

Investigation into Train Flow System on Ukraine's Railways with Methods of Complex Network Analysis

Tatyana But'ko, Andrii Prokhorchenko *

Ukrainian State Academy of Railway Transport, Operational Work Management Department, Kharkov, Ukraine

*Corresponding author: railwayhub@yandex.ua

Received November 08, 2013; Revised November 23, 2013; Accepted December 06, 2013

Abstract The article deals with an improved analysis of the train flow system in Ukraine's railway network. The main objective of the investigation is to reveal the peculiarities of the car flow destination system and to apply up-to-date knowledge for higher efficiency of railway transport. To solve the scientific problem the methods of complex network analysis have been used, thereby determining the basic statistic factors of the network topology. It has been proved that the destination network of Ukraine's train formation plan is characterized by scale invariance. The revealed peculiarities of assortative mixing have made the understanding of the system's processes simpler. The results obtained can be applied in analysis of the transportation system survivability of Ukraine's railway network.

Keywords: train forming plan, complex networks, free-scale networks, railway transport

Cite This Article: Tatyana But'ko, and Andrii Prokhorchenko, "Investigation into Train Flow System on Ukraine's Railways with Methods of Complex Network Analysis." *American Journal of Industrial Engineering* 1, no. 3 (2013): 41-45. doi: 10.12691/ajie-1-3-1.

1. Introduction

A normative document – Train Formation Plan (TFP) – is the basis of organization and movement of loaded and empty cars to their destination points. The document estimates the efficiency of rail transport and influences directly the cost efficiency. Railways of Ukraine in terms of traffic volume and network length are one of the largest in Europe [1,2]. Rail transport played a leading role in the transportation system of Ukraine, and therefore the financial results of its work direct influence on the economy. However, the analysis of TFP on the railways of Ukraine has received much less attention compared to research on the railways of the world [3,4,5,6,7]. The existing approaches towards the TFP analysis in Ukraine's rail network fail to estimate important characteristics of spatial formation of trains, particularly macro parameters of the transportation system, interdependence of railway stations within the network, and their workload. It requires implementation of new systemic TFP analyses. Study of works [8,9,10,11] showed that most widespread is approach on the basis of topology analysis of the train destination network. Thus, implementation of up-to-date research based on the theory of complex networks is promising [12], which enables us to establish interrelations within the train flow system and analyze the rail network as a whole.

2. Methodology Adopted

The proposed approach to the TFP analysis is based on interrelations between railway stations and presented as a

network structure with a directed graph, the nodes corresponding to arrival and departure stations and the arcs corresponding to the TFP destination stations.

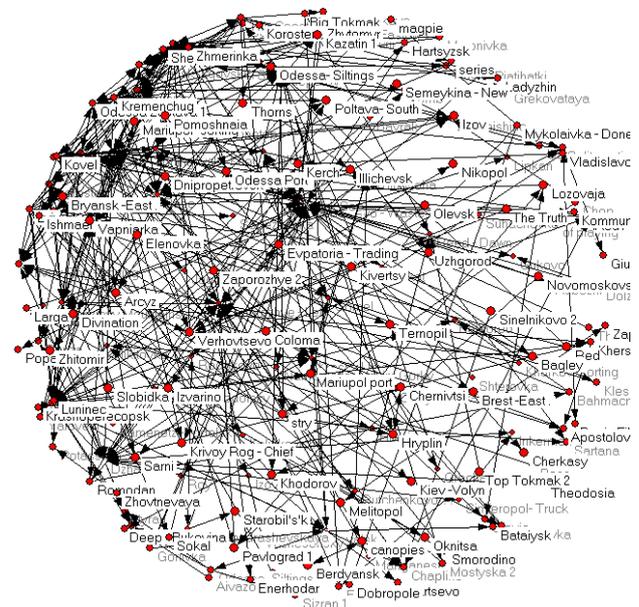


Figure 1. Graph visualization of the TFP destination network

The network analysis is based on the information derived from "The order and direction of car flows in the organization of freight trains on Ukraine's railways for 2012-2013". To summarize the information, the database of railway stations in the network, the TFP destinations in accordance with types of trains (through, divisional, assorted trains), and their distances has been established. The analysis has been carried out with the free software PAJEK (Program for Large Network Analysis, authors V.

Batagel and A. Mrvar, Slovakia) [13]. The analysis of 482 links between 181 railway stations is the object of the applied network research (Figure 1).

3 Topology Analysis of the Freight Car Destination Network on Ukraine's Railways

In the first stage of the investigation it has been proposed to determine the graph metrics indicating influence and dependence of specific stations in the train formation system in order to make the statistic estimate of the network topology. The nodes' centrality metrics in the graph is an important characteristic of each node's relative position [14]. The study uses several centrality indicators, such as degree centrality, closeness centrality, and betweenness centrality.

3.1. Degree Centrality

The order of nodes' connections or degree centrality enables us to determine the stations of the maximum influence within the network. The degree centrality is the amount of TFP destinations (links) of the station within the network. To identify stations of maximum incoming workload it has been proposed to calculate the in-degree (the amount of incoming links). Maximum $\Delta(G)$ and minimum $\delta(G)$ in-degrees of the graph's nodes are $\Delta(G) = 22$ and $\delta(G) = 0$ respectively. Analysis of group distribution testifies to the fact that the network has only 2% of stations of maximum in-degree, the ones of highest destinations within the network. Nizhniodniprovsk yard has the highest in-degree (22 destinations). It should be mentioned that 7% of stations has less in-degrees, anyway they can be considered as stations of high in-degree comparing to others; therefore, such stations are in demand in the network as well. The rest of the stations (91%) have either small (72%) or zero (19%) in-degrees.

To analyze the outgoing workload of the station in the network it has been proposed to calculate the out-degree centrality (the amount of outgoing links). So, Znamienka yard has the highest out-degree centrality (23). The group distribution testifies to almost similar trend: 2% of the highest out-degree stations, 8% of high out-degree stations, and 90% of low out-degree stations.

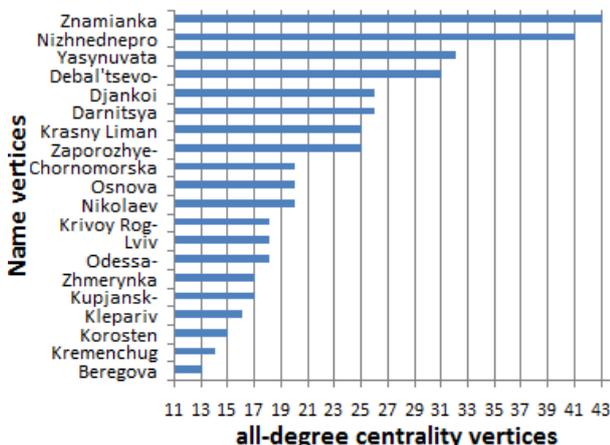


Figure 2. Diagram of all-degree centrality distribution in the TFP destination network

The all-degree centrality ranges from 43 to 1 (e.g. Figure 2, Figure 3). Znamienka yard has the highest all-degree (43). Sorted in descending order the network nodes' distribution identifies only 2% of stations which are in the highest demand in train flow redistribution. Besides, there exists a group of stations (8% out of the total number) with high all-degree centrality. This group includes stations of dominating incoming or outgoing car flows, i.e. they are stations of intensive making-up and breaking-up car flows. All other stations (90%) have low connectivity which testifies to their low influence on car flow redistribution in the network.

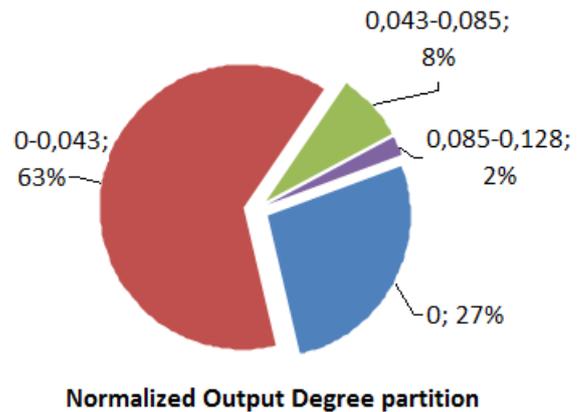


Figure 3. Group distribution of all-degree centrality in the TFP destination network

3.2. Degree Distribution

The design analysis of a node degree has immediately identifies a few hubs of a very high degree comparing to the others, that is the major specific feature of so-called free-scale networks [15,16]. With the empirical distribution of node degree obeying the power law [17], the TFP distribution network belongs to free-scale networks. The power law is often presented as Pareto's Law, but the study considers Zipf's Law, being a discrete analog of Pareto's Law as node degrees are integer values,

$$p_k = \frac{k^{-\alpha}}{\zeta(\alpha)}, \tag{1}$$

where p_k – is the probability that a randomly selected node has a vertex k ; α – is the positive constant; $\zeta(\alpha)$ – is the Riemann zeta function, which helps to satisfy the equation $\sum_k p_k = 1$.

The study has shown that in-degree distribution obeys the power law with the constant ~ 2.34 , out-degree distribution does with the constant ~ 2.07 , and all-degree distribution does with the constant ~ 1.9 . Actually, we can consider the constants' values as an indicator of change in destination network closeness from kernel to periphery. The histogram of probability distribution of the graph's node degrees in two logarithmic coordinates is presented in Figure 4.

The power characteristic has been calculated with the method of Maximum Likelihood Estimation (MLE) and the adequacy was checked with the Kolmogorov-Smirnov test (KS-test). It should be noted that due to statistic fluctuations the right part of the distribution is noisy, and

the most acceptable characteristic for the limited selection of data is the power probability distribution function. Thus, taking into account power distribution it is arguable that such structure as a TFP network is characterized by scale invariance. In the given circumstances increased number

of the TFP destinations does not influence the network stability. The heavy tailed degree distribution is the result of few stations of a great number of connections while the average degree distribution within the network is declining quite rapidly and nonlinearly.

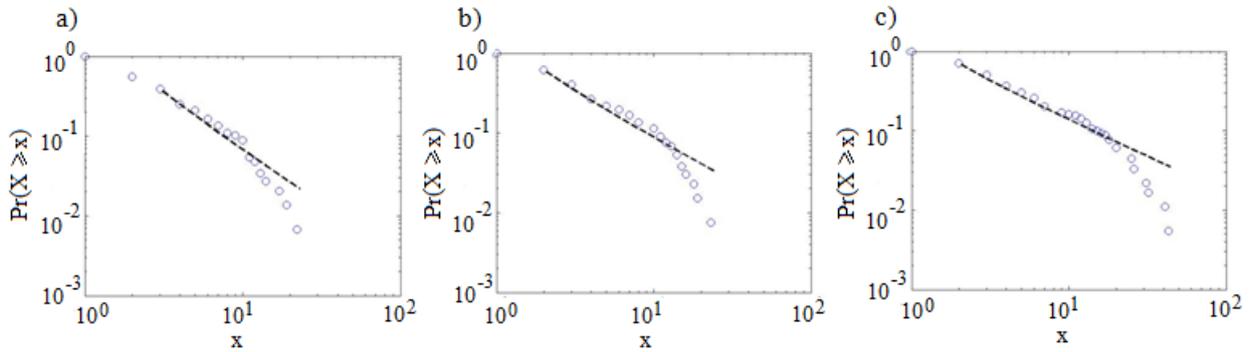


Figure 4. The histogram of probability distribution of graph's node degrees in two logarithmic coordinates: a) in-degree distribution; b) out-degree distribution; c) all-degree distribution

3.3. Closeness Centrality

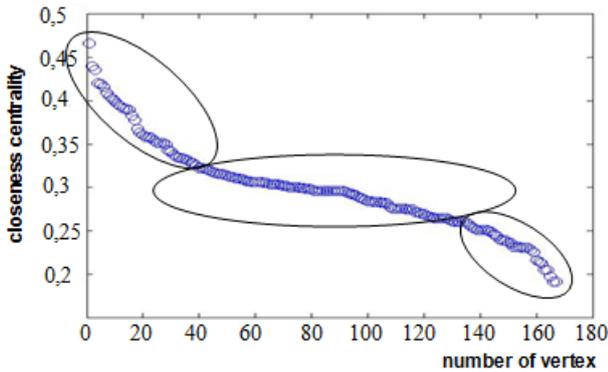


Figure 5. Dependence of networks sorted in descending order on their closeness values

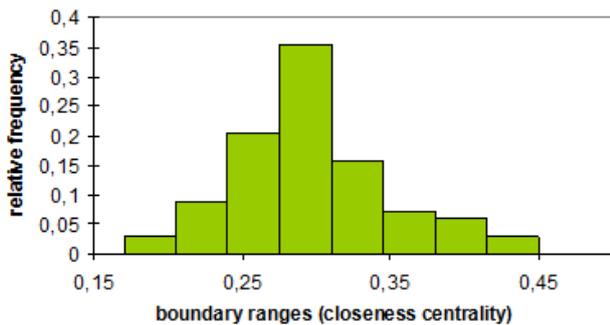


Figure 6. Histogram of station distribution in the TFP destination network by closeness

A high level of congestion is an important feature of free-scale networks. In order to confirm it, a centrality factor as closeness to the other nodes has been analyzed [18]. Small values of closeness indicate the remoteness of the stations from most other stations in the network, while high values testify to the fact that these stations are easily accessible for most other stations in the network, or other stations can be easily reached from them. The all closeness for nodes of the TFP distribution network ranges from 0.4663 to 0.1912. Nyzhniodniprovskya yard is

the closest one to all the others in the network, followed by Znamenka yard (0.4403), and followed by Yasinuvata yard station (0.4357). The dependence of networks sorted in descending order of the nodes on their closeness values (e.g. Figure 5) and the histogram (e.g. Figure 6) testify no sharp decline in closeness values. The average value within the network declines very tightly, but one can immediately identify a cluster of stations of high closeness, followed by a cluster of smaller stations (an average closeness of 0.298) and a cluster of so-called peripheral stations.

3.4. Betweenness Centrality

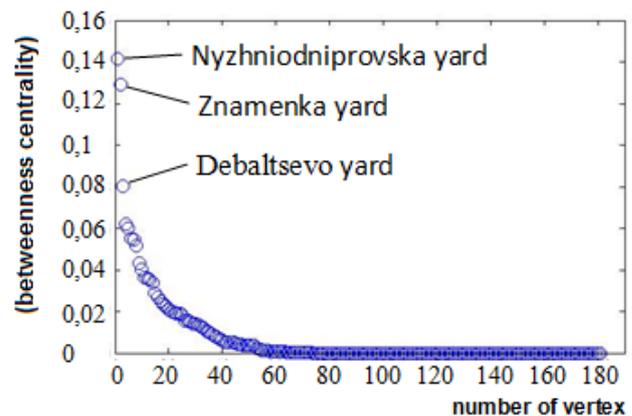


Figure 7. Dependence of networks sorted in descending order on their betweenness centrality

The analysis of betweenness centrality has shown a very dense station allocation within the network. Such high concentration makes it possible to come to the conclusion about the speed of car flows from a specific station to other connected stations. Besides the analysis of nodes' accessibility, another point of interest is few so-called hub vertices, which is a characteristic of free-scale networks. To solve this problem the study applies betweenness centrality which is the indicator of station's quantitative influence on connections with other stations in the TFP destination network. Actually, the indicator determines the frequency with which the station lies on the shortest path between two other stations. The analysis

showed that only 40% of the total number of stations has betweenness centrality (e.g. Figure 7, Figure 8). Several stations have the highest index: Nizhnedneprovsk (0.1417), Znamyanka (0.129), and Debaltsevo (0.804). The existing tariff system adopted at the shortest routes can explain a small number of stations through which most train flows run. It leads to uneven utilization of stations in the network, and operational failures at the station of the highest betweenness centrality could break transportation links and result in station's fragmentation.

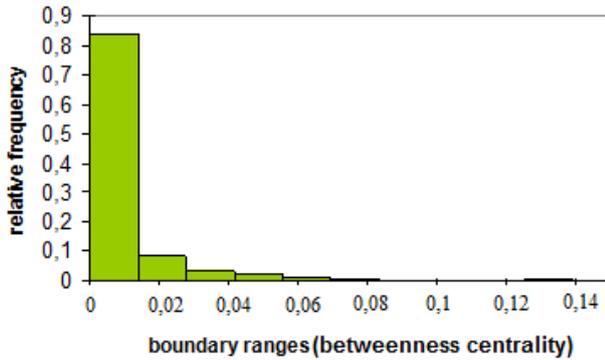


Figure 8. Histogram of station distribution in the TFP destination network by betweenness centrality

Due to high degree of clusters the destination network is extremely resistant to random failures which do not result in change of the network's diameter. If the work in the station fails, the other sorting and divisional stations may distribute its workload to maintain the network operations. At the same time, the network's graph is extremely vulnerable to coordinated attacks. Thus, simultaneous full stop of operations at stations of the highest vertex degrees (10% of total number of stations) leads to graph's disintegration and a huge cluster will fall to pieces. It leads to increase of the diameter.

3.5. Assortative Mixing

To determine assortative mixing [19] of the TFP destination network the study has proposed to trace the correlation of coupled nodes. Parametric Pearson's correlation coefficient and the Spearman's rank correlation coefficient have been used to analyze closeness on the principle of equal values [12]. Little discrepancy between values of two correlation coefficients could consolidate the conclusions. The correlation coefficients have been calculated between nodes' degree centrality and average degrees of their closest neighbors for various links (incoming, outgoing and mutual).

For incoming links Pearson's and Spearman's correlation coefficients are equal to $r_{int}^{Pirs} = 0,23362$;

$r_{int}^{Spir} = 0,51563809$ respectively; for outgoing links

$r_{out}^{Pirs} = 0,31997$; $r_{out}^{Spir} = 0,68164602$; and for mutual links

$r_{all}^{Pirs} = -0,12477$; $r_{all}^{Spir} = -0,08526248$. Medium positive correlation of vertices' parameters for incoming and outgoing links confirms assortative mixing of the TFP destination network. Anyway, the correlation coefficient by mutual links characterizes a weak connection

($-0,05 \leq r_{all} < -0,25$) and is negative, which may define the network as weak disassortative mixing by this

parameter. Assortativity by incoming and outgoing links in the network indicates that stations of a great number of links are interconnected with higher probability than ones with few links. Consequently, it is arguable that large sorting stations are not only local centers, but they also tend to unite into a dense hub surrounded by certain stations which connect them with stations of few links (the small world effect) [20]. The small world effect is clearly understandable for the TFP destination network, as on Ukraine's railways there are few one-destination routes while high volume of transportation requires considerable reforming of car traffic flows on their way to loading/unloading stations. Thus, rapid movement of car traffic flows requires links with hubs.

Disassortative mixing by mutual connectivity is explained by car flow interchanges among stations of a great number of links with stations of few links. Small stations form car flows with the TFP destinations to stations of high mutual connectivity, while large stations form car flows moving to the small stations.

3.6. Density, Diameter and Clustering Coefficient

To confirm the network's characteristics, which have been revealed in the analysis of certain nodes' parameters, it has been proposed to estimate general structural quantitative factors such as density, diameter and clustering coefficient. Density of the TFP destination network is number of destinations to network's general potential ratio. The ratio characterizes speed (intensity) of potential car flow distribution in the network and equals 0.0471262. It means that the graph has 1.47% of arcs out of their potential number. Another indicator of the graph, which shows car flow possibilities among the network's stations, is the diameter which is of great importance in determining the longest route (maximum number of steps between two stations of the network). Taking into account that the graph is disconnected (three disconnected components in its structure), its diameter is infinite and supposed to be a maximum diameter of existing clusters. So, it has been assigned that the graph of the TFP destinations has a maximum diameter of $\delta = 8$, which is between Olewsk station (the Southern-Western Railway) and Starobelsk station (the Donetskay Railway). It means that between two stations re-marshaling of car flow can be carried out less than eight times. All in all, the number of the shortest routes within the network is 15.988. The average length of a path in the TFP destination network is 3.706, which is not the longest regarding the network's diameter. Proximity to the calculated theoretical factor of the average length of a path [21] confirms that the TFP destination network is a free-scale network. This factor has been calculated by the analytical dependence derived experimentally for free-scale networks $l \sim \frac{\ln N}{\ln \ln N} = 3,1544$,

where N – is the number of nodes $N = 181$.

The network transitivity is characterized by the network clustering coefficient which shows the share of cliques of size three over the total number of triplets of nodes. The calculated network clustering coefficient equals 0.13384742 and confirms a high level of interconnections within the network. It is arguable that two destinations of the same station are connected with a probability of 0.13.

3.6.1. Bi-component Networks

For more detailed analysis of the network density it is necessary to identify connected components [22]. In calculating bi-components of the TFP destination network it has been revealed 76 components – separate networks. There exists only one largest network with 99 nodes, whereas all the other networks consist of two (70 dyads), three (4 triads) and four nodes (2 tetrads). One may say that the largest group accounts for 54.7% of the total number of stations in the network that confirms the existence of so-called a giant component.

4. Conclusion

From the above mentioned network's characteristics one may conclude that the most effective way to increase the efficiency of a railway network is to implement projects on improved technological operations for stations of the highest centrality, while investments in a random station of the network will not affect transportation criteria as a whole. According to the centrality analysis there is a need for enlargement of the list of major stations in Ukraine. Thus, the study has shown that such stations as Zdolbuniv, Mykolayev, Chornomorska, Krivoy Rig are also very important in transportation system.

The revealed characteristic of scale invariance in the TFP destination network denies Poisson distribution of links. Thus, the graph of the TFP destination network is not random, and its development is the base of self-organization of complex nonlinear systems. The small world effect and assortative mixing attribute the network to a specific class of complex networks that will make the understanding of its processes easier. Existing knowledge on free-scale networks could be applied to describe survivability of Ukraine's transportation system. Furthermore, it will be possible to design a mathematical model predicting the behavior of a transportation network.

References

- [1] Bukovskiy A.V., Kvartal'na N.A., "Basic premises and trends of reforming railway sector in Ukraine," *Quarterly Scientific Journal: Economic Herald of the Donbas*, 4 (30), 129-135, 2012.
- [2] Golovach, K., "Ukrainian Railways on the path to renewal," *Railway Gazette International*, Apr. 2008. [Online]. Available: <http://www.railwaygazette.com/news/single-view/view/ukrainian-railways-on-the-path-to-renewal.html/>. [Accessed Oct. 8, 2013].
- [3] Li, W., Cai, X., "Empirical analysis of a scale-free railway network in China," *PhysicaA: Statistical Mechanics and its Applications*, 382(2), 693-703, Aug. 2007.
- [4] Bo Zhou, "The forecast of rail freight volume based on complex network theory," *Rail Freight. Jou*, 3, 2008.
- [5] Schwander C., "Network Analysis Applied: the railway network in South East England," in *6th International Space Syntax Symposium*, Istanbul Technical University Faculty of Architecture, 1-17.
- [6] Sen, P., Dasgupta, P., Chatterjee, A., Sreeram, P.A. Mukherjee, G. Manna, S.S., "Small-world properties of the Indian railway network," *Phys Rev E*, 67(3), 1-5, March 2003.
- [7] Seaton, K.A., Hackett, L.M., "Stations, trains and small-world networks", *Physica A: Statistical Mechanics and its Applications*, 339(3), 635-644, Aug. 2004.
- [8] Ghosh, S., Banerjee, A., Shanna, N., Agarwal, S., Ganguly, N., "Statistical analysis of the Indian railway network: a complex network approach," *Acta Physica Polonica B Proceedings Supplement*, 4(2), 123-137, March. 2011.
- [9] Sienkiewicz, J., Janusz A. Holyst "Public transport systems in Poland: from Bialystok to Zielona Gora by bus and tram using universal statistics of complex networks" in *Presented at the 17th Marian Smoluchowski Symposium on Statistical Physics*, Acta Physica Polonica, 1-8.
- [10] Soh, H., Lim, S., Zhang, T., Fu, X., Lee, G.K.K., Hung, T.G.G., Di, P., Prakasam, S., Wong, L., "Weighted complex network analysis of travel routes on the Singapore public transportation system", *Physica A*, 389(24), 5852-5863, Dec. 2010.
- [11] GU, Xuejing, LI, Dewei, QIN, Lu, "Spatial Structural Characteristics of Chinese Railway Passenger Network Based on Complex Network Theory," *National Conference on Information Technology and Computer Science (CITCS 2012)*, Published by Atlantis Press, 750-753.
- [12] Wasserman, S., Faust, K., *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, 1994.
- [13] Batagelj, V., Mrvar, A., *Pajek: Package for Large Networks, Program Version 1.10* (October 25, 2005). University of Ljubljana, Ljubljana.
- [14] W. de Nooy, Mrvar A., Batagelj V., *Exploratory Social Network Analysis with Pajek* (Structural Analysis in the Social Sciences), Cambridge University Press, Cambridge, 2005.
- [15] Barabási, A.L., Albert, R. *Emergence of scaling in random networks*, Science, 1999, 509-512.
- [16] Barabasi, A.-L., *Linked: The New Science of Networks*, Basic Books, Cambridge, 2002.
- [17] Clauset A., Shalizi C.R., Newman M.E.J. "Power-law distributions in empirical data", *SIAM Review*, 51(4). 661-703. 2009.
- [18] Wasserman, S., Faust, K., *Social Network Analysis: Methods and Applications*. Cambridge University Press, New York, 1994.
- [19] Newman, M.E.J., "Assortative mixing in networks," *Physical Review Letters*, 89(20), 208701, May 2002.
- [20] Watts, D.J. "Networks, dynamics, and the small-world phenomenon" *American Journal of Sociology*, 105(2). 493-527. Sep.1999.
- [21] Réka, A., Barabasi, A.-L., "Statistical mechanics of complex networks", *Reviews of Modern Physics*, 74. 47-97. Jan. 2002.
- [22] Newman M.E.J. "The structure and function of complex networks", *SIAM REVIEW*, 45. 167-256. 2003.