



ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ στα
ΠΟΛΥΠΛΟΚΑ ΣΥΣΤΗΜΑΤΑ και ΔΙΚΤΥΑ

ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΤΜΗΜΑ ΓΕΩΛΟΓΙΑΣ

ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ



ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Τίτλος Εργασίας

Στατιστική ανάλυση στο δίκτυο συναλλαγών του Bitcoin

Statistical analysis to Bitcoin transactions network

Καλαμπάκας Αργύριος

ΕΠΙΒΛΕΠΩΝ: Ιωάννης Ε. Αντωνίου, Καθηγητής Α.Π.Θ.

ΣΥΝΕΠΙΒΛΕΠΩΝ: Δρ. Γεώργιος Χρ. Μακρής

Θεσσαλονίκη, Οκτώβριος 2019





ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ στα
ΠΟΛΥΠΛΟΚΑ ΣΥΣΤΗΜΑΤΑ και ΔΙΚΤΥΑ

ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΤΜΗΜΑ ΓΕΩΛΟΓΙΑΣ

ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ



ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Τίτλος Εργασίας

Στατιστική ανάλυση στο δίκτυο συναλλαγών του Bitcoin

Statistical analysis to Bitcoin transactions network

Καλαμπάκας Αργύριος

ΕΠΙΒΛΕΠΩΝ: Ιωάννης Ε. Αντωνίου, Καθηγητής Α.Π.Θ.

ΣΥΝΕΠΙΒΛΕΠΩΝ: Δρ. Γεώργιος Χρ. Μακρής

Εγκρίθηκε από την Τριμελή Εξεταστική Επιτροπή την 25η Οκτωβρίου 2019.

.....
Ιωάννης Ε. Αντωνίου
Καθηγητής Α.Π.Θ.

.....
Δρ. Γεώργιος Χρ. Μακρής

.....
Χαράλαμπος Μπράτσας
ΕΔΙΠ Α.Π.Θ.

Θεσσαλονίκη, Οκτώβριος 2019



.....
Καλαμπάκας Αργύριος
Πτυχιούχος Πληροφορικής Α.Π.Θ.

Copyright © Καλαμπάκας Αργύριος, 2019
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευτεί ότι εκφράζουν τις επίσημες θέσεις του Α.Π.Θ.

ΠΕΡΙΛΗΨΗ

Το θέμα που πραγματεύεται η παρούσα εργασία είναι η ανάλυση δεδομένων σε ένα νεοεμφανιζόμενο σύνολο δεδομένων στην ερευνητική κοινότητα. Ιδιαίτερης ποιότητας καθώς εμπεριέχει οικονομικές συναλλαγές οι οποίες αυξήθηκαν ραγδαία, μαζί με το ενδιαφέρον της κοινωνίας, τον Δεκέμβριο του 2017 εξαιτίας της μεγάλης αύξησης της τιμής του Bitcoin.

Η μέθοδος που ακολουθείται χωρίζεται σε δύο σκέλη, στην στατιστική ανάλυση δικτύων και στην ανάλυση οικονομικής δραστηριότητας. Η στατιστική ανάλυση δικτύων είναι το εργαλείο για την ανάλυση των πολύπλοκων συστημάτων της σύγχρονης εποχής τα οποία αναπαριστώνται ως δίκτυα, είναι αδύνατον να αναλυθούν τμηματικός καθώς οδηγούν σε λανθασμένα συμπεράσματα. Η ανάλυση οικονομικής δραστηριότητας αφορά δείκτες από τον χρηματιστηριακό τομέα και της επιστήμης των οικονομικών, οι οποίοι έχουν μεταφερθεί και αντιστοιχηθεί με το οικονομικό μοντέλο που αντιπροσωπεύει το Bitcoin. Και τα δύο σκέλη της μεθοδολογίας που ακολουθήθηκε είχαν ως προαπαιτούμενο την εξαγωγή των δεδομένων, την μετατροπή τους σε δίκτυο και τον τεμαχισμό τους.

Στα αποτελέσματα διακρίνεται η έκταση των δεδομένων που δημιουργήθηκαν από την στατιστική ανάλυση δικτύων και την ανάλυση οικονομικής δραστηριότητας. Την δεύτερου βαθμού στατιστική ανάλυση χρονικών σειρών που εφαρμόστηκε για να τροφοδοτήσει όσο το δυνατόν περισσότερους συνδυασμούς μοντέλων Τεχνητά Νευρωνικά Δίκτυα.

Συμπερασματικά παρουσιάζονται οι σημαντικές παραδοχές που προέκυψαν για την δομή και την εξέλιξη των ημερήσιων δικτύων συναλλαγών. Οι δυνατότητες και τα προβλήματα που παρουσιάστηκαν κατά την διεξαγωγή της έρευνας. Η καινοτόμα προσέγγιση για την δημιουργία ενός ευρύ μέτρου σύγκρισης μεταξύ των δεδομένων περιγραφικής στατιστικής και των δεδομένων οικονομικής δραστηριότητας από τα μοντέλα ΤΝΔ.

Η ολοκλήρωση της παρούσας εργασίας ήταν ένας αγώνας δρόμου καθώς υπήρχε ο φόβος, της κατάρρευσης της τιμής του Bitcoin, της συγκέντρωσης του 51% της ισχύς του Bitcoin, της κήρυξης του παράνομου. Όλα αυτά θα είχαν ως αποτέλεσμα την ραγδαία μείωση των χρηστών και των συναλλαγών.

ΛΕΞΕΙΣ-ΚΛΕΙΔΙΑ

Bitcoin, Blockchain, Δίκτυα συναλλαγών, Χρονικές σειρές, ΤΝΔ

ABSTRACT

The topic dealt with in the present work is the analysis of data in a newly emerging dataset in the research community. Of particular quality as it involves financial transactions that have grown rapidly, along with the interest of the public, in December 2017 due to the sharp rise in the price of Bitcoin.

The method followed is divided into two parts, statistical network analysis and economic activity analysis. Network statistical analysis is the tool for analyzing the complex systems of modern times which are represented as networks, it is impossible to analyze them in part as they lead to incorrect conclusions. The analysis of economic activity relates to indicators from the stock market and the economics of science, which have been transposed and matched with the economic model represented by Bitcoin. Both strands of the methodology followed included the extraction of data, their conversion to network and their fragmentation.

The results distinguish the extent of the data generated by the statistical analysis of the networks and the analysis of economic activity. The second-order statistical analysis of time series applied to feed as many combinations of Artificial Neural Network models as possible.

In summary, the important assumptions that have emerged for the structure and evolution of daily transactions networks are presented. The opportunities and problems encountered in conducting the research. The innovative approach to create a broad measure of comparison between descriptive statistics and economic activity data from ANN models.

Completion of this work was a race as there was fear, of the collapse of the Bitcoin price, of the 51% concentration of Bitcoin's power and its distortion, of being declared illegal by central banks around the world. All this would result in a sharp decline in users and transactions, that is to say, the data generated and ultimately the interest of the research community.

KEY WORDS

Bitcoin, Blockchain, Transactions Networks, Time Series, ANN

ΠΕΡΙΕΧΟΜΕΝΑ

SUMMARY.....	9
ΠΡΟΛΟΓΟΣ.....	13
ΣΧΕΤΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ.....	14
Κεφάλαιο	Σελίδα
1. ΤΑ ΚΡΥΠΤΟΝΟΜΙΣΜΑΤΑ ΩΣ ΠΗΓΕΣ ΔΕΔΟΜΕΝΩΝ.....	15
1.1 Τα πορτοφόλια του Bitcoin.....	16
1.2 Συναλλαγές με Bitcoin.....	16
1.3 Εξόρυξη Bitcoin.....	17
1.4 Ανταλλακτήρια Bitcoin, επένδυση και κερδοσκοπία.....	18
1.5 Γενικά χαρακτηριστικά για το δίκτυο συναλλαγών του Bitcoin.....	19
2. ΟΙ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ BLOCKCHAIN	20
2.1 Το σύνολο δεδομένων του Bitcoin blockchain.....	21
2.2 Εργαλεία εξόρυξης δεδομένων από το Bitcoin blockchain.....	23
2.3 Το εργαλείο εξόρυξης δεδομένων bitcoin-abe.....	25
2.4 Τα εργαλεία εξόρυξης δεδομένων blockchainToAvro και bitcoin-etl, η υπηρεσία Google BigQuery.....	25
2.5 Το εργαλείο εξόρυξης δεδομένων bitcoind-dump-tsv.....	26
3. ΑΝΑΛΥΣΗ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ TXEDGES.....	28

3.1	Συμπλήρωση δεδομένων και τεμαχισμός του δικτύου συναλλαγών.....	29
3.2	Εφαρμογή περιγραφικής στατιστικής στα ημερήσια δίκτυα συναλλαγών.....	30
3.3	Ανάλυση οικονομικής δραστηριότητας στα ημερήσια δίκτυα συναλλαγών.....	32
4.	ΠΡΟΒΛΕΨΗ ΤΗΣ ΤΙΜΗΣ ΤΟΥ BITCOIN ΜΕ ΤΝΔ.....	34
4.1	Διαχωρισμός των χρονικών σειρών.....	34
4.2	Η δομή των ΤΝΔ.....	35
5.	ΠΕΡΙΓΡΑΦΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ.....	37
5.1	Χρονικές σειρές περιγραφικής στατιστικής και οικονομικής δραστηριότητας.....	38
5.2	Στατιστική ανάλυση χρονικών σειρών.....	52
5.3	ΤΝΔ πρόβλεψης και ταξινόμησης.....	57
6.	ΣΥΜΠΕΡΑΣΜΑΤΑ.	58
6.1	Χρονικές σειρές περιγραφικής στατιστικής και οικονομικής δραστηριότητας.....	59
6.2	Στατιστική ανάλυση χρονικών σειρών.....	60
6.3	ΤΝΔ πρόβλεψης και ταξινόμησης.....	61
	ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ.....	61
	ΠΑΡΑΡΤΗΜΑ.....	65
	ΒΙΒΛΙΟΓΡΑΦΙΑ.....	73

SUMMARY

The purpose of the present work is not to explore another perspective on how the structure and function of Bitcoin and Blockchain are presented. It aims to extend data analysis in the field of networks by applying statistical analysis of networks to the Bitcoin transactions network. It also aims to compare the potentials of the results of network statistical analysis, through forecasting and grading ANN, against known economic activity indicators that are also applied to the Bitcoin transactions network.

Extracting data and processing it to come in an analytical form is itself demanding and important, but the importance of the work lies in the very data that will be analyzed and the resulting ones. The interest shown in Bitcoin blockchain data stems from the technology itself where it is stored. The Bitcoin blockchain offers a very large amount of data publicly, structured and free. This data is of qualitative importance as it is a matter of financial transactions, such data has not until now been given public, structured and free as it was considered a corporate asset mainly by banks but also by various other companies such as online stores or online casinos.

The work assumes that the reader is aware of the terminology and mode of operation of Bitcoin and Blockchain. It is one of the few researches in both Greek and international level that applies statistical analysis of networks to the Bitcoin transactions network and tries to predict the price of Bitcoin in 2018 using ANN. The multitude of researches use data prior to the Bitcoin price drop and its volume of transactions, which are stable in their evolution. Developments and research on Bitcoin and cryptocurrencies are rapid.

Due to the high demand for cryptocurrencies, more than 190 cryptocurrency exchanges have been created. This rapid growth of cryptocurrencies has created a very large number of transactions, namely data that is now publicly available for analysis due to Blockchain technology. The first chapter presents the research on the extent to which the use of cryptocurrencies has been compared with the current banking system, the funds corresponding to cryptocurrencies, and the two main sectors driving the cryptocurrency economy, the mining companies and electronic crypto exchanges.

The advent of Blockchain has renewed the perspective and potential of data analysis. Blockchain provides data that complies with two basic requirements of data analysis. The first

requirement is data security, the second requirement is their structured, abundant and complete state. The second chapter presents the Blockchain database, the Bitcoin blockchain data and how it was extracted.

It was chosen to do the analysis with the programming language R. Due to the large number of links it was decided not to analyze and represent the transactions network as a single one, it was doubtful whether conclusions could be drawn due to the heterogeneity in the evolution of the network, as well as too many computing requirements will limit the capabilities of the iGraph library to be used for statistical network analysis. It was finally decided to split the transactions network into daily transactions networks, where statistical analysis for each daily transactions network would be performed. The results of the statistical analysis for each daily transactions network will be combined with the corresponding calendar order and a time series of results will be created. The third chapter presents the creation of the Bitcoin transactions network, the application of statistical analysis to the Bitcoin transactions network, and the analysis of economic activity in the Bitcoin transactions network.

The analysis yielded 182 time series with 2763 observations each coming from daily transactions networks, gigantic components and ER random networks. In the fourth stage of the analysis, an attempt will be made to predict the price of Bitcoin using the 182 time series in various combinations of artificial neural networks. The fourth chapter presents the datasets to train the ANN and the ANN models to be tested.

The fifth chapter will present the most important results of the analyzes of descriptive statistics and economic activity performed on daily transactions networks, ER random networks and giant components, which produced the initial time series. It will empirically test adaptation to power distribution, compare the evolution of daily transactions networks with random ERs, investigate the small world phenomenon, peak similarity, and time series correlation with Bitcoin prices. The results of the statistical analysis done in the initial time series, which produced the highly correlated time series, the time series with zero or near zero variation, and the PCA of the time series, will then be presented. Closing will present the results with the highest value of successful prediction and classification for the three categories of data given to the 288 ANN that ran in total. Two conclusions are critical and



ΠΡΟΛΟΓΟΣ

Η ανάπτυξη των κρυπτονομισμάτων ταυτίζεται με την ανάπτυξη του ίντερνετ που έγινε σε τρία στάδια. Η πρώτη γενιά κρυπτονομισμάτων, η οποία ξεκίνησε με το Bitcoin, βασίστηκε στην ιδέα ότι υπάρχει η δυνατότητα για ένα νομισματικό σύστημα όπου δεν θα χρειάζεται ο ενδιάμεσος ρόλος των τραπεζών γιατί οι συναλλαγές θα γίνονται απευθείας από τους συμμετέχοντες. Με παρόμοιο τρόπο το ίντερνετ στο πρώτο στάδιο του είχε ως σκοπό να αποκεντρώσει τις διαθέσιμες πληροφορίες χωρίς να χρειάζεται κάποιος μεσάζοντας. Η δεύτερη γενιά κρυπτονομισμάτων ήρθε, με το Ethereum, για να καλύψει την μέχρι τότε ανικανότητα του Blockchain να διαχειρίζεται ψηφιακό περιεχόμενο. Αντίστοιχα το ίντερνετ στο δεύτερο στάδιο του εξελίχθηκε από το στατικό περιεχόμενο ενσωματώνοντας την δημιουργία περιεχομένου και την συνεργασία σε πραγματικό χρόνο. Βρισκόμαστε πλέον στην ανάπτυξη της τρίτης γενιάς κρυπτονομισμάτων τα οποία επικεντρώνονται στην συνεργασία μεταξύ διαφορετικών κρυπτονομισμάτων. Όπως και στο ίντερνετ, στο τρίτο στάδιο του, γίνεται προσπάθεια να γίνει παγκοσμίως συνδεδεμένο, δομημένο, ανοιχτό και ελεύθερο.[1]

Η παρούσα εργασία δεν έχει ως στόχο να διερευνήσει ακόμα μία οπτική στον τρόπο παρουσίασης της δομής και της λειτουργίας του Bitcoin και του Blockchain. Έχει ως στόχο να διευρύνει την ανάλυση δεδομένων στον τομέα των δικτύων, εφαρμόζοντας στατιστική ανάλυση δικτύων στο δίκτυο συναλλαγών του Bitcoin. Επίσης έχει ως στόχο να συγκρίνει τις δυνατότητες των αποτελεσμάτων της στατιστικής ανάλυσης δικτύων, μέσω ΤΝΔ πρόβλεψης και ταξινόμησης, σε σχέση με γνωστούς δείκτες οικονομικής δραστηριότητας που επίσης εφαρμόζονται στο δίκτυο συναλλαγών του Bitcoin.

Η εξαγωγή των δεδομένων και η επεξεργασία τους για να έρθουν σε μία αναλύσιμη μορφή είναι από μόνη της απαιτητική και σημαντική, αλλά η σημαντικότητα της εργασίας βρίσκεται στα ίδια τα δεδομένα που θα αναλυθούν και σε αυτά που θα προκύψουν. Το ενδιαφέρον που παρουσιάζεται σχετικά με τα δεδομένα του Bitcoin blockchain προκύπτει από την ίδια την τεχνολογία όπου είναι αποθηκευμένα. Το Bitcoin blockchain προσφέρει δημόσια, δομημένα και ελεύθερα ένα πολύ μεγάλο όγκο δεδομένων. Τα συγκεκριμένα δεδομένα έχουν ποιοτική σημασία καθώς πρόκειται για οικονομικές συναλλαγές, αντίστοιχα δεδομένα δεν υπήρχαν μέχρι σήμερα δοσμένα δημόσια, δομημένα και ελεύθερα καθώς

θεωρούνταν εταιρικό περιουσιακό στοιχείο κυρίως των τραπεζών αλλά και διάφορων άλλων εταιριών όπως τα ηλεκτρονικά καταστήματα ή τα ηλεκτρονικά καζίνο.

Η εργασία λαμβάνει ως δεδομένο ότι ο αναγνώστης έχει γνώση των ορολογιών και του τρόπου λειτουργίας του Bitcoin και του Blockchain. Αρχικά παρουσιάζεται συνοπτικά εγχώρια και διεθνή γνώση στο θέμα. Στο πρώτο κεφάλαιο παρουσιάζονται, η έρευνα που έγινε σχετικά με την έκταση που έχει πάρει η χρήση κρυπτονομισμάτων και γίνεται σύγκριση με το υφιστάμενο τραπεζικό σύστημα, τα κεφάλαια που αντιστοιχούν σε κρυπτονομίσματα και οι δύο κύριοι τομείς που κινούν την οικονομία των κρυπτονομισμάτων, η εταιρίες εξόρυξης και τα ηλεκτρονικά ανταλλακτήρια κρυπτονομισμάτων. Στο δεύτερο κεφάλαιο παρουσιάζονται, η βάση δεδομένων Blockchain, τα δεδομένα του Bitcoin blockchain και ο τρόπος με τον οποίο εξήχθησαν. Στο τρίτο κεφάλαιο παρουσιάζονται, η δημιουργία του δικτύου συναλλαγών του Bitcoin, η εφαρμογή της στατιστικής ανάλυσης στο δίκτυο συναλλαγών του Bitcoin και η ανάλυση οικονομικής δραστηριότητας στο δίκτυο συναλλαγών του Bitcoin. Στο τέταρτο κεφάλαιο παρουσιάζονται, τα σύνολα δεδομένων με τα οποία θα εκπαιδευτούν τα ΤΝΔ και τα μοντέλα των ΤΝΔ που θα δοκιμαστούν. Στο πέμπτο κεφάλαιο παρουσιάζονται, τα αποτελέσματα των χρονικών σειρών περιγραφικής στατιστικής και οικονομικής δραστηριότητας, τα αποτελέσματα της στατιστικής ανάλυσης στις χρονικές σειρές και τα αποτελέσματα των ΤΝΔ πρόβλεψης και ταξινόμησης. Στο έκτο κεφάλαιο παρουσιάζονται, τα συμπεράσματα για τα αποτελέσματα των χρονικών σειρών περιγραφικής στατιστικής και οικονομικής δραστηριότητας, τα συμπεράσματα για τα αποτελέσματα της στατιστικής ανάλυσης στις χρονικές σειρές και τα συμπεράσματα για τα αποτελέσματα των ΤΝΔ πρόβλεψης και ταξινόμησης. Τέλος ακολουθεί το κεφάλαιο με τις προτάσεις για μελλοντική έρευνα σε σχέση με ανοικτά προβλήματα.

ΣΧΕΤΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

Είναι μία από τις λίγες εργασίες σε ελληνικό αλλά και διεθνή επίπεδο που εφαρμόζει στατιστική ανάλυση δικτύων στο δίκτυο συναλλαγών του Bitcoin και προσπαθεί να προβλέψει την τιμή του Bitcoin στο έτος 2018 με χρήση ΤΝΔ. Η πληθώρα των εργασιών χρησιμοποιούν δεδομένα πριν την εκτίναξη της τιμής του Bitcoin και του όγκου των συναλλαγών του, τα οποία παρουσιάζουν σταθερότητα στην εξέλιξη τους. Αυτή είναι η αιτία

που οι πρώτες εργασίες παρουσίασαν αλγόριθμους απο-ανωνυμοποίησης και εντοπισμού χρηστών από τις συναλλαγές του Bitcoin αντί να προσπαθήσουν να δημιουργήσουν μοντέλα πρόβλεψης τιμής.[7, 8] Έτσι και οι περισσότερες εργασίες προσπαθούν να απο-ανωνυμοποιήσουν τις συναλλαγές, εντοπίζοντας τα δίκτυα χρηστών και τις κοινότητες χρηστών.[2, 6] Νεότερες εργασίες προσπαθούν να εντοπίσουν κερδοσκοπικές συναλλαγές οι οποίες δεν ήταν συχνό φαινόμενο αλλά και μοτίβα στην τοπολογία των δικτύων συναλλαγών και χρηστών που δημιουργούνται από υπηρεσίες ανωνυμοποίησης.[3] Όσες εργασίες εφαρμόζουν κάποιο είδος στατιστικής ανάλυσης στο δίκτυο συναλλαγών του Bitcoin, επικεντρώνονται σε μερικούς δείκτες όπως είναι οι βαθμοί των κόμβων.[4, 5] Υπάρχουν μερικές εργασίες με παρόμοιο τρόπο ανάλυσης της δομής του δικτύου συναλλαγών, μετρώντας τα χαρακτηριστικά του δικτύου, σε συνεχόμενα στιγμιότυπα με την πάροδο του χρόνου και συγκρίνοντας την εξέλιξη τους με την τιμή του Bitcoin.[10]

Οι εξελίξεις και η έρευνα σχετικά με το Bitcoin και τα κρυπτονομίσματα είναι ραγδαίες. Αυτό έχει ως αποτέλεσμα στην παρούσα εργασία να υπάρχουν αρκετές παραπομπές που προέρχονται από οικονομικές και ειδησεογραφικές ιστοσελίδες οι οποίες παρακολουθούν την επικαιρότητα και την έρευνα σχετικά με το Bitcoin και τα κρυπτονομίσματα.

1. ΤΑ ΚΡΥΠΤΟΝΟΜΙΣΜΑΤΑ ΩΣ ΠΗΓΕΣ ΔΕΔΟΜΕΝΩΝ

Το 2008 εμφανίστηκε το πρώτο κρυπτονόμισμα, το Bitcoin[10], ενώ μέχρι το πρώτο τρίμηνο του 2018 έχουν δημοσιευθεί και εισηχθεί σε ανταλλακτήρια κρυπτονομισμάτων πάνω από 1500 κρυπτονομίσματα. Η κεφαλαιοποίηση τους είναι περίπου 370 δις δολάρια, από τα οποία το 90% κατέχουν μόνο 20 κρυπτονομίσματα[11].

Επιπλέον υπάρχει μια πολύ μεγάλη κατηγορία κρυπτονομισμάτων, αυτά που εκδίδονται ως ICO. Στο δεύτερο τρίμηνο του 2018 ενεργά είναι 2000 από τα περίπου 4000 που έχουν δημοσιευθεί από τις αρχές του 2017. Έχουν μαζέψει συνολικά κεφάλαια ύψους περίπου 16 δις δολάρια[12].

Λόγω της μεγάλης ζήτησης κρυπτονομισμάτων που εμφανίστηκε έχουν δημιουργηθεί περισσότερα από 190 ανταλλακτήρια κρυπτονομισμάτων. Αυτή η ραγδαία αύξηση των κρυπτονομισμάτων έχει δημιουργήσει ένα πολύ μεγάλο αριθμό συναλλαγών, δηλαδή

δεδομένων που είναι πλέον δημόσια διαθέσιμα για ανάλυση εξαιτίας της τεχνολογίας Blockchain. Τα αντίστοιχα δεδομένα μέχρι σήμερα ήταν μόνο στην διάθεση των τραπεζών.

1.1 Τα πορτοφόλια του Bitcoin

Ο συνολικός αριθμός πορτοφολιών που υπάρχουν στο τέταρτο τρίμηνο του 2018 είναι περίπου 32 εκατομμύρια. Μόλις το 0.42% του παγκόσμιου πληθυσμού έχει στην κατοχή του πορτοφόλια Bitcoin, υπολογίζοντας ότι πολλά πορτοφόλια ανήκουν στο ίδιο πρόσωπο όπως και ότι αρκετά είναι μη προσβάσιμα, τότε το ποσοστό μειώνεται αρκετά.

Παρατηρείται ότι το τέταρτο τρίμηνο του 2017 υπήρξε ραγδαία αύξηση των πορτοφολιών το οποίο δικαιολογείται από την μεγάλη άνοδο της τιμής του Bitcoin. Η πτώση της τιμής του Bitcoin που ακολούθησε όμως φαίνεται ότι δεν απότρεψε την δημιουργία νέων πορτοφολιών.

Ιδιαίτερη περίπτωση αποτελούν τα πορτοφόλια που παρέχονται από τα ανταλλακτήρια. Κάθε ανταλλακτήριο έχει κάποια πορτοφόλια, τα πρόσωπα όμως που ανοίγουν λογαριασμό στο ανταλλακτήριο δεν αποκτούν δικό τους πορτοφόλι. Αντίθετα αποκτούν ένα λογιστικό υπόλοιπο στην διαδικτυακή πλατφόρμα του εκάστοτε ανταλλακτηρίου. Εύκολα συμπεραίνετε ότι υπάρχει ένας αριθμός προσώπων που κατέχουν Bitcoins αλλά όχι πορτοφόλια.

Είναι βέβαιο ότι το Bitcoin ακόμα δεν έχει υιοθετηθεί παγκοσμίως καθώς ο αριθμός των χρηστών του είναι αρκετά μικρότερος σε σύγκριση με τον αριθμό των χρηστών που έχουν άλλα συστήματα πληρωμών όπως VISA, MASTERCARD και PAYPAL. Συγκεκριμένα το PAYPAL εξυπηρετεί περίπου 250 εκατομμύρια χρήστες, η MASTERCARD εξυπηρετεί περίπου 640 εκατομμύρια χρήστες και η VISA εξυπηρετεί περισσότερους από 1 δις χρήστες παγκοσμίως.[13, 14]

1.2 Συναλλαγές με Bitcoin

Στο τρίτο τρίμηνο του 2018 το ημερήσιο ποσό συναλλαγών με Bitcoin ξεπέρασε τα 8 δις δολάρια, η τιμή αυτή πλησιάζει πολύ το ημερήσιο ποσό συναλλαγών της MASTERCARD που κυμαίνεται κοντά στα 11 δις δολάρια. Την ίδια ώρα η VISA χειρίζεται περισσότερα από 30 δις δολάρια σε ημερήσιες συναλλαγές. Αποτελεί σημαντικό ορόσημο για το Bitcoin η

προσέγγιση του ημερήσιου ποσού συναλλαγών ενός κορυφαίου δικτύου πιστωτικών καρτών.[15]

Επίσης πολλές συναλλαγές στο δίκτυο του Bitcoin δημιουργούνται από υπηρεσίες ανάμειξης με σκοπό την αλλοίωση του ίχνους των συναλλαγών. Από τις 10 πιο έμπιστες υπηρεσίες ανάμειξης, μόνο η πιο γνωστή, για το 2015 διακινούσε 65 χιλιάδες Bitcoins κάθε μήνα.

Αξιοσημείωτη παραμένει η διαφορά του αριθμού συναλλαγών που μπορεί να διεκπεραιώσει το Bitcoin σε σύγκριση με τα συστήματα πιστωτικών καρτών. Το Bitcoin μπορεί να επεξεργαστεί 3 με 4 συναλλαγές ανά δευτερόλεπτο ενώ το PAYPAL επεξεργάζεται περίπου 200 συναλλαγές το δευτερόλεπτο και η VISA μπορεί να χειριστεί περίπου 1600 συναλλαγές το δευτερόλεπτο.[16]

1.3 Εξόρυξη Bitcoin

Τα τελευταία χρόνια μεγάλες ομάδες εξόρυξης κατέλαβαν τον έλεγχο της εξόρυξης, περιορίζοντας την αποκεντρωμένη φύση του δικτύου Bitcoin και κάνοντας την εξόρυξη ασύμφορη για τους άγνωστους miner, αυτούς που δεν συμμετέχουν σε ομάδες.

Το 2018 ήταν χρονιά αλλαγών για τις ομάδες εξόρυξης. Αρκετοί miners αποφάσισαν να εγκαταλείψουν την εξόρυξη εξαιτίας της πτώσης της τιμής του Bitcoin, το ίδιο ισχύει και για αρκετές εταιρίες που διατηρούν φάρμες εξόρυξης Bitcoin. Συγκεκριμένα τα κέρδη της Bitmain, η μεγαλύτερη φάρμα εξόρυξης, κλείνοντας το 2018 μειώθηκαν κατά 700 εκατομμύρια.

Το ίδιο παρατηρείται από την γεωγραφική κατανομή της συνολικής ισχύς εξόρυξης, στην Κίνα όπου δραστηριοποιούνται οι μεγαλύτερες εταιρίες εξόρυξης βρισκόταν το 80% ενώ τώρα έχει μειωθεί στο 60%, μετά είναι η Αμερική με τον Καναδά και ακολουθούν οι Ισλανδία, Σκανδιναβία, Ρωσία, Γεωργία.[17]

Στο πρώτο τρίμηνο του 2018 οι άγνωστοι miners, έλυσαν το 6% των blocks ενώ στο τέταρτο τρίμηνο έφτασαν να έχουν λύσει το 22% των blocks. Αντίστοιχα η Bitmain στο πρώτο τρίμηνο του 2018 έλυσε το 40% των blocks ενώ στο τέταρτο τρίμηνο έλυσε μόλις το 30% των blocks. Τα έσοδα από την εξόρυξη Bitcoin τον Γενάρη του 2018 κυμαίνονταν στα

1.2 δις δολάρια αλλά τον Δεκέμβριο υπέστησαν μείωση ύψους 83%, δηλαδή έφτασαν να είναι μόλις 210 εκατομμύρια δολάρια.

Στις αρχές του 2018 η Bitmain κατείχε το 53% της συνολικής ισχύς του δικτύου Bitcoin. Τώρα το δίκτυο του Bitcoin είναι λιγότερο πιθανών να δεχτεί επίθεση από το 51% των miners. Αυτό οφείλετε στην μείωση της ισχύς, δηλαδή της κυριαρχίας, των ομάδων έναντι της αύξησης της ισχύς των άγνωστων miner.[18, 19]

1.4 Ανταλλακτήρια Bitcoin, επένδυση και κερδοσκοπία

Η αύξηση των ανταλλακτηρίων κρυπτονομισμάτων ήταν εντυπωσιακή καθώς το 2015 υπήρχαν περίπου 70 ενώ στο τέλος του 2018 έχουν ξεπεράσει τα 190. Τον Δεκέμβριο του 2017 που ήταν το αποκορύφωμα της τιμής τους Bitcoin τα ανταλλακτήρια διακίνησαν περισσότερα από 5.5 δις δολάρια.

Η πτώση της τιμής του Bitcoin επηρέασε αρνητικά τον όγκο συναλλαγών των ανταλλακτηρίων κρυπτονομισμάτων. Τα 3 μεγαλύτερα ανταλλακτήρια Binance, Coinbase και OKEx από τον Οκτώβριο του 2018 μέχρι τον Γενάρη του 2019 είδαν μείωση στον ημερήσιο όγκο συναλλαγών μεταξύ Bitcoin και δολαρίου περισσότερο από 40%, 50% και 30% αντίστοιχα.[20]

Επιπλέον αν αναλογιστεί κανείς ότι ένα μεγάλο ποσοστό του ημερήσιου όγκου συναλλαγών των ανταλλακτηρίων κρυπτονομισμάτων είναι εικονικό, τότε γίνεται κατανοητό ότι η μείωση ήταν πολύ μεγαλύτερη. Τα ανταλλακτήρια δημιουργούν αυτοματοποιημένες εικονικές συναλλαγές για να αυξήσουν τον όγκο συναλλαγών τους. Ο αυξημένος όγκος συναλλαγών συνεπάγεται την δημιουργία αίσθησης ότι έχουν μεγάλη ρευστότητα αλλά και την δυνατότητα μεγαλύτερης χρέωσης στις ICO που θέλουν να ενταχθούν στο ανταλλακτήριο.[21]

Επίσης από τον Οκτώβριο του 2018 έως τον Γενάρη του 2019 το ποσοστό μείωσης της τιμής του Bitcoin ταυτίστηκε με το ποσοστό μείωσης της τιμής της μετοχής της Apple. Το ποσοστό μείωσης σε 3 μήνες για το Bitcoin ήταν 40% ενώ για την μετοχή της Apple 38%. Για την μετοχή της Apple γράφτηκε ότι είχε φτάσει σε επίπεδο κερδοσκοπίας, είναι κατανοητό ότι ισχύει το ίδιο για την τιμή του Bitcoin.[22]

Στο τρίτο τρίμηνο του 2018 βρέθηκε ότι το 55% των Bitcoin βρίσκονται σε πορτοφόλια με περισσότερα από 200 Bitcoins. Πολύ ενδιαφέρον είναι ότι το 1/3 αυτών των πορτοφολιών δεν έχουν πραγματοποιήσει ποτέ εξερχόμενη συναλλαγή. Επίσης παρατηρήθηκε ότι το 62% των Bitcoin βρίσκονται σε πορτοφόλια με περισσότερα από 100 Bitcoins. Στα συγκεκριμένα πορτοφόλια αντιστοιχούν μόλις το 0.1% όλων των διευθύνσεων Bitcoin. Ενώ περισσότερα από το 87% των Bitcoin βρίσκονται σε πορτοφόλια με περισσότερα από 10 Bitcoins. Σε αυτά τα πορτοφόλια αντιστοιχεί μόλις 0.7% όλων των διευθύνσεων Bitcoin.

Ωστόσο, αυτά τα στοιχεία δεν αντιπροσωπεύουν απλώς έναν μικρό αριθμό απίστευτα πλούσιων ατόμων, αλλά είναι πολύ πιθανό να ανήκουν σε ανταλλακτήρια κρυπτονομισμάτων. Καθώς περίπου το 4% του συνόλου των Bitcoin ζει σε πορτοφόλια που ανήκουν σε ανταλλακτήρια.[23]

Οι χρήστες του Bitcoin μπορούν να ταξινομηθούν σε τέσσερις κατηγορίες, σε κερδοσκόπους, σε υπηρεσίες, σε επενδυτές και σε χαμένα bitcoin. Από τον Δεκέμβριο του 2017 έως και τον Απρίλιο του 2018 οι μακροπρόθεσμοι επενδυτές πούλησαν Bitcoins αξίας 30 δις δολαρίων σε νέους κερδοσκόπους. Τα Bitcoins που κατέχουν πλέον οι επενδυτές ισούνται με τα Bitcoins που κατέχουν οι κερδοσκόποι μαζί με τις υπηρεσίες.

Όπως γράφτηκε παραπάνω τα Bitcoins είναι συγκεντρωμένα σε λίγες διευθύνσεις. Αυτή η συγκέντρωση μειώθηκε με την είσοδο νέων κερδοσκόπων. Το ένα τρίτο των διαθέσιμων Bitcoin ελέγχεται από 1000 μακροπρόθεσμους επενδυτές και 600 κερδοσκόπους. Οι νέοι κερδοσκόποι ενδέχεται να μειώσουν την τιμή του Bitcoin καθώς εξαιτίας τους η προσφορά για Bitcoins έχει αυξηθεί περίπου 60%. Για να ανακάμψουν οι τιμές, είτε οι κερδοσκόποι θα χρειαστεί να μειώσουν την προσφορά κρατώντας τα Bitcoins τους, είτε η ζήτηση θα πρέπει να αυξηθεί, οδηγούμενη από ένα νέο κύμα κερδοσκόπων ή νέες περιπτώσεις χρήσης του Bitcoin.

Τέλος στην σειρά ταξινόμησης στις τέσσερις κατηγορίες όλων των Bitcoins πρώτη με 35% είναι οι μακροπρόθεσμοι επενδυτές, δεύτερη με 30% είναι τα μη προσβάσιμα Bitcoins και όσα δεν έχουν εξορυχθεί ακόμα, τρίτη είναι οι κερδοσκόποι με 25% ακολουθούν με 10% οι υπηρεσίες.[24]

1.5 Γενικά χαρακτηριστικά για το δίκτυο συναλλαγών του Bitcoin

Επιλέχθηκε το δίκτυο συναλλαγών του Bitcoin επειδή είναι το πιο διαδεδομένο κρυπτονόμισμα και χρησιμοποιείται περισσότερο από όλα τα άλλα. Υπάρχει ένα μεγάλο οικοσύστημα στο οποίο μπορούν να γίνουν αγοραπωλησίες έναντι αγαθών ή υπηρεσιών με Bitcoin. Χαρακτηρίζεται ως ψευδό-ανώνυμο επειδή στις συναλλαγές που καταγράφονται στο Blockchain δεν χρησιμοποιούνται δεδομένα ταυτοποίησης. Όλοι μπορούν να δουν τις συναλλαγές και τις διευθύνσεις που σχετίζονται με αυτές, αλλά κανείς δεν ξέρει σε ποιο πρόσωπο ανήκουν. Θα μπορούσε να θεωρηθεί ότι το δίκτυο συναλλαγών του Bitcoin έχει κοινά χαρακτηριστικά με ένα ανώνυμο κοινωνικό δίκτυο το οποίο περιέχει τις οικονομικές συναλλαγές των μελών του.

Μετά την είσοδο του Bitcoin το 2010 σε ανταλλακτήρια, χρησιμοποιούταν εκτός από μέσω ανταλλαγής και ως μέσω αποταμίευσης. Πλέον λόγω των μεγάλων αυξομειώσεων στην τιμή του θεωρείται ότι είναι ένα ακόμα χρηματοπιστωτικό παράγωγο υψηλού κινδύνου. Συγκεκριμένα στο τέταρτο τρίμηνο του 2018 κατέχει περίπου το 50% την αγοράς κρυπτονομισμάτων, είναι το πιο ακριβό κρυπτονόμισμα παγκοσμίως, με την τιμή του να κυμαίνεται στα 4000 δολάρια. Η κεφαλαιοποίηση του είναι περίπου 102 δις δολάρια ενώ στα τέλη του 2017 είχε φτάσει τα 240 δις δολάρια.

Υπάρχουν περίπου 4400 μηχανήματα ATM παγκοσμίως που ανταλλάσσουν εθνικά νομίσματα με Bitcoin. Τα ATM κατανέμονται γεωγραφικά ανά ήπειρο με πρώτη να είναι η Β. Αμερική με 70%, ακολουθεί η Ευρώπη με 24% και η Ασία με 2.5%, μετά είναι η Ωκεανία και Ν. Αμερική με 1.5% η κάθε μία και τελευταία είναι η Αφρική με 0.5%.[25]

2. ΟΙ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ BLOCKCHAIN

Στο Bitcoin ο όρος Blockchain περιγράφει τόσο το κατάστιχο όσο και την τεχνολογία που είναι υλοποιημένο το κατάστιχο. Μέχρι το 2024 αναμένεται η αγορά κρυπτονομισμάτων που βασίζονται στο Blockchain να ξεπεράσει τα 20 δις δολάρια. Ήδη το 90% των Αμερικάνικων και Ευρωπαϊκών τραπεζών επενδύουν στο Blockchain, γιατί υπολογίζεται ότι οι τράπεζες θα μπορούσαν να εξοικονομήσουν περίπου 10 δις δολάρια ετησίως με την χρήση της τεχνολογίας Blockchain.[26, 27, 28]

Έχει αρχίσει η δημιουργία ιδιωτικών και κοινόχρηστων Blockchains τα οποία αποτελούνται πρακτικά από ένα κλειστό δίκτυο κόμβων. Τα ιδιωτικά Blockchain έχουν ένα

περιέχει τα πραγματικά συνεχιζόμενα δεδομένα των blocks του Bitcoin blockchain σε δυαδική μορφή. Κάθε block που λαμβάνει ένας κόμβος προσαρτάται στο τελευταίο αρχείο blk*.dat μέχρις ότου το αρχείο φτάσει στο επιτρεπόμενο μέγεθος, έπειτα δημιουργείται το επόμενο αρχείο blk*.dat και συνεχίζει η προσάρτηση των υπολειπόμενων δεδομένων.[33]

Το Bitcoin blockchain χρησιμοποιεί ένα σύστημα scripting για συναλλαγές, το οποίο δεν είναι σκόπιμα Turing-πλήρες. Ένα script είναι ουσιαστικά μια λίστα οδηγιών που καταγράφονται με κάθε συναλλαγή, περιγράφει πως ο επόμενος χρήστης που θέλει να ξοδέψει τα bitcoins που απέκτησε μπορεί να αποκτήσει πρόσβαση σε αυτά.[34]

Τα δεδομένα των blocks που βρίσκονται σε δυαδική μορφή στα αρχεία blk*.dat έχουν συγκεκριμένη δομή και σειρά, με πλήρως ορισμένο μέγεθος για κάθε πεδίο σε bytes:

- 1 Ένα μοναδικό αναγνωριστικό της επικοινωνίας μεταξύ των κόμβων ώστε να ξεχωρίζει η αρχή των blocks μεταξύ τους.
- 2 Ένας αριθμός με τα bytes που χρειάζονται ακόμα να διαβαστούν μέχρι να ολοκληρωθεί το block, δηλαδή δηλώνει το μήκος του block σε bytes.
- 3 Η κεφαλή του block, είναι σαν μεταδεδομένα που παρέχουν μια μοναδική περίληψη ολόκληρου του block. Τα συγκεκριμένα πεδία είναι:
 - 3.1 Η έκδοση της δομής του block.
 - 3.2 Το hash του τελευταίου block.
 - 3.3 Την ρίζα του merkle tree όλων των συναλλαγών του block.
 - 3.4 Η χρονική σήμανση που αυτό το block εξορύχτηκε.
 - 3.5 Η δυσκολία εξόρυξης.
 - 3.6 Η λύση του προβλήματος εξόρυξης.
- 4 Ένας ακέραιος αριθμός που χρησιμοποιείται για να υποδείξει το μήκος του επερχόμενου πεδίου.
- 5 Είναι τα δεδομένα των συναλλαγών που υπάρχουν στο block και περιγράφουν την κίνηση των bitcoins με τα ακόλουθα πεδία:
 - 5.1 Η έκδοση της δομής των δεδομένων συναλλαγής.
 - 5.2 Ένας ακέραιος αριθμός που υποδεικνύει τον αριθμό των εισερχόμενων διευθύνσεων.
 - 5.3 Για κάθε εισερχόμενη διεύθυνση υπάρχουν τα παρακάτω πεδία:

- 5.3.1 Ένα μοναδικό αλφαριθμητικό αναγνώρισης μιας υφιστάμενης συναλλαγής.
 - 5.3.2 Ένας ακέραιος αριθμός που υποδεικνύει μια εξερχόμενη διεύθυνση της επιλεγμένης υφιστάμενης συναλλαγής.
 - 5.3.3 Δύο πεδία που σχετίζονται με το hash του ιδιωτικού κλειδιού ώστε να επιτραπεί η μεταφορά των bitcoins.
 - 5.3.4 Ένας ακέραιος αριθμός που υποδεικνύει την θέση της εισερχόμενης διεύθυνσης.
 - 5.4 Ένας ακέραιος αριθμός που υποδεικνύει τον αριθμό των εξερχόμενων διευθύνσεων.
 - 5.5 Για κάθε εξερχόμενη διεύθυνση υπάρχουν τα παρακάτω πεδία:
 - 5.5.1 Το ποσό των bitcoins σε satoshis που θα σταλούν στην εξερχόμενη διεύθυνση.
 - 5.5.2 Δύο πεδία που σχετίζονται με το hash του δημόσιου κλειδιού της εξερχόμενης διεύθυνσης.
 - 5.6 Ο ελάχιστος χρόνος μετά τον οποίο μπορεί να εκτελεστεί η συναλλαγή.
- [35, 36, 37, 38, 39, 40]

Σε κάθε block δημιουργείται μία επιπλέον συναλλαγή, η coinbase, που μπαίνει πάντα πρώτη και αφορά την πληρωμή για την επιτυχή εξόρυξη του block. Η συγκεκριμένη συναλλαγή διαφέρει από τις άλλες συναλλαγές γιατί τα πεδία της εισερχόμενης διεύθυνσης είναι κενά. Η εισερχόμενη διεύθυνση είναι κενή γιατί τα bitcoins που περιέχει η συναλλαγή δεν προέρχονται από κάποια υφιστάμενη διεύθυνση αλλά δημιουργούνται από τον σύστημα και δίνονται ως ανταμοιβή της επιτυχής εξόρυξης.[41]

Η δομή και σειρά των δεδομένων των blocks που βρίσκονται σε δυαδική μορφή στα αρχεία blk*.dat έχει αλλάξει αρκετές φορές μέχρι σήμερα, αυτό συμβαίνει μετά από κάποια πρόταση βελτίωσης των μελών της κοινότητας του Bitcoin προς τους προγραμματιστές μέσω των BIP. Αυτές οι αλλαγές επηρεάζουν την λειτουργία των εργαλείων εξόρυξης δεδομένων από το Bitcoin blockchain.[42]

2.2 Εργαλεία εξόρυξης δεδομένων από το Bitcoin blockchain

Έγινε αναζήτηση στο πρώτο τρίμηνο του 2018 για ελεύθερα και ανοικτού κώδικα εργαλεία εξόρυξης δεδομένων από το Bitcoin blockchain, βρέθηκαν περίπου 40 δοκιμασμένα εργαλεία από τα οποία τα 10 τουλάχιστον δεν υποστηρίζονται πλέον και έχει σταματήσει η ανάπτυξη τους. Αυτά τα εργαλεία εξόρυξης δεδομένων διαφέρουν στην γλώσσα προγραμματισμού που είναι υλοποιημένα, στον τρόπο λειτουργίας τους και στην μορφή που εξάγουν τα δεδομένα. Χωρίστηκαν σε 4 κατηγορίες ανάλογα με την μορφή που εξάγουν τα δεδομένα:

1. Στην πρώτη κατηγορία είναι αυτά που εξάγουν σε αρχεία διάφορων τύπων όπως csv, txt, κλπ.
2. Στην δεύτερη κατηγορία είναι αυτά που εξάγουν σε σχεσιακές βάσεις δεδομένων.
3. Στην τρίτη κατηγορία είναι αυτά που εξάγουν στην Neo4j βάση δεδομένων-γράφων.
4. Στην τέταρτη κατηγορία είναι αυτά που κρατάνε ένα μέρος των δεδομένων στην μνήμη και τα παρουσιάζουν κατά την εκτέλεση τους.

Παρατηρήθηκε ότι η εξαγωγή των περίπου 200 GB δεδομένων του Bitcoin blockchain στην Neo4j βάση δεδομένων χρειάζεται τον πενταπλάσιο χώρο, όπως με το εργαλείο blockchain2graph.[43] Ενώ σε σχεσιακή βάση δεδομένων χρειάζεται περίπου τον τριπλάσιο χώρο, όπως με το εργαλείο bitcoin-abe.[44] Τα εργαλεία που κρατάνε ένα μέρος των δεδομένων στην μνήμη τους χαρακτηρίζονται από την πολύ μεγάλη ταχύτητα τους αλλά δεν είναι εύχρηστα για περαιτέρω ανάλυση καθώς δεν παρέχουν τα δεδομένα σε κάποια δομημένη μορφή, όπως με το εργαλείο Blocksci.[45]

Τα εργαλεία εξόρυξης δεδομένων από το Bitcoin blockchain ταξινομήθηκαν με κριτήρια την συχνότητα αναβαθμίσεων, την ευχρηστία τους και την χρήση τους σε σχετική βιβλιογραφία. Υπάρχουν κάποια εργαλεία που χρησιμοποιούν το κατακευκτωμένο σύστημα αρχείων Hadoop για να εξάγουν τα δεδομένα του Bitcoin blockchain, τα συγκεκριμένα εργαλεία δεν χρησιμοποιήθηκαν γιατί απαιτούσαν λεπτομερείς παραμετροποίηση και αρκετούς πόρους. Επίσης δεν δοκιμάστηκε κάποιο εργαλείο που εξάγει σε Neo4j βάση δεδομένων καθώς δεν υπήρχαν οι διαθέσιμοι πόροι, όπως και δεν δοκιμάστηκε το εργαλείο Blocksci καθώς δεν υπήρχε εύχρηστος τρόπος για να μπορέσουν να χρησιμοποιηθούν τα δεδομένα που επιστρέφει στην στατιστική ανάλυση του δικτύου συναλλαγών του Bitcoin.

Ένα κοινό πρόβλημα που εντοπίστηκε σε όσα εργαλεία εξόρυξης δεδομένων δοκιμάστηκαν είναι η αδυναμία αναγνώρισης κάποιων διευθύνσεων που υπάρχουν στο Bitcoin blockchain. Το κάθε εργαλείο αντιμετώπισε διαφορετικά το συγκεκριμένο πρόβλημα. Αυτό οφείλεται στο Bitcoin blockchain που επιτρέπει, εκτός από τους διάφορους τύπους διευθύνσεων, να μπει οποιαδήποτε κωδικοποίηση αυθαίρετων δεδομένων στην θέση των εξερχόμενων διευθύνσεων.[46, 47]

2.3 Το εργαλείο εξόρυξης δεδομένων bitcoin-abe

Αρχικά δοκιμάστηκε το εργαλείο bitcoin-abe το οποίο χρειάστηκε περίπου 80 ημέρες συνεχής λειτουργίας για να εξάγει τα δεδομένα του Bitcoin blockchain στο δεύτερο τρίμηνο του 2018. Το τελευταίο block μέχρι τότε ήταν περίπου στον αριθμό 538000. Η συνεχής λειτουργία του διακόπηκε όταν έφτασε στα block μετά τον Αύγουστο του 2017 γιατί τότε εφαρμόστηκε το BIP που αφορούσε το λεγόμενο SegWit.[48] Η συγκεκριμένη αναβάθμιση άλλαξε την δομή των block οπότε χρειαζόταν αντίστοιχα αναβάθμιση στο εργαλείο bitcoin-abe για να μπορέσει να συνεχίσει να λειτουργεί. Η αναβάθμιση δεν έγινε ποτέ επίσημα καθώς το εργαλείο είχε σταματήσει να υποστηρίζεται και να αναβαθμίζεται. Μια αναβάθμιση από κάποια μέλη της κοινότητας του εργαλείου bitcoin-abe έδινε μια μερική λύση για να συνεχίσει η λειτουργία του. Εξαιτίας της συγκεκριμένης αναβάθμισης όμως υπήρξε μερική αλλοίωση στα εξαγόμενα δεδομένα και τότε αποφασίστηκε να σταματήσει η εξόρυξη δεδομένων από το Bitcoin blockchain με το συγκεκριμένο εργαλείο.

2.4 Τα εργαλεία εξόρυξης δεδομένων blockchainToAvro και bitcoin-etl, η υπηρεσία Google BigQuery

Η υπηρεσία της Google BigQuery, η οποία λειτουργεί ως μία σχεσιακή βάση δεδομένων στο cloud, προσφέρει τα δεδομένα του Bitcoin blockchain αφού έχουν εξορυχθεί με το εργαλείο blockchainToAvro. Για ευκολότερη χρήση της υπηρεσίας BigQuery χρησιμοποιήθηκε η πλατφόρμα kaggle καθώς συνεργάζεται με την υπηρεσία BigQuery και παρέχει κάποιες διευκολύνσεις. [49, 50, 51]

Υπήρξαν κάποια προβλήματα κατά την ανάκτηση των δεδομένων από την συγκεκριμένη υπηρεσία τα οποία είχαν σαν αποτέλεσμα την απόρριψη της μαζί με τα

δεδομένα που προσφέρει. Το ένα πρόβλημα αφορά τον περίπλοκο τρόπο ανάκτησης των συναλλαγών σε ημερολογιακή σειρά και το άλλο πρόβλημα αφορά τον μεγάλο όγκο δεδομένων που χρειαζόταν να κατέβουν από το cloud ώστε να δημιουργηθεί το δίκτυο συναλλαγών και να ξεκινήσει η στατιστική ανάλυση του. Το 2019 έγινε προσπάθεια βελτίωσης των προβλημάτων που προαναφέρθηκαν χρησιμοποιώντας ένα καινούργιο εργαλείο εξόρυξης το bitcoin-etl.[52]

2.5 Το εργαλείο εξόρυξης δεδομένων bitcoind-dump-tsv

Το λογισμικό bitcoind-dump-tsv είναι μια ιδιαίτερη περίπτωση των εργαλείων εξόρυξης δεδομένων από το Bitcoin blockchain. Αποτελεί μια τροποποιημένη έκδοση του Bitcoin Core το οποίο είναι το λογισμικό που συγχρονίζει και διατηρεί πλήρης κόμβους για την επικύρωση και εγκυρότητα του Bitcoin blockchain. Το bitcoind-dump-tsv λειτουργεί σαν ένας οποιοσδήποτε πλήρης κόμβος, όταν όμως επιλεγθεί η ειδική λειτουργία εξαγωγής διαβάζεται ολόκληρο το Bitcoin blockchain από τα αρχεία δεδομένων και εξάγεται σε αρχεία μορφής tsv.[53, 54]

Όλα τα εξαγόμενα αρχεία χρησιμοποιούν αριθμητικά αναγνωριστικά που ξεκινούν από το 0 για να αναφερθούν σε συναλλαγές, μπλοκ και διευθύνσεις. Η αντιστοίχιση τους με τα πραγματικά hash βρίσκονται στα αρχεία txh.dat για τις συναλλαγές, bh.dat για τα blocks και addresses.dat για τις διευθύνσεις. Η ειδική τιμή -1 για αναγνωριστικό συναλλαγής σημαίνει ότι υπήρξε κάποιο σφάλμα κατά την διαδικασία της εξαγωγής. Η ειδική τιμή -1 για αναγνωριστικό διεύθυνσης σημαίνει ότι δεν ήταν δυνατό να αναγνωριστεί η διεύθυνση, το οποίο δεν είναι απαραίτητα σφάλμα καθώς υπάρχουν μη τυπικές διευθύνσεις.

Όλα τα ποσά είναι στην μικρότερη υποδιαίρεση του Bitcoin το Satoshi. Η coinbase συναλλαγή αναγνωρίζεται από τις μηδενικές εισερχόμενες διευθύνσεις. Οι εισερχόμενες και εξερχόμενες διευθύνσεις μίας συναλλαγής περιλαμβάνουν έναν αριθμό ακολουθίας που ξεκινάει από το 0 για κάθε συναλλαγή, η οποία προσδιορίζει την θέση της εισερχόμενης και εξερχόμενης διεύθυνσης.

Επειδή η διαδικασία εξαγωγής θα διαρκούσε αρκετές μέρες, χρησιμοποιήθηκαν τα εξαγόμενα αρχεία τα οποία δημιουργήθηκαν περίπου στις 09 Φεβρουαρίου 2018 και διανέμονται στο ίντερνετ από την ομάδα που υλοποίησε το εργαλείο εξόρυξης δεδομένων.

Εμπεριέχουν τα δεδομένα από την αρχή του Bitcoin blockchain μέχρι το block 508241. Το σύνολο δεδομένων βρίσκεται στα παρακάτω αρχεία με τα συγκεκριμένα πεδία:

- Το αρχείο bh.dat μεγέθους 44 MB περιέχει τα hash των blocks και την αντιστοίχιση τους στο αναγνωριστικό blockID, το block_timestamp που είναι η ημερομηνία και ώρα που εξορύχτηκαν τα block, τον αριθμό των συναλλαγών σε κάθε block.
- Το αρχείο txh.dat μεγέθους 22.3 GB περιέχει τα hash των συναλλαγών και την αντιστοίχιση τους στο αναγνωριστικό txID.
- Το αρχείο addresses.dat μεγέθους 16.5 GB περιέχει τις διευθύνσεις και την αντιστοίχιση τους στο αναγνωριστικό addrID.
- Το αρχείο tx.dat μεγέθους 6.2 GB περιέχει τα αναγνωριστικά txID και blockID, τον αριθμό εισερχόμενων διευθύνσεων n_inputs και εξερχόμενων διευθύνσεων n_outputs.
- Το αρχείο txin.dat μεγέθους 31 GB περιέχει τα αναγνωριστικά txID, prev_txID(προηγούμενη συναλλαγή) και addrID, την θέση της εισερχόμενης διεύθυνσης input_seq, την θέση της εξερχόμενης διεύθυνσης ως εξερχόμενη στην προηγούμενη συναλλαγή prev_output_seq, το σύνολο των satoshis που στέλνουν τα πρόσωπα.
- Το αρχείο txout.dat μεγέθους 23.8 GB περιέχει τα αναγνωριστικά txID και addrID, την θέση της εξερχόμενης διεύθυνσης output_seq, το σύνολο των satoshis που δέχονται τα πρόσωπα.
- Το αρχείο multiple.dat μεγέθους 21.6 MB περιέχει τα αναγνωριστικά txID και addrID, την θέση της εξερχόμενης διεύθυνσης output_seq, των συναλλαγών όπου πολλαπλές διευθύνσεις λαμβάνουν μαζί το σύνολο των satoshis.
- Το αρχείο nonstandard.dat μεγέθους 55 MB περιέχει το αναγνωριστικό txID και την θέση της εξερχόμενης διεύθυνσης output_seq, των συναλλαγών όπου η εξερχόμενη διεύθυνση δεν μπορεί να αναγνωριστεί.
- Το αρχείο addr_sccs.dat μεγέθους 4.3 GB περιέχει τα αναγνωριστικά addrID και userID. Το userID περιγράφει μια πιθανή ομαδοποίηση διευθύνσεων σε πρόσωπα, χρησιμοποιώντας έναν απλό ευριστικό αλγόριθμο ότι όλες οι εισερχόμενες διευθύνσεις μιας συναλλαγής ελέγχονται από το ίδιο πρόσωπο.[55]

3. ΑΝΑΛΥΣΗ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ TXEDGES

Η ομάδα που υλοποίησε το εργαλείο εξόρυξης δεδομένων bitcoind-dump-tsv έχει υλοποιήσει και το εργαλείο txedges, το οποίο ενώνει τις εισερχόμενες και εξερχόμενες διευθύνσεις των συναλλαγών για να δημιουργήσει ένα κατευθυνόμενο σταθμισμένο δίκτυο μεταξύ διευθύνσεων, με την μορφή λίστας κατευθυνόμενων σταθμισμένων συνδέσεων. Το εξαγόμενο αρχείο txedges.dat μεγέθους 115 GB περιέχει:

- Το αναγνωριστικό συναλλαγής txID
- Το αναγνωριστικό εισερχόμενης διεύθυνσης in_addrID
- Το αναγνωριστικό εξερχόμενης διεύθυνσης out_addrID
- Το πεδίο weight που είναι το σύνολο των satoshis της συναλλαγής.

Δεν συμπεριλαμβάνονται οι coinbase συναλλαγές και οι αμοιβές των miner.[56]

Επίσης χρειάστηκε να βρεθεί η τιμή του Bitcoin σε δολάρια για κάθε ημέρα, σχεδόν όλα τα ανταλλακτήρια παρέχουν λίστες με τις τιμές ανοίγματος και κλεισίματος του Bitcoin σε δολάρια. Επειδή όμως υπάρχουν διαφορές στις τιμές ανά ανταλλακτήριο αποφασίστηκε να χρησιμοποιηθεί η λίστα ημερήσιων τιμών κλεισίματος του Bitcoin σε δολάρια από την ενημερωτική ιστοσελίδα coindesk, με το αρχείο να ονομάζεται BTC_USD-CoinDesk.csv, η οποία μπορεί να θεωρηθεί αξιόπιστη καθώς δεν διαχειρίζεται με κανένα τρόπο κρυπτονομίσματα.

Επιλέχθηκε να γίνει η ανάλυση με την γλώσσα προγραμματισμού R η οποία όμως υποστηρίζει αναπαράσταση ακεραίων αριθμών μέχρι 32 bit, δηλαδή από -2^{31} μέχρι και $2^{31}-1$. Το αρχείο txedges.dat με την λίστα κατευθυνόμενων σταθμισμένων συνδέσεων έχει 2536261805 συνδέσεις, ο συγκεκριμένος αριθμός είναι αρκετά μεγαλύτερος από τον αριθμό 2147483648 που ισούται με το 2^{31} , οπότε και δεν μπορούσε να διαβαστεί το αρχείο txedges.dat στην R. Εξαιτίας αυτού του περιορισμού χωρίστηκε το αρχείο txedges.dat περίπου στις 1500000000 συνδέσεις ώστε να δημιουργηθούν 2 μικρότερα αρχεία, έτσι εξασφαλίστηκε ότι μπορεί να διαβαστεί το κάθε αρχείο στην R.

Επίσης λόγω του πολύ μεγάλου αριθμού συνδέσεων αποφασίστηκε να μην αναλυθεί και αναπαρασταθεί το δίκτυο συναλλαγών ως ένα ενιαίο, υπήρχε αμφιβολία αν θα μπορούσαν να εξαχθούν συμπεράσματα λόγω την ανομοιομορφίας στην εξέλιξη του δικτύου, όπως και ότι οι πάρα πολλές απαιτήσεις σε υπολογιστικούς πόρους θα περιορίσουν τις δυνατότητες της

βιβλιοθήκης iGraph που θα χρησιμοποιηθεί για την στατιστική ανάλυση. Αποφασίστηκε εν τέλει να τεμαχιστεί το δίκτυο συναλλαγών σε ημερήσια δίκτυα συναλλαγών, όπου θα γίνει στατιστική ανάλυση για κάθε ημερήσιο δίκτυο συναλλαγών. Τα αποτελέσματα της στατιστικής ανάλυσης για κάθε ημερήσιο δίκτυο συναλλαγών θα ενωθούν με την αντίστοιχη ημερολογιακή σειρά και θα δημιουργηθεί μία χρονική σειρά.

3.1 Συμπλήρωση δεδομένων και τεμαχισμός του δικτύου συναλλαγών

Χρειάστηκε να προστεθούν μερικά πεδία από το σύνολο δεδομένων του bitcoind-dump-tsn στην λίστα κατευθυνόμενων σταθμισμένων συνδέσεων ώστε να γίνει πιο πλήρης η ανάλυση. Πρώτα διαβάστηκαν στην R τα αρχεία txedges.dat(στην ουσία τα 2 μικρότερα αρχεία που χωρίστηκε), bh.dat, tx.dat και BTC_USD-CoinDesk.csv.

Μετά προστέθηκε στις συνδέσεις του αρχείου txedges.dat το πεδίο blockID από το αρχείο tx.dat και το πεδίο block_timestamp από το αρχείο bh.dat, των οποίων οι τιμές ταυτοποιήθηκαν με τα πεδία txID και blockID αντίστοιχα. Οι συνδέσεις πλέον απέκτησαν την ημερομηνία και ώρα του block στο οποίο ανήκουν με το πεδίο block_timestamp, από το οποίο όμως αφαιρέθηκε η ώρα ώστε να αντιστοιχηθούν οι συνδέσεις στην ίδια χρονική στιγμή για κάθε ημέρα.

Κατά την προσθήκη του πεδίου usd από το αρχείο BTC_USD-CoinDesk.csv παρατηρήθηκε ότι υπάρχουν συνδέσεις πριν από την πρώτη ημέρα που καταγράφηκε τιμή για το Bitcoin. Το οποίο είναι λογικό καθώς το Bitcoin άρχισε να καταγράφεται και να ανταλλάσσεται έναντι δολαρίων περίπου ένα χρόνο μετά την δημιουργία του. Τελικά οι συνδέσεις στις οποίες δεν αντιστοιχεί τιμή απορρίφθηκαν από την ανάλυση. Οι υπόλοιπες συνδέσεις ταυτοποιήθηκαν από το πεδίο block_timestamp την λίστας συνδέσεων με το πεδίο timestamp του αρχείου BTC_USD-CoinDesk.csv.

Το πεδίο weight που περιέχει το σύνολο των satoshis κάθε συναλλαγής μετατράπηκε σε bitcoin, μετά πολλαπλασιάστηκε με την τιμή του Bitcoin και δημιουργήθηκε το πεδίο weight2USD που έχει το σύνολο της συναλλαγής σε δολάρια.

Καθώς ολοκληρώθηκε η συμπλήρωση πεδίων η λίστα κατευθυνόμενων σταθμισμένων συνδέσεων έχει πλέον τα πεδία block_timestamp, blockID, txID, in_addrID, out_addrID,

weight, usd και weight2USD. Κλείνοντας το πρώτο στάδιο της ανάλυσης έγινε ο τεμαχισμός της λίστας συνδέσεων σε 2763 ημερήσιες λίστες συνδέσεων.

3.2 Εφαρμογή περιγραφικής στατιστικής στα ημερήσια δίκτυα συναλλαγών

Το δεύτερο στάδιο της ανάλυσης θα εφαρμοστεί σειριακά σε όλες τις ημερήσιες λίστες συνδέσεων. Αρχικά αφαιρούνται όσες συνδέσεις έχουν ως εισερχόμενη ή εξερχόμενη διεύθυνση την τιμή -1, καθώς αυτή διεύθυνση θα δημιουργούσε έναν κόμβο με συνδέσεις που δεν ανταποκρίνονται στην πραγματικότητα και μπορεί να επηρεάσει τα αποτελέσματα της περιγραφικής στατιστικής. Μετά καταμετράται το πλήθος μοναδικών διευθύνσεων είτε εισερχόμενων είτε εξερχόμενων.

Στην συνέχεια δημιουργείτε το ημερήσιο δίκτυο συναλλαγών με την βιβλιοθήκη iGraph το οποίο έχει για κόμβους εκκίνησης των κατευθυνόμενων συνδέσεων τις τιμές του πεδίου `in_addrID` και για κόμβους τερματισμού των κατευθυνόμενων συνδέσεων τις τιμές του πεδίου `out_addrID`, οι συνδέσεις έχουν μόνο την τιμή του πεδίου `weight`. Από το ημερήσιο δίκτυο συναλλαγών καταμετράται το πλήθος κόμβων, συνδέσεων, πολλαπλών συνδέσεων και βρόγχων συνδέσεων. Μετά υπολογίζεται η πυκνότητα, η μεταβλητότητα, ο συντελεστής σύμπλεξης, ο συντελεστής ομοιότητας βαθμών και το πλήθος συνιστωσών. Επιπλέον καταμετράται το μέγιστο πλήθος Bitcoin, η μέγιστη τιμή σε δολάρια, το συνολικό πλήθος Bitcoin και η συνολική τιμή σε δολάρια από τις συναλλαγές του ημερήσιου δικτύου. Όπως και το πλήθος των κόμβων χωρίς εξερχόμενες συνδέσεις μαζί με το συνολικό πλήθος Bitcoin και την συνολική τιμή σε δολάρια των συγκεκριμένων κόμβων.

Από τα διανύσματα συνιστωσών και βαθμών του ημερήσιου δικτύου συναλλαγών κρατούνται η ελάχιστη τιμή, η τιμή στο 25% του διανύσματος, η ενδιάμεση τιμή, η μέση τιμή, η τιμή στο 75% του διανύσματος, η δεύτερη μεγαλύτερη τιμή, η μέγιστη τιμή, η τυπική απόκλιση, η διακύμανση, η δεύτερη πιο συχνή τιμή, η συχνότερη τιμή και ο συντελεστής μεταβολής.

Στην συνέχεια κατασκευάστηκε Erdos-Renyi τυχαίο δίκτυο με πολλαπλές συνδέσεις και βρόγχους, με τον ίδιο αριθμό κόμβων και συνδέσεων με το ημερήσιο δίκτυο συναλλαγών. Από το τυχαίο δίκτυο καταμετράται το πλήθος πολλαπλών συνδέσεων και βρόγχων συνδέσεων, η πυκνότητα, η μεταβλητότητα, ο συντελεστής σύμπλεξης, ο συντελεστής

ομοιότητας βαθμών, το πλήθος συνιστωσών και το πλήθος των κόμβων χωρίς εξερχόμενες συνδέσεις.

Από τα διανύσματα συνιστωσών και βαθμών του τυχαίου δικτύου κρατούνται η ελάχιστη τιμή, η τιμή στο 25% του διανύσματος, η ενδιάμεση τιμή, η μέση τιμή, η τιμή στο 75% του διανύσματος, η δεύτερη μεγαλύτερη τιμή, η μέγιστη τιμή, η τυπική απόκλιση, η διακύμανση, η δεύτερη πιο συχνή τιμή, η συχνότερη τιμή και ο συντελεστής μεταβολής.

Στην συνέχεια αναλύεται η γιγάντια συνιστώσα του ημερήσιου δικτύου συναλλαγών από την οποία καταμετράται το πλήθος κόμβων, συνδέσεων, πολλαπλών συνδέσεων και βρόγχων συνδέσεων. Επιπλέον καταμετράται το μέγιστο πλήθος Bitcoin, η μέγιστη τιμή σε δολάρια, το συνολικό πλήθος Bitcoin και η συνολική τιμή σε δολάρια από τις συναλλαγές της γιγάντιας συνιστώσας. Επίσης υπολογίζεται η πυκνότητα, η μεταβλητότητα, ο συντελεστής σύμπλεξης, ο συντελεστής ομοιότητας βαθμών και το πλήθος των κόμβων χωρίς εξερχόμενες συνδέσεις μαζί με το συνολικό πλήθος Bitcoin και την συνολική τιμή σε δολάρια των συγκεκριμένων κόμβων.

Από τα διανύσματα βαθμών, βαρών, ισχύος, κύρους με και χωρίς βάρη της γιγάντιας συνιστώσας κρατούνται η ελάχιστη τιμή, η τιμή στο 25% του διανύσματος, η ενδιάμεση τιμή, η μέση τιμή, η τιμή στο 75% του διανύσματος, η δεύτερη μεγαλύτερη τιμή, η μέγιστη τιμή, η τυπική απόκλιση, η διακύμανση, η δεύτερη πιο συχνή τιμή, η συχνότερη τιμή και ο συντελεστής μεταβολής.

Τέλος από την γιγάντια συνιστώσα υπολογίζεται ο συνδετικός και γραμμοσυνδετικός αριθμός, ο δείκτης των hub με και χωρίς βάρη, ο δείκτης των authorities με και χωρίς βάρη, ο αριθμός των τριγώνων και ο αριθμός των κορυφών-τομών.

Κάποιοι σημαντικοί δείκτες όπως τα διανύσματα εκκεντρότητας και ο περιορισμός Burt. Τα μέτρα κεντρικότητας για την ευκολία μετάβασης, την ροή της πληροφορίας, την ιδιοκεντρικότητα με και χωρίς βάρη. Μαζί με τον υπολογισμό της διαμέτρου, της μέσης απόστασης, του αριθμού πλήρη υπογραφημάτων μεγέθους 4, του πλήθους και της δομηκότητας των κοινοτήτων με και χωρίς βάρη και του πλήθους μέγιστης κλίμακας δεν μπόρεσαν να υπολογιστούν.

Ο λόγος που δεν μπόρεσαν να υπολογιστούν οφείλεται στον μεγάλο αριθμό συνδέσεων του ημερήσιου δικτύου συναλλαγών, πρώτων γιατί οι αλγόριθμοι που υπολογίζουν τους

συγκεκριμένους δείκτες χρειάζεται να διατρέξουν αρκετές φορές όλες τις συνδέσεις του δικτύου και δεύτερον γιατί οι συγκεκριμένοι αλγόριθμοι δεν είναι βελτιστοποιημένοι στην βιβλιοθήκη `igraph` ώστε να εκτελούνται παράλληλα.

3.3 Ανάλυση οικονομικής δραστηριότητας στα ημερήσια δίκτυα συναλλαγών

Στο τρίτο στάδιο της ανάλυσης θα χρησιμοποιηθούν τα αποτελέσματα από την εφαρμογή περιγραφικής στατιστικής στα ημερήσια δίκτυα συναλλαγών. Αρχικά υπολογίζεται το συνολικό πλήθος Bitcoin και η συνολική αξία σε δολάρια του δικτύου συναλλαγών για κάθε ημέρα.

Μετά υπολογίζεται ο δείκτης NVT ο οποίος όταν είναι υψηλός, υποδηλώνει ότι η συνολική αξία του δικτύου υπερβαίνει την αξία που μεταδίδεται στο δίκτυό, αυτό μπορεί να συμβεί όταν το δίκτυο είναι σε υψηλή ανάπτυξη και οι επενδυτές το αποτιμούν ως επένδυση με υψηλή απόδοση ή εναλλακτικά όταν η τιμή είναι σε μια μη βιώσιμη φούσκα.[57]

$$NVT = \frac{\text{ΣυνολικήΑξίαΔικτύου}}{\text{ΗμερήσιαΑξίαΔικτύου}} \quad (1)$$

$$NVT_{Smooth} = 14KMO(NVT) \quad (2)$$

Μετά υπολογίζεται ο δείκτης $NVTS$ ο οποίος είναι παράγωγος του NVT και δίνει μεγαλύτερη έμφαση στην πρόβλεψη σημάτων μπροστά από τις κορυφές των τιμών.[58]

$$DNV = 45KMO(\text{ΗμερήσιαΑξίαΔικτύου}) \quad (3)$$

$$NVTS = \frac{\text{ΣυνολικήΑξίαΔικτύου}}{DNV} \quad (4)$$

Επιπλέον υπολογίζεται ο δείκτης PMR που βασίζεται στους νόμους Metcalfe και Zipf, οι οποίοι υπολογίζουν την επίδραση των μοναδικών ημερήσιων διευθύνσεων στα δίκτυα επικοινωνίας.[59, 60]

$$GeneralizedMetcalfe = 15KMO(\text{ΜοναδικέςΗμερήσιεςΔιευθύνσεις}^{1.5}) \quad (5)$$

$$PMR_{GeneralizedMetcalf} = \ln\left(\frac{\text{ΣυνολικήΑξίαΔικτύου}}{GeneralizedMetcalf}\right) \quad (6)$$

$$\ln(DUA) = \ln(\text{ΜοναδικέςΗμερήσιεςΔιευθύνσεις}) \quad (7)$$

$$DUA_{Zipf} = 15KMO(\text{ΜοναδικέςΗμερήσιεςΔιευθύνσεις} \cdot \ln(DUA)) \quad (8)$$

$$PMR_{Zipf} = \ln\left(\frac{\text{ΣυνολικήΑξίαΔικτύου}}{DUA_{Zipf}}\right) \quad (9)$$

$$DUA_{Metcalf} = 15KMO(\text{ΜοναδικέςΗμερήσιεςΔιευθύνσεις}^2) \quad (10)$$

$$PMR_{Metcalf} = \ln\left(\frac{\text{ΣυνολικήΑξίαΔικτύου}}{DUA_{Metcalf}}\right) \quad (11)$$

Επίσης υπολογίζεται ο δείκτης NVM που περιγράφει την αξία του δικτύου σε σχέση με τα ανώτατα και κατώτατα όρια που παρουσιάζει και έτσι ποσοτικοποιεί τυχόν υπερεκτίμηση ή υποτίμηση, ο κανονικοποιημένος NVM θέτει ως όρια το -1 με το 1.[61]

$$\ln(NV) = \ln(\text{ΣυνολικήΑξίαΔικτύου}) \quad (12)$$

$$NVM_{UpperBound} = 15KMO(\ln(\text{ΜοναδικέςΗμερήσιεςΔιευθύνσεις}^2)) \quad (13)$$

$$NVM_{LowerBound} = 15KMO(\ln(\text{ΜοναδικέςΗμερήσιεςΔιευθύνσεις} \cdot \ln(\text{ΜοναδικέςΗμερήσιεςΔιευθύνσεις}))) \quad (14)$$

$$\ln(NVM) = \frac{(NVM_{UpperBound} + NVM_{LowerBound})}{2} \quad (15)$$

$$NVM = \ln(NV) - \ln(NVM) \quad (16)$$

$$NVM_{Normalized} = \frac{NVM}{\frac{(NVM_{UpperBound} - NVM_{LowerBound})}{2}} \quad (17)$$

Στην συνέχεια υπολογίζεται ο δείκτης NVTG και δύο παραλλαγές του που βασίζονται στους νόμος Metcalfe και Zipf. Ο δείκτης NVTG αξιολογεί το δίκτυο συναλλαγών ενός κρυπτονομίσματος, μετρώντας τον λόγο της αξίας συναλλαγών προς την ανάπτυξη του.[62, 63]

$$NVTG_M = \frac{\left(\frac{\text{ΣυνολικήΑξίαΔικτύου}}{45KMO(\text{ΗμερήσιαΑξίαΔικτύου})}\right)}{2 \cdot \text{ΜοναδικέςΗμερήσιεςΔιευθύνσεις}} \quad (18)$$

$$NVTG_Z = \frac{\left(\frac{\text{Συνολική Αξία Δικτύου}}{45KMO(\text{Ημερήσια Αξία Δικτύου})} \right)}{\ln(\text{Μοναδικές Ημερήσιες Διευθύνσεις}) + 1} \quad (19)$$

$$NVTG = \frac{\left(\frac{\text{Συνολική Αξία Δικτύου}}{45KMO(\text{Ημερήσια Αξία Δικτύου})} \right)}{45KMO(2 \cdot \text{Μοναδικές Ημερήσιες Διευθύνσεις})} \quad (20)$$

Έγιναν κάποιες διαφορετικές επιλογές στον τρόπο υπολογισμού των δεικτών οικονομικής δραστηριότητας για λόγους ομοιομορφίας, καθώς οι τρόποι υπολογισμού δεν είναι πλήρως αποσαφηνισμένοι στην βιβλιογραφία που παρουσιάζεται. Οι κινητοί μέσοι όροι εφαρμόζονται με τον μισό ορίζοντα από ότι στην βιβλιογραφία και μόνο σε προγενέστερες παρατηρήσεις, αυτή η επιλογή έγινε καθώς δεν αναφερόταν στην βιβλιογραφία αν οι κινητοί μέσοι όροι ελέγχουν μόνο προγενέστερες παρατηρήσεις ή και μεταγενέστερες. Επίσης χρησιμοποιήθηκε μόνο ο απλός κινητός μέσος όρος για ευκολότερη σύγκριση των αποτελεσμάτων στις περιπτώσεις που κάποιοι δείκτες χρησιμοποιούσαν εκθετικό κινητό μέσο όρο.

Τέλος δημιουργήθηκε μία χρονική σειρά η οποία υποδεικνύει την τάση της τιμής του Bitcoin σε σχέση με την προηγούμενη ημέρα, ο αριθμός 0 σημαίνει ότι η τιμή του Bitcoin παρέμεινε ίδια ή μειώθηκε και ο αριθμός 1 σημαίνει ότι η τιμή του Bitcoin αυξήθηκε. Η περίπτωση να έμεινε ίδια η τιμή του Bitcoin σε σχέση με την προηγούμενη ημέρα είναι σχεδόν μηδενική καθώς συγκρίνονται όλα τα δεκαδικά ψηφία της τιμής του.

4. ΠΡΟΒΛΕΨΗ ΤΗΣ ΤΙΜΗΣ ΤΟΥ BITCOIN ΜΕ ΤΝΔ

Η ανάλυση που έγινε στο προηγούμενο κεφάλαιο απέδωσε 182 χρονικές σειρές με 2763 παρατηρήσεις η κάθε μία, οι οποίες προέρχονται από τα ημερήσια δίκτυα συναλλαγών, τις γιγάντιες συνιστώσες και τα ER τυχαία δίκτυα. Στο τέταρτο στάδιο της ανάλυσης θα γίνει προσπάθεια πρόβλεψης της τιμής του Bitcoin χρησιμοποιώντας τις 182 χρονικές σειρές σε διάφορους συνδυασμούς τεχνικών νευρωνικών δικτύων.

4.1 Διαχωρισμός των χρονικών σειρών

Οι 182 χρονικές σειρές χωρίστηκαν σε 5 γκρουπ χρονικών σειρών. Το πρώτο γκρουπ έχει 4 κοινές χρονικές σειρές, το δεύτερο γκρουπ έχει 21 χρονικές σειρές οικονομικής

ανάλυσης, το τρίτο γκρουπ έχει 137 χρονικές σειρές περιγραφικής στατιστικής, το τέταρτο γκρουπ έχει 158 χρονικές σειρές που συνδυάζουν τις χρονικές σειρές περιγραφικής στατιστικής μαζί με τις χρονικές σειρές οικονομικής ανάλυσης και το πέμπτο γκρουπ έχει 21 χρονικές σειρές οικονομικής ανάλυσης οι οποίες εμπεριέχουν την τιμή του Bitcoin και δεν μπορούν να χρησιμοποιηθούν στα ΤΝΔ.[Παράρτημα]

Από τα 5 γκρουπ χρονικών σειρών θα χρησιμοποιηθούν μόνο τα 3 στα μοντέλα πρόβλεψης, οι χρονικές σειρές περιγραφικής στατιστικής, οι χρονικές σειρές οικονομικής ανάλυσης και συνδυαστικά οι χρονικές σειρές περιγραφικής στατιστικής μαζί με τις χρονικές σειρές οικονομικής ανάλυσης.

Για κάθε ένα από τα 3 επιλεγμένα γκρουπ χρονικών σειρών:

- Υπολογίστηκε η συσχέτιση μεταξύ των χρονικών σειρών, απορρίφθηκαν οι χρονικές σειρές που παρουσιάζουν συσχέτιση μεγαλύτερη από 70%.
- Υπολογίστηκε η διακύμανση των χρονικών σειρών, απορρίφθηκαν οι χρονικές σειρές που παρουσιάζουν μηδενική ή σχεδόν μηδενική διακύμανση.
- Εφαρμόστηκε ανάλυση κύριων συνιστωσών στις χρονικές σειρές με κανονικοποίηση γύρω από το 0 και με μοναδιαία διακύμανση, απορρίφθηκαν οι χρονικές σειρές που παρουσιάζουν τυπική απόκλιση μικρότερη από 1.

Συγκεντρωτικά πλέον οι χρονικές σειρές που θα δοθούν στα ΤΝΔ είναι 12 γκρουπ, 3 γκρουπ είναι οι αρχικές χρονικές σειρές, 3 γκρουπ είναι οι χρονικές σειρές χωρίς υψηλή συσχέτιση, 3 γκρουπ είναι οι χρονικές σειρές χωρίς μηδενική ή σχεδόν μηδενική διακύμανση και 3 γκρουπ είναι οι χρονικές σειρές των οποίων η ανάλυση κύριων συνιστωσών παρουσιάζουν τυπική απόκλιση μεγαλύτερη από 1.

4.2 Η δομή των ΤΝΔ

Αποφασίστηκε να δοκιμαστούν διάφορες παραμετροποιήσεις στα ΤΝΔ λόγω του μεγάλου αριθμού χρονικών σειρών που θα χρησιμοποιηθούν, ώστε να υπάρχει πληρέστερη εικόνα και επιβεβαίωση των αποτελεσμάτων από τα συγκεκριμένα δεδομένα ανεξάρτητα του ΤΝΔ.

Η πρόβλεψη και ταξινόμηση των ΤΝΔ θα γίνει για τις τελευταίες 3 ημέρες από τις 2763, η πρόβλεψη της τιμής ενός τόσο ασταθούς οικονομικού δείκτη όπως το Bitcoin έστω

και για 3 ημέρες είναι πάρα πολύ σημαντικό. Οι μόνοι οικονομικοί δείκτες που μέχρι τώρα μπορούν να προβλεφθούν αρκετά ικανοποιητικά, είναι αυτοί που σχετίζονται με παραγωγικούς τομείς της οικονομίας, οι οποίοι παρουσιάζουν μια σταθερότητα στο μεγάλο βάθος χρόνου που υπάρχουν.

Τα ΤΝΔ θα χωριστούν αρχικά σε 2 ομάδες ανάλογα με το πόσες παρατηρήσεις θα χρησιμοποιηθούν:

1. Θα χρησιμοποιηθούν και οι 2763 παρατηρήσεις. Το 99.9% των παρατηρήσεων θα χρησιμοποιηθεί για την εκπαίδευση και το υπόλοιπο 0.1% που ισούται με τις 3 τελευταίες παρατηρήσεις θα χρησιμοποιηθεί για την πρόβλεψη και ταξινόμηση.
2. Από τις 2763 παρατηρήσεις θα χρησιμοποιηθούν μόνο οι τελευταίες 90, γιατί υπάρχει η υποψία ότι οι πρώτες περίπου 2600 παρατηρήσεις που η τιμή του Bitcoin ήταν πολύ χαμηλή θα επηρεάσει την εκπαίδευση των ΤΝΔ. Το 97% των 90 παρατηρήσεων θα χρησιμοποιηθεί για την εκπαίδευση και το υπόλοιπο 3% που ισούται με τις 3 τελευταίες παρατηρήσεις θα χρησιμοποιηθεί για την πρόβλεψη και ταξινόμηση.

Οι 2 ομάδες ΤΝΔ θα ξανά χωριστούν σε 2 επιπλέον ομάδες ανάλογα με τον τρόπο που θα εκπαιδευτούν τα ΤΝΔ:

1. Θα χρησιμοποιηθεί σταθερό παράθυρο μήκους 15 ημερών, το οποίο θα διατρέχει τις παρατηρήσεις εκπαίδευσης ανά 15, για κάθε επανάληψη της διαδικασίας εκπαίδευσης των ΤΝΔ.
2. Θα χρησιμοποιηθεί αυξανόμενο παράθυρο μήκους 15 ημερών, το οποίο θα ξεκινήσει με τις πρώτες 15 παρατηρήσεις και για κάθε επανάληψη της διαδικασίας εκπαίδευσης των ΤΝΔ θα αυξάνεται κατά 1.

Συγκεντρωτικά υπάρχουν 4 ομάδες ΤΝΔ, πλήρη ημερών σταθερού παραθύρου, πλήρη ημερών αυξανόμενου παραθύρου, 90 ημερών σταθερού παραθύρου, 90 ημερών αυξανόμενου παραθύρου. Για κάθε μία από τις 4 ομάδες ΤΝΔ υπάρχουν 3 ΤΝΔ πρόβλεψης και 3 ΤΝΔ ταξινόμησης. Τα 3 διαφορετικά ΤΝΔ για την πρόβλεψη και την ταξινόμηση αντιστοιχούν σε 3 διαφορετικά μοντέλα:

1. Το μοντέλο nnet από την βιβλιοθήκη nnet, με παραμέτρους τον συνολικό αριθμό κρυφών νευρώνων και τον συντελεστή μείωσης βαρών.

2. Το μοντέλο mlp από την βιβλιοθήκη RSNNS, με παράμετρο τον συνολικό αριθμό κρυφών νευρώνων.
3. Το μοντέλο mlpWeightDecay από την βιβλιοθήκη RSNNS, με παραμέτρους τον συνολικό αριθμό κρυφών νευρώνων και τον συντελεστή μείωσης βαρών.

Συνολικά πλέον τα TND που δοκιμάστηκαν είναι 24, κάθε γκρουπ χρονικών σειρών έχει 12 για την πρόβλεψη και 12 για την ταξινόμηση. Για την ομοιομορφία των TND χρησιμοποιήθηκε η βιβλιοθήκη caret της R στην οποία διαχειρίζονται και παραμετροποιούνται όλα τα TND με τον ίδιο τρόπο ανεξάρτητα του μοντέλου στο οποίο βασίζονται.

5. ΠΕΡΙΓΡΑΦΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

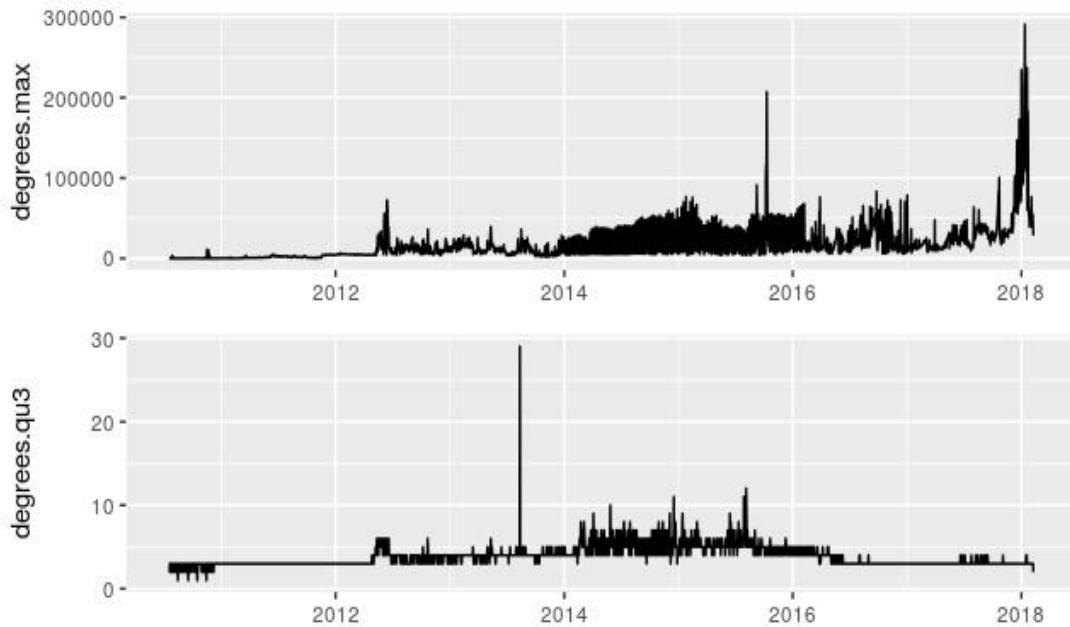
Αρχικά θα παρουσιαστούν τα σημαντικότερα αποτελέσματα των αναλύσεων της περιγραφικής στατιστικής και της οικονομικής δραστηριότητας που έγιναν στα ημερήσια δίκτυα συναλλαγών, στα ER τυχαία δίκτυα και στις γιγάντιες συνιστώσες, τα οποία παρήγαγαν τις αρχικές χρονικές σειρές. Θα γίνει εμπειρικός έλεγχος προσαρμογής σε δυναμοκατανομή, σύγκριση της εξέλιξης των ημερήσιων δικτύων συναλλαγών με τα ER τυχαία δίκτυα, διερεύνηση του φαινομένου μικρού κόσμου, της ομοιότητας βαθμών κορυφών και της συσχέτισης των χρονικών σειρών με την τιμή του Bitcoin.

Μετά θα παρουσιαστούν τα αποτελέσματα της στατιστικής ανάλυσης που έγιναν στις αρχικές χρονικές σειρές, τα οποία παρήγαγαν τις χρονικές σειρές υψηλής συσχέτισης, τις χρονικές σειρές με μηδενική ή σχεδόν μηδενική διακύμανση και την ανάλυση κύριων συνιστωσών των χρονικών σειρών.

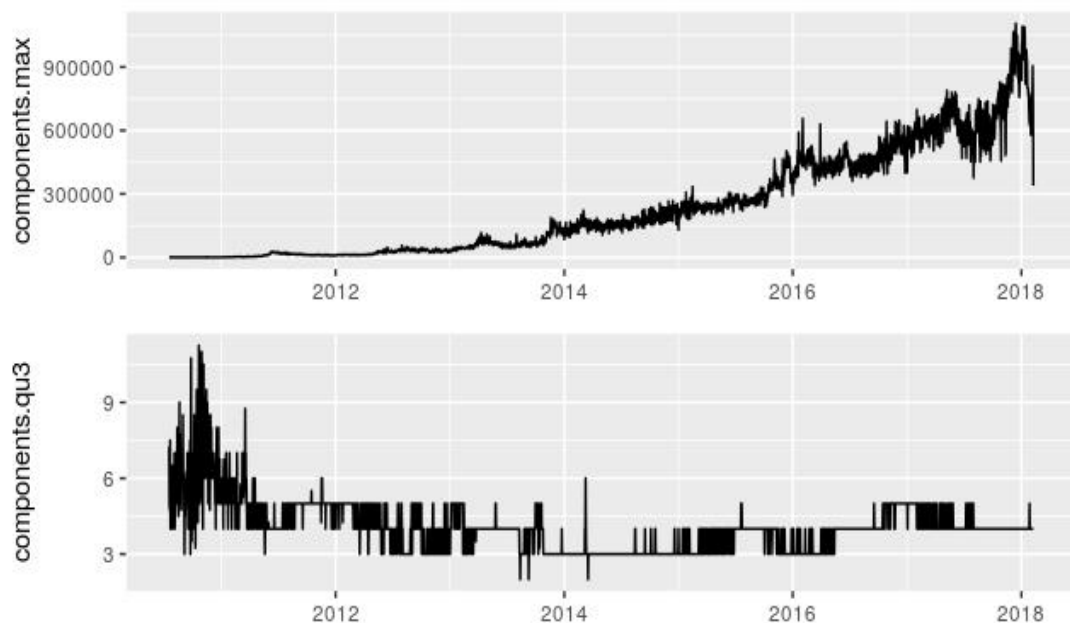
Τέλος θα παρουσιαστούν τα αποτελέσματα με την μεγαλύτερη τιμή επιτυχούς πρόβλεψης και ταξινόμησης, για τις τρεις κατηγορίες δεδομένων οι οποίες δόθηκαν στα 288 TND που έτρεξαν συνολικά. Δύο συμπεράσματα είναι κρίσιμα και πρέπει να διερευνηθούν. Πρώτον αν οι χρονικές σειρές περιγραφικής στατιστικής έχουν καλύτερη τιμή πρόβλεψης και ταξινόμησης σε σύγκριση με τις χρονικές σειρές οικονομικής δραστηριότητας. Δεύτερον αν υπάρχει βελτίωση της τιμής πρόβλεψης και ταξινόμησης όταν χρησιμοποιούνται συνδυαστικά οι χρονικές σειρές περιγραφικής στατιστικής και οικονομικής δραστηριότητας σε σχέση με τις χρονικές σειρές οικονομικής δραστηριότητας όταν είναι μόνες τους.

5.1 Χρονικές σειρές περιγραφικής στατιστικής και οικονομικής δραστηριότητας

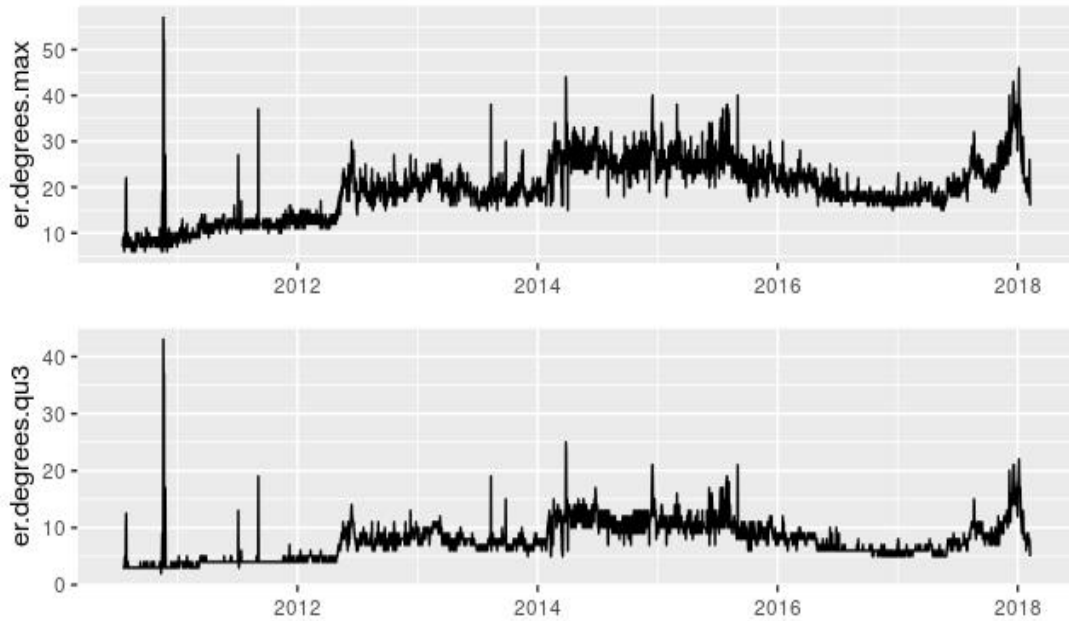
- Εμπειρικός έλεγχος προσαρμογής χρονικών σειρών σε δυναμοκατανομή.



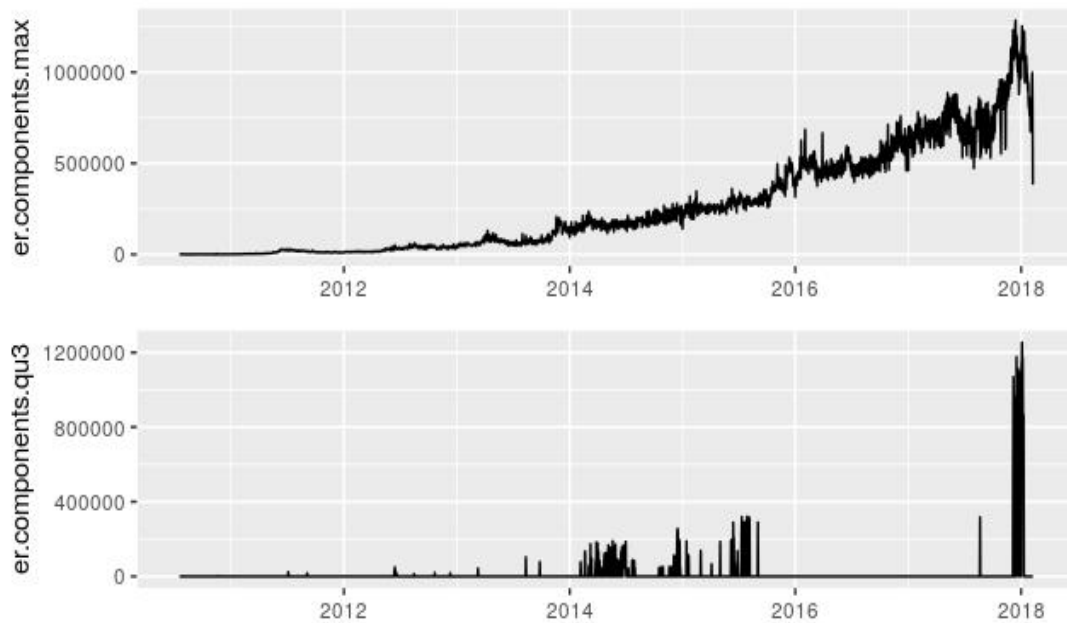
Εικόνα 1: Στο 75% του πλήθους γειτόνων των ημερήσιων δικτύων συναλλαγών οι βαθμοί κυμαίνονται γύρω στο 10, ενώ στο μέγιστο πλήθος γειτόνων των ημερήσιων δικτύων συναλλαγών οι βαθμοί είναι κοντά στις 50000. Τέτοιον βαθμού ανομοιογένεια είναι χαρακτηριστικό της δυναμοκατανομής.



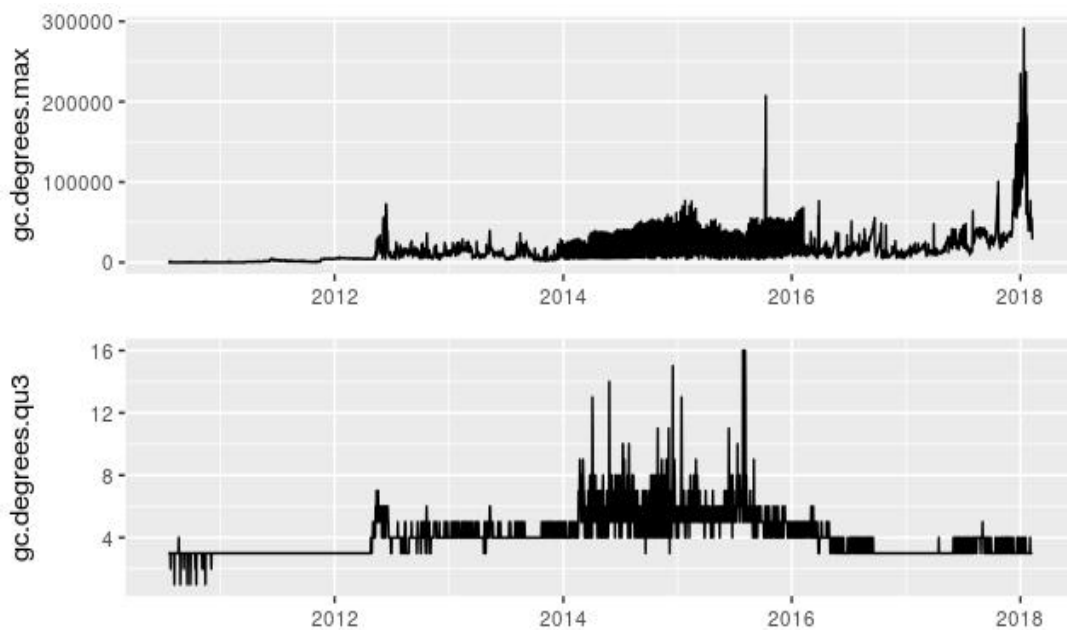
Εικόνα 2: Στο 75% των συνιστωσών των ημερήσιων δικτύων συναλλαγών οι κορυφές κυμαίνονται γύρω στο 5, ενώ στην μέγιστη συνιστώσα των ημερήσιων δικτύων συναλλαγών οι κορυφές είναι πάνω από 150000. Τέτοιου βαθμού ανομοιογένεια είναι χαρακτηριστικό της δυναμοκατανομής.



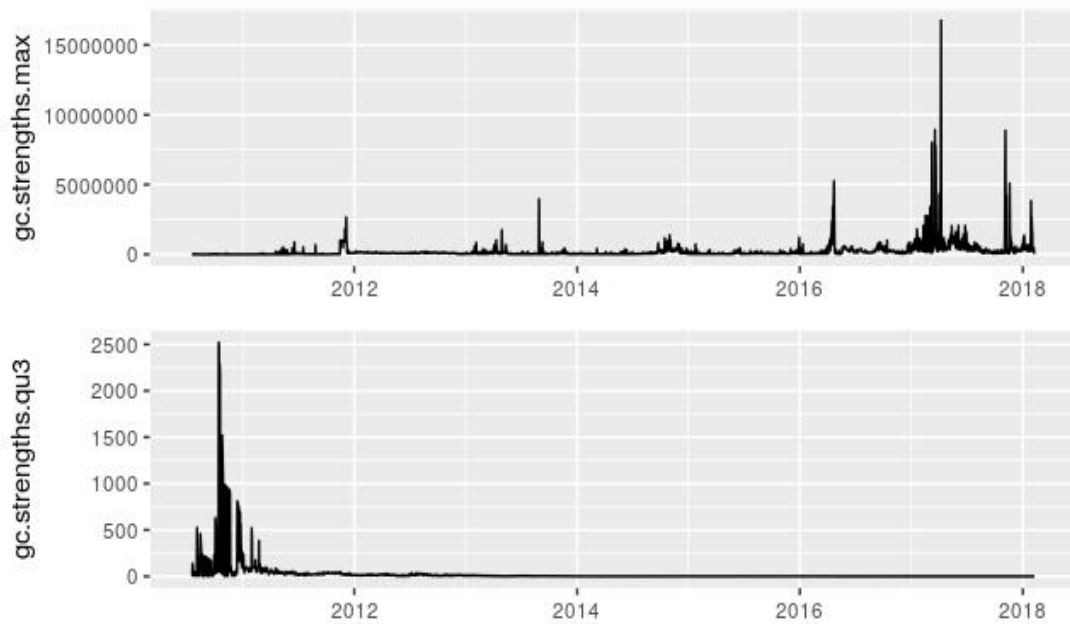
Εικόνα 3: Στο 75% του πλήθους γειτόνων των ER τυχαίων δικτύων οι βαθμοί κυμαίνονται γύρω στο 10, ενώ στο μέγιστο πλήθος γειτόνων των ER τυχαίων δικτύων οι βαθμοί είναι κοντά στο 25. Τέτοιου βαθμού ανομοιογένεια δεν είναι χαρακτηριστικό της δυναμοκατανομής.



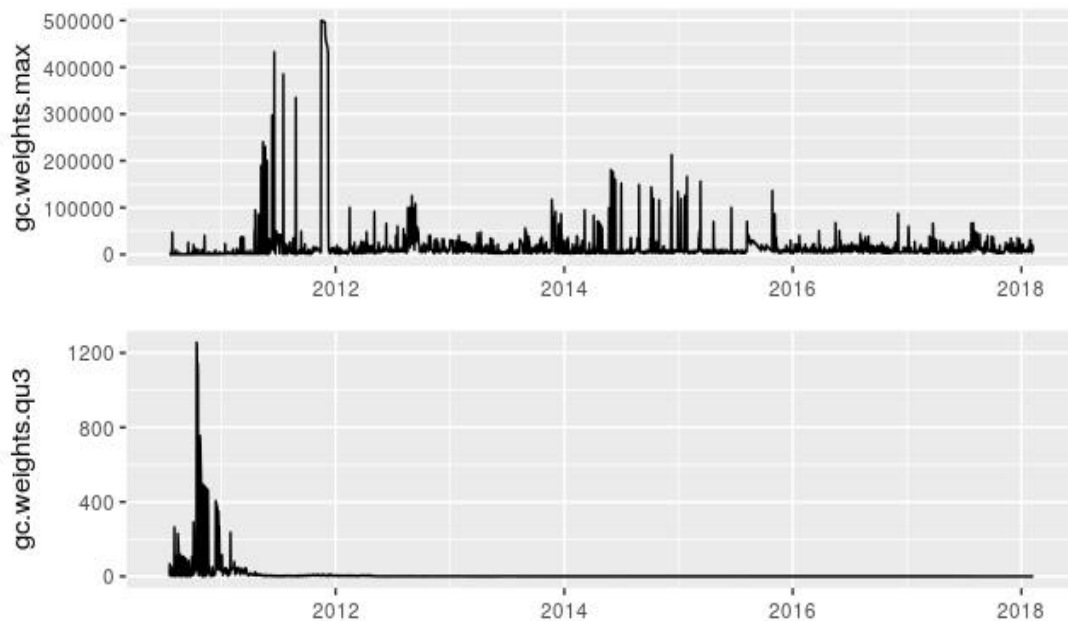
Εικόνα 4: Στο 75% των συνιστωσών των ER τυχαίων δικτύων οι κορυφές κυμαίνονται γύρω στις 200000, ενώ στην μέγιστη συνιστώσα των ER τυχαίων δικτύων οι κορυφές είναι κοντά στις 500000. Τέτοιον βαθμού ανομοιογένεια δεν είναι χαρακτηριστικό της δυναμοκατανομής.



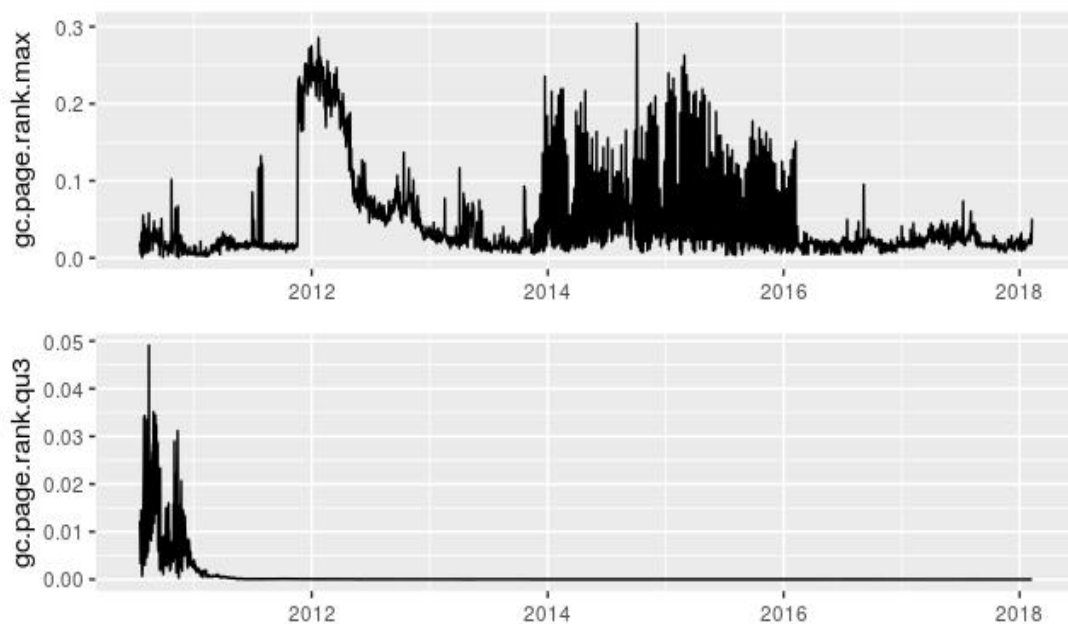
Εικόνα 5: Στο 75% του πλήθους γειτόνων των γιγάντιων συνιστωσών οι βαθμοί κυμαίνονται γύρω στο 6, ενώ στο μέγιστο πλήθος γειτόνων των γιγάντιων συνιστωσών οι βαθμοί είναι κοντά στις 50000. Τέτοιου βαθμού ανομοιογένεια είναι χαρακτηριστικό της δυναμοκατανομής.



Εικόνα 6: Το 75% της ισχύος των κορυφών των γιγάντιων συνιστωσών κυμαίνεται γύρω στο 250, ενώ η μέγιστη ισχύς των κορυφών των γιγάντιων συνιστωσών είναι κοντά στα 2500000. Τέτοιου βαθμού ανομοιογένεια είναι χαρακτηριστικό της δυναμοκατανομής.



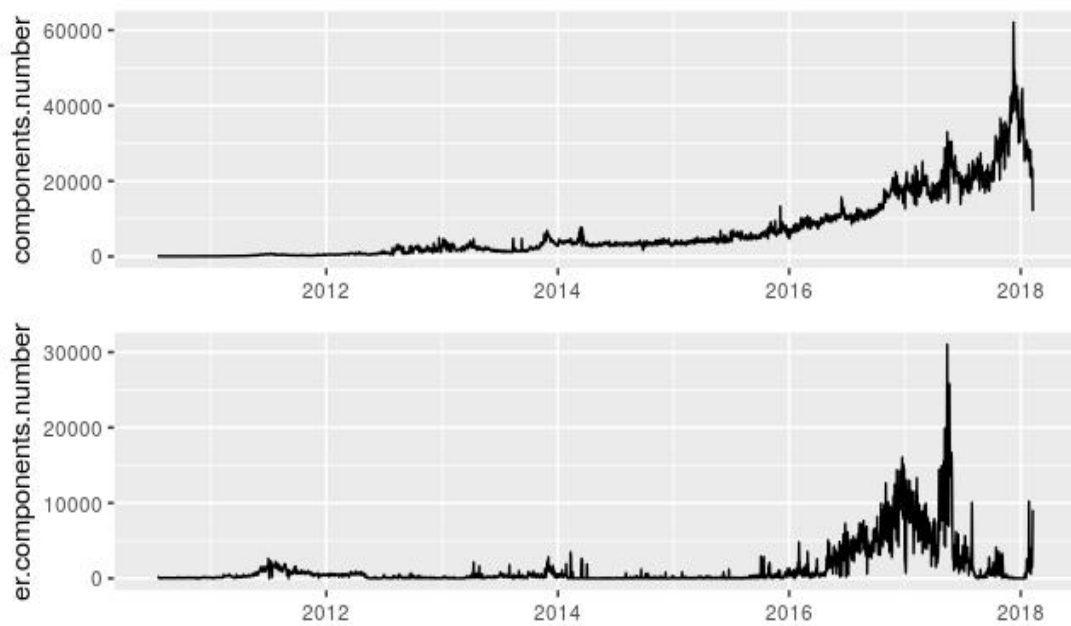
Εικόνα 7: Το 75% των βαρών των ακμών των γιγάντιων συνιστωσών κυμαίνεται γύρω στο 200, ενώ το μέγιστο βάρος των ακμών των γιγάντιων συνιστωσών είναι κοντά στις 50000. Τέτοιου βαθμού ανομοιογένεια είναι χαρακτηριστικό της δυναμοκατανομής.



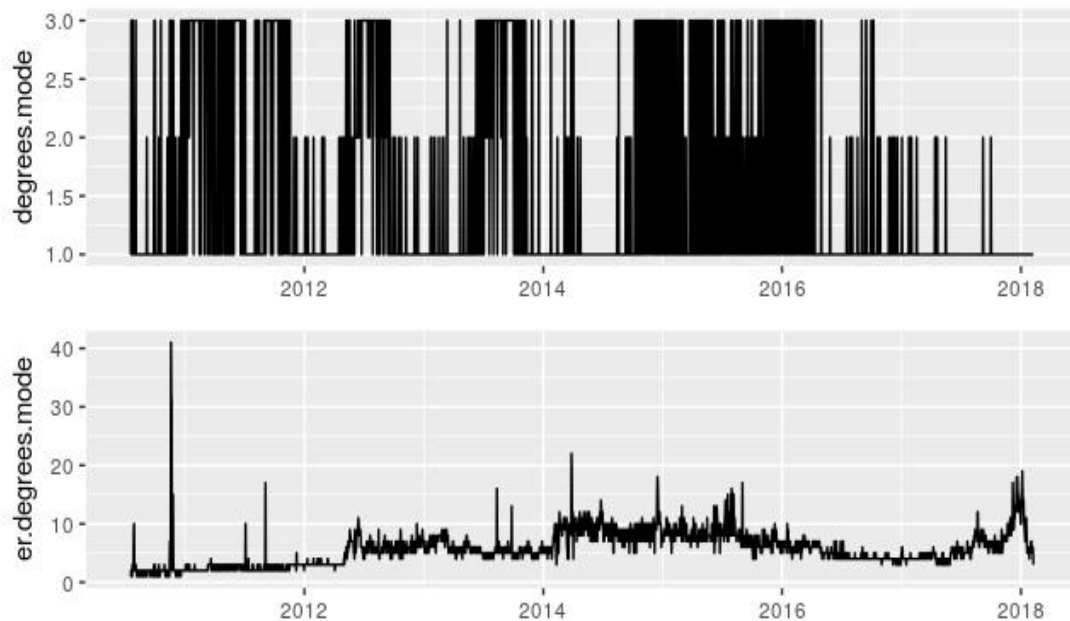
Εικόνα 8: Το 75% του κύρους των κορυφών των γιγάντιων συνιστωσών κυμαίνεται γύρω στο 0.005, ενώ το μέγιστο κύρος των κορυφών των γιγάντιων συνιστωσών είναι κοντά στο 0.1. Τέτοιου βαθμού ανομοιογένεια

μάλλον είναι χαρακτηριστικό της δυναμοκατανομής. Ακριβώς το ίδιο ισχύει και για το σταθμισμένο κύρος κορυφών των γιγάντιων συνιστωσών.

- Σύγκριση της εξέλιξης των ημερήσιων δικτύων συναλλαγών με τα Erdos-Renyi τυχαία δίκτυα όμοιου μεγέθους.



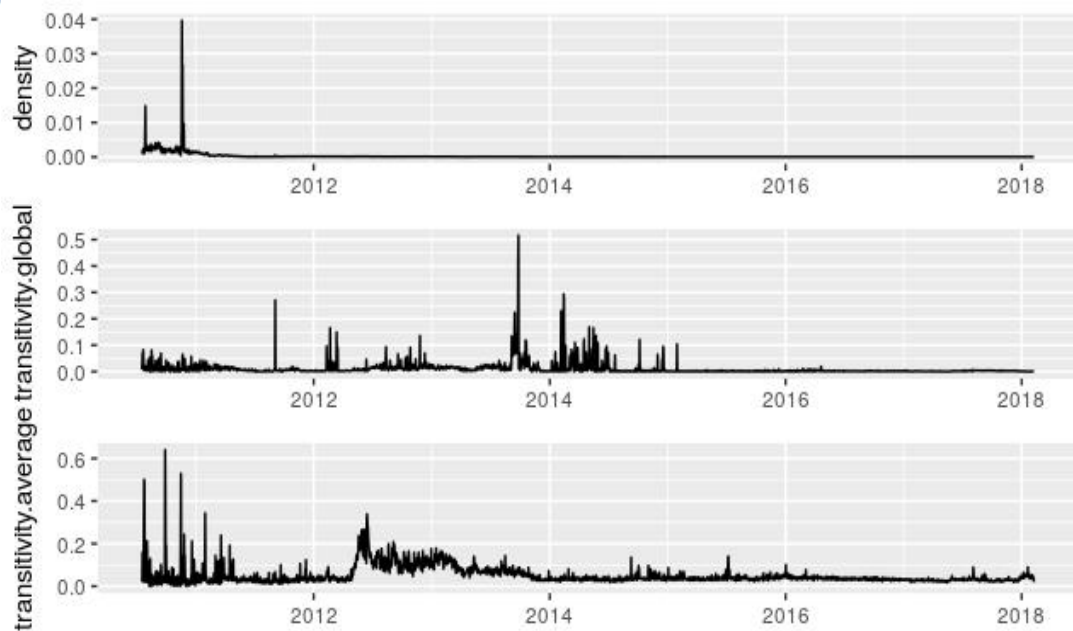
Εικόνα 9: Παρατηρείται από την σύγκριση του πλήθους των συνιστωσών των ημερήσιων δικτύων συναλλαγών με τα Erdos-Renyi τυχαία δίκτυα όμοιου μεγέθους διαφορετικός τρόπος εξέλιξης όμοιας όμως έντασης. Η συσχέτιση των δύο χρονικών σειρών είναι 0.47396.



Εικόνα 10: Παρατηρείται από την σύγκριση των πιο συχνών βαθμών των ημερήσιων δικτύων συναλλαγών με τα Erdos-Renyi τυχαία δίκτυα όμοιου μεγέθους ένας εντελώς διαφορετικός τρόπος εξέλιξης. Η συσχέτιση των δύο χρονικών σειρών είναι μόλις -0.03144 .

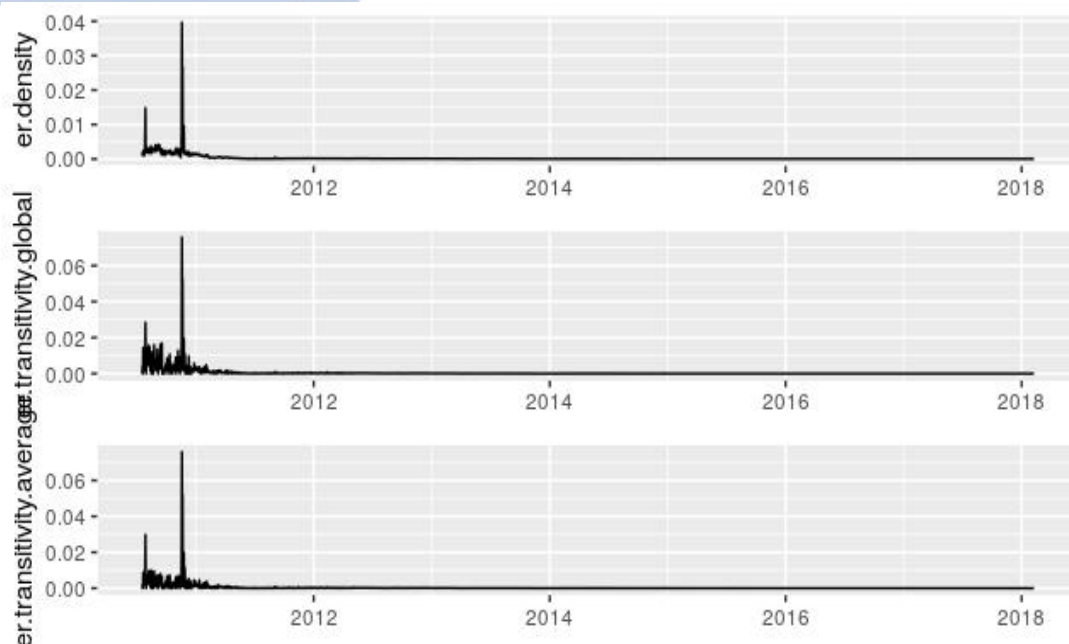
- Φαινόμενο μικρού κόσμου.

Τα ημερήσια δίκτυα συναλλαγών είναι πολύ αραιά, έχουν πυκνότητα πολύ μικρότερη από 0.25. Αν εξελισσότουσαν όπως τα ER τυχαία δίκτυα η μεταβατικότητα και ο συντελεστής σύμπλεξης θα ήταν αρκετά κοντά στην τιμή της πυκνότητας. Η συσχέτιση μεταξύ της πυκνότητας και της μεταβατικότητας είναι 0.067475. Η συσχέτιση μεταξύ της πυκνότητας και του συντελεστή σύμπλεξης είναι -0.0027222 . Η συσχέτιση μεταξύ της μεταβατικότητας και του συντελεστή σύμπλεξης είναι 0.11939. Αυτές οι τρεις συσχετίσεις δείχνουν ότι οι τιμές της μεταβατικότητας και του συντελεστή σύμπλεξης δεν πλησιάζουν ούτε μεταξύ τους ούτε με τις τιμές της πυκνότητας.



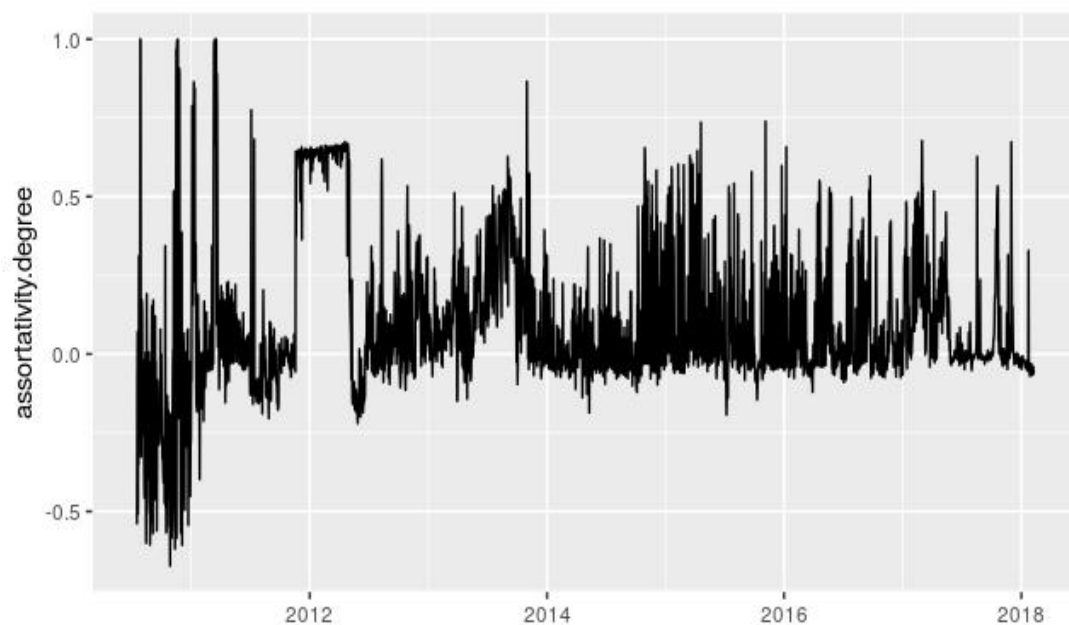
Εικόνα 11: Παρατηρείται στα ημερήσια δίκτυα συναλλαγών ότι οι χρονικές σειρές του συντελεστή σύμπλεξης και της μεταβατικότητας είναι αρκετά διαφορετικές μεταξύ τους αλλά και από την χρονική σειρά της πυκνότητας.

Στα ER τυχαία δίκτυα που δημιουργήθηκαν για αντιπαράβολή παρατηρείται ότι οι τιμές της μεταβατικότητας και του συντελεστή σύμπλεξης είναι αρκετά κοντά στην τιμή της πυκνότητας, το οποίο είναι κάτι αναμενόμενο. Η συσχέτιση μεταξύ της πυκνότητας και της μεταβατικότητας είναι 0.93886. Η συσχέτιση μεταξύ της πυκνότητας και του συντελεστή σύμπλεξης είναι 0.95536. Η συσχέτιση μεταξύ της μεταβατικότητας και του συντελεστή σύμπλεξης είναι 0.96574. Αυτές οι τρεις συσχετίσεις δείχνουν ότι οι τιμές της μεταβατικότητας και του συντελεστή σύμπλεξης πλησιάζουν και μεταξύ τους και με τις τιμές της πυκνότητας.

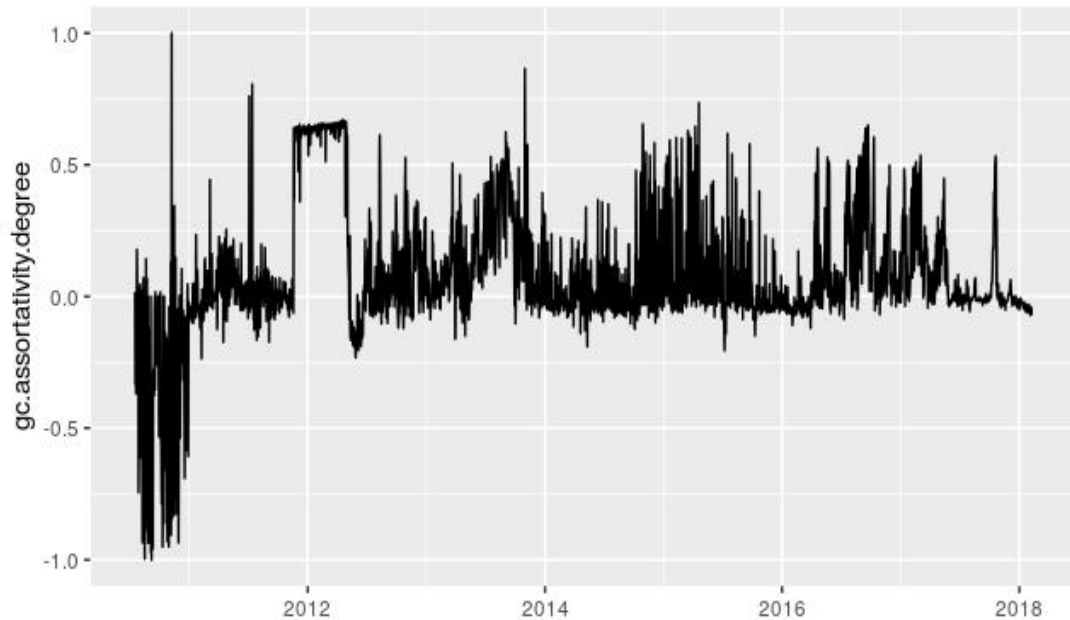


Εικόνα 12: Παρατηρείται στα ER τυχαία δίκτυα ότι οι χρονικές σειρές του συντελεστή σύμπλεξης και της μεταβατικότητας είναι αρκετά όμοιες μεταξύ τους αλλά και με την χρονική σειρά της πυκνότητας.

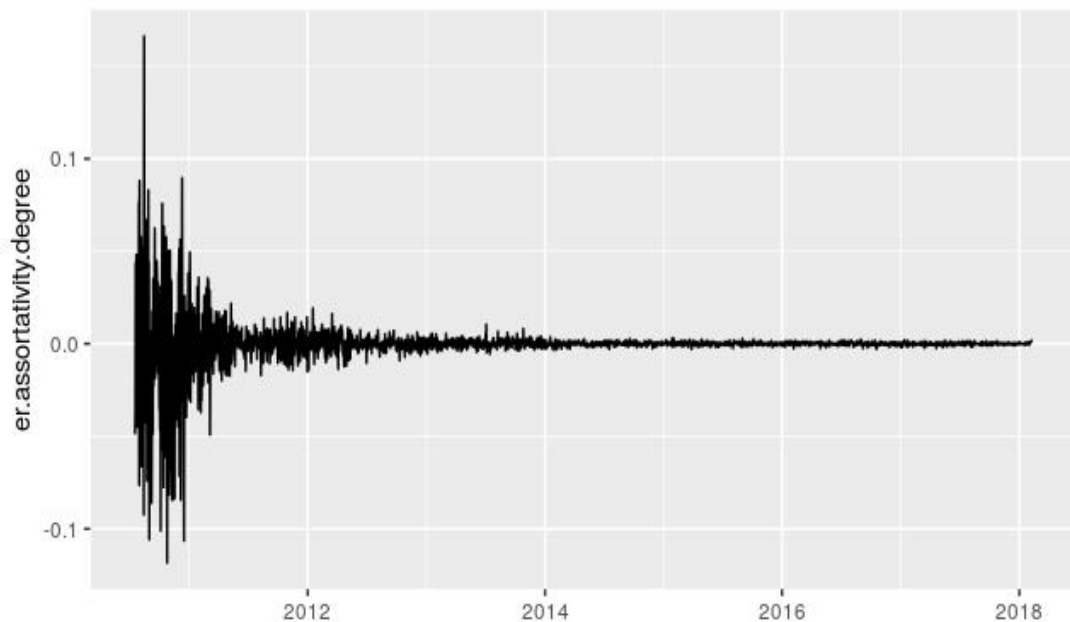
- Ομοιότητα βαθμού κορυφών.



Εικόνα 13: Ο συντελεστής ομοιότητας βαθμού των ημερήσιων δικτύων συναλλαγών μεταβάλλεται από την κατάσταση ανομοιότητας στην κατάσταση ομοιότητας.

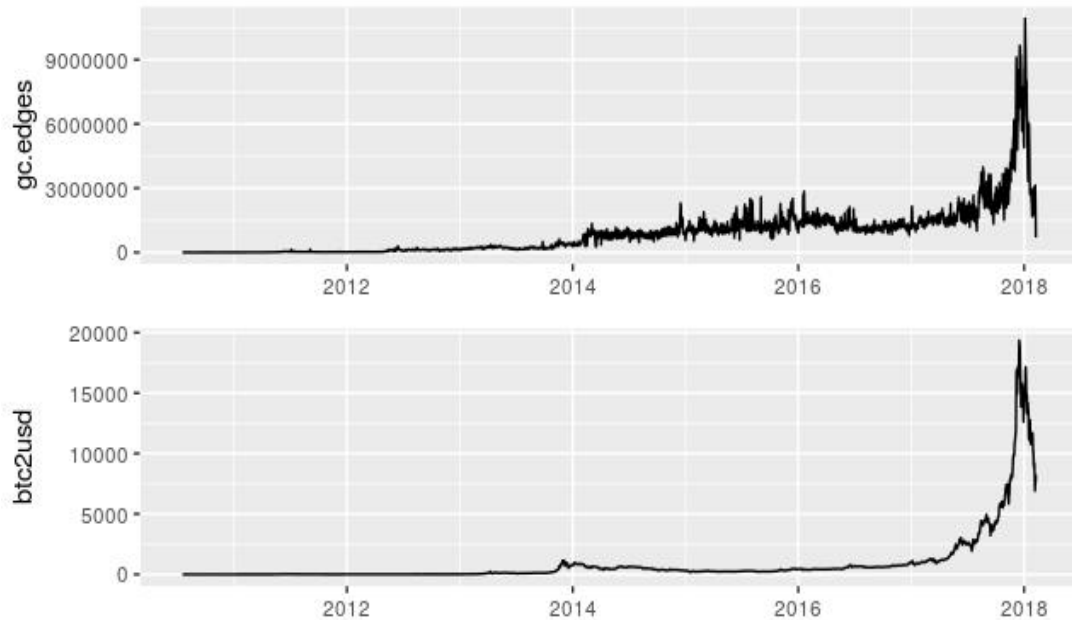


Εικόνα 14: Στον συντελεστή ομοιότητας βαθμού των γιγάντιων συνιστωσών φαίνεται πιο καθαρά η μεταβολή από την κατάσταση ανομοιότητας στην κατάσταση ομοιότητας.

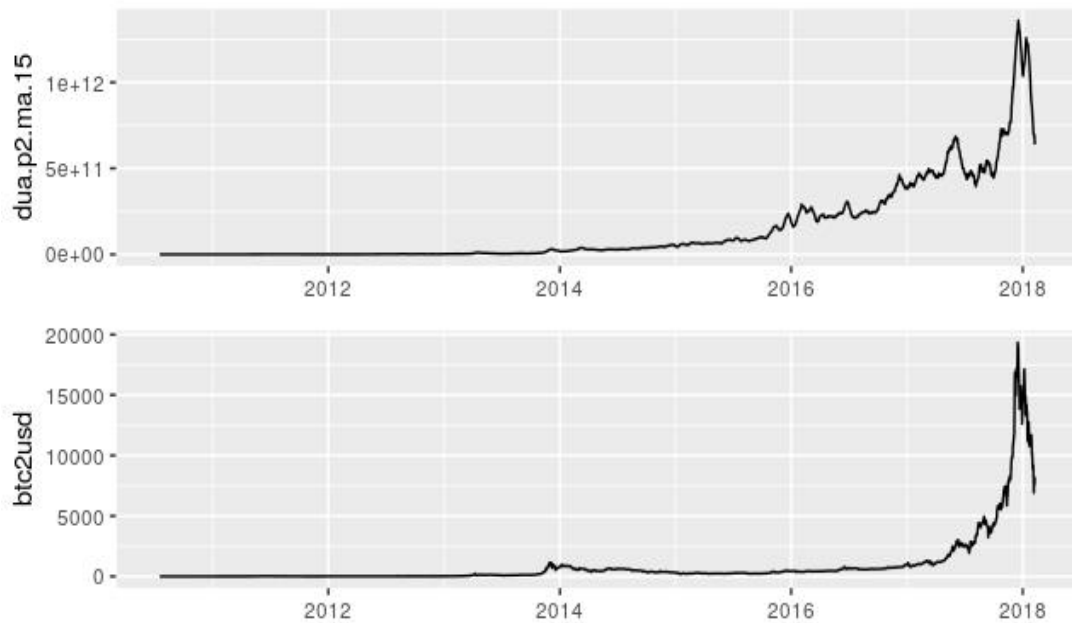


Εικόνα 15: Ο συντελεστής ομοιότητας βαθμού των ER τυχαίων δικτύων τείνει στο μηδέν και την απόλυτη ανομοιογένεια.

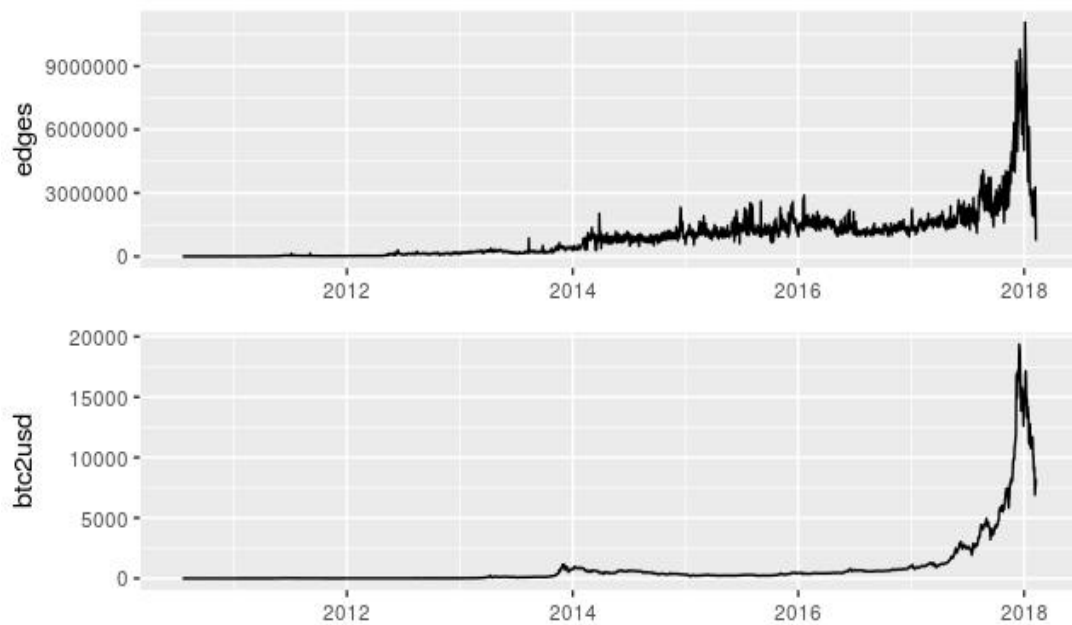
- Χρονικές σειρές οι οποίες δεν εμπεριέχουν την τιμή του Bitcoin αλλά έχουν συσχέτιση μεγαλύτερη από 0.7 με την τιμή του Bitcoin.



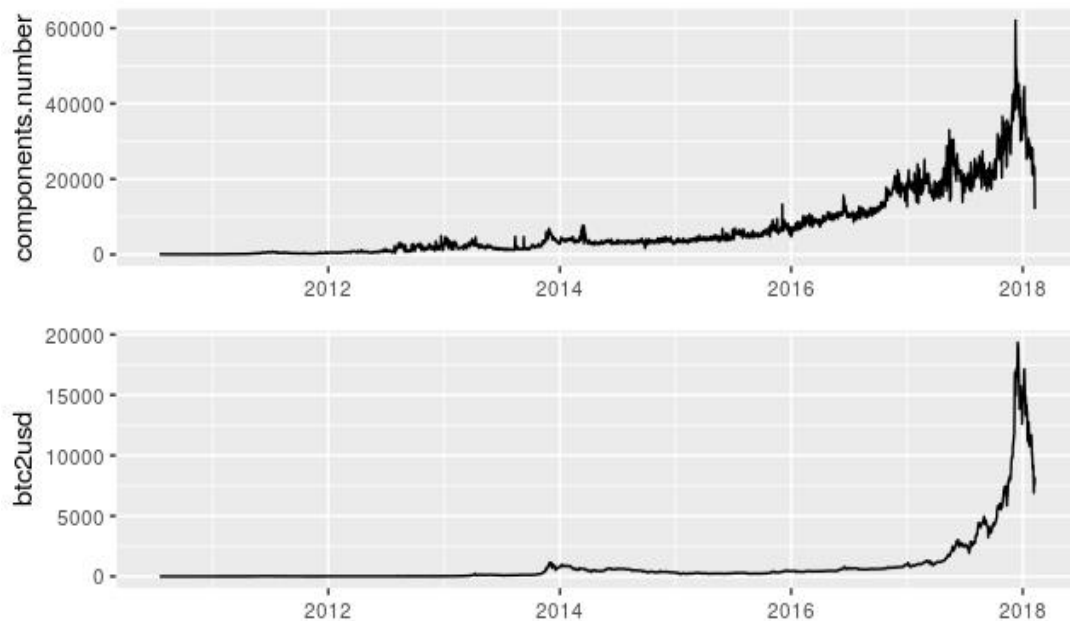
Εικόνα 16: Το πλήθος των ακμών “συναλλαγών” των γιγάντιων συνιστωσών προς την τιμή του Bitcoin με συσχέτιση 0.86210.



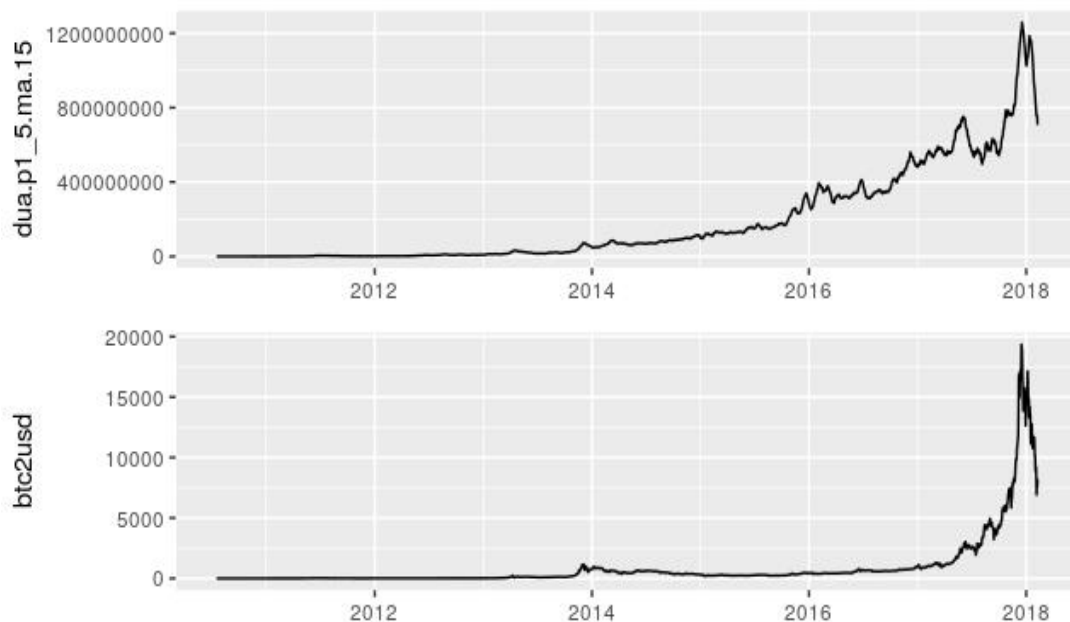
Εικόνα 17: Η χρονική σειρά του πλήθους μοναδικών κορυφών στην δύναμη 2 μέσω φίλτρου μέσου όρου 15 ημερών προς την τιμή του Bitcoin με συσχέτιση 0.86129.



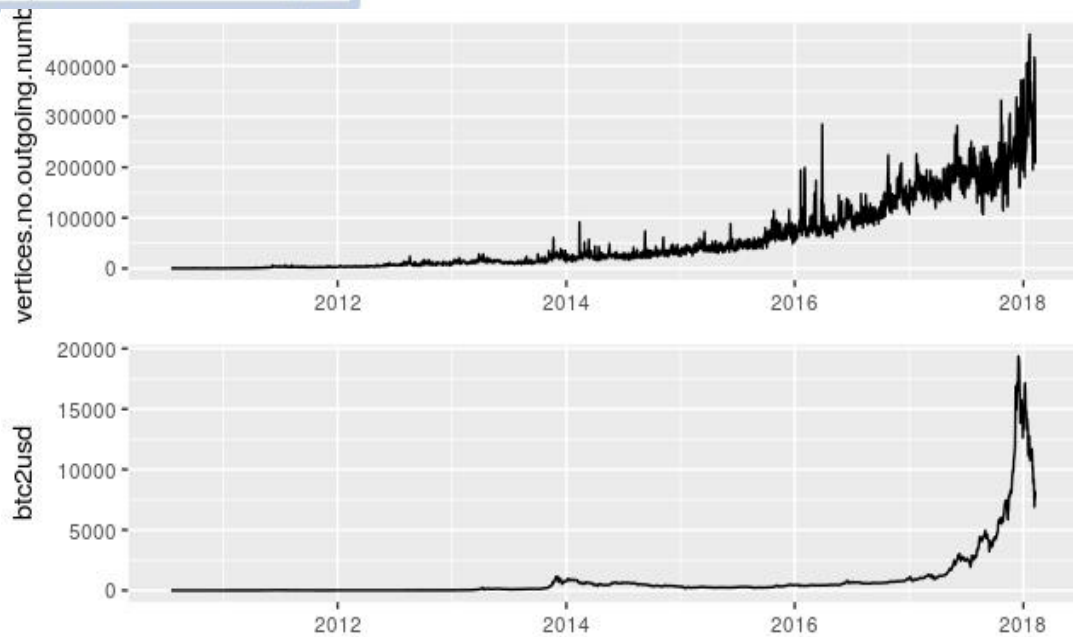
Εικόνα 18: Το πλήθος των ακμών "συναλλαγών" των ημερήσιων δικτύων συναλλαγών προς την τιμή του Bitcoin με συσχέτιση 0.86126.



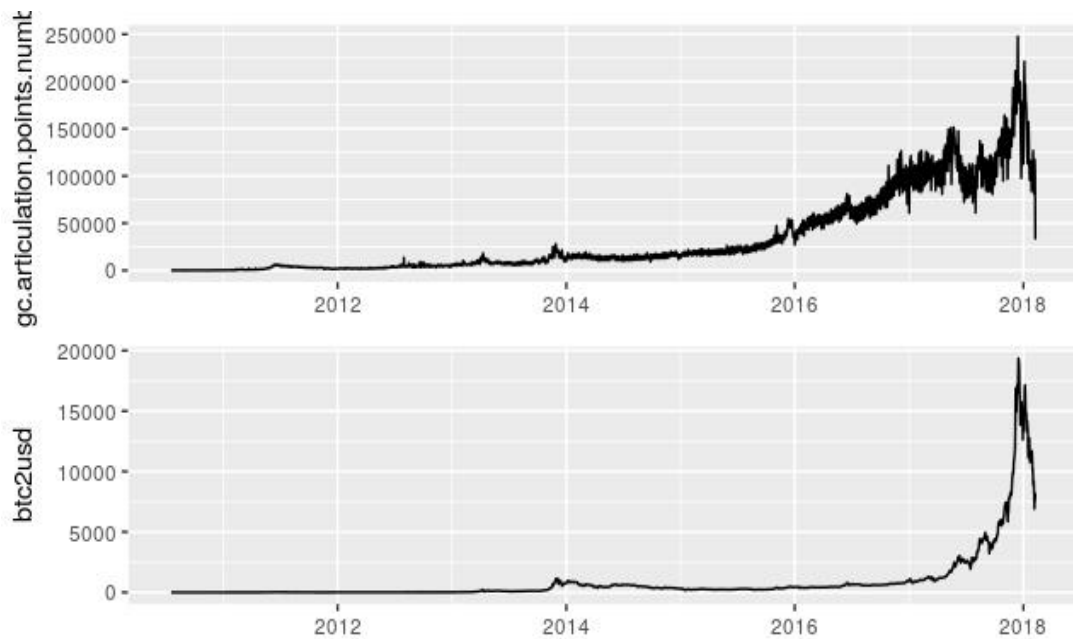
Εικόνα 19: Το πλήθος των συνιστωσών των ημερήσιων δικτύων συναλλαγών προς την τιμή του Bitcoin με συσχέτιση 0.80221.



Εικόνα 20: Η χρονική σειρά του πλήθους μοναδικών κορυφών στην δύναμη 1.5 μέσω φίλτρου μέσου όρου 15 ημερών προς την τιμή του Bitcoin με συσχέτιση 0.79313.



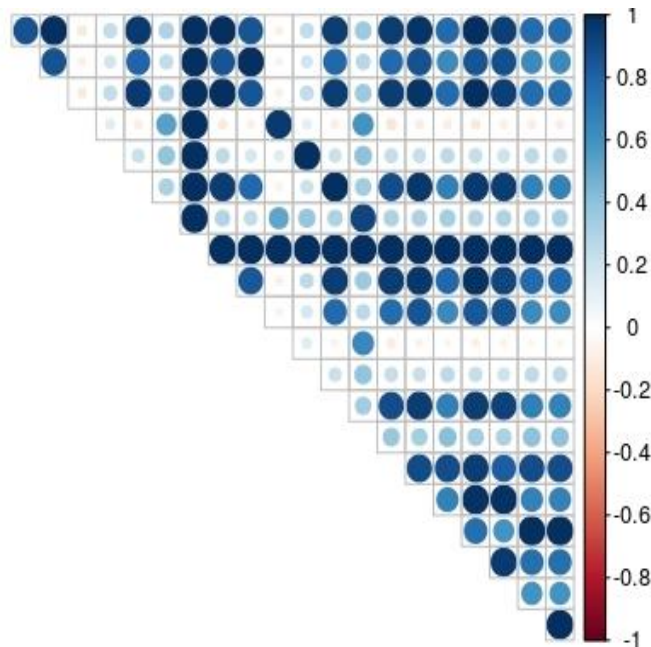
Εικόνα 21: Το πλήθος κορυφών χωρίς εξερχόμενη συναλλαγή “διευθύνσεις αποταμίευσης” ημερήσιων δικτύων συναλλαγών προς την τιμή του Bitcoin με συσχέτιση 0.73070.



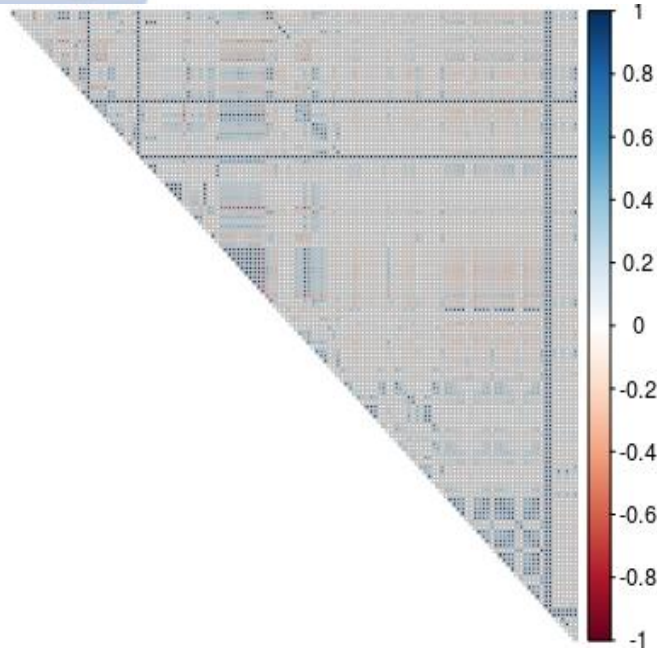
Εικόνα 22: Το πλήθος των κορυφών τομών των γιγάντιων συνιστωσών προς την τιμή του Bitcoin με συσχέτιση 0.71828.

5.2 Στατιστική ανάλυση χρονικών σειρών

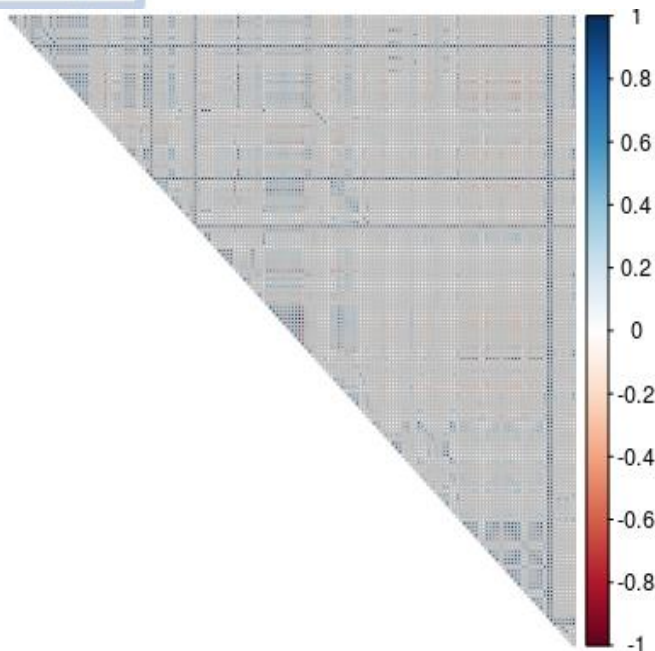
- Χρονικές σειρές υψηλής συσχέτισης.



Εικόνα 23: Άνω τριγωνικός πίνακας συσχέτισης χρονικών σειρών οικονομικής ανάλυσης. Το πλήθος των χρονικών σειρών με συσχέτιση μεγαλύτερη από 0.7 είναι 16. Αυτές οι χρονικές σειρές είναι `er.vertices.no.outgoing.number`, `vertices`, `unique.addresses`, `gc.vertices`, `dua.n_ln.ma.15`, `network.volume`, `dua.p1_5.ma.15`, `vertices.no.outgoing.number`, `gc.vertices.no.outgoing.number`, `dua.p2.ma.15`, `edges`, `dua.ln`, `nvm.upper.bound`, `gc.vertices.no.outgoing.volume`, `daily.volume`, `max.transaction.volume`.



Εικόνα 24: Άνω τριγωνικός πίνακας συσχέτισης χρονικών σειρών περιγραφικής στατιστικής. Το πλήθος των χρονικών σειρών με συσχέτιση μεγαλύτερη από 0.7 είναι 87. Αυτές οι χρονικές σειρές είναι *components.max*, *components.sd*, *components.var*, *components.mode*, *components.cv*, *degrees.min*, *degrees.median*, *er.multiple*, *er.components.qu1*, *er.components.median*, *er.components.mean*, *er.components.qu3*, *er.components.max*, *er.degrees.qu1*, *er.degrees.median*, *er.degrees.mean*, *er.degrees.qu3*, *er.degrees.max2*, *er.degrees.max*, *er.degrees.sd*, *er.degrees.var*, *er.degrees.mode2*, *er.degrees.mode*, *er.degrees.cv*, *gc.multiple*, *gc.assortativity.degree*, *gc.degrees.mean*, *gc.degrees.max2*, *gc.degrees.sd*, *gc.degrees.cv*, *gc.weights.qu1*, *gc.weights.sd*, *gc.weights.var*, *gc.strengths.min*, *gc.strengths.qu3*, *gc.strengths.sd*, *gc.strengths.var*, *gc.strengths.mode2*, *gc.strengths.cv*, *gc.page.rank.median*, *gc.page.rank.mean*, *gc.page.rank.sd*, *gc.page.rank.mode2*, *gc.page.rank.mode*, *gc.page.rank.weighted.min*, *gc.page.rank.weighted.qu1*, *gc.page.rank.weighted.median*, *gc.page.rank.weighted.mean*, *gc.page.rank.weighted.qu3*, *gc.page.rank.weighted.sd*, *gc.page.rank.weighted.var*, *gc.page.rank.weighted.mode2*, *gc.page.rank.weighted.mode*, *gc.page.rank.weighted.cv*, *gc.vertex.connectivity*, *gc.edge.connectivity*, *gc.authority.score*, *gc.articulation.points.number*, *multiple*, *degrees.max2*, *degrees.sd*, *density*, *er.density*, *er.transitivity.global*, *degrees.mean*, *er.components.sd*, *er.components.min*, *er.components.cv*, *er.transitivity.average*, *loop*, *transitivity.global*, *transitivity.average*, *degrees.qu1*, *degrees.qu3*, *degrees.max*, *gc.weights.max2*, *gc.weights.mean*, *gc.weights.min*, *gc.density*, *gc.page.rank.min*, *gc.page.rank.qu3*, *gc.page.rank.qu1*, *gc.page.rank.max2*, *gc.page.rank.max*, *gc.strengths.max*, *gc.hub.score.weighted*, *components.number*.



Εικόνα 25: Άνω τριγωνικός πίνακας συσχέτισης χρονικών σειρών περιγραφικής στατιστικής μαζί με οικονομικής ανάλυσης. Το πλήθος των χρονικών σειρών με συσχέτιση μεγαλύτερη από 0.7 είναι 106. Αυτές οι χρονικές σειρές είναι *unique.addresses*, *er.vertices.no.outgoing.number*, *gc.vertices*, *gc.edges*, *gc.vertices.no.outgoing.volume*, *network.volume*, *dua.p1_5.ma.15*, *dua.ln*, *dua.n.ln.ma.15*, *nvm.upper.bound*, *nvm.lower.bound*, *components.max*, *components.sd*, *components.var*, *components.mode*, *components.cv*, *components.number*, *degrees.min*, *degrees.median*, *er.multiple*, *er.components.qu1*, *er.components.median*, *er.components.mean*, *er.components.qu3*, *er.components.max*, *er.degrees.qu1*, *er.degrees.median*, *er.degrees.mean*, *er.degrees.qu3*, *er.degrees.max2*, *er.degrees.max*, *er.degrees.sd*, *er.degrees.var*, *er.degrees.mode2*, *er.degrees.mode*, *er.degrees.cv*, *gc.multiple*, *gc.transitivity.global*, *gc.assortativity.degree*, *gc.degrees.qu1*, *gc.degrees.mean*, *gc.degrees.max2*, *gc.degrees.sd*, *gc.degrees.cv*, *gc.weights.qu1*, *gc.weights.max2*, *gc.weights.sd*, *gc.strengths.min*, *gc.strengths.qu3*, *gc.strengths.sd*, *gc.strengths.mode2*, *gc.strengths.cv*, *gc.page.rank.median*, *gc.page.rank.mean*, *gc.page.rank.qu3*, *gc.page.rank.sd*, *gc.page.rank.mode2*, *gc.page.rank.mode*, *gc.page.rank.weighted.min*, *gc.page.rank.weighted.qu1*, *gc.page.rank.weighted.median*, *gc.page.rank.weighted.mean*, *gc.page.rank.weighted.qu3*, *gc.page.rank.weighted.sd*, *gc.page.rank.weighted.var*, *gc.page.rank.weighted.mode2*, *gc.page.rank.weighted.mode*, *gc.page.rank.weighted.cv*, *gc.vertex.connectivity*, *gc.edge.connectivity*, *gc.authority.score*, *gc.articulation.points.number*, *vertices*, *edges*, *max.transaction.volume*, *daily.volume*, *vertices.no.outgoing.number*, *gc.vertices.no.outgoing.number*, *multiple*, *degrees.max2*, *degrees.sd*, *density*, *er.density*, *er.transitivity.global*, *degrees.mean*, *er.components.sd*, *er.components.min*, *er.components.cv*, *er.transitivity.average*, *loop*, *transitivity.average*, *degrees.qu3*, *degrees.max*, *gc.max.transaction.volume*, *gc.weights.max*, *gc.daily.volume*, *gc.weights.mean*, *gc.weights.var*, *gc.weights.min*, *gc.density*, *gc.page.rank.min*, *gc.page.rank.qu1*, *gc.page.rank.max2*, *gc.page.rank.max*, *gc.strengths.max*, *gc.hub.score.weighted*.

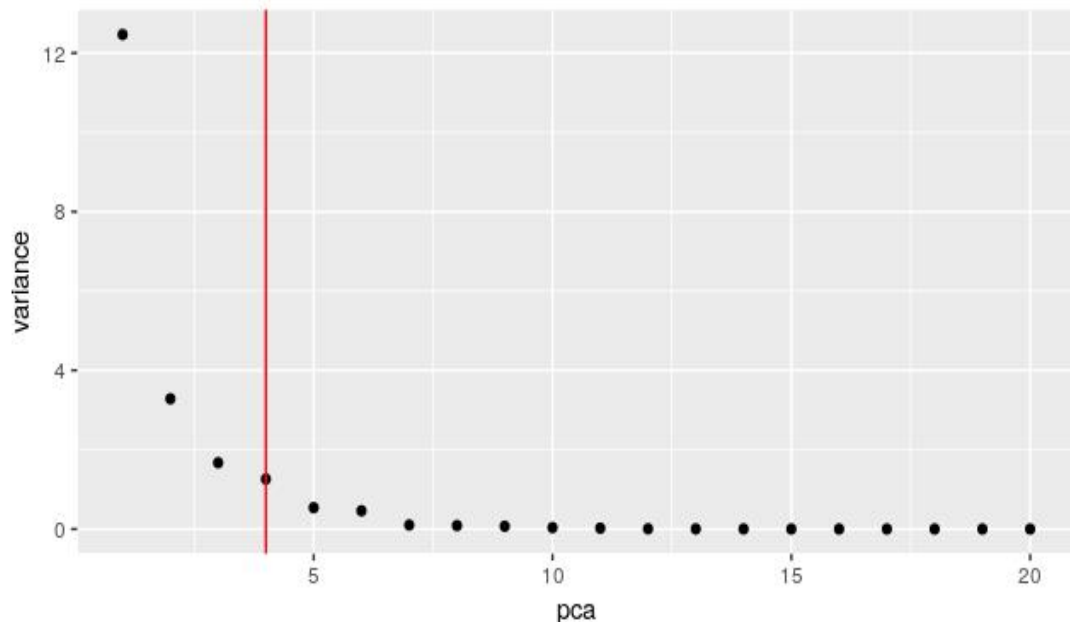
- Χρονικές σειρές με μηδενική ή σχεδόν μηδενική διακύμανση.

Το πλήθος των χρονικών σειρών οικονομικής ανάλυσης με μηδενική ή σχεδόν μηδενική διακύμανση είναι 1. Αυτή η χρονική σειρά είναι `er.vertices.no.outgoing.number`.

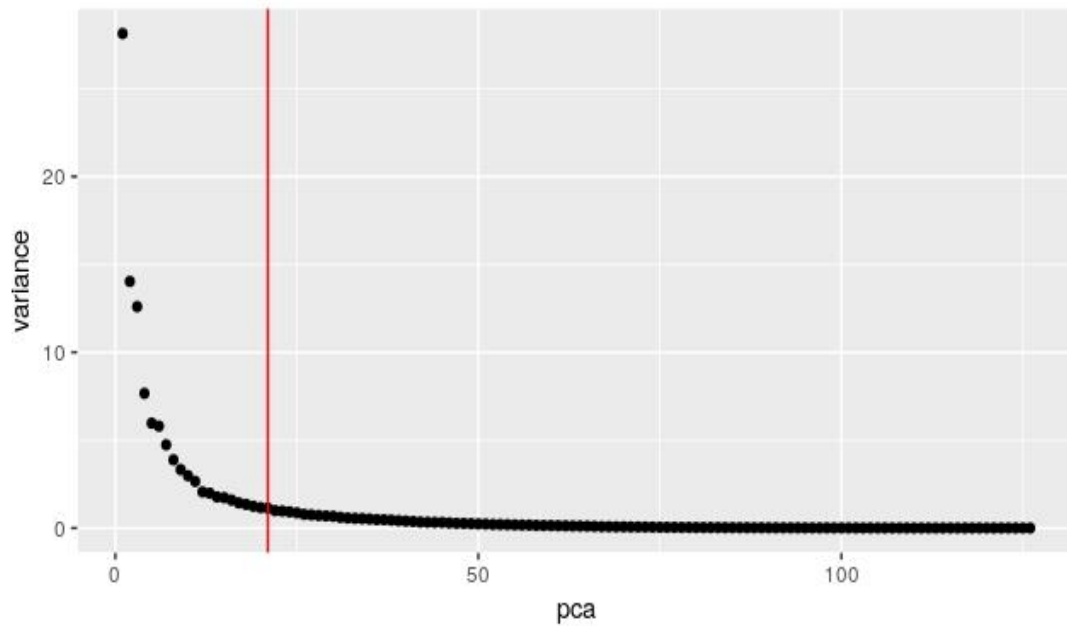
Το πλήθος των χρονικών σειρών περιγραφικής στατιστικής με μηδενική ή σχεδόν μηδενική διακύμανση είναι 11. Αυτές οι χρονικές σειρές είναι `degrees.min`, `er.multiple`, `er.components.min`, `er.components.qu1`, `er.components.median`, `er.components.qu3`, `er.components.mode`, `er.degrees.min`, `gc.degrees.min`, `gc.vertex.connectivity`, `gc.edge.connectivity`.

Το πλήθος των χρονικών σειρών περιγραφικής στατιστικής μαζί με οικονομικής ανάλυσης με μηδενική ή σχεδόν μηδενική διακύμανση είναι 12. Αυτές οι χρονικές σειρές είναι `er.vertices.no.outgoing.number`, `degrees.min`, `er.multiple`, `er.components.min`, `er.components.qu1`, `er.components.median`, `er.components.qu3`, `er.components.mode`, `er.degrees.min`, `gc.degrees.min`, `gc.vertex.connectivity`, `gc.edge.connectivity`.

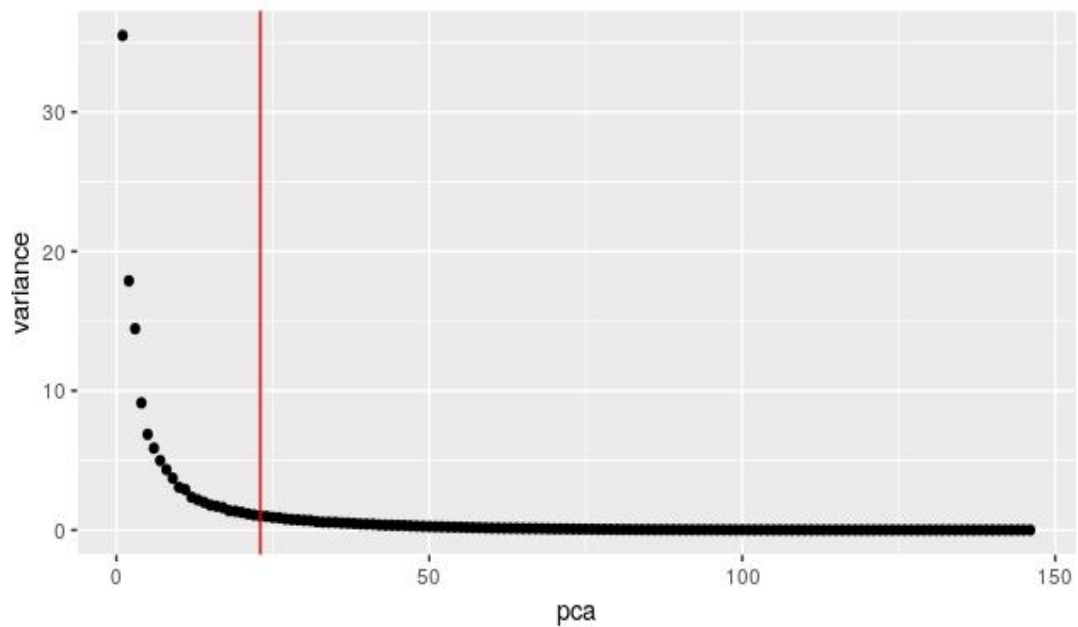
- Ανάλυση κύριων συνιστωσών των χρονικών σειρών.



Εικόνα 26: Από τις 21 χρονικές σειρές οικονομικής ανάλυσης παράχθηκαν 20 κύριες συνιστώσες και επιλέχθηκαν οι 4 πρώτες, οι οποίες έχουν τυπική απόκλιση μεγαλύτερη από 1.



Εικόνα 27: Από τις 137 χρονικές σειρές περιγραφικής στατιστικής παράχθηκαν 126 κύριες συνιστώσες και επιλέχθηκαν οι 21 πρώτες, οι οποίες έχουν τυπική απόκλιση μεγαλύτερη από 1.



Εικόνα 28: Από τις 158 χρονικές σειρές περιγραφικής στατιστικής μαζί με οικονομικής ανάλυσης παράχθηκαν 146 κύριες συνιστώσες και επιλέχθηκαν οι 23 πρώτες, οι οποίες έχουν τυπική απόκλιση μεγαλύτερη από 1.

5.3 ΤΝΔ πρόβλεψης και ταξινόμησης

Το καλύτερο αποτέλεσμα πρόβλεψης στα ΤΝΔ με δεδομένα περιγραφικής στατιστικής είναι για RMSE ίσο με 2019.5. Το ΤΝΔ που υπολόγισε το συγκεκριμένο αποτέλεσμα είναι μοντέλο nnet, με συνολικό αριθμό κρυφών νευρώνων 12 και συντελεστή μείωσης βαρών 0.001. Έχει εκπαιδευτεί στις χρονικές σειρές χωρίς υψηλή συσχέτιση, με αυξανόμενο παράθυρο μήκους 15 ημερών όπου χρησιμοποιήθηκαν και οι 2763 παρατηρήσεις.

Το καλύτερο αποτέλεσμα ταξινόμησης στα ΤΝΔ με δεδομένα περιγραφικής στατιστικής είναι για Accuracy ίσο με 0.66. Το ΤΝΔ που υπολόγισε το συγκεκριμένο αποτέλεσμα είναι μοντέλο nnet, με συνολικό αριθμό κρυφών νευρώνων 3 και συντελεστή μείωσης βαρών 0.01. Έχει εκπαιδευτεί στις χρονικές σειρές χωρίς υψηλή συσχέτιση, με σταθερό παράθυρο μήκους 15 ημερών όπου χρησιμοποιήθηκαν μόνο οι τελευταίες 90 παρατηρήσεις.

Το καλύτερο αποτέλεσμα πρόβλεψης στα ΤΝΔ με δεδομένα οικονομικής δραστηριότητας είναι για RMSE ίσο με 327.2. Το ΤΝΔ που υπολόγισε το συγκεκριμένο αποτέλεσμα είναι μοντέλο mlpWeightDecay, με συνολικό αριθμό κρυφών νευρώνων 9 και συντελεστή μείωσης βαρών 0.01. Έχει εκπαιδευτεί στις χρονικές σειρές χωρίς υψηλή συσχέτιση, με σταθερό παράθυρο μήκους 15 ημερών όπου χρησιμοποιήθηκαν μόνο οι τελευταίες 90 παρατηρήσεις.

Το καλύτερο αποτέλεσμα ταξινόμησης στα ΤΝΔ με δεδομένα οικονομικής δραστηριότητας είναι για Accuracy ίσο με 0.66. Το ΤΝΔ που υπολόγισε το συγκεκριμένο αποτέλεσμα είναι μοντέλο nnet, με συνολικό αριθμό κρυφών νευρώνων 3 και συντελεστή μείωσης βαρών 0.001. Έχει εκπαιδευτεί στις χρονικές σειρές χωρίς υψηλή συσχέτιση, με αυξανόμενο παράθυρο μήκους 15 ημερών όπου χρησιμοποιήθηκαν μόνο οι τελευταίες 90 παρατηρήσεις.

Το καλύτερο αποτέλεσμα πρόβλεψης στα ΤΝΔ με δεδομένα συνδυαστικά περιγραφικής στατιστικής μαζί με οικονομικής ανάλυσης είναι για RMSE ίσο με 1291.9. Το ΤΝΔ που υπολόγισε το συγκεκριμένο αποτέλεσμα είναι μοντέλο nnet, με συνολικό αριθμό κρυφών νευρώνων 6 και συντελεστή μείωσης βαρών 0.001. Έχει εκπαιδευτεί στις αρχικές χρονικές σειρές, με αυξανόμενο παράθυρο μήκους 15 ημερών όπου χρησιμοποιήθηκαν και οι 2763 παρατηρήσεις.

Το καλύτερο αποτέλεσμα ταξινόμησης στα ΤΝΔ με δεδομένα συνδυαστικά περιγραφικής στατιστικής μαζί με οικονομικής ανάλυσης είναι για Accuracy ίσο με 0.66. Το ΤΝΔ που υπολόγισε το συγκεκριμένο αποτέλεσμα είναι μοντέλο `mlpWeightDecay`, με συνολικό αριθμό κρυφών νευρώνων 6 και συντελεστή μείωσης βαρών 0.0001. Έχει εκπαιδευτεί στις αρχικές χρονικές σειρές, με σταθερό παράθυρο μήκους 15 ημερών όπου χρησιμοποιήθηκαν και οι 2763 παρατηρήσεις.

Εντοπίζεται ότι τα δύο από τα τρία καλύτερα ΤΝΔ πρόβλεψης έχουν εκπαιδευτεί με τις χρονικές σειρές χωρίς υψηλή συσχέτιση, με αυξανόμενο παράθυρο και με τις 2763 παρατηρήσεις. Το καλύτερο αποτέλεσμα έχει ληφθεί με δεδομένα οικονομικής δραστηριότητας, από το μοντέλο `mlpWeightDecay` το οποίο εκπαιδεύτηκε στις χρονικές σειρές χωρίς υψηλή συσχέτιση, με σταθερό παράθυρο και μόνο με τις τελευταίες 90 παρατηρήσεις.

Επιπλέον βρέθηκαν οι χρονικές σειρές που συνέβαλαν το μέγιστο στην διαδικασία της εκπαίδευσης σε περισσότερα από 10 ΤΝΔ από το σύνολο των 288 ΤΝΔ. Σε 31 ΤΝΔ είναι η `network.volume`, σε 22 ΤΝΔ είναι η `gc.degrees.var`, σε 17 ΤΝΔ είναι η `gc.degrees.sd` και η `gc.vertices.no.outgoing.volume`, σε 16 ΤΝΔ είναι η `dua.p1_5.ma.15` και η `dua.n_ln.ma.15`, σε 14 ΤΝΔ είναι η `gc.daily.volume`, σε 13 ΤΝΔ είναι η `vertices.no.outgoing.volume`, σε 12 ΤΝΔ είναι η `gc.weights.sum` και η `dua.p2.ma.15` τέλος σε 10 ΤΝΔ είναι η `er.components.max`.

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Εκτιμάται ότι είναι σημαντικά αποτελέσματα ο χαρακτηρισμός των ημερήσιων δικτύων συναλλαγών ως αυτοόμοια δίκτυα, ότι τα ημερήσια δίκτυα συναλλαγών δεν εξελίσσονται όπως τα ER τυχαία δίκτυα και ότι δεν έχουν την ιδιότητα του μικρού κόσμου. Θα ήταν δυνατό να παραχθούν περισσότερα αποτελέσματα αν δεν εμφάνιζαν προβλήματα διάφοροι αλγόριθμοι της `iGraph` στην `R`.

Επίσης σημαντική είναι η ολιστική προσέγγιση που έγινε στον τομέα της ανάλυσης δεδομένων, για την δημιουργία όσο το δυνατό περισσότερων δεδομένων, καθώς χρησιμοποιήθηκε η πλειοψηφία των διαθέσιμων εργαλείων και αλγορίθμων.

Επιπλέον αποτιμάται ως σημαντικό η προσπάθεια για πρόβλεψη της τιμής του Bitcoin και ταξινόμηση της τάσης της τιμής του Bitcoin με χρήση ΤΝΔ, καθώς δεν βρέθηκαν

εργασίες στην εγχώρια και διεθνή βιβλιογραφία με παρόμοιες προσεγγίσεις. Τα αποτελέσματα των ΤΝΔ δημιουργούν ένα ευρύ μέτρο σύγκρισης μεταξύ των δεδομένων περιγραφικής στατιστικής και των δεδομένων οικονομικής δραστηριότητας.

6.1 Χρονικές σειρές περιγραφικής στατιστικής και οικονομικής δραστηριότητας

Ο εμπειρικός έλεγχος προσαρμογής των βαθμών των ημερήσιων δικτύων συναλλαγών και των γιγάντιων συνιστωσών σε δυναμοκατανομή είναι θετικός. Το συμπέρασμα που προκύπτει είναι ότι η μορφή των ημερήσιων δικτύων συναλλαγών και των γιγάντιων συνιστωσών είναι παρόμοια, είτε εστιάζοντας σε συγκεκριμένα μέρη τους είτε παρατηρώντας τα σε μεγαλύτερη κλίμακα. Το αντίθετο συμβαίνει με τον εμπειρικό έλεγχο προσαρμογής των βαθμών των ER τυχαίων δικτύων σε δυναμοκατανομή, ο οποίος είναι αρνητικός και επιβεβαιώνει την διαφορά τους. Τα ημερήσια δίκτυα συναλλαγών και οι γιγάντιες συνιστώσες τους προκύπτει ότι είναι αυτοόμοια δίκτυα, δηλαδή ότι οι περισσότερες συναλλαγές γίνονται από λίγες διευθύνσεις.

Επιπλέον οι εμπειρικοί έλεγχοι προσαρμογής, των συνιστωσών των ημερήσιων δικτύων συναλλαγών, της ισχύος των κορυφών και των βαρών των ακμών των γιγάντιων συνιστωσών, σε δυναμοκατανομή είναι θετικοί. Ενώ οι εμπειρικοί έλεγχοι προσαρμογής, του κύρους των κορυφών και του σταθμισμένου κύρους των κορυφών των γιγάντιων συνιστωσών, σε δυναμοκατανομή τείνουν να είναι θετικοί.

Επίσης είναι σίγουρο ότι τα ημερήσια δίκτυα συναλλαγών και τα ER τυχαία δίκτυα έχουν έναν εντελώς διαφορετικό τρόπο εξέλιξης, δηλαδή τα ημερήσια δίκτυα συναλλαγών δεν μπορούν να αναπαραχθούν με έναν τυχαίο τρόπο όπως τα ER τυχαία δίκτυα. Ακόμα φαίνεται ότι τα ημερήσια δίκτυα συναλλαγών δεν έχουν την ιδιότητα του μικρού κόσμου καθώς η συσχέτιση μεταξύ της μεταβατικότητας και του συντελεστή σύμπλεξης των ημερήσιων δικτύων συναλλαγών είναι πολύ μικρή.

Όσο για τον συντελεστή ομοιότητας βαθμού κορυφών υπάρχει μία εξέλιξη με την πάροδο του χρόνου. Κορυφές μεγάλου βαθμού συνδέονται κατά την διάρκεια των πρώτων ημερών λειτουργίας του Bitcoin, με κορυφές μικρότερου βαθμού καθώς οι τιμές του συντελεστή ομοιότητας βαθμού κορυφών είναι αρνητικές. Αυτό δικαιολογείται γιατί τον πρώτο καιρό λειτουργίας του Bitcoin η εξόρυξη αντάμειβε μεγάλα ποσά bitcoins. Με την

εξέλιξη του Bitcoin και την μείωση της ανταμοιβής των εξορύξεων, αντιστρέφεται ο συντελεστής ομοιότητας βαθμού και γίνεται θετικός, πλέον κορυφές μεγάλου βαθμού συνδέονται με όμοιες κορυφές. Ο συντελεστής ομοιότητας βαθμού κορυφών μπορεί να ερμηνευτεί ως την ομοιότητα των συναλλακτικών συνηθειών που έχουν οι πλούσιες διευθύνσεις μεταξύ τους, το ίδιο ισχύει για την πλευρά των φτωχών διευθύνσεων.

Διερευνώντας τα αποτελέσματα των συσχετίσεων των χρονικών σειρών που δεν εμπεριέχουν την τιμή του Bitcoin με την χρονική σειρά της τιμής του Bitcoin, παρατηρείται ότι δεν υπάρχουν συσχετίσεις μικρότερες από -0.7 . Από τις συνολικά επτά χρονικές σειρές που έχουν συσχέτιση μεγαλύτερη από 0.7 , οι δύο είναι χρονικές σειρές των γιγάντιων συνιστωσών και οι πέντε είναι χρονικές σειρές των ημερήσιων δικτύων συναλλαγών. Επιπλέον, από τις επτά χρονικές σειρές, η μία αφορά συνιστώσες, οι δύο αφορούν ακμές και οι τέσσερις αφορούν κορυφές.

6.2 Στατιστική ανάλυση χρονικών σειρών

Η στατιστική ανάλυση εφαρμόστηκε για να δημιουργήσει περισσότερα γκρουπ χρονικών σειρών τα οποία δόθηκαν στα ΤΝΔ. Παρατηρείται όμως στα αποτελέσματα της στατιστικής ανάλυσης ότι υπάρχει μεγάλη συσχέτιση μεταξύ των χρονικών σειρών άρα και μία ποιοτική ανισότητα στα διαθέσιμα δεδομένα που θα χρησιμοποιήσουν τα ΤΝΔ.

Αυτό οφείλεται στα προβλήματα που εμφανίστηκαν με διάφορους αλγορίθμους της iGraph όπως η εύρεση της διαμέτρου, της μέσης απόστασης, του συντελεστή ANND, των κοινοτήτων και πολλών διαφόρων ακόμα. Τα αποτελέσματα των οποίων θα πρόσφεραν ένα ευρύ φάσμα ποιοτικός διαφορετικών δεδομένων και θα μείωναν με την σειρά τους την υψηλή συσχέτιση που εμφανίστηκε με την στατιστική ανάλυση. Επίσης θα ήταν αυτά που θα συνεισφέροναν το μέγιστο, καθώς η πληροφορία που δίνουν μοιάζει ποιοτικά με την πληροφορία της οικονομικής ανάλυσης.

Αξίζει να γραφτεί ότι τα προβλήματα που εμφανίστηκαν με τους διάφορους αλγορίθμους της iGraph, δημιουργήθηκαν εξαιτίας του πολύ μεγάλου αριθμού ακμών των υπό εξέταση δικτύων αλλά και στην αδυναμία των συγκεκριμένων αλγορίθμων να λειτουργήσουν με παράλληλη επεξεργασία.

6.3 ΤΝΔ πρόβλεψης και ταξινόμησης

Τα αποτελέσματα των ΤΝΔ υποδεικνύουν ότι οι χρονικές σειρές περιγραφικής στατιστικής δεν έχουν καλύτερη τιμή πρόβλεψης από τις χρονικές σειρές οικονομικής δραστηριότητας. Επίσης φαίνεται ότι αντί να υπάρχει βελτίωση της τιμής πρόβλεψης υπάρχει επιδείνωση, όταν χρησιμοποιούνται συνδυαστικά οι χρονικές σειρές περιγραφικής στατιστικής μαζί με οικονομικής δραστηριότητας σε σχέση με τις χρονικές σειρές οικονομικής δραστηριότητας όταν είναι μόνες τους. Οι καλύτερες τιμές ταξινόμησης είναι ίδιες μεταξύ των χρονικών σειρών περιγραφικής στατιστικής, των χρονικών σειρών οικονομικής δραστηριότητας και των χρονικών σειρών περιγραφικής στατιστικής μαζί με οικονομικής δραστηριότητας.

Επίσης από τις συνολικά έντεκα χρονικές σειρές που συνέβαλαν το μέγιστο στην διαδικασία της εκπαίδευσης, η μία είναι χρονική σειρά του ER τυχαίου δικτύου, οι πέντε είναι χρονικές σειρές των γιγάντιων συνιστωσών και οι υπόλοιπες πέντε είναι χρονικές σειρές των ημερήσιων δικτύων συναλλαγών. Επιπλέον, από τις έντεκα χρονικές σειρές, η μία αφορά συνιστώσες, οι τρεις αφορούν ακμές και οι επτά αφορούν κορυφές.

Γενικά από τα αποτελέσματα των ΤΝΔ συμπεραίνεται αδυναμία στην πρόβλεψη και στην ταξινόμηση της τιμής και της τάσης της τιμής του Bitcoin για 3 ημέρες. Αυτό το αποτέλεσμα όμως μπορεί να θεωρηθεί αναμενόμενο και αποδεκτό, καθώς υπάρχει τεράστια δυσκολία πρόβλεψης ακόμα και της τιμής μετοχών που παρουσιάζουν σταθερότητα για περισσότερα από 10 χρόνια. Κάθε οικονομικός δείκτης, όπου το Bitcoin είναι ένας τέτοιος, επηρεάζεται από παράγοντες που δεν αποτυπώνονται στο δίκτυο. Ένας τέτοιος παράγοντας είναι η αναλογία προσφοράς και ζήτησης, η οποία παρακολουθείται και ελέγχεται από τα ανταλλακτήρια κρυπτονομισμάτων.

ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ

Η τρέχουσα εποχή χαρακτηρίζεται από την έντονη εξέλιξη τεχνολογιών, η τεχνολογία Blockchain έχει έρθει στο προσκήνιο μέσα από μια μακρά διαδικασία ανάπτυξης γεμάτη από αποτυχίες και επιτυχίες. Η κοινωνία εισέρχεται τώρα σε αυτό που είναι γνωστό ως τέταρτη βιομηχανική επανάσταση.

Το υπόβαθρο που στηρίζει τον σύγχρονο κόσμο είναι τα μεγάλα δεδομένα, ο παράγοντας που φέρνει στο προσκήνιο το Blockchain και την εφαρμογή του. Ο σύγχρονος κόσμος παράγει μια εκπληκτική ποσότητα πληροφοριών που πρέπει να επεξεργαστούν. Δεδομένου ότι οι πληροφορίες που παράγονται είναι απεριόριστης ποικιλίας θεμάτων, δεν υπάρχει φυσικό ανθρώπινο δυναμικό ικανό να αναλύσει τέτοιες τεράστιες ποσότητες δεδομένων. Η Τεχνητή Νοημοσύνη μπορεί να αναλύσει μεγάλες ποσότητες δεδομένων και το Blockchain μπορεί να χρησιμεύσει ως αμετάβλητο θεμέλιο για την ασφαλή αποθήκευση των δεδομένων για χρήση από διάφορες βιομηχανίες, δεδομένου ότι η τέταρτη βιομηχανική επανάσταση είναι περισσότερο μια σημαντική ανακάλυψη στην ανάλυση δεδομένων.[64]

Η παρούσα έρευνα μπορεί να συνεχιστεί με την εξόρυξη των πιο πρόσφατων δεδομένων από το Bitcoin blockchain με το λογισμικό bitcoind-dump-tsv, τα οποία με την αντίστοιχη ανάλυση θα προσφέρουν τις παρατηρήσεις για το κλείσιμο της πτώσης της τιμής του Bitcoin.

Η ανάλυση που έγινε παρουσιάζει αστοχία στις τιμές των παρατηρήσεων μερικών χρονικών σειρών οι οποίες ξεπέρασαν το όριο των 32bit μεταβλητών ακέραιων αριθμών που έχει η R. Για να μπορέσει να διορθωθεί αυτή η αστοχία χρειάζεται από το επίπεδο της εξόρυξης δεδομένων κιόλας, να αντικατασταθούν οι μεταβλητές ακέραιων αριθμών από μεταβλητές διπλής ακρίβειας. Ακόμα όμως και αυτή η ενέργεια επιφυλάσσει την επανεμφάνιση της ίδια αστοχίας, καθώς μετά από κάποιους υπολογισμούς ή την εφαρμογή κάποιου αλγορίθμου μπορεί να επιστραφεί μεταβλητή ακέραιου αριθμού. Αν δεν γίνει εφικτό να διορθωθεί αυτό το σφάλμα και να εξασφαλιστεί ότι δεν θα ξανά εμφανιστεί, χρειάζεται να μεταφερθεί η συγκεκριμένη ανάλυση κατά προτίμηση στην Python η οποία υποστηρίζει την βιβλιοθήκη iGraph.

Επίσης είναι επιτακτική η βελτίωση των αλγορίθμων της iGraph για να υποστηρίζουν παράλληλη επεξεργασία, με αυτήν την δυνατότητα θα μπορέσουν να τρέξουν οι αλγόριθμοι υπολογισμού αποστάσεων, μοτίβων, κοινοτήτων, συντελεστή ANND, μέτρων κεντρικότητας, κλικών, μέγιστων υπογραφημάτων, ελέγχων προσαρμογής. Οι οποίοι δεν μπόρεσαν να τρέξουν λόγω του μεγάλου όγκου δεδομένων.

Στην ανάλυση οικονομικής δραστηριότητας μπορούν να υπολογιστούν μερικές ακόμα χρονικές σειρές. Όπως το πλήθος Bitcoin που καταστρέφονται ανά ημέρα, το πλήθος Bitcoin

από ρέστα ανά ημέρα, το πλήθος διευθύνσεων που καταστρέφονται ανά ημέρα και το πλήθος διευθύνσεων που δημιουργούνται ανά ημέρα. Αυτές οι χρονικές σειρές χρειάζονται κατά τον υπολογισμό τους να έχουν πρόσβαση σε τιμές που αφορούν την κατάσταση του δικτύου συναλλαγών της προηγούμενης ημέρας.

Επιπλέον μπορούν να υπολογιστούν οι χρονικές σειρές των κινούμενων μέσων όρων για 50, 100 και 200 ημέρες της χρονικής σειράς της τιμής του Bitcoin. Αυτοί οι κινούμενοι μέσοι όροι ελέγχουν αν υπάρχει υπερτίμηση ή υποτίμηση στην τιμή του οικονομικού δείκτη όταν διασταυρώνονται μεταξύ τους. Στο ίδιο μοτίβο με την προηγούμενη πρόταση είναι οι δοκιμές με διαφορετικά φίλτρα χρονικών σειρών αντί του κινούμενου μέσου όρου αλλά και του μήκους των ημερών τους.

Μετά θα μπορούσαν να υπολογιστούν για επιλεγμένες χρονικές σειρές διάφοροι λογάριθμοι, μέσες τιμές και τετράγωνα. Όπως και ο μετασχηματισμός Fourier, η συνδιακύμανση, η διασυσχέτιση, η διασυνδιακύμανση και η αμοιβαία πληροφορία. Η εφαρμογή του μοντέλου ARMA και των διάφορων παραλλαγών του για στοχαστικές διαδικασίες θα μπορούσε να δώσει ενδιαφέρον αποτέλεσμα.

Κλείνοντας, στα ΤΝΔ θα μπορούσαν να χρησιμοποιηθούν μόνο οι χρονικές σειρές που συνεισφέρουν περισσότερο από 70% στην διαδικασία της εκπαίδευσης των ΤΝΔ. Η σύγκριση αυτών των αποτελεσμάτων με τα υπάρχοντα αποτελέσματα θα υποδείξει αν μπορεί να υπάρξει εξοικονόμηση δεδομένων, άρα και πόρων. Παρόμοια θα μπορούσαν να χρησιμοποιηθούν μόνο οι χρονικές σειρές που παρουσιάζουν συσχέτιση με την τιμή του Bitcoin μεγαλύτερη από 0.7.

Τέλος θα ήταν χρήσιμο να δοκιμαστούν πιο εξελιγμένα ΤΝΔ με μεγαλύτερες δυνατότητες παραμετροποίησης και υπολογισμού, ενδεικτικά είναι τα Long short term memory(LSTM) και Recurrent Neural Network(RNN), αλλά και τα Graph Convolutional Networks(GCN) τα οποία δέχονται ως δεδομένα εισόδου δίκτυα.



ΠΑΡΑΡΤΗΜΑ

Παρατίθενται τα ονόματα των χρονικών σειρών ανά γκρουπ, όπως είναι στον κώδικα της ανάλυσης, μαζί με την εννοιολογική μετάφραση τους στα Ελληνικά:

1. 4 κοινές χρονικές σειρές:
 - datetime(ημέρα δικτύου σε αναγνώσιμη μορφή)
 - timestamp(ημέρα δικτύου σε αριθμητική μορφή)
 - btc2usd(τιμή ενός bitcoin σε δολάρια)
 - price.trend(η τάση της τιμής σε σχέση με την προηγούμενη μέρα)
2. 21 χρονικές σειρές οικονομικής ανάλυσης:
 - vertices(πλήθος κορυφών "διευθύνσεων" ημερήσιων δικτύων συναλλαγών)
 - edges(πλήθος ακμών "συναλλαγών" ημερήσιων δικτύων συναλλαγών)
 - unique.addresses(πλήθος μοναδικών κορυφών "διευθύνσεων" ημερήσιων δικτύων συναλλαγών)
 - max.transaction.volume(ποσότητα μέγιστης συναλλαγής bitcoins ημερήσιου δικτύου συναλλαγών)
 - daily.volume(συνολική ποσότητα bitcoins ημερήσιου δικτύου συναλλαγών)
 - vertices.no.outgoing.number(πλήθος κορυφών χωρίς εξερχόμενη συναλλαγή "διευθύνσεις αποταμίευσης" ημερήσιου δικτύου συναλλαγών)
 - vertices.no.outgoing.volume(συνολική ποσότητα bitcoins κορυφών χωρίς εξερχόμενη συναλλαγή "διευθύνσεις αποταμίευσης" ημερήσιου δικτύου συναλλαγών)
 - er.vertices.no.outgoing.number(πλήθος κορυφών χωρίς εξερχόμενη συναλλαγή "διευθύνσεις αποταμίευσης" ER δικτύου)
 - gc.vertices(πλήθος κορυφών "διευθύνσεων" γιγάντιας συνιστώσας)
 - gc.edges(πλήθος ακμών "συναλλαγών" γιγάντιας συνιστώσας)
 - gc.max.transaction.volume(ποσότητα μέγιστης συναλλαγής bitcoins γιγάντιας συνιστώσας)
 - gc.daily.volume(συνολική ποσότητα bitcoins γιγάντιας συνιστώσας)

- `gc.vertices.no.outgoing.number`(πλήθος κορυφών χωρίς εξερχόμενη συναλλαγή "διευθύνσεις αποταμίευσης" γιγάντιας συνιστώσας)
 - `gc.vertices.no.outgoing.volume`(συνολική ποσότητα bitcoins κορυφών χωρίς εξερχόμενη συναλλαγή "διευθύνσεις αποταμίευσης" γιγάντιας συνιστώσας)
 - `network.volume`(συνολική ποσότητα bitcoins)
 - `dua.p1_5.ma.15`(πλήθος μοναδικών κορυφών στην δύναμη 1,5 μέσω φίλτρου μέσου όρου 15 ημερών)
 - `dua.ln`(φυσικός λογάριθμος πλήθους μοναδικών κορυφών)
 - `dua.n_ln.ma.15`(γινόμενο πλήθους μοναδικών κορυφών με `dua.ln` μέσω φίλτρου μέσου όρου 15 ημερών)
 - `dua.p2.ma.15`(πλήθος μοναδικών κορυφών στην δύναμη 2 μέσω φίλτρου μέσου όρου 15 ημερών)
 - `nvm.upper.bound`(φυσικός λογάριθμος πλήθους μοναδικών κορυφών στην δύναμη 2 μέσω φίλτρου μέσου όρου 15 ημερών)
 - `nvm.lower.bound`(φυσικός λογάριθμος γινομένου πλήθους μοναδικών κορυφών με `dua.ln` μέσω φίλτρου μέσου όρου 15 ημερών)
3. 137 χρονικές σειρές περιγραφικής στατιστικής:
- `multiple`(πλήθος πολλαπλών ακμών ημερήσιων δικτύων συναλλαγών)
 - `loop`(πλήθος βρόγχων ημερήσιων δικτύων συναλλαγών)
 - `density`(πυκνότητα ημερήσιων δικτύων συναλλαγών)
 - `transitivity.global`(μεταβατικότητα ημερήσιων δικτύων συναλλαγών)
 - `transitivity.average`(συντελεστής σύμπλεξης ημερήσιων δικτύων συναλλαγών)
 - `assortativity.degree`(συντελεστής ομοιότητας βαθμού κορυφών ημερήσιων δικτύων συναλλαγών)
 - `components.min`(ελάχιστο συνιστωσών ημερήσιων δικτύων συναλλαγών)
 - `components.qul`(πρώτο τέταρτο συνιστωσών ημερήσιων δικτύων συναλλαγών)
 - `components.median`(διάμεσος συνιστωσών ημερήσιων δικτύων συναλλαγών)
 - `components.mean`(μέση τιμή συνιστωσών ημερήσιων δικτύων συναλλαγών)

- `components.qu3`(τρία τέταρτα συνιστωσών ημερήσιων δικτύων συναλλαγών)
- `components.max2`(δεύτερη μεγαλύτερη συνιστώσα ημερήσιων δικτύων συναλλαγών)
- `components.max`(μέγιστο συνιστωσών ημερήσιων δικτύων συναλλαγών)
- `components.sd`(τυπική απόκλιση συνιστωσών ημερήσιων δικτύων συναλλαγών)
- `components.var`(διακύμανση συνιστωσών ημερήσιων δικτύων συναλλαγών)
- `components.mode2`(δεύτερη πιο συχνή συνιστώσα ημερήσιων δικτύων συναλλαγών)
- `components.mode`(συχνότερη συνιστώσα ημερήσιων δικτύων συναλλαγών)
- `components.cv`(σχετική τυπική απόκλιση συνιστωσών ημερήσιων δικτύων συναλλαγών)
- `components.number`(πλήθος συνιστωσών ημερήσιων δικτύων συναλλαγών)
- `degrees.min`(ελάχιστο πλήθος γειτόνων ημερήσιων δικτύων συναλλαγών)
- `degrees.qu1`(πρώτο τέταρτο πλήθους γειτόνων ημερήσιων δικτύων συναλλαγών)
- `degrees.median`(διάμεσος πλήθους γειτόνων ημερήσιων δικτύων συναλλαγών)
- `degrees.mean`(μέση τιμή πλήθους γειτόνων ημερήσιων δικτύων συναλλαγών)
- `degrees.qu3`(τρία τέταρτα πλήθους γειτόνων ημερήσιων δικτύων συναλλαγών)
- `degrees.max2`(δεύτερο μεγαλύτερο πλήθος γειτόνων ημερήσιων δικτύων συναλλαγών)
- `degrees.max`(μέγιστο πλήθος γειτόνων ημερήσιων δικτύων συναλλαγών)
- `degrees.sd`(τυπική απόκλιση πλήθους γειτόνων ημερήσιων δικτύων συναλλαγών)
- `degrees.var`(διακύμανση πλήθους γειτόνων ημερήσιων δικτύων συναλλαγών)
- `degrees.mode2`(δεύτερο πιο συχνό πλήθος γειτόνων ημερήσιων δικτύων συναλλαγών)
- `degrees.mode`(συχνότερο πλήθος γειτόνων ημερήσιων δικτύων συναλλαγών)
- `degrees.cv`(σχετική τυπική απόκλιση πλήθους γειτόνων ημερήσιων δικτύων συναλλαγών)
- `er.multiple`(πλήθος πολλαπλών ακμών ER δικτύου)
- `er.loop`(πλήθος βρόγχων ER δικτύου)
- `er.density`(πυκνότητα ER δικτύου)
- `er.transitivity.global`(μεταβατικότητα ER δικτύου)
- `er.transitivity.average`(συντελεστής σύμπλεξης ER δικτύου)

- `er.assortativity.degree`(συντελεστής ομοιότητας βαθμού κορυφών ER δικτύου)
- `er.components.min`(ελάχιστο συνιστωσών ER δικτύου)
- `er.components.qu1`(πρώτο τέταρτο συνιστωσών ER δικτύου)
- `er.components.median`(διάμεσος συνιστωσών ER δικτύου)
- `er.components.mean`(μέση τιμή συνιστωσών ER δικτύου)
- `er.components.qu3`(τρία τέταρτα συνιστωσών ER δικτύου)
- `er.components.max2`(δεύτερη μεγαλύτερη συνιστώσα ER δικτύου)
- `er.components.max`(μέγιστο συνιστωσών ER δικτύου)
- `er.components.sd`(τυπική απόκλιση συνιστωσών ER δικτύου)
- `er.components.var`(διακύμανση συνιστωσών ER δικτύου)
- `er.components.mode2`(δεύτερη πιο συχνή συνιστώσα ER δικτύου)
- `er.components.mode`(συχνότερη συνιστώσα ER δικτύου)
- `er.components.cv`(σχετική τυπική απόκλιση συνιστωσών ER δικτύου)
- `er.components.number`(πλήθος συνιστωσών ER δικτύου)
- `er.degrees.min`(ελάχιστο πλήθος γειτόνων ER δικτύου)
- `er.degrees.qu1`(πρώτο τέταρτο πλήθους γειτόνων ER δικτύου)
- `er.degrees.median`(διάμεσος πλήθους γειτόνων ER δικτύου)
- `er.degrees.mean`(μέση τιμή πλήθους γειτόνων ER δικτύου)
- `er.degrees.qu3`(τρία τέταρτα πλήθους γειτόνων ER δικτύου)
- `er.degrees.max2`(δεύτερο μεγαλύτερο πλήθος γειτόνων ER δικτύου)
- `er.degrees.max`(μέγιστο πλήθος γειτόνων ER δικτύου)
- `er.degrees.sd`(τυπική απόκλιση πλήθους γειτόνων ER δικτύου)
- `er.degrees.var`(διακύμανση πλήθους γειτόνων ER δικτύου)
- `er.degrees.mode2`(δεύτερο πιο συχνό πλήθος γειτόνων ER δικτύου)
- `er.degrees.mode`(συχνότερο πλήθος γειτόνων ER δικτύου)
- `er.degrees.cv`(σχετική τυπική απόκλιση πλήθους γειτόνων ER δικτύου)
- `gc.multiple`(πλήθος πολλαπλών ακμών γιγάντιας συνιστώσας)
- `gc.loop`(πλήθος βρόγχων γιγάντιας συνιστώσας)
- `gc.density`(πυκνότητα γιγάντιας συνιστώσας)

- gc.transitivity.global(μεταβατικότητα γιγάντιας συνιστώσας)
- gc.transitivity.average(συντελεστής σύμπλεξης γιγάντιας συνιστώσας)
- gc.assortativity.degree(συντελεστής ομοιότητας βαθμού κορυφών γιγάντιας συνιστώσας)
- gc.degrees.min(ελάχιστο πλήθος γειτόνων γιγάντιας συνιστώσας)
- gc.degrees.qu1(πρώτο τέταρτο πλήθους γειτόνων γιγάντιας συνιστώσας)
- gc.degrees.median(διάμεσος πλήθους γειτόνων γιγάντιας συνιστώσας)
- gc.degrees.mean(μέση τιμή πλήθους γειτόνων γιγάντιας συνιστώσας)
- gc.degrees.qu3(τρία τέταρτα πλήθους γειτόνων γιγάντιας συνιστώσας)
- gc.degrees.max2(δεύτερο μεγαλύτερο πλήθος γειτόνων γιγάντιας συνιστώσας)
- gc.degrees.max(μέγιστο πλήθος γειτόνων γιγάντιας συνιστώσας)
- gc.degrees.sd(τυπική απόκλιση πλήθους γειτόνων γιγάντιας συνιστώσας)
- gc.degrees.var(διακύμανση πλήθους γειτόνων γιγάντιας συνιστώσας)
- gc.degrees.mode2(δεύτερο πιο συχνό πλήθος γειτόνων γιγάντιας συνιστώσας)
- gc.degrees.mode(συχνότερο πλήθος γειτόνων γιγάντιας συνιστώσας)
- gc.degrees.cv(σχετική τυπική απόκλιση πλήθους γειτόνων γιγάντιας συνιστώσας)
- gc.weights.min(ελάχιστα satoshis συναλλαγών γιγάντιας συνιστώσας)
- gc.weights.qu1(πρώτο τέταρτο satoshis συναλλαγών γιγάντιας συνιστώσας)
- gc.weights.median(διάμεσος satoshis συναλλαγών γιγάντιας συνιστώσας)
- gc.weights.mean(μέση τιμή satoshis συναλλαγών γιγάντιας συνιστώσας)
- gc.weights.qu3(τρία τέταρτα satoshis συναλλαγών γιγάντιας συνιστώσας)
- gc.weights.max2(δεύτερα μεγαλύτερα satoshis συναλλαγών γιγάντιας συνιστώσας)
- gc.weights.max(μέγιστα satoshis συναλλαγών γιγάντιας συνιστώσας)
- gc.weights.sd(τυπική απόκλιση satoshis συναλλαγών γιγάντιας συνιστώσας)
- gc.weights.var(διακύμανση satoshis συναλλαγών γιγάντιας συνιστώσας)
- gc.weights.mode2(δεύτερα πιο συχνά satoshis συναλλαγών γιγάντιας συνιστώσας)
- gc.weights.mode(συχνότερα satoshis συναλλαγών γιγάντιας συνιστώσας)
- gc.weights.cv(σχετική τυπική απόκλιση satoshis συναλλαγών γιγάντιας συνιστώσας)
- gc.weights.sum(άθροισμα satoshis συναλλαγών γιγάντιας συνιστώσας)

- `gc.strengths.min`(ελάχιστη ισχύς γιγάντιας συνιστώσας)
- `gc.strengths.qu1`(πρώτο τέταρτο ισχύος γιγάντιας συνιστώσας)
- `gc.strengths.median`(διάμεσος ισχύος γιγάντιας συνιστώσας)
- `gc.strengths.mean`(μέση τιμή ισχύος γιγάντιας συνιστώσας)
- `gc.strengths.qu3`(τρία τέταρτα ισχύος γιγάντιας συνιστώσας)
- `gc.strengths.max2`(δεύτερη μεγαλύτερη ισχύς γιγάντιας συνιστώσας)
- `gc.strengths.max`(μέγιστη ισχύς γιγάντιας συνιστώσας)
- `gc.strengths.sd`(τυπική απόκλιση ισχύος γιγάντιας συνιστώσας)
- `gc.strengths.var`(διακύμανση ισχύος γιγάντιας συνιστώσας)
- `gc.strengths.mode2`(δεύτερη πιο συχνή ισχύς γιγάντιας συνιστώσας)
- `gc.strengths.mode`(συχνότερη ισχύς γιγάντιας συνιστώσας)
- `gc.strengths.cv`(σχετική τυπική απόκλιση ισχύος γιγάντιας συνιστώσας)
- `gc.page.rank.min`(ελάχιστο κύρος γιγάντιας συνιστώσας)
- `gc.page.rank.qu1`(πρώτο τέταρτο κύρους γιγάντιας συνιστώσας)
- `gc.page.rank.median`(διάμεσος κύρους γιγάντιας συνιστώσας)
- `gc.page.rank.mean`(μέση τιμή κύρους γιγάντιας συνιστώσας)
- `gc.page.rank.qu3`(τρία τέταρτα κύρους γιγάντιας συνιστώσας)
- `gc.page.rank.max2`(δεύτερο μεγαλύτερο κύρος γιγάντιας συνιστώσας)
- `gc.page.rank.max`(μέγιστο κύρος γιγάντιας συνιστώσας)
- `gc.page.rank.sd`(τυπική απόκλιση κύρους γιγάντιας συνιστώσας)
- `gc.page.rank.var`(διακύμανση κύρους γιγάντιας συνιστώσας)
- `gc.page.rank.mode2`(δεύτερο πιο συχνό κύρος γιγάντιας συνιστώσας)
- `gc.page.rank.mode`(συχνότερο κύρος γιγάντιας συνιστώσας)
- `gc.page.rank.cv`(σχετική τυπική απόκλιση κύρους γιγάντιας συνιστώσας)
- `gc.page.rank.weighted.min`(ελάχιστο σταθμισμένο κύρος γιγάντιας συνιστώσας)
- `gc.page.rank.weighted.qu1`(πρώτο τέταρτο σταθμισμένου κύρους γιγάντιας συνιστώσας)
- `gc.page.rank.weighted.median`(διάμεσος σταθμισμένου κύρους γιγάντιας συνιστώσας)
- `gc.page.rank.weighted.mean`(μέση τιμή σταθμισμένου κύρους γιγάντιας συνιστώσας)

- gc.page.rank.weighted.qu3(τρία τέταρτα σταθμισμένου κύρους γιγάντιας συνιστώσας)
 - gc.page.rank.weighted.max2(δεύτερο μεγαλύτερο σταθμισμένο κύρος γιγάντιας συνιστώσας)
 - gc.page.rank.weighted.max(μέγιστο σταθμισμένο κύρος γιγάντιας συνιστώσας)
 - gc.page.rank.weighted.sd(τυπική απόκλιση σταθμισμένου κύρους γιγάντιας συνιστώσας)
 - gc.page.rank.weighted.var(διακύμανση σταθμισμένου κύρους γιγάντιας συνιστώσας)
 - gc.page.rank.weighted.mode2(δεύτερο πιο συχνό σταθμισμένο κύρος γιγάντιας συνιστώσας)
 - gc.page.rank.weighted.mode(συχνότερο σταθμισμένο κύρος γιγάντιας συνιστώσας)
 - gc.page.rank.weighted.cv(σχετική τυπική απόκλιση σταθμισμένου κύρους γιγάντιας συνιστώσας)
 - gc.vertex.connectivity(ελάχιστο πλήθος κορυφών γιγάντιας συνιστώσας που αν αφαιρεθούν θα γίνει ασθενές συνδεδεμένη)
 - gc.edge.connectivity(ελάχιστο πλήθος ακμών γιγάντιας συνιστώσας που αν αφαιρεθούν θα γίνει ασθενές συνδεδεμένη)
 - gc.hub.score(ιδιοτιμή του μεγαλύτερου hub γιγάντιας συνιστώσας)
 - gc.hub.score.weighted(ιδιοτιμή του μεγαλύτερου σταθμισμένου hub γιγάντιας συνιστώσας)
 - gc.authority.score(ιδιοτιμή του μεγαλύτερου authority γιγάντιας συνιστώσας)
 - gc.authority.score.weighted(ιδιοτιμή του μεγαλύτερου σταθμισμένου authority γιγάντιας συνιστώσας)
 - gc.motifs.3(πλήθος τριγώνων γιγάντιας συνιστώσας)
 - gc.articulation.points.number(πλήθος κορυφών τομών γιγάντιας συνιστώσας)
4. 21 χρονικές σειρές οικονομικής ανάλυσης οι οποίες εμπεριέχουν την τιμή του Bitcoin και δεν μπορούν να χρησιμοποιηθούν στα ΤΝΔ:
- max.transaction.value(τιμή μέγιστης συναλλαγής bitcoins ημερήσιου δικτύου συναλλαγών)

- `daily.value`(συνολική τιμή bitcoins ημερήσιου δικτύου συναλλαγών)
- `vertices.no.outgoing.value`(συνολική τιμή bitcoins κορυφών χωρίς εξερχόμενη συναλλαγή “διευθύνσεις αποταμίευσης” ημερήσιου δικτύου συναλλαγών)
- `gc.max.transaction.value`(τιμή μέγιστης συναλλαγής bitcoins γιγάντιας συνιστώσας)
- `gc.daily.value`(συνολική τιμή bitcoins γιγάντιας συνιστώσας)
- `gc.vertices.no.outgoing.value`(συνολική τιμή bitcoins κορυφών χωρίς εξερχόμενη συναλλαγή “διευθύνσεις αποταμίευσης” γιγάντιας συνιστώσας)
- `network.value`(συνολική τιμή bitcoins)
- `nvt`(λόγος συνολικής τιμής προς ημερήσια συνολική τιμή)
- `nvt.ma.14`(λόγος συνολικής τιμής προς ημερήσια συνολική τιμή μέσω φίλτρου μέσου όρου 14 ημερών)
- `daily.value.ma.45`(ημερήσια συνολική τιμή μέσω φίλτρου μέσου όρου 45 ημερών)
- `nvt.ma.45`(λόγος συνολικής τιμής προς `daily.value.ma.45`)
- `pmr.l_5`(φυσικός λογάριθμος λόγου συνολικής τιμής bitcoins προς `dua.p1_5.ma.15`)
- `pmr.n_ln`(φυσικός λογάριθμος λόγου συνολικής τιμής bitcoins προς `dua.n_ln.ma.15`)
- `pmr.2`(φυσικός λογάριθμος λόγου συνολικής τιμής bitcoins προς `dua.p2.ma.15`)
- `network.value.log`(φυσικός λογάριθμος συνολικής τιμής bitcoins)
- `nvm.log`(πηλίκο του αθροίσματος των `nvm.upper.bound` και `nvm.lower.bound` με το 2)
- `nvm`(διαφορά του `network.value.log` με το `nvm.log`)
- `nvm.norm`(κανονικοποιημένο `nvm`)
- `nvtg.m`(πηλίκο του `nvt.ma.45` με το γινόμενο του πλήθους μοναδικών κορυφών με το 2)
- `nvtg.m.ma45`(πηλίκο του `nvt.ma.45` με το γινόμενο του πλήθους μοναδικών κορυφών με το 2 μέσω φίλτρου μέσου όρου 45 ημερών)
- `nvtg.z`(πηλίκο του `nvt.ma.45` με το άθροισμα του `dua.ln` με το 1)

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Smith, B. (2018). *What are the three generations of blockchain, and how are they similar to the web?*. <https://www.coininsider.com/three-generations-of-blockchain/> (30-03-2018)
- [2] Μπακαβός, Α. (2014). Μελέτη Στόχευσης Χρήσης και Ανάλυση Οικονομικής Δραστηριότητας στο Δίκτυο Συναλλαγών του Bitcoin. <http://ikee.lib.auth.gr/record/136092> (19-12-2017)
- [3] Ron, D., Shamir, A. (2013). *Quantitative Analysis of the Full Bitcoin Transaction Graph*. Financial Cryptography and Data Security Vol. 7859, pp. 6-24.
- [4] Kondor, D., Pósfai, M., Csabai, I., Vattay, G. (2014). *Do the rich get richer? An empirical analysis of the Bitcoin transaction network*. PLoS ONE 9(2): e86197.
- [5] Di Francesco Maesa D., Marino A., Ricci L. (2017). *An analysis of the Bitcoin users graph: inferring unusual behaviours*. Complex Networks & Their Applications V. Studies in Computational Intelligence Vol 693, pp. 749-760.
- [6] Di Francesco Maesa D., Marino A., Ricci L. (2016). *Uncovering the Bitcoin Blockchain: An Analysis of the Full Users Graph*. 537-546. 10.1109/DSAA.2016.52.
- [7] Fergal, R., Harrigan, M. (2011). *An analysis of anonymity in the bitcoin system*. In: Proceeding of 2011 PASSAT/SocialCom 2011, pp. 1318–1326. IEEE
- [8] Androulaki, E., Karame, G., Roeschlin, M., Scherer, T., Capkun, S. (2013). *Evaluating user privacy in bitcoin*. In: Financial Cryptography and Data Security - 17th International Conference, FC 2013, Okinawa, Japan, April 1-5, 2013, Revised Selected Papers, pp. 34–51
- [9] Kondor, D., Csabai, I., Szüle, J., Pósfai, M., Vattay, G. (2014). *Inferring the interplay between network structure and market effects in Bitcoin*. New Journal of Physics. 16. 10.1088/1367-2630/16/12/125003.
- [10] Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. <https://bitcoin.org/bitcoin.pdf> (30-03-2019)

- [11] Frankel, M. (2018). *How Many Cryptocurrencies Are There?*. <https://www.fool.com/investing/2018/03/16/how-many-cryptocurrencies-are-there.aspx> (28-08-2018)
- [12] Pollock, D. (2018). *Bears for Crypto, Bulls for ICOs: 2018 Market Positive Statistics*. <https://cointelegraph.com/news/bears-for-crypto-bulls-for-icos-2018-market-positive-statistics> (28-08-2018)
- [13] Blagojevic, D. (2019). *How Many People Use Bitcoin? How Many Bitcoin Wallets Are There?*. <https://captainaltcoin.com/how-many-people-use-bitcoin/> (11-03-2019)
- [14] Lielacher, A. (2019). *How Many People Use Bitcoin in 2019?*. <https://www.bitcoinmarketjournal.com/how-many-people-use-bitcoin/> (11-03-2019)
- [15] Young, J. (2018). *Big Milestone: Bitcoin on Track to Transfer More Value Than Mastercard Daily*. <https://www.ccn.com/big-milestone-bitcoin-on-track-to-transfer-more-value-than-mastercard-daily> (12-03-2019)
- [16] Vermeulen, J. (2017). *Bitcoin and Ethereum vs Visa and PayPal – Transactions per second*. <https://mybroadband.co.za/news/banking/206742-bitcoin-and-ethereum-vs-visa-and-paypal-transactions-per-second.html> (12-03-2019)
- [17] Zheng, S. (2018). *CoinShares report: Renewable energy accounts for 77.6% of total Bitcoin's energy usage*. <https://www.theblockcrypto.com/2018/11/30/coinshares-report-renewable-energy-accounts-for-77-6-of-total-bitcoins-energy-usage/> (17-03-2019)
- [18] Khatri, Y. (2019). *Bitcoin Is Becoming More Decentralized, Indicates New Research*. <https://www.coindesk.com/bitcoin-is-becoming-more-decentralized-indicates-new-research> (16-03-2019)
- [19] Redman, J. (2019). *Mystery Bitcoin Miners Are Altering Mining Pool Dominance*. <https://news.bitcoin.com/mystery-bitcoin-miners-are-altering-mining-pool-dominance/> (16-03-2019)
- [20] Aki, J. (2019). *Report: Crypto Exchanges Saw Trading Volumes Plummet in January*. <https://bitcoinmagazine.com/articles/report-crypto-exchanges-saw-trading-volumes-plummet-january/> (19-03-2019)

- [21] Madore, P. (2019). *Major Bitcoin Exchanges Turn Blind Eye to Wash Trading: Bombshell Report*. <https://www.ccn.com/major-bitcoin-exchanges-blind-eye-to-wash-trading-report> (29-03-2019)
- [22] Cavicchioli, M. (2019). *The bitcoin bubble vs the Apple stock speculation*. <https://cryptonomist.ch/en/2019/01/04/bitcoin-bubble-vs-apple-stock-speculation/> (22-03-2019)
- [23] Rick, D. (2018). *Diar Report Reveals 30% of Bitcoin Supply is Lost or Yet to be Mined*. <https://www.newsbtc.com/2018/09/18/diar-report-reveals-30-of-bitcoin-supply-is-lost-or-yet-to-be-mined/> (17-03-2019)
- [24] Chainalysis Team. (2018). *Bitcoin's \$30 billion sell-off*. <https://blog.chainalysis.com/reports/money-supply> (31-03-2019)
- [25] Coinatmradar. (2019). *Crypto ATM Distribution by Continents and Countries*. <https://coinatmradar.com/charts/geo-distribution/> (11-03-2019)
- [26] Transparency Market Research. (2017). *Worldwide Blockchain Technology Market is Anticipated to Exhibit a CAGR of 58.7% Between 2016 and 2024*. <https://www.prnewswire.com/news-releases/worldwide-blockchain-technology-market-is-anticipated-to-exhibit-a-cagr-of-587-between-2016-and-2024-elimination-of-third-parties-improves-demand-and-security-of-online-transactions--tmr-611067345.html> (31-03-2019)
- [27] Campbell, R. (2016). *PwC Expert: \$1.4 Billion Invested in Blockchain in 2016*. <https://www.ccn.com/pwc-expert-1-4-billion-invested-blockchain-2016> (31-03-2019)
- [28] Pudwell, S. (2017). *Blockchain Tech 'Could Save' Banks £9.7bn Billion A Year*. <https://www.silicon.co.uk/e-enterprise/blockchain-tech-save-banks-203284> (31-03-2019)
- [29] Thompson, C. (2017). *Private Blockchain or Database?*. <https://medium.com/blockchain-review/private-blockchain-or-database-whats-the-difference-523e7d42edc> (28-08-2018)
- [30] Sharma, T. (2018). *Can a Blockchain Be Queried Like a Normal Database?*. <https://www.blockchain-council.org/blockchain/can-a-blockchain-be-queried-like-a-normal-database/> (31-03-2019)

- [31] Fedak, V. (2018). *Blockchain and Big Data: the match made in heavens*. <https://towardsdatascience.com/blockchain-and-big-data-the-match-made-in-heavens-337887a0ce73> (28-08-2018)
- [32] blockchain.info. (2019). *Bitcoin blockchain size 2010-2019, by quarter*. <https://www.statista.com/statistics/647523/worldwide-bitcoin-blockchain-size/> (31-03-2019)
- [33] Bitcoin Wiki. (2017). *Data directory*. https://en.bitcoin.it/wiki/Data_directory (31-03-2019)
- [34] Bitcoin. (2019). *Script*. <https://en.bitcoin.it/wiki/Script> (04-04-2019)
- [35] Walker, G. (2018). *blk.dat*. <http://learnmeabitcoin.com/glossary/blkdat> (02-04-2019)
- [36] Bitcoin. (2018). *Protocol documentation*. https://en.bitcoin.it/wiki/Protocol_documentation (02-04-2019)
- [37] Walker, G. (2018). *Block Header*. <http://learnmeabitcoin.com/glossary/block-header> (02-04-2019)
- [38] Walker, G. (2018). *Transaction Data*. <http://learnmeabitcoin.com/glossary/transaction-data> (02-04-2019)
- [39] Bitcoin. (2019). *Transaction*. <https://en.bitcoin.it/wiki/Transaction> (02-04-2019)
- [40] Bitcoin. (2018). *Transactions*. <https://bitcoin.org/en/developer-guide#transactions> (04-04-2019)
- [41] Walker, G. (2018). *Coinbase Transaction*. <http://learnmeabitcoin.com/glossary/coinbase-transaction> (04-04-2019)
- [42] Bitcoin. (2016). *Bitcoin Improvement Proposals*. https://en.bitcoin.it/wiki/Bitcoin_Improvement_Proposals (31-03-2019)
- [43] blockchain2graph. (2019). *blockchain2graph*. <https://github.com/straumat/blockchain2graph> (06-04-2019)
- [44] bitcoin-abe. (2018). *bitcoin-abe*. <https://github.com/bitcoin-abe/bitcoin-abe> (06-04-2019)
- [45] Kalodner, H., Goldfeder, S., Chator, A., Möser, M., Narayanan, A. (2017). *Blocksci: Design And Applications Of A Blockchain Analysis Platform*. from Arxiv.Org: <https://arxiv.org/abs/1709.02489>

- [46] Bitcoin. (2018). Address. <https://en.bitcoin.it/wiki/Address> (09-04-2019)
- [47] Bitcoin. (2018). OP_RETURN. https://en.bitcoin.it/wiki/OP_RETURN (09-04-2019)
- [48] Acheson, N. (2018). What is SegWit?. <https://www.coindesk.com/information/what-is-segwit> (07-04-2019)
- [49] Day, A., Bookman, C. (2018). Bitcoin in BigQuery: blockchain analytics on public data. <https://cloud.google.com/blog/products/gcp/bitcoin-in-bigquery-blockchain-analytics-on-public-data> (07-04-2019)
- [50] cobookman. (2018). blockchainToAvro. <https://github.com/cobookman/blockchainToAvro> (07-04-2019)
- [51] Day, A., Risdal, M., Dane, S. (2018). Bitcoin Blockchain. <https://www.kaggle.com/bigquery/bitcoin-blockchain> (07-04-2019)
- [52] Day, A., Medvedev, E., Ak, N., Price, W. (2019). Introducing six new cryptocurrencies in BigQuery Public Datasets—and how to analyze them. <https://cloud.google.com/blog/products/data-analytics/introducing-six-new-cryptocurrencies-in-bigquery-public-datasets-and-how-to-analyze-them> (07-04-2019)
- [53] Kondor, D. (2018). bitcoind-dump-tsv. <https://github.com/dkondor/bitcoin> (13-04-2019)
- [54] Bitcoin. (2019). Bitcoin Core. <https://bitcoincore.org/> (13-04-2019)
- [55] Kondor, D., Csabai, I., Szüle, J., Pósfai, M., Vattay, G. (2015). ELTE Bitcoin Project website and resources <http://www.vo.elte.hu/bitcoin/downloads.htm> (13-04-2019)
- [56] Kondor, D. (2018). txedges. <https://github.com/dkondor/txedges> (13-04-2019)
- [57] Woo, W. (2017). Introducing NVT Ratio (Bitcoin's PE Ratio), use it to detect bubbles. <https://woobull.com/introducing-nvt-ratio-bitcoins-pe-ratio-use-it-to-detect-bubbles/> (23-04-2019)
- [58] Kalichkin, D. (2018). Rethinking Network Value to Transactions (NVT) Ratio. <https://medium.com/cryptolab/https-medium-com-kalichkin-rethinking-nvt-ratio-2cf810df0ab0> (23-04-2019)
- [59] Wikipedia. (2019). Metcalfe's law. https://en.wikipedia.org/wiki/Metcalfe%27s_law (23-04-2019)
- [60] Wikipedia. (2019). Zipf's law. https://en.wikipedia.org/wiki/Zipf%27s_law (23-04-2019)

- [61] Kalichkin, D. (2018). *Rethinking Metcalfe's Law applications to cryptoasset valuation*. <https://medium.com/cryptolab/network-value-to-metcalfe-nvm-ratio-fd59ca3add76> (22-04-2019)
- [62] Arun, V. (2018). *Improvements on the Network Value to Transactions (NVT) Ratio & Introducing Network Value/Transactions to Growth (NVTG) to Value Crypto*. <https://medium.com/ledgercapital/improvements-on-the-network-value-to-transactions-nvt-ratio-introducing-network-569569f6b3e3> (22-04-2019)
- [63] Arun, V. (2018). *Using NVT and NVTG in practice to value the most popular cryptocurrencies*. <https://medium.com/@vikram.arun/using-nvt-and-nvtg-in-practice-to-value-the-most-popular-cryptocurrencies-985cbad9652e> (22-04-2019)
- [64] Magas, J. (2019). *Blockchain and AI: Leading the Way to the Fourth Industrial Revolution Against the Odds*. <https://cointelegraph.com/news/blockchain-and-ai-leading-the-way-to-the-fourth-industrial-revolution-against-the-odds> (06-08-2019)