



**REPUBLIKA E SHQIPËRISË
UNIVERSITETI POLITEKNIK, TIRANË
FAKULTETI I INXHINIERISË MATEMATIKE DHE
INXHINIERISË FIZIKE
DEPARTAMENTI I INXHINIERISË MATEMATIKE**

DISERTACION

**PËR MARRJEN E GRADËS SHKENCORE
DOKTOR**

ANALIZË STATISTIKORE MBI DINAMIKËN E GRAF RRJETAVE TË TELEFONISË FIKSE

**Paraqitur nga:
MSc. Orgeta GJERMËNI**

**Udhëheqës Shkencor:
Akad. Prof. Dr. Neki FRASHËRI**

Tiranë, 2018



**REPUBLIKA E SHQIPËRISË
UNIVERSITETI POLITEKNIK, TIRANË
FAKULTETI I INXHINIERISË MATEMATIKE DHE INXHINIERISË
FIZIKE
DEPARTAMENTI I INXHINIERISË MATEMATIKE**

DISERTACION

i

paraqitur nga:

MSc. Orgeta GJERMËNI

Për marrjen e gradës shkencore

DOKTOR

Programi i studimit: Probabilitet – Statistikë dhe Kërkime Operacionale

me temë:

**ANALIZË STATISTIKORE MBI DINAMIKËN E GRAF RRJETAVE TË
TELEFONISË FIKSE**

Udhëheqës Shkencor:

Akad. Prof. Dr. Neki FRASHËRI

Mbrohet më datë __/__/____ para jurisë:

1. _____
2. _____
3. _____
4. _____
5. _____

Kryetar
Anëtar, oponent
Anëtar, oponent
Anëtar
Anëtar

Tiranë, 2018

**DEDIKUAR
PRINDËRVE TË MI**

FALËNDERIM

Falënderoj përzemërsisht udhëheqësin tim shkencor Akad. Prof. Dr. Neki Frashëri për sugjerimet, vërejtjet dhe kritikën konstruktive në funksion të përmirësimit cilësor të këtij studimi. Gjithashtu, shpreh mirënjohjen time për të gjithë profesorët e mi, që kanë patur ndikim pozitiv dhe frymëzues tek unë në këtë rrugëtim.

Pafundësisht i jam falënderuese familjes time, për përkrahjen e pakushtëzuar të tyre dhe besimin e plotë ndaj meje!

TABELA E PËRMBAJTJES

1.	Një Vështrim i Përgjithshëm	1
1.1.	Hyrje	1
1.2.	Disa Studime të Aplikuara në Graf Rrjeta të Telefonisë	2
1.3.	Qëllimi i Studimit	5
1.4.	Trajtimi i të Dhënave	6
2.	Fenomeni “Botë e Vogël” dhe Lidhja Ndërmjet Përmasës dhe Rendit	11
2.1.	Hyrje	11
2.2.	Metodologjia	12
2.2.1.	Kuptime Bazë	12
2.2.2.	Kontrolli i Hipotezave në lidhje me “Botë e Vogël”	15
2.2.3.	Analiza Korrelative	16
2.2.4.	Analiza e Regresit	18
2.3.	Rezultatet e Hulumtimit	26
2.3.1.	“Botë e Vogël”	26
2.3.2.	Korrelacioni Joparametrik	27
2.3.3.	Lidhja ndërmjet Përmasës dhe Rendit	31
2.3.4.	Distanca Mesatare Gjeodezike, Fuqia Mesatare dhe Densiteti	34
2.4.	Diskutime	35
2.5.	Përfundime	36
3.	Analiza e Mbijetesës në Lidhje me Kohëzgjatjen e Thirrjeve	37
3.1.	Hyrje	37
3.2.	Metodologjia	38
3.2.1.	Terminologji dhe Kuptime Bazë	38
3.2.2.	Vlerësime Joparametrike	39
3.2.3.	Modelet Gjysmëparametrike	40
3.3.	Rezultatet e Hulumtimit dhe Diskutime	42
3.4.	Përfundime	47
4.	Shpërndarjet e Fuqive në Graf Rrjeta	48
4.1.	Hyrje	48
4.2.	Metodologjia	50
4.2.1.	Kuptime Bazë mbi Shpërndarjet dhe Vetitë e Tyre	50
4.2.2.	Vlerësimi dhe Testimi i x_{min} të Vlerësuar	57
4.2.3.	Kontrolli i Hipotezave në Lidhje me Shpërndarjen Hipotetike	58

4.2.4.	Krahasimi Direkt i Shpërndarjeve	60
4.2.5.	Paraqitje e Përmbledhur e Vlerësimeve të Parametrave të Shpërndarjeve	61
4.2.6.	Korrelacioni Ditor Ndërmjet Fuqive Hyrëse dhe Dalëse për Klient	61
4.3.	Rezultatet e Hulumtimit dhe Diskutime	62
4.3.1.	Analiza e Shpërndarjeve Ditore të Fuqive të Kulmeve	62
4.3.2.	Vlerësime për Korrelacionin Ditor Ndërmjet Fuqive Hyrëse dhe Dalëse për Klient	76
4.4.	Përfundime	77
5.	Klasifikimet Preferuese dhe Reciprociteti	78
5.1.	Hyrje	78
5.1.	Metodologjia	79
5.1.1.	Kuptime Bazë	79
5.1.2.	Ndërtimi i Modeleve.....	82
5.2.	Rezultatet e Hulumtimit për Klasifikimet Preferuese.....	84
5.3.	Diskutime	87
5.4.	Përfundime	88
6.	Konkluzione dhe Rekomandime	90
	Referenca.....	92

ABSTRACT

Landline network graph has been studied very little, although it is one of the oldest communication networks still in use. The purpose of this study is the statistical analysis of the dynamics of landline network graph, given their continuous development and temporary connections. The study is focused on the "small world" phenomenon and its correlation to some network graph characteristics, the relationship between the number of different communicating couples and the number of clients, and the dynamics of some other characteristics. Further, it is investigated on the impact of the call start time in its duration, the distribution that best fits the number of distinct communicators and the number of daily calls per client, and whether there is a correlation between incoming and outgoing calls per client. Also, it is checked whether the total number of calls per client and the total duration of calls affect preference and reciprocity in communication. For this purpose several models of network graphs time series have been built. Day-to-day or daily intervals snapshots are used to build network graphs at a time step. The methodology applied during the study analysis combines descriptive and inferential analytical methods. The calculations were carried out on the statistical platform R. The presence of the "small world" phenomenon indicates that information spreads rapidly in the network graph as within a day or even in a daily interval. The relationship between size and order can be used to predict the number of distinct client pairs, which communicate within a day or in a daily interval, based on the number of active clients it has. The call start time was related to a good prognosis for its duration. The high value of the month's day variable was associated with a high risk rate and poor survival communication. Between the power law and the lognormal distribution, the best way to fit the distribution of the number of daily communicators, different from one another, per client, has been found lognormal distribution. Also, this distribution describes better the total number of incoming and outgoing phone calls per client. Both distributions have high-risk to be inadequate to express the number of daily outgoing calls per client. The total number and total duration of calls per client resulted to be non-significant in determining communication preferences, except in case of models built on a daily basis when the features considered were the total number of outgoing calls and the total length of outgoing calls per client. Active clients who had performed a high number of outgoing calls had a slight tendency to communicate with other clients who had performed a low number of outgoing calls. The same trend was observed on a daily basis even in the case of the total duration of outgoing calls. During the day-to-day communication, there was almost no mutual relationship between the clients. Even in the case of daily interval communication, relation reciprocity remained at a very low level. Structure analysis of the clients' network graph helps us to better understand how the operator's clients use landline calls between them. Beyond understanding, interest is also on predicting. Results can be used to build better business or infrastructure policies.

Keywords: landline network graph, "small-world", degree analysis, survival analysis, assortativity coefficient, reciprocity.

PËRMBLEDHJE

Graf rrjeti i telefonisë fikse, pavarësisht se është ndër rrjetat më të vjetra të komunikimit ende në përdorim, është studiuar shumë pak. Qëllimi i këtij punimi është analiza statistikore mbi dinamikën e graf rrjetave të telefonisë fikse, duke marr parasysh zhvillim e tyre të vazhdueshëm dhe lidhjet e përkohshme. Objekte të studimit janë fenomeni “botë e vogël” dhe korrelacioni i tij në raport me disa karakteristika të graf rrjetit, lidhja ndërmjet numrit të çifteve të ndryshme që komunikojnë dhe numrit të klientëve, dhe dinamika e disa karakteristikave të tjera. Më tej, hulumtohet për ndikimin e momentit të fillimit të thirrjes në kohëzgjatjen e saj, për shpërndarjen që përfaqëson mirë numrin e komunikuesve të ndryshëm dhe numrin e thirrjeve ditore për klient dhe shikohet nëse ka korrelacion ndërmjet thirrjeve hyrëse dhe dalëse për klient. Gjithashtu, kontrollohet nëse totali i thirrjeve telefonike për klient dhe kohëzgjatja totale e thirrjeve ndikojnë në preferencat dhe reciprocitetin në komunikim. Për këtë qëllim janë ndërtuar disa modele serish kohore të graf rrjetave. Fotografimet statike ditore ose intervale ditore, janë përdorur për të ndërtuar graf rrjetat në një çast të kohës. Metodologjia e zbatuar gjatë analizës studimore gërshton metodat statistikore përshkuese me ato analitike inferenciale. Njehsimet janë kryer në platformën statistikore R. Prania e fenomenit “botë e vogël” tregon se informacioni përhapet në mënyrë të shpejtë në graf rrjet si brenda një dite apo edhe intervali ditësh. Lidhja e gjetur ndërmjet përmasës dhe rendit mund të përdoret për të parashikuar numrin e çifteve të ndryshme të klientëve të operatorit, që komunikojnë brenda një dite ose intervali ditësh, në bazë të numrit të klientëve aktiv që ai ka. Ora, gjatë së cilës fillonte thirrja telefonike, ishte e lidhur me një prognozë të mirë për kohëzgjatjen. Vlera të larta të ndryshores dita e muajit lidheshin me normë të lartë të rrezikut dhe mbijetesë të dobët të lidhjeve komunikuese. Ndërmjet shpërndarjes power law dhe lognormale, mënyra më e mirë për të përfaqësuar shpërndarjen e numrit të komunikuesve ditor, të ndryshëm nga njëri-tjetri, për klient, u gjet se është shpërndarja lognormale. Gjithashtu, kjo shpërndarje shpreh më mirë numrin e thirrjeve telefonike ditore totale dhe hyrëse për klient. Të dyja shpërndarjet kanë risk të lartë të jenë të papërshtatshme për të shprehur numrin e thirrjeve telefonike ditore dalëse për klient. Totali i thirrjeve për klient dhe kohëzgjatja totale e thirrjeve për klient rezultuan të jenë jo domethënëse për të përcaktuar preferencat në komunikim, me përjashtim të rastit në modelet e ndërtura në bazë ditore kur karakteristika e shqyrtuar ishte totali i numrit të thirrjeve dalëse dhe totali i kohëzgjatjes së thirrjeve dalëse për klient. Klientët aktiv të cilët kishin kryer një numër të lartë thirrjesh dalëse, kishin prirje të lehtë të komunikonin me klientë të tjerë të cilët kishin kryer një numër të ulët thirrjesh dalëse. E njëjta prirje u vu re në bazë ditore edhe në rastin e totalit të kohëzgjatjes së thirrjeve dalëse. Gjatë komunikimit ditor pothuajse nuk ekzistonin marrëdhëniet reciproke ndërmjet klientëve. Edhe në rastin kur komunikimi u pa në interval ditor, reciprociteti i marrëdhënieve qëndronte në nivel të ulët. Analiza e strukturës së graf rrjetit të klientëve të telefonisë na ndihmon të kuptojmë se si klientët e operatorit e përdorin telefoninë fikse midis tyre. Përveç të kuptuarit, interes ka edhe parashikimi. Rezultatet mund të përdoren për të ndërtuar politika biznesi apo infrastrukture më të mira.

Fjalëkyçe: graf rrjet i telefonisë fikse, “botë e vogël”, analiza e fuqive, analiza e mbijetesës, koeficienti i klasifikimeve preferuese, reciprociteti.

LISTA E FIGURAVE

Figurë 1.1: Pamje e të dhënave të regjistruara, të koduara, në lidhje me thirrjet telefonike të kryera.	6
Figurë 1.2: Ilustrim pamor i një graf rrjeti të telefonisë fikse-komponenti i lidhur më i madh i tij [41].	8
Figurë 2.1: Ilustrim pamor për koeficientin e kalimtaritetit.	14
Figurë 2.2: Grafiku i pesëfishtë me pika për: rendin, përmasën, fuqinë mesatare dhe “botë e vogël” sipas të dy përcaktimeve, Modeli 2.1.	28
Figurë 2.3: Grafiku i pesëfishtë me pika për: rendin, përmasën, fuqinë mesatare dhe “botë e vogël” sipas të dy përcaktimeve, Modeli 2.2.	28
Figurë 2.4: Hulimtim pamor për normalitetin e të dhënave, Modeli 2.1.	30
Figurë 2.5: Hulimtim pamor për normalitetin e të dhënave, Modeli 2.2.	30
Figurë 2.6: Paraqitja grafike për përjasjet në Modelin 2.1 dhe Modelin 2.2.	32
Figurë 2.7: Diagnoza për $m5 = \beta_0 + \beta_1n + \beta_2n^2 + \beta_3n^3$ në Modelin 2.1.	33
Figurë 2.8: Diagnoza për $m = \beta_0 + \beta_1n^6 + \beta_2n^7 + \beta_3n^8$ në Modelin 2.2.	34
Figurë 2.9: Distanca mesatare gjeodezike dhe fuqia mesatare e kulmeve në lidhje me kohën.	35
Figurë 2.10: Densiteti i graf rrjetave në lidhje me kohën.	35
Figurë 3.1: Pamje e të dhënave të komunikimit telefonik, të cilat u përdorën për të kryer analizën e mbijetesës. Të dhënat gjenden në [38].	37
Figurë 3.2: Kurbat Kaplan-Meier: A – tregon vlerësimin për kurbën e mbijetesës me interval me besueshmëri 95% duke përdorur log-transformimet; B – tregon një zmadhim të A në intervalin e kohës (140, 160); C – tregon vlerësimin për kurbat e mbijetesës për dy grupet, dhe D – është zmadhim i C në intervalin kohor (0, 1000).	43
Figurë 3.3: Vlerësimet për kurbat bazë të mbijetesës në Modelin 7, të krahasuara me kurbën e mbijetesës në Modelin 1.	46
Figurë 3.4: Funksionet bazë të rrezikut të grumbulluar në Modelin 7, të dhëna në rastin ku $\beta = 0$	46
Figurë 4.1: Histogrami dhe grafiku normal kuantil-kuantil, për vlerat e fuqive të kulmeve për $G1, G15$ dhe $G30$ në rastin e Modelit 4.1.	64
Figurë 4.2: Histogrami dhe grafiku normal kuantil-kuantil, për vlerat e fuqive të kulmeve për $G1, G15$ dhe $G30$ në rastin e Modelit 4.2-Totale.	64
Figurë 4.3: Histogrami dhe grafiku normal kuantil-kuantil, për vlerat e fuqive të kulmeve për $G1, G15$ dhe $G30$ në rastin e Modelit 4.2-Hyrëse.	65
Figurë 4.4: Histogrami dhe grafiku normal kuantil-kuantil, për vlerat e fuqive të kulmeve për $G1, G15$ dhe $G30$ në rastin e Modelit 4.2-Dalëse.	65
Figurë 4.5: Paraqitje grafike e FXx në rastin e Modelit 4.2 për $G1, G15$ dhe $G30$ në plane me boshte të logaritmuara.	72
Figurë 4.6: Paraqitje grafike e FXx në rastin e Modelit 4.1 për $G1, G15$ dhe $G30$ në plane me boshte të logaritmuara.	73
Figurë 4.7: Krahasimi i vlerësimeve për $xmini$ në Modelet 4.1 dhe 4.2 në lidhje me shpërndarjet PL dhe LN.	75
Figurë 5.1: Dy graf rrjeta me shpërndarje të njëjtë të fuqive të kulmeve por me korrelacion të ndryshëm të fuqive. Graf rrjeti (a) është pozitivisht i përcaktuar sipas klasifikimeve preferuese në lidhje me fuqitë e kulmeve, ndërsa graf rrjeti (b) është negativisht i përcaktuar. Shkëputur nga [44].	78
Figurë 5.2: Koeficienti i klasifikimeve preferuese sipas nën modeleve të Modelit 2.1.	85
Figurë 5.3: Koeficienti i klasifikimeve preferuese sipas nën modeleve të Modelit 2.2.	85
Figurë 5.4: Shkalla e reciprocitetit e njehsuar sipas rastit të parë dhe të dytë për Modelin 2.1 dhe Modelin 2.2.	85
Figurë 5.5: Shkalla e reciprocitetit, e njehsuar si korrelacion ndërmjet elementëve të matricës së fqinjësisë së graf rrjetave të orientuara.	87

LISTA E TABELAVE

Tabelë 1.1: Paraqitje e përgjithshme e disa teknikave statistikore topologjike të përdorura në analizimin e të dhënave të thirrjeve komunikuese telefonike [42]. Shkurtime: t-l: telefoni lëvizës; t-f: telefoni fikse; m: muaj; j: javë; d: ditë; -: e pa aplikueshme.....	9
Tabelë 2.1: Vlerësime për disa madhësi strukturore të graf rrjetave, vlerësimet për “botë e vogël” dhe kufiri i sipërm i intervalit me besueshmëri 99% - $CL0.01$ dhe $LT0.01$. M është numri i graf rrjetave ekuivalente të rastit $E - R$, të simuluar. Madhësitë e shoqëruara me shenjën (*) janë të rrumbullakosura në të qindtat, ndërsa madhësitë e shoqëruara me (**) janë të rrumbullakosura në të mijtat.	26
Tabelë 2.2: Vetë strukturore të graf rrjetave, vlerësimet për “botë e vogël” dhe kufiri i sipërm i intervalit me besueshmëri 99% - $CLT0.01$ dhe $CL0.01$. Numri i graf rrjetave ekuivalente të rastit $E - R$, të simuluar për secilin rast është $M=1000$. Madhësitë e shoqëruara me shenjën (*) janë të rrumbullakosura në të qindtat, ndërsa me shenjën (**) në të mijtat.	27
Tabelë 2.3: P-vlerat e gjetura gjatë kontrollit të normalitetit nëpërmjet kriterit Shapiro-Wilk.....	29
Tabelë 2.4: Rezultatet e kontrollit joparametrik të korrelacionit për Modelin 2.1 dhe Modelin 2.2.	29
Tabelë 2.5: Rezultatet për vlerësimet dhe domethënien statistikore të parametrave për përqasjen në Modelin 2.1 dhe Modelin 2.2.	31
Tabelë 3.1: Statistika përshkruese në lidhje me të dhënat për kohëzgjatjen e thirrjeve telefonike.	42
Tabelë 3.2: Modelet e mbijetesës me ndryshore shpjeguese të pavaruara nga koha dhe pa ndërveprim midis tyre.....	45
Tabelë 4.1: Përshkrim numerik i treguesve të vendosjes dhe të formës për vargjet e fuqive në $G1$, $G15$ dhe $G30$ në Modelin 4.1 dhe 4.2.....	63
Tabelë 4.2: Probabilitetet empirike.....	66
Tabelë 4.3: Vlerësime në lidhje me parametrat dhe mirësinë e përqasjes (KS) për shpërndarjet PL dhe LN, në Modelin 4.1.	67
Tabelë 4.4: Vlerësime në lidhje me parametrat dhe mirësinë e përqasjes (KS) për shpërndarjet PL dhe LN, në Modelin 4.2-Totale.....	68
Tabelë 4.5: Vlerësime në lidhje me parametrat dhe mirësinë e përqasjes (KS) për shpërndarjet PL dhe LN, në Modelin 4.2-Hyrëse.....	69
Tabelë 4.6: Vlerësime në lidhje me parametrat dhe mirësinë e përqasjes (KS) për shpërndarjet PL dhe LN, në Modelin 4.2-Dalëse.	70
Tabelë 4.7: Krahësimi i LN kundrejt PL nëpërmjet testit Vuong të raportit të log-përgjasisë $\mathcal{R}$, në Modelin 4.1-Totale dhe Modelin 4.2-Totale.	73
Tabelë 4.8: Krahësimi i LN kundrejt PL nëpërmjet testit Vuong të raportit të log-përgjasisë $\mathcal{R}$, në Modelin 4.2-Hyrëse dhe Modelin 4.2-Dalëse.....	74
Tabelë 4.9: Përshkrimi numerik i parametrave të vlerësuar për shpërndarjes LN në Rastin 1 dhe Rastin 2.....	75
Tabelë 4.10: Analiza e korrelacionit ndërmjet fuqive hyrëse dhe dalëse ($dv - \sim dv+$), në Modelin 4.2. Vlerat e koeficientëve janë rrumbullakosur në të mijtat.	76
Tabelë 5.1: Vlerësimet e bëra për vlerat e koeficientit të klasifikimeve preferuese. Vlerat janë rrumbullakosur në të mijtat.....	84
Tabelë 5.2: Vlerat e shkallës së reciprocitetit. Vlera NAN (= 0.0/0.0) në tabelë nënkupton që në atë rast rezultati nuk mund të përcaktohet si numër. Vlerat janë rrumbullakosur në të dhjetëmijtat.	86
Tabelë 5.3: Koeficienti i korrelacionit për katër zona. Burimi i informacionit [18]	88

1. NJË VËSHTRIM I PËRGJITHSHËM

1.1. HYRJE

Fillesat e analizës së bazuar në graf rrjeta në shkencë, e veçanërisht themelimi matematikor i teorisë së grafeve, datojnë me zgjidhjen e Leonhard Euler për problemin e famshëm të urës Königsberg të publikuar në 1736. Ky problem njihet shpesh edhe si rezultati i parë rigoroz në teorinë e grafeve. Euler provoi se ishte e pamundur të ecje në shtatë urat e qytetit, në mënyrë të tillë që të kaloje në secilën prej tyre vetëm një herë. Kenneth Appel dhe Wolfgang Haken në 1976 vërtetuan teormën e katër ngjyrave. Kjo teoremë ishte formuluar si konjekturë që në vitin 1850. Teorema pohonte se, mjaftonin katër ngjyra për të ngjyrosur një hartë në mënyrë të tillë që dy zona fqinje të mos kishin të njëjtën ngjyrë.

Në shek. XX teoria e grafeve u bë më tepër statistikore dhe algoritmike [1]. Sipas fjalorit anglez të Oxford [2], termi “rrjetë” i referohet një sistemi të ndërlidhur objektesh (subjekte, njësi, individë etj.), ndërsa fjala “graf” vjen nga fakti që është shumë e favorshme përdorimi i paraqitjeve grafike për të pasqyruar këto lidhje. Kështu, përgjatë këtij studimi përdoret termi “graf rrjetë”.

Graf rrjetat shërbejnë si aparat për modelimin e sistemeve reale që janë objekte studimi. Ato janë një mënyrë për të koduar një relacion ndërmjet elementëve të një sistemi. Elementët e sistemit identifikohen nëpërmjet një bashkësie pikash dhe marrëdhëniet ndërmjet tyre nëpërmjet një bashkësie vijash (lidhjesh) bashkuese (të orientuara ose jo), që lidhin secila jo më shumë se dy prej këtyre pikave. Shpeshherë në modelimin e sistemeve reale nëpërmjet graf rrjetave, lidhjet pajisen me karakteristika numerike të cilat i karakterizojnë ato nga pikëpamja sasiore.

Grumbullimi sistematik i të dhënave në graf rrjeta të një forme apo të një tjetre daton që prej viteve 1930, si pjesë e studimeve empirike në graf rrjeta sociale. Rritja e fuqisë kompjuterike dhe revolucioni informatik ka bërë të mundur që të dhënat në lidhje me graf rrjeta reale të një shkallë të lartë¹ të mbledhen në mënyrë sistematike, e prej andej të analizohen vetitë dhe cilësitë strukturore e funksionale të tyre. Numri i madh i elementëve në to shtron nevojën e një analize statistikore, për të bërë një karakterizim matematikor të mirfilltë.

Matematikanët Hungarez Paul Erdős dhe Alfréd Rényi, nëpërmjet punimeve të tyre [3, 4], ishin themeluesit e teorisë të graf rrjetave të rastit. Ata dalluan dy modele graf rrjetash rasti $\mathcal{G}(n, m)$ dhe $\mathcal{G}(n, p)$. Në modelin $\mathcal{G}(n, m)$, çdo graf rrjeti me rend n dhe përmasë m i caktohej probabilitet i njëtrajtshëm, ndërsa në modelin $\mathcal{G}(n, p)$ brinjët në një graf rrjet me rend n ekzistonin me të njëjtën probabilitet p .

Graf rrjetat reale tradicionalisht janë supozuar si graf rrjeta të rastit ku lidhjet apo shkëputjet në to nuk ndikohen nga ndonjë mekanizëm i brendshëm, por janë konsideruar si plotësisht të rastësishme. Përmes studimeve empirike dhe qasjeve analitike [1] është treguar se graf rrjetat reale janë larg të qenit graf rrjeta rasti, por se ato zhvillohen sipas një parimi gjenerues të brendshëm të koduar në strukturën e tyre topologjike. Duke qënë se struktura topologjike e graf rrjetave reale ka një shmangie nga ajo e graf rrjetave të rastit, shtrohet nevoja për matje në terma sasiorë të parimeve të brendshme organizuese të tyre. Megjithatë, graf rrjetat e rastit janë përdorur

¹ Përshtatur nga anglishtja: Large scale networks.

gjerësisht si etalon krahasues, gjatë kontrollit të hipotezave në lidhje me karakteristika specifike të graf rrjetave reale.

Në literaturën e viteve të fundit për graf rrjetat e një shkalle të lartë, analiza statistikore është fokusuar kryesisht në tre veti: “botë e vogël”, koeficienti i grupimeve² dhe shpërndarja e fuqive [5]. Në kërkim të mekanizmave të fshehur, të cilat udhëheqin dinamikën dhe evoluimin e sistemeve komplekse, kërkuesit (fizikantë, biologë, matematikanë, inxhinierë, sociologë etj.) kanë filluar një analizë sistematike dhe karakterizim të paraqitjeve të graf rrjetave. Në lidhje me graf rrjetat e një shkalle të lartë është gjetur se kanë topologji komplekse dhe strukturë heterogjene. Këto cilësi kanë të bëjnë me struktura statistikore të karakterizuara nga: shpërndarje asimetrike e fuqive (me pjesë fundore të anuara, të rrëzuara, në të djathtë³), vetia shkallë e lirë⁴, vlera të larta të koeficientit të grupimeve dhe rend hierarkik [6].

Graf rrjeti i telefonisë [7] fikse është një ndër rrjetat më të vjetra të komunikimit, i cili është ende në përdorim (ndonëse rrjeti i komunikimit nëpërmjet postës është edhe më i vjetër). Ky graf rrjet është studiuar shumë pak, kryesisht për shkak të mungesës së të dhënave në lidhje me strukturën e tij. Graf rrjetat e telefonisë bëjnë pjesë në grupin e graf rrjetave teknologjike por edhe në ato të komunikimit e për rrjedhojë konsiderohen shpeshherë edhe si graf rrjeta sociale.

Epoka në të cilën gjendemi mund të konsiderohet edhe si epoka e realiteteve teknologjike të avancuara, në përdorim të gjerë. Telefonia lëvizëse dominon në komunikimin ndërmjet njerëzve, nisur edhe nga lehtësitë që ajo ofron. Pavarësisht këtij fakti, telefonia fikse ekziston ende në tregun e komunikimeve, e lidhur shpesh herë edhe me shërbime të tjera si internet apo televizion. Ndër avantazhet e telefonisë fikse kundrejt asaj lëvizëse përmendim qartësinë dhe qëndrueshmërinë e lidhjes komunikuese [8]. Në një studim të vitit 2013 mbi impaktin e teknologjive komunikuese në aktivitetin e trurit të njeriut [9] u gjet se ndikimi i rrezatimit elektromagnetik mbi aktivitetin e trurit të njeriut, si rezultat i përdorimit të telefonisë lëvizëse (sistemi i operimit GSM⁵), ishte 50 herë më i madh krahasuar me atë të telefonisë fikse.

1.2. DISA STUDIME TË APLIKUARA NË GRAF RRJETA TË TELEFONISË

Për sa jemi në dijeni, studimet e kryera në graf rrjeta të telefonisë bazohen kryesisht në pamje statike [10, 11, 12] të graf rrjetit dhe shumë pak prej tyre marrin në shqyrtim evoluimin në kohë, ndonëse hulumtimet inferenciale⁶ mbi dinamikën në kohë në graf rrjeta datojnë që prej viteve 1970-1980 [13, 14, 15, 16, 17]. Kryesisht studimet që kanë lidhje me graf rrjetat e telefonisë janë bazuar në të dhëna të telefonisë lëvizëse [12, 18, 19, 20, 21, 22, 23] dhe në disa raste në telefoninë fikse [24, 25, 11, 26]. Në vijim, përmenden disa⁷ punime të cilat janë bazuar në graf rrjeta të telefonisë fikse ose lëvizëse.

Aplikimi i parë i një studimi të mbështetur në thirrje të telefonisë fikse, haset në 1949 ku George Zipf modeloi ndikimin e distancës në komunikim [27].

Cortes *et al.* në [11] studioan lidhjen ndërmjet volumit të komunikimit dhe disa parametrave të tjerë në një graf rrjet të telefonisë. Ata prezantuan një strukturë të

² Përshtatur nga anglishtja: Cluster coefficient.

³ Përshtatur nga anglishtja: Heavy tails.

⁴ Përshtatur nga anglishtja: Scale – free.

⁵ Shkurtim nga anglishtja: Global System for Mobile communication.

⁶ Nënkuptohet, nxjerrje përfundimesh për popullatën në bazë të zgjedhjeve të bëra.

⁷ Në këtë seksion janë përmendur vetëm një pjesë e punimeve, të cilat janë cituar edhe në brendi të disertacionit.

dhënash, e cila sipas tyre kapte evoluimin e graf rrjetit në kohë. Gjithashtu, ata përcaktuan një mënyrë për të zbuluar klientët të cilët nuk kishin ndërmend të paguanin faturat duke u mbështetur në lidhjen e tyre me klientë të tjerë.

Aiello *et al.* [24, 25] ilustruan qëndrueshmërinë e një modeli grafesh rasti të propozuar prej tyre mbi të dhëna të komunikimit telefonik të cilat u kompiluan në bazë të një algoritmi të propozuar nga Abello *et al.* [28]. Thirrjet telefonike ishin të distancave të largëta, të bëra gjatë një ditë. Aiello *et al.* pohuan se vargu i fuqive të graf rrjetit të telefonisë nuk ndiqte në mënyrë perfekte modelin power law (α, β). Ky model nuk mund të kapte sjelljen rastësore të një graf rrjeti real të telefonisë. Numri i kulmeve me një fuqi të caktuar nuk zvogëlohej në mënyrë monotone me rritjen e fuqive. Fuqia hyrëse dhe dalëse e një kulmi nuk ishte e thënë të ishin të njëjtat. Gjithashtu, ata pohuan se shpërndarja e komponentëve, me përmasë më të vogël se komponenti më i madh, ishte afërsisht log-log lineare.

Nanavati *et al.* [12, 18] analizuan vetitë strukturore të katër graf rrjetave të telefonisë me të dhëna nga një operator i telefonisë lëvizëse, të cilat i'u korrespondojnë katër zonave gjeografike. Të dhënat kishin të bënin me thirrje telefonike dhe mesazhe SMS⁸. Qëllimi i studimit të tyre ishte që nëpërmjet analizës strukturore në graf rrjeta të kuptonin sjelljen e konsumatorit dhe t'i ofronin kompanisë informacion biznesi për ta ndihmuar atë në strategjinë e saj ekonomike në kushtet e një konkurrence në rritje në tregun e komunikimit telefonik. Ata prezantuan modelin “Gjueti-Thesari”⁹, i cili përshkruante formën e graf rrjetave të telefonisë duke u bazuar në shpërndarjen e brinjëve. Ky model dallonte nga modelet e tjera, të cilët bazoheshin në shpërndarjen që lidheshin me kulmet. Ata treguan se midis graf rrjetit të thirrjeve telefonike dhe atij të mesazheve kishte si ngjashmëri, po ashtu edhe dallime. Nëpërmjet një analize kohore në shtatë ditë, ata synuan të kuptonin evoluimin në kohë të strukturës së graf rrjetit dhe të shikonin se si ndryshonin parametrat e saj në lidhje me kohën.

Onnela *et al.* në [23] morën në shqyrtim një graf rrjet, me 3.9 milion kulme, të ndërtuar mbi bazën e regjistrimeve të thirrjeve nga një telefoni lëvizëse. Ata e trajtuan graf rrjetin si të peshuar dhe bënë një analizë të detajuar në lidhje me: shpërndarjen e fuqive, fortësive dhe peshave të kulmeve, koeficientin e klasifikimeve preferuese¹⁰ sipas topologjisë, koeficientin e peshuar të klasifikimeve preferuese, koeficientin e grupimeve, koeficientin e peshuar të grupimeve, numrin e fqinjëve të përbashkët të dy kulmeve fqinj. Ndërsa në [22], Onnela *et al.* trajtuan strukturat komunikuese të miliona përdoruesve të telefonisë lëvizëse, në nivel lokal dhe tërësor.

Seshadri *et al.* [19] analizuan një graf rrjet të ndërtuar mbi bazën e të dhënave të telefonisë lëvizëse, të marra nga një operator i madh me më shumë se një milion përdorues dhe dhjetra milion thirrje telefonike. Ata gjetën se: shpërndarja e numrit të thirrjeve telefonike për klient, shpërndarja e kohës totale për klient dhe shpërndarja e numrit të telefonuesve të ndryshëm për klient; ishin asimetrike dhe devijonin nga shpërndarja power law apo ajo lognormale. Ata propozuan shpërndarjen e dyfishtë Pareto lognormale¹¹. Sipas tyre, kjo shpërndarje përqaste më mirë evoluimin e sjelljes së graf rrjetit të telefonisë lëvizëse në kohë.

Dasgupta *et al.* [29] analizuan një graf rrjet të një shkalle të lartë me të dhëna nga komunikimi në një operator shumë i madh i telefonisë lëvizëse. Duke qenë se është treguar se kushton më lirë të mbash një klient ekzistues se sa të fitosh një të ri [30], qëllimi i tyre ishte të studionin evoluimin e humbjes së klientëve, të cilët ndërrojnë

⁸ Shkurtime nga anglishtja i *Short Messaging Service*.

⁹ Përshtatur nga anglishtja: Treasure – Hunt.

¹⁰ Përshtatur nga anglishtja: Assortativity Coefficient.

¹¹ Përshtatur nga anglishtja: Double Pareto Lognormal Distribution.

operator¹², bazuar në të dhënat përgjatë një periudhe katër mujore. Ata propozuan një teknikë për parashikimin e klientëve që ishin potencial për të ndërruar operator duke u nisur nga ekzaminimi i gjendjes aktuale të tyre në graf rrjet.

Ye *et al.* [10] eksploruan modele strukturore kohore të komunikimit në graf rrjetin social Catalano/Vidro. Ky rrjet ishte i ndërtuar mbështetur në të dhëna në lidhje me thirrje nga telefonia lëvizëse, të cilat i takonin një periudhë dhjetë ditore. Ata u përqëndruan më tepër në zbulimin e hierarkisë dhe “aktorëve” më të rëndësishëm në këtë graf rrjet.

Dong *et al.* [20] kryen një studim eksperimental në një graf rrjet social të ndërtuar mbi të dhëna të një telefonie lëvizëse në Kinë. Ata analizuan vetitë statistikore dhe strukturore në graf rrjet duke përdorur një metrikë të përshtatshme për të parë konsumatorët e telefonisë nga këndvështrimi i moshës dhe gjinisë. Ata gjetën se graf rrjeti ishte me shkallë të lirë dhe në të ishte i pranishëm fenomeni “botë e vogël”.

Vaz de Melo *et al.* [31] analizuan një graf rrjet real, të madh, të telefonisë lëvizëse. Ata treguan se shpërndarja log-logjistike, në një version të reduktuar (prerë) të saj, modelonte shumë mirë shpërndarjen e kohëzgjatjes së telefonatave për individ, në dallim me shpërndarjen lognormale apo atë eksponenciale. Gjithashtu, për të modeluar sjelljen kolektive të përdoruesve ata përdorën shpërndarjen MetaDist.

Doran *et al.* [32] përcaktuan një madhësi të re për të matur në mënyrë sasiore se sa shumë devijonte lidhja ndërmjet dy përdoruesve nga mesatarja dhe më pas eksploruan lidhjen ndërmjet sjelljes së telefonuesve dhe strukturës komplekse të graf rrjetit të telefonisë lëvizëse të marr në shqyrtim.

Sridharan dhe Bolot në 2013 [33] propozuan një metodë sistematike për të përdorur strukturat gjeometrike në funksion të kërkimit të informacionit të dobishëm në lidhje me modele të vendodhjes sikurse janë përmasat, forma, struktura e brendshme etj. Ata e aplikuan këtë në një graf rrjet të telefonisë lëvizëse të ndërtuar mbi thirrje të regjistruara përgjatë një muaji.

Dong *et al.* [34] kryen një hulumtim sistematik të shpërndarjes së kohëzgjatjes së telefonatave në një graf rrjet të telefonisë lëvizëse, të pare nga këndvështrimi sociologjik. Ata prezantuan një model graf rrjeti me faktor të varur nga koha për të bërë parashikimin në lidhje me kohëzgjatjen e telefonatave.

Blondel *et al.* në [35] bënë një rishikim të kontributit të dhënë deri në 2015, në graf rrjetat sociale të ndërtuara mbi të dhëna të telefonisë lëvizëse.

Në 2016 Sarmiento *et al.* [36] raportuan trungun e kërkimit në marrjen e mostrave të zgjedhjeve në graf rrjeta dhe i aplikuan ato në një graf rrjet të telefonisë lëvizëse.

Ndër graf rrjetat sociale që morën në shqyrtim Noka (Jani) dhe Hoxha në [21], ishte edhe graf rrjeti i telefonisë lëvizëse me të dhëna nga një operator në Shqipëri. Graf rrjeti ishte statik, i ndërtuar mbi bazën e thirrjeve telefonike dhe SMS, reciprokisht të kryera ndërmjet klientëve të telefonisë. Kjo nënkuptonte se dy klient të operatorit të telefonisë konsideroheshin në një lidhje atëherë dhe vetëm atëherë kur ata kishin patur një komunikim të ndërsjelltë me njëri-tjetrin nëpërmjet telefonatave ose SMS. Ndër të tjera, autorët treguan se fortësia e lidhjeve në graf rrjetin e telefonisë lëvizëse varej nga strukturat lokale përreth tyre¹³. Lidhjet ndërmjet komuniteteve ishin të dobta, ndërkohë që ato silleshin si ura ndërmjet tyre, duke patur fortësinë e duhur për ta mbajtur graf rrjetin të lidhur.

Aksu *et al.* [26] analizuan të dhëna nga komunikimi me anë të telefonisë lëvizëse dhe asaj fikse. Ata ndërtuan katër graf rrjeta të shkallës së lartë, një prej të cilave ishte

¹² Përshtatur nga anglishtja: Churn.

¹³ Për sa jemi në dijeni, Eva Jani (Noka) është e vetmja studiuese që ka zhvilluar hulumtim shkencor në graf rrjeta sociale të telefonisë lëvizëse me të dhëna nga Shqipëria si pjesë e disertacionit të saj të doktoraturës.

nga telefonia fikse dhe tre nga telefonia lëvizëse. Sipas tyre shpërndarja Pareto lognormal (PLN) ishte ajo që përqaste më mirë shpërndajen e fuqive të kulmeve. Në bazë të analizës empirike, ata dolën në përfundimin se ndryshimet në lidhje me densitetin, operatorin apo vendodhjen, nuk ishin të korreluara në lidhje me shpërndarjen e fuqive. Fuqia mesatare e graf rrjetave ishte proporcionale me logaritmin e rendit të graf rrjetit.

1.3. QËLLIMI I STUDIMIT

Qëllimi i këtij punimi është analiza statistikore mbi dinamikën e graf rrjetave të telefonisë fikse. Sfidë më e madhe në këtë fushë qëndron në faktin se graf rrjetat reale janë në zhvillim të vazhdueshëm dhe lidhjet në to janë të përkohshme, e për rrjedhojë teknikat e modelimit të tyre duhet të mbështeten mbi këtë fakt. Studimet e bazuara mbi të dhëna të marra nga baza të dhënash të telefonisë fikse ose asaj lëvizëse, kanë qënë subjekt studimi në disa punime shkencore. Objekti i studimit ka qënë jo vetëm për të zbuluar modele të fshehura nën sjelljen e konsumatorëve të telefonisë, por edhe për të testuar algoritme të ndryshme mbi të dhëna reale.

Analiza e strukturës së graf rrjetit të klientëve të telefonisë na ndihmon për të kuptuar më mirë lidhjet midis tyre. Me klient të operatorit të telefonisë kuptojmë pajtimtarët e tij. Në kontekstin e fushës së komunikimit, zbulimi i modeleve të fshehura nën graf rrjetin e klientëve do të shërbente në disa aspekte. Përmendim këtu: vlerësimin e trafikut të thirrjeve, parashikimin e vëllimit të thirrjeve, monitorimin e infrastrukturës së rrjetave, përcaktimin e strategjive të duhura të biznesit në drejtim të konsumatorit etj.

Në funksion të qëllimit të studimit, në secilin prej kapitujve pasues në këtë disertacion kërkohen përgjigje në lidhje me disa pyetje kërkimore të ngritura.

Në **Kapitullin 2** hulumtohet në lidhje me:

- **P2.1:** A është një seri kohore graf rrjetash të telefonisë fikse, e udhëhequr nga prania e fenomenit “botë e vogël”?
- **P2.2:** A ekziston ndonjë korrelacion për “botë e vogël” në lidhje me rendin, përmasën dhe fuqinë mesatare të kulmeve?
- **P2.3:** Si është lidhja ndërmjet numrit të çifteve të ndryshme që komunikojnë dhe numrit të klientëve?
- **P2.4:** Si evoluon distanca mesatare gjeodezike, fuqia mesatare dhe densiteti në lidhje me kohën?

Në **Kapitullin 3** merren në konsideratë disa faktorë shpjegues (dita e muajit, dita e javës, ora e ditës dhe grupi am/pm) të pavarur në lidhje me kohën, të cilat kanë të bëjnë me momentin e fillimit të thirrjes telefonike. Pyetjet e ngritura janë:

- **P3.1:** Sa përcaktues mund të jetë momenti i fillimit të thirrjes për kohëzgjatjen e saj?
- **P3.2:** Si ndikojnë në kohëzgjatjen e thirrjes faktorët: dita e muajit, dita e javës, ora e ditës dhe grupi? Cili model mund ta përshkruaj më mirë këtë marrëdhënie?

Në **Kapitullin 4** hulumtohet në lidhje me:

- **P4.1:** Cila është mënyra më e mirë për të shprehur numrin e komunikuesve ditor (të ndryshëm nga njëri-tjetri) për klient?

- **P4.2:** Cila është mënyra më e mirë për të shprehur numrin e thirrjeve telefonike ditore (totale, hyrëse dhe dalëse) për klient?
- **P4.3:** A mund të pranojmë se, në thirrjet ditore të një klienti, ekziston ndonjë korrelacion ndërmjet numrit të thirrjeve hyrëse me ato dalëse?

Në **Kapitullin 5** karakteristikat e marra në shqyrtim janë totali i thirrjeve telefonike dhe kohëzgjatja totale e thirrjeve telefonike. Në këtë kapitull hulumtohet në lidhje me:

- **P5.1:** Si është e përcaktuar sipas klasifikimeve preferuese një seri kohore graf rrjetash të telefonisë fikse?
- **P5.2:** A mund të pranojmë që klientët, të cilët kryejnë/marrin më shumë thirrje, komunikojnë më shumë me klientë që gjithashtu kryejnë/marrin më shumë thirrje telefonike?
- **P5.3:** A mund të pranojmë që klientët priren të komunikojnë me ata që janë po aq aktiv sikurse vetë ata, në kuptimin e kohëzgjatjes së përgjithshme të thirrjeve në të cilat ata gjenden të angazhuar?
- **P5.4:** A ekziston ndonjë dallim domethënës në lidhje me klasifikimet preferuese në rastet kur karakteristikat ndahen në tre kategori: totale, hyrëse dhe dalëse?
- **P5.5:** A është reciprok komunikimi ditor ndërmjet klientëve të telefonisë? Po komunikimi në interval ditor?

Në **Konkluzione dhe Sugjerime** paraqitet një përmbledhje e rezultateve të marra dhe jepen sugjerime për hulumtime të mëtejshme.

1.4. TRAJTIMI I TË DHËNAVE

Një regjistrim thirrjesh telefonike, të kryera në një rrjet komunikimi të telefonisë fikse, është siguruar prej një operatori¹⁴ me vendndodhje në jug të Shqipërisë. Në vitin 2014 ky operator mbulonte afërsisht 4% të tregut të telefonisë fikse në Shqipëri. Vrojtmet, pjesë e këtij regjistrimi, përbëjnë të dhënat të cilat pasqyrojnë gjendjen e komunikimit telefonik ndërmjet klientëve (pajtimtarëve) të operatorit përgjatë një muaji. Regjistrime të tilla kompanitë e telefonisë i mbledhin në çdo kohë, si pjesë e procesit të funksionimit të tyre. Paraqitja e të dhënave në momentin fillestar është në trajtën e një liste, sikurse tregohet në Figurën. 1.1.

calldate	callend	connect_duration	caller	called
11/01/2014 02:26	11/01/2014 02:27	73	1954	2056
11/01/2014 05:10	11/01/2014 05:11	32	1867	880
11/01/2014 05:15	11/01/2014 05:17	62	653	1531
11/01/2014 05:26	11/01/2014 05:28	63	2173	1745
11/01/2014 05:26	11/01/2014 05:28	65	2173	1745
11/01/2014 05:27	11/01/2014 05:29	74	2801	2415

Figurë 1.1: Pamje e të dhënave të regjistruara, të koduara, në lidhje me thirrjet telefonike të kryera.

Identitet e klientëve janë zëvendësuar me numra, me qëllim ruajtjen e privatësisë së tyre. Thirrjet telefonike nën shqyrtim janë kryer vetëm brenda rrjetit të klientëve të operatorit dhe jo nga klientët e operatorit në drejtim të një operatori tjetër apo anasjelltas. Arsyeja e këtij kufizimi është shmangia e mungesës së një informacioni të

¹⁴ Emri i kompanisë nuk përmendet për arsye privatësie.

plotë nga operatorët e tjerë. Thirrjet telefonike i përkasin muajit Nëntor 2014 dhe gjenden të publikuara në [37, 38].

Në literaturën e graf rrjetave të telefonisë vihen re disa teknika në marrjen e një zgjedhje të “filtruar”. Lambiote *et al.* [39], për formimin e një lidhje ndërmjet dy kulmeve (klientëve të telefonisë) në graf rrjet, vendosën si kusht reciprocitetin (ekzistencën e të paktën një thirrje në secilin prej drejtimeve) dhe në një interval kohor prej gjashtë muajsh të ishin kryer të paktën gjashtë thirrje në total. Kjo teknikë largonte një pjesë të konsiderueshme të lidhjeve. Ata e trajtuan graf rrjetin si të orientuar dhe u përqëndruan në komponentin më të madh të lidhur të tij. Gjithashtu, Onnela *et al.* [22] përdorën reciprocitetin gjatë analizës së graf rrjetave në lidhje me ndryshimet ndërmjet shpërndarjes së fuqive në dy versione të së njëjtës bashkësi të dhënash. Në njërin graf rrjet ishin marr në shqyrtim të gjitha thirrjet, ndërsa në tjetrin vetëm ato lidhje të cilat ishin reciproke. Graf rrjeti u trajtua si i paorientuar, i peshuar. Teknikën e reciprocitetit e gjejmë të aplikuar edhe nga Noka (Jani) dhe Hoxha [21].

Kovanen *et al.* [40] kanë hetuar nëse lidhjet njerëzore (mbështetur në komunikimin e telefonisë lëvizëse) ishin reciproke apo njëra palë ishte më shumë aktive se sa tjetra. Reciprociteti, për ta, nënkuptonte që çdo dy klient në një lidhje ishin njëkohësisht aktiv, në mënyrë të barabartë, për ruajtjen e lidhjes. Ata gjetën se reciprociteti i thirrjeve shpeshherë është jo i balancuar dhe se lidhjet e njëanshme ishin shumë të zakonshme. Për këtë arsye, filtrimi i të dhënave mbështetur vetëm mbi reciprocitetin pasqyronte shumë pak nga graf rrjeti real.

Ndonëse bashkësia e përgjithshme nën studim është bashkësia e thirrjeve telefonike mujore brenda rrjetit të telefonisë fikse të marr në shqyrtim, nënbashkësia e zgjedhur për tu analizuar është një zgjedhje e “filtruar” prej të dhënave një mujore. Konkretisht, në rastin e këtij studimi vëllimi fillestar i të dhënave ishte 81591 thirrjesh telefonike, ku në 41 prej tyre mungonte kohëzgjatja e thirrjes, ndërsa në 7442 raste kohëzgjatja ishte më e vogël se 10 sekonda. Këto thirrje u përjashtuan nga analiza e të dhënave, me arsyetimin se mund të ishin thirrje të humbura ose thirrje të gabuara të cilat mund të cënonin saktësinë e rezultateve gjatë studimit.

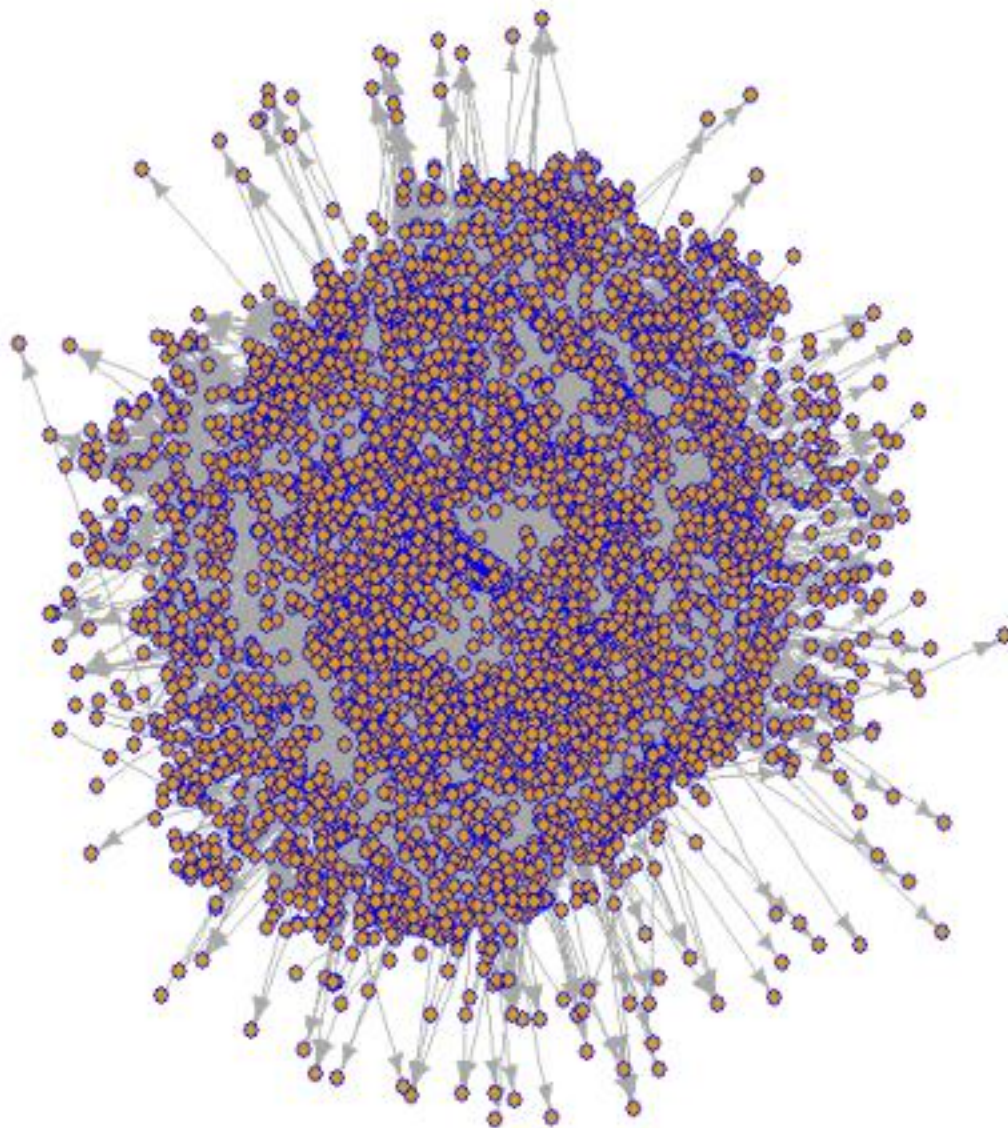
Klientë aktiv janë konsideruar vetëm ata të cilët kanë qenë të përfshirë në të paktën një thirrje telefonike (si telefonues, ose si të telefonuar) me kohëzgjatje të paktën 10 sekonda, të cilët në total ishin 3287. Vëllimi i zgjedhjes së “filtruar”, e cila i'u nënshtua analizës studimore prej nesh, ishte 90.83% e vëllimit fillestar të të dhënave.

Kjo teknikë statistikore e trajtimit të të dhënave, që kanë të bëjnë me komunikimin nëpërmjet telefonisë, është praktikuar edhe nga autorë të tjerë [12, 18]. Zgjedhja e bërë synon të reflektoj më mirë aspektet tërësore të komunikimit. Dasgupta *et al.* [29] përdorën kohëzgjatjen 5 sekonda për të filtruar të dhënat gjatë studimit të tyre.

Graf rrjeti i telefonisë jepet nëpërmjet një strukturë binare $G = (V, E)$, ku bashkësia e kulmeve (që nënkupton klientët aktiv) është shënuar me V dhe bashkësia e brinjëve është shënuar me E . Çdo brinjë paraqet një lidhje komunikuese ndërmjet dy klientëve aktiv. Kështu, nëse v_1 dhe v_2 janë dy kulme të ndryshëm nga njëri-tjetri, atëherë një brinjë $e_{1,2} = (v_1, v_2)$ ndërtohet ndërmjet tyre nëse v_1 dhe v_2 kanë qenë të angazhuar njëkohësisht në të paktën një thirrje telefonike ndis tyre. Në Figurën 1.2. jepet një ilustrim pamor i një graf rrjeti të orientuar.

Thirrjet telefonike mund të shikohen në të njëjtën kohë edhe si të orientuara, edhe si të paorientuara. Thirrjet janë të orientuara për arsye se telefonuesi është ai që bën thirrjen dhe harxhon kohë (dhe para) gjatë asaj lidhje, por biseda gjatë thirrjes është e paorientuar [40]. Kështu, brinjët e graf rrjetit, në varësi të problematikës, gjatë njehsimeve mund të trajtohen si: të orientuara të papeshuara, të orientuara të peshuara, të paorientuara të papeshuara, të paorientuara të peshuara. Në varësi të tyre, merr

emërtimin edhe graf rrjeti. Përgjatë hulumtimeve të bëra, në disa raste brinjët me shumëfishitet (të paktën 2) janë thjeshtuar në një të vetme.



Figurë 1.2: Ilustrim pamor i një graf rrjeti të telefonisë fikse-komponenti i lidhur më i madh i tij [41].

Në Tabelën 1.1. përmenden disa teknika statistikore me natyrë topologjike, të cilat janë përdorur edhe nga autorë të tjerë në studime hulumtuese të kryera mbi të dhëna të komunikimit nëpërmjet telefonisë. Tabela përfshin informacionin si vijon: natyrën e të dhënave të komunikimit telefonik, intervalin kohor, orientimin e lidhjeve, reciprocitetin e lidhjeve, thjeshtimin e lidhjeve, peshën e lidhjeve komunikuese.

Nuk ekziston një teknikë e përcaktuar në mënyrë të vetme për mënyrën se si duhen trajtuar gjatë analizës studimore të dhënat e komunikimit nëpërmjet telefonisë lëvizëse apo asaj fikse. Zgjedhja e teknikës varet nga objektivat e hulumtimit shkencor. Teknika të ndryshme të trajtimit të të dhënave shpjen në karakteristika të ndryshme të graf rrjetave dhe interpretime të ndryshme të së njëjtës bashkësi të dhënash.

Tabelë 1.1: Paraqitje e përgjithshme e disa teknikave statistikore topologjike të përdorura në analizimin e të dhënave të thirrjeve komunikuese telefonike [42]. Shkurtime: t-l: telefoni lëvizëse; t-f: telefoni fikse; m: muaj; j: javë; d: ditë; -: e pa aplikueshme.

Autorë	Të dhëna	Intervali Kohor	I Orientuar	Lidhje Reciproke	I Thjeshtuar	I Peshuar
Nanavati <i>et al.</i> [12, 18]	t-l	1 j, 1 m	po	jo	po	jo
Seshadri <i>et al.</i> [19]	t-l	2 m	jo	po	po	po
Dong <i>et al.</i> [20]	t-l	1 m	jo	jo	po	jo
Onnela <i>et al.</i> [23, 22]	t-l	18 j	jo	Të dyja (po, jo)	-	po
Ye <i>et al.</i> [10]	t-l	10 d	po	jo	jo	jo
Noka (Jani) dhe Hoxha [21]	t-l (thirrje, SMS)	1 m	jo	po	-	po
Dasgupta <i>et al.</i> [29]	t-l (thirrje, SMS)	1m	po	po	-	-
Aiello <i>et al.</i> [24, 25]	t-f	1 d ¹⁵	po	jo	jo	jo
Gjermëni dhe Ramosaço [41]	t-f	1 m	Të dyja (po, jo)	jo	po	po

Mitzenmacher [43] propozon përdorimin e analizës së serive kohore për vrojtimin e sistemit në kohë, për të gjykuar supozimet e bëra në lidhje me modele hipotetike. Me qëllim hulumtimin mbi dinamikën dhe përshkrimin e ndryshueshmërisë që pëson graf rrjeti i telefonisë fikse, gjatë studimit janë dalluar dy këndvështrime të të parit të të dhënave.

- Së pari, është ndërtuar një seri kohore graf rrjetash $G_1, G_2, \dots, G_{30}$, e tillë që çdo graf rrjet G_i bazohet vetëm mbi thirrjet e ditës së i -të të muajit. Ky këndvështrim, nëpërmjet fotografimeve statike ditore, ndihmon në kapjen e efektit të joqëndrueshmërisë së përhershme të lidhjeve komunikuese midis klientëve.
- Së dyti, për të hulumtur mbi dinamikën e topologjisë së graf rrjetit në kohë, është ndërtuar një seri kohore graf rrjetash $G_1 \leq G_2 \leq \dots \leq G_{30}$. Në këtë rast, çdo graf rrjet G_i është ndërtuar bazuar vetëm mbi ato thirrje që janë kryer prej ditës së parë të muajit, deri në ditën e i -të të tij. Graf rrjetat në këtë seri janë të tilla që $G_i \leq G_{i+1}$, për $i = 1, \dots, 29$.

Kështu, objekt i studimit konkret janë tridhjetë graf rrjeta me anë të të cilave synohet të kapet gjendja e rrjetit të komunikimit nëpërmjet telefonisë fikse në një çast të kohës. Çasti i kohës gjatë modeleve që trajtohen mund të jetë një ditë, ose një interval

¹⁵ Thirrjet telefonike i korrespondojnë intervaleve të ndryshme kohore por gjatë studimit ishin marr si thirrje të bëra brenda një dite.

ditor¹⁶. Një teknikë e ngjashme, për të “kapur” evoluimin e graf rrjetit në kohë është prezantuar edhe në [11].

Metodologjia e zbatuar gjatë analizës studimore gërsheton metodat statistikore përshkruese me ato analitike inferenciale. Në secilin prej kapitujve pasues, bëhet një hyrje për pyetjet kërkimore, përmendet metodologjia e ndjekur për atë rast, jepen rezultatet dhe më pas ato krahasohen me punime të tjera me natyrë të ngjashme. Në përfundim të çdo kapitulli jepen përgjigje në lidhje me pyetjet kërkimore përkatëse.

¹⁶ Disa ditë të njëpasnjëshme, duke filluar që nga dita e parë e muajit, [1, *i*].

2. FENOMENI “BOTË E VOGËL” DHE LIDHJA NDËRMJET PËRMASËS DHE RENDIT

2.1. HYRJE

Ithiel Pool dhe Manfred Kochen në 1950 ishin të parët të cilët eksploruan në mënyrë matematikore faktin se në graf rrjetat reale shumica e çifteve të kulmeve, pavarësisht largësisë midis tyre, ishin të lidhur me të paktën një zinxhir kulmesh fqinjë. Numri i kulmeve që mund të arrihen, pas një numri të caktuar hapash në graf rrjet, rritet eksponencialisht në lidhje me numrin e hapave. Kështu që, numri i hapave që duhen për të arritur një kulm të caktuar nisur nga një tjetër kulm në graf rrjet, rritet në mënyrë logaritmike (ose më ngadalë) në lidhje me rendin e graf rrjetit. Duke qënë se funksioni logaritmik është një funksion që rritet ngadalë, rrjedh se numri i hapave për të lidhur dy kulme me njëri-tjetrin është relativisht i vogël pavarësisht rendit të graf rrjetit që mund të jetë shumë i madh [44].

Në të shumtën e rasteve, lidhjet topologjike në graf rrjeta janë supozuar si tërësisht të rregullta ose tërësisht rastësore. Duncan Watts dhe Steven Strogatz [45] treguan se sistemet e botës reale qëndrojnë ndërmjet këtyre dy topologjive strukturore. Sistemet reale karakterizohen nga vlera të larta të koeficientit të grupimeve (ose kalimtaritetit), sikurse ndodh në grilat e rregullta dhe kanë distancë mesatare gjeodezike të vogël, sikurse ndodh në graf rrjetat e rastit [46]. Këto graf rrjeta janë emërtuar “botë e vogël” nga Watts dhe Strogatz, në ngjashmëri me fenomenin “botë e vogël” të trajtuar fillimisht në [47] e më pas më gjerësisht në [48]. Ky fenomen njihet ndryshe edhe si “ndarja në gjashtë shkallë” [49].

Në këtë pjesë, qëllimi i studimit është dhënia përgjigje ndaj pyetjeve të mëposhtme:

- **P_{2.1}:** A është një seri kohore graf rrjetash të telefonisë fikse, e udhëhequr nga prania e fenomenit “botë e vogël”?
- **P_{2.2}:** A ekziston ndonjë korrelacion për “botë e vogël” në lidhje me rendin, përmasën dhe fuqinë mesatare të kulmeve?
- **P_{2.3}:** Si është lidhja ndërmjet numrit të çifteve të ndryshme që komunikojnë dhe numrit të klientëve?

Për më tepër, në këtë pjesë hulumtojmë se si ndryshojnë disa karakteristika të graf rrjetave në lidhje me kohën.

- **P_{2.4}:** Si evoluon distanca mesatare gjeodezike, fuqia mesatare dhe densiteti në lidhje me kohën?

Në shqyrtim janë marr dy modele serish kohore të graf rrjetave të telefonisë fikse:

- **Modeli 2.1:** $G_1, G_2, \dots, G_{30}$, e tillë që çdo graf rrjet G_i , për $i = 1, \dots, 30$ është ndërtuar në bazë vetëm të thirrjeve të ditës së i -të të muajit;
- **Modeli 2.2:** $G_1 \leq G_2 \leq \dots \leq G_{30}$, e tillë që çdo graf rrjet G_i është ndërtuar bazuar vetëm mbi ato thirrje që janë kryer prej ditës së parë të muajit, deri në ditën e i -të të tij. Graf rrjetat në këtë seri janë të tilla që $G_i \leq G_{i+1}$, për $i = 1, \dots, 29$.

Në secilin prej modeleve, çdo graf rrjet G_i është i thjeshtë, i paorientuar, i papeshuar. Brinjët me shumëfishitet më të lartë se një, janë reduktuar në një brinjë të vetme. Modeli 2.1 e trajton fenomenin “botë e vogël” në bazë ditore, ndërsa Modeli 2.2 e trajton atë brenda një intervali ditor $[1, K]$ prej K ditësh të njëpasnjëshme duke filluar nga dita e parë e muajit, $K = 1, \dots, 30$. Disa prej pyetjeve të ngritura në lidhje me modelet e përmendura, janë trajtuar prej nesh edhe në [50, 51].

Fenomeni “botë e vogël” ka një rëndësi të veçantë në lidhje me graf rrjetat e komunikimit, ku “komunikimi” ka kuptim gjithëpërfshirës. Fakti që një graf rrjetë është “botë e vogël” nënkupton se struktura e saj është e atillë që një sistem i shkëmbimit “fqinjë-fqinjë” është i mjaftueshëm për të “transmetuar” informacionin në mënyrë të shpejtë në graf rrjet [52]. Efikasiteti i kësaj strukture transmetimi, në graf rrjetat “botë e vogël”, në nivel lokal edhe tërësor është treguar edhe në [53]. Në kontekstin e fushës së komunikimit telefonik, informacioni që transmetohet mund të jetë: lajm, reklamë, njoftim, shërbim i ri etj.

Ajo që ka dominuar në teorinë shkencore të graf rrjetave dhe aplikime, në lidhje me kontrollin për praninë e fenomenit “botë e vogël” është përdorimi i një teknike standarte sasiore. Kjo teknikë ka të bëjë me krahasimin e distancës mesatare gjeodezike dhe koeficientit të grupimeve (ose kalimtaritetit) të graf rrjetit me ato të një graf rrjeti Erdős-Rényi [46] ekuivalent me të. Telesford *et al.* kanë treguar [54] se kjo teknikë mund të sjell rezultate jo të sakta. Për këtë arsye, kemi aplikuar një teknikë e cila mat në mënyrë sasiore fenomenin “botë e vogël” në një graf rrjet [55, 56]. Ky fenomen është studiuar gjithashtu edhe në graf rrjeta të tjera reale sikurse mund të përmendim: Facebook [57, 58], rrjetat e ushqimeve [59], rrjetat e trurit [60, 56], rrjetat qelizore automatike [61], në parashikimin e performancës akademike [62] etj.

Studimi i lidhjes ndërmjet përmasës dhe rendit në graf rrjeta është i rëndësishëm për operatorin e telefonisë, për të parashikuar numrin e çifteve të ndryshme prej klientëve të tij që komunikojnë, brenda një dite ose intervali ditor, në bazë të numrit të klientëve aktiv që ai ka.

Njehsimet statistikore kompjuterike në këtë pjesë të studimit janë kryer kryesisht bazuar në paketat: igraphdata [63], igraph [64], Kendall [65], ggpbur [66] në platformën statistikore njehsuese R [67].

2.2. METODOLOGJIA

2.2.1. Kuptime Bazë

Graf rrjeti quhet *i thjeshtë* në rastin kur nuk ka leqe apo brinjë të shumëfishta. Le të jetë $G_i = (V_i, E_i)$, graf rrjet i thjeshtë, i paorientuar dhe i papeshuar. Numri i kulmeve $|V_i| = n_i$ quhet *rend*, ndërsa numri i brinjëve $|E_i| = m_i$ quhet *përmasë*. Një *shteg* me gjatësi s nga kulmi u në drejtim të kulmit v në G_i është një varg i renditur kulmesh të ndryshëm nga njëri-tjetri $u = v_0, v_1, \dots, v_s = v$ që plotësojnë kushtin $(v_j, v_{j+1}) \in E_i$ për $j = 0, 1, \dots, s - 1$. Një *rrugë* me gjatësi s nga u në kulmin v në G_i është një varg i renditur kulmesh $u = v_0, v_1, \dots, v_s = v$ që plotësojnë kushtin $(v_j, v_{j+1}) \in E_i$ për $j = 0, 1, \dots, s - 1$. Le të shënojmë me $\bar{l}_{G_i}$ distancën mesatare gjeodezike [68] në G_i , e cila njehsohet nëpërmjet formulës:

$$\bar{l}_{G_i} = \frac{2}{n_i(n_i + 1)} \sum_{(u,v) \in V_i \times V_i} l_{uv},$$

ku l_{uv} është distanca gjeodezike midis dy kulmeve $(u, v) \in V_i \times V_i$. Distanca gjeodezike ndërmjet dy kulmeve është gjatësia e shtegut më të shkurtër që lidh kulmet me njëri-tjetrin. Për të njehsuar $\bar{L}_{G_i}$, në R përdoret algoritmi kërkim së pari në gjerësi¹⁷ me funksion të kohës $O(n_i m_i)$ [68].

Graf rrjeti quhet *i lidhur* kur për çdo dy kulme u dhe v në të, ekziston një shteg nga u në v . Në rastet kur graf rrjeti nuk është i lidhur, gjatë përlllogaritjes së distancës mesatare gjeodezike merren në shqyrtim vetëm distancat përgjatë shtigjeve ekzistuese ndërmjet kulmeve [68].

Densiteti [69] në një graf rrjet G_i tregon afërsinë e tij me një graf rrjet të plotë dhe shërben për të karakterizuar kohezionin. Ai shprehet si raporti i brinjëve ekzistuese mbi brinjët e mundshme në graf rrjet:

$$den(G_i) = \frac{2 |E_i|}{|V_i|(|V_i| - 1)},$$

dhe vlerat e tij variojnë nga 0 në 1. Në rastin kur rendi i graf rrjetit është shumë i madh, përafrojmë $den(G_i) \approx \frac{2 |E_i|}{|V_i|^2}$.

Fuqia e kulmit [68] është numri i brinjëve incidente me të. Fuqia e një kulmi v në G_i është shënuar d_v^i , ndërsa *fuqia mesatare e kulmeve* në graf rrjet është shënuar si vijon:

$$\bar{d}_v^i = \frac{1}{n_i} \sum_{v \in V_i} d_v^i.$$

Një tjetër përcaktim për grafet e rrallë bazohet në krahasimin e fuqisë mesatare të kulmeve me rendin e graf rrjetit. Nëse $\bar{d}_v^i \ll |V_i|$, thuhet se graf rrjeti është i rrallë [6].

Në literaturë dallohen dy mënyra njehsimi të koeficientit të grupimeve në një graf rrjet të thjeshtë, të paorientuar dhe të papeshuar. Këto mënyra bazohen në këndvështrimin lokal dhe atë tërësor. *Koeficienti i grupimit*, i parë sipas këndvështrimit *lokal* (sipas kulmeve) u përkufizua fillimisht nga Watts dhe Strogatz [45, 52] dhe jepet nga formula:

$$cl_i(v) = \frac{\tau_\Delta(v)}{\tau_3(v)},$$

ku $v \in V_i$ dhe $\tau_\Delta(v)$ tregon numrin e trekëndëshave në G_i në të cilat është i përfshirë kulmi v . Ndërkohë, $\tau_3(v)$ tregon numrin e tresheve të kulmeve në G_i , të cilat janë të lidhura në mënyrë të tillë që dy prej brinjëve janë njëkohësisht incidente me kulmin v . *Koeficienti i grupimit* sipas këndvështrimit *tërësor* të graf rrjetit G_i njehsohet sipas formulës që vijon:

$$cl(G_i) = \frac{1}{|V_i|} \sum_{v \in V_i} cl_i(v).$$

Në rastin kur kulmet kanë fuqi të barabartë me zero ose një, koeficienti i grupimit për kulmin merret $cl(v) = 0$ [68]. Ky koeficient mat densitetin e trekëndëshave në graf rrjet.

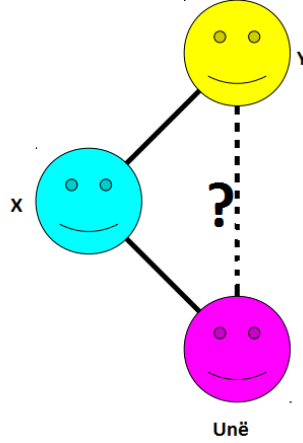
¹⁷ Përshtatur nga anglishtja: Breadth – first search.

Koeficienti i kalimtaritetit¹⁸ në një graf rrjet G_i [70, 52] e trajton graf rrjetin në këndvështrim tërësor. Rrënjët e tij gjenden në sociologji ku shfaqet me emrin “përpjesa e tresheve kalimtare” [69]. Koeficienti i kalimtaritetit jepet nga formula:

$$cl_T(G_i) = \frac{3\tau_\Delta(G_i)}{\tau_3(G_i)},$$

ku $\tau_\Delta(G_i) = \frac{1}{3} \sum_{v \in V_i} \tau_\Delta(v)$ është numri i trekëndëshave në G_i , dhe $\tau_3(G_i) = \sum_{v \in V_i} \tau_3(v)$ është numri i tresheve të lidhura të kulmeve. Vlerat e $cl_T(G_i)$ variojnë nga 0 në 1. Në literaturën matematike dhe fizike ky koeficient haset për herë të parë në [71].

Koeficienti i kalimtaritetit mund të interpretohet si probabiliteti që bashkë-komunikuesi i një bashkë-komunikuesit tim, të jetë gjithashtu bashkë-komunikuesi im. Një ilustrim pamor për këtë jepet në Figurën 2.1. Ndonëse $cl(G_i)$ dhe $cl_T(G_i)$ e shohin njësoj graf rrjetin nisur nga këndvështrimi tërësor, ato nuk janë me vlera të barabarta por shpeshherë kanë vlera të përafërta.



Figurë 2.1: Ilustrim pamor për koeficientin e kalimtaritetit.

Koeficienti i grupimeve dhe ai i kalimtaritetit janë madhësi të cilat matin lidhshmërinë dhe bëjnë pjesë në parametrat të cilat përdoren për të karakterizuar fenomenin “botë e vogël” në graf rrjeta.

Graf rrjeti i rastit Erdős-Rényi [46] ekuivalent me G_i , me të njëjtin rend $|V_i| = n$ dhe përmasë $|E_i| = m$ është shënuar $E - R_i$. Ky graf rrjet është ndërtuar në mënyrë të tillë, që çdo brinjë në të ekziston me të njëjtin probabilitet. Përkufizimet e mëposhtme janë përkufizime gjysmë kategorike për “botë e vogël” [45].

Përkufizim 2.1: Një graf rrjet G_i është “botë e vogël” nëse:

$$\bar{l}_{G_i} \geq \bar{l}_{E-R_i} \text{ dhe } cl_T(G_i) \gg cl_T(E - R_i).$$

Përkufizim 2.2: Një graf rrjet G_i është “botë e vogël” nëse:

$$\bar{l}_{G_i} \geq \bar{l}_{E-R_i} \text{ dhe } cl(G_i) \gg cl(E - R_i).$$

Për të matur në mënyrë sasiore fenomenin “botë e vogël” në një graf rrjet [55, 56] u formulua një përkufizim i ri. Fillimisht janë përcaktuar si vijon madhësitë:

¹⁸ Përshtatur nga anglishtja: Transitivity Coefficient.

$$\lambda_{G_i} = \frac{\bar{l}_{G_i}}{\bar{l}_{E-R_i}},$$

$$\gamma_{G_i}^T = \frac{cl_T(G_i)}{cl_T(E - R_i)},$$

$$S_i^T = \frac{\gamma_{G_i}^T}{\lambda_{G_i}}.$$

Në mënyrë të ngjashme janë përcaktuar:

$$\gamma_{G_i} = \frac{cl(G_i)}{cl(E - R_i)},$$

$$S_i = \frac{\gamma_{G_i}}{\lambda_{G_i}}.$$

Përkufizimet 2.1. dhe 2.2 implikojnë që $\lambda_{G_i} \geq 1$ dhe $\gamma_{G_i}^T \gg 1$ ($\gamma_{G_i} \gg 1$), prej nga rrjedh se $S_i^T > 1$ ($S_i > 1$).

Përkufizim 2.3: Një graf rrjet G_i është “botë e vogël” nëse $S_i^T > 1$ ($S_i > 1$).

2.2.2. Kontrolli i Hipotezave në lidhje me “Botë e Vogël”

Nëpërmjet përkufizimit 2.3 të “botë e vogël”, mund aplikohet një procedurë kontrolli statistikor të përgjithshëm në lidhje me praninë e këtij fenomeni sipas përkufizimit të Watts dhe Strogatz [55, 56]. Për çdo graf rrjet G_i , të serisë kohore të graf rrjetave (Modeli 2.1 dhe Modeli 2.2), është ndërtuar një graf rrjet rasti $E - R_i$ [46] ekuivalent me të, me rend dhe përmasë të njëjtë. Më pas, është njehsuar S_i^T dhe S_i në secilin rast. Për të bërë një kategorizim të qëndrueshëm, graf rrjetat janë kontrolluar për domethënien statistikore në lidhje me praninë e fenomenit “botë e vogël” nëpërmjet metodës Monte Carlo të zgjedhjeve.

Si hipotezë fillestare është ngritur [45] H_0 : “Sistemi G_i është një graf rrjet rasti Erdős-Rényi”. Kështu, për çdo graf rrjet G_i , janë ndërtuar M graf rrjeta rasti ekuivalente me të dhe janë njehsuar S_{ij}^T dhe S_{ij} në lidhje me secilin $E - R_{ij}$, për $j = 1, \dots, M$. Për secilin nga graf rrjetat është vlerësuar kufiri i sipërm i intervalit me besueshmëri 99% për hipotezën fillestare. Procedura [72] është përshkruar si vijon:

1. Le të jetë $\hat{F}$ shpërndarja empirike e M të dhënave S_{ij}^T , ku secilës prej tyre i është caktuar masa probabilitare $\frac{1}{M}$.
2. Një gjenerator rasti është përdorur për të zgjedhur, në mënyrë të pavarur dhe me kthim, M të dhëna të reja S_{ij}^{T*} nga $\hat{F}$. Kjo mënyrë zgjedhje është e tillë që çdo e dhënë e re është një përzgjedhje rastësore e pavarur nga bashkësia e M të dhënave fillestare. Të dhënat e reja, të cilat përbëjnë atë që quhet “zgjedhje bootstrap”, janë një nënbashkësi e të dhënave fillestare.
3. Në zgjedhjen bootstrap, njehsohet vlera mesatare e $\bar{S}_{ij}^T$.

4. Hapi i dytë dhe i tretë përsëriten 10000 herë, ku në secilën prej herëve përdoret një bashkësi e pavarur numrash të rinj rastësor për të gjeneruar një zgjedhje të re bootstrap. Vargu i vlerave mesatare bootstrap që rezulton është shënuar $\overline{S}_{ij}^T, \overline{S}_{ij}^{T^2}, \dots, \overline{S}_{ij}^{T^{10000}}$.

5. Le të jetë $[a^*, b^*]$ intervali qëndror me peshë 68% për vlerat $\overline{S}_{ij}^T$ të tilla që:

$$\frac{\#\{S_{ij}^T < a^*\}}{10000} = 0.16 \text{ dhe } \frac{\#\{S_{ij}^T < b^*\}}{10000} = 0.84 \quad 19.$$

Vlerësuesi bootstrap për devijimin standart σ , është $\hat{\sigma}^{(B)} = \frac{b^* - a^*}{2}$. Ky vlerësim bazohet në faktin se shpërndarja normale vendos 68% të saj në një devijim standart nga mesatarja. Sipas teorisë normale, gjysma e gjatësisë së intervalit, nga percentili i 16-të deri në percentilin e 84-të është një përcaktim i arsyeshëm për të vlerësuar devijimin standart.

Kufiri i sipërm i intervalit me besueshmëri 99% është $CL^{0.01} = 1 + 2.58\hat{\sigma}^{(B)}$, ku për graf rrjetat e rastit $E-R$ nisur nga përkufizimi $S^T = 1$. Një graf rrjet G_i konsiderohet se ka dallim domethënës krahasuar me një graf rrjet rasti, në rast se $S_i^T > CL^{0.01}$ [55, 56]. Në këtë rast hipoteza H_0 refuzohet. Në mënyrë të ngjashme është vepëruar edhe për S_{ij} .

2.2.3. Analiza Korrelative

Analiza korrelative dypërmasore synon të mat fortësinë dhe orientimin e lidhjeve ndërmjet dy ndryshoreve. Si rezultat i kësaj procedure statistikore merret *koeficienti i korrelacionit*, i cili është një madhësi skalare. Gjatë analizës korrelative mund të përftohen rezultate dhe interpretime të ndryshme në varësi të mënyrës së njehsimit të koeficientit të korrelacionit.

Më poshtë i jemi referuar dy madhësive joparametrike për matjen e korrelacionit (empirik), të cilat bazohen tek rangjet: koeficienti i korrelacionit sipas Spearman (ρ) dhe koeficienti i korrelacionit sipas Kendall (τ). Në këto raste nuk merret në konsideratë ndonjë supozim në lidhje me shpërndarjen e të dhënave, sikurse kërkohet në rastin e koeficientit të korrelacionit sipas Pearson ku të dhënat duhet të jenë me shpërndarje normale.

Le të jetë $\{(X, Y)\}$ një popullim dypërmasor në N njësi dhe $\{(X_i, Y_i)\}$, $i = 1, \dots, n$ një zgjedhje rasti e thjeshtë, ku $\{(R_i^X, R_i^Y)\}$ janë rangjet përkatëse të tyre.

Përkufizim 2.4: Vlerësuesi për koeficientin e korrelacionit sipas Spearman $\rho(X, Y)$ [73, 74] ndërmjet X dhe Y përcakohet si koeficienti Pearson i zbatuar për rangjet,

$$\hat{\rho}(X, Y) = \frac{\sum_{i=1}^n (R_i^X - \overline{R^X})(R_i^Y - \overline{R^Y})}{\sqrt{[\sum_{i=1}^n (R_i^X - \overline{R^X})^2] [\sum_{i=1}^n (R_i^Y - \overline{R^Y})^2]}}$$

ku $\overline{R^X}$ është vlerësuesi për vlerën mesatare të R^X dhe $\overline{R^Y}$ për R^Y . Koeficienti i korrelacionit sipas Spearman është një madhësi statistikore që mat fortësinë e një lidhje monotone ndërmjet të dhënash të çiftuara. Për të kontrolluar domethënien statistikore të këtij koeficienti, kryhet kontrolli i hipotezave, $H_0: \rho = 0$ dhe $H_a: \rho \neq 0$. Kriteri statistikor i përdorur në këtë rast është:

¹⁹ Simboli # është përdorur për të treguar numrin e elementëve të bashkësisë.

$$t_c = \frac{\hat{\rho} \sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}},$$

ku $|t_c|$ ka shpërndarje t me $n-2$ shkallë lirie (df²⁰). P-vlera është vlera e α e tillë $|t_c| = t_{1-\frac{\alpha}{2}; n-2}$.

Numri i përgjithshëm i çifteve të mundshme ndërmjet X dhe Y është $\frac{n(n-1)}{2}$, në rastin kur përmasa e zgjedhjes është n . Proçedura fillon me renditjen e çifteve nisur nga vlerat e X . Nëse ndryshoret X dhe Y janë të korreluara, atëherë ato do të duhet të kenë rend të njëjtë rangjesh. Kështu që, për çdo Y_i numërohen herët që $Y_j > Y_i$ (n_c - çifte të pajtueshme) dhe herët që $Y_j < Y_i$ (n_d - çifte të pakrahueshme).

Përkufizim 2.5: Vlerësuesi për koeficientin e korrelacionit sipas Kendall [73, 74] përcaktohet nga formula:

$$\hat{\tau} = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}.$$

Koeficienti i korrelacionit sipas Kendall shpreh diferencën ndërmjet probabilitetit që të dhënat e vrojtuar janë me rend të njëjtë, kundrejt probabilitetit që të dhënat e vrojtuar nuk kanë rend të njëjtë. Për të kontrolluar domethënien [75] e këtij koeficienti kryhet kontrolli i hipotezave, $H_0: \tau = 0$ dhe $H_a: \tau \neq 0$. Kriteri statistikor i përdorur në këtë rast është:

$$Z = \frac{\hat{\tau}}{\hat{\sigma}} = \frac{\hat{\tau}}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}},$$

i cili ka shpërndarje normale standarte $\mathcal{N}(0,1)$. Në formulën e mësipërme, $\hat{\sigma}$ tregon vlerësimin për devijimin standart.

Vlerat që marrin këta koeficientë variojnë ndërmjet -1 dhe 1 . Shenja pozitive e koeficientit të korrelacionit (sipas rangut) tregon se rritja e rangut për njëren ndryshore shoqërohet me rritjen e rangut edhe për ndryshoren tjetër. Ky fakt është tregues i ekzistencës së një lidhje në të njëjtin drejtim. Shenja negative e koeficientit tregon se lidhja midis tyre është në drejtim të kundërt, d.m.th me rritjen e rangjeve të njësive për njëren ndryshore, në përgjithësi rangjet e tyre në lidhje me ndryshoren tjetër zvogëlohen.

Domethënia e koeficientit të korrelacionit midis madhësive të marra në shqyrtim përcaktohet në varësi të probabilitetit kritik (p -vlerës) të vlerësuar nëpërmjet kriterëve statistikore respektive. Koeficienti konsiderohet statistikisht domethënës në rast se $p \leq \alpha$, hipoteza H_0 refuzohet. Në këtë rast, thuhet se ka mjaftueshëm evidenca në nivelin e rëndësisë α , për të pranuar që ekziston një lidhje në të njëjtin drejtim (ose lineare) ndërmjet madhësive në shqyrtim.

Për të kontrolluar normalitetin e të dhënave ngrihet hipoteza fillestare H_0 : “Të dhënat kanë shpërndarje normale”. Paskëtaj, hipoteza H_0 kontrollohet nëpërmjet kriterit statistikor të Shapiro-Wilk të modifikuar [76, 77, 78], i cili sipas [79] vlerësohet si kriteri më i mirë për të kontrolluar normalitetin. Nëse $p \leq \alpha$, atëherë hipoteza H_0

²⁰ Shkurtim nga anglishtja: *degree of freedom*.

refuzohet. Në këtë rast dilet në përfundimin, që gjasat janë që të dhënat nuk kanë shpërndarje normale.

Nëse $\{Y\}$ është një popullim me vlerë mesatare të pritur (pritje matematike) μ_Y dhe devijim standart σ_Y . Teorema e Chebyshev-it pohon se në rast se popullimi ka shpërndarje normale, 95% e të dhënave gjenden në intervalin $(\mu_Y - 1.96 \sigma_Y; \mu_Y + 1.96 \sigma_Y)$ ndërsa 99.7% e të dhënave gjenden në $(\mu_Y - 3\sigma_Y; \mu_Y + 3\sigma_Y)$. Nëse shpërndarja e popullimit është e pa përcaktuar, teorema pohon se për çdo numër pozitiv c , përpjesa e vlerave të popullimit që gjendet në intervalin $(\mu_Y - c\sigma_Y; \mu_Y + c\sigma_Y)$ është më e madhe ose e barabartë me $1 - \frac{1}{c^2}$ [80].

Lidhja e “botë e vogël” me rendin, përmasën dhe fuqinë mesatare është studiuar nëpërmjet koeficientit të korrelacionit sipas Kendall dhe sipas Spearman. Gjithashtu, është parë edhe lidhja ndërmjet përmasës dhe rendit, e cila analizohet më hollësisht në seksionin vijues.

2.2.4. Analiza e Regresit

Në përgjithësi, regresioni linear [80] nënkupton që funksioni i regresit është njëherësh linear në lidhje me parametrat e panjohur β_i . Në regresionin jo linear, funksioni i regresit nuk është njëherësh linear në lidhje me parametrat e panjohur β_i .

Gjatë aplikimit të modeleve të regresit linear supozohet se ekziston një marrëdhënie lineare ndërmjet ndryshores së varur dhe ndryshoreve të pavarura. Duke qënë se ky supozim jo gjithmonë është i vërtetë, aplikohen transformimet në disa (ose në të gjitha) ndryshore. Transformimet mund të jenë të tipit fuqi ose logaritëm dhe përdoren për të korigjuar modelin në rastet kur cënohet lineariteti, ose në rastin kur termi i gabimeve në model nuk ka shpërndarje normale dhe variancë konstante. Transformimet në lidhje me ndryshoren e varur shërbejnë për të korigjuar mungesën e normalitetit dhe/ose të variancës konstante të gabimeve, ndërsa transformimi i ndryshores së pavarur shërben kryesisht për të korigjuar mungesën e linearitetit. Është e mundur që ekuacioni linear i regresit në madhësitë e transformuara, të marr trajtën e një regresit linear të shumëfishtë të ndryshoreve të përfuara si rezultat i transformimeve të ndryshoreve bazë të vrojtura.

Në rastin e këtij studimi, popullimi i cili studiohet përbëhet nga graf rrjetat (ditore ose intervale) e telefonisë fikse në një operator të zgjedhur. Për secilin graf rrjet shqyrtohet numri i brinjëve dhe numri i kulmeve në të. Parametri i interesit për ne është numri i brinjëve në graf rrjeta. Zgjedhja e graf rrjetave (ditore ose intervale) konsiderohet si rastësore duke qënë se, rastësisht kemi zgjedhur një muaj të vitit dhe mbi bazën e tij janë përcaktuar graf rrjetat.

Për të gjetur lidhjen ndërmjet përmasës dhe rendit në graf rrjeta është aplikuar analiza e regresit. Rendi është konsideruar ndryshore e pavarur, ndërsa përmasa ndryshore e varur. Më herët, në [50, 51] është trajtuar prej nesh regresit linear i thjeshtë në madhësitë e log-transformuara. Për të bërë një përmirësim të rezultateve të gjetura, në këtë pjesë është aplikuar regresit linear në madhësitë e transformuara. Pas transformimeve të bëra, modeli i regresit është sjell në trajtën e një regresit linear të shumëfishtë.

2.2.4.1. Regresi Linear i Shumëfishtë [80, 81, 82]

Le të jetë $\{(Y, X_1, \dots, X_k)\}$ një popullim me $(k + 1)$ ndryshore, ku $X_1, \dots, X_k$ janë ndryshore të pavarura, ndërsa Y ndryshore e varur prej tyre.

Përkufizim 2.6: Funkzioni $\mu_Y(X_1, \dots, X_k)$ është quajtur funksioni *i regresit linear të shumëfishtë* të Y në lidhje me $X_1, X_2, \dots, X_k$ në rastin kur:

$$\mu_Y(X_1, \dots, X_k) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k,$$

dhe është mesatarja e vlerave në nënpopullimin e Y , për çdo kombinim të vlerave të $X_1, X_2, \dots, X_k$. Ky funksion njihet edhe me termin *ekuacioni teorik i regresit* dhe është më i miri për të parashikuar vlerat e Y duke përdorur ato të $X_1, X_2, \dots, X_k$. Koeficientët $\beta_0, \beta_1, \dots, \beta_k$ janë parametra të panjohur. Termi i shumëfishtë nënkupton që ekziston më shumë se një ndryshore e pavarur, ndërsa termi linear nënkupton që funksioni $\mu_Y(X_1, \dots, X_k)$ është linear në lidhje me parametrat $\beta_0, \beta_1, \dots, \beta_k$. Në rast se $k = 1$, kemi të bëjmë me regresin linear të thjeshtë.

Gjatë modeleve të regresit mund të hasim në ndryshore të vrojtuar bazë dhe ndryshore të përfutur. Ndryshoret e përfutur janë funksione të ndryshoreve bazë. Për shembull në barazimin,

$$\mu_Y(Z_1) = \beta_0 + \beta_1 Z_1^2 + \beta_2 \sqrt{Z_1},$$

mund të përcaktojmë ndryshoret X_1, X_2 si:

$$X_1 = Z_1^2 \quad X_2 = \sqrt{Z_1},$$

dhe vlerat e tyre x_1, x_2 si:

$$x_1 = z_1^2 \quad x_2 = \sqrt{z_1}.$$

Kështu që, funksioni i regresit mund të shkruhet si:

$$\mu_Y(X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2.$$

Modeli i regresit të popullimit. Le të jetë Y_i vlera e vërtetë e ndryshores së varur në njësinë e i -të në popullim dhe $\mu_Y(X_{i,1}, \dots, X_{i,k})$ vlera e parashikuar nga funksioni i regresit linear të shumëfishtë,

$$\mu_Y(X_{i,1}, \dots, X_{i,k}) = \beta_0 + \beta_1 X_{i,1} + \dots + \beta_k X_{i,k}.$$

Gabimi i parashikimit për elementin e i -të të popullimit është shënuar me E_i dhe jepet si diferenca ndërmjet vlerës së vërtetë dhe vlerës së parashikuar:

$$E_i = Y_i - \mu_Y(X_{i,1}, \dots, X_{i,k}) = Y_i - (\beta_0 + \beta_1 X_{i,1} + \dots + \beta_k X_{i,k}).$$

Në përgjithësi ekuacioni shkruhet si:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \dots + \beta_k X_{i,k} + E_i,$$

për $i = 1, \dots, N$ dhe quhet *modeli i regresit të popullimit*. Përdoret simboli $Y(x_1, \dots, x_k)$ për të shënuar vlerën e Y për një njësi të zgjedhur rastësisht, në të cilën $X_1 = x_1, \dots, X_k = x_k$.

Supozimet në modelin e regresit. Supozimet në lidhje me popullimin $\{(Y, X_1, \dots, X_k)\}$ me $k + 1$ ndryshore dhe metodën e përdorur për të përfutur një zgjedhje nga ky popullim jepen si vijon:

- *Supozimi 1.* Vlera mesatare $\mu_Y(x_1, \dots, x_k)$ e nënpopullimit të vlerave të Y me $X_1 = x_1, \dots, X_k = x_k$ është $\mu_Y(x_1, \dots, x_k) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$, ku $\beta_0, \beta_1, \dots, \beta_k$ janë parametra të panjohur dhe $x_1, \dots, x_k$ i përkasin bashkësisë së vlerave të lejuara për ndryshoret e pavarura.
- *Supozimi 2.* Devijimi standart $\sigma(Y|X_1, \dots, X_k)$ i vlerave të Y në nënpopullimin në të cilin $X_1 = x_1, \dots, X_k = x_k$, nuk varet nga vlerat $x_1, x_2, \dots, x_k$. Ky supozim njihet edhe si homogjeniteti i devijimit standart (ose homogjeniteti i variancës, ose homoskedasticiteti).
- *Supozim 3.* Çdo nënpopullim i vlerave të Y , i cili përcaktohet nga vlerat e veçanta të $X_1, X_2, \dots, X_k$ është një popullim me shpërndarje normale.
- *Supozim 4.* Të dhënat të cilat përbëjnë zgjedhjen n përmasore janë përftuar nga një zgjedhje rastësore e thjeshtë ose një zgjedhje me vlera të paracaktuara të $X_1, X_2, \dots, X_k$.
- *Supozim 5.* Të gjitha vlerat e zgjedhjes $y_i, x_{i,1}, \dots, x_{i,k}$ për $i = 1, \dots, n$ janë të vrojtueshme pa gabim.

Vlerësimi pikësor i parametrave. Për të bërë vlerësimin pikësor të parametrave të panjohur β_i në modelin e regresit, është përdorur *metoda e katrorëve më të vegjël*. Vlerësuesit e gjetur të parametrave i shënojmë respektivisht $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$. Vlerësimi korrespondues për funksionin e regresit është shënuar me:

$$\hat{\mu}_Y(X_1, \dots, X_k) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_k X_k,$$

i cili njihet edhe me termin *ekuacionin empirik i regresit*. Vlera e parashikuar e Y për njësinë e i -të, ku vlerat e ndryshoreve të pavarura janë $x_{i,1}, \dots, x_{i,k}$ është:

$$\hat{\mu}_Y(x_{i,1}, \dots, x_{i,k}) = \hat{\beta}_0 + \hat{\beta}_1 x_{i,1} + \dots + \hat{\beta}_k x_{i,k}.$$

Vlerësimi korrespondues i gabimit është shënuar me $\hat{e}_i$ dhe jepet nga:

$$\hat{e}_i = y_i - [\hat{\beta}_0 + \hat{\beta}_1 x_{i,1} + \dots + \hat{\beta}_k x_{i,k}].$$

Në këtë rast $\hat{e}_i$, $i = 1, \dots, n$ quhen *mbetje* dhe luajnë një rol të rëndësishëm në kontrollin e supozimeve. *Parimi i katrorëve më të vegjël* pohon se: vlerësimi më i mirë për funksionin e regresit:

$$\mu_Y(x_1, \dots, x_k) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k,$$

duke përdorur të dhënat e zgjedhjes përftohet duke zgjedhur $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ në mënyrë të tillë që *shuma e katrorëve e mbetur*²¹

$$SKM(X_1, \dots, X_k) = \sum_{i=1}^n \hat{e}_i^2$$

të jetë minimumi i mundshëm.

Vlerësimi pikësor për σ (ose $\sigma_{Y|X_1, \dots, X_k}$). Vlerësimi pikësor jepet nga

²¹ Njihet ndryshe dhe me termin: *shuma e katrorëve të gabimeve të parashikimit ose shuma e katrorëve të shmangieve nga regresi*.

$$\hat{\sigma} = \sqrt{\frac{SKM(X_1, \dots, X_k)}{n - k - 1}} = \sqrt{KMM(X_1, \dots, X_k)},$$

ku KMM quhet *katrori mesatar i mbetjeve*²². Numri $(n - k - 1)$ quhet *numri i shkallëve të lirisë për gabimin*, ose numri i shkallëve të lirisë në lidhje me $\hat{\sigma}^2$. Sa më e vogël të jetë $\hat{\sigma}$, aq më i mirë konsiderohet modeli i regresit. Numri i shkallëve të lirisë për gabimin është sa përmasa e zgjedhjes minus numrin e parametrave β në model. Në përgjithësi, katrori mesatar është një shumë katrorësh e pjestuar nga numri i shkallëve të lirisë që i korrespondon.

Mbetjet e standartizuara. Mbetjet e standartizuara për regresin linear të shumëfishtë përcaktohen si:

$$r_i = \frac{\hat{e}_i}{\hat{\sigma}\sqrt{1 - h_{i,i}}},$$

ku $h_{i,i}$ është element i i -të i diagonales së matricës $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$, ku $\mathbf{X}$ është një matricë me përmasa $n \times (k + 1)$ me element,

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,k} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{i,1} & x_{i,2} & \dots & x_{i,k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \dots & x_{n,k} \end{bmatrix}.$$

Vlera $h_{i,i}$ quhet “*krah*” *force*²³ i njësisë së i -të në zgjedhje. Vrojtimet me ndikim të madh janë rezultat i vlerave ekstreme të ndryshoreve të pavarura, pra i pikave që kanë *krah force* të lartë. Një vrojtim konsiderohet me ndikim të madh në rast se $h_{i,i} > \frac{2(k+1)}{n}$. Për regresin linear të shumëfishtë me k ndryshore të pavarura, nëse vlera e $h_{i,i}$ është e madhe atëherë vrojtimi në njësinë e i -të, $x_{i,1}, \dots, x_{i,k}$, është shumë larg mesatareve të të gjitha vrojtimeve $\bar{x}_1, \dots, \bar{x}_k$. Në këtë rast, vlerat e *krahut* të forcës mund të kenë ndikim në vlerësimet e bëra gjatë regresit.

2.2.4.2. Analiza e Variancës (ANOVA)

Gjatë vlerësimeve në modelet e regresit përdoren disa shuma: *shuma e katrorëve e përgjithshme (SKP)*, *shuma e katrorëve e shkaktuar nga regresi (SKR)*, *shuma e katrorëve e mbetur (SKM)*. Ekuacioni bazë i cili lidh këto shuma me njëra-tjetrën është:

$$SKP(X_1, \dots, X_k) = SKR(X_1, \dots, X_k) + SKM(X_1, \dots, X_k),$$

ku SKP , SKR dhe SKM përcaktohen me formulë si vijon:

²² Njihet ndryshe dhe me termin: *shuma mesatare e katrorëve të mbetur* ose *katrori mesatar i gabimeve*.

²³ Termi i përdorur në anglisht është: *hat values* ose *leverages*.

$$SKP(X_1, \dots, X_k) = \sum_{i=1}^n (Y_i - \bar{Y})^2,$$

$$SKR(X_1, \dots, X_k) = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2,$$

$$SKM(X_1, \dots, X_k) = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2.$$

Në analizën e variancës luajnë rol të rëndësishëm të tre shumat e mësipërme, me anë të cilave përcaktojmë madhësitë si vijon:

- *Katrori mesatar i përgjithshëm*, i cili ka $(n - 1)$ shkallë lirie, jepet nga formula:

$$KMP(X_1, \dots, X_k) = \frac{SKP(X_1, \dots, X_k)}{n - 1}.$$

- *Katrori mesatar i mbetjeve*, i cili ka $(n - k - 1)$ shkallë lirie, jepet nga formula:

$$KMM(X_1, \dots, X_k) = \frac{SKM(X_1, \dots, X_k)}{(n - k - 1)}.$$

- *Katrori mesatar për shkak të regresit*, i cili ka k shkallë lirie, jepet nga formula:

$$KMR(X_1, \dots, X_k) = \frac{SKR(X_1, \dots, X_k)}{k}.$$

Për të kontrolluar domethënien statistikore (vlefshmërinë) të lidhjes midis ndryshores së varur me ato të pavarura, në vend që të kontrollojmë hipotezën se $H_0: \mu_Y = \mu_Y(x_1, \dots, x_k)$ kundrejt $H_a: \mu_Y \neq \mu_Y(x_1, \dots, x_k)$ kontrollojmë $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ kundrejt H_a : “të paktën një prej $\beta_1, \beta_2, \dots, \beta_k$ është i ndryshëm nga zero”. Kriteri statistikor i përdorur në këtë rast është kriteri F , statistika e të cilit jepet nga formula:

$$F_C = \frac{KMR(X_1, \dots, X_k)}{KMM(X_1, \dots, X_k)}.$$

Nëse hipoteza fillestare është e vërtetë, atëherë madhësia F_C ka shpërndarje Fisher $\mathcal{F}(k, n - k - 1)$ me k shkallë lirie për numëruesin dhe $(n - k - 1)$ shkallë lirie për emëruesin. P-vlera e këtij kriteri statistikor është vlera e α e tillë që $F_C = F_{1-\alpha; k, n-k-1}$. Nëse $p \leq \alpha$ (ose $F_C \geq F_{1-\alpha; k, n-k-1}$), atëherë hipoteza H_0 refuzohet.

Nëse pas kryerjes së kriterit F rezultatet tregojnë se lidhja ndërmjet ndryshoreve në regresin linear të shumëfishtë është statistikisht domethënëse, përdorim kriterin t për të përcaktuar domethënien statistikore të secilit prej parametrave më vete. Në këtë rast hipoteza fillestare është $H_0: \beta_i = 0$ kundrejt $H_a: \beta_i \neq 0$. Kriteri statistikor i përdorur në këtë rast është kriteri t , statistika e të cilit jepet nga formula:

$$t_c = \frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}}.$$

Nëse hipoteza fillestare është e vërtetë, madhësia e mësipërme ka shpërndarje t . P-vlera e këtij kriteri statistikor është vlera e α e tillë që $|t_c| = t_{1-\frac{\alpha}{2}; n-k-1}$. Nëse $p \leq \alpha$ (ose $|t_c| \geq t_{1-\frac{\alpha}{2}; n-k-1}$), atëherë hipoteza H_0 refuzohet.

Një tregues tërësor i domethënies së ekuacionit të regresit është *koeficienti i përcaktueshmërisë*, i shumëfishtë $\rho_{Y(X_1, \dots, X_k)}^2$. Në rastin kur kemi të bëjmë me një zgjedhje rastësore të thjeshtë, si vlerësues për këtë koeficient shërben madhësia R^2 e cila vlerësohet nga formula:

$$R^2 = \frac{SKR(X_1, \dots, X_k)}{SKP(X_1, \dots, X_k)},$$

dhe shpreh përpjesën që zë SKR në SKP . Vlerat e këtij koeficienti variojnë midis 0 dhe 1 dhe tregojnë se çfarë pjese ($R^2 \cdot 100\%$) të shpërhapjes së ndryshores së varur shpjegon ekuacioni i regresit [81]. Kjo vlerë nuk mjafton për të arritur në një përfundim mbi rëndësinë statistikore të lidhjes së shprehur nëpërmjet ekuacionit të regresit. Koeficienti i përcaktueshmërisë është i barabartë me katrorin e koeficientit të korrelacionit ndërmjet vlerave të vrojtuar Y_i dhe vlerave të parashikuara nga ekuacioni i regresit $\hat{Y}_i$.

Vlerësimi për *koeficientin e përcaktueshmërisë, të rregulluar* jepet me anë të formulës:

$$R_{rreg}^2 = 1 - \frac{\frac{SKM(X_1, \dots, X_k)}{(n-k-1)}}{\frac{SKP(X_1, \dots, X_k)}{(n-1)}} = 1 - (1 - R^2) \frac{(n-1)}{(n-k-1)}.$$

Ky koeficient është i ndikueshëm nga shtimi i ndryshoreve të pavarura në model. Shtimi i një ndryshore të pavarur në model e rrit vlerën e R_{rreg}^2 vetëm në rast se vlera e kriterit t në atë ndryshore është më e madhe se 1 në vlerë absolute. Kështu që, R_{rreg}^2 mund të përdoret për të vendosur shtimin ose jo të një ndryshoreje në model. Vlerat e R_{rreg}^2 janë të kufizuara nga lart nga 1, por të pakufizuara nga poshtë. Ky koeficient mund të marrë edhe vlera negative.

2.2.4.3. Analiza e mbetjeve

Analiza e mbetjeve luan rol shumë të rëndësishëm për të bërë diagnozën se sa të drejta janë supozimet e bëra në analizën e regresit. Cënimi i linearitetit do të sillte një parashikim shumë të keq të realitetit, ndërsa cënimi i normalitetit ndikon në vlerësimet intervalore të parametrave, në kontrollet e hipotezave për domethënien statistikore. Për të patur vlerësimet më të mira për parametrat e panjohur, popullimi i gabimeve gjatë përqasjes duhet të ketë shpërndarje normale. Ky supozim është shumë i rëndësishëm, veçanërisht në ato raste kur përmasa e zgjedhjes është e vogël. Në momentin që cënohet homogjeniteti i variancës kemi paqëndrueshmëri në vlerësimet e bëra për parametrat e panjohur dhe devijimet standarte. Pavarësia është e rëndësishme sepse nënkupton që gabimet nuk janë të lidhura me njëra-tjetrën. Prania e vrojttimeve me ndikim të lartë

mund të sjell vlerësime të zhvendosura të parametrave dhe të shumave të katrorëve, e për rrjedhojë edhe vlerësime të zhvendosura për devijimet standarte të cilat përdoren në gjetjen e vlerësuesve të parametrave dhe vlerësuesve intervalor. Duke qënë se modeli i regresit përdoret si një përafrim i realitetit, në rastin kur ndonjë prej supozimeve të bëra nuk plotësohet, duhet gjykuar me kujdes për të hedhur poshtë modelin sepse është e vështirë që supozimet të plotësohen në mënyrë rigorozë.

Le të kemi sërish në konsideratë modelin e regresit të popullimit $Y_i = \beta_0 + \beta_1 X_{i,1} + \dots + \beta_k X_{i,k} + E_i$, për $i = 1, \dots, N$. Nëse supozimet e bëra plotësohen, atëherë E_i janë ndryshore rasti të pavarura nga një popullim normal, me vlerë mesatare zero dhe devijim standart konstant σ (homoskedasticiteti). Meqënëse vlerat e E_i janë të panjohura, në vend të tyre analizohen mbetjet $\hat{e}_i$ për $i = 1, \dots, n$, ku n është përmasa e zgjedhjes.

Nëse supozimet e bëra për modelin e regresit plotësohen, mbetjet e standartizuara r_i për $i = 1, \dots, n$, janë ndryshore rasti të pavarura, me shpërndarje normale, me vlerë mesatare zero dhe devijim standart një njësi. Ky fakt mund të përdoret për të kontrolluar supozimet e modelit. Paraqitjet grafike të mbetjeve na ndihmojnë të kontrollojmë modelin dhe të përcaktojmë nëse ndonjë prej supozimeve nuk plotësohet. Edhe nëse procedura e mëposhtme nuk sugjeron se ndonjë prej supozimeve nuk plotësohet, kjo nuk shërben si garanci se të gjitha supozimet e bëra plotësohen.

(1) Grafiku i mbetjeve të standartizuara kundrejt vlerave të përqasura.

Mesatares së vlerësuar të nënpopullimit $\hat{\mu}_Y(x_{i,1}, \dots, x_{i,k})$, së cilës i korrespondon zgjedhja me vlera $(x_{i,1}, \dots, x_{i,k})$, shpesh i referohemi me termat *vlerat e përqasura* ose *përqasjet*. Në rastin kur supozimet plotësohen, vlerat e përqasura janë të pavarura nga devijimet standarte r_i . Kështu që pikat në grafikun e r_i kundrejt $\hat{\mu}_Y(x_{i,1}, \dots, x_{i,k})$ për $i = 1, \dots, n$ do të jenë të shpërndara në mënyrë të njëtrajshme, rastësore rreth boshtit horizontal $r_i = 0$, pa ndjekur ndonjë model të veçantë. Në rast se vihet re ndonjë model i veçantë, ky fakt shërben si tregues se ndonjë ose disa prej supozimeve të bëra nuk kënaqen. Ky grafik është i dobishëm për kontrollin e linearitetit dhe të homoskedasticitetit. Nëse modeli nuk e plotëson supozimin e linearitetit në grafik do të vihet re që mbetjet janë të shpërndara duke ndjekur një model të caktuar, psh. në trajtë parabolë. Në rast se supozimi i linearitetit plotësohet, vija e kuqe (një kurbë polinomiale e lëmuar, e cila na ndihmon të krijojmë idenë për modelin e shpërhapjes së vijave) që kalon përmes pikave në këtë grafik duhet të jetë e drejtë dhe horizontale. Për të kontrolluar supozimin për homoskedasticitetin mbetjet duhet të jenë të shpërndara në mënyrë të njëtrajshme dhe rastësore rrotull drejtëzës $r_i = 0$ e cila paraqitet me vija të ndërprera.

(2) Grafiku Kuantil-Kuantil. Në rastin kur supozimet e bëra në modelin e regresit plotësohen, grafiku i mbetjeve të standartizuara kundrejt kuantileve teorike normale i përafrohet një drejtëze (që paraqitet me vija të ndërprera) e cila kalon në origjinë dhe këndi që formon me boshtin horizontal është 45° . Mbetjet kanë shpërndarje normale nëse ato ndjekin drejtëzën me vija të ndërprera shumë pranë. Përveç mbështetjes në grafik, për të kontrolluar normalitetin është kryer dhe kriteri Shapiro-Wilk [76, 77, 78].

(3) Grafiku i rrënjës katrore të vlerës absolute të mbetjeve të standartizuara kundrejt vlerave të përqasura. Edhe në këtë rast pikat në grafik duhet të jenë të shpërndara në mënyrë rastësore, pa ndjekur ndonjë model të caktuar. Ky

grafik është i dobishëm për të kontrolluar homoskedasticitetin. Nëse vija e kuqe që shohim në grafik është horizontale dhe pikat në të janë të shpërndara në mënyrë rastësore dhe të njëtrajtshme, kjo do të thotë që supozimi për homoskedasticitetin nuk çënohet. Nëse vija e kuqe ka një kënd pozitiv ose pikat nuk janë të shpërndara në mënyrë rastësore dhe të njëtrajtshme, kjo nënkupton se supozimi nuk plotësohet. Përveç mbështetjes në këtë grafik, apo edhe në grafikun e parë, është kryer kontrolli me anë të kriterit Breusch-Pagan [83] të modifikuar nga Cook dhe Weisberg në 1983 [84]. Hipoteza fillestare e këtij kriteri pohon se varianca e gabimeve është konstante dhe statistika e tij ka shpërndarje χ^2 me 1 shkallë lirie. Hipoteza fillestare refuzohet nëse probabiliteti kritik $p \leq \alpha$.

- (4) Grafiku i mbetjeve kundrejt vlerës së krahut të forcës.** Ky grafik na tregon se cilat pika kanë ndikim të madh. Mund të ndodh që përdorimi i krahut të forcës në përcaktimin e ndikimit në ekuacionin e regresit të mos jetë gjithmonë i saktë, sepse mund të kemi krah të lartë force dhe të mos kemi ndikim në ekuacionin e regresit. Pikat me krah të lartë force dhe me ndikim mund ta ndryshojnë rezultatin në varësi të përfshirjes apo përjashtimit të tyre gjatë analizës. Pra, përdorimi i krahut të forcës në përcaktimin e ndikimit në ekuacionin e regresit jo gjithmonë është i saktë. Për këtë arsye, në përcaktimin e vrojtmeve me ndikim, krahas krahut të forcës, përdoret edhe distanca e Cook-ut. Distanca e Cook-ut në vrojtimin e i -të të zgjedhjes jepet nga formula:

$$c_i = \frac{1}{k+1} \left(\frac{h_{i,i}}{1-h_{i,i}} \right) r_i^2,$$

ku $h_{i,i}$ është krah i forcës në njësinë e i -të të zgjedhjes, r_i është mbetja e standartizuar dhe $k+1$ tregon numrin e parametrave të panjohur β në ekuacionin e regresit. Nëse $\hat{\beta}_j$ është vlerësuesi i gjetur për β_j duke përdorur të gjitha të dhënat, shënojmë $\hat{\beta}_{j(-i)}$ vlerësuesin e gjetur për β_j duke përdorur të gjitha të dhënat me përjashtim të vrojtimit të i -të. Atëherë c_i është një madhësi që mat diferencën ndërmjet $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$ dhe $(\hat{\beta}_{0(-i)}, \hat{\beta}_{1(-i)}, \dots, \hat{\beta}_{k(-i)})$. Vlerat e larta për c_i shërbejnë si tregues se vlerësimet për disa ose të gjithë parametrat e panjohur β që do të pësojnë ndryshim në rast se vrojtimi i i -të përjashtohet nga analiza. Përgjithësisht konsiderohen vrojtme me ndikim të madh ato raste në të cilat $c_i \geq F_{0.5:k+1, n-k-1}$. Në grafik, vija e kuqe me ndërprerje përfaqëson distancën e Cook-ut dhe hapësira e cila është subjekt interesi për ne është hapësira jashtë këtyre vijave në këndin e djathtë lart dhe këndin e djathtë poshtë. Nëse ndonjë prej pikave gjendet në këtë pjesë, themi se ai vrojtim ka krah të lartë force dhe ndikim të lartë në model. Në këtë rast, kemi tre mundësi: ose do e justifikojmë përfshirjen e këtij vrojtimi, ose do e përjashtojmë ose do tentojmë një përjasje të re.

Për të kontrolluar nëse mbetjet janë të pavarura nga njëra-tjetra është përdorur kriteri i Durbin-Watson [85, 86, 87]. Hipoteza fillestare e këtij kriteri pohon se mbetjet nuk janë të korreluara. Kriteri statistikor është: $d = \frac{\sum_{i=2}^n (\hat{e}_i - \hat{e}_{i-1})^2}{\sum_{i=1}^n \hat{e}_i^2}$. Vlera kritike e poshtme d_L dhe vlera kritike e sipërme d_U përcaktohen për vlera të ndryshme të k (numri i ndryshoreve të pavarura në model), n (numri i të dhënave në zgjedhje) dhe α (niveli i rëndësisë i zgjedhur) nëpërmjet një tablele [88]. Nëse vlera e $d < d_L$ hipoteza

fillestare refuzohet, nëse $d > d_U$ hipoteza fillestare nuk refuzohet dhe nëse $d_L < d < d_U$ nuk mund të merret një vendim në lidhje me hipotezën fillestare.

2.3. REZULTATET E HULUMTIMIT

2.3.1. “Botë e Vogël”

Disa vlerësime strukturore për graf rrjetat e Modelit 2.1: $G_1, G_2, \dots, G_{30}$ dhe Modelit 2.2: $G_1 \leq G_2 \leq \dots \leq G_{30}$ në lidhje me rendet $|V_i| = n_i$, përmasat $|E_i| = m_i$, fuqinë mesatare të kulmeve $\bar{d}_v^T$, distancën mesatare gjeodezike $\bar{l}_{G_i}$, koeficientin e grupimeve $cl(G_i)$ dhe koeficientin e kalimtaritetit $cl_T(G_i)$ jepen nëpërmjet Tabelës 2.1 dhe Tabelës 2.2.

Tabelë 2.1: Vlerësime për disa madhësi strukturore të graf rrjetave, vlerësimet për “botë e vogël” dhe kufiri i sipërm i intervalit me besueshmëri 99% - $CL^{0.01}$ dhe $L_T^{0.01}$. M është numri i graf rrjetave ekuivalente të rastit $E - R$, të simuluar. Madhësitë e shoqëruara me shenjën (*) janë të rrumbullakosura në të qindtat, ndërsa madhësitë e shoqëruara me (**) janë të rrumbullakosura në të mijtat.

i	$ V_i $	$ E_i $	$\bar{d}_v^{T*}$	$\bar{l}_{G_i}^*$	$cl_T(G_i)^{**}$	$cl(G_i)^{**}$	$S_i^T^*$	$CL_T^{0.01*}$	S_i^*	$CL^{0.01*}$	M
1	1597	2226	2.79	5.75	0.006	0.005	2.78	1.27	2.96	1.65	966
2	1428	1940	2.72	5.99	0.005	0.003	1.70	1.18	2.25	1.38	972
3	1561	2344	3.00	5.38	0.006	0.004	2.74	1.26	2.57	1.32	990
4	1534	2286	2.98	5.37	0.007	0.005	3.46	1.32	2.77	1.43	990
5	1522	2172	2.85	5.72	0.005	0.004	2.25	1.21	2.66	1.41	977
6	1530	2231	2.92	5.52	0.005	0.003	14.16	1.24	23.66	1.26	982
7	1531	2283	2.98	5.44	0.004	0.005	1.49	1.19	2.22	1.47	986
8	1560	2363	3.03	5.35	0.006	0.005	2.80	1.26	4.12	1.49	991
9	1487	2080	2.80	5.72	0.006	0.005	6.48	1.26	11.18	1.64	977
10	1564	2273	2.91	5.50	0.004	0.003	2.92	1.20	1.67	1.28	989
11	1545	2245	2.91	5.59	0.006	0.003	3.87	1.29	1.68	1.30	983
12	1531	2221	2.90	5.52	0.004	0.003	3.32	1.18	1.75	1.32	980
13	1547	2199	2.84	5.70	0.006	0.006	2.82	1.26	3.58	1.70	986
14	1552	2290	2.95	5.60	0.006	0.005	4.41	1.27	4.67	1.45	989
15	1555	2239	2.88	5.58	0.004	0.003	3.57	1.20	4.86	1.30	988
16	1554	2168	2.79	5.78	0.003	0.002	1.39	1.14	0.87	1.18	968
17	1555	2281	2.93	5.55	0.005	0.002	3.12	1.22	1.58	1.25	985
18	1560	2272	2.91	5.53	0.007	0.004	2.22	1.32	1.54	1.43	984
19	1494	2101	2.81	5.70	0.006	0.004	7.63	1.29	9.04	1.45	972
20	1568	2250	2.87	5.62	0.007	0.005	2.99	1.34	2.18	1.57	981
21	1523	2284	3.00	5.41	0.006	0.003	17.78	1.27	38.21	1.32	992
22	1505	2207	2.93	5.43	0.005	0.005	1.82	1.25	2.70	1.46	985
23	1479	2059	2.78	5.81	0.005	0.002	3.60	1.22	1.76	1.24	976
24	1512	2202	2.91	5.46	0.008	0.005	11.09	1.41	9.66	1.52	985
25	1559	2256	2.89	5.46	0.006	0.003	4.07	1.29	2.30	1.38	978
26	1545	2272	2.94	5.41	0.007	0.005	2.38	1.35	3.79	1.62	984
27	1530	2157	2.82	5.60	0.008	0.008	3.41	1.39	6.53	1.87	981
28	1505	2213	2.94	5.43	0.006	0.004	3.74	1.24	5.53	1.44	986
29	1524	2226	2.92	5.46	0.007	0.007	4.67	1.34	10.97	1.67	991
30	1472	2036	2.77	5.98	0.003	0.001	1.72	1.14	0.66	1.11	961

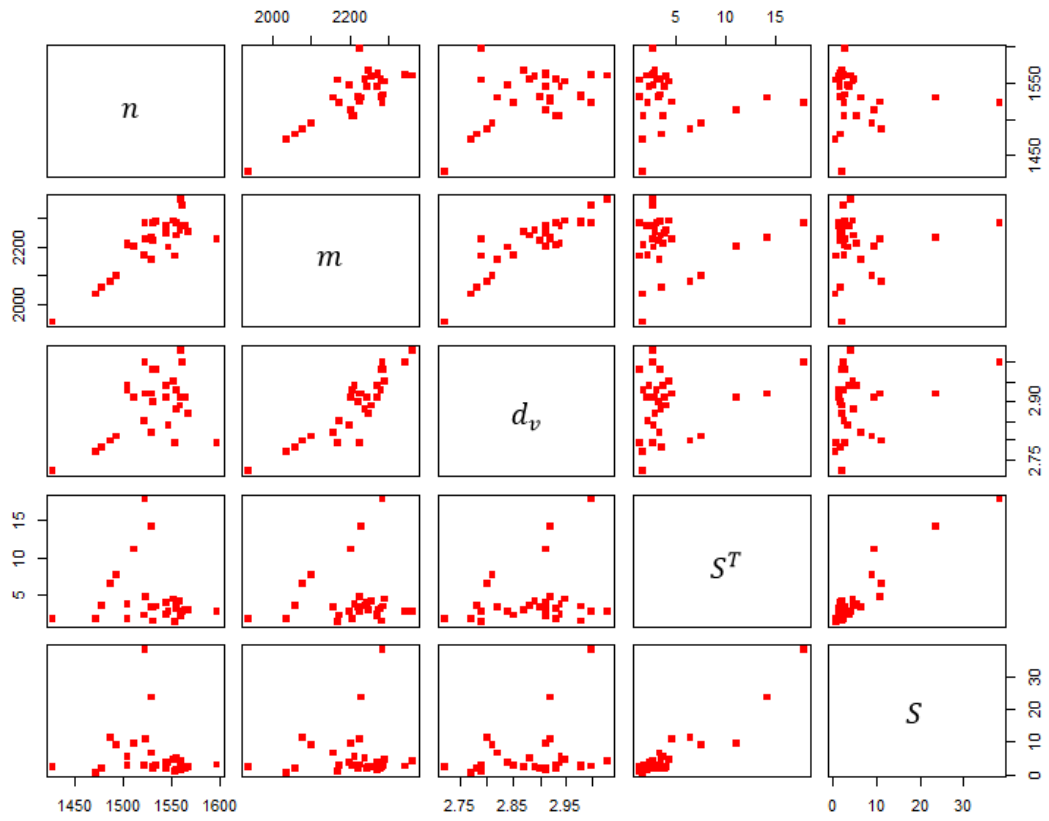
Tabelë 2.2: Vetë strukturore të graf rrjetave, vlerësimet për “botë e vogël” dhe kufiri i sipërm i intervalit me besueshmëri 99% - $CL_T^{0.01}$ dhe $CL^{0.01}$. Numri i graf rrjetave ekuivalente të rastit $E - R$, të simuluar për secilin rast është $M=1000$. Madhësitë e shoqëruara me shenjën (*) janë të rrumbullakosura në të qindtat, ndërsa me shenjën (**) në të mijtat.

i	$ V_i $	$ E_i $	$\bar{d}_v^*$	$\bar{l}_{G_i}^*$	$cl_T(G_i)^{**}$	$cl(G_i)^{**}$	$S_i^T^*$	$CL_T^{0.01*}$	S_i^*	$CL^{0.01*}$
1	1597	2226	2.79	5.75	0.006	0.005	4.64	1.28	8.18	1.58
2	1955	4040	4.13	4.65	0.009	0.009	4.63	1.16	4.44	1.24
3	2226	6238	5.60	4.09	0.014	0.012	5.97	1.10	6.68	1.12
4	2382	8343	7.01	3.78	0.018	0.018	5.80	1.07	6.12	1.09
5	2504	10312	8.24	3.62	0.021	0.023	6.38	1.06	6.29	1.08
6	2592	12296	9.49	3.49	0.024	0.026	6.77	1.05	7.15	1.06
7	2668	14301	10.72	3.38	0.027	0.027	7.17	1.04	7.05	1.04
8	2731	16388	12.00	3.29	0.031	0.031	7.24	1.04	7.54	1.04
9	2781	18203	13.09	3.23	0.034	0.035	7.32	1.03	7.60	1.03
10	2837	20176	14.22	3.18	0.037	0.039	8.39	1.03	8.79	1.03
11	2885	22081	15.31	3.13	0.041	0.041	7.62	1.03	7.87	1.03
12	2928	23964	16.37	3.09	0.044	0.046	7.99	1.02	8.58	1.03
13	2956	25851	17.49	3.05	0.046	0.050	7.87	1.02	8.50	1.02
14	2999	27806	18.54	3.02	0.048	0.052	7.19	1.02	7.78	1.02
15	3026	29688	19.62	2.98	0.051	0.054	7.47	1.02	7.95	1.02
16	3052	31511	20.65	2.96	0.053	0.057	7.64	1.02	8.33	1.02
17	3077	33381	21.70	2.93	0.055	0.061	7.67	1.01	8.48	1.02
18	3112	35240	22.65	2.91	0.057	0.064	7.90	1.01	8.73	1.02
19	3130	36963	23.62	2.89	0.059	0.067	7.76	1.01	8.73	1.02
20	3149	38792	24.64	2.87	0.061	0.069	7.51	1.01	8.39	1.01
21	3167	40659	25.68	2.85	0.064	0.071	7.70	1.01	8.63	1.01
22	3185	42450	26.66	2.83	0.066	0.074	7.89	1.01	8.67	1.01
23	3197	44116	27.60	2.82	0.068	0.076	7.81	1.01	8.78	1.01
24	3213	45876	28.56	2.80	0.071	0.079	8.01	1.01	9.00	1.01
25	3223	47666	29.58	2.79	0.073	0.081	7.79	1.01	8.72	1.01
26	3239	49468	30.55	2.77	0.075	0.084	7.70	1.01	8.67	1.01
27	3255	51195	31.46	2.76	0.077	0.087	7.88	1.01	8.91	1.01
28	3268	52928	32.39	2.74	0.079	0.089	7.90	1.01	8.89	1.01
29	3278	54653	33.35	2.73	0.081	0.091	7.87	1.01	8.82	1.01
30	3287	56259	34.23	2.72	0.083	0.093	7.70	1.01	8.63	1.01

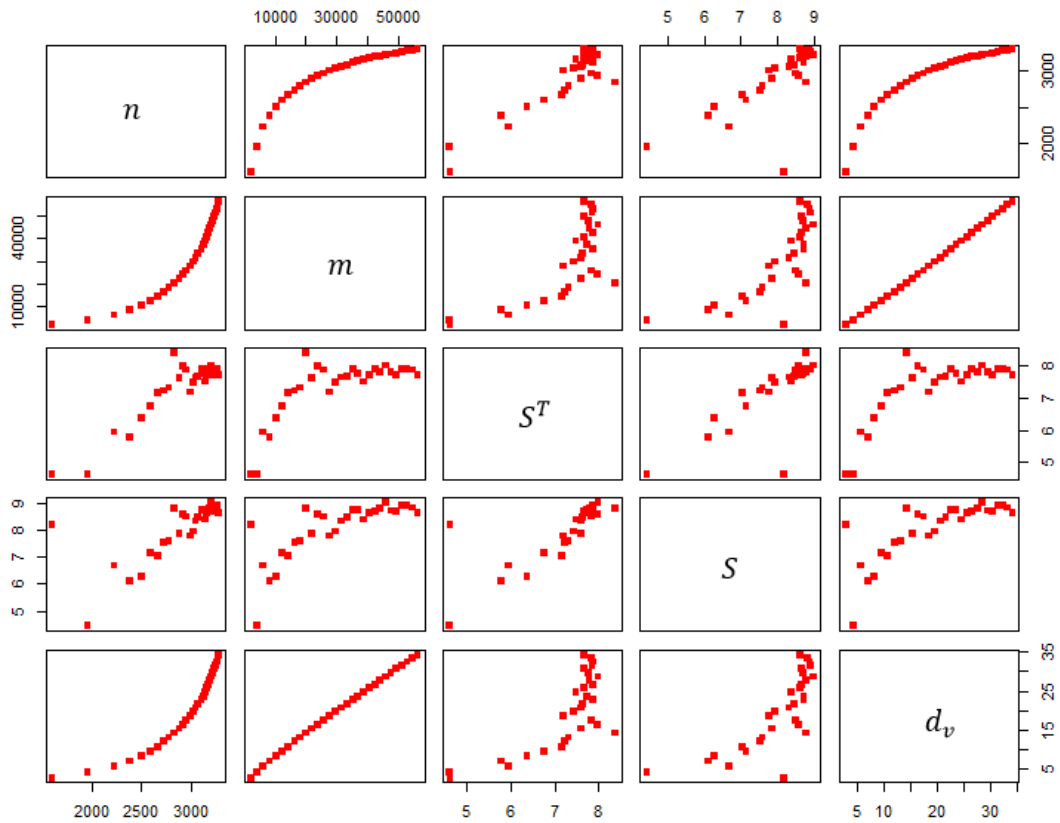
$CL_T^{0.01}$ dhe $CL^{0.01}$ janë njehsuar respektivisht për secilën nga madhësitë që matin “botë e vogël”, S_i^T dhe S_i . Të gjitha graf rrjetat pjesë e Modelit 2.1 nuk janë të lidhura. Në Modelin 2.2 vetëm graf rrjetat $G_1, G_2, \dots, G_{14}, G_{18}$ nuk janë të lidhura, ndërsa $G_{15}, G_{16}, G_{17}, G_{19} \dots, G_{30}$ janë të lidhura.

2.3.2. Korrelacioni Joparametrik

Figura 2.2 dhe Figura 2.3 tregojnë grafikët e pesëfishtë, me pika, të lidhjeve: $S^T \sim n$, $S \sim n$, $S^T \sim m$, $S \sim m$, $S^T \sim \bar{d}_v$, $S \sim \bar{d}_v$ dhe $m \sim n$.



Figurë 2.2: Grafiku i pesëfishtë me pika për: rendin, përmasën, fuqinë mesatare dhe “botë e vogël” sipas të dy përcaktimeve, Modeli 2.1.



Figurë 2.3: Grafiku i pesëfishtë me pika për: rendin, përmasën, fuqinë mesatare dhe “botë e vogël” sipas të dy përcaktimeve, Modeli 2.2.

Tabela 2.3 tregon rezultatet e gjetura pas kryerjes së kriterit Shapiro-Wilk për kontrollin e normalitetit të të dhënave që kanë të bëjnë me: rendin, përmasën, fuqinë mesatare të kulmeve dhe madhësinë “botë e vogël”, sipas të dy përcaktimeve. Në tabelë janë theksuar vlerat të cilat janë më të vogla se niveli $\alpha = 0.01$.

Tabelë 2.3: P-vlerat e gjetura gjate kontrollit të normalitetit nëpërmjet kriterit Shapiro-Wilk.

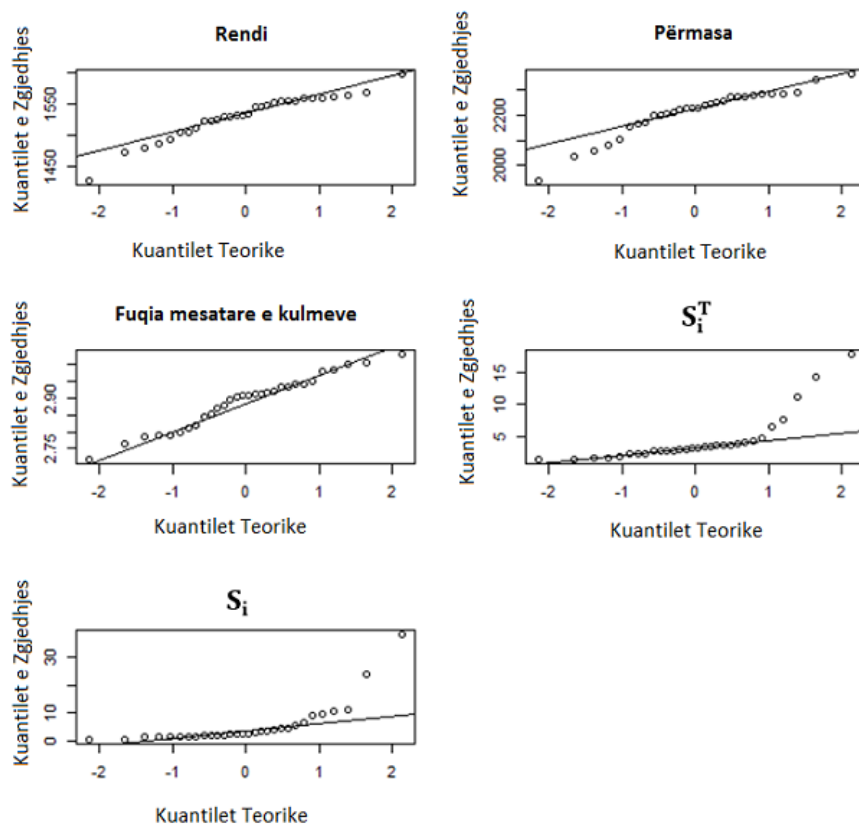
	Modeli 2.1	Modeli 2.2
n	0.0757	0.0002095
m	0.014	0.261
$\bar{d}_v$	0.5064	0.2989
S	4.225×10^{-08}	4.48×10^{-05}
S^T	4.275×10^{-07}	9.189×10^{-06}

Hulumtimi i normalitetit nëpërmjet grafikëve Kuantil-Kuantil jepet nga Figura 2.4 dhe Figura 2.5. Rezultatet e gjetura nga kriteri Shapiro-Wilk në nivelin e rëndësisë $\alpha = 0.01$, konvergjojnë me ato të gjetura nga grafikët Kuantil-Kuantil. Hipoteza se: H_0 : “Të dhënat kanë shpërndarje normale”, në Modelin 2.1 nuk u refuzua vetëm në lidhje me rendin dhe fuqinë mesatare të kulmeve, por dhe në lidhje me përmasën dhe fuqinë mesatare të kulmeve në rastin e Modelit 2.2.

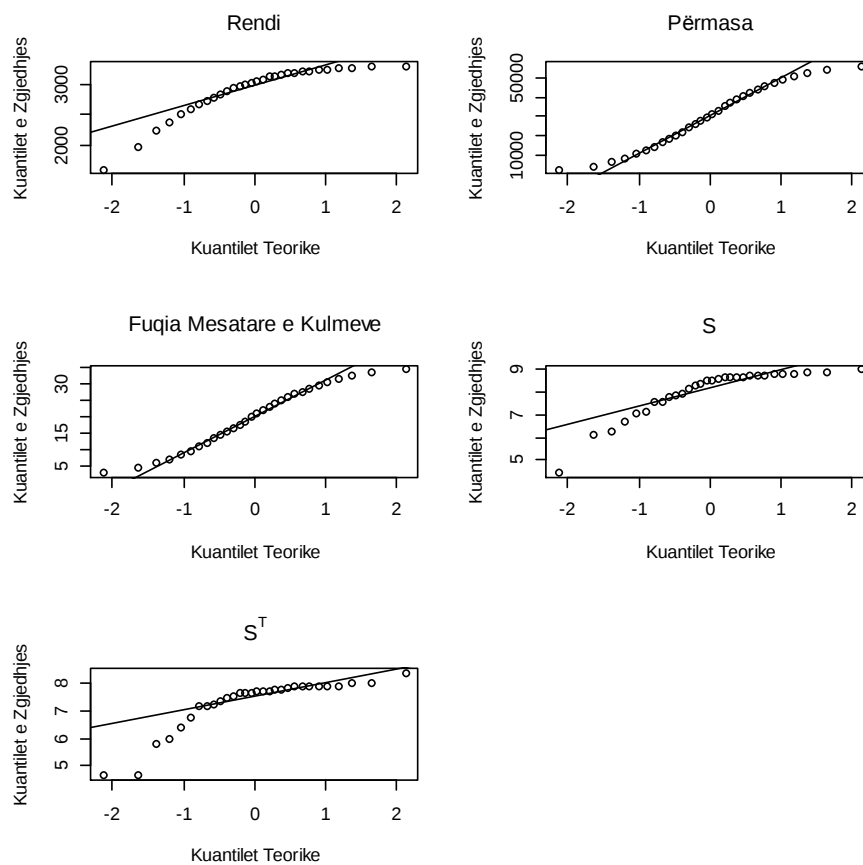
Duke qënë se nuk ekziston një shpërndarje normale dypërmasore në krahasimet e bëra, si vijon: $S^T \sim n$, $S \sim n$, $S^T \sim m$, $S \sim m$, $S^T \sim \bar{d}_v$, $S \sim \bar{d}_v$ dhe $m \sim n$, në vijim është përdorur analiza korrelative joparametrike. Më pas, janë njehsuar dy tipet e koeficientëve të korrelacionit, $\hat{\tau}$ -koeficienti Kendall dhe $\hat{\rho}$ - koeficienti Spearman, vlerat e të cilëve janë të pasqyruara në Tabelën 2.4. Në tabelë, janë theksuar ato vlera të cilat janë më të vogla se niveli i rëndësisë $\alpha = 0.01$.

Tabelë 2.4: Rezultatet e kontrollit joparametrik të korrelacionit për Modelin 2.1 dhe Modelin 2.2.

		Kendall		Spearman	
		$\hat{\tau}$	p - vlera	$\hat{\rho}$	p - vlera
Modeli 2.1	$S^T \sim n$	-0.1319476	0.3087	-0.1892253	0.3166
	$S \sim n$	-0.196764	0.129	-0.2793856	0.1349
	$S^T \sim \bar{d}_v$	0.1034483	0.4358	0.1372636	0.4679
	$S \sim \bar{d}_v$	0.1218391	0.3567	0.2378198	0.2049
	$S^T \sim m$	0.0437789	0.7345	0.05607477	0.7685
	$S \sim m$	-0.002304	0.9858	-0.0015576	0.9935
	$m \sim n$	0.5127665	7.885×10^{-05}	0.6789134	3.717×10^{-05}
Modeli 2.2	$S^T \sim n$	0.5641713	1.318×10^{-05}	0.7028051	1.488×10^{-05}
	$S \sim n$	0.6551366	3.993×10^{-07}	0.822744	2.416×10^{-08}
	$S^T \sim \bar{d}_v$	0.5641713	1.318×10^{-05}	0.7028051	1.488×10^{-05}
	$S \sim \bar{d}_v$	0.6551366	3.993×10^{-07}	0.822744	2.416×10^{-08}
	$S^T \sim m$	0.5641713	1.318×10^{-05}	0.7028051	1.488×10^{-05}
	$S \sim m$	0.6551366	3.993×10^{-07}	0.822744	2.416×10^{-08}
	$m \sim n$	1	1.11×10^{-15}	1	2.2×10^{-16}



Figurë 2.4: Hulumtim pamor për normalitetin e të dhënave, Modeli 2.1.



Figurë 2.5: Hulumtim pamor për normalitetin e të dhënave, Modeli 2.2.

2.3.3. Lidhja ndërmjet Përmasës dhe Rendit

Në Tabelën 2.4 në seksionin 2.3.2 vlerat e gjetura për korrelacionin ndërmjet përmasës dhe rendit janë statistikisht domethënëse. Lidhja është e fortë në të njëjtin drejtim në rastin e Modelit 2.1 dhe perfekte në të njëjtin drejtim në rastin e Modelit 2.2. Pas disa përjasjeve të mundshme për të modeluar varësinë që ekziston ndërmjet përmasës dhe rendit, dolëm në përfundimin se:

- Përjasja më e mirë për Modelin 2.1 është e formës: $m^5 = \beta_0 + \beta_1 n + \beta_2 n^2 + \beta_3 n^3$, e cila trajtohet si regres linear i shumëfishtë $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$, ku $Y = m^5$; $X_1 = n$; $X_2 = n^2$; $X_3 = n^3$.
- Përjasja më e mirë për Modelin 2.2 është e formës: $\sqrt{m} = \beta_0 + \beta_1 n^6 + \beta_2 n^7 + \beta_3 n^8$, trajtohet si regres linear i shumëfishtë $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$, ku $Y = \sqrt{m}$; $X_1 = n^6$; $X_2 = n^7$; $X_3 = n^8$.

Niveli i rëndësisë i zgjedhur prej nesh gjatë kontrollit të hipotezave në vijim në këtë pjesë është $\alpha = 0.05$. Gjatë përjasjes së bërë për Modelin 2.1, vlerësimi për gabimin standart të mbetjeve të modelit është $\hat{\sigma}^2 = 5.727 \times 10^{15}$ në 26 shkallë lirie. Vlerësimi për koeficientin e përcaktueshmërisë është $R^2 = 0.7231$ dhe për koeficientin e përcaktueshmërisë të rregulluar është $R_{reg}^2 = 0.6911$. Vlera e statistikës F është e barabartë me $F_C = 22.63$ në 3 dhe 26 shkallë lirie. P-vlera e gjetur për kriterin e Fisherit është $p = 2.03 \times 10^{-07} < \alpha$.

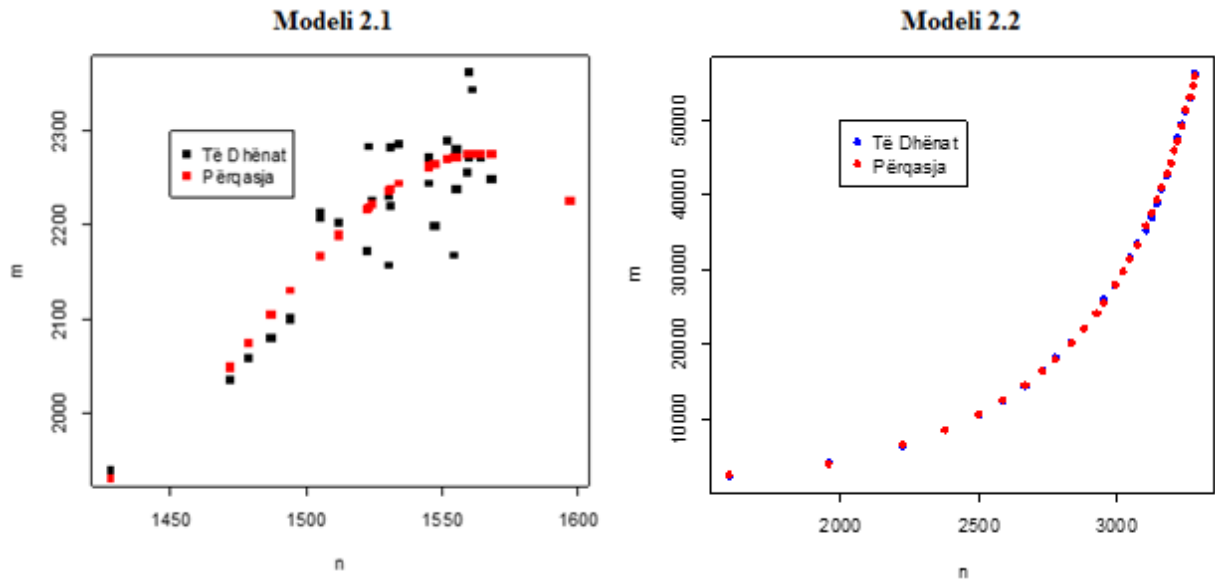
Ndërkohë, gjatë përjasjes së bërë për Modelin 2.2, gabimi standart i mbetjeve të modelit është $\hat{\sigma}^2 = 0.8751$ në 26 shkallë lirie. Vlera e koeficientit të përcaktueshmërisë është $R^2 = 0.9998$ dhe për koeficientin e përcaktueshmërisë të rregulluar $R_{reg}^2 = 0.9997$. Vlera e statistikës F është e barabartë me $F_C = 3.636 \times 10^{04}$ në 3 dhe 26 shkallë lirie. P-vlera e gjetur për kriterin e Fisherit është $p < 2.2 \times 10^{-16}$.

Në bazë të rezultateve të gjetura me anë të kriterit t për të dy modelet, të cilat tregohen në Tabelën 2.5, në të gjitha rastet rezulton se parametrat e vlerësuar janë statistikisht domethënëse në nivelin e rëndësisë të përcaktuar prej nesh.

Tabelë 2.5: Rezultatet për vlerësimet dhe domethënien statistikore të parametrave për përjasjen në Modelin 2.1 dhe Modelin 2.2.

Modeli	Koeficienti	Vlerësimi	Gabimi Standart	t	p-vlera
2.1	$m^5 = \beta_0 + \beta_1 n + \beta_2 n^2 + \beta_3 n^3$				
	β_0	8.157×10^{19}	3.930×10^{19}	2.075	0.0480
	β_1	-1.644×10^{17}	7.803×10^{16}	-2.107	0.0449
	β_2	1.103×10^{14}	5.162×10^{13}	2.137	0.0422
	β_3	-2.461×10^{10}	1.138×10^{10}	-2.163	0.0399
2.2	$\sqrt{m} = \beta_0 + \beta_1 n^6 + \beta_2 n^7 + \beta_3 n^8$				
	β_0	3.772×10^{01}	1.018×10^{00}	37.06	$< 2 \times 10^{-16}$
	β_1	1.714×10^{-18}	8.295×10^{-20}	20.66	$< 2 \times 10^{-16}$
	β_2	-9.163×10^{-22}	5.171×10^{-23}	-17.72	4.92×10^{-16}
	β_3	1.347×10^{-25}	8.151×10^{-27}	16.53	2.62×10^{-15}

Figura 2.6 tregon grafikisht përjasjet e gjetura për varësinë teorike të vlerësuar si më e mira në raport me atë empirike.

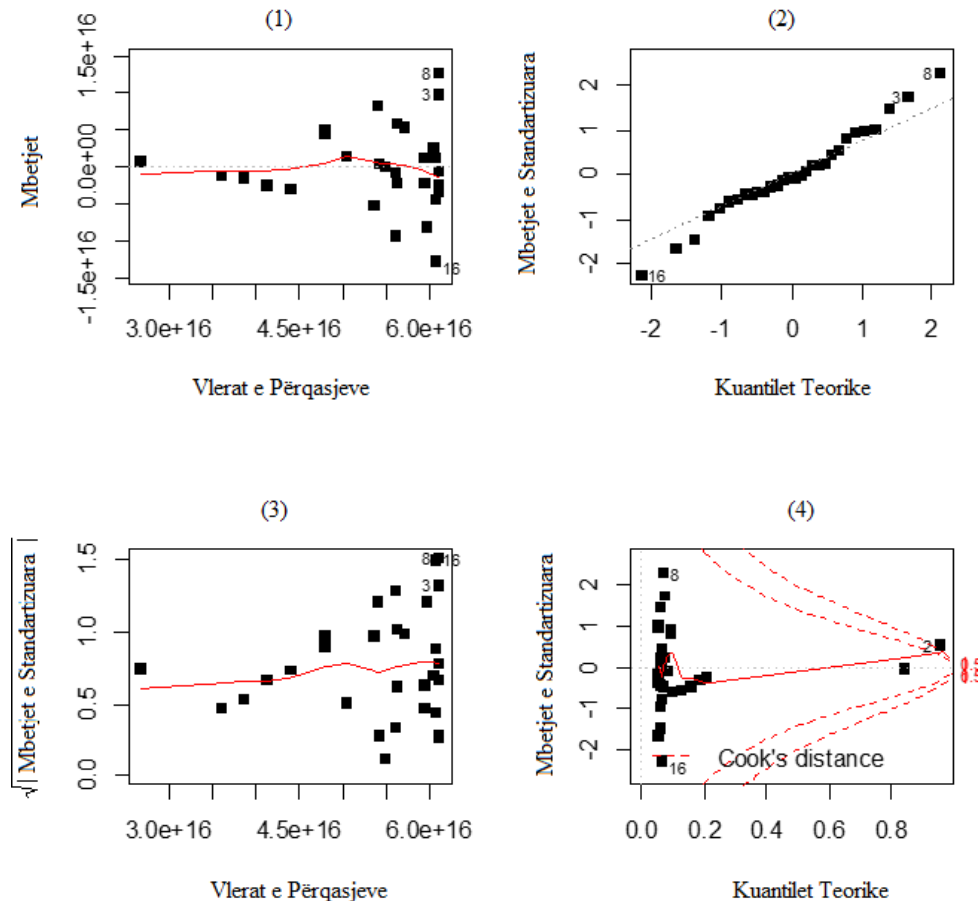


Figurë 2.6: Paraqitja grafike për përqasjet në Modelin 2.1 dhe Modelin 2.2.

Diagnoza e modelit nëpërmjet analizës së mbetjeve jepet grafikisht si më poshtë në Figurën 2.7 dhe Figurën 2.8. Vijat e kuqe në grafik, na ndihmojnë të krijojmë idenë e modeleve që mund të ndjek shpërhapja e pikave.

Komente për Figurën 2.7:

- **(1)** Vemë re se vija e kuqe është afërsisht e drejtë dhe horizontale, pikat nuk ndjekin ndonjë model të caktuar në shpërhapjen e tyre. Ky fakt, konfirmon supozimin për linearitetin. Vrojtimi i 3, 8 dhe 16 dallojnë për vlera të mëdha të mbetjeve. Pikat nuk janë të shpërndara në mënyrë të njëtrajtshme dhe rastësore përreth $r_i = 0$ në të gjithë hapësirën e vlerave të përqasjeve. Kjo nënkupton se devijimet standarte të vlerave të $Y = m^5$ në nënpopullime të ndryshme, pësojnë rritje me rritjen e vlerave të përqasjeve.
- **(2)** Mund të themi se mbetjet kanë shpërndarje normale pasi vihet re se pikat ndjekin drejtëzën me vija të ndërprera, me përjashtim ndoshta në rastin e vrojtimit të 3, 8 dhe të 16. Sipas p -vlerës së gjetur nga kriteri Shapiro-Wilk ($p = 0.7355$) nuk refuzohet hipoteza se mbetjet e modelit kanë shpërndarje normale.
- **(3)** Vrojtimet të cilat dallojnë për vlera të mëdha të mbetjeve janë sërish 3, 8 dhe 16. Vija e kuqe është afërsisht horizontale, por pikat nuk janë të shpërndara njëtrajtësisht. Kjo nënkupton që homoskedasticiteti mund të mos plotësohet, por në bazë të kontrollit me anë të kriterit Breusch-Pagan/Cook dhe Weisberg ($p = 0.06953466 > 0.01$), hipoteza se varianca e gabimeve është konstante nuk mund të refuzohet.
- **(4)** Vihet re se vrojtimi i 2-të është jashtë vijave të kuqe të ndërprera, në këndin e djathtë lart. Ai konsiderohet si vrojtimi me më shumë ndikim në ekuacionin e regresit, ndonëse vrojtimi 2 nuk ka spikatur në grafikët në Figurën 2.7.

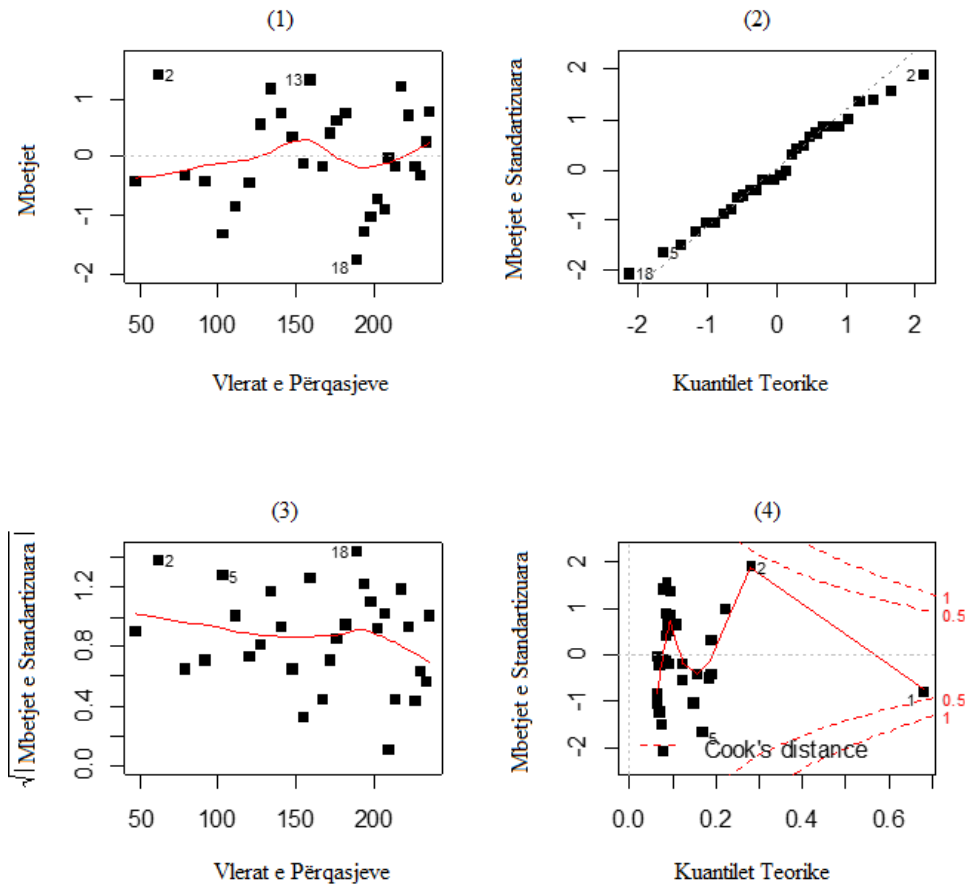


Figurë 2.7: Diagnoza për $m^5 = \beta_0 + \beta_1 n + \beta_2 n^2 + \beta_3 n^3$ në Modelin 2.1.

Komente për Figurën 2.8:

- (1) Vrojtimi i 2, 13 dhe 18 dallojnë si vrojtme të cilat kanë vlera të mëdha të mbetjeve. Pikat janë të shpërndara afërsisht në mënyrë të njëtrajtshme dhe rastësore përreth $r_i = 0$, duke pranuar supozimin për homoskedasticitetin. Vihet re se vija e kuqe është deri diku horizontale, që nënkupton se pikat nuk ndjekin ndonjë model të caktuar në shpërhapjen e tyre. Ky fakt konfirmon supozimin për linearitetin.
- (2) Në këtë rast, pikat ndjekin drejtëzën me vija të ndërprera, me përjashtim ndoshta në rastin e vrojtimit të 2 dhe të 18. Sipas p -vlerës së gjetur nga kriteri Shapiro-Wilk ($p = 0.6903$), nuk refuzohet hipoteza se mbetjet e modelit kanë shpërndarje normale.
- (3) Në këtë rast dallojnë vrojtmet 2, 5 dhe 18. Vija e kuqe është afërsisht horizontale deri në vlerën 200 të përqasjeve, më pas shikohet një trend shumë i lehtë zbritës i saj. Pikat duken sikur janë të shpërndara në mënyrë të njëtrajtshme. Deri tek vlera 200 mund ta konsiderojmë homoskedasticitetin të plotësuar. Në bazë të kontrollit me anë të kriterit Breusch-Pagan/Cook dhe Weisberg ($p = 0.7198454 > 0.01$), hipoteza se varianca e gabimeve është konstante nuk mund të refuzohet.
- (4) Nuk ka vrojtme që të gjenden jashtë vijave të kuqe të ndërprera. Të gjitha vrojtmet janë brenda distancës së Cook-ut. Pavarësisht se në

grafikët në Figurën 2.8 u dalluan disa vrojtime, por ato nuk kanë ndikim shumë të madh në ekuacionin e regresit.



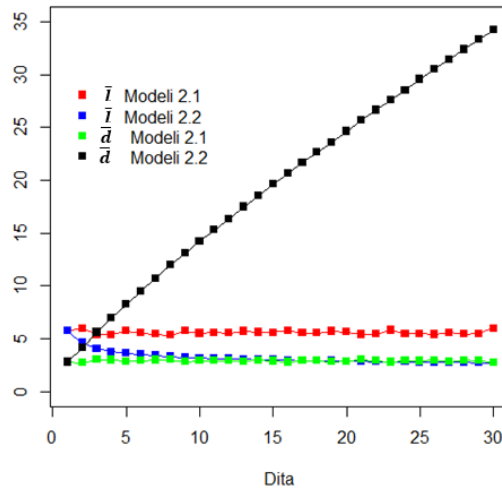
Figurë 2.8: Diagnoza për $\sqrt{m} = \beta_0 + \beta_1 n^6 + \beta_2 n^7 + \beta_3 n^8$ në Modelin 2.2.

Duke qënë se në përqsjet për të dy modelet, numri i ndryshoreve të pavarura është i njëjtë, $k = 3$, $n = 30$ vlerat e kufirit të poshtëm dhe të sipërm me nivel rëndësie $\alpha = 0.01$ për statistikën e kriterit të Durbin-Watson janë $d_L = 1.006$ dhe $d_U = 1.421$. Në lidhje me supozimin për pavarësinë e gabimeve nga njëra-tjetra, në përqsjen për Modelin 2.1 kemi $d = 1.875 > d_U$. Për rrjedhoj, hipoteza fillestare se mungon korrelacioni ndërmjet mbetjeve nuk mund të refuzohet. Në përqsjen për Modelin 2.2 kemi $d = 1.246$, $d_L < d < d_U$, për rrjedhoj nuk mund të dalim në një përfundim në lidhje me refuzimin apo mos refuzimin e hipotezës fillestare.

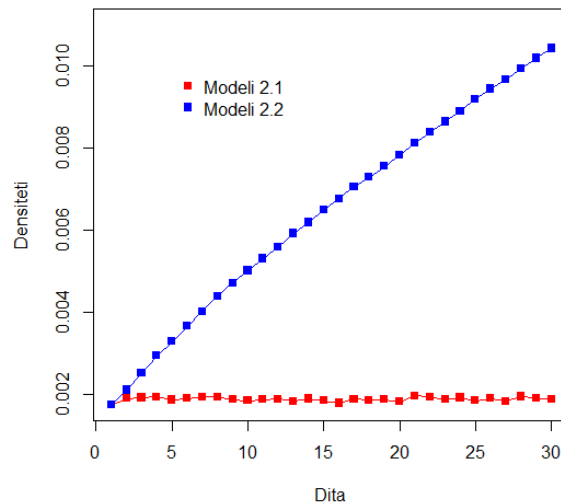
2.3.4. Distanca Mesatare Gjeodezike, Fuqia Mesatare dhe Densiteti

Evoluimi i vlerave në lidhje me distancën mesatare gjeodezike $\bar{l}$, fuqinë mesatare të kulmeve $\bar{a}_v$ në graf rrjetat e serive (Modeli 2.1 dhe Modeli 2.2) tregohet në Figurën 2.9. Në graf rrjetat e Modelit 2.1 dhe Modelit 2.2, në bazë të kriterit Shapiro-Wilk, gjatë kontrollit me nivel rëndësie 0.01, hipoteza se të dhënat kanë shpërndarje normale u refuzua në rastin e distancës mesatare gjeodezike dhe nuk u refuzua në rastin e fuqisë mesatare dhe densitetit.

Për rrjedhoj, sipas teoremës së Chebyshev, të paktën 88.9% e vlerave të distancës mesatare gjenden në intervalin $5.58 \pm 3 \cdot 0.17$ për Modelin 2.1 dhe në $3.20 \pm 3 \cdot 0.65$ për Modelin 2.2. Në rastin e fuqisë mesatare të graf rrjetave, 95% e vlerave gjenden në intervalin $2.89 \pm 1.96 \cdot 0.08$ për Modelin 2.1 dhe në $19.6 \pm 1.96 \cdot 9.4$ për Modelin 2.2. Në lidhje me densitetin e graf rrjetave, 95% e vlerave gjenden në intervalin $0.002 \pm 1.96 \cdot 4.9 \cdot 10^{-05}$ për Modelin 2.1 dhe në $0.006 \pm 1.96 \cdot 0.003$ për Modelin 2.2. Ndryshimi në kohë i distancës mesatare gjeodezike dhe fuqisë mesatare jepet grafikisht në Figurën 2.9, ndërsa ndryshimi i densitetit në Figurën 2.10.



Figurë 2.9: Distanca mesatare gjeodezike dhe fuqia mesatare e kulmeve në lidhje me kohën.



Figurë 2.10: Densiteti i graf rrjetave në lidhje me kohën.

2.4. DISKUTIME

Vlerat për madhësinë “botë e vogël” të njehsuara sipas S_i^T , janë më të mëdha se një, në 100% të çasteve të kohës në të dy modelet e trajtuara. Domethënia e këtij rezultati u konfirmua edhe gjatë kontrollit të hipotezës fillestare, e cila supozonte se graf rrjetat reale të serive kohore ishin graf rrjeta rasti. Gjatë kontrollit vlerat e $S_i^T > CL^{0.01}$ në 100% të çasteve të kohës për të dy modelet. Ndërkohë, vlerat e madhësisë S_i rezultuan

më të mëdha se një, vetëm në 93.3% të rasteve në Modelin 2.1 dhe në 100% të rasteve në Modelin 2.2. Në të njëjtën përqindje respektivisht për secilin prej modeleve, vlerat e S_i ishin më të mëdha se $CL^{0.01}$.

Është shumë e vështirë të krahasosh rezultatet e gjetura për “botë e vogël”, në këtë pjesë të studimit, me studime të tjera të ngjashme në graf rrjeta të telefonisë fikse apo lëvizëse. Në ato studime që ekzistojnë për të cilat jemi në dijeni, nuk janë njehsuar të gjitha madhësitë të cilat ne shqyrtuam në këtë pjesë. Edhe në rast se janë njehsuar madhësitë për të cilat interesohemi, krahasimi është domethënës vetëm nëse teknika e ndërtimit të graf rrjetit do të ishte afërsisht e njëjtë.

Në Modelin 2.2 lidhja e “botë e vogël” me: rendin, përmasën dhe fuqinë mesatare; ishte e fortë dhe statistikisht domethënëse në të njëjtin drejtim. Në Modelin 2.2 nuk kishte korrelacion statistikisht domethënës ndërmjet këtyre madhësive. Korrelacioni ndërmjet përmasës dhe rendit ishte pozitiv, në të njëjtin drejtim, shumë i fortë në Modelin 2.1 dhe perfekt në Modelin 2.2. Lidhjet e gjetura nëpërmjet regresit linear ndërmjet përmasës dhe rendit ishin statistikisht domethënëse. Në Modelin 2.1 lidhja shpjegonte 72.31% të shpërhapjes, ndërsa në Modelin 2.2 shpjegonte 99.98%.

Fuqia mesatare e kulmeve në graf rrjetat e Modelit 2.1 rezultoi mesatarisht 2.89 dhe 19.6 në Modelin 2.2. Fuqitë mesatare të graf rrjetave të telefonisë lëvizëse, me të dhëna nga katër zona të ndryshme shqyrtuar nga Nanavati *et al.* [18], ishin 8.09, 3.6, 4.35, 3.57. Ndërkohë, në graf rrjetin e telefonisë lëvizëse të analizuar nga Noka (Jani) dhe Hoxha [21] fuqia mesatare ishte 2.4.

Në [58], Backstrom *et al.* kanë treguar se graf rrjeti Facebook kishte fuqi mesatare 4.74. Leskovec dhe Horvitz në [89] kanë treguar se graf rrjeti i sistemit të mesazheve direkte në Microsoft Messenger kishte $\bar{l} = 6.6$. Dong *et al.* [20] pohuan se komponenti më i madh i një graf rrjeti social statik të telefonisë lëvizëse kishte $\bar{l} = 5.75$. Ata pohuan se kishin të bënin me një graf rrjetë “botë e vogël”.

Në Modelin 2.1 densiteti është thujse në të njëjtat nivele gjatë gjithë ditëve të muajit. Në Modelin 2.2 kemi një rritje pozitive, thujse lineare të vlerave të densitetit me kalimin e kohës.

2.5. PËRFUNDIME

Në përfundim, mund të themi se prania e fenomenit “botë e vogël” konfirmohet në çdo hap të kohës në seritë kohore të graf rrjetave të marra në shqyrtim, ndërkohë që gjatë njehsimit të madhësisë që mat në mënyrë sasiore këtë fenomen u përdor koeficienti i kalimtaritetit. Gjatë përdorimit të koeficientit të grupimeve, prania e fenomenit konfirmohet vetëm në 93.3% të rasteve në bazë kohore ditore dhe në 100% të rasteve në bazë kohore intervalore. Kjo nënkupton se, informacioni do të përhapej në mënyrë të shpejtë në graf rrjet si brenda një dite (Modeli 2.1), apo edhe brenda disa ditëve (Modeli 2.2). Korrelacioni për “botë e vogël” në lidhje me: rendin, përmasën dhe fuqinë mesatare të kulmeve; në bazë kohore ditore mungonte, por ishte i pranishëm në bazë kohore intervalore. Lidhja e gjetur ndërmjet përmasës dhe rendit në graf rrjeta mund të përdoret për të parashikuar numrin e çifteve të ndryshme të klientëve të operatorit, që komunikojnë brenda një dite ose disa ditësh, në bazë të numrit të klientëve aktiv që ai ka.

3. ANALIZA E MBIJETESËS NË LIDHJE ME KOHËZGJATJEN E THIRRJEVE

3.1. HYRJE

Analiza e mbijetesës është një koleksion procedurash që synojnë analizën e të dhënave në të cilat ndryshorja rezultante për të cilën jemi të interesuar është koha përgjatë së cilës një ngjarje ndodh [90]. Në rastin tonë, me ngjarje nënkuptohet ndërprerja e thirrjes telefonike. Koha derisa përfundon thirrja është koha e mbijetesës, e cila matet në sekonda. Nëse koha derisa ngjarja ndodh njihet saktësisht, të dhënat janë të pacensuruara.

Një graf rrjet $G = (V, E)$ është ndërtuar në bazë të regjistrimeve të thirrjeve të një telefonie fikse, përgjatë një muaji. Çdo brinjë përfaqëson një lidhje komunikuese ndërmjet dy klientëve aktiv të telefonisë dhe është e shoqëruar me informacionin për kohëzgjatjen respektive të thirrjes. Përveç kësaj, disa ndryshore të tjera shpjeguese i janë bashkangjitur çdo brinjë, sikurse janë: dita e muajit, dita e javës, ora e ditës dhe grupi (që nënkupton kategorinë e telefonatave *am*²⁴ ose *pm*²⁵). Këto ndryshore shpjeguese i referohen momentit që thirrja telefonike fillon. Shpërndarja e kohëzgjatjeve të thirrjeve është shumë *asimetrike* (*e anuar*)²⁶ e për rrjedhojë statistika përshkruese si mesatarja dhe varianca nuk janë shumë të dobishme për të përshkruar kohëzgjatjen e tyre.

calldate	callend	connect_duration	caller	called	dita_e_muajit	event	ora	grupi	dita_javes
11/02/2014 00:20	11/02/2014 00:21	17	1020	1913	2	1	0	1	7
11/02/2014 00:20	11/02/2014 00:21	17	1020	1913	2	1	0	1	7
11/04/2014 00:51	11/04/2014 00:52	12	1373	1396	4	1	0	1	2
11/07/2014 00:13	11/07/2014 00:16	191	471	2892	7	1	0	1	5
11/07/2014 00:13	11/07/2014 00:16	191	471	1244	7	1	0	1	5
11/07/2014 00:18	11/07/2014 00:21	175	789	503	7	1	0	1	5
11/07/2014 00:18	11/07/2014 00:21	175	789	103	7	1	0	1	5
11/09/2014 00:26	11/09/2014 00:26	22	548	188	9	1	0	1	7
11/12/2014 00:27	11/12/2014 00:27	12	2884	476	12	1	0	1	3
11/15/2014 0:25	11/15/2014 0:26	44	1224	657	15	1	0	1	6
11/17/2014 0:22	11/17/2014 0:23	73	720	2500	17	1	0	1	1
11/17/2014 0:41	11/17/2014 0:41	16	491	1037	17	1	0	1	1
11/19/2014 0:30	11/19/2014 0:31	54	2578	1936	19	1	0	1	3
11/21/2014 0:11	11/21/2014 0:13	91	1712	892	21	1	0	1	5
11/21/2014 0:28	11/21/2014 0:29	21	649	720	21	1	0	1	5
11/22/2014 0:27	11/22/2014 0:33	351	2729	2348	22	1	0	1	6
11/27/2014 0:45	11/27/2014 0:45	27	2100	1681	27	1	0	1	4
11/27/2014 0:45	11/27/2014 0:46	21	2578	350	27	1	0	1	4
11/27/2014 0:45	11/27/2014 0:46	21	2578	350	27	1	0	1	4
11/27/2014 0:51	11/27/2014 0:51	14	1	743	27	1	0	1	4

Figurë 3.1: Pamje e të dhënave të komunikimit telefonik, të cilat u përdorën për të kryer analizën e mbijetesës. Të dhënat gjenden në [38].

²⁴ Shkurtim nga anglishtja i *Ante meridiem*, nënkuptojmë në shqip *paradite*, përfshin intervalin kohor 00:00 - 11:59.

²⁵ Shkurtim nga anglishtja i *Post meridiem*, nënkuptojmë në shqip *pasdite*, përfshin intervalin kohor 12:00 - 23:59.

²⁶ Përshtatur nga anglishtja: Skewed.

Në këtë pjesë trajtohet problemi i shtruar si vijon nëpërmjet pyetjeve:

- **P_{3.1}:** Sa përcaktues mund të jetë momenti i fillimit të thirrjes për kohëzgjatjen e saj?
- **P_{3.2}:** Si ndikojnë në kohëzgjatjen e thirrjes faktorët: dita e muajit, dita e javës, ora e ditës dhe grupi? Cili model mund ta përshkruaj më mirë këtë marrëdhënie?

Për të zgjidhur këtë problem është përdorur analiza e mbijetesës joparametrike dhe Cox gjysëmparametrike. Rreziku proporcional dhe modelet Cox të shtresëzuara²⁷ i janë përqasur bashkësisë së të dhënave. Kjo problematikë është trajtuar prej nesh edhe në [91].

Aplikime të tjera të analizës së mbijetesës gjenden në trajtimet klinike [92, 93], në studimin e parashikimeve të rreziqeve financiare [94], në shqyrtimin e efekteve të mekanizmave të brendshëm në qeverisjen e korporatave, në lidhje me falimentimin e firmave [95] etj. Aplikime të analizës së mbijetesës janë parë gjithashtu edhe në analizën e të dhënave të komunikimit telefonik [31, 96].

3.2. METODOLOGJIA

3.2.1. Terminologji dhe Kuptime Bazë

Ngjarja për të cilën jemi të interesuar është ndërprerja e thirrjes telefonike. Çdo thirrje telefonike është në korrespondencë një për një me një brinjë të graf rjetit. Me jetëgjatësi të një brinje nënkuptojmë kohëzgjatjen e thirrjes telefonike që i korrespond. Kështu, në vijim me kohë mbijetese [90] i referohemi një ndryshore rasti kohore, duke qenë se ajo tregon kohën gjatë së cilës një thirrje telefonike i mbijeton ndërprerjes së saj. Kjo ndryshore rasti është shënuar me T .

Në përgjithësi $T \geq 0$, por në këtë studim për shkak edhe të mënyrës së trajtimit/filtrimit të të dhënave fillestare kemi që $T \geq 10$. Vlerat e ndryshores së mbijetesës T janë shënuar me t . Ndryshorja e lidhur me ngjarjen është një ndryshore me dy vlera e shënuar me,

$$\delta = \begin{cases} 1 & \text{ngjarja ndodhi} \\ 0 & \text{ngjarja nuk ka ndodhur ende} \end{cases}$$

Në rastin tonë, të dhënat janë të pacensuruara duke qenë se ngjarja ka ndodhur në të gjitha thirrjet telefonike të marra në shqyrtim, pasi ato janë thirrje të kryera, të përfunduara.

Dy madhësi sasiore të cilat shqyrtohen zakonisht në analizën e mbijetesës janë: funksioni i mbijetesës $S(t)$ dhe funksioni i rrezikut i shënuar me $h(t)$ [97]. *Funksioni i mbijetesës jepet me anë të formulës:*

$$S(t) = P(T > t),$$

dhe tregon probabilitetin që një thirrje telefonike ti mbijetoj ndërprerjes përtej çastit të kohës t . *Funksioni i rrezikut, i cili jepet me anë të formulës:*

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t},$$

²⁷ Përshtatur nga anglishtja: Stratified.

është quajtur ndryshe edhe si *norma e rrezikut e kushtëzuar*²⁸. Ai përqëndrohet më shumë tek dështimi, që nënkupton përjetimin e ndërprerjes së thirrjes. Ndërkohë, funksioni i mbijetesës përqëndrohet më shumë tek mbijetesa, që nënkupton mosndërprerjen e thirrjes. *Rreziku i grumbulluar* jepet nga formula:

$$H(t) = \int_0^t h(u)du,$$

dhe përshkruan rrezikun e grumbulluar deri në çastin t të kohës. Ekziston një lidhje ndërmjet tre madhësive $S(t)$, $h(t)$ dhe $H(t)$ e cila jepet si vijon:

- $h(t) = -\frac{\partial \log(S(t))}{\partial t};$
- $H(t) = -\log(S(t));$
- $S(t) = \exp(-H(t)).$

3.2.2. Vlerësime Joparametrike

Vlerësimet joparametrike janë të dobishme për qëllime përshkruese, për të parë formën e funksionit të mbijetesës ose të rrezikut. Metoda Kaplan-Meier (KM) [90, 98] është përdorur për të vlerësuar kurbat e mbijetesës në dy grupe. Probabilitetet e vlerësuara të mbijetesës janë njehsuar nëpërmjet përdorimit të formulës së limitit të prodhimeve²⁹.

Në modelet joparametrike nuk ka asnjë supozim paraprak në lidhje me formën e funksionit të rrezikut apo se si ndryshoret shpjeguese mund të ndikojnë në të [99]. Funksioni i vlerësuar i mbijetesës përmes metodës KM është dhënë në vijim nëpërmjet formulës së limitit të prodhimeve:

$$\hat{S}(t_k) = \prod_{i=1}^k P(T > t_i | T \geq t_i).$$

Në rastin e të dhënave të pacensuruara përdoret formula e mëposhtme,

$$\hat{S}(t_k) = \frac{\text{numri i të mbijetuarve përtej çastit } t_k \text{ të kohës}}{n}.$$

Dy kurba të mbijetesës, nën hipotezën fillestare se vijnë nga i njëjti popullim, $H_0: S_1(t) = S_2(t)$, janë krahasuar nëpërmjet përdorimit të kriterit statistikor log-rang³⁰ [90]. Kriteri statistikor log-rang bazohet në shumën e diferencave ndërmjet rezultateve të vrojtuar me rezultatet e pritshme për një grup të dhënë dhe variancës së tij. Ky kriter ka afërsisht shpërndarje hi-katror me një shkallë lirie në zgjedhje me përmasa të mëdha dhe statistika e kriterit jepet nëpërmjet formulës së përafëruar si vijon:

$$X^2 = \sum_{i=1}^2 \frac{(O_i - E_i)^2}{E_i}.$$

Nëse p -vlara e këtij kriteri statistikor është më pak se niveli i rëndësisë i zgjedhur $\alpha = 0.05$, atëherë hipoteza fillestare do të refuzohet. Në këtë rast, dilet në përfundimin se ka një diferencë statistikisht domethënëse ndërmjet dy kurbave të mbijetesës. Thuhet

²⁸ Përshtatur nga anglishtja: Conditional failure rate.

²⁹ Përshtatur nga anglishtja: Product limit formula.

³⁰ Përshtatur nga anglishtja: Log-rank.

se ka evidence në nivelin e rëndësisë $\alpha = 0.05$ që dy kurbat e mbijetesës vijnë nga popullime të ndryshme.

Kjo teknikë është aplikuar jo vetëm për ndryshoret që marrin vetëm dy vlera të mundshme, por edhe për ndryshoret e tjera të cilat marrin më shumë se dy vlera. Në rastin e dytë, supozohet se k kurba mbijetese vijnë nga një popullim i njëjtë. Kriteri log-rang në këtë rast ka shpërndaje hi-katror me $k - 1$ shkallë lirie dhe statistika e tij është:

$$X^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}.$$

Nëpërmjet kësaj teknike konfirmohet rëndësia e përfshirjes së ndryshoreve në modelet e shqyrtuara të mbijetesës.

3.2.3. Modelet Gjysmëparametrike

Modeli Cox është modeli gjysmëparametrik më i njohur, i cili u prezantua fillimisht në një punim prej Cox në 1972 [97]. Modelet gjysmëparametrike [99] nuk bëjnë supozime në lidhje me formën e funksionit të rrezikut, por bëjnë supozime të forta në lidhje me faktin se si ndryshoret shpjeguese ndikojnë formën e funksionit të rrezikut. Në këto modele supozohet që rreziku është proporcional (PH³¹) në lidhje me kohën.

Në këtë pjesë, është shqyrtuar një problem i cili përfshin katër ndryshore shpjeguese të pavarura nga koha të cilat tentojnë parashikimin e kohës së mbijetesës T , “sekondat derisa thirrja telefonike ndërpritet”. Ndryshoret shpjeguese janë të emërtuara si vijon: X_1 (për përkatësinë e grupit), X_2 (për ditën e muajit), X_3 (për ditën e javës) dhe X_4 (për orën gjatë ditës). Në modelet e ndërtuara janë marrë në konsideratë kombinime të ndryshme të ndryshoreve të pavarura.

Mënyra më e zakonshme për përcaktimin e modelit PH Cox, pa ndërveprim ndërmjet ndryshoreve të pavarura në kohë të marra në konsideratë, është duke shqyrtuar funksionin e rrezikut. Në rastin e përgjithshëm, formula e këtij funksioni jepet si vijon,

$$h(t, \mathbf{X}) = h_0(t) \exp \left[\sum_{i=1}^s \beta_i X_i \right],$$

ku $\mathbf{X} = (X_1, X_2, \dots, X_s)$ është vektori i ndryshoreve shpjeguese dhe $0 \leq h(t, \mathbf{X}) < \infty$. Termi $h_0(t)$ është *funksioni bazë i rrezikut*³², ndërsa termi i dytë është shprehje eksponenciale (me bazë numrin e) e shumës lineare të $\beta_i X_i$, ku β_i janë parametra të panjohur. Kjo shumë varion mbi s ndryshoret shpjeguese. Supozimi PH përqëndrohet në faktin që funksioni $h_0(t)$ është funksion i varur nga koha t , por nuk përfshin ndryshoret shpjeguese $\mathbf{X}$. Modeli Cox PH njihet me emërtimin gjysmëparametrik duke qënë se funksioni bazë i rrezikut është një funksion i pa specifikuar. Ky funksion mund të marrë çfarëdolloj forme, ndërkohë që në model futen linearisht ndryshore shpjeguese.

Metoda e përgjasisë maksimale (ML³³) është përdorur për të vlerësuar parametrat β_i në model. Formula për funksionin e përgjasisë në modelin Cox është quajtur formulë me përgjasi “të pjesshme”, $L(\boldsymbol{\beta})$ (nënkuptomë me $\boldsymbol{\beta}$ koleksionin e parametrave të panjohur). Formula e përgjasisë shqyrton probabilitetet vetëm për ato raste të cilat

³¹ Shkurtim nga anglishtja i *Proportional Hazard*.

³² Përshtatur nga anglishtja: The baseline function.

³³ Shkurtim nga anglishtja i *Maximum Likelihood*.

dështojnë (domethënë e përjetojnë ngjarjen) dhe nuk merr në shqyrtim në mënyrë të veçantë ato raste të cilat janë të censuruara. Përfundimet janë bazuar në teknikën ML, të zbatuar në zgjedhje me përmasa të mëdha. Kontrolli për domethënien statistikore të vlerësimeve të bëra për secilin prej parametrave β_i bëhet nëpërmjet kriterit të Wald-it. Në këtë kriter statistika Z , përftohet duke pjestuar vlerësimin e bërë për parametrin me gabimin standart të tij. Më pas, supozohet se kjo madhësi është një ndryshore afërsisht normale, nën hipotezën fillestare që parametri korrespondues β është zero. Nëse $\alpha = 0.05$ është niveli rëndësisë i zgjedhur dhe $p \leq \alpha$ atëherë vlerësimi konsiderohet statistikisht domethënës për modelin që trajtohet.

Raporti i rrezikut (HR^{34}) është përcaktuar si rreziku për një brinjë i pjestuar nga rreziku për një brinjë të ndryshme nga e para. Ky raport është i pavarur nga koha. Dy brinjët të cilat krahasohen mund të dallohen prej vlerave të ndryshoreve shpjeguese të tyre. Për të vlerësuar raportin e rrezikut në krahasimin e dy brinjëve, le të themi e_1 dhe e_2 , me ndryshore shpjeguese respektivisht $\mathbf{X}_1 = (X_{11}, X_{12}, \dots, X_{1s})$ dhe $\mathbf{X}_2 = (X_{21}, X_{22}, \dots, X_{2s})$, formula është si vijon:

$$\widehat{HR} = \frac{h(t, \mathbf{X}_1)}{h(t, \mathbf{X}_2)} = \exp \left[\sum_{i=1}^s \widehat{\beta}_i (X_{1i} - X_{2i}) \right].$$

Mënyra më e mirë për të menduar në lidhje me HR [100] është të mendojmë sikurse me shanset në regresin e thjeshtë logjistik. Është e mundur që të konvertojmë HR në probabilitete,

$$HR = \frac{p_i}{1 - p_i} \quad \rightarrow \quad p_i = \frac{HR}{1 + HR}.$$

Një vlerësim me besueshmëri 95% në lidhje me intervalin e HR jepet nga $\exp[Koeficient \pm (1.96) Gabim Standart]$.

Kontrollet statistikore për diagnostikimin e PH janë bazuar në mbetjet e shkallëzuara të Schoenfeld dhe janë njehsuar për çdo ndryshore shpjeguese. Njëkohësisht, është kryer dhe një kontroll i përgjithshëm për modelin në tërësi. Supozimi PH është vlerësuar nëpërmjet një kriteri për kontrollin e mirësisë, të zbatuar në secilën ndryshore shpjeguese [101, 90] dhe një p -vlere. Një p -vlere jo domethënëse ($p > 0.10$) sugjeron që supozimi PH është i arsyeshëm. Ndërkohë, $p < 0.05$ sugjeron që ndryshorja e cila është duke u kontrolluar nuk e plotëson supozimin PH.

Kriteri statistikor i raportit të përgjasive (LR^{35}) përdor statistikën e përgjasive. Kriteri LR është diferenca ndërmjet modelit fillestar³⁶ dhe një modeli të specifikuar me ndryshore shpjeguese. Kriteri LR, Wald dhe kriteri hi-katror në lidhje me rezultatet (ose rezultate (log-rang) janë kriteret kontrolli asimptotikisht ekuivalente me njëri-tjetrin nën hipotezën fillestare që të gjithë parametrat β janë 0. Në rast se p -vlere është më pak se niveli i zgjedhur i rëndësisë $\alpha = 0.05$, atëherë modeli është statistikisht domethënës dhe hipoteza fillestare refuzohet. Kjo tregon një përmirësim të modelit fillestar.

Modeli i shtresëzuar Cox [90] modifikon modelin Cox PH duke lejuar kontroll nëpërmjet “shtresëzimit” mbi ndryshoret shpjeguese të cilat janë të përfshira në model, por që nuk plotësojnë supozimin PH. Ndryshoret shpjeguese për të cilat supozohet se e kënaqin supozimin PH, janë përfshirë direkt në model. Ndërkohë, ndryshoret shpjeguese që shtresëzohen nuk janë përfshirë në mënyrë direkte në model. Çdo shtresë

³⁴ Shkurtim nga anglishtja i *Hazard Ratio*.

³⁵ Shkurtim nga anglishtja i *Likelihood Ratio*.

³⁶ Përshtatur nga anglishtja: Null model.

është e lejuar të ketë një funksion bazë rreziku të ndryshëm, ndërkohë që parametrat korresponduese të ndryshoreve shpjeguese të mbetura në model supozohet se qëndrojnë konstant përmes shtresave.

Në përgjithësi shënojmë me $X_1, X_2, \dots, X_p$ për $p \in \mathbb{N}$, ndryshoret të cilat kënaqin supozimin PH dhe me $Z_1, Z_2, \dots, Z_k$ për $k \in \mathbb{N}$, ndryshoret të cilat nuk e kënaqin supozimin PH. Le të jetë Z^* një ndryshore e re ku Z_i tregon çdo vlerë të kategorizuar, e cila është një kombinim i shtresave. Ndryshorja Z^* ka m kategori, ku m është numri i kombinimeve (shtresave). Modeli i përgjithësuar Cox me shtresëzime, me ndryshore shpjeguese të pavarura në kohë të cilat nuk kanë ndërveprim me njëra-tjetrën, jepet si vijon:

$$h_i(t, \mathbf{X}) = h_{0i}(t) \exp[\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p],$$

ku $i = 1, \dots, m$ shtresa të përcaktuara nga Z^* . Për çdo shtresë janë përcaktuar parametra të njëjtë $\beta_1, \beta_2, \dots, \beta_p$, por funksionet bazë të rrezikut $h_{0i}(t)$ mund të jenë të ndryshme në secilën prej tyre. Ndryshoret $X_1, X_2, \dots, X_p$ janë përfshirë direkt në model, ndërsa Z^* shfaqet vetëm nëpërmjet funksioneve të ndryshme $h_{0i}(t)$. Parametrat β_i janë vlerësuar nëpërmjet metodës së përgjasisë maksimale.

Modeli më i mirë për të përqasur funksionin e rrezikut është përzgjedhur nëpërmjet minimizimit të *kriterit të përgjithësuar të informimit*³⁷ (AIC³⁸) [102, 103]. Analiza e mbijetesës në këtë pjesë është kryer nëpërmjet përdorimit të platformës R [67] dhe disa paketave të saj në të, siç janë: igraph [64], igraphdata [63], survival [104, 105], KMsurv [106], coin [107, 108], OIsurv [109], dhe caret [110].

3.3. REZULTATET E HULUMTIMIT DHE DISKUTIME

Të dhënat janë ndarë në dy grupe. Në Grupin 1 janë përfshirë të dhënat të cilat lidhen me të gjitha ato thirrje telefonike që marrin fillim në intervalin e kohës 00:00 am – 11:59 am, ndërkohë në Grupin 2 janë përfshirë të dhënat në lidhje me ato thirrje telefonike të cilat marrin fillim në intervalin e kohës 12:00 pm – 23:59 pm. Disa statistika përshkruese në lidhje me të dhënat e thirrjeve telefonike janë dhënë në Tabelën 3.1.

Tabelë 3.1: Statistika përshkruese në lidhje me të dhënat për kohëzgjatjen e thirrjeve telefonike.

	n	Ngjarjet	Mesatarja	0.95 LCL ³⁹	0.95 UCL ⁴⁰
Të gjitha	74108	74108	61	60	62
Grupi 1	25752	25752	55	55	56
Grupi 2	48356	48356	64	63	65

Kurbat e mbijetesës të vlerësuara nëpërmjet metodës KM janë dhënë në Figurën 3.2. Vihet re se, vlerësimi i bërë për funksionin e mbijetesës në Grupin 2 është mbi atë të Grupit 1. Kjo tregon se Grupi 2 ka një prognozë mbijetese më të mirë se sa Grupi 1. Pas kryerjes së kriteri log-rang për të vlerësuar nëse kurbat KM të dy grupeve janë statistikisht ekuivalente, u gjet se $p = 0 < 0.05$. Për rrjedhoj, hipoteza fillestare refuzohet dhe dilet në përfundimin se kurbat e mbijetesës vijnë nga dy popullime të ndryshme. Kriteri log-rang është aplikuar edhe në ndryshoret e tjera dhe në të gjitha

³⁷ Përshtatur nga anglishtja: The generalized Akaike Information Criterion.

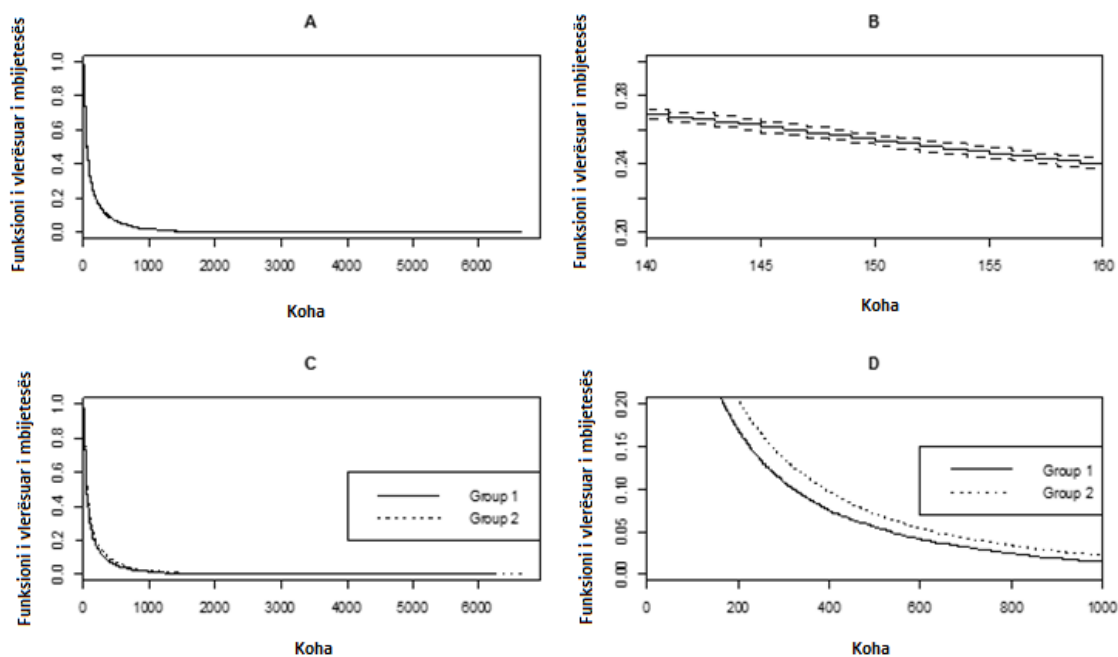
³⁸ Shkurtim nga anglishtja i *Akaike Information Criterion*.

³⁹ Shkurtim nga anglishtja i *Lower Control Limit* nënkuptojmë në shqip *kufirin e poshtëm të kontrollit*.

⁴⁰ Shkurtim nga anglishtja i *Upper Control Limit*, nënkuptojmë në shqip *kufirin e sipërm të kontrollit*.

rastet u gjet një p -vlerë më e vogël se 0.05. Kështu, të katërta ndryshoret shpjeguese, të pavarura nga koha janë marr në shqyrtim për tu përfshirë në modelet e mbijetesës.

Në Tabelën 3.2. janë pasqyruar rezultatet në lidhje me modelet e mbijetesës, pa ndërverpim të ndryshoreve shpjeguese, të cilat i'u përqasën të dhënave. Me Model 1 është shënuar modeli fillestar, i cili nuk ka ndryshore shpjeguese të përfshira në të. Analiza njëpërmasore e mbijetesës është trajtuar në modelet PH Cox 2, 3, 4, dhe 5. Prej saj, rezultoi se vlerësimet $\hat{\beta}$ ishin pozitive për ndryshoret X_2 dhe X_3 në Modelin 3 dhe 4. Kjo nënkupton se, vlera të larta të X_2 dhe X_3 janë të lidhura me mbijetesë të dobët. Ndërkohë, në Modelin 2 dhe 5 vlerësimet $\hat{\beta}$ rezultuan negative, respektivisht për X_1 dhe X_4 . Kjo nënkupton që, të qenit pjesë e Grupit 2 dhe vlera të larta të X_4 lidhen me mbijetesë më të mirë. Në secilën prej modeleve PH Cox Models 2, 3, 4, dhe 5, nisur nga p -vlera ($P(> |z|)$) për kontrollin e domethënies statistikore për $\hat{\beta}$, ka rezultuar që vetëm për ndryshoret shpjeguese X_1, X_2 dhe X_4 vlerësimet për parametrat korrespondues ishin statistikisht domethënëse në nivelin e rëndësisë $\alpha = 0.05$. Për këtë arsye, në Modelin 6, ndryshorja X_3 është larguar.



Figurë 3.2: Kurbat Kaplan-Meier: A – tregon vlerësimin për kurbën e mbijetesës me interval me besueshmëri 95% duke përdorur log-transformimet; B – tregon një zmadhim të A në intervalin e kohës (140, 160); C – tregon vlerësimin për kurbat e mbijetesës për dy grupet, dhe D – është zmadhim i C në intervalin kohor (0, 1000).

Në Modelin 2 PH Cox për ndryshoren X_1 është vlerësuar $\hat{\beta} = -0.113 < 0$ dhe $\widehat{HR} = 0.893 < 1$. Kjo nënkupton që, të qenit në Grupin 2 redukton rrezikun me një faktor 0.893 ose 47%, duke nënkuptuar kështu një prognozë të mirë për mbijetesën. E njëjta gjë ndodh edhe në Modelin 5, ku për ndryshoren X_4 u vlerësuan $\hat{\beta} = -0.027 < 0$ dhe $\widehat{HR} = 0.974 < 1$. Një rritje më 1-njësi e vlerës së ndryshores X_4 , do të zvogëlonte probabilitetin për të përjetuar ngjarjen, duke e reduktuar rrezikun me 49.3%. Një subjekt me vlera të larta të ndryshores X_4 , ka normë rreziku më të ulët krahasuar me një me vlera më të ulta.

Në modelin 3 dhe 4 është vlerësuar se $\hat{\beta} > 0$ dhe $\widehat{HR} > 1$, respektivisht për ndryshoret X_2 dhe X_3 . Kjo sjell që, rritja me 1-njësi për X_2 në Modelin 3, e rrit probabilitetin e përjetimit të ngjarjes me 50%, duke nënkuptuar mbijetesë të dobët. Në rastin e Modelit 4, duke qenë se vlerësimi për parametrin korrespondues të X_3 nuk është statistikisht domethënës dhe vlerësimi me besueshmëri 95% për intervalin e $\widehat{HR}$ është (0.998, 1.005) i cili e përmban vlerën 1, interpretimi për HR në këtë rast nuk do të ishte i saktë.

Ndryshoret shpjeguese X_1, X_3 dhe X_4 , kanë rezultuar se nuk e kënaqin supozimin PH. Ky fakt është statistikisht domethënës bazuar në p (PH)-vlerat e tyre të cilat janë më pak se 0.05. Vetëm në rastin e ndryshores X_2 , supozimi i bërë për PH ishte i arsyeshëm. Në këtë mënyrë, së pari është ndërtuar Modeli 6, i cili është një model PH Cox shumëpërmasor, pa ndërveprim ndërmjet ndryshoreve shpjeguese të përfshira në të dhe që përfshin të gjitha ndryshoret shpjeguese të cilat u vlerësuan si domethënëse në modelet njëpërmasore të konsideruara më herët. Në këtë rast, për të gjitha ndryshoret e përfshira, vlerësimet për parametrat korrespondues të tyre janë statistikisht domethënës, por vetëm në rastin e ndryshores X_2 , supozimi PH rezultoi të jetë i arsyeshëm. Pas kësaj, është ndërtuar Modeli 7 i cili është një model i shtresëzuar Cox. Në këtë model, dy ndryshoret të cilat nuk kënaqin supozimin mbi PH janë shtresëzuar.

Në Modelin 6, i cili është një model shumëpërmasor PH Cox, në rastin e ndryshoreve X_1 dhe X_2 , u pa që $\hat{\beta} > 0$ dhe $\widehat{HR} > 1$, ndërkohë në rastin e ndryshores X_4 , situata ishte e kundërt, $\hat{\beta} < 0$ dhe $\widehat{HR} < 1$. Kjo nënkupton që, ndërkohë që ndryshoret e tjera qëndrojnë konstante, vlera të larta të ndryshores X_1 (ose në rastin tjetër X_2) janë të shoqëruara me rrezik të lartë dhe mbijetesë të dobët. Vlera të larta të ndryshores X_4 janë të lidhura me prognozë të mirë të mbijetesës. Nëse secilës prej ndryshoreve i'u rritet vlera me 1-njësi, ndryshimi i rrezikut është si vijon: për X_1 , rreziku rritet me 56%; për X_2 rritet me 50%; për X_4 reduktohet me 49%. Në modelin 7 rezultoi të ishte e njëjta gjendje për X_2 , sikurse në Modelin 6.

Me përjashtim të Modelit 4, në të gjitha rastet e kritereve të përgjithshme të kontrollit (LR, Wald, Rezultate (log rang) rezultoi që p -vlera ishte më e vogël se niveli rëndësisë $\alpha = 0.05$. Ky rezultat shërben si tregues i faktit që këto modele janë statistikisht domethënëse pasi përmirësojnë modelin fillestar. Modeli më i mirë në përjashtjen e funksionit të rrezikut, prej modeleve të trajtuara, u vlerësua nëpërmjet vlerave të AIC. Në modelin 7 u gjet se vlera e AIC ishte minimalja prej atyre të shqyrtuara. Në këtë mënyrë ky model u vlerësua si më i miri dhe funksioni i rrezikut në të jepet si vijon:

$$\hat{h}_i(t, X_1, X_2, X_4) = \hat{h}_{0i}(t) \exp[0.001 X_2],$$

me $i = 1, \dots, 48$ shtresa të përcaktuara nga kombinimet midis vlerave të X_1 dhe X_4 . Nëse marrim një brinjë të ditës së dhjetë të muajit (nënkuptojmë $X_2 = 10$), funksioni i vlerësuar i rrezikut në këtë rast është:

$$\hat{h}_i(t, X_1, X_2, X_4) = \hat{h}_{0i}(t) \exp[0.001 \cdot 10] = \hat{h}_{0i}(t) \exp[0.01] = \hat{h}_{0i}(t) \cdot 1.01005.$$

Kjo është ekuivalente me një rritje të funksionit të rrezikut me 50%.

Tabelë 3.2: Modelet e mbijetesës me ndryshore shpjeguese të pavaruara nga koha dhe pa ndërveprim midis tyre. Shkurtimi df: shkallë lirie⁴¹.

Ndryshore	⁴² $\hat{\beta}$	⁴³ Gabimi Standart	⁴⁴ $P(> z)$	⁴⁵ Raporti i rrezikut	⁴⁶ 95% CI	⁴⁷ p (PH)	
Model 1							
Modeli Fillestar							
AIC= 1514532	Log Përgjasia= -757266.1						
Model 2~ X_1							
X_1 Grupi	-0.113	0.008	<2e-16	0.893	0.880	0.907	0.006
AIC= 1514322	LR = 212.4 me 1 df, p=0		Wald = 215 me 1 df, p=0		Rezultate (log-rang) = 215.2 me 1 df, p=0		
Model 3~ X_2							
X_2 Dita e muajit	0.001	0.0004	0.004	1.001	1	1.002	0.935
AIC= 1514526	LR = 8.17 me 1 df, p=0.0043		Wald = 8.17 me 1 df, p=0.0043		Rezultate (log-rang) = 8.17 me 1 df, p=0.0043		
Model 4~ X_3							
X_3 Dita e javës	0.001	0.002	0.456	1.001	0.998	1.005	0.0001
AIC= 1514534	LR = 0.56 me 1 df, p=0.456		Wald = 0.55 me 1 df, p=0.456		Rezultate (log-rang) = 0.55 me 1 df, p=0.456		
Model 5~ X_4							
X_4 Ora e ditës	-0.027	0.001	<2e-16	0.974	0.972	0.975	1.17e-13
AIC= 1513624	LR = 910.1 me 1 df, p=0		Wald = 910.5 me 1 df, p=0		Rezultate (log-rang) = 912.8 me 1 df, p=0		
Model 6~ $X_1 + X_2 + X_4$							
X_1	0.261	1.364e-02	<2e-16	1.298	1.264	1.334	3.06e-07
X_2	0.001	4.267e-04	0.0024	1.001	1.000	1.002	8.40e-01
X_4	-0.052	1.589e-03	<2e-16	0.950	0.947	0.952	0.00e+00
Tërësor							0.00e+00
AIC= 1513257	LR = 1281 me 3 df, p=0		Wald =1262 me 3 df, p=0		Rezultate (log-rang) = 1265 me 3 df, p=0		
Model 7~strata(X_1) + X_2 + strata(X_4)							
X_2	0.001	4.276e-04	0.00314	1.001	1	1.002	0.894
AIC= 1110069	LR = 8.72 me 1 df, p=0.00314		Wald =8.72 me 1 df, p=0.00314		Rezultate (log-rang) = 8.72 me 1 df, p=0.00314		

Në Figurën 3.3. jepen vlerësimet për kurbat bazë të mbijetesës të Modelit 7, të cilat i korrespondojnë vlerësimeve për funksionet bazë të rreziqeve. Një kornizë të dhënash

⁴¹ Degree of freedom.

⁴² Vlerësimet për parametrat respektiv të secilës ndryshore të përfshirë në modele.

⁴³ Gabimi standart i parametrave të vlerësuar.

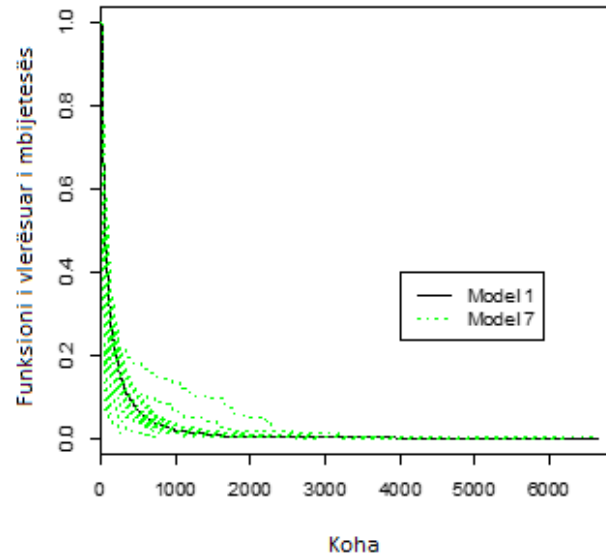
⁴⁴ p-vlera për kontrollin e domethënies statistikore të secilit parametër.

⁴⁵ Vlerësimi për HR në secilën ndryshore.

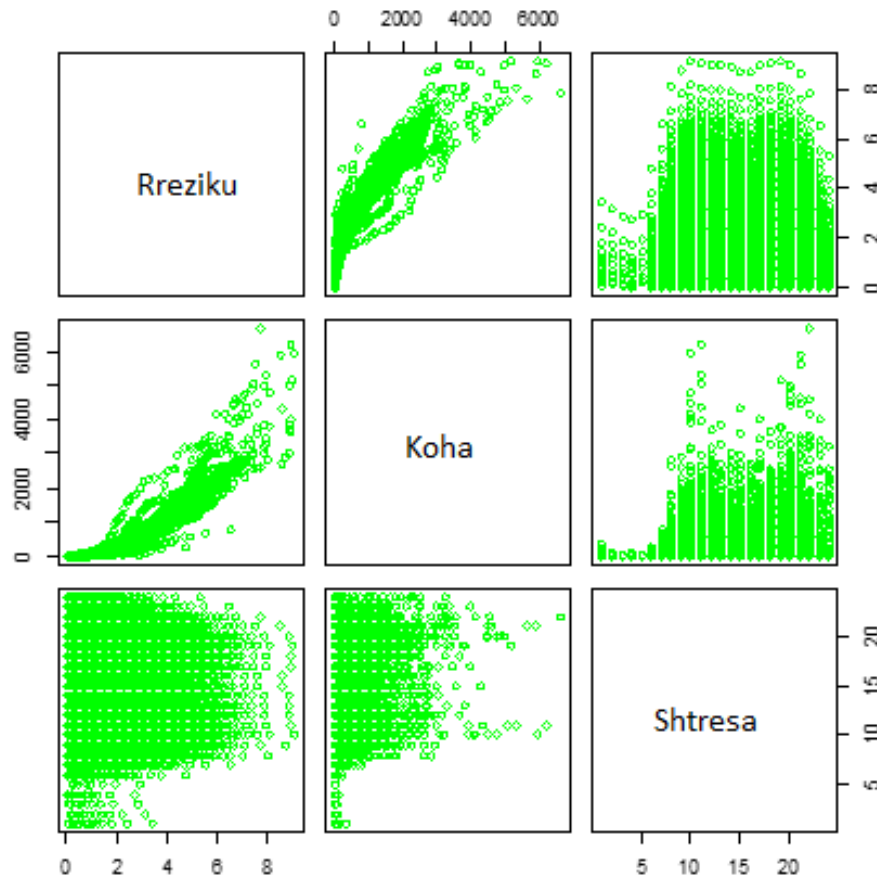
⁴⁶ Intervali me besueshmëri 95% për $HR = e^{Koeficient}$, i dhënë nga hapësira e vlerave në $\exp[Koeficient. \pm(1.96) Gabim Standart]$

⁴⁷ p-vlera për kontrollin e supozimit mbi PH në secilën ndryshore.

e cila përmban kohën, funksionin bazë të rrezikut të grumbulluar dhe shtresat në lidhje me Modelin 7 jepet në Figurën 3.4.



Figurë 3.3: Vlerësimet për kurbat bazë të mbijetesës në Modelin 7, të krahasuara me kurbën e mbijetesës në Modelin 1.



Figurë 3.4: Funksionet bazë të rrezikut të grumbulluar në Modelin 7, të dhëna në rastin ku $\beta = 0$.

3.4. PËRFUNDIME

Si përfundim, pas analizës së rezultateve të studimit u gjet se mënyra më e mirë për të shpjeguar normën e rrezikut të kushtëzuar të një lidhje komunikuese ishte nëpërmjet modelit të shtresëzuar Cox. Ndryshorja në lidhje me “përkatësinë në grup” dhe ndryshorja në lidhje me “orët e ditës” ishin të lidhura me një prognozë të mirë për kohën gjatë së cilës një thirrje telefonike i mbijetonte ndërprerjes së saj. Duke qënë se këto ndryshore nuk e plotësonin supozimin në lidhje me rrezikun proporcional, u vendosën në shtresa. Vlera të larta të ndryshores “dita e muajit” kishin lidhje me normë të lartë të rrezikut dhe mbijetesë të dobët të lidhjeve komunikuese. Ndryshorja “dita e javës” rezultoi pa ndonjë domethënie statistikore në lidhje me prognozën e mbijetesës apo normën e kushtëzuar të dështimeve të lidhjeve komunikuese. Do të qe interesante për të parë më tej në lidhje me përfshirjen e ndërveprimit ndërmjet ndryshoreve shpjeguese në modelet e mbijetesës dhe gjithashtu aplikimi i analizës parametrike në të dhënat e komunikimit telefonik. Në dijeninë tonë, mungojnë aplikimet e analizës së mbijetesës në graf rrjeta të telefonisë fikse në këtë mënyrë, ndaj është e pamundur të bëjmë krahasimet me studime të tjera.

4. SHPËRNDARJET E FUQIVE NË GRAF RRJETA

4.1. HYRJE

Modeli më i përhapur për të modeluar shpërndarjen e fuqive në graf rrjeta ka qënë ai i shpërndarjes power law (PL). Graf rrjetat në të cilat shpërndarja e fuqive të kulmeve është shpërndarje PL janë emërtuar prej Barabasi dhe Albert 1999 me emrin “scale free” [1, 111]. Funksioni i densitetit të shpërndarjes PL gëzon vetinë $f(ax) = bf(x)$. Një vend të rëndësishëm në modelimin e shpërndarjes së fuqive ka zënë gjithashtu edhe shpërndarja lognormale (LN).

Mitzenmacher në artikullin e tij [112] ka paraqitur një historik të shkurtër në lidhje me modelet gjeneruese që konvergjojnë në shpërndarjen power law (PL) dhe lognormale (LN). Sipas tij, këto dy shpërndarje kanë një lidhje natyrore shumë të mirë me njëra-tjetrën dhe se nuk është çudi që shpërndarja LN të ketë lindur si një alternativë ndaj shpërndarjes PL. Ai zbuloi së pari se, shpërndarja PL po gjente një zbatim shumë të madh në studimet shkencore në atë kohë, për të përshkruar dinamikën e rritjes së grafeve Web, ndonëse modele shumë të ngjashme më këtë shpërndarje datonin të paktën që në vitet 1950 apo edhe në vitet 1920. Së dyti, pyetja nëse disa shpërndarje empirike (të hasura në ekonomi, biologji, kimi, astronomi dhe teori informacioni), modeloheshin më mirë me shpërndarjen PL apo atë LN datonte që në vitet 1950. Së treti, këto dy shpërndarje ishin pashmangshmërisht të lidhura me njëra-tjetrën, nisur nga fakti se modele gjeneruese të ngjashme, si rrjedhojë e variacionit të parametrave, mund të shpinin në njërin ose tjetrën prej tyre.

Duke qënë se fuqitë e kulmeve janë madhësi sasiore diskrete, shpërndarjet teorike pritet të jenë diskrete. Shpesh ndodh që një shpërndarje të mos përqs të gjitha të dhënat, por vetëm një pjesë të tyre duke filluar nga një vlerë x_{min} . Në rastet kur x_{min} është vlerë e madhe, diferenca ndërmjet një shpërndarje diskrete dhe një shpërndarje të vazhdueshme nga ana e formës funksionale bëhet e papërfillshme. Kështu që, në këto raste një shpërndarje diskrete mund të përafrohet me një shpërndarje të vazhdueshme.

Në [41, 113], me të dhënat mujore të komunikimit nëpërmjet telefonisë fikse u ndërtua një graf rrjet statik. Përshkrimi i disa karakteristikave strukturore, ndër to edhe i vargut të fuqive u krye në [113], ndërsa në [41] studimi u përqëndrua në komponentin e lidhur më të madh të tij. Në këtë punim u analizuan pjesët fundore të disa shpërndarjeve empirike të cilat kishin të bënin me: numrin e thirrjeve telefonike për klient, kohëzgjatjen totale të komunikimeve (në sekonda) për klient dhe për numrin e komunikuesve (të ndryshëm nga njëri-tjetri) për klient. Graf rrjeti u shqyrtua në dy raste, i orientuar dhe i paorientuar. Në pjesët fundore të shpërndarjeve u bë një krahasim ndërmjet përqsjeve të bëra nëpërmjet shpërndarjes PL dhe asaj LN. Nga studimi nuk u arrit në një përfundim të përcaktuar saktë në lidhje me dominancën e njërës mbi tjetrën.

Analiza e pjesëve fundore të shpërndarjeve në lidhje me fuqitë e kulmeve, në graf rrjetat e komunikimit, është e rëndësishme sepse ajo jep informacion në lidhje me *kulmet qendrore*⁴⁸ apo fenomenet e rralla. Kulmet qendrore kanë lidhshmëri të lartë dhe supozohet se luajnë rolin e kulmeve kryesor në lidhje me konvergjencën apo divergjencën e informacionit në graf rrjeta. Marrja në shqyrtim e graf rrjetit si statik, të ndërtuar mbi të dhënat mujore dhe përqëndrimi vetëm në komponentin e lidhur më të madh të tij mund të ketë ndikuar në rezultatet e gjetura në punimin [41]. Gjithashtu,

⁴⁸ Përshtatur nga anglishtja: hubs.

shpërndarja e fuqive është parë edhe në një studim krahasues ndërmjet katër graf rrjetave, të cilat i takonin të njëjtës ditë jave të muajit [114].

Në këtë pjesë të disertacionit qëllimi kryesor është gjetja e disa rezultateve të reja në lidhje me shpërndarjet ditore të fuqive të kulmeve, duke marrë në konsideratë evoluimin e graf rrjetit në kohë. Duke qënë se lidhjet komunikuese telefonike kanë një kohë mbijetese, me evoluim të graf rrjetit në kohë nënkuptojmë ndryshimin e vazhdueshëm që bën struktura topologjike e tij. Fotografimet ditore janë përdorur për të treguar gjendjen e strukturës topologjike të graf rrjetit në një çast të kohës. Në këtë pjesë është trajtuar problematika e përshkruar nëpërmjet pyetjeve si vijon:

- **P4.1:** Cila është mënyra më e mirë për të shprehur numrin e komunikuesve ditor (të ndryshëm nga njëri-tjetri) për klient?

Kjo problematikë është trajtuar më parë prej nesh edhe në [42].

- **P4.2:** Cila është mënyra më e mirë për të shprehur numrin e thirrjeve telefonike ditore (totale, hyrëse dhe dalëse) për klient?
- **P4.3:** A mund të pranojmë se, në thirrjet ditore të një klienti, ekziston ndonjë korrelacion ndërmjet numrit të thirrjeve hyrëse me ato dalëse?

Le të jenë G_i , për $i = 1, \dots, 30$ graf rrjeta të cilat janë ndërtuar në bazë vetëm të thirrjeve të ditës së i -të të muajit. Për t'ju përgjigjur pyetjes **P4.1** është ndërtuar modeli i serisë kohore të graf rrjetave si vijon:

- **Modeli 4.1**⁴⁹: Secili graf rrjet G_i në serinë $G_1, G_2, \dots, G_{30}$ është i thjeshtë, i paorientuar dhe i papeshuar. Një brinjë paraqet një lidhje komunikuese ndërmjet dy klientëve të telefonisë. Kështu, nëse v_1 dhe v_2 janë kulme, atëherë do të ekzistoj një brinjë e paorientuar (v_1, v_2) ndërmjet tyre vetëm në qoftë se v_1 ka bërë ose marr të paktën një thirrje prej v_2 ose anasjelltas. Lidhjet e shumëfishta ndërmjet çdo dy kulmeve janë thjeshtuar në një të vetme.

Për t'ju përgjigjur pyetjes **P4.2** dhe **P4.3** është ndërtuar modeli i serisë kohore të graf rrjetave si vijon:

- **Modeli 4.2:** Secili graf rrjet G_i në serinë $G_1, G_2, \dots, G_{30}$ është parë në dy këndvështrime: i paorientuar, i papeshuar (Modeli 4.2-Totale) dhe i orientuar, i papeshuar (Modeli 4.2-Hyrëse, Modeli 4.2-Dalëse). Në rastin e modelit të orientuar, nëse v_1 dhe v_2 janë kulme, atëherë ekziston një brinjë e orientuar (v_1, v_2) ndërmjet tyre vetëm në qoftë se v_1 ka bërë ose marr të paktën një thirrje prej v_2 ose anasjelltas. Shuma e numrit të thirrjeve hyrëse me ato dalëse përbën numrin total të thirrjeve për klient.

Shpërndarjet PL dhe LN janë krahasuar në vargun e fuqive për çdo G_i , në secilin prej modeleve të mësipërme. Analiza synon të përcaktoj shpërndarjen e cila përfaqëson më mirë të dhënat në lidhje me vargun e fuqive, në secilin hap të kohës. Gjithashtu, janë marrë në analizë edhe parametrat e vlerësuar për shpërndarjen që është gjykuar si shpërndarja më e mirë për përfaqësimin ndaj të dhënave.

Njehsimet statistikore në lidhje me këto shpërndarje janë bërë duke u mbështetur në paketat: *powerLaw* [115], *fBasics* [116], *igraphdata* [63], *igraph* [64]; në platformën

⁴⁹ Në vijim ky model, emërtohet edhe si Modeli 4.1-Totale.

njehësuese statistikore R [67]. Regjistrimet e të dhënave të cilat janë përdorur për këtë pjesë të studimit gjenden në [37].

4.2. METODOLOGJIA

4.2.1. Kuptime Bazë mbi Shpërndarjet dhe Vetitë e Tyre

Është shënuar me $\{d_v^i\}_{v \in V_i}$ vargu i fuqive të kulmeve në G_i . Përpjesa e kulmeve v në graf rrjetin G_i për të cilat $d_v^i = x$ është shënuar me $\hat{f}_x^i$. Kjo interpretohet gjithashtu edhe si – probabiliteti empirik që një kulm, i zgjedhur rastësisht në mënyrë të njëtrajtshme, ka fuqi të barabartë me x në ditën e i -të. Bashkësia e $\{\hat{f}_x^i\}_{x \geq 0}$ përcakton shpërndarjen empirike ditore të fuqive të kulmeve në graf rrjetin G_i .

Në rastin e Modelit 4.1-Totale merret në analizë vargu i fuqive $\{d_v^i\}_{v \in V_i}$, për çdo graf rrjet G_i të Modelit 4.1. Fuqia e kulmit d_v^i tregon numrin e brinjëve incidente me kulmin v , ku praktikisht nënkuptohet numri i komunikuesve të ndryshëm ditor, me të cilët klienti “ v ” ka qënë i angazhuar në të paktën një thirrje telefonike gjatë ditës së i -të.

Në Modelin 4.2-Totale studiohet $\{d_v^i\}_{v \in V_i}$, për çdo graf rrjet G_i të Modelit 4.2. Në këtë rast d_v^i është tregues i numrit të përgjithshëm të telefonatave në të cilat klienti “ v ” ka qënë i angazhuar (si telefonues apo i telefonuar) gjatë ditës së i -të. Në Modelin 4.2-Hyrëse studiohet $\{d_v^{i-}\}_{v \in V_i}$, për çdo graf rrjet G_i të Modelit 4.2. Fuqia hyrëse d_v^{i-} tregon numrin e brinjëve të orientuara me mbarim në kulmin v , ku nënkuptohet numri i thirrjeve telefonike ditore hyrëse në drejtim të klientit “ v ”. Në Modelin 4.2-Dalëse studiohet $\{d_v^{i+}\}_{v \in V_i}$, për çdo graf rrjet G_i të Modelit 4.2. Fuqia dalëse d_v^{i+} , tregon numrin e brinjëve të orientuara me fillim në kulmin v dhe nënkuptojmë në këtë rast numrin e thirrjeve dalëse prej klientit “ v ” në ditën e i -të. Në një graf rrjet të orientuar, shuma e fuqisë hyrëse me fuqinë dalëse të një kulmi përbën fuqinë e kulmit.

Për të kontrolluar normalitetin e të dhënave në vargun e fuqive (fuqive hyrëse, fuqive dalëse) të secilit graf rrjet, në modelet e marra në shqyrtim, është ndërtuar histogrami me boshte të logaritmuara, grafiku normal kuantil-kuantil e më tej është kryer kriteri Shapiro-Wilk. Evidencat mbi mungesën e normalitetit të të dhënave janë bazuar gjithashtu edhe në p -vlerën e kriterit Shapiro-Wilk [76, 77, 78], ku niveli i zgjedhur i rëndësisë është $\alpha = 0.01$.

Treguesit e formës, sikurse janë koeficienti i asimetrisë [117] dhe ai i sheshtësisë, janë përdorur për të përcaktuar nëse shpërndarja empirike është me pjesë fundore të anuara, të rrëzuara në të djathtë. Le të jetë $\{x_1, x_2, \dots, x_n\}$ një bashkësi të dhënash njëpërmasore.

Koeficienti i asimetrisë⁵⁰, γ_1 , vlerësohet nëpërmjet formulës:

$$\gamma_1 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{3/2}},$$

⁵⁰ Ekzistojnë disa formula për përllogaritjen e koeficientit të asimetrisë. Gjatë këtij studimi është përdorur kjo formulë.

ku $\bar{x}$ i referohet mesatares së të dhënave. Në varësi të vlerave të γ_1 dallojmë rastet si vijon:

- $\gamma_1 = 0$, shpërndarje *simetrike* (psh. shpërndarja normale);
- $\gamma_1 < 0$, shpërndarje *asimetrike negative*, pjesa fundore e shpërndarjes në të majtë është më e gjatë krahasuar me atë në të djathtë;
- $\gamma_1 > 0$, shpërndarje *asimetrike pozitive*, pjesa fundore e shpërndarjes në të djathtë është më e gjatë krahasuar me atë në të majtë.

Koeficienti i sheshtësisë, γ_2 , vlerësohet nëpërmjet formulës:

$$\gamma_2 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^2} - 3.$$

Në varësi të vlerave të γ_2 dallojmë rastet si vijon:

- $\gamma_2 = 0$, shpërndarje *mesokurtic* (psh. shpërndarja normale standarte);
- $\gamma_2 < 0$, shpërndarje *platykurtic*, me pjesë fundore lehtësisht të anuara të rrëzuara⁵¹;
- $\gamma_2 > 0$, shpërndarje *leptokurtic*, me pjesë fundore të anuara, të rrëzuara.

Rritja e vlerave të koeficientit të sheshtësisë lidhet me “zhvendosjen e masës probabilitare nga supet e shpërndarjes në drejtim të qendrës dhe pjesës kulmore të saj [118]. Shpërndarjet leptokurtic janë pjesë e shpërndarjeve me pjesë fundore të rrëzuara [119]. Shpërndarjet në të cilat funksionet probabilitare “zbresin” më ngadalë se një funksion eksponencial janë quajtur shpërndarje me pjesë fundore të rrëzura [120].

Shpërndarjet PL dhe LN janë me pjesë fundore të anuara, të rrëzuara në të djathtë. Janë zgjedhur për tu përqasur këto dy shpërndarje për të dhënat $x \geq x_{min}$, sepse jo gjithmonë mund të kemi një përqasje shumë të mirë për të gjitha të dhënat. Proçedura e vlerësimit të parametrave të tyre është bazuar në *metodën e përgjasisë maksimale* (ML) [121, 122]. Kjo teknikë është aplikuar edhe në [123, 124].

Problemet që shfaqeshin gjatë përqasjes së shpërndarjeve empirike nëpërmjet shpërndarjes PL, janë trajtuar prej Goldstein *et. al.* [125] dhe Newman [126]. Ata treguan se metoda më e mirë për të vlerësuar parametrin e shpërndarjes ishte metoda ML. Gjatë krahasimit u morën në shqyrtim disa metoda si: modeli linear në madhësitë e log-transformuara, modeli linear në pesë pikat e para në madhësitë e log-transformuara, modeli linear në madhësitë e log-transformuara në histogramë dhe metoda ML.

Nëse bashkësia e të dhënave ndaj të cilave bëhet përqasja është më e vogël se 50, duhet treguar tepër kujdes sepse në këto raste është e vështirë të rrëzohet poshtë hipoteza se të dhënat kanë shpërndarje hipotetike PL (ose LN) [121, 122]. Metoda ML krijon mundësinë e përdorimit të kriterit Kolmogorov-Smirnov (KS) [127], për të testuar mirësinë e përqasjes së shpërndarjes teorike (hipotetike) ndaj asaj empirike.

Në shumë aplikime, sjellja e të dhënave diskrete sipas një shpërndarje PL është përafruar me shpërndarjen e vazhdueshme PL duke qënë se kjo e fundit shfaq lehtësira më të mëdha nga pikëpamja matematikore [122]. Metodën që janë përdorur për të bërë këtë përafrim variojnë. Në përafrimin i cili ka rezultuar i saktë në shumë aplikime, numrat e plotë janë konsideruar si vlera të rrumbullakosura të një ndryshoreje të

⁵¹ Përshtatur nga anglishtja: *Light-tailed*.

vazhduar. Përafrimi i cili konsideron të dhënat diskrete si rezultat i marrjes së pjesës së plotë nga të dhënat e vazhduara, ka rezultuar të mos jetë i saktë.

4.2.1.1. Shpërndarja Power Law (PL)

Thuhet se një ndryshore rasti jonegative X , ka shpërndarje PL atëherë kur funksioni i masës probabilitare (ose densitetit probabilitar) [121, 122, 126] jepet me formulën:

$$f(x) = cx^{-\alpha},$$

për ndonjë konstante $c > 0$ dhe $\alpha > 0$. Konstantja α quhet eksponent i shpërndarjes PL. Zakonisht vlerat e saj luhaten në intervalin (2, 3), por ka edhe përjashtime.

Në situata reale, sjellja PL e të dhënave zakonisht shfaqet në pjesën fundore të shpërndarjes. Për këtë arsye, konstantja c e normalizimit përcaktohet nisur nga kushti i normalizimit për vlera të ndryshores $x \geq x_{min}$. Grafiku i *funksionit komplementar të shpërndarjes*⁵² $\overline{F_X}(x) = P(X \geq x)$, në një plan koordinativ me boshte të logaritmuara ka trajtë asimptotikisht lineare. Kjo mund të shërbej edhe si test empirik për të kontrolluar praninë e shpërndarjes PL.

Shpërndarja e Vazhdueshme PL. Le të jetë X një ndryshore rasti jonegative e vazhdueshme me shpërndarje PL. Nisur nga kushti i normalizimit për vlera të ndryshores $x \geq x_{min}$, kemi:

$$1 = \int_{x_{min}}^{\infty} f(x)dx = c \int_{x_{min}}^{\infty} x^{-\alpha} dx = \frac{c}{1-\alpha} [x^{-\alpha+1}]_{x_{min}}^{\infty}.$$

Ky përcaktim ka vend vetëm për ato raste kur $\alpha > 1$, për të shmangur divergjencën. Rastet e shpërndarjes PL me vlerë të konstantes eksponenciale më të vogël se 1, nuk mund të normalizohen dhe nuk hasen në realitet. Konstantja e normalizimit shprehet nga formula:

$$c = (\alpha - 1)x_{min}^{\alpha-1}.$$

Si përfundim, funksioni i densitetit probabilitar të shpërndarjes së vazhduar PL për $x \geq x_{min}$ merr trajtën:

$$f(x) = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha}.$$

Për vlera të caktuara të eksponentit α , momentet e ndryshores së vazhdueshme mund të jenë të pafundme [126]. Dallohen rastet:

- Momenti i rendit të parë,

$$E(X) = \begin{cases} \infty & \alpha \leq 2 \\ \frac{\alpha - 1}{\alpha - 2} x_{min} & \alpha > 2 \end{cases};$$

- Momenti i rendit të dytë,

⁵² Zakonisht zgjidhet të punohet me funksionin komplementar të shpërndarjes dhe jo me funksionin e densitetit probabilitar të shpërndarjes për arsye se: paraqitja pamore e funksionit komplementar të shpërndarjes është më e mirë se sa ajo e densitetit probabilitar të shpërndarjes duke patur parasysh luhatjet e mëdha që vijnë si rezultat i përmasave të fundme të zgjedhjes, veçanërisht në pjesën fundore të shpërndarjes empirike.

$$E(X^2) = \begin{cases} \infty & \alpha \leq 3 \\ \frac{\alpha - 1}{\alpha - 3} x_{min}^2 & \alpha > 3 \end{cases} ;$$

- Momenti i rendit të k-të,

$$E(X^k) = \begin{cases} \infty & \alpha \leq k + 1 \\ \frac{\alpha - 1}{\alpha - 1 - k} x_{min}^k & \alpha > k + 1 \end{cases} .$$

Le të kemi një bashkësi të dhënash $\mathbf{x} = (x_1, x_2, \dots, x_n)$, të cilat konsistojnë në n vrojtime të pavarura nga njëra-tjetra, me shpërndarje PL dhe plotësojnë kushtin $x_i \geq x_{min}$ (x_{min} supozohet i njohur). Për të përcaktuar vlerën e parametrin α zbatohet metoda e përgjasisë maksimale [121] si vijon. Funkzioni probabilitar i përgjasisë është:

$$f(\mathbf{x}) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \frac{\alpha - 1}{x_{min}} \left(\frac{x_i}{x_{min}} \right)^{-\alpha} .$$

Gjasat më të mëdha janë që të dhënat të gjenerohen nga modeli i cili ka parametrin α që maksimizon këtë funksion. Në vijim, për të përcaktuar vlerën e parametrin, marrim logaritmin e këtij funksioni, i cili maksimizohet në të njëjtën pikë sikurse vetë funksioni:

$$L = \ln f(\mathbf{x}) = \ln \left[\prod_{i=1}^n \frac{\alpha - 1}{x_{min}} \left(\frac{x_i}{x_{min}} \right)^{-\alpha} \right] = \sum_{i=1}^n \left[\ln(\alpha - 1) - \ln x_{min} - \alpha \ln \left(\frac{x_i}{x_{min}} \right) \right],$$

$$L = n \ln(\alpha - 1) - n \ln x_{min} - \alpha \sum_{i=1}^n \ln \left(\frac{x_i}{x_{min}} \right).$$

Pasi zgjidhet në lidhje me parametrin α ekuacioni:

$$\frac{\partial L}{\partial \alpha} = 0,$$

rezulton që

$$\hat{\alpha} = 1 + n \left[\sum_{i=1}^n \ln \frac{x_i}{x_{min}} \right]^{-1} .$$

Ky vlerësim për parametrin α përputhet edhe me vlerësimin e Hill [128]. Në mënyrë të ngjashme vlerësohet gabimi standart $\hat{\sigma}$ në lidhje me $\hat{\alpha}$. Në rastin e shpërndarjes së vazhdueshme PL kemi që,

$$\hat{\sigma} = \frac{\hat{\alpha} - 1}{\sqrt{n}} + O\left(\frac{1}{n}\right).$$

Funksioni komplementar i shpërndarjes në rastin e shpërndarjes së vazhdueshme PL për vlera të ndryshores $x \geq x_{min}$, jepet nga formula:

$$\overline{F_X(x)} = \left(\frac{x}{x_{min}} \right)^{-(\alpha-1)}.$$

Ky funksion është sërish PL, por me eksponent $\alpha - 1$, në rastin kur funksioni i densitetit korrespondues është me konstante eksponenciale α .

Shpërndarja Diskrete PL. Le të jetë X një ndryshore rasti diskrete me shpërndarje PL. Vlera më e vogël që mund të marr ndryshorja është 1. Në vlerën zero funksioni i masës probabilitare nuk është i përcaktuar. Për të përcaktuar konstanten c , përdoret kushti i normalizimit:

$$1 = \sum_{x=1}^{\infty} f(x) = c \sum_{x=1}^{\infty} x^{-\alpha} = c\zeta(\alpha),$$

ku $\zeta(\alpha) = \sum_{x=1}^{\infty} x^{-\alpha}$ është funksioni ζ -Riemann. Prej këtij, rrjedh se konstantja $c = \frac{1}{\zeta(\alpha)}$ dhe se funksioni i masës probabilitare të shpërndarjes PL jepet nga formula:

$$f(x) = \frac{x^{-\alpha}}{\zeta(\alpha)}.$$

Për $x \geq x_{min}$, funksioni i masës probabilitare merr trajtën:

$$f(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{min})},$$

ku $\zeta(\alpha, x_{min}) = \sum_{x=x_{min}}^{\infty} x^{-\alpha}$ është funksioni i përgjithësuar ζ -Riemann.

Dallojmë rastet si vijon për momentet e ndryshores diskrete:

- Momenti i rendit të parë,

$$E(X) = \frac{\alpha - 1}{\alpha - 2};$$

- Momenti i rendit të dytë,

$$E(X^2) = \frac{(\alpha - 1)^2}{(\alpha - 2)(\alpha - 3)}.$$

Sikurse në shpërndarjen e vazhueshme PL, edhe në këtë rast ndiqet e njëjta logjikë [121]. Për $x \geq x_{min}$ vlerësimi i konstantes eksponenciale α kryhet si vijon:

$$f(x, \alpha) = \prod_{i=1}^n f(x_i, \alpha) = \prod_{i=1}^n \frac{x_i^{-\alpha}}{\zeta(\alpha, x_{min})},$$

$$L = \ln f(x, \alpha) = \ln \left[\prod_{i=1}^n \frac{x_i^{-\alpha}}{\zeta(\alpha, x_{min})} \right] = \sum_{i=1}^n [\ln x_i^{-\alpha} - \ln \zeta(\alpha, x_{min})],$$

$$L = -\alpha \sum_{i=1}^n \ln x_i - n \ln \zeta(\alpha, x_{min}),$$

$$\frac{\partial L}{\partial \alpha} = 0,$$

$$-\frac{n}{\zeta(\alpha, x_{min})} \frac{\partial}{\partial \alpha} [\zeta(\alpha, x_{min})] - \sum_{i=1}^n \ln x_i = 0.$$

Vlerësimi për konstanten eksponenciale α është zgjidhja e marr nga ekuacioni:

$$\frac{\zeta'(\hat{\alpha}, x_{min})}{\zeta(\hat{\alpha}, x_{min})} = -\frac{1}{n} \sum_{i=1}^n \ln x_i.$$

Një tabelë për vlerat e këtij raporti mund të gjendet në [129], ose ato mund të gjenden nëpërmjet metodave numerike të instaluar në softe kompjuterike. Gabimi standart i vlerësuar, në lidhje me $\hat{\alpha}$ në rastin diskret, është

$$\hat{\sigma} = \frac{1}{\sqrt{n \left[\frac{\zeta''(\hat{\alpha}, x_{min})}{\zeta(\hat{\alpha}, x_{min})} - \left(\frac{\zeta'(\hat{\alpha}, x_{min})}{\zeta(\hat{\alpha}, x_{min})} \right)^2 \right]}}.$$

Clauset *et al.* [121] pohon se në rastet kur $x_{min} \geq 6$, mund të bëhet një përafrim i arsyeshëm përgjatë metodës ML në vlerësimin e parametrin α për shpërndarjen diskrete PL. Tregohet se:

$$\frac{\zeta'(\hat{\alpha}, x_{min})}{\zeta(\hat{\alpha}, x_{min})} = - \left[\frac{1}{\hat{\alpha} - 1} + \ln \left(x_{min} - \frac{1}{2} \right) \right] [1 + O(x_{min}^{-2})],$$

prej nga, duke shpërfillur termin x_{min}^{-2} nëpërmjet krahasimit me madhësi të rendit të parë, kemi që

$$\hat{\alpha} \cong 1 + n \left[\sum_{i=1}^n \ln \left(\frac{x_i}{x_{min} - \frac{1}{2}} \right) \right]^{-1},$$

e cila përngjan me vlerësimin e bërë për α nëpërmjet metodës ML në rastin e shpërndarjes së vazhduar, me një diferencë me $\frac{1}{2}$ në emërues. Funkzioni komplementar i shpërndarjes, për vlera të ndryshores $x \geq x_{min}$, në rastin e shpërndarjes diskrete PL jepet nga formula:

$$\overline{F_X(x)} = \frac{\zeta(\alpha, x)}{\zeta(\alpha, x_{min})}.$$

4.2.1.2. Shpërndarja Lognormale (LN)

Në shumë studime shkencore shpërndarjet empirike janë modeluar nëpërmjet shpërndarjes LN [130]. Një ndryshore rasti X thuhet se ka shpërndarje LN [112] me parametra μ dhe σ^2 nëse ndryshorja $Y = \ln X$ ka shpërndarje normale me pritje matematike μ dhe variancë σ^2 . Funkzioni i densitetit të shpërndarjes normale të ndryshores Y jepet nga formula:

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(y - \mu)^2}{2\sigma^2} \right],$$

ku $\mu = E(Y)$ është pritja matematike, σ është devijimi standart i ndryshores Y , ndërkohë që σ^2 është varianca e kësaj ndryshore. Vlerat e ndryshores variojnë $y \in (-\infty, +\infty)$.

Funksioni i densitetit probabilitar të shpërndarjes LN jepet nga formula:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp \left[-\frac{(\ln x - \mu)^2}{2\sigma^2} \right],$$

dhe funksioni komplementar i shpërndarjes jepet nga formula:

$$\overline{F_X(x)} = \int_x^\infty \frac{1}{\sqrt{2\pi}\sigma t} \exp \left[-\frac{(\ln t - \mu)^2}{2\sigma^2} \right] dt.$$

Shpërndarja LN është asimetrike. Pritja matematike e saj është $E(X) = \exp \left[\mu + \frac{1}{2}\sigma^2 \right]$, Mensorja $(X) = \exp(\mu)$ dhe Moda $(X) = \exp[\mu - \sigma^2]$. Në dallim nga shpërndarja PL, në shpërndarjen LN kemi që $E(X) < \infty$ dhe $\sigma^2 < \infty$.

Pavarësisht faktit se funksionet e momenteve në shpërndarjen LN janë të fundme, ajo është shumë e ngjashme me shpërndarjen PL. Nëse X ka shpërndarje LN, atëherë grafiku i funksionit komplementar të shpërndarjes (ose ai i funksionit të densitetit probabilitar) në një plan koordinativ me boshte të logaritmuara paraqet një sjellje afërsisht lineare në pjesën më të madhe të tij. Në rastin kur varianca e shpërndarjes normale korresponduese është shumë e madhe, lineariteti në këtë grafik është shumë i theksuar. Ky fakt gjen shpjegim edhe duke parë logaritmin e funksionit të densitetit të shpërndarjes lognormale:

$$\begin{aligned} \ln f(x) &= -\ln x - \ln \sqrt{2\pi}\sigma - \frac{(\ln x - \mu)^2}{2\sigma^2}, \\ \ln f(x) &= -\frac{(\ln x)^2}{2\sigma^2} + \left(\frac{\mu}{\sigma^2} - 1 \right) \ln x - \ln \sqrt{2\pi}\sigma - \frac{\mu^2}{2\sigma^2}. \end{aligned}$$

Në rastin kur σ është mjaft e madhe, termi kuadratik $\frac{(\ln x)^2}{2\sigma^2}$ do të ketë një vlerë mjaft të vogël për një hapësirë të konsiderueshme vlerash të ndryshores X . Në këtë mënyrë, logaritmi i funksionit të densitetit të kësaj shpërndarjeje ka trajtë thuajse lineare në një plan koordinativ me boshte të logaritmuara. Në rast se interesohemi të përqasim shpërndarjen empirike me anë të shpërndarjes LN për $x \geq x_{min}$, konstantja c e normalizimit përcaktohet nisur nga kushti i normalizimit:

$$1 = \int_{x_{min}}^\infty f(x, x_{min}) dx = c \int_{x_{min}}^\infty \frac{1}{\sqrt{2\pi}\sigma x} \exp \left[-\frac{(\ln x - \mu)^2}{2\sigma^2} \right] dx,$$

dhe jepet nga formula:

$$c = 2 \left[\operatorname{erfc} \left(\frac{\ln x_{min} - \mu}{\sqrt{2}\sigma} \right) \right]^{-1}.$$

Për rrjedhoj, funksioni i densitetit probabilitar merr trajtën:

$$f(x) = \frac{2}{\sqrt{2\pi}\sigma x} \left[\operatorname{erfc} \left(\frac{\ln x_{\min} - \mu}{\sqrt{2}\sigma} \right) \right]^{-1} \exp \left[-\frac{(\ln x - \mu)^2}{2\sigma^2} \right],$$

ku erfc ⁵³ është funksioni komplementar i gabimit Gaussian⁵⁴:

$$\operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt,$$

dhe erf ⁵⁵ është funksioni i gabimit Gaussian⁵⁶,

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

Funksioni komplementar i shpërndarjes LN, për vlera të ndryshores $x \geq x_{\min}$ jepet nga formula:

$$\overline{F_X(x)} = \int_x^{\infty} \frac{2}{\sqrt{2\pi}\sigma x} \left[\operatorname{erfc} \left(\frac{\ln x_{\min} - \mu}{\sqrt{2}\sigma} \right) \right]^{-1} \exp \left[-\frac{(\ln x - \mu)^2}{2\sigma^2} \right] dt.$$

4.2.2. Vlerësimi dhe Testimi i $x_{\min}$ të Vlerësuar

Saktësia në vlerësimin e parametrave të panjohur të një shpërndarje varet në mënyrë direkte nga saktësia e vlerësimit të kufirit të poshtëm $x_{\min}$ të të dhënave $\mathbf{x} = (x_1, x_2, \dots, x_n)$, mbi të cilat do të bëhet përqasja. Për të përqasur shpërndarjen PL ose LN, zakonisht $x_{\min}$ është vlerësuar nisur nga grafiku i funksionit komplementar ose funksionit të densitetit probabilitar të shpërndarjes empirike të ndërtuar në një plan koordinativ me boshte të logaritmuara. Nisur nga paraqitja grafike, është përcaktuar pika prej nga ku grafiku mund të konsiderohej linear ose afërsisht linear.

Përgjatë këtij studimi, vlerësimet e bëra janë bazuar mbi funksionin komplementar të shpërndarjes. Ky funksion është më i “fuqishëm” ndaj luhatjeve në pjesën fundore të grafikut, të cilat vijnë si pasojë e përmasave të fundme të zgjedhjes. Një metodë tjetër sugjeron ndërtimin e grafikut të pikave $(\hat{x}_{\min}, \hat{\alpha})$, dhe shikohet se për cfarë vlerash kemi qëndrueshmëri. Vlerësimi pamor shoqërohet me subjektivitet, e për rrjedhojë është një metodë që mbart gabime.

Në vijim, për të vlerësuar $x_{\min}$ [131], është ndjekur ideja themelore e cila pohon se: zgjidhet ajo vlerë për $\hat{x}_{\min}$, në mënyrë që shpërndarja empirike dhe ajo teorike hipotetike⁵⁷ të jenë sa më shumë të ngjashme me njëra-tjetrën për të gjitha $x \geq \hat{x}_{\min}$. Ndonëse ka disa mënyra për të matur distancën ndërmjet një shpërndarje teorike dhe asaj empirike [121, 122], kriteri Kolmogorov-Smirnov (KS) [127] është më i përdoruri për të dhënat të cilat nuk janë me shpërndarje normale. Ai jep rezultate më të mira se sa *Kriteri Bayesian i Informimit*⁵⁸ dhe ka lehtësi nga pikëpamja matematikore. Statistika e kriterit KS jepet nga:

⁵³ Shkurtime nga anglishtja i *The complementary Gaussian error function*.

⁵⁴ Përshtatur nga anglishtja: *The complementary Gaussian error function*.

⁵⁵ Shkurtime nga anglishtja i *The Gauss error function*.

⁵⁶ Përshtatur nga anglishtja: *The Gauss error function*.

⁵⁷ Me “hipotetike” në këtë rast nënkuptohet një shpërndarje e caktuar teorike për të cilën hamendësojmë se mund të përqas mirë shpërndarjen empirike të të dhënave.

⁵⁸ Përshtatur nga anglishtja: *Bayesian Information Criterion*.

$$D = \max_{x \geq x_{min}} |S(x) - P(x)|,$$

ku $S(x)$ është funksioni komplementar i shpërndarjes empirike për të dhënat të cilat plotësojnë kushtin $x \geq x_{min}$, ndërsa $P(x)$ është funksioni komplementar i shpërndarjes sipas modelit teorik hipotetik që përfaqëson më mirë të dhënat. Vlerësimi për $\hat{x}_{min}$ bazohet tek ajo vlerë e x_{min} që minimizon statistikën KS:

$$\hat{x}_{min} = \inf_{x_{min} \in \mathcal{X}} (\max_{x \geq x_{min}} |S(x) - P(x|\hat{\alpha}, \hat{x}_{min})|).$$

Kriteri KS bën vlerësime të mira për x_{min} , është i thjeshtë për tu përdorur dhe ka një “kosto” kompjuterike shumë të vogël. Përparësia më e madhe e kësaj metode është se: shpërndarjet tërësisht gjatë vlerësimit të x_{min} , atë pjesë të të dhënave të cilat nuk ndjekin shpërndarjen teorike hipotetike. Kjo krijon lehtësi në kontrollin e mirësisë së përfaqësjes.

Për të matur saktësinë në vlerësimin e x_{min} është përdorur metoda joparametrike “bootstrap” [132]. Nisur nga një zgjedhje me përmasë n , është gjeneruar një bashkësi sintetike të dhënash me shpërndarje të njëjtë sikurse shpërndarja fillestare, duke ndërtuar një bashkësi të re pikash x_i , për $i = 1, \dots, n$ në mënyrë rastësore dhe të njëtrajshme nga të dhënat fillestare (me zëvendësim). Paskëta, është vlerësuar x_{min} dhe α në këtë bashkësi zëvendësuese. Ky proces është përsëritur një numër të madh herësh (psh. 1000 ose më shumë) dhe mbi bashkësinë e $\{x_{min}^i\}$ dhe $\{\alpha^i\}$, për $i = 1, \dots, 1000$ është marrë devijimi standart i vlerësimeve të bëra. Në këtë mënyrë është bërë një vlerësim mbi pasigurinë në lidhje me parametrat e vlerësuar të shpërndarjes.

Në eksperimentet e bëra [122] është gjetur se vlerësimet e bëra janë asimptotikisht të qëndrueshme ($\hat{x}_{min} \rightarrow x_{min}$ dhe $\hat{\alpha} \rightarrow \alpha$) për $n_{fund} \rightarrow \infty$. Zvogëlimi i numrit të të dhënave në pjesën fundore të shpërndarjes, rrit gabimin statistikor gjatë metodës ML në vlerësimin e parametrave të shpërndarjes.

4.2.3. Kontrolli i Hipotezave në Lidhje me Shpërndarjen Hipotetike

Në shumë studime empirike kontrolli mbi një shpërndarje të caktuar, nëse ajo përfaqëson mirë të dhënat ose jo, është bazuar në paraqitjen pamore të shpërndarjes modeluese mbi të dhënat. Kjo procedurë priret shumë nga subjektiviteti dhe si e tillë mbart gabime të mëdha.

Kur synohet modelimi i të dhënave nëpërmjet një shpërndarje të caktuar, sfida qëndron në bërjen e dallimit, nëse devijimi i të dhënave nga shpërndarja që synohet të përfaqësojë ato është si rezultat i procesit të grumbullimit të tyre, apo është tregues që kemi të bëjmë me një shpërndarje të papërshtatshme. Përfaqësja themelore për këtë rast është të marrim një numër shumë të madh bashkësish sintetike nga shpërndarja teorike për të cilën hamendësojmë dhe ti krahasojmë rezultatet me matje të ngjashme në të dhënat empirike. Nëse të dhënat empirike largojnë shumë në krahasim me ato sintetike, atëherë shpërndarja për të cilën hamendësojmë nuk është e duhura. Në lidhje me këtë përfaqësim duhen pasur parasysh dy aspekte:

- Mirësia e përfaqësjes që bën një shpërndarje teorike hipotetike varet nga mënyra se si ne matim distancën midis shpërndarjes empirike dhe asaj teorike. Në këtë studim është përdorur statistika KS.
- Kur synojmë të kontrollojmë një hipotezë në lidhje me një shpërndarje të caktuar, është e rëndësishme të merren përmasa të mëdha të zgjedhjes. Nëse zgjedhja ka përmasë të vogël, ekzistojnë mundësi që të dhënat

empirike të modelohen nga një shpërndarje tjetër, që nuk është shpërndarja e vërtetë e të dhënave.

Procedura që është ndjekur për të kontrolluar mirësinë e përqasjes së një shpërndarje teorike hipotetike ndaj të dhënave empirike, është e njëjtë me procedurën e ndjekur edhe nga Clauset *et al.* në [121]. Kjo procedurë kalon në këto hapa:

- Përqasim shpërndarjen teorike hipotetike mbi të dhënat empirike dhe përlllogaritim statistikën KS për këtë përqasje. Nëse vlera e saj është e vogël, atëherë ajo shërben si tregues se shpërndarja hipotetike nuk mund të rrëzohet poshtë si një shpërndarje e mundshme për të modeluar të dhënat. Përndryshe, nëse vlera e saj është e madhe, atëherë shpërndarja është e papërshtatshme.
- Gjenerojmë një numër shumë të madh bashkësish sintetike me të dhëna me shpërndarje me parametra dhe x_{min} të njëjtë me shpërndarjen hipotetike me anë të së cilës synojmë të modelojmë të dhënat. Për secilën nga bashkësitë sintetike, përqasim shpërndarjen hipotetike dhe përlllogaritim statistikën KS në përputhje me modelin e shpërndarjes që përshtatet më mirë me të dhënat.
- Numërohet se sa herë statistika KS ka rezultuar më e madhe se vlera e saj në të dhënat empirike. Përpjesa e këtij numri herësh mbi numrin total të herëve, përbën atë që quhet p -vlera e kriterit KS.

Gjatë gjenerimit të bashkësive sintetike të të dhënave bëhet kujdes në mënyrë që shpërndarja e tyre, për të gjitha ato të dhëna që do të jenë më të vogla se x_{min} të jetë e njëjtë me shpërndarjen empirike të të dhënave dhe për të gjitha ato që do të jenë më të mëdha se x_{min} të ndjekin shpërndarjen teorike hipotetike. Kjo është shumë e rëndësishme në mënyrë që të gjendet një p -vlerë e saktë. Gjenerimi i bashkësive të të dhënave që plotësojnë kushtin e mësipërm realizohet nëpërmjet një qasje gjysëmparametrike.

Supozojmë se nga të dhënat e vrojtuar, të cilat janë n në total, me n_{fund} është shënuar numri i atyre që plotësojnë kushtin $x \geq x_{min}$. Bashkësia sintetike prej n të dhënash gjenerohet si vijon: me probabilitet $\frac{n_{fund}}{n}$ gjenerohet një numër rasti x_i nga shpërndarja teorike hipotetike (me parametrat e vlerësuar) dhe që plotëson kushtin $x_i \geq x_{min}$. Përndryshe, me probabilitet $1 - \frac{n_{fund}}{n}$ zgjidhet në mënyrë rastësore, të njëtrajtshme, një element nga bashkësia e të dhënave të vrojtuar i tillë që $x < x_{min}$ dhe në këtë rast x_i i jepet vlera e atij elementi. Ky proces përsëritet për $i = 1, \dots, n$ derisa bashkësia sintetike të ketë n të dhëna në total.

Për të patur një p -vlerë me saktësi deri në dy shifra pas presjes dhjetore, duhet të gjenerohen 2500 bashkësi sintetike të dhënash. Hipoteza fillestare (H_0 : “Të dhënat modelohen sipas shpërndarjes hipotetike”) refuzohet në rast se $p \leq 0.1$. Sipas Clauset *et al.* [122], sugjerohet të përdoret niveli $\alpha = 0.1$ dhe jo ai 0.05, pasi një nivel rëndësie më i vogël se 0.1 do të kualifikonte si shpërndarje të mundshme për të modeluar të dhënat disa shpërndarje të papërshtatshme. Gjithashtu, për $n_{fund} \leq 100$ në rastin e shpërndarjes PL dhe për $n_{fund} \leq 300$ në rastin e shpërndarjes LN, është e vështirë që $p \leq 0.1$ [121, 122]. Në këto raste, bashkësi më të mëdha të dhënash do të duheshin për të refuzuar hipotezën fillestare, pavarësisht se shpërndarja hipotetike nuk do ti modelonte më së miri ato.

4.2.4. Krahassimi Direkt i Shpërndarjeve

Për të krahasuar në mënyrë direkte dy shpërndarje teorike hipotetike, të cilat njëkohësisht kanë rezultuar si të mundshme për të modeluar shpërndarjen empirike, është përdorur kriteri i raportit të përgjasive i propozuar nga Vuong [133], detaje të të cilit jepen në [121]. Le të jetë $\mathbf{x} = (x_1, x_2, \dots, x_n)$, një bashkësi të dhënash të cilat konsistojnë në n vrojtme të pavarura nga njëra-tjetra. Marrim në shqyrtim dy shpërndarje të cilat modelojnë të dhënat. Le të jenë $f_1(x)$ dhe $f_2(x)$ funksionet e densiteteve/masave probabilitare të tyre. Funksionet e përgjasive në lidhje me të dy shpërndarjet e marra në shqyrtim janë:

$$L_1 = f_1(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f_1(x_i),$$

dhe

$$L_2 = f_2(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f_2(x_i).$$

Funksioni i raportit të përgjasive jepet nga formula:

$$R = \frac{L_1}{L_2} = \prod_{i=1}^n \frac{f_1(x_i)}{f_2(x_i)}.$$

Nëse marrim logaritmet e të dyja anëve në funksionin e raportit të përgjasive kemi që funksioni i logaritmit të raportit të përgjasive është:

$$\mathcal{R} = \ln(R) = \sum_{i=1}^n [\ln f_1(x_i) - \ln f_2(x_i)] = \sum_{i=1}^n [L_i^{(1)} - L_i^{(2)}],$$

ku $L_i^{(j)} = \ln f_j(x_i)$, mund të shikohet si funksioni i log-përgjasisë për një matje të vetme x_i , për $i = 1, \dots, n$ nga shpërndarja j , për $j = 1, 2$. Diferencat $L_i^{(1)} - L_i^{(2)}$ janë të pavarura si rrjedhim i pavarësisë së x_i . Në saj të teoremës qendrore limite, shuma e tyre $\mathcal{R}$ ka shpërndarje normale për vlera shumë të mëdha të përmasës së zgjedhjes n , me variancë $n\sigma^2$ (σ^2 është varianca e një termi të vetëm). Sikurse ndodh zakonisht kjo variancë, mund të përafrohet me variancën e të dhënave:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n \left[\left(L_i^{(1)} - L_i^{(2)} \right) - \left(\bar{L}^{(1)} - \bar{L}^{(2)} \right) \right]^2,$$

ku

$$\bar{L}^{(1)} = \frac{1}{n} \sum_{i=1}^n L_i^{(1)}, \quad \bar{L}^{(2)} = \frac{1}{n} \sum_{i=1}^n L_i^{(2)}.$$

Në rastin kur vlera e pritur e logaritmit të raportit të përgjasive është vlera zero, atëherë shenja e vrojtuar e $\mathcal{R}$ konsiderohet rezultat i luhatjeve dhe nuk mund të shërbej si një tregues i besueshëm për të përcaktuar se cili nga modelet e shpërndarjeve është

më i miri. Probabiliteti që logaritmi i raportit të përgjasive të matura është me një magnitudë të madhe po aq sa, apo edhe më të madhe, në krahasim me vlerën e vërtetuar $|\mathcal{R}|$ jepet nëpërmjet formulës:

$$p = \frac{1}{\sqrt{2\pi n\sigma^2}} \left[\int_{-\infty}^{-|\mathcal{R}|} e^{-\frac{t^2}{2n\sigma^2}} dt + \int_{|\mathcal{R}|}^{\infty} e^{-\frac{t^2}{2n\sigma^2}} dt \right] = \left| \operatorname{erfc} \left(\frac{|\mathcal{R}|}{\sqrt{2n\sigma^2}} \right) \right|.$$

Në këtë rast p -vlera bën një vlerësim për probabilitetin që ne kemi matur një vlerë të caktuar $\mathcal{R}$, ndërkohë që vlera e vërtetë e $\mathcal{R}$ është afërsisht zero. Nëse p -vlera është e vogël (përgjithësisht $p < 0.1$), atëherë vlera që kemi gjetur për $\mathcal{R}$ nuk konsiderohet si një rezultat i rastësishëm. Shenja e $\mathcal{R}$ në këtë rast shërben si tregues i besueshëm për të përcaktuar se cila shpërndarje prej atyre dy që krahasohen është më e mira. Nëse p -vlera e përlogarituar është e madhe, atëherë kriteri i përgjasisë së raportit është i papërshtatshëm për të bërë dallimin midis dy shpërndarjeve. p -vlera është efektive veçanërisht në ato raste kur numri i të dhënave të marra në shqyrtim nuk është i mjaftueshëm për të bërë dallimin midis dy shpërndarjeve [121]. E meta e metodës së Vuong është se kjo metodë nuk mund të tregoj se cila prej dy shpërndarjeve që krahasohen është më e “dobta” në përqasjen e të dhënave. Kjo metodë mundet të na tregoj se cila ndër dy shpërndarjet që krahasohen është e “keqja” më e vogël.

4.2.5. Paraqitje e Përmbledhur e Vlerësimeve të Parametrave të Shpërndarjeve

Krahasimi i vlerësimeve të bëra për $x_{\min}^i$ për $i = 1, \dots, 30$, për shpërndarjet PL dhe LN në modelet e serive kohore të graf rrjetave të trajtuara në këtë pjesë, është kryer nëpërmjet grafikëve me kuti. Për të vlerësuar shpërndarjen e cila përqas më mirë të dhënat, janë përdorur karakteristikat si vijon: hapësira e të dhënave që ajo përqas më mirë, besueshmëria e vlerësimeve për të, numri i refuzimeve dhe vlerat e statistikës KS. Për këtë janë realizuar përqasjet në tre raste:

- Rasti 1: përqasja e bërë që nga 1;
- Rasti 2: përqasja e bërë që nga $x_{\min}$ i vlerësuar për shpërndarjen;
- Rasti 3: përqasja e bërë që nga $\max(x_{\min}^{PL}, x_{\min}^{LN})$, në rastin kur të dyja shpërndarjet janë vlerësuar si njësoj të mundshme.

Për shpërndarjen e vlerësuar si më e mira për të modeluar të dhënat është bërë një përshkrim numerik i vlerësuesve të parametrave të saj në Rastin 1 dhe Rastin 2. Gjithashtu, për të tre rastet e mësipërme jepen grafikisht funksionet komplementare të shpërndarjes ($(\overline{F_X}(x) = P(X \geq x))$, ku $X = d_v$) në grafikë me boshte të logaritmuara për graf rrjetat⁵⁹ G_1 , G_{15} dhe G_{30} .

4.2.6. Korrelacioni Ditor Ndërmjet Fuqive Hyrëse dhe Dalëse për Klient

A mund të thuhet se kulmet qendrore hyrëse⁶⁰ janë njëkohësisht edhe kulme qendrore dalëse⁶¹? Nëse dy kulme të ndryshëm nga njëri-tjetri kanë respektivisht njëri fuqi hyrëse të lartë dhe tjetri fuqi hyrëse të ulët, a mund të supozojmë se të dy këta kulme kanë

⁵⁹ Në pamundësi hapësire për të paraqitur rastet për të 30-të graf rrjetat e serisë, janë zgjedhur për ilustrim pamor tre prej tyre.

⁶⁰ Kulmet me fuqi hyrëse të lartë.

⁶¹ Kulme me fuqi dalëse të lartë.

mundësi për të patur fuqi dalëse të lartë? A mund të pranojmë se mungon korrelacioni ndërmjet fuqive hyrëse dhe dalëse për kulm?

Pyetjet e mësipërme lidhen ngushtë me pyetjen kryesore **P4.3**. Për t'iu përgjigjur këtyre pyetjeve është përdorur analiza korrelative⁶² në Modelin 4.2 ndërmjet fuqive hyrëse dhe dalëse në secilën graf rrjet G_i . Koeficientët e korrelacionit sipas Kendall dhe Spearman janë njehsuar në shpërndarjen e bashkuar empirike të fuqive hyrëse dhe dalëse dhe më pas është kontrolluar domethënia statistikore e tyre.

Kulmet qëndrore hyrëse, të cilët mund ti quajmë “dëgjues”, në një graf rrjet të telefonisë përfaqësojnë individë/subjekte të cilat marrin telefonata nga shumë të tjerë. Kulmet qëndrore dalëse, të cilat mund ti quajmë “folës”, përfaqësojnë individë/subjekte të cilat kryejnë telefonata në drejtim të shumë individëve/subjekteve të tjerë. Nëse jemi të interesuar për të përhapur në mënyrë të shpejtë një mesazh në graf rrjet, atëherë do të qe me vend që “folësit” të ishin “dëgjues”, përndryshe informacioni do të kishte vështirësi të përhapej në mënyrë të shpejt në graf rrjet.

4.3. REZULTATET E HULUMTIMIT DHE DISKUTIME

Në graf rrjetat e rastit çdo brinjë është e pranishme ose jo me të njëjtin probabilitet. Për rrjedhoj, shpërndarja e fuqive të kulmeve është binomiale, ose shpërndarje Poisson në rastin kur rendi i graf rrjetit është shumë i madh [68, 1]. Këto shpërndarje nuk janë asimetrike, të anuara në të djathtë. Shpërndarja e fuqive të kulmeve në graf rrjetat reale të komunikimit nëpërmjet telefonisë shmanget dallueshëm nga këto shpërndarje.

Fuqia e kulmeve është tregues se sa i rëndësishëm apo aktiv është një klient i telefonisë. Rezultatet e hulumtimit për shpërndarjen ditore të numrit të komunikuesve të ndryshëm për klient, jepen nëpërmjet njehsimeve në Modelin 4.1. Ndërsa, hulumtimi në lidhje me numrin e thirrjeve telefonike ditore (totale, hyrëse dhe dalëse) për klient jepet nëpërmjet njehsimeve në Modelin 4.2.

4.3.1. Analiza e Shpërndarjeve Ditore të Fuqive të Kulmeve

Bashkësia e të dhënave të cilat analizohen në këtë pjesë është vargu i fuqive $\{d_v^i\}_{v \in V_i}$ të secilit graf rrjet G_i , për $i = \overline{1,30}$ në Modelet 4.1 dhe 4.2. Në Modelin 4.1, analizohen fuqitë d_v^i , të cilat tregojnë numrin e komunikuesve të ndryshëm ditor të kulmit (klientit) v në graf rrjetin G_i . Ndërsa në Modelin 4.2, analizohen fuqitë: d_v^i , numri total i thirrjeve telefonike ditore (Modeli 4.2-Totale), d_v^{i-} , numri i thirrjeve hyrëse (Modeli 4.2-Hyrëse) dhe d_v^{i+} numri dalëse (Modeli 4.2-Dalëse) prej v në graf rrjetin G_i .

4.3.1.1. Përshkrimi i Vlerave të Fuqive të Kulmeve

Tabela 4.1 bën një përshkrim numerik të treguesve të vendosjes si: mesatarja, mesorja, moda, kuartili i parë (Q_1), kuartili i tretë (Q_3), minimumi (min.), maksimumi (max.), moda; dhe treguesve të formës si: koeficientit të asimetrisë (γ_1) dhe atij të sheshtësisë (γ_2). Ilustrimi jepet për graf rrjetat G_1 , G_{15} dhe G_{30} të serive.

⁶² Më tepër detaje teorike për këtë metodë jepen në seksionin, 2.2.3 Analiza Korrelative.

Tabelë 4.1: Përshkrim numerik i treguesve të vendosjes dhe të formës për vargjet e fuqive në G_1 , G_{15} dhe G_{30} në Modelin 4.1 dhe 4.2.

		Min.	Q_1	Mesorja	Q_3	Max.	Mesatarja	γ_1	γ_2	Moda
Modeli 4.1	d_v^1	1	1	2	3	66	2.79	7.25	108.41	1
	d_v^{15}	1	1	2	3	60	2.89	5.88	71.46	1
	d_v^{30}	1	1	2	3	49	2.77	5.28	57.92	1
Modeli 4.2	d_v^{1-}	0	0	1	2	82	1.53	15.00	369.74	1
	d_v^{1+}	0	0	1	2	26	1.53	3.70	24.11	0
	d_v^1	1	1	2	4	88	3.7	8.92	158.03	1
	d_v^{15-}	0	0	1	2	79	1.60	14.57	362.49	1
	d_v^{15+}	0	0	1	2	39	1.60	5.03	51.00	0
	d_v^{15}	1	1	2	4	84	3.20	7.65	119.34	1
	d_v^{30-}	0	0	1	2	64	1.54	11.09	212.99	1
	d_v^{30+}	0	0	1	2	20	1.54	2.98	15.08	0
	d_v^{30}	1	1	2	4	66	3.08	6.43	83.99	1

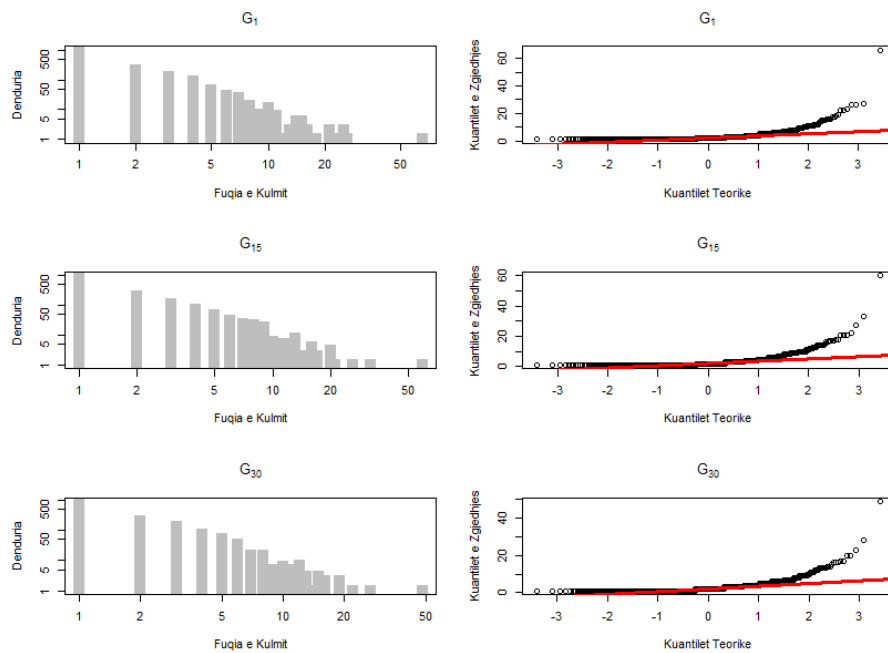
Figurat e mëposhtme (Figura 4.1, Figura 4.2, Figura 4.3, dhe Figura 4.4) japin histogramet në boshte të logaritmuara dhe grafikët normal kuantil-kuantil. Histogramet bëjnë një vlerësim pamor mbi shpërndarjen empirike të të dhënave. Në të gjitha rastet në grafikët normal kuantil-kuantil, u pa një shmangie e fortë nga vija e drejtë.

Gjithashtu, pas kryerjes së kontrollit për normalitetin me anë të kriterit Shapiro-Wilk në secilin prej graf rrjetave në Modelin 4.1 dhe 4.2, në vargjet korresponduese të fuqive të kulmeve, u gjet se p -vlerat në të gjitha rastet janë më pak se $2.2 \times 10^{-16} < 0.01$. Kjo nënkupton që të dhënat nuk i përkasin një popullimi me shpërndarje normale.

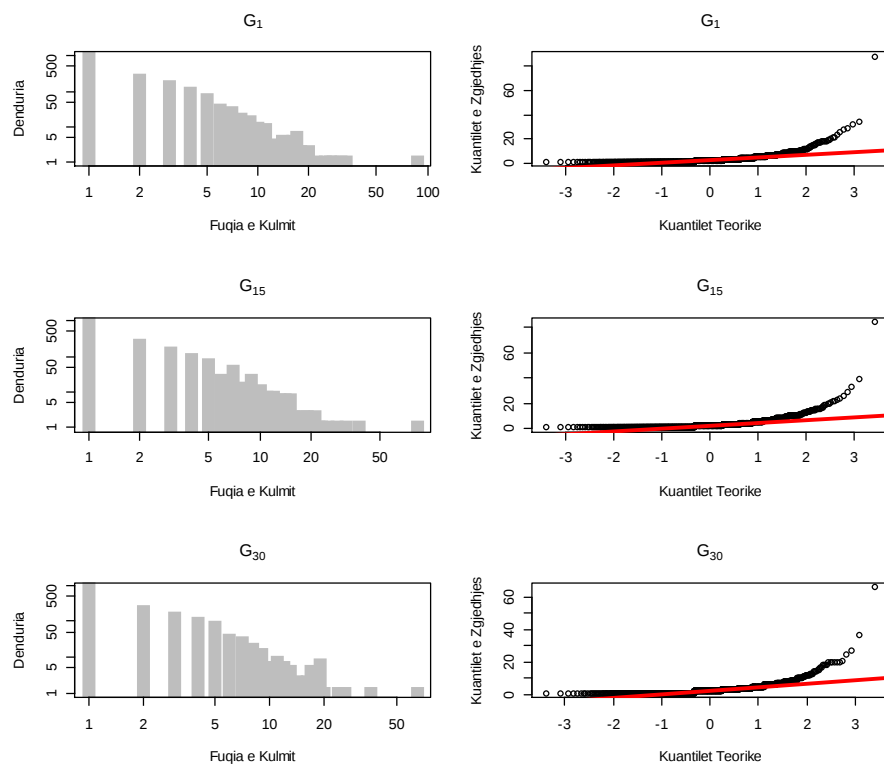
Për më tepër, në të gjitha rastet për $i = 1, \dots, 30$, në vargjet e fuqive $\{d_v^i\}_{v \in V_i}$, $\{d_v^{i-}\}_{v \in V_i}$ dhe $\{d_v^{i+}\}_{v \in V_i}$ rezulton se:

- Vargu i fuqive është njëmodal;
- Moda është 1 për $\{d_v^i\}_{v \in V_i}$, $\{d_v^{i-}\}_{v \in V_i}$ dhe 0 për $\{d_v^{i+}\}_{v \in V_i}$;
- Mesatarja $>$ mesorja $>$ moda për $\{d_v^i\}_{v \in V_i}$, $\{d_v^{i-}\}_{v \in V_i}$ dhe mesatarja $>$ mesorja $\geq$ moda për $\{d_v^{i+}\}_{v \in V_i}$;
- Kulmi i të dhënave (vlera që ka dëndurinë më të madhe) është në të majtë dhe pjesa fundore e shpërndarjes në të djathtë është më e gjatë;
- Koeficienti i asimetrisë është më i madh se 0;
- Koeficienti i sheshtësisë është më i madh se 0.

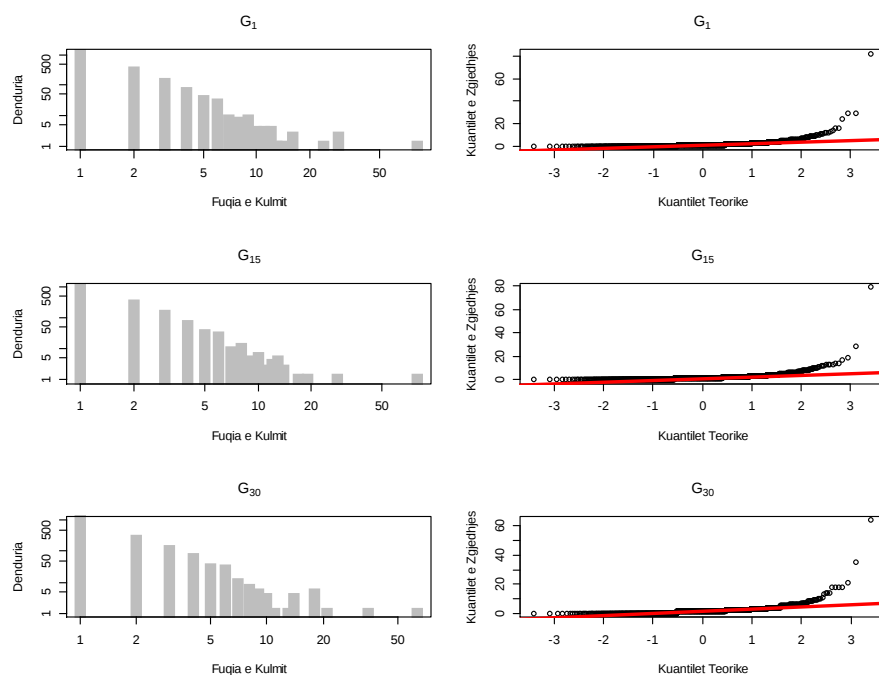
Kjo nënkupton që për vargjet e fuqive $(\{d_v^i\}_{v \in V_i}, \{d_v^{i-}\}_{v \in V_i} \text{ dhe } \{d_v^{i+}\}_{v \in V_i})$, është e pranishme një shpërndarje asimetrike pozitive e tipit leptokurtic. Për rrjedhoj, shpërndarjet empirike janë shpërndarje të anuara, me pjesë fundore të rrëzuara ku pjesa fundore e shpërndarjes në të djathtë është më e gjatë krahasuar me atë në të majtë. Shpërndarjet e fuqive të kulmeve [134] të cilat janë shpërndarje asimetrike me pjesë fundore të rrëzuara janë tregues për një sjellje heterogjene të sistemit.



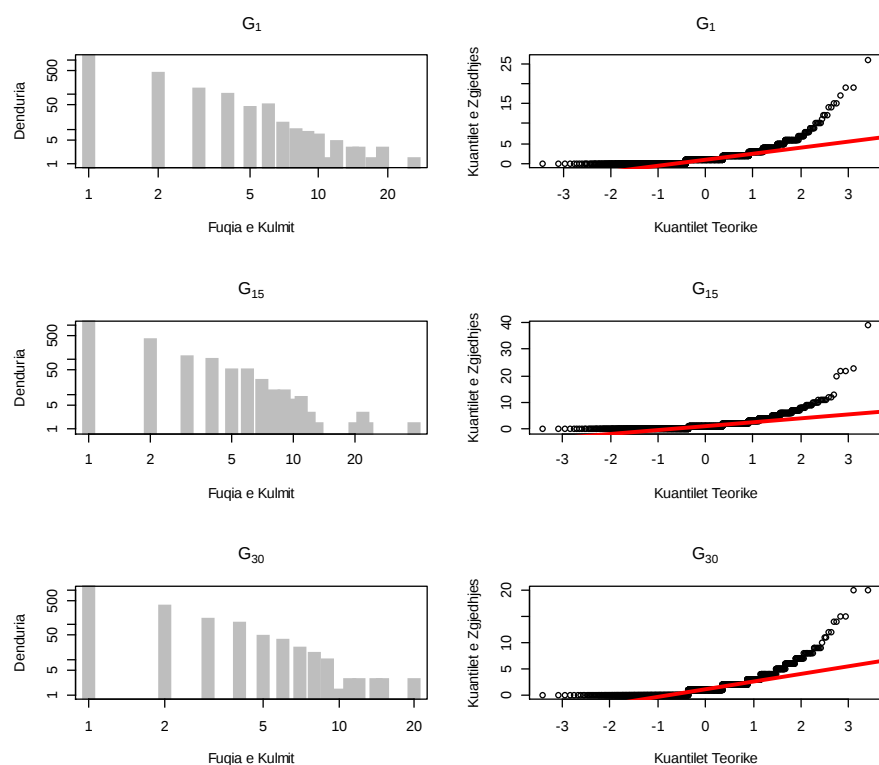
Figurë 4.1: Histogrami dhe grafiku normal kuantil-kuantil, për vlerat e fuqive të kulmeve për G_1 , G_{15} dhe G_{30} në rastin e Modelit 4.1.



Figurë 4.2: Histogrami dhe grafiku normal kuantil-kuantil, për vlerat e fuqive të kulmeve për G_1 , G_{15} dhe G_{30} në rastin e Modelit 4.2-Totale.



Figurë 4.3: Histogrami dhe grafiku normal kuantil-kuantil, për vlerat e fuqive të kulmeve për G_1, G_{15} dhe G_{30} në rastin e Modelit 4.2-Hyrëse.



Figurë 4.4: Histogrami dhe grafiku normal kuantil-kuantil, për vlerat e fuqive të kulmeve për G_1, G_{15} dhe G_{30} në rastin e Modelit 4.2-Dalëse.

Tabela 4.2 jep vlerat $\hat{f}_x^l$, probabilitetet empirike ditore që një kulm i zgjedhur rastësisht në graf rrjetat e serive të ketë fuqinë të barabartë me x . Sikurse vihet re në tabelë, vlerat e $\hat{f}_x^l$ gjatë 30 ditëve nuk kanë ndryshim statistikisht domethënës midis

tyre. Për shembull, në rastin e Modelit 4.1-Totale vlerat e $\hat{f}_1^i$ janë shumë pranë njëra-tjetrës për $i = 1, \dots, 30$. E njëjta gjë ndodh edhe në modelet e tjera, për të gjitha vlerat e x të cilat ne kemi zgjedhur të paraqesim në tabelë. Në rastin e Modelit 4.1-Totale dhe Modelit 4.2-Totale vlerat e zgjedhura për x (1, 2, 3) përmbledhin mesatarisht shpërndarjen empirike ditore respektivisht në 75% dhe 72%. Për Modelin 4.2-Hyrëse dhe Modelin 4.2-Dalëse vlerat e zgjedhura të x (0, 1, 2, 3) përmbledhin mesatarisht shpërndarjen empirike ditore respektivisht në vlerat 89.8% dhe 87.5%.

Tabelë 4.2: Probabilitetet empirike.

i	Modeli 4.1-Totale			Modeli 4.2-Totale			Modeli 4.2-Hyrëse				Modeli 4.2-Dalëse			
	$\hat{f}_1^i$	$\hat{f}_2^i$	$\hat{f}_3^i$	$\hat{f}_1^i$	$\hat{f}_2^i$	$\hat{f}_3^i$	$\hat{f}_0^i$	$\hat{f}_1^i$	$\hat{f}_2^i$	$\hat{f}_3^i$	$\hat{f}_0^i$	$\hat{f}_1^i$	$\hat{f}_2^i$	$\hat{f}_3^i$
1	0.41	0.22	0.13	0.38	0.23	0.13	0.30	0.37	0.16	0.07	0.34	0.30	0.18	0.07
2	0.40	0.23	0.14	0.36	0.23	0.14	0.30	0.38	0.16	0.07	0.36	0.27	0.18	0.08
3	0.40	0.21	0.14	0.36	0.21	0.13	0.27	0.37	0.17	0.08	0.35	0.28	0.17	0.07
4	0.40	0.19	0.14	0.37	0.20	0.14	0.28	0.36	0.17	0.08	0.36	0.26	0.17	0.07
5	0.40	0.21	0.13	0.36	0.22	0.12	0.29	0.38	0.16	0.07	0.35	0.27	0.19	0.06
6	0.38	0.25	0.12	0.36	0.24	0.13	0.27	0.37	0.17	0.08	0.37	0.28	0.16	0.07
7	0.38	0.22	0.13	0.35	0.23	0.12	0.28	0.36	0.17	0.07	0.35	0.27	0.18	0.06
8	0.38	0.22	0.13	0.34	0.22	0.13	0.27	0.35	0.19	0.08	0.35	0.26	0.17	0.07
9	0.40	0.24	0.13	0.36	0.24	0.13	0.31	0.37	0.16	0.07	0.34	0.30	0.17	0.07
10	0.40	0.23	0.13	0.37	0.23	0.12	0.29	0.37	0.17	0.07	0.36	0.29	0.17	0.06
11	0.40	0.22	0.14	0.36	0.23	0.13	0.28	0.39	0.16	0.07	0.34	0.30	0.17	0.07
12	0.41	0.21	0.14	0.38	0.21	0.12	0.28	0.37	0.17	0.07	0.36	0.28	0.16	0.07
13	0.39	0.24	0.13	0.36	0.24	0.13	0.29	0.39	0.16	0.07	0.33	0.29	0.19	0.07
14	0.39	0.21	0.13	0.36	0.21	0.13	0.29	0.36	0.17	0.08	0.35	0.27	0.18	0.08
15	0.41	0.23	0.13	0.37	0.23	0.13	0.28	0.37	0.17	0.07	0.36	0.28	0.17	0.06
16	0.41	0.21	0.12	0.38	0.22	0.13	0.30	0.37	0.17	0.07	0.36	0.27	0.18	0.07
17	0.40	0.21	0.13	0.38	0.21	0.13	0.29	0.36	0.17	0.08	0.35	0.28	0.17	0.07
18	0.40	0.21	0.14	0.36	0.22	0.13	0.30	0.35	0.17	0.07	0.34	0.28	0.18	0.07
19	0.40	0.23	0.13	0.37	0.23	0.13	0.29	0.37	0.17	0.07	0.35	0.29	0.18	0.06
20	0.38	0.24	0.13	0.35	0.23	0.13	0.29	0.37	0.16	0.09	0.33	0.30	0.19	0.07
21	0.38	0.23	0.12	0.35	0.23	0.12	0.27	0.38	0.17	0.09	0.34	0.27	0.20	0.07
22	0.41	0.21	0.12	0.37	0.21	0.12	0.29	0.36	0.16	0.08	0.36	0.27	0.18	0.06
23	0.42	0.21	0.12	0.39	0.21	0.12	0.28	0.38	0.16	0.07	0.37	0.28	0.16	0.07
24	0.40	0.23	0.13	0.37	0.23	0.13	0.29	0.36	0.17	0.09	0.35	0.30	0.17	0.06
25	0.40	0.23	0.12	0.37	0.23	0.11	0.29	0.37	0.15	0.08	0.36	0.29	0.17	0.06
26	0.41	0.22	0.12	0.37	0.22	0.12	0.30	0.37	0.15	0.07	0.36	0.27	0.17	0.07
27	0.42	0.21	0.15	0.38	0.21	0.14	0.29	0.38	0.15	0.08	0.36	0.27	0.19	0.06
28	0.39	0.24	0.13	0.35	0.23	0.15	0.27	0.37	0.17	0.09	0.37	0.27	0.17	0.06
29	0.40	0.22	0.14	0.37	0.21	0.14	0.30	0.35	0.17	0.07	0.35	0.29	0.17	0.07
30	0.40	0.22	0.13	0.37	0.22	0.13	0.30	0.38	0.15	0.07	0.36	0.28	0.17	0.07

Bazuar në vlerësimet empirike të mundësive, mund të themi që mesatarisht 13% e klientëve komunikojnë gjatë ditës me tre klientë të ndryshëm dhe e njëjta përqindje e klientëve janë të angazhuar në tre thirrje (hyrëse ose dalëse) telefonike në ditë.

Mesatarisht 7.56% e klientëve marrin tre thirrje hyrëse në ditë dhe 6.73% e klientëve kryejnë tre thirrje dalëse në ditë.

4.3.1.2. Përqasjet ndaj Shpërndarjeve Empirike Ditore të Fuqive të Kulmeve

Në këtë pjesë jepen rezultatet e vlerësimeve të bëra për parametrat e panjohur të shpërndarjeve dhe për x_{min} përgjatë përqasjeve ndaj shpërndarjeve empirike ditore të fuqive të kulmeve. Tabelat në vijim (Tabela 4.3, Tabela 4.4, Tabela 4.5 dhe Tabela 4.6) japin rezultatet për vlerësimet e bëra në Modelin 4.1 dhe Modelin 4.2 në lidhje me: x_{min} , α në rastin e shpërndarjes PL; dhe x_{min} , μ , σ për shpërndarjen LN.

Tabelë 4.3: Vlerësime në lidhje me parametrat dhe mirësinë e përqasjes (KS) për shpërndarjet PL dhe LN, në Modelin 4.1.

$G_i(n)$	Modeli 4.1-Totale						
	PL			LN			
	x_{min} (n_{fund})	α	$KS(p)$	x_{min} (n_{fund})	μ	σ	$KS(p)$
1 (1597)	5 (245)	3.27	0.02 (0.665)	1 (1597)	0.48	0.93	0.01 (0.196)
2 (1428)	6 (153)	3.95	0.04 (0.274)	1 (1428)	0.52	0.88	0.01 (0.016)
3 (1561)	5 (276)	3.07	0.03 (0.195)	3 (619)	0.38	0.99	0.01 (0.793)
4 (1534)	8 (108)	3.78	0.03 (0.813)	2 (906)	0.83	0.84	0.01 (0.239)
5 (1522)	7 (117)	3.64	0.03 (0.459)	2 (913)	0.79	0.81	0.01 (0.426)
6 (1530)	7 (132)	3.49	0.03 (0.509)	1 (1530)	0.54	0.93	0.01 (0.184)
7 (1531)	9 (74)	3.66	0.02 (0.895)	4 (400)	0.12	1.03	0.01 (0.703)
8 (1560)	6 (199)	3.37	0.03 (0.388)	1 (1560)	0.56	0.95	0.01 (0.251)
9 (1487)	7 (110)	3.64	0.03 (0.801)	1 (1487)	0.52	0.90	0.00 (0.948)
10 (1564)	9 (76)	3.69	0.03 (0.846)	1 (1564)	0.50	0.95	0.00 (0.876)
11 (1545)	7 (131)	3.67	0.03 (0.819)	2 (932)	0.71	0.86	0.01 (0.424)
12 (1531)	8 (97)	3.46	0.04 (0.274)	1 (1531)	0.49	0.96	0.01 (0.037)
13 (1547)	7 (121)	3.58	0.04 (0.359)	1 (1547)	0.53	0.92	0.01 (0.653)
14 (1552)	7 (133)	3.69	0.02 (0.988)	2 (949)	0.84	0.81	0.01 (0.247)
15 (1555)	9 (79)	3.88	0.04 (0.569)	1 (1555)	0.47	0.97	0.00 (0.969)
16 (1554)	8 (90)	4.11	0.02 (0.992)	2 (910)	0.78	0.81	0.01 (0.556)
17 (1555)	6 (195)	3.43	0.03 (0.483)	2 (926)	0.72	0.88	0.01 (0.413)
18 (1560)	7 (126)	3.42	0.03 (0.653)	4 (397)	-0.52	1.16	0.01 (0.895)
19 (1494)	5 (235)	3.24	0.03 (0.204)	1 (1494)	0.51	0.92	0.00 (0.898)
20 (1568)	5 (257)	3.31	0.04 (0.062)	1 (1568)	0.56	0.90	0.01 (0.162)
21 (1523)	7 (137)	3.61	0.04 (0.192)	1 (1523)	0.57	0.93	0.02 (0.009)
22 (1505)	7 (130)	3.49	0.03 (0.489)	2 (894)	0.74	0.87	0.01 (0.464)
23 (1479)	7 (117)	3.87	0.04 (0.377)	2 (859)	0.81	0.80	0.01 (0.223)
24 (1512)	5 (247)	3.14	0.02 (0.682)	1 (1512)	0.5	0.95	0.01 (0.622)
25 (1559)	5 (260)	3.15	0.03 (0.204)	1 (1559)	0.48	0.96	0.01 (0.379)
26 (1545)	10 (60)	3.78	0.03 (0.702)	1 (1545)	0.47	0.99	0.01 (0.467)
27 (1530)	6 (163)	3.35	0.04 (0.130)	3 (577)	0.10	1.03	0.01 (0.871)
28 (1505)	6 (176)	3.18	0.04 (0.253)	1 (1505)	0.52	0.95	0.01 (0.590)
29 (1524)	7 (127)	3.49	0.03 (0.528)	2 (917)	0.66	0.89	0.01 (0.740)
30 (1472)	6 (151)	3.58	0.02 (0.819)	2 (882)	0.74	0.80	0.01 (0.488)

Gjithashtu, jepet informacion në lidhje me numrin e përgjithshëm të të dhënave n në $\{d_v^i\}_{v \in V_i}$, dhe numrin e të dhënave të cilat gjenden në pjesën fundore ($x \geq x_{min}$) të shpërndarjes empirike, n_{fund} . Për secilën tabelë bëhet një përmbledhje e rasteve të refuzimeve ($p \leq 0.10$) për shpërndarjet e përqasura, e cila bazohet në rezultatet e arritura për p -vlerën dhe besueshmërinë e saj.

Në Modelin 4.1-Totale, shpërndarja PL u refuzua vetëm një herë në tridhjetë raste (konkretisht në G_{20}) dhe në gjashtë raste të tjera, p -vlера e njehsuar nuk ishte e besueshme (konkretisht në $G_7, G_{10}, G_{12}, G_{15}, G_{16}, G_{26}$). Shpërndarja LN u refuzua në tre raste nga tridhjetë në total (konkretisht në G_2, G_{12}, G_{21}). Në të gjitha rastet, p -vlера ishte gjithmonë e besueshme.

Tabelë 4.4: Vlerësime në lidhje me parametrat dhe mirësinë e përqasjes (KS) për shpërndarjet PL dhe LN, në Modelin 4.2-Totale.

$G_i(n)$	Modeli 4.2-Totale						
	PL			LN			
	x_{min} (n_{fund})	α	$KS(p)$	x_{min} (n_{fund})	μ	σ	$KS(p)$
1 (1597)	5 (286)	3.08	0.02 (0.762)	1 (1597)	0.57	0.95	0.01 (0.364)
2 (1428)	8 (100)	4.09	0.02 (0.92)	1 (1428)	0.62	0.91	0.01 (0.561)
3 (1561)	9 (110)	3.43	0.04 (0.58)	1 (1561)	0.6	1.0	0.01 (0.577)
4 (1534)	8 (134)	3.41	0.03 (0.722)	2 (968)	0.80	0.91	0.01 (0.427)
5 (1522)	7 (155)	3.46	0.04 (0.314)	2 (967)	0.78	0.88	0.02 (0.035)
6 (1530)	8 (126)	3.39	0.03 (0.522)	1 (1530)	0.59	0.98	0.01 (0.698)
7 (1531)	5 (323)	3.01	0.03 (0.405)	1 (1531)	0.64	0.97	0.01 (0.516)
8 (1560)	7 (194)	3.31	0.04 (0.231)	1 (1560)	0.66	0.99	0.01 (0.196)
9 (1487)	7 (142)	3.41	0.03 (0.63)	1 (1487)	0.60	0.94	0.00 (0.965)
10 (1564)	6 (224)	3.1	0.04 (0.251)	1 (1564)	0.57	0.99	0.00 (0.961)
11 (1545)	7 (172)	3.45	0.02 (0.88)	1 (1545)	0.61	0.97	0.01 (0.638)
12 (1531)	5 (301)	2.99	0.03 (0.105)	1 (1531)	0.55	1.01	0.02 (0.003)
13 (1547)	8 (116)	3.49	0.03 (0.754)	1 (1547)	0.61	0.94	0.00 (0.983)
14 (1552)	5 (324)	3.07	0.04 (0.035)	2 (999)	0.81	0.89	0.01 (0.352)
15 (1555)	10 (81)	3.66	0.04 (0.709)	1 (1555)	0.56	0.99	0.01 (0.747)
16 (1554)	8 (114)	3.63	0.03 (0.851)	1 (1554)	0.57	0.96	0.01 (0.07)
17 (1555)	6 (238)	3.19	0.04 (0.238)	1 (1555)	0.56	1.01	0.01 (0.295)
18 (1560)	6 (221)	3.1	0.02 (0.599)	1 (1560)	0.61	0.98	0.01 (0.14)
19 (1494)	5 (290)	3.06	0.03 (0.197)	1 (1494)	0.59	0.96	0.01 (0.64)
20 (1568)	6 (217)	3.21	0.03 (0.495)	1 (1568)	0.63	0.94	0.01 (0.337)
21 (1523)	7 (170)	3.32	0.03 (0.463)	4 (459)	0.11	1.09	0.01 (0.432)
22 (1505)	6 (228)	3.19	0.03 (0.349)	2 (944)	0.75	0.93	0.01 (0.158)
23 (1479)	6 (206)	3.4	0.03 (0.502)	3 (588)	0.98	0.80	0.02 (0.084)
24 (1512)	6 (210)	3.09	0.02 (0.712)	1 (1512)	0.58	0.98	0.01 (0.475)
25 (1559)	5 (314)	3.02	0.03 (0.503)	1 (1559)	0.56	1.00	0.01 (0.147)
26 (1545)	6 (222)	3.03	0.03 (0.246)	1 (1545)	0.56	1.01	0.01 (0.805)
27 (1530)	10 (70)	3.41	0.05 (0.184)	1 (1530)	0.56	0.99	0.01 (0.37)
28 (1505)	6 (213)	3.01	0.04 (0.292)	2 (974)	0.42	1.04	0.01 (0.245)
29 (1524)	9 (99)	3.56	0.05 (0.233)	2 (957)	0.70	0.94	0.01 (0.786)
30 (1472)	6 (196)	3.43	0.03 (0.633)	1 (1472)	0.60	0.94	0.01 (0.063)

Në Modelin 4.2-Totale, shpërndarja PL u refuzua vetëm një herë në tridhjetë raste (konkretisht në G_{14}) dhe në tre raste të tjera, p -vlera e njehsuar nuk ishte e besueshme (konkretisht në G_{15}, G_{27}, G_{29}). Shpërndarja LN u refuzua në pesë raste nga tridhjetë në total (konkretisht në $G_5, G_{12}, G_{16}, G_{23}, G_{30}$). Në të gjitha rastet, p -vlera ishte gjithmonë e besueshme.

Në Modelin 4.2-Hyrëse dhe Modelin 4.2-Dalëse përjasjet janë shqyrtuar vetëm në atë pjesë të të dhënave ku $d_v^{i-} \geq 1$ dhe $d_v^{i+} \geq 1$, për arsye se shpërndarja PL nuk është e përcaktuar për vlera jopozitive të ndryshores. Në Modelin 4.2-Hyrëse, shpërndarja PL u refuzua shtatë herë në tridhjetë raste (konkretisht në $G_4, G_8, G_{12}, G_{13}, G_{14}, G_{21}, G_{24}$) dhe në pesë raste të tjera, p -vlera e njehsuar nuk ishte e besueshme (konkretisht në $G_3, G_5, G_{17}, G_{19}, G_{27}$). Shpërndarja LN u refuzua në tre raste nga tridhjetë në total (konkretisht në G_{17}, G_{20}, G_{21}). Në të gjitha rastet, p -vlera ishte gjithmonë e besueshme.

Tabelë 4.5: Vlerësime në lidhje me parametrat dhe mirësinë e përjasjes (KS) për shpërndarjet PL dhe LN, në Modelin 4.2-Hyrëse.

$G_i(n)$	Modeli 4.2-Hyrëse							
	PL				LN			
	x_{min} (n_{fund})	α	$KS(p)$		x_{min} (n_{fund})	μ	σ	$KS(p)$
1 (1111)	4 (148)	3.1	0.03 (0.515)		1 (1111)	0.05	0.97	0.01 (0.617)
2 (1006)	4 (137)	3.37	0.03 (0.54)		1 (1006)	0.09	0.94	0.01 (0.449)
3 (1143)	7 (40)	3.08	0.04 (0.313)		1 (1143)	0.17	0.94	0.01 (0.44)
4 (1107)	3 (287)	2.96	0.03 (0.097)		1 (1107)	0.21	0.92	0.01 (0.52)
5 (1086)	5 (99)	3.38	0.02 (0.832)		1 (1086)	0.06	0.99	0.01 (0.797)
6 (1121)	3 (273)	2.99	0.03 (0.175)		1 (1121)	0.17	0.91	0.01 (0.382)
7 (1101)	4 (172)	3.27	0.03 (0.6)		1 (1101)	0.21	0.93	0.01 (0.169)
8 (1135)	2 (586)	2.64	0.04 (0.001)		2 (586)	-1.49	1.39	0.01 (0.408)
9 (1028)	3 (248)	2.9	0.04 (0.142)		1 (1028)	0.04	0.99	0.01 (0.3)
10 (1113)	4 (163)	3.18	0.03 (0.522)		1 (1113)	0.11	0.96	0.01 (0.324)
11 (1107)	4 (165)	3.12	0.01 (0.992)		1 (1107)	-0.02	1.03	0.01 (0.414)
12 (1099)	3 (269)	2.9	0.04 (0.035)		1 (1099)	0.12	0.95	0.01 (0.514)
13 (1092)	3 (253)	2.81	0.04 (0.022)		1 (1092)	-0.07	1.04	0.01 (0.609)
14 (1108)	3 (287)	2.94	0.04 (0.087)		1 (1108)	0.16	0.95	0.01 (0.15)
15 (1115)	3 (270)	2.92	0.02 (0.483)		2 (538)	-0.65	1.17	0.01 (0.736)
16 (1092)	5 (101)	3.52	0.03 (0.601)		1 (1092)	0.11	0.95	0.01 (0.253)
17 (1105)	7 (44)	3.26	0.04 (0.503)		1 (1105)	0.18	0.93	0.01 (0.074)
18 (1097)	4 (167)	3.1	0.02 (0.836)		1 (1097)	0.15	0.96	0.01 (0.16)
19 (1061)	5 (92)	3.33	0.02 (0.844)		1 (1061)	0.10	0.95	0.01 (0.489)
20 (1118)	3 (287)	3.04	0.03 (0.294)		1 (1118)	0.13	0.95	0.02 (0.008)
21 (1115)	3 (281)	3	0.04 (0.073)		1 (1115)	0.14	0.95	0.02 (0.008)
22 (1072)	5 (102)	3.39	0.03 (0.732)		1 (1072)	0.16	0.95	0.01 (0.653)
23 (1058)	4 (144)	3.33	0.03 (0.617)		1 (1058)	0.10	0.94	0.01 (0.743)
24 (1078)	3 (283)	2.99	0.03 (0.089)		1 (1078)	0.21	0.92	0.01 (0.126)
25 (1113)	5 (100)	3.42	0.03 (0.657)		1 (1113)	0.13	0.96	0.01 (0.236)
26 (1082)	4 (168)	3	0.04 (0.205)		1 (1082)	0.05	1.03	0 (0.986)
27 (1085)	6 (74)	3.69	0.03 (0.762)		1 (1085)	0.04	1.00	0.01 (0.596)
28 (1094)	3 (278)	3.03	0.03 (0.106)		1 (1094)	0.19	0.91	0.01 (0.308)
29 (1065)	4 (161)	3.11	0.03 (0.55)		2 (534)	-0.74	1.21	0.01 (0.532)
30 (1026)	4 (143)	3.15	0.03 (0.476)		1 (1026)	0.03	0.99	0.01 (0.472)

Në Modelin 4.2-Dalëse, shpërndarja PL u refuzua shtatë herë në tridhjetë raste (konkretisht në $G_1, G_4, G_6, G_7, G_{12}, G_{17}, G_{25}$) dhe në pesëmbëdhjetë raste të tjera, p -vlera e njehsuar nuk ishte e besueshme (konkretisht në $G_2, G_3, G_5, G_9, G_{10}, G_{11}, G_{13}, G_{14}, G_{15}, G_{19}, G_{22}, G_{23}, G_{27}, G_{29}, G_{30}$). Shpërndarja LN u refuzua në katërmëdhjetë raste nga tridhjetë në total (konkretisht në $G_1, G_6, G_7, G_8, G_{10}, G_{13}, G_{14}, G_{16}, G_{21}, G_{22}, G_{25}, G_{27}, G_{28}, G_{29}$) dhe p -vlera nuk ishte e besueshme në pesë raste (konkretisht në $G_5, G_{18}, G_{19}, G_{20}, G_{26}$).

Tabelë 4.6: Vlerësime në lidhje me parametrat dhe mirësinë e përqasjes (KS) për shpërndarjet PL dhe LN, në Modelin 4.2-Dalëse.

$G_i(n)$	Modeli 4.2-Dalëse						
	PL			LN			
	x_{min} (n_{fund})	α	$KS(p)$	x_{min} (n_{fund})	μ	σ	$KS(p)$
1 (1054)	2 (570)	2.62	0.05 (0)	1 (1054)	0.35	0.86	0.02 (0.026)
2 (921)	6 (66)	4.85	0.02 (0.927)	1 (921)	0.48	0.78	0.01 (0.411)
3 (1009)	6 (98)	3.44	0.04 (0.588)	2 (576)	-0.04	1.08	0.02 (0.105)
4 (979)	2 (575)	2.48	0.05 (0)	2 (575)	0.03	1.04	0.02 (0.164)
5 (989)	6 (78)	3.99	0.03 (0.713)	4 (196)	-0.08	0.95	0.02 (0.29)
6 (969)	2 (545)	2.49	0.05 (0)	1 (969)	0.38	0.92	0.01 (0.062)
7 (995)	2 (574)	2.51	0.05 (0)	2 (574)	-0.40	1.17	0.02 (0.024)
8 (1011)	4 (225)	3.18	0.03 (0.368)	1 (1011)	0.49	0.88	0.01 (0.085)
9 (974)	6 (71)	3.6	0.04 (0.733)	1 (974)	0.35	0.88	0.01 (0.148)
10 (1008)	8 (53)	4.18	0.04 (0.776)	2 (557)	-0.28	1.14	0.02 (0.039)
11 (1024)	6 (87)	3.97	0.04 (0.627)	1 (1024)	0.38	0.89	0.01 (0.56)
12 (976)	4 (190)	3.06	0.05 (0.065)	1 (976)	0.38	0.92	0.01 (0.143)
13 (1034)	7 (52)	3.94	0.04 (0.466)	2 (581)	-0.22	1.04	0.02 (0.05)
14 (1010)	6 (95)	3.79	0.03 (0.798)	1 (1010)	0.44	0.87	0.02 (0.008)
15 (991)	6 (92)	3.92	0.03 (0.804)	3 (289)	1.09	0.69	0.02 (0.13)
16 (1000)	4(192)	3.5	0.03 (0.339)	1 (1000)	0.45	0.82	0.02 (0.029)
17 (1009)	2 (575)	2.51	0.06 (0)	1 (1009)	0.42	0.88	0.01 (0.225)
18 (1022)	4 (197)	3.19	0.02 (0.663)	4 (197)	-1.97	1.37	0.01 (0.587)
19 (970)	9 (30)	4.87	0.05 (0.41)	5 (111)	1.41	0.59	0.02 (0.546)
20 (1052)	4 (182)	3.2	0.02 (0.841)	4 (182)	-3.26	1.57	0.02 (0.678)
21 (1004)	4 (189)	3.17	0.03 (0.552)	2 (599)	-0.60	1.17	0.02 (0.01)
22 (959)	6 (86)	3.75	0.03 (0.857)	1 (959)	0.46	0.87	0.02 (0.002)
23 (935)	6 (79)	4.06	0.05 (0.199)	1 (935)	0.33	0.88	0.01 (0.126)
24 (983)	4 (178)	2.99	0.03 (0.41)	2 (535)	-0.67	1.24	0.02 (0.13)
25 (1001)	4 (198)	3.1	0.06 (0.006)	1 (1001)	0.32	0.95	0.02 (0.006)
26 (993)	4 (198)	3.13	0.03 (0.336)	4 (198)	-2.47	1.48	0.02 (0.423)
27 (984)	6 (79)	3.46	0.04 (0.692)	2 (564)	-0.72	1.22	0.03 (0.001)
28 (951)	4 (191)	2.97	0.04 (0.291)	2 (540)	-0.44	1.20	0.03 (0.008)
29 (994)	6 (87)	3.93	0.03 (0.74)	1 (994)	0.41	0.88	0.01 (0.091)
30 (947)	6 (71)	4.09	0.03 (0.688)	1 (947)	0.43	0.84	0.01 (0.104)

Në Figurën 4.5 dhe 4.6 jepet një paraqitje grafike e funksionit komplementar të shpërndarjes ($\overline{F_X(x)} = P(X \geq x)$, ku $X = d_v$) në plane me boshte të logaritmuara, për të tria rastet, në lidhje me G_1, G_{15} dhe G_{30} . Në Rastin 1, vihet re se PL bën një përqasje shumë të dobët kur është përqasur prej $x_{min} = 1$, ndërkohë në Rastin 2 vihet re se

hapësira e përqasjes së LN është më e madhe krahasuar me hapësirën që përqas PL, duke qënë se $x_{min}^{LN} < x_{min}^{PL}$. Ndërsa, në Rastin 3 vihet re se nuk ka ndonjë ndryshim statistikisht domethënës ndërmjet dy modeleve. Në këto figura janë dhënë gjithashtu dhe parametrat e shpërndarjeve të përqasura në trajtën $PL(\alpha, x_{min})$ dhe $LN(\mu, \sigma, x_{min})$.

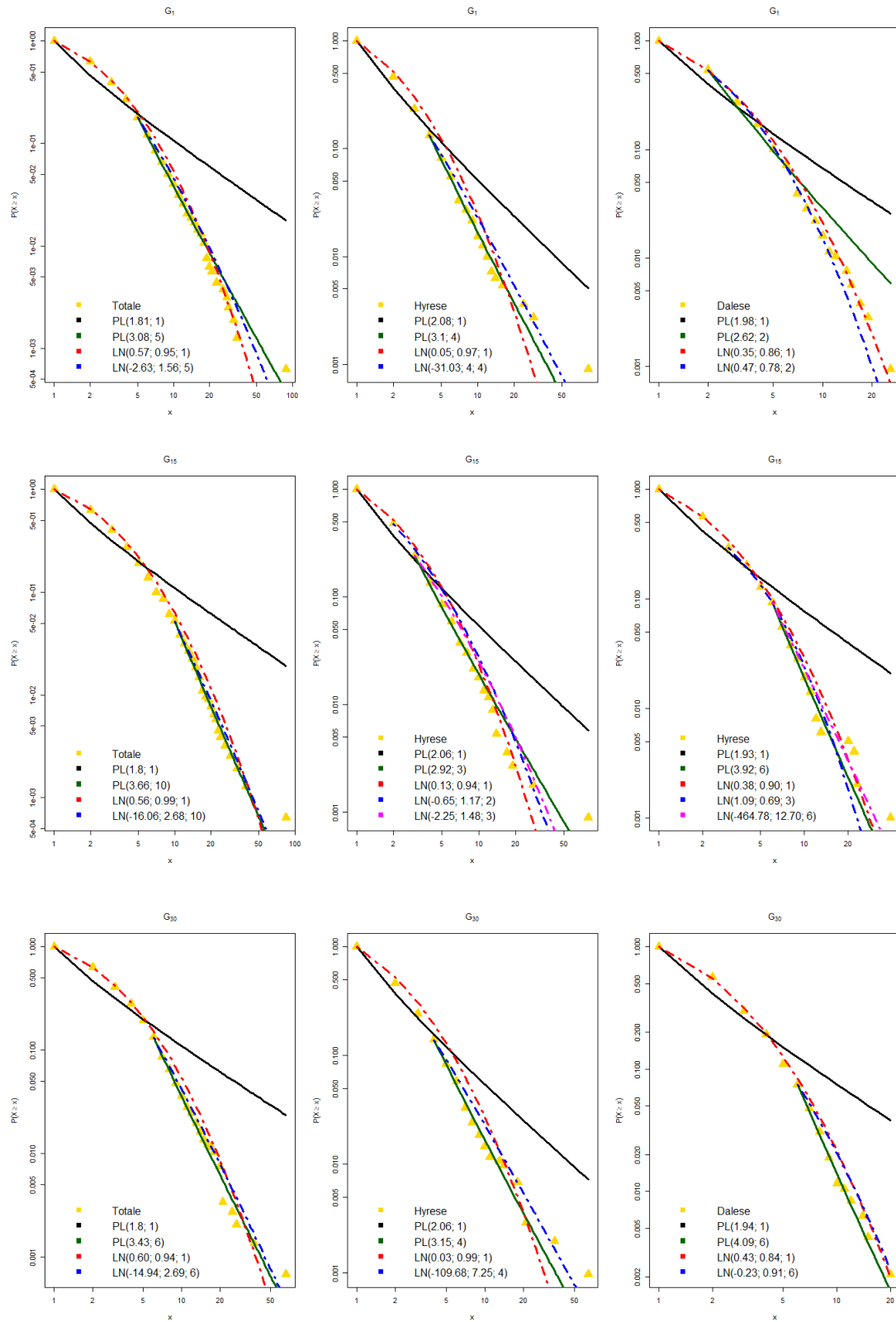
Në shpërndarjen PL, në të shumtën e herëve kulmet kanë fuqi të ulët, por ekziston mundësia që të gjenden kulme me vlera të larta të fuqive. Për më tepër, vlerat e ndërmjetme janë të gjitha të mundshme dhe fuqia mesatare nuk paraqet ndonjë vlerë të veçantë për shpërndarjen. Ndërkohë, në graf rrjetat e rastit shumica e kulmeve kanë afërsisht të njëjtën fuqi, të përafërt me fuqinë mesatare të kulmeve, situatë e cila nuk ndodh në graf rrjetat e telefonisë fikse.

Tabela 4.7 dhe Tabela 4.8 japin rezultatet në lidhje me krahasimin e bërë ndërmjet shpërndarjeve LN dhe PL, në rastin kur të dyja janë të mundshme për të modeluar të dhënat. Duke qënë se në të gjitha rastet kemi që $\max(x_{min}^{PL}, x_{min}^{LN}) = x_{min}^{PL}$, themi se LN është e kushtëzuar nga PL. Me shënimin “-” në tabelë janë paraqitur të gjitha ato raste ku nuk është bërë krahasimi midis LN dhe PL, duke qënë se nuk u vlerësuan si shpërndarje të mundshme njëherazi për atë rast, për të dhënat $x \geq \max(x_{min}^{PL}, x_{min}^{LN})$. Me përjashtim të një rasti (konkretisht në G_{18} në Modelin 4.2-Hyrëse, në të cilin favorizohet PL), shenja e raportit të log-përgjasisë, $\mathcal{R}$, nuk është e besueshme për të përcaktuar se cili nga modelet është më i mirë se tjetri, sepse $p > 0.1$. Kështu, në pothuajse 100% të rasteve të shqyrtuara, të dyja modelet konsiderohen njësoj të mundshme për të modeluar të dhënat në pjesën fundore të shpërndarjeve empirike, për $x \geq \max(x_{min}^{PL}, x_{min}^{LN})$.

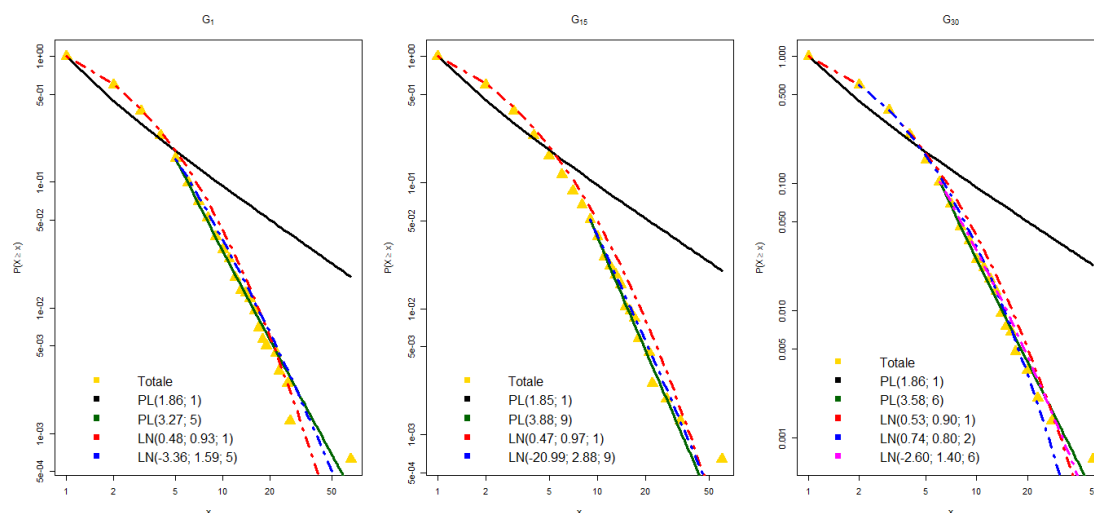
Vihet re se në modelet e trajtuara, vlerat e KS për LN janë gjithmonë më të vogla se vlerat e KS për PL, respektivisht në secilën prej G_i . Hapësira e të dhënave që përqas shpërndarja LN është gjithmonë më e madhe se ajo që përqas shpërndarja PL. Ky përfundim vihet re edhe nga Figura 4.7, e cila krahason grafikisht, nëpërmjet diagrameve kuti, vlerësimet e bëra për x_{min}^i për $i = 1, \dots, 30$ për të dyja shpërndarjet e përqasura. Ndonëse numri i refuzimeve të shpërndarjes PL është kryesisht më i ulët se ai i shpërndarjes LN, në shpërndarjen PL numri i p -vlerave jo të besueshme është kryesisht më i lartë. Nëse marrim totalin e refuzimeve dhe rasteve për p -vlera jo të besueshme, shpërndarja PL ka numrin më të lartë se sa LN. Gjithashtu, vihet re që në rastin e Modelit 4.2 Dalëse, të dyja shpërndarjet kanë një numër të lartë refuzimesh dhe p -vlerash jo të besueshme. Dalim në përfundimin se të dyja këto shpërndarje kanë mundësi të vogël përqasje për vargun e fuqive dalëse, por shpërndarja LN vlerësohet si më e mira në përqasjen e vargut të fuqive totale dhe fuqive hyrëse në graf rrjetat e telefonisë fikse.

Në [19], Seshadri *et al.* kanë treguar se në një graf rrjet të telefonisë lëvizëse shpërndarja e fuqive të kulmeve kishte një ndryshim statistikisht domethënës nga shpërndarja PL dhe ajo LN. Në lidhje me shpërndarjen e numrit të klientëve të ndryshëm për klient, ata raportuan se përqasja më e mirë e mundshme nëpërmjet shpërndarjes PL ishte për $\alpha = 3.5$, ndërsa për shpërndarjen LN kishte parametrat $\mu = 1.6389$ dhe $\sigma = 1.5454$. Në studimin e tyre ata propozojnë përdorimin e shpërndarjes së dyfishtë Pareto lognormale (DPLN) si më të mirën. Ata gjithashtu pohuan se në një interval ndarës 6 mujor ndërmjet dy graf rrjetave të paorientuara që trajtuan, vunë re se parametrat e DPLN nuk kishin ndryshim statistikisht domethënës. Aksu *et al.* [26] propozuan shpërndarjen Pareto lognormal (PLN) si shpërndarja teorike që përqaste më mirë shpërndajen empirike të fuqive të kulmeve. Ndonëse këto studime i takojnë graf

rrjetave të komunikimit nëpërmjet telefonisë lëvizëse, do të qe interesante të hulumtohej për përjasjen e mundshme të këtyre shpërndarjeve në të dhënat e telefonisë fikse. Kjo mund të shërbej si subjekt i një analize të mëtejshme.



Figurë 4.5: Paraqitje grafike e $\overline{F_X(x)}$ në rastin e Modelit 4.2 për G_1 , G_{15} dhe G_{30} në plane me boshte të logaritmuara.



Figurë 4.6: Paraqitje grafike e $\overline{F_X}(x)$ në rastin e Modelit 4.1 për G_1, G_{15} dhe G_{30} në plane me boshte të logaritmuara.

Tabelë 4.7: Krahasimi i LN kundrejt PL nëpërmjet testit Vuong të raportit të log-përgjasisë $\mathcal{R}$, në Modelin 4.1-Totale dhe Modelin 4.2-Totale.

G_i	Modeli 4.1-Totale					Modeli 4.2-Totale				
	LN e kushtëzuar nga PL					LN kundrejt PL				
	x_{min} (n_{fund})	μ	σ	$\mathcal{R}$	p	x_{min} (n_{fund})	μ	σ	$\mathcal{R}$	p
1	5 (245)	-3.36	1.59	0.516	0.606	5 (286)	-2.63	1.56	0.66	0.509
2	-	-	-	-	-	8 (100)	-355.89	10.78	0.182	0.855
3	5 (276)	-0.02	1.08	1.232	0.218	9 (110)	-3.56	1.64	0.268	0.789
4	8 (108)	-490.96	13.36	0.268	0.788	8 (134)	-239.62	10.05	-0.203	0.839
5	7 (117)	-48.44	4.40	0.087	0.931	-	-	-	-	-
6	7 (132)	-5.14	1.77	0.328	0.743	8 (126)	-7.03	2.03	0.261	0.794
7	9 (74)	-10.23	2.22	0.114	0.909	5 (323)	-1.34	1.37	0.881	0.379
8	6 (199)	-7.20	2.03	0.253	0.8	7 (194)	-22.38	3.30	0.105	0.916
9	7 (110)	-505.22	13.89	-0.007	0.994	7 (142)	-340.33	11.96	-0.259	0.795
10	9 (76)	-512.04	13.85	0.424	0.671	6 (224)	-1.42	1.38	0.645	0.519
11	7 (131)	-519.53	14.02	0.459	0.647	7 (172)	-564.30	15.24	0.347	0.728
12	-	-	-	-	-	-	-	-	-	-
13	7 (121)	-123.13	7.00	0.342	0.732	8 (116)	-358.11	12.06	0.105	0.916
14	7 (133)	-266.38	10.02	0.052	0.958	5 (324)	-0.91	1.27	0.886	0.376
15	9 (79)	-20.99	2.88	0.063	0.95	10 (81)	-16.06	2.68	0.076	0.94
16	8 (90)	-440.36	11.97	-0.032	0.975	-	-	-	-	-
17	6 (195)	-2.79	1.48	0.447	0.655	6 (238)	-0.99	1.27	0.704	0.481
18	7 (126)	-1.35	1.29	0.458	0.647	6 (221)	-4.12	1.79	0.446	0.656
19	5 (235)	-0.21	1.07	0.928	0.353	5 (290)	-0.39	1.16	1.015	0.31
20	-	-	-	-	-	6 (217)	-8.73	2.27	0.251	0.802
21	-	-	-	-	-	7 (170)	-8.01	2.15	0.208	0.835
22	7 (130)	-2.56	1.45	0.361	0.718	6 (228)	-1.60	1.38	0.62	0.535
23	7 (117)	-5.83	1.71	0.243	0.808	-	-	-	-	-
24	5 (247)	-6.89	2.09	0.399	0.69	6 (210)	-13.03	2.74	0.219	0.826
25	5 (260)	-0.66	1.19	0.855	0.393	5 (314)	-2.30	1.54	0.68	0.496
26	10 (60)	-490.04	13.32	0.081	0.936	6 (222)	-1.14	1.36	0.696	0.487
27	6 (163)	-3.12	1.55	0.0378	0.705	10 (70)	-291.21	11.06	-0.406	0.685
28	6 (176)	-0.26	1.14	0.772	0.44	6 (213)	-1.02	1.35	0.78	0.435
29	7 (127)	-6.08	1.88	0.185	0.85	9 (99)	-49.10	4.51	-0.036	0.971
30	5 (245)	-3.36	1.59	0.516	0.606	-	-	-	-	-

Tabelë 4.8: Krahاسimi i LN kundrejt PL nëpërmjet testit Vuong të raportit të log-përgjasisë $\mathcal{R}$, në Modelin 4.2-Hyrëse dhe Modelin 4.2-Dalëse.

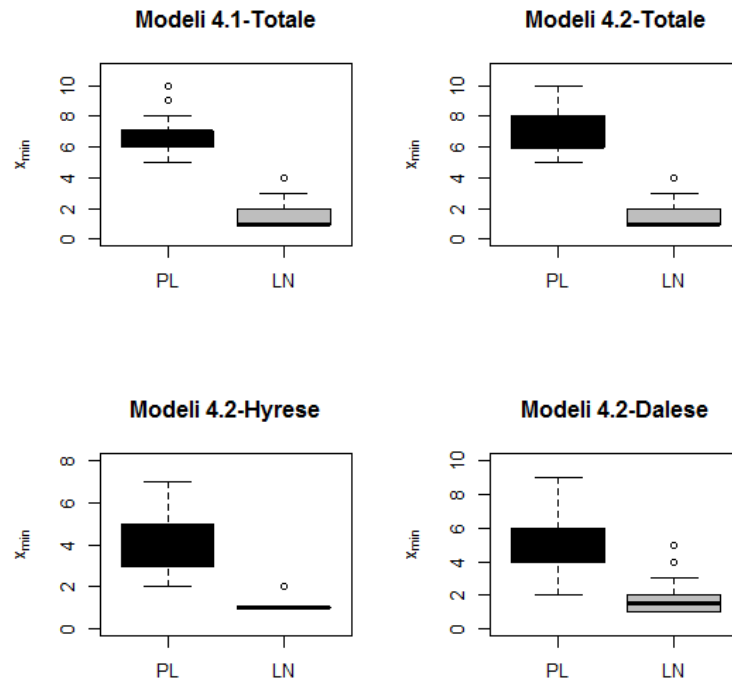
G_i	Modeli 4.2-Hyrëse					Modeli 4.2-Dalëse				
	LN e kushtëzuar nga PL					LN kundrejt PL				
	x_{min} (n_{fund})	μ	σ	$\mathcal{R}$	p	x_{min} (n_{fund})	μ	σ	$\mathcal{R}$	p
1	4 (148)	-31.03	4	-0.173	0.863	-	-	-	-	-
2	4 (137)	-571.37	15.64	-0.239	0.711	6 (66)	-355.79	9.66	0.198	0.843
3	7 (40)	-650.90	17.76	0.064	0.949	6 (98)	-0.27	1.06	0.718	0.473
4	-	-	-	-	-	-	-	-	-	-
5	5 (99)	-569.74	15.57	0.056	0.955	6 (78)	-442.28	12.24	0.437	0.662
6	3 (273)	-2.81	1.55	0.554	0.58	-	-	-	-	-
7	4 (172)	-528.47	15.39	-0.081	0.936	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-
9	3 (248)	-2.82	1.60	0.427	0.67	6 (71)	-5.03	1.70	0.199	0.842
10	4 (163)	-398.84	13.62	0.036	0.971	-	-	-	-	-
11	4 (165)	-642.32	17.51	0.16	0.873	6 (87)	-457.56	12.48	0.59	0.555
12	-	-	-	-	-	-	-	-	-	-
13	-	-	-	-	-	-	-	-	-	-
14	-	-	-	-	-	-	-	-	-	-
15	3 (270)	-2.25	1.48	0.611	0.541	6 (92)	-464.78	12.70	0.118	0.906
16	5 (101)	-553.52	14.92	0.492	0.623	-	-	-	-	-
17	-	-	-	-	-	-	-	-	-	-
18	4 (167)	-446.56	17.71	-2.492	0.013	4 (197)	-1.97	1.37	0.744	0.457
19	5 (92)	-594.53	16.06	0.414	0.679	9 (30)	-361.12	9.71	0.389	0.697
20	-	-	-	-	-	4 (182)	-3.26	1.57	0.48	0.631
21	-	-	-	-	-	-	-	-	-	-
22	5 (102)	-579.64	15.67	0.297	0.766	-	-	-	-	-
23	4 (144)	-200.37	9.39	-0.579	0.563	6 (79)	-453.85	12.27	0.017	0.986
24	-	-	-	-	-	4 (178)	-2.89	1.61	0.558	0.577
25	5 (100)	-568.33	15.42	0.02	0.984	-	-	-	-	-
26	4 (168)	-2.80	1.59	0.356	0.722	4 (198)	-2.47	1.48	0.653	0.514
27	6 (74)	-479.39	13.41	0.633	0.527	-	-	-	-	-
28	3 (278)	-5.17	1.88	0.235	0.814	-	-	-	-	-
29	4 (161)	-12.43	2.65	0.081	0.936	-	-	-	-	-
30	4 (143)	-109.68	7.25	-0.663	0.507	6 (71)	-0.23	0.91	0.537	0.591

Tabela 4.9 jep disa karakteristika statistikore numerike në lidhje me vlerësimet e bëra për parametrat e shpërndarjeve ditore LN në rastin e fuqive totale dhe hyrëse. Rasti 1 jepet me shënimin LN_1 ; Rasti 2 jepet nga $LN_{x_{min}}$. Në bazë të rezultateve të gjetura në bazë të p -vlerës së kriterit Shapiro-Wilk nuk mund të themi se shpërndarja e parametrave të vlerësuar të LN është normale në të gjitha rastet.

Për përqasjet e bëra për LN_1 të paktën 88.9% e vlerësimeve të bëra për parametrat e shpërndarjes gjenden në intervalet si vijon: në Modelin 4.1-Totale, $\mu \in 0.51 \pm 3 \cdot 0.03$ dhe $\sigma \in 0.94 \pm 3 \cdot 0.03$; në Modelin 4.2-Totale, $\mu \in 0.59 \pm 3 \cdot 0.03$ dhe $\sigma \in 0.97 \pm 3 \cdot 0.03$; në Modelin 4.2-Hyrëse, $\mu \in 0.12 \pm 3 \cdot 0.07$ dhe $\sigma \in 0.96 \pm 3 \cdot 0.03$.

Për përqasjet e bëra për $LN_{x_{min}}$ të paktën 88.9% e vlerësimeve të bëra për parametrat e shpërndarjes gjenden në intervalet si vijon: në Modelin 4.1-Totale, $\mu \in 0.53 \pm 3 \cdot 0.27$ dhe $\sigma \in 0.92 \pm 3 \cdot 0.08$; në Modelin 4.2-Totale, $\mu \in 0.61 \pm 3 \cdot 0.14$

dhe $\sigma \in 0.97 \pm 3 \cdot 0.05$; në Modelin 4.2-Hyrëse, $\mu \in 0.004 \pm 3 \cdot 0.36$ dhe $\sigma \in 0.99 \pm 3 \cdot 0.10$.



Figurë 4.7: Krahasoni i vlerësimeve për x_{min}^i në Modelet 4.1 dhe 4.2 në lidhje me shpërndarjet PL dhe LN.

Tabelë 4.9: Përshkrimi numerik i parametrave të vlerësuar për shpërndarjes LN në Rastin 1 dhe Rastin 2.

		Min.	Q ₁	Mesorja	Q ₃	Max.	Mesatarja	Dev. Std.	Y ₁	Y ₂	p ⁶³	
Modeli 4.1- Totale	μ	LN ₁	0.47	0.48	0.51	0.53	0.57	0.51	0.03	0.31	-1.09	0.045
		LN _{x_{min}}	-0.52	0.48	0.52	0.71	0.84	0.53	0.27	-2.04	5.67	2.2×10 ⁻⁰⁵
	σ	LN ₁	0.88	0.93	0.95	0.94	0.99	0.94	0.03	-0.28	-0.31	0.43
		LN _{x_{min}}	0.80	0.87	0.93	0.96	1.16	0.92	0.08	0.64	0.85	0.09
Modeli 4.2- Totale	μ	LN ₁	0.55	0.57	0.59	0.61	0.66	0.59	0.03	0.37	-0.90	1.24×10 ⁻⁰⁴
		LN _{x_{min}}	0.11	0.56	0.60	0.64	0.98	0.61	0.14	-0.72	4.21	1.24×10 ⁻⁰⁴
	σ	LN ₁	0.91	0.96	0.98	0.99	1.01	0.97	0.03	-0.49	-0.49	0.1322
		LN _{x_{min}}	0.80	0.94	0.97	0.99	1.09	0.97	0.05	-0.58	1.47	0.1533
Modeli 4.2- Hyërse	μ	LN ₁	-0.07	0.07	0.13	0.17	0.22	0.12	0.07	-0.66	-0.17	0.1652
		LN _{x_{min}}	-1.49	0.04	0.11	0.16	0.21	0.004	0.36	-2.95	8.50	1.051×10 ⁻⁰⁸
	σ	LN ₁	0.91	0.94	0.95	0.96	1.04	0.96	0.03	0.92	0.04	5.976×10 ⁻⁰³
		LN _{x_{min}}	0.91	0.94	0.95	0.99	1.39	0.99	0.10	2.50	6.27	3.084×10 ⁻⁰⁷

⁶³ p-vlera sipas kriterit Shapiro-Wilk.

4.3.2. Vlerësime për Korrelacionin Ditor Ndërmjet Fuqive Hyrëse dhe Dalëse për Klient

Tabela 4.10 tregon vlerat për koeficientin e korrelacionit sipas Kendall dhe Spearman të njehsuar ndërmjet fuqive hyrëse dhe dalëse në secilin prej graf rrjetave në Modelin 4.2.

Tabelë 4.10: Analiza e korrelacionit ndërmjet fuqive hyrëse dhe dalëse ($d_v^- \sim d_v^+$), në Modelin 4.2. Vlerat e koeficientëve janë rumbullakosur në të mijtat.

G_i	$\hat{\tau}$ (p -vlera)	$\hat{\rho}$ (p -vlera)
1	-0.130 (2.867×10^{-10})	-0.147 (3.792×10^{-09})
2	-0.145 (3.167×10^{-11})	-0.165 (3.136×10^{-10})
3	-0.062 (2.815×10^{-03})	-0.062 (1.394×10^{-02})
4	-0.079 (1.425×10^{-04})	-0.086 (7.154×10^{-04})
5	-0.112 (9.872×10^{-08})	-0.128 (5.711×10^{-07})
6	-0.093 (9.052×10^{-06})	-0.099 (1.052×10^{-04})
7	-0.096 (3.405×10^{-06})	-0.108 (2.219×10^{-05})
8	-0.073 (3.622×10^{-04})	-0.077 (2.236×10^{-03})
9	-0.172 (8.889×10^{-16})	-0.200 (6.246×10^{-15})
10	-0.119 (9.243×10^{-09})	-0.135 (8.007×10^{-08})
11	-0.105 (5.392×10^{-07})	-0.115 (5.902×10^{-06})
12	-0.098 (2.437×10^{-06})	-0.108 (2.307×10^{-05})
13	-0.119 (1.175×10^{-08})	-0.136 (8.284×10^{-08})
14	-0.117 (1.738×10^{-08})	-0.135 (1.017×10^{-07})
15	-0.104 (5.733×10^{-07})	-0.114 (6.99×10^{-06})
16	-0.122 (4.987×10^{-09})	-0.136 (7.78×10^{-08})
17	-0.066 (1.466×10^{-03})	-0.065 (9.815×10^{-03})
18	-0.107 (2.363×10^{-07})	-0.120 (2.065×10^{-06})
19	-0.122 (1.126×10^{-08})	-0.138 (7.738×10^{-08})
20	-0.090 (1.481×10^{-05})	-0.099 (9.08×10^{-05})
21	-0.053 (1.208×10^{-02})	-0.051 (4.799×10^{-02})
22	-0.097 (4.22×10^{-06})	-0.105 (4.241×10^{-05})
23	-0.123 (8.835×10^{-09})	-0.139 (7.953×10^{-08})
24	-0.098 (3.093×10^{-06})	-0.108 (2.715×10^{-05})
25	-0.108 (2.225×10^{-07})	-0.119 (2.679×10^{-06})
26	-0.120 (8.911×10^{-09})	-0.134 (1.315×10^{-07})
27	-0.114 (6.69×10^{-08})	-0.127 (6.031×10^{-07})
28	-0.116 (4.121×10^{-08})	-0.128 (5.692×10^{-07})
29	-0.098 (3.015×10^{-06})	-0.107 (2.72×10^{-05})
30	-0.173 (8.808×10^{-16})	-0.201 (6.057×10^{-15})

Vihet re se në të gjitha rastet, p -vlera e gjetur është më e vogël se 0.01. Kjo nënkupton që vlerat e njehsuara për koeficientët e korrelacionit janë statistikisht

domethënëse në nivelin e rëndësisë $\alpha = 0.01$. Në përputhje me vlerat e gjetura të koeficientëve, të cilat janë shumë afër zeros, mund të themi se thujse mungon një lidhje ndërmjet vlerës së fuqive ditore hyrëse dhe atyre dalëse. Në këtë rast nuk mund të themi se kulmet qëndrore hyrëse janë gjithashtu edhe kulme qëndrore dalëse. Ndonëse dy kulme të ndryshëm nga njëri-tjetri, mund të kenë njëri fuqi hyrëse të lartë dhe tjetri fuqi hyrëse të ulët, ekziston mundësia që të dy këta kulme të kenë fuqi dalëse të lartë. Vemë re se rezultati i gjetur prej nesh në graf rrjetat e komunikimit të telefonisë fikse dallon shumë nga ai i gjetur në [135] në një graf rrjet komunikimi nëpërmjet adresave e-mail, ku është treguar se korrelacioni ndërmjet fuqive hyrëse dhe dalëse është shumë i forte, 0.59.

4.4. PËRFUNDIME

Ndërmjet shpërndarjes PL dhe LN, mënyra më e mirë për të përqasur shpërndarjen e numrit të komunikuesve ditor, të ndryshëm nga njëri-tjetri për klient, u gjet se është shpërndarja LN. Gjithashtu, LN është shpërndarja e cila shpreh më mirë numrin e thirrjeve telefonike ditore totale dhe hyrëse për klient. Të dyja shpërndarjet kanë rrisht të lartë të jenë të papërshtatshme për të shprehur numrin e thirrjeve telefonike ditore dalëse për klient. Një shpërndarje e ndryshme prej tyre, duhet tentuar për këtë qëllim. Në thirrjet ditore të një klienti, mungon korrelacioni ndërmjet numrit të thirrjeve hyrëse me numrin e thirrjeve dalëse. Duke ditur se fuqia e kulmeve në graf rrjetat e komunikimit interpretohet edhe si tregues se sa i “rëndësishëm” është një klient, vihet re se shumica e klientëve kanë një numër të vogël kontaktesh ditore, ndërkohë që një përpjesë tejet e vogël e tyre janë shumë të lidhur, luajnë rol qëndror. Gjatë ditës, mesatarisht 62% e klientëve aktiv kanë komunikim me jo më shumë se dy klientë të ndryshëm dhe 58.83% e klientëve gjenden në komunikim në jo më shumë se dy thirrje (hyrëse ose dalëse). Mesatarisht 82.2% e klientëve aktiv marrin jo më shumë se dy thirrje hyrëse gjatë ditës dhe 80.8 % e klientëve aktiv kryejnë jo më shumë se dy thirrje dalëse gjatë ditës. Këto përqindje janë shumë të rëndësishme nëse nga pikëpamja e politikave të tarifimit shikohet mundësia e tarifimit të diferencuar për komunikimin brenda rrjetit.

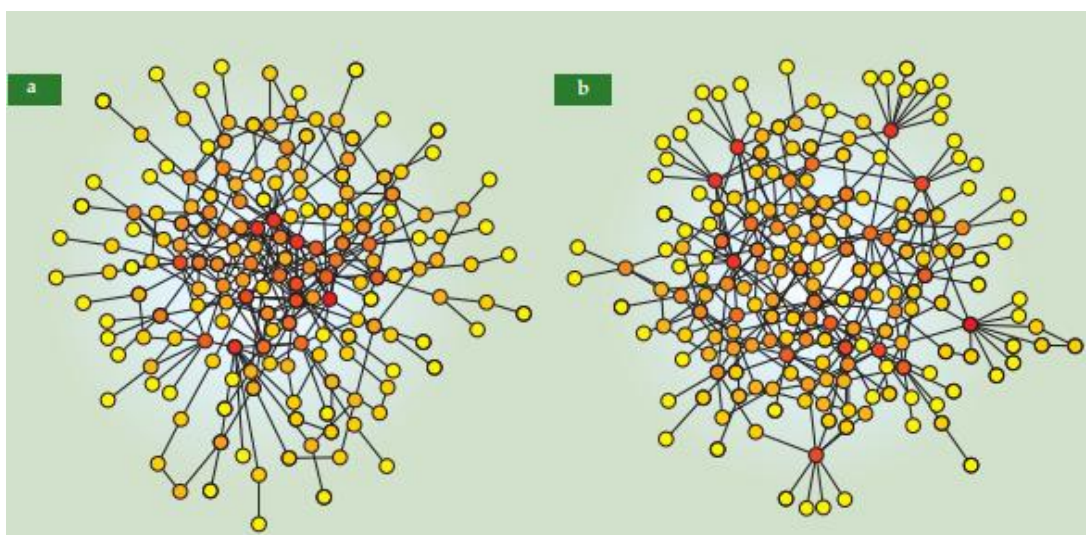
5. KLASIFIKIMET PREFERUESE DHE RECIPROCITETI

5.1. HYRJE

Një cilësi e rëndësishme në graf rrjeta është edhe korrelacioni ndërmjet kulmeve fqinjë në lidhje me një karakteristikë të caktuar. Pavarësisht se dy graf rrjeta mund të kenë varg të njëjtë të fuqive, ato mund të ndryshojnë në mënyrën se si çiftet e kulmeve të tyre lidhen me anë të brinjëve [136]. Lidhjet ndërmjet kulmeve mund të krijohen sipas një parimi parapëlqyes në përputhje me një karakteristikë të caktuar që ata kanë. Kjo karakteristikë mund të jetë nominale, diskrete ose e vazhdueshme.

Koeficienti *i klasifikimeve preferuese* është përkufizuar nga Newman [137]. Ky koeficient mat nivelin e ngjashmërisë ndërmjet kulmeve fqinjë të graf rrjetit bazuar në vlerat e karakteristikës së marr në shqyrtim. Nëse kulmet “preferojnë” të lidhen me kulme të cilët janë të ngjashëm me ta, thuhet se graf rrjeti është *pozitivisht i përcaktuar* sipas klasifikimeve preferuese. Në rast se ata “preferojnë” të lidhen me kulme që janë ndryshe nga ata në lidhje me karakteristikën e marr në shqyrtim, kjo nënkupton që graf rrjeti është *negativisht i përcaktuar* sipas klasifikimeve preferuese.

Në rastin kur karakteristika e marr në shqyrtim është një madhësi strukturore (si psh. fuqia e kulmeve, fortësia e kulmeve), koeficienti i klasifikimeve preferuese (sipas topologjisë) ndikon në mënyrën e sjelljes së graf rrjetit. Nëse karakteristika është fuqia e kulmeve [44] dhe koeficienti është pozitiv, kulmet me fuqi të lartë janë të lidhura me njëra-tjetrën duke u “mpiksur” dhe krijuar një bërthamë e cila në periferi të saj ka kulme me fuqi të vogël. Këto njihen ndryshe edhe si komunitete në graf rrjet. Kulmet me fuqi të lartë tentojnë të jenë të shpërndara gjerësisht nëpër graf rrjet nëse koeficienti është negativ. Kjo situatë përshkruhet në Figurën 5.1.



Figurë 5.1: Dy graf rrjeta me shpërndarje të njëjtë të fuqive të kulmeve por me korrelacion të ndryshëm të fuqive. Graf rrjeti (a) është pozitivisht i përcaktuar sipas klasifikimeve preferuese në lidhje me fuqitë e kulmeve, ndërsa graf rrjeti (b) është negativisht i përcaktuar. Shkëputur nga [44].

Të qenit graf rrjet pozitivisht i përcaktuar sipas klasifikimeve preferuese [18] ndihmon në përhapjen e një “informacioni” në mënyrë epidemike në të. Në të kundërt, nëse graf rrjeti është negativisht i përcaktuar në këtë kuptim, përhapja epidemike do të ishte e kufizuar në segmente të vogla të popullatës në shqyrtim.

Në rrjetin e telefonisë, kjo veti mund të shërbej për përhapjen efektive të informacionit që jepet nëpërmjet fushatave që kanë të bëjnë me ofrimin e shërbimeve të reja. Këto shërbime mund të jenë në lidhje me vetë kompaninë telefonike apo edhe nga kompani të tjera jo telefonike. Për shembull në ato zona, graf rrjetat e të cilave janë pozitivisht të përcaktuara sipas klasifikimeve preferuese, përhapja e informacionit për shërbimet e reja mund të mbështetet edhe në përhapjen “gojë pas goje” [18]. Ndërsa zonat të cilat janë negativisht të përcaktuara, sipas këtij kuptimi, nuk mund të mbështeten në faktin e mësipërm dhe do të kërkonin shpenzime të tjera për këtë qëllim.

Termi *reciprocitet* [40] përdoret për të matur shkallën e simetrisë së marrëdhënieve në një graf rrjet. Një marrëdhënie komunikuese e cila ka reciprocitet të lartë, nënkupton që të dy bashkë-komunikuesit janë njëkohësisht të interesuar për ta ruajtur atë lidhje komunikuese. Nga ana tjetër, nëse reciprociteti është i ulët, njëri prej bashkë-komunikuesve është më tepër aktiv në ruajtjen e lidhjes komunikuese. Reciprocitetin në komunikim e studiojmë duke u nisur nga komunikimet në një graf rrjet aktuale.

Nëse në graf rrjet ndodhin procese të cilat kanë të bëjnë me përhapje [138], atëherë prania e lidhjeve simetrike do të përshpejtonte procesin e përhapjes dhe do të rritë mundësinë e arritjes së target kulmeve nisur nga një kulm fillestar. Gjithashtu, studimi i reciprocitetit të komunikimit ndërmjet klientëve të operatorit është i rëndësishëm sepse mund të shërbej si një tregues i mirë për të parashikuar praninë e lidhjeve komunikuese ndërmjet klientëve në të ardhmen. Hidalgo dhe Rodriguez-Sickert treguan se reciprociteti është parashikues shumë i fortë i qëndrueshmërisë së lidhjeve komunikuese një graf rrjet të telefonisë lëvizëse [139].

Në këtë pjesë, karakteristikat e marra në shqyrtim janë totali i thirrjeve telefonike dhe kohëzgjatja totale e thirrjeve telefonike. Në vijim hulumtohet në lidhje me:

- **P5.1:** Si është e përcaktuar sipas klasifikimeve preferuese një seri kohore graf rrjetash të telefonisë fikse?
- **P5.2:** A mund të pranojmë që klientët, të cilët kryejnë/marrin më shumë thirrje, komunikojnë më shumë me klientë që gjithashtu kryejnë/marrin më shumë thirrje telefonike?
- **P5.3:** A mund të pranojmë që klientët priren të komunikojnë me ata që janë po aq aktiv sikurse vetë ata në kuptimin e kohëzgjatjes së përgjithshme të thirrjeve në të cilat ata gjenden të angazhuar?
- **P5.4:** A ekziston ndonjë dallim domethënës në lidhje me klasifikimet preferuese në rastet kur karakteristikat ndahen në tre kategori: totale, hyrëse dhe dalëse?
- **P5.5:** A është reciprok komunikimi ditor ndërmjet klientëve të telefonisë? Po komunikimi brenda disa ditëve?

5.2. METODOLOGJIA

5.2.1. Kuptime Bazë

5.2.1.1. Koeficienti i Klasifikimeve Preferuese

Le të jetë $G = (V, E)$ graf rrjet i orientuar. Për një karakteristikë të caktuar të kulmeve, për çdo brinjë $e \in E$, përcaktojmë një çift (s_e, t_e) të cilat janë vlerat që merr karakteristika në kulmet që janë incidente me brinjën e . *Koeficienti i klasifikimeve preferuese* sipas topologjisë ndryshe njihet edhe si koeficienti i Pearson për të matur korrelacionin që ekziston ndërmjet kulmeve fqinjë në lidhje me një karakteristikë të

caktuar. Ky koeficient është përcaktuar nga Newman [137] dhe në rastin e një karakteristike diskrete ose të vazhdueshme jepet nga:

$$r = \frac{\sum_{s,t} st(f_{st} - f_{s+}f_{+t})}{\sigma_{\{f_{s+}\}} \sigma_{\{f_{+s}\}}},$$

ku shuma bëhet sipas të gjithë çifteve (s, t) , të cilat janë të përcaktuara në mënyrë të vetme. Përpjesa e brinjëve të G të cilat lidhin një kulm me vlerë të karakteristikës të barabartë me s , me një kulm me vlerë të karakteristikës të barabartë me t është shënuar me f_{st} , ku $\sum_{s,t} f_{st} = 1$.

Le të jetë $\mathbf{f} = [f_{st}]$ matrica rezultante. Me $f_{s+} = \sum_t f_{st}$ është shënuar shuma marxhinale sipas rrjeshtit të s -të dhe me $f_{+t} = \sum_s f_{st}$ shuma marxhinale sipas kolonës së t -të. Në fjalë të tjera, f_{s+} është përpjesa e brinjëve që fillon me një kulm, vlera e karakteristikës së të cilit është s dhe f_{+t} është përpjesa e brinjëve që përfundon në një kulm vlera e karakteristikës së të cilit është t . Për më tepër, le të jenë $\sigma_{\{f_{s+}\}}$ dhe $\sigma_{\{f_{+s}\}}$ devijimet standarte korresponduese respektivisht të shpërndarjeve të frekuencave $\{f_{s+}\}$ dhe $\{f_{+s}\}$.

Në graf rrjetat e paorientuara, duke qënë se $\mathbf{f}$ është simetrike, kemi që $f_{s+} = f_{+s}$ dhe $\sigma_s = \sigma_t$. Për rrjedhoj koeficienti jepet nga formula:

$$r = \frac{\sum_{s,t} st(f_{st} - f_{s+}f_{+t})}{\sigma_{\{f_{s+}\}}^2}.$$

Vlerat e r variojnë ndërmjet -1 dhe 1 . Në rastin kur $0 < r \leq 1$ ($0 \ll \hat{r} \leq 1$), graf rrjeti është pozitivisht i përcaktuar sipas klasifikimeve preferuese dhe kur $-1 \leq r < 0$ ($-1 \leq \hat{r} \ll 0$) është negativisht i përcaktuar. Nëse $r = 1$, është pozitivisht i përcaktuar në mënyrë perfekte dhe kur $r = -1$ është negativisht i përcaktuar në mënyrë perfekte. Në rastin kur $r = 0$ ($\hat{r} \approx 0$), mungon klasifikimi preferencial dhe graf rrjetat mund të quhen neutrale nga kjo pikëpamje.

Nëse vlerat e koeficientit janë shumë afër 1 , kjo nënkupton që kulmet e lidhura tentojnë të kenë vlera shumë të ngjashme të karakteristikës. Në rastin kur vlerat e koeficientit janë shumë afër -1 , kulmet e lidhura tentojnë të kenë vlera shumë të ndryshme të karakteristikës në shqyrtim. Nëse vlerat e njehsuara për koeficientin e klasifikimeve preferuese gjenden në intervalin $(-0.05, 0.05)$, ne i kemi konsideruar ato si jo statistikisht të ndryshme nga vlera 0 .

5.2.1.2. Shkalla e Reciprocitetit

Ekzistojnë dy interpretime për *reciprocitetin e lidhjeve* në një graf rrjet, të cilat dallojnë nga njësia që marrim në shqyrtim e cila mund të jetë:

1. çifte kulmesh të parenditura,
2. brinjë të orientuara.

Sipas rastit të parë, *shkalla e reciprocitetit* (simetrisë, dyanshmërisë) interpretohet si probabiliteti i ekzistencës së një lidhje reciproke ndërmjet një çifti kulmesh. Kulmet v_i dhe v_j janë në një *marrëdhënie reciproke* me njëri-tjetrin atëherë dhe vetëm atëherë kur të dy brinjët me orientime të kundërta të cilat lidhin v_i me v_j gjenden në E . Çifti i kulmeve në këtë rast quhet *çift kulmesh simetrik*, ndërsa çiftet e brinjëve (v_i, v_j) dhe (v_j, v_i) quhen *çift brinjësh simetrike*.

Le të jenë v_i dhe v_j dy kulme në graf rrjetin $G = (V, E)$ i cili është i thjeshtë, i orientuar, i papeshuar. Çiftet (e parenditura) të kulmeve mund ti kategorizojmë në tre gjendje të mundshme:

- a. të palidhura me brinjë midis tyre;
- b. të lidhura me anë të vetëm një brinje të orientuar midis tyre;
- c. të lidhura me anë të dy brinjëve me orientime të kundërta midis tyre.

Shkalla e reciprocitetit në këtë rast [140] përcaktohet si përpjesa që zënë çiftet simetrike të kulmeve në numrin total të çifteve të parenditura në graf rrjet:

$$Rec_1(G) = \frac{n(c)}{n(b) + n(c)},$$

ku $n(b)$ është numri i çifteve të parenditura të kulmeve të cilat lidhen vetëm me anë të një brinje të orientuar midis tyre, ndërsa $n(c)$ është numri i atyre që lidhen me anë të dy brinjëve të orientuara në drejtime të kundërta.

Sipas rastit të dytë, shkalla e reciprocitetit [69] në një graf rrjet G në përputhje me përkufizimin klasik jepet nga formula:

$$Rec_2(G) = \frac{B^{\leftrightarrow}}{B},$$

ku me $B = \sum_{i \neq j} a_{ij}$ është shënuar numri total i brinjëve të orientuara (v_i, v_j) . Madhësia $B^{\leftrightarrow} = \sum_{i \neq j} a_{ij} a_{ji}$ tregon numrin e brinjëve të orientuara (v_i, v_j) , për të cilat ekziston edhe një brinjë me orientim të kundërt. Me a_{ij} janë shënuar elementët e matricës së fqinjësisë së graf rrjetit, të cilët marrin vlerat $a_{ij} = 1$ në rastin kur ekziston një brinjë e orientuar me fillim në kulmin v_i dhe mbarim në kulmin v_j dhe $a_{ij} = 0$ në rast të kundërt. Duke qënë se $i \neq j$, leqet nuk merren në shqyrtim.

Sipas të dy përcaktimeve, vlerat e shkallës së reciprocitetit variojnë nga 0 (nuk ekzistojnë marrëdhënie reciproke) në 1 (të gjitha marrëdhëniet midis kulmeve në graf rrjet janë reciproke).

Vlera të mëdha të shkallës së reciprocitetit janë shumë domethënëse në rastin kur ato hasen në një graf rrjet të rrallë, ndërkohë që vlerat e vogla do të ishin domethënëse nëse do të shfaqeshin në një graf rrjet të dendur. Për këtë arsye, vlera e shkallës së reciprocitetit krahasohet me vlerën e densitetit, e cila gjithashtu përputhet me reciprocitetin e një graf rrjete rasti ekuivalente. Problemi i reciprocitetit të lidhjeve, prania jo rastësore e tyre ndërmjet një çifti kulmesh, është trajtuar edhe nga Garlaschelli dhe Loffredo [138]. Ata propozuan një madhësi të re për të matur shkallën e reciprocitetit e cila jepet si korrelacioni ndërmjet elementëve të matricës së fqinjësisë së graf rrjetës së orientuar:

$$Rec_2^{GL}(G) = \frac{\sum_{i \neq j} (a_{ij} - \bar{a})(a_{ji} - \bar{a})}{\sum_{i \neq j} (a_{ij} - \bar{a})^2},$$

ku $\bar{a} = \frac{1}{|V|(|V|-1)} \sum_{i \neq j} a_{ij}$. Kjo madhësi mund të reduktohet në formulën e mëposhtme:

$$Rec_2^{GL}(G) = \frac{Rec_2(G) - den(G)}{1 - den(G)},$$

e cila në vetvete gërsheton madhësinë $Rec_2(G)$ dhe densitetin e graf rrjetit të thjeshtë të orientuar $den(G) = \frac{B}{|V|(|V|-1)}$.

Nëse $Rec_2^{GL}(G) > 0$ atëherë graf rrjeti është pozitivisht i përcaktuar në lidhje me reciprocitetin (reciprok) dhe kur $Rec_2^{GL}(G) < 0$ është negativisht i përcaktuar (anti reciprok). Në rast se $Rec_2^{GL}(G) = 0$ thuhet se graf rrjeti është neutral (areciprok). Devijimi standart $\sigma_{Rec_2^{GL}(G)}$ shprehet nga formula:

$$\sigma_{Rec_2^{GL}(G)} = \sqrt{B^{\leftrightarrow}(Rec_2^{GL}(G) - Rec_2^{GL}(G)^{\leftrightarrow})^2 + (B - B^{\leftrightarrow})(Rec_2^{GL}(G) - Rec_2^{GL}(G)^{\rightarrow})^2},$$

ku

$$Rec_2^{GL}(G)^{\leftrightarrow} = \frac{\frac{B^{\leftrightarrow} - 2}{B - 2} - \frac{B - 2}{|V|(|V| - 1)}}{1 - \frac{B - 2}{|V|(|V| - 1)}},$$

është vlera e $Rec_2^{GL}(G)$ në rastin kur largohet nga graf rrjeti një çift brinjësh simetrike dhe

$$Rec_2^{GL}(G)^{\rightarrow} = \frac{\frac{B^{\leftrightarrow}}{B - 1} - \frac{B - 1}{|V|(|V| - 1)}}{1 - \frac{B - 1}{|V|(|V| - 1)}},$$

është vlera e $Rec_2^{GL}(G)$ në rastin kur largohet nga graf rrjeti një brinjë, ndërmjet dy kulmeve të lidhur vetëm me anë të një brinje.

5.2.2. Ndërtimi i Modeleve

Karakteristikat që janë marr në shqyrtim në secilin kulm (klient aktiv) janë:

- **K₁**- totali i thirrjeve telefonike⁶⁴;
- **K₂**- totali i kohëzgjatjes së thirrjeve telefonike në të cilat ka qënë i angazhuar.

Modelet e përdorura në këtë pjesë janë modelet e përmendura më parë edhe në Kapitullin 2.

- **Modeli 2.1:** $G_1, G_2, \dots, G_{30}$, është i tillë që çdo graf rrjet G_i , për $i = 1, \dots, 30$ është ndërtuar në bazë vetëm të thirrjeve të ditës së i -të të muajit;
- **Modeli 2.2:** $G_1 \leq G_2 \leq \dots \leq G_{30}$, është i tillë që çdo graf rrjet G_i është ndërtuar bazuar vetëm mbi ato thirrje që janë kryer prej ditës së parë të muajit, deri në ditën e i -të të tij. Graf rrjetat në këtë seri janë të tilla që $G_i \leq G_{i+1}$, për $i = 1, \dots, 29$.

Në rastin e analizës për koeficientin e klasifikimeve preferuese, në dy modelet e mësipërme janë bërë njehsimet në lidhje me karakteristikat **K₁** dhe **K₂** për secilin graf rrjet. Më pas, graf rrjetat në të dy modelet janë thjeshtuar, janë trajtuar si të paorientuara, të papeshuara dhe kulmet e tyre janë pajisur me një karakteristikë numerike diskrete

⁶⁴ Vlerat për karakteristikën **K₁** mund të shikohen edhe si fuqitë e kulmeve në graf rrjetat e pathjeshtuara.

(K_1 ose K_2). Në përputhje me këto karakteristika dallojmë nën modelet e mëposhtme. Nën modelet e Modelit 2.1 janë:

- **Modeli 5.1.1-Totale:** Kulmet janë të pajisur me një karakteristikë numerike që tregon numrin total të thirrjeve telefonike (hyrëse ose dalëse) në të cilat ata janë gjendur në komunikim gjatë ditës së i -të.
- **Modeli 5.1.1-Hyrëse:** Kulmet janë të pajisur me një karakteristikë numerike që tregon numrin total të thirrjeve telefonike hyrëse në të cilat ata janë gjendur në komunikim gjatë ditës së i -të.
- **Modeli 5.1.1-Dalëse:** Kulmet janë të pajisur me një karakteristikë numerike që tregon numrin total të thirrjeve telefonike dalëse në të cilat ata janë gjendur në komunikim gjatë ditës së i -të.
- **Modeli 5.1.2-Totale:** Kulmet janë të pajisur me një karakteristikë numerike që tregon kohëzgjatjen totale të thirrjeve telefonike (hyrëse ose dalëse) në të cilat ata janë gjendur në komunikim gjatë ditës së i -të.
- **Modeli 5.1.2-Hyrëse:** Kulmet janë të pajisur me një karakteristikë numerike që tregon kohëzgjatjen totale të thirrjeve telefonike hyrëse në të cilat ata janë gjendur në komunikim gjatë ditës së i -të.
- **Modeli 5.1.2-Dalëse:** Kulmet janë të pajisur me një karakteristikë numerike që tregon kohëzgjatjen totale të thirrjeve telefonike dalëse në të cilat ata janë gjendur në komunikim gjatë ditës së i -të.

Në mënyrë të ngjashme edhe për Modelin 2.2 janë përcaktuar nën modelet si vijon:

- **Modeli 5.2.1-Totale:** Kulmet janë të pajisur me një karakteristikë numerike që tregon numrin total të thirrjeve telefonike (hyrëse ose dalëse) në të cilat ata janë gjendur në komunikim nga dita e parë deri në ditën e i -të.
- **Modeli 5.2.1-Hyrëse:** Kulmet janë të pajisur me një karakteristikë numerike që tregon numrin total të thirrjeve telefonike hyrëse në të cilat ata janë gjendur në komunikim nga dita e parë deri në ditën e i -të.
- **Modeli 5.2.1-Dalëse:** Kulmet janë të pajisur me një karakteristikë numerike që tregon numrin total të thirrjeve telefonike dalëse në të cilat ata janë gjendur në komunikim nga dita e parë deri në ditën e i -të.
- **Modeli 5.2.2-Totale:** Kulmet janë të pajisur me një karakteristikë numerike që tregon numrin total të thirrjeve telefonike (hyrëse ose dalëse) në të cilat ata janë gjendur në komunikim nga dita e parë deri në ditën e i -të.
- **Modeli 5.2.2-Hyrëse:** Kulmet janë të pajisur me një karakteristikë numerike që tregon numrin total të thirrjeve telefonike hyrëse në të cilat ata janë gjendur në komunikim nga dita e parë deri në ditën e i -të.
- **Modeli 5.2.2-Dalëse:** Kulmet janë të pajisur me një karakteristikë numerike që tregon numrin total të thirrjeve telefonike dalëse në të cilat ata janë gjendur në komunikim nga dita e parë deri në ditën e i -të.

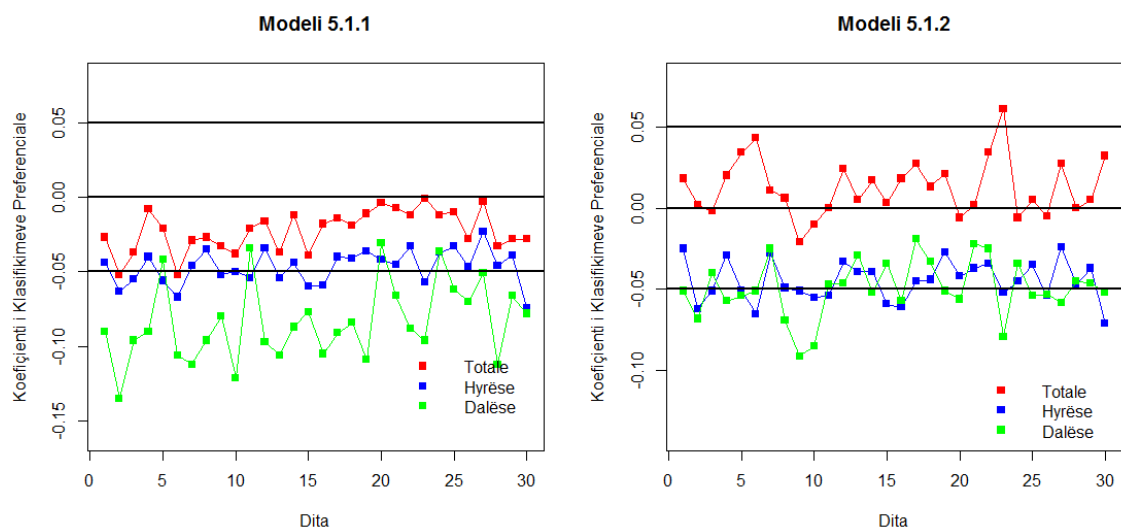
Ndërkohë, në rastin e studimit për shkallën e reciprocitetit midis kulmeve, graf rrjetat në të dy modelet (Modeli 2.1 dhe Modeli 2.2) janë thjeshtuar dhe trajtuar si të orientuara, të papeshuara.

5.3. REZULTATET E HULUMTIMIT PËR KLASIFIKIMET PREFERUESE

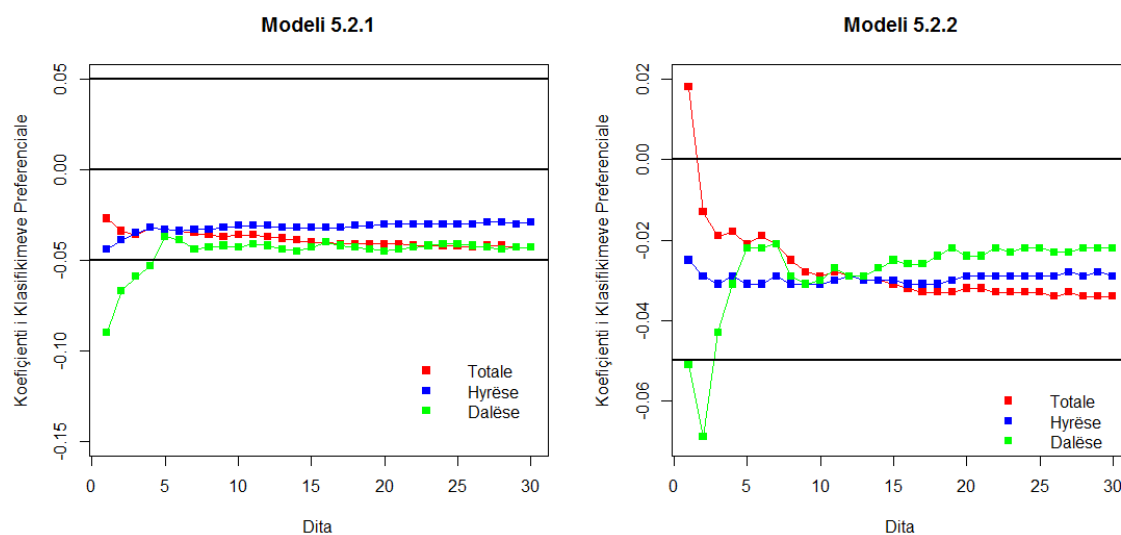
Në Tabelën 5.1 janë dhënë vlerat e gjetura për koeficientin e klasifikimeve preferuese në secilin prej graf rrjetave të marra në shqyrtim në nën modelet e përcaktuara. Vizualisht vlerat e koeficientit të klasifikimeve preferuese janë dhënë në Figurën 5.2 dhe Figurën 5.3. Përqindja e vlerave të $\hat{r}$ të cilat gjenden në intervalin $(-0.05, 0.05)$ sipas modeleve përkatëse është: Modeli 5.1.1-Totale (93.3%), Hyrëse (60%), Dalëse (13.3%); Modeli 5.1.2-Totale (96.7%), Hyrëse (60%), Dalëse (43.3%); Modeli 5.2.1-Totale (100%), Hyrëse (100%), Dalëse (86.7%); Modeli 5.2.2-Totale(100%), Hyrëse (100%), Dalëse (93.3%).

Tabelë 5.1: Vlerësimet e bëra për vlerat e koeficientit të klasifikimeve preferuese. Vlerat janë rumbullakosur në të mijtat.

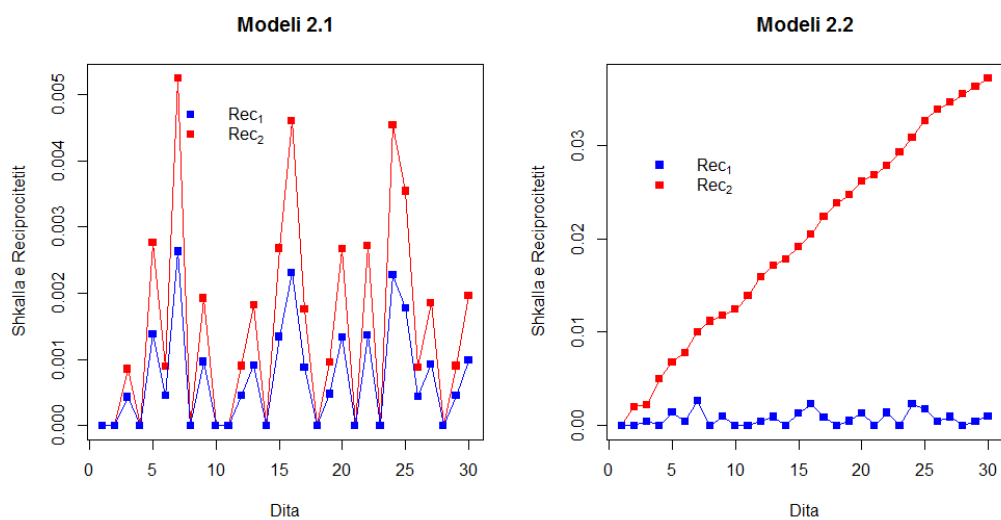
G_i	Modeli 5.1.1			Modeli 5.1.2			Modeli 5.2.1			Modeli 5.2.2		
	Totale	Hyrëse	Dalëse	Totale	Hyrëse	Dalëse	Totale	Hyrëse	Dalëse	Totale	Hyrëse	Dalëse
1	-0.027	-0.044	-0.090	0.018	-0.025	-0.051	-0.027	-0.044	-0.090	0.018	-0.025	-0.051
2	-0.052	-0.063	-0.135	0.002	-0.062	-0.068	-0.034	-0.039	-0.067	-0.013	-0.029	-0.069
3	-0.037	-0.055	-0.096	-0.002	-0.051	-0.040	-0.036	-0.035	-0.059	-0.019	-0.031	-0.043
4	-0.008	-0.040	-0.090	0.020	-0.029	-0.057	-0.032	-0.032	-0.053	-0.018	-0.029	-0.031
5	-0.021	-0.056	-0.042	0.034	-0.051	-0.054	-0.033	-0.033	-0.037	-0.021	-0.031	-0.022
6	-0.052	-0.067	-0.106	0.043	-0.065	-0.051	-0.034	-0.034	-0.039	-0.019	-0.031	-0.022
7	-0.029	-0.046	-0.112	0.011	-0.028	-0.025	-0.035	-0.033	-0.044	-0.021	-0.029	-0.021
8	-0.027	-0.035	-0.096	0.006	-0.049	-0.069	-0.036	-0.033	-0.043	-0.025	-0.031	-0.029
9	-0.033	-0.052	-0.080	-0.021	-0.051	-0.091	-0.037	-0.032	-0.042	-0.028	-0.031	-0.031
10	-0.038	-0.050	-0.121	-0.010	-0.055	-0.085	-0.036	-0.031	-0.043	-0.029	-0.031	-0.030
11	-0.021	-0.054	-0.034	0.000	-0.054	-0.047	-0.036	-0.031	-0.041	-0.028	-0.030	-0.027
12	-0.016	-0.034	-0.097	0.024	-0.033	-0.046	-0.037	-0.031	-0.042	-0.029	-0.029	-0.029
13	-0.037	-0.054	-0.106	0.005	-0.039	-0.029	-0.038	-0.032	-0.044	-0.030	-0.030	-0.029
14	-0.012	-0.044	-0.087	0.017	-0.039	-0.052	-0.039	-0.032	-0.045	-0.030	-0.030	-0.027
15	-0.039	-0.060	-0.077	0.003	-0.059	-0.034	-0.040	-0.032	-0.043	-0.031	-0.030	-0.025
16	-0.018	-0.059	-0.105	0.018	-0.061	-0.057	-0.040	-0.032	-0.040	-0.032	-0.031	-0.026
17	-0.014	-0.040	-0.091	0.027	-0.045	-0.019	-0.041	-0.032	-0.042	-0.033	-0.031	-0.026
18	-0.019	-0.041	-0.084	0.013	-0.044	-0.033	-0.041	-0.031	-0.043	-0.033	-0.031	-0.024
19	-0.011	-0.036	-0.109	0.021	-0.027	-0.051	-0.041	-0.031	-0.044	-0.033	-0.030	-0.022
20	-0.004	-0.042	-0.031	-0.006	-0.042	-0.056	-0.041	-0.030	-0.045	-0.032	-0.029	-0.024
21	-0.007	-0.045	-0.066	0.002	-0.037	-0.022	-0.041	-0.030	-0.044	-0.032	-0.029	-0.024
22	-0.012	-0.033	-0.088	0.034	-0.034	-0.025	-0.042	-0.030	-0.043	-0.033	-0.029	-0.022
23	-0.001	-0.057	-0.096	0.061	-0.052	-0.079	-0.042	-0.030	-0.042	-0.033	-0.029	-0.023
24	-0.012	-0.037	-0.036	-0.006	-0.045	-0.034	-0.042	-0.030	-0.041	-0.033	-0.029	-0.022
25	-0.010	-0.033	-0.062	0.005	-0.035	-0.054	-0.042	-0.030	-0.041	-0.033	-0.029	-0.022
26	-0.028	-0.047	-0.070	-0.005	-0.054	-0.053	-0.043	-0.030	-0.042	-0.034	-0.029	-0.023
27	-0.003	-0.023	-0.051	0.027	-0.024	-0.058	-0.042	-0.029	-0.043	-0.033	-0.028	-0.023
28	-0.033	-0.046	-0.112	0.000	-0.047	-0.045	-0.042	-0.029	-0.044	-0.034	-0.029	-0.022
29	-0.028	-0.039	-0.066	0.005	-0.037	-0.046	-0.043	-0.030	-0.043	-0.034	-0.028	-0.022
30	-0.028	-0.074	-0.078	0.032	-0.071	-0.052	-0.043	-0.029	-0.043	-0.034	-0.029	-0.022



Figurë 5.2: Koeficienti i klasifikimeve preferuese sipas nën modeleve të Modelit 2.1.



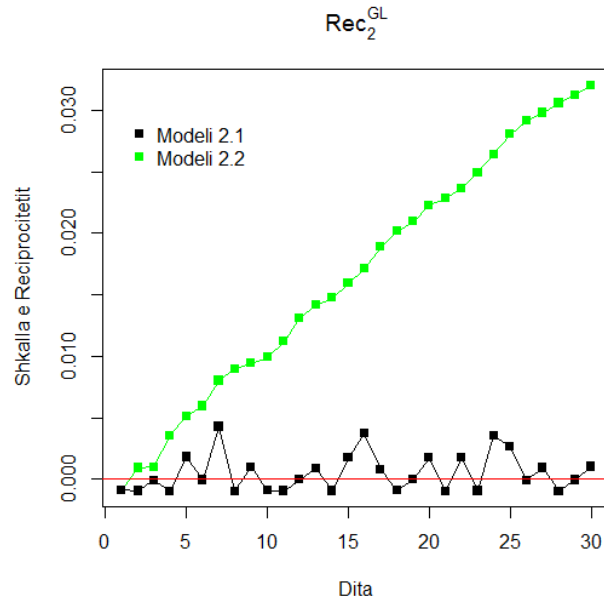
Figurë 5.3: Koeficienti i klasifikimeve preferuese sipas nën modeleve të Modelit 2.2.



Figurë 5.4: Shkalla e reciprocitetit e njehsuar sipas rastit të parë dhe të dytë për Modelin 2.1 dhe Modelin 2.2.

Tabelë 5.2: Vlerat e shkallës së reciprocitetit. Vlera NaN (= 0.0/0.0) në tabelë nënkupton që në atë rast rezultati nuk mund të përcaktohet si numër. Vlerat janë rumbullakosur në të dhjetëmijtat.

G_i	Modeli 2.1				Modeli 2.2			
	Rec_1	Rec_2	Rec_2^{GL}	$\sigma_{Rec_2^{GL}}$	Rec_1	Rec_2	Rec_2^{GL}	$\sigma_{Rec_2^{GL}}$
1	0	0	-0.00087	NaN	0	0	-0.00087	NaN
2	0	0	-0.00095	NaN	0.00099	0.00198	0.00092	0.00140
3	0.00043	0.00085	-0.00011	0.00121	0.00112	0.00224	0.00098	0.00119
4	0	0	-0.00097	NaN	0.00252	0.00502	0.00355	0.00147
5	0.001383	0.00276	0.00182	0.00225	0.00339	0.00677	0.00512	0.00135
6	0.00045	0.00090	-0.00005	0.00127	0.00390	0.00778	0.00595	0.00086
7	0.00263	0.00524	0.00428	0.00303	0.00503	0.01002	0.00802	NaN
8	0	0	-0.00097	NaN	0.00561	0.01117	0.00897	NaN
9	0.00096	0.00192	0.00098	0.00192	0.00593	0.01180	0.00945	NaN
10	0	0	-0.00093	NaN	0.00625	0.01241	0.00991	NaN
11	0	0	-0.00094	NaN	0.00697	0.01385	0.01121	NaN
12	0.00045	0.00090	-0.00005	0.00127	0.00801	0.01590	0.01312	NaN
13	0.00091	0.00182	0.00090	0.00182	0.00863	0.01711	0.01416	NaN
14	0	0	-0.00095	NaN	0.00899	0.01782	0.01475	NaN
15	0.00134	0.00268	0.00175	0.00219	0.00967	0.01915	0.01593	NaN
16	0.00231	0.00460	0.00370	0.00291	0.01035	0.02048	0.01712	NaN
17	0.00088	0.00175	0.00081	0.00175	0.01132	0.02239	0.01889	NaN
18	0	0	-0.00094	NaN	0.01203	0.02378	0.02017	NaN
19	0.00048	0.00095	0.00001	0.00135	0.01250	0.02469	0.02095	NaN
20	0.001338	0.00267	0.00175	0.00217	0.01325	0.02615	0.02228	NaN
21	0	0	-0.00099	NaN	0.01360	0.02684	0.02282	NaN
22	0.00136	0.00271	0.00174	0.00222	0.01411	0.02783	0.02368	NaN
23	0	0	-0.00094	NaN	0.01482	0.02922	0.02494	NaN
24	0.00227	0.00453	0.00357	0.00287	0.01565	0.03082	0.02642	NaN
25	0.00177	0.00354	0.00261	0.00250	0.01657	0.03261	0.02807	NaN
26	0.00044	0.00088	-0.00007	0.00124	0.01720	0.03382	0.02917	NaN
27	0.00093	0.00185	0.00093	0.00185	0.01760	0.03459	0.02982	NaN
28	0	0	-0.00099	NaN	0.01806	0.03548	0.03059	NaN
29	0.00045	0.00090	-0.00006	0.00127	0.01848	0.03629	0.03127	NaN
30	0.00098	0.00196	0.00102	0.00196	0.01893	0.03716	0.03202	NaN



Figurë 5.5: Shkalla e reciprocitetit, e njehsuar si korrelacion ndërmjet elementëve të matricës së fqinjësisë së graf rrjetave të orientuara.

5.4. DISKUTIME

Në graf rrjetat e rastit mungon korrelacioni ndërmjet fuqive të kulmeve fqinjë [7]. Në disa raste të marra në shqyrtim, në lidhje me klasifikimet preferuese sipas fuqive të kulmeve, graf rrjetat teknologjike dhe biologjike kanë rezultuar negativisht të përcaktuara [137] ndërsa graf rrjetat sociale kanë rezultuar pozitivisht të përcaktuara. Ndërkohë, në seritë kohore të graf rrjetave të cilat ne kemi trajtuar vihet re një luhajtje midis shenjës pozitive dhe negative për vlerësimet empirike të bëra për koeficientin e klasifikimeve preferuese. Në të shumtën e rasteve, në graf rrjetat e modeleve të marra në shqyrtim, vlerësimet e bëra për $\hat{r}$ janë në intervalin $(-0.05, 0.05)$. Ky rezultat nënkupton se vlerat e gjetura nuk janë statistikisht shumë të ndryshme nga vlera zero, e për rrjedhojë graf rrjetat mund të konsiderohen si përafërsisht neutrale në lidhje me koeficientin e klasifikimeve preferuese për karakteristikat numerike të marra në shqyrtim.

Dallues është rasti i Modelit 5.1.1-Dalëse dhe Modelit 5.1.2-Dalëse. Në Modelin 5.1.1-Dalëse, 86.7% e graf rrjetave rezultojnë me vlerë në kufijtë $-0.135 \leq \hat{r} \leq -0.051$. Ndërsa në Modelin 5.1.2-Dalëse vlerësimet në 56.7% të graf rrjetave rezultojnë në kufijtë $-0.091 \leq \hat{r} \leq -0.051$. Mund të themi se këto graf rrjeta, të cilat janë të ndërtuara në bazë ditore, janë lehtësisht të përcaktuara negativisht sipas klasifikimeve preferuese në lidhje me numrin total të thirrjeve dalëse dhe kohëzgjatjen totale të thirrjeve dalëse. Kulmet me vlerë të lartë të këtyre karakteristikave kanë prirje shumë të lehtë të kenë lidhje me kulme të cilët kanë vlerë të ulët të saj.

Vlerat e koeficientit të klasifikimeve preferuese sipas fuqive, në rastin e një graf rrjeti të telefonisë lëvizëse të studiuar në katër zona [18] të ndryshme, nga Nanavati *et al.* ishin si në Tabelën 5.4. Graf rrjetat ishin pozitivisht të përcaktuara në rastin e thirrjeve hyrëse dhe lehtësisht, negativisht të përcaktuara në rastin e thirrjeve dalëse. Vlerat e koeficientit dallonin nga zona në zonë.

Tabelë 5.3: Koeficienti i korrelacionit për katër zona. Burimi i informacionit [18].

	Zona A	Zona B	Zona C	Zona D
Hyrëse	0.293915	0.091079	0.175268	0.121410
Dalëse	-0.091477	-0.081860	-0.088832	-0.040468

Në rastin tonë, në vlerë absolute vlerat e koeficientit të klasifikimeve preferuese, në lidhje me karakteristikën $\mathbf{K}_1$ (totalin e thirrjeve), janë kryesisht më të vogla se vlerat e gjetura në studimin e sipër përmendur. Në rastin e graf rrjetit të telefonisë lëvizëse të marr në shqyrtim nga Noka (Jani) [141] intensiteti i përdorimit të telefonit celular (totali i thirrjeve) nuk u vlerësua si faktor i rëndësishëm në përzgjedhjen e individëve me të cilët klientët e telefonisë komunikojnë.

Sipas [138] graf rrjetat reale janë gjithmonë të përcaktuara pozitivisht ose negativisht në lidhje me reciprocitetin por rezultatet e gjetura prej nesh kundërshtojnë disi këtë pohim. Nga studimi ynë, vlerat e shkallës së reciprocitetit sipas të dy përcaktimeve Rec_1 dhe Rec_2 , në graf rrjetat e Modelit 2.1, ose janë zero ose shumë afër zeros. Kjo nënkupton që në graf rrjetat e Modelit 2.1 pothuajse nuk ekzistojnë marrëdhëniet reciproke. Vlerat e Rec_2^{GL} për Modelin 2.1 përveç se qëndrojnë shumë afër zeros, kanë edhe luhajtje në shenjë. Sipas përkufizimit të Garlaschelli dhe Loffredo [138], ne mund ti konsiderojmë këto graf rrjeta si afërsisht neutrale në lidhje me reciprocitetin e marrëdhënieve.

Vlerat e Rec_1 dhe Rec_2 në Modelin 2.2, pësojnë një rritje në raport me graf rrjetat e Modelit 2.1, por gjithësesi ato qëndrojnë shumë afër zeros. Në këtë model, vlerat e Rec_2^{GL} kanë një rritje të lehtë duke filluar nga G_1 deri tek G_{30} por vazhdojnë të qëndrojnë shumë afër zeros.

Shpeshherë graf rrjetat e komunikimit ndërtohen mbi bazën e lidhjeve reciproke, duke anashkaluar ato lidhje të cilat nuk plotësojnë këtë kriter. Nga rezultatet e gjetura u vu re se, reciprociteti është në një shkallë tejte të ulët, pothuajse jo ekzistent. Kështu që, ndërtimi i modeleve për graf rrjetin e telefonisë fikse mbështetur vetëm në thirrjet reciproke, do të pasqyronte fare pak nga graf rrjeti real.

Në lidhje me reciprocitetin e komunikimit, në graf rrjetin e ndërtuar nga adresat email [138] ka rezultuar se $Rec_2^{GL} = 0.231$. Ndërsa Kovanen *et al.* [40] në studimin e tyre në një graf rrjet të telefonisë lëvizëse treguan se zakonisht marrëdhëniet komunikuese ishin jo reciproke.

5.5. PËRFUNDIME

Gjatë këtij kapitulli u shqyrtua vlera e koeficientit të klasifikimeve preferuese në lidhje me dy karakteristika: totalin e thirrjeve për klient dhe kohëzgjatjen totale të thirrjeve për klient; dhe reciprociteti i marrëdhënieve komunikuese. Seritë kohore të graf rrjetave të telefonisë fikse, si ato në bazë ditore apo në bazë intervali ditor, përgjithësisht rezultuan si përafërsisht neutrale në lidhje me klasifikimet preferuese sipas karakteristikave të marra në shqyrtim. Në bazë intervali ditor, të gjitha karakteristikat rezultuan të jenë jo domethënëse për të përcaktuar preferencat në komunikim. Në bazë ditore, numri total i thirrjeve (hyrëse ose dalëse), numri total i thirrjeve hyrëse, kohëzgjatja totale e thirrjeve (hyrëse ose dalëse) dhe kohëzgjatja totale e thirrjeve hyrëse rezultuan jo domethënëse për të përcaktuar preferencat në komunikim ndërmjet

klientëve të telefonisë përgjatë ditës. Dallues ishte rasti në modelet e ndërtura në bazë ditore kur shqyrtohej totali i numrit të thirrjeve dalëse dhe totali i kohëzgjatjes së thirrjeve dalëse për klient. Klientët aktiv të cilët kishin kryer një numër të lartë thirrjesh dalëse, kishin një prirje të lehtë të komunikojnë me klientë të tjerë të cilët kishin kryer një numër të ulët thirrjesh dalëse. E njëjta prirje u vu re në bazë ditore edhe në rastin e totalit të kohëzgjatjes së thirrjeve dalëse. Gjatë komunikimit ditor nëpërmjet telefonisë fikse pothuajse nuk ekzistonin marrëdhëniet reciproke ndërmjet klientëve. Edhe në rastin kur komunikimi u pa në interval ditor, reciprociteti i marrëdhënieve vazhdoi të qëndronte në nivele shumë të ulta.

6. KONKLUZIONE DHE REKOMANDIME

Qëllimi i këtij studimi ishte analiza statistikore mbi dinamikën e graf rrjetave të telefonisë fikse. Si përparësi, përmendim faktin që graf rrjeti i telefonisë fikse analizohet nga pikëpamja statistikore përgjatë zhvillimit dinamik të tij, në evoluim në lidhje me kohën. Në dijeninë tonë një studim i tillë në graf rrjeta të telefonisë fikse është i trajtuar për herë të parë. Analiza statistikore e strukturës së graf rrjetit të klientëve të telefonisë na ndihmon të kuptojmë se si klientët e operatorit e përdorin telefoninë fikse midis tyre. Përveç të kuptuarit, interes ka edhe parashikimi. Kjo shërben për të ndërtuar politika biznesi apo infrastrukture më të mira.

Prania e fenomenit “botë e vogël” u konfirmua në çdo hap të kohës në seritë kohore të graf rrjetave të marra në shqyrtim, në rastin kur gjatë njehsimit të madhësisë që mat në mënyrë sasiore këtë fenomen u përdor koeficienti i kalimtaritetit. Ndërkohë, në rastin kur u përdor koeficienti i grupimeve, prania e fenomenit konfirmohet vetëm në 93.3% të rasteve në bazë kohore ditore dhe në 100% të rasteve në bazë kohore intervale. Kjo nënkupton se, informacioni do të përhapej në mënyrë të shpejtë në graf rrjet si brenda një dite apo edhe brenda një intervali ditor. Korrelacioni për “botë e vogël” në lidhje me: rendin, përmasën dhe fuqinë mesatare të kulmeve; mungonte në bazë kohore ditore, por ishte i pranishëm në bazë kohore intervale. Lidhja e gjetur ndërmjet përmasës dhe rendit në graf rrjeta mund të përdoret për të parashikuar numrin e çifteve të ndryshme të klientëve të operatorit, që komunikojnë brenda një dite ose intervali ditor, në bazë të numrit të klientëve aktiv që ai ka.

Mënyra më e mirë për të shpjeguar normën e rrezikut të kushtëzuar të një lidhje komunikuese ishte nëpërmjet modelit të shtresëzuar Cox. Rezultoi se ndryshorja në lidhje me “përkatësinë në grup (am/pm)” dhe ndryshorja në lidhje me “orët e ditës” ishin të lidhura me një prognozë të mirë për kohën gjatë së cilës një thirrje telefonike i mbijetonte ndërprerjes së saj. Duke qënë se këto ndryshore nuk e plotësonin supozimin në lidhje me rrezikun proporcional, u vendosën në shtresa. Vlera të larta të ndryshores “dita e muajit” kishin lidhje me normë të lartë të rrezikut dhe mbijetesë të dobët të lidhjeve komunikuese. Ndryshorja “dita e javës” rezultoi pa ndonjë domethënie statistikore në lidhje me prognozën e mbijetesës apo normën e kushtëzuar të dështimeve të lidhjeve komunikuese. Do të qe interesante për të parë më tej në lidhje me përfshirjen e ndërveprimit ndërmjet ndryshoreve shpjeguese në modelet e mbijetesës dhe gjithashtu aplikimi i analizës parametrike në të dhënat e komunikimit telefonik.

Midis shpërndarjes power law (PL) dhe lognormale (LN), mënyra më e mirë për të përqasur shpërndarjen e numrit të komunikuesve ditor, të ndryshëm nga njëri-tjetri, për klient u gjet se është shpërndarja LN. Gjithashtu, LN është shpërndarja e cila shpreh më mirë numrin e thirrjeve telefonike ditore totale dhe hyrëse për klient. Të dyja shpërndarjet kanë rrisht të lartë të jenë të papërshtatshme për të shprehur numrin e thirrjeve telefonike ditore dalëse për klient. Një shpërndarje e ndryshme prej tyre, duhet tentuar për këtë qëllim. Në thirrjet ditore të një klienti, mungon korrelacioni ndërmjet numrit të thirrjeve hyrëse me numrin e thirrjeve dalëse. Duke ditur se fuqia e kulmeve në graf rrjetat e komunikimit interpretohet edhe si tregues se sa i “rëndësishëm” është një klient, vihet re se shumica e klientëve kanë një numër të vogël kontaktesh ditore, ndërkohë që një përpjesë shumë e vogël e tyre janë shumë të lidhur, luajnë rol qendror. Gjatë ditës, mesatarisht 62% e klientëve aktiv kanë komunikim me jo më shumë se dy

klientë të ndryshëm dhe 58.83% e klientëve gjenden në komunikim në jo më shumë se dy thirrje (hyrëse ose dalëse). Mesatarisht 82.2% e klientëve aktiv marrin jo më shumë se dy thirrje hyrëse gjatë ditës dhe 80.8 % e klientëve aktiv kryejnë jo më shumë se dy thirrje dalëse gjatë ditës. Këto përqindje janë shumë të rëndësishme nëse nga pikëpamja e politikave të tarifimit shikohet mundësia e tarifimit të diferencuar për komunikimin brenda rrjetit.

Pasi u shqyrtuan vlerësimet e bëra në lidhje me koeficientin e klasifikimeve preferuese në lidhje me dy karakteristika: totalin e thirrjeve për klient dhe kohëzgjatjen totale të thirrjeve për klient; rezultoi se graf rrjetat si në bazë ditore apo në bazë intervali ditor ishin përafërsisht neutrale në lidhje me klasifikimet preferuese sipas karakteristikave të marra në shqyrtim. Në bazë intervali ditor, të gjitha karakteristikat rezultuan të jenë jo domethënëse për të përcaktuar preferencat në komunikim. Në bazë ditore, numri total i thirrjeve (hyrëse ose dalëse), numri total i thirrjeve hyrëse, kohëzgjatja totale e thirrjeve (hyrëse ose dalëse) dhe kohëzgjatja totale e thirrjeve hyrëse rezultuan jo domethënëse për të përcaktuar preferencat në komunikim ndërmjet klientëve të telefonisë përgjatë ditës. Dallues ishte rasti në modelet e ndërtura në bazë ditore kur shqyrtohej totali i numrit të thirrjeve dalëse dhe totali i kohëzgjatjes së thirrjeve dalëse për klient. Klientët aktiv të cilët kishin kryer një numër të lartë thirrjesh dalëse, kishin një prirje të lehtë të komunikonin me klientë të tjerë të cilët kishin kryer një numër të ulët thirrjesh dalëse. E njëjta prirje u vu re në bazë ditore edhe në rastin e totalit të kohëzgjatjes së thirrjeve dalëse. Gjatë komunikimit ditor nëpërmjet telefonisë fikse pothuajse nuk ekzistonin marrëdhëniet reciproke ndërmjet klientëve. Edhe në rastin kur komunikimi u pa në interval ditor, reciprociteti i marrëdhënieve qëndronte në nivele shumë të ulta.

Si përfundim, mund të themi se rezultatet e studimit ndihmojnë për të kuptuar se si graf rrjetat e telefonisë fikse vetorganizohen dhe evoluojnë. Ato nuk mund të konsiderohen graf rrjeta rasti. Mbetet për tu hulumtuar më tej se cilat modele strukturore janë më të favorshme për ti gjeneruar ato, duke patur parasysh këto karakteristika të nxjerra nga studimi.

REFERENCA

- [1] Albert, R., Barabási, A.-L., "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, pp. 47-97, 2002.
- [2] "www.oxforddictionaries.com," Oxford, [Online]. Available: <http://www.oxforddictionaries.com/definition/english/network>.
- [3] Erdős, P., Rényi, A., "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci.*, vol. 5, pp. 17-61, 1960.
- [4] Erdős, P., Rényi, A., "On the strength of connectedness of random graphs," *Acta Math. Acad. Sci. Hungar.*, vol. 12, pp. 261-267, 1961.
- [5] G. Caldarelli and A. Vespignani, Eds., *Large Scale Structure and Dynamics of Complex Networks- From Information Technology to Finance and Natural Science*, vol. 2, Singapore: World Scientific Publishing Co. Pte. Ltd, 2007.
- [6] Barrat, A., Barthélemy, M., Vespignani, A., *Dynamical Processes on Complex Networks*, New York: Cambridge University Press, 2008.
- [7] M.-E. Newman, *Networks: An Introduction*, New York: Oxford University Press Inc., 2010.
- [8] "Compare Home Phone Service," Cellulardeals, [Online]. Available: <http://www.cellulardeals.com/compare-home-phone-service>. [Accessed 2015].
- [9] Tyagi, A., Duhan, M., Bhatia, D., "Impact of Communication Technology on Human Brain Activity: Mobile Phone vs. Landline Phone," *Int. J. Medical Engineering and Informatics*, vol. 5, no. 1, pp. 20-30, 2013.
- [10] Ye, Q., Zhu, T., Hu, D., Wu, B., Du, N., Wang, B., "Cell Phone Mini Challenge Award: Social Network Accuracy - Exploring Temporal Communication in Mobile Call Graphs," in *IEEE Symposium on Visual Analytics Science and Technology*, Columbus, Ohio, USA, 2008.
- [11] Cortes, C., Pregibon, D., Volinsky, C., "Communities of Interest," in *Advances in Intelligent Data Analysis.*, Berlin, Heidelberg, 2001.
- [12] Nanavati, A.-A., Singh, R., Chakraborty, D., Dasgupta, K., Mukherjea, S., Das, G., Gurumurthy, S., Joshi, A., "Analyzing the Structure and Evolution of Massive Telecom Graphs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 5, pp. 703-718, 2008.
- [13] Holland, P.-W., Leinhardt, S., "A Dynamic Model for Social Networks," *Journal of Mathematical Sociology*, vol. 5, no. 1, pp. 5-20, 1977.
- [14] Runger, G., Wasserman, S., "Longitudinal Analysis of Friendship Networks," *Social Networks*, vol. 2, no. 2, pp. 143-154, 1980.
- [15] Wasserman, S., Iacobucci, D., "Statistical Analysis of Discrete Relational Data," *British Journal of Mathematical and Statistical Psychology*, vol. 39, no. 1, pp. 41-46, 1986.

- [16] Wasserman, S., Iacobucci, D., "Sequential Social Network Data," *Psychometrika*, vol. 53, no. 2, pp. 261-282, 1988.
- [17] Iacobucci, D., Wasserman, S., "A General Framework for Statistical Analysis of Sequential Dyadic Interaction Data," *Psychological Bulletin*, vol. 103, no. 3, p. 379, 1988.
- [18] Nanavati, A.-A., Gurumurthy, G.-D., Das, G., Chakraborty, D., Dasgupta, K., Mukherjee, S., Joshi, A., "On the Structural Properties of Massive Telecom Call Graphs: Finding and Implications," in *CIKM'06*, Arlington, Virginia, USA, 2006.
- [19] Seshadri, M., Machiraju, S., Sridharan, A., Bolot, J., Faloutsos, Ch., Leskovec, J., "Mobile Call Graphs: Beyond Power-Law and Lognormal Distributions," in *KDD'08*, Las Vegas, Nevada, USA, 2008.
- [20] Dong, Zh.-B., Song, G.-J., Xie, K.-Q., Wang, J.-Y., "An Experimental Study of Large - Scale Mobile Social Network," in *WWW 2009*, Madrid, Spain, 2009.
- [21] Noka (Jani), E., Hoxha, F., "Comparative Analysis of the Structural and Weighted Properties in Albanian Social Networks," *Journal of Multidisciplinary Engineering Science and Technology (JMEST)*, vol. 3, no. 4, pp. 4505-4509, 2016.
- [22] Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., Barabási, A.-L., "Structure and Tie Strengths in Mobile Communication Networks," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 104, no. 18, pp. 7332-7336, 2007.
- [23] Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Argollo de Mendez, M., Kaski, K., Barabási, A.-L., Kertész, J., "Analysis of a large - scale weighted network of one-to-one human communication," ArXiv, 2007.
- [24] Aiello, W., Chung, F., Lu, L., "A random graph model for massive graphs," in *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, New York, 2000.
- [25] Aiello, W., Chung, F., Lu, L., "A random graph model for power law graphs," *Experimental Mathematics*, vol. 10, no. 1, pp. 53-66, 2001.
- [26] Aksu, H., Korpeoglu, I., Ulusoy, O., "An Analysis of Social Networks based on Tera-scale Telecommunication Datasets," *IEEE Transactions on Emerging Topics in Computing*, vol. PP, no. 99, 2016.
- [27] G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*, Martino Fine Books, 2012 Reprint of 1949 Edition.
- [28] Abello, J., Buchsbaum, A.- L., Westbrook, J.-R., "A Functional Approach to External Graph Algorithms," in *Proceedings of the 6th European Symposium on ALgorithms*, Venice, Italy, 1998.
- [29] Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjee, S., Nanavati, A.- A., Joshi, A., "Social ties and their relevance to churn in mobile telecom networks," in *Extending database technology: Advances in database technology, EDBT '08*, New York, NY, USA, 2008.
- [30] T. Euler, "Churn prediction in telecommunication using Miningmart," in *Proceedings of the Workshop on Data Mining and Business (DMBiz)*, 2005.

- [31] Vaz de Melo, P.-O.-S., Akoglu, L., Faloutsos, C., Loureiro, A.-A.-F. , "Surprising Patterns for the Call Duration Distribution of Mobile Phone Users," in *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2010.*, Berlin, Heidelberg, 2010.
- [32] Doran, D., Mendiratta, V., Phadke, Ch., Uzunalioglu, H., "the Importance of Outlying Relationships in Mobile Call Graphs," in *11th International Conference on Machine Learning and Applications*, Boca Raton, Florida, USA, 2012.
- [33] Sridharan, A., Bolot, J., "Location Patterns of Mobile Users: a Large - Scale Study," in *The 32nd IEEE International Conference on Computer Communications*, Turin, Italy, 2013.
- [34] Dong, Y., Tang, J., Lou, T., Wu, B., Chawla, N. - V., "How Long Will She Call Me? Distribution, Social theory and Duration Prediction," in *ECML PKDD 2013 Proceedings, Part II, of the European Conference on Machine Learning and Knowledge Discovery in Databases*, Prague, Czech Republic, 2013.
- [35] Blondel, V.- D., Decuyper, A., Krings, G., "A survey of results on mobile phone datasets analysis," *EPJ Data Science*, vol. 4, no. 10, 2015.
- [36] Sarmiento, R., Oliveira, M., Cordeiro, M., Tabassum, S., Gama J., "Social Network Analysis in Streaming Call Graphs," in *Big Data Analysis: New Algorithms for a New Society. Studies in Big Data*, vol. 16, N. S. J. Japkowicz, Ed., Cham, Springer, 2016, pp. 239-261.
- [37] O. Gjermëni, "Temporal Statistical Analysis of Degree Distributions in an Undirected Landline Phone Call Network Graph Series," 2017. [Online]. Available: https://www.researchgate.net/publication/318649255_Temporal_Statistical_Analysis_of_Degree_Distributions_in_an_Undirected_Landline_Phone_Call_Network_Graph_Series.
- [38] O. Gjermëni, "Survival Analysis Data Set," 2017. [Online]. Available: https://www.researchgate.net/publication/320760614_Survival_Analysis_Data_Set.
- [39] Lambiotte, R., Blondel, V.-D., de Kerchove, C., Huens, E., Prieur, E., Smoreda, Z., Van Dooren, P., "Geographical dispersal of mobile communication networks," *Physica A: Statistical Mechanics and its Applications* , vol. 387, no. 21, pp. 5317-5325, 2008.
- [40] Kovanen, L., Saramäki, J., Kaski, K., "Reciprocity of mobile phone calls," *Dynamics of Socio-Economic Systems*, vol. 2, no. 2, pp. 138-151, 2011.
- [41] Gjermëni, O., Ramosaço, M., Zotaj, D., "Power - law versus Lognormal Distribution in a Phone Call Network Graph," in *International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE)*, Sofia, 2015.
- [42] O. Gjermëni, "Temporal Statistical Analysis of Degree Distributions in an Undirected Landline Phone Call Network Graph Series," *Data*, 2017.
- [43] M. Mitzenmacher, "Editorial: The Future of Power Law Research," *Internet Mathematics*, vol. 2, no. 4, pp. 525-534, 2005.
- [44] M. Newman, "The physics of networks," *Physics Today*, vol. 61, no. 11, pp. 33-38, 2008.
- [45] Watts, D.- J., Strogatz, S.-H., "Collective Dynamics of 'Small-World' Networks," *Nature*, vol. 393, pp. 440-442, 1998.
- [46] B. Bollöbas, *Random Graphs*, Cambridge: Cambridge University Press, 2001.
- [47] S. Milgram, "The Small - World Problem," *Psychology Today*, vol. 1, no. 1, pp. 61-67, 1967.

- [48] Travers, J., Milgram, S., "An Experimental Study of the Small World Problem," *Sociometry*, vol. 32, no. 4, pp. 425-443, 1969.
- [49] J. Guare, *Six Degrees of Separation: A Play*, New York: Vintage Books, 1990.
- [50] Gjermëni, O., Ramosaço, M., "From Small World Phenomenon to Correlation Analysis in a Temporal Landline Phone Call Network Graph Series," *International Journal of Applied Physics and Mathematics*, vol. 7, no. 4, pp. 275-283, 2017.
- [51] O. Gjermëni, "Small World Phenomenon in a Landline Phone Call Network Graph," in *The 4th International Scientific Conference "Scientific Challenges for Sustainable Development"*, Struga, 2017.
- [52] E. D. Kolaczyk, "Descriptive Analysis of Network Graph Characteristics," in *Statistical Analysis of Network Data*, New York, Springer Science+Business Media, LLC, 2009, pp. 79-122.
- [53] Latora, V., Marchiori, M., "Efficient Behavior of Small-World Networks," *Physical Review Letters*, vol. 87, no. 19, pp. 198701-1 - 198701-4, 2001.
- [54] Telesford, Q.-K., Joyce, K.-E., Hayasaka, S., Burdette, J.-H., Laurienti, P.-J., "The Ubiquity of Small-World Networks," *BRAIN CONNECTIVITY*, vol. 1, no. 5, pp. 367-375, 2011.
- [55] Humphries, M.- D., Gurney, K., "Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence," *PLoS ONE*, vol. 3, no. 4, p. e0002051, 2008.
- [56] Humphries, M.- D., Gurney, K., Prescott, T. -J., "The brain reticular formation is a small - world, not scale - free, network," *Proc. R. Soc. B*, vol. 273, p. 503–511, 2006.
- [57] Wohlgemuth, J., Matache, M.-T., "Small-World Properties of Facebook Group," *Complex Systems*, vol. 23, pp. 197-225, 2014.
- [58] Backstrom, L., Boldi, P., Rosa, M., Ugander, J., Vigna, S., "Four Degrees of Separation," in *WebSci 2012*, Evanston, Illinois, USA, 2012.
- [59] Montoya, J.-M., Sole, R.-V., "Small World Patterns in Food Webs," *J. theor. Biol.*, vol. 214, no. 3, pp. 405-412, 2004.
- [60] Bassett, D.-S., Bullmore, E., "Small-World Brain Networks," *The Neuroscientist*, vol. 12, no. 6, pp. 512-523, 2006.
- [61] Tomassini, M., Giacobini, M., Darabos, C., "Evolution and Dynamics of Small-World Cellular Automata," *Complex Systems*, vol. 15, no. 4, pp. 261-284, 2005.
- [62] de-Marcos, L., Garcia-Lopez, E., Garcia-Cabot, A., Medina, J.-A., Domínguez, A., Martínez-Herráiz, J.-J., Díez-Folledo, T., "Social network analysis of a gamified e-learning course: Small-world phenomenon and network metrics as predictors of academic performance," *Computers in Human Behavior*, vol. 60, pp. 312-321, 2016.
- [63] G. Csardi, "igraphdata: A Collection of Network Data Sets for the 'igraph' Package," 2015. [Online]. Available: <https://CRAN.R-project.org/package=igraphdata>.
- [64] Csardi, G., Nepusz, T., "The igraph software package for complex network research," *InterJournal*, vol. Complex Systems, p. 1695, 2006.
- [65] A.-I. McLeod, *Kendall: Kendall rank correlation and Mann-Kendall trend test. R package version 2.2*, 2011.

- [66] A. Kassambara, *ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.1.2.*, 2017.
- [67] R. C. Team, "R: A Language and Environment for Statistical Computing". Vienna, Austria 21 June 2016.
- [68] M.-E.-J. Newman, "The Structure and Function of Complex Networks," *SIAM Review*, vol. 45, no. 2, pp. 167-256, 2003.
- [69] Wasserman S., Faust K., *Social Network Analysis: Methods and Applications*, Cambridge Univeristy Press, 1994.
- [70] M.-E.-J. Newman, "The Structure and Function of Complex Networks," *SIAM Review*, vol. 45, no. 2, pp. 167-256, 2003.
- [71] Barrat, A., Weigt, M., "On the properties of small-world networks," *Eur. Phys. J. B* , vol. 13, pp. 547-560, 2000.
- [72] B. Efron, "Computers and Theory of Statistics: Thinking the Unthinkable," *SIAM Review*, vol. 21, pp. 460-480, 1979.
- [73] Sprent, P., Smeeton, N.- C., "Correlation and Concordance," in *Applied Nonparametric Statistical Methods*, Chapman&Hall/CRC, 2001, pp. 245-278.
- [74] Hollander, M., Wolfe, D. -A., "Kendall and Spearman tests," in *Nonparametric Statistical Methods*, New York, John Wiley & Sons, 1973, p. 185–194.
- [75] H. Abdi, "The Kendall rank correlation coefficient," in *Encyclopedia of Measurement and Statistics*, N. Salkind, Ed., Sage, 2007, pp. 508-510.
- [76] P. Royston, "An extension of Shapiro and Wilk's W test for normality to large samples," *Applied Statistics*, vol. 31, no. 2, pp. 115-124, 1982.
- [77] P. Royston, "Algorithm AS 181: The W test for Normality," *Applied Statistics*, vol. 31, no. 2, p. 176–180, 1982.
- [78] P. Royston, "Remark AS R94: A remark on Algorithm AS 181: The W test for normality," *Applied Statistics*, vol. 44, no. 4, p. 547–551, 1995.
- [79] Razali, N.-M, Wah, Y.-B., "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests," *Journal of Statistical Modeling and Analysis*, vol. 2, no. 1, pp. 21-33, 2011.
- [80] Graybill, F.-A., Iyer, K., *Regression Analysis: Concepts and Applications*, Duxbury Press, 1994.
- [81] M. Naqo, *Statistika Matematike*, Tiranë: SHBLU, 2005.
- [82] Anderson, D.-R., Sweeney, D.-J., Williams, Th.-A., *Statistics for Business and Economics*, USA: Thomson South-Western, 2008.
- [83] Breusch, T.-S., Pagan, A.-R., "A simple test for heteroscedasticity and random coefficient variation," *Econometrica* , vol. 47, p. 1287–1294, 1979.
- [84] Cook, R.-D., Weisberg, S., "Diagnostics for heteroscedasticity in regression," *Biometrika* , vol. 70, pp. 1-10, 1983.

- [85] Durbin, J., Watson, G.-S., "Testing for Serial Correlation in Least Squares Regression I," *Biometrika*, vol. 37, pp. 409-428, 1950.
- [86] Durbin, J., Watson, G.-S., "Testing for Serial Correlation in Least Squares Regression II," *Biometrika*, vol. 38, pp. 159-178, 1951.
- [87] Durbin, J., Watson, G.-S., "Testing for Serial Correlation in Least Squares Regression III," *Biometrika*, vol. 58, pp. 1-19, 1971.
- [88] "www.real-statistics.com/statistics-tables/durbin-watson-table/," [Online]. Available: <http://www.real-statistics.com/statistics-tables/durbin-watson-table/>.
- [89] Leskovec, J., Horvitz, E., "Planetary-Scale Views on a Large Instant-Messaging Network," in *WWW 2008*, Beijing, China, 2008.
- [90] Kleinbaum, D.-G., Klein, M., *Survival Analysis: A Self - Learning Text*, 3 ed., New York: Springer, 2012.
- [91] O. Gjermëni, "Survival Analysis in a Land - Line Phone Call Duration Data," *International Bulletin of Mathematical Research*, vol. 4, no. 3, pp. 35-44, September 2017.
- [92] Giolo, S.-R., Krieger, J.-E., Mansur, A.-J., Pereira, A.-C., "Survival Analysis of Patients with Heart Failure: Implications of Time-Varying Regression Effects in Modeling Mortality," *PLoS ONE*, vol. 7, no. 6, p. e37392, 2012.
- [93] Ahmad, T., Munir, A., Bhatti, SH., Aftab, M., Raza, M.-A., "Survival analysis of heart failure patients: A case study," *PLoS ONE*, vol. 12, no. 7, p. e0181001, 2017.
- [94] E.-K. Laitinen, "Survival Analysis and Financial Distress Prediction: Finnish Evidence," *Review of Accounting and Finance*, vol. 4, no. 4, pp. 76-90, 2005.
- [95] Mokarami, M., Motefares, Z., "A study of the relationship between corporate governance features," *European Online Journal of Natural and Social Sciences and bankruptcy by using survival analysis*, vol. 2, no. 3, pp. 881-887, 2013.
- [96] Zhang, G., Chen, Y., "An Intergrated Data Mining and Survival Analysis Model for Customer Segmentation," in *IFIP — The International Federation for Information Processing*, Boston MA, 2007.
- [97] D.-R. Cox, "Regression models and life tables," *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 34, no. 2, pp. 187-220, 1972.
- [98] Kaplan, E.-L., Meier, P., "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, vol. 53, no. 282, pp. 457-481, 1958.
- [99] M. Mills, *Introducing Survival and Event History*, Los Angeles, London, New Delhi, Singapore: Sage Publications, 2011.
- [100] Spruance, S.-L., Reid, J.-E., Grace, M., Samore, M., "Hazard Ratio in Clinical Trials," *ANTIMICROBIAL AGENTS AND CHEMOTHERAPY*, vol. 48, no. 8, p. 2787-2792, 2004.
- [101] Grambsch, P., Therneau, T. , "Proportional hazards tests and diagnostics based on weighted residuals," *Biometrika*, vol. 81, pp. 515-526, 1994.
- [102] Venables, W.-N., Ripley, B.-D., *Modern Applied Statistics with S*, 4th ed., New York: Springer, 2002.

- [103] Choi, I., Wells, B.-J., Yu, Ch., Kattan, M.-W. , "An empirical approach to model selection through validation for censored survival data," *Journal of Biomedical Informatics*, vol. 44, p. 595–606, 2011.
- [104] Therneau, T.- M., Grambsch, P.- M. , *Modeling Survival Data: Extending the Cox Model*, New York: Springer, 2000.
- [105] T. Therneau, "A Package for Survival Analysis in S, version 2.38," 2015. [Online]. Available: <https://CRAN.R-project.org/package=survival>.
- [106] Original by Klein, Moeschberger and modifications by J. Yan , " KMsurv: Data sets from Klein and Moeschberger (1997), Survival Analysis. R package version 0.1-5," 2012. [Online]. Available: <https://CRAN.R-project.org/package=KMsurv>.
- [107] Hothorn, T., Hornik, K., van de Wiel, M.- A., Zeileis, A., " A Lego System for Conditional Inference," *The American*, vol. 60, no. 3, pp. 257-263, 2006.
- [108] Hothorn, T., Hornik, K., van de Wiel, M.- A., Zeileis, A., "Implementing a Class of Permutation Tests: The coin Package," *Journal of Statistical Software*, vol. 28, no. 8, pp. 1-23, 2008.
- [109] D.-M. Diez, "OIsurv: Survival analysis supplement to OpenIntro guide. R package version 0.2," 2013. [Online]. Available: <https://CRAN.R-project.org/package=OIsurv>.
- [110] Kuhn, M. et al., "caret: Classification and Regression Training. R package version 6.0-76," 2017. [Online]. Available: <https://CRAN.R-project.org/package=caret>.
- [111] Barabási, A.-L., Albert, R., "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509-512, 1999.
- [112] M. Mitzenmacher, "A Brief History of Generative Models for Power Law and Lognormal," *Internet Mathematics*, vol. 1, no. 2, pp. 226-251, 2003.
- [113] O. Gjermëni, "Descriptive Analysis of Characteristics: A Case Study of a Phone Call Network Graph," in *The 5-th UBT International Conference on Computer Science and Communication Engineering*, Durrës, Albania, 2016.
- [114] Gjermëni, O., Zotaj, D., "A Comparative Study on Four Phone Call Networks," in *The 6-th International Conference on Information systems and Technology Innovations: inducing Modern Business Solution*, Tiranë, Albania, 2015.
- [115] C.-S. Gillespie, "Fitting Heavy Tailed Distributions: The powerLaw Package," *Journal of Statistical Software*, vol. 64, no. 2, pp. 1-16, 2015.
- [116] Rmetrics Core Team, Wuertz, D., Setz, T., Chalabi, Y., "fBasics: Rmetrics - Markets and Basic Statistics," R package 3011.87, 2014.
- [117] M.-G. Bulmer, *Principles of Statistics*, New York: Dover Publications, 1979.
- [118] Balanda, K.-P., MacGillivray, H.-L., "Kurtosis: A Critical Review," *The American Statistician*, vol. 42, no. 2, pp. 11-119, 1988.
- [119] Hanusz, Z., Tarasińska, J., "Impact of Alternative Distributions on Quantile - Quantile Normality Plot," *Colloquium Biometricum*, vol. 45, pp. 67-78, 2015.
- [120] Foss, S., Korshunov, D., Zachary, S., *An Introduction to Heavy-Tailed and Subexponential Distributions*, New York: Springer Science+Business Media, 2013.

- [121] Clauset, A., Shalizi, C.-R., Newman, M.-E.-J., "Power - Law Distributions in Empirical Data," *SIAM REVIEW*, vol. 51, no. 4, pp. 661-703, 2009.
- [122] Clauset, A., Shalizi, C.-R., Newman, M.-E.-J., "Power - Law Distribution in Empirical Data," www.santafe.edu, 2007.
- [123] J. Shim, "Toward a more nuanced understanding of long-tail," *Journal of Business Venturing Insights*, vol. 6, pp. 21-27, 2016.
- [124] O. Gjermëni, "Power Law Distribution as a Component of the Vertex Degree distribution on a Social University Network Course," *European Scientific Journal*, vol. 11, no. 20, pp. 129-137, 2015.
- [125] Goldstein, M.-L., Morris, S.- A., Yen, G.-G., "Problems with fitting to the power - law distribution," *The European Physical Journal B*, vol. 41, pp. 255-258, 2004.
- [126] M.-E.-J. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemporary Physics*, vol. 46, no. 5, pp. 323-351, 2005.
- [127] Press, W.-H., Teukolsky, S.-A., Vetterling, W.-T., Flannery, B.-P, *Numerical Recipes in C: The Art of Scientific computing*, vol. 2nd edition, Cambridge, England: Cambridge University Press, 1992.
- [128] B.-M. Hill, "A Simple General Approach to Inference about the Tail of a Distribution," *The Annals of Statistics*, vol. 3, no. 5, pp. 1163-1174, 1975.
- [129] A. Walther, "Anschauliches zur Riemannischen Zetafunktion," *Acta Mathematica*, vol. 48, no. (3-4), pp. 393-400, 1926.
- [130] Limpert, E., Stahel, W.-A., Abbt, M., "Log-normal distributions across the sciences: Keys and clues," *Bioscience*, vol. 51, pp. 341-352, 2001.
- [131] Clauset, A., Young, M., Gleditsch, K.-S., "On the frequency of severe terrorist events," *J. Conflict Resolution*, vol. 51, pp. 58-87, 2007.
- [132] Efron, B., Tibshirani, R.-J., *An Introduction to the Bootstrap*, New York: Chapman and Hall, 1992.
- [133] Q.-H. Vuong, "Likelihood ratio tests for model selection and non - nested hypothesis," *Econometrica*, vol. 57, pp. 307-333, 1989.
- [134] Caldarelli, G., Vespignani, A., "Preliminaries and Basic Definitions in Network Theory," in *Large Scale Structure and Dynamics of Complex Networks: From Information Technology to Finance and Natural Science*, G. Caldarelli and A. Vespignani, Eds., Singapore, World Scientific Publishing Co. Pte. Ltd., 2007, pp. 5-16.
- [135] Newman, M.-E.-J., Forrest, S., Balthrop, J., "Email networks and the spread of computer viruses," *Phys. Rev. E* , vol. 66, p. 035101, 2002.
- [136] Doyle, J.-C., Alderson, D.-L., Li, L., Low, S., Roughan, M., Shalunov, S., Tanaka, R., "The "robust yet fragile" nature of the Internet," *PNAS*, vol. 102, no. 41, pp. 14497-14502, 2005.
- [137] M.-E.-J. Newman, "Mixing patterns in networks," *Physical Review E*, vol. 67, no. 2, p. 026126, 4 Feb 2003.
- [138] Garlaschelli, D., Loffredo, M.-I., "Patterns of link reciprocity in directed networks," *Physical Review Letters*, vol. 93, no. 26, p. 268701, 2004.

- [139] Hidalgo, A.-A., Rodriguez-Sickert, C., "The dynamics of a mobile phone network," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 12, pp. 3017-3024, 2008.
- [140] Kolaczyk, E.-D., Csárdi, G., *Statistical Analysis of Network Data with R*, New York: Springer, 2014.
- [141] E. Noka (Jani), *Analiza Matematikore e Rrjetave Sociale: Struktura dhe Vetitë Teorike të Rrjetave Sociale Shqiptare*, Tiranë: Universiteti i Tiranës, Fakulteti i Shkencave të Natyrës, 2016.