Latent Dirichlet Allocation) clustering algorithm (Zhao et al., 2012) to cluster all the Information Activity Documents, where each generated cluster belongs to some activity field. The seven activity fields described in this paper includes – Diet, Entertainment, Shopping, Sports, Travelling, Work/Study and Others. This paper also proposes a graphical inference model to infer the relationship strengths between different users on various activity fields incorporating user's profile information, the interaction activities among different users and all the activity fields. The work done in (Zhao et al., 2012) can be useful to design personalized people search system where searched results are based on relationship strengths between query sender and discovered people, making the search efficient by enhancing speed and accuracy.

A summary of the techniques used is provided in Table 1 below. It is important to note that the dataset used in all the techniques mentioned above is the Facebook dataset.

Approach	References	Parameters Involved	Objective
Structure Based Analysis	(Viswanath et al., 2009)	Time	• Study the evolution of Online Relationships over time
	(Humski et al., 2019)	Multiple edges in ESG	• Generation of Synthetic Graph • Study of types of relationship
	(De Nooy, 2012)	Degree Centrality, Clustering coefficient	• Detection of dense substructures • Identification of popular nodes
	(Satapathy et al., 2016)	Likes, Shares and Comments on User Posts	• Determining User Ranking or Reputation
	(Wang et al., 2013)	Degree Distribution, Clustering coefficient	• Characterizing Generic Social Networks • Synthesizing new social networks
	(Krid et al., 2016)	Number of nodes of the network, Proportion of private profiles, public pages in the network, groups in the network, Min-Max degree of isolated private profiles, Min-Max number of members of every group.	• Determining Highly Connected Sub-Networks or Clusters
	(Jiang & Wang, 2011)	Social distance and Node Degree	• Classifying Users as Trustworthy or deceitful. • Implementing the entire framework in a distributed system.
Content Based Analysis	(Reza Hassanzadeh, 2013)	Membership score of each user	• Detecting Anomalies and Outliers in OSN
	(Catanese et al., 2010)	Maximum geodesic Distance, number of connected components and the maximum number of vertices and edges in a connected component	• Improving the Visualization of OSN data • development of a better graph mining algorithm.
	(Anwar & Abulaish, 2012)	Message-similarity relationship technique	• Modelling cross-thread community

		interactions and interests of Web forum users in social graph
(Surma, 2016)	Reciprocity_Likes_Sent, Reciprocity_Likes_Recieved	• Verifying the Principle of Reciprocity
(Zhao et al., 2012)	User's Profile information and Various Interaction Activities	• Estimating full spectrum of relationship strength based upon fields of interest

Table 1: A Brief Comparison of Existing SNA Methodologies

3 Integrated Framework

All the different representations of OSN and their findings, as discussed in the previous section, give an insight to an integrated framework that leaves immense scope for exploration. As stated earlier, a well-represented data can be well analysed. So far, we have seen that a network of entities can be analysed either based on the structural property, or based on various interaction activity. This identifies an opportunity to integrate both of these aspects and therefore a research problem can be framed. If a model can represent both the structural and interaction aspects simultaneously then the further insight analysis will be possible. This will help knowledge mining in multiple aspects.

3.1 Problem Description

We consider the problem of suggesting friends on a social network platform. Facebook can be thought as an example, where the network is represented using graphs. Social actors are denoted by nodes and the association between these entities are depicted by links. Each user type node is accompanied with multiple attributes including name, unique ID, location, relationship status, interests, workplace, etc. Again, links or ties are also attributed with information like relationship strength, activity field, etc. Given a pair of user type nodes along with all users specific and existing link specific information a model is framed in this work to integrate multiple parameters for analysis. The main motivation of this research work is to design a model that comes out with relevant friend suggestions and it will be obviously better if more numbers of parameters are considered for analysis. Moreover, this learning cannot be limited to user pair information rather it needs other pertinent information about the underlying network.

3.2 Proposed Model

A framework of an integrated model for Friend Suggestion is presented in Fig. 1 which takes user pairs from an OSN Graph as input. OSN Graph is defined as the graphical representation of an Online Social Network data; here Facebook is taken as instance. The edge weights are obtained after running a text-mining polarity classifier that reads the comment threads, post shares, photo tagging, post tagging, etc. crawled from user walls (which is the visualization of user's activity in Facebook) mainly. All the user attributes that completely define a user type node is taken into consideration along with the fully labelled multiple edges lying between the corresponding nodes. Multiple edges are very important in this OSN Graph as they distinctively denote separate interaction activities between two users by their labels.

The key components of the model are listed below along with their functionalities described in brief:

i. **Polarized Friendship Score Model** – A weighted average of all the edge weights gives an overall friendship score which can be positive or negative. Sign of edge weight depicts the polarity associated with each interaction. By polarity we mean, a friendship score obtained between two friends is based on all positive as well as negative interactions. By positive or negative interactions, we mean those where both of the users like or dislike a given activity field or agree or disagree on a particular post shared on wall. After generation of polarized friendship score named as F-score (well-known statistical model), for each user pair, the OSN Graph is updated.

ii. **Recommender System** – It is the most widespread application of Machine Learning technologies in business. This model learns after several stages of information filtering from user activities in a network over time to predict user preferences. A recommender system in this context suggests friends or predict establishment of links in future that best suits a particular user type. Friend suggestions are not always made by users, they are mostly automated. Machine learning algorithms capture this part of OSN which learns the system well in order to make more effective and relevant friend suggestions. In Fig. 1, we see the model for recommender system takes both structure-based and content-based information from polarized OSN Graph. Content Based information includes user node properties like interests, location, workplace, etc. and structure-based information include friend count, popularity, frequency of activity or interaction, etc. Both are combined to generate friend suggestions.

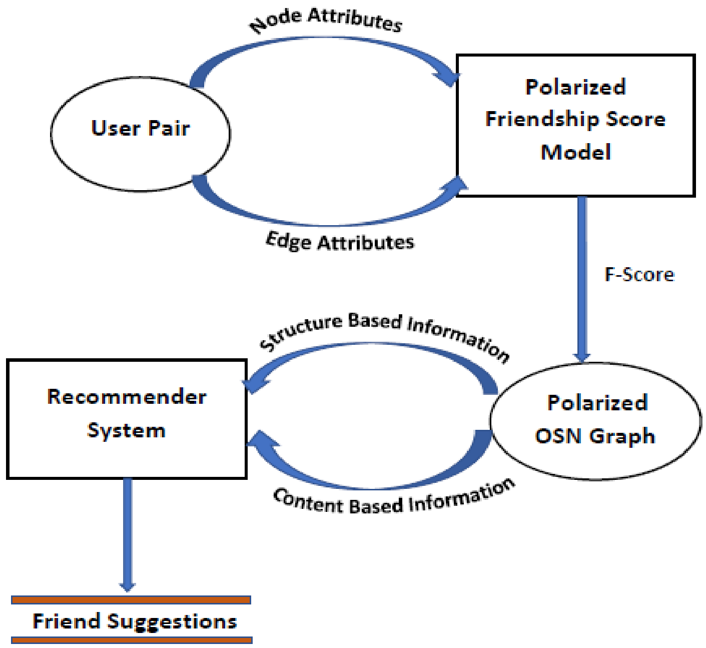

Figure 1: Integrated Framework for a Novel Friend Suggestion System

3.3 Scope of the Model

For designing a recommender system that sincerely makes friend suggestions, we need all the activity related information crawled from the social media dataset along with some inherent node properties. There are multiple direct and indirect parameters that play a major role in this prediction model. Parameters that control the score generation belong to different degree of importance. Some has high impact while other parameters share low. Incorporating all parameters and setting impact factor or value to each, for weighted score calculation is a real challenge. We have seen that this type of network is dynamic in nature and it constantly keeps changing with time. New users enter the network establishing more links while few perform "Unfriend" or "Block" operation breaking ties and creating structural holes in the network. Whatever model is designed for friend suggestion, one must consider this varying social structure for meaningful result.

3.4 Incorporating the Proposed Framework in Big Data Environment

The graphical model of Facebook OSN is best suited to be represented using a Graph Database. Neo4j (Warchał, 2012) is a popular graph database tool which can be used to model any dynamic system where the data topology is difficult to predict. Neo4j (Miller, 2013) finds its applications in the field of real time recommendations, master data management, fraud detection, graph-based search and many more. Social media analysis is a problem area where relationships in data vary both in meaning and value with time. Neo4j (Miller, 2013) implements the graph data model in which fundamental parts are nodes and relations between them. The most important performance benefit of this tool lies in its flexibility in representing semi-structured data and efficient query processing. OSN graphs are normally additive in nature, i.e. new users along with new labels, establishing new relationship ties, keep adding to the overall network without disturbing existing queries.

There are primarily three types of nodes considered in this paper- *User*, *Page* and *Group*. It is to be noted that relationships in Facebook also have types based on the node pairs involved, i.e. *Friendship*, *Followership* and *Membership*. These node types and relationship types can be well reflected using Neo4j which provides scope for node labelling and edge labelling. Moreover, a node or a relation can have multiple labels too. For example, a Page node can be either of celebrity, sports, political interest or food type. Consider the OSN subgraph comprising of node and relationship structure shown in Fig. 2.

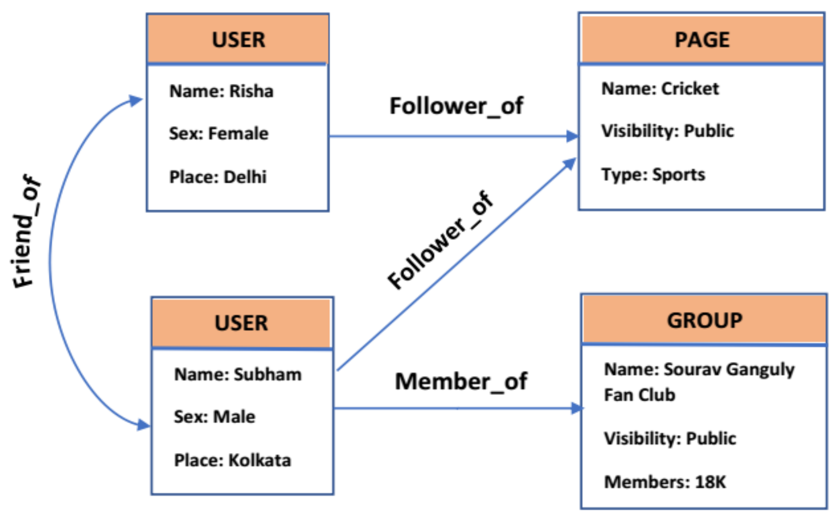

Figure 2: Neo4j Graph Relationship Example

4 Illustration of the Framework in Real Life Scenario

Let us take an example from Facebook OSN dataset and apply the above proposed model for making friend suggestions. Social actors in Facebook are the User Profiles. Each User Profile is connected to many other such profiles. We take any two User type nodes having the following attributes:

Node1, Node2: {UserID, Name, Location, Sex, Relationship *Status, Workplace, Interests, Friend_Count}*

We consider the friendship links between these two User type nodes on multiple grounds, each indicating distinct interaction activity. However, we restrict the case study to 4 interaction activities or interests as {Sports, Music, Food, Movies}. Each edge is defined to have the following attributes:

Edge$[1..n]_{n=4}$ *: { LinkID, Activity_Field, Relationship_strength}*

Activity_Field attribute have values belonging to set {Sports, Music, Food, Movies}.

Relationship_strength gives the intensity of friendship between the two, based on that particular field. Value of this attribute is computed based on a text mining polarity classifier which works on data crawled from both the User Node Profiles.

Positive polarity: [+1 to +n] -> low to high in positive interaction intensity

Neutral polarity: [0] -> No interaction at all on that particular field or multiple interactions with equal magnitude and opposite polarity, which cancel each other out.

Negative polarity: [-1 to -n] -> low to high in negative interaction intensity

Polarized Friendship Score Model works on the above data for the given User pair {Node1, Node2} and computes the F-score by performing weighted average on all the Relationship_strength collected from n connecting edges. Each edge $e_i \in Edge[1..n]_{n=4}$ carries an impact factor based on the frequency and variation of interactions between the two nodes.

The above process is repeated between every User Pair of the entire dataset, thereby obtaining an updated and Polarized OSN Graph. This graph carries not only content-based information, but also Structure based information which identifies popular or weak nodes. Friend Suggestions can be meant for making the less popular users connect to a greater number of friends, or in some other sense, it can be needed to filter the friend suggestions offered to popular users. Thus, based on the demand of application, specific User type nodes are selected and given a set of Friend Suggestion List having people matching to the polarized F-score by a certain threshold value (here we take, threshold value as +3).

Let us consider the following example of Friend Suggestion for Node1, taking Node2 as the mutual friend:

$\text{F-score}_{\text{Node1, Node2}}$: +4
$\text{F-score}_{\text{Node2, Node3}}$: +3.5
$\text{F-score}_{\text{Node2, Node4}}$: +2.3
$\text{F-score}_{\text{Node2, Node5}}$: -4
$\text{F-score}_{\text{Node2, Node6}}$: +4.7

Here, Node3 and Node6 will be added to the Friend Suggestion List of Node1, discarding Node4 and Node5. It is seen that Node4 has a positive polarity but below the mentioned threshold value (i.e. +3), and Node5 is sharing negative polarity with Node 2. While, on the other hand, Node3 and Node6 has positive polarity above the threshold value.

5 Conclusion

In this paper we have reviewed the various techniques available for analyzing social network data using graphs and its properties. The inherent structure of a social network makes it appropriate to be represented or modeled using graph. Along with structural properties, content-based and interaction-based studies between different node types of a graph are also considered for SNA. It helps us to draw valid inferences about the network in general. By taking the interactions between users into account, we realize that the relationships don't remain static but tend to evolve over time. Also, we figure out that opinions associated with a particular interaction activity field between two entities can have a positive or negative polarity. This is important for calculation of friendship score or relationship strength among users in OSN. Using the findings mentioned above, we have proposed an integrated framework that would help provide better friend suggestions to the users in the network. The proposed recommender system will consider various user attributes, activities and past relationships with other users, thereby improving the probability of a user positively responding to friend suggestions. The proposed Recommender System will calculate the compatibility between users by taking into account their areas of interests, their opinions about certain topics and the evolution of their relationship with similar users in the past. This concept can also be applied to provide dating suggestions to users by analyzing their compatibility.

References

Aggarwal, C. C. (2011). An Introduction to Social Network Data Analytics. *Social Network Data Analytics*, 1–15. https://doi.org/10.1007/978-1-4419-8462-3_1

Anwar, T., & Abulaish, M. (2012). Mining an enriched social graph to model cross-thread community interactions and interests. *MSM'12 - Proceedings of 3rd International Workshop on Modeling Social Media*, 35–38. https://doi.org/10.1145/2310057.2310066

Bhattacharya, A., Sen, S., Sarkar, A., & Debnath, N. C. (2016). Hierarchical graph based approach for service composition. *Proceedings of the IEEE International Conference on Industrial Technology*, *2016-May*, 1718–1722. https://doi.org/10.1109/ICIT.2016.7475022

Bonchi, F., Castillo, C., Gionis, A., & Jaimes, A. (2011). Social network analysis and mining for business applications. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 1–37. https://doi.org/10.1145/1961189.1961194

Catanese, S., De Meo, P., Ferrara, E., & Fiumara, G. (2010). Analyzing the facebook friendship graph. *CEUR Workshop Proceedings*, *685*, 14–19.

De Nooy, W. (2012). Social network analysis, graph theoretical approaches to. *Computational Complexity: Theory, Techniques, and Applications*, *9781461418*, 2864–2877. https://doi.org/10.1007/978-1-4614-1800-9_176

Ganguli, R., & Roy, S. (2018). Designing a Graph Anonymization Framework for Secure Packet Transmission in the IP Over Ethernet LAN. *International Journal of Computer Sciences and Engineering*, *6*(9), 650–654. https://doi.org/10.26438/ijcse/v6i9.650654

Ganguli, Runa, & Roy, S. (2017). A Study on Course Timetable Scheduling using Graph Coloring Approach. *International Journal of Computational and Applied Mathematics*, *12*(2), 469–485.

Humski, L., Pintar, D., & Vranic, M. (2019). Analysis of Facebook Interaction as Basis for Synthetic Expanded Social Graph Generation. *IEEE Access*, *7*(c), 6622–6636. https://doi.org/10.1109/ACCESS.2018.2886468

Jiang, W., & Wang, G. (2011). SWTrust: Generating trusted graph for trust evaluation in online social networks. *Proc. 10th IEEE Int. Conf. on Trust, Security and Privacy in Computing and Communications, TrustCom 2011, 8th IEEE Int. Conf. on Embedded Software and Systems, ICESS 2011, 6th Int. Conf. on FCST 2011*, 320–327. https://doi.org/10.1109/TrustCom.2011.251

Koutra, D. (2017). Inferring, summarizing and mining multi-source graph data. *IEEE International Conference on Data Mining Workshops, ICDMW*, *2017-Novem*, 978. https://doi.org/10.1109/ICDMW.2017.136

Krid, D. K., Grati, N. B. S., & Robbana, R. (2016). A new model for Online Social Networks case of Facebook. *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, *2016-July*. https://doi.org/10.1109/AICCSA.2015.7507177

Kurkowski, S., Camp, T., & Colagrosso, M. (2005). MANET simulation studies: the incredibles. *ACM SIGMOBILE Mobile Computing and Communications Review*, *9*(4), 50. https://doi.org/10.1145/1096166.1096174

McKnight, W. (2014). Graph Databases. In *Information Management*. https://doi.org/10.1016/b978-0-

12-408056-0.00012-6

Miller, J. J. (2013). Graph database applications and concepts with Neo4j. *Proceedings of the Southern Association for Information Systems Conference, Atlanta, GA, USA*, *2324*, 36.

Reza Hassanzadeh, R. N. (2013). *A Semi-Supervised Graph-Based Algorithm for Detecting Outliers in Online-Social-Networks*. 577–582.

Riaz, F., & Ali, K. M. (2011). Applications of graph theory in computer science. *Proceedings - 3rd International Conference on Computational Intelligence, Communication Systems and Networks, CICSyN 2011*, *2*(9), 142–145. https://doi.org/10.1109/CICSyN.2011.40

Roy, S., Sen, S., Sarkar, A., Chaki, N., & Debnath, N. C. (2013). Dynamic query path selection from lattice of cuboids using memory hierarchy. *Proceedings - International Symposium on Computers and Communications*, 49–54. https://doi.org/10.1109/ISCC.2013.6754923

Satapathy, S. C., Joshi, A., Modi, N., & Pathak, N. (2016). Preface. *Advances in Intelligent Systems and Computing*, *409*, v–vi. https://doi.org/10.1007/978-981-10-0135-2

Shrivastava, S., & Pal, S. N. (2009). Graph mining framework for finding and visualizing substructures using graph database. *Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining, ASONAM 2009*, 379–380. https://doi.org/10.1109/ASONAM.2009.16

Surma, J. (2016). Social exchange in online social networks. the reciprocity phenomenon on Facebook. *Computer Communications*, *73*, 342–346. https://doi.org/10.1016/j.comcom.2015.06.017

Tang, D. Q., & Tan, Y. (2011). Graph-based bioinformatics mining research and application. *Proceedings - 2011 4th International Symposium on Knowledge Acquisition and Modeling, KAM 2011*, 286–290. https://doi.org/10.1109/KAM.2011.83

Viswanath, B., Mislove, A., Cha, M., & Gummadi, K. P. (2009). On the evolution of user interaction in facebook. *SIGCOMM 2009 - Proceedings of the 2009 SIGCOMM Conference and Co-Located Workshops, Proceedings of the 2nd ACM Workshop on Online Social Networks, WOSN 2009*, 37–42. https://doi.org/10.1145/1592665.1592675

Wang, T., Krim, H., & Viniotis, Y. (2013). A generalized Markov graph model: Application to social network analysis. *IEEE Journal on Selected Topics in Signal Processing*, *7*(2), 318–332. https://doi.org/10.1109/JSTSP.2013.2246767

Warchał, Ł. (2012). Using Neo4j graph database in social network analysis. *Studia Informatica*, *33*(2), 271–279. https://doi.org/10.21936/si2012_v33.n2A.147

Zhao, X., Yuan, J., Li, G., Chen, X., & Li, Z. (2012). Relationship strength estimation for online social networks with the study on Facebook. *Neurocomputing*, *95*, 89–97. https://doi.org/10.1016/j.neucom.2011.06.036