

Pattern Analysis in Dynamic Social Network

Maneesha P Rajeev
Department of Computer Science
Cochin University of Science and Technology

ABSTRACT

Online social networking has become a very popular web application. Many popular online social networks such as twitter, Facebook, LinkedIn, are extremely rich in content and typically contain a tremendous amount of data which can be utilized for pattern analysis. Most of the current efforts in analyzing social networks are done with respect to static structures. Static analysis is done in batch mode over particular snapshot, assuming that the social network changes slowly over time. However, analyzing the dynamic aspects of social networks could provide insight into structural and behavioral characteristics of network. Here, patterns of interaction between users are found based on their 'Like' characteristics.

Keywords

Pattern analysis, Gephi, Facebook, Social Network Analysis.

1. INTRODUCTION

Social network analysis views social relationships in terms of network theory, consisting of nodes representing individual actors within the network and ties which represent relationships between the individuals [2]. Social networks are formed by many types of social interactions daily and continuously between people from different walks of life. Hence social network analysis has become an interesting field of research in order to understand the social structure, social relationships and social behaviors [1]. Static analysis is done in batch mode over particular snapshot, assuming that the social network changes slowly over time [2]. However, analyzing the dynamic aspects of social networks could provide insight into the structural and behavioral characteristics of network [3].

In this paper, analysis of data is done using Facebook data. The patterns of interaction between the users are found based on users 'Like' characteristics. The resulting pattern will help in answering questions such as, what is the key factor that makes a person popular in a network, or how active is a relationship between a popular person and a less popular person is [15]. The main objective of the project is to collect Facebook data for a time period and analyze the data. The collected Facebook data includes user profile details, group details, likes, page fan etc. This is done using Graph API which is a simple HTTP-based API that gives access to the Facebook databases. Facebook provides an SQL-like interface called Facebook Query Language (FQL) that helps us to access these databases. Data collected using Graph API is in JSON format. It is stored locally in a MongoDB database. MongoDB is a NoSQL database engine that stores data in a semi structured JSON format called documents [6].

To understand the network data visual representations of social networks are important. With respect to visualization, network analysis tools are used to change the layout, colors, size and other properties of the network representation [4]. Here, Gephi is used for visualizing the network. The software provides a powerful and interactive visualization and exploration tool kit to explore, understand and manipulate graph structures, to reveal hidden insights of pattern [13].

2. RELATED WORK

People sharing certain similarities tend to form communities in social media. At the same time, they participate in various online activities such as content sharing, tagging, tweeting, etc. These activities leave traces of their social life which would help in understanding the emerging social structures. In this work [8], they advance further to explore different group profiling strategies to construct descriptions of a group, helping explain the group formation.

In this paper [8] they have presented three sensible methods: aggregation, differentiation and egocentric differentiation for group profiling. One possible future direction they suggest on group profiling is to study the internal structure of groups and how they affect the group profiling performance. Another interesting future task is to study the group interaction and how the group profiling can be useful in link prediction.

In paper [3], they have discussed primarily the problem of community detection within social networks. They have discussed community discovery in new domains such as heterogeneous social network, directed social network and dynamic social network.

In paper [1], they define a framework to find evolution of patterns which are regular in dynamic networks. These patterns can be used to characterize the local properties of dynamic networks, and predict future behavior. In this framework, different snapshots of the dynamic network are transformed to a summary graph, and then occurrence rules are discovered for searching for evolution patterns. They have taken protein-protein interaction dataset.

3. IMPLEMENTATION

Data extraction and data analysis are the two important phases of the project. The data extraction process is done for finite time period of six months. Many data analysis techniques are applied on the collected data to find out hidden patterns. Graph Patterns, statistical values of social network analysis and visualization of the network constitutes the result.

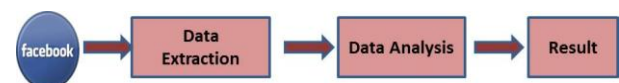


Fig 1: System Architecture

3.1 Data Extraction

'DataHub' application was created using Facebook developer platform. Facebook has got rich amount of social network data. Most of the third party users make use of this valuable data with help of an interface that Facebook provides for analytics called Graph API. Graph API authenticates users who access these data and establishes a transparent data flow portal where users' confidential data remains protected. Facebook supports Heroku server which is a Platform as a service (Pass) where developers can deploy their Facebook applications.

The data is extracted from Facebook users who have accepted the application request. The application is developed in PHP

using Facebook’s Graph API. Backend is implemented using MongoDB. Facebook’s Query language named as FQL is used to fetch data from Facebook server.

The collected data is stored in JSON format which is stored in Mongo DB. Whenever a user adds the application, an access token will be generated for the user and is stored in the database. This access token will be valid for 2 months’ time period. During this time period, the details of the user can be collected using this access token. If a token expires, then the token should be renewed by the user.

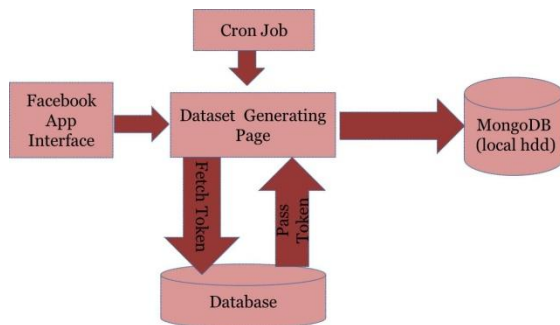


Fig 2: Data Extraction.

3.2 Database

Since the data from Facebook is in JSON format we use MongoDB for storage. Mongo Engine accepts JSON formatted strings as input. It has become easy to use the interface through which data can be queried using simple javascript statements. Moreover, MongoDB provides platform independent data migration and manipulation facilities so that data is easily portable and accessible from anywhere [6]. When the cron job is executed, dataset of all the registered users will be generated. For each user, dataset will be written in JSON Format into a separate text file. All of the generated text files will be compressed into a zip file and will be sent to mail. From there it is transferred into MongoDB.

3.3 Data Collected

Upon registering App in Facebook, they provide access token for Graph API. Graph API triggers queries in the form of FQL. The JSON response from Graph API is stored in MongoDB. The FQL tables that are used to get data are:

- 1) User
 - a) About me: More information about the user being queried.
 - b) Activities: The user’s activities.
 - c) Affiliations: The networks to which the user being queried belongs. The status field within this field will only return results in English.
 - d) Age range: The user’s age range.
 - e) birthday date: The user’s birthday in MM/DD/YYYY format
 - f) Books: The user’s favorite books.
 - g) Current location: The user’s current location.
 - h) Online presence: The user’s Facebook Chat status. Returns a string, one of active, idle, offline, or error (when Facebook can’t determine presence information on the server side).
- 2) Group
 - a) gid: The group ID of the group being queried.
 - b) Name: The name of the group being queried.
 - c) Privacy: The privacy setting of the group being queried, indicating whether the group is OPEN, CLOSED, or SECRET.
 - d) Update time: The last time the group being queried was updated.
- 3) page fan
 - a) Created time: The UNIX time when the user liked the Page
 - b) Page id: The ID of the Page being queried.
 - c) Profile section: The profile section on the user’s profile which contains the Page.

3.4 Tools Used

Gephi: Gephi is an open source data visualization tool developed in Java. It provides several functionalities for the manipulation of network data as well as its representation [13]. The user can interact with the graph representation; manipulate the structures, shapes and colors to reveal hidden properties.

Graph API: It is the interface provided by the Facebook for accessing their database. Graph API works with third party application once the application is registered in Facebook developer platform. It authenticates and validates requests coming from applications and respond in Http-JSON format.

4. RESULT

Social influence can play a crucial role in a range of behavioral phenomena. Identifying the influence factor from the vast availability of data from online social network remains a challenge [10].

This paper analyzes different patterns evolved with respect to users ‘Like’. The user ‘Like’ behavior is analyzed by using most frequently liked objects. The ‘Like’ is classified into two groups:

- 1) Object Likes: Object can be a video, note, link, photo or album. Object id, type of object liked is used for analysis.
- 2) Page Likes: The page id and the time when the user liked the page is used for analysis.

The pattern of community with respect to the authors Facebook network is shown. This is done using Force Atlas’ algorithm in Gephi. Force Atlas makes the connected nodes attracted to each other and pushes the unconnected node apart to create clusters of connection. The blue and violet represents clusters of B.Tech friends, juniors and seniors, which is the largest community. The red color represents cluster of school friends. The green cluster represents M.Tech friends. Yellow cluster represents relatives. With respect to each community we analyze different ‘Like’ patterns.

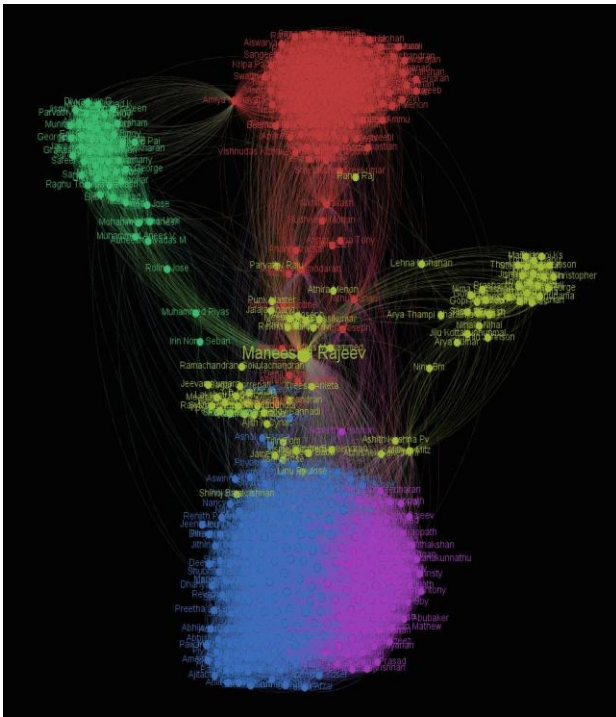


Fig 3: Community Detection.

4.1 Object Likes

4.1.1 Like Pattern of Object 1

Object id = 704662899591652, Type Photo. The dataset value is plotted in Gephi. Here each node represents the person in network. This graph shows which user belonging to which community has liked the object in particular days. The color code of the above graph representing the dates during which the users liked the object is also given.

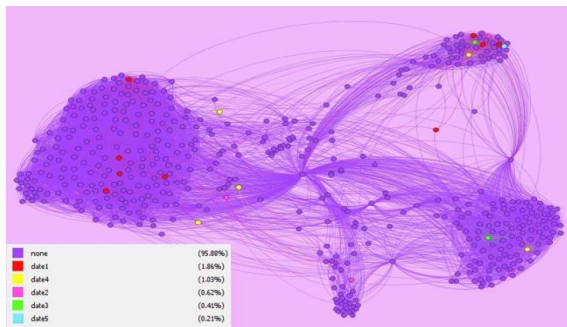


Fig 4: Like Pattern of Object 1

The above graph clearly shows that this particular photo was liked by 9 people on the same day the photo was uploaded. It is mainly from two different communities as can be seen. Considering the two adjacent dates, date 1 and date 2, the object was liked by nodes from two communities on date 1 and it was liked by nodes from three communities on date 2. By comparing each community it is observed that the likes of object 1 were spread among large cluster within no time.

4.1.2 Like Pattern of Object 2

Object id = 10201129045548588, Type Photo. The dataset is plotted in Gephi with respect to this object. An interesting pattern is obtained from this graph. This graph shows that the object was liked only in a particular community. This object was posted on a private group; hence it was not liked by other nodes in the network.

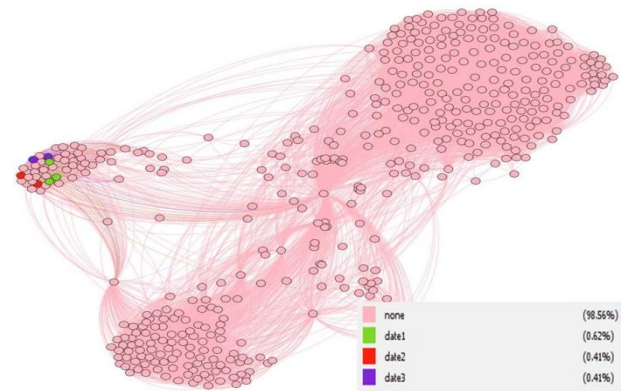


Fig 4: Like Pattern of Object 2

4.1.3 Like Pattern of Object 3

Object id = 10203102020125918, Type Photo. In this graph also, it is seen that object was liked only in a particular community. Following the same procedure, analyzing different object likes with respect to the data set were completed.

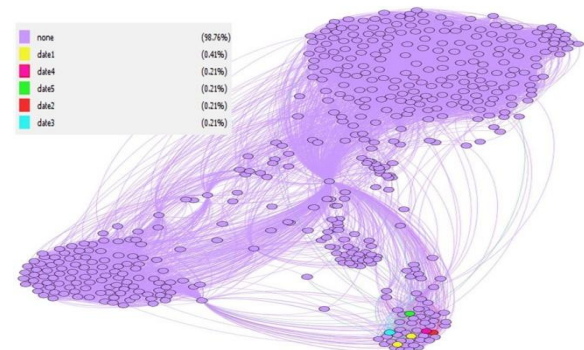


Fig 4: Like Pattern of Object 3

4.2 Page Likes

Now analyzing data set of page 'Likes' the following patterns were obtained.

4.2.1 Like Pattern of Page 1

Page id = 344128252278047, Page name: Sachin Tendulkar. The dataset is plotted in Gephi. Here each node represents the person in network. The graph shows which user belonging to which community has liked the page in particular days.

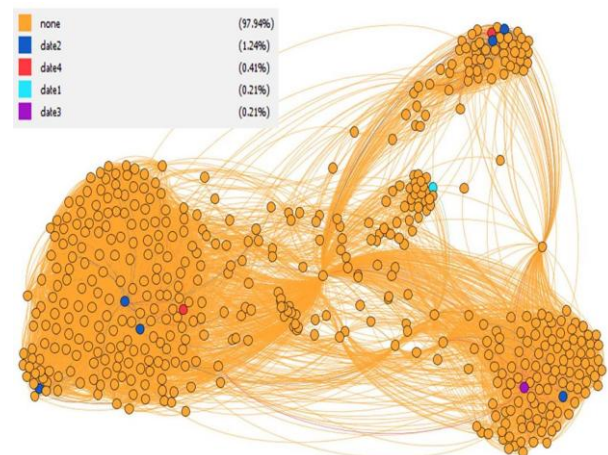


Fig 4: Like Pattern of Page 1

From this graph it is observed that this page was liked by people from almost all communities implying that it is very popular. During a particular day, this page was liked by 6 people from three different communities. So some event might have happened for this condition to occur. Comparing the dates it can be concluded that page was liked by people when Sachin Tendulkar was about to retire from cricket.

4.2.2 Like Pattern of Page 2

Page id = 365947683460934 Page name: Mohanlal. This is a page of Malayalam actor.

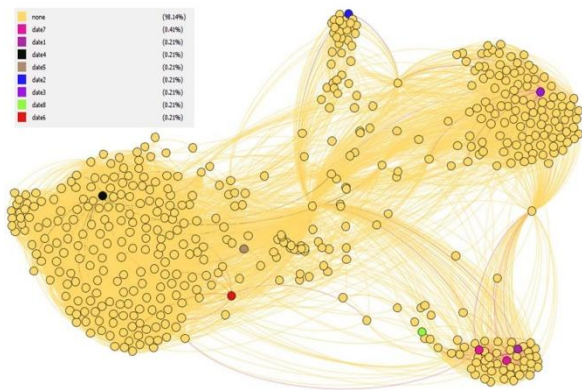


Fig 6: Like Pattern of Page 2

The nodes that liked the page belonged to different communities. Comparing all the adjacent dates, it is observed how the ‘Like’ pattern of this particular page propagated through the network.

4.2.3 Like Pattern of Page 3

Page id = 331288023558058 Page name: Facebook for Every Phone.

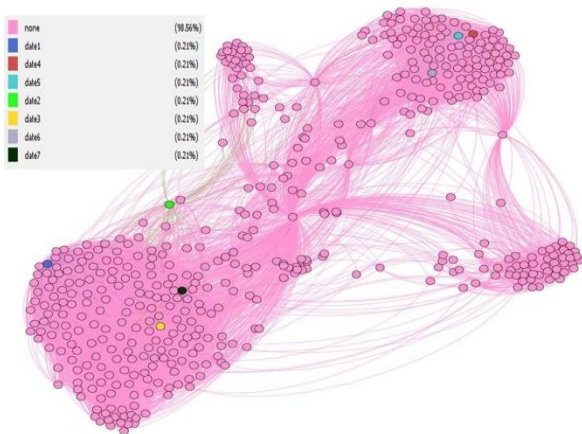


Fig 7: Like Pattern of Page 3

From this graph it is observed that this particular page was liked by people from same community during the first three consecutive days.

All the statistical values like degree, betweenness centrality, page rank, clustering coefficient, number of triangles and eigenvector centrality of each object and pages with respect to different user’s network were examined. For example, the values shown in the tables below represent the statistical values of object one, with respect to social network of three different users whom, indeed, has liked the object.

Table 1: Statistical Value of Object 1 with respect to user1

User 1 (Ma)							
User	Date	Degree	BC	PR	CC	Triangles	EC
F	date2	34	8.829	0.00217	0.846	475	0.0448
A	date3	43	32.034	0.00275	0.676	611	0.0510
J	date1	41	30.869	0.00257	0.725	595	0.0501
G	date1	48	35.613	0.00299	0.637	719	0.0545
Mu	date2	31	1.442	0.00197	0.939	437	0.0429
S	date3	46	32.080	0.00289	0.666	690	0.0533
P	date1	35	12.908	0.00223	0.774	461	0.0454

Table 2: Statistical Value of Object 1 with respect to user2

User 2 (A)							
User	Date	Degree	BC	PR	CC	Triangles	EC
F	date2	34	5.448	0.00128	0.855	480	0.0114
J	date1	42	23.936	0.00157	0.709	611	0.0125
Ma	date1	43	32.0165	0.00163	0.676	611	0.0127
Mu	date2	32	3.299	0.00121	0.899	446	0.0111
S	date1	57	206.206	0.00220	0.471	753	0.0155
P	date3	33	5.0845	0.00124	0.839	443	0.0113

Table 3: Statistical Value of Object 1 with respect to user3

User 3 (F)							
User	Date	Degree	BC	PR	CC	Triangles	EC
A	date3	34	5.523	0.00173	0.855	480	0.0261
J	date1	34	4.0789	0.00173	0.882	495	0.0262
G	date1	35	4.879	0.00177	0.863	514	0.0266
Ma	date1	34	8.581	0.00174	0.846	475	0.0261
Mu	date2	33	3.364	0.00170	0.899	475	0.0247
S	date1	36	6.941	0.00182	0.834	526	0.0269
P	date3	34	9.684	0.00175	0.818	459	0.0260

Where BC: Betweenness Centrality, PR: Page Rank, CC: Clustering Coefficient, EC: Eigenvector Centrality.

4.2.4 Definitions

Betweenness Centrality: It identifies an entity’s position within a network in terms of its ability to make connections to other pairs or groups in a network. It is the number of shortest paths that pass through a node divided by all shortest paths in the network. It shows which nodes are more likely to be in communication paths between other nodes. It is also useful in determining points where the network would break apart. A node with high betweenness has great influence over what flows and does not in the network.

Page Rank: It is an iterative algorithm that measures the importance of each node within the network. The metric assigns each node a probability, that is, the probability of being at that page after many clicks.

Clustering Coefficient: The clustering coefficient, when applied to a single node, is a measure of how complete the neighborhood of a node is. When applied to an entire network, it is the average clustering coefficient over all of the nodes in the network. The clustering coefficient, along with the mean shortest path, can indicate a “small-world” effect.

Eigen vector centrality: This measure takes into account, not only the number of links that each person has, but also the number of links of the connected people, and their links too, and so on, throughout the network. By ranking people by

eigenvector centrality measure, the important people in the network could be found out.

It was analyzed from these statistical values that the users who liked the object or page on the same day and with higher degree value had higher betweenness centrality measure, higher page ranking measure, lower clustering coefficient measure, higher number of triangles and higher eigen vector centrality measure.

5. CONCLUSION

Analyzing the dynamic aspects of social networks could provide insight into structural and behavioral characteristics of network. In this paper, data analysis using Facebook user data, who has accepted the application request, was done. The patterns of interaction between the users are found based on users' 'Like' characteristics. From this approach it is measured how frequent an object or page 'Like' happen in different communities.

An application named "DataHub" was created to collect data from Facebook users. The collected Facebook data includes user profile details, group details, likes, etc. The collected information was stored in the database. With the help of Gephi different patterns of network was plotted with respect to time period. It was observed that some object and page like occur only in one community implying it is a closed group or the object and page liked is of some interest to people belonging to that particular community.

One of the main areas that can be explored is by finding patterns not only with respect to one users' network alone, but also with other users' network pattern too. For this, it is required to create more than one same Facebook application for extracting data and deploying it among more than five active people in Facebook. Another work which can be done is, identifying the features of community by analyzing the patterns. That is by finding out which type of page, object or post is being liked more in a particular community. Using this it is able to predict which object or page is going to be a hit in that community. Association rule mining can be applied to predict the similar behavior of communities. Association rule mining tells the possibility of a new object to become a hit after another object is being liked.

6. ACKNOWLEDGEMENT

I thank Dr. G Santhosh Kumar, my Guide for helping me conceive the idea of the project. He also guided me to execute this particular project. I express my gratitude to Mr. Sarath P.K, my senior, Department of Computer Science, for helping me in the initial stages this work.

7. REFERENCES

- [1] Guimin Qin, Lin Gao, Jianye Yang, Jiajia Li, 2011. Evolution Pattern Discovery in Dynamic Networks. National Key Natural Science Foundation of China.
- [2] Charu C. Aggarwal, 2011. An Introduction to Social Network Data Analytics. Springer Science Business media.
- [3] S. Parthasarathy, Y. Ruan, V. Satuluri, 2011. Community Discovery in Social networks: Applications, Methods and Emerging Trends. The Ohio State University.
- [4] Carlos D. Correa, 2011. Visualizing Social Networks. Kwan-Liu Ma University of California.
- [5] Mohsen Jamali and Hassan Abolhassani, 2006. Different Aspects of Social Network Analysis. Web Intelligence Research Laboratory, Sharif University of Technology, Iran.
- [6] www.momngodb.org MongoDB
- [7] Theus Hossmann, Franck Legendre, George Nomiko, 2011. Stumbl: Using Facebook to Collect Rich Datasets for Opportunistic Networking Re-search. Thrasylvolos Spyropoulos Communication Systems Group, Switzerland, IEEE.
- [8] Lei Tang, Xufei Wang, Huan Liu, 2012. Understanding Emerging Social Structures A Group Profiling Approach. Arizona State University.
- [9] Anu Vaidyanathan, Malcolm Shore and Mark Billingham, 2008. Data in Social Network Analysis. University of Canterbury, Christchurch, New Zealand.
- [10] Eytan Bakshy, Itamar Rosenn, Cameron Marlow, 2012. The Role of Social Networks in Information Diffusion. Facebook.
- [11] Mary McGlohon, 2008. Statistical properties of social networks. , Carnegie Mellon University.
- [12] Smriti Bhagat, Graham Cormode, S. Muthukrishnan, 2011. Node Classification in Social Networks. AT and T Labs research.
- [13] Mathieu Bastian and Sebastien Heymann, Mathieu Jacomy, 2009. Gephi: An Open Source Software Exploring and Manipulating Network. Gephi Web Atlas, ICWSM conference.
- [14] Alan Keller Gomes, Maria da Graca C, 2011. Social Interactions Representation as Users Behavioral Contingencies and Evaluation in Social Networks. IEEE Computer Society.
- [15] Khanh Nguyen, Duc A. Tran, 2011. An Analysis of Activities in Facebook. University of Massachusetts, Boston.
- [16] Ryan Skraba, Johann Stan, Abderrahmane Maaradji, 2009. Developing Compelling Social Enabled Applications with Context Based Social Interaction Analysis. Advances in Social Network Analysis and Mining IEEE.
- [17] Lei Tang, Xufei Wang, Huan Liu, Lei Wang, 2010. A Multi Resolution Approach to Learning with Overlapping Communities. 1st workshop on social media analytics (SOMA, 10).
- [18] Mohammad Ali Abbasi, Sun Ki Chai, Kiran Sahoo, 2008. Real World Behaviour Analysis through a Social Media Lens. Computer Science and Engineering, Arizona State University.
- [19] Eric Sun, Itamar Rosenn, Cameron A Marlow, Thomas M, 2009. Modeling Contagion Through Facebook News Feed. Department of Statistics, Stanford University, Facebook.
- [20] Duncan J. Watts, Steven H. Strogatz, 1998. Collective Dynamics of Small World Networks. Cornell University, Nature vol 393.
- [21] Keith M Hampton, Lauren Sessions Goulet, Cameron Marlow, 2011. Why Most Facebook Users Get More Than They Give. Pew Research Center's Internet and American Life Project.

- [22] Moria Burke and Robert Kraut, Cameron Marlow, 2011. Social Capital on Facebook: Differentiating Uses and Users. Carnegie Mellon University ACM.
- [23] Lars Backstrom, Ravi Kumar, Cameron Marlow, 2008. Preferential Behaviour in Online Groups. WSDM '08, ACM 978.
- [24] Johan Ugander, Lars Backstrom, Cameron Marlow, Jon Kleinberg.
- [25] Structural Diversity in Social Contagion. PNAS and Facebook, 21 February 2011.
- [26] Wenfei Fan, 2012. Graph Pattern Matching Revised for Social Network Analysis. University of Edinburgh and Harbin Institute of Technology, ICDT Berlin, Germany.
- [27] Mansoureh Takaffoli, Justin Fagnan, Farzad Sangi, Osmar R. Zaiane
- [28] Tracking changes in dynamic information networks. International Conference on Computational Aspects of Social Networks (CASoN) , 2011
- [29] Hao Yun Huang, Qize Le, Jitesh H. Panchal, 2011. Analysis of the Structure and Evolution of an Open-Source Community. Journal of Computing and Information Science in Engineering, Vol. 11
- [30] Thomas Steiner , Ruben Verborghy, Joaquim Gabarr Valls, 2011. Adding Meaning to Facebook Microposts via a Mash-up API and Tracking Its Data Provenance. Rik Van de Walley Universitat , Spain,IEEE.
- [31] Mohammad J. Zaki, Mohhamad Al Hasan, 2009. A Survey of Link Prediction in Social Networks. Department of computer science Indiana - Purdue University.