

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΜΣ ΔΙΚΤΥΩΜΕΝΑ ΣΥΣΤΗΜΑΤΑ ΜΕΓΑΛΟΥ ΟΓΚΟΥ ΔΕΔΟΜΕΝΩΝ



Διπλωματική Εργασία

Σύγχρονες Μέθοδοι Ανάλυσης Δεδομένων Δικτύων
Modern Network Data Analysis Techniques

Μακρίδου Αικατερίνη
Επιβλέπων: Καρυώτης Βασίλειος

9 Φεβρουαρίου

Επιβλέπων

Καρυώτης Βασίλειος, Αναπληρωτής Καθηγητής,
Ιόνιο Πανεπιστήμιο

Τριμελής Επιτροπή

Καρυώτης Βασίλειος, Αναπληρωτής Καθηγητής,
Ιόνιο Πανεπιστήμιο

Κουρουθανάσης Παναγιώτης, Καθηγητής,
Ιόνιο Πανεπιστήμιο

Έξαρχος Θεμιστοκλής, Αναπληρωτής Καθηγητής,
Ιόνιο Πανεπιστήμιο

Περίληψη

Η παρούσα διπλωματική εργασία, εστιάζει στην ανάλυση αθλητικών κοινωνικών δικτύων και γραφημάτων γνώσης με στόχο την κατανόηση των βασικών δομικών και κεντρικών χαρακτηριστικών τους. Η βιβλιογραφική ανασκόπηση, επικεντρώνεται σε βασικές έννοιες της ανάλυσης γραφημάτων, όπως ο βαθμός, ο μέσος βαθμός, και οι κεντρικότητες δικτύων. Στη συνέχεια, η έρευνα επικεντρώνεται στην ανάλυση πέντε ολλανδικών clubs της πλατφόρμας Strava και των σχηματιζόμενων δικτύων kudos. Παρουσιάζεται η μεθοδολογία συλλογής και ανάλυσης των δεδομένων, καθώς και τα αποτελέσματα με σημαντικές μετρικές, των οποίων έγινε παρουσίαση τους στο κεφάλαιο της βιβλιογραφικής ανασκόπησης. Εξετάζεται επίσης η εύρεση κοινοτήτων μέσω μεθόδων όπως η μέθοδος Girvan-Newman, η μέθοδος Modularity Maximisation και η μέθοδος Spectral Clustering. Τα συμπεράσματα αναδεικνύουν τα κυριότερα ευρήματα, ενώ προτείνονται πιθανές κατευθύνσεις για μελλοντικές έρευνες, ενισχύοντας τη συνεισφορά της εργασίας στον τομέα της ανάλυσης αθλητικών κοινωνικών δικτύων και γραφημάτων γνώσης. Στην αναλυτική παρουσίαση των μεθόδων χρησιμοποιούνται εικόνες, γραφήματα, και πίνακες προκειμένου να κατανοηθούν ευκολότερα οι διαδικασίες.

Abstract

This thesis focuses on the analysis of sports social networks and knowledge graphs in order to understand their basic structural and central characteristics. The literature review, focuses on key concepts of graph analysis, such as degree, average degree, and network centralities. Then, the research focuses on the analysis of five Dutch clubs of the Strava platform and the formed kudos networks. The methodology of data collection and analysis is presented, as well as the results with important metrics, which were presented in the literature review chapter. Community detection through methods such as Girvan-Newman method, Modularity Maximisation method and Spectral Clustering method are also considered. The conclusions highlight the main findings and suggest possible directions for future research, enhancing the contribution of the paper to the field of sports social network and knowledge graph analysis. Figures, graphs, and tables are used in the detailed presentation of the methods in order to make the processes easier to understand.

Ευχαριστίες

Με την παρούσα διπλωματική εργασία ολοκληρώνονται οι σπουδές μου στο μεταπτυχιακό πρόγραμμα σπουδών «Δικτυωμένα Συστήματα Μεγάλου Όγκου Δεδομένων» του Ιονίου Πανεπιστημίου.

Υπήρξε σημαντική η συμβολή όλων των καθηγητών, στα γνωστικά αντικείμενα που περιελάμβανε το παρόν πρόγραμμα, στους οποίους και εκφράζω τις εγκάρδιες και ειλικρινείς μου ευχαριστίες.

Ιδιαίτερες ευχαριστίες, θα ήθελα να απευθύνω στον καθηγητή και επιβλέποντα της παρούσας διπλωματικής εργασίας, κύριο Καρυώτη Βασίλειο, ο οποίος τόσο καθοριστικά και μεθοδικά, με καθοδήγησε προς την διεκπεραίωση της.

Τέλος, αλλά με ιδιαίτερη σημασία για μένα, θα ήθελα ευχαριστήσω την οικογένεια μου, που με στήριξε και συνεχίζει να με στηρίζει σε όλες μου τις προσπάθειες.

Περιεχόμενα

1	ΕΙΣΑΓΩΓΗ	1
1.1	ΕΙΣΑΓΩΓΗ	1
1.2	ΣΚΟΠΟΣ ΚΑΙ ΣΤΟΧΟΙ ΕΡΓΑΣΙΑΣ	3
1.3	ΔΟΜΗ ΕΡΓΑΣΙΑΣ	4
2	ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ	5
2.1	ΑΝΑΛΥΣΗ ΚΟΙΝΩΝΙΚΩΝ ΔΙΚΤΥΩΝ	5
2.2	ΓΡΑΦΟΙ ΓΝΩΣΗΣ	8
2.3	ΘΕΩΡΙΑ ΓΡΑΦΗΜΑΤΩΝ	12
2.3.1	Βαθμός - Degree	13
2.3.2	Μέσος Βαθμός και Κατανομή Μέσου Βαθμού - Average Degree and Average Degree Distribution	14
2.3.2.1	Κανονικό Γράφημα - Regular Graph	15
2.3.2.2	Τυχαίο Γράφημα - Random Graph	16
2.3.2.3	Τυχαίο Γεωμετρικό Γράφημα - Random Geometric Graph	17
2.3.2.4	Γράφημα Μικρού-Κόσμου - Small-World Graph	18
2.3.2.5	Γράφημα Ελεύθερης Κλίμακας - Scale-Free Graph	19
2.3.3	κεντρικότητες δικτύων	20
2.3.3.1	κεντρικότητα Βαθμού - Degree Centrality	21
2.3.3.2	Ενδιαμεσική κεντρικότητα - Betweenness Centrality	22
2.3.3.3	κεντρικότητα Εγγύτητας - Closeness Centrality	22
2.3.3.4	κεντρικότητα Ιδιοδιανύσματος - Eigenvector Centrality	23
2.3.4	συντελεστής Ομαδοποίησης - Clustering Coefficient	24

2.3.5	Ακτίνα και Διάμετρος - Radius and Diameter	24
2.3.6	Εντοπισμός Κοινοτήτων	25
2.3.6.1	Μέθοδος <i>Girvan - Newman</i>	27
2.3.6.2	Μέθοδος <i>Modularity Maximization</i>	28
2.3.6.3	Μέθοδος <i>Spectral Clustering</i>	29
2.4	ΑΝΑΛΥΣΗ ΑΘΛΗΤΙΚΩΝ ΔΙΚΤΥΩΝ	31
3	ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ	36
3.1	ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ	40
3.2	ΜΕΘΟΔΟΛΟΓΙΑ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ	42
3.2.1	Γραφήματα	42
3.2.2	Ανάλυση Γράφων των Clubs	46
4	ΜΕΘΟΔΟΛΟΓΙΑ	52
4.1	ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ - ΔΗΜΙΟΥΡΓΙΑ ΓΡΑΦΩΝ	52
4.1.1	Προεπεξεργασία Δεδομένων	53
4.1.2	Εξαγωγή μετρικών	54
4.1.3	Δημιουργία Γράφων	55
4.1.4	Δημιουργία Κατανομών Βαθμού Κόμβων	56
4.1.5	Εντοπισμός Κοινοτήτων	56
5	ΑΠΟΤΕΛΕΣΜΑΤΑ	59
5.1	ΠΑΡΟΥΣΙΑΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ	59
6	ΕΠΙΛΟΓΟΣ	82
6.1	ΣΥΝΟΨΗ ΕΡΓΑΣΙΑΣ	82
6.2	ΣΥΝΟΨΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ	83
6.3	ΠΕΡΙΟΡΙΣΜΟΙ ΚΑΙ ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ	84

Κατάλογος Σχημάτων

2.1	Αναπαράσταση ενός κοινωνικού δικτύου [17].	7
2.2	Κατευθυνόμενο δίκτυο [65].	11
2.3	Κατευθυνόμενα δίκτυα με βάρη [55].	12
2.4	Παραδείγματα τοπολογιών και κατανομών βαθμών μοντέλων [56].	15
2.5	Τοπολογία κανονικών γραφημάτων [69].	16
2.6	Τοπολογία τυχαίου γραφήματος [72].	17
2.7	Κατανομή βαθμού κόμβων σε ένα τυχαίο γράφημα [67].	17
2.8	Τοπολογία τυχαίου γεωμετρικού γραφήματος [25].	18
2.9	Κατανομή βαθμού κόμβων σε ένα τυχαίο γεωμετρικό γράφημα σε σύγκριση με ένα τυχαίο γράφημα [63].	18
2.10	Τοπολογία γραφήματος μικρού κόσμου [72].	19
2.11	Τοπολογία γραφήματος ελεύθερης κλίμακας [72].	20
2.12	Κατανομή βαθμού κόμβων σε γραφήματα μικρού κόσμου και γραφήματα χωρίς κλίμακα [72].	21
2.13	Δομές Κοινοτήτων [65].	27
3.14	Προφίλ χρήστη στο Strava.	37
3.15	Προφίλ χρήστη στο Strava και διάφορες πληροφορίες δραστηριοτήτων.	38
3.16	Δημοσίευση δραστηριότητας στο Strava.	39
3.17	Αναπαράσταση δικτύων φιλίας - friendships graph.	43
3.18	δίκτυο kudos του Club 1.	44
3.19	δίκτυο kudos του Club 2.	45
3.20	δίκτυο kudos του Club 3.	46
3.21	δίκτυο kudos του Club 4.	47
3.22	δίκτυο kudos του Club 5.	48

3.23	<i>Average kudo's degree ανά μήνα του Club 1.</i>	49
3.24	<i>Average kudo's degree ανά μήνα του Club 2.</i>	49
3.25	<i>Average kudo's degree ανά μήνα του Club 3.</i>	50
3.26	<i>Average kudo's degree ανά μήνα του Club 4.</i>	50
3.27	<i>Average kudo's degree ανά μήνα του Club 5.</i>	51
5.28	<i>Κοινότητες που σχηματίζονται με τις διαφορετικές μεθόδους τον Μήνα 9.</i>	81

Κατάλογος Πινάκων

3.1	Χαρακτηριστικά Dataset και Περιγραφή.	42
3.2	Χαρακτηριστικά γράφων φιλίας.	43
3.3	Αριθμός μελών με <i>kudos</i> ανά Club.	46
3.4	Μετρικές γραφημάτων για τον προσδιορισμό τοπολογιών ανά Club.	48
3.5	Τύπος Δικτύων ανά Club.	51
5.1	Κοινότητες Club 1 ανά μήνα με τη μέθοδο Modularity Maximization. . . .	60
5.2	Κοινότητες Club 4 ανά μήνα με τη μέθοδο Modularity Maximization. . . .	60
5.3	Κοινότητες Club 2 ανά μήνα με τη μέθοδο Modularity Maximization. . . .	61
5.4	Κοινότητες Club 3 ανά μήνες 1-6 με τη μέθοδο Modularity Maximization. . . .	62
5.5	Κοινότητες Club 3 ανά μήνες 7-12 με τη μέθοδο Modularity Maximization. . . .	63
5.6	Κοινότητες Club 5 ανά μήνες 1-9 με τη μέθοδο Modularity Maximization. . . .	64
5.7	Κοινότητες Club 5 ανά μήνες 9-12 με τη μέθοδο Modularity Maximization. . . .	65
5.8	Κοινότητες Club 1 ανά μήνα με τη μέθοδο Girvan-Newman.	65
5.9	Κοινότητες Club 4 ανά μήνα με τη μέθοδο Girvan-Newman.	66
5.10	Κοινότητες Club 2 ανά μήνα με τη μέθοδο Girvan-Newman.	67
5.11	Κοινότητες Club 3 ανά μήνες 1-6 με τη μέθοδο Girvan-Newman.	68
5.12	Κοινότητες Club 3 ανά μήνες 7-12 με τη μέθοδο Girvan-Newman.	69
5.13	Κοινότητες Club 5 ανά μήνες 1-8 με τη μέθοδο Girvan-Newman.	70
5.14	Κοινότητες Club 5 ανά μήνες 9-12 με τη μέθοδο Girvan-Newman.	71
5.15	Κοινότητες Club 1 ανά μήνα με τη μέθοδο Spectral Clustering.	72
5.16	Κοινότητες Club 4 ανά μήνα με τη μέθοδο Spectral Clustering.	72
5.17	Κοινότητες Club 2 ανά μήνα με τη μέθοδο Spectral Clustering.	73
5.18	Κοινότητες Club 3 ανά μήνες 1-6 με τη μέθοδο Spectral Clustering.	74
5.19	Κοινότητες Club 3 ανά μήνες 7-12 με τη μέθοδο Spectral Clustering.	75

5.20 Κοινότητες Club 5 ανά μήνες 1-8 με τη μέθοδο <i>Spectral Clustering</i>	76
5.21 Κοινότητες Club 5 ανά μήνες 9-12 με τη μέθοδο <i>Spectral Clustering</i>	77
5.22 Χρόνοι εκτέλεσης (<i>sec</i>) των αλγορίθμων εντοπισμού κοινοτήτων.	79
5.23 Μέσος όρος <i>Modularity</i> των αλγορίθμων εντοπισμού κοινοτήτων ανά <i>club</i>	80

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

1.1 Εισαγωγή

Την τελευταία δεκαετία η αύξηση της χρήσης έξυπνων συσκευών (*wearables*) όπως τα έξυπνα ρολόγια, τα ασύρματα ακουστικά κλπ, και εφαρμογών όπως *Facebook*, *X* (πρώην *Twitter*), κλπ, έχει δημιουργήσει μια πληθώρα δυνατοτήτων, για τη συλλογή και ανάλυση δεδομένων, που σχετίζονται με την καθημερινή μας δραστηριότητα. Κινητά τηλέφωνα και έξυπνα ρολόγια, έχουν επικρατήσει σε αυτήν την εξέλιξη, συλλέγοντας σημαντικές πληροφορίες που είναι, όχι μόνο χρήσιμες για τους χρήστες, αλλά και για τους ερευνητές σε διάφορους τομείς. Η τεχνολογία και η ευκολία χρήσης των έξυπνων συσκευών, έχουν καταστήσει τη συλλογή πληροφοριών σχετικά με την υγεία, την φυσική κατάσταση και την καθημερινότητα μια αρκετά απλή διαδικασία για όλους. Μέσα από τη διερεύνηση και την ανάλυση τέτοιων δεδομένων, ερευνητές και επαγγελματίες, μπορούν να αποκτήσουν βαθύτερη κατανόηση της ανθρώπινης συμπεριφοράς, των αθλητικών επιδόσεων και της υγείας γενικότερα. Ακόμη, παρέχουν τη δυνατότητα επίβλεψης της υγείας ατόμων που αντιμετωπίζουν χρόνιες παθήσεις, επιτρέποντας την παρατήρηση δεικτών υγείας και συμμετοχής σε προγράμματα διαχείρισης. Δίνουν τη δυνατότητα ιχνηλάτησης της σωματικής δραστηριότητας και της διατροφής, προσφέροντας σημαντικές πληροφορίες για την υποστήριξη και των υγειών ατόμων. Ως αποτέλεσμα τα παραπάνω επιτρέπουν στους επαγγελματίες υγείας να παρακολουθούν την πρόοδο των ασθενών απομακρυσμένα, βελτιώνοντας την δυνατότητα παροχής προσαρμοσμένης φροντίδας.

Αναλύοντας τέτοια δεδομένα, που συλλέγονται από φορητές συσκευές, μπορούμε να ανιχνεύ-

σουμε τάσεις και προειδοποιητικά σημάδια, σχετικά με την υγεία. Αυτό, μπορεί να οδηγήσει σε προσωποποιημένες υποδείξεις, σχετικές με τη βελτίωση του τρόπου ζωής και την πρόληψη ασθενειών. Επιπλέον, στον τομέα του αθλητισμού, βοηθά στη βελτίωση της απόδοσης των αθλητών. Ερευνητές και προπονητές, μπορούν να εξάγουν σημαντικές πληροφορίες για την αποτελεσματικότητα των προπονητικών προγραμμάτων και τη βελτίωση των τεχνικών αθλητικών επιδόσεων. Τέλος, η ανάλυση δεδομένων στα αθλητικά κοινωνικά δίκτυα βοηθά σε μεγάλο βαθμό, ώστε να μην προκαταβάλλονται τα αποτελέσματα. Η ανάλυση αυτών των δεδομένων, μπορεί να αποκαλύψει τάσεις στον τρόπο που οι άνθρωποι ενσωματώνουν τη φυσική δραστηριότητα στην καθημερινή τους ζωή και πώς αλληλεπιδρούν σε αυτό το πλαίσιο. Αυτές οι πληροφορίες, μπορούν να χρησιμοποιηθούν για την ανάπτυξη κοινωνικών πολιτικών και προγραμμάτων υγείας, που προωθούν την υγιή ζωή και την ευεξία.

Τα κοινωνικά δίκτυα (*Twitter, Facebook, Be Real*, κλπ), είναι ένα παγκόσμιο φαινόμενο το οποίο έχει αναδειχθεί τα τελευταία χρόνια, ενώ έχουν συμβάλλει στην δημιουργία ενός καινούργιου τρόπου επικοινωνίας, κοινωνικής δικτύωσης και ανταλλαγής πληροφορίας. Δίνουν ευκαιρία, για τη σύνδεση και την επικοινωνία ανθρώπων, ανεξαρτήτως αποστάσεων και χωροταξικών περιορισμών. Ακόμη, έχουν γίνει πλατφόρμες διαμοιρασμού πληροφοριών και περιεχομένου, επιτρέποντας στους χρήστες να εκφράζονται, να αναδεικνύουν τα ενδιαφέροντά τους και να συνομιλούν. Αποτελούν, σημαντικά εργαλεία για τις επιχειρήσεις και τον επαγγελματικό κόσμο. Επιτρέπουν την διαφήμιση, την προώθηση των προϊόντων και την επικοινωνία με το κοινό, ενώ παράλληλα παρέχουν δυνατότητες για δικτύωση και ανάπτυξη των επαγγελματικών σχέσεων. Αυτή η δυνατότητα, να επηρεάζουν τη δημόσια γνώμη και να προωθούν ιδέες και αξίες, έχει καταστήσει τα κοινωνικά δίκτυα ισχυρά εργαλεία για την πολιτική, την κοινωνία και τον πολιτισμό. Τα κοινωνικά δίκτυα, έχουν διαμορφώσει τον τρόπο ζωής, εργασίας, επικοινωνίας και έκφρασης της σύγχρονης κοινωνίας, έτσι λοιπόν, είναι ένας τρόπος για τον άνθρωπο να αισθάνεται συνδεδεμένος με τον κόσμο, να ενημερώνεται για τα τελευταία νέα και να επικοινωνεί με άλλους σε παγκόσμιο επίπεδο.

Η ανάγκη για κοινωνική σύνδεση και συμμετοχή σε κοινωνικές ομάδες, έχει φέρει στο προσκήνιο την δημιουργία αθλητικών κοινωνικών δικτύων. Αυτά τα δίκτυα, δίνουν την δυνατότητα σε άτομα με αθλητικά ενδιαφέροντα να συνδεθούν, να ανταλλάξουν εμπειρίες, και να ανταγωνιστούν, στην πραγματικότητα ή σε εικονικά περιβάλλοντα, σε διάφορες αθλητικές δραστηριότητες. Στο πλαίσιο αυτών των αθλητικών κοινωνικών δικτύων, οι χρήστες μπορούν να διαμοιραστούν δεδομένα που σχετίζονται με τη φυσική τους κατάσταση, όπως οι καρδιακοί παλμοί, η απόσταση που διανύουν, κ.α.

Αυτά τα δεδομένα, συλλέγονται μέσω αισθητήρων που βρίσκονται στα κινητά τηλέφωνα, έξυπνα ρολόγια, και άλλες φορετές συσκευές (*wearables*). Ένα από τα κορυφαία παραδείγματα αθλητικού κοινωνικού δικτύου, που έχει μεγάλη δημοτικότητα την τελευταία δεκαετία, είναι το *Strava*¹. Η πλατφόρμα αυτή, επιτρέπει στους χρήστες της να καταγράφουν, να ανεβάζουν και να κοινοποιούν τις αθλητικές τους δραστηριότητες, ανεξάρτητα από το είδος του αθλήματος (τρέξιμο, ποδηλασία, κολύμπι, κ.α.). Ακόμη, παρέχει έναν χώρο όπου οι χρήστες μπορούν να καταγράψουν την απόσταση που διανύθηκε, τον χρόνο που απαιτήθηκε, την ταχύτητα, και άλλα γενικά στοιχεία που σχετίζονται με τη δραστηριότητά τους. Το *Strava*, απευθύνεται σε αθλητές κάθε επιπέδου, ερασιτέχνες και επαγγελματίες. Προσφέρει κοινωνικές δυνατότητες, όπως τη δυνατότητα να ακολουθείς άλλους χρήστες, να δημιουργείς ή να γίνεσαι μέλος σε κοινότητες και να συμμετέχεις σε προκλήσεις, όπου οι χρήστες μπορούν να συνεργαστούν για κοινούς στόχους ή να θέσουν τους δικούς τους. Αυτές οι λειτουργίες, δημιουργούν ένα δυναμικό κοινωνικό περιβάλλον, που ενθαρρύνει την αλληλεπίδραση και την επίτευξη αθλητικών στόχων [30].

1.2 Σκοπός και Στόχοι Εργασίας

Η ανάλυση δικτύων στον χώρο του αθλητισμού, ανοίγει νέες προοπτικές, για την κατανόηση των σχέσεων και των παραμέτρων που επηρεάζουν τους αθλητές και τον αθλητισμό. Με τη χρήση μεθόδων ανάλυσης σύνθετων δικτύων (*CNA*), επιτρέπεται και διευκολύνεται η ανάλυση των πολύπλοκων δικτύων που αναπτύσσονται γενικότερα στον αθλητισμό και ειδικότερα σε αθλητικά κοινωνικά δίκτυα. Η κατασκευή γράφων, από δεδομένα αγώνων, τουρνουά και αθλητικών δραστηριοτήτων επιτρέπει την ανάδειξη διασυνδέσεων μεταξύ αθλητών, ομάδων, και ακόμη και του κοινού. Η ανάλυση των δομών αυτών των δικτύων, μπορεί να αποκαλύψει κρίσιμες πληροφορίες σχετικά με την ιεραρχία των αθλητών, τη σύνθεση των ομάδων, και τον τρόπο με τον οποίο αλληλεπιδρούν μεταξύ τους. Επιπλέον, η εφαρμογή ειδικών μετρικών μπορεί να αποκαλύψει τη σημαντικότητα συγκεκριμένων αθλητών, να συμβάλλει στη διαμόρφωση ομάδων και την ανακάλυψη ειδικών κοινοτήτων ενδιαφέροντος για περαιτέρω ανάλυση. Συνολικά, η Ανάλυση Κοινωνικών Δικτύων (*Complex Network Analysis - CNA*), είναι ένας νέος κλάδος της Επιστήμης που μπορεί να συμβάλλει και στο χώρο του αθλητισμού, όχι μόνο παρέχοντας απλώς μια γραφική αναπαράσταση συνδέσεων, αλλά βοηθά στην ανακάλυψη βα-

¹<https://www.strava.com/>

θύτερων διασυνδέσεων και μοτίβων που επηρεάζουν τον χώρο του αθλητισμού.

Ενώ, η ανάλυση δικτύων είναι μια σημαντική μεθοδολογική προσέγγιση σε πολλούς κλάδους των κοινωνικών και φυσικών επιστημών, πρόσφατα βρέθηκε στο επίκεντρο των ερευνητών ο κλάδος του αθλητισμού. Σκοπός της παρούσας εργασίας, είναι η κατανόηση και ανάλυση αθλητικών δικτύων, που αναπτύσσονται συγκεκριμένα στην πλατφόρμα *Strava*, αλλά και ως μεθοδολογία γενικότερα, όπου υπάρχουν διαθέσιμα τα αντίστοιχα δεδομένα, όπου μπορεί να γίνει εφαρμογή τεχνικών *CNA* και δημιουργία μιας λεπτομερούς εικόνας των δομών και των αλληλεπιδράσεων, που δημιουργούνται ανάμεσα στους αθλητές. Με την εφαρμογή και παρουσίαση διαφόρων μετρικών *CNA*, επιδιώκεται η εμβάθυνση και η κατανόηση των συσχετίσεων και των πιθανών κοινοτήτων εντός των δικτύων αυτών. Τέλος, παρουσιάζουμε κάποια συμπεράσματα, που αναδεικνύουν τη σημασία της ανάλυσης στον χώρο των αθλητικών δικτύων.

1.3 Δομή Εργασίας

Η παρούσα διπλωματική εργασία, επικεντρώνεται στην ανάλυση αθλητικών κοινωνικών δικτύων της πατφόρμας *Strava*, ενώ παράλληλα παρουσιάζονται κάποιες βασικές έννοιες της ανάλυσης δικτύων. Αρχικά, στο δεύτερο κεφάλαιο περιέχεται μια βιβλιογραφική ανασκόπηση, που περιλαμβάνει πληροφορίες για την ανάλυση κοινωνικών δικτύων, εισαγωγή στην θεωρία γράφων και γράφων γνώσης, παρουσίαση των βασικών μετρικών και εννοιών. Στο τρίτο κεφάλαιο, παρουσιάζονται η περιγραφή των δεδομένων που χρησιμοποιήθηκαν στην εργασία, ο τρόπος συλλογής τους, όπως επίσης η εξαγωγή των γραφημάτων και των βασικών χαρακτηριστικών τους (π.χ. *Average Degree*, *Shortest Path*, *Clustering Coefficient*, κλπ) και διαγράμματα των χαρακτηριστών αυτών. Στο τέταρτο κεφάλαιο, περιλαμβάνονται οι μεθοδολογίες που ακολουθήθηκαν για την εξαγωγή των χαρακτηριστικών του τρίτου κεφαλαίου και οι αλγόριθμοι για την εξαγωγή κοινοτήτων των γραφημάτων. Οι μεθοδολογίες αυτές, συμπεριλαμβάνουν τον κώδικα και την επεξήγηση του. Στο πέμπτο κεφάλαιο, παρατίθενται με λεπτομέρεια τα αποτελέσματα από τον εντοπισμό κοινοτήτων, με την μορφή αναλυτικών πινάκων και χρωματισμένων γραφημάτων. Τέλος, στο έκτο κεφάλαιο της εργασίας, υπάρχουν τα κυριότερα ευρήματα, η συνεισφορά της εργασίας αυτής καθώς και προτάσεις για μελλοντικές έρευνες.

ΚΕΦΑΛΑΙΟ 2

ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

2.1 Ανάλυση κοινωνικών δικτύων

Την τελευταία δεκαετία, η μελέτη περίπλοκων δικτύων έχει γίνει ένα αναπτυσσόμενο πεδίο έρευνας με έντονο διεπιστημονικό χαρακτήρα. Πολλά διαφορετικά φαινόμενα στον φυσικό, βιολογικό και κοινωνικό κόσμο μπορούν να γίνουν κατανοητά ως δίκτυα. Βασίζονται κυρίως σε κάποιο περίπλοκο (καθώς και εξελισσόμενο), μοτίβο διμερών συνδέσεων μεταξύ μεμονωμένων οντοτήτων και η συνολική απόδοση του συστήματος διαμορφώνεται σε μεγάλο βαθμό από την περίπλοκη αρχιτεκτονική των συνδέσεων που περιλαμβάνονται [68] [37]. Οι άνθρωποι, οραματίζονται συνήθως την κοινωνική δικτύωση ως ομάδες συναδέλφων, ομάδες φοιτητών που συμμετέχουν σε κοινές δραστηριότητες κλπ. Αυτές οι μικρές ομάδες είναι όλες κοινωνικά δίκτυα. Τυπικά, μπορούμε να ορίσουμε το κοινωνικό δίκτυο ως *”Μια κοινωνική δομή αποτελούμενη από κόμβους (συνήθως άτομα ή επιχειρήσεις) οι οποίοι συνδέονται μεταξύ τους με έναν ή περισσότερους τύπους αλληλεξάρτησης, όπως αξίες, οράματα, ιδέες, οικονομικές συναλλαγές, φιλία, συγγένεια, αντιπάθεια, συγκρούσεις ή επιγραμματικές (web) επαφές”* [1].

Αντίστοιχα οι πλατφόρμες κοινωνικής δικτύωσης, μπορούν να οριστούν ως *”διαδικτυακές υπηρεσίες που επιτρέπουν σε άτομα:*

1. να δημιουργούν ένα δημόσιο ή ημι-δημόσιο προφίλ μέσα σε ένα περιορισμένο σύστημα,
2. να δημοσιοποιούν μια λίστα άλλων χρηστών με τους οποίους μοιράζονται μια επαφή/σχέση και

3. να βλέπουν και να διαμοιράζονται τις λίστες των επαφών/σχέσεων τους και αυτές που δημιουργούνται από άλλους μέσα στο σύστημα” [15].

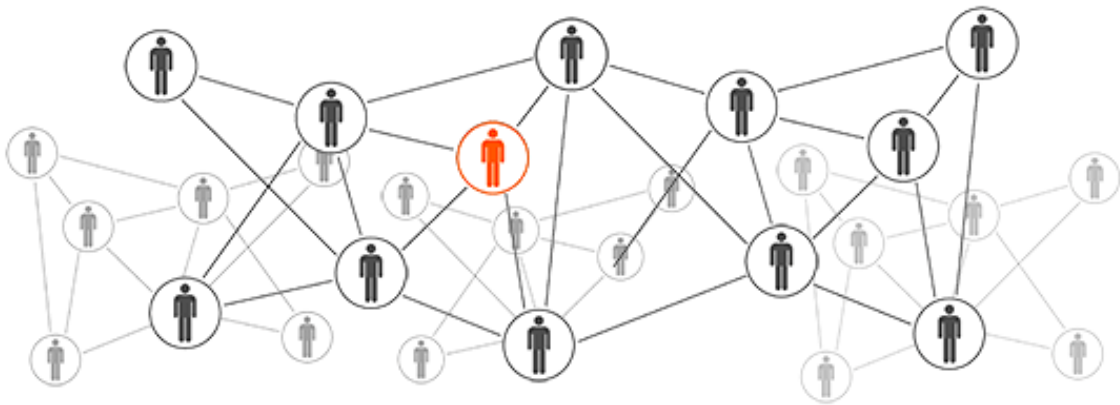
Η έντονη ανάγκη επικοινωνίας του ατόμου με τους ομοίους του, έχει βρει έκφραση μέσω των κοινωνικών δικτύων. Έχουν ενσωματωθεί στην καθημερινή ζωή, σχηματίζοντας μια νέα πραγματικότητα που πριν από λίγα χρόνια δεν υπήρχε. Η συνεχής ανάπτυξη των κοινωνικών δικτύων ακολουθεί τον σύγχρονο τρόπο ζωής, δημιουργώντας και εμπλουτίζοντας έναν ψηφιακό χώρο όπου οι άνθρωποι επικοινωνούν, ενημερώνονται, κλπ. Η ανάπτυξη τους, συνεχίζει να διευρύνεται καθημερινά, αντικατοπτρίζοντας τη σημαντικότητα τους στη σύγχρονη εποχή [6].

Στην εποχή μας, τα δεδομένα που παράγονται από τις εφαρμογές μπορούν να αναπαρασταθούν ως ένα δίκτυο διασυνδεδεμένων αντικειμένων. Ο κύριος στόχος, είναι η εξαγωγή όσο το δυνατόν περισσότερων πληροφοριών. Ένα κοινωνικό δίκτυο, μπορεί να κατασκευαστεί από σχεσιακά δεδομένα ¹ και μπορεί να οριστεί, ως ένα σύνολο κοινωνικών οντοτήτων, όπως άνθρωποι, ομάδες και οργανισμοί, με σχέσεις ή αλληλεπιδράσεις που υπάρχουν μεταξύ τους. Αυτά τα δίκτυα μοντελοποιούνται τις περισσότερες φορές με γραφήματα. Η δομή τέτοιων δικτύων είναι το αντικείμενο μελέτης της ανάλυσης κοινωνικών δικτύων (*Social Network Analysis - SNA*). Έτσι, μέθοδοι και τεχνικές σχεδιάστηκαν για να ανακαλύψουν πρότυπα και αλληλεπιδράσεις μεταξύ κοινωνικών παραγόντων, σε τέτοια δίκτυα. Ως εκ τούτου, η εστίαση του *SNA* είναι στις σχέσεις που δημιουργούνται μεταξύ κοινωνικών οντοτήτων. Στην πραγματικότητα, ο κύριος στόχος αυτής της τεχνικής είναι να εξεταστεί το περιεχόμενο και τα μοτίβα σχέσεων, με σκοπό την εξαγωγή σημαντικών συμπερασμάτων για τις σχέσεις μεταξύ οντοτήτων και τις επιπτώσεις τους [65].

Τα δύο απαραίτητα στοιχεία κάθε κοινωνικού δικτύου, είναι οι οντότητες (κόμβοι) και οι σχέσεις (συνδέσεις), μεταξύ τους. Ο συνδυασμός τους, αποτελεί από κοινού ένα δίκτυο. Οι οντότητες μπορεί να είναι μεμονωμένα φυσικά πρόσωπα ή συλλογικοί παράγοντες όπως ομάδες και επίσημες οργανώσεις (Σχήμα 2.1) [17]. Συνήθη παραδείγματα μεμονωμένων παραγόντων, περιλαμβάνουν κυρίως ξεχωριστά άτομα που δραστηριοποιούνται σε ένα κοινό αντικείμενο, π.χ. συναδέλφους. Συλλογικοί παράγοντες, μπορεί να είναι εταιρείες που δραστηριοποιούνται σε έναν κοινό κλάδο, εθελοντικές ομάδες, κόμματα, κλπ. Άλλοι τύποι οντοτήτων, στερούνται ανθρώπινης φύσης όπως βιβλία, αθλήματα, συστήματα υγειονομικής περίθαλψης (γιατρούς, νοσηλευτές, ασθενείς, κλινικές, νοσοκομεία, κ.α). Στη συνέχεια, ως σχέση ορίζεται γενικά

¹ Δεδομένα οργανωμένα σε συσχετισμένους πίνακες που παρέχει ταυτόχρονα ένα μηχανισμό για ανάγνωση, εγγραφή, τροποποίηση ή και πιο πολύπλοκες διαδικασίες πάνω στα δεδομένα

ένα συγκεκριμένο είδος επαφής, σύνδεσης ή δεσμού μεταξύ ενός ζευγαριού οντοτήτων. Μια σχέση δεν είναι χαρακτηριστικό μιας οντότητας, αλλά μια κοινή ιδιότητα μεταξύ οντοτήτων. Μια τεράστια ποικιλία σχέσεων μεταξύ ατομικών και συλλογικών οντοτήτων, μπορεί να σχετίζεται με την αναπαράσταση δομών δικτύου και την εξήγηση των επιπτώσεών τους. Ποιο συγκεκριμένο είδος σχέσης πρέπει να μετρήσει ένας ερευνητής δικτύου, εξαρτάται από τις συγκεκριμένες εφαρμογές του κάθε ερευνητικού έργου.



Σχήμα 2.1: Αναπαράσταση ενός κοινωνικού δικτύου [17].

Η προσέγγιση αυτή, έχει αλλάξει τον τρόπο, όπου τα δίκτυα εξετάζονται για την παροχή μεθόδων και εργαλείων που χρησιμεύουν στην μοντελοποίηση, την ανάλυση και την επίλυση συμβατικών προβλημάτων. Αν αναφερθούμε, για παράδειγμα, στον Παγκόσμιο Ιστό, μπορεί να αναπαρασταθεί ως ένα πολύπλοκο δίκτυο θεωρώντας τα συστήματα ή τις συσκευές ως κόμβους και την επικοινωνία μεταξύ τους. Ένα κοινωνικό δίκτυο, είναι μια δομή που αποτελείται από ένα σύνολο οντοτήτων, μερικά από τα μέλη των οποίων συνδέονται με ένα σύνολο από μία ή περισσότερες σχέσεις. Αυτά τα δύο θεμελιώδη στοιχεία, είναι κοινά στους περισσότερους ορισμούς δικτύου. Διαφορετικοί τύποι σχέσεων προσδιορίζουν διαφορετικά δίκτυα, ακόμη και όταν οι παρατηρήσεις περιορίζονται στο ίδιο σύνολο οντοτήτων. Η δήλωση ότι υπάρχουν συνδέσεις μεταξύ των μελών ενός δικτύου, δεν απαιτεί όλα τα μέλη να έχουν άμεσες σχέσεις με όλα τα άλλα [44].

Εκτός των κοινωνικών δικτύων, πολλά δίκτυα δεδομένων, όπως αναφέρθηκε και παραπάνω, σχηματίζονται επίσης μεταξύ αντικειμένων διαφορετικών των κοινωνικών οντοτήτων. Τα μη κοινωνικά δίκτυα είναι διαδεδομένα σε πολλούς τομείς: δίκτυα τεχνολογίας, υπολογιστικά συστήματα, Παγκόσμιος Ιστός, τηλεφωνικά δίκτυα, δίκτυα μεταφορών, αισθητήρες,

λέξεις/κείμενα, γεωγραφικές τοποθεσίες, εγκεφαλικοί νευρώνες, κλπ. Μοιράζονται κάποιες ομοιότητες με τα κοινωνικά δίκτυα, εκτός από το ότι τα αντικείμενα τους είναι φυσικές οντότητες, όπως υπολογιστές και μετασχηματιστές, και οι σχέσεις μεταξύ τους είναι γραμμές μετάδοσης όπως καλώδια *Ethernet*, ασύρματες συνδέσεις, αεροπορικές διαδρομές, κλπ. Τα δίκτυα αυτά, αποτελούν αντικείμενο μελετών από πολλούς κλάδους εκτός από τις κοινωνικές επιστήμες. Αν και οι τεχνικές *SNA*, δημιουργήθηκαν κυρίως για την ανάλυση κοινωνικών δικτύων, μπορούν να χρησιμοποιηθούν και σε μη κοινωνικά δεδομένα. Τέτοιες τεχνικές, περιλαμβάνουν τον εντοπισμό δημοφιλών/σημαντικών κόμβων, την ανίχνευση κοινοτήτων, κλπ. Αυτές οι διαδικασίες είναι χρήσιμες στην εξαγωγή γνώσης.

2.2 Γράφοι Γνώσης

Τα τελευταία χρόνια, τα γραφήματα γνώσης (KGs - knowledge graphs), έχουν γίνει η βάση πολλών πληροφοριακών συστημάτων που απαιτούν πρόσβαση σε δομημένη γνώση. Οι κόμβοι σε τέτοια γραφήματα αντιπροσωπεύουν οντότητες και συνδέονται με ακμές που αντιπροσωπεύουν σχέσεις. Η έννοια των συνδεδεμένων δεδομένων [11], εμφανίστηκε το 2009. Προτείνεται, η σύνδεση διαφορετικών συνόλων δεδομένων μεταξύ τους στον Σημασιολογικό Ιστό, ώστε να αντιμετωπίζονται ως ένα μεγάλο, παγκόσμιο γράφημα γνώσης. Μέχρι το 2014, περίπου 1000 σύνολα δεδομένων συνδέονται μεταξύ τους στο Νέφος Συνδεδεμένων Ανοικτών Δεδομένων (Linked Open Data cloud), οι περισσότεροι μεταξύ τους σύνδεσμοι, συνδέουν πανομοιότυπες οντότητες [61].

Πιο συγκεκριμένα, ως γράφημα γνώσης ορίζεται ως μια μεγάλης κλίμακας βάση γνώσεων, που αποτελείται από μεγάλο αριθμό οντοτήτων και σχέσεων που υπάρχουν μεταξύ τους. Τα γραφήματα γνώσης, είναι μια νέα μέθοδος αναπαράστασης γνώσης. Στην ουσία, ο σημασιολογικός ιστός είναι μια πρώιμη μορφή του γραφήματος γνώσης, η οποία είναι μια έννοια που περιγράφει οντότητες και σχέσεις μεταξύ οντοτήτων στον πραγματικό κόσμο και είναι επίσης μια δικτυωμένη βάση γνώσης που αποτελείται από οντότητες, ιδιότητες και σχέσεις. Ένα γράφημα γνώσης είναι μια συλλογή εννοιών, οντοτήτων και των σχέσεών τους στον αφηρημένο φυσικό κόσμο [18]. Ως οντότητα (*entity*), ορίζεται ένα αντικείμενο του πραγματικού κόσμου, που συγκεντρώνει κάποιο επιστημονικό ή μη ενδιαφέρον, το οποίο ξεχωρίζει από τα υπόλοιπα [27], μπορεί να είναι άνθρωποι, μέρη, αντικείμενα, γεγονότα, έννοιες κλπ. Πολύ γνωστά παραδείγματα χρήσης των γραφημάτων γνώσης τα τελευταία χρόνια, ως σημασιολογικά δίκτυα, είναι η επεξεργασία φυσικής γλώσσας, τα έξυπνα αντικείμενα, κλπ. Μαζί με τα μεγάλα δεδομένα και τη βαθιά μάθηση, τα γραφήματα γνώσης έχουν πλέον γίνει ένας

από τους βασικούς παράγοντες ανάπτυξης της τεχνητής νοημοσύνης [21].

Τα γραφήματα γνώσης, έχουν προκαλέσει μεγάλο ενδιαφέρον, τόσο στον βιομηχανικό, όσο και στον ακαδημαϊκό κόσμο. Παρέχουν σημασιολογικά δομημένες πληροφορίες, που μπορούν να ερμηνευθούν από υπολογιστές και μια τέτοια ιδιότητα θεωρείται ότι είναι πολλά υποσχόμενη, για την κατασκευή πιο έξυπνων μηχανών από πολλούς ανθρώπους. Πολλές ανασκοπήσεις, σχετικά με τα γραφήματα γνώσης επικεντρώθηκαν στις τεχνικές κατασκευής [76], [79], ωστόσο, αν και τα γραφήματα γνώσης έχουν ήδη υποστηρίξει πολλαπλές εφαρμογές "Big Data" σε διάφορους εμπορικούς και επιστημονικούς τομείς, δεν υπάρχει κριτική για την εφαρμογή τους.

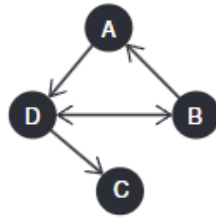
Οι έρευνες για KG, μπορούν να ταξινομηθούν σε δύο κατηγορίες, τις έρευνες για τις τεχνικές κατασκευής KGs και την εφαρμογή KGs. Οι μελέτες για τις τεχνικές κατασκευής, επικεντρώνονται στην εξαγωγή, αναπαράσταση και σύνθεση της γνώσης στα γραφήματα [76]. Τα γραφήματα γνώσης, μπορούν να χωριστούν σε γραφήματα γενικής γνώσης και γραφήματα γνώσης τομέα. Για παράδειγμα, τα γραφήματα γνώσης που χρησιμοποιούνται στην τοπογραφία είναι γραφήματα γνώσης τομέα. Χρησιμοποιούνται, επίσης, γραφήματα γνώσης γεωεπιστημών, τα οποία έχουν μελετηθεί λεπτομερώς. Οι Xu et al. [10], πρότειναν την έννοια, το πλαίσιο, τη θεωρία και τα χαρακτηριστικά των γραφημάτων γεωεπιστήμης. Ο Jiang, πρότεινε τη διαδικασία κατασκευής γραφημάτων γνώσης και διερεύνησε τη βασική τεχνολογία των γραφημάτων γεωγραφικής γνώσης. Οι Lu et al. εξέτασαν την πρόοδο της έρευνας σε θέματα που σχετίζονται με γραφήματα γεωγραφικής γνώσης και ανέλυσαν τα βασικά ζητήματα της τρέχουσας κατασκευής γραφημάτων γνώσης της γεωεπιστήμης [43]. Στον ιατρικό τομέα, έχει επίσης διερευνηθεί η κατασκευή γραφημάτων γνώσης. Υπάρχουν σχετικά πλήρεις βάσεις ιατρικής γνώσης, σε αυτόν τον τομέα, όπως το ICD-11 [36], που χρησιμοποιεί μια δομή δέντρου για να περιγράψει ασθένειες και το UMLS [80], [12], που χρησιμοποιεί μια δομή για την αποθήκευση ιατρικών πληροφοριών. Κινέζοι μελετητές, έχουν κατασκευάσει επίσης γραφήματα γνώσης σχετικά με την παραδοσιακή κινεζική ιατρική [40], [60], [78]. Τα παραπάνω γραφήματα γνώσης, περιέχουν μόνο ιατρικές έννοιες.

Η μεγάλη αξία των γραφημάτων γνώσης έχει βρεθεί σε ένα ευρύ φάσμα εφαρμογών του πραγματικού κόσμου, συμπεριλαμβανομένης, ακόμη, της κατανόησης κειμένου, των συστημάτων συστάσεων και των συστημάτων απαντήσεων σε ερωτήσεις φυσικής γλώσσας. Έχουν δημιουργηθεί όλο και περισσότερα γραφήματα γνώσης, που καλύπτουν τη γνώση της κοινής λογικής (π.χ. Cyc [45], ConceptNet [41]), τη λεξιλογική γνώση (π.χ. WordNet [47], BabelNet [50]), την εγκυκλοπεδική γνώση (π.χ. Freebase [13], DBpedia [5], YAGO [64], WikiData

[70], CN-DBpedia [75]), τη γνώση της ταξινόμησης (π.χ. Probase [74]) και την γεωγραφική γνώση (π.χ. GeoNames [73])

Η εξαγωγή πληροφοριών είναι ένα θεμελιώδες στοιχείο κατασκευής γραφήματος γνώσης, είτε είναι συγκεκριμένος ο τομέας, είτε γενικός [46], [49]. Όπως υποδηλώνει το όνομα, ο στόχος ενός συστήματος εξαγωγής πληροφορίας, είναι να εξάγει χρήσιμες πληροφορίες από "ακατέργαστα" δεδομένα. Οι πληροφορίες έχουν πολλές διαστάσεις, αλλά η πιο σημαντική, για υπολογιστικούς σκοπούς, είναι ότι μπορούν να επεξεργαστούν και να αιτιολογηθούν από μηχανές. Ένας από τους λόγους για τους οποίους η εξαγωγή πληροφοριών αναγνωρίστηκε ως πρώιμο πρόβλημα σε κλάδους, όπως η επεξεργασία φυσικής γλώσσας, ήταν επειδή οι μηχανές δεν είναι καλές στην κατανόηση της φυσικής γλώσσας όπως τα Αγγλικά ή τα Ελληνικά λόγω προβλημάτων όπως η ασάφεια και η παρατυπία. Ακόμη και σήμερα, παρά τη μεγάλη πρόοδο, οι μηχανές εξακολουθούν να μην μπορούν να διαβάσουν και να κατανοήσουν γλώσσες, τόσο καλά όσο οι άνθρωποι. Έτσι, ένας πρωταρχικός στόχος είναι, να εξαχθούν τα βασικά κομμάτια πληροφοριών, όπως οντότητες, σχέσεις και ιδιότητες, συνήθως, από κείμενο φυσικής γλώσσας [38]. Τα γραφήματα προϋπάρχουν και αποθηκεύονται με βάση δύο τύπους δομών δεδομένων: ο πρώτος είναι δομή λίστας και ο δεύτερος είναι δομή πίνακα. Αυτές οι δομές είναι κατάλληλες για την αποθήκευση γραφημάτων σε υπολογιστές προκειμένου να αναλυθούν περαιτέρω χρησιμοποιώντας αυτόματα εργαλεία. Η δομή λίστας, όπως λίστες περιπτώσεων και λίστες γειτνίασης, είναι κατάλληλες για την αποθήκευση αραιών γράφων, καθώς μειώνουν τον απαιτούμενο χώρο αποθήκευσης. Από την άλλη πλευρά, η δομή πινάκων περιλαμβάνουν τους πίνακες γειτνίασης ή κοινωνιομετρικοί πίνακες, τους πίνακες απόστασης (πανομοιότυποι με τους πίνακες γειτνίασης με τη διαφορά ότι οι εγγραφές του πίνακα είναι τα μήκη από τα συντομότερα μονοπάτια μεταξύ ζευγών κορυφών), κλπ. Μπορούν να χρησιμοποιηθούν διάφοροι τύποι γραφημάτων, για τη μοντελοποίηση διαφορετικών ειδών κοινωνικών δικτύων. Για παράδειγμα, κάποια γραφήματα μπορούν να ταξινομηθούν ανάλογα με την κατεύθυνση των συνδέσμων τους. Αυτό μας οδηγεί, στη διαφοροποίηση μεταξύ μη κατευθυνόμενων και κατευθυνόμενων γραφημάτων. Τα μη κατευθυνόμενα γραφήματα (ή δίκτυα), είναι γραφήματα των οποίων οι ακμές συνδέουν μη διατεταγμένα ζεύγη κορυφών ή, με άλλα λόγια, η κάθε ακμή του γραφήματος συνδέει ταυτόχρονα δύο κορυφές (κόμβους). Άλλος ένας τύπος γραφήματος είναι το κατευθυνόμενο γράφημα (ή κατευθυνόμενο δίκτυο). Τα κατευθυνόμενα γραφήματα, ή στη συντομογραφικά *digraphs*, μπορούν να οριστούν ως γραφήματα των οποίων όλες οι ακμές έχουν έναν προσανατολισμό που έχει εκχωρηθεί (όμοια με τόξα), όπως φαίνεται στο Σχήμα 2.2.

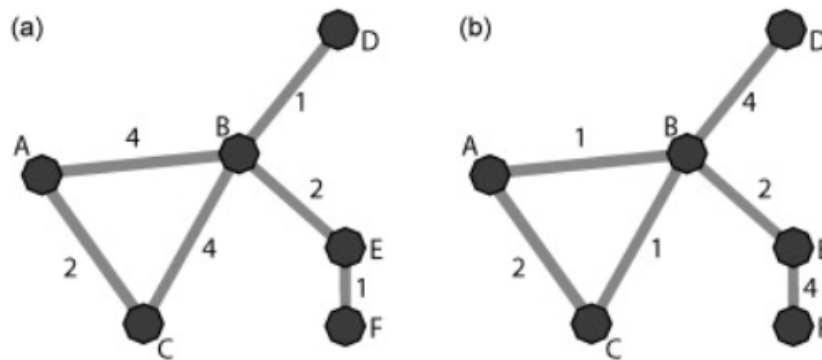
Γραφικά, οι κατευθυνόμενες άκρες απεικονίζονται με βέλη, υποδεικνύοντας την κατεύθυνση



Σχήμα 2.2: Κατευθυνόμενο δίκτυο [65].

της σύνδεσης. Αυτός ο τύπος γραφημάτων μπορεί να είναι κυκλικός, δηλαδή γραφήματα που περιέχουν κλειστούς βρόχους ακμών ή άκυκλος (π.χ. δέντρα). Ένα άτυπο παράδειγμα ενός μη κατευθυνόμενου γραφήματος είναι το *Facebook*, δεδομένου ότι, σε αυτό το κοινωνικό δίκτυο, ο δεσμός φιλίας είναι αμοιβαίος (π.χ. εάν δεχτώ ένα αίτημα φιλίας από κάποιο άτομο, τότε θεωρείται ότι είμαστε φίλοι). Ομοίως, το *X* (πρώην *Twitter*) είναι ένα παράδειγμα κατευθυνόμενου γραφήματος, καθώς ένα άτομο μπορεί να ακολουθείται από άλλους χρήστες, χωρίς αυτό το άτομο να τους ακολουθεί απαραίτητα. Σε αυτή την περίπτωση, οι δεσμοί μεταξύ ζευγαριών ατόμων κατευθύνονται, πράγμα που σημαίνει ότι δημιουργούνται μονόδρομες σχέσεις. Αναφορικά με τις τιμές που αποδίδονται στις ακμές, γίνεται διάκριση μεταξύ γραφημάτων με βάρη και μη (Σχήμα 2.3). Εκτός και αν αναφέρεται ρητά, πάντα υποθέτουμε ότι τα γραφήματα δεν έχουν βάρη. Τα γραφήματα χωρίς βάρη είναι δυαδικά, καθώς οι ακμές τους είτε υπάρχουν, είτε όχι. Σε άλλα είδη δικτύων, το βάρος μιας σχέσης μπορεί να αντιπροσωπεύει μια ποικιλία πραγμάτων. Για παράδειγμα, μια σχέση μπορεί να αντιπροσωπεύει τον αριθμό των θέσεων μεταξύ των αεροδρομίων, τον αριθμό των προϊόντων που ανταλλάσσονται, κλπ. Για μη κατευθυνόμενα και χωρίς βάρη γραφήματα, οι πίνακες γειτνίασης είναι δυαδικοί και συμμετρικοί. Το $a_{ij} = 1$, αντιπροσωπεύει την παρουσία μιας ακμής μεταξύ των κορυφών, ενώ το $a_{ij} = 0$ αντιπροσωπεύει την απουσία ακμής μεταξύ του ζεύγους κορυφών (i, j) . Για κατευθυνόμενα, με βάρη γραφήματα, οι εγγραφές των πινάκων που τα περιγράφουν λαμβάνουν τιμές από το διάστημα $[0, max(weight)]$ και είναι, όχι σε όλες τις περιπτώσεις, αλλά κατά κανόνα ασύμμετροι. Και οι δύο περιπτώσεις γραφημάτων, αφορούν μη αρνητικούς πίνακες [65].

Κατά την αναπαράσταση γραφημάτων, η βασική πρόκληση προκύπτει από την ανάγκη εξαγωγής σημαντικής πληροφορίας, με ταυτόχρονη απομόνωση και απόρριψη "θορύβων" και περιττών πληροφοριών. Κατά την κατασκευή των γραφημάτων, οι οντότητες παρουσιάζονται ως κόμβοι (*nodes*), οι κατηγορίες τους χαρακτηρίζονται με ετικέτες (*labels*), ενώ οι αλλη-



Σχήμα 2.3: Κατευθυνόμενα δίκτυα με βάρη [55].

λεπιδράσεις μεταξύ κόμβων αναπαρίστανται με συνδέσεις (*edges*):

- **Ορθή εξαγωγή οντοτήτων.** Συνήθως σε γραφήματα αναπαράστασης φυσικής γλώσσας, όπου η ίδια έννοια μπορεί να εκφραστεί με διάφορους τρόπους. Αυτό, συχνά, οδηγεί στην περιττή δημιουργία κόμβων που στην πραγματικότητα, θα έπρεπε η πληροφορία να συγκεντρώνονται σε έναν.
- **Ορθή εξαγωγή ετικετών.** Η χρήση κατάλληλων ετικετών, σχετικών ή γειτονικών οντοτήτων, μπορεί να βοηθήσει στον προσδιορισμό ετικέτας μιας γειτονικής οντότητας ενδιαφέροντος.
- **Ορθός προσδιορισμός σχέσεων.** Σε αρκετές περιπτώσεις, μπορεί να υπάρχουν σχέσεις, οι οποίες είναι αληθινές, ενώ κάποιες να είναι ψευδείς ή αντιφατικές.

Η αντιμετώπιση αυτών των προκλήσεων, απαιτεί προηγμένες τεχνικές αναγνώρισης και επεξεργασίας πληροφοριών. Προηγμένοι αλγόριθμοι μηχανικής μάθησης και διάφορες τεχνικές εξόρυξης δεδομένων μπορούν να εφαρμοστούν σε τέτοιες περιπτώσεις [57].

2.3 Θεωρία Γραφημάτων

Ακόμη, υπάρχουν ορισμένες μετρικές γραφημάτων, που χρησιμοποιούνται στην ανάλυση των δικτύων και μας παρέχουν πληροφορίες σχετικά με το ρόλο των κόμβων στο δίκτυο. Η μελέτη του ρόλου των κόμβων και του τρόπου αλληλεπίδρασής τους στο πλαίσιο του δικτύου, στοχεύει στην κατανόηση της συμπεριφοράς των κοινωνικών και μη συστημάτων που

δημιούργησαν τα εκάστοτε δίκτυα, κάτι που είναι συνήθως ο τελικός στόχος μιας ανάλυσης. Τα μέτρα αυτά, μπορούν να χωριστούν ανάλογα με το επίπεδο ανάλυσης που θέλει να εκτελέσει κάποιος: σε επίπεδο μικρής έκτασης, όπως κόμβοι, ή σε επίπεδο ολόκληρου του δικτύου. Η πρώτη, διερευνά γενικές μετρήσεις κεντρικότητας ως τρόπο κατανόησης της θέσης μιας κορυφής στη συνολική δομή του γραφήματος και, επομένως, βοηθά στον εντοπισμό των βασικών παραγόντων του δικτύου. Ενώ, το τελευταίο παρέχει πιο συμπαγείς πληροφορίες και επιτρέπει την αξιολόγηση της συνολικής δομής του δικτύου, δίνοντας πληροφορίες για σημαντικές ιδιότητες των υποκείμενων δομών δικτύου. Στη συνέχεια, ακολουθούν κάποιες από τις βασικές μετρικές που χρησιμοποιούνται για την ανάλυση γραφημάτων.

2.3.1 Βαθμός - Degree

Ο βαθμός ενός κόμβου, που συνήθως συμβολίζεται ως k_v , είναι ένα μέτρο της άμεσης γειτνίασης και της συμμετοχής του κόμβου στο δίκτυο και υπολογίζεται ως ο αριθμός των ακμών που προσπίπτουν σε έναν δεδομένο κόμβο ή, ομοίως, ως ο αριθμός των γειτόνων του κόμβου v . Η γειτονιά N_v , ορίζεται από το σύνολο των κόμβων που συνδέονται άμεσα με τον v . Ο βαθμός, μπορεί να υπολογιστεί με, τουλάχιστον, δύο διαφορετικούς τρόπους: με βάση τον πίνακα γειτνίασης ή με βάση τη γειτονιά ενός κόμβου. Στις παρακάτω εξισώσεις, παρουσιάζονται καθεμία από τις εναλλακτικές λύσεις για μη κατευθυνόμενα δίκτυα.

1. $k_i = \sum_{j=1}^n a_{ij}$, $0 < k_i < n$, όπου a_{ij} είναι η καταχώρηση της σειράς και της στήλης του πίνακα γειτνίασης.
2. $k_v = |N(v)|$, $0 < k_v < n$, όπου $|N(v)|$ είναι η γειτονιά του κόμβου v .

Παρά την απλότητά του, ο βαθμός είναι ένα αποτελεσματικό μέτρο για την αξιολόγηση της σημασίας και της επιρροής ενός κόμβου σε ένα κοινωνικό δίκτυο. Ωστόσο, έχει κάποιους περιορισμούς. Το κυριότερο είναι ότι δεν λαμβάνει υπόψη την καθολική δομή του δικτύου. Για τα κατευθυνόμενα δίκτυα, υπάρχουν δύο παραλλαγές βαθμού: *in-degree*, που συμβολίζεται με k_v^+ , και *out-degree*, με k_v^- . Το πρώτο δίνεται από τον αριθμό των εισερχόμενων κόμβων (δηλαδή, τον αριθμό των ακμών που καταλήγουν στο κόμβο v) και το δεύτερο, από τον αριθμό των εξερχόμενων κόμβων, (δηλαδή, τον αριθμό των ακμών που ξεκινούν από το κόμβο v). Το μέτρο των κατευθυνόμενων δικτύων αναφέρεται επίσης ως κύρος (*prestige*). Αυτή η έκφραση, χρησιμοποιείται ιδιαίτερα στη βιβλιογραφία των κοινωνικών δικτύων, δεδομένου ότι αναπτύχθηκε για τη μέτρηση της θέσης ή της σημασίας των κόμβων στο δίκτυο. Υπάρχουν δύο τύποι μέτρησης κύρους: υποστήριξη και επιρροή. Η πρώτη σχετίζεται με την κεντρική

θέση εντός βαθμού, η οποία θεωρείται ως μέτρο υποστήριξης, και η δεύτερη σχετίζεται με την κεντρικότητα του *in-degree*, η οποία χαρακτηρίζεται ως μέτρο επιρροής. Το *in-degree* ενός κόμβου είναι το άθροισμα των βαρών των ακμών που καταλήγουν στον κόμβο αυτό. Υπολογίζεται από τον τύπο:

$$k_i^- = \sum_{j=1}^n a_{ji},$$

ο εξωτερικός βαθμός ενός κόμβου, *out-degree* είναι το άθροισμα των βαρών των ακμών που εξέρχονται από τον κόμβο και υπολογίζονται, ομοίως, από τον τύπο:

$$k_i^+ = \sum_{j=1}^n a_{ij},$$

στα δίκτυα με βάρη, ο ισχύς είναι το ισοδύναμο του βαθμού, που υπολογίζεται ως το άθροισμα των βαρών w των ακμών δίπλα σε έναν δεδομένο κόμβο και υπολογίζεται από τον τύπο:

$$k_i^w = \sum_{j=1}^n a_{ij}^w.$$

Μια σημαντική ερευνητική προσπάθεια έγινε για τη μελέτη της κατανομής βαθμών πολλών τύπων δικτύων, η οποία κατέστησε δυνατή την ταξινόμηση ενός δικτύου με βάση αυτή την κατανομή. Για παράδειγμα, οι *Barabási* και *Albert* [9], ανακάλυψαν ότι τα περισσότερα δίκτυα πραγματικού κόσμου ακολουθούν εκθετική κατανομή (νόμου δύναμης), τουλάχιστον ασυμπτωτικά. Αυτό σημαίνει ότι, σε αυτά τα δίκτυα, η κατανομή του βαθμού κόμβων είναι πολύ ετερογενής και εξαιρετικά στρεβλωμένη προς τα δεξιά, με μια μεγάλη πλειονότητα των κορυφών να έχουν χαμηλό βαθμό και έναν μικρό αριθμό με υψηλό βαθμό. Αυτά τα δίκτυα είναι γνωστά ως χωρίς - κλίμακα (*Scale-Free*), μια έκφραση που επινοήθηκε από τους ίδιους ερευνητές, παράδειγμα τέτοιου δικτύου υπάρχει στα Σχήματα 2.12 και 2.11.

2.3.2 Μέσος Βαθμός και Κατανομή Μέσου Βαθμού - Average Degree and Average Degree Distribution

Συγκρίνοντας μεγάλα δίκτυα, είναι εύκολο να προσδιοριστεί ότι δύο δίκτυα είναι διαφορετικά, απλά και μόνο με μια σύντομη λίστα ιδιοτήτων στις οποίες διαφέρουν μεταξύ τους. Είναι όμως δύσκολο να αποδειχθεί, ότι είναι παρόμοια ανάμεσα στις πολλές ιδιότητες που συγκεντρώνει το καθένα.

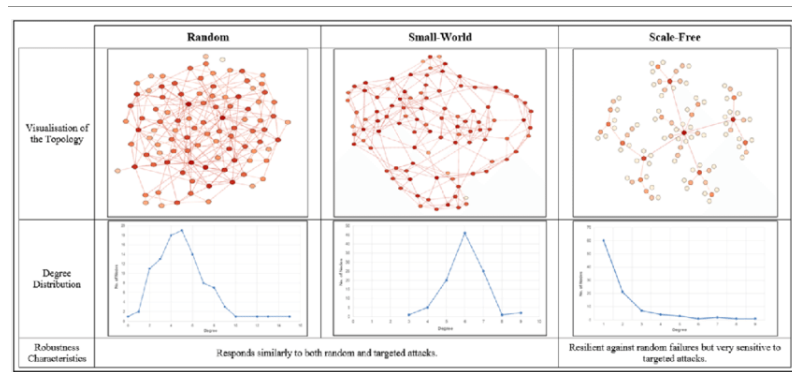
Ο Μέσος Βαθμός (Average Degree) είναι ένας όρος που χρησιμοποιείται στην θεωρία των

γράφων για να περιγράψει το μέσο αριθμό συνδέσεων που έχει ο κάθε κόμβος [37]. Συνήθως, υπολογίζεται από τον τύπο (2.1):

$$\text{Average Degree} = \frac{\text{Συνολικός αριθμός συνδέσεων}}{\text{Συνολικός αριθμός κόμβων}}, \quad (2.1)$$

Το μέτρο αυτό, υποδηλώνει τον βαθμό συνδεσιμότητας των κόμβων ενός γράφου και η τιμή του μπορεί να περιέχει σημαντική πληροφορία, για τον χαρακτήρα του δικτύου. Ανάλογα με την κατανομή του βαθμού των κόμβων, κάθε δικτύου, αναγνωρίζονται πέντε μοντέλα τα οποία υποδηλώνουν τον τύπο του δικτύου και μας παρέχουν χρήσιμες πληροφορίες (Σχήμα 2.4).

Τα μοντέλα αυτά περιλαμβάνουν τα κανονικά γραφήματα, τα τυχαία γραφήματα, τα τυχαία γεωμετρικά, τα γραφήματα μικρού κόσμου και τα γραφήματα ελεύθερης κλίμακας, όπως περιγράφονται στις παρακάτω ενότητες με παραδείγματα των τοπολογιών και των κατανομών των βαθμών των δικτύων που περιγράφουν. Στα πλαίσια της ανάλυσης γράφων, ο μέσος βαθμός



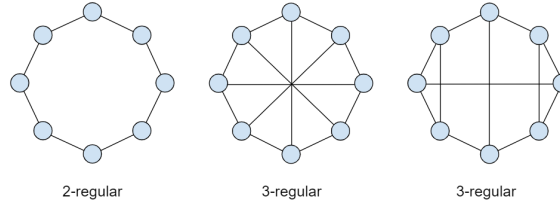
Σχήμα 2.4: Παραδείγματα τοπολογιών και κατανομών βαθμών μοντέλων [56].

μπορεί να δώσει μια εικόνα για το πόσο "συνδεδεμένος" είναι ο γράφος γενικά. Ένας υψηλός μέσος βαθμός υπονοεί ότι οι κόμβοι του γράφου έχουν σε γενικές γραμμές περισσότερες συνδέσεις, ενώ ένας χαμηλός μέσος βαθμός υπονοεί λιγότερες συνδέσεις κατά μέσο όρο. Ο υψηλός μέσος βαθμός υποδεικνύει ότι τα μέλη έχουν τάση να έχουν περισσότερες συνδέσεις, δημιουργώντας ένα πυκνό δίκτυο. Αντίθετα, ένας χαμηλός μέσος βαθμός υποδεικνύει ότι τα μέλη έχουν λιγότερες συνδέσεις και το δίκτυο είναι πιο αραιό, μπορεί δηλαδή να παρέχει ενδιαφέρουσες πληροφορίες για το επίπεδο επικοινωνίας και επιρροής σε ένα δίκτυο.

2.3.2.1 Κανονικό Γράφημα - Regular Graph

Στη θεωρία γραφημάτων, ένα κανονικό γράφημα είναι ένα γράφημα όπου κάθε κορυφή έχει τον ίδιο αριθμό γειτόνων. Δηλαδή, κάθε κορυφή έχει τον ίδιο βαθμό. Ένα κανονικό κατευθυνό-

μενο γράφημα, πρέπει επίσης να ικανοποιεί την ισχυρότερη προϋπόθεση ότι το indegree και το outdegree κάθε εσωτερικής κορυφής είναι ίσα μεταξύ τους [20]. Ένα κανονικό γράφημα με κορυφές βαθμού k , ονομάζεται k -κανονικό γράφημα ή κανονικό γράφημα βαθμού k [8]. Τα παραδείγματα γραφημάτων στο Σχήμα 2.5, είναι κανονικά αφού επιβεβαιώνεται με μια



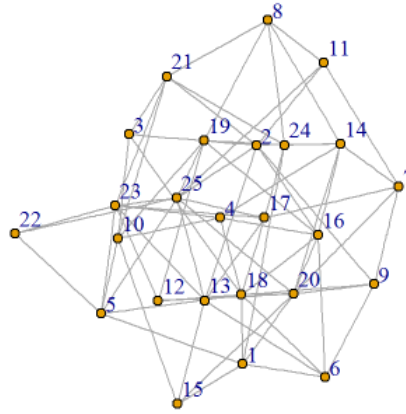
Σχήμα 2.5: Τοπολογία κανονικών γραφημάτων [69].

απλή παρατήρηση, ότι κάθε κορυφή έχει ακριβώς τον ίδιο αριθμό ακμών. Το πρώτο, είναι 2-κανονικό (δύο συνδέσεις ανά κορυφή) και τα δύο επόμενα είναι 3-κανονικά (τρεις συνδέσεις ανά κορυφή).

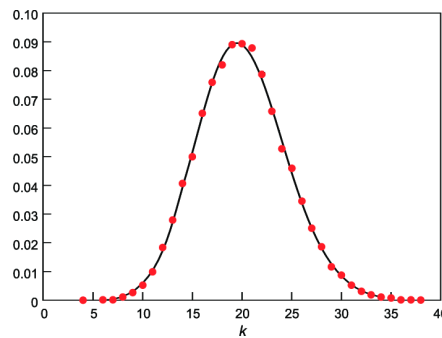
2.3.2.2 Τυχαίο Γράφημα - Random Graph

Ένα τυχαίο γράφημα, είναι ένα γράφημα στο οποίο ιδιότητες όπως ο αριθμός των κορυφών του γραφήματος, οι ακμές του γραφήματος και οι συνδέσεις μεταξύ τους καθορίζονται με κάποιο τυχαίο τρόπο. Σε αυτό το είδος, οι συνδέσεις μεταξύ των κόμβων δημιουργούνται τυχαία με ομοιόμορφη (ίση) πιθανότητα, οδηγώντας σε ένα πιο αραιό δίκτυο (Σχήμα 2.6). Η κατανομή βαθμού κόμβων είναι διωνυμική (*Poisson*) (Σχήμα 2.7).

Οι τυχαίοι γράφοι, αποτελούν σημαντικό πεδίο στη θεωρία γράφων. Ένα δημοφιλές μοντέλο τυχαίων γράφων, είναι το μοντέλο Erdős-Rényi [28], γνωστό και ως μοντέλο $G(n, p)$, όπου n είναι ο αριθμός των κόμβων και p είναι η πιθανότητα σύνδεσης δύο κόμβων. Βασικές ιδέες του μοντέλου αυτού είναι η προσθήκη κόμβων, δηλαδή ο γράφος ξεκινά με έναν μικρό αριθμό κόμβων και κατόπιν προστίθενται σταδιακά νέοι κόμβοι. Κάθε νέος κόμβος, συνδέεται με ήδη υπάρχοντες, με την πιθανότητα σύνδεσης να είναι ανάλογη προς τον αριθμό των συνδέσεων που έχει ήδη ένας κόμβος. Όπως επίσης και ο αυξανόμενος βαθμός κόμβων, οι κόμβοι που έχουν ήδη πολλές συνδέσεις, τείνουν να παίρνουν περισσότερες νέες συνδέσεις. Αυτό οδηγεί, σε κοινωνικούς γράφους με λίγους κεντρικούς κόμβους που έχουν πολλές συνδέσεις. Οι τυχαίοι γράφοι, έχουν εφαρμογές σε πολλούς τομείς, όπως *social networks*, *neural networks*, κ.α.



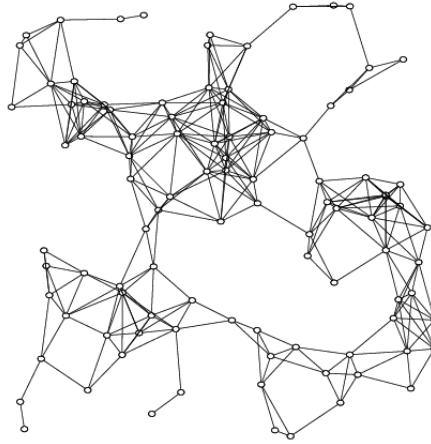
Σχήμα 2.6: Τοπολογία τυχαίου γραφήματος [72].



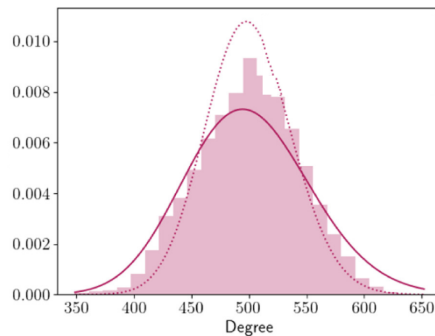
Σχήμα 2.7: Κατανομή βαθμού κόμβων σε ένα τυχαίο γράφημα [67].

2.3.2.3 Τυχαίο Γεωμετρικό Γράφημα - Random Geometric Graph

Οι τυχαίοι γεωμετρικοί γράφοι (Σχήμα 2.8), είναι ένα είδος τυχαίου γράφου που κατασκευάζεται με βάση γεωμετρικά κριτήρια. Συγκεκριμένα, οι κόμβοι του γράφου τοποθετούνται σε έναν χώρο και δύο κόμβοι συνδέονται με μια ακμή αν η απόσταση μεταξύ των κόμβων είναι μικρότερη από κάποια καθορισμένη (συντεταγμένες, ομοιότητες), π.χ. μικρότερη από μια συγκεκριμένη ακτίνα γειτονιάς, r . Δημοφιλείς κόμβοι (αυτοί με πολλές συνδέσεις), είναι ιδιαίτερα πιθανό να συνδέονται με άλλους δημοφιλείς κόμβους. Με το τρόπο αυτό, δημιουργείται ένα δίκτυο με χαρακτηριστικά τόσο τυχαία όσο και γεωμετρικά. Η κατανομή ενός τυχαίου γεωμετρικού δικτύου είναι ακολουθεί την κατανομή των τυχαίων γραφημάτων, απλώς φαίνεται να δημιουργεί μια πιο ομαλή καμπύλη (Σχήμα 2.9). Παραδείγματα τέτοιων δικτύων, περιλαμβάνουν *wireless mobile networks*, *geographic information systems*, κ.α. [24].



Σχήμα 2.8: Τοπολογία τυχαίου γεωμετρικού γραφήματος [25].



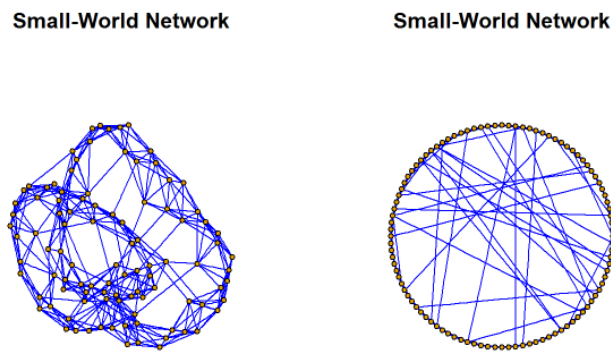
Σχήμα 2.9: Κατανομή βαθμού κόμβων σε ένα τυχαίο γεωμετρικό γράφημα σε σύγκριση με ένα τυχαίο γράφημα [63].

2.3.2.4 Γράφημα Μικρού-Κόσμου - Small-World Graph

Τα δίκτυα του μικρού κόσμου, χρησιμοποιούνται και μελετώνται σε πολλούς κλάδους, συμπεριλαμβανομένων των κοινωνικών και φυσικών επιστημών. Αυτά τα δίκτυα, διαθέτουν μια εντυπωσιακή ιδιότητα, το λεγόμενο φαινόμενο του μικρού κόσμου, που επίσης συχνά αναφέρεται ως "έξι βαθμοί διαχωρισμού" (μεταξύ δύο ατόμων στις Ηνωμένες Πολιτείες) [29]. Σε αυτό το είδος δικτύου, οι κόμβοι έχουν τάση να συνδέονται τόσο τοπικά όσο και με κάποιους απομακρυσμένους κόμβους, οι περισσότεροι δεν είναι μεταξύ τους γείτονες αλλά μπορούν να προσεγγιστούν από οποιονδήποτε άλλο κόμβο, σε έναν σύντομο αριθμό βημάτων (*affiliation networks*, *επιδημιολογία*, κ.α). Σε ένα παράδειγμα κοινωνικών δικτύων, η υψηλή ομαδοποίηση υποδηλώνει τη μεγάλη πιθανότητα δύο φίλοι ενός ατόμου, να είναι και φίλοι μεταξύ τους. Οι

χαμηλές αποστάσεις, από την άλλη πλευρά, σημαίνουν ότι υπάρχει μια σύντομη αλυσίδα κοινωνικών συνδέσεων μεταξύ οποιονδήποτε δύο ανθρώπων.

Τα γραφήματα αυτά, τείνουν να περιέχουν κλίκες, δηλαδή υποδίκτυα που έχουν συνδέσεις μεταξύ, σχεδόν, οποιονδήποτε δύο κόμβων εντός των κλικών αυτών (Σχήμα 2.10). Αυτό, προκύπτει από την ιδιότητα του υψηλού συντελεστή ομαδοποίησης (clustering coefficient), όπως αναφέρθηκε και παραπάνω. Ακόμη, τα περισσότερα ζεύγη κόμβων θα συνδέονται με τουλάχιστον ένα σύντομο μονοπάτι. Το γεγονός αυτό, παρατηρείται από την ιδιότητα ότι το μέσο μήκος μονοπατιού είναι μικρό. Συνήθως, υπάρχει μεγάλος αριθμός κόμβων με μεγάλο αριθμό συνδέσεων. Αυτοί οι κόμβοι, χρησιμεύουν ως κοινές συνδέσεις που μεσολαβούν στα σύντομα μονοπάτια διαδρομών μεταξύ άκρων.



Σχήμα 2.10: Τοπολογία γραφήματος μικρού κόσμου [72].

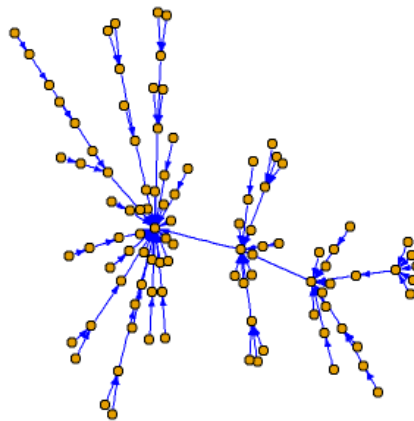
2.3.2.5 Γράφημα Ελεύθερης Κλίμακας - Scale-Free Graph

Τα γραφήματα ελεύθερης κλίμακας, παρουσιάζουν ένα είδος δικτύου, όπου οι κόμβοι δεν έχουν ίσο βαθμό συνδέσεων, αλλά ακολουθούν μια δυναμική κατανομή. Αυτό σημαίνει, ότι υπάρχουν λίγοι κόμβοι με πολύ υψηλό βαθμό συνδέσεων, ενώ οι περισσότεροι έχουν συγκριτικά χαμηλό (Σχήμα 2.12). Η έννοια της ελεύθερης κλίμακας, προέρχεται από το γεγονός ότι η κατανομή των συνδέσεων διατηρείται ανεπηρέαστη κατά μεγάλο βαθμό από την κλίμακα του δικτύου. Στα δίκτυα αυτά περιλαμβάνονται λίγοι κεντρικοί κόμβοι (hubs) (Σχήμα 2.11), οι οποίοι έχουν πολλές συνδέσεις, ενώ η πλειονότητα των κόμβων έχει μικρότερο αριθμό συνδέσεων. Είναι ο πιο συνήθης τύπος δικτύου στη φύση (*internet, road maps, citation networks, κ.α.*).

Ένα ακόμη σημαντικό χαρακτηριστικό των δικτύων χωρίς κλίμακα είναι η κατανομή του συντε-

λεστή ομαδοποίησης, ο οποίος μειώνεται καθώς αυξάνεται ο βαθμός του κόμβου. Αυτή η κατανομή, είναι επίσης δυναμική. Αυτό σημαίνει, ότι οι κόμβοι χαμηλού βαθμού ανήκουν σε πολύ πυκνές ομάδες. Για παράδειγμα, σε ένα κοινωνικό δίκτυο στο οποίο οι κόμβοι είναι άνθρωποι και οι σύνδεσμοι είναι σχέσεις γνωριμίας, διαπιστώνεται ότι οι άνθρωποι τείνουν να σχηματίζουν κοινότητες, δηλαδή μικρές ομάδες στις οποίες όλοι γνωρίζονται μεταξύ τους. Επιπλέον, τα μέλη μιας τέτοιας κοινότητας έχουν λίγες σχέσεις γνωριμίας με άτομα εκτός αυτής. Μερικοί άνθρωποι, ωστόσο, συνδέονται με μεγάλο αριθμό κοινοτήτων (π.χ. διασημότητες, πολιτικοί). Αυτοί οι άνθρωποι μπορεί να θεωρηθούν οι κεντρικοί κόμβοι (hubs).

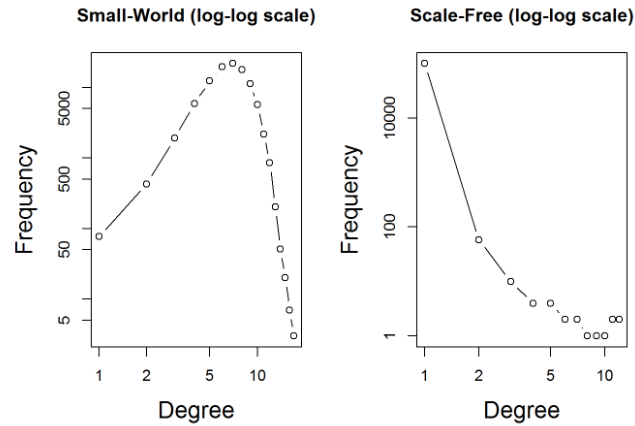
Scale-Free Network



Σχήμα 2.11: Τοπολογία γραφήματος ελεύθερης κλίμακας [72].

2.3.3 κεντρικότητες δικτύων

Η κεντρικότητα, είναι μια γενική μετρική του τρόπου με τον οποίο η θέση ενός κόμβου, διαδραματίζει σημαντικό ρόλο στη συνολική δομή του κοινωνικού δικτύου και μπορεί να υπολογιστεί χρησιμοποιώντας διάφορες μετρήσεις. Οι πιο ευρέως χρησιμοποιούμενες, είναι η κεντρικότητα βαθμού (*degree centrality*), η ενδιαμεσική κεντρικότητα (*betweenness centrality*), η κεντρικότητα εγγύτητας (*closeness centrality*) και η κεντρικότητα ιδιοδιανύσματος (*eigenvector centrality*). Οι τρεις πρώτες, προτάθηκαν από τον Freeman [31] και σχεδιάστηκαν μόνο για τα δίκτυα που δεν έχουν βάρη συνδέσεων. Οι Brin και Page [16] κατέληξαν σε επεκτάσεις για δίκτυα με βάρη. Η τέταρτη μετρική (αυτή της κεντρικότητας ιδιοδιανύσματος) προτάθηκε



Σχήμα 2.12: Κατανομή βαθμού κόμβων σε γραφήματα μικρού κόσμου και γραφήματα χωρίς κλίμακα [72].

αργότερα από τον Bonacich [14] και έχει τα θεμέλιά της στη θεωρία φασματικών γραφημάτων. Έγινε ιδιαίτερα δημοφιλής αφού χρησιμοποιήθηκε ως βάση του γνωστού αλγόριθμου *PageRank*² της Google. Η υψηλότερη κεντρικότητα, συνδέεται με ισχυρούς κόμβους στο δίκτυο, καθώς η κεντρική θέση προσφέρει πολλά πλεονεκτήματα, όπως ευκολότερη και ταχύτερη πρόσβαση σε άλλους κόμβους του δικτύου (χρήσιμο για πρόσβαση σε πόρους όπως πληροφορίες) και δυνατότητα άσκησης επιρροής πάνω από τη ροή μεταξύ άλλων κόμβων [31]. Οι τοπικές μετρήσεις, σε επίπεδο κόμβων (π.χ. κεντρικότητες) ίσως πρέπει να κανονικοποιηθούν προκειμένου να πραγματοποιηθούν συγκρίσεις δικτύων με διαφορετικές τάξεις και μεγέθη.

2.3.3.1 κεντρικότητα Βαθμού - Degree Centrality

Η κεντρικότητα βαθμού ενός κόμβου v , συνήθως συμβολίζεται ως $C_d(v)$, και είναι ένα μέτρο άμεσης γειτονίας και συμμετοχής του κόμβου στο δίκτυο. Υπολογίζεται, ως ο αριθμός των ακμών που προσπίπτουν σε έναν κόμβο ή, πιο απλά, ο αριθμός των γειτόνων του κόμβου v . Η κεντρικότητα βαθμού $C_d(v)$ ενός κόμβου v υπολογίζεται ως ο λόγος του αριθμού των ακμών που συνδέουν τον κόμβο με τον συνολικό αριθμό κόμβων στο γράφο:

$$C_d(i) = \frac{k_i}{n},$$

όπου:

²<https://en.wikipedia.org/wiki/PageRank>

- $C_d(i)$, είναι η κεντρικότητα βαθμού του κόμβου i .
- k_i , είναι ο αριθμός των ακμών που συνδέουν τον κόμβο i .
- n , είναι ο συνολικός αριθμός των κόμβων στο γράφο.

2.3.3.2 Ενδιαμεσική κεντρικότητα - Betweenness Centrality

Η ενδιαμεσική κεντρικότητα $C_b(v)$, μετρά το ποσοστό των συντομότερων διαδρομών (shortest paths) στο δίκτυο, που διέρχονται από έναν συγκεκριμένο κόμβο. Οι κόμβοι με υψηλό βαθμό ενδιαμεσικής κεντρικότητας, φαίνονται να καταλαμβάνουν κρίσιμους ρόλους στη δομή του δικτύου. Οι κόμβοι αυτοί, έχουν συνήθως μια θέση, στο δίκτυο, που τους επιτρέπει να λειτουργούν ως διεπαφή μεταξύ στενά συνδεδεμένων ομάδων, αποτελώντας κομβικά σημεία για την σύνδεση μεταξύ διαφορετικών περιοχών του δικτύου. Η ενδιαμεσική κεντρικότητα ενός κόμβου, μπορεί να υπολογιστεί με τον παρακάτω τρόπο:

$$C_b(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}},$$

όπου:

- v , είναι ο κόμβος για τον οποίο υπολογίζεται η ενδιαμεσική κεντρικότητα.
- σ_{st} , είναι ο συνολικός αριθμός συντομότερων μονοπατιών από τον κόμβο s στον κόμβο t .
- $\sigma_{st}(v)$, είναι ο αριθμός των συντομότερων μονοπατιών από τον κόμβο s στον κόμβο t που περνούν από τον κόμβο v .

2.3.3.3 κεντρικότητα Εγγύτητας - Closeness Centrality

Η κεντρικότητα εγγύτητας, είναι ένα μέτρο της συνολικής θέσης ενός κόμβου στο δίκτυο. Τυπικά, είναι το μέσο μήκος όλων των συντομότερων μονοπατιών από έναν κόμβο προς όλους τους άλλους κόμβους του δικτύου. Λόγω του ορισμού του, συνήθως αυτό το μέτρο υπολογίζεται μόνο για κόμβους εντός του μεγαλύτερου στοιχείου του δικτύου. Στο πλαίσιο των κοινωνικών δικτύων, η εγγύτητα είναι ένα μέτρο προσβασιμότητας που μετρά πόσο γρήγορα ένας συγκεκριμένος κόμβος μπορεί να προσεγγίσει όλους τους υπόλοιπους στο δίκτυο. Η κεντρικότητα εγγύτητας $C_c(v)$ ενός κόμβου v , υπολογίζεται ως η αντίστροφη του μέσου

μήκους των συντομότερων μονοπατιών από τον κόμβο v προς όλους τους άλλους κόμβους στο γράφο:

$$C_c(v) = \frac{1}{\frac{1}{n-1} \sum_{u \neq v} p(v, u)},$$

όπου:

- $C_c(v)$, είναι η κεντρικότητα εγγύτητας του κόμβου v .
- n , είναι ο συνολικός αριθμός των κόμβων στο γράφο.
- $p(v, u)$, είναι το μήκος του συντομότερου μονοπατιού από τον κόμβο v προς τον κόμβο u .

2.3.3.4 κεντρικότητα Ιδιοδιανύσματος - Eigenvector Centrality

Η κεντρικότητα ιδιοδιανύσματος, είναι ένα μέτρο κεντρικότητας σε ένα γράφο, το οποίο αξιολογεί τη σημαντικότητα ενός κόμβου, λαμβάνοντας υπόψη τη σύνδεσή του με άλλους κόμβους που έχουν επίσης υψηλή σημασία στο δίκτυο. Βασίζεται στην έννοια των ιδιοδιανυσμάτων και των ιδιοτιμών ενός γράφου. Κάθε κόμβος, λαμβάνει υπόψη του την κεντρικότητα των γειτονικών κόμβων, αλλά με βάρη που εξαρτώνται από την κεντρικότητα τους. Υπολογίζεται, με τη χρήση του μεγαλύτερου ιδιοδιανύσματος του πίνακα γειτνίασης του γράφου. Η εκτίμηση αυτή της κεντρικότητας, βασίζεται στην προσέγγιση του διανύσματος κεντρικότητας, που προκύπτει από το μεγαλύτερο ιδιοδιανυσματικό ζεύγος. Συνολικά, παρέχει μια μετρική για τον τρόπο με τον οποίο η επιρροή ενός κόμβου, διαδίδεται στο γενικό δίκτυο, λαμβάνοντας υπόψη τη συνδεσιμότητά του με άλλους σημαντικούς κόμβους. Η κεντρικότητα ιδιοδιανύσματος $C_e(v)$ ενός κόμβου v υπολογίζεται με τη χρήση της εξίσωσης:

$$C_e(v) = \frac{1}{\lambda} \sum_u A_{uv} C_e(u),$$

όπου:

- $C_e(v)$, είναι η κεντρικότητα ιδιοδιανύσματος του κόμβου v .
- λ , είναι η μέγιστη ιδιοτιμή του πίνακα γειτνίασης A του γράφου.
- A_{uv} , είναι το στοιχείο στη θέση (u, v) του πίνακα γειτνίασης.
- Το άθροισμα εκτελείται σε όλους τους κόμβους u στο γράφο.

2.3.4 συντελεστής Ομαδοποίησης - Clustering Coefficient

Τα κοινωνικά δίκτυα, τις περισσότερες φορές είναι μεταβατικά, γεγονός που σημαίνει ότι, για παράδειγμα, οι φίλοι ενός συγκεκριμένου ατόμου, είναι επίσης πιθανό να είναι φίλοι μεταξύ τους. Αυτή η ιδιότητα της μεταβατικότητας, ποσοτικοποιείται με έναν συντελεστή ομαδοποίησης, που μπορεί να είναι είτε καθολικός, δηλαδή να υπολογίζεται για ολόκληρο το δίκτυο, είτε τοπικός, δηλαδή να υπολογίζεται για κάθε κόμβο. Ο συντελεστής ομαδοποίησης, μετρά τον βαθμό που οι γείτονες ενός κόμβου σχηματίζουν συνδέσεις μεταξύ τους. Αυτό το μέτρο, δίνει μια ιδέα για το πόσο "κοινωνικά συνδεδεμένος" είναι ένας κόμβος στο δίκτυο. Για έναν κόμβο, μπορεί να υπολογιστεί, ως ο λόγος του πραγματικού αριθμού συνδέσεων μεταξύ των γειτόνων του κόμβου και του μέγιστου δυνατού αριθμού συνδέσεων μεταξύ αυτών των γειτόνων. Ο συντελεστής ομαδοποίησης για έναν κόμβο v υπολογίζεται ως:

$$CC(v) = \frac{\text{links between existing neighbors}(v)}{\text{links between neighbors (all possible)}},$$

όπου:

- $CC(v)$, είναι ο συντελεστής ομαδοποίησης για τον κόμβο v .
- $\text{links between existing neighbors}(v)$, είναι ο αριθμός των συνδέσεων που περιλαμβάνουν τον κόμβο v .
- $\text{links between neighbors (all possible)}$, είναι οι δυνατές συνδέσεις μεταξύ των γειτόνων τους (όλες οι δυνατές).

Ο συντελεστής ομαδοποίησης, κυμαίνεται μεταξύ των τιμών 0 και 1, με υψηλότερες τιμές υποδηλώνοντας υψηλό βαθμό τοπικής συνδεσιμότητας.

2.3.5 Ακτίνα και Διάμετρος - Radius and Diameter

Η εκκεντρότητα ενός κόμβου σε ένα γράφημα, ορίζεται ως το μήκος μιας μακρύτερης συντομότερης διαδρομής, που ξεκινά από τον κόμβο αυτό. Η κατανομή της εκκεντρότητας σε όλους τους κόμβους, είναι μια σχετική περιγραφική ιδιότητα του γραφήματος και οι ακραίες τιμές, επιτρέπουν την εξαγωγή μέτρων, όπως η ακτίνα και η διάμετρος ενός γραφήματος. Η παραδοσιακή μέθοδος, υπολογίζει την εκκεντρότητα κάθε κόμβου εκτελώντας τον αλγόριθμο All Pairs Shortest Path (APSP) [66]. Η διάμετρος D , δίνεται από τη μέγιστη εκκεντρότητα του συνόλου των κορυφών στο δίκτυο και, αναλόγως, η ακτίνα R μπορεί να οριστεί ως η ελάχιστη εκκεντρότητα του συνόλου των κορυφών. Τα πιο αραιά δίκτυα, έχουν γενικά μεγαλύτερη

διάμετρο από τους πλήρεις πίνακες, λόγω της ύπαρξης λιγότερων μονοπατιών μεταξύ ζευγών κόμβων. Οι *Leskovec*, *Kleinberg* και *Faloutsos* [39] ανακάλυψαν ότι, για ορισμένους τύπους δικτύων του πραγματικού κόσμου, η πραγματική διάμετρος συρρικνώνεται με την πάροδο του χρόνου, σε αντίθεση με τη συμβατική ιδέα της αύξησης των διαμέτρων. Στο πλαίσιο του SNA, αυτή η μέτρηση δίνει μια προσέγγιση σχετικά με την εγγύτητα των ζευγαριών κόμβων στο δίκτυο, υποδεικνύοντας πόσο μακριά είναι δύο κόμβοι, στη χειρότερη των περιπτώσεων. Με πιο απλά λόγια, η διάμετρος ενός γράφου είναι η μεγαλύτερη απόσταση μεταξύ δύο κόμβων στο γράφο και η ακτίνα είναι η ελάχιστη απόσταση από έναν κόμβο προς όλους τους υπόλοιπους κόμβους. Η διάμετρος, υπολογίζεται από τον τύπο:

$$D = \max_{u,v \in V} d(u,v),$$

- όπου V , είναι το σύνολο των κόμβων και $d(u,v)$ είναι η απόσταση μεταξύ των κόμβων u και v .

Η ακτίνα υπολογίζεται από τον τύπο:

$$R = \min_{v \in V} \max_{u \in V} d(u,v),$$

- όπου V , είναι το σύνολο των κόμβων και $d(u,v)$ είναι η απόσταση μεταξύ των κόμβων u και v .

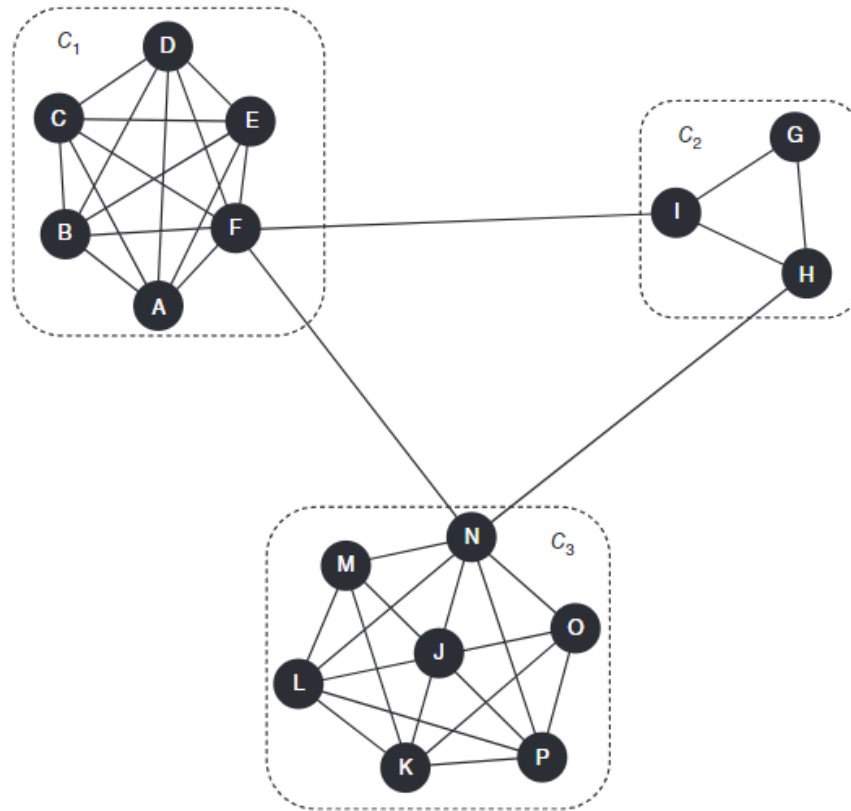
2.3.6 Εντοπισμός Κοινοτήτων

Ένα ακόμη από τα χαρακτηριστικά των κοινωνικών δικτύων, είναι ότι τείνουν να παρουσιάζουν δομές κοινοτήτων. Αυτή η ιδιότητα, συνήθως προκύπτει ως συνέπεια της συνολικής και τοπικής ετερογένειας της κατανομής των ακμών (συνδέσεων) σε ένα γράφημα. Έτσι, συχνά βρίσκουμε υψηλές συγκεντρώσεις ακμών σε ορισμένες περιοχές του γραφήματος, που ονομάζονται κοινότητες, και χαμηλή συγκέντρωση ακμών μεταξύ αυτών των περιοχών. Οι κοινότητες, μπορούν να οριστούν ευθέως ως ομάδες κόμβων που παρουσιάζουν κοινά χαρακτηριστικά. Ένας πληρέστερος ορισμός, βασίζεται στην έννοια της πυκνότητας: οι κοινότητες μπορούν να γίνουν κατανοητές ως πυκνά συνδεδεμένες ομάδες κορυφών (κόμβων) στο δίκτυο, με πιο αραιές συνδέσεις μεταξύ των ομάδων αυτών. Οι συνδέσεις μπορούν να είναι κατευθυνόμενες (*Twitter*) και μη κατευθυνόμενες (*Facebook*).

Σύμφωνα με τους *Newman* και *Girvan* [52], υπάρχουν δύο βασικές γραμμές έρευνας για την ανακάλυψη κοινοτήτων στα δεδομένα δικτύου. Η πρώτη, προέκυψε αρχικά στον τομέα των

Επιστημών υπολογιστών λόγω της ανάγκης να βρεθεί ο καλύτερος τρόπος για την κατανομή εργασιών στους επεξεργαστές, έτσι ώστε να ελαχιστοποιηθούν οι μεταξύ τους επικοινωνίες. Αυτή η εργασία βελτιστοποίησης δικτύου, αποσκοπούσε στην ενίσχυση του υπολογισμού, σε ένα παράλληλο υπολογιστικό περιβάλλον. Ενώ η δεύτερη, προέκυψε από την ανακάλυψη κοινοτήτων εντός της κοινωνίας, προκειμένου να απλοποιηθεί η ανάλυση των κοινωνικών φαινομένων μέσα από τη διάταξη των ανθρώπων, σύμφωνα με τις ομοιότητές τους. Η βασική διαδικασία πίσω από τους αλγόριθμους ανίχνευσης κοινοτήτων, βασίζεται στη διαίρεση του αρχικού γραφήματος, σε ένα σύνολο ασύνδετων υπογραφών, μέσω της βελτιστοποίησης μιας δεδομένης αντικειμενικής συνάρτησης (π.χ. *modularity*). Στόχος και των δύο προσεγγίσεων, είναι η ανακάλυψη ομάδων σχετικών κορυφών (κόμβων) στο δίκτυο και της αντίστοιχης ιεραρχικής οργάνωσης, με βάση μόνο τις πληροφορίες που παρέχονται από την τοπολογία του δικτύου. Αυτό συνήθως γίνεται με την επαναληπτική αφαίρεση των ακμών μεταξύ ομάδων κορυφών, όπως προτείνουν οι *Girvan* και *Newman* [32].

Για την καλύτερη κατανόηση των εισαγόμενων εννοιών, το Σχήμα 2.13 απεικονίζει ένα απλό δίκτυο που περιλαμβάνει τρεις κοινότητες, που ονομάζονται $C1$, $C2$ και $C3$. Σε αυτή την εικόνα, αντιπροσωπεύεται μια ιδανική κατάσταση, αφού κάθε κοινότητα είναι η ίδια της ένα πλήρες γράφημα ή μια κλίκα, διαφορετικού μεγέθους ($C1 = K_6$, $C2 = K_3$ και $C3 = K_7$). Επίσης, η πυκνότητα των δεσμών μεταξύ των κοινοτήτων είναι πολύ χαμηλή. Οι λίγοι δεσμοί που υπάρχουν μεταξύ των κοινοτήτων ονομάζονται γέφυρες, αφού είναι οι μόνες διαθέσιμες συνδέσεις μεταξύ διαφορετικών τμημάτων του δικτύου. Στην πραγματική ζωή, υπάρχουν αρκετά παραδείγματα τέτοιων ομάδων. Η κοινωνία είναι ένα πλούσιο περιβάλλον για την εύρεση και κατανόηση των κοινοτήτων, αφού οι άνθρωποι έχουν τη φυσική τάση να σχηματίζουν ομάδες. Αυτές οι ομάδες, μπορεί να είναι οικογένειες, φίλοι, εργατικές ή/και θρησκευτικές ομάδες, πόλεις, έθνη και ούτω καθεξής. Εάν εξετάσουμε επίσης ομάδες που σχηματίζονται από εταιρείες ή από πελάτες ενός δεδομένου προϊόντος, μπορούμε να εντοπίσουμε κοινότητες που σχετίζονται με τους τομείς της οικονομίας και των επιχειρήσεων. Η βιολογία, είναι ένα ακόμη παράδειγμα, όπου οι μέθοδοι για την εύρεση κοινοτήτων είναι χρήσιμες, ειδικά στο πλαίσιο των μεταβολικών δικτύων. Όπως, στα δίκτυα αλληλεπίδρασης πρωτεΐνης, μπορούμε να βρούμε ομάδες πρωτεϊνών με παρόμοιες λειτουργίες μέσα στο κύτταρο [2] [3] [33]. Μπορούμε, επίσης, να βρούμε εικονικές διαδικτυακές κοινότητες στο Διαδίκτυο ή ομάδες ιστοσελίδων, που σχετίζονται με θέματα, οι οποίες μπορεί να είναι χρήσιμες για την ανάπτυξη αυτόματων και αποτελεσματικών συστημάτων προτάσεων (*recommendation systems*). Η σημασία της μελέτης αυτών των κοινοτήτων, είναι σημαντική σε τομείς όπως το *SNA*.



Σχήμα 2.13: Δομές Κοινοτήτων [65].

2.3.6.1 Μέθοδος Girvan - Newman

Μεταξύ των πιο δημοφιλών αλγορίθμων, για την επίλυση προβλημάτων ανίχνευσης κοινότητας είναι αυτός που επινοήθηκε από τους *Girvan* και *Newman* [32] και είναι γνωστός ως αλγόριθμος *Girvan–Newman*. Ο αλγόριθμος *Girvan–Newman*, είναι μια ιεραρχική τεχνική που αποδομεί το αρχικό πλήρες δίκτυο σε προοδευτικά μικρότερες συνεκτικές συνιστώσες, μέχρι το σημείο όπου δεν υπάρχουν ακμές για αφαίρεση και κάθε κόμβος αντιπροσωπεύει τον εαυτό του ως μια κοινότητα.

Έχοντας κατά νου ότι οι κοινότητες είναι συνεκτικές ομάδες κόμβων, με πιο αραιές συνδέσεις μεταξύ τους, το κριτήριο για την αφαίρεση των ακμών, που προτείνουν, είναι το μέτρο της edge-betweenness centrality. Ο λόγος πίσω από αυτήν την επιλογή, σχετίζεται με το γεγονός ότι αυτό το μέτρο κεντρικότητας είναι σε θέση να προσδιορίσει ακμές που εντοπίζονται σε μεγάλο αριθμό συντομότερων διαδρομών μεταξύ κόμβων και, επομένως, πιστεύεται ότι συνδέουν διαφορετικές μη επικαλυπτόμενες κοινότητες. Έτσι, η κύρια ιδέα αυτού του αλγορίθμου είναι

ότι αν βρεθούν και αφαιρεθούν γέφυρες, απομονώνονται οι υπάρχουσες κοινότητες σε ένα δίκτυο. Η είσοδος του αλγορίθμου, είναι ένα πλήρες γράφημα και η έξοδος είναι μια ιεραρχική δομή, όπως μια δομή δέντρου, όπου οι κοινότητες σε οποιοδήποτε επίπεδο αντιστοιχούν σε μια οριζόντια τομή μέσω αυτού του ιεραρχικού δέντρου. Τα βήματα του αλγορίθμου μπορούν να εκφραστούν ως εξής:

1. υπολογισμός της ενδιαμεσότητας ακμής όλων των συνδέσεων του δικτύου.
2. Αφαίρεση της ακμής με την υψηλότερη τιμή ενδιαμεσότητας. Αυτό το βήμα, μπορεί να προκαλέσει το διαχωρισμό του δικτύου σε ξεχωριστά αποσυνδεδεμένα μέρη, τα οποία αποτελούν το πρώτο επίπεδο περιοχών στη διαμέριση του γραφήματος.
3. Επανάληψη του προηγούμενου βήματος, μέχρι να μην υπάρχουν ακμές για αφαίρεση στο γράφημα. Λόγω της δημοτικότητάς του, σχεδόν όλες οι τυπικές βιβλιοθήκες λογισμικού έχουν εφαρμόσει αυτόν τον αλγόριθμο.

2.3.6.2 Μέθοδος Modularity Maximization

Μια ακόμη συχνά χρησιμοποιούμενη και πολύ δημοφιλής μέθοδος για την ανίχνευση κοινοτήτων σε δίκτυα, είναι το *modularity maximization*. Σύμφωνα με τον Newman [51], το *modularity* είναι, μέχρι μια πολλαπλασιαστική σταθερά, ο αριθμός των συνδέσεων που εμπίπτουν σε ομάδες μείον τον αναμενόμενο αριθμό, σε ένα ισοδύναμο δίκτυο με ακμές τοποθετημένες τυχαία. Με βάση αυτόν τον ορισμό, μπορούμε να συμπεράνουμε ότι το *modularity*, είναι ένα μέτρο, που λαμβάνει ρητά υπόψη την ετερογένεια των ακμών. Η βασική ιδέα, είναι ότι ένα δίκτυο εμφανίζει ουσιαστική δομή κοινότητας, εάν ο αριθμός των ακμών μεταξύ των κοινοτήτων είναι μικρότερος από το αναμενόμενο, με βάση τυχαία επιλογή. Με την υπόθεση αυτή, όσο υψηλότερη είναι η τιμή του, τόσο καλύτερα επιτυγχάνεται και η κατάτμηση, αυτό σημαίνει ότι οι κοινότητες που βρέθηκαν είναι εσωτερικά πυκνά συνδεδεμένες και εξωτερικά αραιά συνδεδεμένες, καθώς υπάρχουν περισσότερες ακμές που εμπίπτουν στις ομάδες, από ό,τι θα περίμενε κανείς τυχαία. Το *modularity maximization*, είναι ένα πρόβλημα βελτιστοποίησης που αφορά την εύρεση της βέλτιστης κατανομής των κόμβων του γράφου σε κοινότητες. Υπολογίζεται από τον τύπο:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j),$$

όπου:

- Q , είναι το *modularity* ενός διαμοιρασμού.
- A_{ij} , είναι το στοιχείο στη θέση (i, j) του πίνακα γειτνίασης.
- k_i , είναι ο βαθμός του κόμβου i .
- m , είναι το συνολικό πλήθος των ακμών στο γράφο.
- $\delta(c_i, c_j)$, είναι η συνάρτηση Kronecker που ισούται με 1 αν οι κόμβοι i και j ανήκουν στην ίδια κοινότητα και 0 διαφορετικά.

Από τον τύπο, μπορούμε να συμπεράνουμε ότι $Q \in [-1, 1]$ είναι είτε αρνητικό είτε θετικό. Εάν είναι θετικό, τότε υπάρχει η δυνατότητα εύρεσης κοινοτικών δομών στο δίκτυο. Εάν το Q δεν είναι μόνο θετικό, αλλά και μεγάλο, τότε το αντίστοιχο κομμάτι του γραφήματος μπορεί να αντικατοπτρίζει την πραγματική δομή της κοινότητας. Σύμφωνα με τους *Clauset, Newman* και *Moore* [22], στην πράξη διαπιστώθηκε ότι περίπου μια τιμή 0,3 είναι ένας καλός δείκτης της ύπαρξης ουσιαστικών κοινοτήτων. Ακολουθώντας αυτό το σκεπτικό και γνωρίζοντας ότι όσο υψηλότερο είναι το *modularity*, τόσο καλύτερη είναι η εκάστοτε διαίρεση δικτύου, αυτό που θα μεγιστοποιούσε αυτό το μέτρο, είναι ο υπολογισμός του για κάθε πιθανή διαίρεση του δικτύου και επιλέγοντας τη διαίρεση που επιστρέφει την υψηλότερη τιμή. Αυτή η απλή ιδέα, οδήγησε σε μια νέα κατηγορία μεθόδων, των οποίων τα θεμέλια τίθενται στο *modularity maximization*. Αν και αυτή η προσέγγιση είναι αρκετά ελκυστική, η εξαντλητική αναζήτηση σε όλες τις πιθανές διαιρέσεις είναι συνήθως δυσεπίλυτη και χρονοβόρα. Αυτή η ανεπιθύμητη επίδραση της υπολογιστικής αναποτελεσματικότητας, έχει επιλυθεί με την προσαρμογή ορισμένων ευρετικών μεθόδων σε αυτό το συγκεκριμένο πρόβλημα βελτιστοποίησης. Ακολουθώντας αυτή τη στρατηγική, μπορεί κανείς να επιτύχει μια αρκετά καλή προσέγγιση του καθολικού βέλτιστου (σε αυτή την περίπτωση, τη μέγιστη τιμή του *modularity*) σε αποδεκτό χρόνο.

2.3.6.3 Μέθοδος Spectral Clustering

Το *spectral clustering*, είναι μια ακόμη τεχνική στον χώρο της ανίχνευσης κοινοτήτων σε δίκτυα. Αντί να εξετάζει απευθείας τις συνδέσεις μεταξύ των κόμβων, χρησιμοποιεί τη φασματική ανάλυση των δεδομένων του δικτύου. Αρχικά, δημιουργείται ένας πίνακας ομοιότητας που χρησιμοποιείται και ως είσοδος, όπου οι κόμβοι αντιπροσωπεύουν τα δείγματα και οι ακμές αντιπροσωπεύουν τις ομοιότητες μεταξύ τους. Στη συνέχεια, εφαρμόζεται φασματική ανάλυση στον πίνακα ομοιότητας για την εξαγωγή των ιδιοδιανυσμάτων. Αυτά τα ιδιοδιανύσματα, αποτελούν τις νέες αναπαραστάσεις των κόμβων. Έπειτα, οι κόμβοι κατη-

γοριοποιούνται σε ομάδες με βάση τις τιμές των κυριότερων ιδιοδιανυσμάτων. Η απόσταση μεταξύ τους μετρείται στο νέο φασματικό χώρο. Τέλος, οι ομάδες κόμβων που προκύπτουν αντιπροσωπεύουν τις κοινότητες στο δίκτυο. Οι κόμβοι που ανήκουν στην ίδια ομάδα θεωρούνται μέλη μιας κοινότητας. Η μέθοδος αυτή, παρέχει το πλεονέκτημα να είναι ευαίσθητη και σε μη γραφικές δομές και μπορεί να ανακαλύψει πιο σύνθετες κοινότητες σε σχέση με τις παραδοσιακές μεθόδους. Επιπλέον, η χρήση της βοηθάει στη μείωση του θορύβου και των ακραίων τιμών στην ανίχνευση κοινοτήτων.

Η διαδικασία spectral clustering, πραγματοποιείται ως εξής:

1. δημιουργία Γράφου Ομοιότητας: Αρχικά, δημιουργείται ο πίνακας ομοιότητας W όπου W_{ij} αντιπροσωπεύει την ομοιότητα μεταξύ των δειγμάτων i και j .
2. υπολογισμός Πίνακα Κανονικοποίησης: δημιουργείται ο πίνακας κανονικοποίησης D , ο οποίος είναι διαγώνιος και $D_{ii} = \sum_j W_{ij}$.
3. υπολογισμός Πίνακα Laplacian: Ο πίνακας Laplacian υπολογίζεται ως $L = D - W$.
4. υπολογισμός Κυρίων Συνιστώσεων: Εφαρμόζεται φασματική ανάλυση στον πίνακα Laplacian για να βρεθούν οι κυρίως συνιστώσες (eigenvectors) του L .
5. Κατηγοριοποίηση με K-Means: Χρησιμοποιώντας τις κυρίως συνιστώσες, εκτελείται η μέθοδος K-Means για να κατηγοριοποιήσει τα δείγματα σε συστάδες.

Ο αλγόριθμος K-Means, θεωρείται ένας από τους πιο ισχυρούς και δημοφιλείς αλγόριθμους εξόρυξης δεδομένων στην ερευνητική κοινότητα. Εξαρτάται από την τιμή του k , το οποίο πρέπει πάντα να προσδιορίζεται προκειμένου να πραγματοποιηθεί οποιαδήποτε ανάλυση ομαδοποίησης. Η ομαδοποίηση με διαφορετικές τιμές k , θα παράγει διαφορετικά αποτελέσματα. Κατά τη διαδικασία του αλγορίθμου, τα δείγματα ανατίθενται σε ομάδες (clusters) με βάση την απόστασή τους από τυχαία σημεία k . Στη συνέχεια, τα σημεία αυτά ανανεώνονται, υπολογίζοντας το μέσο όρο των δειγμάτων που έχουν ανατεθεί σε κάθε ομάδα. Η διαδικασία της εκχώρησης και ανανέωσης των σημείων, επαναλαμβάνεται μέχρι να σταθεροποιηθούν οι ομάδες [26]. Ο τελικός αλγόριθμος, συνοψίζεται από την παρακάτω εξίσωση:

$$\begin{aligned} &\text{Maximize} \quad \text{Tr}(B^T L B) \\ &\text{Subject to} \quad B^T B = I, \end{aligned}$$

όπου B , είναι ο πίνακας με τα eigenvectors που προέκυψαν από τη φασματική ανάλυση.

Στα παραπάνω υποκεφάλαια, παρουσιάστηκε μια συνοπτική επισκόπηση των βασικών μεθόδων

ανάλυσης δικτύων, με έμφαση στους στόχους τους και τις δυνατότητες εφαρμογής τους. Αναφέρθηκε επίσης, ένας αριθμός μετρικών *SNA*, εξετάζοντας τις εφαρμογές τους σε διάφορους τύπους δικτύων. Η ανάλυση δικτύων, βρίσκεται στο επίκεντρο του ενδιαφέροντος λόγω της αυξανόμενης δημοτικότητάς, ενώ παράλληλα αντιμετωπίζει προκλήσεις λόγω της αυξανόμενης πολυπλοκότητας των διαθέσιμων δεδομένων. Ειδικότερα, τα δυναμικά δίκτυα που εξελίσσονται σε πραγματικό χρόνο απαιτούν αλγορίθμους που είναι ταχείς, αυξητικοί και αποδοτικοί σε μνήμη για τον υπολογισμό μέτρων *SNA*. Το ενδιαφέρον των πρόσφατων ερευνών, επικεντρώνεται επίσης στη σχεδίαση αλγορίθμων που μπορούν να ανταποκριθούν σε γραφήματα μεγάλων διαστάσεων, μια πρόκληση που παραμένει ανοιχτή. Τελευταία, οι τάσεις περιλαμβάνουν βελτιώσεις στους τομείς της χρηστικότητας και επεκτασιμότητας, με εφαρμογές σε κοινωνικά δίκτυα, έξυπνες πόλεις, βιοπληροφορική, επιστήμες της γης, μάρκετινγκ και πολλούς άλλους τομείς. Επιπλέον, οι αυξημένες απαιτήσεις από αυτές τις εφαρμογές στον πραγματικό κόσμο προκαλούν εξελίξεις στις τοπολογίες δικτύων, όπως τα πολυεπίπεδα, τα ετερογενή, και τα εξελισσόμενα δίκτυα. Οι παραπάνω υποενότητες, βοηθούν στην κατανόηση των απλών εννοιών που αναφέρονται στην ανάλυση δικτύων.

2.4 Ανάλυση Αθλητικών δικτύων

Για τη διεξαγωγή μιας Ανάλυσης κοινωνικών δικτύων (*SNA*), οι ερευνητές προσδιορίζουν ένα δίκτυο, αποφασίζοντας ποιους κόμβους και ποιες συνδέσεις (δηλαδή δεδομένα δικτύου) θα αναλύσουν. Αν και η ανάλυση αθλητικών δικτύων, είναι σχετικά νέα προσέγγιση στην αθλητική έρευνα, οι βασικές ιδέες εφαρμόζονται σε μια σειρά από κλασικές μελέτες. Κοινωνικοί παράγοντες όπως άτομα, ομάδες, οργανώσεις ή γκρουπ αντιπροσωπεύουν κόμβους κατάλληλους για ανάλυση. Οι συνδέσεις μεταξύ κόμβων, μπορεί να είναι κατευθυνόμενες ή μη. Σε ορισμένες αναλύσεις το βάρος της σύνδεσης είναι σχετικό. Μπορεί να ποσοτικοποιηθεί και να σταθμιστεί με τα ανάλογα βάρη συνδέσεων. Δεδομένου ότι η αθλητική έρευνα είναι ένας σχετικά νέος επιστημονικός κλάδος, η υιοθέτηση ερευνητικών μεθόδων και θεωριών από άλλους κλάδους έχει αποδειχθεί χρήσιμη [71]. Στην έρευνα του ο *Zachary* [77] κατασκεύασε το *Zachary's karate club*, είναι ένα κοινωνικό δίκτυο ενός πανεπιστημιακού συλλόγου καράτε. Το δίκτυο έγινε δημοφιλές παράδειγμα κοινοτικής δομής στα δίκτυα μετά τη χρήση του από τους *Girvan* και *Mark Newman* το 2002. Αυτό το κοινωνικό δίκτυο μελετήθηκε από τον *Zachary*, για μια περίοδο τριών ετών από το 1970 έως το 1972. Αναπαριστά 34 μέλη μιας λέσχης καράτε, και τις σχέσεις μεταξύ ζευγαριών μελών που αλληλεπιδρούσαν εκτός του συλλόγου. Κατά τη διάρκεια της μελέτης, προέκυψε μια σύγκρουση μεταξύ του εκπαιδευτή

και του προέδρου φοιτητή του *club*, η οποία οδήγησε στη διάσπαση του συλλόγου στα δύο. Με τον τρόπο αυτό, σχηματίστηκαν δύο υποομάδες και η σύσταση του δικτύου άλλαξε. Η έρευνα αυτή, αποτελεί πρότυπο για τον τρόπο με τον οποίο τα κοινωνικά δίκτυα μπορούν να αναπαρασταθούν και να μελετηθούν, ενισχύοντας την κατανόηση των δομών και των αλληλεπιδράσεων σε ανάλογα περιβάλλοντα, συμβάλλοντας στην εξέλιξη της αθλητικής έρευνας και της *SNA* γενικότερα.

Στις αρχές της δεκαετίας του 1990, ο Nixon [53], χρησιμοποίησε τις ιδέες της ανάλυσης κοινωνικών δικτύων για να αναπτύξει ένα πλαίσιο για την κατανόηση του κινδύνου των αθλητών, που παίζουν ενώ πονούν ή έχουν τραυματιστεί. Ο Nixon εισήγαγε τον όρο αθλητικό δίκτυο, για να περιγράψει αθλητικές ομάδες, που χαρακτηρίζονται όχι μόνο από ένα σύστημα αλληλεπίδρασης μεταξύ μιας ομάδας ανθρώπων στον αθλητισμό, αλλά και από κοινό όραμα, τελετουργίες, κανόνες κ.λπ. Αυτή η μελέτη, παρέχει μια εξαιρετική εξήγηση, δηλαδή πώς η δομική ενσωμάτωση περιορίζει συμπεριφορά των αθλητών όταν αντιμετωπίζουν πόνο και τραυματισμούς. Ωστόσο, η ανάλυσή του παραμένει σε εννοιολογικό επίπεδο χωρίς ρητή αναφορά σε μεθόδους *SNA*. Σε μια επόμενη εργασία του, ο Nixon διεύρυνε την προσέγγισή του για να παρέχει μια επισκόπηση "της πιθανής αξίας μιας προσέγγισης δομικών κοινωνικών δικτύων" [54]. Αν και δεν εφάρμοσε πάλι καμία επίσημη μέθοδο *SNA*, το έργο του ξεχωρίζει ως η πρώτη σημαντική συνεισφορά, στην έρευνα των κοινωνικών δικτύων στον αθλητικό τομέα, βασιζόμενη σε μια θεωρητική προσέγγιση της οπτικής του κοινωνικού δικτύου [59].

Με βάση τον ορισμό του μεγάλου όγκου δεδομένων, μπορεί να καθοριστεί και η πληθώρα των αθλητικών δεδομένων. Εκατοντάδες εκατομμύρια αθλητικά δεδομένα δημιουργούνται κάθε μέρα από άτομα, σχολεία, διάφορες εκδηλώσεις, κοινότητες, κλπ. Σε αυτά, περιλαμβάνονται διάφορες οντότητες και σχέσεις, γεγονότα που τα καθιστά ανοιχτά σε κάθε είδους πρόκληση. Τα δεδομένα αυτά, συνήθως περιλαμβάνουν: δεδομένα σχετικά με την φυσική κατάσταση των ανθρώπων που περιγράφουν, συμπεριφορές της σωματικής άσκησης, προσωπικά στοιχεία - δεδομένα και αποτελέσματα διαγωνισμών. Οι έρευνες που έχουν γίνει, συνήθως επικεντρώνονται στην αξιολόγηση (απόδοσης, φυσικής κατάστασης, αποτελεσμάτων προπονητικής κλπ) και στην πρόβλεψη (πιθανών αθλητών, απόδοσης, φυσικής κατάστασης, κλπ). Η πληροφορία αυτή, μπορεί να αξιοποιηθεί για την υποστήριξη των αθλητών στη βελτίωση των προπονητικών τους συνηθειών, στην ανάπτυξη νέων στρατηγικών ομαδικών αθλημάτων, και στην εξαγωγή χρήσιμης πληροφορίας που συμβάλλει στην επίτευξη αθλητικών στόχων, [7].

Στον κλάδο του αθλητισμού, σημαντικές έρευνες έχουν γίνει που περιλαμβάνουν την ανάλυση δικτύων και συνεισφέρουν στην κατανόηση των παραγόντων, που επηρεάζουν τη συμπεριφορά

και την απόδοση των αθλητών. Η ενσωμάτωση της κοινωνικής δικτύωσης στην ανάλυση της απόδοσης, ανοίγει νέες προοπτικές για την κατανόηση των σχέσεων και των δικτύων στην αθλητική κοινότητα. Η αθλητική απόδοση μιας ομάδας, για παράδειγμα, εξαρτάται από την συνεργασία και τις αλληλεπιδράσεις μεταξύ συμπαικτών και αντιπάλων, έτσι υπάρχει ανάγκη προσδιορισμού της ατομικής και ομαδικής συνεισφοράς για την επίτευξη απόδοσης υψηλών προδιαγραφών. Η πολυπλοκότητα αυτών των αλληλεπιδράσεων μεταξύ των παικτών, αναλύεται χρησιμοποιώντας νέες ερευνητικές μεθόδους όπως τα δικτυωμένα συστήματα. Ένα τέτοιο παράδειγμα, είναι η έρευνα που έγινε για το Βραζιλιάνικο πρωτάθλημα ποδοσφαίρου[4], στην οποία ως σχέσεις μεταξύ συμπαικτών ορίστηκαν οι πάσες που ανταλλάσσουν μεταξύ τους και θεωρείται ευρέως ως η πιο γνωστή μορφή αλληλεπίδρασης σε τέτοιους αγώνες. Με τον τρόπο αυτό, δημιουργήθηκε ένα δίκτυο και εξήχθησαν σημαντικά συμπεράσματα για την έκβαση των αγώνων. Τα αποτελέσματα διαμορφώθηκαν με βάση τις απαιτήσεις τρεξίματος, το σύστημα και την τακτική της ομάδας, την τοποθεσία του αγώνα (εντός ή εκτός έδρας), τις πάσες που λαμβάνουν και εκτελούν οι παίκτες, κ.α. Η έρευνα αυτή, αποτελεί ένα επιτυχημένο παράδειγμα πώς οι αναλύσεις δικτύων μπορούν να αποτελέσουν σημαντικό παράγοντα αθλητικής απόδοσης. Πρέπει να σημειωθεί ότι οι επιδόσεις στον αθλητισμό δεν περιορίζονται μόνο στα ατομικά επιτεύγματα, αλλά διαμορφώνονται σε σημαντικό βαθμό από τις αλληλεπιδράσεις και τη συνεργασία του συνόλου. Επομένως, η κατανόηση και η εφαρμογή αποτελεσμάτων αναλύσεων αθλητικών δικτύων είναι απαραίτητη για τη βελτίωση της ομαδικής και ατομικής απόδοσης. Η σωστή αξιολόγηση των δικτυακών σχέσεων και η γνώση των παραγόντων που προκύπτουν μέσω αυτών των αναλύσεων, παρέχουν ένα εργαλείο για την επίτευξη υψηλής αθλητικής επίδοσης και την ανάπτυξη στρατηγικών για την επιτυχία των ομαδικών αθλημάτων.

Πολλές διαφορές υπάρχουν μεταξύ ομαδικών και ατομικών αθλημάτων, με κύρια διαφορά τους ότι υπάρχει μεγαλύτερη αλληλεπίδραση και σχέσεις στα ομαδικά σπορ. Σε αντίθεση, τα ατομικά, συχνά, χαρακτηρίζονται από την ανεξαρτησία και την προσωπική ευθύνη του αθλητή. Η ανάγκη για συνεχή ανάπτυξη και η επίτευξη ατομικών στόχων, αποτελούν τα κύρια κίνητρα. Για τέτοιου είδους σπορ, δημιουργούνται διάφορα γκρουπ και ομάδες με κοινούς αθλητικούς στόχους. Έτσι, οι αθλητές βρίσκουν κοινό έδαφος σε εκδηλώσεις, διοργανώσεις ή προπονητικά περιβάλλοντα. Ενώ, ο αθλητής ανταγωνίζεται και προπονείται για την ατομική του επίδοση, η αμοιβαία αναγνώριση των προσπαθειών και επιτευγμάτων μεταξύ των μελών μιας τέτοιας κοινότητας, μπορεί να δημιουργήσει διάφορες αλληλεπιδράσεις και σχέσεις μεταξύ τους.

Σε μια από τις μελέτες που εξετάζουν τη συνοχή και το κοινωνικό τρέξιμο, οι Carnes και Mahoney [19], διερεύνησαν τα αποτελέσματα ως προς την συνοχή και την κοινωνικοποίηση των αθλητών, κατά την ολοκλήρωση υψηλής έντασης προπονήσεων τρεξίματος. Ανακάλυψαν μια αντίστροφη σχέση, μεταξύ της συνοχής και της προσπάθειας που κατέβαλαν οι αθλητές για την ολοκλήρωση απαιτητικών προπονήσεων. Όταν έτρεχαν με άλλα άτομα, οι συμμετέχοντες ένιωθαν ότι οι προπονήσεις ήταν ευκολότερες, λόγω της συνοχής και της κοινωνικοποίησης. Οι ερευνητές, υποστήριξαν ότι οι ομάδες τρεξίματος θα μπορούσαν να ενθαρρύνουν τους δρομείς να καταβάλουν περισσότερη προσπάθεια και να τους βοηθήσουν να βιώσουν λιγότερες αρνητικές στάσεις.

Μία ακόμη έρευνα που έχει γίνει στα πλαίσια αυτά [35], εντοπίζει υποομάδες που δημιουργούνται σε τέτοια γκρουπ και αναλύει σχέσεις φιλίας, εμπιστοσύνης, αποτελεσματικότητας και συμβουλών. Παράγοντες που παίζουν πρωταρχικό ρόλο στη δημιουργία τέτοιων σχέσεων είναι το φύλο και οι αθλητικοί στόχοι (μαραθώνιος, ημιμαραθώνιος, πόσες μέρες τρέχουν μόνοι, κλπ). Η δυναμική των σχέσεων μεταξύ αθλητών, εμφανίζει ποικίλες διακυμάνσεις, εντοπίζοντας διάφορα επίπεδα κοινωνικής δικτύωσης στον αθλητικό χώρο. Παρατηρείται ότι ορισμένοι αθλητές απολαμβάνουν υψηλή δημοτικότητα, ενώ άλλοι επιλέγουν να λειτουργούν πιο απομονωμένα. Στα πλαίσια αυτά, η διαμόρφωση σχέσεων συνήθως προκύπτει από τον τύπο της αθλητικής ομάδας και τους εμπλεκόμενους παράγοντες. Πιο συγκεκριμένα, στο ερασιτεχνικό γκρουπ, εντοπίζεται μια τάση για αυτόνομη δράση, όπου 2-3 άτομα ίσως αναλαμβάνουν ηγετικό ρόλο και παρακινούν τα υπόλοιπα μέλη. Από την άλλη πλευρά, στην οργανωμένη ομάδα η επίδραση του προπονητή είναι πιο σημαντική. Στις περιπτώσεις αυτές, η δυναμική των σχέσεων επηρεάζεται σημαντικά από την καθοδήγηση και τις εντολές του γυμναστή. Η δημιουργία γκρουπ και η αλληλεπίδραση στο ερασιτεχνικό περιβάλλον, αναδεικνύουν τη σημασία της κοινότητας στον αθλητισμό. Η ανάλυση σχέσεων φιλίας, εμπιστοσύνης, αποτελεσματικότητας και συμβουλών σε αυτά τα γκρουπ αποκαλύπτει τη δυναμική των αθλητικών κοινοτήτων. Παράλληλα, οι διαφορές που παρατηρούνται μεταξύ ερασιτεχνικών ομάδων και επίσημα οργανωμένων γκρουπ αναδεικνύουν τη σημασία του προπονητικού παράγοντα στη διαμόρφωση των σχέσεων. Φαίνεται ότι η κοινωνική διασύνδεση στον αθλητικό χώρο, αποτελεί σύνθετο και δυναμικό πεδίο μελέτης, προσφέροντας κατανόηση των αθλητικών δικτύων και των διαδικασιών που διαμορφώνουν τις σχέσεις μεταξύ των αθλητών.

Ένα ακόμη παράδειγμα έρευνας στον τομέα της ανάλυσης αθλητικών δεδομένων, είναι η μελέτη που πραγματοποίησε ο *Radicchi* [58], εστιάζοντας στο δίκτυο επαγγελματιών αθλητών τένις. Με τη συλλογή και την ανάλυση ενός εκτεταμένου συνόλου δεδομένων από επίσημες δι-

ατάξεις αγώνων και τουρνουά, κατασκεύασε ένα δίκτυο που αντιπροσωπεύει τις δυναμικές σχέσεις μεταξύ των αθλητών. Στο πλαίσιο αυτό, το δίκτυο διαμορφώθηκε με τους αθλητές να αντιπροσωπεύονται ως κόμβοι του δικτύου, ως κατευθυνόμενες συνδέσεις μεταξύ τους οι νικητήριοι αγώνες και βάρη αυτών οι ήττες του εκάστοτε παίκτη. Τα αποτελέσματα από την έρευνα, κατέδειξαν τον καλύτερο παίκτη στην ιστορία του τένις, κατατάξεις καλύτερων παικτών ανά δεκαετίες και τους καλύτερους αθλητές σε συγκεκριμένες επιφάνειες γηπέδων. Με την ανάλυση τέτοιων δικτύων, δημιουργούνται νέες προοπτικές για την κατανόηση των αθλητικών δραστηριοτήτων και τη διαμόρφωση των αποτελεσμάτων.

Η πολυπλοκότητα των κοινωνικών δικτύων στον αθλητικό χώρο, αναδεικνύεται με τον καλύτερο δυνατό τρόπο, παρουσιάζοντας τους παράγοντες που παίζουν καθοριστικό ρόλο στη διαμόρφωση των σχέσεων μεταξύ αθλητών, όπως το άθλημα και οι αθλητικοί στόχοι. Η επίδραση των παραγόντων αυτών, στη δημιουργία αθλητικών κοινοτήτων ανακαλύπτεται μέσα από τις δομές των ερασιτεχνικών γκρουπ και των επίσημα οργανωμένων ομάδων. Η σύγχρονη ανάλυση δικτύων στον αθλητισμό, αποτελεί ένα πλούσιο υλικό για την κατανόηση των πολυπλοκοτήτων που υπάρχουν στις αθλητικές κοινότητες. Παρατηρούνται διακυμάνσεις στις σχέσεις και την κοινωνική δικτύωση των ομάδων, προσφέροντας μια ευκαιρία για βαθύτερη κατανόηση του κόσμου του αθλητισμού. Η ερευνητική προσέγγιση στην ανάλυση αθλητικών δεδομένων, όπως η μελέτη που αναφέρθηκε για το δίκτυο των επαγγελματιών αθλητών τένις [58], αποτελεί παράδειγμα καινοτόμου εφαρμογής, που ανοίγει νέες προοπτικές για την κατανόηση των αθλητικών δραστηριοτήτων και των αποτελεσμάτων. Συνολικά, η ανάλυση δικτύων στον αθλητισμό αποδεικνύεται ως ένα ισχυρό εργαλείο, για την εμβάθυνση στην κατανόηση των κοινωνικών σχέσεων στο χώρο του αθλητισμού.

ΚΕΦΑΛΑΙΟ 3

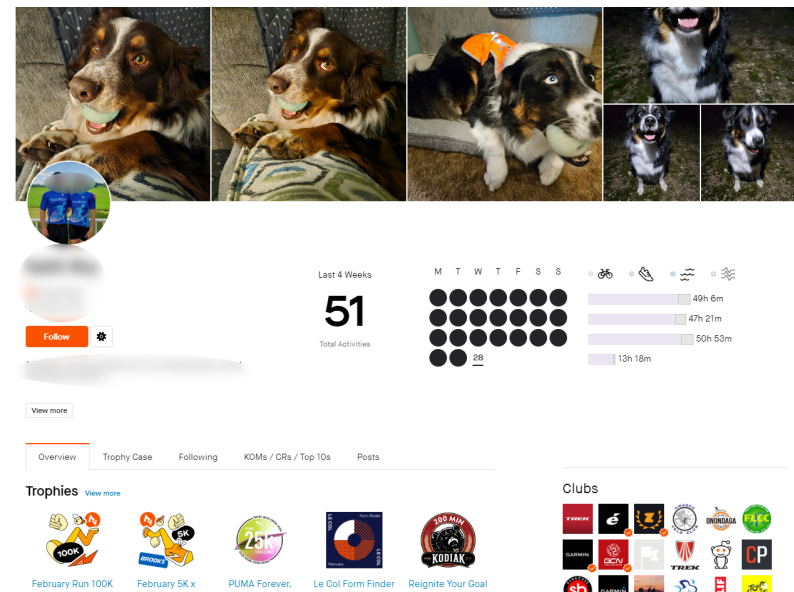
ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

Το *Strava*¹, ιδρύθηκε το 2009 από τους *Mark Gainey* και *Michael Horvath* και μέχρι σήμερα η έδρα του είναι στο Σαν Φρανσίσκο της Καλιφόρνια. Αρχικά, ήταν δημοφιλές στους ποδηλάτες και έπειτα στους δρομείς, μέχρι το 2017 πάνω από ένα δισεκατομμύριο δραστηριότητες, έχουν μεταφορτωθεί στην υπηρεσία. Μέχρι το 2020, είχε περισσότερους από 50 εκατομμύρια χρήστες και τρία δισεκατομμύρια δραστηριότητες είχαν δημοσιευθεί. Πλέον, έχει γίνει ένα κοινωνικό δίκτυο με μεγάλη δημοτικότητα, αφού το 2022 είχε 95 εκατομμύρια εγγεγραμμένους χρήστες, αύξηση 25% σε σχέση με το προηγούμενο έτος.

Αποθηκεύει λεπτομερείς πληροφορίες δραστηριοτήτων ενός χρήστη (Σχήμα 3.15), οι οποίες στη συνέχεια μπορούν να κοινοποιηθούν είτε ιδιωτικά, μεταξύ των ακολούθων του χρήστη, είτε δημόσια. Εάν ο χρήστης κοινοποιεί δημόσια, το *Strava* ομαδοποιεί αυτόματα δραστηριότητες, που πραγματοποιούνται την ίδια στιγμή ή στον ίδιο χώρο (όπως η συμμετοχή σε μαραθώνιο, αθλητικό ή ομαδικό γύρο). Οι καταγεγραμμένες πληροφορίες μιας δραστηριότητας μπορεί να περιλαμβάνουν μια διαδρομή, υψόμετρο (καθαρό και μονής κατεύθυνσης), ταχύτητα (μέσος όρος, ελάχιστος, μέγιστος), χρόνο (συνολικός και χρόνος κίνησης), ισχύ και καρδιακό ρυθμό (Σχήμα 3.16). Οι δραστηριότητες, μπορούν να καταγραφούν χρησιμοποιώντας την εφαρμογή που είναι διαθέσιμη για κινητά τηλέφωνα ή από συσκευές που κατασκευάζονται από τρίτα μέρη όπως *Garmin*, *Google Fit*, *Suunto* και *Wahoo*. Οι δραστηριότητες, μπορούν επίσης να εισαχθούν χειροκίνητα μέσω του ιστοτόπου *Strava*.

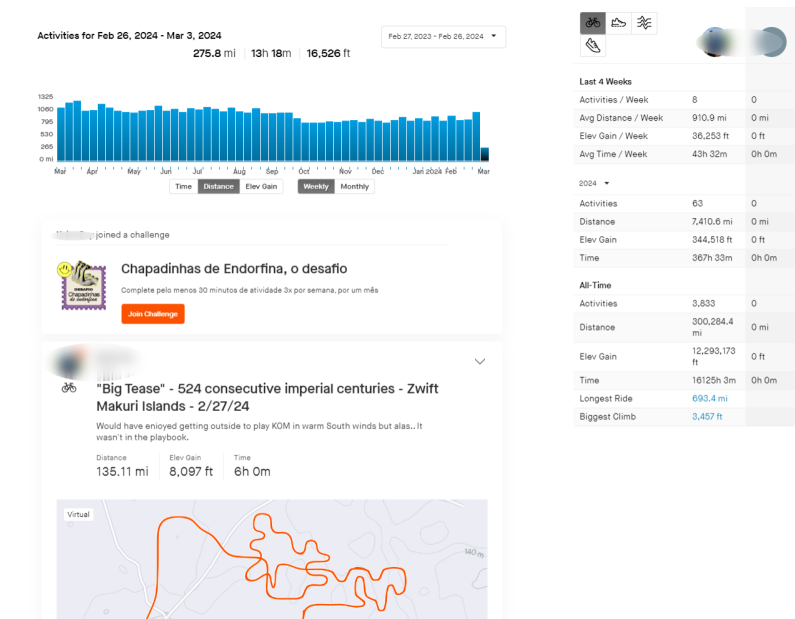
Ωστόσο, ο διαμοιρασμός τόσο προσωπικών δεδομένων έχει δημιουργήσει θέματα ασφαλείας

¹<https://www.strava.com/>



Σχήμα 3.14: Προφίλ χρήστη στο Strava.

για τους χρήστες. Οι ερευνητές, έχουν επιστήσει την προσοχή στο ποιος παρακολουθεί τον εαυτό του (αυτο-παρακολούθηση), γιατί το κάνει και πώς είναι γίνεται η διαδικασία αυτή σωστά στο πλαίσιο του αθλητισμού και της σωματικής δραστηριότητας. Μέχρι σήμερα, περιορισμένη έρευνα έχει επικεντρωθεί στην αυτο-παρακολούθηση ως κοινωνική πρακτική και υπήρξε ελάχιστη δέσμευση με τις συγκεκριμένες διαδικτυακές πλατφόρμες που χρησιμοποιούν τα άτομα για να διαμοιράζονται τέτοιες πληροφορίες [23]. Τα δεδομένα του χρήστη, είναι ρυθμισμένα ως προεπιλογή στην πλατφόρμα ως ‘public-by-default’ ή ‘private-through-effort’, αυτό σημαίνει, αρχικά, ότι κοινοποιούνται όλα δημόσια, για να γίνει ιδιωτική η κοινοποίηση, δηλαδή μεταξύ ακολούθων, θα πρέπει να γίνουν οι κατάλληλες ενέργειες. Ωστόσο, αν γίνουν αυτές οι αλλαγές τότε ο χρήστης περιορίζεται από κάποιες βασικές λειτουργίες της πλατφόρμας. Για παράδειγμα, το *leaderboard*, όπου δημιουργεί μια κατάταξη μεταξύ μελών μιας ομάδας και επιτρέπει στους χρήστες να ανταγωνιστούν μεταξύ τους. Το Strava, παρέχει μια ρύθμιση που ονομάζεται “Privacy Zone” (Ζώνη Απορρήτου). Όταν ενεργοποιείται, αποκρύπτονται τυχόν πληροφορίες που σχετίζονται με την ακριβή τοποθεσία εκκίνησης ή τερματισμού. Αυτό γίνεται, για να διατηρηθεί η ιδιωτικότητα του χρήστη και να αποτραπεί ο ακριβής εντοπισμός της αρχικής ή τελικής τοποθεσίας του. Έτσι, άνθρωποι που παρακολουθούν τις δραστηριότητες ενός χρήστη, θα βλέπουν μια περιοχή γενικά γύρω από την ακριβή τοποθεσία εκκίνησης και τερματισμού, αντί για τις ακριβείς συντεταγμένες. Οι χρήστες, μπορούν να

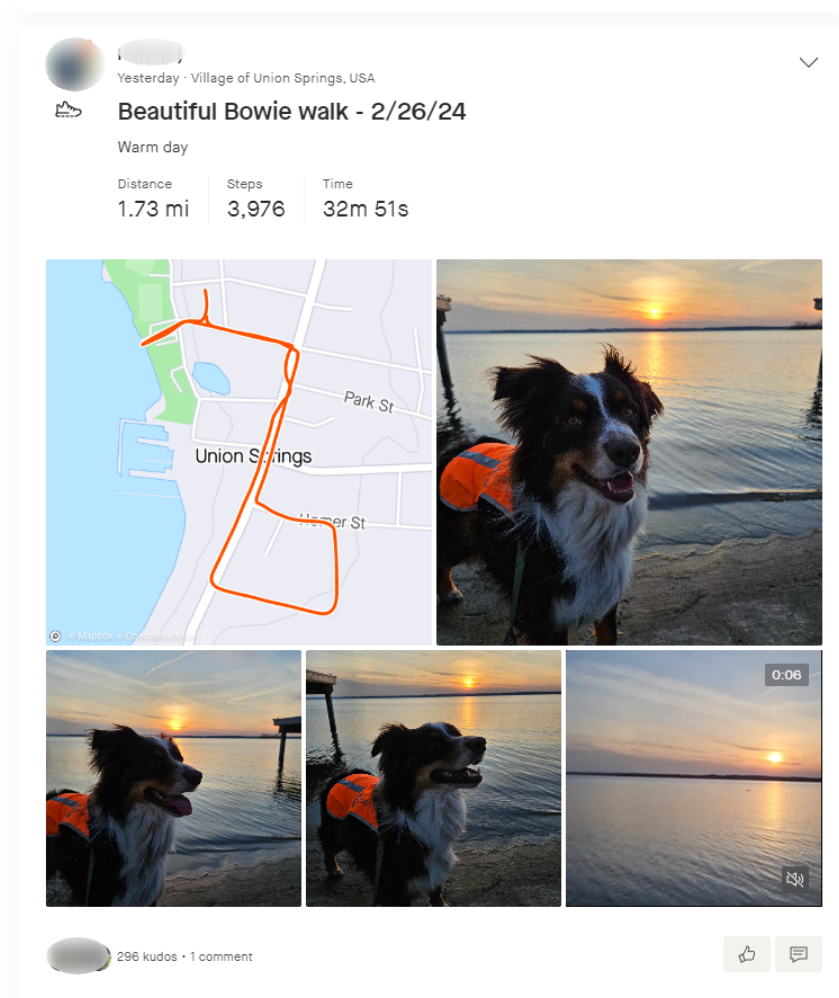


Σχήμα 3.15: Προφίλ χρήστη στο Strava και διάφορες πληροφορίες δραστηριοτήτων.

ορίσουν το μέγεθος της "Privacy Zone", δηλαδή την ακτίνα γύρω από την ακριβή τοποθεσία τους που θα παραμένει ιδιωτική. Η λειτουργία αυτή, δημιουργήθηκε επειδή οι περισσότερες δημοσιεύσεις περιέχουν ακριβείς τοποθεσίες μέσω GPS, που υποδεικνύουν δημόσια ακριβώς την χρονική στιγμή και την τοποθεσία που οι χρήστες αθλούνται, αυτό από μόνο μπορεί να ελλοχεύει κινδύνους στα άτομα που διαμοιράζονται τέτοιες πληροφορίες.

Ένα ακόμη σημαντικό ζήτημα που προκύπτει, αφορά τη σχέση μεταξύ της αυτο - παρακολούθησης και της φυσικής δραστηριότητας. Ορισμένοι αθλητές, νιώθουν ότι δεν πρέπει ή δεν μπορούν να ασχολούνται, χωρίς την καταγραφή και την κοινοποίηση των δραστηριοτήτων τους. Αυτή η εξάρτηση, υποδηλώνει τα πραγματικά κίνητρα των αθλητών. Δηλαδή, η σωματική δραστηριότητα, μπορεί να γίνεται κυρίως για τον αντίκτυπο που έχει στον περίγυρο των χρηστών και όχι αποκλειστικά για την ίδια την δραστηριότητα. Αυτό το φαινόμενο, θέτει ένα ερώτημα σχετικά με την πραγματική φύση της αυτο-παρακολούθησης και την επίδρασή της στην συμπεριφορά αυτών που την χρησιμοποιούν. Είναι σημαντικό, να καταλάβουμε τα κίνητρα της άθλησης και πώς μπορεί να διατηρηθεί μια ισορροπημένη και υγιή σχέση με την φυσική δραστηριότητα, ανεξάρτητα από την παρακολούθηση και την κοινοποίηση της [23].

Για τους λόγους αυτούς, το 2021 έγινε ενημέρωση των Πολιτικών Προστασίας Απορρήτου των δεδομένων της πλατφόρμας. Τότε, ενσωματώθηκε το "Privacy Zone" το οποίο επικεντρώνεται



Σχήμα 3.16: Δημοσίευση δραστηριότητας στο Stava.

νεται γύρω από τις διευθύνσεις που αποθηκεύει ο χρήστης μαζί με μια καθορισμένη ακτίνα γύρω από αυτήν. Ο αλγόριθμος, τροποποιεί ελαφρώς κάθε φορά την απόσταση αυτή, ώστε να μην είναι προβλέψιμο το σημείο εκκίνησης. Το πρόβλημα με την λειτουργία αυτή, είναι ότι δεν είναι εφικτό να δημιουργούνται “Privacy Zones” κάθε φορά που αλλάζει και το σημείο εκκίνησης. Για να επιλυθεί αυτό, προστέθηκαν άλλα δύο χαρακτηριστικά. Αρχικά, οι χρήστες να μπορούν να αποκρύψουν έως και 1 μίλι το σημείο εκκίνησης/τερματισμού και ακόμη να αποκρύπτουν εντελώς τον χάρτη των δραστηριοτήτων τους. Τέλος, ο χρήστης μπορεί να επεξεργάζεται ποια σημεία της διαδρομής του θέλει να διαμοιραστεί δημόσια ενώ ο ίδιος έχει πρόσβαση στα στοιχεία ολόκληρης της διαδρομής.

Σχετικές ρυθμίσεις, που αφορούν τα βιομετρικά δεδομένα προστέθηκαν, όπως επίσης και εξατομικευμένοι χάρτες με στατιστικά στοιχεία. Πιο συγκεκριμένα, ενεργοποίηση / απενεργοποίηση ορατότητας ατόμων, που έχουν πρόσβαση σε βιομετρικά δεδομένα όπως θερμίδες και καρδιακοί παλμοί. Δυνατότητα προσθήκης ιδιωτικών σημειώσεων, που αφορούν μόνο τον χρήστη. Επιλογές χάρτη, που περιλαμβάνουν την δυνατότητα προσθήκης χρωματισμένων στατιστικών χαρτών εκτός από τις βασικές επιλογές χάρτη. Όλα τα παραπάνω, παλαιότερα ήταν περιορισμένης χρήσης σε επαγγελματίες αθλητές ή επαληθευμένους λογαριασμούς, τώρα όμως όλοι οι χρήστες έχουν πρόσβαση σε αυτές τις ρυθμίσεις [48].

Η εξέλιξη αυτή των ρυθμίσεων, όπως αναφέρεται παραπάνω, αναδεικνύει τη σημασία της συλλογής δεδομένων που αφορούν τη φυσική δραστηριότητα. Η δυνατότητα παρακολούθησης και εξατομίκευσης των πληροφοριών για θερμίδες και καρδιακούς παλμούς, δίνει στους χρήστες ενισχυμένο έλεγχο πάνω στην υγεία και την απόδοσή τους. Η προσθήκη ιδιωτικών σημειώσεων, ενισχύει την ατομική εμπειρία του χρήστη, ενώ οι χάρτες με στατιστικά στοιχεία, προσφέρουν αρκετή συγκεντρωτική πληροφορία. Η συλλογή τέτοιων δεδομένων, όπως παρουσιάζεται και στην εργασία των *Franken, Bekhuis*, και *Tolsma* [30], αλλά και η ανάλυσή τους, αποτελεί μια σημαντική πηγή πληροφοριών, για την κατανόηση της επιρροής των αθλητών σε κοινωνικά δίκτυα, όπως το *Strava*, και τον τρόπο που οι χρήστες αλληλεπιδρούν μεταξύ τους.

3.1 Συλλογή Δεδομένων

Επί του παρόντος, οι εφαρμογές μεγάλων αθλητικών δεδομένων είναι δημοφιλείς αλλά αντιμετωπίζουν προβλήματα και προκλήσεις ασφάλειας. Η συλλογή τους, είναι ένα κρίσιμο έργο, για την επεξεργασία δεδομένων και η ασφαλής διαχείρισή τους, είναι ένα σημαντικό βήμα για όλα τα είδη εφαρμογών των δεδομένων αυτών, που μπορούν να παρέχουν πληροφορία χρήσιμη για ανάλυση. Η παρούσα διπλωματική εργασία, βασίζεται σε δεδομένα που συλλέχθηκαν από τους *Franken R*, *Bekhuis H* και *Tolsma J*, στην έρευνα τους με τίτλο "Kudos make you run! How runners influence each other on the online social network Strava" [30]. Τα δεδομένα, συλλέχθηκαν το διάστημα Ιανουάριος - Δεκέμβριος 2019 μέσω ενός *web scraping* εργαλείου. Τα εργαλεία *web scraping*, είναι προγράμματα που επιτρέπουν την αυτοματοποιημένη περιήγηση, ανάλυση και εξαγωγή δεδομένων από ιστοσελίδες, αντλώντας πληροφορίες μέσω του κώδικα HTML. Υπάρχει πληθώρα τέτοιων εργαλείων, όπως το *Octaparse* ², το οποίο αποτελείται

²<https://www.octoparse.com/>

από γραφικό περιβάλλον και επιτρέπει στους χρήστες να ορίζουν μόνοι τους, ή αυτόματα, μονοπάτια για την εξαγωγή δεδομένων χωρίς κώδικα ή το *Strava Clubs Reports* ³, το οποίο χρησιμοποιείται αποκλειστικά στην αναφερόμενη πλατφόρμα, συνοψίζει και εξάγει το .csv των δραστηριοτήτων από την πρόσφατη δραστηριότητα των Clubs ή από τη σελίδα προφίλ του χρήστη. Ο τρόπος που συλλέχθηκαν τα δεδομένα, στην εργασία αναφοράς, δεν περιγράφεται ακριβώς ούτε αναφέρεται συγκεκριμένα πιο εργαλείο χρησιμοποιήθηκε. Ωστόσο, είναι διαθέσιμα *online* σε ένα *GitHub repository* (<https://github.com/robfranken/Strava>), όπου οι ίδιοι οι συγγραφείς παρουσιάζουν τις αναλύσεις που έχουν συμπεριλάβει στην εργασία τους και τα αποτελέσματα.

Το σύνολο δεδομένων αυτό, περιλαμβάνει πληροφορίες σχετικές με τα *clubs*, που υπάρχουν στην πλατφόρμα του *Strava*, πιο συγκεκριμένα περιλαμβάνονται 5 Ολλανδικά *clubs* (το μικρότερο σε αριθμό περιλαμβάνει 9 άτομα ενώ το μεγαλύτερο 159). Η κάθε μια ομάδα-*club*, αποτελείται από 13 χαρακτηριστικά, όπως παρουσιάζεται στον Πίνακα 3.1. Κάποια από τα δεδομένα αυτά (*friendship*, *kudos*, *winter*, *male*, *female*, *other*), είναι σε μορφή πινάκων γειτνίασης (*adjacency matrix*), δηλαδή είναι συμμετρικοί πίνακες όπου οι γραμμές και οι στήλες αντιστοιχούν σε κόμβους του γράφου. Τα στοιχεία του πίνακα υιοθετούν την τιμή 1 όταν υπάρχει σύνδεση (ακμή) μεταξύ των αντίστοιχων κόμβων και την τιμή 0 όταν δεν υπάρχει σύνδεση. Τέτοιοι πίνακες, παρέχουν μια απλή αναπαράσταση της δομής των γράφων, καθιστώντας δυνατή την ανάλυση των σχέσεων μεταξύ των κόμβων.

Τα χαρακτηριστικά *freq-run*, *time-run*, *freq-other*, *time-other*, *novice*, *composition* και *net-size*, λαμβάνουν αμέριστε τιμές. Οι τιμές των *freq-run*, *freq-other* και *time-other*, παίρνουν τιμές 0-7, το *time-run* τιμές 0-9, το *netsize* κάθε φορά έχει μία από τις τιμές 27, 58, 159, 9, 76 και αντιπροσωπεύει το μέγεθος του δικτύου, το *novice* λαμβάνει αμέριστε τιμές ανάλογα με τα χρόνια που είναι ενεργός κάθε χρήστης στην πλατφόρμα και τέλος το *composition* μια λίστα με τιμές ανάλογα με τον αριθμό των μηνών που κατά την περίοδο δειγματοληψίας οι χρήστες παρουσιάζουν κενές τιμές. Στο σύνολο δεδομένων, οι κενές τιμές που υπάρχουν λαμβάνουν την τιμή *NaN*.

³<https://chromewebstore.google.com/detail/strava-clubs-reports/lgflepkbhehloedhbiajhlaecldnijpjd?pli=1>

Πίνακας 3.1: Χαρακτηριστικά Dataset και Περιγραφή.

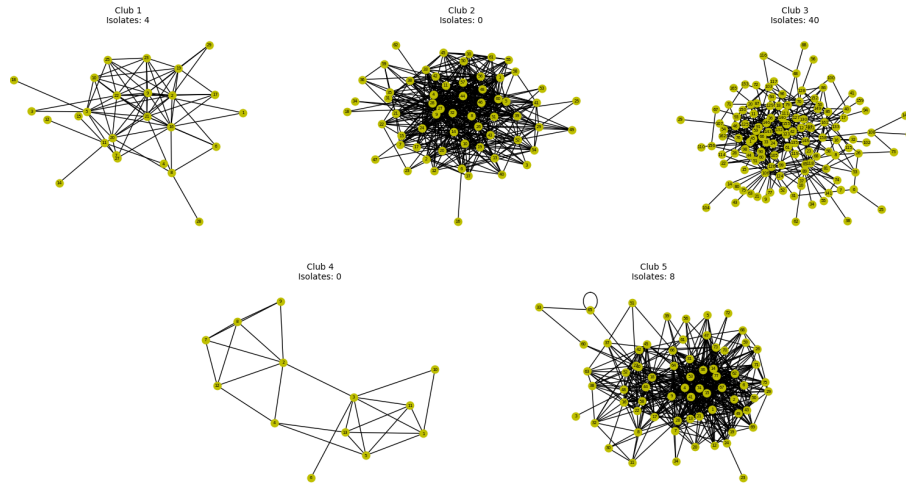
Χαρακτηριστικό	Περιγραφή
friendship	Αντιπροσωπεύει τις φιλίες μεταξύ των μελών.
kudos	Kudos μεταξύ μελών σε 12 μηνιαίες παρατηρήσεις (12 monthly time points).
freq_run	Συχνότητα τρεξίματος ανα μήνα (0 - 7+ φορές την εβδομάδα).
time_run	Χρόνος τρεξίματος σε ώρες ανά μήνα (0 - 9+ ώρες την εβδομάδα).
freq_other	Συχνότητα άλλης δραστηριότητας (εκτός από τρέξιμο) ανα μήνα (0 - 7+ φορές την εβδομάδα).
time_other	Χρόνος άλλης δραστηριότητας σε ώρες ανά μήνα (0 - 7+ ώρες την εβδομάδα).
winter	Αν τα δείγματα λήφθηκαν χειμώνα, γιατί επηρεάζει την δραστηριότητα των αθλητών.
male	Αφορά το φύλο του αθλητή (άνδρας).
female	Αφορά το φύλο του αθλητή (γυναίκα).
other	Αφορά το φύλο του αθλητή (άλλο).
composition	Σε πόσους από τους 12 μήνες οι αθλητές παρουσιάζουν κενές τιμές, έχουν υποθέσει no leavers εξωγενής αλλαγές χαρακτηριστικών.
novice	Πόσα χρόνια είναι ενεργοί οι χρήστες στο Strava.
netsize	Το μέγεθος του δικτύου.

3.2 Μεθοδολογία ανάλυσης δεδομένων

3.2.1 Γραφήματα

Με βάση τα παραπάνω, μπορούν να σχηματιστούν γραφήματα από τους πίνακες γειτνίασης, όπου τα μέλη των *group* θα αναπαριστώνται ως κόμβοι του δικτύου και οι σχέσεις φιλίας/kudos μεταξύ τους ως συνδέσεις μεταξύ των κόμβων αυτών. Για την δημιουργία των παρακάτω γραφημάτων, ιστογραμμάτων, κλπ και για την επεξεργασία των δεδομένων, έγινε χρήση της *Python* και των βιβλιοθηκών *Networkx*, *Numpy*, *Pandas* και *Matplotlib*.

Στα δίκτυα φιλίας, όπως παρουσιάζεται στους γράφους στο Σχήμα 3.17, κάθε μέλος του *club* αναδεικνύεται ως ένας κόμβος στο γράφημα, ενώ οι συνδέσεις αντιπροσωπεύουν τις φιλίες μεταξύ τους, δημιουργώντας ένα δίκτυο σχέσεων φιλίας. Πάνω από κάθε γράφημα με τον όρο



Σχήμα 3.17: Αναπαράσταση δικτύων φιλίας - *friendships graph*.

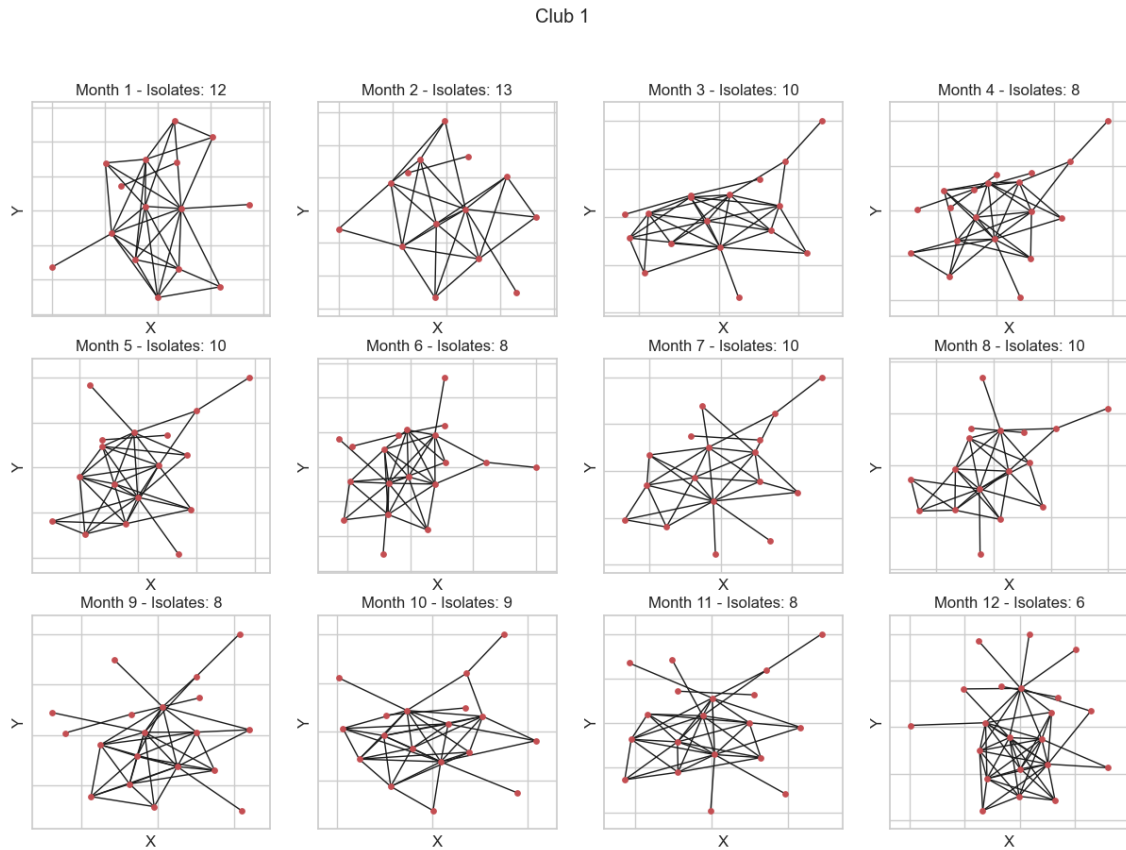
“*isolates*” αναφέρεται ο αριθμός των απομονωμένων κόμβων, οι κόμβοι αυτοί έχουν αφαιρεθεί από την αναπαράσταση του δικτύου. Από την αναπαράσταση κάθε *club*, προκύπτουν κάποια βασικά χαρακτηριστικά, όπως φαίνονται στον Πίνακα 3.2, για την επεξεργασία που θα γίνει στη συνέχεια.

Πίνακας 3.2: Χαρακτηριστικά γράφων φιλίας.

Graph	Nodes	Edges	Isolates
Club 1	30	74	4
Club 2	62	548	0
Club 3	165	420	40
Club 4	13	27	0
Club 5	77	554	8

Εκτός από τα δίκτυα φιλίας, δημιουργήθηκαν και δίκτυα *kudos*. Τα *kudos* είναι ένας τρόπος για τα μέλη της κοινότητας του *Strava*, να αλληλεπιδρούν, να παρέχουν υποστήριξη και να ενθαρρύνουν ο ένας τον άλλον. Τα μέλη κάθε *club*, μπορούν να κάνουν *kudos* μεταξύ τους ανεξάρτητα από το αν είναι φίλοι ή όχι. Στο *dataset*, περιλαμβάνονται τα *kudos* που κάνουν τα μέλη μεταξύ τους ανά 12 μήνες. Για την αναπαράσταση των δικτύων αυτών, αρχικά έγινε διαχωρισμός των μηνών (*kudos* ανά μήνα), στη συνέχεια οι κόμβοι που δεν είχαν συνδέσεις *kudos*, αφαιρέθηκαν από τους γράφους για λόγους οπτικοποίησης και προστέθηκαν με το όρο

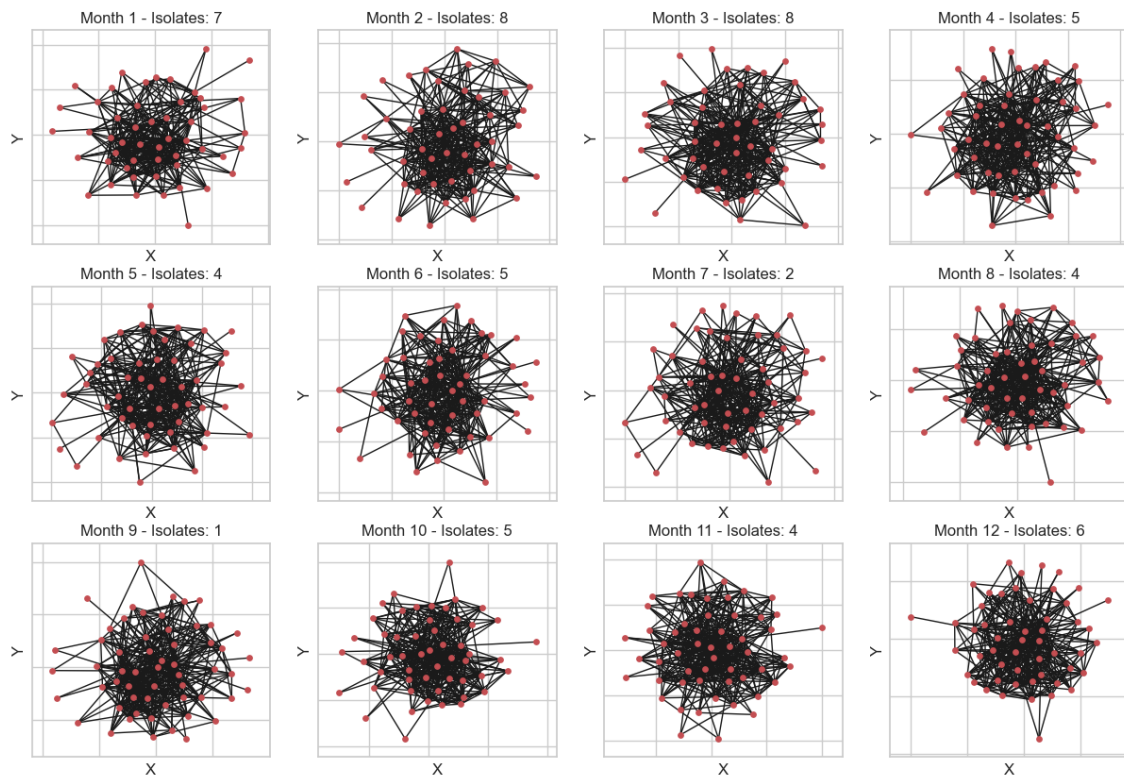
”isolates” πάνω από κάθε γράφημα. Τέλος, στο *dataset* υπήρχαν πολλές εγγραφές ανά *club*, που παρουσίαζαν κενές τιμές (*NaN values*), οι τιμές αυτές αντικαταστάθηκαν με 0 και οι χρήστες που παρουσίαζαν τέτοιες τιμές χαρακτηρίστηκαν ως ανενεργοί τους μήνες αυτούς. Έτσι, για καθένα από τα *clubs*, δημιουργήθηκαν 12 γραφήματα που παρουσιάζουν τα μέλη ως κόμβους και τα *kudos* ως συνδέσεις. Πάνω από κάθε γράφο, προστέθηκε με τον όρο ”isolates” ο αριθμός των μελών που δεν έλαβαν ούτε έδωσαν *kudos*.



Σχήμα 3.18: δίκτυο *kudos* του *Club 1*.

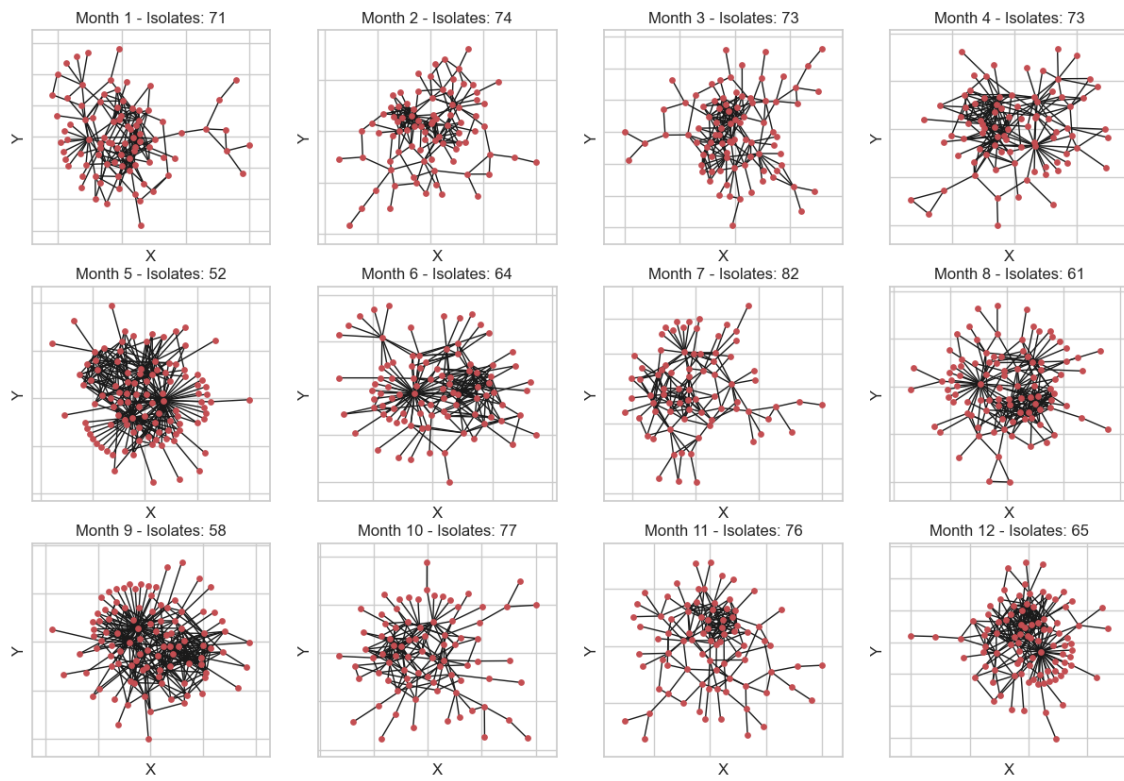
Στα Σχήματα 3.18 - 3.22, παρουσιάζονται τα δίκτυα *kudos* για κάθε *club*, με διαφορετική αναπαράσταση για κάθε μήνα παρατήρησης. Κάθε κόμβος, αντιπροσωπεύει ένα μέλος και οι συνδέσεις αντιπροσωπεύουν τις αλληλεπιδράσεις - *kudos* μεταξύ τους. Στον Πίνακα 3.3, παρουσιάζεται ο συνολικός αριθμός των μελών που λαμβάνουν *kudos* σε κάθε *club* για όλους τους μήνες. Στα γραφήματα αυτά, προστέθηκε ο αριθμός των ”isolates” πάνω από κάθε αναπαράσταση, ο όρος αυτός αναφέρεται στον αριθμό των μελών που δεν έχουν λάβει *kudos*

Club 2

Σχήμα 3.19: δίκτυο *kudos* του Club 2.

από κανένα άλλο μέλος τον συγκεκριμένο μήνα, οι κόμβοι αυτοί έχουν αφαιρεθεί από τα τελικά γραφήματα για λόγους οπτικοποίησης. Να σημειωθεί πως οι απομονωμένοι χρήστες - κόμβοι, κάθε μήνα μεταβάλλονται και δεν είναι οι ίδιοι, δηλαδή κάποιος χρήστης μπορεί τον ένα μήνα να μην λάβει ούτε να δώσει *kudos*, όμως τον επόμενο να συμμετέχει ενεργά. Αυτό, σημαίνει πως ο αριθμός των "isolates", είναι μεταβαλλόμενος και δεν αντιπροσωπεύει συγκεκριμένους κόμβους - χρήστες.

Club 3

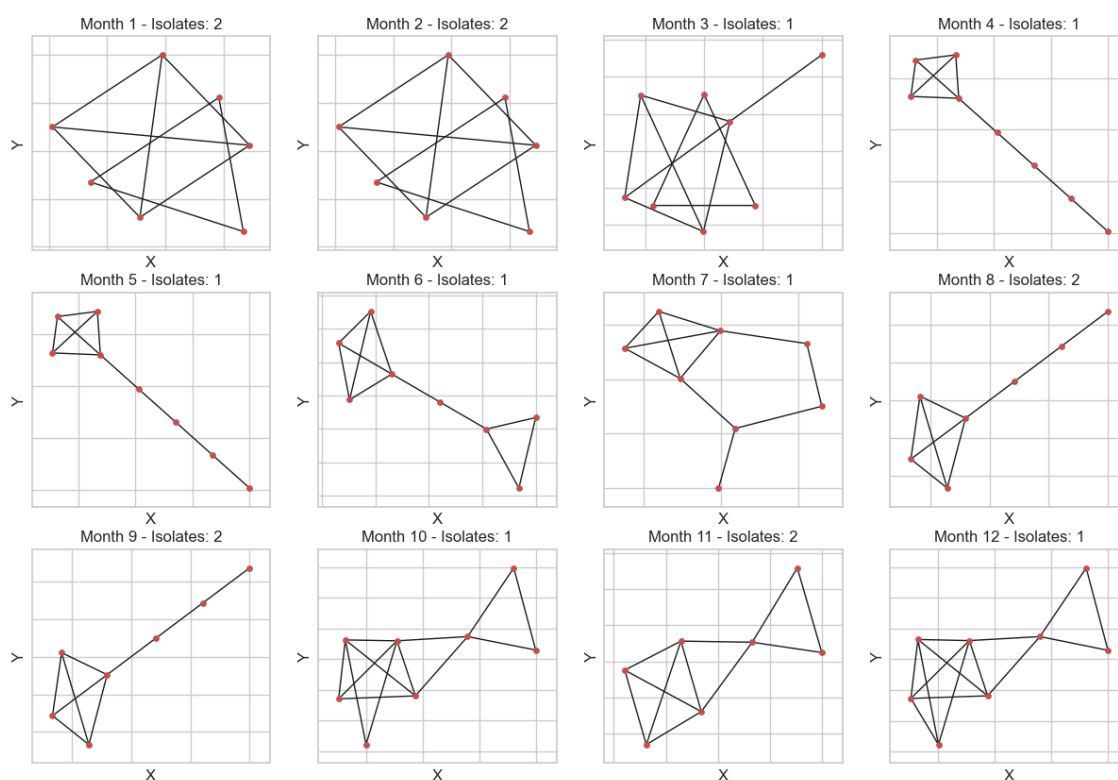
Σχήμα 3.20: δίκτυο *kudos* του Club 3.Πίνακας 3.3: Αριθμός μελών με *kudos* ανά Club.

Club	Members
Club 1	27
Club 2	58
Club 3	159
Club 4	9
Club 5	76

3.2.2 Ανάλυση Γράφων των Clubs

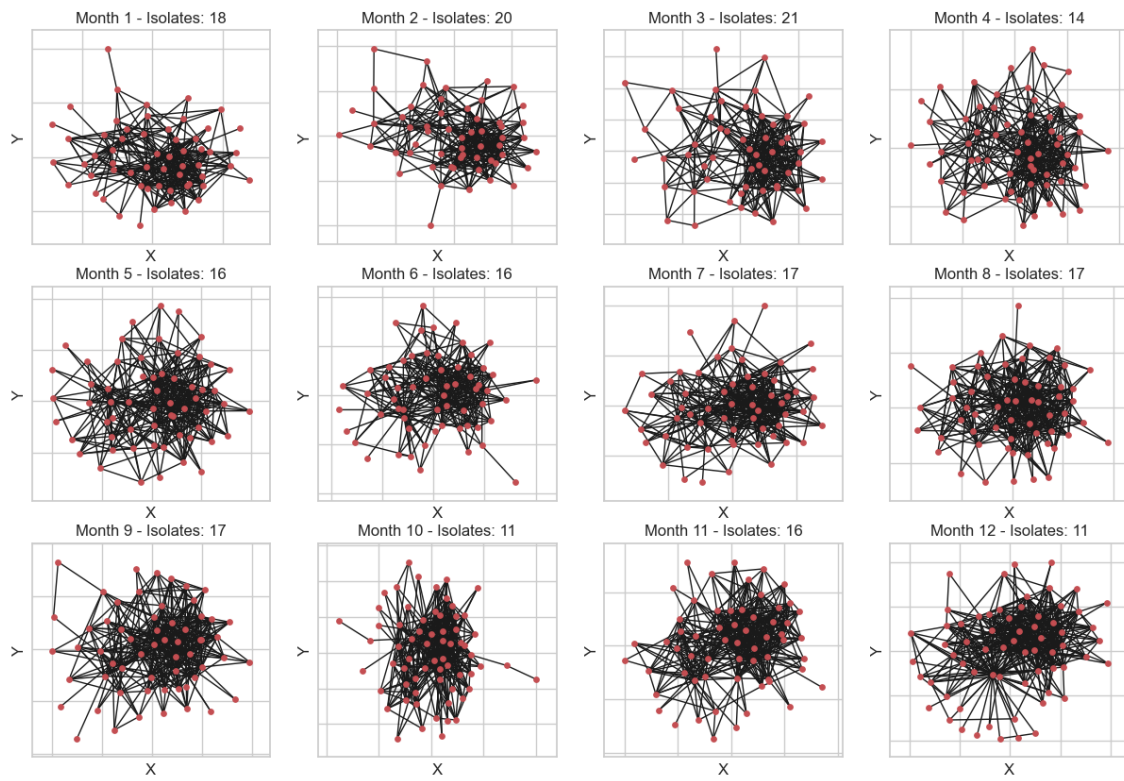
Όπως αναφέρθηκε, στο Κεφάλαιο 2, η κατανομή το μέσου βαθμού αντικατοπτρίζει τον κοινωνικό ή δικτυακό αντίκτυπο του κάθε στοιχείου. Στα δίκτυα φιλίας, που σχηματίστηκαν,

Club 4

Σχήμα 3.21: δίκτυο *kudos* του Club 4.

τον υψηλότερο μέσο βαθμό συγκέντρωσε το Club 2 με τιμή 17.68, ακολουθεί το Club 5 με 14.39, το Club 3 με 5.09 και τέλος το Club 1 και 4 με τιμές 4.93 και 4.15 αντίστοιχα (Πίνακας 3.4). Τα clubs, που συγκέντρωσαν το μεγαλύτερο *average degree*, παρουσιάζουν τον μεγαλύτερο αριθμό συνδέσεων αναλογικά με τον αριθμό των κόμβων τους και δεν έχουν κανέναν “*isolate*” κόμβο, είναι πλήρως συνδεδεμένα και πυκνά δίκτυα σχετικά με τα υπόλοιπα clubs.

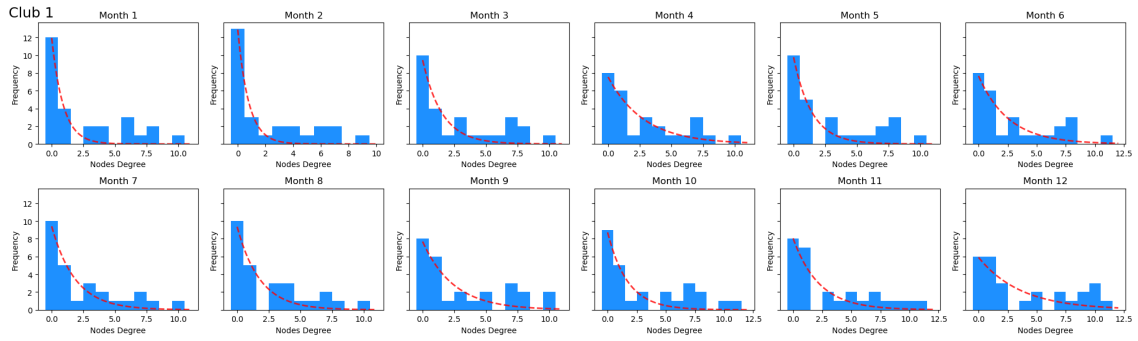
Club 5

Σχήμα 3.22: δίκτυο *kudos* του Club 5.

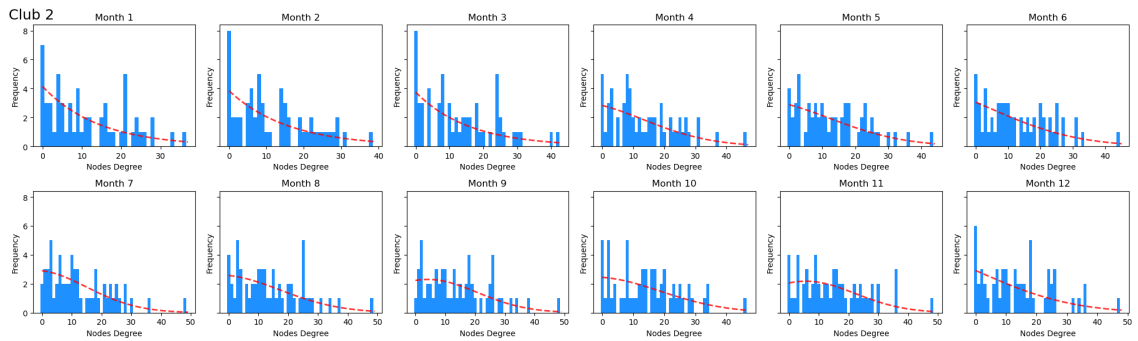
Πίνακας 3.4: Μετρικές γραφημάτων για τον προσδιορισμό τοπολογιών ανά Club.

Club	Average Degree	Shortest Path	Clustering Coefficient
Club 1	4.93	29.25	0.37
Club 2	17.68	137.75	0.59
Club 3	5.09	1046.75	0.14
Club 4	4.15	3.0	0.69
Club 5	14.39	237.5	0.57

Εκτός του μέσου βαθμού των γράφων φιλίας, για να γίνει αναγνώριση του τύπου δικτύου των γραφημάτων *kudos*, πρέπει να παρατηρηθεί η κατανομή βαθμού των κόμβων ανά μήνα του κάθε *club*. Στα Σχήματα 3.23 - 3.27, παρουσιάζονται οι κατανομές βαθμού κόμβων ανά μήνα, στα



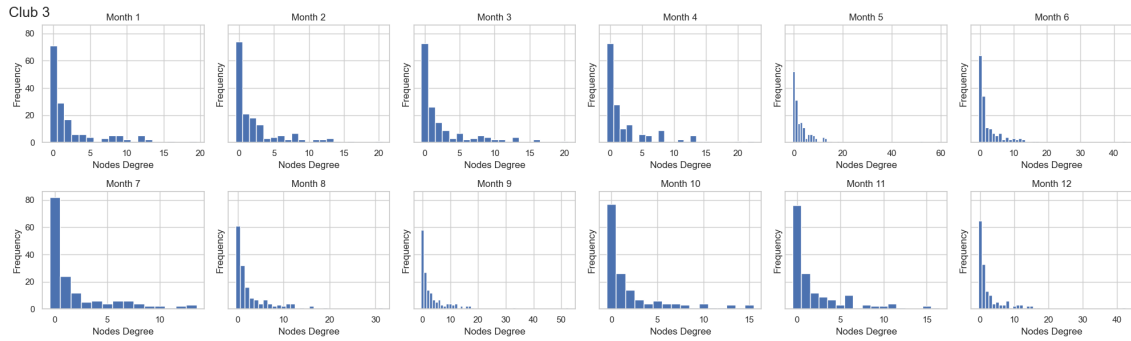
Σχήμα 3.23: *Average kudo's degree ανά μήνα του Club 1.*



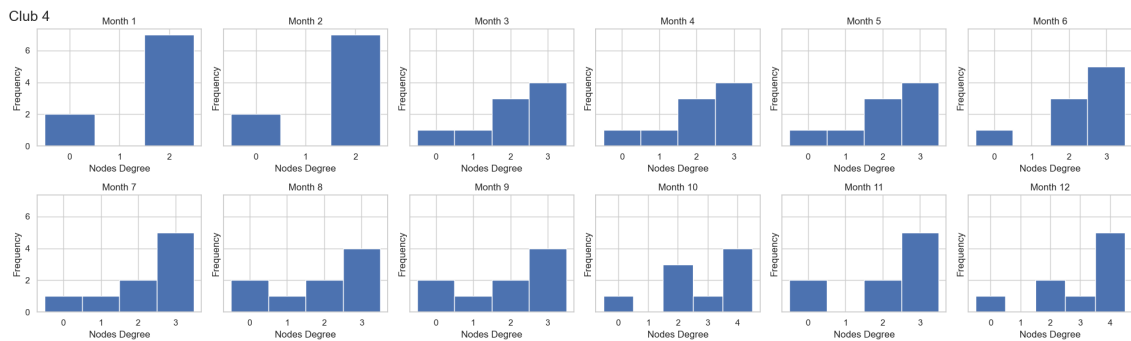
Σχήμα 3.24: *Average kudo's degree ανά μήνα του Club 2.*

γραφήματα *kudos*. Πρόσθετες μετρικές, που βοηθούν στην αναγνώριση τύπων δικτύου είναι ο συντελεστής ομαδοποίησης (*clustering coefficient*) και το μέσο μήκος μονοπατιού (*average shortest path*). Στα περισσότερα δίκτυα πραγματικού κόσμου (π.χ. κοινωνικά δίκτυα), οι κόμβοι παρουσιάζουν υψηλή συνδεσιμότητα και τείνουν να δημιουργούν ομάδες, με τον όρο του συντελεστή ομαδοποίησης, παρέχονται πληροφορίες σχετικά με το πόσο συνδεδεμένο είναι ένα γράφημα. Μια υψηλή τιμή, υποδηλώνει ότι οι γειτονικοί κόμβοι έχουν τάση να συνδέονται μεταξύ τους, ενώ μια χαμηλή δείχνει λιγότερη τάση για συνδέσεις μεταξύ των γειτόνων. Η επόμενη σημαντική μετρική που χρησιμοποιήθηκε, είναι το μέσο μήκος μονοπατιού, το οποίο περιγράφει τον βαθμό συνδεσιμότητας μεταξύ ζευγαριών κόμβων. Παρέχει πληροφορίες, για το πόσο εύκολα επικοινωνούν οι κόμβοι στο γράφημα. Ένα μικρό μέσο μήκος μονοπατιού, υποδηλώνει ότι οι κόμβοι είναι σχετικά κοντά μεταξύ τους, ενώ ένα μεγαλύτερο μέσο μήκος, ότι η επικοινωνία είναι πιο δύσκολη και πιο απομακρυσμένη. Με βάση τα παραπάνω δημιουργήθηκε ο Πίνακας 3.4 όπου περιλαμβάνει τις παραπάνω μετρικές.

Για να γίνει προσδιορισμός των τοπολογιών, λήφθηκαν υπ'όψιν όλες οι προαναφερόμενες



Σχήμα 3.25: *Average kudo's degree ανά μήνα του Club 3.*



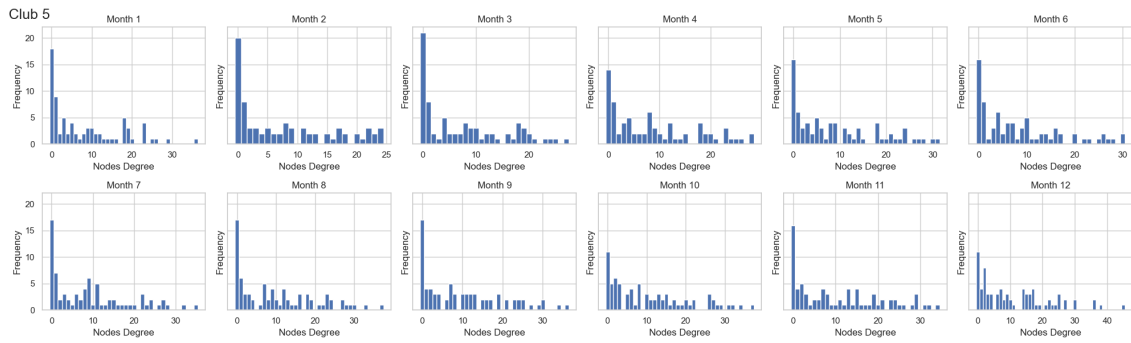
Σχήμα 3.26: *Average kudo's degree ανά μήνα του Club 4.*

μετρικές. Σε κάποια από τα *clubs*, όπως τα 3 και 5, μπορούμε να αναγνωρίσουμε εύκολα τον τύπο δικτύου από την κατανομή βαθμού κόμβων, όπως φαίνεται στα Σχήματα 3.25 και 3.27, όπου ο μέσος βαθμός υποδηλώνει *Scale Free* δίκτυα, αφού σε όλους τους μήνες ακολουθούν εκθετική κατανομή. Στο Σχήμα 3.26, στο οποίο παρουσιάζεται το *club 4*, λόγω του περιορισμένου μεγέθους του ($n = 13$), δεν είναι ξεκάθαρο σε ποια κατηγορία δικτύου ανήκει, σε συνδυασμό όμως με το σχετικά υψηλό *Clustering Coefficient* και το χαμηλό *Average Shortest Path* μπορούμε να το χαρακτηρίσουμε ως ένα δίκτυο *Small World*. Στα υπόλοιπα δίκτυα (*club 1*, *club 2*), η εικόνα κάθε *club* δεν είναι ξεκάθαρη ή μεταβάλλεται ανά μήνα.

Για τον προσδιορισμό των υπολειπομένων δύο *clubs*, έγινε χρήση ενός πρόσθετου εργαλείου που ονομάζεται *scipy.optimize.curve_fit*⁴. Το *scipy.optimize.curve_fit*, είναι ένα πακέτο της βιβλιοθήκης *SciPy*⁵ της *Python*, που χρησιμοποιείται για την προσαρμογή (*fitting*) της καμπύλης σε πειραματικά δεδομένα. Συγκεκριμένα, αναζητά την καλύτερη προσαρμογή (*fit*)

⁴https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html

⁵<https://scipy.org/>



Σχήμα 3.27: *Average kudo's degree ανά μήνα του Club 5.*

μιας συνάρτησης, σε ένα σύνολο δεδομένων με χρήση της μεθόδου ελαχίστων τετραγώνων. Έτσι, στα ιστογράμματα 3.23 και 3.24, σχηματίζεται η καμπύλη που αντιστοιχεί στην κάθε κατανομή. Με τον τρόπο αυτό, μπορούμε να χαρακτηρίσουμε το *club 1* ως ένα δίκτυο *Scale Free*, αφού ακολουθεί εκθετική κατανομή. Το *club 2*, σύμφωνα με τις καμπύλες που προκύπτουν τους πρώτους μήνες παρατήρησης ξεκινά ως ένα δίκτυο *Scale Free* και καταλήγει ως *Random* δίκτυο. Τα αποτελέσματα διαμορφώνονται όπως παρουσιάζονται στον Πίνακα 3.5, στην πρώτη στήλη αναφέρεται το *club* και στην δεύτερη ο τύπος δικτύου.

Πίνακας 3.5: *Τύπος Δικτύων ανά Club.*

Club	Τύπος Δικτύου
Club 1	Scale Free
Club 2	Scale Free & Random
Club 3	Scale Free
Club 4	Small World
Club 5	Scale Free

Η παραπάνω ανάλυση, αποτελεί το πρώτο βήμα προς την κατανόηση των τοπολογιών και των σχέσεων που δημιουργούνται στα γραφήματα *kudos*. Οι διαφορετικές ιδιότητες, που παρουσιάζουν οι διαφορετικοί τύποι δικτύων *kudos*, οι οποίες αναφέρονται στον Πίνακα 3.5, επεξηγούνται αναλυτικά και με εικόνες στο Κεφάλαιο της Βιβλιογραφικής Ανασκόπησης και πιο συγκεκριμένα στο Υποκεφάλαιο 2.3.2. Στο επόμενο κεφάλαιο ακολουθεί η μεθοδολογία που ακολουθήθηκε, καθώς περιγράφεται ο τρόπος που δημιουργήθηκαν τα παραπάνω γραφήματα και προσφέρει πλήρη κατανόηση, σχετικά με τη διαδικασία εξαγωγής των αποτελεσμάτων.

ΚΕΦΑΛΑΙΟ 4

ΜΕΘΟΔΟΛΟΓΙΑ

4.1 Επεξεργασία Δεδομένων - Δημιουργία Γράφων

Στο κεφάλαιο αυτό, γίνεται αναλυτική παρουσίαση της ανάλυσης της μεθοδολογίας που ακολουθήθηκε στην παρούσα διπλωματική. Τα δεδομένα που χρησιμοποιήθηκαν, αρχικά ήταν σε μορφή *.RData*, έτσι μέσα από το περιβάλλον *RStudio*¹, έγινε η εξαγωγή τους σε *.csv* αρχείο. Αυτό συνέβη, γιατί στις αναλύσεις που έγιναν στην εργασία αναφοράς, χρησιμοποιήθηκε η γλώσσα προγραμματισμού *R*, η οποία δίνει την δυνατότητα για υπολογιστική στατιστική και δημιουργία γραφημάτων, καθώς χρησιμοποιείται κυρίως για ανάλυση δεδομένων. Ακόμη, χρησιμοποιήθηκε το περιβάλλον της *RStudio*, για να γίνει σωστή προβολή των δεδομένων και να βοηθήσει στην κατανόηση της δομής τους. Στην εργασία αυτή, έγινε χρήση της γλώσσας προγραμματισμού *Python*, για τον λόγο αυτό έγινε και η μετατροπή του τύπου σε *.csv* αρχείο. Ακόμη, χρησιμοποιήθηκε ως περιβάλλον ανάπτυξης του κώδικα το *Visual Studio Code*² της *Microsoft*, του οποίου η χρήση είναι πολύ ευέλικτη και *Jupyter Notebook*³, που επιτρέπει τη δημιουργία και την εκτέλεση κώδικα σε μορφή κελιών, για γρήγορη προεπισκόπηση και οπτικοποίηση των αποτελεσμάτων με διαδραστική ανάπτυξη.

Κάθε φορά στον κώδικα, οι επαναλήψεις που υπάρχουν στις γραμμές 1 και 3 αφορούν την ανάγνωση των αρχείων από τον φάκελο που είναι αποθηκευμένα και τον σωστό διαχωρισμό

¹<https://rstudio-education.github.io/hopr/starting.html>

²<https://code.visualstudio.com/>

³<https://jupyter.org/>

τους ανά *club*, αυτή η μεθοδολογία επέτρεψε την ομαλή επεξεργασία και ανάλυση των δεδομένων, ενσωματώνοντας τα εργαλεία και τις τεχνικές που χρειαζόταν κάθε φορά.

4.1.1 Προεπεξεργασία Δεδομένων

Αρχικά, έγινε εγκατάσταση και χρήση των βιβλιοθηκών *pandas*⁴, *networkx*⁵, *matplotlib.pyplot*⁶, *scipy.optimize*⁷ και *numpy*⁸, που χρησιμοποιούνται για την ανάλυση και επεξεργασία γράφων και δεδομένων γενικότερα. Η *pandas*, συνήθως χρησιμοποιείται για τη διευκόλυνση της εργασίας με δομές δεδομένων όπως τα *DataFrame*, ενώ, η *networkx*, είναι μια εξειδικευμένη βιβλιοθήκη που αφορά την επεξεργασία και ανάλυση γράφων. Η διαδικασία αυτή, παρουσιάζεται παρακάτω στο Listing 4.1, στα αρχεία των *kudos*, που έγινε η λήψη τους για κάθε *club* ξεχωριστά (*kudo_list*), έπρεπε να γίνει διαχωρισμός των μηνών παρακολούθησης όπως φαίνεται στην γραμμή 4. Κάθε *club* έχει έναν πίνακα γειτνίασης, στον οποίο δεν ξεχωρίζουν οι δώδεκα μήνες παρακολούθησης. Ακόμη, το σύνολο δεδομένων έχει αρκετές τιμές *NaN*, δηλαδή κατά την δειγματοληψία υπήρχαν κενές τιμές. Για να αντιμετωπιστεί αυτό, έγινε αντικατάσταση τους με το 0 (γραμμή 3). Στη συνέχεια, δημιουργήθηκε ένας φάκελος, που ονομάστηκε *'kudo_notnull'*. Περιέχει συνολικά 60 αρχεία *.csv* με πίνακες γειτνίασης, όπου κάθε αρχείο αντιπροσωπεύει συγκεκριμένο μήνα του κάθε *club*, χωρίς να περιέχει τις κενές τιμές. Αυτά τα αρχεία, με την σειρά τους, έχουν ονομαστεί *c{j}_kudo{i}.csv* στη γραμμή 6, όπου *j* είναι ο αριθμός του *club* και *i* ο μήνας παρατήρησης. Με τον τρόπο αυτό, κάθε σύνολο δεδομένων, είναι έτοιμο για την ανάλυση.

Listing 4.1: Προεπεξεργασία δεδομένων

```

1 kudo_list = [kudo_1, kudo_2, kudo_3, kudo_4, kudo_5]
2 for j, kudo in enumerate(kudo_list, start=1):
3     kudo.fillna(0, inplace=True)
4     reshaped_data = np.array_split(kudo, 12, axis=1)
5     for i, matrix in enumerate(reshaped_data, start=1):
6         filename = f'c{j}_kudo{i}.csv'
7         np.savetxt(filename, matrix, delimiter=',', fmt='%f')
```

⁴<https://pandas.pydata.org/>

⁵<https://networkx.org/>

⁶<https://matplotlib.org/>

⁷<https://docs.scipy.org/doc/scipy/reference/optimize.html>

⁸<https://numpy.org/>

4.1.2 Εξαγωγή μετρικών

Μετά την προεπεξεργασία δεδομένων, ακολούθησε η εξαγωγή κάποιων βασικών μετρικών για τον προσδιορισμό των τοπολογιών. Οι μετρικές αυτές είναι το *average_degree*, δηλαδή μέσος αριθμός των συνδέσεων που έχει κάθε κόμβος στο γράφο, το *shortest_path*, η ελάχιστη αριθμητική απόσταση μεταξύ δύο κόμβων στον γράφο και *clustering_coefficient* ο μέσος συντελεστής συνδεσιμότητας των κόμβων στον γράφο. Η σημαντικότητα των μετρικών αυτών παρουσιάζεται στο Κεφάλαιο 3. Ο υπολογισμός τους γίνεται σε κάθε γράφο του κάθε *club* και στο τέλος εξάγεται ο μέσος όρος των μηνών, τα αποτελέσματα παρουσιάζονται στον Πίνακα 3.4. Οι τιμές αυτών των μετρικών χρησιμοποιούνται στη συνέχεια για τη σύγκριση των γράφων και των διαφόρων *clubs*, καθώς και για την κατανόηση της τοπολογικής δομής των δικτύων. Στο Listing 4.2, η εξωτερική επανάληψη στην δεύτερη γραμμή αφορά κάθε *club* (από 1- 5), η επανάληψη στη γραμμή 3 αφορά όλους τους μήνες παρατήρησης του κάθε *club*. Στην γραμμή 6 γίνονται οι σχηματισμοί των γράφων και στην 7 - 8 και 11 από έτοιμες συναρτήσεις ο υπολογισμός *average_degree*, *clustering_coefficient* και *shortest_path*. Έπειτα, στις γραμμές 9 και 10 γίνεται διαίρεση με το 12, με σκοπό να βγει ένας συνολικός μέσος όρος για κάθε *club*, όπου θα περιλαμβάνονται όλοι οι μήνες.

Listing 4.2: Δημιουργία γράφων φιλίας και εξαγωγή μετρικών.

```

1 folder_path = '../kudo-notnull/'
2 for j in range(1, 6):
3     for i in range(1, 13):
4         file_path = CreateFilePath(folder_path, j, i)
5         df = ReadFileToDataFrame(file_path)
6         G = CreateGraphFromDataFrame(df)
7         avarage_degree = CalculateAverageDegree(G)
8         shortest_path = CalculateAverageShortestPath(G)
9         avarage_degree = CalculateListAverage(avarage_degree)/12
10        shortest_path = CalculateListAverage(shortest_path)/12
11        clustering_coefficient = CalculateClusteringCoefficient(G)
12        PrintResults(j, avarage_degree, shortest_path, ←
            clustering_coefficient)

```

4.1.3 Δημιουργία Γράφων

Μετά την εξαγωγή μετρικών, έλαβε χώρα η δημιουργία των γράφων. Κάθε κόμβος του γράφου, αντιστοιχεί σε ένα μέλος του *club*, ενώ οι συνδέσεις μεταξύ τους αποτελούνται από τα *kudos* που ανταλλάσσουν τα μέλη μεταξύ τους. Τα *kudos*, που πραγματοποιούν τα μέλη μεταξύ τους, είναι χωρισμένα ανά μήνα, με 12 συνολικούς μήνες παρατήρησης, ο διαχωρισμός αυτός έγινε στο Listing 4.1. Η δημιουργία γραφημάτων παρουσιάζεται στο Listing 4.3. Στις σειρές 5 και 6, γίνεται ο προσδιορισμός του κάθε αρχείου που περιέχει τα δεδομένα που θέλουμε να μετατρέψουμε σε γραφήματα, όπου αυτό λαμβάνει χώρα στην γραμμή 7. Στη συνέχεια, στις γραμμές 8 και 9 βρίσκουμε τους απομονωμένους κόμβους τους οποίους και αφαιρούμε από τα γραφήματα. Όμως είναι μια πληροφορία που δεν θέλουμε να χαθεί και έτσι αργότερα στην γραμμή 13, την επαναχρησιμοποιούμε για να εμφανίζεται πάνω από κάθε γράφο. Η δημιουργία των γραφημάτων και το αισθητικό κομμάτι υλοποιούνται στις γραμμές 10 - 16 και 2 - 3 αντίστοιχα, παρέχοντας ένα οπτικά ελκυστικό αποτέλεσμα, όπως φαίνεται στις Εικόνες 3.18 - 3.22. Η παραπάνω διαδικασία, διασφαλίζει την κατανόηση και την οπτικοποίηση των δικτύων και των σχέσεών τους για κάθε μήνα και κάθε *club*.

Listing 4.3: Δημιουργία γραφημάτων.

```

1 for j in range(1, 6):
2     fig, axs = plt.subplots(3, 4, figsize=(15, 10))
3     fig.suptitle(f'Club {j}')
4     for i in range(1, 13):
5         file_path = os.path.join(folder_path, f'c{j}_kudo{i}.csv')
6         df = pd.read_csv(file_path, header=None)
7         G = nx.from_pandas_adjacency(df)
8         isolates = list(nx.isolates(G))
9         G.remove_nodes_from(isolates)
10        ax = axs[(i-1)//4, (i-1)%4]
11        pos = nx.layout.kamada_kawai_layout(G)
12        nx.draw_networkx(G, pos, ax=ax, with_labels=False, node_size=
            =15, node_color='r')
13        ax.set_title(f'Month {i} - Isolates: {len(isolates)}')
14    for ax in axs.flat:
15        ax.set(xlabel='X', ylabel='Y')
16    plt.show()

```

4.1.4 Δημιουργία Κατανομών Βαθμού Κόμβων

Η κατανομή βαθμού κόμβων όπως φαίνεται στο Listing 4.4, χρησιμοποιήθηκε για τον προσδιορισμό του τύπου των γράφων που δημιουργήθηκαν με τον τρόπο που περιγράφεται στο Listing 4.3. Ο υπολογισμός της κατανομής των βαθμών γίνεται στην γραμμή 4, οι υπόλοιπες γραμμές 5 - 11 και 2 αφορούν το κομμάτι της εμφάνισης των ιστογραμμάτων όπως παρουσιάζεται στις Εικόνες 3.23 - 3.27, όπου κάθε *club* αποτελείται από 12 ιστογράμματα που το καθένα αναπαριστά την κατανομή βαθμού των κόμβων του γραφήματος ανά διαφορετικό μήνα. Η κατανόηση της κατανομής αυτής είναι βασικό κομμάτι για τον προσδιορισμό της δομής των γράφων και την αλλαγή τους με την πάροδο των μηνών. Κάθε ιστόγραμμα αποτελεί μια χρονική στιγμή (ανά μήνα) και συνεισφέρει στον χαρακτηρισμό της δικτυακής δομής του κάθε *club*.

Listing 4.4: Δημιουργία Κατανομών Βαθμού Κόμβων.

```

1 for j in range(1, 6):
2     fig, axes = plt.subplots(nrows=2, ncols=6, figsize=(20, 6), ←
        sharey=True)
3     for i in range(1, 13):
4         degree_distribution = [G.degree(node) for node in G.nodes()]
5         axes[(i-1)//6, (i-1)%6].hist(degree_distribution, bins=range(←
            min(degree_distribution), max(degree_distribution) + 1), ←
            align='left', color='b')
6         axes[(i-1)//6, (i-1)%6].set_title(f'Month {i}')
7         axes[(i-1)//6, (i-1)%6].set_xlabel('Nodes Degree')
8         axes[(i-1)//6, (i-1)%6].set_ylabel('Frequency')
9     fig.text(0.005, 0.99, f'Club {j}', ha='left', va='top', fontsize←
        =18)
10    plt.tight_layout()
11    plt.show()

```

4.1.5 Εντοπισμός Κοινοτήτων

Στη συνέχεια, προχωρήσαμε στον εντοπισμό κοινοτήτων με την μέθοδο *Modularity Maximization*, στο Listing 4.5. Η βιβλιοθήκη *networkx* περιλαμβάνει ένα έτοιμο πακέτο που ονομάζεται *algorithms.community* και παρέχει διάφορες υλοποιήσεις αλγορίθμων ανίχνευσης κοινοτήτων σε γράφους. Όπως προαναφέρθηκε οι κοινότητες είναι σύνολα κόμβων σε ένα

γράφο με συχνές συνδέσεις μεταξύ τους, ενώ οι συνδέσεις μεταξύ των κοινοτήτων είναι πιο σπάνιες. Ο αλγόριθμος που εφαρμόστηκε στη γραμμή 3 του *Listing* επιδιώκει να βρει τη βέλτιστη κατανομή των κόμβων σε κοινότητες, βάσει του κριτηρίου της μετρικής *modularity*. Η μετρική αυτή, αποτελεί ένα μέτρο που εκτιμά το πόσο έντονα ένας γράφος διαμορφώνεται σε κοινότητες. Οι κοινότητες σε ένα γράφο, χαρακτηρίζονται από έντονες συνδέσεις μεταξύ των κόμβων τους, ενώ οι συνδέσεις μεταξύ διαφορετικών κοινοτήτων είναι πιο σπάνιες. Η διαδικασία αυτή, επιτρέπει να προσδιοριστούν οι δομές κοινοτήτων εντός των γράφων, προσφέροντας μια πολύτιμη εικόνα των μελών που συνδέονται στενά μεταξύ τους και σχηματίζουν κοινότητες. Αυτή η κατανόηση είναι επίσης πολύ σημαντική για την ερμηνεία της δομής του γράφου.

Listing 4.5: Εντοπισμός Κοινοτήτων με τη μέθοδο *Modularity Maximization*.

```

1 for j in range(1, 6):
2     for i in range(1, 13):
3         communities = sorted(nx_comm.greedy_modularity_communities(G) ←
                               , key=len, reverse=True)
4         print(communities)
5         print(f"The graph c{j}_kudo{i} has {len(communities)} ←
               communities.")

```

Ακόμη, έγινε εντοπισμός κοινοτήτων με τη μέθοδο *Girvan-Newman*, που είναι εξίσου μια αποτελεσματική μέθοδος που χρησιμοποιείται στην ανάλυση γραφημάτων, όπως παρουσιάζεται στο Listing 4.6. Η υλοποίηση έγινε μέσω της βιβλιοθήκης *algorithms.community*, η οποία παρέχει διάφορες υλοποιήσεις αλγορίθμων ανίχνευσης κοινοτήτων σε γράφους. Ο αλγόριθμος *Girvan-Newman* είναι γνωστός για την ικανότητά του να ανιχνεύει αλλαγές σε ένα γράφο, προκειμένου να εντοπίσει κοινότητες. Με την μεταβλητή $k = 3$ ορίστηκε ο αριθμός επαναλήψεων της διαδικασίας. Πρακτικά, το $k = 3$ σημαίνει ότι ο αλγόριθμος *girvan_newman* θα εκτελεστεί τρεις φορές (δηλαδή, θα "κόψει" τρεις φορές τις συνδέσεις στο γράφο), και κάθε φορά εξάγονται νέες κοινότητες. Ο αλγόριθμος *Girvan-Newman*, που χρησιμοποιήθηκε, επιστρέφει μια γεννήτρια που παράγει σειρές από κοινότητες. Στον κώδικα, υπάρχει το *next(communities)* για να ληφθεί η πρώτη σειρά κοινοτήτων από τη γεννήτρια.

Ο ακόλουθος κώδικας παρουσιάζει τον εντοπισμό κοινοτήτων με τη μέθοδο *Girvan-Newman*, όπως περιγράφηκε:

Listing 4.6: Εντοπισμός Κοινοτήτων με τη μέθοδο *Girvan-Newman*.

```

1 for j in range(1, 6):
2     for i in range(1, 13):
3         communities_generator = girvan_newman(G)
4         k = 3
5         communities = next(communities_generator)
6         for _ in range(k - 1):
7             communities = next(communities_generator)
8         print(communities)
9         print(f"The graph c{j}_kudo{i} has {len(communities)} ←
              communities.")

```

Τέλος, μια ακόμη μέθοδος που χρησιμοποιήθηκε για τον εντοπισμό κοινοτήτων ήταν η *Spectral Clustering*. Ο μόνος περιορισμός της μεθόδου αυτής ήταν ότι έπρεπε εξαρχής να τεθεί ο αριθμός κοινοτήτων (*clusters*) που θέλουμε ως έξοδο για κάθε *club*. Για τον λόγο αυτό, βρέθηκε ο γενικός μέσος όρος κοινοτήτων των *clubs* και έτσι χρησιμοποιήθηκε ο αριθμός 4. Ο κώδικας διαμορφώθηκε ως εξής:

Listing 4.7: Εντοπισμός Κοινοτήτων με τη μέθοδο *Spectral Clustering*.

```

1 for j in range(1, 6):
2     for i in range(1, 13):
3         communities_generator = spectral_clustering(adjacency_matrix, ←
              n_clusters=4, random_state=42)
4         print(f"The graph c{j}_kudo{i} has {len(communities)} ←
              communities.")

```

ΚΕΦΑΛΑΙΟ 5

ΑΠΟΤΕΛΕΣΜΑΤΑ

5.1 Παρουσίαση αποτελεσμάτων

Οι κοινότητες, αρχικά, προσδιορίστηκαν με την μέθοδο *Modularity Maximization*, το εργαλείο αυτό είναι σημαντικό για την ανάλυση και κατανόηση των δομών και των διασυνδέσεων μεταξύ των στοιχείων σε ένα δίκτυο (πίνακες 5.1 - 5.7). Στη συνέχεια, έγινε εντοπισμός κοινοτήτων με άλλες δύο μεθόδους *Girvan-Newman* (πίνακες 5.8 - 5.14) και *Spectral Clustering* (πίνακες 5.15 - 5.21). Η υλοποίηση των αλγορίθμων αυτών, παρουσιάζεται στα Listings 4.5 - 4.7. Οι πίνακες αυτοί, παρουσιάζουν τα αποτελέσματα από τους αλγορίθμους, για εντοπισμό κοινοτήτων, ανά *club* και παρέχουν συγκεκριμένες πληροφορίες σχετικά με τον αριθμό και την διαμόρφωση των κοινοτήτων κάθε μήνα. Στην πρώτη στήλη, βρίσκονται οι μήνες παρατήρησης, στην δεύτερη ο συνολικός αριθμός των κοινοτήτων ανά μήνα και στην τρίτη οι κόμβοι που αποτελούν κάθε κοινότητα. Οι κοινότητες, διαχωρίζονται σε παρενθέσεις, κάθε παρένθεση δηλαδή περιλαμβάνει μια διαφορετική κοινότητα του γράφου. Με την μέθοδο *Spectral Clustering*, ο αριθμός των κοινοτήτων παραμένει σταθερός ($\text{clusters} = 4$), αφού έχει τεθεί από την αρχή της διαδικασίας εύρεσης των κοινοτήτων. Ο αριθμός 4 προκύπτει από τον μέσο όρο των κοινοτήτων που σχηματίζονται με τις άλλες μεθόδους.

Πίνακας 5.1: Κοινότητες *Club 1* ανά μήνα με τη μέθοδο *Modularity Maximization*.

Μήνας	Communities	Κόμβοι
1	3	(17, 18, 19, 3, 6, 23, 8) (0, 16, 1, 20, 21, 7) (9, 2)
2	3	(0, 16, 1, 19, 20, 21, 7) (17, 18, 3, 6, 23) (9, 2)
3	3	(0, 16, 1, 19, 20, 21, 7) (17, 18, 3, 6, 23, 8, 24) (9, 2, 10)
4	5	(0, 16, 1, 20, 7) (17, 19, 3, 23, 8) (24, 18, 21, 6) (9, 2, 10) (13, 5)
5	3	(0, 1, 7, 8, 16, 19, 20, 21) (17, 18, 3, 6, 23, 24, 12) (9, 2)
6	4	(16, 1, 18, 19, 20, 21, 8) (17, 3, 6, 23, 24, 12) (9, 2, 10) (0, 25, 7)
7	3	(16, 1, 0, 19, 20, 21, 7, 25) (17, 18, 3, 6, 23, 24, 12) (9, 2)
8	4	(16, 1, 19, 20, 21) (17, 18, 6, 24, 12) (0, 3, 7, 23, 25) (9, 2)
9	4	(0, 16, 1, 19, 20, 21, 7, 8) (17, 3, 12, 23) (24, 18, 4, 6) (9, 2, 10)
10	4	(0, 1, 16, 20, 7, 25) (17, 19, 3, 23, 8) (18, 4, 21, 6, 24) (9, 2)
11	4	(0, 1, 7, 14, 16, 19, 20, 21, 25) (17, 3, 12, 23) (24, 18, 4, 6) (9, 2)
12	4	(16, 1, 19, 20, 21, 7, 25) (0, 17, 3, 23, 12, 14) (18, 4, 6, 24, 13, 15) (9, 2)

Πίνακας 5.2: Κοινότητες *Club 4* ανά μήνα με τη μέθοδο *Modularity Maximization*.

Μήνας	Communities	Κόμβοι
1	2	(0, 3, 4, 7) (8, 2, 6)
2	2	(0, 3, 4, 7) (8, 2, 6)
3	2	(0, 1, 3, 4, 7) (8, 2, 6)
4	2	(0, 3, 4, 7) (8, 1, 2, 6)
5	2	(0, 3, 4, 7) (8, 1, 2, 6)
6	2	(0, 1, 3, 4, 7) (8, 2, 6)
7	2	(0, 3, 4, 7) (8, 1, 2, 6)
8	2	(0, 3, 4, 7) (8, 1, 2)
9	2	(0, 3, 4, 7) (8, 1, 2)
10	2	(0, 3, 4, 5, 7) (8, 1, 2)
11	2	(0, 3, 4, 7) (8, 1, 2)
12	2	(0, 3, 4, 5, 7) (8, 1, 2)

Πίνακας 5.3: Κοινότητες *Club 2* ανά μήνα με τη μέθοδο *Modularity Maximization*.

Μήνας	Communities	Κόμβοι
1	3	(0, 3, 4, 10, 19, 21, 23, 25, 26, 27, 28, 33, 34, 37, 39, 40, 41, 43, 45, 50, 51, 52, 55, 56) (1, 2, 6, 7, 8, 9, 11, 12, 13, 15, 16, 17, 22, 30, 31, 38, 42, 47, 48, 49, 53, 54) (18, 20, 24, 46, 14)
2	2	(0, 3, 4, 10, 14, 19, 20, 21, 23, 24, 25, 26, 27, 28, 33, 34, 37, 39, 40, 41, 43, 45, 46, 50, 51, 52, 55, 56) (1, 2, 6, 7, 8, 9, 11, 12, 13, 15, 16, 17, 18, 22, 30, 31, 38, 42, 47, 49, 53, 54)
3	3	(1, 2, 6, 8, 9, 11, 12, 13, 15, 16, 17, 18, 20, 22, 26, 29, 30, 31, 37, 38, 47, 48, 49, 54) (0, 3, 4, 7, 10, 19, 23, 25, 27, 28, 34, 39, 40, 41, 43, 45, 46, 50, 51, 52, 53, 55, 56) (24, 21, 14)
4	2	(1, 2, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 26, 30, 31, 37, 38, 47, 48, 53, 54, 57) (0, 3, 4, 10, 18, 24, 25, 27, 28, 33, 34, 36, 39, 40, 41, 43, 45, 46, 49, 50, 51, 52, 55, 56)
5	3	(1, 2, 5, 6, 8, 9, 11, 12, 13, 14, 15, 16, 17, 22, 24, 30, 31, 38, 42, 47, 48, 49, 53, 54, 57) (0, 3, 10, 19, 23, 25, 26, 27, 28, 33, 37, 40, 41, 43, 45, 50, 51, 52, 55) (34, 4, 36, 7, 39, 46, 18, 20, 21, 56)
6	4	(1, 2, 5, 7, 9, 11, 12, 13, 15, 16, 20, 22, 24, 38, 44, 47, 49, 54, 57) (0, 33, 3, 40, 41, 45, 46, 50, 51, 52, 23, 18, 55, 56, 25, 26, 27, 28) (34, 36, 4, 37, 6, 8, 42, 10, 43, 14, 17, 19, 21, 30) (53, 39)
7	3	(0, 3, 4, 10, 23, 25, 26, 27, 28, 34, 35, 36, 37, 39, 40, 41, 43, 45, 46, 50, 51, 52, 53, 55) (1, 2, 5, 7, 9, 12, 15, 16, 18, 20, 21, 22, 24, 44, 47, 49, 57) (32, 33, 38, 6, 8, 42, 11, 13, 14, 48, 17, 19, 54, 30, 31)
8	3	(0, 8, 10, 14, 17, 18, 23, 25, 26, 27, 28, 30, 31, 36, 37, 39, 40, 41, 42, 45, 46, 50, 51, 52, 53, 55, 56) (1, 2, 5, 7, 9, 11, 12, 13, 15, 16, 20, 22, 24, 32, 38, 44, 47, 48, 49) (33, 34, 3, 4, 6, 43, 19, 21)
9	4	(1, 2, 5, 6, 7, 8, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 30, 31, 32, 35, 38, 42, 47, 49) (0, 3, 4, 10, 23, 25, 27, 28, 33, 34, 37, 40, 41, 43, 45, 46, 50, 51, 52, 55, 56) (36, 39, 9, 44, 48, 18, 53, 54, 24, 57) (26, 22)
10	3	(0, 3, 4, 10, 18, 21, 23, 24, 25, 26, 27, 28, 34, 36, 37, 40, 41, 43, 44, 45, 46, 50, 51, 52, 55, 56) (1, 2, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 19, 20, 22, 30, 31, 32, 35, 38, 42, 47, 49) (53, 39)
11	3	(1, 2, 5, 6, 8, 9, 11, 12, 13, 15, 16, 17, 18, 22, 27, 30, 31, 32, 35, 38, 54) (34, 4, 37, 36, 39, 41, 10, 43, 44, 14, 46, 47, 49, 19, 21, 53, 24) (0, 3, 7, 20, 23, 25, 26, 28, 40, 42, 45, 50, 51, 52, 55, 56)
12	4	(0, 3, 7, 18, 20, 23, 25, 26, 27, 28, 40, 41, 45, 50, 51, 52, 55, 56) (32, 37, 6, 8, 9, 10, 43, 42, 14, 17, 19, 54, 30, 31) (2, 34, 35, 4, 36, 12, 13, 44, 46, 47, 16, 21, 24) (1, 5, 38, 22, 53, 39, 15)

Πίνακας 5.4: Κοινότητες *Club 3* ανά μήνες 1-6 με τη μέθοδο *Modularity Maximization*.

Μήνας	Communities	Κόμβοι
1	11	(129, 66, 74, 11, 141, 15, 148, 21, 149, 23, 30, 99, 42, 112, 49, 55, 56, 121, 126) (134, 71, 136, 10, 139, 80, 81, 84, 92, 95, 31, 34, 40, 43, 44, 117, 124, 61) (64, 65, 3, 68, 9, 82, 83, 153, 96, 97, 108, 109, 110, 46, 113, 52, 118) (100, 36, 37, 6, 135, 41, 79, 50, 54, 120, 157) (5, 7, 76, 151, 24, 57, 154, 88, 29) (143, 17, 19, 111) (94, 38) (28, 102) (105, 85) (131, 22) (146, 101)
2	7	(64, 131, 3, 6, 135, 9, 76, 79, 83, 20, 147, 22, 87, 95, 32, 36, 100, 41, 108, 50, 118, 120) (134, 71, 136, 10, 139, 14, 17, 84, 21, 26, 92, 31, 34, 99, 40, 43, 44, 117, 119, 61) (129, 66, 74, 11, 141, 142, 15, 148, 149, 23, 30, 57, 42, 112, 49, 55, 56, 121, 126) (5, 7, 80, 81, 113, 151, 24, 154, 88, 124, 29) (65, 97, 101, 110, 46, 82, 146, 94) (143, 19, 111) (28, 102)
3	9	(134, 71, 10, 139, 14, 17, 81, 84, 21, 92, 31, 34, 101, 40, 43, 44, 117, 119, 61) (129, 136, 74, 11, 141, 142, 15, 148, 149, 23, 30, 121, 42, 112, 49, 55, 56, 57, 126) (6, 135, 9, 79, 83, 91, 96, 32, 100, 36, 37, 41, 50, 54, 120, 125) (5, 7, 76, 80, 113, 151, 24, 154, 88, 124, 29, 95) (97, 82, 65, 146, 46, 94, 110) (64, 131, 147, 118, 22, 108) (143, 19, 111) (28, 102) (105, 85)
4	10	(3, 5, 6, 135, 7, 9, 151, 152, 24, 154, 26, 29, 32, 36, 39, 50, 52, 68, 76, 79, 80, 83, 88, 91, 95, 100, 106, 113, 117, 118, 120, 125) (134, 136, 10, 139, 17, 81, 84, 92, 93, 31, 34, 101, 40, 43, 44, 119, 121, 124, 61) (129, 71, 74, 11, 141, 142, 15, 148, 21, 23, 30, 42, 112, 49, 55, 56, 57, 126) (65, 110, 94, 97) (143, 19, 111) (82, 46) (28, 102) (105, 85) (108, 87) (131, 22)
5	6	(65, 68, 5, 135, 7, 137, 140, 143, 82, 19, 146, 24, 89, 88, 154, 156, 93, 29, 94, 95, 97, 36, 101, 106, 46, 111, 110, 125) (64, 130, 131, 3, 133, 72, 8, 10, 73, 17, 18, 22, 151, 152, 153, 26, 28, 157, 96, 33, 102, 39, 41, 113, 115, 54, 58, 62) (134, 71, 136, 139, 81, 84, 149, 23, 92, 31, 121, 34, 40, 42, 43, 44, 49, 117, 119, 56, 57, 124, 61) (6, 9, 76, 79, 83, 148, 86, 30, 32, 100, 37, 108, 50, 52, 118, 120) (129, 74, 11, 141, 142, 15, 112, 55, 87, 126) (105, 85)
6	7	(3, 4, 68, 5, 7, 72, 80, 16, 81, 151, 152, 24, 88, 154, 28, 93, 95, 102, 103, 106, 113, 58, 123, 124, 62) (131, 134, 71, 10, 139, 14, 84, 22, 92, 31, 34, 101, 39, 40, 42, 43, 108, 44, 119, 61) (129, 136, 74, 11, 141, 142, 15, 148, 23, 30, 57, 112, 49, 55, 56, 121, 126) (32, 100, 6, 41, 9, 76, 109, 79, 50, 83, 52, 147, 118, 120) (64, 33, 2, 36, 37, 135, 54, 157, 125) (65, 97, 110, 143, 111, 46, 82, 146) (105, 85)

Πίνακας 5.5: Κοινότητες *Club 3* ανά μήνες 7-12 με τη μέθοδο *Modularity Maximization*.

Μήνας	Communities	Κόμβοι
7	6	(3, 68, 5, 7, 79, 80, 81, 151, 24, 88, 154, 29, 95, 113, 51, 124, 63) (34, 101, 134, 71, 40, 136, 10, 43, 44, 139, 84, 119, 92, 61, 31) (129, 121, 74, 42, 11, 141, 142, 15, 112, 55, 23, 56, 57, 126, 30) (32, 131, 100, 102, 6, 9, 108, 49, 50, 83, 148, 22, 118, 120, 28) (64, 36, 37, 54, 135, 157, 91, 125) (143, 97, 82, 65, 46, 110, 111)
8	8	(64, 2, 3, 69, 70, 6, 135, 9, 79, 80, 18, 83, 151, 153, 157, 32, 33, 100, 36, 37, 103, 39, 105, 41, 113, 50, 115, 54, 120, 125, 63) (129, 131, 136, 74, 11, 141, 142, 15, 22, 23, 30, 121, 42, 112, 49, 51, 55, 56, 57, 126) (134, 71, 10, 139, 81, 84, 26, 92, 95, 31, 34, 101, 43, 44, 117, 124, 61) (68, 133, 5, 7, 8, 154, 24, 58, 88, 29) (97, 65, 46, 143, 111, 110, 82, 19, 146) (148, 132, 118, 150, 72, 108) (138, 28, 102) (40, 123)
9	6	(133, 6, 135, 5, 9, 138, 7, 12, 145, 147, 148, 22, 151, 152, 150, 24, 154, 28, 29, 33, 36, 39, 41, 50, 51, 54, 58, 62, 64, 66, 68, 70, 72, 73, 79, 88, 89, 93, 95, 96, 100, 102, 103, 108, 113, 118, 120, 125) (0, 134, 71, 136, 10, 139, 84, 23, 26, 155, 92, 31, 32, 34, 101, 40, 43, 44, 109, 49, 119, 56, 57, 123, 61) (129, 3, 69, 42, 74, 11, 141, 142, 15, 112, 55, 121, 126, 30) (97, 65, 46, 143, 111, 110, 82, 19, 146) (80, 81, 124) (98, 13)
10	8	(0, 71, 136, 84, 155, 92, 158, 31, 34, 101, 40, 43, 44, 49, 119, 123) (32, 129, 11, 141, 142, 15, 112, 51, 55, 23, 121, 126, 30) (134, 102, 10, 139, 79, 124, 81, 117, 151, 28, 61, 95) (97, 65, 74, 110, 46, 111, 143, 82, 146, 19, 57) (64, 33, 131, 103, 108, 148, 21, 22, 118, 56, 149) (68, 5, 7, 145, 24, 154, 88, 29) (50, 147, 100, 120, 9, 59, 41) (153, 125, 69, 135)
11	8	(136, 11, 76, 141, 142, 21, 149, 23, 30, 57, 112, 49, 51, 55, 56, 121, 126) (0, 134, 71, 10, 139, 84, 22, 155, 92, 31, 34, 101, 40, 43, 44, 119, 61) (100, 69, 102, 70, 6, 41, 9, 135, 79, 50, 120, 91, 28, 125) (65, 97, 74, 110, 46, 143, 17, 111, 82, 19, 117, 95) (78, 48, 145, 80, 81, 113, 151, 124) (68, 5, 7, 24, 154, 123, 29, 158) (148, 118, 59, 108, 93) (98, 13)
12	6	(131, 3, 132, 135, 7, 8, 9, 146, 147, 148, 22, 151, 154, 26, 28, 157, 32, 33, 35, 58, 72, 87, 95, 96, 102, 107, 108, 109, 113, 114, 118, 121, 124, 125) (0, 134, 71, 136, 137, 10, 139, 84, 155, 92, 31, 34, 101, 103, 40, 43, 44, 117, 119, 61) (65, 74, 78, 142, 80, 17, 82, 143, 19, 145, 97, 46, 110, 48, 111, 53, 122) (129, 11, 141, 112, 49, 51, 23, 149, 30, 55, 56, 57, 126) (50, 100, 69, 120, 41, 79) (24, 123, 68, 5)

Πίνακας 5.6: Κοινότητες *Club 5* ανά μήνες 1-9 με τη μέθοδο *Modularity Maximization*.

Μήνας	Communities	Κόμβοι
1	4	(0, 65, 3, 67, 5, 6, 70, 8, 73, 11, 75, 12, 13, 14, 18, 19, 20, 26, 27, 34, 37, 39, 41, 45, 47, 48, 51, 57, 60, 62) (64, 7, 72, 10, 74, 15, 16, 17, 21, 25, 28, 30, 32, 35, 36, 43, 46, 49, 50, 55, 56, 58, 61, 63) (59, 38) (1, 68)
2	5	(64, 0, 65, 3, 67, 5, 6, 70, 72, 73, 11, 12, 75, 13, 14, 18, 19, 20, 26, 27, 34, 39, 41, 45, 47, 48, 49, 57, 60, 62) (7, 8, 10, 15, 16, 17, 21, 25, 28, 30, 35, 36, 37, 43, 46, 50, 51, 56, 58, 61) (59, 38) (63, 31) (1, 68)
3	5	(64, 0, 65, 3, 5, 70, 6, 72, 73, 11, 12, 75, 13, 14, 18, 20, 22, 26, 34, 39, 41, 45, 47, 48, 51, 60, 62) (7, 8, 10, 15, 16, 17, 19, 21, 25, 27, 30, 35, 36, 37, 43, 46, 49, 50, 56, 57, 58, 61) (59, 38) (24, 66) (1, 68)
4	6	(64, 0, 65, 67, 3, 69, 6, 5, 71, 72, 73, 11, 75, 13, 12, 14, 18, 20, 22, 26, 29, 34, 41, 45, 47, 48, 51, 60, 62) (7, 10, 15, 16, 17, 21, 25, 28, 30, 32, 35, 36, 37, 42, 43, 46, 49, 50, 56, 58, 61) (19, 70, 39, 8, 57, 27) (59, 38) (24, 66) (1, 68)
5	6	(64, 0, 65, 3, 5, 70, 71, 72, 73, 11, 75, 12, 14, 13, 18, 20, 22, 26, 29, 34, 39, 41, 45, 47, 48, 51, 60, 62) (6, 7, 10, 15, 16, 17, 21, 25, 28, 30, 32, 35, 36, 37, 42, 43, 46, 49, 50, 56, 58, 61) (8, 27, 19, 57) (59, 38) (24, 66) (1, 68)
6	6	(64, 0, 65, 3, 5, 6, 70, 71, 72, 73, 11, 75, 13, 12, 14, 18, 20, 22, 29, 34, 39, 41, 45, 47, 48, 51, 60, 62) (7, 10, 15, 16, 17, 21, 25, 28, 30, 32, 35, 36, 37, 42, 43, 46, 49, 50, 56, 58, 61) (8, 19, 27, 57) (24, 26, 66) (59, 38) (1, 68)
7	5	(0, 65, 66, 69, 6, 70, 5, 72, 8, 11, 75, 13, 14, 18, 19, 20, 22, 24, 26, 27, 34, 39, 41, 45, 47, 48, 51, 57, 60, 62) (7, 10, 15, 16, 17, 21, 25, 30, 35, 36, 42, 43, 46, 50, 54, 56, 58, 61) (64, 49, 3, 37, 71, 12, 29) (59, 38) (1, 68)
8	5	(64, 0, 65, 3, 69, 70, 6, 71, 72, 11, 75, 13, 18, 20, 22, 26, 37, 39, 41, 47, 48, 60, 62) (7, 15, 16, 17, 21, 25, 28, 30, 35, 36, 42, 43, 46, 49, 50, 56, 58, 61) (66, 34, 67, 5, 8, 12, 45, 14, 51, 19, 54, 24, 57, 27) (59, 38) (1, 68)
9	4	(64, 4, 7, 8, 10, 15, 16, 17, 21, 25, 28, 30, 35, 36, 42, 43, 46, 49, 50, 55, 56, 58, 61) (0, 65, 3, 5, 70, 69, 72, 73, 11, 75, 13, 14, 18, 20, 22, 26, 39, 41, 47, 48, 57, 60, 62) (34, 66, 37, 71, 12, 45, 51, 19, 54, 24, 27) (1, 68)

Πίνακας 5.7: Κοινότητες *Club 5* ανά μήνες 9-12 με τη μέθοδο *Modularity Maximization*.

Μήνας	Communities	Κόμβοι
10	4	(0, 3, 5, 8, 11, 12, 13, 14, 18, 19, 20, 22, 24, 27, 29, 34, 37, 38, 39, 41, 45, 47, 48, 51, 54, 57, 59, 60, 62, 65, 66, 69, 70, 71, 72, 73, 75) (2, 4, 7, 10, 15, 16, 17, 21, 25, 26, 28, 30, 32, 35, 36, 40, 42, 43, 44, 46, 50, 56, 58, 61) (64, 49) (1, 68)
11	4	(0, 65, 66, 3, 5, 71, 72, 8, 11, 75, 12, 13, 14, 18, 19, 20, 22, 27, 29, 34, 41, 45, 47, 48, 51, 57, 60) (2, 7, 15, 16, 17, 21, 25, 26, 28, 30, 32, 35, 36, 37, 40, 42, 43, 46, 49, 50, 54, 56, 58, 61) (64, 4, 6, 39, 70, 73, 62) (1, 68)
12	2	(0, 3, 5, 6, 8, 11, 12, 13, 14, 18, 19, 20, 22, 26, 27, 28, 29, 34, 37, 39, 41, 45, 47, 48, 51, 57, 60, 62, 64, 65, 66, 69, 70, 71, 72, 73, 75) (1, 2, 4, 68, 7, 10, 74, 15, 16, 17, 21, 25, 30, 32, 33, 35, 36, 38, 40, 42, 43, 46, 49, 50, 55, 56, 58, 61)

Πίνακας 5.8: Κοινότητες *Club 1* ανά μήνα με τη μέθοδο *Girvan-Newman*.

Μήνας	Communities	Κόμβοι
1	5	(0)(1)(9, 2)(3, 7, 8, 16, 17, 18, 19, 20, 21, 23)(6)
2	5	(0)(1)(9, 2)(3, 7, 16, 17, 18, 19, 20, 21, 23)(6)
3	5	(0)(1)(9, 2, 10)(3, 7, 8, 16, 17, 18, 19, 20, 21, 23)(24, 6)
4	6	(0)(1)(9, 2, 10), (3, 6, 7, 8, 16, 17, 18, 19, 20, 21, 23)(13, 5)(24)
5	5	(0)(1, 3, 6, 7, 8, 16, 17, 18, 19, 20, 21, 23)(9, 2)(12)(24)
6	5	(0)(1, 3, 6, 7, 8, 12, 16, 17, 18, 19, 20, 21, 23)(9, 2, 10)(24)(25)
7	5	(0)(1, 3, 7, 12, 16, 17, 18, 19, 20, 21, 23)(9, 2)(24, 6)(25)
8	5	(0)(1, 3, 7, 16, 17, 18, 19, 20, 21, 23, 25)(9, 2)(24, 6)(12)
9	5	(0)(1, 3, 7, 8, 12, 16, 17, 18, 19, 20, 21, 23)(9, 2, 10)(4)(24, 6)
10	5	(0)(1, 3, 7, 8, 16, 17, 18, 19, 20, 21, 23, 25)(9, 2)(4)(24, 6)
11	5	(0)(1, 3, 7, 12, 14, 16, 17, 18, 19, 20, 21, 23, 25)(9, 2)(4)(24, 6)
12	5	(0, 1, 3, 4, 6, 7, 12, 14, 16, 17, 18, 19, 20, 21, 23, 25)(9, 2)(13)(15)(24)

Πίνακας 5.9: Κοινότητες *Club 4* ανά μήνα με τη μέθοδο *Girvan-Newman*.

Μήνας	Communities	Κόμβοι
1	5	(0)(2)(3)(4, 7)(8, 6)
2	5	(0)(2)(3)(4, 7)(8, 6)
3	5	(0)(1)(2)(3, 4, 7)(8, 6)
4	4	(0)(1, 2)(3, 4, 7)(8, 6)
5	4	(0)(1, 2)(3, 4, 7)(8, 6)
6	4	(0)(1)(8, 2, 6)(3, 4, 7)
7	4	(0)(1, 2)(3, 4, 7)(8, 6)
8	4	(0)(1)(8, 2)(3, 4, 7)
9	4	(0)(1)(8, 2)(3, 4, 7)
10	4	(0)(8, 1, 2)(3, 4, 7)(5)
11	4	(0)(1)(8, 2)(3, 4, 7)
12	4	(0)(8, 1, 2)(3, 4, 7)(5)

Πίνακας 5.10: Κοινότητες *Club 2* ανά μήνα με τη μέθοδο *Girvan-Newman*.

Μήνας	Communities	Κόμβοι
1	4	(0, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 30, 31, 33, 34, 37, 38, 39, 40, 41, 43, 45, 46, 47, 49, 50, 51, 52, 53, 54, 55, 56)(20)(42)(48)
2	4	(0, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 31, 33, 34, 37, 38, 39, 40, 41, 43, 45, 46, 47, 49, 50, 51, 52, 53, 55, 56)(30)(42)(54)
3	4	(0, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 34, 37, 38, 39, 40, 41, 43, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56)(29)(30)(31)
4	4	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 33, 34, 36, 37, 38, 39, 40, 41, 43, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56)(30)(31)(57)
5	4	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 31, 33, 34, 36, 37, 38, 39, 40, 41, 43, 45, 46, 47, 48, 49, 50, 51, 52, 53, 55, 56)(54, 30)(42)(57)
6	4	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 33, 34, 36, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47, 49, 50, 51, 52, 53, 55, 56)(42)(54)(57)
7	4	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, 34, 36, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55)(35)(42)(57)
8	4	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 32, 33, 34, 36, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 55)(31)(42)(56)
9	4	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 49, 50, 51, 52, 53, 54, 55, 56)(20)(48)(57)
10	4	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 32, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 49, 50, 51, 52, 53, 55, 56)(20)(30)(31)
11	4	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 30, 32, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 49, 50, 51, 52, 53, 55, 56)(20)(31)(54)
12	4	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 30, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 50, 51, 52, 53, 54, 55, 56)(20)(31)(32)

Πίνακας 5.11: *Κοινότητες Club 3 ανά μήνες 1-6 με τη μέθοδο Girvan-Newman.*

Μήνας	Communities	Κόμβοι
1	8	(64, 97, 65, 3, 68, 9, 109, 110, 113, 50, 83, 52, 153)(5, 7, 76, 151, 24, 154, 88, 29)(36, 37, 6, 135, 46, 82, 157)(129, 134, 136, 10, 139, 11, 141, 15, 143, 17, 146, 19, 148, 21, 149, 23, 30, 31, 34, 40, 41, 42, 43, 44, 49, 54, 55, 56, 57, 61, 66, 71, 74, 79, 80, 81, 84, 92, 95, 96, 99, 100, 101, 108, 111, 112, 117, 118, 120, 121, 124, 126)(131, 22)(28, 102)(94, 38)(105, 85)
2	5	(3, 131, 6, 135, 9, 147, 20, 22, 32, 36, 41, 50, 64, 79, 83, 87, 95, 100, 108, 118, 120)(5, 7, 76, 80, 113, 151, 24, 154, 88, 124, 29)(129, 134, 136, 10, 139, 11, 141, 14, 143, 142, 17, 15, 19, 148, 21, 149, 23, 26, 30, 31, 34, 40, 42, 43, 44, 49, 55, 56, 57, 61, 66, 71, 74, 81, 84, 92, 99, 101, 111, 112, 117, 119, 121, 126)(28, 102)(65, 97, 110, 46, 82, 146, 94)
3	6	(5, 7, 76, 24, 154, 88, 29)(96, 36, 37, 6, 135, 9, 79, 113, 54, 91, 125)(129, 131, 134, 136, 10, 139, 11, 141, 14, 143, 142, 17, 15, 19, 148, 21, 149, 23, 151, 147, 22, 30, 31, 32, 34, 40, 41, 42, 43, 44, 49, 50, 55, 56, 57, 61, 64, 71, 74, 80, 81, 83, 84, 92, 95, 100, 101, 108, 111, 112, 117, 118, 119, 120, 121, 124, 126)(28, 102)(65, 97, 110, 46, 82, 146, 94)(105, 85)
4	8	(32, 3, 100, 36, 6, 135, 9, 79, 113, 50, 83, 52, 125, 120, 91, 29)(129, 5, 134, 7, 136, 10, 11, 139, 141, 142, 15, 17, 148, 21, 23, 151, 24, 154, 26, 152, 30, 31, 34, 39, 40, 42, 43, 44, 49, 55, 56, 57, 61, 68, 71, 74, 76, 80, 81, 84, 87, 88, 92, 93, 95, 101, 106, 108, 112, 117, 118, 119, 121, 124, 126)(19, 111, 143)(131, 22)(28, 102)(82, 46)(65, 110, 97, 94)(105, 85)
5	5	(130, 3, 5, 6, 133, 8, 9, 135, 7, 139, 137, 140, 143, 18, 146, 151, 24, 153, 154, 26, 152, 157, 28, 29, 156, 33, 36, 37, 39, 40, 41, 46, 50, 52, 54, 58, 62, 64, 65, 68, 72, 73, 76, 79, 82, 83, 88, 89, 93, 94, 95, 96, 97, 100, 101, 102, 106, 110, 111, 113, 115, 117, 120, 125)(17, 10)(129, 134, 136, 11, 141, 142, 15, 19, 148, 149, 23, 30, 31, 32, 34, 42, 43, 44, 49, 55, 56, 57, 61, 71, 74, 81, 84, 86, 87, 92, 108, 112, 118, 119, 121, 124, 126)(131, 22)(105, 85)
6	5	(2, 3, 4, 5, 135, 7, 16, 151, 152, 24, 154, 28, 157, 33, 36, 37, 54, 62, 64, 68, 72, 80, 88, 95, 102, 103, 106, 113, 123, 124, 125)(32, 100, 6, 9, 41, 76, 79, 50, 83, 52, 120)(129, 131, 134, 136, 10, 139, 11, 141, 14, 142, 15, 147, 148, 22, 23, 30, 31, 34, 39, 40, 42, 43, 44, 49, 55, 56, 57, 58, 61, 71, 81, 84, 92, 93, 101, 108, 109, 112, 118, 119, 121, 126)(65, 97, 74, 46, 111, 110, 143, 82, 146)(105, 85)

Πίνακας 5.12: Κοινότητες *Club 3* ανά μήνες 7-12 με τη μέθοδο *Girvan-Newman*.

Μήνας	Communities	Κόμβοι
7	4	(3, 134, 136, 10, 139, 151, 31, 32, 34, 40, 43, 44, 49, 50, 51, 56, 61, 71, 79, 80, 81, 83, 84, 92, 95, 100, 101, 108, 113, 118, 119, 124)(68, 5, 7, 24, 154, 88, 29, 63)(64, 131, 36, 37, 102, 6, 135, 9, 125, 22, 54, 120, 91, 28, 157)(129, 11, 141, 142, 15, 143, 148, 23, 30, 42, 46, 55, 57, 65, 74, 82, 97, 110, 111, 112, 121, 126)
8	4	(2, 3, 6, 9, 10, 11, 15, 18, 19, 23, 26, 30, 31, 32, 33, 34, 36, 37, 39, 40, 41, 42, 43, 44, 46, 49, 50, 51, 54, 55, 56, 57, 61, 63, 64, 65, 69, 70, 71, 72, 74, 79, 80, 81, 82, 83, 84, 88, 92, 95, 97, 100, 101, 103, 105, 108, 110, 111, 112, 113, 115, 117, 118, 120, 121, 123, 124, 125, 126, 129, 132, 134, 135, 136, 139, 141, 142, 143, 146, 148, 150, 151, 153, 154, 157)(68, 5, 133, 7, 8, 24, 58, 29)(131, 22)(138, 28, 102)
9	5	(0, 129, 3, 134, 136, 10, 139, 11, 141, 142, 15, 23, 26, 155, 30, 31, 32, 34, 39, 40, 42, 43, 44, 49, 51, 55, 56, 57, 61, 69, 71, 81, 84, 92, 101, 109, 112, 119, 121, 123, 126)(5, 6, 7, 135, 9, 133, 138, 12, 145, 147, 148, 22, 151, 24, 152, 154, 150, 28, 29, 33, 36, 41, 50, 54, 58, 62, 64, 66, 68, 70, 72, 73, 79, 88, 89, 93, 95, 96, 100, 102, 103, 108, 113, 118, 120, 125)(98, 13)(65, 97, 74, 46, 111, 143, 110, 146, 19, 82)(80, 124)
10	4	(0, 129, 131, 134, 136, 9, 10, 139, 11, 141, 143, 15, 146, 147, 19, 21, 149, 23, 148, 22, 151, 155, 28, 30, 31, 158, 33, 34, 32, 40, 41, 43, 44, 49, 50, 51, 55, 56, 57, 59, 61, 64, 65, 71, 79, 81, 84, 92, 95, 100, 101, 102, 103, 108, 111, 112, 117, 118, 119, 120, 121, 123, 124)(68, 5, 7, 145, 24, 154, 88, 29)(97, 74, 142, 46, 110, 82, 126)(153, 125, 69, 135)
11	5	(0, 134, 136, 9, 10, 139, 11, 141, 142, 148, 21, 149, 23, 22, 151, 155, 28, 30, 31, 34, 40, 41, 43, 44, 49, 50, 51, 55, 56, 57, 59, 61, 69, 70, 71, 76, 79, 80, 81, 84, 92, 93, 95, 100, 101, 102, 108, 112, 113, 117, 118, 119, 120, 121, 124, 126)(68, 5, 7, 78, 48, 145, 24, 154, 123, 29, 158)(91, 125, 6, 135)(98, 13)(65, 97, 74, 46, 111, 143, 17, 82, 19, 110)
12	4	(0, 129, 5, 134, 7, 136, 137, 10, 139, 11, 141, 145, 149, 23, 24, 154, 155, 30, 31, 34, 40, 41, 43, 44, 49, 50, 51, 53, 55, 56, 57, 61, 68, 69, 71, 78, 79, 80, 84, 92, 95, 100, 101, 103, 112, 117, 119, 122, 123)(3, 132, 131, 135, 8, 9, 146, 147, 148, 22, 26, 157, 32, 33, 35, 58, 72, 87, 96, 107, 108, 114, 118, 120, 121, 125)(65, 97, 74, 109, 46, 143, 111, 17, 48, 19, 110, 142, 82, 126)(102, 28, 113, 151, 124)

Πίνακας 5.13: *Κοινότητες Club 5 ανά μήνες 1-8 με τη μέθοδο Girvan-Newman.*

Μήνας	Communities	Κόμβοι
1	6	(0, 3, 5, 6, 7, 8, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 25, 26, 27, 28, 30, 34, 35, 36, 37, 39, 41, 43, 45, 46, 47, 48, 49, 50, 51, 55, 56, 57, 58, 60, 61, 62, 63, 64, 65, 67, 72, 73, 74, 75)(1, 68)(10)(32)(59, 38)(70)
2	7	(0, 3, 5, 6, 8, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 25, 26, 27, 30, 34, 35, 36, 37, 39, 41, 43, 45, 46, 47, 48, 50, 51, 56, 57, 58, 60, 61, 62, 64, 65, 67, 70, 72, 73, 75)(1, 68)(10, 7)(28)(63, 31)(59, 38)(49)
3	7	(0, 3, 5, 6, 8, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 27, 34, 35, 36, 39, 41, 43, 45, 46, 47, 48, 49, 50, 51, 56, 57, 60, 61, 62, 64, 65, 70, 72, 73, 75)(1, 68)(10, 30, 7)(24, 66)(37)(59, 38)(58)
4	7	(0, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 27, 30, 34, 35, 36, 37, 39, 41, 42, 43, 45, 46, 47, 48, 49, 50, 51, 56, 57, 58, 60, 61, 62, 64, 65, 67, 69, 70, 72, 73, 75)(1, 68)(24, 66)(28)(29, 71)(32)(59, 38)
5	7	(0, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 27, 30, 34, 35, 36, 37, 39, 41, 42, 43, 45, 46, 47, 48, 49, 50, 51, 56, 57, 58, 60, 61, 62, 64, 65, 70, 71, 72, 73, 75)(1, 68)(24, 66)(28)(29)(32)(59, 38)
6	6	(0, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 26, 27, 29, 30, 34, 35, 36, 37, 39, 41, 42, 43, 45, 46, 47, 48, 49, 50, 51, 56, 57, 60, 61, 62, 64, 65, 66, 70, 71, 72, 73, 75)(1, 68)(28)(32)(59, 38)(58)
7	6	(0, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 26, 27, 30, 34, 35, 36, 39, 41, 42, 43, 45, 46, 47, 48, 49, 50, 51, 56, 57, 58, 60, 61, 62, 64, 65, 66, 69, 70, 71, 72, 75)(1, 68)(29)(37)(59, 38)(54)
8	6	(0, 3, 5, 6, 7, 8, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 26, 27, 30, 34, 36, 39, 41, 42, 43, 45, 46, 47, 48, 49, 50, 51, 54, 56, 57, 58, 60, 61, 62, 64, 65, 66, 67, 69, 70, 71, 72, 75)(1, 68)(28)(35)(37)(59, 38)

Πίνακας 5.14: Κοινότητες *Club 5* ανά μήνες 9-12 με τη μέθοδο Girvan-Newman.

Μήνας	Communities	Κόμβοι
9	5	(0, 3, 5, 7, 8, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 26, 27, 30, 34, 35, 36, 37, 39, 41, 42, 43, 45, 46, 47, 48, 49, 50, 51, 54, 56, 57, 58, 60, 61, 62, 64, 65, 66, 69, 70, 71, 72, 73, 75)(1, 68)(4)(10, 28)(55)
10	5	(0, 3, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 26, 27, 28, 29, 30, 34, 35, 36, 37, 39, 40, 41, 42, 43, 45, 46, 47, 48, 49, 50, 51, 54, 56, 57, 58, 60, 61, 62, 64, 65, 66, 69, 70, 71, 72, 73, 75)(1, 68)(2, 44)(32)(59, 38)
11	5	(0, 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 27, 29, 30, 32, 34, 35, 36, 37, 39, 40, 41, 42, 43, 45, 46, 47, 48, 49, 50, 51, 54, 56, 57, 58, 60, 61, 62, 64, 65, 66, 70, 71, 72, 75)(1, 68)(2)(28)(73)
12	4	(0, 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 27, 28, 29, 30, 32, 34, 35, 36, 37, 39, 40, 41, 42, 43, 45, 46, 47, 48, 49, 50, 51, 56, 57, 58, 60, 61, 62, 64, 65, 66, 68, 69, 70, 71, 72, 73, 74, 75)(33)(38)(55)

Πίνακας 5.15: Κοινότητες *Club 1* ανά μήνα με τη μέθοδο *Spectral Clustering*.

Μήνας	Communities	Κόμβοι
1	4	(1, 16, 18, 19, 20, 21)(2, 9)(0, 3, 7, 8, 17, 23)(6)
2	4	(0, 3, 7, 23)(2, 9)(6, 17, 18, 19, 21)(1, 16, 20)
3	4	(0, 1, 7, 8, 16, 18, 19, 20, 21)(2, 9, 10)(6, 24)(3, 17, 23)
4	4	(0, 1, 3, 7, 8, 16, 17, 18, 19, 20, 21, 23)(2, 9, 10)(5, 13)(6, 24)
5	4	(0, 1, 7, 8, 16, 18, 19, 20, 21)(2, 9)(6, 24)(3, 12, 17, 23)
6	4	(3, 12, 17, 23)(2, 9, 10)(0, 1, 7, 8, 16, 18, 19, 20, 21, 25)(6, 24)
7	4	(2, 9)(0, 1, 7, 16, 19, 20, 21, 25)(6, 24)(3, 12, 17, 18, 23)
8	4	(0, 1, 7, 16, 18, 19, 20, 21)(2, 9)(6, 24)(3, 12, 17, 23, 25)
9	4	(0, 1, 4, 7, 8, 16, 18, 19, 20, 21)(2, 9, 10)(6, 24)(3, 12, 17, 23)
10	4	(3, 4, 8, 17, 18, 19, 21, 23)(2, 9)(6, 24)(0, 1, 7, 16, 20, 25)
11	4	(2, 9)(0, 1, 3, 4, 7, 14, 16, 18, 19, 20, 21, 25)(6, 24)(12, 17, 23)
12	4	(4, 6, 13, 15, 24, 25)(0, 1, 3, 7, 14, 16, 17, 18, 19, 20, 21, 23)(2, 9)(12)

Πίνακας 5.16: Κοινότητες *Club 4* ανά μήνα με τη μέθοδο *Spectral Clustering*.

Μήνας	Communities	Κόμβοι
1	4	(2, 6, 8)(3, 4)(0)(7)
2	4	(2, 6, 8)(3, 4)(0)(7)
3	4	(0, 4, 7)(1)(2, 6, 8)(3)
4	4	(3, 4, 7)(6, 8)(2)(0, 1)
5	4	(3, 4, 7)(6, 8)(2)(0, 1)
6	4	(3, 4, 7)(6, 8)(1, 2)(0)
7	4	(2, 8)(0, 1)(6)(3, 4, 7)
8	4	(4)(0, 1)(2, 8)(3, 7))
9	4	(7)(0, 1)(2, 8)(3, 4)
10	4	(3, 4)(0, 1, 7)(2, 8)(5)
11	4	(3, 4)(2, 8)(1, 7)(0)
12	4	(3, 4, 7)(0, 1)(2, 8)(5)

Πίνακας 5.17: Κοινότητες *Club 2* ανά μήνα με τη μέθοδο *Spectral Clustering*.

Μήνας	Communities	Κόμβοι
1	4	(0, 1, 2, 3, 4, 7, 8, 9, 10, 11, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 25, 26, 27, 28, 33, 34, 37, 39, 40, 41, 43, 45, 46, 50, 51, 52, 53, 55, 56)(12, 24, 49)(38, 47, 48)(6, 17, 30, 31, 42, 54)
2	4	(0, 3, 4, 7, 8, 9, 10, 14, 18, 19, 21, 23, 25, 26, 27, 28, 33, 34, 37, 39, 40, 41, 43, 45, 46, 50, 51, 52, 55, 56)(6, 17, 30, 31, 42, 54)(2, 12, 20, 24, 47, 49)(1, 11, 13, 15, 16, 22, 38, 53)
3	4	(0, 3, 4, 7, 9, 10, 14, 18, 19, 20, 21, 23, 27, 28, 29, 34, 37, 39, 40, 41, 43, 45, 46, 51, 52, 55, 56)(1, 2, 6, 8, 11, 13, 15, 16, 17, 22, 26, 30, 31, 38, 48, 53, 54)(25, 50)(12, 24, 47, 49)
4	4	(1, 2, 5, 9, 11, 13, 15, 16, 22, 38, 48, 57)(0, 7, 18, 20, 25, 26, 27, 28, 36, 40, 41, 45, 46, 50, 51, 52, 53, 55, 56)(3, 4, 6, 8, 10, 14, 17, 19, 21, 23, 30, 31, 33, 34, 37, 39, 43, 54)(12, 24, 47, 49)
5	4	(2, 6, 12, 17, 21, 24, 31, 34, 36, 39, 47, 49, 53)(0, 3, 4, 7, 8, 9, 10, 14, 18, 19, 23, 25, 26, 27, 28, 33, 37, 40, 41, 42, 43, 45, 46, 50, 51, 52, 55, 56)(30, 54)(1, 5, 11, 13, 15, 16, 20, 22, 38, 48, 57)
6	4	(8, 10, 17, 30, 37, 39, 42, 53)(1, 5, 22)(2, 9, 11, 12, 13, 15, 16, 20, 21, 24, 34, 36, 38, 44, 47, 49, 54, 57)(0, 3, 4, 6, 7, 14, 18, 19, 23, 25, 26, 27, 28, 33, 40, 41, 43, 45, 46, 50, 51, 52, 55, 56)
7	4	(0, 3, 4, 8, 10, 14, 21, 23, 25, 26, 27, 28, 33, 34, 35, 36, 37, 39, 40, 41, 42, 43, 45, 46, 50, 51, 52, 53, 55)(30, 54)(1, 2, 5, 6, 7, 9, 11, 12, 13, 15, 16, 17, 18, 19, 20, 22, 31, 32, 38, 48, 49, 57)(24, 44, 47)
8	4	(1, 2, 5, 6, 7, 11, 12, 13, 15, 16, 20, 22, 32, 38, 48, 49) (17, 30, 31, 42)(0, 3, 4, 8, 9, 10, 14, 18, 19, 21, 23, 25, 26, 27, 28, 33, 34, 36, 37, 39, 40, 41, 43, 45, 46, 50, 51, 52, 53, 55, 56)(24, 44, 47)
9	4	(17, 30, 31)(0, 3, 4, 6, 7, 8, 9, 10, 14, 18, 19, 20, 23, 25, 27, 28, 33, 34, 37, 38, 39, 40, 41, 42, 43, 45, 46, 48, 50, 51, 52, 53, 54, 55, 56, 57)(2, 11, 12, 13, 15, 21, 24, 35, 36, 44, 47, 49)(1, 5, 16, 22, 26, 32)
10	4	(0, 3, 7, 8, 9, 18, 19, 20, 23, 25, 26, 27, 28, 37, 40, 41, 43, 45, 50, 51, 52, 55, 56)(2, 4, 10, 14, 21, 24, 34, 36, 44, 46, 47)(1, 5, 6, 11, 12, 13, 15, 16, 22, 32, 35, 38, 49)(17, 30, 31, 39, 42, 53)
11	4	(0, 3, 4, 6, 7, 8, 9, 10, 14, 17, 18, 19, 23, 25, 26, 27, 28, 37, 39, 40, 41, 42, 43, 45, 46, 50, 51, 52, 53, 55, 56)(30, 31, 54)(1, 5, 11, 12, 13, 15, 16, 20, 22, 32, 35, 38)(2, 21, 24, 34, 36, 44, 47, 49)
12	4	(0, 3, 4, 7, 8, 9, 10, 14, 18, 19, 21, 23, 25, 26, 27, 28, 34, 36, 37, 39, 40, 41, 42, 43, 45, 46, 50, 51, 52, 53, 54, 55, 56)(31)(1, 2, 5, 12, 13, 15, 16, 20, 22, 24, 35, 38, 44, 47)(6, 17, 30, 32)

Πίνακας 5.18: Κοινότητες *Club 3* ανά μήνες 1-6 με τη μέθοδο *Spectral Clustering*.

Μήνας	Communities	Κόμβοι
1	4	(3, 5, 6, 7, 9, 10, 11, 15, 17, 19, 21, 23, 24, 29, 30, 31, 34, 36, 37, 38, 40, 41, 42, 43, 44, 46, 49, 50, 52, 54, 55, 56, 57, 61, 64, 65, 66, 68, 71, 74, 76, 79, 80, 81, 82, 83, 84, 88, 92, 94, 95, 96, 97, 99, 100, 101, 108, 109, 110, 111, 112, 113, 117, 118, 120, 121, 124, 126, 129, 134, 135, 136, 139, 141, 143, 146, 148, 149, 151, 153, 154, 157)(85, 105)(28, 102)(22, 131)
2	4	(3, 6, 9, 10, 11, 14, 15, 17, 19, 20, 21, 22, 23, 26, 30, 31, 32, 34, 36, 40, 41, 42, 43, 44, 49, 50, 55, 56, 57, 61, 64, 66, 71, 74, 79, 81, 83, 84, 87, 92, 95, 99, 100, 101, 108, 111, 112, 117, 118, 119, 120, 121, 126, 129, 131, 134, 135, 136, 139, 141, 142, 143, 146, 147, 148, 149)(46, 65, 82, 94, 97, 110)(28, 102)(5, 7, 24, 29, 76, 80, 88, 113, 124, 151, 154)
3	4	(65, 94, 97, 110, 146)(6, 9, 10, 11, 14, 15, 17, 19, 21, 22, 23, 30, 31, 32, 34, 36, 37, 40, 41, 42, 43, 44, 46, 49, 50, 54, 55, 56, 57, 61, 64, 71, 74, 79, 80, 81, 82, 83, 84, 91, 92, 95, 96, 100, 101, 108, 111, 112, 113, 117, 118, 119, 120, 121, 124, 125, 126, 129, 131, 134, 135, 136, 139, 141, 142, 143, 147, 148, 149, 151)(28, 85, 102, 105)(5, 7, 24, 29, 76, 88, 154)
4	4	(46, 82)(3, 5, 6, 7, 9, 10, 11, 15, 17, 19, 21, 23, 24, 26, 29, 30, 31, 32, 34, 36, 39, 40, 42, 43, 44, 49, 50, 52, 55, 56, 57, 61, 65, 68, 71, 74, 76, 79, 80, 81, 83, 84, 85, 87, 88, 91, 92, 93, 94, 95, 97, 100, 101, 105, 106, 108, 110, 111, 112, 113, 117, 118, 119, 120, 121, 124, 125, 126, 129, 134, 135, 136, 139, 141, 142, 143, 148, 151, 152, 154)(22, 131)(28, 102)
5	4	(3, 5, 6, 7, 8, 9, 18, 22, 24, 26, 28, 29, 33, 36, 37, 39, 41, 46, 52, 54, 58, 62, 64, 65, 68, 72, 73, 79, 82, 88, 89, 93, 94, 95, 96, 97, 101, 102, 106, 110, 111, 113, 115, 125, 130, 131, 133, 135, 137, 140, 143, 146, 151, 152, 153, 154, 156, 157)(85, 105)(11, 15, 30, 32, 42, 49, 50, 55, 57, 74, 76, 83, 86, 87, 100, 108, 112, 118, 120, 126, 129, 141, 142, 148)(10, 17, 19, 23, 31, 34, 40, 43, 44, 56, 61, 71, 81, 84, 92, 117, 119, 121, 124, 134, 136, 139, 149)
6	4	(3, 4, 9, 10, 11, 14, 15, 16, 22, 23, 28, 30, 31, 32, 34, 39, 40, 41, 42, 43, 44, 49, 50, 52, 55, 56, 57, 58, 61, 62, 71, 72, 76, 80, 81, 83, 84, 92, 93, 95, 100, 101, 102, 103, 106, 108, 109, 112, 113, 118, 119, 121, 123, 124, 126, 129, 131, 134, 136, 139, 141, 142, 147, 148, 151, 152)(85, 105)(5, 7, 24, 46, 65, 68, 74, 82, 88, 97, 110, 111, 143, 146, 154)(2, 6, 33, 36, 37, 54, 64, 79, 120, 125, 135, 157)

Πίνακας 5.19: Κοινότητες *Club 3* ανά μήνες 7-12 με τη μέθοδο *Spectral Clustering*.

Μήνας	Communities	Κόμβοι
7	4	(3, 10, 11, 15, 23, 30, 31, 34, 40, 42, 43, 44, 49, 50, 51, 55, 56, 57, 61, 71, 81, 84, 92, 95, 100, 101, 112, 119, 121, 124, 126, 129, 134, 136, 139, 141, 142, 148)(5, 7, 24, 29, 63, 68, 88, 154)(46, 65, 74, 82, 97, 110, 111, 143)(6, 9, 22, 28, 32, 36, 37, 54, 64, 79, 80, 83, 91, 102, 108, 113, 118, 120, 125, 131, 135, 151, 157)
8	4	(10, 11, 15, 22, 23, 26, 30, 31, 34, 40, 42, 43, 44, 46, 49, 51, 56, 57, 61, 65, 71, 74, 81, 82, 84, 92, 95, 97, 101, 108, 110, 113, 117, 121, 123, 124, 126, 129, 131, 134, 136, 139, 141, 142, 146)(28, 102, 138)(19, 111, 143)(2, 3, 5, 6, 7, 8, 9, 18, 24, 29, 32, 33, 36, 37, 39, 41, 50, 54, 55, 58, 63, 64, 68, 69, 70, 72, 79, 80, 83, 88, 100, 103, 105, 112, 115, 118, 120, 125, 132, 133, 135, 148, 150, 151, 153, 154, 157)
9	4	(3, 5, 6, 7, 9, 11, 12, 15, 22, 24, 28, 29, 30, 33, 36, 39, 41, 42, 50, 51, 54, 55, 58, 62, 64, 66, 68, 69, 70, 72, 73, 79, 80, 88, 89, 93, 95, 96, 100, 102, 103, 108, 112, 113, 118, 120, 124, 125, 129, 133, 135, 138, 141, 145, 147, 148, 150, 151, 152, 154)(13, 98)(0, 10, 23, 26, 31, 32, 34, 40, 43, 44, 49, 56, 57, 61, 71, 81, 84, 92, 101, 109, 119, 121, 123, 126, 134, 136, 139, 142, 155)(19, 46, 65, 74, 82, 97, 110, 111, 143, 146)
10	4	(46, 82, 97, 110)(0, 9, 10, 11, 15, 19, 21, 22, 23, 28, 30, 31, 32, 33, 34, 40, 41, 43, 44, 49, 50, 51, 55, 56, 57, 59, 61, 64, 65, 71, 74, 79, 81, 84, 92, 95, 100, 101, 102, 103, 108, 111, 112, 117, 118, 119, 120, 121, 123, 124, 126, 129, 131, 134, 136, 139, 141, 142, 143, 146, 147, 148, 149, 151, 155, 158)(69, 125, 135, 153)(5, 7, 24, 29, 68, 88, 145, 154)
11	4	((0, 9, 10, 11, 17, 19, 21, 22, 23, 28, 30, 31, 34, 40, 41, 43, 44, 46, 49, 50, 51, 55, 56, 57, 59, 61, 65, 69, 70, 71, 74, 76, 79, 80, 81, 82, 84, 92, 93, 95, 97, 100, 101, 102, 108, 110, 111, 112, 113, 117, 118, 119, 120, 121, 124, 126, 134, 136, 139, 141, 142, 143, 148, 149, 151, 155)(13, 98)(5, 7, 24, 29, 48, 68, 78, 123, 145, 154, 158)(6, 91, 125, 135)
12	4	(0, 3, 7, 8, 9, 10, 11, 22, 23, 26, 30, 31, 32, 33, 34, 35, 40, 41, 43, 44, 48, 49, 50, 51, 55, 56, 57, 58, 61, 69, 71, 72, 78, 79, 80, 84, 87, 92, 95, 96, 100, 101, 103, 107, 108, 109, 112, 113, 114, 117, 118, 119, 120, 121, 123, 125, 126, 129, 131, 132, 134, 135, 136, 137, 139, 141, 145, 146, 147, 148, 149, 154, 155, 157)(53, 122)(5, 24, 28, 68, 102, 124, 151)(17, 19, 46, 65, 74, 82, 97, 110, 111, 142, 143)

Πίνακας 5.20: Κοινότητες *Club 5* ανά μήνες 1-8 με τη μέθοδο *Spectral Clustering*.

Μήνας	Communities	Κόμβοι
1	4	(7, 10, 15, 16, 21, 25, 28, 30, 32, 35, 36, 43, 46, 49, 50, 56, 58, 61, 63, 74)(1, 68)(38, 59)(0, 3, 5, 6, 8, 11, 12, 13, 14, 17, 18, 19, 20, 26, 27, 34, 37, 39, 41, 45, 47, 48, 51, 55, 57, 60, 62, 64, 65, 67, 70, 72, 73, 75)
2	4	(0, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 25, 26, 27, 28, 30, 34, 35, 36, 37, 39, 41, 43, 45, 46, 47, 48, 49, 50, 51, 56, 57, 58, 60, 61, 62, 64, 65, 67, 70, 72, 73, 75)(38, 59)(1, 68)(31, 63)
3	4	(0, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 27, 30, 34, 35, 36, 37, 39, 41, 43, 45, 46, 47, 48, 49, 50, 51, 56, 57, 58, 60, 61, 62, 64, 65, 70, 72, 73, 75)(38, 59)(24, 66)(1, 68)
4	4	(0, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 27, 28, 29, 30, 32, 34, 35, 36, 37, 39, 41, 42, 43, 45, 46, 47, 48, 49, 50, 51, 56, 57, 58, 60, 61, 62, 64, 65, 67, 69, 70, 71, 72, 73, 75)(1, 68)(24, 66)(38, 59)
5	4	(0, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 27, 28, 29, 30, 32, 34, 35, 36, 37, 39, 41, 42, 43, 45, 46, 47, 48, 49, 50, 51, 56, 57, 58, 60, 61, 62, 64, 65, 70, 71, 72, 73, 75)(24, 66)(1, 68)(38, 59)
6	4	(0, 3, 5, 6, 8, 11, 12, 13, 14, 18, 19, 20, 22, 24, 26, 27, 29, 34, 37, 39, 41, 45, 47, 48, 51, 57, 60, 62, 64, 65, 66, 70, 71, 72, 73, 75)(38, 59)(1, 68)(7, 10, 15, 16, 17, 21, 25, 28, 30, 32, 35, 36, 42, 43, 46, 49, 50, 56, 58, 61)
7	4	(7, 10, 15, 16, 21, 25, 30, 35, 36, 42, 43, 46, 49, 50, 56, 58, 61)(1, 68)(38, 59)(0, 3, 5, 6, 8, 11, 12, 13, 14, 17, 18, 19, 20, 22, 24, 26, 27, 29, 34, 37, 39, 41, 45, 47, 48, 51, 54, 57, 60, 62, 64, 65, 66, 69, 70, 71, 72, 75)
8	4	(7, 15, 16, 21, 25, 28, 30, 35, 36, 42, 43, 46, 49, 50, 56, 58, 61)(38, 59)(1, 68)(0, 3, 5, 6, 8, 11, 12, 13, 14, 17, 18, 19, 20, 22, 24, 26, 27, 34, 37, 39, 41, 45, 47, 48, 51, 54, 57, 60, 62, 64, 65, 66, 67, 69, 70, 71, 72, 75)

Πίνακας 5.21: Κοινότητες *Club 5* ανά μήνες 9-12 με τη μέθοδο *Spectral Clustering*.

Μήνας	Communities	Κόμβοι
9	4	(0, 3, 5, 8, 11, 12, 13, 14, 17, 18, 19, 20, 22, 24, 26, 27, 34, 37, 39, 41, 45, 47, 48, 51, 54, 57, 60, 62, 64, 65, 66, 69, 70, 71, 72, 73, 75)(1, 68)(10, 28)(4, 7, 15, 16, 21, 25, 30, 35, 36, 42, 43, 46, 49, 50, 55, 56, 58, 61)
10	4	(0, 3, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 26, 27, 28, 29, 30, 32, 34, 35, 36, 37, 39, 40, 41, 42, 43, 45, 46, 47, 48, 49, 50, 51, 54, 56, 57, 58, 60, 61, 62, 64, 65, 66, 69, 70, 71, 72, 73, 75)(1, 68)(2, 44)(38, 59)
11	4	(2, 7, 15, 16, 17, 28, 37, 58, 73)(1, 68)(0, 3, 5, 6, 8, 11, 12, 13, 14, 18, 19, 20, 22, 26, 27, 29, 34, 39, 41, 45, 47, 48, 51, 57, 60, 62, 64, 65, 66, 70, 71, 72, 75)(4, 21, 25, 30, 32, 35, 36, 40, 42, 43, 46, 49, 50, 54, 56, 61)
12	4	(0, 3, 4, 5, 6, 8, 11, 12, 13, 14, 18, 19, 20, 22, 26, 27, 29, 34, 37, 39, 41, 45, 47, 48, 51, 57, 60, 62, 64, 65, 66, 69, 70, 71, 72, 73, 75)(1, 68)(7, 10, 28)(2, 15, 16, 17, 21, 25, 30, 32, 33, 35, 36, 38, 40, 42, 43, 46, 49, 50, 55, 56, 58, 61, 74)

Αρχικά, μεταξύ των μεθόδων Modularity Maximization και Girvan Newman, όπου δεν έχουν προκαθορισμένο αριθμό κοινοτήτων, παρατηρείται ότι κατά μέσο όρο ο δεύτερος εντοπίζει περισσότερες κοινότητες, όμως με μικρή απόκλιση 1 ή 2 περισσότερες από τον πρώτο. Πιο συγκεκριμένα, το club 1 με τον αλγόριθμο Modularity Maximization έχει, κατά μέσο όρο, 4 κοινότητες, ενώ με τον αλγόριθμο Girvan Newman, έχει μέσο όρο 5 κοινότητες. Το club 2, έχει αντίστοιχα μέσο όρο 3 και 4 κοινότητες ανά αλγόριθμο, το club 4 έχει 2 και 4 κοινότητες μέσο όρο και το club 5, έχει μέσο όρο με τον Modularity Maximization 5 και με τον Girvan Newman 6. Στο club 3, ωστόσο ο αλγόριθμος Modularity Maximization, προσδιορίζει 8 κοινότητες, κατά μέσο όρο, με διαφορά 3 μονάδες από τον Girvan Newman, που εντοπίζει 5 κοινότητες.

Στη συνέχεια, με μια απλή παρατήρηση των πινάκων, ανά μήνα κάθε αλγορίθμου, υπάρχουν αλλαγές στην σύσταση και τον αριθμό των κοινοτήτων, ωστόσο παρατηρούνται κάποιοι βασικοί κόμβοι, που αποτελούν κάποιες ομάδες, οι οποίοι παραμένουν σε σταθερή θέση. Σε κάποιες

περιπτώσεις, συμβαίνει ο αριθμός των κοινοτήτων, να μην αλλάξει ανά μήνα, αλλά οι κόμβοι, που αποτελούν τις κοινότητες αυτές, να είναι μεταβαλλόμενοι. Αντίθετα δεν συμβαίνει το ίδιο, με την αλλαγή των μεθόδων, αφού ανά μέθοδο δεν εντοπίζονται πανομοιότυπες κοινότητες, δηλαδή που να αποτελούνται από κοινούς κόμβους. Ακολουθεί το Σχήμα 5.28, όπου παρουσιάζεται ενδεικτικά ένα παράδειγμα με τις κοινότητες που εμφανίζονται ανά μέθοδο. Στο πάνω μέρος των γραφημάτων, αναγράφεται ο αλγόριθμος και στο πάνω μέρος αριστερά το club, όλα τα παραδείγματα είναι του μήνα 9. Οι διαφορετικές κοινότητες, είναι χρωματισμένες με διαφορετικά χρώματα, η αρίθμηση των κόμβων έχει αφαιρεθεί για λόγους οπτικοποίησης, ωστόσο όλη η πληροφορία παρέχεται αναλυτικά στους πίνακες 5.1 - 5.21.

Με παρατήρηση των μεθόδων προκύπτει, ότι με τον αλγόριθμο Modularity Maximization, μόνο στο club 3 υπάρχει μεγαλύτερος αριθμός κοινοτήτων, ανά μήνα, σε σύγκριση με τα αποτελέσματα του Girvan Newman στα υπόλοιπα clubs. Γενικά, με την μέθοδο αυτή, δεν παρατηρούνται μεγάλες αλλαγές ανά μήνα, σε ότι αφορά τον αριθμό κοινοτήτων, ωστόσο οι κόμβοι που αποτελούν κάθε κοινότητα αλλάζουν δυναμικά.

Για την χρήση του αλγορίθμου Girvan Newman, έχει τεθεί μια σταθερά $k = 3$, που αντιπροσωπεύει τον αριθμό επαναλήψεων του αλγορίθμου και επιλέχθηκε με βάση το κριτήριο modularity, του οποίου αναφορά γίνεται στο Υποκεφάλαιο 2.3.6.2. Με τον αλγόριθμο αυτό, στις περισσότερες περιπτώσεις γίνεται η διαίρεση των γραφημάτων σε περισσότερες ομάδες, σε σύγκριση με τον αλγόριθμο Modularity Maximization. Αρκετές από τις κοινότητες, εμφανίζονται να αποτελούνται από 1 ή 2 κόμβους. Στα clubs 1, 2 και 4, δεν παρατηρούνται μεγάλες αλλαγές στον αριθμό κοινοτήτων ανά μήνα.

Για τον αλγόριθμο Spectral Clustering, ο αριθμός των κοινοτήτων που θα εμφανίζονται, για κάθε μήνα, είναι προκαθορισμένος και ίσος με 4. Το νούμερο αυτό προέκυψε, από τον μέσο όρο των κοινοτήτων, που εντοπίζονται από τις προηγούμενες μεθόδους. Παρόλο, που ο αριθμός αυτός είναι σταθερός, οι κόμβοι που αποτελούν τις κοινότητες αλλάζουν δυναμικά. Ακόμη από τους πίνακες παρατηρείται, κυρίως στο club 3 και στο club 5, ότι εμφανίζονται κάποιες πολύ μικρές κοινότητες, παρά τον μεγάλο αριθμό των συγκεκριμένων clubs, που αποτελούνται από 2 κόμβους.

Επιπλέον, εκτός από την εύρεση κοινοτήτων, σε κάθε εκτέλεση των παραπάνω αλγορίθμων, χρησιμοποιήθηκε μια μεταβλητή, στην οποία αποθηκεύεται ο χρόνος που χρειάζεται σε κάθε αλγόριθμο να βρει τις κοινότητες ανά club. Ο χρόνος εκτέλεσης κάθε αλγορίθμου, φαίνεται στον πίνακα 5.22, όπου στην πρώτη στήλη αναγράφεται ο εκάστοτε αλγόριθμος και στις επόμενες στήλες ο χρόνος εκτέλεσης ανά club. Μεγαλύτερος χρόνος εκτέλεσης, παρατηρεί-

ται γενικότερα στα μεγαλύτερης έκτασης clubs (2, 3 και 5), ωστόσο σημαντικός φαίνεται να είναι ο ρόλος των isolates. Το club 2 ($n = 58$) και το club 5 ($n = 76$), παρουσιάζουν πολύ κοντινούς χρόνους εκτέλεσης και στις 3 περιπτώσεις αλγορίθμων, παρόλο που το club 2 είναι μικρότερο σε έκταση, το club 5 έχει περισσότερους isolates από το club 2 (club 2: 1 - 8 isolates ανά μήνα, club 5: 11 - 21 isolates ανά μήνα). Αναφορικά με τους αλγορίθμους, όπως παρατηρείται, ο Modularity Maximization είναι και ο πιο γρήγορος, με μικρή διαφορά από τον Girvan Newman. Ο αλγόριθμος Spectral Clustering, είναι σημαντικά πιο αργός από τους άλλους δύο. Αυτό συμβαίνει, λόγω της πολυπλοκότητας των αλγορίθμων, αφού οι δύο πρώτοι είναι της τάξης $O(V * E)$, όπου V ο αριθμός των κόμβων και E ο αριθμός των ακμών, ενώ ο τελευταίος είναι της τάξης $O(n^3)$, λόγω του πίνακα ομοιότητας που έχει να δημιουργήσει [34] [62] [42].

Τέλος, μαζί με τον απαιτούμενο χρόνο εκτέλεσης, στον πίνακα 5.23, παρουσιάζεται ο μέσος όρος modularity ανά club. Τα αποτελέσματα αυτά, προκύπτουν από το modularity των κοινοτήτων που έχει κάθε club ανά μήνα, ανά τις διαφορετικές μεθόδους που χρησιμοποιήθηκαν. Το μέτρο αυτό, αξιολογεί πόσο "καλή" είναι κάθε φορά η διαίρεση του δικτύου. Οι τιμές του, κυμαίνονται στο διάστημα $[-1, 1]$ και μια θετική τιμή κοντά στο 0.3, είναι ένας καλός δείκτης της ύπαρξης ουσιαστικών κοινοτήτων. Έτσι, στον πίνακα αυτό, το Modularity Maximization εμφανίζει τις καλύτερες τιμές, άρα και τον καλύτερο σχηματισμό κοινοτήτων και ακολουθεί το Spectral Clustering. Αυτό εξηγείται, λόγω του αλγορίθμου Modularity Maximization, όπου ο ίδιος χρησιμοποιεί το βέλτιστο modularity για τον διαχωρισμό κοινοτήτων.

Πίνακας 5.22: Χρόνοι εκτέλεσης (sec) των αλγορίθμων εντοπισμού κοινοτήτων.

Αλγόριθμος	Club 1	Club 2	Club 3	Club 4	Club 5
Modularity Maximization	3.288	5.306	7.340	3.098	5.567
Girvan - Newman	3.652	6.258	20.541	3.293	6.832
Spectral Clustering	14.460	15.686	17.343	14.439	16.119

Πίνακας 5.23: Μέσος όρος *Modularity* των αλγορίθμων εντοπισμού κοινοτήτων ανά club.

Αλγόριθμος	Club 1	Club 2	Club 3	Club 4	Club 5
Modularity Maximization	0.209	0.167	0.485	0.367	0.250
Girvan - Newman	0.082	0.0004	0.292	0.117	0.15
Spectral Clustering	0.182	0.128	0.209	0.116	0.130

Συνολικά, ο αλγόριθμος Modularity Maximization, πετυχαίνει τα καλύτερα αποτελέσματα σε όλες τις περιπτώσεις, γεγονός φαίνεται ακόμη πιο ξεκάθαρα από το παράδειγμα του club 3. Από άποψη χρόνου και πολυπλοκότητας, είναι ο πιο αποδοτικός. Παρατηρώντας το κριτήριο του modularity, πραγματοποιεί τον βέλτιστο εντοπισμό κοινοτήτων, συγκριτικά με τις άλλες δύο μεθόδους που παρουσιάζονται. Ακολουθεί ο αλγόριθμος Spectral Clustering, όπου από την άποψη της αποτελεσματικότητας είναι ο αμέσως επόμενος, αλλά και ο πιο χρονοβόρος. Ο Girvan Newman, ενώ είναι ο δεύτερος πιο γρήγορος σε εκτέλεση, δεν ήταν τόσο αποτελεσματικός για το συγκεκριμένο σύνολο δεδομένων, αφού ειδικά για τα clubs 1 και 2, οι τιμές του modularity είναι αρκετά χαμηλές.



Σχήμα 5.28: Κοινότητες που σχηματίζονται με τις διαφορετικές μεθόδους τον Μήνα 9.

ΚΕΦΑΛΑΙΟ 6

ΕΠΙΛΟΓΟΣ

6.1 Σύνοψη Εργασίας

Η παρούσα διπλωματική εργασία επικεντρώθηκε στην ανάλυση κοινωνικών δικτύων, με έμφαση στα αθλητικά δίκτυα. Στο πρώτο κεφάλαιο "Εισαγωγή", περιγράφηκε εν συντομία το πρόβλημα που απασχόλησε, ενώ ορίστηκαν επίσης ο σκοπός και οι στόχοι της εργασίας. Επιπλέον, παρουσιάστηκε η δομή της εργασίας για να προετοιμάσει τον αναγνώστη για την ακολουθούμενη ανάπτυξη του κειμένου. Στο δεύτερο κεφάλαιο, με τίτλο "Βιβλιογραφική Ανασκόπηση", παραγμετοποιήθηκε μια εκτενής ανασκόπηση της σχετικής βιβλιογραφίας. Αναλύθηκαν επιστημονικές δημοσιεύσεις, που καλύπτουν ευρεία φάσματα ανάλυσης δικτύων, συμπεριλαμβανομένων και των αθλητικών. Καλύφθηκαν θέματα, όπως η ανάλυση κοινωνικών δικτύων, οι γράφοι γνώσης και η θεωρία γραφημάτων, εξετάστηκαν θεμελιώδεις έννοιες όπως ο βαθμός, ο μέσος βαθμός, και οι κεντρικότητες δικτύων. Επιπλέον, αναλύθηκε η εφαρμογή αυτών των θεωριών, στα αθλητικά δίκτυα, παρουσιάζοντας μια ποικιλία παραδειγμάτων. Στο τρίτο κεφάλαιο, "Ανάλυση Δεδομένων", περιγράφηκε η μεθοδολογία συλλογής και ανάλυσης δεδομένων. Αναλύθηκαν οι διαδικασίες, σχετικά με τη συλλογή, την επεξεργασία και τη δημιουργία γράφων για την περαιτέρω ανάλυση. Όπως, επίσης, έγινε και μια λεπτομερής παρουσίαση των δεδομένων που χρησιμοποιήθηκαν στην εργασία αυτή. Στο τέταρτο κεφάλαιο, με τίτλο "Μεθοδολογία", αναλύθηκε η διαδικασία επεξεργασίας των δεδομένων, περιλαμβάνοντας τη δημιουργία γράφων, την εξαγωγή μετρικών και τον εντοπισμό κοινοτήτων. Στο πέμπτο κεφάλαιο, παρουσιάστηκαν τα αποτελέσματα της ανάλυσης, τα οποία επικεντρώθηκαν, στους αλγόριθμους εντοπισμού κοινοτήτων που χρησιμοποιήθηκαν.

6.2 Σύνοψη Αποτελεσμάτων

Στο πλαίσιο της διπλωματικής εργασίας, η ανάλυση των αθλητικών δικτύων που δημιουργήθηκαν, αποκάλυψε σημαντικά ευρήματα, που διαμορφώνουν τη συνολική κατανόηση της παρούσας έρευνας. Μια πρώτη παρατήρηση, που προκύπτει από την ανάλυση των πινάκων, είναι η δυναμικότητα των κοινοτήτων στα αθλητικά κοινωνικά δίκτυα. Κάθε μήνα, παρατηρούνται αλλαγές στη σύσταση και τον αριθμό των κοινοτήτων. Ενώ ορισμένες κοινότητες, παραμένουν σταθερές, άλλες υφίστανται συνεχείς μεταβολές. Παρατηρείται ότι υπάρχουν βασικοί κόμβοι, που σχηματίζουν μια σταθερή ομάδα και διατηρούνται σε σταθερή θέση. Αντιθέτως, σε άλλες περιπτώσεις, όπου ο αριθμός των κοινοτήτων παραμένει σταθερός ή μπορεί να μεταβάλλεται, οι κόμβοι που απαρτίζουν τις κοινότητες αυτές, επίσης μεταβάλλονται. Με βάση τα παραπάνω, φαίνεται ότι οι σχέσεις "kudos" στο Strava, παίζουν ένα σημαντικό ρόλο στις κοινότητες των αθλητών. Τα "kudos" μπορούν να θεωρηθούν ως ένα είδος αναγνώρισης ή επιβράβευσης για τις αθλητικές επιδόσεις, και αντικατοπτρίζουν τις σχέσεις και τον κοινωνικό δεσμό μεταξύ των μελών της πλατφόρμας. Η δυναμική των κοινοτήτων και οι συνεχείς μεταβολές στη σύστασή τους, ενισχύουν τη σημασία των "kudos" ως μέσο επικοινωνίας και αλληλεπίδρασης μεταξύ των αθλητών. Η ανάλυση των γραφημάτων και των ομάδων, αναδεικνύει τη δομή των σχέσεων, που δημιουργούνται μέσω των "kudos" και τη σημασία τους στη διατήρηση των κοινοτήτων. Επιπλέον, η εφαρμογή αλγορίθμων ανάλυσης κοινοτήτων στα δεδομένα, αποκαλύπτει την πολυπλοκότητα των δικτύων και τη σημασία της επιλογής του κατάλληλου αλγορίθμου. Η εμφάνιση διακυμάνσεων στον αριθμό και τη σύσταση των κοινοτήτων, υποδεικνύει τη δυναμικότητα του αθλητικού κοινωνικού δικτύου. Έτσι, η ανάλυση των "kudos" αναδεικνύει την κοινωνική διάσταση της πλατφόρμας Strava και τον τρόπο με τον οποίο οι αθλητές αλληλεπιδρούν και συνδέονται.

Κατά την εφαρμογή των αλγορίθμων, παρατηρείται ότι ο αλγόριθμος Modularity Maximization επιδεικνύει μεγαλύτερη ευελιξία και αξιοπιστία. Ο Girvan Newman, αν και γρήγορος αλγόριθμος, δεν είναι τόσο αποτελεσματικός στην παρούσα ανάλυση, ενώ το Spectral Clustering, λόγω της φύσης του, εμφανίζει σταθερό αριθμό κοινοτήτων, παρά τις δυναμικές αλλαγές που πραγματοποιούνται μεταξύ των κόμβων, όπως επίσης, είναι ιδιαίτερα αργός σε σύγκριση με τους άλλους δύο. Συνολικά, ο αλγόριθμος Modularity Maximization ξεχωρίζει για την αποτελεσματικότητα και τη δυνατότητα προσαρμογής στις μεταβαλλόμενες κοινωνικές δομές. Ο χρόνος εκτέλεσης και το modularity επιβεβαιώνουν την αποτελεσματικότητά του. Στη

συνέχεια, ακολουθεί ο Spectral Clustering, ενώ ο Girvan Newman ενδείκνυται για γρήγορες αναλύσεις, αλλά υστερεί σε θέματα απόδοσης.

6.3 Περιορισμοί και Προτάσεις για Μελλοντική Έρευνα

Ένας σημαντικός περιορισμός της παρούσας έρευνας, αφορά τα περιορισμένα δεδομένα που χρησιμοποιήθηκαν. Το σύνολο δεδομένων που χρησιμοποιήθηκε, ήταν αρκετά ασαφές, όσον αφορά την πληροφορία που παρέχουν οι ερευνητές, για τα υπόλοιπα χαρακτηριστικά (freq_run, time_run, male, novice, κλπ). Παρόλο που υπάρχουν αρκετά διαφορετικά χαρακτηριστικά, που θα μπορούσαν να είχαν αξιοποιηθεί, αλλάζουν τα αρχικά μεγέθη των πινάκων που τα ορίζουν και λόγω της ανωνυμοποίησης που έχει πραγματοποιηθεί, με μια απλή παρατήρηση των πινάκων που έγινε δεν υπάρχει αντιστοιχία των αρχικών κόμβων κατά την αλλαγή των χαρακτηριστικών. Η ανωνυμοποίηση, παρότι προστατεύει την ιδιωτικότητα των ατόμων, ενδέχεται να επηρεάσει τη συνολική συνοχή των δεδομένων και την αντιστοιχία τους, κατά την αλλαγή των χαρακτηριστικών. Ο περιορισμός αυτός, της ανωνυμοποίησης, επηρεάζει την ικανότητα να συνδέσουμε τις αλληλεπιδράσεις και τα υπόλοιπα χαρακτηριστικά με συγκεκριμένους αθλητές. Τα χαρακτηριστικά που δεν χρησιμοποιήθηκαν, θα μπορούσαν να προσθέσουν βάθος και ποικιλία στην ανάλυση, επιτρέποντας πιο εκτενείς και λεπτομερείς ερμηνείες των κοινοτήτων. Σε περίπτωση που δεν υπήρχε η ανάγκη ανωνυμοποίησης, ή αν οι ερευνητές ήταν σε θέση να παρέχουν διευκρινήσεις για τα δεδομένα αυτά, θα μπορούσε να εξεταστούν ακόμη και οι ατομικές αλληλεπιδράσεις μεταξύ αθλητών. Αυτό θα προσέφερε, μια πιο σφαιρική κατανόηση της κοινωνικής διάστασης της πλατφόρμας.

Η ανάλυση που πραγματοποιήθηκε, αναδεικνύει τον σημαντικό ρόλο των δεδομένων στην επίτευξη συνολικής κατανόησης των αθλητικών κοινοτήτων. Παρά τους περιορισμούς που αντιμετωπίσαμε με την ασάφεια του συνόλου δεδομένων και την ανωνυμοποίηση, υπάρχει μεγάλη ανάγκη, στον συγκεκριμένο τομέα της ανάλυσης αθλητικών κοινωνικών δικτύων, για μελλοντικές έρευνες. Ένα παράδειγμα τέτοιων ερευνών, που στοχεύει στη διεύρυνση του πεδίου αυτού, είναι η εξέταση διαφορετικών αθλημάτων, καθώς η σύγκριση μεταξύ διαφορετικών αθλητικών κοινοτήτων μπορεί να παρουσιάσει δομικές διαφορές και ομοιότητες. Ακόμη, με τα δεδομένα που χρησιμοποιήθηκαν στην εργασία αυτή, θα μπορούσε να πραγματοποιηθεί μια ανάλυση της εξέλιξης των κοινοτήτων κατά τη διάρκεια συγκεκριμένων χρονικών διαστημάτων. Η ανάλυση αυτή, θα μπορούσε να αποκαλύψει πτυχές της δυναμικής των αθλητικών δικτύων. Τέλος, η σύγκριση αποτελεσμάτων με διαφορετικές μεθόδους ανάλυσης, θα ενίσχυε

την αξιοπιστία των ευρημάτων, ενώ, η εξέταση των διαφορετικών παραμέτρων που επηρεάζουν την εξέλιξη των κοινοτήτων θα συνέβαλε στη βαθύτερη κατανόηση της δομής των αθλητικών δικτύων. Σε κάθε περίπτωση λόγω της περιορισμένης έρευνας που υπάρχει στον τομέα αυτό, οποιαδήποτε κατεύθυνση για μελλοντικές προσεγγίσεις, μπορεί να αποκαλύψει νέες πτυχές των αθλητικών κοινοτήτων και αλληλεπιδράσεων, προσφέροντας έτσι πολύτιμες πληροφορίες για τη σχέση μεταξύ αθλητισμού και κοινωνικής δικτύωσης.

Βιβλιογραφία

- [1] Ν Παπαηλιού. “Κοινωνικά δίκτυα & ανάλυση κοινωνικών δικτύων”. In: *Μονάδα Διοίκησης Πληροφοριακών Συστημάτων, ΕΠΙΣΕΥ* (2007).
- [2] David Ahmedt-Aristizabal et al. “Graph-based deep learning for medical diagnosis and analysis: past, present and future”. In: *Sensors* 21.14 (2021), p. 4758.
- [3] Tero Aittokallio and Benno Schwikowski. “Graph-based methods for analysing networks in cell biology”. In: *Briefings in bioinformatics* 7.3 (2006), pp. 243–255.
- [4] Rodrigo Aquino et al. “Influence of situational variables, team formation, and playing position on match running performance and social network analysis in brazilian professional soccer players”. In: *The Journal of Strength & Conditioning Research* 34.3 (2020), pp. 808–817.
- [5] Sören Auer et al. “Dbpedia: A nucleus for a web of open data”. In: *international semantic web conference*. Springer. 2007, pp. 722–735.
- [6] Αριστείδης Βαγγελάτος, Φραγκίσκος Φώσκολος, and Θεόδωρος Κομνηνός. “Τα κοινωνικά δίκτυα στο σχολείο του σήμερα”. In: *Πρακτικά Εργασιών 7ου Πανελλήνιου Συνεδρίου με Διεθνή Συμμετοχή «Οι ΤΠΕ στην Εκπαίδευση 2* (2010), pp. 791–794.
- [7] Zhongbo Bai and Xiaomei Bai. “Sports big data: management, analysis, applications, and challenges”. In: *Complexity* 2021 (2021), pp. 1–11.
- [8] Ravindra B Bapat. *Graphs and matrices*. Vol. 27. Springer, 2010.
- [9] Albert-László Barabási and Réka Albert. “Emergence of scaling in random networks”. In: *science* 286.5439 (1999), pp. 509–512.
- [10] JIANG Bingchuan et al. “Geographic knowledge graph building extracted from multi-sourced heterogeneous data”. In: *Acta Geodaetica et Cartographica Sinica* 47.8 (2018), p. 1051.

- [11] Christian Bizer, Tom Heath, and Tim Berners-Lee. “Linked data: The story so far”. In: *Semantic services, interoperability and web applications: emerging concepts*. IGI global, 2011, pp. 205–227.
- [12] Olivier Bodenreider. “The unified medical language system (UMLS): integrating biomedical terminology”. In: *Nucleic acids research* 32.suppl.1 (2004), pp. D267–D270.
- [13] Kurt Bollacker et al. “Freebase: a collaboratively created graph database for structuring human knowledge”. In: *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. 2008, pp. 1247–1250.
- [14] Phillip Bonacich. “Power and centrality: A family of measures”. In: *American journal of sociology* 92.5 (1987), pp. 1170–1182.
- [15] Danah M Boyd and Nicole B Ellison. “Social network sites: Definition, history, and scholarship”. In: *Journal of computer-mediated Communication* 13.1 (2007), pp. 210–230.
- [16] Sergey Brin and Lawrence Page. “The anatomy of a large-scale hypertextual web search engine”. In: *Computer networks and ISDN systems* 30.1-7 (1998), pp. 107–117.
- [17] Viviane Callier. *Science of Teamwork*. ASU - Ask A Biologist. Dec. 2012.
- [18] Qian Cao and Yi-ming Zhao. “The technical realization process and related applications of the knowledge graph”. In: *Inf. Stud. Theory Appl* 38 (2015), pp. 127–132.
- [19] Andrew J Carnes and Sara E Mahoney. “Cohesion is Associated with Perceived Exertion and Enjoyment during Group Running.” In: *Journal of Exercise Physiology Online* 19.6 (2016).
- [20] Wai-Kai Chen. *Graph theory and its engineering applications*. Vol. 5. World Scientific, 1997.
- [21] Zhe Chen et al. “Knowledge Graph Completion: A Review”. In: *IEEE Access* 8 (2020), pp. 192435–192456. DOI: 10.1109/ACCESS.2020.3030076.
- [22] Aaron Clauset, Mark EJ Newman, and Cristopher Moore. “Finding community structure in very large networks”. In: *Physical review E* 70.6 (2004), p. 066111.
- [23] Jesse Couture. “Reflections from the ‘Strava-sphere’: Kudos, community, and (self-) surveillance on a social network for athletes”. In: *Qualitative Research in Sport, Exercise and Health* 13.1 (2021), pp. 184–200.

- [24] Jesper Dall and Michael Christensen. “Random geometric graphs”. In: *Physical review E* 66.1 (2002), p. 016121.
- [25] Josep Daz et al. “Linear Orderings of Random Geometric Graphs”. In: (Apr. 2001).
- [26] Inderjit S Dhillon, Yuqiang Guan, and Brian Kulis. “Kernel k-means: spectral clustering and normalized cuts”. In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. 2004, pp. 551–556.
- [27] Richard C Dorf. *The engineering handbook*. CRC press, 2018.
- [28] Paul Erdős, Alfréd Rényi, et al. “On the evolution of random graphs”. In: *Publ. math. inst. hung. acad. sci* 5.1 (1960), pp. 17–60.
- [29] Illés J Farkas et al. “Spectra of “real-world” graphs: Beyond the semicircle law”. In: *Physical Review E* 64.2 (2001), p. 026704.
- [30] Rob Franken, Hidde Bekhuis, and Jochem Tolsma. “Kudos make you run! How runners influence each other on the online social network Strava”. In: *Social Networks* 72 (2023), pp. 151–164.
- [31] Linton C Freeman et al. “Centrality in social networks: Conceptual clarification”. In: *Social network: critical concepts in sociology*. Londres: Routledge 1 (2002), pp. 238–263.
- [32] Michelle Girvan and Mark EJ Newman. “Community structure in social and biological networks”. In: *Proceedings of the national academy of sciences* 99.12 (2002), pp. 7821–7826.
- [33] Alessandro Giuliani et al. “Proteins as networks: usefulness of graph theory in protein science”. In: *Current Protein and Peptide Science* 9.1 (2008), pp. 28–38.
- [34] Ali Fatih Gündüz and Ahmet Karadoğan. “Community Detection in Social Media Network with Maximum Modularity Using Girvan-Newman Algorithm”. In: *International Symposium on Multidisciplinary Studies and Innovative Technologies*. 2017, pp. 222–225.
- [35] Marion E Hambrick, Samuel H Schmidt, and Alicia M Cintron. “Cohesion and leadership in individual sports: a social network analysis of participation in recreational running groups”. In: *Managing Sport and Leisure* 23.3 (2018), pp. 225–239.
- [36] WHO ICD. “for Mortality and Morbidity Statistics”. In: *Autism spectrum disorder. Versión 5* (11), p. 2021.
- [37] Vasileios Karyotis, Eleni Stai, and Symeon Papavassiliou. *Evolutionary dynamics of complex communications networks*. CRC Press, 2013.

- [38] Mayank Kejriwal. *Domain-specific knowledge graph construction*. Springer, 2019.
- [39] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. “Graphs over time: densification laws, shrinking diameters and possible explanations”. In: *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*. 2005, pp. 177–187.
- [40] Jia Lirong et al. “Construction of traditional Chinese medicine knowledge graph”. In: *Journal of Medical Informatics* (2015), pp. 51–53.
- [41] Hugo Liu and Push Singh. “ConceptNet—a practical commonsense reasoning toolkit”. In: *BT technology journal* 22.4 (2004), pp. 211–226.
- [42] Jialu Liu and Jiawei Han. “Spectral clustering”. In: *Data clustering*. Chapman and Hall/CRC, 2018, pp. 177–200.
- [43] Feng Lu, Li Yu, and PY Qiu. “On geographic knowledge graph”. In: *J. Geo Inf. Sci* 19.6 (2017), pp. 723–734.
- [44] Hafiz Abid Mahmood Malik. “Complex network formation and analysis of online social media systems”. In: *Cmes-Comr Model Engg & Sci* 130.3 (2022), pp. 1737–1750.
- [45] Cynthia Matuszek et al. “An introduction to the syntax and content of Cyc”. In: *UMBC Computer Science and Electrical Engineering Department Collection* (2006).
- [46] Andrew McCallum. “Information extraction: Distilling structured data from unstructured text”. In: *Queue* 3.9 (2005), pp. 48–57.
- [47] George A Miller. “WordNet: a lexical database for English”. In: *Communications of the ACM* 38.11 (1995), pp. 39–41.
- [48] Jaron Mink et al. “Users Can Deduce Sensitive Locations Protected by Privacy Zones on Fitness Tracking Apps”. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–21.
- [49] Raymond J Mooney and Razvan Bunescu. “Mining knowledge from text using information extraction”. In: *ACM SIGKDD explorations newsletter* 7.1 (2005), pp. 3–10.
- [50] Roberto Navigli and Simone Paolo Ponzetto. “BabelNet: Building a very large multilingual semantic network”. In: *Proceedings of the 48th annual meeting of the association for computational linguistics*. 2010, pp. 216–225.
- [51] Mark EJ Newman. “Modularity and community structure in networks”. In: *Proceedings of the national academy of sciences* 103.23 (2006), pp. 8577–8582.

- [52] Mark EJ Newman and Michelle Girvan. “Finding and evaluating community structure in networks”. In: *Physical review E* 69.2 (2004), p. 026113.
- [53] Howard L Nixon. “A social network analysis of influences on athletes to play with pain and injuries”. In: *Journal of sport and social issues* 16.2 (1992), pp. 127–135.
- [54] Howard L Nixon. “Social network analysis of sport: Emphasizing social structure in sport sociology”. In: *Sociology of Sport Journal* 10.3 (1993), pp. 315–321.
- [55] Tore Opsahl and Pietro Panzarasa. “Clustering in weighted networks”. In: *Social networks* 31.2 (2009), pp. 155–163.
- [56] Supun Perera, Michael Bell, and Michiel Bliemer. “Network science approach to modelling the topology and robustness of supply chain networks: a review and perspective”. In: *Applied Network Science* 2 (Oct. 2017), p. 33. DOI: 10.1007/s41109-017-0053-0.
- [57] Jay Pujara et al. “Knowledge graph identification”. In: *The Semantic Web–ISWC 2013: 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I* 12. Springer. 2013, pp. 542–557.
- [58] Filippo Radicchi. “Who is the best player ever? A complex network analysis of the history of professional tennis”. In: *PloS one* 6.2 (2011), e17249.
- [59] Martin Roderick. “The sociology of risk, pain, and injury: A comment on the work of Howard L. Nixon II”. In: *Sociology of Sport Journal* 15.1 (1998), pp. 64–79.
- [60] Tong Ruan et al. “Construction of traditional Chinese medicine knowledge graph and its application”. In: *Journal of Medical Informatics* 37.4 (2016), pp. 8–13.
- [61] Max Schmachtenberg, Christian Bizer, and Heiko Paulheim. “Adoption of the linked data best practices in different topical domains”. In: *The Semantic Web–ISWC 2014: 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part I* 13. Springer. 2014, pp. 245–260.
- [62] Philipp Schuetz and Amedeo Caflisch. “Efficient modularity optimization by multistep greedy algorithm and vertex mover refinement”. In: *Physical Review E* 77.4 (2008), p. 046112.
- [63] Clara Stegehuis and Lotte Weedage. “Degree distributions in AB random geometric graphs”. In: *Physica A: Statistical Mechanics and its Applications* 586 (2022), p. 126460.

- [64] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. “Yago: a core of semantic knowledge”. In: *Proceedings of the 16th international conference on World Wide Web*. 2007, pp. 697–706.
- [65] Shazia Tabassum et al. “Social network analysis: An overview”. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8.5 (2018), e1256.
- [66] Frank W Takes and Walter A Kusters. “Computing the eccentricity distribution of large graphs”. In: *Algorithms* 6.1 (2013), pp. 100–118.
- [67] S.A. Timashev. “Cyber Reliability, Resilience, and Safety of Physical Infrastructures”. In: *IOP Conference Series: Materials Science and Engineering* 481 (Mar. 2019), p. 012009. DOI: 10.1088/1757-899X/481/1/012009.
- [68] Fernando Vega-Redondo. *Complex social networks*. 44. Cambridge University Press, 2007.
- [69] Thomas C Vilhena. *Data Replication in Random Regular Graphs*. Online source. 2020. URL: <https://thomasvilhena.com/2020/05/data-replication-in-random-regular-graphs>.
- [70] Denny Vrandečić and Markus Krötzsch. “Wikidata: a free collaborative knowledge-base”. In: *Communications of the ACM* 57.10 (2014), pp. 78–85.
- [71] Hagen Wäsche et al. “Social network analysis in sport research: an emerging paradigm”. In: *European Journal for Sport and Society* 14.2 (2017), pp. 138–165.
- [72] WEN, Tzai-Hung. *The Small-World vs. Scale-Free Graphs*. NTU Geography. 2011. URL: https://homepage.ntu.edu.tw/~wenthung/R_Network/LabM4.html%7D.
- [73] M WICK. “GeoNames Home [online]. Switzerland: The GeoNames Geographical Database”. In: <http://www.geonames.org> (2010).
- [74] Wentao Wu et al. “Probase: A probabilistic taxonomy for text understanding”. In: *Proceedings of the 2012 ACM SIGMOD international conference on management of data*. 2012, pp. 481–492.
- [75] Bo Xu et al. “CN-DBpedia: A never-ending Chinese knowledge extraction system”. In: *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer. 2017, pp. 428–438.
- [76] Zenglin Xu et al. “Review on knowledge graph techniques”. In: *Journal of University of Electronic Science and Technology of China* 45.4 (2016), pp. 589–606.
- [77] Wayne W Zachary. “Modeling social network processes using constrained flow representations”. In: *Social Networks* 6.3 (1984), pp. 259–292.

- [78] Dezheng Zhang et al. “Construction of knowledge graph of traditional Chinese medicine based on the ontology”. In: *Technology intelligence engineering* 3.1 (2017), pp. 35–42.
- [79] Zhanfang Zhao, Sung-Kook Han, and In-Mi So. “Architecture of knowledge graph construction techniques”. In: *International Journal of Pure and Applied Mathematics* 118.19 (2018), pp. 1869–1883.
- [80] Shuren Zhou et al. “Double-domain imaging and adaption for person re-identification”. In: *IEEE Access* 7 (2019), pp. 103336–103345.