



ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ στα
ΠΟΛΥΠΛΟΚΑ ΣΥΣΤΗΜΑΤΑ και ΔΙΚΤΥΑ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ
ΤΜΗΜΑ ΓΕΩΛΟΓΙΑΣ
ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ



ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Τίτλος Εργασίας

Διαχρονική Ανάλυση Οικονομικών Δικτύων και
Δικτύων Εμπορίου

Temporal Analysis of Financial and Trade Networks

Παναγιώτης Α. Παναγιώτου

ΕΠΙΒΛΕΠΩΝ: Νικόλαος Βαρσακέλης, Αν. Καθηγητής Α.Π.Θ.

ΣΥΝΕΠΙΒΛΕΠΩΝ: Ιωάννης Αντωνίου, Καθηγητής Α.Π.Θ.

Θεσσαλονίκη, Σεπτέμβριος 2017









Ψηφιακή συλλογή
Διαχρονική Ανάλυση Οικονομικών Δικτύων και Δικτύων Εμπορίου
ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ στα
ΠΟΛΥΠΛΟΚΑ ΣΥΣΤΗΜΑΤΑ και ΔΙΚΤΥΑ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ
ΤΜΗΜΑ ΓΕΩΛΟΓΙΑΣ
ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Τίτλος Εργασίας

Διαχρονική Ανάλυση Οικονομικών Δικτύων και
Δικτύων Εμπορίου

Temporal Analysis of Financial and Trade Networks

Παναγιώτης Α. Παναγιώτου

ΕΠΙΒΛΕΠΩΝ: Νικόλαος Βαρσακέλης
Αν. Καθηγητής Α.Π.Θ.

Εγκρίθηκε από την Τριμελή Εξεταστική Επιτροπή την 12η Σεπτεμβρίου 2017.

.....
Ν. Βαρσακέλης
Αν. Καθηγητής Α.Π.Θ.

.....
Ι. Αντωνίου
Αν. Καθηγητής Α.Π.Θ.

.....
Β. Καραγιάννης
Διδάκτορας Α.Π.Θ.

Θεσσαλονίκη , Σεπτέμβριος 2017



.....
Παναγιώτης Α. Παναγιώτου
Πτυχιούχος Ηλεκτρολόγος Μηχανικός, Πανεπιστήμιο Πατρών
Μεταπτυχιακός Φοιτητής, Πολύπλοκα Συστήματα & Δίκτυα, Α.Π.Θ.

Copyright © Παναγιώτης Α. Παναγιώτου, 2017
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευτεί ότι εκφράζουν τις επίσημες θέσεις του Α.Π.Θ.

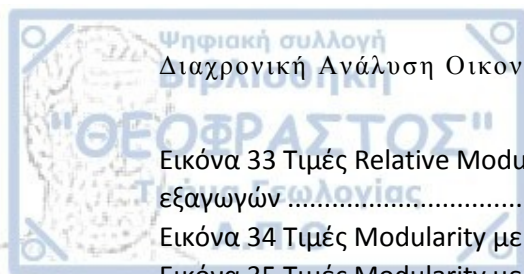


Περιεχόμενα

Ευρετήριο Εικόνων.....	6
Ευρετήριο Πινάκων.....	7
ΠΕΡΙΛΗΨΗ.....	8
ABSTRACT.....	9
ΕΥΧΑΡΙΣΤΙΕΣ.....	10
1. Εισαγωγή.....	11
2. Μεθοδολογία.....	14
2.1 Modularity & Nestedness σε διμερή δίκτυα.....	15
2.1.1 Modularity.....	15
2.2 Nestedness.....	17
2.3 Modularity Detection Algorithms.....	18
2.3.1 Newman’s “Leading Eigenvector”.....	18
2.3.2 Ο αλγόριθμος Fast-Greedy.....	23
2.3.4 Barber’s “Adaptive BRIM”.....	24
2.3.5 “Label Propagation” & Murata’s “LP&BRIM”.....	27
2.3.6 “Label Propagation”.....	27
2.3.7 Murata’s “LP&BRIM”.....	29
2.4 Nestedness Detection Algorithms.....	31
2.4.1 NTC (Nestedness Temperature Calculator).....	31
2.4.3 NODF.....	32
2.5 Ερμηνεία Nestedness και Modularity.....	33
2.5 Σχηματική Περίληψη της μεθοδολογίας.....	35
3. Ανάλυση.....	36
3.1 Προέλευση και παρουσίαση των Data Set.....	36
3.2 Διαχρονικά Δίκτυα μετοχών του χρηματιστηρίου (S&P500).....	39
3.3 Δίκτυα Παγκόσμιου Εμπορίου.....	50
3.3.1 Δίκτυα Εξαγωγών (Exports).....	51
3.3.2 Τιμές Relative Modularity (Q_r) για τα δίκτυα εξαγωγών (exports).....	54
3.3.3 Δίκτυα Εισαγωγών (Imports).....	58
3.3.4 Τιμές Relative modularity (Q_r) για τα δίκτυα εισαγωγών (imports).....	61
3.3.5 Τιμές Nestedness για τα δίκτυα εξαγωγών.....	65
3.3.6 Τιμές Nestedness για τα δίκτυα εισαγωγών.....	66
4. Συμπεράσματα.....	68
Παράρτημα.....	69
Βιβλιογραφία.....	76

Ευρετήριο Εικόνων

Εικόνα 1 Το δίκτυο που δημιουργείται κατά την εξάπλωση μιας μεταδοτικής νόσου.....	11
Εικόνα 2 Δίκτυο επαφών από τον ιστότοπο του myspace	12
Εικόνα 3 Το δίκτυο του συστήματος λειτουργίας της αεροπορικής εταιρίας Lufthansa	12
Εικόνα 4 Διμερές δίκτυο "φυτών-επικοινωνιών" [51]. Στο συγκεκριμένο δίκτυο υπάρχει μια εμφανής τάση για δημιουργία συστάδων (clusters) και στα δύο σύνολα κόμβων.	13
Εικόνα 5 Σε πολλά δίκτυα δημιουργούνται ομάδες κόμβων ή αλλιώς communities , ήτοι σύνολα κορυφών (με σκίαση) μέσα στα οποία υπάρχει μεγάλη συνδετικότητα και μόνο μικρός αριθμός από ακμές μεταξύ κορυφών που βρίσκονται σε διαφορετικές ομάδες. [31]	15
Εικόνα 6 Παραδείγματα Εμφωλευμένων δομών και εικονική αναπαράσταση της μήτρας εμφωλευσιμότητας (nestedness matrix)	18
Εικόνα 7 Διάγραμμα ροής του αλγορίθμου Adaptive Brim [27]	26
Εικόνα 8 Δόμηση μίας κοινότητας με τον αλγόριθμο LP. Στο συγκεκριμένο παράδειγμα, οι ετικέτες διαδίδονται σύγχρονα. Μετά από τρεις επαναλήψεις ο αλγόριθμος έχει δομήσει από το πυκνά συνδεδεμένο γκρουπ τεσσάρων κόμβων μία κοινότητα με την ετικέτα "{c}" [27]	28
Εικόνα 9 Το φαινόμενο της ταλάντωσης των ετικετών λαμβάνει χώρα, όταν ο αλγόριθμος LP με synchronous updating εφαρμόζεται σε διμερή δίκτυα [27].....	28
Εικόνα 10 Αναπαράσταση του asynchronous updating. Σε κάθε επανάληψη, η σκίαση με κίτρινο δείχνει τον κόμβο του οποίου η ετικέτα ανανεώνεται [27].....	28
Εικόνα 11 Διάγραμμα ροής του αλγορίθμου LP&BRIM [27]	29
Εικόνα 12 Σχηματική αναπαράσταση του αλγορίθμου LP&BRIM [27].....	30
Εικόνα 13 Ο αλγόριθμος LP&BRIM σε μορφή ψευδοκώδικα [27]	30
Εικόνα 14 Unexpectedness: η παρουσία κάτω από την isocline/απουσία πάνω από την isocline [72]	31
Εικόνα 15 Modularity και Nestedness διαφόρων δικτύων.....	34
Εικόνα 16 Nested & Modular δίκτυο	34
Εικόνα 17 Σχηματική αναπαράσταση της προτεινόμενης μεθοδολογίας.....	35
Εικόνα 18 Πορεία του δείκτη S&P500 στο χρόνο	40
Εικόνα 19 Τιμές modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα μετοχών .	41
Εικόνα 20 Τιμές modularity με τον αλγόριθμο fast-greedy για τα δίκτυα μετοχών	42
Εικόνα 21 Τιμές modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα μετοχών.....	43
Εικόνα 22 Τιμές Qr με τον αλγόριθμο Leading Eigenvector για τα δίκτυα μετοχών	45
Εικόνα 23 Τιμές Qr με τον αλγόριθμο LP&BRIM για τα δίκτυα μετοχών	46
Εικόνα 24 Τιμές Qr με τον αλγόριθμο fast-greedy για τα δίκτυα μετοχών.....	47
Εικόνα 25 Τιμές Nestedness με τον αλγόριθμο NTC για τα δίκτυα μετοχών	48
Εικόνα 26 NTC nested sorting	48
Εικόνα 27 Τιμές Nestedness με τον αλγόριθμο NODF για τα δίκτυα μετοχών	49
Εικόνα 28 Τιμές Modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα εξαγωγών	51
Εικόνα 29 Τιμές Modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα εξαγωγών.....	52
Εικόνα 30 Τιμές Modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εξαγωγών.....	53
Εικόνα 31 Τιμές Relative Modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα εξαγωγών	54
Εικόνα 32 Τιμές Relative Modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα εξαγωγών..	56



Εικόνα 33 Τιμές Relative Modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εισαγωγών	57
Εικόνα 34 Τιμές Modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εισαγωγών	58
Εικόνα 35 Τιμές Modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα εισαγωγών	59
Εικόνα 36 Τιμές Modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εισαγωγών	60
Εικόνα 37 Τιμές Relative Modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα εισαγωγών.....	62
Εικόνα 38 Τιμές Relative Modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα εισαγωγών	63
Εικόνα 39 Τιμές Relative Modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εισαγωγών.....	64
Εικόνα 40 Τιμές Nestedness με τους αλγορίθμους NODF&NTC για τα δίκτυα εισαγωγών ...	65
Εικόνα 41 Τιμές Nestedness με τους αλγορίθμους NODF&NTC για τα δίκτυα εισαγωγών....	66

Ευρετήριο Πινάκων

Πίνακας 1 Προϊόντα στο σύστημα ταξινόμησης HS.....	38
Πίνακας 2 Πυκνότητα δικτύου ανά έτος.....	40
Πίνακας 3 Τιμές modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα μετοχών .	41
Πίνακας 4 Τιμές modularity με τον αλγόριθμο fast-greedy για τα δίκτυα μετοχών	42
Πίνακας 5 Τιμές modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα μετοχών	43
Πίνακας 6 Τιμές Qr με τον αλγόριθμο LP&BRIM για τα δίκτυα μετοχών.....	46
Πίνακας 7 Τιμές Qr με τον αλγόριθμο fast-greedy για τα δίκτυα μετοχών.....	47
Πίνακας 8 Τιμές Nestedness NTC & NODF.....	49
Πίνακας 9 Οι πυκνότητες των δικτύων για εξαγωγές (exports) και εισαγωγές (imports)	51
Πίνακας 10 Τιμές modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα εισαγωγών	52
Πίνακας 11 Τιμές modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα εισαγωγών.....	53
Πίνακας 12 Τιμές modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εισαγωγών.....	54
Πίνακας 13 Τιμές relative modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα εισαγωγών	55
Πίνακας 14 Τιμές relative modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα εισαγωγών.....	56
Πίνακας 15 Τιμές relative modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εισαγωγών	57
Πίνακας 16 Τιμές modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα εισαγωγών.....	59
Πίνακας 17 Τιμές modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα εισαγωγών	60
Πίνακας 18 Τιμές modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εισαγωγών.....	61
Πίνακας 19 Τιμές relative modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα εισαγωγών.....	62
Πίνακας 20 Τιμές relative modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα εισαγωγών.....	63
Πίνακας 21 Τιμές relative modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εισαγωγών.....	64
Πίνακας 22 Τιμές nestedness με τους αλγορίθμους NODF & NTC για τα δίκτυα εισαγωγών ..	65
Πίνακας 23 Τιμές nestedness με τον αλγόριθμο NTC για τα δίκτυα εισαγωγών	66
Πίνακας 24 Τιμές nestedness με τον αλγόριθμο NODF για τα δίκτυα εισαγωγών	67

ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία πραγματεύεται τη διαχρονική ανάλυση δικτύων του χρηματιστηρίου και δικτύων του παγκόσμιου εμπορίου, κάνοντας χρήση αλγορίθμων που στοχεύουν στην ανίχνευση και αξιολόγηση κοινοτήτων. Επιπλέον, διερευνάται στα εν λόγω δίκτυα η ύπαρξη εμφωλευσιμότητας αναφορικά με τη δομή τους. Πιο συγκεκριμένα, εξετάζεται η επίδραση χρηματοοικονομικών εταιριών σε μετοχές άλλων κλάδων του αμερικανικού χρηματιστηρίου με τη βοήθεια των βιολογικών host-parasite δικτύων (διμερή δίκτυα), στα οποία οι χρηματοοικονομικές έχουν τον παρασιτικό ρόλο ενώ οι υπόλοιπες διαδραματίζουν το ρόλο του ξενιστή (host). Εξετάζοντας την εξέλιξη της τοπολογίας του δικτύου από το 2000 έως το 2015, υπολογίζεται η εμφωλευσιμότητα και η κοινοτικοποίηση ή δομικότητα, που στην έκταση του κειμένου απαντώνται ως nestedness και modularity αντίστοιχα. Κατά πλήρη αντιστοιχία, η ανάλυση εφαρμόζεται και σε δίκτυα του παγκόσμιου εμπορίου, με τα προϊόντα να λειτουργούν ως παράσιτα και τις χώρες που τα εισάγουν (ή εξάγουν) ως ξενιστές. Τα αποτελέσματα δείχνουν ότι το δίκτυο γίνεται nested και anti-modular και, επομένως, ιδιαίτερα εύαλωτο σε διάδοση κρίσης αρκετά πριν το 2008. Παρόμοια συμπεριφορά δεν παρατηρείται κατά τη διάρκεια της κρίσης dot com, η οποία προηγείται. Αυτό αποτελεί επιπλέον θετικό στοιχείο για την καταλληλότητα της μεθόδου να εντοπίζει τη δυνατότητα διάδοσης κρίσης προερχόμενης από συγκεκριμένους κλάδους της οικονομίας.

Για την ανάλυση των δεδομένων ως προς τη δόμηση κοινοτήτων χρησιμοποιήθηκαν κάποιοι από τους αλγορίθμους, που είναι γνωστοί από τη βιβλιογραφία ως modularity detection algorithms. Αρχικά χρησιμοποιείται ο αλγόριθμος Leading Eigenvector του M.J. Newman, που βασίζεται στο οδηγό ιδιοδιάνυσμα της μήτρας γειτονικότητας. Έπειτα, εφαρμόζεται ο αλγόριθμος fast-greedy και αναλύεται ο αλγόριθμος Adaptive BRIM (Bipartite Recursive Identification of Modules) του M. Barber, που βασίζεται σε επαναληπτική μέθοδο για την αξιολόγηση μιας κοινότητας, αφού αυτή βρεθεί. Τέλος, γίνεται χρήση του LP&BRIM algorithm (Label Propagation με εξακρίβωση BRIM), που προτάθηκε ως συνδυασμός αλγορίθμων από τον P. Murata.

Η ύπαρξη εμφωλευμένων δομών και ο βαθμός τους διερευνήθηκαν χρησιμοποιώντας δύο μέτρα: NTC (Nestedness Temperature Calculator) και NODF (Nestedness based on Overlap and Decerasing Fill). Το πρώτο μέτρο θεμελιώθηκε από το μοτίβο, σύμφωνα με το οποίο τα νησιά δέχονται αποίκους από τη στεριά και βασίζεται στην αναδιάταξη του πίνακα εμφωλευσιμότητας, ενώ το δεύτερο αποτελεί μέτρο που προκύπτει από σύγκριση στηλών της μήτρας γειτονικότητας, έως ότου υπολογιστεί το ποσοστό επικάλυψης στοιχείων μετά από φθίνουσα ταξινόμηση.

Για τη δημιουργία των δικτύων και την ανάλυση των δεδομένων χρησιμοποιήθηκε το λογισμικό R Studio και το πακέτο Falcon αυτού, καθώς και το πακέτο BiMat του λογισμικού MatLab για ανάλυση διμερών δικτύων.

Λέξεις-Κλειδιά: δίκτυα χρηματιστηρίου, nestedness, modularity, χρηματιστηριακή κρίση, δίκτυα εμπορίου



ABSTRACT

The present thesis deals with a temporal analysis of financial networks and world trade networks with the use of algorithms aiming to community detection and evaluation. Moreover, the property of nestedness in regards with the networks' structure is investigated for the aforementioned networks. More specifically, the effect of financial firms on stocks of other stock market sectors is examined with the help of host-parasite networks (bipartite networks), in which the financials act as parasites while the rest stocks have the roll of the host. By examining the evolution of the networks' topology from 2000 until 2015, the values of nestedness and modularity are calculated. The analysis is implemented for networks emerging from international world trade data, with the trade products acting as parasites and the countries importing (or exporting) them acting as hosts. The results indicate that the networks become nested and anti-modular before 2008. This behaviour is not observed during the pre-occurring dot com crisis. This is additional positive evidence for the functionality of the method to detect a potential outraging crisis in specific economic sectors of the stock market.

For the analysis of the data some algorithms known from the literature as community detection algorithms are used. First and foremost, the Leading Eigenvector algorithm of M.J. Newman is used. This algorithm is based on the leading eigenvector of the adjacency matrix. Secondly, the fast-greedy algorithm is used for the detection and evaluation of communities. Then, the Adaptive BRIM algorithm (Bipartite Recursive Identification of Modules) is described and analysed for a deeper insight in the community detection procedure for bipartite networks. Finally, the LP&BRIM algorithm (Label Propagation & BRIM) is implemented on the dataset. This is an algorithm suggested by P. Murata, using the Label Propagation algorithm for community detection and combining it with Barber's Adaptive BRIM for the evaluation and verification.

For investigating the existence of nested structures in the temporal networks, two nestedness metrics are used: NTC (Nestedness Temperature Calculator) and NODF (Nestedness based on Overlap and Decreasing Fill). The first one is a metric emerging from the patterns according to which, islands are colonized by species and it is based on the rearrangement of the nestedness matrix. On the other hand, NODF is based on the rearrangement and comparison of the adjacency matrix's elements until the percentage of decreasing fill is calculated.

For the analysis of the data and the creation of the corresponding networks the statistical software of RStudio is used, as well as the package RFalcon which is used for analysing bipartite networks for community detection and nested structures. Moreover, the package BiMat of MatLab software was taken advantage of, for calculation and visualization of nested structures and some meta-statistics evaluation. The BiMat package, is a special package used for the analysis of bipartite networks.

Key-Words: stock market networks, nestedness, modularity, stock market crisis, trade networks

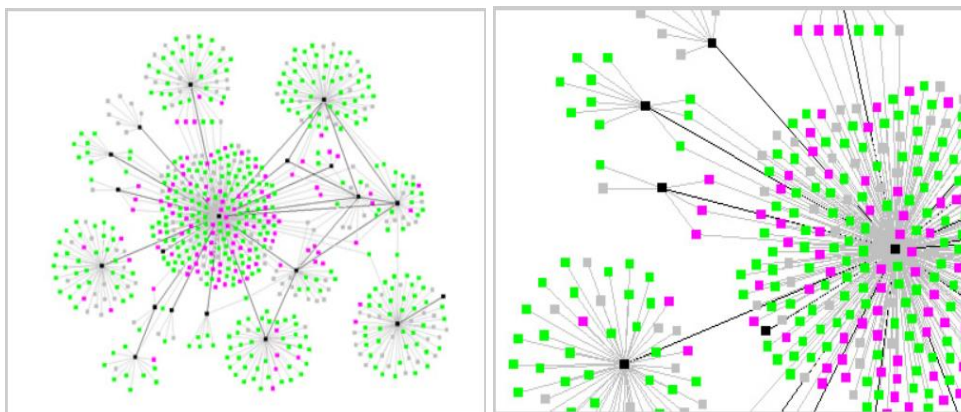


ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω τα μέλη της τριμελούς επιτροπής και καθηγητές μου στο ΔΠΜΣ-ΠΣΔ για την καθοδήγηση και υποστήριξη τους σε όλο τον κύκλο των μεταπτυχιακών σπουδών, καθώς οι συμβουλές και διδαχές τους οδήγησαν στην εξέλιξη της πορείας μου και των γνώσεών μου. Ευχαριστώ τους γονείς μου, Αχιλλέα και Ευαγγελία, για τη στήριξη τους κατά τη διάρκεια του μεταπτυχιακού αλλά και όλων των σπουδών μου. Επίσης, ευχαριστώ τα ξαδέλφια μου Βασίλη, Άννα και Νίκο για την πολύτιμη βοήθειά τους, όπου αυτή χρειάστηκε. Τέλος, οφείλω να ευχαριστήσω τους φίλους και συναδέλφους Προκόπιο Καραδήμο, Αστέριο Μπατζιάκα, Δημοσθένη Κούλικα και Σάββα Παραγκαμιάν των οποίων οι συμβουλές και παρατηρήσεις κατά τη διάρκεια της κοινής μας πορείας στον κύκλο του ΔΠΜΣ αποδείχθηκαν σημαντικές μαζί με την ανταλλαγή απόψεων και ιδεών με στόχο την από κοινού εξέλιξη των γνώσεών μας και της εμπειρίας μας.

1. Εισαγωγή

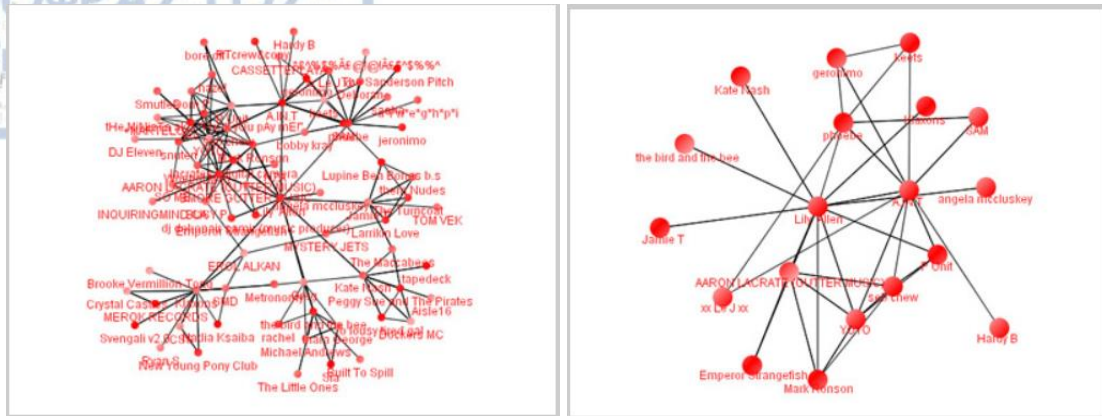
Η επιστήμη των δικτύων (network science) έχει συνεισφέρει σημαντικά την τελευταία δεκαετία στην πρόοδο που σημειώνεται σχετικά με τη μοντελοποίηση και την κατανόηση των πολύπλοκων συστημάτων. Η νέα αυτή προσέγγιση βρίσκει εφαρμογές σε αρκετά πεδία των φυσικών, κοινωνικών και τεχνολογικών επιστημών [15]-[46],[47]-[49]. Ο λόγος είναι ότι κάθε σύστημα αποτελείται από επιμέρους στοιχεία, που καλούνται Δομικές Μονάδες ή Δομικά Στοιχεία (components), και από τις μεταξύ τους αλληλεπιδράσεις. Τα συστήματα αυτά αναπαρίστανται ως δίκτυα αποτελούμενα από κόμβους που συνδέονται μεταξύ τους με ακμές. Οι κόμβοι του δικτύου απεικονίζουν τα αλληλεπιδρώντα Δομικά Στοιχεία του Συστήματος, ενώ οι ακμές τη μεταξύ τους αλληλεπίδραση [7],[17],[18]. Στις εικόνες 1-3 παρουσιάζονται παραδείγματα σύγχρονων δικτύων που έχουν προκύψει από ανάλυση πραγματικών δεδομένων για βιολογικά δίκτυα (Εικόνα 1), κοινωνικά δίκτυα (Εικόνα 2) και οικονομικά δίκτυα επιχειρήσεων (Εικόνα 3).



Εικόνα 1 Το δίκτυο που δημιουργείται κατά την εξάπλωση μιας μεταδοτικής νόσου

Το δίκτυο περιγράφει τις επιρροές που ασκούνται μεταξύ ανθρώπινων κοινοτήτων κατά την έξαρση μιας μεταδοτικής νόσου. Οι κόμβοι είναι οι άνθρωποι που επηρέασε. Τα χρώματα των κόμβων δείχνουν το ρόλο του κάθε ανθρώπου ως μονάδας στη διαδικασία μετάδοσης. Με μαύρο είναι τα άτομα που έχουν ήδη μια κλινική νόσο (και πιο πιθανοί για μόλυνση), με πράσινο οι άνθρωποι που έχουν εκτεθεί στη νόσο αλλά δεν έχουν μολυνθεί, με μωβ απεικονίζονται τα άτομα που έχουν μολυνθεί και μεταδίδουν τη νόσο, με γκρι είναι οι άνθρωποι της κοινότητας που η κατάστασή τους ως προς τον ιό είναι ακόμα άγνωστη.

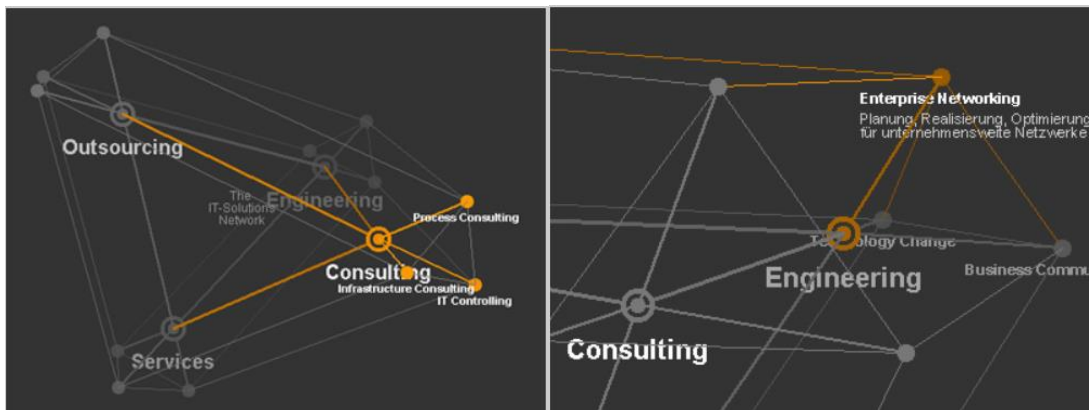
Πηγή: <http://www.visualcomplexity.com>



Εικόνα 2 Δίκτυο επαφών από τον ιστότοπο του myspace

Οι κορυφές του δικτύου παριστάνουν τα προφίλ των χρηστών του myspace και οι ακμές την ύπαρξη σύνδεσης, δηλαδή την ύπαρξη φίλας-επαφής στο myspace. Ο δημιουργός του δικτύου παρέχει πλέον σε ιστότοπο εφαρμογή για εποπτεία του προσωπικού δικτύου κάθε χρήστη σε πραγματικό χρόνο για το δίκτυο των επαφών του.

Πηγή: <http://www.visualcomplexity.com>



Εικόνα 3 Το δίκτυο του συστήματος λειτουργίας της αεροπορικής εταιρίας Lufthansa

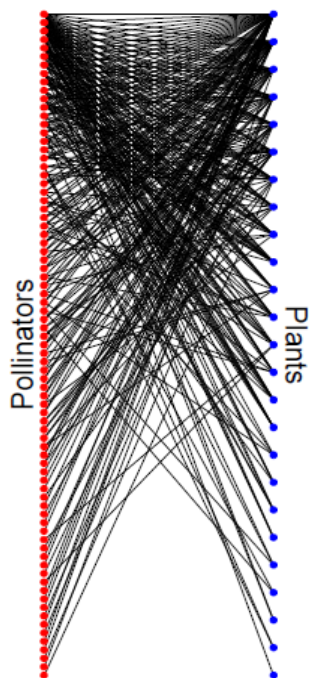
Το δίκτυο απεικονίζει τη διασύνδεση των τμημάτων λειτουργίας της εταιρίας και παρέχει σύγχρονες υπηρεσίες για τους πελάτες της ανά τον κόσμο. Οι κόμβοι παριστάνουν τα τμήματα της εταιρίας και οι συνδέσεις την μεταξύ τους αλληλεπίδραση, σύμφωνα με την οποία οι χρήστες μπορούν να προσπελάσουν πολύπλοκα ζητήματα μέσα σε μια απλή ιεραρχία πληροφορίας. Τα διαφορετικά χρώματα και επίπεδα του 3-D μοντέλου δηλώνουν αυτή την ιεραρχία. Η εφαρμογή παρέχεται στους πελάτες μετά τη δημιουργία της από τον Cerative Director, Jeremy Abbett.

Πηγή: <http://www.visualcomplexity.com>

Ένα δίκτυο χαρακτηρίζεται ως μονομερές (unipartite), όταν κάθε κόμβος μπορεί δυνητικά να συνδεθεί με οποιοδήποτε άλλο κόμβο. Παραδείγματα τέτοιων δικτύων υπάρχουν σε βιολογικά δίκτυα μεταβολισμού [19], σε οικολογικά δίκτυα τροφικών αλυσίδων [20],[21] ή κατά τη σύναψη φιλίας/δημιουργία επαφής σε κοινωνικά δίκτυα [22]. Οι αλληλεπιδράσεις μεταξύ των κόμβων σε τέτοια δίκτυα είναι ισχυρά δομημένες, κάτι που σημαίνει ότι διαφέρουν από τα ιδεατά μοντέλα δικτύων στα οποία η πιθανότητα αλληλεπίδρασης μεταξύ δύο οποιωνδήποτε κόμβων είναι σταθερή (πχ Erdos-Renyi model) [23]. Τα δίκτυα των εικόνων 1-3 είναι μονομερή δίκτυα.

Μία άλλη κατηγορία δικτύων είναι τα διμερή δίκτυα, στα οποία οι κόμβοι χωρίζονται σε δύο διακριτά σύνολα (δύο διαφορετικές ομάδες). Έτσι, τυχόν αλληλεπιδράσεις μπορούν να υπάρξουν μόνο μεταξύ κόμβων διαφορετικού συνόλου, ενώ συνδέσεις μεταξύ κόμβων του ίδιου συνόλου απαγορεύονται [24]. Ένα από τα πιο συνηθισμένα παραδείγματα διμερούς δικτύου είναι τα δίκτυα “φυτών-επικονιαστών”: στο ένα σύνολο κόμβων προβάλλονται τα φυτά, ενώ στο άλλο οι επικονιαστές, και οι ακμές εκφράζουν την επικονίαση [21],[33]. Πράγματι, πληθώρα μελετών επικεντρώνεται στη μοντελοποίηση των διμερών δικτύων. Έτσι, αναδύθηκε εκτενής βιβλιογραφία, προσφέροντας τεχνικές για την ανάλυση των δικτύων “φυτών-επικονιαστών” [6], [8], [10],[25]-[30]. Παράλληλα, οι τεχνικές ανάλυσης και οι μέθοδοι που αναπτύχθηκαν για την ανάλυση τέτοιων δικτύων εφαρμόζονται πλέον και σε διμερή δίκτυα άλλων επιστημονικών πεδίων. Παραδείγματα από τη βιβλιογραφία αποτελούν τα

ανταγωνιστικά δίκτυα (“ξενιστές-παράσιτα”) [35],[36], [38],[43],[44], τα οικονομικά δίκτυα (δίκτυα χρηματιστηρίου) [32],[34] ή και κάποια κοινωνικά δίκτυα (δίκτυο δημοσιεύσεων επιστημονικών άρθρων σε περιοδικά) [25],[39]. Στην Εικόνα 4 παρουσιάζεται ένα κλασικό διμερές δίκτυο “φυτών-επικονιαστών”.



Εικόνα 4 Διμερές δίκτυο “φυτών-επικονιαστών” [51]. Στο συγκεκριμένο δίκτυο υπάρχει μια εμφανής τάση για δημιουργία συστάδων (clusters) και στα δύο σύνολα κόμβων.

2. Μεθοδολογία

Όπως ήδη αναφέρθηκε, υπάρχουν πολλές περιπτώσεις όπου τα διαθέσιμα πραγματικά δεδομένα (real world data) κωδικοποιούνται ως διμερή δίκτυα (bipartite networks) [4]-[9],[27]. Η ποσοτικοποίηση και ο υπολογισμός των ιδιοτήτων ενός διμερούς δικτύου μας οδηγεί σε ενδιαφέροντα συμπεράσματα τόσο για τη δομή όσο και για τη συμπεριφορά του δικτύου, ειδικά αν το δίκτυο εξετάζεται διαχρονικά. Η εκτίμηση της δομής διάφορων δικτύων οδήγησε τους ερευνητές στην ανεύρεση μέτρων για τις ιδιότητες των δικτύων όπως: modularity, small-world structure, hierarchy [18]. Κατά την απόπειρα να μετρηθούν αυτές οι έννοιες, συντάχθηκαν διάφοροι αλγόριθμοι με σκοπό την ποσοτικοποίηση και τον χαρακτηρισμό της δομής -κυρίως μονομερών-δικτύων [37],[40],[41],[45].

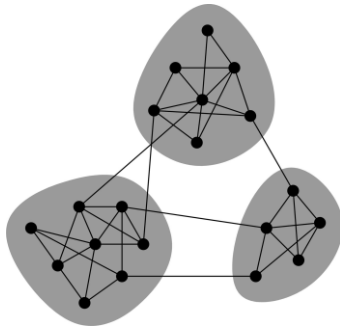
Μεγάλο ενδιαφέρον στην ανάλυση δικτύων παρουσιάζει η ανίχνευση κοινοτήτων (community detection) και η ανίχνευση εμφωλευμένων δομών (nested structures) μέσα σε ένα δεδομένο δίκτυο. Στην παρούσα εργασία μελετώνται οι δύο αυτές ιδιότητες των δικτύων: **a.** της δημιουργίας κοινοτήτων (**modularity**) και **b.** της εμφωλευσιμότητας (**nestedness**). Οι ιδιότητες αυτές απαντώνται τόσο σε μονομερή όσο και σε διμερή δίκτυα και είναι σημαντικές για το χαρακτηρισμό του δικτύου ποσοτικά και ποιοτικά [4],[14],[27],[33]. Σε επόμενες ενότητες (2.1 έως 2.4) περιγράφονται οι δύο αυτές έννοιες, καθώς και προσεγγίσεις που έχουν γίνει με στόχο την ανεύρεση μέτρων για τον ακριβή υπολογισμό τους. Επίσης, στις ενότητες αυτές γίνεται επισκόπηση και σύγκριση κάποιων αλγορίθμων που έχουν δομηθεί για την ανίχνευση των δύο αυτών ιδιοτήτων, εντοπίζοντας ποιοι από αυτούς χρησιμοποιούνται την τελευταία δεκαετία κατά κόρο και γιατί. Οι αλγόριθμοι που παρουσιάζονται για την ανίχνευση κοινοτήτων, ανήκουν στην κατηγορία των φασματικών αλγορίθμων (spectral algorithms). Οι αλγόριθμοι αυτοί βασίζονται στην αξιοποίηση των ιδιοτιμών και ιδιοδιανυσμάτων της μήτρας γειτνίασης (adjacency matrix), καθώς και σε ιδιότητες μητρών που προκύπτουν από αυτή. Οι αλγόριθμοι που παρουσιάζονται και χρησιμοποιούνται για ανίχνευση εμφωλευσιμότητας, είναι δύο αλγόριθμοι που αναπτύχθηκαν το 1993 και το 2009 αντίστοιχα. Τέλος, στην Ενότητα 3 εφαρμόζονται οι αλγόριθμοι αυτοί σε πραγματικά δεδομένα, όπου γίνεται ανάλυση δύο διαφορετικών ειδών διμερών δικτύων: δικτύων από τα χρηματοοικονομικά και δικτύων του παγκόσμιου εμπορίου.

Τα δεδομένα που αξιοποιήθηκαν για την μελέτη αυτή, υποβλήθηκαν σε αναλύσεις για ανίχνευση κοινοτήτων (modularity) με τρεις διαφορετικούς αλγορίθμους [7]: Leading Eigenvector [31], fast-greedy algorithm [6], LP&BRIM [27]. Για την ανίχνευση εμφωλευμένων δομών (nestedness) εφαρμόστηκαν οι αλγόριθμοι NestednessNODF [52] και NestednessNTC [53,54]. Η υλοποίηση των αλγορίθμων πραγματοποιήθηκε στο περιβάλλον του λογισμικών R και Matlab. Επίσης, αξιοποιήθηκε το πακέτο "BiMat Master", ένα ελεύθερο λογισμικό που δημιουργήθηκε το 2014 από ειδικούς αναλυτές δικτύων και δημοσιεύτηκε το 2015 για να διευκολύνει την οπτικοποίηση και τη μελέτη διμερών δικτύων [7], [8].

2.1 Modularity & Nestedness σε διμερή δίκτυα

2.1.1 Modularity

Έχει παρατηρηθεί ότι τα δίκτυα διαμερίζονται με φυσικό τρόπο σε *κοινότητες* (communities ή modules). Αυτό σημαίνει ότι δημιουργούνται ομάδες κόμβων που έχουν από κοινού κάποιο ρόλο μέσα στο δίκτυο ή –έστω– κάποια κοινή ιδιότητα. Οι συνδέσεις είναι αρκετά πυκνές μεταξύ των κόμβων της ομάδας και αρκετά αραιές μεταξύ των διάφορων ομάδων [4]-[10],[13]-[15],[27]- [31]. Μία τέτοια δομή δικτύου παρουσιάζεται στην Εικόνα 5. Η ιδιότητα αυτή των δικτύων απαντάται στη βιβλιογραφία ως δόμηση κοινοτήτων (*community structure*) (Girvan and Newman, 2002) ή συσταδοποίηση (*clustering*) (Danon et al., 2007; Fortunato and Castellano, 2009; Newman, 2004a; Porter et al., 2009; Schaefer, 2007).



Εικόνα 5 Σε πολλά δίκτυα δημιουργούνται ομάδες κόμβων ή αλλιώς communities, ήτοι σύνολα κορυφών (με σκίαση) μέσα στα οποία υπάρχει μεγάλη συνδετικότητα και μόνο μικρός αριθμός από ακμές μεταξύ κορυφών που βρίσκονται σε διαφορετικές ομάδες. [31]

Στη σύγχρονη κοινωνία απαντώνται πάρα πολλά δίκτυα που οργανώνονται σε κοινότητες: οικογένειες, κύκλοι φιλίας, δίκτυα εργασιακών σχέσεων, πόλεις και χωριά [15]. Η διάδοση του παγκόσμιου ιστού οδήγησε στη δημιουργία πραγματικών ομάδων, που υφίστανται σαν οντότητες μέσα στο Web. Συστηματική μελέτη έχει λάβει χώρα για τα κοινωνικά δίκτυα (Coleman, 1964; Freeman, 2004; Kottak, 2004; Moody and White, 2003). Κοινότητες, επίσης, εμφανίζονται σε πολλά συστήματα δικτύων από τη βιολογία, την επιστήμη των Η/Υ, την οικονομία, την πολιτική κ.ο.κ. [9],[15],[50]. Στα δίκτυα αλληλεπίδρασης πρωτεϊνών αναγνωρίζονται κοινότητες στις οποίες ανήκουν πρωτεΐνες που εκτελούν κοινές λειτουργίες μέσα στο κύτταρο (Chen and Yuan, 2006; Rives and Galitski, 2003; Spirin and Mirny, 2003). Στο δίκτυο του World Wide Web παρατηρείται σύσταση κοινοτήτων από ιστοσελίδες με κοινό ή σχετικά κοινό θέμα (Dourisboure et al., 2007; Flake et al., 2002). Στα δίκτυα μεταβολισμού δομούνται κοινότητες από τις οποίες εντοπίζονται κύκλοι ή μονοπάτια αλληλεπίδρασης (Gumerea and Amaral, 2005; Palla et al., 2005). Σε δίκτυα τροφικών αλυσίδων οι κοινότητες παρέχουν πληροφορία για τη διαμέριση των ειδών (“predator-prey” networks) (Krause et al., 2003; Pimm, 1979).

Η ανίχνευση κοινοτήτων είναι σημαντική και για άλλους λόγους. Η αναγνώριση και η οριοθέτησή τους επιτρέπει την ταξινόμηση των κόμβων ανάλογα με την επιρροή που έχουν μέσα στην κοινότητα. Έτσι, κόμβοι με κεντρική θέση στην κοινότητα -π.χ. αν μοιράζονται πολλές ακμές με τους κόμβους της ομάδας τους- πιθανότατα να έχουν καθοριστικό ρόλο στον έλεγχο ή την ευστάθεια της κοινότητας. Από την άλλη, κόμβοι που βρίσκονται στα όρια της κοινότητας έχουν καθοριστικό ρόλο ως γέφυρες, δηλαδή ως διαμεσολαβητές της πληροφορίας που μεταδίδεται από μία κοινότητα σε μία άλλη και είναι οδηγοί των αλληλεπιδράσεων από κοινότητα σε κοινότητα (Csermely, 2008). Τέτοιου είδους διαχωρισμοί είναι ιδιαίτερης σημασίας για την ανάλυση κοινωνικών δικτύων (Burt, 1976; Freeman, 1977; Granovetter, 1973) και δικτύων μεταβολισμού (Guimera and Amaral, 2005).

Επιπλέον, οι κοινότητες μας παρέχουν σημαντική πληροφορία για την ιεραρχία του συστήματος που μελετάμε. Τα δίκτυα του πραγματικού κόσμου συνίστανται από κοινότητες που εμπεριέχουν μικρότερες κοινότητες και αυτές με τη σειρά τους ακόμη μικρότερες κοινότητες κ.ο.κ. Χαρακτηριστικό παράδειγμα είναι τα δίκτυα επιχειρήσεων: ακολουθούν δομή τύπου πυραμίδας, από τους εργάτες που βρίσκονται στη βάση της έως το συμβούλιο διευθυντών στην κορυφή της πυραμίδας, με τα ενδιάμεσα επίπεδα να αντιστοιχούν στις εργασιακές ομάδες, τα επιμέρους τμήματα της επιχείρησης και το management [15]. Έμφαση στο ζωτικό ρόλο της ιεραρχίας στα πολύπλοκα συστήματα δόθηκε από τον Simon το 1962, από τον οποίο διαπιστώθηκε ότι η δημιουργία και η εξέλιξη ενός συστήματος που οργανώνεται σε μικρότερα ευσταθή υποσυστήματα είναι πολύ πιο γρήγορη έναντι ενός αδόμητου συστήματος. Αυτό διότι είναι πολύ πιο εύκολο να εξετασθούν τα επιμέρους δομικά στοιχεία έως ότου να συναρμολογήσουν τελικά το δίκτυο ως όλον. Με τον τρόπο αυτό περιορίζεται η πιθανότητα λάθους κατά τη διαδικασία κατασκευής ενός συστήματος.

Στόχος της ανίχνευσης κοινοτήτων, λοιπόν, είναι η αναγνώριση υποσυστημάτων και πιθανής ιεραρχίας μεταξύ τους αξιοποιώντας μόνο την τοπολογία του δικτύου. Η πρώτη ανάλυση συσταδοποιημένων δομών πραγματοποιήθηκε από τους Weiss & Jacobson το 1955, οι οποίοι μελέτησαν τη σύσταση κοινοτήτων σε ένα δίκτυο εργαζομένων κυβερνητικής υπηρεσίας. Στη μελέτη τους, τα σύνολα κόμβων που αντιστοιχούν στους εργαζόμενους δημιουργήθηκαν διαγράφοντας από το καθένα τους κόμβους που δρούσαν στο όριο της κοινότητας ως γέφυρες, δηλαδή αφαιρώντας από την κάθε ομάδα τους εργαζόμενους που συνεργάζονται με μέλη άλλων κοινοτήτων. Η ιδέα τους αποτέλεσε το θεμέλιο των μεθόδων που χρησιμοποιήθηκαν για εύρεση κοινοτήτων τις επόμενες δεκαετίες. Οι μέθοδοι αυτές είναι ευρέως γνωστές ως Graph Partitioning και βασίζονται στην ιδέα διαγραφής γεφυροποιών κόμβων [15],[31],[55],[56]. Στις μεθόδους αυτές το πρόβλημα της ανίχνευσης ανάγεται σε κλασική ανάλυση ιδιοτιμών και ιδιοδιανυσμάτων της μήτρας του Laplace (Laplacian Matrix) [59]. Ωστόσο, στις αρχές του 2000 διαπιστώθηκε πως η προσέγγιση των εν λόγω μεθόδων και των θυγατρικών τους υστερεί για διάφορους λόγους, οι κυριότεροι από τους οποίους είναι: 1) η απώλεια πληροφορίας λόγω αξιοποίησης ελάχιστων ιδιοτιμών της μήτρας Laplace, 2) το πλήθος και το μέγεθος των κοινοτήτων είναι προκαθορισμένο, 3) οι αλγόριθμοι εργάζονται για εύρεση βέλτιστου διαχωρισμού του δικτύου, ανεξάρτητα από το αν υπάρχει βέλτιστος διαχωρισμός ή όχι (υπάρχει το ενδεχόμενο το δίκτυο να μη διαιρείται παρά μόνο σε μία και μοναδική κοινότητα). 4.)

Σύγκριση για τα μέτρα που προέκυπταν αναφορικά με το αν θα χρησιμοποιούνται αυτούσια ή με κάποια κανονικοποίηση ώστε να προκύπτει κάποιος χαρακτηριστικός συντελεστής [13],[14],[31],[39],[41],[58].

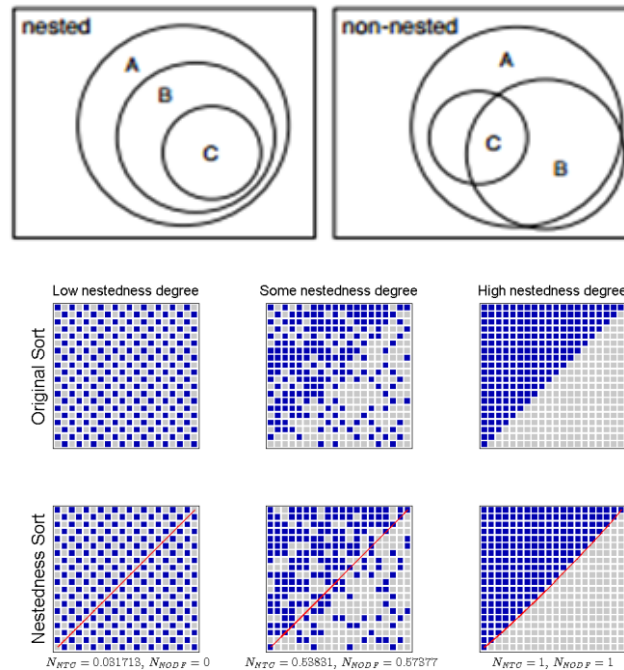
Σύντομα, λοιπόν, άρχισαν να κερδίζουν έδαφος άλλες τεχνικές και αλγόριθμοι που χρησιμοποιούσαν καλύτερη προσέγγιση για την ανίχνευση κοινοτήτων. Στις αρχές της δεκαετίας του '90 αναδύθηκαν οι τεχνικές του hierarchical clustering και του partitional clustering [15], ενώ στα τέλη της ίδιας δεκαετίας με τις δυνατότητες που παρείχαν οι υπολογιστές κάνουν την εμφάνισή τους αλγόριθμοι με διάφορες προσεγγίσεις (δυναμικός προγραμματισμός, παράλληλος προγραμματισμός, simulated annealing κ.α) [15],[30],[31].

Στις αρχές του 2002, οι Girvan & Newman σε δημοσίευσή τους ορίζουν (σχεδόν) εκ νέου τη μαθηματική έκφραση του modularity και πλέον γίνεται λόγος για τεχνικές που από εδώ και στο εξής απαντώνται ως modularity optimization techniques [1-7], [15],[27],[31],[38],[55]. Διατυπώνουν έναν αλγόριθμο που βασίζεται στην ενδιάμεση κεντρικότητα (Betweenness Centrality). Τέσσερα χρόνια αργότερα, το 2006, ο Newman προσαρμόζει τον τελευταίο του ορισμό με μία σταθερά κανονικοποίησης και περιγράφει τον αλγόριθμο Leading Eigenvector [31]. Πρόκειται για έναν αλγόριθμο βασισμένο στη φιλοσοφία του Graph Partitioning [56] με κάποιες τροποποιήσεις για βελτίωση και καλύτερη προσέγγιση. Υπολογίζει το πλήθος κοινοτήτων και το μέγεθος τους κάνοντας φασματική ανάλυση ιδιοτιμών και ιδιοδιανυσμάτων (eigenspectrum analysis) από την οποία οδηγείται σε τιμή modularity, την οποία και προσπαθεί να βελτιστοποιήσει. [10],[18],[31],[55]. Αφού ο αλγόριθμος καταλήξει σε κάποια τιμή modularity, αρχίζει και βελτιστοποιεί την ποσότητα αυτή με τρόπο που θα δούμε αναλυτικά αμέσως πιο κάτω (παράγραφος α₁). Από το 2006 και έπειτα, η επιστημονική κοινότητα ακολουθεί τον ορισμό του Newman για το modularity και οι αλγόριθμοι που φτιάχνονται έχουν ως θεμέλιο το μεγαλύτερο μέρος της δουλειάς του Newman. Λίγο αργότερα την ίδια χρονιά (2006) ο M. Barber -βασισμένος απόλυτα στη μέχρι τώρα δουλειά του Newman- ορίζει στο [4] modularity για διμερή δίκτυα μόνο (Bipartite modularity). Με χρήση επαναληπτικής μεθόδου οδηγείται στον αλγόριθμο Adaptive BRIM που στοχεύει στην ανίχνευση κοινοτήτων μέσω μεγιστοποίησης του modularity [27],[42],[50]. Τέλος, το 2009 ο Tsuyoshi Murata στο [27] και μετά από ενδελεχή επισκόπηση των εργασιών των δύο προηγούμενων, βρίσκει ευάλωτα σημεία στους αλγόριθμους και των δύο όταν πρόκειται για αναλύσεις που αφορούν σε large-scale bipartite networks. Έτσι, προτείνει έναν αλγόριθμο που συνδυάζει τον Adaptive Brim με μία κατηγορία αλγορίθμων γνωστών ως LP Algorithms (Label Propagation Algorithms) δίνοντας καλύτερη ακρίβεια στον αλγόριθμο του Barber για την εύρεση Bipartite Modularity [9],[13],[27].

2.2 Nestedness

Η εμφωλευσιμότητα (nestedness) είναι μια ιδιότητα που αφορά τη δομή ενός οικολογικού συστήματος και πιο συγκεκριμένα τον τρόπο με τον οποίο είναι σχηματισμένες οι συνδέσεις. Ένα διμερές δίκτυο είναι nested, αν οι συνδέσεις των κόμβων μικρού βαθμού (specialists) είναι υποσύνολο των κόμβων μεγαλύτερου

βαθμού, οι οποίες με τη σειρά τους είναι υποσύνολο των συνδέσεων κάποιου κόμβου μεγαλύτερου βαθμού κοκ. Οι κόμβοι με τις περισσότερες συνδέσεις ενός nested δικτύου καλούνται generalists.



Εικόνα 6 Παραδείγματα Εμφωλευμένων δομών και εικονική αναπαράσταση της μήτρας εμφωλευσιμότητας (nestedness matrix)

Ενδεικτική για την κατανόηση του nestedness είναι η παραπάνω εικόνα. Δεξιά απεικονίζεται ένα τέλεια nested δίκτυο, αφού οι συνδέσεις κάθε κόμβου είναι υποσύνολο των συνδέσεων ενός άλλου. Στη μέση απεικονίζεται ένα δίκτυο με κάποιο βαθμό nestedness, ενώ από το αριστερό απουσιάζει εντελώς η ιδιότητα αυτή.

Η ιδιότητα nestedness εισήχθη ως ιδέα πολλά χρόνια πριν. Η αρχική ιδέα αφορούσε νησιά που δέχονται είδη-αποίκους από τη στεριά (Hulten 1937, Darlington 1957). Χαρακτηριστικό είναι το immigrant pattern του Darlington (1957) , σύμφωνα με το οποίο:

- οι καλύτεροι διασκορπιστές (dispersers) αναμένεται να αποικίσουν στα πιο πολλά νησιά, συμπεριλαμβανομένων των πιο απομακρυσμένων
- οι χειρότεροι στα πιο κοντινά

Ο επίσημος ορισμός της nestedness δόθηκε από τους Atmar και Patterson το 1986. Σύμφωνα με αυτόν σε ένα nested δίκτυο τα όντα που ζουν σε species-poor τοποθεσίες θα είναι υποσύνολο των όντων που ζουν σε species-rich.

2.3 Modularity Detection Algorithms

2.3.1 Newman's "Leading Eigenvector"

Ο αλγόριθμος Leading Eigenvector (οδηγό ιδιοδιάνυσμα), που αναπτύχθηκε από τον M.J. Newman το 2006, είναι μια τεχνική Ανάλυσης Δεδομένων και εντοπίζει τις

κοινοτήτες μέσα σε ένα δίκτυο. Από το 2006 αποτελεί σημείο αναφοράς για τις περισσότερες έρευνες προς αυτή την κατεύθυνση και εφαρμόζεται τόσο σε μονομερή όσο και σε διμερή δίκτυα με την εξής διαφορά: επειδή αναπτύχθηκε αρχικά για μονομερή δίκτυα, κατά την αναζήτηση κοινοτήτων σε ένα μονομερές δίκτυο με n κόμβους εργάζεται πάνω στη μήτρα γειννίαςης A (adjacency matrix), οπότε για ανάλυση μονομερούς δικτύου δέχεται ως είσοδο κανονικά τη μήτρα A διαστάσεων $n \times n$. Για διμερή δίκτυα, όμως, με n κόμβους στο ένα σύνολο και m στο άλλο διατάσσει τους κόμβους με τρόπο, ώστε να εργαστεί πάνω σε μία μήτρα διαστάσεων $(m+n) \times (m+n)$ (την οποία ορίζουμε παρακάτω) αντί για τη μήτρα γειννίαςης του διμερούς δικτύου.

Έστω, λοιπόν, διμερές δίκτυο με μήτρα γειννίαςης A και τα στοιχεία της στη διάταξη που αρμόζει για ανάλυση διμερούς δικτύου. Ας κάνουμε την απλή θεώρηση για διαχωρισμό σε δύο κοινότητες: module1 και module2. Ο αναλυτικός τύπος για το modularity κατά τον Newman είναι ο εξής:

$$Q = \frac{1}{4m} \sum_{ij} A_{ij} - \frac{k_i k_j}{2m} \quad , \quad (1)$$

μόνο εφόσον τα i, j ανήκουν στο ίδιο module. Αν τα i, j ανήκουν σε διαφορετικά modules, τότε προσθέτουμε στο Q το μηδέν.

Ο όρος $\frac{k_i k_j}{2m}$ εκφράζει το πλήθος των ακμών μεταξύ των κόμβων i, j αν οι ακμές τοποθετηθούν με τυχαίο τρόπο. Τα k_i, k_j είναι οι βαθμοί των κόμβων και $m = \frac{1}{2} \sum_i k_i$ είναι το πλήθος των ακμών στο δίκτυο που μελετάμε. Το μέτρο Q , με λίγα λόγια συγκρίνει το πλήθος των πραγματικών ακμών μεταξύ των κόμβων με το πλήθος των ακμών που αναμένεται αν αυτές τοποθετηθούν με τυχαίο τρόπο.

Για να γίνει το άθροισμα της σχέσης (1) πιο «ομοιόμορφο», δηλαδή για να αθροίζουμε πάνω σε όλα τα πιθανά i, j χωρίς να κάνουμε τον έλεγχο σε ποιο module ανήκουν εισάγουμε μία ποσότητα $\delta(i, j)$ τέτοια ώστε:

$$\delta(i, j) = \begin{cases} 1, & \text{αν } i, j \text{ ανήκουν στο ίδιο module} \\ 0, & \text{αν ανήκουν σε διαφορετικό module} \end{cases}$$

Η σχέση (1) μπορεί πλέον να γραφτεί:

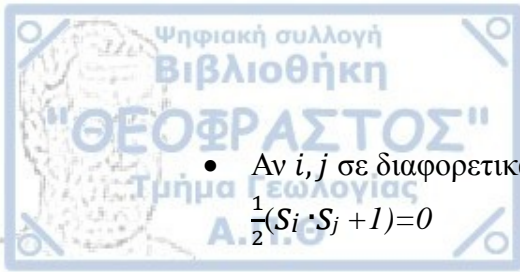
$$Q = \frac{1}{4m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta(i, j) \quad (2)$$

Στο σημείο αυτό, ορίζεται ένα διάνυσμα S , τέτοιο ώστε:

- $S_i = 1$, αν ο κόμβος i ανήκει στο module1
- $S_i = -1$, αν ο κόμβος i ανήκει στο module2

Για δύο κόμβους i, j σύμφωνα με τα ανωτέρω έχουμε:

- Αν i, j στο ίδιο module, τότε: $S_i = S_j = 1$ ή $S_i = S_j = -1$ και $\frac{1}{2}(S_i \cdot S_j + 1) = 1$



- Αν i, j σε διαφορετικά modules, τότε: $S_i=1, S_j=-1$ ή $S_i=-1, S_j=1$ και $\frac{1}{2}(S_i \cdot S_j + 1) = 0$

Έτσι, η σχέση (2) για το Q γίνεται:

$$Q = \frac{1}{4m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) (S_i \cdot S_j + 1) = \frac{1}{4m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) S_i \cdot S_j \quad (3)$$

Η δεύτερη ισότητα προκύπτει από την παρατήρηση ότι $2m = \sum_i k_i = \sum_{ij} A_{ij}$. [60]

Η εξίσωση (3) υπό μορφή μητρών γράφεται: $Q = \frac{1}{4m} \mathbf{s}^T \mathbf{B} \mathbf{s}$ (4)

όπου το \mathbf{s} είναι διάνυσμα στήλης με στοιχεία τα \mathbf{s}_i και η μήτρα \mathbf{B} είναι πραγματική συμμετρική μήτρα με στοιχεία $B_{ij} = A_{ij} - \frac{k_i k_j}{2m}$ (5)

Η μήτρα \mathbf{B} καλείται *modularity matrix*.

Το σημείο αυτό είναι ιδιαίτερης σημασίας, καθώς η ιδέα του Newman για τον αλγόριθμο αυτό είναι ότι μπορούμε να αναλύσουμε την εξίσωση (4) ως γραμμικό συνδυασμό των ιδιοτιμών β_i και των κανονικοποιημένων ιδιοδιανυσμάτων (normalized eigenvectors) \mathbf{u}_i που αντιστοιχούν σε αυτές, ώστε να μεγιστοποιηθεί κατά το βέλτιστο το modularity Q . Γιατί, όμως, υπάρχει αυτή η δυνατότητα;

Η μήτρα \mathbf{A} είναι διαστάσεων $(m+n) \times (m+n)$. Η μήτρα \mathbf{B} είναι επίσης συμμετρική, καθώς προκύπτει από τη μήτρα \mathbf{A} σύμφωνα με τη σχέση (5). Αφού, λοιπόν, η μήτρα \mathbf{B} είναι συμμετρική και πραγματική, διαγωνοποιείται. Άρα, σύμφωνα με το Spectral Theorem [61], υπάρχει βάση του R^n , και μάλιστα ορθοκανονική που αποτελείται από τα

ιδιοδιανύσματα του \mathbf{B} . Επιπλέον, το διάνυσμα $\mathbf{s} \in R^n$, καθώς $\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}$

Αφού υπάρχει βάση του R^n που αποτελείται από τα ιδιοδιανύσματα του πίνακα \mathbf{B} και $\mathbf{s} \in R^n$, τότε το \mathbf{s} μπορεί να γραφεί σαν γραμμικός συνδυασμός των ιδιοτιμών β_i και των ιδιοδιανυσμάτων \mathbf{u}_i του \mathbf{B} .

Θέλουμε, λοιπόν, να γράψουμε τα \mathbf{s} ως γραμμικό συνδυασμό των ιδιοδιανυσμάτων του \mathbf{B} , έτσι ώστε $\mathbf{s} = \sum_{i=1}^n \alpha_i \mathbf{u}_i$, όπου $\alpha_i = \mathbf{u}_i^T \cdot \mathbf{s}$. Έτσι, με αντικατάσταση στη σχέση (4) προκύπτει ότι:

$$Q = \frac{1}{4m} \sum_i \alpha_i \mathbf{u}_i^T \cdot \mathbf{B} \cdot \sum_j \alpha_j \mathbf{u}_j \quad (6)$$

Όμως, $\mathbf{B}\mathbf{u}_j = \beta_j \mathbf{u}_j$ και $\mathbf{u}_i^T \beta_j = \begin{cases} 1, & \text{αν } i = j \\ 0, & \text{αν } i \neq j \end{cases}$ (ορθοκανονική βάση)

Άρα μένουν στο άθροισμα της (6) μόνο οι όροι για $i=j$ και έτσι έχουμε τον τελικό τύπο του modularity εκφρασμένο σε γραμμικό συνδυασμό ιδιοτιμών και ιδιοδιανυσμάτων:

$$Q = \frac{1}{4m} \sum_{i=1}^n \alpha_i^2 \beta_i = \frac{1}{4m} \sum_{i=1}^n (\mathbf{u}_i^T \cdot \mathbf{s})^2 \beta_i \quad (7)$$

Από εδώ και στο εξής τα πράγματα είναι απλά: Τα ιδιοδιανύσματα β_i διατάσσονται σε φθίνουσα σειρά $\beta_1 \geq \beta_2 \geq \dots \geq \beta_n$. Πλέον, το πρόβλημα της μεγιστοποίησης του modularity Q για βέλτιστο διαχωρισμό του δικτύου ανάγεται στην κατάλληλη επιλογή των στοιχείων του διανύσματος \mathbf{s} . Καθώς το β_1 είναι μέγιστη ιδιοτιμή, αρκεί να πάρει το άθροισμα της σχέσης (7) τη μέγιστη δυνατή τιμή. Οπότε αρκεί να μεγιστοποιηθεί το εσωτερικό γινόμενο $\mathbf{u}_i^T \cdot \mathbf{s}$. Αυτό επιτυγχάνεται ως εξής: Αφού το \mathbf{s} έχει τιμές ± 1 , ο αλγόριθμος επιλέγει τιμές για το \mathbf{s} έτσι ώστε: $\mathbf{s} = \begin{cases} 1, & \text{αν } \mathbf{u}_i > 0 \\ -1, & \text{αν } \mathbf{u}_i < 0 \end{cases}$

Με τον τρόπο αυτό, ο αλγόριθμος Leading Eigenvector επιλέγει να αθροίζει πάντα θετικές τιμές που προφανώς θα δώσουνε μέγιστο για το επιθυμητό γινόμενο

$\mathbf{u}_i^T \cdot \mathbf{s}$, οπότε μεγιστοποιείται η σχέση (7).

*Μερικές παρατηρήσεις από τον Newman [31]:

- Κατά αντιστοιχία με τη μήτρα του Laplace $L=D-A$, όπου D μήτρα με στοιχεία τους βαθμούς των κόμβων του δικτύου και A η μήτρα γειτνίασης [59], η μήτρα B έχει τις εξής ιδιότητες:
 - Τα στοιχεία κάθε γραμμής αθροίζουν σε μηδέν
 - $\mathbf{B} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = 0$
 - Σύμφωνα με την 2η ιδιότητα, υπάρχει ιδιοτιμή ίση με το μηδέν, στην οποία αντιστοιχεί το ιδιοδιάνυσμα $\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ (trivial solution)
 - Όλα τα άλλα ιδιοδιανύσματα της μήτρας B είναι κάθετα στο ιδιοδιάνυσμα που προαναφέρθηκε, γιατί τα ιδιοδιανύσματα που αντιστοιχούν σε διαφορετικές ιδιοτιμές είναι πάντα κάθετα.
- Αν σε κάποια ιδιοτιμή αντιστοιχούν περισσότερα από ένα ιδιοδιανύσματα, τότε πρέπει να εφαρμοσθεί ο αλγόριθμος Gram-Schmidt. Πρόκειται για έναν αλγόριθμο που λαμβάνει σαν εισόδους ιδιοδιανύσματα γραμμικά ανεξάρτητα μεταξύ τους και τα μετατρέπει σε ορθοκανονικά [62].

- Αν η μήτρα για κάποιο λόγο δεν είναι τετραγωνική, θα χρειαστεί να γίνει SVD (Singular Value Decomposition), ώστε να ληφθούν τα απαραίτητα ιδιοδιάνυσματα που απαιτούνται για τον αλγόριθμο Leading Eigenvector [63].
- Αν υπάρχει θετική ιδιοτιμή, τότε η οδηγούσα ιδιοτιμή β_l που αντιστοιχεί στο οδηγό ιδιοδιάνυσμα \mathbf{u}_l είναι σίγουρα θετική ιδιοτιμή. Τότε, το \mathbf{u}_l είναι κάθετο

στο $\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$. Άρα το \mathbf{u}_l έχει και θετικά και αρνητικά στοιχεία. Συνεπώς, ο

αλγόριθμος δημιουργώντας το διάνυσμα \mathbf{s} θα κάνει την ταξινόμηση ως εξής: αν το i στοιχείο του \mathbf{u}_l είναι θετικό, τότε ο κόμβος i καταχωρείται στο module1, ενώ αν είναι αρνητικό καταχωρείται στο module2 (για την απλή θεώρηση διαχωρισμού σε δύο κοινότητες). Εφόσον, λοιπόν, το \mathbf{u}_l έχει και θετικά και αρνητικά στοιχεία, τότε θα υπάρχουν κόμβοι και στις δύο κοινότητες. Η συγκεκριμένη παρατήρηση αποτελεί και τον πυρήνα του εν λόγω αλγορίθμου [7],[8].

- Αν όλες οι ιδιοτιμές είναι αρνητικές, τότε $\mathbf{u}_l = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ και τότε όλοι οι κόμβοι ανήκουν στην ίδια κοινότητα που είναι και η μοναδική κοινότητα του δικτύου.

Διαχωρισμός του δικτύου σε περισσότερες από δύο κοινότητες

Αφού ολοκληρωθεί η πιο πάνω διαδικασία, ο αλγόριθμος θα έχει δώσει ένα διαχωρισμό του δικτύου σε δύο κοινότητες. Στη συνέχεια, ο Newman [31] εξηγεί πώς ο αλγόριθμος μπορεί να εφαρμοστεί επαναληπτικά σε κάθε module για περαιτέρω διαχωρισμούς σε κοινότητες. Ωστόσο, η προσέγγιση να απομονωθεί κάθε module και να εφαρμοστεί ο αλγόριθμος δεν είναι σωστή, και αυτό γιατί για να απομονωθεί ένα module πρέπει να διαγραφούν οι ακμές που το συνδέουν με άλλα. Η διαγραφή ακμών, όμως, προκαλεί προφανώς αλλαγή στους βαθμούς των κόμβων (κάτι που αποτελεί μειονέκτημα στη μέθοδο Graph Partitioning), οπότε κάθε απόπειρα μεγιστοποίησης του modularity θα μεγιστοποιεί τη λάθος ποσότητα [12],[15],[31],[58]. Για το λόγο αυτό, ο Newman εισάγει στον αλγόριθμό του μία ποσότητα ΔQ , την οποία καλεί *generalized modularity*, ώστε να ποσοτικοποιήσει την αλλαγή που προκύπτει στο modularity Q με κάθε επανάληψη του αλγορίθμου για απόπειρα διαίρεσης του δικτύου σε επιμέρους κοινότητες.

Έστω ότι το δίκτυο έχει διαχωριστεί για πρώτη φορά σε communities (modules) και επιθυμείται διαίρεση στα δύο ενός module g το οποίο έχει μέγεθος n_g , τότε η επιπλέον συνεισφορά ΔQ στο Q είναι:

$$\Delta Q = \frac{1}{2m} \left[\frac{1}{2} \sum_{i,j \in g} B_{ij} (s_i s_j + 1) - \sum_{i,j \in g} B_{ij} \right]$$

$$\begin{aligned}
 &= \frac{1}{4m} \left[\sum_{i,g \in g} B_{ij} s_i s_j - \sum_{i,j \in g} B_{ij} \right] \\
 &= \frac{1}{4m} \sum_{i,g \in g} [B_{ij} - \delta_{ij} \sum_{k \in g} B_{ik}] s_i s_j \delta \\
 &= \frac{1}{4m} \mathbf{s}^T \mathbf{B}^{(g)} \mathbf{s}, \quad (8)
 \end{aligned}$$

Όπου δ_{ij} είναι το δ του Kronecker, κάνουμε χρήση της ιδιότητας $s_i^2 = 1$ και $\mathbf{B}^{(g)}$ είναι μήτρα διαστάσεων $n_g \times n_g$ και στοιχεία:

$$B_{ij}^{(g)} = B_{ij} - \delta_{ij} \sum_{k \in g} B_{ik} \quad (9)$$

Επειδή η σχέση (8) είναι ίδιας μορφής με τη σχέση (4) ο αλγόριθμος εφαρμόζεται στη φάση αυτή για μεγιστοποίηση της ποσότητας του generalized modularity ΔQ μέσω ανάλυσης του νέου modularity matrix $B_{ij}^{(g)}$. Οι γραμμές και οι στήλες της μήτρας αυτής αθροίζουν στο μηδέν και η ποσότητα ΔQ είναι ορθώς μηδέν αν το module g είναι αδιαίρετο. Επίσης, για ένα πλήρες δίκτυο η εξίσωση (9) επανέρχεται στον προηγούμενο ορισμό του modularity matrix που δίνει η εξίσωση (3), διότι ο όρος $\sum_{k \in g} B_{ik}$ στην περίπτωση αυτή ισούται με μηδέν.

Ένα καλό ερώτημα είναι πότε ο αλγόριθμος θα σταματήσει τη διαδικασία υποδιαίρεσης. Η απάντηση βρίσκεται στον ίδιο τον αλγόριθμο: ουσιαστικά σε κάθε επανάληψη του αλγορίθμου (δηλαδή σε κάθε διαχωρισμό) με τη μέθοδο Leading Eigenvector γίνεται optimization του ΔQ και έλεγχος της τιμής του. Ο διαχωρισμός γίνεται δεκτός από τον αλγόριθμο αν και μόνο αν $\Delta Q > 0$, δηλαδή αν η συνεισφορά του ΔQ οδηγεί σε αύξηση του Q και όχι σε μείωση. Αν η ποσότητα ΔQ έχει μηδενική ή αρνητική συνεισφορά στο modularity, τότε το module παραμένει ως έχει. Μόλις ολόκληρο το δίκτυο έχει διαιρεθεί σε τέτοιου είδους αδιαίρετα υποδίκτυα, ο αλγόριθμος σταματά.

Άμεση απόρροια της μεθόδου αυτής είναι ότι οι κοινότητες (communities) ενός δικτύου είναι εξ ορισμού αδιαίρετες. Αρκετοί συγγραφείς πρότειναν ορισμούς για το modularity με την πάροδο των ετών [64],[65],[66]. Η μέθοδος του Newman έχει τεράστια συνεισφορά, διότι ουσιαστικά προσφέρει έναν νέο επί της αρχής ορισμό του community ως αδιαίρετου υπογραφήματος ενός δικτύου.

2.3.2 Ο αλγόριθμος Fast-Greedy

Ο αλγόριθμος fast-greedy είναι άλλη μία προσέγγιση για αξιολόγηση της ιεραρχίας ενός δικτύου μέσω του διαχωρισμού σε κοινότητες. Παρόμοια με τον αλγόριθμο Leading Eigenvector του Newman, αυτό που προσπαθεί να κάνει ο αλγόριθμος, είναι να βελτιστοποιήσει μία συνάρτηση σύμφωνα με την οποία ορίζεται

η ιδιότητα του modularity με έναν «άπληστο» (greedy) τρόπο, ώστε να οδηγηθεί γρήγορα σε σύγκλιση. Κατά την αρχικοποίηση, θεωρεί ότι κάθε κόμβος του δικτύου ανήκει σε μία ξεχωριστή κοινότητα (όπως και στον αλγόριθμο του Newman). Έπειτα, οι κοινότητες προκύπτουν επαναληπτικά από αξιολόγηση της δομής τους έως ότου κάθε προκύπτουσα κοινότητα να είναι κατά το βέλτιστο ισχυρά δομημένη, έστω και τοπικά (π.χ. ο διαχωρισμός κοινοτήτων τοπικά έχει επιφέρει τη μέγιστη δυνατή αύξηση στη συνάρτηση του modularity) [99].

Ο αλγόριθμος σταματά την αναζήτηση για βέλτιστο διαχωρισμό σε κοινότητες, όταν η συνάρτηση του modularity δεν επιδέχεται επιπλέον εξέτασης για τοπικά μέγιστα, ήτοι θεωρεί ότι έχει κάνει τον καλύτερο δυνατό διαχωρισμό. Η μέθοδος αυτή εφαρμόζεται γρήγορα όσο αφορά την υπολογιστική ισχύ και προτιμάται, διότι δεν προαπαιτεί κάποιον καθορισμό ή ρύθμιση παραμέτρων [99]-[101].

Ωστόσο, ο αλγόριθμος αυτός είναι γνωστό ότι υστερεί λόγω περιορισμού από κάποιο όριο ανάλυσης (resolution limitation). Πιο συγκεκριμένα, ανάλογα με το πλήθος των κόμβων και των ακμών, κοινότητες με τιμή modularity κάτω από ένα συγκεκριμένο όριο που θέτει ο αλγόριθμος θα θεωρηθούν «ασθενείς» και θα ενταχθούν στην αμέσως επόμενη μεγαλύτερη γειτονική κοινότητα [100].

2.3.4 Barber's "Adaptive BRIM"

Ο αλγόριθμος BRIM (Bipartite Recursive Identification of Modules) είναι επίσης αλγόριθμος για community detection και μάλιστα αποκλειστικά για διμερή δίκτυα. Αναπτύχθηκε από τον Barber το 2006, ο οποίος την ίδια χρονιά και βασισμένος στη δουλειά του Newman (αλγόριθμος Leading Eigenvector), ορίζει bipartite modularity, δηλαδή modularity αποκλειστικά για διμερή δίκτυα, όπου αντί για μήτρες χρησιμοποιεί στην ανάλυσή του μήτρες μητρώων, δηλαδή μήτρες με στοιχεία πίνακες.

Αρχικά, κατά πλήρη αντιστοιχία με τον Newman, εκφράζει το modularity ενός δικτύου ως (βλ. Σχέση 2):

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - P_{ij}) \delta(g_i g_j) \quad (10),$$

όπου A η μήτρα γειννίασης του δικτύου και $P_{ij} = \frac{k_i k_j}{2m}$ με k_i, k_j τους βαθμούς των κορυφών και m το πλήθος των ακμών.

Για διμερή δίκτυα, έστω ότι p είναι το πλήθος των κόκκινων κόμβων και q το πλήθος των μπλε κόμβων. Υποθέτοντας ότι k_i, d_j είναι ο βαθμός του i κόκκινου κόμβου και του j κόμβου με g_i, h_j τα communities στα οποία ο i κόκκινος κόμβος και ο j μπλε κόμβος ανήκουν κατά αντιστοιχία, τότε -καθώς δεν υπάρχουν ακμές μεταξύ ίδιου χρώματος κόμβων- η μήτρα γειννίασης κατά Barber του διμερούς δικτύου παίρνει την μορφή τετραγωνικής μήτρας και είναι:

$$A = \begin{bmatrix} O_{p \times p} & \tilde{A}_{p \times q} \\ \tilde{A}_{q \times p}^T & O_{q \times q} \end{bmatrix} \quad (11)$$

Η ποσότητα P_{ij} εκφράζει, όπως και στο μοντέλο του Newman, τις τυχαίες συνδέσεις στο δίκτυο [4],[7],[27],[31] και για διμερή δίκτυα πρέπει να αποκλειστούν οι ακμές μεταξύ κόμβων του ίδιου χρώματος και συνεπώς είναι:

$$P = \begin{bmatrix} O_{p \times p} & \hat{P}_{p \times q} \\ \hat{P}_{q \times p}^T & O_{q \times q} \end{bmatrix}, \quad (12)$$

$$\text{όπου } \hat{P}_{ij} = \frac{k_i d_j}{m} \quad (13)$$

Θεωρώντας $\hat{B} = \tilde{A} - \hat{P}$, έχουμε:

$$\begin{aligned} Q_b &= \frac{1}{2m} \sum_{i=1}^p \sum_{j=1}^q (A_{ij} - P_{ij}) \delta(g_i, h_j) \\ &= \frac{1}{m} \sum_{i=1}^p \sum_{j=1}^q \hat{B}_{ij} \delta(g_i, h_j) \end{aligned} \quad (14)$$

Η τελευταία σχέση καλείται Bipartite Modularity κατά Barber και προς αποφυγή σύγχυσης χρησιμοποιείται απλά ο όρος Bipartite Modularity.

Έστω ότι το δίκτυο έχει c communities (modules). Παρόμοια με τον αλγόριθμο του Newman, ορίζεται και εδώ ένα διάνυσμα S διαστάσεων $n \times c$:

$$S_{ij} \begin{cases} 1, & \text{αν ο κόμβος } i \text{ ανήκει στο community } j \\ 0, & \text{στην αντίθετη περίπτωση} \end{cases} \quad (15)$$

Στα διμερή δίκτυα, το διάνυσμα S μπορεί να διαχωριστεί ως εξής:

$$S = \begin{bmatrix} R_{p \times c} \\ T_{q \times c} \end{bmatrix}, \quad (16)$$

όπου R μήτρα κατάταξης για τους κόκκινους κόμβους και T για τους μπλε.

Μία ισοδύναμη έκφραση για το modularity, λοιπόν, είναι:

$$Q_b = \frac{1}{m} \sum_{i=1}^p [\sum_{k=1}^c R_{ik} (\widehat{BT})_{ik}], \quad (17)$$

καθώς κάθε γραμμή της μήτρας R περιέχει ένα μόνο στοιχείο ίσο με τη μονάδα και όλα τα άλλα στοιχεία είναι μηδενικά. Για παράδειγμα, αν $R_{ik}=1$, τότε ο κόμβος της i γραμμής ανήκει στο k community με $R_{ij}=0$ για κάθε $j \neq k$. Έτσι, η μεγιστοποίηση του Q_b γίνεται απλή: αρκεί να καταταχθεί ο i κόκκινος κόμβος στο community k , έτσι ώστε ο όρος $(\widehat{BT})_{ik}$ να είναι το μέγιστο της i -οστής γραμμής του \widehat{BT} . Παρατηρούμε ότι ο όρος \widehat{BT} καθορίζεται από το διαχωρισμό των μπλε κόμβων (λόγω της μήτρας T). Οπότε ο διαχωρισμός των κόκκινων κόμβων επάγεται από το διαχωρισμό των μπλε

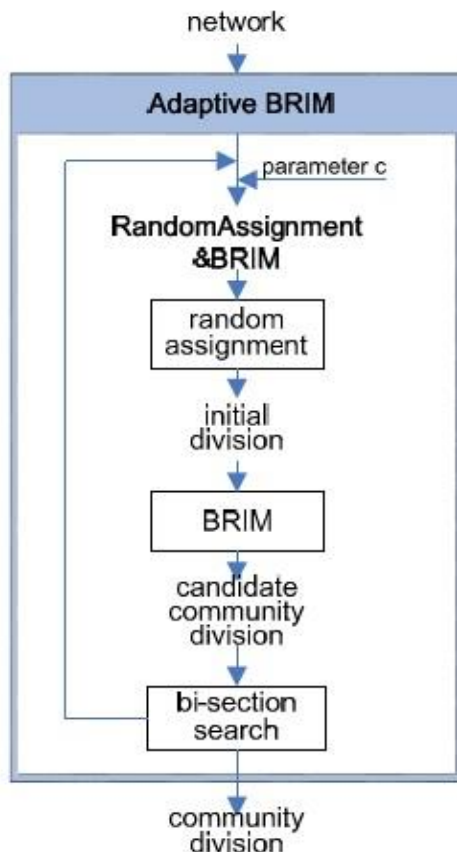
κόμβων με στόχο την μεγιστοποίηση του modularity. Όμοια με πριν, λοιπόν, συνάγεται ότι:

$$Q_b = \frac{1}{m} \sum_{j=1}^q [\sum_{k=1}^c T_{jk} (\hat{B}^T R)_{jk}] \quad (18)$$

Οι δύο διαδικασίες που προκύπτουν από τις σχέσεις (17) και (18), στοιχειοθετούν τον αλγόριθμο standard BRIM: ξεκινώντας από έναν διαχωρισμό σε communities των κόκκινων κόμβων, επάγεται ο διαχωρισμός των μπλε κόμβων σε communities. Από τον διαχωρισμό που προέκυψε για τους μπλε κόμβους επάγεται ο επόμενος διαχωρισμός για τους κόκκινους και συνεχίζεται η ίδια διαδικασία επαναληπτικά. Έτσι, η bipartite modularity αυξάνεται έως ότου ο αλγόριθμος βρει τοπικό μέγιστο και τότε επιτυγχάνεται ο διαχωρισμός σε κοινότητες.

Ωστόσο, η επιλογή μιας προκαθορισμένης τιμής του c περιορίζει την αποτελεσματικότητα του αλγορίθμου. Για το λόγο αυτό, ο Barber προσαρμόζει (adaption, εξού και adaptive BRIM) στον αλγόριθμο με μέθοδο “διαίρει και βασίλευε” ώστε να φτάσει στο βέλτιστο πλήθος $c = c^*$ από communities: ξεκινώντας από $c=1$ και $Q_b(1) = 0$ ο adaptive BRIM αναζητά το βέλτιστο c διπλασιάζοντας επαναληπτικά τον αριθμό των modules όσο το modularity αυξάνει, $Q_b(2c) > Q_b(c)$. Σε κάποιο σημείο, η αναζήτηση εντοπίζει μέγιστο, π.χ. $Q_b(2c) < Q_b(c)$, οπότε η τιμή c^* παρεμβάλλεται σε κάποια από τις ενδιάμεσες τιμές του διαστήματος $(c/2, c)$.

Στην επόμενη εικόνα φαίνεται το διάγραμμα ροής του αλγορίθμου:



Εικόνα 7 Διάγραμμα ροής του αλγορίθμου Adaptive Brim [27]

2.3.5 “Label Propagation” & Murata’s “LP&BRIM”

Ο αλγόριθμος Adaptive BRIM εφαρμόζεται με πολύ καλή ακρίβεια σε διμερή δίκτυα μικρής κλίμακας, δεν είναι όμως κατάλληλος για διμερή δίκτυα μεγάλης κλίμακας λόγω του μεγάλου χρόνου που απαιτεί η αναζήτηση της κατάλληλης τιμής του c . Για να ξεπεραστεί αυτό το εμπόδιο, οι Xin Liu & Tsuyoshi Murata [27] το 2009 δημοσιεύουν τον αλγόριθμο LP&BRIM. Συνδυάζοντας τη στρατηγική του Barber [4] (Adaptive BRIM) με τον αλγόριθμο LP [67],[68], ένα σθεναρό αλγόριθμο για ανίχνευση κοινοτήτων, οδηγούν τον LP ώστε να κάνει τον αρχικό και χρονοβόρο διαχωρισμό του δικτύου σε κοινότητες και έπειτα αξιοποιούν τον Adaptive BRIM για να κάνει τις τελικές υποδιαιρέσεις των κοινοτήτων που βρήκε ο πρώτος.

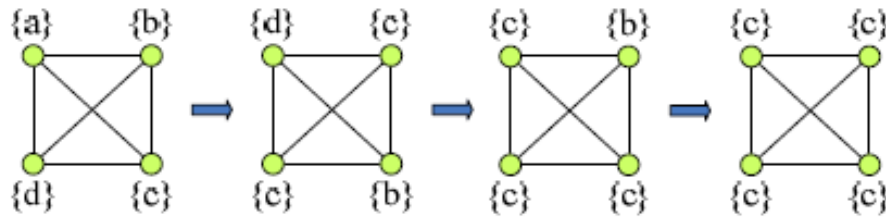
2.3.6 “Label Propagation”

Οι Raghavan και Albert, πρότειναν το 2007 έναν αλγόριθμο για community detection σε μονομερή δίκτυα, γνωστό ως LP (Label Propagation – Διάδοση Ετικέτας) [67]. Τα πιο σημαντικά από τα πλεονεκτήματά του –πέραν της ταχύτητας- είναι οι χαμηλότερες απαιτήσεις σε υπολογιστική ισχύ από οποιονδήποτε άλλο αλγόριθμο [7],[27],[67], καθώς και το γεγονός ότι δε χρειάζεται να είναι γνωστή καμία παράμετρος εκ των προτέρων (όπως λ.χ. η παράμετρος c στον Adaptive BRIM). Μοναδικό του μειονέκτημα αποτελεί το γεγονός ότι λόγω της φύσης του αλγορίθμου και της στρατηγικής του, δεν παρέχει μοναδική λύση -όπως οι αλγόριθμοι που εξετάστηκαν μέχρι τώρα- αλλά μία υπέρθεση της οικογένειας των λύσεων. Αυτό συμβαίνει, γιατί κάθε φορά που εφαρμόζεται ο αλγόριθμος και σε κάθε του επανάληψη, ο υπό εξέταση κόμβος δε λαμβάνεται με κάποια σειρά ή με κάποιον κανόνα αλλά τυχαία. Η βασική ιδέα του αλγορίθμου είναι απλή: αρχικά, σε κάθε κόμβο ανατίθεται ένα αναγνωριστικό ή αλλιώς μία ετικέτα (label). Η ετικέτα αντιπροσωπεύει την κοινότητα στην οποία ανήκει ο κόμβος. Σε κάθε βήμα, όλοι οι κόμβοι ανανεώνουν τις ετικέτες τους αποκτώντας την ετικέτα που φέρει η πλειοψηφία των γειτόνων του (maximal label). Αν ένας κόμβος έχει δύο ή περισσότερα maximal labels, τότε η ανάθεση ετικέτας στο συγκεκριμένο κόμβο επιλέγεται τυχαία [27]. Με αυτή την επαναληπτική διαδικασία πυκνά συνδεδεμένοι κόμβοι φτάνουν πολύ γρήγορα στην απόκτηση μίας (κοινής) ετικέτας και έτσι δομούνται οι κοινότητες (Εικόνα7).

Η διαδικασία σε 5 βήματα [68]:

1. Αρχικοποίηση: ανάθεση ετικετών στους κόμβους του δικτύου (ή έστω στους περισσότερους). Για έναν κόμβο x , $C_x(0)=X$ (ετικέτα)
2. $t=1$
3. Διάταξη των κόμβων με τυχαίο τρόπο και επανάθεση της ετικέτας X σε τυχαίο κόμβο.
4. Για κάθε $x \in X$, $C_x(t) = f(C_{xi1}(t), \dots, C_{xim}(t), C_{xi(m+1)}(t-1), \dots, C_{xik}(t-1))$, όπου η f εντοπίζει την ετικέτα που εμφανίζεται με μεγαλύτερη συχνότητα και την αναθέτει στον κόμβο x .

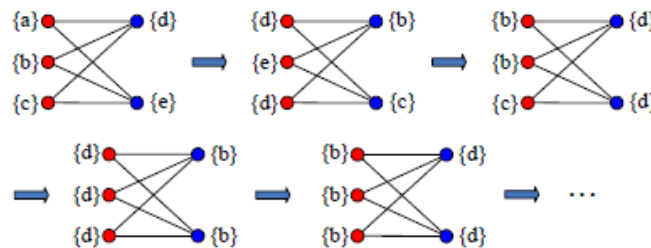
5. Αν κάθε κόμβος στο δίκτυο φέρει την ετικέτα που φέρει η πλειοψηφία των γειτόνων του, τότε ο αλγόριθμος σταματά. Διαφορετικά, θέτει $t' = t+1$ και επιστρέφει στο Βήμα3.



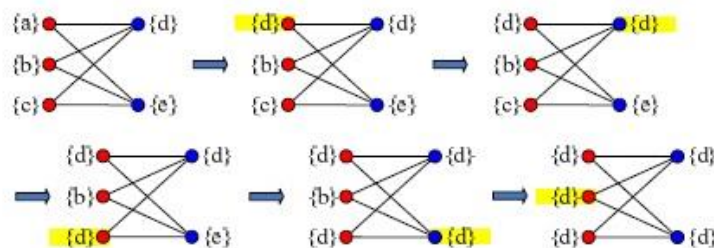
Εικόνα 8 Δόμηση μίας κοινότητας με τον αλγόριθμο LP. Στο συγκεκριμένο παράδειγμα, οι ετικέτες διαδίδονται σύγχρονα. Μετά από τρεις επαναλήψεις ο αλγόριθμος έχει δομήσει από το πυκνά συνδεδεμένο γκρουπ τεσσάρων κόμβων μία κοινότητα με την ετικέτα "{c}" [27]

Η ανανέωση των κόμβων στον αλγόριθμο LP μπορεί να γίνει με δύο τρόπους: σύγχρονα ή ασύγχρονα (synchronous updating ή asynchronous updating) [67]. Κατά το synchronous updating, ο κόμβος x λαμβάνει στην t -οστή επανάληψη την ετικέτα που έχει η πλειοψηφία των γειτόνων του κατά την $(t-1)$ -οστή επανάληψη, όπως ακριβώς περιγράφει το Βήμα 4. Ωστόσο, σε διμερή δίκτυα η σύγχρονη ανανέωση μπορεί να οδηγήσει σε ταλάντωση των ετικετών από το ένα σύνολο κόμβων στο άλλο (oscillation of labels) με αποτέλεσμα ο αλγόριθμος να παγιδευτεί σε ατέρμων βρόχο (Εικόνα8). Για να λυθεί το πρόβλημα αυτό οι δημιουργοί του αλγόριθμου συνιστούν σε διμερή δίκτυα asynchronous updating (Εικόνα9), όπου:

$C_x(t) = f(C_{x_{i1}}(t), \dots, C_{x_{im}}(t), C_{x_{i(m+1)}}(t-1), \dots, C_{x_{ik}}(t-1))$, με x_{i1}, \dots, x_{im} τους γείτονες του x που έχουν ήδη ανανεωθεί για την τρέχουσα επανάληψη του αλγορίθμου, ενώ $x_{i(m+1)}, \dots, x_{ik}$ είναι οι γείτονες που δεν έχουν ανανεωθεί κατά την τρέχουσα επανάληψη. Όπως ήδη αναφέρθηκε, η σειρά με την οποία επιλέγονται οι κόμβοι σε κάθε επανάληψη του αλγορίθμου είναι τυχαία.



Εικόνα 9 Το φαινόμενο της ταλάντωσης των ετικετών λαμβάνει χώρα, όταν ο αλγόριθμος LP με synchronous updating εφαρμόζεται σε διμερή δίκτυα [27]



Εικόνα 10 Αναπαράσταση του asynchronous updating. Σε κάθε επανάληψη, η σκίαση με κίτρινο δείχνει τον κόμβο του οποίου η ετικέτα ανανεώνεται [27]

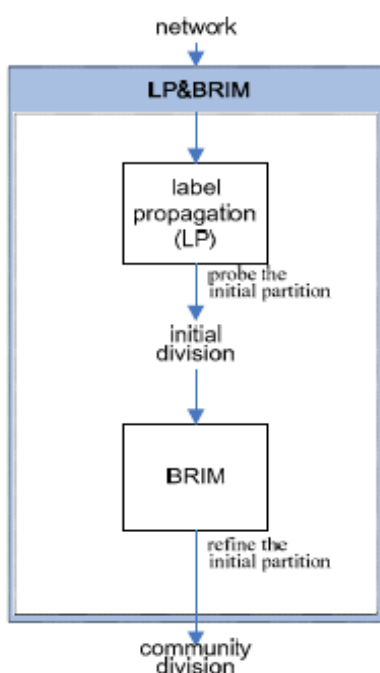
2.3.7 Murata's "LP&BRIM"

Δεν είναι δύσκολο να παρατηρήσει κανείς, ότι ο διαχωρισμός κοινοτήτων που προκύπτουν από τον αλγόριθμο BRIM εξαρτάται αυστηρά από την αρχική διαμέριση του δικτύου. Χωρίς παραβίαση της γενικότητας, διαπιστώνεται ότι μία καλή αρχική διαμέριση οδηγεί τον BRIM στον βέλτιστο διαχωρισμό, ενώ μία κακή αρχική διαμέριση τον οδηγεί στο να συγκλίνει γρήγορα και να παγιδευτεί σε τοπικό ελάχιστο. Ακόμη, όπως αναφέρθηκε, δεν είναι κατάλληλος για μεγάλης κλίμακας δίκτυα, διότι η αναζήτηση της κατάλληλης τιμής του c είναι εξαιρετικά χρονοβόρα καθώς οι πιθανές τιμές αυξάνουν γραμμικά σε σχέση με την κλίμακα του δικτύου.

Από την άλλη, ο αλγόριθμος LP (Label Propagation) αποτελεί καλή επιλογή για την ανάλυση δεδομένων που μοντελοποιούνται με τεχνικές δικτύων, και ακόμα καλύτερη επιλογή στην περίπτωση των big data. Ο κύριος λόγος είναι, πως ο αλγόριθμος αυτός είναι πάρα πολύ γρήγορος (algorithm for fast community detection). Ακόμη, δεν έχει απαιτήσεις υπολογιστικής ισχύος και αρχικής παραμετροποίησης. Τέλος, η ποιότητα των communities επαληθεύεται από τον ίδιο τον αλγόριθμο και όχι συγκρίνοντας με μοντέλο που προκύπτει από τυχαιοποίηση.

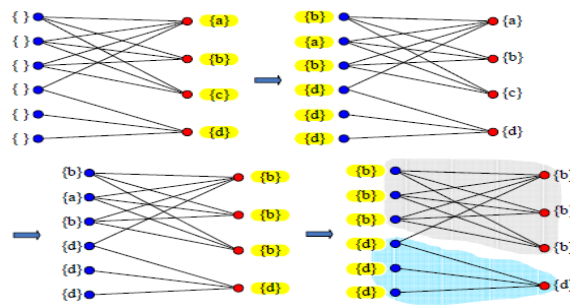
Ο συνδυασμός των δύο αλγορίθμων παρέχει αρχικά μέσω του LP μία διαμέριση με πολύ καλή ακρίβεια την οποία και διαδίδει όπως περιγράψαμε στην παράγραφο *c1*. Στη συνέχεια, καλεί τον BRIM για να επαληθεύσει την αρχική διαμέριση και πιθανώς να τη διορθώσει. Στο πρώτο μέρος του αλγορίθμου, ο LP προσπαθεί να βρει την αρχική διαμέριση με μικροκοσμική προσέγγιση: οι κοινότητες σχηματίζονται με φυσικό τρόπο από αυτοοργάνωση (they naturally merge), η οποία έχει ως καθοδήγηση τη δομή του δικτύου. Στο δεύτερο μέρος του αλγορίθμου, ο BRIM στοχεύει να αξιολογήσει την αρχική διαμέριση από μακροσκοπική σκοπιά: στοχεύει στη μεγιστοποίηση του global metric για το modularity κατά τη διάρκεια επαγωγής των ετικετών. Η ένωση των δύο

αλγορίθμων έχει ως αποτέλεσμα να εκμεταλλευτούν ο ένας τον άλλον με πολύ καλή ακρίβεια και αποτελεσματικότητα. Στην Εικόνα10 φαίνεται το διάγραμμα ροής του αλγορίθμου LP&BRIM.



Εικόνα 11 Διάγραμμα ροής του αλγορίθμου LP&BRIM [27]

Σύμφωνα με τον ορισμό κατά Barber [4], η κοινότητα σε ένα διμερές δίκτυο (Bipartite Community) είναι ένα σύνολο πυκνά συνδεδεμένων κόμβων. Επειδή, λοιπόν, δεν μπορεί να υπάρξει ακμή παρά μόνο μεταξύ κόκκινων και μπλε κόμβων και όχι μεταξύ κόμβων του ίδιου είδους, οποιαδήποτε κοινότητα άξια σημασίας για τον υπολογισμό του Bipartite Modularity θα πρέπει να περιέχει και κόκκινους και μπλε κόμβους ειδάλλως δε θα υπάρχει ακμή εντός του community. Ένεκα της ιδιότητας αυτής, το πλήθος των communities σε ένα διμερές δίκτυο είναι το πολύ ίσο με το πλήθος των κόκκινων (μπλε) κόμβων. Υπό το πρίσμα της απόρροιας αυτής, αντί ο αλγόριθμος να αναθέτει ετικέτες σε όλους τους κόμβους του δικτύου, μπορεί για διμερές δίκτυο να αναθέσει ετικέτες μόνο σε κάθε κόκκινο (μπλε) κόμβο αφήνοντας χωρίς ετικέτες (unlabeled) τους κόμβους του άλλου είδους. Εφόσον η μόνη πληροφορία που χρειάζεται ο κάθε κόμβος για να ανανεώσει το αναγνωριστικό του αφορά τους γειτονές του και εφόσον οι γείτονες κάθε κόκκινου (μπλε) κόμβου είναι μπλε (κόκκινοι), η διαδικασία της ανανέωσης των αναγνωριστικών γίνεται ως ακολούθως: ξεκινώντας από τους κόκκινους (ή τους μπλε) κόμβους, γίνεται διάδοση των αναγνωριστικών από τους κόκκινους προς τους μπλε για ανανέωση, έπειτα τα αναγνωριστικά διαχέονται ξανά στους κόκκινους κόμβους και η διαδικασία συνεχίζεται επαναληπτικά όπως στην εικόνα:



Εικόνα 12 Σχηματική αναπαράσταση του αλγορίθμου LP&BRIM [27]

Τέλος, στην Εικόνα 14 δίνεται σε απλή μορφή ψευδοκώδικα η λειτουργία του αλγορίθμου LP&BRIM, όπως παρέχεται από τους δημιουργούς του [27],[67]:

```

input: bipartite network G
output: community division
function:
  initiate each red node with a unique label;
  while (not every node has a label that is the maximum label of its neighbors)
    propagate labels from red nodes to blue nodes;
    propagate labels from blue nodes to red nodes;
  end while
divisionLP = the division corresponding to the largest Bipartitemodularity in the above while loop
divisionred = the division for red nodes in divisionLP;
divisionblue = the division for blue nodes in divisionLP;
inducingBlueFlag=true;
while (true)
  preBipartitemodularity = Bipartitemodularity(divisionred, divisionblue);
  if (inducingBlueFlag == true)
    inducing divisionblue from divisionred;
  else
    inducing divisionred from divisionblue;
  newBipartitemodularity = Bipartitemodularity(divisionred, divisionblue);
  if (newBipartitemodularity == preBipartitemodularity)
    break;
  else
    preBipartitemodularity = newBipartitemodularity;
    inducingBlueFlag = !inducingBlueFlag;
end while
division = (divisionred, divisionblue);
end function

```

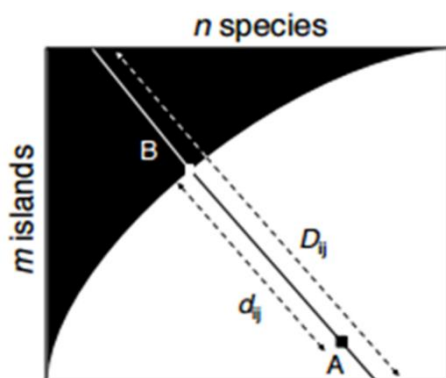
Εικόνα 13 Ο αλγόριθμος LP&BRIM σε μορφή ψευδοκώδικα [27]

2.4 Nestedness Detection Algorithms

2.4.1 NTC (Nestedness Temperature Calculator)

Η ιδέα αρχικά εφαρμόστηκε από τους Atman και Patterson [71] για να περιγράψει τις συγκεντρώσεις ειδών σε νησιά. Οι ίδιοι συγγραφείς το 1993 [72] παρουσιάζουν την πρώτη προσπάθεια υπολογισμού nestedness με τη μέθοδο NestednessNTC (Nestedness Temperature Calculator). Σκοπός των συγγραφέων ήταν να εκτιμήσουν την σειρά με την οποία μπορούσαν να εξαφανιστούν τα είδη (countdown to extinction). Όσο πιο nested είναι ένα δίκτυο, τόσο πιο βέβαιοι μπορούμε να είμαστε για την ακολουθία αυτή. Αρχικά, δηλαδή, αναμένεται να εξαφανιστούν οι specialists κοκ. Μεγαλύτερο nestedness σημαίνει περισσότερη τάξη και λιγότερη αβεβαιότητα ως προς τη σειρά εξαφάνισης των ειδών. Ορμώμενοι από αυτή την εννοιολογική σύνδεση, οι Atman και Patterson μίλησαν για θερμοκρασία (temperature) συστήματος, συνδέοντας έτσι τις έννοιες heat, order, expectedness και unexpectedness. Αναπόφευκτα έτσι αναφέρουν στο άρθρο τους και την εντροπία, τόσο τη θερμοδυναμική όσο και του Shannon. Για τους συγγραφείς, «κρύο» θεωρείται ένα highly nested δίκτυο, στο οποίο αφού οι συνδέσεις παρουσιάζουν μεγάλη τάξη (όπως συμβαίνει στα μόρια κρύων σωμάτων) και μικρή αβεβαιότητα ως προς τη σειρά εξαφάνισης (κατά μία έννοια μικρή εντροπία πληροφορίας). Από την άλλη ένα «ζεστό» δίκτυο παρουσιάζει συνδεσιμότητα με μεγάλη αταξία (όπως ένα ζεστό σώμα στη θερμοδυναμική) και μεγάλη αβεβαιότητα ως προς την εξαφάνιση (μεγάλη εντροπία Shannon).

Αν και ασαφώς ορισμένο από τη βιβλιογραφία σαν μέτρο εκτίμησης της εμφωλευσιμότητας, η τεχνική υπολογισμού της βασίζεται σε μία αναδιάταξη της μήτρας γειτνίασης, έτσι ώστε οι κατά το δυνατόν περισσότερες συνδέσεις να βρίσκονται πάνω από μία χαρακτηριστική καμπύλη που καλείται isocline (*isocline: καμπύλη που χωρίζει τις presences από τις absences ενός τέλει nested πίνακα ίδιου μεγέθους/πυκνότητας). Έτσι, χαρακτηρίζεται η ύπαρξη στοιχείου πάνω ή κάτω από αυτή την καμπύλη ως expected ή unexpected αντίστοιχα. Ο υπολογισμός της θερμοκρασίας του δικτύου βασίζεται στον υπολογισμό της μέσης απόστασης από ένα perfectly nested δίκτυο (αντιστοιχία στο isocline).



Εικόνα 14 Unexpectedness: η παρουσία κάτω από την isocline/απουσία πάνω από την isocline [72]

$$u_{ij} = \left(\frac{d_{ij}}{D_{ij}} \right)^2$$

$$U = \frac{1}{m * n} \sum \sum u_{ij} \quad T = \frac{100}{U_{max}} * U$$

2.4.3 NODF

Α. Το μέτρο NODF παρουσιάστηκε το 2009 (*Bastolla et al, 2009*) και αποτελεί ένα πολύ καλό μέτρο για τον υπολογισμό της εμφωλευσιμότητας. Τα αρχικά του σημαίνουν Nestedness Metric based on Overlap and Deacrising Fill. Παρακάτω αναλύεται ο τύπος που υπολογίζει το NODF και δίνεται ένα παράδειγμα για την καλύτερη κατανόησή του.

$$NODF = \frac{\sum N_{i,i'}}{\frac{n(n-1)}{2} + \frac{m(m-1)}{2}}, N_{i,i'} = DF_{i,i'} * PO_{i,i'} \quad (19)$$

Στην πιο πάνω σχέση το n είναι ο αριθμός των στηλών και το m είναι ο αριθμός των γραμμών, οπότε ο παρονομαστής εκφράζει το συνολικό πλήθος για τα πιθανά ζεύγη που ενώνονται με ακμή στο διμερές δίκτυο. Περαιτέρω ανάλυσης χρήζει ο αριθμητής, δηλαδή η επεξήγηση του όρου $N_{i,i'}$:

Όπως αναγράφεται και δίπλα από τη σχέση, $N_{i,i'} = D.F._{i,i'} * P.O._{i,i'}$

$PO_{i,i'}$ σημαίνει Percentage Overlap, δηλαδή τι ποσοστό των στοιχείων της στήλης i και i' επικαλύπτεται (overlap). Πιο συγκεκριμένα, για ένα διμερές δίκτυο του οποίου η μήτρα γειννίασης περιέχει μόνο 0 και 1 για τη δήλωση ύπαρξης ή απουσίας αλληλεπίδρασης, το PO_{12} εκφράζει το ποσοστό των "1" της στήλης 2 που εμπεριέχονται στη στήλη 1. Αυτό προκύπτει από το κλάσμα με αριθμητή το πλήθος των κοινών "1" στις δύο στήλες και παρονομαστή το πλήθος των πιθανών αλληλεπιδράσεων που θα μπορούσαν να υπάρχουν (πλήθος "1" στη στήλη 2).

Η τιμή του όρου DF ισούται με 100 ή με μηδέν και αυτό καθορίζεται ως εξής:

Ορίζουμε τον όρο $MT_i =$ Marginal Total (άθροισμα των στοιχείων της στήλης i)

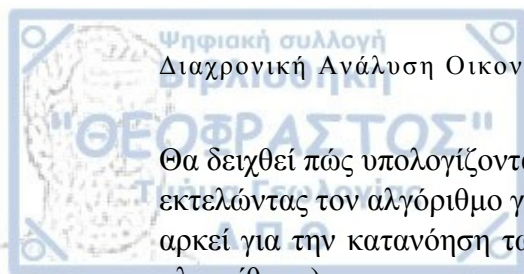
Δεδομένης της μήτρας γειννίασης ενός διμερούς δικτύου συγκρίνουμε κάθε γραμμή του πίνακα με κάθε άλλη ως εξής:

- Αν $MT_1 \geq MT_2$, τότε $DF=0$
- Αν $MT_1 < MT_2$, τότε $DF=100$.

Για την κατανόηση της άνωθεν διαδικασίας παραθέτουμε το παράδειγμα που ακολουθεί:

Έστω διμερές δίκτυο με μήτρα γειννίασης την εξής (τα μηδενικά που δηλώνουν την απουσία αλληλεπίδρασης μεταξύ των κόμβων r_i και c_j παραλείπονται):

	c1	c2	c3	c4	c5
r1	1		1	1	1
r2	1	1	1		
r3		1	1	1	
r4	1	1	1		
r5	1	1	1		



Θα δειχθεί πώς υπολογίζονται τα μεγέθη που αναφέρθηκαν για κατανόηση του τύπου, εκτελώντας τον αλγόριθμο για σύγκριση της πρώτης στήλης με άλλες δύο καθώς αυτό αρκεί για την κατανόηση των μεγεθών (δεν έχει νόημα η εκπόνηση ολόκληρου του αλγορίθμου).

c1	c2
1	
1	1
	1
1	1
1	1

Για τη στήλη c1 έχουμε $MT_1=4$, ενώ για τη στήλη c2 έχουμε επίσης $MT_2=4$. Άρα αφού $MT_1=MT_2$, έχουμε $DF_{1,2}=0$. Οι από κοινού αλληλεπιδράσεις, δηλαδή το ποσοστό των άσων που επικαλύπτονται στις δύο στήλες είναι $PO_{1,2}=3/4$, δηλαδή ποσοστό με αριθμητή τις από κοινού θέσεις που υπάρχει "1" και παρονομαστή το συνολικό πλήθος των πιθανών αλληλεπιδράσεων. Ας πάρουμε τώρα την επόμενη στήλη (c3) για σύγκριση με τη c1. Έτσι, προκύπτει ότι $N_{1,2}=0*3/4=0$.

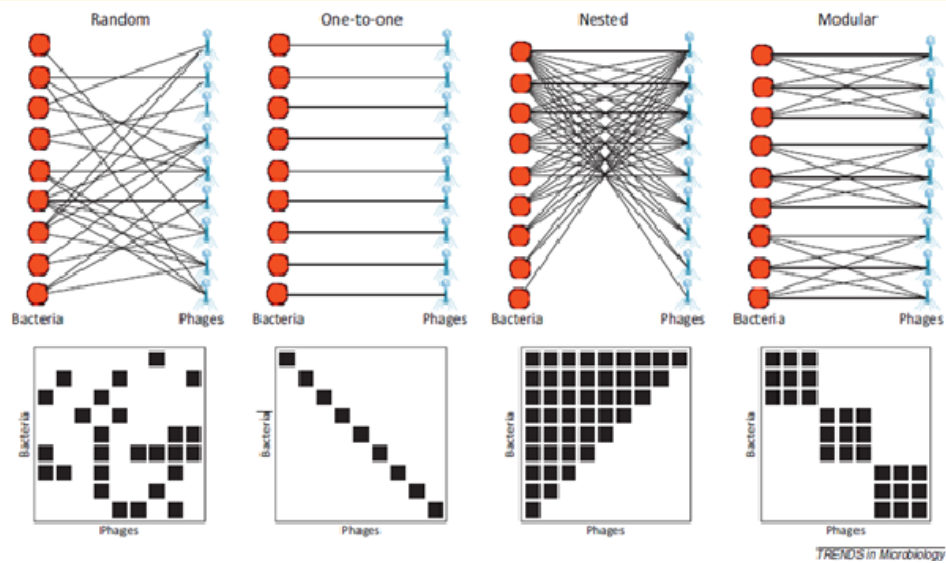
c1	c3
1	1
1	1
	1
1	1
1	1

Για τη στήλη c1, όπως και πριν, έχουμε $MT_1=4$, ενώ για τη στήλη c3 έχουμε $MT_3=5$. Έτσι, αφού $MT_1 < MT_3$, έχουμε $DF_{1,3}=100$. Εδώ το ποσοστό επικάλυψης είναι $PO=4/5$, δηλαδή 4 θέσεις, στις οποίες υπάρχει "1", διαιρούμενο με 5 που είναι το συνολικό πλήθος αλληλεπιδράσεων που θα μπορούσε να έχει η στήλη 3 με την 1 (αν η στήλη 1 δεν είχε κενή καταχώρηση στη θέση 3. Υπολογίζοντας τελικά το $N_{1,3}$ έχουμε $N_{1,3}=100*2/3=67$ (γίνεται πάντα στρογγυλοποίηση προς τα πάνω καθώς το N είναι ακέραιος πάντα).

Ο αλγόριθμος NODF συνεχίζει την διαδικασία για κάθε στήλη και τελικά υπολογίζει τον αριθμητή $\sum N_{i,i'}$ του κλάσματος που περιγράφει η σχέση (19) αθροίζοντας όλα τα N, για να καταλήξει τελικά στον υπολογισμό του συνολικού NODF για όλο το δίκτυο.

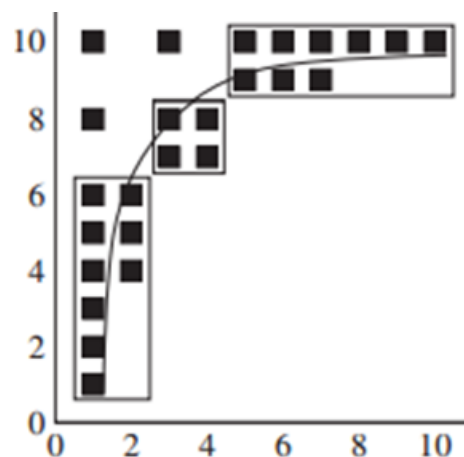
2.5 Ερμηνεία Nestedness και Modularity

Όπως φαίνεται και στην παραπάνω εικόνα, οι συνδέσεις ενός διμερούς δικτύου μπορεί να μη φανερώνουν συγκεκριμένο pattern (random), να είναι ένα προς ένα (one to one) αλλά και να φανερώνουν nested ή modular δίκτυα. Η βιβλιογραφία έχει δείξει ότι οι τελευταίες δύο ιδιότητες δεν είναι απαραίτητα αμοιβαίως αποκλειόμενες. Μία μελέτη 95 βιολογικών δικτύων διαφόρων ειδών οδήγησε στο συμπέρασμα ότι οι τιμές των δύο αυτών δεικτών μπορεί να έχουν θετικό συντελεστή συσχέτισης σε αραιά δίκτυα (Fortuna et al., 2010). Όμως καθώς η πυκνότητα του δικτύου αυξάνεται, οι δείκτες τείνουν να έχουν αρνητική συσχέτιση. Επομένως οι συγγραφείς μπόρεσαν να εντοπίσουν μόνο αραιά δίκτυα που είναι και nested και modular.



Εικόνα 15 Modularity και Nestedness διαφόρων δικτύων [Weitz et al. Trends Microbiology 21(2):82, 2013]

Οι δύο ιδιότητες αυτές ερμηνεύονται με διαφορετικό τρόπο ανάλογα με το είδος του δικτύου. Οι Fortuna et. al (2010) έχουν προβεί στην ακόλουθη βιβλιογραφική επισκόπηση όσον αφορά nestedness και modularity σε **mutualistic** δίκτυα. Έχειδειχθεί ότι σε τέτοια δίκτυα η nestedness μειώνει τον ανταγωνισμό και ενισχύει τη συνύπαρξη μεταξύ των ειδών (Bastolla et al., 2009). Επιπλέον, θεωρείται ιδιότητα που ενισχύει τη robustness του δικτύου απέναντι σε τυχαίες εξαφανίσεις ειδών (Burgos et al., 2007; Memmott, Waser, & Price, 2004) και τόπων κατοικίας ("habitats") τους (Fortuna & Bascompte, 2006). Η ιδιότητα modularity αποτελεί στη βιβλιογραφία ένδειξη σταθερότητας καθώς η επίδραση από διαταραχές (perturbations) απορροφάται μέσα στις κοινότητες και διαδίδεται σε όλο το δίκτυο (Krause, Frank, Mason, Ulanowicz, & Taylor, 2003; Teng & McCann, 2004).



Εικόνα 16 Nested & Modular δίκτυο [Fortuna et al., Journal of Animal Ecology, 79 811-817, 2010]

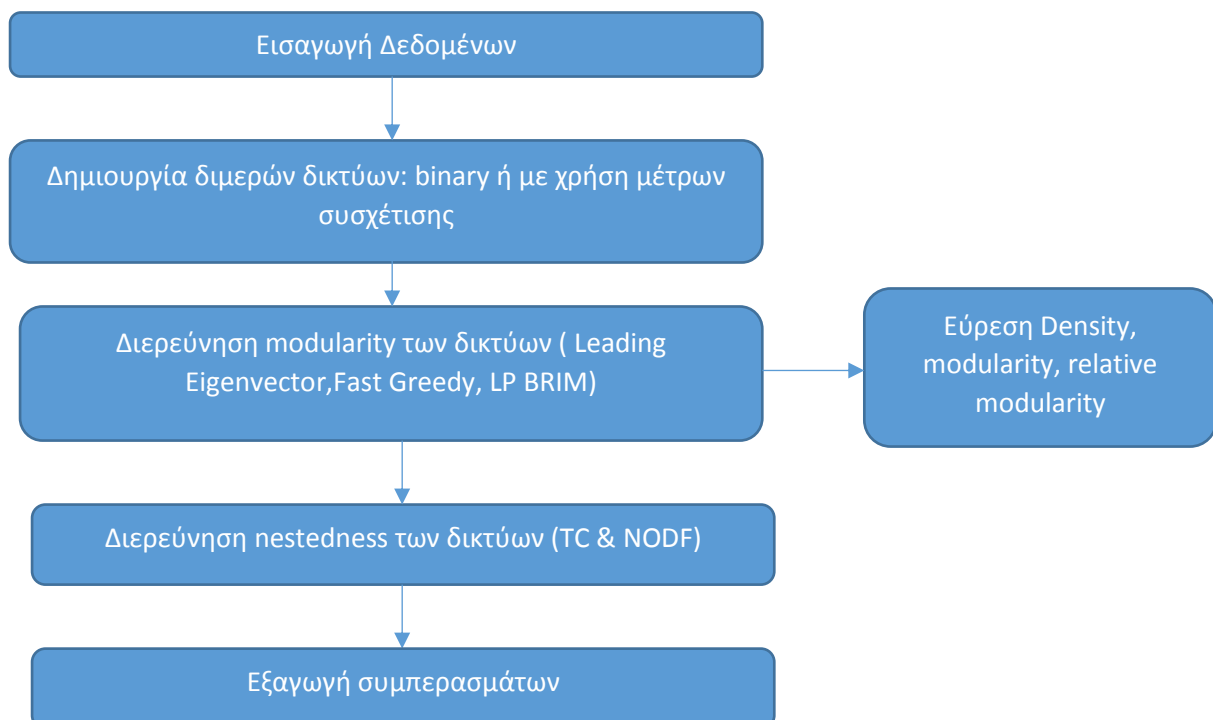
Επιπλέον έρευνα δείχνει πώς ερμηνεύονται οι δύο δείκτες σε antagonistic δίκτυα, και πιο συγκεκριμένα σε δίκτυα μόλυνσης (host-parasite). «Σε ένα τέλει nested δίκτυο, το πιο specialized παράσιτο θα μολύνει μόνο τον πιο δεκτικό σε μόλυνση host. Έπειτα, το αμέσως λιγότερο specializes θα μολύνει τον ίδιο host καθώς και άλλον έναν κ.ο.κ.» (Flores, Valverde & Weitz, 2013). «Σε ένα τέλει modular δίκτυο, δε θα υπάρχουν μολύνσεις μεταξύ παρασίτων μίας κοινότητας και hosts μιας άλλης» (Flores, Valverde & Weitz, 2013). Από τα παραπάνω οδηγούμαστε το συμπέρασμα ότι η nestedness μπορεί να ευνοήσει τη μαζική μόλυνση –ειδικά αν αυτή προέρχεται από generalist παράσιτο- ενώ η modularity δρα αποτρεπτικά απέναντι της. Το συμπέρασμα αυτό ήταν κομβικό για το ερώτημα της παρούσας διατριβής. Στις εφαρμογές παρακάτω παρομοιάζεται η βιολογική μόλυνση με την μετάδοση κρίσης στην οικονομία και το εμπόριο.

2.5 Σχηματική Περίληψη της μεθοδολογίας

Συνοπτικά η μεθοδολογία που θα ακολουθεί παρουσιάζεται στο παρακάτω σχήμα. Η μεθοδολογία αυτή θα εφαρμοστεί:

- α) Σε δεδομένα από το αμερικάνικο χρηματιστήριο και πιο συγκεκριμένα σε εταιρίες από τους δείκτες S&P 100 και S&P 500, για τα έτη 2000 έως 2015
- β) Σε δεδομένα παγκόσμιου εμπορίου σε σχέση με τις εξαγωγές και τις εισαγωγές όλων των κρατών, για τα έτη 2000 έως 2014

Εικόνα 17 Σχηματική αναπαράσταση της προτεινόμενης μεθοδολογίας



3 Ανάλυση

3.1 Προέλευση και παρουσίαση των Data Set

Στην εργασία γίνεται ανάλυση 2 διαφορετικών ειδών διμερών δικτύων από τα χρηματοοικονομικά.

Αρχικά αναλύθηκαν temporal (διαχρονικά) διμερή δίκτυα **μετοχών εταιριών και χρηματοοικονομικών οργανισμών**. Γνωρίζοντας ότι η κρίση του 2007-2008 προκλήθηκε από τον χρηματοοικονομικό τομέα, μελετάται διμερές δίκτυο στο οποίο το ένα ανεξάρτητο σετ κόμβων αποτελείται από χρηματοοικονομικές εταιρίες, ενώ το άλλο από εταιρίες κάθε είδους εκτός από χρηματοοικονομικές. Τα διμερή αυτά δίκτυα δημιουργήθηκαν κατά αναλογία με τα βιολογικά host parasite δίκτυα. Η λογική της μεθόδου βασίζεται, όπως δηλώθηκε και προηγουμένως, στον παραλληλισμό του βακτηριακού infection με τη μετάδοση μιας κρίσης στην οικονομία. Το ρόλο των parasites έχουν επομένως οι χρηματοοικονομικοί οργανισμοί και των hosts οι μετοχές όλων των άλλων ειδών. Εξετάζεται η διαχρονική εξέλιξη του δικτύου αυτού όσον αφορά τις ιδιότητες nestedness και modularity. Μία σημαντική αλλαγή στους δείκτες αυτούς ερμηνεύεται ως μια αλλαγή στο πόσο μαζική μπορεί να είναι μια μετάδοση σε όλη την αγορά σε περίπτωση που κάποιος ή κάποιοι χρηματοοικονομικοί οργανισμοί αντιμετωπίζουν προβλήματα και ως εκ τούτου δουν τις χρηματιστηριακές τιμές τους να σημειώνουν ραγδαία πτώση.

Οι εταιρίες-hosts είναι αυτές που το 2008 συμπεριλαμβάνονταν στον δείκτη υψηλής κεφαλαιοποίησης *S&P 100* και δεν ήταν τράπεζες. Η μία ομάδα κόμβων αποτελείται, δηλαδή, από εταιρίες διαφόρων ειδών οι οποίες, όμως, είναι «βαριά χαρτιά» και αντιπροσωπεύουν έτσι σε μεγάλο βαθμό ολόκληρη την αγορά. Από την άλλη βρίσκονται όλες οι εταιρίες που σύμφωνα με την *Global Industry Classification Standard (GICS)* της *Standard & Poor's* ανήκουν στην κατηγορία *financials* και ήταν το 2008 στον δείκτη *S&P 500*. Η δεύτερη ομάδα εταιριών (τα parasites) προέρχεται, δηλαδή, από τον κλάδο που πυροδότησε την κρίση. Είναι 68 μέχρι και το 8^ο δίκτυο και στη συνέχεια είναι 67 λόγω της χρεοκοπίας της *Lehman Brothers*. Στην πραγματικότητα τα διμερή αυτά δίκτυα είναι υπογραφήματα των πραγματικών δικτύων, αφού πρακτικά υπάρχει σχέση και μεταξύ των κόμβων και στις δύο ομάδες. Όμως, η απεικόνιση των σχέσεων αυτών δε θα βοηθούσε στο συγκεκριμένο ερώτημα. Γνωρίζοντας δηλαδή ότι η κρίση του 2008 προήλθε από τον χρηματοοικονομικό τομέα, είναι επιθυμητό τα δίκτυα να απεικονίζουν μόνο την επίδραση αυτού του τομέα στην υπόλοιπη αγορά. Τα δεδομένα αφορούν ημερήσιες χρηματιστηριακές τιμές από τις 12/9/2000 μέχρι τις 11/9/2015. Το χρονικό εύρος επιλέχθηκε έτσι ώστε να περιλαμβάνει 2 κρίσεις: τη χρηματοοικονομική του 2008 αλλά και την dot com που οριοθετείται από το 2000 μέχρι το 2002 και δεν προήλθε από τον χρηματοοικονομικό τομέα, αλλά από τον τεχνολογικό. Πηγή των δεδομένων είναι η βάση δεδομένων του *Yahoo Finance*. Τα δεδομένα έχουν χωριστεί σε 15 ίσα χρονικά παράθυρα, ένα για κάθε χρόνο.

Οι ακμές στα binary (0-1) δίκτυα δείχνουν επίδραση από τις χρηματοοικονομικές (financial) στις υπόλοιπες εταιρίες. Για το λόγο αυτό οι ακμές υπάρχουν για στατιστικά σημαντικές συσχετίσεις των λογαριθμικών αποδόσεων μιας financial στο χρόνο t-1 με τις αποδόσεις μιας από τις εταιρίες των υπόλοιπων ειδών στο χρόνο t. Επιλέχθηκαν οι λογαριθμικές αποδόσεις των τιμών και όχι οι ίδιες οι τιμές καθώς οι τελευταίες χαρακτηρίζονται από μεγάλη αυτοσυσχέτιση η οποία δημιουργεί ψευδείς συσχετίσεις (spurious correlations). Οι λογαριθμικές αποδόσεις υπολογίζονται σύμφωνα με τη σχέση (1), ενώ ακολουθεί ο συντελεστής Pearson (2)

$$(1) \quad \ln \frac{P_t}{P_{t-1}}$$

$$(2) \quad \rho(\text{fin}_{t-1}, \text{firm}_t) = \frac{\text{cov}(\text{fin}_{t-1}, \text{firm}_t)}{\sigma_{\text{fin}_{t-1}} \sigma_{\text{firm}_t}}$$

, όπου:

- p_t : η ημερήσια τιμή μιας μετοχής στο χρόνο t
- p_{t-1} : η ημερήσια τιμή μιας μετοχής στο χρόνο t-1
- cov_{ij} : συνδιακύμανση
- σ_i : τυπική απόκλιση
- fin_{t-1} : η ημερήσια απόδοση μιας financial μετοχής στο χρόνο t-1
- firm_t : η ημερήσια απόδοση μιας μη financial μετοχής στο χρόνο t

Οι συνδέσεις και με τα δύο μέτρα υποβλήθηκαν σε μη παραμετρικό έλεγχο στατιστικής σημαντικότητας ($p=0.05$). Ο μη παραμετρικός έλεγχος θεωρήθηκε προτιμητέος, διότι δεν υποθέτει κάποια συγκεκριμένη κατανομή και οι παρατηρήσεις έχουν σχετικά μικρό αριθμό ανά δίκτυο. Αφού εξήχθησαν τα διμερή αυτά δίκτυα, εξετάστηκαν διαχρονικά οι ιδιότητες nestedness και modularity. Για ανίχνευση κοινοτήτων αξιοποιήθηκαν οι αλγόριθμοι Leading Eigenvector, fast-greedy και LP&Brim, ενώ για nestedness detection εφαρμόστηκαν οι αλγόριθμοι NestednessNTC και NestednessNODF.

Η ίδια ακριβώς ανάλυση εφαρμόστηκε και στα δίκτυα που προέκυψαν από το Dataset του Παγκόσμιου Εμπορίου. Πηγή των δεδομένων αποτέλεσε το «The observatory of economic complexity» του MIT (<http://atlas.media.mit.edu>) και είναι διαθέσιμα ανοιχτά και δωρεάν. Τα δεδομένα αφορούν το ύψος των εισαγωγών και των εξαγωγών για το σύνολο των κρατών ανά την υφήλιο για τα έτη 2000 έως 2014. Τα εμπορεύματα περιγράφονται με βάση το Harmonized System (HS) του Διεθνούς οργανισμού Τελωνείων (World Customs Organization – WCO) και χωρίζονται σε 15 κατηγορίες, οι οποίες παρουσιάζονται στον Πίνακα 1. Τα χρονικά διμερή δίκτυα που σχηματίστηκαν ανά έτος έχουν ως κόμβους τις χώρες και τις κατηγορίες των προϊόντων. Ανάμεσα στα δύο αυτά είδη κόμβων, υπάρχει ακμή μόνο εάν το εκάστοτε κράτος εισήγαγε ή εξήγαγε προϊόντα από την αντίστοιχη κατηγορία. Σε αντιστοιχία με τα δεδομένα των μετοχών, θεωρούμε hosts τα κράτη και parasites τα προϊόντα. Συνεπώς, στην περίπτωση αυτή θα χρειαστεί να εξετασθούνε δύο διμερή δίκτυα: imports και exports, δηλαδή ένα διμερές δίκτυο για τις εισαγωγές και ένα για τις

εξαγωγές. Στόχος είναι να εξετασθεί κατά πόσο και με τι τρόπο η οικονομική κρίση του 2008 επηρέασε το παγκόσμιο εμπόριο και τις διεθνείς εισαγωγές-εξαγωγές. Τα δίκτυα δημιουργήθηκαν με τους ίδιους δείκτες συσχέτισης (Pearson Correlation και Spearman Correlation), καθώς επίσης και οι ακμές (μη παραμετρικός έλεγχος στατιστικής σημαντικότητας $p=0.05$). Οι ιδιότητες της εμφωλευσιμότητας και της κοινοτικοποίησης εξετάστηκαν με τους ίδιους αλγορίθμους.

Πίνακας 1 Προϊόντα στο σύστημα ταξινόμησης HS

Κατηγορίες προϊόντων με βάση το Harmonized System
Ζώα και Ζωικά προϊόντα
Φυτικά προϊόντα
Τροφές
Ορυκτά προϊόντα
Χημικά και συναφείς βιομηχανίες
Πλαστικά / Ελαστικές ύλες
Δέρματα ζώων, γούνες
Ξύλο και προϊόντα ξύλου
Προϊόντα κλωστοϋφαντουργίας
Υποδήματα
Πέτρα / Γυαλί
Μέταλλα
Μηχανήματα / ηλεκτρικά είδη
Προϊόντα σχετιζόμενα με την μετακίνηση
Διάφορα

Η ανάλυση όλων των δικτύων έγινε με χρήση των λογισμικών RStudio και Matlab, καθώς και του εργαλείου BiMat [7]. Πρόκειται για ένα εργαλείο που δημιουργήθηκε από ειδικούς αναλυτές δικτύων το 2014 για βιολογικά και οικολογικά δίκτυα και δημοσιεύθηκε το 2015 [8]. Έκτοτε, χρησιμοποιείται για ανάλυση διμερών δικτύων γενικότερα. Είναι ένα ελεύθερο λογισμικό, κυρίως για ανάλυση διμερών δικτύων, παρέχοντας όμως και κάποιες δυνατότητες και για ανάλυση μονομερών. Το BiMat παρέχει τη δυνατότητα στατικών αναλύσεων, δηλαδή ανάλυση για ένα δίκτυο κατά την οποία ο χρήστης μπορεί να εξάγει δεδομένα για διάφορα μέτρα όπως modularity, nestedness, Q_b values, και πολλά άλλα. Επίσης, υπάρχει η δυνατότητα διαχρονικών αναλύσεων, δηλαδή η ανάλυση πολλών δικτύων στην πάροδο του χρόνου είτε αυτά είναι διαφορετικά δίκτυα είτε το ίδιο δίκτυο σε πολλές χρονικές στιγμές. Αυτή η δυνατότητα παρέχει στο χρήστη το σημαντικό πλεονέκτημα για στατιστική ανάλυση δικτύων καλώντας απευθείας εντολές, είτε με τεχνικές αντικειμενοστραφούς προγραμματισμού (OOP-Object Oriented Programming approach), είτε με απλή κλήση συναρτήσεων (Functional Approach) [7],[8]. Με κλήση συναρτήσεων όπως: `internal.statistics()`; `MetaAnalysis()`; `MetaStatistics()`; επιστρέφονται στο χρήστη τιμές για το χαρακτηρισμό του δικτύου, που έχει δοθεί ως είσοδος, αναφορικά με nestedness και modularity. Μάλιστα, οι τιμές αυτές συγκρίνονται αυτόματα από το πακέτο BiMat με τιμές nestedness και modularity που έχουν προκύψει με στατιστική σημαντικότητα για n τυχαία δίκτυα που έχει ορίσει ο χρήστης (null model). Ο αριθμός n είναι ο αριθμός των επαναλήψεων (replicates) που θα τρέξει ο αλγόριθμος για τα τυχαία δίκτυα. Σε κάθεμια από αυτές θα προκύπτει ένα τυχαίο δίκτυο και τελικά από όλα τα τυχαία δίκτυα θα προκύπτουν τιμές nestedness & modularity οι οποίες χρησιμεύουν

στο να εκτιμηθεί η στατιστική σημαντικότητα των τιμών του υπό εξέταση πραγματικού δικτύου με έναν two-tailed έλεγχο. Πιο συγκεκριμένα, θέτοντας ένα συγκεκριμένο p-value, το δίκτυο θεωρείται modular ή nested αν η αντίστοιχη τιμή είναι μεγαλύτερη από το $(100-p/2)\%$ των τυχαίων παρατηρήσεων. Anti-modular ή anti-nested θεωρούμε το δίκτυο, αν οι τιμές αυτές είναι μικρότερες από το $(100-p/2)\%$ των τυχαίων τιμών. Παρόμοια συμπεράσματα μπορούν να εξαχθούν με χρήση z-scores. Η οπτικοποίηση των εν λόγω μεγεθών γραφικά γίνεται με κλήση συναρτήσεων όπως: `bipartite.plotter()`; `mstat.MetaStatistics.Plotter()`; `mstat.bipartite.printer()`;

Για τον έλεγχο της στατιστικής σημαντικότητας των τιμών για nestedness και modularity, χρησιμοποιήθηκε null model που δημιουργεί 500 τυχαία δίκτυα παρόμοιου μεγέθους και πυκνότητας με το παρατηρούμενο. Το μοντέλο κάνει έναν two-tailed έλεγχο, όπως περιγράφηκε παραπάνω, με $p=5\%$.

Το null model που επιλέχθηκε, υπολογίζει τα τυχαία δίκτυα με σεβασμό στο πλήθος των ακμών και τους βαθμούς των κορυφών (AverageNullModel) [7],[8]. Στο null model η πιθανότητα ύπαρξης ακμής είναι η εξής:

$$P_{ij} = \frac{\left(\frac{k_i}{n} + \frac{d_j}{m}\right)}{2},$$

όπου:

- k_i : ο βαθμός της κορυφής i στις στήλες
- d_j : ο βαθμός της κορυφής j στις γραμμές
- n : ο αριθμός των κορυφών στις γραμμές
- m : ο αριθμός των κορυφών στις στήλες

3.2 Διαχρονικά Δίκτυα μετοχών του χρηματιστηρίου (S&P500)

Για να γίνουν περισσότερο αντιληπτά τα αποτελέσματα που περιγράφονται στη συνέχεια, η παρακάτω εικόνα απεικονίζει την πορεία του βασικότερου χρηματιστηριακού δείκτη της Αμερικής για την εξεταζόμενη περίοδο. Βλέπουμε ότι από το 2000 μέχρι το 2002 το χρηματιστήριο σημειώνει έντονα πτωτική πορεία. Αυτό οφείλεται σε μία από τις 2 κρίσεις που το σημάδεψαν τον αιώνα αυτό, την κρίση dot com τα αίτια της οποίας βρίσκονται στη φούσκα που είχε δημιουργηθεί από την υπερβολική προτίμηση των επενδυτών σε εταιρίες που σχετίζονταν με την τεχνολογία και κυρίως το ίντερνετ. Από το 2003 μέχρι το τέλος του 2007 το χρηματιστήριο σημειώνει ανοδική πορεία. Αυτή ανακόπτεται στις αρχές του 2008, ενώ από τον Σεπτέμβριο του ίδιου έτους σημειώνεται ραγδαία πτώση λόγω της κατάρρευσης της Lehman Brothers.



Εικόνα 18 Πορεία του δείκτη S&P500 στο χρόνο

Από το 2009 και μετά, ύστερα και από τις παρεμβάσεις του κράτους και της κεντρικής τράπεζας των ΗΠΑ (Federal Reserve), το χρηματιστήριο αρχίζει να παίρνει και πάλι την ανιούσα, με την κρίση να περνάει στην Ευρωζώνη και τυχόν διορθώσεις (βραχυπρόθεσμες διακοπές της ανοδικής τάσης) του S&P 500 να οφείλονται εν πολλοίς στον ευρωπαϊκό παράγοντα.

Πίνακας 2 Πυκνότητα δικτύου ανά έτος

year	density
2000	0.29641
2001	0.27974
2002	0.35741
2003	0.40773
2004	0.34662
2005	0.32843
2006	0.34956
2007	0.34706
2008	0.44292
2009	0.44651
2010	0.33802
2011	0.35741
2012	0.32124
2013	0.39281
2014	0.33649
2015	0.30512

Σημαντική για την ανάλυση είναι η μελέτη της πυκνότητας κάθε δικτύου. Από τον Πίνακα 2 φαίνεται ότι υψηλότερα σε πυκνότητα είναι τα δίκτυα 07-08, 08-09 -δηλαδή

αυτό που προηγείται της χρηματοοικονομικής κρίσης του 2008 και το δίκτυο της κρίσης- αλλά και το 10-11, για το οποίο γίνεται ειδική μνεία στη συνέχεια.

Modularity of stock trade network (P-LE)

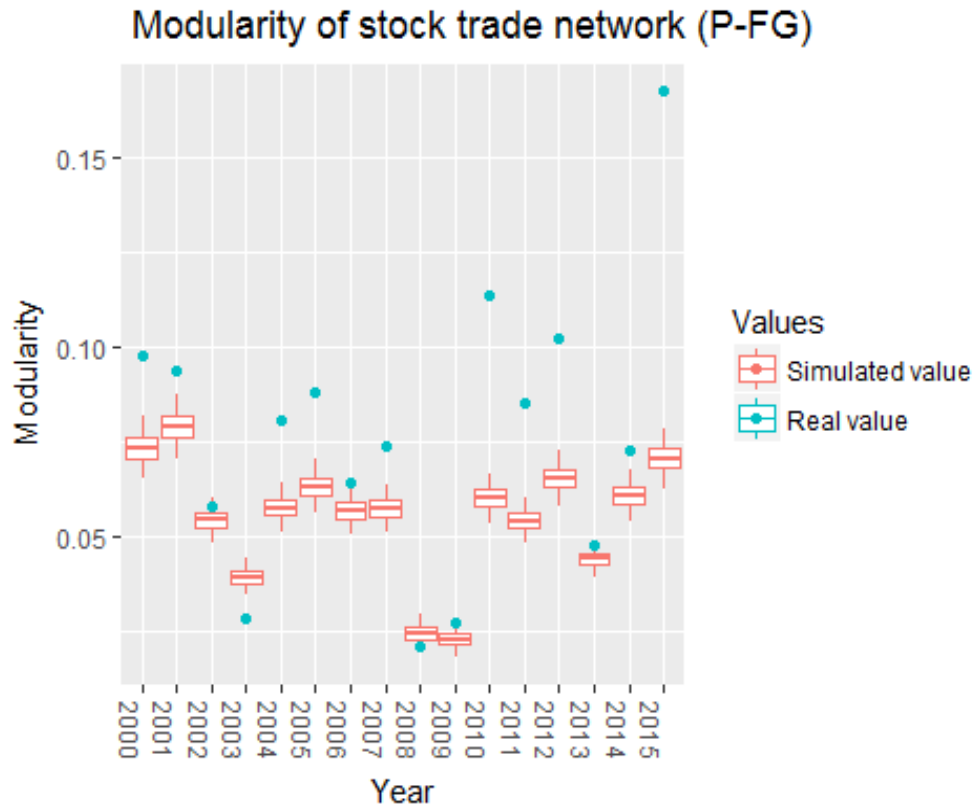


Εικόνα 19 Τιμές modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα μετοχών

Πίνακας 3 Τιμές modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα μετοχών

Leading Eigenvector - Modularity

year	Real	mean	sd	Percent	z_score
2000	0.0998449	0.0655640	0.0041699	100.0	0.62248197
2001	0.1067485	0.0706929	0.0051403	100.0	0.79484763
2002	0.0693675	0.0478935	0.0040245	100.0	-0.13846250
2003	0.0284116	0.0343301	0.0040747	6.4	-1.16102761
2004	0.0540746	0.0507897	0.0036156	83.6	-0.52028700
2005	0.0839581	0.0559786	0.0039190	100.0	0.22582919
2006	0.0682720	0.0498739	0.0039969	100.0	-0.16581285
2007	0.0799393	0.0508848	0.0035638	100.0	0.12549101
2008	0.0158022	0.0232293	0.0052517	9.0	-1.47585230
2009	0.0133966	0.0219297	0.0056188	8.4	-1.53591399
2010	0.1229293	0.0531306	0.0038967	100.0	1.19884210
2011	0.0790808	0.0479655	0.0035554	100.0	0.10405475
2012	0.0950743	0.0578126	0.0039972	100.0	0.50337363
2013	0.0424521	0.0383604	0.0037316	90.4	-0.81047198
2014	0.0717323	0.0537162	0.0042444	100.0	-0.07941952
2015	0.1675267	0.0627674	0.0042016	100.0	2.31232747



Εικόνα 20 Τιμές modularity με τον αλγόριθμο fast-greedy για τα δίκτυα μετοχών

Πίνακας 4 Τιμές modularity με τον αλγόριθμο fast-greedy για τα δίκτυα μετοχών

Fast Greedy - Modularity

year	Real	mean	sd	Percent	z_score
2000	0.0978673	0.0733342	0.0027802	100.0	0.57988156
2001	0.0937571	0.0789208	0.0028585	100.0	0.46922770
2002	0.0578135	0.0542340	0.0020329	94.8	-0.49842154
2003	0.0282263	0.0391414	0.0015785	0.0	-1.29495081
2004	0.0804610	0.0574248	0.0021725	100.0	0.11128060
2005	0.0882577	0.0630270	0.0023807	100.0	0.32117649
2006	0.0641843	0.0566584	0.0021684	100.0	-0.32691299
2007	0.0739539	0.0572502	0.0021288	100.0	-0.06389973
2008	0.0206538	0.0242643	0.0017525	0.8	-1.49881328
2009	0.0267329	0.0226388	0.0016090	98.6	-1.33515449
2010	0.1135784	0.0599317	0.0022342	100.0	1.00284591
2011	0.0852349	0.0540708	0.0020232	100.0	0.23979862
2012	0.1024317	0.0652805	0.0024640	100.0	0.70276245
2013	0.0473863	0.0438427	0.0016958	98.3	-0.77913677
2014	0.0727137	0.0606445	0.0022762	100.0	-0.09728820
2015	0.1679871	0.0703654	0.0025779	100.0	2.46760449

Modularity of stock trade network (P-LPBrim)



Εικόνα 21 Τιμές modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα μετοχών

Πίνακας 5 Τιμές modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα μετοχών

lpbrim - Modularity

year	Real	mean	sd	Percent	z_score
2000	0.1075742	0.0761656	0.0035632	100.0	0.62990405
2001	0.1098004	0.0825067	0.0034457	100.0	0.68712333
2002	0.0748468	0.0565425	0.0027318	100.0	-0.21125519
2003	0.0270225	0.0406069	0.0022035	0.0	-1.44043399
2004	0.0820120	0.0597385	0.0028869	100.0	-0.02709482
2005	0.0871257	0.0650211	0.0032470	100.0	0.10433726
2006	0.0760730	0.0589546	0.0027744	100.0	-0.17973847
2007	0.0917040	0.0595361	0.0027342	100.0	0.22200864
2008	0.0221621	0.0296049	0.0014941	0.0	-1.56535620
2009	0.0285792	0.0282814	0.0014366	59.2	-1.40042366
2010	0.1262961	0.0621383	0.0030534	100.0	1.11109435
2011	0.0852995	0.0567698	0.0026701	100.0	0.05740021
2012	0.1148786	0.0675547	0.0033416	100.0	0.81764305
2013	0.0476740	0.0450004	0.0024377	88.4	-0.90964883
2014	0.0799978	0.0625136	0.0030766	100.0	-0.07886384
2015	0.1680131	0.0728115	0.0034606	100.0	2.18330412

Όσον αφορά το modularity, παρατηρούμε ότι οι τιμές δε διαφέρουν σημαντικά μεταξύ των τριών αλγορίθμων ενώ ακολουθούν και παρόμοια πορεία στο χρόνο. Γενικά το

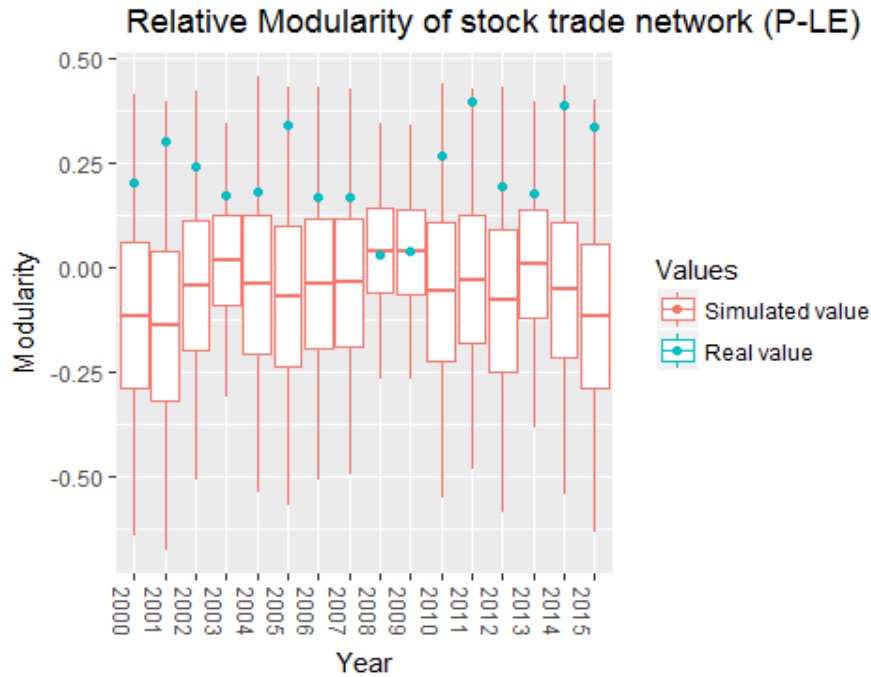
modularity είναι υψηλότερο σε περιόδους που το χρηματιστήριο ανεβαίνει και χαμηλότερο στις περιόδους των 2 κρίσεων. Οι τιμές αυτές όμως δεν είναι στατιστικά σημαντικές, αφού βρίσκονται μέσα στα όρια των αντίστοιχων τιμών των τυχαίων δικτύων. Οι μεταβολές δηλαδή του modularity αντικατοπτρίζουν εν πολλοίς τις μεταβολές στην πυκνότητα των δικτύων, η οποία είναι μεγαλύτερη στις περιόδους των κρίσεων. Σε τέτοιες περιόδους οι αποδόσεις περισσότερων μετοχών συσχετίζονται λόγω του αρνητικού οικονομικού περιβάλλοντος που δρα ως κοινός παράγοντας.

Ο Leading Eigenvector και δευτερευόντως ο fast-greedy εμφανίζουν ένα ιδιαίτερα ενδιαφέρον στοιχείο. Τα δίκτυα 3, 7, 8, 9 είναι ή προσεγγίζουν τον χαρακτηρισμό του anti-modular. Εμφανίζουν δηλαδή τιμές modularity μικρότερες ή κοντά στο 2,5% των χαμηλότερων παρατήρησεων. Δεν υπάρχει, επομένως, κανένας σχηματισμός κοινοτήτων στα δίκτυα αυτά. Τα δίκτυα αυτά, όπως έχει ήδη δείχθει, ανήκουν στην κρίση dot com (3) και στην κρίση του 2008 (7,8,9). Το anti-modularity μας δείχνει ότι η τοπολογία των δικτύων αυτών είναι τέτοια που δε μπορεί να απορροφήσει μια κρίση αν αυτή προκύψει. Υπάρχει όμως μια βασική διαφορά μεταξύ των δύο κρίσεων. Το δίκτυο 3 χρονολογείται μόλις στο τέλος της dot com και απεικονίζει ουσιαστικά το μεταβατικό στάδιο μέχρι η αγορά να ξαναπάρει τα πάνω της. Στο σημείο αυτό, δηλαδή, η κρίση έχει ήδη διαδοθεί από τον τεχνολογικό τομέα σε όλη την αγορά, οι αλληλεξαρτήσεις μεταξύ των μετοχών έχουν πληθύνει και επομένως ακόμη και ένα δίκτυο με χρηματοοικονομικές μετοχές στη θέση των hosts έχει πλέον τοπολογία που ευνοεί τη διάδοση κρίσης. Και ενώ στην κρίση dot com εντοπίζεται τέτοιο δίκτυο αμέσως μετά το πέρας της, το δίκτυο 7 που αντιστοιχεί στο 2006-2007 έχει διαμορφωθεί με παρόμοιο τρόπο αρκετά πριν εμφανιστεί η κρίση. Αν και το χρηματιστήριο την περίοδο αυτή ακόμη ανεβαίνει, οι συσχετισμοί στην αγορά αλλάζουν. Το ειδικά διαμορφωμένο δίκτυο αφομοιώνει τις πληροφορίες για την κρίση στα subprime δάνεια και ήδη από το 2006-2007 δείχνει ότι η τοπολογία του δικτύου είναι τέτοια που αν κάποιος κόμβος-τράπεζα αντιμετωπίσει προβλήματα, αυτά είναι εύκολο να διαδοθούν στην υπόλοιπη αγορά και δε θα απορροφηθούν σε κάποια κοινότητα.

Στη διαφορά που μόλις αναφέρθηκε, συνίσταται και η αξία της μεθόδου. Αν η εικόνα των δικτύων ήταν παρόμοια στις δύο κρίσεις, η μέθοδος δε θα προσέδιδε στην ουσία τίποτα. Το γεγονός, όμως, ότι το έχον αποκλειστικά χρηματοοικονομικές μετοχές σε ένα από τα δύο set δίκτυο γίνεται anti-modular μόνο μετά το τέλος μιας μη χρηματοοικονομικής -αλλά αρκετά πριν την αρχή μιας χρηματοοικονομικής κρίσης- είναι μια πολύ ισχυρή ένδειξη ότι η μέθοδος έχει πραγματική αξία: την απεικόνιση της δυνατότητας διάδοσης μιας κρίσης από έναν συγκεκριμένο κλάδο της οικονομίας.

Συγκριτικές Τιμές Relative modularity (Qr) για τα δίκτυα μετοχών

Σημαντικό είναι να παρατηρήσουμε για τα προαναφερθέντα δίκτυα τις τιμές του μέτρου Q_r , που χρησιμεύει στην αξιολόγηση των κοινοτήτων εν συγκρίσει με μία σειρά από τυχαία δίκτυα (simulated values) με σεβασμό στο πλήθος των κόμβων και ακμών.



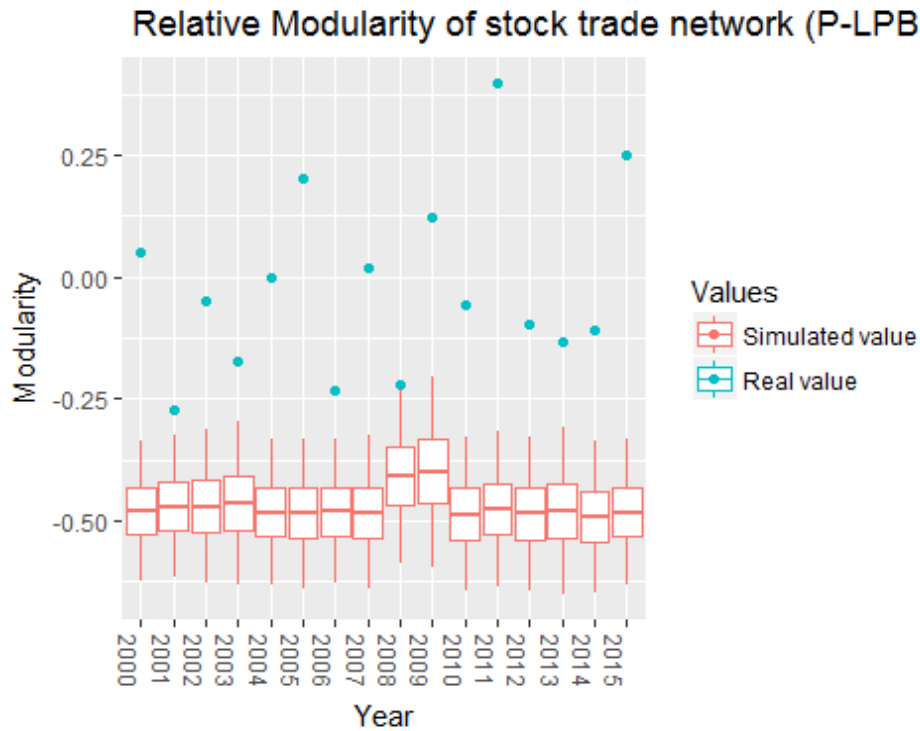
Εικόνα 22 Τιμές Qr με τον αλγόριθμο Leading Eigenvector για τα δίκτυα μετοχών

Leading Eigenvector

year	Real	mean	sd	Percent	z_score
2000	0.2039691	-0.1138714	0.1761182	100.0	-0.2043623
2001	0.3006231	-0.1392188	0.1790321	99.8	0.6874632
2002	0.2429137	-0.0417153	0.1555703	100.0	0.1549796
2003	0.1728560	0.0188945	0.1088350	100.0	-0.4914431
2004	0.1829038	-0.0395286	0.1659956	99.8	-0.3987317
2005	0.3406302	-0.0690687	0.1669937	100.0	1.0566089
2006	0.1692116	-0.0389878	0.1562769	99.9	-0.5250700
2007	0.1688638	-0.0351965	0.1535325	100.0	-0.5282793
2008	0.0319725	0.0406950	0.1017550	12.5	-1.7913751
2009	0.0392779	0.0375926	0.1016199	23.1	-1.7239681
2010	0.2690944	-0.0569874	0.1651150	99.9	0.3965486
2011	0.3983542	-0.0277123	0.1513428	100.0	1.5892278
2012	0.1963377	-0.0776202	0.1697202	99.9	-0.2747771
2013	0.1774820	0.0085324	0.1297156	99.9	-0.4487592
2014	0.3881515	-0.0526475	0.1626746	100.0	1.4950880
2015	0.3352374	-0.1144955	0.1723738	100.0	1.0068498

Πίνακας 6. Τιμές Qr με τον αλγόριθμο Leading Eigenvector για τα δίκτυα μετοχών

Στους Πίνακες 6 έως 8 και στις Εικόνες 22 έως 24 αντίστοιχα, παρουσιάζονται οι τιμές του Qr για τα δίκτυα μετοχών ανά έτος. Είναι προφανές ότι ο αλγόριθμος Leading Eigenvector αξιολογείται να έχει την καλύτερη προσέγγιση στο διαχωρισμό κοινοτήτων. Ακολουθεί ο αλγόριθμος fast-greedy, με διαφορά στα δίκτυα των ετών 2010 και έπειτα.



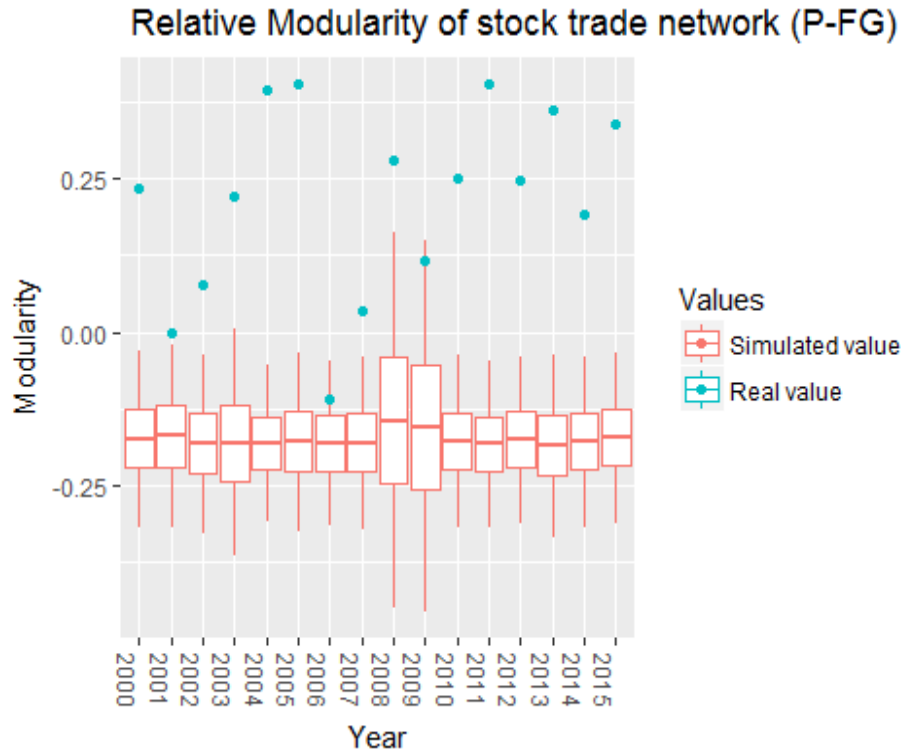
Εικόνα 23 Τιμές Q_r με τον αλγόριθμο LP&BRIM για τα δίκτυα μετοχών

Πίνακας 6 Τιμές Q_r με τον αλγόριθμο LP&BRIM για τα δίκτυα μετοχών

Label propag brim

year	Real	mean	sd	Percent	z_score
2000	0.0518192	-0.4801551	0.0484170	100.0	0.37594615
2001	-0.2710280	-0.4709759	0.0487988	100.0	-1.34690037
2002	-0.0484608	-0.4704419	0.0524950	100.0	-0.15918957
2003	-0.1744590	-0.4632578	0.0556914	100.0	-0.83156793
2004	-0.0018856	-0.4832269	0.0497727	100.0	0.08935514
2005	0.2013267	-0.4849035	0.0513168	100.0	1.17378020
2006	-0.2321596	-0.4812708	0.0496155	100.0	-1.13948203
2007	0.0207156	-0.4833365	0.0526480	100.0	0.20996471
2008	-0.2218396	-0.4094958	0.0601212	99.9	-1.08441074
2009	0.1246646	-0.3994960	0.0653278	100.0	0.76467922
2010	-0.0563970	-0.4868018	0.0525424	100.0	-0.20154042
2011	0.3995733	-0.4761737	0.0533707	100.0	2.23170619
2012	-0.0952865	-0.4861702	0.0522487	100.0	-0.40907092
2013	-0.1331115	-0.4807499	0.0573230	100.0	-0.61092049
2014	-0.1103917	-0.4915267	0.0519195	100.0	-0.48967839
2015	0.2488397	-0.4824513	0.0498123	100.0	1.42732925

Για τον αλγόριθμο LP&BRIM, η αξιολόγηση μέσω του μέτρου Q_r δε δίνει ικανοποιητικές ενδείξεις για το διαχωρισμό κοινοτήτων στα διαχρονικά δίκτυα. Πιο συγκεκριμένα, από την ανάλυση των πραγματικών δεδομένων βλέπουμε ότι ο αλγόριθμος προσδίδει σχετικά μεγάλη modular ιδιότητα στα δίκτυα των ετών πριν από το 2008, ενώ το Q_r βρίσκει αυτά τα δίκτυα να είναι modular προ κρίσης, anti-modular κατά την κρίση και ξανά modular μετά την πάροδό της.

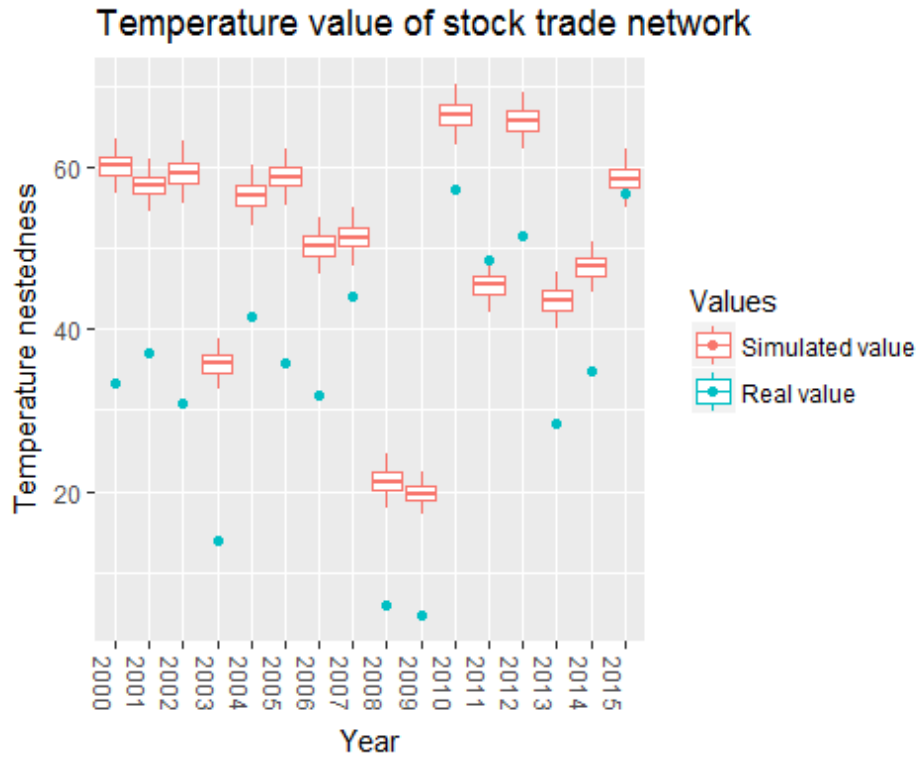


Εικόνα 24 Τιμές Qr με τον αλγόριθμο fast-greedy για τα δίκτυα μετοχών

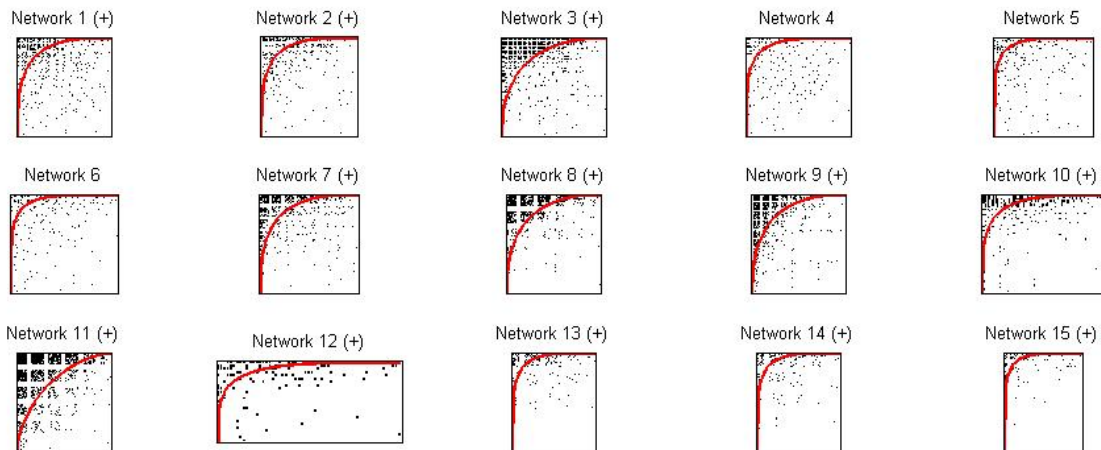
Πίνακας 7 Τιμές Qr με τον αλγόριθμο fast-greedy για τα δίκτυα μετοχών

Fast greedy

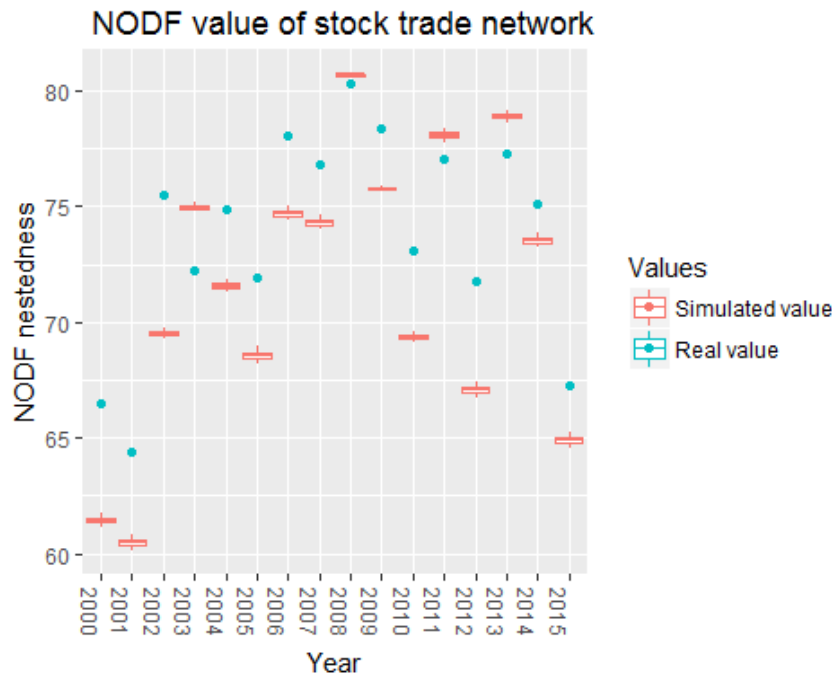
year	Real	mean	sd	Percent	z_score
2000	0.2341051	-0.1731356	0.0478929	100.0	0.12473704
2001	-0.0007788	-0.1694688	0.0490642	99.9	-1.39366991
2002	0.0765011	-0.1813374	0.0483297	100.0	-0.89409416
2003	0.2220144	-0.1800134	0.0614655	100.0	0.04657684
2004	0.3928347	-0.1815462	0.0427395	100.0	1.15084433
2005	0.4043118	-0.1783516	0.0486338	100.0	1.22503790
2006	-0.1087566	-0.1808669	0.0449386	93.9	-2.09169235
2007	0.0345261	-0.1804444	0.0469304	99.9	-1.16544162
2008	0.2798819	-0.1439626	0.1020239	100.0	0.42066138
2009	0.1173457	-0.1538897	0.1005835	100.0	-0.63005410
2010	0.2484692	-0.1777525	0.0466467	100.0	0.21759385
2011	0.4038403	-0.1824304	0.0448964	100.0	1.22199000
2012	0.2458460	-0.1752384	0.0449819	100.0	0.20063634
2013	0.3594010	-0.1852634	0.0494400	100.0	0.93471222
2014	0.1900291	-0.1790282	0.0458368	100.0	-0.16019205
2015	0.3373795	-0.1714995	0.0462837	100.0	0.79235431



Εικόνα 25 Τιμές Nestedness με τον αλγόριθμο NTC για τα δίκτυα μετοχών



Εικόνα 26 NTC nested sorting



Εικόνα 27 Τιμές Nestedness με τον αλγόριθμο NODF για τα δίκτυα μετοχών

Πίνακας 8 Τιμές Nestedness NTC & NODF

NODF							NTC						
year	Real	Mean	SD	percent	z_score	p_val	year	Real	Mean	SD	percent	z_score	p_val
2000	66.47393	61.44372	0.10593	100	-1.59668	0.00200	2000	33.27033	60.15311	1.11227	0.0	-0.09503	0.00200
2001	64.42022	60.49254	0.12105	100	-2.04552	0.00200	2001	37.12408	57.70420	1.07481	0.0	0.14605	0.00200
2002	75.46332	69.50362	0.07388	100	0.36793	0.00200	2002	30.80525	59.31988	1.27675	0.0	-0.24924	0.00200
2003	72.22462	74.96965	0.06118	100	-0.33988	0.00200	2003	14.03226	35.73252	1.03134	0.0	-1.29851	0.00200
2004	74.85097	71.57650	0.09678	100	0.23411	0.00200	2004	41.64579	56.49648	1.27680	0.0	0.42891	0.00200
2005	71.95138	68.57223	0.13304	100	-0.39959	0.00200	2005	35.71989	58.77089	1.14324	0.0	0.05821	0.00200
2006	78.08338	74.71476	0.10760	100	0.94054	0.00200	2006	31.92358	50.27975	1.14809	0.0	-0.17928	0.00200
2007	76.84011	74.30281	0.10504	100	0.66883	0.00200	2007	44.03433	51.37949	1.17398	0.0	0.57833	0.00200
2008	80.32654	80.66588	0.04133	100	1.43078	0.58084	2008	5.90593	21.17466	1.12717	0.0	-1.80687	0.00200
2009	78.34240	75.76500	0.03766	100	0.99715	0.00200	2009	4.68789	19.78151	0.89603	0.0	-1.88307	0.00200
2010	73.07464	69.36839	0.08488	100	-0.15411	0.00200	2010	57.15947	66.43952	1.27219	0.0	1.39940	0.00200
2011	77.05065	78.09265	0.10496	0	0.71484	0.06986	2011	48.57910	45.44008	1.13570	99.8	0.86264	0.00599
2012	71.77267	67.08340	0.11944	100	-0.43865	0.00200	2012	51.58906	65.67970	1.17193	0.0	1.05094	0.00200
2013	77.24831	78.89037	0.08659	0	0.75804	0.01796	2013	28.47202	43.51635	1.16623	0.0	-0.39520	0.00200
2014	75.11092	73.55583	0.11243	100	0.29092	0.00200	2014	34.89609	47.70550	1.03791	0.0	0.00667	0.00200
2015	67.24244	64.90720	0.10761	100	-1.42872	0.00200	2015	56.78616	58.59758	1.20874	5.2	1.37605	0.10579

Σε αντίθεση με το modularity, οι τιμές του nestedness διαφέρουν σημαντικά μεταξύ των αλγορίθμων. Ο NTC δίνει πάρα πολύ υψηλές τιμές (πάνω από 0.8) για όλα τα δίκτυα, σε αντίθεση με τον NODF, του οποίου οι τιμές είναι σαφώς μικρότερες. Σε κάθε περίπτωση, όπως και στο modularity, δε μας ενδιαφέρουν οι τιμές από μόνες τους αλλά το πού βρίσκονται κάθε φορά σε σχέση με τα τυχαία δίκτυα. Ένα πρώτο συμπέρασμα, επομένως, είναι ότι το δίκτυο είναι nested διαχρονικά. Μία πιθανή εξήγηση γι' αυτό είναι ο συσχετισμός που έχει βρεθεί στη βιβλιογραφία μεταξύ nestedness και diassortativity (Johnson et al, 2013), καθώς τα χρηματοοικονομικά δίκτυα είναι γενικά diassortative [71].

Αν και όλα τα δίκτυα είναι nested, υπάρχουν χρονιές στις οποίες η παρατηρούμενη τιμή απέχει πάρα πολύ σε σχέση με το κατώφλι του 97,5% των τυχαίων δικτύων. Τέτοια είναι αρχικά τα δίκτυα 3, 7, 8, 9 σε αντίθεση με τα δίκτυα 1,2,4,5,6. Μέχρι και το 9^ο δίκτυο τα αποτελέσματα έχουν παρόμοια λογική με το modularity. Από την dot com, δομή ιδιαίτερα επιρρεπής σε διάδοση εμφανίζεται μόνο μετά το τέλος της κρίσης. Αντίθετα, το δίκτυο γίνεται ιδιαίτερα nested ήδη από το 2006-2007. Βλέπουμε δηλαδή και πάλι ότι το διμερές δίκτυο που έχει μόνο χρηματοοικονομικές μετοχές ως ένα από τα δύο ανεξάρτητα sets έχει τα επιθυμητά αποτελέσματα εγκαίρως μόνο για την χρηματοοικονομική κρίση και όχι για την τεχνολογική.

Διαφορετική εικόνα δίνει το nestedness σε σχέση με το modularity μετά την κρίση. Ενώ το modularity επιστρέφει ουσιαστικά σε επίπεδα προ κρίσης από το δίκτυο 10 και μετά (με εξαίρεση το 11 για το οποίο θα δοθεί ερμηνεία αμέσως παρακάτω), δε συμβαίνει κάτι τέτοιο με το nestedness. Τα δίκτυα παραμένουν ιδιαίτερα nested και μετά την κρίση, με τις παρατηρούμενες τιμές να απέχουν πολύ από τις τυχαίες. Η παρατήρηση αυτή έχει σαφή οικονομική ερμηνεία. Από την κρίση και μετά ο χρηματοοικονομικός κλάδος έχει έρθει στο προσκήνιο. Η σημασία του είναι βαρύνουσα παγκοσμίως, ενώ πολλοί είναι αυτοί που ψάχνουν την επόμενη Lehman Brothers. Τα αποτελέσματα της ανάλυσης δεν αποτελούν κάποιο είδος προοικονομίας για κάποια κρίση, δείχνουν όμως ότι αν αυτή χτυπήσει κάποιες τράπεζες, η συνδεσιμότητα του δικτύου είναι τέτοια που θα ευνοήσει τη μαζική εξάπλωση του προβλήματος και όχι τον περιορισμό του.

Τέλος, είναι σκόπιμο να γίνει ξεχωριστή αναφορά στο δίκτυο 11. Το δίκτυο αυτό είναι το πυκνότερο όλων (25,80%), έχει χαμηλό modularity και ιδιαίτερα υψηλό nestedness. Ουσιαστικά έχει εικόνα παρόμοια με τα δίκτυα 7,8,9 (και 3), παρόλο που η χρηματοοικονομική κρίση για την Αμερική θεωρητικά έχει παρέλθει. Στην πραγματικότητα το διάστημα αυτό κρίνεται αρκετά ταραχώδες. Υπήρχαν εντονότατοι φόβοι για διάδοση της ευρωπαϊκής πλέον κρίσης χρέους σε Ιταλία και Ισπανία, γεγονός που θα έπληττε στη συνέχεια και την παγκόσμια οικονομία και επομένως και την Αμερική. Είναι χαρακτηριστικό ότι το 2011 η πιστοληπτική ικανότητα της Αμερικής υποβαθμίστηκε για πρώτη φορά από το 1941. Το δίκτυο 11 είναι επίσης το πιο ενδεικτικό για να επιβεβαιώσει τον Fortuna (2009), ότι δηλαδή σε ιδιαίτερα πυκνά δίκτυα οι τιμές του nestedness και του modularity έχουν αρνητική συσχέτιση.

3.3 Δίκτυα Παγκόσμιου Εμπορίου

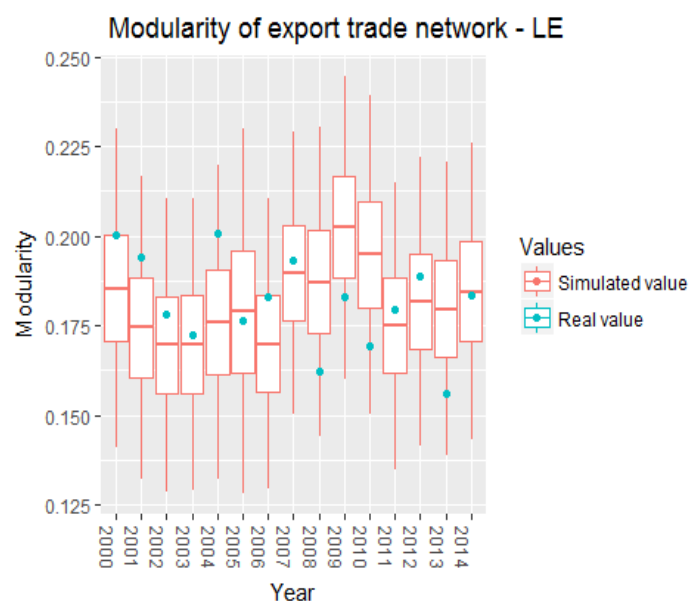
Για την ανάλυση των δικτύων του εμπορίου είναι επίσης απαραίτητη, όπως και στα δίκτυα μετοχών, η μελέτη της πυκνότητας κάθε δικτύου. Στον Πίνακα 10 περιέχονται οι πυκνότητες των δικτύων ανά έτος για τα δίκτυα των εξαγωγών (exports) και εισαγωγών (imports). Υψηλότερα σε πυκνότητα είναι τα δίκτυα 01-02, 02-03 για τις εξαγωγές, ενώ για τα δίκτυα των εισαγωγών υψηλότερα σε πυκνότητα είναι τα δίκτυα των ετών 13 και 14.

Πίνακας 9 Οι πυκνότητες των δικτύων για εξαγωγές (exports) και εισαγωγές (imports)

Export Network		Import Network	
year	density	year	density
2000	0.10112	2000	0.04730
2001	0.12035	2001	0.04906
2002	0.12021	2002	0.04899
2003	0.11314	2003	0.05127
2004	0.10293	2004	0.05274
2005	0.10057	2005	0.05179
2006	0.10889	2006	0.05307
2007	0.09535	2007	0.05051
2008	0.09666	2008	0.05046
2009	0.08739	2009	0.04913
2010	0.08671	2010	0.05112
2011	0.10137	2011	0.05295
2012	0.09597	2012	0.05057
2013	0.10780	2013	0.05308
2014	0.10970	2014	0.05683

3.3.1 Δίκτυα Εξαγωγών (Exports)

Από την εικόνα 28 παρατηρεί κανείς ότι τα δίκτυα εξαγωγών χαρακτηρίζονται από τον αλγόριθμο Leading Eigenvector αρκετά modular για τις περιόδους 2000 έως και το 2007. Έπειτα, για τα έτη 08-09, 09-10 και 10-11, δομούνται ως anti-modular και επανέρχονται σε modular δομή έως το 2014 με εξαίρεση το 2013, στο οποίο παρατηρείται η μεγαλύτερη anti-modular τιμή.

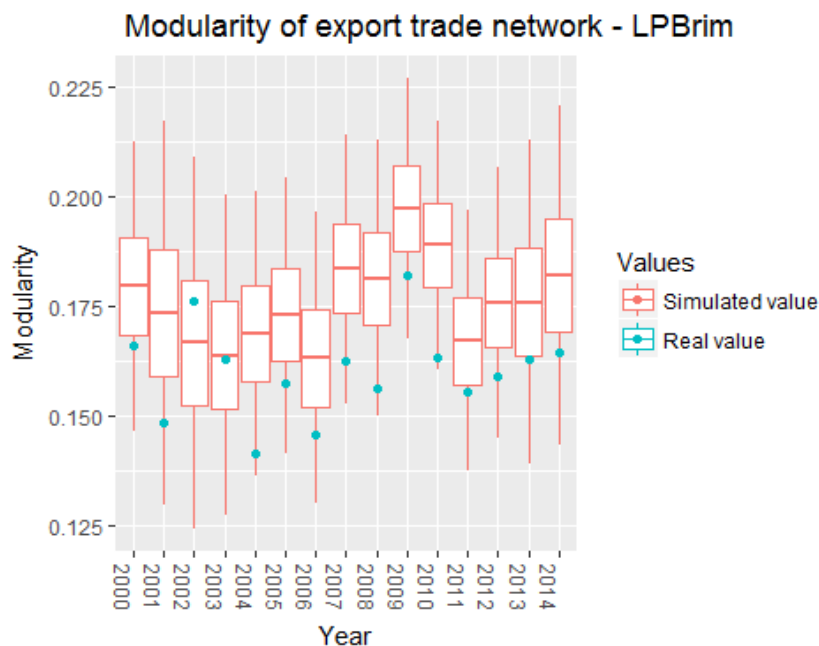


Εικόνα 28 Τιμές Modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα εξαγωγών

Πίνακας 10 Τιμές modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα εξαγωγών

Leading Eigenvector - Modularity

year	Real	mean	sd	Percent	z_score
2000	0.2001481	0.1852641	0.0148357	85.8	1.4402085
2001	0.1942005	0.1743990	0.0140858	92.0	0.9839764
2002	0.1783749	0.1695369	0.0136576	72.7	-0.2299941
2003	0.1723427	0.1697336	0.0135314	57.2	-0.6927206
2004	0.2006976	0.1759963	0.0146431	96.8	1.4823641
2005	0.1765019	0.1789412	0.0169620	40.0	-0.3736702
2006	0.1828489	0.1699331	0.0134872	83.8	0.1132017
2007	0.1932333	0.1897483	0.0131426	59.0	0.9097760
2008	0.1622754	0.1870954	0.0143821	4.0	-1.4649785
2009	0.1830939	0.2024489	0.0140826	8.0	0.1319924
2010	0.1691752	0.1948860	0.0148343	3.3	-0.9356961
2011	0.1795694	0.1750226	0.0133690	62.1	-0.1383684
2012	0.1886998	0.1816411	0.0134278	70.3	0.5620207
2013	0.1558160	0.1796803	0.0136614	4.8	-1.9604740
2014	0.1836201	0.1845752	0.0137787	46.9	0.1723620



Εικόνα 29 Τιμές Modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα εξαγωγών

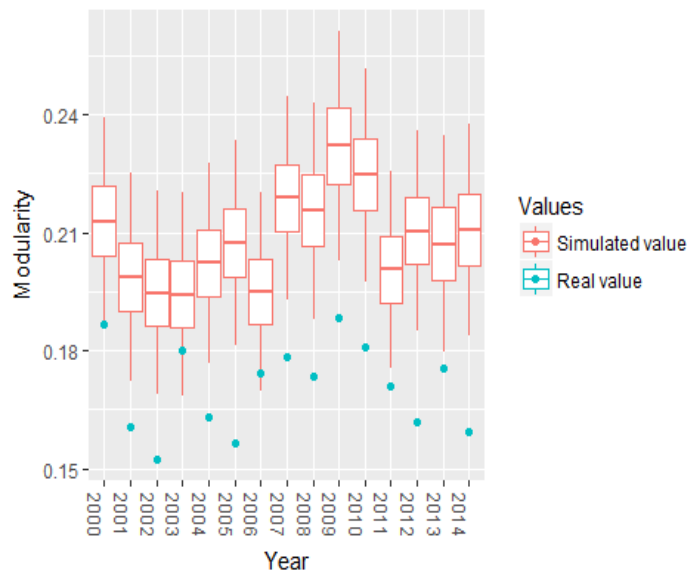
Ο αλγόριθμος LP&BRIM, όπως φαίνεται και από την Εικόνα 29, βρίσκει τα δίκτυα εξαγωγών κάθε έτους anti-modular, εκτός από τα έτη 2002 και 2003. Όπως είναι γνωστό από τη βιβλιογραφία [86]-[88], την περίοδο αυτή έλαβε χώρα στο παγκόσμιο εμπόριο μία αναδιάταξη κοινοτήτων, λόγω της μεγαλύτερης ροής εμπορευμάτων προς την ανατολική Ασία και τα Ηνωμένα Αραβικά Εμιράτα [91]-[94].

Πίνακας 11 Τιμές modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα εξαγωγών

Lpbrim - Modularity

year	Real	mean	sd	Percent	z_score
2000	0.1660132	0.1794809	0.0109973	8.3	0.5470037
2001	0.1485903	0.1734791	0.0145532	0.8	-1.1118486
2002	0.1761919	0.1666165	0.0141391	76.4	1.5161320
2003	0.1627155	0.1638241	0.0121633	54.3	0.2330268
2004	0.1414556	0.1687622	0.0108192	0.0	-1.7911534
2005	0.1573790	0.1728633	0.0105133	3.0	-0.2750665
2006	0.1458959	0.1631379	0.0110709	2.5	-1.3683822
2007	0.1626722	0.1834791	0.0102196	0.1	0.2289046
2008	0.1560482	0.1812711	0.0104612	0.2	-0.4017712
2009	0.1818653	0.1972097	0.0099130	3.7	2.0563077
2010	0.1633935	0.1888492	0.0094720	0.0	0.2975795
2011	0.1556213	0.1670160	0.0098971	8.9	-0.4424136
2012	0.1589716	0.1757777	0.0102871	1.5	-0.1234344
2013	0.1627953	0.1758665	0.0122567	12.1	0.2406302
2014	0.1644113	0.1819627	0.0128752	5.7	0.3944855

Modularity of export trade network - FG



Εικόνα 30 Τιμές Modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εξαγωγών

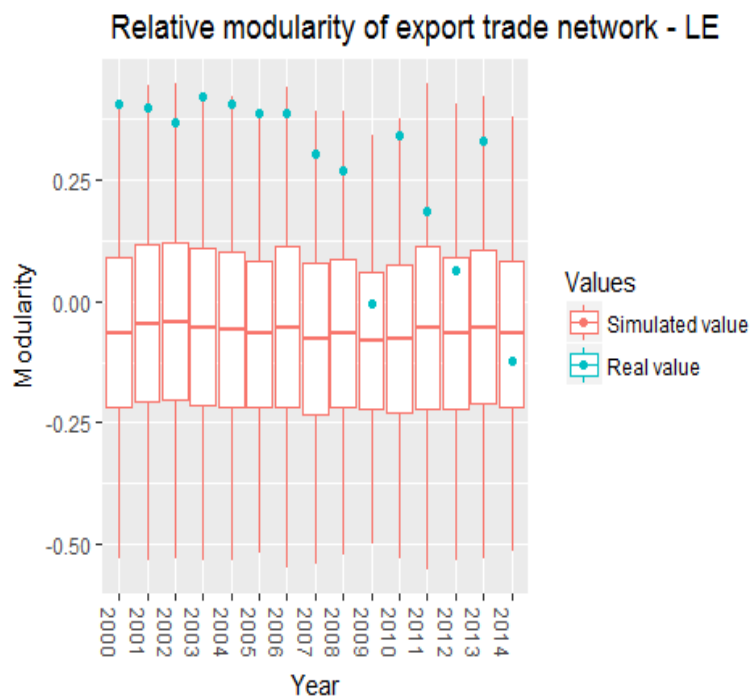
Τέλος, με τον αλγόριθμο fast-greedy, όλα τα έτη του παγκόσμιου εμπορίου βρίσκονται να έχουν μόνο anti-modular δομή και μάλιστα σε ακραίες τιμές, εν συγκρίσει με το null model. Τα λιγότερο anti-modular δίκτυα είναι αυτά των ετών 2001 και 2003, τα οποία υπόκεινται στις αλλαγές που προαναφέρθηκαν αναφορικά με τη δομή και τη ροή του παγκόσμιου εμπορίου [91]-[94].

Πίνακας 12 Τιμές modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εξαγωγών

Fast Greedy - Modularity

year	Real	mean	sd	Percent	z_score
2000	0.1867849	0.2127584	0.0088570	0.1	1.4209158471
2001	0.1605751	0.1987166	0.0087934	0.0	-0.9059947433
2002	0.1522702	0.1946547	0.0086073	0.0	-1.6433074013
2003	0.1800877	0.1942780	0.0086451	5.0	0.8263350979
2004	0.1629197	0.2022215	0.0084477	0.0	-0.6978425554
2005	0.1562624	0.2073516	0.0087054	0.0	-1.2888730345
2006	0.1743138	0.1948837	0.0084037	0.2	0.3137258224
2007	0.1782698	0.2186984	0.0086541	0.0	0.6649453844
2008	0.1732213	0.2155464	0.0091631	0.0	0.2167380173
2009	0.1885061	0.2320210	0.0097203	0.0	1.5737226136
2010	0.1808711	0.2245585	0.0090681	0.0	0.8958814769
2011	0.1707849	0.2005143	0.0083915	0.0	0.0004318251
2012	0.1619887	0.2104100	0.0084278	0.0	-0.7804911374
2013	0.1752871	0.2070786	0.0091511	0.0	0.4001400697
2014	0.1595576	0.2105971	0.0089574	0.0	-0.9963272824

3.3.2 Τιμές Relative Modularity (Q_r) για τα δίκτυα εξαγωγών (exports)



Εικόνα 31 Τιμές Relative Modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα εξαγωγών

Όσον αφορά τις τιμές του μέτρου Q_r για το Relative Modularity, ο αλγόριθμος Leading Eigenvector αξιολογεί τα δίκτυα κάθε χρονιάς ως extremely modular. Σε αυτό έρχεται να διαφωνήσει ο αμέσως επόμενος αλγόριθμος (LP&BRIM), ο οποίος βρίσκει τα



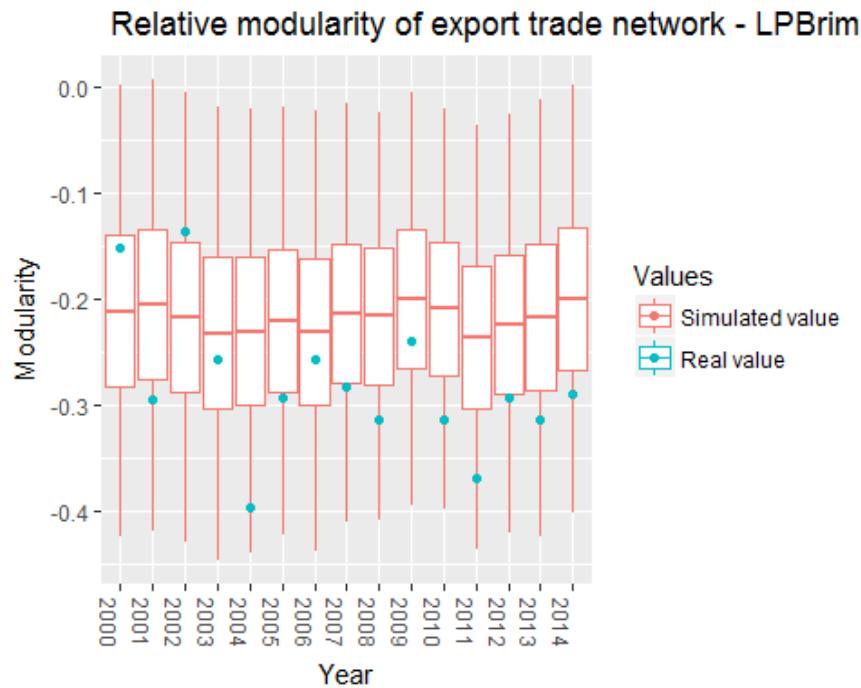
τυχαία δίκτυα που χρησιμοποιούνται για την αξιολόγηση μέσω του Qr να είναι modular για τα έτη 2000, 2002, 2003, 2006 και 2009, ενώ όλα τα υπόλοιπα antimodular.

Το αποτέλεσμα αυτό, συνάδει με την αξιολόγηση των communities από την ανάλυση των πραγματικών δεδομένων και πιθανώς να εξηγείται από το γεγονός ότι το 2009, σύμφωνα με τη βιβλιογραφία, δομήθηκε ένα ισχυρό cluster εξαγωγών από τις χώρες της Αμερικής, της Κίνας, της Ιαπωνίας και της Γερμανίας [87], [88].

Πίνακας 13 Τμές relative modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα εξαγωγών

Leading Eigenvector

year	Real	mean	sd	Percent	z_score
2000	0.4072022	-0.0647978	0.1547981	99.9	0.77028454
2001	0.3994169	-0.0447988	0.1633111	99.9	0.72449242
2002	0.3684211	-0.0409751	0.1634104	99.9	0.54217897
2003	0.4233766	-0.0522545	0.1609812	100.0	0.86542022
2004	0.4060914	-0.0572893	0.1592231	100.0	0.76375070
2005	0.3870130	-0.0667948	0.1504330	100.0	0.65153422
2006	0.3869347	-0.0532462	0.1647361	99.8	0.65107359
2007	0.3041096	-0.0765644	0.1555187	97.1	0.16390762
2008	0.2702703	-0.0644000	0.1521953	97.3	-0.03513069
2009	-0.0057143	-0.0798286	0.1399847	74.6	-1.65843460
2010	0.3403694	-0.0771135	0.1507270	98.0	0.37718290
2011	0.1871921	-0.0530000	0.1667430	92.6	-0.52378516
2012	0.0636132	-0.0648193	0.1564761	74.5	-1.25065950
2013	0.3296089	-0.0529553	0.1588304	96.9	0.31389135
2014	-0.1242604	-0.0666509	0.1492719	38.9	-2.35570658

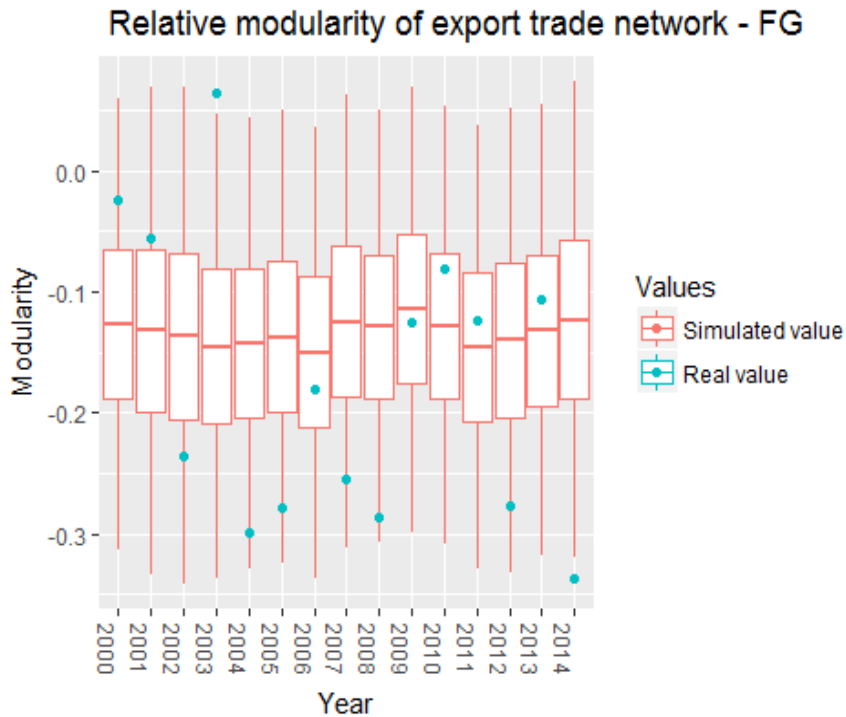


Εικόνα 32 Τιμές Relative Modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα εξαγωγών

Πίνακας 14 Τιμές relative modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα εξαγωγών

Label propag brim

year	Real	mean	sd	Percent	z_score
2000	-0.1523546	-0.2114183	0.0710206	79.4	1.86974800
2001	-0.2944606	-0.2055743	0.0707589	7.8	-0.21184360
2002	-0.1357341	-0.2174017	0.0704725	85.7	2.11320762
2003	-0.2571429	-0.2326805	0.0712383	35.9	0.33479305
2004	-0.3959391	-0.2304924	0.0697807	0.0	-1.69831548
2005	-0.2935065	-0.2208364	0.0672346	13.3	-0.19786709
2006	-0.2562814	-0.2306231	0.0690734	36.9	0.34741170
2007	-0.2821918	-0.2136384	0.0656216	15.8	-0.03212744
2008	-0.3135135	-0.2163892	0.0638081	3.4	-0.49093300
2009	-0.2400000	-0.2002629	0.0647073	28.5	0.58590426
2010	-0.3139842	-0.2095409	0.0626694	2.2	-0.49782724
2011	-0.3694581	-0.2360148	0.0666680	0.9	-1.31041833
2012	-0.2926209	-0.2236845	0.0656477	14.8	-0.18489427
2013	-0.3128492	-0.2180615	0.0687467	6.5	-0.48120148
2014	-0.2899408	-0.2004734	0.0672208	7.5	-0.14563668



Εικόνα 33 Τιμές Relative Modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εξαγωγών

Τέλος, με τον αλγόριθμο fast-greedy, τα έτη 2002, 2004-2005, 2007-2008, 2012 και 2014 του παγκόσμιου εμπορίου φαίνεται να δομούν anti-modular δίκτυα σύμφωνα με το μέτρο Q_r . Τα λιγότερο anti-modular δίκτυα είναι αυτά των ετών 2002 και 2007, ενώ τα τυχαία δίκτυα όλων των υπόλοιπων ετών αξιολογούνται ως modular.

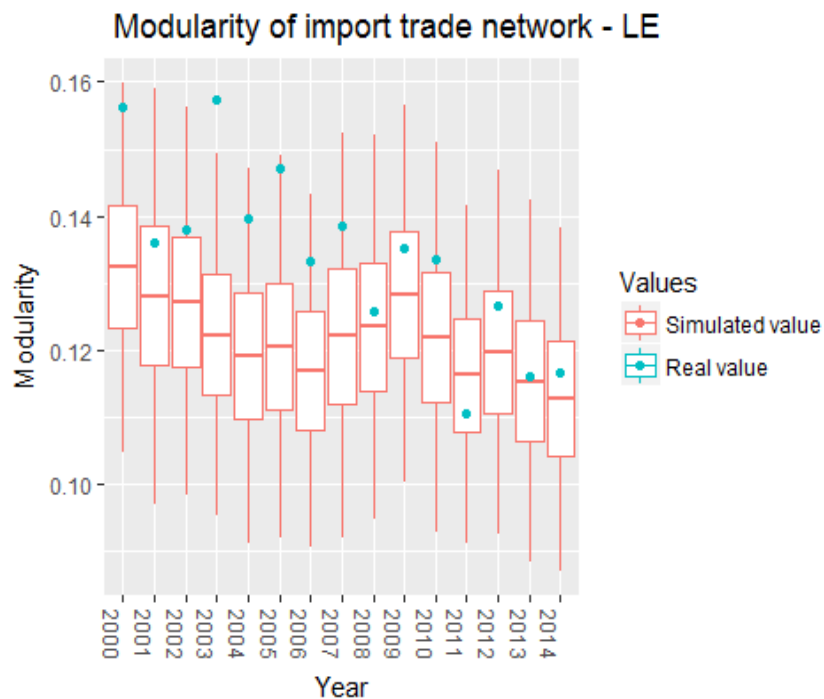
Πίνακας 15 Τιμές relative modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εξαγωγών

Fast greedy

year	Real	mean	sd	Percent	z_score
2000	-0.0249307	-0.1267645	0.0621412	92.9	1.24801263
2001	-0.0553936	-0.1322099	0.0671021	85.9	0.99204776
2002	-0.2354571	-0.1366039	0.0683695	5.6	-0.52094074
2003	0.0649351	-0.1454026	0.0639082	99.4	2.00311269
2004	-0.2994924	-0.1426041	0.0619791	0.1	-1.05899936
2005	-0.2779221	-0.1372779	0.0622827	0.5	-0.87775424
2006	-0.1809045	-0.1497940	0.0620914	31.8	-0.06256146
2007	-0.2547945	-0.1247562	0.0624465	0.9	-0.68342428
2008	-0.2864865	-0.1287892	0.0593848	0.2	-0.94971692
2009	-0.1257143	-0.1142571	0.0612885	41.7	0.40117608
2010	-0.0817942	-0.1276992	0.0604434	79.9	0.77021589
2011	-0.1231527	-0.1458079	0.0608770	67.7	0.42269980
2012	-0.2773537	-0.1396590	0.0639072	0.7	-0.87297834
2013	-0.1061453	-0.1319162	0.0621361	66.5	0.56560545
2014	-0.3372781	-0.1230059	0.0656998	0.0	-1.37649496

3.3.3 Δίκτυα Εισαγωγών (Imports)

Όπως φαίνεται από την Εικόνα 34 και τον Πίνακα 16, η ιδιότητα του modularity είναι έντονη στο δίκτυο κάθε έτους για τις εισαγωγές, σύμφωνα με τον αλγόριθμο Leading Eigenvector. Αυτό έρχεται να επιβεβαιωθεί από τους δύο επόμενους αλγορίθμους, με εξαίρεση τα δίκτυα των ετών 2002, 2011 και 2014 που χαρακτηρίζονται από τον αλγόριθμο fast-greedy ως anti-modular (Εικόνα 36).

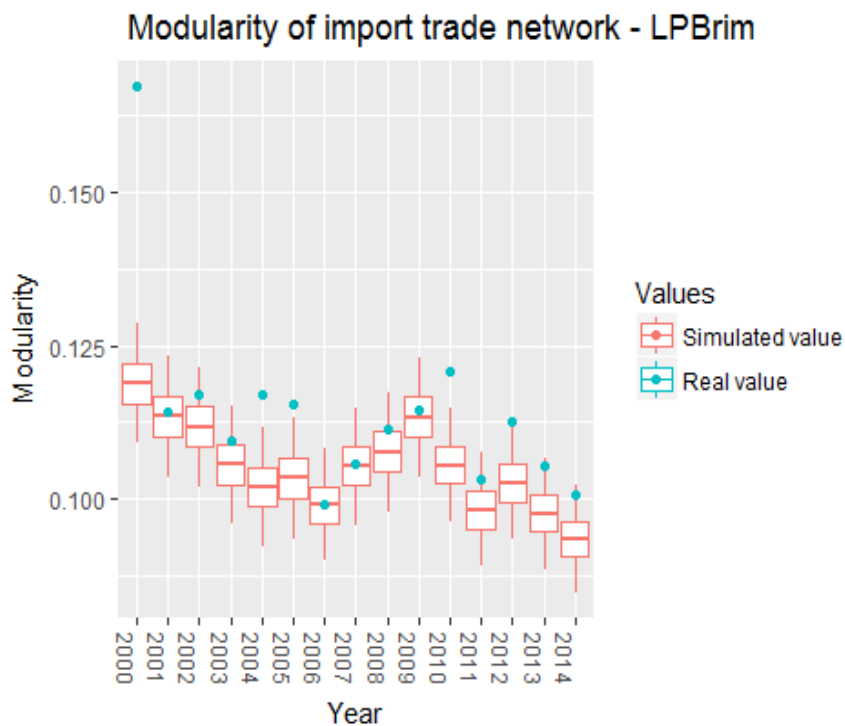


Εικόνα 34 Τιμές Modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εισαγωγών

Πίνακας 16 Τιμές modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα εισαγωγών

Leading Eigenvector – Modularity

year	Real	mean	sd	Percent	z_score
2000	0.1561547	0.1323835	0.0091712	99.5	1.63119499
2001	0.1359550	0.1280962	0.0103382	78.6	0.14396566
2002	0.1378949	0.1272383	0.0096223	86.8	0.28679481
2003	0.1574394	0.1222633	0.0089733	100.0	1.72578200
2004	0.1395268	0.1191839	0.0093330	99.1	0.40694508
2005	0.1470356	0.1205386	0.0094708	99.7	0.95979359
2006	0.1333173	0.1169031	0.0087654	97.6	-0.05023447
2007	0.1386385	0.1221166	0.0100746	96.7	0.34154349
2008	0.1257463	0.1234914	0.0095502	57.9	-0.60766025
2009	0.1351484	0.1283709	0.0093796	77.3	0.08458239
2010	0.1334035	0.1218449	0.0096682	91.2	-0.04389039
2011	0.1105086	0.1163310	0.0084173	23.3	-1.72955691
2012	0.1265414	0.1196413	0.0090549	78.5	-0.54912323
2013	0.1161681	0.1153995	0.0089939	52.2	-1.31287333
2014	0.1165159	0.1126804	0.0085609	66.9	-1.28726344



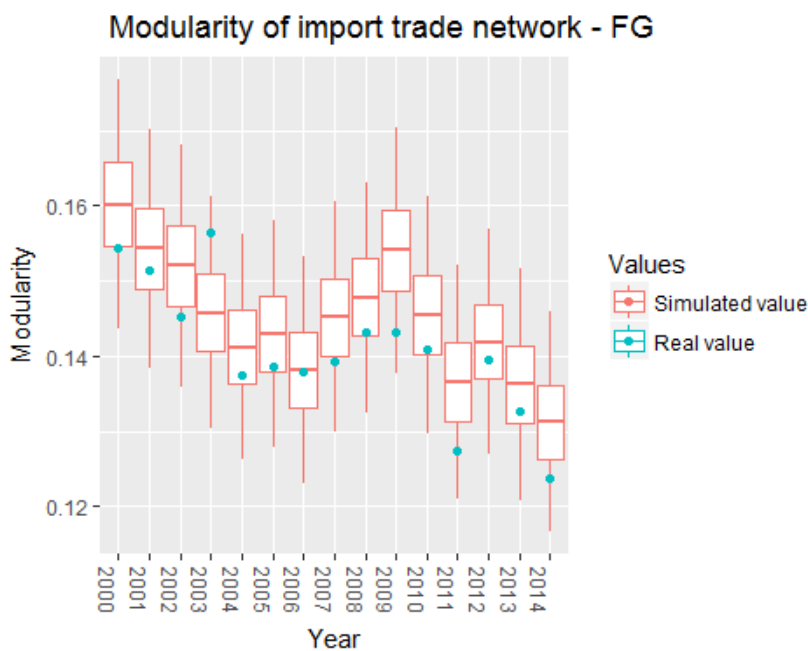
Εικόνα 35 Τιμές Modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα εισαγωγών

Πίνακας 17 Τιμές modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα εισαγωγών

Lpbrim - Modularity

year	Real	mean	sd	Percent	z_score
2000	0.1671858	0.1188427	0.0032307	100.0	3.3098121941
2001	0.1142676	0.1135542	0.0032904	61.4	-0.0008086425
2002	0.1170730	0.1118213	0.0032419	93.4	0.1746994053
2003	0.1094752	0.1056771	0.0032104	88.2	-0.3006247923
2004	0.1169066	0.1020037	0.0032682	100.0	0.1642924063
2005	0.1155725	0.1035243	0.0032828	99.8	0.0808272557
2006	0.0991926	0.0991407	0.0029997	52.8	-0.9439189988
2007	0.1059095	0.1053797	0.0031620	60.2	-0.5236979731
2008	0.1115165	0.1077198	0.0032460	86.9	-0.1729223722
2009	0.1143968	0.1133251	0.0032556	64.6	0.0072752043
2010	0.1208694	0.1056088	0.0031149	100.0	0.4122092592
2011	0.1032177	0.0983540	0.0031029	93.4	-0.6920995583
2012	0.1125778	0.1026225	0.0029964	100.0	-0.1065242193
2013	0.1053752	0.0976665	0.0030001	99.0	-0.5571272057
2014	0.1006716	0.0935580	0.0029322	99.0	-0.8513919629

Το γεγονός ότι και οι τρεις αλγόριθμοι δίνουν παρόμοια αποτελέσματα, υποδεικνύει ότι η διάδοση κρίσης κατά την περίοδο 2007-2008 δεν επηρεάζει τις εισαγωγές. Αντιθέτως, παρατηρείται μία αύξηση στις εισαγωγές και αύξηση του modularity, κυρίως κατά τα έτη 2003-2004 και 2009-2011. Η κοινοτικοποίηση αυτή χαρακτηρίζεται στη βιβλιογραφία ως trade integration, και χαρακτηρίζεται από μεγαλύτερο μερίδιο εισαγωγών στις μεγαλύτερες χώρες, χωρίς απαραίτητα αυτή να συνάδει με την οικονομική ανάπτυξη [88]-[91].



Εικόνα 36 Τιμές Modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εισαγωγών

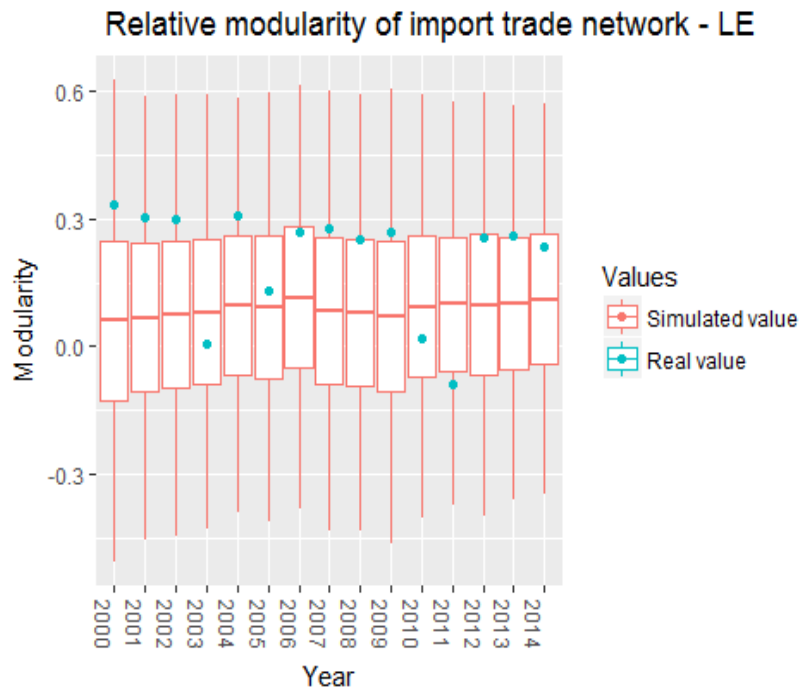
Πίνακας 18 Τιμές modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εισαγωγών

Fast Greedy - Modularity

year	Real	mean	sd	Percent	z_score
2000	0.1544711	0.1602057	0.0055577	15.8	1.513861897
2001	0.1514569	0.1542922	0.0053335	30.3	1.181038648
2002	0.1452841	0.1520307	0.0053810	9.4	0.499462318
2003	0.1565068	0.1458106	0.0051369	98.2	1.738632275
2004	0.1375439	0.1412090	0.0049723	23.7	-0.355187695
2005	0.1386770	0.1428856	0.0050422	21.2	-0.230077299
2006	0.1378423	0.1381836	0.0050468	47.3	-0.322241096
2007	0.1393669	0.1452194	0.0051349	13.1	-0.153895323
2008	0.1431458	0.1478066	0.0051122	18.0	0.263359433
2009	0.1431143	0.1540917	0.0054413	2.2	0.259873417
2010	0.1408329	0.1454286	0.0052618	18.3	0.007966582
2011	0.1273203	0.1365142	0.0052089	3.3	-1.484047094
2012	0.1396079	0.1419025	0.0050144	32.4	-0.127287684
2013	0.1325698	0.1362096	0.0051169	25.2	-0.904417478
2014	0.1236706	0.1312189	0.0048732	5.2	-1.887040901

3.3.4 Τιμές Relative modularity (Q_r) για τα δίκτυα εισαγωγών (imports)

Στις Εικόνες 37-39 και τους Πίνακες 19-21 παρουσιάζονται οι τιμές του μέτρου Q_r με τους τρεις αλγορίθμους. Παρατηρούμε ότι η αξιολόγηση κοινοτήτων είναι ικανοποιητική συγκρίνοντας τις τιμές αυτές με τις τιμές του modularity που παρουσιάστηκαν στην προηγούμενη παράγραφο.



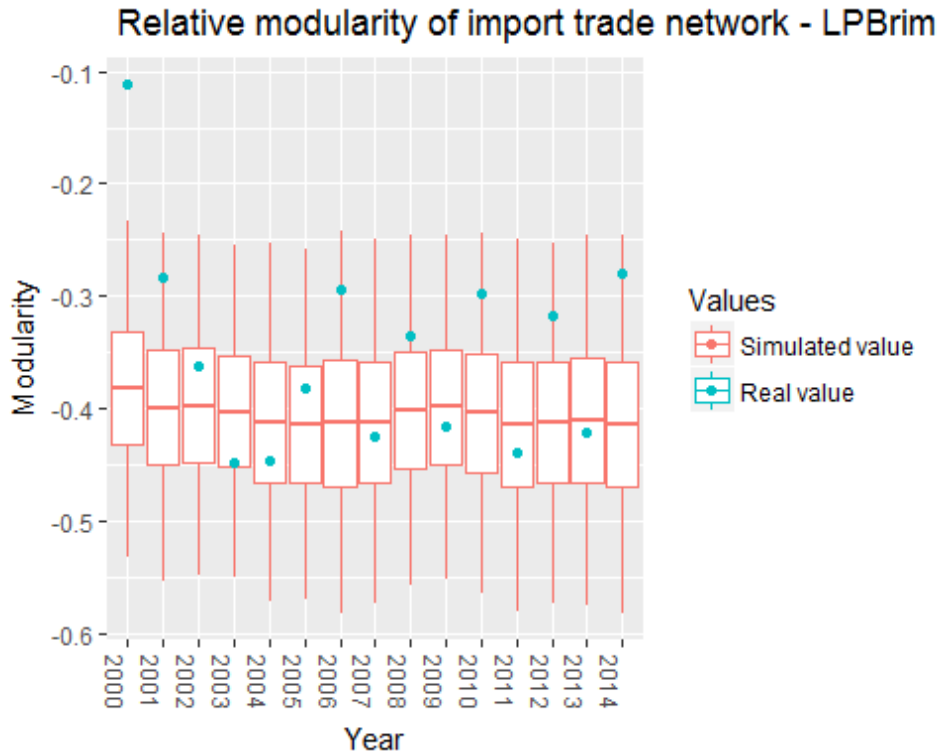
Εικόνα 37 Τιμές Relative Modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα εισαγωγών

Πίνακας 19 Τιμές relative modularity με τον αλγόριθμο Leading Eigenvector για τα δίκτυα εισαγωγών

Leading Eigenvector

year	Real	mean	sd	Percent	z_score
2000	0.3313343	0.0598201	0.1893344	99.8	0.9454080
2001	0.3017493	0.0667420	0.1742252	99.9	0.7172801
2002	0.2996441	0.0736612	0.1729878	99.9	0.7010476
2003	0.0069156	0.0804066	0.1700392	34.9	-1.5561564
2004	0.3084746	0.0964569	0.1620514	100.0	0.7691384
2005	0.1310251	0.0913265	0.1681478	50.6	-0.5991589
2006	0.2706965	0.1138200	0.1660317	99.3	0.4778346
2007	0.2776646	0.0838276	0.1722126	99.7	0.5315656
2008	0.2529617	0.0791540	0.1716999	82.6	0.3410832
2009	0.2707424	0.0700509	0.1784210	92.8	0.4781885
2010	0.0165062	0.0931032	0.1663950	36.1	-1.4822044
2011	-0.0906736	0.0995298	0.1579138	9.9	-2.3086582
2012	0.2572178	0.0974882	0.1664521	94.0	0.3739022
2013	0.2609819	0.1015788	0.1546758	97.9	0.4029266
2014	0.2356771	0.1108737	0.1527099	81.1	0.2078033

Και οι τρεις αλγόριθμοι επιστρέφουν τιμές του Q_r που χαρακτηρίζουν τα δίκτυα modular, χωρίς να εντοπίζεται anti-modular δίκτυο σε καμία χρονολογία.

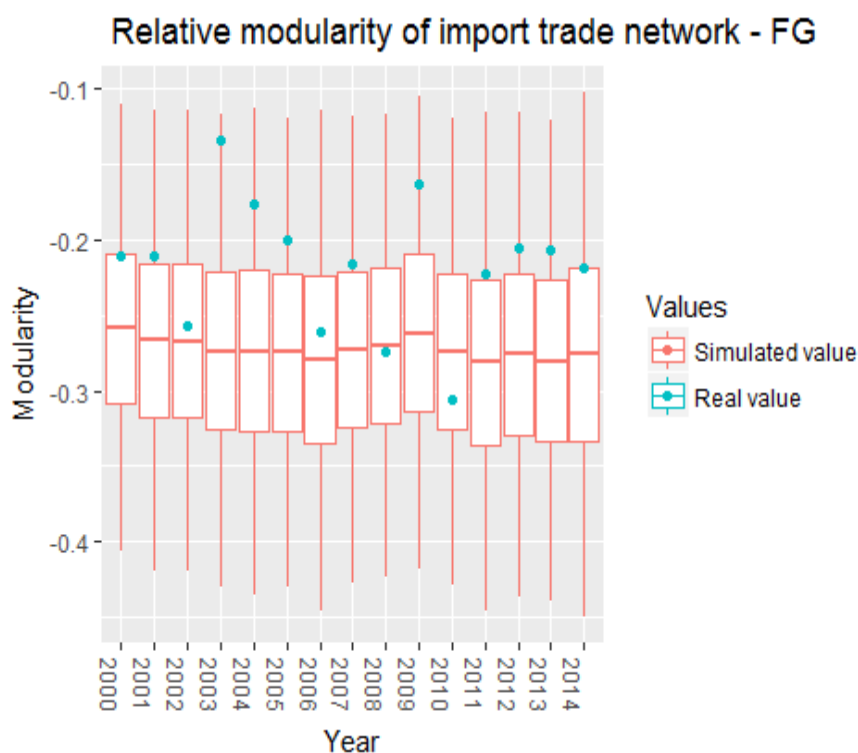


Εικόνα 38 Τιμές Relative Modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα εισαγωγών

Πίνακας 20 Τιμές relative modularity με τον αλγόριθμο LP&BRIM για τα δίκτυα εισαγωγών

Label propag brim

year	Real	mean	sd	Percent	z_score
2000	-0.1109445	-0.3819805	0.0498190	100.0	2.6246528
2001	-0.2827988	-0.3987784	0.0515058	98.0	0.7407264
2002	-0.3622776	-0.3970762	0.0505449	72.6	-0.1305472
2003	-0.4481328	-0.4030581	0.0492250	14.9	-1.0717216
2004	-0.4454237	-0.4122169	0.0531710	27.5	-1.0420240
2005	-0.3808554	-0.4140434	0.0518278	71.4	-0.3342036
2006	-0.2943495	-0.4124665	0.0566420	96.3	0.6141035
2007	-0.4243041	-0.4120462	0.0539794	46.6	-0.8105038
2008	-0.3351916	-0.4009798	0.0520165	90.2	0.1663783
2009	-0.4163028	-0.3981587	0.0509300	38.7	-0.7227900
2010	-0.2971114	-0.4037111	0.0532871	96.0	0.5838269
2011	-0.4391192	-0.4145039	0.0553648	37.4	-0.9729113
2012	-0.3175853	-0.4124646	0.0534096	95.1	0.3593851
2013	-0.4211886	-0.4101266	0.0549408	46.7	-0.7763505
2014	-0.2799479	-0.4137227	0.0561154	97.8	0.7719791



Εικόνα 39 Τιμές Relative Modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εισαγωγών

Πίνακας 21 Τιμές relative modularity με τον αλγόριθμο fast-greedy για τα δίκτυα εισαγωγών

Fast greedy

year	Real	mean	sd	Percent	z_score
2000	-0.2113943	-0.2588381	0.0492576	85.9	0.14500421
2001	-0.2113703	-0.2666968	0.0508136	87.3	0.14555579
2002	-0.2569395	-0.2672100	0.0507054	60.6	-0.89998144
2003	-0.1341632	-0.2737593	0.0521704	99.4	1.91698867
2004	-0.1769492	-0.2738088	0.0535863	95.5	0.93531117
2005	-0.2002716	-0.2747644	0.0517611	88.6	0.40020372
2006	-0.2614980	-0.2798581	0.0551928	69.7	-1.00457194
2007	-0.2165648	-0.2726232	0.0514289	84.2	0.02637194
2008	-0.2738676	-0.2702871	0.0509464	50.0	-1.28837836
2009	-0.1630277	-0.2617831	0.0519979	94.7	1.25472494
2010	-0.3053645	-0.2736066	0.0514547	22.7	-2.01104125
2011	-0.2227979	-0.2813860	0.0549665	80.1	-0.11663970
2012	-0.2060367	-0.2759777	0.0534687	86.0	0.26792761
2013	-0.2067183	-0.2803579	0.0530629	88.7	0.25228901
2014	-0.2187500	-0.2760990	0.0578957	78.0	-0.02376436

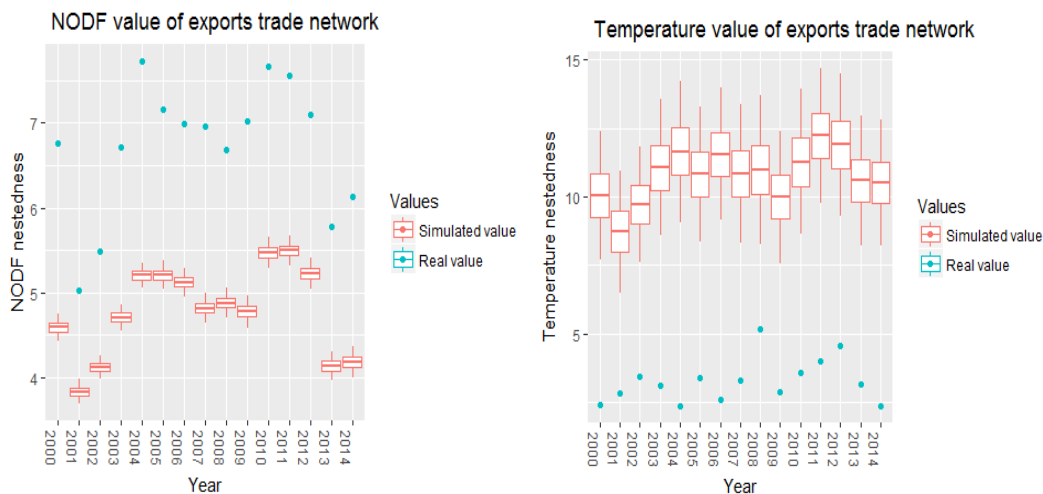
3.3.5 Τιμές Nestedness για τα δίκτυα εξαγωγών

Ενδιαφέρον παρουσιάζει η σύγκριση των τιμών της εμφωλευσιμότητας με τους αλγόριθμους NTC και NODF. Οι τιμές παρουσιάζονται συγκριτικά στον Πίνακα 22. Εύκολα παρατηρεί κανείς ότι με τον αλγόριθμο NODF, τα δίκτυα χαρακτηρίζονται ως απόλυτα nested, ενώ με τον αλγόριθμο Nestedness Temperature βρίσκονται πολύ χαμηλές τιμές εμφωλευσιμότητας.

Πίνακας 22 Τιμές nestedness με τους αλγόριθμους NODF & NTC για τα δίκτυα εξαγωγών

Exports – NODF							Exports - temperature								
year	Real	Mean	SD	percent	z_score	p_val	year	Real	Mean	SD	percent	z_score	p_val		
2000	2000	6.76423	4.59271	0.05519	100	0.05970	0.002	2000	2000	2.41696	10.04658	0.78411	0	-1.06085	0.002
2001	2001	5.02796	3.83315	0.04771	100	-2.1328	0.002	2001	2001	2.83534	8.72977	0.74009	0	-0.54241	0.002
2002	2002	5.48059	4.11844	0.04555	100	-1.5612	0.002	2002	2002	3.44182	9.73597	0.70391	0	0.20913	0.002
2003	2003	6.71898	4.70941	0.05043	100	0.00256	0.002	2003	2003	3.11699	11.07158	0.82626	0	-0.19339	0.002
2004	2004	7.73068	5.20478	0.04950	100	1.28014	0.002	2004	2004	2.35602	11.65602	0.85612	0	-1.13638	0.002
2005	2005	7.15386	5.20479	0.05583	100	0.55173	0.002	2005	2005	3.38937	10.82705	0.81987	0	0.14413	0.002
2006	2006	6.98529	5.12266	0.05625	100	0.33885	0.002	2006	2006	2.58526	11.55844	0.80534	0	-0.85230	0.002
2007	2007	6.95542	4.81775	0.05760	100	0.30113	0.002	2007	2007	3.31372	10.83058	0.84331	0	0.05038	0.002
2008	2008	6.68264	4.87881	0.05860	100	-0.0433	0.002	2008	2008	5.15353	10.98185	0.90730	0	2.33024	0.002
2009	2009	7.02169	4.77849	0.06354	100	0.38482	0.002	2009	2009	2.86691	9.98981	0.80054	0	-0.50329	0.002
2010	2010	7.66393	5.47130	0.06266	100	1.19584	0.002	2010	2010	3.57757	11.27075	0.88199	0	0.37734	0.002
2011	2011	7.55631	5.49576	0.05933	100	1.05994	0.002	2011	2011	3.99193	12.23585	0.82180	0	0.89081	0.002
2012	2012	7.09352	5.22948	0.06127	100	0.47553	0.002	2012	2012	4.55474	11.90664	0.86273	0	1.58824	0.002
2013	2013	5.78076	4.13690	0.05422	100	-1.1822	0.002	2013	2013	3.13875	10.59218	0.78690	0	-0.16643	0.002
2014	2014	6.13848	4.18611	0.06026	100	-0.7305	0.002	2014	2014	2.35695	10.51105	0.76369	0	-1.13523	0.002

Η απόκλιση αυτή πιθανώς να οφείλεται στο γεγονός ότι ο αλγόριθμος NTC βασίζεται σε εμπειρικό υπολογισμό της μήτρας εμφωλευσιμότητας, κάτι που μπορεί να εμποδίζει τον αλγόριθμο να έρθει σε σύγκλιση.

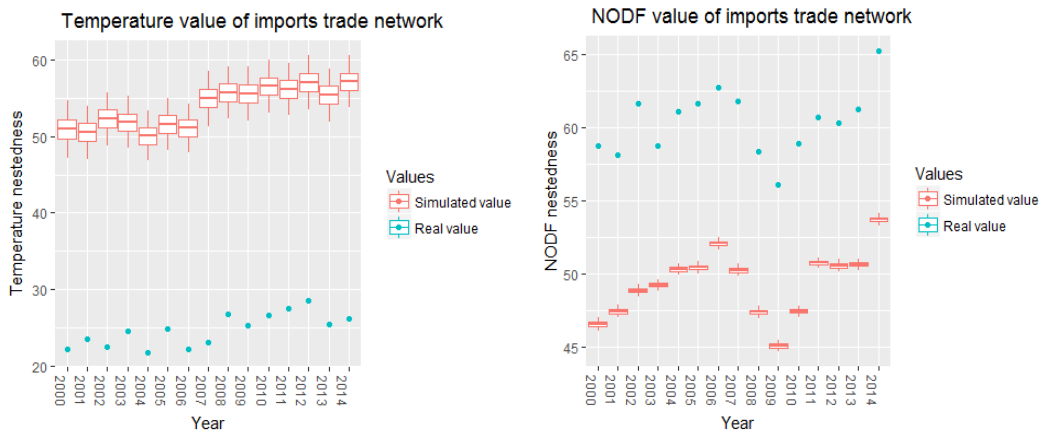


Εικόνα 40 Τιμές Nestedness με τους αλγόριθμους NODF&NTC για τα δίκτυα εξαγωγών

Παρά τη διαφορά στις τιμές από τον κάθε αλγόριθμο, αξίζει να παρατηρηθεί ότι στα δίκτυα των ετών 2005, 2006 και 2007 εμφανίζονται και με τους δύο αλγόριθμους οι μεγαλύτερες τιμές nestedness. Το ίδιο συμβαίνει και στο δίκτυο του έτους 2012-2013, κατά το οποίο παρατηρήθηκε μεγάλη αύξηση στις εξαγωγές της Κίνας και χαμηλότερη κοινοτικοποίηση σε σχέση με άλλα έτη [90], [93].

3.3.6 Τιμές Nestedness για τα δίκτυα εισαγωγών

Ακριβώς ίδια συμπεριφορά παρατηρείται από τους δύο αλγόριθμους κατά την ανάλυση των δικτύων εισαγωγών. Οι τιμές παρουσιάζονται στην Εικόνα 43 και τους Πίνακες 23&24. Ο αλγόριθμος NODF χαρακτηρίζει τα δίκτυα ως απόλυτα εμφωλεύσιμα σε όλα τα έτη, ενώ ο αλγόριθμος NTC επιστρέφει πολύ χαμηλές τιμές για την ιδιότητα.



Εικόνα 41 Τιμές Nestedness με τους αλγόριθμους NODF&NTC για τα δίκτυα εισαγωγών

Ωστόσο, και πάλι παρατηρούμε στα έτη 2005-2007 τις πιο ακραίες τιμές, ενώ φαίνεται τα δίκτυα των εισαγωγών να ξαναπαίρνουν nested δομή το έτος 2014.

Πίνακας 23 Τιμές nestedness με τον αλγόριθμο NTC για τα δίκτυα εισαγωγών

Imports NTC							
	year	Real	Mean	SD	percent	z_score	p_val
2000	2000	22.13602	50.95523	1.25508	0	-1.20699	0.002
2001	2001	23.52765	50.51779	1.16907	0	-0.55538	0.002
2002	2002	22.42155	52.33997	1.15914	0	-1.07329	0.002
2003	2003	24.54267	51.86193	1.12222	0	-0.08011	0.002
2004	2004	21.70987	50.10700	1.09606	0	-1.40653	0.002
2005	2005	24.90632	51.61240	1.15157	0	0.09016	0.002
2006	2006	22.22071	51.11649	1.06747	0	-1.16734	0.002
2007	2007	23.06966	54.97107	1.21706	0	-0.76983	0.002
2008	2008	26.82463	55.76103	1.15140	0	0.98838	0.002
2009	2009	25.27903	55.57754	1.18632	0	0.26468	0.002
2010	2010	26.57275	56.57058	1.15101	0	0.87044	0.002
2011	2011	27.45798	56.19070	1.13732	0	1.28493	0.002
2012	2012	28.48317	57.09555	1.18900	0	1.76497	0.002
2013	2013	25.38369	55.41580	1.16608	0	0.31368	0.002
2014	2014	26.17080	57.18450	1.13694	0	0.68223	0.002

Imports NODF

year	Real	Mean	SD	percent	z_score	p_val
2000	58.79083	46.52858	0.14911	100	-0.69948	0.002
2001	58.11077	47.42144	0.14516	100	-1.00007	0.002
2002	61.69127	48.84275	0.14452	100	0.58251	0.002
2003	58.74336	49.23589	0.12859	100	-0.72046	0.002
2004	61.10728	50.29790	0.12980	100	0.32439	0.002
2005	61.62541	50.42682	0.13510	100	0.55340	0.002
2006	62.78834	52.06834	0.13580	100	1.06742	0.002
2007	61.81768	50.23685	0.14435	100	0.63839	0.002
2008	58.35176	47.36546	0.13336	100	-0.89355	0.002
2009	56.08558	45.06599	0.13862	100	-1.89520	0.002
2010	58.90624	47.41923	0.13177	100	-0.64847	0.002
2011	60.70648	50.75142	0.12273	100	0.14724	0.002
2012	60.31692	50.55162	0.14492	100	-0.02495	0.002
2013	61.28029	50.64299	0.12909	100	0.40086	0.002
2014	65.27826	53.70746	0.13785	100	2.16796	0.002

4 Συμπεράσματα

Στην παρούσα διπλωματική εργασία μελετήθηκαν διαχρονικά δίκτυα από τα χρηματοοικονομικά και από τις εισαγωγές-εξαγωγές του παγκόσμιου εμπορίου. Στα δίκτυα αυτά, έχοντας δανειστεί τα βιολογικά μοντέλα αναπαράστασης διμερών δικτύων host-parasites, εξετάστηκαν οι ιδιότητες της εμφωλευσιμότητας (nestedness) και της κοινοτικοποίησης (modularity). Για την αξιολόγηση του modularity χρησιμοποιήθηκαν οι αλγόριθμοι Leading Eigenvector, fast-greedy και LP&BRIM, ενώ για την ιδιότητα nestedness οι αλγόριθμοι NTC και NODF.

Από τη σύγκριση που παρουσιάστηκε στην προηγούμενη ενότητα, για τα δίκτυα χρηματιστηρίου, προκύπτουν οι εξής παρατηρήσεις:

- Πολύ υψηλό density (25%), χαμηλό modularity, nestedness πολύ υψηλότερη του στατιστικά σημαντικού.
- Επιβεβαίωση του Fortuna (2009) ότι σε υψηλή πυκνότητα δικτύου το nestedness και το modularity έχουν αρνητική σχέση.
- Τα μέτρα δίνουν πολύ πιο σαφή εικόνα πριν και στην κρίση του 2008 από την dotcom.
- Χαμηλό modularity (Newman, fast-greedy), υψηλό nestedness (διευκόλυνση διάδοσης κρίσης).
- Υψηλό nestedness και μετά την κρίση.

Τέλος, για τα δίκτυα των εξαγωγών και εισαγωγών που προκύπτουν από τα δεδομένα του παγκόσμιου εμπορίου, επισημαίνονται οι εξής παρατηρήσεις:

- Παρόμοια με τα δίκτυα των μετοχών, τα δίκτυα του εμπορίου παρατηρούνται να έχουν υψηλό density στις περιόδους 2002-2003 και 20013-2014.
- Η χρηματιστηριακή κρίση δεν φαίνεται να επηρεάζει τα δίκτυα των ετών 2007-2008 όσο αφορά το modularity.
- Η ιδιότητα της εμφωλευσιμότητας εμφανίζεται έντονα στα δίκτυα εισαγωγών και εξαγωγών τις περιόδους 2005-2007 και 2013-2014.
- Παρατηρείται έντονη κοινοτικοποίηση τα έτη 2001-2002 και 2009-2010, όπου δομήθηκαν κοινότητες μεταξύ μεγάλων χωρών στις εξαγωγές και μεταξύ μικρών χωρών στις εισαγωγές.
- Η κρίση dot com δε φαίνεται να επηρεάζει καμία από τις δύο ιδιότητες, κάτι που επιβεβαιώνει την αξία της μεθόδου (όπως και στα δίκτυα των μετοχών).

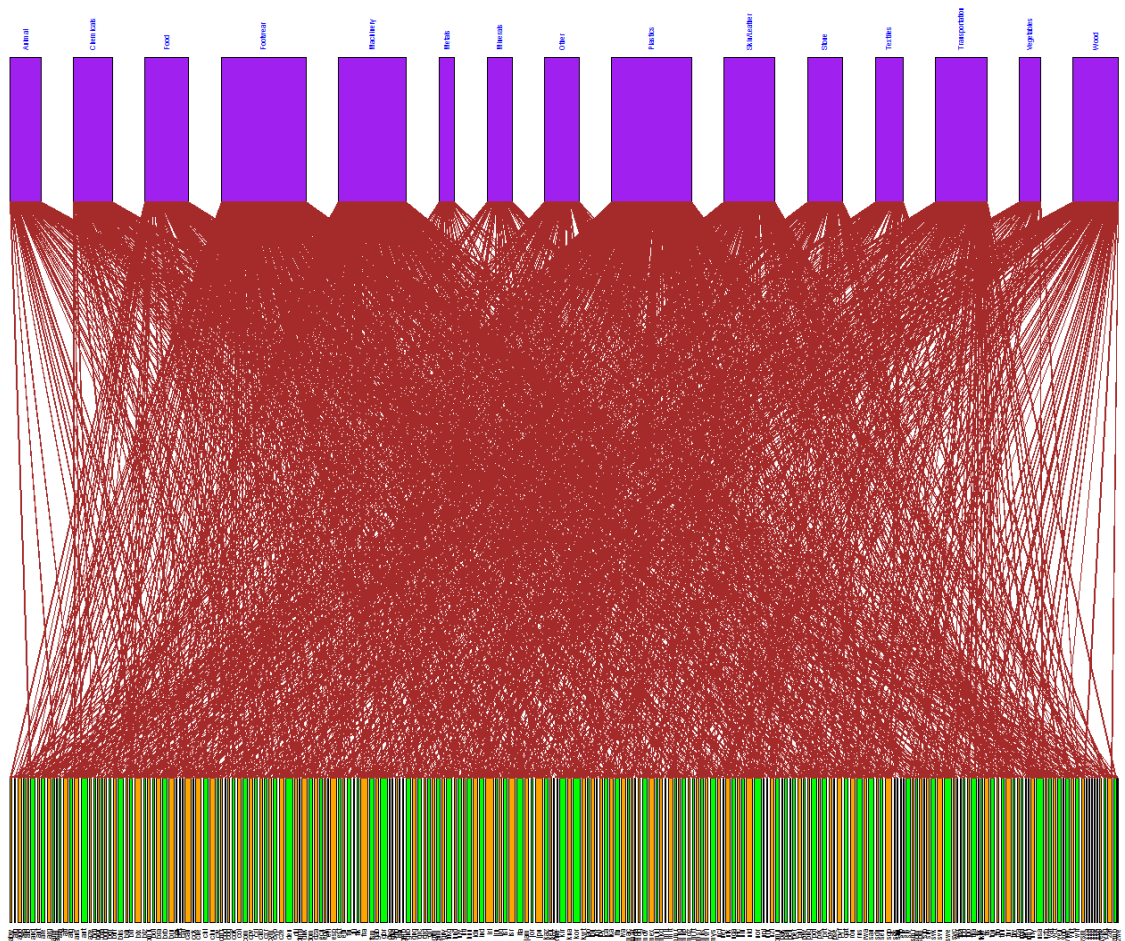
Γενικότερα διαπιστώνεται η χρησιμότητα των δικτύων ως εργαλεία ανάλυσης οικονομικών δεδομένων, είτε αυτά αφορούν χρηματιστηριακές μετοχές είτε το παγκόσμιο εμπόριο. Με βάση τα συμπεράσματα αυτά, ενδιαφέρον για μελλοντική μελέτη θα παρουσιάζε η ανάλυση δεδομένων σε μεγαλύτερο βάθος χρόνου καθώς και η μελέτη των δικτύων εμπορίου με βάση δείκτες συσχέτισης.

Παράρτημα

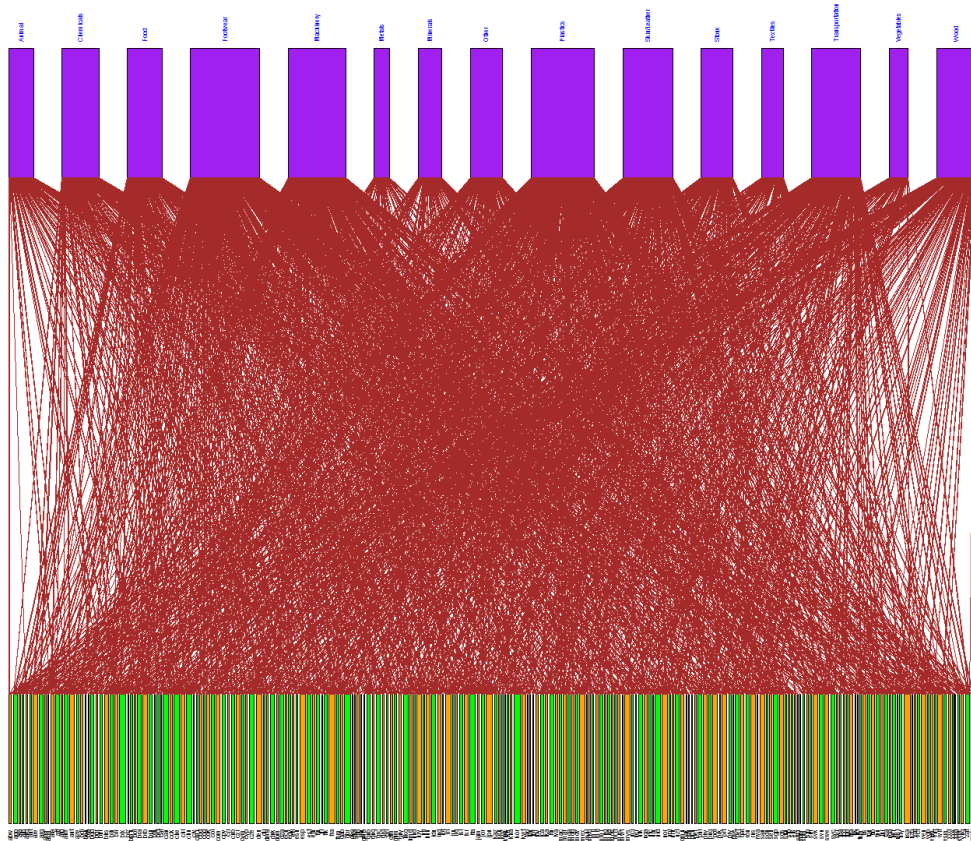
Στο παράρτημα παρουσιάζονται επιλεκτικά μερικές εικόνες ως «στιγμιότυπα» για τα δίκτυα κάποιων ετών που αναφέρθηκαν στην Ενότητα 3.

imports networks plot

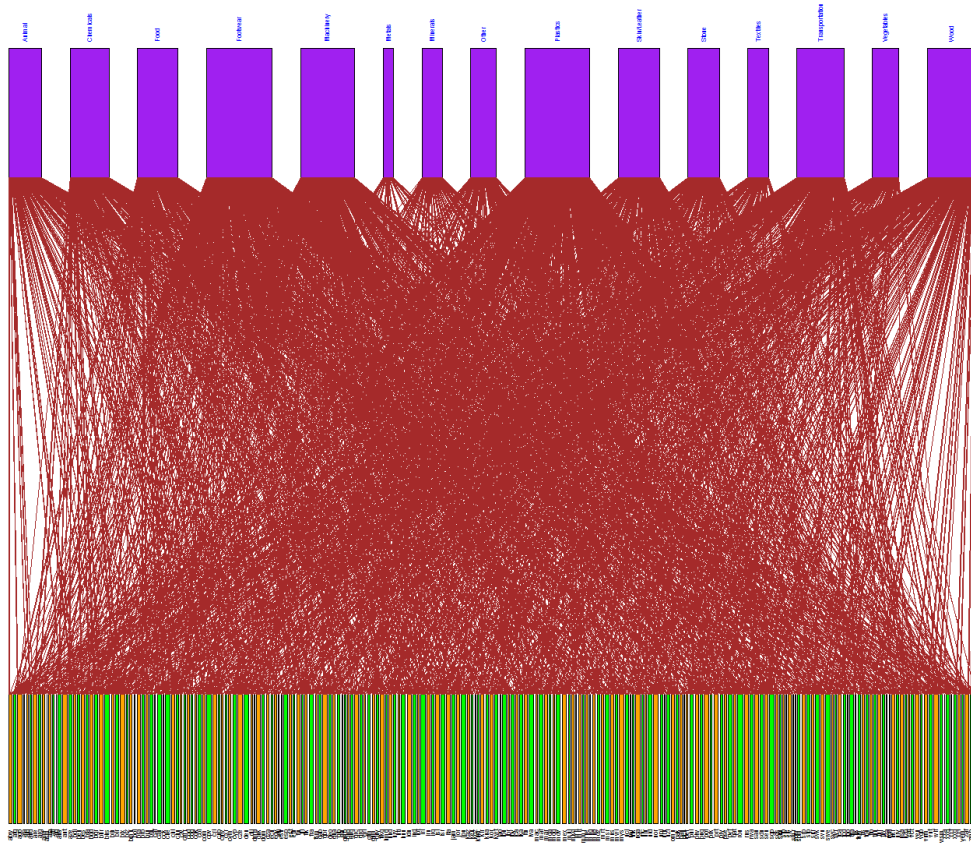
2000



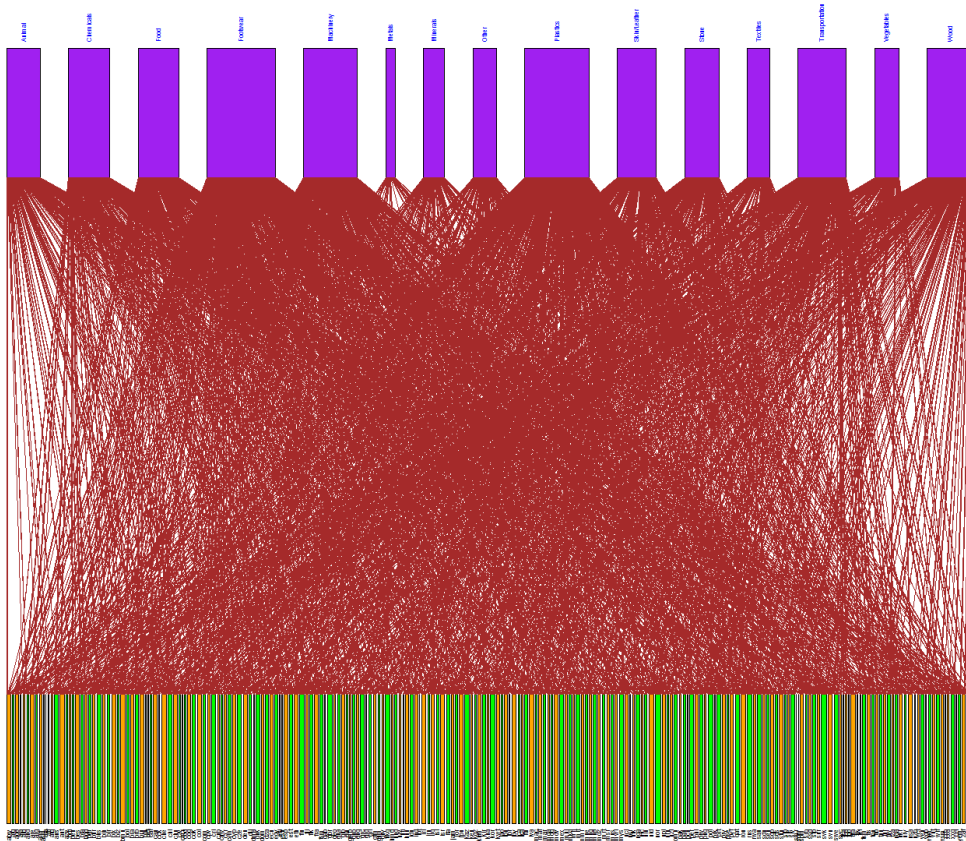
2001



2011

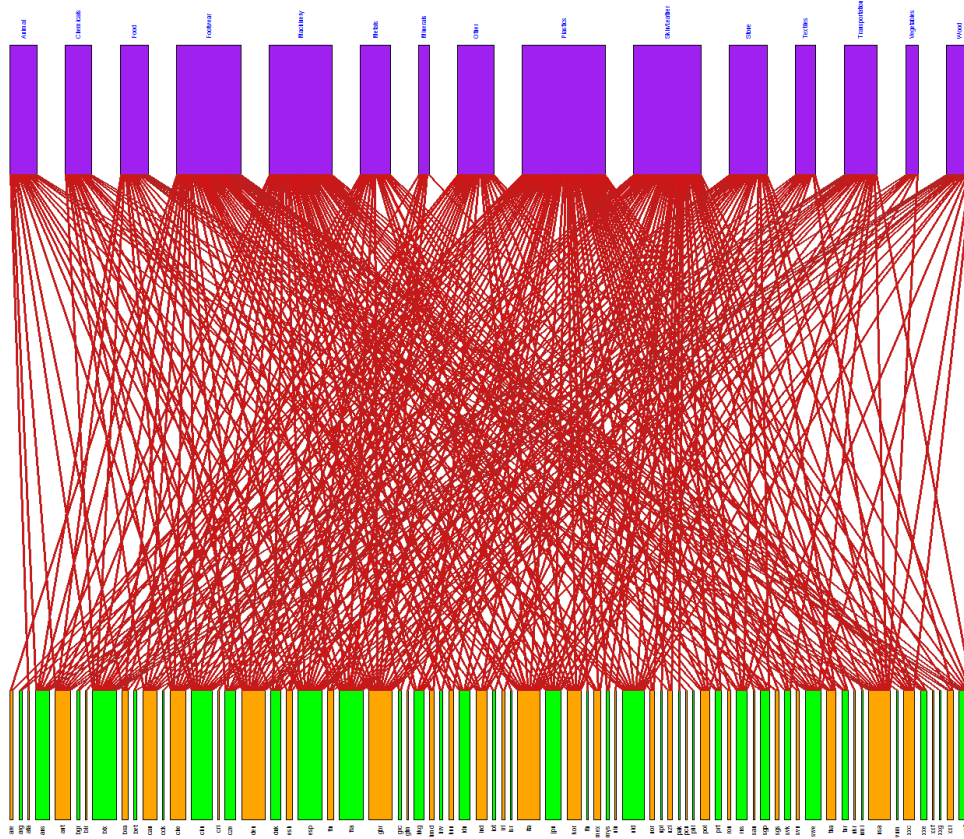


2013

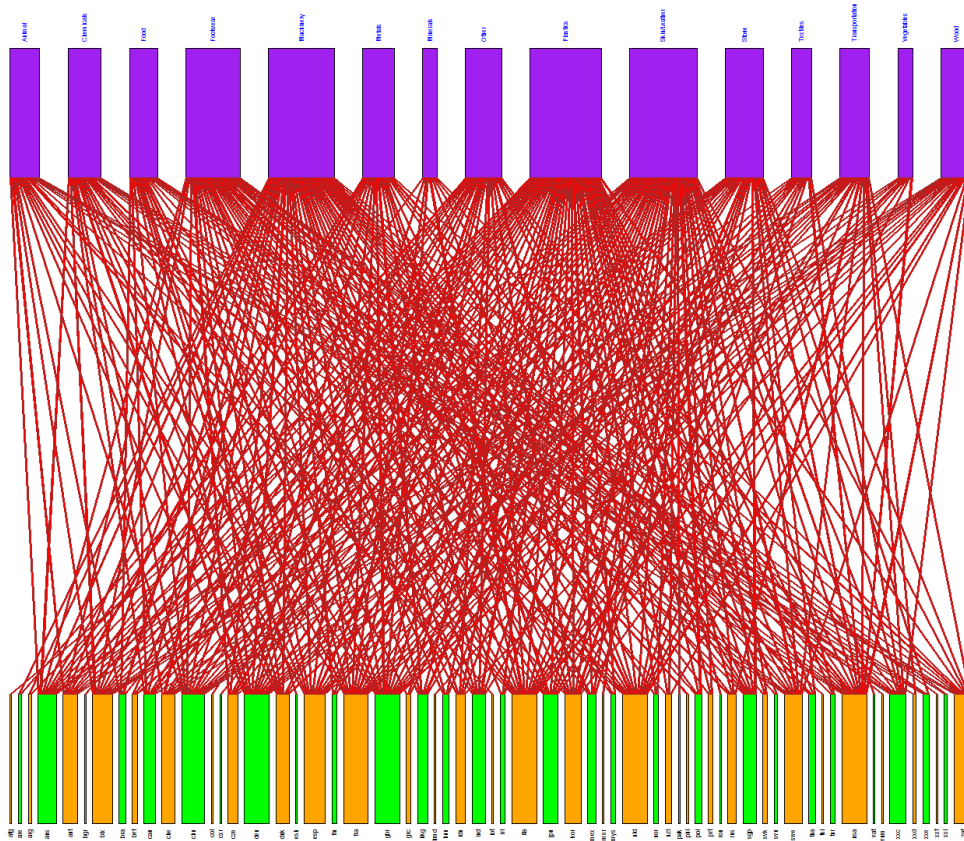


exports networks plot

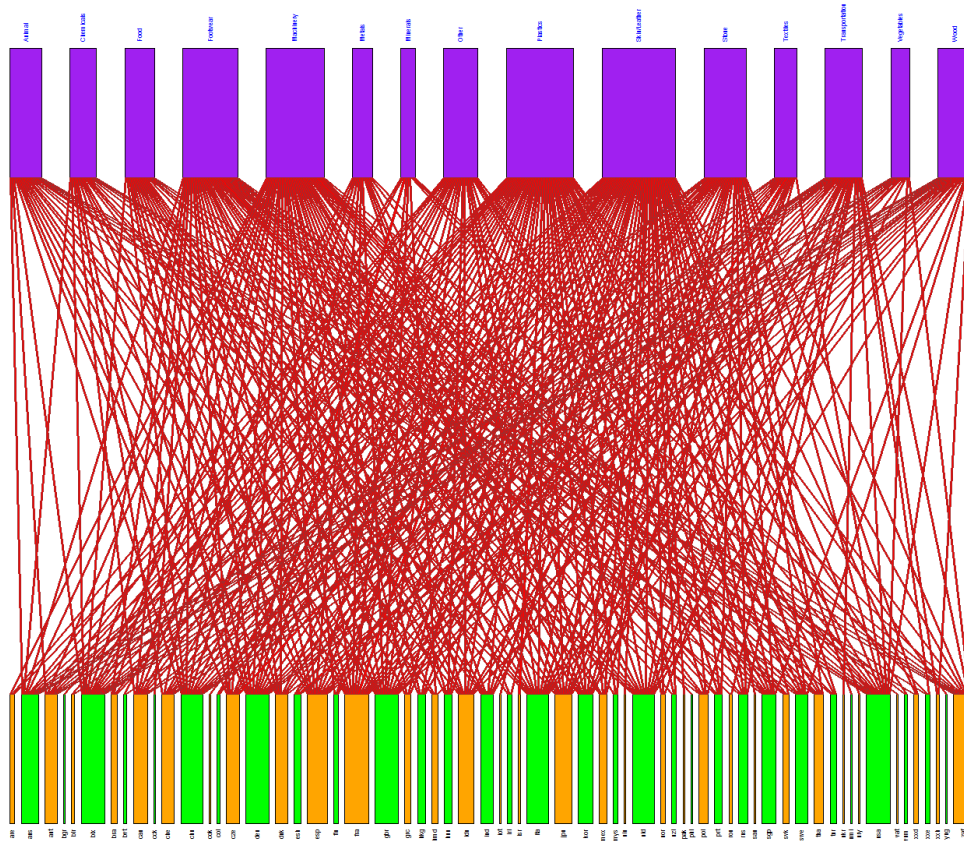
2000

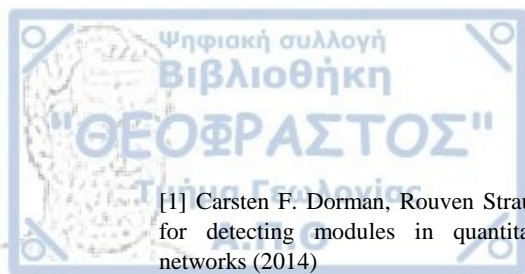


2001



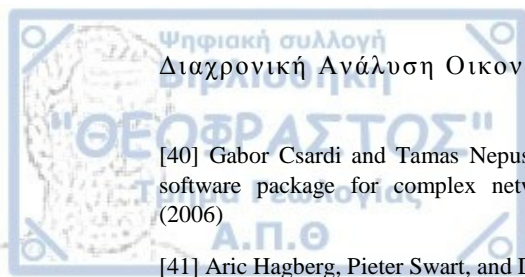
2002



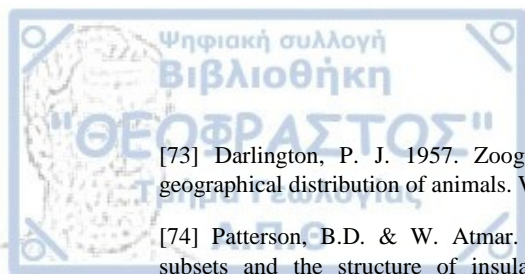


Βιβλιογραφία

- [1] Carsten F. Dorman, Rouven Strauss: A method for detecting modules in quantitative bipartite networks (2014)
- [2] M. E. J. Newman et al: Finding community structure in very large networks (2004)
- [3] Phillip Staniczenko et al: The ghost of nestedness in ecological networks (2012)
- [4] Michael J. Barber: Modularity and community detection in bipartite networks (2007)
- [5] M. E. J. Newman: Communities, modules and large-scale structure in networks (2001)
- [6] M. Girvan and M. E. J. Newman: Community structure in social and biological networks (2001)
- [7] Cesar O. Flores et al: BiMAT : a MATLAB R package to facilitate the analysis and visualization of bipartite networks (2014)
- [8] Cesar O. Flores et al: BiMat Start Guide (2015)
- [9] Tsuyoshi Murata: Modularity for Bipartite Networks, Chapter 7 (2010)
- [10] Taher Alzahrani and K.J. Horadam: Community Detection in Bipartite Networks: Algorithms and Case Studies , Chapter2 (2016)
- [11] Miguel A. Fortuna et al: Nestedness versus Modularity in ecological networks: two sides of the same coin? (2010)
- [12] M. E. J. Newman: Spectral methods for community detection and graph partitioning (2013)
- [13] Tsuyoshi Murata: Community Detection from Bipartite Networks based on Bipartite Modularities (2009)
- [14] Roger Guimera et al: Module identification in bipartite and directed networks (2007)
- [15] Santo Fortunato: Community Detection in Graphs, Complex Networks and Systems Lagrange Laboratory, ISI Foundation, Torino (2010)
- [16] M. E. J. Newman: Detecting community structure in networks (2004)
- [17] <http://cosynet.auth.gr/content/%CF%83%CE%BA%CE%BF%CF%80%CE%BF%CF%82-scope-0>
- [18] Mark Newman. Networks: an introduction. Oxford University Press (2010)
- [19] Hawoong Jeong et al: The large-scale organization of metabolic networks. Nature (2000)
- [20] Joel E Cohen. Food webs and niche space. Number 11. Princeton University Press (1978)
- [21] Jennifer A Dunne. The network structure of food webs. In Ecological Networks: Linking Structure to Dynamics in Food Webs, pages 27-86. Oxford University Press (2006)
- [22] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-world networks (1998)
- [23] Paul Erdos and Alfréd Rényi: On the evolution of random graphs (1960)
- [24] Gary Chartrand: Introductory graph theory (1985)
- [25] Jordi Bascompte and Pedro Jordano. Plant-animal mutualistic networks: the architecture of biodiversity (2007)
- [26] Jordi Bascompte, Pedro Jordano, Carlos J Melián, and Jens M Olesen: The nested assembly of plant-animal mutualistic networks (2003)
- [27] Xin Liu, Tsuyoshi Murata: Community Detection in Large-scale Bipartite Networks (2009)
- [28] Ugo Bastolla, Jordi Bascompte et al: The architecture of mutualistic networks minimizes competition and increases biodiversity (2009)
- [29] Lucas N Joppa, Jordi Bascompte, Jose M Montoya et al: Reciprocal specialization in ecological networks (2009)
- [30] Daniel B Stouffer and Jordi Bascompte. Compartmentalization increases food-web persistence (2011)
- [31] M. E. J. Newman: Modularity and community structure in networks (2006)
- [32] Luca Marotta et al: Bank-Firm Credit Network in Japan: An Analysis of a Bipartite Network (2015)
- [33] J. Olesen, J. Bascompte et al: The modularity of pollination networks (2007)
- [34] Masayasu Kanno: The network structure and systemic risk in the Japanese interbank market (2015)
- [35] Cesar O Flores et al: Statistical structure of host-phage interactions (2011)
- [36] Cesar O Flores et al: Multi-scale structure and geographic drivers of crossinfection within marine bacteria and phages (2013)
- [37] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy: Gephi: an open source software for exploring and manipulating networks (2009)
- [38] Joshua S Weitz, Timothee Poisot et al: Phage-bacteria infection networks (2013)
- [39] M. E. J. Newman, Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), 64 (2001)



- [40] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research (2006)
- [41] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx (2008)
- [42] M. J. Barber, A. Krueger, T. Krueger, and T. Roediger-Schluga, *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 73 (2006)
- [43] Timothee Poisot et al: Resource availability affects the structure of a natural bacteria-bacteriophage community (2011)
- [44] Timothee Poisot et al: The structure of natural microbial enemy-victim networks (2013)
- [45] Paul Shannon, Andrew Markiel et al: Cytoscape: a software environment for integrated models of biomolecular interaction networks (2003)
- [46] C. Christensen, and R. Albert, *International Journal of Bifurcation and Chaos* 17, 2201–2214 (2007)
- [47] S. N. Dorogovtsev, and J. F. F. Mendes, “The shortest path to complex networks,” in *Complex Systems and Inter-disciplinary Science*, edited by N. Johnson, J. Efstathiou, and F. Reed-Tsochas, World Scientific (2004)
- [48] M. E. J. Newman, *SIAM Review* 45, 167–256 (2003)
- [49] R. Albert, and A.-L. Barabási, *Reviews of Modern Physics* 74, 47 (2002)
- [50] M. Barber, M. Faris, L. Streit, O. Strogan: Searching for communities in Bipartite Networks (2004)
- [51] Jane Memmott. The structure of a plant-pollinator food web. *Ecology Letters*, 2(5):276-280 (1999)
- [52] Mario Almeida-Neto, Paulo Guimaraes, Paulo R Guimaraes, Rafael D Loyola, and Werner Ulrich. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement (2008)
- [53] Wirt Atmar and Bruce D Patterson. The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia*, 96:373-382 (1993)
- [54] M A Rodriguez-Giron_es and L Santamar__a. A new algorithm to calculate the nestedness temperature of presence-absence matrices (2006)
- [55] M.J. Newman: Finding community structure in networks using the eigenvectors and eigenvalues (2006)
- [56] Fiedler, M. (1973) *Czech. Math. J.* **23**, 298–305.
- [57] Pothén, A., Simon, H. & Liou, K.-P. (1990) 430–452.
- [58] Leonid E. Zhukov, “Introduction to Network Science”, Lectures, School of Data Analysis and Artificial Intelligence, Department of Computer Science, National Research University Higher School of Economics
- [59] Chung, F. R. K. (1997) *Spectral Graph Theory*, CBMS Regional Conference Series in Mathematics (Am. Math. Soc., Providence, RI), no. 92
- [60] * Ο παράγοντας $\frac{1}{4m}$ χρησιμοποιείται από τον Newman για συμβατότητα σε πηηγούμενο ορισμό του modularity που είχε δοθεί πάλι από τον Newman
- [61] Peter Petersen, *Linear Algebra*, Springer, 2012
- [62] Harvey Mudd College Math Tutorial, The Gram-Schmidt Algorithm
- [63] Haruo Yanai , Kei Takeuchi, Yoshio Takane: *Singular Value Decomposition*, Springer, 2011
- [64] Flake, G. W., Lawrence, S. R., Giles, C. L. & Coetzee, F. M. (2002) *IEEE Computer* **35**, 66–71
- [65] Wasserman, S. & Faust, K. (1994) *Social Network Analysis* (Cambridge Univ. Press, Cambridge, U.K.).
- [66] Radicchi, F., Castellano, C., Cecconi, F., Loreto, V. & Parisi, D. (2004) *Proc.Natl. Acad. Sci. USA* **101**, 2658–2663.
- [67] U. N. Raghavan, R. Albert, and S. Kumara, “Near linear time algorithm to detect community structures in large-scale networks,” *Phys. Rev. E*, vol. 76, no. 3, 036106, Sep. (2007)
- [68] https://en.wikipedia.org/wiki/Label_Propagation_Algorithm
- [69] I. X. Y. Leung, P. Hui, P. Lio, and J. Crowcroft, “Towards real time community detection in large networks,” *Phys. Rev. E*, vol. 79, no. 6, 066107, 2009.
- [70] X. Liu, T. Murata, “How does label propagation algorithm work in bipartite networks?” *Proceedings of the 2009 International Workshop on Intelligent Web Interaction (IWI 09)*, Sep. 2009, in press.
- [71] Cai, S., Zhou, Y., Zhou, T. & Zhou, P. Hierarchical organization and disassortative mixing of correlation-based weighted financial networks. *Int. J. Mod. Phys. C* 21, 433–441 (2010).
- [72] M.Meyer, A.Brintrup, K.Windt, “Linking Product and Machine Network Structure Using Nested Pattern Analysis” *Variety Management in Manufacturing. Proceedings of the 47th CIRP Conference on Manufacturing Systems*, 17 (2014) 278 – 283



- [73] Darlington, P. J. 1957. Zoogeography: the geographical distribution of animals. Wiley.
- [74] Patterson, B.D. & W. Atmar. 1986. Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological Journal of the Linnean Society* 28:65-82
- [75] Atman, W., and B. D. Patterson. 1993. The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia* 96:373–382
- [76] Ulrich W, Almeida-Neto M: On the meanings of nestedness: back to the basics. *Ecography*. 2012; 35(10): 865–871
- [77] Weitz et.al, Phage-bacteria infection networks. *Trends Microbiol*. 2013 Feb;21(2):82-91.
- [78] Memmott, J., Waser, N.M. & Price, M.V. (2004) Tolerance of pollination networks to species extinctions. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, 271, 2605–2611.
- [79] Burgos, E., Ceva, H., Perazzo, R.P.J., Devoto, M., Medan, D., Zimmermann, M. & Maria Delbue, A. (2007) Why nestedness in mutualistic networks? *Journal of Theoretical Biology*, 249, 307–313
- [80] Krause, A.E., Frank, K.J., Mason, D.M., Ulanowicz, R.E. & Taylor, W.W. (2003) Compartments revealed in food web structure. *Nature*, 426, 282–285
- [81] Teng, J. & McCann, K.S. (2004) Dynamics of compartmented and reticulate food webs in relation to energetic flow. *American Naturalist*, 164, 85–100
- [82] Flores et al, Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages, *ISME J*. 2013 Mar; 7(3): 520–532.
- [83] Jonhson S, Domínguez-García V, Muñoz MA (2013) Factors Determining Nestedness in Complex Networks. *PLoS ONE* 8(9): e74025. doi:10.1371/journal.pone.0074025
- [84] Christiano Silva, Thiago, Zhao, Liang. Machine Learning in Complex Networks
- [85] Carlos Aller, Lorenzo Ductor, M.J. Herreras, “The world trade network and the environment”, Elsevier, *Energy Economics*, September 2015, <http://dx.doi.org/10.1016/j.eneco.2015.09.008>
- [86] Luca De Benedictis, Lucia Tajoli, “The world trade network”, January 2008
- [87] Dr. William Chow, “An Anatomy of the World Trade Network”, July 2013
- [88] K. Bhattacharya, G. Mukherjee, J. Saramaki, K. Kaski and S. S. Manna, “The International Trade Network: weighted network analysis and modelling”, *Journal of Statistical Mechanics: Theory and Experiment*, February 2008, doi:10.1088/1742-5468/2008/02/P02002
- [89] Jae Woo Lee, Seong Eun Maeng, Gyeong-Gyun Ha, Moon Hyeok Lee, Eun Seong Cho, “Applications of Complex Networks on Analysis of World Trade Network”, IC-MSQUARE 2012: International Conference on Mathematical Modelling in Physical Sciences IOP Publishing, doi:10.1088/1742-6596/410/1/012063
- [90] Christina Kao, Lili Yang, Ye Yuan, “Predicting Crisis in the Global Trade Network”, December 2015
- [91] Luca De Benedictis, Silvia Nenci, Lucia Tajoli, Claudio Vicarelli, “Network Analysis of World Trade using the BACI-CEPII dataset”, October 2014
- [92] Berk Coker, Charissa Sonder Plattner, Aristidis Papaioannou, “Clustering in Bipartite Graphs: State-Based Trade Networks”, December 2016
- [93] Zhu Z, Cerina F, Chessa A, Caldarelli G, Riccaboni M (2014) The Rise of China in the International Trade Network: A Community Core Detection Approach. *PLoS ONE* 9(8): e105496. doi:10.1371/journal.pone.0105496
- [94] Fabio Saracco, Riccardo Di Clemente, Andrea Gabrielli and Tiziano Squartini, “Detecting early signs of the 2007-2008 crisis in the world trade”, July 2016
- [95] Fabio Saracco, Riccardo Di Clemente, Andrea Gabrielli and Tiziano Squartini, “Randomizing Bipartite Networks: the case of the World Trade Web”. June 2015
- [96] Caldarelli G, Cristelli M, Gabrielli A, Pietronero L, Scala A, et al. (2012) A Network Analysis of Countries’ Export Flows: Firm Grounds for the Building Blocks of the Economy. *PLoS ONE* 7(10): e47278. doi:10.1371/journal.pone.0047278
- [97] Matthew Elliott and Jonathon Hazell, “Endogenous Financial Networks: Efficient Modularity and Why Shareholders Prevent It”, February 2016
- [98] Beckett SJ. 2016 Improved community detection in weighted bipartite networks. *R. Soc. open sci.* 3: 140536. <http://dx.doi.org/10.1098/rsos.140536>
- [99] https://en.wikipedia.org/wiki/Greedy_algorithm
- [100] Alexander Petukhov, Inna Kozlov, “Fast greedy algorithm for subspace clustering from corrupted and incomplete data”, June 2013
- [101] Ravi Kumar, Benjamin Moseley, Sergei Vassilvitskii, Andrea Vattani, “Fast Greedy Algorithms in MapReduce and Streaming”, June 2013