DETERMINING THE OPINION LEADERS IN ONLINE SOCIAL NETWORKS: FACEBOOK GROUP CASE

MURAT BERKANT KURU

B.S., Mathematical Engineering Işık University, 2009

Submitted to the Graduate School of Science and Engineering
in partial fulfillment of the requirements for the degree of
Master of Science
in
Information Technology

IŞIK UNIVERSITY
2014

IŞIK UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

DETERMINING THE OPINION LEADERS IN ONLINE SOCIAL NETWORKS

MURAT BERKANT KURU

APPROVED BY:

Assist. Prof. FATİH ÖZAYDIN _______________
(Thesis Supervisor)

Assist. Prof. ASLI T. ÇELİKEL _______________

Assist. Prof. CÜNEYT SEVGİ _______________

APPROVAL DATE:

DETERMINING THE OPINION LEADERS IN ONLINE SOCIAL NETWORKS: FACEBOOK GROUP CASE

Abstract

As online social networks have become the most visited web portals, marketing professionals are having more attention over them. Besides advertising options of the OSNs, advertisers try to leverage content marketing by creating organic word of mouth and earned media. Because of this situation, opinion leaders or key influencers in OSNs become the center of marketing plans but the main issue is how to detect those nodes. In literature, there are different studies in different domains and use several different methods to find out who are those opinion leaders. The common methodology among those studies is centrality and also degree centrality is called a simple and better way to determine the key nodes. In this study, the opinion leaders identification methodology is defined as a combination of degree centrality and engagement rate algorithms. The method is implemented on Facebook domain through a special community by matching existed attributes with the algorithms to clarify how it can be used in marketers' real time operations.

ÇEVRİMİÇİ SOSYAL AĞLARDA KANAAT ÖNDERLERİNİN TESPİTİ: FACEBOOK GRUP VAKASI

Özet

Çevrimiçi sosyal ağların en çok ziyaret edilen web portalları haline gelmesiyle birlikte pazarlama profesyonellerinin de odak noktası bu yöne doğru dönmeye başladı. Çevrimiçi sosyal ağlardaki reklam seçeneklerinin yanı sıra, reklamverenler içerik pazarlaması üzerine yoğunlaşarak, organik yani kulaktan kulağa yayılımı arttırmaya ve ek medya kazanımı yaratmaya büyük önem gösteriyorlar. Bu durum nedeniyle, çevrimiçi sosyal ağlardaki kanaat önderleri pazarlama planlarının merkezinde yer almaya başlasa da kimlerin gerçekten kanaat önderi olduğuna karar vermek hala bir problem olmayı sürdürüyor. Farklı platformlara odaklanan ve farklı metodlar ile kanaat önderlerini belirleyen birçok çalışma içerisinde merkeziyet metodunun öne çıktığını hatta kademe merkeziyeti metodunun diğerlerine gore daha basit ve etkili bir yöntem olarak tercih edildiğini görüyoruz.

Bu çalışmada, kanaat önderlerini belirleyebilmek için kademe merkeziyeti ve etkileşim oranı metodlarının kombinasyonu ile yeni bir algoritma oluşturulmuştur. Ayrıca, algoritmanın Facebook içerisinden özel bir topluluğun verileri ile eşleştirilmesi yöntemiyle pazarlama alanındaki operasyonlara gerçek hayatta nasıl entegre olabileceğine açıklık getirilmiştir.

To the memory of my mother

Acknowledgements

I would like to thank my major supervisor Asst. Prof. Fatih Özaydın for all of his support in my thesis process. Moreover, I would like to thank all of my faculty staff for their support in my graduate study and thanks to my dear family for their encouragement. Finally, I am also sincerely grateful to my dear partner Güneş Merve Akyol for her great moral support from the beginning to date.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BBSs	Bulleting Board Systems
BTK	Bilgi Teknolojileri Kurumu (Information Technologies and Communication Office)
CRM	Customer Relationship Management
GWI	Global Web Index
IAB	Interactive Advertising Bureau
IEEE	The Institute of Electrical and Electronics Engineers
JC	Jaccard Coefficent
KS	Knowledge Score
MC	Matching Coefficent
OECD	Organisation for Economic Co-operation and Development
OSNs	Online Social Networks
PCC	Principle Component Centrality
SC	Synthesize Centrality
SNM	Social Network Marketing
TGI	Target Group Index

Chapter 1
Introduction

Social and digital media, online startups, e-commerce, mobile apps, fiber connection and cloud; those are some of the currently rising terms in Turkey. The pushing power behind those and lots of similar online terms is the rise of the internet connection in Turkey. According to Information Technologies and Communication Office (BTK) 2013, 4th quarter report, there are currently 32.566.534 broadband connection subscribers in Turkey. In comparison with the last year's same period report, there were 27.589.309 subscribers which means there is 18,04% year over growth in terms of broadband subscribers. The biggest growth has been happened in fiber connection in one year with 85% through the rise from 645.092 to 1.193.704 subscribers. The mobile subscribers' growth is following fiber with 24,53% year over growth. Furthermore, mobile connection has the biggest portion through the broadband subscribers with 69%. There are currently 22.472.129 subscribers who connect internet through their cell phones. [1]

All of those online users upload, download and broadcast lots of information through their connections that is called "Big Data", which is also one of the rising topics for the digital world. In Turkey, over 32 million broadband subscribers create 3.033.830 terabyte traffic through fixed connections and 141.637 terabyte from mobile connections. The rise of mobile traffic in 2013 is 105%. [1]

Although, Turkish broadband connection subscribers growth from 6 million to over 32 million in last five years, penetration ratios are still below the average of OECD countries, which are used for benchmarking local data in global scale. In OECD countries, the average fixed broadband connection penetration is 26,7% while the ratio is 11% for Turkey. Similarly, the average mobile broadband connection penetration is 68,4% for OECD and 31,5% for Turkey. [1]

From the aspect of user profile, Turkey has a young internet population. Based on the Interactive Advertising Bureau (IAB) internet measurements research data, the 45% of internet users in Turkey are 24 years old or younger. One quarter of all users are between the age 25 – 34 and remaining 20% are older than 34. There is a similar data for mobile users based on Target Group Index (TGI) research shows that 41% of mobile internet users in Turkey are under the age 25, another 27% are between 25 – 34 years old and remaining 32% are over 35. [2]

Millward Brown 2014 Adreaction research also shows that second screen usage is a popular behavior among Turkish people. People who own smartphone, tablet device, laptop and TV consume 6,5 hours content in 5 hours. The consumption distribution among screens is 132 minutes for smartphones, 111 minutes for TV, 109 minutes for laptops and 39 minutes for tablets. TV and digital screens are consumed together in 36% of total screen time. People mostly claim that they don’t find TV interesting although keep it on. Other reasons for the second screen consumptions are filling the ad breaks, keep up with friends and follow up what they see on TV. [3]

Besides demographic distribution and second screen behavior of internet users in Turkey, another important point is what they spend their online time for. According to IAB Gemius October ’13 data, an average user spend his/her day on search engines, social networks, video/tv content, news and entertainment portals. The average daily time spend for a user is 3 hours 12 minutes. In this period, the users visit 17 web portals and 137 page views. Social networks have the biggest portion among visited portals with 403 million page views in daily basis. Search engines have the second place with 242 million page views and news portals following them with 103 million page views. [4] Although those portals are common among all users, there are women-kind and men-kind portals too. While youth, student, cooking, food and health portals mostly visited by the women, gaming, automotive and lifestyle portals usually attracts male users. [5]

As it is mentioned above, social networks have an important positioning among internet users but this is not a special situation for Turkey. All over the world, in every sixty seconds, Facebook users share 2.460.000 content, Twitter users tweet 277.000 times, Instagram users post 216.000 new photos, Vine users share 8.333 videos, Pinterest users pin 3.472 images and YouTube users upload 72 hours of new video. [6] In global scale, 1,8 billion of all global internet population, which are 2,5 billion people, are using social networks. There are 1,2 billion active Facebook users, 300 million active Google+ users, 259 million active LinkedIn users and 232 million active Twitter users worldwide. [7] To illustrate the rise of the social networks over the last several years, In January 2009, Twitter users tweet just 2 million times a day and now this number is 500 million tweets per day. [8] Similarly, Facebook had 360 million monthly users in 2009 and now it has over 1,2 billion monthly users. [9] In terms of local active users for Turkey, Global Web Index (GWI) data shows that there are 21.600.000 active users on Facebook, 12.240.000 active users on Twitter, 10.080.000 active users on Google+ and 3.960.000 active users on LinkedIn and Instagram. [7] Furthermore, Socialbakers data shows that Turkey is the sixth most crowded country on Facebook in terms of number of users. [10]

All of those statistics make social media one of the main channels of media strategies. The latest data for Europe market by IAB Europe shows that digital media investments have risen by 11,5% and reach over 24,3 billion euros in 2013. The United Kingdom has the largest volume among Europe market with 6,6 billion Euros. Germany is the second largest with 4,6 billion Euros and France is the third with 2,8 billion euros. While classified advertisements in listing portals have been the largest portion with 49%, 32% of total digital media investment has been spent for “Display” advertisements that include click-based advertisements like social network advertisements. The rest of the investments have been shared among search, mobile and other channels like email marketing or in-game advertising. [11]

On the other hand, the data of the Turkey for the same period have been published as 943 million TL digital media investment that corresponds to 395 million Euros in

January 2013. The distribution among channels was completely different from Europe. Search engine ad expenses had the leadership with 48% and display ads had the second position with 40%. [12] Another data for the same period from Nielsen research shows that digital channels had the 9% share among total media investments in Turkey while TV had the 56% and “Print” had the 24% share. [13] In terms of year over growth, digital media investments in Turkey have reached 1.17 billion TL with 24% increase. Similar to last year, search has the biggest share with 50% and display ads have the second position with 38%. [14] Although digital media investments have been increasing, the share of digital in total media is still about 10%. [15] According to Zenith Optimedia Media Agency predictions, this share will increase to 14% till 2016 while TV keeps the leadership with share of 56%. [16]

“Search advertising” or “paid search media buying” refers to advertisement places on the search result pages. Search engine users are targeted according to their search keywords and directed to the related page just like organic search results. Unlike paid search, display advertising has various ad models and channels in it. Affiliate marketing is a conversion-based model that is mostly preferred by e-commerce portals. In this model, advertisers are charged only if a predefined action happens such as registration, add to cart or buy a product. Sponsorships are also another model under the display ads. Brands can be the sponsor of online portals by providing some member advantages to increase sales of covering the portal to increase awareness. Video advertisements are also another popular model among display advertising. Advertisers can broadcast their TV copies or another short video at before, mid or after the actual video that a user wants to play on online video portals.

Besides those different models, impression and click based banner advertisings have the biggest share on display ads. All of the banners in any online portal, text links inside content, “Rich media implementations” like expandable banners and all social ads like promoted tweets or posts are included display models too. Since there are lots of different portals in different categories and interests, it is an issue to reach right people at the right time for advertisers but this situation is also the reason why social ads have been preferred more and more.

Online users share lots of their personal data while registering to social portals and portals use those data together with the users' content and interactions on the portal for targeting the ads. In Turkey, there are four social networks that advertisers can broadcast their ads on: Facebook, Twitter, LinkedIn and Foursquare.

Among those four portals, Facebook has the most widen targeting options for business ads. Brands can target possible customers in key locations by country, county, city and postcode or by using general demographics like age and gender. Moreover, users can be targeted according to their interests by using terms people have shared in their Facebook Timelines. These may be drawn from their listed interests, activities, education, job titles, pages they like or groups to which they belong. Besides, every connection between two users or a user and a brand page is also another source of targeting for Facebook. Brands have the opportunity to target either users who like their page or friends of them and also can exclude their fans in their target audience. In more advance, brands can use data matching method to find Facebook users from their CRM data and target them or users who have similar specialties with them. Lastly, thanks to Facebook exchange mechanism, brands can track users who visited their web sites and show them the products they have already looked for or a similar one as a Facebook ad. [17]

In Twitter, since there is not that much demographic user info, content is more important for targeting users. Brands can promote their tweet by using keywords that are used in timeline or searched for, and using interests according to the pages that the users follow, using location and gender data. Brands can also target users similar to their own follower to extend their follower base or reach their message to a broader audience and choose the devices that their message will be appear on. [18]

Besides those portals for entertainment, LinkedIn is a social platform for professionals and have the most data about education and career of its users. Advertisers can choose where the target audience is located, what are the job functions of them in which industry and their positions, titles and skills. Those are very valuable information that any brand cannot collect that amount of data from people in certain time. Besides, demographic options such as age and gender are still exist in LinkedIn too. [19]

The last social network that has advertising options is Foursquare that is a location based social service. The Foursquare target its ads based on two criteria - whether a person is nearby a location, and if they are likely to become a customer for a brand. The criteria of if someone is likely to become a customer is based on where they've checked in previously (for instance, at similar places in another part of town) or because they're searching for something related to a business (for instance, 'pizza' if you're a pizza parlor). [20]

Besides all of those targeting options, there are also several social networks that don't have ad options but actively used by a specific audience. Similarly, since the ads are signed as "Sponsored" may have a low effect on customers' decisions. This situation makes the content as key component for the social network marketing. If a brand has an engaging content can catch users' attention. Moreover, in all of four portals, there are ad options to boost the reach and the impact of the broadcasted content. In terms of credibility of the content, people trust some users more than the others and those people are called as key influencers or opinion leaders.

Marketing through the content which is organically distributed is called Word of Mouth Marketing and it is getting more popular day by day since the users getting used to ads get engaged more difficultly. To spread a message far away from the origin, it is an important issue to find out the key influencers in your topic and reach them your messages directly.

The motivation behind this study is to imply the networking algorithms into the online social networks to identify the opinion leaders who have a special impact among other users. The Facebook will be used as a domain and the application of the algorithms will be exemplified on a focus group that is part of the Facebook community.

The remainder of this thesis is organized as follows. Chapter 2 provides the research background, which covers the previous studies about detecting the opinion leaders or key influencers in social networks or communities from several academic studies. Chapter 3 presents the research motivation and methodology. After giving the methodology, details of the empirical study will be discussed in Chapter 4 as research domain, focus group, available data, application, results and the further discussions. Finally, in Chapter 5 general conclusions of the study will be presented.

Chapter 2

Research Background

Being an opinion leader is not a new term or established together with online social networks. It is common concept for every formation that people come together. There are “offline” opinion leaders in all part of society like schools, companies or sport teams as much as all online gathering portals such as Bulleting Board Sites (BBSs), Blogs or Online Social Networks (OSNs). In literature, we can say that researchers have been studying on detecting online opinion leaders in different parts of online universe.

In [21] Bodendorf and Kaiser use text-mining approach to detect opinions and relationships among forum users. They choose sample forums which opinions on Apple’s iPhone are exchanged as domain to illustrate their text mining approach. In the paper, the relationship between two forum users is defined as existence of referral from one user to another in at least one posting. These referrals are including mentioning of the recipient’s name or quotation in the sender’s posting, appearance of specific words in both sender’s and recipient’s postings and distance of the postings in the forum. Although those indicators are strong enough to identify the relationships between users, to detect which users are opinion leaders Bodendorf and Kaiser works with another indicator: centrality.

There are three different types of centrality in common: Degree, Closeness and Betweenness. Degree centrality measures how frequently a user communicates directly with others. It is calculated as the fraction of the sum of all outgoing and incoming relationships of one user and the sum of all relationships among all users in the network. Closeness centrality measures the closeness of one user to all other users. It is defined as the inverse sum of all path distance from the observed user to all others. Betweenness centrality shows how often a specific user can be found on the shortest connecting path of all pairs of the users. The betweenness centrality of a user is calculated as the fraction of the number of the shortest paths the user is lying on and the number of the all of the shortest paths. According to [21], while degree centrality is an indicator for local opinion leaders since it is concentrated on the direct relationship, closeness and betweenness centrality are global opinion leader indicators by looking at indirect relationships.

As stated in [22], the blogosphere implicitly stores a great deal of social network related information that can be used to evaluate the influence on customers' decisions. In that paper, this influence value is divided into three main categories: network based, content based and activeness based values. The network based value has subdivisions as social connection which refers to explicit relationship links or visits and social interaction that includes amount of comments, citations and blogroll. While social connection value calculation is based on Page Rank algorithm (inbound/outbound links) and number of visits on a blog site, social interaction is calculated by number of member's recommendations (blogroll), comments and the number of citations.

Similarly, content-based value is also divided into three parts as subjective degree, length and living time of a blog. Subjective degree refers to occurrences of subjective words in blog posts, while length is length of a post and living time is existence time of a blog in the network. Subjective degree and living time are indexed to the max value in the network and length calculated as blog average. Finally activeness value is connected to content posts and replies. Sum of all content on the blog and the collected comment replies in a given period is defined as activeness value of a blog. The importance of the paper [22] is they are not only making experiments on their own method, they also compare the results with other methods. They crawl one year data from a weblog community for the experiment and results show that their model could give better results than other approaches such as betweenness, review rating and popular author ratings. Besides, they claim that their approach has an average accuracy improvement of 85%.

In paper [23], Y. Cho et al. prefer to construct their own hypothetical network, which has N entities, number of neighbour is limited to 6 and the maximum distance between entities is 10. Three main concepts are evaluated in the paper to detect opinion leaders: Intimacy, Sociality and Centrality. Intimacy is defined as the strength of a tie between users regarding to their relationship status. Correspondingly, sociality is defined as the total sum of the intimacy values of an entity. Likewise paper [21], Y. Cho et al. also mainly concentrates on centrality in their research. They accept sociality as a type of centrality together with Send/Receive nomination centrality (degree centrality), Distance centrality (sum of length of all send nominations) and Rank-nomination centrality (sum of length of all send nominations of neighbours). They simulate this network in Matlab with 10.000 different entities with different social characteristics. As a result of several different scenario simulations, they suggest that market leader firms should use opinion leaders with higher distance centrality to obtain maximum penetrations while market follower firms should use opinion leaders with high sociality centrality, which has the shortest peak time in simulations. Moreover, firms should choose opinion leaders according to peak time as the best marketing strategy by considering this peak time will change according to the characteristics of the market.

In [24] Z. Zhai et al. works on BBS by constructing the reply networks using users' starting / replying interaction information in order to identify opinion leaders. They listed different opinion leadership ranking methods as System Information Based System which is using systems' stats according to users' online time, post and login counts, Simple Statistical Measures that reduce each followers reply to 1 in a single article chain and calculate replies in all articles, Z-Score which is calculated by the combination of started and replied topics, Page Rank which is calculated by reply count of each user and prominence of the user's followers and Interest based algorithms which are enclosed in a board or article chain in BBS. In this research, writers worked on a crawl data from one of the biggest BBS in China and applied those different algorithms above. As a result, they claim that interest-field based algorithms are sensitive to the high status nodes in the communication network and their performance relies on the quality of field discovery.

In [25], researchers concentrate on exploring the identification method of opinion leaders based on Online Customer Reviews analysis. They use RFM (Recency, Frequency, Monentary) model in the marketing field by adding the Sentiment parameter on it. To clarify, recency is the time between the last review time of a user to the current time, frequency is the number of reviews of a user in a given period, monentary is the normalization of number of users who find a user’s review useful and finally sentiment is the normalization of number of sentiment words existing in user’s reviews. To complete this model, they consider degree centrality too. As a result of experiment that based on a review portal, they claim that RFMS method is better than the existing ones because they consider all four key factors affecting the influence of online community review releasers.

The study of H. Zhou et al. [26] concentrates on opinion mining and utilize its results for social network analysis. They create an OpinionRank algorithm based on the PageRank algorithm to order the users in opinion networks in terms of their influence. As a result of tests with real-world datasets the OpinionRank approach performs better than the alternative methods since it includes the sentiment information.

An alternative solution for the problem to identify the key influencers in social communities is the social hubs that are defined in the study of M. Ulyas and H. Radha [27]. The central nodes of influential neighborhoods are called as social hubs by principal component centrality (PCC). Thanks to comparison of the PCC and eigenvector centrality (EVC), they analyze actual influence of a node by its positioning in a network. The relationship of 70.000 Orkut users and 143.020 Facebook users are analyzed to test performance of the PCC and it shows that most of the influencer nodes are in the same neighborhood, social hubs, and including more parameters to PCC to evaluate add new influencer nodes without replacing the old ones.

Bulleting Board Systems are similar to social networks with its mechanism that users post and reply articles from each other. Therefor, Yu Xioa and Lin Xia claim in their study [28] that they can analyze the common characteristics of opinion leaders in BBS by applying social network analysis methods but results shows that those characteristics are scale-free. Furthermore, they come up with a LeaderRank algorithm that process datasets to find out interest communities and than mine the emotional relationships in those communities to identify the opinion leaders. Moreover, they additional take away from the study is that the identified opinion leaders choose few BBS boards to actively participate.

Users who have capabilities to change the behaviors and perceptions of others in social networks are call opinion leaders and they are important since this influencer effect. Y.Li et al. propose an improved mix framework to solve the problem of identification of those influencers in their study [29]. The framework analyzes textual content, user behavior and time to rank users according to four features: expertise, novelty, influence and activity. They validated this framework with an experimental study on real datasets and the results of this experiments shows that the framework works well to identify the opinion leaders in online learning communities like longevity and centrality.

The study of Y. S. Kim and and V. L. Tran [30] approaches opinion leaders in social networks from the social network marketing (SNM) aspect. They consider both trust and distrust relationship among users to investigate the impact of the sizes of opinion leaders on the outcome of SNM. The main assumption of the study is that there are three trust metrics, knowledge score (KS), matching coefficient (MC), and Jaccard coefficient (JC), that one of them is utilized by SNM managers to reflect both the trust relationship and the outcome of the campaigns in terms of engaged users. As a result of the analysis, they claim that although in- and out-degree centrality both have effect on those three metrics MC is mostly out-degree centrality oriented while KS and JC are in-degree centrality oriented. The study also shows that there is direct proportion between the size of opinion leaders and the size of engaged users.

Blogosphere is also another social network-like community that have influential bloggers that have important position in sales and advertising. For evaluating the strength of the influence between bloggers, Y. Li et al. develop a marketing influential value (MIV) model in their study [31]. They discover bloggers who have influential potential by utilizing an artificial neural network (ANN) and analyze blogs in terms of network-based, content-based, and activeness-based factors. Their experimental study shows that out-degree and betweenness centrality algorithms and review rating and popular author approaches performs better than other attributes to identify influencer bloggers among whole blogosphere.

In a study of S. Aral and D. Walker [32], vivo ramdomized methodology is used in a method to identify influence in networks. They worked on a sample of 1.3 million Facebook users and find out that older users are less open to be influenced than younger ones, women are influencer on men more the other women and married users are the least susceptible group to influence about an offered product adoption. The analysis also shows that influential nodes are not open to be influenced as much as noninfluential ones.

H. Liu et al. propose the term synthesize centrality (SC), which includes degree centrality, betwenness centrality and closeness centrality, as an innovation to find opinion leaders in their study [33]. Pagerank algorithm by Stanford University and hyperlink induced topic search (HTIC) are used to find out influential nodes and following experimental study shows that SC results overlap the same results with a higher accuracy.

The viral marketing concept is based on creating brand awareness through a self-replicating and spreading message. In their study about identification of influencers [34], C. Kiss and M. Bichler claims that the measure of centrality provides enough data to select influencers and use them to create a viral marketing campaigns in customer networks. In decision support system applications, there are usage stimulation and churn management attributes and centrality of a customer has direct effects on those attributes. To decide which centrality measure should be used among several alternatives, a computational experiment is implemented on a telecom company call data and found that simple out-degree centrality is achieved very good results.

Identifying the influencers or opinion leaders is also an important issue for microblogs as social networking services. In their study [35], W. Chen et al. find out the principal factors that indicating influence by the analysis of the characteristics of interactive behaviors and the spreading way of information. Those factors are included the quality and quantity of followers, the quality of content and the ratio of reshares and the similarity between users in terms of interests. They combined User Relative Influence Measure model and User Network Global Influence Model and create InfluencerRank algorithm on time complexity. The InfluencerRank algorithm is implemented on Weibo data sets and validate that the algorithm is quite effective to identify influencers in microblogs.

N. Booth et al. defines influencers as “somebodies” in their study Mapping and leveraging influencers in social media to shape corporate brand perceptions [36]. They use the index valuation algorithm to measure a cross-section of variables to rank the somebodies around a particular social conversation around a single topic such as product or service. Beyond the influencers, the index algorithm reveals the conversation points to concentrate on trigger the influencers that have certain attributes such as subject and tone of voice.

Chapter 3

Research Motivations and Methodology

This chapter includes the motivations about comprehension of the study and the methodology, which is chosen for the implementation.

3.1 Research Motivation

As it was mentioned before, social media is taking more shares of marketing budget from other digital media channels as much as mass media channels increasingly. It can be easily said that the main reason of this rise is the reach of the Facebook. Together with the Google, Facebook is one of the most effective ways to reach most of the internet users in Turkey. Besides the reach, the digital trend in marketing turns into content from display ads because of the contamination of banners in web sites. Internet users know where the banners are located in a site, know there are different ads in rotation but rarely beware about what is written in there. So brands are trying to be integrated in content to get attention of the users and trying to do it without irritating them. Although advertorial content is one of the soft ways of reaching that goal, actually it is just another article in a web portal, which is known it is broadcasted by a brand. However, brands need to go one step further and become inspiring source for the content instead staying as content provider.

At that point, the question is who should brands trigger to create content on behalf of them and the answer has few key characteristics: credible, active content provider, trendsetter in other words key influencer or opinion leader. Because of this reason, detecting who is the opinion leader is getting more important day by day and need to be defined on real metrics from the current social networking tools.

The purpose of this study is to match available attributes in OSNs with the algorithms about comparison of the nodes in online networks to define central nodes that is called as opinion leaders or key influencers in the literature.

3.2 Research Methodology

As covered in the research background section, the research starts with literature review. The search on literature was based on the keywords: “Online Social Networks”, “Opinion Leaders”, “Key Influencers” and “Centrality”. The literature research was conducted on the electronic databases such as IEEE Electronic Library, Science Direct and Emerald. As a result of literature review, it is clear that the studies about determining influencer users are started with the BBSs and spread together with the rise of blogosphere and OSNs.

The main goal of this research is defining a combined methodology with the degree centrality which is a common method in research background to determine the central nodes that can affect other users opinions and behaviors and the engagement rate which is a new trend in insight generation to determine who are the opinion leaders in a group of users around similar interest. The methodology is going to be implemented through an empirical study to determine whom the opinion leader in the focus group. Facebook will be used for this empirical study because of its penetration among online users and one of the most preferred online social networks from brands.

Chapter 4

Empirical Study

This chapter covers the explanation of the chosen domain for empirical study, properties of the focus group, available data and attributes in the scope, the algorithm which is implemented for analyzing the available data and results that is extracted from the data.

4.1 Facebook Domain

In this study, Facebook platform is evaluated on as the research area. There are several reasons behind why we choose that platform for the study. First of all, it has more than 90% coverage on Turkish internet users. Secondly, it has an important marketing positioning among top brands. Lots of brands from every sector have at least fan pages on Facebook and try to communicate with their fans (potential customers) and also lots of them make media investment on Facebook ads to support this asset. However, Facebook ads also have same problems with banner ads that they lose their efficiency in comparison with the last years. Moreover, Facebook fan pages can be accepted as special communities of brands and to analyze those communities there are certain metrics such as like, comment or share.

4.2 Focus Group

To implement the methodology, the study is going to be concentrate on a closed Facebook group page that has the people with same interest. So, a master class communication group is chosen that is actively used for six months period. In that group, the educational positions are equal, there are common interests like entrepreneurship, investment, generate a new business, every user has same interaction rights without any restrictions or any moderators and all users know each other in their offline social lives.

One of the main reasons behind choosing this group to work on is to equalize all other attributes except the activations among the network. By the implementation of this methodlogy, it will be valid that the opinion leaders or the key influencers will distinguish with their activities from other users.

4.3 Available Data

In this study, before implementing opinion leader defining methodology, the Facebook focus group is analyzed for the available data. As a result of the analysis we reach four key components to use for our methodology:

- Posts: All content that a user directly share on the main feed of the group
- Comments: Users' replies to a distributed content by another user. In that component, each user comment calculated once in a post to prevent spam comments to give extra credit to the main content (post) distributor.
- Post likes and Comment likes: The like action of a user and separated into to as related with the liked content

According to those components, following parameters are developed to use in further calculations:

- Posts (P): Number of posts of a user
- Comments (C): Number of comments that a user posts
- Comment Replies (CR): Number of comments that a user gets
- Unique Comment Replies (UCR): Number of unique comments that a user gets
- Post Like (PL): Number of posts that a user liked

- Comment Like (CL): Number of comments that a user liked
- Incoming Post Like (IPL): Number of post likes that a user gets
- Incoming Comment Like (ICL): Number of comment likes that a user gets

4.4 Opinion Leader Defining Algorithm

After defining the parameters, the opinion leader-defining algorithm is set to match with the available attributes. As mentioned before, two main approaches are defined to implement on the focus group: Degree Centrality and Engagement Rate.

4.4.1 Degree Centrality

As it is defined before, degree centrality is the frequency of a user's communication with others. In other words, degree centrality is all depends on the volume that a user contributes to a community area. It is usually abstracted in two aspects: incoming and outgoing centrality. Outgoing centrality is the total interaction volume of a user with others and incoming centrality is the total interactions, which are directed to that user.

OSNs are actually not ordinary web portals, where only the information is collected. They are online platforms that provide infrastructure share our information with other users and vice versa. So they are exist as long as users continue to share content and this situation is valid every part or community in those portals. Therefore, it is obvious that the most important components are user interactions in OSNs and users are as valuable as their contributions in other words degree centrality. So in comparison with other centrality measures such as closeness or betweenness centrality which are all related with the users connection with each other, it can be that if there is no content distribution, friendship or other relations doesn't mean a lot.

4.4.2 Engagement Rate

Engagement rate is a rising parameter, which is used to analyze efficiency of especially Facebook fan pages. It basically means that the ratio of the response you get from others to the number of your contributions. Since every brand has different target segment and different communication style, this parameter creates a chance to evaluate the performance of fan pages fairly. In other words, engagement rate is become a new marketing goal for the brands which has Facebook fan page as a marketing asset.

In parallel to fan pages, if the digital trend is turning to content and each user (or i.e. Facebook fan) is becoming a marketing asset for brands, they should have engagement ratio too. Since the users and brands are similar nodes, which have same interaction options, it is easy to convert the engagement rate method with a user-centric way.

4.4.3 Redefining the Algorithm with Available Data

The methodology is converted by using available parameters for the final calculation. In degree centrality, there are two different approaches, which are outgoing and incoming degree centralities. Moreover, in engagement rate, we set again two different approaches according interacting content Post Engagement Rate and Comment Engagement Rate. Besides, there are two options to evaluate the incoming interactions for incoming degree centrality and post engagement rate. It can be either calculated by counting all the incoming interactions one by one and get the total amount, or split the interactions per posts and count that how many unique interactions happens. The main reason behind this discrimination is to prevent manipulation of the results by spam interactions.

Finally, to determine the opinion leaders, following metrics are calculated for each user (x):

- Outgoing Degree Centrality: This metric refers to the total contribution of a user to the group. So the methodology includes all available components as follows where O refers to Outgoing Degree Centrality:

$O_x=(P_x+C_x+PL_x+CL_x)$

- Incoming Degree Centrality: This metric is very similar to the first one but since it only includes reactions of other users to the user, there is no Post component and there are incoming parameters in the formula where I refers to Incoming Degree Centrality:

$I_x=(CR_x+IPL_x+ICL_x)$

- Unique Incoming Degree Centrality: The only difference between Incoming Degree Centrality and this metric is that all interactions coming from the same user into the same element as only one in here. In the formula below, UI refers to Incoming Degree Centrality:

$UI_x=(UCR_x+IPL_x+ICL_x)$

- Post Engagement Rate: In the focus Facebook group, the ratio of total number of comments and likes that a user gets over his/her posts to the number of distributed posts by a user gives us the Post Engagement Rate (PER).

$PER_x=(CR_x+IPL_x)/P_x$

- Unique Post Engagement Rate: Similar to UI, incoming interactions are counted as unique if they are coming from the same user. In the formula below, UPER refers to Unique Post Engagement Rate

$UPER_x=(UCR_x+IPL_x)/P_x$

- Comment Engagement Rate: In the focus Facebook group, the ratio of total number of likes that a user gets over his/her comments to the number of comments by a user gives us the Comment Engagement Rate (CER).

$CER_x=ICL_x/P_x$

To fill out those formulas, all available data is crawled from our Facebook focus group, counted all parameters above and calculated our four key metrics. Moreover, normalization is used to turn those key metrics into values for each user. So, minimum and maximum values are calculated for all four metrics among 26 active users. After this process final values are calculated for each user by following methods:

- Outgoing Degree Centrality Value

  $(O_x-O_{min})/(O_{max}-O_{min})$

- Incoming Degree Centrality Value

  $(I_x-I_{min})/(I_{max}-I_{min})$

- Unique Incoming Degree Centrality Value

  $(UI_x-UI_{min})/(UI_{max}-UI_{min})$

- Post Engagement Rate Value

  $(PER_x-PER_{min})/(PER_{max}-PER_{min})$

- Unique Post Engagement Rate Value

  $(UPER_x-UPER_{min})/(UPER_{max}-UPER_{min})$

- Outgoing Degree Centrality Value

  $(CER_x-CER_{min})/(CER_{max}-CER_{min})$

At the end, all of those values are sum up for each user to determine their final scores and ordered according to their total scores, which also mean their influencer orders.

4.5 Results

As it is mentioned before, the empirical study has been performed on a closed Facebook group that includes 26 users, who are all friends with each other, from same education level and same interests. All available data has been collected for six months period. The final raw data has 1.024 total interactions, which include all outgoing posts from users, and reactions to them such as like and comments.

The raw data was used to create a scoreboard by calculating personal results of the centrality and engagement rate formulations. Firstly, total interactions were separated as personal incoming and outgoing interactions. Than initial attributes – outgoing degree centrality, incoming degree centrality, post engagement and comment engagement – are calculated for each user. In these calculations, the results were normalized according to minimum and maximum values of the whole group and defined as a number between 0 and 1. At the end, a total influencer score is defined as sum of all attributes for that user. At this point, we assume weights of every single attribute on total score are equal to each other.

As explained in Chapter 3, there are two different approaches for incoming degree centrality and post engagement. First way of the calculation is including all available data without considering where those interactions coming from and the second way is to count all interactions coming from the same user into the same element as one interaction.

Therefore, each user has two different final score as total influencer score and total unique influencer score. Those two versions of scores create two different orders for the users as Total Influencer Score Table (Table 1) and Total Unique Influencer Score Table (Table 2).

Table 1 - Total Influencer Score Table

Node ID	Degree Centrality Outgoing	Degree Centrality Incoming	Post Engagement Rating	Comment Engagement Rating	Total Score
1	0,12	0,24	0,32	0,33	1,01
2	0,04	0,21	0,46	0,55	1,25
3	0,09	0,27	0,33	0,53	1,22
4	0,00	0,00	0,00	0,67	0,67
5	0,02	0,02	0,06	0,00	0,09
6	0,07	0,34	0,39	0,51	1,30
7	0,07	0,09	0,00	0,67	0,82
8	0,11	0,34	0,29	0,50	1,23
9	0,02	0,03	0,11	0,33	0,49
10	0,22	1,00	0,35	0,57	2,14
11	0,03	0,13	1,11	0,22	1,49
12	0,01	0,03	0,00	1,00	1,05
13	0,02	0,09	0,44	0,89	1,44
14	0,05	0,13	0,61	0,95	1,74
15	0,00	0,02	0,00	0,67	0,69
16	0,02	0,13	0,44	0,56	1,15
17	0,13	0,88	0,62	0,74	2,37
18	0,08	0,17	0,26	0,35	0,86
19	0,00	0,00	0,00	0,67	0,67
20	0,12	0,34	0,26	0,69	1,42
21	0,01	0,01	0,00	0,40	0,42
22	0,08	0,29	0,83	0,52	1,73
23	0,04	0,04	0,11	0,33	0,52
24	0,03	0,12	0,94	0,40	1,49
25	0,03	0,04	0,06	0,58	0,71
26	1,00	0,74	0,28	0,15	2,17

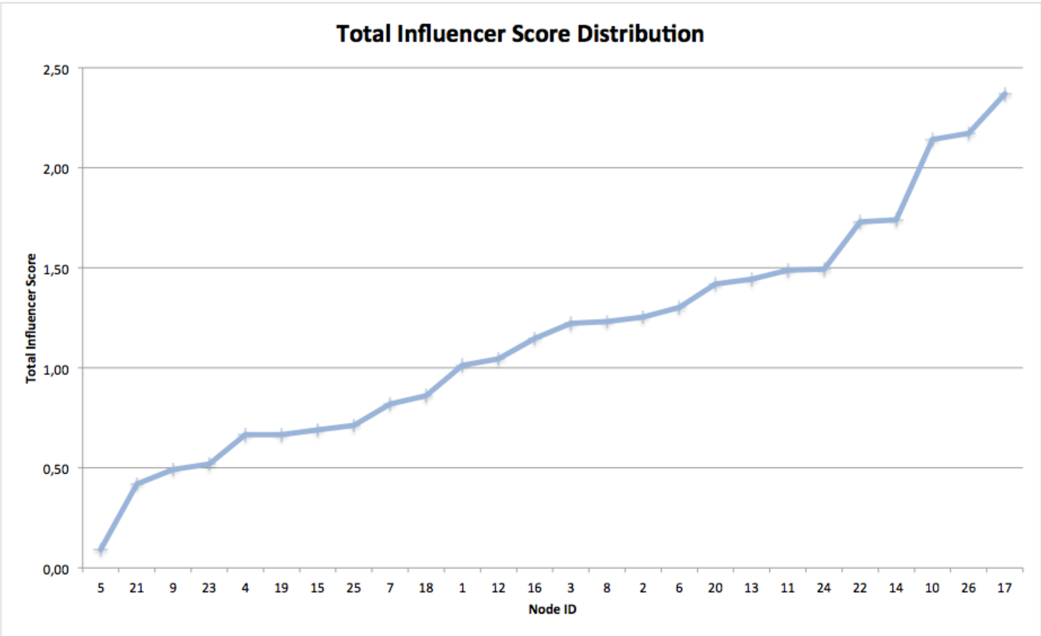
Table 2 - Total Unique Influencer Score Table

Node ID	Degree Centrality Outgoing	Degree Centrality Incoming (Unique)	Post Engagement Rating (Unique)	Comment Engagement Rating	Total Score (Unique)
1	0,12	0,29	0,35	0,33	1,09
2	0,04	0,33	0,73	0,55	1,65
3	0,09	0,29	0,20	0,53	1,12
4	0,00	0,00	0,00	0,67	0,67
5	0,02	0,03	0,09	0,00	0,14
6	0,07	0,45	0,47	0,51	1,49
7	0,07	0,15	0,00	0,67	0,88
8	0,11	0,46	0,32	0,50	1,38
9	0,02	0,04	0,18	0,33	0,58
10	0,22	1,00	0,27	0,57	2,06
11	0,03	0,14	1,18	0,22	1,57
12	0,01	0,05	0,00	1,00	1,07
13	0,02	0,13	0,45	0,89	1,49
14	0,05	0,19	0,82	0,95	2,01
15	0,00	0,03	0,00	0,67	0,70
16	0,02	0,16	0,50	0,56	1,23
17	0,13	0,94	0,49	0,74	2,30
18	0,08	0,20	0,25	0,35	0,88
19	0,00	0,00	0,00	0,67	0,67
20	0,12	0,51	0,33	0,69	1,66
21	0,01	0,02	0,00	0,40	0,43
22	0,08	0,31	0,59	0,52	1,51
23	0,04	0,06	0,18	0,33	0,61
24	0,03	0,07	0,45	0,40	0,96
25	0,03	0,07	0,09	0,58	0,78
26	1,00	0,84	0,26	0,15	2,26

When those orders are visualized in line graphs, the data line is following a stable increase user by user; few users' scores are rising much greater than the others in left hand. We read those line graphs as users who are closer to the left end of the scale are having significantly more score than others and say that they have relatively more influential effect than all other users.

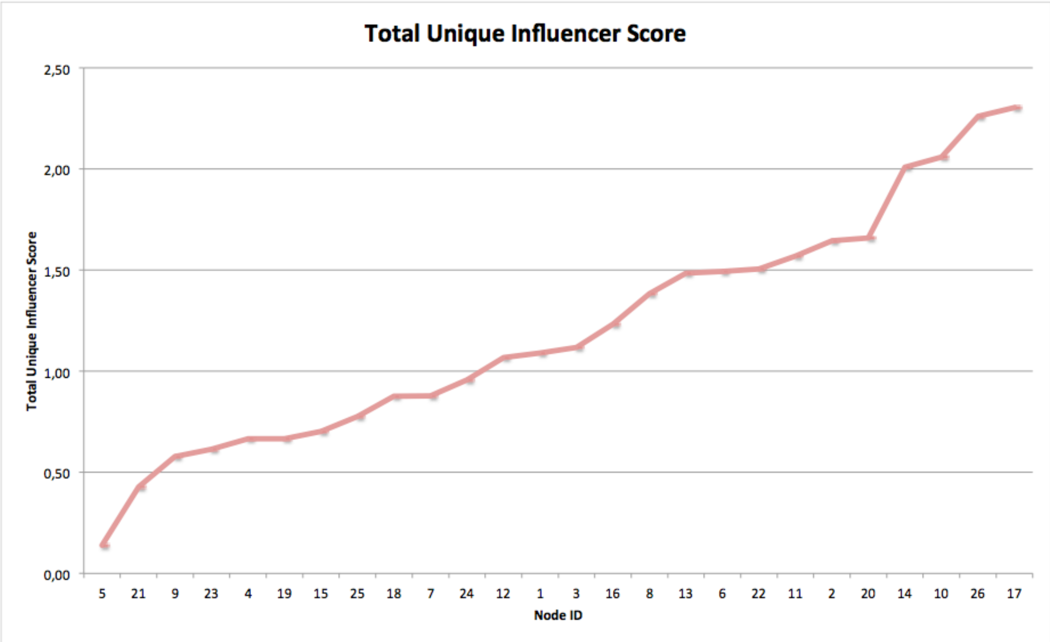
In Figure 1, the total influencer score graph, last five users have higher total scores than the average score of the group. Since, this study is looking for the most influencer users in that group, the last few nodes come up as opinion leaders among the members of the group. In other words, if a company or brand wants to influence that group of users, should work with those users who have the highest scores in the group. How many of those users will be accepted as opinion leaders for that group depends on several variables such as number of total nodes in the group, how many users you want to effect or any other special attributes from the profiles of those users.

Figure 1 - Total Unique Influencer Score Distribution



Similarly, in Figure 2 that represents the total unique influencer scores, there is a certain increase in the scores of the last four nodes. The important point in here is that, although positions of all the nodes are changing graph to graph, the four nodes are always on top and in same order. This situation claims that both algorithms indicate the same people as top influencers according to their degree centrality and engagement rates.

Figure 2 - Total Unique Influencer Score Distribution



To sum up, implementation of the algorithm indicates an influential order among the focus group and helps to analyze who are the most influential nodes among all members of the group. From the practical perspective, the algorithm reveals where to start to find the best opinion leader in a specific group and it becomes easier to decide who should be the brand ambassador to carry brand message to the group.

4.6 Limitations

Online social networks are still developing according to users needs and requests and so new attributes and interactions are added day by day. So it is really difficult to define a stabile algorithm to define opinion leaders in those networks. This is the reason why we choose to work on a close Facebook group to show how our algorithm could be implemented by defining certain attributes.

Moreover, opinion leaders should be determined based on certain topics, so you should choose a group of nodes instead of all users of an online social networks. In this study, we were able to work on a group of nodes that have similar interests to illustrate how the algorithm works. Every implementation should define its own group of nodes around a certain topic or interest to find a relatively influential order.

Crawling data is also another limitation in this study. In our focus group, we collect every data manually and match with the related attributes. To work on a larger group of nodes, an automated data crawl mechanism should be integrated into study to reach result tables.

Chapter 5
Conclusion

5.1 Key Findings

Opinion leaders, or key influencers, have important role in online social networks because of their capability to influence other nodes' behaviors and opinions about a certain topic or product. There are many theories about formation and characteristics of social networks and some of them mention some algorithms to define central nodes. However, just a few of them concentrate on the issue of opinion leader identification in an applicable way.

In this thesis, an explorative study has been performed to identify opinion leaders in Facebook domain by implementing networking and social algorithms into real data.

As a result of the literature review process, degree centrality approach that is the most common in previous studies is chosen to determine the central nodes in online social networks. Besides, another algorithm called engagement rate is combined with that formula to reach better results in terms of detecting the right nodes as influencers.

The output of the analysis showed that, the proposed total influencer scores reveals an influential order or the group of nodes that is worked on that indicated which users have the most influential effect on other users. To follow up this study, brands should make a deep analyze on other attributes of those influential nodes to decide which one could be the best brand ambassador among those group of users. As it is mentioned several times, digital marketing is getting more interactive and users are concentrating on the content of their connections more than the branded content. Therefore, brands should track those opinion leaders on online social networks to persuade to believe their message and carry the message to the other users in a smooth way, which is called word of mouth or viral marketing.

5.2 Further Implementations

Online social networks owe its existence to the content that is distributed by its users. So when there is no content, we can’t talk about any influential relationship. Based on this approach, in the proposed algorithm, outgoing degree centrality may have more weight than the other metrics. The nodes that do not create any content may be discarded from the influential order by adding an exception for the outgoing degree centrality score as if it equals to 0, other metrics may be accepted as 0 too.

In our study, we concentrate on the Facebook and more specifically on a closed Facebook group. This study may be extended to a specific Facebook page or all of the users that have the same interest by crawling the related data. In these further studies, the most important point should be the definition of the attributes that will be crawled from the portal. For instance, for a Facebook page study, shares should be tracked and should be included to the all of four metrics.

Besides Facebook, the algorithm can be easily implemented to other online social networks like Twitter, Instagram, LinkedIn and even blogs. Similarly, the basic attributes should be matched with related metrics in the proposed algorithm. To illustrate, for Twitter, it should include tweets, replies, mentions, favorites and retweets. It would be easier than analyzing a Facebook page since retweeting is a more simple action than shares on Facebook. When a user shares a content to the outside of a certain page it becomes a new entity and should be also tracked for further interactions while a retweet still keeps the whole interactions on the original content.

References

[1] – Turkey Electronic Communication Sector Quarterly Market Data Report, Information and Communications Technologies Authority, Ankara, March 2014

[2] – Vivaki Business Intelligence, *Second Spring in Social Networks*, Connected Vivaki, İstanbul, 2014

[3] – Vivaki Business Intelligence, We Consume 6,5 Hours Content in 5 Hours with Multiscreen Usage, Connected Vivaki, İstanbul, 2014

[4] - Vivaki Business Intelligence, A Day of the Digital Person, Connected Vivaki, İstanbul, 2014

[5] - Vivaki Business Intelligence, Most Visited Web Sites according to Gender, Connected Vivaki, İstanbul, 2014

[6] – James,J. ,*Data Never Sleeps 2.0*, Domo, 2014

[7] – Önerli, B., Internet and Social Media Usage Statistics, Dijital Ajanslar, 2014

[8] – Del Harvey, The strangeness of scale at Twitter, TED 2014, March 2014

[9] – Sedghi,A., Facebook: 10 years of social networking, in numbers, The Guardian, 2014

[10] – Facebook Statistics By Country, Social Bakers, April 2014

[11] – Adex Benchmark 2012, IAB Europe, May 2013

[12] – 2012 Digital Ad Spent Report, IAB Turkey, March 2013

[13] - 2012 Turkey Media Investments Report, Turkish Association of Advertising Agencies, March 2013

[14] – Adex Turkey Report, IAB Turkey, April 2014

[15] - 2013 Turkey Media Investments Report, Turkish Association of Advertising Agencies, April 2013

[16] - Vivaki Business Intelligence, Ad Spent Change in Turkey over Years, Connected Vivaki, 2014

[17] – Online Sales Ad Targeting Details, Facebook for Business, 2014

[18] – Targeting on Twitter, Twitter for Business, 2014

[19] – Targeting Your Ads, LinkedIn Help Center, 2014

[20] – Foursquare Ads, Foursquare Help Center, 2014

[21] – F. Bodendorf and C. Kaiser, Detecting Opinion Leaders and Trends in Online Communities, Fourth International Conference on Digital Society, 2010

[22] – Y. Li et al., Discovering influencers for marketing in the blogosphere, Information Sciences, 2011, 181, 5143–5157

[23] – Y. Cho et al., Identification of effective opinion leaders in the diffusion of technological innovation: A social network approach, Technological Forecasting & Social Change, 2012, 79, 97 – 106

[24] – Z. Zhai et al., Identifying Opinion Leaders in BBS, IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008

[25] – M. Yu-tao et al., Study on the Method of Identifying Opinion Leaders Based on Online Customer Reviews, International Conference on Management Science and Engineering, 2011, Rome

[26] – H. Zhou et al., Finding Leaders from Opinion Networks, ISI 2009, 2009, TX, USA

[27] – M. Ilyas et al., Identifying Influential Nodes in Online Social Networks Using Principal Component Centrality, IEEE International Conference on Communications ICC2011, 2011, Kyoto

[28] – Y. Xiao and L. Xia, Understanding Opinion Leaders in Bulletin Board Systems: Structures and Algorithms, 5th IEEE Workshop on Network Measurements, 2010, Denver, USA

[29] – Y. Li et al., An improved mix framework for opinion leader identification in online learning communities, Knowledge-Based Systems, 2013, 43, 43 – 51

[30] – Y. S. Kim et al., Assessing the ripple effects of online opinion leaders with trust and distrust metrics, Expert Systems with Applications, 2013, 40, 3500 - 3511

[31] – Y. Li et al., Discovering influencers for marketing in the blogosphere, Information Sciences, 2011, 181, 5143 - 5157

[32] – S. Aral and D. Walker, Identifying Influential and Susceptible Members of Social Networks, Science, 2012, 337, 337 - 341

[33] – H. Liu et al., Identifying Top-n Opinion Leaders on Local Social Network, ICSSC 2013, 2013, Shanghai

[34] – C. Kiss and M. Bichler, Identification of influencers — Measuring influence in customer networks, Decision Support Systems, 2008, 46, 233 - 253

[35] – W. Chen et al., InfluenceRank : An Efficient Social Influence Measurement for Millions of Users in Microblog, 2012 Second International Conference on Cloud and Green Computing, 2012, China

[36] – N. Booth and J. A. Matic, Mapping and leveraging influencers in social media to shape corporate brand perceptions, Corporate Communications: An International Journal, 2011, 16, 184 - 191

Curriculum Vitae

Murat Berkant Kuru was born in 21 January 1987, in Istanbul. He received his BS degree in Math Engineering in 2009 from ISIK University, Istanbul and he is still studying MA degree in Entrepreneurship in Özyeğin University, Istanbul. He is currently working as Social Media Supervisor in Starcom MediaVest Group Turkey. His research areas include social media marketing, digital media buying, software programming, entrepreneurship and volunteer management.