STATISTICAL ANALYSIS OF BUS TRANSPORTATION NETWORKS FOR FOUR LARGEST CITIES IN TURKEY

VEYSEL ÇOBAN

B.S., Industrial Engineering, Işık University, 2008

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

IŞIK UNIVERSITY

2011

IŞIK UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

STATISTICAL ANALYSIS OF BUS TRANSPORTATION
NETWORKS FOR FOUR LARGEST CITIES
IN TURKEY

VEYSEL ÇOBAN

APPROVED BY:

Assist. Prof. Tankut Atan (Thesis Supervisor)	(Işık University)	_______________________
Assist. Prof. Selçuk Savaş (Thesis Supervisor)	(Işık University)	_______________________
Prof. Ercan Solak	(Işık University)	_______________________

APPROVAL DATE: 01/06/2011

STATISTICAL ANALYSIS OF BUS TRANSPORTATION NETWORKS FOR FOUR LARGEST CITIES IN TURKEY

Abstract

Network properties for public bus transportation networks (BTNs) of four largest cities (İstanbul, İzmir, Ankara, Bursa) in Turkey have been analyzed under C-, L- and P-Space topologies. Node degree distributions are shown to follow power-law distribution indicating scale-free nature of the networks. Furthermore, the networks have been growing preferentially with the exception of Bursa. Degree-degree correlations of nodes were studied in assortativity. The networks have small shortest-path values. For example, in İstanbul the maximum shortest path was found to be 4 in C-Space meaning that one needs at most 4 transfers to reach any point in the network. Correlation between mean shortest path lengths and nodes of given degrees of single end nodes are shown to follow power-law in L-Space. Degree, eigenvector, closeness, and betweenness centralities of nodes were calculated, and correlation between mean centrality values and degree were well described by power-law. Thus, in summary, BTNs show small-world and scale-free properties. We also investigated the impacts of two recent projects, Metrobüs and the Third Bosphorus Bridge, on bus network properties in İstanbul.

TÜRKİYE’NİN DÖRT BÜYÜK ŞEHRİNİN OTOBÜSLÜ ULAŞIM HATTININ İSTATİSTİKSEL ANALİZİ

Özet

Türkiye’nin dört büyük şehrinin (İstanbul, İzmir, Ankara, Bursa) otobüs ulaşım ağlarının özellikleri C-, L-, ve P- Uzay topolojileri altında incelendi. Düğüm derece dağılımlarının kuvvet yasası dağılımını takip ettiği ve ağın ölçeksiz (scale-free) bir yapıya sahip olduğu gösterildi. Ayrıca Bursa hariç diğer üç şehrin otobüs ulaşım ağlarındaki büyümenin rassal değil tercihli olduğu gözlendi. Düğümlerin dereceleri üzerinden bağlanma eğilimleri derece-derece ilişkilendirme (assortativity) başlığı altında incelendi. Dört büyük şehirde de herhangi iki nokta arasındaki en kısa ulaşım mesafelerinin düşük değerler olduğu ölçüldü. Örneğin, İstanbul C-Uzay topolojisi incelendiğinde bulunan en büyük kısa mesafenin dört olması şehir içinde en fazla dört vesait kullanarak seyahat edilebileceği anlamına gelmektedir. L-Uzay topolojisinde tek düğümlerin ortalama en kısa ulaşım değerleri ve dereceleri ilintilenmiş ve dağılımın kuvvet yasasına uyduğu gösterilmiştir. Düğümlerin derece, özvektör, yakınlık ve arada bulunma merkezilikleri hesaplanmış ve düğüm dereceleri ile olan ilitilenme kuvvet yasası ile tanımlanmıştır. Özetle, Türkiye’nin dört büyük şehrinin otobüs ağlarının küçük dünya (small-world) ve ölçeksiz ağ özelliklerini yansıttığı gösterilmiştir. Ayrıca, Metrobüs hatlarının ve Üçüncü Boğaz Köprüsü’nün İstanbul otobüs ulaşım ağının özelliklerine olan etkileri de istatiksel olarak incelenmiştir.

Table of Contents

List of Tables

List of Figures

List of Symbols

A	Adjacency matrix
$\mathbf{a_{ij}}$	Elements of adjacency matrix - there is an edge between i and j in the graph $a_{ij} = 1$, $a_{ij} = 0$ otherwise
$\boldsymbol{C}$	Centrality - importance or popularity of a vertex and edge in a network
$\mathbf{C_i}$	Local clustering coefficient of node i
$\mathbf{C_r}$ **or** $\mathbf{C_{ER}}$	Clustering value of random networks
$\langle \boldsymbol{C} \rangle$	Mean clustering coefficient - mean probability that two vertices are connected indirectly via one common neighbor vertex for all vertices
$\boldsymbol{C^{norm}}$	Ratio between mean clustering coefficient and Erdös-Rényi random graph of equal size – $C^{norm} = \langle C \rangle / \langle C_{ER} \rangle$
d	Distance of a network - maximum shortest path length between any two vertices
E	Total number of edges in network
$\mathbf{E_i}$	Edges among first neighbors of node i
F	Mean number of bus stations per route
$\boldsymbol{k}$	Node degree – number of edges connected to a vertex
$\mathbf{k_i}$	Degree of a selected vertex i
$\langle \boldsymbol{k} \rangle$	Mean node degree – average number of edges connected to a vertex
$\boldsymbol{k^{max}}$	Maximum node degree – maximal node degree in network
$\bar{\mathbf{k}}$	Exponent value of exponential distribution - $p(k) \sim e^{(-k/\bar{k})}$
$\mathbf{l_{ij}}$	Shortest path length from vertex i to vertex j
$\langle \boldsymbol{l} \rangle$	Average shortest path length - average number of stops to travel from one vertex to another
$\boldsymbol{l^{max}}$	Maximum shortest path length - maximum number of stops to travel between any two stations
N	Number of vertices in network

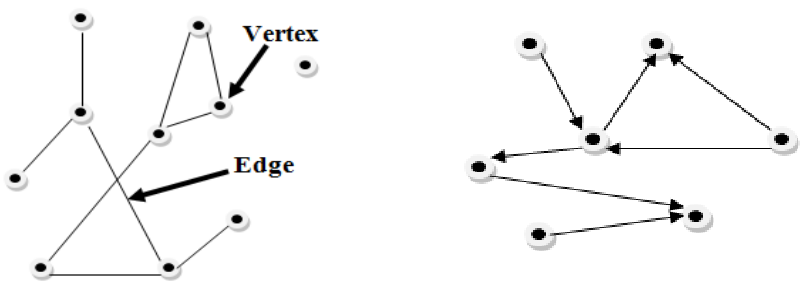
$\boldsymbol{r}$	Pearson correlation coefficient – correlation measure of the strength of linear dependence between two variables $(-1 \leq R \leq +1)$
$\mathbf{r^1}$	Assortativity mixing of nearest neighbor - connectivity tendency between vertices according to their degrees
$\mathbf{r^2}$	Assortativity mixing of next nearest neighbor
P	Population of cities
R	Number of bus routes
$\boldsymbol{z}$	Number of next nearest neighbors of a vertex – mean value denotes as $\langle z \rangle$
$\boldsymbol{\delta}$	Rate between $\langle z \rangle$ and $\langle k \rangle$ ($\delta = \langle z \rangle / \langle k \rangle$)
$\boldsymbol{\gamma}$	Exponent value of power-law distribution - $p(k) \sim k^{-\gamma}$
$\widehat{\boldsymbol{\alpha_k}}$	Maximum likelihood estimate of Hill plot

Chapter1

Network History

1.1 Introduction to Networks

A network (graph) is a set of points (nodes, vertices) that are connected by communication paths (edges, arcs) (Figure 1.1). Points in a network can be of different nature. For example, in social networks they can represent nationality, income, gender, locations, or ages. Moreover, the edges of a network can have different types and they can have different properties (Table 1.1). Edges can be weighted, unweighted, directed (pointing in only one direction -directed graphs (digraphs)), and undirected.



a **b**

Figure 1.1 Undirected and directed network samples. **(a)** Undirected network and **(b)** directed network.

Networks are real-life webs (social, information, technology, biological, physical science, etc.) and their complex structure and the behavior of their points can be analyzed within different fields as mathematics, physics, computer science, sociology, and biology. Networks can be further classified as natural and man-made. Natural networks are without a spider meaning that there isn't a central point that regulates the growth of the network (Figure 1.2). Rather they self-organize, and grow in a decentralized manner. The analysis of these special structures of complex real-world networks created a new research field –network science- and similar architectures were revealed in many real-world networks that differ from random world idea.[1] Network science is a field trying to develop and understand the theoretical and practical structure of natural and artificial networks [1].

Table 1.1 Complex real-world networks categories and examples with their vertices and edges.

NETWORKS	SAMPLES
Social	• Friendship (people - friendship relation) • Business (companies - business dealings) • Movies (actors - collaboration) • Science (scientist - research) • Phone calls (number - cells) • Spread of disease (people - sick)
Information	• Citation (paper - cited) • WWW (html pages – URL links)
Biological	• Food (predator - prey) • Neural (neurons - axons) • Genetic (proteins - dependence)
Technology	• Power grid (power station - lines) • Internet (routers – physical links)
Transportation	• Airline (airports - flight) • Railway (stations - railways) • Public transport (stations - roads)

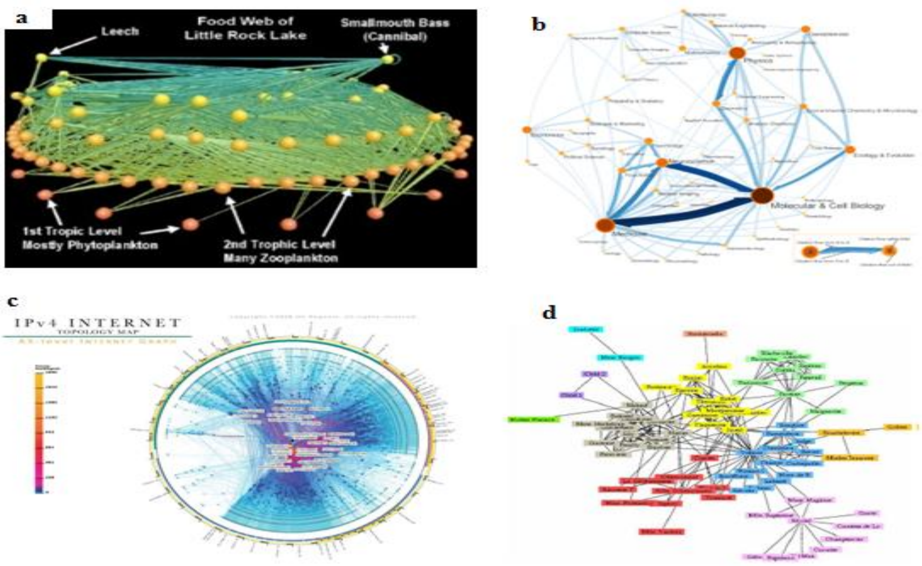


Figure 1.2 Complex real world network samples **(a)** food web **(b)** science and social science citation network **(c)** internet **(d)** literature network.

---

1 “God may not play dice with the universe, but something strange is going on with the prime numbers.” **Paul Erdős**
"God does not play dice with the universe." **Albert Einstein**

The study of networks has a long history in mathematics and the sciences going back as far as 1736 when Leonard Euler worked on the infamous “Königsberg Bridge Mathematical Problem” that asks “Are there any way that could be walk across all seven bridges and never cross the same one twice?”

Recent developments in graph theory brought a new perspective on the complex networks and helped to clarify, simplify their analysis and take applications into engineering, computer science, operation research, and sociology. Some research aims are to;

- Identify and measure the statistical properties of network (path lengths and degree distributions) in order to characterize the structure and behavior of network,
- Model the networks in order to understand the meaning of statistical properties and behaviors of network elements,
- Predict the behavior of networked systems that are based on measured structural and individual properties.

Empirical studies of real-world graphs created new terminology such as path length, clique (clustering), centrality, and connected components.

Moreover, helped by computers, large size real-network databases like WWW, citation network, internet, and social networks could be analyzed. Such studies of real-world networks yielded new and important contributions such as the revelation of the “small-world model” by Watts and Strogatz [2] and power-law degree distribution in networks and preferential attachment model of a growing network by Barabási and Albert [3].

1.2 Network Models

In this section, several theoretical models for network structure will be discussed.

1.2.1 Regular Graph Model

A network where each vertex has the same number of neighbors (the same degree) is called a “regular network”. A regular graph whose vertices have k degree is called a k-regular graph or regular graph of degree k (Figure 1.3).

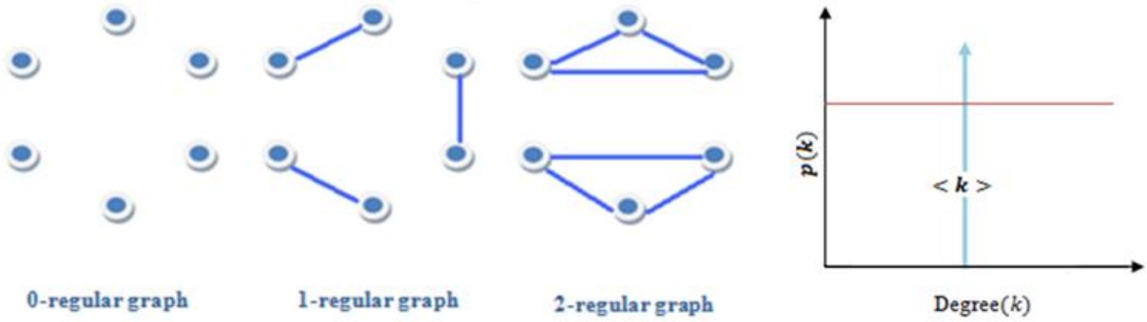


Figure 1.3 Sample regular networks and node degree distribution.

Regular networks have an artificial structure. They have a fixed node degree and high average degree, high clustering and high average path length.

1.2.2 Random Graph Model (Erdös-Rényi)

1.2.2.1 Introduction to Random Graphs

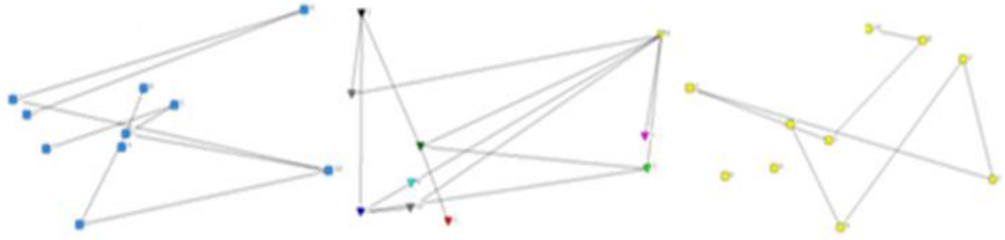
Simplest and oldest network model is random graph (Figure 1.4) which was firstly studied in [4, 5] and extended by the Hungarian mathematicians Paul Erdös and Alfréd Rényi within serial papers. Solomonoff and Rapoport [5] and Paul Erdös and Alfréd Rényi [6] independently described the simplest network model as N vertices that are connected as a pair with probability p ($G(N,p)$). Erdös and Rényi random graph model is also named as “Bernoulli graph” or “Poisson random graph” due to its node degree distribution structure. Erdös and Rényi also utilized another model that uses edges ($G(N,m)$) in their papers.

$G(N,m)$ graphs start with N vertices and exactly m undirected edges. A fixed number of unconnected vertices are connected with undirected m edges where vertices are chosen randomly from a uniform distribution. Degree distributions of random graphs are binomial and Poisson (for large size networks).

Another Erdös-Rényi random graph model is a $G(N,p)$ that starts with N isolated vertices and their pairs are connected with probability p. In this model a graph has maximum possible $\binom{N}{2} = \frac{N(N-1)}{2}$ undirected edges (when p=1), the probability of appearance m edges is $p^m(1-p)^{\left(\frac{1}{2}N(N-1)\right)-m}$ within $\left(\binom{N}{2} \atop m\right)$ different possible edges. Number of edges is defined in $1 \ll m \ll \binom{N}{2}$ and expected (average) number of edges is $\langle m \rangle = p\frac{N(N-1)}{2}$ for random graph models. Also, number of edges in unconnected random graph is $m = 0$ and $m = \binom{N}{2}$ for complete random graph (fully connected).

Procedure of building $G(N, p)$ graph;

- Start with N isolated vertices,
- Connect each of the $\binom{N}{2}$ unconnected vertex pairs with probability p (randomly chosen vertex pairs different from each other).



a **b** **c**

Figure 1.4 Sample random networks.

1.2.2.2 Degree Distribution

Degree distribution of the random graph $G(N, p)$ is a binomial which gives the connection probability of a randomly chosen vertex being connected with k other vertices and also "k_i" reflects the degree of a selected vertex i.

Binomial degree distribution (Figure 1.5):

$$P(k; p, N) = p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{1.1}$$

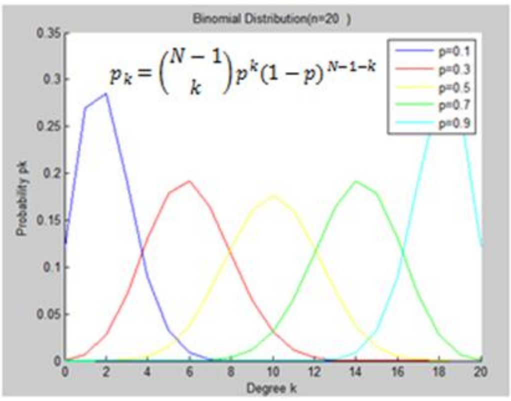


Figure 1.5 Binomial degree distribution of sample random graphs with given p values.

Probability of randomly selected vertex in network will have k neighbors (connected with different k other vertices) in one step. If N goes infinity and p goes to zero degree distribution transform from binomial distribution to Poisson distribution. Poisson degree distribution (Figure 1.6):

$$\lim_{N\to\infty} p_k = \lim_{N\to\infty}\binom{N}{k}p^k(1-p)^{N-k} \quad (1.2)$$

$$= \lim_{N\to\infty}\frac{N!}{(N-k)!k!}(\langle k\rangle/N)^k(1-\langle k\rangle/N)^{N-k}$$

$$p_k = \frac{\langle k\rangle^k e^{-\langle k\rangle}}{k!} \quad (1.3)$$

$$\langle k\rangle = p(N-1) = pN(\text{for large N}) \quad (1.4)$$

Mean node degree ($\langle k\rangle$) of a large graph size N is $\langle k\rangle = \frac{2E}{N} = p(N-1) \approx pN$ constant, $\langle k\rangle$ is also expected number of first neighbors of a random selected vertex ($N^1 = \langle k\rangle$) and number of second neighbors is defined as $N^2 = \langle k\rangle^2$.

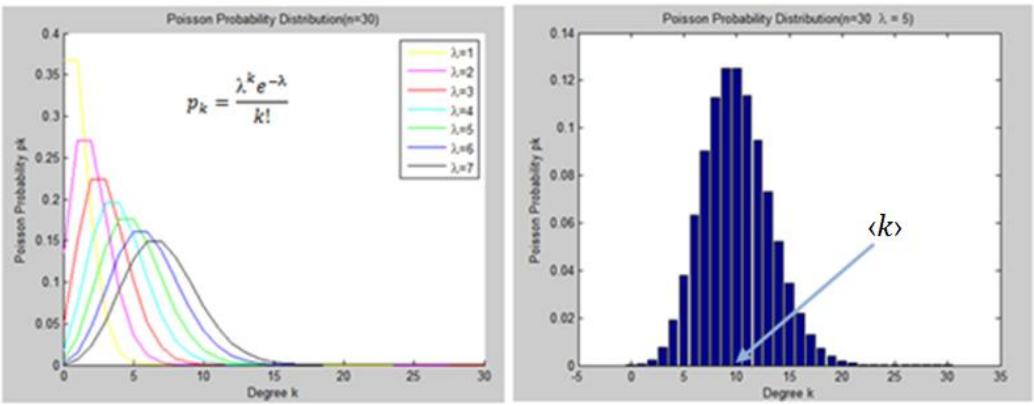


Figure 1.6 Poisson degree distributions of sample random graphs.

When probability distributions of random networks are observed, most of the vertices have approximately the same degree and the probability of very highly and little connected vertices are very small.

Binomial and Poisson distributions are sharply peaked and have tails that decay rapidly with k (1/k!) that is very high decay value than any exponential (Figure 1.5 – Figure 1.6). This behavior shows that appearance of well connected nodes isn't very possible. This property separates them from power-law distributions where some high-degree nodes exist. Random graphs' degree properties do not always follow binomial and Poisson distributions but other distributions such as exponential and power-law.

1.2.2.3 Path Length

Average degree of a random network is approximately calculated as $\langle k \rangle = \mathrm{Np}$ and the number of nodes a distance d (path length) from any node is approximately $\langle k \rangle^{\mathrm{d}}$.When d is equal to l (maximum distance between any two vertices) $\langle k \rangle^{\mathrm{l}}$ is the total number of vertices in network. Estimated maximum distance is,

$$\langle k \rangle^{\mathrm{l}} = \mathrm{N} \rightarrow \mathrm{d_{max}} = \mathrm{l} \approx \frac{\log \mathrm{N}}{\log \langle k \rangle} \quad (1.5)$$

Logarithmic growing of ℓ with N from equation show that ER graphs has "small-world" property and also many real-world networks exhibit same property.

1.2.2.4 Clustering Coefficient

In a random graph, the edges are created independently and distributed randomly so the clustering coefficient is $\mathrm{C} \cong \mathrm{p}(= \frac{\langle k \rangle}{\mathrm{N}})$ that is much smaller than comparable real networks with same number of nodes, and edges and go to zero with N^{-1} in the limit of large system size ($\mathrm{N} \rightarrow \infty$) (Table 1.2). There is no local clustering coefficient in random graphs.

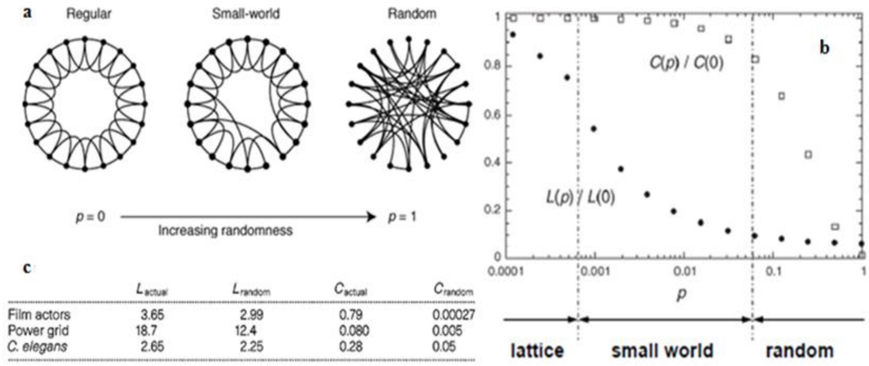
Table 1.2 Real-world network study results on mean degree, real clustering coefficient, and clustering value of random networks with same network size [7].

Network	N (# of vertices)	‹k› mean degree	C	C_{random}
Internet	6,374	3.8	0.24	0.00060
WWW	153,127	35.2	0.11	0.00023
Power grid	4,941	2.7	0.080	0.00054
Neural network	282	14.0	0.28	0.049
Biology collaborations	1,520,251	15.5	0.081	0.000010
Mathematics collaborations	253,339	3.9	0.15	0.000015
Film actor collaborations	449,913	113.4	0.20	0.00025
Company directors	7,673	14.4	0.59	0.0019
Word co-occurrence	460,902	70.1	0.44	0.00015
Metabolic network	315	28.3	0.59	0.090
Food web	134	8.7	0.22	0.065

All studies in network research revealed that the random networks are different from real-world networks with very small clustering coefficient (<<1) which tends to zero with N^{-1} and binomial or Poisson degree distribution that imply to any linking between nodes can occur with equal probability, and have similarity with small average path length (highly connected $\approx$log(N)) between vertices pairs. Also the average distance and clustering coefficient only depend on the number of nodes and edges in the network.

1.2.3 The Small-World Model (Watts-Strogatz)

Small-world models were introduced as a simple model of social networks by Duncan Watts and Steven Strogatz [2] but the name comes from Stanley Milgram's experiment of letter sending [8]. Watts and Strogatz [2] started with a regular network (p=0, ring-like) and reconnected independently each edge with increasing probability p that causes randomness until p=1 (Erdös-Rényi random graph). In these rewired networks, they observed high clustering coefficients, long path lengths for regular structure (p=0); low clustering coefficients, low path lengths for random structure (p=1); and high clustering coefficients, low path lengths for a range of p (small-world structure) compatible with real-world networks (Figure 1.7).



	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

Figure 1.7 Results by Watts and Strogatz [2]. **(a)** Random rewiring Procedure : p is increased the model moves from a regular graph, through intermediate graphs, to a random graph at $p = 1$ **(b)** Description of characteristic path length L(p) and clustering coefficient C(p) within randomly rewired procedure: There is a large intermediate region which shows "small-world" behavior: small l (like random graphs) but large C (like regular graph) **(c)** Empirical examples of small-world networks: Characteristic path length L and clustering coefficient C for three networks were compared to random graphs with same number of vertices (N) and average number of edges per vertex (k). Actors: N=225,226, k=61. Power-grid: N=4941, k=2.67.Celegans: N= 282, k=14. Networks were defined as undirected and unweighted structure. Sample real-world network reflected the small-world phenomenon with $L \gg L_{random}$ and $C \gg C_{random}$ properties.

Table 1.3 Comparison of networks on path length and clustering properties.

	GRAPHS		
	Regular	**Small-World**	**Random**
Path Length	Long	Short $L_{WS} \geq L_{rand}$	Short
Clustering	High	High $C_{WS} \gg C_{rand}$	Low

In rewiring process, average degree (‹k›) is not changed but the degree distributions change with different rewiring parameters p. Degree distribution is similar to the random graph degree distribution as binomial or Poisson that are centered on average degree ‹k› and with exponentially small probability for very high connected nodes. Also, changes are observed and compared for path length and clustering structures in rewiring process (Figure 1.7, Table 1.3). This model provided a new explanation of real-world networks which have two important properties.

The small-world effect: The pairs of vertices in most networks are connected with shortest path in the network and the mean or maximum vertex-vertex distance in the network (except unconnected vertex pairs) increase logarithmically (≈ logN or more slowly) with the all vertices in the network. The first important explanation of small world was made by the social psychologist Stanley Milgram [8] with his famous letter experiment and the experiment was resulted reached their targeted person via about six (~6.5) acquaintances and hence ‘‘six degrees of separation’’. Different field experiments showed similar results such as the actors are connected via three co-stars with each other in Hollywood and chemicals in a cell are separated by three reactions. [9] and [7] reveal empirical evidences in order to show the logarithmic increase of path length. To measure “small-world effect”, one finds the shortest distance between all pairs of vertices in the network and computes their averages. Mean shortest path length between vertex pairs for an undirected network is as follows.

$$l^{-1} = \frac{1}{\frac{1}{2}N(N+1)} \sum_{i \geq j} {l_{ij}}^{-1} \tag{1.6}$$

l_{ij} represents the shortest path length from vertex i to vertex j and there is no contribution of the infinite values ($\frac{1}{l_{ij}} = \frac{1}{\infty} = 0$) to the sum. Also, observations on experiments showed that mean shortest path distances increase as log N and the number of vertices within a distance d of a central vertex grows exponentially with ($\langle k \rangle^{d}$). The result of *l* (mean shortest path=$\langle k \rangle^{l} = N \rightarrow l = \frac{\log N}{\log \langle k \rangle}$) scales logarithmically or grows slower with network size (N) for fixed mean degree in real-world networks. This is a special property of "small-world" networks.

High clustering: There is a high probability that two vertices will be connected directly to each other if they have a common neighboring vertex. For example, in social networks one encounters "cliques" -circles of friends or acquaintances in which every member knows every other member. In a social network, two people are much more likely to be acquainted with one another if they have another common acquaintance or the friend of your friend is likely also to be your friend. "Clustering coefficient" can be defined as the density of triangles in a network.

1.2.4 Scale-Free Network Model (Barabási -Albert)

Price's studies on network of citations between scientific papers in [10] confirmed a power–law degree distribution and its exponent values were found 2.5-3. The next works on degree distribution of real-world networks (citation networks, WWW, Internet, metabolic networks, telephone call graphs, network of human sexual contacts) produced an important "power law" property. Power-law degree distribution was seen a characteristic property for real-world networks and these networks are defined as "scale-free networks" by Barabási and Albert in [3].

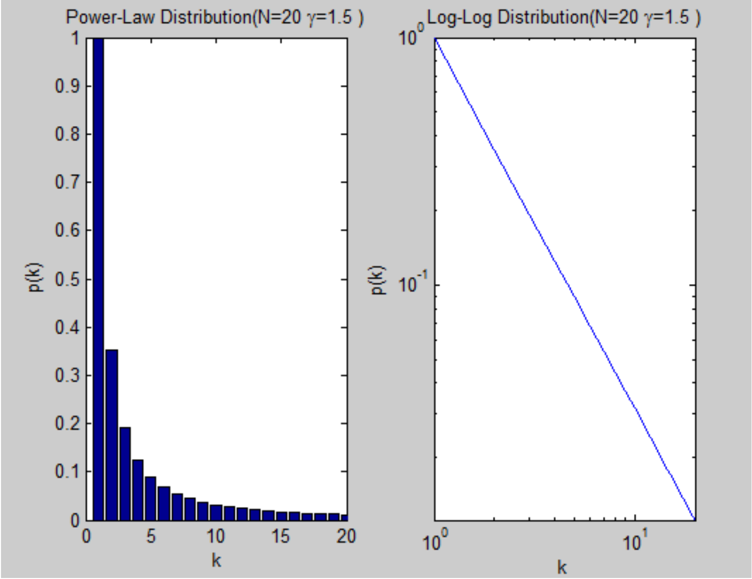


Figure 1.8 Sample power-law degree distribution and log-log fit to decay for a network of size is 20 and exponent value (γ) 1.5.

Power-law distributions are called as “Pareto Distribution” in statistics. Power-law distribution has the following form.

$$p(k) \sim k^{-\gamma} \tag{1.7}$$

Decaying of power-law distribution (Figure 1.8) is much slower than Poisson distribution that can be seen on large k,

$$k^{-\gamma} \gg \frac{1}{k!} \ \ (\text{Poisson distribution } \rightarrow p_k = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!}) \tag{1.8}$$

Power-law exponent ($\gamma > 0$ and usually between 1 and 3) values have special reflections about networks:

$1 \leq \gamma$: requirement for normalizability

$1 < \gamma \leq 2$: both the average degree and standard deviation are infinite, a total number of edges may grow faster than a linear function of a total number of vertices that this is true for many real networks

$2 < \gamma \leq 3$: average degree is finite, but standard deviation is infinite that mean fluctuations in systems are unbounded and depend only on the system size, network doesn’t have percolation threshold with respect to a dilution of its nodes, typical for real networks

$3 < \gamma \leq 4$: both the average degree and standard deviation are finite

$4 < \gamma$: power-law distribution and its properties look like exponential decaying distributions, network closes to random graph structure.

Previous random graph and small-world models didn’t uncover power-law degree distribution, so this property opened a new investigation field about networks. While Barabási and Albert were revealing power-law degree distribution property in their paper [3], they also reached three important results.

- Power-law degree distribution is not only a special property of WWW that is a general property of many real-world networks
- Rather than being static structures, networks grow dynamically via attachment of new vertices that is referred to with different descriptors such as “cumulative advantage”, “rich-get-richer”, and “preferential attachment”. Vertices are connected with new edges to the existing vertices according to their degrees
- Specific models of growing networks can be generated which show power-law degree distributions.

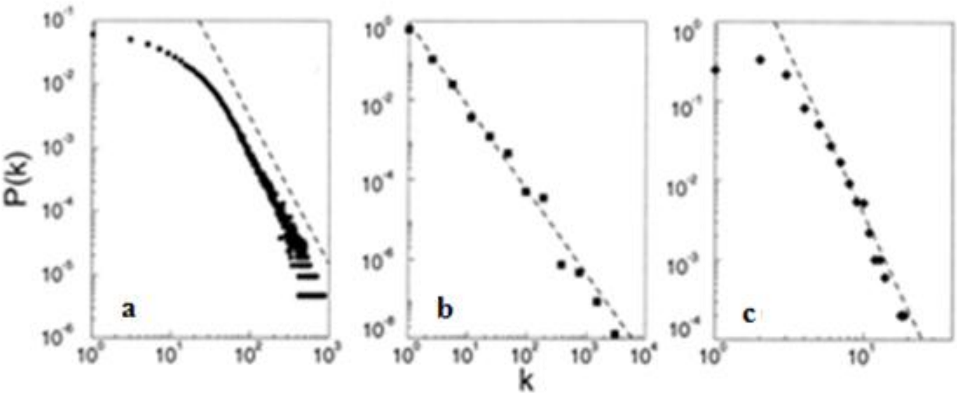


Figure 1.9 Barabási and Albert's real-world network analysis in [3] and observed distribution function of connectivity (average degree distributions) for **(a)** actor collaboration graph with $N = 212{,}250$ vertices, average degree $\langle k \rangle = 28.78$, and $\gamma_{actor} = 2.3$, **(b)** WWW $N = 325{,}729$, $\langle k \rangle = 5.46$, and $\gamma_{WWW} = 2.1$, **(c)** and Power grid, $N = 4941$, $\langle k \rangle = 2.67$, and $\gamma_{powergrid} = 4$.

Barabási and Albert's [3] findings are supported with studies on different real-world networks such as WWW, film actor collaborations, and citation networks (Figure 1.9).

We can observe and compare all network models with same size in Figure 1.10 and classify real-world networks on network model in Table 1.4.

Table 1.4 Different real-world networks and description of their vertices, edges, and models.

Networks	Vertices	Edges	Models
Electronic Circuit	Electronic components (resistors,diodes,capacitors)	Wires	Small world Scale-free
South California Power Grid	Transformers,substations, generators	High-voltage transmission lines	Small world Not Scale-free
Airport Network of World	Airports	Non-stop connections	Single scale network
Metabolic Networks	Substrates,enzymes, intermediate,complexes	Biochemical reactions	Scale-free
Conceptual Network of Language	Words	Appearance in the same entry of a thesaurus of English language	Small world Scale-free
WWW	Sites(pages)	Hyperlinks	Small world Scale-free
Internet	Routers	Wires	Scale-free
Co-authorship Network	Authors of scientific articles	Co-authorship	Scale-free

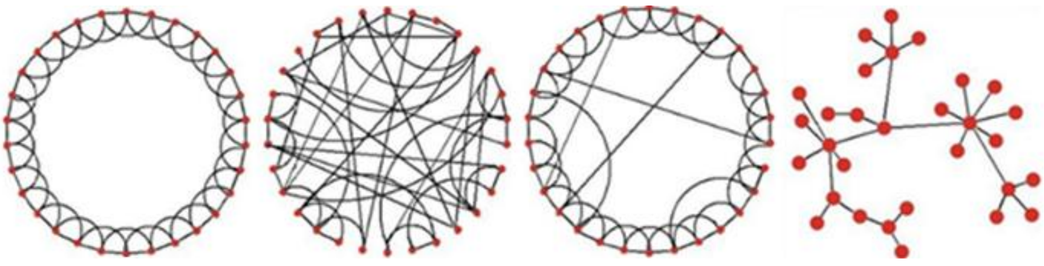


Figure 1.10 Networks are of same size (30 vertices). **(a)** Regular network model, each vertex is linked to its four nearest neighbors **(b)** Random network model, vertices are linked randomly and vertices have three and four edges **(c)** Small-world network model, most vertices are linked only to their nearest neighbors **(d)** Scale-free network model, a few vertices are linked many other vertices.

Chapter2

Network Analysis

2.1 Node Degree Distribution

Node degree of vertex i (k_i) is the total number of connected edges of vertex i to other vertices (nearest neighbors) in an undirected network. All nodes in a network do not have the same number of degree whose spread is characterized by a distribution function p(k) that gives the probability of a randomly chosen node having exactly k edges in an undirected network. Histogram of the degree probability of vertices p(k) versus node degrees is named degree distribution for the network. Average degree for the whole network ($\langle k \rangle$);

$$\langle k \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i \equiv \sum_k kp(k) \equiv \frac{2E}{N} \tag{2.1}$$

where E represents the total number of edges and N is the total number of nodes in network.

In an Erdös-Rényi random graph the edges are placed randomly and equal probability. The largest of nodes have approximately the same degree that closes to the average degree ‹k› of the network. The degree distribution of a random graph is a binomial distribution with a peak at p(‹k›).

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{2.2}$$

The mean number of neighbors of a chosen vertex i in a random graph with degree distribution (Figure 2.1) p_k is found as follows.

$$\langle k \rangle = \sum_k kp_k \tag{2.3}$$

In the limit where N becomes large the degree distribution of a random graph follows Poisson distribution.

$$p_k = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!} \tag{2.4}$$

N 20
p 0.5

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

k	$\binom{N-1}{k}$	p^k	$(1-p)^{N-1-k}$	p_k	$p_k k$
0	1	1	1.90735E-06	1.90735E-06	0
1	19	0.5	3.8147E-06	3.62396E-05	3.62396E-05
2	171	0.25	7.62939E-06	0.000326157	0.000652313
3	969	0.125	1.52588E-05	0.001848221	0.005544662
4	3876	0.0625	3.05176E-05	0.007392883	0.029571533
5	11628	0.03125	6.10352E-05	0.02217865	0.11089325
6	27132	0.015625	0.00012207	0.051750183	0.310501099
7	50388	0.007813	0.000244141	0.096107483	0.67275238
8	75582	0.003906	0.000488281	0.144161224	1.153289795
9	92378	0.001953	0.000976563	0.176197052	1.585773468
10	92378	0.000977	0.001953125	0.176197052	1.76197052
11	75582	0.000488	0.00390625	0.144161224	1.585773468
12	50388	0.000244	0.0078125	0.096107483	1.153289795
13	27132	0.000122	0.015625	0.051750183	0.67275238
14	11628	6.1E-05	0.03125	0.02217865	0.310501099
15	3876	3.05E-05	0.0625	0.007392883	0.11089325
16	969	1.53E-05	0.125	0.001848221	0.029571533
17	171	7.63E-06	0.25	0.000326157	0.005544662
18	19	3.81E-06	0.5	3.62396E-05	0.000652313
19	1	1.91E-06	1	1.90735E-06	3.62396E-05
				$<k>$	9.5

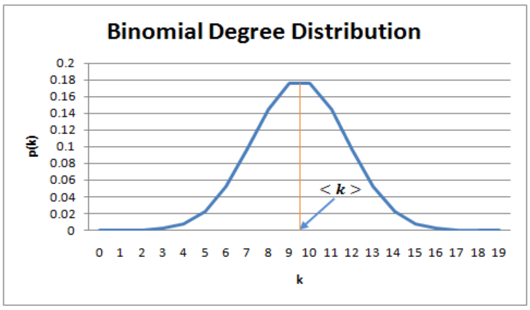


Figure 2.1 Sample binomial degree distributions.

Both binomial and Poisson distributions are strongly peaked about the mean ‹k›, and have a large-k tail that decays rapidly with 1/k!. Such networks are defined as ”homogeneous networks”. But binomial and Poisson degree distributions do not represent real-world network properties that decay slowly in large degree ranges. In real-world networks most nodes have low degrees and a few numbers of nodes (hubs) have high degrees (heterogeneous connectivity) and high clustering. This behavior can be modeled with exponential or power-law distributions.

Exponential degree distribution,

$$p(k) \sim e^{(-k/\langle k \rangle)} \tag{2.5}$$

Cumulative exponential distribution (Figure 2.2) with same exponent,

$$P_k = \sum_{k'=k}^{\infty} p_k \sim \sum_{k'=k}^{\infty} e^{-k'/} \sim e^{-k/\langle k \rangle} \tag{2.6}$$

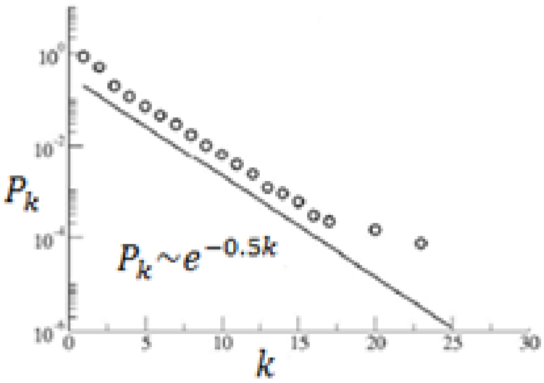


Figure 2.2 Cumulative exponential degree distribution of power grid [11].

Also experiments on large real-world networks (WWW, internet, metabolic networks) showed that their degree distributions significantly different from binomial and Poisson distributions and follow the power-law degree distribution. In power-law distribution (2.7) for networks vertices with small degrees are most frequent and the fraction of highly connected vertices decreases, but is not zero within a logarithmic plot (Figure 2.3). These kinds of networks are called as a “scale-free networks” in [3].

$$p_k \sim k^{-\gamma} \tag{2.7}$$

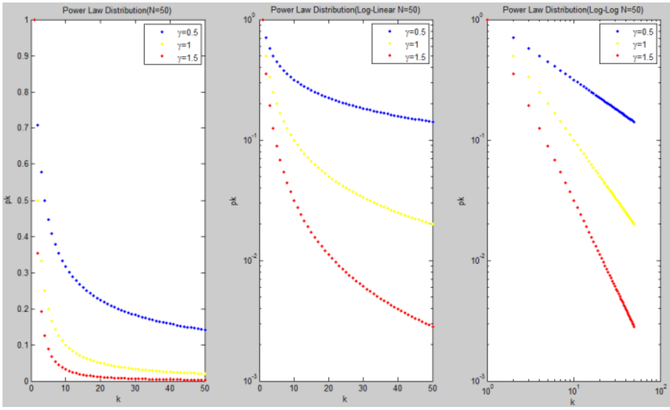


Figure 2.3 Sample power-law degree distributions in linear-linear, semi-log, and log-log forms.

The scale-free degree distribution has a heterogeneous structure whereas random and small-world networks follow a homogeneous topology. Both exponential and power-law degree distributions show a non-equilibrium growth of the network when new vertices and edges are added to the existing network in time. In power-law degree distribution, connections of newly added vertices to existing vertices occur with “preferential attachments” which means that new connection probability to existing node is proportional to the degree of existing vertices. In this way popular nodes become more popular and “hubs” are created, in other words rich get richer.

Investigation of power-law degree distribution is made on cumulative distribution function or probability distribution function and the plot is shown in a log-log scale in order to investigate the linearity of the relationship and to estimate the exponent. Use of cumulative distribution allows removal of the fluctuations in large degree regions and makes fluctuations less pronounced. The probability that the degree is greater than or equal to k is,

$$P_k = \sum_{k'=k}^{k_{max}} p(k') \tag{2.8}$$

Studies in real-world networks also showed no fluctuations and revealed right-skewed distributions that follow power law with a constant exponent γ [12, 13].

As a result, power-law and exponential distributions can be observed when corresponding cumulative distributions are plotted in logarithmic scales (for power laws) or semi-logarithmic scales (for exponentials) in real-network studies.

Analysis of linearity fit is made by calculating the Pearson product-moment correlation coefficient (r) that is a measure of the correlation (strength of the linear dependence) between two variables. It is a value between +1 and −1 where 1 implies the perfect relationship between X and Y (Y increases as X increases), −1 implies the inverse relationship between X and Y (Y decreases as X increases), and 0 implies the no linear correlation between the variables.

$$r = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{N \sum x_i^2 - (\sum x_i)^2} \sqrt{N \sum y_i^2 - (\sum y_i)^2}} \quad (2.9)$$

2.2 Centrality

Centrality is an important analysis method in network studies. It indicates the importance or popularity of a vertex and edge within a network and has different forms: degree centrality (on node degree), closeness centrality (on geodesic path), and betweenness centrality (on geodesic path) that are described by Freeman [14], and eigenvector centrality by Bonacich [15].

2.2.1 Degree Centrality

Degree centrality is measured on the number of edges attached to given vertex i that also represents the degree of a vertex (k_i) like the number of people that knows the given person. Degree centrality shows the potential influence of a vertex in network. For example, if a person who has more connection in an epidemic network gets sick, disease can spread in the network quickly.

Degree centrality of vertex i:

$$C^D(i) = k_i \text{ (degree of a vertex i)} \quad (2.10)$$

The elements of the adjacency matrix (A) of an undirected graph are $a_{ij} = a_{ji} = 1$ if there is an edge between i and j in the graph and $a_{ij} = 0$ otherwise. Then, the degree of a vertex i can be found by

$$k_i = \sum_j A_{ij} \quad (2.11)$$

High degree vertices are called "hubs" that have a high role in the network.

2.2.2 Eigenvector Centrality

Eigenvector centrality of vertex i is the sum of its connections to other nodes that are weighted with their centrality and is also referred as the principal eigenvector of the adjacency matrix. The equation of an eigenvector is;

$$x_i = \frac{1}{\lambda}\sum_{j=1}^{N} a_{ij}x_j \qquad x = x_1, x_2, \ldots\ldots x_N \quad (2.12)$$

$$\lambda x = Ax \text{ (vector centralities)}$$

x_i : centrality of vertex i,

A : adjacency matrix of the network,

λ : the eigenvalue (constant - must be the largest eigenvalue of the adjacency matrix),

x : the eigenvector of a adjacency matrix,

x_i is proportional to the sum of the scores of all nodes which are connected to i. The eigenvector centrality shows the importance of a vertex in the network.

2.2.3 Closeness Centrality

Closeness centrality is the total shortest path distance to all other vertices in the network. A low closeness centrality value (highly central) implies closeness and easy access to other vertices. We can say that vertices are more central if they reach other vertices easily. Closeness provides a centrality measure for a vertex and measures the spread time (its inverse shows the spreading speed) of an effect from a given vertex to other reachable vertices in the network.

Closeness centrality of a vertex i: $C^C(i) = \sum_j d_{ij}$ (2.13)

2.2.4 Betweenness Centrality

The most widely used kind of centrality is "betweenness centrality" that is the number of shortest paths on a network that run through the vertex of interest [16].

Measuring of betweenness of a vertex i is the number of shortest path vertex pairs that pass through between all vertex pairs in network. It also refers the importance of a vertex in traffic on a network.

Betweenness centrality of a vertex i: $C^B(i) = \sum_{j \neq i \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}}$ (2.14)

σ_{jk} : total number of the shortest path between vertices j and k (>0)

$\sigma_{jk}(i)$: number of the shortest path between vertices i and j that pass through a vertex i [16].

Betweenness shows the importance of vertex i for connecting vertice pairs. A vertex with a high betweenness centrality value represents a hub (controller) of the network and its value is strongly related to its degree. In a road network, high betweenness could indicate where alternate routes are needed. Betweenness also measures the resilience of a network since removal of high betweenness nodes can cause connectivity in network to be lost.

2.3 Clustering

Widely used measuring of "clustering coefficient" is defined by Watts and Strogatz [2] on local values as,

$$C_i = \frac{\text{number of triangles connected to vertex i}}{\text{number of triples centered on vertex i}} \quad (2.15)$$

Node has k_i edges. There are $\frac{k_i(k_i-1)}{2}$ possible edges between node i and its next nearest neighbors. The ratio between the actually exist number of edges (E_i) to the possible total number of connection ($\frac{k_i(k_i-1)}{2}$) gives the local clustering coefficient (Figure 2.4, Figure 2.5) of node i,

$$C_i = \frac{E_i}{k_i(k_i-1)/2} = \frac{2E_i}{k_i(k_i-1)} \quad (k_i \neq 0,1) \quad (2.16)$$

E_i denotes the edges among first neighbors of node i.

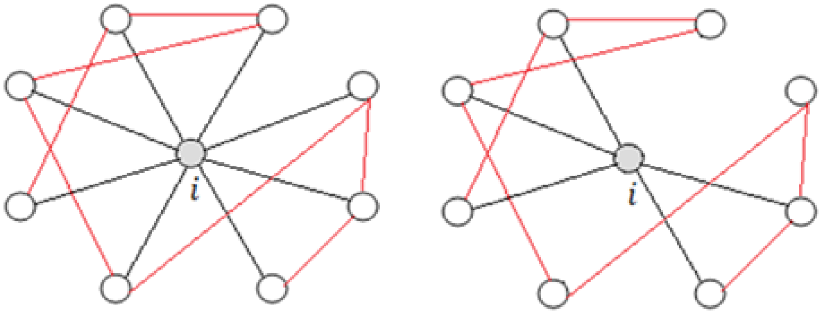


Figure 2.4 Local clustering coefficient example. **(a)** Central vertex i has 8 neighborhood in same network and there are $28(8(8-1)/2)$ possible edges between them. But there is only 7 actual edges and local clustering coefficient of central vertex i (C_i) is 7 by 28 (0.25). **(b)** If a vertex i has a small number neighborhood like $k_i = 5$, there will be 10 possible neighborhood between its neighbors and its clustering coefficient will increase to $7/10(0.7)$.

Clustering coefficient of whole network is the average of all individual local clustering coefficients.

$$C_{WS} = \frac{1}{N}\sum_{i=1}^{N} C_i = \frac{1}{N}\sum_{i=1}^{N} \frac{(\text{number of connected neighbor pairs})}{\frac{1}{2}k_i(k_i-1)} \quad (2.17)$$

$$(1 \geq C_{WS} \geq 0)$$

This definition has problems as it is heavily biased in favor of low degree vertices because of the factor $k_i(k_i-1)$ in the denominator. Differences in the value of clustering coefficients can become very high. For instance, take two vertices in the same network; one having two neighbors that are connected and the other having a hundred with none of them connected. Local clustering coefficients are $C_1 = 1, C_2 = 0$ and average probability is $C = 0.5$, but there is 4951(1 from vertex 1(2(2-1)/2=1) and (100*99)/2=4950 from vertex 2) possible total pairs in all network and only one is connected. As a result average probability of a pair being connected is not 0.5, but 1/4951=0.0002. Correct way to calculate the average probability of a pair of neighbors being connected was expressed by Newman [9] (Figure 2.5) as;

$$C = \frac{3(\text{number of triangles on a graph})}{\text{number of connected triples of vertices}} \quad (1 \geq C \geq 0) \quad (2.18)$$

- **triangles:** three vertices that are each connected to both of the others. If vertices i1, i2, and i3 are connected to each other, there is a triangle

- **connected triple:** a vertex that is connected to a pair of other vertices which may or may not be connected to each other. There is a triple between vertices i1, i2, and i3 around i1 if i1 is connected to i2 and i3
- **the factor of 3 in the numerator:** each triangle contributes three separate connected triples.

Method is also called the “fraction of transitive (closed) triples” by social network analysts and easier to calculate analytically. In 2003 Newman proposed an alternative and numerically equal clustering coefficient definition as:

$$C = \frac{6(\text{number of triangles on a graph})}{\text{number of path of length 2}} \quad (1 \geq C \geq 0) \tag{2.19}$$

This definition shows that C is also the mean probability that the friend of your friend is also your friend. (Number of paths of length 2: number of friends of friends)

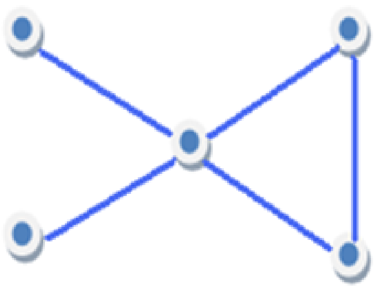


Figure 2.5 Newman’s [9] clustering coefficient equation show (Eq. 2.15) that graph has one triangle and 8 connected triples therefore clustering coefficient is $C = (3 * 1)/8 = 3/8$. Local clustering coefficients are 1,1, 1/6, 0, 0 and mean value $C = 13/30$ on Watts –Strogatz equation (Eq. 2.17) for same graph.

In a random graph, connected probability of two vertices is $p = \frac{\langle k \rangle}{N}$ for all vertex pairs where ‹k› is the mean degree of a vertex. Clustering coefficient of a random graph is found as

$$\langle C_{ER} \rangle \cong \frac{\langle k \rangle}{N} \cong \frac{2E}{N^2} \tag{2.20}$$

The clustering coefficient is quite small for random graphs in comparison with same size real-world networks with vertices (N) and mean degree values (k) ($C \gg C_{rg}$) as shown in Table 2.1 [2].

Table 2.1 Clustering coefficients of real-world networks and comparison with clustering coefficients of random networks on same network sizes.

Networks	N	<k>	Measured Clustering Coefficient	Random Clustering Coefficient
Internet	6374	3.8	0.24	0.0006
WWW	153127	35.2	0.11	0.00023
Power grid	4941	2.7	0.08	0.00054
Biology collaborations	1520251	15.5	0.081	0.00001
Mathematics collaborations	253339	3.9	0.15	0.000015
Film actor collaborations	449913	113.4	0.2	0.00025
Company directors	7673	14.4	0.59	0.0019
Word cooccurrence	460902	70.1	0.44	0.00015
Neural network	282	14.0	0.28	0.049
Metabolic network	315	28.3	0.59	0.09
Food web	134	8.7	0.22	0.065

2.4 Assortativity

Networks consist of different types of vertices, properties and connections such as races, gender, ages, income, education etc. for social networks. Sample observations in social networks [17, 18] show that connections are realized based on the properties of vertices like race and content of Web links. This property of networks and measurement method of connection tendency for the vertices according to their degrees is called as "assortative mixing". If high-degree vertices have a tendency to connect with other high-degree vertices, we can say that the network shows "assortativity (homophily)". On the other hand the connection tendency of low-degree vertices to high-degree vertices is called "disassortativity (disassortative mixing)" [19].

Degree assortativity is also used to show the preferential attachment of the high-degree vertices with other high or low degree vertices. Calculation of the assortativity mixing is done with "assortativity coefficient" and it measures in terms of the mean Pearson correlation coefficient (r) of degree between pairs of linked nodes in [19, 20].

$$r = \frac{M^{-1}\sum_i X_i Y_i - [M^{-1}\sum_i \frac{1}{2}(X_i+Y_i)]^2}{M^{-1}\sum_i \frac{1}{2}(X_i{}^2+Y_i{}^2) - [M^{-1}\sum_i \frac{1}{2}(X_i+Y_i)]^2} \tag{2.21}$$

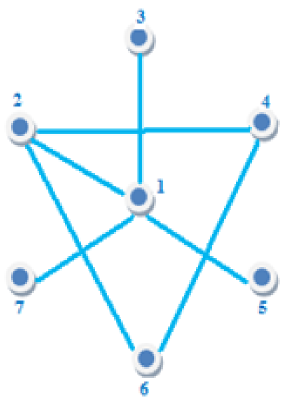
X_i and Y_i refers the connected vertices by edge i within total M edges in network. In this equation X_i and Y_i represent the the nearest vertex degrees and the equation is recalled as "degree assortativity r^1". Values $(-1 \leq r \leq 1)$ represent as;

- $r > 0$: a correlation between nodes of similar degree (assortative)
- $r < 0$: a correlation between nodes of different degree (disassortative)
- $r = 1$: the network is perfect assortative
- $r = -1$: the network is completely disassortative
- $r = 0$: the network is randomly mixed.

Observation on real-world studies showed that social networks are assortative, while technological and biological networks are disassortative.

2.5 Shortest Path Length

The path length is the total number of edges to travel from a starting vertex to a target vertex (Figure 2.6) and the shortest value between these path lengths that connects two vertices within minimum travelling steps in a network is called a "shortest path length" or "geodesic distance". Shortest path length between vertex pairs$(i-j)$ is denoted as $l_{ij} = l_{ji}$ in undirected networks. If there is no path between two vertices, the distance is assumed to be infinite (undefined). The shortest path length between vertex pair i and j is equal to minimum power of the adjacency matrix (A) that will reach a non-zero value for that vertex pair (a_{ij}).



Geodesic Distances

$l_{12} = l_{21} = 1$	$l_{23} = l_{32} = 2$	$l_{34} = l_{43} = 2$
$l_{13} = l_{31} = 1$	$l_{24} = l_{42} = 1$	$l_{35} = l_{53} = 2$
$l_{14} = l_{41} = 2$	$l_{25} = l_{52} = 2$	$l_{36} = l_{63} = 3$
$l_{15} = l_{51} = 1$	$l_{26} = l_{62} = 1$	$l_{37} = l_{73} = 2$
$l_{16} = l_{61} = 2$	$l_{27} = l_{72} = 2$	$l_{45} = l_{54} = 3$
$l_{17} = l_{71} = 1$	$l_{56} = l_{65} = 3$	$l_{46} = l_{64} = 1$
$l_{67} = l_{76} = 3$	$l_{57} = l_{75} = 2$	$l_{47} = l_{74} = 3$

Figure 2.6 Shortest path lengths in sample network.

Diameter of a network is a maximum geodesic distance between vertex pairs ($\max(l_{ij})$ for all i and j). Mean shortest path length between vertex pairs (including loops $\frac{N(N+1)}{2}$) for an undirected network is calculated as follows.

$$l = \frac{1}{\frac{1}{2}N(N+1)} \sum_{i \geq j} l_{ij} \quad (2.22)$$

Previously mentioned condition that if there is any unconnected vertex pairs in a network their shortest path length is defined as infinite causes a problem in mean shortest path calculations. This problem is overcome with “harmonic mean” that removes this infinite value by equating it to zero ($\frac{1}{\infty} = 0$). The “harmonic mean” shortest path length distance between all connected pairs is the reverse of the average of the reverses.

$$l^{-1} = \frac{1}{\frac{1}{2}N(N+1)} \sum_{i \geq j} {l_{ij}}^{-1} \tag{2.23}$$

Table 2.2 Statistical analyzed networks that were published. The properties measured are: network types: directed or undirected; network size: N; total number of edges m; mean node degree ‹k›; mean shortest path length ‹l›; power-law exponent γ;clustering coefficients C^1, C^2; and degree correlation coefficient r [21].

	Networks	Type	N	m	⟨k⟩	⟨l⟩	γ	C^1	C^2	r
Social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1			
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16	
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029
	sexual contacts	undirected	2 810				3.2			
Information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7			
	citation network	directed	783 339	6 716 198	8.57		3.0/–			
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44	
Technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366
Biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	−0.326
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226

We can summarize real-world network properties (Table 2.2) as follows.

- Low average path length (small-world phenomenon)
- High clustering
- Degree distributions follow Power-law (many networks) and they called as “Scale-free Networks”
- Betweenness centrality distribution is a decreasing function (usually power-law). Decreasing function properties of degree and betweenness centrality distribution indicate heterogeneity and existence of hubs.

- The distances are proportional to the logarithm of network size

$$l \approx \frac{\log N}{\log \langle k \rangle} \quad \text{(Small-world property)}$$

- The clustering coefficient is independent of network size and is proportional to average degree (which is larger than comparable random networks)

$$C \propto \langle k \rangle$$

In the next chapter, we will study bus transportation networks of four largest Turkish cities - İstanbul, İzmir, Ankara, and Bursa. This study involves the statistical analysis and comparison of network properties of bus transportation networks for these cities. We will also try to relate our findings with network topologies of cities and other geographical, historical, and economical factors.

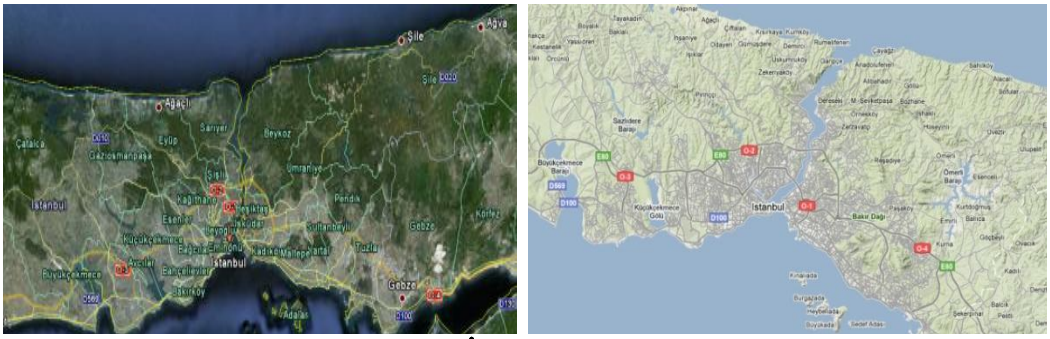
Chapter3

Bus Transportation Network Study on Four Largest Cities of Turkey

3.1 Introduction

Transportation networks are complex structural real-world networks and they have been of interest within statistical physics during the past decade. First transportation network analyses were made on power grids [22, 23] with transformers and transmissions that were defined as nodes and edges respectively. Researchers investigated railway networks [24], airport networks (India [25], China [26]), and public transportation networks (bus and tram networks of Poland [27, 28] and China [29, 30, 31], railway, subway, seaway networks of fourteen cities in the world [32]). In these studies, network efficiency, degree distributions, clustering coefficients, centralities, and assortativity properties were observed. Exponential and power-law distributions were revealed showing correlation between degree and other network properties.

In this study, we investigated bus transportation networks (BTNs) of four largest cities (İstanbul, İzmir, Ankara, Bursa) covering 42% of the population in Turkey. These cities are located in different geographical locations (Figure 3.1) and they have different economical, historical, and cultural backgrounds.



İstanbul

Ankara

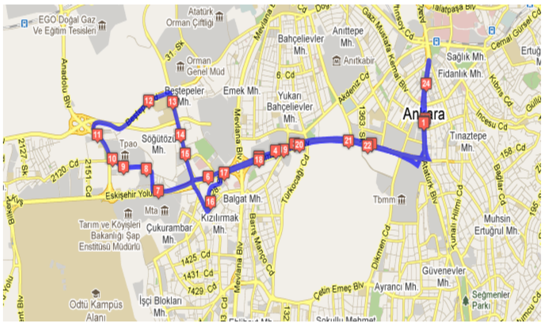
İzmir

Bursa

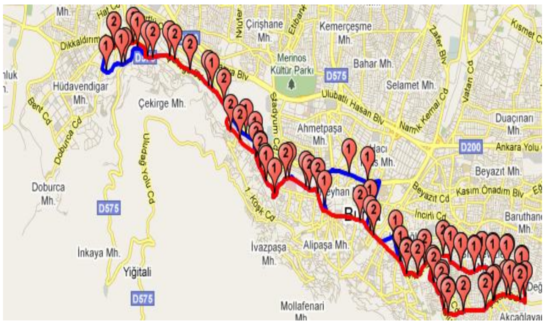
Figure 3.1 Topographic appearances of four largest cities of Turkey.

The specific properties of these resulted in public transportation networks with different characteristics. We only deal with BTNs. While availability of data was the main reason for this, one could also say that bus transport is still a very dominant transportation mode in Turkey in general. Table 3.1 gives a comparison of some route statistics for the cities in Turkey and other cities in the world taken from [32]. Sample routes can be observed in Figure 3.2.

a **b**



c



d

Figure 3.2 Sample routes and stations of bus transportation networks that are taken from **(a)** İETT, **(b)** ESHOT, **(c)** EGO, and **(d)** BURULAŞ – respective transportation departments of municipality in İstanbul, İzmir, Ankara, and Bursa.

Table 3.1 N: number of bus stops; R: number of bus routes; F: mean number of bus stations per route; P: population of cities.

City	N	R	F	P
İstanbul	4726	529	8.9	13.120.596
İzmir	5489	295	18.61	3.606.326
Ankara	4249	376	11.3	4.641.256
Bursa	3914	206	19.0	2.308.574
Sydney	1978	596	3.3	4,575,532
Hong Kong	2024	321	6.3	7,055,071
Taipei	5311	389	13.7	6,900,273
Dallas	5366	117	45.9	6,477,315
Sao Paolo	7215	997	7.2	19.672.582
Los Angeles	44629	1881	23.7	15,250,000

Before going into the details of our analysis, we introduce different network topologies that will be used in this study.

3.2 Bus Network Topology and Representations

Connections on different parts of a network produce differences in topological representation of the network. Bus networks consist of bus stations (or routes) that are defined as vertices. The connections of the vertices are defined via route-route and station-station relations and they are embedded in two-dimensional space.

If routes of a bus network are defined as vertices and linked when they intersect on a common bus station, this topology is called C-Space. Other bus network topologies can be obtained by using the bus stops as vertices. In L-Space, bus stops are only connected when they follow each other on a route. However, in P-Space bus stops are connected if they can be reached via a route regardless of their actual location. Thus in P-space there will be many more connections (edges) than in L-space (Figure 3.3).

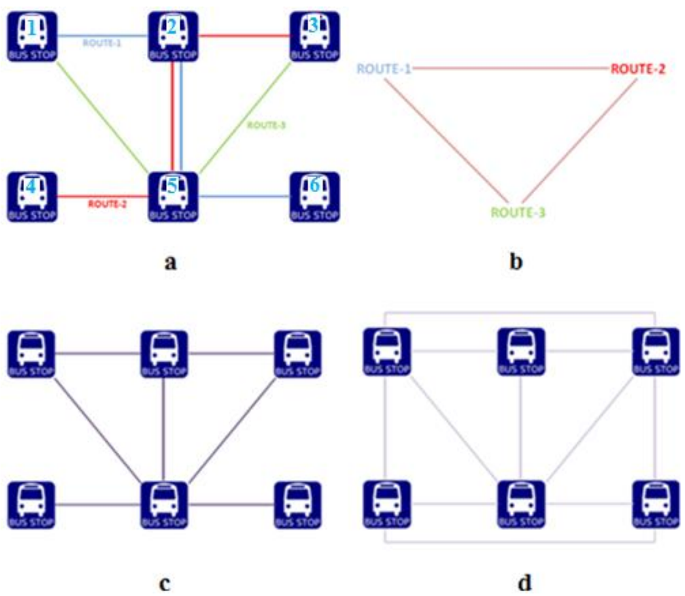


Figure 3.3 (a) Sample bus transportation network is defined on 3 routes and 6 bus stations. **(b)** Relation representation of routes and bus stations are showed in C-Space, **(c)** L-Space, and **(d)** P-Space respectively.

An adjacency matrix can be used to represent and mathematically analyze a network. It is a square, symmetric matrix that is defined as $a_{ij} = 1$ if there is any connection between pairs of vertices i and j; otherwise $a_{ij} = 0$. Vertices of networks don't have loops ($a_{ii} = 0$) and multiple connections, and their edges are undirected (Figure 3.4).

L-Space	1	2	3	4	5	6
1	0	1	0	0	1	0
2	1	0	1	0	1	0
3	0	1	0	0	1	0
4	0	0	0	0	1	0
5	1	1	1	1	0	1
6	0	0	0	0	1	0

P-Space	1	2	3	4	5	6
1	0	1	1	1	1	1
2	1	0	1	1	1	1
3	1	1	0	1	1	1
4	1	1	1	0	1	1
5	1	1	1	1	0	1
6	1	1	1	1	1	0

C-Space	Route-1	Route-2	Route-3
Route-1	0	1	1
Route-2	1	0	1
Route-3	1	1	0

Figure 3.4 Sample adjacency matrices of L-Space, P- Space, and C- Space for networks in Figure 3.3.

In following parts we will use these representations in order to calculate and analyze the properties of bus networks of four largest cities (Table 3.2). All network properties have been defined on their "space" structures and used specific abstracts. Topological representations of networks were used as exponents or subscripts in network property notations. For example, maximum shortest path length in an L-Space representation is shown as l_L^{max} and defined as the maximum number of stops to travel between any two stations.

Table 3.2 Bus network properties for C-, L-, and P-Spaces. ‹k›: average node degree of nearest neighbors for whole network; ‹z›: average node degree of next nearest neighbors for whole network; δ: rate between ‹z› and ‹k› ($\delta = ‹z›/‹k›$); k^{max}: maximum node degree in space; ‹C› mean (global) clustering coefficient of whole network; r^1: assortativity mixing of nearest neighbor; ‹l› : average shortest path length; l^{max}: maximum shortest path length in network.

	C-Space								L-Space								P-Space							
City	$‹k_C›$	k_C^{max}	$‹c_C›$	$r^1{}_C$	$‹l_C›$	l_C^{max}	$‹z›_C$	δ_C	$‹k_L›$	k_L^{max}	$‹c_L›$	$r^1{}_L$	$‹l_L›$	l_L^{max}	$‹z›_L$	δ_L	$‹k_P›$	k_P^{max}	$‹c_P›$	$r^1{}_P$	$‹l_P›$	l_P^{max}	$‹z›_P$	δ_P
İstanbul	88.96	258	0.709	0.210	2.068	4	318.33	3.58	3.367	50	0.087	0.146	10.29	53	15.62	4.64	121.043	1064	0.779	-0.097	2.86	5	1387.619	11.464
İzmir	48.68	133	0.726	0.211	2.253	8	150.66	3.10	2.405	27	0.026	0.108	46.44	212	4.13	1.72	154.576	1696	0.821	-0.016	3.25	9	1353.301	8.755
Ankara	54.40	143	0.698	0.434	2.365	6	156.78	2.88	3.159	49	0.101	0.105	9.629	55	12.97	4.11	78.691	1287	0.833	-0.055	2.94	7	1170.532	14.875
Bursa	70.94	148	0.737	0.204	1.813	6	107.99	1.52	2.472	13	0.015	0.295	29.00	126	4.16	1.68	221.67	1764	0.768	-0.022	2.46	7	2118.699	9.558

3.3 Network Properties

3.3.1 Node Degree Analysis

Node degree and cumulative node degree distributions are investigated using linear-linear, log-linear, and log-log plots for each topology.

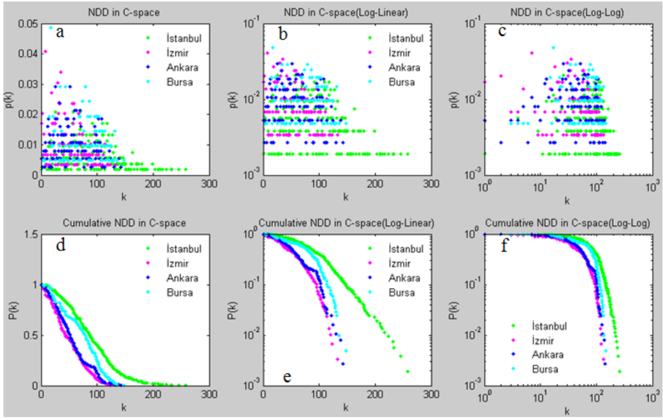


Figure 3.5 Node degree and cumulative node degree distributions in C-Space for four cities. Plots show the distributions in linear-linear, log-linear, and log-log scales respectively.

Degree of a node (route) in C-Space gives the total number of connections to other routes via common bus stations. Figure 3.5 shows the node and cumulative node degree distributions in C-Space whose cumulative log-linear (e) and log-log (f) plots show exponential decay fit for four cities. An exponential fit in C-Space indicates a random route structure and growth. All cities have the same growth structure. Exponent parameters of distributions in C-Space give no special information about the corresponding network so they have not been calculated (Table 3.3).

Table 3.3 Exponent values of exponential ($\bar{k}$) and power-law (γ) degree distributions and Pearson correlation coefficients (R) for L- and P-Spaces.

	L-Space				P-Space			
City	$\bar{k}_L$	$R_L^{\langle k \rangle}$	γ_L	R_L^{γ}	$\bar{k}_P$	$R_P^{\langle k \rangle}$	γ_P	R_P^{γ}
İstanbul	6.67	0.705	2.660	0.995	153.639	0.968	1.873	0.998 (%92)
İzmir	3.12	0.799	3.46	0.997	179.479	0.983	1.973	0.991 (%91)
Ankara	5.80	0.708	2.73	0.991	101.688	0.978	2.535	0.998 (%90)
Bursa	1.32	0.975	4.44	0.995	239.914	0.989	1.661	0.985 (%93)

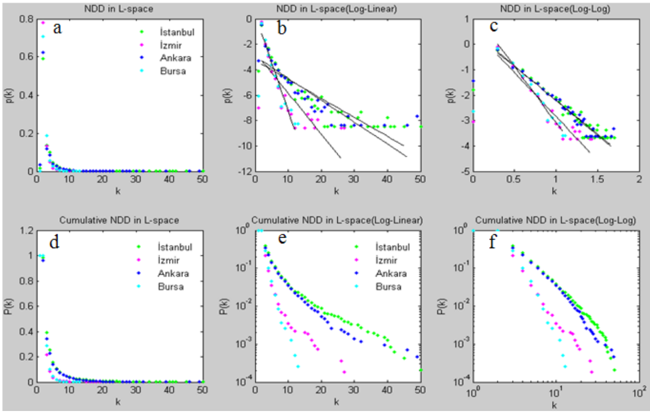


Figure 3.6 Node degree and cumulative node degree distributions in L-Space for four cities. Plots show the distributions in linear-linear, log-linear, and log-log scales respectively.

In Figure 3.6 upper plots (a-b-c) show the node degree distributions and lower plots (d-e-f) represent cumulative node degree distributions in linear-linear, log-linear, and log-log scales for L-Space. Degree $k = 2$ has maximum observed probability. Number of nodes with degree $k = 1$ is smaller than the number of nodes with degree $k = 2$ nodes and probabilities of node degree observations $p(k)$ decrease with k increasing. Maximum node degrees reach 50, 27, 49, and 13 in İstanbul, İzmir, Ankara, and Bursa respectively. Nodes with high degrees are called hubs in L-Space. Fitting parameters and Pearson correlation coefficients are given in Table 3.3. Figure 3.6.b is a log-linear plot in order to observe exponential decay behavior. Exponential decay is observed for four cities but Bursa has the highest Pearson correlation coefficient with $\bar{k}_L = 1.32$and with $R_L^{Bursa} = 0.975$ while $R_L^{İzmir} = 0.799$, $R_L^{Ankara} = 0.708$, and $R_L^{İstanbul} = 0.705$. Figure 3.6.c shows node degree distribution in log-log scale in order to observe power-law decay. All four cities reflected high power –law fitting with different exponent values and their Pearson correlation coefficients exceeded $R_L^{\gamma} = 0.99$. Power-law exponent parameters show that nodes in İstanbul and Ankara ($2 < \gamma_L$ <3) are robustly connected and random removal of their nodes will not damage unity and high connectivity up to a point. Also values close to 3 mean that in İstanbul and Ankara the transportation network has grown preferentially. In İzmir, nodes are added faster than edges. Such network growth can weaken network unity and tends to random growth and exponential degree decay. Although Bursa has a high correlation coefficient $R_L^{\gamma} = 0.995$ which indicates power-law decay, its power-law exponent parameter $\gamma = 4.44 > 4$ and fit to an exponential decay shows that Bursa bus network is also growing randomly.

Good fits of node degree distribution to power-law in L-Space for four cities indicate highly connected bus stops and random connection between routes that was shown in C-Space results. Fitting parameters of similar works for bus transportation networks in L-Space are in line with our study and range from 0.78 to 1.88 for exponential and from 2.72 to 5.49 for power-law [27].

Good fits of node degree distribution to power-law in L-Space for four cities indicates highly connected bus stops. Fitting parameters of similar works for bus transportation networks in L-Space are in line with our study and range from 0.78 to 1.88 for exponential and from 2.72 to 5.49 for power-law [32].

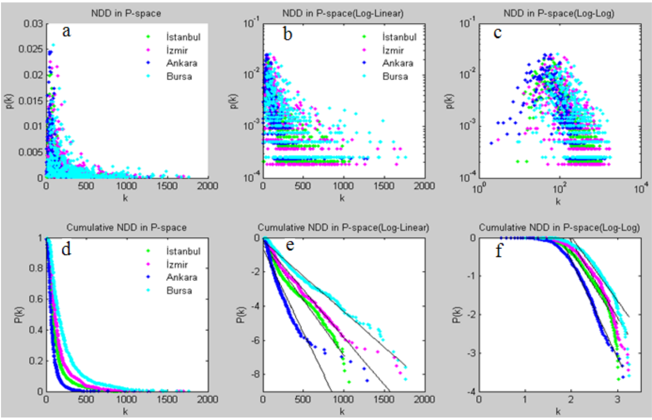


Figure 3.7 Node degree and cumulative node degree distributions in P-Space for four cities. Plots show the distributions in linear-linear, log-linear, and log-log scales respectively.

Node degree and cumulative node degree distributions in P-Space of four cities are plotted in Figure 3.7. Node degree analysis for P-Space is made on semi-log and log-log plots of cumulative node degree distributions. The fitted parameters are shown in Table 3.3. Results show that for all cities networks decay exponentially decay and Bursa has the highest fit in L-Space with $R_L^{\langle k \rangle} = 0.975$ (Table 3.3). A similar study [27] on bus transportation network of six cities from different countries also observed similar results where their power-law parameter range is 3.92-5.66 and exponent parameter range is 38.7-225.0.

But we have to consider that P-Space does not reflect the original network structure since node connections are not consecutive links like in L-Space. So network structure and evolving behavior must be primarily observed in L-Space and then they should be confirmed with subsequent network analysis such as analysis of clustering coefficient in P-Space.

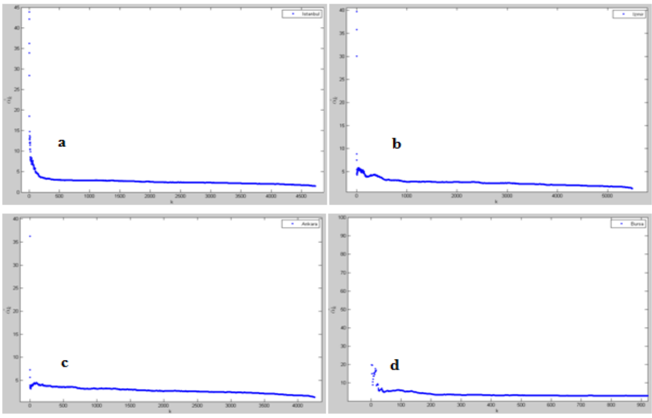


Figure 3.8 Hill plots of maximum likelihood estimates $\widehat{\alpha_k}$ as a function of $\boldsymbol{k}$ for İstanbul, İzmir, Ankara, and Bursa in P-Space.

Credibility of power-law in networks is also measured wit Hill plots where one checks for areas where the plot ‘settles down’ to some stable values of maximum likelihood estimates ($\hat{\alpha}$) away from small to large values on range k. Hill estimator $\hat{\gamma}$ is defined as follows.

$$\widehat{\gamma_k} = \frac{1}{k}\sum_{i=0}^{k-1} log \frac{d_{(N-i)}}{d_{(N-k)}} \text{ (a)} \qquad => \qquad \widehat{\alpha_k} = 1 + \hat{\gamma}_k{}^{-1} \text{ (b)} \tag{3.1}$$

where degrees of vertices are sorted as $d_1 \leq d_1 \leq \cdots \leq d_N$. Hill estimators are measured for chosen k values and plot is drawn with maximum likelihood estimator $\hat{\alpha}$ versus degrees.

Hill plots show that decay in İstanbul and Bursa is much more sharper in small range of values k ($8 \rightarrow 2$ and $20 \rightarrow 2$) than in İzmir and Ankara whose plots decay slightly from 5 to 2. Relation between degree distribution in L-Space (Figure 3.8.c) and hill plots of in P-Space confirm that power-law fits for all cities in L-Space is credible but power-law model has less relevance in Bursa than in the other cities.

In conclusion, İstanbul, İzmir, and Ankara have a scale-free network structure with power-law exponents $\gamma = 2.66 - 3.46$whereas Bursa is closer to a random network structure with a good fit to exponential node degree distribution and a high ($\gamma > 4$) power-law fitting parameter ($\gamma = 4.44$). Hill plots showed power-law credibility for İstanbul, İzmir, and Ankara and less credible results for Bursa corresponding with earlier degree distribution results.

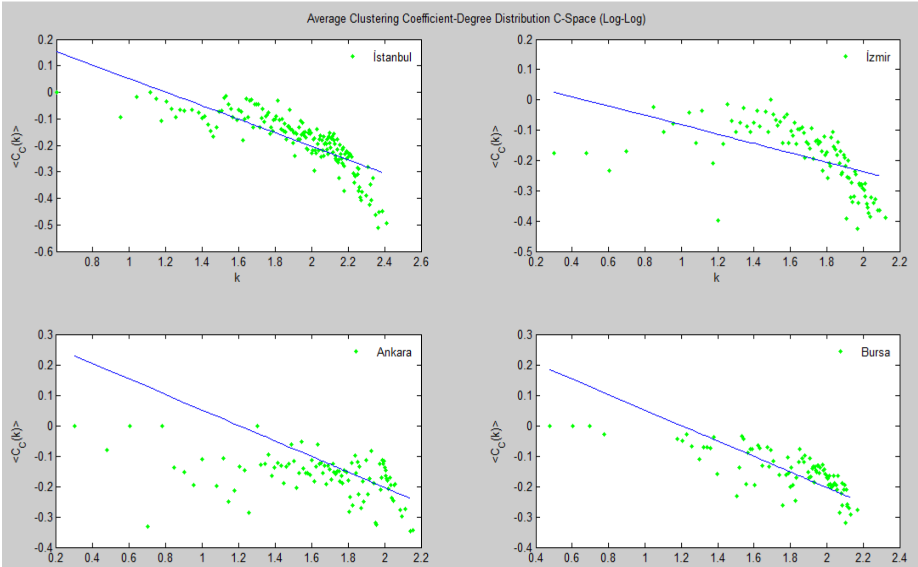
3.3.2 Clustering Coefficient

Clustering coefficients were first calculated locally (Eq. 2.15) and then network clustering coefficients were found on mean local clustering coefficients (Eq. 2.17). Also we compared whole network clustering coefficients of random networks (Eq. 2.20) that have the same network size with our results and showed that Turkish BTNs in this study are not random networks.

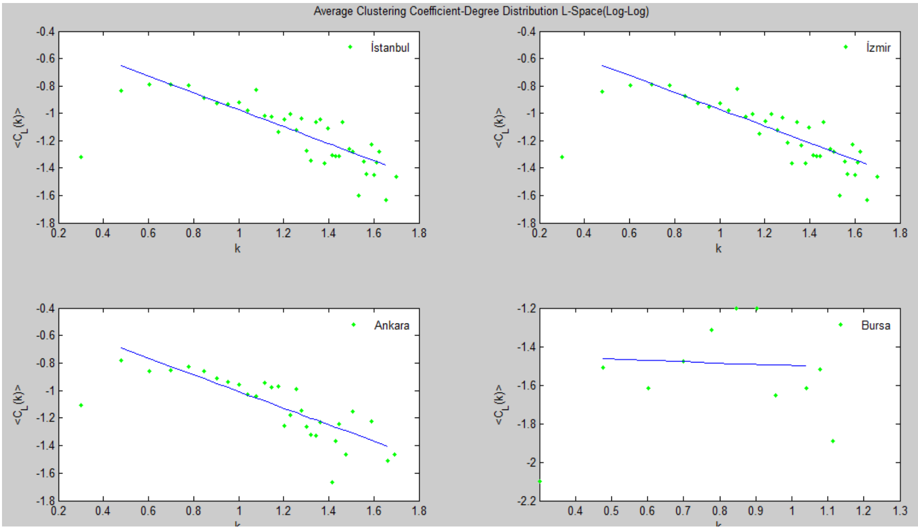
Table 3.4 Mean (global) clustering coefficients of bus networks ($\langle C \rangle$), comparative mean clustering coefficient ($\langle C_{ER} \rangle$) of Erdös-Rényi random network that same size nodes and edges with bus networks, clustering coefficient-degree correlation exponent (β) for bus networks from power-law distribution, and ratio ($C^{norm} = \langle C \rangle / \langle C_{ER} \rangle$ between mean clustering coefficient and Erdös-Rényi random graph of equal size in C-, L-, and P-Spaces.

	C-Space				L- Space				P- Space			
City	$\langle C \rangle_C$	$\langle C_{ER} \rangle_C$	β_C	C_C^{norm}	$\langle C \rangle_L$	$\langle C_{ER} \rangle_L$	β_L	C_L^{norm}	$\langle C \rangle_P$	$\langle C_{ER} \rangle_P$	β_P	C_P^{norm}
İstanbul	0.709	0.168	0.255	4.2	0.087	0.000712	0.614	122.2	0.779	0.026	0.639	30.0
İzmir	0.726	0.165	0.155	4.4	0.026	0.000438	1.346	59.4	0.821	0.028	0.609	29.3
Ankara	0.698	0.145	0.072	4.8	0.101	0.000743	0.604	135.9	0.833	0.019	0.647	43.8
Bursa	0.737	0.344	0.150	2.1	0.015	0.000631	0.067	23.8	0.768	0.057	0.533	13.5

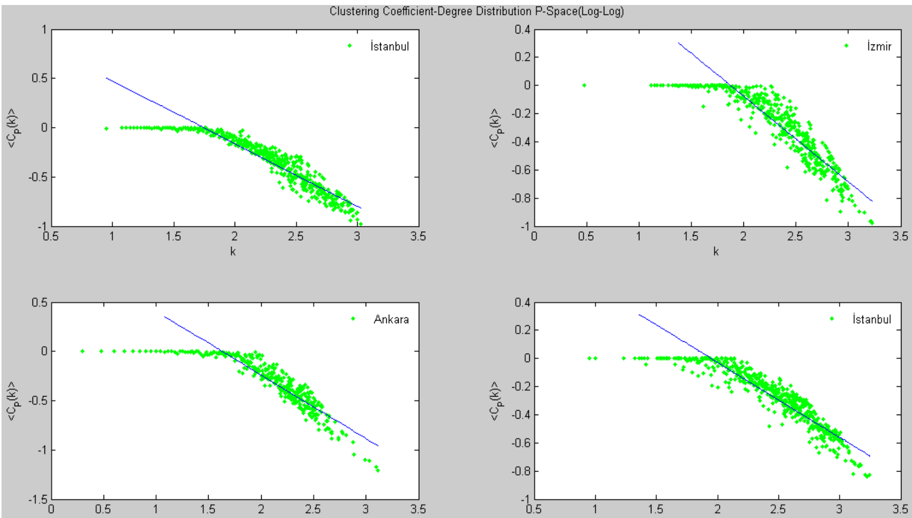
Table 3.4 shows the mean (global) clustering coefficient in C-, L-, P-Spaces for real-world bus networks, and Erdös-Rényi random graph, clustering coefficient-degree correlation exponent parameters (β), and normalized clustering coefficients. BTNs in P-Space have the highest mean clustering coefficient. This result is expected because of all bus-stops on the same route are fully connected with other bus-stops on that route in P-Space representation. The lowest clustering coefficients and the highest differences between clustering coefficient of an Erdös-Rényi random graph and real-world bus transportation networks (C^{norm}) are observed in L-Space. High and approximate clustering coefficient values and the lowest differences between normalized rates are observed in C-Spaces because of their lower node degree. This also shows that the routes are highly connected in all cities.



a



b



c

Figure 3.9 Average clustering coefficients of same degree nodes - degree k correlations for four cities in **(a)** C-, **(b)** L-, and **(c)** P-Spaces.

Strong correlations are observed between clustering coefficients and degree in P-Space while other topologies don't show a meaningful correlation (Figure 3.9). Correlation in P-Space is deduced when average clustering coefficient of all nodes with given degree k decays as a function of k and the decay distribution follows power-law. Correlation equation 3.2 reveals the inverse ratio where increasing node degree decreases mean clustering coefficient values.

$$\langle C(k) \rangle \sim k^{-\beta} \tag{3.2}$$

Fitted power-law exponent parameters are shown in Table 3.4 and a similar study [27] found exponent ranges in 0.65-0.96. Only Bursa is not in this range (0.533). Bursa's weak clustering behavior reflects its irregular network structure and agrees with the previous node degree analysis where its log-linear plot in L-Space indicated an exponential decay.

3.3.3 Assortative Mixing

Connectivity tendency of vertices in BTNs was measured with assortativity values (Eq. 2.21) and degree-degree correlations for each city were computed.

Table 3.5 Assortativity mixing values of whole networks in C-, L-, and P-Spaces are showed within nearest (r^1) and next nearest neighbor (r^2) connections.

	C-Space		L- Space		P- Space	
City	r_C^1	r_C^2	r_L^1	r_L^2	r_P^1	r_P^2
İstanbul	0.210	-0.139	0.146	0.124	-0.097	-0.008
İzmir	0.211	0.009	0.108	0.778	-0.016	0.072
Ankara	0.434	0.004	0.105	0.226	-0.055	0.284
Bursa	0.204	-0.205	0.295	0.597	-0.022	-0.014

Table 3.5 shows that nearest neighbor assortativity (r^1) in C- and L-Spaces reflects the correlations between nodes of similar degree (assortative) and in P-Space the correlation between nodes of different degree (disassortative). In C-Space nearest assortativity values $r_C^1 = 0.2 - 0.4$ with values shown in previous studies $r_C = 0.1 - 0.5$ [27]. r_C^1 shows similarity in İstanbul, İzmir, and Bursa but Ankara value is double of values in Table 3.5 that indicates a high number of similar (high) degree nodes (routes) in Ankara. In L-Space r_L^1 values are similar in İstanbul, İzmir, and Ankara but in Bursa they are almost three times bigger than values in Table 3.5

that also shows density of similar (low) degree nodes. Degree assortativity values, $r_L^1 = 0.1 - 0.3$, are similar to [32] and these small assortative values also indicate a finite preference for assortative mixing. This and former studies demonstrate that assortativity values in L-Space are independent from network size and they are always positive. This can be explained as appearance of a few number of nodes with high degrees (hubs) and they are linked among themselves, and the majority of remaining nodes with small degrees are connected among themselves in networks. Disassortativity in P-Space is observed with very small r_P^1 values between -0.016 and -0.097 that show no preference linkage based on node degree. This also shows an existence of very high degree nodes (hubs) and tendency of small degree nodes toward these hubs in the network. We observed the highest nearest absolute assortativity values in C-Space and the lowest values in P-Space.

Next nearest neighbors' state enlarges connected nodes and degree number of each node. In this state for all topologies, assortativity values did not reveal a uniform structure; there are increases, decreases, and state changes in degree-degree correlations that are shown in Table 3.5. In C-Space assortativity values decrease and İstanbul and Bursa turned to disassortativity from assortativity. Next nearest assortativity in Ankara and İzmir decreased and indicated a randomly mixed state. In L-Space, İzmir remained assortative but İzmir, Ankara, and Bursa increased their assortativity while İstanbul's value decreased. Next nearest assortativity in P-Space indicates randomly mixed structure for İstanbul and Bursa but İzmir and Ankara r_P^2 values show a state change from disasortativity to assortativity. Ankara does not show a linkage preference based on node degree.

3.3.4 Shortest Path

Average and maximum shortest path values are shown in Table 3.6 for all topologies whereas Table 3.7 gives the proportions of certain path lengths in C- and P-Space.

Maximum shortest path length values change in 4-8 range in C-Space and 5-9 range in P-Space. These characteristic values can be read as the length of travelling between two different points. Minimum of maximum shortest path length values are observed in İstanbul and maximum values are observed in İzmir for C- and P-Spaces. These values show that one can travel between any two points with maximum 5 steps for İstanbul and 9, 7, and 7 steps for İzmir, Ankara, and Bursa respectively.

Table 3.6 Average and maximum shortest path length in C-, L-, and P-Spaces for BTNs of four cities; N: number of bus stops; R: number of bus routes; F: number of bus stop per a route (route density).

	Average Shortest Path Length(‹l›)						Maximum Shortest Path Length(l^{max})		
City	N	R	F	C-Space	L- Space	P- Space	C-Space	L- Space	P- Space
İstanbul	4726	529	8.9	2.068	10.294	2.858	4	53	5
İzmir	5489	295	18.6	2.253	46.436	3.246	8	212	9
Ankara	4249	376	11.3	2.365	9.629	2.940	6	55	7
Bursa	3914	206	19	1.813	29.001	2.456	6	126	7

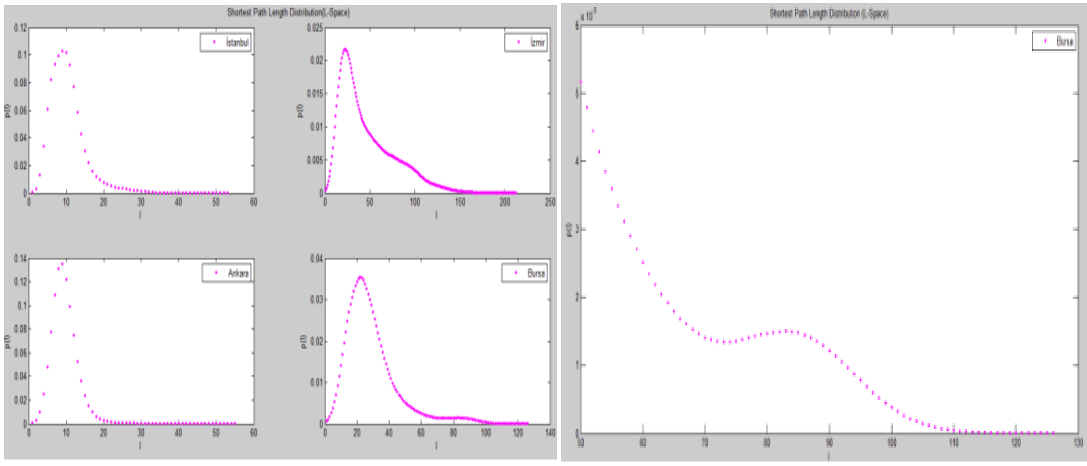
In L-Space we observed very high values for average and maximum shortest path length than in C- and L-Spaces. Maximum shortest path length values in İstanbul and Ankara with 53 and 55 steps are relatively low compared to values in İzmir and Bursa with 212 and 126 respectively. Small average shortest path distances and high clustering coefficients in C-, L-, and P-Spaces signal the small-world network structure for BTNs of our cities.

Table 3.7 Shortest path lengths and proportions in C- and P-Spaces for BTNs of four cities.

	Cities							
	C-Space Proportions				P-Space Proportions			
Shortest Path Lengths	İstanbul	İzmir	Ankara	Bursa	İstanbul	İzmir	Ankara	Bursa
1	0.168	0.16784	0.14508	0.34606	0.026	0.0293	0.018524	0.05665
2	0.603	0.51948	0.41809	0.52678	0.294	0.2562	0.275549	0.541451
3	0.221	0.23196	0.36767	0.10064	0.483	0.3511	0.474157	0.317161
4	0.008	0.05975	0.06553	0.02070	0.193	0.2352	0.211443	0.060463
5		0.01470	0.00359	0.00578	0.005	0.0815	0.019485	0.022754
6		0.00559	0.00004	0.00005		0.0294	0.000834	0.001515
7		0.00066				0.0137	0.000007	0.000006
8		0.00002				0.0036		
9						0.0003		

There is no evidence to show a relation between average and maximum shortest path length (l^{max}) and number of bus stations, bus routes, and route density in network. Although Bursa has high route density (F) with low number of routes, its average shortest path length is the lowest and reaching each node is easier. Efficient usage of routes to connect the bus stations caused easier linkage in Bursa whereas İzmir shows an ineffective design of routes with high route density and high average shortest path length. Greater values in L-Space for İzmir and Bursa show an existence of several non-overlapping connections between routes. Bus stations spread over a wide area because of new settlements and geographic limitations on growth due to nearby mountains.

Mean shortest path lengths between vertex pairs were also obtained using Eq. 2.23 and results are shown in Figure 3.10.



a **b**

Figure 3.10 (a) Mean shortest path length distributions of BTNs in L-Space for İstanbul, İzmir, Ankara, and Bursa. **(b)** Second local maximum in Bursa.

Plots show peak values around the average shortest path lengths of networks and *p(l)* values decrease with increasing path values. Bursa and İzmir have a second local maximum. Second local maximum in path length distribution points to separate communities. In the map of İzmir (Figure 3.1) one can see that two communities are divided geographically by a cove. In Bursa, new settlements created a separate community from dense and old settlements in the center. We expected to see more than one community in İstanbul that is geographically divided into European and Asian sides via Bosphorus, but there is no evidence for this. One can say that there is unity in the BTN of İstanbul despite the two sides of Bosphorus are separated by sea.

Plots in Figure 3.10 show an asymmetric and bimodal (Bursa and İzmir) structure that is in line with the Lavenberg-Marquardt method with its function as;

$$p(l) = Ale^{-Bl^2+Cl} \tag{3.3}$$

where A, B, and C are fitted parameters. Lavenberg-Marquardt method doesn't show any correlation between degree and shortest path length distribution and is used to define fitted curves.

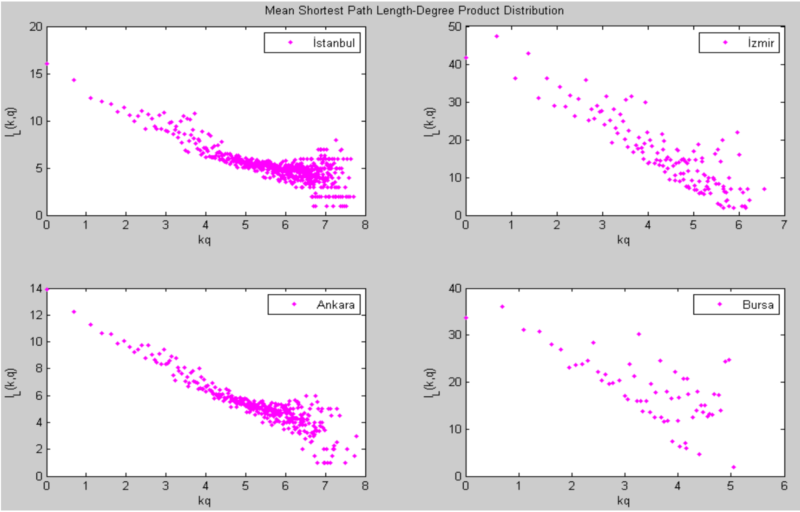


Figure 3.11 Mean shortest path ($l_L(k,q)$) - degree product (kq) of two end nodes in BTNs of four cities.

Through shortest path length studies one can also relate node degree and shortest path length of vertex pairs. Relation between mean shortest path and node degrees (k-q) can be approximated by,

$$l(k,q) = A - B\log(kq) \tag{3.4}$$

Equation 3.4 says that mean shortest path length of two different end nodes with degrees k and q is proportional to the logarithm of their degree product. In Figure 3.11, we do not observe a clear linear correlation between average shortest path lengths on degree product of end nodes k-q in L-Space. Equation (3.4) cannot be used to show correlation in C- and P-Spaces because of small values of shortest path length ($\langle l_C \rangle = 4 - 8, \langle l_P \rangle = 5 - 9$).

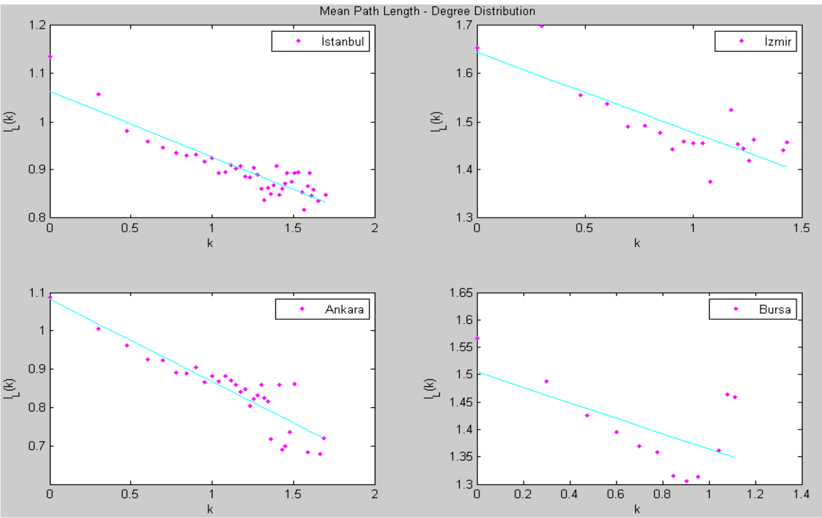


Figure 3.12 Average path length $l(k)$ on the degree of a single end node k.

Also we can define mean shortest path on degree k of a single end node ($l(k)$) in L-Space and fits to power law as,

$$l(k) \sim k^{-\alpha} \tag{3.5}$$

Within our study we observed approximate fitting between average shortest path on node degree with exponents İstanbul=0.136, İzmir = 0.166, Ankara = 0.215, and Bursa = 0.141. Exponents of previous studies were also measured in the range from 0.17 to 0.27 [32]. These approximate relation results (especially in İstanbul and Ankara) are indicative of a scale-free network. We also did not find any linear relation between node degree and path lengths. Separated points in Figure 3.12 (b) and (d) denote the other communities in BTNs of İzmir and Bursa that we also described under assortativity. We could not investigate the relation between mean shortest path and node degree for two- and one-sided end nodes and see whether scale-free and exponential decay behavior were present in C- and P-Spaces due to narrow shortest path ranges which give only a very limited number of points to draw these relationships from.

As a result we uncovered scale-free relation between average lengths of shortest paths of end node degrees for four cities in L-Space that indicates preferentially growing networks. Also, results don't yield linear path length-degree correlations as $l(k) \sim a - bk$ that shows a random network structure and confirm previous results.

3.3.5 Centrality

Centrality study in BTNs of cities was aimed to show the degree-mean centrality correlation and demonstrate the validity of real-world network properties on betweeness centrality for all cities.

Eigenvector centralities of vertices in networks were measured on Eq. 2.12 and observed correlation between average eigenvector centrality and node degree as,

$$\langle C^e(k) \rangle \sim k^{\vartheta} \tag{3.6}$$

Mean eigenvector centrality on degree –degree correlation can be seen in C-, L-, and P-Spaces (Figure 3.13, Table 3.8) where BTNs of İzmir and Bursa show the highest fitting. Fitting parameters in all space don't show the high differences that they change in 1.05-2.22 intervals. These correlations in all spaces also extend to high degree nodes have high statues or importance in network description for eigenvector centrality.

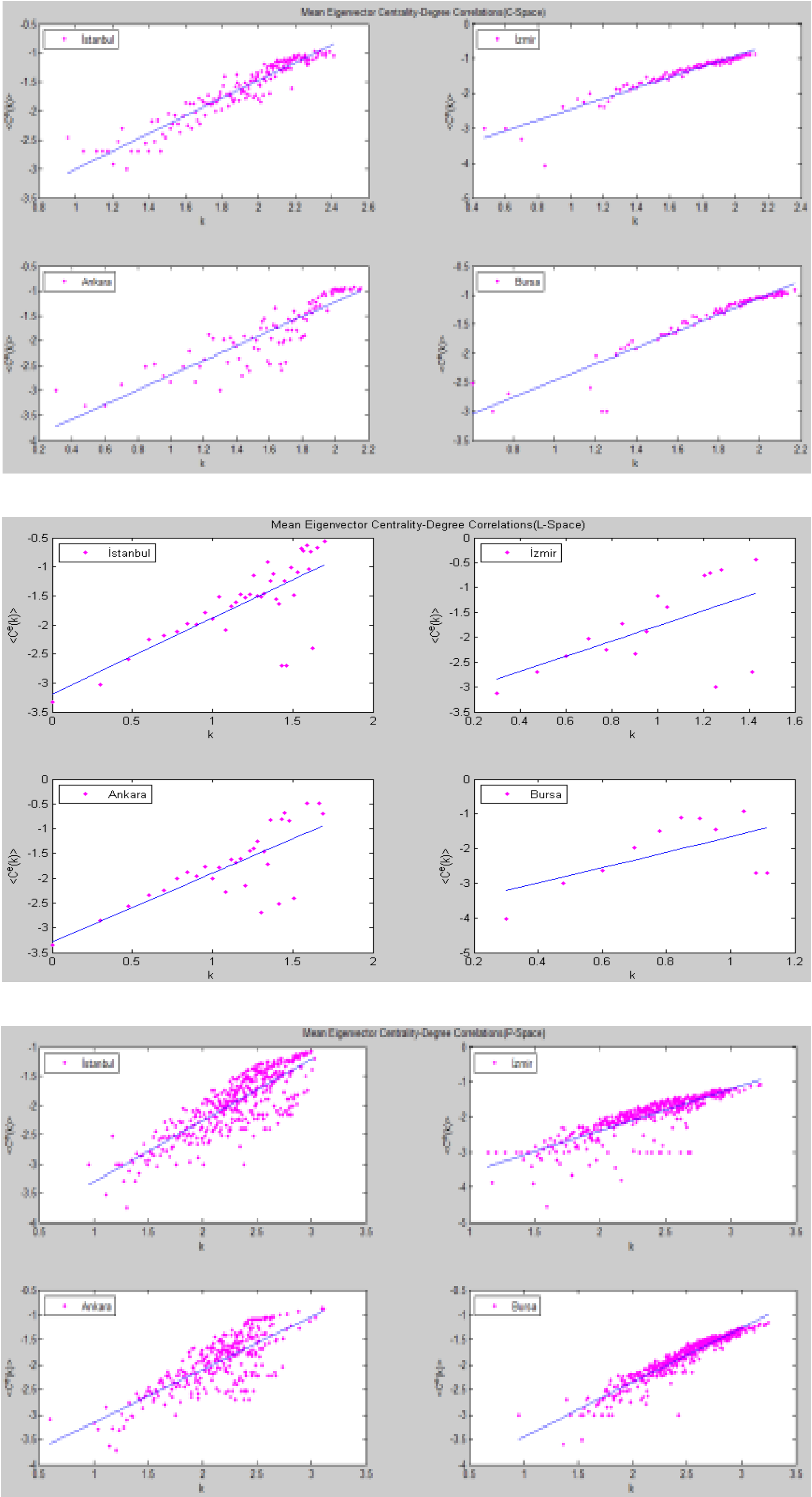


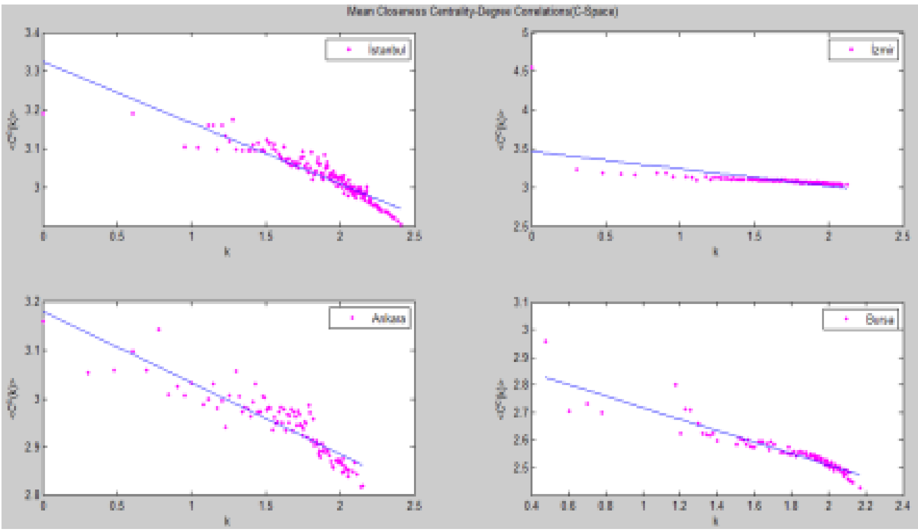
Figure 3.13 Mean eigenvector centrality – degree correlations for four cities in **(a)** C-, **(b)** L-, and **(c)** P-Spaces respectively.

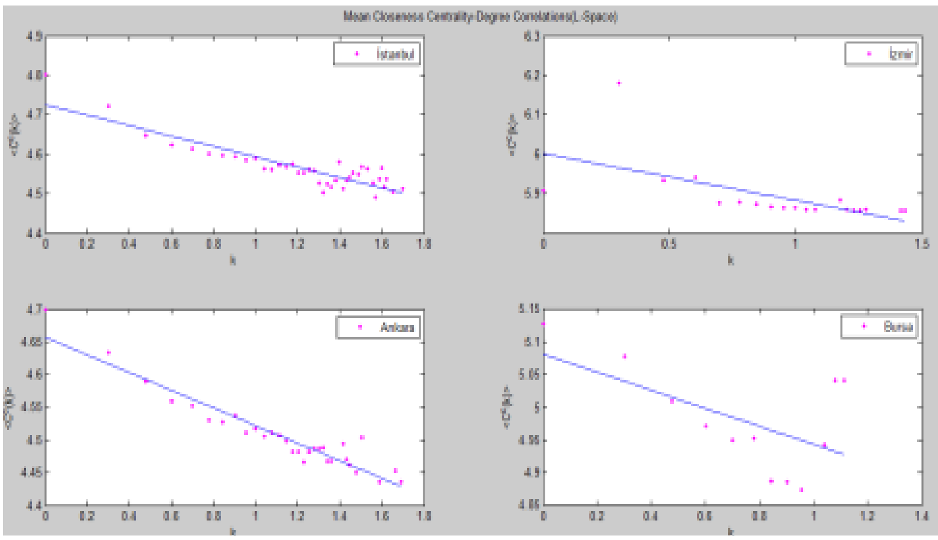
Table 3.8 Fitting parameters of the mean eigenvector centrality-degree on power law for each city in C-, L-, and P-Spaces.

	Eigenvector Centrality Exponents		
City	C-Space	L- Space	P- Space
İstanbul	1.523	1.309	1.054
İzmir	1.530	1.522	1.185
Ankara	1.472	1.393	1.064
Bursa	1.423	2.223	1.090

Closeness centralities were calculated using Eq. 2.13 and Figure 3.14 and values in Table 3.9 show degree-mean closeness correlations. Correlation between average closeness centrality and node degree can be expressed as an inverse relationship via a power-law distribution.

$$\langle C^c(k) \rangle \sim k^{-\omega} \tag{3.7}$$





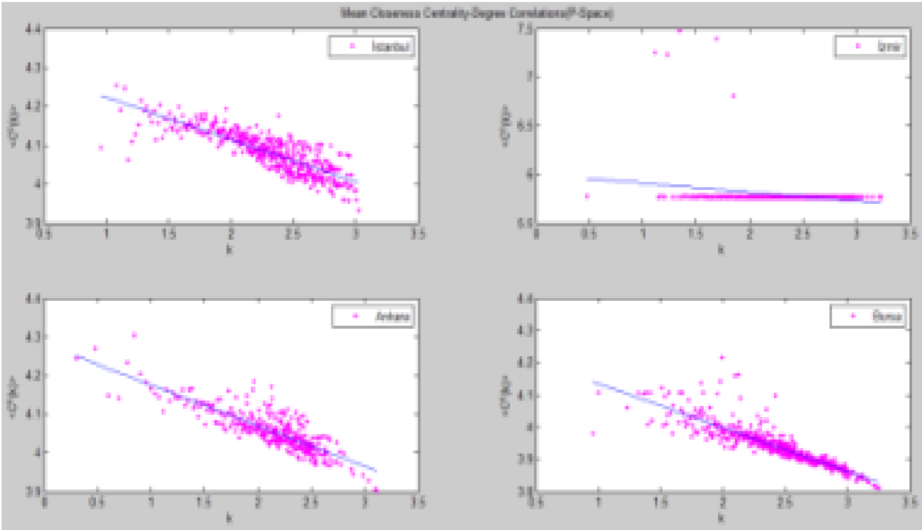


Figure 3.14 Mean closeness centrality – degree correlations for four cities in **(a)** C-, **(b)** L-, and **(c)** P-Spaces respectively.

Table 3.9 Fitting parameters of the mean closeness centrality-degree on power law for each city in C-, L-, and P-Space.

	Closeness Centrality Exponents		
City	C-Space	L- Space	P- Space
İstanbul	0.157	0.132	0.108
İzmir	0.223	0.119	0.091
Ankara	0.147	0.135	0.105
Bursa	0.208	0.138	0.136

In general, all topologies (P-Space is highest) conform to the expected inverse relation between degree and mean closeness centrality. High degree nodes have many contacts within shortest path length and they are quickly reachable. İzmir differs from other cities with a stable mean closeness for high degrees that shows the independence of closeness from node degree and a highly connected network with similar path lengths.

Betweenness centrality calculations were made with Eq. 2.14 and correlations can be seen in Figure 3.15 and Table 3.10. Correlation between average betweenness centrality and node degree was again fit to a power-law distribution.

$$\langle C^b(k)\rangle \sim k^{\varphi} \quad (3.8)$$

Bursa does not have a good power-law fit in L-Space and shows a random distribution. The indication is that there is no central point where traffic intersects. Thus, Bursa has the highest resilience against removing bus stations.

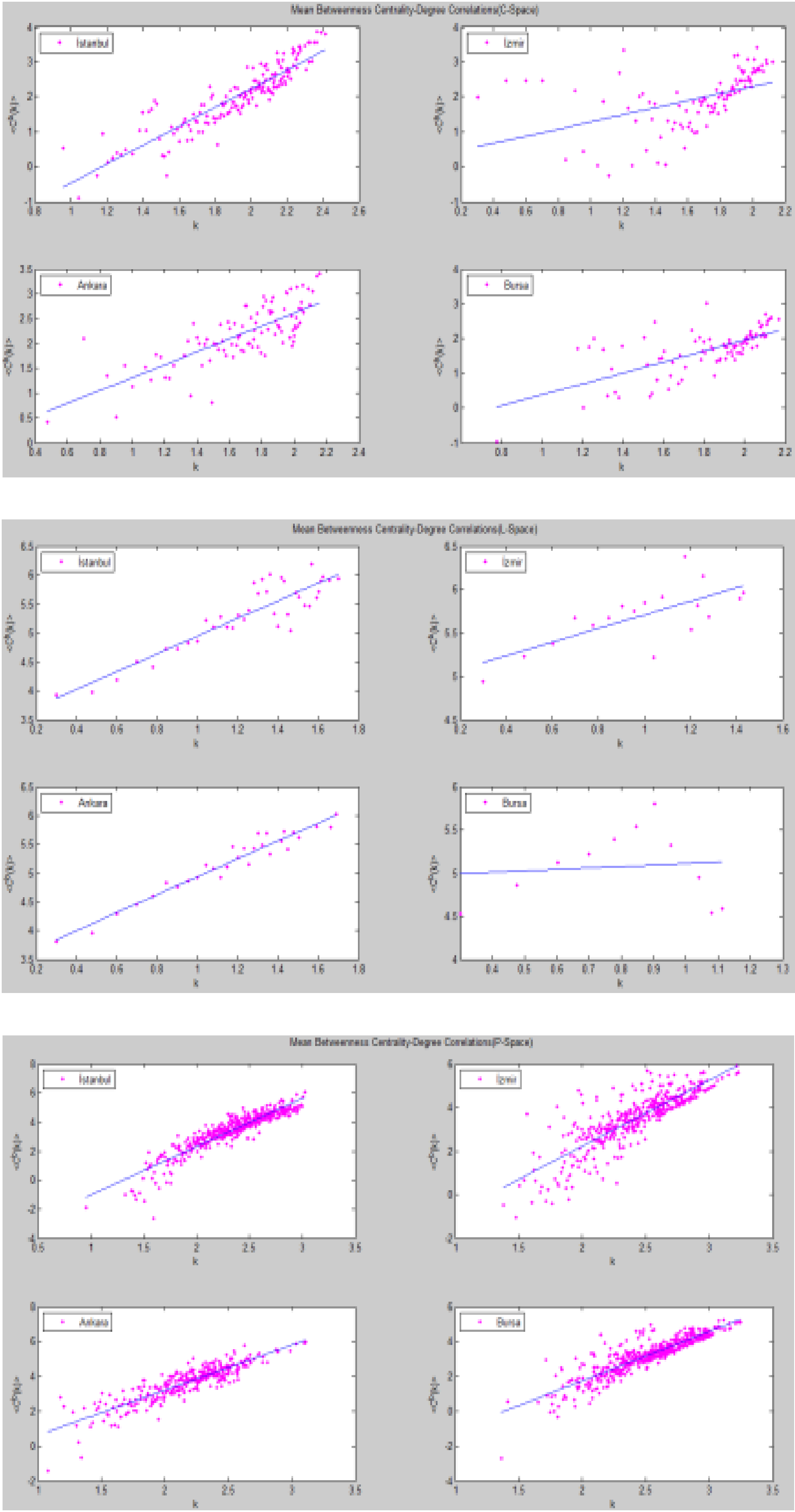


Figure 3.15 Mean betweenness centrality – degree correlations for four cities in **(a)** C-, **(b)** L-, and **(c)** P-Spaces respectively.

Table 3.10 Fitting parameters of the mean betwenness centrality-degree on power law for each city in C-, L-, and P-Space.

	Betweenness Centrality Exponents		
City	C-Space	L- Space	P- Space
İstanbul	2.69	1.53	3.31
İzmir	1.02	0.78	3.03
Ankara	1.30	1.56	2.59
Bursa	1.58	0.17	2.83

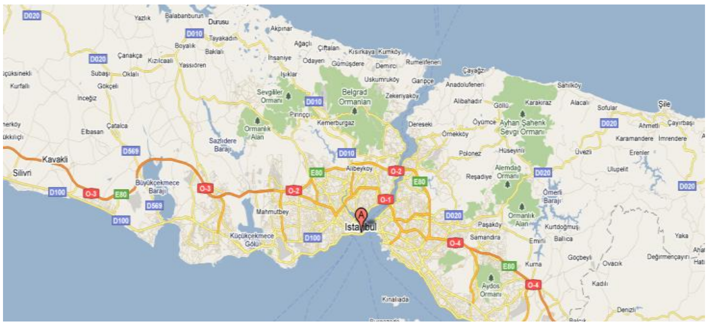
In summary, centrality-degree plots and calculations provided good-fits to power-law distribution showing the importance of high degree nodes in the networks. Nodes in Figure 3.13, 3.14, and 3.15 of C- and P-Space that are further off from fitted lines show existence of sub-networks (communities). These visual observations also conform to assortativity mixing and shortest path length distribution results.

3.4 Further Studies for İstanbul

İstanbul is the most populous (13 million – 18% of Turkey's population) and economically most important city of Turkey. It is the world's 34th largest economy and also Europe's most populous city taking into account the municipal boundaries. İstanbul is a transcontinental city divided into European (Rumeli) and Asian (Anatolia) sides via Bosphorus Channel. There are 39 districts (14 on Asian side and 25 on European side) within the municipal boundaries.

Public transportation in Istanbul comprises of a bus network (bus, metrobus), rail systems (light rail, metro, trams, suburban trains), funiculars, maritime services (ferryboats, sea bus), and cable cars. 3.5 million people are transported daily with 4,891 bus and metrobuses on around 529 routes and 10,272 bus stops. With 29.13% of all transportation, bus transportation has an important role in public transportation with high transportation capacity, bus fleet, route and bus stop numbers.

In this section, we analyze network effects of some new and proposed projects in İstanbul that caught a lot of media attention. Metrobus is a recent project that crosses Bosphorus using the Bosphorus Bridge and reaches the Western boundaries of the city with newly built dedicated lanes. The project dramatically cut down the travel times between Asia and Europe in both directions. The approved third bridge on Bosphorus has long been a much-debated topic. We compared statistical results of the old bus network, currently existing network with metrobus, and a hypothetical network with a third bridge between Garipçe and Poyrazköy (Figure 3.16). Metrobus fleet comprises of 848 busses servicing Avcılar-Zincirlikuyu (route 34), Avcılar-Topkapı (route 34T), and Söğütlüçeşme-Edirnekapı (route 34A) with 98 bus stops (Table 3.11). Third bridge that will be built between Galipçe and Poyrazköy will connect northern districts of İstanbul on both sides. We defined several new connections between routes and bus-stops of the existing structure for C- and P-Space. In L-Space, a new network could not be formed due to the absence of detailed bus-stop and route information that can only be available when the bridge is built and bus transportation starts. Changes in the statistical values of three networks were observed for node degree distribution, clustering coefficient, degree-degree correlation, path length distributions, and centralities.



a

b

c

Figure 3.16 (a) İstanbul and Bosphorus and **(b)** existing bridge traffic and **(c)** metrobus and third bridge routes.

Table 3.11 N: number of bus stops; R: number of bus routes; F: mean number of bus stations per route.

City	N	R	F
İstanbul Metrobus	4726	529	8.9
İstanbul Non-Metrobus	4716	526	9.0
İstanbul Third Bridge	4726	530	8.9

Table 3.12 shows changes in number of routes and bus-stops. For metrobus, we added three new routes and ten new bus-stops and some existing bus-stops on main lines were also adapted to new routes. For third-bridge analysis, we defined one new route that is described in Figure 3.16 (c) and added some new bus stops towards both ends of the route connecting them to existing bus-stops.

Table 3.12 Comparison of exponent values of exponential ($\bar{k}$) and power-law (γ) degree distributions and Pearson correlation coefficients (R) for L- and P-Spaces for three İstanbul BTNs.

	L-Space				P-Space			
City	$\bar{k}_L$	$R_L^{\langle k \rangle}$	γ_L	R_L^{γ}	$\bar{k}_P$	$R_P^{\langle k \rangle}$	γ_P	R_P^{γ}
İstanbul Metrobus	6.671	0.705	2.660	0.995	153.639	0.968	1.873	0.998 (%92)
İstanbul Non-Metrobus	6.663	0.706	2.661	0.995	153.671	0.968	1.883	0.998 (%91)
İstanbul Third Bridge	-	-	-	-	206.488	0.961	1.509	0.978 (%92)

Table 3.12 shows fitted parameters and Pearson correlation coefficients for node degree distributions. Metrobus has no significant effect on log-linear plot that is observed exponential decay behavior. There is no change in the decay of node degree distribution with a good exponential fit in log-linear scale. In log-log scale, good power–law fit with similar exponent values were again observed and Pearson correlation coefficient exceeded $R_L^{\gamma} = 0.995$. Power-law exponent parameters $\gamma_{L=}$2.66 ($2 < \gamma_L$ <3) indicates existence of robustly connected nodes and high resilience against random removal of nodes. Furthermore, values close to 3 reflect preferential network growth.

Cumulative node degree distributions in P-Space are plotted in semi-log and log-log scales. Fitted distribution parameters are shown in Table 3.12. Results are similar in that they all show exponential decay while İstanbul Third Bridge does not have a good fit.

Table 3.13 Bus network properties for C-, L-, and P-Spaces. ‹k›: average node degree of nearest neighbors for whole network; ‹z›: average node degree of next nearest neighbors for whole network; δ: rate between ‹z› and ‹k› ($\delta = ‹z›/‹k›$); k^{max}: maximum node degree in space; ‹C› mean (global) clustering coefficient of whole network; r^1: assortativity mixing of nearest neighbor; ‹l› : average shortest path length ; l^{max}:maximum shortest path length in network.

	C-Space								L-Space								P-Space							
City	$‹k_C›$	k_C^{max}	$‹c_C›$	$r^1{}_C$	$‹l_C›$	l_C^{max}	$‹z›_C$	δ_C	$‹k_L›$	k_L^{max}	$‹c_L›$	$r^1{}_L$	$‹l_L›$	l_L^{max}	$‹z›_L$	δ_L	$‹k_P›$	k_P^{max}	$‹c_P›$	$r^1{}_P$	$‹l_P›$	l_P^{max}	$‹z›_P$	δ_P
İstanbul Metrobus	88.96	258	0.709	0.210	2.068	4	318.33	3.58	3.367	50	0.087	0.146	10.29	53	15.62	4.64	121.043	1064	0.779	-0.097	2.86	5	1387.619	11.464
İstanbul Non- Metrobus	87.31	242	0.71	0.23	2.084	4	310.42	3.56	3.360	50	0.087	0.145	10.81	55	15.51	4.62	121.210	1051	0.779	-0.097	2.86	5	1377.877	11.368
İstanbul 3	89.18	258	0.70	0.210	2.060	4	321.41	3.60	-	-	-	-	-	-	-	-	186.496	1511	0.796	0.287	2.64	5	1826.782	9.795

Table 3.14 Mean (global) clustering coefficients of bus networks (‹C›), comparative mean clustering coefficient (‹C_{ER}›) of Erdös-Rényi random network that same size nodes and edges with bus networks, and clustering coefficient-degree correlation exponent (β) for bus networks from power-law distribution in C-, L-, and P-Spaces.

	C-Space				L- Space				P- Space			
City	‹C›$_C$	‹C_{ER}›$_C$	β_C	C_C^{norm}	‹C›$_L$	‹C_{ER}›$_L$	β_L	C_L^{norm}	‹C›$_P$	‹C_{ER}›$_P$	β_P	C_P^{norm}
İstanbul Metrobus	0.709	0.168	0.255	4.2	0.087	0.000712	0.614	122.2	0.779	0.026	0.639	30.0
İstanbul Non-Metrobus	0.709	0.166	0.260	4.3	0.087	0.000712	0.616	122.2	0.779	0.026	0.640	30.0
İstanbul Third Bridge	0.701	0.169	0.260	4.1	-	-	-		0.796	0.039	0.370	20.4

Clustering analysis is conducted in C- and P-Space for three İstanbul networks and in L-Space for İstanbul Current and Non-Metrobus systems. We don't observe meaningful differences in C-, L-, and P-Space analyses due to metrobus addition. As path length and clustering values remain the same system unity of BTN does not increase. Third bridge results are only given for C- and P-Space because currently future locations of the bus-stops are not known. We observed weakness in C-Space clustering due to the distance of the defined third bridge route with existing routes. But weak route connection increased transportation between far and disconnected bus-stops in north districts of two sides (Asian-European) and increased global clustering in P-Space.

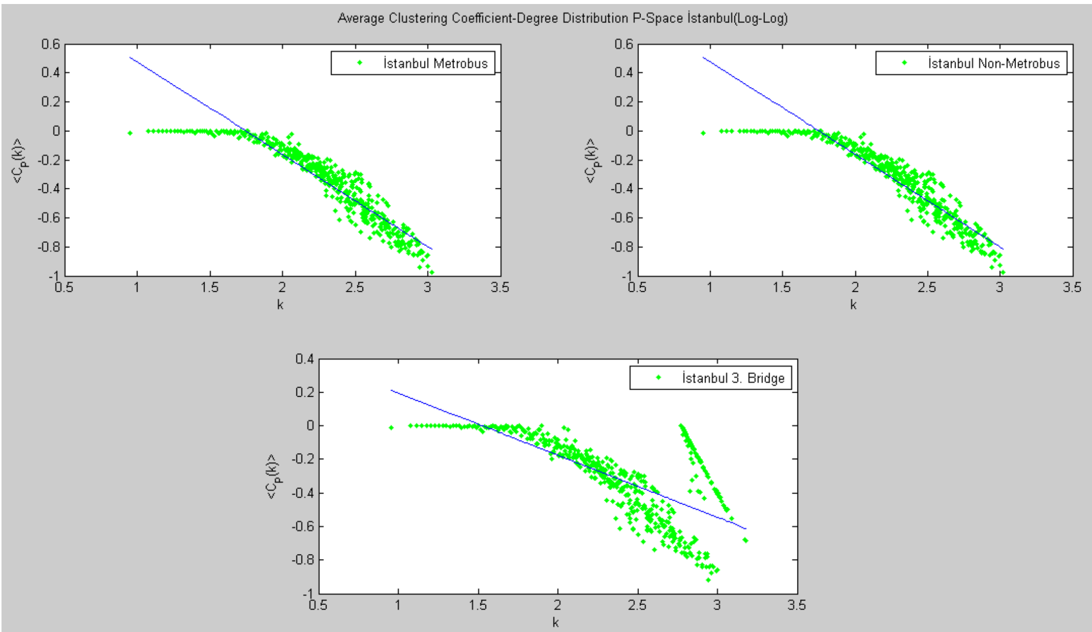


Figure 3.17 Average clustering coefficients of same degree nodes - degree k correlations for İstanbul in C-, L-, and P-Space.

Clustering coefficient – degree correlation plot for İstanbul third bridge network in P-Space emphasizes the distant third bridge route (Figure 3.17c).

Table 3.15 Assortativity mixing values of whole networks in C-, L-, and P-Spaces are showed within nearest (r^1) and next nearest neighbor (r^2) connections.

	C-Space		L- Space		P- Space	
City	r_C^1	r_C^2	r_L^1	r_L^2	r_P^1	r_P^2
İstanbul Metrobus	0.210	-0.139	0.146	0.124	-0.097	-0.008
İstanbul Non-Metrobus	0.229	-0.137	0.145	0.128	-0.097	-0.006
İstanbul Third Bridge	0.214	-0.129	-	-	0.287	-0.050

Table 3.15 gives a comparison of assortativity mixing values in C- and P-Space for three İstanbul networks and in L-Space for current and Non-Metrobus networks. In L-Space metrobus does not make any change and the network stayed assortative which is also true for P-Space. But assortativity status is changed from disassortative to assortative with the third bridge (Table 3.15). Third bridge increases connections between all bus-stops as nodes with high degrees on both sides get connected. Assortativity in C-Space did not change since adding only one distant route does not impact the network much. However, metrobus route created new high degree routes and decreasing assortativity values of same state.

Table 3.16 Average and maximum shortest path length in C-, L-, and P-Spaces for BTNs of four cities; N: number of bus stops; R: number of bus routes; F: number of bus stop per a route (route density).

	$\langle l \rangle$			l^{max}		
City	C-Space	L- Space	P- Space	C-Space	L- Space	P- Space
İstanbul Metrobus	2.068	10.29	2.86	4	53	5
İstanbul Non- Metrobus	2.084	10.81	2.86	4	55	5
İstanbul Third Bridge	2.060	-	2.64	4	-	5

Effects on path lengths are shown in Table 3.16. Adding the new routes for metrobus and third bridge increased connections and decreased number of transits between routes in C-Space. Thus average path lengths between routes decreased. A greater decrease is observed in the third bridge network. The third bridge route connects routes on both sides that were not connected in the old BTN while metrobus connects routes that were also connected in the former network. Comparison of İstanbul Current and Non-Metrobus networks for average shortest path length in L-Space shows a decrease in average shortest path length with metrobus. In P-Space there are no differences between the two networks. However, a decrease in is observed with the construction of the third bridge. Via the third bridge distant points of both sides get connected which decreases mean shortest path. Maximum path values in Table 3.16 show that there are no differences in C- and P-Space but a decrease in L-Space. From P-Space results we observe that one can reach any point in İstanbul with maximum 5 transits.

Chapter4
Conclusion

In this study we conducted a thorough statistical analysis for bus transportation networks (BTNs) of four largest cities (İstanbul, İzmir, Ankara, Bursa) in Turkey in C-, L-, and P-Space. In L- and P-Space network sizes are dictated by the number of bus stops which were 4726, 5489, 4249, and 3914 respectively. In C-Space defined by route-route relations, the numbers of bus routes were 529, 295, 376, and 206 respectively. Node degree distributions of all cities in L-Space showed a good fit to power law distribution with exponents γ ranging from 2.66 to 4.44. In P-Space, cumulative degree distributions show exponential decay with $\bar{\mathrm{k}} = 101.7 - 239.9$. Node degree analysis in L- and P-Space shows that BTNs of İstanbul, İzmir, and Ankara are scale-free. While Bursa's growth seems to be random, new nodes are added preferentially in the other three cities. Global (mean) clustering coefficient analysis gave the highest values in P-Space as $\langle C \rangle_P = 0.78 - 0.83$ while L- and C-Space ranges were $\langle C \rangle_L = 0.08 - 0.10$ and $\langle C \rangle_C = 0.70 - 0.74$. Results show bus network unity for all four cities. Also power-law distribution models correlation between average clustering coefficient for all nodes with given degree k well with exponent values $\beta = 0.53 - 0.65$in P-Space. Assortativity analysis showed that the networks were assortative in C- and L-Space but disassortative in P-Space. We observed no relation of assortativity with other network properties. Shortest path length distribution in L-Space is described in Lavenberg-Marquardt method with $p(l) = Ale^{-Bl^2+Cl}$. Small shortest path length values that reflect small-world behavior were observed in all topologies and they were more apparent in P-Space. No correlations between mean shortest path length and the end node degrees $(l(k,q) - kq)$ were found. L-Space studies revealed correlation between mean shortest path lengths for single end nodes. Degree, closeness, eigenvector, and betweenness centralities were also analyzed. We observed that average centralities among nodes with degree k and node degree correlations fit power-law distributions.

Further studies for İstanbul looked at the effects of metrobus and third bridge construction on existing BTN of İstanbul. Metrobus does not result in significant changes on local and global network characteristics. These results show that the main purpose of metrobus was not to increase connection between routes and bus-stops but it aimed to increase the number of people carried on Avcılar-Zincirlikuyu, Avcılar-Topkapı, and Söğütlüçeşme-Edirnekapı routes, decrease transport times and save fuel. The third bridge to be built between Poyrazköy (in Asia) and Garipçe (in Europe) will create a new route(s) that will connect very distant points of the northern districts on both sides of İstanbul. BTN with the third bridge route added revealed important differences in local and global characteristic of system in C- and P-Space. Connections between routes and bus-stops on two sides increased. Node degrees and global clustering coefficients also increased. Moreover, assortativity mixing changed state from disassortative to assortative, and travelling steps (path length) decreased with additional linkage of northern districts. Eventually third bridge construction will increase network unity and decrease travel times in İstanbul.

Overall, BTNs of all four cities are real-world networks which show good degree fits to power-law distribution (“scale-free”), high clustering (“small world”), low average path length (“small world”), good fits of betweenness centrality-degree correlations to power-law distributions (“scale-free”).

Statistical analysis of networks enables us to analyze several network properties for BTNs of cities. It also helps us to measure the impacts of newly added routes and bus-stops to a certain degree. However, in public projects several interesting questions can arise such as how to add routes while minimizing cost, reducing greenhouse effects, or reducing travel time. Using statistical analysis results of public transportation systems to answer questions about economical and environmental problems of municipalities and governments is an interesting topic for future research.

Appendix A Dictionary

Bipartite graph: Vertices are divided into 2 subsets, and they don't have any edges linking them in their set.

Circuit: A path that starts and ends at the same vertex in network.

Clique (clustering): Highly connected networks where each vertex connect to every other node in other word connection between any node pairs occur in a single step and each node connects to every other nodes. For social network model, a group of people who all know each other.

Complete graph: The total number of vertices in the set of G(V,E) is called the size of the graph (N) and the total number of edges in the set of $E(M)$. Maximum number of edges in graph equal to $N(N-1)/2$, if all possible edges are in connection, graph is called as "complete graph".

Component: A subset of vertices in the graph that they can be reachable from the other vertices on some path through the network.

Connected graph: Each vertex of a network has a path to all other vertices (reachable) in network that any two vertices are attached by a path(no matter how long).If there is more than 1 connected components network is defined as "disconnected network".

Cycle: A path that starts and ends at the same vertex but does not revisit vertices in network.

Degree: The number of edges connected to a vertex that its minimum value is 0 and maximum value is N-1(N total number of vertices in network). A directed graph has an in-degree and an out-degree for each vertex, which are the numbers of in-coming and out-going edges.

$$\text{Degree of node i: } k_i = \sum_j k_{ij}$$

$$\text{Average degree of a network: } = \langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2E}{N}$$

(N is number of nodes in the graph)

Diameter: Maximum shortest path value in network ($diam(G) \rightarrow d = \max_{ij} d_{ij}$).

Directed graphs: A directed graph D (digraph) consists of a non-empty set of nodes (V) and a set of one direction (arrow) connection between pairs of different vertices (Directed Edges). Edges directions are represented as $i < j$ that mean an edge from vertex i to vertex j.

Distance (l_{ij} − distance from vertex i to j): The number of edges along the shortest path connecting two nodes. The mean number of neighbors in a distance "l" is defined as z^l and for a diameter "d" total number of vertices in a graph approximately equal to z^d ($N \approx z^d$). This equation follows the approximate diameter of a graph as $d \approx \frac{\log N}{\log z}$ .Distance of disconnected two vertices is infinity.

Edge: The line connecting two vertices, also called a bond (physics), a link (computer science), or a tie (sociology).

Graph representation: Mathematical representation of graph is made on $N \times N$ adjacency matrix (sociomatrix in social networks) $x = \{x_{ij}\}$ that are defined as $x_{ij} = 1$ if there is any connection between pairs of vertices i and j $((i,j) \in E)$ and otherwise is referred as $x_{ij} = 0((i,j) \notin E)$. In undirected graphs the adjacency matrix is symmetric $x_{ij} = x_{ij}$.

Geodesic distance: Minimum path value within path lengths also called as shortest path.

Path Length: Path is a sequence of nodes in which each node is attached to the next one and total number of edges between any two vertices is defined as "path length". Shortest path is the minimum path value within path lengths also called as "geodesic distance". If there isn't any path between vertex pairs, geodesic distance is defined as "infinite" for this vertex pairs.

Pearson correlation coefficient: A measure of the correlation (linear dependence) between two variables *X* and *Y*, giving a value between +1 and −1.

$$r = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{N \sum x_i{}^2 - (\sum x_i)^2} \sqrt{N \sum y_i{}^2 - (\sum y_i)^2}}$$

Simple graph: Graph does not have loops (self-edges) and multiple edges.

Undirected graph: An undirected graph G= (V, E) consists of non-empty set of vertices and connections between pairs of different nodes (E) are both direction.

Vertex: The fundamental elements of a network, also called a site (physics), a node (computer science), or an actor (sociology).

References

[1] Newman, M., Barabási, A.-L., Watts, D., (eds.), *The Structure and Dynamics of Networks*, Princeton University Press, Princeton, 2006.

[2] Watts, D. J., Strogatz, S. H., “Collective Dynamics of 'Small-World' Networks”, *Nature* 393, 440-442 (1998).

[3] Barabási, A. L., Albert, R., “Emergence of Scaling in Random Networks”, *Science* 286, 509-512 (1999).

[4] Rapoport, A., “Contribution to The Theory of Random and Biased Nets”, *Bulletin of Mathematical Biologysics* 19, 257-277 (1957).

[5] Solomonoff, R., Rapoport, A., “Connectivity of Random Nets”, *Bulletin of Mathematical Biologysics* 13, 107-117 (1951).

[6] Crucitti, P., Latora, Marchiori, V. M., “A Topological Analysis of the Italian Electric Power Grid”, *Physica A* 338, 92-97 (2004).

[7] Albert, R., Barabási, A. L., “Statistical Mechanics of Complex Networks”, *Reviews of Modern Physics* 74 (1), 47-97 (2002).

[8] Milgram, S., “The Small World Problem”, *Psychology Today* 2, 60-67 (1967).

[9] Newman, M. E. J., “The Structure of Scientific Collaboration Networks”, *Proc. Natl. Acad. Sci. USA* 98, 404-409 (2001).

[10] Price, D. J. DE S., “Networks of Scientific Papers”, *Science* 149, 510-515 (1965).

[11] Momtoya, J. M., Solé, R. V., “Small World Patterns in Food Webs”, *Santa Fe Institue Working Paper* 00-10-059 (2000).

[12] Albert, R., Albert, I., Nakarado, G. L., “Structural Vulnerability of the North American Power Grid”, *Physical Review E* 69, 1-4 (2004).

[13] Newman, M. E. J., “Scientific Collaboration Networks: I. Network construction and Fundamental Results”, *Phys. Rev. E* 64, 016131 (2001).

[14] Barabási, A.-L., et al., “Deterministic Scale-Free Networks”, *Physica A* 299, 559-564 (2001).

[15] Freeman, L. C., “Centrality in Social Networks: Conceptual Clarification”, *Social Networks* 1, 215-239 (1979).

[16] Freeman, L. C., “A Set of Measures of Centrality Based Upon Betweenness”, *Sociometry* 40, 35-41 (1977).

[17] Moddy, J., “Race, School Integration, and Friendship Segregation in America”, *Am. J. Sociol* 107, 679-716 (2001).

[18] Morris, M., *Data Driven Network Models for The Spread of İnfectious Disease, In D. Mollison (ed.), Epidemic Models: Their Structure and Relation to Data*, 302-322. Cambridge University Press, Cambridge, 1995.

[19] Newman, M. E. J., “Assortative Mixing in Networks”, *Phys. Rev. Lett.* 89, 208701 (2002).

[20] Newman, M. E. J., “Mixing Patterns in Networks”, *Phys. Rev. E.* 67, 026126 (2003).

[21] Newman, M. E. J., “The Structure and Function of Complex Networks”, *Computer Physics Communications* 147, 40-45 (2003).

[22] Albert, R., Albert, I., Nakarado, G. L., “Structural Vulnerability of the North American Power Grid”, *Physical Review E,* 69*:* 1-4 (2004).

[23] Crucitti, P., Latora, Marchiori, V. M., *“*A Topological Analysis of the Italian Electric Power Grid”, *Physica A* 338, 92-97 (2004).

[24] Sen, P., Dasgupta, S., Chatterjee, et al., “Small-World Properties of the Indian Railway Network”, *Physical Review E* 67, 036106 (2003).

[25] Li, W., Cai, X., “Statistical Analysis of Airport Network of China”, *Physical Review E* 69, 046106 (2004).

[26] Zhang, P., Chen, K., et al., “Model and Empirical Study on Some Collaboration Networks”, *Physica A* 360, 599-616 (2006).

[27] Sienkiewicz, J., Holyst, J. A, “Statistical Analysis of 22 Public Transport Networks in Poland”, *Phys. Rev. E* 72, 046127 (2005).

[28] Sienkiewicz, J., Holyst, J. A., “Public Transport Systems in Poland: From Bialystok to Zielona Gora by Bus and Tram Using Universal Statistics of Complex Networks”, *Physics*/0503099 (2005).

[29] Xinping, X., Junhui, H., Feng, L., Lianshou, L., “Scaling and Correlations in Three Bus-Transport Networks of China”, *Physica A* 374, 441 (2007).

[30] Chen, Y.-Z., Li, N., He, D.-R., “A Study on Some Urban Bus Transport Networks”, *Physica A* 376, 747-754 (2007).

[31] Xu, X., Hu, J., Liu, F., Liu, L., “Scaling and Correlations in 3 Bus-Transport Networks of China”, Physica A 374 (1), 441-448 (2007).

[32] Ferber, C. Von, Holovatch, T., Holovatch, Yu., and Palchykov, V., “Public Transport Networks: Empirical Analysis and Modeling” , *Eur. Phys. J. B* 68, 261-375 (2008).