



E-İşte Sürdürülebilir Bağlantılılığı İzlemek için Ağ Tabanlı Teorinin Kullanımı

Araştırma Makalesi/Research Article

 Ziya Nazım PERDAHÇI^{1*},  Mehmet Nafiz AYDIN²

¹Enformatik Bölümü, Mimar Sinan Güzel Sanatlar Üniversitesi, İstanbul, Türkiye

²Yönetim Bilişim Sistemleri Bölümü, Kadir Has Üniversitesi, İstanbul, Türkiye

nz.perdahci@msgsu.edu.tr, mehmet.aydin@khas.edu.tr

(Geliş/Received: 15.11.2019; Kabul/Accepted: 16.03.2020)

DOI: 10.17671/gazibtd.647396

Özet— Çevrimiçi Etkileşimli Platformlar, yüz binlerce birbirine bağlı kullanıcı oluşturabilen iletişim, bilgi alışverişi ve bağlantılar gibi çeşitli hizmetler sunar. Son çalışmalar, platform kullanıcılarının birbirine bağlılığının, yani kullanıcıların birbirine bağlılığa yönelik ortak eğiliminin büyümesinin ve sürdürülebilirliğinin, bu tür e-İşletmelerde başarılı olmak için iki temel bileşen olduğunu göstermektedir. Bu makale, çevrimiçi etkileşimli platformların yöneticilerine, kullanıcıların birbirine bağlılığının büyümesini ve sürdürülebilirliğini izlemek için analitik bir araç sağlamak amacıyla e-İş bağlamında “ağ korelasyonu” kavramını araştırmayı amaçlamaktadır. Hem statik hem de zamana bağlı ağ korelasyon analizi yapmak için Doktorsitesi.com’u dijital iz verisi kaynağı ve kullanıcı meta verileri (rol, cinsiyet) kaynağı olarak kullanarak kullanıcıların birbirine ne ölçüde bağlı olduğuna ilişkin ağ korelasyonlarının ve kullanıcı meta verilerine ilişkin ağ korelasyonlarının, kullanıcıların birbirine bağlı büyümesini nicel olarak sürdürmenin yollarını izlememize ve önerebilmemize yardımcı olabileceğini gösteriyoruz. Birlikte ele alındığında, sonuçlarımız ağ korelasyon analizlerinin, çevrimiçi etkileşimli platformlarda sürdürülebilir kullanıcı bağlantılılığını daha iyi tanımlamak, tavsiye etmek ve ön görmek için analitik sunabildiğini göstermektedir. Buna göre, platform yöneticileri, yeni kullanıcıların, birbirine bağlılık analitiğine dayalı olarak birbirine bağlılıklarını artırmalarına yardımcı olan bir platform özelliği eklemeyi düşünebilir.

Anahtar Kelimeler— sürdürülebilir e-iş, ağ teorisi, ağ korelasyonu, bağlantılılık

Using the Network-Based Theory for Monitoring Sustainable Interconnectedness on e-Business

Abstract— Online Interactive Platforms offer a variety of services including communication, information exchange, and connections that can create hundreds of thousands of interconnected users. Recent studies suggest that growth and sustainability of platform users’ interconnectedness, that is, users’ collective tendency to interconnectedness, are the two essential ingredients to succeed in such e-Businesses. This paper aims to explore the notion of “network correlation” in e-Business context in an attempt to provide managers of online interactive platforms with an analytical tool for monitoring growth and sustainability of users’ interconnectedness. We leverage Doktorsitesi.com as a digital trace data source and user metadata (role, gender) source to conduct both static and temporal network correlation analysis, and show that network correlations regarding the extent to which users are interconnected and network correlations regarding user metadata can help us to monitor and to suggest ways to quantitatively sustain growth of users’ interconnectedness. Taken together, our results suggest that network correlation analyses is capable of offering analytics to better describe, prescribe, and predict sustainable users’ interconnectedness on online interactive platforms. Accordingly, platform managers may consider the addition of a platform feature that helps newcomers with growing their interconnectedness based on the interconnectedness analytics.

Keywords— sustainable e-business, theory of network, network correlation, interconnectedness

1. INTRODUCTION

Online interactive platforms (OIPs) provide users with interconnected services including communication (e.g., WhatsApp, Snapchat), information exchange (e.g., WebMD, Healthtap, Instagram), and connections (e.g., LinkedIn, Facebook). Electronic business (e-Business) is a promising application domain in which one can make use of OIPs for establishing innovative services leading to data-driven business. The idea of users' interconnectedness signifies the underlying (social as well as economic) values for users such as bringing information seekers and providers together along with multitude of services. It is this interconnectedness aspect that is examined as research matters related to, for instance, virtual communities [1], information sharing [2], business intelligence [3, 4], and robustness in large social networks [5].

e-Business studies focusing on OIPs confirm that growth and sustainability of users' interconnectedness (that is, collective tendency toward interconnectedness) are essential to achieve successful e-Business [6,7]. Lifecycle models of OIPs typically indicate various stages across timescale for growth and sustainability of interconnectedness [8]. Additionally, real-time data-driven business intelligence [9] utilizing metadata (e.g., profile such as age, gender, profession) [10] is essential to monitor the progress of collective behavior [11] that can be associated with service attractiveness and sustainable interconnectedness [12]. The current state of things in practice as well as in e-Business publications in this regard appear to be mainly concerned with conventional key web and social media performance indicators such as benchmarks on active users, sources of Web traffic, navigation behavior of users, engagement, and conversion metrics [1, 13]. These measures are useful but limited in truly reflecting an appropriate ground to utilize metadata for examining users' interconnectedness of various kinds of relations.

Users' interconnectedness, or interconnectedness for short, naturally leads to the establishment of things and their relations that one can conceive as a network. E-Business data-driven analytics focusing on interconnectedness on OIP necessitates network understanding, which calls for a network-based theory. A network can exhibit various correlations between its elements (i.e., nodes) and their interconnectedness (i.e., edges). These correlations can be concerned with either metadata of users or to the extent of users' interconnectedness. Essentially, correlations as network analytics attempt to exhibit how and on which metadata, users are selective in establishing interconnectedness collectively.

Typically, platform users may provide metadata such as gender, age and profession. For a researcher, the availability of data is one thing but its quality in general and validity in particular is utmost importance to achieve scientific rigor. As shall be seen later on, we obtained the digital trace data of one of the leading European online health platforms and considered gender and role as user

metadata and the established ties as interconnectedness data to examine network correlations over time.

The basic premise of this research is that static and dynamic network-based correlation analysis helps managers to describe, predict, and prescribe their business with respect to growth and sustainability of user interconnectedness. [12] have emphasized the importance of a network-based approach, rooted in network science, to better examine the notion of interconnectedness by employing dynamic (temporal) analysis. From a network theory point of view, growth of e-Business platforms that promote interconnectedness with respect to reciprocal communication and exchange of information can be measured by monitoring the emergence and evolution of the network components in general, and the Giant Component (GC) [14], in particular. [12] discovered a phenomenon, as they observed evolution of the GC, which was born about sixteen months since the service launched. This phenomenon relates to a metadata based quantity that is approximately conserved over time. This approximately conserved quantity (ACQ) is related to the ratio of users of two different metadata, namely, patients and physicians, which maintains a nearly constant value in a narrow band around four as soon as sustainable stage is achieved. This research is aimed to explore the reasons for observing ACQ by examining degree correlation and user metadata correlations over time for the GC. We argue that this exploration helps to better describe, prescribe, and predict sustainable interconnectedness on OIPs.

2. RESEARCH BACKGROUND

Scholars have examined OIP as e-Business practice from different perspectives. In the domain of marketing research and information systems research, virtual community platforms have been studied by examining behaviors of users and value of product and customers [15]. A life-cycle perspective identifies distinctive lifecycle stages of evolving communities and suggests what conditions to be incorporated for each stage so that lasting interactions can be achieved [8]. Roles of members have been discussed as an important factor for especially a growth stage of platforms, but detailed analysis of how role attribute affects formation of network is limited.

The last two decades have witnessed a significant increase in the attention of scholars to better understand complex systems characterized by many entities and their interactions [16]. Network Science has advanced our comprehension of the actual nature of these systems in diverse contexts such as socio-technological, health, biological, and political sciences where they are revisited and represented as interaction networks. Although the research questions discussed are very different from each other and therefore the research groups embroider the networks differently, it has been observed that networks used in the representation of complex systems in various research areas such as social and management sciences, engineering, natural sciences and life sciences have some common characteristics. Thus, scientists have taken

seriously the task of developing new theoretical calculations and a scientific approach to better understand complex systems through networks. Network Science offers academics and practitioners a new mathematical framework and scientific methodology for better exploration of real-world complex systems.

Theoretical accounts associated with complex systems have been examined as social network theory [17, 18, 19], theory of network [20], and network science [21]. Scholars including [22] employ theory of network to examine communication and interaction patterns realized on interactive health platforms. Graph theory is used as a means to create mathematical models that represent complex systems, which manifest a sheer number of interactions enabled by interactive platforms. These mathematical models help us discover phenomena that lead to better understanding and prediction of complex systems. For example, the preferential attachment model of [23] has been tested with respect to explanatory and predictive power of model in various contexts. As to be discussed later on, heavy-tailed degree distributions of networks help us in characterizing them as real-world networks. Scholars conduct macro and micro levels of analysis to identify properties of networks.

At the macroscopic level, various measures including average path length and clustering coefficient have been used to examine structural characteristics of networks. For instance, a short path length irrespective to network size signifies high possibility for existence of small-world phenomenon [11]. At the microscopic level, those network characteristics that take into account node and edge attributes are subject to investigation. For instance, one of the most studied issues is to identify importance of entities in the network. Network measures of centrality such as degree, closeness, betweenness, and eigenvector centrality can be used for this purpose.

A node's degree signifies connectivity, giving us the number of interactions it has with other nodes. The degree distribution is a fundamental network property which tells the probability (p_k) that a node selected at random will have a particular degree. In this article, a set of interconnected nodes, a set having all constituent parts (nodes) connected is referred to as a network component. A single conspicuously large network component qualifies as the giant component (GC) if it fulfils two conditions: (i) it must comprise more than half of the nodes and links and (ii) no other component comprises more than a small amount per mill of the nodes and edges. The GC will eventually emerge if average degree connectivity per node is greater than one [24]. Nodes can be very selective in their ties. Two most commonly used terms for selective linking in literature are "homophily" [25, 26] and assortative mixing [27]. Assortative mixing coefficients [28] are much like Pearson's correlation coefficient for a certain characteristic of nodes that have common edges. Each coefficient is normalized to the range between -1 and 1, inclusive. The coefficient is equal to 1 when assortative mixing is impeccable, that is, when edges form only among vertices

of the same category. It takes negative values when nodes mix disassortatively. Node characteristics can be discrete or scalar. While gender and race are among the examples of discrete characteristics, degree and age are scalar characteristics. In what follows, user metadata or metadata will refer to the node characteristics while interconnectedness will refer to all structural network metrics. For each metadata, the corresponding mixing matrix quantifies the magnitude of the assortative mixing. By a mixing matrix we mean a table on which each cell provides the fraction of edges that link pairs of user metadata (e.g. role-to-role, role-to-gender). While in the case of metadata characteristics, the correlation coefficient obtained from the mixing matrix is the sole reference, things might get a little more involved with interconnectedness characteristics, where the correlation function ($k_{nn}(k)$) captures the nature of correlations better. We first probe correlations by taking the mean of a certain connectivity characteristic of the neighbors of each node, and then taking the mean of this quantity over the nodes with the same characteristic value gives us the correlation function. The plot of a correlation function might exhibit assortative behavior if it is an increasing function of a certain characteristic and disassortative otherwise. The slope of the line fitted to this plot gives us the correlation exponent, which is a better measure of correlations especially for disassortative behavior [29].

For a real-world complex network, its degree distribution rather than individual degrees of its nodes rose to prominence due to the discovery of the scale-free property of these distributions [23]. The coexistence of few nodes with large degrees ("hubs") and many nodes with small degrees ("spoke") is a signature of scale-free networks. Indeed in [30] researches the hub-and-spoke character of the tie network extensively, where the authors conclude that cooperation between preferential attachment and homophily could explain the formation of ties.

Real-world complex networks are dynamic in nature, growing out of events that occur over a period. Although static network analysis of a single snapshot of aggregated events provides us with valuable insights into the structure of established ties on an e-business platform, it fails to address some key business questions like customer loyalty, long-term sustainability of a service, and typical timescale for service maturity. In this regard, the authors in [31] discuss the development of hubs, and [12] discuss the development of the giant component (GC). The former article is a first step towards addressing customer loyalty while the latter is a pioneering research addressing long-term sustainability of a service.

Unlike most of the social networks [27], the majority of hub nodes of the tie network tend to behave disassortatively by both metadata and interconnectedness characteristics. In other words, highly tied nodes tend to avoid tying to each other. In [21], the author explains vividly the theoretical reasons behind the difficulty for a disassortative network to form a GC, which is in line with

our findings of the timescale of the service/product maturity [12].

While the observation of hub-and-spoke character, assortative mixing of hubs, and the discovery of a phenomena related to a stationary state coinciding with the emergence of the GC have been researched, the mechanisms behind them, and the way interactions make a network up as it grows by the addition of nodes and edges, have remained largely elusive.

How is the timescale for the emergence of the GC affected by the decisions on the service rules? In the present study, we delve further into the development of GC-wide assortative mixing both according to user metadata and node interconnectedness by taking snapshots of aggregated events as a function of time for investigating the timescale of GC development. Besides GC, we shall utilise additional basic network concepts. One of the basic concepts is the arithmetic mean degree of a node, which is twice the number of undirected edges divided by the number of nodes. The length of the shortest walk, i.e., the number of edges traversed along the walk, is the distance between two nodes. The mean path length is the arithmetic mean of all distances. The network diameter is the largest distance while the radius is the smallest distance.

3. RESEARCH CONTEXT AND METHOD

Health industry has been going through transformation by means of emerging OIP. In particular, social networks and virtual communities supporting OIPs (e.g., WebMD, Healthline, HealthTap, WebMD, Doktorsitesi) in health domain have been increasing as they bring knowledgeable health experts (physicians and other relevant actors) and health information seekers (patients and alike) together. It is the fact that in the health domain, peers' opinions for medical practitioners and patients' experience for "like-minded others" are deemed to be useful for decision-making [32]. Utilizing OIPs as data sources for conducting network science-driven research is scarce with few exceptions such as [30, 33]. In [30] the researchers employ one of the main online interactive health information exchange Web site as a data source for constructing an overall network to find out if and how health information seekers' and advice providers' interactions are established, and [33] is a pioneering work elaborating the development of a real-world GC.

In this study, along with the present research focus, we continue to leverage Doktorsitesi.com as a data source similar to the work [34]. As one of the pioneering online interactive health networks (www.doktorsitesi.com) in Europe (founded in 2004), Doktorsitesi.com allows its members to access the true identities of physicians but not patients or alike. The platform clears health experts and health information seekers (presumably patients or their close relatives) to access to patients' anonymous user profiles where optional metadata about the users, namely gender, educational level, and age appear.

The health OIP interconnectedness service under investigation is the Connections, which provides the platform users with a feature that allows them to establish private ties. A private tie signifies privacy, enabling users exchange messages through the messages service. The connections protocol is simple: members can make a request for a tie and wait for its approval, or give approve of a connection request. Digital trace data [35] pertaining to the Connections service were not readily available, and we made a formal request of it to Doktorsitesi.com. The dataset thus obtained comprise records of members who used the service; either to seek approval for a tie or to approve of a tie, over the five and a half years period from January 2009 to July 2014. It is worth noting that this service initialized in December 17, 2008 and it was dormant for about three moths (May 25, 2012 and July 2, 2012) due to platform migration. In October 2013, the graphical user interface was completely changed after which the service was no longer promoted though it was still active. This management decision was based on conventional performance metrics. For articulation of management and system related decisions and service rules, we held seven one-hour in-depth meetings with the Chief Executive Officer (CEO) and several in person and online meetings with the system administrators.

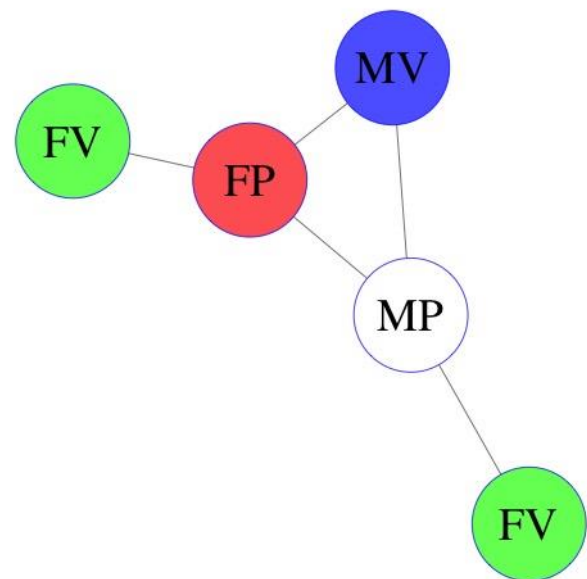


Figure 1. Network model of the Connections service

A traditional relational database system records the digital trace of tying events that occur at particular points in time as longitudinal data. This necessitates doing a preliminary work on the transformation of the records into an appropriate form amenable to network analysis. What we do is to aggregate those events that grant approval for ties to construct a network. The ties network thus obtained, by its nature, is dynamic that evolve in time. Snapshots taken at regular intervals of time will measure the network, thereby giving us the opportunity to observe the timescale for the emergence of a well-connected network fragment whose relative size is large enough to qualify as a virtual community.

We have done all preliminary work on the preparation of the raw data as well as the network analysis with R and Python. R is also our sole choice for the visual analysis of network diagrams. For the sake of doing reproducible research and validity of the results, we adhere mostly to generally accepted software libraries of the research field, namely igraph [36] and NetworkX [37].

In the present study, we model the private ties (i.e., interconnectedness) as an undirected network. Our approach presumes that every single tie represents a channel for reciprocal communication; one through which two platform members agree to exchange health-related information.

A graph of undirected edges models the network of ties (see Figure 1). On each of the graph illustrations (see Figures 1 and 2) red nodes, white nodes, green nodes and blue nodes represent female practitioners (FP), male practitioners (MP), female visitors (FV), and male visitors (MV), respectively. Black edges between the nodes represent reciprocal connections as ties. Our convention on colored nodes serves two purposes: providing a decent visualization [38] and playing a vital role in network correlation analysis.

4. FINDINGS

4.1. Static Analysis of Degree Correlation and User Metadata Correlations

All network measures and diagrams belong to the snapshot of the tie network taken as of July 2014. In this regard, this is the last cross section of the structure of the Connection network. We kindly refer the interested reader to [12] for a detailed discussion of the time development of the GC of this network.

Table 1. Summary of Network Analysis measures for the overall tie network and GC

Connections Network	Network Statistics and Measures					
	<i>N</i>	<i>E</i>	<i>AD</i>	<i>APL</i>	<i>D</i>	<i>R</i>
Overall	11559	27500	2.4	5.72	17	1
GC	8209	23606	2.9	5.73	17	9

The network of ties focuses on reciprocated connections. Table 1 presents basic characteristics of the network, where *N*, *E*, *AD*, *APL*, *D*, *R* stand for the number of nodes, the number of edges, average degree, average path length, diameter and radius, respectively. To better visualize highly tied nodes, we provide network models by Fruchterman-Reingold layout algorithm [39] (see Figure 2). In addition, the algorithm provides a visual representation that brings out the giant component (GC). In the figure, GC is apparent at the center. Much smaller fragments (network components of a few nodes and ties) are scattered around the GC like the asteroid belt.

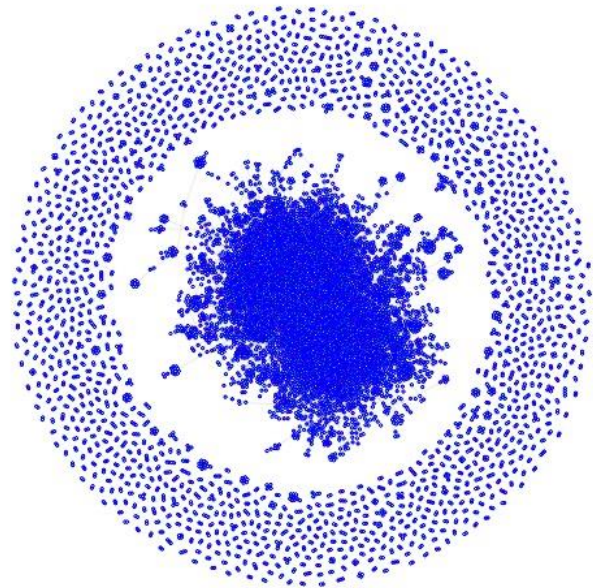


Figure 2. Graph of users' interconnectedness

The network of ties has $N = 11559$ platform members represented by colored nodes and $L = 13750$ ties between them represented by undirected edges (see Figure 2). The mean number of 2.38 ties indicates that a typical platform member secures two to three private ties, but the larger number of members have secured only one tie nevertheless.

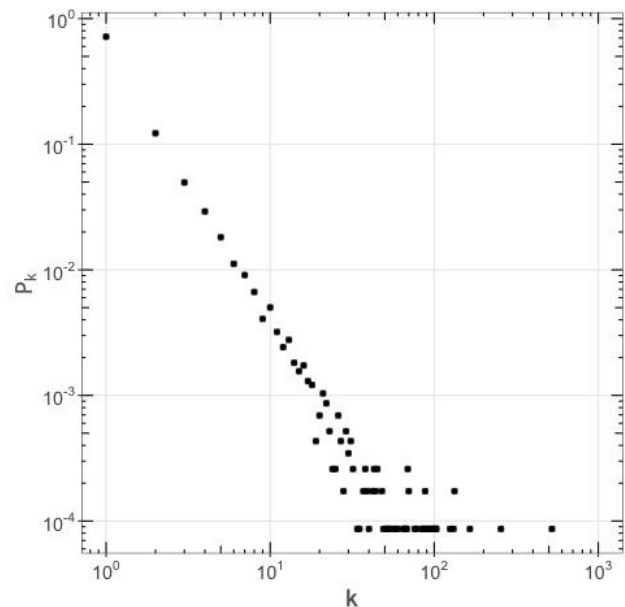


Figure 3. Interconnectedness degree distribution. We observe a nonrandom pattern which is usual for real-world networks telling us that the established ties result from conscious choices of the users rather than random choices.

The log-log plot of the degree distribution (see Figure 3) reveals the hubs and spokes: 72% members (spokes) have established only one tie (the highest probability on the left),

12% of them in 2 ties. Members with more than ten ties constitute a mere 2.8% of the users. There appear a small number of conspicuous hubs on the far right of the distribution with very high degrees and the lowest probability on the left most of the plot. 71% of all members and 86% of all connections are absorbed by the GC (see Figure 2).

The average distance (mean path) between the users of the GC in the tie network is 5.73, meaning that members are, on the average, "less than six people apart" should they seek for more health-related information. There are also remotely tied nodes as extremum distances range from network diameter of 17 to network radius of nine.

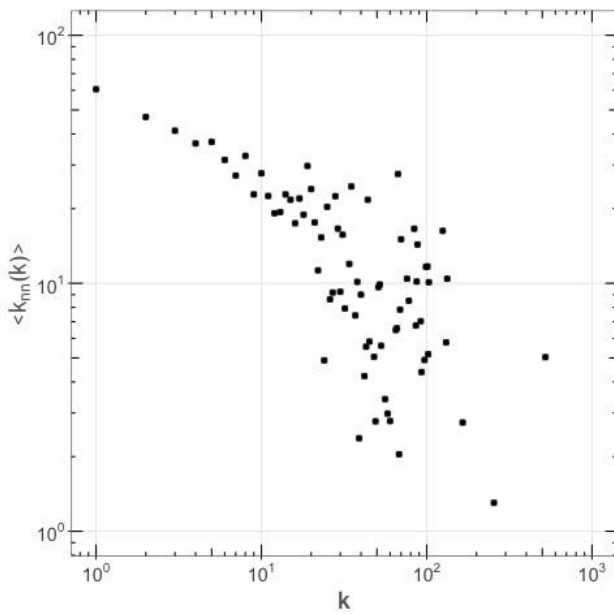


Figure 4. Interconnectedness degree correlation function. We observe that the users are interconnected in such a way that negatively correlated with their own degrees for all degree range.

The degree distribution helps us with recognizing the hub-and-spoke connectivity character, but it fails to convey the complete structure of the network. In particular, it fails to address the question of whether there exist ties among hubs, for which we resort to the degree correlation function (see Figure 4). There is an apparent decrease in the number of neighbors of nodes as their degrees increase. For example, hubs with more several tens of degrees have neighbors with no more than two to three degrees (see the lower right of the plot). The overall network is clearly disassortative with respect to degree (interconnectedness). The value of correlation exponent is -0.53.

As for the assortative mixing by users' metadata, we focus on the role, gender and role-gender of the platform members. Table 2 lists the mixing matrix for the role-gender case where m, f, p, and v stand for male, female, practitioner, and visitor, respectively. For instance, according to the mixing matrix, the fraction of interconnectedness between men practitioners and women

visitors mixing is 0.3332 (0.1666 plus 0.1666), that is about 33 percent of all interconnections.

Table 2. The mixing matrix for role-gender ties in the GC

User Metadata					
Gender	Role	m		f	
		p	v	p	v
m	p	0.0151	0.0493	0.0165	0.1666
	v	0.0493	0.0148	0.0292	0.1491
f	p	0.0165	0.0292	0.0088	0.0471
	v	0.1666	0.1491	0.0471	0.0453

Note that one can obtain the other two mixing matrices (gender-to-gender mixing matrix and role-to-role mixing matrix) from this matrix easily by aggregating the corresponding table elements. Note also that the matrix is symmetric as the network of ties is undirected. The values of the assortativity coefficients with respect to role metadata, gender metadata, and combined role-gender metadata are -0.28, -0.44, and -0.30 respectively.

In summary, it is worth noticing that the Connections network exhibits disassortative correlations (i.e. negative values of correlation coefficients and exponents) with respect to both interconnectedness and user metadata. Cross sectional analysis gives us valuable insights in the structure of interconnectedness for just a single time window of 66 months, lacking the vital information about how the network has evolved. Specifically, was the network disassortative all the time, or were there some fluctuations over time?

4.2. Temporal Analysis of Degree Correlation and User Metadata Correlations

Figure 5 displays proportion of nodes and edges the GC accumulates over time. To investigate the transition from small-sized network fragments to the GC we have observed proportion of nodes and edges of the largest component one week apart for the first three months. What we observe can be described as a fast growth as the percentage of nodes and edges jump from 10% to 50% within only two months.

It is worth noticing that the yet-to-be born giant component (Figure 5) signifies the beginning of the second stage of the network evolution, which involves a spike followed by one year of fluctuations before the giant regime debuts at around the middle of 2010. To put it differently, it has taken one-and-half year for the largest network component to reach its mature state as the giant.

Plot of degree correlation and metadata correlation values illustrate that the largest component of the network starts disassortative and keeps showing disassortative behavior by all kinds of characteristics. Role-gender correlation coefficient is especially notable as it fluctuates in a narrow band between -0.31 and -0.27, which we contend that is an approximately conserved quantity of the network. It is a

remarkable phenomenon to observe that this quantity stays in this narrow band since the beginning of the tie network, much earlier than the emergence of the GC.

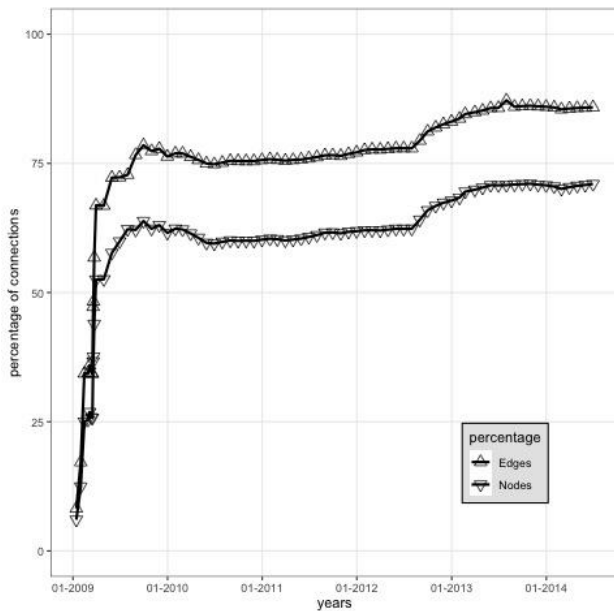


Figure 5. Proportion of nodes and edges accumulated by GC across timeline. We observe that it has taken sixteen months for the Connections service to unite the majority of the users around a shared service of interconnectedness.

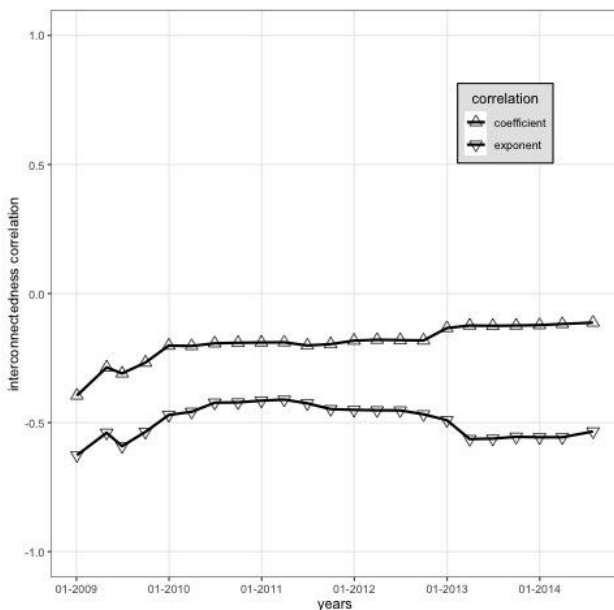


Figure 6. Development of interconnectedness correlation in terms of degree correlation coefficient and correlation exponent indicate that both highly and slightly interconnected users avoid connecting to each other for all time; four trends identified are in parallel with managerial decisions.

We include correlation coefficients in the interconnectedness degree correlation plot just to illustrate that it has always been negative, and note that the

correlation exponent also fluctuates in a rather narrow band between -0.41 and -0.63.

These results are encouraging enough to merit further discussion.

5. DISCUSSION

In the following, we shall discuss our findings in the context of OIP as eBusiness for managerial implications and research venues for open-ended problems.

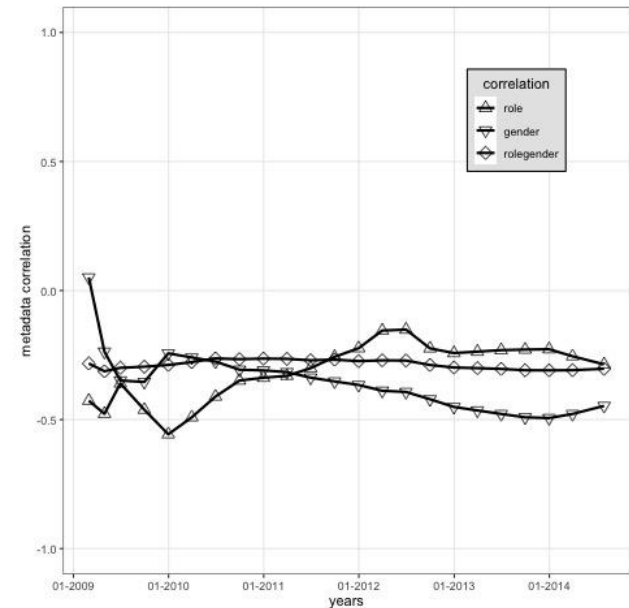


Figure 7. Development of the user metadata correlation. Irrespective of the nodes are of two or three node types according to metadata, we observe that most ties are established between users of opposite metadata for all time.

This research suggests and employs a number of network metrics for better understand assortative characteristics of the network model of OIPs. In particular, interconnectedness and user metadata correlations (i.e., assortativity measures) as network analytics prove to be useful for monitoring sustainability of an e-Business service. Table 3 summarizes the findings and their implications for practitioners regarding monitoring sustainable interconnectedness. While we interpret the findings, we employ the phenomena rooted in network theory [23, 40].

5.1. Degree correlation related implications

According to the theory of network, one should pay attention to both the algebraic sign and the absolute value of the correlation coefficients, and in case the sign of the coefficient is negative, then one should take into account the value of the interconnectedness correlation exponent rather than the value of the interconnectedness coefficient. In the present study, this is certainly the case that the sign is negative all the time (see Fig. 6).

Table 3. Network-based correlations for monitoring sustainable interconnectedness on e-Business

Network Measures	Temporal Analysis for Monthly Time Window	Overall Tendency (66 months)	Importation of the Phenomena rooted in Network Theory	Monitoring Interconnectedness on e-Business
<i>Interconnectedness Correlation Coefficient Value</i>	Increasing function of time	Disassortative by Interconnections	Hub-spoke and scale free [23]	Disassortative characteristic of the network that delays formation of the GC, leads to prolonged service maturity
<i>Interconnectedness Correlation Exponent Value</i>	Fluctuates in a fairly narrow band	Disassortative by Interconnections	Giant Component [33] Assortative Mixing [27, 28]	Timescale of the order of a year for a GC formation may not support a generic startup motto "start small and grow fast" Due to hub-spoke and scale free characteristics, managers should keep watch largely interconnected users closely to maintain interconnectedness
<i>Role Metadata Correlation Value</i>	Large aperiodic fluctuation between -0.10, and -0.55	Disassortative by Role	A metadata based approximately conserved quantity (ACQ) [12]	Managers should take into account both domain insight and metadata to decide and collect some of the metadata that correlates with the service successfully The underlying reason for observing ACQ is the invariance of combined user metadata correlation signifies the fact that the interconnectedness has a self-sustaining nature
<i>Gender Metadata Correlation Value</i>	Medium aperiodic fluctuation between -0.25, and -0.50	Disassortative by Gender		
<i>RoleGender Metadata Correlation Value</i>	Approximately constant at around -0.30 even before the giant component formation	Disassortative by RoleGender		

A close examination of coefficient values and exponents (Fig. 6) shows

1. From the debut of the Connections service until the formation of GC at around mid 2010: Values of exponent and coefficient mimic each other with an overall increasing trend.
2. From mid 2010 to mid 2012: Stable in a quite narrow band.
3. From mid 2012 to mid 2013: Sudden change of hearts in both values, the coefficient starts increasing while the exponent drops to values that are more negative.
4. From mid 2013 to mid 2014: Stable in a quite narrow band

Believing that the third interval begs additional interpretation we made several meetings with both the CEO and the system administrators. Eventually we learnt that two different decisions took place around the time. First, the CEO decided to shut down the Connections service (between May and June 2012) during a migration of the information system. About a year later, in October 2013, they decided to update the Web site with a new GUI. The makeover was a complete visual overhaul of the website with the decision to downgrade the Connections service based on conventional web analytics [41]. We ascribe the sudden change in the interconnectedness correlation values to these decisions.

Additional insights can be discerned by the evolution of the interconnectedness exponent. In contrast to the corresponding interconnectedness coefficient, which is an increasing function of time (Fig. 6), the value of the exponent fluctuates in a narrow band. The underlying

reason for these all-time negative values is the hub-spoke interconnectedness (i.e., minimized hub-to-hub interconnectedness and spoke-to-spoke interconnectedness). This suggests that the majority number of highly interconnected users prefer to link themselves to slightly interconnected users. In other words, both highly and slightly interconnected users avoid connecting to each other.

In the eBusiness context, this might be of interest to platform managers to see that the natural occurrence of interconnectedness is in favor of hub-spoke; that is, highly interconnected users-slightly connected users. This prevailing choice results in formation of the GC with the cost of certain prolonged time. If there is no intervention on the system (i.e., *ceteris paribus*), the findings corroborate that disassortative tendency will continue in the near future.

This can be a crucial decision point for OIP eBusiness startups whose motto is often "start small and grow fast" [42]. That is, the timescale of the business growth is of primary importance. The case at hand suggests just the opposite of the motto because the timescale of the GC indicates a slow growth because of disassortative tendency in terms of interconnectedness. Disassortative characteristic of the network delays the formation of the GC. In the present study, indeed it took about 16 months for formation of the GC.

Knowing this fact, the startup owner should either immediately quit such a service or consider the following options. In case the decision made to encourage or discourage whom-to-whom interactions from interconnectedness point of view, the following

prescriptions will be consistent with the assortative mixing phenomenon [27, 28].

- Promotion driven connectedness for disassortativity. Promote highly interconnected user-slightly interconnected user interactions. If this choice is implemented and the strategy works, as the network gets more disassortative, the service sustainability will be achieved by slow growth of the GC. Thus, managers should be prepared for this timeline and make strategic and operational plans accordingly.
- Promotion driven connectedness for assortativity. Promote highly interconnected user-highly interconnected user interactions or slightly interconnected user-slightly connected user interactions. If this choice is implemented, as the network gets assortative it will take less time to reach giant formation. However, misalignment of service promotions with users' selective behavior can bring unprecedented situations. The side effect would be that this aggressive or unexpected tactic might result in unfavorable result for service sustainability.

Overall, one can suggest that the CEO could have utilized these network metrics to assess the impact of managerial decisions on interconnectedness.

5.2 User metadata correlations related implications

The above considerations do not take into account user metadata correlations. A sheer consideration of interconnectedness without incorporating metadata correlations, however, would yield a partial picture of the mixing preferences of the users.

Regarding the network wide correlations of user metadata, we observe that tying with respect to role and gender is disassortative. That is, mixing preferences of platform users correlate with opposite metadata (individual attributes). Specifically, role metadata correlations exhibit large aperiodic fluctuations between -0.10 and -0.55 whereas gender metadata correlations show relatively medium aperiodic fluctuations between -0.25, and -0.50 after a very brief initial period of positive correlations (Fig. 7). Contrary to these two aperiodic correlations between metadata and interconnectedness, the combined metadata correlations of role-gender and interconnectedness leads to the "metadata based approximately conserved quantity (ACQ)" phenomenon [12] where the correlation coefficient starts near -0.27 and shows an approximately constant trend around the value of -0.30. Remarkably, the trend reveals itself much before the formation of the GC, which first emerges in the middle of 2010. The combined metadata correlation signifies the fact that the interconnectedness has a self-sustaining nature, which assumes a steady state value even before the GC formation and retains it as the GC evolves.

We believe that the ACQ phenomenon [12] is a signature for interconnectedness. ACQs are not new to theoretical complex systems literature as [43, 44] confirm their existence algebraically. They have been also emphasized in a real-world complex system [12, 45]. However, it seems that, they have not been paid the attention they deserve, yet. We believe that the underlying reason for the ACQ should, somehow, be related to assortative mixing by node metadata as the role-gender combined metadata correlation exhibits nearly a constant value at around 0.30 over the whole time window. This is a call for research to articulate this phenomenon further as this observation is purely accidental begging for a theoretical explanation. On the other hand, this would be of utmost importance to managers.

We have already discussed (Fig. 6) along with four-time windows above that interconnectedness correlations should reflect managerial decisions such as a major overhaul of GUI design. Similarly, some correlations should remain invariant so long as no data related actions (e.g., service rules) on the platform are realized. Examples of such interventions could be making the service paid or unpaid, disclosing or hiding metadata among users, mandatory collection of some metadata.

In the present case, knowing the fact that the interconnectedness rules (i.e., the rules of establishing ties) have not been changed, we expect that some metadata correlations or combined metadata correlations should reflect it. Indeed, the combined metadata correlation of gender and role remains invariant, which can be ascribed to the unchanged rules.

According to the theory of network, the exact relation between data (interconnectedness of users) and user metadata is an open issue and no conjecture has been made as to which metadata is to be collected that reflects nature of data [46]. Thus, we cannot prescribe which metadata are relevant to a particular business model and therefore need to be collected, yet we can suggest the way to assess relevance of metadata to the data that the business generates in terms of correlations. In the present case, we believe that the rich domain insight of the CEO helped their business to collect some of the metadata that correlate with the service successfully.

If the observed mixing values are the result of some underlying causes of user metadata correlations, then to the mixing matrix of observed ties there corresponds a mixing matrix of potential ties. Thanks to the rareness of some of the observed mixing values (e.g. practitioner-to-practitioner ties occur seldom, of the order of per mill), the chances of establishing their first ties for newcomers will be quite low. This is especially crucial to the growth of the service. The CEO should be aware of this fact and act accordingly. The manager may consider the addition of a platform feature that helps the newcomers with establishing their first ties based on the mixing matrix of observed ties based on their metadata.

Accordingly, one can consider the following suggestions to platform managers: do not promote connectivity among the same type of metadata, like patient-to-patient or physician-to-physician, or male patient-to-male patient connectivity, instead promote connectivity between physician-to-patient, female patient-to-male patient.

6. CONCLUSION

Online interactive platforms (OIPs) including social networking applications, community support applications provide new opportunities and challenges for e-Business practitioners and interested academics. At the hearth of most these OIPs is the Connection services providing user interconnectedness that can best described as a network of complex interactions. Growth and sustainability of interconnectedness are essential to achieve successful e-Business. One of the key issues is to better describe, prescribe and eventually predict sustainable interconnectedness for OIPs.

It has been suggested by [12] that from the network theory point of view, sustainable interconnectedness on OIPs is strongly associated with the emergence of the GC as well as existence of metadata based ACQs. The challenge is then to explore underlying reasons behind ACQs so that we can better describe, prescribe, and predict sustainable interconnectedness on OIPs.

In this research, to better describe sustainable interconnectedness we conduct both static and temporal network analysis for an OIP. Regarding the static analysis, we explore the structural properties of the GC in terms of the extent of users' interconnectedness, interconnectedness correlation among users, and metadata correlations. The extent of user interconnectedness shows us the scale free nature of interconnectedness as well as the existence of hubs. Interconnectedness correlation among users and metadata correlations both display disassortative nature of the Connections service. That is, in general both hubs and the same metadata type (gender, role) avoid connecting each other.

Regarding the temporal analysis, interconnectedness coefficient and exponent show a disassortative interconnectedness trend over time. Theory of network confirms the fact that this disassortative characteristic of the network delays formation of the GC, leading to prolonged service maturity. Furthermore, one can foresee that this trend will continue. In line with this prediction, we further provide managers with growth related suggestions for especially existing users and their connections.

We found that the underlying reason for observing ACQ is invariance of combined user metadata correlation, which is derived from the mixing matrix of observed ties. This signifies the fact that the interconnectedness has a self-sustaining nature. This is especially crucial to the growth of the service. The CEO should be aware of this fact and act accordingly. The manager may consider the addition of

a platform feature that helps the newcomers with establishing their first ties based on the mixing matrix of observed ties based on their metadata.

We agree to the call by [47] that network science will be one of the main contributors to social media analytics [48] in the near future, and the work is underway to propose generic performance analytics for OIPs that offer interconnectedness services.

REFERENCES

- [1] R. Benbunan-Fich, M. Koufaris, "An Empirical Examination of the Sustainability of Social Bookmarking Websites", *Information Systems and e-Business Management*, 8(2), 131–148, 2010.
- [2] D. K. Liou, W. H. Chih, L. C. Hsu, C. Y. Huang, "Investigating Information Sharing Behavior: The Mediating Roles of the Desire to Share Information in Virtual Communities", *Information Systems and e-Business Management*, 14(2), 187–216, 2016.
- [3] B. K. Samanthula, W. Jiang, "Interest-driven Private Friend Recommendation", *Knowledge and Information Systems*, 42(3), 663–687, 2015.
- [4] M. Chau, J. Xu, "Business Intelligence in Blogs: Understanding Consumer Interactions and Communities", *MIS Quarterly*, 36(4), 1189–1216, 2012.
- [5] S. Kisilevich, C. S. Ang, M. Last, "Large-scale analysis of self-disclosure patterns among online social networks users: a Russian context", *Knowledge and information systems*, 32(3), 609–628, 2012.
- [6] M. K. Foster, A. Francescucci, B. C. West, "Why Users Participate in Online Social Networks", *International Journal of e-Business Management*, 4(1), 3–19, 2010.
- [7] C. Ridings, M. Wasko, "Online Discussion Group Sustainability: Investigating the Interplay Between Structural Dynamics and Social Dynamics Over Time", *Journal of the Association for Information Systems*, 11(2), 95–121, 2010.
- [8] A. Iriberry, G. Leroy, "A Life-cycle Perspective on Online Community Success", *ACM Computing Surveys (CSUR)*, 41(2), 11, 1–29, 2009.
- [9] H. Chen, R. H. Chiang, V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact", *MIS Quarterly*, 36(4), 1165–1188, 2012.
- [10] T. W. Wang, Y. Verbitskiy, W. Yeoh, "Depicting Data Quality Issues in Business Intelligence Environment through a Metadata Framework", *International Journal of Business Intelligence Research (IJBIR)*, 7(2), 20–31, 2016.
- [11] D. J. Watts, S. H. Strogatz, "Collective Dynamics of 'small-world' Networks", *Nature*, 393(6684), 440–442, 1998.
- [12] M. N. Aydin, N. Z. Perdahci, "Dynamic Network Analysis of Online Interactive Platform", *Inf Syst Front*, 21(2), 229–240, 2019.
- [13] G. F. Khan, "Social Media Analytics", **Social Media for Government**, Springer, Singapore, https://doi.org/10.1007/978-981-10-2942-4_6, 2017.

- [14] F. D. Malliaros, V. Megalooikonomou, C. Faloutsos, “Estimating Robustness in Large Social Graphs”, *Knowledge and Information Systems*, 45(3), 645–678, 2015.
- [15] J. Cao, K. A. Basoglu, H. Sheng, P. B. Lowry, “A Systematic Review of Social Networking Research in Information Systems”, *Communications of the Association for Information Systems*, 36(37), 727–758, 2015.
- [16] A. Vespignani, “Twenty Years of Network Science”, *Nature*, 558, 528–529, 2018.
- [17] S. Milgram, “The Small World Problem”, *Psychology Today*, 2(1), 60–67, 1967.
- [18] S. Wasserman, **Social Network Analysis: Methods and Applications**, Vol. 8, Cambridge University press, 1994.
- [19] L. Freeman, **The Development of Social Network Analysis. A Study in the Sociology of Science**, BookSurge, LLC North Charleston, A.B.D., 2004.
- [20] S. P. Borgatti, D. S. Halgin, “On Network Theory”, *Organization Science*, 22(5), 1168–1181, 2011.
- [21] A. L. Barabási, **Network Science**, Cambridge University Press, Cambridge, B.K., 2016.
- [22] T. van Mierlo, “The 1% rule in four digital health social networks: An observational study”, *Journal of medical Internet research*, 16(2), 2014
- [23] A. L. Barabási, R. Albert, “Emergence of Scaling in Random Networks”, *Science*, 286(5439), 509–512, 1999.
- [24] P. Erdős, A. Rényi, “On the Evolution of Random Graphs”, *Publ. Math. Inst. Hung. Acad. Sci.*, 5(1), 17–60, 1960.
- [25] M., McPherson, L. Smith-Lovin, J. M. Cook, “Birds of a Feather: Homophily in Social Networks”, *Annual Review of Sociology*, 27(1), 415–444, 2001.
- [26] K. Zhang, D. Lo, E. P. Lim, P. K. Prasetyo, “Mining Indirect Antagonistic Communities from Social Interactions”, *Knowledge and Information Systems*, 35(3), 553–583, 2013.
- [27] M. E. Newman, “Assortative Mixing in Networks”, *Physical Review Letters*, 89(20), 208701-1–208701-4, 2002.
- [28] M. E. Newman, “Mixing Patterns in Networks”, *Physical Review E*, 67(2), 026126-1–026126-13, 2003.
- [29] A. L. Barabási, “Scale-free Networks: A Decade and Beyond”, *Science*, 325(5939), 412–413, 2009.
- [30] M. N. Aydin, N. Z. Perdahçı, “Network Analysis of an Interactive Health Network”, *Journal of Internet Social Networking & Virtual Communities*, 2016(2016), 1–17, 2016.
- [31] J. Leskovec, J. Kleinberg, C. Faloutsos, “Graph Evolution: Densification and Shrinking Diameters”, **ACM Transactions on Knowledge Discovery from Data (TKDD)**, Cilt: 1, Editör: Jiawei Han, ACM New York, NY, A.B.D., 1–41, 2007.
- [32] G. T. Bosslet, A. M. Torke, S. E. Hickman, C. L. Terry, P. R. Helft, “The Patient–doctor Relationship and Online Social Networks: Results of a National Survey”, *Journal of General Internal Medicine*, 26(10), 1168–1174, 2011.
- [33] R. Kumar, J. Novak, A. Tomkins, “Structure and Evolution of Online Social Networks.”, **Link mining: models, algorithms, and applications**, Cilt: 1, Editör: P. Yu, J. Han, C. Faloutsos, Springer, New York, NY, A.B.D., 337–357, 2010.
- [34] G. Ünel, “Information Passing in Healthcare Social Networks”, *Bilişim Teknolojileri Dergisi*, 10(1), 99–104., 2017
- [35] J. Howison, A. Wiggins, K. Crowston, “Validity Issues in the Use of Social Network Analysis with Digital Trace Data”, *Journal of the Association for Information Systems*, 12(12), 767–797, 2011.
- [36] G. Csardi, T. Nepusz, “The iGraph Software Package for Complex Network Research”, *Inter Journal, Complex Systems*, 1695(5), 1–9, 2006
- [37] D. A. Schult, P. Swart, “Exploring Network Structure, Dynamics, and Function Using NetworkX”, **The 7th Python in Science Conferences (SciPy 2008)**, Pasadena, Kaliforniya, 11–16, 19-24 Ağustos, 2008.
- [38] X. Chen, C.-Z. Yang, “Visualization of Social Networks”, **Handbook of Social Network Technologies and Applications**, Cilt 1, Editör: Furht, B., Springer, New York, NY, A.B.D., 585–610, 2010.
- [39] T. M. Fruchterman, E. M. Reingold, “Graph Drawing by Force-Directed Placement.”, *Software: Practice and experience*, 21(11), 1129–1164, 1991
- [40] G. C. Kane, M. Alavi, G. Labianca, S. P. Borgatti, “What’s Different About Social Media Networks? A Framework and Research Agenda”, *MIS Quarterly*, 38(1), 275–304, 2014
- [41] V. Ö. Budak, E. Kartal, S. Gülseçen, “Site-içi Aramalar ve Apriori Algoritması Kullanılarak Web Sitesi Ziyaretçilerinin İhtiyaç Tespitine Yönelik Bir Örnek Olay İncelemesi”, *Bilişim Teknolojileri Dergisi*, 11(2), 211-222, 2018.
- [42] S. Koushik, J. Birkinshaw, S. Crainer, “Using Web 2.0 to Create Management 2.0”, *Business Strategy Review*, 20(2), 20–23, 2009
- [43] İ. Birol, A. Hacınlıyan, “Approximately Conserved Quantity in the Hénon-Heiles Problem”, *Physical Review E*, 52(5), 4750–4753, 1995
- [44] N. Z., Perdahçı, A. Hacınlıyan, “Normal Forms and Nonlocal Chaotic Behavior in Sprott Systems”, *International Journal of Engineering Science*, 41(10), 1085-1108, 2003
- [45] H. B. Hu, X. F. Wang, “Disassortative Mixing in Online Social Networks”, *EPL (Europhysics Letters)*, 86(1), 18003–18009, 2009
- [46] L. Peel, D. B. Larremore, A. Clauset, “The Ground Truth About Metadata and Community Detection in Networks”, *Science Advances*, 3(5), e1602548, 1–8, 2017
- [47] S. Stieglitz, L. Dang-Xuan, A. Bruns, C. Neuberger, “Social Media Analytics”, *Bus Inf Syst Eng*, 6(2), 89–96, 2014
- [48] B. Karaöz, U. T. Gürsoy, “Adaptif Öğrenme Sözlüğü Temelli Duygu Analiz Algoritması Önerisi”, *Bilişim Teknolojileri Dergisi*, 11(3), 245–253, 2018.