AKADEMİK BAKIŞ DERGİSİ
Sayı: 59 Ocak - Şubat 2017
Uluslararası Hakemli Sosyal Bilimler E-Dergisi
ISSN:1694-528X İktisat ve Girişimcilik Üniversitesi, Türk Dünyası Kırgız – Türk Sosyal Bilimler Enstitüsü, Celalabat – KIRGIZİSTAN
http://www.akademikbakis.org

TESTING TWO HYPOTHESES FOR THE FRACTIONS OF DRIVER NODES

Necmi GÜRSAKAL* **Selim TÜZÜNTÜRK****

Abstract
Networks are used to represent complex systems in the real world. Recently, the focus of interest in the area of network science has shifted to the controllability of complex networks. In this context, the new concept of a subset of nodes called "driver nodes" is becoming pronounced in the network world. Driver nodes belong to the intersection of network science and the control theory of engineering. The growing interest in this field has resulted in an opportunity for scientists to explain how to control the dynamics of complex systems. The scope of this study is to directly test (1) the fractions of driver nodes' distributions of real networks and fully randomized networks and (2) the statistically significant difference in the mean fractions of driver nodes between natural and manmade networks and plus between natural and fully randomized networks. On the basis of the sample results, it is found that whereas real networks follow a largest extreme value distribution, fully randomized networks follow a gamma distribution. In addition, whereas a statistically significant difference was found between natural and manmade networks, no difference was found between natural and fully randomized networks.

Keywords: Networks, Complex Systems, Driver Nodes, Controllability, Critical Nodes.

SÜRÜCÜ DÜĞÜMLERİN ORANLARI İÇİN İKİ HİPOTEZİN TEST EDİLMESİ

Öz
Ağlar gerçek dünyadaki karmaşık sistemlerin temsil edilmesinde kullanılır. Yakın dönemde ağ bilimi alanındaki ilgi odağı karmaşık ağların kontrol edilebilirliğine yöneldi. Bu çerçevede, ağ dünyasında yeni bir kavram düğümlerin alt kümesi olarak adlandırılan "sürücü düğümler" telaffuz edilmeye başlandı. Sürücü düğümler ağ bilimi ve mühendisliğin kontrol teorisi ile ilgilidir. Bu alana artan ilgi bilim adamlarına karmaşık dinamik sistemlerin nasıl kontrol edileceğinin izah edilmesine olanak tanıdı. Bu çalışmanın kapsamı (1) gerçek ağların ve tam rassallaşmış ağların sürücü düğümlerinin oranlarının dağılımlarının ve (2) doğal ve insan yapımı ağlar arasında ve ayrıca doğal ve tam rassallaşmış ağlar arasında sürücü düğümlerin ortalama oranlarının istatistiksel anlamlı farklılığının doğrudan test edilmesidir. Örneklem sonuçlarına göre, gerçek ağlar en büyük ekstrem değer dağılımı izlerken, tam rassallaşmış ağlar bir gama dağılımı izlemektedir. Ayrıca, doğal ve insan yapısı ağlar arasında istatistiksel bir anlamlı farklılık bulunurken, doğal ve tam rassallaşmış ağlar arasında bir fark bulunmamıştır.

Anahtar Kelimeler: Ağlar, Karmaşık Sistemler, Sürücü Düğümler, Kontrol Edilebilirlik, Kritik Düğümler.

1. Introduction

Networks of interconnected nodes are ubiquitous and lie at the heart of complex systems. Natural and manmade systems can be analyzed using networks and graphs. Networks are used to represent many nonlinear complex systems in the real world (Asgari et al., 2013: 1). They are useful for bringing scientists a holistic view of "everything depends on and interacts on everything". Concepts and tools of network science are continuously developing in the study of complex networks.

In 1736, Euler focused on networks and laid the foundations of graph theory. Thereafter, the contributions of Paul Erdős and Alfred Rényi in 1960, Mark Granovetter in

* Prof. Dr., Uludağ Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, negursakal@gmail.com
** Yrd. Doç. Dr., Uludağ Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, selimtuzunturk@uludag.edu.tr

1973, Duncan Watts in 1998, Albert-László Barabási and Reka Albert in 1999 and Jure Leskovec in 2000s, along with those of other network scientists, gave rise to the development of this field.

As is widely known, the main purpose of science is to understand the relationships between variables and to estimate and/or control them. Having the understanding and ability required to control systems are the main objectives of science. Network science is data-driven and computational in nature.

The 1998-2010 periods can be described as a **descriptive phase** in network science. After 2010, network science left the descriptive phase and moved to new phase. Now in a **control phase**, network scientists focus particularly on how to control complex systems. Therefore, the controllability of massive and complex networks such as biological networks, physical networks and social networks is of primary and vital importance for scientists and human beings. Through these developments, the **driver node** concept has emerged.

In light of these developments, it is observed in related literature that some scientists are trying to form a bridge between network science and the control theory of engineering. Recent opinions of scientists are shown as follows. Trafton (2011) expressed the following opinions about driver nodes:

> *"At first glance, a diagram of the complex network of genes that regulate cellular metabolism might seem hopelessly complex, and efforts to control such a system futile. However, an MIT researcher has come up with a new computational model that can analyze any type of complex network - biological, social or electronic- and reveal the* ***critical points*** *that can be used to control the entire system. These critical points are called* ***driver nodes****."*

Liu, Slotine and Barabasi (2011: 167) defined driver nodes as follows:

> *"If we wish to* ***control a system****, we first need to identify the set of nodes that, if driven by different signals, can offer full control over the network. We will call these* ***driver nodes****."*

Liu, Slotine and Barabasi (2011: 171) and Liu (2014: 342) underlined that sparse and heterogeneous networks are harder to **control** than dense and homogeneous networks.

Banerjee and Roy (2012: 1) argued that an effective understanding of controllability in directed networks can be reached using distance-based measures of closeness centrality (CC) and betweenness centrality (BC) and may not require the knowledge of local connectivity measures such as in-degree and out-degree.

Yazıcıoğlu and Egerstedt (2013: 3802) devised a topological leader selection scheme and a network assembly scheme, both of which achieved **complete controllability**.

To control a system, we need to control a fixed fraction of nodes. These developments have presented interesting results in brain research. Kumar et al. (2013: 585) found the following results for brain networks:

> *"We have shown knowing the network connectivity of neurons and neuronal populations can help in choosing the most appropriate network*

AKADEMİK BAKIŞ DERGİSİ
Sayı: 59 Ocak - Şubat 2017
Uluslararası Hakemli Sosyal Bilimler E-Dergisi
ISSN:1694-528X İktisat ve Girişimcilik Üniversitesi, Türk Dünyası Kırgız – Türk Sosyal Bilimler Enstitüsü, Celalabat – KIRGIZİSTAN
http://www.akademikbakis.org

> *node(s) for activity modulation to help understand the function and dynamics of networks in the brain.”*

Yuan et al. (2013: 3) proposed that in many real situations, it is not possible to have exact link weights. Therefore, some researchers think that it is reasonable to assume random-weight distributions rather than identical weights for real networks (Yuan et al., 2013: 3). However, Sabattini, Secchi and Fantuzzi (2014:1841) found a result that makes this approach arguable. They found that introducing random edge weights always ensures the controllability of the networked system.

Jia and Barabasi (2013:1-2) classified nodes as critical, redundant and intermittent and defined **critical nodes** as follows:

> *“Nodes that always need to* ***control*** *and a node is* ***critical*** *if and only if it has no incoming links”.*

Jia and Barabasi (2013:2) found that although the number of driver nodes sufficient and necessary for control (N_D) is primarily fixed by the network’s degree distribution, there are multiple minimum driver node sets[1] (MDS) with the same N_D that can maintain control. Jia and Barabasi (2013:4) defined the concept of **control capacity** as “a centrality measure quantifying a node's likelihood of being a driver node”. Examining the role of individual nodes in controlling a network requires scientists to understand control capacity.

Menichetti, Dall’Asta and Bianconi (2014:1) stated that the relation between the controllability of a network and its structure has recently started to be investigated. As a result of these investigations, driver nodes are considered the change agents in complex networks.

According to the definition by Zhang et al. (2014: 1), a system is called **controllable** if it can be driven from any initial state to any desired state in a finite time. These researchers stated that by using **driver nodes**, it may be possible for us to reach any desired state in a finite time.

Banerjee et al. (2014) explored driver nodes such as people. The researchers simply studied **diffusion centrality** by tracking sources of gossip in a community, determining which people were most central in a network. **Driver nodes** are expected to open ways to select leaders scientifically.

The scopes of networks to be controlled have been changing in recent studies. Sometimes, the studies aim to control the entire network, and sometimes, the control aim is narrow. Gao et al. (2014:1) described this change as follows:

> *“It is typically neither feasible nor necessary to control the entire network, prompting us to explore target control: the efficient control of a preselected subset of nodes.”*

Compared to entire control, target control is found to be more practical. It localizes the control problem. Gao et al. (2014:1) expressed their opinions as follows:

[1] Minimum driver node set represents the smallest set of nodes through which we can yield control over the whole system.

AKADEMİK BAKIŞ DERGİSİ
Sayı: 59 Ocak - Şubat 2017
Uluslararası Hakemli Sosyal Bilimler E-Dergisi
ISSN:1694-528X İktisat ve Girişimcilik Üniversitesi, Türk Dünyası Kırgız – Türk Sosyal Bilimler Enstitüsü, Celalabat – KIRGIZİSTAN
http://www.akademikbakis.org

> *".....for directed tree networks, one node can control a set of target nodes if the path length to each target node is unique and degree heterogeneous networks are target controllable with higher efficiency."*

Bayer's website indicates that understanding biochemical processes in the body may give scientists valuable clues as to how a disease can be cured. The following opinions were found on Bayer's website entitled, "Bayer: Science for a better life":

> *"Target drugs require highly understanding of biochemical networks. The signaling cascades involve proteins can be potential sites of action for drugs. Drugs either switch proteins off or enhance their function. However, only few protein molecules are suitable as targets for drugs. It is a difficult and complex task to detect them among the countless proteins that are produced by the body. Also number of targets has been increased from one target to multitarget."*

Depending on the scientists' various points of view summarized above, the objective of this study is to attract the attention of readers to driver nodes, particularly to the fractions of driver nodes. This study aimed to examine their calculation operations and probability distributions and similarities in natural and manmade networks. To do so, Section 2 addresses the question "How can fractions of driver nodes are calculated theoretically?" Section 3 covers the applications: At first, an attempt was made to determine the probability distribution of the fraction of the driver nodes. Then, statistical hypothesis tests were performed to determine whether there was a significant difference between fractions of driver nodes in natural and manmade networks. The last section covers the conclusion.

2. Calculation of fractions of driver nodes in networks

In the following paragraph, Zhang et al. (2014: 7) briefly summarize how to calculate minimum driver nodes:

> *"The minimal driver nodes set (MDS) can be obtained by finding the maximal matching of network. However, the MDS's of a network are not unique, and have very different topological features existed. Thus, one important research direction in the controllability of complex networks is analyzing the topological features of all of the possible MDS's."*

The main problem here is that minimal driver nodes set to control a network are not unique. Therefore, it is important to characterize these driver nodes and select the right driver nodes (Mahia, Fulwani and Singh, 2014: 1). A pseudocode of a driver node algorithm and igraph maximum.bipartite.matching codes can be found, respectively, in Khazanchi (2014) and on Inside-R web link (see references).

Complete information regarding the weights of the links of a network is needed. Thus, Rathore et al. (2015: 2) explain that the concept of **structural controllability** is used to overcome this limitation of weights. A relatively efficient algorithm called the **Maximum Matching Algorithm** is used to compute the driver nodes. Driver nodes can be identified

AKADEMİK BAKIŞ DERGİSİ
Sayı: 59 Ocak - Şubat 2017
Uluslararası Hakemli Sosyal Bilimler E-Dergisi
ISSN:1694-528X İktisat ve Girişimcilik Üniversitesi, Türk Dünyası Kırgız – Türk Sosyal Bilimler Enstitüsü, Celalabat – KIRGIZİSTAN
http://www.akademikbakis.org

only for directed graphs. First, a directed graph is converted to a bipartite graph, and second, maximum matching is found in the converted bipartite graph.

Structural network controllability is a new field that guides a system's behavior towards a desired state through the appropriate management of input variables (Asgari et al., 2013:10). The relations among structural network controllability, topological parameters and network medicine (metabolic drug targets) have been studied many times in the last 5 years. Researchers are studying the relationship between the topological features and the functions of biological and social networks. Within this context, many papers have been written on the controllability of networks.

Liu, Slotine and Barabasi (2011: 167) developed analytical tools to study the controllability of arbitrary complex networks. In a directed network using a subset of nodes called **driver nodes**, the researchers explained how to control the dynamics of the system. They explained that degree distribution is the primary concept to determine the number of driver nodes. Dense networks can be controlled using a few driver nodes, but sparse networks require a high number of driver nodes. Additionally, in both types of networks (dense or sparse), driver nodes tend to avoid hubs. The researchers also introduced the **control centrality** concept.

How can a set of driver nodes be found in directed networks? It requires structural controllability and matching. A linear control system (A, B) is structurally controllable (Wang, Gao and Gao, 2012: 6) if and only if the structured matrix [A; B] is irreducible and has generic rank N. For an undirected graph, a matching M is an independent edge set without common vertices. For a directed graph, if no two edges in M share a common starting or ending vertex, then M is a matching, and M has a maximum when it contains the maximum number of edges (Nikolopoulos and Palios, 2005:69). Maximum matching is used to determine the minimum set of driver nodes (N_D) for the networks:

$$\frac{dx(t)}{dt} = Ax\ \ t\ \ + Bu(t) \qquad (1)$$

Vector x (t) represents the states of N nodes in times t:

$$x\ \ t\ \ = (x_1\ \ t\ \ , \dots \dots \dots \dots . x_N\ \ t\ \ )^T \qquad (2)$$

For example $x_i(t)$ can be the traffic flow of node "i" in a communication network.

The system is controlled using the time-dependent input vector $u(t)=(u_1(t),\dots,u_M(t))^T$ imposed by the controller, where in general, the same signal $u_i(t)$ can drive multiple nodes.

The mathematical condition for controllability can be explained as follows: The system described by equation (2) is said to be controllable if it can be driven from any initial state to any desired final state in finite time, which is possible if and only if the NxNM controllability matrix has full rank. Liu, Slotine and Barabasi (2011: 167) state:

$$C= (B, AB, A^2B,\dots, A^{N-1}B) \qquad (3)$$

$$\mathrm{rank}(C)=N \qquad (4)$$

AKADEMİK BAKIŞ DERGİSİ
Sayı: 59 Ocak - Şubat 2017
Uluslararası Hakemli Sosyal Bilimler E-Dergisi
ISSN:1694-528X İktisat ve Girişimcilik Üniversitesi, Türk Dünyası Kırgız – Türk Sosyal Bilimler Enstitüsü, Celalabat – KIRGIZİSTAN
http://www.akademikbakis.org

a_{ij} weights of our adjacency matrix must be known for the applications of these equations.

As time has gone by, other researchers have studied and explored the relationships among structural network controllability, topological parameters, and network medicine (metabolic drug targets). The empirical results are surprising, and sometimes, the found nodes are not the expected ones. A major result of Delpini et al.'s (2013: 5) analysis is that the banks that are more relevant to the overall state of the credit network are neither the most connected lenders nor the top ones.

3. Testing two hypotheses for the fractions of driver nodes

To test the two hypotheses—(1) the fractions of driver nodes' distributions of real networks and fully randomized networks and (2) the statistically significant difference of mean fractions of driver nodes between natural and manmade networks and between natural and fully randomized networks—the data below in Table 1, taken from Liu, Slotine and Barabasi (2011: 169), were used.

Table 1. The characteristics of the real networks

Type	Name	N	L	(FDN) $n_D \equiv N_D/N$	FDNER	Density	Mean Degree
Regulatory	TRN-Yeast-1 (N)	4441	12873	0.965	0.083000	0.000652	5.797
	TRN-Yeast-2 (N)	688	1079	0.821	0.303000	0.002282	3.136
	TRN-EC-1 (N)	1550	3340	0.891	0.188000	0.001391	4.309
	TRN-EC-2 (N)	418	519	0.751	0.380000	0.002977	2.483
	Ownership-USCor(M)	7253	6726	0.820	0.480000	0.000127	1.854
Trust	College student (M)	32	96	0.188	0.082000	0.096770	6.000
	Prison inmate (M)	67	182	0.134	0.103000	0.041157	5.432
	Slashdot(M)	82168	948464	0.045	0.000017	0.000140	23.086
	WikiVote (M)	7115	103689	0.666	0.000140	0.002048	29.146
	Epinions(M)	75888	508837	0.549	0.001000	0.000088	13.410
Food web	Ythan(N)	135	601	0.511	0.016000	0.033222	8.903
	Little Rock(N)	183	2494	0.541	0.0050	0.0744	27.25

					00	72	6
	Grassland(N)	88	137	0.523	0.301000	0.017894	3.113
	Seagrass(N)	49	226	0.265	0.203000	0.096088	9.224
Power grid	Texas(M)	4889	5855	0.325	0.396000	0.000245	2.395
Metabolic	Escherichia coli(N)	2275	5763	0.382	0.129000	0.001113	5.066
	Saccharomyces cerevisae(N)	1511	3833	0.329	0.130000	0.001679	5.073
	Caenorhabditis elegans(N)	1173	2864	0.302	0.144000	0.002083	4.883
Electronic circuits	s838(M)	512	819	0.232	0.293000	0.003130	3.199
	s420(M)	252	399	0.234	0.298000	0.006308	3.166
	s208(M)	122	189	0.238	0.301000	0.012603	3.098
Neural	Caenorhabditis elegans(N)	297	2345	0.165	0.003000	0.026674	15.791
Citation	ArXiv-HepTh(M)	27770	352807	0.216	0.000036	0.000457	25.409
	ArXiv-HepPh(M)	34546	421578	0.232	0.000030	0.000353	24.406
World Wide Web	nd.edu(M)	325729	1497134	0.677	0.012000	0.000014	9.192
	stanford.edu(M)	281903	2312497	0.317	0.000300	0.000029	16.406
	Political Blogs(M)	1224	19025	0.356	0.000800	0.012709	31.086
Internet	p2p-1(M)	10876	39994	0.552	0.001000	0.000338	7.354
	p2p-2(M)	8846	31839	0.578	0.002000	0.000406	7.198
	p2p-3(M)	8717	31525	0.577	0.002000	0.000414	7.232
Social communication	UClonline(M)	1899	20296	0.323	0.706000	0.005631	21.375
	Email-epoch(M)	3188	39256	0.426	0.000300	0.003863	24.627
	Cellphone(M)	36595	91826	0.204	0.133000	0.000068	5.018
Intra-organizational	Freemans-2(M)	34	830	0.029	0.029000	0.739750	48.823

AKADEMİK BAKIŞ DERGİSİ
Sayı: 59 Ocak - Şubat 2017
Uluslararası Hakemli Sosyal Bilimler E-Dergisi
ISSN:1694-528X İktisat ve Girişimcilik Üniversitesi, Türk Dünyası Kırgız – Türk Sosyal Bilimler Enstitüsü, Celalabat – KIRGIZİSTAN
http://www.akademikbakis.org

	Freemans-1(M)	34	695	0.029	0.029000	0.619429	40.882
	Manufacturing (M)	77	2228	0.013	0.013000	0.380724	57.870
	Consulting(M)	46	879	0.043	0.022000	0.424637	38.217

Source: Liu, Slotine and Barabasi (2011: 169)

This table shows twelve different types (regulatory, trust, food web, power grid, metabolic, electronic circuits, neural, citation, world wide web, internet, social communication and intra-organizational) of networks. The second column contains the names of 37 networks. The third column contains the numbers of nodes (symbolized with “N”) of the related networks. The fourth column contains the numbers of links (symbolized with “L”). The fifth column contains **fractions of driver nodes in real networks** that are symbolized with “FDN”. The sixth column contains fractions **of driver nodes in fully randomized networks** that are symbolized with “FDNER” . The last two columns show the density and mean degree values of the related networks, which we calculated with a network package called PAJEK and added at the end of this table using the calculation of correlation coefficients. The correlation coefficients are shown in Table 2.

Table 2. Correlation table for FDN, FDNER, Density, Mean Degree

	N	**L**	**FDN**	**FDNER**	**Density**
L	**0.943 (p-value=0.000)**				
FDN	0.079 (p-value=0.641)	-0.018 (p-value=0.917)			
FDNER	-0.240 (p-value=0.153)	-0.273 (p-value=0.102)	0.190 (p-value=0.260)		
Density	-0.148 (p-value=0.382)	-0.153 (p-value=0.365)	**-0.498 (p-value=0.002)**	-0.218 (p-value=0.194)	
Mean Degree	-0.026 (p-value=0.880)	0.041 (p-value=0.808)	**-0.496 (p-value=0.002)**	**-0.423 (p-value=0.009)**	**0.731 (p-value=0.000)**

Table 2 shows that as the density and mean degree increase, the fractions of driver nodes (FDN) decrease. The correlation coefficients are not high but significant at the 1% level. For high density and high mean degree networks, the FDN should be low. FDNER has a low negative correlation with mean degree but no significant correlation with the density.

It is shown that the structural controllability of a network depends strongly on the fraction of low in-degree and low out-degree nodes. A strategy was proposed to improve the structural controllability of networks by adding links to low degree nodes (Menichetti,

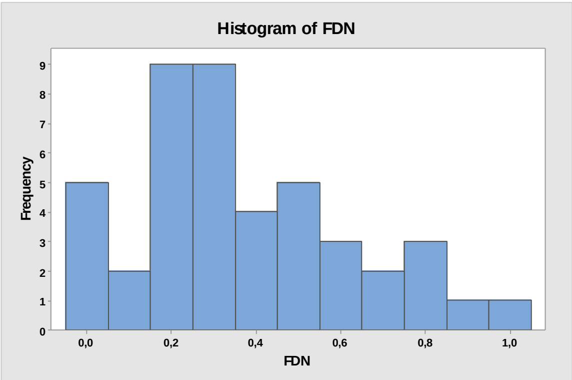
AKADEMİK BAKIŞ DERGİSİ
Sayı: 59 Ocak - Şubat 2017
Uluslararası Hakemli Sosyal Bilimler E-Dergisi
ISSN:1694-528X İktisat ve Girişimcilik Üniversitesi, Türk Dünyası Kırgız – Türk Sosyal Bilimler Enstitüsü, Celalabat – KIRGIZİSTAN
http://www.akademikbakis.org

Dall'Asta and Bianconi, 2014:4)). Our "mean degree increase, fractions of driver nodes (FDN) decrease" finding is parallel to abovementioned result.

Additionally, it is interesting to note that FDN and FDNER values (which are generated and calculated from the same networks) have no significant correlation. No statistically significant correlation is found between the fractions of driver nodes for real networks and the fractions of driver nodes for fully randomized ER networks. As a result, FDNER cannot be used instead of FDN in empirical problems.

The hierarchical structure of a network can be measured by the global reaching centrality (GRC) value. Food webs have the largest GRC, and networks of ıntra-organizational trust have the smallest (Mones, Vicsek and Vicsek, 2012: 4). For real networks, the Pearson correlations of the GRC and n_D are above 0.5, which is a relatively small value but nonetheless indicates a weak relation between the hierarchical structure of a network and a network that is easy or hard to control (low n_D-high n_D; Mones, Vicsek and Vicsek, 2012: 1-10).

Figure 1 shows the histograms of the fractions of driver nodes of real networks (FDN) and fully randomized networks (FDNER) that were drawn with the MINITAB package.



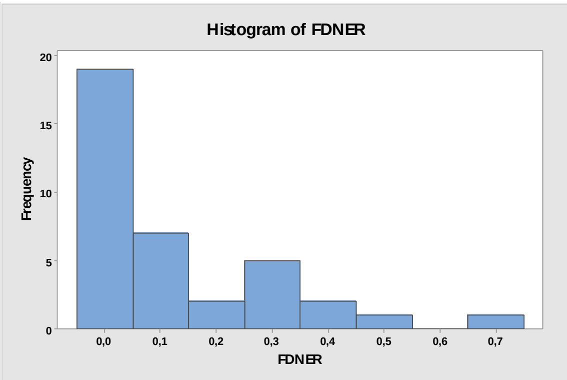


Figure 1. Histograms of FDN and FDNER (the number of networks is 37 for each)

With the naked eye, it is seen that there is a large distributional difference between the two histograms. To decide in more formal way, an Anderson Darling (AD) test was performed, and the probability plots were drawn with the MINITAB package. The following two figures (Figure 2 and Figure 3) present the results:

AKADEMİK BAKIŞ DERGİSİ
Sayı: 59 Ocak - Şubat 2017
Uluslararası Hakemli Sosyal Bilimler E-Dergisi
ISSN:1694-528X İktisat ve Girişimcilik Üniversitesi, Türk Dünyası Kırgız – Türk Sosyal Bilimler Enstitüsü, Celalabat – KIRGIZİSTAN
http://www.akademikbakis.org

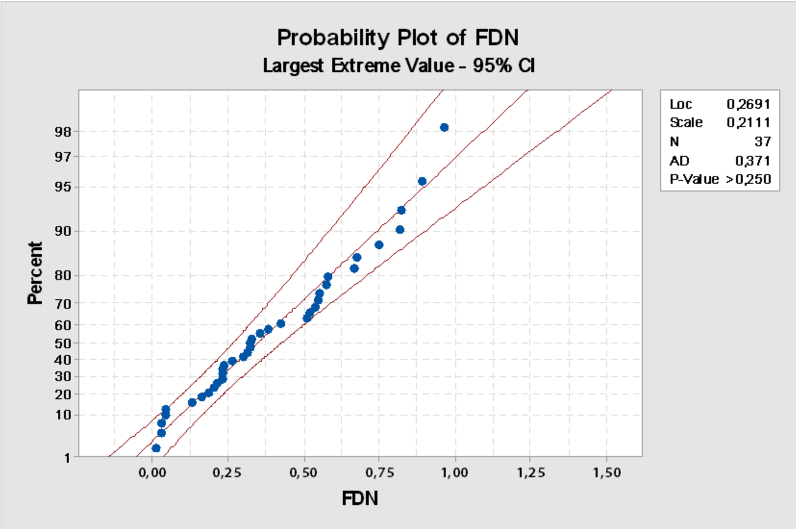


Figure 2. Distributions of FDN and FDNER (the number of networks is 37)

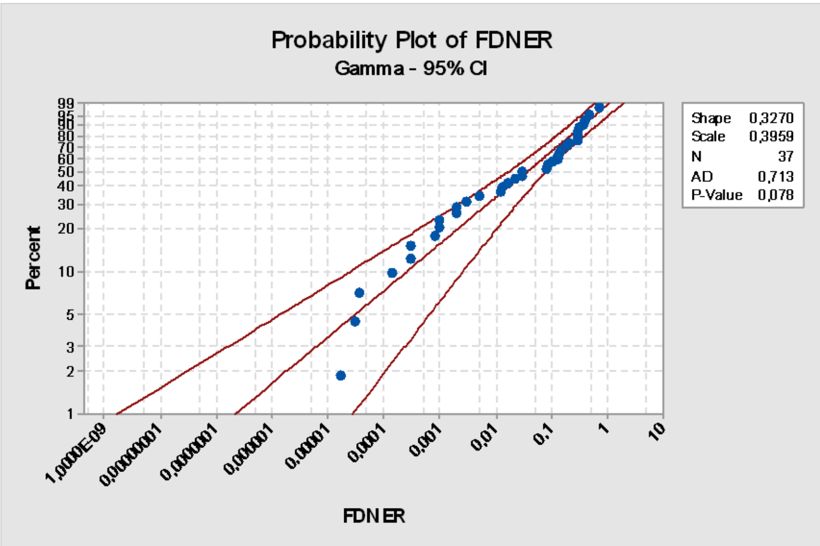


Figure 3. Distributions of FDNER (the number of networks is 37)

In Figure 2, the 0.250 p-value of FDN, which is greater than 0.05, shows that the null hypothesis, which indicates that the sample data will follow an large extreme value distribution, cannot be rejected. In Figure 3, the 0.078 p-value of FDNER, which is greater than 0,05, shows that the null hypothesis, which indicates the sample data will follow a gamma distribution, cannot be rejected. To conclude, according to the AD test results observed in the figures above (Figure 2 and Figure 3) and on the basis of sample data, the fractions of driver nodes’ distribution of real networks (FDN) found **a large extreme value**

AKADEMİK BAKIŞ DERGİSİ
Sayı: 59 Ocak - Şubat 2017
Uluslararası Hakemli Sosyal Bilimler E-Dergisi
ISSN:1694-528X İktisat ve Girişimcilik Üniversitesi, Türk Dünyası Kırgız – Türk Sosyal Bilimler Enstitüsü, Celalabat – KIRGIZİSTAN
http://www.akademikbakis.org

distribution[2], and the fractions of driver nodes' distribution of fully randomized networks (FDNER) found a **gamma distribution**.

In addition, **student t tests** were performed to determine the statistically significant difference in the mean fractions of driver nodes between natural and manmade networks and between natural and fully randomized networks. The results show that the fractions of driver nodes have a statistically significant difference between natural and manmade networks (p-value=0,025) but not fully randomized ER networks (p-value=0,437).

By following Šubelj and Bajec's (2012: 1-8) construct dependency network data from the source code of different Java projects, 7 more fractions of driver nodes were calculated for the confirmation of the previous analysis, i.e., whether the enlarged data follow the largest extreme value distribution. Therefore, the 7 additional fractions of the driver nodes of real networks (FDN) added up to 37 FDN, based on Liu, Slotine and Barabasi (2011: 169), leading to a total of 44 data. Again, the Anderson Darling (AD) test was performed. The following figure (Figure 4) presents the result:

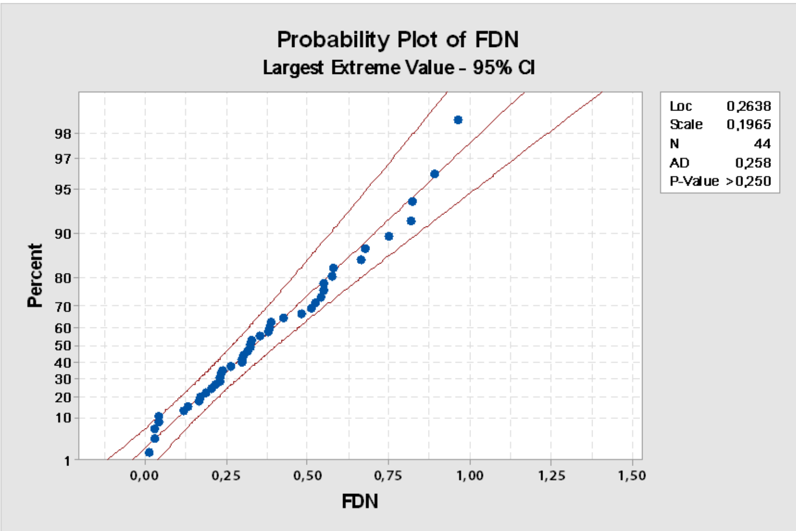


Figure 4. Distribution of FDN (the number of networks is 44)

In Figure 4, the 0.250 p-value of FDN, which is greater than 0.05, shows that the null hypothesis, which indicates that the sample data will follow a large extreme value distribution, cannot be rejected. Therefore, on the basis of enlarged sample data, the fractions of driver nodes' distribution of real networks (FDN) were found to have a **large extreme value distribution**.

Additionally, the same result was found: The fractions of driver nodes display a statistically significant difference between natural and manmade networks (p-value=0.018)

[2] This is also called standard Gumbel (maximum) distribution with location and scale parameters.

AKADEMİK BAKIŞ DERGİSİ
Sayı: 59 Ocak - Şubat 2017
Uluslararası Hakemli Sosyal Bilimler E-Dergisi
ISSN:1694-528X İktisat ve Girişimcilik Üniversitesi, Türk Dünyası Kırgız – Türk Sosyal Bilimler Enstitüsü, Celalabat – KIRGIZİSTAN
http://www.akademikbakis.org

for the enlarged data. Boxplots of natural (N) and manmade (m) networks are observed in Figure 5.

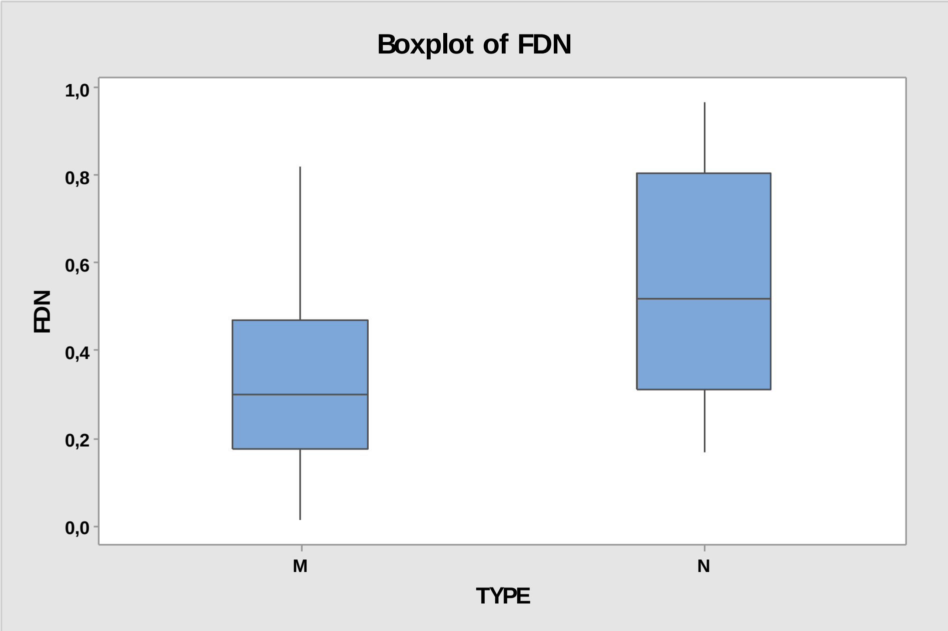


Figure 5. Boxplots of two types of FDNs (n=44)

These boxplots confirm the difference with different averages. The lines in the middle of the boxes show the medians of M and N. The median of N is much higher than N.

In Table 3 below, the descriptive FDN statistics for different types of networks are summarized.

Table 3. Descriptive FDN Statistics for Different Types of Networks

Type	Number of networks studied	Mean	Standard Deviation	Standard error of the Mean
Natural	12	0.537	0.265	0.076
Man-made	32	0.314	0.210	0.037

As expected, manmade networks are easy to control because less than 1/3 of the number of nodes must be driver nodes to control a manmade network.

4. Summary and Conclusion

New developments in network science have revealed the concept of **driver nodes**. This change is interpreted as showing that network science has left the descriptive phase and moved on to a new phase, called the **control phase**. Network scientists' focus points have shifted to determining how to control complex systems, such as biological networks, physical networks and social networks that have a primary and vital importance for scientists and human beings. In this context, analyzing and examining driver nodes of any network come into prominence. This paper addresses this importance and presents results of the fractions of driver nodes' distributions of real networks and fully randomized networks and the statistically significant difference in mean fractions of driver nodes between natural and manmade networks and between natural and fully randomized networks. On the basis of the sample results, whereas real networks follow a largest extreme value distribution, fully randomized networks follow a gamma distribution. In addition, whereas a statistically significant difference was found between natural and manmade networks, no difference was found between natural and fully randomized networks.

References

Asgari Yazdan, Ali Salehzadeh-Yazdi, Falk Schreiber and Ali Masoudi-Nejad, (2013), "Controllability in Cancer Metabolic Networks According to Drug Targets as Driver Nodes", PLOS ONE, Vol. 8, Issue 11, pp. 1-12.

Banerjee Abhijit, Arun G. Chandrasekhar, Esther Duflo and Matthew O. Jackson, (2014), "Gossip: Identifying Central Individuals in a Social Network", MIT Department of Economics Working Paper, No. 14-15, pp. 1-30.

Banerjee Soumya Jyoti and Soumen Roy, (2012), "Key to Network Controllability", (Accessed on 22 th of June 2015 via http://arxiv.org/pdf/1209.3737.pdf).

Bayer Web Site, , (Accessed on 25 th of June 2015 via http://www.bayer.com/).

Delpini Danilo, Stefano Battiston, Massimo Riccaboni, Giampaolo Gabbi, Fabio Pammolli and Guido Caldarelli, (2013), "Evolution of Controllability in Interbank Networks", Scientific Reports, 3: 1626.

Gao Jianxi, Yang-Yu Liu, Raissa M. D'souza and Albert Laszlo Barabari, (2014), "Target control of complex networks", Nature Communications, Vol. 5, No. 5415, pp. 1-8.

Inside-R web (Accessed on 22 th of June 2015 vie http://www.inside-r.org/).

Jia Tao and Barabasi, (2013), "Control Capacity and A Random Sampling Method in Exploring Controllability of Complex Networks", Scientific Reports 3, No. 2354, pp. 1-6.

Khazanchi Rohan, (2014), "On Identifying and Analyzing Significant Nodes in Protein-Protein Interaction Networks", Washington University Undergraduate Research Digest, Vol. 9, Issue 2, pp. 17-28.

Kumar Arvind, Ioannis Vlachos, Ad Aertsen and Clemens Boucsein, (2013), "Challenges of understanding brain function by selective modulation of neuronal subpopulations", Trends Neurosci, Vol. 36, No. 10, pp. 579-586.

Liu Yang Yu, (2014), "Theoretical progress and practical challenges in controlling complex networks", National Science Review, Vol. 1, No. 3, pp. 341-343.

Liu Yang Yu, Jean-Jacques Slotine and Albert Laszlo Barabasi, (2011), "Controllability of complex networks", NATURE, Vol. 473, pp. 167-173.

Šubelj Lovro and Marko Bajec's, (2012), Software Systems through Complex Networks Science: Review, Analysis and Applications, (Accessed on 22 th of June 2015 via http://arxiv.org/abs/1208.2518).

Mahia Ram Niwash, Deepak Fulwani and Mahaveer Singh, (2014), Characterization of Driver Nodes of Anti-Stable Networks, (Accessed on 22 th of June 2015 via http://arxiv.org/pdf/1410.3251v1.pdf).

Menichetti Giulia, Luca Dall'Asta and Ginestra Bianconi, (2014), "Network Controllability Is Determined by the Density of Low In-Degree and Out-Degree Nodes", Phys. Rev. Lett., Vol 113, Issue 7, 078701, pp.1-5.

Mones Enys, Lilla Vicsek and Tamas Vicsek, (2012), "Hierarch Measure for Complex Networks", PLOS ONE, Vol. 7, Issue 3, pp. 1-10.

Nikolopoulos Stavros D. and Leonidas Palios, (2005), "Maximum-Size Subgraphs of P4-Sparse Graphs Admitting a Perfect Matching". In Editors Panayiotis Bozanis and Elias N. Houtis, Advances in Informatics, 10 th Panhellenic Conference on Informatics, PCI 2005 Volos, Greece, November 2005 Proceedings, Springer, pp. 68-79.

Rathore Tapish, Siva Pratheek, Srikanth M. and Ganesh Bagler, (2015), "Study on Controllability of Complex Networks", Simon Fraser University, Report (Accessed on 22 th of June 2015 via http://www.sfu.ca/~smuralid/papers/Report.pdf).

Sabattini Lorenzo, Cristian Secchi and Cesare Fantuzzi, (2014), "Controllability and Observability Preservation for Networked Systems with Time Varying Topologies", Preprints of the 19th World Congress, The International Federation of Automatic Control, Cape Town, South Africa, August 24-25, pp. 1837-1842.

Trafton Anne, (2011), "How to control complex networks: New algorithm offers ability to influence systems such as living cells or social networks", MIT News Office. (Accessed on 22 th of June 2015 via http://newsoffice.mit.edu/2011/network-control-0512).

Wang Bingbo, Lin Gao and Yong Gao, (2012), "Control range: a controllability-based index for node significance in directed networks", Journal of Statistical Mechanics: Theory and Experiment, P04011, pp. 1-22.

Yazıcıoğlu A. Yasin and Magnus Egerstedt, (2013), "Leader Selection and Network Assembly for Controllability of Leader-Follower Networks", American Control Conference (ACC), Washington, DC, USA, June 17-19, pp. 3802-3807.

Yuan Zhengzhong, Chen Zhao, Zengru Dİ, Wen-Xu Wang and Ying-Cheng Lai, (2013), "Exact controllability of complex networks", Nature Communications, Vol. 4, No. 2447, pp. 1-9.

Zhang Xizhe, Tianyang Lv, Xue Ying Yang and Bin Zhang, (2014), "Structural Controllability of Complex Networks Based on Preferential Matching", PLOS ONE, Vol. 9, Issue 11, pp. 1-8.