Universidade de Aveiro
Departamento de Electrónica, Telecomunicações e Informática
2015

Sudhir Kumar Routray

Análise Estatística e Modelação de Redes Óticas de Transporte

Statistical Analysis and Modeling of Optical Transport Networks

Sudhir Kumar Routray

Análise Estatística e Modelação de Redes Óticas de Transporte

Statistical Analysis and Modeling of Optical Transport Networks

Tese apresentada às Universidades do Minho, Aveiro e Porto para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Engenharia Eletrotécnica no âmbito do programa doutoral MAP-Tele, realizada sob a orientação científica do Doutor Armando Humberto Moreira Nolasco Pinto, Professor Associado do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro e do Doutor José Rodrigues Ferreira da Rocha, Professor Catedrático do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro.

This research was partially supported by the PTDC/EEI – TEL/3303/2012 (ONECI) and PTDC/EEI – TEL/3283/2012 (DiNEq) projects, funded by the Fundação para Ciência e a Tecnologia — FCT, and the grants from the Department of Electronics, Telecommunication and Informatics of the University of Aveiro (Ref. No. – 2.14.300.1.4/15557/2012).

o júri / the jury

presidente / president

Doutor António Carlos Mendes de Sousa
Professor Catedrático da Universidade de Aveiro

vogais / examiners committee

Doutor Mário Marques Freire
Professor Catedrático da Universidade da Beira Interior

Doutor Henrique José Almeida Silva
Professor Associado da Universidade de Coimbra

Doutor Henrique Manuel de Castro Faria Salgado
Professor Associado da Universidade do Porto

Doutor Armando Humberto Moreira Nolasco Pinto
Professor Associado da Universidade de Aveiro

Doutor Amaro Fernandes de Sousa
Professor Auxiliar da Universidade de Aveiro

agradecimentos / acknowledgments

First of all, I would like to thank my thesis advisor, Professor Armando Nolasco Pinto, and my co-advisor, Professor José Rodrigues Ferreira da Rocha, for their guidance and advice throughout the entire duration of my PhD studies which finally resulted as this thesis. I thank Professor Gokhan Sahin for his guidance and helpful advice at critical junctures of my PhD. I feel very proud, honored and privileged for having the opportunity to work with these professors.

I would like to thank the University of Aveiro, Instituto de Telecomunicações, and all the associated administrative staff, for the excellent facilities offered during the course of this research.

Pursuing PhD in Portugal would not have been possible without the financial support from the projects of Fundação para Ciência e a Tecnologia — FCT, and the University of Aveiro. In this regard, I would like to thank my thesis advisors who took care of my financial needs during my PhD. I am very delighted for contributing to the ONECI and DiNEq projects.

I thank all my colleagues at Instituto de Telecomunicações, and the colleagues and teachers of MAP-tele doctoral programme for the helpful scientific discussions, and for providing a friendly atmosphere. I also use this opportunity to express my gratitude to everyone who supported me throughout the course of this PhD work.

Last but not the least, I thank my father who is always there to support my endeavors, and my wife whose love and dedication have always encouraged my efforts.

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

instituto de telecomunicações

To my father,

and

to Sasmita.

Palavras-chave Redes óticas de transporte, análise estatística, modelação estatística, comprimentos das ligações, comprimentos das ligações parâmetros relacionados, comprimentos dos caminhos

Resumo A análise estatística e modelação de redes é atualmente uma parte integrante da ciência e engenharia de redes. No caso das redes óticas de transporte (OTN), a modelação estatistica pode ser usada para o planeamento e dimensionamento quando a informação completa não está disponível ou o seu processamento é muito demorado. As redes óticas constituem atualmente o núcleo central das redes que suportam a Internet. Portanto, as características estatísticas dessas redes devem ser estudadas por forma a compreender sua natureza e estimar os seus parâmetros. Em ciência e tecnologia, a análise e modelação de redes é usada para vários fins, tais como análise de estabilidade, fiabilidade e evolução a longo prazo. O conhecimento dos modelos estatísticos ajuda na estimativa de vários parâmetros críticos das redes.

O trabalho apresentado nesta tese está focado na análise e modelação dos comprimentos das ligaçães e dos caminhos mais curtos em OTN. Os parâmetros usados nos modelos apresentados nesta tese podem ser estimados a partir de informação muito simples das redes, tais como a sua área de cobertura e o número de nós, sendo que ambas podem ser obtidas a partir da localização dos nós. Estes modelos podem ser aplicados para estimar parâmetros-chave das redes.

Nesta tese, demonstramos que o comprimento dos ligações em OTN segue uma distribuição do tipo general extreme value. Os parâmetros da distribuição podem ser estimados a partir do comprimento médio das ligações. Por sua vez mostramos que o comprimento médio das ligações pode ser estimado com um erro médio de 11% sendo apenas conhecida a área de cobertura da rede. Mostramos como é possivel aplicar o modelo desenvolvido à estimação de parâmetros dependentes do comprimento das ligações. Mostramos também que o comprimento dos caminhos mais curtos segue uma distribuição do tipo Johnson S_B. Os parâmetros usados neste modelo podem ser estimados a partir da área convexa e do número de nós da rede. Aplicamos ainda este modelo para estimar diversos parâmetros dependentes do caminho mais curto.

Keywords Optical transport networks, statistical analysis, statistical modeling, link lengths, link length related parameters, shortest path lengths

Abstract Statistical analysis and modeling of networks is now an integral part of network science and engineering. In case of optical transport networks (OTNs), it can be used for the planning and dimensioning when the complete information is not available or is difficult to process. The core networks around the world today are almost optical and they form the backbone of the Internet. Therefore, the statistical characteristics of these networks must be studied to understand their nature and to estimate their parameters. In science and technology, network analysis and modeling are used for several purposes such as the analysis of their stability, reliability and long term evolution. Knowledge of the statistical models helps in the estimation of several critical parameters of the networks.

The work presented in this thesis is focused on the analysis and modeling of link lengths and shortest path lengths in OTNs. The parameters used in the models presented in this thesis can be estimated from the very basic information of the networks such as the coverage area and the number of nodes, both of which can be found from the node locations. These models can be applied to estimate key parameters of the networks.

In this thesis, we have shown that the link lengths of the OTNs follow general extreme value distribution. The parameters of the proposed distribution can be estimated from the average link lengths of the networks. We develop expressions for the average link lengths of OTNs which can be estimated with an average error of just 11%. We apply the developed model to estimate link length dependent parameters in OTNs. We show that the shortest path lengths of the OTNs follow Johnson S_B distribution. We estimate the parameters of the developed model from the convex area and the number of nodes of the network. We also apply this model to estimate several shortest path-dependent parameters in OTNs.

Contents

List of Figures

List of Tables

List of Acronyms

1R	Reamplification
2R	Reamplification and Reshaping
3R	Reamplification, Reshaping and Re-timing
8QAM	8 Quadrature Amplitude Modulation
16QAM	16 Quadrature Amplitude Modulation
AD	Anderson-Darling
Avg. KSS	Average Kolmogorov Smirnov Statistic
BPSK	Binary Phase Shift Keying
CAPEX	Capital Expenditure
CDF	Cumulative Distribution Function
CE	Circumferential Ellipse
CI	Confidence Interval
EDFA	Erbium Doped Fiber Amplifier
EO	Electrical to Optical
EON	Elastic Optical Networking
EXC	Electrical Cross Connect
GEV	General Extreme Value
H. KSS	Highest Kolmogorov Smirnov Statistic
ITU	International Telecommunication Union
IP	Internet Protocol

KS	Kolmogorov Smirnov
KSS	Kolmogorov Smirnov Statistic
L. KSS	Lowest Kolmogorov Smirnov Statistic
MANEX	Management Expenditure
NB	N (number of) Links (or Paths) with BPSK Modulation
NE	N (number of) Links (or Paths) with 8QAM Modulation
No. I/P	Number of Input Parameters
NQ	N (number of) Links (or Paths) with QPSK Modulation
NS	N (number of) Links (or Paths) with 16QAM Modulation
OADM	Optical Add Drop Multiplexer
OE	Optical to Electrical
OEO	Optical Electrical Optical
OFDM	Orthogonal Frequency Division Multiplexing
OLT	Optical Line Terminal
OM	Other Method
OPEX	Operational Expenditure
OPNET	Optimized Network Engineering Tool (A Software)
OSC	Optical Supervisory Channel
OTN	Optical Transport Network
OXC	Optical Cross Connect
PDF	Probability Distribution Function
QPSK	Quadrature Phase Shift Keying
ROADM	Reconfigurable Optical Add Drop Multiplexer
SDH	Synchronous Digital Hierarchy
SDN	Software Defined Networking
SOA	Semiconductor Optical Amplifier
SONET	Synchronous Optical Network

UALL	Using Average Link Length
UASPL	Using Average Shortest Path Length
UELL	Using Exact Link Length
UESPL	Using Exact Shortest Path Length
ULLD	Using Link Length Distribution
USPLD	Using Shortest Path Length Distribution
WDM	Wavelength Division Multiplexing

N. B.: Please note that in the majority of the cases, the acronyms are defined at the first instance they appear in a chapter. The short names of the optical transport networks (OTNs) used for this thesis work are not listed in this section. For more information about these OTNs, please follow the references provided in the respective chapters.

CHAPTER 1

Introduction

THESE DAYS COMMUNICATION networks are found everywhere. Be it the wireless network of mobile phones, or the landline connections, or the versatile Internet, or the networks of several other broadcasting services such as the television and radio, the communication networks are omnipresent. In addition to this, networks in general, are also found in different scales: in the society, within human metabolism, in the brain of animals, in the micro-world in the forms of bacteria and virus, and in many other forms [1] – [15]. Due to their vital presence and importance in the world, networks are being studied carefully to understand their nature and behavior in different contexts [4] – [18].

In the modern world, telecommunications networks have basic roles in the lives of the human beings. With the proliferation of the Internet, telecommunications now furnish several vital functions in the personal lives of people and business processes. The global communication in such a fast changing world is possible due to the high speed core networks. These days, the core networks are predominantly optical, as no other medium can provide such high bandwidths, and handle such high traffic demands. The long distance optical communication systems have been evolved in several forms of the optical transport networks (OTNs). Overall, the core networks, high speed metro and access networks, and inter-cloud networks are supported completely by optical infrastructure [19]. In this scenario, the role of the optical networks becomes very important. Thus, these networks must be studied systematically from their structure and behavior points of views. Statistical analysis and modeling of the network parameters are integral parts of these critical studies [20] – [23]. They provide several critical behavioral aspects of the optical networks, and are quite helpful in the estimation of network parameters with

incomplete information [20],[21].

In this thesis, we study the statistical properties of the links and shortest paths of the OTNs, analyze their statistics, and provide novel models for them which can be utilized for practical purposes. OTN in the International Telecommunication Union (ITU) standards nomenclature, includes several standards for the optical communication of various types of data in the optical networks [24]. However, in this thesis, we use OTN as the core optical transport network that facilitates end-to-end communications, unless otherwise explained. In some cases, we also use 'network' and 'optical network' for these core optical transport networks.

1.1 Motivation

Dimensioning and cost estimation of OTNs are required in several occasions for the design and planning related operations [20], [21]. Mainly, in the early stage of planning, network overhauling and optimization of the OTNs, there are needs of the estimation of network parameters [20] – [23], [25] – [27]. These estimations have to be accurate for the overall effectiveness. In most of these cases, complete information is not available or the detailed estimation is too complicated, or too much time consuming. The parameters available at the early stage are the basic information of the OTNs, such as the coverage area and the number of nodes [25]. Some extra knowledge of the OTN, such as the statistical distribution, and the first and second moments of the network parameters can be essential in network planning and design [25]. They are useful for the determination of several network parameters such as the modulation and demodulation schemes needed, compensation techniques at the transmitting and receiving ends, number of repeaters and amplifiers along the optical links and paths, and total length of fiber needed [28]. Thus statistical models with appropriate accuracy are needed to be found for the statistical distribution of link lengths and the shortest path lengths. In many cases of OTNs, fast and reliable estimations using statistical modeling are quite effective and popular. These estimations play critical roles in situations where, getting extra information becomes too complicated [20], [25].

1.2 Objectives

The main motivation for the subject areas of this thesis have been addressed in the previous paragraph, which emphasizes the role of statistical analysis and modeling of OTN parameters. There are several fundamental research problems associated with

them. In order to solve those problems, a set of well-defined objectives are framed for this work. The main objectives of this work are listed below.

1. To develop a statistical model for link lengths in OTNs which can be applied for the cases of estimations with incomplete information.

2. Application of the developed link statistical model for the estimation of link length dependent parameters of OTNs.

3. Development of a statistical model for the shortest path lengths in OTNs.

4. Application of the developed shortest path statistical model for the estimation of shortest path dependent parameters of OTNs such as the number and types of modulation formats needed in transparent optical networks.

1.3 Main Achievements

In this thesis work, statistical models for the measurement and estimation of network parameters have been developed for the OTNs. These models have been used for several real applications of the OTNs. The main contributions achieved in this thesis are mentioned in the following list.

1. Development of a statistical model for link lengths in OTNs [25]. This model can be estimated from the average link length of the OTNs. The estimation formula for average link length of the OTNs used for dimensioning was also improved for better accuracy [25].

2. Development of a novel method for the estimation of the convex area of OTNs, which is used in the estimation of their parameters [29].

3. Development of new application methods for the estimation of link length dependent parameters in the OTNs using the above statistical model, which improves the estimation accuracy significantly [28].

4. Development of a statistical model for the shortest path lengths in OTNs [30]. Estimation of the above mentioned model without complete information of the OTNs [30]. Application of this model for the estimation of the shortest path length dependent parameters of the transparent OTNs [30].

1.4 Thesis Outline

This thesis is organized as eight separate chapters and four appendices. In Chapter 2, we present the importance of statistical analysis and modeling of networks in different disciplines. Then we present the main features of OTNs that are relevant for this work. In this chapter, the fundamentals of OTNs and their components are described along with their associated enabling technologies. The needs of statistical modeling in OTNs are explained briefly.

In Chapter 3, we present the statistical model developed for the link lengths of OTNs. This model can be estimated with incomplete information of the OTNs. Only the node locations are needed for this estimation. In this chapter, we also present the improved expressions for the average link lengths of OTNs based on different areas. We use the expression based on the convex area to estimate the link statistical model.

In Chapter 4, we develop an estimation method for the convex area of the OTNs. Convex area is a basic parameter in the estimation of average link length and several other statistical parameters of the OTNs. We provide a novel method of estimation of convex area using a simple planar map. In this chapter, we show the simplified expression for the convex area obtained from the circumferential ellipse of the OTN.

In Chapter 5, we present several applications of the statistical models developed for the link lengths for the OTNs. We developed the expressions for the estimation of OTN parameters from the statistical model. We evaluate the effectiveness of the developed expressions using the errors associated with these estimations with respect to their exact values.

In Chapter 6, we present the statistical model developed for the shortest path lengths between the node pairs of the real OTNs. This model can be estimated with incomplete information of the OTNs. Only the node locations are needed for this estimation of the statistical model for the shortest path lengths. In this chapter, we also present the expressions developed for the average, standard deviation, median and an upper bound of the shortest path lengths of OTNs.

In Chapter 7, we present the applications of the statistical model developed for the shortest path lengths for the OTNs. In this chapter, we developed the expressions for the parameters of the statistical model. These expressions depend on the basic information of the OTNs. We evaluate the effectiveness of the developed expressions using the statistical goodness to fit tests with respect to the optimized model.

In Chapter 8, we summarize the main points of the thesis with the achievements and their applications in the OTNs. We provide the possible future directions, and the

common outlook for the work done in this thesis framework.

In Appendix A, we show the significance and measurement of 'Kolmogorov-Smirnov Statistic,' which has been used to show the statistical 'goodness-of-fit' for different statistical distributions. We illustrate it using appropriate graphical presentation. In Appendix B, we estimate the shape factor of General Extreme Value (GEV) distribution which is used for the estimation of the link length distribution. In Appendix C, we show the mathematical expressions for moments of Johnson S_B distribution. In Appendix D, we show the relationships between the median and parameters of Johnson S_B distribution which can be used for the estimation of the shortest path length distribution.

1.5 List of Publications

The major findings of this work have been reported in relevant journals and conference. The main published works are listed below.

1.5.1 Papers in Journals

4. S. K. Routray, G. Sahin, J. R. F. da Rocha, and A. N. Pinto, "Statistical Analysis and Modeling of Shortest Path Lengths in Optical Transport Networks," *IEEE/OSA Journal of Lightwave Technology*, accepted in Mar. 2015.

3. S. K. Routray, G. Sahin, J. R. F. da Rocha, and A. N. Pinto, "Estimation of Link-Dependent Parameters of Optical Transport Networks from the Statistical Models," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 6, no. 7, pp. 601 – 609, Jul. 2014.

2. S. K. Routray, "Changing Trends of Optical Communication," *IEEE Potentials Magazine*, vol. 33, no. 1, pp. 28 – 33, Feb. 2014.

1. S. K. Routray, R. Morais, J. R. F. da Rocha, and A. N. Pinto, "Statistical Model for Link Lengths in Optical Transport Networks," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 5, no. 7, pp. 762 – 773, Jul. 2013.

1.5.2 Papers in Conferences

1. S. K. Routray, J. R. F. da Rocha, and A. N. Pinto, "Estimating the Parameters of Optical Transport Networks from Their Circumferential Ellipses," in Proceedings of IEEE International Conference on Telecommunications (ICT), pp. 119 – 123, Lisbon, Portugal, 4 – 7 May 2014.

References

[1] D. P. Bertsekas, R. G. Gallager, and P. Humblet, *Data Networks,* ed. 2, Englewood Cliffs, NJ: Prentice Hall, 1992.

[2] R. Ramaswamy, K. N. Shivrajan, and G. H. Sasaki, *Optical Networks: A Practical Perspective*, San Francisco, CA: Morgan Kaufman, 2009.

[3] M. Schwatrz, *Telecommunication Networks: Protocols, Modeling and Analysis,* Englewood Cliffs, NJ: Prentice Hall, 1987.

[4] M. Newman, *Networks: An Introduction,* ed. 1, New York, NY: Oxford University Press, 2010.

[5] S. N. Dorogovtsev, and J. F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW,* Oxford, UK: Oxford University Press, 2013.

[6] A. L. Barabási, and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509 – 512, 1999.

[7] R. Albert, H. Jeong, and A. L. Barabási, "Internet: Diameter of the world-wide web," *Nature*, vol. 401, pp. 130 – 131, 1999.

[8] D. J. Watts, and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440 – 442, 1998.

[9] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, "Classes of small-world networks," *Proceedings of the National Academy of Sciences*, vol. 97, no. 21, pp. 11149 – 11152, 2000.

[10] Y. Y. Ahn, J. P. Bagrow, and S. Lehmann, "Link communities reveal multiscale complexity in networks," *Nature*, vol. 466, pp. 761 – 764. 2010.

[11] C. Moore, and M. E. Newman, "Epidemics and percolation in small-world networks," *Physical Review E*, vol. 61, no. 5, 5678, 2000.

[12] A. L. Barabási, N. Gulbahce, and J. Loscalzo, "Network medicine: a network-based approach to human disease," *Nature Reviews Genetics*, vol. 12, no. 1, pp. 56 – 68, 2011.

[13] M. Vidal, M. E. Cusick, and A. L. Barabási, "Interactome networks and human disease," *Cell*, vol. 144, no. 6, pp. 986 – 998, 2011.

[14] A. Wilhite, "Bilateral trade and 'small-world' networks," *Computational Economics*, vol. 18, no. 1, pp. 49 – 64, 2001.

[15] E. Bullmore, and O. Sporns, "The economy of brain network organization," *Nature Reviews Neuroscience*, vol. 13, no. 5, pp. 336 – 349, 2012.

[16] B. Barzel, and A. L. Barabási, "Universality in network dynamics," *Nature physics*, vol. 9, no. 10, pp. 673 – 681, 2013.

[17] M. Schich, C. Song, Y. Y. Ahn, A. Mirsky, M. Martino, A. L. Barabási, and D. Helbing, "A network framework of cultural history," *Science*, vol. 345, pp. 558 – 562, 2014.

[18] S. Suweis, F. Simini, J. R. Banavar, and A. Maritan, "Emergence of structural and dynamical properties of ecological mutualistic networks," *Nature*, vol. 500, pp. 449 – 452, 2013.

[19] A. Q. Lawey, T. E. H. El-Gorashi, and J. M. H. Elmirghani, "Distributed Energy Efficient Clouds Over Core Networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 32, no. 7, pp. 1261 – 1281, Apr. 2014.

[20] S. K. Korotky, "Network global expectation model: A statistical formalism for quickly quantifying network needs and costs," *IEEE/OSA Journal of Lightwave Technology*, vol. 22, no. 3, pp. 703 – 722, Mar. 2004.

[21] J. F. Labourdette, E. Bouillet, R. Ramamurthy, and A. A. Akyamaç, "Fast approximate dimensioning and performance analysis of mesh optical networks," *IEEE/ACM Transactions in Networking*, vol. 13, no. 4, pp. 906 – 917, Aug. 2005.

[22] W. Van Heddeghem, F. Idzikowski, W. Vereecken, D. Colle, M. Pickavet, and P. Demeester, "Power consumption modeling in optical multilayer networks," *Photonic Network Communications*, vol. 24, no. 2, pp. 86 – 102, 2012.

[23] J. P. Cardenas, A. Santiago, J. C. Losada, R. M. Benito, and M. L. Mouronte, "On the topology of optical transport networks," *Journal of Physics: Conference Series*, vol. 246, 012013, 2010.

[24] ITU Tutorial on Optical Transport Networks, *ITU*, 2002.

[25] S. K. Routray, R. M. Morais, J. R. F. da Rocha, and A. N. Pinto, "Statistical Model for Link Lengths in Optical Transport Networks," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 5, no. 7, pp. 762 – 773, Jul. 2013.

[26] C. Pavan, R. M. Morais, J. R. F. da Rocha, and A. N. Pinto, "Generating realistic optical transport network topologies," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 2, no. 1, pp. 80 – 90, Jan. 2010.

[27] L. Velasco, M. Klinkowski, M. Ruiz, J. Comellas, "Modeling the routing and spectrum allocation problem for flexgrid optical networks," *Photonic Network Communications*, vol. 24, no.- 3, pp. 177 – 186, 2012.

[28] S. K. Routray, G. Sahin, J. R. F. da Rocha, and A. N. Pinto, "Estimation of Link-Dependent Parameters of Optical Transport Networks from the Statistical Models," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 6, no. 7, pp. 601 – 609, July 2014.

[29] S. K. Routray, J. R. F. da Rocha, and A. N. Pinto, "Estimating the Parameters of Optical Transport Networks from Their Circumferential Ellipses," in Proceedings of IEEE Int. Conf. on Telecommunications 2014, Lisbon, Portugal.

[30] S. K. Routray, G. Sahin, J. R. F. da Rocha, and A. N. Pinto, "Statistical Analysis and Modeling of Shortest Path Lengths in Optical Transport Networks," *IEEE/OSA Journal of Lightwave Technology*, accepted in Mar. 2015.

CHAPTER 2

Statistical Modeling in Optical Transport Networks

2.1 Introduction

IN THE MODERN connected world, the communication traffic is quite large. The services over the Internet are growing every year, and they create very large demands for bandwidth. It is not possible to provide that large bandwidth through the wired metallic conductors in the long range. Right now, only the optical fibers can provide that with an affordable budget. Today the main traffic around the world, whether data, video or voice all are carried predominantly through the optical fiber core networks. Due to the robust nature, and the ability to carry high data rates, optical transport networks (OTNs) are the unrivaled choice for data communications. OTNs are quite versatile, and they are meant to carry all types of traffic such as the packets from various digital devices, Internet Protocol (IP) traffic as well as the OTN, SONET/SDH traffic [1]. OTNs are upgraded regularly to cope with the high demands for increasing traffic. Over time, their features have been changed, and now they have become quite flexible in both the management and operation fronts. Throughout the world, OTNs serve the role of core networks, including the intercontinental carriers. In this chapter, we provide the main features of OTNs in the core networks, and emphasize the need for their statistical modeling.

2.1.1 Motivation

Statistics has vital roles in science and engineering. It is used in several studies, measurements, characterization, analysis, modeling and applications in science and engineering. With the growing complexities in the studies of networks and systems, statistics provide new methods to unearth their mysteries. One of the prime motivations for the statistical modeling of networks is their characterization, which is used in several fields of science, engineering and social science. However, depending on the contexts and areas of application, the motivations for the statistical analysis and modeling of networks are different.

The mathematicians solve the problems related to graphs using the concepts of graph theory which was started accidentally by Euler when he tried to solve the puzzle popularly known as the 'Seven Bridges of Königsberg' [2]. In fact, a lot of network analysis is done in graph theory. In some cases statistical modeling is also used to solve the problems of graphs. For physicists, issues like statistical mechanics (origin and growth) of the networks are of great interest. Generally, a common network modeling aims to explain the time dependent degree distributions of networks and their structural changes such as the percolations [3] which have extensive uses in physics, chemistry and material science. In computational biology, there are numerous applications of statistical network analysis and modeling such as the spread of epidemics, growth of bacteria, rapid infections of viruses (like HIV and Ebola), protein-protein interactions in human metabolism. Communication networks are one of the main users of network statistical models. Since the early days of telegraph to the recent advances of the Internet, we find a lot of applications of network statistics in communication engineering. It helped understanding several complexities of the networks and their properties.

Sociology and Psychology are among the oldest disciplines in which the network concepts were applied [4]. The researches on the 'small-world' behavior (in a small-world network most of the nodes are not neighbors of one another; however, they can be reached from every other node by a small number of hops or links) were being done in Sociology much before the arrival of the modern social networking sites like LinkedIn, Facebook, MySpace and Google+ [4],[5]. The idea of 'six degrees of separation' was borrowed from the 'small-world' concept which was very popular in 1990 [6]. The brain cells are also found to be linked in a 'small-world' pattern [3]. In business and international trade, researchers use the 'small-world' models to increase the efficiency of trade and logistics [8].

The task of finding hidden states and groups are popular in communication networks,

and the sleeper cells of terrorists which are analyzed using the statistical and stochastic methods [3]. Computer networks use the network models and their properties to study the propagation of viruses. In management and businesses, several functions such as marketing and strategy use the concepts of network statistics and also use the modeling aspects. In machine learning, network models are used to predict the missing links. These methods are also applicable to business models to predict the risks and opportunities, and to terrorist networks to find out their missing links [3]. Besides these issues, the uses and applications of network parameters also motivates the use of statistical analysis and modeling. For instance, optimal routing and backup paths need the structural information of the networks and their paths.

For OTNs, the main motivation comes from the utilities of the statistical models in network measurement, planning, design and management. Furthermore, the issues like visualization, security and survivability analysis also rely on the statistical models. There are several statistical models of networks which are applied to the Internet, and other complex networks. The popular statistical models such as the 'small-world', 'scale-free', 'random' and 'complex' have some characteristics common with the OTNs. However, there is no specific model available for the OTNs. Mainly for the early stage planning there are needs for these models.

2.1.2 Related Work

Study of networks in mathematics has its roots in graph theory which was started during the times of Euler [2] in the 18th century. Systematic analysis and characterization of networks in different forms started around 1930s in a few areas of science [4]. Early statistical analysis of the networks are found mainly in the literature of social sciences [7]. The social scientists look for the nexus between different groups of different entities and find out the reason behind their relationships using the networks [3]. These relationships are analyzed in many different aspects such as the strategic alliances, partnerships, and also for rivalries [3]. Problems related to the networks such as the estimation of the shortest paths, diameter and clustering became prominent with the communication networks and their applications [3]. An extensive review of popular statistical models of networks is presented in [3] which includes both static and dynamic types. Network models and their statistical properties were being used in mathematics, social science and business theories before the other branches of science [2] – [8]. In the 1990s, Internet was studied extensively and statistical network models found their ways into communication and computer engineering [4]. Since then, it has been used in several new areas of science, arts and social science in various forms for many different

purposes [1] – [36].

2.1.3 Chapter Outline

The rest part of this chapter is organized in four different sections. In section 2.2, we classify the OTNs on the basis of transparency and present the features of each class. In section 2.3, we separate the elements of the OTNs on the basis of their presence in nodes and links, and then present the functions of the main elements. In section 2.4, we present the needs and uses of statistical analysis and modeling in OTNs using relevant literature. In section 2.5, we conclude the chapter with a few points on the utilities of the work done in this thesis.

2.2 Classification of OTNs

OTNs can be classified in to several categories on different bases. With respect to the work presented in this thesis, we classify them on the basis of transparency. In communication engineering, transparency can be explained as the physical medium (it is optical fiber in case of the OTNs) that can support end-to-end communication of data independent of the bit rates and the signal formats [9]. On the basis of transparency, OTNs can be either fully transparent, or partially transparent (i.e., translucent), or not at all transparent (i.e., opaque). Broadly, OTNs are now divided in to three categories, depending on their ability to handle the changes in the network data rates and types of regenerations along the channel: transparent, translucent and opaque.

This classification can also be explained on the basis of the types of repeaters used along the links. There are three types of repeaters used in optical networks: 1R, 2R and 3R [10]. In 1R repeaters, only optical amplifiers or 1R regenerators are used. In 2R repeaters, both regeneration and reshaping is done, and thus O-E, O-E-O and E-O conversions are required. Similarly in the 3R repeaters, regeneration, reshaping and re-timing is done which also needs O-E, E-O and O-E-O conversions. Depending on the types of the repeaters used, the data rates in the links of the OTNs have different limits. Despite all these diversities in the links of the OTNs, new techniques are available to groom the traffic to optimum data rates, and the overall performance is maintained at the optimum level [1],[10].

2.2.1 Transparent OTNs

In case of the transparent OTNs, the processing along the channel is all-optical. There is no conversion of the optical signal to the electrical domain in the channel be-

tween the source and the destination. The repeaters used along the channel for the transparent OTNs are all-optical. In fact, for the transparent OTNs, only optical amplifiers (1R regenerators) are used along the channel. It has the ability to handle the changes in data rates, protocols, and they can adapt with the changing modulation formats as well. In terms of their adaptability to changes, transparent OTNs are the most flexible. To a large extent, they are future proof as they can handle the increasing demand of the data rates, the changes in the protocols and other basic changes. That is why they are more popular for the long term planning and deployment of long range communications. Erbium doped fiber amplifiers (EDFAs), and other modern optical amplifiers such as the Raman are the main components that enabled the all-optical transparency in the core networks. More and more core optical networks are heading towards transparency.

2.2.1.1 Motivation for Transparent OTNs

These OTNs use only the optical amplifiers (1R repeaters) to strengthen the signal in the fiber when it is weakened over some distance. 1R repeaters do not limit the data rates and the protocols. Thus the limits of the transparent OTNs are determined by signal degradations (nonlinearities of the fiber in most of the cases). Due to these flexibilities, these OTNs are popular for the large and long term operation of networks.

There are enough motivation and rationale behind the deployment of transparent optical networks. Presently, the data rate demand is increasing very fast. This leads to the need of higher network capacity, which in return implies an increase in the bandwidth. But it is not possible to put new fiber in the core network every now and then. The alternative is to have the main trunks, which are able to adapt with the changing traffic and very much durable as far as the changes in the data rates are concerned. The solutions based on this front very much indicate towards the transparent optical networks [1]. These networks gain popularity in the core and other other durable networks. In addition to the data rates, in several other fronts elasticity is required in the modern core optical networks. Only the transparent optical networks can fit into those requirements. Software defined networking (SDN) is a necessity in the modern core networks to optimize the resource allocation and spending. Implementation of SDN is the best over the transparent optical networks. It gives the flexibility needed for the optimization. Furthermore, the latency and delay can be reduced significantly in the transparent optical networks, which are important requirements for the real-time applications and businesses. The extra cost factor that comes into picture for the signal processing in the transparent optical networks over the opaque can frequently be surpassed by the

benefits, because in the core networks the demand for adaptability, and the future proof abilities are very important [1]. Overall, the cost of data transmission tends to be much cheaper in transparent optical networks, when all the factors like adaptability, reconfigurability and longevity are taken into account [1], [9].

2.2.2 Opaque OTNs

In case of the opaque OTNs the ability to adapt to the changes are very limited and they cannot adapt to changes in the data rates as O-E-O conversions (3R regeneration) are used along the channel. 3R regenerations have re-timing and reshaping devices, and these devices are not able to handle the changes in the signal formats and data rates. So, opaque networks have fixed limits for these changes. As the major changes in the adaptable networking technologies emerge, opaque networks are getting out-of-date with every passing year. In Figure 2.1, we show the differences between the opaque and transparent switchings in OTNs. In case of opaque switching, the incoming and outgoing signal are processed in the electrical domain in the switching devices (such as the 2R/3R regenerators and the cross-connects) of the OTNs; whereas in case of transparent switching, the signal remains in the optical domain throughout.

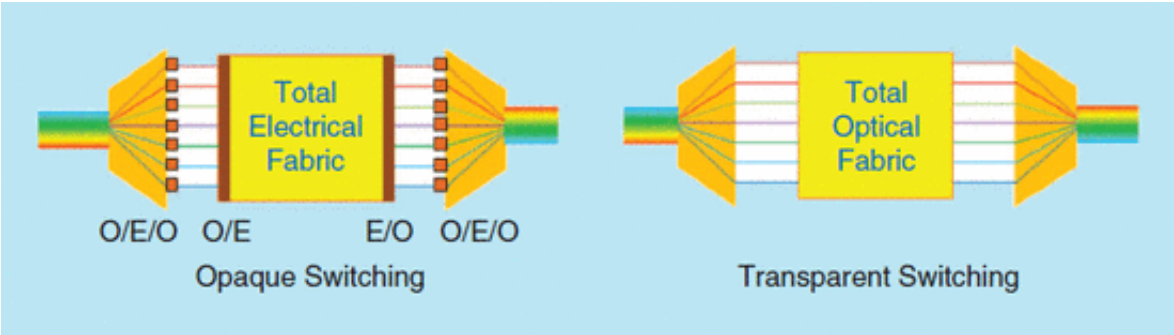

Figure 2.1: Opaque and transparent switchings in OTNs. In opaque OTNs O-E-O conversion is needed; whereas, in case of transparent OTNs the fabric is totally optical [37].

2.2.3 Translucent OTNs

Translucent OTNs are mixed hybrid of the transparent and opaque OTNs. In these OTNs, some of the links have 1R, some have 2R, and some may also have 3R repeaters. Thus translucent networks are not entirely transparent, but they are able to handle the changes in data rates and protocols to some extent (in the links having only 1R repeaters). Translucent networks can play a compromising role in an OTN with respect to the data rates. In the small and highly demanded links they normally provide trans-

parency, and the long and lowly demanded links may be left as opaque channels. As it is very difficult and expensive to convert the whole opaque networks into transparent, most of the old networks are now translucent and their degree of transparency increases over time [11]. As an intermediate choice better than the fully opaque networks, translucent networks used to be very popular in the core OTNs [12] in the last decade.

2.3 Elements of OTNs

OTNs are consist of several network elements. In a typical wavelength routed OTN, several multi-wavelength cross-connect switches are interconnected by the optical fibers with several regeneration facilities along the channel. However, here, we present the network elements and their features, which are relevant to this thesis work. In this work, we broadly separate the elements as: elements of nodes, and elements of links. Nodes are the points where the information is either originated or terminated. Nodes are also the places, where the links are connected with each other. In the nodes, several cross-connects are placed, which work as switches. Both optical and electrical cross-connects are normally found in the nodes. In addition to these switching elements, several signal processing related components are also placed in the nodes according to the requirements of the OTNs. Links are physical connections through which the communication traffic moves directly from one node to the other. Along the links, several network elements such as the optical amplifiers, re-timing devices, add-drop multiplexers, and optical filters are placed according to the requirements.

2.3.1 Elements of Nodes

Nodes are responsible for several important functions of the OTNs such as reception, transmission, internal nodal signal processing, and routing. The nodes in the OTNs have several electronic and optical components to furnish these functions. The main components of a typical node of OTNs are briefly described below.

1. OXCs: Optical Cross-Connects are the high speed optical switches used in the OTNs at the nodes. The cross-connects provide the interconnections between the links, routing, forwarding, regeneration and conversion of wavelengths if required, and other signal switching related functions at the node. In addition to these functions, OXCs also provide traffic provisioning, grooming and restoration facilities [13]. Depending on the networks and nodal arrangements, they can be either transparent or translucent or opaque. Depending on their transparency, the in-

ternal switching fabric of the OXCs can be either completely optical, or partially optical.

2. EXCs: Electrical Cross-Connects are the digital switching devices that provide switching facilities at the nodes in the electrical domain. Normally, the light paths which are not to be processed optically, are bypassed to the EXCs. In majority of the cases, EXCs are used for the traffic that is either dropped or added to a core network at a node. The traffic dropped by the EXC, is usually meant for the local destination, and the ones either bypassed or added is usually meant for the remote destination. EXCs used to perform all the switchings at the nodes before the invention of the OXCs [13].

3. Transponders: Transponders or optical transponders are the devices that receive the light signal, processes it, and then transmit it to the channel. A typical optical transponder adapts the signal coming in from a client of the OTN into a signal suitable for use within the OTN. Similarly, in the reverse direction, it adapts the signal from the OTN into a signal format that is suitable for the client [1]. The interface between the client and the transponder depends on the client specifications such as the bit rate, and distance and/or loss between the client and the transponder [1]. In several cases, the transponders are found inside the optical line terminals (OLTs). However, they can also exist outside the OLTs. Depending on the link lengths, the transponders can be short reach, medium reach or long reach.

4. OLTs: OLTs are the devices used at the ends of a WDM link, typically before and after the OXCs. They are used at either end of a point-to-point link to multiplex, and demultiplex wavelengths using the three functional elements (within an OLT): transponders, wavelength multiplexers, and optionally, optical amplifiers [1]. The OLTs may also terminate the optical supervisory channels (OSCs), which are carried on a separate wavelength, different from the wavelengths carrying the actual traffic, and used for monitoring the performances of the optical amplifiers and other management related functions [1].

2.3.2 Elements of Links

A typical optical link consists of optical fiber and its associated components that help in the communication of the light signal. In case of short links, in which the nodes are not too far apart, the signal processing components are placed in the nodes. However, in case of long and extra-long links, the fibers are provided with several signal regener-

ators or repeaters. The common link related elements are the regenerators of different types. These regenerators include the optical amplifiers, reshaping devices, and the re-timing devices. Along the channel, sometimes wavelength changing is required. For that purpose, optical add / drop multiplexers (OADMs) and reconfigurable optical add / drop multiplexers (ROADMs) are used. The commonly used optical components of a typical OTN are presented below.

1. Optical Amplifiers: These amplifiers (also known as 1R regenerators) work in the optical domain, and strengthen the incoming weak light signal to a stronger one. These amplifiers are deployed in periodic intervals along the optical fiber links. The power needed for this amplification is provided by the pumps that allow the incoming photons to be multiplied by several times as they pass through the gain block of the optical amplifiers [1]. Multiple gain blocks and automatic gain control blocks are provided, where more gain is required [1]. The EDFA, Raman and the semiconductor optical amplifier (SOA) are the popular ones used in the OTNs.

2. 2R Regenerators: These regenerators use optical amplifiers, and reshaping devices for strengthening and repairing the signal. In addition to the strength, these devices also rectify the shape of the signal that usually get distorted after traveling some distance in the fiber. The reshaping is done in the electrical domain using wave shaping circuits. So, O-E-O conversion is mandatory for these regenerators.

3. 3R Regenerators: These regenerators use optical amplifiers, reshaping devices, and re-timing devices for the strengthening and conditioning of the signal. In addition to the strength, these devices rectify the shape and timing of the signal that usually get distorted, and time-shifted after traveling some distance in the fiber. The reshaping and re-timing are done in the electrical domain using wave shaping and re-timing circuits. O-E-O conversion is also mandatory for these regenerators.

4. OADMs: Optical Add / Drop Multiplexers are the devices used in the WDM systems for adding, dropping, forwarding and routing of different light paths into or out of optical channels. These devices are similar to the OXCs to some extent. However, OADMs are smaller switching elements having limited scopes, and used in the links; whereas OXCs are placed in the nodes, and provide large scale switching functions.

5. ROADMs: These are Reconfigurable Optical Add / Drop Multiplexers. In terms of the function, ROADMs are the OADMs which can be reconfigured. Here, reconfigurability means the ability to select the desired wavelengths to be dropped

and added to the optical channel, without having prior plan during the deployment phase as it is done in case of the OADMs [1]. This reconfigurability, provides a lot of advantages in the network operations such as the flexibility of adding and dropping the light paths without the change of wavelengths, remote configuration, and remote power balancing. Overall, it increases the flexibility in the OTNs, and thus very much important.

2.3.3 Key Enabling Technologies for OTN Deployment

In typical OTNs, the above mentioned components are required for the basic functioning. In addition to the components, the enabling technologies too affect the cost and function of the OTNs. The communication in the OTNs is enabled by some key technologies those make the system work continuously. In OTNs, the switching is provided by selecting or rejecting the wavelengths. So, the frequency division multiplexing technique used in the optical domain is commonly known as the wavelength division multiplexing (WDM). The main enabling technologies in the optical domain are: WDM, optical amplifiers, ROADM, coherent receivers, efficient modulation formats, and the low-cost access area techniques. However, with the progress in research, now several new advanced technologies such as the orthogonal frequency division multiplexing (OFDM), digital signal processing, compensation techniques and SDN have also been used. Depending on the enabling technologies used, the span and capacity of the OTNs, the capital, operational and other associated expenditures for the deployment are found to be different in cost and performance.

2.4 Statistical Analysis and Modeling in OTNs

Statistical analysis is vital for the characterization of networks, and now it has crossed beyond its traditional fields of graph theory and circuit theory. A new area of study on networks, called 'network science and engineering' has emerged which is being used in almost every field, where networks are used [14] – [25]. OTNs are large core networks, and they carry huge traffic around the world. They are the backbone of the modern communication systems, and thus the reliability and safety of these networks are of immense importance. In order to study the characteristics of the OTNs systematically, their statistical analysis is necessary. In the last decade, statistical analysis, modeling, formalism and estimation related issues of OTNs have become very popular [26] – [30]. Statistical modeling of network parameters is helpful in the early stage of planning and design when there is not enough data available.

The knowledge of statistical distribution of OTN parameters provide clear idea about its properties. In the novel paradigms, such as EON and SDN, there are always the need of statistical data of the OTNs such as the average link length, shortest path lengths, back up paths, and the express links [32] – [34]. For flexible operations and management of the modern OTNs, statistical information is necessary [34]. Besides the above reasons, the study of network statistics and stochastic natures have now spread in almost all the associated fields of networks, and all types of contemporary networks such as the Internet, metabolic networks, social networks, cellular networks, trade networks and several other complex networks are being studied and modeled for their physical characterization and application [14] – [31]. However, in optical communication engineering, there are not too many statistical models for their characterization. Therefore, we aim to fill this void, and apply the developed models for the related engineering estimations.

2.4.1 Statistical Modeling of Node Related Parameters

Nodes may serve as either the source or sink of information. In fact, in the modern OTNs, the nodes are both source and sink of information, and they also serve as the main signal processing hubs. Nodes also route the traffic from one link to the other. In the large OTNs, very often their topologies are survivable (a topology is survivable, if its each node is connected with two or more links [27]). OTN topology survivability can be quantified in terms of nodal degree. Nodal degree of a node is the number of links connected to that node. A survivable topology has nodal degree always larger than or equal to 2 for each node.

A lot of work have been done on the nodal degrees of the networks of different types [17] – [31]. Mainly, in the complex networks, such as the Internet, it is a popular topic as it provides a lot of information about the structural mechanics of the networks and their stability related issues [14] – [17]. In case of OTNs too, statistics of nodal degrees provides several information about the network and its features. In addition to survivability, nodal degrees and the mean nodal degree provide other information about the structure and function of the OTNs. When the mean nodal degree is exactly equal to 2, the survivable network becomes a simple ring. A central node (one node connected with almost all other nodes) has high nodal degree, and typically OTNs having central nodes exhibit high mean nodal degrees. In a similar fashion, the overall statistical distribution of the nodal degrees of an OTN provides the connection related characteristics of the network. In the early days of optical communications, OTNs used to have a Poisson's distribution for the nodal degrees [27], [31]. However, from the measurement, we found that several networks in the recent times have shown a deviation from that, and

they exhibit scale-free characteristics.

2.4.2 Statistical Modeling of Links and its Related Parameters

Links in OTNs play the important role of connecting the nodes through the fiber channels. The lengths of the links are the lengths of the fibers that carry information from one node to the other. Total sum of link lengths provides the length of the fiber needed for a network. In addition to these directly related parameters, there are several other link-dependent parameters such as the number of amplifiers needed in a network, types of modulation and demodulation schemes required at the ends of the links (i.e., nodes), types of compensation schemes needed, and the number of preamplifiers needed. However, there used to be no statistical model for link lengths of OTNs. In this thesis work, we develop a statistical model for the link lengths in OTNs which is presented in Chapter 3 (reported in [35]), and its applications for the real networks in Chapter 5 (reported in [36]).

In case of routing and forwarding of traffic in OTNs, shortest paths between the nodes play key roles. In case of the transparent OTNs, shortest path lengths are needed in determining the types of modulation-demodulation schemes, and compensation methods to be used at the nodes. In addition to that, there are several other cases in which shortest path lengths play vital roles in the network related estimations. These information can be used in both the data and control plane related operations [38]. For shortest path lengths of OTNs, unfortunately, there used to be no statistical model. We develop such a model for shortest path lengths in OTNs which is presented in Chapter 6 and its applications in Chapter 7.

2.5 Chapter Summary

OTNs are the backbone of the modern communication networks. They carry almost all the traffic in the core, which runs the whole Internet and several other networks. Without the OTNs, there would not be the Internet as we see it today. Considering on the essence of the OTNs in modern communication, their statistical modeling and characterization become very important. Statistical modeling helps understanding the OTNs better. In the planning and design stage, it provides very important information to the network architects. Statistical modeling can be used for the estimation of the network parameters from the basic information. The statistical models also explain a lot about the structure and dynamics of the networks. In the present scenario, statistical analysis and modeling of OTNs has a great role to play in both their study and engineering

applications.

References

[1] R. Ramaswamy, K. N. Shivrajan, and G. H. Sasaki, *Optical Networks: A Practical Perspective*, San Francisco, CA: Morgan Kaufman, 2009.

[2] G. L. Alexanderson, "Euler and Königsberg's Bridges: A historical view," *Bulletin of the American Mathematical Society*, vol. 43, no. 4, pp. 567 – 573, Oct. 2006.

[3] A. Goldenberg, A. X. Zheng, S. E. Fienberg, E. M. Airoldi, "A survey of statistical network models," *Foundations and Trends® in Machine Learning*, vol. 2, no. 2, pp. 129 – 233, Jun. 2010.

[4] E. D. Kolaczyk, *Statistical Analysis of Network Data,* New York, NY: Springer, 2009.

[5] S. Milgram, "The small world problem," *Psychology today*, vol. 2, no. 1, pp. 60 – 67, 1967.

[6] J. Guare, *Six degrees of separation: A play*, New York, NY: Random House LLC, 1990.

[7] S. E. Fienberg, "A brief history of statistical models for network analysis and open challenges," *Journal of Computational and Graphical Statistics*, vol. 21, no. 4, pp. 825 – 839, Dec. 2012.

[8] A. Wilhite, "Bilateral trade and 'small-world' networks," *Computational Economics*, vol. 18, no. 1, pp. 49 – 64, 2001.

[9] B. Ramamurthy, H. Feng, D. Dutta, J. P. Heritage, B. Mukherjee, "Transparent vs. Opaque vs. Translucent Wavelength-Routed Optical Networks," in the Proceedings of the OFC, pp. 59 – 61, Mar. 1999.

[10] B. Mukherjee, "WDM optical communication networks: progress and challenges," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 10, pp. 1810 – 1824, Oct. 2000.

[11] G. Shen, and R. Tucker, "Translucent optical networks: the way forward ," *IEEE Communications Magazine*, vol. 45, no. 2, pp. 48 – 54, Feb. 2007.

[12] X. Yang, and B. Ramamurthy, "Dynamic routing in translucent WDM optical networks: the intradomain case," *IEEE/OSA Journal of Lightwave Technology*, vol. 23, no. 3, pp. 955 – 971, Mar. 2005.

[13] N. A. Jackman, S. H. Patel, B. P. Mikkelsen, and S. K.Korotky, "Optical cross connects for optical networking," *Bell Labs Technical Journal*, vol. 4, no. 1, pp. 262 – 281, Jan. 1999.

[14] A. L. Barabási, and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509 – 512, 1999.

[15] R. Albert, H. Jeong, and A. L. Barabási, "Internet: Diameter of the world-wide web," *Nature*, vol. 401, pp. 130 – 131, 1999.

[16] L. A. Adamic, and B. A. Huberman, "Power-law distribution of the world wide web," *Science*, vol. 287, pp. 2115 – 2115, 2000.

[17] R. Albert, and A. L. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, pp. 47 – 97, Jan. 2002.

[18] Y. Y. Ahn, J. P. Bagrow, and S. Lehmann, "Link communities reveal multiscale complexity in networks," *Nature*, vol. 466, pp. 761 – 764. 2010.

[19] A. L. Barabási, N. Gulbahce, and J. Loscalzo, "Network medicine: a network-based approach to human disease," *Nature Reviews Genetics*, vol. 12, no. 1, pp. 56 – 68, 2011.

[20] M. Vidal, M. E. Cusick, and A. L. Barabási, "Interactome networks and human disease," *Cell*, vol. 144, no. 6, pp. 986 – 998, 2011.

[21] E. Bullmore, and O. Sporns, "The economy of brain network organization," *Nature Reviews Neuroscience*, vol. 13, no. 5, pp. 336 – 349, 2012.

[22] B. Barzel, and A. L. Barabási, "Universality in network dynamics," *Nature physics*, vol. 9, no. 10, pp. 673 – 681, 2013.

[23] M. Schich, C. Song, Y. Y. Ahn, A. Mirsky, M. Martino, A. L. Barabási, and D. Helbing, "A network framework of cultural history," *Science*, vol. 345, pp. 558 – 562, 2014.

[24] S. Suweis, F. Simini, J. R. Banavar, and A. Maritan, "Emergence of structural and dynamical properties of ecological mutualistic networks," *Nature*, vol. 500, pp. 449 – 452, 2013.

[25] W. Van Heddeghem, F. Idzikowski, W. Vereecken, D. Colle, M. Pickavet, and P. Demeester, "Power consumption modeling in optical multilayer networks," *Photonic Network Communications*, vol. 24, no. 2, pp. 86 – 102, 2012.

[26] S. K. Korotky, "Network global expectation model: A statistical formalism for quickly quantifying network needs and costs," *IEEE/OSA Journal of Lightwave Technology*, vol. 22, no. 3, pp. 703 – 722, Mar. 2004.

[27] C. Pavan, R. M. Morais, J. R. F. da Rocha and A. N. Pinto, "Generating Realistic Optical Transport Network Topologies," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 2, no. 1, pp. 80 – 90, Jan. 2010.

[28] E. Bouillet, *Path routing in mesh optical networks*, Chichester, England: John Wiley & Sons, 2007.

[29] J. F. Labourdette, E. Bouillet, R. Ramamurthy, and A. A. Akyamaç, "Fast approximate dimensioning and performance analysis of mesh optical networks," *IEEE/ACM Transactions in Networking*, vol. 13, no. 4, pp. 906 – 917, Aug. 2005.

[30] K. M. Sivalingam, and S. Subramaniam,(Eds.) *Emerging optical network technologies: architectures, protocols and performance,*New York, NY: Springer, 2005.

[31] B. Waxman, "Routing of multipoint connections," *IEEE Journal of Selected Areas in Communications*, vol. 6, no. 9, pp. 1617 – 1622, Dec. 1988.

[32] M. Channegowda, R. Nejabati, and D. Simeonidou, "Software defined optical networks technology and infrastructure: Enabling software-defined optical network operations," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 5, no. 10, pp. A274-A282, Oct. 2013.

[33] R. Ramamurthy, J. F. Labourdette, S. Chaudhuri, R. Levy, "Routing light-paths in optical mesh networks with express links," in the Proceedings of OFC, pp. 503 – 504, 2002.

[34] L. Velasco, M. Klinkowski, M. Ruiz, J. Comellas, "Modeling the routing and spectrum allocation problem for flexgrid optical networks," *Photonic Network Communications*, vol. 24, no. 3, pp. 177 – 186, 2012.

[35] S. K. Routray, R. M. Morais, J. R. F. da Rocha, and A. N. Pinto, "Statistical Model for Link Lengths in Optical Transport Networks," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 5, no. 7, pp. 762 – 773, Jul. 2013.

[36] S. K. Routray, G. Sahin, J. R. F. da Rocha, and A. N. Pinto, "Estimation of Link-Dependent Parameters of Optical Transport Networks from the Statistical Models," *Journal of Optical Communication and Networking*, vol. 6, no. 7, pp. 601 – 609, July 2014.

[37] S. K. Routray, "The Changing Trends of Optical Communication," *IEEE Potentials Magazine*, vol. 33, no. 1, pp. 27 – 33, Feb. 2014.

[38] R. Casellas, R. Muñoz, R. Martìnez, and R. Vilalta, "Applications and Status of Path Computation Elements," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 5, no. 10, pp. A192 – A203, Oct. 2013.

CHAPTER 3

Statistical Modeling of Link Lengths

3.1 Introduction

IN THE FIRST two chapters, we have discussed the importance of statistical analysis and modeling of network parameters in optical transport networks (OTNs). In this chapter, we develop a statistical model based on probability density function (PDF) for the link lengths in OTNs. In oder to make the distribution model useful in the early stages of network planning, we develop expressions for the estimation of its parameters with only the knowledge of the node locations. The parameters of the developed model are found to depend on the average link length of the OTN. In this chapter, we also improve the expressions for the average link length of the OTN for better accuracy. The validity of this model for link lengths is tested by comparing it with the statistical properties of link lengths in real OTNs.

3.1.1 Motivation

Very often, in the early stage planning and dimensioning, the network engineers do not have complete information of the OTNs. These cases have to be dealt with partial or incomplete information of the OTNs. In the majority of such cases, parameters available at the early stage are the coverage area, and the number of nodes of the network. In these situations, some knowledge of the statistical distribution is required for network

planing and design. Several fast estimations without complete information are also possible through these models. For instance, determination of compensation techniques at the receiving end, number of repeaters and amplifiers along the optical links, and total length of fiber needed can be estimated using a link statistical model. Thus, there is need of a statistical model of link lengths, which can provide the estimations related to the OTNs with partial information. In addition to that, the model needs to be as accurate as possible so that the outcomes of the estimations are quite reliable for practical applications. That is why, the developed model should be tested over the existing real OTNs.

3.1.2 Related Work

Though there was no prior model available for the link lengths of the OTNs, some statistical formalism for the parameters of OTNs were proposed in [1] – [4] . In [1], a semi-empirical expression is proposed for the average link length of OTNs. This expression in [1], depends only on the coverage area and the number of network nodes. Topological analysis are helpful in the study of the structures of OTNs. In [3], SDH network topologies of Telefónica in Spain (these networks are commercial OTNs) are analyzed with respect to their pattern of randomness with some known parameters, and the roles of their link lengths in it. The obtained empirical results show that the OTNs have characteristic traits of complex systems at different scales, both at national and provincial levels [3]. It is also found that the OTNs exhibit several scale-free and small-world network properties [3]. In [4], artificial OTN topologies are generated through simulation using the real characteristics of the OTNs. These topologies are survivable in nature (i.e., every node is connected with 2 or more links). OTN parameters and statistics such as the nodal degree and number of hops are estimated and compared with the real networks in [4]. In [5], artificial OTN topologies are generated and optimized for CAPEX using genetic algorithms. Accuracy of average link length estimation provides a good basis for dimensioning and calculation of other link and network related parameters, such as the network traffic gain [6]. In [7] and [8], the role of links and link lengths are analyzed in various network traffic modeling, and its related statistical distributions and comparisons. In [1] – [7], several expressions are proposed for the quick estimation of network attributes with incomplete information and their subsequent applications in different cases. Majority of these cases just take the very basic parameters to estimate the network requirements. In [10], network topologies and the role of link lengths in their growth and expansion are addressed in general for different types of networks.

3.1.3 Chapter Outline

We organize the rest of this chapter in 4 different sections. In section 3.2, we analyze the real OTN topologies, and develop a new model to describe the statistical properties of their link lengths. In this section, we present the expressions developed for the estimation of the parameters of the proposed statistical model, which depend on the average link length. In section 3.3, we present different approaches for the estimation of average link length of OTNs. We propose three different areas for these estimations. In this section, we evaluate the expressions for their accuracy. We found the estimation based on the convex area of the OTNs provide the best estimate for average link length. In section, 3.4, we estimate the developed model from the average link length. We show that the resulting model for estimating the average link length depends only on the node locations. In section 3.5, we present the main summary and a discussion of the effectiveness of the proposed model. The work presented in this chapter is reported in [11].

3.2 Statistical Analysis of Links in Real OTNs

For this study of the statistical behavior of the link lengths, we used the topologies of the real OTNs. The topologies used for this study are widely known, and have been used for the previous studies of related works on OTNs. In this study, we analyzed 40 real OTNs to develop a general statistical model for the link length distribution. Out of these 40 OTNs, 29 are mentioned and used in [4], and the rest are listed in [12] – [23]. The graphics of these topologies, and other basic parameters of the OTNs can be found in [12].

3.2.1 Measurement of Exact Link Lengths

Accurate measurement of the link lengths of the OTNs is essential for their statistical analysis. In this work, we used two software tools for the measurement of link lengths. We used OPNET Transport Planner 15 (a commercial software tool for network dimensioning and optimization) to find out the exact great circle (the shortest distance along the surface of earth) link lengths of the real OTNs. The measurement accuracy of the link lengths were cross-checked using Google Earth Professional (version 6.2). The measurements using OPNET Transport Planner 15 have very small errors (the maximum error is less than 1%) with respect to the measurements of Google Earth Professional.

Table 3.1: Best Fitting Distributions and Their Average (Avg. KSS), Lowest (L. KSS) and Highest (H. KSS) KSS Values

#	Distribution	Avg. KSS	L. KSS	H. KSS
1	General Extreme Value	0.1061	0.0360	0.1785
2	Log-Logistic (3P)	0.1096	0.0476	0.2154
3	Pearson 5 (3P)	0.1098	0.0411	0.1877
4	Log-Normal (3P)	0.1123	0.0411	0.1997
5	Log Pearson 3	0.1133	0.0469	0.2101
6	Frechet (3P)	0.1145	0.0413	0.2034
7	Inverse Gaussian (3P)	0.1159	0.0403	0.2276
8	General Pareto	0.1164	0.0610	0.2107
9	Burr	0.1181	0.0337	0.3539
10	Pearson 6	0.1226	0.0419	0.2246
11	WeiBull (3P)	0.1251	0.0531	0.2581
12	Weibull	0.1318	0.0429	0.2567
13	Gamma	0.1324	0.0389	0.4169
14	Log-Gamma	0.1359	0.0731	0.2461
15	Log-Logistic	0.1392	0.0709	0.2815

(3P) *indicates the 3-parameter version of the distribution*

3.2.2 Selection of Appropriate Distribution

In order to get an appropriate distribution, we analyzed all the exact link lengths obtained from the above measurements, and studied their statistical properties for 40 real OTNs. From the analysis, we found that the link length statistics do not fit with the commonly used distributions such as the normal or Poisson. In addition to that, none of the one- and two-parameter distributions are found to be suitable for the link lengths. For these reasons, we extended our analysis to a wide range of distributions. For this statistical analysis, we used EasyFit software (of Mathwave.com), and also cross-checked using Matlab, wherever possible. Out of the 61 distributions used for this study, the top 15 best fitting are presented in Table 3.1. We measure the validity or the 'goodness of fit' of the distributions in terms of Kolmogorov-Smirnov (KS) statistic (KSS value), which is illustrated in Appendix A. Because KS test is suitable for small samples, and can also be applied to large ones (fits our case, in which, the number of links of networks, L, range from 11 to 171). The confidence interval (CI), of all the KS tests are set to 0.95 (i.e., significance level = 0.05). The smaller the KSS is, the smaller is the maximum difference between the hypothesized cumulative distribution function (CDF) and the empirical distribution function of the real data (thus better fitting is the distribution relative to the sample). The ranking of the best-fitting distributions in Table 3.1 has been done as per the average KSS values (column 'Avg. KSS' in Table 3.1). These values were obtained for each distribution, by taking the average over the KSS values obtained for each one of the 40 networks studied.

As can be seen in Table 3.1, the average KSS value of General Extreme Value (GEV) distribution is the smallest among the 61 distributions analyzed for this study. In this distribution, the probability density function (PDF) increases initially and then decreases with the increase of link lengths. This behavior can be observed in Figure 3.1, which shows the real link length histogram and the best fit distribution for the USA100 network (having 171 links). The exact link lengths of the 40 real networks follow the GEV distribution to a large extent. It is better than its nearest rivals in almost all aspects. As can be seen in Table 3.1, the highest value of KSS (H. KSS), is the lowest for this distribution. In case of lowest value of KSS (L. KSS), the GEV distribution is very much close to the Burr distribution (L. KSS column of Table 3.1), but is lower than the values for all other distributions.

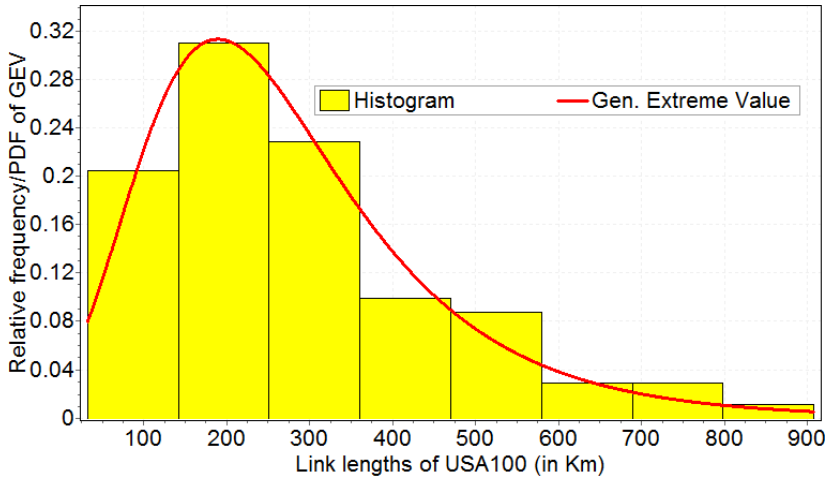

Figure 3.1: Comparison of the General Extreme Value distribution with the link length histogram of USA100 [1] network (171 links)

Basic attributes of the 40 real OTNs such as the number of nodes (N), links (L), and average nodal degree (average number of links connected to any node, $\langle D \rangle$) are presented in Table 3.2. The average and standard deviation of link lengths of the real topologies are calculated using expressions (3.1) and (3.2), respectively.

$$\langle l \rangle = \frac{1}{L} \sum_{i=1}^{L} l_i \tag{3.1}$$

$$\sigma_l = \left(\frac{1}{L} \sum_{i=1}^{L} (\langle l \rangle - l_i)^2 \right)^{\frac{1}{2}} \tag{3.2}$$

In expressions (3.1) and (3.2), l_i represents individual link lengths, and $\langle l \rangle$, is the average link length in kilometer. KSS values of General Extreme Value distribution for each network are listed in the last column of Table 3.2 (i.e., column 'KSS_G'). We

also calculated the average and standard deviation of the average link lengths of the 40 networks (shown at the bottom of '$\langle l \rangle$' column in Table 3.2). Those numbers give the idea about the link length diversity of the 40 real OTNs.

Table 3.2: Real OTN Topologies and Their Attributes ($\langle l \rangle$ and σ_l are in km).

#	Network	N	L	$\langle D \rangle$	$\langle l \rangle$	σ_l	KSS_G
1	VIA Network[4]	9	12	2.67	571	338	0.1744
2	BREN[4]	10	11	2.20	94	33	0.1785
3	RNP[4]	10	12	2.40	748	517	0.1709
4	Abilene Core[15]	10	13	2.60	1067	516	0.1266
5	LEARN[13]	10	12	2.40	189	111	0.1476
6	CompuServe[14]	11	14	2.55	1161	859	0.1543
7	vBNS[4]	12	17	2.83	965	563	0.1005
8	CESNET[4]	12	19	3.17	91	31	0.1425
9	NSFNET[4]	14	21	3.00	1086	707	0.1014
10	ITALY[4]	14	29	4.14	280	146	0.0783
11	ACONET[4]	15	22	2.93	119	111	0.1397
12	MZIMA[4]	15	19	2.53	852	430	0.0896
13	GARR-B[4]	16	27	3.37	224	140	0.1354
14	ARNES[4]	17	20	3.25	38	19	0.1388
15	GERMANY[4]	17	26	3.06	143	78	0.1057
16	REDIRIS[4]	17	28	3.29	319	131	0.0820
17	LambadaRail[4]	19	23	2.42	671	378	0.0901
18	MEMOREX[4]	19	24	2.53	137	68	0.0865
19	CANARIE[4]	19	26	2.74	668	632	0.0685
20	EON[4]	19	37	3.89	754	356	0.1043
21	ARPANET[4]	20	32	3.20	839	491	0.0739
22	OPTOSunet[17]	20	24	2.40	100	34	0.1219
23	Hibernia USA[18]	20	27	2.70	279	207	0.0823
24	PIONIER[4]	21	25	2.38	131	55	0.1418
25	COX [4]	24	40	3.33	662	428	0.0673
26	SANET[4]	25	28	2.24	36	14	0.1553
27	NEWNET [4]	26	31	2.38	528	268	0.1183
28	PORTUGAL[4]	26	36	2.77	203	325	0.1593
29	RENATER[4]	27	35	2.59	155	66	0.0790
30	IBN31[19]	31	51	3.29	131	59	0.0914
31	BULGARIA[20]	32	33	2.06	51	21	0.1039
32	GEANT2[4]	32	52	3.25	661	455	0.0716
33	LONI[4]	33	37	2.24	62	15	0.0886
34	METRONA[4]	33	41	2.48	73	40	0.0798
35	COST37[21]	37	57	3.08	439	248	0.0840
36	CERNET[22]	37	53	2.86	636	503	0.0698
37	OMNICOM[4]	38	54	2.84	298	153	0.0796
38	INTERNET2[4]	56	61	2.18	334	186	0.0694
39	CORONET[23]	75	99	2.64	326	265	0.0546
40	USA100[1]	100	171	3.42	310	174	0.0360
	Average:				410.78		0.1061
	Standard Deviation:				332.83		—

3.2.3 Basics of GEV Distribution of Link Lengths

The General Extreme Value distribution [24] has three parameters: α, β and ξ. The parameter α, is the location factor, β, is the scale, and ξ, is the shape factor. Its CDF, $F(l)$, and PDF, $f(l)$, as functions of link length l are shown in (3.3) and (3.4), in which, $t = 1 + \xi(l - \alpha)/\beta$.

$$F(l;\alpha,\beta,\xi) = F(t;\xi) = \exp\left\{-[t]^{-1/\xi}\right\} \tag{3.3}$$

$$f(l;\alpha,\beta,\xi) = f(t;\beta,\xi) = \frac{1}{\beta}[t]^{(-1/\xi)-1}\exp\left\{-[t]^{-1/\xi}\right\} \tag{3.4}$$

In Figure 3.1, we show the PDF of this distribution as fitted with the link length histogram of USA100 network (α = 196.66, β = 128.89, and ξ = 0.0597). As can be seen in Table 3.2, the KSS_G value for this network is 0.0360, which indicates the appropriateness of the GEV distribution for this network. The KSS_G values for the other networks in Table 3.2 are also generally low, resulting in good average value of 0.1061 (last row of column 'KSS_G'), when all 40 networks are considered. The higher KSS values are obtained for the networks having fewer links or irregular shapes (few links are either too long or too short than the average, e.g., VIA Network, BREN, RNP, PORTUGAL, ACONET).

3.2.4 GEV Distribution Parameters and $\langle l \rangle$

The relationship between the first moment of the link lengths, $\langle l \rangle$, and parameters of the GEV distribution are key to understand and estimate the parameters of the distribution. The location factor, α, gives the location of the distribution. The longer are the link lengths, the larger is the α of the distribution. The scale factor, β, gives the spread out of the distribution (β is proportional to the standard deviation). The shape factor, ξ, controls the shape of the distribution (both PDF and CDF). There are the following bounds on l and ξ with respect to their domains:

i. $l \in (\alpha - \frac{\beta}{\xi}, \infty)$, when, $\xi > 0$;

ii. $l \in (-\infty, \infty)$, when, $\xi = 0$;

iii. $l \in (-\infty, \alpha - \frac{\beta}{\xi})$, when, $\xi < 0$.

The three bounds shown in expressions i - iii above are theoretical ones [24]. In case of OTNs, l is always greater than 0. Out of the 40 OTNs of Table 3.2, the optimum PDF fitting led to $\xi > 0$, in 26 networks (i.e., case i), and in the rest 14 networks, $\xi < 0$, (i.e., case iii).

In order to determine the relation between the statistical parameters of GEV distribution, α and β, and the average link length ($\langle l \rangle$) of the networks, we started by plotting the value of each parameter as a function of the average link length. Then a linear regression over the plotted data was used to get the approximate expressions for α and β. As shown in Figure 3.2 and Figure 3.3, the linear regressions result in expressions (3.5) and (3.6), with good coefficients of determination (i.e., R^2).

$$\alpha \approx 0.6577\langle l \rangle + 8.67 \tag{3.5}$$

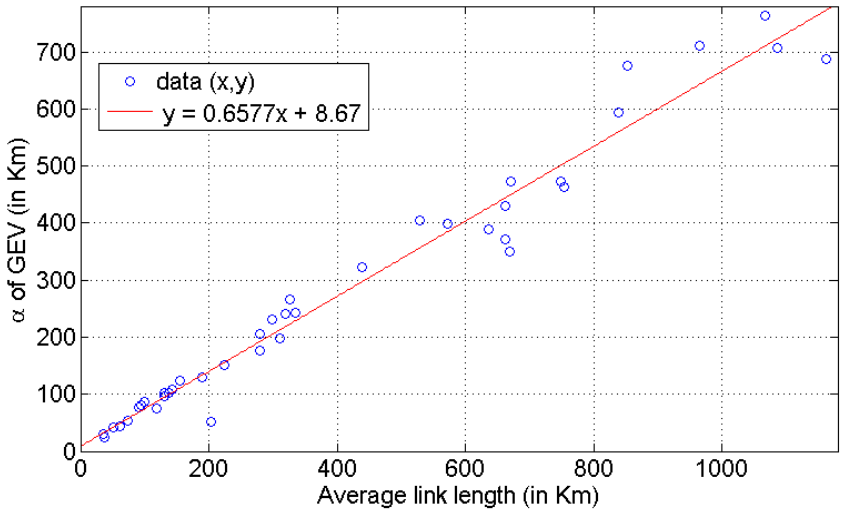

Figure 3.2: Plot of α of GEV vs. $\langle l \rangle$ of 40 real OTNs (R^2 = 0.9676)

$$\beta \approx 0.441\langle l \rangle - 12.37 \tag{3.6}$$

From the analysis, we found that $\xi < 1$, for all the 40 OTNs. Under this condition, we show the exact expression for the mean, $E(l)$, of the GEV distribution in terms of α, β and ξ in (3.7). The error in the value of the first moment, obtained from (3.7) with respect to the exact average link length (i.e., $\langle l \rangle$ in Table 3.2) is very small.

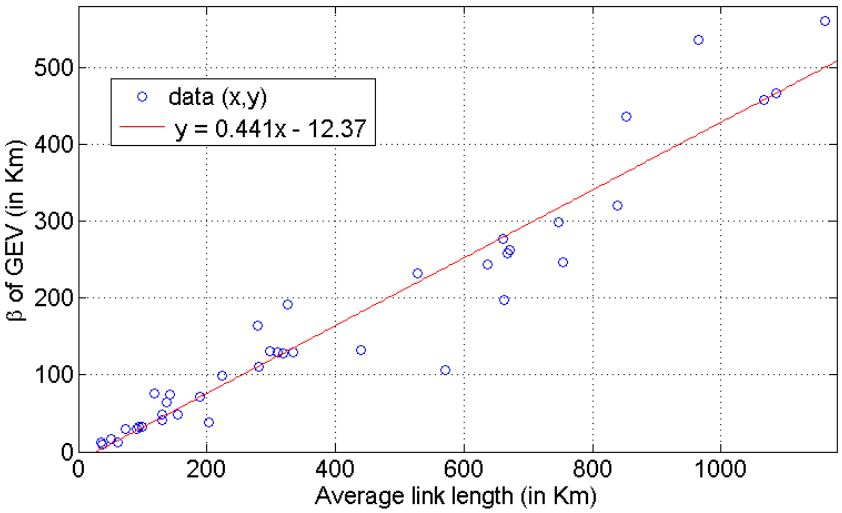

Figure 3.3: Plot of β of GEV vs. $\langle l \rangle$ of 40 real networks (R^2 = 0.9225)

$$E(l) = \alpha - \frac{\beta}{\xi} + \frac{\beta}{\xi}\Gamma(1-\xi) \tag{3.7}$$

In expression (3.7), $\Gamma(\cdot)$, represents the mathematical gamma function (see Appendix B for more information). Assuming, $\langle l \rangle$ is equal to $E(l)$ of the GEV distribution, from expression (3.7), we obtain:

$$\frac{\langle l \rangle - \alpha}{\beta} = \frac{\Gamma(1-\xi) - 1}{\xi}. \tag{3.8}$$

Estimation of the shape parameter, ξ using (3.8) is quite accurate, if the exact values of $\langle l \rangle$, α and β are used. But the gamma function, $\Gamma(\cdot)$ in (3.8) needs to be resolved for a simpler expression. In order to obtain an explicit expression for ξ, we substitute the right hand side of expression (3.8) with the following simple approximation (which is same as expression B.3 in Appendix B):

$$\frac{\Gamma(1-\xi) - 1}{\xi} \cong \frac{1}{1-\xi} - 0.425. \tag{3.9}$$

Using this substitution along with expressions (3.5) and (3.6) in (3.8), we obtain the following approximate expression for ξ (same as expression B.5 in Appendix B):

$$\xi \approx \frac{0.0887\langle l \rangle - 1.557}{0.5297\langle l \rangle - 13.927}. \tag{3.10}$$

For large OTNs, the expression in (3.10) provides good results. From this expression, it can be seen that for large OTNs (for which, $\langle l \rangle \gg 26.3$ km), we can use the approximation $\xi \approx 0.167$. In our model, we use this value for all the 40 OTNs. The best-fit and the estimated distributions (using (3.5), (3.6), and $\xi = 0.167$) of USA100 network are shown in Figure 3.4. The impact of using an estimated average link length is studied in the next section, after the development of an expression for this parameter.

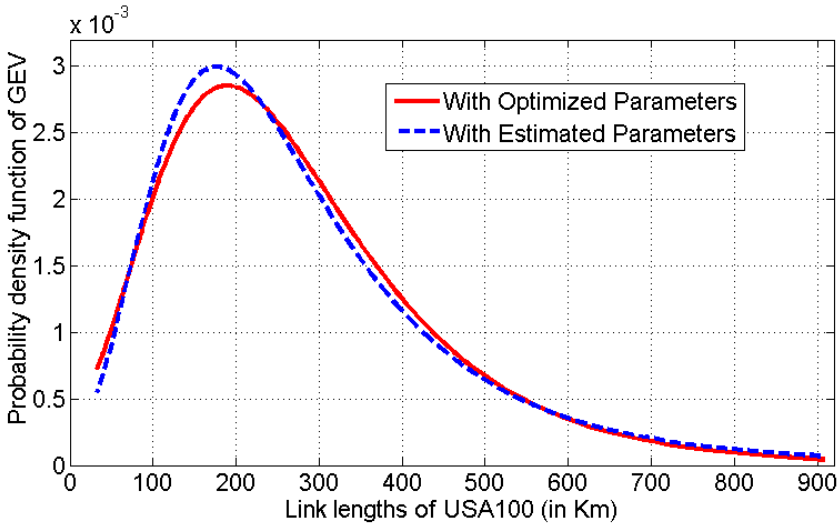

Figure 3.4: Optimized vs. approximate GEV distribution of link lengths of USA100[1] network. This estimation uses exact average link length, $\langle l \rangle$.

GEV distribution is the generalized version of three lognormal distributions. As its name suggests, it is the generalized form of three extreme value distributions: Gumble,

Fréchet and Weibull [24]. In terms of the shape, the PDF of GEV distribution is similar to that of normal distribution (i.e., symmetrical about the mode), when $\xi \approx -0.2064$ (as in the case of MZIMA). When, ξ is closer to 1, its PDF curve is tilted towards the left (i.e., towards the shorter link lengths), and when its value is closer to -1, it is tilted towards the right (i.e., towards the longer link lengths). Overall, the negative values of ξ moves the peak of the PDF curve (i.e., the mode) away from the smaller values.

3.2.5 Relationship Between σ_l and $\langle l \rangle$

We have presented the relationships between $\langle l \rangle$ and the parameters of the GEV distribution in the previous subsection. These relationships reveal several information of the distribution, and provide outlook for their estimations. In this subsection, we verify the relationships between the first moment (i.e., $\langle l \rangle$), and the standard deviation (i.e., σ_l) of the distribution. As shown in the Table 3.2, the standard deviation of the link lengths of the 40 real OTNs is large. For $\xi < 1$, and $\xi \neq 0$, the standard deviation of the GEV distribution is given by expression (3.11).

$$\sigma_G = \frac{\beta}{\xi}\sqrt{\Gamma(1-2\xi) - [\Gamma(1-\xi)]^2} \tag{3.11}$$

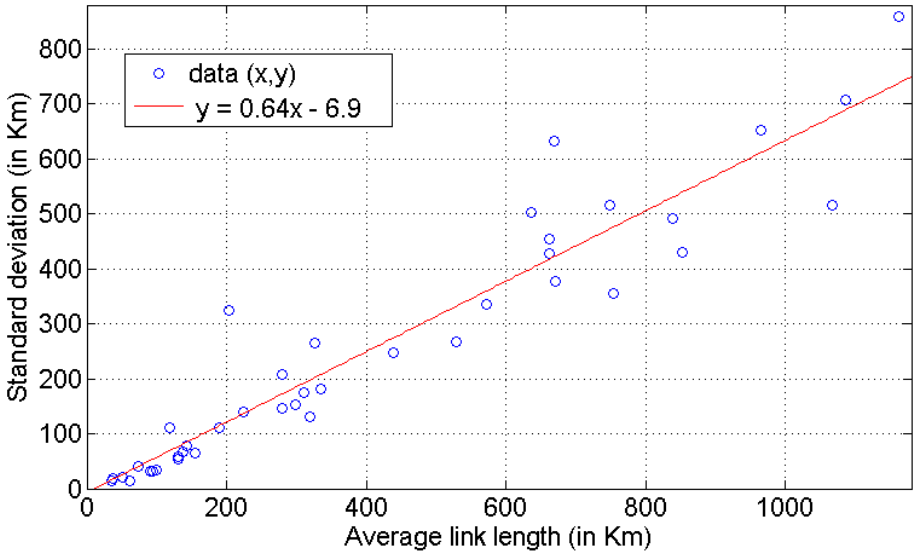

Figure 3.5: Standard deviation vs. average link length of 40 real networks follows a linear trend (R^2 = 0.9011)

The complexity of this expression is that it has two gamma functions. Instead of using the gamma functions along with the estimates of β and ξ, we can approximate (3.11) by a linear regression. We found that in real OTNs, the standard deviation, σ_l, follows an approximate linear trend with the average link length as shown in expression

(3.12), which corresponds to the linear regression in Figure 3.5. This expression is free from both the gamma functions, and the parameters of the GEV distribution.

$$\sigma_l \approx 0.64\langle l \rangle - 6.9 \tag{3.12}$$

We also found that σ_l follows a linear trend with the scale factor of the GEV distribution, β. It again confirms the linear dependance of σ_l on $\langle l \rangle$.

3.3 Estimation of Average Link Length

In the previous section, we have developed the expressions for the parameters of the GEV distribution. Clearly, the proposed model for link length distribution depends significantly on the average link length. In the previous section, it is shown that the expressions (3.5), (3.6) and (3.10), need appropriate estimates of the average link length. Expression (3.1) needs the total information of the link lengths, and thus is not suitable for the estimations, where the complete information is not available. Instead, expression (3.13), first proposed for OTNs in [1], is preferred in these cases for quick estimation.

$$\langle l_x \rangle \cong \frac{\sqrt{A_x}}{\sqrt{N} - 1} \tag{3.13}$$

In expression (3.13), $\langle l_x \rangle$ is the estimated average link length of the network whose area is A_x and number of nodes N. This expression, depends only on the territorial coverage area of the network, and the number of nodes. The subscript x represents the kind of area. In this section, we use three different areas, viz. exact area (x=e), convex area (x=c) and geographical area (x=g) of the networks. The 'area of a network' normally indicates its area of coverage, which may be quite complex to estimate. There are many networks, which do not completely cover the whole geography of a country or province. In that case, taking the whole geographical area of the country or province is not logically correct. Furthermore, there are networks, which cover more than one country. In such cases, the sum of the areas of all the concerned countries can give big errors. Additionally, in the planning stages of a network, it is difficult to predict the boundary links since there is not enough information. This makes the estimation of exact area quite difficult.

Overall, all these complex issues point to the use of some kind of area, which does not need the link information, rather only the node locations. To deal with this constraint, we introduced the convex area of the networks. Issues related to different area definitions are explained in the following subsections.

3.3.1 Different Areas of OTNs

In [1], the area is not defined rigorously where expression (3.13) was proposed for the OTNs. The areas of networks can be of different types depending on the contexts in which it is measured. It may be the exact area (i.e., area confined by the boundary links), or the convex area (i.e., area which can be found from the nodes alone), or the geographical area (i.e., the area of the country or province or countries or provinces where the network operates). These areas are described briefly in this subsection. Exact and convex areas are illustrated with examples in Figure 3.6.

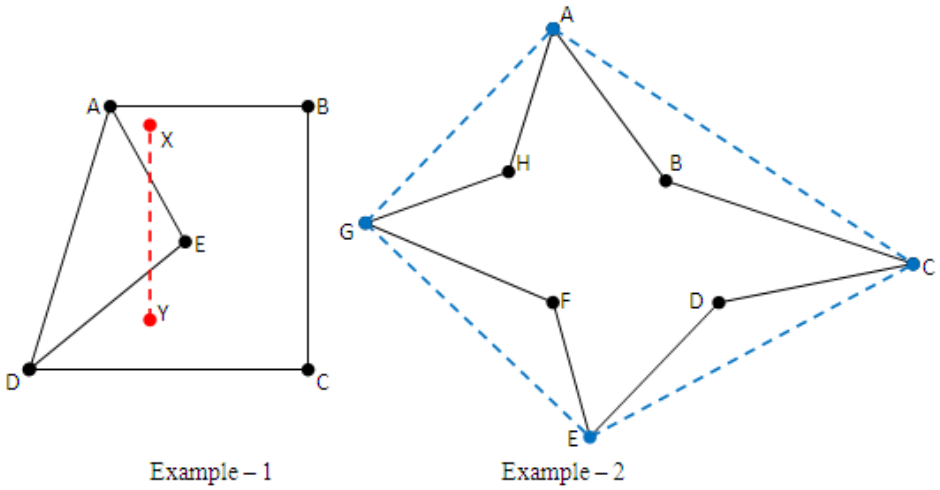

Figure 3.6: Depiction of convex and exact areas: ABCDA and ACEGA are convex areas of ABCDEA and ABCDEFGHA respectively.

3.3.1.1 Convex Area

Convex area is the smallest area that includes all the nodes and line segments between any two points of that area. This area (denoted as A_c) is formed by convexing the boundary of the network (it is a convex set). That means the line segment joining any two interior points of convex area always remains inside that area. It is also true that, when the sides of a non-convex polygon is elongated in both directions, at least one of them passes through the interior of the polygon. In case of convex polygon, this does not happen. The concepts of convexity and concavity are explained in Figure 3.6. The line segment XY in Example – 1, has some points outside the polygon ABCDEA. So the corresponding area is not convex; but ABCDA defines a convex area. Similarly, in Example – 2, the convex area corresponding to polygon ABCDEFGHA is the quadrilateral ACEGA. In other words, ABCDEFGHA defines the exact area in Example – 2; while ACEGA defines the convex area. We use this area (i.e., A_c) to estimate the average link length, $\langle l_c \rangle$.

3.3.1.2 Exact Area

This is the area confined by the boundary links of the network (denoted as A_e). In order to calculate this area, the knowledge of the boundary links is essential. For instance, the polygon ABCDEFGH in Example – 2 of Figure 3.6 defines the exact area. This area is always smaller than or equal to the convex area ($A_e \leq A_c$). When the inside of a network is convex, the exact area becomes the convex area. We use this area (i.e., A_e) to estimate the average link length, $\langle l_e \rangle$.

3.3.1.3 Geographical Area

This is the area mentioned in the geographical maps (denoted as A_g). In fact, it is the total land area or the territorial area of a certain country or province or countries where the network operates. We use this area to estimate the average link length, $\langle l_g \rangle$. We collected these values from the World Atlas [25], and did some necessary changes where appropriate. For example, we use the area of mainland USA for American networks (which excludes the areas of Alaska and islands such as Hawaii). For OMNICOM, and other such European networks (having nodes in Russia), we use the European part of the Russian area, instead of the whole area of Russia. Similar appropriate steps are adopted for other such cases as well.

3.3.1.4 Measurement of Different Areas

We measure the three areas presented in this section using appropriate tools. Exact and convex areas, (A_e and A_c) used for the estimations are measured using Google Earth Professional (version 6.2) which are shown in Table 3.3, columns 'A_e' and 'A_c'. Most of the geographical areas, (A_g) are calculated using the data from the World Atlas [25] which are shown in Table 3.3, column 'A_g'. Using the expressions for statistical distribution, the link length bound probability, $Pr(l_i \leq \sqrt{A_c}) \geq 0.8$, is found for all the 40 networks presented in Table 3.3. It indicates, at least 80% links of the 40 OTNs are smaller than the square root of their convex areas. Clearly, from Table 3.3, it is observed that there can be big differences between the three areas presented.

3.3.2 Results from Different Areas

The average link lengths estimated using expression (3.13) are approximate and deviate from the exact values. We determine these deviations in terms of percentage of errors (i.e., E_c, E_e and E_g) in Table 3.4. These errors are estimated according to expression (3.14).

Table 3.3: Different Areas of the 40 Real OTNs (all areas are in square km)

#	Network	A_c	A_e	A_g
1	VIA Network[4]	1267876	824464	1915456
2	BREN[4]	42196	33780	110883
3	RNP[4]	1814676	1077222	8514965
4	Abilene Core[15]	5698507	3728429	8080464
5	LEARN[13]	183296	148137	696241
6	CompuServe[14]	5442148	4661688	8080464
7	vBNS[4]	6279993	4182861	8080464
8	CESNET[4]	45957	32357	78909
9	NSFNET[4]	6007605	5005030	8080464
10	ITALY[4]	436577	291715	301230
11	ACONET[4]	52826	39297	83858
12	MZIMA[4]	6805153	4278891	8080464
13	GARR-B[4]	476100	242064	301230
14	ARNES[4]	12645	7112	20273
15	GERMANY[4]	196675	157319	357021
16	REDIRIS[4]	463539	367094	504882
17	LambadaRail[4]	6899337	4393019	8080464
18	MEMOREX[4]	253850	141524	981834
19	CANARIE[4]	3462688	2095793	9985140
20	EON[4]	5800056	2524156	6550943
21	ARPANET[4]	5975012	5640345	8080464
22	OPTOSunet[17]	122980	81063	449964
23	Hibernia USA[18]	1257762	531658	2697216
24	PIONIER[4]	194178	163888	312685
25	COX [4]	4870065	3531799	8080464
26	SANET[4]	24336	18496	48845
27	NEWNET [4]	7033452	4200490	8080464
28	PORTUGAL[4]	888824	612952	92391
29	RENATER[4]	462096	320900	642346
30	IBN31[19]	426658	200382	300448
31	BULGARIA[20]	95108	73138	113477
32	GEANT2[4]	7703318	3383604	770326
33	LONI [4]	68008	52069	128578
34	METRONA[4]	158837	75727	244820
35	COST37[21]	5371917	3329573	6228201
36	CERNET[22]	6719547	2628942	9644552
37	OMNICOM[4]	2596279	2091020	2646447
38	INTERNET2[4]	7169590	5974161	8080464
39	CORONET[23]	7903482	6046394	8080464
40	USA100[1]	8191044	5992704	8080464

$$E_x = \frac{\langle l \rangle - \langle l_x \rangle}{\langle l \rangle} \tag{3.14}$$

We present the effectiveness of expression (3.13), using exact area, geographical area and convex area of the 40 OTNs in Table 3.4. It can be observed that there are both positive and negative errors. So average error does not provide a representative evaluation criterion. However, average absolute error (average of absolute error values) is a good parameter to present the overall effectiveness. The values for the average and absolute averages for all 40 OTNs are shown in the last two rows of Table 3.4. From this data, it is clear that the average absolute error in the average link length is higher in the cases of exact and geographical areas (see the last row under columns '$E_c(\%)$', '$E_e(\%)$' and '$E_g(\%)$' in Table 3.4).

In terms of the individual network errors, convex area provides better results in 33 cases, out of total 40 in comparison to the exact area (columns '$E_c(\%)$', '$E_e(\%)$' and '$E_g(\%)$' in Table 3.4). Exact area gives better results only in case of 7 networks, and

Table 3.4: Estimation of Average Link Lengths from Different Areas and Their Comparisons ($\langle l_c \rangle$, $\langle l_e \rangle$, $\langle l_g \rangle$ and $\langle l \rangle$ are in km)

#	Network	$\langle l_c \rangle$	$\langle l_e \rangle$	$\langle l_g \rangle$	$\langle l \rangle$	$E_c(\%)$	$E_e(\%)$	$E_g(\%)$	Diff.1	Diff.2	Best Area
1	VIA Network[4]	563	454	692	571	1.40	20.49	-21.19	19.09	19.79	A_c
2	BREN[4]	95	85	154	94	-1.06	9.57	-63.83	7.86	62.77	A_c
3	RNP[4]	623	480	1350	748	16.71	35.83	-80.48	19.12	63.77	A_c
4	Abilene Core[15]	1104	893	1315	1067	-3.47	16.31	-23.24	12.84	19.77	A_c
5	LEARN[13]	198	178	386	189	-4.76	5.82	-104.23	1.06	99.47	A_c
6	CompuServe[14]	1007	932	1227	1161	13.26	19.72	-5.68	6.46	-7.58	A_g
7	vBNS[4]	1017	830	1154	965	-5.39	13.99	-19.59	8.60	14.20	A_c
8	CESNET[4]	87	73	114	91	4.40	19.78	-25.27	15.38	20.87	A_c
9	NSFNET[4]	894	816	1037	1086	17.68	24.86	4.51	7.18	-13.17	A_g
10	ITALY[4]	241	197	200	280	13.93	29.64	28.57	15.71	14.64	A_c
11	ACONET[4]	80	69	101	119	32.77	42.02	15.13	9.25	-17.64	A_g
12	MZIMA[4]	908	720	989	852	-6.57	15.49	-16.08	8.92	9.51	A_c
13	GARR-B[4]	230	164	183	224	-2.68	26.79	18.30	24.11	15.62	A_c
14	ARNES[4]	36	27	46	38	5.26	28.95	-21.05	23.69	15.79	A_c
15	GERMANY[4]	142	127	191	143	0.70	11.19	-33.57	10.49	32.87	A_c
16	REDIRIS[4]	218	194	228	319	31.66	39.18	28.53	7.52	-3.13	A_g
17	LambadaRail[4]	782	624	846	671	-16.54	7.00	-26.08	-9.54	9.54	A_e
18	MEMOREX[4]	150	112	295	137	-9.49	18.25	-115.33	8.76	105.84	A_c
19	CANARIE[4]	554	431	941	668	17.07	35.48	-40.87	18.41	23.80	A_c
20	EON[4]	717	473	762	754	4.91	37.27	-1.06	32.36	-3.85	A_g
21	ARPANET[4]	704	684	819	839	16.09	18.47	2.38	2.38	-13.71	A_g
22	OPTOSunet[17]	101	82	193	100	-1.00	18.00	-93.00	17.00	92.00	A_c
23	Hibernia USA[18]	323	210	473	279	-15.77	24.73	-69.53	8.96	53.76	A_c
24	PIONIER[4]	123	113	156	131	6.11	13.74	-19.08	7.63	12.97	A_c
25	COX[4]	566	482	729	662	14.50	27.19	-10.12	12.69	-4.38	A_g
26	SANET[4]	39	34	55	36	-8.33	5.56	-52.78	-2.77	44.45	A_e
27	NEWNET[4]	647	500	693	528	-22.54	5.30	-31.25	-17.24	8.71	A_e
28	PORTUGAL[4]	230	191	74	203	-13.30	5.91	63.55	-7.39	50.25	A_e
29	RENATER[4]	162	135	191	155	-4.52	12.90	-23.23	8.38	18.71	A_c
30	IBN31[19]	143	98	120	131	-9.16	25.19	8.40	16.03	-0.76	A_g
31	BULGARIA[20]	65	57	71	51	-27.45	-11.76	-39.22	-15.69	11.77	A_e
32	GEANT2[4]	596	395	596	661	9.83	40.24	9.83	30.41	0.00	$A_g = A_c$
33	LONI[4]	56	49	77	62	9.68	20.97	-24.19	11.29	14.51	A_c
34	METRONA[4]	84	58	104	73	-15.07	20.55	-42.47	5.48	27.40	A_c
35	COST37[21]	456	359	491	439	-3.87	18.22	-11.85	14.35	7.98	A_c
36	CERNET[22]	510	319	611	636	19.81	49.84	3.93	30.03	-15.83	A_g
37	OMNICOM[4]	312	280	315	298	-4.70	6.04	-5.70	1.34	1.00	A_c
38	INTERNET2[4]	413	377	438	334	-23.65	-12.87	-31.14	-10.78	7.49	A_e
39	CORONET[23]	367	321	371	326	-12.58	1.53	-13.80	-11.05	1.22	A_e
40	USA100[1]	318	272	316	310	-2.58	12.26	-1.94	9.68	-0.64	A_g
Average:		396.53	322.38	477.60	410.78	0.53	18.99	-22.09	8.97	19.99	A_c
Absolute Average:		396.53	322.38	477.60	410.78	11.26	20.22	31.25	12.69	24.03	A_c

out of these 7 networks, some have specific irregularities with respect to the node and link distributions. For example, networks such as PORTUGAL, LambdaRail, NEWNET and SANET, have a few very long links, and the rest of the links are quite shorter than these long links. The other 3 of these 7 networks, such as BULGARIA, INTERNET2 and CORONET have quite large voids in their convex areas that leads to increased errors. For some ring-like networks, (such as INTERNET2, BULGARIA, LambdaRail, SANET, LONI and NEWNET) better results are obtained with exact areas.

The expression for average link length in [1] uses the area of coverage, which is very often the area of the country (used to estimate E_g and $\langle l_g \rangle$ in Table 3.4) or province or the sum of the areas of the countries, which are served by the network. However, this area is not very effective as can be seen from the column '$E_g(\%)$' in Table 3.4.

We present the errors (in $\%$, i.e., $E_c(\%)$ and $E_e(\%)$) in histograms, and find them to follow approximate normal distributions. In Figures 3.7 and 3.8, we present these error

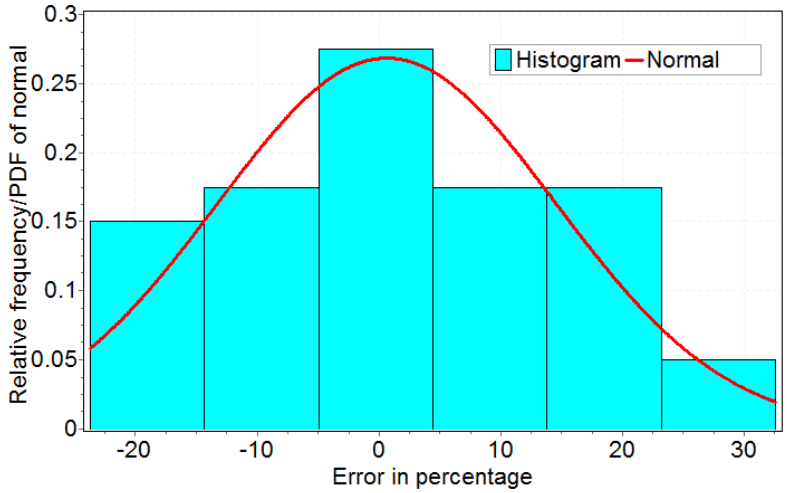

Figure 3.7: Distribution of estimated errors with convex area ($E_c(\%)$) for 40 OTNs (fitted to normal distribution [μ=0.53, σ=14.15])

distributions for 40 OTNs from which, it is quite clear that $E_c(\%)$ is symmetrical about 0.53%, whereas $E_e(\%)$ is shifted by 18.46% to the right with respect to the former.

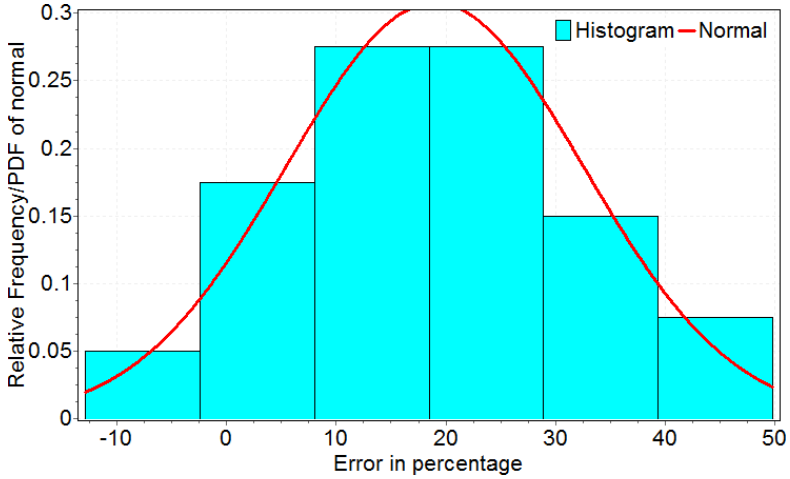

Figure 3.8: Distribution of estimated errors with exact area ($E_e(\%)$) for 40 OTNs (fitted to normal distribution [μ=18.99, σ=13.67])

We summarize the results of the evaluation using different areas in the last three columns in Table 3.4. Columns under headings 'Diff.1' and 'Diff.2' present the differences between the errors in average link length estimations, when using different areas. Diff.1 is the difference between the absolute values of the errors from exact and convex areas (i.e., Diff.1 = $|E_e(\%)|$ - $|E_c(\%)|$). Diff.2 is the difference between the absolute values of the errors from geographical and convex areas (i.e., Diff.2 = $|E_g(\%)|$ - $|E_c(\%)|$). When Diff.1 is positive, it shows that A_c is better than A_e and vice versa. Similarly, when Diff.2 is

positive, A_c is better than A_g and vice versa. In the last column, we present the best area in terms of least percentage of absolute error. For 22 OTNs convex area, A_c is the best, for 11 OTNs country area, A_g is the best, and for the rest 7 OTNs, exact area, A_e is the best. When considering the average absolute error over all 40 OTNs, the convex area provides the best estimate (last row, under columns 'Diff.1' and 'Diff.2' in Table 3.4).

We have also considered the effect of ellipticity (i.e., eccentricity of the circumferential ellipse which is presented in detail in Chapter 4) of the networks on their average link lengths, in this analysis. OTNs with higher eccentricity, typically greater than 0.7, have higher error percentages with all types of areas (such as RNP, CANARIE and BULGARIA). However, ellipticity alone is not a very dominant factor.

3.3.3 Proposed Changes for Accuracy

The formula (3.13), is an approximate expression. It assumes that the average link length of a topology is proportional to the square root of the area and inversely proportional with the square root of the number of nodes. In fact, this expression is correct for a square planar topology, with the sides of the square as its links. For other kind of topologies (even the square topology with a diagonal as a link), it gives error, whose amount depends on the shape and other attributes of the network topology. Thus expression (3.13) needs to be changed for better accuracy.

Using formula (3.13), we get an average absolute error of 20.22% when the exact area is used (last row, column '$E_e(\%)$', in Table 3.4). However, using a multiplying factor, k_e = 1.22, this average error can be reduced to 12.65%. But, the individual estimates of some of the networks (those with negative errors) get worse. With convex area, we have average absolute error of 11.26% (last row, column '$E_c(\%)$', in Table 3.4). In this case too, if we provide a multiplying factor, the error reduces slightly to a minimum of 10.90% with k_c = 0.97. These modified expressions are presented in (3.15) and (3.16). The effects of multiplying factors on the exact and convex areas are presented in Figures 3.9 and 3.10, respectively. The lowest points of the curves shown in Figures 3.9 and 3.10 are the best multiplying factors for the corresponding areas.

$$\langle l_e^{'} \rangle \cong \frac{k_e \sqrt{A_e}}{\sqrt{N} - 1} \tag{3.15}$$

With k_e=1.22, we get 12.65% average absolute error with exact area, A_e.

$$\langle l_c^{'} \rangle \cong \frac{k_c \sqrt{A_c}}{\sqrt{N} - 1} \tag{3.16}$$

With k_c=0.97, we get 10.90% average absolute error with convex area A_c. Using nonlinear regression, we get some better expressions, which can reduce the error to around 9%.

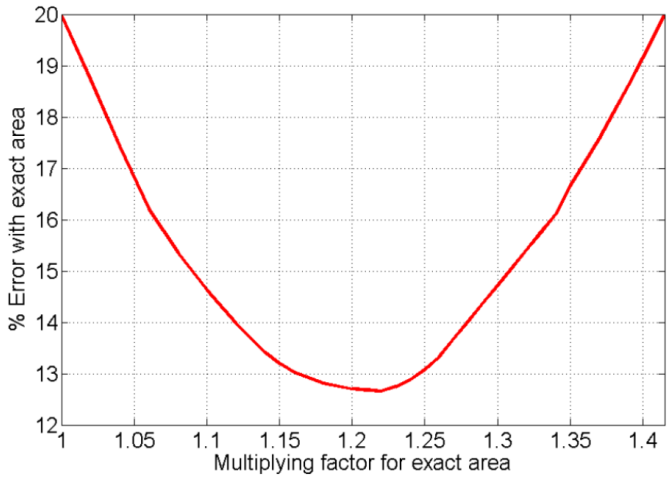

Figure 3.9: Effect of multiplying factors, k_e when using the exact area (k_e=1.22 gives minimum error).

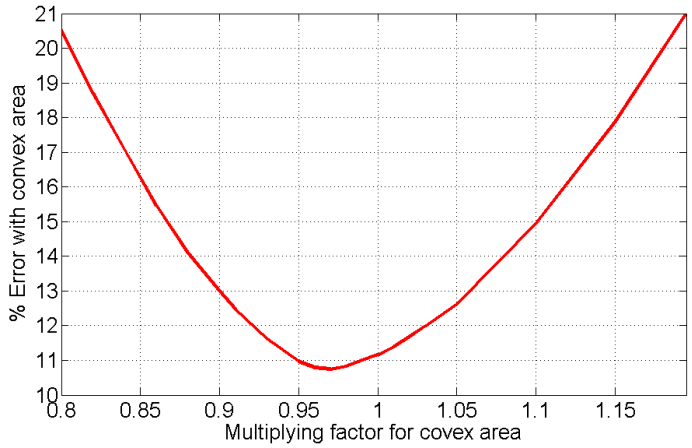

Figure 3.10: Effect of multiplying factors, k_c when using the convex area (k_c = 0.97 gives minimum error).

But these expressions are either too complex (polynomials of the order of 7 and higher) or need extra information such as the shape of the exact area. Therefore, use of convex area for the better approximation of average link length as shown in (3.16) is preferred.

The results obtained for average link length estimations using the best fitting multiplying factors (k_e=1.22 and k_c=0.97) are presented in Table 3.5. Results under column 'Diff.' are the differences between the two absolute errors using multiplying factors k_e=1.22 and k_c=0.97 with exact and convex areas respectively (i.e., Diff.=$|E_e^{'}(\%)|$ - $|E_c^{'}(\%)|$). A positive difference indicates that convex area gives better output than exact area and vice versa. As can be seen from Table 3.5 (row 'Absolute Average', column 'Diff.'), the convex area provides an improvement of 7.27% in overall absolute value,

compared with the exact area. Using approximation (3.16), link length distribution parameters α_c', β_c', ξ_c' and the standard deviation, σ_{lc}' of the link lengths can be estimated.

3.4 Development of the Model from $\langle l_c' \rangle$

We analyze all the relevant issues for the development of the link length statistical model based on its PDF in the previous sections. The parameters of the corresponding GEV distribution can be estimated using the modified expressions for the average link length developed in the last section with reasonable accuracy. From the results obtained in the last section, we are now confident that for obtaining the GEV distribution for the link lengths, just the node locations are enough. Following this approach, we start by estimating the average link length, $\langle l_c' \rangle$, using equation (3.16). Then, we use $\langle l_c' \rangle$ in the place of $\langle l \rangle$ in expressions (3.5) and (3.6) to estimate the corresponding location factor, α_c', and scale factor, β_c'. We use the approximate value of ξ_c', which is equal to 0.167. Eventually, we obtain the PDF of the GEV distribution for the link lengths as defined by equation (3.4).

We evaluate the developed model using the KSS values. In Table 3.5, the KSS values obtained from the estimated CDFs are presented as KSS_{GE}. The average value of KSS_{GE} for 40 OTNs is 0.2076 (row 'Average', column 'KSS_{GE}' in Table 3.5). It is considered as a good value, since the corresponding distribution parameters were obtained without the knowledge of the networks details, rather just the nodes' locations.

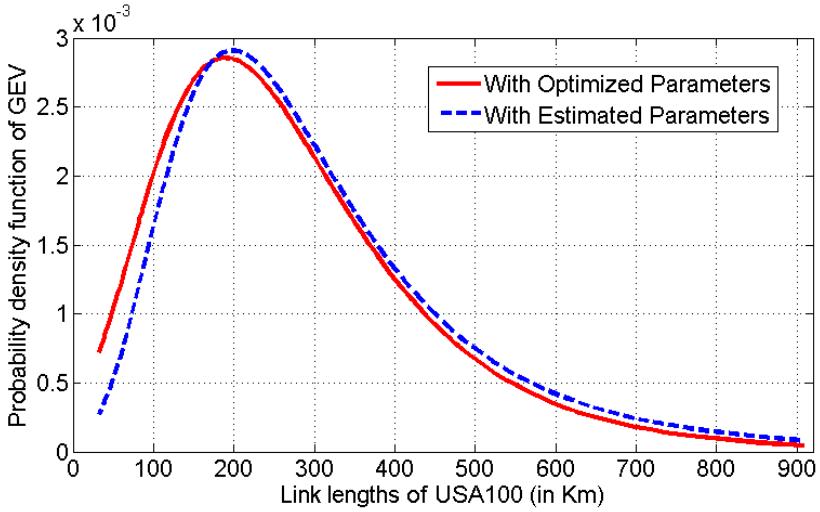

Figure 3.11: Exact vs. estimated GEV distribution of link lengths of USA100 [1] network (171 links). This estimation uses estimated average link length, $\langle l_c' \rangle$.

In addition to this, it can be observed from Table 3.5 that 34 networks provide 'acceptable' model at CI = 0.95. A model is 'acceptable,' when its KSS_{GE} is smaller than

Table 3.5: Comparison of Errors with Optimized Multiplying Factors in Expressions (3.15) and (3.16) (columns '$E_c^{'}(\%)$', '$E_e^{'}(\%)$', 'Diff.' and 'Best Area'). Estimation of the Parameters of the Proposed Model from the Average Link Length, (columns '$\langle l_c^{'} \rangle$', '$\alpha_c^{'}$', '$\beta_c^{'}$', '$\xi_c^{'}$'), the KSS Values of the networks, (column 'KSS_{GE}'), and Their Evaluation (column 'Acceptable?')

#	Network	$E_c^{'}(\%)$ (k_c=0.97)	$E_e^{'}(\%)$ (k_e=1.22)	Diff.	Best Area	$\langle l_c^{'} \rangle$	$\alpha_c^{'}$	$\beta_c^{'}$	$\xi_c^{'}$	KSS_{GE}	Acceptable? (CI = 0.95)
1	VIA Network[4]	4.38	2.98	-1.40	A_e	546	368	228	0.167	0.2692	Yes
2	BREN[4]	2.13	-10.64	8.51	A_c	92	69	28	0.167	0.3084	Yes
3	RNP[4]	19.25	-21.66	2.41	A_c	604	406	254	0.167	0.2231	Yes
4	Abilene Core[15]	- 0.37	-2.06	1.69	A_c	1071	713	460	0.167	0.1351	Yes
5	LEARN[13]	-1.59	-14.81	13.22	A_c	192	135	72	0.167	0.1243	Yes
6	CompuServe[14]	15.85	2.07	-13.78	A_e	977	651	418	0.167	0.1872	Yes
7	vBNS[4]	-2.18	-4.97	2.79	A_c	986	657	422	0.167	0.1438	Yes
8	CESNET[4]	7.69	2.20	-5.49	A_e	84	64	25	0.167	0.3254	No
9	NSFNET[4]	20.17	8.29	-11.88	A_e	867	579	370	0.167	0.1849	Yes
10	ITALY[4]	16.43	-14.29	-2.14	A_e	234	163	91	0.167	0.1951	Yes
11	ACONET[4]	34.45	-29.41	-5.04	A_e	78	60	22	0.167	0.4421	No
12	MZIMA[4]	-3.40	-3.05	-0.35	A_e	881	588	376	0.167	0.1468	Yes
13	GARR-B[4]	0.45	10.71	10.26	A_c	223	155	86	0.167	0.1551	Yes
14	ARNES[4]	7.89	-13.16	5.27	A_c	35	32	3	0.167	0.5283	No
15	GERMANY[4]	3.50	-8.39	4.89	A_c	138	99	48	0.167	0.1731	Yes
16	REDIRIS[4]	33.86	25.71	-8.15	A_e	211	147	81	0.167	0.3782	No
17	LambadaRail[4]	-13.11	-13.41	0.30	A_c	759	508	322	0.167	0.1235	Yes
18	MEMOREX[4]	-6.57	0.00	-6.57	A_e	146	105	52	0.167	0.1203	Yes
19	CANARIE[4]	19.61	21.26	1.65	A_c	537	362	224	0.167	0.1170	Yes
20	EON[4]	7.82	23.47	15.65	A_c	695	466	294	0.167	0.1223	Yes
21	ARPANET[4]	18.59	0.60	-17.99	A_e	683	458	289	0.167	0.2062	Yes
22	OPTOSunet[17]	2.00	0.00	-2.00	A_e	98	73	31	0.167	0.2619	Yes
23	Hibernia USA[18]	-12.19	8.24	-3.95	A_e	313	215	126	0.167	0.1756	Yes
24	PIONIER[4]	9.16	-5.34	-3.82	A_e	119	87	40	0.167	0.2047	Yes
25	COX[4]	17.07	11.18	-5.89	A_e	549	370	230	0.167	0.0956	Yes
26	SANET[4]	-5.56	-13.89	8.33	A_c	38	34	4	0.167	0.3885	No
27	NEWNET[4]	-18.94	-15.53	-3.41	A_e	628	422	265	0.167	0.1178	Yes
28	PORTUGAL[4]	-9.85	-14.78	4.93	A_c	223	155	86	0.167	0.6653	No
29	RENATER[4]	-1.29	-6.45	5.16	A_c	157	112	57	0.167	0.1488	Yes
30	IBN31[19]	-6.11	8.40	2.29	A_c	139	100	49	0.167	0.1067	Yes
31	BULGARIA[20]	-23.53	-37.25	13.72	A_c	62	49	15	0.167	0.2929	No
32	GEANT2[4]	12.56	27.08	14.52	A_c	578	389	243	0.167	0.1053	Yes
33	LONI[4]	12.90	3.23	-9.67	A_e	54	44	11	0.167	0.0902	Yes
34	METRONA[4]	-10.96	2.74	-8.22	A_e	81	62	23	0.167	0.2278	Yes
35	COST37[21]	-0.68	0.23	-0.45	A_e	442	299	183	0.167	0.1447	Yes
36	CERNET[22]	22.17	38.84	16.67	A_c	495	334	206	0.167	0.1726	Yes
37	OMNICOM[4]	-1.68	-14.77	13.09	A_c	303	208	122	0.167	0.1411	Yes
38	INTERNET2[4]	-20.06	-37.72	17.66	A_c	401	272	164	0.167	0.1615	Yes
39	CORONET[23]	-9.20	-20.25	11.05	A_c	356	243	145	0.167	0.1221	Yes
40	USA100[1]	0.65	-7.10	6.45	A_c	308	212	123	0.167	0.0707	Yes
Average:		-3.53	1.13	1.76	A_c	384.70	–	–	–	0.2076	Yes
Absolute Average:		10.90	12.65	7.27	A_c	384.70	–	–	–	–	Yes
Standard Deviation:		13.72	16.70	8.86	–	303.25	–	–	–	–	–

the critical value of the corresponding KS statistic at a specific CI (here, it is 0.95). Only for 6 OTNs, the KSS_{GE} exceeds the critical value (at, CI = 0.95). However, one of the 6 OTNs (i.e., CESNET), passes the KS test with CI = 0.99, and becomes 'acceptable'. The networks which do not pass the KS test in Table 3.5, in fact, have some kind of irregularity, as explained in subsection 3.3.2 for the 'high percentage of error in the average link lengths of seven networks'. A few other networks in which, there are comparatively large voids in their convex areas and small number of links, too give large values of KSS_{GE} (such as VIA Network and BREN). However, if the exceptionally long links are excluded (normally, they are around 10% of the total number of links in the 6 OTNs that did not give 'acceptable' KSS values); the rest of the samples provide small KSS values,

and pass the KS test. In Figure 3.11, we present the estimated PDF, for the USA100 network (in blue dotted curve). For comparison, its optimized PDF is also shown (solid red curve). The latter represents the best-fitting link length distribution for the network, obtained using the optimized values of the GEV parameters.

3.5 Chapter Summary

In this chapter, we analyzed 40 real OTNs for the development of a statistical model for their link lengths. We also developed an improved expression for the average link length estimation. We perform these estimations without prior knowledge of network configuration details, except for the node locations. Out of the 61 distributions analyzed in this study, the GEV distribution was found to provide the best fitting distribution model for the link length statistics of real OTNs. We found that all the three parameters of the GEV distribution can be estimated from the average link length of the OTNs. For the estimation of average link length, we used the 'convex area' as it provides the best estimate. The multiplying factors obtained for the improvement in the accuracy of the average link lengths are quite impressive with the exact area. For the convex area we get a very small improvement. So, the expression for the average link length using the convex area can be used without any multiplying factor.

The validity of the developed statistical models have been checked using the KSS values. The estimated PDF model for the link lengths has good accuracy (in terms of their KSS values) for large networks (having large number of links). For smaller networks, or for networks with some specificities (e.g., just a few links with a much longer span compared to the majority of links), the KSS values obtained were larger than the other networks. Overall, the KSS value of 0.2076, obtained by averaging over all 40 networks, is good, considering that the estimations need only the knowledge of node locations.

References

[1] S. K. Korotky, "Network Global Expectation Model: A Statistical formalism for Quickly Quantifying Network Needs and Costs," *IEEE/OSA Journal of Lightwave Technology*, vol. 22, no. 3, pp. 703 – 722, Mar. 2004.

[2] J. F. Labourdette, E. Bouillet, R. Ramamurthy, and A. A. Akyamaç, "Fast Approximate Dimensioning and Performance Analysis of Mesh Optical Networks," *IEEE/ACM Transactions on Networking*, vol. 13, no. 4, pp. 906 – 917, Aug. 2005.

[3] J. P. Cardenas, A. Santiago, J. C. Losada, R. M. Benito and M. L. Mouronte, "On the Topology of Optical Transport Networks," *Journal of Physics: Conference Series 246*, pp. 1 – 7, 2010.

[4] C. Pavan, R. M. Morais, J. R. F. da Rocha and A. N. Pinto, "Generating Realistic Optical Transport Network Topologies," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 2, no. 1, pp. 80 – 90, Jan. 2010.

[5] R. M. Morais, C. Pavan, C. Requejo and A. N. Pinto, "Genetic Algorithm for the Topological Design of Survivable Optical Transport Networks," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 2, no. 1, pp. 80 – 90, Jan. 2010.

[6] S. K. Korotky, R. Essiambre, and R. W. Tkach, "Expectations of Optical Network Traffic Gain Afforded by Bit Rate Adaptive Transmission," *Bell Labs Technical Journal,* vol. 14, no. 4, pp. 285 – 296, Feb. 2010.

[7] S. K. Korotky and K. N. Oikonomou, "Scaling of Most-Likely Traffic Patterns of Hose- and Cost-Constrained Ring and Mesh Networks," in Proceedings of International Conference on Optical Fiber Communications (OFC), pp. 1 – 3, Mar. 2006.

[8] A. Dwivedi and R. E. Wagner, "Traffic model for USA long-distance optical network," in Proceedings of International Conference on Optical Fiber Communications (OFC), pp.156 – 158 2000.

[9] A. N. Pinto, C. Pavan, and R. M. Morais, "Dimensioning optical networks: A practical approach," in Proceedings of 12th International Conference on Transparent Optical Networks (ICTON), pp. 1 – 4, Jan. 2010.

[10] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, "Small but Slow World: How Network Topology and Burstiness Slow Down Spreading," *Phys. Rev. E*, vol. 83, no. 2, pp. 025102: 1 – 4, Feb. 2011.

[11] S. K. Routray, R. M. Morais, J. R. F. da Rocha, and A. N. Pinto, "Statistical Model for Link Lengths in Optical Transport Networks," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 5, no. 7, pp. 762 – 773, Jul. 2013.

[12] Reference Optical Networks, Apr. 2013 [Online]. Available: http://www.av.it.pt/anp/on/refnet2.html.

[13] The Lonestar Education And Research Network (LEARN). [Online]. Available: http://www.tx-learn.org/images/map-large.jpg.

[14] The CompuServe Backbone Optical Transport Network. [Online]. Available: http://www.nthelp.com/images/compuserve.jpg.

[15] The Abilene Backbone Optical Transport Network. [Online]. Available: http://www.internet2.edu/2004AR/abilene_map_large.cfm.

[16] GARR-B, The Italian Research and Education Network (NREN). [Online]. Available: http://www.garr.it/index.php/rete/garr-x#mappa.

[17] Swedish University Computer Network, OptoSUNET. [Online]. Available: http://www.sunet.se/download/18.6d7c8917128274d3dd0800 06061/opto-sunetbroschyr_eng.pdf.

[18] The High Speed Hibernia Atlantic USA Networks. [Online]. Available: http://www.hiberniaatlantic.com/US_network.html.

[19] The 31 Node Italian Backbone Optical Transport Network, IBN31. [Online]. Available: http://www.opticsinfobase.org/abstract.cfm?URI= OFC–2012-JTh2A.37.

[20] Bulgarian Backbone Optical Transport Network. [Online]. Available: http://www.gcn.bg/bg_backbone.html.

[21] European Core Optical Transport Network (COST). [Online]. Available: http://www.cost37.bg/bg_backbone.htm.

[22] The Chinese Education and Research Network (CERNET). [Online]. Available: http://www.edu.cn/20060111/3170194.shtml.

[23] The Next Generation Core Optical Networks (CORONET). [Online]. Available: http://monarchna.com/topology.html.

[24] S. Kotz, S. Nadarajah, *Extreme Value Distributions: Theory and Applications,* London, England: Imperial College Press, 2000.

[25] The World Atlas. [Online]. Available: http://www.worldatlas.com/.

CHAPTER 4

Estimation of Convex Area of Optical Transport Network

4.1 Introduction

In CHAPTER 3, we have come accross cases in which the coverage areas of the optical transport networks (OTNs) are needed for parameter estimations. During the planning stages, and also for the fast estimation of the link related parameters, the coverage areas of OTNs are used in majority of the cases, when the main parameters are not known. However, the coverage area of a core network is quite complex to determine. In order to solve this problem, we provide a method for its estimation, in which, we use the area of circumferential ellipse (CE) of the OTN.

4.1.1 Motivation

Coverage area of core networks cannot be determined from the geographical or country areas [1]. These areas give large errors, and we have come across it in Chapter 3. Convex area (the smallest area that includes all the nodes and line segments between any two points of that area) can be used as coverage area. In the previous chapter, we have seen the importance of convex area as the coverage area of the OTNs in the estimation of the models for link lengths. As presented in Chapter 3, this parameter is essential in the estimation of the average link length, which is used in the estimation of other related parameters of the OTNs. The accuracy of the estimations using this area

has been shown in Chapter 3. However, estimation of convex area from a planar map is not trivial. In this chapter, we show an efficient method to estimate the convex area from a planar map.

4.1.2 Related Work

Depending on the context, where it is used and how it is defined, coverage areas can be of different types. Three types of areas for the OTNs have been defined in [1], viz. exact area (that covers the area inside the topology formed by the boundary links), convex area (that is the largest convex set formed by the nodes of the topology), and the geographical area (as defined by the geography of the region in which the network provides the services). Convex area is found to be the most suitable out of these three, for the link related estimations. Using this area, average link lengths of an OTNs can be estimated with quite good accuracies [1]. The distribution model of the link lengths can be estimated in terms of probability density function [1], using the average link length. Several link-related parameters can be estimated from the link length distributions [2]. Approximate area of a graph can be estimated using the geometric measure theory [3]. However, in that method, a large number of inputs are required for the complex calculations, and yet, the outcome is not exact [3],[4]. Furthermore, the input parameters can be found out, only if the whole topology of the graph is known [4]. These methods are not very helpful in the practical estimations for graphs like the OTNs, in which the exact areas do not provide good estimations [1]. Coverage areas are used to estimate the average link lengths of OTNs [1], [5]. In [5], [6] fast estimation of several OTN parameters are performed without complete information. Eccentricity of network ellipse (also known as ellipticity) is researched for the study of several properties associated with the networks, and graphs [7] – [14]. In [10] – [12], several approaches are followed to determine the diameter, radius and eccentricities of the networks including the distance between the node pairs. Fast estimation of these parameters is required in several fields such as networking, distributed computing and graph theory [15], [16]. Such fast estimation methods have been emphasized in different fields [15] – [18]. Shortest paths between node pairs, and diameters of networks are estimated in [17], [18]. These methods are helpful in the estimation of the parameters of the CEs. Fast calculation of CAPEX is required in several estimations of the OTNs, especially during the planning phase [5]. Methods for fast calculation of CAPEX in optical multilayer networks are presented in [5], [19].

4.1.3 Chapter Outline

In this chapter, we present several theoretical aspects of CE, and demonstrate their utilities for OTNs. We apply the CE concepts to the OTNs and estimate several parameters commonly used for the OTNs with incomplete information. We show that the convex area and average link length can be estimated using the circumferential ellipses with quite good accuracies. The rest parts of this chapter are organized in 3 sections. In section 4.2, we define, illustrate and explain the concepts of CE. Then we show some common properties of CE, which are required for the applications for the OTNs. In section 4.3, we apply the concepts of CE to the 40 real OTNs used in [1], and measure their parameters for evaluation. In section 4.4, we conclude with a few remarks on the utility of the CE parameters for the OTN related estimations. The work presented in this chapter is reported in [20].

4.2 Concepts of CE of OTN

For regular structures like squares, or rectangles, or regular polygons, we can always find a circumferential circle (also known as circumcircle). However, for general graphs which are commonly irregular, it is not always possible to get a circumferential circle. Instead we can have CEs. For an OTN, the CE can be defined as the smallest ellipse (in area) that circumscribes or surrounds the whole OTN. So, it just touches three or more nodes of the OTN. The common points between the CE, and the OTN topology are only a finite number of nodal points. If the CE is belittled slightly in any manner, then it must enter into the OTN by intersecting at least a link (at least two links for survival OTNs) of the topology. Therefore, for every network topology, there is a unique CE. In Figure 4.1, we show the CE of BREN network, which touches the topology at four nodal points.

4.2.1 Properties of CE of OTN

Though CE has several properties (some of which are common for the network lying inside it), here we consider only a few important ones needed for the estimation of convex area of OTN (A_C). Every closed network has a unique CE. In contrast, the same CE can surround several different networks (theoretically infinite). Under such circumstances, the convex areas of the networks may be different from one another. In case of the OTNs, the area of their CEs (A_{CE}) and their convex areas have some restrictions as the nodes of the topologies are located mainly at the big cities (i.e., the node locations are not completely random as in a general case). Thus the centroid of the convex area of

an OTN topology is typically very close to the centroid of its CE. The CE contains the whole convex area of the OTN within it (thus, the exact area of the OTN as well). Thus, $A_{CE} \geq A_C$, and geometrically the inside of the convex area is a subset of the area of the CE (i.e., $A_C \subseteq A_{CE}$).

4.2.2 Estimation of CE of OTN

Here, we use planar map to estimate the CE of an OTN. To get the CE, the major axis of the ellipse has to be either approximately oriented along the chord joining the node pairs for which the distance is the longest or approximately parallel to it. If there is more than one longest distance of the OTN (i.e., the longest chords are equal in length), then the major axis has to pass approximately in between those chords. Only the major and minor axes of the CE are to be known. Area of the CE is given by:

$$A_{CE} = \pi ab. \tag{4.1}$$

In (4.1), a (= OA = OB, in Figure 4.1), and b (= OC = OD, in Figure 4.1), are the semi-major and semi-minor axes, respectively (shown as the '- · -' dotted blue lines in Figure 4.1). Every ellipse has its own unique circumferential rectangle. In Figure 4.1, we have shown it in dark red dotted lines (- - -). So, a is half of the length of the rectangle, and b is the half of the breadth of it. Knowledge of a and b are enough to estimate the CE and the other parameters of it, such as the eccentricity and the separation between its foci.

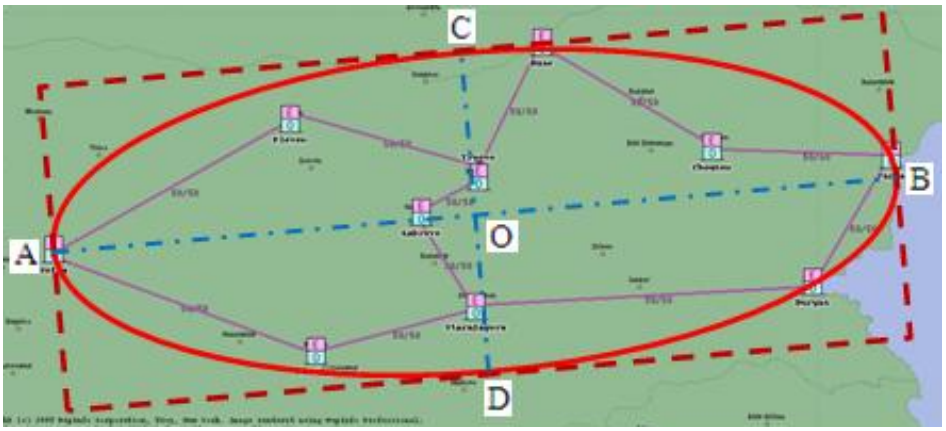

Figure 4.1: CE of BREN. In this case, the CE touches the topology only at 4 points.

A few specific points are to be noted while estimating the CE. The planar map used for the estimation must have the same scale along the length and the width. Instead of locating the nearest place, the exact end points should be used while calculating the length of the major and minor axes. For instance, if one of the end points of the axes of

the CE is located inside a sea or river (in Figure 4.1, B is in the sea and we considered its exact location, not its nearest point on land), that exact point should be used instead of its nearest town or city. Reliable and accurate maps such as the Google Earth provide good accuracies in the overall estimation.

4.3 Estimation of OTN Parameters from CE

The CE is very closely associated with its OTN topology, and bears several common properties. So it can be used for the network parameter estimations such as the convex area and average link length. The statistical model for link lengths in [1] depends on the average link length of the OTN. Average link length of an OTN is a key factor, and it is estimated from the convex area and the number of nodes of the OTN [1]. In this section, we find out the areas of CEs of 40 real OTNs, and evaluate their utilities in estimating the convex areas and average link lengths of the OTNs.

We measured the link lengths of the OTNs using OPNET Transport Planner 15 and cross-checked them using Google Earth Professional (version 6.2). Averaging the link lengths, we calculated the exact average link length (denoted as $\langle l \rangle$ in Table 4.1). We measured the semi-major axes (a), and semi-minor axes (b) of the CEs using Google Earth Professional. The convex areas (column 'A_C' in Table 4.1) of the OTNs were obtained using Google Earth Professional, and the areas of their CEs (column 'A_{CE}' in Table 4.1) are obtained using (4.1). For $\langle l_c \rangle$ (the estimated value of average link length using A_C) and $\langle l_{CE} \rangle$ (a similar estimate like $\langle l_c \rangle$, in which A_{CE} is used in place of A_C), we use the formula proposed in [1] and [5]. Accordingly, $\langle l_c \rangle$ and $\langle l_{CE} \rangle$ are presented in (4.2a) and (4.2b) respectively. e is the eccentricity (=$\sqrt{1 - \frac{b^2}{a^2}}$) of the CE of the OTNs (see Table 4.1).

$$\langle l_c \rangle = \frac{\sqrt{A_C}}{\sqrt{N} - 1}; \tag{4.2a}$$

$$\langle l_{CE} \rangle = \frac{\sqrt{A_{CE}}}{\sqrt{N} - 1}. \tag{4.2b}$$

For the correlations between the parameters of the OTNs, and their CEs, we used regressions (both linear and nonlinear) between their parameters. We found significantly strong correlations between their respective areas. This indicates that CE is useful in the estimation of average link lengths of the OTNs. The regressions are evaluated in terms of their coefficient of determination (R^2), which were found to be quite high.

Table 4.1: Comparison between the parameters of the 40 real OTNs and their CEs. Parameters of the 40 real OTNs used in this study and their CEs are provided in the first 8 columns (column '*T*' – column 'A_{CE}'). In the rest 5 columns (column '$\langle l_c \rangle$' – column '*e*') we present the evaluated parameters. ($\langle l \rangle$, a, b, $\langle l_c \rangle$, $\langle l_{CE} \rangle$, $\langle l_{ce} \rangle$ are in km; A_C and A_{CE} are in sq. km).

T	*Network*	*N*	$\langle l \rangle$	*a*	*b*	A_C	A_{CE}	$\langle l_c \rangle$	$\langle l_{CE} \rangle$	$\langle l_{ce} \rangle$	$Error(\%)$	*e*
1	VIA Network[1]	9	571	1178	512	1267876	1893847	563	688	583	-2.10	0.9006
2	BREN[1]	10	94	193	103	42196	62420	95	116	100	-6.38	0.8457
3	RNP[1]	10	748	1732	534	1814676	2904148	623	788	668	10.70	0.9513
4	Abilene Core[1]	10	1067	2053	1328	5698507	8560846	1104	1353	1145	-7.31	0.7625
5	LEARN[1]	10	189	386	233	183296	282405	198	246	210	-11.11	0.7973
6	CompuServe[1]	11	1161	2434	1220	5442148	9324167	1007	1318	1115	3.96	0.8653
7	vBNS[1]	12	965	2430	1276	6279993	9736135	1017	1266	1072	-11.09	0.8510
8	CESNET[1]	12	91	197	112	45957	69281	87	107	93	-2.20	0.8227
9	NFSNET[1]	14	1086	2318	1328	6007605	9665875	894	1134	960	11.60	0.8196
10	ITALY Net[1]	14	280	567	417	436577	742418	241	314	268	4.29	0.6776
11	ACONET[1]	15	119	269	110	52826	92913	80	106	92	2.69	0.9126
12	MZIMA[1]	15	852	2322	1410	6805153	10280423	908	1116	945	-10.92	0.7945
13	GARR-B[1]	16	224	548	382	476100	657315	230	270	230	-2.68	0.7170
14	ARNES[1]	17	38	95	55	12641	16407	36	41	37	2.63	0.8157
15	GERMANY[1]	17	143	350	266	196675	292334	142	173	148	-3.50	0.6499
16	REDIRIS[1]	17	319	520	387	463539	631894	218	255	218	31.66	0.6679
17	LambadaRail[1]	19	671	2389	1292	6899337	9691886	782	927	785	-16.99	0.8411
18	MEMOREX[1]	19	137	479	291	253850	437681	150	197	169	-23.36	0.9557
19	CANARIE[1]	19	668	2299	590	3462688	4259127	554	614	521	22.01	0.9666
20	EON[1]	19	754	1933	1343	5800056	8151500	717	850	720	4.51	0.7192
21	ARPANET[1]	20	839	2197	1270	5975012	8761197	704	852	722	13.95	0.8160
22	OPTOSunet[1]	20	100	308	158	122980	152805	101	113	98	2.00	0.8584
23	Hibernia USA[1]	20	279	1243	580	1257762	2263752	323	433	368	-31.90	0.8845
24	PIONIER[1]	21	131	393	223	194178	275186	123	146	126	3.82	0.8234
25	COX[1]	24	662	2258	890	4870065	6310207	566	644	546	17.52	0.9190
26	SANET[1]	25	36	161	80	24336	40443	39	50	45	-25.00	0.8678
27	NEWNet[1]	26	528	2463	1228	7033452	9497131	647	752	637	-20.64	0.8668
28	PORTUGAL[1]	26	203	969	489	888824	1487861	230	298	254	-25.12	0.8633
29	RENATER[1]	27	155	467	416	462096	610014	162	186	159	-2.58	0.4544
30	IBN31[1]	31	131	546	394	426658	675489	143	180	154	-17.56	0.6923
31	BULGARIA[1]	32	51	248	171	95108	133161	65	78	68	-33.33	0.7243
32	GEANT2[1]	32	661	2009	1747	7703318	11020530	596	713	604	8.62	0.4938
33	LONI[1]	33	62	241	162	68008	122592	56	74	65	-4.84	0.7404
34	METRONA[1]	33	73	324	273	158837	277739	84	111	96	-31.51	0.5386
35	COST37[1]	37	439	1798	1417	5371917	7999985	456	556	472	-7.52	0.6156
36	CERNET[1]	37	636	2054	1688	6719547	10886857	510	649	550	13.52	0.5698
37	OMNICOM[1]	38	298	1282	888	2596279	3574626	312	366	311	-4.36	0.7213
38	INTERNET2[1]	56	334	2441	1168	7169590	8952416	413	462	393	-17.66	0.8781
39	CORONET[1]	75	326	2476	1390	7903482	10806750	367	429	365	-11.96	0.8276
40	USA100[1]	100	310	2605	1391	8191044	11377963	318	375	319	-2.90	0.8455
	Absolute Average:										12.70	

4.3.1 Estimation of Convex Area from CE

Convex area of an OTN is the smallest possible convex set which contains all the nodes of the OTN [1]. It is always greater than or equal to the exact area [1]. It provides better accuracies for the estimations of the OTNs than other types of areas [1]. However, estimation of convex area is not very straightforward. It needs the location of all the nodes (though only the peripheral nodes are enough for its estimation). From the locations of the nodes, the lengths of the sides of the convex polygon are calculated for

the estimation of the convex area. In order to avoid these complex steps, we propose an alternative method.

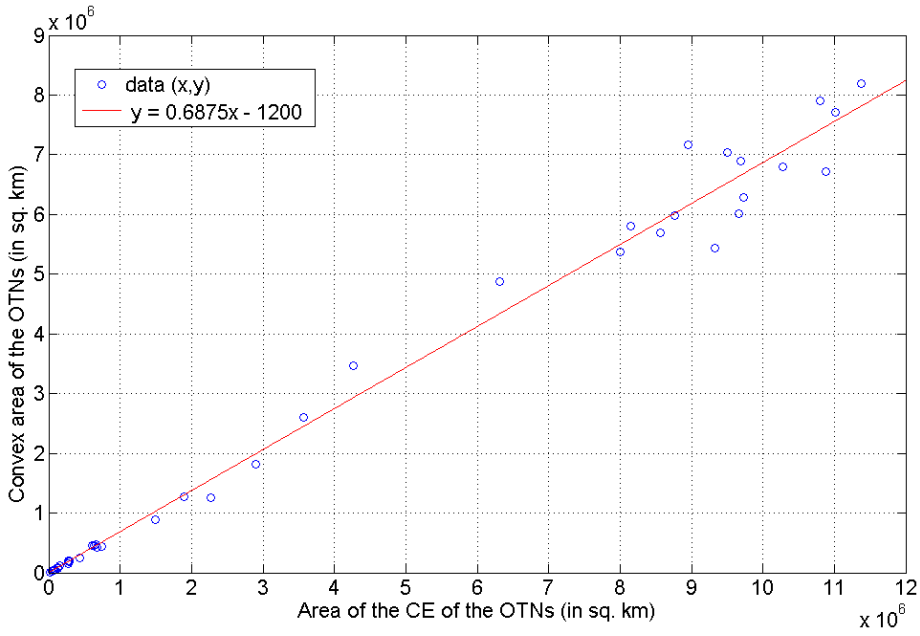

Figure 4.2: Relationship between the area of the CE (A_{CE}) and the convex area (A_C) of its OTN (R^2 = 0.9870)

$$A_C \cong 0.6875 A_{CE} - 1200 \tag{4.3}$$

Here, we use the area of CE of OTN (A_{CE}) to estimate its convex area (A_C). The area of the CE is estimated using (4.1), which just needs the values of a and b. From the linear regression of these two parameters, we found a simplified expression for convex area as shown in (4.3), and presented graphically in Figure 4.2. The coefficient of determination for this regression is 0.9870, which shows the justification of its use.

4.3.2 Estimating $\langle l \rangle$ from A_{CE}

Using (4.2a) and (4.2b), we calculated both $\langle l \rangle$, $\langle l_{CE} \rangle$ and their regressions show linear trends between these parameters (see Figure 4.3). From the linear regression between the exact average link lengths and the estimated average link lengths using the CE, we can estimate the average link length as shown in (4.4), denoted as $\langle l_{ce} \rangle$.

$$\langle l \rangle \cong 0.8446 \langle l_{CE} \rangle + 2.3 = \langle l_{ce} \rangle \tag{4.4}$$

Now substituting the formula for $\langle l_{CE} \rangle$ (from 4.2b), we get,

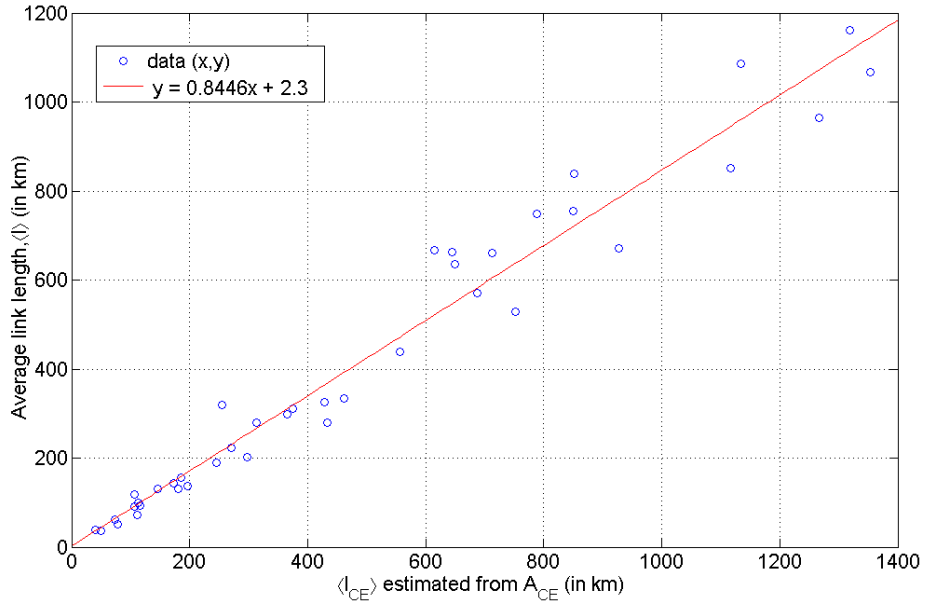

Figure 4.3: Relationships between the exact average link length ($\langle l \rangle$) of the OTNs and $\langle l_{CE} \rangle$ parameter estimated using A_{CE} (R^2 = 0.9613)

$$\langle l_{ce} \rangle = 0.8446 \frac{\sqrt{A_{CE}}}{\sqrt{N} - 1} + 2.3. \tag{4.5}$$

The plot of $\langle l_c \rangle$ versus $\langle l_{CE} \rangle$ too follow a linear trend (see Figure 4.4). These results show that the average link length, and the distribution of link lengths can be estimated from the CE. Estimation of CE parameters a and b is also quite straightforward and less complex than estimating A_C.

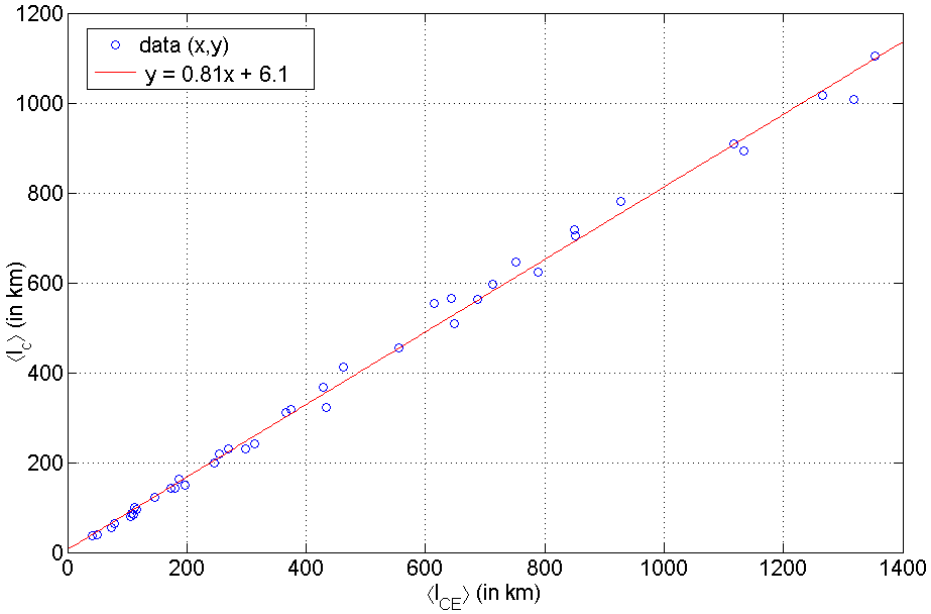

Figure 4.4: Relationships between $\langle l_{CE} \rangle$ parameter and $\langle l_c \rangle$ of the OTNs (R^2 = 0.9955)

In some cases, the rectangular area of a planar map that spreads the whole geography of the coverage area was used for the fast estimations of the parameters of the OTNs [19]. As shown here, the rectangular areas will incur larger errors than the area of CE, and thus should be replaced by the convex area or its estimate from the CE. In fact, the rectangular area and the area of the CE are much larger than the convex area of the OTN (as shown in the linear regression). So, the expression proposed in (4.5) should be used as it provides more accurate outcome.

4.3.3 Evaluation of the Method Based on CE

Evaluation of the proposed method is necessary for judging its effectiveness. We evaluate the performance of the method developed in this chapter for the estimation of the average link length of OTNs in terms of relative errors (in $\%$). The error is estimated using expression (4.6).

$$Error(\%) = \frac{\langle l \rangle - \langle l_{ce} \rangle}{\langle l \rangle} \times 100. \tag{4.6}$$

As can be seen in Table 4.1, in the last row under the column 'Error (%)', the absolute average of the errors for 40 OTNs is just 12.70$\%$. Using convex area, the average absolute error for these 40 real OTNs is 11.26$\%$ [1]. Thus, the error observed using the area of CE, is a little higher than the error obtained from the convex area. Overall, this method is very much satisfactory as the estimation of A_{CE} is much simpler than the estimation of A_C.

4.4 Chapter Summary

Convex area is necessary for the OTN related estimations. In this chapter, we showed the effectiveness of the estimation methods based on the CE of the OTNs. Its utilities in the network related estimations have also been showed using the cases of 40 real OTNs. Only the major and minor axes are needed for the estimations of the CEs. These ellipses are instrumental in the estimation of the convex areas of the OTNs, from which the average link lengths can be estimated. Considering the simplicities of these methods based on the CEs, the average error of 12.70$\%$ in the average link lengths is quite good. The link-related parameters of OTNs, in which average link lengths are used can be estimated using the results obtained in this work.

References

[1] S. K. Routray, R. M. Morais, J. R. F. da Rocha, and A. N. Pinto, "Statistical model for link lengths in optical transport networks," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 5, no. 7, pp. 762 – 773, Jul. 2013.

[2] S. K. Routray, G. Sahin, J. R. F. da Rocha, and A. N. Pinto, "Estimation of Link-Dependent Parameters of Optical Transport Networks from the Statistical Models," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 6, no. 7, pp. 601 – 609, July 2014.

[3] D. Mucci, "Approximation in area of graphs with isolated singularities," *Manuscripta Mathematica*, vol. 88, no. 1, pp. 135 – 146, 1995.

[4] H. Federer, *Geometric measure theory,* New York, NY: Springer Verlag, 2013.

[5] S. K. Korotky, "Network global expectation model: A statistical formalism for quickly quantifying network needs and costs," *IEEE/OSA Journal of Lightwave Technology*, vol. 22, no. 3, pp. 703 – 722, Mar. 2004.

[6] J. F. Labourdette, E. Bouillet, R. Ramamurthy, and A. A. Akyamaç, "Fast approximate dimensioning and performance analysis of mesh optical networks," *IEEE/ACM Transactions on Networking*, vol. 3, no. 4, pp. 906 – 917, Aug. 2005.

[7] F. Garin, D. Varagnolo, and K. H. Johansson, "Distributed estimation of diameter, radius and eccentricities in anonymous networks," in Proceedings of 3rd IFAC Workshop on Distributed Estimation and Control in Networked Systems, pp. 13 – 18, 2012.

[8] U. Zwick, "Exact and approximate distances in graphs – a survey," in *Algorithms – ESA 2001*, vol. 2161 of Lecture Notes in Computer Science, pp. 33 – 48. Springer, Heidelberg, 2001.

[9] D. Varagnolo, G. Pillonetto, and L. Schenato, "Distributed statistical estimation of the number of nodes in sensor networks," in Proceedings of IEEE Conference on Decision and Control, pp. 1498 – 1503, Atlanta, USA, 2010.

[10] M. Elkin, "Computing almost shortest paths," in Proceedings of the Twentieth Annual ACM Symposium on Principles of Distributed Computing (PODC), pp. 53 – 62, ACM Press, New York, USA, 2001.

[11] B. Deb, S. Bhatnagar, and B. Nath, "STREAM: Sensor topology retrieval at multiple resolutions," *Telecommunication Systems*, vol. 26, no. 2–4, pp. 285 – 320, 2004.

[12] P. Crescenzi, R. Grossi, C. Imbrenda, L. Lanzi, and A. Marino, "Finding the diameter in real-world graphs - experimentally turning a lower bound into an upper bound," in Proceedings of European Conference on Algorithms, 2010.

[13] D. G. Corneil, F. F. Dragan, M. Habib, and C. Paul, "Diameter determination on restricted graph families," *Discrete Applied Mathematics*, vol. 113, no. 2-3, pp. 143 – 166, 2001.

[14] J. C. S. Cardoso, C. Baquero, and P. S. Almeida, "Probabilistic estimation of network size and diameter," in Proceedings of Fourth IEEE Latin-American Symposium on Dependable Computing, pp. 33 – 40, Joao Pessoa, Brazil, 2009.

[15] K. Boitmanis, K. Freivalds, P. Ledins, and R. Opmanis, "Fast and simple approximation of the diameter and radius of a graph," in *Experimental Algorithms*, vol. 4007 of Lecture Notes in Computer Science, pp. 98 – 108, 2006.

[16] C. Baquero, P. S. S. Almeida, R. Menezes, and P. Jesus, "Extrema propagation: Fast distributed estimation of sums and network sizes," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 4, pp. 668 – 675, 2012.

[17] P. S. Almeida, C. Baquero, and A. Cunha, "Fast distributed computation of distances in networks," in Proceedings of 51st IEEE Conference on Decision and Control (CDC), pp. 5215 – 5220, 2012.

[18] D. Aingworth, C. Chekuri, R. Motwani, and P. Indyk, "Fast estimation of diameter and shortest paths (without matrix multiplication)" *SIAM Journal on Computing*, vol. 28, no. 4, pp. 1167 – 1181, 1999.

[19] C. Pavan, R. M. Morais, and A. N. Pinto, "Fast calculation of capex in optical multilayer networks," in Proceedings of the 14th European Conference on Networks and Optical Communications (NOC/OC&I), Valladolid, Spain, pp. 325 – 332, Jun. 2009.

[20] S. K. Routray, J. R. F. da Rocha, and A. N. Pinto, "Estimating the Parameters of Optical Transport Networks from their Circumferential Ellipses," in Proceedings of IEEE Internatinal Conference on Telecommunications (ICT), pp. 119 – 123, Lisbon, Portugal, 4-7 May 2014.

CHAPTER 5

Estimation of Link Related Parameters

5.1 Introduction

In CHAPTER 3, we presented a statistical model for the link lengths in optical transport networks (OTNs). This model can be estimated from the very basic information of the OTNs. Thus it is suitable for the estimations without complete information. In addition to the knowledge of the distribution of link lengths, this model can be applied to estimate link-related parameters of the OTNs. In this chapter, we develop these novel methods for the estimation of link-dependent parameters of OTNs.

5.1.1 Motivation

There are several link-dependent parameters which are needed for the early stage planning and dimensioning. Estimation of these network parameters is quite important at the beginning, because it gives an approximate idea about the initial capital needed for the deployment of the links of the OTNs. Using the statistical model developed for links lengths in Chapter 3, the link-dependent parameters of OTNs can be estimated with better accuracy. The methods we propose, are based on the statistical distribution model of the link lengths, which only needs an estimate of the average link length. Using these methods, we achieve significant improvement in the estimation accuracies, compared to the previous methods based on just the average link length. For these im-

provement, we do not need any extra network information. Overall, these new methods are very much attractive for the estimations with incomplete information.

5.1.2 Related Work

As presented in Chapter 2, the parameters of the OTNs can be broadly divided in to link-dependent and node-dependent parameters. Link-dependent parameters are used for the estimation of the link-related components of deployment expenditure [1]. In Chapter 3, a statistical model for the link lengths of the OTNs has been proposed. This model shows that the link lengths of the OTNs follow the General Extreme Value (GEV) distribution, and the corresponding GEV model can be estimated from the average link length of the OTN. This work also provides expressions for the average link lengths, and shows that the estimation from the convex area is better than the alternative estimates. As defined in Chapter 3, convex area is the smallest area that includes all the nodes and line segments between any two points of that area. The average link length based on the convex area can be estimated using only the knowledge of the node locations, and it provides better accuracy than the other alternatives [2]. In [1] and [3], simplified empirical and approximate expressions are proposed for the estimation of the parameters of OTNs. Several network parameters are estimated in [4] and [5], including some link-dependent parameters for which average link length is used. In [5], different costs of OTNs, and their associated parameters are evaluated for the OTNs. In [6], parameters of survivable OTNs are estimated using several analytical expressions without complete information. A statistical model for fast estimation of network parameters is proposed in [7]. In [8] and [9], the network parameter estimations are extended for multi-layer optical networks and several modified expressions are provided. Basics of GEV distribution and its parameters can be found in [2], and [10]. In [11] – [13], distance dependent resource allocation is proposed for better performances of the OTNs. These methods are focused on the optimized resource allocations in the OTNs.

5.1.3 Chapter Outline

The rest of this chapter is organized in 4 different sections. In section 5.2, we present the link length statistical model, its significance, how it can be estimated, and its relationship with the link-dependent parameters. We explain the sources of errors, when we use just the average link length for such estimations. In section 5.3, we provide our analytical methods, and present the new expressions for estimating the number of amplifiers, total fiber link length, and modulation schemes. In section 5.4, we evaluate the

performance of our proposed methods using 40 real OTNs. Concluding remarks on the overall effectiveness of the new methods developed in this chapter are presented in the last section. The work presented in this chapter is reported in [14].

5.2 Link Lengths and Link-Dependent Parameters

In Chapter 3, we developed a statistical model for link lengths in OTNs [2]. We have shown using statistical analysis and appropriate validation that the link lengths (l) of OTNs follow the GEV distribution. This distribution has three parameters (i.e., location parameter, α, scale parameter, β, and shape parameter, ξ) [2], [10]. All these parameters are found to depend on the average link length of the OTN [2]. In Chapter 3, we have shown that the average link length depends on the convex area and the number of nodes of the network [2]. Overall, the GEV distribution of the link lengths can be estimated with good accuracy, just from the knowledge of the nodes' locations of OTN.

5.2.1 Properties of GEV Distribution for Link Lengths

The link length model presented in Chapter 3 is based on the probability density function (PDF) of GEV distribution. The PDF, *f(l)* and cumulative distribution function (CDF), *F(l)* of the GEV distribution are shown in (5.1) and (5.2), respectively.

$$f(l;\alpha,\beta,\xi) = f(t;\beta,\xi) = \frac{1}{\beta}t^{(-1/\xi)-1}\exp(-t^{-1/\xi}) \tag{5.1}$$

$$F(l;\alpha,\beta,\xi) = F(t;\xi) = \exp(-t^{-1/\xi}) \tag{5.2}$$

In (5.1) and (5.2), $t = 1+\xi((l-\alpha)/\beta)$. Equations (5.1) and (5.2) are useful in calculating the probability of link lengths in a specific range. We later use them to estimate the CAPEX parameters of the OTNs that depend on the link lengths. The three parameters of the GEV distribution, α, β and ξ depend on the average link length, $\langle l\rangle$ as shown in expressions (5.3), (5.4) and (5.5), in which α, β and $\langle l\rangle$ are in km [2].

$$\alpha \approx 0.6577\langle l\rangle + 8.67 \tag{5.3}$$

$$\beta \approx 0.441\langle l\rangle - 12.37 \tag{5.4}$$

$$\xi \approx \frac{0.0887\langle l\rangle - 1.557}{0.5297\langle l\rangle - 13.927} \tag{5.5}$$

Using these three expressions, we can estimate the CDF, and PDF of the link lengths. The parameter $\langle l_c \rangle$ (the estimated version of $\langle l \rangle$), can be estimated using the convex area (A_c) and the number of nodes of the OTN (N) as shown in (5.6) [2].

$$\langle l \rangle \cong \langle l_c \rangle = \frac{\sqrt{A_c}}{\sqrt{N} - 1} \tag{5.6}$$

5.2.2 Link-Dependent CAPEX Parameters

In Chapter 2, we have seen that the components of CAPEX of OTNs can be broadly divided into link-dependent parameters, C_L, and node-dependent parameters, C_N [1]. Therefore we can write:

$$CAPEX = C_L + C_N. \tag{5.7}$$

In this chapter, we focus only on the former. The C_L parameters include number of optical amplifiers (1R regenerators) needed along the links, total length of fiber needed for the whole network, type of compensation schemes needed at different nodes, number and location of regenerators (both 2R and 3R) needed, and the choice of optical modulation and demodulation formats to be used in each link.

5.2.3 Errors in C_L Parameter Estimation

Previously used methods for the estimation of C_L parameters rely on the average link length of the OTN. If $K(l)$ denote the parameter value for a link of length l, the typical approach for estimating the parameter value for the whole OTN uses the expression $LK(\langle l_c \rangle)$, where $\langle l_c \rangle$ is the estimated average link length, and L is the number of links in the OTN. For instance, the total number of amplifiers needed along the channels in an OTN with a specific span (distance between consecutive amplifiers) length, as presented in [6] and [7], can be estimated as:

$$A'_{T_E} = L\left\lfloor \frac{\langle l_c \rangle}{span} \right\rfloor. \tag{5.8}$$

In (5.8), the floor function, $\lfloor a \rfloor$ represents the largest integer less than a. A similar expression can be used for the estimation of compensation methods required at the ends of the links between two nodes. The total length of the fiber needed in the network is computed from the estimated average link length as:

$$F'_{T_E} = L\langle l_c \rangle. \tag{5.9}$$

The use of estimated average link length for the above parameters as shown in (5.8) and (5.9) can lead to significant errors. We illustrate it in Figure 5.1, which shows the PDF of link lengths of an OTN (follows GEV distribution). The use of average link length (or its estimate) for cost estimation amounts to the assumption that all links have lengths equal to the average length, leading to significant error for certain type of costs. In comparison to that, the knowledge of the link length distribution provides more information about the links, and decreases the estimation errors.

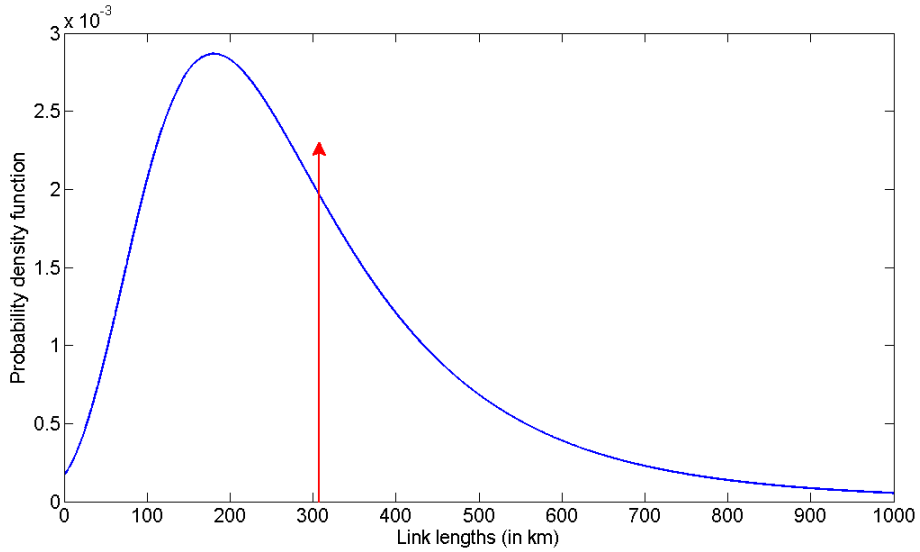

Figure 5.1: Estimated link length distribution for the USA100 network [2]. Its average link length is 310 km, and GEV distribution parameters are: α=213 km, β=124 km and ξ = 0.167. (The vertical red arrow shows its average link length.)

5.3 Estimations Using Link Statistical Model of OTNs

We have explained the possibility of errors in the calculations of C_L parameters using just the average link length in the previous section. In this section, we calculate the parameters more accurately using the PDF of link length distribution.

5.3.1 Number of Links in Specific Ranges

We use the PDF of GEV distribution as defined in expressions (5.1) and (5.3) – (5.5), for the estimation of fraction of the total links in a specific length range. We check the sum of links in different intervals with the total number of links of the 40 real OTNs. For any distribution, the net sum of the area under its PDF is 1. For GEV distribution too:

$$\int_{-\infty}^{\infty} f(l;\alpha,\beta,\xi)dl = 1. \tag{5.10}$$

In case of the OTNs, link lengths are positive entities and thus, we have:

$$\int_{0}^{\infty} f(l;\alpha,\beta,\xi)dl \cong 1. \tag{5.11}$$

It is also true that $\int_{-\infty}^{0} f(l;\alpha,\beta,\xi)dl \cong 0$, for the values of α, β and ξ obtained using (5.3), (5.4), and (5.5), respectively. In the estimation of link-dependent parameters such as the total number of amplifiers for the network, we need a real and finite upper limit. The upper limit of the integral in (5.11) can be replaced by the longest link length (*LLL*) of the corresponding network. However, it is not possible to estimate the longest link length from just the node locations. In the 40 real OTNs used in this analysis, the ratio of the longest link length (*LLL*) to that of the average link length ($\langle l_c \rangle$) varies from 1.4 to 7.05. After testing those 40 OTNs using linear and nonlinear regressions, we do not get a good single general expression for *LLL* (because, the coefficients of determination for the regressions are quite small). After checking several expressions for *LLL*, we chose the approximate values of this parameter as shown below:

i. $LLL \cong 2.7\langle l_c \rangle$, when, $\langle l_c \rangle \leq 100$;

ii. $LLL \cong 5\langle l_c \rangle$, when, $\langle l_c \rangle > 100$.

These values are realistic for the estimation of link probabilities of all the 40 real OTNs used in this study. With these values we confirm that:

$$\int_{0}^{LLL} f(l)dl \cong 1. \tag{5.12}$$

We can estimate the number of links with lengths in a certain length range using (5.1), as illustrated in Figure 5.2. This number can be estimated using (5.13).

$$L_{l_i - l_j} = L \int_{l_i}^{l_j} f(l;\alpha,\beta,\xi)dl \tag{5.13}$$

In (5.13), $L_{l_i - l_j}$ is the number of links with lengths in the interval l_i, to l_j. As the number of links in any interval is an integer, we round it to $\lfloor (L_{l_i - l_j}) \rceil$, where $\lfloor a \rceil$ is the nearest integer of a (provides the same result as rounding). Because of (5.12),

$$\left[\int_{l_0=0}^{l_1} f(l)dl + \int_{l_1}^{l_2} f(l)dl + ... + \int_{l_{X-1}}^{l_X = LLL} f(l)dl \right] \cong 1. \tag{5.14}$$

The subscript X, in (5.14) represents the number of partitions made in between 0 $(= l_0)$ and *LLL* $(= l_X)$. X = 16, for the case shown in Figure 5.2 (i.e., the link lengths between 0

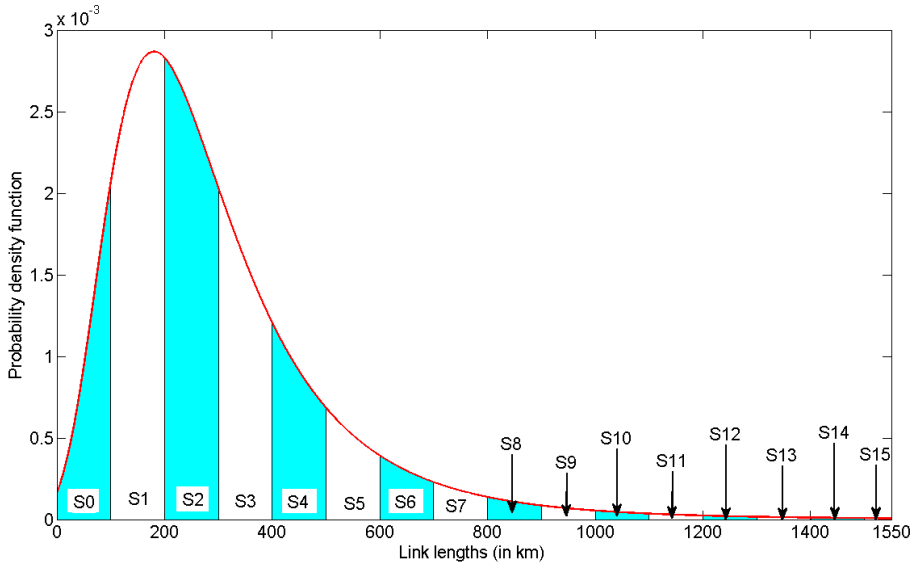

Figure 5.2: An OTN with average link length 310 km (so, *LLL* = 1550 km), and GEV distribution parameters α=213, β=124, and ξ=0.167. The proportion of links in each interval of 100 km is given by the area under its corresponding interval (i.e., S0, S1, ..., S15).

and 1550 km are partitioned into S0, S1, ..., S15). In general, for a fixed span length, X is the smallest integer not smaller than $\frac{LLL}{span}$ (i.e., X = $\lceil \frac{LLL}{span} \rceil$). Using expressions (5.12), (5.13) and (5.14), the number of links in different intervals obeys the condition:

$$L = \lfloor (L_{l_0-l_1} + L_{l_1-l_2} + ... + L_{l_{X-1}-l_X}) \rceil. \tag{5.15}$$

Expression (5.15) holds good for all the 40 real OTNs used in this chapter for different span lengths between 0 and *LLL*. It vindicates the choices made for *LLL* in (5.12), for OTNs with different values of $\langle l_c \rangle$.

5.3.2 Total Number of Amplifiers

Optical amplifiers are needed to strengthen the weak signal in OTNs. Its exact number is essential for planners to estimate the C_L part of the total CAPEX. The estimation in (5.8) is based on the average link length, and the outcome can be erroneous in the majority of the cases.

Here, we use the statistical model which takes the distribution of the link lengths into account. In this method, we reduce the errors significantly. We make partitions of the link length distribution as per length intervals that fall between integer multiples of the span length. These partitions determine the number of amplifiers needed for the whole network. This method uses expressions (5.12) – (5.15), and is illustrated

in Figure 5.2, which considers spans of 100 km between amplifiers. The links having lengths between $l_0(=0)$ and $l_1(=100)$ do not need any amplifier as they are too short. However, the links between $l_1(=100)$ and $l_2(=200)$ need one amplifier each, the links between $l_2(=200)$ and $l_3(=300)$ need two amplifiers each, and so on till the last interval, $l_{X-1} - l_X$ which needs $X-1$ amplifiers per link. In general, the number of amplifiers required for the length span, $l_i - l_{i+1}$ is:

$$A_{l_i - l_{i+1}} = iL_{l_i - l_{i+1}}. \tag{5.16}$$

The total number of amplifiers estimated in this method (A_{T_E}), after rounding, is:

$$A_{T_E} = \left\lfloor \sum_{i=0}^{X-1} A_{l_i - l_{i+1}} \right\rceil = \left\lfloor \sum_{i=0}^{X-1} iL_{l_i - l_{i+1}} \right\rceil; \tag{5.17}$$

which can be simplified as:

$$A_{T_E} = \lfloor 0 \cdot L_{l_0 - l_1} + 1 \cdot L_{l_1 - l_2} + ... + (X-1) \cdot L_{l_{X-1} - l_X} \rceil.$$

The estimation, obtained using equation (5.17) is more accurate than the common method based on equation (5.8), where just the average link length is used (as shown later in section 5.4).

5.3.3 Types of Modulation Schemes

Distance dependent modulation-demodulation schemes are found to be more effective than the singular choices (one type for the whole network) [11]. Like the number of optical amplifiers, types of modulation and demodulation schemes at each node depend on the link lengths.

A similar estimation as formulated in subsection 5.3.2, can be used to determine the type of modulation. The number of links corresponding to each type of modulation format can be estimated by applying the half distance law proposed in [11] and [12], to the link distribution model. This law lowers the spectral efficiency with the increasing link lengths for overall optimum performance. According to that, the modulation format must be changed when the length of the span (initial span was chosen to be 375 km in [12]) is doubled as shown in Figure 5.3. So links with lengths of 0 – 375 km, 375 – 750 km, 750 – 1500 km, and 1500 – 3000 km are to be modulated using 16QAM, 8QAM, QPSK and BPSK, respectively.

As a result, the number of links that use 16QAM, 8QAM, QPSK and BPSK modulation formats in a network are respectively given by $N_{16QAM}(=NS)$, $N_{8QAM}(=NE)$,

$N_{QPSK}(= NQ)$ and $N_{BPSK}(= NB)$ in equation (5.18). The corresponding results are presented in Table 5.2, and has been discussed in the next section.

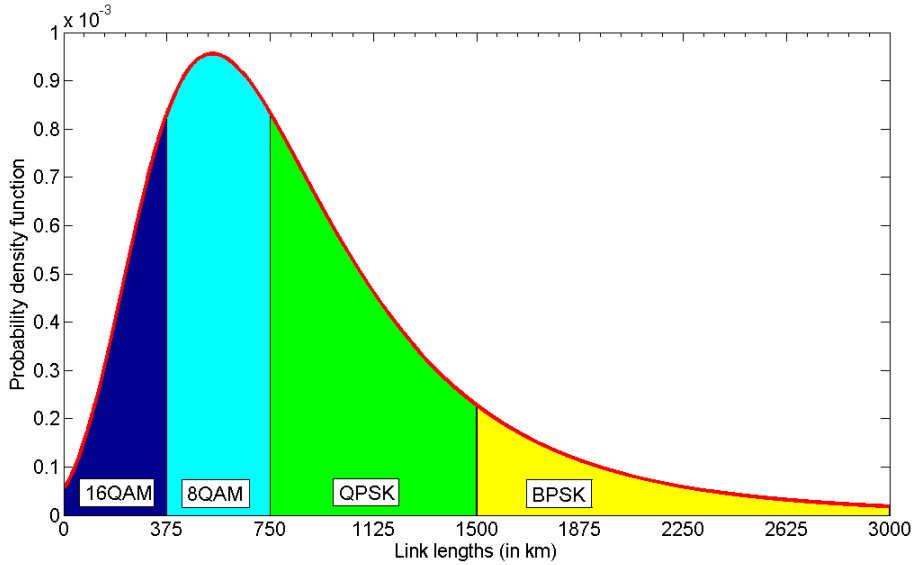

Figure 5.3: An OTN with average link length 930 km, and GEV distribution parameters α=600, β=390, and ξ=0.167. It shows the link probabilities according to the half distance law proposed in [12].

$$N_{16QAM} = NS = \left\lfloor L \int_0^{375} f(l;\alpha,\beta,\xi)dl \right\rfloor; \tag{5.18a}$$

$$N_{8QAM} = NE = \left\lfloor L \int_{375}^{750} f(l;\alpha,\beta,\xi)dl \right\rfloor; \tag{5.18b}$$

$$N_{QPSK} = NQ = \left\lfloor L \int_{750}^{1500} f(l;\alpha,\beta,\xi)dl \right\rfloor; \tag{5.18c}$$

$$N_{BPSK} = NB = \left\lfloor L \int_{1500}^{3000} f(l;\alpha,\beta,\xi)dl \right\rfloor. \tag{5.18d}$$

Estimations for other link-dependent parameters such as the compensation methods required at the ends of the links, the number and location of regenerators and pre-amplifiers needed can be done using similar methods.

5.3.4 Total Length of Fiber

Cost of fiber is related to the length of the fiber the operators need to deploy in the network for the planned operations. It is a significant part of the CAPEX. In general, estimation of the total fiber cost is obtained by multiplying the cost per unit length by the average link length ($\langle l \rangle$) and the number of links (L). Corresponding total length of fiber, F_T, is given by

$$F_T = L\langle l \rangle. \tag{5.19}$$

In the planning phase, the exact value of $\langle l \rangle$ is not known, and we estimate its value $\langle l_c \rangle$. It gives errors because,

$$\langle l \rangle \neq \langle l_c \rangle. \tag{5.20}$$

So, using $\langle l_c \rangle$ we have:

$$Error = L\left(\langle l \rangle - \langle l_c \rangle\right). \tag{5.21}$$

Using the distribution of the link lengths (estimated from $\langle l_c \rangle$) this error can be reduced. The accuracy in this case is dependent on the accuracy of the estimated average link length. Considering the link lengths as a random variable l, with probability density function *f(l)*, and by setting the lower limit to 0 (as negative link lengths are not possible) the total length of fiber (F_{T_E}) is given by:

$$F_{T_E} = L \int_0^\infty l f(l) dl. \tag{5.22}$$

The entity shown in the integral in (5.22) represents the expected value of l (thus, the multiplication of L, and the expected value of l is the total length of fiber).

5.4 Evaluation of the Developed Methods

In this section, we evaluate the performances of the proposed statistical methods by comparing them with alternative methods based on the same input parameters in 40 real OTNs. We start by calculating the exact parameters, i.e., the total number of amplifiers, A_T and the total fiber length, F_T as shown in Table 5.1. These values are obtained using the complete information of the OTNs. These exact results are compared with the results A_{T_E} and F_{T_E} obtained using our proposed estimation methods based on (5.17) and (5.22), respectively, which only need partial information of the OTN. The results are shown in Table 5.1, which also provides the estimations A'_{T_E} and F'_{T_E} based on (5.8) and (5.9), respectively and hence represent the commonly used methods. These estimations (A'_{T_E} and F'_{T_E}) use the same information as our proposed statistical methods. In Table 5.1, N and L are the number of nodes and links of the networks, respectively. The parameter $\langle l_c \rangle$ is the estimated average link length, obtained using (5.6), based on just the location of the nodes of the OTNs. The span length for evaluating A_T, A_{T_E} and A'_{T_E} was taken to be 100 km.

The performances of the estimation methods are evaluated by the percentage of error using expression (5.23). The errors are estimated with respect to the exact values (i.e., A_T and F_T).

$$Error(\%) = \frac{Exact - Estimated}{Exact} \times 100. \tag{5.23}$$

In Table 5.1, column 'E_{T_E}' represents the percentage errors of the number of amplifiers estimations performed with our proposed method (A_{T_E}) compared with the exact calculations (A_T). Similarly, column 'E'_{T_E}' shows the errors in the number of amplifiers based on the common method (A'_{T_E}). Columns 'E_{F_E}' and 'E'_{F_E}' show the percentage errors related to the total fiber length when using the proposed and the common methods, respectively.

As can be seen from Table 5.1, it is clear that the estimation of number of amplifiers based on the proposed method provides, in general, significant error reduction, when compared with the common method. For the estimation of number of amplifiers, in six cases both methods provide same results; in 28 cases the proposed method is better than the common method, and only in six cases it gives worse results. Overall performance in terms of the absolute average error (see row 'Absolute Average' under columns '$E_{T_E}(\%)$' and '$E'_{T_E}(\%)$') with the proposed method is quite better than the common method: the average error is 14.74% for our proposed distribution based method, whereas for the common method based on the average link length it is 31.10%. We have found that these differences between the two methods are almost independent of the span lengths.

On the other hand, for the total length of fiber, we obtain an improvement of just 1.25% (= $E_{F_E}(\%) - E'_{F_E}(\%)$), (see row 'Absolute Average' under columns '$E_{F_E}(\%)$' and '$E'_{F_E}(\%)$').

Using the half distance law proposed in [12], we estimate the types of modulation formats required for the links of the 40 real OTNs in three different methods (see Table 5.2), viz.: UELL (Using Exact Link Lengths), UALL (Using Average Link Length), and ULLD (Using Link Length Distribution). According to this law, BPSK (B in Table 5.2), QPSK (Q in Table 5.2), 8QAM (E in Table 5.2), and 16QAM (S in Table 5.2) formats can be used for transmissions up to 3000 km, 1500 km, 750 km, and 375 km, respectively. The numbers that precede S, E, Q and B in Table 5.2 are the number of links that need to use that modulation format. In column 'UELL', is the exact number for the various modulation formats required according to the half distance law. The sum of the numbers preceding S, E, Q and B under UELL column is equal to the total number of links (L) of the associated OTN. In column 'UALL', we provide the estimation using $\langle l_c \rangle$ alone (i.e., $\langle l_c \rangle$ determines whether it is S or E or Q or B for all the links). In column 'ULLD' is our

Table 5.1: Comparison between different methods for 40 real OTNs. The first five columns represent the basic network information, the next five (A_{T_E}, $A^{'}_{T_E}$, A_T, $E_{T_E}(\%)$ and $E^{'}_{T_E}(\%)$) are the estimations related to the optical amplifiers, and the last five columns (F_{T_E}, $F^{'}_{T_E}$, F_T, $E_{F_E}(\%)$, and $E^{'}_{F_E}(\%)$) are the estimations related to the total fiber length. ($\langle l_c \rangle$, F_{T_E}, $F^{'}_{T_E}$ and F_T are in km).

No.	*Network*	N	L	$\langle l_c \rangle$	A_{T_E}	$A^{'}_{T_E}$	A_T	$E_{T_E}(\%)$	$E^{'}_{T_E}(\%)$	F_{T_E}	$F^{'}_{T_E}$	F_T	$E_{F_E}(\%)$	$E^{'}_{F_E}(\%)$
1	Via NET[2]	9	12	563	61	60	64	4.69	6.25	6737	6756	6852	-1.68	-1.4
2	BREN[2]	10	11	95	4	0	4	0	100	1034	1045	1034	0	1.06
3	RNP[2]	10	12	623	68	72	86	20.93	16.28	7456	7476	8976	-16.93	-16.71
4	Abilene Core[2]	10	13	1104	127	143	134	5.22	-6.72	14258	14352	13871	2.79	3.47
5	LEARN[2]	10	12	198	17	12	16	-6.25	25	2362	2376	2268	4.14	4.76
6	CompuServe[2]	11	14	1007	127	140	155	18.06	9.68	14068	14098	16254	-13.45	-13.26
7	vBNS[2]	12	17	1017	150	170	155	3.23	-9.68	17229	17289	16405	5.02	5.39
8	CESNET[2]	12	19	87	5	0	6	16.67	100	1634	1653	1729	-5.49	-4.4
9	NSFNET[2]	14	21	894	173	168	218	20.64	22.94	18717	18774	22806	-17.93	-17.68
10	ITALY[2]	14	29	241	55	58	67	17.91	13.43	4556	4579	5320	-14.36	-13.93
11	ACONET[2]	15	22	80	5	0	20	75	100	1738	1760	2618	-33.61	-32.77
12	MZIMA[2]	15	19	908	154	171	150	-2.67	-14	16141	17252	16188	-0.29	6.57
13	GARR-B[2]	16	27	230	48	54	46	-4.35	-17.39	6115	6210	6048	1.11	2.68
14	ARNES[2]	17	20	36	0	0	0	0	0	716	720	760	-5.79	-5.26
15	GERMANY[2]	17	26	142	23	26	23	0	-13.04	3660	3692	3718	-1.56	-0.7
16	REDIRIS[2]	17	28	218	47	56	66	28.79	15.15	6071	6104	8932	-32.03	-31.66
17	Lambada Rail[2]	19	23	782	157	161	142	-10.56	-13.38	17237	17986	15433	11.69	16.54
18	MEMOREX[2]	19	24	150	23	24	21	-9.52	-14.29	3345	3600	3288	1.73	9.49
19	CANARIE[2]	19	26	554	130	130	160	18.75	18.75	14414	14404	17368	-17.01	-17.07
20	EON[2]	19	37	717	238	259	220	-8.18	-17.73	26462	26529	27898	-5.15	-4.91
21	ARPANET[2]	20	32	704	207	224	253	18.18	11.46	22512	22528	26848	-16.15	-16.09
22	OPTOSunet[2]	20	24	101	11	24	9	-22.22	-166.67	2398	2424	2400	-0.08	1
23	Hibernia USA[2]	20	27	323	72	81	65	-10.77	-24.62	8366	8721	7533	11.06	15.77
24	PIONIER[2]	21	25	123	17	25	16	-6.25	-56.25	3047	3075	3275	-6.96	-6.11
25	COX[2]	24	40	566	203	200	211	3.79	5.21	22576	22640	26480	-14.74	-14.5
26	SANET[2]	25	28	39	0	0	0	0	0	1065	1092	1008	5.65	8.33
27	NEWNET[2]	26	31	647	176	186	150	-17.33	-24	19587	20057	16368	19.67	22.54
28	PORTUGAL[2]	26	36	230	64	72	45	-42.22	-60	8153	8280	7308	11.56	13.3
29	RENATER[2]	27	35	162	37	35	36	-2.78	2.78	5620	5670	5425	3.59	4.52
30	IBN31[2]	31	51	143	46	51	40	-15	-27.5	7221	7293	6681	8.08	9.16
31	BULGARIA[2]	32	33	65	2	0	1	-100	100	2080	2145	1683	23.59	27.45
32	GEANT2[2]	32	52	596	279	260	315	11.43	17.46	30906	30992	34372	-10.08	-9.83
33	LONI[2]	33	37	56	0	0	0	0	0	2036	2072	2294	-11.25	-9.68
34	Metrona[2]	33	41	84	10	0	10	0	100	3352	3444	2993	11.99	15.07
35	COST37[2]	37	57	456	227	228	223	-1.79	-2.24	25364	25992	25023	1.36	3.87
36	CERNET[2]	37	53	510	243	265	314	22.61	15.61	26948	27030	33708	-20.05	-19.81
37	OMNICOM[2]	38	54	312	138	162	133	-3.76	-21.8	16759	16848	16092	4.14	4.7
38	INTERNET2[2]	56	61	413	218	244	186	-17.2	-31.18	25416	27258	20374	24.75	33.79
39	CORONET[2]	75	99	367	309	297	338	8.58	12.13	36115	36333	32274	11.9	12.58
40	USA100 [1]	100	171	318	446	513	390	-14.36	-31.54	54089	54378	53010	2.04	2.58
	Absolute Average:							14.74	31.10				10.26	11.51

proposed estimation using (5.18).

We evaluate the performances of estimation of modulation formats using the total number of correct modulation formats predicted by the methods. We denote it as C, the number of links for which the modulation formats have been estimated correctly by ULLD or UALL with respect to UELL. It is estimated according to expression (5.24). *Min*[], in (5.24), is the minimum selecting function, whose output is the smallest number among its arguments in []. Z $\in \{S, E, Q, B\}$, and $N_{Z/UELL}$ is the number under column 'UELL' in the Z category. $N_{Z/OM}$, is the corresponding number in the Z category in other methods (OMs) under column 'ULLD' and 'UALL' for the estimation of C_{ULLD} and C_{UALL}, respectively. We estimate the correct predictions in each modulation category

(i.e., S, E, Q and B) and add them together as shown in (5.24) to get the total number of correct predictions (C).

$$C = \sum_{Z \in \{S,E,Q,B\}} Min[N_{Z/UELL}, N_{Z/OM}]. \tag{5.24}$$

In Table 5.2, the errors are estimated using expression (5.25). The errors in UALL and ULLD are estimated with respect to UELL (columns '$E_{UALL}(\%)$' and '$E_{ULLD}(\%)$'). For instance, as shown in Table 5.2, Abilene Core (row 4) has 13 links and its $\langle l_c \rangle$ is 1104 km. According to the half distance law, the UALL estimation is 13Q, meaning that all the 13 links will have the QPSK modulation format. However, the ULLD estimation results in 1S – 4E – 5Q – 3B, that is, 1 link will have 16QAM, 4 links will have 8QAM, 5 links will have QPSK and 3 links will have BPSK. So, the value of *C* for Abilene Core in UALL (=C_{UALL} in Table 5.2) is 6 (= Min [13, 6], because between '13Q' and '2S – 3E – 6Q – 2B,' only 6 links are found to have common modulation formats as estimated in the respective methods), and *C* for ULLD (=C_{ULLD} in Table 5.2) is 11 (= Min [1, 2] + Min [4, 3] + Min [5, 6] + Min [3, 2], because between '1S – 4E – 5Q – 3B' and '2S – 3E – 6Q – 2B,' 11 links are found to have common modulation formats as estimated using the respective methods).

$$Error\ (\%) = \frac{L - C}{L} \times 100. \tag{5.25}$$

The estimation of ULLD values uses the same information as for the UALL values (i.e., node locations and number of links). However, ULLD provides better estimations (see the average error values in columns '$E_{ULLD}(\%)$' and '$E_{UALL}(\%)$', last row in Table 5.2), because it successfully estimates the number of links in the four specified length ranges. Out of the 40 networks, only in case of 3 ULLD provides a larger error than the UALL. In fact, these errors occur in case of a few small networks with large number of links, and are quite small in value (see the corresponding errors in column '$E_{ULLD}(\%)$'). It is worth noting that the proposed ULLD method is highly accurate in predicting the set of modulations schemes with an average error of only 6.64%, whereas the common method entails an average error of 32.11% (see row 'Average', columns 'E_{ULLD}' and 'E_{UALL}' in Table 5.2). This is significant, given that the method only relies on the knowledge of the node locations and the total number of links.

Table 5.2: Choosing modulation formats for 40 real OTNs using the average link length $\langle l_c \rangle$ (column UALL), link length distribution (column ULLD) and exact link lengths (column UELL). S, E, Q and B stand for 16QAM, 8QAM, QPSK and BPSK respectively ($\langle l_c \rangle$ is in km).

No.	*Network*	$\langle l_c \rangle$	*UELL*	*ULLD*	*UALL*	C_{ULLD}	C_{UALL}	$E_{ULLD}(\%)$	$E_{UALL}(\%)$
1	Via NET[2]	563	4S – 6E – 1Q – 1B	4S – 5E – 2Q – 1B	12E	11	6	8.33	50.00
2	BREN[2]	95	11S	11S	11S	11	11	0.00	0.00
3	RNP[2]	623	4S – 4E – 2Q – 2B	4S – 5E – 3Q	12E	10	4	16.67	66.67
4	Abilene Core[2]	1104	2S – 3E – 6Q – 2B	1S – 4E – 5Q – 3B	13Q	11	6	15.38	53.85
5	LEARN[2]	198	11S – 1E	11S – 1E	12S	12	11	0.00	8.33
6	CompuServe[2]	1007	3S – 3E – 5Q – 3B	2S – 4E – 6Q – 2B	14Q	12	5	14.29	64.29
7	vBNS[2]	1017	3S – 4E – 6Q – 4B	2S – 5E – 7Q – 3B	17Q	15	6	11.76	64.71
8	CESNET[2]	87	19S	19S	19S	19	19	0.00	0.00
9	NSFNET[2]	894	3S – 7E – 6Q – 5B	3S – 7E – 8Q – 3B	21Q	19	6	9.52	71.43
10	ITALY[2]	241	22S – 7E	25S – 4E	29S	26	22	10.34	24.14
11	ACONET[2]	80	20S – 2E	22S	22S	20	20	9.09	9.09
12	MZIMA[2]	908	4S – 5E – 8Q – 2B	3S – 7E – 7Q – 2B	19Q	17	8	10.53	57.89
13	GARR-B[2]	230	21S – 6E	24S – 3E	27S	24	21	11.11	22.22
14	ARNES[2]	36	20S	20S	20S	20	20	0.00	0.00
15	GERMANY[2]	142	26S	25S – 1E	26S	25	26	3.85	0.00
16	REDIRIS[2]	218	21S – 7E	25S – 3E	28S	24	21	14.29	25.00
17	Lambada Rail[2]	782	6S – 10E – 6Q – 1B	5S – 9E – 7Q – 2B	23Q	21	6	8.70	73.91
18	MEMOREX[2]	150	24S	23S – 1E	24S	23	24	4.17	0.00
19	CANARIE[2]	554	10S – 10E – 3Q – 3B	10S – 11E – 5Q – 1B	26E	24	10	7.69	61.54
20	EON[2]	717	9S – 17E – 9Q – 2B	9S – 15E – 11Q – 2B	37E	35	17	5.41	54.05
21	ARPANET[2]	704	4S – 12E – 13Q – 3B	8S – 13E – 9Q – 2B	32E	27	12	15.63	62.50
22	OPTOSunet[2]	101	24S	24S	24S	24	24	0.00	0.00
23	Hibernia USA[2]	323	20S – 6E – 1Q	19S – 7E – 1Q	27S	27	12	15.63	62.50
24	PIONIER[2]	123	25S	25S	25S	25	25	0.00	0.00
25	COX[2]	566	14S – 17E – 7Q – 2B	14S – 17E – 8Q – 1B	40E	39	17	2.50	57.50
26	SANET[2]	39	28S	28S	28S	28	28	0.00	0.00
27	NEWNET[2]	647	12S – 12E – 7Q	9S – 13E – 8Q – 1B	31E	28	12	9.68	61.29
28	PORTUGAL[2]	230	32S – 4Q	32S – 4E	36S	32	32	11.11	11.11
29	RENATER[2]	162	34S – 1E	34S – 1E	35S	35	34	0.00	2.86
30	IBN31[2]	143	51S	50S – 1E	51S	50	51	1.96	0.00
31	BULGARIA[2]	65	33S	33S	33S	33	33	0.00	0.00
32	GEANT2[2]	596	15S – 22E – 10Q – 5B	17S – 22E – 11Q – 2B	52E	49	22	5.77	57.69
33	LONI[2]	56	37S	37S	37S	37	37	0.00	0.00
34	Metrona[2]	84	41S	41S	41S	41	41	0.00	0.00
35	COST37[2]	456	29S – 22E – 5Q – 1B	28S – 22E – 7Q – 1B	57E	56	22	1.75	61.40
36	CERNET[2]	510	17S – 23E – 9Q – 4B	22S – 22E – 8Q – 1B	53E	48	23	9.43	56.60
37	OMNICOM[2]	312	37S – 16E – 1Q	40S – 12E – 2Q	54S	50	37	7.41	31.48
38	INTERNET2[2]	413	46S – 12E – 3Q	34S – 21E – 5Q – 1B	61E	49	12	19.67	80.33
39	CORONET[2]	367	53S – 35E – 11Q	63S – 30E – 6Q – 1B	99S	89	53	10.10	46.46
40	USA100 [1]	318	133S – 35E – 3Q	123S – 41E – 7Q	171S	161	133	5.85	22.22
	Average:							6.64	32.11

5.5 Chapter Summary

In the initial phases, when designing OTNs, estimation of link-dependent parameters is an important step. At the early stage, details of network topology may be not known. So it is mandatory to have estimation methods based only on partial knowledge of the network configuration. In this chapter, it is shown that the most important link-dependent network parameters can be estimated with improved accuracy by using methods based on the link length statistical model, when compared with the alternative methods based on the average link length. In the estimation of the total number of amplifiers, the error (compared with exact solution) is reduced significantly from 31.10% for the method based on the average link length, to 14.74% achieved with our proposed method. The improvement is very significant when dealing with the type of modulation

formats to be used on various links: here the proposed method implies an error reduction from 32.11% to 6.64%. Of course this method for the modulation formats is suitable for the opaque OTNs (as the shortest path lengths are required for the transparent OTNs, it will be evaluated in Chapter 7). For total fiber length, the accuracy of the proposed method is not very significant as we got an average improvement of just 1.26%. Overall, the improvements are obtained without the requirement of additional information.

References

[1] S. K. Korotky, "Network Global Expectation Model: A Statistical formalism for Quickly Quantifying Network Needs and Costs," *IEEE/OSA Journal of Lightwave Technology*, vol. 22, no. 3, pp. 703 – 722, March 2004.

[2] S. K. Routray, R. M. Morais, J. R. F. da Rocha, and A. N. Pinto, "Statistical Model for Link Lengths in Optical Transport Networks," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 5, no. 7, pp. 762 – 773, Jul. 2013.

[3] J. F. Labourdette, E. Bouillet, R. Ramamurthy, and A. A. Akyamaç, "Fast approximate dimensioning and performance analysis of mesh optical networks," *IEEE/ACM Transactions on Networking*, vol. 13, no. 4, pp. 906 – 917, Aug. 2005.

[4] A. N. Pinto, C. Pavan, and R. M. Morais, "Dimensioning optical networks: A practical approach," in Proceedings of International Conference on Transparent Networks (ICTON), Munich, Germany, pp. 1 – 4, June 2010.

[5] A. N. Pinto, C. Pavan, R. M. Morais, and A. Correia, "Cost evaluation in optical networks," in Proceedings of International Conference on Transparent Networks (ICTON), Stockholm, Sweden, pp. 1 – 4, June 2011.

[6] C. Pavan, R. M. Morais, and A. N. Pinto, "Estimation of CapEx in Survivable Optical Transport Networks," *Proc European Conf. on Networks and Optical Communications (NOC)*, Faro, Portugal, vol. 1, pp. 263 – 268, June, 2010.

[7] A. N. Pinto, C. Pavan, and R. M. Morais, "A Statistical Model to CAPEX Fast Calculation in Optical Transport Networks," in Proceedings of International Conference on Transparent Networks (ICTON), Ilha de São Miguel - Açores, Portugal, Vol. 1, pp. 1 – 4, June, 2009.

[8] C. Pavan, R. Morais, A. Correia, and A. Pinto, "Dimensioning of optical networks with incomplete information," *Proc of the 6th Advanced International Conference on Telecommunications, AICT '08*, Athens, Greece, pp. 261– 264, June 2008.

[9] C. Pavan, R. M. Morais, and A. N. Pinto, "Estimating capex in optical multilayer networks," in Proceedings of the 7th Conference on Telecommunications (ConfTele'09), Santa Maria da Feira, Portugal, pp. 335 – 338, May 2009.

[10] S. Kotz, S. Nadarajah. *Extreme Value Distributions: Theory and Applications*, London, England: Imperial College Press, 2000.

[11] K. Chirstodoulopoulos, I. Tomkos, and E. A. Varvarigos, "Elastic Bandwidth Allocation in Flexible OFDM based Optical Networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 29, no. 9, pp. 1354 – 1366, May 2011.

[12] A. Bocoi, M. Schuster, F. Rambach, M. Kiese, C.-A. Bunge, and B. Spinnler, "Reach-Dependent Capacity in Optical Networks Enabled by OFDM," in Proceedings of International Conference on Optical Fiber Communication (OFC), pp. 1 – 3, Paper OMQ4, 2009.

[13] M. Jinno, B. Kozicki, H. Takara, A. Watanabe, Y. Sone, T. Tanaka, A. Hirano, "Distance-Adaptive Spectrum Resource Allocation in Spectrum-Sliced Elastic Optical Path Network," *IEEE Communication Magazine*, vol. 48, no. 8, pp. 138 – 145, Aug. 2010.

[14] S. K. Routray, G. Sahin, J. R. F. da Rocha, and A. N. Pinto, "Estimation of Link-Dependent Parameters of Optical Transport Networks from the Statistical Models," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 6, no. 7, pp. 601 – 609, July 2014.

CHAPTER 6

Statistical Modeling of Shortest Path Lengths

6.1 Introduction

SHORTEST PATH LENGTHS are important physical parameters of optical transport networks (OTNs). They exhibit several structural and dynamical characteristics of the OTNs. In this chapter, we analyze the shortest path lengths between the node pairs of real OTNs. Using the results of the analysis, we present a statistical characterization of physical shortest path lengths and their key properties in OTNs. The prime focus of this chapter is to obtain a statistical model for the shortest path lengths of OTNs that can be used for several practical applications. We propose a statistical model and show that the key statistical parameters such as the mean, median, standard deviation and an upper bound of the shortest path lengths can be estimated from the convex area of the OTN, which depends only on the node locations. We also provide the formalism for its estimation from basic information of the OTNs.

6.1.1 Motivation

Estimation of the shortest paths between two nodes in a network is a classic and fundamental problem in graph theory, network science, electrical engineering, computer science and several other related fields [1]. Shortest path lengths between the node pairs of the OTNs are required for estimating several characteristics and parameters of the

OTNs. Knowledge of the shortest path lengths between any two nodes in OTNs is used for routing, traffic management, control and several other network management and design purposes. The shortest path lengths can also be used for estimating the associated components of capital expenditure (CAPEX), operational expenditure (OPEX), and management expenditure (MANEX) of the OTNs [2].

6.1.2 Related Work

Shortest path lengths are fundamental entities in many areas of science, mathematics and engineering [1]. Several studies have been conducted on the topological analysis of OTNs. Main topological characteristics, such as the physical lengths of links and paths, and the number of hops of each path play important roles in understanding the behavior of graphs and networks that arise in various contexts, including OTNs [3]. Shortest path lengths in the context of large graphs such as the Internet have been the subject of several research efforts [4]. In [4] and [5], the authors presented several properties of the shortest path lengths in large graphs collected from different websites, and provided approximate expressions for such estimations. Improvement of path length estimation is studied in [6] via network distance-based coordinate systems, with an emphasis on the accuracy of the associated parameters. In [7], the shortest path lengths in random networks are found to follow Weibull distribution. In the case of OTNs, there have been several efforts focusing on topological properties related to hop lengths of paths, link lengths, as well as their applications in network design and optimization. Several expressions are proposed for fast estimation of parameters such as the number of hops and performance analysis of OTNs in [2] and [3]. The authors in [3] analyzed the hop lengths of paths and provided an approximate expression for the average number of hops. Survivable hierarchical optical networks are analyzed, and their design aspects are presented with dedicated wavelength path protection in [8]. Several OTN parameters are estimated and dimensioning models are proposed in [9] – [12]. In these studies, OTN costs and related parameters are estimated with incomplete information. In [13] a statistical model for link lengths is provided (also presented in Chapter 3), which depends only on the node locations of the OTNs. It is also shown that the convex area (the bounded interior area of the smallest convex set comprising of all the nodes) is more effective than the other areas (i.e., exact area and the geographical area) in the network-related estimations in OTNs. Key parameters of the OTNs such as the average link length, and the link length distribution model are estimated with good accuracy using the convex area. The model developed in [13] can also be used to estimate the link-dependent parameters of the OTNs [14] which also presented in Chapter 5. Con-

vex area can be estimated from the circumferential ellipse of the OTN with reasonable accuracy [15].

6.1.3 Chapter Outline

The rest of this chapter is organized in 4 sections. In section 6.2, real OTN topologies are analyzed and a new model is proposed for the shortest path lengths between the nodes. In section 6.3, expressions are developed for the mean, median and standard deviation of the shortest path lengths, using only the convex area and the number of nodes of the OTNs. In section 6.4, estimation of the model proposed in section 6.2 is presented, requiring only the basic parameters of the OTNs. In section 6.5, the main concluding remarks on the effectiveness of the proposed models are presented. The work presented in this chapter is reported in [16].

6.2 Analysis of Shortest Path Lengths in OTNs

For this study, we analyzed 40 real OTNs. The total number of shortest paths between unique node pairs in an OTN is determined by its total number of nodes. For a network with N nodes, there are $N(N-1)/2$ $(= P)$ shortest path lengths. In general, the actual shortest path lengths in an OTN would depend on factors such as the link lengths, the node adjacencies, and the total number of links, L. The smallest shortest path length is equal to the shortest link length of the network.

In order to identify a general statistical distribution for the shortest path lengths, we analyzed a diverse set of OTNs. We included real networks of all the 6 continents in this analysis, with OTNs of small, medium as well as large sizes (having average link lengths spanning from 35 km to 1190 km). These 40 networks are quite well known, and have been used in several studies, and described in [13] – [19]. Graphical topologies of these networks can be found in [20] and [21]. Out of these 40 OTNs, 30 are mentioned and used in [13] – [15], and the rest are listed in [22] – [31].

6.2.1 Measurement of Link Lengths, Shortest Path Lengths and Convex Areas

For the statistical analysis of the shortest path lengths in 40 real OTNs, we need the exact lengths of shortest paths between the node pairs of the OTNs, and hence the exact link lengths. In this study, Google Earth Professional (version 6.2) was used to measure the exact (great circle) link lengths of the 40 real OTNs. Accuracy of the measurement

of link lengths was cross-checked using OPNET Transport Planner 15 (a commercial software tool for network dimensioning). Both Google Earth Professional and OPNET Transport Planner 15 provided almost the same results. The shortest path lengths between each pair of the nodes of the OTNs were calculated by adding up the associated link lengths. As shown in Chapter 3, we measured the convex areas of the OTNs using Google Earth Professional (version 6.2).

6.2.2 Statistical Distributions of Shortest Path Lengths

All the shortest path lengths in all of the 40 real OTNs were calculated, and their statistical properties were studied for the development of a suitable distribution. It was found from the analysis that the shortest path length statistics do not fit with the commonly used distributions such as normal and Rayleigh. Besides that, we found that none of the one-, and two-parameter distributions is suitable for the shortest path lengths of the 40 real OTNs. Therefore, we extended this analysis to a wide pool of 65 distributions. We used, EasyFit (a software from Mathwave.com) for this statistical analysis (also, cross-checked using Matlab, wherever possible). For each statistical distribution, we found the optimal parameters for each network's shortest path length distribution, and evaluated its validity. Out of the 65 distributions analyzed for this study, the top 15 best fitting are presented in Table 6.1.

Table 6.1: Best Fitting Distributions and Their Number of Input Parameters (No. I/P), Average (Avg. KSS), Lowest (L. KSS) and Highest (H. KSS) KSS Values for the Shortest Path Lengths of the OTNs

No.	Distribution	No. I/P	Avg. KSS	L. KSS	H. KSS
1	Wakeby	5	0.0408	0.0129	0.0837
2	Johnson S_B	4	0.0423	0.0166	0.0788
3	General Pareto	4	0.0533	0.0288	0.0863
4	Beta	4	0.0560	0.0308	0.1492
5	Gamma (4P)	4	0.0577	0.0199	0.1210
6	Log Pearson 3	3	0.0585	0.0202	0.1255
7	GEV	3	0.0589	0.0252	0.1073
8	Error	3	0.0615	0.0215	0.1260
9	Weibull (3P)	3	0.0653	0.0160	0.1343
10	Weibull	2	0.0677	0.0170	0.1715
11	Log-Normal (3P)	3	0.0691	0.0259	0.1373
12	Pearson 6	4	0.0692	0.0263	0.1405
13	Gamma	3	0.0702	0.0276	0.1450
14	Log-Logistic (3P)	3	0.0706	0.0327	0.1309
15	Uniform	2	0.0727	0.0295	0.1259

(nP) *in* (·) *is the n-parameter version of the distribution*

We measured the validity or the 'goodness-of-fit' of the distributions in terms of the Kolmogorov-Smirnov (KS) statistic (KSS value). We choose the KS test because it is

suitable for small number of samples, and can also be applied to medium and large ones (suitable for our case, since the number of shortest paths of the networks, P, ranges from 21 to 528). Furthermore, this test is also appropriate for bounded distributions, for which the Anderson-Darling (AD) test does not provide significant outlook on the fitting. In our pool of 65 different distributions, several are bounded, giving us further motivation to choose the KS test for the goodness-of-fit. The smaller the Kolmogorov-Smirnov statistic is, the smaller the maximum difference is between the hypothesized cumulative distribution function (CDF) and the empirical distribution function of the real data (thus better fitting the distribution is for the samples). The ranking of the distributions in Table 6.1 has been done as per their average KSS values (Avg. KSS in Table 6.1). These values were obtained for each of the 65 distributions, by taking the average over the KSS values obtained for each one of the 40 studied networks. The lowest value of KSS (L. KSS in Table 6.1) and the highest value of KSS (H. KSS in Table 6.1) among the 40 OTNs have also been presented in Table 6.1. When the confidence interval (CI) of all the KS tests are set to 0.95 (significance level = 0.05), the results obtained for all the networks are statistically acceptable for the first two distributions (i.e., Wakeby and Johnson S_B) shown in Table 6.1.

6.2.3 Selection of the Suitable Distribution

As can be seen from Table 6.1, Wakeby distribution has the lowest Avg. KSS and L. KSS for the 40 real OTNs out of the 65 distributions used for this analysis. However, this distribution needs five input parameters, and its H. KSS is larger than the H. KSS value for the Johnson S_B distribution. Furthermore, Wakeby distribution does not have a well-defined expression for its probability density function (PDF). It is normally expressed in terms of its quantile function. The average KSS value of Johnson S_B distribution is the second smallest among the 65 distributions analyzed (see Table 6.1). It is better than all of its nearest rivals in almost all aspects except for the Wakeby distribution. As can be seen in Table 6.1, the highest value of KSS (H. KSS), is the lowest for this distribution. In terms of the lowest value of KSS (L. KSS), the Johnson S_B distribution is very much close to Wakeby distribution (L. KSS column of Table 6.1), and it is lower than the L. KSS of all other distributions. In this distribution, the PDF can be either unimodal, in which case, it increases initially up to the mode and then decreases with the increase of shortest path lengths; or bimodal with two different modes. The availability of the bimodal version of Johnson S_B distribution makes it a better choice than the other distributions including the Wakeby, because this property is exhibited by 6 networks with large convex area and small number of nodes. Therefore, this distribution is preferred for the shortest

path lengths of the OTNs. In Figure 6.1, we show the shortest path length histogram for a real OTN and its best-fit Johnson S_B distribution.

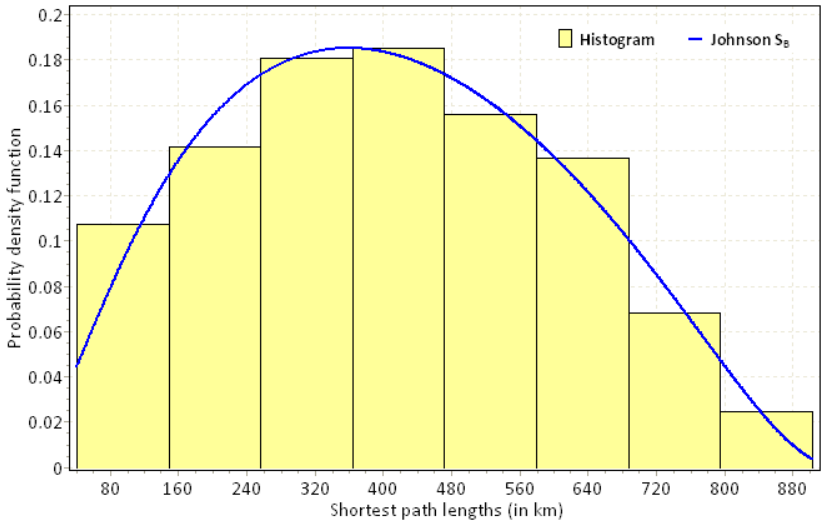

Figure 6.1: Johnson S_B distribution fitted to the path length histograms of Pionier network [13]. The KSS for this fitting is: 0.0255.

6.2.4 Basics of Johnson S_B Distribution

Johnson S_B distribution is one of the transformed distributions of the 'Johnson System' family. It is the only bounded distribution of this family of distributions ('S' stands for 'system', and the subscript 'B' indicates the boundedness). Johnson S_B is related to the normal distribution in several ways [32] . It can be regarded as a modified lognormal distribution [33]. However, it is different from the commonly known lognormal family of distributions because it has four input parameters, rather than the typical value of three. It is a true four-parameter lognormal distribution [33], with the parameters: δ, γ, λ, and ζ. Its PDF and CDF are shown in (6.1) and (6.2).

$$f(z;\gamma,\delta,\lambda,\zeta) = \frac{\delta}{\lambda\sqrt{2\pi}z(1-z)} exp\left[-\frac{1}{2}\left(\gamma + \delta log\left(\frac{z}{1-z}\right)\right)^2\right] \tag{6.1}$$

$$F(z;\gamma,\delta,\lambda,\zeta) = \Phi\left(\gamma + \delta log\left(\frac{z}{1-z}\right)\right) \tag{6.2}$$

Here, $z = \frac{p-\zeta}{\lambda}$, in which p represents the shortest path length variable, and $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-(\frac{t^2}{2})}dt$. This integral presented as $\Phi(x)$, is the CDF of a standard normal distribution, whose mean is 0, and standard deviation is 1. From this, it is clearly indicated that $x\left[= \gamma + \delta log\left(\frac{z}{1-z}\right) = \gamma + \delta log\left(\frac{p-\zeta}{\zeta+\lambda-p}\right)\right]$, is a unit normal variable, and thus p is a lognormal variable.

Table 6.2: Parameters of shortest path lengths of 40 real OTNs. The first 10 columns show the network attributes, obtained from measurement. The last 5 columns (γ, δ, λ, ζ, and KSS_{JS_B}) are obtained from their distribution fitting ($\langle A_c \rangle$ is in Sq. km; $\langle p \rangle$, m, σ_p, $p_M + p_m$, λ and ζ are in km).

#	Network	N	L	$\langle D \rangle$	A_c	$\langle p \rangle$	m	σ_p	$p_M + p_m$	γ	δ	λ	ζ	KSS_{JS_B}
1	SANREN[23]	7	7	2.00	508348	874	913	432	1628	-0.1918	0.7582	1735	-75.46	0.0788
2	Via NET[13]	9	12	2.67	1267876	1211	1035	698	2959	0.3948	0.5895	2491	239.25	0.0615
3	BREN[13]	10	11	2.20	42196	210	210	101	488	0.5892	1.2047	577	-18.45	0.0520
4	RNP[13]	10	12	2.40	1814676	1558	1468	913	3814	0.3848	0.7198	3626	113.62	0.0546
5	LEARN[13]	10	12	2.40	183296	464	409	255	1030	0.2848	0.7345	10105	30.29	0.0686
6	Abilene Core[13]	10	13	2.60	5698507	2204	2025	1174	4964	0.3598	0.6334	4327	459.96	0.0699
7	SINET[24]	11	13	2.36	406969	580	499	346	1580	1.0951	1.0312	1976	5.80	0.0483
8	CompuServe[13]	11	14	2.55	5442148	2420	2276	1429	5375	0.1811	0.5259	4695	323.84	0.0619
9	vBNS[13]	12	17	2.83	6279993	2351	2160	1418	5107	0.1690	0.5512	4763	202.75	0.0585
10	CESNET[13]	12	19	3.17	45957	194	194	90	432	0.3194	1.1362	480	-17.53	0.0601
11	AARNET[22]	13	13	2.00	7591320	2481	2305	1495	6378	0.6582	0.9136	7176	-94.21	0.0468
12	FLRNET[30]	13	14	2.15	156097	422	406	238	1026	0.7292	1.2512	1426	-112.72	0.0533
13	NSFNET[13]	14	21	3.00	6007605	2304	2221	1186	4681	0.1563	0.7792	4832	71.31	0.0502
14	ITALY[13]	14	29	4.14	436577	524	497	256	1137	0.2106	0.8126	1076	39.89	0.0384
15	HEANET[31]	15	15	2.00	55361	297	289	154	630	0.1115	0.7372	604	11.49	0.0302
16	MZIMA[13]	15	19	2.53	6805153	2419	2344	1279	5330	0.2350	0.8527	5549	-56.88	0.0530
17	ACONET[13]	15	22	2.80	52826	311	297	169	694	0.2398	0.9029	762	-29.90	0.0485
18	KAREN[28]	16	17	2.13	349015	555	503	325	1418	0.8270	0.9464	1652	12.24	0.0402
19	BELNET[25]	16	18	2.25	17466	139	132	70	341	0.7989	1.3918	455	-30.95	0.0312
20	ERNET[27]	16	18	2.25	2382858	1713	1830	799	3612	-0.1750	0.9891	3824	-338.47	0.0406
21	GARR-B[13]	16	27	3.38	476100	549	503	292	1191	0.3441	0.8786	1304	-2.81	0.0571
22	ARNES[13]	17	20	2.35	12641	111	109	59	269	0.4956	1.0633	306	-12.08	0.0372
23	GERMANY[13]	17	26	3.06	196675	351	345	182	835	0.6316	1.4154	1181	-122.52	0.0373
24	REDIRIS[13]	17	28	3.29	463539	600	616	241	1181	-0.1013	1.2510	1374	-111.38	0.0432
25	CALREN[29]	19	22	2.32	168087	466	445	264	1040	0.1598	0.7141	1020	-2.74	0.0337
26	Lambada Rail[13]	19	23	2.42	6899337	2305	2153	1311	5665	0.5712	0.9301	6268	-65.16	0.0316
27	MEMOREX[13]	19	24	2.53	253850	396	374	193	965	0.4723	1.2133	1097	-60.51	0.0302
28	CANARIE[13]	19	26	2.74	3462688	1928	1795	1098	4339	0.2453	0.7019	4228	80.45	0.0283
29	EON[13]	19	37	3.89	5800056	1495	1397	807	4114	1.0033	1.2391	5042	-165.75	0.0433
30	OPTOSunet[13]	20	24	2.40	122980	380	372	189	838	0.0908	0.7806	767	12.91	0.0313
31	Hibernia USA[13]	20	27	2.70	1257762	783	731	458	2314	1.5218	1.5730	3715	-299.42	0.0360
32	ARPANET[13]	20	32	3.20	5975012	2339	2143	1253	4881	0.2193	0.6260	4512	353.12	0.0363
33	PIONIER[13]	21	25	2.38	194178	409	409	199	944	0.2234	1.0486	994	-43.57	0.0255
34	SANET[13]	25	28	2.24	24336	174	174	87	424	0.5721	1.3383	540	-44.72	0.0410
35	FUNET[26]	26	27	2.08	210828	441	429	231	1019	0.2186	0.8658	1010	-13.33	0.0290
36	NEWNET[13]	26	31	2.38	7033452	2303	2144	1277	5169	0.3212	0.7962	5343	69.82	0.0257
37	RENATER[13]	27	35	2.59	462096	543	541	243	1231	0.0108	1.0640	1222	-65.50	0.0190
38	BULGARIA[13]	32	33	2.06	95108	294	276	164	834	1.0845	1.3301	1092	-62.22	0.0167
39	LONI[13]	33	37	2.24	68008	288	278	155	698	0.3380	1.0449	782	-50.17	0.0257
40	Metrona[13]	33	41	2.48	158837	293	276	154	771	0.9401	1.2894	979	-44.20	0.0173
	Average:				1971995	992	938	542	2284	0.4185	0.9656	2395	2.15	0.0423
	Lowest:				12641	111	109	59	269	-0.1918	0.5259	306	-338.47	0.0166
	Highest:				7591320	2481	2344	1495	6378	1.5218	1.5730	7176	459.96	0.0788

The parameters δ and γ are known as the shape parameters since they determine the shape of the distribution. Always, $\delta > 0$, and, γ can be any real number. The shape parameter δ is key in determining the modality of the PDF (both the number, and magnitude of the modes). The second shape parameter, γ determines the orientation (i.e., left or right tilting) of the PDFs. For unimodal distributions, when $\gamma > 0$, the PDF is left oriented (the mode is smaller than the mean and median) and for $\gamma < 0$, the PDF is right oriented (the mode is larger than the mean and median). The scale parameter, λ, and the location parameter, ζ present the location and spread of the distribution. In

other words, ζ is the starting point, and $\zeta + \lambda$ is the end point of the PDF and CDF of the Johnson S_B distribution.

The necessary and sufficient conditions for the bimodality of the PDF of the Johnson S_B distribution [34] are:

i. $\delta < \frac{1}{\sqrt{2}}$, and
ii. $|\gamma| < \delta^{-1}\sqrt{1-2\delta^2} - 2\delta tanh^{-1}\sqrt{1-2\delta^2}$.

Out of the 40 OTNs presented in Table 6.2, the first condition is satisfied in 6 networks. Out of those 6, only 2 networks satisfy the second condition. These 2 networks that satisfy both conditions exhibit clear bimodal PDFs. However, the remaining 4 networks that satisfy the first condition, but not the second exhibit hybrid characteristics. While the second mode is not clear for those networks, their shortest path length distributions are different from the pure unimodal and the pure bimodal PDFs.

We calculated the mean ($\langle p \rangle$), and the standard deviation (σ_p) of the shortest path lengths using the equations shown in (6.3) and (6.4), in which, p_i represents the individual shortest path lengths between node pairs.

$$\langle p \rangle = \frac{1}{P}\sum_{i=1}^{P} p_i \tag{6.3}$$

$$\sigma_p = \left(\frac{1}{P}\sum_{i=1}^{P}(p_i - \langle p \rangle)^2\right)^{\frac{1}{2}} \tag{6.4}$$

We show the parameters for the 40 real OTNs in Table 6.2. In the first five columns, we present the basic attributes of the OTNs including the number of nodes (N), number of links (L), and mean nodal degree, ($\langle D \rangle = 2L/N$). We present the measured parameters such as the convex area (A_c), mean ($\langle p \rangle$), median (m), standard deviation (σ_p), and the sum of the smallest (p_m) and the largest shortest path lengths (p_M), which are also defined mathematically in 6.5(a) – 6.5(c).

$$p_m = min(p_i) \tag{6.5a}$$

$$p_M = max(p_i) \tag{6.5b}$$

$$m = median(p_i) \tag{6.5c}$$

We present the parameters of Johnson S_B distribution (γ, δ, λ, and ζ) obtained from the statistical fitting using EasyFit in the next four columns. In the last column, we present the KSS values ('goodness-of-fit' results) for the shortest path lengths of the OTNs, when considering the Johnson SB distribution. This distribution was chosen due

to its low KSS values when considering all the 40 real OTNs (see last column, Table 6.1). The values shown in the last three rows of right most column of Table 6.2 are precisely the values presented in Table 6.1, 3rd row, columns 4 – 6.

6.3 Analysis of the Statistical Parameters of the Shortest Path Lengths

The moments of a distribution and other statistical parameters such as the median are quite useful in the applications of the distribution. However, Johnson S_B distribution does not have simple expressions for calculating its moments. We show the expressions for the moments of this distribution in Appendix C. The expressions presented in [32] – [38] for the moments and other associated statistical parameters, like the median, are based on several assumptions such as, $\sigma_p \leq 0.2\langle p \rangle$. These assumptions are not consistent with the shortest path lengths of the OTNs (see Table 6.2). Therefore these expressions do not provide good accuracies for the shortest path lengths of the OTNs. There are several instances in which the moments of Johnson S_B distribution have been estimated for the real data [39] – [41]. These estimation methods are also based on several assumptions (some of which are similar to the cases in [32] and [33]), and are not satisfied by the OTNs. So these methods are not suitable for OTNs.

So, instead of using the expressions presented for the moments of Johnson S_B distribution in Appendix C, which are not suitable for this case, we developed alternative estimation expressions that are quite simple. Convex area, which is a key parameter in the estimation of the average link length and the link-related parameters [13] – [14], is found to be instrumental in the estimation of the first two moments. So, we develop expressions for the average, median and standard deviation of the shortest path lengths using the convex area. We also observed a linear trend between the sum of the smallest and the largest values of the shortest path lengths with the square root of the convex area. We show the expressions obtained from linear regression in (6.6) – (6.9), and also in Figure 6.2. The coefficients of determination (i.e., R^2) for these regressions are larger than 0.95, which support the linear trends. We use subscript 'c' in expressions (6.6) – (6.9) to indicate that these expressions depend on the convex area of the OTNs. Although the sum of p_m and p_M provide a good linear trend with the square root of convex area, p_m and p_M, individually do not provide good coefficient of determination for the associated regressions with $\sqrt{A_c}$.

$$\langle p_c \rangle \cong 0.8988\sqrt{A_c} + 36.8090 \tag{6.6}$$

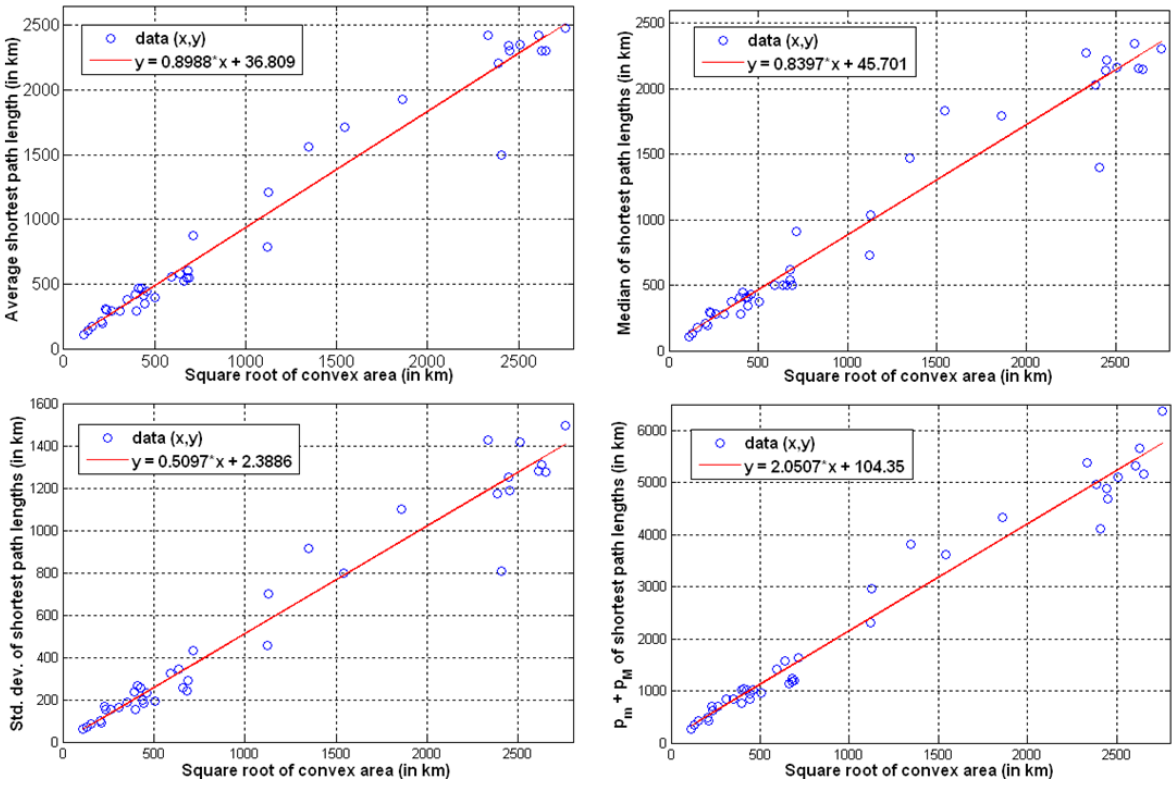

Figure 6.2: Linear regression between the mean, median, standard deviation and the sum of the smallest and largest shortest path lengths with the square root of the convex area of the OTNs shown in the clockwise order, starting from the top left. The coefficients of determination (R^2) for these regressions are 0.9623, 0.9536, 0.9522 and, 0.9596, respectively.

$$m_c \cong 0.8397\sqrt{A_c} + 45.7010 \tag{6.7}$$

$$\sigma_{pc} \cong 0.5079\sqrt{A_c} + 2.3886 \tag{6.8}$$

$$p_{m_c} + p_{M_c} \cong 2.0507\sqrt{A_c} + 104.3500 \tag{6.9}$$

We measure the accuracy of these estimations in terms of the percentage of error as shown in (6.10). Subscript 'x' in (6.10) represents the parameters with which the error is associated. In Table 6.3, we show the values obtained with expressions (6.6) – (6.9), for the 40 OTNs. The corresponding percentage errors are also shown in the last 4 columns of Table 6.3. These values were obtained by using the estimated values from Table 6.3, and the corresponding exact values from Table 6.2 in (6.10). The low average absolute

Table 6.3: Accuracy of the estimated parameters of shortest path lengths of forty real optical transport networks. The last 4 columns (E_{p_c}, E_{m_c}, $E_{\sigma_{pc}}$, and $E_{p_M+p_m}$) are the errors in the estimation of $\langle p_c \rangle$, m_c, σ_{pc}, and $p_M + p_m$ respectively ($\langle p_c \rangle$, m_c, σ_{pc}, and $p_M + p_m$ are in km).

#	Network	$\langle p_c \rangle$	m_c	σ_{pc}	$p_{m_c} + p_{M_c}$	$E_{p_c}(\%)$	$E_{m_c}(\%)$	$E_{\sigma_{pc}}(\%)$	$E_{p_{M_c}+p_{m_c}}(\%)$
1	SANREN[23]	678	644	365	1566	22.43	29.46	15.51	3.81
2	Via NET[13]	1049	991	574	2413	13.38	4.25	17.77	18.45
3	BREN[13]	221	218	107	525	-5.24	-3.81	-5.94	-7.58
4	RNP[13]	1247	1177	687	2867	19.96	19.82	24.75	24.83
5	LEARN[13]	421	405	220	982	9.27	0.98	13.73	4.66
6	Abilene Core[13]	2182	2050	1215	4999	1.00	-1.23	-3.49	-0.71
7	SINET[24]	610	581	326	1413	-5.17	-16.43	5.78	10.57
8	CompuServe[13]	2134	2005	1187	4889	11.82	11.91	16.93	9.04
9	vBNS[13]	2289	2150	1275	5243	2.64	0.46	10.08	-2.66
10	CESNET[13]	229	225	111	543	-18.04	-15.98	-23.33	-25.69
11	AARNET[22]	2513	2359	1402	5754	-1.29	-2.34	6.22	9.78
12	FLRNET[30]	392	377	203	914	7.11	7.14	14.71	10.92
13	NSFNET[13]	2240	2104	1247	5131	2.78	5.27	-5.14	-9.61
14	ITALY[13]	631	601	338	1460	-20.42	-20.93	-32.03	-28.41
15	HEANET[31]	248	243	122	586	16.50	15.92	20.78	6.98
16	MZIMA[13]	2382	2236	1327	5455	1.53	4.61	-3.75	-2.35
17	ACONET[13]	244	239	119	576	21.54	19.53	29.59	17.00
18	KAREN[28]	568	542	303	1316	-2.34	-7.75	6.77	7.19
19	BELNET[25]	155	157	69	375	-11.51	-18.94	1.43	-9.97
20	ERNET[27]	1425	1342	787	3271	16.81	26.67	1.50	9.44
21	GARR-B[13]	657	625	353	1519	-19.67	-24.25	-20.89	-27.54
22	ARNES[13]	137	140	59	334	-23.42	-28.44	0.00	-24.16
23	GERMANY[13]	435	418	227	1013	-23.93	-21.16	-24.73	-21.32
24	REDIRIS[13]	649	618	348	1501	-8.17	-0.32	-44.4	-27.10
25	CALREN[29]	405	390	211	945	13.09	12.36	20.08	9.13
26	Lambada Rail[13]	2398	2252	1337	5492	-4.03	-4.6	-1.98	3.05
27	MEMOREX[13]	490	469	258	1138	-23.74	-25.4	-33.68	-17.93
28	CANARIE[13]	1709	1608	948	3921	11.36	10.42	13.66	9.63
29	EON[13]	2201	2068	1225	5042	-47.22	-48.03	-51.80	-22.56
30	OPTOSunet[13]	352	340	181	824	7.37	8.6	4.23	1.67
31	Hibernia USA[13]	1044	987	572	2403	-33.33	-35.02	-24.89	-3.85
32	ARPANET[13]	2233	2098	1244	5116	4.53	2.10	0.72	-4.81
33	PIONIER[13]	433	416	226	1009	-5.87	-1.71	-13.57	-6.89
34	SANET[13]	177	177	82	424	-1.72	-1.72	5.75	0.00
35	FUNET[26]	449	431	236	1046	-1.81	-0.47	-2.16	-2.65
36	NEWNET[13]	2420	2273	1349	5543	-5.08	-6.02	-5.64	-7.24
37	RENATER[13]	648	617	348	1499	-19.34	-14.05	-43.21	-21.77
38	BULGARIA[13]	314	304	159	736	-6.8	-10.14	3.05	11.75
39	LONI[13]	271	265	135	640	5.90	4.68	12.90	8.31
40	Metrona[13]	395	381	205	923	-34.81	-38.04	-33.12	-19.71
	Absolute Average Errors:					12.80	13.27	15.49	11.77

errors (see last row, Table 6.3), show that these estimations are quite reliable and can be used for practical applications. The average, median, and the standard deviation of the path lengths can be estimated with average absolute errors of 12.8%, 13.27%, and 15.49%, respectively.

$$E_x(\%) = \frac{Exact - Estimated}{Exact} \times 100 \tag{6.10}$$

These outcomes are significant, given that the estimations rely only on the knowledge of the convex area of the network, which in turn can be obtained from the location of nodes [13]. In other words, only limited network information is needed for the methods

presented in this section. We note that our expression for $p_m + p_M$, (i.e. the sum of the smallest and the largest shortest path length) in the network, also yields a low average error of 11.77%. Since this expression could practically serve as an approximate upper bound on the maximum shortest path length, it is useful, for instance in estimating the system reach needed for near-transparent operation in an OTN.

6.4 Estimation of the Johnson S_B Distribution Model

There are several cases in which the parameters of Johnson S_B distribution have been estimated for the fitting with real data. These estimations are based on several assumptions which do not hold good for the shortest path lengths of OTNs. The methods provided in [32] have been used for the estimation of the parameters, and it is found to be statistically acceptable (the KSS obtained with these parameters exceed the critical value at 95% CI) for only 19 networks out of total 40. Hence, we do not find these methods suitable for the shortest path length distributions. The method described in [34], for the median of Johnson S_B (also shown in Appendix D) can be used for the estimation of a shape parameter (estimation of shaparters is critical for Johnson S_B distribution). However, this method does not produce satisfactory results (only 21 networks out of total 40 are statistically acceptable).

Tables to facilitate the fitting of the Johnson S_B curves are presented in [35] and [36]. The tables given in [36] estimate the two shape parameters when both terminals (i.e., ζ and λ) are known. In [36], approximate expressions for the shape parameters have been mentioned for small values of the standard deviation. However, these expressions do not provide satisfactory results for all the 40 OTNs. Similarly, the values obtained from the tables are also not very effective with the estimated values of the mean, median, standard deviation, ζ and λ; though they provide good results with the exact values of these parameters. There are several instances of the applications of Johnson S_B distribution, in which both the traditional and new methods are used for the estimation of its parameters [39] – [41]. Therefore, we choose to use the convex area and the number of nodes for these estimations (these parameters can be obtained from the node locations).

6.4.1 Estimation of the Parameters of Johnson S_B Distribution

We develop new expressions for the estimation of the parameters of the Johnson S_B distribution for the shortest path lengths in the OTNs. Out of the four parameters, λ can be estimated from either the mean of the distribution, or the square root of the convex area of the OTN using linear regression. However, such an expression results in large

errors for some small networks, and the networks with small mean nodal degrees. As this parameter represents the end point of the Johnson S_B distribution, its accuracy is very important in estimating the PDF and CDF of the distribution. That is why, we use the following estimation formulas for λ as shown in (6.11), obtained from two regressions.

$$\lambda_c \cong \begin{cases} 10.67(\sqrt{A_c})^{0.7564}, & \text{if } \sqrt{A_c} < 1000; \\ 0.0016A_c - 4.076\sqrt{A_c} + 5600, & \text{if } \sqrt{A_c} \geq 1000. \end{cases} \tag{6.11}$$

The shape parameter, δ, determines the modality of Johnson S_B distribution. Small changes in the values of δ around the critical point, which determines the modality ($\delta = \frac{1}{\sqrt{2}}$) changes the distribution to a large extent. Using the estimated values of σ_p and λ as described in [34], we found that their ratio (i.e., σ_p/λ) does not vary as it should; rather it is very much confined to a narrow range of values. Therefore the values of δ estimated in this method gives large errors.

Thus we looked for alternative methods to have good accuracy in the estimation of δ. The ratio of the square root of the convex area of the OTN ($\sqrt{A_C}$), to its number of nodes (N), was found to be a good parameter for the estimation of δ. We found that when this ratio (i.e., $\sqrt{A_C}/N$) is less than 70, the PDF of the distribution is unimodal, and when this ratio is larger than 70 the distribution can be either unimodal or bimodal. This ratio is less than 70 in 25 of the 40 real OTNs, and larger than 70 in the remaining 15. Accordingly, we developed two expressions for δ based on the values of $\sqrt{A_C}/N$ as shown in (6.12), in which, $t = \sqrt{A_C}/N$.

$$\delta_c \cong \begin{cases} cosec\left(\frac{t+30}{40}\right) + 0.25cos(62.46t) & \\ -0.1cos(112.62t), & \text{if } t < 70; \\ 0.8sin\left(\frac{t+126.5}{162}\right) + 0.17cos(50.258t) & \\ +0.12cos(149.226t), & \text{if } t \geq 70. \end{cases} \tag{6.12}$$

The location parameter, ζ, presents the starting point of the PDF of Johnson S_B distribution. For the 40 real OTNs, we observed that ζ is located around 0; but taking $\zeta = 0$, results in a large deviation in the overall PDF (the PDF of the shortest path lengths of the OTN is shifted to the right for the negative values of ζ, and vice versa). However, using the knowledge of δ_c, estimated in (6.12), helps in more accurate estimation of ζ. For all the bimodal PDFs $\zeta > 0$, and for the majority of the unimodal PDFs $\zeta < 0$. Accordingly, we developed expression (6.13) for the estimation of ζ, which depends on the values of $\sqrt{A_c}$, and δ_c.

Table 6.4: Performance evaluation of the estimated parameters for forty real optical transport networks. KSS_{JS_B} is the KSS obtained from optimized distribution parameters, and KSS'_{JS_B} is obtained from the estimated parameters. The CI for the KSS values and the acceptability is set at 0.95 (λ_c and ζ_c are in km, and A_c is in sq. km).

#	Network	N	L	A_c	KSS_{JS_B}	δ_c	γ_c	λ_c	ζ_c	KSS'_{JS_B}	Acceptable?
1	SANREN[23]	7	7	508348	0.0788	0.8496	0.0866	1535	-78.43	0.2841	Yes
2	Via NET[13]	9	12	1267876	0.0615	0.6978	0.2641	3039	112.60	0.1405	Yes
3	BREN[13]	10	11	42196	0.0520	1.2003	0.3394	598	-22.55	0.1513	Yes
4	RNP[13]	10	12	1814676	0.0546	0.7222	0.1973	3031	134.70	0.0997	Yes
5	LEARN[13]	10	12	183296	0.0686	0.7277	0.2654	1044	42.80	0.1054	Yes
6	Abilene Core[13]	10	13	5698507	0.0699	0.6295	0.2311	4987	238.70	0.1296	Yes
7	SINET[24]	11	13	406969	0.0483	1.0921	0.2614	1412	-70.18	0.0984	Yes
8	CompuServe[13]	11	14	5442148	0.0619	0.5991	0.2341	4799	233.30	0.0854	Yes
9	vBNS[13]	12	17	6279993	0.0585	0.5511	0.2338	5434	250.60	0.1384	Yes
10	CESNET[13]	12	19	45957	0.0601	1.1688	0.1940	618	-23.54	0.2997	No
11	AARNET[22]	13	13	7591320	0.0468	0.9192	0.2339	6515	-303.05	0.1001	Yes
12	FLRNET[30]	13	14	156097	0.0533	1.3498	0.1053	982	-43.45	0.1387	Yes
13	NSFNET[13]	14	21	6007605	0.0502	0.7712	0.2287	5222	245.10	0.2369	Yes
14	ITALY[13]	14	29	436577	0.0384	0.9473	0.2442	1450	-72.71	0.1193	Yes
15	HEANET[31]	15	15	55361	0.0302	1.0668	0.1953	663	-25.85	0.1121	Yes
16	MZIMA[13]	15	19	6805153	0.0530	0.8551	0.2301	5857	-286.99	0.0764	Yes
17	ACONET[13]	15	22	52826	0.0485	0.8928	0.1903	652	-25.30	0.1188	Yes
18	KAREN[28]	16	17	349015	0.0402	0.9861	0.1771	1332	-65.01	0.0748	Yes
19	BELNET[25]	16	18	17466	0.0312	1.4033	0.6750	429	-14.52	0.1144	Yes
20	ERNET[27]	16	18	2382858	0.0406	1.0002	-0.4290	3121	-169.84	0.1161	Yes
21	GARR-B[13]	16	27	476100	0.0571	0.9379	0.2580	1498	-75.90	0.1020	Yes
22	ARNES[13]	17	20	12641	0.0372	0.9393	0.7728	379	-12.32	0.0857	Yes
23	GERMANY[13]	17	26	196675	0.0373	1.1674	0.6475	1071	-48.73	0.0472	Yes
24	REDIRIS[13]	17	28	463539	0.0432	0.9518	0.2159	1483	-74.91	0.1097	Yes
25	CALREN[29]	19	22	168087	0.0337	0.7622	0.3649	1010	41.00	0.0643	Yes
26	Lambada Rail[13]	19	23	6899337	0.0316	0.8956	0.1978	5934	-288.97	0.0815	Yes
27	MEMOREX[13]	19	24	253850	0.0302	1.0799	0.6726	1181	-55.44	0.0840	Yes
28	CANARIE[13]	19	26	3462688	0.0283	0.5675	-0.0949	3556	186.10	0.0878	Yes
29	EON[13]	19	37	5800056	0.0433	0.9247	0.6414	5063	-264.88	0.0922	Yes
30	OPTOSunet[13]	20	24	122980	0.0313	0.9264	0.1795	898	-38.61	0.0438	Yes
31	Hibernia USA[13]	20	27	1257762	0.0360	1.3611	1.1888	3041	-123.31	0.0670	Yes
32	ARPANET[13]	20	32	5975012	0.0363	0.6313	0.2417	5195	244.40	0.0977	Yes
33	PIONIER[13]	21	25	194178	0.0255	1.1392	0.3820	1068	-48.51	0.0332	Yes
34	SANET[13]	25	28	24336	0.0410	1.4592	0.7604	486	-17.16	0.0770	Yes
35	FUNET[26]	26	27	210828	0.0290	0.8997	0.1889	1100	-50.49	0.0528	Yes
36	NEWNET[13]	26	31	7033452	0.0257	0.7927	0.0884	6043	265.20	0.2220	No
37	RENATER[13]	27	35	462096	0.0190	0.9080	0.6182	1481	-74.80	0.1813	No
38	BULGARIA[13]	32	33	95108	0.0167	1.2420	0.5842	814	-33.88	0.0454	Yes
39	LONI[13]	33	37	68008	0.0257	1.0442	0.6721	718	-28.71	0.1543	No
40	Metrona[13]	33	41	158837	0.0173	1.2814	0.3863	990	-43.89	0.2396	No
	Absolute Average:									0.0951	Yes

$$\zeta_c \approx \begin{cases} 0.1\sqrt{A_c}, & \text{if } \delta_c < 0.82; \\ -0.11\sqrt{A_c}, & \text{if } \delta_c \geq 0.82. \end{cases} \tag{6.13}$$

The second shape parameter, γ, can be estimated in two ways as presented in [34]: one using the values of μ_c and δ_c; and the other using the values of the median and δ_c (as shown in Appendix D). Here, μ_c is the transformed mean, obtained from $\langle p_c \rangle$, λ_c and ζ_c (i.e., $\mu_c = \frac{\langle p_c \rangle - \zeta_c}{\lambda_c}$). This first method, provides 21 'Acceptable' values (measured using KSS values) of γ out of 40 OTNs. A statistic is 'Acceptable' at a certain confidence level (or

significance level), if it is smaller than the corresponding rejection threshold. However, γ using the second method, which needs the estimated values of δ_c, and the median, m_c, provides 18 'Acceptable' KSS values out of 40. That is why, we look for alternative expressions for the estimation of γ. Using $\sqrt{A_C}$ and N, we developed better expression for γ_c as shown in (6.14), which provides 'Acceptable' KSS values for 35 OTNs at C.I. = 95%, when used along with (6.11), (6.12) and (6.13). These estimations are further explained in the next subsection.

$$\gamma_c \approx \begin{cases} 0.12 + |1.02 - 0.06t| + 0.03sin(4t), & \text{if } t < 29; \\ 0.235 + sinc(2t - 112.4), & \text{if } 29 \leq t \leq 96; \\ 0.24 - \frac{1}{t-95} + 0.45sinc(2t - 253), & \text{if } t > 96. \end{cases} \tag{6.14}$$

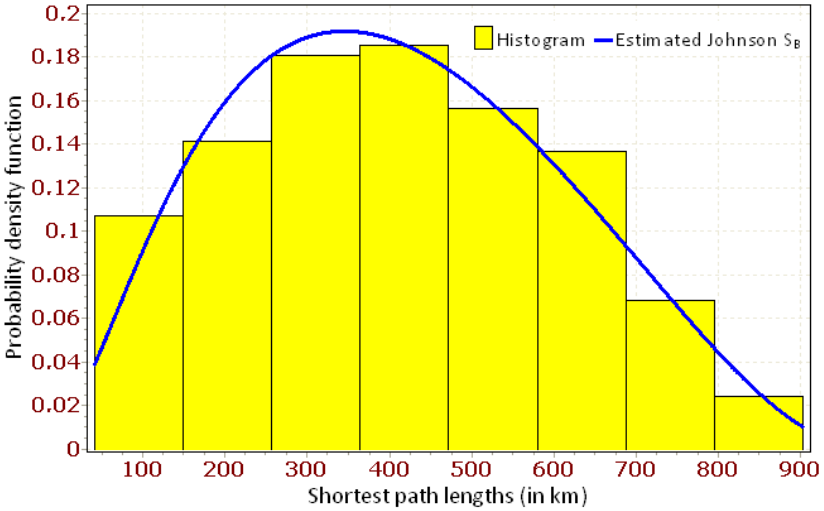

Figure 6.3: Estimated Johnson S_B distribution fitted with the path length histograms of Pionier network. The KSS for this fitting is: 0.0332.

6.4.2 Estimation of the Proposed Model

Using the parameters shown in Table 6.4, the PDF of Johnson S_B distribution for the shortest path lengths can be estimated. The location and scale parameters are not much changed from the optimized values shown in Table 6.2. The two shape parameters too are well within the acceptable ranges for 35 OTNs out of total 40. We estimate the PDFs of all the 40 OTNs and measure their accuracy in terms of the KSS values (i.e., KSS'_{JS_B}). These values can be found from the 11th column of Table 6.4. In the last column, we show whether it is acceptable or not? This acceptability is measured in terms of the KSS values shown. We show the the evaluation of all the fittings in Table

6.4. As can be seen from the last column (column 'Acceptable?') of Table 6.4, 35 OTNs have acceptable models at 95% CI. In Figure 6.3, we present the estimated shortest path distribution model for Pionier network. The KSS for this fitting is quite small, though it is a little higher than the optimum value shown in Figure 6.1 for the same network. The expressions developed for γ_c and δ_c are obtained after checking different functions and iterating their arguments for more than 10 million times.

6.5 Chapter Summary

Shortest path lengths between the node pairs of 40 real OTNs were found to be quite random, and to broadly follow the Johnson S_B distribution. This distribution is considered a better choice than other distributions, because in addition to yielding low KSS values in all OTNs, it also accommodates both unimodal and bimodal PDFs, which are commonly observed in the OTNs. The bimodality is found in case of the networks having large convex areas with small number of nodes. We have also shown that the mean, median and standard deviation of the shortest path lengths of OTNs can be estimated from their convex areas. The expressions used for estimation of these parameters are found to be linear. Moreover, these key metrics of the shortest paths (mean, median, standard deviation, and upper bound on the maximum) can be estimated without full knowledge of the network topology. This is because, the convex area can be computed simply from node locations.

References

[1] J. Hershberger, and S. Suri, "Vickrey prices and shortest paths: What is an edge worth?," In Proceedings of 42nd IEEE Symposium on Foundations of Computer Science, pp. 252 – 259, Oct. 2001.

[2] S. K. Korotky, "Network Global Expectation Model: A Statistical formalism for Quickly Quantifying Network Needs and Costs," *IEEE/OSA Journal of Lightwave Technology*, vol. 22, no. 3, pp. 703 – 722, Mar. 2004.

[3] J. F. Labourdette, E. Bouillet, R. Ramamurthy, and A. A. Akyamaç, "Fast approximate dimensioning and performance analysis of mesh optical networks," *IEEE/ACM Journal of Transactions on Networking*, vol. 3, no. 4, pp. 906 – 917, Aug. 2005.

[4] A. Gubichev, S. Bedathur, S. Seufert, and G. Weikum, "Fast and accurate estimation of shortest paths in large graphs," in Proceedings of ACM Conference CIKM, pp. 906 – 917, Oct. 2010.

[5] M. Potamias, F. Bonchi, C. Castillo, and A. Gionis, "Fast shortest path distance estimation in large networks," in Proceedings of IEEE/ACM CKIM Conference, vol. 3, no. 4, pp. 906 – 917, Nov. 2009.

[6] S. Lee, and S. Sahu, "Accuracy of path length estimation via network distance based coordinate systems," in Proceedings of IEEE ICUFN Conference, pp. 906 – 917, Aug. 2010.

[7] C. Bauckhage, K. Kersting, B. Rastegarpanah, "The Weibull as a Model of Shortest Path Distributions in Random Networks," in Proceedings of International Workshop on Mining and Learning with Graphs, Chicago, IL, USA, ACM, Aug. 2013.

[8] Y. Yamada, H. Hasegawa, and K.-I. Sato, "Survivable hierarchial optical path network design with dedicated wavelength path protection," *IEEE/OSA Journal of Lightwave Technology*, vol. 29, no. 21, pp. 3196 – 3209, Nov. 2011.

[9] A. N. Pinto, C. Pavan, and R. M. Morais, "Dimensioning optical networks: A practical approach," in Proceedings of International Conference on Transparent Networks (ICTON), Munich, Germany, pp. 1 – 4, June 2010 .

[10] A. N. Pinto, C. Pavan, R. M. Morais, and A. Correia, "Cost evaluation in optical networks," in Proceedings of International Conference on Transparent Networks (ICTON), Stockholm, Sweden, pp. 1 – 4, June 2011 .

[11] C. Pavan, R. M. Morais, and A. N. Pinto, "Estimation of CapEx in Survivable Optical Transport Networks," in Proceedings of European Conference on Networks and Optical Communications (NOC), Faro, Portugal, Vol. 1, pp. 263 – 268, June, 2010.

[12] A. N. Pinto, C. Pavan, and R. M. Morais, "A Statistical Model to CAPEX Fast Calculation in Optical Transport Networks," in Proceedings of International Conf. on Transparent Networks (ICTON), Ilha de São Miguel - Açores, Portugal, Vol. 1, pp. 1 – 4, June, 2009.

[13] S. K. Routray, R. M. Morais, J. R. F. da Rocha, and A. N. Pinto, "Statistical

Model for Link Lengths in Optical Transport Networks," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 5, no. 7, pp. 762 – 773, Jul. 2013.

[14] S. K. Routray, G. Sahin, J. R. F. da Rocha, and A. N. Pinto, "Estimation of Link-Dependent Parameters in Optical Transport Networks from Statistical Models," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 6, no. 7, pp. 601 – 609, July 2014.

[15] S. K. Routray, J. R. F. da Rocha, and A. N. Pinto, "Estimating the Parameters of Optical Transport Networks from their Circumferential Ellipses," in Proceedings of 21st IEEE International Conference on Telecommunications (ICT), Lisbon, 4 – 7 May 2014.

[16] S. K. Routray, G. Sahin, J. R. F. da Rocha, and A. N. Pinto, "Statistical Analysis and Modeling of Shortest Path Lengths in Optical Transport Networks," *IEEE/OSA Journal of Lightwave Technology*, accepted in Mar. 2015.

[17] C. Pavan, R. M. Morais, J. R. F. da Rocha, and A. N. Pinto, "Generating realistic optical transport network topologies," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 2, no. 1, pp. 80 – 90, Jan. 2010.

[18] R. M. Morais, C. Pavan, C. Requejo, and A. N. Pinto, "Genetic algorithm for the topological design of survivable optical transport networks," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 3, no. 1, pp. 17 – 26, Jan. 2011.

[19] S. Knight, H. X. Nguyen, N. Falkner, R. Bowden, and M. Roughan, "The Internet topology zoo," *IEEE Journal on Selected Areas in Communication*, vol. 29, no. 9, pp. 1765 – 1775, Sep. 2011.

[20] Reference Optical Networks, Apr. 2013 [Online]. Available: http://www.av.it.pt/anp/on/refnet2.html.

[21] The Internet Topology Zoo of Networks, Apr. 2013 [Online]. Available: http://www.topology-zoo.org/.

[22] The Australian Research and Education Network (AARNET). [Online]. Available: http://www.topology-zoo.org/.

[23] The South African National Research Network (SANReN). [Online]. Available: http://www.sanren.ac.za/design/backbone/.

[24] The Japanese Science Information Network (SINET). [Online]. Available: http://www.sinet.ad.jp/Network/.

[25] The Belgian National Research Network (BELNET). [Online]. Available: http://www.belnet.be/en/about-us/belnet-network.

[26] The Finnish University and Research Network (FUNET). [Online]. Available: http://www.csc.fi/hallinto/funet/esittely/runkoverkko.

[27] The Indian Educational and Reserch Network (ERNET). [Online]. Available: http://www.topology-zoo.org/.

[28] The Kiwi Advanced Research and Education Network (KAREN). [Online]. Available: http://www.topology-zoo.org/.

[29] The California Research and Education Network (CALREN). [Online]. Available: http://www.cenic.org/page_id=66/.

[30] The Florida LambdaRail Network (FLRNET). [Online]. Available: http://www.flrnet.org/?page_id=247.

[31] The Iris National Education and Research Network (HEANET). [Online]. Available: http://www.heanet.ie/about.

[32] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Vol. 1*, Chichester, UK: Wiley-Interscience, 1994.

[33] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Vol. 2*, Chichester, UK: Wiley-Interscience, 1995.

[34] N. L. Johnson, "Systems of frequency curves generated by methods of translation," *Biometrika*, vol. 36, no. 1/2, pp. 149 – 176, Jun. 1949.

[35] N. L. Johnson, and J. O. Kitchen, "Some notes on tables to facilitate fitting S_B curves," *Biometrika*, vol. 58, no. 1, pp. 223 – 226, Apr. 1971.

[36] N. L. Johnson, and J. O. Kitchen, "Tables to facilitate fitting S_B curves II: both terminals known," *Biometrika*, vol. 58, no. 3, pp. 149 – 176, Dec. 1971.

[37] N. L. Johnson, S. Kotz, and C.B. Read, *Encyclopedia of Statistical Sciences*, New York, NY: John Wiley & Sons, 1989.

[38] S. Kotz, and J. R. van Dorp, *Beyond Beta: Other Continuous Families of Distributions with Bounded Support and Applications*, Singapore: World Scientific Pub Co Inc, 2004.

[39] M. R. Flynn, "Fitting human exposure data with the Johnson distribution," *Journal of Exposure Science and Environmental Epidemiology*, vol. 16, no. 1, pp. 56 – 62, Jan. 2006.

[40] T. F. Fonseca, C. P. Marques, and B. R. Parresol, "Describing Maritime Pine Diameter Distributions with Johnson's SB Distribution Using a New All-Parameter Recovery Approach," *Forest Science*, vol. 55, no. 4, pp. 367 – 373, Aug. 2009.

[41] A. Mateus and M. Tomé, "Estimating the parameters of the Johnson's S_B distribution using an approach of method of moments," in Proceedings of ICNAAM, pp. 1483 – 1485, Sep. 2011.

CHAPTER 7

Estimation of the Shortest Path Related Parameters

7.1 Introduction

SHORTEST PATH LENGTHS in optical transport networks (OTNs) are required for many applications such as traffic management, routing, restoration, protection, control and several other network management related operations. Estimations related to the shortest path lengths are also quite important in the early stage planning and dimensioning of the OTNs. In this chapter, we apply the statistical model for the shortest path lengths developed in the previous chapter to 40 real OTNs. We estimate the modulation formats required for the shortest paths between the node pairs assuming the network to be transparent. We evaluate the outcome obtained from the statistical model by comparing it with the exact estimations, which use the exact shortest path lengths between the node pairs of the OTNs.

7.1.1 Motivation

In Chapter 6, we described the importance of shortest path lengths in various fields of science and engineering. Information of the shortest path lengths are essential for several functions of the OTNs. Using realistic models, this shortest path data can be used for the estimation of associated components of CAPEX, OPEX and MANEX of the OTNs. In the modern OTNs, there are several complexities, and the changes in network

traffic dynamics are quite frequent. In such changing conditions, it is always desirable to have a statistical model for the shortest path lengths. In [1] – [5], several cases are presented in which the dimensioning and estimation of OTN parameters are required without complete information. Estimation of some of those parameters can be done using the shortest path model presented in Chapter 6. All these issues, points towards the usefulness and applications of the statistical model developed in the previous chapter.

7.1.2 Related Work

Several empirical and semi-empirical expressions are proposed for the fast estimation of the parameters of the OTNs in [1]. Expressions in [1] depend on the basic parameters of the networks such as the number of nodes, number of links and coverage area. In [2], fast approximate dimensioning and performance analysis of optical network topologies are presented. In this work, path lengths are analyzed and an approximate expression for the average number of hops is derived. In [6] a statistical model for link lengths is provided, which depends only on the convex area (bounded interior of the smallest convex set comprising of all the nodes) and the number of nodes of the OTNs. This model can be used to estimate the link-dependent parameters of the OTNs [7]. In [8] a Weibull model is presented for the shortest path distributions in random networks. Shortest path lengths are studied in many disciplines of science and engineering where they have several important utilities [9]. Real OTN topologies used for this study on the application of shortest path model can be found in [10]. Several practical applications of Johnson S_B distribution are presented in [11] – [15].

7.1.3 Chapter Outline

Rest of this chapter is organized in 3 different sections. In section 7.2, we present the general applicability issues of the shortest paths, and the role of the model developed in the previous chapter. In this section, we explain the utilities of the developed model. In section 7.3, we applied the model to determine the modulation formats needed in transparent optical networks. In section 7.4, we present the concluding remarks, and the effectiveness of the proposed method. The work presented in this chapter is reported in [16].

7.2 Utility of Shortest Path Length Models

In Chapter 6, we have described the importance of shortest path lengths in different disciplines of science, engineering and social sciences. In this section, we provide a list of

popular shortest path algorithms and their applications in different fields. Shortest path related problems are not very new. However, their significance have been broadened since the arrival of the digital communication networks in the 1950s [17]. That is clear from the following chronological list of shortest path algorithms which are always being tested for different practical applications in networks.

1. 1955: Shimbel algorithm [18] was proposed mainly for information networks.

2. 1958: Bellman-Ford algorithm (also known as Bellman-Ford-Moore algorithm) [19] – [21] was proposed for random networks. However, the main aim of this algorithm was to find an economical solution for transportation. Later in Moore's version it was applied to the routing of long distance telephone calls.

3. 1959: Dijkstra's algorithm [22] was proposed for the efficient routing in telecommunication networks. This is a simpler and faster version of Bellman-Ford algorithm.

4. 1983: Gabow algorithm [23] was proposed for the large telecommunication networks. This algorithm reduced the computation time and resources in finding the shortest paths with respect to its predecessors.

5. 1989: Gabow and Tarjan algorithm [24] was proposed to improve the previous version of the Gabow algorithm.

6. 1993: Goldberg algorithm [25] was proposed for large computer networks such as the Internet. It is also used in telecommunications and search algorithms.

7. 2005: Sankowski, Yuster and Zwick algorithm [26] was proposed for the improvement of the computation times of the shortest paths in large graphs. Now it is applied by several web based firms and telecommunication companies.

These algorithms have different levels of efficiency in terms of the estimation time of the shortest paths. These algorithms are helpful for the routing and traffic management related functions. However, there are several operations in which the knowledge of all the shortest paths is required. Shortest path trees for the large backbone networks are designed using Dijkstra's algorithm [17]. The computation of the tree needs complete information. It is thus useful only when the whole topology is known. For the early stage planning and design related functions, the shortest paths are to be estimated without the information of the complete topology. In such cases, the distribution of the shortest path lengths is essential for the estimations.

Table 7.1: Basic attributes, exact (N, L, P and $\langle p \rangle$) and estimated ($\langle p_c \rangle$, δ_c, γ_c, λ_c and ζ_c) parameters of 40 real OTNs ($\langle p \rangle$, $\langle p_c \rangle$, λ_c and ζ_c are in km).

#	Network	N	L	P	$\langle p \rangle$	$\langle p_c \rangle$	δ_c	γ_c	λ_c	ζ_c
1	SANREN[10]	7	7	21	874	678	0.8496	0.0866	1535	-78.43
2	Via NET[6]	9	12	36	1211	1049	0.6978	0.2641	3039	112.60
3	BREN[6]	10	11	45	210	221	1.2003	0.3394	598	-22.55
4	RNP[6]	10	12	45	1558	1247	0.7222	0.1973	3031	134.70
5	LEARN[6]	10	12	45	464	421	0.7277	0.2654	1044	42.80
6	Abilene Core[6]	10	13	45	2204	2182	0.6295	0.2311	4987	238.70
7	SINET[10]	11	13	55	580	610	1.0921	0.2614	1412	-70.18
8	CompuServe[6]	11	14	55	2420	2134	0.5991	0.2341	4799	233.30
9	vBNS[6]	12	17	66	2351	2289	0.5511	0.2338	5434	250.60
10	CESNET[6]	12	19	66	194	229	1.1688	0.1940	618	-23.54
11	AARNET[10]	13	13	78	2481	2513	0.9192	0.2339	6515	-303.05
12	FLRNET[10]	13	14	78	422	392	1.3498	0.1053	982	-43.45
13	NSFNET[6]	14	21	91	2304	2240	0.7712	0.2287	5222	245.10
14	ITALY[6]	14	29	91	524	631	0.9473	0.2442	1450	-72.71
15	HEANET[10]	15	15	105	297	248	1.0668	0.1953	663	-25.85
16	MZIMA[6]	15	19	105	2419	2382	0.8551	0.2301	5857	-286.99
17	ACONET[6]	15	22	105	311	244	0.8928	0.1903	652	-25.30
18	KAREN[10]	16	17	120	555	568	0.9861	0.1771	1332	-65.01
19	BELNET[10]	16	18	120	139	155	1.4033	0.6750	429	-14.52
20	ERNET[10]	16	18	120	1713	1425	1.0002	-0.4290	3121	-169.84
21	GARR-B[6]	16	27	120	549	657	0.9379	0.2580	1498	-75.90
22	ARNES[6]	17	20	136	111	137	0.9393	0.7728	379	-12.32
23	GERMANY[6]	17	26	136	351	435	1.1674	0.6475	1071	-48.73
24	REDIRIS[6]	17	28	136	600	649	0.9518	0.2159	1483	-74.91
25	CALREN[10]	19	22	171	466	405	0.7622	0.3649	1010	41.00
26	Lambada Rail[6]	19	23	171	2305	2398	0.8956	0.1978	5934	-288.97
27	MEMOREX[6]	19	24	171	396	490	1.0799	0.6726	1181	-55.44
28	CANARIE[6]	19	26	171	1928	1709	0.5675	-0.0949	3556	186.10
29	EON[6]	19	37	171	1495	2201	0.9247	0.6414	5063	-264.88
30	OPTOSunet[6]	20	24	190	380	352	0.9264	0.1795	898	-38.61
31	Hibernia USA[6]	20	27	190	783	1044	1.3611	1.1888	3041	-123.31
32	ARPANET[6]	20	32	190	2339	2233	0.6313	0.2417	5195	244.40
33	PIONIER[6]	21	25	210	409	433	1.1392	0.3820	1068	-48.51
34	SANET[6]	25	28	300	174	177	1.4592	0.7604	486	-17.16
35	FUNET[10]	26	27	325	441	449	0.8997	0.1889	1100	-50.49
36	NEWNET[6]	26	31	325	2303	2420	0.7927	0.0884	6043	265.20
37	RENATER[6]	27	35	351	543	648	0.9080	0.6182	1481	-74.80
38	BULGARIA[6]	32	33	496	294	314	1.2420	0.5842	814	-33.88
39	LONI[6]	33	37	528	288	271	1.0442	0.6721	718	-28.71
40	Metrona[6]	33	41	528	293	395	1.2814	0.3863	990	-43.89

7.2.1 Utilities in OTNs

OTNs, like other kinds of graphs and networks have several topology-dependent properties. Shortest path lengths are key topological parameters, and they determine several operations of OTNs as mentioned in the first section of this chapter. In the case of transparent networks, the optical signals of the lightpaths are not processed at each node; rather they are allowed to pass through the intermediate nodes with only 1R processing. In such cases, the lightpaths are often established along the shortest paths. For the translucent networks too, along the transparent paths (where, none of the intermediate nodes is opaque), similar procedures are followed. The statistical model (for shortest path lengths) can be used for the estimation of the associated shortest path dependent parameters. In Table 7.1, we present the parameters of the 40 real OTNs

and the associated Johnson S_B distribution parameters which are applied in the next section. These parameters have already been defined in Chapter 6.

7.3 Estimation of Path-dependent Parameters in OTNs

We apply the developed model for the shortest path lengths in OTNs to show their utility in the practical situations. We use the model to estimate the type of modulation schemes required at the ends of the shortest paths between the node pairs, as required in transparent OTNs. In [7], the types of modulation schemes needed at the ends of the links are estimated using the statistical model of link lengths. However, the approach in [7], is suitable for the opaque networks. In case of the transparent networks, we need the shortest path lengths instead of link lengths for these estimations.

We estimate the number and types of modulation schemes needed for the 40 networks used in this study, assuming them to be transparent with three different methods. In the first method, we use the exact shortest path lengths for this calculation using the half distance law described in [27] and [28]. This is the exact estimation as it uses the exact shortest path lengths. In the second method, we use the statistical model developed for the shortest path lengths in the previous chapter and the half distance law. This method needs just the node locations (or alternatively, the convex area and the number of nodes). In the third method, we use just the estimated average shortest path length and the same half distance law. The third method uses the same information as the second.

The number of shortest paths corresponding to each type of modulation format can be estimated by applying the half distance law proposed in [27] and [28], to the Johnson S_B shortest path length distribution model. According to that law, the modulation format must be changed when the length of the span (initial span was chosen to be 375 km in [27]) is doubled as shown in Figure 7.1. For shortest paths with lengths of 0 – 375 km, 375 – 750 km, 750 – 1500 km, and more than 1500 km, the modulations are chosen to be 16QAM, 8QAM, QPSK and BPSK respectively. In our case, we made a small change in the lower limit of 16QAM (in [27] it was 0 – 375 km), and the upper limit of BPSK (in [27] it was 1500 – 3000 km). We changed the lower and upper limits of the shortest path lengths to ζ_c and $\lambda_c - \zeta_c$, in place of '0' and '3000', respectively. Accordingly, the number of shortest paths that need 16QAM, 8QAM, QPSK and BPSK modulation formats in a network are respectively given by $N_{16QAM}(= NS)$, $N_{8QAM}(= NE)$, $N_{QPSK}(= NQ)$ and $N_{BPSK}(= NB)$ as shown in (7.1) – (7.4). In these equations, $\lfloor a \rceil$, represents the rounded value of a (i.e., the nearest integer of a). The outcomes of these equations for the real

OTNs are shown in Table 7.2.

Table 7.2: Selection of modulation formats for 40 real OTNs using the exact information (column 'UESPL'), the statistical model for the shortest path lengths (column 'USPLD'), and the average shortest path lengths (column 'UASPL'). S, E, Q, and B stand for 16QAM, 8QAM, QPSK, and BPSK, respectively.

#	Network	$UESPL$	$USPLD$	$UASPL$	C_D	C_L	$E_{USPLD}(\%)$	$E_{UASPL}(\%)$
1	SANREN[10]	3S – 6E – 11Q – 1B	5S – 7E – 9Q – 0B	21E	18	6	14.2857	71.4286
2	Via NET[6]	4S – 8E – 10Q – 14B	3S – 6E – 11Q – 16B	36Q	33	10	8.3333	72.2222
3	BREN[6]	43S – 2E – 0Q – 0B	39S – 6E – 0Q – 0B	45S	41	43	8.8889	4.4444
4	RNP[6]	4S – 8E – 11 Q – 22B	3S – 7E – 14Q – 22B	45Q	43	11	4.4444	75.5556
5	LEARN[6]	18S – 20E – 7Q – 0B	17S – 18E –9Q – 0B	45E	42	20	6.6667	55.5556
6	Abilene Core[6]	0S – 5E – 10Q – 30B	1S – 5E – 9Q – 30B	45B	44	10	2.2222	77.7778
7	SINET[10]	19S – 21E – 15Q – 1B	15S – 25E – 15Q – 0B	55E	50	21	7.2727	61.8182
8	CompuServe[6]	3S – 7E – 11Q – 34B	2S – 7E – 11Q – 36B	55B	53	11	1.8182	80.0000
9	vBNS[6]	4S – 6E – 13Q – 43B	2S – 8E – 12Q – 44B	66B	63	6	4.5455	90.9091
10	CESNET[6]	61S – 5E – 0Q – 0B	54S – 12E –0Q – 0B	66S	59	61	10.6061	7.5758
11	AARNET[10]	3S – 8E – 16Q – 51B	3S – 5E – 12Q – 58B	78B	71	16	8.9744	79.4872
12	FLRNET[10]	36S – 35E – 7Q – 0B	30S – 47E – 2Q – 0B	78E	67	35	14.1026	55.1282
13	NSFNET[6]	2S – 10E – 15Q – 64B	0S – 6E – 17Q – 68B	91B	85	64	6.5934	29.6703
14	ITALY[6]	31S – 41E – 19Q – 0B	27S – 35E – 28Q – 0B	91E	80	41	10.9890	54.9451
15	HEANET[10]	71S – 34E – 0Q – 0B	78S – 27E – 0Q – 0B	105S	97	71	6.6667	32.3810
16	MZIMA[6]	4S – 6E – 21Q – 74B	7S – 8E – 19Q – 72B	105B	101	74	3.8095	29.5238
17	ACONET[6]	69S – 36E – 0Q – 0B	76S – 29E – 0Q – 0B	105S	98	69	6.6667	34.2857
18	KAREN[10]	39S – 51E – 30Q – 0B	36S – 52E – 32Q – 0B	120E	116	51	2.5000	57.5000
19	BELNET[10]	120S – 0E – 0Q – 0B	120S – 0E – 0Q – 0B	120S	120	120	0.0000	0.0000
20	ERNET[10]	6S – 12E – 27Q – 75B	3S – 9E – 35Q – 74B	120Q	112	27	5.8333	77.5000
21	GARR-B[6]	36S – 50E – 34Q – 0B	36S – 45E – 39Q – 0B	120E	114	50	4.1667	58.3333
22	ARNES[6]	136S – 0E – 0Q – 0B	136S – 0E – 0Q – 0B	136S	136	136	0.0000	0.0000
23	GERMANY[6]	76S – 57E – 3Q – 0B	76S – 56E – 4Q – 0B	136E	132	57	0.7353	58.0882
24	REDIRIS[6]	29S – 67E – 40Q – 0B	38S – 52E – 45Q – 0B	136E	119	67	11.0294	50.7353
25	CALREN[10]	73S – 69E – 29Q – 0B	74S – 71E – 26Q – 0B	171E	165	69	1.7543	59.6491
26	Lambada Rail[6]	9S – 15E – 32Q – 115B	8S – 12E – 30Q – 121B	171B	164	115	3.5088	32.7485
27	MEMOREX[6]	86S – 79E – 6Q – 0B	90S – 69E – 12Q – 0B	171E	161	79	5.8480	53.8012
28	CANARIE[6]	10S – 19E – 36Q – 106B	7S – 18E – 34Q – 112B	171B	165	106	3.5088	38.0117
29	EON[6]	9S – 26E – 58Q – 78B	22S – 23E – 45Q – 81B	171B	155	78	9.3567	54.3860
30	OPTOSunet[6]	96S – 93E – 1Q – 0B	98S – 88E – 4Q – 0B	190S	185	96	2.6316	49.4737
31	Hibernia USA[6]	40S – 58E – 77Q – 15B	29S – 63E – 83Q – 16B	190Q	180	77	5.7895	59.4737
32	ARPANET[6]	4S – 15E – 39Q – 132B	4S – 20E – 37Q – 130B	190B	186	132	2.1053	30.5263
33	PIONIER[6]	94S – 107E – 9Q – 0B	97S – 102E – 11Q – 0B	210E	203	107	2.8310	49.0476
34	SANET[6]	293S – 7E – 0Q – 0B	299S – 1E – 0Q – 0B	300S	294	293	2.0000	2.3333
35	FUNET[10]	125S – 149E – 51Q – 0B	133S – 146E – 46Q – 0B	325E	314	149	2.4615	54.1538
36	NEWNET[6]	11S – 23E – 66Q – 225B	0S – 10E – 42Q – 273B	325B	278	225	14.7692	30.7692
37	RENATER[6]	96S – 178E – 77Q – 0B	157S – 123E – 72Q – 0B	351E	293	178	17.0940	49.2877
38	BULGARIA[6]	354S – 140E – 2Q – 0B	359S – 137E – 0Q – 0B	496S	495	354	1.0080	28.6290
39	LONI[6]	384S – 144E – 0Q – 0B	435S – 93E – 0Q – 0B	528S	477	384	9.6591	27.2727
40	Metrona[6]	371S – 154E – 3Q – 0B	262S – 259E – 8Q – 0B	528E	415	154	20.6440	70.8333
	Average:						6.3920	47.6316

$$N_{16QAM} = NS = \left\lfloor P \int_{\zeta_c}^{375} f(p;\delta,\gamma,\lambda,\zeta)dp \right\rfloor \tag{7.1}$$

$$N_{8QAM} = NE = \left\lfloor P \int_{375}^{750} f(p;\delta,\gamma,\lambda,\zeta)dp \right\rfloor \tag{7.2}$$

$$N_{QPSK} = NQ = \left\lfloor P \int_{750}^{1500} f(p;\delta,\gamma,\lambda,\zeta)dp \right\rfloor \tag{7.3}$$

$$N_{BPSK} = NB = \left\lfloor P \int_{1500}^{\lambda_c - \zeta_c} f(p; \delta, \gamma, \lambda, \zeta) dp \right\rfloor \tag{7.4}$$

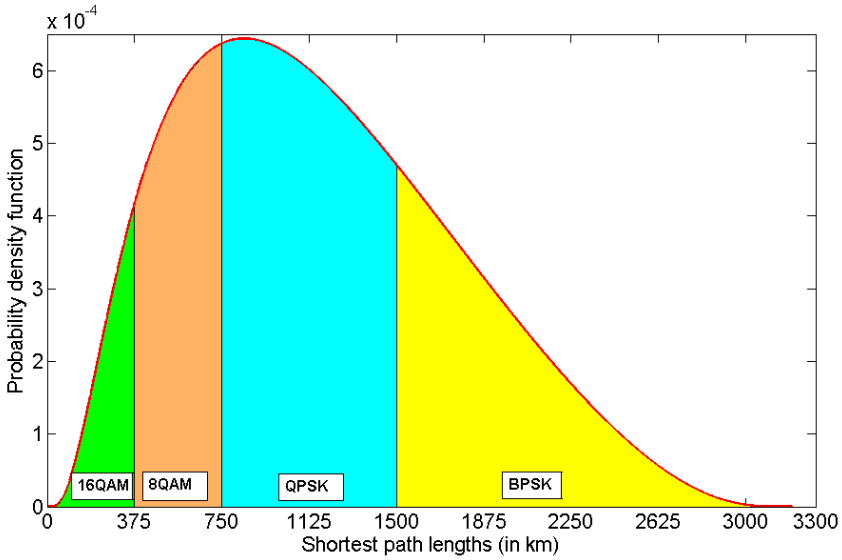

Figure 7.1: An optical transport network with average link length 930 km, and Johnson S_B distribution parameters γ=0.7010, δ=1.1030, λ = 3200, and ζ = 2.67. It shows the shortest path probabilities in different intervals of shortest path lengths according to the half distance law proposed in [27].

We evaluate the accuracy of the second and third methods by comparing their outcomes with the outcome of the first method. The basis for the evaluation of performances of the estimations using the second and the third methods is the number of correctly predicted modulation formats. We denote by C, the number of shortest paths for which the modulation formats have been estimated correctly by USPLD or UASPL with respect to UESPL. In Table 7.2, we denote the values of C obtained from USPLD and UASPL as C_D, and C_L, respectively.

$$C = \sum_{Z \epsilon \{S,E,Q,B\}} Min[N_{Z/UESPL}, N_{Z/OM}]. \tag{7.5}$$

The values of C are estimated according to equation (7.5). *Min*[ ], in (7.5), is the minimum selecting function, whose output is the smallest number among its arguments in [ ]. Z ϵ $\{S, E, Q, B\}$, and $N_{Z/UESPL}$ is the number under column UESPL in the Z category in Table 7.2. $N_{Z/OM}$, is the corresponding number in the Z category in other methods (OM can be either USPLD or UASPL). We estimate the correct predictions in each modulation category (i.e., S, E, Q and B) and add them together as shown in (7.5) to get the total number of correct predictions. The comparison is then evaluated by the

percentage of error in the estimation using (7.6). We show these results in Table 7.2, in which E_{USPLD} and E_{UASPL} represent the errors in USPLD and UASPL, respectively.

$$Error\ (\%) = \frac{P - C}{P} \times 100. \tag{7.6}$$

It can be observed in Table 7.2, the estimations using the distribution (USPLD) are more accurate than the estimations using just the average shortest path length (UASPL). The average error in the case of USPLD is 6.39%, while for UASPL it is 47.63%.

7.4 Chapter Summary

We applied the Johnson S_B distribution model obtained for the shortest path lengths between the node pairs of 40 real OTNs to the practical problem of determining the type of modulation formats required for different shortest paths. This model is found to be quite accurate for this estimation, as the percentage of error estimated with respect to the exact calculations are found to be quite low. As presented in this chapter, the average error for the correctly estimated number of modulation formats is just 6.4%. It is also noteworthy that the maximum error obtained in this case is lower than 21%. Overall, considering the information required for these estimations (i.e., convex area and number of nodes), the results are quite good and can be used for the cases of early stage planning and dimensioning.

References

[1] S. K. Korotky, "Network Global Expectation Model: A Statistical formalism for Quickly Quantifying Network Needs and Costs," *IEEE/OSA Journal of Lightwave Technology*, vol. 22, no. 3, pp. 703 – 722, March 2004.

[2] J. F. Labourdette, E. Bouillet, R. Ramamurthy, and A. A. Akyamaç, "Fast approximate dimensioning and performance analysis of mesh optical networks," *IEEE/ACM Journal of Transactions on Networking*, vol. 3, no. 4, pp. 906 – 917, Aug. 2005.

[3] A. N. Pinto, C. Pavan, and R. M. Morais, "Dimensioning optical networks: A practical approach," in Proceedings of of International Conference on Transparent Networks (ICTON), Munich, Germany, pp. 1 – 4, June 2010 .

[4] A. N. Pinto, C. Pavan, R. M. Morais, and A. Correia, "Cost evaluation in optical networks," in Proceedings of of International Conference on Transparent Networks (ICTON), Stockholm, Sweden, pp. 1 – 4, June 2011 .

[5] C. Pavan, R. M. Morais, and A. N. Pinto, "Estimation of CapEx in Survivable Optical Transport Networks," in Proceedings of of European Conference on Networks and Optical Communications (NOC), Faro, Portugal, Vol. 1, pp. 263 – 268, June, 2010.

[6] S. K. Routray, R. M. Morais, J. R. F. da Rocha, and A. N. Pinto, "Statistical Model for Link Lengths in Optical Transport Networks," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 5, no. 7, pp. 762 – 773, July 2013.

[7] S. K. Routray, G. Sahin, J. R. F. da Rocha, and A. N. Pinto, "Estimation of Link-Dependent Parameters in Optical Transport Networks from Statistical Models," *IEEE/OSA Journal of Optical Communication and Networking*, vol. 6, no. 6, pp. 601 – 609, June 2014.

[8] C. Bauckhage, K. Kersting, B. Rastegarpanah, "The Weibull as a Model of Shortest Path Distributions in Random Networks," in Proceedings of ACM International Workshop on Mining and Learning with Graphs, Chicago, IL, USA, Aug. 2013.

[9] J. Hershberger, and S. Suri, "Vickrey prices and shortest paths: What is an edge worth?," In Proceedings of 42nd IEEE Symposium on Foundations of Computer Science, pp. 252 – 259, Oct. 2001.

[10] Reference Optical Networks, Apr. 2013 [Online]. Available: http://www.av.it.pt/anp/on/refnet2.html.

[11] N. L. Johnson, "Systems of frequency curves generated by methods of translation," *Biometrika*, vol. 36, no. 1/2, pp. 149 – 176, Jun. 1949.

[12] S. Kotz, and J. R. van Dorp, *Beyond Beta: Other Continuous Families of Distributions with Bounded Support and Applications*, Singapore: World Scientific, 2004.

[13] M. R. Flynn, "Fitting human exposure data with the Johnson distribution," *Journal of Exposure Science and Environmental Epidemiology*, vol. 16, no. 1, pp. 56 – 62, Jan. 2006.

[14] T. F. Fonseca, C. P. Marques, and B. R. Parresol, "Describing Maritime Pine Diameter Distributions with Johnson's SB Distribution Using a New All-Parameter Recovery Approach," *Forest Science*, vol. 55, no. 4, pp. 367 – 373, Aug. 2009.

[15] A. Mateus and M. Tomé, "Estimating the parameters of the Johnson's S_B distribution using an approach of method of moments," in Proceedings of ICNAAM, pp. 1483 – 1485, Sep. 2011.

[16] S. K. Routray, G. Sahin, J. R. F. da Rocha, and A. N. Pinto, "Statistical Analysis and Modeling of Shortest Path Lengths in Optical Transport Networks," *IEEE/OSA Journal of Lightwave Technology*, accepted in Mar. 2015.

[17] A. Schrijver, "On the history of the shortest path problem," *Documenta Mathematica*, pp. 155 – 167, 2012.

[18] A. Shimbel, "Structural parameters of communication networks," *Bulletin of Mathematical Biophysics*, vol. 15, pp. 501 – 507, 1953.

[19] L. R. Ford, Jr, "Network Flow Theory," *Paper P–923*, The RAND Corporation, Santa Monica, California, Aug. 1956.

[20] R. Bellman, "On a routing problem," *Quarterly of Applied Mathematics*, vol. 16, pp. 87 – 90, 1958.

[21] E. F. Moore, "The shortest path through a maze," in Proceedings of International Symposium on the Theory of Switching, Apr. 1957, Part II (in The Annals of the Computation Laboratory of Harvard University Volume XXX, *Ed. H. Aiken*), Cambridge, MA: Harvard University Press, pp. 285 – 292, 1959.

[22] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, pp. 269 – 271, 1959.

[23] H. N. Gabow, R. E. Tarjan, "A linear-time algorithm for a special case of disjoint set union," in Proceedings of the 15th Annual ACM Symposium on Theory of Computing, pp. 246 – 251, Dec. 1983.

[24] H. N. Gabow, R. E. Tarjan, "Faster scaling algorithms for network problems," *SIAM Journal on Computing*, vol. 18, no. 5, pp. 1013 – 1036, 1989.

[25] A. V. Goldberg, "Scaling algorithms for the shortest paths problem," In Proceedings of the fourth annual ACM-SIAM Symposium on Discrete algorithms, pp. 222 – 231, Jan. 1993.

[26] P. Sankowski, "Shortest paths in matrix multiplication time," in *Algorithms – ESA*, pp. 770 – 778, Berlin, Germany: Springer, 2005.

[27] K. Chirstodoulopoulos, I. Tomkos, and E. A. Varvarigos, "Elastic Bandwidth Allocation in Flexible OFDM based Optical Networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 29, no. 9, pp. 1354 – 1366, May 2011.

[28] A. Bocoi, M. Schuster, F. Rambach, M. Kiese, C.-A. Bunge, and B. Spinnler, "Reach-Dependent Capacity in Optical Networks Enabled by OFDM," in Proceedings of of Optical Fiber Communication Conference (OFC), pp. 1 – 3, Paper OMQ4, 2009.

CHAPTER 8

Conclusions

THIS THESIS WORK is devoted to the analysis and development of statistical models for the link lengths and the shortest path lengths in optical transport networks (OTNs). The findings presented in this thesis are based on the measurements of parameters of real OTNs. In the framework of this work, we have obtained new statistical models for the OTNs which are found to be quite realistic and accurate. The results of the estimation methods are validated by comparing the results obtained from the exact calculations of the real OTNs. The performances of the newly proposed methods were also compared with the available alternatives and found to be better. In this chapter, we summarize the work done in this thesis, and provide the potential future work which may be carried on as its follow up.

8.1 Concluding Remarks

The link lengths in OTNs broadly exhibit the lognormal distribution. Overall, we found the GEV distribution to be the accurate model for the link length statistics of OTNs. The model can be estimated from incomplete information of the OTNs. The parameters of the associated GEV distribution can be estimated from the average link length of the OTN. We also developed expressions for average link lengths, based only on the knowledge of network coverage area and number of nodes, and improved them for better accuracy. We have shown that the GEV distribution estimates the link statistics of OTNs with good accuracy and thus can be used for the real applications.

Estimation of link-dependent parameters with incomplete information is essential in

the initial phases for the design of OTNs. It is helpful in the estimation of the initial capital needed for the network. In Chapter 5, we showed that the common link length dependent OTN parameters can be estimated with significant accuracy by using the model developed in Chapter 3. Knowing the number of links in an OTN, the model can estimate the link-dependent OTN parameters with minor errors which is not possible by the previous methods based on the average link length.

We have used the convex area for the estimation of the models, parameters and their subsequent applications for the parameter estimations to the estimation of link length and shortest path dependent parameters.. We used measuring tools to find out the convex area for these cases. In order to simplify the estimation of the convex area, we developed a method, in which just a planar map is used. We showed the effectiveness of the estimation methods based on the circumferential ellipse (CE) in case of the OTNs. Its utilities in the network related estimations have also been showed using real OTNs. CEs do not need the total coverage area for the estimations. Instead, only the major and minor axes are needed for their estimations. These ellipses are instrumental in estimating the convex area of the OTNs, from which the average link length can be estimated.

In Chapter 6, we analyzed the shortest path lengths between the node pairs of OTNs. Shortest path lengths between the node pairs of real OTNs were found to follow the Johnson S_B distribution. This distribution is considered a better choice than other distributions, because in addition to yielding low KSS values in all OTNs, it also accommodates both unimodal and bimodal PDFs, which are commonly observed in the OTNs. The bimodality is found in case of the networks having large convex areas with small number of nodes. We have also shown that the mean, median, standard deviation, and an approximate upper bound of the shortest path lengths of OTNs can be estimated from their convex areas. We developed expressions for the estimation of the parameters of the Johnson S_B distribution model from the information of the node locations. The developed models are found to be statistically acceptable.

We applied the model for the shortest path lengths developed in Chapter 6, and found it to be quite good for the estimations of the modulation formats for shortest paths between the node pairs of the OTNs. The percentage of error estimated with respect to the exact calculations are found to be quite low, which is remarkable considering the information it requires for these estimation (i.e., convex area and number of nodes). Thus overall, the model is quite good and can be used for the practical estimation of shortest path related parameters.

8.2 Potential Future Work

The developed models can be used for several new estimations of the OTNs. These methods are useful where the planning and dimensioning is difficult to perform, because the information available is limited or incomplete. The potential applications are numerous. Some of the prime examples of future work are:

1. The statistical analysis and modeling of number of hops between nodes and the minimum hop paths (the shortest path in terms of the number of hops) in OTNs;
2. Statistical analysis and modeling of the traffic in OTNs;
3. Evaluation of quality of signal in different types of OTNs using the estimations from the statistical distributions;
4. Applications in the resource consumption related areas including energy consumption and floor space occupation in OTNs, and its optimization.

APPENDIX A

Kolmogorov-Smirnov Statistic

MEASURMENT OF DISTRIBUTION fittings are necessary in statistical assessments. 'Goodnees-of-fit' tests are used for this purpose. In this thesis, we use Kolmogorov-Smirnov test which is popular for small and medium sized samples, and can also be used for large samples. In this test, fitting is quantified as Kolmogorov-Smirnov statistic (KSS). It is defined as the maximum difference (D_{max}) between the theoretical CDF (also known as hypothesized CDF, $G(x)$) and the empirical CDF (also known as empirical distribution function, $\hat{F}(x)$), which is mathematically defined in (A.1).

$$KSS = D_{max} = \sup_x(|\hat{F}(x) - G(x)|) \tag{A.1}$$

In Figure A.1, D $(= |\hat{F}(x) - G(x)|)$ is shown, whose maximum value is equal to the KSS for the fitting ($\hat{F}(x)$ is black steps and $G(x)$ is blue curve). We also show the PDF fitting for this data set whose samples (i.e., x) vary from –2 to 9.

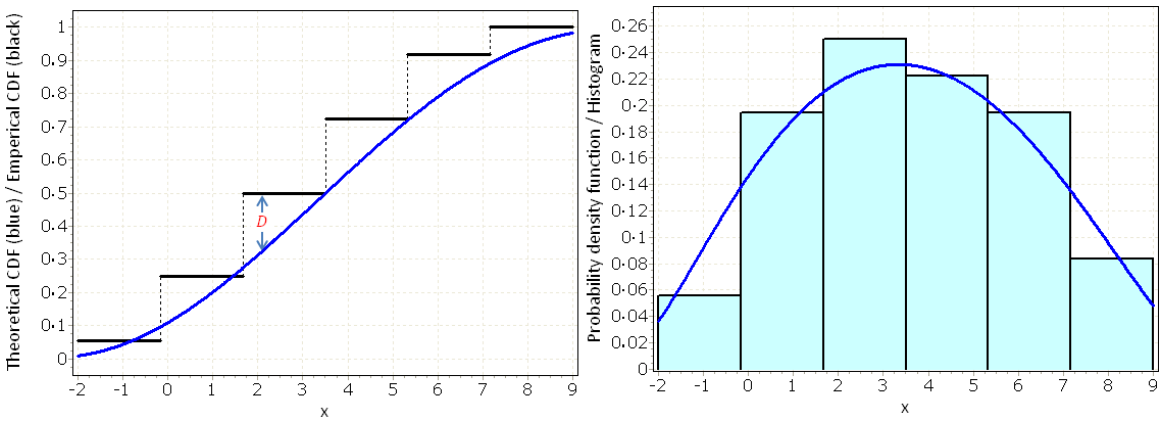

Figure A.1: KSS is the maximum value of D (=$|\hat{F}(x) - G(x)|$).

APPENDIX B

Estimation of the Shape Factor (ξ) of GEV Distribution

THE MATHEMATICAL GAMMA function is found in several cases such as the statistical distributions and probability of certain natural events. In the General Extreme Value (GEV) distribution, its parameters α, β and ξ are related to its *mean* and *variance* through the Gamma function. This function is defined for all complex numbers except zero and the negative integers.
For $x > 0$, it is defined as,

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt. \tag{B.1}$$

In the case of OTNs, whose link lengths follow the GEV distribution, we have [from expression (3.7)],

$$E(l) = \alpha - \frac{\beta}{\xi} + \frac{\beta}{\xi}\Gamma(1-\xi).$$

Making, $E(l) = \langle l \rangle$ in the above equation, we obtain,

$$\frac{\langle l \rangle - \alpha}{\beta} = \frac{\Gamma(1-\xi) - 1}{\xi}, \tag{B.2}$$

after separating $\langle l \rangle$, α and β from ξ terms. Still we cannot find an explicit expression for ξ from (B.2) due to the presence of $\Gamma(\cdot)$ terms on the right hand side.

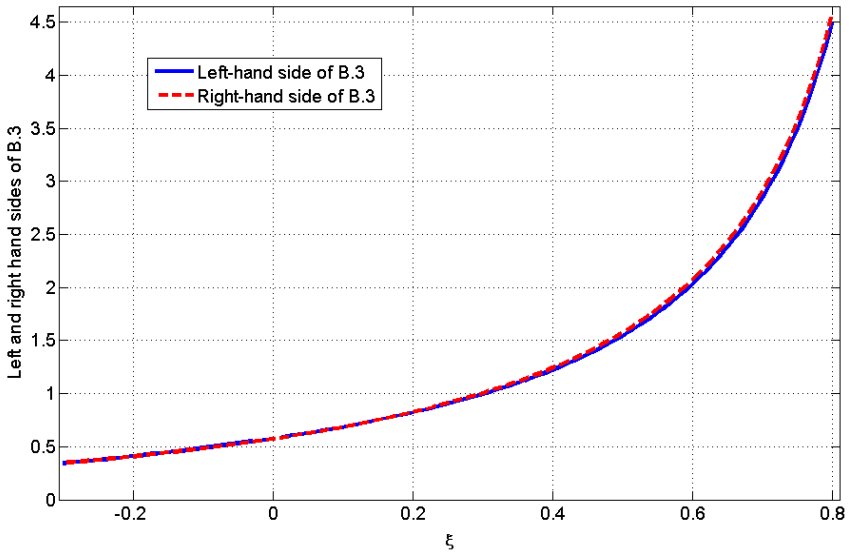

Figure B.1: Plot of $\frac{\Gamma(1-\xi)-1}{\xi}$, (blue, solid curve) and its approximation $\frac{1}{1-\xi} - 0.425$, (red, dotted curve) with respect to ξ.

For the 40 networks studied in Chapter 3, the minimum value of the shape factor, ξ is –0.2064 (for MZIMA) and the maximum value is 0.7317 (for PORTUGAL). So, the ξ bound for this case is: $-0.2064 \leq \xi \leq 0.7317$. For this range of values, it can be shown (see Figure B.1) that right hand side of (B.3) provides a good approximation for the right-hand side of (B.2).

$$\frac{\Gamma(1-\xi)-1}{\xi} \cong \frac{1}{1-\xi} - 0.425 \tag{B.3}$$

Substitution of approximation (B.3) into (B.2) results in,

$$\frac{\langle l \rangle - \alpha}{\beta} \cong \frac{1}{1-\xi} - 0.425. \tag{B.4}$$

So, now we can write the approximate expression for ξ as shown below, where expressions (3.5) and (3.6) are used for α and β respectively.

$$\xi \approx \frac{\frac{\langle l \rangle - \alpha}{\beta} - 0.575}{\frac{\langle l \rangle - \alpha}{\beta} + 0.425} = \frac{0.0887\langle l \rangle - 1.558}{0.5297\langle l \rangle - 13.927} \tag{B.5}$$

For large networks, when, $\langle l \rangle \gg 26.3$ km, we can approximate ξ as shown in (B.6),

$$\xi \approx \frac{0.0887}{0.5297} = 0.167. \tag{B.6}$$

APPENDIX C

Moments of Johnson S_B Distribution

THE MOMENTS OF the statistical distributions are helpful in the estimations of its parameters. However, sometimes the expressions for the moments become too complex and they make the parameter estimation quite complex. Johnson S_B distribution has four parameters, and its moment generating function depends on its parameters as shown in [1], in which r stands for the rth moment.

$$\mu_r(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} (1 + e^{-\frac{x-\gamma}{\delta}})^{-r} dx. \tag{C.1}$$

Here, $z = \frac{p-\zeta}{\lambda}$, and $x = \gamma + \delta ln\left(\frac{z}{1-z}\right) = \gamma + \delta ln\left(\frac{p-\zeta}{\zeta+\lambda-p}\right)$, (please note that ζ, λ, γ and δ in this appendix are the four parameters of Johnson $\mathbf{S_B}$ distribution for the shortest path lengths of OTNs, and p is the path length variable as described in Chapter 6). When, $r = 1$, we get the mean of the distribution.

$$\mu_1(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} (1 + e^{-\frac{x-\gamma}{\delta}})^{-1} dx \tag{C.2}$$

However, this integral in (C.2), which represents the mean, does not yield a clear and simple output. Using the results obtained in [2] and [3] it can be simplified to the expression shown in (C.3):

$$\mu_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\gamma^2} \frac{\frac{1}{\delta}(0.5 + D) - 2\pi\delta E}{1 + 2F}, \tag{C.3}$$

in which D, E, and F are shown in C.4 (a) – C.4(c).

$$D = \sum_{n=1}^{\infty} e^{-\frac{n^2}{2\delta^2}} cosh\left[\frac{n(1-2\gamma\delta)}{2\delta^2}\right] sech\left[\frac{n}{2\delta^2}\right] \tag{C.4a}$$

$$E = \sum_{n=1}^{\infty} \frac{e^{-\frac{(2n-1)^2\pi^2\delta^2}{2}}}{sinh\left[(2n-1)\pi^2\delta^2\right]} sin\left[(2n-1)\pi\gamma\delta\right] \tag{C.4b}$$

$$F = \sum_{n=1}^{\infty} e^{-2n^2\pi^2\delta^2} cos(2n\pi\gamma\delta) \tag{C.4c}$$

As presented in [1] and [4], the higher moments of this distribution can be estimated recursively from the first moment as:

$$\mu_{r+1}(x) = \mu_r + \frac{\delta}{r}\frac{d\mu_r}{d\gamma}. \tag{C.5}$$

So, standard deviation depends on the mean as:

$$\sigma = \sqrt{\mu_1 + \delta\frac{d\mu_1}{d\gamma} - \mu_1^2}. \tag{C.6}$$

As can be seen, the expressions in (C.3), (C.5) and (C.6) are interdependent and implicit with respect to the parameters of the distribution. This shows that it is not that simple to be used for the derivation of the mean and standard deviation from the parameters of the distribution. Though the higher moments of the Johnson S_B distribution can be deduced from the mean as shown in [4], the mean as presented in (C.3) itself needs complex calculations, and thus does not helpful in the direct estimation.

References

[1] N. L. Johnson, "Systems of frequency curves generated by methods of translation," *Biometrika*, vol. 36, no. 1/2, pp. 149 – 176, Jun. 1949.

[2] L. J. Mordell, "The value of the definite integral $\int_{-\infty}^{\infty} \frac{e^{at^2+bt}}{e^{ct}+d} dt$," *Quarterly Journal of Mathematics*, vol. 68, pp. 329 – 342, 1920.

[3] L. J. Mordell, "The definite integral $\int_{-\infty}^{\infty} \frac{e^{at^2+bt}}{e^{ct}+d} dt$ and the analytic theory of numbers," *Acta Mathematica*, vol. 61, No. 1, pp. 323 – 360, Jul. 1933.

[4] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Vol. 1*, Chichester, UK: Wiley-Interscience, 1994.

APPENDIX D

Median and Parameters of Johnson S_B Distribution

MEDIAN OF A statistical distribution is one of its key parameters. It equally divides the area under the probability density function (PDF) of the distribution into two halves. In Chapter 6, we have shown the PDF and cumulative distribution function (CDF) of the Johnson S_B distribution of the shortest path lengths of optical transport networks (OTNs) as shown in (D.1) and (D.2) respectively.

$$f\left(z;\gamma,\delta,\lambda,\zeta\right) = \frac{\delta}{\lambda\sqrt{2\pi}z(1-z)} exp\left[-\frac{1}{2}\left(\gamma + \delta log\left(\frac{z}{1-z}\right)\right)^2\right] \tag{D.1}$$

$$F(z;\gamma,\delta,\lambda,\zeta) = \Phi\left(\gamma + \delta log\left(\frac{z}{1-z}\right)\right) \tag{D.2}$$

Here, $z = \frac{p-\zeta}{\lambda}$, in which p represents the shortest path length variable (please note that, ζ, in this appendix stands for the location parameter of the Johnson S_B distribution for the shortest path lengths of OTNs). In Chapter 6, we have seen that the path length, p, is a log-normal variable in the Johnson S_B distribution model. As shown in Chapter 6, we present the associated normal variable x as:

$$x = \gamma + \delta log\left(\frac{z}{1-z}\right) = \gamma + \delta log\left(\frac{p-\zeta}{\zeta+\lambda-p}\right), \tag{D.3}$$

in which, p has the following bounds: $\zeta < p < \zeta + \lambda$.
Now, reverse calculating the parameter, z, as shown in [1], and [2] we have:

$$z = \frac{1}{1 + e^{(x-\gamma)/\delta}}. \tag{D.4}$$

Putting, x = 0, we get the median value of z, which is also associated with the median of the shortest path lengths m as:

$$z_{med} = \frac{m - \zeta}{\lambda} = \frac{1}{1 + e^{-\gamma/\delta}}. \tag{D.5}$$

Therefore, we can estimate the median of the shortest path lengths using (D.6).

$$m = \frac{\lambda}{1 + e^{-\gamma/\delta}} + \zeta. \tag{D.6}$$

Using the estimated values of the parameters of the Johnson S_B distribution, we have:

$$m_c = \frac{\lambda_c}{1 + e^{-\gamma_c/\delta_c}} + \zeta_c. \tag{D.7}$$

Reverse calculating for the ratios of the two shape parameters of Johnson S_B distribution from (D.7), we get:

$$\frac{\gamma_c}{\delta_c} = -log\frac{(\lambda_c + \zeta_c - m_c)}{(m_c - \zeta_c)}. \tag{D.8}$$

Therefore, we have:

$$\gamma_c = \delta_c \cdot log\frac{(m_c - \zeta_c)}{(\lambda_c + \zeta_c - m_c)}. \tag{D.9}$$

References

[1] N. L. Johnson, “Systems of frequency curves generated by methods of translation,” *Biometrika*, vol. 36, no. 1/2, pp. 149 – 176, Jun. 1949.

[2] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Vol. 1*, Chichester, UK: Wiley-Interscience, 1994.