

POLITECNICO DI MILANO
Scuola di Ingegneria dei Sistemi



POLO TERRITORIALE DI COMO

**Master of Science in
Management, Economics and Industrial Engineering**

Using Social Media Data Analysis for Decision Support

Supervisor : *Ms. Michela Arnaboldi*

**Master Graduation Thesis by:
Barira Hanif (755869)
Farhan Javed (755875)**

Academic Year 2011/2012

POLITECNICO DI MILANO
Scuola di Ingegneria dei Sistemi



POLO TERRITORIALE DI COMO

**Master of Science in
Management, Economics and Industrial Engineering**

Using Social Media Data Analysis for Decision Support

Supervisor : *Ms. Michela Arnaboldi*

**Master Graduation Thesis by:
Barira Hanif (755869)
Farhan Javed (755875)**

Academic Year 2011/2012

Acknowledgements

We would like to thank Prof. Michela Arnaboldi for giving us the opportunity to work on this research project. We express our deep sense of gratitude & indebtedness towards our mentor and guide for her support.

I would also like to acknowledge the encouraging attitude of my friends & other staff members of Politecnico Di Milano who helped me to complete the dissertation.

This research work relied heavily on research tools and real data. We are in debt to Branding Bees for their full support in allowing us to use their tools and data. We would like to thank Saad Salman, CEO Pak Servers for allowing us to carry out his digital media strategy.

Finally, We would like to thank our family for their love and support. This dissertation is dedicated to them.

Table of Contents

<i>i Acknowledgements</i>	2
<i>ii Table of Contents</i>	3
1 Introduction	4
2 Problem Statement	5
3 Proposed Framework	6
4 Literature Review	9
4.1 Social Network	9
4.2 Social Network Analysis	10
4.3 Social CRM	24
4.4 Social Network Marketing	26
5 Methodology	39
5.1 Categorization of Data in Web 2.0	39
5.2 Data Extraction and Analysis	46
5.3 Web 2.0 Category C	69
6 Case Studies	70
6.1 Pak Servers	70
6.2 Chicken Hut	75
<i>References</i>	79

1. Introduction:

The concept of social media and data analytics is gaining popularity as the Web 2.0 is expanding and social communities on the internet are evolving. The potential of this social cloud on the internet can be estimated by its exponential growth rate and its popularity among the masses. The convenient access to the users, their data and network is the key feature of a decision support model based on social media and the traditional ways cannot measure or monitor the processes effectively. Hence, in this paper, we discuss the extensive data sources available on Web 2.0 cloud, categorization of these data sources according to their typology and analysis, analysis of the web 2.0 data for mining the relevant information and then measuring the KPIs to track the performance.

2. Problem Statement:

The vast amount of data and the growing numbers of web platforms are the future of technology. This amount of data is growing exponentially and the effective measures to utilize this data for future decision support and analysis is becoming a challenge. This is because of the typology of data these platforms are creating as well as dynamics of the data and platforms. Many tools are also being created to make use of this data and the utilization of these tools to the best is becoming an expertise.

The most popular among these web 2.0 platforms are social media, social networking site etc. This dynamic platforms are the touch points of the companies with their consumers and the data from these social platforms could serve as source to know about the consumer, their preferences, likes, dislike, trends, behaviors, opinions etc.

Another trend in web 2.0 is the growth of e-commerce. Nearly all the businesses are introducing their businesses online to take advantage of the growth of online consumers. The reason is that they can not only market their products effectively online, but also track their performance, online sales as well as measure KPIs for their company.

The data from these two web 2.0 platforms is tremendous, but are not complete enough to support a decision if analyzed separately. However, if we integrate the data from all the web 2.0 sources, we can not only extract useful information required for decision support, but also can track the performance of post-decision.

The research is based on finding an effective method to collect the web 2.0 data, integrate it and analyze it using the effective tools available for web data analysis. Also, the research includes measuring the results of the decision by measuring ROI of the web 2.0 and social media investments.

3. Proposed Framework:

The study and proposed research on the project will be carried out in the following simplified steps.

STEP1	Categorization of Web 2.0 Sources into pre-defined categories.
STEP 2	Implementation of relevant data extraction techniques from the users touch-points on these web categories using different data extraction tools. This requires thorough know-how of the existing web platforms and their technical and non-technical requirements.
STEP 3	Integration of the data from all these sources from web categories to build a dedicated Database for decision support.
STEP 4	Using BI and data analytics tools to extract the useful information .Data mining and Analysis for information extraction and trend finding. Analyze and extract useful information from this data by identifying the correct metrics and measuring them. The analysis will be performed at various levels with the help of different social media and web 2.0 tools. The results derived from these tools will be compared to identify any inconsistencies.
STEP 5	Deducing Results for supporting decision.
STEP 6	Take measures to keep monitoring the results on these platforms Measuring the results of the decision.

The methodology will be based on web 2.0 data analytics, business intelligence and the use of relevant tools. The objectives of the proposed thesis can be achieved in the following steps:

1. *Development of different user applications for various web 2.0 platforms to develop an effective database.*
2. *Extract data from the social media cloud with the help of sophisticated Social media tools and applications.*
3. *Measurement of customer engagement on these social platforms, identify influencers and potential brand advocates. Take advantage of the ideas and innovative users online by means of crowd sourcing.*
4. *Analyze the data from the social CRM database to identify and measure some key metrics for tracking performance and effectiveness of a decision.*

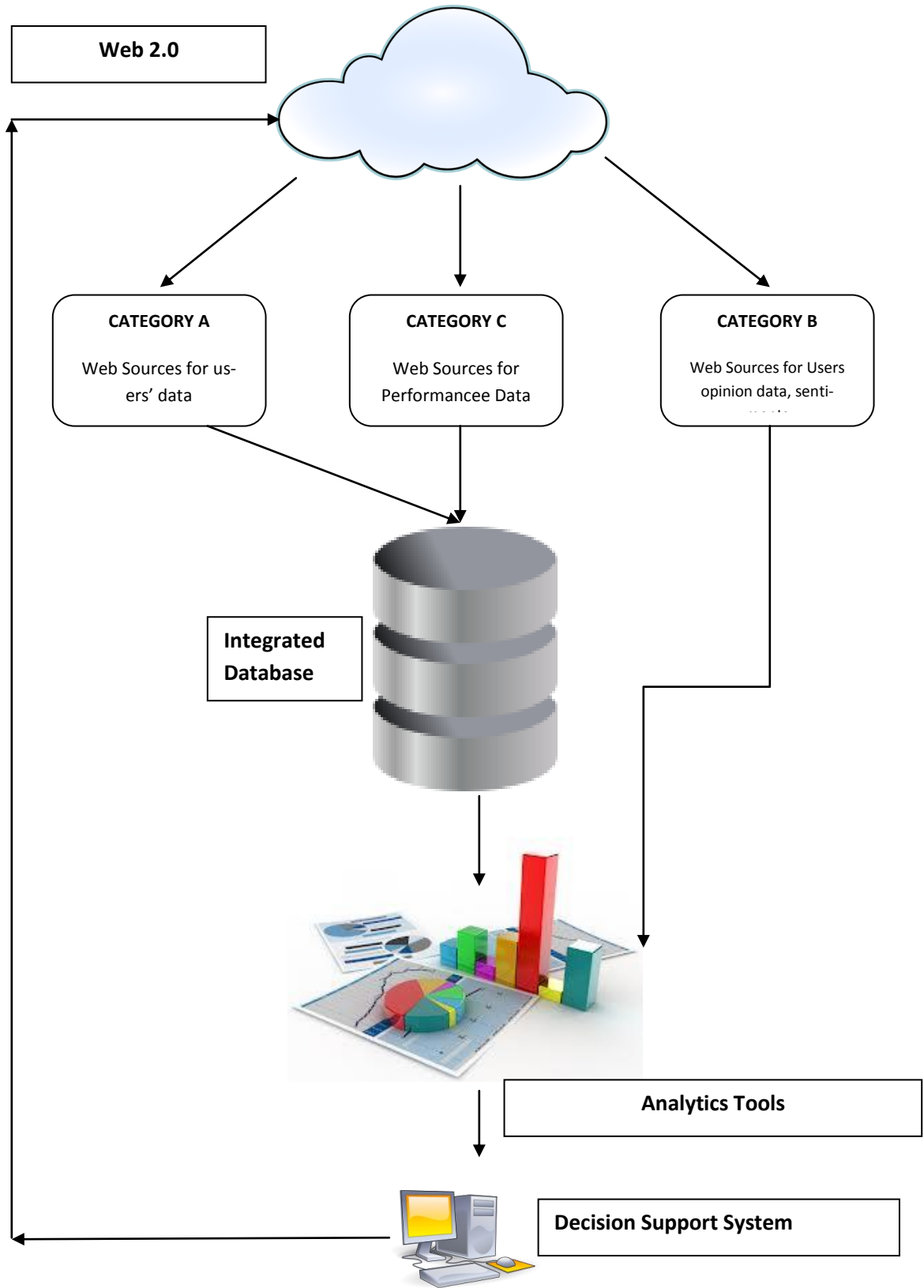


Fig (3.1) Methodology MAP

4. LITERATURE REVIEW:

4.1. *Social Network:*

Social networks are defined as web based platforms aimed to fulfill the socializing requirements of people by bringing them together in a network which to share their common interests. These networks are designed in such a way where the connection between users is established after mutual acceptance.

Once the users establish a connection, they can interact with each other. Such type of connections form a network in which each user is a node and each node is well-connected to the other node (**FIG 4.1**). These networks can be public as well as private. Following are the mostly used social network types:

ISN: A closed social network designed not for public but for a certain community, e.g a social network designed for the employees of a company.

ESN: It is a publicly open platform for all people to join and interact with each other and also the most common type. Examples are facebook, MySpace, LinkedIn etc.

Social networks are growing with an exponential growth rate with respect to the number of users. This is the reason why new social platforms are being developed every day and the current social platforms are evolving each day to satisfy this enormous number of users and keep up with the fast pace of the online world of social networks.

The importance of these social networks can be emphasized by the fact that users are sharing an enormous amount of data on these networks which includes their personal information, interests, activities, network connections etc. which can be useful for companies to know them better.

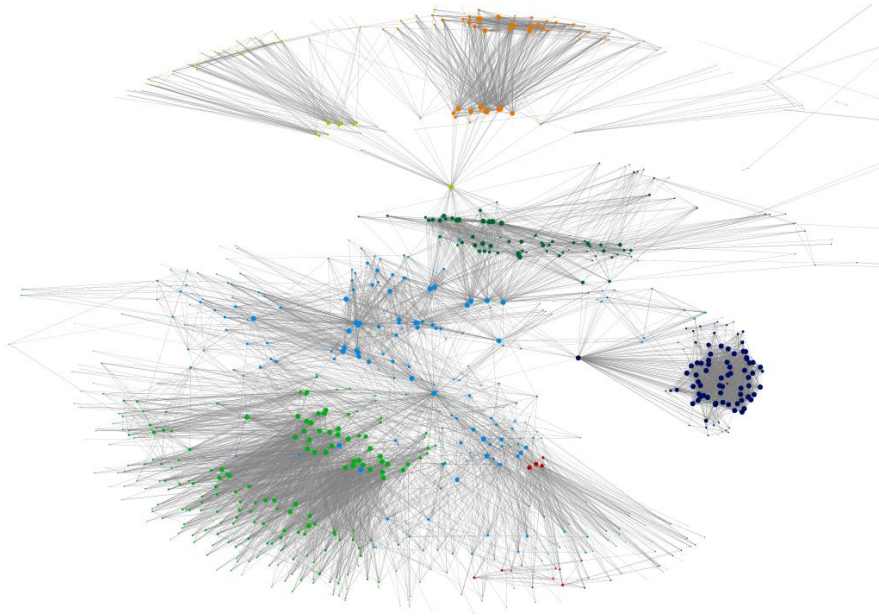


FIG (4.1) Visual Representation of network connections (Nodes) in a Social Network

4.2 Social Network Analysis (SNA)

SNA is an analysis in which social relationships (offline or online) are analysed under the theories of social networks which makes these ties among individuals as a complex network of connections. Each individual plays an important role in this network and SNA identifies the roles and importance of these individuals (referred to as Nodes in SNA terminology). Social networks extend to different levels, from individuals to groups, families, companies, nations etc. And the resulting networks are very complex and difficult to understand. Hence, SNA techniques have been introduced to understand and resolve this complex system in order to make use of it.

Social network Analysis Metrics are the metrics used to define the structure of a social network, the kinds of relationships between them and the ties between them.

4.2.1. SNA Objectives:

The main objectives of performing the above mentioned SNA techniques are to understand the social network structure, the flows of data, communication and information within a network.

Following could be the main objectives to perform SNA techniques:

4.2.1.1. Social Network Graphical Representation:

Social networks can be represented as a combination of graphs and matrices. The graphical methods used to represent the elements of social networks and matrices can be used to represent the relationships between these elements. The main reason for this formal representation is that it provides many benefits and simplifies the analysis.

Advantages of graphical and matrices Representation:

Following are the advantages of representing Social Networks as graphs and Matrices:

- Graphs and matrices are compact and systematic, hence simplify our interpretations of patterns among them.
- They can be saved as data structures for computing purposes especially if the data set is very large.
- We can apply mathematical rules and conventions on the graphs and matrices and it is easier to model the network.

4.2.1.1.A. GRAPHICAL REPRESENTATION:

Social network can be represented as graph with nodes as edges and the ties between them as links. By definition, a social network is composed of the following main constituents which could be represented graphically:

- **Nodes:**

There are certain node types in a social network. These nodes can be persons, groups, key words, web pages, events etc. before performing the analysis, it is important to identify which kinds of nodes will be the most useful and which nodes can be isolated.

- **Links:**

The interaction or connections between the nodes are called links. Links can be different depending upon the relationship between these nodes. Some examples of links are: friendships, family ties, work relationship etc.

- **Population:**

The population of a social network is the total number of individuals forming the network. In technical terms, the total number of nodes that comprise a social network is the total population of a social network. This number could be increasing or decreasing depending on the growth rate of the network. E.g. the current population of facebook users is 664,032,460. The

population of a social network is extremely diverse, hence before performing SNA, one must consider the part of the network within some Boundary.

- **Boundaries:**

There are two types of boundaries in a social network:

- a. *Priori*: Boundaries created by the actors themselves (clusters, groups formed on the basis of a common interest)
- b. Boundaries created by analysts to simplify analysis (based on geography or ecology)

- **Data Structures:**

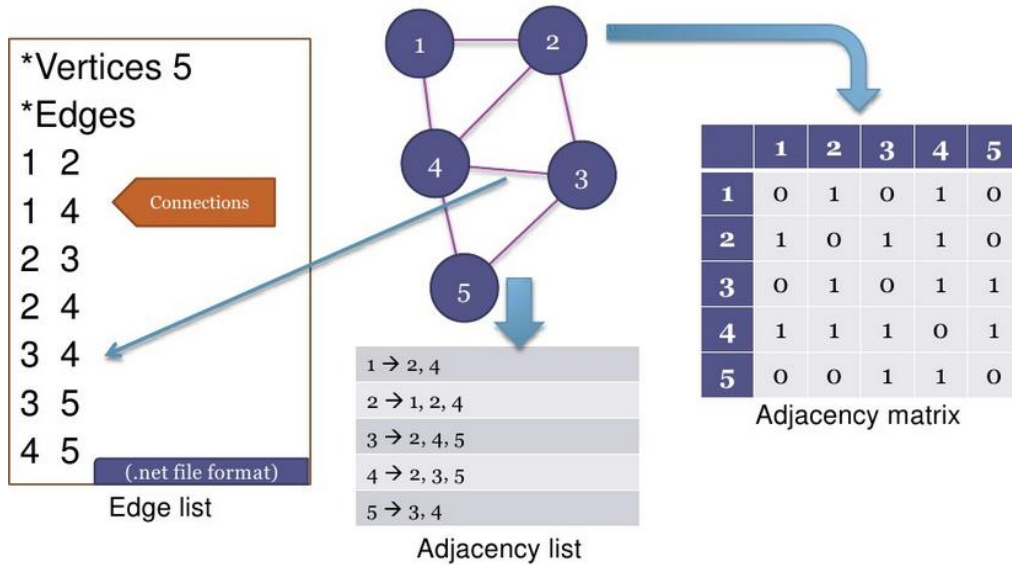
There are many methods to store the above mentioned parts of a social network as data structures for computing purposes.

- **Dynamic Data:**

The nodes in a social network are constantly communicating with other nodes via links. This data is dynamic since it is being generated and changed every second. This data is very useful since it provides us the information of the users on the social network. It can be easily accessed via various tools by accessing different APIs of social network and could be used for further analysis.

Apart from that, a graphical representation of the social network could also help us defining:

- Link orientation
Directed Links (arcs, e.g emails sent), Undirected Links(edges, e.g co-authorship etc.)
- Weights on edges
- With typed nodes



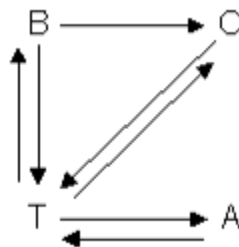
Fig(4.2) Network representation as Graph

3.2.1.1.B. REPRESENTATION OF SOCIAL RELATIONS AS MATRICES:

Social network data is not like conventional data where each unit is independent and not a part of a probability samples. Representing social relationships as matrices could be useful because it makes computation easier if the data set is large and actors are many. Large data sets can be stored as arrays and matrices. Following are the mostly used matrices by Social network Analysts:

i. Adjacency Matrix:

Adjacency matrix is a square matrix ($i=j$) which represents the adjacency of nodes i.e. who is adjacent to whom in the social space. It also tells us how many paths of different lengths exist between players.



Fig(4.3) An example to Represent the social relations in a matrix

The figure above represents a simple social network. We can represent this social network as a simple binary matrix:

	B	C	T	A
B	-	1	1	0
C	0	-	1	0
T	1	1	-	1
A	0	0	1	-

Representation of Relations in Fig(4.3) as a matrix

Other types of matrix representations are:

ii. Incidence Matrix:

The incidence matrix or un-oriented incidence matrix is the visual representation of relationships between any objects A and B in a social domain. Each row is represented by an element of A and each column is represented by element of B.

Example:

If the elements of A are represented by (A1, A2, A3)

And Elements of B are represented by (B1, B2, B3) , a 3x3 matrix if the elements of A&B are not related can be represented as follows:

	B1	B2	B3
A1	0	0	0
A2	0	0	0
A3	0	0	0

Incidence matrix if Elements of A&B are not related

	B1	B2	B3
A1	0	1	0
A2	0	0	0

A3	0	0	0
----	---	---	---

If elements of A A1 is related somehow to B2 and incident is occurred and the value at this position will be equal to 1(Shown in the matrix above).

	B1	B2	B3
A1	1	1	1
A2	1	1	1
A3	1	1	1

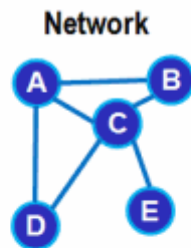
Incidence matrix if Elements of A&B are all related

iii. Laplacian Matrix:

also called admittance matrix or Kirchoff matrix contains both adjacency and degree information of nodes in social space.

iv. Distance matrix:

If a social network consists of 'N' number of nodes, this NxN matrix represents the distance between each node.



	A	B	C	D	E	Σ
A	0	1	1	1	2	5
B	1	0	1	2	2	6
C	1	1	0	1	1	4
D	1	2	1	0	2	6
E	2	2	1	2	0	7
Σ	5	6	4	6	7	28

Fig(4.4) A Simple Example of Distance Matrix

4.2.1.2 Visualization & Characterization by indicators:

Visualization of a Social Network gives an overall visual representation of the network, how the network looks like, how convergent it is etc. There are many approaches to visually represent it. One of the main objectives of SNA could be to measure some key metrics of a social network. These measures give the characteristics of a social network in quantitative measures. We can measure these metrics at global and local levels in a network.

a. Indicators at Global Level:

These indicators are measured for the whole social network. Some examples are:

- Number of nodes
 - Number of edges
 - Diameter
- } **Density**

b. Indicators at Local Level:

We can measure some indicators at local level, i.e. at the node level (or a sub-group of nodes) some examples are:

- Number of Neighbors (**Degree**)
- Distance: Measurement of the shortest path between subgroups, nodes etc.
- Measure whether a node is central or not (betweenness centrality etc.) which are discussed below in detail.

4.2.1.3 Community detection:

SNA techniques can be used to identify the sub-networks or clusters formed within a network which are well-connected with each other than the rest. These communities share a common interest and could be useful to identify the commonality among them.

There are many tools and algorithms available to identify the communities in a social network.

4.2.1.4 Other Objectives:

- Identify important nodes or edges in terms of communications and information flows within a network because of these nodes. Measure the relative positioning of a node within a network with respect to other neighboring nodes.
- Determine Influencers within a sub-network or community

- Determine the nodes which generate the maximum flow of information to the other nodes. These nodes are the source of information for many other nodes and are well placed and well-connected in a network. These nodes can be identified after performing the position analysis.

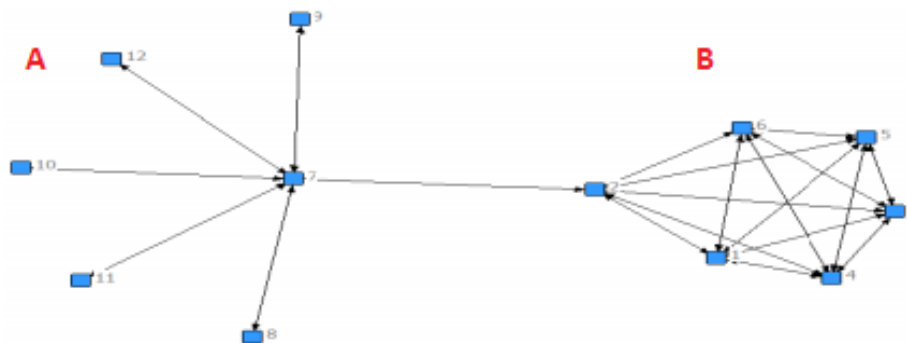
4.2.1. SNA techniques:

Based on the objectives above, there are many techniques to achieve these objectives. These techniques can fall under any of the objective category defined above, but the techniques are randomly described below:

4.2.1.1. Clustering coefficient:

This coefficient is used to measure the level of convergence of a social network and defines the structure of a network. It measure how well connected the neighbors of a node are.

Clustering coefficient = No. of connections in a neighborhood of a node/no. of connections if the neighborhood was fully connected



Fig(4.5) An Example to Understand Clustering

In the figure above, we can visualize the two sub-networks. Sub-Network A is clearly divergent whereas Subnetwork B is convergent to form a cluster. The overall clustering coefficient of the network is medium. But if we divide the network into sub-networks, the scenario will be different.

Network A: In network A, the nodes are not well-connected with each other but divergent from the node 7. Node 7 plays a key role in the network because the communication is limited and only via node 7. It is very important to identify these types of nodes in a divergent network which are the key source of information and communication for many other nodes. The clustering coefficient for A is lower.

Network B: The clustering coefficient of sub-network B is clearly higher than A. This means that each node in the sub-network is well-connected with the other, accessibility is higher, information is well-distributed among the sub-network and communication between nodes is efficient.

4.2.1.2. Centrality measure in SNA:

Centrality measure in a social network is an important aspect of SNA which helps in assessing the relationships and the flow of information between the nodes, the location of the node in a social network and evaluation of its importance. However we can measure the centrality on the individual and network level.

3.2.2.2.A. INDIVIDUAL CENTRALITY MEASURE:

Centrality Measure of a node is the level of the power of the node in a social network. It measures the tendency of the node of being central and well-connected to the rest of the nodes. Following are the three most important individual centrality measures:

- **Degree Centrality:**

The degree of a node is the number of direct connections the node has in a social network. The degree centrality can also measure the network activity of a node.

$D_c = d(i)/n-1$ where n = no. of users in a network and $d(i)$ is the degree of the node

- **Betweenness Centrality:**

It measures the tendency of a node to be the 'tier' or the 'link' between other important nodes. If a node serves the purpose of being the pathway of communication between many nodes, its betweenness centrality is higher. This node will control what flows between the network.

- **Closeness Centrality:**

This measures the accessibility of a node to the other parts of the network in a direct and indirect way.

- **EigenVector Centrality:**

A node's eigenvector centrality is the sum of the eigenvector centralities of all the nodes connected to it. It is similar to the methodology of Google's website ranking system.

We can use the above measures to determine the following useful aspects of a node:

Centrality Measure	Interpretation
Degree Centrality	No. Of people directly reached by the person
Betweenness Centrality	Likelihood of being the most direct route between two people
Closeness Centrality	Reach of the person to everyone in the network
EigenVector Centrality	Connection strength with other well-connected connections

Table() Using individual Centrality Measures

4.2.2.2.B. NETWORK CENTRALITY MEASURE:

The network centrality measure indexes the tendency of a single node in a network to be more central than all the other nodes of the network. The graph centrality measure of such a type of analysis are based on referencing the centrality of the nodes in a network to the most central node. The network centrality measure is a useful method to understand the social network structure.

- **Bridge Centrality (BC):**

This metric of SNA discriminates the bridge nodes in a network and measures how well-located a node is between well-connected regions. BC makes use of Betweenness centrality and Bridge Centrality coefficient. It differs from Betweenness centrality because it measure the importance of a node with respect to information flow through the node.

- **Cliques:**

Cliques are the subnetworks within a network in which the nodes are well connected to one other than the rest. This cluster of sub-networks are formed because of common interests, hobbies etc. N-clans, n-cliques, k-plex are few of the Clique metrics .

4.2.1.3. Mathematical Operations on Matrices:

As discussed before, we can represent the relationships between nodes and vertices in the form of matrices. Once we have achieved this, it is easier to perform some operations on these matrices to detect some unusual patterns in the network. Some operations include the following:

- **Inverse of a Matrix:** $A^{-1} = 1/|A| * [A^t]$
- **Transpose of Matrix:** Switching the rows and columns of a matrix is called transpose of a matrix and is represented by A^T
- **Matrix Addition and Subtraction:** Subtraction and addition of corresponding elements of a matrix.
- **Matrix Correlation and Regression:** This operation determines the correlation between different matrices
- **Matrix multiplication:** For multiplication of two matrices, the columns of the first matrix must be equal to the rows of the other matrix.

The above mentioned operations if performed single can not give interesting results but if combined together can help us detect various patterns across social network space.

4.2.1.4. Modality and Levels of Social network Analysis:

It is important to understand that there are networks embedded inside a network and communities within a vast network. Hence, an individual is a part of a network and many other sub-networks which are embedded in a network and so-on. Such types of networks are called 'Multi-Modal'. Usually data of a social network gives you information about various types of entities (individuals, communities, organizations etc.), hence we can assume that any social network is multi-modal by nature.








4.2.2. SNA Tools:

Various software tools and data mining techniques have made it easier to implement the SNA techniques and perform various types of analyses efficiently and accurately. Some examples of softwares used nowadays are summarized in the following table mapped with the objectives of SNA they serve:

Tools	SNA Objectives
Gephi	Interactive graphic and visualisation platform
GUESS	Used for visualization
Tulip	Visualization, clustering, and extension plug-ins. Can handle upto 1 million edges and 4 million vertices
GraphViz	Visualization
UCInet	Statistical and matricial analysis, calculation of indicators, hierarchichal clustering, uses Pajek and Netdraw for visualization
Pajek	Windows program for analysing large networks
Igraph	Creation and manipulating graphs
NetworkX	Study of structure, dynamics and manipulation of social networks
JUNG	Java Universal Network /Graph framework
MatlabBGL	Network Analysis, Graph-

Table() Tools available for SNA

The tools mentioned above can fulfill some of the objectives based on their flexibility. However, their performance varies depending upon the functions. Some are good for one objective and vice versa. We have ranked the tools according to their functionality on the table below:

	Pajek	Gephi	NetworkX	igraph
 Input/output	+	++	+	+
 Attribute handling	+	+	++	--
 Bipartite graphs	+	-	+	+
 Temporality	+	+	+	-
 Visualization	++	++	-	++
 Indicators	+	+	++	++
 Clustering	+	--	--	++

++ Mature Functionality

--Weak Functionality

Table() how to choose the right SNA tool

4.2.3. Uses of SNA:

A lot of research is being carried out to use these social networks to facilitate people. Some of the research is being carried out in the following areas:

4.2.4.a. Modeling Organization as Social Network:

An organization can be modeled as an interlocking connection of networks of entities like people, tasks, groups and resources. This matrix of organization can:

- Determine the relationships between these entities
- Using SNA techniques to identify the links of the entities outside organization
- Data collection which doesn't include the attributes of these entities but also the ties between them
- Determine the team leader or the influencer among these entities
- Measure the performances of individual entities

4.2.4.b. Organizational Risk Assessment:

Organizational risk is defined as the set of all the possible risks which an organization can face during its operation. A complete set of organization risk factors includes all the possible risks coming from its various units such and can be tagged as:

- Market/Reputation risk
- Financial risk
- Operational risk
- Legal/regulatory risk
- Strategic risk
- Technology risk
- People/culture risks
- Fraud risk

Since SNA techniques can be used to model an organization as a large social network, we can also model the risk associated with each unit with the help of Social Network Analysis. By analyzing the socio-cultural trends within the social network in an organization can help us estimate the risk at a vast perspective. Apart from that, studying the social patterns among various units within an organization with the help of SNA techniques available can also help us model the systems of risk and also rate them by assigning them proportional weights.

4.2.4.c. Collaboration schemes in knowledge sharing:

Determine an effective communication between researchers, doctors on a social network to share their research on diseases like cancer, malaria, and dengue. SNA also helps in determining their links, the most effective nodes, and the isolated ones. The isolated one once determined can also be brought together so that knowledge sharing is maximized.

4.2.4.d. Social Capital:

Social capital is a term coined for the value that social network potentially holds for businesses. However, a definite way of measuring this value has not been proposed so far. Many companies and Social Media Firms are trying to estimate the return on investment (ROI) of a social media campaign for businesses.

Many other uses of SNA are psychology, electronic communication etc.

4.2.4.e. Marketing:

SNA techniques are very useful for personalized marketing and targeted advertisements. This topic is discussed in detail in the next chapters. Cluster identifications (groups, communities etc. based on common interest, age groups, locations etc.) in a social network can help in identifying the right targets for a product.

4.2.4. Future Study and Research Topics:

- Dynamic Fuzzy Social Network Analysis
- Contextualized Analysis(Analysis by type, node)
- Correlation Analysis (among different social media performance KPIs and business value)
- Regression Analysis (Determining a model based on the correlation of the KPIs and the variables)

4.3. Social CRM:

CRM:

Customer relationship management (commonly known as CRM) is defined as the framework of a company to sustain, improve and expand its customers' portfolio. This objective is achieved by using different processes and Information Systems frameworks deployed and used by various departments of a company like sales, marketing and support. The traditional CRM includes the thorough processes of data collection of the customers, identifying their needs and devising the company strategy to fulfill these needs. This process is usually customer centric, with limited touch points of the customers with the company, aimed to create value for the company by enhancing the relationship with the customers.

CRM can have different aspects depending on the functions it serves.

	CRM Aspects		
	CRM Operational	Collaborative CRM	Analytical CRM
CRM Functionalities	Marketing Marketing Plan Campaign Management Lead Management Sales Opportunity Management Service Order Management Activity Management Service Order Management Service Contract Management Warranty Management	Social CRM Call Center E-Commerce Digital media Web <u><i>(CRM functions to be developed)</i></u>	Data Collection Information Search Analysis Algorithm Data mining Business Intelligence <u><i>(CRM functions to be developed)</i></u>

4.3.1. CRM based on Social Media:

Social CRM is the term coined for CRM techniques/ methodologies designed for the social customer. The online presence of the customers has provided companies with a vast number of possibilities of engaging them on their social platforms hence raising the opportunities of knowing the customers, their preferences and needs by many-folds. The technological flexibility and ease with which the companies can interact with the users is replacing the conventional methods of CRM with techniques using the social networks.

Social CRM can be defined as ‘the effective use of social networks and their evolution to engage the social customer on these platforms, monitor these networks to identify their needs, re-

spond to these needs, strengthen relationships with the customer, identify and target potential customers.

4.4. Social Network Marketing:

Before understanding the marketing techniques on Social Media, we must understand the evolution of web platforms, and the background of the development of social media. Social media is completely dependent on the technological frameworks of web and internet platforms. In the following section, we define the boundaries of Social media and how technology plays and will keep playing its role in the evolution of social media.

4.4.1. Web 2.0:

Web 2.0 is a second generation of world wide web (www)and the term was coined in 2004 in a conference brainstorming session between O'Reilly and MediaLive International to name this latest medium of internet cloud. web 2.0 by definition is ' a dynamic internet platform developed for information sharing, interaction, collaboration and communication across the internet cloud (worl wide web)'.

Web 2.0 is much more dynamic as compared to web 1.0 and has features which are more user-centric and technically user- friendly. The features of web 2.0 depict web as a platform where the users are the content generators which makes web 2.0 a more interactive platform than web 1.0

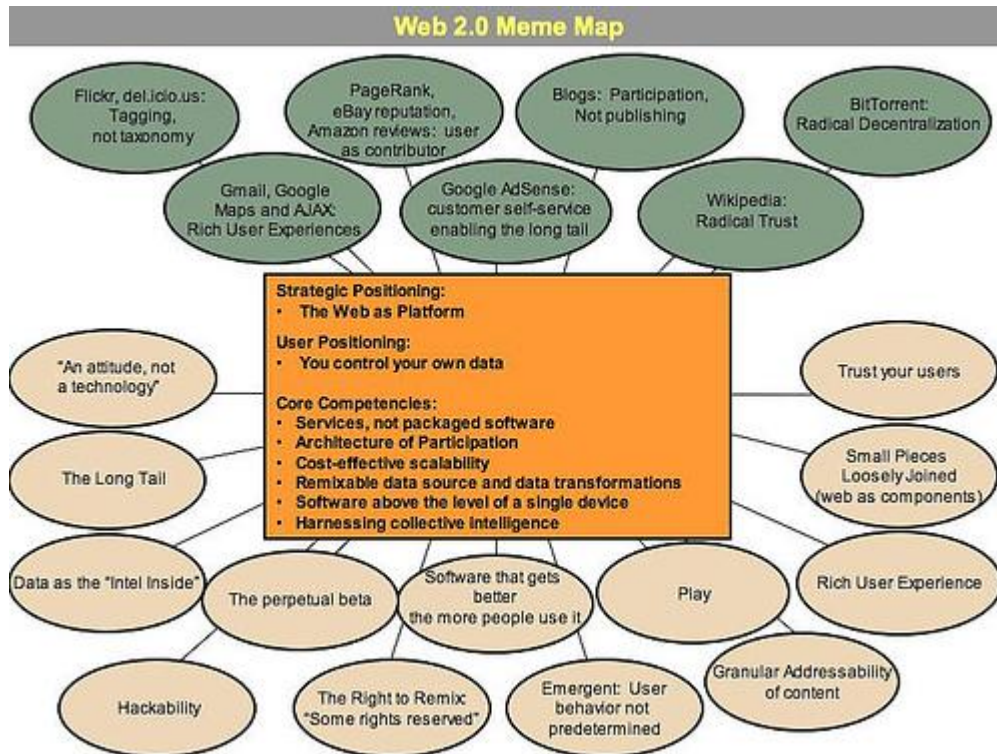


Fig (4.6) Web as platform (web 2.0 map developed at a brainstorming session at FOO Camp during a conference at O'Reilly Media)

Some features of web 2.0 are discussed below:

-Users as Content Creators:

users are the key content creators of the web. This feature of web 2.0 led to many leading website services such as Wikipedia, Flickr etc. Also it paved the way for creating the blogosphere which is a platform for the wisdom of the crowds.

-Users as Co-Developers:

Web 2.0 applications help the users to develop dynamic systems and its key competence is operations. Hence various software development languages (also called scripting languages) like Perl, Python and PHP are being used. Open source web development modules also help the users to actively participate in the development phase empowering them to tailor the applications according to their requirements.

-Easily Programmable Modules:

The idea of web 2.0 is syndication, and it support lightweight modules which can be easily used and developed by the users. Hence, it supports the feature mentioned above i.e users as co-developers.

-Electronic Devices Accessibility:

Another key competence of web 2.0 platforms is that they are not limited to PC. The applications developed on web 2.0 can be run on any mobile device and electronic platform that makes it more feasible and accessible for all kinds of users at any time.

-Database management:

Database management is the key aspect of a web 2.0 companies and the challenges are to collect data , control it and display it. Hence, every key application of web 2.0 comes up with a different data class and data management system, hence these softwares can be better termed as ‘Infowares’.

Web 1.0	Web 2.0
DoubleClick	Google AdSense
Ofoto	Flickr
Akamai	BitTorrent
mp3.com	Napster
Britannica Online	Wikipedia
personal websites	blogging
evite	upcoming.org and EVDB
domain name speculation	search engine optimization
page views	cost per click
screen scraping	web services
publishing	participation
content management systems	wikis
directories (taxonomy)	tagging ("folksonomy")
stickiness	syndication

4.4.2. Social Media:

With the evolution of web 2.0, new horizons of application development and web-based communities were developed, also known as 'Social Media'. Social Media is a platform based on web (Web 2.0) and mobile technology which allows dynamic and two-way communication between communities, organizations and individuals. Social Media websites can provide different services. Following are some examples of social media websites based on their functions:

- **Social Bookmarking:**

We can tag websites or rate them by tagging them (examples are: Del.icio.us, Blinklist, Simpy)

- **Social News:**

This type of social website rates the articles on their websites by users' votes and comments (Examples are: Digg, Propeller, Reddit)

- **Social Networking:**

This is probably the most famous kind of social website based on web 2.0. it allows networking among individuals and groups of communities formed on the basis of a common interest. (Examples are: Facebook, Orkut, MySpace, Hi5, Last.FM etc.)

- **Social Photo and Video Sharing:**

This type of social website allows users to create content by posting photos and videos as well as interact by posting comments (YouTube, Flickr)

- **Wikis:**

This type allows users to create content by writing the articles for the web-based information websites (Wikipedia, Wikia)

Apart from the websites mentioned above, any website which allows users to interact, communicate and develop content can be categorized as social media.

4.4.3. E-commerce:

E-commerce is the short for electronic commerce which is the name given to the business developed using the internet and other modes of latest digital communications like mobile, emails etc.

With the evolution of e-commerce, technologies like automated data collections systems, inventory management, Electronic Data Interchange (EDI) and Internet marketing also evolved. Now the biggest challenge is to integrate the social media with the e-commerce and develop a correlation between e-commerce KPIs with the Social Media Marketing efforts.

4.4.4. Online Advertising Frameworks:

There are many online advertising models being followed and implemented where the client is being charged differently on different terms by the advertiser. However, with the evolution of social networks, new ADV models are being materialized in order to take advantage of the presence of a large amount of potential customers on these social networks.

Before discussing the SNS based models, we must understand the traditional web-based ADV models:

4.4.4.1. Web-based Traditional ADV Models:

There are many adv models being used in the internet marketing paradigm. Some of the most commonly used are mentioned below:

- ***PPC (Pay-per-Click):***

The most famous model for this type of ADV is the Google's ad words model. The ads are displayed with some search results and the publisher charges for the number of clicks. The ads are arranged by the order of their number of clicks and this model is based on ranks-by revenue. The client can also buy some keywords in order to display their ads on the search results of a certain keyword.

- ***PPM (Pay-per-Impression):***

It is a time-based business model which is used to display ads on various online spaces for creating brand awareness or promoting the brand. An example of this kind of model is banner ads on MSN IM application interface. Banners are usually displayed on the interface so that customers using the ads can see it.

- ***PPA (Pay-per-Action):***

This kind of model is an advanced level of the PPC and PPM because it is based on the action (like purchase, call to the sales center etc.) of the customers upon seeing the ad online. The amount to be paid for an action (anything beyond a click) in PPA should be defined by the advertiser and any action by the customer should be reported.

4.4.4.2. SNS –based ADV Models:

These type of models are the new emerging type of social network based ADV. There are a lot of advantages of SN-based advertising among which following are the most important:

Advantages of Social Media Marketing:

- The large presence of online customers on a single web platform (i.e. web based social network).
- Ads are easier to propagate because of the large network (hence higher approach)
- Personalized Advertising: Vast amount of accurate and personal user data available publicly which makes it easier to target the customer accordingly by the means of profiling.

There are many social media Adv techniques which use the current internet marketing models for payments but target the customers present on social media.

Some of the most commonly used SNS based Adv models are mentioned below:

- **Facebook Ads:**

This model is similar to PPC & PPM in terms of functionality but facebook makes use of the advantages of SNS in terms of customer targeting. It is a form of personalized advertising where the ads are displayed to the users according to their personal preferences, their browsing behavior etc. facebook profiles profiling of users

- **Contextual ADV Models:**

This type of advertising is usually used for websites and blogs spaces where the ads are displayed on the relevant websites. The relevance is determined by buying some keywords. By buying Goolge's AdWords, the advertiser can link to the Google's Ad sense which will advertise the ad on the publisher's blog post. The advertisers have to bid for the Adwords to buy them, and the competition for certain words is higher than the others.

- **Status Ad:**

This is the type of advertising where the publisher uses its status space available on various social networks (e.g 132 characters in Twitter, facebook status etc.)for advertising. The publisher can not only promote the ad to his/ her friends in his circles but also to extended circles in the network.

4.4.4.3. Multi-Level Marketing:

Multi-level marketing is an SNS based marketing technique proposed to extend the marketing in the different social circles in a social network. The vast social network makes it possible to achieve this goal. It follows the traditional PPC or PPM Adv models for rewarding mechanism. Principal of multi-level marketing in a social network is based on using the personal status space of an individual as the ad space instead of a web ad space. To explain it in a simplified way, we have divided the procedure in the following levels:

Level 1:

- User1 can use his personal status space to market an ad. Since his friends and community are more interested in his status than other random ads ap-

pearing on the web, the chances of the ad being viewed and clicked are higher. The more his friends click, the user will accumulate rewards.

Level 2:

- A friend of user1, user2, also wants to promote the ad and earn money. Hence, user1 gives him his personal number and user2 and receives the ad message to display on his status. He also receives his personal id number. User1 also gets reward on the basis if performance of user2.

Hence, the chain follows and somehow it is similar to **Markov Chain** in the distribution methodology.

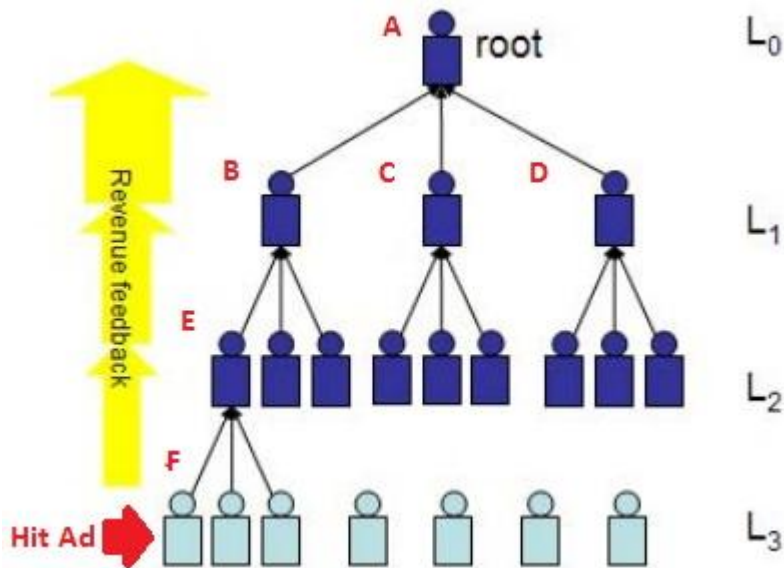


Fig (4.7) Multi-Level Marketing

Feedback Formula of Multi-Level marketing:

Consider the figure above for calculating the feedback revenue formulas of multi-level marketing. If A invited B, and B invited E and E invited F, and because of the promotion of F’s ad, someone Hit the Ad, the following procedure will be followed to distribute the fee among these publishers at different levels:

X: fee of ad hit 1st time

P: Rate of publisher

1-P: Rate of fee retained for upper layer L0

E.g if $P=0.7$,

L3: Publisher 'F' gets = $0.7X$, the rest $0.3X$ must be distributed among E,B and A

L2: Now, $X=0.3X$, Publisher 'E' gets = $0.7(0.3X)$, the rest $(1-(0.7(0.3X)))$ must be distributed among B&A

L1: Now, $X= (1-(0.7(0.3X)))$, Publisher 'B' gets = $0.7((1-(0.7(0.3X))))$, the rest $1-((1-(0.7(0.3X))))$ is for A

L0: A gets= $0.7(1-((1-(0.7(0.3X))))))$

4.4.4.4. Text Mining and Sentiment Analysis on Social Media:

One of the measures that Social Media marketing can provide is analyzing the feedback or the opinions of online customers. Nowadays, customers are very keen about expressing their feelings about a product on social media platforms, be it positive or negative. Such platforms could be a page on facebook, blog or any other social platform. Different types of analysis can be performed on the textual data available on these platforms, generated continuously by users, using text mining techniques.

- Subjective Analysis:

A subjective analysis is the precursor to sentiment Analysis. In this type of analysis, the subject of a small phrase or sentence is deduced and Subjective units are separated from the corpora.

- ***Sentiment Analysis:***

The sentiments and the polarity of the opinions on a certain product is estimated using this type of analysis on a certain social media platform.

- ***Discourse Analysis:***

This analysis is analyzing the sentiments of people on different sets of social media platforms (blogs, social networks etc) and summarizing them. Sentiment analysis is a subset of discourse analysis. This type of text analysis is usually carried out on political issues, news etc.

4.4.4.5. SEO (Search Engine Optimization):

Search Engine Optimization is one of the key marketing strategies for an internet based business or E-commerce. This technique uses the power of a search engine (e.g. Yahoo or Google) to direct the users online searching for a certain word towards their website, hence increasing the online visibility of the website. The more the users are directed towards the website, the more chances of getting potential customers. There are many techniques for SEO, some of which are also considered to be illegal (Termed as Black Hat Techniques in SEO) and could lead to the banning of the website operation. Companies can assess the performance of their website by using web analytics KPIs or also by evaluating its ranking in SERP (Search Engine Research Page).

Many companies try to achieve the objective of prominent online presence by using keywords, and using aggressive inbound marketing strategies. Some of the SEO methodologies are mentioned below:

- Cross Linking
- Indexing
- Title Tag & Meta Data
- URL Normalization
- Canonical Link Element
- 301 Redirects

- Key-words:



4.5. E-Commerce Data Analysis:

4.5.1. Web Log:

A Web log is a recorded file generated which contains records of all the data viewed by each visitor to your Web site. The data is in the raw format and contains information of each transaction , each visitor and other important information of the the visitor like the following:

- URL which referred the visitor to the website
- Arcane (i.e. the OS of user and browser) etc.

A sample log file is shown in an image below:

```

1 #Software: Microsoft Internet Information Services X.X-
2 #Version: X-
3 #Date: 2010-03-24 07:00:01-
4 #Fields: date time s-sitename s-computername s-ip cs-method cs-uri-stem cs-uri-query s-port cs-
5 2010-03-24 07:00:01 ZZZZC941948879 RUFFLES 222.222.222.222 GET / - 80 - 220.181.7.113 HTTP/1.1
6 2010-03-24 07:00:23 ZZZZC941948879 RUFFLES 222.222.222.222 GET /2009/12/im_not_mean_im_just_ar
7 2010-03-24 07:00:32 ZZZZC941948879 RUFFLES 222.222.222.222 GET /terminal-blank.gif - 80 - 217.
8 2010-03-24 07:00:32 ZZZZC941948879 RUFFLES 222.222.222.222 GET /grep-options.gif - 80 - 217.23
9 2010-03-24 07:00:32 ZZZZC941948879 RUFFLES 222.222.222.222 GET /terminal-cat.gif - 80 - 217.23
10 2010-03-24 07:00:32 ZZZZC941948879 RUFFLES 222.222.222.222 GET /terminal-pwd-cd.gif - 80 - 217.
11 2010-03-24 07:00:39 ZZZZC941948879 RUFFLES 222.222.222.222 GET /robots.txt - 80 - 95.55.207.95
12 2010-03-24 07:00:39 ZZZZC941948879 RUFFLES 222.222.222.222 GET /rss-short.xml - 80 - 173.45.23
13 2010-03-24 07:00:43 ZZZZC941948879 RUFFLES 222.222.222.222 GET /2009/08/22-things-you-dont-knc
14 2010-03-24 07:00:44 ZZZZC941948879 RUFFLES 222.222.222.222 GET /screen.css - 80 - 98.88.35.133
15 2010-03-24 07:00:44 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/rss-header-red.gif - 80 -
16 2010-03-24 07:00:44 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/logo.jpg - 80 - 98.88.35.133
17 2010-03-24 07:00:44 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/input-emailsend.jpg - 80 -
18 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /images/cm-ebook-banner.gif - 80 -
19 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/bg.jpg - 80 - 98.88.35.133
20 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/bg-top.jpg - 80 - 98.88.35.133
21 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /21things/checkout-login.gif -
22 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/topnav-contact.jpg - 80 -
23 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /21things/portent-email-sub.gif -
24 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /rss-header.jpg - 80 - 98.88.35.133

```

Fig (4.8) Multi-Level Marketing

As shown in the figure above, the data in a web log file is extensive, hence we need some analytical tool or software to derive the information from such raw set of data. There are however two formats of a log file:

Format Type	Description	Advantages	Disadvantages
Common Log Format	The standardized Log format which contains the data in a standard format.	Supported by nearly all web servers	Inflexible, Doesn't contain Browser info, time taken, Referer fields
Extended Log Format	Similar to Common Log Format but adds information like Browser information	More information than CLF	Lesser information than the rest formats
W3C Standards		Flexible and much extensive	Not supported by many web servers

4.5.2. Web Log Analyzer:

Web Log Analyzer or web log analysis software is a tool which parses a web log file into meaningful indicators which could be useful to measure in order to track the performance of an e-commerce website.

These indicators are as follows:

- Number of visits and number of unique visitors
- Visits duration and last visits
- Authenticated users, and last authenticated visits
- Days of week and rush hours
- Domains/countries of host's visitors
- Hosts list
- Number total pageviews
- Most viewed, entry and exit pages
- Files type
- OS used
- Browsers used
- Robots
- HTTP referrer
- Search engines, key phrases and keywords used to find the analyzed web site
- HTTP errors
- Some of the log analyzers also report on who's on the site, conversion tracking, visit time and page navigation.

The data from log files could be stored in a dedicated database off server and could be used for further analysis. Some examples of free or open source and proprietary web log analyzers and their features are mentioned below in the following table:

Name	Platform	Supported Databases	Tracking Method	Opensource/ Proprietary
Analog	C	Log-file based	Web Log	Open Source
AWStats	Perl	Log-file based	Web Log	Open Source
Mint	PHP	MySQL	Cookies via Javascript	Proprietary
LogZilla	Linux	MySQL	Sys log ng	Mixed
Splunk	Windows/Solar/BSD/Linux	Proprietary Database	Web Log files	Proprietary

5. METHODOLOGY:

The methodology is based on using a web based social media data analytics tool to support a decision. Companies need to foresee the public reaction to a certain campaign, product or a decision which can be crucial to the company. Hence social media data from various sources will be used to monitor, evaluate and support a decision. The results after the decision is taken, will be then compared with the forecasts reports etc.

5.1. Categorization of Data sources on the Web 2.0:

As discussed before, the Web 2.0 is a vast ocean of information which is constantly growing at an enormous speed. This is due to the crowd-sourcing feature of the web 2.0 which has given the masses the liberty to contribute their share of drops in this vast ocean by creating its content. The enormity of web 2.0 can be estimated by the fact that its growth rate is exponential and the statistics mentioned below:

Name	Stats	Category
Google	<p>1,000,000,000,000 (one trillion) - approximate number of unique URLs in Google's index (source)</p> <p>2,000,000,000 (two billion) – very rough number of Google searches daily (source)</p> <p>\$110,000,000 – approximately amount of money lost by Google annually due to the “I’m Feeling Lucky” button (source)</p> <p>(source: Google)</p>	C2C, general web
Wikipedia	<p>2,695,205 - the number of articles in English on Wikipedia</p> <p>684,000,000 – the number of visitors to Wikipedia in the last year</p> <p>75,000 - the number of active contributors to Wikipedia</p> <p>10,000,000 – the number of total articles in Wikipedia in all languages</p>	C2C, general web

260 – the number of languages articles have been written in on Wikipedia

Youtube

70,000,000 – number of total videos on YouTube (March 2008)
200,000 – number of video publishers on YouTube (March 2008)
100,000,000 – number of YouTube videos viewed per day (this stat from 2006 is the most recent I could locate)

C2C, general web

Blogosphere

133,000,000 – number of blogs indexed by Technorati since 2002
346,000,000 – number of people globally who read blogs (comScore March 2008)
900,000 – average number of blog posts in a 24 hour period

C2C, Expert opinion

Twitter

1,111,991,000 – number of Tweets to date (see an up to the minute count [here](#))
3,000,000 – number of Tweets/day(March 2008) (from TechCrunch)

C2C, General Web, Could be used for determining influencers, Also expert opinions

Facebook

200,000,000 – number of active users
100,000,000 - number of users who log on to Facebook at least once each day

C2C, General Web, Could be used for determining influencers, Also expert opinions

Digg

236,000,000 – number of visitors attracted annually by 2008 (according to a Compete survey)
56% - percentage of Digg's frontpage content allegedly controlled by top 100 users

C2C, General Web, Could be used for determining influencers, Also expert opinions

The above table shows masses of data which needs to be categorized in order to reduce processing time and reduce the resources used for information analysis. Hence, to make the process effective, we categorize each web source and analyze it separately because the results of analysis for each category could be very different.

It is very important to arrange and categorize this vast data and information source to deduce correct results and to save time and resources. In order to simplify the analysis, we can divide the web sources in the following two main categories:

	Description	Use in the methodology	Examples
CATEGORY A:	Web Sources to Collect users data.	Social CRM, Marketing. (Clustering of users based of communities, interests, location and other relevant characteristics)	LinkedIn, Facebook, Twitter, Myspace
CATEGORY B:	Web Sources to measure Opinions.	Opinions of experts and masses, Sentiment Analysis, Future Pattern of trends etc.	Blogs, Websites, Social Media Fan Pages and Groups, Interactive forums on the websites etc.
CATEGORY C:	Web Sources to measure performance	To measure the financial performance like sales, conversion rates etc.	E-commerce websites

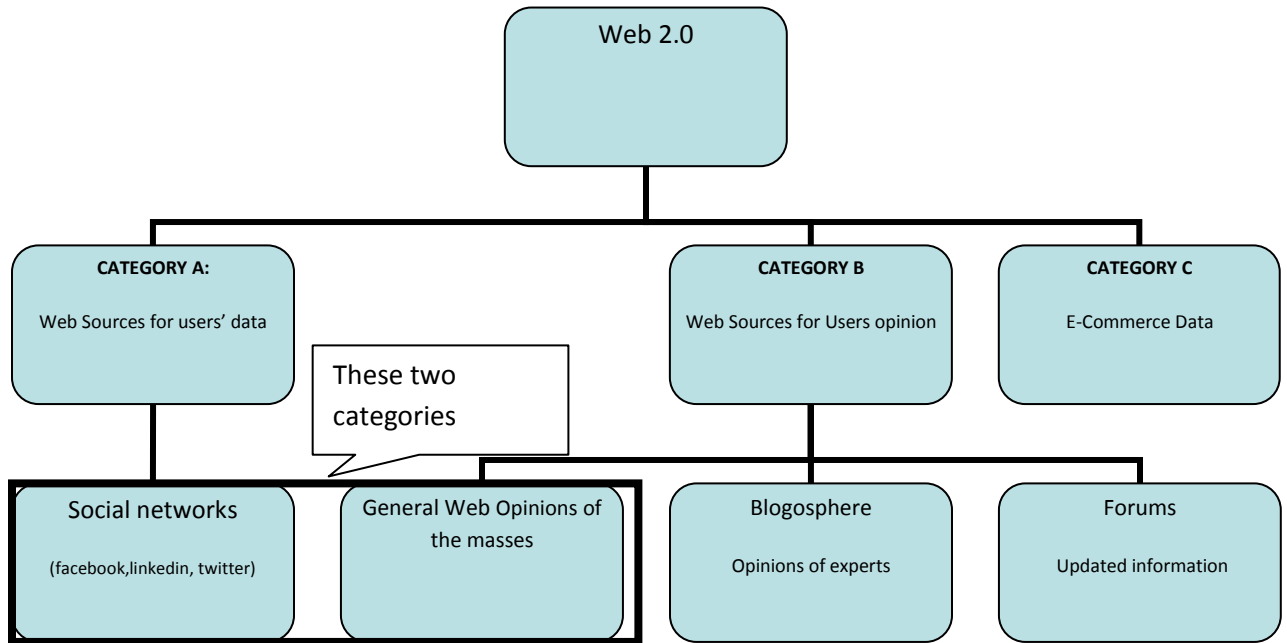


Fig (5.1) Multi-Level Marketing

It is important to categorize the web because each segment will be evaluated and processed separately for the analysis depending on its relevance. This will have the advantage of deriving accurate results, save time as well as resources.

5.1.1. Web Sources to measure Opinions.

5.1.1.1. General Web: The Websites (Masses, crowds), which represent the opinions of majority. Social media websites used for networking are also included in this category. The general web can be further divided into main subcategories depending on the purpose they serve. The table below shows the name, purpose and usage of the subcategory in our methodology:

Sub-category	Description	Examples	Use in the methodology
C2C	Usually aimed at connecting people among themselves, forming a huge network	Facebook, Twitter, YouTube, SlideShare, GoogleDocs, Twitter, DIGG , Delicious, Wikipedia, Linux.	Marketing, solving marketing related problems, Market Analysis, Advertising, Collective Intelligence
B2B	Integrated web platforms for different companies for collaboration and mutual benefits.	NA	NA
B2C	Platforms where companies provide an opportunity to consumers to share their knowledge and ideas about products	Lego, Procter and Gamble's Connect and Develop and GoldCorp	Customer relationship management, quality control of products etc.
C2B	There are also many forums out there that helps people solve their issues for any organization.	Whirlpool : a telecommunication forum for Australians where consumers help consumers solve problems that they have with Telco companies	Sometimes, problems are solved by free for the companies
Enterprise 2.0 (Internal Business)	These are web platforms for internal use of the companies	Jive's Clearspace, SocialText and Telegant	Collaboration and internal communication among employees

5.1.1.2. Blogosphere: These web 2.0 platforms represent the 'Niche' of the society, i.e intellect or the experts. Their opinions are valued separately than the general mass opinions. These platforms are assessed differently than the general web because we need to categorize the mass opinions and the expert opinion. (Experts, intellects)

5.1.1.3. Forums: Forums represent the updated information from the crowds, daily users. These forums could give us important information like

- The feedback of a certain product and also serve as an integral part of a social CRM platform.
- These forums are also considered to be the 'touch points' of a company to its consumer.

5.1.2. Web Sources to Collect users data:

The second category of web sources comprises of web sites which are hosting the users' data. These data sources are usually social networking sites which contain all the possible characteristics of an individual related to his personal life, professional life, associations, education, network of colleagues, friends, families, groups, events, preferences and choices. The type of data from these web sources may differ depending on the type of the networking site. Some characteristics / advantages, disadvantages of such web information sources are given below:

NAME	DESCRIPTION	Use
1. Id	The user's Facebook ID	Node
2. name	The user's full name	-
3. gender	The user's gender: female or male	This information could be used for user clustering based on gender
4. Email	User's address	Emails could be used for email marketing
5. languages	The user's languages	This information could be used for user clustering based on languages
6. link	The URL of the profile for the user on Facebook	-
7. username	The user's Facebook username	-

8. education	List of education's history : school, year , type , with who	Information could be used to assess the capabilities of a user
9. location	Location of user	-
10. work	List of work' s history: employer, location, position, start end date	
11. Bio	The user's biography	
12. birthday	The user's birthday	
13. Home town	Where the user lives	
14. Interested in	User's hobbies, sports and spare time activities	

NAME	DESCRIPTION	RETURN INFORMATION
Activities	The activities listed on the user's profile.	activity id, name, category and create_time
Interested objects	User's: books, albums, games, interests, movies, video.	Category, description, created time, likes, comments on these, etc.
Check-ins	The places that the user has checked-into.	Place, like, created time, message
Events	The events this user is attending.	event id, name, start_time, end_time, location
Feed	The user's: posts, notes, links, statuses	Can get comments, likes, link on a post
Friends	The user's friends, friend list	
Groups	The Groups that the user belongs to.	Group's owner, all members, group's wall, docs

Likes	All the pages this user has liked.
Mutual friends	The mutual friends between two users.
Tagged	Posts the user is tagged in.
Family	The user's family relationships

The categorization of the web 2.0 in these sources is the most important because we will extract and analyze the data from each source separately using different tools and methods. Although these two categories overlap, but the assessment of data and information extraction from these two types of data sources could be different from each other. This information will be critically evaluated for important decision making.

5.2. Web 2.0 Categories' Data Extraction & Analysis:

Once we have categorized the web 2.0 data sources for separate evaluation, we can start the process of data extraction. The data extraction tools and methods are different in both cases as well as the data types for each category. In some cases, we need to store the data and build a dedicated database off the server, and in some cases we only need to evaluate the data available on web 2.0 cloud without storing it on a local server.

5.2.1. CATEGORY A :

5.2.1.1. Web 2.0 Data Extraction:

As discussed before, this category includes the social networking sites which host the users' personal data. The most important feature of web 2.0 in this category is that social network sites are growing dynamically and hence a magnanimous amount of data is available on the web 2.0 cloud. The most important factors of data extraction on these sites are as follows:

- These sites provide integrated environment and platforms for developers hence making the data extraction easy.
- These web platform have their peculiar data extraction APIs which provide data in their particular data format.

Designing applications specifically for these social media platforms is quite easy because of the above mentioned developer-friendly feature. An application designed in any ded-

icated web-programming language like Java, PHP, Python could be developed to access the APIs of these networks.

Some characteristics of data extraction methodologies for some famous networking platforms are describes below in this table:

Platform	Api Description	Restriction Of Data Request
Facebook	<p>Facebook data are accessible through different methods that can be summed up into two main branches Graph API and FQL (Facebook Query Language).</p> <p>The Graph API is a way that tries to simplify the vision of each object contained in Facebook. The idea is that a user, a picture, a comment, a page are regarded as single elements in the platform, connected by some particular relations (e.g. the list of photo elements in an album of a user). The Graph API is very useful for gathering individual data of an object.</p> <p>FQL, based on the Graph API, allows to make queries on the platform for gathering bunches of data and turns out to be more readable in some ways when it comes to obtain the data you want.</p>	<p>Data access need permission to fetch non public data: nowadays Facebook has about 60 different permission levels. Facebook has a rate limit of 600 calls/600 minutes.</p>
Twitter	<p>The REST API enables developers to access some of the core primitives of Twitter including timelines, status updates, and user information. In addition to offering programmatic access to the timeline, status, and user objects, this API also enables developers a multitude of integration opportunities to interact with Twitter.</p> <p>The Streaming API is the real-time sample of the Twitter Firehose. This API is for those developers with data intensive needs. The Streaming API is most suited for data mining and analytics research. Streaming API allows for large quantities of key-</p>	<p>Authentication is required to access the data by API methods. The data retrieval from twitter is restricted by rate limiting: data calls are permitted 150 requests per hour.</p>

words to be specified and tracked, retrieving geo-tagged tweets from a certain region, or have the public statuses of a user set returned.

LinkedIn Uses Share API for the developers to access the data which includes data like IDs of people, content creators, groups, connections, companies, jobs via job search and job API.

The data extracted from such networking sites must be stored in an off server data base. The application could be hosted on an online server (e.g Google App Engine Server) which has certain feasible features. The tested methodology has made use of an development of an application in a dedicated web programming language (Python in our case).

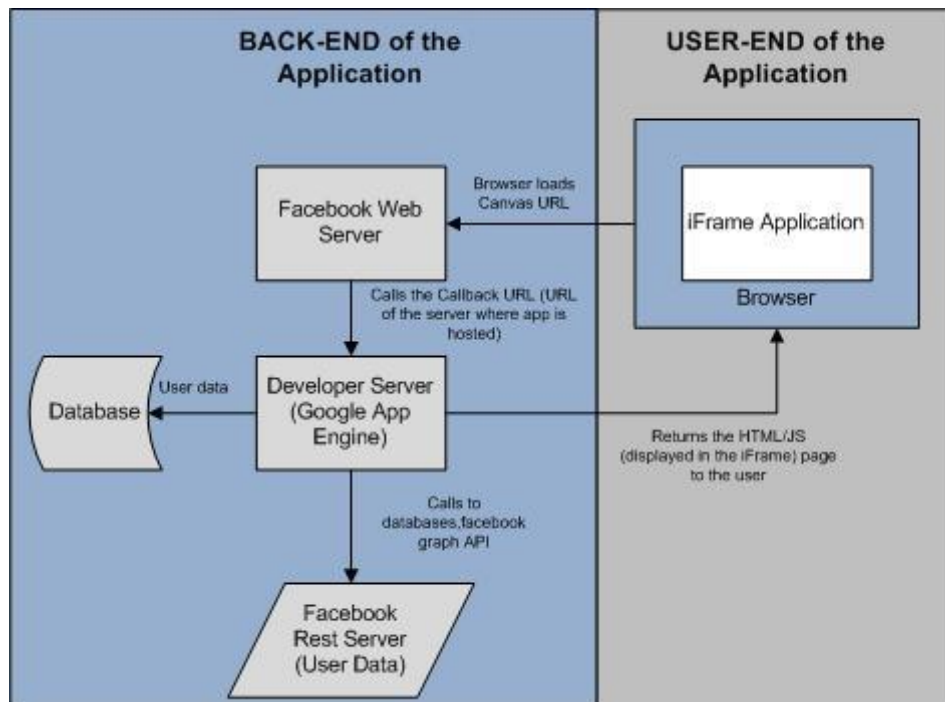


Fig (5.2) Architecture of a web-based application to collect users' data

The data schema of the data captured from a social networking site as a result of the application use given below:

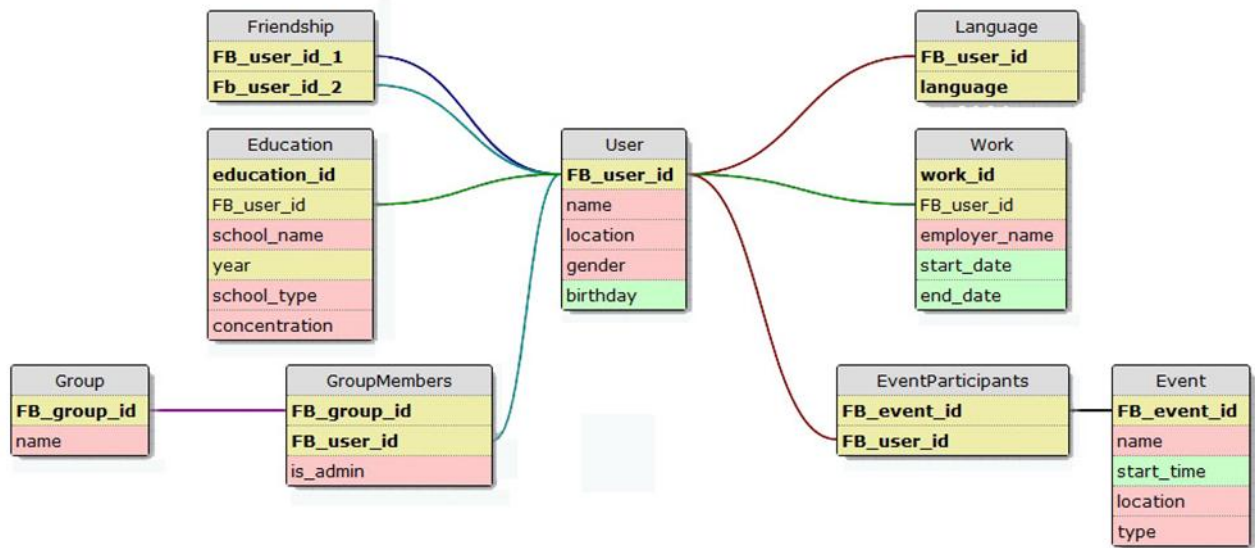


Fig (5.3) Data Schema of users' data captured from a Social networking Site

5.2.1.2. Data Analysis:

The users' data has been analyzed to find the communities, influencers and placement of users in the network. SNA tool Gephi has been used to visually analyze the users data which contains users personal information as well as their network. The tool GEPHI helps us to visually see the connections and their characteristics.

The data we extracted gave us a network of users friends, education and groups. We have analyzed the education network and friends network. For the time being, we are only considering the friends network and the overall network of friends was visualized as follows:

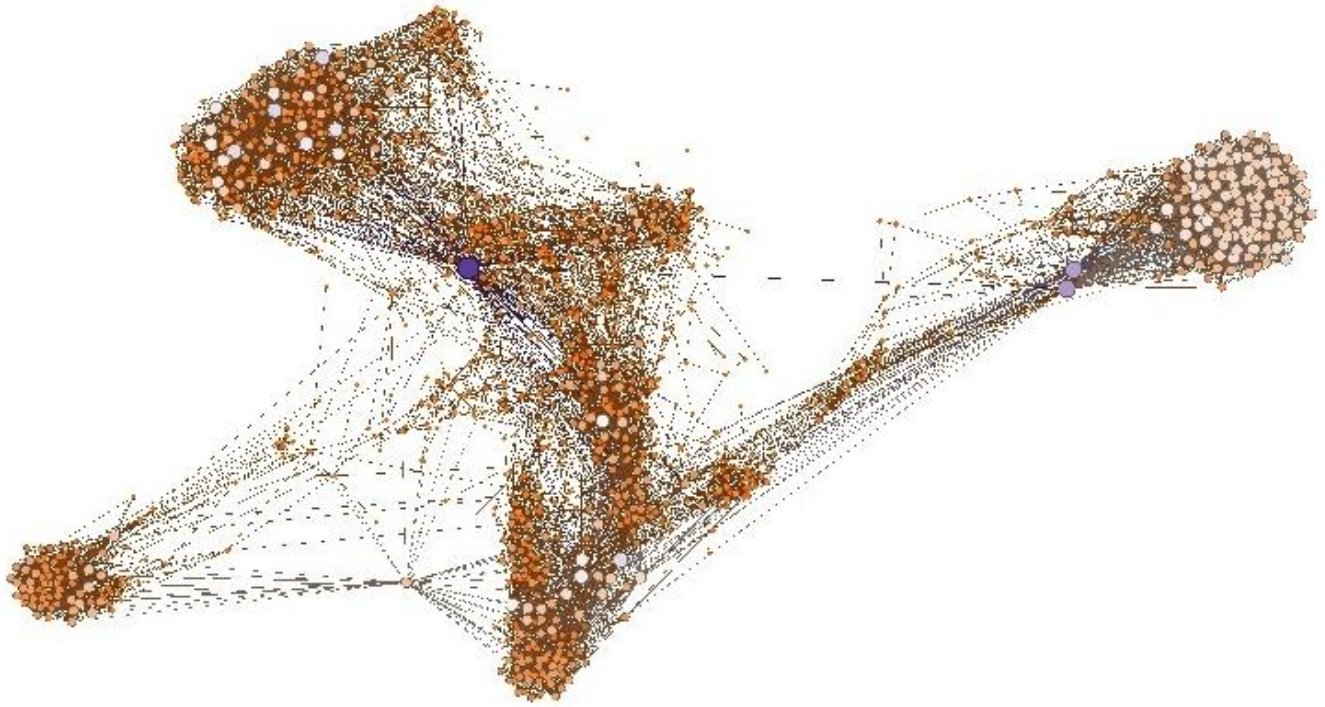


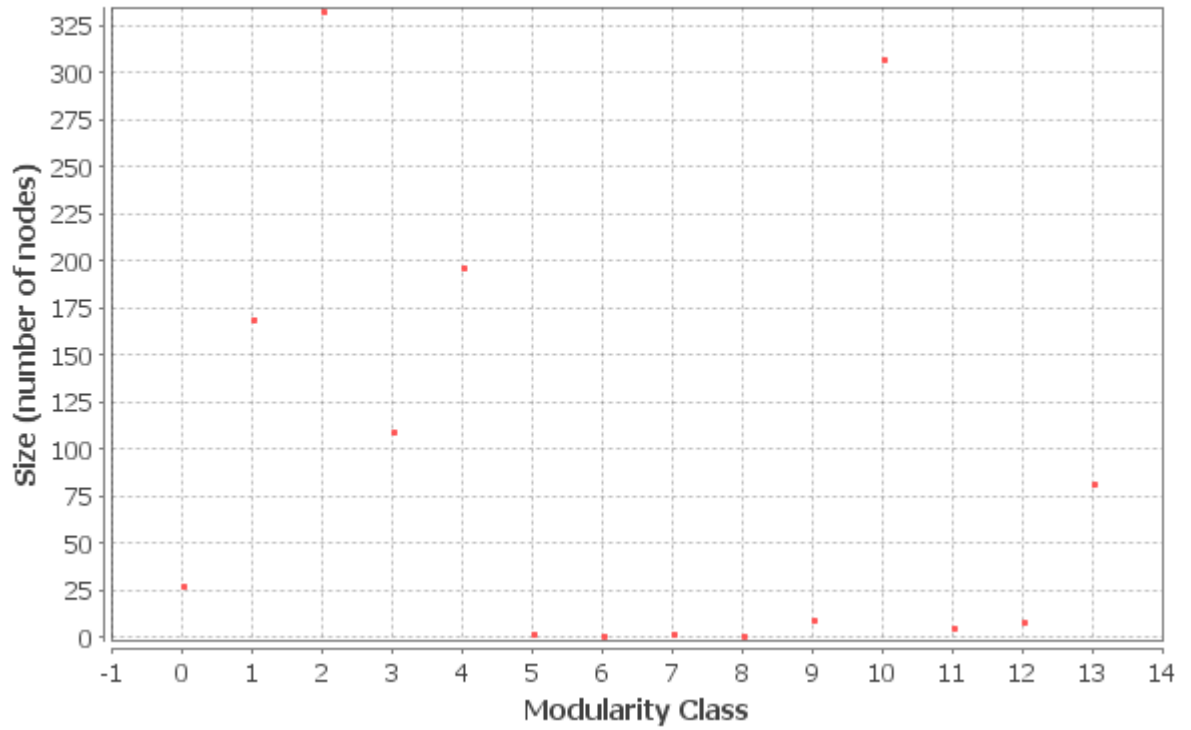
Fig (5.4) *Users' facebook friends network visualized in SNA tool*

The further analysis of the friends network has been divided into the following parts:

5.2.1.2.1. Network Overview:

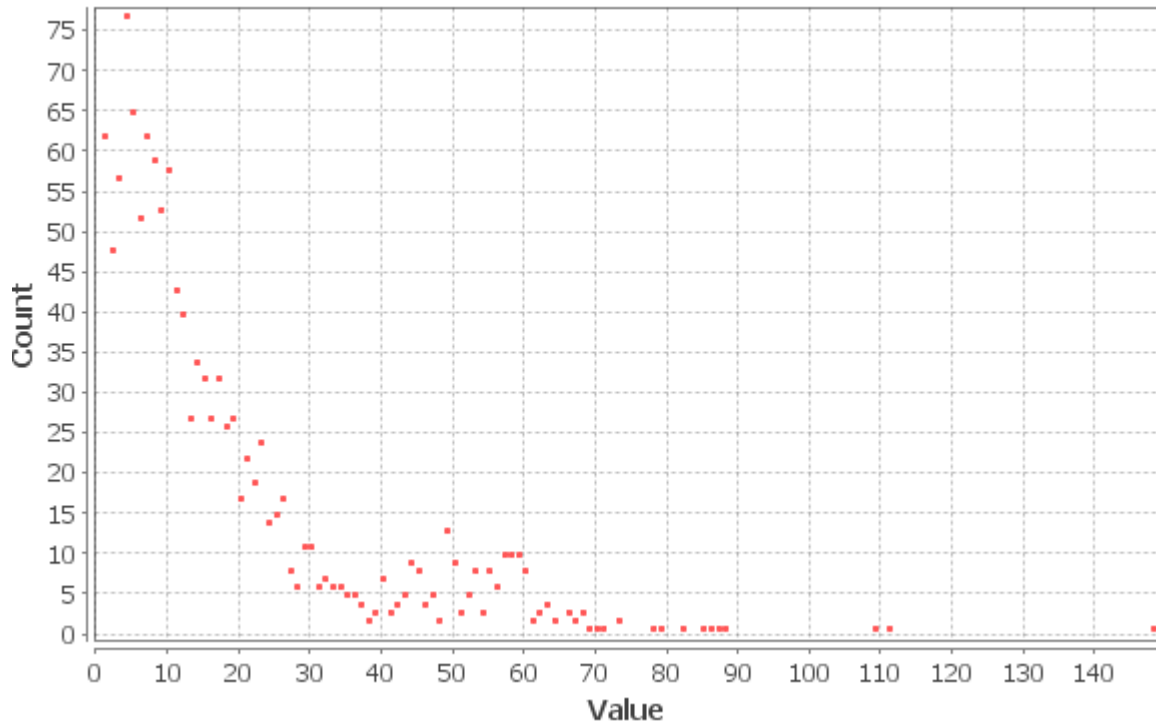
Modularity:

Size Distribution



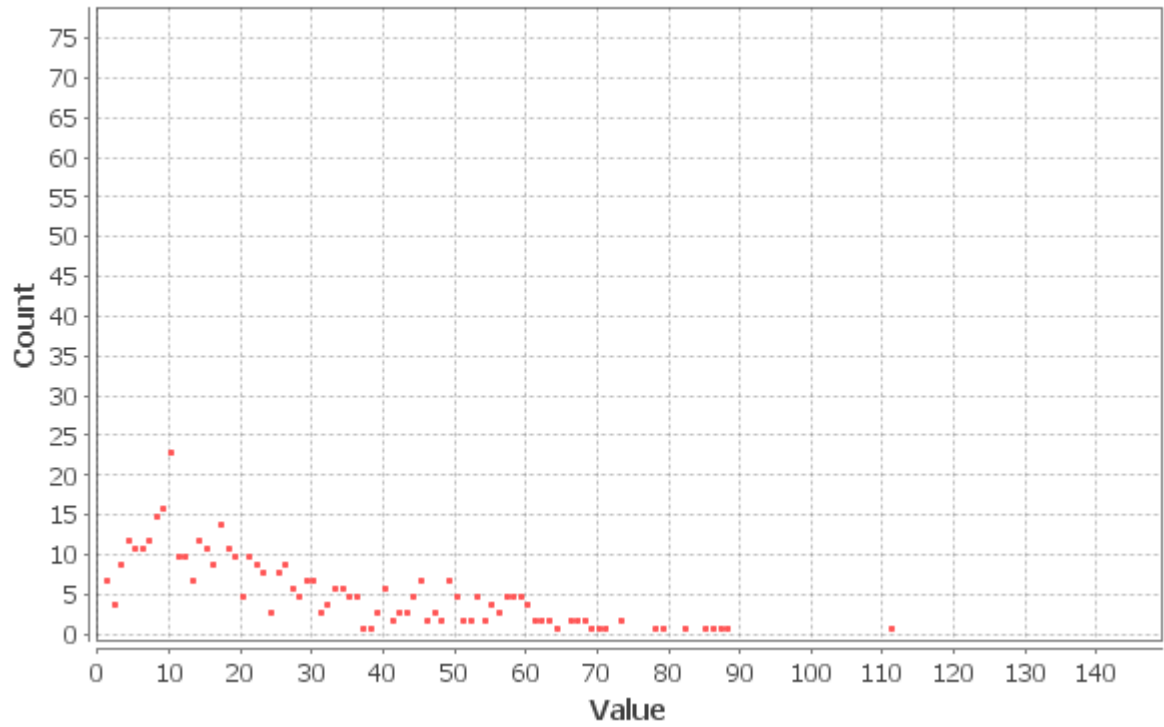
Weighted Degree Distribution:

Degree Distribution



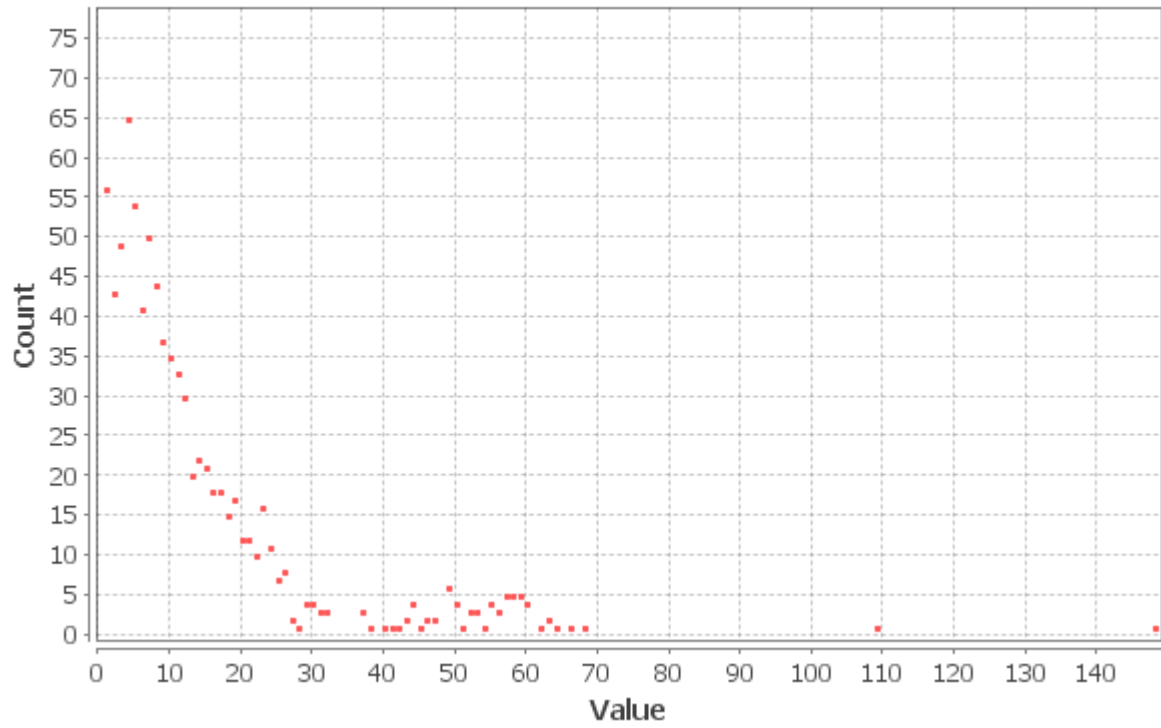
Indegree

In-Degree Distribution

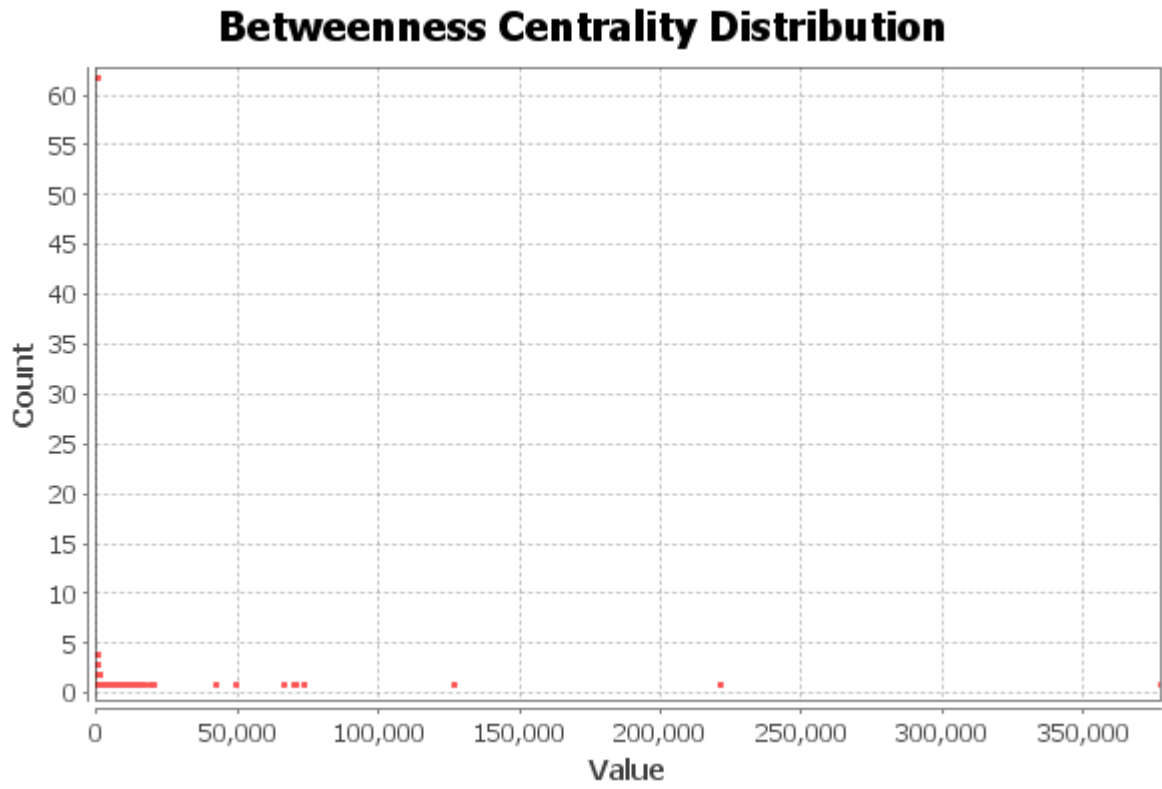


Outdegree Distribution:

Out-Degree Distribution

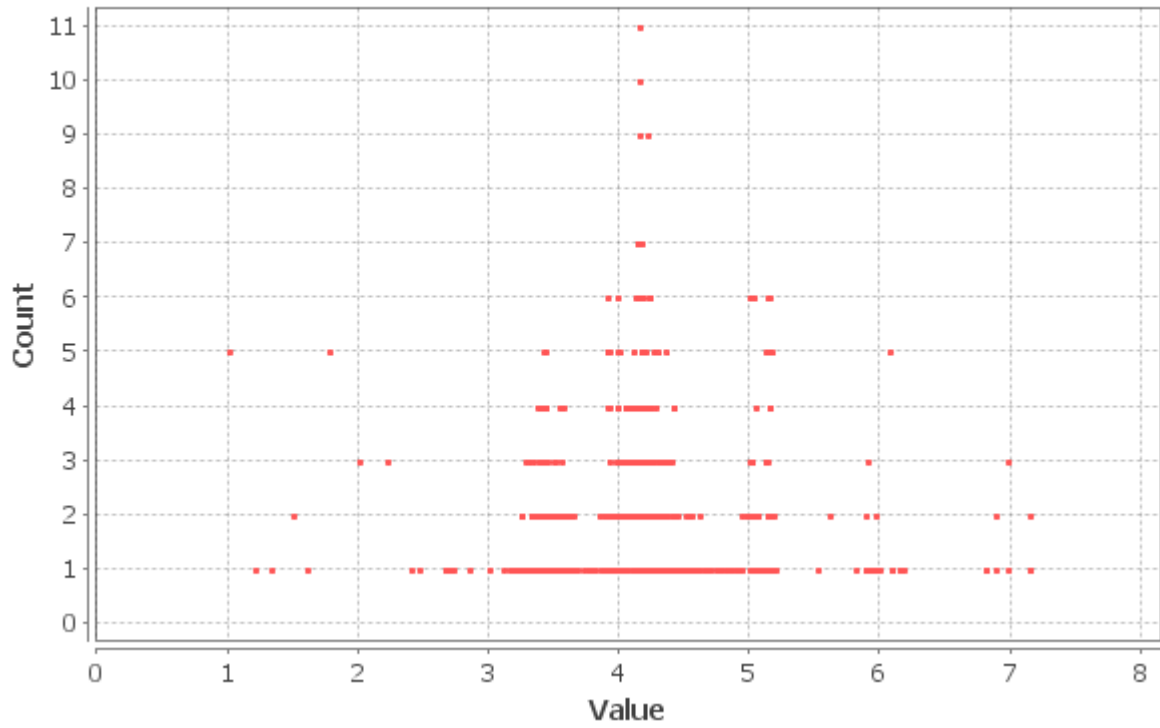


Betweenness centrality Distribution:



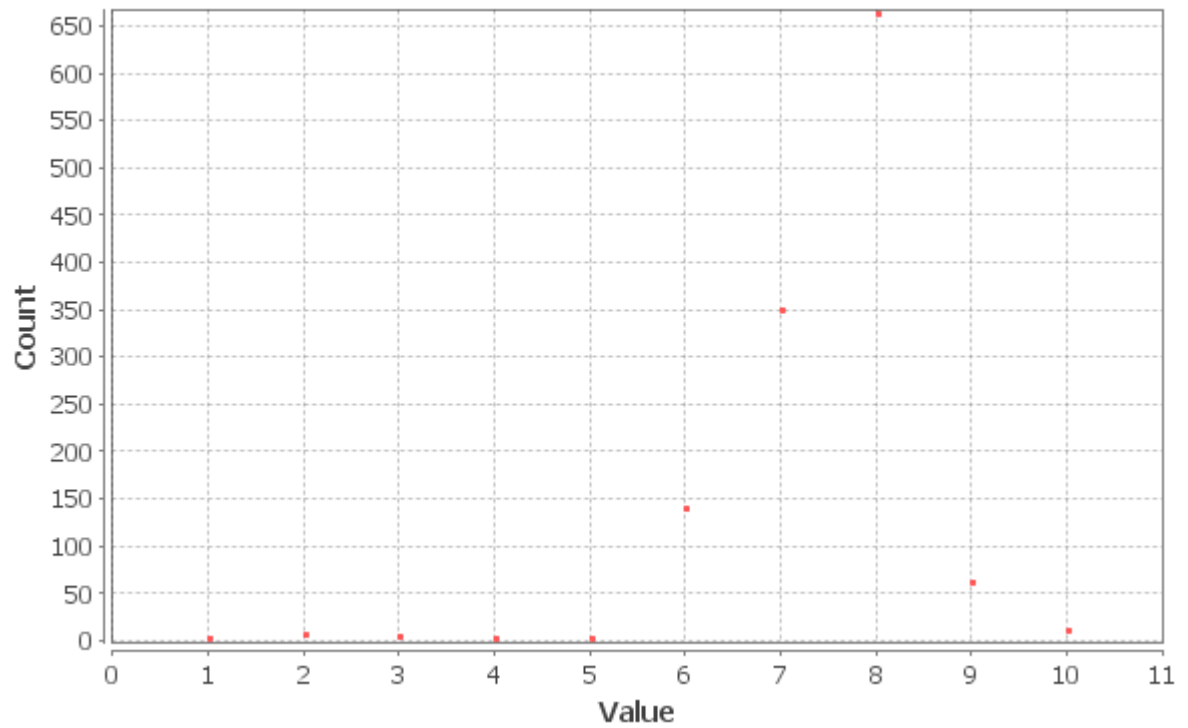
Closeness Centrality Distribution:

Closeness Centrality Distribution



Eccentricity:

Eccentricity Distribution



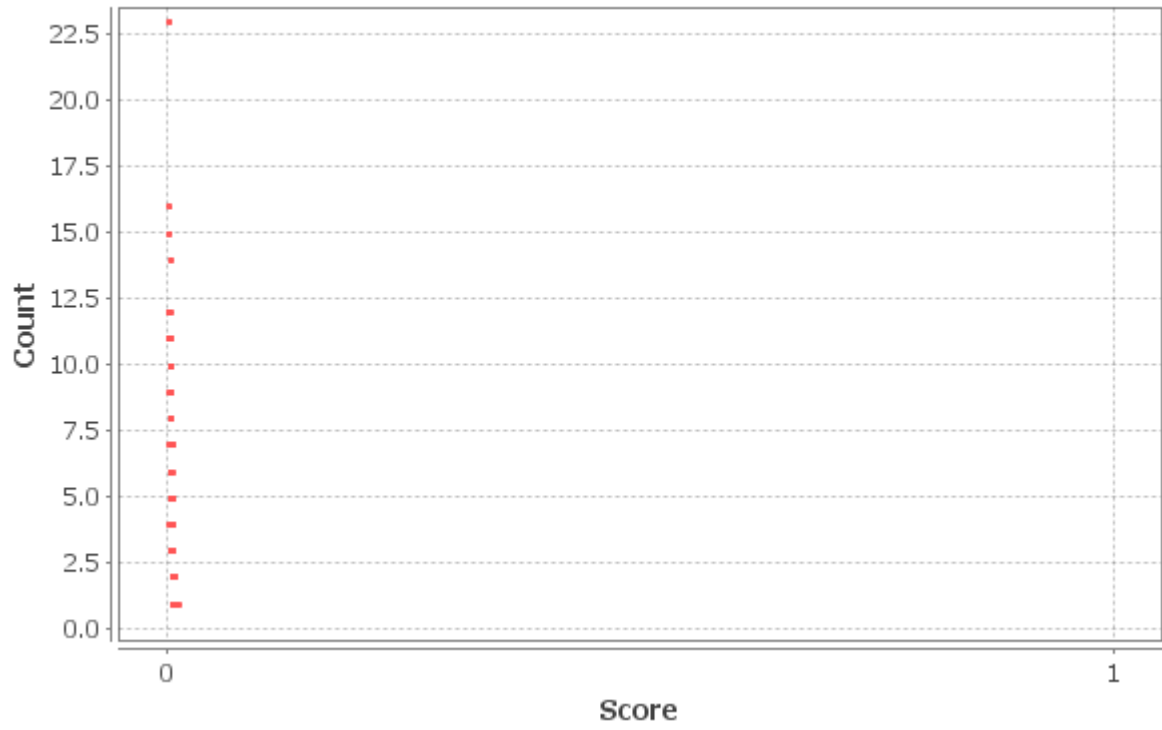
Hubs:

Hubs Distribution



Authorities:

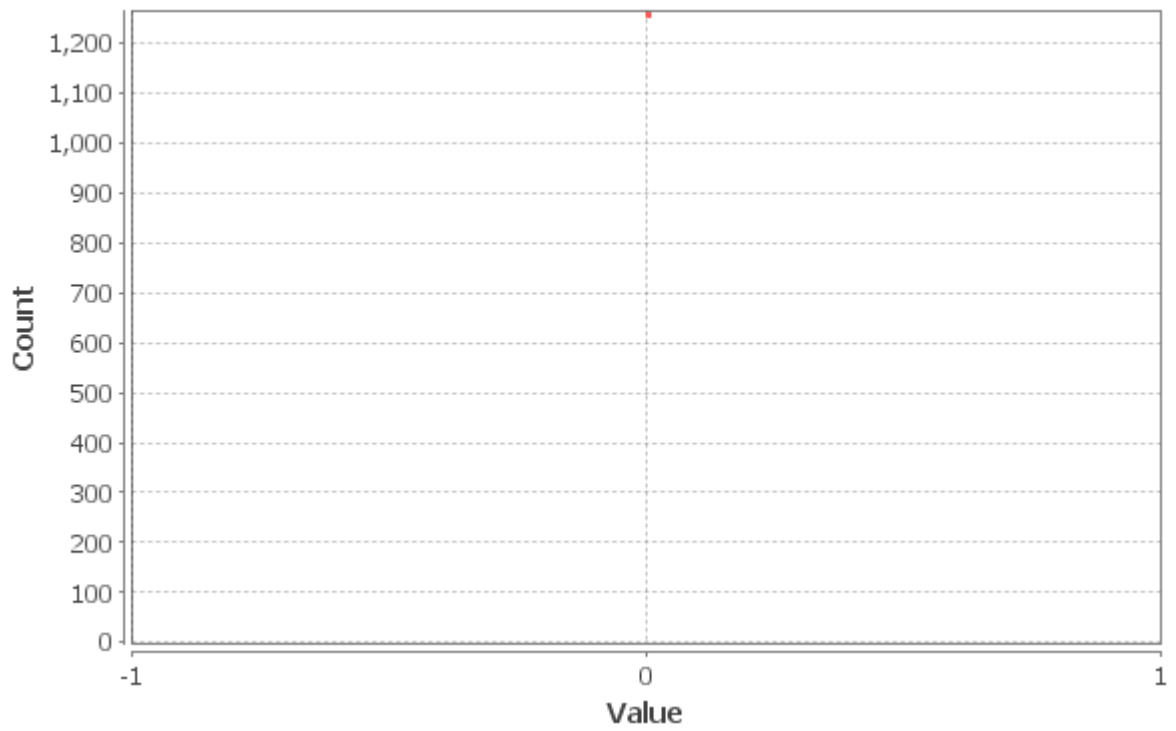
Authority Distribution



5.2.1.2.2.. Nodes Overview:

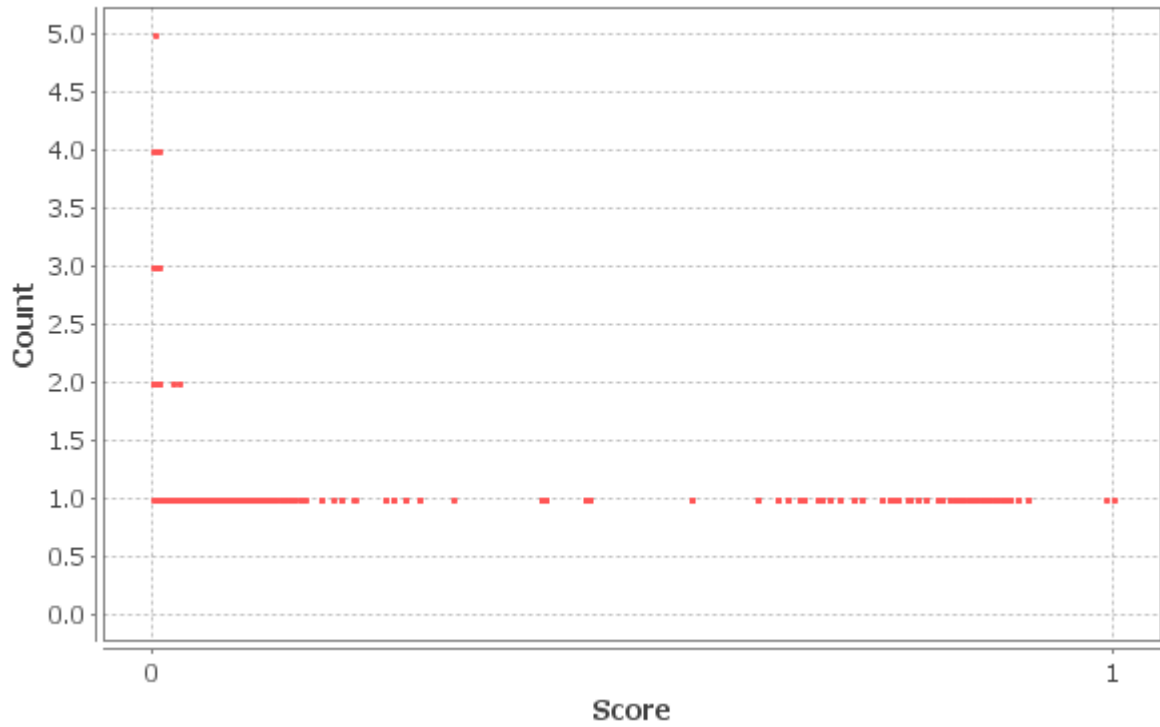
Clustering Coefficient:

Clustering Coefficient Distribution



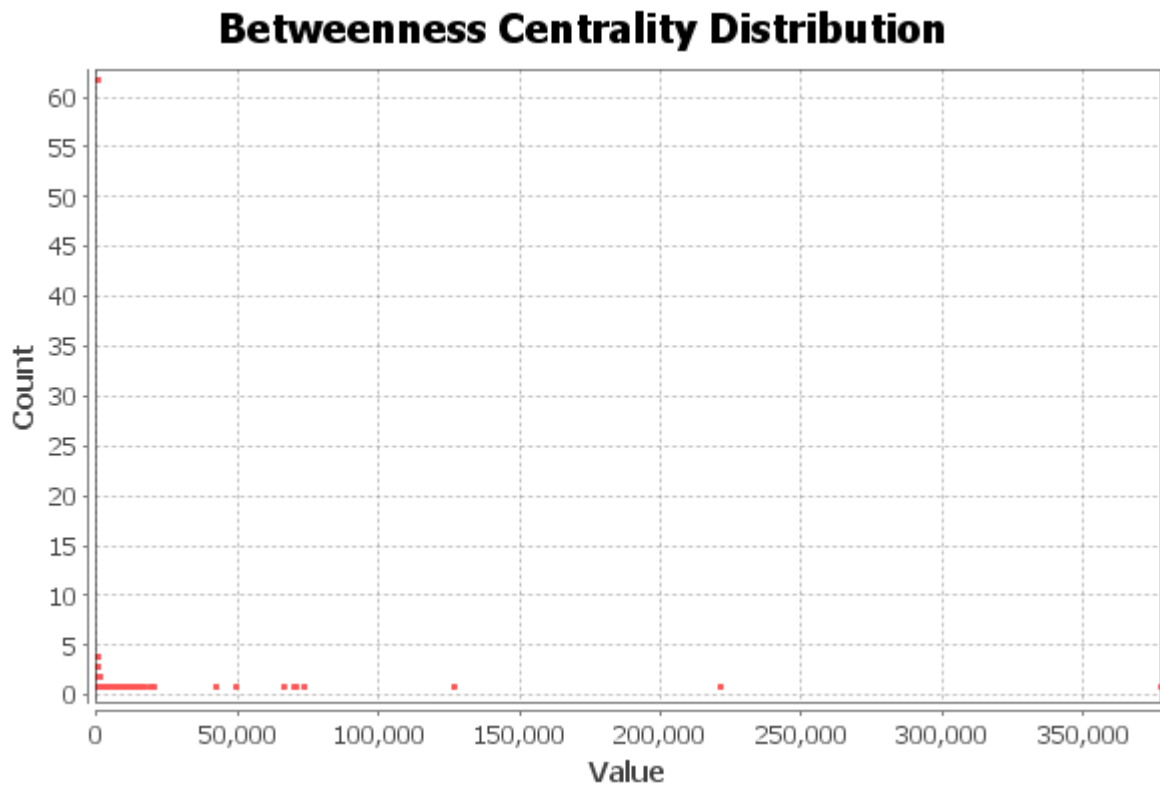
Eigen-vector Centralities:

Eigenvector Centrality Distribution



1.2.1.2.c. Edges Overview:

Betweenness Centrality:

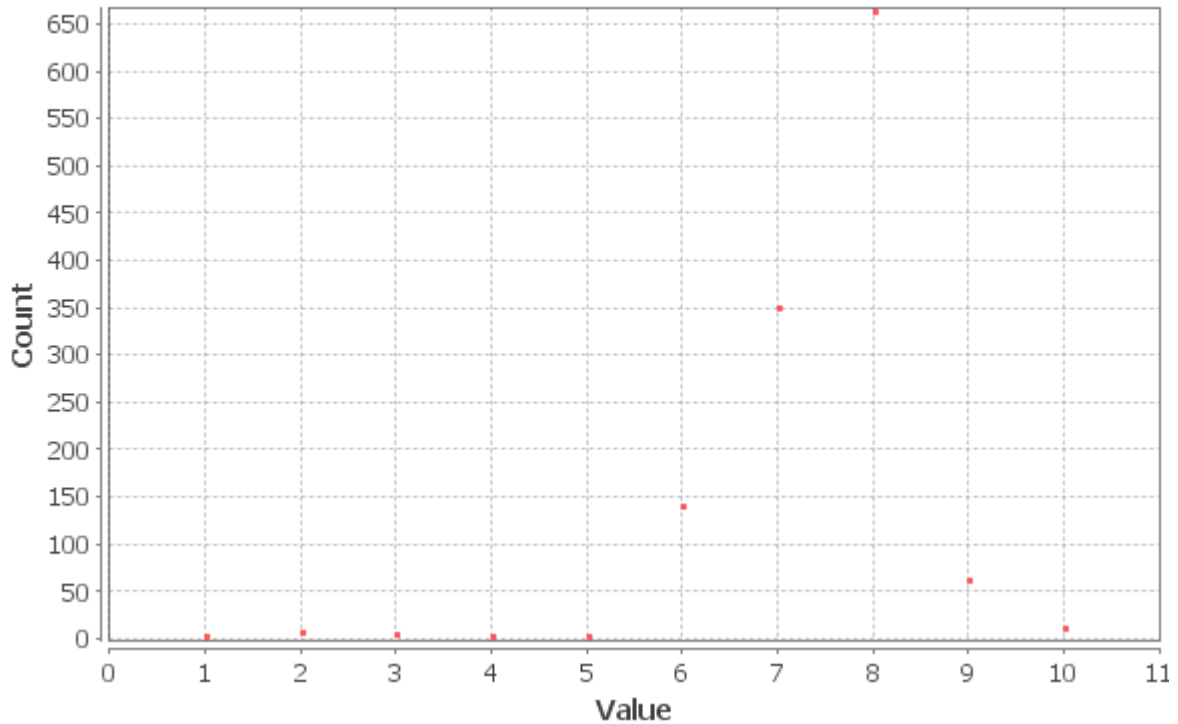


Closeness Centrality:

-

Eccentricity:

Eccentricity Distribution



5.2.2. CATEGORY B:

5.2.2.1. Web 2.0 Data Extraction:

The data on web 2.0 is dynamic, distributed in the form of a structured information spread throughout the semantic Web and growing constantly with a significant growth. Hence, to extract all of the data and storing it is out of question. Hence the data only relevant to a certain website will be extracted and used. For data extraction, we can use the RDF tool available for data extraction of semantic web.

However, in this methodology, Web 2.0 category will not be extracted and stored in an offline Database, but, will directly be analyzed using a relevant analysis tool.

5.2.2.2. Data Analysis:

We analyze the data sources in this category to extract useful information on the web i.e the opinions of the masses on the web. There are hundreds of thousands of web platforms available, hence it is important to select only the ones actually talking about the topic of interest.

In order to save time and resources, we must select only the web spheres which are relevant to the topic of the interest i.e web sources actually talking about the topic. If a

website is actually referring/linking to another topic on the web sphere, it means that both the websites share a common topic. Hence we can further analyze this linking by using a tool, (in our case we have used the tool GEPHI):

5.2.2.2. a. Finding the Giant Component in Web Network:

Recalling the SNA principles, if we consider:

- A website as a node and
- A reference link to the website from another website as an edge, then we have a network of similar websites which are linked together.

Community detection:

If we consider the giant component as a community of websites actually talking about the topic, we can detect the web sphere relevant to our topic of interest. Recalling the SNA techniques, we can use k-cliques method of k-node methods in SNA tool to detect the community of websites within the network which are actually talking more about the topic.

Weighting strategy for detecting similarity of topics on the web

Edge weight represents the similarity between nodes, so higher similarity means smaller edge length. The weighting strategy uses SNA techniques to weight certain documents on web sphere by the similarity of topics. It is achieved by automated textual analysis functionality algorithms. The similarity of topics is determined using similar texts, vocabulary etc. and various web-based tools have this capability. We can also cluster nodes based on the exchange of similarity of information between them (emails, links) by using the term 'frequency-inverse document frequency'.

5.2.2.2. b. Finding the trends on the web:

As we have the network of the connected website and their giant component, we can actually measure quantitatively the relevance of a topic on the web sphere using SNA concepts.

Betweenness centrality: We can use the concept of betweenness centrality of SNA to find the trends of a certain topic on the selected but relevant information sphere on the web. It can be measured periodically to determine the temporal trend and

the resulting numerical value of betweenness centrality will determine the importance of a concept on the web sphere. A numerical value '0' will mean that the concept is of no importance whereas a value above '0' will mean that the concept holds importance with reference to other concepts.

Degree of separation: 'Degree of Separation' means the number of links between websites not directly linked to each other. We can measure the 'degree of separation' parameter in order to find the relevance between two concepts on a web information sphere. This implies that the importance of a concept relies on the linking structure of the temporal network and the betweenness of the other concepts in the network.

5.2.2.3. Data Analysis using a Dashboard Tool:

For the section of data analysis, we consider the social networking part, blogosphere as well as forums, all the types of web in Category B, for analysis. As mentioned before, the social media data overlaps in both the categories and the type of data analysis performed for each category is different.

All the mentioned data analysis processes for category 2 mentioned in section 1.2.2.2 are available in a social media dashboard tool, which uses the above methods to mine the web and then using sophisticated textual based algorithms, gives relevant results. It is very important to select keywords and web platforms for analysis.

To use a dashboard means careful selection of:

- i- Keywords
- ii- Relevant web platforms

Measuring the defined parameters using the data available from the 'relevant' web sources. This can be achieved using various web tools and social media tools. We have used the following tools for our methodology:

Dashboard Tool	Description
-----------------------	--------------------

Mutual Minds

MutualMind monitors the social web for keywords and topics of interest and provides real-time results and analysis. We offer comprehensive monitoring coverage including millions of blogs, forums and news sites and of course social networks such as Facebook, Twitter, YouTube, Flickr and more.

Social bakers

Application organizing sophisticated metrics into easy to understand dashboards to compare social media effectiveness to others in the relevant industry.

After we have selected the relevant web sources, keywords relevant to the brand and industry, we can perform the following steps using the tool and we have elaborated the methods by using the results of a sample run. We can measure some key Social Media metrics using the dash board tool and we have given the results below:

i) Social media Metrics:

All the parameters which define the user engagement on the social touch-points of the company are termed as 'Social Media Metrics'. Following are the results of measurement of a Brand using the above mentioned tools:

- Sources:

This Key metric graph measures and displays the percentage of web sources used for measuring the results. i.e. the web platforms used for mining and deducing the results

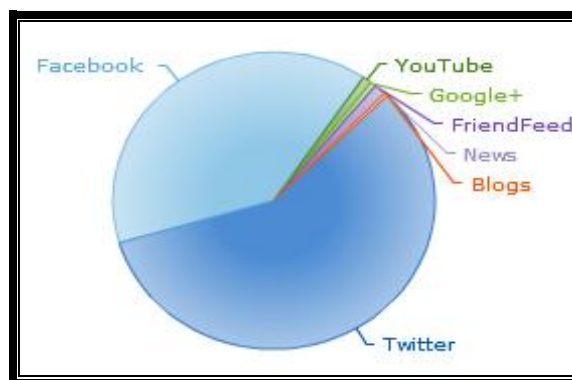


Fig (5.5) Volume of Sources

- Fan growth:

The number of followers on the company pages on these famous platforms. This is important to measure because these fans can be converted to customers given the conversion rate and could be useful to track for future prediction of sales. The tool also measures the fan growth of the competitors.

- People talking about the brand:

Before we analyze what the people are talking about the brand, it is relevant to know how many are talking about it.

- Page Engagement rates:

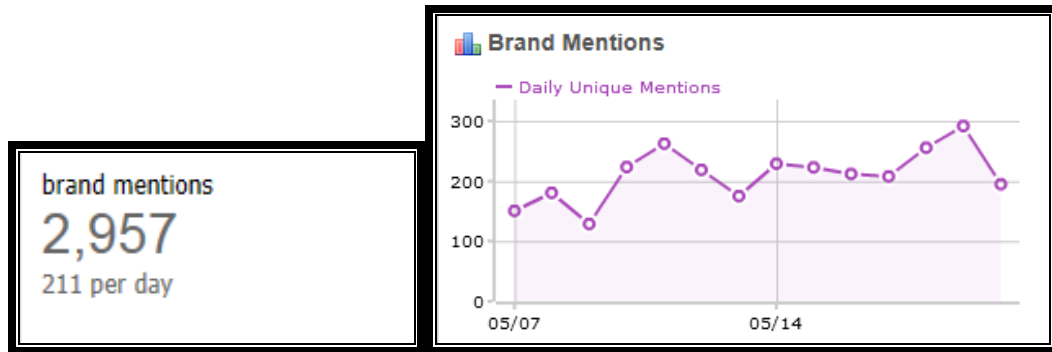
It measures the number of people actually engaged at the social media platforms of the company.

We conducted the analysis for a telecomm company based in Pakistan along with the competition analysis. The company wanted to measure the Social Media metrics for its brands as well as its competitors. Following are the results of the weekly measurement of the above defined key metrics of the tool for the company along with the others in the industry:

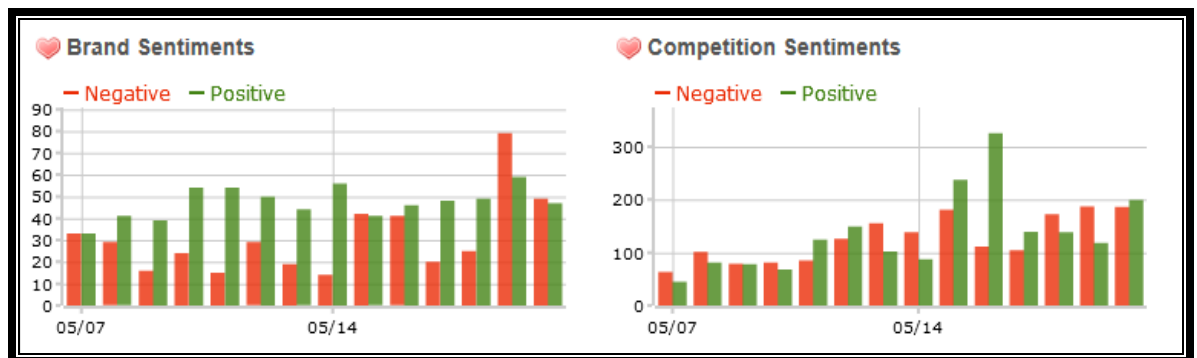
Daily Fan Growth	07/05/2012	08/05/2012	09/05/2012	10/05/2012	11/05/2012	12/05/2012	13/05/2012
Simply Mobilink	39	69	59	61	88	57	84
Jazba Official	59	67	68	58	98	5282	1438
Mobilink Jazz	39	42	55	27	44	41	50
Ufone	9192	9152	8613	5844	5394	5121	5431
Telenor Pakistan	140	139	137	146	187	148	173
Telenor Talkshawk	55	318	341	381	202	91	95
Djuice Pakistan	77	75	109	91	115	103	150
Zong	970	4105	2885	2901	2726	4358	4940
M9 by Zong	7	28	28	16	41	33	45
Warid Official	146	1868	1455	1510	550	75	776
Glow by Warid	28	36	22	19	28	31	39
People Talking About	07/05/2012	08/05/2012	09/05/2012	10/05/2012	11/05/2012	12/05/2012	13/05/2012
Simply Mobilink	1832	1832	1473	1428	1335	1213	2028
Jazba Official	19854	14590	10544	10186	10564	11750	13947
Mobilink Jazz	2880	2044	1932	1912	1912	1889	1949
Ufone	41253	49204	57233	65262	69601	72203	67365
Telenor Pakistan	2397	2397	2097	1945	1960	1895	1940
Telenor Talkshawk	4279	3842	4300	4300	4244	4478	4371
Djuice Pakistan	3812	5023	5416	5457	5457	6527	4740
Zong	36350	28490	25994	29987	34926	36654	37322
M9 by Zong	741	665	507	507	512	639	557
Warid Official	10300	10300	9512	9736	11430	10666	10666
Glow by Warid	4815	4815	4255	4134	3192	3192	1602
Daily Page Engagement Rate	07/05/2012	08/05/2012	09/05/2012	10/05/2012	11/05/2012	12/05/2012	13/05/2012
Simply Mobilink	0.10	0.04		0.08		0.25	0.05
Jazba Official	0.59	0.67	0.14	0.48	0.08	0.52	0.63
Mobilink Jazz	0.23	1.44	0.34	3.23	1.49	0.44	0.16
Ufone	0.08					0.05	
Telenor Pakistan	0.30	0.10			0.12	0.50	0.19
Telenor Talkshawk	0.15	0.13		0.10	0.20		0.28
Djuice Pakistan	0.05	0.04		0.03	0.08		0.19
Zong	1.24	1.10	1.59	0.55	1.36	0.36	0.43
M9 by Zong	0.11	0.15	0.17	0.01	0.01	0.01	0.01
Warid Official	0.65	0.99	0.32	1.56	0.40	0.59	0.51
Glow by Warid	0.20	0.20	0.33	0.22	0.21	0.26	0.37
Average Post Engagement Rate	07/05/2012	08/05/2012	09/05/2012	10/05/2012	11/05/2012	12/05/2012	13/05/2012
Simply Mobilink	0.02	0.04		0.03		0.25	0.03
Jazba Official	0.13	0.32	0.05	0.09	0.08	0.06	0.10
Mobilink Jazz	0.23	0.18	0.11	0.35	0.37	0.22	0.13
Ufone	0.06					0.05	
Telenor Pakistan	0.28	0.10			0.11	0.20	0.06
Telenor Talkshawk	0.15	0.13		0.10	0.10		0.14
Djuice Pakistan	0.04	0.02		0.03	0.08		0.07
Zong	0.12	0.11	0.19	0.13	0.15	0.17	0.14
M9 by Zong	0.01	0.02	0.06	0.01	0.01	0.01	0.01
Warid Official	0.21	0.25	0.16	0.38	0.20	0.20	0.07
Glow by Warid	0.07	0.10	0.11	0.07	0.10	0.09	0.09
Twitter Follower Growth daily	07/05/2012	08/05/2012	09/05/2012	10/05/2012	11/05/2012	12/05/2012	13/05/2012
Simply Mobilink	0	2	0	0	1	1	0
Jazba Official	1	0	1	2	2	7	2
Ufone	6	2	2	9	9	8	1
Telenor Pakistan	14	10	1	15	8	1	1
Djuice Pakistan	2	2	-1	2	3	-1	-1
Zong	-1	0	3	6	2	2	3
M9 by Zong							
Warid Official	1	3	1	2	1	-1	0
Glow by Warid	0	0	0	-1	2	0	0
Mentions	07/05/2012	08/05/2012	09/05/2012	10/05/2012	11/05/2012	12/05/2012	13/05/2012
Simply Mobilink						1	1
Jazba Official				2		5	3
Ufone				4	2		
Telenor Pakistan					4	2	3
Djuice Pakistan				1			
Zong				1			1
Warid Official				2			
Glow by Warid							

ii) Brand Monitoring:

The brand monitoring feature listens to the mention of the brand name of the company as well as its peers.

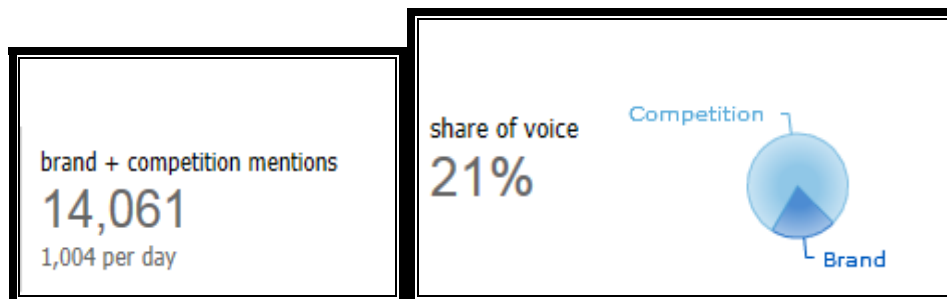


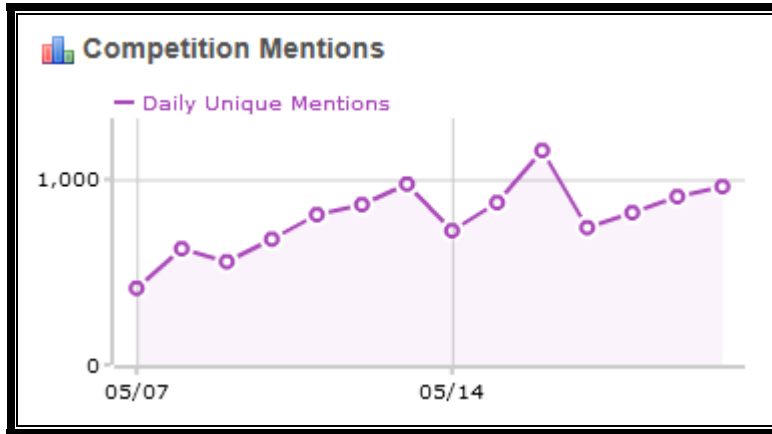
iii) Sentiment Analysis:



iv) Competitive Analysis:

The competitive analysis, as the name suggests, analyzes the social metrics of the competitors.





v) Key influencers:

Identifying the potential advocates or potential threats to the brand on these social platforms.

	ID	Followers	Friends	On Topic Mentions/retweets	Klout
1	Nadeemparacha	18,180	205	20	63
2	SanaKazmi	9,853	838	13	59
3	SamraMuslim	1631	245	9	53
4	shahidmasooddr	7936	247	8	65
5	q_malick	302	649	11	54
6	SaadGH	3,520	622	6	68
8	Asadmemon	245	291	4	63
7	nubeals	527	459		53

vi) Trend Analysis:










The trend on the web about the topics can help shape the strategy and improve marketing strategy and targeting techniques.

> Competition Topics



























1. [number](#)
2. [prison break](#)
3. [333181# bluetooth #2820](#)
4. [333191# name check](#)
5. [code 333181# bluetooth](#)
6. [name check 333141#](#)
7. [topup 333191# name](#)
8. [wasim akram's new](#)
9. [damn funny](#)
10. [samne mobile screen](#)

Brand Topics

1. [official](#)
2. [atif aslam](#)
3. [juro gey](#)
4. [mai koi problem](#)
5. [network mai koi](#)
6. [wasim akram ditches](#)
7. [hona chaahiye](#)
8. [paane](#)
9. [baqi nahin hai](#)
10. [paana pyaar karne](#)

★	 sanakazmi damn, no more jazz. RT @Umar_Khan10 Have you guys seen Wasim Bhai's Ufone wala commercial?	May 15, 2012 @ 12:20 PM
		 Pakistan  60
★	 samramuslim Mobilink is the worst telecom service providerEVER! #MobilinkSucks	May 08, 2012 @ 03:29 PM
		 54
★	 samramuslim The #WasimAkram #TVC by #Ufone! This is how good #advertising is done #Pakistan http://www.youtube.com/watch?v=Z7Lea8lVNkc&feature=youtube_gdata_player	May 16, 2012 @ 02:05 PM
		 53
★	 rasikh @taalz Mobilink sucks everywhere	May 14, 2012 @ 12:59 AM
		 Pakistan  54

Top twitter posts

★	 chowrangi.com PTA bans Twitter in Pakistan	May 20, 2012 @ 05:53 AM
	 	
★	 propakistani.pk Ufone Announces The Return of Wasim Akram as Brand Ambassador	May 18, 2012 @ 04:53 AM
		 Pakistan
★	 geotauaisay.com Wasim Akram has once again joined Ufone as its Brand Ambassador	May 18, 2012 @ 04:05 AM
		
★	 propakistani.pk Send Free SMS to Any Ufone Number from Gmail Chat	May 18, 2012 @ 02:22 AM
	 	 Pakistan
★	 propakistani.pk Mobilink Super Engineer Contest 2012 Commencing Today	May 18, 2012 @ 12:10 AM
		 Pakistan
★	 forum.gsmhosting.com Wanted!!!	May 16, 2012 @ 03:18 AM
		
★	 propakistani.pk Ufone Offers Daily Mobile Internet Package	May 15, 2012 @ 11:59 PM
		 Pakistan
★	 propakistani.pk Wasim Akram Ditches Mobilink to Join Ufone	May 15, 2012 @ 11:40 PM
	 	 Pakistan
★	 propakistani.pk Ufone Develops Play Area and Hosts a Carnival at the SOS Village	May 15, 2012 @ 11:00 PM
		 Pakistan

Top Blogposts

5.3. Web 2.0 Category C:

As mentioned before, this category of web 2.0 classification contains e-commerce websites. This data source will provide us with useful information regarding the performance of the website and tracking of the results of any decision made online.

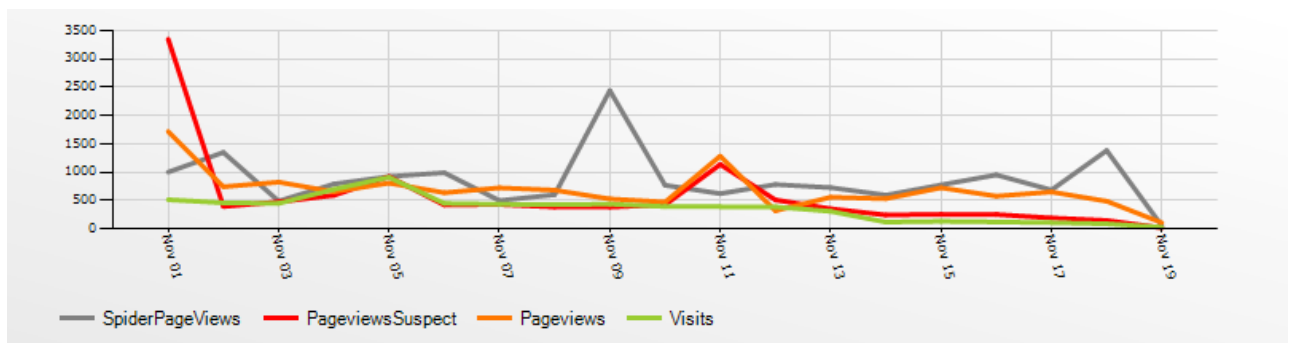
5.3.1. Data Extraction:

Web log files, cookies are primary sources of collecting data. Web server logs are used to record and accumulate users' interaction data. The downloading of a web log file of an e-commerce site is simple and all the data relevant to products, users, sales, and track records for a certain span of time.

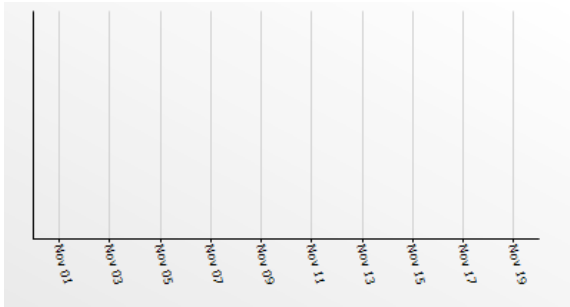
However, it is quite difficult to fetch product details along with customer buying behavior. The main purpose in this step is to acquire purchase data and path traversal patterns of past users which drive to find out navigation behavior of future users. This is why it is suggested that data from other web sources should be used with this data.

5.3.2. Data Analysis:

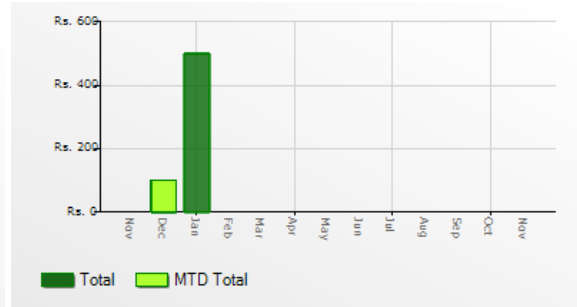
The web log files are usually used for web server debugging, hence it cannot be mainly used to assess the performance of a website, buying behavior of customers, their preferences, their characteristics and their network. Many development platforms or CMS provide basic analytics for visual tracking of the performance, but it is not enough for high level clustering and performance. We have given a sample of an e-commerce client and their sales data to give an idea how limited usefulness of a web log file could be:



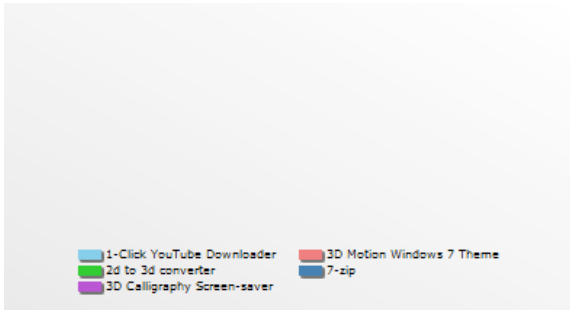
Visits and Page Views



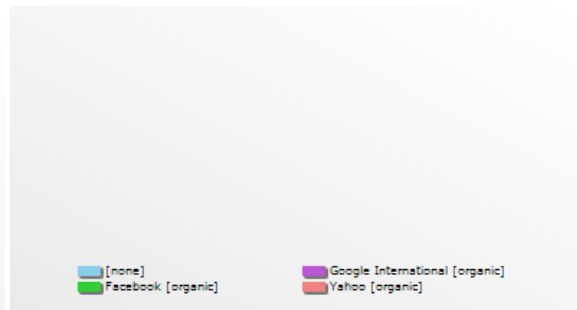
Total Sales



MTD Sales for the last 12 months



Top Products



Traffic Sources by Sales

NOTE: The Sales graph of the client is empty because the client is a startup

5.3.3. Ecommerce KPIs:

The performance of an e-commerce website can be measured by defining the key indicators necessary and then measuring them. We can only consider the KPIs of e-commerce relevant to our methodology and which help us to measure the performance of an e-commerce website. Following are the main KPIs defined to measure the performance of an e-commerce website:

- **Number of Unique Visitors**
- **Number of Visits per Day**
- **Page Views per session**
- **Conversion Rates:**

$$\text{Conversion rate} = \frac{\text{No. of goals achieved}}{\text{Visits}}$$

The 'goals' could vary depending on the type of e-commerce website. It could be a purchase, first order, sign up for the website etc.

- **Periodic Sales**

All the above mentioned KPIs could be measured with the help of data extracted from the web log file. The next step is to integrate the information of the performance of the e-commerce website which has been extracted from the data, with all the other data and information sources.

The e-commerce data information is highly related to financial performance and key measure and as seen above, the data available provides no information about the customer behavior or patterns of the customer, his interests, likes etc.

6. Decision Making and Monitoring:

In the proposed methodology, we can further analyze all the data information and build a dedicated database and keep updating the database. After extracting the useful information required, we can give weights to the information provided by these defined categories of web sources.

6. Case Studies:

6.1 PAK SERVERS

Pak Servers is a web hosting company based in Pakistan. Pak Server has been able to build a reliable infrastructure to provide efficient services at competing rates since 2008. The core service of Pak Servers remains web hosting but they also provide auxiliary support in domain registration, reselling accounts and consultancy.



Figure 6.1 Pak Server provides web hosting.

6.1.1 Challenge:

The business domain of web hosting has saturated since the internet bubble and the competition has brought the hosting companies to use price as bait against quality of services.

Pak Servers positioned itself to provide better services over price, they have invested heavily on their hardware infrastructure to provide better services in backup, online security, data migration, live customer support and reliable bandwidth. The company wants to justify their price and packages for the quality of services they are providing.

6.1.2 Solution:

Pak Servers had set up facebook page since one year but it was a neglected area of their marketing strategy. It was decided to work out a plan for Pak Servers and base its campaign on facebook using paid ads and creative ads focusing on their values that they wished to highlight. The campaign was launched on 12 November 2012, with a small budget. The creatives and their messages were prepared keeping in mind the whole drive. Furthermore, interactive status updates were made to create an input from clients to help Pak Servers listen to the market.

Face book ads were setup with \$50 daily budget. The target was based on Pak Servers market and following parameters were set for the reach:

- *who live in Pakistan*

- who like #PHP, #Symfony, #Blog, #XML, #Freelancer.com, #MySQL, #Yii, #PhpBB, #World Wide Web, #Elnce, #WordPress, #HTML, #Joomla, #PhpMyAdmin, #Domain name, #Propel (PHP), #Freelancer, #Software engineering, #Web hosting service, #VBulletin, #ODesk, #Home Shopping Network, #Zend Framework, #CakePHP, #Home shopping, #Website, #Web page, #CPanel, #WordPress.com, #MyBB or #Codeigniter
- who are not already connected to Pak Servers

This gave us a market of 851,260 users. The first ad was kept generic to build up substantial fan base. The ad text was “Looking for quality web hosting services? Click to know why you need to host with us!” along with a generic image of Pak Servers Logo.

The page was able to build up 4000+ fans in three days. The ads were posted with a gap of days along with filler posts.

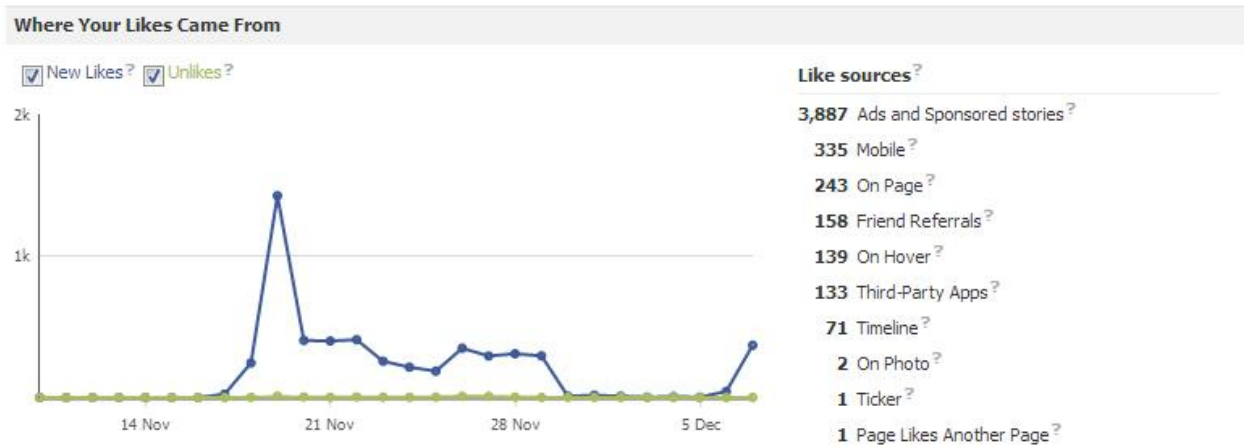


Figure 6.2 Daily likes on Pak Servers page along with source

Once a substantial number was built up, following question was polled to get input from users in terms of web hosting values.



Figure 6. Poll on Pak Servers page

This poll not only helped Pak servers assess the market but also conveyed to the users the values matrix they should use for choosing domain and hosting service. Clearly, users went for Reliability, Bandwidth and Customer Support. Reliability and Customer Support. Are normally overlooked when choosing the service, people indulge in technical specifications that these are taken for granted.

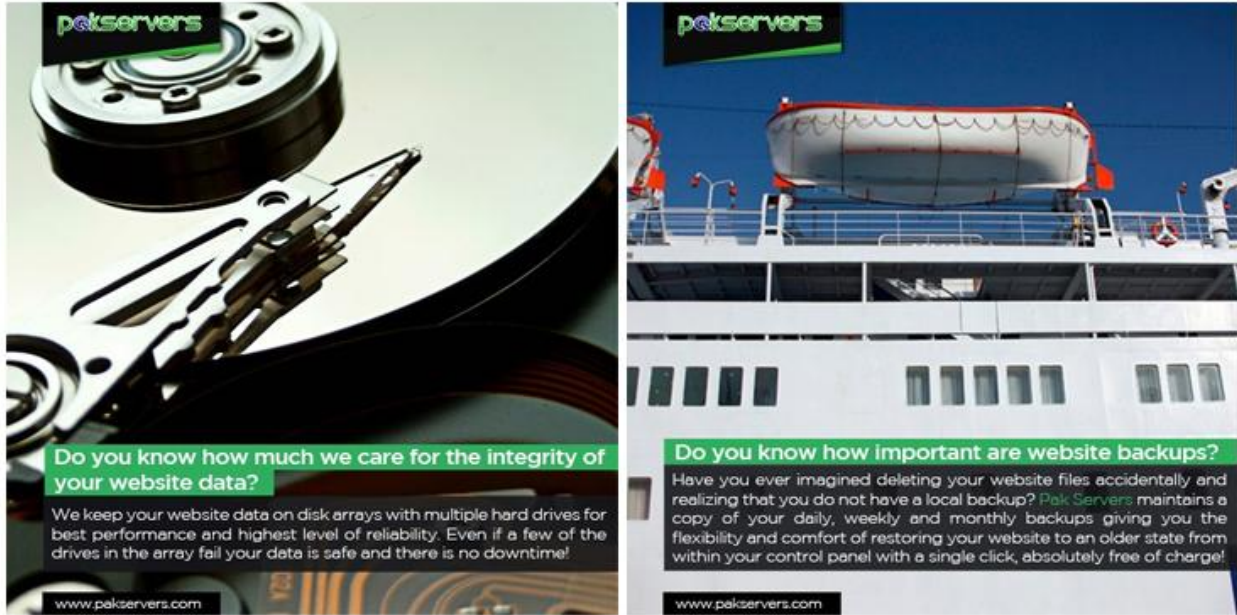


Figure 6.2 Ad Creative highlighting Data loss protection and Back up Plans



Figure 6.3 Ad Creative highlighting Uptime and Customer Support.

Campaign ?	Status ?	Start date ?	End date ?	Budget ?	Remaining ?	Spent ?
Promoting "Ever wondered what happens if one of the disk..."	▶ ▼	09/12/2012 20:48	12/12/2012 20:48	\$30.00 Lifetime	\$26.46 Lifetime	\$3.54
Promoting "How worried are you about the loss of your..."	▶ ▼	08/12/2012 18:25	11/12/2012 18:25	\$10.00 Lifetime	\$2.78 Lifetime	\$7.22
Promoting "We have grown to 5000 fans! Thank you everyone..."	▶ ▼	08/12/2012 10:30	11/12/2012 10:30	\$20.00 Lifetime	\$20.00 Lifetime	\$0.00
Promoting /PakServers/posts/445162415544944	▶ ▼	08/12/2012 01:27	11/12/2012 01:27	\$20.00 Lifetime	\$3.17 Lifetime	\$16.83
PS Q4 2012	▶ ▼	18/11/2012 05:51	Ongoing	\$50.00 Daily	\$45.12 Today	\$272.10
Promoting "3,000 fans and counting! This time YOU get to..."	✔ ▼	23/11/2012 21:49	26/11/2012 21:49	\$5.00 Lifetime	\$0.00 Lifetime	\$5.00
Promoting "How many times have you felt the need to contact..."	✔ ▼	20/11/2012 17:44	23/11/2012 17:44	\$5.00 Lifetime	\$0.00 Lifetime	\$5.00
Promoting "How serious do you consider your website's uptime?"	✔ ▼	29/11/2012 07:58	02/12/2012 07:58	\$5.00 Lifetime	\$0.96 Lifetime	\$4.04
Promoting "We have crossed 2,000 fans! Thank you for showing..."	✔ ▼	20/11/2012 21:52	23/11/2012 21:52	\$5.00 Lifetime	\$0.00 Lifetime	\$5.00
Promoting "You asked, we have heard! Watch out for the .CO..."	✔ ▼	25/11/2012 04:36	28/11/2012 04:36	\$5.00 Lifetime	\$0.00 Lifetime	\$5.00
Promoting /PakServers/posts/437745342953318	✔ ▼	19/11/2012 14:29	22/11/2012 14:29	\$5.00 Lifetime	\$0.00 Lifetime	\$5.00
Promoting /PakServers/posts/438278346233351	✔ ▼	20/11/2012 22:40	23/11/2012 22:40	\$25.00 Lifetime	\$0.00 Lifetime	\$25.00
Promoting /PakServers/posts/439654509429068	✔ ▼	24/11/2012 09:48	27/11/2012 09:48	\$5.00 Lifetime	\$0.00 Lifetime	\$5.00

Figure 6.4 Ad Campaigns for Pak Server.

In less than two weeks Pak Servers was able to make fan base of 6000+ fans. The company managed to offer two domain offers during the black Friday and cyber Monday, 100+ claims were made, out of which 40% got converted. The company managed to get a spike of 10-11% in their sales.

Figure 6.5 Pak Servers deals.

Pak Servers has decided to run a consistent social media channel for their offerings and customer insights.

6.2 CHICKEN HUT

Chicken Hut, Como is a small fast food franchise that sells fried chicken and patties. Chicken hut is based in United Kingdom and has chains all over the world. In Italy They cover more than 5 cities and constantly expanding, since their entry in 2009. In November 2011, a small franchise was open on Via Milano. The operations started in January 2012.



Figure 6.6 Chicken Hut Como

6.2.1 Challenge:

The product offering of chicken hut is good as there are no other specific fried chicken outlets in Como. The major and the foremost problem Chicken Hut has its location. The shop is located at Via Milano, alta, it is almost at the end of city street and has very few walk through in front of it. Chicken Hut tried out traditional techniques to reach out but they failed in building up consistent customers.

6.2.2 Solution:

It was decided to focus on home delivery first. The idea was to invoke interest in Chicken hut near lunch and dinner timings over social media to trigger orders. Chicken Hut’s facebook page was branded with proper images. The image posts were made to make user act on the phone and order from chicken hut.



Figure 6.7 Facebook page reach of Chicken Hut

Chicken hut ads were made to run an hour before lunch, targeting people living in Como and 10kms of the city. The ads directed users to a banner with menu item and call numbers. The action resulted in higher number of calls at lunch time.

	Lunch Calls (week 21-27 Nov) Without facebook ads	Lunch Calls (week 28-4 Dec Nov) With facebook ads
Day 1	4	8
Day 2	6	12
Day 3	5	4
Day 4	5	7
Day 5	4	8
Day 6	3	6
Day 7	5	8

The ads did help with orders but we could not acquire fans as our banners focused on delivery only. The rise of fans was not substantial. In order to increase in house activity, deal was offered to clients eating at the Chicken Hut. Two deals were setup, one on facebook and one on foursquare. Any client who checks in at the location was given free chicken piece. The banner promoting the deal was placed in front of the cashier where the client has the maximum stand time and also on tables.

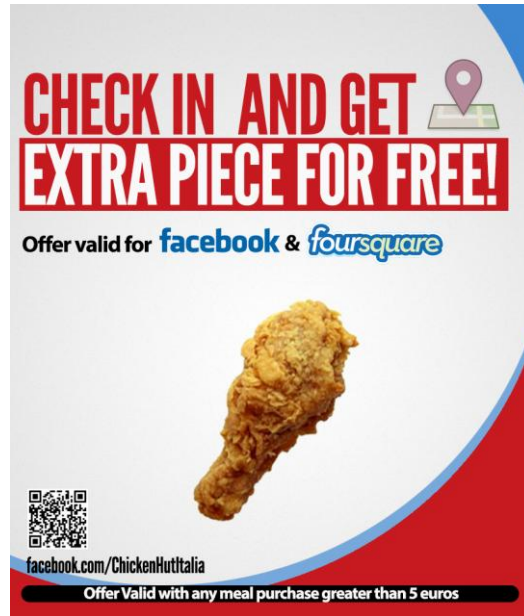


Fig 6.8 Check In banner in front of counter (English version for International Students)

Two versions were made, one targeting local Italian customers, another in English, focusing on International students. The campaign was able to generate 76 check-ins in one week. The location was able to get visibility on foursquare and facebook checkins, these tools are used by tourists to identify food outlets.

Chicken hut has realized importance of Social Media Marketing and has decided to use facebook and foursquare instead of printed advertisements (pubblicita).

References:

- [1] *The Design and Implement of CRM Data Mining System for Medium-small enterprises Based on Weka* (Shen Yan¹, Song Shun Lin², Li Shou Wei)
- [2] *A CRM Model Based on Mining Unstructured Customers' Data* (Deng Shaoling, Li Yan)
- [3] *A Process Model for Successful CRM System Development* (Hee-Woong Kim, National University of Singapore)
- [4] *Guide to Understanding Social CRM* By Chess Media Group, in Collaboration with Mitch Lieberman
- [5] *2011 IEEE International Conference on Granular Computing, A social Network perspective*, Dushyant Kumar Sharma
- [6] *Conceptual Modeling of Enterprise Application System Using Social Networking and Web 2.0*
"Social CRM System" Subaji Mohan, Eunmi Choi , Dugki Min, 2 June 2010
- [7] "Inbound Marketing", Dharmesh Shah, Brian Halligan
- [8] Martin Everett, Mitchell Centre for Social Network Analysis, University of Manchester
- [9] *Positional Analysis in Fuzzy Social Networks*, Tuan-Fang Fan, Churn-Jung Liau, Tsau-Young Lin, 2007 IEEE International Conference on Granular Computing
- [10] *A Social Network Analysis Platform for Organizational Risk Analysis – ORA*, Yin Qiuju, Chen Qingqing (2012 Second International Conference on Intelligent System Design and Engineering Applications)
- [11] *Trust-based Dynamic Web Service Composition using Social Network Analysis*, Srividya K Bansal, Ajay Bansal , M. Brian Blake, IEEE 2010 Publication
- [12] *A multi-level marketing framework for advertising in social network services* , Wen-Tai Hsieh , Jay Stu, Chun-Ming Liang, Seng-Cho Timothy Chou, IEEE 2010
- [13] *Controversy is Marketing: Mining Sentiments in Social Media* ,Zhu Zhang, Xin Li, Proceedings of the 43rd Hawaii International Conference on System Sciences – 2010