

Ατομική Διπλωματική Εργασία

**Η ΕΞΕΛΙΞΗ ΤΩΝ ΕΠΙΣΤΗΜΟΝΙΚΩΝ ΣΥΝΕΡΓΑΣΙΩΝ ΤΟΥ
ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΚΥΠΡΟΥ**

Χριστίνα Θεοδώρου

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Σεπτέμβριος 2010

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Η εξέλιξη των επιστημονικών συνεργασιών του Πανεπιστημίου Κύπρου

Χριστίνα Θεοδώρου

Επιβλέπων Καθηγητής

Μάριος Δικαιάκος

Η Ατομική Διπλωματική Εργασία υποβλήθηκε προς μερική εκπλήρωση των απαιτήσεων απόκτησης του πτυχίου Πληροφορικής του Τμήματος Πληροφορικής του Πανεπιστημίου Κύπρου

Σεπτέμβριος 2010

Ευχαριστίες

Θα ήθελα κατ' αρχή να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κ. Μάριο Δικαϊάκο για τη βοήθεια και τις κατευθύνσεις που μου έχει προσφέρει για την περάτωση της εργασίας αυτής.

Επίσης θα ήθελα να ευχαριστήσω το διευθυντή βιβλιοθήκης του Πανεπιστημίου Κύπρου, τον Δρ. Φίλιππο Τσιμπόγλου για την βοήθεια του στην επιλογή των βάσεων δεδομένων από τις οποίες θα ανακτούσα τα απαιτούμενα δεδομένα της εργασίας μου.

Ειδικά θα ήθελα να ευχαριστήσω τον κ. Λούη Προκοπίου για την πολύτιμη βοήθεια που μου πρόσφερε κατά τη διάρκεια ανάκτησης δεδομένων από τις βάσεις δεδομένων που προσφέρει η βιβλιοθήκη του Πανεπιστημίου Κύπρου, εξηγώντας μου τις λειτουργίες που προσφέρουν.

Περίληψη

Ο σκοπός της διπλωματικής αυτής εργασίας είναι η μελέτη των επιστημονικών συνεργασιών των επιστημόνων του Πανεπιστημίου Κύπρου με άλλους επιστήμονες (είτε με επιστήμονες του ίδιου του πανεπιστημίου είτε και με επιστήμονες εκτός Πανεπιστημίου Κύπρου) και παρουσίαση της εξέλιξης αυτών των συνεργασιών.

Για τη μελέτη της εξέλιξης των επιστημονικών συνεργασιών του Πανεπιστημίου Κύπρου, δημιουργήθηκε ένα δίκτυο συνεργασιών (collaboration network), του οποίου διερευνήθηκε η δομή. Δύο επιστήμονες θεωρούνται συνδεδεμένοι εάν έχουν συγγράψει ένα άρθρο από κοινού.

Η κατασκευή του δικτύου συνεργασίας έγινε με βάση στοιχεία που συλλέχθηκαν από βάσεις δεδομένων. Συγκεκριμένα, για κάθε άρθρο συλλέχθηκαν ο τίτλος του άρθρου, τα ονοματεπώνυμα των επιστημόνων που συνεργάστηκαν για τη συγγραφή του άρθρου, η χρονιά δημοσίευσης του άρθρου και τα ονόματα των ιδρυμάτων-οργανισμών στα οποία ανήκουν οι συγγραφείς του άρθρου. Στη συνέχεια, τα δεδομένα αυτά επεξεργάστηκαν και αποθηκεύτηκαν σε μια βάση δεδομένων.

Έπειτα, ανακτώντας τα απαραίτητα δεδομένα από τη βάση δεδομένων· δημιουργήθηκε το δίκτυο επιστημονικών συνεργασιών του Πανεπιστημίου Κύπρου. Το δίκτυο αναλύθηκε και έγινε η εξαγωγή κάποιων στατιστικών στοιχείων όσον αφορά τη δομή του δικτύου γιατί η δομή δικτύων που απεικονίζουν συνεργασίες έχουν σημαντικές συνέπειες για τη διάδοση των πληροφοριών.

Στο τέλος, παρουσιάζονται και σχολιάζονται τα αποτελέσματα της ανάλυσης των δικτύων συνεργασίας μέσω κάποιων γραφικών παραστάσεων και πινάκων. Εντοπίζονται εμφανείς διαφορές στις συνεργασίες ανάλογα με το τμήμα στο οποίο ανήκουν οι επιστήμονες του Πανεπιστημίου Κύπρου καθώς επίσης ανάλογα με τη χρονιά δημοσίευσης των άρθρων και παρουσιάζονται ανά τμήμα οι μεγαλύτερες συνεργασίες του Πανεπιστημίου Κύπρου με άλλα ιδρύματα-οργανισμούς.

Περιεχόμενα

Κεφάλαιο 1	Εισαγωγή.....	1
	1.1 Γενικά	1
	1.2 Κοινωνικά δίκτυα	2
	1.3 Εισαγωγή στους γράφους	3
	1.4 Στόχος της παρούσας εργασίας	4
	1.5 Μεθοδολογία υλοποίησης	5
	1.6 Περιεχόμενο της εργασίας ανά κεφάλαιο	7
Κεφάλαιο 2	Βιβλιογραφία και Σχεδίαση Βάσης Δεδομένων	7
	2.1 Γενικά	7
	2.2 Ανάλυση δεδομένων-Στατιστικά Στοιχεία	10
	2.3 Διάγραμμα σχέσης οντοτήτων(ER-diagram)	13
	2.4 Σχεσιακό μοντέλο	15
Κεφάλαιο 3	Εξόρυξη Δεδομένων και Αποθήκευση.....	18
	3.1 Γενικά	18
	3.2 Ανάλυση συστήματος εξόρυξης βασικών δεδομένων	19
	3.3 Δεδομένα εισόδου-Scopus	24
	3.4 Ανάλυση συστήματος εξόρυξης δεδομένων Scopus	25
	3.5 Δεδομένα εισόδου-ISI(Web of Knowledge)	33
	3.6 Ανάλυση συστήματος εξόρυξης δεδομένων ISI	35
	3.7 Επεξεργασία δεδομένων ιδρυμάτων-οργανισμών	46
	3.7 Δεδομένα εξόδου	47
Κεφάλαιο 4	Δημιουργία Δικτύου Συνεργασίας.....	51
	4.1 Επιλογή βιβλιοθήκης SNAP	51
	4.2 Δημιουργία δικτύου συνεργασίας	53
	4.2.1 Περιγραφή κλάσεων	54
	4.2.2 Περιγραφή δημιουργίας δικτύου συνεργασίας	56
	4.2.3 Περιγραφή δημιουργίας γράφου	57

Κεφάλαιο 5	Ανάλυση Δικτύου Συνεργασίας	59
5.1	Κριτήρια ανάλυσης δικτύου συνεργασίας	59
5.2	Περιγραφή εξόρυξης γενικών στατιστικών δεδομένων	63
5.3	Περιγραφή ανάλυσης δικτύου συνεργασίας	65
5.3.1	Ανάλυση δικτύου συνεργασίας επιστημόνων	65
5.3.2	Ανάλυση δικτύου συνεργασίας ιδρυμάτων-οργανισμών	68
5.4	Αποτελέσματα εξόρυξης γενικών στατιστικών δεδομένων	71
5.5	Αποτελέσματα επιστημονικού δικτύου συνεργασίας	72
5.6	Αποτελέσματα δικτύου συνεργασίας ιδρυμάτων	102
ΚΕΦΑΛΑΙΟ 6	Συμπεράσματα	111
6.1	Τελικά συμπεράσματα	111
Βιβλιογραφία	115
Παράρτημα Α	A-1
Παράρτημα Β	B-3
Παράρτημα Γ	Γ-11

Κεφάλαιο 1

Εισαγωγή

1.1 Γενικά	1
1.2 Κοινωνικά δίκτυα	2
1.3 Εισαγωγή στους γράφους	3
1.4 Στόχος της παρούσας εργασίας	4
1.5 Μεθοδολογία υλοποίησης	5
1.6 Περιεχόμενο της εργασίας ανά κεφάλαιο	7

1.1 Γενικά

Ο σκοπός της διπλωματικής αυτής εργασίας είναι η μελέτη των επιστημονικών συνεργασιών των επιστημόνων του Πανεπιστημίου Κύπρου με άλλους επιστήμονες (είτε με επιστήμονες το υ ίδιου του πανεπιστημίου είτε και με επιστήμονες εκτός Πανεπιστημίου Κύπρου) και παρουσίαση της εξέλιξης αυτών των συνεργασιών. Για τη μελέτη των συνεργασιών αυτών, θα συλλεχθούν δεδομένα από βάσεις δεδομένων που αφορούν δημοσιεύσεις των επιστημόνων του Πανεπιστημίου Κύπρου, με τα οποία θα δημιουργηθεί ένα δίκτυο συνεργασίας.

Τώρα ένα δίκτυο συνεργασίας επιστημόνων είναι ένα πολύπλοκο σύστημα αφού είναι δύσκολο τέτοια συστήματα να τα μοντελοποιήσεις και να τα απεικονίσεις. Έτσι, πρέπει να μελετηθούν οι προϋποθέσεις που πρέπει να έχει ένα δίκτυο συνεργασίας επιστημόνων προκειμένου να μοντελοποιηθεί και να αναπαραστηθεί. Άξιο αναφοράς είναι το γεγονός ότι τέτοιου είδους δίκτυα είναι γνωστά ως *Social networks* αφού είναι δίκτυα τα οποία απεικονίζουν τις σχέσεις κάποιων ατόμων και πιο συγκεκριμένα *Collaboration networks* αφού η σχέση των ατόμων αυτών είναι συγκεκριμένη, είναι σχέση συνεργασίας.

1.2 Κοινωνικά δίκτυα

Τα κοινωνικά δίκτυα (social networks) είναι συλλογές από ανθρώπους, στα οποία ο καθένας γνωρίζει ένα υποσύνολο των υπολοίπων. Ένα τέτοιο δίκτυο μπορεί να αναπαραστηθεί από ένα σύνολο σημείων (κορυφές) που υποδηλώνουν τους ανθρώπους, που συνδέονται με γραμμές (ακμές) που υποδηλώνουν τη γνωριμία τους. Κοινωνικό δίκτυο θα μπορούσε να κατασκευαστεί για μια εταιρεία-επιχείρηση, για ένα σχολείο ή πανεπιστήμιο ή για οποιαδήποτε άλλη κοινότητα.

Εδώ και 50 χρόνια τα κοινωνικά δίκτυα (social networks) έχουν αποτελέσει αντικείμενο προς μελέτη τόσο στις εμπειρικές όσο και στις θεωρητικές σπουδές. Αυτό οφείλεται στο γεγονός ότι ο άνθρωπος πάντα ήθελε να μάθει για τα πρότυπα της ανθρώπινης αλληλεπίδρασης, αλλά και γιατί η δομή αυτών των δικτύων έχει σημαντικές συνέπειες για την εξάπλωση των πληροφοριών και των ασθενειών. Για παράδειγμα, για τη διάδοση μιας φήμης ή γρίπης του τρέχοντος έτους.

Ο **Stanley Milgram** ήταν ο πρώτος άνθρωπος που μελέτησε τη δομή των κοινωνικών δικτύων[1]. Πώς; Αποφάσισε να επιλέξει τυχαία κάποιον από έναν τηλεφωνικό κατάλογο Νεμπράσκα και να του στείλει ένα γράμμα, ο οποίος στη συνέχεια να το έστειλε σε ένα φίλο χρηματιστή του Milgram. Οι οδηγίες που είχε δώσει ο Milgram ήταν το γράμμα να δίνεται σε άτομο το οποίο θα ήταν πιο κοντά σε αυτόν τον χρηματιστή είτε κοινωνικά είτε γεωγραφικά.

Σε κάποια φάση το γράμμα έφτασε στον προορισμό του και ο Milgram ανακάλυψε ότι ο μέσος αριθμός των ενδιάμεσων 'σταθμών' ήταν μόνο έξι. Αυτό είχε ως αποτέλεσμα ο John Guare το 1990 να δώσει τον τίτλο «*Six Degrees of Separation*».

Φυσικά υπήρχε η πιθανότητα στο πείραμα του Milgram, το γράμμα να ακολουθήσει μεγαλύτερη διαδρομή και στην πορεία είτε να ξεχαστεί είτε να χαθεί. Το αποτέλεσμα όμως του πειράματός του είναι απόδειξη της υπόθεσης «*small worlds*». Δηλαδή, τα πλείστα ζευγάρια από ανθρώπους σε έναν πληθυσμό μπορούν να συνδεθούν από μια μικρή αλυσίδα ενδιάμεσων γνωριμιών, ακόμη και αν το μέγεθος του πληθυσμού είναι αρκετά μεγάλο[2].

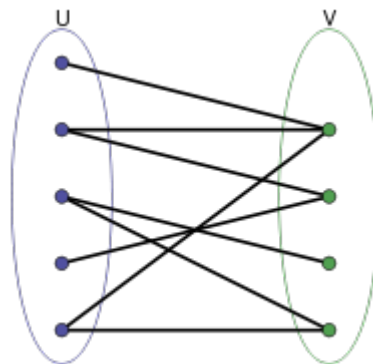
1.3 Εισαγωγή στους γράφους

Στην επιστήμη της Πληροφορικής, οι επιστημονικές συνεργασίες των επιστημόνων του Πανεπιστημίου Κύπρου με άλλους επιστήμονες μπορούν να αναπαραστηθούν σαν ένας γράφος που βασικά απεικονίζει ένα δίκτυο συνεργασίας των επιστημόνων.

Ένας **γράφος** G είναι μια *αφηρημένη δομή δεδομένων*, η οποία αποτελείται από ένα πεπερασμένο σύνολο από διατεταγμένα ζεύγη (V, E) όπου V (**vertices**) είναι οι κορυφές-κόμβοι του και E (**edges**) είναι οι ακμές-τόξα του. Ο γράφος μπορεί να είναι κατευθυνόμενος ή και όχι. Στην προκειμένη περίπτωση, ο γράφος δεν θα είναι κατευθυνόμενος αφού μια ακμή μεταξύ δυο κόμβων θα αναπαριστά μια συνεργασία μεταξύ δυο επιστημόνων.

Επομένως, η μελέτη των επιστημονικών συνεργασιών των επιστημόνων του Πανεπιστημίου Κύπρου με άλλους επιστήμονες και η παρουσίαση της εξέλιξης αυτών των συνεργασιών, θα γίνει μέσω μιας εφαρμογής, η οποία θα δημιουργά ενός τέτοιου είδους γράφο, και συγκεκριμένα ένα διμερή γράφο.

Ένας *διμερής γράφος* $G = (U, V, E)$ είναι ένα γράφημα του οποίου οι κορυφές μπορούν να χωριστούν σε δυο ξεχωριστά σύνολα U και V , με τέτοιο τρόπο έτσι ώστε κάθε ακμή E να ενώνει ένα κόμβο από το σύνολο U με ένα κόμβο στο σύνολο V . Έτσι, τα σύνολα U και V είναι ανεξάρτητα σύνολα και σε αυτό το γράφημα δεν παρατηρείτε κανέναν κύκλο.



Διμερής Γράφος
<http://www.answers.com/topic/completebipartite-graph>

Στην προκειμένη περίπτωση, θα δημιουργηθεί ένας διμερής γράφος στον οποίο οι κόμβοι που θα ανήκουν στο σύνολο U θα είναι οι ίδιοι οι επιστήμονες (επιστήμονες του Πανεπιστημίου Κύπρου και οι άλλοι επιστήμονες-συνεργάτες) και οι κόμβοι που θα ανήκουν στο σύνολο V θα είναι οι μελέτες τους, δηλαδή αυτό που τους συσχετίζει, οι οποίες θα είναι άρθρα τα οποία έχουν γράψει μαζί (συνεργάστηκαν για την συγγραφή τους), τα οποία έχουν δημοσιευτεί στο διαδίκτυο ή σε επιστημονικά περιοδικά. Έτσι σε κάθε κόμβο τύπου άρθρο θα υπάρχουν ακμές E με τον κάθε κόμβο τύπου επιστήμονα, ο οποίος αντιστοιχεί σε ένα από τους συγγραφείς του συγκεκριμένου άρθρου.

1.4 Στόχος της παρούσας εργασίας

Ο σκοπός της διπλωματικής αυτής εργασίας είναι η μελέτη των επιστημονικών συνεργασιών των επιστημόνων του Πανεπιστημίου Κύπρου με άλλους επιστήμονες (είτε με επιστήμονες του ίδιου του πανεπιστημίου είτε και με επιστήμονες εκτός Πανεπιστημίου Κύπρου) και παρουσίαση της εξέλιξης αυτών των συνεργασιών.

Για να μελετηθεί η εξέλιξη των επιστημονικών συνεργασιών, θα πρέπει να πραγματοποιηθούν τα εξής:

1. Συλλογή των απαιτούμενων δεδομένων από βάσεις δεδομένων που προσφέρει μέσω διαδικτύου η βιβλιοθήκη του Πανεπιστημίου Κύπρου.
2. Αποθήκευση των δεδομένων σε μια βάση δεδομένων μέσω κάποιων scripts.
3. Αποθήκευση των δεδομένων που βρίσκονται στους πίνακες στη βάση δεδομένων σε αρχεία κειμένου.
4. Δημιουργία γράφου μέσω μιας εφαρμογής, η οποία θα χρησιμοποιεί τα αρχεία κειμένου με τα δεδομένα της βάσης δεδομένων.
5. Ανάλυση του γράφου βάση κάποιων δομικών προδιαγραφών.

Πραγματοποιώντας τα πιο πάνω βήματα, τότε είμαστε σε θέση να εξάξουμε συμπεράσματα για την εξέλιξη των επιστημονικών συνεργασιών των επιστημόνων του Πανεπιστημίου Κύπρου για κάθε τμήμα της επιλογής μας ή ακόμη για τμήμα και χρονιά ξεχωριστά. Έτσι, θα μπορούμε να συμπεράνουμε κατά πόσο με την πάροδο του χρόνου εξελίσσονται οι επιστημονικές συνεργασίες γενικά και ειδικά.

1.5 Μεθοδολογία υλοποίησης

Κατά τη διάρκεια του χειμερινού εξαμήνου ακαδημαϊκού έτους 2009-2010, μελετήθηκε το βιβλίο «*Linked*» του συγγραφέα *Barabasi, Albert-Laszlo*, το οποίο εξηγούσε τον τρόπο με τον οποίο όλα συνδέονται με όλα τα υπόλοιπα και τι σημαίνει αυτό για τις επιχειρήσεις, την επιστήμη αλλά και την καθημερινή ζωή. Επιπλέον μελετήθηκαν άρθρα τα οποία ασχολήθηκαν με σχετικές εφαρμογές ανάλυσης γράφων καθώς επίσης μελετήθηκαν εργαλεία που προσφέρονται για ανάλυση γράφων προκειμένου να γίνει η επιλογή του κατάλληλου εργαλείου-λογισμικού.

Για τη συλλογή των απαιτούμενων δεδομένων για να μπορεί να δημιουργηθεί το δίκτυο συνεργασίας, έπρεπε να επιλεγθούν μία ή περισσότερες βάσεις δεδομένων από τις οποίες θα παίρναμε ένα ικανοποιητικό αριθμό δεδομένων προκειμένου ο γράφος που θα δημιουργόταν στη συνέχεια να είναι όσο το δυνατόν πιο ακριβής και πιο κοντά στην πραγματικότητα. Αφού έγινε η επιλογή των βάσεων δεδομένων, τότε έγινε η συλλογή των απαραίτητων δεδομένων.

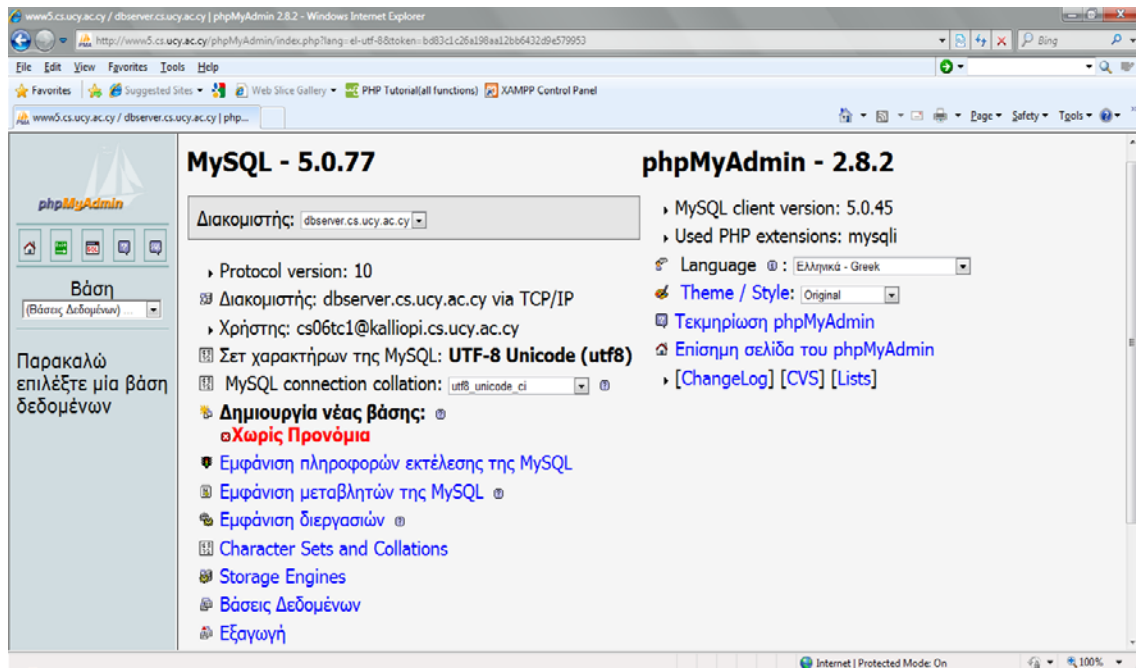
Μετά τη συλλογή των δεδομένων από τις βάσεις δεδομένων, κατά τη διάρκεια του εαρινού εξαμήνου ακαδημαϊκού έτους 2009-2010 μέσω κάποιων scripts επιλέχθηκαν τα απαιτούμενα δεδομένα και αποθηκεύτηκαν σε μια βάση δεδομένων. Συγκεκριμένα, για τη φάση αυτή ζητήθηκε χώρος για δημιουργία βάσης δεδομένων από το Πανεπιστήμιο Κύπρου. Για τη διαχείριση της βάσης δεδομένων χρησιμοποιήθηκε το λογισμικό *phpMyAdmin 2.8.2*, το οποίο υποστηρίζει τα εξής:

- **PHP 5.0 (scripting language)**

Η PHP είναι μια ευρέως γνωστή και πολυχρησιμοποιούμενη, γενικού σκοπού, object-oriented γλώσσα προγραμματισμού, η οποία χρησιμοποιείται για τη δημιουργία script.

- **MySQL 5.0.77**

MySQL είναι ένα σχεσιακό σύστημα διαχείρισης βάσεων δεδομένων (Relational Database Management System- RDBMS) που λειτουργεί ως server παρέχοντας σε πολλούς χρήστες πρόσβαση σε μια σειρά από βάσεις δεδομένων.



<http://www5.cs.ucy.ac.cy/phpMyAdmin>

Στη συνέχεια, έγινε η επιλογή του εργαλείου-λογισμικού μέσω του οποίου δημιουργήθηκε και αναλύθηκε το δίκτυο συνεργασίας. Συγκεκριμένα, επιλέχθηκε να χρησιμοποιήσουμε τη βιβλιοθήκη **SNAP library**, SNAP Ver. 2010-04-20.

Η βιβλιοθήκη SNAP (Stanford Network Analysis Platform) είναι μια πλατφόρμα για γενικού σκοπού ανάλυση δικτύων και είναι γραμμένη σε C++. Μπορεί εύκολα να διασχίσει κλιμακωτά και ογκώδη δίκτυα με εκατοντάδες εκατομμύρια κόμβους και δισεκατομμύρια ακμές. Χειρίζεται αποτελεσματικά μεγάλα γραφήματα, υπολογίζει δομικές ιδιότητες, δημιουργεί τυπικά και τυχαία γραφήματα, και υποστηρίζει χαρακτηριστικά πάνω σε κόμβους και ακμές.

Μελετήθηκαν καλά τα αρχεία που υπάρχουν στη βιβλιοθήκη SNAP library, προκειμένου να είμαστε σε θέση για τη δημιουργία και ανάλυση του δικτύου συνεργασίας. Μετά τη δημιουργία της εφαρμογής, η οποία δημιουργά και αναλύει ένα δίκτυο συνεργασίας, τότε τρέξαμε την εφαρμογή χρησιμοποιώντας δύο τύπων αρχεία. Το πρώτο τύπου αρχείου έχει τα δεδομένα όλων των άρθρων που δημοσιεύτηκαν από τους επιστήμονες του Πανεπιστημίου Κύπρου συγκεκριμένου τμήματος από το 1992

εώς το 2009 και το δεύτερου τύπου αρχείου έχει τα δεδομένα όλων των άρθρων που δημοσιεύτηκαν από τους επιστήμονες του Πανεπιστημίου Κύπρου συγκεκριμένου τμήματος και συγκεκριμένης χρονιάς δημοσίευσης.

Ανάλογα με την κάθε επιλογή, καταλάβαμε και συμπεράναμε την εξέλιξη των επιστημονικών συνεργασιών του Πανεπιστημίου Κύπρου με βάση τα κριτήρια που θέσαμε.

1.6 Περιεχόμενο της εργασίας ανά κεφάλαιο

Στο κεφάλαιο 2 που ακολουθεί παρουσιάζονται τα κριτήρια με τα οποία επιλέχθηκαν οι βάσεις δεδομένων από τις οποίες έγινε η συλλογή των δεδομένων και των στατιστικών στοιχείων που παρέχουν οι αντίστοιχες βάσεις δεδομένων, τα οποία οδήγησαν στην επιλογή των τμημάτων που τελικά αναλύθηκαν. Επιπλέον, παρουσιάζεται το διάγραμμα σχέσης οντοτήτων της βάσης δεδομένων που δημιουργήθηκε καθώς επίσης και το αντίστοιχο σχεσιακό μοντέλο στο οποίο μπορούμε να δούμε τους πίνακες της βάσης δεδομένων.

Στο κεφάλαιο 3 αναφέρεται ο τρόπος με τον οποίο έγινε η συλλογή των κατάλληλων δεδομένων για τη δημιουργία του γράφου, παρουσιάζονται η μορφή που έχουν τα δεδομένα που παίρνουμε από την κάθε βάση δεδομένων, η ανάλυση της αντίστοιχης εφαρμογής που υλοποιήθηκε και η μορφή δεδομένων εξόδου.

Στο κεφάλαιο 4 αναφέρονται οι λόγοι κατά τους οποίους επιλέχθηκε η χρήση του συγκεκριμένου λογισμικού για τη δημιουργία δικτύου συνεργασίας. Επιπλέον, γίνεται η ανάλυση της εφαρμογής που δημιουργεί το δίκτυο συνεργασίας.

Στο κεφάλαιο 5 επεξηγείτε ο όρος ‘ανάλυση δικτύου συνεργασίας’ και παρουσιάζονται τα κριτήρια με τα οποία θα γίνει η ανάλυση του δικτύου. Επιπρόσθετα, παρουσιάζονται τα αποτελέσματα των δικτύων συνεργασίας.

Τέλος, στο κεφάλαιο 6 θα επεξηγηθούν τα συμπεράσματα που προκύπτουν από τα αποτελέσματα της ανάλυσης των δικτύων συνεργασίας που υπάρχουν στο κεφάλαιο 5.

Κεφάλαιο 2

Βιβλιογραφία και Σχεδίαση Βάσης Δεδομένων

2.1 Γενικά	8
2.2 Ανάλυση δεδομένων – Στατιστικά Στοιχεία	10
2.3 Διάγραμμα σχέσης οντοτήτων (ER - diagram)	13
2.4 Σχεσιακό μοντέλο	15

2.1 Γενικά

Για τη δημιουργία αυτού του γράφου που θα αναπαριστά το δίκτυο συνεργασίας των επιστημόνων του Πανεπιστημίου Κύπρου είναι απαραίτητη η συλλογή σχετικών δεδομένων. Με τον όρο δεδομένα, εννοούμε:

- Τους αριθμούς εκείνους που ταυτοποιούν μοναδικά τον τίτλο των άρθρων που συγγράφηκαν από τουλάχιστον δυο επιστήμονες, από τους οποίους τουλάχιστον ο ένας είναι επιστήμονας του Πανεπιστημίου Κύπρου.
- Τη χρονιά δημοσίευσης του κάθε άρθρου.
- Τους αριθμούς εκείνους που ταυτοποιούν μοναδικά τα ονόματα των επιστημόνων για κάθε άρθρο, που συνεργάστηκαν για την συγγραφή του άρθρου.
- Την τιμή που καθορίζει εάν κάποιος επιστήμονας ανήκει στο Πανεπιστήμιο Κύπρου.
- Τον αριθμό εκείνο που ταυτοποιεί μοναδικά ένα ίδρυμα-οργανισμό, του οποίου επιστήμονας συνεργάστηκε με επιστήμονα/ες άλλου ιδρύματος-οργανισμού για τη συγγραφή κάποιου άρθρου.

Τα δεδομένα που ψάχνουμε είναι επιστημονικά άρθρα τα οποία βρίσκονται στο διαδίκτυο ή σε επιστημονικά περιοδικά.

Η επιλογή των δεδομένων έγινε μέσα από δυο βάσεις δεδομένων, οι οποίες περιέχουν και άρθρα σε ηλεκτρονική μορφή τα οποία έχουν δημοσιευτεί σε επιστημονικά περιοδικά. Να σημειωθεί ότι τα άρθρα που έχουν επιλεγεί, δημοσιεύτηκαν από το 1992 έως το 2009.

-Από ποιες βάσεις δεδομένων έγινε η επιλογή των απαιτούμενων δεδομένων;

Η επιλογή των δεδομένων έγινε μέσα από τις εξής βάσεις δεδομένων:

- Citation Index(ISI-Web of Knowledge)
- Scopus

-Γιατί επιλέχθηκαν αυτές οι βάσεις δεδομένων;

Η επιλογή των δεδομένων έγινε από τις επιστημονικές βάσεις δεδομένων Scopus και Citation Index(ISI-Web of Knowledge) γιατί είναι βάσεις δεδομένων με μεγάλη ερευνητική βιβλιογραφία και πολλά ποιοτικά δεδομένα από το διαδίκτυο. Αυτές οι βάσεις είναι σχεδιασμένες για να βρίσκουν τις πληροφορίες που χρειάζονται οι επιστήμονες και είναι γρήγορες, εύκολες και ολοκληρωμένες. Επιπλέον, αυτές οι βάσεις δεδομένων ενημερώνονται καθημερινά και έτσι ο όγκος των δεδομένων τους καθημερινώς αυξάνεται. Και οι δυο αυτές βάσεις δεδομένων καλύπτουν ένα ευρύ φάσμα επιστημονικών δεδομένων που ψάχνουμε να βρούμε, καλύπτουν πολλούς τομείς και είναι περιεκτικές. Επίσης, παρέχουν και τη δυνατότητα έρευνας με βάση τον οργανισμό από τον οποίο προέρχεται ο συγγραφέας του κάθε άρθρου. Στη συγκεκριμένη περίπτωση, οργανισμός θα είναι το Πανεπιστήμιο Κύπρου.

Να σημειωθεί ότι η βάση δεδομένων Scopus είναι μια πρόσφατη βάση δεδομένων αφού δημιουργήθηκε στα τέλη της δεκαετίας του '80 και μια ευρέως γνωστή βάση δεδομένων. Αποτελεί μια μεγάλη πηγή από περιοδικά αφού πάρα πολλά άρθρα τα οποία έχουν δημοσιευτεί σε επιστημονικά περιοδικά βρίσκονται σε αυτή τη βάση σε ηλεκτρονική μορφή.

Επιπλέον, να σημειωθεί ότι η βάση δεδομένων Citation Index (ISI-Web of Knowledge) έχει δημιουργηθεί το 1973 και είναι μια ευρέως γνωστή βάση δεδομένων. Αυτή η βάση δεδομένων είναι μια βάση δεδομένων με παραθέματα (citations), η οποία παρέχει μια πολυεπιστημονική κάλυψη με περισσότερο από

10,000 περιοδικά με μεγάλο αντίκτυπο(high-impact journals) και παρέχει μια διεθνή κάλυψη πρακτικών για περισσότερες από 120,000 διασκέψεις(conferences). Παρέχει πρόσβαση σε πολλές βάσεις δεδομένων και άλλους πόρους και καλύπτει περισσότερα από 11.000 κορυφαία έντυπα στην επιστήμη , την τεχνολογία , τις κοινωνικές επιστήμες , τέχνες και ανθρωπιστικές επιστήμες.

Παρατηρήσεις:

- Επειδή η αναζήτηση δεδομένων και στις δυο περιπτώσεις έγινε με βάση το όνομα οργανισμού (Affiliation Name = University of Cyprus) τότε απαραίτητη προϋπόθεση είναι ότι όλοι οι επιστήμονες του Πανεπιστημίου Κύπρου έχουν δώσει ως όνομα ιδρύματος στο οποίο ανήκουν το Πανεπιστήμιο Κύπρου.
- Να σημειωθεί ότι τα δεδομένα που λήφθηκαν με βάση ως οργανισμό το Πανεπιστήμιο Κύπρου αφορούν όχι μόνο τους καθηγητές του Πανεπιστημίου Κύπρου αλλά και τις τέσσερις βαθμίδες προσωπικού του Πανεπιστημίου Κύπρου.

2.2 Ανάλυση δεδομένων – Στατιστικά Στοιχεία

Έτσι λοιπόν, με βάση το όνομα του ιδρύματος-οργανισμού που ανήκουν οι επιστήμονες του Πανεπιστημίου Κύπρου έγινε η αναζήτηση και λήφθηκαν τα εξής:

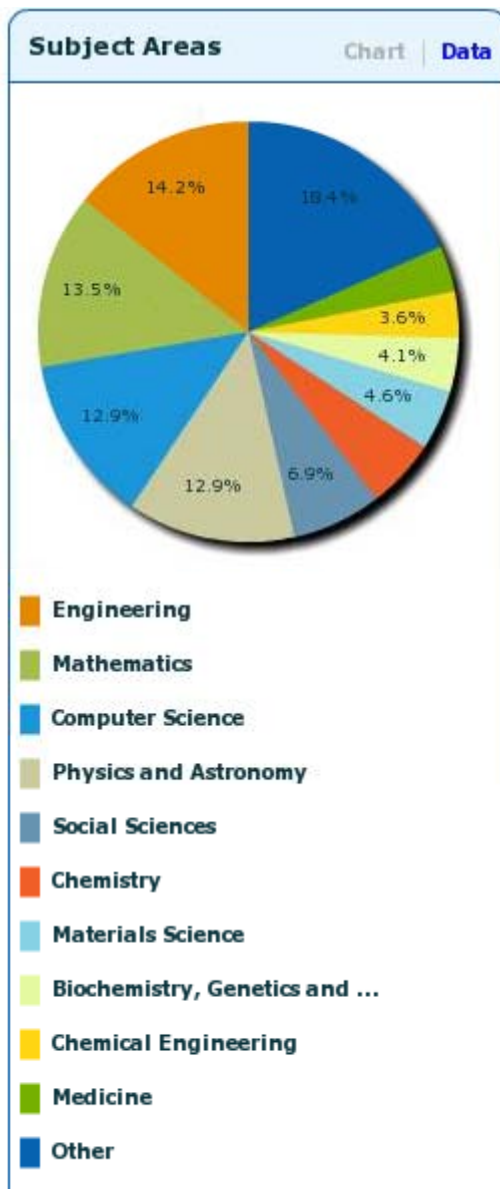
- **3202** δεδομένα από τη βάση δεδομένων Scopus
(Ημερομηνία εξαγωγής δεδομένων: **18-12-09**)
- **2768** δεδομένα από τη βάση δεδομένων Citation Index(ISI-Web of Knowledge)
(Ημερομηνία εξαγωγής δεδομένων: **17-06-10**)

Όπως έχει προαναφερθεί, και οι δυο αυτές βάσεις δεδομένων καλύπτουν πολλούς τομείς. Στην προκειμένη περίπτωση, τα τμήματα που θα μελετηθούν είναι τα τμήματα εκείνα που ανήκουν στη **Σχολή Θετικών και Εφαρμοσμένων Επιστημών**.

Συγκεκριμένα, θα μελετηθούν το τμήμα *Βιολογικών Επιστημών*, το τμήμα *Χημείας* , το τμήμα *Πληροφορικής*, το τμήμα *Μαθηματικών και Στατιστικής* και το τμήμα *Φυσικής*.

Και οι δύο βάσεις δεδομένων που επιλέχθηκαν παρέχουν τη δυνατότητα σε κάθε χρήστη να δει τις επιπλέον πληροφορίες για τον οργανισμό και συγκεκριμένα για ποιους τομείς καλύπτει και με πόσο ποσοστό.

Η βάση δεδομένων Scopus δίνει τις εξής πληροφορίες:



Τμήμα Πληροφορικής → 12,9%
Τμήμα Μαθηματικών
και Στατιστικής → 13,5%
Τμήμα Φυσικής → 12,9%
Τμήμα Βιολογικών Επιστημών → 4,1%
Τμήμα Χημείας → 5,4%

Στατιστικά στοιχεία βάσης δεδομένων
Scopus για άρθρα των επιστημόνων που
ανήκουν στο Πανεπιστήμιο Κύπρου

Η βάση δεδομένων Citation Index(ISI-Web of Knowledge) δίνει τις εξής πληροφορίες:

Field: Subject Area	Record Count	% of 2961	Bar Chart
PHYSICS, MULTIDISCIPLINARY	301	10.1655 %	■
COMPUTER SCIENCE, THEORY & METHODS	176	5.9439 %	■
MATHEMATICS, APPLIED	151	5.0996 %	■
PHYSICS, PARTICLES & FIELDS	149	5.0321 %	■
ENGINEERING, ELECTRICAL & ELECTRONIC	148	4.9983 %	■
ECONOMICS	146	4.9308 %	■
STATISTICS & PROBABILITY	136	4.5930 %	■
CHEMISTRY, PHYSICAL	134	4.5255 %	■
MATHEMATICS	123	4.1540 %	■
PHYSICS, APPLIED	102	3.4448 %	■
EDUCATION & EDUCATIONAL RESEARCH	99	3.3435 %	■
MANAGEMENT	87	2.9382 %	■
MECHANICS	83	2.8031 %	■
MATERIALS SCIENCE, MULTIDISCIPLINARY	81	2.7356 %	■
COMPUTER SCIENCE, INFORMATION SYSTEMS	80	2.7018 %	■
POLYMER SCIENCE	80	2.7018 %	■
COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	76	2.5667 %	■
COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	72	2.4316 %	■
ENGINEERING, CHEMICAL	68	2.2965 %	■
OPERATIONS RESEARCH & MANAGEMENT SCIENCE	67	2.2627 %	■
MATHEMATICS, INTERDISCIPLINARY APPLICATIONS	66	2.2290 %	■
CHEMISTRY, MULTIDISCIPLINARY	64	2.1614 %	■
ASTRONOMY & ASTROPHYSICS	59	1.9926 %	■
PHYSICS, CONDENSED MATTER	58	1.9588 %	■
BIOCHEMISTRY & MOLECULAR BIOLOGY	57	1.9250 %	■

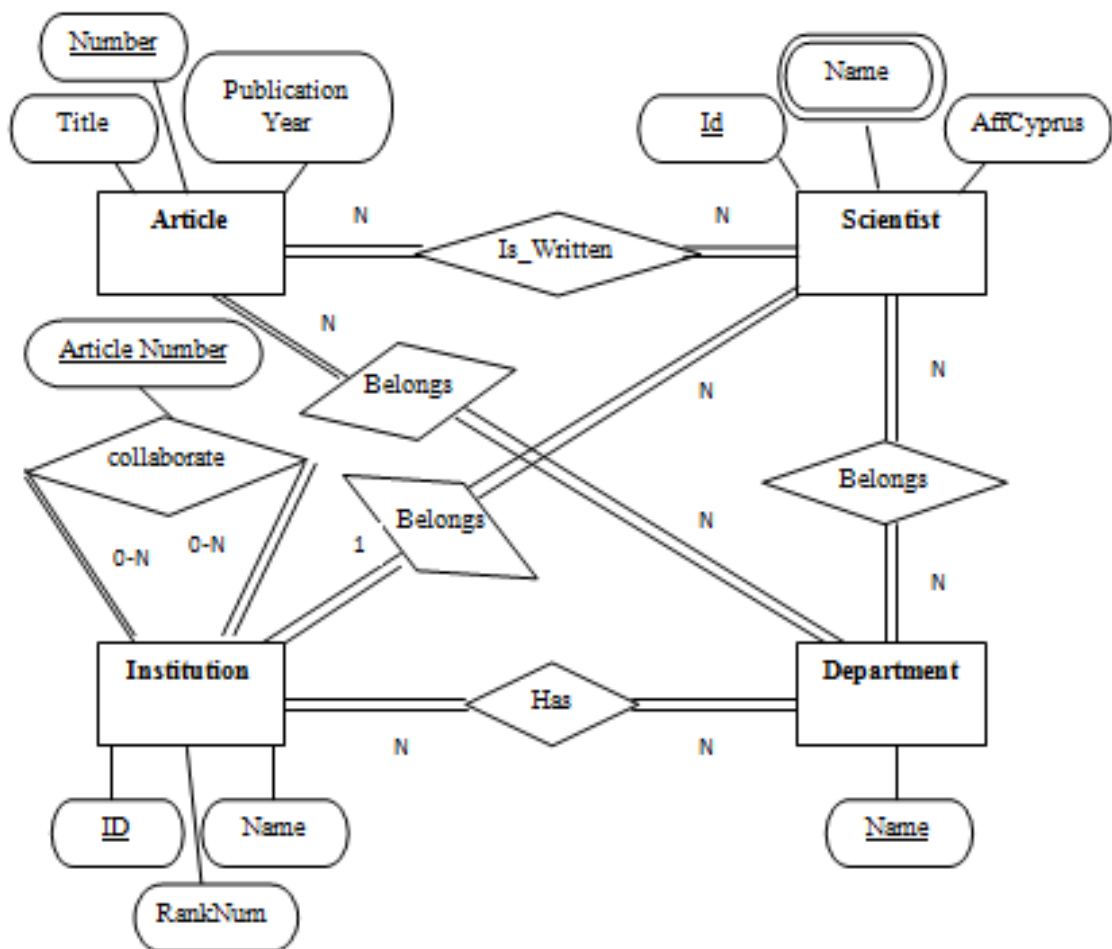
Στατιστικά στοιχεία βάσης δεδομένων Citation Index(ISI-Web of Knowledge)
για άρθρα των επιστημόνων που ανήκουν στο Πανεπιστήμιο Κύπρου

Τμήμα Πληροφορικής	→ 13,6%
Τμήμα Μαθηματικών και Στατιστικής	→ 13,8%
Τμήμα Φυσικής	→ 20,6%
Τμήμα Βιολογικών Επιστημών	→ 1,9%
Τμήμα Χημείας	→ 2,2%

Από τα στατιστικά στοιχεία των δυο αυτών βάσεων δεδομένων παρατηρείτε πως τα πέντε τμήματα που επιλέχθηκαν καλύπτονται επαρκώς. Προφανώς, θα υπάρξουν διπλότυπα αφού παίρνουμε δεδομένα από δυο διαφορετικές βάσεις δεδομένων, άρα πιθανώς κάποιο άρθρο να εμφανίζεται δυο φορές μέσα στα δεδομένα, μια φορά στη

βάση δεδομένων Scopus και μια φορά στη δεδομένων Citation Index (ISI-Web of Knowledge). Γι' αυτό θα χρειαστεί να ελεγχθούν τα δεδομένα που συλλέχθηκαν και από τις δυο βάσεις δεδομένων και τα διπλότυπα να απαλειφθούν μέσω κάποιων scripts. Στη συνέχεια τα δεδομένα θα πρέπει να κατηγοριοποιηθούν έτσι ώστε να είναι γνωστό ποια δεδομένα συσχετίζονται με το κάθε τμήμα.

2.3 Διάγραμμα σχέσης οντοτήτων (ER - diagram)



Σχήμα 2.1: Διάγραμμα σχέσης οντοτήτων

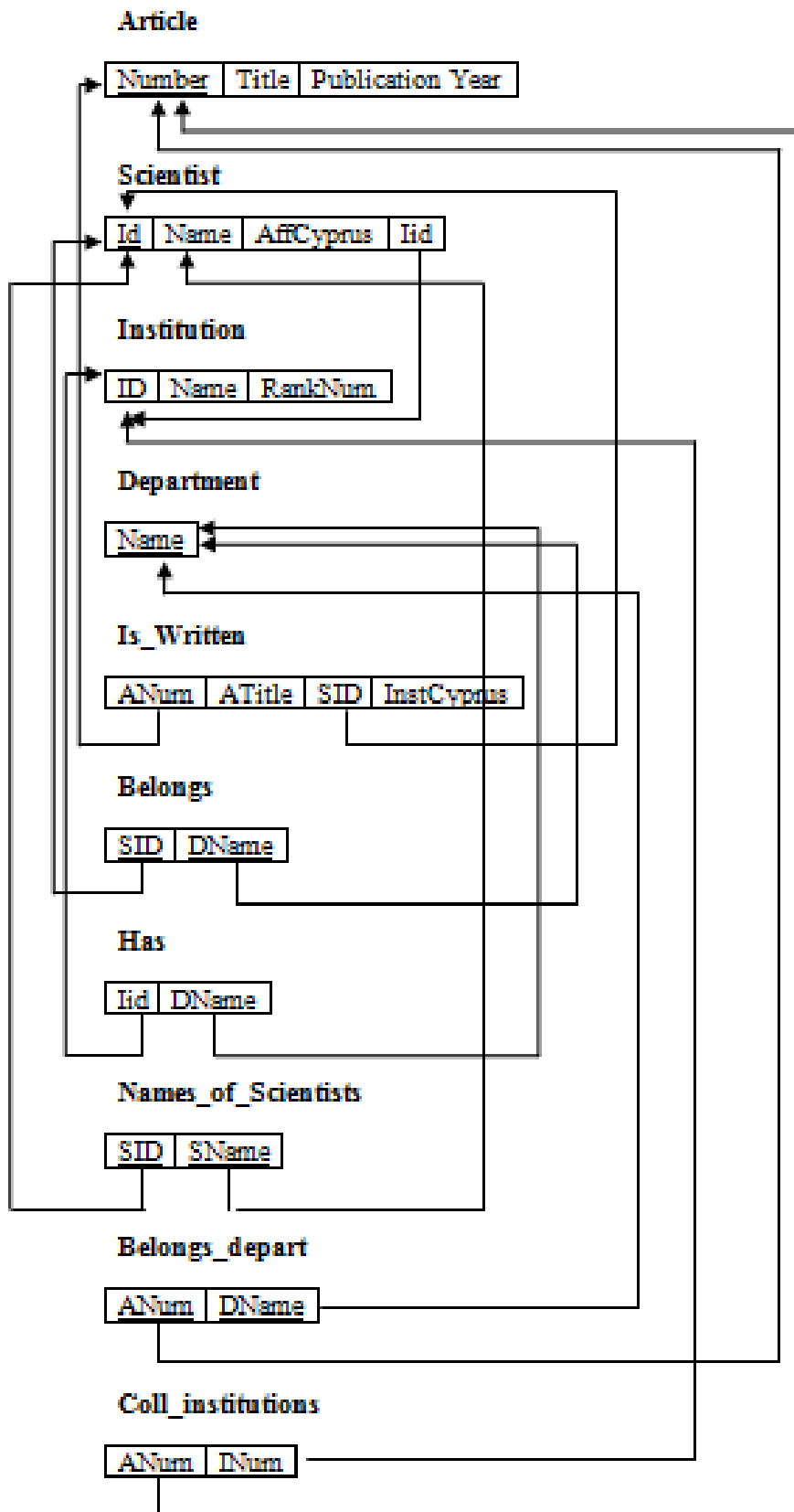
Παρατηρήσεις:

- Ένα άρθρο μπορεί να έχει περισσότερους από ένα συγγραφείς. Μόνο έτσι μπορεί να υπάρξει συνεργασία επιστημόνων για τη συγγραφή ενός άρθρου.

Περιγραφή ERD του σχήματος 2.1:

- Κάθε άρθρο μπορεί να έχει περισσότερους από ένα συγγραφείς και κάθε επιστήμονας μπορεί να συγγράψει πολλά άρθρα.
- Κάθε επιστήμονας μπορεί να ανήκει σε πολλά τμήματα και σε κάθε τμήμα μπορεί να ανήκουν πολλοί επιστήμονες.
- Κάθε επιστήμονας μπορεί να ανήκει σε ένα ίδρυμα-οργανισμό και σε ίδρυμα-οργανισμό μπορεί να ανήκουν πολλοί επιστήμονες.
- Κάθε άρθρο μπορεί να ανήκει σε πολλά τμήματα και σε κάθε τμήμα μπορεί να ανήκουν πολλά άρθρα.
- Κάθε ίδρυμα-οργανισμός μπορεί να έχει πολλά τμήματα και κάθε τμήμα μπορεί να το έχουν πολλά ιδρύματα-οργανισμοί.
- Κάθε ίδρυμα-οργανισμός μπορεί να συνεργάζεται με πολλά ιδρύματα-οργανισμούς.

2.4 Σχεσιακό μοντέλο



Σχήμα 2.2: Σχεσιακό μοντέλο

Περιγραφή σχεσιακού μοντέλου του σχήματος 2.2:

Έχουμε τέσσερις πίνακες που απεικονίζουν τις οντότητες του *article*, *scientist*, *institution*, *department* και έξι πίνακες που απεικονίζουν τις σχέσεις μεταξύ των οντοτήτων στο ERD (βλέπε Σχήμα 2.1).

Πίνακες που απεικονίζουν τις οντότητες του ERD:

-Στον πίνακα *Article* έχουμε ως **πρωτεύον κλειδί** το πεδίο *Number* και ως επιπλέον πεδία τα *Title* και *Publication Year*.

-Στον πίνακα *Scientist* έχουμε ως **πρωτεύον κλειδί** το πεδίο *Id* και ως επιπλέον πεδία τα *Name*, *AffCyprus* και *IName*. Ο πίνακας *Scientist* συσχετίζεται με τον πίνακα *Institution* μέσω του **ξένου κλειδιού** *IName*.

-Στον πίνακα *Institution* έχουμε ως **πρωτεύον κλειδί** το πεδίο *Id* και ως επιπλέον πεδία το *Name* και το *RankNum*.

-Στον πίνακα *Department* έχουμε ως **πρωτεύον κλειδί** το πεδίο *Name*.

Πίνακες που απεικονίζουν τις σχέσεις μεταξύ των οντοτήτων του ERD:

-Στον πίνακα *Is_Written* έχουμε ως **πρωτεύοντα κλειδιά** τα πεδία *ANum* και *SID* και ως επιπλέον πεδία τα *ATitle* και *InstCyprus*. Ο πίνακας αυτός προέκυψε από τη N-N σχέση μεταξύ των οντοτήτων *Article* και *Scientist* και συνεπώς έχει ως πρωτεύοντα κλειδιά το πρωτεύον κλειδί της κάθε οντότητας από την οποία προέκυψε.

-Στον πίνακα *Belongs* έχουμε ως **πρωτεύοντα κλειδιά** τα πεδία *DName* και *SID*. Ο πίνακας αυτός προέκυψε από τη N-N σχέση μεταξύ των οντοτήτων *Department* και *Scientist* και συνεπώς έχει ως πρωτεύοντα κλειδιά το πρωτεύον κλειδί της κάθε οντότητας από την οποία προέκυψε.

-Στον πίνακα *Has* έχουμε ως **πρωτεύοντα κλειδιά** τα πεδία *Iid* και *DName*. Ο πίνακας αυτός προέκυψε από τη N-N σχέση μεταξύ των οντοτήτων *Institution* και *Department* και συνεπώς έχει ως πρωτεύοντα κλειδιά το πρωτεύον κλειδί της κάθε οντότητας από την οποία προέκυψε.

-Στον πίνακα *Names_of_Scientists* έχουμε ως **πρωτεύοντα κλειδιά** τα πεδία *SID* και *SName*. Ο πίνακας αυτός προέκυψε από το γεγονός ότι το χαρακτηριστικό *Name* παίρνει πολλαπλές τιμές (multi-valued) στην οντότητα *Scientist*. Έτσι, πήραμε το

πρωτεύον κλειδί της οντότητας *Scientist* και μαζί με αυτό το χαρακτηριστικό δημιουργήσαμε το νέο αυτό πίνακα.

-Στον πίνακα *Belongs_depart* έχουμε ως **πρωτεύοντα κλειδιά** τα πεδία *A Num* και *DNames*. Ο πίνακας αυτός προέκυψε από τη N-N μεταξύ των οντοτήτων *Article* και *Department* και συνεπώς έχει ως πρωτεύοντα κλειδιά το πρωτεύον κλειδί της κάθε οντότητας από την οποία προέκυψε.

-Στον πίνακα *Coll_institutions* έχουμε ως **πρωτεύοντα κλειδιά** τα πεδία *A Num* και *I Num*. Ο πίνακας αυτός προέκυψε επειδή είναι η οντότητα *Institution* είναι αυτοαναφορική οντότητα.

Κεφάλαιο 3

Εξόρυξη Δεδομένων και Αποθήκευση

3.1 Γενικά	18
3.2 Ανάλυση συστήματος εξόρυξης βασικών δεδομένων	19
3.3 Δεδομένα εισόδου-Scopus	24
3.4 Ανάλυση συστήματος εξόρυξης δεδομένων Scopus	25
3.5 Δεδομένα εισόδου-ISI(Web of Knowledge)	33
3.6 Ανάλυση συστήματος εξόρυξης δεδομένων ISI	35
3.7 Επεξεργασία Δεδομένων Ιδρυμάτων-Οργανισμών	46
3.8 Δεδομένα εξόδου	47

3.1 Γενικά

Η επιλογή των δεδομένων έγινε μέσα από τις βάσεις δεδομένων Scopus και Citation Index(ISI-Web of Knowledge). Αυτή η επιλογή έγινε με βάση το όνομα του ιδρύματος στο οποίο ανήκουν οι επιστήμονες.

Συγκεκριμένα, για τη βάση δεδομένων Scopus ακολουθήθηκαν τα εξής βήματα:

1. Επίσκεψη στην αρχική σελίδα της βάσης δεδομένων Scopus <http://www.scopus.com/home.url>
2. Επιλογή *Affiliation Search*
3. Συμπλήρωση ετικέτας, βάζοντας «University of Cyprus» και πάτημα κουμπιού *Search*
4. Εμφανίζονται τα πέντε πανεπιστήμια της Κύπρου. Επιλογή του πρώτου συνδέσμου ο οποίος αναγράφει «University of Cyprus» και πάτημα κουμπιού *Show documents*
5. Για την εξόρυξη των δεδομένων, επιλέχθηκαν τα δεδομένα ανά χρονιά δημοσίευσης και στη συνέχεια πάτημα κουμπιού *Output*

6. Στην ετικέτα *Export Format* επιλέγουμε *Text (ASCII format)* και στην ετικέτα *Output* επιλέγουμε *Specified fields to be Exported*.
7. Επιλέγουμε *Author(s)*, *Document title* και *Affiliations* και πατάμε το κουμπί *Export*

Για τη βάση δεδομένων Citation Index (ISI-Web of Knowledge) ακολουθήθηκαν τα εξής βήματα:

1. Επίσκεψη στη σελίδα με όλες τις βάσεις δεδομένων που παρέχει η βιβλιοθήκη του Πανεπιστημίου Κύπρου http://library.ucy.ac.cy/electronic_services/dbases_atoz_gr.htm
2. Επιλογή της βάσης δεδομένων Citation Index που βρίσκεται στο section C
3. Επιλογή *Advanced Search*
4. Στην ετικέτα γράφουμε το εξής: **OG = Univ Cyprus**
5. Επιλογή του αριθμού που βρίσκεται στη στήλη Results
6. Επιλογή χρονιά δημοσίευσης άρθρου (π.χ. 2009) και πάτημα κουμπιού *Refine*
7. Επιλογή *Show 50 per page*
8. Πηγαίνουμε στο τέλος της σελίδας και επιλέγουμε:
 - Step 1 – *All records on page*
 - Step 2 – *Full Record (όχι plus abstract)*
 - Step 3 – *Save to other Reference Software : Save to HTML*
9. Πάτημα κουμπιού *Save*
10. Πάτημα κουμπιού *Save*

3.2 Ανάλυση συστήματος εξόρυξης βασικών δεδομένων

Με τον όρο βασικά δεδομένα εννοούμε τα εξής:

- A. Τα ονόματα των ιδρυμάτων-οργανισμών με τα οποία επιστήμονες του Πανεπιστημίου Κύπρου συνεργάστηκαν με επιστήμονες αυτών των ιδρυμάτων-οργανισμών.
- B. Τα ονοματεπώνυμα των επιστημόνων που ανήκουν στο Πανεπιστήμιο Κύπρου.
- C. Τα τμήματα στα οποία ανήκουν οι επιστήμονες που ανήκουν στο Πανεπιστήμιο Κύπρου.

Αυτά τα δεδομένα συλλέχθηκαν από τη βάση δεδομένων Scopus αφού μας προσφέρει τη δυνατότητα και να δούμε περισσότερες λεπτομέρειες για το συγκεκριμένο ίδρυμα-οργανισμό. Στην προκειμένη περίπτωση για το Πανεπιστήμιο Κύπρου. Συγκεκριμένα από τον εξής σύνδεσμο:

<http://www.scopus.com/affil/profile.url?afid=60071343&offset=1&sid=xniPCZkkFbl8J-ENI6bXCsf%3a50&origin=AffiliationNamesList&txGid=xniPCZkkFbl8J-ENI6bXCsf%3a7>

A. Τα ονόματα των ιδρυμάτων-οργανισμών με τα οποία επιστήμονες του Πανεπιστημίου Κύπρου συνεργάστηκαν με επιστήμονες αυτών των ιδρυμάτων-οργανισμών είναι αποθηκευμένα σε αρχείο κειμένου, στο οποίο σε κάθε γραμμή υπάρχει το όνομα ενός ιδρύματος .

Ακολουθεί ο ψευδοκώδικας για την εισαγωγή των ονομάτων ιδρυμάτων-οργανισμών στον πίνακα institution της τοπικής βάσης δεδομένων.

```
<?php
//syndesi me MySQL Database
$s = mysql_connect(servername, username, password);

mysql_select_db("collaboration", $s); //epilogi vasis dedomenon

$file = fopen("universities.txt", "r");

while(!feof($file))
{
    $inst = fgets($file);

    //insert data into table institution
    mysql_query("INSERT INTO institution (Name) VALUES ('$inst')");
}

fclose($file); //close file

mysql_close($s); //kleisimo syndesis
?>
```

Αρχικά, σε κάθε script γίνεται η σύνδεση με τη βάση δεδομένων δίνοντας το όνομα του server καθώς επίσης και το όνομα του χρήστη και το συνθηματικό του (κωδικό

πρόσβασης). Στη συνέχεια, επιλέγεται η βάση δεδομένων που θα διαχειριστεί και ανοίγεται το αρχείο κειμένου που περιέχει τα ονόματα των ιδρυμάτων-οργανισμών. Το αρχείο κειμένου διαβάζεται από την πρώτη γραμμή του μέχρι την τελευταία. Στον πίνακα *institution* προστίθεται το όνομα του ιδρύματος που βρίσκεται στη συγκεκριμένη γραμμή. Στο τέλος, το αρχείο κλείνει, όπως επίσης και η σύνδεση με την MySQL βάση δεδομένων.



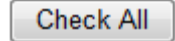

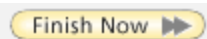

- B.** Τα ονοματεπώνυμα των επιστημόνων που ανήκουν στο Πανεπιστήμιο Κύπρου.
- C.** Τα τμήματα στα οποία ανήκουν οι επιστήμονες που ανήκουν στο Πανεπιστήμιο Κύπρου.

Να σημειωθεί πως η βάση δεδομένων Scopus στις πλείστες περιπτώσεις για κάθε επιστήμονα έδινε περισσότερα από ένα τμήματα ή δεν έδινε όλα τα πιθανά ονοματεπώνυμα που αντιστοιχούν στο ίδιο καθεαυτό άτομο. Έτσι, για να διευκρινιστεί σε ποιο τμήμα ακριβώς ανήκει ο κάθε επιστήμονας του Πανεπιστημίου Κύπρου και για να έχουμε όλους τους τρόπους με τους οποίους έδωσε το όνομά του ο κάθε επιστήμονας, χρησιμοποιήθηκε η λειτουργία **Author Search** από τη βάση δεδομένων Scopus και η λειτουργία **Author Finder** από τη βάση δεδομένων *Citation Index (ISI-Web of Knowledge)*. Τα βήματα που ακολουθήθηκαν για την κάθε βάση δεδομένων είναι τα εξής:

Scopus:

1. Επίσκεψη στην αρχική σελίδα της βάσης δεδομένων Scopus <http://www.scopus.com/home.url>
2. Επιλογή *Author Search*
3. Συμπλήρωση ονοματεπώνυμου του επιστήμονα στην ετικέτα *Author*
4. Συμπλήρωση ονόματος του ιδρύματος, στο οποίο ανήκει ο επιστήμονας στην ετικέτα *Affiliation*. Στην προκειμένη περίπτωση «*University of Cyprus*»
5. Πάτημα κουμπιού *Search*
6. Επιλογή του επιστήμονα (αριστερά από το ονοματεπώνυμο του)
7. Για κάθε άρθρο, επιλογή του κουμπιού **Abstract + Refs**

Citation Index (ISI-Web of Knowledge):

1. Επίσκεψη στη σελίδα με όλες τις βάσεις δεδομένων που παρέχει η βιβλιοθήκη του Πανεπιστημίου Κύπρου http://library.ucy.ac.cy/electronic_services/dbases_atoz_gr.htm
2. Επιλογή της βάσης δεδομένων Citation Index που βρίσκεται στο section C
3. Επιλογή **Author Finder** που βρίσκεται στη κέντρο της σελίδας
4. Συμπλήρωση στην ετικέτα το ονοματεπώνυμο του επιστήμονα
5. Πάτημα κουμπιού 
6. Πάτημα κουμπιού 
7. Πάτημα κουμπιού 
8. Πάτημα κουμπιού 
9. Επιλογή ιδρύματος **UNIV CYPRUS**
10. Πάτημα κουμπιού 
11. Για κάθε άρθρο, επιλογή του κουμπιού 

Μετά το «καθάρισμα» των δεδομένων για τους επιστήμονες του Πανεπιστημίου Κύπρου, τα ονοματεπώνυμα και τα τμήματα αυτών των επιστημόνων αποθηκεύτηκαν σε αρχεία τύπου κειμένου.

Στο αρχείο με τα ονόματα των επιστημόνων, σε κάθε γραμμή έχουμε το όνομα/τα ονόματα του κάθε επιστήμονα (δηλαδή τους διάφορους τρόπους με τους οποίους είναι γραμμένοι οι συγκεκριμένοι επιστήμονες και αναφέρονται στο ίδιο καθαυτό άτομο). Όταν ένας επιστήμονας έχει περισσότερα από ένα ονόματα τότε όλα τα ονόματα βρίσκονται σε μια γραμμή και διαχωρίζονται από ένα ελληνικό ερωτηματικό.

Παράδειγμα:

Achilleos, Evdokia C.; Achilleos, E. C.; Achilleos, E.C.; Achilleos, Evdokia

Όλα τα πιο πάνω ονόματα αναφέρονται σε ένα άτομο, στην Αχιλλέως Ευδοκία.

Στο αρχείο με τα τμήματα στα οποία ανήκει ο κάθε επιστήμονας, σε κάθε γραμμή έχουμε το τμήμα του κάθε επιστήμονα.

Στην επόμενη σελίδα ακολουθεί ο ψευδοκώδικας για την εισαγωγή των ονομάτων των επιστημόνων του Πανεπιστημίου Κύπρου και των τμημάτων στα οποία ανήκουν.

```

<?php
$s = mysql_connect(servername, username, password);
mysql_select_db("collaboration", $s);

$file1 = fopen("AuthorsNames.txt","r");
while(!feof($file1)){
    $inst1 = fgets($file1);
    mysql_query("INSERT INTO scientist (Name, AffCyprus, IName ) VALUES
('$inst1', '1', 'University of Cyprus')");
}

$file2 = fopen("AuthorsNames.txt","r");
while(!feof($file2)){
    $inst2 = fgets($file2);
    $array1 = explode(";", $inst2);
    while(1){
        mysql_query("INSERT INTO names_of_scientists (SID, SName) VALUES
('$count1', '$array1[$c]')");
        $c++;
        if($array1[$c] == NULL) break;
        $count1++; //gia na afkiso to SID
    }
}

$file3 = fopen("AuthorsDeparts.txt","r");
while(!feof($file3)){
    $inst3 = fgets($file3);
    mysql_query("INSERT INTO belongs (SID, DName) VALUES ('$count1', '$inst3
')");
    $count1++; //gia na afkiso to SID
}

$result1 = mysql_query("SELECT DName FROM belongs");
while($row1 = mysql_fetch_array($result1)){
    $result2 = mysql_query("SELECT Name FROM department");
    while($row2 = mysql_fetch_array($result2)){
        $check = strcmp($row1['DName'], $row2['Name']);
        if($check == 0) break;
    }
    if($check != 0){
        if(strcmp($row1[DName], "") != 0 )
            mysql_query("INSERT INTO department (Name) VALUES
('$row1[DName]')");
    }
}

fclose($file1); fclose($file2); fclose($file3);
mysql_close($s);
?>

```

Αρχικά, στο πρώτο κουτί γίνεται η σύνδεση με τη βάση δεδομένων δίνοντας το όνομα του server καθώς επίσης και το όνομα του χρήστη και το συνθηματικό του (κωδικό πρόσβασης) και στη συνέχεια επιλέγεται η βάση δεδομένων που θα διαχειριστούμε. Στο δεύτερο κουτί εισάγονται τα ονόματα των επιστημόνων στον πίνακα *scientists*. Αν κάποιος έχει περισσότερα από ένα ονόματα, σε αυτή τη φάση τα ονόματα δεν έσπασαν ένα-ένα, αλλά μπήκαν όλα σε μια εγγραφή. Στο τρίτο κουτί στον πίνακα *names_of_scientists* εισάγονται τα ονόματα του κάθε επιστήμονα ένα-ένα, αν βέβαια ένας επιστήμονας έχει περισσότερα από ένα ονόματα. Στο τέταρτο κουτί γίνεται η εισαγωγή του τμήματος στο οποίο ανήκει ο κάθε επιστήμονας στον πίνακα *belongs*. **Να σημειωθεί το γεγονός ότι υπάρχει αντιστοιχία των αρχείων AuthorsNames.txt και AuthorsDeparts.txt.** Δηλαδή, ο επιστήμονας με το όνομα στη γραμμή 7 του αρχείου AuthorsNames.txt ανήκει στο τμήμα το οποίο βρίσκεται στη γραμμή 7 του αρχείου AuthorsDeparts.txt. Στο τέταρτο κουτί γίνεται η εισαγωγή των ονομάτων των τμημάτων στον πίνακα *departments*. Στο τέλος, κλείνει η σύνδεση με την MySQL βάση δεδομένων.

3.3 Δεδομένα εισόδου-Scopus

Τα δεδομένα που λήφθηκαν από τη βιβλιοθήκη Scopus είναι αποθηκευμένα σε αρχεία κειμένου και έχουν την εξής μορφή:

```
Scopus
EXPORT DATE: 18 December 2009

Llewellyn, P.L., Theocharis, C.R.
A diffuse reflectance Fourier transform infra-red study of carbon dioxide adsorption on Silicalite-I
.
http://www.scopus.com/inward/record.url?eid=2-s2.0-0026340597&partnerID=40
AFFILIATIONS: University of Cyprus, PO Box 537, Nicosia, Cyprus
```

Οι πρώτες δυο γραμμές κάθε αρχείου κειμένου πάντα είναι το όνομα της βάσης και η ημερομηνία εξόρυξης των δεδομένων. Η τρίτη γραμμή είναι κενή. Στην τέταρτη γραμμή υπάρχουν όλα τα ονόματα των επιστημόνων όσα και αν είναι. Κάθε δυο κόμματα έχουμε ένα ονοματεπώνυμο. Αρχικά το επώνυμο και μετά το όνομα. Στην πέμπτη γραμμή έχουμε τον τίτλο του άρθρου και στην έκτη γραμμή μια τελεία. Στην έβδομη γραμμή υπάρχει το URL, το οποίο παραπέμπει στο συγκεκριμένο άρθρο. Στην

όγδοη γραμμή υπάρχουν τα affiliations, δηλαδή τα ιδρύματα στα οποία ανήκουν οι επιστήμονες οι οποίοι συνεργάστηκαν για τη συγγραφή του συγγραφμένου άρθρου. Τα affiliations δεν βρίσκονται όλα σε μια γραμμή όπως τα ονόματα των επιστημόνων, αλλά από το σημείο στο οποίο υπάρχει η λέξη «AFFILIATIONS» μέχρι την επόμενη κενή γραμμή, κάθε νέα γραμμή αντιστοιχεί σε ένα νέο ίδρυμα. Καταλαμβαίνουμε πως πήραμε όλα τα δεδομένα για το συγκεκριμένο άρθρο μέχρι την επόμενη κενή γραμμή αφού μεταξύ κάθε δυο άρθρων υπάρχει πάντα κενή γραμμή ή όταν φτάσουμε στο τέλος του αρχείου. Αφού συναντήσουμε την επόμενη κενή γραμμή, την θεωρούμε ως τρίτη και ακολουθούμε το ίδιο σκεπτικό.

3.4 Ανάλυση συστήματος εξόρυξης δεδομένων Scopus

Η εξαγωγή των δεδομένων από τη βάση δεδομένων Scopus έγινε με βάση τη χρονιά δημοσίευσης. Άρα, έχουμε 18 αρχεία κειμένου (από το 1992 έως 2009) με τα δεδομένα των άρθρων στα οποία έχουμε τουλάχιστον ένα επιστήμονα ο οποίος ανήκει στο Πανεπιστήμιο Κύπρου. **Απαραίτητη προϋπόθεση** για να την επεξεργασία και να την αποθήκευση των πληροφοριών στη βάση δεδομένων είναι οι συγγραφείς κάθε άρθρου να είναι τουλάχιστον δυο αλλιώς δεν υπάρχει συνεργασία επιστημόνων.

Το script αυτό έχει σαν σκοπό να διαβάσει τα δεδομένα που υπάρχουν στα αρχεία κειμένου, να γίνει η επεξεργασία τους και να προστεθούν στους σωστούς πίνακες της βάσης δεδομένων τα κατάλληλα δεδομένα. Συγκεκριμένα:

1. Στον πίνακα *article* πρέπει να αποθηκευτούν ο αριθμός που ταυτοποιεί μοναδικά το συγκεκριμένο άρθρο, ο τίτλος του άρθρου και η χρονιά δημοσίευσης του.
2. Στον πίνακα *scientist* πρέπει να αποθηκευτούν ο αριθμός που ταυτοποιεί μοναδικά τον συγκεκριμένο επιστήμονα, το όνομα του και η τιμή εκείνη που δείχνει κατά πόσο αυτός ο επιστήμονας ανήκει στο Πανεπιστήμιο Κύπρου.
3. Στον πίνακα *is_written* πρέπει να αποθηκευτούν ο αριθμός που ταυτοποιεί μοναδικά το συγκεκριμένο άρθρο, ο τίτλος του άρθρου, ο αριθμός που ταυτοποιεί μοναδικά τον επιστήμονα που αποτελεί ένα από τους συγγραφείς του συγκεκριμένου άρθρου και η τιμή εκείνη που δείχνει κατά πόσο αυτός ο επιστήμονας ανήκει στο Πανεπιστήμιο Κύπρου.

4. Στον πίνακα *institution* πρέπει να αποθηκευτεί το όνομα του νέου ιδρύματος-οργανισμού.
5. Στον πίνακα *coll_institutions* πρέπει να αποθηκευτούν ο αριθμός που ταυτοποιεί μοναδικά το συγκεκριμένο άρθρο και το όνομα του ιδρύματος-οργανισμού το οποίο 'συμμετείχε' στη συγγραφή του συγκεκριμένου άρθρου.
6. Στον πίνακα *belongs_depart* πρέπει να αποθηκευτούν ο αριθμός που ταυτοποιεί μοναδικά το ένα άρθρο και το όνομα του τμήματος από το οποίο προέρχεται ο επιστήμονας του Πανεπιστημίου Κύπρου.
7. Στον πίνακα *names_of_scientists* πρέπει να αποθηκευτούν ο αριθμός που ταυτοποιεί μοναδικά έναν επιστήμονα και το όνομα αυτού επιστήμονα.

Ανάλυση της κεντρικής ιδέας του script

Το αρχείο κειμένου διαβάζεται από την πρώτη γραμμή του μέχρι την τελευταία. Εάν είμαστε στην πρώτη ή στη δεύτερη τότε αγνοούνται αυτές τις γραμμές γιατί κρατούν πληροφορίες για τη βάση δεδομένων Scopus. Η τρίτη γραμμή είναι πάντα κενή.

Στην τέταρτη γραμμή υπάρχουν όλα τα ονόματα των επιστημόνων που συνεργάστηκαν για τη συγγραφή αυτού του άρθρου. Ελέγχεται αν υπάρχουν τουλάχιστον δυο ονόματα έτσι ώστε να υπάρχει συνεργασία. Αν όχι, τότε αγνοούνται οι πληροφορίες που ακολουθούν, δεν αποθηκεύεται καμιά πληροφορία στη βάση δεδομένων μέχρι την επόμενη κενή γραμμή. Αν υπάρχει συνεργασία τότε αποθηκεύονται τα κατάλληλα δεδομένα στους πίνακες *scientist* και *names_of_scientists*.

Στην πέμπτη γραμμή βρίσκεται ο τίτλος του άρθρου. Αν υπάρχει συνεργασία τότε αποθηκεύονται τα κατάλληλα δεδομένα στους πίνακες *title* και *is_written*.

Στην έκτη και έβδομη γραμμή υπάρχουν πληροφορίες που δεν χρειάζεται να επεξεργαστούν και έτσι αγνοούνται.

Στην όγδοη γραμμή υπάρχουν τα στοιχεία για τους οργανισμούς-ιδρύματα (affiliations). Αυτά τα στοιχεία δεν βρίσκονται όλα σε μια γραμμή αλλά θεωρούμε πως οι γραμμές που ακολουθούν και δεν είναι κενές ανήκουν σε αυτή την κατηγορία. Αν η επόμενη γραμμή είναι κενή τότε ακολουθεί η γραμμή τύπου τέσσερα με τα ονόματα των επιστημόνων του επόμενου άρθρου. Όταν είμαστε στη γραμμή οκτώ και υπάρχει συνεργασία τότε αποθηκεύονται τα κατάλληλα δεδομένα στους πίνακες *coll_institutions*, *belongs_depart* και *institution*.

Ακολουθεί ο ψευδοκώδικας με την κεντρική ιδέα του script.

```
while(!feof($file)){
    $inst = fgets($file);

    if($r <= 3 || $r == 6 || $r == 7){
        $r++;
        continue;
    }

    if($flag1 == 1){
        if($r == 4) {
            if($numOfScientists == 1) {
                echo "No collaboration";
                $flag1 = 0;
            }

            mysql_query("INSERT INTO scientist (Name, AffCyprus) VALUES
('$stringNew', '0')");
            mysql_query("INSERT INTO names_of_scientists (SID, SName) VALUES
('$sidnew', '$stringNew')");
        }
        else if($r == 5){
            mysql_query("INSERT INTO article (Number, Title, PubYear) VALUES
('$anum','$title', '$year')");
            mysql_query("INSERT INTO is_written (ANum, ATitle, SID, InstCyprus)
VALUES ('$anum', '$title', '$r1[SID]', '$r1[AffCyprus]')");
        }
        else if(($r == 8){
            if (lineEmpty() == 0){
                $r=4; continue;
            }

            mysql_query("INSERT INTO coll_institutions (ANum, INum) VALUES
('$anum', '$sidTable[$r]')");
            mysql_query("INSERT INTO belongs_depart (ANum, DName) VALUES
('$anum', '$depart[$r]')");
            mysql_query("INSERT INTO institution (Name) VALUES ('$affiliation')");
        }
    }
    else {
        if(($r == 8){
            if (lineEmpty() == 0 ) $r=4;
        }
    }
}
```

Διεξοδική περιγραφή του script:

Το αρχείο κειμένου διαβάζεται από την πρώτη γραμμή του μέχρι την τελευταία.

Εάν είμαστε στην πρώτη γραμμή του αρχείου, τότε καλείται η συνάρτηση **setInstitutions()** στην οποία αρχικοποιείται ο πίνακας *institutions*. Ο πίνακας αυτός έχει 26 θέσεις στις οποίες υπάρχει ένας πίνακας, ο οποίος αντιστοιχεί σε ένα γράμμα τις αγγλικής αλφαβήτου(A-Z). Στον κάθε πίνακα αποθηκεύονται οι αριθμοί εκείνοι που ταυτοποιούν μοναδικά το ίδρυμα που ξεκινά από το συγκεκριμένο γράμμα. Για παράδειγμα, στον πίνακα *institutions* στη θέση μηδέν υπάρχει ο πίνακας με τα ids των ιδρυμάτων που το όνομά τους ξεκινά από 'A'.

Στη συνέχεια, καλείται η συνάρτηση **setSidLet()** κατά την οποία αρχικοποιείται ο πίνακας *letters*. Ο πίνακας αυτός έχει 26 θέσεις στις οποίες υπάρχει ένας πίνακας, ο οποίος αντιστοιχεί σε ένα γράμμα τις αγγλικής αλφαβήτου(A-Z). Στη συνάρτηση *setSidLet()*, βρίσκονται οι αριθμοί που ταυτοποιούν μοναδικά τους επιστήμονες και τους οποίους το επίθετό τους ξεκινά από συγκεκριμένο γράμμα και αυτός ο αριθμός προστίθεται στον αντίστοιχο υποπίνακα.

Τα δεδομένα που βρίσκονται στην πρώτη γραμμή και στη δεύτερη αγνοούνται γιατί κρατούν πληροφορίες για τη βάση δεδομένων Scopus που δεν χρειάζεται να επεξεργαστούν .

Η τρίτη γραμμή είναι πάντα κενή. Κάθε φορά που υπάρχει κενή γραμμή θεωρούμε ότι είμαστε στην τρίτη γραμμή. Η γραμμή αυτή είναι το σημάδι που μας δίνεται για να καταλάβουμε ότι ακολουθεί νέο άρθρο το οποίο πρέπει να επεξεργαστούμε τα δεδομένα του.

Εάν είμαστε στην τέταρτη γραμμή τότε σε αυτήν τη γραμμή υπάρχουν όλα τα ονοματεπώνυμα των επιστημόνων που συνεργάστηκαν για τη συγγραφή του συγκεκριμένου άρθρου. Αρχικά, γίνεται έλεγχος αν υπάρχει συνεργασία. Αυτό διαπιστώνεται αν υπάρχουν τουλάχιστον δυο ονοματεπώνυμα. Όλα τα ονοματεπώνυμα τα παίρνουμε ως ένα string, το οποίο το σπάει σε μικρότερες συμβολοσειρές με βάση την εμφάνιση συγκεκριμένου χαρακτήρα ', '. Είναι γνωστό το γεγονός ότι ένα πλήρες

ονοματεπώνυμο ανακτάται παίρνοντας δυο διαδοχικά strings αφού μετά από κάθε επίθετο υπάρχει κόμμα, ακολουθεί το όνομα, έπειτα κόμμα κοκ.

Αν δεν υπάρχει συνεργασία τότε οι επόμενες γραμμές αγνοούνται μέχρι την επόμενη κενή γραμμή η οποία σηματοδοτεί την αρχή των δεδομένων του επόμενου άρθρου.

Αν τώρα υπάρχει συνεργασία, τότε καλείται η συνάρτηση **CommaOccurrencesFunction(\$inst)**. Η συνάρτηση αυτή παίρνει ως παράμετρο τη συμβολοσειρά που κρατά όλα τα ονοματεπώνυμα των επιστημόνων που συνεργάστηκαν για τη συγγραφή του συγκεκριμένου άρθρου και επιστρέφει ένα πίνακα, ο οποίος σε κάθε του θέση έχει ολόκληρο το ονοματεπώνυμο του κάθε επιστήμονα. Στη συνέχεια, παίρνουμε ένα-ένα τα ονοματεπώνυμα που υπάρχουν στον πίνακα που επέστρεψε η συνάρτηση CommaOccurrencesFunction(\$inst) και γίνονται κάποιιοι ελέγχοι προτού αποφανθεί το πρόγραμμα αν θα αποθηκεύσει τα ονόματα στη βάση δεδομένων.

Αρχικά, επιλέγεται το πρώτο γράμμα του ονοματεπώνυμου, αν είναι πεζό τότε μετατρέπεται σε κεφαλαίο και επιλέγεται ο πίνακας του συγκεκριμένου γράμματος που περιέχει τις 'ταυτότητες' των επιστημόνων που τα επίθετά τους ξεκινούν από αυτό το γράμμα. Επιλέγονται μία-μία οι 'ταυτότητες' των επιστημόνων και επιλέγονται τα ονοματεπώνυμα που έχει ο επιστήμονας με τη συγκεκριμένη ταυτότητα από τον πίνακα *names_of_scientists* (Υπενθύμιση: ένας επιστήμονας μπορεί να έχει πολλά ονόματα, δηλαδή το όνομά του μπορεί να γράφεται με πολλούς τρόπους και να όλα τα ονόματα να αναφέρονται σε αυτό το άτομο). Έπειτα, ελέγχεται αν η συμβολοσειρά που έχουμε υπάρχει μέσα σε μια εγγραφή στον πίνακα *names_of_scientists*. Αν η συμβολοσειρά που έχουμε δεν είναι ίδια με το ονοματεπώνυμο εκείνης της εγγραφής του πίνακα, τότε επιλέγουμε την επόμενη εγγραφή, αλλιώς είναι γνωστό ότι υπάρχει αυτός ο επιστήμονας ήδη στον πίνακα *names_of_scientists*. Τώρα, αν ο επιστήμονας υπάρχει ήδη στον πίνακα τότε την 'ταυτότητα' αυτού του επιστήμονα προστίθεται στον πίνακα που κρατά τις 'ταυτότητες' των επιστημόνων που συνεργάστηκαν για τη συγγραφή του συγκεκριμένου άρθρου και προσθέτω ένα ακόμη στοιχείο στον πίνακα που κρατά τα ονοματεπώνυμα των επιστημόνων που συνεργάστηκαν για τη συγγραφή του συγκεκριμένου άρθρου.

Αν ο επιστήμονας αυτός δεν υπάρχει σε καμία εγγραφή από τις εγγραφές που έχουν επιλεγεί, τότε πρέπει να αποθηκευτούν τα στοιχεία αυτού του επιστήμονα στους δυο πίνακες της βάσης δεδομένων που κρατούν στοιχεία για τους επιστήμονες. Συγκεκριμένα, στον πίνακα *scientist* αποθηκεύονται το ονοματεπώνυμο του επιστήμονα και η τιμή που δηλώνει ότι δεν ανήκει στο Πανεπιστήμιο Κύπρου, καθώς επίσης στον πίνακα *names_of_scientists* αποθηκεύονται η 'ταυτότητα' του επιστήμονα και το ονοματεπώνυμο του επιστήμονα. Τέλος, σε αυτή τη φάση, ενημερώνεται ο πίνακας *letters* προσθέτοντας την 'ταυτότητα' του συγκεκριμένου επιστήμονα στο σωστό πίνακα που αντιστοιχεί στο γράμμα που ξεκινά το ονοματεπώνυμο του επιστήμονα.

Εάν είμαστε στην πέμπτη γραμμή τότε σε αυτήν τη γραμμή υπάρχει ο τίτλος του άρθρου. Στον πίνακα *article* προστίθονται ο αριθμός ταυτοποίησης του άρθρου, ο τίτλος και η χρονιά δημοσίευσης του άρθρου. Στη συνέχεια, ενημερώνεται και ο πίνακας *is_written*. Επιλέγεται μια-μια 'ταυτότητα' των επιστημόνων που συνεργάστηκαν για αυτό το άρθρο από τον πίνακα που κρατά αυτά τα στοιχεία και προστίθεται ο αριθμός ταυτοποίησης του άρθρου, ο τίτλος του άρθρου, η 'ταυτότητα' του επιστήμονα που είναι ένας από το \mathfrak{u} επιστήμονες που συνεργάστηκε για τη συγγραφή του συγκεκριμένου άρθρου και αν ανήκει στους επιστήμονες που ανήκουν στο Πανεπιστήμιο Κύπρου.

Εάν είμαστε στην όγδοη γραμμή τότε σε αυτήν τη γραμμή υπάρχουν οι πληροφορίες για τα ιδρύματα-οργανισμούς που συνεργάστηκαν για τη συγγραφή αυτού του άρθρου. Αυτές οι πληροφορίες δεν βρίσκονται όλες σε μια γραμμή, αλλά σε κάθε γραμμή που ακολουθεί υπάρχουν οι πληροφορίες για το κάθε ίδρυμα στο οποίο ανήκει ένας επιστήμονας, ο οποίος είναι ένας από τους συγγραφείς. Να σημειωθεί ότι μπορεί περισσότεροι από ένα επιστήμονες από κάποιο ίδρυμα να συνεργάστηκαν για το ίδιο άρθρο και το όνομα του ιδρύματος να αναφέρεται μόνο μια φορά, συνεπώς δεν μπορεί να γίνει αντιστοίχιση επιστήμονα και ιδρύματος. Καταλαμβάνουμε ότι είμαστε στην όγδοη γραμμή όταν εμφανίζεται η ετικέτα 'AFFILIATIONS' και υπάρχουν οι σχετικές πληροφορίες μέχρι να εμφανιστεί μια κενή γραμμή.

Αρχικά, ελέγχεται κατά πόσο υπάρχει κενή γραμμή.

Αν όχι, τότε υπάρχουν πληροφορίες για κάποιο ίδρυμα. Αν βρισκόμαστε στη γραμμή με την ετικέτα τότε πρέπει να αφαιρεθεί προκειμένου να απομονωθούν οι πληροφορίες για το πρώτο ίδρυμα. Παρατηρείτε ότι πάντα σε κάθε γραμμή που υπάρχουν πληροφορίες για το ίδρυμα-οργανισμό, το όνομα του ιδρύματος βρίσκεται κάπου μεταξύ των τριών πρώτων κομμάτων. Άρα, θα χρειαστεί να σπάσει η συμβολοσειρά σε μικρότερες συμβολοσειρές κάθε φορά που συναντούμε το χαρακτήρα ‘,’ (κόμμα). Οι τρεις πρώτες μικρότερες συμβολοσειρές αποθηκεύονται σε τρεις διαφορετικές μεταβλητές. Για να ξεχωρίσουμε σε ποια από τις τρεις μεταβλητές υπάρχει το όνομα του ιδρύματος-οργανισμού αρκεί να γίνει έλεγχος σε ποια μεταβλητή υπάρχει ένα από τα εξής substrings:

- Univ (Universita, University, Univ., Universidad etc)
- Inst (Inst., Institute, Institut)
- School

Υπάρχουν και κάποιες εξαιρέσεις: CIEMAT, CERN, Bulgarian Academy of Sciences.

Αν δεν υπάρχει σε καμιά μεταβλητή κάποιο από τα πιο πάνω substrings τότε πάντα το όνομα του ιδρύματος είναι στην πρώτη μεταβλητή.

Στην περίπτωση που έχουμε ως όνομα ιδρύματος:

- Univ of Cyprus
- Cyprus Univ.
- University of Cyprus

τότε αναφερόμαστε στο Πανεπιστήμιο Κύπρου. Αν υπάρχει τμήμα, θα πρέπει να αποθηκευτεί για να ξέρουμε σε ποιο τμήμα πιθανότατα αναφέρεται το άρθρο. Το όνομα του τμήματος το βρίσκουμε σε κάποια από τις τρεις μεταβλητές, σε αυτή την οποία υπάρχει το substring "Dep". Όταν το βρούμε, τότε το όνομα του τμήματος προστίθεται στον πίνακα που κρατά τα τμήματα του άρθρου.

Στη συνέχεια, καλείται η συνάρτηση **updateInstitution(\$affiliation)**, η οποία παίρνει ως παράμετρο το όνομα του ιδρύματος-οργανισμού και επιστρέφει ένα πίνακα με τους αριθμούς που ταυτοποιούν τα ιδρύματα-οργανισμούς που ‘συνεργάστηκαν’ για τη συγγραφή του συγκεκριμένου άρθρου. Μέσα σε αυτή τη συνάρτηση, στον πίνακα *institution* προστίθεται το νέο ίδρυμα-οργανισμό αν δεν υπάρχει ήδη στον πίνακα. Συγκεκριμένα, τον αριθμό ταυτοποίησης του ιδρύματος και το όνομα του ιδρύματος. Επιπλέον, προστίθεται ένας ακόμη αριθμός που ταυτοποιεί το ίδρυμα στον πίνακα που αντιστοιχεί στο γράμμα από το οποίο ξεκινά το όνομα του ιδρύματος.

Αν ναι, τότε γνωρίζουμε ότι ακολουθεί γραμμή τύπου τέσσερα, δηλαδή ακολουθούν τα ονοματεπώνυμα των επιστημόνων που συνεργάστηκαν για τη συγγραφή του επόμενου άρθρου. Προτού διαβαστεί η επόμενη γραμμή, θα πρέπει να αποθηκευτούν τα δεδομένα που έχουμε πάρει στον πίνακα *coll_institutions* και στον πίνακα *belongs_depart*. Συγκεκριμένα, στον πίνακα *coll_institutions* θα προστεθούν ο αριθμός που ταυτοποιεί το άρθρο και οι αριθμοί που ταυτοποιούν τα ιδρύματα-οργανισμούς και στον πίνακα *belongs_depart* θα προστεθούν ο αριθμός που ταυτοποιεί το άρθρο και τα ονόματα των τμημάτων από τα οποία προέρχεται ο επιστήμονας του Πανεπιστημίου Κύπρου.

Στη συνέχεια, καλείται η συνάρτηση **correspondingDepart()** στην οποία ελέγχεται αν υπάρχει σωστή αντιστοιχία ονόματος επιστήμονα Πανεπιστημίου Κύπρου και τμήματος. Για παράδειγμα, υπάρχει περίπτωση να εμφανίζεται δυο φορές το όνομα 'Αντρέας Αντρέου' αλλά αυτό το όνομα να ανήκει σε δύο διαφορετικά άτομα. Πως μπορούμε να τα ξεχωρίσουμε; Μπορούμε να καταλάβουμε σε ποιο επιστήμονα αναφέρεται το άρθρο από το τμήμα στο οποίο ανήκει ο επιστήμονας. Αν δηλαδή είναι γνωστό ότι ο πρώτος Αντρέας Αντρέου ανήκει στο τμήμα Μαθηματικών και ο δεύτερος στο τμήμα Φυσικής, και ότι σε αυτό το άρθρο ο επιστήμονας του Πανεπιστημίου Κύπρου ανήκει στο τμήμα Φυσικής, τότε θα πρέπει να επιλεγεί η 'ταυτότητα' του δεύτερου επιστήμονα και όχι του πρώτου που επιλέχθηκε στην αρχή. Συνεπώς, στον πίνακα *names_of_scientists* ελέγχονται οι επόμενες 'ταυτότητες' αν έχουν το ίδιο ονοματεπώνυμο μέχρι την επόμενη 'ταυτότητα' στην οποία δεν υπάρχει το ίδιο ονοματεπώνυμο. Αν υπάρχουν ίδια ονοματεπώνυμα, τότε ελέγχονται τα τμήματα των επιστημόνων που έχουν το ίδιο ονοματεπώνυμο με τον επιστήμονα που επιλέχθηκε στην αρχή, για να ελεγχθεί αν το άρθρο γράφτηκε από εκείνο που επιλέχθηκε στην αρχή ή από κάποιο άλλο με το ίδιο ονοματεπώνυμο. Ο έλεγχος αυτός γίνεται κάθε φορά που καλείται η συνάρτηση **belongsFunction(\$sid, \$sid2, \$departName)**. Σε αυτή τη συνάρτηση αν συμπεράνουμε ότι όντως έχει γίνει λάθος επιλογή από την αρχή τότε ενημερώνεται ο πίνακας *is_written* της βάσης δεδομένων.

Έπειτα, αδειάζουμε τον πίνακα που κρατά τους αριθμούς που ταυτοποιούν τα ιδρύματα-οργανισμούς, αδειάζουμε τον πίνακα που κρατά τα τμήματα των επιστημόνων που συνεργάστηκαν για τη συγγραφή αυτού του άρθρου και ανήκουν στο Πανεπιστήμιο Κύπρου και αυξάνουμε τον αριθμό των άρθρων.

3.5 Δεδομένα εισόδου - ISI(Web of Knowledge)

Τα δεδομένα που λήφθηκαν από τη βιβλιοθήκη ISI(Web of Knowledge) είναι αποθηκευμένα σε αρχεία κειμένου και έχουν την εξής μορφή:

FN	ISI Export Format
VR	1.0
PT	J
AU	Kalaitzidakis, P Mamuneas, TP Savvides, A Stengos, T
TI	Measures of human capital and nonlinearities in economic growth
SO	JOURNAL OF ECONOMIC GROWTH
LA	English
DT	Article
DE	cross country growth regressions; human capital; semi-parametric additive linear model
ID	REGRESSION; EMPIRICS
AB	In this paper we study the relationship between human capital accumulation and economic growth using various measures of human capital frequently employed by researchers. We use semiparametric estimation techniques to uncover any nonlinearities that may exist. Using mean years of schooling measures of human capital we find a nonlinear effect on economic growth. There seem to be important differences in the growth effect of educational attainment by gender and level of education. Enrollment rates do not yield a nonlinear effect.
C1	Univ Crete, Rethimnon 74100, Greece. Univ Cyprus, CY-1678 Nicosia, Cyprus. Univ Leicester, Leicester LE1 7RH, Leics, England. Oklahoma State Univ, Stillwater, OK 74078 USA. Univ Guelph, Guelph, ON N1G 2W1, Canada.
RP	Kalaitzidakis, P, Univ Crete, Rethimnon 74100, Greece.
NR	28
TC	22
PU	KLUWER ACADEMIC PUBL
PI	DORDRECHT
PA	SPIBOULEVARD 50, PO BOX 17, 3300 AA DORDRECHT, NETHERLANDS
SN	1381-4338
J9	J ECON GROWTH
JI	J. Econ. Growth
PD	SEP
PY	2001
VL	6
IS	3
BP	229
EP	254
PG	26
SC	Economics
GA	458ZZ
UT	ISI:000170226500004
ER	
EF	

Οι πρώτες δυο γραμμές κάθε αρχείου κειμένου πάντα είναι το όνομα ISI Export Format και το version και η τελευταία γραμμή γράφει EF(end of file). Όπως φαίνεται, κάθε γραμμή ξεκινά με κάποια ετικέτα (tag). Οι ετικέτες με τα δεδομένα που μας ενδιαφέρουν είναι οι εξής:

- **AU** – Από την γραμμή που υπάρχει αυτή η ετικέτα μέχρι να την επόμενη ετικέτα (συνήθως AF ή CA ή TI) ξέρουμε ότι σε κάθε γραμμή έχουμε το ονοματεπώνυμο κάποιου επιστήμονα.
- **AF** – Από την γραμμή που υπάρχει αυτή η ετικέτα μέχρι να την επόμενη ετικέτα (συνήθως CA ή TI) ξέρουμε ότι σε κάθε γραμμή έχουμε ολόκληρο το ονοματεπώνυμο κάποιου επιστήμονα. Αυτή η ετικέτα δεν εμφανίζεται πάντα.
- **TI** – Σε αυτήν την γραμμή υπάρχει το όνομα του άρθρου.
- **C1** – Από την γραμμή που υπάρχει αυτή η ετικέτα μέχρι να την επόμενη ετικέτα ξέρουμε ότι σε κάθε γραμμή που ακολουθεί έχουμε τα ονόματα των ιδρυμάτων-οργανισμών που επιστήμονές τους συνεργάστηκαν για τη συγγραφή αυτού του άρθρου.
- **RP** – Σε αυτήν την ετικέτα υπάρχει το όνομα του ιδρύματος-οργανισμού που ανήκει ο επιστήμονας που δημοσίευσε το άρθρο . Η πληροφορία αυτής της ετικέτας χρησιμοποιείται όταν δεν εμφανίζεται η ετικέτα C1.
- **PY** – Σε αυτήν την γραμμή υπάρχει η χρονιά δημοσίευσης του άρθρου.
- **ER** – Όταν συναντάμε αυτή την ετικέτα, ξέρουμε ότι πήραμε όλα τα δεδομένα για αυτό το άρθρο και ακολουθεί το επόμενο άρθρο. Ουσιαστικά, παίζει το ρόλο της κενής γραμμής που είχαμε στα δεδομένα εισόδου-Scopus.
- **EF** – Όταν συναντάμε αυτήν την ετικέτα γνωρίζουμε πως έχουμε φτάσει στο τέλος του αρχείου κειμένου.

Να σημειωθεί πως δεν υπάρχει σταθερός αριθμός επαναλαμβανόμενων ετικέτων. Απλά είναι γνωστό ότι πριν και μετά τα δεδομένα του κάθε άρθρου υπάρχει η ετικέτα ER.

3.6 Ανάλυση συστήματος εξόρυξης δεδομένων ISI

Η εξαγωγή των δεδομένων από τη βάση δεδομένων ISI(Web of Knowledge) έγινε με βάση τη χρονιά δημοσίευσης. Επειδή σε κάθε σελίδα μπορούσαν να εμφανιστούν το πολύ 50 άρθρα, έτσι έχουμε για κάθε χρονιά τουλάχιστον ένα αρχείο, το οποίο περιέχει πληροφορίες για το πολύ 50 άρθρα. Συνολικά έχουμε 62 αρχεία κειμένου με τα δεδομένα των άρθρων στα οποία υπάρχει τουλάχιστον ένας επιστήμονας ο οποίος ανήκει στο Πανεπιστήμιο Κύπρου. **Απαραίτητη προϋπόθεση** για την επεξεργασία και την αποθήκευση των πληροφοριών στη βάση δεδομένων είναι οι συγγραφείς κάθε άρθρου να είναι τουλάχιστον δυο αλλιώς δεν υπάρχει συνεργασία επιστημόνων.

Το script αυτό έχει σαν σκοπό να διαβάσει τα δεδομένα που υπάρχουν στα αρχεία κειμένου, να τα επεξεργαστεί και να βάλει στους σωστούς πίνακες της βάσης δεδομένων τα κατάλληλα δεδομένα. Συγκεκριμένα:

1. Στον πίνακα *article* πρέπει να αποθηκευτούν ο αριθμός που ταυτοποιεί μοναδικά το συγκεκριμένο άρθρο, ο τίτλος του άρθρου και η χρονιά δημοσίευσης του.
2. Στον πίνακα *scientist* πρέπει να αποθηκευτούν ο αριθμός που ταυτοποιεί μοναδικά τον συγκεκριμένο επιστήμονα, το όνομα του και η τιμή εκείνη που δείχνει κατά πόσο αυτός ο επιστήμονας ανήκει στο Πανεπιστήμιο Κύπρου.
3. Στον πίνακα *is_written* πρέπει να αποθηκευτούν ο αριθμός που ταυτοποιεί μοναδικά το συγκεκριμένο άρθρο, ο τίτλος του άρθρου, ο αριθμός που ταυτοποιεί μοναδικά τον επιστήμονα που αποτελεί ένα από τους συγγραφείς του συγκεκριμένου άρθρου και η τιμή εκείνη που δείχνει κατά πόσο αυτός ο επιστήμονας ανήκει στο Πανεπιστήμιο Κύπρου.
4. Στον πίνακα *names_of_scientists* πρέπει να αποθηκευτούν ο αριθμός που ταυτοποιεί μοναδικά έναν επιστήμονα και το όνομα αυτού επιστήμονα.
5. Στον πίνακα *institution* πρέπει να αποθηκευτούν ο αριθμός που ταυτοποιεί ένα ίδρυμα-οργανισμό, καθώς επίσης και το όνομα αυτού του ιδρύματος-οργανισμού.
6. Στον πίνακα *coll_institutions* πρέπει να αποθηκευτούν ο αριθμός που ταυτοποιεί ένα άρθρο και το όνομα του ιδρύματος-οργανισμού που συμμετείχε στη συγγραφή αυτού του άρθρου.

Στόχος αυτού του script είναι αν υπάρχουν άρθρα τα οποία δεν υπήρχαν στη βάση δεδομένων Scopus να προστεθούν για να έχει η βάση δεδομένων όσο το δυνατόν περισσότερα άρθρα για την κάθε χρονιά. Έτσι, πρέπει να γίνει έλεγχος αν υπάρχουν διπλότυπα, να μην προστεθούν αλλά μόνο τα άρθρα εκείνα που δεν υπάρχουν ήδη στη βάση δεδομένων.

Ανάλυση της κεντρικής ιδέας του script

Το αρχείο κειμένου διαβάζεται από την πρώτη γραμμή του μέχρι την τελευταία. Υπάρχουν πολλές και διαφορετικές ετικέτες, αλλά αυτές που μας ενδιαφέρουν είναι μόνο επτά. Συγκεκριμένα:

- **AU** – Από αυτήν την ετικέτα μέχρι την επόμενη έχουμε όλα τα ονοματεπώνυμα των επιστημόνων που συνεργάστηκαν για τη συγγραφή του συγκεκριμένου άρθρου. Όταν συναντάμε αυτήν την ετικέτα, τότε είμαστε στη γραμμή τύπου ένα.
- **AF** – Από αυτήν την ετικέτα μέχρι την επόμενη έχουμε όλα τα πλήρη ονοματεπώνυμα των επιστημόνων που συνεργάστηκαν για τη συγγραφή του συγκεκριμένου άρθρου. Όταν συναντάμε αυτήν την ετικέτα, τότε είμαστε στη γραμμή τύπου δύο.
- **TI** – Σε αυτήν την ετικέτα έχουμε το όνομα του άρθρου. Όταν συναντάμε αυτήν την ετικέτα, τότε είμαστε στη γραμμή τύπου τρία.
- **C1** – Από την γραμμή που υπάρχει αυτή η ετικέτα μέχρι να την επόμενη
- ετικέτα ξέρουμε ότι σε κάθε γραμμή που ακολουθεί έχουμε τα ονόματα των ιδρυμάτων-οργανισμών που επιστήμονές τους συνεργάστηκαν για τη συγγραφή αυτού του άρθρου. Όταν συναντάμε αυτήν την ετικέτα, τότε είμαστε στη γραμμή τύπου τέσσερα.
- **RP** – Σε αυτήν την ετικέτα υπάρχει το όνομα του ιδρύματος-οργανισμού που ανήκει ο επιστήμονας που δημοσίευσε το άρθρο. Όταν συναντάμε αυτήν την ετικέτα, τότε είμαστε στη γραμμή τύπου πέντε.
- **PY** – Σε αυτήν την ετικέτα έχουμε τη χρονιά δημοσίευσης του άρθρου. Όταν συναντάμε αυτήν την ετικέτα, τότε είμαστε στη γραμμή τύπου έξι.
- **ER** – Όταν συναντάμε αυτή την ετικέτα, ξέρουμε ότι πήραμε όλα τα δεδομένα για αυτό το άρθρο και ακολουθεί το επόμενο άρθρο. Επιπλέον, γνωρίζουμε ότι είμαστε στη γραμμή τύπου επτά.

Εάν είμαστε σε οποιαδήποτε άλλου τύπου γραμμή εκτός από αυτές τις επτά τότε οι πληροφορίες που βρίσκονται σε αυτές τις γραμμές απλά αγνοούνται.

Στην γραμμή τύπου ένα υπάρχουν τα ονόματα των επιστημόνων που συνεργάστηκαν για τη συγγραφή αυτού του άρθρου. Τα ονόματα αυτά αν είναι περισσότερα από ένα τότε δεν βρίσκονται σε μια γραμμή, αλλά σε κάθε γραμμή που ακολουθεί μέχρι την επόμενη ετικέτα υπάρχει ένα ονοματεπώνυμο που αντιστοιχεί σε ένα επιστήμονα.

Στην γραμμή τύπου δύο υπάρχουν τα πλήρη ονόματα των επιστημόνων που συνεργάστηκαν για τη συγγραφή αυτού του άρθρου. Αυτή η τύπο υ γραμμή δεν βρίσκεται πάντα σε όλα τα αρχεία κειμένου. Τα ονόματα αυτά αν είναι περισσότερα από ένα τότε δεν βρίσκονται σε μια γραμμή, αλλά σε κάθε γραμμή που ακολουθεί μέχρι την επόμενη ετικέτα υπάρχει ένα ονοματεπώνυμο που αντιστοιχεί σε ένα επιστήμονα. Ο αριθμός των πλήρη ονομάτων είναι ίσος με τον αριθμό των ονομάτων που υπάρχουν στη γραμμή τύπου ένα.

Να σημειωθεί ότι στις γραμμές τύπου ένα και δύο πάντα γίνεται έλεγχος αν υπάρχει συνεργασία επιστημόνων. Βασικά γίνεται έλεγχος αν υπάρχουν τουλάχιστον δυο ονοματεπώνυμα προκειμένου να υπάρχει συνεργασία. Αν όχι, τότε αγνοούνται οι πληροφορίες που ακολουθούν, δεν αποθηκεύεται τίποτα στη βάση δεδομένων μέχρι να συναντήσουμε την ετικέτα ER.

Στη γραμμή τύπου τρία υπάρχει ο τίτλος του άρθρου.

Στη γραμμή τύπου τέσσερα βρίσκονται τα ονόματα των ιδρυμάτων-οργανισμών που επιστήμονες τους συνεργάστηκαν για τη συγγραφή αυτού του άρθρου. Τα ονόματα αυτά αν είναι περισσότερα από ένα τότε δεν βρίσκονται σε μια γραμμή, αλλά σε κάθε γραμμή που ακολουθεί μέχρι την επόμενη ετικέτα.

Στη γραμμή τύπου πέντε υπάρχει το όνομα του ιδρύματος-οργανισμού που ανήκει ο επιστήμονας που δημοσίευσε το άρθρο. υπάρχει το τμήμα του επιστήμονα που δημοσίευσε το άρθρο. Η πληροφορία αυτής της ετικέτας χρησιμοποιείται όταν δεν εμφανίζεται η ετικέτα C1 και υπάρχει συνεργασία επιστημόνων.

Στην γραμμή τύπου έξι υπάρχει η χρονιά δημοσίευσης του άρθρου.

Στην γραμμή τύπου επτά αν υπάρχει συνεργασία επιστημόνων για τη συγγραφή αυτού του άρθρου τότε αποθηκεύονται τα δεδομένα που έχουν αποθηκευτεί προσωρινά σε τοπικές μεταβλητές στα προηγούμενα βήματα. Συγκεκριμένα, στον πίνακα *article* αποθηκεύονται ο αριθμός ταυτοποίησης του άρθρου, ο τίτλος και η χρονιά δημοσίευσης του. Στον πίνακα *scientist* αποθηκεύονται ο αριθμός ταυτοποίησης του επιστήμονα, το ονοματεπώνυμό του και η τιμή εκείνη που δηλώνει αν ο συγκεκριμένος επιστήμονας ανήκει στο Πανεπιστήμιο Κύπρου. Επιπλέον, στον πίνακα *names_of_scientists* αποθηκεύονται ο αριθμός ταυτοποίησης του επιστήμονα και το ονοματεπώνυμό του. Επιπρόσθετα, στον πίνακα *institution* αποθηκεύονται ο αριθμός που ταυτοποιεί ένα ίδρυμα-οργανισμό, καθώς επίσης και το όνομά αυτού του ιδρύματος-οργανισμού. Στον πίνακα *coll_institutions* αποθηκεύονται ο αριθμός που ταυτοποιεί ένα άρθρο και το όνομα του ιδρύματος-οργανισμού που συμμετείχε στη συγγραφή αυτού του άρθρου. Τέλος, στον πίνακα *is_writtten* αποθηκεύονται ο αριθμός ταυτοποίησης του άρθρου, ο τίτλος του άρθρου, ο αριθμός ταυτοποίησης του επιστήμονα που συνεργάστηκε με άλλους επιστήμονες για τη συγγραφή αυτού του άρθρου και η τιμή εκείνη που δηλώνει αν ο συγκεκριμένος επιστήμονας ανήκει στο Πανεπιστήμιο Κύπρου.

Ακολουθεί ο ψευδοκώδικας υπάρχει η κεντρική ιδέα του script.

```
while(!feof($file))
{
    $inst = fgets($file);

    if($r == 1) //AU
    {
        array_push($auTable,$inst);
    }
    else if($r == 2) //AF
    {
        if(sizeof($auTable) != 1){
            array_push($afTable,$inst);
        }
    }
    else if($r == 3) //TI
    {
        if(sizeof($auTable) != 1){
            $title = substr($inst, 4);
        }
    }
    else if($r == 4) //C1
    {
        if(sizeof($auTable) != 1){
            array_push($c1Table,$inst);
        }
    }
    else if($r == 5) //RP
    {
        $array = explode(",",$inst);
        $rp = trim($array[2]);
    }
    else if($r == 6) //PY
    {
        if(sizeof($auTable) != 1){
            $py = substr($inst, 4);
        }
    }
    else if($r == 7) //ER
    {
        mysql_query("INSERT INTO article (Number, Title, PubYear) VALUES ('$anum','$title', '$py')");
        mysql_query("INSERT INTO scientist (Name, AffCyprus) VALUES ('$stringNew', '0')");
        mysql_query("INSERT INTO names_of_scientists (SID, SName) VALUES ('$sidnew',
'$stringNew')");
        mysql_query("INSERT INTO institution (Name) VALUES ('$c1Table[$u]')");
        mysql_query("INSERT INTO coll_institutions (ANum, INum) VALUES ('$anum', '$idTable[$r]')");
        mysql_query("INSERT INTO is_written (ANum, ATitle, SID, InstCyprus) VALUES ('$anum', '$title',
'$r1[SID]', '$r1[AffCyprus]')");
    }
}
```

Διεξοδική περιγραφή του script:

Διαβάζουμε το αρχείο κειμένου γραμμή προς γραμμή.

Εάν είμαστε στην πρώτη γραμμή του αρχείου υ αρχικά παίρνουμε από τη βάση δεδομένων τον αριθμό που ταυτοποιεί το τελευταίο άρθρο που αποθηκεύτηκε στη βάση δεδομένων. Επιπλέον, καλούνται η συνάρτηση **setSidLet()**, στην οποία αρχικοποιείται ο πίνακας *letters* και η συνάρτηση **setInstitutions()**, στην οποία αρχικοποιείται ο πίνακας *institutions*. Οι δυο αυτοί πίνακες έχουν 26 θέσεις στις οποίες υπάρχει ένας πίνακας, ο οποίος αντιστοιχεί σε ένα γράμμα τις αγγλικής αλφαβήτου (A-Z). Στον κάθε πίνακα *letters* αποθηκεύονται οι αριθμοί εκείνοι που ταυτοποιούν μοναδικά τον επιστήμονα του οποίου το όνομα ξεκινά από το συγκεκριμένο γράμμα. Για παράδειγμα, στον πίνακα *letters* στη θέση μηδέν υπάρχει ο πίνακας με τα ids των επιστημόνων που το επίθετό τους ξεκινά από 'A'. Στον κάθε πίνακα *institutions* αποθηκεύονται οι αριθμοί εκείνοι που ταυτοποιούν μοναδικά το ίδρυμα-οργανισμό του οποίου το όνομα ξεκινά από το συγκεκριμένο γράμμα. Για παράδειγμα, στον πίνακα *institutions* στη θέση 21 υπάρχει ο πίνακας με τα ids των επιστημόνων που το όνομά τους ξεκινά από 'U' (UNIV CYPRUS).

Ακολούθως, γίνονται κάποιοι ελέγχοι για να καταλάβουμε σε τι τύπου γραμμή βρισκόμαστε γιατί όπως αναφέρθηκε προηγουμένως δεν εμφανίζονται πάντα σταθερός αριθμός ετικετών σε όλα τα αρχεία κειμένου. Συνεπώς, πρέπει να αναγνωρίζουμε σε τι τύπου γραμμή βρισκόμαστε κάθε φορά. Στις γραμμές εκείνες που οι πληροφορίες που παρέχουν πρέπει να επεξεργαστούν, τους δίνεται ένας αριθμός από το ένα έως το επτά ενώ στις υπόλοιπες δίνεται ο αριθμός μηδέν.

Όταν βρισκόμαστε σε γραμμή τύπου μηδέν απλά αγνοείτε και διαβάζεται η επόμενη γραμμή.

Κάθε φορά που βρισκόμαστε στη γραμμή τύπου ένα πρέπει να γίνεται έλεγχος αν υπάρχει ετικέτα για να συμπεράνουμε αν αλλάξαμε τύπο γραμμής. Όμοιος έλεγχος γίνεται και όταν βρισκόμαστε στη γραμμή τύπου δύο.

Όταν βρισκόμαστε στη γραμμή τύπου ένα τότε έχουμε τα ονοματεπώνυμα των επιστημόνων που συνεργάστηκαν για τη συγγραφή του συγκεκριμένου άρθρου. Τα

ονοματεπώνυμα δεν βρίσκονται όλα σε μια γραμμή, αλλά ένα σε κάθε γραμμή. Καταλαμβάνουμε πως μια επόμενη γραμμή δεν έχει κάποιο ονοματεπώνυμο ενός επιστήμονα όταν έχει ετικέτα. Αν τώρα, είμαστε στο πρώτο ονοματεπώνυμο θα πρέπει να αφαιρεθεί η ετικέτα AU για να ληφθεί η σωστή συμβολοσειρά. Το ονοματεπώνυμο του κάθε επιστήμονα προστίθεται σε ένα πίνακα, ο οποίος κρατάει όλα τα ονοματεπώνυμα των επιστημόνων που συνεργάστηκαν για τη συγγραφή του συγκεκριμένου άρθρου.

Όταν βρισκόμαστε στη γραμμή τύπου δύο τότε έχουμε τα πλήρη ονοματεπώνυμα των επιστημόνων που συνεργάστηκαν για τη συγγραφή του συγκεκριμένου άρθρου. Αρχικά, ελέγχεται το μέγεθος του πίνακα με ονοματεπώνυμα των επιστημόνων αν είναι μεγαλύτερος από ένα προκειμένου να υπάρχει συνεργασία επιστημόνων. Αν υπάρχει συνεργασία επιστημόνων τότε μπαίνουμε στη διαδικασία να αποθηκεύσουμε τα πλήρη ονοματεπώνυμα. Τα ονοματεπώνυμα όμως δεν βρίσκονται όλα σε μια γραμμή, αλλά ένα σε κάθε γραμμή. Καταλαμβάνουμε πως μια επόμενη γραμμή δεν έχει ένα πλήρες ονοματεπώνυμο ενός επιστήμονα όταν υπάρχει ετικέτα. Αν τώρα, είμαστε στο πρώτο πλήρες ονοματεπώνυμο θα πρέπει να αφαιρεθεί η ετικέτα AF για να πάρουμε τη σωστή συμβολοσειρά. Ακολούθως, το πλήρες ονοματεπώνυμο του κάθε επιστήμονα προστίθεται σε ένα πίνακα, ο οποίος κρατάει όλα τα πλήρη ονοματεπώνυμα των επιστημόνων που συνεργάστηκαν για τη συγγραφή του συγκεκριμένου άρθρου.

Όταν βρισκόμαστε στη γραμμή τύπου τρία τότε έχουμε τον τίτλο του άρθρου. Αρχικά, ελέγχεται αν υπάρχει συνεργασία επιστημόνων. Αν υπάρχει συνεργασία επιστημόνων τότε θα πρέπει να αφαιρεθεί η ετικέτα TI για να ληφθεί η σωστή συμβολοσειρά. Ο τίτλος του άρθρου αποθηκεύεται σε μια καθολική μεταβλητή.

Αν τώρα βρισκόμαστε στη γραμμή τύπου τέσσερα τότε έχουμε τα ονόματα των ιδρυμάτων-οργανισμών που επιστήμονές τους συνεργάστηκαν για τη συγγραφή αυτού του άρθρου. Αρχικά, ελέγχεται αν υπάρχει συνεργασία επιστημόνων. Αν υπάρχει συνεργασία επιστημόνων τότε μπαίνουμε στη διαδικασία να αποθηκεύσουμε τα ονόματα των ιδρυμάτων-οργανισμών. Τα ονόματα όμως δεν βρίσκονται όλα σε μια γραμμή, αλλά ένα σε κάθε γραμμή. Καταλαμβάνουμε πως μια επόμενη γραμμή δεν έχει ένα όνομα ενός ιδρύματος-οργανισμού όταν υπάρχει ετικέτα. Αν τώρα, είμαστε

στο πρώτο όνομα θα πρέπει να αφαιρεθεί η ετικέτα C1 για να πάρουμε τη σωστή συμβολοσειρά. Είτε έχουμε το πρώτο όνομα είτε κάποιο άλλο όνομα ιδρύματος-οργανισμού, θα πρέπει να επεξεργαστεί η συμβολοσειρά και να απομονώσουμε το όνομα του ιδρύματος-οργανισμού προκειμένου να πάρουμε μόνο την πληροφορία που μας ενδιαφέρει. Η νέα συμβολοσειρά προστίθεται σε ένα πίνακα, ο οποίος κρατάει όλα τα ονόματα των ιδρυμάτων-οργανισμών που επιστήμονές τους συνεργάστηκαν για τη συγγραφή αυτού του άρθρου. Επιπρόσθετα, αν το όνομα που παίρνουμε είναι “UNIV CYPRUS”, τότε προτού διαβαστεί η επόμενη γραμμή, θα πρέπει να επεξεργαστεί η αρχική συμβολοσειρά για να απομονώσουμε το όνομα του τμήματος που αναφέρεται αυτό το άρθρο. Το όνομα του τμήματος αποθηκεύεται σε μια καθολική μεταβλητή. Αν το άρθρο αναφέρεται σε ένα από τα πέντε τμήματα που θα μελετηθούν στη συνέχεια, τότε η καθολική σημαία του αντίστοιχου τμήματος αρχικοποιείται με ένα. Αυτό συμβαίνει προκειμένου αν το άρθρο δεν υπάρχει ήδη στη βάση δεδομένων, κατά τη φάση αποθήκευσης των δεδομένων στη βάση δεδομένων να γίνει έλεγχος αν επιλέχθηκαν οι σωστοί επιστήμονες. Συγκεκριμένα, αν ο επιστήμονας με το όνομα που πήραμε προηγουμένως ήδη υπάρχει στη βάση δεδομένων, μπορεί να μην είναι αυτός αλλά κάποιος άλλος με το ίδιο ονοματεπώνυμο. Ο έλεγχος γίνεται με βάση το τμήμα που ανήκει ο επιστήμονας. Στο τέλος επιλέγεται ο επιστήμονας που ανήκει στο τμήμα εκείνο που αναφέρεται το άρθρο.

Αν τώρα βρισκόμαστε στη γραμμή τύπου πέντε τότε έχουμε το όνομα του ιδρύματος-οργανισμού που ανήκει ο επιστήμονας που δημοσίευσε το άρθρο. Η συμβολοσειρά επεξεργάζεται προκειμένου να απομονώσουμε το όνομα του ιδρύματος-οργανισμού. Η πληροφορία αυτή χρησιμοποιείται μόνο όταν δεν εμφανίζεται η ετικέτα C1.

Όταν βρισκόμαστε στη γραμμή τύπου έξι τότε έχουμε τη χρονιά δημοσίευσης του άρθρου. Αν υπάρχει συνεργασία επιστημόνων τότε θα πρέπει να αφαιρεθεί η ετικέτα PY για να πάρουμε τη σωστή συμβολοσειρά. Η χρονιά δημοσίευσης του άρθρου αποθηκεύεται σε μια καθολική μεταβλητή.

Όταν βρισκόμαστε στη γραμμή τύπου επτά τότε καταλαβαίνουμε πως επεξεργάστηκαν όλες τις πληροφορίες για αυτό το άρθρο και προτού προχωρήσουμε στις πληροφορίες του επόμενου άρθρου θα πρέπει να αποθηκεύσουμε τα δεδομένα που

συλλέχθηκαν για αυτό το άρθρο στους σωστούς πίνακες της βάσης δεδομένων. Αν υπάρχει συνεργασία επιστημόνων τότε καλείται η συνάρτηση **insertData(\$auTable, \$afTable, \$title, \$py)**, η οποία παίρνει ως παραμέτρους τον πίνακα με τα ονοματεπώνυμα των επιστημόνων, τον πίνακα με τα πλήρη ονοματεπώνυμα των επιστημόνων, τον τίτλο του άρθρου και τη χρονιά δημοσίευσης του άρθρου και ενημερώνει τους πίνακες *article*, *scientist*, *names_of_scientists*, *institution*, *coll_institutions* και *is_written*.

Αρχικά, στη συνάρτηση αυτή, επιλέγονται από τον πίνακα *article* κατά αλφαβητική σειρά τα άρθρα που δημοσιεύτηκαν τη χρονιά που δημοσιεύτηκε το άρθρο που θέλουμε να προσθέσουμε τα δεδομένα του στη βάση δεδομένων. Στη συνέχεια, ελέγχεται αν ο τίτλος του άρθρου μας είναι ίδιος με ένα από τα άρθρα που ήδη υπάρχουν στον πίνακα *article*. Αν ήδη υπάρχει αυτό το άρθρο στη βάση δεδομένων τότε δεν χρειάζεται να αποθηκευτεί τίποτα στη βάση δεδομένων και προχωρούμε στις πληροφορίες του επόμενου άρθρου. Στην περίπτωση που το άρθρο δεν υπάρχει στον πίνακα *article*, τότε στον πίνακα *article* προστίθονται ο αριθμός που ταυτοποιεί το άρθρο, ο τίτλος του άρθρου και η χρονιά δημοσίευσής του. Στη συνέχεια, καλείται η συνάρτηση **updateScientist(\$auTable, \$afTable)**.

Η συνάρτηση **updateScientist(\$auTable, \$afTable)** παίρνει ως παραμέτρους τους πίνακες με τα ονοματεπώνυμα (πλήρη και μη) των επιστημόνων που συνεργάστηκαν για τη συγγραφή του συγκεκριμένου άρθρου και επιστρέφει έναν πίνακα με όλες τις 'ταυτότητες' αυτών των επιστημόνων. Στη συνάρτηση αυτή, αρχικά επιλέγονται ένα-ένα τα ονοματεπώνυμα των επιστημόνων από τους δυο πίνακες και αποθηκεύονται σε δυο μεταβλητές. Επιλέγουμε το γράμμα από το οποίο ξεκινά το επίθετο του επιστήμονα και επιλέγεται το αντίστοιχος πίνακας που περιέχει τις 'ταυτότητες' των επιστημόνων που το επίθετό τους ξεκινά από αυτό το γράμμα. Στη συνέχεια, γίνεται έλεγχος αν τα στοιχεία αυτού του επιστήμονα υπάρχουν ήδη στη βάση δεδομένων μας. Συγκεκριμένα, παίρνουμε μία-μία τις 'ταυτότητες' από τον πίνακα που επιλέχθηκαν και από τον πίνακα *names_of_scientists* επιλέγονται τα δεδομένα για την κάθε ταυτότητα ξεχωριστά. Ακολούθως, γίνεται έλεγχος αν το όνομα που έχουμε είναι ίδιο με κάποιο από αυτά που επιλέχθηκαν. Αν ναι, τότε ο επιστήμονας ήδη υπάρχει στον πίνακα *names_of_scientists*. Αν η σημαία ενός από τα πέντε τμήματα που θα μελετηθούν είναι

ιση με ένα τότε γίνεται έλεγχος αν επιλέχθηκε ο σωστός επιστήμονας γιατί κάποιοι επιστήμονες τυγχάνει να έχουν το ίδιο ονοματεπώνυμο, δηλαδή αυτός που ανήκει στο τμήμα που αναφέρεται το συγκεκριμένο άρθρο. Έπειτα, προστίθεται η ‘ταυτότητα’ του συγκεκριμένου επιστήμονα στον πίνακα με τις ‘ταυτότητες’ των επιστημόνων που συνεργάστηκαν για τη συγγραφή του συγκεκριμένου άρθρου και σταματάμε τον έλεγχο. Αν μετά από τον έλεγχο, διαπιστώσουμε πως αυτός ο επιστήμονας δεν υπάρχει μέσα σε αυτόν τον πίνακα τότε πρέπει να προστεθούν τα στοιχεία του στη βάση δεδομένων. Αν ο πίνακας που έχει τα πλήρη ονοματεπώνυμα των επιστημόνων δεν είναι άδειος, τότε στον πίνακα *scientist* το όνομα του επιστήμονα θα πρέπει να έχει την εξής μορφή:

Surname 1, Name1; Surname2, Name2

Συνεπώς, δημιουργείτε μια καινούργια συμβολοσειρά που περιέχει τα δύο ονοματεπώνυμα με βάση την πιο πάνω μορφή και στον πίνακα *scientist* προστίθονται το όνομα του επιστήμονα (νέα συμβολοσειρά) και η τιμή που δηλώνει ότι αυτός ο επιστήμονας δεν ανήκει στο Πανεπιστήμιο Κύπρου. Αν τώρα, ο πίνακας με πλήρη ονοματεπώνυμα των επιστημόνων είναι άδειος, τότε απευθείας στον πίνακα *scientist* προστίθονται το όνομα του επιστήμονα και η τιμή που δηλώνει ότι αυτός ο επιστήμονας δεν ανήκει στο Πανεπιστήμιο Κύπρου. Έπειτα, προστίθεται ακόμη μία ‘ταυτότητα’ στον πίνακα με τις ‘ταυτότητες’ των επιστημόνων που συνεργάστηκαν για τη συγγραφή αυτού του άρθρου. Στη συνέχεια, ενημερώνεται ο πίνακας *names_of_scientists*. Αν ο πίνακας που έχει τα πλήρη ονοματεπώνυμα των επιστημόνων είναι άδειος, τότε απευθείας προστίθονται η ‘ταυτότητα’ και το ονοματεπώνυμο του επιστήμονα. Σε αντίθετη περίπτωση προστίθονται δύο νέες εγγραφές στον πίνακα *names_of_scientists*. Στην πρώτη βάζουμε την ‘ταυτότητα’ του επιστήμονα και το ονοματεπώνυμο του επιστήμονα και στη δεύτερη βάζουμε την ‘ταυτότητα’ του επιστήμονα και το πλήρες ονοματεπώνυμο του επιστήμονα. Στο τέλος, ενημερώνεται ο αντίστοιχος πίνακας που περιέχει τις ‘ταυτότητες’ των επιστημόνων που το επίθετό τους ξεκινά από το γράμμα που ξεκινά το επίθετο του νέου επιστήμονα και επιστρέφουμε στη συνάρτηση `insertData($auTable, $afTable, $stitle, $spy)` τον πίνακα με τις ‘ταυτότητες’ όλων εκείνων των επιστημόνων που συνεργάστηκαν για τη συγγραφή του συγκεκριμένου άρθρου.

Ακολούθως, στη συνάρτηση *insertData(\$auTable, \$afTable, \$title, \$py)* αποθηκεύονται τα κατάλληλα δεδομένα στον πίνακα *is_written*. Για κάθε ‘ταυτότητα’ που έχουμε από τον πίνακα που πήραμε από τη συνάρτηση *updateScientist(\$auTable, \$afTable)* επιλέγονται οι πληροφορίες του επιστήμονα με την αντίστοιχη ‘ταυτότητα’ και στον πίνακα *is_written* τον αριθμό που ταυτοποιεί το άρθρο, το ν τίτλο του άρθρο *y* την ‘ταυτότητα’ του επιστήμονα και την τιμή που δηλώνει αν ο επιστήμονας ανήκει στο Πανεπιστήμιο Κύπρου.

Στη συνάρτηση *insertData(\$auTable, \$afTable, \$title, \$py)* απέμεινε να αποθηκευτούν τα κατάλληλα δεδομένα στους πίνακες *institution* και *coll_institutions*. Προτού όμως αποθηκευτούν τα δεδομένα γίνεται έλεγχος αν ο πίνακας με τα ονόματα των ιδρυμάτων-οργανισμών είναι άδειος. Αν ναι, τότε σε αυτόν τον πίνακα προστίθεται η πληροφορία που πήραμε από την ετικέτα RP. Στη συνέχεια, καλείται η συνάρτηση **updateInstitution()**.

Στη συνάρτηση *updateInstitution()*, αρχικά επιλέγονται ένα-ένα τα ονόματα των ιδρυμάτων-οργανισμών από τον πίνακα *\$c1Table*. Επιλέγουμε το γράμμα από το οποίο ξεκινά το όνομα του ιδρύματος-οργανισμού και επιλέγεται το αντίστοιχος πίνακας που περιέχει τις ‘ταυτότητες’ των ιδρυμάτων-οργανισμών που το όνομά τους ξεκινά από αυτό το γράμμα. Στη συνέχεια, γίνεται έλεγχος αν αυτό το ίδρυμα-οργανισμός υπάρχει ήδη στη βάση δεδομένων. Συγκεκριμένα, παίρνουμε μία-μία τις ‘ταυτότητες’ από τον πίνακα που επιλέχθηκαν και από τον πίνακα *institution* επιλέγονται τα δεδομένα για την κάθε ταυτότητα ξεχωριστά. Ακολούθως, γίνεται έλεγχος αν το όνομα που έχουμε είναι ίδιο με κάποιο από αυτά που επιλέχθηκαν. Αν ναι, τότε το ίδρυμα-οργανισμός ήδη υπάρχει στον πίνακα *institution* και αυτόματα η ‘ταυτότητα’ του ιδρύματος-οργανισμού αποθηκεύεται στον πίνακα *\$idTable* που κρατάει τις ‘ταυτότητες’ των ιδρυμάτων-οργανισμών που επιστήμονές τους συνεργάστηκαν για τη συγγραφή του συγκεκριμένου άρθρου. Αν όχι, τότε προστίθεται στον πίνακα *institution* το όνομα του «νέου» ιδρύματος-οργανισμού και η ‘ταυτότητα’ του αποθηκεύεται στον πίνακα *\$idTable* και στον πίνακα με τα ιδρύματα που το όνομά τους ξεκινά με το ίδιο γράμμα αυτού του ιδρύματος.

Τέλος, αυξάνεται ο αριθμός που ταυτοποιεί τα άρθρα.

Στο τέλος της φάσης για τη γραμμή τύπου επτά, αδειάζονται οι πίνακες που κρατούν τα ονοματεπώνυμα των επιστημόνων που συνεργάστηκαν για τη συγγραφή αυτού του άρθρου και που κρατούν τα ονόματα των ιδρυμάτων-οργανισμών, και μηδενίζονται οι καθολικές μεταβλητές.

3.7 Επεξεργασία Δεδομένων Ιδρυμάτων-Οργανισμών

Αφού αποθηκεύτηκαν στη βάση δεδομένων τα ονόματα των ιδρυμάτων-οργανισμών τόσο από τα αρχεία δεδομένων της βάσης δεδομένων Scopus όσο και από τα αρχεία δεδομένων της βάσης δεδομένων *Citation Index (ISI-Web of Knowledge)*, παρατηρήθηκε πως υπήρχαν ονόματα ιδρυμάτων-οργανισμών με διαφορετικές ταυτότητες που αντιστοιχούσαν στο ίδιο ίδρυμα-οργανισμό. Το πρόβλημα αυτό ήταν προβλέψιμο λόγω του ανθρώπινου παράγοντα. Ο κάθε επιστήμονας μπορεί να έγραφε το όνομα του ιδρύματος στο οποίο ανήκει ολογράφως ή με συντομογραφία ή ακόμη μπορούσε να το γράψει και λανθασμένα. Έτσι ήταν αναγκαίο να επεξεργαστούν τα δεδομένα των ιδρυμάτων-οργανισμών προκειμένου να έχουμε μια συγκεκριμένη ταυτότητα που να αντιστοιχεί σε κάθε ίδρυμα-οργανισμό. Σε αυτή τη φάση τροποποιήθηκε ο πίνακας *coll_institutions* και προστέθηκε ένας νέος πίνακας *new_institutions* στον οποίο έχουμε τις ταυτότητες και τα ονόματα μόνο των ιδρυμάτων-οργανισμών που απομείναν μετά την επεξεργασία των δεδομένων του πίνακα *coll_institutions*.

Στη συνέχεια, στον πίνακα *new_institutions* έπρεπε να προστεθούν τα ranking numbers (αριθμοί κατάταξης) των πανεπιστημίων. Με βάση τη γνωστή **λίστα της Shanghai** που δημοσιεύεται κάθε χρόνο από το *Shanghai Jiaotong University* (www.arwu.org) πήραμε τη λίστα των 500 καλύτερων πανεπιστημίων παγκοσμίως για το 2009. Με βάση λοιπόν αυτή τη λίστα, στα πρώτα 10 καλύτερα πανεπιστήμια βάλουμε ως ranking number τον αριθμό 1, από το 101-200 τον αριθμό 2, από το 201-300 τον αριθμό 3, από τον αριθμό 301-400 τον αριθμό 4 και από τον αριθμό 401-500 τον αριθμό 5. Στις ταυτότητες των ιδρυμάτων εκείνων όπου το όνομα που ο επιστήμονας έδωσε ήταν λανθασμένο-μη υπαρκτό δόθηκε ως ranking number ο αριθμός -1. Πότε συνέβαινε αυτό; Υπήρχαν περιπτώσεις που ο επιστήμονας έδινε τη διεύθυνση του

ιδρύματος ή μόνο το τμήμα στο οποίο ανήκει. Στα υπόλοιπα ιδρύματα που δεν είχε δοθεί ranking number, προστέθηκε ο αριθμός 6.

3.8 Δεδομένα εξόδου

Για την επόμενη φάση, αυτή της δημιουργίας του δικτύου συνεργασίας του Πανεπιστημίου Κύπρου· θα πρέπει να επιλεχθούν τα απαραίτητα δεδομένα από τη βάση δεδομένων. Όπως έχει προαναφερθεί, θα μελετηθεί η εξέλιξη των επιστημονικών συνεργασιών του Πανεπιστημίου Κύπρου για κάθε τμήμα που έχει επιλεγεί, για τμήμα και χρονιά ξεχωριστά, όπως επίσης και η εξέλιξη των συνεργασιών του Πανεπιστημίου Κύπρου για κάθε τμήμα με άλλα ιδρύματα-οργανισμούς.

-Για τη δημιουργία των αρχείων κειμένου που περιέχουν τις πληροφορίες που συνδέουν άρθρα-επιστήμονες και άρθρα-ιδρύματα ακολουθήθηκε η εξής διαδικασία:

Αρχικά, αποθηκεύονται σε πέντε διαφορετικούς πίνακες οι ταυτότητες των επιστημόνων του Πανεπιστημίου Κύπρου που ανήκο **v** στα πέντε τμήματα που επιλέχθηκαν να μελετηθούν (Πληροφορικής, Φυσικής, Μαθηματικών, Χημείας, Βιολογίας). Η επιλογή των ταυτοτήτων γίνεται με βάση τον πίνακα *belongs* της βάσης δεδομένων.

Στη συνέχεια, αποθηκεύονται σε ένα πίνακα οι ταυτότητες των άρθρων εκείνων που ο αριθμός των συγγραφέων τους δεν ξεπερνάει τους 100. Αυτό γίνεται γιατί υπάρχουν άρθρα στα οποία συνεργάζονται περισσότεροι από τους 250 επιστήμονες και αυτά τα άρθρα ως επί των πλείστων αναφέρονται στο τμήμα Φυσικής. Έτσι, για να έχουμε σωστά αποτελέσματα στη συνέχεια επιλέγονται μόνο τα άρθρα εκείνα που συμμετέχουν λιγότεροι από 100 επιστήμονες. Η επιλογή γίνεται με βάση πόσες εγγραφές υπάρχουν για το κάθε άρθρο στον πίνακα *is_written*.

Ακολούθως, επιλέγονται όλοι οι αριθμοί που ταυτοποιούν τα άρθρα εκείνα στα οποία συμμετέχουν επιστήμονες που ανήκουν σε ένα από τα πέντε τμήματα. Για παράδειγμα, για κάθε επιστήμονα που ανήκει στο τμήμα Πληροφορικής του Πανεπιστημίου Κύπρου επιλέγονται οι εγγραφές του πίνακα *is_written* στις οποίες παρουσιάζεται η ταυτότητα

του συγκεκριμένου επιστήμονα. Για κάθε μία από τις επιλεγμένες εγγραφές ελέγχεται κατά πόσο ο αριθμός του άρθρου υπάρχει στον πίνακα με τις ταυτότητες των άρθρων με λιγότερους από 100 επιστήμονες. Αν ναι, τότε η ταυτότητα του συγκεκριμένου άρθρου προστίθεται στον πίνακα με τους αριθμούς των άρθρων που αναφέρονται στο τμήμα Πληροφορικής. Στο τέλος ο πίνακας με τις ταυτότητες των άρθρων που αναφέρονται στο τμήμα Πληροφορικής επεξεργάζεται προκειμένου να υπάρχει μόνο μια φορά η ταυτότητα κάθε άρθρου. Η ίδια διαδικασία ακολουθείται και για τα υπόλοιπα τμήματα.

Έπειτα, επιλέγονται όλα τα άρθρα όλων των χρονιών (1992-2009), τα οποία έχουν ένα τουλάχιστον επιστήμονα, ο οποίος ανήκει στο Πανεπιστήμιο Κύπρου σε συγκεκριμένο τμήμα. Συγκεκριμένα, για το τμήμα της Πληροφορικής, αρχικά ανοίγεται ένα αρχείο κειμένου στο οποίο θα γραφτούν τα δεδομένα για τα άρθρα που αναφέρονται στο τμήμα της Πληροφορικής από το 1992 έως 2009. Για κάθε αριθμό που ταυτοποιεί ένα άρθρο το οποίο αναφέρεται στο τμήμα Πληροφορικής επιλέγεται μία εγγραφή από τον πίνακα *article* η οποία αντιστοιχεί στο ίδιο άρθρο. Αυτό γίνεται για να έχουμε και την πληροφορία δημοσίευσης του άρθρου που δεν υπάρχει στον πίνακα *is_written*. Έπειτα, δημιουργείται μια νέα συμβολοσειρά η οποία έχει όλες τις απαιτούμενες πληροφορίες και η οποία έχει την εξής μορφή:

ANum|SID|InstCyprus|PubYear

ANum: Αριθμός άρθρου

SID: Ταυτότητα επιστήμονα

InstCyprus: Τιμή που υποδηλώνει αν ο επιστήμονας ανήκει στο Πανεπιστήμιο Κύπρου

PubYear: Χρονιά δημοσίευσης άρθρου

Αυτή η συμβολοσειρά γράφεται στο αρχείο κειμένου και αφού γραφτούν όλα τα δεδομένα κλείνει το αρχείο κειμένου. Η ίδια διαδικασία ακολουθείται και για τα υπόλοιπα τμήματα.

Στη συνέχεια, ακολουθεί η διαδικασία για τη δημιουργία αρχείων κειμένου που περιέχουν τις πληροφορίες για τα άρθρα που αναφέρονται σε ένα συγκεκριμένο τμήμα μιας συγκεκριμένης χρονιάς.(1992-2009). Αρχικά, επιλέγονται όλες οι εγγραφές του πίνακα *article*. Για κάθε εγγραφή, ανάλογα με τη χρονιά δημοσίευσης του άρθρου προστίθεται ο αριθμός του άρθρου σε πίνακα που αναφέρεται στο αντίστοιχο έτος.

Έτσι, στο τέλος θα δημιουργηθούν 18 πίνακες, οι οποίοι θα κρατούν τους αριθμούς των άρθρων που δημοσιεύτηκαν την αντίστοιχη χρονιά. Ακολούθως, για κάθε χρονιά από το 1992 έως το 2009, για κάθε ένα από τα πέντε τμήματα δημιουργείται το αντίστοιχο αρχείο κειμένου(π.χ. cs1992.txt, phy2003.txt) και καλείται η συνάρτηση **yearDepartFunction(\$anumYear, \$departAnumArray)**.

Η συνάρτηση **yearDepartFunction(\$anumYear, \$departAnumArray)** παίρνει δύο παραμέτρους, ένα πίνακα ο οποίος περιλαμβάνει τους αριθμούς των άρθρων συγκεκριμένης χρονιάς και ένα πίνακα ο οποίος περιλαμβάνει τους αριθμούς των άρθρων που αναφέρονται σε συγκεκριμένο τμήμα. Στο σώμα αυτής της συνάρτησης δημιουργείται ένας νέος πίνακας στον οποίο αποθηκεύονται οι αριθμοί των άρθρων που εμφανίζονται και στους δύο πίνακες. Έτσι, αυτός ο πίνακας κρατάει μόνο τους αριθμούς των άρθρων που δημοσιεύτηκαν συγκεκριμένη χρονιά και αναφέρονται σε συγκεκριμένο τμήμα. Στο τέλος, αυτή η συνάρτηση επιστρέφει το νέο πίνακα.

Το αντίστοιχο αρχείο κειμένου ανοίγει. Για κάθε αριθμό άρθρου του νέου πίνακα που επιστρέφει η συνάρτηση *yearDepartFunction*, επιλέγονται οι εγγραφές του πίνακα *is_written* στις οποίες παρουσιάζεται ο ίδιος αριθμός άρθρου. Για κάθε επιλεγμένη εγγραφή δημιουργείται μια νέα συμβολοσειρά η οποία έχει όλες τις απαιτούμενες πληροφορίες και η οποία έχει την εξής μορφή:

A Num|SID|InstCyprus|PubYear

A Num: Αριθμός άρθρου

SID: Ταυτότητα επιστήμονα

InstCyprus: Τιμή που υποδηλώνει αν ο επιστήμονας ανήκει στο Πανεπιστήμιο Κύπρου

PubYear: Χρονιά δημοσίευσης άρθρου

Αυτή η συμβολοσειρά γράφεται στο αρχείο κειμένου και αφού γραφτούν όλα τα δεδομένα κλείνει το αρχείο κειμένου.

Στη συνέχεια, ακολουθεί η διαδικασία για τη δημιουργία αρχείων κειμένου που περιέχουν τις πληροφορίες που συνδέουν άρθρα με ιδρύματα-οργανισμούς. Για κάθε άρθρο που αναφέρεται σε ένα από τα πέντε τμήματα, επιλέγονται οι εγγραφές από τον πίνακα *coll_institutions* στις οποίες παρουσιάζεται ο ίδιος αριθμός άρθρου. Έπειτα, για κάθε επιλεγμένη εγγραφή, επιλέγεται μία εγγραφή του πίνακα *new_institutions* στην

οποία εμφανίζεται ο ίδιος αριθμός που αντιστοιχεί στο ίδρυμα-οργανισμό και μία εγγραφή του πίνακα *is_written* στην οποία εμφανίζεται ο ίδιος αριθμός άρθρου. Ακολούθως, δημιουργείται μια νέα συμβολοσειρά η οποία έχει όλες τις απαιτούμενες πληροφορίες και η οποία έχει την εξής μορφή:

A Num|ID|RankNum|PubYear

A Num: Αριθμός άρθρου

ID: Ταυτότητα ιδρύματος-οργανισμού

RankNum: Αριθμός κατάταξης ιδρύματος

PubYear: Χρονιά δημοσίευσης άρθρου

Αυτή η συμβολοσειρά γράφεται στο αρχείο κειμένου και αφού γραφτούν όλα τα δεδομένα κλείνει το αρχείο κειμένου.

Κεφάλαιο 4

Δημιουργία Δικτύου Συνεργασίας

4.1 Επιλογή βιβλιοθήκης SNAP library	51
4.2 Δημιουργία δικτύου συνεργασίας	53
4.2.1 Περιγραφή κλάσεων	54
4.2.2 Περιγραφή δημιουργίας δικτύου συνεργασίας	56
4.2.3 Περιγραφή δημιουργίας γράφου	57

4.1 Επιλογή βιβλιοθήκης *SNAP library*

Όπως έχει αναφερθεί, για τη μελέτη της εξέλιξης των επιστημονικών συνεργασιών του Πανεπιστημίου Κύπρου θα πρέπει να δημιουργήσουμε ένα γράφο, τον οποίο ακολούθως θα αναλύσουμε. Στην προκειμένη περίπτωση, δημιουργήθηκε ένα δίκτυο συνεργασίας. Ένα τέτοιο δίκτυο είναι ένας μη κατευθυνόμενος γράφος με δεδομένα στους κόμβους ή/και στις ακμές του.

Για τη δημιουργία(και την ανάλυση) του δικτύου συνεργασίας επιλέχθηκε η βιβλιοθήκη **SNAP library**, SNAP Ver. 2010-04-20. Η βιβλιοθήκη SNAP υλοποιήθηκε από τον Jure Leskovec, ο οποίος είναι επίκουρος καθηγητής του τμήματος Πληροφορικής του Stanford University.

Η βιβλιοθήκη SNAP (Stanford Network Analysis Platform) είναι μια πλατφόρμα για γενικού σκοπού ανάλυση δικτύων και είναι γραμμένη σε C++. Μπορεί εύκολα να διασχίσει κλιμακωτά και ογκώδη δίκτυα με εκατοντάδες εκατομμύρια κόμβους και δισεκατομμύρια ακμές. Χειρίζεται αποτελεσματικά μεγάλα γραφήματα, υπολογίζει δομικές ιδιότητες, δημιουργεί τυπικά και τυχαία γραφήματα, και υποστηρίζει χαρακτηριστικά πάνω σε κόμβους και ακμές.

Ακολουθεί η βασική σύνταξη της βιβλιοθήκης SNAP που χρησιμοποιήθηκε για τη δημιουργία και ανάλυση του δικτύου συνεργασίας των επιστημονικών συνεργασιών του Πανεπιστημίου Κύπρου.

TUNGraph: undirected graph (single edge between an unordered pair of nodes)

TNodeNet<TNodeData>: like TUNGraph but with TNodeData object for each node

TNodeEDatNet<TNodeData, TEdgeData>: like TUNGraph but with TNodeData on each node and TEdgeData on each edge

Prefix P in the class name stands for a pointer, while T means a type.

Create a graph:

```
PNGraph Graph = TUNGraph::New();
```

```
Graph->AddNode(1);
```

Iterators:

BegNI() : iterator to first node

EndNI() : iterator to one past last node

GetNI(u) : iterator to node with id u

BegEI() : iterator to first edge

EndEI() : iterator to one past last edge

Functionality of node iterators:

GetId() : return node id

GetOutDeg() : return out-degree of a node

GetInDeg() : return in-degree of a node

GetOutNId(e) : return node id of the endpoint of e-th out-edge

GetInNId(e) : return node id of the endpoint of e-th in-edge

Iterators over networks also allow for easy access to the data:

GetDat() : return data type TNodeData associated with the node

GetOutNDat(e) : return data associated with node at endpoint of e-th out-edge

GetInNDat(e) : return data associated with node at endpoint of e-th in-edge

GetOutEDat(e) : return data associated with e-th out-edge

GetInEDat(e) : return data associated with e-th in-edge

4.2 Δημιουργία δικτύου συνεργασίας

Το δίκτυο συνεργασίας των επιστημονικών συνεργασιών του Πανεπιστημίου Κύπρου είναι ένας μη κατευθυνόμενος, διμερής γράφος με δεδομένα στους κόμβους του. Αυτό το δίκτυο συνεργασίας αποτελείται από δύο διμερείς γράφους. Ο πρώτος γράφος απεικονίζει τις επιστημονικές συνεργασίες και ο δεύτερος απεικονίζει τις συνεργασίες των ιδρυμάτων.

Συγκεκριμένα, στον πρώτο διμερή γράφο $G = (U, V, E)$ οι κόμβοι που ανήκουν στο σύνολο U είναι οι ίδιοι οι επιστήμονες (επιστήμονες του Πανεπιστημίου Κύπρου και οι άλλοι επιστήμονες-συνεργάτες) και οι κόμβοι που ανήκουν στο σύνολο V είναι οι μελέτες τους, δηλαδή αυτό που τους συσχετίζει, οι οποίες είναι άρθρα τα οποία έχουν γράψει μαζί, τα οποία έχουν δημοσιευτεί στο διαδίκτυο ή σε επιστημονικά περιοδικά. Έτσι σε κάθε κόμβο τύπου άρθρο υπάρχουν ακμές E με τον κάθε κόμβο τύπου u επιστήμονα, ο οποίος αντιστοιχεί σε ένα από τους συγγραφείς του συγκεκριμένου άρθρου.

- Κάθε κόμβος τύπου επιστήμονας κρατάει πληροφορίες σχετικές με την ‘ταυτότητά’ του επιστήμονα και την τιμή που δηλώνει αν ο συγκεκριμένος επιστήμονας ανήκει στο Πανεπιστήμιο Κύπρου.
- Κάθε κόμβος τύπου άρθρο κρατάει πληροφορίες σχετικές με τον αριθμό που ταυτοποιεί το άρθρο και τη χρονιά δημοσίευσης του άρθρου.

Στο δεύτερο διμερή γράφο $G = (U, V, E)$ οι κόμβοι που ανήκουν στο σύνολο U είναι τα ιδρύματα-οργανισμοί (πανεπιστήμια) και οι κόμβοι που ανήκουν στο σύνολο V είναι τα άρθρα των επιστημόνων που ανήκουν σε αυτά τα πανεπιστήμια. Έτσι σε κάθε κόμβο τύπου άρθρο υπάρχουν ακμές E με τον κάθε κόμβο τύπου πανεπιστήμιο.

- Κάθε κόμβος τύπου ίδρυμα κρατάει πληροφορίες σχετικές με την ‘ταυτότητά’ του πανεπιστημίου και τον αριθμό κατάταξής του στη λίστα των καλύτερων πανεπιστημίων στον κόσμο (με βάση τη λίστα της Shanghai για το 2009).
- Κάθε κόμβος τύπου άρθρο κρατάει πληροφορίες σχετικές με τον αριθμό που ταυτοποιεί το άρθρο και τη χρονιά δημοσίευσης του άρθρου.

4.2.1 Περιγραφή κλάσεων

Για τη δημιουργία του δικτύου συνεργασίας επιστημονικών συνεργασιών του Πανεπιστημίου Κύπρου ορίστηκαν οι δύο κύριες κλάσεις του συστήματος:

1. TUcyCollNode
2. TUcyCollNet

1. *TUcyCollNode*

Η κλάση αυτή αναφέρεται στον κόμβο του δικτύου.

Η κλάση TUcyCollNode έχει ως ιδιωτικά χαρακτηριστικά της μια ακέραια τιμή, η οποία αναφέρεται στην ‘ταυτότητα’ επιστήμονα ή άρθρου ή του ιδρύματος, τη χρονιά δημοσίευσης του άρθρου, την τιμή εκείνη που δηλώνει αν ένας επιστήμονας ανήκει στο Πανεπιστήμιο Κύπρου, τον αριθμό κατάταξης του ιδρύματος και τον τύπο του κόμβου. Αυτό που κάνει ένα κόμβο να ξεχωρίζει ο τύπος του, είναι ποιος constructor καλείται κατά τη δημιουργία ενός αντικειμένου αυτής της κλάσης.

Στην προκειμένη περίπτωση υπάρχουν τέσσερις constructors.

- Εάν κατά τη δημιουργία αντικειμένου της κλάσης TUcyCollNode δεν δοθούν παραμέτροι, τότε αυτόματα ο τύπος του κόμβου αυτού δηλώνεται ως απροσδιόριστος.
- Εάν κατά τη δημιουργία αντικειμένου της κλάσης TUcyCollNode δοθούν τρεις παραμέτροι και δύο τύπου integer και μία τύπου ‘τύπος κόμβου’, τότε αυτόματα δημιουργείται ένας κόμβος τύπου άρθρο.
- Εάν κατά τη δημιουργία αντικειμένου της κλάσης TUcyCollNode δοθούν δύο παραμέτροι εκ των οποίων η πρώτη τύπου integer και η δεύτερη τύπου boolean, τότε αυτόματα δημιουργείται ένας κόμβος τύπου επιστήμονας.
- Τέλος, αν κατά τη δημιουργία αντικειμένου της κλάσης TUcyCollNode δοθούν δύο παραμέτροι τύπου integer, τότε αυτόματα δημιουργείται ένας κόμβος τύπου ίδρυμα.

Στο σώμα της κλάσης `TUcyCollNode` δηλώθηκε πως η κλάση `TUcyCollNet` είναι friend αυτής της κλάσης και αυτό έχει ως αποτέλεσμα όλα τα μέλη της κλάσης `TUcyCollNet` να μπορούν να έχουν πρόσβαση στα ιδιωτικά χαρακτηριστικά (private members) της κλάσης `TUcyCollNode`.

2. *TUcyCollNet*

Η κλάση αυτή αναφέρεται στο δικτύου συνεργασίας και κληρονομεί όλα τα μέλη της κλάσης `TNodeNet<TUcyCollNet>`, στα οποία μπορεί να έχει πρόσβαση.

Στο ιδιωτικό κομμάτι της κλάσης δηλώνεται ένας hash table, ο οποίος κρατά το κλειδί το κάθε κόμβου που στο δίκτυο τον κάνει μοναδικό.

Στο δημόσιο κομμάτι της κλάσης ορίστηκαν και υλοποιήθηκαν οι μέθοδοι αυτής της κλάσης. Κάποιες από αυτές ήταν η δημιουργία ενός δικτύου με ή χωρίς παραμέτρους. Επιπλέον, ορίστηκαν και υλοποιήθηκαν οι έξι βασικές μέθοδοι. Μέσω αυτών των μεθόδων δημιουργείται τόσο ο γράφος που απεικονίζει τη συνεργασία επιστημόνων όσο και ο γράφος που απεικονίζει τις συνεργασίες των ιδρυμάτων, διαβάζονται τα δεδομένα από κάποιο αρχείο κειμένου και δημιουργείται το δίκτυο επιστημονικών συνεργασιών του Πανεπιστημίου Κύπρου και το δίκτυο συνεργασίας ιδρυμάτων, υπολογίζονται και τυπώνονται τα στατιστικά στοιχεία του δικτύου συνεργασίας τόσο των επιστημόνων όσο και των ιδρυμάτων.

Στο σώμα της κλάσης `TUcyCollNet` δηλώθηκε πως η κλάση `TPt<TUcyCollNet>` είναι friend αυτής της κλάσης και αυτό έχει ως αποτέλεσμα όλα τα μέλη της κλάσης `TPt<TUcyCollNet>` να μπορούν να έχουν πρόσβαση στα ιδιωτικά χαρακτηριστικά (private members) της κλάσης `TUcyCollNet`.

Στο κύριο σώμα του κώδικα υπάρχει η κύρια συνάρτηση μέσω της οποίας στην οθόνη του χρήστη εμφανίζεται αρχικά το κυρίως μενού. Στο κυρίως μενού ο χρήστης έχει να επιλέξει αν θέλει να δημιουργήσει και να αναλύσει ένα δίκτυο επιστημονικών συνεργασιών ή ένα δίκτυο συνεργασίας ιδρυμάτων. Ανάλογα με την επιλογή του, εμφανίζεται το κατάλληλο υπομενού. Αν επιλέξει να δημιουργηθεί και να αναλυθεί ένα

επιστημονικό δίκτυο συνεργασίας, τότε ο χρήστης μπορεί να επιλέξει αν το επιστημονικό δίκτυο συνεργασίας θα αναφέρεται σε ένα τμήμα όλων των χρονιών(1992-2009) ή σε ένα τμήμα συγκεκριμένης χρονιάς. Αν τώρα επιλέξει να δημιουργηθεί και να αναλυθεί ένα δίκτυο συνεργασίας ιδρυμάτων, τότε ο χρήστης μπορεί να επιλέξει αν το επιστημονικό δίκτυο συνεργασίας θα αναφέρεται σε ένα συγκεκριμένο τμήμα. Σε κάθε επιλογή του χρήστη γίνεται έλεγχος λάθους δεδομένων εισόδου και σε κάθε μενού ο χρήστης μπορεί να τερματίσει το πρόγραμμα.

4.2.2 Περιγραφή δημιουργίας δικτύου συνεργασίας

Για τη δημιουργία τόσο του δικτύου συνεργασίας των επιστημόνων όσο και του δικτύου συνεργασίας των ιδρυμάτων ακολουθήθηκε ο ίδιος τρόπος σκέψης με μικρές διαφορές γι' αυτό ακολουθεί μόνο η περιγραφή του επιστημονικού δικτύου συνεργασίας.

Το δίκτυο συνεργασίας δημιουργείται με την κλήση της μεθόδου

PUcyCollNet TUcyCollNet::LoadTxt(TStr& FileName). Αυτή η μέθοδος παίρνει ως παράμετρο το όνομα του αρχείου κειμένου, από το οποίο διαβάζονται τα δεδομένα για τη δημιουργία του δικτύου συνεργασίας και επιστρέφει το δείκτη που δείχνει προς στο δίκτυο συνεργασίας.

Στο σώμα αυτής της μεθόδου, αρχικά δηλώνεται ένας δείκτης, ο οποίος δείχνει σε ένα δίκτυο τύπου TUcyCollNet. Στη συνέχεια, το αρχείο κειμένου που δίνεται μέσω της παραμέτρου ανοίγει. Το αρχείο αυτό διαβάζεται γραμμή-γραμμή. Κάθε γραμμή σπάει σε μικρότερες συμβολοσειρές, οι οποίες αποθηκεύονται σε μια μεταβλητή τύπου vector. Κάθε φορά που εμφανίζεται ο χαρακτήρας ' ' τότε συμβολοσειρά που αντιστοιχεί σε ολόκληρη τη γραμμή σπάει σε μικρότερες συμβολοσειρές.

Η μεταβλητή τύπου Vector έχει μέγεθος τέσσερα. Στην πρώτη θέση υπάρχει ο αριθμός που ταυτοποιεί μοναδικά το άρθρο, στη δεύτερη θέση υπάρχει η 'ταυτότητα' του επιστήμονα που είναι ένας από τους συγγραφείς του συγκεκριμένου άρθρου, στην τρίτη θέση υπάρχει η τιμή που καθορίζει αν ο συγκεκριμένος επιστήμονας ανήκει στο Πανεπιστήμιο Κύπρου και στην τέταρτη θέση υπάρχει η χρονιά δημοσίευσης του

άρθρου. Αφού επεξεργαστούν και αποθηκευτούν οι πιο πάνω τιμές στις ανάλογες τοπικές μεταβλητές, στη συνέχεια δημιουργούνται οι δύο τύπου κόμβοι. Κατά τη φάση δημιουργίας των κόμβων του δικτύου, καλείται ο κατάλληλος constructor για τη δημιουργία του αντίστοιχου τύπου κόμβου. Ο κόμβος που δημιουργείται, προστίθεται στο δίκτυο συνεργασίας και τα αντίστοιχα δεδομένα που κρατά ο συγκεκριμένος κόμβος αποθηκεύονται στον ανάλογο hash table. Συγκεκριμένα, υπάρχουν δύο hash tables, ο ScientistNidH και ο ArticleNidH. Ο πρώτος κρατά τα δεδομένα που αναφέρονται σε ένα επιστήμονα ('ταυτότητα', αν ανήκει στο Πανεπιστήμιο Κύπρου) και ο δεύτερος κρατά τα δεδομένα που αναφέρονται σε ένα άρθρο ('ταυτότητα' άρθρου, χρονιά δημοσίευσης).

Ακολούθως, δημιουργούνται οι ακμές του δικτύου συνεργασίας. Οι ακμές αυτές συνδέουν ένα κόμβο τύπου επιστήμονα με ένα κόμβο τύπου άρθρο.

Στο τέλος, αφού διαβαστούν όλες οι γραμμές του αρχείου κειμένου, το αρχείο κειμένου κλείνει και η μέθοδος επιστρέφει το δείκτη που δείχνει στο δίκτυο συνεργασίας που δημιουργήθηκε.

4.2.3 Περιγραφή δημιουργίας γράφου

Για τη δημιουργία τόσο του γράφου των επιστημόνων όσο και του γράφου των ιδρυμάτων ακολουθήθηκε ο ίδιος τρόπος σκέψης με μικρές διαφορές γι' αυτό ακολουθεί μόνο η περιγραφή του γράφου επιστημόνων.

Ο γράφος που δημιουργείται με την κλήση της μεθόδου

PUNGraph TUcyCollNet::GetScientistGraph(), απεικονίζει τη συνεργασία των επιστημόνων. Συγκεκριμένα, αυτός ο γράφος δεν είναι διμερής, αλλά ένας κανονικός μη κατευθυνόμενος γράφος, ο οποίος αποτελείται από κόμβους, οι οποίοι αντιστοιχούν σε ένα επιστήμονα, και από ακμές, οι οποίες συνδέουν δύο κόμβους εάν δυο επιστήμονες συνεργάστηκαν για τη συγγραφή ενός άρθρου. Η δημιουργία του γράφου γίνεται με βάση το δίκτυο συνεργασίας που προηγουμένως έχει δημιουργηθεί. Συνεπώς, πριν τη δημιουργία του γράφου είναι αναγκαία η δημιουργία του δικτύου επιστημονικών συνεργασιών του Πανεπιστημίου Κύπρου. Κατά τη δημιουργία του

γράφου είναι αναγκαίο να κρατάμε τον αριθμό εκείνο που αντιστοιχεί στις πόσες φορές εμφανίζεται η κάθε ακμή μεταξύ κάθε συγκεκριμένου ζεύγους κόμβων γιατί η κάθε ακμή δημιουργείται μόνο μια φορά. Αυτό καθίσταται αναγκαίο γιατί στη συνέχεια αυτή η πληροφορία θα είναι χρήσιμη για την ανάλυση του δικτύου συνεργασίας. Αφού δημιουργηθεί ο γράφος, η μέθοδος επιστρέφει το δείκτη που δείχνει σε αυτό το γράφο.

Κεφάλαιο 5

Ανάλυση Δικτύου Συνεργασίας

5.1 Κριτήρια ανάλυσης δικτύου συνεργασίας	59
5.2 Περιγραφή εξόρυξης γενικών στατιστικών δεδομένων	63
5.3 Περιγραφή ανάλυσης δικτύου συνεργασίας	65
5.2.1 Ανάλυση δικτύου συνεργασίας επιστημόνων	65
5.2.2 Ανάλυση δικτύου συνεργασίας ιδρυμάτων-οργανισμών	68
5.4 Αποτελέσματα εξόρυξης γενικών στατιστικών δεδομένων	71
5.5 Αποτελέσματα επιστημονικού δικτύου συνεργασίας	72
5.6 Αποτελέσματα δικτύου συνεργασίας ιδρυμάτων	102

5.1 Κριτήρια ανάλυσης δικτύου συνεργασίας

Μετά τη δημιουργία του δικτύου συνεργασίας του Πανεπιστημίου Κύπρου, θα πρέπει να είναι κατανοητό το πρότυπο σύνδεσης και επικοινωνίας μεταξύ των επιστημόνων (ανά τμήμα και ανά τμήμα και χρονιά) και μεταξύ των ιδρυμάτων(ανά τμήμα). Συνεπώς, αυτό για να επιτευχθεί θα πρέπει μετά την ανάλυση του δικτύου συνεργασίας να έχουμε και τα εξής στατιστικά δεδομένα:

Γενικά στατιστικά δεδομένα:

- Αριθμός τμημάτων προς μελέτη
- Αριθμός όλων των επιστημόνων
- Αριθμός των επιστημόνων που ανήκουν στο Πανεπιστήμιο Κύπρου
- Αριθμός των επιστημόνων του Πανεπιστήμιο Κύπρου που ανήκουν σε συγκεκριμένο τμήμα προς μελέτη
- Αριθμός συνολικών δημοσιεύσεων
- Αριθμός συνολικών δημοσιεύσεων ανά τμήμα προς μελέτη

- Αριθμός ιδρυμάτων
- Αριθμός ακμών μεταξύ κόμβου τύπου άρθρο και τύπου συγγραφέας
- Αριθμός ακμών μεταξύ κόμβου τύπου άρθρο και τύπου ίδρυμα

Ειδικά στατιστικά δεδομένα για το δίκτυο συνεργασίας επιστημόνων:

- Αριθμός επιστημόνων
- Αριθμός άρθρων
- Αριθμός ακμών στο γράφο
- Μέσος όρος clustering coefficient του δικτύου
- Αριθμός άρθρων για κάθε επιστήμονα που ανήκει στο Πανεπιστήμιο Κύπρου (διάμεση τιμή)
- Αριθμός επιστημόνων-συγγραφέων για κάθε άρθρο(διάμεση τιμή)
- Αριθμός εσωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου (διάμεση τιμή)
- Αριθμός εξωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου (διάμεση τιμή)
- Degree Distribution
- Giant component – αριθμός κόμβων
- Giant component – μέσος όρος κόμβων

Ειδικά στατιστικά δεδομένα για το δίκτυο συνεργασίας ιδρυμάτων:

- Αριθμός ιδρυμάτων
- Αριθμός άρθρων
- Αριθμός ακμών στο γράφο
- Μέσος όρος clustering coefficient του δικτύου
- Αριθμός άρθρων για κάθε ίδρυμα (διάμεση τιμή)
- Αριθμός ιδρυμάτων για κάθε άρθρο (διάμεση τιμή)
- Degree Distribution
- Τα 10 πρώτα ιδρύματα(ονόματα και αριθμό κατάταξης παγκοσμίως), τα οποία συνεργάστηκαν περισσότερο με το Πανεπιστήμιο Κύπρου.

Επεξήγηση ειδικών κριτηρίων δικτύου συνεργασίας

- **Αριθμός επιστημόνων**
Δίνεται ο συνολικός αριθμός των επιστημόνων που συνεργάστηκαν με τουλάχιστον ένα επιστήμονα για τη συγγραφή ενός άρθρου.
- **Αριθμός άρθρων**
Δίνεται ο συνολικός αριθμός άρθρων που επεξεργάστηκαν για τη δημιουργία του δικτύου συνεργασίας.
- **Αριθμός ιδρυμάτων**
Δίνεται ο συνολικός αριθμός ιδρυμάτων-οργανισμών που οι επιστήμονες τους συνεργάστηκαν για τη συγγραφή άρθρων.
- **Αριθμός ακμών**
Δίνεται ο συνολικός αριθμός ακμών του δικτύου συνεργασίας. Οι ακμές αυτές είτε συνδέουν ένα κόμβο τύπου επιστήμονα με ένα κόμβο τύπου επιστήμονα είτε συνδέουν ένα κόμβο τύπου ίδρυμα με ένα κόμβο τύπου ίδρυμα.
- **Αριθμός άρθρων για κάθε επιστήμονα που ανήκει στο Πανεπιστήμιο Κύπρου**
Δίνεται ο αριθμός των άρθρων που μπορεί ένας επιστήμονας του Πανεπιστημίου Κύπρου να συνεργάστηκε με έναν άλλο τουλάχιστον επιστήμονα (οποιοδήποτε άλλο ιδρύματος ή οργανισμού) για τη συγγραφή αυτών των άρθρων. Ο αριθμός αυτός αντιστοιχεί στη διάμεση τιμή της κατανομής.
- **Αριθμός επιστημόνων-συγγραφέων για κάθε άρθρο**
Για κάθε άρθρο δίνεται ο αριθμός των επιστημόνων που συνεργάστηκαν για τη συγγραφή ενός άρθρου. Ο αριθμός αυτός αντιστοιχεί στη διάμεση τιμή της κατανομής και σίγουρα είναι ίσος ή μεγαλύτερος από δύο αφού για κάθε άρθρο ο αριθμός των συγγραφέων είναι τουλάχιστον δύο.

- Αριθμός εσωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου
Δίνεται ο αριθμός επιστημόνων, οι οποίοι ανήκουν στο Πανεπιστήμιο Κύπρου και συνεργάστηκαν με ένα επιστήμονα, ο οποίος επίσης ανήκει στο Πανεπιστήμιο Κύπρου. Ο αριθμός αυτός αντιστοιχεί στη διάμεση τιμή της κατανομής.
- Αριθμός εξωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου
Δίνεται ο αριθμός επιστημόνων, οι οποίοι δεν ανήκουν στο Πανεπιστήμιο Κύπρου και συνεργάστηκαν με ένα επιστήμονα, ο οποίος ανήκει στο Πανεπιστήμιο Κύπρου. Ο αριθμός αυτός αντιστοιχεί στη διάμεση τιμή της κατανομής.
- Μέσος όρος clustering coefficient του δικτύου
Δίνεται μια τιμή, η οποία βασικά εκτιμά το βαθμό με τον οποίο κάποιοι κόμβοι του δικτύου τείνουν να γίνονται ένα σύμπλεγμα μαζί, να δημιουργούν έτσι κάποιες ομάδες οι οποίες χαρακτηρίζονται σχετικά από μεγάλη πυκνότητα από δεσμούς. Με πιο απλά λόγια, είναι η πιθανότητα δυο διαφορετικοί συνεργάτες κάποιου επιστήμονα, να είναι τελικά και μεταξύ τους συνεργάτες. Αυτός ο συντελεστής για ένα γράφο συνεργασίας είναι το κατά μέσο όρο κλάσμα ζευγαριών επιστημόνων που επίσης συνεργάστηκαν με κάποιο άλλο επιστήμονα [3]. Μαθηματικά, έχει οριστεί ως:

$$C = \frac{3 \times \text{number of triangles}}{\text{Number of connected triple of vertices}} = \frac{\text{number of closed triples}}{\text{Number of connected triple of vertices}}$$

- Degree Distribution
Δίνεται ο διάμεσος αριθμός συνδέσεων των κόμβων είτε του γράφου επιστημόνων είτε του γράφου ιδρυμάτων-οργανισμών.
- Giant component (αριθμός κόμβων - μέσος όρος κόμβων)
Δίνεται ο αριθμός των κόμβων, οι οποίοι ανήκουν σε έναν υπογράφο(subgraph), ο οποίος περιέχει την πλειοψηφία των κόμβων ολόκληρου του γράφου. Συνεπώς,

οι κόμβοι που ανήκουν σε αυτόν τον υπογράφο(subgraph) είναι κόμβοι που παρατηρούνται σε μονοπάτια με το μικρότερο αριθμό ‘ενδιάμεσων κόμβων’ μεταξύ άλλων κόμβων. Τέτοιοι κόμβοι τείνουν να είναι πιο ισχυροί (ασκούν περισσότερη επιρροή) στο δίκτυο και συμπεριφέρονται ως information hubs.

5.2 Περιγραφή εξόρυξης γενικών στατιστικών δεδομένων

Στην προκειμένη περίπτωση, για την εξόρυξη των γενικών στατιστικών δεδομένων αντλήθηκαν πληροφορίες από τη βάση δεδομένων και υπολογίστηκαν οι τιμές μέσω ενός PHP script.

- Αριθμός τμημάτων προς μελέτη
Ο αριθμός των τμημάτων καθορίστηκε από την επιλογή της σχολής προς μελέτη. Στην προκειμένη περίπτωση, επιλέχθηκε η Σχολή Θετικών και Εφαρμοσμένων Επιστημών άρα τα τμήματα της μελέτης είναι πέντε (τμήμα Βιολογικών Επιστημών, τμήμα Χημείας, τμήμα Πληροφορικής, τμήμα Μαθηματικών και Στατιστικής και τμήμα Φυσικής).
- Αριθμός όλων των επιστημόνων
Ο αριθμός όλων των επιστημόνων ισούται με τον αριθμό όλων των εγγραφών του πίνακα *scientist* της βάσης δεδομένων.
- Αριθμός των επιστημόνων που ανήκουν στο Πανεπιστήμιο Κύπρου
Ο αριθμός όλων των επιστημόνων που ανήκουν στο Πανεπιστήμιο Κύπρου ισούται με τον αριθμό όλων των εγγραφών του πίνακα *scientist* της βάσης δεδομένων που στο πεδίο *AffCyprus* υπάρχει η τιμή 1.
- Αριθμός των επιστημόνων του Πανεπιστήμιο Κύπρου που ανήκουν σε συγκεκριμένο τμήμα προς μελέτη
Ο αριθμός όλων των επιστημόνων ισούται με τον αριθμό όλων των εγγραφών του πίνακα *belongs* της βάσης δεδομένων όπου στο πεδίο *DName* υπάρχει το όνομα του αντίστοιχου τμήματος.

- Αριθμός συνολικών δημοσιεύσεων
Ο αριθμός όλων των δημοσιεύσεων ισούται με τον αριθμό όλων των άρθρων του πίνακα *article* της βάσης δεδομένων. Από τον αριθμό αυτό αφαιρέθηκαν τα άρθρα τα οποία έχουν περισσότερους από 100 συγγραφείς.
- Αριθμός συνολικών δημοσιεύσεων ανά τμήμα προς μελέτη
Για τον υπολογισμό του αριθμού συνολικών δημοσιεύσεων για κάθε τμήμα, αρχικά αποθηκεύτηκαν σε πίνακες οι ταυτότητες των επιστημόνων του Πανεπιστημίου Κύπρου που ανήκουν στα αντίστοιχα τμήματα και επιλέχθηκαν οι εγγραφές του πίνακα *is_written* της βάσης δεδομένων, στις οποίες στο πεδίο *SID* υπάρχει η ταυτότητα των αντίστοιχων επιστημόνων. Έτσι, ο αριθμός των συνολικών δημοσιεύσεων ανά τμήμα ισούται με τον αριθμό των επιλεγμένων εγγραφών του πίνακα *is_written* αφαιρώντας τα διπλότυπα και τις εγγραφές εκείνες που συναντούμε ταυτότητες άρθρων που έχουν περισσότερους από 100 συγγραφείς.
- Αριθμός ιδρυμάτων
Ο αριθμός των ιδρυμάτων ισούται με τον αριθμό όλων των εγγραφών του πίνακα *new_institutions* της βάσης δεδομένων.
- Αριθμός ακμών μεταξύ κόμβου τύπου άρθρο και τύπου συγγραφέας
Ο αριθμός ακμών μεταξύ κόμβου τύπου άρθρο και τύπου συγγραφέας ισούται με τον αριθμό όλων των εγγραφών του πίνακα *is_written* της βάσης δεδομένων, οι οποίες αναφέρονται στα πέντε τμήματα που έχουν επιλεγθεί να μελετηθούν και σε άρθρα που δεν έχουν περισσότερους από 100 επιστήμονες.
- Αριθμός ακμών μεταξύ κόμβου τύπου άρθρο και τύπου ίδρυμα
Ο αριθμός ακμών μεταξύ κόμβου τύπου άρθρο και τύπου ίδρυμα ισούται με τον αριθμό όλων των εγγραφών του πίνακα *new_coll_institutions* της βάσης δεδομένων, οι οποίες αναφέρονται στα πέντε τμήματα που έχουν επιλεγθεί να μελετηθούν και σε άρθρα που δεν έχουν περισσότερους από 100 επιστήμονες.

5.3 Περιγραφή ανάλυσης δικτύου συνεργασίας

Για την ανάλυση του δικτύου συνεργασίας του Πανεπιστημίου Κύπρου χρειάστηκε να υπολογιστούν στατιστικά δεδομένα τόσο για το δίκτυο συνεργασίας επιστημόνων όσο και για το δίκτυο συνεργασίας ιδρυμάτων. Για τα δύο αυτά δίκτυα ακολουθήθηκε διαφορετική διαδικασία.

5.3.1 Ανάλυση δικτύου συνεργασίας επιστημόνων

Για την ανάλυση του δικτύου συνεργασίας επιστημόνων που δημιουργείται στην προηγούμενη φάση, καλείται η μέθοδος **void printStatistics(const PUcyCollNet& net, const PUNGraph& G, const int depart, const int year)** της κλάσης *TUcyCollNet*. Η μέθοδος αυτή υπολογίζει τα στατιστικά στοιχεία που προκύπτουν με βάση τα κριτήρια ανάλυσης του δικτύου συνεργασίας επιστημόνων.

Η μέθοδος αυτή παίρνει ως παραμέτρους ένα δείκτη προς το δίκτυο συνεργασίας, ένα άλλο δείκτη προς το γράφο των επιστημόνων, ένα ακέραιο αριθμό που αντιστοιχεί στο τμήμα επιλογής του χρήστη και ένα ακέραιο αριθμό που αντιστοιχεί στη χρονιά δημοσίευσης του άρθρων. Στο σώμα αυτής της μεθόδου υπολογίζονται στατιστικά δεδομένα, τα οποία στη συνέχεια αποθηκεύονται σε αρχεία κειμένου.

Αρχικά, αν η μέθοδος παίρνει ως χρονιά δημοσίευσης των άρθρων την τιμή 0, τότε αυτόματα δημιουργούνται τα αρχεία κειμένου, στα οποία δίνονται οι κατάλληλοι ονοματισμοί προκειμένου να καταλαβαίνουμε πως περιέχουν τα στατιστικά δεδομένα για άρθρα συγκεκριμένου τμήματος αλλά όλων των χρονιών (1992-2009). Σε αντίθετη περίπτωση, δημιουργούνται τα αρχεία κειμένου που περιέχουν τα στατιστικά δεδομένα για άρθρα συγκεκριμένου τμήματος και συγκεκριμένης χρονιάς δημοσίευσης. Και σε αυτά τα αρχεία δίνονται κατάλληλοι ονοματισμοί προκειμένου να καταλαβαίνουμε σε ποιο τμήμα και χρονιά δημοσίευσης αναφέρονται.

Στη συνέχεια, υπολογίζεται ο συνολικός αριθμός των ακμών του επιστημονικού γράφου. Η βιβλιοθήκη SNAP παρέχει τη συνάρτηση `GetEdges()`, η οποία υπολογίζει τον αριθμό των ακμών ενός γράφου αλλά στην προκειμένη περίπτωση δεν μας ικανοποιεί. Γιατί; Κατά τη δημιουργία του γράφου μό v μια ακμή δημιουργόταν μεταξύ δυο επιστημόνων, ακόμη κι αν συνεργαστήκαν περισσότερες από μία φορές. Γι' αυτό και στη μέθο `ο δ PUNGraph TUcyCollNet::GetScientistGraph()` υπολογίζεται το βάρος κάθε ακμής. Έτσι, με βάσει αυτή την πληροφορία· αθροίζοντας τα βάρη όλων των ακμών του γράφου επιστημόνων υπολογίζεται ο συνολικός αριθμός ακμών. Ακολούθως, υπολογίζεται ο μέσος όρος `clustering coefficient` χρησιμοποιώντας τη συνάρτηση `GetClustCf(TUNGraph G)` που παρέχει η βιβλιοθήκη SNAP.

Έπειτα, υπολογίζονται ο αριθμός των επιστημόνων και ο αριθμός των άρθρων. Ο αριθμός των επιστημόνων ισούται με το μέγεθος του hash table `ScientistNidH` και ο αριθμός των άρθρων ισούται με το μέγεθος του hash table `ArticleNidH`. Η βιβλιοθήκη SNAP παρέχει τη δυνατότητα στο χρήστη να μάθει γρήγορα το μέγεθος ενός hash table, καλώντας απλά τη μέθοδο `Len()` (π.χ. `ScientistNidH.Len()`). Αφού υπολογιστούν αυτές οι τέσσερις τιμές, τότε τυπώνονται σε ένα αρχείο κειμένου. Ο λόγος για τον οποίο τυπώνονται μόνο αυτά τα δεδομένα σε αυτό το αρχείο κειμένου· είναι γιατί τα υπόλοιπα στατιστικά αφού τυπωθούν στα υπόλοιπα αρχεία κειμένου θα τύχουν επιπλέον επεξεργασίας προκειμένου να πάρουμε την τελική πληροφορία που μας ενδιαφέρει.

Ακολούθως, αποθηκεύονται τα δεδομένα για την εξεύρεση της μέσης τιμής του αριθμού των επιστημόνων ανά άρθρο. Διασχίζοντας το δίκτυο συνεργασίας, όταν υπάρχει κόμβος τύπου άρθρο αποθηκεύονται σε ένα αρχείο κειμένου δυο στήλες από ακέραιους αριθμούς. Η πρώτη στήλη αντιστοιχεί στην ταυτότητα του άρθρου και η δεύτερη στο v αριθμό των ακμών που v έρχονται από κόμβους τύπου επιστήμονας προς αυτό τον κόμβο τύπου άρθρο. Ο αριθμός του κόμβου τύπου άρθρο είναι ίσος με την τιμή που επιστρέφει η μέθοδος `GetId()` και ο αριθμός των ακμών που έρχονται από κόμβους τύπου επιστήμονας προς αυτό τον κόμβο τύπου άρθρο ισούται με την τιμή που επιστρέφει η μέθοδος `GetInDeg()`.

Έπειτα, αποθηκεύονται τα δεδομένα για την εξεύρεση της διάμεσης τιμής του αριθμού των άρθρων ανά επιστήμονα που ανήκει στο Πανεπιστήμιο Κύπρου. Διασχίζοντας το δίκτυο συνεργασίας, όταν υπάρχει κόμβος τύπου επιστήμονας, ο οποίος ανήκει στο Πανεπιστήμιο Κύπρου, αποθηκεύονται σε ένα αρχείο κειμένου δυο στήλες από ακέραιους αριθμούς. Η πρώτη στήλη αντιστοιχεί στην ταυτότητα του επιστήμονα και η δεύτερη στον αριθμό των ακμών που φεύγουν από κόμβους τύπου επιστήμονας προς κόμβο τύπου άρθρο. Ο αριθμός του κόμβου τύπου επιστήμονας είναι ίσος με την τιμή που επιστρέφει η μέθοδος *GetId()* και ο αριθμός των ακμών που φεύγουν από κόμβους τύπου επιστήμονας προς κόμβο τύπου άρθρο ισούται με την τιμή που επιστρέφει η μέθοδος *GetOutDeg()*.

Στη συνέχεια, υπολογίζονται ο αριθμός εσωτερικών και εξωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου. Για κάθε επιστήμονα του Πανεπιστημίου Κύπρου και για κάθε ακμή του επιστημονικού γράφου, ελέγχεται κατά πόσο ο επιστήμονας με τη συγκεκριμένη ταυτότητα είναι μέρος της κάθε ακμής. Αν ναι, είτε είναι η πηγή είτε ο προορισμός, ελέγχεται κατά πόσο ο άλλος επιστήμονας ανήκει στο Πανεπιστήμιο Κύπρου. Αυτό γίνεται με την κλήση των μεθόδων *GetDat().GetFromUCY()*. Με τη μέθοδο *GetDat()* αποκτούμε πρόσβαση στα δεδομένα του κόμβου και με τη μέθοδο *GetFromUCY()* παίρνουμε την τιμή ένα ή μηδέν. Αν πάρουμε την τιμή μηδέν, αυξάνεται ο μετρητής των επιστημόνων που συνεργάστηκε ο επιστήμονας του Πανεπιστημίου Κύπρου και δεν ανήκουν στο Πανεπιστήμιο Κύπρου. Αν πάρουμε την τιμή ένα, αυξάνεται ο μετρητής των επιστημόνων που συνεργάστηκε ο επιστήμονας του Πανεπιστημίου Κύπρου και ανήκουν στο Πανεπιστήμιο Κύπρου. Προτού συνεχίσει ο έλεγχος για τον επόμενο επιστήμονα του Πανεπιστημίου Κύπρου, δημιουργούνται δύο αρχεία κειμένου με δύο στήλες από ακέραιους αριθμούς αντίστοιχα. Στο πρώτο αρχείο κειμένου, η πρώτη στήλη αντιστοιχεί στην ταυτότητα του επιστήμονα του Πανεπιστημίου Κύπρου και η δεύτερη στον αριθμό των εσωτερικών συνεργατών του και στο δεύτερο αρχείο κειμένου, η πρώτη στήλη αντιστοιχεί στην ταυτότητα του επιστήμονα του Πανεπιστημίου Κύπρου και η δεύτερη στον αριθμό των εξωτερικών συνεργατών του.

Τέλος, υπολογίζονται το degree κάθε κόμβου τύπου επιστήμονα στο γράφο συνεργασίας επιστημόνων και το giant component του γράφου επιστημόνων. Συγκεκριμένα, για τον υπολογισμό του degree κάθε κόμβου τύπου επιστήμονα στο γράφο συνεργασίας επιστημόνων, δημιουργείται ένα αρχείο κειμένου τυπώνονται στο οποίο τυπώνονται δύο στήλες από ακέραιους αριθμούς. Η πρώτη στήλη αντιστοιχεί στην ταυτότητα του επιστήμονα και η δεύτερη αντιστοιχεί στον αριθμό ακμών που εξέρχονται από τον κόμβο τύπου επιστήμονα (degree). Η δεύτερη τιμή ανακτάται από τη κλήση της μεθόδου *GetOutDeg()*. Για τον υπολογισμό του giant component του γράφου επιστημόνων καλείται η μέθοδος *TSnap:: GetMxScc(G)*. Η μέθοδος αυτή επιστρέφει ένα γράφο ο οποίος περιέχει την πλειοψηφία των κόμβων που απέχουν μεταξύ τους πολύ λίγα βήματα και που ανήκουν στον αρχικό γράφο. Στη συνέχεια, υπολογίζεται ο αριθμός των κόμβων που βρίσκονται στο νέο γράφο με την κλήση της μεθόδου *G2->GetNodes()* και υπολογίζεται ο μέσος όρος των κόμβων που βρίσκονται στο νέο γράφο σε σύγκριση με τον αρχικό μέσω της εξής πράξης: $G2->GetNodes() / G->GetNodes()$. Τέλος, δημιουργείται ένα αρχείο κειμένου, στο οποίο τυπώνονται ο αριθμός των κόμβων και ο μέσος αριθμός των κόμβων που ανήκουν στο giant component.

Στη συνέχεια, για την εξεύρεση της διάμεσης τιμής του αριθμού άρθρων για κάθε επιστήμονα που ανήκει στο Πανεπιστήμιο Κύπρου, του αριθμού επιστημόνων-συγγραφέων για κάθε άρθρο, του αριθμού εσωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου, του αριθμού εξωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου και του degree distribution, τα δεδομένα από τα αντίστοιχα αρχεία κειμένου επεξεργάζονται με τη χρήση του προγράμματος Microsoft Office Excel 2007 και συγκεκριμένα με τη χρήση της συνάρτησης Median.

5.3.2 Ανάλυση δικτύου συνεργασίας ιδρυμάτων-οργανισμών

Για την ανάλυση του δικτύου συνεργασίας των ιδρυμάτων-οργανισμών που δημιουργείται στην προηγούμενη φάση, καλείται η μέθοδος **void printStatistics2(const PUcyCollNet& net, const PUNGraph& G, const int selection)** της κλάσης *TUcyCollNet*. Η μέθοδος αυτή υπολογίζει τα στατιστικά

στοιχεία που προκύπτουν με βάση τα κριτήρια ανάλυσης του δικτύου συνεργασίας ιδρυμάτων-οργανισμών.

Η μέθοδος αυτή παίρνει ως παραμέτρους ένα δείκτη προς το δίκτυο συνεργασίας, ένα άλλο δείκτη προς το γράφο των επιστημόνων και ένα ακέραιο αριθμό που αντιστοιχεί στο τμήμα επιλογής του χρήστη. Στο σώμα αυτής της μεθόδου υπολογίζονται στατιστικά δεδομένα, τα οποία στη συνέχεια αποθηκεύονται σε αρχεία κειμένου.

Αρχικά, ανάλογα με τον αριθμό επιλογής του χρήστη από το μενού δημιουργούνται τα αρχεία κειμένου, στα οποία δίνονται οι κατάλληλοι ονοματισμοί προκειμένου να καταλαμβαίνουμε πως περιέχουν τα στατιστικά δεδομένα για άρθρα συγκεκριμένου τμήματος.

Στη συνέχεια, υπολογίζεται ο συνολικός αριθμός των ακμών του γράφου ιδρυμάτων-οργανισμών. Για τον ίδιο λόγο που έχει προαναφερθεί, για τον υπολογισμό του συνολικού αριθμού ακμών του γράφου αυτού δεν χρησιμοποιήθηκε η συνάρτηση `GetEdges()` που παρέχει η βιβλιοθήκη SNAP, αλλά το βάρος της κάθε ακμής που υπολογίστηκε στο σώμα της μεθόδου `PUNGraph TUcyCollNet::GetScientistGraph()`. Έτσι, με βάση αυτή την πληροφορία, αθροίζοντας τα βάρη όλων των ακμών του επιστημονικού γράφου υπολογίζεται ο συνολικός αριθμός ακμών. Ακολούθως, υπολογίζεται ο μέσος όρος clustering coefficient χρησιμοποιώντας τη συνάρτηση `GetClustCf(TUNGraph G)` που παρέχει η βιβλιοθήκη SNAP.

Έπειτα, υπολογίζονται ο αριθμός των ιδρυμάτων-οργανισμών και ο αριθμός των άρθρων. Ο αριθμός των ιδρυμάτων-οργανισμών ισούται με το μέγεθος του hash table `InstitutionNIdH` και ο αριθμός των άρθρων ισούται με το μέγεθος του hash table `ArticleNIdH`. Η βιβλιοθήκη SNAP παρέχει τη δυνατότητα στο χρήστη να μάθει γρήγορα το μέγεθος ενός hash table, καλώντας απλά τη μέθοδο `Len()` (π.χ. `InstitutionNIdH.Len()`). Αφού υπολογιστούν αυτές οι τέσσερις τιμές, τότε τυπώνονται σε ένα αρχείο κειμένου. Ο λόγος για τον οποίο τυπώνονται μόνο αυτά τα δεδομένα σε αυτό το αρχείο κειμένου είναι γιατί τα υπόλοιπα στατιστικά αφού

τυπωθούν στα υπόλοιπα αρχεία κειμένου θα τύχουν επιπλέον επεξεργασίας προκειμένου να πάρουμε την τελική πληροφορία που μας ενδιαφέρει.

Στη συνέχεια, για το τμήμα επιλογής του χρήστη υπολογίζονται τα δέκα πρώτα ιδρύματα-οργανισμοί με τα οποία το Πανεπιστήμιο Κύπρου είχε τις περισσότερες συνεργασίες. Για κάθε ακμή του γράφου ιδρυμάτων, στην οποία το Πανεπιστήμιο Κύπρου είναι η πηγή ή ο προ ρισμός, αποθηκεύονται δεδομένα σε δύο μονοδιάστατους πίνακες. Στον πρώτο πίνακα αποθηκεύεται η ταυτότητα του ιδρυματος-οργανισμού με το οποίο συνεργάστηκε το Πανεπιστήμιο Κύπρου και στο δεύτερο πίνακα αποθηκεύονται το βάρος της αντίστοιχης ακμής. Ακολούθως, καλείται η συνάρτηση **void quick(int collInstDeg[], int collInst[], int count)**. Η συνάρτηση αυτή παίρνει ως παραμέτρους τους δύο μονοδιάστατους πίνακες καθώς και το μέγεθός το u . Στο σώμα αυτής της συνάρτησης καλείται η αναδρομική συνάρτηση **void qs(int collInstDeg[], int collInst[], int left, int right)**. Η συνάρτηση **qs** χρησιμοποιώντας τη μέθοδο επιλογής, ταξινομεί γρήγορα ένα πεπερασμένο σύνολο από ακέραιους αριθμούς (χρόνος ταξινόμησης : $O(n \log(n))$). Στην προκειμένη περίπτωση, ταξινομεί τον πίνακα με τα βάρη των ακμών. Σε κάθε μετατόπιση κάποιας τιμής, αυτόματα γίνεται και η αλλαγή στον πίνακα με τις ταυτότητες των ιδρυμάτων-οργανισμών προκειμένου στη τέλος να έχουμε στις αντίστοιχες θέσεις των δύο πινάκων τα δεδομένα που αναφέρονται στο ίδιο ίδρυμα-οργανισμό. Αφού ο πίνακας με τα βάρη των ακμών είναι ταξινομημένος κατά αύξουσα σειρά, επιλέγονται οι τελευταίες 10 θέσεις του πίνακα. Για κάθε μια από τις δέκα ταυτότητες, αποκτούμε πρό σβαση στον κόμβο τύπου ίδρυμα με τη συγκεκριμένη ταυτότητα και ανακτούμε το ranking number του ιδρυματος-οργανισμού. Έπειτα, σε ένα αρχείο κειμένου τυπώνονται δεδομένα σε τρεις στήλες ακέραιων αριθμών. Η πρώτη στήλη αντιστοιχεί στην ταυτότητα του ιδρυματος-οργανισμού, η δεύτερη στο ranking number του και η τρίτη στο βάρος ακμής του συγκεκριμένου ιδρυματος-οργανισμού με το Πανεπιστήμιο Κύπρου.

Ακολούθως, αποθηκεύονται τα δεδομένα για την εξεύρεση της διάμεσης τιμής του αριθμού των ιδρυμάτων-οργανισμών ανά άρθρο. Διασχίζοντας το δίκτυο συνεργασίας, όταν υπάρχει κόμβος τύπου ίδρυμα αποθηκεύονται σε ένα αρχείο κειμένου δυο στήλες από ακέραιους αριθμούς. Η πρώτη στήλη αντιστοιχεί στην

ταυτότητα του ιδρύματος και η δεύτερη στον αριθμό των ακμών που έρχονται από κόμβους τύπου ίδρυμα προς αυτό τον κόμβο τύπου άρθρο. Ο αριθμός του κόμβου τύπου άρθρο είναι ίσος με την τιμή που επιστρέφει η μέθοδος *GetId()* και ο αριθμός των ακμών που έρχονται από κόμβους τύπου ίδρυμα προς αυτό τον κόμβο τύπου άρθρο ισούται με την τιμή που επιστρέφει η μέθοδος *GetInDeg()*.

Έπειτα, αποθηκεύονται τα δεδομένα για την εξεύρεση της διάμεσης τιμής του αριθμού των άρθρων ανά ίδρυμα-οργανισμός. Διασχίζοντας το δίκτυο συνεργασίας, όταν υπάρχει κόμβος τύπου ίδρυμα αποθηκεύονται σε ένα αρχείο κειμένου δυο στήλες από ακέραιους αριθμούς. Η πρώτη στήλη αντιστοιχεί στην ταυτότητα του ιδρύματος και η δεύτερη στον αριθμό των ακμών που φεύγουν από κόμβους τύπου ίδρυμα προς κόμβο τύπου άρθρο. Ο αριθμός του κόμβου τύπου ίδρυμα είναι ίσος με την τιμή που επιστρέφει η μέθοδος *GetId()* και ο αριθμός των ακμών που φεύγουν από κόμβους τύπου ίδρυμα προς κόμβο τύπου άρθρο ισούται με την τιμή που επιστρέφει η μέθοδος *GetOutDeg()*.

Τέλος, υπολογίζεται το degree κάθε κόμβου τύπου ίδρυμα στο γράφο συνεργασίας ιδρυμάτων. Συγκεκριμένα, για τον υπολογισμό του degree κάθε κόμβου τύπου ίδρυμα στο γράφο συνεργασίας ιδρυμάτων, δημιουργείται ένα αρχείο κειμένου τυπώνονται στο οποίο τυπώνονται δύο στήλες από ακέραιους αριθμούς. Η πρώτη στήλη αντιστοιχεί στην ταυτότητα του ιδρύματος και η δεύτερη αντιστοιχεί στον αριθμό ακμών που εξέρχονται από τον κόμβο τύπου ίδρυμα (degree). Η δεύτερη τιμή ανακτάται από τη κλήση της μεθόδου *GetOutDeg()*.

Στη συνέχεια, για την εξεύρεση της διάμεσης τιμής του αριθμού άρθρων για κάθε ίδρυμα-οργανισμό, του αριθμού ιδρυμάτων-οργανισμών για κάθε άρθρο και του degree κάθε κόμβου τύπου ίδρυμα, τα δεδομένα από τα αντίστοιχα αρχεία κειμένου επεξεργάζονται με τη χρήση του προγράμματος Microsoft Office Excel 2007 και συγκεκριμένα με τη χρήση της συνάρτησης Median.

5.4 Αποτελέσματα εξόρυξης γενικών στατιστικών δεδομένων

Μέσω ενός PHP υπολογίστηκαν τα γενικά στατιστικά δεδομένα και τα αποτελέσματα δίνονται από τον πίνακα 5.1.

Αριθμός τμημάτων προς μελέτη	5
Αριθμός όλων των επιστημόνων	19522
Αριθμός των επιστημόνων που ανήκουν στο Πανεπιστήμιο Κύπρου	762
Αριθμός των επιστημόνων που ανήκουν στα τμήματα προς μελέτη	416
Αριθμός των επιστημόνων του Πανεπιστήμιο Κύπρου που ανήκουν στο τμήμα Βιολογίας	32
Αριθμός των επιστημόνων του Πανεπιστήμιο Κύπρου που ανήκουν στο τμήμα Χημείας	89
Αριθμός των επιστημόνων του Πανεπιστήμιο Κύπρου που ανήκουν στο τμήμα Πληροφορικής	180
Αριθμός των επιστημόνων του Πανεπιστήμιο Κύπρου που ανήκουν στο τμήμα Μαθηματικών και Στατιστικής	50
Αριθμός των επιστημόνων του Πανεπιστήμιο Κύπρου που ανήκουν στο τμήμα Φυσικής	65
Αριθμός συνολικών δημοσιεύσεων	3219
Αριθμός συνολικών δημοσιεύσεων τμήματος Βιολογίας	114
Αριθμός συνολικών δημοσιεύσεων τμήματος Χημείας	487
Αριθμός συνολικών δημοσιεύσεων τμήματος Πληροφορικής	777
Αριθμός συνολικών δημοσιεύσεων τμήματος Μαθηματικών και Στατιστικής	386
Αριθμός συνολικών δημοσιεύσεων τμήματος Φυσικής	367
Αριθμός ιδρυμάτων	1872
Αριθμός ακμών μεταξύ κόμβου τύπου άρθρο και τύπου συγγραφέας	5278
Αριθμός ακμών μεταξύ κόμβου τύπου άρθρο και τύπου ίδρυμα	3313

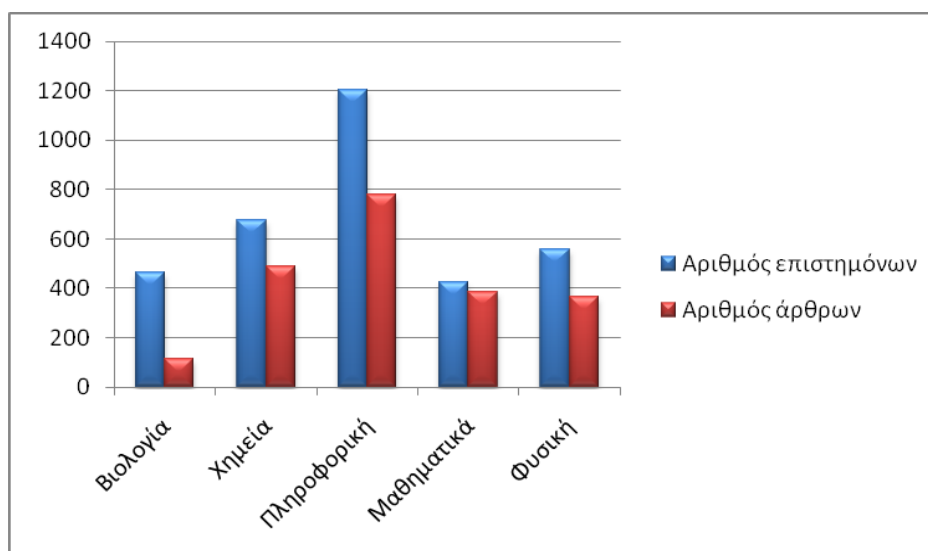
Πίνακας 5.1: Γενικά στατιστικά δεδομένα

5.5 Αποτελέσματα επιστημονικού δικτύου συνεργασίας

Χρησιμοποιώντας τη βιβλιοθήκη SNAP δημιουργήθηκε πρόγραμμα μέσω του οποίου αναλύθηκαν τόσο τα δίκτυα συνεργασίας επιστημόνων συγκεκριμένου τμήματος για όλες τις χρονιές δημοσίευσης άρθρων όσο και τα δίκτυα συνεργασίας επιστημόνων συγκεκριμένου τμήματος και συγκεκριμένης χρονιάς δημοσίευσης άρθρων.

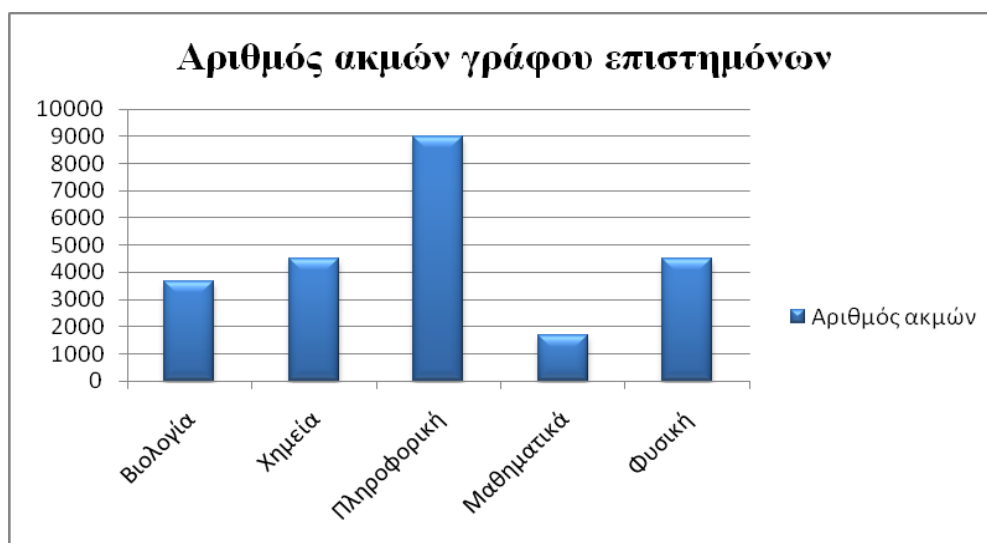
Επιλέγοντας την ανάλυση δικτύου συνεργασίας επιστημόνων συγκεκριμένου τμήματος για όλες τις χρονιές δημοσίευσης άρθρων(1992-2009), τα αποτελέσματα που προέκυψαν παρουσιάζονται με τις εξής γραφικές παραστάσεις. Αναλυτικά τα αποτελέσματα παρουσιάζονται στο παράρτημα Α.

❖ Συνολικός αριθμός επιστημόνων και άρθρων ανά τμήμα



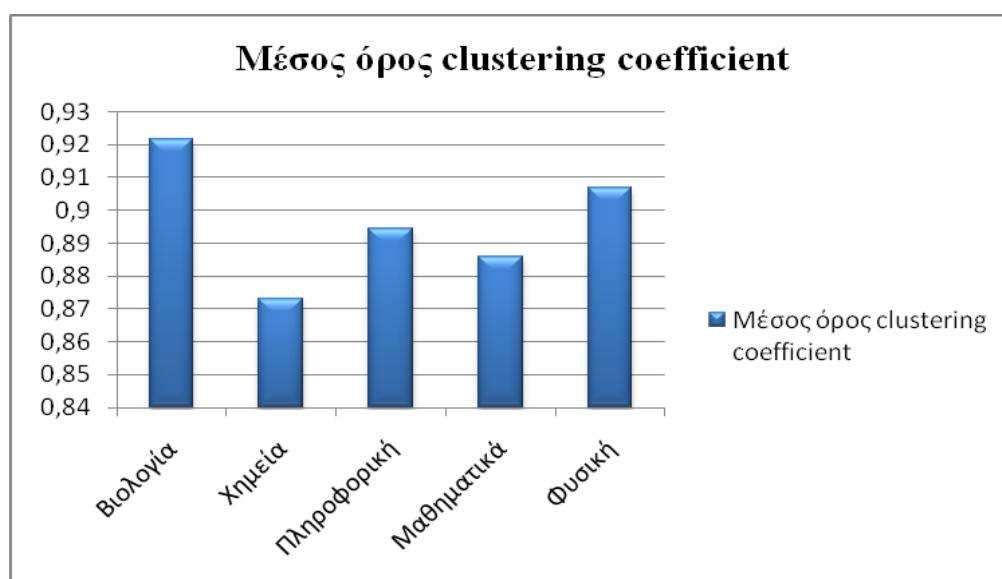
Παρατηρούμε πως από το 1992 έως το 2009 το τμήμα με τους λιγότερους επιστήμονες είναι το τμήμα Μαθηματικών και Στατιστικής με 424 επιστήμονες και το τμήμα με τους περισσότερους επιστήμονες είναι το τμήμα Πληροφορικής με 1205 επιστήμονες. Άξιο σχολιασμού είναι το γεγονός ότι ο αριθμός των επιστημόνων του τμήματος Πληροφορικής είναι κατά πολύ μεγαλύτερος και από το δεύτερο μεγαλύτερο σε αριθμό επιστημόνων τμήμα, αυτό της Χημείας. Επιπλέον, παρατηρούμε πως από το 1992-2009 το τμήμα με τα λιγότερα δημοσιευμένα άρθρα είναι το τμήμα Βιολογίας με 114 άρθρα και το τμήμα με τα περισσότερα δημοσιευμένα άρθρα είναι το τμήμα Πληροφορικής με 777 άρθρα.

❖ **Αριθμός ακμών επιστημονικού γράφου ανά τμήμα**



Παρατηρούμε πως από το 1992 έως το 2009 ο γράφος με τις λιγότερες ακμές είναι αυτός του τμήματος Μαθηματικών και Στατιστικής με 1671 ακμές και ο γράφος με τις περισσότερες ακμές είναι αυτός του τμήματος Πληροφορικής με 8983 ακμές.

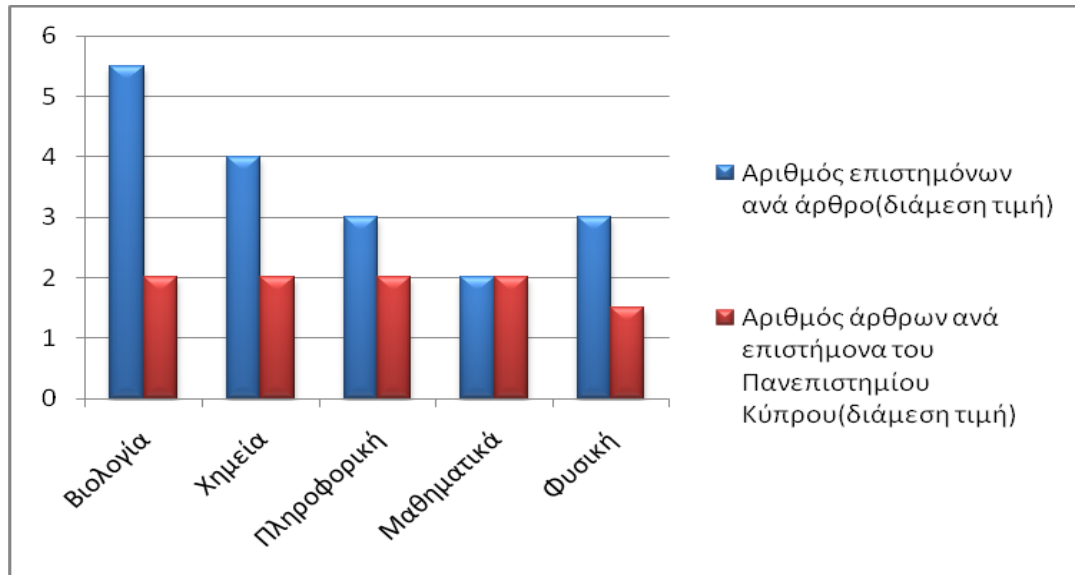
❖ **Μέσος όρος clustering coefficient**



Παρατηρούμε πως από το 1992 έως το 2009 ο γράφος με το μικρότερο μέσο όρο clustering coefficient είναι αυτός του τμήματος Χημείας ίσος με 0,8731 και ο γράφος με το μεγαλύτερο είναι αυτός του τμήματος Βιολογίας ίσος με 0,9215. Η διαφορά των μέσων όρων δεν είναι αρκετά μεγάλη και αυτό δηλώνει πως οι γράφοι και των πέντε

τμημάτων δημιουργούν πολλές ομάδες κόμβων που χαρακτηρίζονται σχετικά από μεγάλη πυκνότητα από δεσμούς.

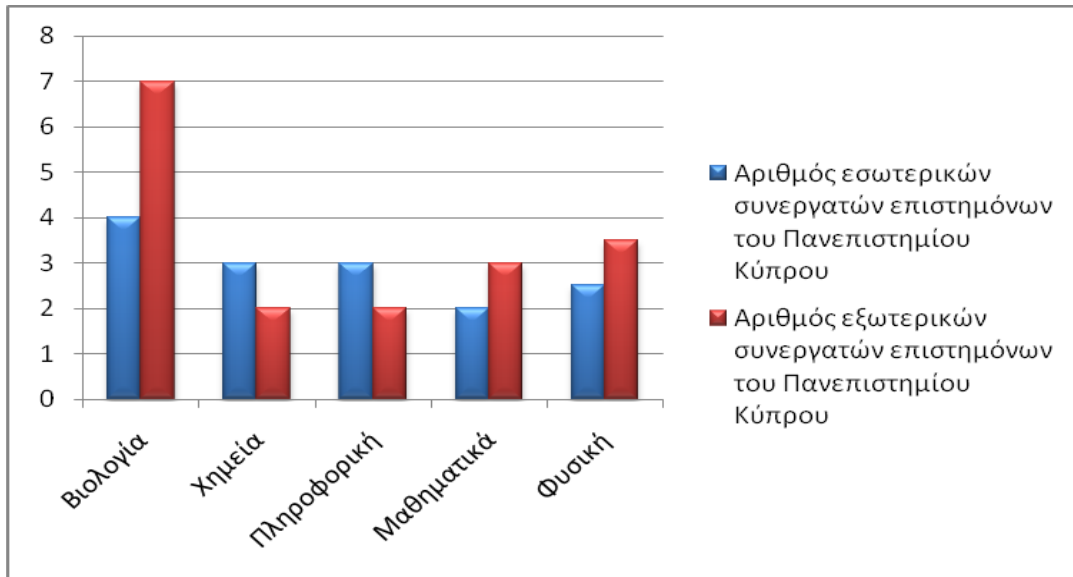
❖ **Αριθμός επιστημόνων ανά άρθρο και αριθμός άρθρων ανά επιστήμονα του Πανεπιστημίου Κύπρου**



Παρατηρούμε πως τα άρθρα που έχουν δημοσιευτεί από το 1992 έως το 2009 και αναφέρονται στο τμήμα Μαθηματικών και Στατιστικής τα έχουν συγγράψει συνήθως 2 επιστήμονες, στα οποία τουλάχιστον ο ένας ανήκει στο Πανεπιστήμιο Κύπρου. Ενώ, για τα άρθρα του τμήματος Βιολογίας έχουν συνεργαστεί συνήθως 5-6 επιστήμονες. Επιπλέον, με βάση τα αποτελέσματα, ο αριθμός των άρθρων που αντιστοιχεί για κάθε επιστήμονα του Πανεπιστημίου Κύπρου είναι ίσος με δύο και στα πέντε τμήματα. Κανείς θα αναρωτηθεί μα πως αυτό να είναι αλήθεια αφού οι καθηγητές που ανήκουν σε αυτά τα τμήματα συνήθως γράφουν δύο άρθρα το χρόνο και όχι δύο από το 1992 έως το 2009. Απλά σε αυτά τα αποτελέσματα έπαιξε σημαντικό ρόλο το γεγονός ότι επιστήμονες του Πανεπιστημίου Κύπρου θεωρούνται και οι τέσσερις βαθμίδες ακαδημαϊκού προσωπικού του Πανεπιστημίου Κύπρου συμπεριλαμβανομένου προπτυχιακών και μεταπτυχιακών φοιτητές. Το αποτέλεσμα επηρέασε αρκετά το γεγονός ότι για παράδειγμα προπτυχιακοί ή μεταπτυχιακοί φοιτητές συνεργάστηκαν με κάποιο καθηγητή του τμήματος τους για τη συγγραφή ενός ή δύο άρθρων. Έτσι υπήρξε περισσότερες φορές ο αριθμός ένα και δύο και συνεπώς η διάμεση τιμή συνήθως ήταν ίση με δύο. Αναφορικά, ο αριθμός των άρθρων από το 1992 έως το 2009 επιστημόνων

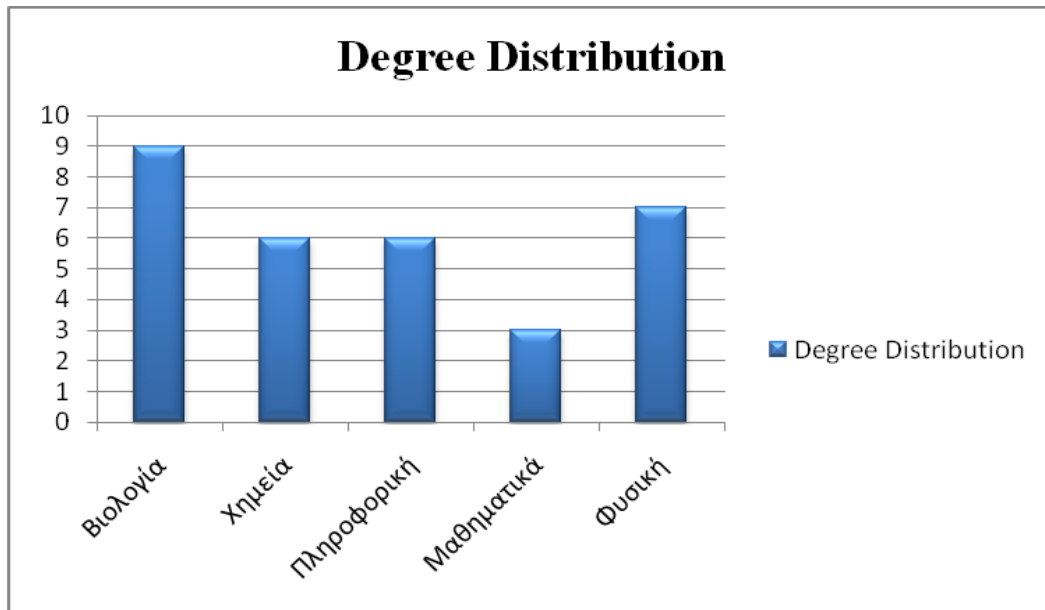
του τμήματος Βιολογίας κυμαίνεται από 1 έως 40, του τμήματος Χημείας κυμαίνεται από 1 έως 91, του τμήματος Πληροφορικής κυμαίνεται από 1 έως 114 του τμήματος Μαθηματικών και Στατιστικής κυμαίνεται από 1 έως 71 και του τμήματος Φυσικής κυμαίνεται από 1 έως 99.

❖ **Αριθμός εσωτερικών και εξωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου**



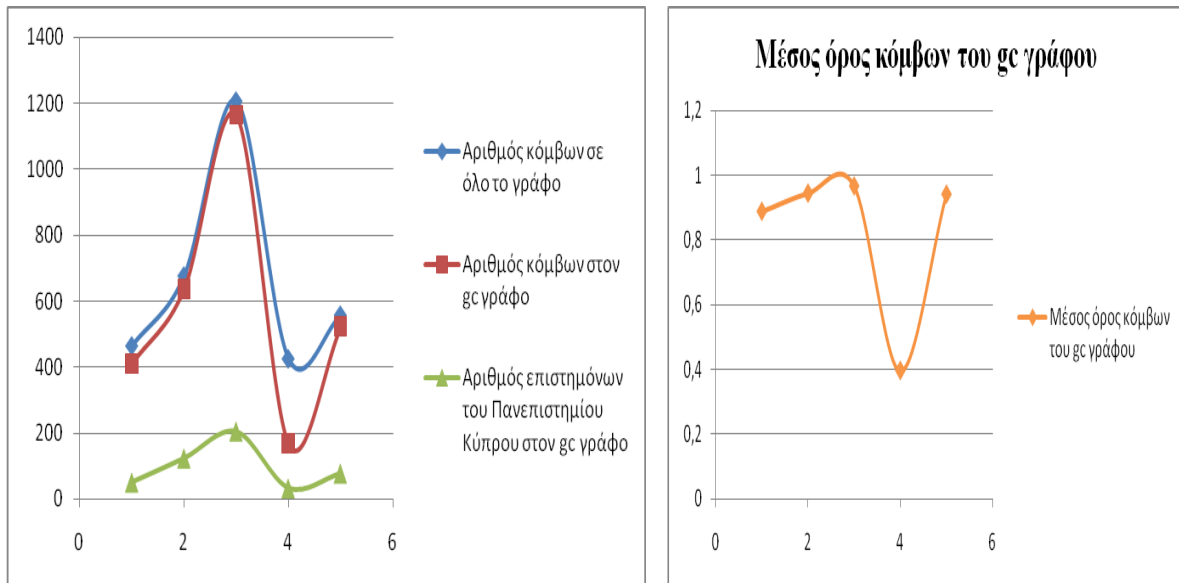
Παρατηρούμε πως οι επιστήμονες του τμήματος Βιολογίας για τη συγγραφή ενός άρθρου τείνουν να συνεργάζονται με πολλούς επιστήμονες και του Πανεπιστημίου Κύπρου και με επιστήμονες άλλων ιδρυμάτων οργανισμών. Αυτό το είχαμε παρατηρήσει και στην προηγούμενη γραφική παράσταση. Επίσης, παρατηρούμε πως οι επιστήμονες του τμήματος Χημείας και Πληροφορικής του Πανεπιστημίου Κύπρου έχουν περισσότερους εσωτερικούς συνεργάτες και οι επιστήμονες του τμήματος Φυσικής και Μαθηματικών και Στατιστικής του Πανεπιστημίου Κύπρου έχουν περισσότερους εξωτερικούς συνεργάτες.

❖ Degree Distribution



Από αυτή τη γραφική παράσταση μπορούμε να δούμε τον διάμεσο αριθμό συνδέσεων κόμβου τύπου επιστήμονα. Παρατηρούμε πως ο διάμεσος αριθμός συνδέσεων επιστημόνων του τμήματος Βιολογίας είναι ο ίσος με εννέα, του τμήματος Χημείας είναι ίσος με έξι, του τμήματος Πληροφορικής είναι ίσος με έξι, του τμήματος Μαθηματικών και Στατιστικής είναι ίσος με τρία και του τμήματος Φυσικής είναι ίσος με επτά. Σε κανένα γράφο από τα πέντε τμήματα, ο διάμεσος αριθμός συνδέσεων δεν υπερβαίνει το μέσο αριθμό συνδέσεων. Συνεπώς, και στις πέντε περιπτώσεις οι κατανομές είναι δεξιά ασύμμετρες, πράγμα που σημαίνει ότι η πλειοψηφία των κόμβων και στους πέντε γράφους έχουν low-degree.

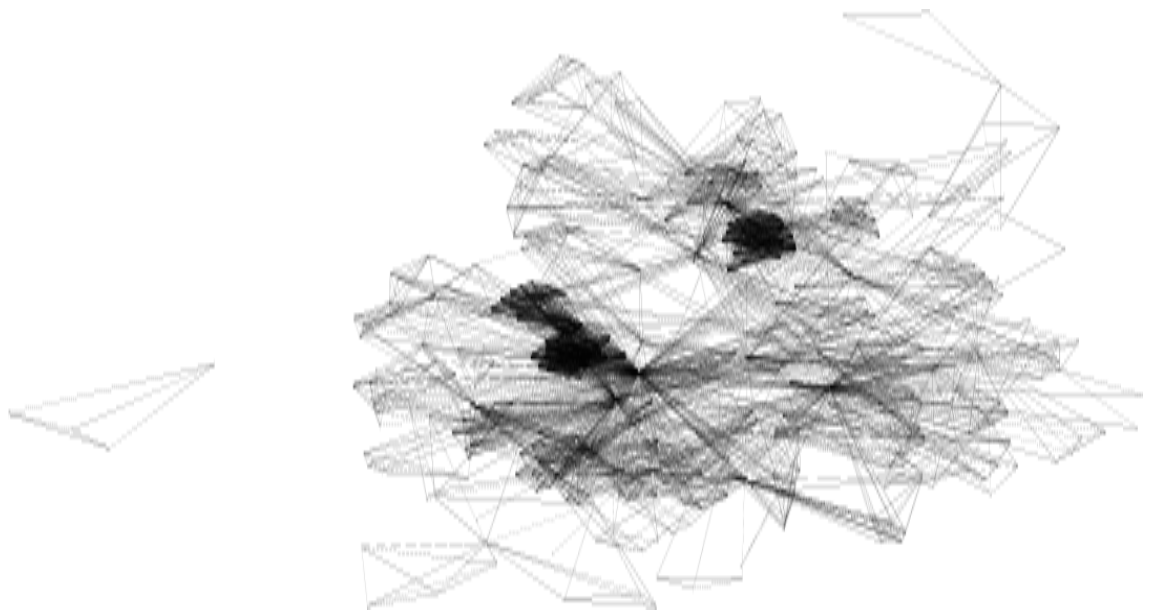
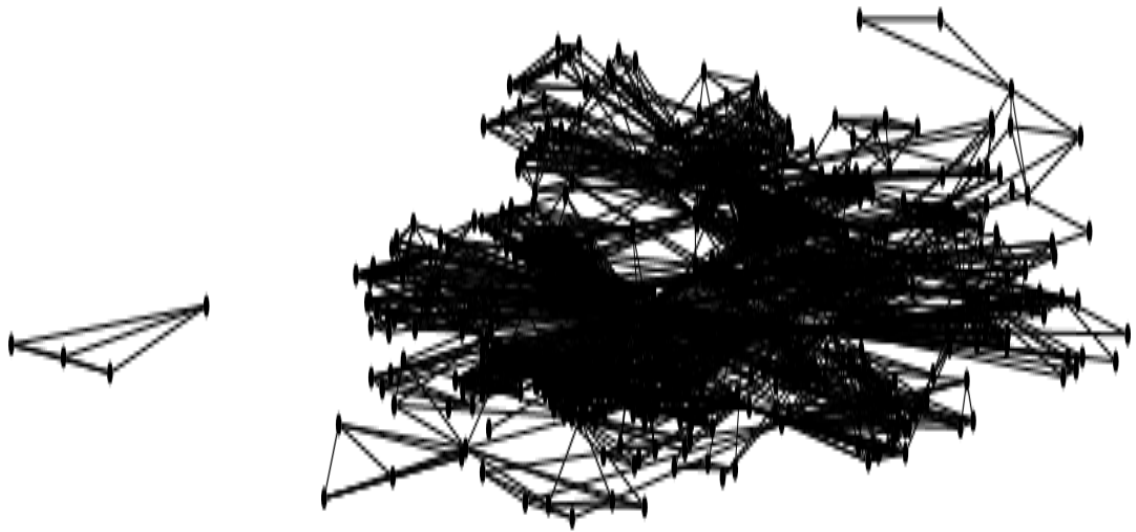
❖ Giant component (αριθμός κόμβων-μέσος όρος κόμβων)



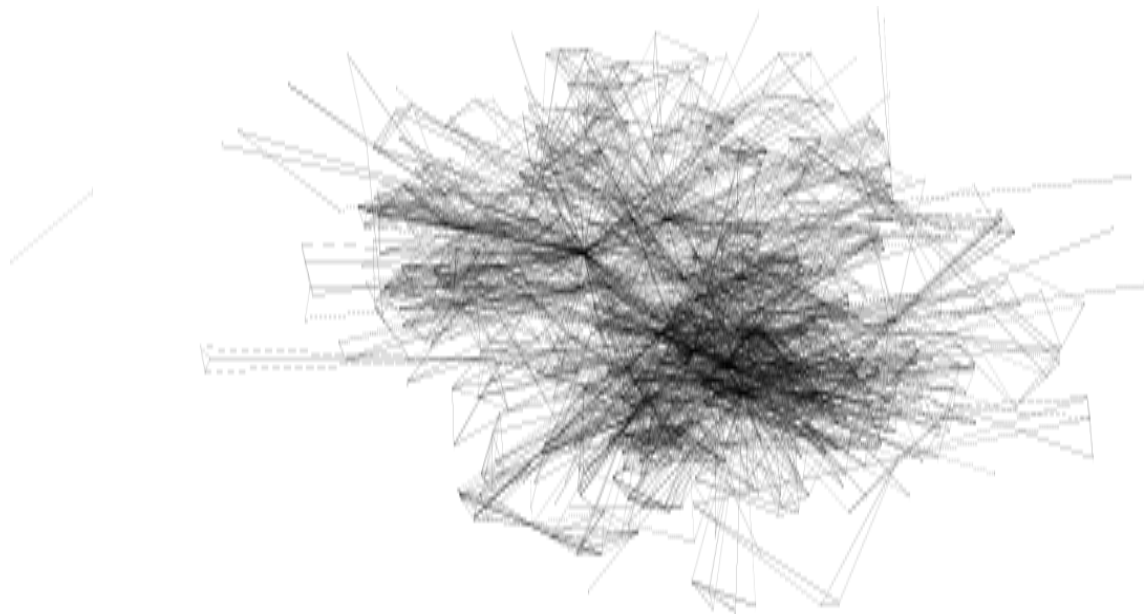
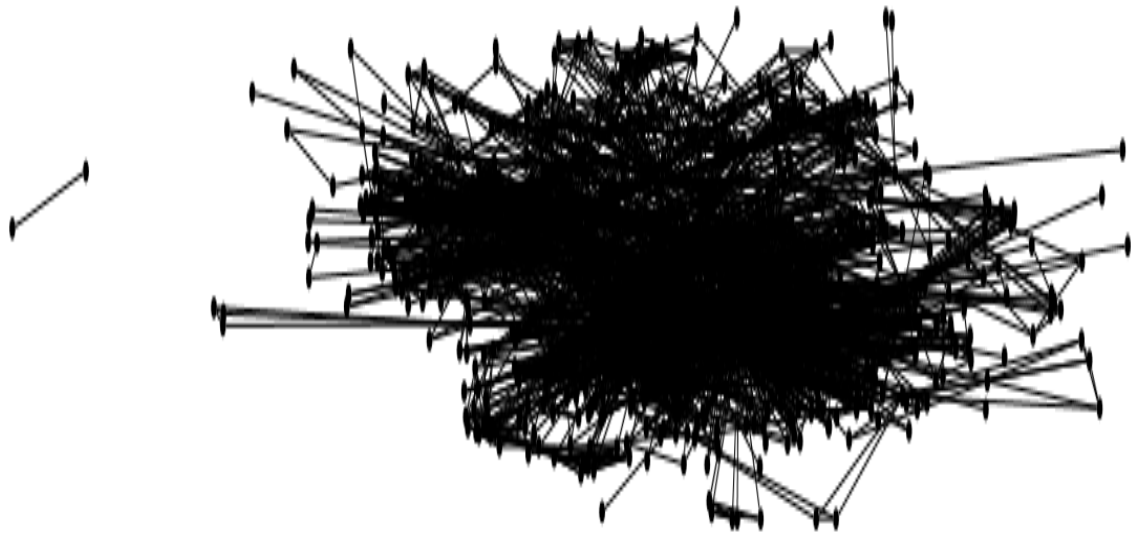
Από αυτές τις γραφικές παραστάσεις μπορούμε να δούμε τον αριθμό και το μέσο όρο των κόμβων που ανήκουν σε ένα υπογράφο(subgraph), ο οποίος περιέχει την πλειοψηφία των κόμβων ολόκληρου του γράφου. Οι κόμβοι που ανήκουν σε αυτόν τον υπογράφο(subgraph) είναι κόμβοι που παρατηρούνται σε μονοπάτια με το μικρότερο αριθμό 'ενδιάμεσων κόμβων' μεταξύ άλλων κόμβων. Παρατηρούμε πως στους τέσσερις από τους πέντε γράφους που απεικονίζουν τις συνεργασίες επιστημόνων το giant component καλύπτει σχεδόν ολόκληρο το γράφο, περισσότερο από 90%. Δηλαδή, περισσότερο από το 90% των επιστημόνων σε αυτούς τους γράφους είναι μέλη μιας ενιαίας κοινότητας, άρα απέχουν μερικά βήματα μακριά από οποιοδήποτε άλλο επιστήμονα αυτής της κοινότητας. Μόνο στο γράφο που απεικονίζει τις συνεργασίες των επιστημόνων του τμήματος Μαθηματικών και Στατιστικής του Πανεπιστημίου Κύπρου δεν παρατηρείται αυτό. Στη συγκεκριμένη περίπτωση, μόνο το 40% του ολικού γράφου καλύπτει το giant component.

Ακολουθούν οι γράφοι που απεικονίζουν τις συνεργασίες των επιστημόνων των πέντε αυτών τμημάτων του Πανεπιστημίου Κύπρου για την περίοδο 1992-2009. Ο δεύτερος γράφος κάθε σελίδας είναι ο ίδιος με τον πρώτο, αλλά απεικονίζεται σε μεγαλύτερη κλίμακα προκειμένου να ξεχωρίζουν που δημιουργούνται κοινότητες.

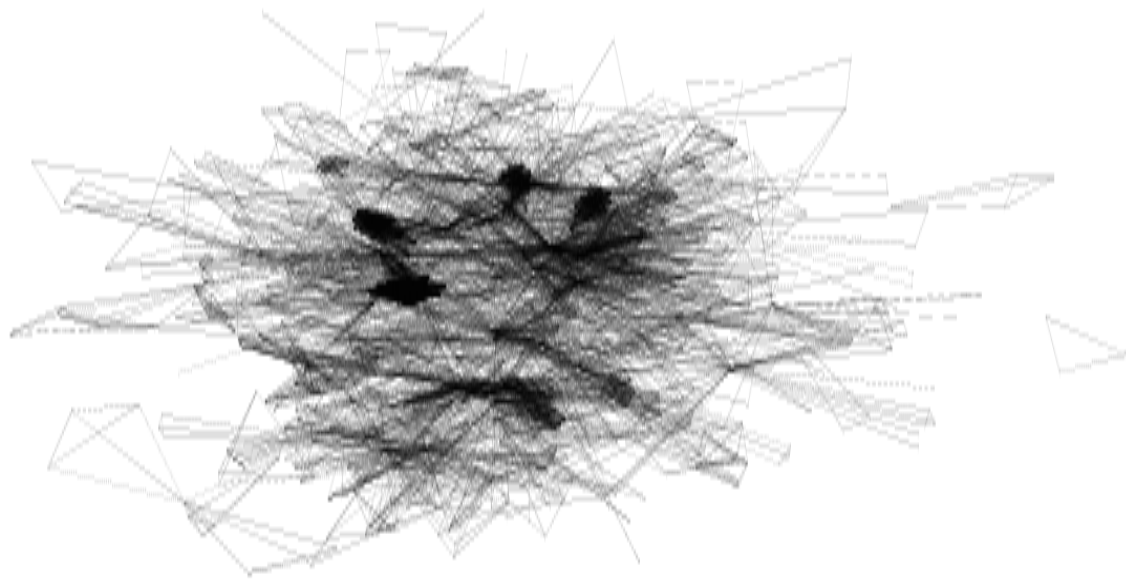
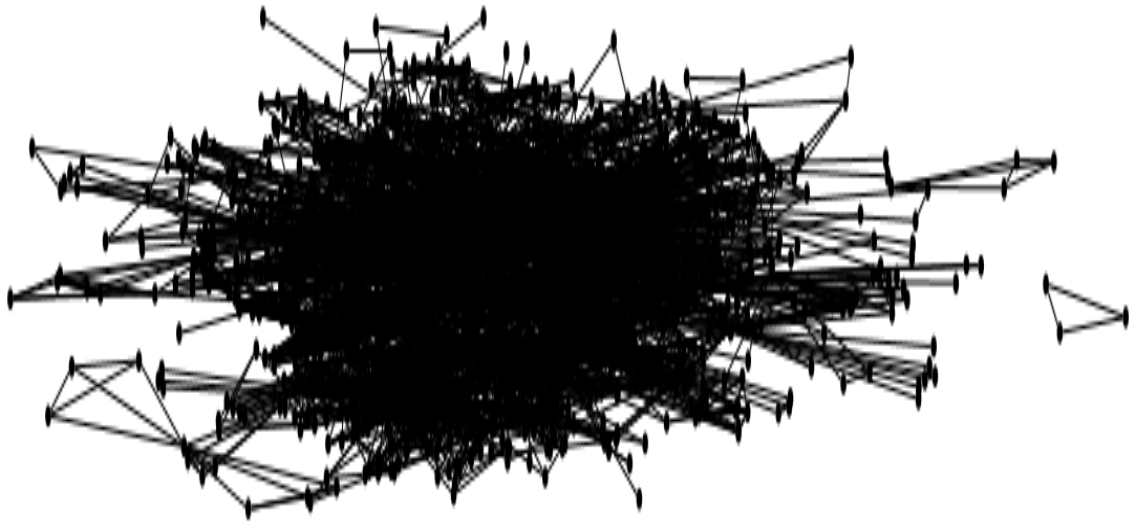
-Γράφος επιστημόνων για το τμήμα Βιολογίας



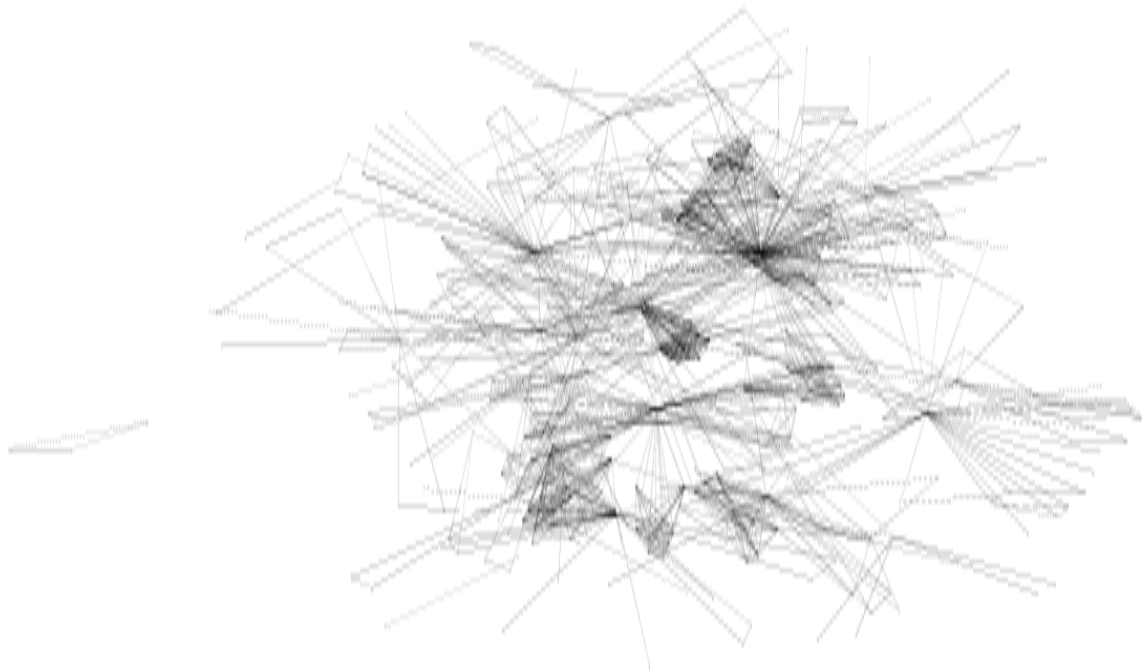
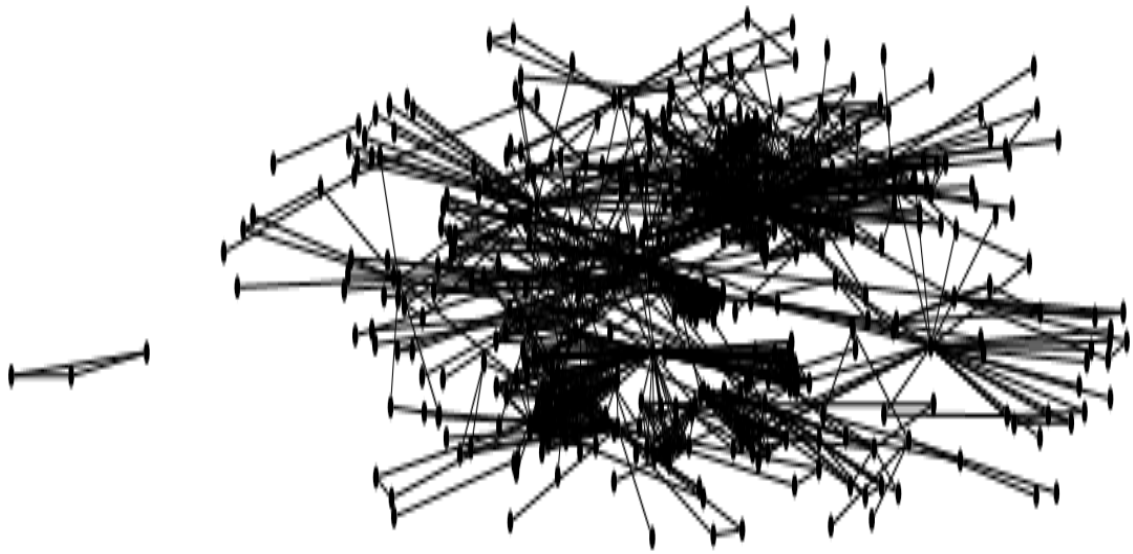
-Γράφος επιστημόνων για το τμήμα Χημείας



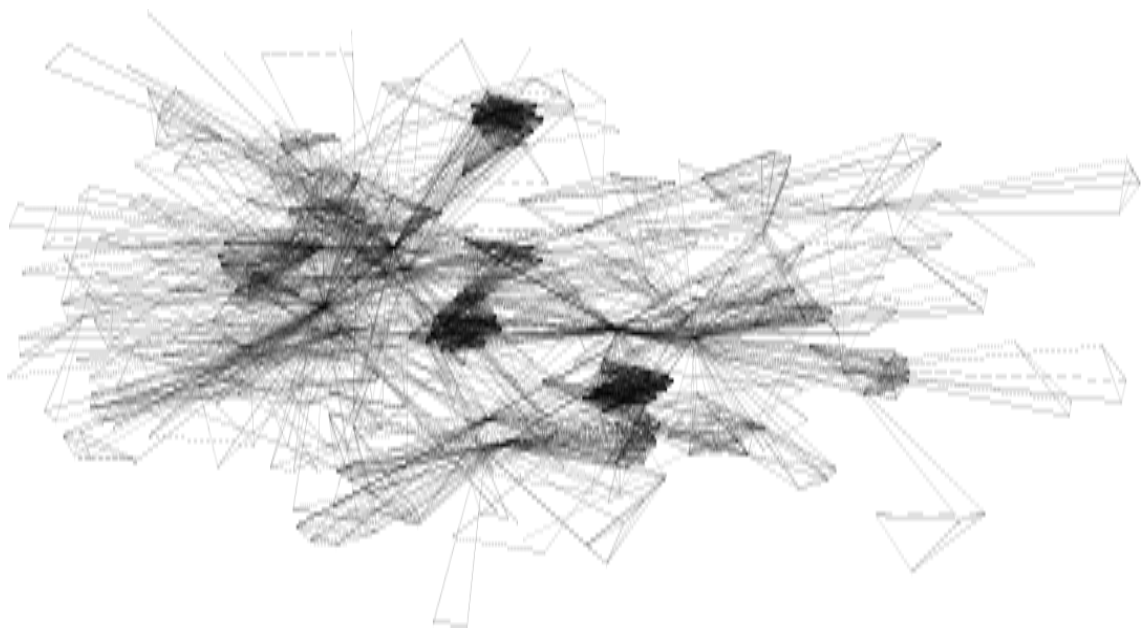
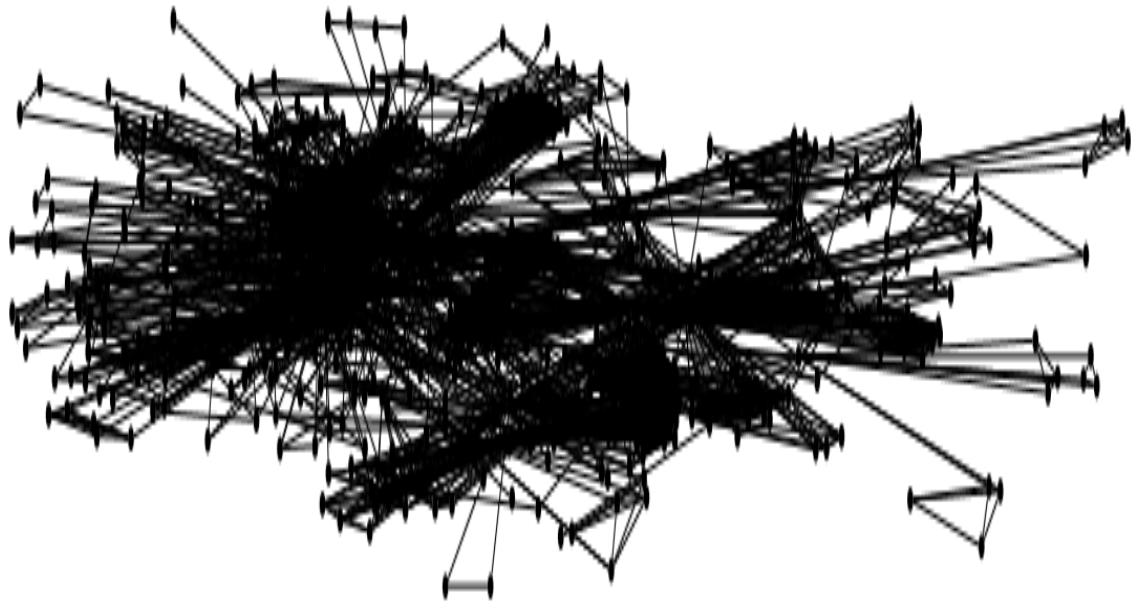
-Γράφος επιστημόνων για το τμήμα Πληροφορικής



-Γράφος επιστημόνων για το τμήμα Μαθηματικών και Στατιστικής



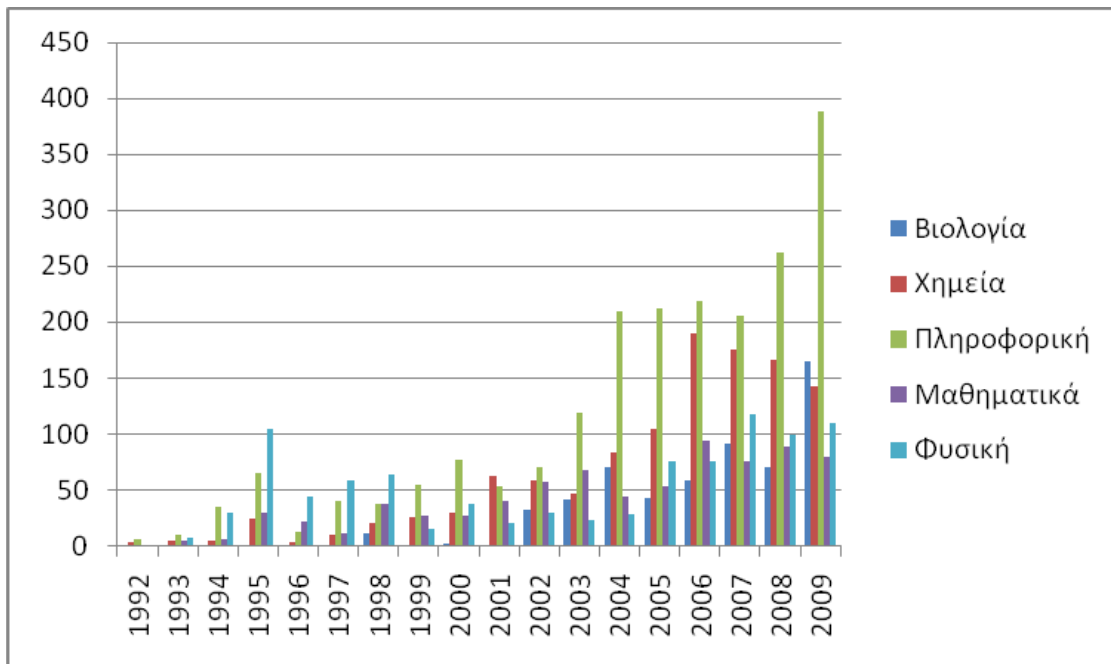
-Γράφος επιστημόνων για το τμήμα Φυσικής



Επιλέγοντας την ανάλυση δικτύου συνεργασίας επιστημόνων συγκεκριμένου τμήματος και συγκεκριμένης χρονιάς δημοσίευσης άρθρων(1992-2009), τα αποτελέσματα που προέκυψαν παρουσιάζονται με τις εξής γραφικές παραστάσεις.

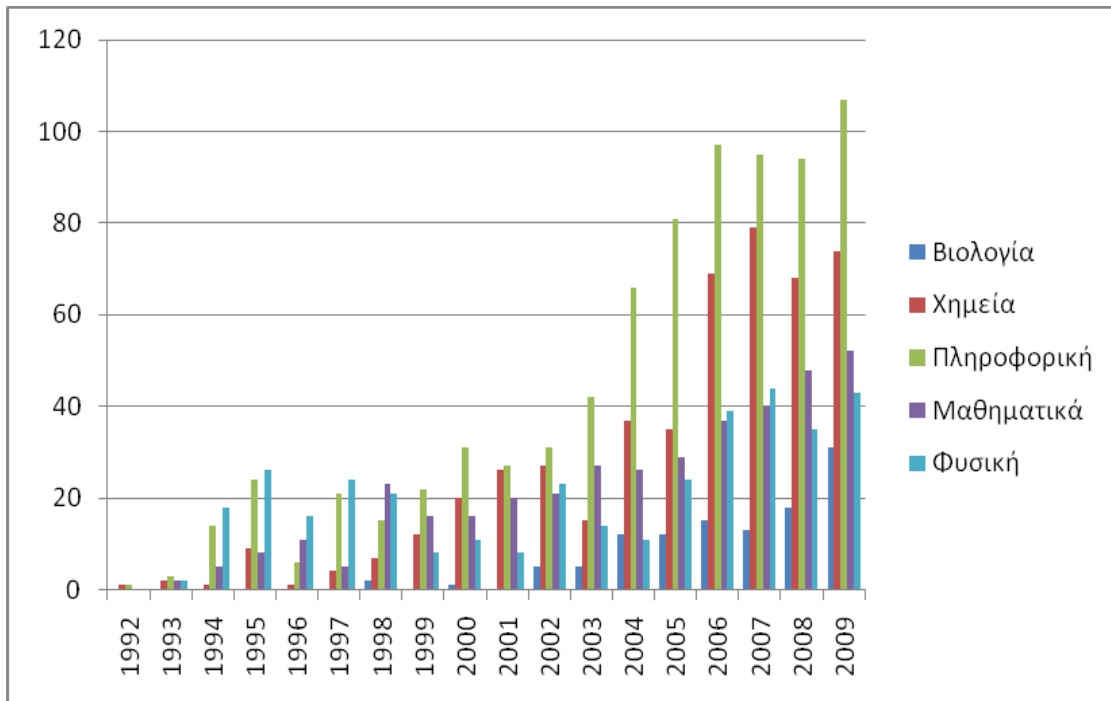
Αναλυτικά τα αποτελέσματα παρουσιάζονται στο παράρτημα Β.

❖ **Συνολικός αριθμός επιστημόνων ανά τμήμα και χρονιά**



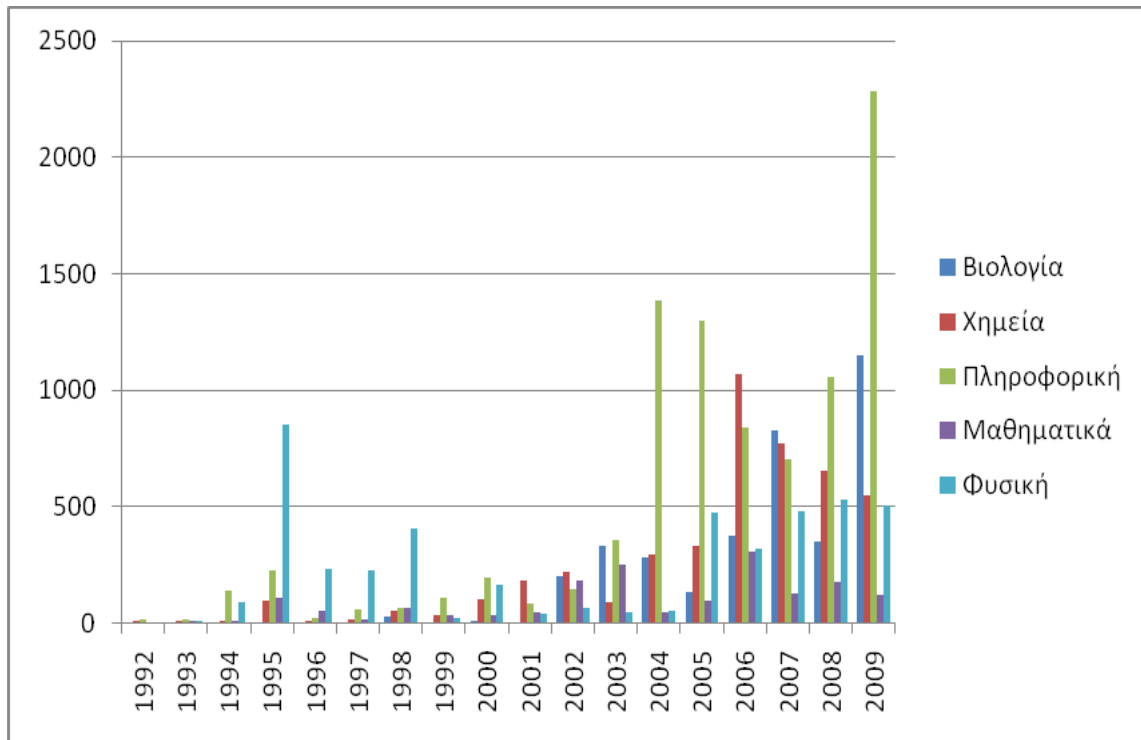
Παρατηρούμε ότι χρόνο με το χρόνο ο συνολικός αριθμός των επιστημόνων και των πέντε τμημάτων αυξάνεται. Το τμήμα της Πληροφορικής του Πανεπιστημίου Κύπρου όμως, είναι το τμήμα που ξεχωρίζει από τα υπόλοιπα τμήματα. Το 1992 έξι επιστήμονες της Πληροφορικής δημοσίευσαν άρθρο τους και φτάσαμε το 2009 να έχουν δημοσιεύσει άρθρα τους 389 επιστήμονες της Πληροφορικής. Έντονη δραστηριότητα και από τα πέντε τμήματα ξεκινά να παρουσιάζεται μετά την αλλαγή της χιλιετίας.

❖ **Συνολικός αριθμός άρθρων ανά τμήμα και χρονιά**



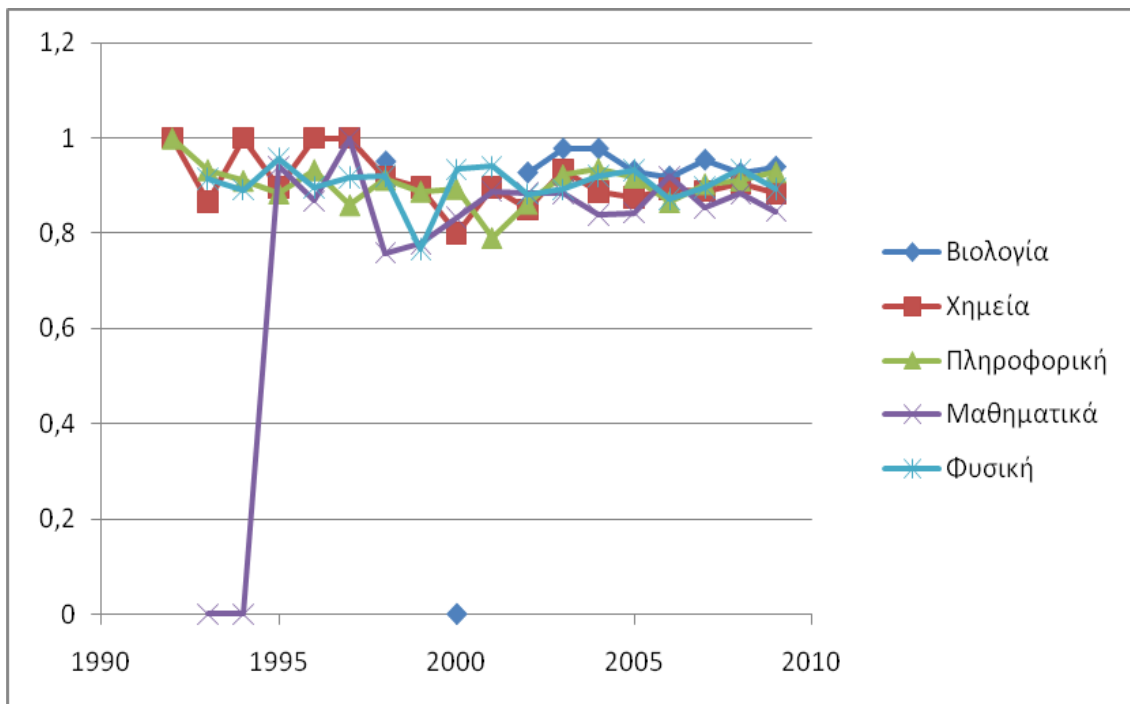
Παρατηρούμε ότι χρόνο με το χρόνο αυξάνεται ο συνολικός αριθμός των άρθρων και των πέντε τμημάτων. Και πάλι το τμήμα της Πληροφορικής του Πανεπιστημίου Κύπρου είναι το τμήμα που ξεχωρίζει από τα υπόλοιπα τμήματα γιατί προφανώς αφού χρόνο με χρόνο αυξάνονται οι επιστήμονές του, τότε θα δημοσιεύονται και περισσότερα άρθρα. Το 1992 δημοσιεύτηκε ένα άρθρο το οποίο αναφέρεται στο τμήμα της Πληροφορικής και το 2009 έχουν δημοσιευτεί 107 άρθρα. Έντονη δραστηριότητα παρατηρείται να έχει και το τμήμα Χημείας του Πανεπιστημίου Κύπρου. Το 1992 δημοσιεύτηκε ένα άρθρο το οποίο αναφέρεται στο τμήμα της Χημείας και το 2009 έχουν δημοσιευτεί 74 άρθρα. Εντούτοις, το τμήμα με τα λιγότερα δημοσιευμένα άρθρα είναι το τμήμα Βιολογίας του Πανεπιστημίου Κύπρου. Σε σύγκριση με τα αποτελέσματα του συνολικού αριθμού άρθρων των πέντε τμημάτων για όλες τις χρονιές, παρατηρούμε πως κάθε χρονιά, ειδικά από το 2000 και μετά που δραστηριοποιήθηκαν και τα πέντε τμήματα, σχεδόν πάντα τα περισσότερα άρθρα αναφέρονται στο τμήμα Πληροφορικής, έπειτα στο τμήμα Χημείας, στη συνέχεια στο τμήμα Μαθηματικών και Στατιστικής, ακολούθως στο τμήμα Φυσικής και τέλος στο τμήμα Βιολογίας.

❖ Αριθμός ακμών γράφου επιστημόνων ανά τμήμα και χρονιά



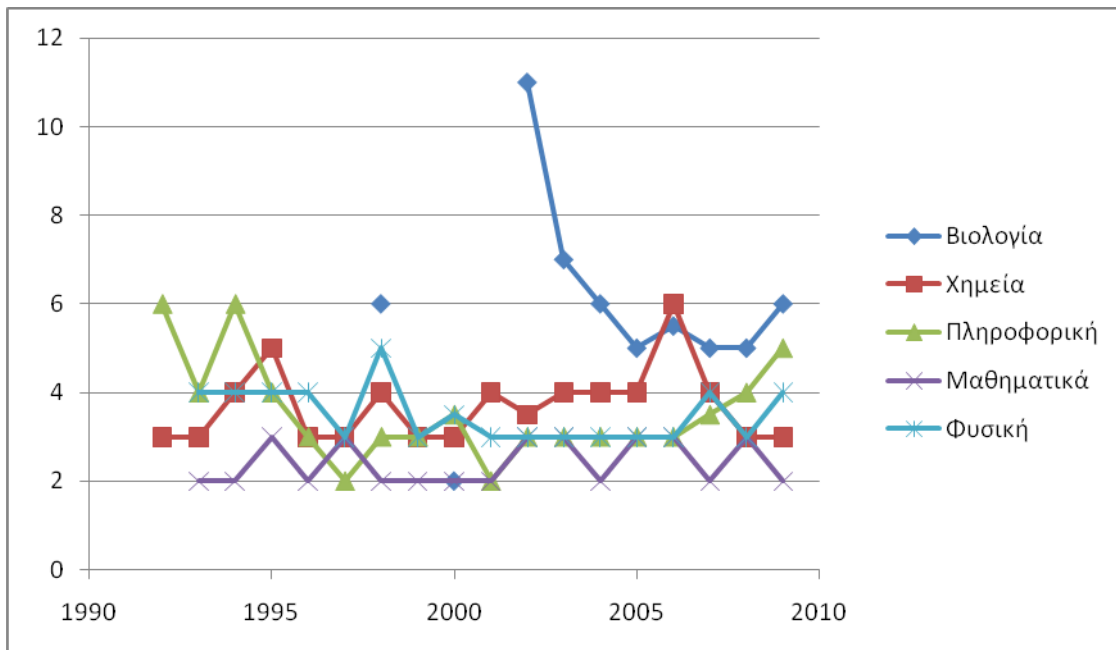
Από αυτή τη γραφική παράσταση μπορούμε να καταλάβουμε πόσες διαφορετικές ακμές δημιουργήθηκαν κάθε χρονιά στον αντίστοιχο γράφο που απεικονίζει τη συνεργασία των επιστημόνων συγκεκριμένου τμήματος. Παρατηρούμε πως από το 1994 μέχρι το 2000 το τμήμα Φυσικής είχε περισσότερες νέες συνεργασίες σε σχέση με τα υπόλοιπα τμήματα. Από το 2000 και μετά, ο αριθμός των ακμών των τμημάτων Πληροφορικής και Χημείας παρουσιάζει έντονες αλλαγές μέχρι που την τελευταία πενταετία ο αριθμός των ακμών των δύο αυτών τμημάτων είναι από τους πιο μεγάλους. Αυτό δείχνει πως αφού ο αριθμός των ακμών είναι μεγάλος, τότε νέες συνεργασίες είχαν γίνει με νέους επιστήμονες. Από την άλλη το τμήμα Μαθηματικών και Στατιστικής είναι το τμήμα με το λιγότερο σχεδόν πάντα αριθμό ακμών. Αυτό μας οδηγεί στο συμπέρασμα πως πιθανότατα ίδιοι επιστήμονες να συνεργάζονται με τους ίδιους επιστήμονες εντός και εκτός Πανεπιστημίου Κύπρου.

❖ Μέσος όρος clustering coefficient



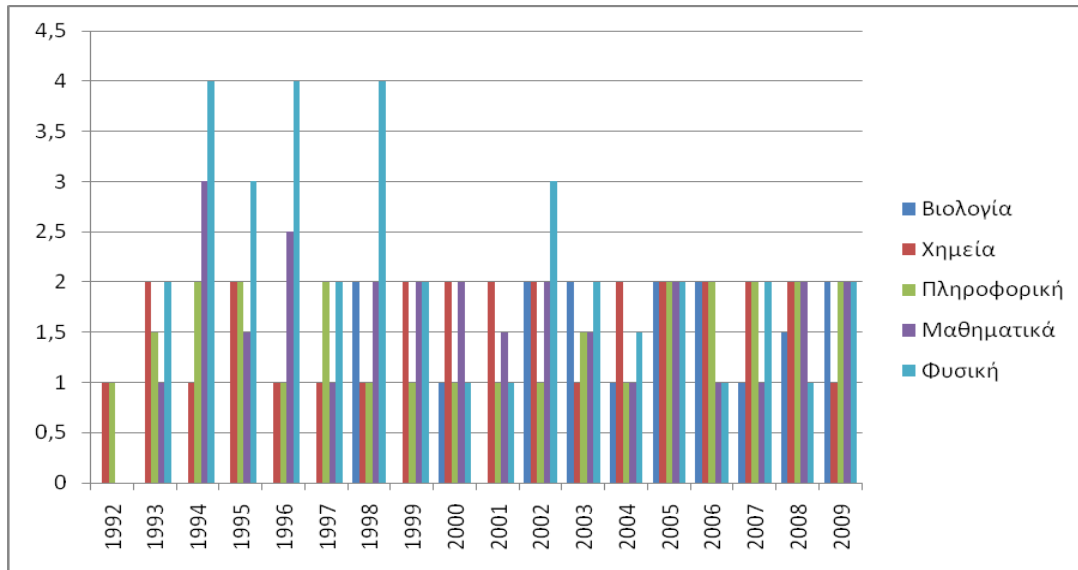
Από αυτή τη γραφική παράσταση μπορούμε να παρατηρήσουμε το βαθμό με τον οποίο οι κόμβοι του κάθε γράφου επιστημόνων τείνουν να γίνονται ένα σύμπλεγμα μαζί, να δημιουργούν έτσι κάποιες ομάδες οι οποίες χαρακτηρίζονται σχετικά από μεγάλη πυκνότητα από δεσμούς. Έτσι, μπορούμε να συμπεράνουμε πόσο πιθανό είναι δυο διαφορετικοί συνεργάτες κάποιου επιστήμονα, να είναι τελικά και μεταξύ τους συνεργάτες. Όσο πιο μεγάλη είναι αυτή η τιμή, τόσο αυξάνεται η πιθανότητα να έχουν συνεργαστεί οι επιστήμονες για τη συγγραφή και άλλου άρθρου. Στην προκειμένη περίπτωση, παρατηρούμε πως και στα πέντε τμήματα δύο επιστήμονες τυπικά έχουν περισσότερο από 80% πιθανότητα να συνεργαστούν αν έχουν ήδη συνεργαστεί με ένα τρίτο άτομο. Σίγουρα, αυτό οφείλεται και στο γεγονός ότι για τη συγγραφή ενός άρθρου συνεργάζονται περισσότεροι από δύο επιστήμονες. Συνεπώς, αυτό δείχνει πως οι επιστήμονες τείνουν να προτείνουν τους συνεργάτες τους σε άλλο επιστήμονα και αυτό έχει ως αποτέλεσμα τη δημιουργία νέων συνεργασιών. Παρατηρούμε πως υπάρχουν περιπτώσεις που ο μέσος όρος clustering coefficient είναι ίσος με μηδέν και αυτό οφείλεται στο γεγονός ότι δεν δημιουργήθηκε καμιά τριπλέττα ακμών, κανένα κλειστό τρίγωνο ακμών. Στο τμήμα Χημείας, μεταξύ του 1992 και του 1997, παρατηρείται πως τέσσερις φορές ο μέσος όρος clustering coefficient είναι ίσος με ένα αφού δημιουργούνται όλοι οι πιθανοί συνδυασμοί κλειστών τριγώνων ακμών. Αυτό οφείλεται στο γεγονός στο μικρό αριθμό επιστημόνων και δημοσιευμένων άρθρων.

❖ Αριθμός επιστημόνων ανά άρθρο



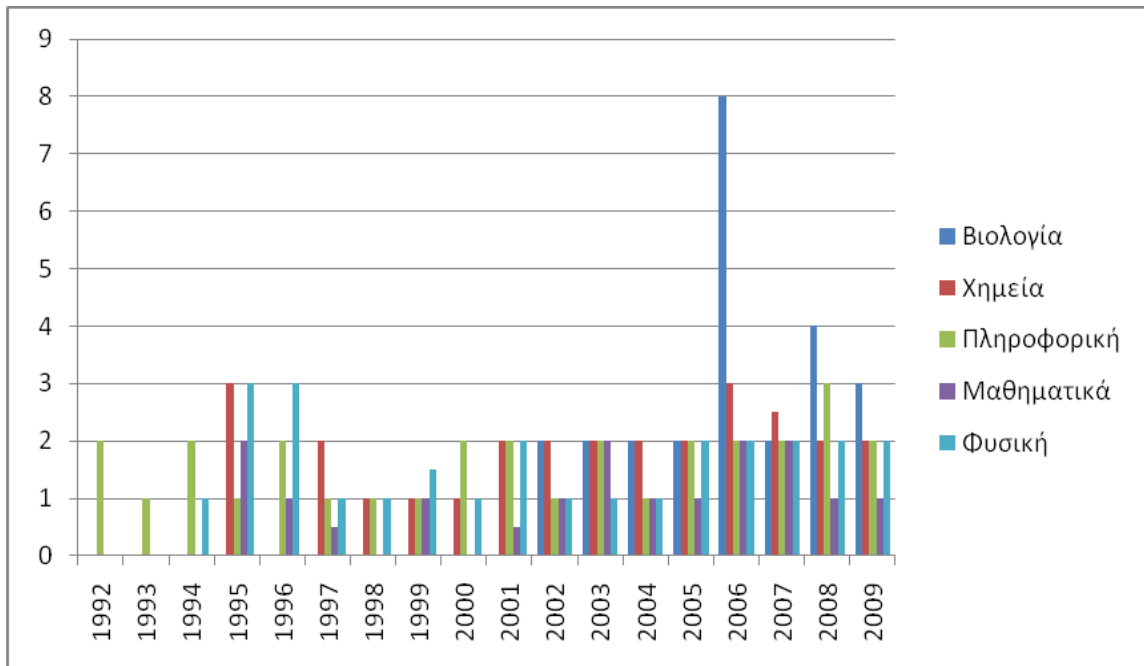
Από αυτή τη γραφική παράσταση μπορούμε να δούμε το διάμεσο αριθμό επιστημόνων ανά άρθρο. Βλέπουμε πως για τη συγγραφή άρθρων που αναφέρομαι στο τμήμα Βιολογίας και στο Χημείας συνεργάζονται περισσότεροι επιστήμονες από ότι στα άλλα τμήματα. Στο τμήμα Βιολογίας συνήθως συνεργάζονται 5-6 επιστήμονες και στο τμήμα Χημείας συνεργάζονται 3-4 επιστήμονες. Πιθανότατα αυτό οφείλεται στο γεγονός ότι τα άρθρα αυτά είναι περισσότερο πειραματικά. Για τα άρθρα του τμήματος Πληροφορικής και του τμήματος Φυσικής συνήθως συνεργάζονται τρεις επιστήμονες. Από την άλλη το τμήμα Μαθηματικών και Στατιστικής έχει τις μικρότερες συνεργασίες σε σύγκριση με τα υπόλοιπα τμήματα μιας και που τα άρθρα που αναφέρονται σε αυτό το τμήμα είναι περισσότερο θεωρητικά από αυτά των άλλων τμημάτων. Για τα άρθρα του τμήματος Μαθηματικών και Στατιστικής συνήθως συνεργάζονται μόνο δύο επιστήμονες τόσοι όσοι για να ικανοποιείται η αναγκαία συνθήκη συνεργασίας.

❖ Αριθμός άρθρων ανά επιστήμονα του Πανεπιστημίου Κύπρου



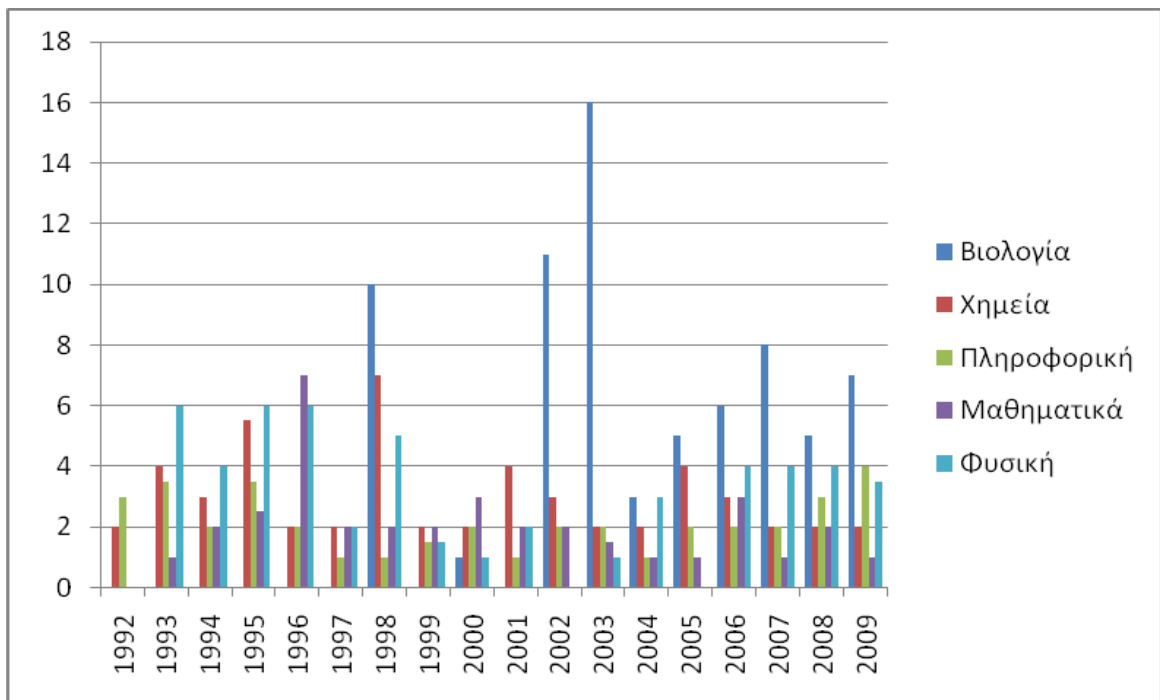
Από αυτή τη γραφική παράσταση μπορούμε να δούμε το διάμεσο αριθμό άρθρων που συγγράφει ένας επιστήμονας συγκεκριμένου τμήματος του Πανεπιστημίου Κύπρου με τουλάχιστον ακόμη έναν άλλο επιστήμονα. Παρατηρούμε πως ο διάμεσος αριθμός άρθρων κάθε τμήματος είναι ίσος με δύο. Άρα, κάθε επιστήμονας του Πανεπιστημίου Κύπρου ενός από τα πέντε τμήματα προς μελέτη, τείνει να συγγράφει δύο άρθρα το χρόνο.

❖ Αριθμός εσωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου



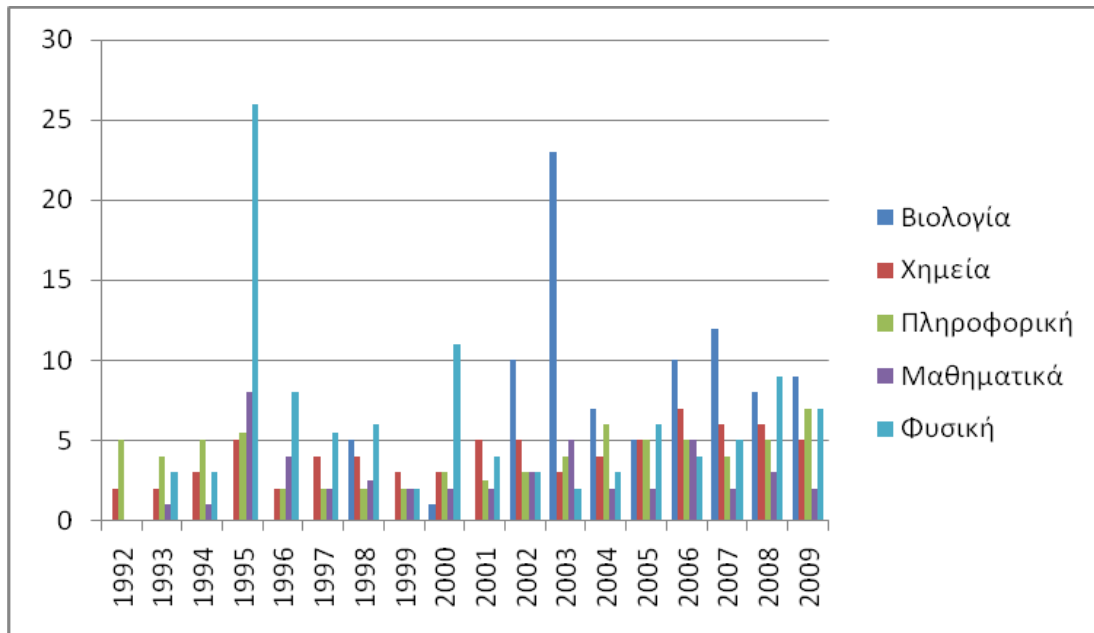
Από αυτή τη γραφική παράσταση μπορούμε να δούμε το διάμεσο αριθμό εσωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου. Παρατηρούμε πως τις πλείστες χρονιές ο διάμεσος αριθμός εσωτερικών επιστημόνων για τα τμήματα Βιολογίας, Χημείας και Πληροφορικής είναι ίσος με δύο ενώ για τα τμήματα Φυσικής και Μαθηματικών και Στατιστικής είναι ίσος με ένα.

❖ Αριθμός εξωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου



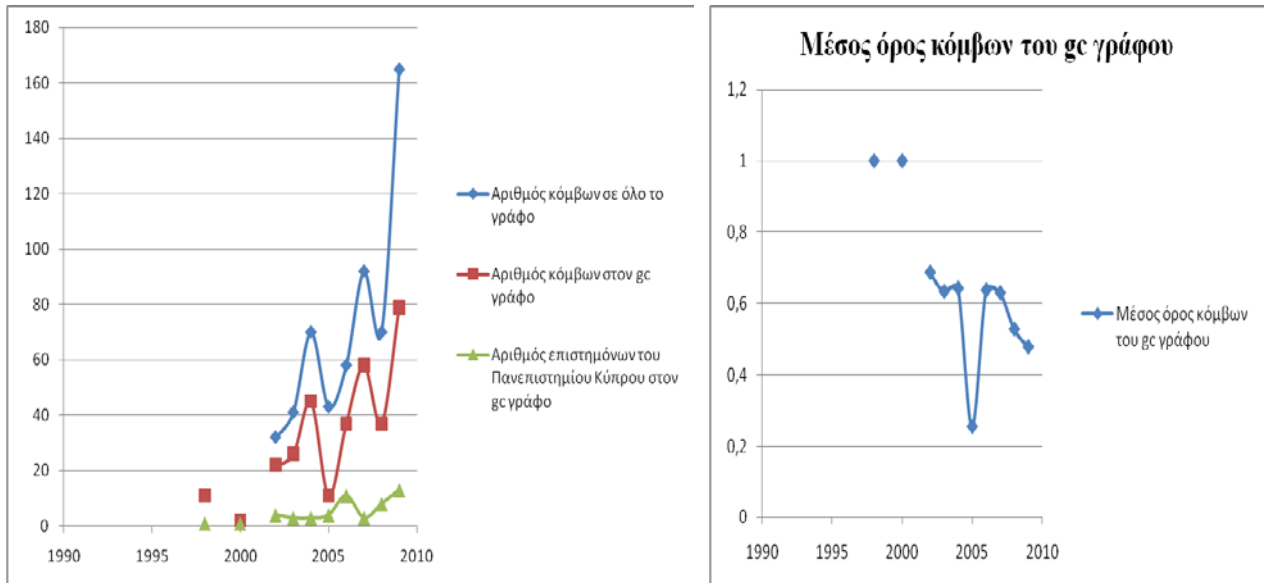
Από αυτή τη γραφική παράσταση μπορούμε να δούμε το διάμεσο αριθμό εξωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου. Παρατηρούμε πως τις πλείστες χρονιές ο διάμεσος αριθμός εξωτερικών επιστημόνων για τα τμήματα Χημείας, Πληροφορικής και Μαθηματικών και Στατιστικής είναι ίσος με δύο. Από την άλλη όμως, οι επιστήμονες του τμήματος Φυσικής του Πανεπιστημίου Κύπρου τείνουν να συνεργάζονται με 3-4 επιστήμονες άλλων ιδρυμάτων-οργανισμών. Άξιο αναφοράς είναι πως για το τμήμα Βιολογίας υπάρχουν χρονιές(1992-1997,1999,2001) που δεν έχουν δημοσιευτεί άρθρα που να αναφέρονται σε αυτό το τμήμα, αλλά τις χρο νές που υπάρχουν τέτοια άρθρα παρατηρείται πως οι επιστήμονες αυτού του τμήματος τείνουν να συνεργάζονται με έξι επιστήμονες άλλων ιδρυμάτων-οργανισμών.

❖ Degree Distribution



Με βάσει αυτή τη γραφική παράσταση μπορο ύμε να δούμε τον διάμεσο αριθμό συνδέσεων κόμβου τύπου επιστήμονα ανά τμήμα και χρονιά. Παρατηρούμε πως ο διάμεσος αριθμός συνδέσεων επιστημόνων και των πέντε τμημάτων κυμαίνεται από δύο έως επτά στις πλείστες περιπτώσεις. Ο αριθμός αυτός δεν είναι μεγάλος και δεν υπερβαίνει το μέσο αριθμό συνδέσεων. Συνεπώς, και στις πέντε περιπτώσεις οι κατανομές είναι δεξιά ασύμμετρες, πράγμα που σημαίνει ότι η πλειοψηφία των κόμβων και στους πέντε γράφους έχουν low-degree.

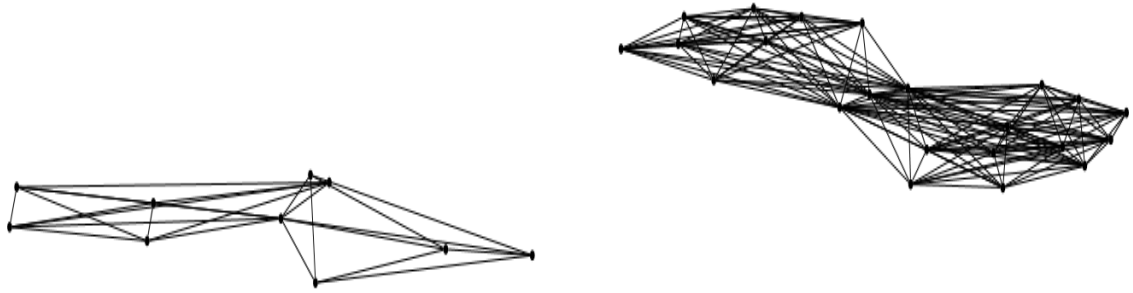
❖ **Giant component (αριθμός κόμβων-μέσος όρος κόμβων)**
τμήματος Βιολογίας



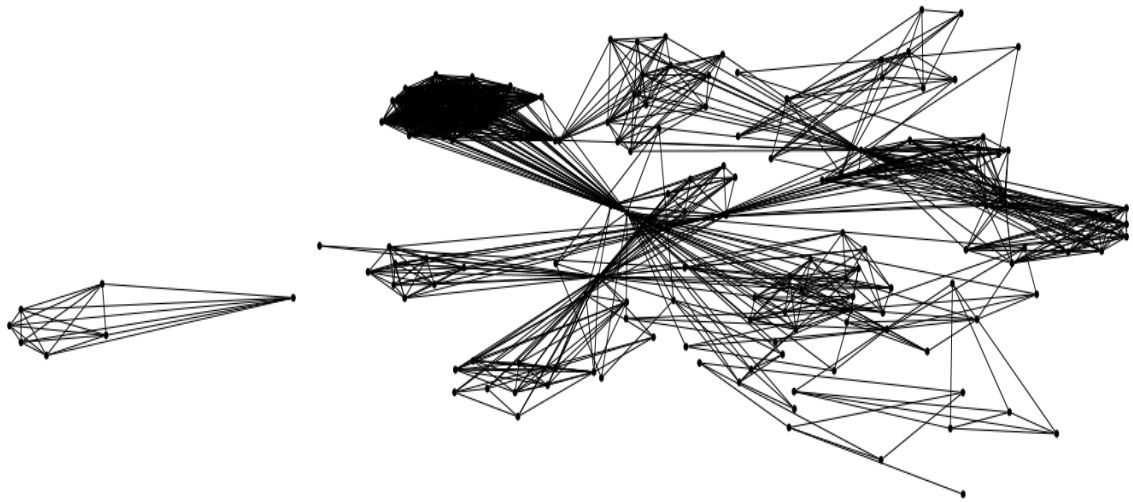
Από τις πιο πάνω γραφικές παραστάσεις μπορούμε να δούμε τον αριθμό και το μέσο όρο των κόμβων που ανήκουν σε ένα υπογράφο(subgraph), ο οποίος περιέχει την πλειοψηφία των κόμβων ολόκληρου του γράφου. Από το 2002, το τμήμα της Βιολογίας του Πανεπιστημίου Κύπρου άρχισε να δραστηριοποιείται. Όλες τις χρονιές εκτός από το 2005 το giant component καλύπτει περισσότερο από το 50%. Ο λόγος για τον οποίο σε αυτό το δίκτυο συνεργασίας δεν παρατηρείται κάποια ισορροπία είναι γιατί γενικά τα δίκτυα συνεργασίας συνεχώς εξελίσσονται. Καθώς περνάει ο καιρός, μπορεί κάποια μικρά components σταδιακά να συνδέονται με μεγαλύτερα, εντούτοις μπορεί να εμφανίζονται και νέοι επιστήμονες ή να υπάρχουν και πολλά ημι-απομονωμένα άτομα και ομάδες που ξεκινούν τις έρευνες τους ανεξάρτητα. Συνεπώς, οι επιστήμονες που ανήκουν στο giant component, μπορεί να θεωρηθεί ότι αποτελούν μέρος μιας κοινότητας, με την έννοια ότι όλοι απέχουν μερικά βήματα – ενδιάμεσες γνωριμίες μακριά από οποιοδήποτε άλλο επιστήμονα αυτής της κοινότητας.

Ακολουθούν οι γράφοι του 2002 και του 2009 για το τμήμα Βιολογίας του Πανεπιστημίου Κύπρου.

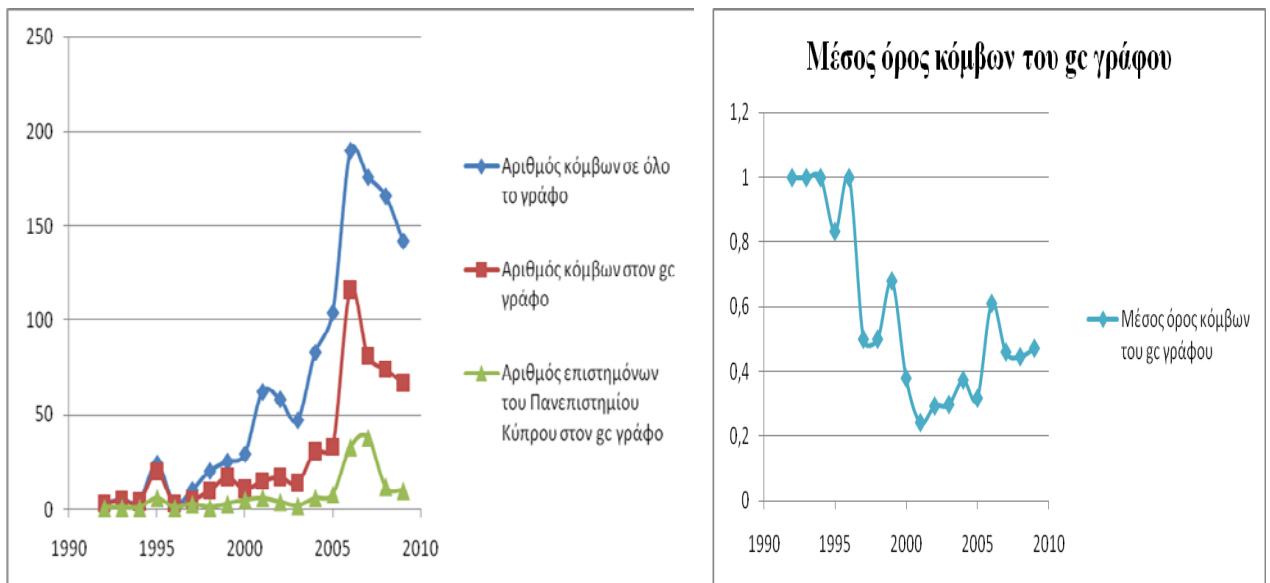
- Γράφος του 2002 για το τμήμα Βιολογίας



- Γράφος του 2009 για το τμήμα Βιολογίας



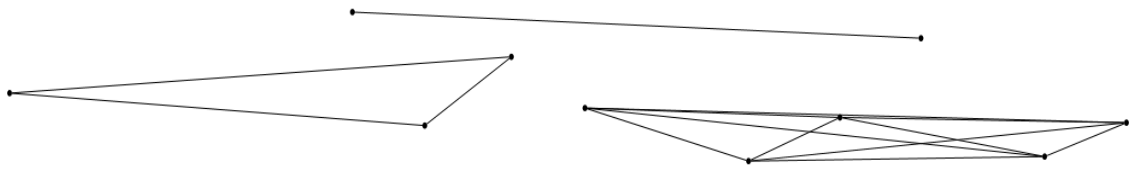
❖ **Giant component (αριθμός κόμβων-μέσος όρος κόμβων)**
τμήματος Χημείας



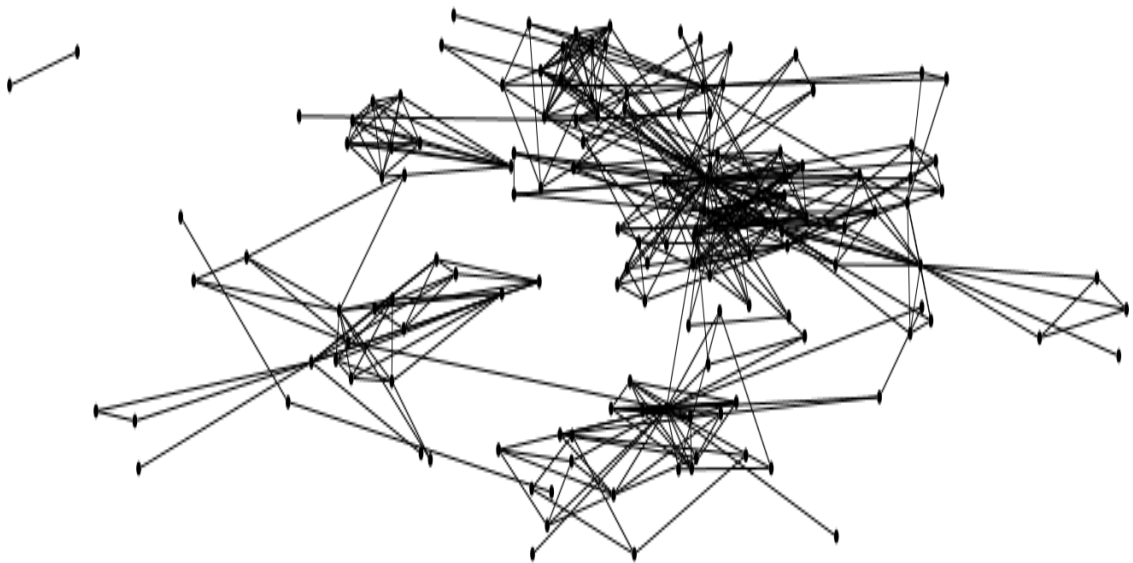
Από τις πιο πάνω γραφικές παραστάσεις μπορούμε να δούμε τον αριθμό και το μέσο όρο των κόμβων που ανήκουν σε ένα υπογράφο(subgraph), ο οποίος περιέχει την πλειοψηφία των κόμβων ολόκληρου του γράφου. Από το 1997, το τμήμα της Χημείας του Πανεπιστημίου Κύπρου άρχισε να δραστηριοποιείται πιο έντονα. Παρατηρείται πως δεν υπάρχει κάποια ισορροπία είναι γιατί γενικά τα δίκτυα συνεργασίας συνεχώς εξελίσσονται. Καθώς περνάει ο καιρός, μπορεί κάποια μικρά components σταδιακά να συνδέονται με μεγαλύτερα, εντούτοις μπορεί να εμφανίζονται και νέοι επιστήμονες ή να υπάρχουν και πολλά ημι-απομονωμένα άτομα και ομάδες που ξεκινούν τις έρευνες τους ανεξάρτητα. Τα τελευταία τρία χρόνια (2007-2009) που δημοσιεύτηκαν περισσότερα άρθρα και υπήρξαν περισσότερες συνεργασίες επιστημόνων, το 45% των επιστημόνων του ολικού γράφου αποτελούν μέρος μιας κοινότητας, με την έννοια ότι όλοι απέχουν μερικά βήματα – ενδιάμεσες γνωριμίες μακριά από οποιοδήποτε άλλο επιστήμονα αυτής της κοινότητας. Από το 1992 μέχρι 1996, παρατηρείται πως κάποιες χρονιές το giant component καλύπτει όλο το γράφο. Ο λόγος για τον οποίο συμβαίνει αυτό είναι γιατί υπήρξε δημοσίευση ενός ή δύο άρθρων μόνο.

Ακολουθούν οι γράφοι του 1997 και του 2009 για το τμήμα Χημείας του Πανεπιστημίου Κύπρου.

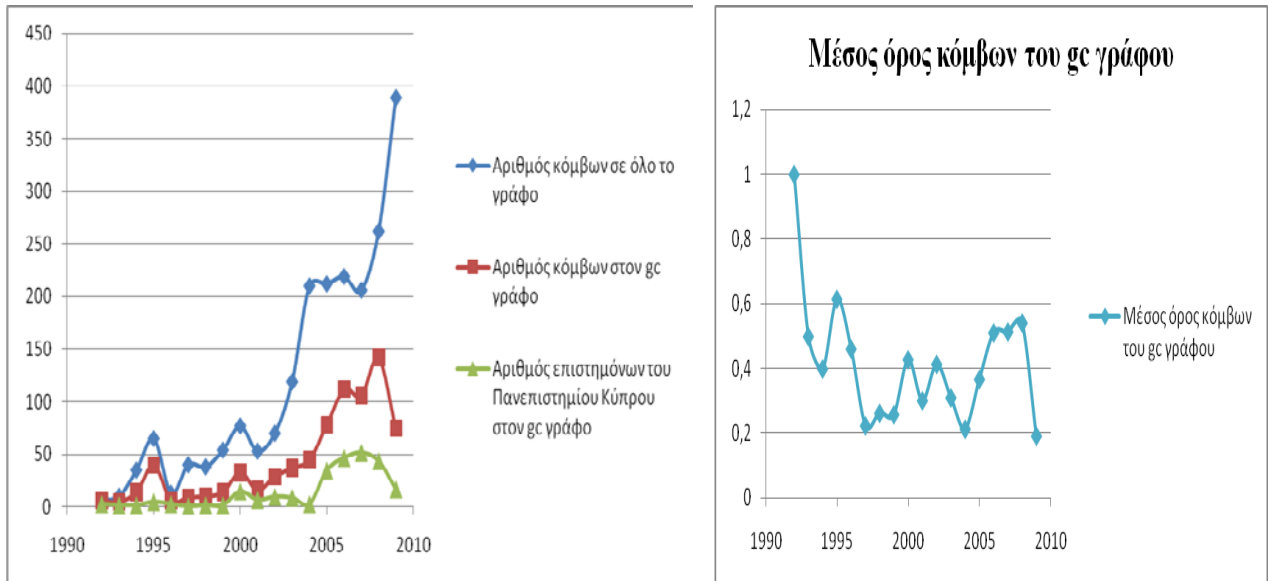
- Γράφος του 1997 για το τμήμα Χημείας



- Γράφος του 2009 για το τμήμα Χημείας



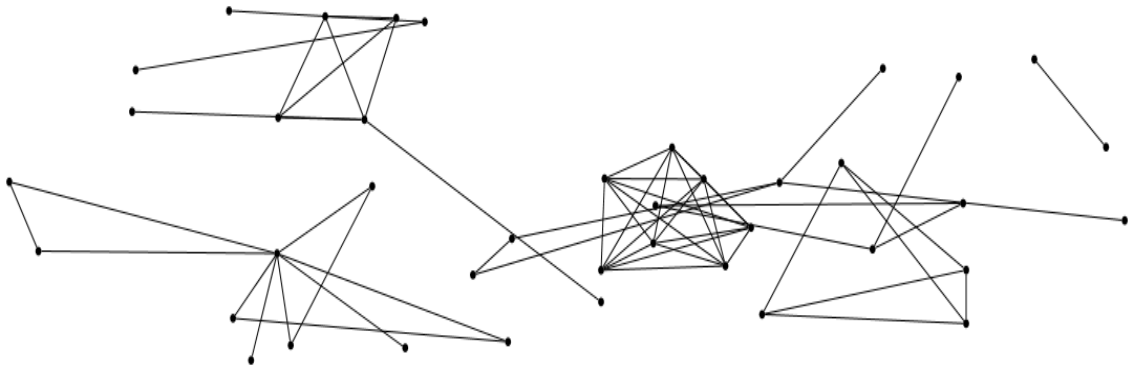
❖ **Giant component (αριθμός κόμβων-μέσος όρος κόμβων)**
τμήματος Πληροφορικής



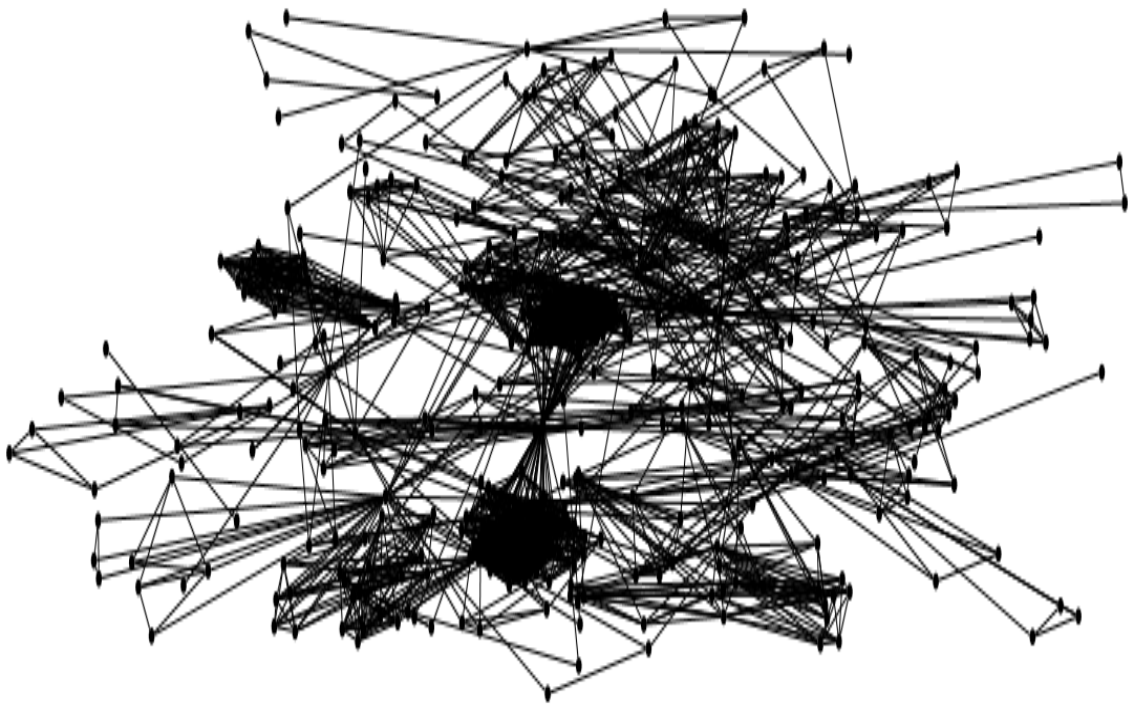
Από τις πιο πάνω γραφικές παραστάσεις μπορούμε να δούμε τον αριθμό και το μέσο όρο των κόμβων που ανήκουν σε ένα υπογράφο(subgraph), ο οποίος περιέχει την πλειοψηφία των κόμβων ολόκληρου του γράφου. Από το 1993, το τμήμα της Πληροφορικής του Πανεπιστημίου Κύπρου άρχισε να δραστηριοποιείται. Το 2006, 2007 και 2008 που δημοσιεύτηκαν περισσότερα άρθρα και υπήρξαν περισσότερες συνεργασίες επιστημόνων, περισσότερο από 51% των επιστημόνων του ολικού γράφου αποτελούν μέρος μιας κοινότητας, με την έννοια ότι όλοι απέχουν μερικά βήματα – ενδιάμεσες γνωριμίες μακριά από οποιοδήποτε άλλο επιστήμονα αυτής της κοινότητας. Παρόλο που το 2009 το τμήμα Πληροφορικής δημοσίευσε τα περισσότερα άρθρα και είχε τις περισσότερες συνεργασίες, εντούτοις ο μέσος όρος κόμβων που ανήκουν στο giant component δεν είναι πολύ μεγάλος, μόνο 20% του γράφου. Αυτό συμβαίνει γιατί εμφανίζονται νέοι επιστήμονες ή υπάρχουν πολλά ημι-απομονωμένα άτομα και ομάδες που ξεκινούν τις έρευνες τους ανεξάρτητα. Το 1992 παρατηρείται πως το giant component καλύπτει όλο το γράφο γιατί υπήρξε δημοσίευση μόνο ενός άρθρου.

Ακολουθούν οι γράφοι του 1997 και του 2009 για το τμήμα Πληροφορικής του Πανεπιστημίου Κύπρου.

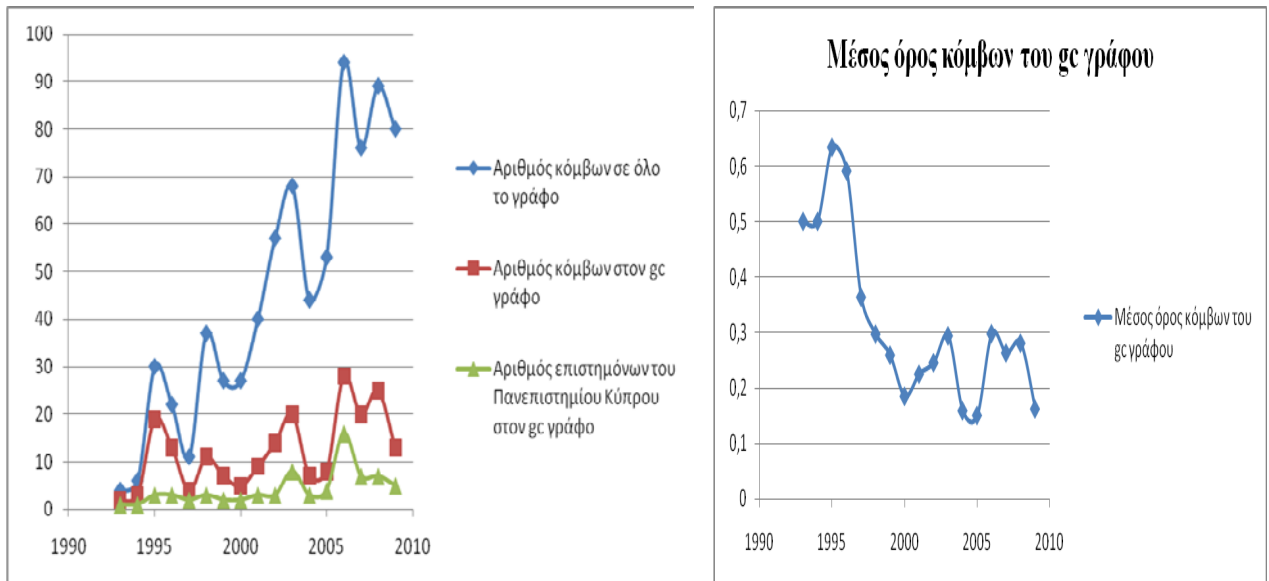
- Γράφος του 1997 για το τμήμα Πληροφορικής



- Γράφος του 2009 για το τμήμα Πληροφορικής



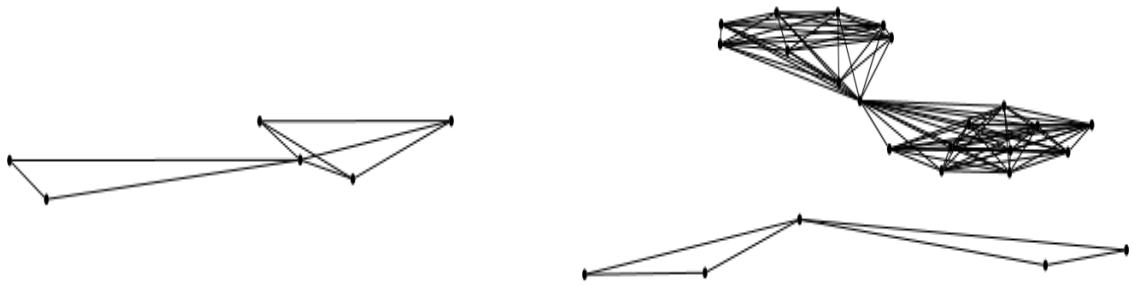
❖ **Giant component (αριθμός κόμβων-μέσος όρος κόμβων)**
τμήματος Μαθηματικών και Στατιστικής



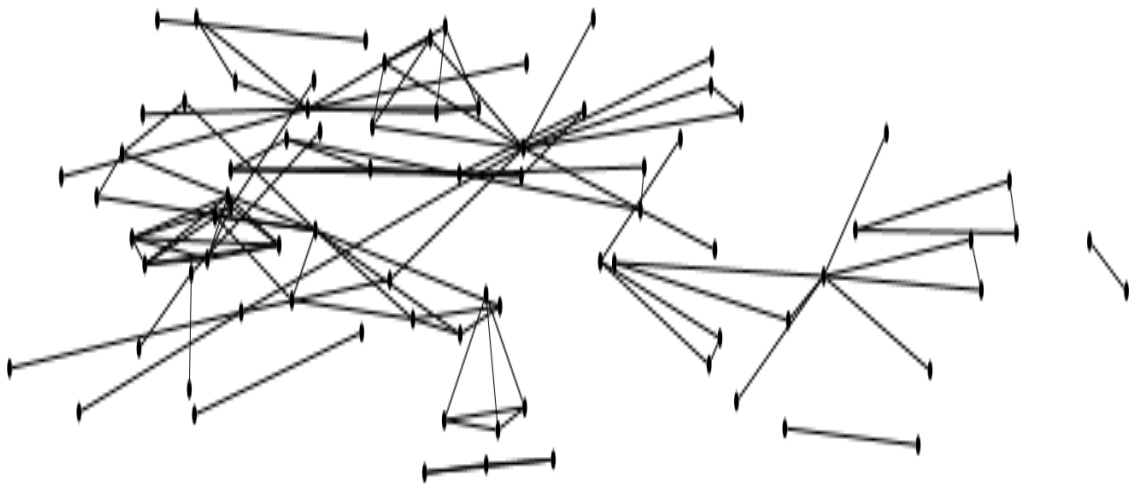
Από τις πιο πάνω γραφικές παραστάσεις μπορούμε να δούμε τον αριθμό και το μέσο όρο των κόμβων που ανήκουν σε ένα υπογράφο (subgraph), ο οποίος περιέχει την πλειοψηφία των κόμβων ολόκληρου του γράφου. Από το 1995, το τμήμα Μαθηματικών και Στατιστικής του Πανεπιστημίου Κύπρου άρχισε να δραστηριοποιείται. Όπως έχει παρατηρηθεί και από τα αποτελέσματα του giant component όλων των χρονιών, το τμήμα Μαθηματικών και Στατιστικής είναι το τμήμα με το μικρότερο giant component. Αυτό ίσως και να οφείλεται στο γεγονός ότι οι επιστήμονες αυτού του τμήματος συνεργάζονται με ένα ή δύο ως επί των πλείστον επιστήμονες. Ο μέσος όρος κόμβων που ανήκουν στο giant component κυμαίνεται από 15% - 30%. Δηλαδή, το 15%-30% των επιστημόνων είναι μέρος κοινοτήτων, με την έννοια ότι αυτοί οι επιστήμονες απέχουν μερικά βήματα – ενδιάμεσες γνωριμίες μακριά από οποιοδήποτε άλλο επιστήμονα της συγκεκριμένης κοινότητας.

Ακολουθούν οι γράφοι του 1995 και του 2009 για το τμήμα Μαθηματικών και Στατιστικής του Πανεπιστημίου Κύπρου.

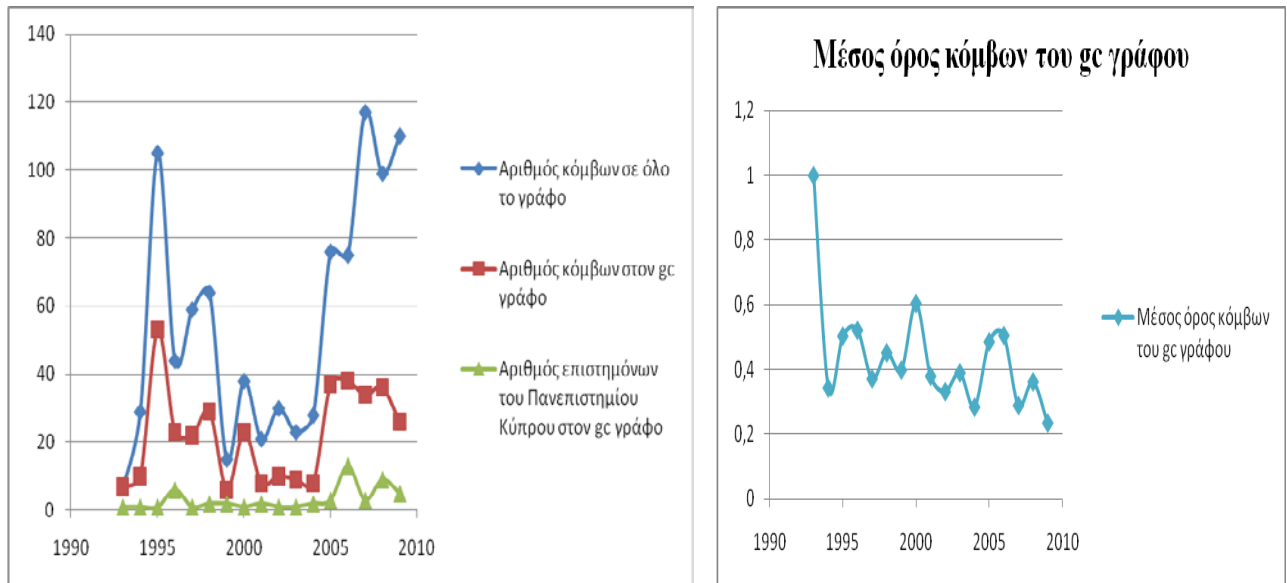
- Γράφος τον 1995 για το τμήμα Μαθηματικών και Στατιστικής



- Γράφος τον 2009 για το τμήμα Μαθηματικών και Στατιστικής



❖ **Giant component (αριθμός κόμβων-μέσος όρος κόμβων)**
τιμή ματος Φυσικής

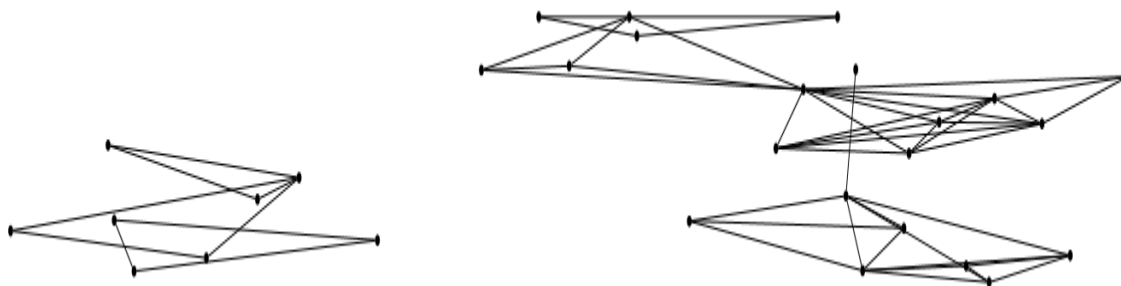


Από τις πιο πάνω γραφικές

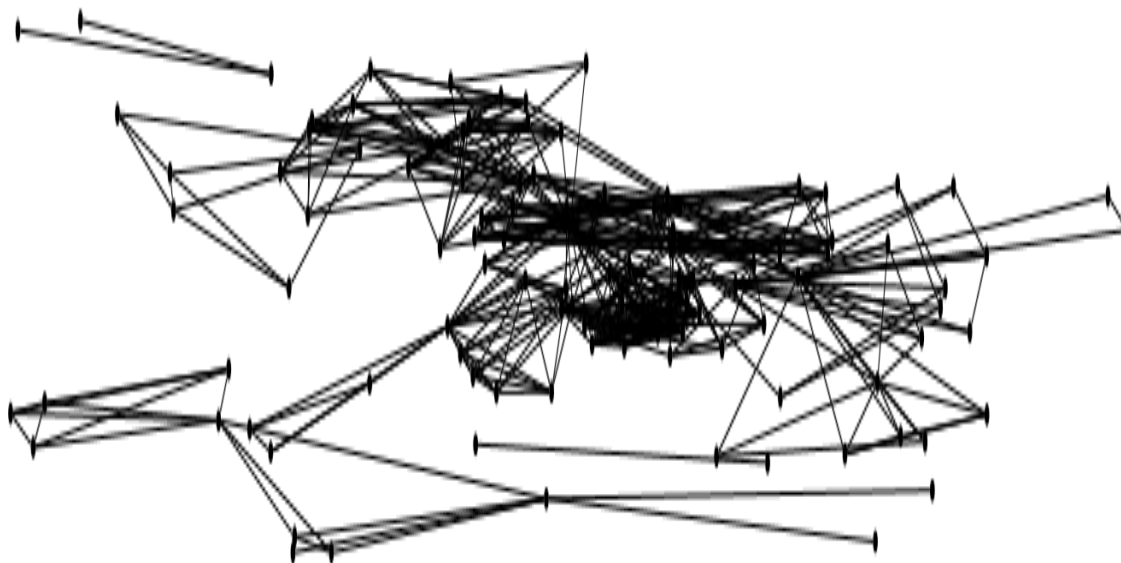
παραστάσεις μπορούμε να δούμε τον αριθμό και το μέσο όρο των κόμβων που ανήκουν σε ένα υπογράφο(subgraph), ο οποίος περιέχει την πλειοψηφία των κόμβων ολόκληρου του γράφου. Από το 1994, το τμήμα της Φυσικής του Πανεπιστημίου Κύπρου άρχισε να δραστηριοποιείται. Παρατηρείται πως κάθε χρονιά, από το 1995 έως το 2009, ο μέσος όρος κόμβων στο giant component γράφο δεν είναι ούτε σταθερός, αλλά ούτε παρατηρείται κάποια ισορροπία. Άλλοτε το ποσοστό είναι μεγάλο (60% το 2000) και άλλοτε μικρό (24% το 2009). Στη γραφική παράσταση με τα ποσοστά των αριθμών κόμβων στο gc γράφο, παρατηρούνται σκαμπανεβάσματα. Αυτό οφείλεται στο γεγονός ότι δίκτυα συνεργασίας συνεχώς εξελίσσονται. Όταν το μέγεθος του giant component μεγαλώνει αυτό οφείλεται στο γεγονός ότι νέοι επιστήμονες συνεργάστηκαν με επιστήμονες που ήδη ανήκουν στο giant component και πλέον ανήκουν και αυτή στην ίδια κοινότητα με τους 'παλιούς' επιστήμονες. Επιπρόσθετα, μπορεί να οφείλεται και στην απορρόφηση μερικών μικρών, μη συνδεδεμένων components στο giant component χάρις σε μία ή περισσότερες νέες συνεργασίες. Όταν το μέγεθος του giant component μικραίνει αυτό οφείλεται στο γεγονός ότι εμφανίζονται νέοι επιστήμονες ή υπάρχουν και πολλά ημι-απομονωμένα άτομα και ομάδες που ξεκινούν τις έρευνες τους ανεξάρτητα.

Ακολουθούν οι γράφοι του 1994 και του 2009 για το τμήμα Φυσικής του Πανεπιστημίου Κύπρου.

- Γράφος του 1994 για το τμήμα Φυσικής



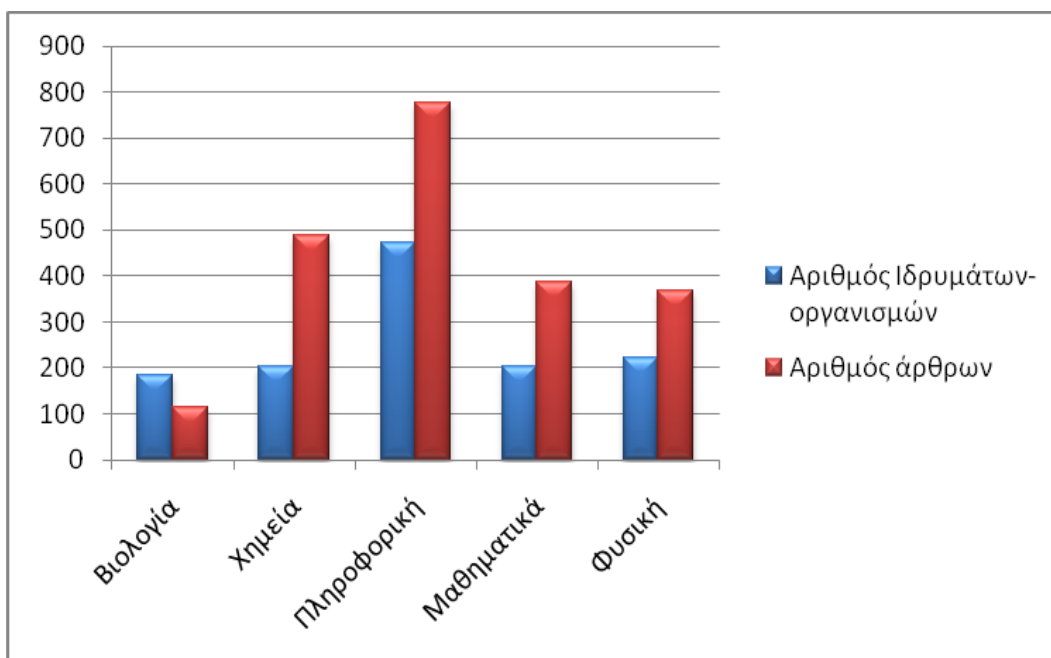
- Γράφος του 2009 για το τμήμα Φυσικής



5.6 Αποτελέσματα δικτύου συνεργασίας ιδρυμάτων-οργανισμών

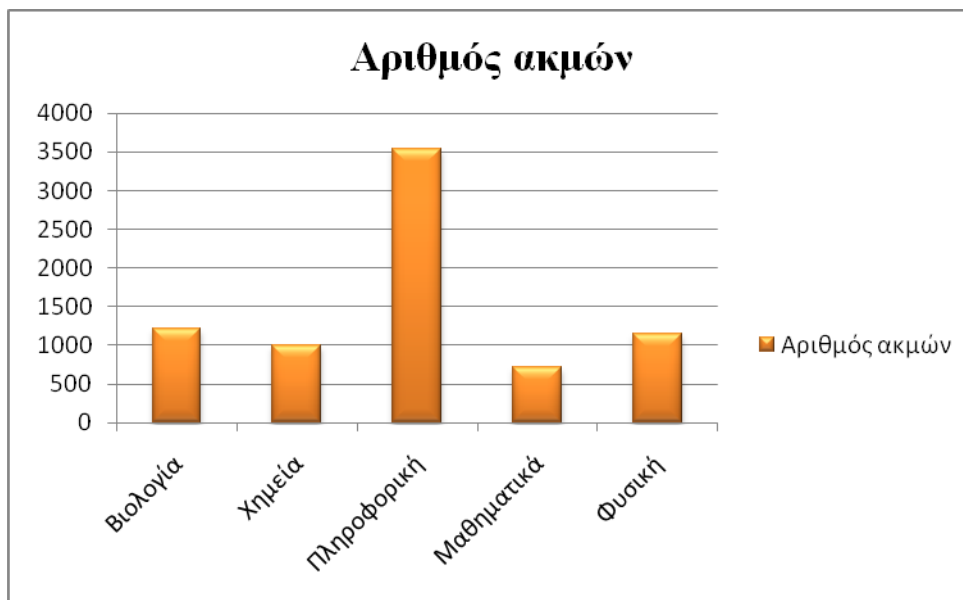
Χρησιμοποιώντας τη βιβλιοθήκη SNAP δημιουργήθηκε πρόγραμμα μέσω του οποίου αναλύθηκαν τα πέντε δίκτυα συνεργασίας ιδρυμάτων-οργανισμών που αντιστοιχούν στα πέντε τμήματα προς μελέτη. Τα αποτελέσματα που προέκυψαν παρουσιάζονται με τις εξής γραφικές παραστάσεις. Αναλυτικά τα αποτελέσματα παρουσιάζονται στο παράρτημα Γ.

❖ Συνολικός αριθμός ιδρυμάτων-οργανισμών και άρθρων



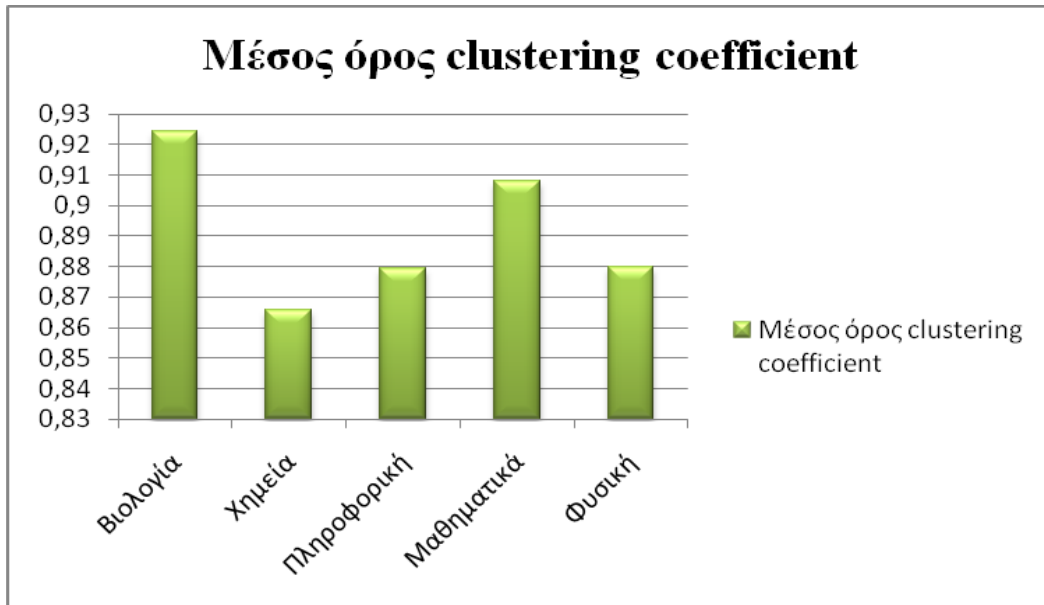
Από αυτή τη γραφική παράσταση, για κάθε ένα από τα πέντε τμήματα μπορούμε να δούμε τόσο το συνολικό αριθμό ιδρυμάτων που συνεργάστηκαν για τη συγγραφή κάποιων άρθρων όσο και τον αριθμό των άρθρων που δημοσιεύτηκαν. Ο αριθμός των άρθρων του κάθε τμήματος σαφώς είναι ίδιος με αυτόν που υπολογίστηκε στο δίκτυο συνεργασίας επιστημόνων. Εδώ, παρατηρούμε αφού για το τμήμα Πληροφορικής υπάρχουν πληροφορίες για πολλά άρθρα ήταν φυσικό επακόλουθο να συνεργαστούν πολλά ιδρύματα-οργανισμοί. Τα υπόλοιπα τμήματα έχουν συνεργαστεί περίπου με 200 ιδρύματα-οργανισμοί.

❖ Αριθμός ακμών γράφου ιδρυμάτων-οργανισμών



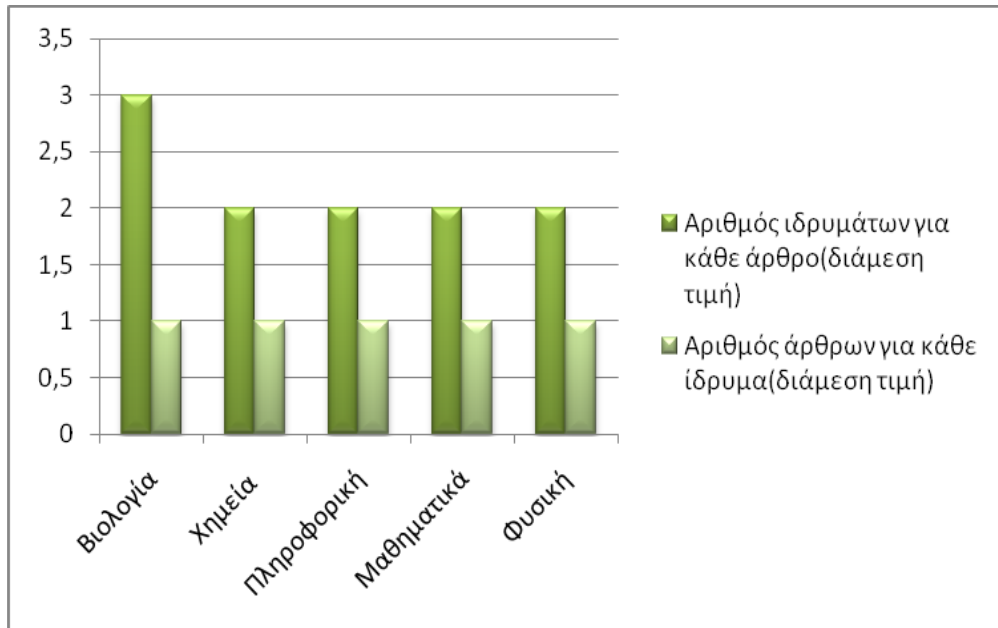
Η γραφική αυτή παράσταση παρουσιάζει τον αριθμό ακμών που δημιουργήθηκαν για κάθε ένα γράφο που απεικονίζει τη συνεργασία των ιδρυμάτων-οργανισμών του αντίστοιχου τμήματος. Εντύπωση δημιουργεί ο τεράστιος αριθμός ακμών που παρουσιάζεται στο γράφο συνεργασίας ιδρυμάτων-οργανισμών μιας και ο αριθμός των ακμών είναι ίσος με 3534 ακμές. Η διαφορά είναι μεγάλη από το δεύτερο μεγαλύτερο σε αριθμό ακμών, αυτών του γράφου συνεργασίας του τμήματος Βιολογίας (1211 ακμές). Συνεπώς, αφού στο γράφο συνεργασίας του τμήματος Πληροφορικής συνεργάστηκαν πολλά διαφορετικά ιδρύματα-οργανισμοί(473 ιδρύματα) ήταν φυσικό επακόλουθο να δημιουργηθούν πολλές διαφορετικές ακμές μεταξύ των ιδρυμάτων. Αφού τα άλλα τέσσερα τμήματα έχουν συνεργαστεί με περίπου 200 ιδρύματα, τότε και ο αριθμός ακμών των γράφων τους είναι περίπου ο ίδιος. Ο αριθμός ακμών των γράφων συνεργασίας των τμημάτων Βιολογίας, Χημείας και Φυσικής είναι περίπου 1000 διαφορετικές ακμές και του τμήματος Μαθηματικών και Στατιστικής είναι 722.

❖ Μέσος όρος clustering coefficient



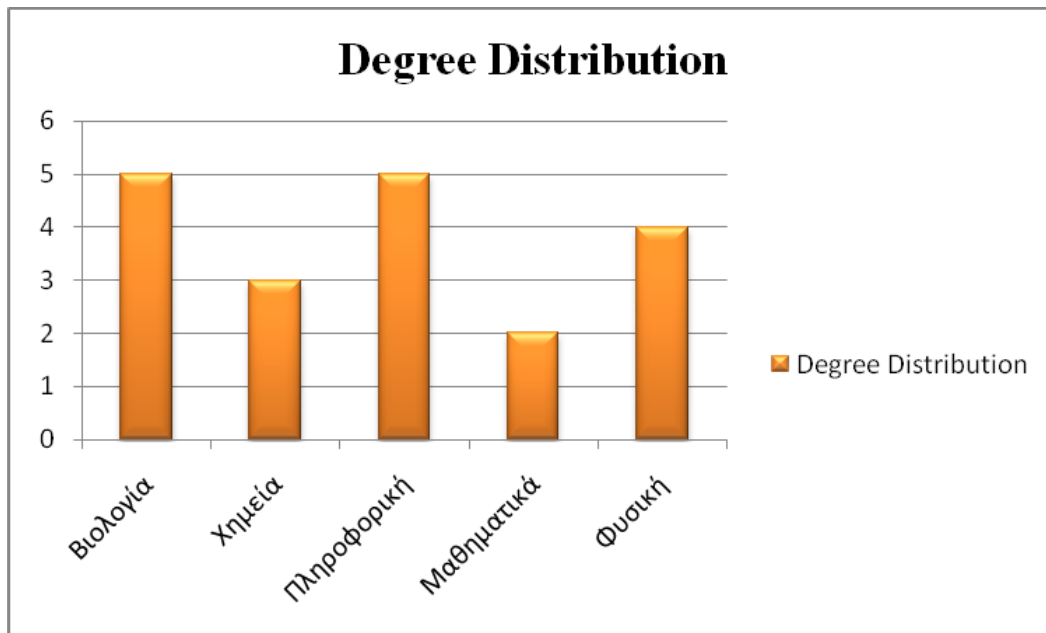
Από αυτή τη γραφική παράσταση μπορούμε να παρατηρήσουμε το βαθμό με τον οποίο οι κόμβοι του κάθε γράφου ιδρυμάτων τείνουν να γίνονται ένα σύμπλεγμα μαζί, να δημιουργούν έτσι κάποιες ομάδες οι οποίες χαρακτηρίζονται σχετικά από μεγάλη πυκνότητα από δεσμούς. Παρατηρούμε πως και στα πέντε τμήματα δύο ιδρύματα-οργανισμοί τυπικά έχουν περισσότερο από 80% πιθανότητα να συνεργαστούν αν έχουν ήδη συνεργαστεί με ένα άλλο ίδρυμα-οργανισμό. Είναι γνωστό πως όσο πιο μεγάλη είναι αυτή η τιμή, τόσο μεγαλύτερη είναι και η πιθανότητα ιδρύματα-οργανισμοί να έχουν συνεργαστεί για τη συγγραφή και άλλου άρθρου. Συγκεκριμένα, παρατηρούμε πως στο τμήμα Βιολογίας υπάρχουν πολλές ομάδες που χαρακτηρίζονται από μεγάλη πυκνότητα από δεσμούς και δύο ιδρύματα-οργανισμοί. Μπορεί το τμήμα Χημείας να έχει το μικρότερο μέσο όρο clustering coefficient (87%), εντούτοις είναι αρκετά μεγάλος. Συνεπώς, αυτό δείχνει πως οι επιστήμονες των ιδρυμάτων-οργανισμών και των πέντε τμημάτων τείνουν να προτείνουν τους συνεργάτες τους σε επιστήμονες άλλων ιδρυμάτων-οργανισμών και αυτό έχει ως αποτέλεσμα τη δημιουργία νέων συνεργασιών και αύξηση των συμπλεγμάτων.

❖ **Αριθμός άρθρων για κάθε ίδρυμα και αριθμός ιδρυμάτων για κάθε άρθρο**



Από αυτή τη γραφική παράσταση βλέπουμε πως ο διάμεσος αριθμός ιδρυμάτων-οργανισμών που συνεργαστήκαν για τη συγγραφή άρθρων που αναφέρονται στο τμήμα Βιολογίας είναι ίσος με τρία ενώ για τα υπόλοιπα τμήματα είναι ίσος με 2. Από την άλλη παρατηρούμε πως ο διάμεσος αριθμός άρθρων για κάθε ίδρυμα-οργανισμό είναι και για τα πέντε τμήματα ίσος με ένα. Αυτό οφείλεται στο γεγονός ότι ο αριθμός των ιδρυμάτων με τα οποία συνεργάστηκε το Πανεπιστήμιο Κύπρου είναι μεγάλος, με αποτέλεσμα να υπάρχουν πολλά ιδρύματα-οργανισμοί με τα οποία το Πανεπιστήμιο Κύπρου να συνεργάστηκε μόνο μία φορά και αυτό να επηρέασε το τελικό αποτέλεσμα. Αναφορικά, ένα ίδρυμα μπορεί να συνεργάζεται με κάποιο άλλο ίδρυμα για τη συγγραφή ενός έως 15 άρθρων του τμήματος Βιολογίας, ενός έως 41 άρθρων του τμήματος Χημείας, ενός έως 49 άρθρων του τμήματος Πληροφορικής, ενός έως 18 άρθρων του τμήματος Μαθηματικών και Στατιστικής και ενός έως 18 άρθρων του τμήματος Φυσικής. Δεν έχει συμπεριληφθεί το Πανεπιστήμιο Κύπρου αφού αυτό έχει συνεργαστεί με όλα τα υπόλοιπα ιδρύματα-οργανισμούς.

❖ Degree Distribution



Με βάση αυτή τη γραφική παράσταση μπορούμε να δούμε τον διάμεσο αριθμό συνδέσεων κόμβου τύπου ίδρυμα. Παρατηρούμε πως ο διάμεσος αριθμός συνδέσεων ιδρυμάτων-οργανισμών για το τμήμα Βιολογίας και Πληροφορικής είναι ίσος με 5, για το τμήμα Φυσικής είναι ίσος με τέσσερα, για το τμήμα Χημείας είναι ίσος με τρία και για το τμήμα Μαθηματικών και Στατιστικής είναι ίσος με δύο. Ο διάμεσος αριθμός συνδέσεων ιδρυμάτων-κόμβων για τα πέντε τμήματα δεν είναι μεγάλος και δεν υπερβαίνει το μέσο αριθμό συνδέσεων. Συνεπώς, και στις πέντε περιπτώσεις οι κατανομές είναι δεξιά ασύμμετρες, πράγμα που σημαίνει ότι η πλειοψηφία των κόμβων και στους πέντε γράφους έχουν low-degree.

❖ **Τα 10 πρώτα ιδρύματα-οργανισμοί, τα οποία συνεργάστηκαν περισσότερο με το Πανεπιστήμιο Κύπρου**

Τα πρώτα ιδρύματα-οργανισμοί, τα οποία συνεργάστηκαν περισσότερο με το Πανεπιστήμιο Κύπρου παρουσιάζονται ανά τμήμα στους πίνακες Πίνακας 5.2-5.6 .

Τμήμα Βιολογίας

	Όνομα Ιδρύματος-Οργανισμού	Αριθμός κατάταξης	Βάρος ακμής
1	Nicosia General Hospital	6	15
2	Niels Bohr Institute, University of Copenhagen	1(N ^ο 43)	15
3	Cyprus Institute of Neurology and Genetics	6	14
4	University of Nicosia	6	9
5	National and Kapodistrian University of Athens	3	8
6	Yasoo Health Inc.	6	5
7	Aristotle University of Thessaloniki	4	5
8	Ippokrateion Private Hospital	6	5
9	European Molecular Biology Lab(EMBL)	6	4
10	Democritus University of Thrace	6	4

Πίνακας 5.2: Τα πρώτα δέκα ιδρύματα-οργανισμοί με τα οποία συνεργάστηκε περισσότερο το τμήμα Βιολογίας του Πανεπιστημίου Κύπρου

Τμήμα Χημείας

	Όνομα Ιδρύματος-Οργανισμού	Αριθμός κατάταξης	Βάρος ακμής
1	University of Patras	6	41
2	University of Ioannina	6	27
3	University of Crete	6	21
4	University of Florida	1(N ^ο 58)	14
5	National and Kapodistrian University of Athens	3	12
6	Northeastern University	5	12
7	University of Manchester	1 (N ^ο 41)	11
8	National Technical University of Athens (NTUA)	6	11
9	Institute of Catalysis and Petrochemistry (ICP)	6	10
10	Aristotle University of Thessaloniki	4	10

Πίνακας 5.3: Τα πρώτα δέκα ιδρύματα-οργανισμοί με τα οποία συνεργάστηκε περισσότερο το τμήμα Χημείας του Πανεπιστημίου Κύπρου

Τμήμα Πληροφορικής

	Όνομα Ιδρύματος-Οργανισμού	Αριθμός κατάταξης	Βάρος ακμής
1	University of Patras	6	49
2	Cyprus Institute of Neurology and Genetics	6	39
3	European Organization for Nuclear Research (CERN)	6	35
4	National and Kapodistrian University of Athens	3	30
5	University of New Mexico	3	29
6	University of Paderborn	6	23
7	Paul Scherrer Institute	6	20
8	University of Pittsburgh	1(N ^ο 50)	19
9	University of Ioannina	6	17
10	University of Nicosia	6	17

Πίνακας 5.4: Τα πρώτα δέκα ιδρύματα-οργανισμοί με τα οποία συνεργάστηκε περισσότερο το τμήμα Πληροφορικής του Πανεπιστημίου Κύπρου

Τμήμα Μαθηματικών και Στατιστικής

	Όνομα Ιδρύματος-Οργανισμού	Αριθμός κατάταξης	Βάρος ακμής
1	Colorado School of Mines	6	18
2	Niels Bohr Institute, University of Copenhagen	1(N ^ο 43)	13
3	Institute for Mathematics and its Applications(IMA), Univeristy of Minnesota	1(N ^ο 28)	10
4	University of Maryland, College Park	1(N ^ο 37)	9
5	University of Copenhagen	1(N ^ο 43)	9
6	University of California , Riverside	2	9
7	Electricity Authority of Cyprus	6	8
8	University Joseph Fourier	2	7
9	University of South Carolina – Columbia	3	7
10	Nicosia General Hospital	6	6

Πίνακας 5.5: Τα πρώτα δέκα ιδρύματα-οργανισμοί με τα οποία συνεργάστηκε περισσότερο το τμήμα Μαθηματικών και Στατιστικής του Πανεπιστημίου Κύπρου

Τμήμα Φυσικής

	Όνομα Ιδρύματος-Οργανισμού	Αριθμός κατάταξης	Βάρος ακμής
1	University of Toronto	1(N ^ο 27)	18
2	University of Pisa	2	17
3	Università di Genova e Sezione INFN	4	15
4	National and Kapodistrian University of Athens	3	13
5	Institute of Nuclear Physics-National Center for Scientific Research “Demokritos”	6	10
6	GSI Helmholtz Centre for Heavy Ion Research GmbH (Gesellschaft für Schwerionenforschung)	6	10
7	University of Crete	6	9
8	University of Liverpool	2	9
9	Korea Advanced Institute of Science and Technology (KAIST)	2	8
10	Yerevan State University	6	8

Πίνακας 5.6: Τα πρώτα δέκα ιδρύματα-οργανισμοί με τα οποία συνεργάστηκε περισσότερο το τμήμα Φυσικής του Πανεπιστημίου Κύπρου

Κεφάλαιο 6

Συμπεράσματα

6.1 Τελικά Συμπεράσματα

111

6.1 Τελικά Συμπεράσματα

Η δομή των δικτύων συνεργασίας του Πανεπιστημίου Κύπρου έχει διερευνηθεί. Δύο επιστήμονες θεωρούνται συνδεδεμένοι εφόσον έχουν συγγράψει μαζί ένα άρθρο, συνεπώς και δύο ιδρύματα-οργανισμοί θεωρούνται συνδεδεμένοι εφόσον επιστήμονές τους έχουν συνεργαστεί με άλλους επιστήμονες άλλων ιδρυμάτων-οργανισμών για τη συγγραφή κάποιου άρθρου. Τα δίκτυα αυτά έχουν δημιουργηθεί με βάση κάποια δεδομένα που έχουν αντληθεί από τις βάσεις *Scopus* και *Citation Index (ISI-Web of Knowledge)* που παρέχει η βιβλιοθήκη του Πανεπιστημίου Κύπρου.

Έχει αποδειχθεί ότι τα δίκτυα συνεργασίας των επιστημόνων και των ιδρυμάτων-οργανισμών και των πέντε τμημάτων (τμήμα Βιολογίας, τμήμα Χημείας, τμήμα Πληροφορικής, τμήμα Μαθηματικών και Στατιστικής, τμήμα Φυσικής) χαρακτηρίζονται σχετικά από μεγάλη πυκνότητα από δεσμούς. Αυτό σημαίνει πως δύο επιστήμονες είναι πολύ περισσότερο πιθανόν να έχουν συνεργαστεί εάν έχουν ήδη προηγουμένως συνεργαστεί με ένα τρίτο κοινό επιστήμονα. (Το ίδιο ισχύει και για τα ιδρύματα-οργανισμούς.) Αυτό δείχνει πως η διαδικασία των επιστημόνων να παρουσιάζουν και να προτείνουν σε άλλους επιστήμονες τους συνεργάτες τους, είναι σημαντική για την ανάπτυξη των επιστημονικών κοινοτήτων και συνεπώς και των κοινοτήτων των ιδρυμάτων.

Επιπλέον, έχουν μελετηθεί και οι κατανομές τόσο του αριθμού επιστημόνων που συνεργάζονται για τη συγγραφή άρθρου όσο και του αριθμού άρθρων που συγγράφει

έναν επιστήμονα του Πανεπιστημίου Κύπρου με άλλους επιστήμονες εντός ή/και εκτός Πανεπιστημίου Κύπρου. Στην πρώτη περίπτωση, παρατηρείται πως ο διάμεσος αριθμός επιστημόνων ανά άρθρο δεν είναι ίδιος για όλα τα τμήματα. Για τα άρθρα που είναι πιο πειραματικά, όπως για παράδειγμα αυτά που αναφέρονται στο τμήμα Βιολογίας και Χημείας, ο αριθμός των επιστημόνων ανά άρθρο είναι μεγάλος, περισσότερα από τέσσερα άτομα. Σε αντίθεση, για τα άρθρα που είναι πιο θεωρητικά, όπως για παράδειγμα αυτά που αναφέρονται στο τμήμα Μαθηματικών και Στατιστικής, ο αριθμός των επιστημόνων ανά άρθρο είναι μόλις δύο, ίσος με τον αναγκαίο αριθμό επιστημόνων για να υπάρχει συνεργασία. Στη δεύτερη περίπτωση, παρατηρείται πως ο διάμεσος αριθμός άρθρων ανά επιστήμονα του Πανεπιστημίου Κύπρου είναι ίδιος για όλα τα τμήματα, ίσος με δύο άρθρα το χρόνο.

Επιπρόσθετα, έχουν μελετηθεί και οι κατανομές τόσο του αριθμού εσωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου όσο και των εξωτερικών συνεργατών τους ανά άρθρο. Παρατήθηκε πως οι επιστήμονες που ανήκουν στα τμήματα Βιολογίας, Φυσικής και Χημείας, τείνουν να συνεργάζονται περισσότερο με επιστήμονες άλλων ιδρυμάτων-οργανισμών. Παρόλο που το τμήμα Πληροφορικής του Πανεπιστημίου Κύπρου έχει τους περισσότερους επιστήμονες από τα υπόλοιπα τμήματα, εντούτοις ο αριθμός τόσο των εσωτερικών όσο και των εξωτερικών συνεργατών είναι ίσος με δύο και στις δύο περιπτώσεις. Στην περίπτωση άρθρων που αναφέρονται στο τμήμα Μαθηματικών και Στατιστικής, αφού ο αριθμός των επιστημόνων που συνεργάζονται για τη συγγραφή τους είναι μικρός, ήταν λοιπόν φυσικό επακόλουθο το τμήμα αυτό να έχει το μικρότερο αριθμό εσωτερικών και εξωτερικών συνεργατών. Συνεπώς, σε πειραματικά άρθρα οι επιστήμονες του Πανεπιστημίου Κύπρου τείνουν να συνεργάζονται περισσότερο με επιστήμονες άλλων ιδρυμάτων-οργανισμών, σε αντίθεση με θεωρητικά άρθρα που οι επιστήμονες του Πανεπιστημίου Κύπρου συνεργάζονται με ένα μικρό αριθμό εσωτερικών και εξωτερικών συνεργατών.

Επιπλέον, έχει υπολογιστεί και ο διάμεσος αριθμός κόμβων σε γράφους που απεικονίζουν τις συνεργασίες των επιστημόνων και τις συνεργασίες των ιδρυμάτων-οργανισμών. Και στις δύο περιπτώσεις, και για τα πέντε τμήματα, η πλειοψηφία των κόμβων και στους πέντε γράφους έχουν low-degree.

Με βάση την ανάλυση των δικτύων συνεργασίας ιδρυμάτων-οργανισμών έχει δημιουργηθεί η δεκάδα των ιδρυμάτων-οργανισμών, με τα οποία συνεργάστηκε περισσότερο κάθε ένα από τα πέντε προς μελέτη τμήματα του Πανεπιστημίου Κύπρου. Έχει λοιπόν αποδειχθεί πως οι επιστήμονες του Πανεπιστημίου Κύπρου και των πέντε τμημάτων, ιδιαίτερα τα τελευταία χρόνια, συνεργάζονται με επιστήμονες που ανήκουν σε παγκοσμίως αναγνωρισμένα πανεπιστήμια, τα οποία κατέχουν υψηλές θέσεις στην 500άδα των καλύτερων πανεπιστημίων παγκοσμίως (με βάση τη λίστα της Shanghai). Στη δεκάδα των τμημάτων υπάρχουν μεγάλα ερευνητικά κέντρα και ιδρύματα. Επιπλέον, παρατηρείται πως οι επιστήμονες του Πανεπιστημίου Κύπρου συνεργάζονται με πολλά εκπαιδευτικά ιδρύματα της Ελλάδας αλλά και με οργανισμούς που βρίσκονται στην Κύπρο.

Επιπρόσθετα, μελετήθηκε η ύπαρξη ενός μεγάλου connected component (giant component) σε γράφους που αναπαριστούν συνεργασίες γιατί η ύπαρξή του έχει κοινωνική σημασία. Το ποσοστό των επιστημόνων που ανήκουν στο giant component είναι μέλη μιας ενιαίας κοινότητας. Κοινότητα μπορεί να θεωρηθεί ως ένα σύνολο από πυκνά συνδεδεμένες κορυφές που έχουν συνδέσεις με κορυφές που ανήκουν σε άλλες κοινότητες, η οποία δείχνει τη δομή των ομάδων γύρω από αυτές. Όπως γνωρίζουμε από προσωπική εμπειρία, κάποια άτομα έχουν στενότερους δεσμούς σε σύγκριση με κάποια άλλα και θα μπορούσαμε να πούμε ότι αυτοί ανήκουν σε μια πιο στενή κοινότητα. Στην προκειμένη περίπτωση, στους γράφους δεν παρατηρείται κάποια ισορροπία αφού τα δίκτυα συνεργασίας εξελίσσονται συνεχώς. Όταν το ποσοστό των επιστημόνων που ανήκουν στο giant component είναι μεγάλο, παρατηρείται πως μικρά components συνδέονται με μεγαλύτερα και και νέοι επιστήμονες συνεργάζονται με επιστήμονες που ήδη ανήκουν στο giant component. Όταν το ποσοστό είναι μικρό, αυτό μπορεί να οφείλεται στο γεγονός ότι νέοι επιστήμονες, πολλά ημι-απομονωμένα άτομα και ομάδες ξεκινούν τις έρευνες ανεξάρτητα.

Εν κατακλείδι, η εργασία αυτή αναπόφευκτα αντιπροσωπεύει μια πρώτη ματιά των δικτύων συνεργασίας του Πανεπιστημίου Κύπρου. Έχει αποδειχθεί πως χρονιά με τη χρονιά τα πέντε η δραστηριότητα αυτών των τμημάτων γίνεται περισσότερο έντονη και σίγουρα με το πέρασμα των χρόνων ολοένα και περισσότερα άρθρα θα συγγράφονται από επιστήμονες του Πανεπιστημίου Κύπρου, 'νέους' ή 'παλιούς'. Επίσης, ολοένα θα

αυξάνονται οι συνεργασίες με ιδρύματα-οργανισμούς που έχει ήδη συνεργαστεί το Πανεπιστήμιο Κύπρου και θα δημιουργούνται συνεργασίες με ιδρύματα-οργανισμούς που δεν έχει συνεργαστεί ξανά το Πανεπιστήμιο Κύπρου. Αναμφίβολα, η μελέτη τέτοιων δικτύων συνεργασίας είναι πολλή σημαντική γιατί η δομή τέτοιων δικτύων που απεικονίζουν συνεργασίες, μπορούν να ερμηνεύσει τον τρόπο με τον οποίο μπορούν να διαδοθούν πληρο φρίες και συνεπώς να βοηθήσουν στη συνεχή ανάπτυξη των αντίστοιχων κλάδων.

Βιβλιογραφία

- [1] Barabasi, A.-L. (2003). *Linked: how everything is connected to everything else and what it means for business, science, and everyday life*. New York: A Plume Book
- [2] Newman, M.E.J. (2001). The structure of scientific collaboration networks, *PNAS*, 98(2001), 404-409
- [3] Tomassini, M., Luthi, L. (2007). Empirical analysis of the evolution of a scientific collaboration network, *Physica A*, 385(2007), 750-764

Παράρτημα Α

Ακολουθούν αναλυτικά τα αποτελέσματα ανάλυσης δικτύου συνεργασίας επιστημόνων συγκεκριμένου τμήματος για όλες τις χρονιές δημοσίευσης άρθρων(1992-2009).

❖ *Συνολικός αριθμός επιστημόνων και άρθρων ανά τμήμα*

	Αριθμός επιστημόνων	Αριθμός άρθρων
Βιολογία	463	114
Χημεία	675	487
Πληροφορική	1205	777
Μαθηματικά	424	386
Φυσική	556	367

❖ *Αριθμός ακμών επιστημονικού γράφου ανά τμήμα*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
Αριθμός ακμών	3678	4490	8983	1671	4512

❖ *Μέσος όρος clustering coefficient*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
Μέσος όρος clustering coefficient	0,9215	0,8731	0,8943	0,886	0,9068

❖ *Αριθμός επιστημόνων ανά άρθρο και αριθμός άρθρων ανά επιστήμονα του Πανεπιστημίου Κύπρου*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
Αριθμός επιστημόνων ανά άρθρο(διάμεση τιμή)	5,5	4	3	2	3
Αριθμός άρθρων ανά επιστήμονα του Πανεπιστημίου Κύπρου(διάμεση τιμή)	2	2	2	2	1,5

❖ *Αριθμός εσωτερικών και εξωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
Αριθμός εσωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου	4	3	3	2	2,5
Αριθμός εξωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου	7	2	2	3	3,5

❖ *Degree Distribution*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
Degree Distribution	9	6	6	3	7

❖ *Giant component*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
Αριθμός κόμβων σε όλο το γράφο	463	675	1205	424	556
Αριθμός κόμβων στον gc γράφο	411	637	1164	168	523
Μέσος όρος κόμβων του gc γράφου	0,89	0,94	0,97	0,4	0,941
Αριθμός επιστημόνων του Πανεπιστημίου Κύπρου στον gc γράφο	51	124	205	33	78

Παράρτημα Β

Ακολουθούν αναλυτικά τα αποτελέσματα ανάλυσης δικτύου συνεργασίας επιστημόνων συγκεκριμένου τμήματος συγκεκριμένης χρονιάς δημοσίευσης άρθρων(1992-2009).

❖ *Συνολικός αριθμός επιστημόνων ανά τμήμα και χρονιά*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
1992		3	6		
1993		5	10	4	7
1994		4	35	6	29
1995		24	65	30	105
1996		3	13	22	44
1997		10	40	11	59
1998	11	20	38	37	64
1999		25	54	27	15
2000	2	29	77	27	38
2001		62	53	40	21
2002	32	58	70	57	30
2003	41	47	119	68	23
2004	70	83	210	44	28
2005	43	104	212	53	76
2006	58	190	219	94	75
2007	92	176	206	76	117
2008	70	166	262	89	99
2009	165	142	389	80	110

❖ *Συνολικός αριθμός άρθρων ανά τμήμα και χρονιά*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
1992		1	1		
1993		2	3	2	2
1994		1	14	5	18
1995		9	24	8	26
1996		1	6	11	16
1997		4	21	5	24
1998	2	7	15	23	21
1999		12	22	16	8
2000	1	20	31	16	11
2001		26	27	20	8
2002	5	27	31	21	23
2003	5	15	42	27	14
2004	12	37	66	26	11
2005	12	35	81	29	24
2006	15	69	97	37	39
2007	13	79	95	40	44

2008	18	68	94	48	35
2009	31	74	107	52	43

❖ *Αριθμός ακμών επιστημονικού γράφου ανά τμήμα και χρονιά*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
1992		3	15		
1993		6	16	2	12
1994		6	139	5	89
1995		96	224	108	854
1996		3	19	52	232
1997		15	61	14	227
1998	30	54	62	67	404
1999		36	110	34	19
2000	1	105	198	31	162
2001		185	83	47	39
2002	201	219	145	183	67
2003	333	89	353	252	48
2004	283	296	1384	48	55
2005	130	334	1297	98	475
2006	373	1071	837	305	318
2007	827	772	705	125	478
2008	350	654	1054	177	528
2009	1150	546	2281	123	505

❖ *Μέσος όρος clustering coefficient ανά τμήμα και χρονιά*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
1992		1	1		
1993		0,8667	0,9333	0	0,9143
1994		1	0,9104	0	0,8911
1995		0,8955	0,8838	0,9403	0,9571
1996		1	0,9333	0,8677	0,8953
1997		1	0,8589	1	0,918
1998	0,9495	0,9183	0,9129	0,7586	0,9195
1999		0,8962	0,8879	0,7778	0,7667
2000	0	0,799	0,8934	0,8333	0,934
2001		0,8953	0,7895	0,8881	0,9398
2002	0,9263	0,8494	0,862	0,8838	0,8825
2003	0,9766	0,9325	0,9223	0,884	0,8917
2004	0,9765	0,8873	0,9359	0,8387	0,92
2005	0,9299	0,8751	0,9169	0,8431	0,9317
2006	0,918	0,8928	0,8655	0,9188	0,8716
2007	0,953	0,8884	0,9034	0,8537	0,8957
2008	0,9253	0,9056	0,9113	0,8828	0,9342
2009	0,9385	0,8846	0,9287	0,8462	0,8936

❖ *Αριθμός επιστημόνων ανά άρθρο*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
1992		3	6		
1993		3	4	2	4
1994		4	6	2	4
1995		5	4	3	4
1996		3	3	2	4
1997		3	2	3	3
1998	6	4	3	2	5
1999		3	3	2	3
2000	2	3	3,5	2	3,5
2001		4	2	2	3
2002	11	3,5	3	3	3
2003	7	4	3	3	3
2004	6	4	3	2	3
2005	5	4	3	3	3
2006	5,5	6	3	3	3
2007	5	4	3,5	2	4
2008	5	3	4	3	3
2009	6	3	5	2	4

❖ *Αριθμός άρθρων ανά επιστήμονα του Πανεπιστημίου Κύπρου*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
1992		1	1		
1993		2	1,5	1	2
1994		1	2	3	4
1995		2	2	1,5	3
1996		1	1	2,5	4
1997		1	2	1	2
1998	2	1	1	2	4
1999		2	1	2	2
2000	1	2	1	2	1
2001		2	1	1,5	1
2002	2	2	1	2	3
2003	2	1	1,5	1,5	2
2004	1	2	1	1	1,5
2005	2	2	2	2	2
2006	2	2	2	1	1
2007	1	2	2	1	2
2008	1,5	2	2	2	1
2009	2	1	2	2	2

❖ *Αριθμός εσωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
1992		0	2		
1993		0	1	0	0
1994		0	2	0	1
1995		3	1	2	3
1996		0	2	1	3
1997		2	1	0,5	1
1998	0	1	1	0	1
1999		1	1	1	1,5
2000	0	1	2	0	1
2001		2	2	0,5	2
2002	2	2	1	1	1
2003	2	2	2	2	1
2004	2	2	1	1	1
2005	2	2	2	1	2
2006	8	3	2	2	2
2007	2	2,5	2	2	2
2008	4	2	3	1	2
2009	3	2	2	1	2

❖ *Αριθμός εξωτερικών συνεργατών επιστημόνων του Πανεπιστημίου Κύπρου*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
1992		2	3		
1993		4	3,5	1	6
1994		3	2	2	4
1995		5,5	3,5	2,5	6
1996		2	2	7	6
1997		2	1	2	2
1998	10	7	1	2	5
1999		2	1,5	2	1,5
2000	1	2	2	3	1
2001		4	1	2	2
2002	11	3	2	2	0
2003	16	2	2	1,5	1
2004	3	2	1	1	3
2005	5	4	2	1	0
2006	6	3	2	3	4
2007	8	2	2	1	4
2008	5	2	3	2	4
2009	7	2	4	1	3,5

❖ *Degree Distribution*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
1992		2	5		
1993		2	4	1	3
1994		3	5	1	3
1995		5	5,5	8	26
1996		2	2	4	8
1997		4	2	2	5,5
1998	5	4	2	2,5	6
1999		3	2	2	2
2000	1	3	3	2	11
2001		5	2,5	2	4
2002	10	5	3	3	3
2003	23	3	4	5	2
2004	7	4	6	2	3
2005	5	5	5	2	6
2006	10	7	5	5	4
2007	12	6	4	2	5
2008	8	6	5	3	9
2009	9	5	7	2	7

❖ *Giant component (αριθμός κόμβων-μέσος όρος κόμβων) τμήματος Βιολογίας*

	Αριθμός κόμβων σε όλο το γράφο	Αριθμός κόμβων στον gc γράφο	Μέσος όρος κόμβων του gc γράφου	Αριθμός επιστημόνων του Πανεπιστημίου Κύπρου στον gc γράφο
1992				
1993				
1994				
1995				
1996				
1997				
1998	11	11		1
1999				
2000	2	2		1
2001				
2002	32	22	0,6875	4
2003	41	26	0,6341	3
2004	70	45	0,6429	3
2005	43	11	0,2558	4
2006	58	37	0,6379	11
2007	92	58	0,6304	3
2008	70	37	0,5286	8
2009	165	79	0,4788	13

❖ *Giant component (αριθμός κόμβων-μέσος όρος κόμβων) τμήματος Χημείας*

	Αριθμός κόμβων σε όλο το γράφο	Αριθμός κόμβων στον gc γράφο	Μέσος όρος κόμβων του gc γράφου	Αριθμός επιστημόνων του Πανεπιστημίου Κύπρου στον gc γράφο
1992	3	3	1	1
1993	5	5	1	1
1994	4	4	1	1
1995	24	20	0,8333	6
1996	3	3	1	1
1997	10	5	0,5	3
1998	20	10	0,5	1
1999	25	17	0,68	3
2000	29	11	0,3793	5
2001	62	15	0,2419	6
2002	58	17	0,2931	4
2003	47	14	0,2979	2
2004	83	31	0,3735	6
2005	104	33	0,3173	8
2006	190	116	0,6105	33
2007	176	81	0,4602	38
2008	166	74	0,4458	12
2009	142	67	0,4718	10

❖ *Giant component (αριθμός κόμβων-μέσος όρος κόμβων) τμήματος Πληροφορικής*

	Αριθμός κόμβων σε όλο το γράφο	Αριθμός κόμβων στον gc γράφο	Μέσος όρος κόμβων του gc γράφου	Αριθμός επιστημόνων του Πανεπιστημίου Κύπρου στον gc γράφο
1992	6	6	1	3
1993	10	5	0,5	2
1994	35	14	0,4	2
1995	65	40	0,6154	5
1996	13	6	0,4615	3
1997	40	9	0,225	1
1998	38	10	0,2632	2
1999	54	14	0,2593	2
2000	77	33	0,4286	15
2001	53	16	0,3019	7
2002	70	29	0,4143	10
2003	119	37	0,3109	9
2004	210	45	0,2143	3
2005	212	78	0,3679	35
2006	219	112	0,5114	47
2007	206	106	0,5146	52

2008	262	142	0,542	44
2009	389	75	0,1928	17

❖ *Giant component (αριθμός κόμβων-μέσος όρος κόμβων)*
τμήματος Μαθηματικών και Στατιστικής

	Αριθμός κόμβων σε όλο το γράφο	Αριθμός κόμβων στον gc γράφο	Μέσος όρος κόμβων του gc γράφου	Αριθμός επιστημόνων του Πανεπιστημίου Κύπρου στον gc γράφο
1992				
1993	4	2	0,5	1
1994	6	3	0,5	1
1995	30	19	0,6333	3
1996	22	13	0,5909	3
1997	11	4	0,3636	2
1998	37	11	0,2973	3
1999	27	7	0,2593	2
2000	27	5	0,1852	2
2001	40	9	0,225	3
2002	57	14	0,2456	3
2003	68	20	0,2941	8
2004	44	7	0,1591	3
2005	53	8	0,1509	4
2006	94	28	0,2979	16
2007	76	20	0,2632	7
2008	89	25	0,2809	7
2009	80	13	0,1625	5

❖ *Giant component (αριθμός κόμβων-μέσος όρος κόμβων)*
τμήματος Φυσικής

	Αριθμός κόμβων σε όλο το γράφο	Αριθμός κόμβων στον gc γράφο	Μέσος όρος κόμβων του gc γράφου	Αριθμός επιστημόνων του Πανεπιστημίου Κύπρου στον gc γράφο
1992				
1993	7	7	1	1
1994	29	10	0,3448	1
1995	105	53	0,5048	1
1996	44	23	0,5227	6
1997	59	22	0,3729	1
1998	64	29	0,4531	2
1999	15	6	0,4	2
2000	38	23	0,6053	1
2001	21	8	0,381	2
2002	30	10	0,3333	1

2003	23	9	0,3913	1
2004	28	8	0,2857	2
2005	76	37	0,4868	3
2006	75	38	0,5067	13
2007	117	34	0,2906	3
2008	99	36	0,3636	9
2009	110	26	0,2364	5

Παράρτημα Γ

Ακολουθούν αναλυτικά τα αποτελέσματα ανάλυσης των πέντε δικτύων συνεργασίας ιδρυμάτων-οργανισμών (Βιολογίας, Χημείας, Πληροφορικής, Μαθηματικών και Στατιστικής, Φυσικής).

❖ *Συνολικός αριθμός ιδρυμάτων-οργανισμών και άρθρων*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
Αριθμός Ιδρυμάτων-οργανισμών	185	204	473	204	223
Αριθμός άρθρων	114	487	777	386	367

❖ *Αριθμός ακμών γράφου ιδρυμάτων-οργανισμών*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
Αριθμός ακμών	1211	992	3534	722	1140

❖ *Μέσος όρος clustering coefficient*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
Μέσος όρος clustering coefficient	0,9243	0,8659	0,8793	0,9082	0,88

❖ *Αριθμός άρθρων για κάθε ίδρυμα και αριθμός ιδρυμάτων για κάθε άρθρο (διάμεση τιμή)*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική	
Αριθμός ιδρυμάτων για κάθε άρθρο(διάμεση τιμή)		3	2	2	2	2
Αριθμός άρθρων για κάθε ίδρυμα(διάμεση τιμή)		1	1	1	1	1

❖ *Degree Distribution*

	Βιολογία	Χημεία	Πληροφορική	Μαθηματικά	Φυσική
Degree Distribution	5	3	5	2	4