Analysis of the Digital Divide in the Telecommunication Services Using Knowledge Discovery in Databases

Vjekoslav Klarić
Hrvatski Telekom d.d., Croatia

Abstract

In this paper, the digital divide in the telecommunication services between 28 member states of the European Union in the period from 2007 to 2011 has been analysed using methods of knowledge discovery in databases. A brief history of telecommunications and essential telecommunication services has been presented, as well as their characteristics. Data related to the indicators of the digital divide in the telecommunication services has been collected and subjected to descriptive analysis. After choosing an optimal set of variables and a method, cluster analysis has been performed for each of the five years of the study period. Groups of EU member states at a similar stage of development of telecommunications have been identified and the characteristics of each of the groups have been described. Additionally, the paper presents some conclusions about changes in the scale of the digital divide in the telecommunications services between EU countries during the study period.

Keywords: digital divide, telecommunications, knowledge discovery in databases, data mining, cluster analysis
JEL classification: L86, L96, O33
Acknowledgments: The author would like to thank Mirjana Pejić Bach for her advice and guidance during the process of research and writing this paper.

Introduction

Communication at a distance has been present since the emergence of the first communities. The early means of telecommunication were capable of conveying just a few simple messages, but those messages could have literally been the difference between life and death. Some of the examples are the talking drums in Africa, smoke signals of the Native Americans, the hydraulic telegraph in Ancient Greece and chains of beacons in the medieval times.

Modern means of telecommunication started appearing in the 19th century, with the development of electronics. The invention of the telegraph in the first half of the 19th century has been followed by the invention of the telephone, the story of which demonstrates the importance of financing in the modern telecommunications. Many individuals contributed to the invention of the telephone: Mazzenti can be credited for the concept of the modern telephone, Reiss for the name, and the credit for its invention would probably belong to Meucci, had he been able to pay 10 US dollars for the rights to his patent in 1874. However, the person commonly regarded as the inventor of the telephone is Alexander Graham Bell, whose research has been funded by his father-in-law Gardiner Hubbard, the first president of the National Geographic society, and Thomas Sanders, a wealthy leather merchant. After Bell's invention in 1876, the three men founded Bell Telephone Company in 1877. This company in time evolved into AT&T, today one of world's largest

telephone companies. A natural step in the development of the telephone was the introduction of the mobile phone. Mobile phones were intended for commercial use from the get go. They have been developed for the use on the trains and ships in the 1920s and for the use in cars in the 1940s, but the first true hand-held device did not appear until 1973. First digital mobile phone networks appeared in the 1990s, and data traffic using mobile phones became available in the late 1990s and early 2000s. Data traffic itself started developing intensively in the 1960s with the appearance of ARPANET, a network mostly under control of the US Department of Defence. ARPANET was superseded by the supercomputer network called NSFNEt, but the unlimited commercial use of Internet was not possible until decommissioning of NSFNet in 1995.

It is apparent that the modern means of telecommunication differ greatly from the ones in the first communities. They are complex and their introduction, development and maintenance require significant material and human resources, so their usage is not free of charge, which means that they are not equally available to all. Inequality in the development and availability of goods and services related to the information and communication technology is called "digital divide". This term is more often used to describe the inequalities related to the information technology, rather than the ones related to the traditional telecommunication services. Therefore, most analyses on the subject of the digital divide focus primarily on the broadband, its coverage and applications in areas such as e-business and e-learning. For instance, Cruz-Jesus, Oliveira and Bacao (2012) studied the digital divide between 27 Member States of the European Union in the period 2008-2010, using data related to the Internet and Internet-based services. In their work, they used both factor and cluster analysis and determined the digital divide between EU countries exists.

When measuring the information society, UN's International Telecommunications Union (ITU), calculates the ICT Development Index using indicators reflecting the levels of access to ICT and use of ICT as well as skills required for the effective use of ICT (ITU, 2013), therein combining indicators related to information and communication technology with, e.g. adult literacy rate and gross enrolment ratio.

The telecommunications services, themselves, have found their way into everyday lives of a large portion of the world population during the 20th century, becoming their indispensable part. This resulted in the development of some telecommunications services markets to the point at which they have a strong two-way relationship with associated economies. Not only are those markets being influenced by economic trends, but also vice-versa, the rest of the market has become susceptible to changes in the telecommunications services markets.

Such trends are particularly noticeable in the developed regions of the world, which certainly includes the European Union. However, as with any other entity, the European Union is not entirely homogenous. Given the relatively short history of the European Union, in particular in its present form, as well as the diversity of its Member States, their resources, history and populations, disparities in the level of economic and technological development are to be expected.

The objectives of this paper are to choose indicators based on which it is possible to determine the extent of the digital divide in the telecommunication services in the European Union, to group 28 Member States of the EU according to the selected indicators, to determine if the divide is present and, if it really is present, to examine the changes in the extent of the divide in the period 2007-2011.

Methodology

In telecommunications, as one of technologically most developed industries, capabilities of data collection are substantial, but due to its profitability, each piece of information about telecommunications markets is considered valuable and seldom available free of charge. Despite such a situation, given the nature of the analysis in which the countries are grouped according to certain indicators, it is necessary to ensure that values of each indicator, in every single point in time, for all countries are comparable. Since one of the goals in this paper is to determine if any trends regarding the extent of the potential digital divide in the telecommunications services are noticeable, retaining certain level of consistency through time is also desirable. Keeping this in mind, as well as the fact that new systemized data is becoming available with a lag of a couple of years, 2011 has been chosen as the final year of this analysis. The eight initial indicators chosen for the analysis are given in Table 1. The subsequent three variables in the table have been used only in the repeated analysis for the year 2011. This analysis has been conducted in order to determine if introducing additional variables would seriously affect the results.

Table 1
Definitions and Sources of Variables Chosen for the Analysis of the Digital Divide between the 28 Member States of the EU between 2007 and 2011

Indicator	Definition	Source and period	Unit
Mobile-cellular telephone subscriptions per 100 inhabitants	Number of postpaid and active prepaid subscriptins to a public mobile telephone service that provides access to PSTN using cellular technology, divided by the popuation and multipiled by 100. Subscriptions offering only data services not included	ITU, 2007-2011	Number (per 100 inhab.)
Prepaid mobile-cellular subscriptions share in total mobile-cellular subscriptions	Number of active prepaid mobile-cellular telephone subscriptions divided by the number of all mobile-cellular telephone subscriptions and multiplied by 100	ITU for the 2007-2010 period, European Comission for 2011, with correction for Estonia and extrapolation of missing data	Percent.
Fixed telephone subscriptions per 100 inhabitants	Sum of numbers of active analogue fixed telephone lines, voice over IP subscriptions, fixed wireless local loop subscriptions, ISDN voice-channel equivalents and fixed public payphones, divided by the popuation and multipiled by 100	ITU, 2007-2011	Number (per 100 inhab.)
Fixed broadband subscriptions per 100 inhabitants	Subscriptions to high spped access to the public Internet, at downstream speed speeds equal to, or greater than, 256 kbit/s, divided by the popuation and multipiled by 100	ITU, 2007-2011	Number (per 100 inhab.)
DSL subscriptions per 100 inhabitants	Number of Digital Subscription Lines devided by the popuation and multipiled by 100	European Comission for the DSL share in fixed broadband subscriptions, ITU for the fixed broaband subscriptions per 100 inhabitants, 2007-2011	Number (per 100 inhab.)
Internet users per 100 inhabitants	Proportion of individuals that used the Internet in the last 12 months, multiplied by 100	ITU, based on surveys generally carried out by national statistical offices, 2007-2011	Number (per 100 inhab.)

Indicator	Definition	Source and period	Unit
Enterprises having a fixed broadband connection	Percentage of enterprises belonging to all manufacturing and service sectors, excluding the financial sector, with 10 or more persons employed, having DSL, xDSL, cable leased lines, Frame Relay, Metro-Ethernet, PLC-Powerline communications, fixed wireless connections, or similar connections *Note: Brakes in series, due to changes in methodology in 2009 and addition of a sector in 2010*	European Commision, 2007-2011	Percent.
Households having a broadband connection	Percentage of households with at least one member aged 16-74 with DSL, wired fixed (cable, fiber, Ethernet, PLC), fixed wireless or mobile wireless connection.	European Commision via Eurostat, with extrapolation of the missing data, 2007-2011	Percent.
Mobile broadband subscriptions per 100 inhabitants	*Number of mobile broadband subscriptions, dedicated to data traffic, divided by the population and multiplied by 100 Subscriptions offering only data services not included*	*Eurostat, except for Croatia. HAKOM, Croatian Regulatory Authority for Network Industries, for Croatia, 2011*	*Number (per 100 inhab.)*
Market share of the biggest mobile telecommunications operator	*Number of active SIM cards belonging to the biggest mobile telecommunications operator in the market, divided by the total number of active SIM cards in the same market*	*Eurostat, except for Croatia. HAKOM, and Hrvatski Telekom d.d., Croatia's biggest mobile telecommunications operator, for Croatia, 2011*	*Percent.*
Total revenues of the telecommunications sector per inhabitant	*All wholesale and retail telecommunications revenues of all the telecommunications operators, divided by the population*	*European Commission for the revenues, European Commission based on Eurostat estimate for the population, 2011*	*Euro*

Source: ITU, European Commission, Eurostat, own elaboration

A total of 8 potential outliers have been identified using methods of descriptive statistics, but all of those could be explained using qualitative analysis, so none of them has been dismissed as incorrect. Since the purpose of data in the analyses such as this one is to find differences and similarities between the objects of the analysis, excluding or changing some of the data without conclusive proof of their inaccuracy, just because they are different from most, might mean dismissing or reducing the influence of important evidence of dissimilarities between the observed objects, and thereby directly influencing the results. And if some data proves to be inaccurate, or causes issues later in the analysis, the process of knowledge discovery in databases is iterative, so this data can simply be dismissed later.

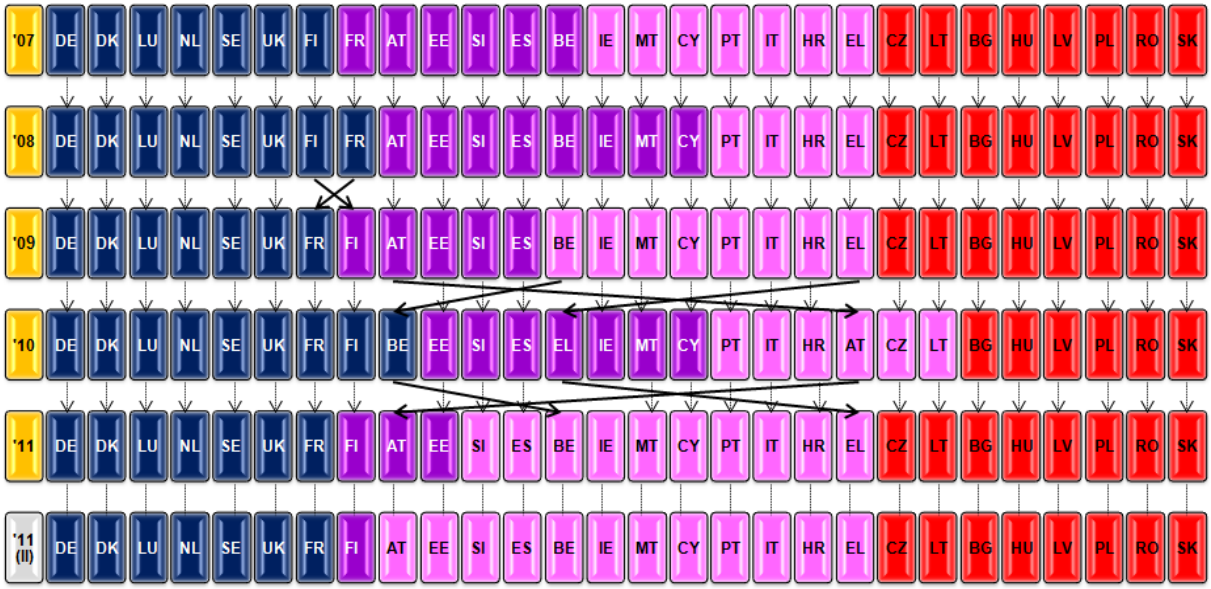
Even though the measurement units of all of the variables are basically the same, the descriptive methods showed that the levels of values the variables take on differ greatly, so the data has been standardized prior to clustering.

Clustering using different methods and types of distances, as well as on a reduced dataset, has been attempted, partly due to detected multicollinearity. In the end, a combination of Ward's method, squared Euclidian distances and 8 initial variables has been chosen as representative and suitable for analysis. It is interesting to mention that different combinations of methods, distances and sets of variables resulted in the formation of similar groups of countries, which could indicate the existence of natural clusters. The analysis has been conducted using R software.

Results

The results of the cluster analysis conducted using Ward's method with squared Euclidian distances, based on the chosen variables is shown in Figure 1. The number of clusters has been determined based on the plot of linkage distances across the steps. The clusters on the left are, in general, characterized by higher levels of fixed telephone penetration and Internet availability and usage, as well as lower share of prepaid subscriptions in mobile subscriptions, but differences are not significant when it comes to the penetration of mobile subscriptions.

Figure 1
Clusters of 28 EU Member States Based on Indicators of the Digital Divide in Telecommunications Services, between 2007 and 2011



Source: Author's illustration

The differences are more pronounced when comparing the most developed, left-most cluster, to the rest of the clusters, and when comparing the right-most, least developed cluster to the rest of the cluster, than when comparing the two middle clusters amongst themselves. This is consistent with the rather frequent changes in the structure of the two middle clusters, while the clusters on the left and right are much more stable.

The results of the analysis have been fairly consistent through the years, perhaps with the exception of year 2010. Removing any one of three variables from the analysis in 2010 makes the clusters for 2010 "fall in line", one of the three being the percentage of enterprises with a fixed broadband connection, which does have a break in series in 2010. In that case, though, one would expect that 2010 would be similar to 2011, which it is not. Still, given the results, there are doubts regarding data for 2010.

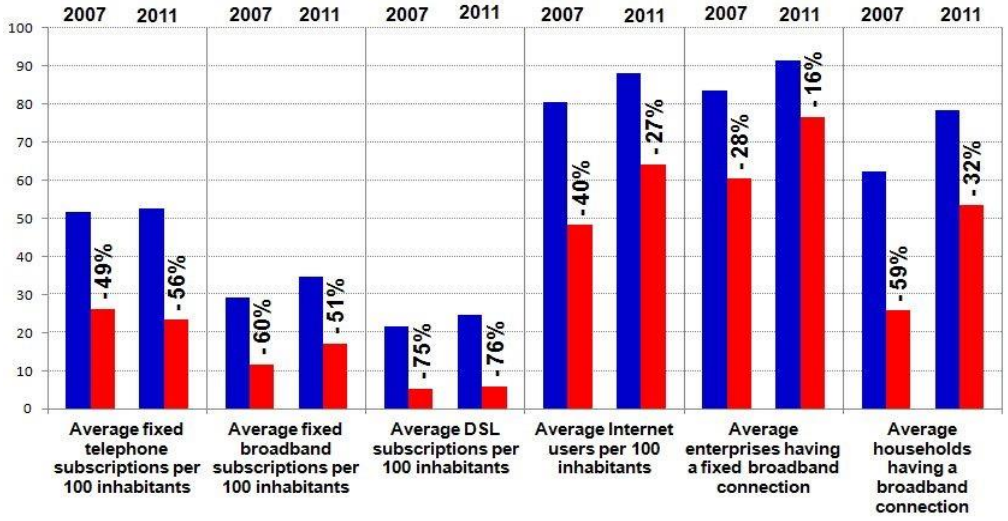
The result of the repeated analysis for 2011, conducted using an increased number of variables, shows Finland in a cluster on the middle-left. However, in this case it cannot be argued that this single-member cluster is less developed then the left-most cluster. Finland is, in fact, characterized by extremely high values in mobile broadband penetration and mobile telephone penetration in general, as well as high telecommunications revenue.

Discussion

The comparison of characteristics of the most and the least developed cluster at the beginning and the end of the observed period are is shown in Figure 2. Difference in fixed telephone penetration rates increased and probably never will be closed, since this particular service is slowly becoming obsolete. However, for most of the variables shown, differences in average values decreased, but they were still substantial in 2011. Perhaps more important than the decrease in differences is the increase in the average value of pretty much every variable, that is, in the penetration, accessibility and usage of telecommunication services in both clusters.

Figure 2

Average Values of Indicators Related to Fixed Telephony and the Internet in the Most and Least Developed Cluster of EU Member States

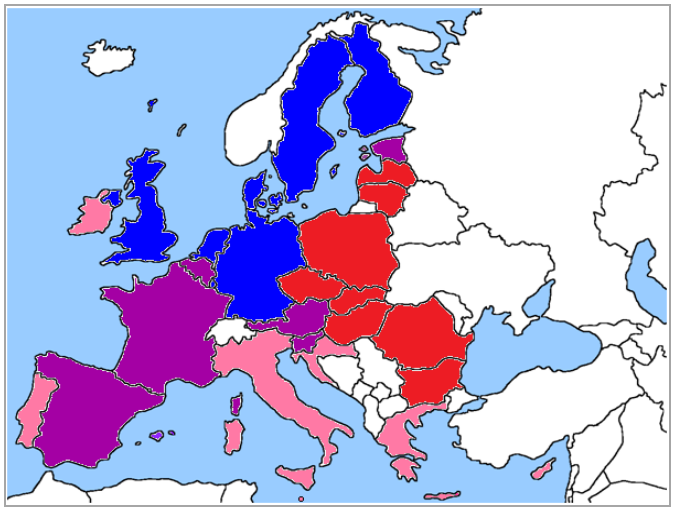


Source: Author's illustration

Final linkage distances, or, in other words, distances at which a single, unified cluster has been formed, decreased steadily during the observed period. Since the cluster analysis for the entire period has been conducted using the same method, same distances and the same variables, under the assumption of the temporal consistency of data, final linkage distances can be considered as indicators of trends in the extent of the digital divide. Distance required to complete the agglomeration process was 204.00 in 2007 and 159.65 in 2011, which is a decrease of 22%.

Looking at the countries that are forming each of the clusters, certain geographic patterns become apparent. As seen in figure 3, the least-developed cluster is located in the east of the European Union, while the most developed countries are grouped in the northwest. One of the "middle" clusters in 2007 consisted of Mediterranean and Atlantic countries, and the other was located in the southwest and the centre of the EU, with addition of the relatively highly digitalized Estonia.

Figure 3
Geographic Position of Clusters of 28 EU Member States in 2007



Source: Author's illustration

Conclusion

In this paper the digital divide in telecommunications services between 28 Member States of the EU during the five-year period, from 2007 to 2011, has been researched. Although Europe as a whole has always been among the leaders in the telecommunications development, it is internally diverse enough to expect signs of a digital divide, which is a phenomenon rooted in modern telecommunications from the very beginning.

Data for the analysis have been collected with the intention to preserve their consistency through both space, and time. The availability of data and the requirements placed on them limited the number of variables in the analysis to eight during most of the period. A greater number of variables would certainly be preferable and it would probably further increase the consistency of the results. The same constraints also limited the scope of analysis, both in regard to the observed period and the observed area. Comparing the EU Member States to other, both European and non-European, countries would also be of interest, while the increased duration of the observed period in further analysis would be helpful in recognizing the trends in the digital divide. Finding ways of making the telecommunications indicators publicly available and comparable on the global scale would be extremely beneficial in that respect.

After its extraction, data has been subjected to methods of descriptive statistics. Finally, cluster analysis has been performed for each year of the observed period. During the analysis, a group of most developed Member States in regards to telecommunications, located in the northwest of the EU, stood out. At the other end of the spectrum, a group of least developed countries, located in the east of the EU, emerged. When comparing the clusters, differences in values of indicators related to Internet and fixed telephony, as well as telecommunications revenue are substantial, but when it comes to penetration of mobile telephone subscriptions, they are not as noticeable. The share of prepaid subscriptions in total mobile phone subscriptions turned out to be generally lower at on the markets at a higher level of telecommunications development, and also, it decreases over time.

Distances required to complete the agglomeration process in the cluster analysis might give certain indications of the changes in the extent of the digital divide in

telecommunications. Those distances decreased slightly, and dropped by the total of 22% during the five-year period, which indicates an increase in the homogeneity of the EU countries in terms of level of telecommunication development. Another positive fact is that most of the indicators of the level of telecommunications development have increased in value in all of the clusters. However, their values, both at the beginning, and at the end of the observed period, were significantly higher in the cluster formed by the most developed countries, then in the cluster of the less developed ones. It is, therefore, still necessary to put a lot of effort into reducing, if not eliminating, the digital divide in telecommunications services between the countries of the EU, which is without a doubt existent.

References

1. Acilar, A. (2011), "Exploring the Aspects of Digital Divide in a Developing Country", Issues in Informing Science and Information Technology, Vol. 8, pp. 231-244.
2. Carveth, R., Kretchmer, S.B. (2002), "The Digital Divide in Western Europe: Problems and Prospects", Informing Science Special Series on the Digital Divide, Vol. 5 No 3, 2002, pp. 115-123.
3. Boje, C., Dragulanescu, N.G. (2003), " "Digital Divide" in Eastern European Countries and its Social Impact", presented at: 2003 ASEE Annual Conference & Exposition, Nashville, TN.
4. Clarke, G.R.G. (2001), "Bridging the Digital Divide: How Enterprise Ownership and Foreign Competition Affect Internet Access in Eastern Europe and Central Asia", World Bank Policy Research Working Paper No. 2629, Washington D.C., World Bank.
5. Cruz-Jesus, F., Oliveira, T., Bacao, F. (2012), "Digital devide across the European Union", Information & Management, Vol. 49, pp. 278-291.
6. Davis, E. (2011), "The life of information, from drums to Wikipedia", The Times Literary Supplement, available at: http://www.the-tls.co.uk/tls/public/article760768.ece (accessed July 1st 2015).
7. Digital Agenda for Europe (2015), European Commission, available at: http://ec.europa.eu/digital-agenda/en (accessed July 1st 2015)
8. DiMaggio, P., Hargittai, E., Celeste, C., Shafer, S. (2004), "Digital Inequality: From Unequal Access to Differentiated Use", in Neckerman, K.M. (Ed.), Social Inequality, pp. 355-400, New York, Russell Sage Foundation.
9. Eurostat Statistics Database (2015), available at: http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database (accessed July 1st 2015).
10. HAKOM (2012), "Godišnji usporedni podaci tržišta elektroničkih komunikacija u Republici Hrvatskoj", Zagreb, HAKOM.
11. HAKOM (2012), "Tromjesečni podaci tržišta elektroničkih komunikacija u Republici Hrvatskoj - 2011", Zagreb, HAKOM.
12. Fong, M.W.L. (2011), "Digital Divide: The Case of Developing Countries", Issues in Informing Science and Information Technology, Vol. 6, pp. 471-478.
13. Friedman, W.H. (2001), "The digital divide", AMCIS 2001 Proceedings, Paper 401, available at: http://ils.indiana.edu/faculty/hrosenba/www/I574/pdf/friedman_dig-div.pdf (accessed July 1st 2015).
14. Fuchs, C., Horak, E. (2007), "Informational capitalism and the digital divide in Africa", Masaryk University of Law and Technology, Vol. 1 No. 2, pp. 11-32.
15. Hargittai, E. (2003), "The Digital Divide and What to Do About It", in: Jones, D.C. (ed.) (2003), New Economy Handbook, pp. 822-841, San Diego, CA, Academic Press.
16. Huurdeman, A. (2003), "The Worldwide History of Telecommunications", John Wiley & Sons.
17. International Telecommunication Union (ITU) (2011), "Yearbook of Statistics: Telecommunication/ICT indicators 2001-2010", Geneva, ITU.

18. International Telecommunication Union (ITU) (2013), "Measuring the Information Society (without Annex 4)", Geneva, ITU.
19. International Telecommunication Union (ITU) Database (2015), available at: http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx (accessed July 1st 2015).
20. Internet Society (2015), "Brief History of the Internet – Internet Timeline", available at: http://www.Internetsociety.org/Internet/what-Internet/history-Internet/brief-history-Internet (accessed July 1st 2015).
21. OECD (2001), "Understanding the Digital Divide", OECD Digital Economy Papers, No. 49, OECD Publishing, available at: http://www.oecd-ilibrary.org/science-and-technology/understanding-the-digital-divide_236405667766 (accessed July 1st 2015).
22. Team Core (2013), "R: A language and environment for statistical computing", Foundation for Statistical Computing, Vienna, Austria, available at: http://www.R-project.org/ (accessed July 1st 2015).
23. van Dijk, J. (2008), "The Digital Divide in Europe", in Chadwick, A., Howard, P.N. (Ed.), "The Handbook of Internet Politics", London, Routledge.

About the author

Vjekoslav Klarić received a graduate degree from the Department of Mathematics at the University of Zagreb's Faculty of Science in 2010. In 2015 he completed a postgraduate specialist study in Statistical Methods for Economic Analysis and Forecasting, organized by the Faculty of Business and Economics, at the University of Zagreb. Since 2011, he has been working for at Hrvatski Telekom d.d., the largest telecommunications operator in Croatia. He is a member of the Croatian Mathematical Society and the Croatian Operational Research Society. So far he has published one scientific paper. Author can be contacted at **vjekoslav_klaric@yahoo.com**