SOPRONI EGYETEM

FAIPARI MÉRNÖKI ÉS KREATÍVIPARI KAR

AZ ALKALMAZOTT MŰVÉSZET LÉTMÓDJAI ÉS A KREATÍV IPAR KIHÍVÁSAI NAPJAINKBAN

Faipari Mérnöki és Kreatívipari Kar Tudományos Kiadványa

Szerkesztette: Márfai Molnár László és Pásztory Zoltán

AZ ALKALMAZOTT MŰVÉSZET LÉTMÓDJAI ÉS A KREATÍV IPAR KIHÍVÁSAI NAPJAINKBAN

FAIPARI MÉRNÖKI ÉS KREATÍVIPARI KAR TUDOMÁNYOS KIADVÁNYA

Szerkesztette: Márfai Molnár László és Pásztory Zoltán

SOPRONI EGYETEM KIADÓ

SOPRON, 2023

A kötet első 12 írása a Sopronban 2022. október 28-án *Az alkalmazott művészet létmódjai napjainkban* címmel megrendezett tudományos konferencia előadásainak szerkesztett anyagát tartalmazza.

A konferencia támogatói:

MTA VEAB Soproni Tudós Társaság Művészeti és Irodalomtudományi Szakbizottság

Magyar Tudományos Akadémia VEAB Képzőművészet, Művészetelmélet és Design Munkabizottság

Soproni Egyetem Faipari Mérnöki és Kreatívipari Kar

Felelős kiadó: Prof. Dr. Fábián Attila

a Soproni Egyetem rektora

Szerkesztette:

Dr. Márfai Molnár László és Dr. Pásztory Zoltán

Lektorálta:

Dr. Börcsök Zoltán

ISBN 978-963-334-453-8 (pdf)

https://doi.org/10.35511/978-963-334-453-8

Creative Commons licenc: BY-NC-SA 2.5

Nevezd meg! Ne add el! Így add tovább! 2.5 Hungary
Attribution – Non commercial – Share Alike 2.5 HUngary

Tartalom

Műszaki szekció

Foregin languages section

Comparative social network analysis (SNA) of FP7 and Horizon 2020 projects on remote sensing

Zsolt Tóth

PhD, University Associate Professor, University of Sopron, Faculty of Wood Engineering and Creative Industry, email: toth.zsolt@uni-sopron.hu

DOI: https://doi.org/10.35511/978-963-334-453-8.Toth_Zs-2

Abstract

As part of the data collection, data from all FP7 and Horizon 2020 projects were downloaded from the CORDIS data tables and converted into a relational table. After data cleaning and linking the tables, we queried the remote sensing projects of the two programmes, mapped the network of project participants in the two programmes, and compared the network indicators and the distributions of degree rates. A scale freedom of degrees seems to be a relevant assumption.

Keywords: remote sensing, social network analysis, projects, FP7, Horizon 2020

Literary background

The network research methods used in this paper are based on the model describing random networks (graphs) (Erdős and Rényi, 1960), on the configuration model for modelling networks with a fixed degree number distribution, but which otherwise consist of completely random links (Bollobás, 1980) (Molloy and Reed, 1995) (Newman, 2010), on the small world model built around the six-step distance question (Watts and Strogatz, 1998), on the model of scale-free networks (Barabási and Albert, 1999) and on the Barabási-Albert model describing the formation of scale-free networks (Albert and Barabási, 2002).

In the literature, the analysis of network cooperation on the basis of project collaborations is mostly focused on the analysis of a single project (Vonortas and Okamura, 2013) (Sobhani, 2012) (Li, et al., 2011) or makes relevant methodological-efficiency recommendations (Hansen, et al., 2012) (Liang, et al., 2009), but there are also examples of research carried out on a wider scale (Enger, 2018) (JuanolaFeliu and Samitier, 2016) (Balland, et al., 2019).

However, a complete comparative SNA analysis of remote sensing research, mainly with drones and satellites, carried out in the framework of the two major European R&D programmes (FP7 and Horizon 2020) has not yet been found by the author.

Methods

The downloaded data tables were converted into a relational database, and the SQL queries of the database were processed using statistical programming tools.

Summary of the statistical programming methods used in the paper

SQL and R codes were used in the statistical programming process. To assess the scientific context of the project, the .csv data tables imported from CORDIS were downloaded for both FP7 and H2020 projects, and from these, after preparing the tables, correcting some columns, and creating primary and foreign keys, a SQLite database was compiled. A SQL query was used to define the remote sensing-related data of FP7 and H2020 projects, mostly describing UAV surveys, in a summary table.

The following analysis process was then carried out for both tables (queries).

The topic was identified from the CORDIS SciVocCodes. In the next phase, the packages to be used for data analysis were selected (R RSQLite, igraph, vcd, poweRlaw, fitdistrplus). Then a formal class for a SQLite database connection was defined and the table of project participants for "remote sensing" was loaded into the data frame. After data cleaning, we wrote rcn-pairs describing the network of project participants using a much faster algorithm than used in our previous research. ("rcn" is the participant identifier.) The full R code can be studied at the reference given in the bibliography. (Tóth, 2022) Related R documentation is worth studying.

The rcn pairs were first converted into a list object, then an intermediate graph was created, and finally a matrix was defined from it. From the matrix, a graph could be constructed with the appropriate function to write the net. From the graph, an adjacency matrix describing the relationships of the project participants in binary was created. The matrix required optimisation due to its size and structure, so a network graph was created from it, which could be directly plotted and analysed. The network was then plotted after the parameters were entered. The very complex network that was plotted was unsuitable for visual evaluation due to its complexity, and was therefore characterised in terms of density, diameters and transitivity. Based on the betweenness index, the order of the central actors was also plotted. Then, in order to possibly verify the existence of distributions that are noteworthy in network research, the frequency of the degree numbers was examined. Before examining the notable

distributions, there was considered it useful to divide the network into clusters, to map the cluster network and to calculate some indicators. The goodness of fit of the distributions was also tested for the notable distributions of continuous and discrete variables, since the degree number as a discrete variable is considered a continuous variable due to the large number of elements in the discrete variable, according to the evaluation of statistical software, and therefore can be followed more or less by both methodological groups, the reliability of which is well supported by the literature. (Delignette-Muller and Dutang, 2015)

The Poisson distribution was tested using the likelihood ratio method and illustrated with a rootogram. (Wainer, 1974) The power distribution with appropriate parameters is the main measure of scale independence, and was therefore tested with several package methods, including bootstrap. To test the other notable distributions, we used a Cullen-Frey chart based on kurtosis and skewness. (Bailer and Bailer, 2001)

Results

Figure 1 shows the structure of the two project networks. For visually larger networks, essentially no relevant information can be extracted. What we can see from the figures is perhaps that the H2020 projects have more nodes and subnets that are not directly or indirectly connected to the central, closely connected nodes.

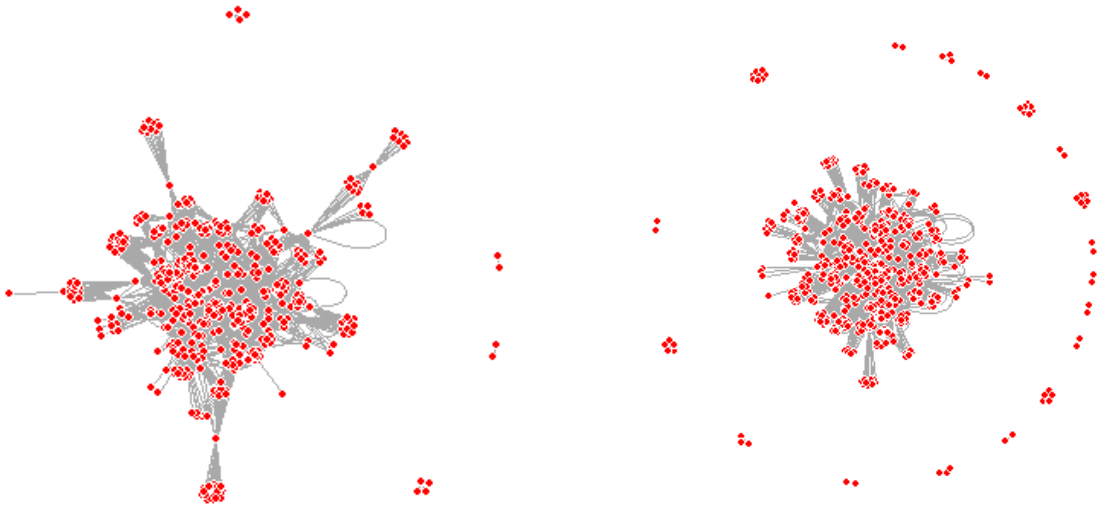

Figure 1. Participants' network in FP7 and Horizon 2020 projects on remote sensing

Based on the values obtained in Table 1, the remote sensing-related projects of the two funding programmes show a slight difference.

	Edge density	Transitivity	Diameter
FP7	0.045	0.68	6
Horizon 2020	0.033	0.61	5

Table 1. Network measures in FP7 and Horizon 2020 projects on remote sensing

As number of edges divided by maximal number of edges, decrease of edge density shows that the nodes were less interconnected in Horizon 2020 than in FP7. This may indicate a substantially larger number of elements or a strengthening of the centre and marginalisation of the periphery. However, a small increase in the number of networks' participants may not have this effect. The visual impression and the changes of other two indicators suggest the phenomenon of concentration-peripherization dual effect.

The histogram of the degrees (number of participants' connections) indicated scale freedom in the case of few divisors in both programmes. (Figure 2) Subsequent statistical analysis has clearly not confirmed this, but did not reject, as well. The visual similarity of histograms is apparent.

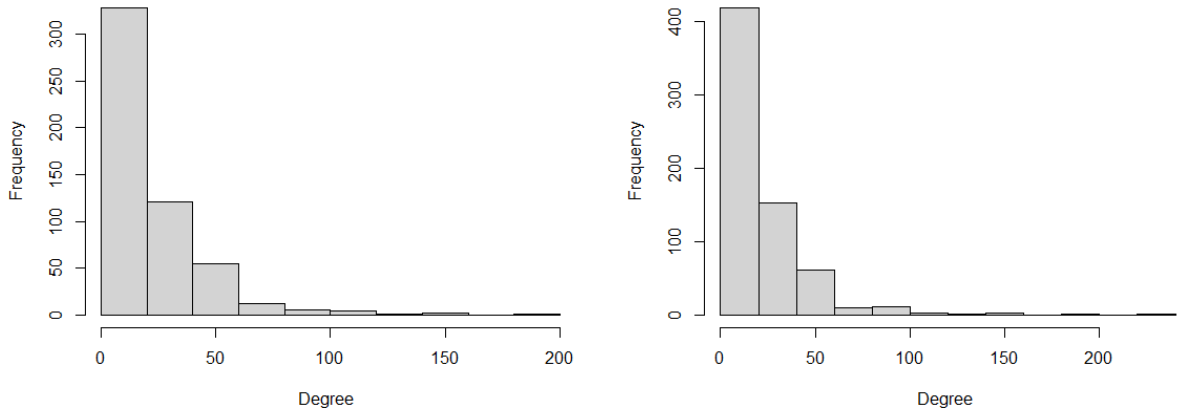

Figure 2. Frequency of degrees in FP7 and Horizon 2020 projects on remote sensing

According to the first five highest betweenness values of project participants (Table 2) the most active project participant's role is unquestionable, but the next places indicate a structural swift. The network of projects related to remote sensing does not seem to be a private club.

FP7		Horizon 2020	
RCN	Betweenness	RCN	Betweenness
1905609	27010.16	1905609	29935.07
1949732	15083.31	1919568	22338.78
1905572	14733.30	1909988	15266.06
1905912	10684.28	1905675	12484.88
1930250	9940.00	1905579	10200.24

Table 2. First five participants (RCNs) by betweenness values in FP7 and Horizon 2020 projects on remote sensing

The networks of FP7 projects on remote sensing can be divided into 51 clusters by cluster edge betweenness. The largest cluster includes 63 participants. The same indicators are in order 62 and 73 in Horizon 2020 projects. (Figure 3) There is no radical difference between the two samples and it can explained with the slight increase of participants. (In order: N = 801, N = 1004)

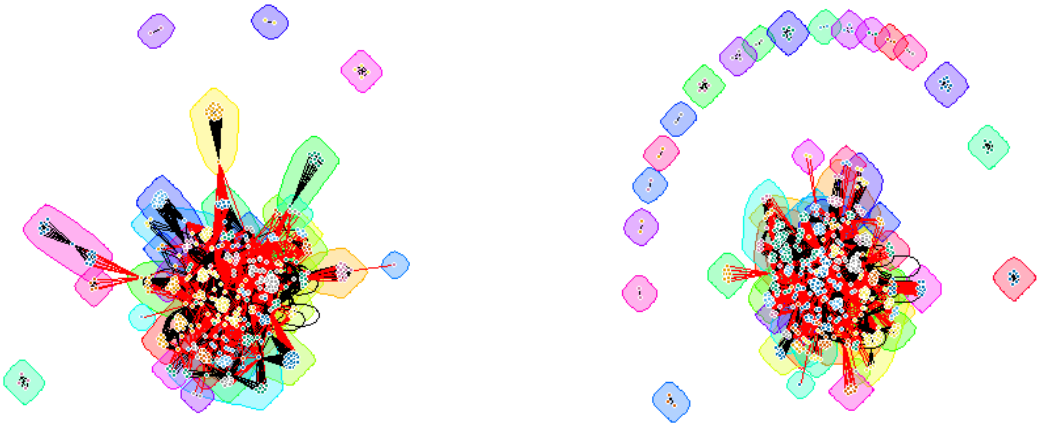

Figure 3. Clusters of FP7 and Horizon 2020 projects on remote sensing

Results of Goodness of fit analysis of Poisson distribution are clear in both cases. (Table 3) There is no Poisson distribution as we interpret the degree as discrete value. According to $P(>\chi 2)$ values calculated by maximum likelihood method H_0 (degrees follow a Poisson distribution) is rejected in both cases. As a benchmark, the analysis also included a random number generator to generate approximate data serie for Poisson distribution (n = 50, set.seed(123), $\lambda = 3$).

	Programme	p	χ^2	df	$P(>\chi^2)$
H_0: Poisson	FP7	0.95	6860.88	70	0
	Horizon 2020	0.95	8848.06	76	0
	Benchmark	0.95	5.35	6	0.5

Table 3. Results of maximum likelihood test of Poisson distribution of degrees in FP7 and Horizon 2020 projects on remote sensing

The results above can be represented with rootograms. (Figure 4)

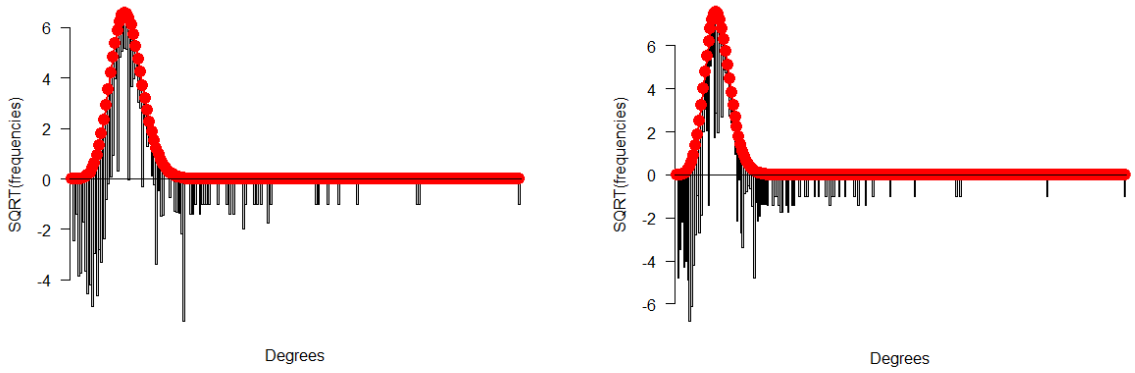

Figure 4. Rootograms of FP7 and Horizon 2020 projects on remote sensing

The next phase the H_0: data is generated from a power law distribution was tested, the result can be seen in Table 4. As a benchmark, the analysis also included a random number generator to generate approximate data serie for power lawn distribution (n = 100, set.seed(123), $\lambda = 3$, $x_{min} = 1$).

We can be quite confident that we cannot reject the null hypothesis at the 5% level of significance given these ranges, which we can infer as evidence that the data follows a power law distribution. (From x_{min} of course.) So, we can suppose that with the participants who are not the members of any projects but collaborate in this area, the degrees would have followed a power fit distribution. So many participants have few connections, and few participants have many, especially in the first case. This distribution suggests a notable distribution in social network analysis.

	Programme	Level of sig.	x_{min}	p-value
H_0: power law	FP7	0.05	68	0.554
	Horizon 2020	0.05	31	0.016
	Benchmark	0.05	1	0.99

Table 4. Results of Kolmogorov-Smirnov test of power law distribution of degrees in FP7 and Horizon 2020 projects on remote sensing

Examining the other frequent distributions was a surprise, as the Cullen-Frey diagram shows a gamma distribution for discrete variables treated as continuous in FP7, and this is confirmed by the bootsrap methods. When the degree number is treated as a discrete distribution, the distribution of degree points towards a negative binomial distribution, but this is unlikely based on kurtosis and skewness. (Figure 5)

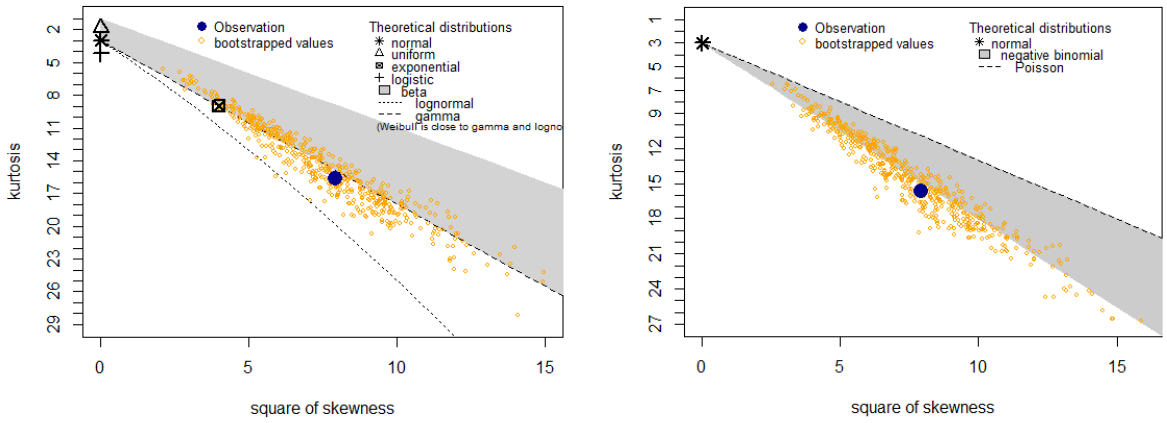

Figure 5. Cullen and Frey diagrams of FP7 projects on remote sensing

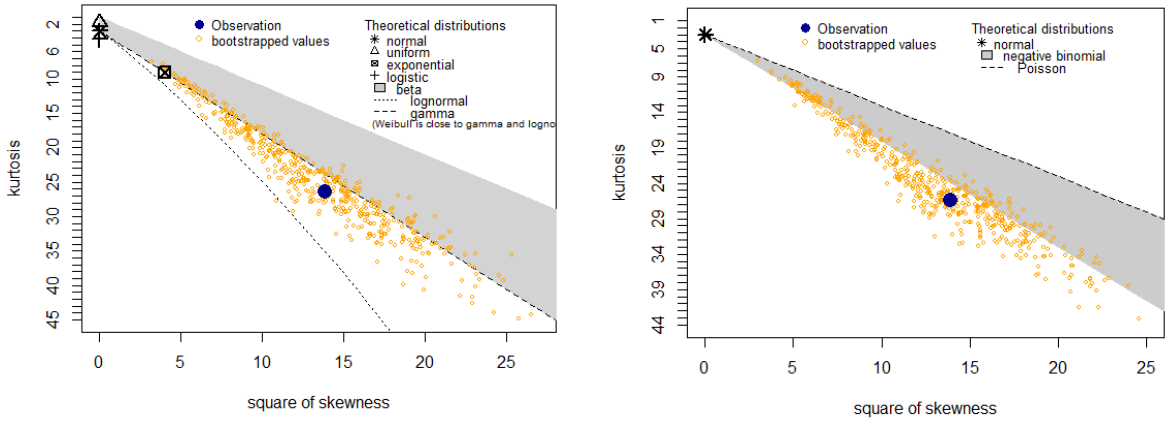

Figure 6. Cullen and Frey diagrams of H2020 projects on remote sensing

The projects in Horizon 2020 do not show any notable distribution. (Figure 6) However, this does not mean that the distributions of the two samples are significantly different, on the contrary, they are surprisingly similar.

The corresponding calculated values are given in the Table 5.

Name	FP7	Horizon 2020
Minimum	1	1
Maximum	196	237
Median	17	17
Mean	23.74	21.46
Estimated standard deviation	22.09	22.23
Estimated skewness	2.82	3.72
Estimated kurtosis	15.67	26.37

Table 5. Summary statistics related to Cullen and Frey plots in FP7 and Horizon 2020 projects on remote sensing

Acknowledgments

This article was made in frame of the project TKP2021-NVA-13 which has been implemented with the support provided by the Ministry of Innovation and Technology of Hungary (successor: Ministry of Culture and Innovation of Hungary) from the National Research, Development and Innovation Fund, financed under the TKP2021-NVA funding scheme.

Bibliography

Albert, R. & Barabási, A.-L., 2002. Statistical mechanics of complex networks. Reviews of Modern Physics, *Issue* 74, pp. 47-97. DOI: https://doi.org/10.1103/RevModPhys.74.47

Bailer, A. J. & Bailer, A. J., 2001. Probabilistic Techniques in Exposure Assessment. A Handbook for Dealing with Variability and Uncertainty in Models and Inputs. A. C. Cullen and H. C. Frey, Plenum Press, New York and London, 1999. No. of pages: ix + 335. Price: $99.50. ISBN 0-306-45956-6. *Statistics in Medicine*, 20(14), pp. 2211-2213. DOI: https://doi.org/10.1002/sim.958

Balland, P.-A., Boschma, R. & Ravet, J., 2019. *Network dynamics in collaborative research in the EU, 2003-2017*. DOI: https://doi.org/10.1080/09654313.2019.1641187

Barabási, A.-L. & Albert, R., 1999. *Emergence of scaling in random networks. Science, Issue* 286, pp. 509-512. DOI: https://doi.org/10.1126/science.286.5439.509

Bollobás, B., 1980. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, 1(4), pp. 311-316. DOI: https://doi.org/10.1016/S0195-6698(80)80030-8

Delignette-Muller, M. L. & Dutang, C., 2015. fitdistrplus : *An R Package for Fitting Distributions*. DOI: https://doi.org/10.18637/jss.v064.i04

Enger, S. G., 2018. Closed clubs: Network centrality and participation in Horizon 2020. *Science and Public Policy*, 45(6), pp. 884-896. DOI: https://doi.org/10.1093/scipol/scy029

Erdős, P. & Rényi, A., 1960. *On The Evolution of Random Graphs.* Magyar Tud. Akad. Mat. Kutató Int. Közl., Issue 5, pp. 17-61.

Hansen, D. L. et al., 2012. *Do You Know the Way to SNA?: A Process Model for Analyzing and Visualizing Social Media Network Data.* DOI: https://doi.org/10.1109/SocialInformatics.2012.26

JuanolaFeliu, C. P. E. & Samitier, J., 2016. Nano-enabled medical devices: Mapping the cross-fertilization of key enabling technologies in H2020 projects. *Journal of Nanomedicine & Nanotechnology*.

Li, Y. et al., 2011. Social network analysis and organizational control in complex projects: *construction of EXPO 2010 in China.* DOI: https://doi.org/10.1080/21573727.2011.601453

Liang, C., Wiewiora, A., Gable, G. G. & Trigunarsyah, B., 2009. *Project team's internal and external social networks and their influence on project performance.*

Molloy, M. & Reed, B. A., 1995. *Critical point for random graphs with a given degree sequence. Random Structures & Algorithms,* 6(2–3), p. 161–180. DOI: https://doi.org/10.1002/rsa.3240060204

Newman, M. E. J., 2010. Networks: *An Introduction. Oxford: Oxford University Press.*

Sobhani, F. M., 2012. A Framework for Stakeholder Analysis in ERP Project: A Case Study on SAP Implementation Using SNA (Social Network Analysis).

Tóth, Z., 2022. Social Network Analysis of Remote Sensing Projects in FP7 with R. Zenodo.

Vonortas, N. S. & Okamura, K., 2013. Network structure and robustness: lessons for research programme design. *Economics of Innovation and New Technology*, 22(4), pp. 392-411. DOI: https://doi.org/10.1080/10438599.2012.757897

Wainer, H., 1974. The Suspended Rootogram and other Visual Displays: An Empirical Validation. *The American Statistician, 2*8(4), pp. 143-145. DOI: https://doi.org/10.1080/00031305.1974.10479098

Watts, D. J. & Strogatz, S. H., 1998. Collective dynamics of 'small-world' networks. *Nature,* Issue 393, p. 440–442. DOI: https://doi.org/10.1038/30918