



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

Διπλωματική Εργασία
Τυχαία και Κοινωνικά Δίκτυα
Random and Social Networks
ΧΑΤΖΗΕΛΕΥΘΕΡΙΟΥ ΛΙΒΙΑ ΕΛΕΝΑ

Επιβλέπων: Λουλάκης Μιχαήλ, Επίκουρος Καθηγητής Ε.Μ.Π

Τριμελής Επιτροπή:
Λουλάκης Μιχαήλ, Επίκουρος Καθηγητής Ε.Μ.Π.
Σπηλιώτης Ιωάννης, Αναπληρωτής Καθηγητής Ε.Μ.Π.
Συμβώνης Αντώνιος, Καθηγητής Ε.Μ.Π.

Αθήνα, 2015

Ευχαριστίες

Η εκπόνηση της παρούσας διπλωματικής εργασίας πραγματοποιήθηκε υπό την επίβλεψη του Επ. Καθηγητή του Ε.Μ.Π., Μιχαήλ Λουλάκη. Τον ευχαριστώ ειλικρινά για όλα όσα απλόχερα μου προσέφερε. Για το υπέροχο θέμα το οποίο μου έδωσε την ευκαιρία να εξερευνήσω, τις πολύτιμες συμβουλές του, το συνεχές ενδιαφέρον, τις ατέρμονες συζητήσεις.. Και, φυσικά, για τρόπο του, που έκανε ακόμα και τα πιο ασήμαντα να γίνονται τόσο διασκεδαστικά και ενδιαφέροντα! Σας ευχαριστώ πολύ, υπήρξατε πραγματικός Δάσκαλος για μένα!

Ακόμη, θα ήθελα να ευχαριστήσω τους γονείς μου, Γαβριέλλα και Νίκο, και την αδερφή μου, Λέλα, που με τη διαρκή τους φροντίδα με βοήθησαν να ολοκληρώσω επιτυχώς και αυτό το κομμάτι των σπουδών μου.

Επίσης, ευχαριστώ όλες μου τις φίλες και τους φίλους που έκαναν τα φοιτητικά μου χρόνια τόσο όμορφα. Ιδιαίτερα τις φίλες μου Έφη, Ιωάννα, Σοφία, Ειρήνη και το φίλο μου Μάριο, για την παρουσία τους στις καλές και τις κακές στιγμές.

Τέλος, θα ήθελα να εκφράσω την ευγνωμοσύνη μου προς τη γιαγιά μου, Λίβια, που υπήρξε μια από τις πιο δυνατές επιρροές στη μέχρι τώρα πορεία της ζωής μου, καθώς και τη λύπη μου που έφυγε τόσο απρόσμενα. Ciao Nonnis..

Χατζηελευθερίου Λίβια Έλενα
Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών
Αθήνα, 2015

Περίληψη

Η ραγδαία ανάπτυξη της τεχνολογίας οδηγεί εκ των πραγμάτων στη δημιουργία καινούριων δομών και αντικειμένων που χρειάζονται μελέτη. Για να περιγραφούν αυτά, ως κομμάτι της πραγματικότητας, απαιτείται η γνώση μεγάλου πλήθους παραμέτρων. Το γεγονός αυτό κάνει την ακριβή περιγραφή της πραγματικότητας ένα άπιαστο όνειρο. Λύση στο πρόβλημα αυτό είναι η χρήση μοντέλων. Εξ'άλλου, "All models are wrong, but some of them are useful", όπως είχε πει και ο Άγγλος μαθηματικός G.Box.

Στην εργασία αυτή θα παρουσιαστούν μοντέλα που επιχειρούν να περιγράψουν δίκτυα. Τα μοντέλα που θα παρουσιαστούν είναι τυχαία κ συνήθως είναι χρήσιμο και βολικό να μελετώνται οι οριακές τους συμπεριφορές.

Στο πρώτο κεφάλαιο της εργασίας αυτής αναφέρονται κάποιες εισαγωγικές έννοιες σχετικές με τη Θεωρία Γραφημάτων και τη Θεωρία Πιθανοτήτων.

Στο δεύτερο κεφάλαιο, μελετάται το μοντέλο των Erdős και Rényi για τους τυχαίους γράφους. Σύμφωνα με αυτό, για κάθε ζεύγος i, j κόμβων ενός κενού γραφήματος, έχουμε δημιουργία της ακμής ij με πιθανότητα p . Το μοντέλο αυτό αφορά σε γράφους σταθερού μεγέθους. Μπορεί, όμως μια μικρή αλλαγή στην τιμή της πιθανότητας p να ευθύνεται για την αλλαγή της δομής ολόκληρου του γραφήματος; Στο κεφάλαιο αυτό μελετώνται οι οριακές συναρτήσεις για την πιθανότητα p για διάφορες ιδιότητες, όπως για την κατανομή των βαθμών, τη δομή, τη συνεκτικότητα, τις αποστάσεις σε γραφήματα που προκύπτουν από τους κανόνες που αναφέρθηκαν προηγουμένως.

Στο τέλος του τρίτου κεφαλαίου, γίνεται μελέτη κάποιων μέτρων κεντρικότητας στα γραφήματα. Θεωρήθηκε ένα γράφημα και σε αυτό υπολογίστηκαν οι τιμές για κάποια από τα μέτρα αυτά.

Τα τυχαία ER γραφήματα, που μελετήθηκαν στο προπροηγούμενο κεφάλαιο, δεν είναι, συνήθως, ικανά να περιγράψουν επιτυχώς πραγματικά και κοινωνικά δίκτυα. Παρ'όλ' αυτά, είναι πολύ χρήσιμα σαν σημεία αναφοράς, για τη σύγκριση των δεύτερων με τα πρώτα, από την οποία μπορούμε να λάβουμε πληροφορίες σχετικά με το κατά πόσο ένα δίκτυο έχει δημιουργηθεί με τρόπο τυχαίο. Οι δύο βασικές συμπεριφορές που παρατηρούνται σε πραγματικά και κοινωνικά δίκτυα και που καταστούν το προηγούμενο μοντέλο ανίκανο για την περιγραφή τους είναι η γιγάντωση και η προτίμηση για τη σύναψη συνδέσεων με τις κεντρικότερες οντότητες στο γράφημα. Έτσι, στο τέταρτο κεφάλαιο, γίνεται μελέτη του μοντέλου των Barabási και Albert.

Το μοντέλο αυτό, ή αλλιώς το Preferential Attachment μοντέλο, ορίζεται με τρόπο τέτοιο που ενσωματώνει τις δύο συμπεριφορές που παρατηρούνται σε κοινωνικά και πραγματικά δίκτυα και που αναφέρθηκαν προηγουμένως. Μια βασική διαφορά με το ER μοντέλο, είναι ότι στο BA παράγει ακολουθίες γραφημάτων και όχι γραφήματα. Έτσι, με βάση τις διάφορες παραμέτρους του μοντέλου, εξετάζεται η συμπεριφορά και η χρονική εξέλιξη του γραφήματος, στα οποία κυριαχεί ο κανόνας ότι ο πλούσιος γίνεται πλουσιότερος (πιο γρήγορα από το φτωχό). Παρατηρούμε ότι ισχύουν εντελώς

διαφορετικά αποτελέσματα από αυτά που ισχύουν για ER γραφήματα.

Τέλος, στο τέταρτο κεφάλαιο, γίνεται ανάλυση ενός πραγματικού κοινωνικού δικτύου. Το δίκτυο αυτό έχει για κόμβους τεταρτοετείς φοιτητές της ΣΕΜΦΕ. Ακμή από τον i κόμβο προς τον j κόμβο υπάρχει όταν ο j αναφέρεται από τον i στην απάντησή του στην ερώτηση "Σε ποιον συμφοιτητή σου απευθύνεσαι όταν έχεις απορίες σχετικά με μαθήματα της σχολής;". Μελετώνται διάφορες ιδιότητες του γραφήματος και συσχετίσεις μεταξύ διαφόρων ποσοτήτων. Η ανάλυση γίνεται με χρήση του Rajeck και της R.

Abstract

The great technological developments in recent decades have led practically to the creation of new structures and objects that require further study. In order to describe them as part of the reality, we need to know a big amount of parameters. For this reason the exact description of reality looks like an elusive dream. This obstacle can be overcome with the use of models. Besides, according to the English mathematician G.Box “All models are wrong, but some of them are useful”.

In this thesis we present models which try to describe networks. These models are random and usually it is useful and convenient for us to study their limit behaviours. In the first chapter, some introductory concepts are mentioned which are related to Graph and Probability theories.

In the second chapter, we study the Erdős and Rényi model about random graphs. According to this, for every couple of i, j nodes in an empty graph, we have the creation of the ij edge with probability p . This model concerns graphs with fixed size. However, is it possible to have a big change to the graph structure due to a small change in the probability? In this chapter, we study the threshold functions for the probability p for some properties such as degree distribution, structure, connectedness, the distances in graphs that derive from the rules we have referred to.

To end the second chapter, we study some centrality measures in graphs. We consider a graph and we calculate the values for some of these centrality measures.

The random ER graphs we studied above cannot describe successfully real and social networks. Nevertheless, they are very useful as benchmarks, for instance for the comparison of social with random networks from which we can get information about how much randomly this network has been created. The two basic behaviours we observe in real and social networks that make the previous model unable to describe them are the growth and preference for the junction of links with more central entities in the graph. Thus, in this fourth chapter we review the Barabási and Albert model.

This model, which is also known as Preferential Attachment model, is defined in such a manner that incorporates these two previously referred behaviours that are met in social and real networks. A basic distinction between these two models is that while the ER model creates random graphs, the BA model creates sequences of random graphs. As a result, taking into account the various parameters of the model, the behavior and the time evolution of a graph is examined. In this kind of graphs, due to the preferential attachment property, in which there is the dominating rule that “the rich gets richer” (faster than the poor). From this, we can notice that completely different results occur in this model in contrast to ER graphs.

Finally, in the fourth chapter we analyse a real social network. This network includes 4th-grade students of the School of Applied Mathematical and Physical Sciences as nodes. There is an arc from the i node to the j node if the j is mentioned by the i in his/her answer to the question “From which student do you ask information when you

don't understand something at university subjects?". We study several properties of the graph and relations between some quantities. The analysis is achieved with the use of the Pajek and R softwares.

Περιεχόμενα

1	Εισαγωγή	15
1.1	Θεωρία Γραφημάτων-Γενικά	15
1.2	Πιθανότητες-Κατανομές	21
1.2.1	Η Πιθανότητα	21
1.3	Τυχαίες Μεταβλητές	23
1.3.1	Κατανομές	25
1.4	Χρήσιμες Σχέσεις	27
2	Random Graphs: The Poisson Model	29
2.1	Γενικά	29
2.2	Ορισμός Τυχαίων Γράφων	31
2.3	Η Κατανομή των Βαθμών στο γράφημα	33
2.4	Οριακές Συναρτήσεις και Μεταβολή Φάσης (Threshold Functions and Phase Transition)	38
2.5	Οριακές Συναρτήσεις και Δομή του Γραφήματος	40
2.6	Οριακές Συναρτήσεις και Διάμετρος	49
2.7	«Το γράφημα»	60
2.8	Μέτρα Κεντρικότητας	62
2.8.1	Σύγκριση μεταξύ μέτρων.	89
2.8.2	Clustering Coefficient	90
2.8.3	Τυχαίοι ER γράφοι και κεντρικότητα.	90
2.8.4	Clustering Coefficient σε τυχαίους Erdős Rényi γράφους.	91
2.8.5	Πραγματικά δίκτυα και μέτρα κεντρικότητας.	91
2.8.6	Πραγματικά δίκτυα και Clustering Coefficient.	91
3	Random Graphs: To Preferential Attachment μοντέλο	93
3.1	Εισαγωγή	93
3.2	Ορισμός του Μοντέλου	95
3.3	Η κατανομή των βαθμών	97
3.4	Συνεκτικότητα, Αποστάσεις και Διάμετρος.	102
3.4.1	Για το πλήθος των συνεκτικών συνιστωσών	102
3.4.2	Διάμετρος	107
4	Το «πείραμα»	109

Κατάλογος σχημάτων

1.1	Γραφική αναπαράσταση γράφου	16
1.2	K_5	18
1.3	Διμερές γράφημα	18
1.4	K_4	19
1.5	K_5 και K_5 μείον μια ακμή	20
2.1	ER Γράφημα με $p=0.015$, $v=60$	34
2.2	Η κατανομή των βαθμών για ER Γράφημα με $p=0.015$, $v=60$ και η σύγκριση της με της Poisson.	34
2.3	ER Γράφημα με $p=0.07$, $v=60$	35
2.4	Κατανομή των βαθμών για το ER Γράφημα με $p=0.07$, $v=60$ και σύγκριση με Poisson.	35
2.5	Τυχαίος ER γράφος με $\nu = 100$, $p = 0.035$	41
2.6	Τυχαίος ER γράφος με $\nu = 100$, $c = 1/4$, $N_c = 175$	44
2.7	Η συνάρτηση $x(c) = \sum_{k=1}^{+\infty} \frac{\binom{k}{k-1}}{k!} (2ce^{-2c})^k$	46
2.8	Η σχέση του λόγου του μέσου πλήθους των συνιστωσών και του πλήθους των κόμβων του γραφήματος είναι φίνουσα καθώς το πλήθος των κόμβων αυξάνει.	46
2.9	Εφαρμογή του Θεωρήματος 2.5.4, για $c = 1/4$	47
2.10	Εφαρμογή του Θεωρήματος 2.5.4, για $c = 1/2$	48
2.11	Εφαρμογή του Θεωρήματος 2.5.4, για $c = 2$	48
2.12	Μερίδιο των κόμβων του γραφήματος που βρίσκονται σε κάποιο απομονωμένο δένδρο συναρτήσει της σταθεράς c	49
2.13	Εφαρμογή του παραπάνω Θεωρήματος	55
2.14	Γραφική παράσταση της συνάρτησης G	58
2.15	Γράφημα με μαθήματα της ΣΕΜΦΕ	61
2.16	Παιχνίδι	62
2.17	Κανονικοποιημένοι Εσωτερικοί Βαθμοί των κόμβων του "Γραφήματος"	67
2.18	Κανονικοποιημένοι Εξωτερικοί Βαθμοί των κόμβων του "Γραφήματος"	68
2.19	Γράφημα για την κανονικοποιημένη (εσωτερική) εγγύτητα των κόμβων του «Γραφήματος».	75
2.20	Γράφημα για την κανονικοποιημένη (εξωτερική) εγγύτητα των κόμβων του «Γραφήματος»	76
2.21	Οι κόμβοι του «Γραφήματος» και η συμμετοχή τους στα γεωδαισιακά μονοπάτια του.	80
2.22	Γράφημα	82
2.23	Γράφημα	86
2.24	Γράφημα.	89

2.25	Σύγκριση μεταξύ του συντελεστή clustering για πραγματικά δίκτυα και τυχαίους ER γράφους.	92
3.1	Κατανομή των βαθμών σε γράφημα που έχει κατασκευαστεί με βάση τις προηγούμενες διαδικασίες. Το (a) έχει $N = m_o + t = 300\,000$ και $m_o = m = 1$ για τους κύκλους, $m_o = m = 5$ για τους ρόμβους και $m_o = m = 7$ για τα τρίγωνα. Η κλίση της διακεκομμένης γραμμής είναι 2.9. Το μικρό γράφημα στο εσωτερικό προκύπτει από τα δεδομένα του μεγάλου διαιρεμένα με τη σταθερά m^2 και η κλίση της διακεκομμένης γραμμής στο μικρό γράφημα είναι 3. Το (b) έχει $m_o = m = 5$ και $N = 100\,000$ για τους κύκλους, $N = 150\,000$ για τα τετράγωνα και $N = 200\,000$ για τους ρόμβους. Το μικρό πλαίσιο στο εσωτερικό του δείχνει τη χρονική εξέλιξη των βαθμών δύο κόμβων που προστέθηκαν στο σύστημα τις χρονικές στιγμές $t_1 = 5$ και $t_2 = 95$, με $m_o = m = 5$ και κλίση της διακεκομμένης γραμμής ίση με 0.5, όπως προβλέπεται από τη μέθοδο.	98
4.1	Γράφημα συσχέτισης φοιτητών 4ου έτους.	111
4.2	Κατευθύνσεις που έχουν ακολουθηθεί από τους φοιτητές.	113
4.3	Κατευθύνσεις και ασθενώς συνεκτικές συνιστώσες.	114
4.4	Εσωτερικός Βαθμός φοιτητών.	115
4.5	Διαμεσολαβησιμότητα φοιτητών.	117
4.6	Γράφημα με Articulation Points, κρίσιμους για τη διάδοση της πληροφορίας φοιτητές στο γράφημα.	118
4.7	Γράφημα με τους φοιτητές που μεταδίδουν πληροφορία πιο συχνά.	119
4.8	Γράφημα με τις ισχυρά συνεκτικές συνιστώσες του γραφήματός μας.	120
4.9	Αποτελέσματα που αφορούν στη συσχέτιση βαθμολογίας και εσωτερικών βαθμών.	120

Κατάλογος πινάκων

2.1	Εσωτερικοί βαθμοί των κόμβων του "Γραφήματος"	65
2.2	Πίνακας με τους εξωτερικούς βαθμούς των κόμβων του "Γραφήματος"	65
2.3	In-Closeness Κεντρικότητα των κόμβων του "Γραφήματος"	71
2.4	Out-Closeness Κεντρικότητα των κόμβων του "Γραφήματος"	71
2.5	Betweenness Κεντρικότητα των κόμβων του "Γραφήματος"	78
2.6	Betweenness Κεντρικότητα των κόμβων της μη κατευθυνόμενης παραλλαγής του "Γραφήματος"	79
2.7	Μέτρα κεντρικότητας για τους κόμβους του Γραφήματος 2.8.1	89
2.8	Πίνακας σύγκρισης. Οι στήλες είναι από τα αριστερά προς τα δεξιά: Δίκτυο, Μέγεθος του δικτύου, μέσος βαθμός των κόμβων, μέσο μήκος μονοπατιών, μέσο μήκος μονοπατιών σε τυχαίο ER γράφο ίδιου μεγέθους, συντελεστής Clustering για τα πραγματικά δίκτυα της πρώτης στήλης και μετά για τυχαία ER γραφήματα ίδιας τάξης. Τέλος, στη δεξιά στήλη, αναφέρονται οι πηγές από όπου έχουν προκύψει αυτά τα αποτελέσματα. Ο Πίνακας αυτός έχει βρεθεί στο [1]	92
4.1	Πίνακας με τους εσωτερικούς βαθμούς των κόμβων.	112
4.2	Ασθενώς Συνεκτικές Συνιστώσες.	112
4.3	Για τις Ασθενώς Συνεκτικές Συνιστώσες.	113
4.4	Εσωτερικός Βαθμός φοιτητών.	116
4.5	Εγγύτητα φοιτητών.	116
4.6	Πίνακες με τους 10 κεντρικότερους φοιτητές για κάθε μέτρο.	117
4.7	Πίνακας με τις αποστάσεις από τον κόμβο M3.	119

Κεφάλαιο 1

Εισαγωγή

1.1 Θεωρία Γραφημάτων-Γενικά

Θα ήταν χρήσιμο να ξεκινήσουμε με κάποιους βασικούς ορισμούς που συναντώνται συχνά στη Θεωρία Γραφημάτων, και που θα συναντήσουμε και στη συνέχεια.

• Έστω V ένα μη κενό σύνολο από διακριτά σημεία K_1, K_2, \dots, K_v και E ένα σύνολο από μη διατεταγμένα ζευγάρια (K_i, K_j) όπου $i \neq j$ και $K_i, K_j \in V$.

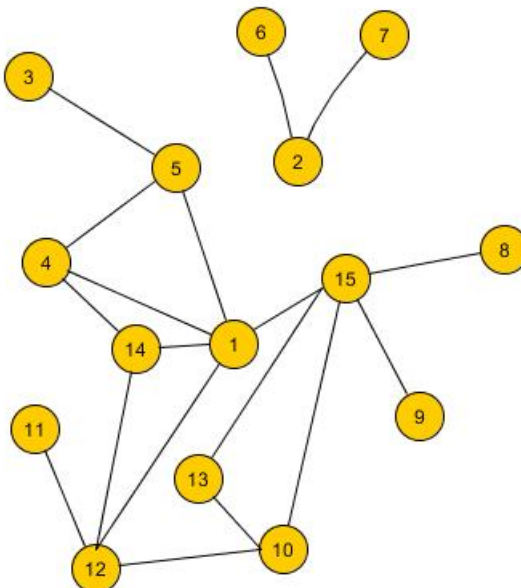
Γράφος, ή **γράφημα**, καλείται το σύνολο $G = \{V, E\}$. Τα σημεία καλούνται κορυφές, ή κόμβοι, και τα ζευγάρια (K_i, K_j) καλούνται πλευρές, ή ακμές. Τάξη του γραφήματος ονομάζεται το $|V|$ και μέγεθος το $|E|$. Αν το ζεύγος $(K_i, K_j) \in E$, τότε οι δύο κορυφές ονομάζονται γειτονικές. Δύο πλευρές ονομάζονται γειτονικές σε περίπτωση που έχουν ένα κοινό άκρο.

Για παράδειγμα, στο Σχήμα 1.1 της επόμενης σελίδας έχουμε τη γραφική αναπαράσταση του γράφου

$$G = \{\{1, 2, \dots, 15\} \{1 - 4, 1 - 5, 1 - 14, 1 - 12, 1 - 15, 2 - 6, 2 - 7, 3 - 5, 4 - 5, 4 - 14, 8 - 15, 9 - 15, 10 - 15, 10 - 12, 10 - 13, 11 - 12, 12 - 14, 13 - 15\}\}.$$

• **«Δίκτυο»** μπορούμε να πούμε ότι είναι ένα σύνολο με οντότητες και συσχετίσεις μεταξύ των οντοτήτων αυτών. Ένα δίκτυο μπορεί να αναπαρασταθεί από ένα γράφο. Ένας γράφος με τη σειρά του, πέρα από γραφικά, μπορεί να αναπαρασταθεί από έναν πίνακα, τον πίνακα γειννίας. Ένας από τους διαχωρισμούς που μπορούμε να κάνουμε στους γράφους είναι να δούμε αν αυτός είναι **«κατευθυνόμενος»** ή **«μη κατευθυνόμενος»**. Στους μη κατευθυνόμενους γράφους όταν υπάρχει συσχέτιση μεταξύ δύο οντοτήτων, τότε αυτή είναι συμμετρική και συνήθως υπονοείται ότι η σχέση που υπάρχει ανάμεσα στους δύο κόμβους χρειάζεται τη «συμφωνία» και των δύο πλευρών για να συνεχίσει να υπάρχει: αν η Αλίκη είναι φίλη του Βασιλή, τότε (θα περιμένουμε) και ο Βασίλης να είναι φίλος της Αλίκης. Ή, στη γραφική αναπαράστασή του γραφήματος θα υπάρχει ακμή ανάμεσα στους κόμβους της Αλίκης και του Βασιλή. Στους κατευθυνόμενους γράφους η σχέση μεταξύ δύο κόμβων που ενώνονται δεν είναι απαραίτητα συμμετρική: Η Γεωργία είναι κόρη της Δήμητρας, αλλά η Δήμητρα δεν μπορεί να είναι κόρη της Γεωργίας. Στην περίπτωση αυτή, τα ζεύγη $K_i K_j$ είναι διατεταγμένα και θεωρούμε ότι υπάρχει ακμή με αρχή το K_i και πέρας το K_j . Ο διαχωρισμός ανάμεσα

σε κατευθυνόμενα και μη γραφήματα δεν είναι τυπικός. Όπως θα φανεί και στη συνέχεια, τα αποτελέσματα των αναλύσεων μπορεί να είναι πολύ διαφορετικά για τους δύο αυτούς τύπους γραφημάτων.



Σχήμα 1.1: Γραφική αναπαράσταση γράφου

- Μια ακμή, πέρα από κατεύθυνση, μπορεί να έχει και «**Βάρος**». Το βάρος της ακμής μπορεί να αναπαριστά κάποιο κόστος, λχ για μετακίνηση από μία πόλη σε μια άλλη, κάποιο κέρδος, τις ώρες ενασχόλησης με ένα αντικείμενο και πολλά ακόμη.

- Ένας πίνακας λέγεται «**Στοχαστικός**» αν για κάθε κόμβο i ισχύει $\sum_j \alpha_{ij} = 1$, όπου j κάθε άλλος κόμβος του γραφήματος. Δηλαδή, αν το άθροισμα των βαρών κάθε γραμμής, ή και στήλης αν είναι συμμετρικός ο πίνακας, είναι ίσο με τη μονάδα. Με έναν τέτοιο πίνακα μας δίνεται η πιθανότητα με την οποία, ξεκινώντας από τον κόμβο i , θα βρεθούμε σε καθέναν από τους υπόλοιπους κόμβους του γραφήματος στο επόμενο βήμα. Για μη κατευθυνόμενο γράφημα, αν δύο κόμβοι i, j δεν συνδέονται άμεσα, το στοιχείο α_{ij} θα είναι ίσο με μηδέν, αφού δεν υπάρχει ακμή που να συνδέει τους κόμβους αυτούς και άρα δεν υπάρχει τρόπος να πάμε από τον έναν στον άλλον με ένα βήμα. Για κατευθυνόμενο γράφημα, μπορεί να μπορούμε από τον i να πάμε στον j με ένα βήμα αλλά όχι αντίστροφα. Επομένως, για κατευθυνόμενα γραφήματα, δεν είναι απαραίτητο να ισχύει και $\sum_i \alpha_{ij} = 1$ ταυτόχρονα, όπως για παράδειγμα στον παρακάτω πίνακα.

$$\begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 2/3 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

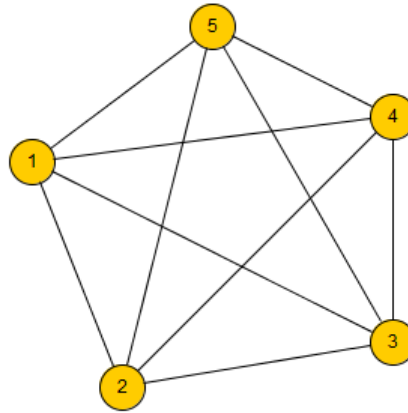
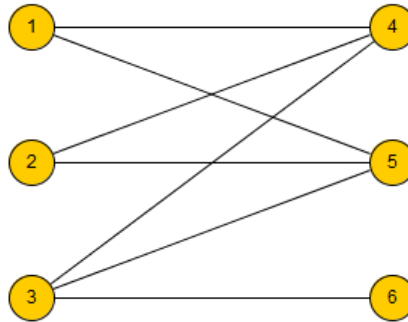
• **«Δυνάμεις Πίνακα»:** Έστω ένας πίνακας γειτνίασης ενός μη κατευθυνόμενου γράφου. Ο πίνακας αυτός αποτελείται από "1" και "0", ανάλογα με το αν συνδέονται ή όχι δύο κόμβοι με ακμή. Υψώνοντας τον πίνακα αυτόν στο τετράγωνο λαμβάνουμε έναν καινούριο πίνακα, που όμως μας δείχνει από ποιον κόμβο προς τα ποιον κόμβο υπάρχουν μονοπάτια μήκους 2 και πόσα είναι αυτά. Υψώνοντας στην n -οστή δύναμη, λαμβάνουμε πληροφορία για την ύπαρξη μονοπατιών μήκους n ανάμεσα σε δύο κόμβους και για το πλήθος τους. Στον πίνακα αυτόν λαμβάνονται υπ' όψιν όλα τα μονοπάτια μήκους n , ακόμη και αν αυτά περιέχουν κύκλους.

• Ένας μη κατευθυνόμενος γράφος λέγεται **«συνεκτικός»** εάν από κάθε σημείο του υπάρχει μονοπάτι προς κάθε άλλο σημείο του. Τα μεγιστικά συνεκτικά υπογραφήματα ενός γραφήματος καλούνται συνεκτικές «συνιστώσες» του. Οποιοσδήποτε συνεκτικές συνιστώσες του γραφήματος είναι ξένες μεταξύ τους. Εάν ένα γράφημα δεν είναι συνεκτικό ονομάζεται μη συνεκτικό. Έτσι, το γράφημα της Εικόνας 1 είναι μη συνεκτικό, αποτελούμενο από 2 συνιστώσες. **«Γέφυρα»** λέγεται οποιοσδήποτε κόμβος ενός συνεκτικού γραφήματος που, αν αφαιρεθεί από το γράφημα, το καθιστά μη συνεκτικό. Ένα κατευθυνόμενο γράφημα λέγεται **«ασθενώς συνεκτικό»** αν το αντίστοιχο μη κατευθυνόμενο γράφημα θα ήταν συνεκτικό και **«ισχυρά συνεκτικό»** αν υπάρχει προσανατολισμένο μονοπάτι από κάθε κόμβο του γραφήματος προς κάθε άλλον. Η ύπαρξη συνιστωσών σε ένα γράφημα καθορίζει πράγματα όπως η διάδοση φαινομένων σε αυτό, όπως θα μπορούσε να είναι μια πληροφορία ή κάποια ασθένεια.

• **«Διμερές»** ονομάζεται το γράφημα εκείνο του οποίου οι κορυφές μπορούν να χωριστούν σε δύο ξένες ομάδες A και B έτσι ώστε κάθε ακμή να ενώνει κόμβους που ανήκουν σε διαφορετική ομάδα. Τριμερές ονομάζεται το γράφημα όταν οι κορυφές του χωρίζονται σε 3 ομάδες, και παρόμοια, n -μερές όταν χωρίζονται σε n ομάδες. **«Πλήρες»** ονομάζεται ένα γράφημα στο οποίο υπάρχει ακμή από κάθε κορυφή προς κάθε κορυφή. Τέτοια γραφήματα ονομάζονται και **«Κλίκες»** και έχουν $\binom{n}{2}$ ακμές, όπου n το πλήθος των κόμβων τους. Το K_5 (Σχήμα 1.2) είναι το πλήρες γράφημα με 5 κόμβους και $\binom{5}{2} = 10$ πλευρές και το $K_{3,3}$ (Σχήμα 1.3) είναι ένα από τα διμερή με 3 και 3 κόμβους. Αν υπήρχαν και οι πλευρές 1-6, 2-6 το γράφημα θα ήταν διμερές και πλήρες και θα είχε $|A| * |B| = 3 * 3 = 9$ ακμές.

• Σε ένα γράφημα θεωρούμε μια πεπερασμένη ακολουθία στην οποία έχουμε εναλλάξ κόμβους και ακμές, ξεκινώντας και τελειώνοντας σε κάποιον κόμβο. Κάθε πλευρά που υπάρχει στην ακολουθία ξεκινάει από την κορυφή που βρίσκεται πριν από αυτήν και προσπίπτει στην κορυφή που είναι μετά από αυτήν. Την ονομάζουμε **«δρόμο»**. Ένας τέτοιος στο γράφημα του Σχήματος 1.1 είναι η ακολουθία 12, 12 – 14, 14, 14 – 1, 1, 1 – 15, 15. Ο δρόμος στον οποίο κάθε ακμή εμφανίζεται μια μόνο φορά ονομάζεται **«δρομίσκος»** (μια κορυφή μπορεί και να υπάρχει περισσότερες από μια φορές). Αν και οι ακμές εμφανίζονται από μια μόνο φορά, τότε η ακολουθία λέγεται **«μονοπάτι»**. Η ακολουθία που ξεκινάει και τελειώνει στον ίδιο κόμβο λέγεται **«κλειστός δρόμος»**, **«κλειστός δρομίσκος»** και **«κλειστό μονοπάτι»**, αντίστοιχα. Το κλειστό μονοπάτι λέγεται και **«κύκλος»**. **«Βρόγχος»** λέγεται μια ακμή που ξεκινά και προσπίπτει στον ίδιο κόμβο. Γραφήματα με βρόγχους ονομάζονται **«πολλαπλά»**. Γραφήματα στα οποία δεν υπάρχουν κύκλοι ονομάζονται **«ακυκλικά»**, ή **«δέντρα»**. Στην περίπτωση κατευθυνόμενου γραφήματος, θα πρέπει σε όλα αυτά να ληφθεί υπ' όψιν ότι οι πλευρές πλέον είναι προσανατολισμένες.

• Γραφήματα που μπορούν να αναπαρασταθούν στο επίπεδο χωρίς οι πλευρές τους να τέμνονται ονομάζονται **«επίπεδα»**, ή **«εμφυτεύσιμα στο επίπεδο»**. Δεν είναι απα-

Σχήμα 1.2: K_5 

Σχήμα 1.3: Διμερές γράφημα

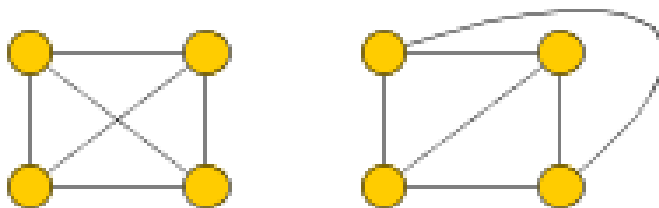
ραίτητο κάθε αναπαράστασή τους να μην έχει τεμνόμενες πλευρές, αρκεί να υπάρχει μια αναπαράσταση στην οποία να μην τέμνονται. Τέτοιο γράφημα είναι το K_4 . Στο ακόλουθο σχήμα φαίνονται δύο αναπαραστάσεις του. Στην πρώτη οι δύο εσωτερικές του τετραγώνου πλευρές τεμνονται, ενώ στην άλλη όχι.

Μη επίπεδα γραφήματα είναι, προφανώς, και όσα περιέχουν σαν υπογράφημά τους κάποιο μη επίπεδο γράφημα, όπως λχ τα $K_{3,3}$ και K_5 . Η συνθήκη αυτή είναι κανή και αναγκαία.

• Δύο γραφήματα $G=(V,E)$ και $G'=(V',E')$ ονομάζονται «**ισομορφικά**» εάν υπάρχει μια αμφιμονοσήμαντη απεικόνιση $f : V \rightarrow V'$ τέτοια ώστε μια ακμή ij να υπάρχει στο G αν και μόνο αν υπάρχει η ακμή $f(i)f(j)$ υπάρχει στο γράφημα G' . Δηλαδή η f είναι τέτοια ώστε να αντιστοιχίζει κορυφές του ενός με κορυφές του άλλου, διατηρώντας γειτνιάσεις και μη γειτνιάσεις. Για παράδειγμα, τα ακόλουθα δύο υπογραφήματα του γραφήματος του Σχήματος 1.1, έστω G_1 και G_2 , είναι ισομορφικά:

$$G_1 = \{\{1, 2, 3, 4, 5, 6, 7, 14\}, \{3-5, 4-5, 4-14, 14-1, 1-5, 2-6, 2-7\}\}$$

και

Σχήμα 1.4: K_4

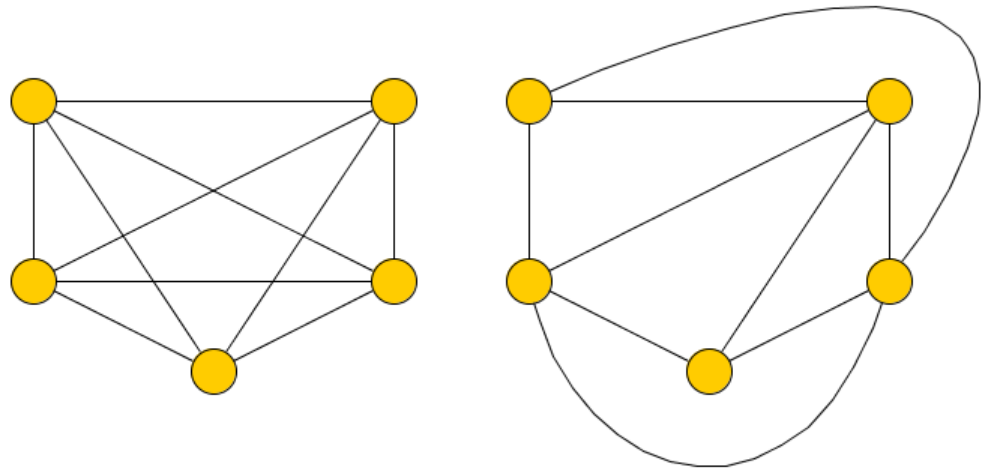
$$G_2 = \{\{11, 12, 14, 4, 1, 12, 9, 8\}, \{11 - 12, 12 - 14, 14 - 4, 4 - 1, 1 - 12, 9 - 15, 8 - 15\}\}.$$

•«**Δρομίσκος Euler**» ονομάζεται ένας δρομίσκος που διέρχεται από όλες τις πλευρές ενός γραφήματος. Γραφήματα στα οποία υπάρχει τέτοιος δρομίσκος ονομάζονται «**Γραφήματα Euler**». Γραφήματα που περιέχουν κύκλο ο οποίος περνά από κάθε κορυφή τους, ονομάζονται «**Γραφήματα Hamilton**».

•Θεωρούμε $\Lambda(v)$ τη λίστα με όλα τα γραφήματα με v κόμβους. Η αναφορά σε μία συγκεκριμένη ιδιότητα που κάποια από αυτά έχουν, πρακτικά είναι αναφορά στη λίστα αυτών που την έχουν. Δηλαδή, η ιδιότητα $A(v)$ (των γραφημάτων με v κόμβους) περιγράφεται σαν εκείνο το υποσύνολο του $\Lambda(v)$ που περιέχει εκείνα ακριβώς τα γραφήματα για τα οποία ισχύει η ιδιότητα. Για παράδειγμα, η ιδιότητα να έχουν όλοι οι κόμβοι δεσμούς με δύο ακριβώς άλλους κόμβους του γραφήματος μπορεί να περιγραφεί σαν $A(v) = \{i : [i \in G(v)] \& [deg_i = 2]\}$.

Μια ιδιότητα A καλείται «**μονότονη**» όταν για κάθε γράφημα G που την ικανοποιεί, κάθε υπογράφημά του ικανοποιεί και αυτό την ιδιότητα. Για παράδειγμα, έστω A η ιδιότητα ένα (συνεκτικό) γράφημα να είναι επίπεδο. Ικανή και αναγκαία συνθήκη για να συμβαίνει αυτό είναι να ικανοποιείται η ανισότητα του Euler: $N \leq 3v + 6$, όπου N το πλήθος των ακμών του γραφήματος και v των κορυφών του. Είναι προφανές ότι και κάθε υπογράφημα ενός γραφήματος που ικανοποιεί την ιδιότητα αυτή την ικανοποιεί και αυτό: Αν σε ένα γράφημα οι πλευρές του μπορούν να ζωγραφιστούν κατά τρόπο τέτοιο που να μην τέμνονται οι πλευρές του, τότε όσες πλευρές και αν αφαιρέσουμε, πάλι θα μπορούμε να το ζωγραφίσουμε έτσι ώστε να μην τέμνονται αυτές οι πλευρές που έμειναν. Όμοια, αν αφαιρεθούν κορυφές, τότε μαζί τους εξ'ορισμού αφαιρούνται και οι ακμές που προσπίπτουν σε αυτές, οπότε το συμπέρασμα παραμένει ίδιο. Αντίθετα, η ιδιότητα το γράφημα να μην είναι επίπεδο δεν είναι μονότονη, γιατί μπορεί μετά την αφαίρεση κάποιας ακμής το γράφημα να ικανοποιεί τη συνθήκη του Euler, και άρα, πλέον να είναι επίπεδο. Αυτό συμβαίνει για παράδειγμα με το K_5 : όσο είναι πλήρες δεν ικανοποιεί τη συνθήκη, αλλά με την αφαίρεση μιας ακμής την ικανοποιεί, πράγμα που μπορούμε να επιβεβαιώσουμε και γραφικά, μιας και το μέγεθός του είναι αρκετά μικρό, στο Σχήμα 1.5.

Μία ακόμη ιδιότητα που δεν είναι μονότονη είναι η ιδιότητα του να είναι ισομορφικό ένα γράφημα με κάποιο άλλο. Είναι προφανές ότι αν ένα γράφημα G_1 είναι ισομορφικό με ένα γράφημα G_2 δεν μπορεί να είναι ισομορφικό με κανένα (μη τετριμμένο) υπογράφημα του G_2 , έστω G_3 , αφού το G_3 θα έχει σίγουρα λιγότερες ακμές.



Σχήμα 1.5: K_5 και K_5 μείον μια ακμή

1.2 Πιθανότητες-Κατανομές

1.2.1 Η Πιθανότητα

«Πείραμα τύχης» ονομάζουμε ένα πείραμα το οποίο μπορεί να δώσει διαφορετικά αποτελέσματα ακόμα και αν εκτελείται, φαινομενικά, υπό τις ίδιες συνθήκες. Σε τέτοια πειράματα υπάρχει αδυναμία πρόβλεψης του αποτελέσματος και το σύνολο των δυνατών αποτελεσμάτων ενός τέτοιου πειράματος λέγεται δειγματικός χώρος και θα συμβολίζεται Ω . Παράδειγμα ενός πειράματος τύχης είναι η ρίψη ενός νομίσματος, με δειγματικό χώρο τον $\Omega = \{\text{«κορώνα»}, \text{«γράμματα»}\}$. Το αδύνατο ενδεχόμενο θα συμβολίζεται \emptyset και το βέβαιο Ω . Ένα αδύνατο ενδεχόμενο θα μπορούσε να είναι το $A = \{\text{«κορώνα» και «γράμματα» ταυτόχρονα}\}$, ενώ ένα βέβαιο είναι το $B = \{\text{να φέρω ή «κορώνα» ή «γράμματα»}\}$.

•Τι είναι σ-άλγεβρα;

Ορισμός 1.2.1. Μια οικογένεια \mathcal{F} υποσυνόλων ενός μη κενού συνόλου X είναι μια σ-άλγεβρα αν η \mathcal{F} έχει τις παρακάτω ιδιότητες:

- 1 $X \in \mathcal{F}$.
- 2 Αν $A \in \mathcal{F}$, τότε $A^c \in \mathcal{F}$, όπου A^c το συμπλήρωμα του A ως προς το X .
- 3 Αν $A_n \in \mathcal{F}$, $n \in \mathbb{N}$, τότε και $\bigcup_{n=1}^{+\infty} A_n \in \mathcal{F}$.

Παρατήρηση 1. Από τον ορισμό εύκολα προκύπτει ότι για κάθε αριθμήσιμο στο πλήθος συνδυασμό πράξεων μεταξύ στοιχείων της \mathcal{F} το αποτέλεσμα είναι στοιχείο της \mathcal{F} .

•Τι είναι μέτρο;

Ορισμός 1.2.2. Μια συνολοσυνάρτηση μ πάνω στη σ-άλγεβρα \mathcal{F} στο X είναι ένα μέτρο αν ικανοποιεί τις παρακάτω προϋποθέσεις:

1. Το $\mu(A) \in [0, +\infty]$ για $A \in \mathcal{F}$.
2. $\mu(\emptyset) = 0$.
3. Αν A_1, A_2, \dots είναι ακολουθία ξένων υποσυνόλων της \mathcal{F} και αν $\bigcup_{\kappa=1}^{+\infty} A_\kappa \in \mathcal{F}$, τότε $\mu(\bigcup_{\kappa=1}^{+\infty} A_\kappa) = \sum_{\kappa=1}^{+\infty} \mu(A_\kappa)$ (σ-αθροιστικότητα).

Το (X, \mathcal{F}) λέγεται μετρήσιμος χώρος και η τριάδα (X, \mathcal{F}, μ) λέγεται χώρος μέτρου και τα στοιχεία της \mathcal{F} λέγονται μετρήσιμα.

Αν, επιπλέον, $\mu(X) = 1$ τότε το μ λέγεται μέτρο πιθανότητας και ο χώρος (X, \mathcal{F}, μ) λέγεται χώρος μέτρου πιθανότητας. Στο υπόλοιπο της εργασίας αυτής θα χρησιμοποιείται το μέτρο πιθανότητας P , χωρίς να διευκρινίζεται. Σε κάθε κεφάλαιο θα διευκρινίζεται ο χ.π. που θα εννοείται.

Ορισμός 1.2.3. Έστω (X, \mathcal{F}) μετρήσιμος χώρος. Μια συνάρτηση $f : X \rightarrow [-\infty, +\infty]$ λέγεται μετρήσιμη ή \mathcal{F} -μετρήσιμη αν $\forall \beta \in \mathbb{R}$ το $[f \leq \beta] = f^{-1}([-\infty, \beta])$ είναι μετρήσιμο, δηλαδή, ανήκει στην \mathcal{F} .

Από αυτό, η συνάρτηση $P : \Omega \rightarrow \mathbb{R}$ θα είναι μετρήσιμη αν ισχύει το παραπάνω. Ισοδύναμα, η συνάρτηση πιθανότητας P θα αντιστοιχεί έναν αριθμό του $[0, 1]$ μόνο σε μετρήσιμα υποσύνολα του Ω , που θα ονομάζουμε ενδεχόμενα.

Θεωρώντας ότι όλα τα ενδεχόμενα ενός συνόλου Ω είναι το ίδιο «δυνατά», η «πιθανότητα» του ενδεχομένου E είναι ίση με:

$$P(E) = \frac{N(E)}{N(\Omega)}, E \subset \Omega,$$

όπου $N(E)$ και $N(\Omega)$ το πλήθος των στοιχείων του E και του Ω αντίστοιχα.

Οι τιμές 0 και 1 λαμβάνονται για το αδύνατο και το βέβαιο ενδεχόμενο αντίστοιχα.

Για παράδειγμα, έστω ο δειγματικός χώρος $\Omega = \{1, 2, 3, 4, 5, 6\}$, που θα μπορούσε να είναι όλες οι πιθανές τιμές που μπορεί να φέρει ένα ζάρι κατά τη ρίψη του, και τα ενδεχόμενα $A = \{1, 3, 5\}$, $B = \{2, 4, 6\}$. Η πιθανότητα να φέρει το ζάρι περιττό αριθμό κατά τη ρίψη (ενδεχόμενο A) είναι $P(A) = \frac{3}{6} = 0.5$ και το ίδιο και η πιθανότητα να φέρει άρτιο αριθμό (ενδεχόμενο B).

Κάποιες από τις **ιδιότητες της πιθανότητας** είναι:

1. $P(E) \geq 0, \forall E \subset \Omega$
2. $P(E_1 \cup E_2) = P(E_1) + P(E_2), \gamma\iota\alpha E_1 \cap E_2 = \emptyset$
3. $P(\Omega) = 1$

Ως «**συμπληρωματικό**» ενδεχόμενο E' ενός ενδεχομένου E ορίζεται το ενδεχόμενο εκείνο που πραγματοποιείται όταν δεν πραγματοποιείται το E . Επειδή

$$N(E \cup E') = N(E) + N(E') = N(\Omega),$$

ισχύει πώς

$$P(E) = 1 - P(E')$$

Έστω (Ω, \mathcal{F}, P) ένας χώρος πιθανότητας ενός πειράματος και $A \in \mathcal{F}$, δηλαδή το A είναι ένα ενδεχόμενο του πειράματος τύχης στο οποίο αντιστοιχεί αυτός ο χώρος πιθανότητας. Έστω ότι $A \neq \Omega$. Έστω ότι κατά την εκτέλεση του πειράματος πραγματοποιείται το ενδεχόμενο A . Τότε κάθε ενδεχόμενο ασυμβίβαστο με το A θα πρέπει να έχει μηδενική πιθανότητα να συμβεί, αφού το A έχει ήδη πραγματοποιηθεί. Πώς, λοιπόν, τροποποιούνται οι πιθανότητες των ενδεχομένων εάν γνωρίζουμε ότι ένα ενδεχόμενο έχει ήδη πραγματοποιηθεί;

Οι τροποποιημένες αυτές πιθανότητες ονομάζονται **δεσμευμένες**. Ακολουθώντας τον κλασικό ορισμό της πιθανότητας, η δεσμευμένη πιθανότητα ενός ενδεχομένου B , όταν το ενδεχόμενο A έχει ήδη πραγματοποιηθεί είναι

$$P(B|A) = \frac{N(AB)}{N(A)}.$$

Επειδή:

$$\frac{N(AB)}{N(A)} = \frac{\frac{N(AB)}{N(\Omega)}}{\frac{N(A)}{N(\Omega)}} = \frac{P(AB)}{P(A)}$$

τελικά έχουμε ότι:

$$P(B|A) = \frac{P(AB)}{P(A)}$$

Και τα τρία αξιώματα του ορισμού της Πιθανότητας κατά Kolmogorov για την πιθανότητα ικανοποιούνται, αφού:

1. $P(B|A) \geq 0, \forall A \in \mathcal{F}$, αφού $P(AB) \geq 0$ και $P(A) \geq 0 \forall A, B \in \mathcal{F}$,
2. Αν $B_i, A \in \mathcal{F}$, για $i = 1, 2, \dots$, και $B_i \cap B_j = \emptyset$ για $i \neq j$ τότε να ισχύει

$$P\left(\bigcup_{i=1}^{\infty} B_i | A\right) = \sum_{i=1}^{\infty} P(B_i | A)$$

αφού αν τα ενδεχόμενα B_i, B_j είναι ανεξάρτητα, τότε ανεξάρτητα θα συμβούν και με δεδομένη την πραγματοποίηση του ενδεχομένου A

3. $P(\Omega) = 1$

Συνεπώς, όλοι οι νόμοι και οι ιδιότητες των πιθανοτήτων ισχύουν και για δεσμευμένες πιθανότητες.

Τέλος, θα μπορούσαμε να πούμε ότι όλες οι πιθανότητες μπορούν να θεωρηθούν δεσμευμένες, θεωρώντας ότι έχει ήδη πραγματοποιηθεί το βέβαιο ενδεχόμενο: $P(A) = P(A|\Omega)$

Δύο ενδεχόμενα καλούνται «*ανεξάρτητα*» όταν $P(AB) = P(A)P(B)$. Αυτό σημαίνει ότι η πραγματοποίηση του ενός είναι ανεξάρτητη από την πραγματοποίηση του άλλου και δεν παρέχει καμία πληροφορία σχετικά με αυτήν. Η σχέση ανεξαρτησίας είναι συμμετρική: Αν το ενδεχόμενο A είναι ανεξάρτητο από το B , τότε και το B είναι ανεξάρτητο από το A . Έτσι, έχουμε $P(A|B) = P(A)$ και $P(B|A) = P(B)$. Από αυτά και τον ορισμό της δεσμευμένης πιθανότητας προκύπτει η σχέση $P(AB) = P(A)P(B)$.

1.3 Τυχαίες Μεταβλητές

Όπως αναφέρθηκε προηγουμένως, δεν μπορούμε ποτέ να είμαστε σίγουροι για το αποτέλεσμα ενός πειράματος. Έτσι, θα μπορούσαμε να πούμε ότι δεν έχει νόημα να ψάχνουμε ποιο ακριβώς ενδεχόμενο θα είναι το αποτέλεσμα του πειράματος, όσο έχει νόημα να μας ενδιαφέρουν κάποια χαρακτηριστικά του αποτελέσματος και πόσο πιθανό είναι αυτά να συμβούν.

Για παράδειγμα, έστω ότι έχουμε ένα δείγμα από φοιτητές του Πολυτεχνείου και μας ενδιαφέρει να δούμε αν καπνίζουν ή όχι. Ο δειγματικός μας χώρος είναι το σύνολο $\Omega = \{\{K, M\}^n\}^4$, όπου με K συμβολίζουμε την περίπτωση ο φοιτητής που επιλέχθηκε να απάντησε ότι καπνίζει ενώ με M την περίπτωση να απάντησε πως δεν καπνίζει. Αυτό που στην πραγματικότητα μας ενδιαφέρει είναι να ξέρουμε πόσοι είναι αυτοί που καπνίζουν, χωρίς να μας ενδιαφέρει η σειρά με την οποία βρέθηκαν αυτοί στο δείγμα. Έτσι, θεωρούμε μία αντιστοίχιση X η οποία για κάθε πιθανή n -άδα αποτελεσμάτων θα μας επιστρέψει το πλήθος αυτών που καπνίζουν. Είναι προφανές ότι για τη X ισχύει ότι $X : \Omega \rightarrow \mathbb{R}$. Παρότι η τιμή της είναι «τυχαία», αφού μόνο μετά την εκτέλεση του πειράματος ξέρουμε το αποτέλεσμα, μας ενδιαφέρει να μάθουμε την

πιθανότητα με την οποία η X θα λάβει την τιμή k , με $k \in \{0, 1, 2, \dots, n\}$.

Έστω (Ω, \mathcal{F}, P) ένας χώρος πιθανότητας και $X : \Omega \rightarrow \mathbb{R}$ μια συνάρτηση. Για κάθε $B \subseteq \mathbb{R}$ ορίζουμε $X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}$. Το σύνολο αυτό είναι το υποσύνολο του Ω που περιέχει τα ω εκείνα για τα οποία $X(\omega) \in B$. Υποσύνολα Borel του \mathbb{R} καλούνται τα σύνολα που ανήκουν στη σ -άλγεβρα Borel. Η σ -άλγεβρα Borel ενός συνόλου A είναι η τομή όλων των σ -άλγεβρών που περιέχουν, με την έννοια του εκλεισμού, την κλάση A : είναι η μικρότερη σ -άλγεβρα υποσυνόλων του A που περιέχει όλα τα υποσύνολα του A . Η X λέγεται «*τυχαία μεταβλητή*» εάν για κάθε υποσύνολο Borel του \mathbb{R} , το σύνολο $X^{-1}(\omega) \in \mathcal{F}$. Σύμφωνα με αυτά, αν το B είναι ένα υποσύνολο Borel του \mathbb{R} , το σύνολο $X^{-1}(B)$ είναι ενδεχόμενο και, άρα, δικαιούται να έχει πιθανότητα. Ορίζουμε

$$P_X(B) := P(X \in B) = P(X^{-1}(B))$$

Η συνολοσυνάρτηση $P_X(\cdot)$ έχει ως πεδίο ορισμού τη σ -άλγεβρα Borel του \mathbb{R} και είναι ένα μέτρο πιθανότητας πάνω στο χώρο $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, όπου με $\mathcal{B}(\mathbb{R})$ συμβολίζουμε τη σ -άλγεβρα Borel του \mathbb{R} . Η συνολοσυνάρτηση αυτή καλείται «*κατανομή*» της τυχαίας μεταβλητής X και καθορίζεται από την συνάρτηση κατανομής της: $F_X(x) := P(X \leq x) := P(X \in (-\infty, x])$, με $x \in \mathbb{R}$.

Άμεσες *συνέπειες του ορισμού* της πιθανότητας είναι οι ακόλουθες ιδιότητες για τις συναρτήσεις κατανομής πιθανότητας των τυχαίων μεταβλητών:

1. $0 \leq F(x) \leq 1, \forall x \in \mathbb{R}$
2. Είναι αύξουσα, δηλαδή $F(x_1) \leq F(x_2)$, για $x_1 \leq x_2$.
3. Η F είναι δεξιά συνεχής, δηλαδή $\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$
4. $\lim_{x \rightarrow -\infty} F(x) = F(-\infty) = 0$ και $\lim_{x \rightarrow \infty} F(x) = F(\infty) = 1$

Αυτές οι 4 σχέσεις είναι ικανές και αναγκαίες συνθήκες για να είναι μια συνάρτηση, συνάρτηση κατανομής πιθανότητας.

Αν το σύνολο τιμών (*σ.τ.*) μιας τυχαίας μεταβλητής (*τ.μ.*) X , είναι πεπερασμένο ή αριθμήσιμο τότε λέγεται ότι η $\tau.μ.$ είναι διακριτή, και η κατανομή της το ίδιο. Η κατανομή της διακριτής $\tau.μ.$ X καθορίζεται από τη συνάρτηση μάζας πιθανότητας:

$$f_X(x) := P(X = x), \text{ για } x \in \mathbb{R}.$$

Οι συναρτήσεις μάζας και κατανομής συνδέονται μέσω των σχέσεων:

$$F_X(x) = \sum_{(x_i \leq x)} f_X(x_i) \text{ και } f_X(x) = F_X(x) - \lim_{x \rightarrow k^-} F_X(k)$$

Για τη συνάρτηση μάζας πιθανότητας ισχύει ακόμα ότι

$$f_X(x) \geq 0, \forall x \in \mathbb{R} \text{ και } \sum_{x \in \sigma.τ} f_X(x) = 1$$

$${}^0\Omega = \{\{K, M\}^n\} = \{(x_1, x_2, \dots, x_n) : x_i = K, M, \text{ με } i = 1..n\}$$

Αν το σύνολο τιμών (σ.τ.) μιας τυχαίας μεταβλητής (τ.μ.) X , δεν είναι πεπερασμένο ή αριθμήσιμο τότε λέγεται ότι η τ.μ. είναι *συνεχής*, και η κατανομή της το ίδιο. Η κατανομή της συνεχούς τ.μ. X καθορίζεται από τη *συνάρτηση πυκνότητας πιθανότητας*: $f_X(x)$. Οι συναρτήσεις μάζας και κατανομής συνδέονται μέσω της:

$$F_X(x) = \int_{-\infty}^x f_X(y) dy \text{ για } x \in \mathbb{R}.$$

Από αυτό προκύπτει ότι η συνάρτηση κατανομής μιας συνεχούς τ.μ. είναι συνεχής και όπου η f_X είναι θετική η F είναι γνησίως αύξουσα. Για την συνάρτηση πυκνότητας πιθανότητας ισχύει ακόμα ότι:

$$1. f_X(x) \geq 0, \forall x \in \mathbb{R}$$

και

$$2. \int_{-\infty}^{+\infty} f_X(x) dx = 1.$$

Η f_X δεν είναι μοναδική: μπορεί να υπάρχουν διαφορετικές συναρτήσεις f που να διαφέρουν σε πεπερασμένο αριθμό σημείων και άρα το ολοκλήρωμα τα έχει ίδια τιμή.

Έστω χώρος πιθανότητας (Ω, \mathcal{F}, P) και μία τ.μ. $X : \Omega \rightarrow \mathbb{R}$. *Μέση τιμή* της τ.μ. X ως προς μέτρο P ονομάζεται το ολοκλήρωμα Lebesgue:

$$E[X] = \int_{\Omega} x dP.$$

Λέμε ότι η μέση τιμή της μεταβλητής X υπάρχει αν και μόνο αν το ολοκλήρωμα υπάρχει. Η ποσότητα:

$$Var(X) = E[(X - E[X])^2]$$

λέγεται διασπορά της X .

1.3.1 Κατανομές

Κατανομή Bernoulli

Η κατανομή Bernoulli είναι μια διακριτή κατανομή. Έχει Δειγματικό χώρο τον $\Omega = \{0,1\}$ στον οποίο το 1, ή η επιτυχία, συμβαίνει με πιθανότητα p , ενώ η αποτυχία με πιθανότητα $1-p$. Επειδή το p είναι πιθανότητα, θα είναι $0 < p < 1$.

Η συνάρτηση μάζας πιθανότητας είναι

$$P(X=x) = \begin{cases} p, & \text{για επιτυχία} \\ 1-p, & \text{για αποτυχία} \end{cases}$$

Η κατανομή αυτή θα μπορούσε να περιγράψει πειράματα όπως η ρίψη ενός κέρματος. Επιτυχία μπορεί να θεωρηθεί το να έρθει «κορώνα» και αποτυχία το να έρθει «γράμματα». Η μέση τιμή μιας τυχαίας μεταβλητής που ακολουθεί την Bernoulli είναι p και η διασπορά της $p(1-p)$.

Διωνυμική Κατανομή (Binomial Distribution)

Η διωνυμική κατανομή είναι και αυτή διακριτή. Μέσω αυτής περιγράφεται η πιθανότητα να έχουμε ακριβώς κ επιτυχίες σε ένα σύνολο με ν ανεξάρτητες δοκιμές Bernoulli με πιθανότητα επιτυχίας ίση με p . Η πιθανότητα που ψάχνουμε είναι :

$$P(X = \kappa | \nu, p) = \binom{\nu}{\kappa} p^{\kappa} (1-p)^{\nu-\kappa}$$

και η συνάρτηση κατανομής πιθανότητας είναι

$$F(x) = P(X \leq x) = \sum_{\kappa \leq x} \binom{\nu}{\kappa} p^{\kappa} (1-p)^{\nu-\kappa}$$

Η μέση τιμή μιας τυχαίας μεταβλητής που ακολουθεί την Διωνυμική είναι νp και η διασπορά της $\nu p(1-p)$.

Κατανομή Poisson

Η πιθανότητα για μία τ.μ. X που ακολουθεί αυτή την κατανομή με παράμετρο λ να λάβει την τιμή x είναι

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

Η μέση τιμή και η διασπορά παίρνουν και οι δύο την τιμή της παραμέτρου λ .

Η κατανομή αυτή αφορά στον αριθμό συμβάντων σε ορισμένο χρονικό ή χωρικό διάστημα. Η Poisson είναι άμεσα συνδεδεμένη με τη Διωνυμική: καθώς ο αριθμός ν των δοκιμών Bernoulli απειρίζεται και ταυτόχρονα η πιθανότητα επιτυχίας p κάθε δοκιμής τείνει στο μηδέν έτσι ώστε $\nu p \rightarrow \lambda$ έχουμε:

$$\lim_{n \rightarrow \infty} P_{\text{Binomial}}(X = \kappa | \nu, p) = \lim_{n \rightarrow \infty} \binom{\nu}{\kappa} p^{\kappa} (1-p)^{\nu-\kappa} = \lim_{n \rightarrow \infty} \frac{\nu!}{\kappa! (\nu-\kappa)!}$$

$$p^{\kappa} (1-p)^{\nu-\kappa} = \lim_{n \rightarrow \infty} \frac{(1*2*\dots*(\nu-\kappa)(\nu-\kappa+1)\dots(\nu-1)\nu)}{(\kappa! * 1*2*3*\dots(\nu-\kappa))} p^{\kappa} (1-p)^{\nu-\kappa} =$$

$$\lim_{n \rightarrow \infty} \frac{((\nu-\kappa+1)\dots(\nu-1)\nu)}{\kappa!} p^{\kappa} (1-p)^{\nu-\kappa} \stackrel{(*)}{=} \lim_{n \rightarrow \infty} \frac{((\nu-\kappa+1)\dots(\nu-1)\nu)}{\kappa!} \left(\frac{\lambda}{\nu}\right)^{\kappa} \left(1 - \frac{\lambda}{\nu}\right)^{\nu-\kappa} =$$

$$\lim_{n \rightarrow \infty} \frac{((\nu-\kappa+1)\dots(\nu-1)\nu)}{\nu^{\kappa}} \frac{\lambda^{\kappa}}{\kappa!} \left(1 - \frac{\lambda}{\nu}\right)^{\nu} \left(1 - \frac{\lambda}{\nu}\right)^{-\kappa} = 1 * \frac{\lambda^{\kappa}}{\kappa!} * e^{-\nu} * 1 = \frac{\lambda^{\kappa}}{\kappa!} e^{-\nu}$$

που είναι η κατανομή Poisson.

(*): Όπου p βάζουμε τον αναμενόμενο αριθμό επιτυχιών προς το συνολικό πλήθος δοκιμών, δηλαδή κάνουμε την αντικατάσταση $\lambda = \nu p = E(X) \Rightarrow p = \lambda/\nu$.

1.4 Χρήσιμες Σχέσεις

Παρατίθενται κάποιες σχέσεις που χρησιμοποιούνται συχνά παρακάτω. Οι σχέσεις αυτές δεν θα αποδειχθούν.

- Ισχύει ότι $\lim_{\beta \rightarrow \infty} \left(1 + \frac{\alpha}{\beta}\right)^\beta = e^\alpha, \alpha, \beta \in \mathbb{R}$.

- Έστω $f : G \rightarrow \mathbb{R}$ αναλυτική συνάρτηση στο ανοιχτό σύνολο G και έστω $z_0 \in \mathbb{R}$. Αν δ είναι η απόσταση του x_0 από το σύνορο ∂G του G , τότε

$$f(x) = \sum_{\nu=0}^{+\infty} \frac{f^{(\nu)}(x_0)}{\nu!} (x - x_0)^\nu, \forall x \in D(x_0, \delta)$$

- Η εκθετική συνάρτηση είναι αναλυτική και άρα προσεγγίζεται ως εξής με σειρά Taylor, με κέντρο το 0:

$$e^x = \sum_{\kappa=0}^{+\infty} \frac{f^{(\kappa)}(0)}{\kappa!} (x - 0)^\kappa = \sum_{\kappa=0}^{+\infty} \frac{1}{\kappa!} x^\kappa = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

Ορισμός 1.4.1. Έστω (Ω, \mathcal{F}, P) χώρος πιθανότητας. Λέμε ότι η ιδιότητα E ισχύει «σχεδόν παντού» σε ένα A , ή «σχεδόν βεβαίως» για τα στοιχεία του A που είναι μετρήσιμο υποσύνολο του Ω αν υπάρχει $N \subset A$ μετρήσιμο με $P(N) = 0$, και η E ισχύει σε κάθε σημείο του $A \setminus N$.

- Θεωρούμε τα σύμβολα « O » και « o ». Για τις συναρτήσεις $\alpha(\nu)$ και $\beta(\nu)$, με $\beta(\nu) > 0$ για $\nu=1,2,\dots$, λέμε ότι:

1. $\alpha(\nu) = o(\beta(\nu))$, όταν

$$\lim_{\nu \rightarrow \infty} \frac{|\alpha(\nu)|}{\beta(\nu)} = 0.$$

Για παράδειγμα, έστω οι συναρτήσεις $\alpha(\nu) = 5\nu + 2$ και $\beta(\nu) = 8\nu^2 + 15$. Τότε είναι $\alpha(\nu) = o(\beta(\nu))$.

2. $\alpha(\nu) = O(\beta(\nu))$, όταν

$$\lim_{\nu \rightarrow \infty} \frac{|\alpha(\nu)|}{\beta(\nu)} = \kappa,$$

με κ κάποιον πραγματικό, φραγμένο, αριθμό. Τέτοιες είναι οι $\alpha(\nu)$ του (1) και η $\gamma(\nu) = 7\nu - 9$.

3. $\alpha(\nu) = O_\varepsilon(\beta(\nu))$, όταν

$$\lim_{\nu \rightarrow \infty} \frac{|\alpha(\nu)|}{\beta(\nu)} \leq \kappa(\varepsilon),$$

με κ έναν θετικό αριθμό που εξαρτάται από το ε .

4. $\alpha(\nu) \sim \beta(\nu)$, όταν

$$\lim_{\nu \rightarrow \infty} \frac{|\alpha(\nu)|}{\beta(\nu)} = 1.$$

Τέτοιες είναι η $\beta(\nu)$ του (1) και η $\delta(\nu) = 8\nu^2 - 1000$

Κεφάλαιο 2

Random Graphs: The Poisson Model

2.1 Γενικά

Η ραγδαία ανάπτυξη της τεχνολογίας οδηγεί εκ των πραγμάτων στη δημιουργία καινούριων δομών και αντικειμένων που χρειάζονται μελέτη. Για να περιγραφούν αυτά, ως κομμάτι της πραγματικότητας, απαιτείται η γνώση μεγάλου πλήθους παραμέτρων. Το γεγονός αυτό κάνει την ακριβή περιγραφή της πραγματικότητας ένα άπιαστο όνειρο. Λύση στο πρόβλημα αυτό είναι η χρήση μοντέλων. Εξ' άλλου, "All models are wrong, but some of them are usefull", όπως είχε πει και ο Άγγλος μαθηματικός G.Box.

Στην εργασία αυτή θα παρουσιαστούν μοντέλα που επιχειρούν αν περιγράψουν δίκτυα. Τα μοντέλα που θα παρουσιαστούν είναι τυχαία κ συνήθως είναι χρήσιμο και βολικό να μελετώνται οι οριακές τους συμπεριφορές.

Η τυχαιότητα παίζει πολύ σημαντικό ρόλο στη ζωή μας. Πολλά από αυτά που παρατηρούμε γύρω μας, ενώ μπορεί να έχουν δημιουργηθεί-κατασκευαστεί από κάποιον καλά καθορισμένο κανόνα, δεν μπορούμε να βρούμε ποιος είναι αυτός. Έτσι, πολλές φορές, καταφεύγουμε στη χρήση κάποιων μοντέλων μέσω των οποίων, εν τέλει, προσπαθούμε να περιγράψουμε σε όσο το δυνατόν ικανοποιητικότερο βαθμό αυτό που παρατηρήσαμε.

Όπως έχει αναφερθεί, ένα δίκτυο μπορεί να αναπαρασταθεί από ένα γράφο. Άρα, η διαδικασία δημιουργίας δικτύων μπορεί να αναχθεί στη διαδικασία δημιουργίας γράφων. Τυχαίος γράφος είναι ένας γράφος που έχει δημιουργηθεί «*με τυχαίο τρόπο*».

Υπάρχουν διάφορα μοντέλα μέσω των οποίων μπορούμε να μοντελοποιήσουμε τα τυχαία δίκτυα, δηλαδή, υπάρχουν διάφοροι «κανόνες» που μπορούμε να ακολουθήσουμε για να κατασκευάσουμε τελικά έναν τυχαίο γράφο.

Μπορούμε να σκεφτούμε πολλούς τρόπους να διαχωρίσουμε ένα δίκτυο από ένα άλλο. Ένας από αυτούς τα διαχωρίζει σε αυτά που όσο περνάει ο χρόνος μένουν ίδια, που είναι δηλαδή *στατικά*, και αυτά που με την πάροδο του χρόνου εξελίσσονται, τα *δυναμικά*. Στην πρώτη κατηγορία, ο γράφος μένει ο ίδιος με την πάροδο του χρόνου: το πλήθος των κόμβων παραμένει σταθερό από μετά τη στιγμή της δημιουργίας του και οι ακμές που εμφανίζονται μέσω κάποιας (στοχαστικής) διαδικασίας μένουν ως έχουν. Στην δεύτερη κατηγορία ανήκουν γραφήματα που όσο περνάει ο χρόνος αλλάζουν μορφή: θεωρούμε ότι σε κάθε χρονική στιγμή ένας νέος κόμβος προστίθεται στο γράφημα. Μαζί με την «είσοδο» του καινούριου κόμβου στο γράφημα έχουμε και τη

δημιουργία ακμών που ενώνουν τον καινούριο αυτό κόμβο με τους υπόλοιπους του ήδη υπάρχοντος γραφήματος. Όπως προκύπτει από το διαχωρισμό αυτό, τα πρώτα γραφήματα είναι σταθερού μεγέθους, ενώ τα δεύτερα όσο περνάει ο καιρός γιγαντώνονται.

Τα τυχαία δίκτυα θεωρούνται πολλές φορές σημεία αναφοράς: Η σύγκριση ενός δικτύου με ένα τυχαίο δίκτυο της ίδιας τάξης μεγέθους μπορεί να μας βοηθήσει να δούμε αν το δίκτυό μας είναι τυχαίο ή αν έχει δομηθεί με έναν ίσως όχι και τόσο τυχαίο τρόπο, και ενδεχομένως να αναζητήσουμε το «μοτίβο» που το δημιουργεί. Μπορούμε να βρούμε τις ομοιότητες ή τις διαφορές ενός γράφου που έχουμε στα χέρια μας, με έναν γράφο της ίδιας τάξης, που όμως έχει δημιουργηθεί με τρόπο τυχαίο. Το είδος των στοιχείων που ενδιαφέρει να δούμε αν είναι ίδια ή παρόμοια ανάμεσα σε ένα γράφο και έναν τυχαίο γράφο της ίδιας τάξης αφορούν κυρίως στη δομή, το πόσες συνιστώσες έχει και αν έχουν περίπου το ίδιο μέγεθος, συνεκτικές συνιστώσες που υπάρχουν, αν υπάρχουν περισσότερες από μια, η κατανομή των βαθμών στους κόμβους του γραφήματος και το μέγεθος των αποστάσεων, όπως λ.χ. της διαμέτρου του. Για να γίνει μια τέτοια σύγκριση, τα εργαλεία που χρησιμοποιούνται συνήθως είναι διάφορες ιδιότητες που μπορεί να έχει το γράφημα και κάποια «φράγματα», ή «thresholds», που θα αναλυθούν στη συνέχεια.

Όπως γενικότερα συμβαίνει σε πολλές αναλύσεις στην επιστήμη των Μαθηματικών, έτσι και στο κομμάτι της Θεωρίας Γραφημάτων με το οποίο ασχολείται αυτή η εργασία, μελετάται η οριακή συμπεριφορά των γράφων. Δηλαδή η συμπεριφορά εκείνη που έχει ένα γράφημα μεγάλων διαστάσεων, ή ισοδύναμα, ένα γράφημα για το οποίο $n \rightarrow +\infty$. Αυτό δεν συμβαίνει απαραίτητα γιατί θεωρούμε ότι θα έχουμε στα χέρια μας γράφους τόσο μεγάλων διαστάσεων, αλλά γιατί σε τέτοιες οριακές συνθήκες τα αποτελέσματα είναι μακράν πιο «εύχρηστα» και «κομψά» σε σχέση με αποτελέσματα που θα προκύψουν από αναλύσεις αντίστοιχες αλλά σε μικρότερη κλίμακα, οι οποίες μπορούμε να θεωρήσουμε με σχεδόν αμελητέο σφάλμα ότι είναι ειδικές περιπτώσεις των αναλύσεων στο όριο. Έτσι, η οριακή μελέτη ιδιοτήτων και συμπεριφορών μπορεί να θεωρηθεί ένας τρόπος να μελετήσουμε πολλές διαφορετικές υποπεριπτώσεις του ίδιου προβλήματος ταυτόχρονα, με αποτελέσματα που είναι ικανοποιητικά κοντά στο αποτέλεσμα που θα έδινε η μελέτη κάθε ξεχωριστού υποπροβλήματος. Το κέρδος που γεννιέται από την επιλογή αυτή αποτελεί όχι μόνο το γεγονός πως τη μελέτη πολλών υποπεριπτώσεων την ανάγουμε στη μελέτη μιας γενικότερης περίπτωσης, αλλά και το γεγονός ότι οι χειρισμοί που απαιτείται να γίνουν κατά τη διάρκεια της μελέτης αυτής είναι συνήθως πολύ ευκολότεροι συγκριτικά με αυτούς που θα απαιτούνταν σε διαφορετική περίπτωση.

Τα τυχαία δίκτυα θεωρούνται από πολλούς το πιο θεμελιώδες μοντέλο δικτύων γιατί το μόνο που προϋποθέτει είναι μια τυχαία διαδικασία μέσα από την οποία αυτά γεννιούνται. Το περιεχόμενο της εργασίας αυτής αφορά, μεταξύ άλλων, και στην περιγραφή και την ανάλυση κάποιων ιδιοτήτων δύο εκ των βασικότερων μοντέλων που υπάρχουν για τα τυχαία δίκτυα. Γραφήματα που έχουν προκύψει από τις διαδικασίες που θα περιγραφούν αμέσως μετά ονομάζονται «τυχαία γραφήματα Poisson» και στη συνέχεια αναλύεται το «Preferential Attachment» μοντέλο.

2.2 Ορισμός Τυχαίων Γράφων

Το μοντέλο για τους «Poisson τυχαίους γράφους» θεωρείται από τα πιο απλά μοντέλα τυχαίων γράφων. Αλλιώς ονομάζεται και Μοντέλο των Erdős και Renyi, και ένας γράφος που ανήκει στην κατηγορία αυτή θα αναφέρεται στη συνέχεια ως «ER γράφος». Ένας τέτοιος μπορεί να οριστεί με πάνω από έναν τρόπους:

Λαδικασία 1. Έστω το κενό γράφημα με n κόμβους, $E(n)$. Σε αυτό δημιουργούμε ακμές ανάμεσα στους κόμβους. Η δημιουργία, ή μη, καθεμιάς από τις $\binom{n}{2}$ ακμές συμβαίνει με πιθανότητα $p > 0$, ή $1 - p$ αντίστοιχα, κάθε φορά και είναι ανεξάρτητη από την ύπαρξη ή την μη ύπαρξη άλλων ακμών.

Ένας τέτοιος γράφος συχνά αναφέρεται και ως $ER_n(p)$ και έστω ότι έχει N ακμές. Από την παραπάνω διαδικασία φαίνεται ότι η δημιουργία κάθε ακμής είναι μια Bernoulli τυχαία μεταβλητή, με πιθανότητα «επιτυχίας», ή δημιουργίας της ακμής, ίση με p , και πιθανότητα «αποτυχίας», ή μη δημιουργίας της ακμής, ίση με $1 - p$. Οι τ.μ. αυτές είναι ανεξάρτητες, αφού η δημιουργία ή μη μιας ακμής συμβαίνει ανεξάρτητα από την ύπαρξη ή μη των άλλων ακμών, και ισόνομες, αφού όλες ακολουθούν την Bernoulli με πιθανότητα επιτυχίας p . Το συνολικό πλήθος των ακμών στο γράφημα που προκύπτει είναι το άθροισμα των πολλών αυτών Bernoulli τυχαίων μεταβλητών. Μιας και, όπως δείχθηκε σε προηγούμενη ενότητα, το άθροισμα ισόνομων και ανεξάρτητων Bernoulli τ.μ. ακολουθεί τη Διωνυμική κατανομή, το πλήθος N ακολουθεί τη Διωνυμική με $\binom{n}{2}$ «προσπάθειες» και p πιθανότητα επιτυχίας σε κάθε προσπάθεια. Οι γράφοι που συνήθως επιλέγεται να κατασκευαστούν είναι μεγάλου μεγέθους.

Έτσι, εν τέλει, η τ.μ. N προσεγγίζεται ικανοποιητικά και από την κατανομή Poisson με παράμετρο

$$\lambda = p \binom{n}{2}.$$

Με βάση αυτά, η αναμενόμενη τιμή του πλήθους N είναι ίση με $\lambda = p \binom{n}{2}$. Έτσι, αν θελήσουμε μέσω αυτής της μεθόδου να βρεθούμε με ένα γράφημα που θα έχει περίπου N ακμές, θα πρέπει να επιλέξουμε την πιθανότητα p ίση με

$$\frac{N}{\binom{n}{2}}.$$

Από τα παραπάνω προκύπτει επίσης ότι για δεδομένη p η πιθανότητα να κατασκευαστεί γράφημα με M ακμές και n κόμβους είναι

$$P(N = M) = e^{-p \binom{n}{2}} \frac{[p \binom{n}{2}]^M}{M!}.$$

Ένας διαφορετικός τρόπος να ορίσουμε έναν ER τυχαίο γράφο είναι:

Λαδικασία 2. Αρχικά διαλέγουμε μία από τις $\binom{n}{2}$ πιθανές ακμές, στη συνέχεια μία από τις $\binom{n}{2} - 1$ πιθανές ακμές που απομένουν και, συνεχίζοντας έτσι, λήγουμε τη διαδικασία όταν φτάσουμε στο σημείο να επιλέξουμε μία από τις $-(N - 1)$ εναπομείνουσες πιθανές ακμές.

Τέτοια γραφήματα συμβολίζονται με $ER_{\nu, N}$. Κατά τη διαδικασία αυτή, έχοντας n κόμβους θέλουμε με τυχαίο τρόπο να φτιάξουμε ένα γράφημα που να έχει N ακριβώς πλευρές. Δηλαδή το N από τυχαία μεταβλητή που ήταν ακολουθώντας την πρώτη διαδικασία κατασκευής τώρα θέλουμε να είναι σταθερά. Αφού όλες οι πιθανές ακμές είναι $\binom{\nu}{2}$ και θέλουμε στο γράφο μας να έχουμε N ακμές, και αφού η δημιουργία κάθε ακμής είναι ισοπίθανη με τη δημιουργία κάποιας άλλης, η επιλογή N από τις $\binom{\nu}{2}$ συνολικά πιθανές ακμές γίνεται με $\binom{\nu}{N}$ διαφορετικούς τρόπους. Έτσι, αφού όλα γραφήματα που παράγονται από αυτή τη διαδικασία είναι ισοπίθανο να παραχθούν, η πιθανότητα δημιουργίας ενός συγκεκριμένου γράφου από το σύνολο αυτών με n κόμβους και N ακμές είναι

$$\frac{1}{\binom{\nu}{N}}.$$

Ισοδύναμα, μπορούμε να πούμε ότι διαλέγουμε από τα συνολικά $\binom{\nu}{N}$ δυνατά γραφήματα με n κόμβους και N ακμές αυτό το ένα που θέλουμε.

Κατά τις διαδικασίες αυτές, θεωρήθηκε ότι οι κόμβοι είναι μεταξύ τους διακριτοί και, επομένως, από το σύνολο των γράφων που θα δημιουργηθούν με τους δύο αυτούς τρόπους σίγουρα θα υπάρχουν και κάποιοι που θα είναι μεταξύ τους ισομορφικοί.

Οι δύο τρόποι δημιουργίας γραφημάτων είναι ισοδύναμοι στις περισσότερες των περιπτώσεων, όπως μπορεί εύκολα να αποδειχθεί με οριακά Θεωρήματα¹, οπότε στο εξής θα χρησιμοποιούνται εναλλάξ χωρίς αυτό να βλάπτει τη γενικότητα των αποτελεσμάτων.

Από τις παραπάνω διαδικασίες φαίνεται πως στη φάση αυτή θα αναλυθούν γραφήματα σταθερού μεγέθους, δηλαδή γραφήματα που είναι καλά καθορισμένα και ως προς τους κόμβους και ως προς τις ακμές τους, από την αρχή κιόλας. Στη συνέχεια θα αναφερθούν κάποιες ιδιότητές τους και κάποια θεωρήματα που έχουν γραφεί για τέτοιου είδους γράφους.

¹Έστω η τ.μ. $I_{i,j}$ που δείχνει την ύπαρξη ακμής ανάμεσα στους κόμβους i, j του γραφηματος. Τότε, από τον Ασθενή Νόμο των Μεγάλων Αριθμών, έχουμε:

$$P\left(\left\|\sum I_{i,j} - \frac{\nu(\nu-1)}{2}p\right\| \geq \varepsilon \frac{\nu(\nu-1)}{2}\right) \rightarrow 0, \forall \varepsilon > 0, \text{ όταν } \nu \rightarrow +\infty.$$

Επομένως, το πλήθος των ακμών στο γράφημα είναι, από τη μια, τ.μ. αλλά, από την άλλη, είναι πολύ συγκεντρωμένη γύρω από τη μέση της τιμή.

2.3 Η Κατανομή των Βαθμών στο γράφημα

Όπως και τα μέτρα κεντρικότητας, έτσι και η κατανομή των βαθμών των κόμβων ενός γράφου είναι μέτρο που μπορεί να βοηθήσει στην καλύτερη αντίληψη ιδιοτήτων, και ενδεχομένως ιδιαιτεροτήτων, που μπορεί το γράφημά μας να έχει. Είναι, αμφότερα, χαρακτηριστικά που μπορούν να περιγράψουν, σε συνδυασμό και με άλλα, κατά κάποιο τρόπο έμμεσα ένα γράφημα. Σε μικρά γραφήματα αυτό μπορεί να μην είναι τόσο εμφανές, αφού συνήθως μπορούμε να καταλάβουμε τη δομή του γραφήματος και οπτικά, αλλά σε μεγάλα γραφήματα μια αναπαράσταση γραφική αποκτά χαοτικές διαστάσεις και έτσι η χρήση τέτοιων μέτρων για τη λεπτομερή εξέτασή τους καταλήγει να είναι απαραίτητη.

Έστω P το μέτρο πιθανότητας που περιγράφει το βαθμό των κόμβων του γραφήματος. Ανάλογα με το αν αναφερόμαστε στην περιγραφή κάποιων δεδομένων ή αν θέλουμε να περιγράψουμε την κατανομή αυτή για τυχαία δίκτυα, το μέτρο P είναι η καταγραφή των εμφανιζόμενων συχνοτήτων ή μια συνάρτηση κατανομής πιθανότητας για δίκτυα που δεν έχουν δημιουργηθεί ακόμη, αντίστοιχα.

Έστω, λοιπόν, ότι κάθε κόμβος του γραφήματος έχει κάποιο βαθμό d . Το αν πολλοί κόμβοι έχουν μικρό βαθμό ή όχι, το αν υπάρχουν κόμβοι με μεγάλους βαθμούς, πόσοι είναι αυτοί κοκ, μπορεί να επηρεάσει κάποιες από τις ιδιότητες που αναμένουμε το γράφημα να έχει. Ισοδύναμα, ο τρόπος με τον οποίο κατανέμονται οι βαθμοί ενός γραφήματος, μπορεί να καθορίσει κάποιες από τις ιδιότητες του γραφήματος και γενικά η γνώση της κατανομής των βαθμών μας δίνει πληροφορίες για τη δομή ενός γραφήματος.

Στο $ER_\nu(p)$ μοντέλο, έστω η τυχαία μεταβλητή D , που δηλώνει το βαθμό ενός τυχόντα κόμβου, δηλαδή το πλήθος των γειτόνων του κόμβου αυτού. Η D ακολουθεί τη Διωνυμική κατανομή, αφού κάθε ένας από τους υπόλοιπους κόμβους του γραφήματος ή θα είναι γείτονάς του, ή δεν θα είναι. Η πιθανότητα, λοιπόν, η D να πάρει την τιμή d είναι:

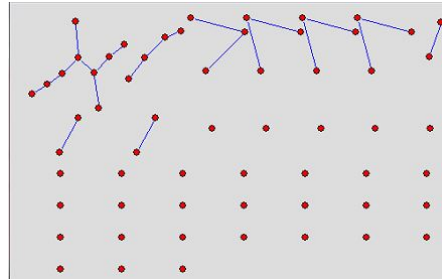
$$\begin{aligned} P(D = d) &= \binom{\nu - 1}{d} p^d (1 - p)^{(\nu - d - 1)} \\ &= \frac{(\nu - 1)!}{d!(\nu - d - 1)!} p^d (1 - p)^{(\nu - d - 1)} \end{aligned}$$

αφού με πιθανότητα p κάποιοι d κόμβοι θα είναι γείτονές του και με πιθανότητα $(1-p)$ οι υπόλοιποι $(\nu-d-1)$ δεν θα είναι γείτονες. Επειδή οι συνολικοί τρόποι να επιλεγθούν d από τους $(\nu-1)$ κόμβους του γραφήματος που απομένουν είναι $\binom{\nu-1}{d}$, προκύπτει η παραπάνω σχέση. Καθώς, όμως, τα γραφήματα για τα οποία μιλάμε έχουν μεγάλες διαστάσεις και, όπως έχει δειχθεί και σε προηγούμενη ενότητα, η Διωνυμική κατανομή για μεγάλο πλήθος «προσπαθειών» ν με πιθανότητα επιτυχίας p , περιγράφεται από την κατανομή Poisson με παράμετρο $\lambda = p\nu$, τελικά έχουμε ότι:

$$P(D = d) = \frac{(\nu - 1)^d}{d!} p^d e^{-(\nu-1)p}$$

αφού η αναμενόμενη τιμή της D είναι $E[D] = (\nu - 1)p$, μιας και στην πραγματικότητα ο κάθε κόμβος έχει $\nu - 1$ πιθανούς γείτονες, άρα $\nu - 1$ προσπάθειες.

Με χρήση του προγράμματος Pajek φτιάχνουμε τα επόμενα παραδείγματα τυχαίων γραφημάτων. Αρχικά φτιάχνουμε ένα γράφημα με 60 κόμβους και με πιθανότητα εμφάνισης ακμής ίση με $p=0.015$. Η γραφική αναπαράσταση ενός τέτοιου δικτύου είναι αυτή του Σχήματος (2.3). Στο γράφημα αυτό μπορούμε να

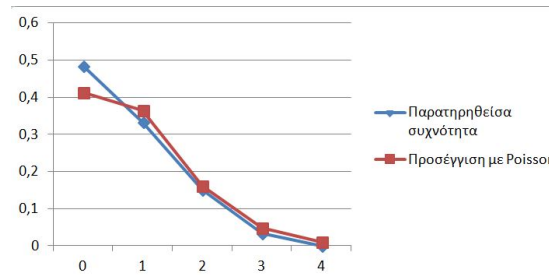


Σχήμα 2.1: ER Γράφημα με $p=0.015$, $n=60$.

αρκετά καλή.

παρατηρήσουμε πως υπάρχουν πολλοί μεμονωμένοι κόμβοι, και άρα πολλές μικρές ξεχωριστές συνιστώσες, και πως ίσα που ξεκινάει να υπάρχει μια κάπως μεγαλύτερη συνιστώσα στο γράφημα. Δεν υπάρχει κανένας κύκλος, και γενικότερα θα λέγαμε πως οι συνιστώσες έχουν δενδρική δομή.

Για τις συγκεκριμένες τιμές των p και n , στο τυχαίο γράφημα που προέκυψε, έχουμε 29 απομονωμένους κόμβους, 20 κόμβους με βαθμό ένα, 9 με βαθμό 2 και 2 με βαθμό 3. Κατασκευάζουμε το Γράφημα (2.3), που δείχνει ότι η προσέγγιση με την κατανομή Poisson είναι όντως αρ-

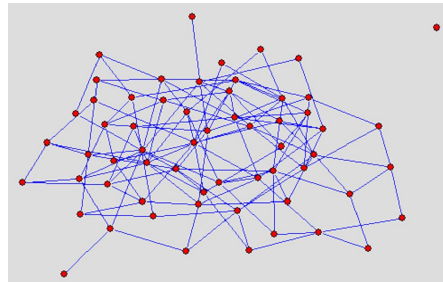


Σχήμα 2.2: Η κατανομή των βαθμών για ER Γράφημα με $p=0.015$, $n=60$ και η σύγκρισή της με της Poisson.

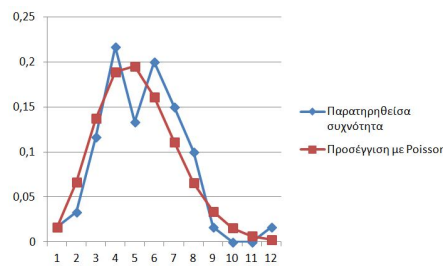
Δοκιμάζουμε το ίδιο για τιμή $p = 0.07$ και το ίδιο πλήθος κόμβων, και έχουμε την εικόνα του Σχήματος (2.3)

Αυτή τη φορά το γράφημα είναι πολύ διαφορετικό. Υπάρχει μια μεγάλη συνεκτική συνιστώσα και ένας μόνο απομονωμένος κόμβος. Το γράφημα έχει 1 απομονωμένο κόμβο, 2 κόμβους με βαθμό 1, 7 κόμβους με βαθμό 2, 13 κόμβους με βαθμό 3, 8 με βαθμό 4, 12 με βαθμό 5, 9 με βαθμό 6, 6 με βαθμό 7, 1 με βαθμό 8 και έναν με βαθμό 11. Έτσι, είμαστε σε θέση να κατασκευάσουμε το επόμενο Γράφημα, το συγκριτικό γράφημα ανάμεσα στις αναμενόμενες και τις παρατηρηθείσες τιμές των βαθμών των κόμβων του γραφήματος.

Βλέπουμε πως και αυτή τη φορά η προσέγγιση πλησιάζει αρκετά στις τιμές που παρατηρήθηκαν, αν και είναι λογικό να περιμένουμε καλύτερη προσέγγιση όσο το n μεγαλώνει.



Σχήμα 2.3: ER Γράφημα με $p=0.07$, $v=60$.



Σχήμα 2.4: Κατανομή των βαθμών για το ER Γράφημα με $p=0.07$, $v=60$ και σύγκριση με Poisson.

Στα δύο τελευταία γραφήματα (2.3) και (2.3) με πορτοκαλί χρωματισμό έχουμε την κατανομή των πιθανοτήτων που περιμένουμε για τους βαθμούς των κόμβων, ενώ με γαλάζιο έχουμε τους ίδιους τους βαθμούς, όπως παρατηρήθηκαν στα τυχαία γραφήματα που δημιουργήθηκαν.

Πιο συγκεκριμένα, για την κατανομή των βαθμών των κόμβων κάποιων από τους ER γράφους μεγάλων διαστάσεων ισχύει το ακόλουθο ενδιαφέρον αποτέλεσμα:

Θεώρημα 2.3.1. Έστω Deg_x ο βαθμός του x κόμβου ενός τυχαίου Erdős-Rényi γραφήματος $ER_{\nu, N(\nu)}$. Έστω επίσης

$$Deg_{MAXx} = \max_{1 \leq x \leq \nu} Deg_x \text{ και } Deg_{MINx} = \min_{1 \leq x \leq \nu} Deg_x$$

ο μέγιστος και ο ελάχιστος βαθμός για τους κόμβους του γραφήματος, ανάμεσα σε κάθε κόμβο x από τους ν διακριτούς κόμβους του γραφήματος.

Για

$$\lim_{\nu \rightarrow +\infty} \frac{N(\nu)}{\nu \log \nu} = +\infty$$

και για κάθε $\varepsilon > 0$ ισχύει

$$\lim_{\nu \rightarrow \infty} P \left(\left| \frac{Deg_{MAXx}}{Deg_{MINx}} - 1 \right| < \varepsilon \right) = 1.$$

Κάτι τέτοιο σημαίνει πως οι βαθμοί όλων των κόμβων του τυχόντος γραφήματος που πληροί τις προϋποθέσεις του παραπάνω Θεωρήματος είναι σχεδόν βεβαίως ίσοι μεταξύ τους: Έστω η σχέση

$$\left| \frac{Deg_{MAXx}}{Deg_{MINx}} - 1 \right| < \varepsilon, \forall \varepsilon > 0. \quad (2.1)$$

Αν ισχύει το παραπάνω συμπέρασμα με πιθανότητα 1, θα υπάρχει μόνο ένα σύνολο N στο οποίο θα ανήκουν οι γράφοι για τους οποίους δεν ισχύει η σχέση (2.1), και για το οποίο θα είναι $P(N) = 0$. Δηλαδή, για σχεδόν όλους τους γράφους θα έχουμε $Deg_{MAXx} = Deg_{MINx}$ και επομένως δεν θα υπάρχουν κόμβοι πιο καλά συνδεδεμένοι σε σχέση με άλλους στο γράφημα. Ισοδύναμα, δεν θα υπάρχουν πιο κεντρικοί, σε σχέση με το βαθμό, κόμβοι στο γράφημα: θα είναι όλοι το ίδιο «κεντρικοί».

Απόδειξη:

Σε τυχόν $ER_{\nu, N}$ γράφο υπάρχουν συνολικά

$$\binom{\nu}{2}$$

τρόποι να διαλέξουμε N από τις $\binom{\nu}{2}$ πιθανές ακμές. Οι συνολικά πιθανοί τρόποι ² να

²Στην πραγματικότητα, παρότι η δημιουργία των ακμών ανάμεσα στους κόμβους είναι ανεξάρτητη διαδικασία για την κάθε πλευρά, υπάρχει καποιου είδους εξάρτηση στους βαθμούς των κόμβων, πράμα που θα έχει επίπτωση στην κατανομή τους.

Πράγματι, έστω ένα γράφημα $\Gamma_{\nu, E}$ και δύο κόμβοι τυχαία επιλεγμένοι, έστω x, y με βαθμούς Deg_x και Deg_y αντίστοιχα. Τότε είναι:

$$Deg_x = \sum_{y \in V} \mathbb{K}_E(x, y) \text{ και } Deg_y = \sum_{x \in V} \mathbb{K}_E(y, x).$$

Οι τ.μ. Deg_x και Deg_y ακολουθούν τη Διωνυμική κατανομή, οπότε για τις αναμενόμενες τιμές τους ισχύει ότι:

$$E[Deg_x] = \sum_{y \in V, y \neq x} P[(x, y) \in E] = (\nu - 1)p$$

και

$$E[Deg_y] = \sum_{x \in V, x \neq y} P[(x, y) \in E] = (\nu - 1)p.$$

Ακόμη,

$$\begin{aligned} Deg_x Deg_y &= \sum_{u \in V} \mathbb{K}_E(x, u) \sum_{v \in V} \mathbb{K}_E(y, v) \\ &= \sum_{u, v \in V * V, u \neq x, v \neq y} \mathbb{K}_E(x, u) \mathbb{K}_E(y, v) \\ &= \sum_{u=y, v=x} \mathbb{K}_E(x, y) \mathbb{K}_E(y, x) + \sum_{(x, u) \neq (v, y)} \mathbb{K}_E(x, u) \mathbb{K}_E(y, v) \\ &= \mathbb{K}_E(x, y) \mathbb{K}_E(y, x) + \sum_{(x, u) \neq (v, y)} \mathbb{K}_E(x, u) \mathbb{K}_E(y, v) \\ &= \mathbb{K}_E(y, x) + \sum_{(x, u) \neq (v, y)} \mathbb{K}_E(x, u) \mathbb{K}_E(y, v), \end{aligned}$$

μιας και μια ακμή ή θα ενώνει τους x, y , ή δεν θα τους ενώνει.

Σύμφωνα με τα παραπάνω, λοιπόν, έχουμε:

$$\mathbb{E}[Deg_x Deg_y] = \mathbb{1}_E(y, x) + \sum_{(x, u) \neq (v, y)} \mathbb{1}_E(x, u) \mathbb{1}_E(y, v),$$

κατανεμηθούν οι ακμές στο γράφημα όταν ένας υπό εξέταση τυχαίος κόμβος είναι γείτονας με ακριβώς k άλλους κόμβους στο γράφημα είναι

$$\binom{\nu-1}{k} \binom{\binom{\nu-1}{2}}{N-k}$$

αφού με τον πρώτο παράγοντα απαριθμούνται οι πιθανοί τρόποι να επιλεγθούν οι k κόμβοι-γείτονες του κόμβου και με τον δεύτερο οι συνολικοί τρόποι να μοιραστούν οι υπόλοιπες ακμές στο γράφημα. Έτσι, η πιθανότητα ένας τυχόν κόμβος που θα επιλεγεί να έχει βαθμό ακριβώς k είναι:

$$\frac{\binom{\nu-1}{k} \binom{\binom{\nu-1}{2}}{N-k}}{\binom{\binom{\nu}{2}}{N}}$$

$$\begin{aligned} &= p + \sum_{(x,u) \neq (v,y)} p^2 \\ &= p + ((\nu-1)^2 - 1)p^2, \text{ διότι το ζεύγος } (x,y) \text{ θα πρέπει να μετρηθεί μια μόνο φορά.} \end{aligned}$$

Έτσι, έχουμε:

$$\begin{aligned} Cov[Deg_x Deg_y] &= \mathbb{E}[Deg_x Deg_y] - \mathbb{E}[Deg_x] \mathbb{E}[Deg_y] \\ &= p + ((\nu-1)^2 - 1)p^2 - [(\nu-1)p]^2 \\ &= p + [(\nu-1)p]^2 - p^2 - [(\nu-1)p]^2 \\ &= p(1-p). \end{aligned} \tag{2.2}$$

(2.3)

Επειδή είναι $0 < p < 1$, η τιμή της συνδιασποράς των βαθμών δύο τυχαίων κόμβων του γραφήματος είναι φραγμένη και μάλιστα μεγιστοποιείται για $p = 1/2$, λαμβάνοντας τιμή ίση με $1/4$.

Έτσι,

$$Cov \left[\frac{Deg_x}{\nu-1}, \frac{Deg_x}{\nu-1} \right] = \frac{p(1-p)}{(\nu-1)^2},$$

το οποίο παίρνει εξαιρετικά μικρές τιμές για μεγάλα ν . Δηλαδή η συνδιασπορά είναι μια πρακτικά αμελητέα ποσότητα που μπορούμε να αγνοήσουμε.

2.4 Οριακές Συναρτήσεις και Μεταβολή Φάσης (Threshold Functions and Phase Transition)

Χαρακτηριστικά όπως αυτό που περιγράφηκε προηγουμένως, εξετάζονται πολλές φορές στους γράφους. Από την προηγούμενη διαδικασία φάνηκε ότι για διαφορετικές τιμές της πιθανότητας p , με την οποία δημιουργούνται οι ακμές, χαρακτηριστικά όπως το πλήθος και το μέγεθος των συνεκτικών συνιστωσών ή το πλήθος των απομονωμένων κόμβων λαμβάνουν διαφορετική τιμή. Επειδή, όπως ήδη εξηγήθηκε, δεν είναι πάντα εύκολο να γίνει μια διαδικασία μελέτης των ιδιοτήτων των τυχαίων γράφων όπως αυτή που περιγράφηκε προηγουμένως, θα χρησιμοποιούνται στις αναλύσεις μεγάλων διαστάσεων γράφοι χωρίς αυτό να σημειώνεται κάθε φορά.

Είναι σημαντικό να μπορέσουμε να καταλάβουμε πώς, με μια μικρή αλλαγή στην τιμή μιας παραμέτρου, μπορεί να υπάρξει σημαντική αλλαγή στην ίδια τη δομή του γραφήματός μας. Για παράδειγμα, δεδομένου του πλήθους των κόμβων ενός γραφήματος, η αλλαγή της πιθανότητας p με την οποία εμφανίζονται οι ακμές μπορεί να συνεπάγεται σοβαρές αλλαγές στη δομή του γραφήματος που δημιουργείται.

Όπως έχει ήδη αναφερθεί, ενδιαφέρει ιδιαίτερα να βρεθεί μια «τυπική δομή», ένα χαρακτηριστικό που ίσως να υπάρχει στο γράφο μας. Λέγοντας «τυπική δομή», εννοείται μία ιδιότητα που μας ενδιαφέρει αν εμφανίζεται ή όχι στο γράφημα με πιθανότητα που τείνει στο 1. Υπάρχει εξάρτηση ανάμεσα στην τιμή της πιθανότητας p , το μέγεθος του γράφου και την εμφάνιση κάποιας ιδιότητας;

Φαίνεται λογικό να υπάρχει κάποια εξάρτηση. Για παράδειγμα, έστω ότι θέλουμε να δούμε αν ισχύει η ιδιότητα «Ο μέσος βαθμός των κόμβων του γραφήματος να είναι πεπερασμένος». Αν η p είναι σταθερή, και ανεξάρτητη από το n , τότε αυτό θα σημαίνει πως όσο μεγαλώνει το πλήθος των κόμβων τόσο θα μεγαλώνει και ο μέσος βαθμός τους, αφού κάθε κόμβος θα έχει αναμενόμενο αριθμό γειτόνων ίσο με pn . Έτσι, παρά το γεγονός ότι το p παραμένει σταθερό, η ιδιότητα που μελετάμε θα πάψει να ισχύει καθώς το n τείνει στο άπειρο. Καταλήγουμε, λοιπόν, ότι μάλλον θα πρέπει να υπάρχει κάποια εξάρτηση. Για να δούμε αν αυτό συμβαίνει για κάθε ιδιότητα, δηλαδή να δούμε αν υπάρχει εξάρτηση ανάμεσα στην πιθανότητα δημιουργίας p , το μέγεθος του γράφου και την εμφάνιση κάποιας ιδιότητας, ας θεωρήσουμε ότι η πιθανότητα p εξαρτάται από το n : υποθέτουμε, δηλαδή, ότι είναι $p = p(n)$.

Έχοντας, λοιπόν, αυτό σαν δεδομένο, αναρωτιόμαστε ποια είναι κάθε φορά η τιμή της πιθανότητας με την οποία δημιουργούνται οι ακμές στο γράφημα, πάνω από την οποία μια ιδιότητα ισχύει σχεδόν βεβαίως και κάτω από την οποία, σχεδόν βεβαίως, δεν ισχύει. Ταυτόχρονα, θεωρούμε ότι $n \rightarrow +\infty$. Όπως θα εξετασθεί και παρακάτω, πολλές φορές είμαστε σε θέση να βρούμε μια τιμή της πιθανότητας για την οποία συμβαίνει αυτό που περιγράφηκε προηγουμένως. Ακόμη, όμως, και να μην μπορούσαμε να βρούμε μία ακριβή τιμή της, στη βιβλιογραφία αναλύεται ότι θα μπορούσαμε να βρούμε κάποιο διάστημα μέσα στο οποίο η πιθανότητα αυτή κυμαίνεται.

Τι ονομάζεται «*οριακή συνάρτηση*» και τι λαμβάνουμε γνωρίζοντας μια τέτοια;

Ορισμός 2.4.1. Η συνάρτηση $A(v)$ καλείται «*οριακή συνάρτηση*» («*threshold function*») για την ιδιότητα A εάν είναι μονότονη για $v \rightarrow +\infty$ και ισχύει το ακόλουθο:

$$\lim_{v \rightarrow \infty} P_{v,A(v)}(A) = \begin{cases} 0, & \text{για } \lim_{v \rightarrow +\infty} \frac{N(v)}{A(v)} = 0 \\ 1, & \text{για } \lim_{v \rightarrow +\infty} \frac{N(v)}{A(v)} = +\infty \end{cases}$$

όπου

$$P_{v,A(v)}(A) = \frac{A_{v,A(v)}}{\binom{v}{N}}$$

η πιθανότητα επιλέγοντας τυχαία ένα από όλα τα γραφήματα με v κόμβους και N ακμές να βρούμε κάποιο γράφημα που να ικανοποιεί την ιδιότητα A . Τότε λέμε επίσης, ότι για πιθανότητα p που έχει τιμή ίση με $A(v)$, συμβαίνει «*μεταβολή φάσης*» στο γράφημα όσων αφορά στην ιδιότητα A , ή ότι έχουμε «*phase transition*».

Αναζητούμε μια συνάρτηση που να περιγράφει την τιμή της πιθανότητας. Θέλουμε για πιθανότητα μεγαλύτερη από $A(v)$, δηλαδή για τιμή της συνάρτησης μεγαλύτερη από $A(v)$, η ιδιότητα A να ικανοποιείται με πιθανότητα 1 στα γραφήματα, ενώ για πιθανότητα μικρότερη από $A(v)$ η ιδιότητα A να μην ικανοποιείται, σχεδόν βεβαίως. Ψάχνουμε την τιμή της πιθανότητας με την οποία οι πλευρές παράγονται στο γράφημα για την οποία συμβαίνει αυτή η μεταβολή φάσης, δηλαδή από τη δημιουργία γραφημάτων που ικανοποιούν την ιδιότητα, περνάμε στη δημιουργία γραφημάτων που δεν την ικανοποιούν, σχεδόν βεβαίως την πιθανότητα εκείνη που θα κάνει τη δομή των γραφημάτων να αλλάζει από το να περιέχει κάποια στοιχεία στο να μην τα περιέχει.

Από τον τρόπο που έχει οριστεί η συνάρτηση $A(v)$, μπορούμε να παρατηρήσουμε διάφορα πράγματα. Αρχικά, μπορούμε να δούμε ότι δεν είναι απαραίτητα μοναδική για όλες τις περιπτώσεις. Ένα ακόμη πράγμα που μπορούμε να παρατηρήσουμε είναι ότι οι συναρτήσεις αυτές αναφέρονται στο τι θα γίνει ασυμπτωτικά, δεν μας προσφέρουν κάποια αίσθηση της «ταχύτητας» με την οποία αυτό θα συμβεί.

Γενικά το ζήτημα εύρεσης τέτοιων συναρτήσεων είναι ένα από αυτά που απασχολούν στην ελληνική και ξενόγλωσση βιβλιογραφία. Πολλά είναι τα θεωρήματα που έχουν κατά καιρούς διατυπωθεί γι' αυτές, αλλά στην παρούσα εργασία θα γίνει αναφορά σε κάποια μόνο από αυτά.

2.5 Οριακές Συναρτήσεις και Δομή του Γραφήματος

Στην παράγραφο αυτή θα παρουσιαστούν κάποια από τα πιο ενδιαφέροντα Θεωρήματα σχετικά με τη δομή των ER τυχαίων γράφων που συναντήθηκαν στη βιβλιογραφία και έκαναν εντύπωση.

Όπως έχει και προηγουμένως αναφερθεί, στόχος γενικά είναι να παρατηρήσουμε αν υπάρχουν ή όχι συγκεκριμένες δομές στους τυχαίους γράφους, καθώς και να βρεθεί και η πιθανότητα με την οποία αυτό συμβαίνει. Με στόχο αυτό, θα παρουσιαστούν κάποια από τα αποτελέσματα που συναντήθηκαν στη βιβλιογραφία μαζί με κάποια σχόλια. Παρ'ότι εξαιρετικά ενδιαφέρουσες, δεν θα αναφερθούν κάποιες από τις αποδείξεις των σχετικών Θεωρημάτων.

Στην ενότητα αυτή θα αναφερθούν κάποια χαρακτηριστικά που αποκτούν οι τυχαίοι γράφοι όταν η πιθανότητα ύπαρξης ακμής p γίνεται μεγαλύτερη από κάποια τιμή «φράγμα». Τα αποτελέσματα που παρουσιάζονται αφορούν κυρίως στη δομή των τυχαίων ER γραφημάτων και το πλήθος των συνεκτικών συνιστώσών τους.

Πριν ξεκινήσουμε, θα οριστεί ο παρακάτω τύπος γραφήματος:

Ορισμός 2.5.1. Το γράφημα $ER_{\nu, N}$ θα λέγεται ότι είναι τύπου A αν αποτελείται από μια συνεκτική συνιστώσα $\nu - k$ κόμβων και από k απομονωμένους κόμβους. Κάθε γράφος που δεν είναι τύπου A θα λέγεται ότι είναι τύπου \bar{A} .

$$\bullet \quad N_c = \left\lfloor \frac{1}{2} \nu \log \nu + c\nu \right\rfloor.$$

Ισχύει ότι:

Θεώρημα 2.5.1. Έστω $P(\bar{A}, \nu, N_c)$ η πιθανότητα να είναι ο γράφος ER_{ν, N_c} τύπου \bar{A} , όπου $N_c = \left\lfloor \frac{1}{2} \nu \log \nu + c\nu \right\rfloor$. Έχουμε:

$$\lim_{\nu \rightarrow +\infty} P(\bar{A}, \nu, N_c) = 0 \quad (2.4)$$

Απόδειξη. Η απόδειξη υπάρχει στο [16]. □

Σύμφωνα με το παραπάνω Θεώρημα, ένας ER τυχαίος γράφος με ν κόμβους και πλήθος ακμών ίσο με N_c για κάποια σταθερά c , είναι σχεδόν βεβαίως γράφος τύπου A . Επειδή, όπως αναφέρεται και στη βιβλιογραφία, γράφοι που έχουν παραχθεί σύμφωνα με τις Διαδικασίες 1 και 2 προηγούμενης ενότητας, για μεγάλες τιμές του ν , μπορούν να θεωρηθούν γράφοι με αμελητέες διαφορές ως προς τη δομή και γενικότερα τις περισσότερες ιδιότητές τους, έχουμε ότι και για τους γράφους με τιμή της της πιθανότητας p περίπου ίση με

$$p = \frac{N_c}{\binom{\nu}{2}} = \frac{\frac{1}{2} \nu \log \nu + c\nu}{\frac{\nu!}{2!(\nu-2)!}} = \frac{\frac{1}{2} \nu \log \nu + c\nu}{\frac{\nu(\nu-1)}{2}} = \frac{\log \nu}{\nu-1} + 2 \frac{c}{\nu-1},$$

που ασυμπτωτικά είναι ίση με $\frac{\log \nu}{\nu}$, θα ισχύει το ίδιο αποτέλεσμα. Τουτ'έστιν, γράφοι που έχουν παραχθεί μέσω της Διαδικασίας 1 για αυτήν την τιμή της πιθανότητας p ή πλήθος ακμών ίσο με N_c , είναι σχεδόν βεβαίως γράφοι τύπου A.

Παρατηρούμε πως, σε κάθε περίπτωση, σχηματίζονται με πιθανότητα 1 γράφοι που αποτελούνται από **μια μεγάλη συνεκτική συνιστώσα** και **κάποιους απομονωμένους κόμβους**. Άρα, στο σύνολο των ER γραφημάτων που προκύπτουν μέσω των Διαδικασιών 1 και 2 για την προαναφερθείσα τιμή της πιθανότητας, θα εμφανίζονται σχεδόν βεβαίως γράφοι στους οποίους δεν υπάρχουν συνιστώσες με 2-3 ή και παραπάνω κόμβους, εκτός από την κύρια συνιστώσα. Έτσι, κατά τα προηγούμενα, η πιθανότητα εμφάνισης ενός γραφήματος με πολλές μικρές συνιστώσες (για το συγκεκριμένο πλήθος ακμών) είναι μηδέν.

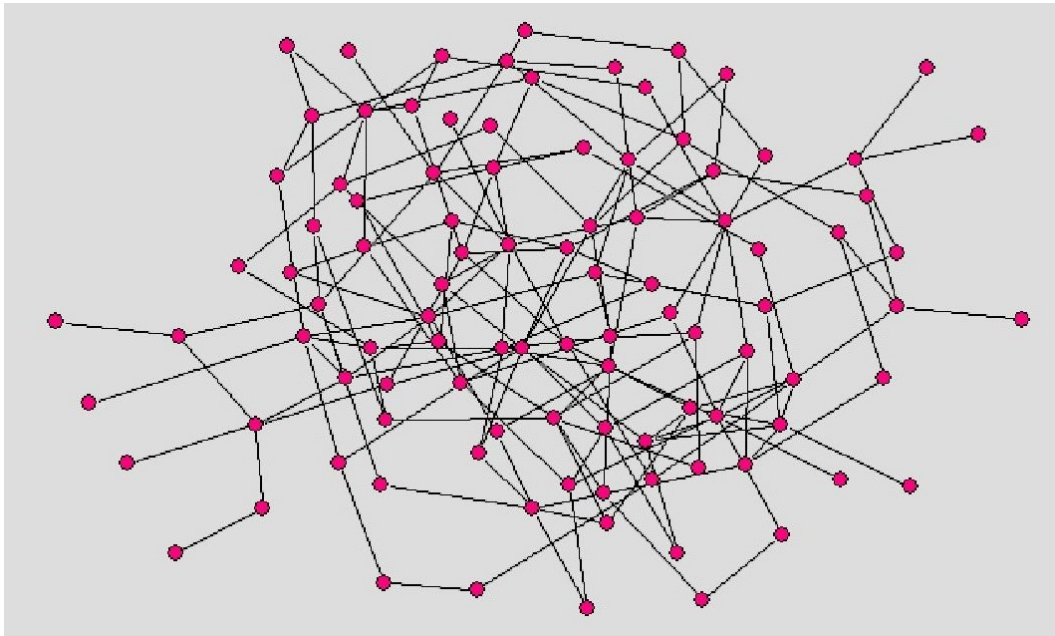
Εφαρμογή: Χρησιμοποιώντας κατάλληλο πρόγραμμα μπορούμε να κάνουμε κάποιες δοκιμές για την οπτική επαλήθευση του παραπάνω αποτελέσματος. Σε ένα γράφο με 100 κόμβους και για $c = 1/4$, έχουμε

$$N_c = \frac{1}{2} \nu \log \nu + c\nu = 175,$$

ή

$$p = \frac{N_c}{\binom{\nu}{2}} = 0.035.$$

Επιλέγουμε να κατασκευάσουμε το γράφημα δίνοντας στο πρόγραμμα τιμή για την πιθανότητα p , δηλαδή σύμφωνα με την πρώτη διαδικασία. Μεταξύ άλλων, λαμβάνουμε και το χαρακτηριστικό γράφημα που ακολουθεί.



Σχήμα 2.5: Τυχαίος ER γράφος με $\nu = 100$, $p = 0.035$.

Παρατηρούμε ότι ο γράφος που προέκυψε είναι συνεκτικός, πράγμα που αναμενόταν, με βάση το Θεώρημα 2.5.3, από την τιμή της πιθανότητας p . Ακόμη, φαίνεται να υπάρχουν αρκετοί κόμβοι με βαθμό ένα, πράγμα που δείχνει ότι, πράγματι, πριν ενωθούν με την κυρίως συνεκτική συνιστώσα, το γράφημα ήταν τύπου A . Πάντως, από την απεικόνιση αυτή φαίνεται ότι δεν είναι και πολύ πιθανό να μην ήταν τέτοιου τύπου, δηλαδή να αποτελούνταν από δύο και παραπάνω συνεκτικές συνιστώσες, μιας και στην κυρίως συνιστώσα φαίνεται ότι οι κόμβοι είναι καλά συνδεδεμένοι μεταξύ τους.

◇

Ένα άλλο αποτέλεσμα σχετικό με τη δομή των ER τυχαίων γράφων είναι το ακόλουθο:

Θεώρημα 2.5.2. Έστω $P_k(\nu, N_c)$ η πιθανότητα ο γράφος ER_{ν, N_c} να έχει στη μέγιστη συνεκτική του συνιστώσα $\nu - k$ κόμβους, όπου N_c αυτό του Θεωρήματος 2.5.1. Τότε,

$$\lim_{\nu \rightarrow +\infty} P_k(\nu, N_c) = \frac{(e^{-2c})^k e^{e^{-2c}}}{k!}. \quad (2.5)$$

Μέσω του Θεωρήματος αυτού βλέπουμε ότι το ενδεχόμενο ένας τυχαίος ER γράφος να αποτελείται από μια μεγάλη συνεκτική συνιστώσα και k απομονωμένους κόμβους είναι μια τυχαία μεταβλητή που ακολουθεί την κατανομή Poisson με παράμετρο e^{-2c} . Ακολουθεί η :

Απόδειξη. Σύμφωνα με το Θεώρημα 2.5.1 με πιθανότητα 1 θα προκύψουν γράφοι τύπου A . Έτσι, δίχως βλάβη της γενικότητας, μπορούμε για την απόδειξη του προηγούμενου αποτελέσματος να θεωρήσουμε μόνο γραφήματα τέτοιου τύπου.

Το πλήθος των συνεκτικών γραφημάτων που αποτελούνται από $\nu - k$ κόμβους και N_c ακμές είναι ίσο με $\binom{\nu-k}{N_c}$, ενώ το συνολικό πλήθος των γραφημάτων με ν κόμβους και N_c ακμές είναι $\binom{\nu}{N_c}$. Όλοι οι πιθανοί γράφοι με ν κόμβους εκ των οποίων οι $\nu - k$ να είναι στην ίδια συνεκτική συνιστώσα είναι $\binom{\nu}{k} \binom{\nu-k}{N_c}$.

Για την πιθανότητα να είναι ένας γράφος με ν κορυφές και N_c ακμές συνεκτικός είναι από Θεώρημα των Erdős και Renyi στο [16], ισχύει ότι: $\lim_{\nu \rightarrow +\infty} P(ER_{\nu, N_c} \text{ is connected}) = e^{e^{-2c}}$, οπότε, το αντίστοιχο αποτέλεσμα για γράφο με $\nu - k$ κόμβους θα δίνεται από τη σχέση: $\lim_{\nu \rightarrow +\infty} P(ER_{\nu-k, N_c} \text{ is connected}) = e^{e^{-2c}}$.

Ακόμη, μέσω του τύπου του Stirling³ μπορεί να αποδειχθεί ότι ισχύει, για τη συγκεκριμένη τιμή του πλήθους N_c , το:

$$\lim_{\nu \rightarrow \infty} \frac{\binom{\nu}{k} \binom{\nu-k}{N_c}}{\binom{\nu}{N_c}} = \frac{e^{-2kc}}{k!}.$$

Έτσι, για την πιθανότητα να έχει ο τυχαίος ER γράφος k απομονωμένους κόμβους και μια μεγάλη συνεκτική συνιστώσα αποτελούμενη από τους υπόλοιπους $\nu - k$ κόμβους

³ $\lim_{n \rightarrow +\infty} \frac{n!}{\left(\frac{n}{e}\right)^{\sqrt{2n\pi}}} = 1$

του γραφήματος έχουμε:

$$P_k(\nu, N_c) \sim \frac{\binom{\nu}{k} \binom{\nu-k}{N_c}}{\binom{\nu}{N_c}} P_0(\nu - k, N_c),$$

όπου $P_a(b, c)$ η πιθανότητα ο γράφος να είναι τύπου A με b κόμβους, c ακμές και a απομονωμένους κόμβους. Στη συνέχεια, συνδιάζοντας τα προηγούμενα και λαμβάνοντας υπ'όψιν ότι $N_c = \frac{1}{2}\nu \log \nu + c\nu$, καταλήγουμε στο αποτέλεσμα του Θεωρήματος. \square

Το παραπάνω, σε συνδιασμό με το προηγούμενό του, οδηγούν στο συμπέρασμα πως:

Πόρισμα 2.5.1. Έστω $\Pi_k(\nu, N_c)$ η πιθανότητα ο γράφος ER_{ν, N_c} να αποτελείται από $k + 1$ ζένες μεταξύ τους, συνεκτικές συνιστώσες, όπου N_c αυτό του Θεωρήματος 2.5.1. Τότε,

$$\lim_{\nu \rightarrow +\infty} \Pi_k(\nu, N_c) = \frac{(e^{-2c})^k e^{-2c}}{k!}. \quad (2.6)$$

Απόδειξη. Το αποτέλεσμα είναι προφανές από το συνδιασμό των προηγούμενων δύο Θεωρημάτων. \square

Το Πόρισμα αυτό μας αναδεικνύει το γεγονός ότι το πλήθος των συνεκτικών συνιστωσών του γράφου μειωμένο κατά ένα είναι μια τ.μ. που ακολουθεί την Κατανομή Poisson με παράμετρο e^{-2c} .

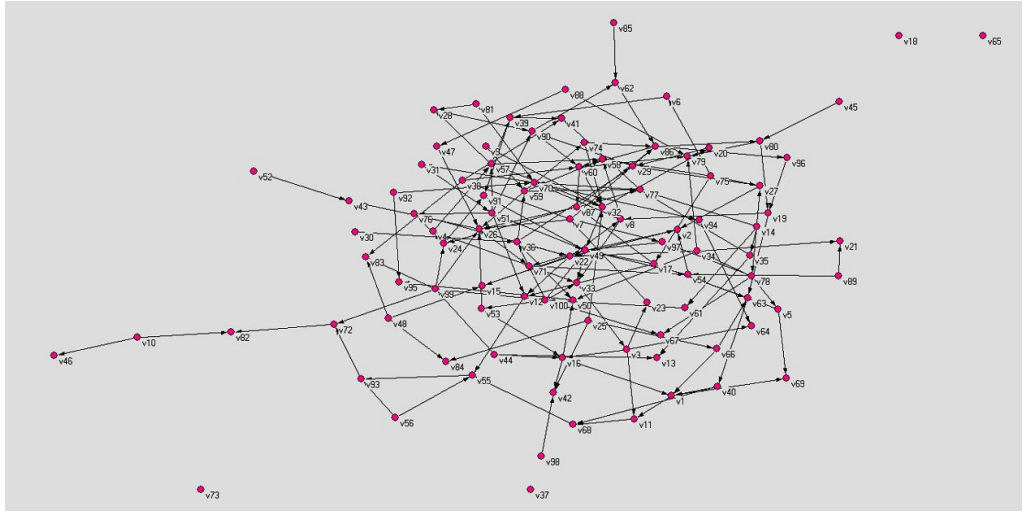
Τα αποτελέσματα αυτά είναι άμεσα εξαρτημένα και η σύνδεση με τα προηγούμενα είναι προφανής: το γεγονός ένας γράφος να έχει $k + 1$ συνιστώσες, σε συνδιασμό με το γεγονός ότι με πιθανότητα 1 οι k θα είναι μεμονωμένοι κόμβοι και η μια που απομένει θα είναι μια μεγάλη συνεκτική συνιστώσα, δίνει άμεσα ότι η τελευταία θα έχει, με πιθανότητα 1, $\nu - k$ κόμβους.

Εφαρμογή: Δημιουργούμε έναν τυχαίο ER γράφο με τη βοήθεια προγράμματος. Επιλέγουμε αυτός να έχει 100 κόμβους και 175 ακμές. Αυτή τη φορά κατασκευάζουμε το γράφο σύμφωνα με τη δεύτερη διαδικασία. Τα περισσότερα από τα αποτελέσματα που σχετίζονται με αυτούς, όπως έχει προαναφερθεί, είναι ίδια για γράφους ίδιου μεγέθους, με όποια από τις δύο διαδικασίες κατασκευής και αν έχουν προκύψει. Ακολουθεί ο γράφος που κατασκευάστηκε.

Παρατηρούμε ότι ο γράφος έχει 4 απομονωμένους κόμβους και μια μεγάλη συνεκτική συνιστώσα, δηλαδή είναι τύπου A , όπως ακριβώς αναμενόταν. Ο γράφος είναι τύπου A με 5 συνεκτικές συνιστώσες, με τη μεγαλύτερη από αυτές να αποτελείται από $175 - 4 = 171$ κόμβους.

Ακόμη, με βάση το Πόρισμα 2.5.3, περιμέναμε ο γράφος που θα προκύψει να είναι συνεκτικός, πράγμα που δεν βλέπουμε να συμβαίνει. Έτσι, για το γράφο που λάβαμε αναρωτιώμαστε την πιθανότητα που είχε να εμφανιστεί. Για να γίνει αυτό, χρειαζόμαστε την παράμετρο c . Παρατηρούμε ότι για τον γράφο αυτό προκύπτει ίδια τιμή για την πιθανότητα p και την παράμετρο c με αυτή του προηγούμενου, $p = 0.035$ και $c = 1/4$. Τώρα μπορούμε να προχωρήσουμε στον υπολογισμό της πιθανότητας. Είναι:

$$\lim_{\nu \rightarrow +\infty} \Pi_k(100, 175) = \frac{(e^{-2c})^k e^{-2c}}{k!} = \frac{(e^{-\frac{1}{2}})^4 e^{-\frac{1}{2}}}{4!}$$



Σχήμα 2.6: Τυχαίος ER γράφος με $\nu = 100$, $c = 1/4$, $N_c = 175$.

$$= 0.0103.$$

Η πιθανότητα αυτή είναι πολύ μικρή, πράγμα που δεν αντίκειται στο Πόρισμα 2.5.3. Το γεγονός ότι δεν είναι πάρα πολύ μικρή μπορεί να δικαιολογηθεί από το γεγονός ότι η τάξη (100 κόμβοι) του γραφήματος δεν είναι πάρα πολύ μεγάλη, δηλαδή, η πιθανότητα εμφάνισης γράφων ακόμη μεγαλύτερων αλλά με, κατά τα άλλα, ίδια χαρακτηριστικά θα είναι ακόμη μικρότερη από 0.010.

◇

Το ενδεχόμενο το γράφημα που θα προκύψει μέσω κάποιας διαδικασίας να είναι συνεκτικό είναι ισοδύναμο με το ενδεχόμενο το γράφημα που θα προκύψει να αποτελείται από μια μεγάλη συνεκτική συνιστώσα αποτελούμενη από ν κόμβους και κανέναν απομονωμένο κόμβο. Έτσι, προκύπτει το ακόλουθο:

Πόρισμα 2.5.2. Έστω $P_0(\nu, N_c)$ η πιθανότητα ο γράφος ER_{ν, N_c} να είναι συνεκτικός, όπου N_c αυτό του Θεωρήματος 2.5.1. Τότε,

$$\lim_{\nu \rightarrow +\infty} P_0(\nu, N_c) = e^{e^{-2c}}. \quad (2.7)$$

Απόδειξη. Άμεση από τα προηγούμενα. □

Με βάση το τελευταίο μπορούμε να συμπεράνουμε ότι ισχύει το ακόλουθο.

Θεώρημα 2.5.3. Η συνάρτηση

$$N(\nu) = \frac{1}{2} \nu \log \nu$$

είναι οριακή συνάρτηση για την ιδιότητα της συνεκτικότητας.

Απόδειξη. Πράγματι, για να ισχύει κάτι τέτοιο, αρκεί να ισχύει ότι

$$\lim_{\nu \rightarrow \infty} P_{ER_{\nu, N_c}}(ER_{\nu, N_c} \text{ is connected}) = \begin{cases} 1, & \text{για } N > (\frac{1}{2} + \epsilon) \nu \log \nu \\ 0, & \text{για } N < (\frac{1}{2} - \epsilon) \nu \log \nu \end{cases}$$

για $\epsilon > 0$. Το παραπάνω εύκολα επαληθεύεται με πράξεις και θα παραληφθεί. \square

• $N = N(c) \sim c\nu, c > 0.$

Για πλήθος ακμών διαφορετικό από N_c όπως αυτό που μέχρι στιγμής εξετάστηκε, ισχύουν διαφορετικά αποτελέσματα. Στη συνέχεια της ενότητας παρουσιάζεται ένα μόνο Θεώρημα, σχετικό με ER γράφους με ν κόμβους και $N = N(c) \sim c\nu$ ακμές, όπου c θετική σταθερά, δηλαδή για γραφήματα συγκριτικά πιο αραιά από τα προηγούμενα (για ίδιες τιμές της παραμέτρου c). Τα υπόλοιπα Θεωρήματα που αφορούν γράφους τέτοιου μεγέθους παρουσιάζονται στην επόμενη ενότητα μαζί με κάποια άλλα που σχετίζονται με τη Διάμετρο των τυχαίων ER γράφων.

Ισχύουν τα ακόλουθα:

Θεώρημα 2.5.4. *Εστω $Z_{\nu, N}$ το πλήθος των συνεκτικών συνιστωσών του ER γραφήματος $\Gamma_{\nu, N}$ και έστω $N = N(c) \sim c\nu$. Τότε:*

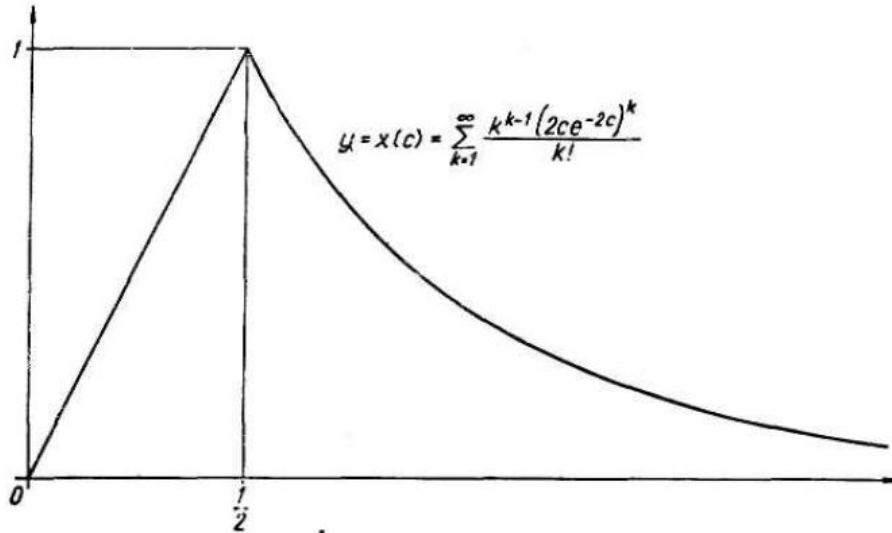
- $E(Z_{\nu, N_c}) = \nu - N + O(1)$, για $0 < c < 1/2$ και τον όρο $O(1)$ να εξαρτάται αποκλειστικά από την παράμετρο c ,
- $E(Z_{\nu, N_c}) = \nu - N + O(\log \nu)$, για $c \sim 1/2$,
- $\lim_{\nu \rightarrow \infty} \frac{E(Z_{\nu, N_c})}{\nu} = \frac{1}{2} \left(x(c) - \frac{x^2(c)}{2} \right)$, για $c > 1/2$, με $x(c) = \sum_{k=1}^{+\infty} \frac{(k)^{k-1}}{k!} (2ce^{-2c})^k$, όπως στο σχήμα 2.5

Απόδειξη. Η απόδειξη υπάρχει στην 6η παράγραφο του [15] και θα παραληφθεί. \square

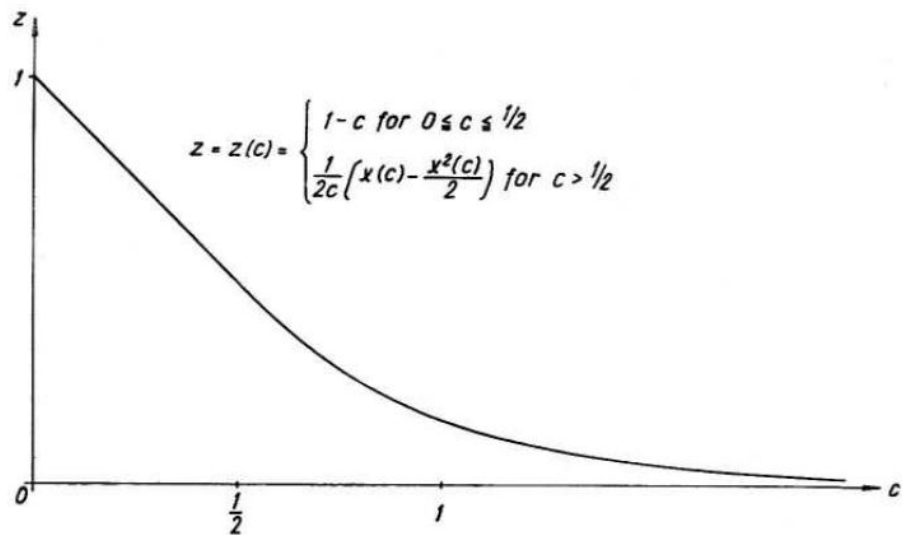
Από το Θεώρημα αυτό δίνεται ένα ακόμα παράδειγμα των μεταβολών φάσης στα τυχαία γραφήματα. Βλέπουμε ότι οι μεταβολές αυτές μπορεί να εξαρτώνται από την αλλαγή στην τιμή μιας και μόνο παραμέτρου. Στο συγκεκριμένο αποτέλεσμα, για τιμές της παραμέτρου c μικρότερες ή περίπου ίσες με $1/2$, το πλήθος των συνιστωσών είναι μια φθίνουσα γραμμική συνάρτηση του πλήθους των ακμών στο γράφημα, για οποιοδήποτε σταθερό ν . Αντίθετα, για τιμές της c μεγαλύτερες του $1/2$, η μείωση αυτή δεν συμβαίνει με τον ίδιο ρυθμό, αλλά πραγματοποιείται με έναν συγκριτικά πιο αργό. Το γράφημα του μέσου πλήθους των συνιστωσών ενός ER γραφήματος με τα παραπάνω χαρακτηριστικά διαιρεμένο με το πλήθος των κόμβων του, φαίνεται στο Σχήμα (2.5).

Η σχέση από την οποία προκύπτει αυτό είναι η

$$z = z(c) = \lim_{\frac{N(\nu)}{\nu} \rightarrow c} \frac{E(Z_{\nu, N(\nu)})}{\nu} = \begin{cases} 1 - c, & \text{για } c \in [0, 1/2] \\ \frac{1}{2} \left(x(c) - \frac{x^2(c)}{2} \right), & \text{για } c > 1/2 \end{cases}$$



Σχήμα 2.7: Η συνάρτηση $x(c) = \sum_{k=1}^{+\infty} \frac{(k)^{k-1}}{k!} (2ce^{-2c})^k$.



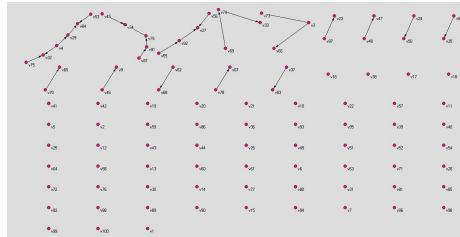
Σχήμα 2.8: Η σχέση του λόγου του μέσου πλήθους των συνιστωσών και του πλήθους των κόμβων του γραφήματος είναι φίνουσα καθώς το πλήθος των κόμβων αυξάνει.

Ίσως να μπορούσαμε να πούμε ότι ένα τέτοιο αποτέλεσμα είναι διαισθητικά αναμενόμενο. Όταν $c \leq 1/2$ οι ακμές του γραφήματος που προκύπτει είναι λιγότερες από ότι όταν $c > 1/2$. Έτσι, οι συνεκτικές συνιστώσες που προκύπτουν στην πρώτη περίπτωση περιμένουμε να είναι περισσότερες από ότι στη δεύτερη. Έστω ότι θέλουμε να εισάγουμε μια καινούρια ακμή σε δύο ήδη έτοιμα γραφήματα που έχουν προκύψει, το πρώτο με $c \leq 1/2$ και το δεύτερο με $c > 1/2$. Στην πρώτη περίπτωση υπάρχουν περισσότερες συνιστώσες από ότι στο δεύτερο, άρα είναι πιο πιθανό το ζευγάρι των κόμβων ανάμεσα στους οποίους θα διαλέξουμε να τοποθετήσουμε την ακμή να είναι σε διαφορετικές συνιστώσες από ότι στο δεύτερο γράφημα, μιας και στο τελευταίο περισσότεροι κόμβοι βρίσκονται σε ίδιες συνεκτικές συνιστώσες σε σχέση με το πρώτο.

Εφαρμογή: Κατασκευάζουμε με κατάλληλο πρόγραμμα τυχαίους ER γράφους επιλέγοντας αυτοί να έχουν χαρακτηριστικά που ικανοποιούν τις προϋποθέσεις του προηγούμενου Θεωρήματος. Επιλέχθηκε, δίχως βλάβη της γενικότητας, η κατασκευή να γίνει μέσω της δεύτερης Διαδικασίας, για γράφους που αποτελούνται από 100 κόμβους. Λαμβάνουμε:

- Για $c = 1/4$:

Το γράφημα που θα προκύψει θα έχει $N(c) = \frac{1}{4}\nu = 25$ ακμές. Το αναμενόμενο πλήθος συνιστωσών είναι ίσο με $E(Z_{\nu, N(c)}) = \nu - \frac{1}{4}\nu + O(1) = 100 - 25 + O(1) = 75$ συνιστώσες. Η γραφική αναπαράσταση του γραφήματος που προέκυψε είναι αυτή του Σχήματος (2.5).



Σχήμα 2.9: Εφαρμογή του Θεωρήματος 2.5.4, για $c = 1/4$.

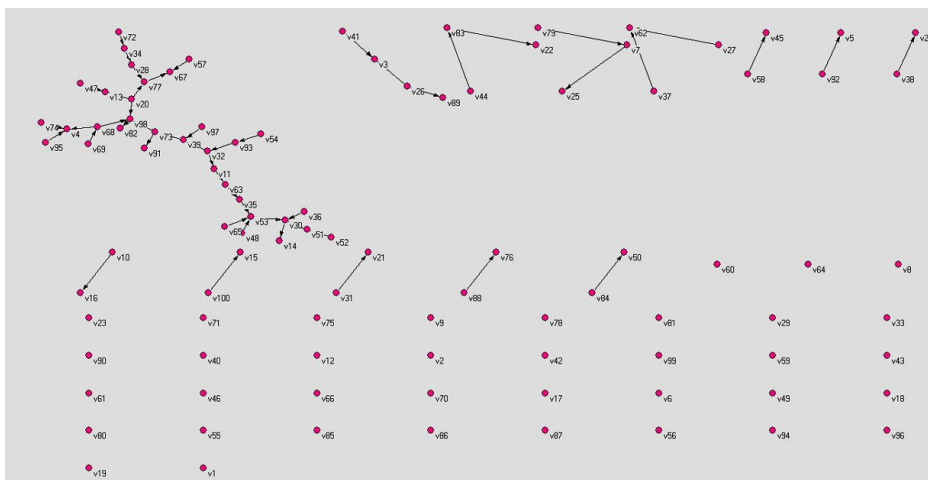
Παρατηρούμε ότι το γράφημα που προέκυψε έχει 61 απομονωμένους κόμβους και 14 συνιστώσες που αποτελούνται από μερικούς κόμβους η καθεμία. Η μεγαλύτερη από αυτές αποτελείται από 6 κόμβους, είναι αρκετά μικρή. Στο σύνολό τους οι συνιστώσες είναι 75 και έτσι επιβεβαιώνεται το Θεώρημα για $c = 1/4$.

- Για $c = 1/2$:

Το γράφημα που θα προκύψει θα έχει $N(c) = \frac{1}{2}\nu = 50$ ακμές. Το αναμενόμενο πλήθος συνιστωσών αυτή τη φορά είναι $E(Z_{\nu, N(c)}) = \nu - \frac{1}{2}\nu + O(\log 100) = 100 - 50 + 2 = 52$ συνιστώσες. Η γραφική αναπαράσταση του γραφήματος που προέκυψε είναι αυτή του Σχήματος (2.5). Παρατηρούμε ότι το γράφημα που προέκυψε έχει 37 απομονωμένους κόμβους, 13 μικρές συνιστώσες που αποτελούνται από μερικούς κόμβους η καθεμία και μια μεγάλη συνιστώσα που αποτελείται από 34 κόμβους. Στο σύνολό τους οι συνιστώσες είναι 50, πολύ κοντά στο αναμενόμενο πλήθος.

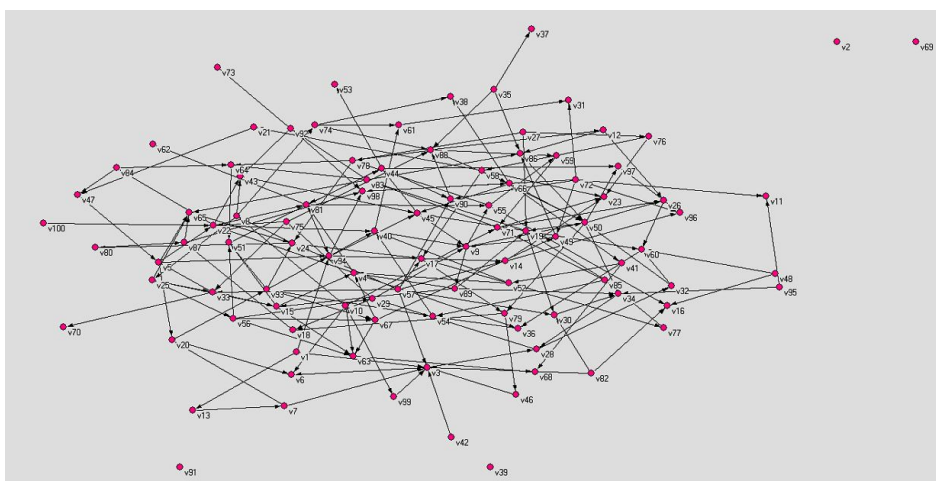
- Για $c = 2$:

Το γράφημα που θα προκύψει θα έχει $N(c) = 2\nu = 200$ ακμές. Το αναμενόμενο πλήθος συνιστωσών διαιρεμένο με το πλήθος των κόμβων είναι $\frac{E(Z_{\nu, N(c)})}{\nu} = \frac{1}{2} \left(x(c) - \frac{x^2(c)}{2} \right)$ συνιστώσες. Η γραφική αναπαράσταση του γραφήματος που προέκυψε είναι αυτή του



Σχήμα 2.10: Εφαρμογή του Θεωρήματος 2.5.4, για $c = 1/2$.

Σχήματος (2.5). Παρατηρούμε ότι το γράφημα που προέκυψε έχει μια μεγάλη συνε-



Σχήμα 2.11: Εφαρμογή του Θεωρήματος 2.5.4, για $c = 2$.

κτική συνιστώσα και 4 απομονωμένους κόμβους. Στο σύνολό τους οι συνιστώσες είναι 5, άρα ο λόγος $\frac{E(Z_{\nu, N(c)})}{\nu}$ του Θεωρήματος παίρνει τιμή ίση με 0.05. Από το γράφημα 2.5 ο αναμενόμενος αριθμός είναι πολύ πολύ κοντά στο 0.05!

Σε κάθε περίπτωση, τα γραφήματα που προέκυψαν επιβεβαιώνουν το αποτέλεσμα του Θεωρήματος.

◇

2.6 Οριακές Συναρτήσεις και Διάμετρος

Συμπληρωματικά με τα Θεωρήματα που αναπτύχθηκαν στην προηγούμενη ενότητα, ισχύουν και τα ακόλουθα Θεωρήματα που αφορούν στη δομή ενός τυχαίου ER γράφου. Τα αποτελέσματα αυτά, ξεκαθαρίζοντας κάπως πιο ειδικά τη δομή που περιμένουμε ένας τέτοιος γράφος να έχει, μας επιτρέπουν να βγάλουμε κάποια συμπεράσματα σχετικά με τη διάμετρό τους, πράγμα ιδιαίτερα χρήσιμο σε πολλές περιπτώσεις.

Θεώρημα 2.6.1. Έστω $V_{\nu, N}$ το πλήθος των κόμβων του ER_{ν, N_c} που ανήκουν σε κάποιο απομονωμένο δέντρο στον ER_{ν, N_c} . Έστω επιπλέον, ότι

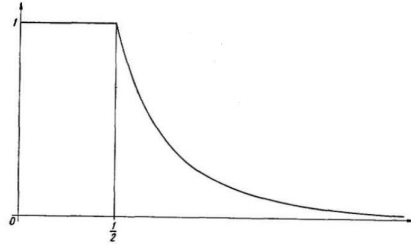
$$\lim_{\nu \rightarrow +\infty} \frac{N(c)}{\nu} = c > 0, \quad (2.8)$$

και $x(c)$ η συνάρτηση όπως ορίστηκε στην προηγούμενη ενότητα. Τότε, ισχύει ότι:

$$\lim_{\nu \rightarrow \infty} \frac{E(V_{\nu, N})}{\nu} = \begin{cases} 1, & \text{για } c \leq 1/2, \\ \frac{x(c)}{2c}, & \text{για } c > 1/2. \end{cases} \quad (2.9)$$

Απόδειξη. Για την απόδειξη του παρόντος αποτελέσματος, μπορούμε να ανατρέξουμε στο [15], στις σελίδες 37-40. \square

Σύμφωνα με το παραπάνω Θεώρημα, για γράφους στους οποίους ισχύει ότι $N(c) = O(\nu) = c\nu$ το πλήθος των κόμβων που βρίσκονται σε κάποιο απομονωμένο δέντρο φαίνεται στο ακόλουθο γράφημα για τις διάφορες τιμές της παραμέτρου c .



Σχήμα 2.12: Μερίδιο των κόμβων του γραφήματος που βρίσκονται σε κάποιο απομονωμένο δέντρο συναρτήσει της σταθεράς c .

Όπως είναι λογικό, όσο η τιμή της πιθανότητας p μεγαλώνει, τόσο πιο πολλές ακμές αναμένεται να έχει το γράφημα που θα προκύψει. Έτσι για τιμές της p για τις οποίες είναι $N_c = c\nu$ με $c \geq 1/2$ από ένα σημείο και μετά θα έχουμε ένωση αυτών των ξένων δένδρων από τα οποία αποτελείται το γράφημα και στη συνέχεια δημιουργία κύκλων σε αυτά. Είναι, λοιπόν, απόλυτα αναμενόμενη η μορφή του γραφήματος (2.6).

Εφαρμογή: Θα επιβεβαιώσουμε το αποτέλεσμα του Θεωρήματος μέσω των γραφημάτων της προηγούμενης Εφαρμογής.

Στην πρώτη περίπτωση, για τιμές της παραμέτρου $c \leq 1/2$, από το Θεώρημα περιμένουμε όλοι οι κόμβοι του γραφήματος να βρίσκονται σε κάποιο απομονωμένο δέντρο, σχεδόν βεβαίως. Από τους γράφους των Σχημάτων 2.5 και 2.5 βλέπουμε ότι αυτό επιβεβαιώνεται.

Για την περίπτωση στην οποία είναι $c > 1/2$, περιμένουμε το μερίδιο των κόμβων που βρίσκονται σε απομονωμένα δέντρα να είναι όσο φαίνεται στο Σχήμα 2.6, δηλαδή πολύ πολύ μικρό. Πράγματι, από το Σχήμα 2.5 παρατηρούμε ότι όλοι οι κόμβοι (εκτός από τους 4 απομονωμένους) βρίσκονται σε μια μεγάλη συνεκτική συνιστώσα, με πολλούς κύκλους. Το μερίδιο για το οποίο γίνεται λόγος είναι λοιπόν ίσο με $\frac{4}{100} = 0.04$, δηλαδή όντως πολύ μικρό.

Έτσι, επιβεβαιώθηκε και οπτικά το Θεώρημα για τιμές της παραμέτρου c μεγαλύτερες και μικρότερες από την τιμή στην οποία συμβαίνει μεταβολή φάσης.

◇

Εξετάζοντας το γράφημα για τιμές της p μικρότερες από αυτές που απαιτούνται για το προηγούμενο, ισχύει το ακόλουθο:

Θεώρημα 2.6.2. Για $N(\nu) = o(\nu)$ έχουμε ότι:

$$\lim_{\nu \rightarrow +\infty} P[ER_{\nu, N(\nu)} \text{ is union of disjoint trees}] = 1 \quad (2.10)$$

Απόδειξη. Για να αποτελείται ένα γράφημα μόνο από ξένα μεταξύ τους δέντρα θα πρέπει σίγουρα να περιέχει το πολύ ένα κύκλο. Το πλήθος των «υποψήφιων» γραφημάτων $ER_{n, N}$ που περιέχουν τουλάχιστον ένα κύκλο μπορεί να υπολογιστεί, από θεώρημα των Erdős και Renyi⁴, για κάθε τιμή του μήκους των κύκλων. Έστω T η ιδιότητα ένα γράφημα να είναι ένωση ξένων δέντρων και \bar{T} η συμπληρωματική της. Τότε είναι:

$$P_{\nu, N(\nu)}(\bar{T}) \leq \sum_{k=3}^{\nu} \binom{\nu}{k} (k-1)! \frac{\binom{\binom{\nu}{2} - k}{N-k}}{\binom{\binom{\nu}{2}}{N}}.$$

Επίσης:

$$\binom{\nu}{k} (k-1)! \frac{\binom{\binom{\nu}{2} - k}{N-k}}{\binom{\binom{\nu}{2}}{N}} =$$

⁴Θεώρημα στο [15].

$$\begin{aligned}
&= \frac{\nu!}{(\nu-k)!k!} (k-1)! \frac{\frac{\left(\left(\frac{\nu}{2}\right)-k\right)!}{\left[\left(\frac{\nu}{2}\right)-k-(N-k)\right]!(N-k)!}}{\frac{\left(\frac{\nu}{2}\right)!}{\left[\left(\frac{\nu}{2}\right)-N\right]!N!}} = \\
&= \frac{\nu!}{(\nu-k)!k} \frac{\frac{\left(\left(\frac{\nu}{2}\right)-k\right)!}{\left[\left(\frac{\nu}{2}\right)-N\right]!(N-k)!}}{\frac{\left(\frac{\nu}{2}\right)!}{\left[\left(\frac{\nu}{2}\right)-N\right]!N!}} = \\
&= \frac{\nu! \left(\left(\frac{\nu}{2}\right)-k\right)!N! \left[\left(\frac{\nu}{2}\right)-N\right]!}{k(\nu-k)! \left(\frac{\nu}{2}\right)! \left(\left(\frac{\nu}{2}\right)-N\right)! (N-k)!} \\
&= \frac{\nu! \left(\left(\frac{\nu}{2}\right)-k\right)!N!}{k(\nu-k)! \left(\frac{\nu}{2}\right)! (N-k)!} \quad [\text{Χρήση του τύπου } Stirling, \text{ για } \nu \rightarrow +\infty] \\
&\rightarrow \frac{\left[\left(\frac{\nu}{e}\right)^\nu \sqrt{2\pi\nu} \frac{\nu!}{\left(\frac{\nu}{e}\right)^\nu \sqrt{2\pi\nu}}\right] \left[\left(\frac{N}{e}\right)^N \sqrt{2\pi N} \frac{N!}{\left(\frac{N}{e}\right)^N \sqrt{2\pi N}}\right]}{k \left[\left(\frac{\nu-k}{e}\right)^{\nu-k} \sqrt{2\pi(\nu-k)} \frac{(\nu-k)!}{\left(\frac{\nu-k}{e}\right)^{\nu-k} \sqrt{2\pi(\nu-k)}}\right]}.
\end{aligned}$$

$$\begin{aligned}
& \left[\frac{\sqrt{2\pi} \binom{\nu}{2}^{-k} \left(\frac{\binom{\nu}{2}^{-k}}{e} \right) \binom{\nu}{2}^{-k}}{\left(\frac{\binom{\nu}{2}^{-k}}{e} \right) \binom{\nu}{2}^{-k} \sqrt{2\pi} \binom{\nu}{2}^{-k}} \frac{\binom{\nu}{2}^{-k}!}{\left(\frac{\binom{\nu}{2}^{-k}}{e} \right) \binom{\nu}{2}^{-k}} \right] = \\
& \left[\frac{\binom{\nu}{2}!}{\left(\frac{\binom{\nu}{2}}{e} \right) \binom{\nu}{2} \sqrt{2\pi} \binom{\nu}{2}} \left[\frac{\binom{\nu}{2}}{e} \right] \binom{\nu}{2} \sqrt{2\pi} \binom{\nu}{2} \right] \left[\frac{(N-k)! \left(\frac{N-k}{e} \right)^{N-k} \sqrt{2\pi(N-k)}}{\left(\frac{N-k}{e} \right)^{N-k} \sqrt{2\pi(N-k)}} \right] = \\
& = \frac{\left(\frac{\nu}{e} \right)^\nu \sqrt{2\pi} \sqrt{\nu} \left(\frac{N}{e} \right)^N \sqrt{2\pi} \sqrt{N} \left(\frac{\binom{\nu}{2}^{-k}}{e} \right) \binom{\nu}{2}^{-k} \sqrt{2\pi} \sqrt{\binom{\nu}{2}^{-k}}}{k \left(\frac{\nu-k}{e} \right)^{\nu-k} \sqrt{2\pi} \sqrt{\nu-k} \left(\frac{\binom{\nu}{2}}{e} \right) \binom{\nu}{2} \sqrt{2\pi} \sqrt{\binom{\nu}{2}} \left(\frac{N-k}{e} \right)^{N-k} \sqrt{2\pi} \sqrt{N-k}} = \\
& = \frac{\nu^{\nu+\frac{1}{2}} N^{N+\frac{1}{2}} \left[\binom{\nu}{2}^{-k} \right] \binom{\nu}{2}^{-k+\frac{1}{2}} e^{\nu-k} \binom{\nu}{2}^{\nu-k} \binom{\nu}{2}^{+N-k}}{k e^{\nu+N} \binom{\nu}{2}^{-k} (\nu-k)^{\nu-k+\frac{1}{2}} \binom{\nu}{2} \binom{\nu}{2}^{+\frac{1}{2}} (N-k)^{N-k+\frac{1}{2}}} =
\end{aligned}$$

$$\begin{aligned}
&= \frac{e^{-k}}{k} \left(\frac{\nu}{\nu-k} \right)^{\nu+\frac{1}{2}} \frac{1}{(\nu-k)^k} \left(\frac{\binom{\nu}{2} - k}{\binom{\nu}{2}} \right)^{\binom{\nu}{2} + \frac{1}{2}} \left[\binom{\nu}{2} - k \right]^{-k} \\
&\quad \cdot \left(\frac{N}{N-k} \right)^{N+\frac{1}{2}} \frac{1}{(N-k)^{-k}} = \\
&= \frac{e^{-k}}{k} \left[\frac{(\nu-k)(N-k)}{\binom{\nu}{2} - k} \right]^k \frac{1}{\left(\frac{\nu-k}{\nu} \right)^{\nu+\frac{1}{2}}} \left(\frac{\binom{\nu}{2} - k}{\binom{\nu}{2}} \right)^{\binom{\nu}{2} + \frac{1}{2}} \frac{1}{\left(\frac{N-k}{N} \right)^{N+\frac{1}{2}}} = \\
&= \frac{e^{-k}}{k} \left[\frac{(\nu-k)(N-k)}{\binom{\nu}{2} - k} \right]^k \frac{1}{\left(1 - \frac{k}{\nu} \right)^{\nu+\frac{1}{2}}} \left(1 - \frac{k}{\binom{\nu}{2}} \right)^{\binom{\nu}{2} + \frac{1}{2}} \frac{1}{\left(1 - \frac{k}{N} \right)^{N+\frac{1}{2}}} = \\
&= \frac{e^{-k}}{k} \left[\frac{(\nu-k)(N-k)}{\binom{\nu}{2} - k} \right]^k \frac{\left(1 - \frac{k}{\binom{\nu}{2}} \right)^{\frac{1}{2}} \left(1 - \frac{k}{\binom{\nu}{2}} \right)^{\binom{\nu}{2}}}{\left(1 - \frac{k}{\nu} \right)^{\frac{1}{2}} \left(1 - \frac{k}{N} \right)^{\frac{1}{2}} \left(1 - \frac{k}{\nu} \right)^{\nu}} \frac{1}{\left(1 - \frac{k}{N} \right)^N}.
\end{aligned}$$

Ισχύει ότι:

$$\left(1 + \frac{C}{n} \right)^n = e^C, \gamma\iota\alpha \frac{C}{n} \xrightarrow{n \rightarrow +\infty} 0.$$

Οπότε, επειδή:

$$N = O(\nu) \xrightarrow{k \leq N} \frac{k}{\nu} \rightarrow 0, \text{ όταν } \nu \rightarrow +\infty$$

και

$$\frac{k}{\binom{\nu}{2}} \rightarrow 0, \text{ όταν } n \rightarrow +\infty,$$

για $\nu \rightarrow +\infty$ έχουμε:

$$\left(1 - \frac{k}{\nu}\right)^\nu \rightarrow e^{-k}$$

και

$$\left(1 - \frac{k}{\binom{\nu}{2}}\right)^{\binom{\nu}{2}} \rightarrow e^{-k}.$$

Έτσι, το προηγούμενο γίνεται:

$$\frac{e^{-k}}{k} \left[\frac{(\nu - k)(N - k)}{\binom{\nu}{2} - k} \right]^k \frac{\left(1 - \frac{k}{\binom{\nu}{2}}\right)^{\frac{1}{2}}}{\left(1 - \frac{k}{\nu}\right)^{\frac{1}{2}} \left(1 - \frac{k}{N}\right)^{\frac{1}{2}}} \frac{e^{-k}}{e^{-k}} \frac{1}{\left(1 - \frac{k}{N}\right)^N}$$

και έτσι τελικά:

$$\frac{e^{-k}}{k} \left(\frac{1 - \frac{k}{\binom{\nu}{2}}}{\left(1 - \frac{k}{\nu}\right) \left(1 - \frac{k}{N}\right)} \right)^{\frac{1}{2}} \left[\frac{(\nu - k)(N - k)}{\binom{\nu}{2} - k} \right]^k \frac{1}{\left(1 - \frac{k}{N}\right)^N}.$$

Επειδή $(a_n)_{n \geq 1} \rightarrow 1$ για $n \rightarrow +\infty$ συνεπάγεται ότι $(a_n)^{\frac{1}{2}} \rightarrow 1$ για $n \rightarrow +\infty$, τελικά έχουμε:

$$\frac{e^{-k}}{k} \left(\frac{1}{1 - \frac{k}{N}} \right)^{\frac{1}{2}} \left[\frac{(\nu - k)(N - k)}{\binom{\nu}{2} - k} \right]^k \frac{1}{\left(1 - \frac{k}{N}\right)^N}.$$

Δηλαδή για την απόδειξη του θεωρήματος θέλουμε:

$$\sum_{k=3}^n \frac{e^{-k}}{k} \left(\frac{1}{1 - \frac{k}{N}} \right)^{\frac{1}{2}} \left[\frac{(\nu - k)(N - k)}{\binom{\nu}{2} - k} \right]^k \frac{1}{\left(1 - \frac{k}{N}\right)^N} = O\left(\frac{N}{\nu}\right).$$

Επειδή στα γραφήματα που θεωρούμε δεν μπορούμε να έχουμε παραπάνω από N ακμές, τελικά:

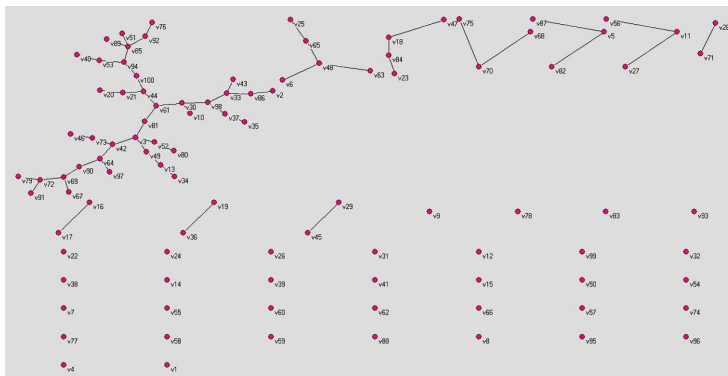
$$\sum_{k=3}^N \frac{e^{-k}}{k} \left(\frac{1}{1 - \frac{k}{N}} \right)^{\frac{1}{2}} \left[\frac{(\nu - k)(N - k)}{\binom{\nu}{2} - k} \right]^k \frac{1}{\left(1 - \frac{k}{N}\right)^N} = O\left(\frac{N}{\nu}\right).$$

Ακόμη, έχουμε ότι $\frac{1}{ke^k} < +\infty$, $\left(1 - \frac{k}{N}\right) < +\infty$, $(\nu - k) = O(\nu)$, $(N - k) = O(N)$, $\binom{\nu}{2} - k = \frac{\nu(\nu-1)}{2} - k = O(\nu^2)$, $\forall k = 3, \dots, N$. Άρα, $(\nu - k)^k =$

$O(\nu^k), (N - k)^k = O(N^k)$ και $\left[\binom{\nu}{2} - k \right]^k = O(\nu^{2k})$ και, τελικά, παίρνουμε το ζητούμενο. \square

Από τα προηγούμενα παρατηρούμε ότι για $N_c = c\nu$ με $0 \leq c \leq 1/2$ η αναμενόμενη τιμή του πλήθους των κόμβων που ανήκουν σε κάποιο δέντρο τείνει στη μονάδα. Δηλαδή, το γράφημα δεν είναι ακόμη συνεκτικό για αυτές τις τιμές της πιθανότητας p και, επιπλέον, κάθε μια από τις συνεκτικές συνιστώσες είναι δέντρο. Από αυτό με τρόπο φυσικό υποψιαζόμαστε ότι για ακόμη μικρότερες τιμές της p το συμπέρασμα αυτό ενισχύεται. Πιο συγκεκριμένα, για τιμές της p τέτοιες ώστε να είναι $N(\nu) = o(n)$ το γράφημα που θα προκύψει θα είναι πιο αραιό από ότι όταν $N_c = c\nu$ με $0 \leq c \leq 1/2$, άρα η πρόταση του προηγούμενου Θεωρήματος είναι κάτι που περιμέναμε να ισχύει με βάση τα προηγούμενα αποτελέσματα που αναφέρθηκαν.

Εφαρμογή: Αυτή τη φορά επιλέγουμε να δημιουργήσουμε έναν τυχαίο ER γράφο μέσω της Διαδικασίας 1, δηλαδή επιλέγοντας την πιθανότητα p με την οποία μια ακμή δημιουργείται ή όχι. Ο γράφος θα έχει $\nu = 100$ κόμβους και έστω ότι θέλουμε να έχει περίπου $\nu^{3/4} = o(\nu)$ ακμές, δηλαδή να είναι $p = \frac{N}{\binom{\nu}{2}} = \frac{\nu^{3/4}}{\nu(\nu-1)/2} = O(\nu^{-5/4}) = 0.011$. Για την τιμή αυτή της πιθανότητας ικανοποιείται η προϋπόθεση να είναι $N = o(\nu)$. Ο γράφος που προκύπτει είναι αυτός του Σχήματος (2.6).



Σχήμα 2.13: Εφαρμογή του παραπάνω Θεωρήματος

Αρχικά παρατηρούμε ότι υπάρχει μια μεγάλη συνεκτική συνιστώσα και ότι αυτή είναι αρκετά μεγαλύτερη από τις υπόλοιπες. Στη συνέχεια παρατηρούμε ότι, όπως περιμέναμε, το γράφημα είναι όντως ένωση ξένων δέντρων. Αποτελείται από ένα μεγάλο δέντρο με 40 κόμβους, κάποια μικρά με το πολύ 5 κόμβους το καθένα και 36 απομονωμένους κόμβους.

\diamond

Κάτι παρόμοιο με το προηγούμενο διατυπώνεται στο ακόλουθο Θεώρημα των Erdős και Renyi:

Θεώρημα 2.6.3. Έστω ER_{ν, N_c} τυχαίος ER γράφος για τον οποίο είναι $N_c(\nu) \sim c\nu$, με $0 < c < 1/2$.

Τότε, για $\nu \rightarrow +\infty$ είναι $P[B] \rightarrow 1$, όπου B το ενδεχόμενο όλες οι συνιστώσες είτε να είναι δένδρα είτε να περιέχουν έναν ακριβώς κύκλο.

Απόδειξη. Έστω $X_{\nu, N}$ το πλήθος των σημείων του $ER_{\nu, N}$ που ανήκουν σε κάποια συνιστώσα που περιέχει περισσότερες ακμές από ότι κόμβους και ταυτόχρονα λιγότερους από $\sqrt{\log \nu}$ κόμβους. Τότε έχουμε ότι:

• Για $k \leq 3$ δεν μπορεί να υπάρξει γράφημα με παραπάνω ακμές από ότι κόμβους.

• Οι τρόποι να επιλεγθούν k κόμβοι από τους ν είναι $\binom{\nu}{k}$.

• Οι τρόποι να επιλέξουμε ή όχι k κόμβους είναι $2^{\binom{k}{2}}$.

• Οι τρόποι να επιλέξουμε ή όχι τις υπόλοιπες $(N - k - 1)$ ακμές που απομένουν στο γράφημα για τους υπόλοιπους $\binom{\nu - k}{2}$ κόμβους είναι $\binom{\binom{\nu - k}{2}}{N - k - 1}$.

Συνδυάζοντας τα παραπάνω και αφού η αναμενόμενη τιμή του ενδεχομένου A που παίρνει τιμές a είναι $E(A) = \sum_{a \in A} aP(a)$, έχουμε:

$$E(X_{\nu, N}) \leq \sum_{k=4}^{\lfloor \sqrt{\log \nu} \rfloor} k \binom{\nu}{2} 2^{\binom{k}{2}} \frac{\binom{\binom{\nu - k}{2}}{N - k - 1}}{\binom{\binom{\nu}{2}}{N}}.$$

Με χρήση του τύπου του Stirling, καταλήγουμε ότι το άθροισμα αυτό είναι $O\left(\nu^{\lfloor \frac{\log 2}{2} - 1 \rfloor}\right)$ και άρα:

$$P(X_{\nu, N} \geq 1) = O\left(\frac{1}{\nu^{1 - \frac{\log 2}{2}}}\right).$$

Δηλαδή δείξαμε ότι με πιθανότητα 1 δεν υπάρχουν κόμβοι που να ανήκουν σε συνιστώσες μεγέθους μικρότερου ή ίσου του $\sqrt{\log \nu}$ και να έχουν περισσότερες ακμές από ότι κόμβους.

Σε ένα από τα θεωρήματα των Erdős και Renyi στο [15] αναφέρεται ότι η πιθανότητα μια συνεκτική συνιστώσα να αποτελείται από πάνω από $\sqrt{\log \nu}$ κόμβους, όταν $N(\nu) \sim c\nu$ με $0 < c < \frac{1}{2}$, τείνει στο 0.

Από τα δύο προηγούμενα, συμπεραίνουμε το θεώρημα. \square

Εφαρμογή: Ένα παράδειγμα γράφου που ικανοποιεί τις απαιτήσεις του Θεωρήματος είναι αυτό του Σχήματος 2.5 της προηγούμενης παραγράφου. Γι' αυτόν είναι $\nu = 100$, $c = 1/4$ και προκύπτει $N = 175$ ακμές. Παρατηρούμε ότι υπάρχουν πολλοί απομονωμένοι κόμβοι και αρκετά μικρά δέντρα, όπως περιμέναμε.

\diamond

Πιο συγκεκριμένα για την ύπαρξη κύκλων σε ER τυχαίο γράφο, ισχύει το:

Θεώρημα 2.6.4. Έστω K η ιδιότητα ενός γράφου να περιέχει τουλάχιστον έναν κύκλο. Έστω ότι είναι $N_c(\nu) \sim c\nu$, με $0 < c < 1/2$.

Τότε, ισχύει ότι:

$$\lim_{\nu \rightarrow \infty} P_{\nu, N_c(\nu)}[K] = 1 - \sqrt{1 - 2ce^{c+e^2}} \quad (2.11)$$

Απόδειξη. Για την απόδειξη του Θεωρήματος αυτού μπορούμε να ανατρέξουμε στο άρθρο των Erdős και Renyi, [15], στην πέμπτη ενότητα. \square

Από το παραπάνω αποτέλεσμα βλέπουμε πως για $c \rightarrow 1/2$ η προηγούμενη σχέση δίνει πιθανότητα ίση με 1 στο όριο. Δηλαδή όλοι οι (μεγάλοι) γράφοι $ER_{\nu}, N(\nu)$ για $c \rightarrow 1/2$ θα περιέχουν έναν κύκλο σχεδόν βεβαίως. Από την ίδια σχέση φαίνεται πως για τιμές της παραμέτρου c μικρότερες από $1/2$ το ίδιο όριο τείνει σε αριθμό μικρότερο της μονάδας και άρα η πιθανότητα ο γράφος που θα προκύψει να περιέχει κύκλο είναι μικρότερη της μονάδας. Αυτό είναι αναμενόμενο, αφού αν για $c = 1/2$ έχουμε οριακά πλήθος ακμών που να μας δίνουν σ.β. γράφο με τουλάχιστον έναν κύκλο, τότε για $c < 1/2$, δηλαδή για μικρότερο αναμενόμενο πλήθος ακμών, η πιθανότητα να έχουμε πάλι τουλάχιστον έναν κύκλο θα είναι μειωμένη.

Η τελευταία Εφαρμογή συμφωνεί με το αποτέλεσμα και αυτού του Θεωρήματος, αφού οριακά έχουμε έναν κύκλο στο γράφημα που δημιουργήθηκε.

Είναι πολύ ενδιαφέρον το ότι για κάποιες τιμές της πιθανότητας p το γράφημα αποτελείται από συνιστώσες που με πιθανότητα 1 είτε είναι δέντρα, είτε περιέχουν ακριβώς έναν κύκλο. Το επόμενο Θεώρημα δίνει πληροφορίες για το μέγεθος της μεγαλύτερης συνεκτικής συνιστώσας ενός τυχαίου ER γράφου όταν $N \sim cn, c > 1/2$.

Θεώρημα 2.6.5. Έστω $\rho_{\nu, N}$ το μέγεθος της μεγαλύτερης συνεκτικής συνιστώσας του τυχαίου ER γραφήματος $ER_{\nu, N}$. Για $N(\nu) \sim cn, c > 1/2$ και $\forall \varepsilon > 0$. Τότε ισχύει ότι:

$$\lim_{\nu \rightarrow +\infty} P\left(\left|\frac{\rho_{\nu, N}}{\nu} - G(c)\right| < \varepsilon\right) = 1, \quad (2.12)$$

όπου $G(c) = 1 - \frac{x(c)}{2c}$, όπου $x(c)$ όπως έχει ήδη οριστεί.

Από την παραπάνω σχέση μπορούμε να καθορίσουμε το αναμενόμενο ποσοστό των κόμβων που ανήκουν στη μεγαλύτερη συνεκτική συνιστώσα για ER γράφους με τα χαρακτηριστικά που αναφέρθηκαν.

Γραφική αναπαράσταση της συνάρτησης G έχουμε στο παρακάτω σχήμα:

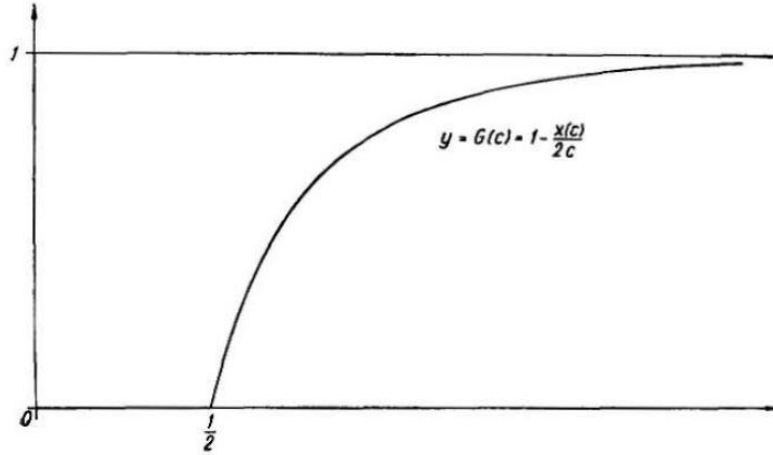
Απόδειξη. Το παραπάνω Θεώρημα προκύπτει σαν Πόρισμα Θεωρήματος των Erdős και Renyi που αναφέρεται στην 9η παράγραφο του [15], για κατάλληλες τιμές των παραμέτρων που εμφανίζονται σε αυτό. \square

Στηριζόμενοι σε μεγάλο βαθμό στα πρηγούμενα, οι Chung και Lu αποδεικνύουν τα ακόλουθα αποτελέσματα.

Θεώρημα 2.6.6. Έστω $1 \leq c \leq \nu p < \log \nu$ για κάποια σταθερά c . Τότε, το μέγεθος των συνιστωσών του γραφήματος $ER_{\nu, p}$, πέραν της μεγάλης συνεκτικής συνιστώσας, είναι σχεδόν βεβαίως το πολύ

$$\frac{\log \nu}{\nu p - 1 - \log(\nu p)}(1 + o(1))$$

Απόδειξη. Η απόδειξη μπορεί να βρεθεί στο [?]. \square



Σχήμα 2.14: Γραφική παράσταση της συνάρτησης G .

Δηλαδή, για κατάλληλες τιμές της παραμέτρου p το μέγεθος των συνιστωσών του γραφήματος πέραν της μεγάλης συνεκτικής συνιστώσας, είναι σχεδόν βεβαίως καθορισμένο.

Θεώρημα 2.6.7. Έστω $\nu p > 3.513$. Τότε η διάμετρος του γραφήματος $ER_{\nu,p}$ είναι ίση με τη διάμετρο της μεγαλύτερης συνεκτικής συνιστώσας.

Λήμμα 2.6.1. Έστω $p > \frac{1}{\nu}$. Τότε, με πιθανότητα τουλάχιστον $1 - o(1)$ ισχύει ότι:

$$|\Gamma_i(x)| \leq 2i^2 \log(\nu(\nu p)^i) \text{ και } |N_i(x)| \leq 2i^3 \log(\nu(\nu p)^i), \forall 1 \leq i \leq \nu,$$

όπου $|\Gamma_i(x)|$ και $|N_i(x)|$ τα σύνολα των κόμβων του γραφήματος σε απόσταση i και μικρότερη ή ίση του i από τον κόμβο x , αντίστοιχα.

Απόδειξη. (Θεωρήματος) Από το προηγούμενο Θεώρημα, η διάμετρος των μικρότερων συνιστωσών του γραφήματος είναι το πολύ $\frac{\log \nu}{\nu p - 1 - \log(\nu p)}(1 + o(1))$.

Από το παραπάνω Λήμμα έχουμε ότι για κάθε κόμβο x του γραφήματός μας, με πιθανότητα τουλάχιστον $1 - o(1)$ ισχύει ότι

$$|N_i(x)| = \sum_{j=1}^i |\Gamma_j(x)| \leq 2i^3 \log(\nu(\nu p)^i).$$

Από αυτό, η διάμετρος τους $ER_{\nu,p}$ είναι τουλάχιστον $\frac{\log \nu}{\log(\nu p)}(1 + o(1))$. Όμως, για $\nu p > 3.513$ είναι $\nu p - 1 - \log(\nu p) > \log(\nu p)$ και, έτσι, καταλήγουμε στο συμπέρασμα του Θεωρήματος. \square

Στο Θεώρημα που ακολουθεί βρίσκονται οι τιμές της διαμέτρου για τις διάφορες τιμές της πιθανότητας p .

Θεώρημα 2.6.8.

Έστω το $ER_{\nu,p}$. Τότε για την τιμή της διαμέτρου είναι:

- Αν $p \geq \frac{c \log \nu}{\nu}$ με $c > 8$, τότε η διάμετρος είναι σχεδόν βεβαίως συγκεντρωμένη γύρω από το πολύ 2 ακέραιες τιμές κοντά στο $\frac{\log \nu}{\log \nu p}$.
- Αν $p \geq \frac{c \log \nu}{\nu}$ με $2 < c < 8$, τότε η διάμετρος είναι σχεδόν βεβαίως συγκεντρωμένη γύρω από το πολύ 3 ακέραιες τιμές κοντά στο $\frac{\log \nu}{\log \nu p}$.
- Αν $p \geq \frac{c \log \nu}{\nu}$ με $c_o < c \leq 1$, τότε η διάμετρος είναι σχεδόν βεβαίως συγκεντρωμένη γύρω από το πολύ $2\lfloor \frac{1}{c} \rfloor + 4$ ακέραιες τιμές κοντά στο $\frac{\log \nu}{\log \nu p}$ και ισχύει ότι

$$\left\lfloor \frac{\log \left(\frac{c\nu}{11} \right)}{\log(\nu p)} \right\rfloor \leq \text{diam}(ER_{\nu,p}) \leq \left\lceil \frac{\log \left(\frac{33c^2}{400} \nu \log \nu \right)}{\log(\nu p)} \right\rceil + 2\lfloor \frac{1}{c} \rfloor + 2.$$

- Αν $\log \nu > \nu p \rightarrow +\infty$, τότε είναι $\text{diam}(ER_{\nu,p}) = \frac{\log \nu}{\log(\nu p)}(1 + o(1))$, σχεδόν βεβαίως.
- Αν $1 < c \leq \nu p$, τότε σχεδόν βεβαίως ισχύει ότι

$$\frac{\log \nu}{\log(\nu p)}(1 + o(1)) \leq \text{diam}(ER_{\nu,p}) \leq \frac{\log \nu}{\log(\nu p)} + 2 \frac{\frac{10c}{(\sqrt{c-1})^2 + 1}}{c - \log(2c)} \frac{\log \nu}{\nu p} + 1.$$

Απόδειξη. Οι αποδείξεις των παραπάνω βρίσκονται στο [?] και στηρίζονται σε μεγάλο βαθμό στις ακόλουθες εύκολα επαληθεύσιμες παρατηρήσεις:

Παρατήρηση 2. Έστω $\varepsilon > 0$ και ακέραιος k τέτοιος ώστε να ικανοποιείται ένα από τα παρακάτω:

- Όταν $ER_{\nu,p}$ είναι συνεκτικός, τότε σχεδόν βεβαίως υπάρχει ένας κόμβος i που ικανοποιεί την $|N_k(i)| < (1 - \varepsilon)\nu$.
- Όταν $ER_{\nu,p}$ δεν είναι συνεκτικός, τότε για κάθε κόμβο i στο γράφημα ισχύει σχεδόν βεβαίως ότι $|N_k(i)| < \nu^{(1-\varepsilon)}$, ή $|N_k(i)|$ μικρότερο από κάθε άλλο κάτω γράφημα για το μέγεθος της μεγάλης συνεκτικής συνιστώσας.

Τότε, ισχύει ότι $\text{diam}(ER_{\nu,p}) > k$.

Παρατήρηση 3. Έστω k_1, k_2 τέτοια ώστε $|\Gamma_{k_1}(i)| |\Gamma_{k_2}(j)| p > (2 + \varepsilon) \log \nu, \forall (i, j) \in ER_{\nu,p}$. Αν $\Gamma_{k_1}(i) \cap \Gamma_{k_2}(j) \neq \emptyset$, τότε $d(i, j) \leq k_1 + k_2$. Αν $\Gamma_{k_1}(i) \cap \Gamma_{k_2}(j) = \emptyset$, τότε η πιθανότητα να υπάρχει ακμή που να συνδέει τους i, j είναι τουλάχιστον

$$1 - (1 - p)^{|\Gamma_{k_1}(i)| |\Gamma_{k_2}(j)|} \geq 1 - e^{-|\Gamma_{k_1}(i)| |\Gamma_{k_2}(j)| p} = 1 - o(\nu^{-2})$$

και, από το γεγονός ότι στο γράφημα υπάρχουν το πολύ ν^2 τέτοια ζεύγη, έχουμε ότι η διάμετρος της μεγαλύτερης συνεκτικής συνιστώσας είναι, σχεδόν βεβαίως, το πολύ $k_1 + k_2 + 1$.

□

2.7 «Το γράφημα»

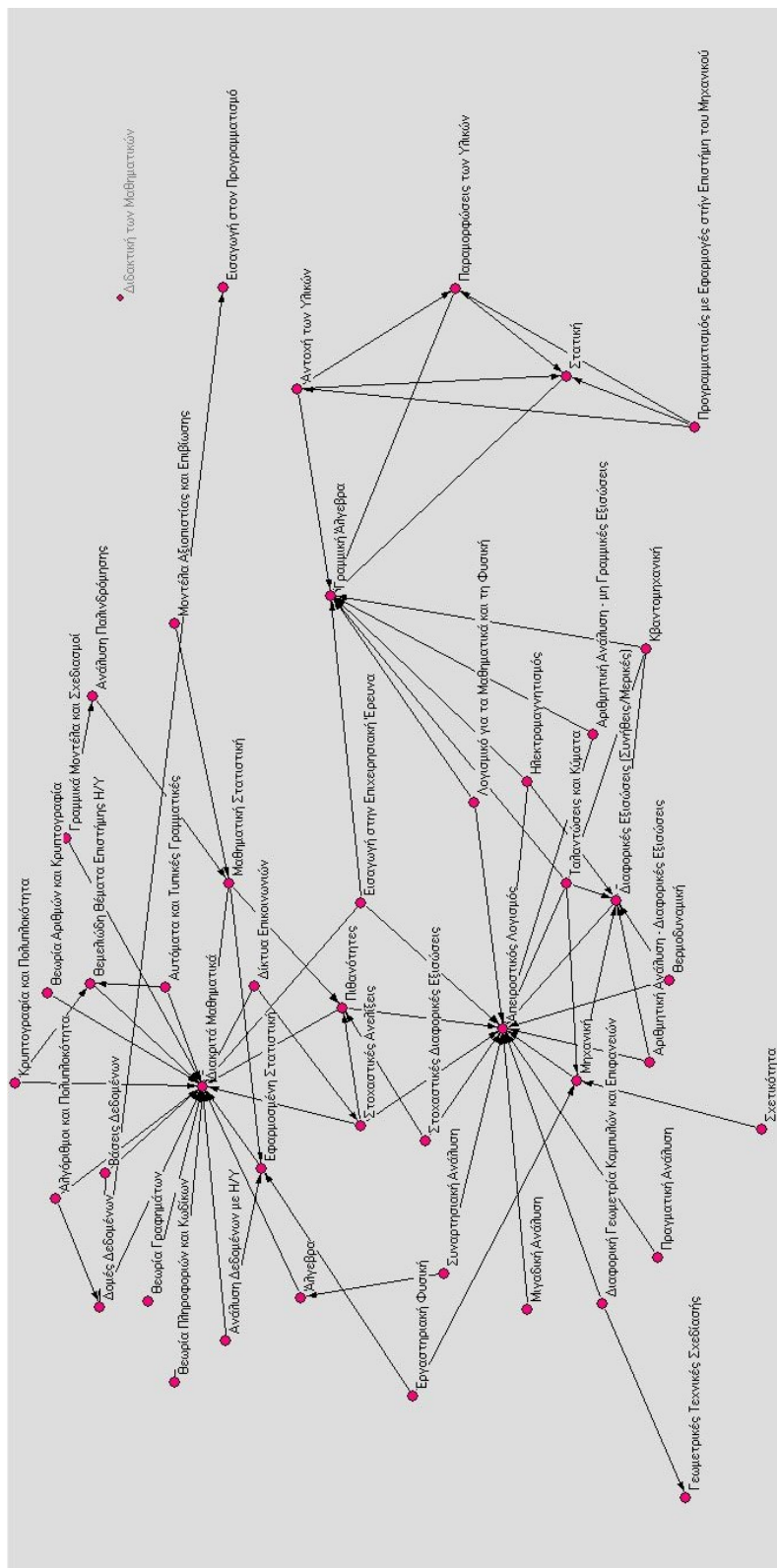
Έστω το γράφημα της επόμενης σελίδας. Στο γράφημα αυτό κόμβοι είναι κάποια από τα μαθήματα της Σχολής ΕΜΦΕ. Είναι τα υποχρεωτικά μαθήματα του κορμού και κάποια από τα μαθήματα που προσφέρονται κυρίως για φοιτητές που ακολουθούν την κατεύθυνση του Μαθηματικού Εφαρμογών, με ροές τη Στατιστική και τα Μαθηματικά Πληροφορικής. Ακμή από τον κόμβο A προς τον κόμβο B υπάρχει όταν το μάθημα A προϋποθέτει τη γνώση του μαθήματος B . Το γράφημα είναι, προφανώς, κατευθυνόμενο. Ο διαχωρισμός των γραφημάτων σε κατευθυνόμενα και μη δεν είναι απλά κάτι τεχνικό, αλλά αντικατοπτρίζει τη σχέση μεταξύ των κόμβων και είναι βασικό διαχωριστικό στοιχείο για την ανάλυση τους.

Ανάμεσα στα πολλά που θα μπορούσαμε να παρατηρήσουμε για τη δομή του, είναι και το γεγονός ότι υπάρχουν κάποιοι κόμβοι για τους οποίους μόνο ο εσωτερικός βαθμός⁵ είναι θετικός, ενώ έχουν εξωτερικό βαθμό⁶ ίσο με μηδέν. Τέτοιοι κόμβοι ονομάζονται «καταβόθρες» (sink) και θα μπορούσαν να χαρακτηριστούν μαθήματα-«βάση» για τα υπόλοιπα. Τέτοια μαθήματα είναι, όπως θα δούμε, ο Απειροστικός Λογισμός, η Γραμμική Άλγεβρα και τα Διακριτά Μαθηματικά. Σε κατευθυνόμενα γραφήματα το in-degree και το out-degree ενός κόμβου μπορεί να διαφέρουν, πράγμα που σε μη κατευθυνόμενα γραφήματα εξ'ορισμού δεν μπορεί να συμβαίνει. Είναι λογικό στο γράφημά μας να υπάρχουν τέτοιοι κόμβοι, μιας και θα πρέπει να υπάρχουν μαθήματα θεμελιώδη, πάνω στα οποία ο φοιτητής θα στηρίζει τη γνώση που θα αποκτήσει μελλοντικά.

Στην επόμενη σελίδα βρίσκεται η γραφική αναπαράσταση του γραφήματος που περιγράφηκε προηγουμένως.

⁵ Εσωτερικός βαθμός του κόμβου i ή $In-degree(i)$: Το πλήθος των ακμών που προσπίπτουν στον κόμβο i .

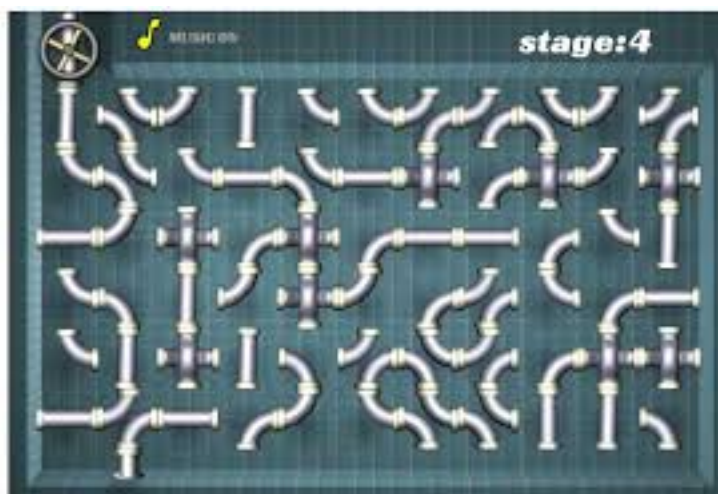
⁶ Εξωτερικός βαθμός του κόμβου i ή $Out-degree(i)$: Το πλήθος των ακμών που ξεκινούν από τον κόμβο i .



Σχήμα 2.15: Γράφημα με μαθήματα της ΣΕΜΦΕ

2.8 Μέτρα Κεντρικότητας

Είναι λογικό να προσπαθούμε να βρούμε τρόπους να μετράμε αποστάσεις στα γραφήματα. Τρόπους και μέτρα που να μας υποδεικνύουν αν υπάρχει σύντομος δρόμος ανάμεσα σε δύο σημεία, ενδεχομένως και ποιος είναι ο συντομότερος. Μέτρα που να δείχνουν αν και πόσο κοντινοί είναι οι κόμβοι ενός γραφήματος ή πόσο γρήγορα και εύκολα μπορεί να διαδοθεί ένα φαινόμενο ανάμεσά τους. Ανάλογα με την περίπτωση, βέβαια, μπορεί και να ευχόμαστε η διάδοση να είναι όσο το δυνατόν πιο αργή και πιο δύσκολη. Πόσο γρήγορα θα καταφέρουμε να διαδώσουμε μια είδηση σε έναν ολόκληρο πληθυσμό; Πώς πρέπει να φτιάξουμε



Σχήμα 2.16: Παιχνίδι

το υδραυλικό αυτό κύκλωμα για να κερδίσουμε στο παιχνίδι ⁷ αν το νερό αρχίσει να ρέει άμεσα; Τότε οι πολλές διακλαδώσεις δεν είναι επιθυμητές, διότι θα χρειαστεί να αντιμετωπιστούν πολλά «μέτωπα»! Ταυτόχρονα, για γραφήματα όπως αυτά που μπορεί να σχετίζονται με τη διάδοση κάποιας πληροφορίας, κάποιας ασθένειας σε πληθυσμούς και γενικά για τη διάδοση κάποιου φαινομένου, μας ενδιαφέρει να ξέρουμε και τη «θέση» της «πηγής» του φαινομένου στο γράφημά μας, ή αντίστοιχα, από ποιόν κόμβο θα ξεκινήσει να ρέει το νερό· προφανώς είναι κάμποσα αυτά που μας ενδιαφέρουν: Σε πόσους κόμβους μπορούμε να έχουμε πρόσβαση από τον κόμβο-πηγή; Πόσο εύκολα μπορεί κανείς, ξεκινώντας από την πηγή, να φτάσει ακόμα και στον πιο απρόσιτο κόμβο του γραφήματος για να διαδοθεί το φαινόμενο; Χρειάζεται η πηγή να παίξει συχνά το ρόλο του διαμεσολαβητή ανάμεσα σε κόμβους του γραφήματος; Πόσο «Prestige» έχει; Μπορεί να «επιρρεάσει» σημαντικούς κόμβους του γραφήματος; Πότε κάποιος κόμβος λέγεται «σημαντικός»; Απάντηση σε όλα αυτά μπορούμε να λάβουμε κοιτώντας

⁷Το «ταμπλό» αποτελείται από τετράγωνα που βρίσκονται το ένα δίπλα στο άλλο και που μπορούν να περιστραφούν γύρω από το κέντρο τους. Κάθε τετράγωνο περιέχει ένα κομμάτι σωλήνα. Ο παίκτης έχει λίγα δευτερόλεπτα στη διάθεσή του για να περιστρέψει όπως αυτός θεωρεί καλύτερα τα τετράγωνα. Έπειτα από λίγο αρχίζει να ρέει νερό από την αντλία, πάνω αριστερά. Σκοπός είναι το νερό, μέσω των σωλήνων, να βγει από την κάτω δεξιά γωνία του πλέγματος, και μόνο από εκεί.

διάφορα μέτρα που αφορούν στους κόμβους του γραφήματος, ή και σε ολόκληρο το γράφημα, τα λεγόμενα *μέτρα κεντρικότητας*, όπως επίσης και μέτρα για το σύνολο του γραφήματος, όπως για παράδειγμα *μέτρα απόστασης*. Τέτοια είναι η διάμετρος και το μέσο μήκος μονοπατιού.

Υπάρχουν διάφορα χαρακτηριστικά-μέτρα μέσω των οποίων μπορούμε να προσδιορίσουμε τη θέση ενός κόμβου μέσα στο γράφημα: να δούμε πόσο «κεντρικός» είναι και πόσο η συμπεριφορά του, ενδεχομένως, μπορεί να «επηρεάσει» άλλους κόμβους στο γράφημα. Οι πλευρές από τις οποίες κοιτάμε την κεντρικότητα ενός κόμβου είναι τέσσερις: ο *βαθμός* του κόμβου, η *εγγύτητα*, δηλαδή οι αποστάσεις του από τους άλλους κόμβους του γραφήματος, η ιδιότητά του να βρίσκεται *ανάμεσα* στους υπόλοιπους και μια ιδιότητα που σχετίζεται με κάποιο *ιδιοδιάνυσμα* που προκύπτει από τη θέση του και αφορά στο «Prestige» που ο κόμβος έχει στο γράφημα. Τα μέτρα αυτά είναι διαφορετικά μεταξύ τους, αλλά δεν υπάρχει κάποιο που να είναι σημαντικότερο από τα άλλα. Η «κεντρικότητα» ενός κόμβου κρίνεται με βάση το σύνολο των κριτηρίων, ένα από αυτά δεν είναι ικανό από μόνο του να προσδιορίσει τον κεντρικότερο κόμβο: δεν υπάρχει ένα κριτήριο που να είναι καλύτερο/ισχυρότερο από τα υπόλοιπα. Η κεντρικότητα δεν καθορίζεται από το πλήθος των δεσμών, αλλά και από το πού και σε ποιον μπορούμε να οδηγηθούμε μέσω των δεσμών αυτών. Είναι σημαντικό να διαλέξουμε το κατάλληλο κάθε φορά κριτήριο με βάση τις διαφορετικές απαιτήσεις του εκάστοτε προβλήματος, με βάση κάποια συγκεκριμένη ιδιότητα που μπορεί να θέλουμε ή να μην θέλουμε να εντοπιστεί στους κόμβους του γραφήματος.

Στη συνέχεια θα παρουσιαστούν κάποια μέτρα κεντρικότητας και θα υπολογιστούν οι τιμές τους για τους κόμβους του γραφήματος των μαθημάτων που παρουσιάστηκε προηγουμένως. Αυτό θα γίνει με χρήση των πακέτων Rajeck και R. Το γράφημα που θα αναλυθεί είναι κατευθυνόμενο χωρίς βάρη στις ακμές, αλλά οι υπολογισμοί που θα παρουσιαστούν μπορούν εύκολα να γενικευτούν και σε γραφήματα με βάρη στις ακμές.

• Centrality: Degree. Ο βαθμός του κόμβου.

Το μέτρο αυτό «φωτογραφίζει» πόσο πολύ ή λίγο συνδεδεμένος είναι ένας κόμβος με τους υπόλοιπους κόμβους του γραφήματος. Η τιμή του ισούται με το πλήθος των γειτόνων του κόμβου στον οποίο αναφερόμαστε. Μπορούμε να κανονικοποιήσουμε την τιμή του μέτρου αυτού, διαιρώντας με το πλήθος των πιθανών γειτόνων⁸ του, δηλαδή με $n-1$. Τότε, κατά κάποιον τρόπο, βλέπουμε πόσο μεγάλη είναι η «τοπική» γειτονιά του κόμβου, σε σχέση με την «ευρύτερη» γειτονιά, το γράφημα ολόκληρο. Για έναν κόμβο, λοιπόν, έχουμε:

$$Degree_i = \text{πλήθος των γειτόνων του } i.$$

$$NormalizedDegree_i = \frac{\text{Πλήθος των γειτόνων του } i}{\text{Πλήθος όλων των πιθανών γειτόνων του } i}$$

Προφανώς, οι τιμές που μπορεί να πάρει ο βαθμός ενός κόμβου είναι από 0 έως $n-1$, ενώ οι τιμές του κανονικοποιημένου βαθμού είναι στο διάστημα $[0,1]$. Επίσης, σε αντίθεση με τις τιμές του βαθμού που είναι ακέραιες, οι τιμές του κανονικοποιημένου βαθμού είναι πραγματικές.

Ο βαθμός του κόμβου μπορεί να φανεί ιδιαίτερα χρήσιμος ακόμα και στα πιο «ιδιαίτερα» δίκτυα. Για παράδειγμα, σε ένα δίκτυο στο οποίο μελετάται η διάδοση κουτσομπολιών, κάποιος κόμβος με μεγαλύτερο βαθμό από κάποιον άλλον είναι πιο πιθανό να ακούσει κάποιο κουτσομπολιό από τον δεύτερο. Το μέγεθος του βαθμού ενός κόμβου μπορεί να ερμηνευτεί και σαν την δυνατότητα κάποιου κόμβου να αλληλεπιδράσει άμεσα με τους υπόλοιπους κόμβους του γραφήματος.

Το μέτρο αυτό, παρότι μας δίνει κάποια αίσθηση κεντρικότητας, δεν μπορεί να προσδιορίσει τη θέση που έχει ο κόμβος μας στο γράφημα. Για παράδειγμα, μπορεί δύο κόμβοι με τον ίδιο βαθμό να βρίσκονται ο ένας «στην καρδιά» του γραφήματος και ο άλλος, κατά κάποιο τρόπο, κοντά στα άκρα του. Τέτοιοι κόμβοι είναι Πιθανότητες και η Στατική στο γράφημα που μελετάμε: και οι δύο έχουν τιμή για τον κανονικοποιημένο εσωτερικό βαθμό 0.07 αλλά η Στατική φαίνεται πιο «περιθωριακός» κόμβος σε σχέση με τις Πιθανότητες.

Στην επόμενη σελίδα δίνονται σε πίνακα οι 5 πιο κεντρικοί κόμβοι όσων αφορά στους εσωτερικούς και εξωτερικούς, κανονικοποιημένους και μη, βαθμούς των κόμβων του γραφήματος και έπειτα σε γράφημα οι κανονικοποιημένοι. Στα γραφήματα αυτά σε διαφορετικές τιμές του μέτρου αντιστοιχούν διαφορετικά χρώματα.

Από τον εσωτερικό βαθμό ενός κόμβου i φαίνεται πόσα είναι τα μαθήματα που «χρειάζονται» το i . Όσο πιο μεγάλος είναι ο εσωτερικός βαθμός τόσο πιο «θεμελιώδης» μπορούμε να πούμε ότι είναι η γνώση του μαθήματος αυτού για τα υπόλοιπα και άρα τόσο πιο «βασικό» μάθημα μπορεί να θεωρηθεί. Αντίθετα, όσο μεγαλύτερο εξωτερικό βαθμό έχει ένας κόμβος τόσο περισσότερη γνώση είναι καλό να έχει ο φοιτητής και, άρα, τόσο πιο «περίπλοκο» θα μπορούσε να θεωρηθεί το μάθημα αυτό.

Όπως φαίνεται από τον παραπάνω πίνακα, μεγαλύτερο (κανονικοποιημένο) εσωτερικό βαθμό έχουν τα Διακριτά Μαθηματικά (0.40), με μικρή διαφορά δεύτερος είναι ο Απειροστικός Λογισμός (0.36) και μετά με συγκριτικά μεγάλη διαφορά ακολουθούν η Γραμμική Άλγεβρα (0.20) και οι Συνήθειες και Μερικές Διαφορικές Εξισώσεις (0.13).

⁸Όλων των κόμβων του γραφήματος, ακόμα και αν ανήκουν σε διαφορετικές συστητώσεις.

Μάθημα	N-In-Degree	In-Degree
Διακριτά Μαθηματικά	0.400000	18
Απειροστικός Λογισμός	0.355556	16
Γραμμική Άλγεβρα	0.200000	9
Διαφορικές Εξισώσεις (Συνήθειες/Μερικές)	0.133333	6
Μηχανική	0.066667	3
Στατική	0.066667	3
Εφαρμοσμένη Στατιστική	0.066667	3
Πιθανότητες	0.066667	3

Πίνακας 2.1: Εσωτερικοί βαθμοί των κόμβων του "Γραφήματος"

Ακολουθεί ο αντίστοιχος πίνακας για τους εξωτερικούς βαθμούς:

Μάθημα	N-Out-Degree	Out-Degree
Ταλαντώσεις και Κύματα	0.088889	4
Μαθηματική Στατιστική	0.066667	3
Αντοχή των Υλικών	0.066667	3
Στοχαστικές Ανελίζεις	0.066667	3
Ηλεκτρομαγνητισμός	0.066667	3
Κβαντομηχανική	0.066667	3
Προγραμ.: Εφ. στην Επιστήμη του Μηχανικού	0.066667	3
Εισαγωγή στην Επιχειρησιακή Έρευνα	0.066667	3

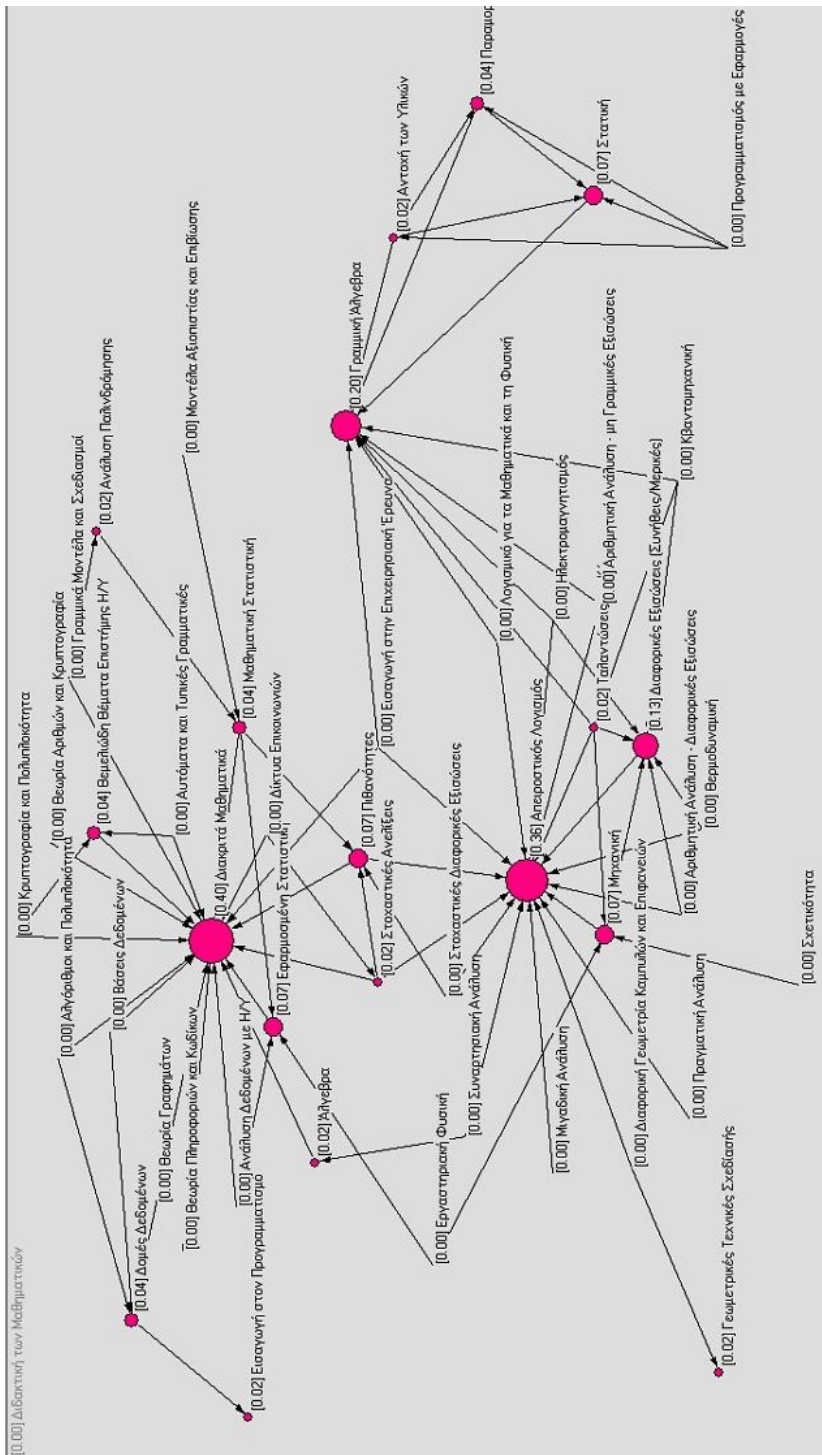
Πίνακας 2.2: Πίνακας με τους εξωτερικούς βαθμούς των κόμβων του "Γραφήματος"

Από αυτόν τον δεύτερο πίνακα μπορούμε να δούμε ότι όλοι οι κόμβοι έχουν αισθητά μικρότερες τιμές για τον εξωτερικό τους βαθμό, συγκριτικά με τους εσωτερικούς βαθμούς. Ο κόμβος "Ταλαντώσεις και Κύματα", με τον μεγαλύτερο εξωτερικό βαθμό, έχει τιμή μόλις 0.09. Ακολουθούν με 0.07 η Μαθηματική Στατιστική, η Αντοχή των υλικών, οι Στοχαστικές Ανελίζεις, ο Ηλεκτρομαγνητισμός, η Κβαντομηχανική, Ο Προγραμματισμός με Εφαρμογές στην Επιστήμη του Μηχανικού και η Εισαγωγή στην Επιχειρησιακή Έρευνα.

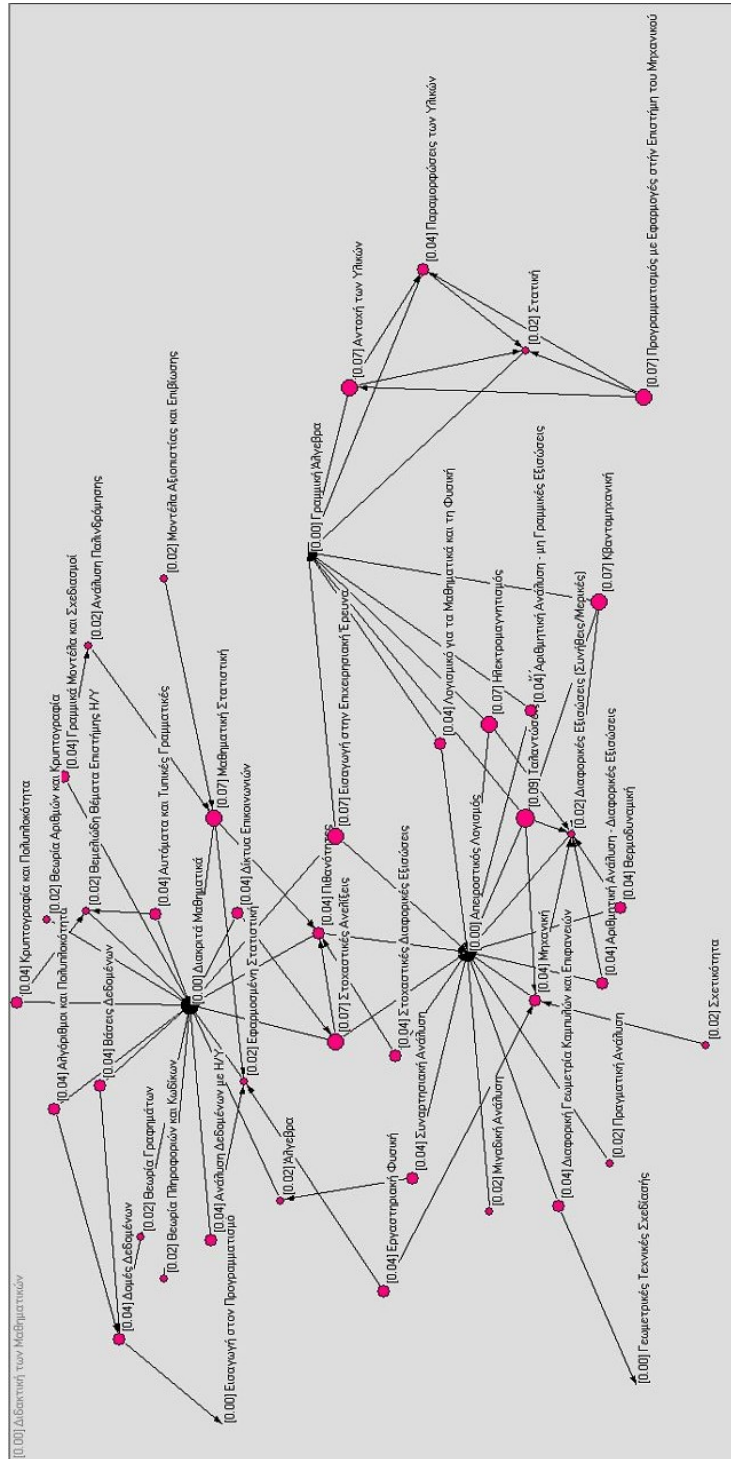
Συγκρίνοντας τις τιμές των δύο μέτρων συνολικότερα, κυρίως μέσω των μη κανονικοποιημένων τιμών τους, ίσως μπορούμε να καταλήξουμε στο ότι το πλήθος των μαθημάτων που απαιτούν γνώση από άλλα είναι κατά πολύ μεγαλύτερο από αυτό αυτών που χρησιμεύουν στα υπόλοιπα. Αυτό μπορεί να προκύψει από το γεγονός ότι όλοι οι κόμβοι έχουν εξωτερικό βαθμό από 0 μέχρι 4, ενώ αντίθετα, υπάρχουν 2 κόμβοι με μεγάλο εσωτερικό βαθμό (που άρα μαζεύουν μεγάλο μερίδιο των ακμών που εξέρχονται από τους υπόλοιπους) και όλοι οι άλλοι έχουν σχετικά αρκετά μικρότερο εσωτερικό βαθμό.

Ίσως αξίζει να παρατηρήσουμε και ότι οι κόμβοι με μεγάλο εσωτερικό βαθμό έχουν μικρό εξωτερικό: Τα Διακριτά Μαθηματικά, ο Απειροστικός Λογισμός και η Γραμμική Άλγεβρα έχουν εξωτερικό βαθμό ίσο με 0 και οι Συνήθειες και Μερικές Διαφορικές Εξισώσεις έχουν εξωτερικό βαθμό ίσο με 1. Από αυτό το γεγονός, τα μαθήματα αυτά θα μπορούσαμε να τα χαρακτηρίσουμε μαθήματα-«βάση», με τη βοήθεια των οποίων κανείς καταφέρνει να παρακολουθήσει πιο εύκολα τη ροή των υπολοίπων (και χωρίς τη βοήθεια των οποίων μάλλον θα δυσκολευτεί να το καταφέρει).

Στις επόμενες σελίδες ακολουθούν τα γραφήματα των κανονικοποιημένων εσωτερικών και εξωτερικών βαθμών των κόμβων του "Γραφήματος".



Σχήμα 2.17: Κανονικοποιημένοι Εσωτερικοί Βαθμοί των κόμβων του "Γραφήματος"



Σχήμα 2.18: Κανονικοποιημένοι Εξωτερικοί Βαθμοί των κόμβων του "Γραφήματος"

• Centrality: Closeness

Το μέτρο αυτό «φωτογραφίζει» πόσο πολύ ή λίγο «κοντά» είναι ο κόμβος μας στους άλλους κόμβους του γραφήματος. Προτάθηκε από τον Gert Sabidussi περίπου το 1966 και δείχνει την «ευκολία» που έχει κανείς ξεκινώντας από έναν συγκεκριμένο κόμβο i , τον οποίο μελετάμε ως προς Closeness Centrality, να φτάσει στους υπόλοιπους κόμβους j του γραφήματος. Είναι φυσικό πως με βάση την περιγραφή αυτή ενδιαφερόμαστε για τα ελάχιστα μονοπάτια του i από και προς τους υπόλοιπους κόμβους j και, προφανώς, για το μήκος τους. Όσο πιο μικρά είναι τα ελάχιστα μονοπάτια από έναν κόμβο προς τους υπόλοιπους του γραφήματος, σε σχέση με τις τιμές των αντίστοιχων μονοπατιών για άλλους κόμβους, τόσο πιο κεντρικός θεωρείται ότι είναι ως προς αυτό το μέτρο. Η τιμή του μέτρου δίνεται από τον τύπο:

$$Closeness_i = \frac{\nu - 1}{\sum_j I(i, j)},$$

όπου $I(i, j)$ το μήκος του ελάχιστου μονοπατιού από τον κόμβο i στον κόμβο j , δηλαδή ο ελάχιστος αριθμός βημάτων που απαιτείται να κάνουμε για να πάμε από τον i στον j . Προσθέτουμε τα μήκη των ελάχιστων μονοπατιών από τον κόμβο i , του οποίου την κεντρικότητα ως προς closeness μελετάμε, προς κάθε άλλο κόμβο j του γραφήματος. Στη συνέχεια, διαιρώντας το πλήθος των πιθανών γειτόνων του i με το άθροισμα αυτό, εμπλέκουμε κάθε κόμβο του γραφήματος σε μια τιμή που τελικά μας δείχνει πόσο κοντά είναι ο i σε όλους τους υπόλοιπους κόμβους του γραφήματος. Επειδή όσο μεγαλύτερη είναι η τιμή του παρονομαστή τόσο πιο μικρή είναι η τιμή του κλάσματος, από τον ορισμό της κιάλας, η τιμή αυτού του μέτρου φανερώνει πόσο «μακριά» είναι ο προς μελέτη κόμβος από τους υπόλοιπους. Για να μεγαλώσει ο παρονομαστής του κλάσματος, πρέπει να μεγαλώσουν τα μήκη των ελάχιστων μονοπατιών από τον κόμβο μας προς τους άλλους. Άρα, όσο μεγαλύτερη είναι η σχετική απόσταση μεταξύ του προς μελέτη κόμβου και των υπόλοιπων, τόσο πιο μικρή είναι η τιμή του μέτρου αυτού. Μάλιστα, επειδή τα μήκη των ελάχιστων μονοπατιών είναι υψωμένα στην πρώτη δύναμη, η τιμή του μέτρου και του αθροίσματος των αποστάσεων είναι μεγέθη αντιστρόφως ανάλογα. Πράγματι, αν το μήκος όλων των ελάχιστων μονοπατιών διπλασιαστεί, διπλασιάζεται και το άθροισμα των μηκών. Επειδή το ν είναι σταθερά του γραφήματος, το $\nu-1$ παραμένει και αυτό σταθερό και, επομένως, η τιμή του κλάσματος υποδιπλασιάζεται. Για το γράφημά μας, η τιμή του μέτρου αυτού για κάθε κόμβο βρίσκεται στον πίνακα και το γράφημα των επόμενων σελίδων.

Για να υπολογιστεί η τιμή του μέτρου πρέπει να υπάρχουν ελάχιστα μονοπατια ανάμεσα στους κόμβους του γραφήματος. Αυτό αμέσως σημαίνει ότι τα γραφήματα για τα οποία υφίσταται αυτό το μέτρο είναι μόνο τα γραφήματα εκείνα για τα οποία υπάρχει τουλάχιστον ένα μονοπάτι από κάθε κόμβο σε κάθε κόμβο: γραφήματα που είναι, δηλαδή, συνεκτικά. Έστω, όμως, ότι το γράφημα του οποίου τους κόμβους μελετάμε δεν είναι συνεκτικό μπορεί να γίνει κάποια διόρθωση έτσι ώστε να βρεθεί κάποια συνάρτηση που να μοντελοποιεί την ιδέα του μέτρου που προηγουμένως περιγράφηκε;

Μια λύση θα ήταν να υπολογιστεί η τιμή του μέτρου για κάθε συνεκτική συνιστώσα χωριστά. Αυτό, όμως, ίσως δεν είναι κατάλληλο μέτρο γιατί θα έδινε μεγαλύτερη τιμή για κόμβους που ανήκουν σε μικρότερες συνιστώσες, αφού σε γενικές γραμμές θα ήταν πιο «κοντά» στους υπόλοιπους της συνιστώσας τους, πράγμα που θα «αδικούσε» τυχόν «κεντρικότερους» (Σε σχέση με το μέτρο αυτό) κόμβους που όμως ανήκουν σε μεγα-

λύτερη συνιστώσα (και άρα είναι πιο κοντά, ακριβώς λόγω αυτού, σε περισσότερους κόμβους).

Η απόσταση δύο κόμβων ανάμεσα στους οποίους δεν υπάρχει μονοπάτι μπορεί να θεωρηθεί ίση με άπειρο. Έτσι, η απόσταση μεταξύ κόμβων που βρίσκονται σε διαφορετικές συνιστώσες του γραφήματος είναι άπειρη, για γραφήματα μη κατευθυνόμενα. Στον τύπο που δόθηκε προηγουμένως, έχουμε αντιστροφή του αθροίσματος των μηκών των ελάχιστων μονοπατιών, πράγμα που δημιουργεί πρόβλημα στη διαφοροποίηση της τιμής του μέτρου για δύο κόμβους ενός μη συνεκτικού γραφήματος (όταν ένα κοινό τους χαρακτηριστικό είναι ότι ανήκουν σε διαφορετικές συνιστώσες (που μπορεί να έχουν διαφορετικό μέγεθος)). Λαμβάνοντας υπ' όψιν ότι $\lim_{k \rightarrow \infty} \frac{1}{k} = 0$, έχουμε ότι η αρχική σχέση μπορεί να τροποποιηθεί στην άθροιση των αντιστρόφων των αριθμών: έτσι, η Closeness Centrality ενός κόμβου μπορεί να αναπαρασταθεί τελικά από τη σχέση

$$Closeness_i = \sum_j \frac{\nu - 1}{I(i, j)},$$

και η ιδέα που θέλει να ποσοτικοποιηθεί από το μέτρο έχει διατηρηθεί.

Το άθροισμα $\sum_j I(i, j)$ έχει ελάχιστη τιμή ίση με $\nu - 1$, αφού ο i έχει το πολύ $\nu - 1$ γείτονες, σε απόσταση τουλάχιστον 1 τον καθένα⁹. Έτσι, οι τιμές που μπορεί το μέτρο αυτό να πάρει είναι πραγματικοί αριθμοί που κυμαίνονται στο διάστημα $[0, 1]$. Την τιμή 0 το μέτρο τη λαμβάνει όταν ο υπό εξέταση κόμβος είναι απομονωμένος.

Το μέτρο αυτό μπορεί σε ένα ανθρώπινο δίκτυο στο οποίο μελετάμε τη διάδοση μιας πληροφορίας να ερμηνευτεί ως το γεγονός πως κάποιος πιο «κεντρικός» άνθρωπος είναι λογικό να ακούσει νωρίτερα από τους υπόλοιπους ανθρώπους-κόμβους του γραφήματος το σύνολο των πληροφοριών που έχουν διαδοθεί στο γράφημα. Κόμβοι με μεγάλη τιμή για το μέτρο αυτό μπορούν να αλληλεπιδράσουν εύκολότερα από άλλους με τους υπόλοιπους κόμβους του γραφήματος. Είναι μέτρο που, σε αντίθεση με το βαθμό, λαμβάνει υπ' όψιν και τις άμεσες αλλά και τις έμμεσες συνδέσεις που ο κόμβος έχει.

Τα κατευθυνόμενα γραφήματα ενδέχεται να μην είναι ισχυρά συνεκτικά ακόμη και στις συνεκτικές συνιστώσες του αντίστοιχου μη κατευθυνόμενου γραφήματος. Σε αυτά μπορούμε να υπολογίσουμε το μέτρο αυτό θεωρώντας, αντί για όλα τα μονοπάτια που διέρχονται από τον υπό μελέτη κόμβο i , είτε αυτά που εισέρχονται στον i από τους υπόλοιπους κόμβους του γραφήματος (και τις αποστάσεις κάθε άλλου κόμβου j από τον i) -και τότε έχουμε την In-Closeness Centrality-, είτε αυτά που εξέρχονται από τον i , λαμβάνοντας υπ' όψιν τις αποστάσεις του κόμβου i από κάθε κόμβο j του γραφήματος. Στη δεύτερη περίπτωση έχουμε την Out-Closeness Centrality.

Όπως και με τον βαθμό των κόμβων, έτσι και με το μέτρο αυτό μπορούμε να έχουμε κάποια αίσθηση κεντρικότητας, αλλά ούτε αυτό το μέτρο μπορεί να θεωρηθεί απόλυτο κριτήριο κεντρικότητας. Για παράδειγμα, έστω οι κόμβοι "Παραμορφώσεις των Υλικών" και "Δομές Δεδομένων". Και οι δύο αυτοί κόμβοι έχουν τιμή 0.065217 όσον αφορά στα ελάχιστα μονοπάτια που περνώντας από τον προς εξέταση κόμβο "εισερχονται" σε αυτόν (In-Closeness Centrality). Σύμφωνα με το μέτρο αυτό, λοιπόν, θα είναι το ίδιο «κεντρικοί» αυτοί οι δύο κόμβοι. Παρ' όλ' αυτά, από την αρχική αναπαράσταση του «Γραφήματος» φαίνεται ότι οι κόμβοι αυτοί ίσως να μην είναι εντελώς ισότιμοι, πράγμα που μπορούμε να επιβεβαιώσουμε και από τον υπολογισμό κάποιων άλλων μέτρων που θα γίνει παρακάτω.

⁹ Αυτό μπορεί να πραγματοποιηθεί, για παράδειγμα, με μια διάταξη τύπου άστρου, για τον κεντρικό κόμβο.

Μάθημα	In-Closeness Centrality
Διακριτά Μαθηματικά	0.428571
Απειροστικός Λογισμός	0.362319
Γραμμική Άλγεβρα	0.217391
Διαφορικές Εξισώσεις (Συνήθειες/Μερικές)	0.156522
Πιθανότητες	0.101449
Εφαρμοσμένη Στατιστική	0.091304

Πίνακας 2.3: In-Closeness Κεντρικότητα των κόμβων του "Γραφήματος"

Όπως φαίνεται από τον Πίνακα (2.8), μεγαλύτερη τιμή και για το μέτρο αυτό έχουν τα Διακριτά Μαθηματικά (0.43), με μικρή διαφορά δεύτερος είναι ο Απειροστικός Λογισμός (0.36) και μετά με συγκριτικά μεγάλη διαφορά ακολουθούν η Γραμμική Άλγεβρα (0.21) και οι Συνήθειες και Μερικές Διαφορικές Εξισώσεις (0.16).

Ακολουθεί ο αντίστοιχος πίνακας για την Out-Closeness κεντρικότητα:

Μάθημα	Out-Closeness Centrality
Ταλαντώσεις και Κύματα	0.108696
Κβαντομηχανική	0.093168
Μαθηματική Στατιστική	0.086957
Αντοχή των Υλικών	0.086957
Στοχαστικές Ανελιξίες	0.086957
Ηλεκτρομαγνητισμός	0.086957
Προγραμ.: Εφ. στην Επιστήμη του Μηχανικού	0.086957
Εισαγωγή στην Επιχειρησιακή Έρευνα	0.086957
Εργαστηριακή Φυσική	0.086957

Πίνακας 2.4: Out-Closeness Κεντρικότητα των κόμβων του "Γραφήματος"

Από αυτόν, τον δεύτερο, πίνακα μεγαλύτερη τιμή έχουν οι "Ταλαντώσεις και Κύματα" (0.11), ακολουθεί η "Κβαντομηχανική" (0.093) και τα "Μαθηματική Στατιστική", "Αντοχή Υλικών", "Στοχαστικές Ανελιξίες", "Ηλεκτρομαγνητισμός", "Προγραμματισμός με Εφαρμογές στην Επιστήμη του Μηχανικού", "Επιχειρησιακή Έρευνα" και "Εργαστηριακή Φυσική" (0.086). Αυτό που μπορούμε αμέσως να παρατηρήσουμε είναι ότι οι περισσότεροι κόμβοι έχουν πολύ κοντινές τιμές για το μέτρο αυτό, πράγμα που μαρτυρά ότι τα ελάχιστα μονοπάτια που εξέρχονται από τους περισσότερους κόμβους του γραφήματος είναι, μάλλον, παρόμοιων μεγεθών. Στη συνέχεια, μπορούμε να επιβεβαιώσουμε ότι αυτό όντως ισχύει, με τη βοήθεια του «Γραφήματος».

Παρατηρούμε, ακόμη, ότι οι κόμβοι που έχουν τις μεγαλύτερες τιμές για την In-Closeness κεντρικότητα έχουν μηδενικές τιμές για την Out-Closeness κεντρικότητα, και αντίστροφα. Τέτοιοι κόμβοι είναι τα "Διακριτά Μαθηματικά", ο "Απειροστικός Λογισμός" και η "Γραμμική Άλγεβρα". Αυτό δείχνει ότι υπάρχουν πολλά μονοπάτια που καταλήγουν σε αυτούς και καθόλου μονοπάτια που να ξεκινούν από αυτούς.

Εξαίρεση για την παραπάνω παρατήρηση αποτελεί η "Διδακτική των Μαθηματικών" που έχει μηδενική τιμή και για τα δύο αυτά μέτρα. Αυτό μας «προδίδει», κατά κάποιο τρόπο, ότι όλοι οι κόμβοι «εύκολα» ή «δύσκολα» φτάνουν στους υπόλοιπους κόμβους, εκτός από τη Διδακτική. Η «ευκολία» που έχει η Διδακτική να φτάσει στους

άλλους κόμβους, καθώς και αυτή που έχουν οι άλλοι κόμβοι να φτάσουν σε αυτή, είναι 0.0, πράγμα που δείχνει ότι ο παρονομαστής είναι, σε κάθε περίπτωση, πολύ μεγαλύτερος από τον αριθμητή: Ο κόμβος είναι είτε πολύ περιθωριακός, είτε μεμονωμένος. Βλέποντας, όμως, ότι δεν υπάρχουν άλλοι τόσο περιθωριακοί κόμβοι στο γράφημα (με έναν τουλάχιστον εκ των οποίων η Διδακτική θα έπρεπε να συνδέεται σε περίπτωση που ήταν περιθωριακή), καταλήγουμε πως δεν είναι: Η Διδακτική είναι αποκομμένη από τους υπόλοιπους κόμβους του Γραφήματος.

Το γεγονός ότι όλοι οι υπόλοιποι έχουν μη μηδενικές τιμές για τουλάχιστον ένα από τα δύο αυτά μέτρα δείχνει ότι ανήκουν όλοι στην ίδια συνεκτική συνιστώσα¹⁰ του γραφήματος, διότι σε διαφορετική περίπτωση όλοι οι παρονομαστές θα απειρίζονταν και έτσι τα κλάσματα θα μηδενίζοντουσαν. Άρα, από το δεδομένο αυτό θα μπορούσαμε να μαντέψουμε, αν δεν το ξέραμε ήδη, ότι το γράφημα χωρίζεται σε δύο συνιστώσες: στη μία ανήκει η Διδακτική και στην άλλη όλοι οι υπόλοιποι κόμβοι του γραφήματος.

Συγκεκριμένα για το δικό μας γράφημα, το να έχει ένας κόμβος μεγάλη τιμή για την In- ή την Out- Closeness κεντρικότητα σημαίνει ότι ο κόμβος αυτός είτε είναι ο ίδιος «εύκολα», σχετικά, προσβάσιμος από μεγάλο μέρος των υπολοίπων κόμβων, είτε ότι ο ίδιος έχει «εύκολα» πρόσβαση σε μεγάλο μερίδιο των υπολοίπων, αντίστοιχα. Ισοδύναμα, για την In-Closeness ότι για όσα μαθήματα αυτό αποτελεί μάθημα-«βάση». Στο «Γράφημα» μεγάλη τιμή για το μέτρο αυτό σημαίνει πως ο υπό εξέταση κόμβος δεν θα έχει «πολλά» «επίπεδα» γνώσης που καλό είναι κανείς να γνωρίζει πιο πριν.

Στις επόμενες σελίδες ακολουθούν τα γραφήματα των μέτρων που περιγράψαμε μόλις, για τους κόμβους του «Γραφήματος».

• Decay Centrality

Ένας άλλος τρόπος να ποσοτικοποιηθεί η ίδια ιδέα είναι ο ακόλουθος. Έστω μια παράμετρος «Φθοράς» (Decay parameter) δ , $\delta \in (0, 1)$. Θεωρούμε κόστος για τη διάσχιση της κάθε ακμής ίσο με $1 - \delta$ επί του ποσού με το οποίο ξεκινάμε στο κάθε βήμα.

Ισχυριζόμαστε ότι μετά από κ βήματα το ποσό που μας έχει απομείνει είναι ίσο με δ^κ και θα αποδείξουμε με τη χρήση επαγωγής ότι αυτό ισχύει για κάθε Φυσικό αριθμό κ , δηλαδή για όσα βήματα και αν κάνουμε. Πράγματι, στο πρώτο βήμα θα έχουμε $1 * \delta = \delta$, και άρα το 1 ανήκει στο σύνολο των αριθμών που ικανοποιούν τον ισχυρισμό μας. Υποθέτουμε ότι ο αριθμός κ τον ικανοποιεί. Επομένως, αν δειχθεί ότι και ο αριθμός $\kappa + 1$ ανήκει στο σύνολο των αριθμών που τον ικανοποιούν, τότε ο ισχυρισμός αποδείχθηκε, αφού το σύνολο που θα προκύψει δεν είναι άλλο από το σύνολο των Φυσικών αριθμών. Πράγματι, μετά το $\kappa + 1$ βήμα θα έχουμε διαθέσιμο ποσό ίσο με

$$\begin{aligned} \pi &= \delta^\kappa - \delta^\kappa(1 - \delta) \\ &= \delta^\kappa(1 - (1 - \delta)) \\ &= \delta^{\kappa+1} \end{aligned}$$

και ο ισχυρισμός αποδείχθηκε.

¹⁰Στην ίδια ασθενώς συνεκτική συνιστώσα του γραφήματος, αφού το γράφημα είναι κατευθυνόμενο.

Η Decay Closeness Centrality του κόμβου i ορίζεται ως το άθροισμα

$$Decay_i = \sum_{i \neq j} \delta^{I(i,j)}.$$

Ορίζουμε τα σύνολα $N_k = \{j \in G : I(i, j) = k\}$. Κάνοντας χρήση των συνόλων αυτών, η Decay Κεντρικότητα ενός κόμβου μπορεί να γραφεί και ως:

$$\begin{aligned} Decay_i &= \sum_{i \neq j} \delta^{I(i,j)} \\ &= \delta^1 |N_1| + \delta^2 |N_2| + \delta^3 |N_3| + \dots \\ &= \sum_{k=1}^{\infty} \delta^k |N_k| \end{aligned}$$

Από τον ορισμό κιάλας φαίνεται ότι η τιμή του μέτρου αυτού είναι συνάρτηση του δ : θα λάβουμε διαφορετική τιμή για το μέτρο εάν αλλάξουμε την τιμή της παραμέτρου. Για παράδειγμα, όταν $\delta \rightarrow 1$, λαμβάνουμε

$$\begin{aligned} Decay_i &= \lim_{\delta \rightarrow 1} \sum_{k=1}^{\infty} \delta^k |N_k| \\ &= \sum_{k=1}^{\infty} |N_k| \lim_{\delta \rightarrow 1} \delta^k \\ &\approx \sum_{k=1}^{\infty} |N_k| \cdot 1 \\ &= \sum_{k=1}^{\infty} |N_k| \\ &= \eta - 1 \end{aligned}$$

Στην περίπτωση αυτή, το μέτρο δίνει τιμή ίση με το μέγεθος της συνιστώσας στην οποία ο κόμβος βρίσκεται, μέγεθος που έχει συμβολιστεί με η . Αντίστοιχα, όταν $\delta \rightarrow 0$, έχουμε

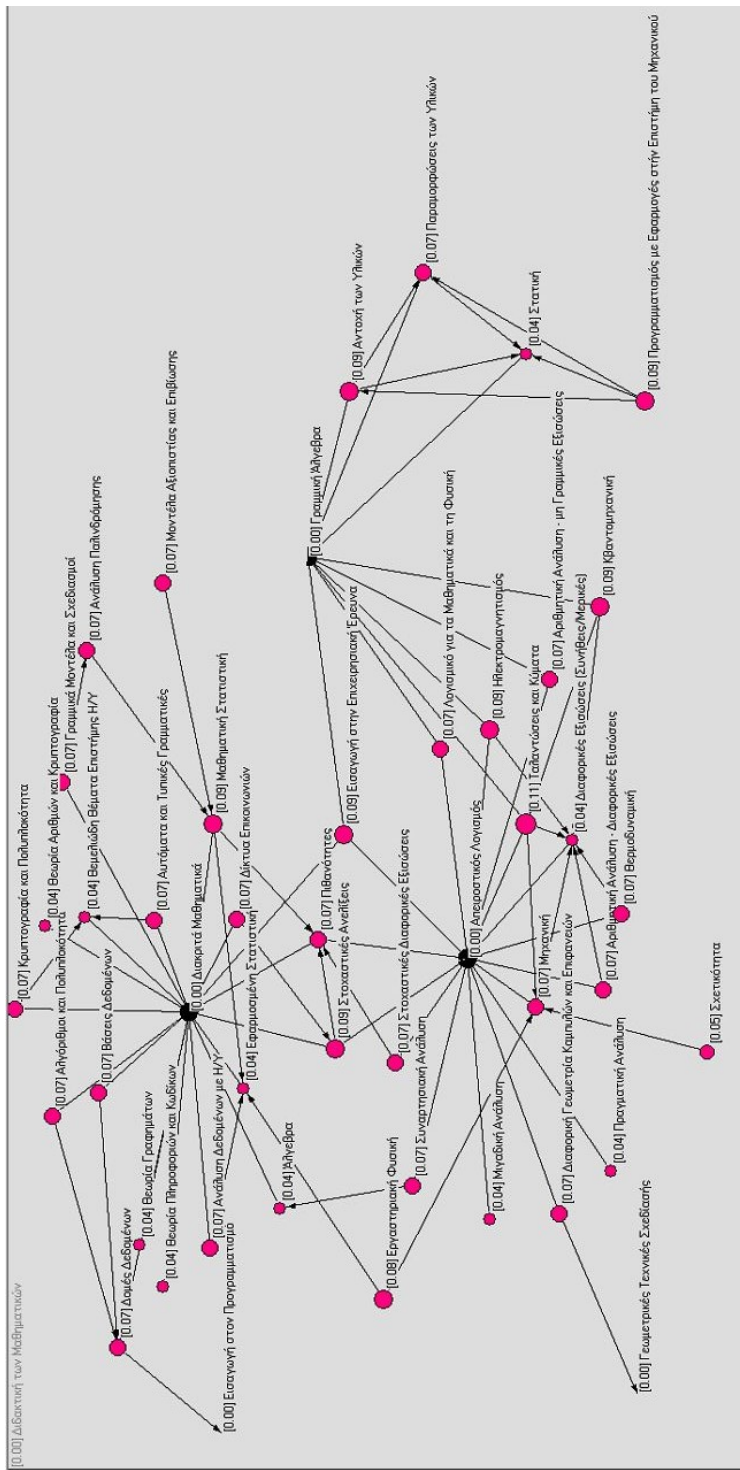
$$\begin{aligned} Decay_i &= \lim_{\delta \rightarrow 0} \sum_{k=1}^{\infty} \delta^k |N_k| \\ &= \lim_{\delta \rightarrow 0} (\delta^1 |N_1| + \delta^2 |N_2| + \delta^3 |N_3| + \dots) \\ &\approx \xi |N_1| + O(\delta^2) \\ &\approx \xi |N_1| \end{aligned}$$

όπου ξ μικρή σταθερά. Καταλήγουμε, έτσι, σε αριθμό ανάλογο του βαθμού του κόμβου, αφού δίνεται πολύ μεγαλύτερο βάρος σε κοντινούς κόμβους από ότι σε μακρυνότερους.

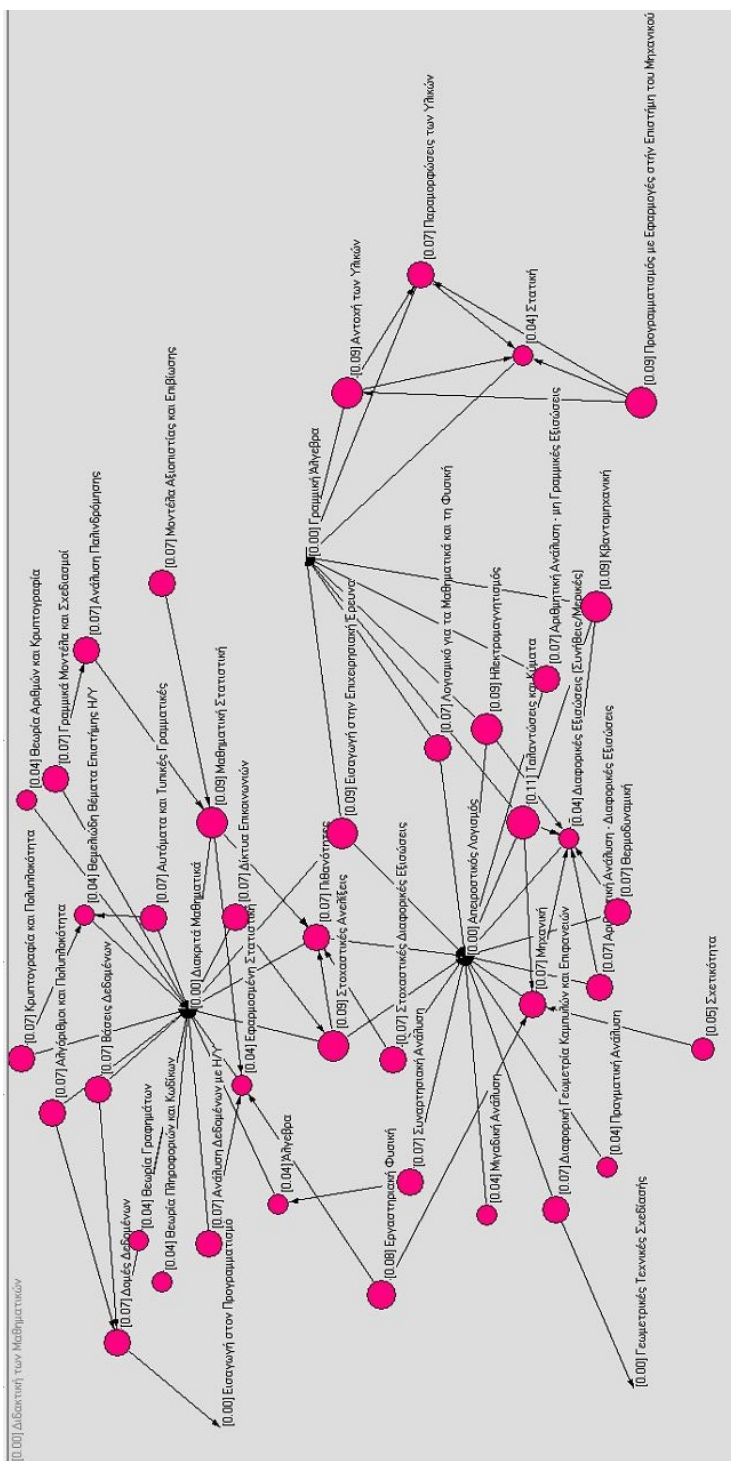
Ξέροντας τη συνάρτηση της Decay κεντρικότητας ενός κόμβου συναρτήσει της παραμέτρου δ , δηλαδή έχοντας μια σχέση της μορφής

$$Decay_i = a\delta^1 + b\delta^2 + c\delta^3 + \dots + n\delta^n$$

είμαστε σε θέση να καταλάβουμε ότι ο κόμβος αυτός έχει a γείτονες, b κόμβους σε απόσταση 2, c σε απόσταση 3 και n σε απόσταση n : η μέγιστη απόσταση οποιουδήποτε κόμβου από τον i είναι n βήματα.



Σχήμα 2.19: Γράφημα για την κανονικοποιημένη (εσωτερική) εγγύτητα των κόμβων του «Γραφήματος».



Σχήμα 2.20: Γράφημα για την κανονικοποιημένη (εξωτερική) εγγύτητα των κόμβων του «Γραφήματος»

• **Centrality: Betweenness**

Το μέτρο αυτό προτάθηκε από τον Freeman το 1977 και μας δείχνει πόσο ο κόμβος μας παίζει το ρόλο του «διαμεσολαβητή», ή της «γέφυρας»¹¹ στο γράφημα. Δείχνει, δηλαδή, πόσο πιθανό είναι να είναι αναγκασμένος κανείς να περάσει από τον k κόμβο εάν περιηγείται στο γράφημα και θέλει να πάει από τον i στον j μέσω κάποιου ελάχιστου μονοπατιού (για κάθε ζευγάρι i, j του γραφήματος!). Ας ορίσουμε το πλήθος των ελάχιστων μονοπατιών ανάμεσα στους κόμβους i και j ως

$$NoI(i, j) = \sum_{\text{paths between } i \text{ and } j} 1_{\{\text{path's length} = I(i, j)\}}$$

Έστω το πλήθος των ελάχιστων αυτών μονοπατιών που περνάνε από τον κόμβο k :

$$NoI(i, j)_k = \sum_{\text{paths between } i \text{ and } j} 1_{\{\text{path's length} = I(i, j)\} \& \text{path lies on } k}$$

Το μέτρο Betweenness ενός κόμβου i ορίζεται ως:

$$Betweenness_i = \frac{\sum_{i \neq j, k} \frac{NoI(k, j)_i}{NoI(i, j)}}{\binom{\nu-1}{2}}$$

και στην ουσία του είναι η αναλογία των ελάχιστων μονοπατιών ανάμεσα στους κόμβους j, k που περνάνε από τον κόμβο i , προς το πλήθος των δυνατών διακριτών ζευγών κόμβων στο γράφο, που είναι $\binom{\nu-1}{2}$.

Αυτό το μέτρο, «φωτογραφίζει» τη θέση του κόμβου i ως διαμεσολαβητή ανάμεσα σε άλλους κόμβους, εννοώντας ότι μας δίνει το μέρος εκείνο των ελάχιστων μονοπατιών του γραφήματος που για να πάνε από τον k κόμβο στον j είναι αναγκασμένα να περάσουν από τον i . Με αυτή την έννοια, ο i μπορεί να έχει κάποιου είδους «εξουσία» πάνω στους k, j . Προφανώς, όσο μεγαλύτερη είναι η τιμή αυτού του μέτρου, τόσο πιο «δυνατός» είναι ο κόμβος i τόσο πιο αναγκαίος μας είναι για την πλοήγησή μας στο γράφημα: στην περίπτωση που για κάποιο λόγο από ένα, συνεκτικό, γράφημα διώξουμε κόμβο με μεγάλη τιμή στο μέτρο αυτό υπάρχει κίνδυνος να πάψει το γράφημα να είναι συνεκτικό. Η ύπαρξη τέτοιων κόμβων είναι κρίσιμη για τη διάδοση διάφορων φαινομένων στο γράφημα.

Το μέτρο αυτό παίρνει τιμές στο διάστημα $[0, \frac{2}{\nu-1}]$. Πράγματι, το κάτω άκρο του διαστήματος επιτυγχάνεται εάν κανένα από τα μονοπάτια που ενώνουν τους κόμβους k, j δεν διέρχεται από τον κόμβο i . Για το άνω άκρο του διαστήματος εξετάζεται η περίπτωση που όλα τα μονοπάτια ανάμεσα στους k, j διέρχονται από τον κόμβο i . Τότε θα είναι $NoI(i, j) = NoI(i, j)_k$ και άρα το κλάσμα $\frac{NoI(k, j)_i}{NoI(i, j)}$ του οποίου οι τιμές

¹¹Ένας κόμβος i του γραφήματος G λέγεται «γέφυρα» εάν το γράφημα $G \setminus \{i\}$ έχει περισσότερες συνιστώσες από το G . Σε κατευθυνόμενα γραφήματα, θεωρούμε ως γέφυρες κόμβους που αν αφαιρεθούν από το γράφημα τότε εκείνο έχει παραπάνω ασθενώς συνεκτικές συνιστώσες.

αθροίζονται στον αριθμητή παίρνει την τιμή 1. Στην περίπτωση αυτή, το άθροισμα γίνεται ίσο με ν , δηλαδή με το πλήθος των κόμβων του γραφήματος. Όσον αφορά στις τιμές του παρονομαστή, είναι $\binom{\nu-1}{2} = \frac{\nu(\nu-1)}{2}$. Έτσι, τελικά, η μέγιστη τιμή που μπορεί να λάβει το μέτρο αυτό είναι $\frac{\nu}{\frac{\nu(\nu-1)}{2}}$, ή $\frac{2}{\nu-1}$.

Από τη σχέση με την οποία το μέτρο αυτό έχει οριστεί, φαίνεται ότι όσο μεγαλώνει το πλήθος των μονοπατιών ανάμεσα στους i, j που διέρχονται από τον k , τόσο μεγαλύτερη είναι η τιμή του αριθμητή, και, από τη στιγμή που το πλήθος των μονοπατιών ανάμεσα στους i, j δεν έχει νόημα να θεωρηθεί μεταβλητό και ο παρονομαστής του κλάσματος είναι σταθερά του γραφήματος, τόσο πιο μεγάλη η τιμή του μέτρου. Δηλαδή, όσο μεγαλύτερη τιμή έχει το μέτρο, τόσο πιο κεντρικός είναι ο κόμβος σε σχέση με το κριτήριο αυτό.

Ακολουθεί πίνακας με τις τιμές του μέτρου αυτού για τους κόμβους του «Γραφήματος».

Μάθημα	Betweenness Centrality
Μαθηματική Στατιστική	0.005556
Πιθανότητες	0.002525
Ανάλυση Παλινδρόμησης	0.002020
Μηχανική	0.002020
Δομές Δεδομένων	0.001010
Στοχαστικές Ανεξίξεις	0.001010

Πίνακας 2.5: Betweenness Κεντρικότητα των κόμβων του "Γραφήματος"

Το μέτρο αυτό, όπως και το πρηγούμενο, ίσως να μην έχει κάποιο ιδιαίτερο νόημα στο συγκεκριμένο γραφημα που μελετάται. Για τη χρησιμότητά του, όμως, μπορούμε να πειστούμε θεωρώντας κάποιο άλλο γράφημα. Έστω ότι οι κόμβοι ήταν πόλεις και οι ακμές δρόμοι ανάμεσα στις πόλεις: δηλαδή, έστω ότι το γράφημα ήταν το οδικό δίκτυο κάποιων ν πόλεων. Τότε με το μέτρο αυτό, όπως περιγράφηκε και προηγουμένως, βλέπουμε κατα πόσο η υπό εξέταση πόλη ανήκει στη διαδρομή ενός τυχόντα ταξιδιώτη που περιπλανιέται στις πόλεις του γραφήματος. Μια τέτοια πληροφορία μπορεί να φανεί ιδιαίτερα χρήσιμη, ειδικά όταν θέλουμε να διαδώσουμε κάποια πληροφορία στους ταξιδιώτες: Θα είναι πολύ περισσότερο αποδοτικό να βάλουμε μια διαφήμιση σε πολλά σημεία μιας πόλης από την οποία ο ταξιδιώτης είναι «αναγκασμένος» να περάσει, από ότι αν την βάζαμε σε πολλά σημεία περισσότερων πόλεων από τις οποίες ο ταξιδιώτης δεν θα περάσει, εκτός αν κατευθύνεται εκεί.

Από τον Πίνακα 2.8, λοιπόν, θεωρώντας ότι οι κόμβοι είναι πόλεις, αν θέλαμε να διαφημίσουμε κάτι, θα κάναμε διάφορες ενέργειες για το σκοπό αυτό κυρίως στις πόλεις "Μαθηματική Στατιστική" και "Πιθανότητες". Οι τιμές του μέτρου αυτού για το γράφημα είναι πάρα πολύ μικρές, πράγμα που οφείλεται στο γεγονός ότι αυτό είναι κατευθυνόμενο, και στην ίδια του τη φύση¹².

Όπως είναι λογικό, κόμβοι που είναι απομονωμένοι έχουν μηδενική τιμή για το μέτρο αφού από τη στιγμή που είναι απομονωμένοι δεν μπορούν σε καμία περίπτωση να

¹²Ότι, δηλαδή, είναι ένα γράφημα που έχει προκύψει από μαθήματα της ΣΕΜΦΕ, με τα κριτήρια που αναφέρθηκαν σε προηγούμενη ενότητα και, άρα, δεν έχει πολύ νόημα η μελέτη του μέτρου αυτού στο συγκεκριμένο γράφημα

δράσουν σαν ενδιάμεσος σταθμός για να πάμε από έναν κόμβο σε κάποιον άλλο. Όπως αναμενόταν, η "Διακριτά Μαθηματικά" έχει και γι' αυτό το μέτρο κεντρικότητας τιμή ίση με μηδέν.

Αναμενόμενο, επίσης, είναι να λάβουμε στην μη-κατευθυνόμενη παραλλαγή του «Γραφήματος» αποτέλεσμα που να καταλήγει στο συμπέρασμα ότι πιο κεντρικοί είναι οι κόμβοι "Διακριτά Μαθηματικά" και "Απειροστικός Λογισμός". Ακολουθεί πίνακας με τους 5 κεντρικότερους, σύμφωνα με το μέτρο αυτό, κόμβους στο γράφημα. Ένα τέτοιο αποτέλεσμα όσων αφορά στα Διακριτά Μαθηματικά και τον Απειροστικό Λογισμό, αν παρατηρήσουμε ότι στο γράφημα αυτό:

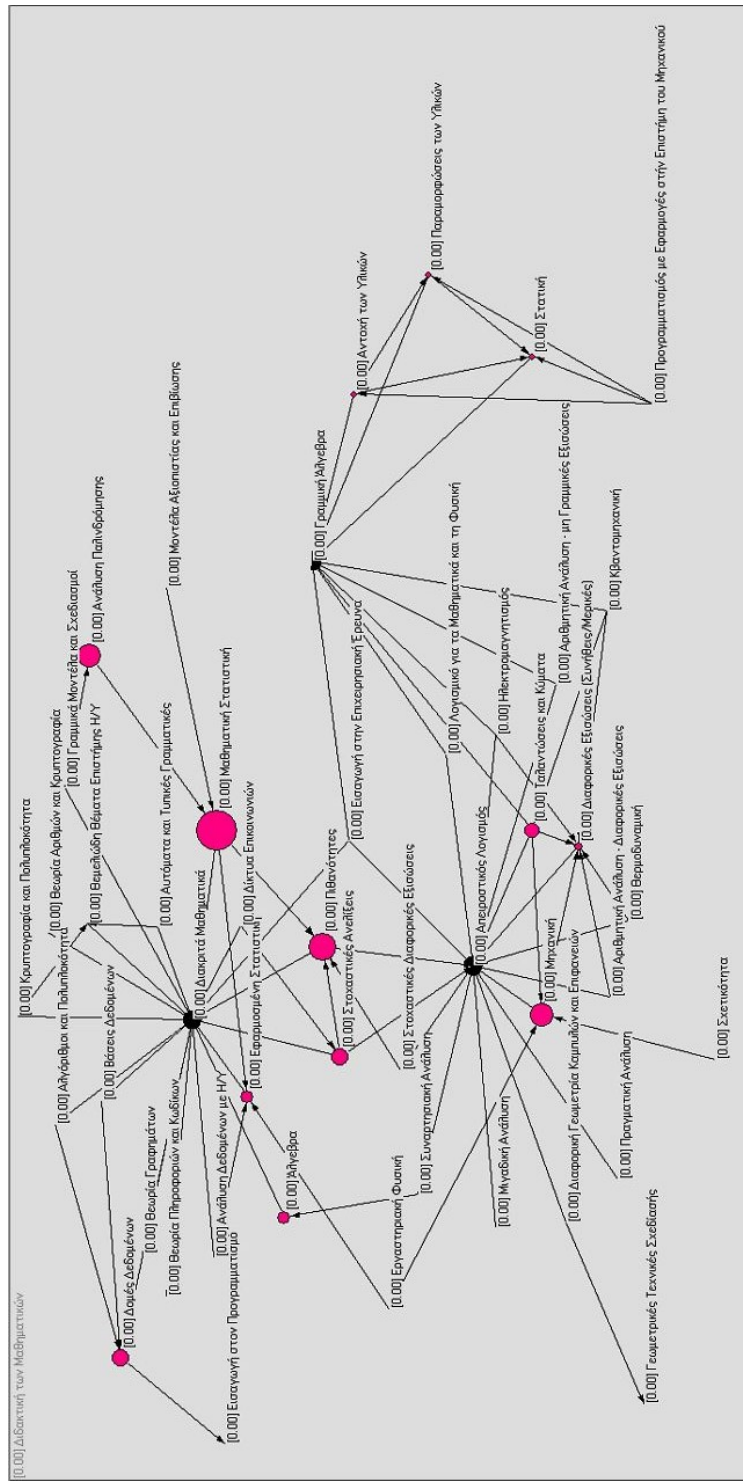
- Τα δύο αυτά μαθήματα είναι συνδεδεμένα με πολύ περισσότερους κόμβους σε σχέση με τα υπόλοιπα, πράγμα που τα καθιστά και πιο «πιθανά» να επιλεγούν ως κόμβοι του ενδεχόμενου μονοπατιού ανάμεσα στον i και το j μάθημα, και
- τα (πολλά) μαθήματα με τα οποία είναι συνδεδεμένα τα Διακριτά και ο Απειροστικός στην πλειοψηφία τους δεν είναι συνδεδεμένα μεταξύ τους: Θα αναγκαστούμε να χρησιμοποιήσουμε τα πρώτα σαν ενδιάμεσους σταθμούς για να πάμε από το ένα στο άλλο.

Μάθημα	Betweenness Centrality
Διακριτά Μαθηματικά	0.532193
Απειροστικός Λογισμός	0.439462
Γραμμική Άλγεβρα	0.208167
Εισαγωγή στην Επιχειρησιακή Έρευνα	0.201394
Πιθανότητες	0.126130

Πίνακας 2.6: Betweenness Κεντρικότητα των κόμβων της μη κατευθυνόμενης παραλλαγής του "Γραφήματος"

Σαν συμπέρασμα, θα μπορούσαμε να καταλήξουμε στο ότι για να έχει ένας κόμβος μη μηδενική τιμή για το μέτρο αυτό θα πρέπει να έχει πάνω από έναν γείτονες που δεν θα συνδέονται άμεσα μεταξύ τους, για να υπάρχει λόγος να περάσουν (και) από αυτόν. Η τιμή του μέτρου μεγιστοποιείται αν μεγιστοποιηθεί ο αριθμητής, ο οποίος με τη σειρά του μεγιστοποιείται εάν το πλήθος των ελάχιστων μονοπατιών από τον j στον k που περνάνε από τον i κόμβο γίνει ίσο με το πλήθος των ελάχιστων μονοπατιών ανάμεσα στους j, k . Κάτι τέτοιο μπορεί να γίνει μόνο εάν έχουμε δομή «άστρου» και μόνο για τον κεντρικό κόμβο, αφού όλοι οι υπόλοιποι αναγκαστικά θα πρέπει να περάσουν από αυτόν. Προφανώς, οι υπόλοιποι κόμβοι, αφού δεν συνδέονται, έχουν Betweenness Centrality ίση με μηδέν. Τέλος, σε κάθε άλλη περίπτωση η τιμή του μέτρου θα κυμαίνεται κάπου ανάμεσα. Αυτά μπορούμε να τα επιβεβαιώσουμε και από τα σχετικά γραφήματα.

Στις επόμενες σελίδες ακολουθούν τα γραφήματα των μέτρων που περιγράψαμε μόλις, για τους κόμβους του «Γραφήματος».



Σχήμα 2.21: Οι κόμβοι του «Γραφήματος» και η συμμετοχή τους στα γεωδαισιακά μονοπάτια του.

• Centrality: Eigenvector

Πέρα από το να βρούμε άμεσα την κεντρικότητα του κάθε κόμβου, μπορούμε να λάβουμε υπ' όψιν τις κεντρικότητες των γειτόνων του για να καταλήξουμε τελικά στην κεντρικότητα του υπό μελέτη κόμβου. Αυτή η ιδέα εισήχθηκε πρώτα από τον Bonacich ήδη από το 1972 και θεωρεί ότι όσο πιο «σημαντικοί» είναι οι γείτονες ενός κόμβου, τόσο πιο σημαντικός είναι και αυτός. Η δυσκολία στη θεωρία αυτή είναι ότι το μέτρο αυτό είναι αυτοαναφορικό, αφού για τον υπολογισμό του κάθε κόμβου χρειαζόμαστε τον υπολογισμό του μέτρου για τους γείτονές του οι οποίοι με τη σειρά τους χρειάζονται την κεντρικότητα των γειτόνων τους, δηλαδή και του υπό μελέτη κόμβου. Η σημασία των μέτρων που θα παρουσιαστούν φαίνεται κυρίως μέσω εφαρμογών. Έστω, ακόμη, ένα κοινωνικό δίκτυο. Η κεντρικότητα κατά Bonacich ενός συγκεκριμένου κόμβου-ατόμου i μπορεί να ερμηνευθεί σαν τη μέση τιμή των αντίστοιχων κεντρικοτήτων των γειτόνων-φίλων του, πολλαπλασιασμένη επί κάποιον συντελεστή που αντιστοιχεί σε βάρη που μπορούμε να ορίσουμε στο κοινωνικό αυτό δίκτυο.

• **Katz Prestige**

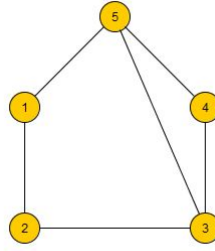
Η τιμή του μέτρου αυτού για τον κόμβο i είναι το άθροισμα της τιμής του μέτρου των γειτόνων του i , διαιρεμένο με τον αντίστοιχο βαθμό του κάθε κόμβου. Έτσι, ο i «μαζεύει» κεντρικότητα από την κεντρικότητα των γειτόνων του. Όπως πολλές φορές συμβαίνει και στην κοινωνία μας, έτσι και στα γραφήματα, όσο πιο πολύ κύρος έχουν οι γείτονες του i , τόσο πιο πολύ κύρος θα έχει και ο ίδιος ο i . Η, κατά κάποιο τρόπο, κανονικοποίηση που γίνεται μέσω της διαίρεσης του κύρους των γειτόνων του i με τον αντίστοιχο βαθμό τους μπορεί να ερμηνευθεί σαν μια μικρή «διόρθωση» που κάνουμε, θεωρώντας ότι ο i τελικά εκμεταλεύεται το κύρος του κάθε j γείτονά του στο χρόνο που του αντιστοιχεί, δηλαδή μόνο κατά χρόνο deg_j^{-1} .

Τελικά, η τιμή του μέτρου αυτού για τους κόμβους i του γραφήματος ορίζεται από τη σχέση:

$$Katz_i = \sum_{j \neq i} \frac{Katz_j}{deg_j} 1_{\{j-is-i's-Neighbor\}} \quad (2.13)$$

Η σχέση αυτή είναι αυτοαναφορική και μας επιστρέφει πλήθος εξισώσεων όσο και αγνώστων, δηλαδή n .

Όπως είπαμε προηγουμένως, θέλουμε το κύρος του i να το μοιράσουμε εξίσου στους j γείτονές του. Έστω ότι έχουμε το γράφημα του Σχήματος 2.22. Το γράφημα αυτό έχει πίνακα γειτνίασης τον G :



Σχήμα 2.22: Γράφημα

$$G = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Διαιρώντας κάθε στοιχείο του πίνακα με το βαθμό του αντίστοιχου κόμβου, φτιάχνουμε τον

$$\widehat{g}_{i,j} = \frac{g_{i,j}}{d_j}.$$

Έχουμε:

$$\widehat{G} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix}$$

Παρατηρούμε ότι ο πίνακας γειτνίασης g διαιρώντας με το βαθμό των κόμβων μεταμορφώθηκε στον στοχαστικό πίνακα \widehat{g} . Αυτό μας χαροποιεί ιδιαίτερα, μιας και οι στοχαστικοί πίνακες έχουν μια ιδιοτιμή τους, και μάλιστα τη μεγαλύτερη σε μέτρο, ίση με τη μονάδα ¹³. Έτσι, από Perron-Frobenius ¹⁴ το σύστημα θα λύνεται και μάλιστα σε κάθε ιδιοδιάνυσμα θα έχουμε θετικές συντεταγμένες.

¹³Επιβεβαιώνουμε εύκολα ότι το 1 είναι ιδιοτιμή του ιδιοδιανύσματος $[1 \dots 1]$ για τυχόν στοχαστικό πίνακα αφού, ως τέτοιος, έχει άθροισμα κάθε γραμμής ίσο με τη μονάδα. Θα δειχθεί ότι η ιδιοτιμή αυτή είναι και η μεγαλύτερη, δηλαδή εν τέλει, κάθε ιδιοτιμή του πίνακα θα περιέχεται στο μοναδιαίο δίσκο. Πράγματι, έστω $\lambda > 1$ που αντιστοιχεί στο ιδιοδιάνυσμα v , για το οποίο $gv = \lambda v$. Οι γραμμές του g έχουν όλες μη αρνητικά στοιχεία και έχουν άθροισμα ίσο με τη μονάδα. Έτσι, κάθε στοιχείο του διανύσματος gv θα είναι γραμμικός συνδυασμός των στοιχείων του v και αναγκαστικά μικρότερο από το μέγιστο στοιχείο του (του v). Όμως, αφού το $\lambda > 1$, θα υπάρχει τουλάχιστον ένα στοιχείο του λv που να είναι μεγαλύτερο από το μέγιστο στοιχείο του v . Καταλήξαμε σε άτοπο με μόνη υπόθεση ότι υπάρχει $\lambda > 1$. Άρα, κάθε ιδιοτιμή κάθε στοχαστικού πίνακα περιέχεται στο μοναδιαίο δίσκο.

¹⁴Εστω

- $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_\nu)$ διάνυσμα μη αρνητικών αριθμών,
- $A = [\alpha_{ij}]$ ανάγωγος πίνακας με μη αρνητικά στοιχεία,
- $M_\lambda = \inf\{\mu : \mu \lambda_i > \sum_{j=1}^{\nu} |\alpha_{ij}| \lambda_j, 1 \leq i \leq \nu\}$

Εάν κάθε στοιχείο α_{ij} του A είναι μη αρνητικό, τότε $\min\{M\}$ είναι ιδιοτιμή του A και κάθε ιδιοτιμή του A βρίσκεται εντός και επί του συνόρου του δίσκου $|z| \leq 1$.

Εφαρμόζοντας την ίδια διαδικασία που εφαρμόστηκε για το Γράφημα του Σχήματος 2.22 και σε τυχόν γράφημα, μετασχηματίζουμε τη σχέση με την οποία ορίστηκε η κεντρικότητα κατά Katz (Σχέση 2.13) στην

$$Katz(G) = \widehat{G}Katz(G) \quad (2.14)$$

ή, ισοδύναμα,

$$(I - \widehat{G}Katz(G)) = 0 \quad (2.15)$$

όπου $Katz(G)$ είναι το διάνυσμα με στοιχεία τις κεντρικότητες κατά Katz για κάθε κόμβο του G και I ο μοναδιαίος πίνακας διάστασης v . Μέσω της παραπάνω διαδικασίας ανάχθηκε το αυτοαναφορικό πρόβλημα εύρεσης των κεντρικότητων κατά Katz στο πρόβλημα εύρεσης ιδιοτιμών-ιδιοδιανυσμάτων για το σύστημα της Σχέσης 2.15 που, όπως δείχθηκε στις αντίστοιχες υποσημειώσεις των σελίδων αυτών, έχει λύση.

Το μέτρο αυτό έχει άλλη τιμή και, όπως είναι λογικό, ελαφρώς παραλλαγμένη ερμηνεία για γραφήματα μη κατευθυνόμενα και άλλη για κατευθυνόμενα. Στην περίπτωση των πρώτων, όπου ο εσωτερικός και ο εξωτερικός βαθμός των κόμβων ταυτίζονται, το μέτρο αυτό δίνει τιμή ανάλογη του βαθμού του κόμβου για κάθε κόμβο. Αυτό μπορεί να φανεί εύκολα από τη Σχέση 2.14. Όσον αφορά στην ερμηνεία, μπορούμε να πούμε ότι το μέτρο, μιας και οι ακμές δεν έχουν κατεύθυνση, λαμβάνει υπ' όψιν τη «δύναμη» όλων των γειτόνων του υπό εξέταση κόμβου. Στην περίπτωση κατευθυνόμενου γραφήματος, το μέτρο δίνει τιμή ανάλογη του εσωτερικού ή του εξωτερικού βαθμού του κόμβου, ανάλογα με την κανονικοποίηση που επιλέχθηκε να γίνει στα αρχικά βήματα της διαδικασίας. Έτσι, μπορούμε να πούμε ότι για την περίπτωση αυτή λαμβάνει υπ' όψιν μόνο τους κόμβους-γειτονες από τους οποίους είτε εισέρχεται ακμή, είτε εξέρχεται, δηλαδή εν τέλει, ότι για τον υπολογισμό της τιμής του μέτρου στεκόμαστε μόνο στους γείτονες αυτούς που, κατά κάποιο τρόπο, «μας εμπιστεύονται» ή «τους εμπιστευόμαστε», αντίστοιχα ¹⁵.

• Bonacich Centrality

Υπάρχει και ακόμη ένας άλλος τρόπος να ληφθεί υπ' όψιν η «δύναμη» των γειτόνων ενός κόμβου για τον υπολογισμό της δικής του εισήχθει πάλι από τον Katz, βασισμένος σε ιδέα του Bonacich. Η αρχική ιδέα του τελευταίου θεωρούσε την κεντρικότητα ενός κόμβου ανάλογη με το άθροισμα των κεντρικότητων των γειτόνων του.

Αρχικά μπορούμε να παρατηρήσουμε ότι το αποτέλεσμα του γινομένου του πίνακα γειτνίασης G ενός γραφήματος v κόμβων με το μοναδιαίο διάνυσμα διάστασης v , δίνει ως αποτέλεσμα διάνυσμα διάστασης v . Κάθε συντεταγμένη του διανύσματος αυτού μπορεί να ερμηνευτεί σαν το πλήθος των μονοπατιών μήκους 1 από τον κόμβο που αντιστοιχεί στην εκάστοτε συντεταγμένη του διανύσματος. Ακόμη, όπως είχε αναφερθεί και σε προηγούμενη παράγραφο, για τις δυνάμεις πινάκων: μπορούμε να πούμε ότι στον πίνακα G^k το στοιχείο g_{ij} δείχνει το πλήθος των μονοπατιών μήκους k από τον i κόμβο στον j . Συνδυάζοντας αυτά τα δύο καταλήγουμε στο ότι το διάνυσμα $G^k \mathbf{1}$ δίνει το πλήθος των μονοπατιών μήκους k από τον κόμβο που αντιστοιχεί στην κάθε συντεταγμένη προς το σύνολο των κόμβων του γραφήματος, όπου $\mathbf{1}$ το μοναδιαίο διάνυσμα διάστασης v .

¹⁵Θεωρώντας ότι μια εισερχόμενη ακμή μπορεί να «μετραφραστεί» ως η εμπιστοσύνη που κάποιος μας δείχνει, και η εξερχόμενη ακμή σαν το γεγονός ότι εμπιστευόμαστε εμείς κάποιον.

Επιστρέφοντας στην ιδέα του Katz: Θεωρούμε παράμετρο α , με $0 < \alpha < 1$. Θεωρούμε, ακόμη, ότι βαδίζουμε σε ένα μονοπάτι ξεκινώντας από τον υπό εξέταση κόμβο i , ξεκινώντας με κάποιο διαθέσιμο υποθετικό ποσό ίσο με 1. Θεωρούμε ότι κάθε φορά που επισκεπτόμαστε κάποιον κόμβο επιβραβεύουμε με ποσό ίσο με α επί του ποσού με το οποίο βρισκόμασταν στο προηγούμενο βήμα. Έτσι, για κόμβους που απέχουν από τον κόμβο εκκίνησης απόσταση 1 θα έχουμε επιβράβευση ίση με α , για κόμβους που απέχουν 2 βήματα θα έχουμε επιβράβευση ίση με α^2 και παρόμοια για κόμβους που απέχουν κ βήματα θα έχουμε επιβράβευση ίση με α^κ . Η συνολική επιβράβευση που λαμβάνουμε από τη βόλτα πάνω στο γράφημα ολόκληρο είναι η κεντρικότητα που περιέγραψε ο Katz:

$$\begin{aligned} Katz_i^2 &= Katz_i^2(G, \alpha) \\ &= \alpha G \mathbf{1} + \alpha^2 G^2 \mathbf{1} + \alpha^3 G^3 \mathbf{1} + \dots \\ &= (1 + \alpha G + \alpha^2 G^2 + \dots) \alpha G \mathbf{1} \end{aligned}$$

Λόγω της τελευταίας, έχουμε

$$\alpha G Katz_i^2(G, \alpha) = (\alpha G + \alpha^2 G^2 + \alpha^3 G^3 + \dots) \alpha G \mathbf{1}$$

και άρα ισχύει ότι

$$Katz_i^2(G, \alpha) - \alpha G Katz_i^2(G, \alpha) = \alpha G \mathbf{1}.$$

Έτσι, τελικά καταλήγουμε στην

$$(I - \alpha G) Katz_i^2(G, \alpha) = \alpha G \mathbf{1}. \quad (2.16)$$

Η κεντρικότητα κατά Bonacich είναι μια επέκταση αυτής, με μόνη διαφορά ότι αυτή τη φορά εισάγεται μια καινούρια παράμετρος, β που ερμηνεύεται σαν την «αξία» που έχει ο κόμβος αρχικά, πρωτού ξεκινήσει τον περίπατο που περιγράφηκε προηγουμένως. Η σχέση που εκφράζει το μέτρο αυτό είναι η:

$$Bonacich_i(G, \alpha, \beta) = (I - \alpha G)^{-1} \beta G \mathbf{1}. \quad (2.17)$$

• Ο αλγόριθμος PageRank της Google

Είναι γεγονός πως τα τελευταία χρόνια το ποσό της διαθέσιμης πληροφορίας στον ιστό είναι τεράστιο. Πώς μπορεί όμως ένας χρήστης του διαδικτύου να εκμεταλευτεί όλη αυτή την πληροφορία; Η σημασία της ύπαρξης των μηχανών αναζήτησης είναι προφανής: αν αυτές δεν υπήρχαν, οι χρήστες θα είμασταν καταδικασμένοι να επισκεπτόμαστε μόνο ιστοσελίδες που ήδη γνωρίζουμε, πράγμα που θα περιόριζε κατά πολύ τη διάδοση της πληροφορίας. Η εφεύρεση των μηχανών αναζήτησης ήταν καθοριστικής σημασίας: μέσω αυτών, πλέον, ο ιστός θα μπορούσε να γίνει χώρος στον οποίο να μπορεί ο καθένας σχετικά εύκολα να βρει αυτό που ζητά. Η πληροφορία είναι τόση

πολλή που δεν μπορεί, προφανώς, να βρίσκεται σε κάποια βάση δεδομένων για να ψάξουμε εκεί: οι μηχανές καλούνται να «βγάλουν άκρη» μέσα στο «χάος». Για να γίνει αυτό θα πρέπει να συνδυαστούν αποτελέσματα από έρευνα σε πολλούς τομείς αλλά δεν είναι στους στόχους αυτής της εργασίας να αναλυθεί αυτό. Παρ'ότι μερίδα των μελετητών ισχυρίζεται ότι η χρήση ενός όχι και τόσο καλού αλγορίθμου αναζήτησης σε μεγάλη ποσότητα διαθέσιμης πληροφορίας μπορεί να ισοσταθμίσει, αν όχι να κερδίσει, τη χρήση ενός καλού αλγορίθμου, είναι λογικό να προσπαθούμε να βρούμε «καλούς» αλγορίθμους για τις αναζητήσεις μας.

Ακόμη, όταν αναζητούμε μια πληροφορία μέσω κάποιας μηχανής αναζήτησης, θέλουμε τα αποτελέσματα να είναι όσο το δυνατόν περισσότερο ακριβή και κοντά σε αυτό που ψάχνουμε. Έτσι, πέρα από πρόκληση, ήταν και ανάγκη να υπάρξει κάποιος αλγόριθμος που να μπορεί να διαχειριστεί γρήγορα και αποτελεσματικά τον τεράστιο όγκο πληροφορίας που είναι στον ιστό. Για χρόνια οι μηχανές αναζήτησης έκαναν το καλύτερο δυνατό, αλλά χωρίς αυτό να είναι αρκετά ικανοποιητικό. Τα τελευταία χρόνια εμφανίστηκε η μηχανή αναζήτησης της εταιρίας *Google* που φαίνεται να άλλαξε τα δεδομένα στην αναζήτηση, χάρη στο αλγόριθμο που χρησιμοποιεί.

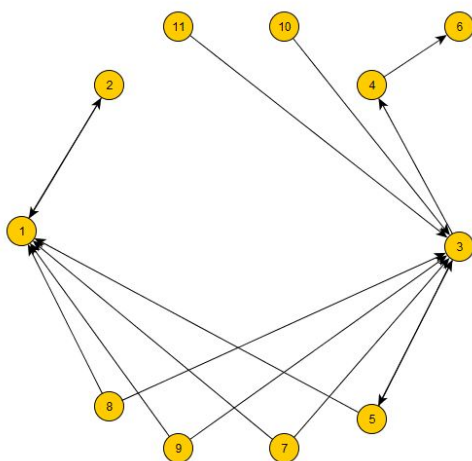
Ο αλγόριθμος αυτός καλείται PageRank και η αντιστοιχία των προηγούμενων με το θέμα μας είναι η ακόλουθη: οι σελίδες του ιστού αντιμετωπίζονται ως κόμβοι σε ένα τεράστιο γράφημα. Το περιεχόμενο των σελίδων είναι κάτι που δεν λαμβάνει υπ'όψιν του ο αλγόριθμος, αυτό με το οποίο ασχολείται είναι οι συσχετίσεις μεταξύ αυτών¹⁶. Μέσω της δομής των διάφορων συνδέσεων ο αλγόριθμος καθορίζει ποιες σελίδες είναι σημαντικότερες από κάποιες άλλες. Ακμή ανάμεσα σε δύο κόμβους του γραφήματος, έστω ακμή από τον *i* προς τον *j*, υπάρχει αν η σελίδα *i* περιέχει σύνδεσμο ("link") προς τη σελίδα *j* και, άρα, το γράφημα αυτό είναι κατευθυνόμενο. Η ύπαρξη συνδέσμου προς μια σελίδα ερμηνεύεται από τον αλγόριθμο ως «εμπιστοσύνη» αυτής προς την πρώτη, κάτι σαν καλή «ψήφος». Είναι λογικό να θεωρούμε πως όσο πιο πολλούς συνδέσμους έχουμε προς την *i* σελίδα, ή ισοδύναμα, όσο πιο πολλές εισερχόμενες ακμές έχει ο κόμβος *i*, τόσο πιο «αξιόπιστος» θα είναι, αφού πολλοί κόμβοι τον εμπιστεύονται. Δυστυχώς, όμως, η εμπιστοσύνη που δείχνει μια σελίδα σε μια άλλη φαίνεται να μην είναι το μόνο κριτήριο για τον αλγόριθμο. Όπως στην πραγματική ζωή η γνώμη των «ειδικών» επί ενός θέματος μετράει περισσότερο από τη γνώμη κάποιων που δεν έχουν τόσο βαθιά γνώση επ' αυτού, και ας είναι περισσότεροι, έτσι και για τον αλγόριθμο η ύπαρξη συνδέσμου προς τη σελίδα *i* σε σελίδες που είναι σημαντικές για το θέμα που πραγματεύεται η *i* είναι πιο σημαντική από την ύπαρξη συνδέσμου σε όχι και τόσο σημαντικές σελίδες, ακόμη και αν οι τελευταίες είναι περισσότερες. Έτσι, ο αλγόριθμος επιπλέον λαμβάνει υπ'όψιν και την αξιοπιστία των σελίδων που εμπιστεύονται την υπό εξέταση σελίδα, δηλαδή, εν τέλει, αν κόμβοι κεντρικοί στο γράφημα έχουν εξερχόμενη ακμή προς τον κόμβο που μελετάται. Το ζητούμενο, λοιπόν, για να γίνει ο *i* κόμβος «σημαντικός» μετατοπίζεται στην ύπαρξη πολλών ακμών που να προσπίπτουν σε αυτόν και που να προέρχονται από «σημαντικούς» κόμβους του γραφήματος. Από την περιγραφή αυτή φαίνεται ότι ο αλγόριθμος ανήκει στην κατηγορία των μέτρων που αναλύθηκαν αμέσως πριν, αφού για να είναι σημαντικός κάποιος θα πρέπει οι γείτονές του να είναι σημαντικοί, πράγμα που καταλήγει να είναι αυτοαναφορικό.

Τρόπος λειτουργίας του αλγορίθμου

Έστω ότι ο τυχόν χρήστης ξεκινάει την περιήγησή του στο διαδίκτυο από την *i* σελίδα, δηλαδή από τον *i* κόμβο του γραφήματος με κόμβους όλες τις υπάρχουσες

¹⁶ Αν και ίσως δεν μπορούμε να αρνηθούμε το γεγονός ότι το βασικότερο ρόλο στην αναζήτηση μιας πληροφορίας παίζουν οι λέξεις στο query του χρήστη προς τη μηχανή αναζήτησης και οι λέξεις των σελίδων του ιστού που τελικά αντιστοιχίζονται με τις λέξεις των queries.

σελίδες. Απλουστεύοντας κατά πολύ την πραγματικότητα, έστω ότι το γράφημα του Σχήματος 2.23.



Σχήμα 2.23: Γράφημα

Το γράφημα αυτό θυμίζει πολύ «το Γράφημα» που έχει γενικότερα μελετηθεί μέχρι τώρα, αφού μπορούμε να δούμε ότι οι κόμβοι 1 και 3 μας θυμίζουν τους κόμβους "Διακριτά Μαθηματικά" και "Απειροστικός Λογισμός" του γραφήματος αυτού. Στο απλουστευμένο παράδειγμα, έστω ότι ο χρήστης ξεκινάει από τον κόμβο 7. Συνεχίζει ακολουθώντας κάποιον από τους συνδέσμους που ο 7 περιέχει, δηλαδή μεταπηδά σε κάποιον από τους γείτονές του. Ας πούμε ότι στη συνέχεια επισκέπτεται τον 3, μετά τον 5, τον 3 ξανά, τον 4 τον 1 και τον 2. Από το σχήμα βλέπουμε ότι αν ο χρήστης ακολουθήσει αυτή τη διαδικασία, μετά τον 2 θα πρέπει να ξαναπάει στον 1 και μετά θα είναι αναγκασμένος να βρεθεί ξανά στον 2, από όπου θα πρέπει να μεταβεί στον 1. Κάπως έτσι, η διαδικασία μετατράπηκε σε έναν κύκλο που στην πραγματικότητα όχι μόνο δεν έχει νόημα να συμβαίνει, αλλά είναι και πολύ διαφορετικός από τη συμπεριφορά του μέσου χρήστη. Αυτό είναι κάτι που πρέπει να ληφθεί υπ' όψιν από τον αλγόριθμο. Για να αποφευχθούν, λοιπόν, τέτοιες αφύσικες και περίπλοκες καταστάσεις θεωρούμε ότι όταν έρθει η στιγμή για το χρήστη να διαλέξει κάποια σελίδα στην οποία θα μεταβεί, διαλέγει

~ με πιθανότητα p κάποιον από τους συνδέσμους που προτίνοται στη σελίδα στην οποία μέχρι τώρα βρισκόταν, και

~ με πιθανότητα $1 - p$ διαλέγει τυχαία έναν κόμβο του γραφήματος (δηλαδή απλά πληκτρολογεί μια καινούρια διεύθυνση url) και μεταβαίνει σε αυτόν.

Σύμφωνα με τα παραπάνω, λοιπόν, υπάρχουν δύο τρόποι να βρεθεί κανείς σε έναν κόμβο: είτε μεταβαίνοντας τυχαία, είτε μεταβαίνοντας μέσω κάποιου συνδέσμου.

Σε αναλογία με όλα αυτά, θεωρώντας ότι έχουμε κάποιον χρήστη που περιηγείται για κάποιο χρονικό διάστημα, μπορούμε να πούμε ότι η τιμή του μέτρου αυτού για τη σελίδα i δείχνει να μπορεί να είναι η πιθανότητα σε μία τυχαία στιγμή ο χρήστης να βρίσκεται στην i σελίδα. Έτσι, μπορούμε να θέσουμε τον περιορισμό το άθροισμα των τιμών του μέτρου για τους κόμβους όλου του γραφήματος να είναι ίσο με τη μο-

νάδα. Συνενώνοντας τις αντίστοιχες τιμές για όλους τους κόμβους του γραφήματος θα φτιάξουμε ένα ιδιοδιάνυσμα του πίνακα γειννίασης του γραφήματος.

Ο αλγόριθμος PageRank (PR) «τρέχει» σε βήματα. Ξεκινάμε αρχικοποιώντας την τιμή του μέτρου για όλους τους κόμβους. Η πιθανότητα να βρεθούμε σε κάποιον κόμβο, διαλέγοντας τυχαία, είναι η ίδια για όλους τους κόμβους. Επομένως, η τιμή του μέτρου θα πρέπει να είναι η ίδια για όλους τους κόμβους. Συγκεκριμένα, θα είναι ίση με $\frac{1}{v}$, όπου v το πλήθος των κόμβων του γραφήματος. Στο προηγούμενο παράδειγμα, η τιμή αυτή είναι 0.091 για κάθε κόμβο.

Όπως ειπώθηκε προηγουμένως, σε μια σελίδα i μπορούμε να βρεθούμε είτε τυχαία, είτε μεταβαίνοντας σε αυτήν από κάποιο σύνδεσμο. Η πιθανότητα να διαλέξει με τυχαίο τρόπο κάποια σελίδα ο χρήστης είναι $1 - p$, άρα, η πιθανότητα να βρεθεί στην i σελίδα τυχαία είναι

$$(1 - p) \frac{1}{v}. \quad (2.18)$$

Αντίστοιχα, η πιθανότητα να βρεθεί κάπου μέσω κάποιου συνδέσμου είναι p . Η πιθανότητα, δεδομένου ότι βρισκόμαστε στον κόμβο j , να βρεθούμε στον κόμβο i είναι $\frac{1}{outDegree(j)}$. Όμως, η πιθανότητα να βρεθούμε στον j είναι $PR(j)$, και άρα, η πιθανότητα να βρεθούμε στον i μέσω κάποιου συνδέσμου παίρνει την ακόλουθη τιμή:

$$p \sum_{j \in G} \frac{PR(j)}{outDegree(j)}. \quad (2.19)$$

Τελικά, η πιθανότητα να βρεθούμε στον i κόμβο είναι το άθροισμα των Σχέσεων (2.18) και (2.19) και έτσι είναι:

$$PR(i) = \frac{1 - p}{v} + p \sum_{j \in G} \frac{PR(j)}{outDegree(j)}. \quad (2.20)$$

Πρόβλημα στην παραπάνω διαδικασία αποτελούν οι κόμβοι-καταβόθρες, δηλαδή αυτοί που έχουν εξωτερικό βαθμό μηδεν. Ο λόγος είναι ότι η ποσότητα του PR που «μάζεψαν» οι κόμβοι αυτοί δεν θα μοιραστεί σε κανέναν άλλον (ακριβώς λόγω της φύσης των κόμβων αυτών: όταν σε κάποιο σύννηθες γράφημα βρεθεί κανείς σε τέτοιο κόμβο, εκεί θα παραμείνει, αφού δεν υπάρχει ακμή που να εξέρχεται από τον κόμβο αυτό.). Λύση στο πρόβλημα μπορεί να δοθεί αν θεωρήσουμε ότι οι καταβόθρες είναι κόμβοι που έχουν εξερχόμενες ακμές προς κάθε κόμβο του γραφήματος. Αυτό, βέβαια, θα γίνει μόνο υπολογιστικά, δεν θα προστεθούν στην πραγματικότητα οι ακμές στο γράφημα που αναπαριστά τον ιστό γιατί κάτι τέτοιο θα κόστιζε υπερβολικά πολύ. Έτσι, τελικά, ίσως μπορούμε να καταλήξουμε στην ακόλουθη σχέση για την περιγραφή της τιμής του PR για κάθε κόμβο:

$$PR(i) = \frac{1 - p}{v} + p \left(\sum_{j \in G} \frac{PR(j)}{outDegree(j)} + \sum_{outDegree(k)=0} \frac{PR(k)}{outDegree(k)} \right) \quad (2.21)$$

Πάντως, σε κάθε επανάληψη το άθροισμα όλων των πιθανοτήτων θα πρέπει να είναι ίσο με ένα, αφού θεωρούμε ότι κάθε χρονική στιγμή ο χρήστης επισκέπτεται κάποια σελίδα. Από τη σχέση (??) μπορούμε να συμπεράνουμε ότι ο κάθε κόμβος j «μοιράζει» ισόποσα το PR του στους κόμβους στους οποίους μπορεί (με ένα βήμα) να μεταβεί.

Από την σχέση υπολογισμού του PR μπορούμε να παρατηρήσουμε ότι:

- Κόμβοι χωρίς εισερχόμενες ακμές θα έχουν σε κάθε επανάληψη ίδια τιμή για το μέτρο αυτό, μιας και ο μόνος τρόπος να βρεθεί κανείς σε αυτές είναι να μεταπηδήσει τυχαία εκεί.

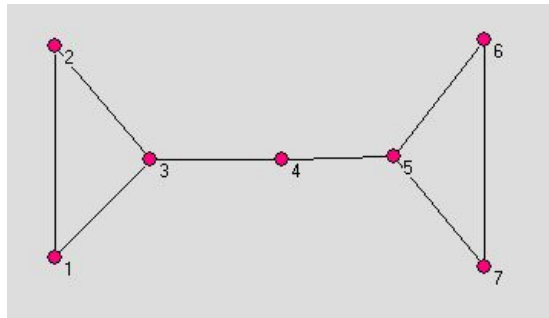
- Κόμβοι που έχουν ίδιο πλήθος εισερχόμενων και από ίδιου μεγέθους κόμβους με ίσο πλήθος εξερχόμενων ακμών θα έχουν την ίδια PR τιμή, αφού πλέον όλοι οι όροι στη σχέση (2.20) θα είναι ίδιοι.
- Μια εισερχόμενη ακμή από έναν σημαντικό κόμβο με λίγες εξερχόμενες ακμές μπορεί να είναι πιο σημαντική από περισσότερες ακμές, όταν αυτές προέρχονται από λιγότερο σημαντικούς κόμβους.

Αποδεικνύεται ότι ο αλγόριθμος αυτός συγκλίνει σε λίγα μόλις βήματα.

2.8.1 Σύγκριση μεταξύ μέτρων.

Στην παράγραφο αυτή θα υπολογιστούν οι τιμές για κάποια από τα μέτρα κεντρικότητας που παρουσιάστηκαν στα προηγούμενα για ένα καινούριο γράφημα. Ακόμη θαδειχθεί ότι ακόμα και σε ένα μικρό γράφημα μπορούμε να έχουμε μεγάλη διαφορά ανάμεσα στα διάφορα μέτρα για κάθε κόμβο. Τέλος, θα επιβεβαιωθεί ότι κάθε μέτρο έχει δική του, ξεχωριστή σημασία και δεν υπάρχει κάποιο που να είναι πάντοτε καλύτερο από τα υπόλοιπα.

Έστω το γράφημα του Σχήματος ().



Σχήμα 2.24: Γράφημα.

Οι τιμές για τα μέτρα του βαθμού, της εγγύτητας και της διαμεσολαβησιμότητας φαίνονται στον ακόλουθο πίνακα:

	Κόμβοι 1,2,6,7	Κόμβοι 3,5	Κόμβος 4
Βαθμός	0,33	0,50	0,33
Εγγύτητα	0,40	0,55	0,60
Διαμεσολαβησιμότητα	0,00	0,53	0,60

Πίνακας 2.7: Μέτρα κεντρικότητας για τους κόμβους του Γραφήματος 2.8.1

Από τα παραπάνω φαίνεται ότι ο κόμβος 4 έχει τη μεγαλύτερη τιμή για το μέτρο της διαμεσολαβησιμότητας και της εγγύτητας. Δηλαδή, είναι αυτός που καλύτερα από όλους δρα σαν γέφυρα ανάμεσα στους κόμβους του γραφήματος και αυτός που, αθροιστικά, είναι πιο κοντά σε όλους τους υπόλοιπους. Οι κόμβοι 1,2,6,7 παρ'ότι έχουν τον ίδιο βαθμό με τον 4, είναι πολύ λιγότερο κεντρικοί. Αυτό διότι είναι αυτοί με το μικρότερο άθροισμα των αντίστροφων των με'γιστων αποστάσεων από τους άλλους και γιατί δεν χρειάζεται ποτέ να περάσει κανείς από αυτούς για να κατευθυνθεί προς οποιονδήποτε άλλον κόμβο του γραφήματος.

2.8.2 Clustering Coefficient

Για κάθε γράφημα G μπορούμε να ορίσουμε και έναν ακόμη συντελεστή:

$$Cl_G = \frac{E(\Delta_G)}{E(W_G)},$$

όπου

$$\Delta_G = \sum_{i,j,k \in G} I_{\{ij, jk, ki \text{ ανήκουν στο γράφημα.}\}} \text{ και } \Delta_G = \sum_{i,j,k \in G} I_{\{ij, jk \text{ ανήκουν στο γράφημα.}\}}.$$

Το μέτρο αυτό, σε αντίθεση με τα προηγούμενα, αφορά ολόκληρο το γράφημα και όχι κάθε κόμβο χωριστά. Από τον ορισμό του, είναι ο λόγος του πλήθους των τριγώνων που υπάρχουν στο γράφημα προς το πλήθος το μονοπατιών μήκους δύο. Μια ερμηνεία που θα μπορούσε να δωθεί σε αυτό το μέτρο είναι, κατά κάποιο τρόπο, το πόσο πιθανό είναι αν δύο κόμβοι είναι συνδεδεμένοι με έναν τρίτο, να είναι και μεταξύ τους συνδεδεμένοι. Αυτό θα είχε περισσότερο ενδιαφέρον αν το δίκτυο αποτελούνταν από ανθρώπους που γνωρίζονται: Εάν δύο άνθρωποι γνωρίζουν έναν τρίτο, πόσο πιθανό είναι να γνωρίζονται και μεταξύ τους;

2.8.3 Τυχαίοι ER γράφοι και κεντρικότητα.

Για το βαθμό. Όπως αναφέρθηκε και σε προηγούμενη ενότητα, για γράφους που δεν είναι πολύ αραιοί, σ.β. έχουμε ότι όλοι οι κόμβοι έχουν τον ίδιο βαθμό. Δηλαδή, δεν υπάρχουν κεντρικότεροι κόμβοι. Για τιμές της πιθανότητας μικρότερες της οριακής, δηλαδή για πιθανότητα που δίνει $N(\nu) \sim o(\nu \log \nu)$, έχουμε πάλι το ίδιο αποτέλεσμα, μιας και η πιθανότητα δημιουργίας των ακμών είναι σταθερή και ίδια για κάθε ακμή που πρόκειται να ενώσει κάθε διαφορετικό ζευγάρι κόμβων.

Για την εγγύτητα και τη διαμεσολαβησιμότητα. Για το μέτρο αυτό, θα πρέπει να δούμε τη δομή του γραφήματος. Ανάλογα με την τιμή της πιθανότητας p , έχουμε γραφήματα που μπορεί να είναι συνεκτικά ή μη συνεκτικά και στα οποία να υπάρχει μια μεγάλη συνεκτική συνιστώσα, να μην υπάρχει καμία ή μόλις να έχει αρχίσει να εμφανίζεται. Ανάλογα με την τιμή της πιθανότητας, λοιπόν, περιμένουμε να έχουν διαφορετική τιμή τα μέτρα αυτά.

Για παράδειγμα, για τιμή της πιθανότητας τέτοια ώστε να υπάρχουν στο γράφημα πολλά μικρά απομονωμένα δένδρα, τα μέτρα αυτά θα παίρνουν πολύ μικρές τιμές. Ο λόγος είναι ότι στο γράφημα θα υπάρχουν πολλοί κόμβοι, αλλά σε κάθε συνεκτική συνιστώσα λίγοι μόνο. Από τον τρόπο που έχουν οριστεί τα δύο αυτά μέτρα, προκύπτει άμεσα ότι

- Οι μέσες αποστάσεις από κάθε κόμβοι της μιας συνιστώσας προς κάθε άλλον κάθε άλλης συνεκτικές συνιστώσας είναι πολύ μεγάλες.
- Οι στιγμές που θα χρειαστεί ένας κόμβος να αποτελέσει γέφυρα ανάμεσα σε δύο άλλους είναι πολύ περιορισμένες, μιας και οι κόμβοι που ανήκουν σε ίδιες συνιστώσες είναι πολύ πολύ λίγοι.

Για τιμή της πιθανότητας τέτοια ώστε να υπάρχει μια μεγάλη συνεκτική συνιστώσα και ίσως κάποιες μικρές, έχουμε ότι οι αποστάσεις ανάμεσα σε κόμβους της μεγάλης συνιστώσας είναι της τάξης της διαμέτρου. Για τους κόμβους που βρίσκονται σε μικρότερες συνιστώσες οι αποστάσεις είναι πιο μικρής τάξης. Παρ'ότι, όμως, οι αποστάσεις για τη δεύτερη περίπτωση είναι μικρότερες, αυτές αφορούν πολύ μικρότερο ποσοστό

των κόμβων στο γράφημα. Έτσι, καταλήγουμε ότι οι τιμές για τα δύο αυτά μετρα είναι μεγαλύτερες για τους κόμβους που ανήκουν στη μεγάλη συνεκτική συνιστώσα και κατά πολύ μικρότερες για τους κόμβους που βρίσκονται στις μικρότερες ή που ίσως είναι απομονωμένοι.

Για τιμή της πιθανότητας τέτοια ώστε να έχουμε μια μεγάλη ενιαία συνεκτική συνιστώσα, με δεδομένο ότι η πιθανότητα δημιουργίας των ακμών είναι σταθερή και ίδια για κάθε ακμή, περιμένουμε πάλι όλοι οι κόμβοι να έχουν περίπου ίδιες τιμές για το μέτρο αυτό.

2.8.4 Clustering Coefficient σε τυχαίους Erdős Rényi γράφους.

Δεν είναι δύσκολο να δούμε ότι σε τυχαίους ER γράφους είναι $Cl_G = p$. Αυτό είναι αρκετά απλό, μιας και αν επιλέξουμε στην τύχη έναν κόμβο i και όλους του τους γείτονες, τότε η πιθανότητα δύο από αυτούς να είναι συνδεδεμένοι είναι p , όση ακριβώς η πιθανότητα δύο τυχαία επιλεγμένοι κόμβοι να είναι συνδεδεμένοι.

Ακόμη, σχετικά με το αποτέλεσμα αυτό, να παρατηρήσουμε ότι για $\lambda = p\nu$ ισχύουν και τα εξής:

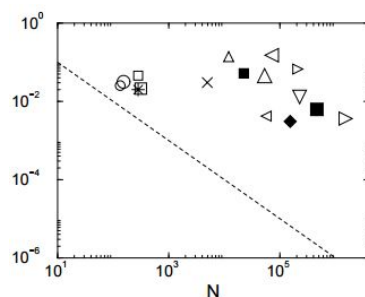
- $\frac{W_G}{\nu} \rightarrow \lambda^2$, κατά πιθανότητα. Η απόδειξη γίνεται με χρήση δεύτερης ροπής και της ανισότητας του Chebyceν.
- $\frac{\nu\Delta_G}{W_G} \rightarrow \frac{3}{\lambda^2} Y$, κατά κατανομή. Η τ.μ. Y ακολουθεί μια Poisson με παράμετρο $\frac{\lambda^3}{6}$.

2.8.5 Πραγματικά δίκτυα και μέτρα κεντρικότητας.

Ενώ οι αποστάσεις σε τυχαία γραφήματα και πραγματικά δίκτυα φαίνεται να ακολουθούν ίδιες συμπεριφορές, για τα μέτρα κεντρικότητας παρατηρούνται κάπως διαφορετικά αποτελέσματα. Στους τυχαίους ER γράφους, όπως είδαμε, δεν υπάρχουν αρκετά πιο κεντρικοί κόμβοι σε σχέση με άλλους. Αντίθετα, σε πραγματικά δίκτυα παρατηρείται ότι υπάρχουν κόμβοι που κυριαρχούν στο γράφημα. Από ότο φαίνεται, αυτό είναι ένα από τα δομικά συστατικά των πραγματικών δικτύων, πράγμα που, όπως θα αναλυθεί σε επόμενη ενότητα, θα πρέπει να ενσωματωθεί σε όποιο μοντέλο προσπαθεί να τα περιγράψει.

2.8.6 Πραγματικά δίκτυα και Clustering Coefficient.

Πάλι σε αντίθεση με αυτό που συμβαίνει για τυχαίους ER γράφους, ο αυτελεστής αυτός είναι διαφορετικός για τις δύο περιπτώσεις. Είδαμε ότι για ER γραφήματα ο αυτελεστής ισούται με την πιθανότητα p , που μπορεί να ποσοτικοποιηθεί και εκ των υστέρων, αν θεωρήσουμε ότι $p = \frac{D_{avg}}{\nu}$, δηλαδή ότι είναι ο λόγος του μέσου βαθμού στο γράφημα δια του πλήθους των κόμβων σε αυτό. Από αυτό, προκύπτει ότι ο λόγος $\frac{Cl_{ER}}{D_{avg}}$ είναι μια ποσότητα που εξαρτάται από τον πλήθος των κόμβων ν και, άρα, είναι διαφορετικός για γραφήματα διαφορετικού μεταξύ τους μεγέθους. Στο γράφημα του Σχήματος (2.8.6), φαίνεται με διακεκομμένη γραμμή ο λόγος στον οποίο μόλις αναφερθήκαμε και τιμές για το λόγο αυτό για διάφορα γραφήματα, τυχαία και πραγματικά. Οι «κουκκίδες» που βρίσκονται στο δεξί κομμάτι της εικόνας αφορούν τυχαίους γράφους. Μπορούμε να δούμε ότι η τιμή για το κλάσμα δεν είναι σταθερή, όπως ακριβώς περιμέναμε. Αντίθετα, για πραγματικά δίκτυα παρατηρήθηκαν οι «κουκκίδες» που βρίσκονται στο αριστερό μέρος της εικόνας. Παρατηρούμε ότι είναι όλες συγκεντρωμένες



Σχήμα 2.25: Σύγκριση μεταξύ του συντελεστή clustering για πραγματικά δίκτυα και τυχαίους ER γράφους.

γύρω από μια τιμή. Το γράφημα αυτό μας υποδεικνύει ότι τα πραγματικά δίκτυα δεν ακολουθούν τον κανόνα των τυχαίων γραφημάτων, καθώς φαίνεται ο λόγος να είναι σταθερός και ανεξάρτητος από το πλήθος ν των κόμβων στο γράφημα.

Όπως φαίνεται και από τον Πίνακα (2.8.6) έχουν παρατηρηθεί τιμές αρκετά διαφο-

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference
WWW, site level, undir.	153, 127	35.21	3.1	3.35	0.1078	0.00023	Adamic 1999
Internet, domain level	3015 - 6209	3.52 - 4.11	3.7 - 3.76	6.36 - 6.18	0.18 - 0.3	0.001	Yook <i>et al.</i> 2001a, Pastor-Satorras <i>et al.</i> 2001
Movie actors	225, 226	61	3.65	2.99	0.79	0.00027	Watts, Strogatz 1998
LANL coauthorship	52, 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman 2001a,b
MEDLINE coauthorship	1, 520, 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman 2001a,b
SPIRES coauthorship	56, 627	173	4.0	2.12	0.726	0.003	Newman 2001a,b,c
NCSTRL coauthorship	11, 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman 2001a,b
Math coauthorship	70, 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási <i>et al.</i> 2001
Neurosci. coauthorship	209, 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási <i>et al.</i> 2001
<i>E. coli</i> , substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner, Fell 2000
<i>E. coli</i> , reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner, Fell 2000
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya, Solé 2000
Silwood park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya, Solé 2000
Words, cooccurrence	460,902	70.13	2.67	3.03	0.437	0.0001	Cancho, Solé 2001
Words, synonyms	22, 311	13.48	4.5	3.84	0.7	0.0006	Yook <i>et al.</i> 2001
Power grid	4, 941	2.67	18.7	12.4	0.08	0.005	Watts, Strogatz 1998
<i>C. Elegans</i>	282	14	2.65	2.25	0.28	0.05	Watts, Strogatz 1998

Πίνακας 2.8: Πίνακας σύγκρισης. Οι στήλες είναι από τα αριστερά προς τα δεξιά: Δίκτυο, Μέγεθος του δικτύου, μέσος βαθμός των κόμβων, μέσο μήκος μονοπατιών, μέσο μήκος μονοπατιών σε τυχαίο ER γράφο ίδιου μεγέθους, συντελεστής Clustering για τα πραγματικά δίκτυα της πρώτης στήλης και μετά για τυχαία ER γραφήματα ίδιας τάξης. Τέλος, στη δεξιότερη στήλη, αναφέρονται οι πηγές από όπου έχουν προκύψει αυτά τα αποτελέσματα. Ο Πίνακας αυτός έχει βρεθεί στο [1]

ρετικές ανάμεσα σε πραγματικά και τυχαία γραφήματα για το συντελεστή του Clustering.

Κεφάλαιο 3

Random Graphs: To Preferential Attachment μοντέλο

3.1 Εισαγωγή

Η ανάλυση των κοινωνικών δικτύων (Social Network Analysis-SNA) συνδιάζει στοιχεία της Θεωρίας Γραφημάτων και της Κοινωνιολογίας. Στόχος είναι η μοντελοποίηση των χαρακτηριστικών των δικτύων που απαρτίζονται από ανθρώπους και ένα μοντέλο που απ'ότι φαίνεται τα καταφέρνει αρκετά καλά είναι αυτό που θα μελετηθεί στη συνέχεια.

Σε αντίθεση με το μοντέλο που περιγράφηκε προηγουμένως, γραφήματα που κατασκευάστηκαν με βάση αυτό της παρούσας ενότητας είναι γραφήματα που δεν έχουν σταθερό μέγεθος. Πιο συγκεκριμένα, θεωρούμε ότι υπάρχει μια μεταβλητή που αντιστοιχεί, κατά κάποιον τρόπο, στο χρόνο και που όλο και αυξάνει. Τα γραφήματα για τα οποία θα γίνει λόγος στην ενότητα αυτή είναι γραφήματα που όσο περνά ο χρόνος γιγαντώνονται.

Υπάρχουν δύο βασικά χαρακτηριστικά που παρατηρούνται σε πραγματικά δίκτυα και που το προηγούμενο μοντέλο, όπως και πολλά άλλα, δεν καταφέρνουν να μοντελοποιήσουν επιτυχώς. Το πρώτο που μπορούμε να παρατηρήσουμε είναι ότι στους Poisson τυχαίους γράφους θεωρούμε ότι το μέγεθος των γράφων που δημιουργούνται είναι σταθερό και εκ των προτέρων καλά καθορισμένο. Στην πραγματικότητα, όμως, τα περισσότερα «αντικείμενα» που θα θέλαμε να μοντελοποιήσουμε είναι κάθε άλλο παρά σταθερού μεγέθους. Για παράδειγμα μπορούμε να δούμε το γράφημα συνεργασιών μεταξύ ηθοποιών σε διάφορες ταινίες, ή το γράφημα που προκύπτει με κόμβους μαθηματικούς και ακμή ανάμεσά τους αν έχουν συνεργαστεί στην συγγραφή κάποιου άρθρου, ή ακόμη, ένα γράφημα που αναπαριστά ανθρώπινες κοινότητες. Σε αυτό θα έχουμε κόμβους που αντιστοιχούν σε ανθρώπους και ακμές που αντιστοιχούν στην ύπαρξη κάποιου δεσμού μεταξύ τους. Ένα τέτοιο γράφημα δεν θα μπορούσε να είναι σταθερού μεγέθους, οπότε το προηγούμενο μοντέλο κρίνεται εντελώς ακατάλληλο για την μοντελοποίηση ενός τέτοιου αντικειμένου.

Το δεύτερο χαρακτηριστικό που συχνά συναντάμε σε πραγματικά δίκτυα είναι ότι οι ακμές μεταξύ των κόμβων δεν προστίθεται πάντα με τρόπο ομοιόμορφο, αλλά η πι-

θανότητα μια ακμή που ακολουθεί τον καινούργιο κάθε φορά κομβο να προστεθεί σε κόμβο που έχει ήδη μεγάλο βαθμό, είναι μεγαλύτερη από αυτή που αντιστοιχεί σε κάποιον κόμβο με μικρότερο βαθμό. Για παράδειγμα, ένας νέος ηθοποιός είναι πιο πιθανό να συνεργαστεί με ηθοποιούς που ήδη έχουν διαδραματίσει κάποιους ρόλους και κάπως λιγότερο πιθανό να συνεργαστεί με ηθοποιούς που δεν έχουν εμφανιστεί ξανά σε κάποιο έργο. Αντίστοιχα, το ίδιο συμβαίνει και με τη συγγραφή άρθρων, για τους μαθηματικούς για παράδειγμα. Ακόμη πιο φυσικά μπορούμε να δεχθούμε το χαρακτηριστικό αυτό για το γράφημα που αναπαριστά κοινότητες ανθρώπων. Ένας άνθρωπος που εμφανίζεται στην κοινότητα περιμένουμε διαισθητικά να εμφανίσει δεσμούς πιο εύκολα με ανθρώπους που είναι κοινωνικοί ή «σημαντικοί» για την κοινότητα που μελετάμε σε σχέση με το να εμφανίσει δεσμούς με ανθρώπους που είναι κάπως πιο απομονωμένοι συγκριτικά με τους πρώτους.

Συνέπεια αυτών των δύο ιδιοτήτων των μεγάλων πραγματικών δικτύων είναι ότι δημιουργούνται ιδιαίτερες κατανομές σχετικά με τους βαθμούς των κόμβων στο γράφημα, πράγμα που θα αναλυθεί σε επόμενη ενότητα. Οι αλληλεπιδράσεις μεταξύ των βαθμών και το μεγάλο μέγεθος των πραγματικών δικτύων προς το παρόν δυσκολεύουν την ανακάλυψη των τοπολογικών ιδιοτήτων τέτοιων γραφημάτων, αλλά με τη βοήθεια της τεχνολογίας γίνονται ολοένα και περισσότερα βήματα προς αυτή την κατεύθυνση.

Οι Albert και Barabási δείχνουν ότι και τα δύο συστατικά που περιγράφηκαν μόλις πριν, είναι αναγκαία για να μοντελοποιηθούν οι εκθετικές κατανομές που παρατηρήθηκαν σε πραγματικά δίκτυα.

Τέλος, στην βιβλιογραφία εκτιμάται πως αν σε κάποιο μοντέλο ληφθεί με κάποιον τρόπο υπ' όψιν ότι σε πραγματικά δίκτυα, συνήθως ακμές μεταξύ των ήδη υπάρχοντων κόμβων προστίθενται συνεχώς και χωρίς την είσοδο κάποιου καινούριου κόμβου, τότε ο εκθετικός χαρακτήρας του μοντέλου συνεχίζει να υπάρχει, με την προϋπόθεση ότι το πλήθος των ακμών που εισέρχονται μαζί με τους καινούριους κόμβους είναι αρκετά μεγάλο. Αντίθετα, πιστεύεται ότι αν θεωρηθεί ότι υπάρχει η δυνατότητα να υπάρξουν επανασυνδέσεις κόμβων, και επιπλέον αυτές να κυριαρχούν τη μεγέθυνση του γραφήματος, οι πιο συνδεδεμένοι κόμβοι του γραφήματος θα αποκτήσουν εν τέλει όλες τις ακμές. Η δυνατότητα αυτή υπάρχει προφανώς σε γραφήματα που αναπαριστούν τμήματα του www. Τα παραπάνω δείχνουν ότι για τα τυχαία δίκτυα υπάρχουν κάποιες καταστάσεις που ενδεχομένως θα «απορροφήσουν» κάθε άλλη, αν ο ένας παράγοντας είναι μεγαλύτερος από τον άλλο.

3.2 Ορισμός του Μοντέλου

Έχουν προταθεί μοντέλα που να ικανοποιούν τις παραπάνω απαιτήσεις από αρκετούς μαθηματικούς. Τα πρώτα μοντέλα που γράφηκαν λαμβάνοντας υπ' όψιν τα παραπάνω, είναι από τους Yule και Simon. Οι πρώτοι, όμως που ασχολήθηκαν κάπως πιο συστηματικά με το θέμα αυτό ήταν οι Albert και Barabási στο [?]. Ελάχιστα αργότερα (2001) ασχολήθηκαν και οι Bollobás, Riordan, Spencer και Tusnády, στο [9].

Το μοντέλο έτσι όπως ορίστηκε από τους πρώτους είναι:

Ορισμός 3.2.1. *Ξεκινάμε με m_0 αρχικούς κόμβους στο γράφημα. Κάθε χρονική στιγμή προσθέτουμε έναν κόμβο και μαζί με αυτόν προσθέτουμε $m < m_0$ ακμές από τον καινούριο κόμβο στους υπόλοιπους (ή στον εαυτό του). Η πιθανότητα Π με την οποία η κάθε μια από τις m ακμές συνδέεται με τους υπόλοιπους εξαρτάται από τον βαθμό των τελευταίων και δίνεται από τη σχέση $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$. Έτσι, μετά από t χρονικές στιγμές αυτή η διαδικασία θα μας οδηγήσει σε ένα γράφημα με $t + m_0$ κόμβους και mt ακμές.*

Από το παραπάνω προκύπτει ότι υπάρχει μια βασική διαφορά με το προηγούμενο μοντέλο που μελετήθηκε. Ενώ το προηγούμενο μοντέλο αφορά γραφήματα που προκύπτουν με τις διαδικασίες που αναλύθηκαν και θεωρούμε τα ίδια τα γραφήματα ως τυχαίες μεταβλητές, αυτή τη φορά το μοντέλο είναι μια διαδικασία δημιουργίας τ.μ., δηλαδή μια ακολουθία από τυχαία γραφήματα που έχουν σαν δείκτη μια μεταβλητή χρόνου t , έστω $(PA_t)_{t \in \mathbb{N}}$. Έτσι, το μοντέλο είναι ρητά καθορισμένο έτσι ώστε να περιλαμβάνει τη χρονική εξέλιξη των γραφημάτων που προκύπτουν.

Από τη στιγμή που το αντικείμενο μελέτης είναι η ακολουθία $(PA_0, PA_1, \dots, PA_t, \dots)$, για να είναι καλά καθορισμένο το μοντέλο θα πρέπει να καθοριστούν καλά, αφενός, από τι ακριβώς ξεκινά η ακολουθία και, αφετέρου, με ποιον τρόπο λαμβάνουμε το PA_{t+1} εάν γνωρίζουμε το PA_t . Στη διαδικασία που περιγράφηκε παραπάνω, η κάθε ακμή είναι ανεξάρτητη από την παρουσία κάθε άλλης ακμής, πράγμα που συνεπάγεται ότι είναι δυνατόν τα γραφήματα να περιέχουν πολλαπλές ακμές, σε αντίθεση με τα προηγούμενα. Επομένως, η παραπάνω περιγραφή περιγράφει καλά το μοντέλο, με παραμέτρους το αρχικό γράφημα PA_0 και το πλήθος m των ακμών που προστίθεται με την εισαγωγή κάθε νέου κόμβου στο γράφημα. Το αρχικό γράφημα PA_0 , το οποίο δεν είναι καλά καθορισμένο στον παραπάνω ορισμό, μιας και δεν καθορίζεται με ποιον τρόπο συνδέουμε τις πρώτες ακμές μεταξύ των πρώτων κόμβων, από αρκετούς στη βιβλιογραφία θεωρείται ότι δεν παίζει βασικό ρόλο στην εξέλιξη της ακολουθίας PA_t . Με βάση αυτό, μπορούμε να θεωρήσουμε ως μοναδική παράμετρο του μοντέλου το πλήθος m των ακμών που προστίθενται σε κάθε χρονική στιγμή. Παρ' όλα αυτά, ένας κάπως πιο αυστηρός ορισμός δίνεται από τους Bollobás, Riordan, Spencer και Tusnády στο [10].

Ένας γενικότερος ορισμός για αυτού του είδους τα γραφήματα δίνεται στο [?]

Ορισμός 3.2.2. *Όπως και πριν, σκοπός είναι να προσδιοριστεί η ακολουθία γραφημάτων $(PA_0, PA_1, \dots, PA_t, \dots)$. Θεωρούμε μία ακόμα παράμετρο $\delta \geq -1$ για το μοντέλο μας. Για $m = 1$ έχουμε την ακολουθία $(PA_0(1, \delta), PA_1(1, \delta), \dots, PA_t(1, \delta), \dots)$. Το αρχικό γράφημα $PA_0(1, \delta)$ αποτελείται από ένα μονό κόμβο με ένα βρόγχο. Οι κόμβοι των γραφημάτων αυτών θα συμβολίζονται με $\{v_1^{(1)}, v_2^{(1)}, \dots, v_t^{(1)}\}$ και θα θεωρείται ότι κάθε βρόγχος αυξάνει τον βαθμό του κόμβου στον οποίο προσπίπτει κατά δύο. Έχοντας*

το γράφημα $PA_t(1, \delta)$ προσθέτουμε ένα κόμβο $v_{t+1}^{(1)}$ με μία μόνο ακμή. Η πιθανότητα με την οποία η καινούργια ακμή θα επιλέξει να συνδεθεί με τον κόμβο i είναι

$$P\left(v_{t+1}^{(1)} \rightarrow v_i^{(1)} \mid PA_t(1, \delta)\right) = \begin{cases} \frac{1+\delta}{i(2+\delta)+(1+\delta)}, & \text{για } i = t+1 \\ \frac{D_i(t)+\delta}{t(2+\delta)+(1+\delta)} & \text{για } i = 1, \dots, t \end{cases}$$

Ο τρόπος με τον οποίο από το $PA_t(1, \delta)$ αποκτούμε το $PA_t(m, \delta)$ έχει ως εξής: ξεκινάμε με $PA_{mt}(\delta/m, 1)$. Οι κόμβοι $v_1^{(1)}, v_2^{(1)}, \dots, v_m^{(1)}$ του γραφήματος $PA_t(\delta, m)$ ενώνονται και σχηματίζουν τον κόμβο $v_1^{(m)}$ του γραφήματος $PA_t(\delta, m)$. Οι κόμβοι $v_{m+1}^{(1)}, v_{m+2}^{(1)}, \dots, v_{2m}^{(1)}$ του γραφήματος $PA_{mt}(\delta/m, 1)$ σχηματίζουν τον κόμβο $v_2^{(m)}$ του γραφήματος $PA_{mt}(\delta/m, 1)$ και γενικότερα οι κόμβοι $v_{(j-1)m+1}^{(1)}, \dots, v_{jm}^{(1)}$ του γραφήματος $PA_{mt}(\delta/m, 1)$ συγχωνεύονται και σχηματίζουν τον κόμβο $v_j^{(m)}$ του γραφήματος $PA_t(\delta, m)$. Σχετικά με αυτόν τον τρόπο δημιουργίας του γραφήματος $PA_t(\delta, m)$ από το γράφημα $PA_{mt}(\delta/m, 1)$ μπορούν να βρεθούν περισσότερα στην παράγραφο 6 του άρθρου [9] Θεωρούμε ότι οι βαθμοί των κόμβων ανανεώνονται μετά από την εισαγωγή κάθε νέας ακμής στο γράφημα.

Παρατηρούμε ότι για $\delta = 0$ το παραπάνω μοντέλο καταλήγει να είναι αυτό που ορίστηκε από τους BA και που παρουσιάστηκε στην αρχή της παραγράφου. Παρατηρούμε, ακόμη, ότι όσο η τιμή της παραμέτρου δ μεγαλώνει, τόσο η επιλογή των κόμβων που προϋπάρχουν στο γράφημα και με τους οποίους θα συνδεθεί ένας καινούργιος κόμβος τείνει να γίνει ομοιόμορφη. Δηλαδή, όσο πιο μεγάλη είναι η τιμή του δ τόσο μικρότερο "πλεονέκτημα" να επιλεγούν έχουν οι κόμβοι με μεγάλο βαθμό συγκριτικά με αυτούς που έχουν μικρότερο.

Ισοδύναμα, το μοντέλο για $m \geq 2$ μπορεί να οριστεί από το $PA_t(\delta, m)$ που θεωρούμε ότι περιλαμβάνει ένα κόμβο και m βρόγχους. Αν στο $PA_t(\delta, m)$ προσθέσουμε ένα κόμβο με m ακμές, τότε καταλήγουμε στο $PA_{t+1}(\delta, m)$.

3.3 Η κατανομή των βαθμών

Όπως περιγράφηκε και στο προηγούμενο μοντέλο, έτσι και σε αυτό, είναι ενδιαφέρον και χρήσιμο να δούμε την κατανομή των βαθμών ανάμεσα στους κόμβους του γραφήματος. Στο προηγούμενο μοντέλο που περιγράφηκε, η πιθανότητα να βρεθούν κόμβοι με αρκετά μεγαλύτερο βαθμό συγκριτικά με τους υπόλοιπους του γραφήματος μειώνεται με εκθετικό ρυθμό όσο ο βαθμός του κόμβου μεγαλώνει, πράγμα που κάνει πρακτικά απόντες τέτοιους κόμβους. Αντίθετα, σε Preferential Attachment μοντέλα υπάρχουν κάποιοι πολύ ισχυρά συνδεδεμένοι κόμβοι που "κυριαρχούν", κατά κάποιο τρόπο, στο γράφημα.

Οι Albert και Barabási στο [2] αναφέρουν σχετικά: «*Exploring several large databases describing the topology of large networks (...) we show that, independent of the system and identity of its constituents, the probability $P(k)$ that a vertex in the network interacts with k other vertices decays as a power law, following $P(k) \sim K^{-\gamma}$.*»

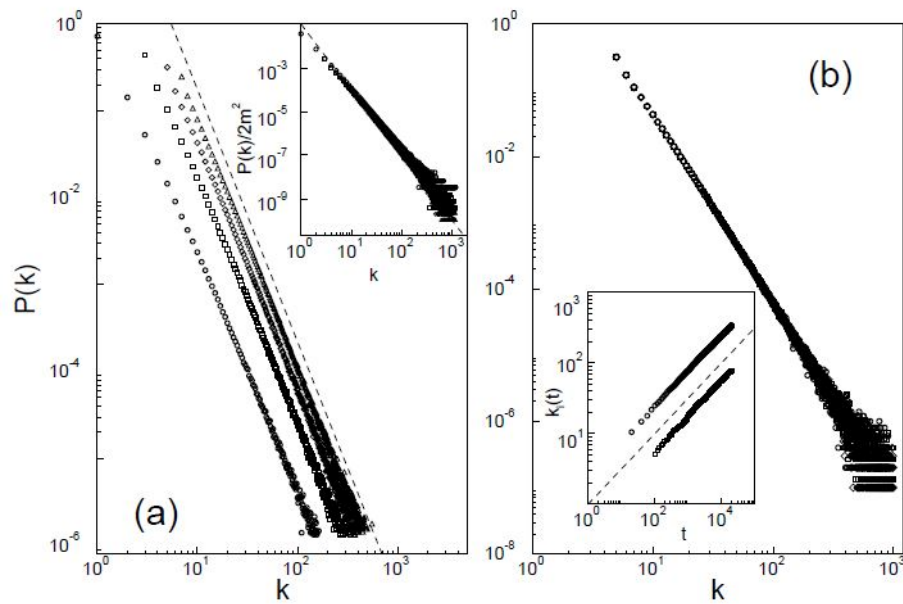
Έπειτα από μελέτες σε διάφορα πειραματικά δεδομένα, ανάμεσά τους και σε δεδομένα γραφημάτων μεταξύ ηθοποιών ή μαθηματικών όπως αυτά έχουν ήδη περιγραφεί, υποψιάζονται ότι οι δύο παράγοντες στους οποίους οφείλεται η συμπεριφορά αυτή είναι και ότι στα πραγματικά δίκτυα έχουμε μεγένθυση όσο περνάει ο χρόνος και ότι η επιλογή των ακμών με τις οποίες πρόκειται να συνδεθεί ο κόμβος που μόλις εισέρχεται δεν είναι ομοιόμορφη αλλά γίνεται δείχνοντας κάποια προτίμηση στους κόμβους που ήδη έχουν μεγαλύτερο βαθμό σε σχέση με άλλους. Έπειτα από πειραματισμούς αφαίρεσης πότε του ενός παράγοντα και πότε του άλλου από τα μοντέλα τους, καταλήγουν ότι πράγματι αυτοί οι δύο είναι οι παράγοντες που αρκεί και πρέπει να είναι στο μοντέλο για να περιγράφονται καλά δεδομένα που παρατηρούνται σε πραγματικά δίκτυα. Κατέληξαν πως η κατανομή της «επιθυμίας» με την οποία οι κόμβοι επιλέγουν να συνδεθούν τοπικά είναι ανεξάρτητη του μεγέθους του γραφήματος και ακολουθεί έναν κανόνα εκθετικό με τάξη ανάμεσα στο 2.1 και το 4, για τα δεδομένα με τα οποία ασχολήθηκαν. Συγκεκριμένα, στο μοντέλο αυτό, οδηγούμαστε σε μια «scale invariant» κατανομή, που στο γράφημα $P(k) - k$, όπου k ο βαθμός των κόμβων, είναι ευθεία με σταθερή κλίση.

Από ότι φαίνεται και στο γράφημα του Σχήματος (3.3a) το δίκτυο που αναπαρίσταται φαίνεται να έχει κατανομή που περιγράφεται από μια ευθεία με κλίση $\gamma = 2.9 \pm 0.1$, όπως αναφέρουν οι συγγραφείς του άρθρου στο οποίο βρέθηκε. Από την επεξεργασία διαφόρων δεδομένων, η κλίση αυτή φαίνεται να είναι ανεξάρτητη από το m , το οποίο αποτελεί την μοναδική παράμετρο του μοντέλου. Ακόμη, από τον εκθετικό κανόνα που παρατηρήθηκε σε πραγματικά δίκτυα, φαίνεται να περιγράφονται γραφήματα διαφορετικών μεγεθών και σε διαφορετική στιγμή της εξέλιξής τους. Έτσι, περιμένουμε το μοντέλο να είναι ανεξάρτητο και από τον χρόνο t . Αυτό επιβεβαιώνεται και από το δίκτυο του Σχήματος (3.3b), στην οποία φαίνεται ότι παρά το γεγονός ότι έχει υπάρξει αύξηση της τάξης του γραφήματος, αυτό «οργανώνεται σε μια στάσιμη scale-free κατάσταση».

Στη συνέχεια περιγράφεται μια μέθοδος για τον υπολογισμό πιθανότητας $P(k)$ να συνδεθεί ένας καινούριος κόμβος σε κάποιον ήδη υπάρχον στο γράφημα βαθμού k . Από τον προσδιορισμό της $P(k)$ μπορούμε να προσδιορίσουμε την τιμή της σταθεράς γ , δηλαδή τον εκθετικό κανόνα που ακολουθείται.

Εάν υποθέσουμε ότι το k είναι συνεχές, η πιθανότητα $P(k_i) = \frac{k_i}{\sum_j k_j}$ μπορεί να εμνηυθεί ως μια τιμή για τη μεταβολή του βαθμού k_i του κόμβου i . Άρα:

$$\frac{\partial k_i}{\partial t} = aP(k_i) = a \frac{k_i}{\sum_j^{m_0+t-1} k_j}.$$



Σχήμα 3.1: Κατανομή των βαθμών σε γράφημα που έχει κατασκευαστεί με βάση τις προηγούμενες διαδικασίες. Το (a) έχει $N = m_o + t = 300\,000$ και $m_o = m = 1$ για τους κύκλους, $m_o = m = 5$ για τους ρόμβους και $m_o = m = 7$ για τα τρίγωνα. Η κλίση της διακεκομμένης γραμμής είναι 2.9. Το μικρό γράφημα στο εσωτερικό προκύπτει από τα δεδομένα του μεγάλου διαιρεμένα με τη σταθερά m^2 και η κλίση της διακεκομμένης γραμμής στο μικρό γράφημα είναι 3. Το (b) έχει $m_o = m = 5$ και $N = 100\,000$ για τους κύκλους, $N = 150\,000$ για τα τετράγωνα και $N = 200\,000$ για τους ρόμβους. Το μικρό πλαίσιο στο εσωτερικό του δείχνει τη χρονική εξέλιξη των βαθμών δύο κόμβων που προστέθηκαν στο σύστημα τις χρονικές στιγμές $t_1 = 5$ και $t_2 = 95$, με $m_o = m = 5$ και κλίση της διακεκομμένης γραμμής ίση με 0.5, όπως προβλέπεται από τη μέθοδο.

Επειδή το άθροισμα των βαθμών των κόμβων είναι $\sum_{j=1}^t k_j = 2mt$ και σε κάθε χρονική στιγμή, μαζί με τον καινούριο κόμβο εισέρχονται στο γράφημα m ακμές, η διαφορά στο άθροισμα των βαθμών των κόμβων (που ήδη υπήρχαν στο γράφημα) είναι $\Delta k = m$. Συγκρίνοντας με τα παραπάνω, προκύπτει ότι $a = m$ και έτσι έχουμε ότι:

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t}.$$

Για την λύση αυτής της διαφορικής εξίσωσης θεωρούμε ως αρχική συνθήκη την $k_i(t_i) = m$, δηλαδή το γεγονός ότι ο τυχόν κόμβος i έχει m συνδέσεις από την χρονική στιγμή που θα εισαχθεί στο γράφημα. Έτσι, φτάνουμε στην λύση της διαφορικής:

$$k_i(t) = m \left(\frac{t}{t_i} \right)^{\frac{1}{2}}. \quad (3.1)$$

Όμως, οι «γηραιότεροι» κόμβοι (δηλαδή μικρό t_i) λαμβάνουν με μεγαλύτερη πιθανότητα ακμές από τους κόμβους που εισέρχονται σε σχέση με τους «νεότερους» (με μεγαλύτερο t_i) κόμβους. Η ιδιότητα αυτή, δηλαδή «ο πλούσιος να γίνεται πλουσιότερος» πιο γρήγορα από τους «φτωχούς» είναι κάτι που έτσι και αλλιώς παρατηρείται σε πραγματικά δίκτυα, πολλών ειδών.

Χρησιμοποιώντας τα παραπάνω, υπολογίζουμε αναλυτικά την σταθερά γ . Από την Σχέση (3.1) έχουμε:

$$P(k_i(t) < k) = P\left(t_i > \frac{m^2 t}{k^2}\right) = 1 - P\left(t_i \leq \frac{m^2 t}{k^2}\right) = 1 - \frac{m^2 t}{k^2(t + m_0)},$$

αφού η πυκνότητα της πιθανότητας του t_i του κόμβου i είναι, μιας και οι κόμβοι εισάγονται στο γράφημα ανα σταθερά χρονικά διαστήματα, $P_i(t_i) = \frac{1}{m_0 + t}$. Με βάση αυτά, έχουμε:

$$P(k) = \frac{\partial P(k_i < k)}{\partial t} = \frac{2m^2 t}{m_0 + t} \cdot \frac{1}{k^3},$$

από όπου η εκθετική σταθερά είναι $\gamma = 3$, ανεξάρτητη του m . Δηλαδή, είναι:

$$P(k) \sim Ak^{-\gamma},$$

όπου A σταθερά ανάλογη του m^2 .

Έτσι, ο συνδυασμός των παραγόντων που αναφέρθηκαν προηγουμένως, μας οδηγεί πιο κοντά στην κατανόηση της δυναμικής των συνδέσεων μεταξύ των κόμβων. Μέσω της διαδικασίας αυτής, οι κόμβοι που έχουν πολλές συνδέσεις είναι πιθανότατα αυτοί που εισήχθησαν στο μοντέλο στα αρχικά στάδια. Άρα κάποιοι από τους «γηραιότερους» κόμβους του γραφήματος θα αποκτούν όλο και περισσότερες συνδέσεις, πράγμα που είναι υπεύθυνο για το γεγονός ότι σε τέτοιου τύπου γραφήματα υπάρχουν «hubs», δηλαδή πολύ κεντρικότεροι σε σχέση με τον εσωτερικό βαθμό τους, κόμβοι.

• **Σημείωση 1:** Τα παραπάνω ισχύουν για γραφήματα που ανήκουν σε ακολουθία $PA_t(m, 0)$.

Από τα προηγούμενα, υποδεικνύεται ότι το ποσοστό των κόμβων του $PA_t(m, 0)$ βαθμού k μειώνεται με ρυθμό k^{-3} καθώς το $k \rightarrow +\infty$. Έστω X_k^δ το πλήθος των κόμβων του $PA_t(m, \delta)$ με εσωτερικό βαθμό ίσο με k . Τότε ισχύει το ακόλουθο:

Θεώρημα 3.3.1. Έστω $m \geq 1$, $PA_t(m, 0)_{t \geq 0}$ η διαδικασία που έχει περιγραφεί σε προηγούμενο ορισμό και $\varepsilon > 0$. Τότε:

$$\lim_{n \rightarrow +\infty} P \left[(1 - \varepsilon) a_{m,k} \leq \frac{X_k^0}{n} \leq (1 + \varepsilon) a_{m,k} \right] = 1, \forall 0 \leq k \leq n^{\frac{1}{15}},$$

$$\text{όπου } a_{m,k} = \frac{2m(m+1)}{(k+m)(k+m+1)(k+m+2)}.$$

Απόδειξη. Μπορεί να βρεθεί στο [10]. \square

Στο θεώρημα αυτό οι Bollóbas, Riordan, Spencer και Tusnadý δείχνουν ότι, για μεγάλο t , με πιθανότητα 1 η κατανομή των βαθμών στους γράφους αυτούς ακολουθεί εκθετικό κανόνα με $\gamma = 3$, αφού $a_{m,k} \sim \frac{1}{k^3}$, όπως ακριβώς αποδείχθηκε από τους Albert και Barabasi στην προηγούμενη διαδικασία. Το παραπάνω θεώρημα δείχνει ότι για τιμές του βαθμού k στο διάστημα $0 \leq k \leq t^{\frac{1}{15}}$ το ποσοστό των κόμβων με βαθμό k είναι με πιθανότητα 1 ίσο με $a_{m,k}$.

• **Σημείωση 2:** Η ύπαρξη αυτού του εκθετικού κανόνα, που κυριαρχεί και την εξέλιξη του παγκόσμιου ιστού, επιτρέπει να προσδιοριστεί το αναμενόμενο πλήθος ιστοσελίδων προς τις οποίες υπάρχει <link>, χωρίς αυτό βέβαια να γίνει εξαντλητικά, πράγμα αδύνατο δεδομένου του μεγέθους του. Ακόμη, δείχνει ότι η πιθανότητα να βρεθεί κανείς σε ιστοσελίδες με μεγάλο εσωτερικό βαθμο είναι αρκετά μεγάλη με δεδομένο ότι αυτές κυριαρχούν στον ιστό, πράγμα που μπορεί εύκολα να επιβεβαιωθεί από μια απλή χρήση του ιστού.

Έστω η χρονική στιγμή $k \geq 2$. Στα επόμενα λαμβάνουμε υπ' όψιν ότι μόνο τις χρονικές στιγμές $t \geq k$ για τις οποίες ο νέος κόμβος $(PA_t(1, \delta))_{t \geq 1}$ συνδέεται με τους k αρχικούς κόμβους του γραφήματος. Δηλαδή, οι μόνες χρονικές στιγμές t που δεν λαμβάνουμε υπ' όψιν είναι αυτές κατά τις οποίες ο καινούριος κόμβος t επιλέγει να φτιάξει βρόγχο. Τότε ισχύει το ακόλουθο:

Θεώρημα 3.3.2. Για $t \rightarrow +\infty$, στο $PA_t(1, \delta)$ ισχύει ότι:

$$\frac{D_k(t)}{D_1(t) + D_2(t) + \dots + D_k(t)} \xrightarrow{\sigma, \beta} B_k \sim B(1 + \delta, (2 + \delta)(k - 1)).$$

Απόδειξη. Έστω οι χρόνοι διακοπής $(T_n(k))_{k \geq 2n-1}$, όπου το συνολικό άθροισμα των βαθμών κόμβων $1, \dots, n$ του γραφήματος ισούται με k , δηλαδή:

$$T_n(k) = \inf \{t : D_1(t) + \dots + D_k(t) = k\}.$$

Θα είναι $T_n(2n-1) = n-1$.

Αφού $D_j(t) \xrightarrow{\sigma, \beta} +\infty$ για $t \rightarrow +\infty$, ισχύει πως $T_n(k) < +\infty, \forall k$. Ακόμη, αφού $\lim_{k \rightarrow +\infty} T_n(k) = +\infty$, ισχύει:

$$\begin{aligned} \lim_{t \rightarrow +\infty} \frac{D_n(t)}{D_1(t) + \dots + D_n(t)} &\stackrel{(1)}{=} \lim_{k \rightarrow +\infty} \frac{D_n(T_n(k))}{D_1(T_n(k)) + \dots + D_n(T_n(k))} \\ &= \lim_{k \rightarrow +\infty} \frac{D_n(T_n(k))}{k}. \end{aligned}$$

Οι ακολουθία τ.μ $(D_1(T_n(k)), \dots, D_{n-1}(T_n(k)), D_n(T_n(k)))_{k \geq 2n-1}$ αποτελεί ένα σύστημα δοχείων Polya ¹

$$D_n(T_n(2n-1)) = 1, D_1(T_n(2n-1)) + \dots + D_{n-1}(T_n(2n-1)) = 2n-2.$$

Η ακμή του νέου κόμβου τη στιγμή $T_n(k)$, συνδέεται με τον κόμβο n με πιθανότητα:

$$\frac{D_n(t) + \delta}{k + n\delta}$$

που είναι η πιθανότητα ενός P.U.S. με $a_b = \delta, a_r = (k-1)\delta, b_0 = 1, r_0 = 2(k-1)$ και άρα από Θεώρημα σχετικό με τα P.U.S., προκύπτει το αποτέλεσμα αυτού του θεωρήματος. \square

¹**Συστήματα Δοχείων του Polya, Polya Urn Schemes (P.U.S.):**

Θεωρούμε ένα δοχείο το οποίο έχει αρχικά b_0 μπλε μπάλες και r_0 κόκκινες μπάλες. Κάθε φορά που τραβάμε μια μπάλα i χρώματος, την επανατοποθετούμε στο δοχείο μαζί με άλλες l μπάλες του ίδιου χρώματος. Η εφαρμογή των ιδιοτήτων των Polya Urn Schemes για $l = 1$ φαίνεται να έχει κάποια συμμετρία με τα γραφήματα που μελετώνται στην ενότητα αυτή.

Έστω μια συνάρτηση βάρους για κάθε χρώμα $W_b, W_r : N \rightarrow (0, +\infty)$ τέτοια ώστε $W_b(k) = a_b + k$ και $W_r(k) = a_r + k$ με $a_b, a_r \geq 0$, δηλαδή γραμμικές με την ίδια κλίση. Θεωρούμε ότι η πιθανότητα να τραβήξουμε τυχαία μια μπάλα κατά την n -οστή δοκιμή και αυτή η μπάλα να είναι κόκκινη δίνεται από τη σχέση:

$$\frac{W_r(R_n)}{W_b(b_n) + W_r(r_n)} = \frac{R_n + a_r}{B_n + R_n + a_b + a_r}.$$

Είναι προφανές ότι: $B_n + R_n = b_0 + r_0 + n$.

Θεώρημα 3.3.3 (De Finetti). Έστω $(X_i)_{i \geq 1}$ μια άπειρη ακολουθία από ανταλλάξιμες τυχαίες μεταβλητές και $X_i \in \{0, 1\}$. Τότε υπάρχει τυχαία μεταβλητή U με $P(U \in [0, 1]) = 1$ τέτοια ώστε $\forall 1 \leq k \leq n$ να ισχύει

$$P(X_1 = \dots = X_k = 1, X_{k+1} = \dots = X_n = 0) = E[U^k (1-U)^{n-k}].$$

Θεώρημα 3.3.4. Έστω $((B_n, R_n)_n)_{n=1}^{+\infty}$ ένα $((B_n, R_n)_n)_{n \geq 1}$ ένα P.U.S., όπως αυτό περιγράφηκε προηγουμένως. Τότε:

$$\frac{B_n}{B_n + R_n} \xrightarrow{\sigma\beta} U \sim B(b_0 + a_b, r_0 + a_r)$$

και

$$P(B_n = b_0 + k) = E[P(\text{Bin}(n, U) = k)].$$

Απόδειξη. Εφαρμογή του Θεωρήματος De Finetti σε άπειρες ανταλλάξιμες ακολουθίες από μπλε και κόκκινες μπάλες. \square

3.4 Συνεκτικότητα, Αποστάσεις και Διάμετρος.

3.4.1 Για το πλήθος των συνεκτικών συνιστωσών

Όπως και στην αντίστοιχη ενότητα για το μοντέλο των Erdős και Renyi, έτσι και για το παρόν, θα μελετηθούν η συνεκτικότητα, οι αποστάσεις και η διάμετρος των γραφημάτων που προκύπτουν από τις διαδικασίες που περιγράφηκαν αρχικά. Μελετάμε τη ακριβώς συμβαίνει στο γράφημα για δύο περιπτώσεις: για $m = 1$ και για $m \geq 2$.

- Για $m = 1$:

Έστω $N_t = I_1 + I_2 + \dots + I_t$ το πλήθος των συνεκτικών συνιστωσών του $PA_t(1, \delta)$, όπου I_i η δείκτρια συνάρτηση για το ενδεχόμενο η i -οστή ακμή να δημιουργήσει βρόγχο.

Θεώρημα 3.4.1. Για το πλήθος των συνεκτικών συνιστωσών N_t στο $PA_t(1, \delta)$ ισχύει ότι:

$$E(N_t) = \frac{1 + \delta}{2 + \delta} \log t (1 + O(1)),$$

και

$$V(N_t) = \frac{1 + \delta}{2 + \delta} \log t (1 + O(1)).$$

Λήμμα 3.4.1 (Θεώρημα Stoltz - Cesáro για σύγκλιση ακολουθιών). Έστω $(a_n)_{n \geq 1}$, $(b_n)_{n \geq 1}$ μη φραγμένες ακολουθίες πραγματικών αριθμών, με $(b_n)_{n \geq 1}$ γνησίως μονότονη. Αν το $\lim_{n \rightarrow +\infty} \frac{a_{n+1} - a_n}{b_{n+1} - b_n} = l$ υπάρχει, τότε $\lim_{n \rightarrow +\infty} \frac{a_n}{b_n} = l$.

Απόδειξη. (Θεωρήματος) Από τον ορισμό του μοντέλου, εύκολα προκύπτει ότι είναι:

$$P(I_{t+1} = 1 \mid \mathcal{F}_t) = \frac{1 + \delta}{t(2 + \delta) + (1 + \delta)}$$

και

$$P(I_{t+1} = 0 \mid \mathcal{F}_t) = 1 - \frac{1 + \delta}{t(2 + \delta) + (1 + \delta)},$$

η πιθανότητα να δημιουργηθεί βρόγχος και να μη δημιουργηθεί αντίστοιχα. Ισοδύναμα, την χρονική στιγμή $t + 1$ το γράφημα να έχει μια επιπλέον συνεκτική συνιστώσα σε σχέση με την χρονική στιγμή t ή να έχει το ίδιο πλήθος συνιστωσών. Έχουμε:

$$\begin{aligned} E(N_{t+1} \mid \mathcal{F}_t) &= E(N_t + I_{t+1} \mid \mathcal{F}_t) \\ &= N_t + \frac{1 + \delta}{t(2 + \delta) + (1 + \delta)} + 0 \cdot \left(\frac{1 + \delta}{t(2 + \delta) + (1 + \delta)} \right) \\ &= N_t + \frac{1 + \delta}{t(2 + \delta) + (1 + \delta)}. \end{aligned}$$

Ισχυριζόμαστε ότι η ποσότητα:

$$M_t = N_t - \sum_{k=1}^{t-1} \frac{1 + \delta}{(k-1)(2 + \delta) + (1 + \delta)}$$

είναι Martingale². Πράγματι, έχουμε:

$$E(M_{t+1} - M_t \mid \mathcal{F}_t) = E\left(N_{t+1} - N_t - \frac{1 + \delta}{t(2 + \delta) + (1 + \delta)} \mid \mathcal{F}_t\right) = 0.$$

²Μια ακολουθία τ.μ. $((X_n)_{n \in \mathbb{N}})$ καλείται Martingale όταν $E[X_{n+1} \mid X_1, \dots, X_n] = X_n$.

Επειδή λοιπόν $E(M_{t+1} - M_t | \mathcal{F}_t) = 0$, ισχύει ότι $E(M_{t+1} | \mathcal{F}_t) = M_t$. Ακόμη ισχύει ότι:

$$M_t = (M_t - M_{t-1}) + (M_{t-1} - M_{t-2}) + \dots + (M_2 - M_1) + M_1 \quad (3.2)$$

και

$$M_t - M_1 = (M_t - M_{t-1}) + \dots + (M_2 - M_1). \quad (3.3)$$

Επομένως, γνωρίζοντας το αρχικό γράφημα, έχουμε:

$$\begin{aligned} E(M_t | \mathcal{F}_1) &= \\ &= E(M_t - M_{t-1} | \mathcal{F}_1) + E(M_{t-1} - M_{t-2} | \mathcal{F}_1) \\ &\quad + \dots + E(M_2 - M_1 | \mathcal{F}_1) + E(M_1 | \mathcal{F}_1) \\ &= 0 + 0 + \dots + 0 + E(M_1 | \mathcal{F}_1). \end{aligned}$$

Αλλά $E(M_1 | \mathcal{F}_1) = 0$ και, άρα, $E(M_t | \mathcal{F}_t) = 0$. Έτσι:

$$E(N_t) = \sum_{k=1}^{t-1} \frac{1 + \delta}{k(2 + \delta) + (1 + \delta)}.$$

Με βάση τα παραπάνω λοιπόν για την πρώτη σχέση αρκεί να δειχθεί ότι:

$$\frac{1}{\log t} \sum_{k=1}^{t-1} \frac{1 + \delta}{k(2 + \delta) + (1 + \delta)} \xrightarrow{t \rightarrow +\infty} 1.$$

Κάνοντας χρήση του Θεωρήματος Stoltz - Cesàro για την σύγκλιση ακολουθιών και θεωρώντας: $a_t = \sum_{j=1}^k \frac{1}{j}$ και $b_t = \log t$ έχουμε:

$$\lim_{t \rightarrow +\infty} \frac{a_{t+1} - a_t}{b_{t+1} - b_t} = \lim_{t \rightarrow +\infty} \frac{\frac{1+\delta}{t(2+\delta)+(1+\delta)}}{\log(1+\frac{1}{t})} \xrightarrow{D'l'H} \frac{1+\delta}{2+\delta}, \text{ για } t \rightarrow +\infty$$

και το ζητούμενο αποδείχθηκε.

-Για την απόδειξη της δεύτερης σχέσης, με την βοήθεια της Σχέσης (3.3) έχουμε:

$$\begin{aligned} V(M_t) &= E[(M_t - M_1)^2] = E\left[\left(\sum_{n=1}^{t-1} M_{n+1} - M_n\right)^2\right] = \\ &= E\left[\sum_{n=1}^{t-1} (M_{n+1} - M_n)^2 + 2 \sum_{n=0, i>j}^{t-1} (M_{i+1} - M_i)(M_{j+1} - M_j)\right]. \end{aligned}$$

Αλλά για κάθε τυχαία μεταβλητή $X \in \mathcal{F}$ είναι:

$$E[X] = E[E[X | \mathcal{F}]]$$

και άρα:

$$E[(M_{i+1} - M_i)(M_{j+1} - M_j)] = E[E[(M_{i+1} - M_i)(M_{j+1} - M_j) | \mathcal{F}_{j+1}]] =$$

$$= E[(M_{j+1} - M_j) \cdot E[(M_{i+1} - M_i) | \mathcal{F}_{i+1}]] = E[(M_{j+1} - M_j) \cdot 0] = 0$$

και τελικά είναι:

$$\begin{aligned} V(M_n) &= \sum_{n=1}^{t-1} E[(M_{n+1} - M_n)^2] \\ &= \sum_{n=1}^{t-1} \left(N_{t+1} - N_t - \frac{1+\delta}{k(2+\delta) + (1+\delta)} \right) = \sum_{n=1}^{t-1} V(I_n) \\ &= \sum_{n=1}^{t-1} \left[(0 - E(I_n))^2 \left(1 - \frac{1+\delta}{k(2+\delta) + (1+\delta)} \right) \right. \\ &\quad \left. + (1 - E(I_n))^2 \left(\frac{1+\delta}{k(2+\delta) + (1+\delta)} \right) \right] = \\ &= \sum_{n=1}^{t-1} \left[E(I_n)^2 + E(I_n)^2 \left(\frac{1+\delta}{k(2+\delta) + (1+\delta)} \right) \right. \\ &\quad \left. - E(I_n)^2 \left(\frac{1+\delta}{k(2+\delta) + (1+\delta)} \right) \right. \\ &\quad \left. + \frac{1+\delta}{k(2+\delta) + (1+\delta)} - 2E(I_n) \frac{1+\delta}{k(2+\delta) + (1+\delta)} \right] = \\ &= \sum_{n=1}^{t-1} \left[E(I_n)^2 - 2E(I_n) \frac{1+\delta}{k(2+\delta) + (1+\delta)} + \frac{1+\delta}{k(2+\delta) + (1+\delta)} \right] = \\ &= \sum_{n=1}^{t-1} \left[\left(1 \cdot \frac{1+\delta}{k(2+\delta) + (1+\delta)} + 0 \cdot \frac{1+\delta}{k(2+\delta) + (1+\delta)} \right)^2 + \frac{1+\delta}{k(2+\delta) + (1+\delta)} \right. \\ &\quad \left. - 2 \frac{1+\delta}{k(2+\delta) + (1+\delta)} \left(1 \cdot \frac{1+\delta}{k(2+\delta) + (1+\delta)} + 0 \cdot \frac{1+\delta}{k(2+\delta) + (1+\delta)} \right) \right] = \\ &= \sum_{n=1}^{t-1} \left[\left(\frac{1+\delta}{k(2+\delta) + (1+\delta)} \right)^2 - 2 \left(\frac{1+\delta}{k(2+\delta) + (1+\delta)} \right)^2 + \frac{1+\delta}{k(2+\delta) + (1+\delta)} \right] \\ &= \sum_{n=1}^{t-1} \left[\frac{1+\delta}{k(2+\delta) + (1+\delta)} - \left(\frac{1+\delta}{k(2+\delta) + (1+\delta)} \right)^2 \right] \\ &= \sum_{n=1}^{t-1} \frac{1+\delta}{k(2+\delta) + (1+\delta)} - \sum_{n=1}^{t-1} \left(\frac{1+\delta}{k(2+\delta) + (1+\delta)} \right)^2. \end{aligned}$$

Θεωρώντας τις ακολουθίες:

$$a_t = \sum_{j=1}^t \left(\frac{1}{j} - \frac{1}{j^2} \right)$$

και

$$b_t = \log t$$

και κάνοντας χρήση του Θεωρήματος Stoltz-Cesáro αποδεικνύεται το ζητούμενο. \square

Από τα παραπάνω παρατηρούμε ότι ικανοποιούνται οι υποθέσεις του Κεντρικού Οριακού Θεωρήματος, και άρα ισχύει ότι:

$$\frac{N_t - E(N_t)}{\sqrt{Var(N_t)}} \xrightarrow{t \rightarrow +\infty} U \sim N(0, 1).$$

• Για $m \geq 2$:

Θεώρημα 3.4.2. Για τυχόν $m \geq 2$ ισχύει ότι το γράφημα $PA_t(m, \delta)$ είναι σχεδόν βεβαίως συνεκτικό.

Απόδειξη. Έστω N_t = το πλήθος των συνεκτικών συνιστωσών του $PA_t(m, \delta)$. Τότε το ενδεχόμενο $I_t = N_t - N_{t-1} = 1$ ισοδυναμεί με το ενδεχόμενο ο καινούριος κόμβος t στο γράφημα να ενώσει όλες του (τις m) ακμές στον εαυτό του και το $I_t = N_t - N_{t-1} < 0$ με το ενδεχόμενο m ακμές που θα προστεθούν να ενώσουν τον καινούριο κόμβο με δύο διαφορετικές μεταξύ τους συνεκτικές του προϋπάρχοντος γραφήματος.

Η πιθανότητα να συμβεί το πρώτο είναι:

$$P[I_t = 1] = \prod_{e=1}^m \frac{2e - 1 + \delta}{(2m + \delta)t + (2e - 1 + \delta)}$$

Επειδή το $\sum_{t=2}^{+\infty} P(I_t = 1) < +\infty$, από το πρώτο Λήμμα Borel-Cantelli έχουμε ότι $P(\limsup_t I_t = 1) = 0$ και, άρα, το ενδεχόμενο $I_t = 1$ κάποια χρονική στιγμή θα σταματήσει να πραγματοποιείται. Αντίθετα, το δεύτερο ενδεχόμενο πραγματοποιείται με μεγαλύτερη πιθανότητα, πράγμα που οδηγεί στο να είναι $\lim_{t \rightarrow +\infty} N_t = 1$, με μεγάλη πιθανότητα.

Άρα, αφού $N_t \leq 1 + \sum_{t=2}^{+\infty} I_t$ και το I_t πραγματοποιείται σβ πεπερασμένες φορές, θα είναι $\lim_{t \rightarrow +\infty} N_t < +\infty$, δηλαδή το $PA_t(m, \delta)$ έχει σβ - πεπερασμένες στο πλήθος συνιστώσες και είναι $P[PA_t(m, \delta) = disconnected] \in (0, 1)$.

Έστω $N \geq 1$ αρκετά μεγάλο. Τότε από τα παραπάνω προκύπτει ότι:

$$\lim_{N \rightarrow +\infty} P \left[\sum_{t=N}^{+\infty} I_t = 0 \right] = 1.$$

Αυτό σημαίνει ότι, με πιθανότητα που τείνει στο 1, από την χρονική στιγμή $t = N$ και μετά δεν θα έχουμε δημιουργία καινούριων συνεκτικών συνιστωσών. Αυτό που μένει συνεπώς να δειχθεί, προκειμένου το $PA_t(m, \delta)$ να είναι με μεγάλη πιθανότητα συνεκτικό, είναι ότι οι κόμβοι $1, \dots, N$ ανήκουν στην ίδια συνεκτική συνιστώσα.

Ισχυρισμός 1:

Για κάποιο τυχαίο αλλά συγκεκριμένο κόμβο $i \in 1, \dots, N$, το πλήθος των κόμβων που συνδέονται με τον i φράσσεται από κάτω από το B_{t-1} , όπου B_{t-1} είναι το πλήθος των μπλε μπαλών σε ένα Polya Urn Schemes.

Ισχυρισμός 2:

Το μέγεθος της συνεκτικής συνιστώσας του i -κόμβου, $\forall i \in 1, \dots, N$, είναι με μεγάλη πιθανότητα τουλάχιστον $\varepsilon > 0$.

Ισχυρισμός 3:

Εάν οι ισχυρισμοί 1 και 2 ευσταθούν, τότε για κάποιο t αρκετά μεγάλο, ισχύει ότι με μεγάλη πιθανότητα όλοι οι κόμβοι $1 \dots, N$ είναι στην ίδια συνεκτική συνιστώσα με τον i κόμβο και άρα το θεώρημα αποδείχθηκε.

Για τον ισχυρισμό 1:

Έστω ένα P.U.S. και έστω ότι αρχικά στο δοχείο περιέχονται μια μπλε ($b_0 = 1$) και i

κόκκινες μπάλες ($r_0 = i$). Το περιεχόμενο του δοχείου $((B_n, R_n))_{n \geq 1}$ περιγράφεται ως εξής: Από τον ορισμό των Συστημάτων Δοχείων Polya, η πιθανότητα να τραβήξουμε μπάλα χρώματος μπλε είναι ανάλογη με το πλήθος των ήδη υπάρχοντων μπλε μπαλών στο δοχείο, και αν αυτό συμβεί, τότε επανατοποθετούμε την μπλε μπάλα μαζί με μια ακόμη μπλε μπάλα στο δοχείο. Τα ίδια ισχύουν και για τις κόκκινες μπάλες.

Ο ισχυρισμός θα αποδειχθεί με επαγωγή ως προς το χρόνο t , όπου $t \geq i$. *Επαγωγική βάση:* Την χρονική στιγμή i το πλήθος των γειτόνων του i είναι τουλάχιστον ένα. Γείτονας θεωρείται και ο ίδιος ο κόμβος i σε περίπτωση που κατά την είσοδο του στο γράφημα έχει δημιουργήσει m βρόγχους. Άρα, αφού $B_0 = b_0 = 1$, για $t = i$ ο ισχυρισμός είναι αληθής. *Επαγωγική Υπόθεση:* Έστω ότι ισχύει πως $C_i(t) \geq B_{t-i}$, όπου $C_i(t)$ είναι το πλήθος των κόμβων στην συνεκτική συνιστώσα του i . *Επαγωγικό Βήμα:* Θα δείξουμε ότι αν η πλάγια επαγωγική υπόθεση ευσταθεί, τότε ισχύει και ότι $C_i(t+1) \geq B_{t+1-i}$. Πράγματι, όταν $C_i(t) \geq B_{t-i}$, το συνολικό βάρος των ακμών του κόμβου i την χρονική στιγμή t είναι τουλάχιστον $B_{t-i}(2m + \delta)$. Ταυτόχρονα, έχουμε ότι: $B_{t-i} + R_{t-i} = (t-i) + b_0 + r_0 = t+1$. Οπότε, το βάρος των ακμών που προσπίπτουν σε όλους κόμβους του γραφήματος είναι $t(2m + \delta) + (1 + \delta) < (t+1)(2m + \delta) = (2m + \delta)(B_{t-i} + R_{t-i})$.

Σε αντιστοιχία με τα προηγούμενα, θεωρούμε ότι «τραβάμε» μια κόκκινη μπάλα την χρονική στιγμή $t+1$, όταν η πρώτη ακμή από τον $t+1$ κόμβο στο γράφημα συνδέει τον $t+1$ με κάποιον j , που βρίσκεται στην ίδια συνεκτική συνιστώσα με τον i (στο γράφημα G_t). Αν αυτό συμβεί, τότε ο $t+1$ θα βρίσκεται στην ίδια συνεκτική συνιστώσα του i και συνεπώς το πλήθος των κόμβων $C_i(t)$ στην συνιστώσα αυτή αυξάνεται κατά τουλάχιστον ένα. Δηλαδή, $C_i(t+1) \geq B_{t+1-i}$ και ο πρώτος ισχυρισμός αποδείχθηκε.

Για τον ισχυρισμό 2:

Από το Θεώρημα της σχετικής με τα P.U.S. υποσημείωσης, έχουμε ότι:

$$\frac{B_t}{t} \xrightarrow{\sigma\beta} U \sim B(1, i). \quad (3.4)$$

Αφού $\frac{C_i(t)}{t-i} \geq \frac{B_{t-i}}{t-i}, \forall i = 1, \dots, N$, έπεται ότι:

$$\liminf_{t \rightarrow +\infty} \frac{C_i(t)}{t-i} \geq \liminf_{t \rightarrow +\infty} \frac{B_{t-i}}{t-i}, \forall i = 1, \dots, N. \quad (3.5)$$

Από (1) $\implies P(U = 0) = 0$ και άρα $\forall \delta, \exists \varepsilon : P(U \geq \varepsilon) \geq 1 - \delta$. Δηλαδή:

$$\begin{aligned} P\left(\lim_{t \rightarrow +\infty} \frac{B_t}{t} \geq \varepsilon\right) &\geq 1 - \delta \\ &\implies P\left(\liminf_{t \rightarrow +\infty} \frac{B_t}{t} \geq \varepsilon\right) \geq 1 - \delta \stackrel{(2)}{\implies} \\ &\implies P\left(\liminf_{t \rightarrow +\infty} \frac{C_i(t)}{t-i} \geq \varepsilon\right) \geq 1 - \delta, \forall i = 1, \dots, N \implies \\ &\implies P\left(\liminf_{t \rightarrow +\infty} \frac{C_i(t)}{t} \geq \varepsilon\right) \geq 1 - \delta, \forall i = 1, \dots, N \implies \\ &\implies P\left(\liminf_{t \rightarrow +\infty} \frac{C_i(t)}{t} \geq \varepsilon, \forall i = 1, \dots, N\right) = 1 - O(1), \end{aligned}$$

καθώς $\varepsilon \rightarrow 0$, και έτσι αποδείχθηκε ο δεύτερος ισχυρισμός.

Για τον ισχυρισμό 3:

Για τον 3ο ισχυρισμό, αρκεί να δείξουμε ότι για $i \neq j$ και $\frac{C_i(t)}{t} \geq \varepsilon$, $\frac{C_j(t)}{t} \geq \varepsilon$, με μεγάλη πιθανότητα, ισχύει ότι τα i, j ανήκουν στην ίδια συνεκτική συνιστώσα στο $PA_{2t}(m, \delta)$. Έστω $l \in t+1, \dots, 2t$. Για να είναι τα i, j στην ίδια συνεκτική συνιστώσα, αρκεί μια από τις m ακμές να συνδέεται με κάποιον κόμβο στην συνιστώσα του i και μια από τις υπόλοιπες $m-1$ να συνδέεται με κάποιο κόμβο στην συνιστώσα του j . Τότε, ο κόμβος l ονομάζεται t -connector για τους κόμβους i, j . Έτσι έχουμε:

$$P[l \text{ είναι } t\text{-connector για } i, j \mid PA_t(m, \delta)] \geq \frac{C_i(t)(2m+\delta)}{l(2m+\delta) + (1+\delta)} \cdot \frac{C_j(t)(2m+\delta)}{l(2m+\delta) + (1+\delta) + 1} \quad (3.6)$$

αφού το «βάρος» των συνιστωσών των i, j κόμβων κατά την επιλογή των υποψήφιων κόμβων με τους οποίους ο l θα συνδέσει τις m ακμές του είναι $C_i(t)(2m+\delta)$ και $C_j(t)(2m+\delta)$ αντίστοιχα. Επίσης, αφού $l \in t+1, \dots, 2t$ είναι $t < 2t+1$, και έτσι:

$$\begin{aligned} \frac{C_k(t)(2m+\delta)}{l(2m+\delta) + (1+\delta)} &\geq \frac{C_k(t)(2m+\delta)}{(2t+1)(2m+\delta) + (1+\delta)} \geq \frac{C_k(t)(2m+\delta)}{(2t+1)(2m+\delta)} = \\ &= \frac{C_k(t)}{(2t+1)} \geq \frac{\varepsilon}{2}, \text{ για } k = i, j, \end{aligned}$$

αφού $\frac{C_k(t)}{t} \geq \varepsilon$ για $k = i, j$. Επομένως, από την Σχέση (3.6) έχουμε:

$$P[l \text{ είναι } t\text{-connector για } i, j \mid PA_t(m, \delta)] \geq \frac{\varepsilon^2}{4},$$

δηλαδή η πιθανότητα κάποιος από τους κόμβους $t+1, \dots, 2t$ να είναι t -connector για τους i, j που ανήκουν σε διαφορετική συνεκτική συνιστώσα μέχρι την χρονική στιγμή t , είναι μεγαλύτερη από $\frac{\varepsilon^2}{4}$, ανεξάρτητα από τις συνδέσεις που εισήχθησαν από τους κόμβους $t+1, \dots, l-1$.

Από αυτό, η πιθανότητα να μην είναι κανείς από τους $t+1, \dots, 2t$ t -connector για τους i, j είναι:

$$\left(1 - \frac{\varepsilon^2}{4}\right)^t \leq \exp\left(-\frac{\varepsilon^2 t}{4}\right),$$

που τείνει εκθετικά γρήγορα στο 0 καθώς $t \rightarrow +\infty$. Δηλαδή η πιθανότητα να μην υπάρχει t -connector για τους $\forall i, j$ στο $1, \dots, N$ τείνει στο 0 και άρα το $PA_t(m, \delta)$ είναι σχεδόν βεβαίως συνεκτικό. \square

3.4.2 Διάμετρος

Εξετάζουμε πάλι δύο περιπτώσεις:

- Για $m = 1$:

Στην περίπτωση που $m = 1$ ο γράφος που προκύπτει έχει αρκετά απλή δομή: κάθε κόμβος που εισέρχεται στο γράφημα έχει μια ακμή με την οποία είτε θα ενωθεί με κάποιο παλαιότερο κόμβο (κρατώντας το πλήθος των συνεκτικών συνιστωσών σταθερό), είτε θα εισάγει ένα βρόγχο (αυξάνοντας κατά έναν το πλήθος των συνεκτικών συνιστωσών του γραφήματος). Πάντως, σε κάθε περίπτωση, μέσω της διαδικασίας αυτής δεν μπορούμε να έχουμε στο γράφημα κύκλους μήκους $l \geq 2$. Δηλαδή το γράφημα θα είναι ένωση ξένων δέντρων που το καθένα από αυτά θα περιέχει έναν ακριβώς βρόγχο, που ανήκει στον παλαιότερο κόμβο της συνιστώσας αυτής. Ισχύει το ακόλουθο:

Θεώρημα 3.4.3. Έστω γ η λύση της εξίσωσης $\gamma e^{1+\gamma} = 1$ και $\epsilon > 0$. Τότε για σχεδόν κάθε $PA_t(1, \delta) \in \mathcal{F}_n$ η μεγαλύτερη απόσταση μεταξύ δύο κόμβων στην ίδια συνεκτική συνιστώσα του $PA_t(1, \delta)$ είναι μεταξύ $(\gamma^{-1} - \epsilon) \log n$ και $(\gamma^{-1} + \epsilon) \log n$.

Απόδειξη. Η απόδειξη μπορεί να βρεθεί στο [31]. □

- Για $m \geq 2$:

Για γράφους που έχουν προέλθει με τις μεθόδους που αναλύθηκαν προηγουμένως, ισχύει το ακόλουθο:

Θεώρημα 3.4.4. Έστω $\epsilon > 0$. Τότε το $PA_t(m, 0)$ είναι σβ συνεκτικό με διάμετρο:

$$\frac{(1 - \epsilon) \log n}{\log(\log n)} \leq \text{diam}G \leq \frac{(1 + \epsilon) \log n}{\log(\log n)}.$$

Απόδειξη. Για το κάτω φράγμα, υπάρχει απόδειξη στην παράγραφο 4 του [9], ενώ για το άνω φράγμα στην παράγραφο 8, του ίδιου. □

- *Σημείωση:*

Έστω το δίκτυο του παγκόσμιου ιστού, www. Το δίκτυο αυτό είναι πολύ μεγάλου μεγέθους και, όπως είναι φυσικό, είναι σημαντικό να μπορούμε να βρούμε ελάχιστα μονοπάτια ανάμεσα σε δύο ιστοσελίδες - κόμβους, δηλαδή το ελάχιστο πλήθος links που πρέπει να ακολουθηθούν προκειμένου να μεταβεί κανείς από την μια σελίδα στην άλλη. Έπειτα από μελέτες σε δεδομένα για $N = 8 \cdot 10^8$ κόμβους, βρέθηκε ότι το αναμενόμενο πλήθος links για την μετάβαση από μια τυχαία ιστοσελίδα σε μια δεύτερη, επίσης τυχαία, είναι μόλις 18.59! Δηλαδή δύο τυχαία επιλεγμένες ιστοσελίδες στο www είναι μόλις 19 clicks μακριά η μια από την άλλη πράγμα αρκετά ενδιαφέρον δεδομένης της πολύ μεγάλης διάστασης του παγκόσμιου ιστού.

Κεφάλαιο 4

Το «πείραμα»

• Τα δεδομένα

Στο κεφάλαιο αυτό θα παρουσιαστούν τα αποτελέσματα της ανάλυσης ενός κοινωνικού δικτύου. Τα δεδομένα αφορούν φοιτητές που βρίσκονται στο τέλος του 4ο έτους του Προπτυχιακού Τμήματος της Σχολής Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών του Εθνικού Μετσόβιου Πολυτεχνείου. Στο γράφημα που κατασκευάστηκε, κάθε κόμβος αντιστοιχεί σε έναν φοιτητή ή μια φοιτήτρια και μια ακμή από τον i κόμβο στον j αντιστοιχεί στην αναφορά του/της j από τον i φοιτητή/φοιτήτρια στην ερώτηση «Με ποιους συμφοιτητές σου μοιράζεσαι απορίες σχετικά με τα μαθήματα στη σχολή;». Είναι προφανές ότι το γράφημα που προέκυψε από τα δεδομένα που συλλέχθηκαν είναι κατευθυνόμενο, αφού κάποιος φοιτητής μπορεί να έχει αναφερθεί από κάποιον άλλον, χωρίς αυτό να τον υποχρεώνει να αναφέρει τον δεύτερο στα άτομα με τα οποία μοιράζεται τις απορίες του. Επίσης, είναι προφανές ότι στο γράφημα αυτό δεν έχει νόημα να υπάρχουν βρόγχοι, μιας και είναι αυτονόητο ότι κανείς, πρώτου μοιραστεί τις απορίες του, τις έχει επεξεργαστεί ο ίδιος.

Σκοπός είναι να βρεθούν δομές στο γράφημα, αν υπάρχουν, και ίσως και κάποιοι κόμβοι κεντρικότεροι συγκριτικά με τους άλλους. Η ίδια η δομή της Σχολής μας υποδεικνύει κάποια από τα αποτελέσματα που, πριν καν κάνουμε την ανάλυση, περιμένουμε να έχουμε. Η Σχολή είναι πενταετής με τα πρώτα δύο χρόνια της φοίτησης να είναι κοινά για όλους. Στον τρίτο χρόνο φοίτησης, κάθε φοιτήτρια και κάθε φοιτητής καλείται να διαλέξει σε ποια από τις υπάρχουσες κατευθύνσεις θα συνεχίσει τις σπουδές της/του. Οι επιλογές είναι δύο: κανείς μπορεί να συνεχίσει είτε στην κατεύθυνση του Μαθηματικού Εφαρμογών, είτε στην κατεύθυνση του Φυσικού Εφαρμογών. Αφού κανείς επιλέξει κατεύθυνση, καλείται να επιλέξει και δύο ροές από τις υπάρχουσες της κατεύθυνσής του. Οι ροές της κατεύθυνσης «Μαθηματικού» είναι οι: Εφαρμοσμένη Ανάλυση, Στατιστική, Μαθηματικά Πληροφορικής και Εφαρμοσμένη Μηχανική Υπολογιστική Προσομοίωση, ενώ αυτές του «Φυσικού» είναι οι: Υπολογιστική και Θεωρητική Φυσική, Πυρηνική Φυσική και Στοιχειώση Σωματίδια, Οπτοηλεκτρονική και Λείζερ, Προηγμένα Τεχνολογικά Υλικά και Μηχανική των Υλικών.

Έτσι, λοιπόν, στο δείγμα μας ανήκουν άτομα τα οποία έχουν ήδη επιλέξει κατεύθυνση και ροές ήδη από 4 εξάμηνα. Με δεδομένο αυτό και το γεγονός ότι τα μαθήματα για τις δύο κατευθύνσεις δεν είναι κοινά, οπότε ούτε και οι σχετικές με τα μαθήματα απορίες, περιμένουμε να υπάρχει διαχωρισμός στα δεδομένα σε δύο μεγάλες ομάδες: τους «Μαθηματικούς» και τους «Φυσικούς».

Τα στοιχεία που έπρεπε οι ερωτηθέντες να δώσουν ήταν το ονοματεπώνυμό τους, η

κατεύθυνση την οποία επέλεξαν, ο μέσος όρος των μέχρι τότε περασμένων μαθημάτων, το πλήθος τους και, φυσικά, οι άνθρωποι στους οποίους απευθύνονται όταν έχουν κάποια απορία. Για λόγους ιδιωτικότητας, τα ονόματα των συμμετεχόντων έχουν αλλάξει και στη θέση τους έχουν χρησιμοποιηθεί ψευδώνυμα. Κάθε ομοιότητα μεταξύ των αποτελεσμάτων που θα παρουσιαστούν και των φοιτητών και φοιτητριών που απάντησαν στα ερωτηματολόγια ή που με κάποιο τρόπο σχετίζονται με αυτά είναι τυχαίο γεγονός και δεν ανταποκρίνεται στην πραγματικότητα.

• Μια πρώτη προσέγγιση στα δεδομένα

Τα δεδομένα μας αφορούν σε 53 άτομα που απάντησαν στα ερωτηματολόγια που τους δόθηκαν και 26 άτομα που δεν απάντησαν στα ερωτηματολόγια, αλλά υπήρξε αναφορά από κάποιον ή κάποια που απάντησε προς αυτά. Δηλαδή το γράφημα αποτελείται από συνολικά 79 κόμβους. Μεταξύ των κόμβων αυτών υπάρχουν συνολικά 129 ακμές. Δεν υπάρχουν βρόγχοι ή πολλαπλές ακμές. Από αυτά, ο μέσος όρος των (εσωτερικών και εξωτερικών) βαθμών του κάθε κόμβου είναι 3.27. Δηλαδή ο κάθε φοιτητής και η κάθε φοιτήτρια ρωτάει και ρωτάται κατά μέσο όρο από 3.27 φοιτητές.

Στο δείγμα υπάρχουν 46 φοιτητές και φοιτήτριες που επέλεξαν την κατεύθυνση του Μαθηματικού Εφαρμογών και 12 που επέλεξαν αυτή του Φυσικού Εφαρμογών. Ακόμη, δεν έχουμε το δεδομένο αυτό για 20 φοιτητές και 1 κόμβος του γραφήματος αντιστοιχεί στην αναζήτηση πληροφοριών στον παγκόσμιο ιστό.

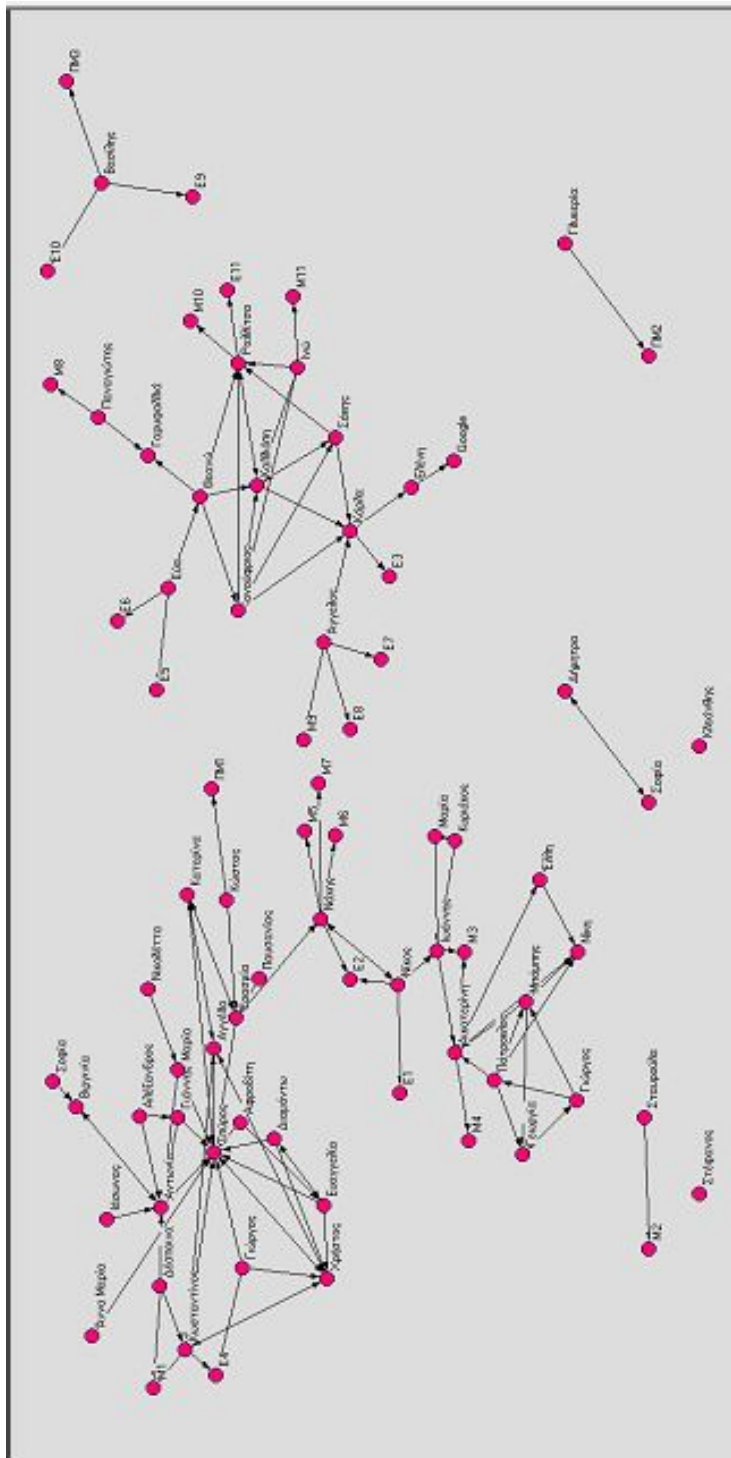
Μια πρώτη προσέγγιση και στους βαθμούς των ερωτηθέντων δίνει ότι ο μεγαλύτερος μέσος όρος στο δείγμα είναι 9, ο μέσος όρος των μέσων όρων του δείγματος είναι 3.7 με τυπική απόκλιση 3.6, καθώς σε αυτό υπολογίζονται ως 0 οι βαθμοί όσων δεν έδωσαν αυτό το δεδομένο, που είναι λίγο λιγότεροι από τους μισούς. Η διάμεσος τιμή στο συγκεκριμένο δεδομένο είναι 6. Το γεγονός πως η διάμεσος είναι πολύ διαφορετική από τη μέση τιμή είναι αναμενόμενο με δεδομένο ότι λίγοι λιγότεροι από τους μισούς δεν έχουν απαντήσει, αφού οι τιμές για το συγκεκριμένο δεδομένο είναι είτε 0 είτε ακέραιοι μεγαλύτεροι ή ίσοι του 5.

Στο Σχήμα (4.1) δίνεται μια αναπαράσταση του γραφήματος που μοντελοποιεί το δίκτυο που περιγράφηκε προηγουμένως.

• Μια πιο λεπτομερής προσέγγιση των δεδομένων

Αρχικά αναζητούμε κόμβους πιο κεντρικούς σε σχέση με τους υπόλοιπους. Αρχικά, να παρατηρήσουμε ότι όσο περισσότερες εισερχόμενες ακμές έχει ένας κόμβος, τόσο πιο πολλοί συμφοιτητές του του εμπιστεύονται τις απορίες τους. Το γεγονός ότι κάποιος μπορεί να έχει μεγαλύτερο εσωτερικό βαθμό από άλλους, τον καθιστά ίσως πιο δημοφιλή και πιο προσιτό μεταξύ των υπόλοιπων. Το βασικότερο, όμως, είναι ότι το να απευθύνονται πολλοί φοιτητές στον i φοιτητή/φοιτήτρια δείχνει ότι ίσως είναι πιο μεταδοτικός όταν εξηγεί, σε σχέση με άλλους, και πιο αξιόπιστος όσων αφορά την ποιότητα των απαντήσεων που θα δώσει. Τέλος, το να έχει κάποιος φοιτητής/φοιτήτρια μεγάλο εσωτερικό βαθμό σε σχέση με άλλους, μπορεί να είναι λόγος να τον αυξήσει ακόμη περισσότερο, μιας και κάποιος ενδέχεται να απευθυνθεί κατευθείαν σε αυτόν επηρεαζόμενος από το γεγονός ότι υπάρχουν ήδη πολλοί που τον εμπιστεύονται. Σε καμία περίπτωση δε σημαίνει ότι είναι καλύτερος φοιτητής από άλλους, αν και είναι αναμενόμενο αυτό να συμβαίνει σε κάποιο βαθμό, αφού συνήθως κάνουμε ερωτήσεις σε κάποιον που θεωρούμε ότι θα είναι σε θέση να μας απαντήσει.

Όπως φαίνεται και από τον Πίνακα (4), υπάρχουν 19 φοιτητές που κανείς δεν ρωτάει όταν έχει απορίες, 34 που τους ρωτά ένας συμφοιτητής τους, 11 που τους ρωτούν



Σχήμα 4.1: Γράφημα συσχέτισης φοιτητών 4ου έτους.

Dimension: 79
 The lowest value: 0
 The highest value: 13

Frequency distribution of cluster values:

Cluster	Freq	Freq%	CumFreq	CumFreq%	Representative
0	19	24.0506	19	24.0506	Αννα Μαρία
1	34	43.0380	53	67.0886	Ελένη
2	11	13.9241	64	81.0127	Βιργινία
3	4	5.0633	68	86.0759	Κατερίνα
4	4	5.0633	72	91.1392	Ερασμία
5	4	5.0633	76	96.2025	Αικατερίνη
6	2	2.5316	78	98.7342	Αντωνία
13	1	1.2658	79	100.0000	Σπύρος
Sum	79	100.0000			

Πίνακας 4.1: Πίνακας με τους εσωτερικούς βαθμούς των κόμβων.

Weak Components

working...
 Number of components: 8
 Size of the largest component: 44 vertices (55.696%).
 Time spent: 0:00:00

Πίνακας 4.2: Ασθενώς Συνεκτικές Συνιστώσες.

δύο συμφοιτητές τους 4 που τους ρωτούν 3, ακόμη 4 που ρωτώνται από 4 και ακόμη 4 που ρωτώνται από 5. Υπάρχουν 2 φοιτήτριες/φοιτητές που τους εμπιστεύονται 6 και μόλις ένας που, με μεγάλη διαφορά από τους υπόλοιπους, τον εμπιστεύονται 13 συμφοιτητές του. Στην πρώτη στήλη του πίνακα φαίνεται το πλήθος των συμφοιτητών του *i* φοιτητή/φοιτήτριας που τον/την εμπιστεύεται, στη δεύτερη στήλη φαίνεται η συχνότητα με την οποία αυτό συμβαίνει, στην τρίτη στήλη υπάρχει η σχετική συχνότητα. Στην επόμενη στήλη υπάρχει το άθροισμα των φοιτητών που ρωτώνται από φοιτητές που στο πλήθος είναι το πολύ όσο η αντίστοιχη τιμή της πρώτης στήλης και στην απόμενη στήλη από αυτή υπάρχουν οι σχετικές συχνότητες της ίδιας ποσότητας. Τέλος, στην τελευταία στήλη, υπάρχει ένας τυχαία επιλεγμένος αντιπρόσωπος από κάθε μια από τις παραπάνω ομάδες φοιτητών.

Συνεχίζοντας την ανάλυση του γραφήματος των φοιτητών του 4ου έτους, αναζητούμε τις (ασθενώς) συνεκτικές συνιστώσες του γραφήματος. Από τον πίνακα (4) λαμβάνουμε ότι το πλήθος τους είναι 8 και η μεγαλύτερη από αυτές αποτελείται από 44 κόμβους-φοιτητές.

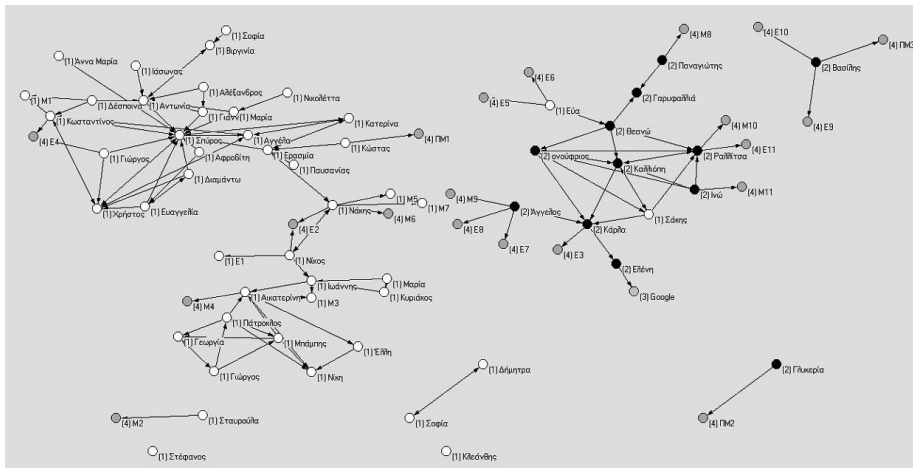
Μια ερμηνεία γι' αυτές είναι ότι αν υπάρχει μια πληροφορία που ξέρουν κάποιοι μόνο στη μια από αυτές, τότε αυτή η πληροφορία δεν μπορεί να περάσει σε φοιτητές εκτός της συνιστώσας στην οποία ανήκουν. Για παράδειγμα, αν η Δήμητρα και η Γλυκερία μάθουν κάτι σχετικό με κάποιο μάθημα και το αναφέρουν μόνο στα άτομα με τα οποία συζητάνε τις απορίες τους και τα οποία τους εμπιστεύονται τις δικές τους, τότε την πληροφορία αυτή τελικά θα την γνωρίζουν μόνο οι δυο τους, η Σοφία και ο/η ΠΜ2.

Από τον Πίνακα (4) λαμβάνουμε κάποιες παραπάνω πληροφορίες για τις ασθενώς συνεκτικές συνιστώσες του γραφήματος. Για παράδειγμα, λαμβάνουμε ότι στην μεγα-

Frequency distribution of cluster values:

Cluster	Freq	Freq%	CumFreq	CumFreq%	Representative
1	44	55.6962	44	55.6962	Βιργιλνία
2	23	29.1139	67	84.8101	Ελένη
3	2	2.5316	69	87.3418	Δήμητρα
4	2	2.5316	71	89.8734	Σταυρούλα
5	1	1.2658	72	91.1392	Στέφανος
6	1	1.2658	73	92.4051	Κλεάνθης
7	2	2.5316	75	94.9367	Γλυκερία
8	4	5.0633	79	100.0000	Βασίλης
Sum	79	100.0000			

Πίνακας 4.3: Για τις Ασθενώς Συνεκτικές Συνιστώσες.

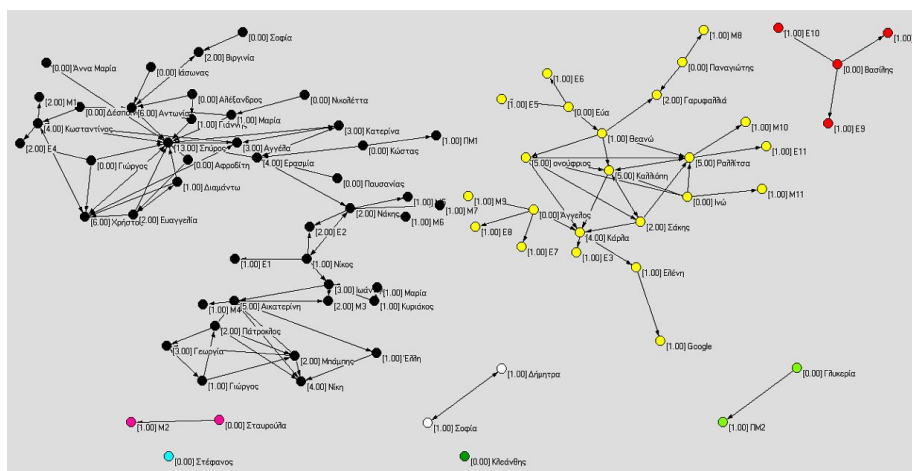


Σχήμα 4.2: Κατευθύνσεις που έχουν ακολουθηθεί από τους φοιτητές.

λύτερη συνεκτική συνιστώσα υπάρχουν 44 φοιτητές, στην αμέσως επόμενη υπάρχουν 23, στη συνέχεια 4 φοιτητές, υπάρχουν 3 συνιστώσες με 2 φοιτητές η καθεμια και 2 απομονωμένοι κόμβοι, δηλαδή φοιτητές που ούτε ρωτάνε τους συμφοιτητές τους ούτε οι συμφοιτητές τους τους ρωτάνε.

Κάνοντας ένα ακόμη βήμα, αναζητούμε τη συσχέτιση των ασθενώς συνεκτικών συνιστωσών του γραφήματος και των κατευθύνσεων που οι φοιτητές έχουν επιλέξει να ακολουθήσουν. Αρχικά, στο Σχήμα (4) μπορούμε να δούμε και οπτικά κάποια από αυτά που περιγράφηκαν προηγουμένως. Στο γράφημα οι άσπροι κόμβοι αναπαριστούν φοιτητές που έχουν επιλέξει την κατεύθυνση Μαθηματικού Εφαρμογών, οι μαύροι κόμβοι αυτούς που έχουν επιλέξει την κατεύθυνση Φυσικού εφαρμογών και οι γκρι αυτούς που δεν γνωρίζουμε ποια από τις δύο κατευθύνσεις έχουν επιλέξει.

Μπορούμε να παρατηρήσουμε ότι υπάρχουν αρκετά περισσότεροι φοιτητές που επέλεξαν την κατεύθυνση Μαθηματικού Εφαρμογών συγκριτικά με το πλήθος αυτών που έχουν επιλέξει αυτή του Φυσικού Εφαρμογών. Ακόμη, ο συνδυασμός του γραφήματος αυτού και του γραφήματος στο Σχήμα (4) δίνει την πληροφορία ότι οι φοιτητές του δείγματος χωρίζονται όντως σε δυο μεγάλες ομάδες, όπως ακριβώς υπονιαζόμε-



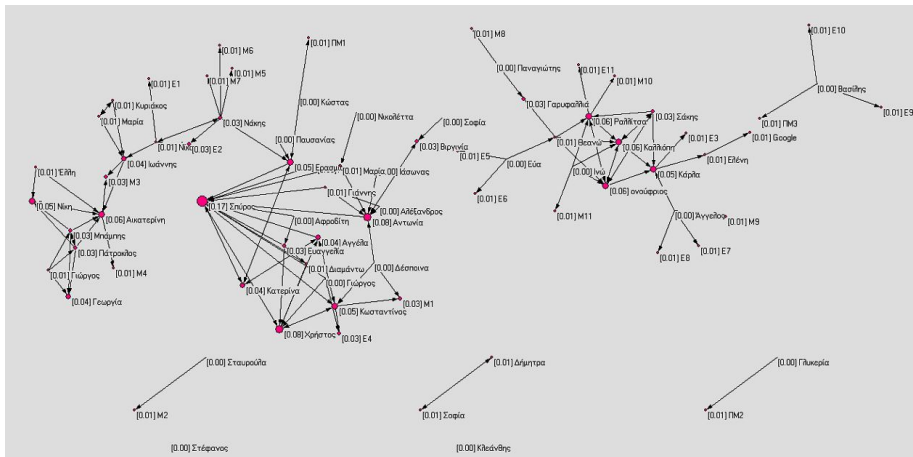
Σχήμα 4.3: Κατευθύνσεις και ασθενώς συνεκτικές συνιστώσες.

στην πριν ξεκινήσουμε την ανάλυση. Παρατηρούμε ότι όλοι οι κόμβοι για τους οποίους έχουμε πληροφορία σχετικά με την κατεύθυνση που ακολούθησαν στη συνιστώσα με το μαύρο χρώμα είναι φοιτητές του Μαθηματικού. Παρόμοια για την κίτρινη συνιστώσα, οι περισσότεροι από τους φοιτητές που βρίσκονται εκεί είναι φοιτητές που επέλεξαν την κατεύθυνση του Φυσικού. Παρατηρούμε, ακόμη, ότι στην κίτρινη συνιστώσα του τελευταίου Σχήματος, υπάρχουν δύο Μαθηματικοί και κάποιος για τους οποίους δε γνωρίζουμε. Το φαινόμενο αυτό, δηλαδή οι κόμβοι ενός γραφήματος να είναι συνδεδεμένοι σε μεγάλο βαθμό με όμοιους τους, καλείται «ομοφιλία».

Σχετικά με τα προηγούμενα, μπορούμε να παρατηρήσουμε και κάποια πιο συγκεκριμένα πράγματα. Για παράδειγμα, έστω ο φοιτητής ονόματι «Σάκης», που ακολούθησε την κατεύθυνση του Μαθηματικού Εφαρμογών. Παρατηρούμε ότι όταν έχει απορίες ρωτάει τις Ραλλίτσα και Κάρλα, που είναι φοιτήτριες στην κατεύθυνση Φυσικού. Ακόμη, βλέπουμε ότι τον ρωτάνε για απορίες η Καλλιόπη και ο Ονούφριος, που και αυτοί με τη σειρά τους είναι στο Φυσικό. Από αυτό υποψιαζόμαστε ότι μπορεί αυτό να συμβαίνει επειδή ίσως και οι 5 αυτοί φοιτητές ασχολούνται με μαθήματα που ανήκουν στον κορμό και, άρα, γι' αυτό μοιράζονται τις απορίες τους παρ' ότι ακολουθούν διαφορετικές κατευθύνσεις. Συμβουλευόμαστε τα δεδομένα μας και παρατηρούμε ότι και οι 5 αυτοί φοιτητές έχουν περάσει λίγα μαθήματα. Το τελευταίο μπορεί να σημαίνει ότι άρχισαν να ασχολούνται με τη σχολή αργότερα σχετικά με τους υπόλοιπους φοιτητές του έτους και ακόμη ασχολούνται με μαθήματα που ανήκουν στον κορμό.

Κάτι άλλο, παρόμοιο με το προηγούμενο, που μπορούμε να παρατηρήσουμε από το τελευταίο γράφημα, είναι πως και η φοιτήτρια ονόματι «Γλυκερία» απευθύνεται για τις απορίες της σε μια μόνο φοιτήτρια, άγνωστης σε εμάς κατεύθυνσης. Το μόνο που γνωρίζουμε για τη δεύτερη είναι ότι ανήκει σε μικρότερο έτος από τη Γλυκερία. Αυτό, όπως και πριν, ίσως προδίδει πως και η Γλυκερία ασχολείται με τα μαθήματα του κορμού, πράγμα που επιβεβαιώνουμε από το δεδομένο ότι έχει περάσει λίγα μόνο μαθήματα.

Με βάση αυτά, μπορούμε να θεωρούμε την κίτρινη συνιστώσα, συνιστώσα Φυσικών και τη μαύρη, συνιστώσα Μαθηματικών. Από τα δύο γραφήματα, παρατηρούμε



Σχήμα 4.4: Εσωτερικός Βαθμός φοιτητών.

ότι ανάμεσα στις δύο συνιστώσες δεν υπάρχει ούτε μια γέφυρα! Το γεγονός ότι μάλλον¹ κανείς Μαθηματικός εκτός από το Σάκη δεν απευθύνεται σε κάποιον Φυσικό για τις απορίες του δείχνει ότι η μεγαλύτερη μερίδα των φοιτητών του Μαθηματικού που έχουμε στο δείγμα μας ασχολούνται με τα μαθήματα του έτους στο οποίο φοιτούν και τυπικά, πράγμα που επιβεβαιώνεται και από τα δεδομένα, μιας και το μεγαλύτερο μερίδιο αυτών έχει μεγάλο πλήθος περασμένων μαθημάτων.

• Αναζητώντας τους «κεντρικότερους» φοιτητές

Στην ενότητα αυτή αναζητώνται οι «κεντρικότεροι» φοιτητές του γραφήματος, όπως αυτοί έχουν περιγραφεί σε προηγούμενη παράγραφο. Στα Σχήματα (4),(4) και (4) δίνονται τα γραφήματα με τους φοιτητές του δικτύου που μελετάται, στα οποία μεγαλύτερες κουκκίδες αντιστοιχούν σε κεντρικότερο κόμβο για το μέτρο του εσωτερικού βαθμού, της εγγύτητας και της διαμεσολαβησιμότητας αντίστοιχα. Τα γραφήματα ακολουθούν και οι αντίστοιχοι πίνακες.

Ακολουθούν οι 10 κεντρικότεροι φοιτητές με βάση τα διάφορα μέτρα.

Παρατηρούμε ότι ο Σπύρος είναι και για τα τρία μέτρα από τους κεντρικότερους κόμβους, όπως και η Αντωνία, η Ερασμία και ο Κωσταντίνος. Σχετικά με τον εσωτερικό βαθμό, δηλαδή το πλήθος των συμφοιτητών και συμφοιτητριών του που εμπιστεύονται κάποιον ή κάποια, παρατηρούμε ότι ο Σπύρος είναι με διαφορά ο κεντρικότερος συγκριτικά με την Αντωνία, που είναι η δεύτερη στην κατάταξη, και το Χρήστο. Σχετικά με το μέρο της διαμεσολαβητικότητας, παρατηρούμε ότι πρώτη είναι η Ερασμία αλλά με μικρή διαφορά από το Σπύρο, που έρχεται δεύτερος. Ο Χρήστος, παρότι είναι στις πρώτες θέσεις για τα μέτρα του εσωτερικού βαθμού και της εγγύτητας, είναι αρκετά χαμηλά για το μέτρο της διαμεσολαβητικότητας.

Ένα ακόμη πράγμα που μπορούμε να παρατηρήσουμε είναι το γεγονός ότι παρ' ότι

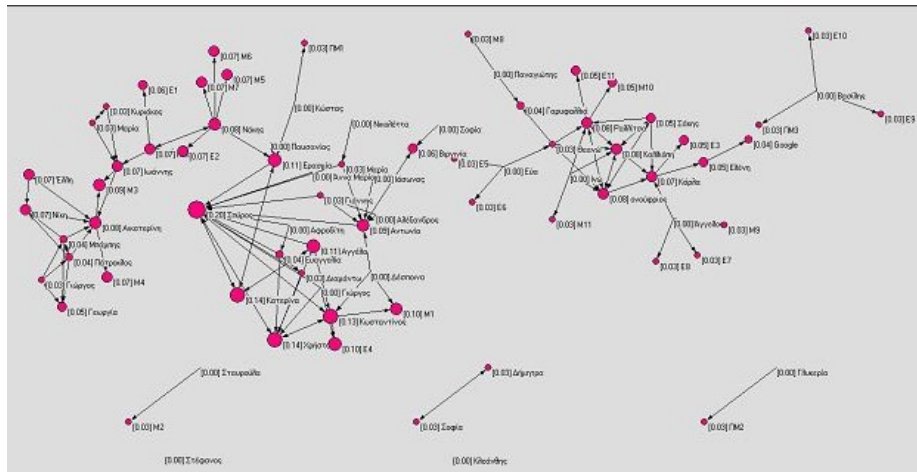
¹ Δεν μπορούμε να το πούμε με βεβαιότητα, καθώς υπάρχουν και κάποιοι γκρι κόμβοι, οι οποίοι δεν γνωρίζουμε ποια κατεύθυνση έχουν ακολουθήσει. Από την άλλη, έχουμε κάποιες παραπάνω ενδείξεις ότι κάτι τέτοιο μπορεί να είναι αληθές, μιας και οι γκρι κόμβοι της συνιστώσας των Μαθηματικών έχουν όλοι τους αρκετά περασμένα μαθήματα.

Dimension: 79
 The lowest value: 0
 The highest value: 13

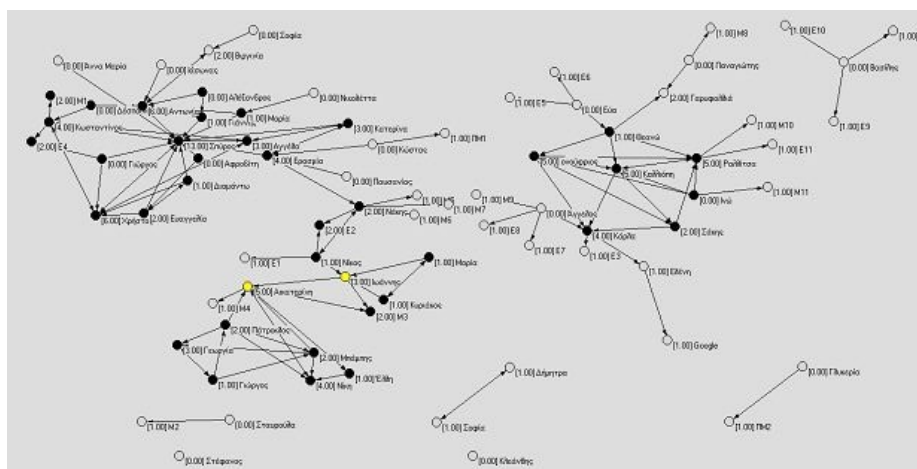
Frequency distribution of cluster values:

Cluster	Freq	Freq%	CumFreq	CumFreq%	Representative
0	19	24.0506	19	24.0506	Αννα Μαρία
1	34	43.0380	53	67.0886	Ελένη
2	11	13.9241	64	81.0127	Βιργινία
3	4	5.0633	68	86.0759	Κατερίνα
4	4	5.0633	72	91.1392	Ερασμία
5	4	5.0633	76	96.2025	Αικατερίνη
6	2	2.5316	78	98.7342	Αντωνία
13	1	1.2658	79	100.0000	Σπύρος
Sum	79	100.0000			

Πίνακας 4.4: Εσωτερικός Βαθμός φοιτητών.



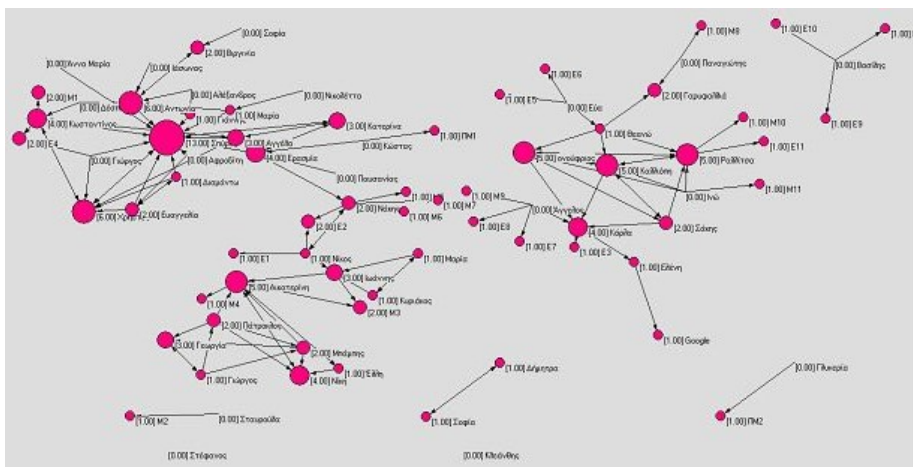
Πίνακας 4.5: Εγγύτητα φοιτητών.



Σχήμα 4.6: Γράφημα με Articulation Points, κρίσιμους για τη διάδοση της πληροφορίας φοιτητές στο γράφημα.

ο Ιωάννης και η Αικατερίνη είναι μόλις στην 6η και 7η θέση στην κατάταξη για τη διαμεσολαβητικότητα, όπως φαίνεται και στο Γράφημα (4), διαδραματίζουν έναν πολύ σημαντικό ρόλο στη διάδοση της πληροφορίας στο γράφημά μας. Από το γράφημα αυτό μπορούμε να παρατηρήσουμε ότι είναι οι μόνοι δύο κόμβοι που αν αφαιρεθούν από το γράφημα τότε αυξάνεται το πλήθος των συνεκτικών συνιστωσών. Η ιδιότητά τους αυτή τους καθιστά ιδιαίτερα σημαντικούς κόμβους του γραφήματος καθώς, αν κάτι τέτοιο συμβεί, τότε θα υπάρξουν ομάδες φοιτητών που θα είναι πλέον μεταξύ τους απομονωμένες. Τέτοιοι κόμβοι ονομάζονται «articulation points».

Κοιτάζοντας το Γράφημα του Πίνακα (??) θα μπορούσαμε να παρατηρήσουμε διάφορα σχετικά με τη διάδοση μια πληροφορίας και τη σχέση της με τον εσωτερικό βαθμό του κάθε κόμβου. Από το γράφημα βλέπουμε, όπως είχαμε δει και σε προηγούμενο, ακριβώς αυτά που περιγράφονται από τον Πίνακα (4). Συμπεραίνουμε, δηλαδή, ότι ο ελάχιστος αριθμός φοιτητών που αρκεί να μάθουν μια πληροφορία για να μεταδοθεί αυτή η πληροφορία στο μέγιστο δυνατό αριθμό φοιτητών του γραφήματος γρήγορα (με λίγα βήματα) είναι αρκετά μικρός, με την προϋπόθεση ότι όταν κανείς μάθει την πληροφορία τη μεταδίδει σε όσους μοιράζονται τις απορίες τους μαζί του). Από το προηγούμενο γράφημα και αυτό του Πίνακα (4), παρατηρούμε ότι αρκεί να ενημερωθούν δύο φοιτητές για ένα ζήτημα και αυτό πολύ γρήγορα θα γίνει γνωστό στις δύο μεγάλες συνεκτικές συνιστώσες. Οι δύο αυτοί φοιτητές θα είναι, προφανώς, ένας από κάθε ασθενώς συνεκτική συνιστώσα. Στο γράφημα του Πίνακα (4) παρατηρούμε ότι μπορεί να υπάρξει διάδοση της πληροφορίας από τη μικρή ροζ συνιστώσα στη μεγάλη γαλάζια, αλλά όχι αντίστροφα. Ακόμη, από την κίτρινη συνιστώσα μπορεί να διαδοθεί πληροφορία προς την Αικατερίνη (και αντίστροφα) και από αυτήν προς τη μαύρη συνιστώσα, αλλά όχι αντίστροφα. Τέλος, παρατηρούμε ότι από τον άσπρο κόμβο M3 μπορεί να διαδοθεί πληροφορία προς την Αικατερίνη και τον Ιωάννη. Με βάση αυτά, μπορούμε με ασφάλεια να πούμε ότι αρκεί να ενημερώσουμε τον M3 για ένα γεγονός και τότε αυτό θα διαδοθεί σε όλη την ασθενώς συνεκτική συνιστώσα του «Μαθηματικού». Τα βήματα που απαιτούνται για να συμβεί αυτό φαίνονται στον Πίνακα (??). Βλέπουμε ότι σε το πολύ 9 βήματα, η πληροφορία θα έχει διαδοθεί σε σχεδόν όλους τους



Σχήμα 4.7: Γράφημα με τους φοιτητές που μεταδίδουν πληροφορία πιο συχνά.

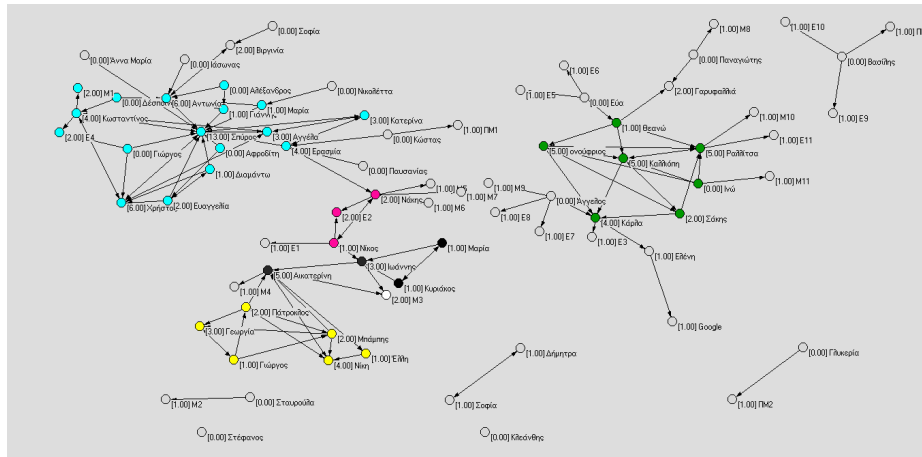
0	M3
1	Αικατερίνη, Ιωάννης
2	Κυριάκος, Νίκη, Έλλη, Νίκος, Μαρία, Πάτροκλος, Μπάμπης
3	Νάκης, Γιώργος
4	Ερασμία, Γεωργία
5	Κατερίνα, Κώστας, Πausανίας
6	Σπύρος, Αγγέλα
7	Άννα Μαρία, Γιάννης, Αντωνία, Μαρία, Κωσταντίνος, Χρήστος, Γιώργος, Ευαγγελία, Αφροδίτη, Διαμάντω
8	Βυργινία, Αλέξανδρος, Ιάσωνας, Δέσποινα, Νικολέττα
9	Σοφία

Πίνακας 4.7: Πίνακας με τις αποστάσεις από τον κόμβο M3.

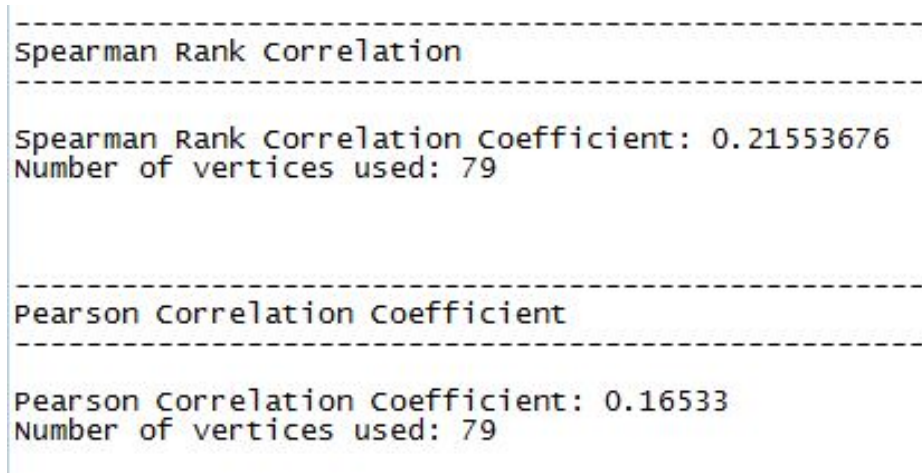
φοιτητές του Μαθηματικού. Αντίστοιχα αποτελέσματα λαμβάνουμε και για την ίδια επεξεργασία της συνεκτικής συνιστώσας των φοιτητών του Φυσικού. Με τον τρόπο αυτό, έχοντας δώσει σε δύο μόνο φοιτητές μια πληροφορία, μπορούμε να είμαστε σίγουροι ότι αυτοί θα μεταδοθεί στο σύνολο σχεδόν των φοιτητών. Να παρατηρήσουμε, τέλος, ότι ο M3 έχει εσωτερικό βαθμό μόλις 2. Από τα παραπάνω συμπεραίνουμε ότι ο εσωτερικός βαθμός δεν σχετίζεται απαραίτητα με τη σημαντικότητά του ως προς τη διάδοση πληροφοριών, και φαινομένων γενικότερα, στο γράφημα.

Τελειώνοντας, θα ήταν ενδιαφέρον να δούμε αν και πόσο σχετίζεται ο μέσος όρος των βαθμών κάποιου φοιτητή στα μαθήματα με τον εσωτερικό του βαθμό στο γράφημα. Αυτό που διαισθητικά περιμένουμε να δούμε είναι ότι θα υπάρχει μια θετική συσχέτιση μεταξύ των δύο ποσοτήτων και αυτό διότι συνήθως τείνουμε να ζητάμε βοήθεια από ανθρώπους που θεωρούμε ότι ίσως να μπορούν να μας τη δώσουν. Αν και δεν είναι απόλυτο, τέτοιοι άνθρωποι περιμένουμε να είναι κάποιοι που έχουν συγκριτικά μεγαλύτερους βαθμούς από άλλους.

Στο Rajek μπορούν να υπολογιστούν δύο τύποι συντελεστών συσχέτισης: ο συντελεστής συσχέτισης ταξινομήσεων του Spearman και ο συντελεστής συσχέτισης του Pearson. Ο πρώτος καθορίζει αν η ταξινόμηση των τιμών της πρώτης ποσότητας (εσωτερικός βαθμός) είναι σύμφωνη με αυτήν της δεύτερης ποσότητας (μέσος όρος βαθμο-



Σχήμα 4.8: Γράφημα με τις ισχυρά συνεκτικές συνιστώσες του γραφήματός μας.



Σχήμα 4.9: Αποτελέσματα που αφορούν στη συσχέτιση βαθμολογίας και εσωτερικών βαθμών.

λογίας) και στον καθορισμό αυτό δεν παίζουν ρόλο οι τάξεις μεγέθους των τιμών των δύο ποσοτήτων, αρκεί να μπορεί να υπάρξει ταξινόμηση ανάμεσά τους. Ο δεύτερος θεωρεί μια γραμμική σχέση μεταξύ των δύο ποσοτήτων, πράγμα που σημαίνει ότι για καθε μοναδιαία αύξηση της μιας ποσότητας υπήθεται σταθερή αυξομείωση της άλλης.

Πράγματι, από τα αποτελέσματα που φαίνονται στο Σχήμα (4), επιβεβαιώνουμε τη διαίσθησή μας. Μεταξύ των δύο ποσοτήτων που προαναφέρθηκαν, υπάρχει μια μικρή θετική συσχέτιση. Ο συντελεστής συσχέτισης του Spearman έχει τιμή 0.215. Δηλαδή, η αύξηση του εσωτερικού βαθμού κατά μια μονάδα είναι ικανή να εξηγήσει κατά 21.5% την αύξηση του μέσου όρου των βαθμολογιών. Ο συντελεστής συσχέτισης του Pearson έχει τιμή 0.165, που σημαίνει ότι για κάθε αύξηση κατά μια μονάδα του μέσου όρου στις βαθμολογίες, αναμένεται να έχουμε αύξηση κατά μια σταθερή ποσότητα. Το ότι ο δεύτερος είναι κάπως μικρότερος από τον πρώτο υποδεικνύει ότι η σχέση που υπάρχει ανάμεσα στις δύο αυτές ποσότητες δεν είναι γραμμική.

Βιβλιογραφία

- [1] R.Albert and A.L.Barabási. *Mechanics Statistics of Complex Networks*, 2001.
- [2] A.L.Barabási and R.Albert. *Emergence of Scaling in Random Networks*. , Elsevier Science, 1999.
- [3] A.L.Barabási, R.Albert, H.Jeong. *Mean-field theory for Scale-free Random Networks*. Elsevier Science, 1999.
- [4] A.L.Barabási, R.Albert, H.Jeong. *Scale-free characteristics of random networks: the topology of the world-wide web.*, PHYSICA A 281, 2000.
- [5] A.L.Barabási, R.Albert, H.Jeong. *Diameter of the world-wide web*. Macmillan Magazines Ltd, 1999.
- [6] Patrick Billingsley. *Probability and measure.*, Third Edition. The University of Chicago, 1995.
- [7] B.Bollobás. *Random Graphs*, Cambridge Studies in Advanced Mathematics, 2001.
- [8] B.Bollobás. *Modern Graph Theory*, Springer, Graduate Texts in Mathematics, 1998.
- [9] B.Bollobás, O.Riordan. *The Diameter of a Scale-free Random Graph.*, COMBINATORICA 24 (1), 2004.
- [10] B.Bollobás, O.Riordan, J.Spencer, G.Tusnády. *The Degree Sequence of a Scale-Free Random Graph Process*. J. Wiley & Sons, Inc, 2001.
- [11] Steve Borgatti. *Lecture on Centrality.*, 2005.
- [12] S.P.Borgatti. *Centrality and network flow.*, Science Direct, 2005.
- [13] F.Chung, L.Lu *The diameter of sparse random graphs*. Academic Press University of California, 2000.
- [14] P.Erdős, A.Rényi. *On the strength of connectedness of a random graph*. Budapest Academy, 1960.
- [15] P. Erdős, A.Rényi. *On the evolution of random graphs.*, Budapest, 1959.
- [16] P. Erdős, A.Rényi. *On random graphs I*. Budapest, 1958.
- [17] P. Erdős and A.Rényi. *On random matrices*. Budapest, 1964.

- [18] A.Goel, R.Goodman. *Lecture 10: PageRank*. Stanford, 2009.
- [19] R.A.Hanneman, M.Riddle *Introduction to social network methods.*, chapter 10: Centrality and power.
- [20] R.v.d.Hofstad. *Random Graphs and Complex Networks. Vol. I*. Department of Mathematics and Computer Science, Eindhoven University of Technology,2014.
- [21] R.v.d.Hofstad. *Random Graphs and Complex Networks. Vol. II*. Department of Mathematics and Computer Science, Eindhoven University of Technology,2014.
- [22] R.v.d.Hofstad. *Mini-Course on Random Graphs and Complex Networks*. Athens, 2014.
- [23] B.A.Huberman, L.A.Adamic *Growth Dynamics of the world-wide web*. Macmillan Magazines Ltd, 1999.
- [24] A.Mrvar, V.Batagelj. *Pajek & Pajek-XXL*. University of Ljubljana, Slovenia, 2003.
- [25] A.Mvar. *Network Analysis using Pajek*. University of Ljubljana, Slovenia, 2003.
- [26] D.A.McFarland, S.Messing, M.Nowak and S.J.Westwood. *Social network analysis labs in R and SoNIA*.
- [27] Dr.Cecilia Mascolo. *Social and Technological Network Analysis, Lecture on Networks, Random Graphs and Metrics*.
- [28] W.d.Nooy, A. Mrvar, V.Batagelj. *Exploratory Social Network Analysis with Pajek.*, Cambridge, 2000.
- [29] M.O.Jackson. *Social and Economic Networks*. Princeton University Press, 2008.
- [30] T.Opsahl. *Closeness centrality in networks with disconnected components.*, 2010
- [31] Pittel. *Note on the height of random recursive trees*, 1994.
- [32] Shlomo Sternberg. *Lecture on the Perron-Frobenius theorem*.
- [33] J.Spencer. *Lectures on Random Graphs*. Courant Istitute of Mathematics.
- [34] www.cs.middlebury.edu/rfaseit/searchengines.
- [35] trec.nist.gov
- [36] www.ndsu.edu/pubweb/novozhil/teaching/767.20data/54.pdfsam.notes.pdf
- [37] *Virtual Laboratories in Probability & Statistics Finite Sampling Models, Pólya's Urn Process*.
- [38] Γ.Κοκολάκης, Ι.Σπηλιώτης. *Εισαγωγή στις Πιθανότητες, 1η έκδοση*. Εκδόσεις Συμμεών, 2002.