ANALYSIS OF TURKISH HIGHWAY TRANSPORTATION NETWORK

2017
M. Sc. Thesis
Computer Engineering

SABAH BASHIR SALEM RASHED

ANALYSIS OF TURKISH HIGHWAY TRANSPORTATION NETWORK

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES OF
KARABUK UNIVERSITY

BY

Sabah Bashir Salem RASHED

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE IN
DEPARTMENT OF
COMPUTER ENGINEERING

March 2017

I certify that in my opinion the thesis submitted by Sabah Bashir Salem RASHED titled “ANALYSIS OF TURKISH HIGHWAY TRANSPORTATION NETWORK” is fully adequate in scope and in quality as a thesis for the degree of Master of Science.

Assist. Prof. Dr. İlker TÜRKER ....

Thesis Advisor, Department of Computer Engineering

This thesis is accepted by the examining committee with a unanimous vote in the Department of Computer Engineering as a master thesis. March 23, 2017

Examining Committee Members (Institutions)	Signature
Chairman : Assist. Prof. Dr. Burhan SELÇUK (KBU)	....
Member : Assist. Prof. Dr. İlker TÜRKER (KBU)	....
Member : Assist. Prof. Dr. Shadi Al SHEHABI (THKU)	....

....../…..../2017

The degree of Master of Science by the thesis submitted is approved by the Administrative Board of the Graduate School of Natural and Applied Sciences, Karabük University.

Prof. Dr. Nevin AYTEMİZ ....

Head of Graduate School of Natural and Applied Sciences

"I declare that all the information within this thesis has been gathered and presented in accordance with academic regulations and ethical principles and I have according to the requirements of these regulations and principles cited all those which do not originate in this work as well."

Sabah Bashir Salem RASHED

ABSTRACT

M. Sc. Thesis

ANALYSIS OF TURKISH HIGHWAY TRANSPORTATION NETWORK

Sabah Bashir Salem RASHED

Karabük University
Graduate School of Natural and Applied Sciences
Department of Computer Engineering

Thesis Advisor:
Assist. Prof. Dr. İlker TÜRKER
March 2017, 51 pages

With the fast growth of population around all over the world, and the rapid increase of cities, towns, and villages; roads that join these places where people live in, become more and more complex. In our study, we have built a database for cities of Turkey in term of the paths that connect these cites based on KGM (Karayolları Genel Müdürlüğü) national map. We then analyzed these links to find out the network properties like degree distributions, shortest paths, centralities etc. By the way, we outlined the connectivity properties between Turkish cities and presented the universal network properties that are also evident in a nation-wide highway transportation network. The major benefit of this study is outlining the important nodes in the transportation map by the means of different centrality measures, together with the modularity and degree correlations. In contrast with the classical studies, our study span the distance distribution in kilometers between the cities, which form a basis for further studies that will involve hop-distances together with the metric distances, such as finding navigation routes that have minimized distances in kilometers. The distribution of these metric distances is also presented together with the percentile plots of several parameters in 3D colored plots to outline the relative occurrence rates of these parameters with respect to others. While some generic properties of complex networks are imitated by this national transportation network, some distinct properties are also observed.

Key Word : Complex network analysis, transportation networks, scale-free networks.

Science Code : 902.2.042

ÖZET

Yüksek Lisans Tezi

TÜRKİYE’NİN KARAYOLLARI ULAŞIM AĞI ANALİZİ

Sabah Bashir Salem RASHED

Karabük Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı:
Yrd. Doç. Dr. İlker TÜRKER
Mart 2017, 51 sayfa

Dünya çapında nüfusun hızlı artışı ve dolayısıyla yerleşim yerleri sayısının hızla artması, bu yerler arasındaki bağlantıyı sağlayan yolların da giderek daha kompleks bir yapı kazanmasını sağlamıştır. Bu çalışmada Karayolları Genel Müdürlüğü (KGM) haritaları kullanılarak Türkiye’deki yerleşim birimlerini birbirine bağlayan yolları tanımlayan bir veri tabanı oluşturulmuştur. Bu bağlantılar üzerinde kompleks ağ analiz araçları kullanılarak derece dağılımı, en kısa yollar, merkezilik değerleri gibi parametrelerin analizi yapılmıştır. Bu şekilde Türkiye’deki yerleşim birimleri arasındaki ağın, evrensel gerçel ağ dinamiklerine benzer özellikler sergilediği gösterilmiştir. Çalışmanın önemli çıktılarından birisi, çeşitli merkezilik ölçekleri, modülerlik ve derece korelasyonları yardımıyla ağda ulaşım açısından önemli düğümlerin çıkartımıdır. Klasik ağ çalışmalarından farklı olarak düğümler arasındaki mesafeler kilometre cinsinden elde edilmiş ve dağılımı incelenmiştir. Bu sayede düğümler arasındaki atlama mesafeleri ile birlikte metrik uzaklıkların da dikkate alındığı, navigasyon benzeri çalışmalar içinde önemli bir veri kaynağı

oluşturulmuştur. Bu uzaklıklara ait dağılımla birlikte çeşitli parametrelere ait 3 boyutlu grafikler üretilerek parametrelerin birbiri ile ilişkisi irdelenmiştir. Gerçel kompleks ağlarda gözlenen temel özellikler bu ağda da görülmekle birlikte bazı çıktıların benzersiz nitelikte olduğu da ortaya koyulmuştur.

Anahtar Sözcükler : Kompleks ağ analizi, ulaşım ağları, ölçeksiz ağlar.
Bilim Kodu **:** 902.2.042

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor Dr. İlker TÜRKER of the Computer Engineering Department at Karabük University, without his helpful advice, valuable comments and guidance for this thesis that could not be completed without. His door was always open for me whenever I needed his help.

I want to thank my family for their support. Finally, I would like to thanks my friends, teachers for everything.

CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

Page

INDEX OF SYMBOLS AND ABBREVIATIONS

SYMBOLS

k_i^{out} : The number of outgoing connections (degree)

k_i^{in} : The number of incoming connections (degree)

k_i : Node degree

l_{ij} : Path length between nodes I and j

$<k>$: Average degree

$\mathrm{P}(k)$: Degree distribution function

$\mathrm{C_c}$: Clustering coefficient

Ci^D : Degree centrality

Ci^C : Closeness centrality

Ci^B : Betweeness centrality

Ci^S : Straightness centrality

Ci^I : Information centrality

ABBREVIATOINS

WWW : World Wide Web

KGM : Karayolları Genel Müdürlüğü (General Directorate of Highways)

BC : Betweenness Centrality

CC : Closeness Centrality

EC : Eigenvector Centrality

CHAPTER 1

INTRODUCTION

1.1. COMPLEX NETWORKS

As a term, complex networks define a wide range of scientific studies that concern web-like structures that consist of interconnected components. Complex, non-uniform interconnectedness is present in such diverse areas as human society and social interaction [1].

In contrast with the past, wide ranges of databases are available nowadays for analyzing and studying complex networks. These kinds of analyzing studies cannot be performed without powerful modern computers. The databases defining complex structures are growing daily. The concept of “complexity” may well refer to a quality of the system or to a quantitative characterization of that system [2, 3]. We can give several examples of these kind of networks such as world wide web (www), transportation systems, food webs, internet, social networks, neural networks, GPS navigations which store the paths for moving patterns, emails data, epidemiology networks etc. To understand the term complex networks, we first need to define the term of networks in detail.

Network is a set of nodes (vertices), connected via links (edges). Network science is considered as one of the main important fundamental areas of the discrete mathematics. In 1735 Euler published the first paper in graph theory when he found solutions for bridges of Königsberg [4]. Since that time, graph theory has evolved so much. With the rapid development of computers and communications, the analysis of complex networks has become more ferformable. Thus, many papers have been published in this field, which have tested a large scale of networks and discovered that these networks in the reality show features lead to important topological structure. The interactions between the components of many complex networks can be described in detail. In this chapter, we exhibit types of complex networks and discuss them.

1.2. COMPLEX NETWORKS FEATURES

Although there are many types of complex networks, most of these networks share common characteristics, some of them mentioned below. We also give a brief list of nework parameters and distributions which are used to define the network dynamics.

1.2.1. The Size of A Complex Network

There is no specific definition to determine whether a network is complex or not. Instead, according to the size of the network we conclude is a complex network or not [5]. What is meant by size is the number of nodes in the network (huge number of nodes means more complex). The world-wide web is an example of such a network composes of numerouse nodes and connections. By the addition of new nodes everyday, it is almost impossible to count the number of web pages. On the other hand, there are many real networks that are sprinkled, which means the number of nodes are countable. As a conclusion, the size of the complex networks expands by the addition of new nodes (which means they are dynamic, not static in term of size), and the size can be labeled as the number of nodes it contains.

1.2.2. Average Path Length

Path length a useful concept used to explain how easy it is to reach from one node to another. It defines how many steps in a shortest way from one node no another is. Averaging the path lengths between all node pairs in a network gives the average path length, which is a parameter used to explain the diffusion capability in a network [6].

To give a measure of the 'linear size' of a system, the ideas of the mean $<L>$ and maximal $<L>_{max}$, shortest paths are helpful. For a connected network of nodes (N), the mean shortest path is defined by:

$$l = \frac{2}{n(n-1)} \sum_{i>j} l(i,j) \tag{1.1}$$

where l represents the length of the shortest paths between node (i) and node (j).

1.2.3. Small World

Recent studies about real networks show that, regardless from the size of the network, most of these networks display short average path length values [5]. The concept of "six degrees of separation" is considered as the most famous manifestation of small-world phenomenon which was discovered by Stanley Milgram (1967), a social psychologist who found that there was a path associated with a typical length of about six between most pairs of people in the United States. The property of small-world is used for most complex networks characterization. For example, the Hollywood characters are approximately 3 co-stars distant from each other, or the chemicals within a cell are isolated by approximately 3 reactions. Erdos-Renyi have stated out that in a random graph, the path length between any two nodes scales as the logarithm of the number of nodes. Thus, random graphs are small worlds as well.

1.2.4. Clustering

Clustering is of the common features of social networks those are formatted just like cliques. For example, if we have group of friends, every member within the group knows all the other members in the same group. Clustering coefficient is used to describe the inherent tendency of clustering [5]. The clustering coefficient can be simply described as follows:

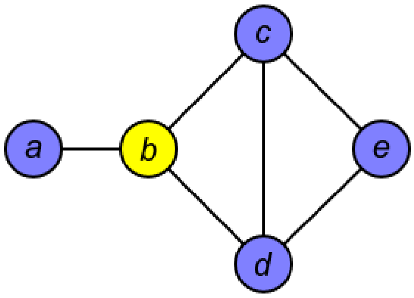

Figure 1.1. Clique of points.

In Fig. 1.1, we have node "b" in a clique of points. K_b is the degree of node b which represents the number of the neighbor nodes connected directly to node b (In our example K_b=3). N_b represents the number of links between the neighbors of b. In this case, N_b=1, defined by the only link between d and c). To calculate the clustering coefficient, we use this equation:

$$C_c(b) = \frac{2N_b}{K_b(K_b - 1)} \tag{1.2}$$

The value of $C_c(b)$ is always between 0 and 1. If $C_c(b)$ is close to 1, it means the network has a high clustering, while the low values close to 0 indicate low clustering. Usually most of the real networks exhibit high clustering.

1.2.5. Degree Distribution

The number of edges for each node are not the same in a network (i.e. nodes have different degrees). In this context, the number of the connections per each node is represented by its degree, denoted by k. The distribution of the node degrees is represented by the distribution function $P(k)$. This function presents a probability for a randomly selected node to have the value of degree k [6].

In a random graph, the edges between nodes are distributed in a random way that makes most of the nodes to have approximately the same degree, very close to the average degree (k) of the network. In the random graphs, the distribution type is a Poisson distribution.

The degree distribution for the large complex networks remarkably deviates from a Poisson distribution. In fact, many complex networks such as world wide web or Internet, their degree distribution has a power-law tail. This type of networks is called scale free. On the other hand, there are some networks that show an exponential tail and the function *P(k)* still deviates from the Poisson distribution. Due to these discoveries, the complex network modeling has been developed recently which classified the modeling according to [6] into three classes:

1. First random graph which are variations of the Erdos-Renyi model, are still generally utilized as a part of numerous fields and serve as a benchmark in vast range of modeling and experimental studies.
2. The second class is motivated by clustering which called small-world models.
3. The third class about the revelation of the power law degree distribution has prompted the development of different scale-free models that, by concentrating on the system elements, expect to offer an all-inclusive hypothesis of system development.

1.2.6. Scale-Free Network

Scale-free network is a graph with a feature that the number of edges k generating from a node shows a power law distribution P(k). This network can be built by adding nodes to the network and connecting it ti the recent nodes with preferential attachment. Preferential attachment means that the probability of connecting to a node is proportional with the degree of that node, resulting the situation that rich get richer. Scale free networks have a power-law degree distribution as:

$$P(k) \sim k^{-\gamma} \tag{1.3}$$

Where γ is labeled as the power-law exponent. This distribution indicates that while k increases, the value of p(k) decays resulting the probability of finding a node with a high degree degreases as well.

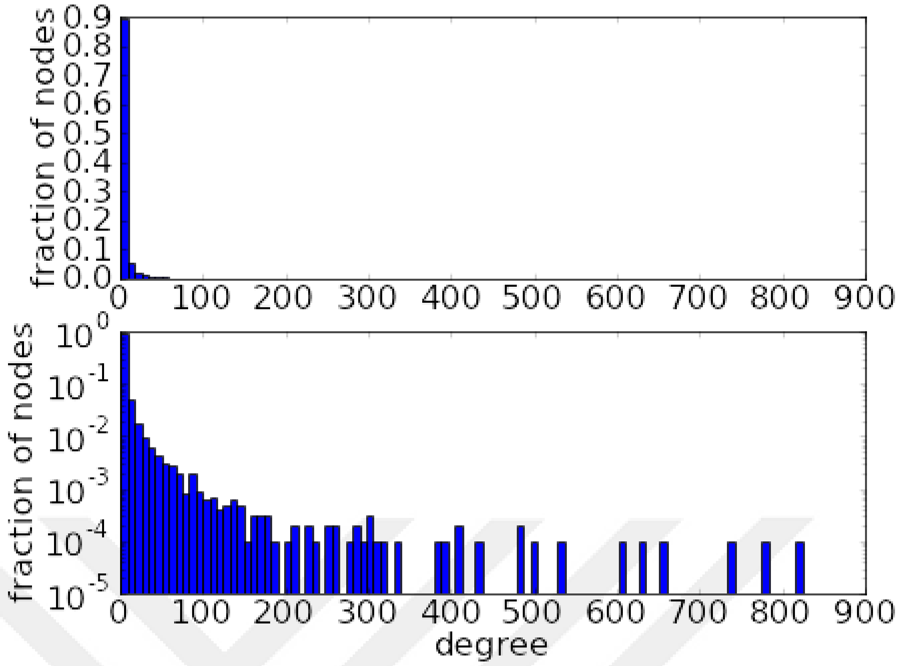

Figure 1.2. The degree distribution of a scale-free network.

The figures above introduce the degree distribution of a scale free network of N=10.000 nodes and power law exponent ($\gamma = 2$). We can recognize that in the first plot, it is impossible to extract the right distribution of degrees whereas the logarithmic scale in the secont plot exposes the long tail of degree distribution. We also notice that although majority of the nodes have little degrees, there are some nodes having degrees greater than 500 which represents the characteristic of a scale-free network.

1.2.7. Centrality

In term of complex networks centrality measures specify the most essential nodes inside a graph. It used to rank the nodes according to their importance. The possibility of centrality was initially connected to human correspondence by Bavelas [7] who was keen on the portrayal of the correspondence in little gatherings of individuals and accepted a connection between basic centrality and impact/ power in gathering forms. From that point forward, different measures of basic centrality have been proposed over a long time to evaluate the significance of a person in a social network. There are some important concepts related to the centrality that must be mentioned [28]:

1. Degree centrality (DC): This term is used to define the centrality of a node based on the number of links connected to that node. The most important node is the node that has the largest number of connections.

$$Ci^D = \frac{\sum_{j \in G} aij}{N-1} = \frac{Ki}{N-1} \tag{1.4}$$

2. Closeness centrality (CC): Evaluates which range node *i* is close to all other nodes as in the equation:

$$Ci^C = \frac{N-1}{\sum_{\substack{j \in G \\ i \neq j}} d_{ij}} \tag{1.5}$$

   where d_{ij} represents the shortest path distance between node *i* and node *j*.

3. Betweenness centrality (BC): When a node located between the shortest paths of large number of nodes, it is considered as a central node. In other words, the node is considered a central node if many shortest paths traversed that node. BC is calculated as follows:

$$Ci^B = \frac{1}{(N-1)(N-2)} \cdot \sum_{\substack{j,k \in G \\ j \neq k \neq i}} \frac{n_{jk}(i)}{n_{jk}} \tag{1.6}$$

   where n_{jk} represents the how many shortest paths between j and k where $n_{jk}(i)$ represents the number of shortest paths that traversed node *i*.

4. Straightness centrality (SC): The equation below describes this term which is basically the relationship between the efficiency and the inverse of the shortest path distance d_{ij}. The distance used here is the Euclidean distance.

$$\mathrm{Ci}^S = \frac{\sum_{j \in G} \frac{dij\ (Eucl)}{dij}}{N-1} \tag{1.7}$$

5. Information centrality (IC): It is the measure of the ability of how the network performs when an important node deactivated. This measurement should be done before the deactivation and after. The information centrality of node *i* represents the deactivated node in the network.

$$\mathrm{Ci}^I = \frac{\Delta E}{E} = \frac{E(G)-E(G\prime)}{E(G)}, E(G) = \frac{\sum_{\substack{i,j \in G \\ i \neq j}} \frac{dij(Eucl)}{dij}}{N(N-1)} \tag{1.8}$$

where G′ is the network with N nodes and K-k*i* edge obtained by removing from G the edges incident in node i. Notice that E(G) is finite even for a non-connected graph.

1.3. TYPES OF COMPLEX NETWORKS

In this section, we highlight on the knowledge about the real networks by presenting several types of complex networks. The mathematical representation of a network is very useful because it makes us able to model the network and extract the properties of it. Although there are many types of complex networks, they can be divided into four main categories as below:

1.3.1. Social Networks

A social network can be defined as a group of people with relationships connecting them [8,9]. For example, the patterns of friendships between individuals [10,11], business relationships between companies [10], and intermarriages between families [12] are the examples of networks that been studied in the past. Facebook is an example of a complex network, where the individuals represent the nodes, the relationships between them represent the links. Other example of this type of networks is the e-mails between individuals where the addresses are the nodes and the path of transferring this message from one address to the destination address represents the link between these nodes. Movies database appear to behave as s social network when two or more actors appear in the same movie. The telephone communications between individuals is another kind of social network, where the telephone number considered are to be the nodes and the directed edges calls from one number to another define the links. Telephone networks are big network having over 50 million nodes distributed around the world. In fact, it takes the second place after the world-wide web in term of size.

1.3.2. Information Networks (Knowledge Networks)

The second category, information networks are about the information and their relationships. Academic papers citation is a good example for this category. It is rare to find an academic paper proposed without no cite by other papers on related topics. Eventually, if we have paper A and the author cited the paper B, both A and B are connected nodes. It is worth mention that the citation network point forward because the citation happens in one direction to another only. In another word, according to the example that we mentioned previously, the author of paper A hasn’t finished his paper yet so the author of paper B cannot cite by paper A (see Fig. 1.3).

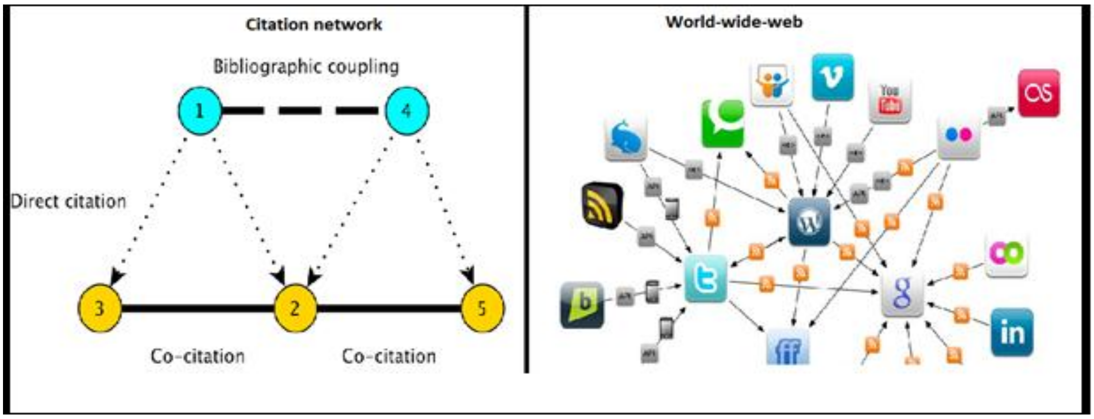

Figure 1.3. Link direction differences.

The second example of this category is the world-wide web, which is a very giant number of web pages connected by hyperlinks. Unlike the citation network, all the web pages can be reached in all directions (cyclic). One important point to notice about the Web is that our data about it come from "crawls" of the network, in which the web pages are found by following hyperlinks from other pages [13]. The web pages represent the nodes in the world-wide web, the edges are the links (URL) that are used to connect these web pages as shown in Fig. 1.4.

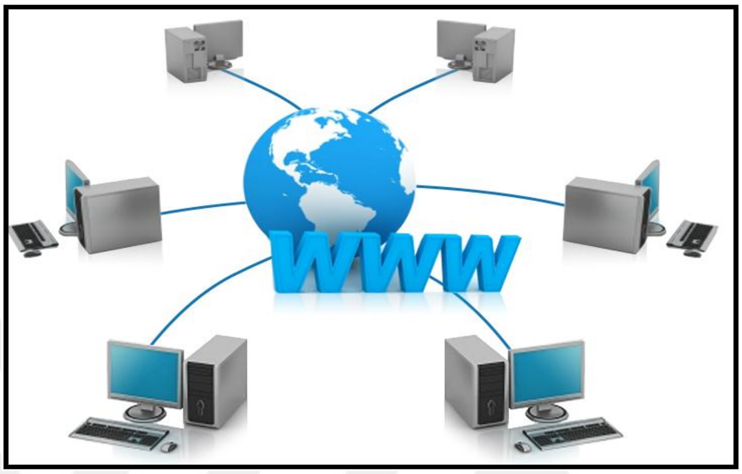

Figure 1.4. Web pages connected by hyperlinks.

The world-wide web network takes the first place comparing with the other types of networks in term of nodes counts, in 1999 the number of nodes were around one billion [29], since that time this number is increasing rapidly. Preference network is one of the information technology networks category. It connects the objects with individuals. The objects could be anything that the individuals like such as movies, books, etc.

1.3.3. Technological Networks

Technological networks are human-made networks to distribute resources or information such as internet, power grid network. As mentioned previously a network can be defined as a set of nodes connected via edges. Based on this definition, Internet is considered as one of the most complicated networks. Computers and routers represent nodes whereas the wires that used to connect these computers and routers represent the edges as shown in the Fig. 1.5.

Figure 1.5. Complex network (Internet).

It is clear to say that Internet is considered one of the most complicated networks due to the tremendous numbers of edges and vertices that are distributed around the world. This number increases daily, together with complexty degree. Different topologies are in use to construct local networks connected to other local networks with different topologies. In 2001 the number of hosts were about 100 million [29]. According to the study in [14], a comparison between several studies was introduced. This comparison focused on the interdomain level of the internet. The papers published in 1999 stated that the number of nodes at that time was 5287.

Another example of technological networks is power grid network, which represents one of the largest and most complex networks [15]. The power stations represent the edges whereas the vertices are represented by the power transmittion lines. Through the analysis of these networks we can find the loads applied on each node, which node has a heavy load and which one is not. Fig. 1.6 shows an illustration of this type network.

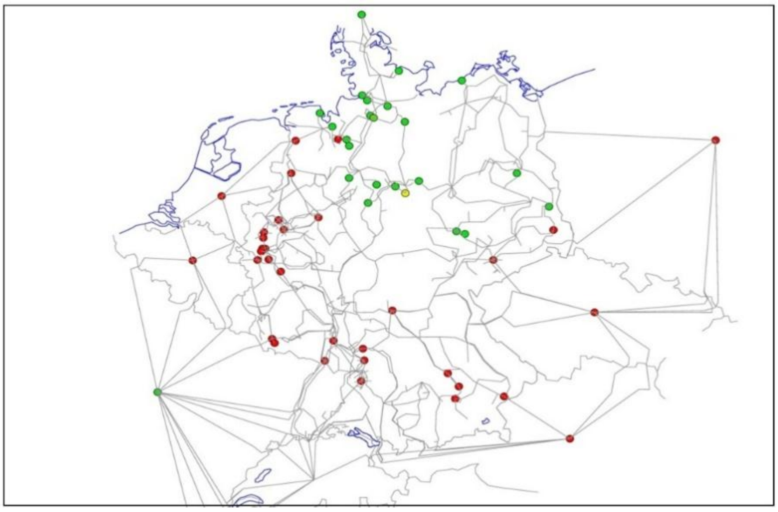

Figure 1.6. Power grid network.

1.3.4. Biological Networks

One of the good examples of this category is the metabolic networks which is a representation of metabolic substrates and products with directed edges joining them, if a known metabolic reaction exists that acts on a given substrate and produces a given product. Metabolic pathways generally have giant maps.

Another famous studied example of this type of network is the food web, in which the vertices represent species in an ecosystem and a directed edge from species A to species B indicates that A preys on B as in Fig. 1.7. Protein interactions analysis gives us a good opportunity to understand the cellular network, where the proteins represent the nodes and the interactions between them represent the links. According to the study published by Cerevisiae and Pylori (Jeong et al., 2001; Wagner, 2001) protein interactions show a scale-free behavior even the sources of gathering data about protein interactions are different.

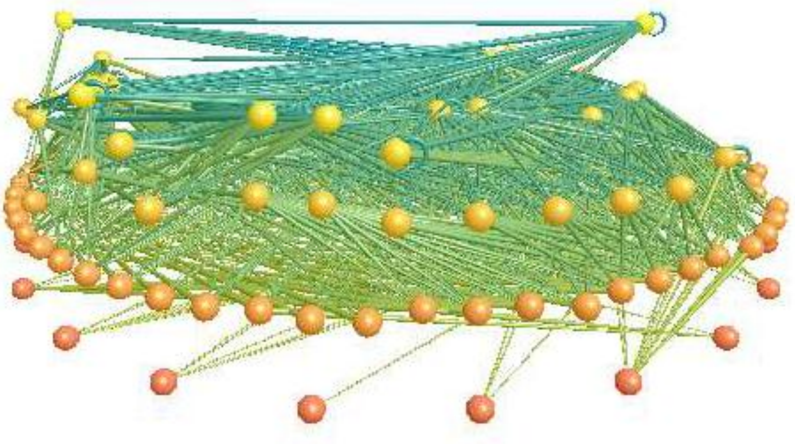

Figure 1.7. A food web [26].

1.4. RELATED WORK

The vast increase of the transportations led to a high degree of complexity for the transportation lines. Many studies introduced recently analyze transportation networks which are considered as one of the complex networks types. We surveyed some of these studies in this section.

In Ref. [16] that the Airport network of India (ANI) is studied. It is stated out that ANI is a small world network with complex dynamics similar to those of greater air transportation networks. It shares the same common characteristics with real networks such as scale free distribution. ANI is expected to grow at a rapid speed with addition of airports and several low-cost air services [17].

Another study analyzed the travel routes of the rail RTS and bus BUS public transportation systems in Singapore from a complex weighted networks perspective. They found that the topological properties may display some deviations from the majority of the network studies.

An analysis study on Pakistan highway was presented in [18] to evaluate the contribution of the weighted complex networks in Pakistan highway. Although the highway in Pakistan is responsible of 75% of transporting, it is still poor, unreliable and inadequate to use, as the author states. The author mentions that the highway exhibits the properties of small-world features, and concludes that the Pakistan national high way network PNHN is a highly clustered network. He found that PNHN properties are similar to the Indian highway network properties. Furthermore, using betweenness centrality, the cities with potential traffic congestion are also identified. The author in [19] presented a study on road networks in India. The author considered that the junction points are the nodes and the links between them are the edges. He explored the topological features of this network and the community structure of it, and discovered that this network has the small world network properties. Based on the centrality and betweenness, he identified the most important cities on the highway, which helps to identify the congestion points on the network. He proposed a novel approach to improve the performance of the network and decrease the number of congestions within the network (see Fig. 1.8).

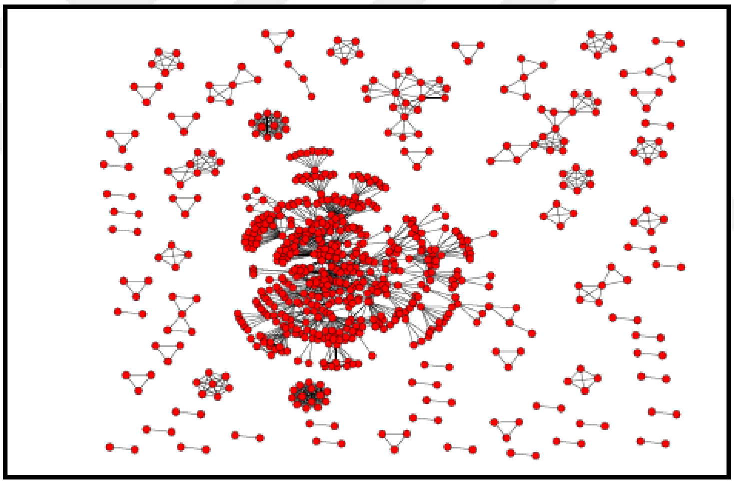

Figure 1.8. Indian highway structure [16].

The author made a comparison of several parameters between the Indian highway network, Erdos-Renyi model and Configuration model as in Table 1.1. He concluded that the distribution of node-connectivity is neither a power law nor normal.

Table 1.1. Comparison of network properties of Indian Highway Network Erdes-Renyi model (p = 0.009186) and Configuration model.

Properties	I H N	Erdos-Renyi	Configuration
V	694	694	694
E	2209	2148	2169
K	6.36	6.19	6.25
C	0.78	0.008	0.015
D	4.75	3.77	3.61

The next study that we want to present is the one presented in [20]. This study about analyzing highway network in Guangdong which is a china's province (GDHN). The author used the L space methodology to present the topology structure of the GDHN province. The author applied the network properties such as closeness centrality, distribution and betweenness centrality, to analyze the GDHN properties. He found that the highway development is unbalanced between the regions, so he proposed some suggestions for future construction.

The study was introduced in [21], analyzing the Korean highway system that includes two types of transportation information as public and private. The author found that there is a similarity between the relationship between the cities connected directly and Newton's law of gravity where the gravity law is ($F \sim P_i P_j / r^2$), where P_i is the population in the first city while P_j represents the population of the second city and r is the distance between them.

Another study proposed in [22] reviews a list of several real networks by the means of some network parameters, as abstracted in Table 1.2. The power-law degree exponents and related parameters of the networks are also presented in Table 1.3.

In Table 1.2, *k* represents the average degree, *l* is the average path length, where *C* is the clustering coefficient. The author also lists the average path length and the clustering coefficient values for a comparable random graph.

Table 1.2. List of networks with their properties [22].

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

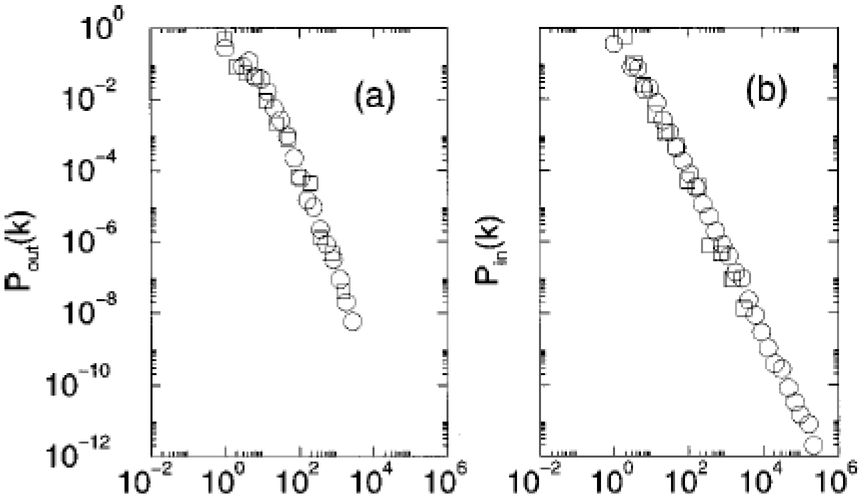

Figure 1.9. Degree distribution of the world-wide web. a) The degree distribution of outgoing edges, b) The degree distribution of incoming edges. Logarithmic scales are used to reduce the noise and display power-law consistencies better.

To exemplify the characteristics of one of the networks in Table 1.3, we present the degree distribution of the world-wide web in Fig. 1.9. Two different measurements were used: the squares represent the 325,729-node sample of Albert et al. (1999); whereas the circles represent 200 million measurements pages by Broder et al. (2000).

Table 1.3. List of network characteristics [22].

Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}	ℓ_{real}	ℓ_{rand}	ℓ_{pow}	Reference	Nr.
WWW	325 729	4.51	900	2.45	2.1	11.2	8.32	4.77	Albert, Jeong, and Barabási 1999	1
WWW	4×10^7	7		2.38	2.1				Kumar *et al.*, 1999	2
WWW	2×10^8	7.5	4000	2.72	2.1	16	8.85	7.61	Broder *et al.*, 2000	3
WWW, site	260 000				1.94				Huberman and Adamic, 2000	4
Internet, domain*	3015–4389	3.42–3.76	30–40	2.1–2.2	2.1–2.2	4	6.3	5.2	Faloutsos, 1999	5
Internet, router*	3888	2.57	30	2.48	2.48	12.15	8.75	7.67	Faloutsos, 1999	6
Internet, router*	150 000	2.66	60	2.4	2.4	11	12.8	7.47	Govindan, 2000	7
Movie actors*	212 250	28.78	900	2.3	2.3	4.54	3.65	4.01	Barabási and Albert, 1999	8
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2	4	2.12	1.95	Newman, 2001b	9
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1	6	5.01	3.86	Barabási *et al.*, 2001	10
Co-authors, math.*	70 975	3.9	120	2.5	2.5	9.5	8.2	6.53	Barabási *et al.*, 2001	11
Sexual contacts*	2810			3.4	3.4				Liljeros *et al.*, 2001	12
Metabolic, *E. coli*	778	7.4	110	2.2	2.2	3.2	3.32	2.89	Jeong *et al.*, 2000	13
Protein, *S. cerev.**	1870	2.39		2.4	2.4				Jeong, Mason, *et al.*, 2001	14
Ythan estuary*	134	8.7	35	1.05	1.05	2.43	2.26	1.71	Montoya and Solé, 2000	14
Silwood Park*	154	4.75	27	1.13	1.13	3.4	3.23	2	Montoya and Solé, 2000	16
Citation	783 339	8.57			3				Redner, 1998	17
Phone call	53×10^6	3.16		2.1	2.1				Aiello *et al.*, 2000	18
Words, co-occurrence*	460 902	70.13		2.7	2.7				Ferrer i Cancho and Solé, 2001	19
Words, synonyms*	22 311	13.48		2.8	2.8				Yook *et al.*, 2001b	20

In Table 1.3, the degree distributions were characterized by scaling exponents of several networks. The size of each network was stated, together with the average degree $<k>$, where κ represents the cutoff for the power low scaling. γ_{in} and γ_{out} represent the in degree and the out degree respectively and L is the path length. As seen from the two tables representing several parameters of several real networks, regardless from what type of networks they are, they represent some universal properties like small average path, high clustering, power-law consistent degree distributions etc. These universal properties of diverse systems indicate that the underlying mechanisms of these systems have strong resemblance.

The above-mentioned parameters are also similar with the transportation network outputs, presented by the mentioned studies in this section. Another study about urban public transportation networks of five Hungarian cities is presented in [23]. A comprehensive investigation was performed on this network to determine the main characteristics of these cities. The author used the weighted links to consider the capacity of different vehicles such as buses, trucks and other kind of vehicles in the crowed hours during the morning. Table 1.4 introduces the characteristics of these five cities.

Table 1.4. Urban public transportation network of Hungarian cities. Links–simple refers to the number of links in the simplified graphs, where a link is defined between two nodes these nodes are the consecutive stations of at least one line. Links-multiple refers to the number of links in the models, each line between two stations is represented by a link. the codes of the vehicle types are as follows: B: bus, E: electric trolleybus, T: tram [23].

City	Area (km^2)	Pop. ($\times$ 1000)	Density (inhab./km^2)	Nodes	Links–simple	Links–multiple	Lines	Diameter	Avg. path length	Vehicle types
Debrecen	461	204	442.5	306	711	1772	53	41	11.7	BET
Győr	174	129	741.4	230	529	1391	43	30	10.8	B
Miskolc	236	161	682.2	257	535	977	35	45	14.5	BT
Pécs	163	147	901.8	256	569	1960	55	36	13	B
Szeged	281	162	576.5	242	558	1192	40	35	11.8	BET

From Table 1.4, we can easily notice that not all the cities have the same vehicle type. The author proposed a comparison between the weighted and unweighted cases in term of several centrality measures of the nodes. Based on that comparison the author got the ability to distinguish the most critical network's paths and stations. As in the figure below in the unweighted, the centrality value is decreasing.

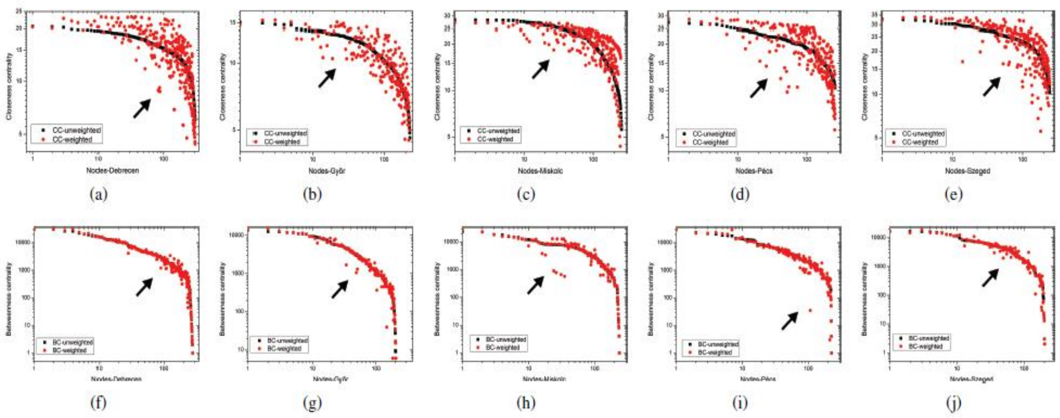

Figure 1.10. Closeness and betweennes centrality frequencies for the five cities.

The study lead the author to expose the clear picture of the variations of transportation network organizations with the influence of some external factors such as geographical or even historical ones. The author chooses the five Hungarian cities based on:

1. The population of each city between 100,000-250,000.
2. All the five cities almost share the same characteristics (economic role, land use) and the public transportation of these five cities are organized in similar way.
3. The geographical conditions are different.

So, according to these criteria the five cities were selected, the author concluded that the way that they followed to analyze the networks can expose efficiently the network structure and its organization. Despite not all the vehicles are full with passengers, the author used count of vehicles as a good approximation to evaluate the maximal load links between stations.

1.5. SIGNIFICANCE OF THE STUDY

The study performed is the first study involving the Turkish highway transportation network, outlining the network structure by the means of network science metrics. We have built a database for cities of Turkey in term of the paths that connect these cites based on KGM (Karayolları Genel Müdürlüğü) national map. The major benefit of this study is outlining the important nodes in the transportation map by the means of different centrality measures, together with the modularity and degree correlations. In contrast with the classical studies, our study span the distance distribution in kilometers between the cities, which form a basis for further studies that will involve hop-distances together with the metric distances, such as finding navigation routes that have minimized distances in kilometers.

CHAPTER 2

TRANSPORTATION NETWORKS

2.1. INTRODUCTION

The patterns in transportation networks can be extracted both manually and by automated services (if available). The users of the systems which can be classified into three types (drivers, pedestrians, and passengers), minimizes the disutility connected with transportation. For instance, drivers driving from a known origin point to a known destination point mostly prefer the shortest paths to reach their destination.

The distance of the path is associated with the travel time according the basic physics laws. The driver example shows that the travel time on each end of any two connected points can be represented mathematically as a function of the total traffic flow in the possible paths. Since the traffic jam is not so easy to determine, the path within the shortest travel time is not always a trivial process. Travelers take into consideration several factors before they make their decision such as congestion points, destination, and how to get to that destination. The most compelling obstacles that face the transportation engineers is how to predict the effect of transportation scenarios.

Worldwide urbanization is combined with huge number of vehicles nowadays. The aggregate vehicle stock is anticipated to increment from 800 million vehicles in 2002 to more than 2 billion vehicles by 2030. Also, many researches have demonstrated that vehicle-miles headed out increment relatively to the accessible miles of roadway, recommending that the development of new streets does not adequately diminish congestion. Congestion is subsequently a critical issue in transportation approach, as it damages our surroundings and economy because of emanations and expanded travel times. There are vital devices utilized by urban organizers and transportation specialists to comprehend congestions such as Microscopic traffic simulators and macroscopic traffic models. Traditionally, microscopic simulators have been utilized to gauge framework execution and direct situation based examinations of urban systems. These simulators display complex choices of individual drivers, for example, the determination of routes and path evolving conduct, and additionally the cooperation's between drivers. Along these lines, the assessment of these models are computationally costly. Plainly visible models have been utilized to decide techniques to diminish congestion, for example, improving sign arrangements for a set of convergences, however depend on a less sensible arrangement of suppositions.

2.2. STATE OF ART IN TRANSPORTATION NETWORKS

With the end goal of analyzing transportation related issues and for transport arranging as a rule, particular types of models have been produced throughout the years; among most regular are travel request models or activity models. These are intended to assess transport free market activity. At exhibit time, we can separate two fundamental sorts of transport issues, one where monetary development and advancement produces such request that it surpasses the supply (limit) of the transport framework within reach. The other kind of issue depends on circumstances where long stretches of under-venture brought about a poor supply framework that can't fulfill the request. Inside these travel deman models (TDMs), we can frame another sub-division by the topographical region they mean to portray and figure; these can shift from a solitary convergence, to a city, nation or a considerably bigger district. This study concentrates on demonstrating a more extensive locale, and subsequently the emphasis will be on territorial level demonstrating. For all intents and purposes each TDM forward elements some variant of a customary 4-stage display that is destined to be depicted in more detail. Before going any further and clarifying this sort, we must say that there are different sorts of models utilized as a part of transport arranging.

2.3. NETWORK CHARACTERISTICS

In our study, we analyzed transportation network of Turkey which connects the cities of Turkey. Our data source is the online road maps located in the website of “Türkiye Karayolları Genel Müdürlüğü” (KGM), which is divided into regions as shown in Fig. 2.1.

Figure 2.1. Map of Turkey.

The features of this map are as follows:

1. It is divided into 18 regions.
2. Each region composed of several provinces.
3. Each province is divided into several districts.
4. The district within each province may contain some towns and/or villages.
5. These (provinces, towns, villages) are connected via roads.
6. The distances between the cities are stated on each link.
7. The link to access this map is in [6].
8. We will use the terms *node* and *link* both when mentioning basic graph and their unpredictable ensembles, systems. The *node* degree advises what number of links are appended to the node (see Fig. 2.2).

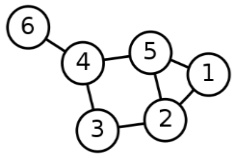

Figure 2.2. Nodes and links.

In this simple network the number of nodes is N=6 and links is M=7. Nodes 2, 4 and 5 have degrees k=3, whereas node 1,3,4 has k=2 and node 6 has k=1. According to this example we have calculated the total number of nodes, links and the degrees for each node.

There are two types of links, directed and undirected links. A directed link means the link has a direction as in or out from/to a node, resulting in-degrees and out-degrees for the nodes correspondingly. An undirected link has no direction, thus results only one-degree parameter for a node. In our network construction procedure, we used undirected links, which is the suitable method for such a transportation network, since the transportation in a specific road is realized in two directions.

2.4. METHODOLOGY

The main sections of the empirical study we performed can be listed as:

1. Manually collecting the node data from the detailed KGM maps, to an MS Access database consisting of three tables corresponding to three levels of cities (See Fig. 2.2 and Fig. 2.3).
2. Manually collecting the links data, corresponding to the connections between the cities with given IDs, to the same database. Links table is responsible for the roads data, while it also contains the metric distance between the localizations in kilometers (See Fig. 2.4).
3. Based on the links table, analyzing the network characteristics using Gephi software and visualizing the results in MATLAB.

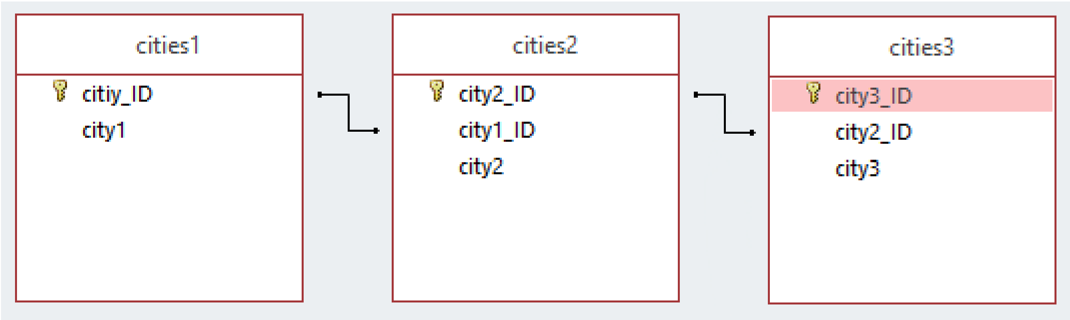

Figure 2.3. The database diagram of cities.

cities1

citiy_ID	city1
1	Adana
2	Adıyaman
3	Afyonkarahisa
4	Ağrı
5	Amasya
6	Ankara
7	Antalya
8	Artvin
9	Aydın

cities2

city2_ID	city1_ID	city2
1	1	Seyhan
2	1	Ceyhan
3	1	Feke
14	1	Karaisalı
15	1	Karataş
16	1	Kozan
17	1	Pozantı
18	1	Saimbeyli
19	1	Tufanbeyli
20	1	Yumurtalık

cities3

city3_ID	city2_ID	city3
1	2	Doruk
3	2	Termık santral
4	2	Narlık
5	2	Kurtpınarı
6	2	Mercimek
7	2	Yılamkale
8	21	Yakapınar
9	2	Kıvrıklı
11	25	Tuzla
23	19	Yılankale
25	27	Cakırhöyük

Figure 2.4. The data collection for cities, in three levels connected to each other.

Links

ID	source	target	distance km
1	1361	591	7
2	591	592	11
3	592	593	10
4	593	1275	17
5	594	593	13
6	11	594	18
7	11	1767	46
8	900	1275	33
9	8	1275	13
10	7	8	17
11	1357	7	18
12	1366	3	18
13	3	4	10
14	1	4	9
15	4	1358	9

Figure 2.5. The data collection for links between cities (based on cities3 table), including the source and target node IDs together with the distances in kilometers.

2.4.1. Building The Database Tables

The highway map mentioned in Fig. 2.1 is an index for more detailed sub-sections, which are available after clicking on a specific region. An a-example of these detailed sub-sections, together with a magnified view are given in Fig. 2.5 and Fig. 2.6. Based on these detailed sectional maps, we constructed our cities1, 2 and 3 tables to contain three levels of settlements as given in Fig. 2.3. Although the real list of these settlements in Turkey are larger than these maps, we only captured the cities those are present in the detailed maps.

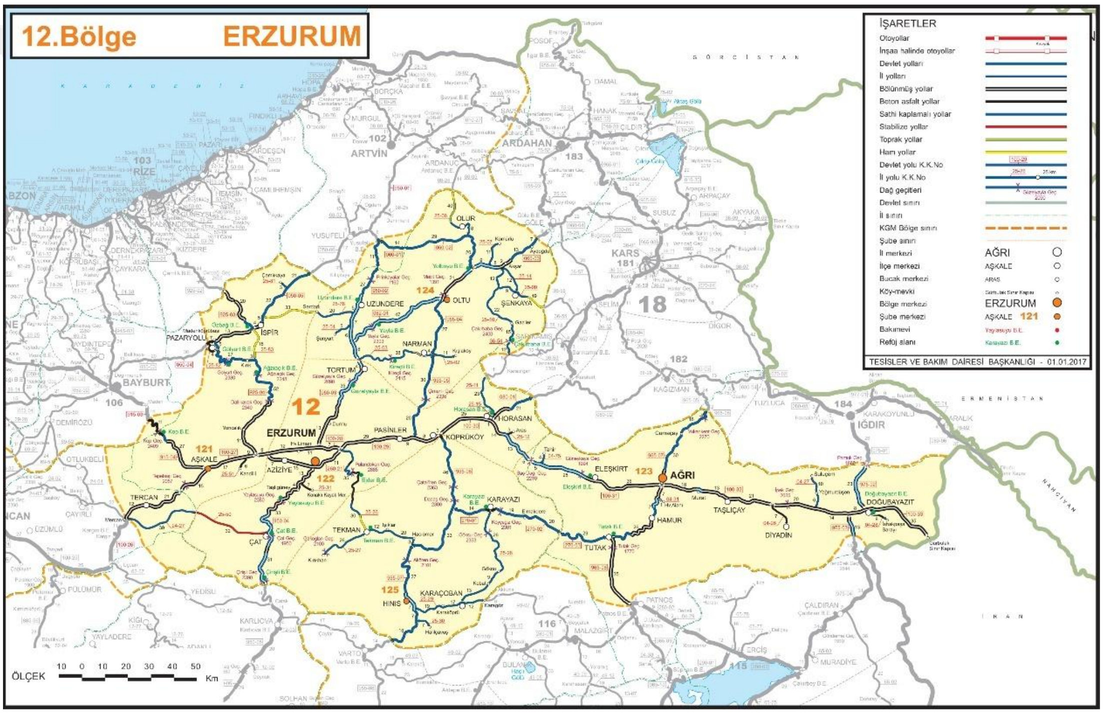

Figure 2.6. A screenshot from the highway map of Turkey, from the 12th region [27].

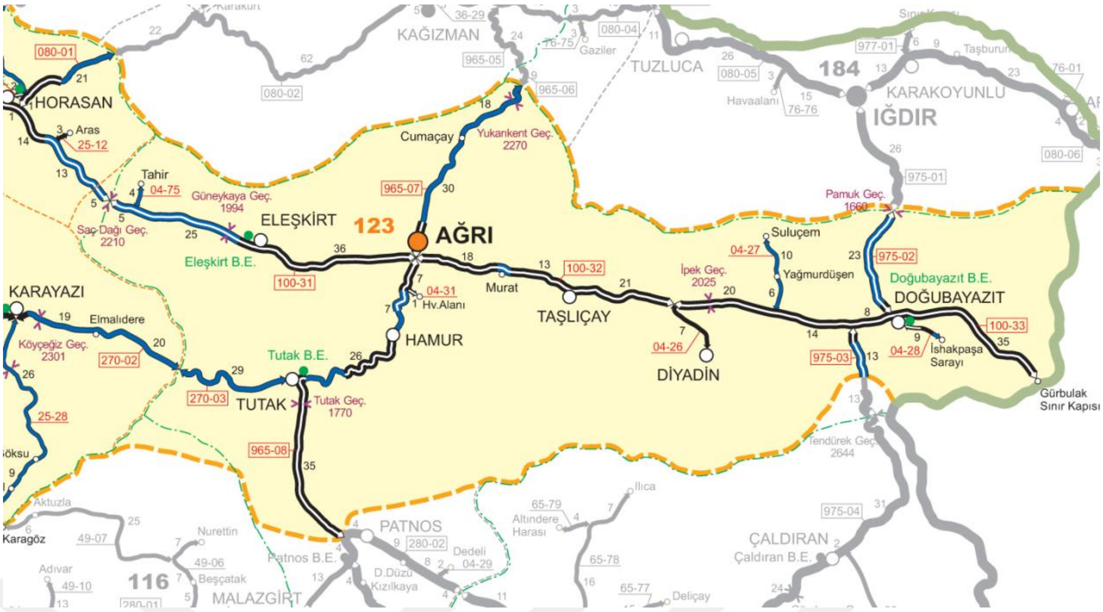

Figure 2.7. A magnified view of the 12th section of the KGM map [27].

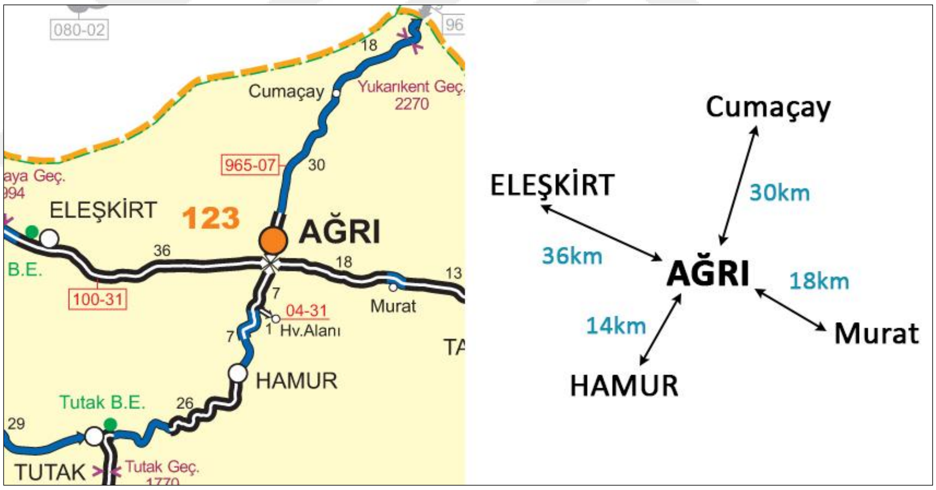

Figure 2.8. Evaluating the map to build the database [27].

According to the procedure illustrated in Fif. 2.7, we build a tree of cities. We have three types of cities:

1. City1: stands for the provenances.
2. City2: represents towns or districts.
3. City3: represents villages or small towns.

We build four tables as follow:

Table 1: contains cities-1.
Table 2: contains cities-2 within each of city-1.
Table 3: contains cities-3 within each of city-2.
Table 4: contains the links and distances between the cities, each city of (1-2-3) represented by its ID. Each city composed of city name and city ID. The IDs in the City3 table are used to represent the cities. Each city in City1 and City2 table has also reflections in City3 table (i.e., Ankara in City1 table, Ankara Merkez in City2 table and the same in City 3 table), so City3 table contains all the city variations.

2.4.2. Network Analysis

We performed network analysis in Gephi Software [3], which has several utilities that output network parameters, distributions, together with the network visualizations. We used the essential utilities in the statistics tab as average degree, average path length, average clustering coefficient, modularity, betweenness centrality, closeness centrality, eigenvector centrality, the distributions of these metrics, together with the degree distribution. We also constructed network visualizations in several aspects provided by GEPHI.

The graphical plots about the distributions are perfprmed in MATLAB since the plotting services in GEPHI are not sufficient by the means of logarithmic scales or customization of the plots. These processes of processing are also mentioned in the next section together with the results.

CHAPTER 3

RESULTS AND DISCUSSIONS

3.1. GEPHI

This section exhibits the results that we achieved. We performed network analysis by using a software called Gephi [3]. Gephi is a software for complex network analysis and visualization. It also has network visualization utilities affording to modify the properties of nodes and edges. That is to assist information experts with making theory, naturally find designs, confine structure singularities or shortcomings amid information sourcing. It is a correlative instrument to conventional insights, as visual speculation with intelligent interfaces is currently perceived to encourage thinking. This is a product for “Exploratory Data Analysis”, a paradigm showed up in the Visual Analytics field of research.

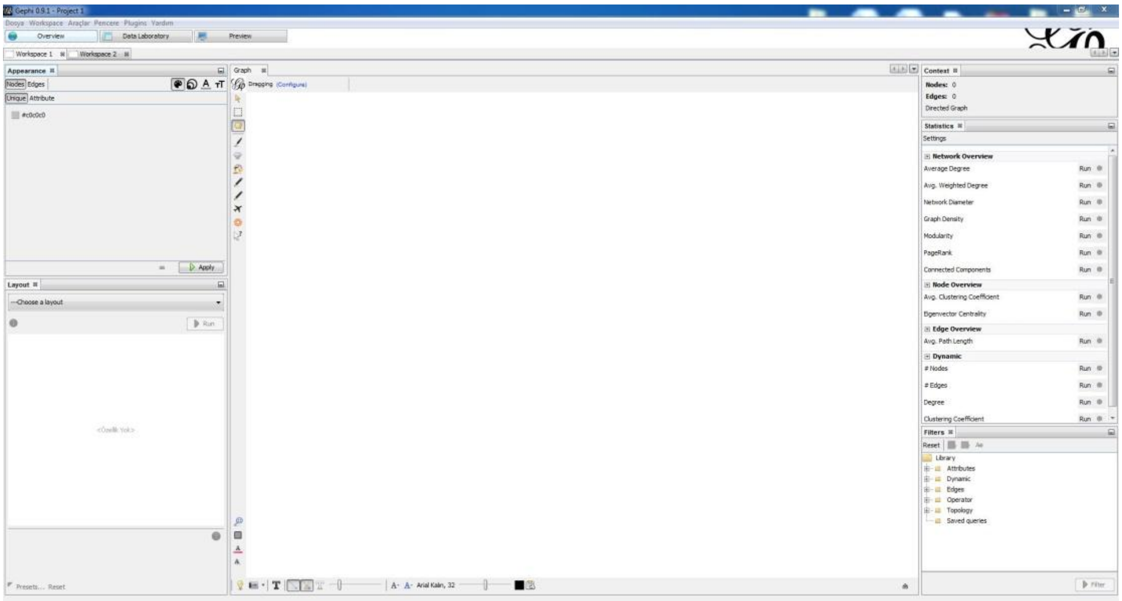

Figure 3.1. Gephi software interface.

Figure 3.1 shows the interface of Gephi software. The processes in Gephi software are realized under three main windows. These are the overview window, the data laboratory window and the preview window.

1. Overview window: This section provides general information about the complex network. Basic visualization settings are available in this section. Also, the layout algorithms for visualization and metric analysis are presented in this window. Generaly includes Partition, Ranking, Layouts, Context, Statistics and Filters sections.
2. Data Laboratory Window: This section contains the database of nodes and links. Adding new nodes and links to the network takes place in this section. The changes that need to be made in existing nodes are done in this section. It is possible to import the data from various external formats such as csv, gexf, or export it in csv format.
3. Preview Window: This section contains detailed visualization settings and network visuals. Changes to the image of the installed network and fine adjustments (size, color, shape, etc.) are made in this section.

3.2. GEPHI RESULTS

Our network is composed of 2543 nodes and 2818 undirected links between these nodes. The parametric and distributional results about the network are presented in the sections below.

3.2.1. Average Degree

Average degree represents the average number of connections per node. Gephi returned 2.221 as average degree, indicating that every node has connections slightly greater than 2, leading ~2 alternative transportation paths for each point. Average weighted degree is the same as this value since the network consists of unweighted links between nodes.

3.2.2. Network Diameter

Network diameter is the greatest of the shortest path length values in the whole network. Our network displays a high diameter of 83, similar with regular networks. This is mostly caused by the spatial dependencies of the paths, those are affected from the metric distances in km (i.e. two distant cities can not be linked with a short number of paths, since there are numerous cities between them).

3.2.3. Modularity

Modularity is a measure describing the ratio of a network to consist of separate modules. A module is described as a cluster that is strongly interconnected, but have sparse connections to the other modules. This parameter gets values between 0 and 1, which is close to 1 for modular networks. Our network displayed a distinct modularity measure of 0.918, indicating that the close cities are strongly interconnected with local roads, having sparse connections to the other modules that spatially locate in distant locations.

We also visualized our network with node sizes proportional with their modularities as in Fig. 3.2.

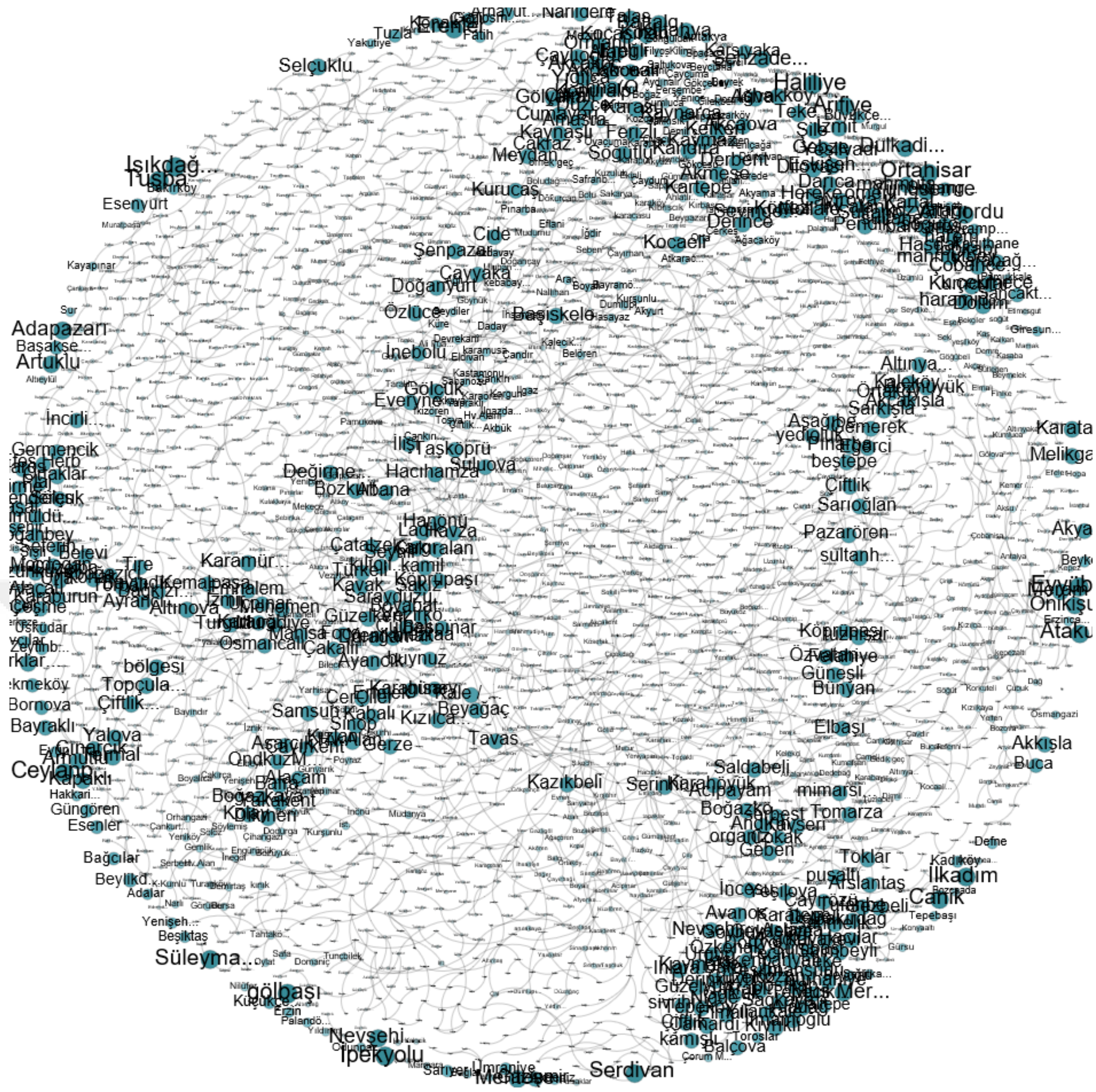

Figure 3.2. Network visualization regarding the modularity measures of the nodes linked with their sizes.

3.2.4. Average Clustering Coefficient

Clustering coefficient catches how much the neighbors of a given node are connected to each other. The range of this parameter lies between 0 and 1, where 0 means there is no links between neighbors of a node, whereas the value 1 means all the neighbors of a node are linked together.

Gephi returned an average clustering coefficient of 0.034 for our network, which means a small portion of the nodes' neighbors are linked, unlike the majority of the real networks. Real networks display high clustering property together with small path lengths. Our network deviates from this behavior by displaying high distances with low clustering. But high modularity property is in a good agreement with other real network types. We can conclude that the transportation network displays specific deviations from the real networks, which is mostly caused by the spatial dependencies.

3.2.5. Average Path Length

This parameter calculates the shortest paths between the nodes and then get the average average value for the whole network [6]. Our network resulted an average path length of 28.304. This is a high value close to regular networks, again originating from the spatial positioning of the nodes, which in turn result high separation between node pairs.

3.2.6. Betweenness Centrality

In previous chapter, we explained the concept of betweennes centrality which is can be defined simply by saying how many times can the node act as a bridge between two other nodes connected with shortest paths [6]. According to this definition, we generated Fig. 3.3 with different sizes of nodes, subject to the betweenness centrality values of them. The nodes with big sizes have higher BC values.

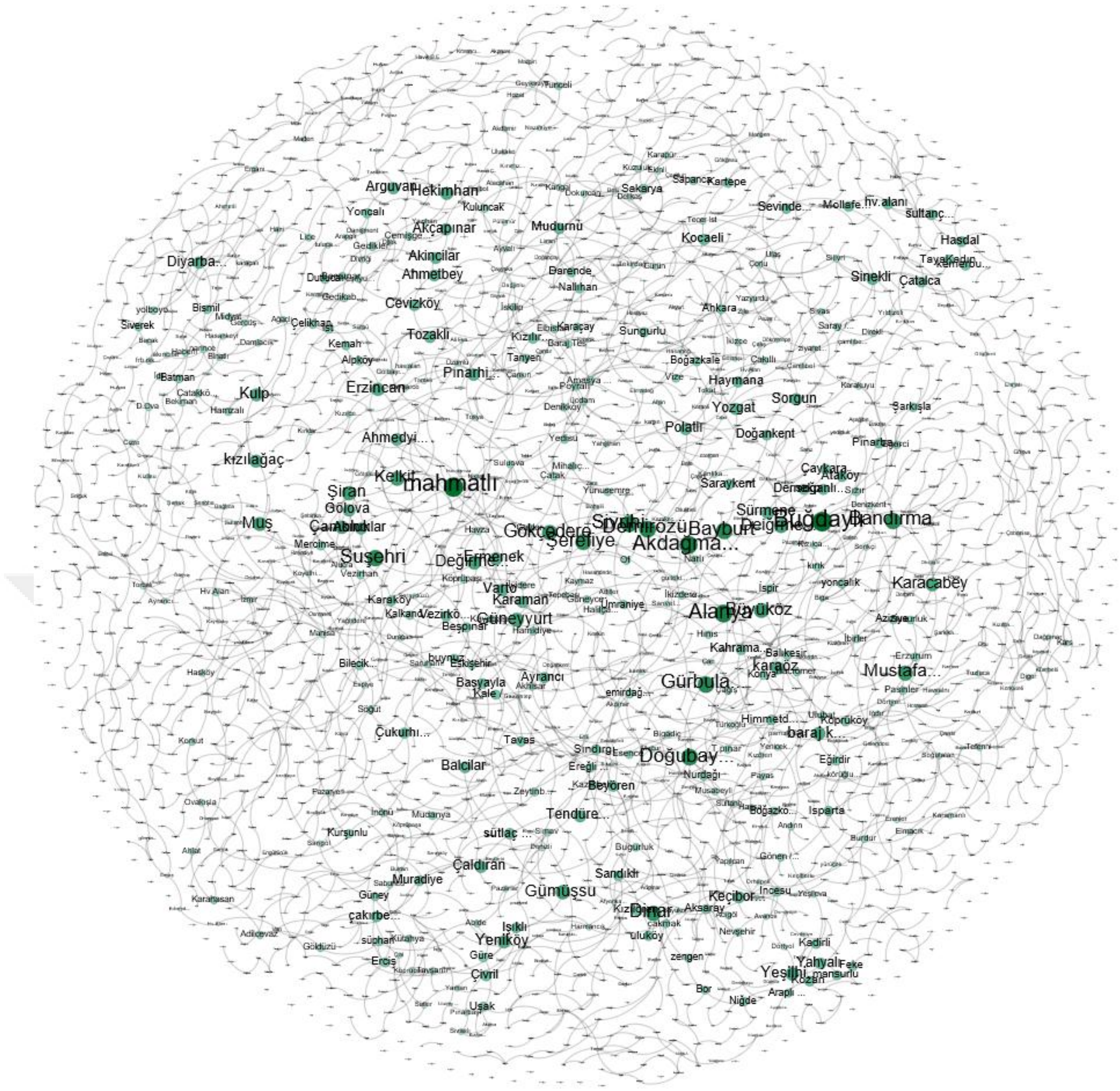

Figure 3.3. Network visualization regarding the betweenness centrality measures of the nodes linked with their sizes.

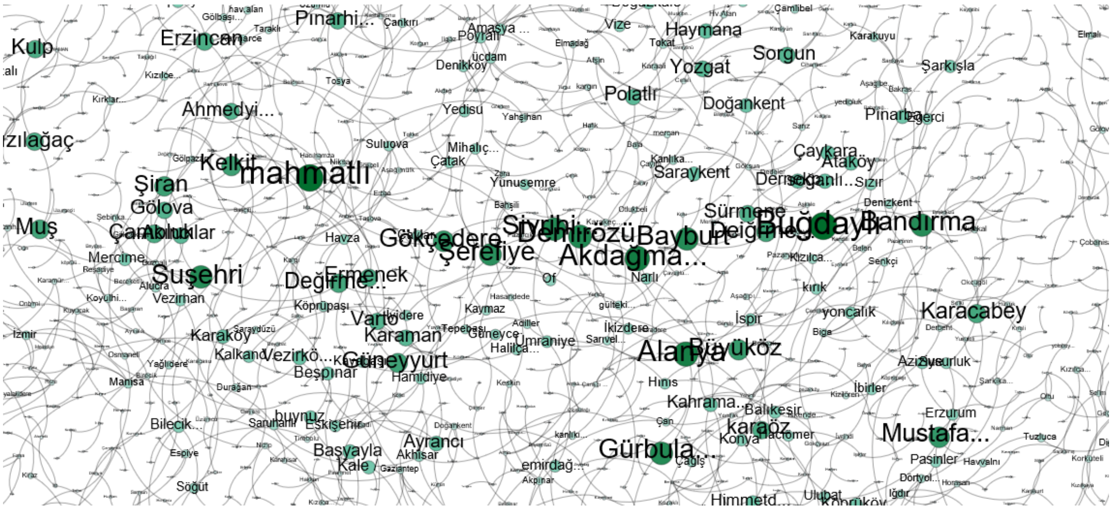

Figure 3.4. A small region of Fig. 3.3 magnified.

3.2.7. Eigenvector Centrality

This measure represents the importance of a node within a network. The nodes with high degree of importance have high influence on the network. The number of links doesn’t effect the importance of a node [6]. We can see in the figure below some nodes are bigger than the others, due to the high EC value that they have.

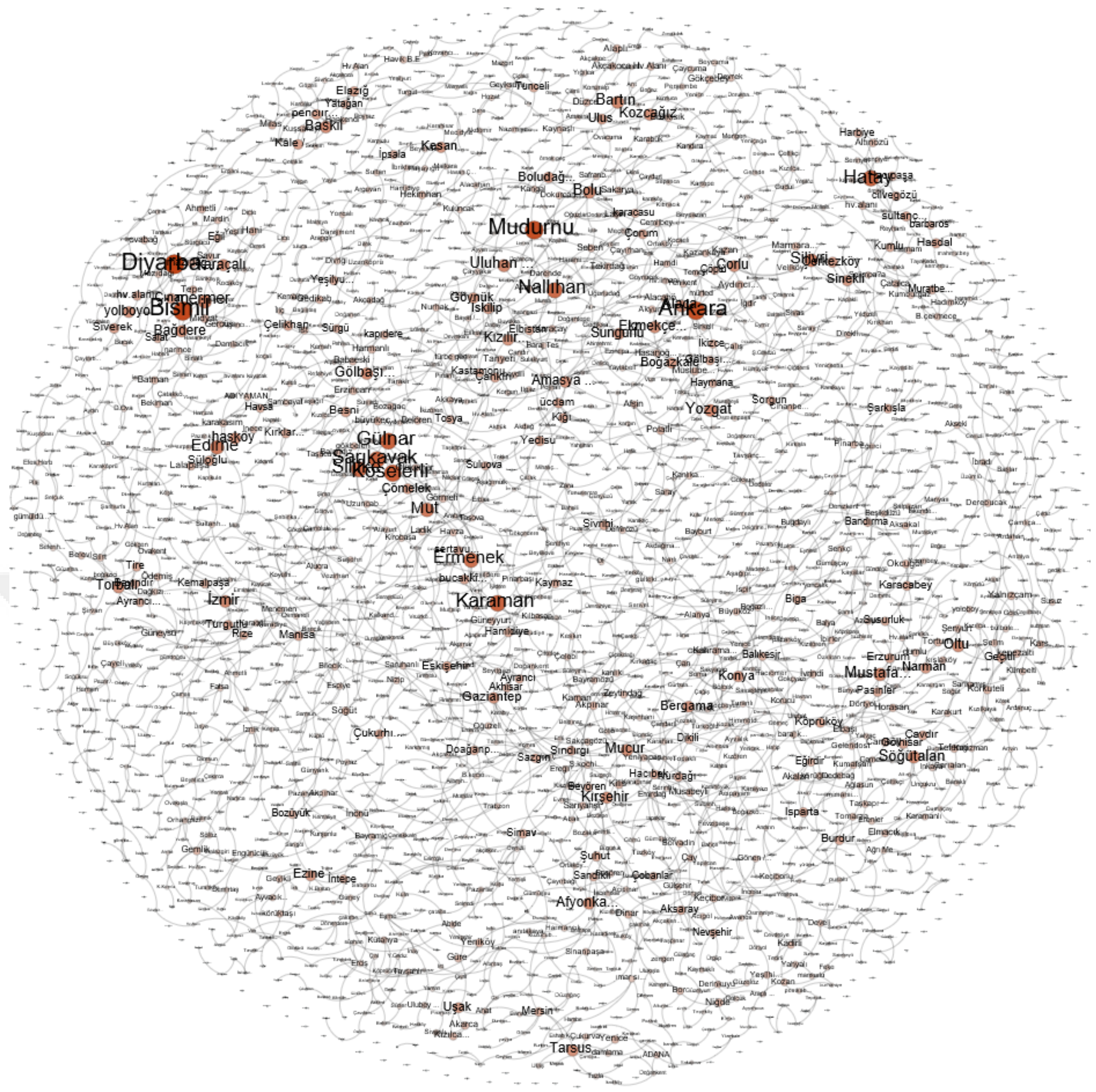

Figure 3.5. Network visualization regarding the eigenvector centrality measures of the nodes linked with their sizes.

3.3. MATLAB ANALYSIS

MATLAB is a powerful tool used in mathematical evaluations in many disciplines. Since Gephi has limited abilities to draw formatted graphs, we employ MATLAB to evaluate the following plots about network characterization.

3.3.1. Degree Distribution

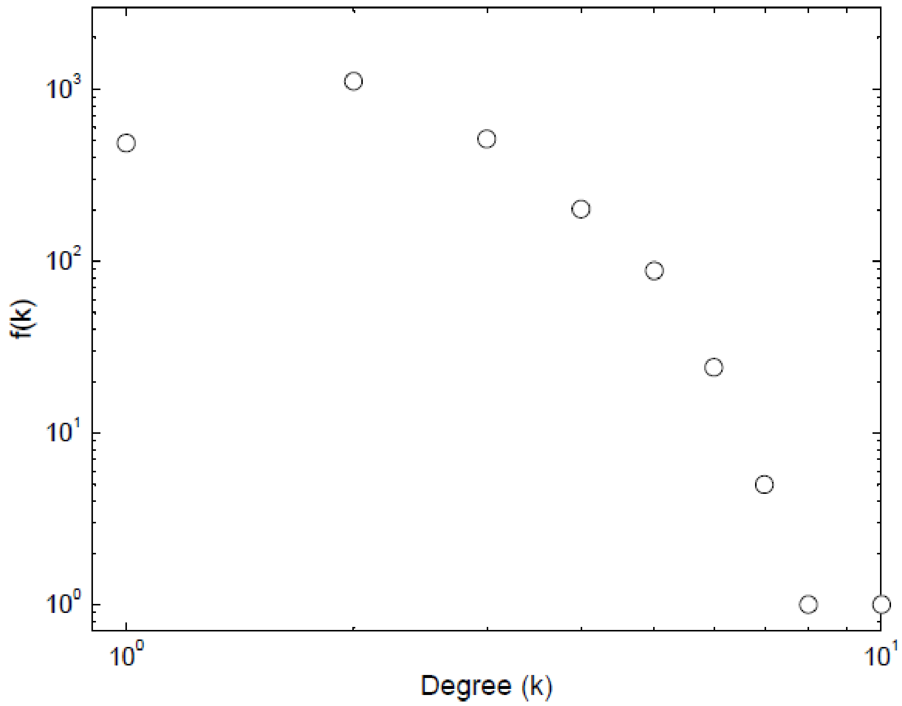

Figure 3.6. Frequencies of degree occurrences.

In Fig. 3.6, we present the frequencies of degree occurrences for all nodes in City-3 table. For example, we can conclude that there are about 500 cities that contain one connection, about 1000 cities that have 2 connections and so on. The cities with the highest number of connections (such as cities with 8 and 10 degrees) have lowest number of occurrences (we have about 1 city which has 8-10 connections).

Since this distribution has limited capabilities in network classification, we convert this plot to a “log- binned distribution”. This means we first get aggregate values for the data points in logarithmic scale, and then normalize these log-binned values to achieve a probability distribution as described in [25]. The resulting degree distribution plot is presented in Fig. 3.7.

Power-law consistency is best diagnosed in log-binned degree distributions. While a small portion of real networks display perfect power-law consistencies, the majority display left-side saturation regions together with right side cotoff regions. Our degree distribution graph display both of these saturations and cotoff regions together with a power-law consistent tail with a slope of -4.2. This region is a sufficient requirement to label the network as scale-free. But we can say that the power-law coefficient (4.2) is sensibally high compared to the other real networks. This type of power-law consistent tails with high coefficients are labeled as random regime of the scale-free networks by Barabasi [6].

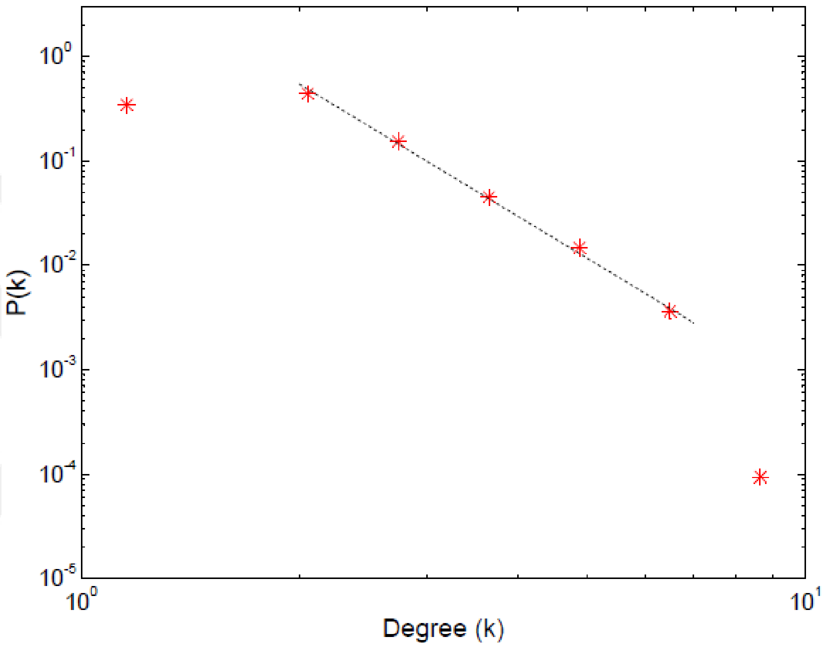

Figure 3.7. Degree distribution with logarithmic binning.

3.3.2. Closeness Centrality

We already explained the concept of closeness centrality in the previous chapter. The figure below shows the closeness centrality distribution of our network. We can say that there are two groups in of cities in our network in term of closeness centrality. The first three scatters indicate that there is a group of cities that show low closeness centrality with high probabilities and these cities have limited opportunities to reach other cities in Turkey. The second group show high values (close to one) with low probabilities, meaning there are small number of cities havinh high opportunities to reach other cities in short pathways. This is an expected situation since the small locations like villages are numerous and they have long hop distances to other locations. Contrary with this, there are small numbers of big cities that are located more centrally in the map, having greater transportation opportunities.

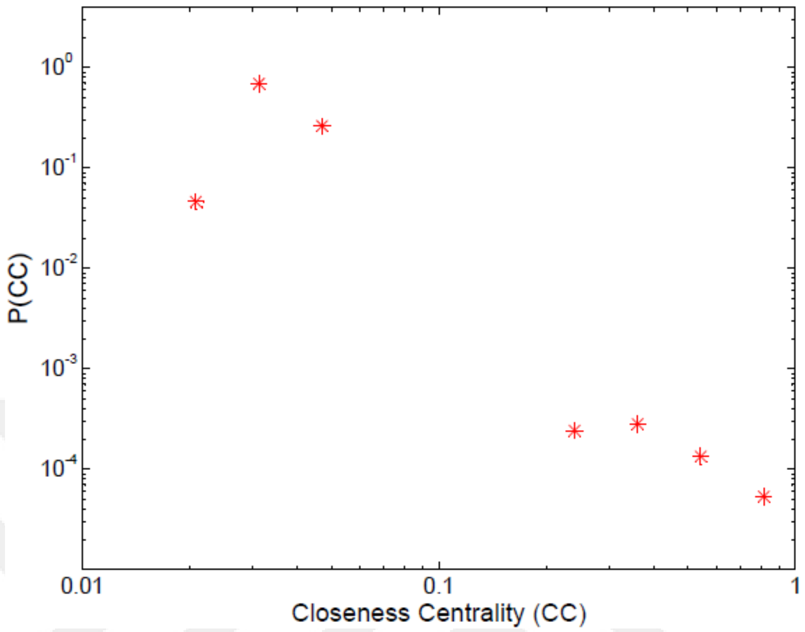

Figure 3.8. Distribution of closeness centrality.

3.3.3. Betweenness Centrality

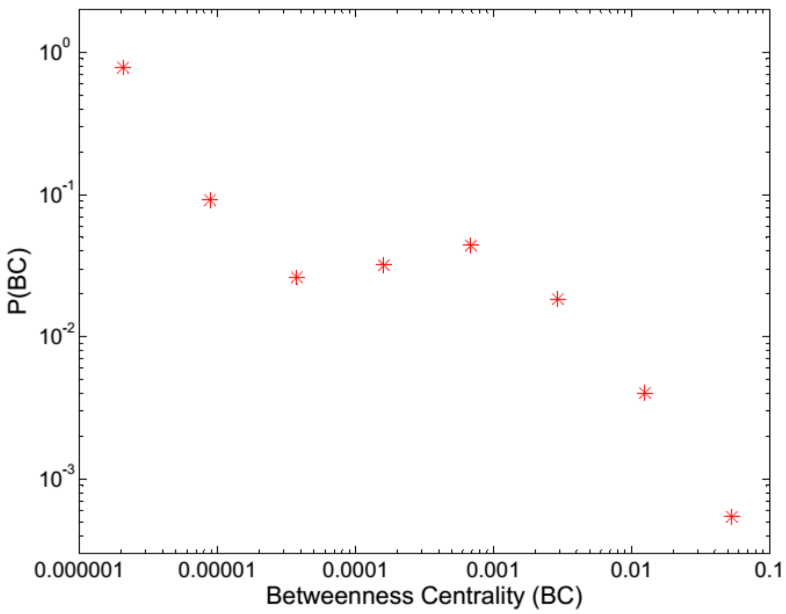

Figure 3.9. Distribution of betweenness centrality.

The betweenness centrality distribution of the network displays a two-region power-law consistency, indicating that there are small numbers of nodes which are located in the shortest pathways of the other nodes, while the majority of the nodes do not yield within these important pathways. The mid-range of the graph displays a different behavior, indicating a portion of nodes to have similar BC values, with similar importances by the means of highway transportation.

3.3.4. Eigenvector Centrality

According to the definition of this concept, eigenvector centrality (EC) increases with connections to important nodes. We present in Fig. 3.9 the log-binned version of eigenvector centrality distribution. This is a smoother distribution compared to CC, more like a bell curve. Unlike the CC distribution, the leftmost side of the graph does not display the highest values, the mid-range makes a peak instead. This indicates that the probability of having very low or high EC values are similar, while EC values around 0.1 are the most propbable ones.

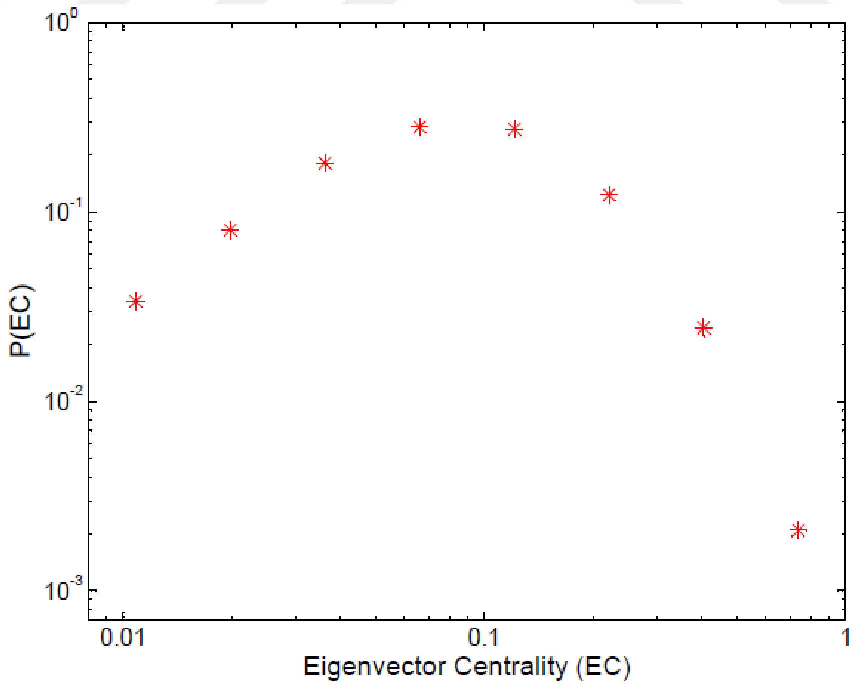

Figure 3.10. Distribution of eigenvector centrality.

3.3.5. Distance In Kilometers

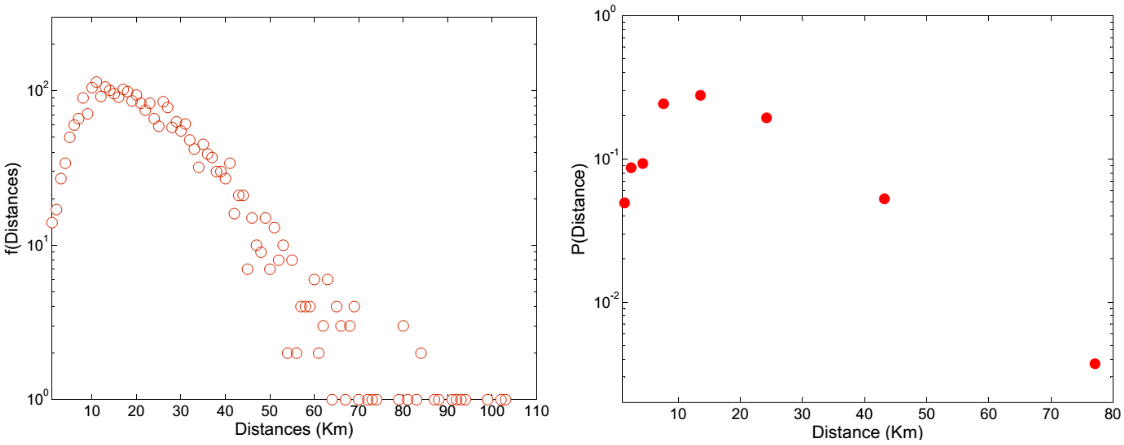

Figure 3.11. Distance_KM: a) Scattered, b) Log-binned distribution.

As mentioned in the methodology section, we have also captured the distances between the nodes in kilometers. The two plots in Fig. 3.10 represent the scattered and log-binned distributions of these distances in km respectively. The plots span the distances between 1 to ~100 kilometers.

The distance distribution plots indicate that the most probable distance between cities is about 10-20 kilometers. The lower distances have comparable occurrences with this peak values, while the higher distances close to 100 have rare occurrences. The distribution, plotted in a semi-logarithmic way, indicates an exponential distribution together with a left-side saturation region.

3.3.6. Modularity Class (MC)

The modularity distribution of the nodes is presented in Fig. 3.11, indicating that a decreasing trend dominates the distribution. The probability of a city to have a high modularity is inversely proportional with the modularity measure. But apart from this distribution, we must remind that the overall modularity is very high for this transportation network.

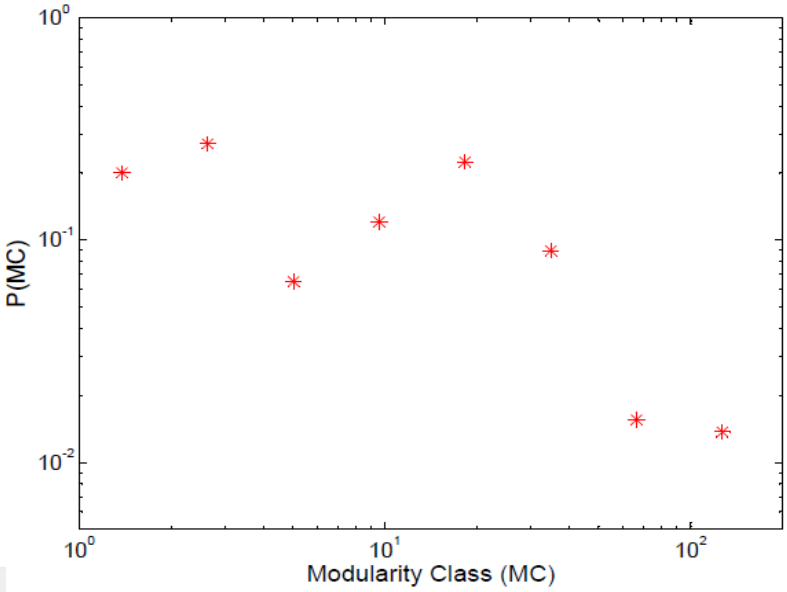

Figure 3.12. Modularity class distribution with log-binning.

3.4. DIMENSIONAL PERCENTILE DISTRIBUTION GRAPHS

In this section, we illustrated several graphs that derive two dimensional distributions of some parameters. These are percentile plots, indicating each axis as percents of the maximal value of it, and the intersections of the two axes are filled with grayscale colors which indicate exponential growth proportional with the color scales in the rightmost side of the plots.

3.4.1. Closeness Centrality Vs. Betweennees Centrality

We can notice from Fig. 3.12 that most of the nodes in our network have CC values between 3-5% according to the maximum CC value in the network, and the CC is stuck into the lower 5% of the CC Scala. On the other hand, the BC values, being mostly cumulated in the lower 50% band of the BC Scala, display a large variety from 0 to 83% with respect to the maximum BC value in the network.

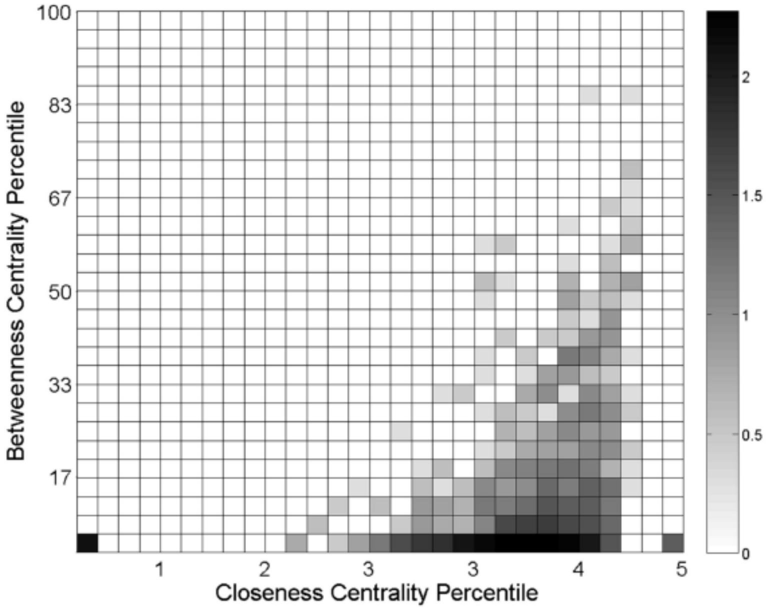

Figure 3.13. Closeness centrality vs. betweenanes centrality percentile.

Having very low CC values, means that, a randomly selected node is not close to the rest of the network, on average. In other words, when you pick a city from the network, it is not very close to the rest of the network. For instance, you may have to jump by 20 nodes to reach a target node. On the other hand, BC is relatively high. High BC means when you navigate from a random node to another, you have to pass from a specific node in a high rate. For example, to navigate from Ankara to Sinop, you have to pass from Karabük. And to navigate from Zonguldak to Kastamonu, you again have to pass from Karabük. This means that Karabük has high betweenness centrality. But it may not have closeness centrality because, it is not close to many cities in the Turkey map.

3.4.2. Eigenvector Centrality Vs. Betweenness Centrality

The eigenvector centrality vs. betweenness centrality occurrrences are illustrated in Fig. 3.13. The majority of the nodes in our network have EC values between 0-50 according to the maximum EC value in the network, thus the EC Scala is stuck into the lower 50% band of the EC Scala. On the other hand, the BC values, being mostly cumulated in the lower 50% band of the BC Scala, display a large variety from 0 to 83% with respect to the maximum BC value in the network. Having reasonable EC values, means that, there are a reasonable number of nodes within the network are important nodes.

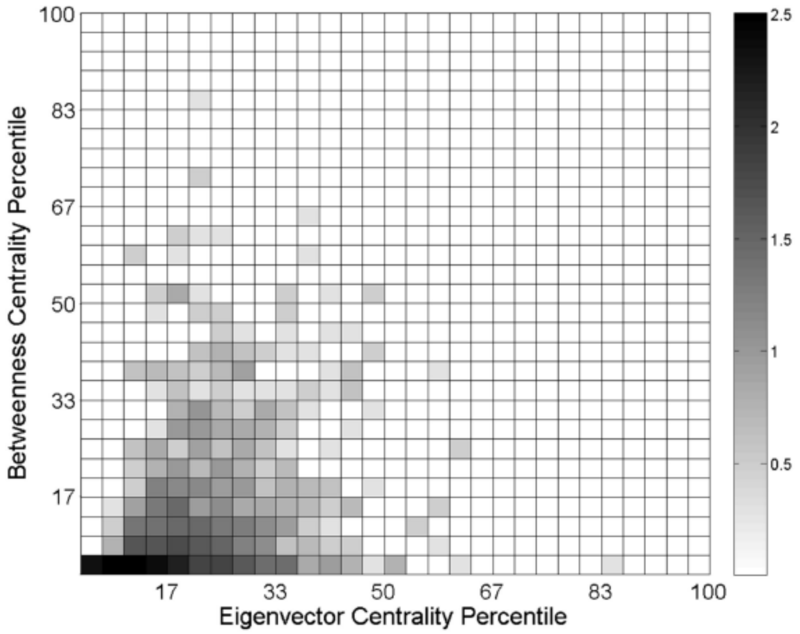

Figure 3.14. Eigenvector centrality and betweenness centrality percentile.

3.4.3. Modularity Class Vs. Betweenness Centrality

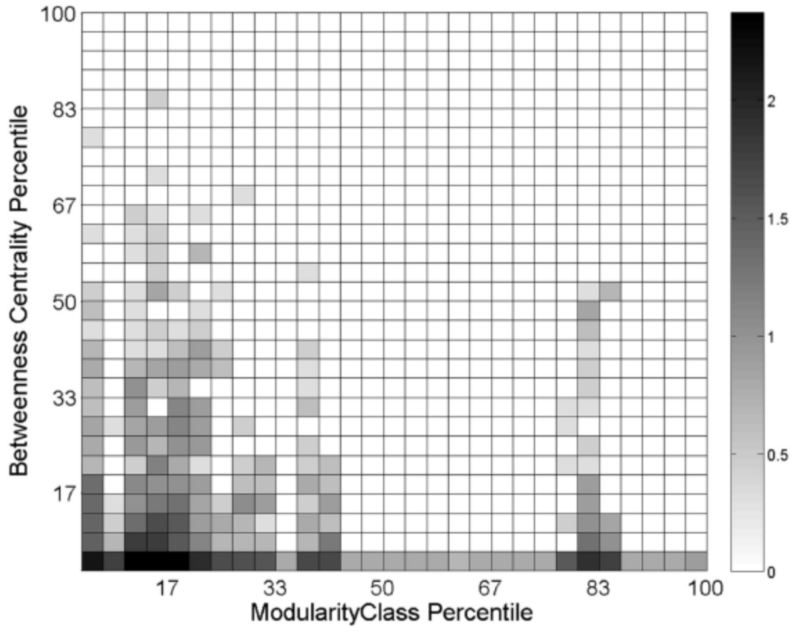

Figure 3.15. Modularity class betweenness centrality.

The modularity class vs. betweenness centrality occurrrences are illustrated in Fig. 3.14. We can conclude from this figure that MC values between 1-42 which means that most of our network's nodes have lower modularity than the nodes that have MC values between 77-83%. The MC Scala seems to be stuck into the lower 83% of the MC Scala. On the other hand, the BC values, being mostly cumulated in the lower 67% band of the BC Scala, display a large variety from 0 to 87% with respect to the maximum BC value in the network. Having two different regions of distribution means there is a group of nodes have high modularity whereas the majority of nodes display low modularity.

3.5. OUTLINING THE IMPORTANT NODES FOR TRANSPORTATION

As mentioned in the previous sections, the importance of a specific node can be described with the centrality measures. Having performed three centrality distributions for the whole network, we present the list of important locations in Turkish transportation network in Tables 3.1, 3.2 and 3.3 below.

Table 3.1. Top 20 cities according to eigenvector centrality.

City	Eigenvector Centrality
Diyarbakır	1
Bismil	0.9679
Mudurnu	0.9310
Ankara	0.9271
Köselerli	0.8531
Gülnar	0.8271
Karaman	0.8162
Sarıkavak	0.8107
Hatay	0.7967
Silifke	0.7798
Nallıhan	0.7764
Ermenek	0.7442
Mut	0.6918
Edirne	0.6858
Mermer	0.6538
Silivri	0.6474
Mustafakemalpaşa	0.6281
Uluhan	0.6179
Bartın	0.6115
Söğütalan	0.6028

Table 3.2. Top 20 cities according to betweenness centrality.

City	Betweenness Centrality
Buğdaylı	0.1098
Mahmatlı	0.1083
Akdağmadeni	0.0998
Alanya	0.0976
Demirözü	0.0923
Doğubayazıt	0.0913
Sivrihisar	0.0909
Gürbulak	0.0903
Bayburt	0.0901
Suşehri	0.0897
Bandırma	0.0893
Şerefiye	0.0871
Dİnar	0.0829
Gökçedere	0.0823
Mustafakemalpaşa	0.0819
Büyüköz	0.0801
Kelkit	0.0769
Şiran	0.0768
Gümüşsu	0.0760
Güneyyurt	0.0751

Table 3.3. Top 20 cities according to closeness centrality.

City	Closeness Centrality
Bademli	1
İstanbul	1
Ortacalar	1
Kalekoy	1
Kalecik	1
Gölbaşı	1
Oymapınar	0.8
Mutki	0.6667
Solhan	0.6667
Meydan	0.5714
Koyunlu	0.5714
Yenibaşak	0.5
Kömür liman	0.4210
Şirinköy	0.4
Barbaros	0.4
Kavakbaşı	0.4
Geyikpınar	0.4
Kuzulimanı	0.3478
Uğurlu	0.3478
Aydıncık	0.2285

CHAPTER 4

CONCLUSION

Studying the highway transportation network of Turkey with the tools of complexity, the first main proceeding of this thesis is extracting the nodes and edges data from the detailed KGM maps. This data will enable studies about this network in some other aspects in the future.

Analyzing the main network parameters, we see that the network has an average degree value slightly over 2, indicating that each location (city) has approximately two conjunctions to other cities. Since the endpoints of some paths, through the villages yield 1 connections, this average degree means that the other cities locating towards the inside of the network should have more than 2 or 3 connections to other cities.

The network yields a noteworthy average path value of 28.309, higher than the several types of real networks. Since the studies transportation network involves from crowded cities to small villages, the pathways connecting two distant villages is naturally expected to have numerous hops. Originating from the spatial structure of the network, the low-degree nodes can not directly connect to a popular node as in a social network. This fact increases the path lengths and network diameter. As expected, the diameter also has a great value of 83.

The spatial structure, combining with the cost of constructing main roads, results a very small average clustering coefficient of 0.034, like a Barabasi-Albert network without clustering property. In fact, there should be secondary roads between close locations but since KGM maps include only main roads, the real clustering between spatially close nodes does not come into the picture. On the other hand, the network displays a distinct modularity measure of 0.918. In combination with very low clustering, this is a very interesting result to expound. We predict that this behavior of very low clustering with very high modularity emerges from the consulting of the main roads which are sufficient to define modules, but insufficient to connect small but close locations.

The degree distribution has a power-law tail with a prominently high exponent of 4.2, together with left-side saturation region and right-side cutoff about $k \approx 7$. We can say that this is scale-free network but the small-worldness does not sufficiently emerge since the average path length values are very high.

A noteworthy outcome of this study is outputting the distance distribution in kilometers, thanks to the collection of the distance data manually from the maps. The distance distribution displays an exponential tail, together with a left-side saturation regime. The most probable distances between the locations seem to major on 10 to 20 kilometers.

The closeness centrality distribution of the network exposes two distinct regions, one consisting of lower values with higher probability, other consisting of higher values with low probability, which in combination exhibit a power-law behavior. The betweenness centrality distribution also displays power-law decay with a fluctuation in the mid-range, which indicates a portion of the network has similar BC values and similar importance by the means of locating among important paths. The reader can also extract detailed information from the two-dimensional plots of centralities which confront the combinations of centrality measures with ocurrance percentiles.

As a result, we can conclude that the transportation network of Turkey exhibits some similar characteristics with the majority of real networks, together with some unique properties only consistent with a small portion of recent studies. We can say that these distinct properties of low clustering and high average path length and diameter mostly depend on the spatial dependencies of the network structure which in turn avoid direct connectons of low-degree nodes to hubs. Low clustering with high modularity is also supposed to be originating from the extraction of the network only regarding with the main roads which are included by KGM maps.

This study provides a data source for not only complex network researchers but also some other disciplines which involve in transportation, navigation, logistics etc. The collected data in nodes and edges approach together with distance data smooth the way for performing navigation-based studies in the future.

REFERENCES

1. Scott, J. P., “Social Network Analysis: A Handbook”, ***SAGE Publications***, 1, 16 (2000).

2. Ottino, J. M., “Complex systems”, ***AIChE J***, 49: 292–299 (2003).

3. Standish, R. K., “Concept and Definition of Complexity, in Intelligent Complex Adaptive Systems”, Ed. by A. Yang, Y. Shan, ***IGI Global***, Hershey, 105–124 (2008).

4. Sevtsuk, A., and Mekonnen, M., "Urban network analysis: A new toolbox for ArcGIS”, ***Revue internationale de géomatique***, 22 (2): 287-305 (2016).

5. Newman, M. E. J., “The structure and function of complex networks”, ***SIAM Rev.,*** 45 (2): 167–256 (2003).

6. Internet: Barabási, A., “Network Science by Albert-László Barabási”, **http://barabasi.com/networksciencebook** (2017).

7. Bavelas, A., "Mathematical model for group structures", ***Hum. Organ,*** 7, 16-30 (1948).

8. Scott, J., “Social Network Analysis: A Handbook”, 2nd ed., ***Sage***, London, (2000).

9. Wasserman, S. and Faust, K., “Social Network Analysis", ***Cambridge University Press,*** Cambridge, UK, (1994).

10. Moreno, J. L., “Who Shall Survive?”, Beacon House, ***Beacon***, NY, (1934).

11. Rapoportand, A. and Horvath, W. J., “A study of a large sociogram”, ***Behavioral Sci***., 6: 279–291 (1961).

12. Padgett, J. F. and Ansell, C. K., “Robust action and the rise of the Medici, 1400–1434, ***Amer. J. Social.,*** 98: 1259–1319 (1993).

13. Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, and Wiener, J., “Graph structure in the web”, ***Computer Networks***, 33: 309–320 (2000).

14. Dorogovtsev, S. N. and Mendes, J. F. F., “Evolution of networks”, ***Oxford University Press,*** UK (2003).

15. Anghel, M., Werley, K. A. and Motter, A. E., “Stochastic model for power grid dynamics”, ***40th Hawaii International Conference on System Sciences***, US (2007).

16. Bagler, G., “Analysis of the airport network of India as a complex weighted network,” ***Physica A***, 387 (12): 2972–2980 (2008).

17. Barrat, A., Barthelemy, M., Pastor-Satorras, R. and Vespignani, A., “The architecture of complex weighted networks”, ***Proc. Natl. Acad. Sci***., 101: 3747–3752 (2004).

18. Mohmand, Y. and Wang, A., “Weighted complex network analysis of Pakistan highways”, ***Discrete Dynamics in Nature and Society***, 1-5 (2013).

19. Mukherjee, S., “Statistical analysis of the road network of India”. ***Pramana - J Phys***, 79 (3): 483-491 (2012).

20. Haibin, C., “Complex network analysis of highway network in Guangdong Province in China”, ***Canadian Social Science***, 2016: 1-4 (2016).

21. Jung, W. S., Wang, F., and EugeneStanley, H., "Gravity model in the Korean highway", ***Europhys. Lett***., 81 (4): 48005 (2008).

22. Albert, R., and Barabasi, A. L., “Statistical mechanics of complex networks”, ***Reviews of Modern Physics,*** 74 (2002).

23. Haznagy, A., Fi, I., London, A. and Nemeth, T., “Complex network analysis of public transportation networks: A comprehensive study”, ***Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2015 International Conference***, Hungary (2015).

24. Bastian, M., Heymann, S., and Jacomy, M. “Gephi: an open source software for exploring and manipulating networks”, ***International AAAI Conference on Weblogs and Social Media***, San Jose, California (2009).

25. Clauset, A., Shalizi, C. R. and Newman, M. E. J., "Power-law distributions in empirical data", ***SIAM Rev.,*** 51: 661 (2009).

26. Martinez, N. D., “Constant connectance in community food webs”, ***American Naturalist*** 139: 1208–1218 (1992).

27. Internet:” Karayolları Genel Müdürlüğü Bilgi İşlem Dairesi, "Haritalar”, **http://www.kgm.gov.tr/Sayfalar/KGM/SiteTr/Root/Haritalar.aspx** (2017).

28. Freeman, L. C. “Centrality in social networks conceptual clarication”, ***Social Networks,*** 1 (3): 215-239 (1979).

29. Barabási, A., Ravasz, E., and Oltvai, Z. “Hierarchical organization of modularity in complex networks”, ***Lecture Notes in Physics***, 46-65 (2003).

RESUME

Sabah Basher Salem Rashed was borned in Libya in 1984. She studied computer engineering in Libya and graduated in 2005-2006. She worked for 5 years as a teacher. Then she moved to Turkey after winning a scholarship from the Libyan government.

ADDRESS INFORMATION

Address : Karabuk University
Department of Computer Engineering
Demir Çelik Campus / KARABÜK
Tel : (545) 726 9115
E-mail : sabahrashed1984@gmail.com