Centrality measures with a new index called E-User (Effective User) Index for determiningthe most effective user in Twitter Online Social Network

Murat Yazici, Merve Sarac
Data Scientist, Murat Yazici (corresponding author)
Department of Research and Development, Etiya Information Technologies Inc.,
Istanbul 34220, Turkey
e-mail: asiscrus.muratyazici@gmail.com
tel no: 0090 539 601 6854
Junior Data Scientist, Merve Sarac
Department of Research and Development, Etiya Information Technologies Inc.,
Istanbul 34220, Turkey
e-mail: merve.sarac@etiya.com

Abstract

In this study, we considered the issue of determination of the most effective user in the twitter online social network. We worked on asocial network graph which have relationships (edges) between users who posteda tweet and other users who re-posted it. In other words, we assume that there is a relationship between User-X and User-Y when User-X posted a tweet and User-Y re-postedit. In Social Network Analysis (SNA), there are four fundamental centrality measures such as Degree Centrality, Closeness Centrality, Betweenness Centrality, and Eigenvector Centralities.We developed a new approach for determiningthe most effective user in Twitter online social network by using an index named E-User (Effective User) Index.Through this index, we think that we are able to obtain more realistic results in SNA for Twitter.We designed a small weighted and directed social network graph by using a simulated data and used it for determining the most effective user in this study. In our graph, weights indicate the number of retweets between a user and other user, and directions indicate which user did retweet to other user's tweet. In the graph, directions can be bidirected. This means that both users did retweet their tweets to each other.

KeywordsSocial network analysis – Centrality measures – Online social networks – Social network graphs – Weighted and directed networks – Twitter

1. Introduction

The idea of centrality as applied to human communication was introduced by Bavelas in 1948. He was specifically concerned with communication in small groups and he hypothesized a relationship between structural centrality and influence in group processes. The first research application of centrality was made under the direction of Bavelas at the Group Networks Laboratory, M.I.T., in the late 1940s. The first studies were conducted by Harold Leavitt (1949) and Sidney Smith (1950). They were reported by Bavelas (1950) and Bavelas and Barrett (1951), and were first described in detail by Leavitt (1951). These reports all concluded that centrality was related to group efficiency in problem-solving, perception of leadership and the personal satisfaction of participants (Freeman 1978/79).

Social networks are systems consisting of two elements: nodes and edges between nodes. Nodes can be individuals, companies, countries, etc., while edgesrefer to the interaction or relation between the nodes. The objective of social network analysis (SNA) is to examine the structure of the network and to analyze the relations between the nodes within the network using graph theory based statistical techniques (Bozdogan and Akbilgic 2013).

SNA views social relationships in terms of network theory consisting of nodes and ties (also called edges, links, or connections). Nodes are the individual actors within the networks, and ties are the relationships between the actors. The resulting graph-based structures are often very complex. There can be many kinds of ties between the nodes. Research in a number of academic fields has shown that social networks operate on many levels, from families up to the level of nations, and play a critical role in determining the way problems are solved, organizations are run, and the degree to which individuals succeed in achiving their goals. In its simplest form, a social network is a map of specified ties, such as friendship, between the nodes being studied. The nodes to which an individual is thus connected are the social contacts of that individual. The network can also be used to measure social capital – the value that an individual gets from the social network. These concepts are often displayed in a social network diagram, where nodes are the points and ties are the lines (Passmore 2011).

In this study, we assume that there is a relationship between a user whoposteda tweet and other user who reposted it. We proposed a new index named E-User (Effective User) Indexfor determining the most effective user in the twitter online social network because of getting more realistic results in SNA for Twitterand also showing the popularity of users.

2. Social Network Analysis (SNA)

SNA has emerged as a key technique in modern sociology. It has also gained asignificant following in anthropology, biology, communication studies, economics, geography, information science, organizational studies, social psychology, and sociolinguistics, and has become a popular topic of speculation and study. People have used the idea of "social network" loosely for over a century to connote complex sets of relationships between members of social systems at all scales, from interpersonal to international. In 1954, J. A. Barnes started using the term systematically to denote patterns of ties, encompassing concepts traditionally used by the public and those used by social scientists: bounded groups (e.g., tribes, families) and social categories (e.g., gender, ethnicity)(Passmore 2011).

In a social network, relations can be any of the gained or defined ties that can be discovered among individuals. Often, individuals are also called *nodes*, while the relations between nodes are called *edges*(Bandyopadhyay et al. 2010).In this study, we will refer nodes as twitter users, and edges as relationsamongtwitter users.

In SNA, there are four fundamental centrality measures such as Degree, Closeness, Betweenness, and Eigenvector centralities.We can say that PageRank Centrality is also important. These centrality measures give us ideasabout networks and provide us to understand structure of networks.The formulas of these centrality measures will be given in following sections.

2.1. Degree Centrality

The simplest definition of point centrality is based onthe idea that important points must be the most active, in the sense that they have the largest number of ties to other points in the graph. Thus a centrality measure for an actor i, is the degree of i, i.e. the number of points adjacent to i. Two points are said adjacent if they are linked by an edge. The degree centrality of i can be defined as (Nieminen, 1974; Freeman, 1979):

$$C_i^D = \frac{k_i}{N-1} = \frac{\sum_{j \in G} a_{ij}}{N-1} \quad (1)$$

where k_i is the degree of point i. Since a given point i can at most be adjacent to $N-1$ other points, $N-1$ is the normalization factor introduced to make the definition independent of the size of the network and have $0 \le C_i^D \le 1$. The degree centrality focuses on the most visible actors in the network. An actor with a large degree is in direct contact to many other actors and being very visible is immediately recognized by others as a hub, a very active point and major channel of communication. (Latora and Marchiori, 2007)

For directed graphs, we can measure degree centrality as In Degree and Out Degree centralities. In Degree centrality represents that the number of time that node i is cited by other nodes,it can be used as an indicator to measure knowledge flow from one target node to later nodes.The higher InDegree Centrality, themore times that node i is cited, the higher momentum of knowledge diffusion from node i to other nodes (Ning and Chun, 2009).

$$In\_Degree(i) = \sum_j m_{ji} \quad (2)$$

where m is the adjacency matrix, in which m_{ji} is 1 if node i is cited by node j.

Out Degree represents that the number of times that node i cites other nodes, it can be used as an indicator to measure knowledge flow received by a target orders. The higher OutDegree Centrality, the more times that node i cites other node, the highermomentum of knowledge convergence from other nodes to node i (Ning and Chun 2009).

$$Out\_Degree(i) = \sum_j m_{ij} \quad (3)$$

In which m_{ij} is 1 if node i cites node j.

A weighted network can be represents mathematically by an adjacency matrix with entries that are not simply zero or 1, but are equal instead to the weights on the edges (Newman 2004).

$$A_{ij} = \text{weight of connection from } i \text{ to } j \tag{4}$$

The degree k_i of a vertex i in a weighted network is the sum of the weights of the edges attached to it:

$$k_i = \sum_j A_{ij} \tag{5}$$

We can also measure In Degree and Out Degree centralities in weighted and directed graphs by using the weighted adjacency matrix.

In a network graph, weighted edges affectsto the size of nodes proportionally.The size of nodes refers to their strength. Opsahl and co-works used a tuning parameter, α, which determines the relative importance of the number of ties compared to tie weights for combining both degree and strength. They proposed a degree centrality measure, which is the product of the number of nodes that a focal node is connected to, and the average weight to these nodes adjusted by the tuning parameter. Their proposed measure is as follows:

$$k_i = C_D(i) = \sum_j^N x_{ij} \tag{6}$$

$$s_i = C_D^w(i) = \sum_j^N w_{ij} \tag{7}$$

$$C_D^{w\alpha}(i) = k_i \times \left(\frac{s_i}{k_i}\right)^{\alpha} = k_i^{(1-\alpha)} \times s_i^{\alpha} \tag{8}$$

where α is a positive tuning parameter that can set according to the research setting and data. If this parameter is between 0 and 1, then having a high degree is taken as favorable, whereas if it is set above 1, a low degree is favorable (Opsahl et al. 2010).Also, k_i is the degree centrality, s_i is the weighted degree centrality, x represents the adjacency matrix, w represents the weighted adjacency matrix.

2.2. Betweenness Centrality

The betweenness centrality is a measure of a node's centrality in a network. It is equal to the number of shortest paths from all vertices to all others that pass through that node. Betweenness centrality is a more useful measure (than just connectivity) of both the load and importance of a node. The former is more global to the network, whereas the latter is only a local effect. Development of betweenness centrality is generally attributed to sociologist Linton Freeman, who has also developed a number of other centrality measures (Freeman 1977).

This centralitywas proposed by Anthonisse J. (1971)and developed by Freeman L. (1977). Itcan be used to find the edges between two communities in a complex network. Betweenness Centrality $C_B(v)$ for node v is:

$$C_B(v) = \sum_{\substack{s \neq v \neq t \in V \\ s \neq t}} \frac{\sigma_{st}(v)/\sigma_{st}}{(n-1)(n-2)} \tag{9}$$

where σ_{st} is the number of shortest geodesic paths from s to t, and $\sigma_{st}(v)$ is the number of shortest geodesic paths from s to t that pass through a node v (Freeman 1977;Zhuge and Zhang 2010).

There are many algorithms such as Bellman and Ford, Landmark, Floyd, and Dijkstra Algorithms for calculating betweenness centrality in weighted graphs. In this study, we used the Dijkstra Algorithm for calculating weighted betweenness centrality of vertex s.It will require three inputs (G,w,s), the graph G, the weights w, and the source vertex s. Pseudo-code Dijkstra's Algorithm is shown in figure 1.

```
dijkstra (G, w, s)

d[s] = 0
for each v ∈ V − {s}
        do d[v] = ∞
S = ∅
Q = V

while Q ≠ ∅
        do u = ExtractMin(Q)
        S = S ∪ {u}
        for each v ∈ adj {u}
                do if d[v] > d[u] + w(u, v)
                        then d[v] = d[u] + w(u, v)
```

Figure 1. Pseudo-code for Dijkstra's Algorithm (Hart, 2013)

Dijkstra's algorithm must first initialize its three important arrays. First, the array S contains the vertices that have already been examined or relaxed. It first starts as the empty set, but as the algorithm progresses, it will fill it with each vertex untill all are examined. Then, the distance array $d[x]$ is defined to be an array of the shortest paths from s to x, or also denoted $\delta(s,x)$ when $x \in S$. Finally, Q is simply the data type used to form the list of vertices (Hart, 2013).

2.3. Closeness Centrality

In topology and related areas in mathematics, closeness is one of the basic concepts in a topological space. Intuitively we say two sets are close if they are arbitrarily near to each other. The concept can be defined naturally in a metric space where a notion of distance between elements of the space is defined, but it can be generalized to topological spaces where we have no concrete way to measure distances. In graph theory closeness is a centrality measure of a vertex within a graph. Vertices that are 'shallow' to other vertices (that is, those that tend to have short geodesic distances to other vertices with in the graph) have higher closeness. Closeness is preferred in network analysis to mean shortest-path length, as it gives higher values to more central vertices, and so is usually positively associated with other measures such as degree. In the network theory, closeness is a sophisticated measure of centrality. It is defined as the mean geodesic distance (i.e., the shortest path) between a vertex v and all other vertices reachable from it(Newman, 2003; Passmore, 2011):

$$\frac{\sum_{t \in V \setminus v} d_G(v,t)}{n-1} \quad (10)$$

Where $n \geq 2$ is the size of the network's 'connectivity components' V reachable from v. Closeness can be regarded as a measure of how long it will take information to spread from a given vertex to other reachable vertices in the network.

Some define closeness to be the reciprocal of this quantity, but the either way the information communicated is the same (this time the estimating the speed instead of the timespan). The closeness $C_C(v)$ for a vertex v is the reciprocal of the sum of geodesic distances to all other vertices of V (Sabidussi, 1966):

$$C_C(v) = \frac{1}{\sum_{t \in V \setminus v} d_G(v,t)} \quad (11)$$

For directed graphs, we can measure closeness centrality as In Closeness and Out Closeness centralities. In closeness centrality represents that the shortest path from other patents to patents i, the higher InCloseness Centrality, the higher influence of patent i on other patents(Ning and Chun, 2009).

$$In\_Closeness(i) = \sum_{j=1}^{N} \frac{1}{d_{ji}} \tag{12}$$

where d_{ij} is the shortest path from patent j to patent i.

Out Closeness represents that the shortest path from patent i to other patents, the higher OutCloseness Centrality, the easier for patent i to be influenced by other patents (Ning and Chun, 2009).

$$Out\_Closeness(i) = \sum_{j=1}^{N} \frac{1}{d_{ij}} \tag{13}$$

where d_{ij} is the shortest path from patent i to patent j .

When we have a weighted graph, we can calculate closeness centrality using the weighted geodesic distance matrix. $C_C^w(i)$ for node i is as follows;

$$C_C^w(i) = \frac{1}{\sum_{j \in V} d_G^w(i,j)} \tag{14}$$

where d_G^w is the weighted geodesic distance matrix, $d_G^w(i, j)$ is a shortest path between node i and j with edge weights, and $C_C^w(i)$ represents the weighted closeness centrality for node i .

We can also measure In Closeness and Out Closeness centrality in weighted and directed graphs by using the weighted geodesic distance matrix.

2.4. Eigenvector Centrality

The basic concept underlying eigenvector centrality is that the “quality” of an edge should matte, i.e., an edge to a highly central node should matter more than an edge to a node with low centrality. As such, a node’s centrality should depend on the centrality of its neighbors. If we let be the centrality score for node I, then we can formalize this concept as follows(Soh H. et al. 2010):

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{N} A_{ij} x_j \tag{15}$$

where λ is a conctant value. Written in vector-matrix notation,

$$\lambda \mathbf{x} = \mathbf{A} . \mathbf{x}$$

and hence $\mathbf{x}$ is an eigenvector of the adjacency matrix $\mathbf{A}$ with eigenvalue λ. Using the Perron–Frobenius theorem, we can show that λ is the largest eigenvalue and $\mathbf{x}$ is the associated eigenvector. If we normalize $\mathbf{x}$, the eigenvector centrality of a node varies in the range (0,1), with larger values indicating higher centrality. We can easily extend this concept to weighted networks by noting that weights should affect the importance of edges (Soh H. et al. 2010),

$$x_i^w = \frac{1}{\lambda} \sum_{j=1}^{N} a_{i,j} w_{i,j} x_j \; . \tag{16}$$

In general, there will be many different eigenvalues for which an eigenvector solution exists. However, the additional requirement that all the entries in the eigenvector be positive implies (by the Perron-Frobenius theorem) that only the greatest eigenvalue results in the desired centrality measure. The component of the related eigenvector then gives the centrality score of the node in the network. Power iteration is one of many eigenvalue algorithms that may be used to find this dominant eigenvector(Passmore 2011).

2.5. PageRank Centrality

The main idea implemented by PageRank is that of “voting” or “recommendation.” When one vertex links to another one, it is basically casting a vote for that other vertex. The higher the number of votes that are cast for a vertex, the higher the importance of the vertex. Moreover, the importance of the vertex casting a vote determines how important the vote itself is, and this information is also taken into account by the ranking algorithm. The PageRank score associated with a vertex V_a is definedusing a recursive function (Sinha and Mihalcea2011):

$$PageRank(V_a) = (1-d) + d \sum_{(V_b,V_a)\in E} \frac{PageRank(V_b)}{\left|Outdegree(V_b)\right|} \tag{17}$$

where d is a parameter that is set between 0 and 1. The typical value for d is 0.85 (Brin and Page, 1998).

In a weighted graph, the decision on what edge to follow during a random walk is also taking into account the weights of outgoing edges, with a higher likelihood of following an edge that has a larger weight (Mihalcea and Tarau, 2004). Given a set of weights w_{ab} associated with edgesconnecting vertices V_a and V_b, the weighted PageRank score is determined as(Sinha and Mihalcea, 2011):

$$PageRank(V_a) = (1-d) + d \sum_{(V_b,V_a)\in E} \frac{w_{ba} PageRank(V_b)}{\sum_{(V_c,V_b)\in E} w_{cb}} \tag{18}$$

PageRank in its traditional sense corresponds to a uniform probability distribution among the vertices in the graph. Instead, biased PageRank, first mentioned in (Brin et al. 1998) and (Haveliwala 1999) and further referencedin (Haveliwala 2003), takes this idea further by introducing the concept of relative importance of the vertices. Instead of assigning the same probability to each vertex that a random surfer could potentially jump to, biased PageRank allows a certain “bias” toward certain vertices. This is done by multiplying the corresponding contributing score of a vertex by its bias weight, determined by whether that vertex belongs to a word in context or whether it is a synonym(Sinha and Mihalcea 2011).

3. E-User Index

In anonline social network, to determine effective users is an important issue about understanding the structure of the network. To determine effective users in a social network by using fundamentals centrality measures such as degree, betweenness, closeness, and eigenvector centralitiesgives to us significant information about the online social network. We think that some inputs are necessary for determining the most effective user intwitter online social network. These inputs are posted tweet count, retweet count, follower count, and number of users who aren’t in followers but do retweet that users have. With these inputs, we are able to obtain more realistic informationabout the network structure. We developed an index called E-User (Effective User) Index which can give more realistic information about the roles of users in twitter online social network.According to E-User Index,a user who has not thehighest centrality measurement value can be more effective than another user who has the highest measurement value.In other words, a user who has10 posted tweets, 50 retweets, and 50 followers will can bemoreeffective thananother user who has 10 posted tweets, 50 retweets, and 2000 followers in the twitter online social network although according to SNA, the user who has 2000 followers is more effective than the other user who has 50 followers.When we look at follower count of both users, we can see that first user who has 50 followers reached more followers with 50 retweets than second user who has 2000 followers with 50 retweets although both users have same posted tweet count, 10.These index also shows uspopularity of users in Twitter. E-User Index formula is as follows:

$$E\text{-}User(i) = \frac{RTCount_i}{PTCount_i \cdot \left(FCount_i + RTunFCount_i\right)} \text{x}100 \tag{19}$$

where $E\text{-}User(i)$ is an index value indicating the level of i^{th} user’s effect, $RTCount_i$ is the posted retweetcount of i^{th} user, $PTCount_i$ is the posted tweet count of i^{th} user, $FCount_i$ is the follower count of i^{th} user, and $RTunFCount_i$ indicates the count of users who are not in i^{th} user’s followers but do retweet i^{th} user’s tweet by using hashtags or word search etc.Because one tweet can do retweet only one time by someone, $E\text{-}User(i)$ can take min. 0 and max. 100 values.The high value of E-User Index indicates the high level of user’s effect. Though this Index, we can not only determine the most effective user but also determine the popularity of users and how their comments, or posts affect other users in the twitter online social network.

4. Centrality Measures with E-User Index for Twitter Online Social Network

The issue of the centrality measures in SNA is an important issue for understanding structure of a network. With centrality measures, researchers can determine effective nodes in a network. In this way, they can get more information about the network and nodes. The obtained information with centrality measures can also give them information about the network development process in the future. The information of nodes' role in a network can be used in various fields such as information sciences, biology, economics, social and political sciences etc. For instance, researcher may want to get information about the political trends of people in a social network. In this respect, to get more realistic results about centrality measures is important for researcher.

Figure 2 represents the process of determining the most effective user in Twitter online social network. In the first step, E-User index values and centrality measurement values are calculated for each user (nodes). In the following step, the calculated values in first step are used in a function. The function is appiled for each user. In the final step, maximum function value is assigned as the most effective user.

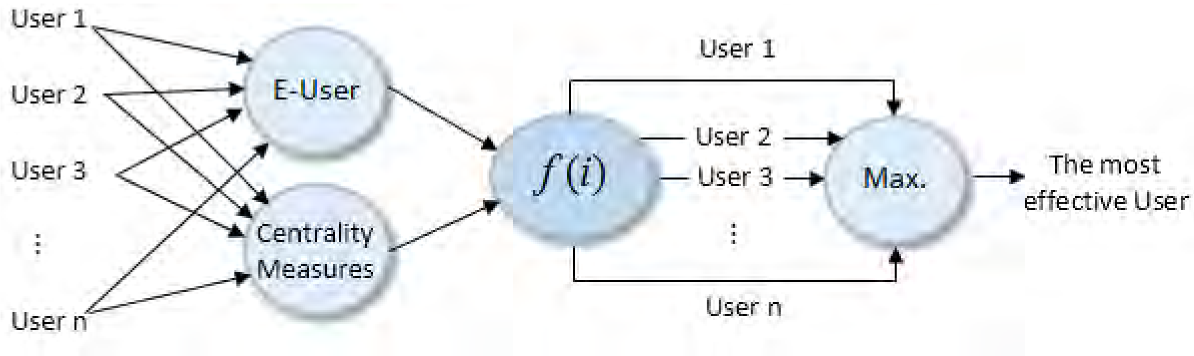

Figure 2. The process of determining most effective user

The function is as follows;

$$f(i)=C_k^{E\text{-}User,\alpha}(i)=\begin{cases} z(C_k(i))\,.\left(\dfrac{z(E\text{-}User(i))}{z(C_k(i))}\right)^{\alpha}=z(C_k(i)^{(1-\alpha)})\,.\,z(E\text{-}User(i)^{\alpha}), & z(C_k(i))\wedge z(E\text{-}User(i))>0 \\ 0 & ,\ \text{others} \end{cases} \tag{7}$$

where $z(C_k(i))$ indicates standardized value of k^{th} type centrality of i^{th} user (node), $z(E\text{-}User(i))$ indicates standardized $E\text{-}User_{Index}$ value of i^{th} user, and a indicates the contribution rate of $z(E\text{-}User(i))$ value for $z(C_k(i))$. a takes a value from 0 to 1. An a value close to 1 shows a high contribution rate of $z(E\text{-}User(i))$ value for $z(C_k(i))$. We recommend that a should take a value in $(0, 0.3]$ for a sufficient contributions of $z(E\text{-}User(i))$. In this study, we will get results for many a values. And, we will show them in table 4.

5. Application

In this section, we designed a small weighted and directed social network graph by using simulated data and used it in our study for comparison on determining the most effective user. In our graph, weights indicate the number of retweets between two users,and directions indicate which user have done retweet. The directions can be bidirected. This means that both users have done retweet to each other's tweets. The graph has two different representations for relationships. One of the representations for relationships is a red directed bridge, the other is agreendirected bridge. The red directed bridge refers to relation which is established between a user and other user who is not a follower of the user but do retweet. The green directed bridge refers to relation which is established between a user and other user who is a follower of yhe user and do retweet.The numbers on bridges (weights) refer to how many times a user have done retweet another user's tweet.

Figure 3indicates a small social network using simulated data, where the most important users according to different centrality measures has been marked.

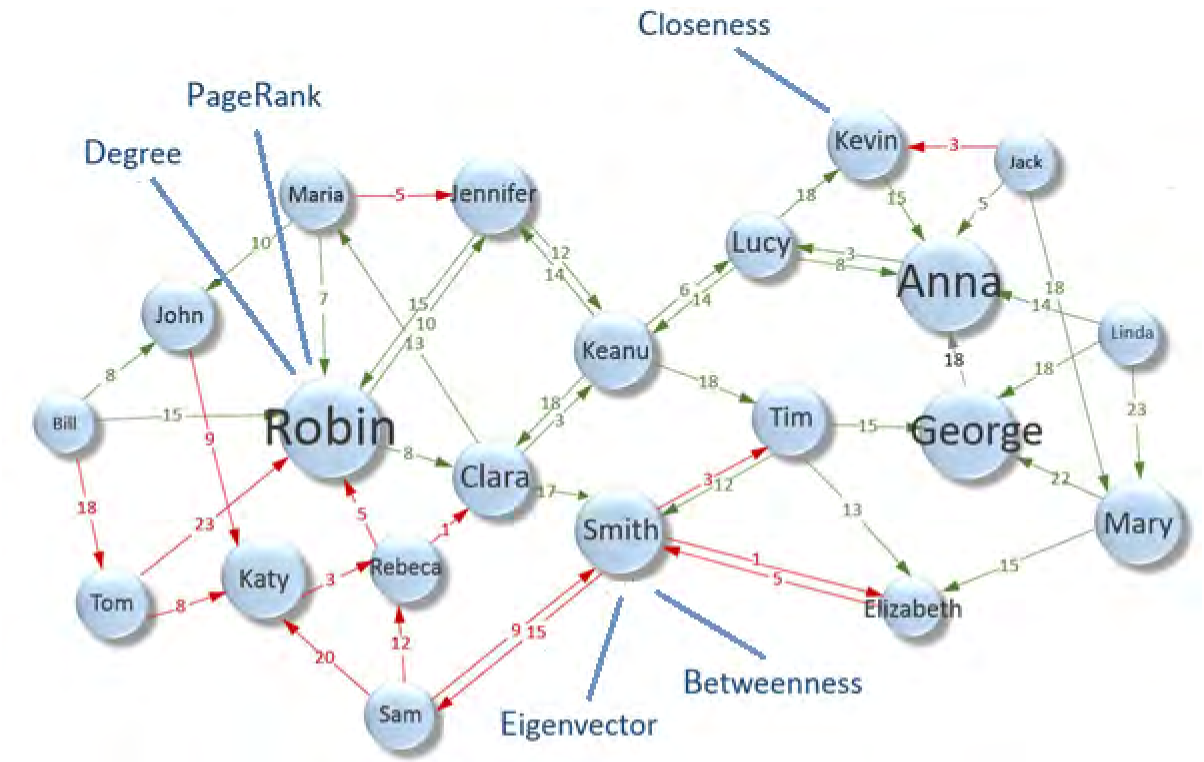

Figure 3. A small social network using simulated data, where the most important users according to different centrality measures has been marked.

Table 1 indicates users' (nodes) followers, who haven't done retweet (unretweet), or have done retweet, and unfollowers who have done retweet. For example, Clara has not done retweet Jennifer's tweets but has followed her, Robin and Keanu have done retweet to Jennifer's tweets and have followed her, and Maria has not followed Jennifer but have done retweet her tweets.

Table 1: The relation between users in Twitter

	Followers		Unfollowers
	UnReTweet	**ReTweet**	**ReTweet**
Kevin		Lucy	Jack
Jack	Maria		
Maria	Tom, Linda, Tim, John, Bill, Jack	Clara	
Jennifer	Clara	Robin, Keanu	Maria
John		Maria	
Lucy		Keanu, Anna	
Linda	Jack		
Anna	Keanu	Lucy, Kevin, Jack, Linda	George
George		Mary, Tim, Linda	
Robin	Sam, Katy	Bill, Maria, Jennifer	Tom, Rebeca
Bill	Maria,Tom,Katy		
Tom	John		Bill
Katy	Bill, Rebeca, Clara		John, Tom, Sam
Rebeca	Clara, Smith		Katy, Sam
Sam	Katy		Smith
Clara	Sam	Robin, Keanu	Rebeca
Keanu		Jennifer, Clara, Lucy	
Smith	Rebeca, Katy	Clara, Tim	Elizabeth, Sam
Tim	Elizabeth	Keanu	Smith
Mary	Tim, Anna, Lucy, Elizabeth, George	Jack, Linda	
Elizabeth	Lucy, Kevin, Jack, Jennifer, Clara, Katy	Tim, Mary	Smith

Table 2 indicates twitter information of users, centrality measured values, and calculated E-User Index Value. In centrality measured values, the largest value and other values which is close to it are expressed with blue color. In E-User Index values, the values are given with rank, and the largest value and other values which is close to it were shown with blue color.

Table 2: Centrality measured values, twitter information, and E-User Index

	Centrality Measured Values									Twitter Informations					
	Degree		Degreeα:0.5		Betweenness	Closeness		Eigenvector	PageRank	RT Count	PT Count	FCount	RTUFC	E-User	
	In	Out	In	Out		In	Out							Index	Rank
Kevin	1.0500	0.7500	0.2291	0.1936	0.0026	**0.0498**	0.0268	0.0026	0.0078	21	5	4	1	84.00	2nd
Jack	0.0000	1.3000	0.0000	0.4416	0.0026	**0.0476**	0.0292	0.0026	0.0071	0	1	1	0	0.00	17th
Maria	0.6500	1.1000	0.1803	0.4062	0.0526	0.0256	0.0375	0.3187	0.0326	13	16	7	0	11.61	16th
Jennifer	1.4500	1.3500	0.4664	0.3674	0.2816	0.0321	0.0295	0.5937	0.0794	29	16	3	1	45.31	10th
John	0.9000	0.4500	0.3000	0.1500	0.0053	0.0218	0.0366	0.1348	0.0209	18	12	2	0	75.00	3rd
Lucy	0.4500	**2.0000**	0.2121	**0.5477**	0.2079	**0.0481**	0.0284	0.1977	0.0690	9	12	2	0	37.50	14th
Linda	0.0026	**2.7500**	0.0000	**0.6423**	0.0026	0.0264	0.0248	0.1506	0.0335	0	10	1	0	0.00	17th
Anna	**3.0000**	0.1500	**0.7746**	0.0866	0.1526	0.0285	0.0275	0.5738	0.0660	60	20	5	1	50.00	9th
George	**2.7500**	0.9000	**0.7416**	0.2121	0.0974	0.0462	0.0199	0.6363	0.0514	55	19	3	0	96.49	1st
Robin	**3.2500**	0.5000	**0.9014**	0.2236	0.2263	0.0313	0.0367	0.5551	0.1110	65	22	5	2	42.21	12th
Bill	0.0026	**2.0500**	0.0000	**0.5545**	0.0026	0.0476	0.0313	0.0026	0.0071	0	9	3	0	0.00	17th
Tom	0.9000	1.5500	0.2121	0.3937	0.0053	0.0480	0.0368	0.0026	0.0098	18	12	1	1	75.00	3rd
Katy	1.8500	0.1500	0.5268	0.0866	0.0447	0.0220	0.0466	0.5883	0.0541	37	11	3	3	56.06	7th
Rebeca	0.7500	0.3000	0.2739	0.1732	0.1447	0.0233	0.0483	0.3969	0.0694	15	10	2	2	37.50	14th
Sam	0.7500	**2.0500**	0.1936	**0.5545**	0.1289	0.0218	0.0418	0.6346	0.0652	15	14	1	1	53.57	8th
Clara	0.9500	1.6500	0.3775	0.4975	0.2026	0.0373	0.0442	0.5794	0.0763	27	12	3	1	56.25	6th
Keanu	1.4500	2.8000	0.4664	**0.7483**	**0.4632**	0.0458	0.0353	0.4921	0.0635	29	15	3	0	64.44	5th
Smith	2.1500	0.9500	0.6557	0.3775	**0.7105**	0.0320	0.0326	1.0000	0.0866	43	17	4	2	42.16	13th
Tim	1.0500	**2.0000**	0.3240	**0.5477**	0.1184	0.0342	0.0272	0.5016	0.0361	21	10	2	1	70.00	4th
Mary	2.0500	1.8500	0.4528	0.4301	0.0526	0.0177	0.0253	0.1465	0.0232	41	13	7	0	45.05	11th
Elizabeth	1.4500	0.2500	0.4664	0.1118	0.0368	0.0322	0.0289	0.4112	0.0290	29	11	8	1	29.29	15th
								Eigenvalue=	**23.6366**	**Max. Index Value=**				**96.49**	**George**

Figure 4 indicates calculated E-User Index values by using twitter information of users. It also shows the popularity of users. We can see that the first user who has the highest value is George with 96.49 index value, the second user who has the 2rd highest value is Kevin with 84 index value, and the third user who has 3rd largest value is John and Tom with 75 index value etc.

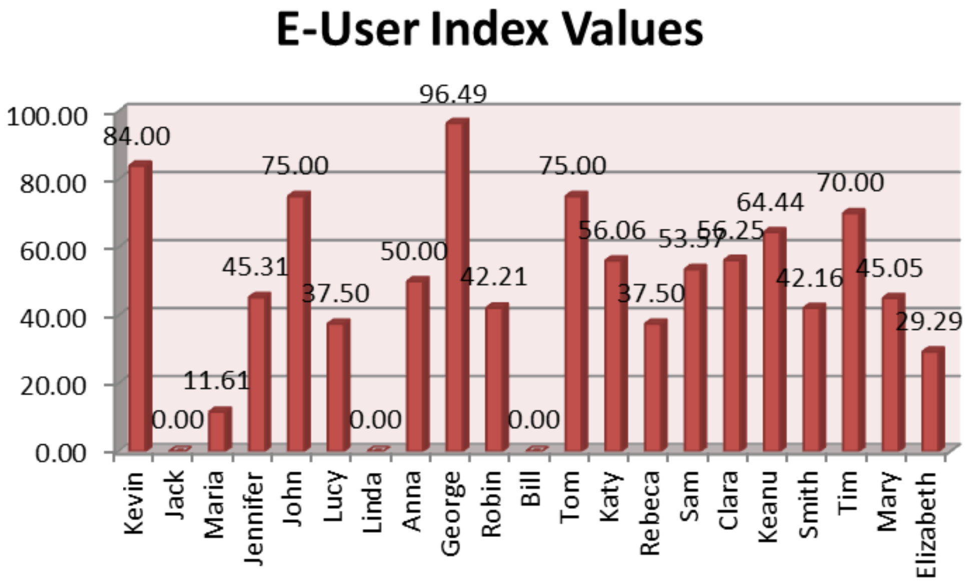

Figure 4. Calculated E-User Index values by using twitter information of users

Table 3 indicates standardized centrality measured values and standardized E-User Index values. These values were used while calculated robust centrality measured values.

Table 3: Standardized centrality measured and E-User Index values

	Standardized Measure Values									
	Degree		Degree, a=0,5		Betweenness	Closeness		Eigenvector	PageRank	E-User Index
	In	Out	In	Out		In	Out			
Kevin	-0.2484	-0.6551	-0.5657	-0.9168	-0.8031	1.4794	-1.5465	-1.895	-1.3672	1.419
Jack	-1.3885	0.0266	-1.4906	0.3803	-0.8031	1.276	-0.4609	-1.395	-1.3913	-1.7393
Maria	-0.6827	-0.2213	-0.7629	0.1952	-0.511	-0.8241	0.597	-0.2181	-0.5147	-1.3029
Jennifer	0.1899	0.0885	0.3919	-0.0077	0.8265	-0.2093	-0.4276	0.8061	1.0942	-0.0356
John	-0.4113	-1.0269	-0.2796	-1.1451	-0.7877	-1.1909	0.4909	-0.9027	-0.9169	1.0806
Lucy	-0.8999	0.8941	-0.6343	0.9356	0.3961	1.3197	-0.5731	-0.6684	0.7366	-0.3293
Linda	-1.3856	1.8236	-1.4906	1.4301	-0.8031	-0.7532	-1.0889	-0.844	-0.4837	-1.7393
Anna	1.8687	-1.3987	1.6361	-1.4768	0.0732	-0.5491	-0.6837	0.7319	0.6335	0.1406
George	1.5973	-0.4692	1.503	-0.8201	-0.2496	1.1396	-1.6613	0.9648	0.1316	1.8887
Robin	2.1402	-0.9649	2.1479	-0.7601	0.5037	-0.2811	0.4996	0.6623	2.1805	-0.1523
Bill	-1.3856	1.8236	-1.4906	0.9712	-0.8031	1.276	-0.1894	-1.395	-1.3913	-1.7393
Tom	-0.4113	0.3364	-0.6343	0.1298	-0.7877	1.3087	0.5083	-1.395	-1.2985	1.0806
Katy	0.6201	-1.3987	0.6358	-1.4768	-0.5571	-1.1658	1.7776	0.7861	0.2244	0.3685
Rebeca	-0.5742	-1.2128	-0.3851	-1.0237	0.0271	-1.0459	1.994	0.0731	0.7504	-0.3293
Sam	-0.5742	0.956	-0.7089	0.9712	-0.0652	-1.1841	1.1508	0.9585	0.606	0.2749
Clara	-0.357	0.4603	0.0332	0.6728	0.3653	0.2927	1.4721	0.7528	0.9876	0.3756
Keanu	0.1859	1.8855	0.3919	1.985	1.8873	1.0993	0.3162	0.4277	0.5476	0.6837
Smith	0.9459	-0.4072	1.1563	0.045	3.3324	-0.2142	-0.0316	2.3192	1.3417	-0.1543
Tim	-0.2484	0.8941	-0.1826	0.9356	-0.1266	0.0001	-0.7175	0.4633	-0.3943	0.8926
Mary	0.8373	0.7082	0.337	0.3203	-0.511	-1.5742	-0.9663	-0.8591	-0.8378	-0.0453
Elizabeth	0.1859	-1.2747	0.3919	-1.3449	-0.6032	-0.1995	-0.5098	0.1265	-0.6384	-0.6379

Figure 5 shows comparison of standardized centrality measures with standardized E-User Index. We took advantage of standardized centrality measured values and E-User Index values while we were creating it.

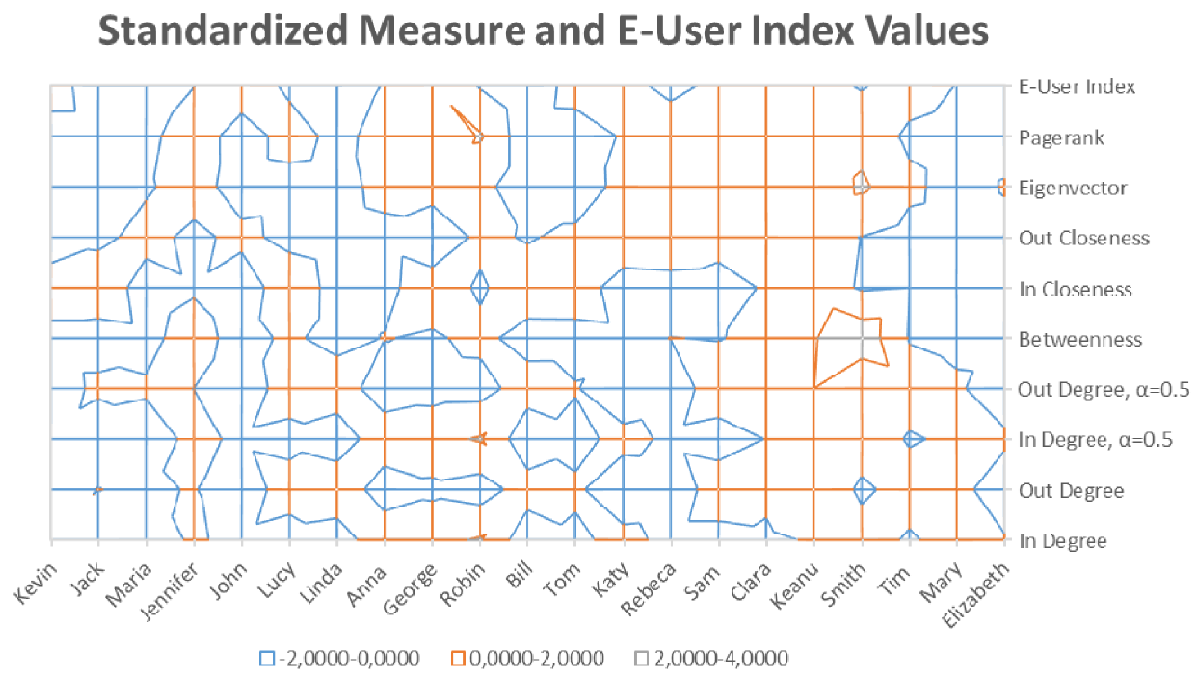

Figure 5. The comparison of standardized centrality measures with standardized E-User Index

Table 4 indicates results of the most effective user for each centrality measure with the contribution rates a of $z(E\text{-}User(i))$ value for $z(C_k(i))$ .We recommend a value a in $(0, 0.3]$ for researchers.

Table 4: Results of the most effective user for each centrality measure with a values

a	The Most Effective User								
	Degree		Degree, a=0,5		Betweenness	Closeness		Eigenvector	PageRank
	In	Out	In	Out		In	Out		
0	Robin	Keanu	Robin	Keanu	Smith	Kevin	Rebeca	Smith	Robin
0.1	George	Keanu	George	Keanu	Keanu	Kevin	Katy	George	Clara
0.2	George	Keanu	George	Keanu	Keanu	Kevin	Katy	George	Clara
0.3	George	Keanu	George	Keanu	Keanu	Kevin	Katy	George	Clara
0.4	George	Keanu	George	Keanu	Keanu	Kevin	Katy	George	Clara
0.5	George	Keanu	George	Keanu	Keanu	George	Katy	George	Keanu
0.6	George	Keanu	George	Keanu	Keanu	George	John	George	George
0.7	George	Keanu	George	Keanu	Keanu	George	John	George	George
0.8	George	Tim	George	Tim	Keanu	George	John	George	George
0.9	George	Tim	George	Tim	Keanu	George	John	George	George

In this study, we see that there are notany differencesamong $a = 0.1$, $a = 0.2$, and $a = 0.3$ results. The calculation with $a = 0$ value indicates centrality measures results without E-User Index values. We also see that there is a difference between $a = 0$ and $a = 0.1$ results. This means that E-User Index with $a = 0.1$ has a significant effect on centrality measures results. Before taking into account E-User index, we can say that Robin is In Degree and PageRank centrality measures results, Keanu is Out Degree centrality measure result, Smith is Betweenness and Eigenvector centrality measures results, Kevin is In Closeness centrality measure result, and Rebeca Out Closeness centrality measure result. After taking into account E-User Index with $a = 0.1$,or $a = 0.2$,or $a = 0.3$, we can say that George is In Degree and Eigenvector centrality measures results, Keanu is Out Degree and Betweenness centrality measures results, Kevin is In Closeness centrality measure result, Katy is Out Closeness centrality measure result, and Clara is PageRank centrality measure result. We see that the result of centrality measurement has changed with E-User Index.

With E-User Index, Degree centrality measure result is George instead of Robin. Robin has 65 retweet count, 22 tweet count, 5 follower count, 2 unfollowers but did retweet count, and 42,21 E-User Index value, but George has 55 retweet count, 19 tweet count, 3 follower count, 0 unfollowers but did retweet count,and 96.49 E-User Index value. So, we can say that George is more effective than Robin. Betweenness centrality measure result is Keanu instead of Smith. Smith has 43 retweet count, 17 tweet count, 4 follower count, 2 unfollowers but did retweet count, and 42.16 E-User Index value, but Keanu has 29 retweet count, 15 tweet count, 3 follower count, 0 unfollowers but did retweet count,and 64.44 E-User Index value. So, we can say that Keanu more effective than Smith. Out Closeness centrality measure result is Katy instead of Rebeca. Rebeca has 15 retweet count, 10 tweet count, 2 follower count, 2 unfollowers but did retweet count, and 37.50 E-User Index value, but Katy has 37 retweet count, 11 tweet count, 3 follower count, 3 unfollowers but did retweet count,and 56.06 E-User Index value. So, we can say that Katy more effective than Rebeca. Eigenvector centrality measure result is George instead of Smith. Smith has 42.16 E-User Index value, but George has 96.49 E-User Index value. So, we can say that George more effective than Smith. PageRank centrality measure result is Clara instead of Robin. Robin has 42.21 E-User Index value, but Clara has 56.25 E-User Index value. So, we can say that Clara more effective than Robin.

6.Conclusions and Future Work

In this study we analyzed the issue that determination of the most effective user in online social networks. According to our research and work which is based on designing a small weighted and directed social network graphformed withsimulated data, we were able to show that we can obtain more realistic results by using E-User Index. Through E-User Index, we have seen that we can determine not only the most important and effective user but also show popularity of users.

As a future point of interest, we would like to study on established different relationships in online social networks.In situations different relationshipscan be established in social online networks, we think that new methods or approaches can be developed.

Acknowledgments

We extend our thanks to Etiya Information Technologies Inc., Istanbul/Turkey and Abdulkerim Mizrakfor supporting our study, and also thank Dr. Oguz Akbilgic for sharing his knowledge with us.

References

[1] S. Bandyopadhyay, A. R. Rao, and B. K. Sinha, Models for social networks with statistical applications. SAGE Publications 2010.

[2] H. Bozdogan andO. Akbilgic, Social network analysis of scientific collaborations between different cross disciplinary fields. Information Services& Use 2013, Vol: 33, pp. 219-233.

[3] U. Brandes,On variants of shortest-path betweenness centrality and their generic computation. Social Networks 2008, 30: 136–145. doi:10.1016/j.socnet.2007.11.001

[4] S. Brin, R. Motwani,L. Page, and T. Winograd,What can you do with a web in your pocket? IEEE Data Engineering Bulletin 1998, 21(2):37–47.

[5] M. Chen, R. A. Chowdhury, V. Ramachandran, D. L. Roche, and L. Tong, Priority queues and dijkstra's algorithm. UTCS Technical Report TR-07-54, 2007 http://www.cs.sunysb.edu/~rezaul/papers/TR-07-54.pdf

[6] E. W. Dijkstra,A note on two problems in connection with graphs. Numerische Mathematik 1959, Vol:1: pp:269–271

[7] C. Freeman, A set of measures of centrality based on Betweenness. Sociometry1977, Vol:40, pp:35-41.

[8] L. C. Freeman, Centrality in social networks conceptual clarification. Social Networks 1978/79, Vol.1, pp.215-239. doi:10.2307/3033543

[9] L. C. Freeman, The development of social network analysis: A study in the sociology of science.Vancouver, CA: Empirical Press 2004.

[10] V. Grienitz andA. M. Schmidt, Scenariobased complexity management by adapting the methods of social network analysis. Proceedings of The International Multi-Conference on Complexity, Informatics and Cybernetics 2010, pp. 61-66.

[11] C. Hart, Graph Theory Topics in Computer Networking. University of Houston-Downtown, Department Computer and Mathematical Sciences, Senior Project 2013.

[12] T. Haveliwala, Efficient computation of pagerank. Technical report 1999.

[13] T. H. Haveliwala, Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search. IEEE Transactions on Knowledge and Data Engineering 2003, 15:784–796.

[14] V. Latora andM. Marchiori, A measure of centrality based on network efficiency. New Journal of Physics 2007, Vol:9, Issue: 6, pp. 188.

[15] R. Mihalcea andP. Tarau, TextRank – bringing order into texts. In Proceedings of the Conference on Empirical Methods in Natural Language Processing 2004.

[16] M. E. J. Newman,A measure of betweenness centrality based on random walks. 2003, http://arxiv.org/abs/cond-mat/0309045

[17] M. E. J. Newman, Analysis of weighted networks. Physical Review E 70, 056131, 2004.

[18] J. Nieminen, On centrality in a graph. Scandinavian Journal of Psychology 1974, Vol:15 pp. 322-336

[19] T. Opsahl, F. Agneessens, and J. Skvoretz, Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks 2010,Vol: 32pp:245-251 doi:10.1016/j.socnet.2010.03.006

[20] D. L. Passmore, Social network analysis: Theory and applications. Creative Commons Attribution-Share Alike 3.0 Unported 2011. http://creativecommons.org/licenses/by-sa/3.0/

[21] G. Sabidussi, The centrality index of a graph. Psychometrika 1966,Vol:31 pp. 581-603

[22] R. Sinha, R. Mihalcea, Using centrality Algorithms on directed graphs for synonym expansion. FLAIRS Conference, AAAI Press 2011. http://www.cse.unt.edu/~rada/papers/sinha.flairs11.pdf

[23] H. Soh, S. Lim, T. Zhang, X. Fu, K. K. G. Lee, T. G. G. Hung, P. Di, S. Prakasam, and L. Wong, Weighted complex network analysis of travel routes on the Singapore public transportation system. Physica A 389 5852-5863, 2010.

[24] H. N. Su, P. C. Lee, Assessment of thermal-stable polymer nanocomposite techniques by patent citation network analysis. Journal of Business Chemistry 2009, Vol. 6, pp. 108-125

[25] H. Zhuge, J. Zhang, Topological centrality and its e-Science applications. Journal of the American Society for Information Science and Technology 2010, Vol:61, pp. 1824-1841