



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ
ΕΠΙΣΤΗΜΩΝ

Πιθανοθεωρητικές Μέθοδοι στην Επιστήμη των Υπολογιστών

Η Μεταβολή Φάσης στο Τυχαίο Γράφημα

Διπλωματική Εργασία

Γεωργίου Αυδή

Επιβλέπων

Μιχαήλ Λουλάκης, Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2021



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ
ΕΠΙΣΤΗΜΩΝ

Πιθανοθεωρητικές Μέθοδοι στην Επιστήμη των Υπολογιστών

Η Μεταβολή Φάσης στο Τυχαίο Γράφημα

Διπλωματική Εργασία

Γεωργίου Αυδή

Επιβλέπων

Μιχαήλ Λουλάκης, Αναπληρωτής Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 6^η Ιουλίου 2021.

.....
Μιχαήλ Λουλάκης
Αν. Καθηγητής Ε.Μ.Π.

.....
Δημήτριος Φουσκάκης
Καθηγητής Ε.Μ.Π.

.....
Αριστείδης Παγουρτζής
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2021

.....
Γεώργιος Αυδής

Διπλωματούχος Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

Copyright © Γεώργιος Αυδής, 2021

Με επιφύλαξη παντός δικαιώματος. All rights Reserved.

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

Οι απόψεις και τα συμπεράσματα που περιέχονται σ' αυτό το έγγραφο εκφράζουν το συγγραφέα και δεν πρέπει να ερμηνευτεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Σε αυτή τη διπλωματική εργασία μελετάται το τυχαίο γράφημα $G(n, p)$ καθώς το n μεγαλώνει και το p είναι μια συνάρτησή του. Συγκεκριμένα, όταν το $p = \lambda/n$ παρατηρείται η εξής συμπεριφορά του μεγέθους της μεγιστοτικής σε μέγεθος συνεκτικής συνιστώσας ($|C_{\max}|$) του τυχαίου γραφήματος: όταν λ είναι μια σταθερά μικρότερη του 1 το $|C_{\max}| = \Theta(\log n)$, όταν είναι μια σταθερά μεγαλύτερη του 1 το $|C_{\max}| = \Theta(n)$ και όταν λ είναι μια κατάλληλα παραμετροποιημένη συνάρτηση του n έτσι ώστε καθώς το $n \rightarrow \infty$ $\lambda \rightarrow 1$ το $|C_{\max}| = \Theta(n^{2/3})$. Αυτό το φαινόμενο δικαιολογεί και τον όρο 'Μεταβολή Φάσης' που έχει δοθεί σε αυτή τη συμπεριφορά του τυχαίου γραφήματος. Επίσης, δίνονται μερικά αποτελέσματα για την κατανομή των βαθμών των κορυφών στο τυχαίο γράφημα καθώς και μια ανάλυση των ουρών αναμονής η οποία κρίνεται θεμελιώδης για τη μελέτη του. Παρόλα αυτά, η κατανομή βαθμών του $G(n, p)$ δε συνάδει με αυτή που παρατηρείται στα πραγματικά δίκτυα και έτσι δίνεται μια εισαγωγή σε μοντέλα που λειτουργούν με επιλεκτική σύνδεση, αποδεικνύοντας πως η κατανομή βαθμών που παράγεται από αυτά προσομοιάζει καλύτερα τα αποτελέσματα που δίνουν τα εμπειρικά δεδομένα.

Λέξεις - Κλειδιά

τυχαίο γράφημα, μοντέλο Erdős-Rényi, ουρές αναμονής, κλαδωτές διαδικασίες, μεταβολή φάσης, κατανομή βαθμών, επιλεκτική σύνδεση, μοντέλο Barabási-Albert

Abstract

In this diploma thesis the random graph $G(n, p)$ is studied as n increases and p is a function of n . In particular, when $p = \lambda/n$ the following behavior of the size of the maximal in size connected component ($|C_{\max}|$) of the random graph is observed: when λ is a constant less than 1 then $|C_{\max}| = \Theta(\log n)$, when λ is a constant greater than 1 then $|C_{\max}| = \Theta(n)$ and when λ is an aptly parametrized function of n such that as $n \rightarrow \infty$, $\lambda \rightarrow 1$ then $|C_{\max}| = \Theta(n^{2/3})$. This phenomenon justifies the term “Phase Transition” which is given to this behavior of the random graph. Moreover, some results concerning the degree distribution of the vertices in the random graph are given along with an analysis of queues which is substantial for the study of the random graph. Nonetheless, the degree distribution of $G(n, p)$ does not comply with the one that is observed in real-world networks and so an introduction to models that function with preferential attachment is given, proving that the degree distribution that is produced by them mimics better the results that are given by empirical data.

Keywords

random graph, Erdős-Rényi model, queues, branching processes, phase transition, degree distribution, preferential attachment, Barabási-Albert model

Ευχαριστίες

Θέλω να ευχαριστήσω θερμά τον κύριο Μιχάλη Λουλάκη ο οποίος ως διδάσκων του μαθήματος των Πιθανοτήτων μου κίνησε το ενδιαφέρον για αυτόν τον επιστημονικό χώρο. Επίσης, τον ευχαριστώ για όλη την καθοδήγηση που μου προσέφερε όσο καιρό προετοιμαζόμουν για τη διπλωματική εργασία, καθώς και για τις χρήσιμες συμβουλές του κατά την εκπόνησή της. Ήρθα έτσι σε επαφή με διάφορα μαθηματικά εργαλεία όπως η πιθανοθεωρητική μέθοδος, το τυχαίο γράφημα, το configuration model, το μοντέλο Barabási-Albert κ.ά., την ύπαρξη των οποίων αγνούσα προηγουμένως.

Οφείλω επίσης ένα ιδιαίτερο ευχαριστώ στον κύριο Στάθη Ζάχο ο οποίος με το χαρακτηριστικό του τρόπο και την αστείρευτη ενέργειά του μου εξήγησε την περιέργεια ώστε να ανακαλύψω το συναρπαστικό κόσμο της Θεωρητικής Επιστήμης των Υπολογιστών. Επιπλέον, ευχαριστώ όλους τους καθηγητές και το βοηθητικό προσωπικό του Εθνικού Μετσόβιου Πολυτεχνείου που με τον έναν ή τον άλλον τρόπο συνεισέφεραν στην εξέλιξή μου κατά τη διάρκεια των προπτυχιακών σπουδών μου.

Τέλος, ευχαριστώ την οικογένειά μου και τους φίλους μου που με βοηθούν όταν το χρειάζομαι.

Περιεχόμενα

1	Εισαγωγή	1
2	Το Τυχαίο Γράφημα	3
2.1	Το κλασικό τυχαίο γράφημα των Erdős και Rényi	3
2.2	Η συχνότερη μορφή του	3
2.3	Κάποιες στοιχειώδεις ιδιότητες του τυχαίου γραφήματος	4
3	Ουρές Αναμονής και Κλαδωτές Διαδικασίες	6
3.1	Ουρές αναμονής	6
3.2	Πιθανότητα επιβίωσης ουράς αναμονής	7
3.3	Συνολικός πληθυσμός ουράς αναμονής	12
3.4	Σχέση Poisson και διωνυμικών ουρών αναμονής	14
3.5	Κλαδωτές διαδικασίες	17
4	Μεταβολή Φάσης στο Τυχαίο Γράφημα	20
4.1	Διαδικασία Γραφήματος	20
4.1.1	Αναζήτηση Κατά Πλάτος	20
4.1.2	Ορισμός Διαδικασίας Γραφήματος	21
4.2	Σύνδεση της διαδικασίας γραφήματος με τη διωνυμική ουρά αναμονής	22
4.3	Η πρώτη περίπτωση ($\lambda < 1$)	23
4.4	Η δεύτερη περίπτωση ($\lambda > 1$)	30
4.5	Κεντρικό Οριακό Θεώρημα για τη γιγαντιαία συνιστώσα	39
4.6	Κρίσιμο διάστημα	47
4.7	Το τυχαίο γράφημα ακολουθεί την κατανομή Poisson ασυμπτωτικά	53
4.8	Μονοτονία στο τυχαίο γράφημα	56
4.9	Παράδειγμα τυχαίου γραφήματος	56
5	Επιλεκτική Τοποθέτηση	61
5.1	Το μοντέλο των Barabási και Albert	61
5.2	Μια νέα θεώρηση της επιλεκτικής σύνδεσης	62
5.2.1	Το εναλλακτικό μοντέλο	62
5.2.2	Η κατανομή βαθμών στο G_m^n	64

Κεφάλαιο 1

Εισαγωγή

Με τον όρο δίκτυα θα αναφερόμαστε σε οποιαδήποτε διαδικασία μπορεί να αναπαρασταθεί ως ένα γράφημα με κάποια σημεία (ή κόμβους ή κορυφές) που ενώνονται μεταξύ τους μέσω συνδέσμων (ή ακμών). Για παράδειγμα, το οδικό δίκτυο είναι ένα δίκτυο στο οποίο κάθε συμβολή δρόμων ή τέλος αδιεξόδου θεωρείται μια κορυφή και κάθε οδικό κομμάτι που ενώνει δύο κορυφές είναι μια ακμή. Ένα άλλο παράδειγμα είναι το οικογενειακό δέντρο κάθε ανθρώπου στο οποίο τα άτομα είναι οι κορυφές και υπάρχει ακμή ανάμεσα σε δύο από αυτά μόνο εάν υπάρχει σχέση γονιού-παιδιού ανάμεσά τους.

Πολλές φορές μελετούμε δίκτυα τα οποία έχουν κάποια χαρακτηριστική ιδιότητα όπως για παράδειγμα μια συγκεκριμένη κατανομή βαθμών στις κορυφές τους ή διάμετρο (μέγιστη απόσταση κορυφών) που ακολουθεί κάποιο συγκεκριμένο κανόνα. Θα θέλαμε να μπορούμε να κάνουμε ένα βήμα πίσω και να δούμε ένα δίκτυο με μια τέτοια ιδιότητα ως προς το σύνολό του, παίρνοντας πληροφορίες και για τα υπόλοιπα χαρακτηριστικά του. Αυτή η γενικότερη γνώση που μπορεί να πάρουμε ενδεχομένως να επιβεβαιώσει ή να απορρίψει κάποιες υποθέσεις που έχουμε κάνει εκ των προτέρων για αυτό. Βρίσκουμε έτσι τον πρώτο λόγο για την ενασχόλησή μας με την ανάπτυξη μοντέλων για τη λειτουργία δικτύων.

Από εκεί και πέρα τα μοντέλα των δικτύων είναι χρήσιμα και για άλλους σκοπούς. Εάν έχουμε φτιάξει ένα μοντέλο μέσω υπολογιστή το οποίο είμαστε σίγουροι πως προσομοιώνει ικανοποιητικά το πραγματικό δίκτυο το οποίο θέλουμε να μελετήσουμε, δε χρειάζεται πια να συλλέγουμε πληροφορίες για το πραγματικό τακτικά, καθώς μπορούμε να προβλέψουμε τη δομή του μέσω του μοντέλου που έχουμε δημιουργήσει. Ένα δίκτυο το οποίο αποτελείται από όλους τους πολίτες των ΗΠΑ ως κορυφές και την ύπαρξη φιλίας ανάμεσα σε δύο από αυτούς να μεταφράζεται ως ακμή θα χρειαζόταν πολύ καιρό για να αναπαρασταθεί επακριβώς σε ένα γράφημα. Επιπροσθέτως, μέχρι να τελειώσει η συλλογή όλων των δεδομένων πιθανώς η σχέση κάποιων να έχει αλλάξει και έτσι το δίκτυο που θα δημιουργούσαμε να μην ήταν ακόμα κι έτσι ακριβές. Με τη δημιουργία ενός “καλού” μοντέλου θα μπορούσαμε να γλιτώσουμε τον κόπο, το χρόνο και το χρήμα που είναι αναγκαία για μια εξαντλητική περιγραφή ενός δικτύου, χωρίς ωστόσο να χάσουμε τα βασικά αποτελέσματά του. Γιατί εάν για κάποιο λόγο μας ενδιέφεραν οι φιλίες στις ΗΠΑ, πιθανότατα δε θα μας απασχολούσε εάν ο Dean είναι φίλος με τον Sal, αλλά ιδιότητες όπως το ποια κατανομή ακολουθούν οι βαθμοί των κορυφών του δικτύου, εάν το δίκτυο είναι συνεκτικό, εάν παρατηρείται κάποια διαφορά στη συμπεριφορά του στα μεγάλα αστικά κέντρα σε σχέση με την επαρχία ή κάποια άλλη ιδιότητα. Ο δεύτερος σκοπός της δημιουργίας μοντέλων είναι δηλαδή η δυνατότητα απεξάρτησης από τη συχνή συλλογή δεδομένων.

Τέλος, με την ανάπτυξη ακριβών μοντέλων για κάποια δίκτυα έχουμε τη δυνατότητα να προβλέπουμε τη συμπεριφορά τους, καθώς και να κάνουμε καιρικές επεμβάσεις κάποιου είδους σε αυτά. Ενδεχομένως να μπορούμε να μοντελοποιήσουμε ένα δυναμικό (αλλάζει η δομή του σε κάθε χρονική στιγμή) δίκτυο όπως αυτό με κορυφές όλους τους Έλληνες πολίτες που είναι υγιείς, προσβεβλημένοι από έναν ιό ή έχουν αναρρώσει από αυτόν και ακμές τις καθημερινές επαφές του καθενός με κάποιον άλλον και έτσι να δούμε τι γενικά μέτρα χρειάζεται να παρθούν ώστε να αποφευχθεί ή να ξεπεραστεί μια εξάπλωση του ιού. Επιπλέον, έχοντας ένα εμβόλιο σε περιορισμένες ποσότητες θα μπορούσαμε να βρούμε ποια είναι η ιδανική στρατηγική εμβολιασμού για την αντιμετώπιση του ιού και να την εφαρμόσουμε.

Οδηγούμαστε έτσι στην ανάγκη εύρεσης μοντέλων τα οποία να μπορούν να εφαρμοστούν με ικανοποιητικά αποτελέσματα σε αρκετές περιπτώσεις στην πραγματική ζωή. Το πιο ευρέως γνωστό και ταυτόχρονα απλό μοντέλο που έχει αναπτυχθεί είναι αυτό του **τυχαίου γραφήματος (random graph)**. Εδώ το πλήθος των κορυφών είναι ίσο με n και καθεμία από τις δυνατές ακμές έχει πιθανότητα να εμφανιστεί στο γράφημα ίση με p ανεξάρτητα από την εμφάνιση ή μη των υπολοίπων ακμών. Αυτό το μοντέλο δεν ακούγεται

πολύ ρεαλιστικό και όντως δεν είναι. Γιατί όμως τότε έχει παρουσιαστεί τόσο μεγάλο ενδιαφέρον για τη μελέτη του; Η απάντηση βρίσκεται στο γεγονός πως η ενασχόληση με ένα τόσο απλό μοντέλο δείχνει να είναι το πρότυπο και για τη μελέτη άλλων πιο σύνθετων μοντέλων, αλλά καθώς και ότι παρόλη την απλότητά του, παρουσιάζει μια εξαιρετικά ενδιαφέρουσα και απρόσμενη συμπεριφορά η οποία αποτελεί και το κύριο θέμα αυτής της διπλωματικής εργασίας.

Αυτή η εργασία χωρίζεται σε τρία κύρια μέρη. Στο πρώτο (κεφάλαιο 2) γίνεται μια συζήτηση για το τυχαίο γράφημα καθώς και μια μικρή ενασχόληση με τις ιδιότητές του οι οποίες είναι εύκολα υπολογίσιμες. Στη συνέχεια (κεφάλαιο 3) κάνουμε λόγο για κάποια γενικά αποτελέσματα πάνω στις ουρές αναμονής τα οποία είναι εξαιρετικά χρήσιμα για το υπόλοιπο μέρος της εργασίας.

Στο τρίτο και τελευταίο μέρος (κεφάλαιο 4) γίνεται ουσιαστικά η ανάλυση του τυχαίου γραφήματος. Βασιζόμενοι σε μια ελαφριά παραλλαγή της αναζήτησης κατά πλάτος (BFS) είμαστε σε θέση να ορίσουμε τη διαδικασία γραφήματος (ενότητα 4.1). Από εκεί και πέρα ασχολούμαστε σχεδόν αποκλειστικά με τη συμπεριφορά του μεγέθους της μεγαλύτερης συνεκτικής συνιστώσας, $|C_{\max}|$. Έχοντας εντοπίσει πως η τιμή της πιθανότητας p γύρω από την οποία παρατηρείται μια διαφορά στο $|C_{\max}|$ είναι η $\frac{1}{n}$ μελετούμε την περίπτωση στην οποία η p είναι μικρότερη αυτής (4.3), μεγαλύτερή της (4.4) ή πολύ κοντά της και τείνοντας σε αυτήν (4.6).

Τα αποτελέσματα είναι εντυπωσιακά. Στην πρώτη περίπτωση το $|C_{\max}|$ είναι τάξης μεγέθους $\log n$. Στη δεύτερη περίπτωση το $|C_{\max}|$ είναι τάξης μεγέθους n , γεγονός που δικαιολογεί την ιδιαίτερη ονομασία που του δίνουμε: η γιγαντιαία συνιστώσα. Κάνοντας μια παραμετροποίηση για το λ βλέπουμε πως στην τρίτη περίπτωση το $|C_{\max}|$ έχει μέγεθος της τάξης $n^{2/3}$.

Κλείνουμε αυτό το κεφάλαιο με την παρατήρηση πως η κατανομή των βαθμών του τυχαίου γραφήματος είναι η Poisson καθώς το πλήθος των κορυφών η τείνει να γίνει πολύ μεγάλο (ενότητα 4.7). Αυτός είναι και ο κύριος λόγος που καθιστά αυτό το μοντέλο μη χρήσιμο για πραγματικές καταστάσεις, καθώς έχει παρατηρηθεί πως στις περισσότερες -αν όχι όλες- καταστάσεις του πραγματικού κόσμου, οι βαθμοί των κόμβων των δικτύων ακολουθούν μια προς τα δεξιά στρεβλή κατανομή, ενώ η Poisson δεν είναι τέτοια.

Τέλος, στο κεφάλαιο 5 δίνεται ένας τρόπος με τον οποίο έχει αντιμετωπιστεί το προαναφερθέν πρόβλημα με την κατανομή βαθμών. Η ιδέα της **επιλεκτικής σύνδεσης (preferential attachment)** εμπεριέχει ένα γράφημα το οποίο μεγαλώνει καθώς περνάει ο χρόνος, και κάθε φορά που προστίθεται μια νέα κορυφή σε αυτό, ενώνεται με m από τις προϋπάρχουσες κορυφές με πιθανότητα ανάλογη στο βαθμό που έχουν αυτές μέχρι εκείνη τη στιγμή. Στη ενότητα 5.1 δίνεται ένα πρωτοποριακό μοντέλο που δόθηκε το 1999, το οποίο καταφέρνει να δημιουργήσει μια κατανομή βαθμών η οποία (τουλάχιστον πειραματικά) είναι δεξιά στρεβλή. Παρόλα αυτά, το συγκεκριμένο μοντέλο παρουσιάζει κάποιες αστοχίες οι οποίες επιλύονται μέσω ενός νέου μοντέλου που εισάγεται στην ενότητα 5.2. Εκεί ξεπερνώνται όλα τα προβλήματα και αποδεικνύεται πως ένα μοντέλο μπορεί να μιμηθεί την κατανομή βαθμών που παρατηρείται συχνά στα δίκτυα της πραγματικής ζωής.

Σε αυτό το σημείο αναφέρουμε πως το υλικό που χρησιμοποιήθηκε για τη συγγραφή της διπλωματικής εργασίας προέρχεται ως επί το πλείστον από τον πρώτο τόμο του βιβλίου του Remco van der Hofstad [13] και το βιβλίο του Mark Newman [12]. Αρκετές εξηγήσεις στις αποδείξεις οφείλονται στο βιβλίο *The Probabilistic Method* των Alon και Spencer [1]. Επίσης, όλες οι γνώσεις περί στοχαστικών διαδικασιών μπορούν να ανακτηθούν από το βιβλίο του Μιχάλη Λουλάκη [11]. Περαιτέρω μελέτη των ιδιοτήτων του τυχαίου γραφήματος κατά τη μεταβολή της φάσης του μπορεί να διενεργηθεί μέσω των βιβλίων των Bollobás [5] και των Janson, Luczak & Rucinski [10].

Κεφάλαιο 2

Το Τυχαίο Γράφημα

2.1 Το κλασικό τυχαίο γράφημα των Erdős και Rényi

Το πιο απλό παράδειγμα ενός μοντέλου κάποιου δικτύου είναι το τυχαίο γράφημα, σύμφωνα με το οποίο έχουμε n κορυφές σε ένα γράφημα και επιλέγουμε m από τις $\binom{n}{2}$ δυνατές ακμές στην τύχη. Αυτό το μοντέλο συμβολίζεται με $G(n, m)$ και μελετήθηκε για πρώτη φορά συστηματικά από τους Erdős και Rényi στη μνημειώδη εργασία τους *On the evolution of random graphs* [7]. Συγκεκριμένα, εφόσον η επιλογή των m ακμών θα γίνει από τις συνολικά $\binom{n}{2}$ δυνατές ακμές, υπάρχουν $\binom{\binom{n}{2}}{m}$ δυνατές επιλογές των m ακμών. Αυτό σημαίνει πως όταν μιλάμε για το τυχαίο γράφημα $G(n, m)$ ουσιαστικά μιλάμε για ένα γράφημα το οποίο το έχουμε επιλέξει τυχαία από όλα τα γραφήματα σε n κορυφές και με m ακμές. Δηλαδή πιθανότητα να επιλεγεί ένα συγκεκριμένο γράφημα G με n κορυφές και m ακμές είναι

$$\mathbb{P}(G) = \frac{1}{\binom{\binom{n}{2}}{m}} \quad (2.1)$$

Αρα θα πρέπει να κρατάμε πάντα στο μυαλό μας πως παρόλο που συζητάμε για το τυχαίο γράφημα, ουσιαστικά μελετούμε έναν χώρο πιθανότητας, στον οποίο ο δειγματικός χώρος είναι όλα τα δυνατά γραφήματα σε n κορυφές και m ακμές και η πιθανότητα που δίνεται στο καθένα ξεχωριστά είναι $\frac{1}{\binom{\binom{n}{2}}{m}}$. Κάποιες ιδιότητες του $G(n, m)$ είναι γρήγορα υπολογίσιμες. Για παράδειγμα, ο μέσος αριθμός των ακμών στο $G(n, m)$ είναι m και ο αναμενόμενος βαθμός μας κορυφής είναι $\frac{2m}{n}$. Κάποιες άλλες όμως δεν είναι το ίδιο εύκολο να υπολογιστούν, γεγονός που καθιστά την επόμενη ενότητα ιδιαίτερας χρήσιμη.

2.2 Η συχνότερη μορφή του

Σε αντίθεση με το μοντέλο που περιγράψαμε προηγουμένως, συνήθως χρησιμοποιείται μια ελαφρώς διαφοροποιημένη εκδοχή του τυχαίου γραφήματος n οποία κάνει πιο εύκολες τις πράξεις για τις ασυμπτωτικές συμπεριφορές του τυχαίου γραφήματος. Αυτό το μοντέλο εισήχθη από τον Edgar Gilbert [8], ωστόσο επειδή έχει την ίδια συμπεριφορά με το μοντέλο των Erdős και Rényi και επειδή το τελευταίο αναλύθηκε εις βάθος και λίγο νωρίτερα, αναφέρεται και αυτό ως **το τυχαίο γράφημα**.

Θεωρούμε πως έχουμε n κορυφές τις οποίες συμβολίζουμε με αριθμούς από το 1 έως και το n . Σε αυτό το τυχαίο γράφημα όμως, θεωρούμε πως όλες οι δυνατές ακμές μεταξύ δύο κορυφών μπορούν να συμπεριληφθούν στο γράφημα, η καθεμία με πιθανότητα p και ανεξάρτητα με το εάν οι υπόλοιπες ακμές θα παρουσιαστούν στο γράφημα.

Αυτό το μοντέλο το συμβολίζουμε με $G(n, p)$. Φυσικά, το $G(n, p)$ δεν είναι ένα γράφημα, αλλά ένας χώρος πιθανότητας που περιέχει κάθε δυνατό γράφημα με n κορυφές στο οποίο δίνεται και η αντίστοιχη πιθανότητα.

Ουσιαστικά, ένα συγκεκριμένο γράφημα στις n κορυφές που έχει m ακμές, έχει πιθανότητα $p^m \cdot (1-p)^{\binom{n}{2}-m}$ να εμφανιστεί, αφού καθεμία από τις m ακμές έχει πιθανότητα p να εμφανιστεί στο γράφημα και η ύπαρξή της είναι ανεξάρτητη από την ύπαρξη των άλλων. Σε κάθε γράφημα οι δυνατές ακμές είναι $\binom{n}{2}$

και άρα οι υπόλοιπες $\binom{n}{2} - m$ ακμές έχουν πιθανότητα $(1-p)$ να μην εμφανιστούν ανεξάρτητα η μία από τις άλλες.

Η βασική διαφορά αυτού του μοντέλου σε σχέση με το $G(n,m)$ είναι πως εδώ κάθε γράφημα έχει μια πιθανότητα εμφάνισης, ενώ στο $G(n,m)$ μόνο τα γραφήματα που περιέχουν m ακμές μπορούν να εμφανιστούν.

Από εδώ και στο εξής, όταν αναφερόμαστε στο "τυχαίο γράφημα" ή στο "γράφημα των Erdős και Rényi" θα εννοούμε το $G(n,p)$, εκτός και αν αναφέρεται ρητά πως εννοείται το $G(n,m)$.

2.3 Κάποιες στοιχειώδεις ιδιότητες του τυχαίου γραφήματος

Ας θεωρήσουμε κάθε δυνατή ακμή ως μια τυχαία μεταβλητή που μπορεί να πάρει τις τιμές 1 ή 0 ανάλογα με το εάν εμφανιστεί ή όχι αντίστοιχα. Αφού σε ένα γράφημα με n κορυφές υπάρχουν $\binom{n}{2}$ δυνατές ακμές θα έχουμε τις τυχαίες μεταβλητές $X_1, X_2, \dots, X_{\binom{n}{2}}$. Η πιθανότητα να εμφανιστεί μια ακμή στο $G(n,p)$ είναι p και άρα οι X_i ακολουθούν την κατανομή Bernoulli με παράμετρο p . Αυτό σημαίνει πως η αναμενόμενη τιμή των τυχαίων μεταβλητών είναι $\mathbb{E}(X_i) = 1 \cdot p + 0 \cdot (1-p) = p$.

Ο συνολικός αριθμός των ακμών που θα εμφανιστούν στο τυχαίο γράφημα θα είναι $X = X_1 + X_2 + \dots + X_{\binom{n}{2}}$, αφού αν μια ακμή παρουσιαστεί στο γράφημα θα συνεισφέρει $X_i = 1$ στο σύνολο των ακμών, ενώ αν δεν εμφανιστεί δε συνεισφέρει τίποτα, $X_i = 0$.

Ο αναμενόμενος αριθμός ακμών στο $G(n,p)$ τότε θα είναι:

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(X_1 + X_2 + \dots + X_{\binom{n}{2}}) \\ &= \mathbb{E}(X_1) + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_{\binom{n}{2}}) \\ &= p + p + \dots + p \\ &= \binom{n}{2} \cdot p \end{aligned} \tag{2.2}$$

όπου χρησιμοποιήθηκε η γραμμικότητα της αναμενόμενης τιμής στη δεύτερη ισότητα.

Μια άλλη ιδιότητα που μελετάται στα μοντέλα γραφημάτων είναι ο **αναμενόμενος βαθμός** μιας κορυφής ενός γραφήματος.

Στο τυχαίο γράφημα κάθε κόμβος έχει $n-1$ πιθανούς γείτονες και άρα ο βαθμός του κόμβου ισούται με το άθροισμα των τυχαίων μεταβλητών για τις δυνατές ακμές μεταξύ αυτού και των υπόλοιπων κορυφών ($d_i = \sum_{j=1}^{n-1} X_j$). Κάθε δυνατή σύνδεση με αυτούς έχει πιθανότητα p να πραγματοποιηθεί ανεξάρτητα από την εμφάνιση των άλλων ακμών. Άρα ο αναμενόμενος βαθμός μιας κορυφής στο τυχαίο γράφημα είναι:

$$\begin{aligned} \mathbb{E}(d_i) &= \mathbb{E}\left(\sum_{j=1}^{n-1} X_j\right) \\ &= \sum_{j=1}^{n-1} (\mathbb{E}(X_j)) \\ &= \sum_{j=1}^{n-1} (p) \\ &= (n-1) \cdot p \end{aligned}$$

Στη συνέχεια ασχολούμαστε με την **κατανομή του βαθμού μιας κορυφής**. Δηλαδή εάν επιλέξουμε έναν κόμβο τυχαία να βρούμε ποια είναι η πιθανότητα να έχει k γείτονες, όπου $k=0,1,\dots,n-1$. Στο τυχαίο γράφημα όλες οι ακμές έχουν την ίδια πιθανότητα να εμφανιστούν και άρα όλοι οι κόμβοι του γραφήματος θα ακολουθούν την ίδια κατανομή γειτόνων.

Ένας κόμβος του γραφήματος έχει $n-1$ πιθανούς γείτονες, τον καθένα με πιθανότητα p . Άρα η πιθανότητα να έχει ακριβώς k συγκεκριμένους από αυτούς είναι $p^k \cdot (1-p)^{(n-1)-k}$. Οι τρόποι με τους οποίους μπορούν να επιλεγούν k γείτονες από τους $n-1$ υποψήφιους όμως είναι $\binom{n-1}{k}$ και άρα η πιθανότητα ένας κόμβος να έχει ακριβώς k γείτονες είναι:

$$p_k = \binom{n-1}{k} \cdot p^k \cdot (1-p)^{n-1-k} \quad (2.3)$$

Έχει δηλαδή διωνυμική κατανομή $\text{Bin}(n-1, p)$, κάτι που φυσικά το περιμέναμε, αφού η κάθε επιλογή μιας ακμής ισοδυναμεί με μια δοκιμή Bernoulli με παράμετρο p και ξέρουμε ότι το άθροισμα μεταβλητών που ακολουθούν την κατανομή Bernoulli ακολουθεί τη διωνυμική κατανομή. Με τους συμβολισμούς που χρησιμοποιήθηκαν προηγουμένως δηλαδή, οι $X_j, j = 1, 2, \dots, n-1$ (οι υποψήφιες ακμές ενός συγκεκριμένου κόμβου) ακολουθούν την κατανομή Bernoulli με πιθανότητα επιτυχίας p και ο βαθμός $d_i = \sum_{j=1}^{n-1} X_j$ ακολουθεί τη διωνυμική κατανομή $\text{Bin}(n-1, p)$.

Κεφάλαιο 3

Ουρές Αναμονής και Κλαδωτές Διαδικασίες

3.1 Ουρές αναμονής

Μια ουρά αναμονής είναι ένας τυχαίος περίπατος στους μη-αρνητικούς ακέραιους αριθμούς και αποτελεί τον πιο απλό τρόπο περιγραφής της εξέλιξης ενός πληθυσμού στο πέρασμα του χρόνου. Την αρχική χρονική στιγμή ($t=0$), υπάρχει ένα άτομο στον πληθυσμό. Αυτό θα δώσει κάποιους απογόνους (μπορεί και κανέναν) και στη συνέχεια θα πεθάνει. Πλέον βρισκόμαστε στην επόμενη χρονική στιγμή ($t=1$). Ο συνολικός αριθμός των ανθρώπων είναι όσοι γεννήθηκαν στο $t=0$ αφαιρώντας έναν (αυτόν που πέθανε). Τώρα επιλέγεται ένα άτομο από τον πληθυσμό τυχαία, δίνει κάποιους απογόνους και πεθαίνει. Αυτή η διαδικασία συνεχίζεται μέχρι να μη μείνει κανείς ζωντανός. Άρα εάν το πλήθος των ζωντανών ανθρώπων τη χρονική στιγμή i συμβολιστεί με S_i , η ουρά αναμονής είναι ένας τυχαίος περίπατος του S_i στους μη-αρνητικούς ακέραιους αριθμούς μέχρι την πρώτη χρονική στιγμή που θα φτάσει στο 0.

Φυσικά, την ουρά αναμονής μπορούμε να τη σκεφτόμαστε και ως μια ουρά εξυπηρέτησης, όπου πριν π.χ. ανοίξει ένα μαγαζί, υπάρχει ένας πελάτης που περιμένει. Τον υποδεχόμαστε στο μαγαζί, τον εξυπηρετούμε και καθώς φεύγει, έρχονται άλλοι πελάτες στην ουρά έξω από το μαγαζί σύμφωνα με μια κατανομή. Αυτή η διαδικασία συνεχίζεται μέχρι την πρώτη φορά που δε θα υπάρχει κανείς που να περιμένει στην ουρά.

Ωστόσο, συχνά θα μιλάμε με τους παραπάνω όρους (απόγονοι, γονείς, ζώντες, αποθνήσκοντες κ.λπ.) έτσι ώστε να γίνει και πιο εύκολα η σύνδεση με τις κλαδωτές διαδικασίες στην ενότητα 3.5.

Την πρώτη χρονική στιγμή που ο πληθυσμός θα φτάσει το μηδέν, τη συμβολίζουμε με T . Ο αριθμός των απογόνων κάθε ατόμου ακολουθεί την ίδια διακριτή τυχαία μεταβλητή X με χώρο καταστάσεων το \mathbb{N}_0 ανεξάρτητα από την τεκνοποίηση των υπολοίπων και ο αριθμός των ζωντανών ατόμων τη χρονική στιγμή i συμβολίζεται με S_i .

Χρησιμοποιώντας αυτούς τους συμβολισμούς, μια ουρά αναμονής περιγράφεται από τις σχέσεις:

$$\begin{aligned} S_0 &= 1, \\ S_i &= S_{i-1} + X_i - 1 \\ &= S_{i-2} + X_{i-1} - 1 + X_i - 1 \\ &= \dots \\ &= X_1 + X_2 + \dots + X_i - (i - 1) \end{aligned} \tag{3.1}$$

όπου οι τυχαίες μεταβλητές X_i ακολουθούν την ίδια κατανομή με την X και ο δείκτης i συμβολίζει τη χρονική στιγμή την οποία επιλέγεται ένα άτομο από τον πληθυσμό.

Επίσης για το T ισχύει πως:

$$T = \inf\{t : S_t = 0\} = \inf\{t : X_1 + X_2 + \dots + X_t = t - 1\} \tag{3.2}$$

Το T θα το λέμε και συνολικό πληθυσμό της ουράς αναμονής. Αν δεν υπάρχει t για το οποίο να ισχύει η παραπάνω σχέση, τότε ορίζουμε πως $T = \infty$.

3.2 Πιθανότητα επιβίωσης ουράς αναμονής

Σε μια ουρά αναμονής, η πιθανότητα επιβίωσής της είναι η πιθανότητα να μη σταματήσει ποτέ να έχει ζωντανό πληθυσμό, δηλαδή η πιθανότητα να ισχύει πως $T = \infty$. Η πιθανότητα αυτή συμβολίζεται με η .

Το συμπληρωματικό ενδεχόμενο του $\{T = \infty\}$ είναι κάποια στιγμή η ουρά αναμονής να σταματήσει, δηλαδή ο συνολικός πληθυσμός της να είναι πεπερασμένος. Την πιθανότητα αυτού του ενδεχομένου τη συμβολίζουμε με η και την αποκαλούμε πιθανότητα εξαφάνισης της ουράς αναμονής. Ισχύει για την η πως:

$$\eta = \mathbb{P}(\exists t : S_t = 0)$$

Το επόμενο θεώρημα είναι ουσιώδες για τη συμπεριφορά των ουρών αναμονής.

Θεώρημα 3.1. Σε μια ουρά αναμονής που ο αριθμός των απογόνων είναι η τυχαία μεταβλητή X ,

- i) αν το $\mathbb{E}(X) < 1$, τότε $\eta = \mathbb{P}(\{T < \infty\}) = 1$,
- ii) αν το $\mathbb{E}(X) = 1$ με $\mathbb{P}(X = 1) < 1$, τότε $\eta = \mathbb{P}(\{T < \infty\}) = 1$ και
- iii) αν το $\mathbb{E}(X) > 1$, τότε $\eta = \mathbb{P}(\{T < \infty\}) < 1$.

Όπως έχουμε πει, η ουρά αναμονής είναι ένας τυχαίος περίπατος στο \mathbb{N}_0 ο οποίος αρχίζει από το 1. Σε κάθε βήμα, έρχονται X απόγονοι και πεθαίνει ένας από τον πληθυσμό. Άρα, αν το $X = 0$, πάμε ένα βήμα πίσω στον περίπατο, αν το $X = 1$ μένουμε εκεί που είμαστε, και αν $X > 1$ κάνουμε $X - 1$ βήματα προς τα δεξιά. Επομένως είναι λογικό που η μέση τιμή που μας ενδιαφέρει για την κατανομή X είναι το 1.

Απόδειξη. Έστω η πιθανότητα $\eta_t = \mathbb{P}(S_t = 0)$.

Ισχύει πως $\{S_t = 0\} \subset \{S_{t+1} = 0\}$, αφού εάν $\{S_t = 0\}$ τότε $\{S_{t+1} = 0\}$, αλλά το αντίστροφο δεν ισχύει κατ'ανάγκη. Οπότε, έχουμε πως

$$\{S_t = 0\} \subset \{S_{t+1} = 0\} \implies \mathbb{P}(\{S_t = 0\}) \leq \mathbb{P}(\{S_{t+1} = 0\}) \implies \eta_t \leq \eta_{t+1}, \forall t \in \mathbb{N}_0 \text{ και}$$

$\bigcup_{t=1}^{\infty} \{S_t = 0\} = \{\exists t : S_t = 0\} \implies \mathbb{P}(\bigcup_{t=1}^{\infty} \{S_t = 0\}) = \mathbb{P}(\{\exists t : S_t = 0\}) \implies \lim_{t \rightarrow \infty} \mathbb{P}(\{S_t = 0\}) = \mathbb{P}(\{\exists t : S_t = 0\}) \implies \lim_{t \rightarrow \infty} \eta_t = \eta$, όπου η πιθανότητα της ένωσης είναι ίση με το όριο της πιθανότητας καθώς η ακολουθία των ενδεχομένων $\{S_t = 0\}$ είναι αύξουσα.

Άρα έχουμε πως $\eta_t \uparrow \eta$.

Σε αυτό το σημείο υπενθυμίζουμε τον τύπο της γεννήτριας συνάρτησης της πιθανότητας μιας τυχαίας μεταβλητής X :

$$G_X(s) = \mathbb{E}(s^X)$$

Άμα η τυχαία μεταβλητή είναι διακριτή και έχει πεδίο ορισμού το \mathbb{N}_0 , τότε:

$$\begin{aligned} G_X(s) &= \mathbb{E}(s^X) \\ &= \sum_{x=0}^{\infty} \mathbb{P}(X = x) s^x \\ &= \sum_{x=0}^{\infty} p_x s^x \\ &= p_0 + p_1 s + p_2 s^2 + \dots \end{aligned}$$

Άρα $\mathbb{P}(X = 0) = G_X(0)$.

Δηλαδή έχουμε ότι

$$\eta_t = \mathbb{P}(S_t = 0) = G_t(0)$$

όπου $G_t(s) = \mathbb{E}(s^{S_t})$.

Δεσμεύοντας ως προς το πλήθος των παιδιών του πρώτου ατόμου του πληθυσμού, έχουμε πως

$$\begin{aligned} G_t(s) &= \mathbb{E}(s^{S_t}) \\ &= \sum_{i=0}^{\infty} \mathbb{P}(X_1 = i) \cdot \mathbb{E}(s^{S_t} | X_1 = i) \\ &= \sum_{i=0}^{\infty} p_i \mathbb{E}(s^{S_{t-1}})^i \\ &= \sum_{i=0}^{\infty} p_i [G_{t-1}(s)]^i \end{aligned}$$

όπου η τρίτη ισότητα ισχύει καθώς δεδομένου πως έχουν γεννηθεί i παιδιά στην πρώτη "γενιά", από τη χρονική στιγμή 1 μέχρι και τη χρονική στιγμή t , το καθένα από αυτά θα είναι σαν το πρώτο άτομο σε έναν ξεχωριστό πληθυσμό και σε συνδυασμό με το γεγονός πως αυτές οι ουρές αναμονής είναι ανεξάρτητες, έχουν τα ίδια χαρακτηριστικά και πως χρειάζονται χρόνο $t-1$ για να φτάσουν στο χρόνο t , το $\mathbb{E}(s^{S_t} | X_1 = i) = \mathbb{E}(s^{S_{t-1}})^i$.

Για $s=0$ παίρνουμε

$$\begin{aligned} G_t(0) &= \sum_{i=0}^{\infty} p_i G_{t-1}(0)^i \\ \eta_t &= \sum_{i=0}^{\infty} p_i \eta_{t-1}^i \\ \eta_t &= G_X(\eta_{t-1}) \end{aligned} \tag{3.3}$$

Όμως καθώς $t \rightarrow \infty$, $\eta_t \uparrow \eta$ και λόγω της συνέχειας της γεννήτριας συνάρτησης πιθανότητας $G_X(s)$, έχουμε πως

$$\eta = G_X(\eta) \tag{3.4}$$

Εναλλακτικά μπορούμε να δούμε αυτή τη σχέση ως εξής: αρχικά έχουμε ένα άτομο στο πληθυσμό. Η πιθανότητα ο πληθυσμός να εξαφανιστεί είναι η . Αν αυτό το άτομο κάνει i παιδιά, τότε το καθένα από αυτά θα είναι σαν η αρχή ενός ξεχωριστού πληθυσμού και άρα κάθε τέτοιος πληθυσμός θα έχει πιθανότητα η να εξαφανιστεί. Για να εξαφανιστεί ο συνολικός πληθυσμός θα πρέπει οι πληθυσμοί και των i παιδιών να εξαφανιστούν, πράγμα που συμβαίνει με πιθανότητα η^i . Επομένως ισχύει πως $\eta = \sum_{i=0}^{\infty} p_i \eta^i = G_X(\eta)$.

Όταν σε μια ουρά αναμονής έχουμε πως $\mathbb{E}(X) = 1$ και $\mathbb{P}(X = 1) = 1$, σε κάθε χρονική στιγμή γεννιέται ένας απόγονος και σε συνδυασμό με το θάνατο του γονέα του, το $S_t = 1$ πάντα, και άρα $T = \infty$.

Επίσης, όταν $\mathbb{P}(X \leq 1) = 1$ με $p = \mathbb{P}(X = 0) > 0$, τότε $\mathbb{P}(S_t = 0) = 1 - \mathbb{P}(S_t \neq 0) = 1 - \mathbb{P}(S_t = 1) = 1 - (1-p)^t \rightarrow 1$, όπου $\mathbb{P}(S_t \neq 0) = \mathbb{P}(S_t = 1)$, γιατί σε κάθε χρονική στιγμή γεννιέται το πολύ ένας απόγονος και σε συνδυασμό με το θάνατο του γονέα του το S_t δεν ξεπερνάει το 1.

Άρα, από εδώ και στο εξής θα θεωρούμε πως $\mathbb{P}(X \leq 1) < 1$.

Τώρα θα δείξουμε πως η $G_X(s)$ είναι αυστηρά αύξουσα και αυστηρά κυρτή όταν το $s > 0$ και το $\mathbb{P}(X \leq 1) < 1$.

$$\begin{aligned}
G'_X(s) &= \left(\sum_{x=0}^{\infty} p_x s^x \right)' \\
&= (p_0 + p_1 s + p_2 s^2 + \dots)' \\
&= p_1 + 2p_2 s + 3p_3 s^2 \dots
\end{aligned}$$

και επειδή $\mathbb{P}(X \leq 1) < 1$, ισχύει ότι $p_2 + p_3 + \dots > 0$ και οπότε τουλάχιστον ένα από τα p_2, p_3, \dots είναι θετικό και άρα το $G'_X(s) > 0$ για $s > 0$, δηλαδή η $G_X(s)$ είναι αυστηρά αύξουσα.

Επίσης,

$$\begin{aligned}
G''_X(s) &= [G'_X(s)]' \\
&= (p_1 + 2p_2 s + 3p_3 s^2 \dots)' \\
&= 2p_2 + 6p_3 s + \dots \\
&= \sum_{x=2}^{\infty} x(x-1)p_x s^{x-2}
\end{aligned}$$

Αφού $p_2 + p_3 + \dots > 0$, τουλάχιστον ένα από τα p_2, p_3, \dots είναι θετικό και άρα το $G''_X(s) > 0$ για $s > 0$, δηλαδή η $G_X(s)$ είναι αυστηρά κυρτή.

Έστω τώρα ότι υπάρχει ένα ψ στο διάστημα $[0,1]$ για το οποίο ισχύει ότι $\psi = G_X(\psi)$.

Θα δείξουμε πως $\eta \leq \psi$. Για κάθε t θα δείξουμε με χρήση της μαθηματικής επαγωγής ότι ισχύει πως $\eta_t \leq \psi$:

1) $\eta_0 = 0$ (αφού $S_0 = 1$) $\leq \psi$ ($\in [0,1]$)

2) Έστω ότι $\eta_t \leq \psi$. Τότε

$$\eta_{t+1} = G_X(\eta_t) \leq G_X(\psi) = \psi$$

λόγω της 3.3 και επειδή η συνάρτηση $G_X(s)$ είναι αύξουσα στο $[0,1]$.

Άρα για κάθε t ισχύει πως $\eta_t \leq \psi$.

Επίσης, $\eta_t \uparrow \eta$ και άρα το $\eta \leq \psi$ για όλα τα ψ για τα οποία $\psi = G_X(\psi)$ στο $[0,1]$.

Άρα, αφού για την η δείξαμε πως ισχύει ότι $\eta = G_X(\eta)$, θα έχουμε ότι η η είναι η μικρότερη λύση της εξίσωσης $s = G_X(s)$, για s στο $[0,1]$.

Το $s = 1$ ικανοποιεί πάντα τη σχέση $s = G_X(s)$, αφού $G_X(1) = p_0 + p_1 \cdot 1 + p_2 \cdot 1^2 + \dots = p_0 + p_1 + p_2 + \dots = 1$. Επίσης, το $G_X(0) = p_0 + p_1 \cdot 0 + p_2 \cdot 0^2 + \dots = p_0 > 0$, γιατί διαφορετικά το $T = \infty$ τετριμμένα.

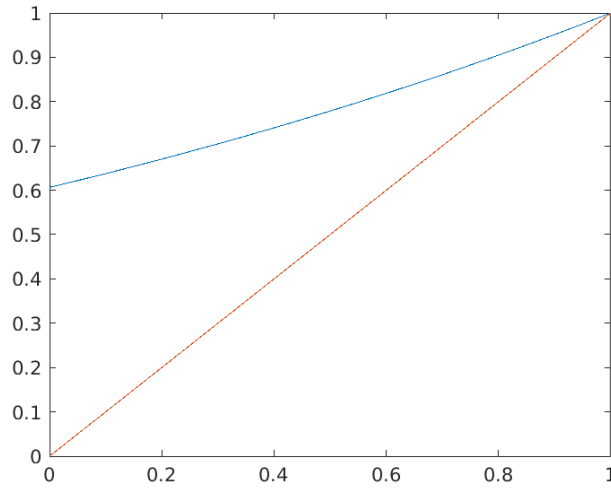
Επειδή η $G_X(s)$ είναι αυστηρά κυρτή και αυστηρά αύξουσα στο $[0,1]$ αυτό σημαίνει πως θα τέμνει την ευθεία $y = s$ είτε μόνο για $s = 1$ ή και για ένα ακόμα σημείο στο $(0,1)$.

Για να έχουν και ένα δεύτερο σημείο τομής θα πρέπει η συνάρτηση $G_X(s)$ να "έρχεται από τα κάτω" στο $s = 1$. Δηλαδή, θα πρέπει $G'_X(1) > 1$. Διαφορετικά, για $G'_X(1) \leq 1$ η εξίσωση $s = G_X(s)$ έχει μόνο μια λύση.

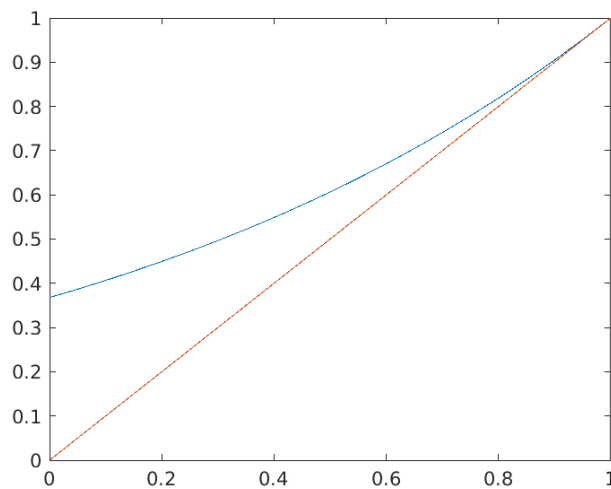
Επίσης, ισχύει πως $\mathbb{E}(X) = \sum_{x=0}^{\infty} x \mathbb{P}(X = x) = G'_X(1)$ και άρα τα ζητούμενα του θεωρήματος απεδεδίχθησαν.

□

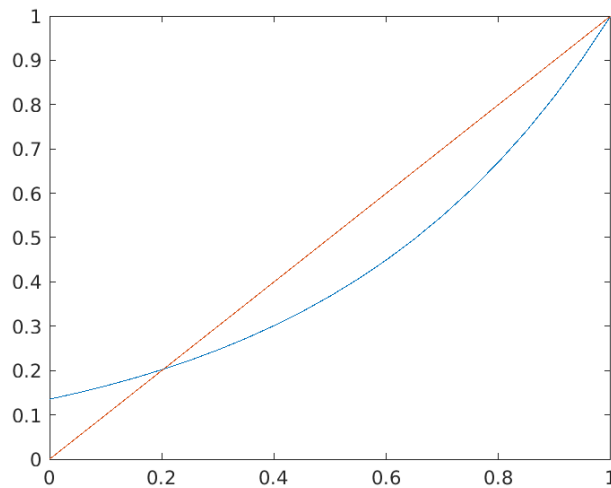
Στα επόμενα σχήματα φαίνονται και τα αποτελέσματα της παραπάνω απόδειξης για την περίπτωση που η τυχαία μεταβλητή $X \sim \text{Po}(\lambda)$ για το $\lambda = \mathbb{E}(X)$ να είναι μικρότερο, ίσο και μεγαλύτερο του 1. Με πορτοκαλί χρώμα βλέπουμε την ταυτοτική συνάρτηση και με μπλε τη $G_X(s) = e^{\lambda(s-1)}$ για την Poisson, έτσι ώστε να είναι εμφανή τα σημεία στο διάστημα $[0,1]$ για τα οποία ισχύει πως $s = G_X(s)$.



Σχήμα 3.1: Η περίπτωση όπου το $\lambda < 1$



Σχήμα 3.2: Η περίπτωση όπου το $\lambda = 1$



Σχήμα 3.3: Η περίπτωση όπου το $\lambda > 1$.

Πρόταση 3.2. Έστω η_λ η πιθανότητα αφανισμού μιας ουράς αναμονής με αριθμό απογόνων X που ακολουθεί κατανομή Poisson με μέσο όρο λ και $\zeta_\lambda = 1 - \eta_\lambda$ η αντίστοιχη πιθανότητα επιβίωσης. Τότε, για όλα τα $\lambda > 1$ ισχύει

$$\frac{d}{d\lambda} \eta_\lambda = -\frac{\eta_\lambda(1-\eta_\lambda)}{1-\lambda\eta_\lambda} \quad (3.5)$$

Απόδειξη. Παρατηρούμε πως η η_λ ικανοποιεί τη σχέση

$$\eta_\lambda = e^{\lambda(1-\eta_\lambda)} \quad (3.6)$$

γιατί για τη διαδικασία Poisson με μέσο όρο λ η γεννήτρια συνάρτηση είναι η

$$G_{\text{po}(\lambda)}(s) = \mathbb{E}(s^X) = \sum_{x=0}^{\infty} s^x e^{-\lambda} \frac{\lambda^x}{x!} = e^{\lambda(s-1)}$$

και από τη σχέση 3.4 παίρνουμε πως $\eta = G_X(\eta)$.

Έστω τώρα η συνάρτηση $f : (0,1) \rightarrow (1, \infty)$ με τύπο

$$f(x) = \frac{\log x}{x-1}$$

Αυτή η συνάρτηση είναι γνησίως φθίνουσα στο $(0,1)$ με $\lim_{x \rightarrow 0^+} f(x) = \infty$ και $\lim_{x \rightarrow 1} f(x) = 1$. Άρα η f είναι 1-1 και επί και είναι συνεχώς παραγωγίσιμη στο $(0,1)$. Δηλαδή και η αντίστροφή της θα είναι συνεχώς παραγωγίσιμη και $f^{-1}(\lambda) = \eta_\lambda$ για το $\lambda > 1$. Οπότε η συνάρτηση που απεικονίζει το λ στο η_λ είναι συνεχώς παραγωγίσιμη.

Λογαριθμίζοντας στη σχέση 3.6 παίρνουμε ότι $\log \eta_\lambda = \lambda(\eta_\lambda - 1)$. Μετά από τη συζήτηση της προηγούμενης παραγράφου μπορούμε να παραγωγίσουμε αυτή τη σχέση και τελικά να πάρουμε ότι

$$\begin{aligned} (\log \eta_\lambda)' &= [\lambda(\eta_\lambda - 1)]' \\ \frac{1}{\eta_\lambda} \frac{d}{d\lambda} \eta_\lambda &= (\eta_\lambda - 1) + \lambda \frac{d}{d\lambda} \eta_\lambda \\ \frac{d}{d\lambda} \eta_\lambda &= -\frac{\eta_\lambda(1 - \eta_\lambda)}{1 - \lambda \eta_\lambda} \end{aligned}$$

□

Πόρισμα 3.3. Επιπλέον, όταν το λ τείνει από τα θετικά στο 1, ισχύει ότι

$$\zeta_\lambda = 2(\lambda - 1)(1 + o(1)) \quad (3.7)$$

Απόδειξη. Θυμίζουμε την εξίσωση που ικανοποιεί η πιθανότητα επιβίωσης:

$$\zeta_\lambda = 1 - e^{-\lambda \zeta_\lambda}$$

Ξέρουμε πως όταν το $\lambda \rightarrow 1+$, το $\zeta_\lambda \rightarrow 0+$. Άρα εάν αναπτύξουμε σε σειρά Taylor τη συνάρτηση $f(x) = e^{-\lambda x}$ στο $x = 0$ παίρνουμε πως

$$\begin{aligned} \zeta_\lambda &= 1 - e^{-\lambda \zeta_\lambda} \iff \\ \zeta_\lambda &= 1 - 1 + \lambda \zeta_\lambda - \frac{\lambda^2 \zeta_\lambda^2}{2} + o(\zeta_\lambda^2) \iff \\ \lambda^2 \frac{\zeta_\lambda}{2} &= \lambda - 1 + o(\zeta_\lambda) \\ \frac{\zeta_\lambda}{2} &= \lambda - 1 + o(\zeta_\lambda) \end{aligned} \quad \text{αφού το } \lambda^2 = 1 + \varepsilon, \varepsilon \rightarrow 0$$

□

3.3 Συνολικός πληθυσμός ουράς αναμονής

Σε έναν τυχαίο περίπατο στους ακεραίους ή σε κάποιο υποσύνολό τους θα συμβολίζουμε με $\mathbb{P}_k(A)$ την πιθανότητα να συμβεί το ενδεχόμενο A στον τυχαίο περίπατο που έχει αρχική θέση τον ακέραιο k . Επίσης, με $Y_i, i \geq 1$ θα συμβολίζουμε τις ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που καθορίζουν την απόσταση που καλύπτεται στο i -οστό βήμα του τυχαίου περιπάτου, με $S_n = k + Y_1 + \dots + Y_n$ τη θέση του περιπάτου μετά από n βήματα και με $H_0 = \inf\{n : S_n\}$ την πρώτη χρονική στιγμή που ο περίπατος θα φτάσει στο 0, δηλαδή το **χρόνο πρώτης άφιξης** στο 0.

Θεώρημα 3.4 (Θεώρημα Χρόνου Πρώτης Άφιξης). Σε έναν τυχαίο περίπατο με $Y_i \geq -1, i \geq 1$ τις ανεξάρτητες και ισόνομες διακριτές τυχαίες μεταβλητές των βημάτων του, n κατανομή του χρόνου πρώτης άφιξης στο 0 δίνεται από τη σχέση

$$\mathbb{P}_k(H_0 = n) = \frac{k}{n} \mathbb{P}_k(S_n = 0) \quad (3.8)$$

όπου $k \in \mathbb{N}_0$.

Απόδειξη. Θα δείξουμε το ζητούμενο με μαθηματική επαγωγή για κάθε $k \geq 0$.

Επαγωγική Βάση: Έστω πως $n=1$. Τότε για $k=0$ το αριστερό μέλος της 3.8 γίνεται $\mathbb{P}_0(H_0 = 1) = 0$, αφού αν ο τυχαίος περίπατος ξεκινάει από το 0, τότε ο χρόνος πρώτης άφιξης στο 0 είναι το 0 και το δεξί $\frac{0}{1} \mathbb{P}_0(S_1 = 0) = 0$ και άρα ισχύει η ισότητα.

Για $k=1$ το αριστερό μέλος της 3.8 γίνεται $\mathbb{P}_1(H_0 = 1) = \mathbb{P}(Y_1 = -1)$ και το δεξί $\frac{1}{1} \mathbb{P}_1(S_1 = 0) = \mathbb{P}_1(1 + Y_1 = 0) = \mathbb{P}_1(Y_1 = -1)$ και άρα ισχύει η ισότητα.

Για $k > 1$ το αριστερό μέλος της 3.8 γίνεται $\mathbb{P}_k(H_0 = 1) = \mathbb{P}(Y_1 = -k) = 0$ και το δεξί $\frac{k}{1} \mathbb{P}_k(S_1 = 0) = \mathbb{P}_k(k + Y_1 = 0) = \mathbb{P}(Y_1 = -k) = 0$ και άρα ισχύει η ισότητα.

Επαγωγικό Βήμα: Έστω ότι η σχέση ισχύει έως και το $n-1$. Θα δείξουμε πως ισχύει και για το n . Αρχικά, για $k=0$ όπως και παραπάνω τα δύο μέλη της 3.8 είναι ίσα, οπότε για τη συνέχεια της απόδειξης θα θεωρούμε πως $k \geq 1$.

Με δέσμευση στο πρώτο βήμα του περιπάτου έχουμε πως

$$\mathbb{P}_k(H_0 = n) = \sum_{s=-1}^{\infty} \mathbb{P}_k(H_0 = n | Y_1 = s) \mathbb{P}(Y_1 = s)$$

Με χρήση της Ισχυρής Μαρκοβιανής Ιδιότητας για τον περίπατο έχουμε ότι

$$\mathbb{P}_k(H_0 = n | Y_1 = s) = \mathbb{P}_{k+s}(H_0 = n - 1) = \frac{k+s}{n-1} \mathbb{P}_{k+s}(S_{n-1} = 0)$$

με την τελευταία ισότητα να ισχύει εξαιτίας της επαγωγικής υπόθεσης και επειδή $k \geq 1$ και $s \geq -1$ καθώς το $Y \geq -1$ και άρα $k+s \geq 0$.

Οπότε

$$\mathbb{P}_k(H_0 = n) = \sum_{s=-1}^{\infty} \frac{k+s}{n-1} \mathbb{P}_{k+s}(S_{n-1} = 0) \mathbb{P}(Y_1 = s)$$

Χρησιμοποιώντας και πάλι την Ισχυρή Μαρκοβιανή Ιδιότητα έχουμε ότι:

$$\begin{aligned}
\mathbb{P}_k(H_0 = n) &= \sum_{s=-1}^{\infty} \frac{k+s}{n-1} \mathbb{P}_{k+s}(S_{n-1} = 0) \mathbb{P}(Y_1 = s) \\
&= \sum_{s=-1}^{\infty} \frac{k+s}{n-1} \mathbb{P}_k(S_n = 0 | Y_1 = s) \mathbb{P}(Y_1 = s) \\
&= \frac{1}{n-1} \left[\sum_{s=-1}^{\infty} k \mathbb{P}_k(S_n = 0 | Y_1 = s) \mathbb{P}(Y_1 = s) + \sum_{s=-1}^{\infty} s \mathbb{P}_k(S_n = 0 | Y_1 = s) \mathbb{P}(Y_1 = s) \right] \\
&= \frac{1}{n-1} \left[k \mathbb{P}_k(S_n = 0) + \sum_{s=-1}^{\infty} s \mathbb{P}_k(Y_1 = s | S_n = 0) \mathbb{P}_k(S_n = 0) \right] \\
&= \frac{1}{n-1} \mathbb{P}_k(S_n = 0) [k + \mathbb{E}_k(Y_1 | S_n = 0)] \tag{3.9}
\end{aligned}$$

όπου η προτελευταία ισότητα είναι αποτέλεσμα της ιδιότητας των δεσμευμένων πιθανοτήτων.

Για να υπολογίσουμε το $\mathbb{E}_k(Y_1 | S_n = 0)$ προσέχουμε πως το $\mathbb{E}_k(Y_i | S_n = 0)$ είναι ανεξάρτητο από το δείκτη i αφού δεδομένου πως $S_n = 0 \iff k + Y_1 + \dots + Y_n = 0 \iff Y_1 + \dots + Y_n = -k$, δεν έχει σημασία ποιο Y_i παίρνει μια συγκεκριμένη τιμή αρκεί να ισχύει η σχέση για το άθροισμά τους. Οπότε ισχύει πως

$$\mathbb{E}_k(Y_1 | S_n = 0) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_k(Y_i | S_n = 0) = \frac{1}{n} \mathbb{E}_k \left(\sum_{i=1}^n (Y_i | S_n = 0) \right) = -\frac{k}{n}$$

καθώς η δεύτερη ισότητα ισχύει εξαιτίας της γραμμικότητας της αναμενόμενης τιμής και η τρίτη αφού $\sum_{i=1}^n Y_i = S_n - k = -k$ δεδομένου πως $S_n = 0$.

Τελικά η σχέση 3.9 θα γίνει

$$\mathbb{P}_k(H_0 = n) = \frac{1}{n-1} \mathbb{P}_k(S_n = 0) \left(k - \frac{k}{n} \right) = \frac{k}{n} \mathbb{P}_k(S_n = 0)$$

□

Το παραπάνω θεώρημα ισχύει για κάθε τυχαίο περίπατο στους ακεραίους με αρχική θέση $k \geq 0$. Εφόσον η ουρά αναμονής είναι ένας τυχαίος περίπατος με αρχική θέση το 1, που έχει αριθμό απογόνων κάθε χρονική στιγμή που ακολουθεί κατανομή όπως και αυτή της τυχαίας μεταβλητής X με $\mathbb{P}(X \geq 0) = 1$ και άρα ο τυχαίος περίπατος κάνει βήμα $Y = X - 1$ με $\mathbb{P}(Y \geq -1) = 1$ και ο αριθμός των ζωντανών ανθρώπων είναι ο $S_t = X_1 + \dots + X_t - (t - 1) = Y_1 + 1 + \dots + Y_t + 1 - t + 1 = 1 + Y_1 + \dots + Y_t$, το θεώρημα 3.4 γράφεται ειδικά ως

$$\mathbb{P}(T = n) = \frac{1}{n} \mathbb{P}(S_n = 0) = \frac{1}{n} \mathbb{P}(X_1 + \dots + X_n = n - 1) \tag{3.10}$$

Το παρακάτω θεώρημα δίνει την κατανομή της πιθανότητας του συνολικού πληθυσμού μιας ουράς αναμονής Poisson, δηλαδή μιας ουράς αναμονής όπου ο αριθμός των απογόνων σε κάθε χρονική στιγμή ακολουθεί μια κατανομή Poisson με μέσο όρο λ .

Θεώρημα 3.5. Για μια ουρά αναμονής Poisson με μέσο όρο λ ισχύει ότι

$$\mathbb{P}_{po}(T_{po} = n) = \frac{(\lambda n)^{n-1}}{n!} e^{-\lambda n}$$

Απόδειξη. Από τη σχέση 3.10 έχουμε πως

$$\mathbb{P}_{po}(T_{po} = n) = \frac{\mathbb{P}_{po}(X_1 + X_2 + \dots + X_n = n - 1)}{n}$$

Όμως σύμφωνα με αυτά που ξέρουμε από εισαγωγικά μαθήματα Πιθανοτήτων, το άθροισμα n τυχαίων μεταβλητών που ακολουθούν την κατανομή Poisson με παράμετρο λ ακολουθεί μια κατανομή Poisson με παράμετρο λn . Άρα

$$\mathbb{P}_{po}(T_{po} = n) = \frac{\frac{(\lambda n)^{n-1}}{(n-1)!} e^{-\lambda n}}{n} = \frac{(\lambda n)^{n-1}}{n!} e^{-\lambda n}$$

□

Πλέον είμαστε έτοιμοι να δείξουμε και ένα ασυμπτωτικό αποτέλεσμα πάνω στην πιθανότητα της ουράς αναμονής Poisson.

Θεώρημα 3.6. Για μια ουρά αναμονής Poisson με παράμετρο λ , ισχύει ότι καθώς το $n \rightarrow \infty$

$$\mathbb{P}_\lambda(T_{po} = n) = \frac{e^{-I_\lambda n}}{\lambda \sqrt{2\pi n^3}} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \quad (3.11)$$

με το $I_\lambda = \lambda - 1 - \log \lambda$. Συγκεκριμένα, όταν το $\lambda 1$ έχουμε πως

$$\mathbb{P}_1(T_{po} = n) = \frac{1}{\sqrt{2\pi n^3}} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \quad (3.12)$$

όπου με τους δείκτες στις πιθανότητες συμβολίζουμε την τιμή του μέσου όρου της Poisson.

Απόδειξη. Από το θεώρημα 3.5 παίρνουμε πως

$$\begin{aligned} \mathbb{P}_\lambda(T_{po} = n) &= \frac{(\lambda n)^{n-1}}{n!} e^{-\lambda n} \\ &= \frac{\lambda^n}{\lambda} e^{-\lambda n + n} \frac{n^{n-1}}{n!} e^{-n} \\ &= \frac{\lambda^n}{\lambda} e^{-(\lambda-1)n} \mathbb{P}_1(T_{po} = n) \\ &= \frac{1}{\lambda} e^{-I_\lambda n} \mathbb{P}_1(T_{po} = n) \end{aligned}$$

Άρα αρκεί να δείξουμε τη 3.12 για να αποδειχθεί το θεώρημα. Από το θεώρημα 3.5 παίρνουμε επίσης πως για $\lambda = 1$

$$\begin{aligned} \mathbb{P}_1(T_{po} = n) &= \frac{n^{n-1}}{n!} e^{-n} \\ &= \frac{1}{\sqrt{2\pi n^3}} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \end{aligned}$$

από την άμεση αντικατάσταση του τύπου του Stirling:

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right)$$

□

3.4 Σχέση Poisson και διωνυμικών ουρών αναμονής

Όπως ξέρουμε και από εισαγωγικά μαθήματα Πιθανοτήτων, μια τυχαία μεταβλητή που ακολουθεί τη διωνυμική κατανομή με παραμέτρους n και p και το n μεγαλώνει ενώ ταυτόχρονα το p μικραίνει έτσι ώστε το γινόμενό τους να είναι σταθερό ($np = n'p'$) ακολουθεί την κατανομή Poisson στο όριο (καθώς $n \rightarrow \infty$) με παράμετρο $\lambda = np$. Όταν όμως μελετούμε τους συνολικούς πληθυσμούς T_{bin} και T_{po} αυτοί δεν είναι ακριβώς τυχαίες μεταβλητές που ακολουθούν τη διωνυμική και την Poisson κατανομή αντίστοιχα όπως φαίνεται από το θεώρημα 3.4. Παρόλα αυτά οι κατανομές τους είναι αρκετά κοντά και το παρακάτω θεώρημα είναι μια σαφέστερη έκφραση αυτού του γεγονότος, αφού εξηγεί ότι οι ουρές των δύο κατανομών διαφέρουν κατά έναν όρο ο οποίος τείνει στο 0, καθώς $n \rightarrow \infty$ τουλάχιστον τόσο γρήγορα όσο το $\frac{1}{n}$.

Θεώρημα 3.7. Μια διωνυμική ουρά αναμονής με παράμετρους n και p και μια ουρά αναμονής Poisson με παράμετρο $\lambda = np$ συνδέονται για $k \geq 1$ με τη σχέση

$$\mathbb{P}_{bin}(T_{bin} \geq k) = \mathbb{P}_{po}(T_{po} \geq k) + e_n(k) \quad (3.13)$$

όπου οι $T_{\text{bin}}, T_{\text{po}}$ είναι αντίστοιχα ο συνολικός πληθυσμός της διωνυμικής ουράς αναμονής με παραμέτρους n και p και της ουράς αναμονής Poisson με παράμετρο $\lambda = np$ και

$$|e_n(k)| \leq \frac{\lambda^2}{n} \sum_{s=1}^{k-1} \mathbb{P}_{\text{po}}(T_{\text{po}} \geq s) \leq \frac{\lambda^2(k-1)}{n} \quad (3.14)$$

Πριν προχωρήσουμε στην απόδειξη, θα διατυπώσουμε και θα αποδείξουμε ένα θεώρημα σύζευξης (coupling) το οποίο θα χρειαστούμε.

Θεώρημα 3.8. Έστω μια ακολουθία $I_i, i = 1, \dots, n$ ανεξάρτητων τυχαίων μεταβλητών που ακολουθούν κατανομή Bernoulli, $\text{Be}(p_i)$ και το $\lambda = \sum_{i=1}^n p_i$. Τότε εάν n τυχαία μεταβλητή $X = \sum_{i=1}^n I_i$ και n τυχαία μεταβλητή Y ακολουθεί την κατανομή Poisson, $\text{Po}(\lambda)$, υπάρχει μια σύζευξη τους (\hat{X}, \hat{Y}) τέτοια ώστε

$$\mathbb{P}(\hat{X} \neq \hat{Y}) \leq \sum_{i=1}^n p_i^2$$

Πιο συγκεκριμένα, εάν κάθε $p_i = p$ $n X \sim \text{Bin}(n, p)$ και $n Y \sim \text{Po}(\lambda)$ με $\lambda = np$, το θεώρημα δίνει

$$\mathbb{P}(\hat{X} \neq \hat{Y}) \leq \sum_{i=1}^n p^2 = np^2 = \frac{\lambda^2}{n}$$

Απόδειξη. Έστω n ακολουθία $J_i, i = 1, \dots, n$ ανεξάρτητων τυχαίων μεταβλητών που ακολουθούν την κατανομή Poisson με παράμετρο p_i : $\text{Po}(p_i)$. Τότε για $x \in \mathbb{N}_0$

$$\mathbb{P}(I_i = x) = p_i \mathbb{1}_{\{x=1\}} + (1 - p_i) \mathbb{1}_{\{x=0\}}$$

$$\mathbb{P}(J_i = x) = e^{-p_i} \frac{p_i^x}{x!}$$

Έστω n σύζευξη (\hat{I}_i, \hat{J}_i) των I_i, J_i με τα ζεύγη να είναι ανεξάρτητα για κάθε i . Έστω επίσης πως για το μέτρο σύζευξης ισχύει ότι

$$\mathbb{P}(\hat{I}_i = \hat{J}_i = x) = \min\{\mathbb{P}(I_i = x), \mathbb{P}(J_i = x)\} = \begin{cases} 1 - p_i & x=0 \\ e^{-p_i} p_i, & x=1 \\ 0, & x \geq 2 \end{cases}$$

όπου χρησιμοποιήσαμε την ανισότητα $1 - p_i \leq e^{-p_i}$ για το $x=0$ και το γεγονός πως $e^{-p_i} < 1$ για $p_i \in (0, 1]$ για το $x=1$. Τότε αφού $1 - p_i \leq e^{-p_i} \implies 1 - e^{-p_i} \leq p_i$ έχουμε ότι

$$\mathbb{P}(\hat{I}_i \neq \hat{J}_i) = 1 - (1 - p_i) - e^{-p_i} p_i = p_i(1 - e^{-p_i}) \leq p_i^2$$

Τώρα, έστω οι τυχαίες μεταβλητές $\hat{X} = \sum_{i=1}^n \hat{I}_i$ και $\hat{Y} = \sum_{i=1}^n \hat{J}_i$. Η \hat{X} θα έχει την ίδια κατανομή όπως και $n X = \sum_{i=1}^n I_i$ και $n \hat{Y}$ όπως $n Y = \sum_{i=1}^n J_i$, δηλαδή την $\text{Po}(p_1 + \dots + p_2)$.

Επίσης στο ενδεχόμενο $\{\hat{X} \neq \hat{Y}\} = \{\sum_{i=1}^n \hat{I}_i \neq \sum_{i=1}^n \hat{J}_i\}$ θα πρέπει οι συντεταγμένες του ζεύγους (\hat{I}_i, \hat{J}_i) να διαφέρουν σε τουλάχιστον ένα i , χωρίς όμως αυτό να συνεπάγεται αναγκαστικά πως το $\{\sum_{i=1}^n \hat{I}_i \neq \sum_{i=1}^n \hat{J}_i\}$. Για παράδειγμα, για το $(\hat{I}_i, \hat{J}_i), i = 1, 2$ θα μπορούσαμε να έχουμε διαφορετικές συντεταγμένες και για $i=1, 2$ χωρίς το $\{\sum_{i=1}^n \hat{I}_i \neq \sum_{i=1}^n \hat{J}_i\}$, π.χ. $\hat{I}_1 = 0, \hat{I}_2 = 1, \hat{J}_1 = 1, \hat{J}_2 = 0$. Οπότε ισχύει ότι $\{\hat{X} \neq \hat{Y}\} = \{\sum_{i=1}^n \hat{I}_i \neq \sum_{i=1}^n \hat{J}_i\} \subset \{\bigcup_{i=1}^n (\hat{I}_i \neq \hat{J}_i)\}$. Μαζί με την ανισότητα Boole και το παραπάνω αποτέλεσμα καταλήγουμε πως

$$\mathbb{P}(\hat{X} \neq \hat{Y}) \leq \mathbb{P}\left(\bigcup_{i=1}^n (\hat{I}_i \neq \hat{J}_i)\right) \leq \sum_{i=1}^n \mathbb{P}(\hat{I}_i \neq \hat{J}_i) \leq \sum_{i=1}^n p_i^2$$

□

Απόδειξη. Θεωρήματος 3.7.

Για τις δύο ουρές αναμονής, ο αριθμός των απογόνων σε κάθε χρονική στιγμή είναι οι τυχαίες μεταβλητές $X = X_{\text{bin}} \sim \text{Bin}(n, p)$ και $Y = X_{\text{po}} \sim \text{Po}(\lambda)$ αντίστοιχα. Χρησιμοποιούμε τη σύζευξη (\hat{X}, \hat{Y}) που δίνει το θεώρημα 3.8 για τις X, Y την οποία από εδώ και πέρα θα συμβολίζουμε χωρίς τα καπελάκια. Για αυτήν, τα ζεύγη (X_i, Y_i) είναι ανεξάρτητα ανά δύο για κάθε i και

$$\mathbb{P}(X_i \neq Y_i) \leq \frac{\lambda^2}{n} \quad (3.15)$$

όπου το \mathbb{P} συμβολίζει το μέτρο πιθανότητας του ζεύγους (X, Y) για τη σύζευξη. Τις πιθανότητες των ουρών τις γράφουμε ως

$$\mathbb{P}_{\text{bin}}(T_{\text{bin}} \geq k) = \mathbb{P}(T_{\text{bin}} \geq k, T_{\text{po}} \geq k) + \mathbb{P}(T_{\text{bin}} \geq k, T_{\text{po}} < k)$$

$$\mathbb{P}_{\text{po}}(T_{\text{po}} \geq k) = \mathbb{P}(T_{\text{po}} \geq k, T_{\text{bin}} \geq k) + \mathbb{P}(T_{\text{po}} \geq k, T_{\text{bin}} < k)$$

Αφαιρώντας τις παραπάνω σχέσεις παίρνουμε ότι

$$|\mathbb{P}_{\text{bin}}(T_{\text{bin}} \geq k) - \mathbb{P}_{\text{po}}(T_{\text{po}} \geq k)| \leq \max\{\mathbb{P}(T_{\text{bin}} \geq k, T_{\text{po}} < k), \mathbb{P}(T_{\text{po}} \geq k, T_{\text{bin}} < k)\} \quad (3.16)$$

Το ενδεχόμενο $\{T_{\text{bin}} \geq k\}$ καθορίζεται αποκλειστικά από τις τιμές των X_1, X_2, \dots, X_{k-1} . Συγκεκριμένα, αν ξέρουμε τις τιμές των X_1, X_2, \dots, X_{k-1} , τότε μπορούμε να προσδιορίσουμε εάν σε κάποιο χρόνο $t < k$ ικανοποιείται η συνθήκη $X_1 + X_2 + \dots + X_t = t - 1$ οπότε και θα ισχύει ότι $T_{\text{bin}} < k$. Άμα δεν υπάρχει τέτοιο t θα έχουμε ότι το $\{T_{\text{bin}} \geq k\}$ ικανοποιείται. Παρόμοια και για τη διαδικασία Poisson αν ξέρουμε το Y_1, Y_2, \dots, Y_{k-1} μπορούμε να προσδιορίσουμε εάν το $\{T_{\text{po}} \geq k\}$ ικανοποιείται.

Όταν ένα από τα ενδεχόμενα $\{T_{\text{bin}} \geq k, T_{\text{po}} < k\}$, $\{T_{\text{po}} \geq k, T_{\text{bin}} < k\}$ ικανοποιείται, τότε οπωσδήποτε υπάρχει τουλάχιστον ένα $s < k$ για το οποίο $X_s \neq Y_s$. Άρα μπορούμε να πούμε ότι

$$\{T_{\text{bin}} \geq k, T_{\text{po}} < k\} \subset \left\{ \bigcup_{s=1}^{k-1} (X_i = Y_i \ \forall i \leq s-1, X_s \neq Y_s, T_{\text{bin}} \geq k) \right\} \quad (3.17)$$

με τη συνθήκη $X_i = Y_i \ \forall i \leq s-1$ να τίθεται σε ισχύ από το $s=2$ και μετά.

Τότε επιστρατεύοντας την ανισότητα του Boole:

$$\begin{aligned} \mathbb{P}(T_{\text{bin}} \geq k, T_{\text{po}} < k) &\leq \mathbb{P}\left(\bigcup_{s=1}^{k-1} (X_i = Y_i \ \forall i \leq s-1, X_s \neq Y_s, T_{\text{bin}} \geq k)\right) \\ &\leq \sum_{s=1}^{k-1} \mathbb{P}(X_i = Y_i \ \forall i \leq s-1, X_s \neq Y_s, T_{\text{bin}} \geq k) \end{aligned} \quad (3.18)$$

Τώρα παρατηρούμε ότι όταν $X_i = Y_i \ \forall i \leq s-1$ και $T_{\text{bin}} \geq k$, τότε θα ισχύει ότι $\{X_1 + X_2 + \dots + X_i \geq i\} = \{Y_1 + Y_2 + \dots + Y_i \geq i\}$ για κάθε $i \leq s-1$ και άρα και ότι $T_{\text{po}} \geq s$. Επιπλέον επειδή όπως είπαμε το ενδεχόμενο $\{T_{\text{po}} \geq s\}$ εξαρτάται μόνο από τα Y_1, Y_2, \dots, Y_{s-1} , είναι ανεξάρτητο από το εάν $X_s \neq Y_s$. Άρα μπορούμε να γράψουμε

$$\mathbb{P}(T_{\text{bin}} \geq k, T_{\text{po}} < k) \leq \sum_{s=1}^{k-1} \mathbb{P}(T_{\text{po}} \geq s, X_s \neq Y_s) \quad (3.19)$$

$$\leq \sum_{s=1}^{k-1} \mathbb{P}(T_{\text{po}} \geq s) \mathbb{P}(X_s \neq Y_s) \quad (3.20)$$

Μαζί με την 3.15 γίνεται

$$\mathbb{P}(T_{\text{bin}} \geq k, T_{\text{po}} < k) \leq \frac{\lambda^2}{n} \sum_{s=1}^{k-1} \mathbb{P}(T_{\text{po}} \geq s)$$

Με παρόμοιο τρόπο και μια μικρή αλλαγή στο επιχειρήμα από τη σχέση 3.18 στη 3.19 παίρνουμε ότι

$$\mathbb{P}(T_{\text{po}} \geq k, T_{\text{bin}} < k) \leq \frac{\lambda^2}{n} \sum_{s=1}^{k-1} \mathbb{P}(T_{\text{po}} \geq s)$$

και άρα η 3.16 δίνει

$$|\mathbb{P}_{\text{bin}}(T_{\text{bin}} \geq k) - \mathbb{P}_{\text{po}}(T_{\text{po}} \geq k)| \leq \frac{\lambda^2}{n} \sum_{s=1}^{k-1} \mathbb{P}_{\text{po}}(T_{\text{po}} \geq s)$$

και το ζητούμενο αποδείχθη. \square

Θεώρημα 3.9. Έστω μια ουρά αναμονής Poisson με μέσο όρο λ . Τότε για κατάλληλα μεγάλο k ισχύει για το συνολικό πληθυσμό T της διαδικασίας ότι

$$\mathbb{P}(k \leq T < \infty) \leq \frac{e^{-kI_\lambda}}{1 - e^{-I_\lambda}} \quad (3.21)$$

όπου το $I_\lambda = \lambda - 1 - \log \lambda$.

Απόδειξη.

$$\begin{aligned} \mathbb{P}(k \leq T < \infty) &= \sum_{n=k}^{\infty} \mathbb{P}(T = n) \\ &= \frac{1}{\lambda} \sum_{n=k}^{\infty} \frac{e^{-nI_\lambda}}{\sqrt{2\pi n^3}} (1 + o(1)) && \text{από το θεώρημα 3.6} \\ &\leq \sum_{n=k}^{\infty} e^{-nI_\lambda} && \text{για κατάλληλα μεγάλο } k \\ &= \frac{e^{-kI_\lambda}}{1 - e^{-I_\lambda}} \end{aligned}$$

\square

3.5 Κλαδωτές διαδικασίες

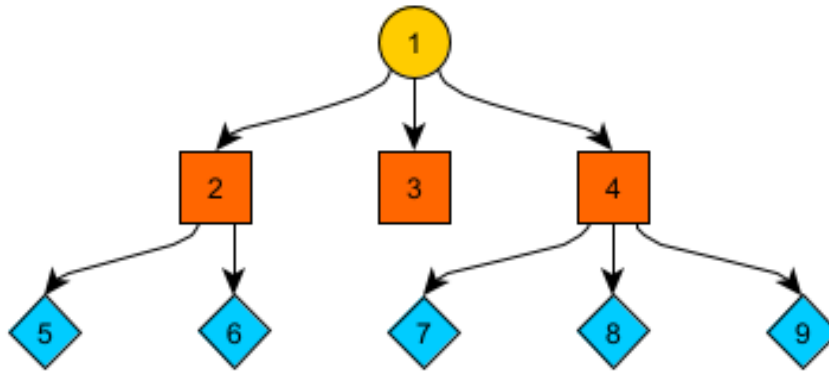
Σε αυτήν την ενότητα θα φανεί χρήσιμο να μιλήσουμε λίγο για τις κλαδωτές διαδικασίες, καθώς θα μας δώσουν κάποια αποτελέσματα για τις ουρές αναμονής αρκετά πιο εύκολα.

Σε μια κλαδωτή διαδικασία ο πληθυσμός μετρείται ανά γενιές. Δηλαδή με Z_n θα συμβολίζουμε τον πληθυσμό της διαδικασίας την n -οστή γενιά (χρονική στιγμή). Αυτό που συμβαίνει είναι πως κάθε ένα από τα Z_n μέλη της n -οστής γενιάς θα γεννά έναν αριθμό απογόνων σύμφωνα με την ίδια πάντα κατανομή, και ύστερα θα "πεθαίνουν", καθιστώντας έτσι τα παιδιά τους (Z_{n+1}) ως τον νέο πληθυσμό. Γράφουμε τυπικά πως

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_{n-1,i} \quad (3.22)$$

με $Z_0 = 1$ και οι $X_{n-1,i}$ για $n, i \in \mathbb{N}$ να είναι τυχαίες μεταβλητές που ακολουθούν όλες την ίδια κατανομή, έστω την X .

Η αρχή μιας κλαδωτής διαδικασίας φαίνεται στο παρακάτω σχήμα.



Σχήμα 3.4

Θεώρημα 3.10. Για κάθε $n \geq 0$ και με το $\mu = \mathbb{E}(Z_1) = \mathbb{E}(X)$ να είναι ο αναμενόμενος αριθμός απογόνων κάθε ατόμου, ισχύει ότι

$$\mathbb{E}(Z_n) = \mu^n \quad (3.23)$$

Απόδειξη. Στην απόδειξη θα ακολουθήσουμε τη μέθοδο της μαθηματικής επαγωγής. Αρχικά, βλέπουμε πως για $n = 0, 1$ έχουμε ότι

$$\mathbb{E}(Z_0) = \mathbb{E}(1) = 1 = \mu^0$$

$$\mathbb{E}(Z_1) = \mathbb{E}(X_{0,1}) = \mu = \mu^1$$

Έστω λοιπόν ότι ισχύει $\mathbb{E}(Z_n) = \mu^n$ για κάποιο γενικό n . Θα δείξουμε πως η σχέση είναι αληθής και για το $n+1$.

$$\begin{aligned} \mathbb{E}(Z_{n+1}) &= \mathbb{E}\left(\sum_{i=1}^{Z_n} X_{n,i}\right) \\ &= \sum_{i=1}^{Z_n} \mathbb{E}(X_{n,i}) \\ &= \sum_{i=1}^{Z_n} \mu \\ &= \mathbb{E}(Z_n) \mu && \text{από την ιδιότητα του Wald} \\ &= \mu^n \mu && \text{από το επαγωγικό βήμα} \\ &= \mu^{n+1} \end{aligned}$$

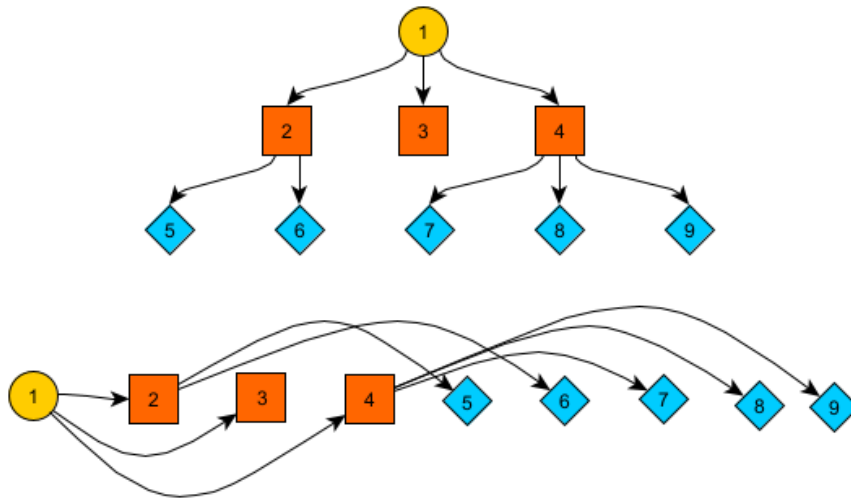
□

Πριν συνεχίσουμε, επισημαίνουμε πως ο συνολικός πληθυσμός της κλαδωτής διαδικασίας είναι ο

$$T = \sum_{n=0}^{\infty} Z_n \quad (3.24)$$

Αυτό σημαίνει πως ο συνολικός πληθυσμός στην κλαδωτή διαδικασία είναι ο ίδιος όπως και στην αναπαράστασή της ως ένας τυχαίος περίπατος (δηλαδή μια ουρά αναμονής), αφού αυτό που μετρούσαμε ως χρόνο στην ουρά αναμονής ($T = \inf\{t : S_t = 0\}$) είναι ακριβώς ισοδύναμο με τον τρόπο μέτρησης του συνολικού πληθυσμού στην κλαδωτή διαδικασία ($T = \sum_{n=0}^{\infty} Z_n$). Αν για παράδειγμα έχουμε την

προηγούμενη κλαδωτή διαδικασία η οποία όμως δε συνεχίζει πέρα από τη δεύτερη γενιά της, αυτή μπορεί να αναπαρασταθεί σε μια σειρά στην οποία κάθε χρονική στιγμή μια κορυφή (από αριστερά προς τα δεξιά) γεννάει απογόνους και στη συνέχεια πεθαίνει, δηλαδή μια αντίστοιχη ουρά αναμονής. Έτσι για το συγκεκριμένο περίπατο θα έχουμε πως $S_t, t \in \{0, \dots, 9\} = (1, 3, 4, 3, 5, 4, 3, 2, 1, 0)$ και άρα το $T = 9$. Στην κλαδωτή διαδικασία το μέτρομα θα γίνει $1 + 3 + 5 = 9$.



Σχήμα 3.5

Το επόμενο θεώρημα είναι και ο βασικός λόγος της ενασχόλησής μας με τις κλαδωτές διαδικασίες, καθώς διαφορετικά θα ήταν δύσκολο να αποδειχθεί αυτό το αποτέλεσμα, ενώ τώρα θα το χρησιμοποιούμε κατευθείαν για τις ουρές αναμονής.

Θεώρημα 3.11. Για μια κλαδωτή διαδικασία όπου ο αριθμός των απογόνων κάθε ατόμου ακολουθεί την κατανομή X με $\mathbb{E}(X) = \mu < 1$ έχουμε πως

$$\mathbb{E}(T) = \frac{1}{1 - \mu} \quad (3.25)$$

Απόδειξη.

$$\begin{aligned} \mathbb{E}(T) &= \mathbb{E}\left(\sum_{n=0}^{\infty} Z_n\right) \\ &= \sum_{n=0}^{\infty} \mathbb{E}(Z_n) \\ &= \sum_{n=0}^{\infty} \mu^n && \text{από την 3.23} \\ &= \frac{1}{1 - \mu} \end{aligned}$$

□

Κεφάλαιο 4

Μεταβολή Φάσης στο Τυχαίο Γράφημα

Έστω ότι είμαστε σε ένα τυχαίο γράφημα $G(n,p)$ με $V=\{1, 2, \dots, n\}$ και κάθε ακμή εμφανίζεται στο γράφημα με πιθανότητα p ανεξάρτητα από την εμφάνιση ή μη των υπολοίπων ακμών. Τη συνεκτική συνιστώσα ενός κόμβου i θα τη συμβολίζουμε με C_i . Επιπλέον, το μέγεθός της το συμβολίζουμε με $|C_i|$. Το μέγεθος της μεγιστοτικής συνεκτικής συνιστώσας το συμβολίζουμε με $|C_{\max}|$ και θα είναι η κύρια ενασχόλησή μας σε αυτό το κεφάλαιο. Συνήθως θα αναφερόμαστε και στην “μεγαλύτερη συνιστώσα”, χωρίς να μας ενδιαφέρει εάν υπάρχει και κάποια άλλη με το ίδιο μέγεθος, καθώς μόνο το μέγεθος θα μας απασχολεί.

Τα αποτελέσματα αυτού του κεφαλαίου έγκεινται στη συμπεριφορά της μεγαλύτερης συνιστώσας σε τρεις διαφορετικές περιοχές για την τιμή του λ , όπου $p = \frac{\lambda}{n}$. Συνοπτικά, όταν το λ είναι μια σταθερά μικρότερη του 1, η μεγαλύτερη συνιστώσα έχει μέγεθος τάξης μεγέθους $\log n$. Όταν το λ είναι μια σταθερά μεγαλύτερη του 1, το $|C_{\max}|$ είναι τάξης μεγέθους n και μάλιστα η δεύτερη σε μέγεθος συνιστώσα είναι μικρότερη από $K \log n$. Αυτός ο σχηματισμός μιας τόσο μεγάλης συνιστώσας μας ωθεί να την καλούμε “γιγαντιαία συνιστώσα”. Στην τρίτη περίπτωση μελετούμε το μέγεθος της μεγαλύτερης συνιστώσας καθώς το λ είναι πολύ κοντά στο 1. Παραμετροποιώντας το λ ως $\lambda = 1 + \theta n^{-1/3}$ παρατηρούμε πως καθώς το n μεγαλώνει, το μέγεθος της μεγαλύτερης συνιστώσας είναι τάξης μεγέθους $n^{2/3}$. Είναι δηλαδή σημαντικά μικρότερη από την περίπτωση όπου το λ είναι μια σταθερά μεγαλύτερη του 1, αλλά ταυτόχρονα αρκετά μεγαλύτερη σε σχέση με την περίπτωση όπου το λ είναι μια σταθερά μικρότερη του 1.

Ωστόσο πριν προχωρήσουμε στην παρουσίαση αυτών των αποτελεσμάτων θα χρειαστεί να ορίσουμε μια διαδικασία παρόμοια με την ουρά αναμονής Poisson και τη διωνυμική ουρά αναμονής, η οποία όμως θα βρίσκεται σε ένα τυχαίο γράφημα και άρα θα πρέπει να λαμβάνει υπόψη πως δε θα μπορεί να επεκτείνεται επ'άπειρον, καθώς υπάρχει πεπερασμένο πλήθος κορυφών στο γράφημα. Αυτή θα μας βοηθήσει να αναλύσουμε το μέγεθος μιας συνεκτικής συνιστώσας στο τυχαίο γράφημα και έτσι να βγάλουμε συμπεράσματα και για τη μεγαλύτερη συνεκτική συνιστώσα.

4.1 Διαδικασία Γραφήματος

Στη συνέχεια θα παρουσιάσουμε έναν τρόπο για την εξερεύνηση της συνεκτικής συνιστώσας ενός κόμβου ο οποίος είναι γνωστός ως Αναζήτηση Κατά Πλάτος (Breadth First Search ή BFS).

4.1.1 Αναζήτηση Κατά Πλάτος

Έστω ότι έχουμε ένα οποιοδήποτε γράφημα και θέλουμε να βρούμε από ποιους κόμβους αποτελείται η συνεκτική συνιστώσα μιας κορυφής i .

Χωρίζουμε όλες τις κορυφές του γραφήματος σε ενεργές, ανενεργές και ουδέτερες. Αρχικά, μόνο η κορυφή i είναι ενεργή και όλες οι υπόλοιπες είναι ουδέτερες. Βρίσκουμε όλους τους γείτονες της i και τους θέτουμε ως ενεργούς με μια οποιαδήποτε διάταξη. Αυτοί αποτελούν το πρώτο “στρώμα” εξερεύνησης του αλγορίθμου. Στη συνέχεια, αφού έχει τελειώσει η εξερεύνηση της κορυφής i , τη θέτουμε ως ανενεργή. Τώρα, διαλέγουμε από τις ενεργές κορυφές την πρώτη στη διάταξη και βρίσκουμε τους γείτονές της που είναι ουδέτεροι. Τους θέτουμε ενεργούς και την κορυφή που μόλις εξερευνήσαμε τη βάζουμε στις ανενεργές. Συνεχίζουμε αυτήν τη διαδικασία μέχρι να μην υπάρχουν ενεργές κορυφές.

Γιατί δουλεύει ο αλγόριθμος;

1) Αν η αρχική κορυφή ενώνεται μέσω κάποιου μονοπατιού με κάποια άλλη κορυφή, τότε κάθε κορυφή αυτού του μονοπατιού θα εξερευνηθεί μέσω της Αναζήτησης Κατά Πλάτος. Η δεύτερη κορυφή α_2 του μονοπατιού θα εξερευνηθεί σίγουρα στην πρώτη εξερεύνηση και θα γίνει ενεργή. Η επόμενη κορυφή α_3 θα εξερευνηθεί κατά την προσπέλαση των γειτόνων της α_2 (ή νωρίτερα) και θα γίνει ενεργή κ.ο.κ. έως ότου και η τελευταία κορυφή του μονοπατιού εξερευνηθεί. Επίσης εάν κάποια κορυφή δεν ανήκει στη συνιστώσα της κορυφής i , δε θα της δοθεί ποτέ η ευκαιρία να εξερευνηθεί, εφόσον δεν είναι γειτονική κορυφή καμίας κορυφής που ανήκει στη C_i .

2) Ο αλγόριθμος θα τερματίσει, αφού σε κάθε βήμα μια κορυφή γίνεται ανενεργή (δε μπορεί να ξαναγίνει ενεργή) και άρα το πολύ σε n εξερευνήσεις ο αριθμός των ενεργών κορυφών θα μηδενιστεί.

Η Αναζήτηση Κατά Πλάτος συνήθως χρησιμοποιείται για την εύρεση της απόστασης και ενός ελαχιστοτικού μονοπατιού ανάμεσα σε δύο κορυφές, αλλά επειδή εμείς θα χρειαστούμε μόνο την έννοια της συνεκτικής συνιστώσας δεν μπαίνουμε σε περισσότερες λεπτομέρειες.

4.1.2 Ορισμός Διαδικασίας Γραφήματος

Για να αναλύσουμε τις συνεκτικές συνιστώσες στο τυχαίο γράφημα $G(n,p)$ θα χρειαστεί να τις επεξηγήσουμε σε όρους μιας διαδικασίας.

Έστω ότι θέλουμε να βρούμε το μέγεθος της συνεκτικής συνιστώσας μιας κορυφής i . Την αρχική στιγμή $t=0$ ο μόνος κόμβος είναι ο i και θέτουμε $S_0 = 1$. Την επόμενη χρονική στιγμή, όπως στην Αναζήτηση Κατά Πλάτος κάποιες κορυφές θα γίνουν ενεργές και μια θα γίνει ανενεργή. Τότε $S_1 = |\text{ενεργές κορυφές στο χρόνο } t=1|$ και αυτή η διαδικασία συνεχίζεται έως ότου το $S_t = 0$. Την πρώτη φορά που θα συμβεί αυτό τη συμβολίζουμε με T και τότε η Αναζήτηση Κατά Πλάτος έχει τελειώσει την εξερεύνηση της συνεκτικής συνιστώσας και άρα το $|C_i| = T$.

Δηλαδή αν συμβολίσουμε με X_t την τυχαία μεταβλητή που δίνει τον αριθμό των κορυφών που θα γίνουν ενεργές τη χρονική στιγμή t , τότε για τον αριθμό των ενεργών κορυφών τη χρονική στιγμή t , S_t , ισχύει πως

$$S_0 = 1, S_t = S_{t-1} + X_t - 1 \quad (4.1)$$

και αν συμβολίσουμε και το T ως

$$T = \inf\{t : S_t = 0\}$$

το συνολικό μέγεθος της συνεκτικής συνιστώσας του κόμβου i είναι $|C_i| = T$.

Αυτή η σχέση είναι η ίδια με αυτή που ισχύει στις ουρές αναμονής (σχέση 3.1). Όπως θα δούμε όμως αμέσως τώρα, η βασική διαφορά είναι στην τυχαία μεταβλητή X_t .

Σε κάθε βήμα, ο αριθμός των ουδέτερων κορυφών στο γράφημα μένει σταθερός ή μειώνεται και άρα όταν εξερευνούμε μια κορυφή, το πλήθος των υπολοίπων κορυφών τις οποίες μπορεί να θέσει ως ενεργές μπορεί να αλλάξει σε κάθε νέα εξερεύνηση. Στις ουρές αναμονής το πλήθος των απογόνων (που συμβολίζει το X_t) είναι σε κάθε χρονική στιγμή τυχαίες μεταβλητές που είναι ανεξάρτητες και ισόνομες.

Συγκεκριμένα τώρα για το τυχαίο γράφημα $G(n,p)$ κάθε ακμή έχει πιθανότητα p να εμφανιστεί στο γράφημα ανεξάρτητα από τις υπόλοιπες. Άρα κατά την εξερεύνηση μιας κορυφής, το πόσες κορυφές θα γίνουν ενεργές εξαρτάται μόνο από το πόσες είναι ουδέτερες, αφού κάθε ουδέτερη κορυφή έχει πιθανότητα p να συνδέεται με την εν λόγω κορυφή που εξερευνείται. Οι ουδέτερες κορυφές πριν την εξερεύνηση της χρονικής στιγμής t είναι $n - S_{t-1} - (t - 1)$, καθώς οι $t-1$ κορυφές είναι ανενεργές και οι υπόλοιπες S_{t-1} είναι ενεργές. Αφού κάθε μία από αυτές έχει πιθανότητα p να συνδέεται με την κορυφή που εξερευνείται, τότε ο αριθμός των νέων κορυφών που θα γίνουν ενεργές τη χρονική στιγμή t ακολουθεί τη διωνυμική κατανομή, δηλαδή

$$X_t \sim \text{Bin}(n - S_{t-1} - (t - 1), p) \quad (4.2)$$

4.2 Σύνδεση της διαδικασίας γραφήματος με τη διωνυμική ουρά αναμονής

Συνήθως παραμετροποιούμε την πιθανότητα εμφάνισης μιας ακμής στο τυχαίο γράφημα $G(n,p)$ ως $p = \frac{\lambda}{n}$, όπου το $\lambda \in [0, n]$. Όπως θα δούμε και στη συνέχεια ιδιαίτερο ενδιαφέρον παρουσιάζει η συμπεριφορά του τυχαίου γραφήματος καθώς το λ γίνεται από μικρότερο του 1, μεγαλύτερό του. Υπενθυμίζεται πως ο αναμενόμενος βαθμός του τυχαίου γραφήματος είναι $(n-1)p$ και άρα για $\lambda=1$ γίνεται $(n-1)\frac{1}{n} \simeq 1$ για $n \rightarrow \infty$. Δηλαδή η συμπεριφορά του τυχαίου γραφήματος αλλάζει δραστικά ανάλογα με το εάν ο αναμενόμενος βαθμός κάθε κορυφής είναι μικρότερος ή μεγαλύτερος από 1.

Σε αυτήν την ενότητα θα δούμε δύο θεωρήματα τα οποία θα αποδειχθούν πολύ χρήσιμα για τη συνέχεια.

Επίσης, από εδώ και στο εξής θα χρησιμοποιείται η παραμετροποίηση $p = \frac{\lambda}{n}$, εκτός και αν δηλώνεται κάτι διαφορετικό και υπενθυμίζεται πως με $\mathbb{P}_{n,p}$ συμβολίζεται το μέτρο πιθανότητας σε μια διωνυμική ουρά αναμονής με παραμέτρους n και p . Επίσης, με \mathbb{P}_λ θα συμβολίζεται το μέτρο πιθανότητας στη διαδικασία γραφήματος, όπου $\lambda = np$.

Τότε παίρνουμε το παρακάτω θεώρημα.

Θεώρημα 4.1. Για κάθε k ισχύει πως

$$\mathbb{P}_\lambda(|C_i| \geq k) \leq \mathbb{P}_{n,p}(T_{\text{bin}} \geq k)$$

όπου $\lambda = np$ και i είναι μια οποιαδήποτε κορυφή του τυχαίου γραφήματος.

Απόδειξη. Έστω $N_t = n - S_t - t$ να είναι ο αριθμός των ουδέτερων κορυφών μετά την εξερεύνηση ενός κόμβου τη χρονική στιγμή t . Τότε δεδομένου του N_{t-1} , όπως είπαμε και παραπάνω, στη διαδικασία γραφήματος το $X_t \sim \text{Bin}(n - S_{t-1} - (t-1), p)$ ή αλλιώς $X_t \sim \text{Bin}(N_{t-1}, p)$.

Έστω η τυχαία μεταβλητή Y_t που ακολουθεί διωνυμική κατανομή με $Y_t \sim \text{Bin}(n - N_{t-1}, p)$.

Τότε η τυχαία μεταβλητή $X'_t = X_t + Y_t$ ακολουθεί τη διωνυμική κατανομή $\text{Bin}(n, p)$. Επίσης, αφού $Y_t \geq 0$ ισχύει πως $X'_t \geq X_t$.

Θέτοντας $S'_t = X'_1 + X'_2 + \dots + X'_t - (t-1) \geq X_1 + X_2 + \dots + X_t - (t-1) = S_t$, έχουμε ότι

$$\mathbb{P}_\lambda(|C_i| \geq k) = \mathbb{P}(S_t > 0, \forall t \leq k-1) \leq \mathbb{P}(S'_t > 0, \forall t \leq k-1) = \mathbb{P}_{n,p}(T_{\text{bin}} \geq k)$$

αφού στη διωνυμική ουρά αναμονής ο αριθμός των απογόνων κάθε στιγμή ακολουθεί την $\text{Bin}(n, p)$ και ο συνολικός πληθυσμός είναι μεγαλύτερος ή ίσος από το k όταν υπάρχουν ζωντανά άτομα σε κάθε χρονική στιγμή μέχρι και το $k-1$. \square

Ένας άλλος τρόπος να σκεφτούμε την παραπάνω απόδειξη είναι ο εξής:

Έχουμε μια διαδικασία γραφήματος και αρχίζουμε την Αναζήτηση Κατά Πλάτος από ένα κόμβο i για να βρούμε το μέγεθος της συνεκτικής συνιστώσας του, C_i . Σε κάθε εξερεύνηση ενός κόμβου j κάνουμε το ακόλουθο. Διαλέγουμε από τους έστω $n-s$ ($s \geq 1$) ουδέτερους κόμβους του γραφήματος τους γείτονές του, όπως δηλαδή γίνεται στη διαδικασία γραφήματος. Όμως επιπλέον έχουμε και ένα εικονικό γράφημα G' στο οποίο σε κάθε βήμα προσθέτουμε s κόμβους στο γράφημα αυθαίρετα έτσι ώστε η επιλογή των γειτόνων να μπορεί να συμπεριλάβει και κόμβους από αυτούς τους επιπλέον κόμβους με πιθανότητα p . Προφανώς η συνεκτική συνιστώσα στο εικονικό γράφημα περιέχει τους κόμβους της συνεκτικής συνιστώσας στο κανονικό γράφημα και μπορεί να έχει και κάποιους παραπάνω και άρα κυριαρχεί επί αυτής στο μέγεθος.

Επειδή αυτή η εικονική διαδικασία είναι μια διωνυμική ουρά αναμονής με παραμέτρους n και p , το ζητούμενο ισχύει.

Θεώρημα 4.2. Για κάθε k ισχύει πως

$$\mathbb{P}_\lambda(|C_i| \geq k) \geq \mathbb{P}_{n-k,p}(T_{\text{bin}} \geq k)$$

όπου $\lambda = np$ και i είναι μια οποιαδήποτε κορυφή του τυχαίου γραφήματος.

Απόδειξη. Έστω ότι υπάρχουν N_t ουδέτερες κορυφές μετά από t εξερευνήσεις. Ας συμβολίσουμε με \mathcal{T}_k τον εξής χρόνο διακοπής

$$\mathcal{T}_k = \min\{t : N_t \leq n - k\}$$

Δηλαδή ο \mathcal{T}_k αναπαριστεί την πρώτη χρονική στιγμή της διαδικασίας γραφήματος κατά την οποία το πλήθος των ουδέτερων κορυφών θα γίνει μικρότερο ή ίσο από $n-k$ ή ισοδύναμα την πρώτη χρονική στιγμή που το πλήθος των ανενεργών και ενεργών κορυφών θα γίνει τουλάχιστον k .

Άρα όταν $|C_i| \geq k$, τότε σίγουρα $\mathcal{T}_k \leq k - 1$, αφού στη χειρότερη περίπτωση, κατά την εύρεση της συνεκτικής συνιστώσας του i θα χρειαστούν $k-1$ εξερευνήσεις έως ότου και οι υπόλοιπες $k-1$ κορυφές της συνεκτικής συνιστώσας γίνουν μη-ουδέτερες.

Εφόσον για μια συνιστώσα με $|C_i| \geq k$, η αναζήτηση θα συνεχίζεται για τουλάχιστον μέχρι τη χρονική στιγμή \mathcal{T}_k , θα πρέπει η διαδικασία γραφήματος με αρχή τον κόμβο i να έχει ενεργές κορυφές για όλη την αναζήτηση μέχρι τουλάχιστον και τον \mathcal{T}_k , δηλαδή

$$\mathbb{P}_\lambda(|C_i| \geq k) = \mathbb{P}_\lambda(S_t > 0, \forall t \leq \mathcal{T}_k)$$

Τώρα ορίζουμε μια ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών X'_t που ακολουθούν τη διωνυμική κατανομή με παραμέτρους $n-k$ και p , $\text{Bin}(n-k, p)$. Για όλα τα $t \leq \mathcal{T}_k$ και δεδομένου του N_{t-1} , ορίζουμε και την $Y_t \sim \text{Bin}(N_{t-1} - (n - k), p)$.

Τότε για την $X_t = X'_t + Y_t$, ισχύει ότι $X_t \geq X'_t$, αφού η $Y_t \geq 0$ και δεδομένου του N_{t-1} , η $X_t \sim \text{Bin}(N_{t-1}, p)$ και άρα είναι όπως τη θέλουμε στη διαδικασία γραφήματος (σχέση 4.2).

Άρα θέτοντας

$$S'_t = X'_1 + X'_2 + \dots + X'_t - (t - 1)$$

θα ισχύει ότι $S'_t \leq S_t$ για κάθε $t \leq \mathcal{T}_k$. Οπότε έχουμε

$$\{S_t > 0, \forall t \leq \mathcal{T}_k\} \supset \{S'_t > 0, \forall t \leq \mathcal{T}_k\} \supset \{S'_t > 0, \forall t \leq k - 1\} = \{T_{\text{bin}} \geq k\}$$

Άρα από στοιχειώδη ιδιότητα των μέτρων πιθανοτήτων θα ισχύει για τις αντίστοιχες πιθανότητες των ενδεχομένων ότι

$$\begin{aligned} \mathbb{P}(S_t > 0, \forall t \leq \mathcal{T}_k) &\geq \mathbb{P}(T_{\text{bin}} \geq k) \iff \\ \mathbb{P}_\lambda(|C_i| \geq k) &\geq \mathbb{P}_{n-k, p}(T_{\text{bin}} \geq k) \end{aligned}$$

□

Διασθητικά μπορεί κάποιος να δει την παραπάνω απόδειξη ως εξής:

Σταματάμε τη διαδικασία γραφήματος την πρώτη χρονική στιγμή, έστω T , που ο αριθμός των ουδέτερων κορυφών θα γίνει μικρότερος ή ίσος από $n - k$. Ξανατρέχουμε πάλι τη διαδικασία γραφήματος μέχρι και τη χρονική στιγμή T , αλλά αυτή τη φορά τρέχουμε και μια διαδικασία παράλληλα σε ένα γράφημα-αντίγραφο G' δίνοντας κάθε φορά την επιλογή για γείτονες από $n-k$ από τις έως τότε ουδέτερες κορυφές. Αυτή η διαδικασία είναι ουσιαστικά μια διωνυμική ουρά αναμονής με παραμέτρους $n - k$ και p και η συνιστώσα που θα δώσει θα είναι η ίδια ή μικρότερη από αυτή που θα δώσει η διαδικασία γραφήματος.

4.3 Η πρώτη περίπτωση ($\lambda < 1$)

Σε αυτήν την ενότητα θα ασχοληθούμε με το τυχαίο γράφημα και τη συμπεριφορά της μεγιστοτικής συνεκτικής συνιστώσας του όταν για τη σταθερά λ ισχύει πως $\lambda < 1$, όπου $\lambda = np$. Το κεντρικό αποτέλεσμα που θα δείξουμε είναι πως το μέγεθος της μεγιστοτικής συνιστώσας είναι τάξης μεγέθους $\log n$ για αυτές τις τιμές του λ .

Πρώτα όμως θα αποδείξουμε ένα θεώρημα που δηλώνει πως για κατάλληλα a , η πιθανότητα το μέγεθος της μεγαλύτερης συνιστώσας να ξεπερνά το $a \log n$ μειώνεται τουλάχιστον εκθετικά καθώς οι κορυφές του τυχαίου γραφήματος γίνονται πάρα πολλές.

Θεώρημα 4.3. Έστω ένα σταθερό $\lambda < 1$. Τότε, για κάθε $\alpha > \frac{1}{1-\lambda}$, υπάρχει ένα $\delta(\alpha, \lambda) > 0$ τέτοιο ώστε

$$\mathbb{P}_\lambda(|C_{\max}| \geq \alpha \log n) = \mathcal{O}(n^{-\delta})$$

όπου το $I_\lambda = \lambda - 1 - \log \lambda$.

Απόδειξη. Από το θεώρημα 4.1 έχουμε ότι

$$\mathbb{P}_\lambda(|C_i| > t) \leq \mathbb{P}_{n,p}(T_{\text{bin}} > t)$$

Θα φράξουμε την $\mathbb{P}_\lambda(|C_i| \geq t)$ από πάνω μέσω αυτής της ανισότητας. Αφού η T_{bin} συμβολίζει το συνολικό πληθυσμό μιας διωνυμικής ουράς αναμονής, ας θυμηθούμε από τη σχέση 3.1 πως αν ο αριθμός των νέων ατόμων σε κάθε χρονική στιγμή συμβολίζεται με X_i και ακολουθεί τη διωνυμική κατανομή με παραμέτρους n και p , τότε το $S_t = X_1 + X_2 + \dots + X_t - (t-1)$ συμβολίζει τον αριθμό των ατόμων του πληθυσμού τη χρονική στιγμή t . Για να επιβιώσει ένας πληθυσμός περισσότερο από t χρονικές στιγμές θα πρέπει τη χρονική στιγμή t το $S_t > 0$ ($\{T > t\} \implies \{S_t > 0\}$). Άρα

$$\mathbb{P}_{n,p}(T > t) \leq \mathbb{P}_{n,p}(S_t > 0) = \mathbb{P}_{n,p}(X_1 + X_2 + \dots + X_t \geq t)$$

Ας εξετάσουμε την τελευταία ποσότητα με τη βοήθεια των φραγμάτων Chernoff. Για $g \geq 0$

$$\begin{aligned} \mathbb{P}(X_1 + \dots + X_t \geq t) &= \mathbb{P}(g(X_1 + \dots + X_t) \geq gt) \\ &= \mathbb{P}\left(e^{g(X_1 + \dots + X_t)} \geq e^{gt}\right) \\ &\leq e^{-tg} \cdot \mathbb{E}\left(e^{g(X_1 + \dots + X_t)}\right) \\ &= e^{-tg} \cdot \mathbb{E}(e^{gX_1}) \cdot \mathbb{E}(e^{gX_2}) \cdot \dots \cdot \mathbb{E}(e^{gX_t}) \\ &= e^{-tg} \cdot [\mathbb{E}(e^{g \cdot X})]^t \\ &= e^{-tg} \cdot [M(g)]^t \\ &= e^{-t(g - \log M(g))} \end{aligned}$$

όπου η ανισότητα προέκυψε από την ανίσωση Markov, η τρίτη ισότητα επειδή οι X_i είναι ανεξάρτητες μεταξύ τους και η τέταρτη ισότητα αφού όλες οι τυχαίες μεταβλητές είναι ισόνομες και ακολουθούν την ίδια κατανομή όπως και η X . Η $M(g)$ είναι η ροπογεννήτρια συνάρτηση μιας μεταβλητής που ακολουθεί τη διωνυμική κατανομή με παραμέτρους n και p . Από εισαγωγικά μαθήματα Πιθανοτήτων ξέρουμε πως για τη ροπογεννήτρια μιας $\text{Bin}(n, p)$ ισχύει ότι

$$M(g) = [(1-p) + pe^g]^n$$

Δηλαδή

$$e^{-t(g - \log M(g))} = e^{-t(g - n \log[(1-p) + pe^g])}$$

Από την ανισότητα $1 + x \leq e^x$, μπορούμε να πάρουμε πως

$$(1-p) + pe^g = 1 + p(e^g - 1) \leq e^{p(e^g - 1)}$$

Κατά σειρά έχουμε

$$\begin{aligned} (1-p) + pe^g &\leq e^{p(e^g - 1)} \\ \log[(1-p) + pe^g] &\leq \log[e^{p(e^g - 1)}] \\ \log[(1-p) + pe^g] &\leq p(e^g - 1) \\ g - n \log[(1-p) + pe^g] &\geq g - np(e^g - 1) \end{aligned}$$

Βρίσκουμε και το σημείο στο οποίο μεγιστοποιείται η $f(g) = g - np(e^g - 1)$:

$$f'(g) = 1 - npe^g$$

Για $f'(g) = 0$, παίρνουμε πως το σημείο στο οποίο θα έχει ακρότατο η f είναι το $g = -\log(np) = -\log(n\frac{\lambda}{n}) = -\log\lambda$. Τελικά

$$g - n\log[(1-p) + pe^g] \geq -\log\lambda - np(e^{-\log\lambda} - 1) = -\log\lambda - n\frac{\lambda}{n}(\frac{1}{\lambda} - 1) = \lambda - 1 - \log\lambda = I_\lambda$$

και άρα

$$\mathbb{P}_\lambda(|C_i| \geq k) \leq e^{-tI_\lambda} \quad (4.3)$$

Επίσης το ενδεχόμενο $\{|C_{\max}| \geq \alpha \log n\}$ είναι η ένωση των ενδεχομένων $\{|C_i| \geq \alpha \log n\}$, για $i = 1, 2, \dots, n$. Όμως λόγω της ανισότητας Boole θα ισχύει ότι

$$\begin{aligned} \mathbb{P}(|C_{\max}| \geq \alpha \log n) &= \mathbb{P}(\cup_{i=1}^n \{|C_i| \geq \alpha \log n\}) \\ &\leq \sum_{i=1}^n \mathbb{P}(|C_i| \geq \alpha \log n) \\ &= n\mathbb{P}(|C_i| \geq \alpha \log n) \\ &\leq n \cdot n^{-\alpha I_\lambda} \\ &= n^{1-\alpha I_\lambda} \\ &= n^{-\delta} \end{aligned}$$

για $\delta = \alpha I_\lambda - 1 > 0$. Η δεύτερη ισότητα από τις παραπάνω ισχύει επειδή κάθε κορυφή έχει την ίδια πιθανότητα να έχει ένα συγκεκριμένο μέγεθος συνεκτικής συνιστώσας με τις άλλες, αφού λόγω της ανεξαρτησίας των πιθανοτήτων για τις ακμές, υπάρχει συμμετρία και όλες οι κορυφές έχουν την ίδια αντιμετώπιση εκ των προτέρων. \square

Το επόμενο θεώρημα μας εξασφαλίζει επιπλέον πως το μέγεθος της μεγιστοτικής συνεκτικής συνιστώσας του τυχαίου γραφήματος με σταθερό $\lambda < 1$ δεν είναι μικρότερο του $\alpha \log n$ για κατάλληλο α σχεδόν σίγουρα.

Θεώρημα 4.4. Για ένα σταθερό $\lambda < 1$ και για κάθε $\alpha < \frac{1}{I_\lambda}$ υπάρχει ένα $\delta = \delta(\alpha, \lambda) > 0$ τέτοιο ώστε

$$\mathbb{P}_\lambda(|C_{\max}| \leq \alpha \log n) = \mathcal{O}(n^{-\delta})$$

Για την απόδειξη του θεωρήματος 4.4 θα χρειαστεί η εισαγωγή μερικών συμβολισμών και μια πρόταση.

Ο αριθμός των κορυφών που ανήκουν σε συνεκτικές συνιστώσες των οποίων το μέγεθος είναι τουλάχιστον k θα συμβολίζεται με

$$Z_{\geq k} = \sum_{v \in V} \mathbb{1}_{\{|C_v| \geq k\}}$$

Τότε το μέγεθος της μεγαλύτερης συνεκτικής συνιστώσας μπορεί να γραφτεί ως

$$|C_{\max}| = \max\{k : Z_{\geq k} \geq k\}$$

Επίσης θα ισχύει και ότι

$$\{|C_{\max}| \geq k\} = \{Z_{\geq k} \geq k\}$$

Επιπροσθέτως ο συμβολισμός

$$\chi_{\geq k}(\lambda) = \mathbb{E}_\lambda(|C_v| \mathbb{1}_{\{|C_v| \geq k\}})$$

θα χρησιμοποιηθεί, όπου το λ δηλώνει πως είμαστε σε ένα τυχαίο γράφημα με παράμετρο $p = \frac{\lambda}{n}$ και η επιλογή του κόμβου v δεν επηρεάζει το αποτέλεσμα, αφού στο τυχαίο γράφημα όλοι οι κόμβοι έχουν την ίδια θεώρηση εκ των προτέρων.

Τέλος, όταν δύο κορυφές i και j βρίσκονται στην ίδια συνεκτική συνιστώσα θα γράφουμε πως $i \leftrightarrow j$.

Πρόταση 4.5. Για κάθε $n, k \in \{0, 1, 2, \dots, n\} = [n]$ και $\lambda > 0$,

$$\text{Var}(Z_{\geq k}) \leq n\chi_{\geq k}(\lambda)$$

Απόδειξη. Αρχικά παρατηρούμε ότι η διασπορά μπορεί να γραφτεί και ως

$$\begin{aligned} \text{Var}(Z_{\geq k}) &= \mathbb{E}(Z_{\geq k}^2) - \mathbb{E}(Z_{\geq k})^2 \\ &= \mathbb{E} \left[\left(\sum_{v \in V} \mathbb{1}_{\{|C_v| \geq k\}} \right)^2 \right] - \left[\mathbb{E} \left(\sum_{v \in V} \mathbb{1}_{\{|C_v| \geq k\}} \right) \right]^2 \\ &= \mathbb{E} \left(\sum_{i, j \in [n]} \mathbb{1}_{\{|C_i| \geq k\}} \mathbb{1}_{\{|C_j| \geq k\}} \right) - \left[\sum_{v \in V} \mathbb{E}(\mathbb{1}_{\{|C_v| \geq k\}}) \right]^2 \\ &= \mathbb{E} \left(\sum_{i, j \in [n]} \mathbb{1}_{\{|C_i| \geq k, |C_j| \geq k\}} \right) - \left[\sum_{v \in V} \mathbb{E}(\mathbb{1}_{\{|C_v| \geq k\}}) \right]^2 \\ &= \sum_{i, j \in [n]} \mathbb{E}(\mathbb{1}_{\{|C_i| \geq k, |C_j| \geq k\}}) - \left[\sum_{v \in V} \mathbb{E}(\mathbb{1}_{\{|C_v| \geq k\}}) \right]^2 \\ &= \sum_{i, j \in [n]} \mathbb{P}(|C_i| \geq k, |C_j| \geq k) - \left[\sum_{i \in [n]} \mathbb{P}(|C_i| \geq k) \right]^2 \end{aligned} \quad (4.4)$$

Επειδή δύο κορυφές είτε θα βρίσκονται στην ίδια συνιστώσα είτε όχι, μπορούμε να σπάσουμε την πιθανότητα $\mathbb{P}(|C_i| \geq k, |C_j| \geq k)$ σε

$$\mathbb{P}(|C_i| \geq k, |C_j| \geq k) = \mathbb{P}(|C_i| \geq k, i \leftrightarrow j) + \mathbb{P}(|C_i| \geq k, |C_j| \geq k, i \nleftrightarrow j) \quad (4.5)$$

Επίσης λόγω της βασικής ιδιότητας των δεσμευμένων πιθανοτήτων μπορούμε να γράψουμε

$$\mathbb{P}(|C_i| = l, |C_j| \geq k, i \nleftrightarrow j) = \mathbb{P}(|C_i| = l, i \nleftrightarrow j) \mathbb{P}(|C_j| \geq k | |C_i| = l, i \nleftrightarrow j) \quad (4.6)$$

Τώρα μπορούμε να σκεφτούμε το εξής: όταν το $|C_i| = l$ και $i \nleftrightarrow j$ όλες οι κορυφές που δεν ανήκουν στη συνεκτική συνιστώσα του i σχηματίζουν ένα τυχαίο γράφημα με παράμετρο $p = \frac{\lambda}{n}$ σε $n-l$ το πλήθος κορυφές (τις εναπομείνουσες). Εφόσον όμως η πιθανότητα η $|C_j| \geq k$ αυξάνεται (βλ. ενότητα 4.8) καθώς συμπεριλαμβάνονται περισσότερες κορυφές στο τυχαίο γράφημα, θα ισχύει ότι η πιθανότητα να ισχύει ότι $|C_j| \geq k$ στο $G(n-l, p)$ θα είναι μικρότερη από την πιθανότητα να ισχύει ότι $|C_j| \geq k$ στο $G(n, p)$. Δηλαδή

$$\mathbb{P}(|C_j| \geq k | |C_i| = l, i \nleftrightarrow j) \leq \mathbb{P}(|C_j| \geq k)$$

Επίσης, η πιθανότητα το $|C_i| = l$ και $i \nleftrightarrow j$ θα είναι μικρότερη από την πιθανότητα απλά το $|C_i| = l$, γιατί στην πρώτη περίπτωση ένας συγκεκριμένος αριθμός κορυφών (n συνεκτική συνιστώσα του j) δε θα είναι ανάμεσα στους υπονήφιους για να βρίσκεται στη συνεκτική συνιστώσα του i , δηλαδή το $|C_i|$ θα βρίσκεται στο $G(n - |C_j|, p)$, ενώ στη δεύτερη η επιλογή γίνεται στο $G(n, p)$.

Οπότε για κάθε $l \geq 1$, η σχέση 4.6 γράφεται και ως

$$\mathbb{P}(|C_i| = l, |C_j| \geq k, i \nleftrightarrow j) \leq \mathbb{P}(|C_i| = l) \mathbb{P}(|C_j| \geq k) \quad (4.7)$$

Οπότε μπορούμε να γράψουμε

$$\begin{aligned} \mathbb{P}(|C_i| \geq k, |C_j| \geq k, i \nleftrightarrow j) &= \sum_{l=k}^n \mathbb{P}(|C_i| = l, |C_j| \geq k, i \nleftrightarrow j) \\ &\leq \sum_{l=k}^n \mathbb{P}(|C_i| = l) \mathbb{P}(|C_j| \geq k) && \text{εξαιτίας της 4.7} \\ &= \mathbb{P}(|C_i| \geq k) \mathbb{P}(|C_j| \geq k) \end{aligned}$$

Μαζεύοντας τα μέχρι τώρα αποτελέσματα έχουμε

$$\begin{aligned}
\text{Var}(Z_{\geq k}) &= \sum_{i,j \in [n]} \mathbb{P}(|C_i| \geq k, |C_j| \geq k) - \left[\sum_{i \in [n]} \mathbb{P}(|C_i| \geq k) \right]^2 && \text{από την 5.2} \\
&= \sum_{i,j \in [n]} [\mathbb{P}(|C_i| \geq k, i \leftrightarrow j) + \mathbb{P}(|C_i| \geq k, |C_j| \geq k, i \not\leftrightarrow j)] - \left[\sum_{i \in [n]} \mathbb{P}(|C_i| \geq k) \right]^2 && \text{από την 4.5} \\
&= \sum_{i,j \in [n]} [\mathbb{P}(|C_i| \geq k, i \leftrightarrow j)] + \sum_{i,j \in [n]} [\mathbb{P}(|C_i| \geq k, |C_j| \geq k, i \not\leftrightarrow j)] - \left[\sum_{i \in [n]} \mathbb{P}(|C_i| \geq k) \right]^2 \\
&\leq \sum_{i,j \in [n]} [\mathbb{P}(|C_i| \geq k, i \leftrightarrow j)] + \sum_{i,j \in [n]} [\mathbb{P}(|C_i| \geq k) \mathbb{P}(|C_j| \geq k)] - \left[\sum_{i \in [n]} \mathbb{P}(|C_i| \geq k) \right]^2 && \text{από την 4.7} \\
&= \sum_{i,j \in [n]} [\mathbb{P}(|C_i| \geq k, i \leftrightarrow j)] && \text{από απλοποίηση} \\
&= \sum_{i=1}^n \sum_{j=1}^n [\mathbb{E}(\mathbb{1}_{\{|C_i| \geq k, i \leftrightarrow j\}})] \\
&= \sum_{i=1}^n \sum_{j=1}^n [\mathbb{E}(\mathbb{1}_{\{|C_i| \geq k\}} \mathbb{1}_{\{i \leftrightarrow j\}})] \\
&= \sum_{i=1}^n \sum_{j=1}^n [\mathbb{E}(\mathbb{1}_{\{|C_i| \geq k\}} \mathbb{1}_{\{j \in C_i\}})] \\
&= \sum_{i=1}^n [\mathbb{E}(\mathbb{1}_{\{|C_i| \geq k\}} \sum_{j=1}^n \mathbb{1}_{\{j \in C_i\}})]
\end{aligned} \tag{4.8}$$

Όμως το $\sum_{j=1}^n \mathbb{1}_{\{j \in C_i\}} = |C_i|$ και τελικά

$$\begin{aligned}
\text{Var}(Z_{\geq k}) &\leq \sum_{i=1}^n [\mathbb{E}(\mathbb{1}_{\{|C_i| \geq k\}} |C_i|)] \\
&= n \mathbb{E}(\mathbb{1}_{\{|C_1| \geq k\}} |C_1|) \\
&= n \chi_{\geq k}(\lambda)
\end{aligned}$$

όπου δεν έχει σημασία ποιον κόμβο (εδώ τον 1) θα βάλουμε αφού όλοι οι κόμβοι δίνουν το ίδιο αποτέλεσμα. \square

Τώρα είμαστε έτοιμοι να αποδείξουμε το ζητούμενο θεώρημα.

Απόδειξη. Θεωρήματος 4.4. Αρχικά παρατηρούμε πως εάν η μεγαλύτερη συνεκτική συνιστώσα περιέχει λιγότερες από k κορυφές, τότε ο αριθμός των κορυφών που ανήκουν σε συνεκτικές συνιστώσες μεγέθους τουλάχιστον k είναι 0:

$$\{|C_{\max}| < k\} \implies \{Z_{\geq k} = 0\}$$

Αλλά και αντίστροφα, εάν ο αριθμός των κορυφών που ανήκουν σε συνεκτικές συνιστώσες μεγέθους τουλάχιστον k είναι 0, τότε η μεγαλύτερη συνιστώσα θα περιέχει το πολύ $k-1$ κορυφές:

$$\{Z_{\geq k} = 0\} \implies \{|C_{\max}| < k\}$$

Επομένως αυτά τα δύο ενδεχόμενα είναι ισοδύναμα και για να αποδείξουμε το ζητούμενο θεώρημα μπορούμε να το δείξουμε για την $\mathbb{P}_\lambda(Z_{\geq k_n} = 0)$ με το $k_n = \alpha \log n$ και $\alpha < \frac{1}{1\lambda}$. Θα χρησιμοποιήσουμε την ανισότητα του Chebyshev και άρα θα χρειαστούμε ένα άνω όριο για τη διασπορά $\text{Var}(Z_{\geq k_n})$ και ένα κάτω όριο για τη μέση τιμή $\mathbb{E}(Z_{\geq k_n})$. Από εδώ και πέρα θα παραλείπεται ο δείκτης λ για τη δήλωση του γεγονότος πως βρισκόμαστε στο τυχαίο γράφημα με $p = \frac{\lambda}{n}$.

Αρχικά βλέπουμε ότι

$$\mathbb{E}(Z_{\geq k_n}) = \mathbb{E} \left(\sum_{v \in V} \mathbb{1}_{\{|C_v| \geq k_n\}} \right) = \sum_{v \in V} \mathbb{E}(\mathbb{1}_{\{|C_v| \geq k_n\}}) = n \mathbb{P}(|C_v| \geq k_n)$$

όπου δεν έχει σημασία για ποια κορυφή v μιλάμε λόγω της συμμετρίας στο τυχαίο γράφημα.

Από το θεώρημα 4.2 έχουμε πως

$$\mathbb{P}(|C_v| \geq k_n) \geq \mathbb{P}_{n-k_n, p}(T_{\text{bin}} \geq k_n) \quad (4.9)$$

Το θεώρημα 3.7 δηλώνει πως για μια διωνυμική με παραμέτρους $(n - k_n, p)$ και μια Poisson με παράμετρο $\lambda_n = (n - k_n)p = (n - k_n)\frac{\lambda}{n}$

$$\mathbb{P}_{\text{bin}}(T_{\text{bin}} \geq k) = \mathbb{P}_{\text{po}}(T_{\text{po}} \geq k) + e_n(k)$$

ή ισοδύναμα

$$\mathbb{P}_{n-k_n, p}(T_{\text{bin}} \geq k_n) = \mathbb{P}_{\text{po}(\lambda_n)}(T_{\text{po}} \geq k_n) + \mathcal{O}\left(\frac{\alpha\lambda^2 \log n}{n}\right)$$

Επίσης το θεώρημα 3.5 μας δίνει ότι

$$\mathbb{P}_{\text{po}(\lambda)}(T_{\text{po}} \geq \alpha \log n) = \sum_{k=\alpha \log n}^{\infty} \mathbb{P}_{\text{po}(\lambda)}(T_{\text{po}} = k) = \sum_{k=\alpha \log n}^{\infty} \frac{(\lambda n k)^{k-1}}{k!} e^{-\lambda n k} \quad (4.10)$$

Ο τύπος του Stirling μας δίνει ότι

$$k! = \left(\frac{k}{e}\right)^k \sqrt{2\pi k}(1 + o(1)) \quad (4.11)$$

Επειδή όμως $I_\lambda = \lambda - 1 - \log \lambda$ θα ισχύει και ότι

$$\begin{aligned} I_{\lambda_n} &= \lambda_n - 1 - \log \lambda_n \\ &= -1 + \frac{\lambda(n - k_n)}{n} - \log \left(\frac{\lambda(n - k_n)}{n}\right) \\ &= -1 + \lambda - \frac{\lambda k_n}{n} - \log \lambda - \log \left(\frac{n - k_n}{n}\right) \\ &= -1 + \lambda - \log \lambda - \lambda \frac{\alpha \log n}{n} - \log \left(1 - \frac{\alpha \log n}{n}\right) \\ &= I_\lambda + o(1) \end{aligned} \quad (4.12)$$

αφού το n υπερिशύει έναντι του $\log n$ καθώς το n τείνει στο άπειρο.

Οπότε θα ισχύει ότι

$$\begin{aligned}
\mathbb{P}_{\text{po}(\lambda)}(T_{\text{po}} \geq \alpha \log n) &= \sum_{k=\alpha \log n}^{\infty} \frac{(\lambda_n k)^{k-1}}{k!} e^{-\lambda_n k} \\
&= \sum_{k=\alpha \log n}^{\infty} \frac{(\lambda_n k)^{k-1}}{\left(\frac{k}{e}\right)^k \sqrt{2\pi k} (1+o(1))} e^{-\lambda_n k} && \text{λόγω της 4.10} \\
&= \sum_{k=\alpha \log n}^{\infty} \frac{(\lambda_n)^k}{\lambda_n \sqrt{2\pi k^3} (1+o(1))} e^{k(1-\lambda_n)} \\
&= \sum_{k=\alpha \log n}^{\infty} \frac{e^{k(1-\lambda_n + \log \lambda_n)}}{\lambda_n \sqrt{2\pi k^3}} (1+o(1)) \\
&= \sum_{k=\alpha \log n}^{\infty} \frac{\frac{n}{n-k_n} e^{-k I_{\lambda_n}}}{\lambda \sqrt{2\pi k^3}} (1+o(1)) \\
&\geq \sum_{k=\alpha \log n}^{\infty} \frac{e^{-k I_{\lambda_n}}}{\lambda \sqrt{2\pi k^3}} (1+o(1)) \\
&\geq \sum_{k=\alpha \log n}^{\infty} \frac{e^{-k I_{\lambda_n}}}{\sqrt{2\pi k^3}} (1+o(1)) && \text{αφού το } \lambda < 1 \\
&\geq e^{-\alpha \log n I_{\lambda} (1+o(1))} && \text{για μεγάλο } n
\end{aligned}$$

Άρα εν τέλει

$$\begin{aligned}
\mathbb{E}(Z_{\geq k_n}) &= n\mathbb{P}(|C_v| \geq k_n) \\
&\geq n\mathbb{P}_{n-k_n, p}(T_{\text{bin}} \geq k_n) \\
&\geq n \left(\mathbb{P}_{\text{po}(\lambda)}(T_{\text{po}} \geq k_n) + \mathcal{O}\left(\frac{\alpha \lambda^2 \log n}{n}\right) \right) \\
&\geq n e^{-\alpha \log n I_{\lambda} (1+o(1))} + n \mathcal{O}\left(\frac{\alpha \lambda^2 \log n}{n}\right) \\
&= n^{(1-I_{\lambda} \alpha)(1+o(1))} + \mathcal{O}(\alpha \lambda^2 \log n) \\
&\geq n^d && (4.13)
\end{aligned}$$

για κατάλληλα μεγάλο n , και επιλέγοντας το d έτσι ώστε $0 < \frac{1-I_{\lambda} \alpha}{2} < d = d(\alpha, \lambda) < 1 - I_{\lambda} \alpha$.

Τώρα θα επικεντρωθούμε στο άνω φράγμα για τη διασπορά. Για κάθε τυχαία μεταβλητή X που παίρνει τιμές στους ακεραίους άμα μετρήσουμε την ακόλουθη αναμενόμενη τιμή με διαφορετικούς τρόπους παίρνουμε πως

$$\begin{aligned}
\mathbb{E}(X \mathbb{1}_{\{X \geq k\}}) &= \sum_{t=k}^{\infty} t \mathbb{P}(X = t) \\
&= \sum_{t=k}^{\infty} \sum_{s=1}^t \mathbb{P}(X = t) \\
&= \sum_{s=1}^{\infty} \sum_{t=k \vee s}^{\infty} \mathbb{P}(X = t) \\
&= k \mathbb{P}(X \geq k) + \sum_{s=k+1}^{\infty} \mathbb{P}(X \geq s)
\end{aligned}$$

Στην περίπτωσή μας

$$\begin{aligned}
\chi_{\geq k}(\lambda) &= \mathbb{E}_\lambda(|C_{\max}| \mathbb{1}_{\{|C_v| \geq k\}}) \\
&= k_n \mathbb{P}(|C_v| \geq k_n) + \sum_{s=k_n+1}^n \mathbb{P}(|C_v| \geq s) \\
&\leq k_n e^{-(k_n-1)I_\lambda} + \sum_{s=k_n+1}^n e^{-I_\lambda(s-1)} && \text{από την 4.3} \\
&\leq k_n e^{-(k_n-1)I_\lambda} + \frac{e^{-I_\lambda k_n}}{1 - e^{-I_\lambda}} && \text{μερικό άθροισμα γεωμετρικής προόδου} \\
&= O(k_n n^{-\alpha I_\lambda}) && (4.14)
\end{aligned}$$

Από την πρόταση 4.5 και τη σχέση 4.14 για $k_n = \alpha \log n$ θα έχουμε τότε πως:

$$\text{Var}(Z_{\geq k_n}) \leq n \chi_{\geq k_n}(\lambda) = O(k_n n^{1-\alpha I_\lambda}) \quad (4.15)$$

Από την ανισότητα Chebyshev και τις σχέσεις 4.13 και 4.15 παίρνουμε

$$\mathbb{P}(Z_{\geq k_n} = 0) \leq \frac{\text{Var}(Z_{\geq k_n})}{[\mathbb{E}(Z_{\geq k_n})]^2} = O(k_n n^{1-\alpha I_\lambda - 2d}) = O(n^{-\delta}) \quad (4.16)$$

για $\delta < 2d - 1 + I_\lambda \alpha$ και n κατάλληλα μεγάλο.

□

Πλέον έχουμε στη διάθεσή μας τα θεωρήματα 4.3 και 4.4 σύμφωνα με τα οποία υπάρχουν $\alpha_2 < \frac{1}{I_\lambda} < \alpha_1$ τέτοια ώστε να ισχύει πως

$$\mathbb{P}(|C_{\max}| \geq \alpha_1 \log n) = O(n^{-\delta_1}) \quad (4.17)$$

και

$$\mathbb{P}(|C_{\max}| \leq \alpha_2 \log n) = O(n^{-\delta_2}) \quad (4.18)$$

για $\delta_1, \delta_2 > 0$.

Αυτό σημαίνει πως n πιθανότητα

$$\mathbb{P}(\alpha_2 \log n \leq |C_{\max}| \leq \alpha_1 \log n) = 1 - \mathbb{P}(|C_{\max}| \geq \alpha_1 \log n) - \mathbb{P}(|C_{\max}| \leq \alpha_2 \log n) \quad (4.19)$$

$$= 1 - O(n^{-\delta_1}) - O(n^{-\delta_2}) \quad (4.20)$$

$$\geq 1 - O(n^{-\delta}) \quad (4.21)$$

με το $\delta = \min\{\delta_1, \delta_2\}$. Δηλαδή το μέγεθος της μεγαλύτερης συνεκτικής συνιστώσας στην περίπτωση που το λ είναι σταθερό και μικρότερο του 1 είναι τάξης μεγέθους $\log n$. Επιπλέον εάν βάλουμε τα α_1, α_2 πολύ κοντά στο $\frac{1}{I_\lambda}$ παίρνουμε πως ασυμπτωτικά το μέγεθος της μεγαλύτερης συνιστώσας θα προσεγγίζει το $\frac{1}{I_\lambda} \log n$.

4.4 Η δεύτερη περίπτωση ($\lambda > 1$)

Όταν σε αντίθεση με την προηγούμενη ενότητα το λ είναι σταθερό μεν, μεγαλύτερο του 1 δε, παρατηρείται μια απότομη διαφορά στη συμπεριφορά της μεγαλύτερης συνεκτικής συνιστώσας του τυχαίου γραφήματος σε n κορυφές και με πιθανότητα $p = \frac{\lambda}{n}$ εμφάνισης κάθε δυνατής ακμής στο γράφημα. Αυτή η διαφορά έγκειται στο γεγονός πως το μέγεθος της μεγαλύτερης συνιστώσας παύει να είναι τάξης $\log n$ και απεναντίας είναι τάξης n , δηλαδή cn για μια σταθερά $c < 1$. Αυτή η τεράστια εναλλαγή της τάξης του μεγέθους της μεγαλύτερης συνιστώσας είναι που κάνει τόσο ενδιαφέρουσα τη μελέτη του τυχαίου γραφήματος, παρόλο που το ίδιο δε μπορεί να χρησιμοποιηθεί ως μοντέλο για την πρόβλεψη κάποιας ποσότητας, αφού στο φυσικό κόσμο σπάνια κάποιο δίκτυο μπορεί να προσομοιωθεί από ένα μοντέλο που δίνει την

ίδια πιθανότητα εμφάνισης σε όλες τις ακμές. Παρόλα αυτά, και σε άλλα πιο ρεαλιστικά μοντέλα παρατηρείται αρκετά συχνά κάποιου είδους αλλαγή στη συμπεριφορά των μεγαλύτερων συνιστώσων, γεγονός που καθιστά απαραίτητη τη μελέτη της μεταβολής φάσης (phase transition) που είναι ο όρος που έχει δοθεί για την περιγραφή αυτού του φαινομένου.

Σε αυτήν την ενότητα θα αποδείξουμε το παραπάνω αποτέλεσμα με τον εξής τρόπο:

Αρχικά θα δείξουμε πως ο αριθμός των κορυφών που ανήκουν σε συνιστώσες μεγέθους μεγαλύτερου από $k_n = K \log n$ δεν απέχει πολύ από το $\zeta_\lambda n$ (πρόγραμμα 4.8), όπου το ζ_λ είναι η πιθανότητα επιβίωσης σε μια ουρά αναμονής Poisson. Για να το πετύχουμε αυτό θα χρειαστούμε τις προτάσεις 4.6 και 4.7. Η 4.6 δίνει την πιθανότητα μια συνεκτική συνιστώσα να έχει μέγεθος μεγαλύτερο από $k_n = K \log n$. Ήδη από εδώ βλέπουμε τη διαφορά σε σχέση με την περίπτωση όπου το $\lambda < 1$, αφού εδώ αυτή η πιθανότητα δεν είναι εκθετικά μικρή. Η πρόταση 4.7 δίνει μια καλύτερη ανισότητα για την ποσότητα $Z_{\geq k}$ σε σχέση με την πρόταση 4.5 και έτσι έχουμε το επιθυμητό αποτέλεσμα με την εφαρμογή της ανισότητας Chebyshev στη συνέχεια.

Στη συνέχεια οι προτάσεις 4.9 και 4.10 είναι αρκετές για να δείξουμε ότι στην ουσία δεν υπάρχουν συνεκτικές συνιστώσες με μέγεθος ανάμεσα σε $K \log n$ και αn με το $\alpha < \zeta_\lambda$. Σε συνδυασμό με το γεγονός πως ο αριθμός των κορυφών που ανήκουν σε συνιστώσες μεγέθους μεγαλύτερου από $k_n = K \log n$ δεν απέχει πολύ από το $\zeta_\lambda n$ δεν υπάρχει άλλη περίπτωση πέρα από αυτές οι συνιστώσες να έχουν και μέγεθος μεγαλύτερο από αn . Επιπλέον όπως γίνεται φανερό και στο λήμμα 4.13 είναι αδύνατο να υπάρχουν περισσότερες από μια τέτοιες συνεκτικές συνιστώσες. Αυτή η παρατήρηση μας οδηγεί στην απόδειξη ύπαρξης μια συνεκτικής συνιστώσας η οποία έχει μέγεθος $\Theta(n)$. Σε αυτή τη συνιστώσα έχει δοθεί ο τίτλος “γιγαντιαία” συνιστώσα, χαρακτηρισμός ο οποίος χρησιμοποιείται για τη μεγαλύτερη συνεκτική συνιστώσα μονάχα όταν είμαστε στην περίπτωση $\lambda > 1$ και το λ είναι μια σταθερά.

Πρόταση 4.6. Έστω ένα σταθερό $\lambda > 1$. Τότε, για $k_n \geq \alpha \log n$ με το $\alpha > \frac{1}{I_\lambda}$ και για αρκετά μεγάλο n θα ισχύει για κάθε κορυφή του τυχαίου γραφήματος ότι

$$\mathbb{P}(|C_v| \geq k_n) = \zeta_\lambda + \mathcal{O}\left(\frac{k_n}{n}\right) \quad (4.22)$$

Απόδειξη. Έχουμε ότι

$$\mathbb{P}(|C_v| \geq k_n) \leq \mathbb{P}_{n, \frac{\lambda}{n}}(T_{\text{bin}} \geq k_n) \leq \mathbb{P}(T_{\text{po}} \geq k_n) + \mathcal{O}\left(\frac{k_n}{n}\right) \quad (4.23)$$

όπου οι δύο ανισότητες εξηγούνται από τα θεωρήματα 4.1 και 3.7.

Επίσης, εξαιτίας του θεωρήματος 3.9 έχουμε και ότι

$$\mathbb{P}(t_{\text{po}} \geq k_n) = \mathbb{P}(T_{\text{po}} = \infty) + \mathbb{P}(k_n \leq T_{\text{po}} < \infty) = \zeta_\lambda + \mathcal{O}(e^{-k_n I_\lambda}) = \zeta_\lambda + \mathcal{O}(n^{-\alpha I_\lambda}) = \zeta_\lambda + o\left(\frac{1}{n}\right)$$

αφού $\alpha > \frac{1}{I_\lambda}$ και άρα η 4.23 γίνεται

$$\mathbb{P}(|C_v| \geq k_n) \leq \zeta_\lambda + \mathcal{O}\left(\frac{k_n}{n}\right) \quad (4.24)$$

Εάν τώρα χρησιμοποιήσουμε τα θεωρήματα 4.2 και 3.7 και θέσουμε $\lambda_n = \lambda(1 - \frac{k_n}{n})$ για την αντίστοιχη διαδικασία Poisson του 3.7 παίρνουμε

$$\mathbb{P}(|C_v| \geq k_n) \geq \mathbb{P}_{n-k_n, \frac{\lambda}{n}}(T_{\text{bin}} \geq k_n) \geq \mathbb{P}(T_{\text{po}} \geq k_n) + \mathcal{O}\left(\frac{k_n}{n}\right) \quad (4.25)$$

Πάλι από το θεώρημα 3.9 και για $k_n \geq \alpha \log n$, $\alpha > \frac{1}{I_\lambda}$ και από την 4.12 παίρνουμε ότι

$$\mathbb{P}(T_{\text{po}} \geq k_n) = \zeta_{\lambda_n} + \mathcal{O}(e^{-k_n I_{\lambda_n}}) = \zeta_{\lambda_n} + o\left(\frac{1}{n}\right) \quad (4.26)$$

Επιπλέον

$$\eta_{\lambda_n} = \eta_\lambda + (\lambda_n - \lambda) \frac{d}{d\lambda} \eta_\lambda|_{\lambda=\lambda_n^*} = \eta_\lambda + \mathcal{O}\left(\frac{k_n}{n}\right)$$

εξαιτίας του Θεωρήματος Μέσης Τιμής για $\lambda_n^* \in (\lambda_n, \lambda)$, της πρότασης 3.2 και επειδή $\lambda_n - \lambda = \frac{k_n}{n}$. Επίσης, αφού $\zeta_\lambda = 1 - \eta_\lambda$ ισχύει και ότι $\zeta_{\lambda_n} = \zeta_\lambda + \mathcal{O}\left(\frac{k_n}{n}\right)$. Οπότε η σχέση 4.25 γίνεται

$$\mathbb{P}(|C_v| \geq k_n) \geq \zeta_\lambda + \mathcal{O}\left(\frac{k_n}{n}\right) \quad (4.27)$$

Οι ανισότητες 4.24 και 4.27 μαζί αποδεικνύουν τη ζητούμενη πρόταση, δηλαδή ότι

$$\mathbb{P}(|C_v| \geq k_n) = \zeta_\lambda + \mathcal{O}\left(\frac{k_n}{n}\right) \quad (4.28)$$

□

Ακόμα, θα χρειαστούμε την παρακάτω πρόταση. Υπενθυμίζεται ότι με $Z_{\geq k}$ συμβολίζεται ο αριθμός των κορυφών που βρίσκονται σε συνεκτική συνιστώσα μεγέθους τουλάχιστον k , δηλαδή $Z_{\geq k} = \sum_{v \in V} \mathbb{1}_{\{|C_v| \geq k\}}$. Επίσης η αναμενόμενη τιμή του μεγέθους των συνιστώσων που έχουν μέγεθος μικρότερο από k συμβολίζεται με $\chi_{<k}(\lambda)$, δηλαδή $\chi_{<k}(\lambda) = \mathbb{E}(|C_v| \mathbb{1}_{\{|C_v| < k\}})$.

Πρόταση 4.7. Για κάθε n και $k \leq n$ ισχύει ότι

$$\text{Var}(Z_{\geq k}) \leq (\lambda k + 1) n \chi_{<k}(\lambda) \quad (4.29)$$

Απόδειξη. Εάν ορίσουμε $Z_{<k} = \sum_{v \in V} \mathbb{1}_{\{|C_v| < k\}}$, τότε ισχύει και ότι $Z_{<k} = n - Z_{\geq k}$. Επειδή το n είναι κάποιο συγκεκριμένο και από βασικές ιδιότητες της διασποράς θα πάρουμε πως

$$\text{Var}(Z_{<k}) = \text{Var}(Z_{\geq k})$$

Οπότε μπορούμε να δείξουμε την πρόταση για τη $Z_{<k}$. Αυτό θα γίνει με ανάλογα επιχειρήματα όπως στην απόδειξη της πρότασης 4.5. Αρχικά γράφουμε τον τύπο της διασποράς (παραλείποντας τις πράξεις που έχουν ήδη γίνει στην προαναφερθείσα πρόταση).

$$\text{Var}(Z_{<k}) = \sum_{i,j \in [n]} \mathbb{P}(|C_i| < k, |C_j| < k) - \left[\sum_{i \in [n]} \mathbb{P}(|C_i| < k) \right]^2 \quad (4.30)$$

Στη συνέχεια "σπάμε" την πρώτη πιθανότητα ανάλογα με το εάν οι κορυφές i και j βρίσκονται στην ίδια συνεκτική συνιστώσα ή όχι.

$$\text{Var}(Z_{<k}) = \sum_{i,j \in [n]} \mathbb{P}(|C_i| < k, |C_j| < k, i \leftrightarrow j) + \sum_{i,j \in [n]} \mathbb{P}(|C_i| < k, |C_j| < k, i \not\leftrightarrow j) - \left[\sum_{i \in [n]} \mathbb{P}(|C_i| < k) \right]^2 \quad (4.31)$$

Θα ασχοληθούμε με τα αυτά τα σπασίματα ξεχωριστά. Πρώτα έχουμε πως

$$\begin{aligned} \sum_{i,j \in [n]} \mathbb{P}(|C_i| < k, |C_j| < k, i \leftrightarrow j) &= \sum_{i,j \in [n]} \mathbb{E}[\mathbb{1}_{\{|C_i| < k\}} \mathbb{1}_{\{i \leftrightarrow j\}}] \\ &= \sum_{i \in [n]} \mathbb{E}[\mathbb{1}_{\{|C_i| < k\}} \sum_{j \in [n]} \mathbb{1}_{\{i \leftrightarrow j\}}] \\ &= \sum_{i \in [n]} \mathbb{E}[|C_i| \mathbb{1}_{\{|C_i| < k\}}] \\ &= n \chi_{<k}(\lambda) \end{aligned} \quad (4.32)$$

Τώρα θα περιεργαστούμε το δεύτερο κομμάτι, χρησιμοποιώντας αρχικά τον ορισμό της δεσμευμένης πιθανότητας δύο φορές κατά σειρά ώστε να πάρουμε

$$\mathbb{P}(|C_i| = l, |C_j| < k, i \leftrightarrow j) = \mathbb{P}(|C_i| = l)\mathbb{P}(i \leftrightarrow j | |C_i| = l)\mathbb{P}(|C_j| < k | |C_i| = l, i \leftrightarrow j) \quad (4.33)$$

για κάθε $l < k$. Επειδή κάθε πιθανότητα είναι μικρότερη ή ίση του 1, μπορούμε να γράψουμε

$$\mathbb{P}(|C_i| = l, |C_j| < k, i \leftrightarrow j) \leq \mathbb{P}(|C_i| = l)\mathbb{P}(|C_j| < k | |C_i| = l, i \leftrightarrow j) \quad (4.34)$$

Δεδομένου πως το μέγεθος της συνεκτικής συνιστώσας της κορυφής i είναι l και του ότι οι κορυφές i και j βρίσκονται σε διαφορετικές συνιστώσες, η πιθανότητα η συνεκτική συνιστώσα της j να έχει οποιοδήποτε μέγεθος (μέχρι και $n - l$) είναι ίση με την πιθανότητα που θα είχε εάν βρισκόταν σε ένα τυχαίο γράφημα $G(n - l, p)$ και άρα θα έχουμε πως $\mathbb{P}(|C_j| < k | |C_i| = l, i \leftrightarrow j) = \mathbb{P}^*(|C_j| < k)$ όπου με \mathbb{P}^* είναι η πιθανότητα στο $G(n - l, p)$.

Ή αλλιώς

$$\begin{aligned} \mathbb{P}(|C_j| < k | |C_i| = l, i \leftrightarrow j) &= \mathbb{P}^*(|C_j| < k) \\ &= \mathbb{P}(|C_j| < k) + \mathbb{P}^*(|C_j| < k) - \mathbb{P}(|C_j| < k) \end{aligned} \quad (4.35)$$

Σε αυτό το σημείο είναι χρήσιμο να κάνουμε την ακόλουθη παρατήρηση. Έχουμε το $G(n - l, p)$ και συμπληρώνουμε τις υπόλοιπες l κορυφές ώστε να δημιουργηθεί το $G(n, p)$ δίνοντας ανεξάρτητα την πιθανότητα p σε κάθε δυνατή ακμή ανάμεσα στις n κορυφές και σε αυτές που προσθήσαμε να εμφανίζεται. Τότε η $\mathbb{P}^*(|C_j| < k) - \mathbb{P}(|C_j| < k)$ δεν είναι παρά η πιθανότητα η κορυφή j να έχει συνεκτική συνιστώσα μικρότερη από k στο $G(n - l, p)$ και μεγαλύτερη ή ίση από k στο $G(n, p)$. Για να συμβαίνει αυτό θα πρέπει τουλάχιστον μια από τις l κορυφές που προσθήσαμε να συνδέεται με τουλάχιστον μια από τις κορυφές που ανήκουν στη συνεκτική συνιστώσα της j στο $G(n - l, p)$. Εφόσον οι κορυφές που ανήκουν στη συνεκτική συνιστώσα της j στο $G(n - l, p)$ είναι λιγότερες από k και οι καινούριες κορυφές του γραφήματος είναι l το πλήθος, αυτή η πιθανότητα είναι μικρότερη από lkp . Δηλαδή η 4.35 γράφεται και ως

$$\mathbb{P}(|C_j| < k | |C_i| = l, i \leftrightarrow j) \leq \mathbb{P}(|C_j| < k) + lkp \quad (4.36)$$

Δηλαδή παίρνοντας το κομμάτι της σχέσης 4.30 για το οποίο δεν έχουμε μιλήσει και συνειδητοποιώντας πως θα πρέπει να καλύψουμε όλο το δυνατό εύρος των μεγεθών l από το 1 έως και το $k-1$, θα έχουμε ότι

$$\begin{aligned} &\sum_{i,j \in [n]} \mathbb{P}(|C_i| < k, |C_j| < k, i \leftrightarrow j) - \left[\sum_{i \in [n]} \mathbb{P}(|C_i| < k) \right]^2 \\ &\leq \sum_{l=1}^{k-1} \sum_{i,j \in [n]} \mathbb{P}(|C_i| = l) (\mathbb{P}(|C_j| < k) + lkp) - \left[\sum_{i \in [n]} \mathbb{P}(|C_i| < k) \right]^2 && \text{από τις 4.34, 4.36} \\ &= \sum_{i,j \in [n]} \mathbb{P}(|C_i| < k) \mathbb{P}(|C_j| < k) + \sum_{l=1}^{k-1} \sum_{i,j \in [n]} \mathbb{P}(|C_i| = l) lkp - \left[\sum_{i \in [n]} \mathbb{P}(|C_i| < k) \right]^2 \\ &= \frac{\lambda k}{n} \sum_{i,j \in [n]} \sum_{l=1}^{k-1} l \mathbb{P}(|C_i| = l) \\ &= \frac{\lambda k}{n} \sum_{i,j \in [n]} \mathbb{E}(|C_i| \mathbb{1}_{\{|C_i| < k\}}) \\ &= \frac{\lambda k}{n} \sum_{i,j \in [n]} \chi_{<k}(\lambda) \\ &= nk \lambda \chi_{<k}(\lambda) \end{aligned} \quad (4.37)$$

Συνδυάζοντας τις σχέσεις 4.32 και 4.37 αποδεικνύεται πως

$$\text{Var}(Z_{\geq k}) \leq (\lambda k + 1) n \chi_{<k}(\lambda)$$

□

Οι δύο προτάσεις 4.6 και 4.7 μας δίνουν ένα ενδιαφέρον αποτέλεσμα: η διαφορά μεταξύ του αριθμού των κορυφών που ανήκουν σε μεγάλες συνεκτικές συνιστώσες και του $n\zeta_\lambda$ δεν είναι πολύ μεγάλη. Στη διατύπωση του ακόλουθου πορίσματος γίνεται πιο συγκεκριμένη η έννοια του “μεγαλείου”.

Πόρισμα 4.8. Έστω το $k_n = K \log n$ και $u \in (\frac{1}{2}, 1)$. Τότε για K κατάλληλα μεγάλο και για κάθε $\delta < 2u - 1$ ισχύει καθώς το $n \rightarrow \infty$ ότι

$$\mathbb{P}(|Z_{\geq k_n} - n\zeta_\lambda| > n^u) = \mathcal{O}(n^{-\delta}) \quad (4.38)$$

Απόδειξη. Αφού κανένας κόμβος του γραφήματος δεν έχει κάποια ξεχωριστή ιδιότητα σε σχέση με τους υπολοίπους έχουμε ότι

$$\mathbb{E}(Z_{\geq k_n}) = \mathbb{E} \left(\sum_{v \in V} \mathbb{1}_{\{|C_v| \geq k\}} \right) = \sum_{v \in V} \mathbb{P}(|C_v| \geq k) = n\mathbb{P}(|C_1| \geq k)$$

Η πρόταση 4.6 καθιστά την παραπάνω σχέση ως

$$\mathbb{E}(Z_{\geq k_n}) = n \left(\zeta_\lambda + \mathcal{O} \left(\frac{k_n}{n} \right) \right) = n\zeta_\lambda + \mathcal{O}(k_n) \quad (4.39)$$

Επειδή το $k_n = o(n^u)$, για κατάλληλα μεγάλο n θα ισχύει ότι

$$\{|Z_{\geq k_n} - n\zeta_\lambda| > n^u\} \subset \{|Z_{\geq k_n} - \mathbb{E}(Z_{\geq k_n})| > \frac{n^u}{2}\}$$

όπου η διαίρεση με το 2 έγινε γιατί η διαφορά των δύο απόλυτων τιμών είναι κάτι αρκετά μικρό ($o(n^u)$) σε σχέση με το n^u , και άρα για μεγάλο n δε θα ρίξει την απόλυτη τιμή κάτω από το $\frac{n^u}{2}$.

Με τη βοήθεια της ανισότητας Chebyshev, της πρότασης 4.7, της παραπάνω σχέσης, του γεγονότος πως $\chi_{<k_n}(\lambda) \leq k_n$ εξορισμού και για n κατάλληλα μεγάλο, έχουμε ότι

$$\begin{aligned} \mathbb{P}(|Z_{\geq k_n} - n\zeta_\lambda| > n^u) &\leq \mathbb{P}(|Z_{\geq k_n} - \mathbb{E}(Z_{\geq k_n})| > \frac{n^u}{2}) \\ &\leq 4n^{-2u} \text{Var}(Z_{\geq k_n}) \\ &\leq 4n^{-2u} (\lambda k_n + 1) n \chi_{<k_n}(\lambda) \\ &\leq 4n^{-2u+1} (\lambda k_n^2 + k_n) \\ &= \mathcal{O}(n^{-\delta}) \end{aligned}$$

για $\delta < 2u - 1$ και αφού το $n^{-\delta}$ υπερσχύει του $(\log n)^2$ για μεγάλο n . □

Πριν προχωρήσουμε στην πρόταση που δηλώνει πως ουσιαστικά δεν υπάρχουν συνιστώσες “μεσαίου” μεγέθους, θα χρειαστούμε την παρακάτω πρόταση:

Πρόταση 4.9. Σε μια διαδικασία γραφήματος στην οποία ο αριθμός των ενεργών κορυφών τη χρονική στιγμή t περιγράφεται από τη σχέση $S_t = S_{t-1} + X_t - 1 = X_1 + \dots + X_t - t + 1$ όπου τα $X_i, i = 1, \dots, t$ είναι ο αριθμός των νέων κορυφών που γίνονται ενεργές τη συγκεκριμένη χρονική στιγμή, ο S_t ακολουθεί μια διωνυμική κατανομή με παραμέτρους $n-1$ και $1 - (1-p)^t$ αφού αφαιρέσουμε το $t-1$ από αυτή. Προς το παρόν, αυτό το αποτέλεσμα θα το χρησιμοποιούμε μόνο όταν ξέρουμε ότι το $S_t \geq 0$, δηλαδή ότι η συνεκτική συνιστώσα του αρχικού κόμβου κατά τη διαδικασία γραφήματος έχει πλήθος τουλάχιστον ίσο με t .

Απόδειξη. Έστω ότι είμαστε στο τέλος της t -οστής εξερεύνησης. Η πιθανότητα ένας κόμβος να μην έχει γίνει ακόμη ενεργός σε κάποια στιγμή μέχρι και τότε είναι $(1-p)^t$, καθώς σε καθεμία από τις t προηγούμενες εξερευνήσεις είχε ανεξάρτητα πιθανότητα p να γίνει ενεργός. Αυτό όμως σημαίνει πως η πιθανότητα να έχει γίνει κάποια στιγμή ενεργός έως εκείνο το σημείο είναι $1 - (1-p)^t$. Άρα το πλήθος

των ενεργών και ανενεργών κορυφών μέχρι και τη χρονική στιγμή t ακολουθεί μια διωνυμική κατανομή με παραμέτρους $n-1$ (οι αρχικά ουδέτερες κορυφές) και πιθανότητα $1 - (1-p)^t$. Οπότε

$$S_t + t - 1 \sim \text{Bin}(n-1, 1 - (1-p)^t) \quad (4.40)$$

Άρα ο αριθμός των ενεργών κορυφών θα είναι

$$S_t \sim \text{Bin}(n-1, 1 - (1-p)^t) - t + 1 \quad (4.41)$$

□

Για τη συνέχεια, θα χρειαστεί να ορίσουμε τη συνάρτηση

$$J(\alpha, \lambda) = I_{g(\alpha, \lambda)} \quad (4.42)$$

όπου $n I_x = x - 1 - \log x$. Η συνάρτηση $g(\alpha, \lambda)$ έχει τύπο

$$g(\alpha, \lambda) = \frac{1 - e^{-\lambda\alpha}}{\alpha} \quad (4.43)$$

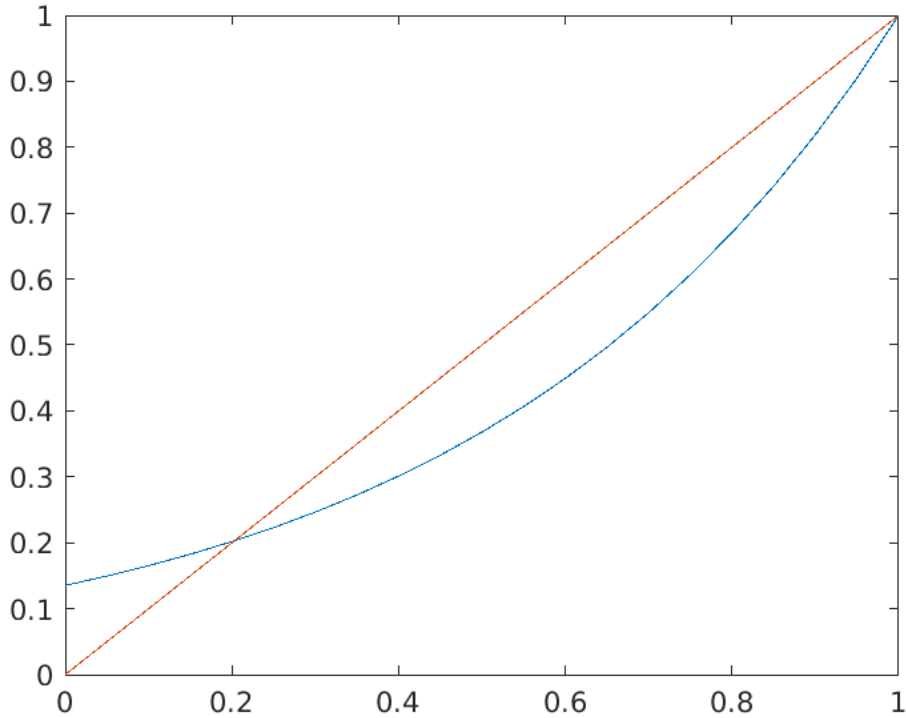
Μερικές παρατηρήσεις πάνω σε αυτές τις συναρτήσεις είναι απαραίτητες πριν συνεχίσουμε. Στην απόδειξη για την πιθανότητα επιβίωσης (3.1) δείξαμε πως $n \eta = \mathbb{P}(T < \infty)$ δίνεται από τη σχέση 3.4: $\eta = G_X(\eta)$ με την G_X να είναι η γεννήτρια συνάρτηση πιθανότητας της τυχαίας μεταβλητής X . Όταν η X ακολουθεί την κατανομή Poisson με παράμετρο λ , η $G_X(s)$ είναι η

$$G_X(s) = e^{\lambda(s-1)} \quad (4.44)$$

Άρα για την πιθανότητα εξαφάνισης στην Poisson(λ) η_λ θα ισχύει πως

$$\begin{aligned} \eta_\lambda &= G_X(\eta_\lambda) \iff \\ \eta_\lambda &= e^{\lambda(\eta_\lambda-1)} \iff \\ 1 - \zeta_\lambda &= e^{\lambda(-\zeta_\lambda)} \end{aligned} \quad \text{αφού } \eta_\lambda = 1 - \zeta_\lambda \quad (4.45)$$

Καταλήγουμε στο συμπέρασμα πως η μοναδική λύση της εξίσωσης $g(\alpha, \lambda) = 1$ στο $[0,1]$ είναι η ζ_λ η οποία όπως είχαμε δει στην απόδειξη του 3.1 είναι μεγαλύτερη του 0, αφού η $\eta_\lambda < 1$ για το $\lambda > 1$. Επιπροσθέτως, σε εκείνη την απόδειξη είχαμε παρατηρήσει πως η συνάρτηση $y_1 = G_X(s)$ είναι αυστηρά κυρτή και αυστηρά αύξουσα και τέμνει την $y_2 = s$ στο $s = \eta_\lambda = 1 - \zeta_\lambda$, όπως στο ακόλουθο σχήμα.



Σχήμα 4.1: Οι y_1 (μπλε) και y_2 (πορτοκαλί) όταν το $\lambda > 1$.

Συγκεντρωτικά έχουμε ότι η $y_1 = e^{\lambda(s-1)}$ και η $y_2 = s$ τέμνονται στο $s = \eta_\lambda$. Όπως φαίνεται και στο σχήμα για την περίπτωση που το $\lambda > 1$ και όπως είχαμε εξηγήσει και στην απόδειξη του θεωρήματος 3.1 για τη γενική περίπτωση (όχι μόνο για την Poisson), για $s < \eta_\lambda$ η y_1 βρίσκεται πιο πάνω από την y_2 και άρα το $\frac{y_1}{y_2} > 1$ για $s \in (0, \eta_\lambda)$. Κάνοντας την παραμετροποίηση $x = 1 - s$ και ανακαλώντας πως $\eta_\lambda = 1 - \zeta_\lambda$, το παραπάνω αποτέλεσμα μπορεί να γραφεί και ως:

Όταν $s < \eta_\lambda \iff 1 - x < 1 - \zeta_\lambda \iff x > \zeta_\lambda$, τότε το $\frac{y_1}{y_2} > 1 \iff \frac{e^{-\lambda x}}{1-x} > 1 \iff g(x, \lambda) < 1$. Με ακριβώς το ίδιο σκεπτικό, όταν $s > \eta_\lambda \iff 1 - x > 1 - \zeta_\lambda \iff x < \zeta_\lambda$, τότε το $\frac{y_1}{y_2} < 1 \iff \frac{e^{-\lambda x}}{1-x} < 1 \iff g(x, \lambda) > 1$.

Με μια απλή ανάλυση της $I_x = x - 1 - \log x$ βλέπουμε ότι $I_x > 0$ για $x \neq 1$. Δηλαδή θα ισχύει πως $J(\alpha, \lambda) = I_{g(\alpha, \lambda)} > 0$, καθώς το $g(\alpha, \lambda) > 1$ για $\alpha < \zeta_\lambda$.

Επίσης, επειδή η $G_X(s)$ είναι αυστηρά κυρτή όπως δείξαμε στο θεώρημα 3.1 και για $\lambda > 1$ ισχύει πως η $G_X(s)$ τέμνει την $y_2 = s$ στο σημείο $s = \eta_\lambda$, θα πρέπει η $G_X(\eta_\lambda) < 1$ και άρα για την περίπτωση που η $X \sim \text{Po}(\lambda)$ ισχύει ότι $\lambda e^{\lambda(\eta_\lambda - 1)} < 1 \iff \lambda e^{-\lambda \zeta_\lambda} - 1 < 0$.

Είμαστε πλέον έτοιμοι να δείξουμε πως οι συνιστώσες “μεσαίου” μεγέθους είναι σχεδόν ανύπαρκτες.

Πρόταση 4.10. Για σταθερό $\lambda > 1$, το k_n να είναι μια συνάρτηση του n τέτοια ώστε $\lim_{n \rightarrow \infty} k_n = \infty$ και για κάθε $\alpha < \zeta_\lambda$ έχουμε πως

$$\mathbb{P}(k_n \leq |C_v| \leq \alpha n) \leq \frac{e^{-k_n J(\alpha, \lambda)}}{1 - e^{-J(\alpha, \lambda)}} \quad (4.46)$$

δηλαδή αυτή η πιθανότητα μειώνεται εκθετικά καθώς μεγαλώνει το μέγεθος του τυχαίου γραφήματος.

Απόδειξη. Δεν κάνει διαφορά στην ανάλυσή μας αν τα $k_n, \alpha n$ είναι ακέραιοι ή όχι, και τα αντιμετωπίζουμε ως ακέραιους, καθώς ασυμπτωτικά δεν έχει καμία σημασία μια τέτοια επιλογή. Το αn θα δούμε ότι θα πρέπει να είναι μικρότερο από το ζ_λ , άρα για να είμαστε τυπικά σωστοί θα υποθέτουμε ότι προσεγγίζεται από το ακέραιο κομμάτι του αριθμού. Επίσης, αφού όλες οι κορυφές έχουν την ίδια αντιμετώπιση, θα μιλάμε και εδώ γενικά για την κορυφή v που είναι μια οποιαδήποτε κορυφή του τυχαίου γραφήματος. Έχουμε λοιπόν ότι

$$\begin{aligned}
\mathbb{P}(k_n \leq |C_v| \leq \alpha n) &= \sum_{t=k_n}^{\alpha n} \mathbb{P}(|C_v| = t) \\
&\leq \sum_{t=k_n}^{\alpha n} \mathbb{P}(S_t = 0) \quad \text{αφού } \{|C_v| = t\} \subset \{S_t = 0\}
\end{aligned} \tag{4.47}$$

Ας ασχοληθούμε λίγο με την $\mathbb{P}(S_t = 0)$.

$$\begin{aligned}
\mathbb{P}(S_t = 0) &= \mathbb{P}(\text{Bin}(n-1, 1 - (1-p)^t) = t-1) \\
&\leq \mathbb{P}(\text{Bin}(n, 1 - (1-p)^t) = t) \\
&\leq \mathbb{P}(\text{Bin}(n, 1 - (1-p)^t) \leq t) \\
&\leq \mathbb{P}(\text{Bin}(n, 1 - e^{-pt}) \leq t)
\end{aligned} \tag{4.48}$$

όπου με $\text{Bin}(a, b)$ συμβολίζουμε μια τυχαία μεταβλητή που ακολουθεί την κατανομή $\text{Bin}(a, b)$. Επίσης, η πρώτη ανισότητα ισχύει, γιατί εάν σε μια διωνυμική διαδικασία προσθέσουμε μια δοκιμή Bernoulli και η πιθανότητα αυτή να πραγματοποιηθεί δεν είναι 1, τότε η συνολική πιθανότητα στη νέα διωνυμική να πραγματοποιηθούν μια παραπάνω δοκιμές θα είναι μικρότερη. Η τελευταία ανισότητα ισχύει καθώς $1 - (1-p)^t \geq 1 - e^{-pt}$ και άρα ανάμεσα σε δύο διωνυμικές διαδικασίες με τον ίδιο αριθμό δοκιμών, αυτή που δίνει μικρότερη πιθανότητα ικανοποίησης της καθεμιάς δοκιμής θα δίνει μεγαλύτερη πιθανότητα στο ενδεχόμενο οι επιτυχίες να είναι το πολύ t το-πλήθος. Για το τελευταίο επιχείρημα που είναι ουσιαστικά ένα επιχείρημα μονοτονίας, μπορεί κανείς να δει και την ενότητα 4.8.

Χρησιμοποιώντας την ανισότητα Markov μπορούμε να δούμε μερικά πράγματα για την τελευταία ανισότητα:

$$\begin{aligned}
\mathbb{P}(\text{Bin}(n, 1 - e^{-pt}) \leq t) &= \mathbb{P}(e^{-\gamma \text{Bin}(n, 1 - e^{-pt})} \geq e^{-\gamma t}) \\
&\leq \frac{\mathbb{E}(e^{-\gamma \text{Bin}(n, 1 - e^{-pt})})}{e^{-\gamma t}} \\
&= e^{\gamma t} [(1 - e^{-pt})e^{-\gamma} + e^{-pt}]^n \\
&= e^{\gamma t} [1 - 1 + (1 - e^{-pt})e^{-\gamma} + e^{-pt}]^n \\
&= e^{\gamma t} [1 - (1 - e^{-pt})(1 - e^{-\gamma})]^n \\
&\leq e^{\gamma t - (1 - e^{-pt})(1 - e^{-\gamma})n}
\end{aligned} \tag{4.49}$$

Το μόνο που μένει να κάνουμε είναι να ελαχιστοποιήσουμε τον εκθέτη. Παραγωγίζοντας ως προς το γ και θέτοντας το αποτέλεσμα ίσο με 0, η κρίσιμη τιμή του γ είναι n

$$\gamma' = \ln \left(\frac{n(1 - e^{-pt})}{t} \right) = \ln \left(\frac{1 - e^{-\frac{\lambda t}{n}}}{\frac{t}{n}} \right) = \ln \left(g \left(\frac{t}{n}, \lambda \right) \right)$$

Από αυτά που δείξαμε για τη συνάρτηση 4.43, καταλήγουμε στο συμπέρασμα πως για να είναι θετικό το γ' πρέπει η ποσότητα του λογαρίθμου να είναι θετική, δηλαδή το $g(\frac{t}{n}, \lambda) > 1$ και άρα το $\frac{t}{n} < \zeta_\lambda \iff \alpha < \zeta_\lambda$. Τότε αντικαθιστώντας το γ' στην 4.49 θα πάρουμε ότι

$$\begin{aligned}
\mathbb{P}(\text{Bin}(n, 1 - e^{-pt}) \leq t) &\leq e^{\ln(g(\frac{t}{n}, \lambda))t - (1 - e^{-\frac{\lambda t}{n}})(1 - e^{-\ln(g(\frac{t}{n}, \lambda))})n} \\
&= e^{\ln(g(\frac{t}{n}, \lambda))t - t g(\frac{t}{n}, \lambda)(1 - \frac{1}{g(\frac{t}{n}, \lambda)})} \\
&= e^{-t I_g(\alpha, \lambda)} \quad \text{αφού το } \alpha = \frac{t}{n} \\
&= e^{-t J(\alpha, \lambda)}
\end{aligned} \tag{4.50}$$

Τελικά, ο συνσυσμασμός των 4.47, 4.48, 4.49 και 4.50, το γεγονός πως για $\alpha < \zeta_\lambda$ το $g(\alpha, \lambda) > 1$ και άρα το $J(\alpha, \lambda) > 0$ καθώς και η ενσωμάτωση ολοκλήρωσης της ουράς στο 4.47 μας δίνουν το επιθυμητό αποτέλεσμα. \square

Πόρισμα 4.11. Για $k_n = K \log n$, $\alpha < \zeta_\lambda$ και $\delta = KJ(\alpha, \lambda) - 1$ με το K να είναι τέτοιο ώστε το δ να είναι θετικό, η πιθανότητα να μην υπάρχει συνεκτική συνιστώσα μεγέθους μεταξύ $K \log n$ και αn είναι τουλάχιστον $1 - \mathcal{O}(n^{-\delta})$.

Απόδειξη. Από την ανισότητα του Boole και από το γεγονός πως όλες οι κορυφές του τυχαίου γραφήματος αντιμετωπίζονται με τον ίδιο τρόπο χωρίς κάποια ειδική γνώση για το γράφημα έχουμε ότι

$$\begin{aligned} \mathbb{P}(\exists v : k_n \leq |C_v| \leq \alpha n) &= \mathbb{P}(\cup_{v \in V} (k_n \leq |C_v| \leq \alpha n)) \\ &\leq \sum_{v \in V} \mathbb{P}(k_n \leq |C_v| \leq \alpha n) \\ &= n \mathbb{P}(k_n \leq |C_v| \leq \alpha n) \\ &\leq n \frac{e^{-k_n J(\alpha, \lambda)}}{1 - e^{-J(\alpha, \lambda)}} \quad , \text{ από την 4.10} \\ &= \frac{1}{1 - e^{-J(\alpha, \lambda)}} n^{1 - KJ(\alpha, \lambda)} \\ &= \mathcal{O}(n^{-\delta}) \end{aligned}$$

\square

Μας έχει μείνει ένα μόνο βήμα μέχρι να δείξουμε το αρχικό ζητούμενο αυτής της ενότητας, δηλαδή πως ο αριθμός των κορυφών που ανήκουν στη μεγαλύτερη συνεκτική συνιστώσα ενός τυχαίου γραφήματος με $\lambda > 1$ είναι πολύ κοντά στο γινόμενο μιας σταθεράς επί το n , ή αλλιώς πως: $|C_{\max}| = \Theta(n)$.

Αυτό το βήμα εμπεριέχει τον ορισμό ενός ενδεχομένου και την απόδειξη ενός λήμματος για αυτό.

Ορισμός 4.12. Για $u \in (\frac{1}{2}, 1)$, για $\alpha \in (\frac{\zeta_\lambda}{2}, \zeta_\lambda)$ και $k_n = K \log n$ με το K να είναι κατάλληλα μεγάλο ορίζουμε το ενδεχόμενο \mathcal{E}_n ως

$$\mathcal{E}_n = \{|Z_{\geq k_n} - n\zeta_\lambda| \leq n^u\} \cap \{\nexists v \in V : k_n \leq |C_v| \leq \alpha n\} \quad (4.51)$$

Λήμμα 4.13. Το \mathcal{E}_n πραγματοποιείται με πιθανότητα τουλάχιστον $1 - \mathcal{O}(n^{-\delta})$ για το μικρότερο δ εκ των δύο που εμφανίζονται στα πορίσματα 4.8 και 4.11. Επιπλέον, όταν πραγματοποιείται το \mathcal{E}_n ισχύει ότι $|C_{\max}| = Z_{\geq k_n}$.

Απόδειξη. Θα υπολογίσουμε την πιθανότητα να μην πραγματοποιηθεί το \mathcal{E}_n . Εφόσον το \mathcal{E}_n είναι η τομή δύο ενδεχομένων, το συμπλήρωμά του θα είναι η ένωση των συμπληρωμάτων των δύο έκαστων ενδεχομένων. Μαζί με την ανισότητα του Boole θα έχουμε τότε πως

$$\begin{aligned} \mathbb{P}(\mathcal{E}_n^c) &\leq \mathbb{P}(|Z_{\geq k_n} - n\zeta_\lambda| > n^u) + \mathbb{P}(\exists v \in V : k_n \leq |C_v| \leq \alpha n) \\ &= \mathcal{O}(n^{-\delta_1}) + \mathcal{O}(n^{-\delta_2}) \quad , \text{ από τα 4.8 και 4.11} \\ &= \mathcal{O}(n^{-\delta_{\min}}) \end{aligned}$$

Τώρα δεδομένου πως πραγματοποιείται το \mathcal{E}_n , θα ισχύει ότι $|Z_{\geq k_n} - n\zeta_\lambda| \leq n^u$ και άρα το $Z_{\geq k_n} \geq 1$ αναγκαστικά, διαφορετικά θα έπρεπε το $n\zeta_\lambda \leq n^u$ το οποίο δεν αληθεύει ασυμπτωτικά. Αυτό σημαίνει ότι υπάρχει τουλάχιστον μια κορυφή που ανήκει σε συνεκτική συνιστώσα μεγέθους τουλάχιστον ίσου με k_n και άρα η μεγαλύτερη συνιστώσα θα έχει μέγεθος τουλάχιστον ίσο με k_n . Αυτό σημαίνει όμως πως επειδή μπορεί να υπάρχουν και άλλες συνιστώσες μεγέθους τουλάχιστον ίσου με k_n , θα έχουμε ότι $|C_{\max}| \leq Z_{\geq k_n}$.

Επιπλέον η πραγματοποίηση του \mathcal{E}_n σηματοδοτεί πως δεν υπάρχουν συνεκτικές συνιστώσες μεγέθους ανάμεσα στο k_n και το αn . Άρα αν έχουμε δύο συνιστώσες μεγαλύτερες από αn , το $Z_{\geq k_n} \geq 2\alpha n > \zeta_\lambda n$. Όμως το \mathcal{E}_n μας δίνει πως πρέπει $Z_{\geq k_n} \leq n\zeta_\lambda + n^u$ κάτι που ασυμπτωτικά αντιβαίνει στο ενδεχόμενο να υπάρχουν δύο συνεκτικές συνιστώσες μεγαλύτερες από αn . Επομένως είναι μόνο μία αυτή και $|C_{\max}| = Z_{\geq k_n}$.

□

Θεώρημα 4.14. Όταν το λ είναι σταθερό με $\lambda > 1$, τότε για κάθε $u \in (\frac{1}{2}, 1)$ υπάρχει $\delta > 0$ τέτοιο ώστε

$$\mathbb{P}(|C_{\max}| - \zeta_\lambda n \leq n^u) \geq 1 - \mathcal{O}(n^{-\delta}) \quad (4.52)$$

Απόδειξη.

$$\begin{aligned} \mathbb{P}(|C_{\max}| - \zeta_\lambda n \leq n^u) &\geq \mathbb{P}(\{|C_{\max}| - \zeta_\lambda n \leq n^u\} \cap \mathcal{E}_n) \\ &= \mathbb{P}(\mathcal{E}_n) && \text{, από το 4.13} \\ &\geq 1 - \mathcal{O}(n^{-\delta}) && \text{, από το 4.13} \end{aligned}$$

□

Βλέπουμε λοιπόν πως όταν το λ είναι σταθερό και $\lambda > 1$ στο τυχαίο γράφημα με n κόμβους και πιθανότητα $p = \frac{\lambda}{n}$ εμφάνισης κάθε δυνατής ακμής του, υπάρχει μια γιγαντιαία συνεκτική συνιστώσα σε αυτό της οποίας το μέγεθος είναι αρκετά κοντά στο $\zeta_\lambda n$. Ωστόσο, οφείλουμε να παρατηρήσουμε πως ο όρος γιγαντιαία μπορεί να μη δικαιολογείται με την έννοια που τον έχουμε στο μυαλό μας. Παρόλο που η μεγαλύτερη συνιστώσα θα έχει μέγεθος $\zeta_\lambda n = \Theta(n)$ αν η πιθανότητα επιβίωσης ζ_λ είναι πολύ μικρή, τότε ίσως η λέξη "γιγαντιαία" να φαντάζει υπερβολική. Για παράδειγμα, σε ένα γράφημα με 1000 κόμβους και τη $\zeta_\lambda = 0.01$, οι 10 κορυφές από τις οποίες θα αποτελείται πάνω-κάτω η μεγαλύτερη συνεκτική συνιστώσα είναι μάλλον πολύ λίγες. Ωστόσο, η ανάλυση που κάναμε για την περίπτωση που το $\lambda > 1$ είναι συναρπαστική περισσότερο ως προς την αλλαγή της τάξης μεγέθους της μεγαλύτερης συνιστώσας του τυχαίου γραφήματος. Από εκεί που για $\lambda < 1$ το μέγεθος αυτής ήταν το πολύ $\log n$, τώρα είναι $\Theta(n)$. Οι τιμές του λ "πολύ κοντά" στο 1 παρουσιάζουν ιδιαίτερο ενδιαφέρον και συζητούνται στη συνέχεια (ενότητα 4.6), όμως για λ σταθερό και μεγαλύτερο ή μικρότερο του 1, η διαφορά στη μεγαλύτερη συνεκτική συνιστώσα είναι ξεκάθαρη όπως φαίνεται και στο παράδειγμα στην ενότητα 4.9.

Επίσης, δεν πρέπει να ξεχνάμε πως το τυχαίο γράφημα είναι ένας χώρος πιθανότητας που περιέχει κάθε δυνατό γράφημα σε n κορυφές. Έτσι, φυσικά δεν αποκλείεται σε ένα $G(n, p_1)$ με $\lambda_1 < 1$ η μεγαλύτερη συνεκτική συνιστώσα να υπερέχει σε μέγεθος της αντίστοιχης σε ένα $G(n, p_2)$ με $\lambda_2 > 1$, απλά κάτι τέτοιο δεν είναι ο κανόνας.

4.5 Κεντρικό Οριακό Θεώρημα για τη γιγαντιαία συνιστώσα

Μπορούμε όμως να δείξουμε και κάτι ισχυρότερο από το θεώρημα 4.14. Συγκεκριμένα ένα Κεντρικό Οριακό Θεώρημα για το μέγεθος της γιγαντιαίας συνιστώσας.

Θεώρημα 4.15. Έστω μια σταθερά $\lambda > 1$. Τότε ισχύει ότι

$$\frac{|C_{\max}| - \zeta_\lambda n}{\sqrt{n}} \xrightarrow{d} Z \quad (4.53)$$

όπου η Z είναι μια τυχαία μεταβλητή η οποία ακολουθεί την κανονική κατανομή με μέσο όρο 0 και διασπορά

$$\sigma_\lambda^2 = \frac{\zeta_\lambda(1 - \zeta_\lambda)}{(1 - \lambda + \lambda\zeta_\lambda)^2} \quad (4.54)$$

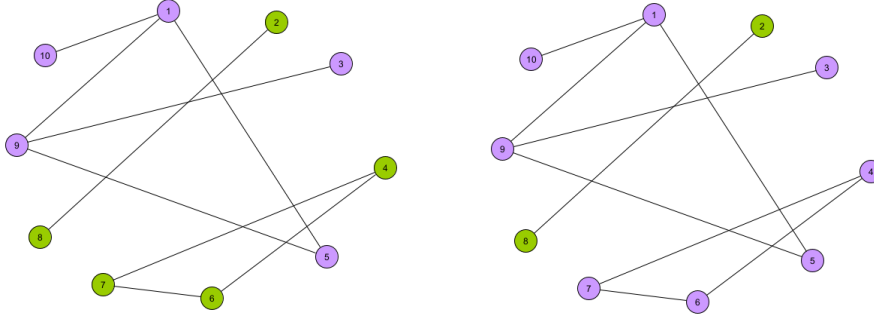
Για να αποδείξουμε το παραπάνω θεώρημα θα χρειαστεί να εισάγουμε την έννοια της ένωσης συνεκτικών συνιστωσών. Έτσι, για τις κορυφές $[k] = \{1, 2, \dots, k\}$ η $C_{[k]}$ είναι το σύνολο που αποτελείται από όλες τις κορυφές που ανήκουν στις συνεκτικές συνιστώσες των κορυφών $[k] = \{1, 2, \dots, k\}$. Πριν προχωρήσουμε στην απόδειξη του θεωρήματος θα δούμε μια πρόταση παρόμοια με την πρόταση 4.9 με τη διαφορά πως εδώ δεν αρχίζουμε την αναζήτηση κατά πλάτος έχοντας μια ενεργή κορυφή, αλλά k τέτοιες. Δηλαδή αυτή η τροποποιημένη διαδικασία γραφήματος έχει

$$S_t = S_{t-1} + X_t - 1 \quad (4.55)$$

και

$$X_t \sim \text{Bin}(n - S_{t-1} - (t - 1), p) \quad (4.56)$$

και αρχικό πλήθος ενεργών κόμβων $S_0 = k$ αντί για $S_0 = 1$. Είναι δηλαδή μια πιο γενική διαδικασία γραφήματος.



Σχήμα 4.2

Στο παραπάνω σχήμα στο αριστερό γράφημα βλέπουμε το αποτέλεσμα της γενικευμένης διαδικασίας γραφήματος εάν οι αρχικές κορυφές είναι οι $\{1, 9\}$, ενώ στο δεξί εάν αυτές είναι οι $\{1, 4, 9\}$. Και στις δύο περιπτώσεις οι μωβ κορυφές είναι αυτές που θα είναι ανενεργές μετά το τέλος της διαδικασίας και με πράσινο χρωματίζονται αυτές που θα έχουν παραμείνει ουδέτερες έως το τέλος.

Πρόταση 4.16. Για $t, k \in [n]$ έχουμε ότι

$$S_t + (t - k) \sim \text{Bin}(n - k, 1 - (1 - p)^t) \quad (4.57)$$

Επιπλέον για τα $t \leq l$, $t, l \in [n]$ και δεδομένου του S_t ισχύει πως

$$S_l + (l - t) - S_t \sim \text{Bin}(n - t - S_t, 1 - (1 - p)^{l-t}) \quad (4.58)$$

Το $S_t + t$ είναι το πλήθος των κορυφών που είναι ενεργές ή ανενεργές μετά τον t -οστό γύρο της εξερεύνησης της διαδικασίας γραφήματος. Αυτό σημαίνει πως το πλήθος των ουδέτερων κορυφών μετά τον t -οστό γύρο είναι $N_t = n - S_t - t$. Για να δείξουμε τη σχέση 4.57 είναι πιο εύκολο να δείξουμε πως $N_t \sim \text{Bin}(n - k, (1 - p)^t)$, το οποίο είναι ισοδύναμο αφού αν το $N_t = n - S_t - t = n - k - S_t - t + k \sim \text{Bin}(n - k, (1 - p)^t)$, τότε το $-(-S_t - t + k) = S_t + t - k \sim \text{Bin}(n - k, 1 - (1 - p)^t)$.

Δεδομένου του πλήθους των ενεργών κορυφών μετά τη χρονική στιγμή $t-1$ το $X_t \sim \text{Bin}(n - S_{t-1} - (t - 1), p) = \text{Bin}(N_{t-1}, p)$. Δηλαδή καθεμία από τις εναπομείνουσες ουδέτερες κορυφές έχει πιθανότητα p να επιλεγεί την επόμενη χρονική στιγμή. Γράφουμε

$$\begin{aligned} N_t &= n - t - S_t \\ &= n - t - S_{t-1} - X_t + 1 && \text{από την 4.55} \\ &= N_{t-1} - X_t \sim \text{Bin}(N_{t-1}, 1 - p) && \text{αφού } X_t \sim \text{Bin}(N_{t-1}, p) \end{aligned} \quad (4.59)$$

Επιπλέον αρχικά υπάρχουν k ενεργές κορυφές και άρα $N_0 = n - k$. Αυτό το γεγονός σε συνδυασμό με μια αναδρομή επί της 4.59 μας δίνει ότι $N_t \sim \text{Bin}(n - k, (1 - p)^t)$, όπου χρησιμοποιούμε πως μια M_2 με $M_2 | M_1 \sim \text{Bin}(M_1, q_2)$ και $M_1 \sim \text{Bin}(m, q_1)$ ακολουθεί διωνυμική κατανομή $M_2 \sim \text{Bin}(m, q_1 q_2)$.

Παρατηρούμε πως η 4.59 μέσω της ίδιας αναδρομής μας δίνει επίσης πως για $l \geq t$

$$\begin{aligned} N_l &\sim \text{Bin}(N_t, (1 - p)^{l-t}) \\ n - l - S_l &\sim \text{Bin}(n - t - S_t, (1 - p)^{l-t}) \\ n - l - S_l &\sim n - t - S_t - \text{Bin}(n - t - S_t, 1 - (1 - p)^{l-t}) \\ S_l + (l - t) - S_t &\sim \text{Bin}(n - t - S_t, 1 - (1 - p)^{l-t}) \end{aligned}$$

Μπορούμε τώρα να πάρουμε ένα χρήσιμο πόρισμα που μας ενημερώνει για το Κεντρικό Οριακό Θεώρημα στην ποσότητα S_{nt} με το $t \in (0, 1)$.

Πόρισμα 4.17. Έστω $k = k_n = o(\sqrt{n})$ και $S_0 = k$. Τότε για κάθε $t \in (0, 1)$ η τυχαία μεταβλητή $\frac{S_{nt} - n\mu_t}{\sqrt{nv_t}}$ συγκλίνει σε μια τυχαία μεταβλητή που ακολουθεί την τυποποιημένη κανονική κατανομή, όπου

$$\mu_t = 1 - t - e^{-\lambda t} \quad (4.60)$$

$$v_t = e^{-\lambda t}(1 - e^{-\lambda t}) \quad (4.61)$$

είναι ο ασυμπτωτικός μέσος και η ασυμπτωτική διασπορά του S_{nt} .

Απόδειξη. Έχουμε ότι

$$\begin{aligned} \mathbb{E}(S_{nt}) &= (n - k) \left(1 - \left(1 - \frac{\lambda}{n} \right)^{nt} \right) - (nt - k) && \text{λόγω της 4.57} \\ &= n \left(1 - \left(1 - \frac{\lambda}{n} \right)^{nt} \right) + k \left(1 - \frac{\lambda}{n} \right)^{nt} - nt \\ &= n(1 - e^{-\lambda t} - t) + ke^{-\lambda t} + o(1) && \text{αφού } \left(1 + \frac{x}{n} \right)^n \rightarrow e^x \\ &= n\mu_t + o(\sqrt{n}) \end{aligned} \quad (4.62)$$

και

$$\begin{aligned} \text{Var}(S_{nt}) &= (n - k) \left(1 - \frac{\lambda}{n} \right) \left(1 - \left(1 - \frac{\lambda}{n} \right)^{nt} \right) && \text{λόγω της 4.57} \\ &= n(e^{-\lambda t}(1 - e^{-\lambda t})) + o(\sqrt{n}) \\ &= nv_t + o(\sqrt{n}) \end{aligned} \quad (4.63)$$

Για τη διωνυμική κατανομή το Κεντρικό Οριακό Θεώρημα διατυπώνεται ως εξής:

$$\frac{S_{nt} - \mathbb{E}(S_{nt})}{\sqrt{\text{Var}(S_{nt})}} \xrightarrow{d} Z \quad (4.64)$$

όπου Z είναι μια τυχαία μεταβλητή που ακολουθεί την τυποποιημένη κανονική κατανομή. Η τυχαία μεταβλητή που μας ενδιαφέρει είναι η

$$\frac{S_{nt} - n\mu_t}{\sqrt{nv_t}} = \sqrt{\frac{\text{Var}(S_{nt})}{nv_t}} \frac{S_{nt} - \mathbb{E}(S_{nt})}{\sqrt{\text{Var}(S_{nt})}} + \frac{\mathbb{E}(S_{nt}) - n\mu_t}{\sqrt{nv_t}} \quad (4.65)$$

Ο τελευταίος όρος τείνει στο 0 εξαιτίας της 4.62, ενώ το $\sqrt{\frac{\text{Var}(S_{nt})}{nv_t}}$ τείνει στο 1 εξαιτίας της 4.63 και άρα σε συνδυασμό με την 4.64 το ζητούμενο αποδεικνύεται. \square

Έστω ότι στη γενικευμένη διαδικασία γραφήματος επιλέγουμε $k = k_n = o(\sqrt{n})$, με $k \rightarrow \infty$ αρχικές κορυφές. Το μέγεθος της ένωσης των συνεκτικών συνιστωσών τους είναι $|C_{[k]}| = \min\{m : S_m = 0\}$. Υπενθυμίζουμε πως από την προηγούμενη ενότητα ξέρουμε πως η δεύτερη σε μέγεθος συνιστώσα θα είναι μικρότερη από $K \log n$ εξαιτίας του πορίσματος 4.11 και του θεωρήματος 4.14 με πιθανότητα που τείνει στο 1. Επίσης η γιγαντιαία συνιστώσα έχει μέγεθος $\zeta_\lambda n(1 + o(1))$ από το θεώρημα 4.14. Τέλος, η $C_{[k]}$ θα περιέχει τουλάχιστον μια κορυφή που ανήκει στη γιγαντιαία συνιστώσα με μεγάλη πιθανότητα.

Το τελευταίο συμπέρασμα μπορούμε να το δούμε και με τον εξής τρόπο. Από τις n κορυφές του γραφήματος οι περίπου $\zeta_\lambda n$ θα ανήκουν στη γιγαντιαία συνιστώσα και οι υπόλοιπες $(1 - \zeta_\lambda)n$ όχι. Άρα εάν επιλέξουμε τυχαία μια κορυφή από το τυχαίο γράφημα, η πιθανότητα αυτή να μην ανήκει στη γιγαντιαία συνιστώσα είναι $\frac{(1 - \zeta_\lambda)n}{n} = (1 - \zeta_\lambda)$. Άρα η πιθανότητα καμία από τις k κορυφές να μην ανήκει στη γιγαντιαία συνιστώσα θα είναι της μορφής $(1 - \zeta_\lambda)^k$ και επειδή $k \rightarrow \infty$ και το $\zeta_\lambda > 0$ για $\lambda > 1$ αυτή η πιθανότητα τείνει στο 0.

Δηλαδή $|C_{\max}| \leq |C_{[k]}|$. Επίσης στην περίπτωση που μόνο μια από τις k κορυφές ανήκει στη γιγα-ντιαία συνιστώσα και οι υπόλοιπες ανήκουν σε ξεχωριστές συνιστώσες μεγέθους μέχρι $K\log n$ θα έχουμε ότι $|C_{[k]}| \leq |C_{\max}| + (k-1)K\log n$.

Τελικά συνδυάζοντας αυτές τις δύο παρατηρήσεις βλέπουμε πως ισχύει ότι αφού το $k_n = o(\sqrt{n})$, το $|C_{[k]}|$ θα είναι της μορφής $|C_{\max}| + o(\sqrt{n})$. Αυτό σημαίνει πως άμα δείξουμε το Κεντρικό Οριακό Θεώρημα για το $|C_{[k]}|$, τότε αυτό έπεται και για το $|C_{\max}|$ και άρα το θεώρημα 4.15 αποδεικνύεται.

Θα δείξουμε το Κεντρικό Οριακό Θεώρημα για το $|C_{[k]}|$ χρησιμοποιώντας άνω και κάτω φράγματα στην πιθανότητα

$$\mathbb{P}\left(\frac{|C_{[k]}| - \zeta_\lambda n}{\sqrt{n}} > x\right)$$

Πρώτα θα δούμε το **άνω φράγμα**. Έχουμε ότι το $|C_{[k]}| = \min\{m : S_m = 0\}$ και άρα

$$\mathbb{P}(|C_{[k]}| > l) = \mathbb{P}(S_m > 0, \forall m \leq l) \quad (4.66)$$

Αυτό σημαίνει πως για $l = \zeta_\lambda n + x\sqrt{n}$ θα ισχύει ότι

$$\mathbb{P}\left(\frac{|C_{[k]}| - \zeta_\lambda n}{\sqrt{n}} > x\right) = \mathbb{P}(S_m > 0, \forall m \leq l) \leq \mathbb{P}(S_l > 0) \quad (4.67)$$

Αν δούμε τον ασυμπτωτικό μέσο $\mu_t = 1 - t - e^{-\lambda t}$ ως συνάρτηση ως προς το t και και τον αναπτύξουμε ως προς το ζ_λ θα πάρουμε ότι

$$\mu_t = \mu(t) = \mu(\zeta_\lambda) + \mu'(\zeta_\lambda)(t - \zeta_\lambda) + \dots$$

Όμως από τη σχέση 4.45 παίρνουμε πως το $\mu(\zeta_\lambda) = 0$. Τελικά γράφοντας $l = \zeta_\lambda n + x\sqrt{n} = n(\zeta_\lambda + \frac{x}{\sqrt{n}})$ μπορούμε να πάρουμε

$$\begin{aligned} \mathbb{E}(S_l) &= n\mu(\zeta_\lambda + \frac{x}{\sqrt{n}}) + o(\sqrt{n}) && \text{από την 4.62} \\ &= n(\mu'(\zeta_\lambda)(\zeta_\lambda + \frac{x}{\sqrt{n}} - \zeta_\lambda) + \dots) + o(\sqrt{n}) \\ &= n\mu'(\zeta_\lambda)(\frac{x}{\sqrt{n}}) + o(\sqrt{n}) \\ &= \sqrt{n}x(\lambda e^{-\lambda\zeta_\lambda} - 1) + o(\sqrt{n}) \end{aligned}$$

όπου το $\lambda e^{-\lambda\zeta_\lambda} - 1 < 0$ όταν το $\lambda > 1$, από τη συζήτηση κάτω από τη σχέση 4.45.

Με ένα παρόμοιο ανάπτυγμα Taylor δείχνεται και ότι

$$\text{Var}(S_l) = n\nu_{\zeta_\lambda} + o(\sqrt{n})$$

Εν τέλει ισχύει πως

$$\mathbb{P}(S_l > 0) = \mathbb{P}\left(\frac{S_l - \mathbb{E}(S_l)}{\sqrt{\text{Var}(S_l)}} > \frac{x(1 - \lambda e^{-\lambda\zeta_\lambda})}{\sqrt{\nu_{\zeta_\lambda}}}\right) + o(1) \quad (4.68)$$

Από το πόρισμα 4.17 έχουμε ότι

$$\mathbb{P}\left(\frac{S_l - \mathbb{E}(S_l)}{\sqrt{\text{Var}(S_l)}} > \frac{x(1 - \lambda e^{-\lambda\zeta_\lambda})}{\sqrt{\nu_{\zeta_\lambda}}}\right) \rightarrow \mathbb{P}\left(Z > \frac{x(1 - \lambda e^{-\lambda\zeta_\lambda})}{\sqrt{\nu_{\zeta_\lambda}}}\right) = \mathbb{P}(Z' > x) \quad (4.69)$$

Με τη Z' να ακολουθεί κανονική κατανομή με μέσο όρο το 0 και διασπορά $\frac{\nu_{\zeta_\lambda}}{(1 - \lambda e^{-\lambda\zeta_\lambda})^2}$. Χρησιμοποιώντας και τη σχέση 4.45 η διασπορά της Z' είναι n

$$\frac{v_{\zeta_\lambda}}{(1 - \lambda e^{-\lambda \zeta_\lambda})^2} = \frac{\zeta_\lambda(1 - \zeta_\lambda)}{(1 - \lambda + \lambda \zeta_\lambda)^2} = \sigma_\lambda^2$$

όπου η σ_λ^2 είναι η ίδια με αυτή της σχέσης 4.54. Τελικά έχουμε ότι

$$\mathbb{P}\left(\frac{|C_{[k]}| - \zeta_\lambda n}{\sqrt{n}} > x\right) \leq \mathbb{P}(Z' > x) \quad (4.70)$$

με την Z' να είναι η επιθυμητή για το θεώρημα 4.15.

Μας μένει να δείξουμε το αντίστοιχο **κάτω φράγμα**. Θα χρησιμοποιήσουμε και πάλι πως για $l = \zeta_\lambda n + x\sqrt{n}$ ισχύει ότι $\mathbb{P}\left(\frac{|C_{[k]}| - \zeta_\lambda n}{\sqrt{n}} > x\right) = \mathbb{P}(S_m > 0, \forall m \leq l)$. Τότε, για κάθε $\varepsilon > 0$ μπορούμε να φράξουμε αυτήν την πιθανότητα από κάτω παίρνοντας την τομή με το ενδεχόμενο $S_l > \varepsilon\sqrt{n}$.

$$\begin{aligned} \mathbb{P}(S_m > 0, \forall m \leq l) &\geq \mathbb{P}(\{S_m > 0, \forall m \leq l\} \cap \{S_l > \varepsilon\sqrt{n}\}) \\ &= \mathbb{P}(S_l > \varepsilon\sqrt{n}) - \mathbb{P}(S_l > \varepsilon\sqrt{n}, \exists m < l : S_m = 0) \end{aligned} \quad (4.71)$$

Για να μη δημιουργηθεί σύγχυση σχετικά με τη δυνατότητα μια διαδικασία γραφήματος να δίνει θετικό πληθυσμό ($S_l > \varepsilon\sqrt{n}$) σε μια χρονική στιγμή μετά τον πρώτο μηδενισμό του ($\exists m < l : S_l = 0$), θα δώσουμε τη δυνατότητα η διαδικασία γραφήματος να συνεχίζεται σύμφωνα με τον τύπο της ακόμα και αν δεν έχει κάποιο φυσικό νόημα μιας και η ένωση των συνεκτικών συνιστωσών θα έχει ήδη βρεθεί.

Δηλαδή, πλέον μιλάμε για τη διαδικασία γραφήματος με $S_0 = k$ αρχικές ενεργές κορυφές, αναδρομική σχέση $S_t = S_{t-1} + X_t - 1$, όπου $X_t \sim \text{Bin}(n - S_{t-1} - (t-1), p)$, η οποία συνεχίζεται για κάθε χρονική στιγμή $t \in \{0, 1, \dots, n\}$.

Προσέξτε πως αυτός ο ορισμός δεν αντιτίθεται στα έως τώρα λεγόμενά μας για την S_t όπως για παράδειγμα στην πρόταση 4.16.

Συνεχίζοντας, την πιθανότητα $\mathbb{P}(S_l > \varepsilon\sqrt{n})$ μπορούμε να την αντιμετωπίσουμε ακριβώς με τον ίδιο τρόπο όπως μεταχειριστήκαμε την πιθανότητα $\mathbb{P}(S_l > 0)$ στο άνω φράγμα προηγουμένως και να λάβουμε

$$\mathbb{P}(S_l > \varepsilon\sqrt{n}) \rightarrow \mathbb{P}\left(Z > \frac{x(1 - \lambda e^{-\lambda \zeta_\lambda}) + \varepsilon}{\sqrt{v_{\zeta_\lambda}}}\right) = \mathbb{P}\left(Z' > x + \frac{\varepsilon}{1 - \lambda e^{-\lambda \zeta_\lambda}}\right) \quad (4.72)$$

με τη Z' να είναι μια τυχαία μεταβλητή που ακολουθεί την κανονική κατανομή με μέσο όρο το 0 και διασπορά τη σ_λ^2 . Όταν το ε τείνει στο 0 από τα θετικά αυτή η πιθανότητα τείνει στην $\mathbb{P}(Z' > x)$ όπως θα θέλαμε. Άρα το μόνο που έχει μείνει να δείξουμε είναι το πως η πιθανότητα

$$\mathbb{P}(S_l > \varepsilon\sqrt{n}, \exists m < l : S_m = 0) = o(1)$$

Αρχικά, έχουμε πως

$$\mathbb{P}(S_l > \varepsilon\sqrt{n}, \exists m < l : S_m = 0) \leq \sum_{m=k}^{l-1} \mathbb{P}(S_m = 0, S_l > \varepsilon\sqrt{n}) \quad (4.73)$$

εξαιτίας της ανισότητας Boole και του γεγονότος πως οι αρχικά ενεργές κορυφές ήταν k και άρα πριν το πέρας της k -οστής εξερεύνησης δε γίνεται το $S_m = 0$ εξ'ορισμού.

Σε αυτό το σημείο μπορούμε για τα m για τα οποία ισχύει ότι $m \leq \alpha n$ για $\alpha < \zeta_\lambda$ να δούμε πως

$$\mathbb{P}(S_m = 0) \leq e^{-mJ(\frac{\alpha}{n}, \lambda)} \quad (4.74)$$

με την απόδειξη να είναι σχεδόν η ίδια με αυτή του πορίσματος 4.10 (βλ. σχέσεις 4.48 και 4.50). Συγκεκριμένα, επειδή πλέον είμαστε στη γενικευμένη διαδικασία γραφήματος στην οποία οι αρχικά ενεργές κορυφές είναι k το πλήθος, εξαιτίας της πρότασης 4.16 η ισότητα στην αρχή της σχέσης 4.48 θα γίνει

$$\begin{aligned}\mathbb{P}(S_m = 0) &= \mathbb{P}(\text{Bin}(n-k, 1-(1-p)^m) = m-k) \\ &\leq \mathbb{P}(\text{Bin}(n, 1-(1-p)^m) = m)\end{aligned}$$

και από εκεί και πέρα η απόδειξη είναι η ίδια και τελικά καταλήγει στην 4.50 για $t = m$ και $\alpha = \frac{m}{n}$ ή αλλιώς στην 4.74.

Μπορούμε τότε να σπάσουμε το άθροισμα των πιθανοτήτων ανάλογα με το εάν το $m \leq \alpha n$ ή όχι και να πάρουμε

$$\begin{aligned}\mathbb{P}(S_l > \varepsilon\sqrt{n}, \exists m < l : S_l = 0) &\leq \sum_{m=k}^{l-1} \mathbb{P}(S_m = 0, S_l > \varepsilon\sqrt{n}) && \text{από την 4.73} \\ &= \sum_{m=k}^{\alpha n} \mathbb{P}(S_m = 0, S_l > \varepsilon\sqrt{n}) + \sum_{m=\alpha n+1}^{l-1} \mathbb{P}(S_m = 0, S_l > \varepsilon\sqrt{n}) \\ &\leq \sum_{m=k}^{\alpha n} e^{-mJ(\frac{m}{n}, \lambda)} + \sum_{m=\alpha n+1}^{l-1} \mathbb{P}(S_m = 0, S_l > \varepsilon\sqrt{n}) && \text{από την 4.74} \\ &= \sum_{m=\alpha n+1}^{l-1} \mathbb{P}(S_m = 0, S_l > \varepsilon\sqrt{n}) + o(1) && \text{γιατί το } k \rightarrow \infty\end{aligned}\tag{4.75}$$

Είχαμε δει στην πρόταση 4.16 τη σχέση 4.58 σύμφωνα με την οποία για $l \geq m$ και δεδομένου του S_m ισχύει ότι

$$S_l + (l-m) - S_m \sim \text{Bin}(n-m-S_m, 1-(1-p)^{l-m})$$

Αυτό σημαίνει πως δεδομένου ότι το $S_m = 0$, το $S_l + (l-m) \sim \text{Bin}(n-m, 1-(1-p)^{l-m})$ και άρα αν μια τυχαία μεταβλητή $Y \sim \text{Bin}(n-m, 1-(1-p)^{l-m})$ η πιθανότητα $\mathbb{P}(S_m = 0, S_l > \varepsilon\sqrt{n})$ γράφεται

$$\begin{aligned}\mathbb{P}(S_m = 0, S_l > \varepsilon\sqrt{n}) &= \mathbb{P}(S_l > \varepsilon\sqrt{n} | S_m = 0) \mathbb{P}(S_m = 0) \\ &= \mathbb{P}(Y > l-m + \varepsilon\sqrt{n}) \mathbb{P}(S_m = 0) \\ &\leq \mathbb{P}(Y > l-m + \varepsilon\sqrt{n})\end{aligned}\tag{4.76}$$

Έστω το $\alpha = \zeta_\lambda - \varepsilon$ για $\varepsilon > 0$. Χρησιμοποιούμε την ανισότητα $1 - (1-a)^b \leq ab$ για κάθε $0 \leq a \leq 1$ και $b \geq 1$ ώστε να γράψουμε

$$1 - (1-p)^{l-m} = 1 - \left(1 - \frac{\lambda}{n}\right)^{l-m} \leq \frac{\lambda(l-m)}{n}\tag{4.77}$$

Με το $m \geq \alpha n = (\zeta_\lambda - \varepsilon)n$ και άρα $n-m \leq n(1-\zeta_\lambda + \varepsilon)$, για την αναμενόμενη τιμή της Y θα έχουμε ότι

$$\mathbb{E}(Y) = (n-m)(1-(1-p)^{l-m}) \leq (1-\zeta_\lambda + \varepsilon)\lambda(l-m)\tag{4.78}$$

Καθώς όμως το $\lambda > 1$ θα ισχύει ότι $\lambda(1-\zeta_\lambda) = \lambda e^{-\lambda\zeta_\lambda} < 1$ λόγω της συζήτησης κάτω από τη σχέση 4.45, και μπορούμε να πάρουμε το ε να είναι όσο μικρό χρειάζεται ώστε $\lambda(1-\zeta_\lambda + \varepsilon) < 1 - \varepsilon$ και άρα

$$\mathbb{E}(Y) \leq (1-\varepsilon)(l-m)\tag{4.79}$$

Δηλαδή η σχέση 4.76 γίνεται

$$\mathbb{P}(S_m = 0, S_l > \varepsilon\sqrt{n}) \leq \mathbb{P}[Y - \mathbb{E}(Y) > \varepsilon(l - m + \sqrt{n})] \quad (4.80)$$

Πριν συνεχίσουμε, θα χρειαστεί να αποδείξουμε και μια ανισότητα.

Πρόταση 4.18. Έστω ότι τα $X_i, i = 1, \dots, n$ είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν κατανομές Bernoulli, $X_i \sim \text{Be}(p_i)$. Γράφοντας $X = \sum_{i=1}^n X_i$ και $\lambda = \mathbb{E}(X) = \sum_{i=1}^n p_i$, θα ισχύει πως

$$\mathbb{P}(X \geq \mathbb{E}(X) + t) \leq \exp\left(-\frac{t^2}{2(\lambda + \frac{t}{3})}\right) \quad t \geq 0 \quad (4.81)$$

Απόδειξη. Έστω η τυχαία μεταβλητή $Y \sim \text{Bin}(n, \frac{\lambda}{n})$. Τότε αφού η συνάρτηση $g(x) = \log x$ είναι κοίλη, από την ανισότητα του Jensen για κοίλες συναρτήσεις παίρνουμε ότι για $x_1, x_2, \dots, x_n > 0$ ισχύει ότι

$$\sum_{i=1}^n \frac{\log x_i}{n} \leq \log\left(\frac{1}{n} \sum_{i=1}^n x_i\right) \quad (4.82)$$

Άρα για κάθε πραγματικό u θα έχουμε

$$\begin{aligned} \mathbb{E}(e^{uX}) &= \prod_{i=1}^n (1 + (e^u - 1)p_i) && \text{εξαιτίας της ανεξαρτησίας και της Bernoulli} \\ &= \exp\left(n \sum_{i=1}^n \frac{1}{n} \log(1 + (e^u - 1)p_i)\right) \\ &\leq \exp\left(n \cdot \log\left(1 + \frac{(e^u - 1)\lambda}{n}\right)\right) && \text{από την 4.82} \\ &= \left(1 + \frac{(e^u - 1)\lambda}{n}\right)^n \\ &= \mathbb{E}(e^{uY}) \end{aligned} \quad (4.83)$$

Επιπλέον, η ανισότητα Markov μας δίνει

$$\begin{aligned} \mathbb{P}(X \geq \mathbb{E}(X) + t) &\leq e^{-u(\mathbb{E}(X)+t)} \mathbb{E}(e^{uX}) \\ &\leq e^{-u(\mathbb{E}(X)+t)} \mathbb{E}(e^{uY}) && \text{από την 4.83} \\ &= e^{-u(\lambda+t)} (1 + (e^u - 1)p)^n && \text{γιατί } \frac{\lambda}{n} = p \text{ και } \mathbb{E}(X) = \lambda \end{aligned} \quad (4.84)$$

Όταν το $t > n - \lambda$, φυσικά το $\mathbb{P}(X \geq \mathbb{E}(X) + t) = \mathbb{P}(X \geq \lambda + t) = 0$ και άρα ικανοποιείται η 4.81. Διαφορετικά, ψάχνουμε να βρούμε το ελάχιστο του δεξιού μέλους της ανισότητας 4.84. Παραγωγίζοντας θα βρούμε πως θα πρέπει το $e^u = \frac{(\lambda+t)(1-p)}{p(n-t-\lambda)}$. Αντικαθιστώντας στην 4.84 και αλλάζοντας τα p με $\frac{\lambda}{n}$ παίρνουμε τελικά πως

$$\mathbb{P}(X \geq \mathbb{E}(X) + t) \leq \left(\frac{\lambda}{\lambda+t}\right)^{\lambda+t} \left(\frac{n-\lambda}{n-t-\lambda}\right)^{n-t-\lambda} \quad (4.85)$$

για $0 \leq t \leq n - \lambda$. Εισάγοντας τη συνάρτηση $\varphi(x) = (1+x)\log(1+x) - x$ για $x \geq -1$ η 4.86 γράφεται

$$\mathbb{P}(X \geq \mathbb{E}(X) + t) \leq \exp\left(-\lambda\varphi\left(\frac{t}{\lambda}\right) - (n-\lambda)\varphi\left(-\frac{t}{n-\lambda}\right)\right) \quad (4.86)$$

Εφόσον η $\varphi(x)$ είναι μη-αρνητική για κάθε x που μπορεί να προκύψει για $0 \leq t \leq n - \lambda$, μπορούμε να παρακάμψουμε το δεύτερο όρο για να πάρουμε ένα "χειρότερο" αποτέλεσμα. Επίσης, παρατηρούμε ότι το

$$\varphi(x) \geq \frac{x^2}{2(1+x/3)}$$

και άρα πως

$$\begin{aligned} \mathbb{P}(X \geq \mathbb{E}(X) + t) &\leq \exp\left(-\lambda\varphi\left(\frac{t}{\lambda}\right)\right) \\ &\leq \exp\left(-\lambda\frac{\left(\frac{t}{\lambda}\right)^2}{2(1+t/(3\lambda))}\right) \\ &\leq \exp\left(-\frac{t^2}{2(\lambda + \frac{t}{3})}\right) \end{aligned}$$

□

Παίρνοντας στην παραπάνω πρόταση το $t = \varepsilon(l - m + \sqrt{n})$ και λαμβάνοντας υπόψιν την 4.79 έχουμε ότι η 4.80 δίνει

$$\mathbb{P}(S_m = 0, S_l > \varepsilon\sqrt{n}) \leq \exp\left(-\frac{t^2}{2((1-\varepsilon)(l-m) + \frac{t}{3})}\right) \quad (4.87)$$

Χωρίζουμε αυτήν την πιθανότητα ανάλογα με το εάν το $l - m \geq \varepsilon\sqrt{n}$ ή όχι.

Για $l - m \geq \varepsilon\sqrt{n}$ και αφού το $t \geq \varepsilon(l - m)$ και το $(1 - \varepsilon) + \frac{\varepsilon}{3} \leq 1$ έχουμε πως

$$\begin{aligned} \mathbb{P}(S_m = 0, S_l > \varepsilon\sqrt{n}) &\leq \exp\left(-\frac{t}{2\left(\frac{(1-\varepsilon)(l-m)}{t} + \frac{1}{3}\right)}\right) \\ &\leq \exp\left(-\frac{t}{2\left(\frac{1-\varepsilon}{\varepsilon} + \frac{1}{3}\right)}\right) \\ &\leq \exp\left(-\frac{\varepsilon t}{2}\right) \\ &\leq \exp\left(-\frac{\varepsilon^2(l-m)}{2}\right) \\ &\leq \exp\left(-\frac{\varepsilon^3\sqrt{n}}{2}\right) \end{aligned} \quad (4.88)$$

Στην περίπτωση που το $l - m < \varepsilon\sqrt{n}$ και αφού το $t \geq \varepsilon\sqrt{n}$ έχουμε πως

$$\begin{aligned} \mathbb{P}(S_m = 0, S_l > \varepsilon\sqrt{n}) &\leq \exp\left(-\frac{t^2}{2((1-\varepsilon)\varepsilon\sqrt{n} + \frac{t}{3})}\right) \\ &\leq \exp\left(-\frac{t^2}{2((1-\varepsilon)t + \frac{t}{3})}\right) \\ &\leq \exp\left(-\frac{3t}{8}\right) \\ &\leq \exp\left(-\frac{3\varepsilon\sqrt{n}}{8}\right) \end{aligned} \quad (4.89)$$

Από τις σχέσεις 4.88 και 4.89, η 4.75 θα μας δώσει ότι

$$\mathbb{P}(S_l > \varepsilon\sqrt{n}, \exists m < l : S_l = 0) \leq n \left(\exp\left(-\frac{\varepsilon^2\sqrt{n}}{2}\right) + \exp\left(-\frac{3\varepsilon\sqrt{n}}{8}\right) \right) + o(1) = o(1) \quad (4.90)$$

Καταφέραμε δηλαδή να δείξουμε το στόχο μας: από τις σχέσεις 4.71 και 4.72 η πιθανότητα

$$\begin{aligned} \mathbb{P}(S_m > 0, \forall m \leq l) &\geq \mathbb{P}(S_l > \varepsilon\sqrt{n}) - \mathbb{P}(S_l > \varepsilon\sqrt{n}, \exists m < l : S_l = 0) \\ &= \mathbb{P}\left(Z' > x + \frac{\varepsilon}{1 - \lambda e^{-\lambda \zeta_\lambda}}\right) + o(1) \\ &\rightarrow \mathbb{P}(Z' > x) \end{aligned}$$

με τη Z' να είναι μια τυχαία μεταβλητή που ακολουθεί κανονική κατανομή με μέσο όρο το 0 και διασπορά τη σ_λ^2 .

Επομένως αποδείχτηκε και το κάτω φράγμα για την πιθανότητα

$$\mathbb{P}\left(\frac{|C_{[k]}| - \zeta_\lambda n}{\sqrt{n}} > x\right)$$

και άρα ισχύει το Κεντρικό Οριακό Θεώρημα 4.15.

4.6 Κρίσιμο διάστημα

Όπως πάντα, βρισκόμαστε σε ένα τυχαίο γράφημα σε n κορυφές και με πιθανότητα $p = \frac{\lambda}{n}$ εμφάνισης κάθε δυνατής ακμής σε αυτό, δηλαδή στο $G(n, \frac{\lambda}{n})$. Με την έννοια "**Κρίσιμο Διάστημα**" (Critical Window) θα αναφερόμαστε στην κρίσιμη τιμή του λ για το μέγεθος της μεγαλύτερης συνιστώσας και στις τιμές κοντά σε αυτή. Όπως έχουμε δει η κρίσιμη τιμή του λ είναι το 1. Στις προηγούμενες ενότητες εξηγήσαμε αναλυτικά ότι εάν το λ είναι μια σταθερά με $\lambda < 1$, τότε η μεγαλύτερη συνεκτική συνιστώσα θα είναι μικρότερη από $K \log n$ με το K να είναι μια σταθερά με μεγάλη πιθανότητα. Οπότε σε συνδυασμό πως με μεγάλη πιθανότητα υπάρχει και μια συνεκτική συνιστώσα μεγέθους μεγαλύτερου από $D \log n$ με το $D < K$ ισχύει σε αυτήν την περίπτωση πως η $|C_{\max}| = \Theta(\log n)$. Στον αντίποδα, όταν το λ είναι και πάλι μια σταθερά αλλά αυτή τη φορά έχουμε ότι το $\lambda > 1$, είδαμε πως η μεγαλύτερη συνιστώσα, η γιγαντιαία, θα έχει μέγεθος ίσο με περίπου $\zeta_\lambda n$, και άρα πως $|C_{\max}| = \Theta(n)$ με τη ζ_λ να είναι η πιθανότητα επιβίωσης σε μια ουρά αναμονής Poisson. Γεννιέται επομένως το ερώτημα τι ισχύει όταν το $\lambda = 1$. Η $|C_{\max}| = \Theta(\log n)$; Μήπως $|C_{\max}| = \Theta(n)$;

Όπως θα δούμε σε αυτή την ενότητα η απάντηση είναι διαφορετική. Η μεγαλύτερη συνεκτική συνιστώσα - *κυρίαρχη* όπως συχνά αποκαλείται - έχει μέγεθος τάξης $\Theta(n^{\frac{2}{3}})$. Είναι δηλαδή μεν σημαντικά μεγαλύτερη από αυτήν όταν το λ είναι μια σταθερά μικρότερη του 1, αλλά ταυτόχρονα και σημαντικά μικρότερη από τη γιγαντιαία συνιστώσα για το $\lambda > 1$. Ίσως όμως το πιο συναρπαστικό γεγονός να είναι πως αυτή η ανάλυση για τη μεγαλύτερη συνιστώσα ισχύει όχι μόνο για το $\lambda = 1$, αλλά και μια περιοχή γύρω από αυτό, εξ ου και η ονομασία **Κρίσιμο Διάστημα**. Συγκεκριμένα, αν παραμετροποιήσουμε την πιθανότητα $p = \frac{\lambda}{n}$, αλλά αφήσουμε το λ να μην είναι μια σταθερά, αλλά να είναι έχει τη μορφή $\lambda = 1 + \theta n^{-\frac{1}{3}}$ με το $\theta \in \mathbb{R}$, τότε το μέγεθος της κυρίαρχης συνιστώσας είναι τάξης $\Theta(n^{\frac{2}{3}})$. Αυτό το αποτέλεσμα δόθηκε για πρώτη φορά από τον Βέλα Bollobás [4] το 1984.

Γιατί όμως χρησιμοποιείται αυτή η συγκεκριμένη παραμετροποίηση; Γιατί το $\lambda = 1 + \theta n^{-\frac{1}{3}}$ και ο εκθέτης να μην είναι ένα άλλο κλάσμα μικρότερο του 0; Σε γενικές γραμμές ο λόγος είναι ο εξής: Έστω πως δεν ξέρουμε ποια είναι η κατάλληλη παραμετροποίηση και γράφουμε πως το $\lambda = 1 + \varepsilon(n)$ με το $\varepsilon = \varepsilon(n) \rightarrow 0$ και $\varepsilon > 0$. Από το θεώρημα 3.9 παίρνουμε πως η πιθανότητα μια ουρά αναμονής Poisson με παράμετρο το λ να ξεπερνάει το $A\varepsilon^{-2}$ με την A να είναι μια θετική σταθερά είναι εκθετικό στην A , καθώς

$$\mathbb{P}(A\varepsilon^{-2} \leq T < \infty) \leq \frac{e^{-A\varepsilon^{-2}I_\lambda}}{1 - e^{-I_\lambda}} = ce^{-\frac{A}{2}(1+o(1))} \rightarrow e^{-\frac{A}{2}}$$

όπου έχει χρησιμοποιηθεί πως το $I_\lambda = \lambda - 1 - \log \lambda$ και το ανάπτυγμα Taylor του $\log(1 + \varepsilon)$ στο 0. Προσέξτε πως κάποια μεγαλύτερη δύναμη του ε θα έδινε ένα τετραμμένο άνω όριο για την πιθανότητα και άρα το $A\varepsilon^{-2}$ είναι το καλύτερο που μπορούμε να πάρουμε. Επίσης, από το πόρισμα 3.3 η πιθανότητα ο συνολικός πληθυσμός να είναι άπειρος είναι

$$\mathbb{P}(T_{po} = \infty) = \zeta_\lambda \simeq 2(\lambda - 1) = 2\varepsilon$$

Για μεγάλο n η ουρά αναμονής Poisson και η διαδικασία γραφήματος είναι αρκετά παρόμοιες (ενότητα 11.6, [1]) και άρα θεωρούμε πως η διαδικασία γραφήματος έχει την ίδια συμπεριφορά με την ουρά αναμονής Poisson. Άρα το μέγεθος της μεγαλύτερης συνεκτικής συνιστώσας του $G(n, \frac{\lambda}{n})$ έχει μέγεθος $\simeq \zeta_\lambda n = 2\varepsilon n$ και η δεύτερη σε μέγεθος συνιστώσα περίπου ε^{-2} . Για να μη διαφέρουν πολύ η πρώτη με τη δεύτερη συνιστώσα θα πρέπει το $\varepsilon = \Theta(n^{-1/3})$. Εάν το $\varepsilon \gg n^{-1/3}$ τότε θα σχηματιζόταν η γιγαντιαία συνιστώσα. Εάν το $\varepsilon \ll n^{-1/3}$ το επιχείρημα καταρρέει καθώς η “δεύτερη” συνιστώσα θα είχε μέγεθος μεγαλύτερο από την “πρώτη”.

Όταν το $\varepsilon(n) \leq 0$ η ουρά αναμονής Poisson δε δίνει άπειρο πληθυσμό με πιθανότητα 1 (θεώρημα 3.1) και έτσι δεν έχουμε κάποιο ειδικό χαρακτηριστικό για τις συνιστώσες (έχουν όλες περίπου το ίδιο μέγεθος ε^{-2}).

Αυτό σημαίνει πως το $\lambda = 1 + \theta n^{-\frac{1}{3}}$ με το θ να είναι σταθερό και $\theta \in \mathbb{R}$.

Για να αποδείξουμε το κεντρικό θεώρημα αυτού του κεφαλαίου, θα χρειαστούμε δύο προτάσεις.

Πρόταση 4.19. Έστω ότι το $\lambda = 1 + \theta n^{-\frac{1}{3}}$, με το $\theta \in \mathbb{R}$ και το $r > 0$. Τότε για $k = rn^{\frac{2}{3}}$ υπάρχουν σταθερές c_1, c_2 όπου $0 < c_1 < c_2$ με τη $c_1 = c_1(r, \theta)$ και τη c_2 να είναι ανεξάρτητη και από το r και από το θ , τέτοιες ώστε για κατάλληλα μεγάλο n να ισχύει πως

$$\frac{c_1}{\sqrt{k}} \leq \mathbb{P}(|C_i| \geq k) \leq c_2 \frac{f(r, \theta)}{\sqrt{k}} \quad (4.91)$$

με τη $|C_i|$ να είναι η συνεκτική συνιστώσα οποιουδήποτε τυχαία επιλεγμένου κόμβου στο τυχαίο γράφημα $G(n, \frac{\lambda}{n})$.

Απόδειξη. Θα αποδείξουμε αρχικά το άνω φράγμα. Από το θεώρημα 4.1 έχουμε ότι

$$\mathbb{P}_\lambda(|C_i| \geq k) \leq \mathbb{P}_{n,p}(T_{\text{bin}} \geq k) \quad (4.92)$$

με την $\mathbb{P}_{n,p}$ να συμβολίζει την πιθανότητα σε μια διωνυμική ουρά αναμονής με παραμέτρους n και $p = \frac{\lambda}{n} = \frac{1+\theta n^{-\frac{1}{3}}}{n}$. Χρησιμοποιώντας και το θεώρημα 3.7 παίρνουμε

$$\mathbb{P}_\lambda(|C_i| \geq k) \leq \mathbb{P}_{p_o(\lambda)}(T_{p_o} \geq k) + e_n(k) \quad (4.93)$$

με την $\mathbb{P}_{p_o(\lambda)}$ να δηλώνει την πιθανότητα σε μια ουρά αναμονής Poisson με μέσο όρο $\lambda = 1 + \theta n^{-\frac{1}{3}}$. Επίσης από το ίδιο θεώρημα είχαμε δει πως

$$|e_n(k)| \leq \frac{\lambda^2}{n} \sum_{s=1}^{k-1} \mathbb{P}_{p_o(\lambda)}(T_{p_o} \geq s) \quad (4.94)$$

Μπορούμε να γράψουμε

$$\begin{aligned}
\mathbb{P}_{\text{po}(\lambda)}(\Gamma_{\text{po}} \geq s) &= \mathbb{P}_{\text{po}(\lambda)}(\Gamma_{\text{po}} = \infty) + \sum_{t=s}^{\infty} \mathbb{P}_{\text{po}(\lambda)}(\Gamma_{\text{po}} = t) \\
&= \zeta_{\lambda} + \sum_{t=s}^{\infty} \mathbb{P}_{\text{po}(\lambda)}(\Gamma_{\text{po}} = t) \\
&\leq 2(\lambda - 1) + \sum_{t=s}^{\infty} \frac{e^{-I_{\lambda} t}}{\lambda \sqrt{2\pi t^3}} \left(1 + \mathcal{O}\left(\frac{1}{t}\right)\right) && \text{από τα 3.3,3.11} \\
&\leq 2(\lambda - 1) + \sum_{t=s}^{\infty} \frac{1}{\lambda \sqrt{2\pi t^3}} \left(1 + \mathcal{O}\left(\frac{1}{t}\right)\right) && \text{αφού το } I_{\lambda} \geq 0 \\
&\leq 2(\lambda - 1) + K_1 \sum_{t=s}^{\infty} \frac{1}{\lambda \sqrt{2\pi t^{\frac{3}{2}}}} && \text{εξ'ορισμού του } \mathcal{O}\left(\frac{1}{t}\right) \\
&\leq 2\theta_+ n^{-1/3} + K_2 \sum_{t=s}^{\infty} \frac{1}{\lambda t^{\frac{3}{2}}} \\
&\leq 2\theta_+ n^{-1/3} + \frac{K_2}{\lambda} \frac{1}{\sqrt{s}} \left[\frac{1}{s^2} + \frac{\sqrt{s}}{\sqrt{s+1}(s+1)^2} + \dots \right] \\
&\leq 2\theta_+ n^{-1/3} + \frac{K_2}{\lambda} \frac{1}{\sqrt{s}} \left[\frac{1}{s^2} + \frac{1}{(s+1)^2} + \dots \right] \\
&\leq 2\theta_+ n^{-1/3} + \frac{K_3}{\lambda \sqrt{s}}
\end{aligned} \tag{4.95}$$

με το $\theta_+ = \max\{\theta, 0\}$. Επισημαίνεται πως το K_3 είναι ανεξάρτητο του θ και του r και πως το $K_2 = \frac{K_1}{\sqrt{2\pi}} > \frac{1}{\sqrt{2\pi}}$. Επίσης, η 4.94 με τη βοήθεια της 4.95 δίνει ότι

$$\begin{aligned}
|e_n(k)| &\leq \frac{\lambda^2}{n} \sum_{s=1}^{k-1} \left(2\theta_+ n^{-1/3} + \frac{K_3}{\lambda \sqrt{s}}\right) \\
&\leq \frac{\lambda^2}{n} \sum_{s=1}^k \left(2\theta_+ n^{-1/3} + \frac{K_3}{\lambda \sqrt{s}}\right) \\
&\leq 2\lambda^2 \theta_+ k n^{-4/3} + \frac{K_3 \lambda k}{n} \\
&\leq 2\lambda^2 \theta_+ r n^{-2/3} + K_3 \lambda r n^{-1/3} \\
&\leq K_4 \left(\lambda^2 \theta_+ r n^{-1/3} + \lambda r n^{-1/3}\right)
\end{aligned} \tag{4.96}$$

Τελικά έχουμε ότι

$$\begin{aligned}
\mathbb{P}_{\lambda}(|C_i| \geq k) &\leq 2\theta_+ n^{-1/3} + \frac{K_3}{\lambda \sqrt{k}} + K_4 \left(\lambda^2 \theta_+ r n^{-1/3} + \lambda r n^{-1/3}\right) \\
&= \frac{2\theta_+ \sqrt{r}}{\sqrt{k}} + \frac{K_3}{\lambda \sqrt{k}} + \frac{K_4 \lambda^2 \theta_+ r^{3/2}}{\sqrt{k}} + \frac{\lambda r^{3/2}}{\sqrt{k}} \\
&\leq K_5 \frac{f(r, \theta)}{\sqrt{k}}
\end{aligned}$$

όπου και πάλι επισημαίνουμε πως το $K_5 > \frac{1}{\sqrt{2\pi}}$. Ας πάμε τώρα να δείξουμε το κάτω φράγμα της επιθυμητής σχέσης 4.91. Μπορούμε να δείξουμε το κάτω φράγμα μόνο για την περίπτωση που το $\theta \leq 0$, καθώς εάν στο γράφημα το λ (δηλαδή το θ) γίνει μεγαλύτερο (και άρα μεγαλώσει το $p = \frac{\lambda}{n}$), τότε το γράφημα θα δώσει περισσότερη πιθανότητα στο ενδεχόμενο το $\{|C_i| \geq k\}$. Αυτή η ιδιότητα λέγεται μονοτονία στα γραφήματα και περιγράφεται στην ενότητα 4.8. Αυτή τη φορά θα χρησιμοποιήσουμε το θεώρημα 4.2

$$\mathbb{P}_\lambda(|C_i| \geq k) \geq \mathbb{P}_{n-k,p}(T_{bin} \geq k) \quad (4.97)$$

όπου το $n - k = n - rn^{\frac{2}{3}}$ και το $p = \frac{\lambda}{n} = \frac{1+\theta n^{-\frac{1}{3}}}{n}$. Σε συνδυασμό με τις σχέσεις από το θεώρημα 3.7 για το $\lambda_n = \frac{(n-k)\lambda}{n} = 1 + (\theta - r)n^{-\frac{1}{3}}$ παίρνουμε ότι

$$\begin{aligned} \mathbb{P}_\lambda(|C_i| \geq k) &\geq \mathbb{P}_{p_o(\lambda_n)}(T_{p_o} \geq k) - e_n(k) \\ &\geq \mathbb{P}_{p_o(\lambda_n)}(T_{p_o} \geq k) - K_4 \left(\lambda^2 \theta_+ r n^{-1/3} + \lambda r n^{-1/3} \right) && \text{από την 4.96} \\ &\geq \sum_{t=k}^{\infty} \mathbb{P}_{p_o(\lambda_n)}(T_{p_o} = t) - K_4 \left(\lambda^2 \theta_+ r n^{-1/3} + \lambda r n^{-1/3} \right) \\ &\geq \sum_{t=k}^{\infty} \frac{e^{-I_{\lambda_n} t}}{\lambda_n \sqrt{2\pi n^3}} \left(1 + \mathcal{O}\left(\frac{1}{n}\right) \right) - K_4 \left(\lambda^2 \theta_+ r n^{-1/3} + \lambda r n^{-1/3} \right) && \text{από την 3.11} \end{aligned} \quad (4.98)$$

Αναπτύσσοντας το $\log \lambda_n$ σε Taylor έχουμε πως

$$I_{\lambda_n} = \lambda_n - 1 - \log \lambda_n = \frac{(\lambda_n - 1)^2}{2} + \mathcal{O}(|\lambda_n - 1|^3) = \frac{(\lambda_n - 1)^2}{2} + o(1) \quad (4.99)$$

αφού το $\lambda_n = \frac{(n-k)\lambda}{n} = 1 + (\theta - r)n^{-\frac{1}{3}}$.

Δηλαδή η 4.98 γίνεται

$$\begin{aligned} \mathbb{P}_\lambda(|C_i| \geq k) &\geq \sum_{t=k}^{\infty} \frac{e^{-I_{\lambda_n} t}}{\lambda_n \sqrt{2\pi t^3}} \left(1 + \mathcal{O}\left(\frac{1}{t}\right) \right) - K_4 \left(\lambda^2 \theta_+ r n^{-1/3} + \lambda r n^{-1/3} \right) \\ &\geq \sum_{t=k}^{2k} \frac{e^{-[(\frac{\lambda_n - 1}{2})^2 + o(1)]t}}{\sqrt{2\pi t^3}} - K_4 \left(\lambda^2 \theta_+ r n^{-1/3} + \lambda r n^{-1/3} \right) && \text{αφού το } \lambda_n < 1 \\ &\geq K_6 \frac{e^{-(\theta-r)^2 n^{-2/3} r n^{2/3}}}{\sqrt{k}} - K_4 \left(\lambda^2 \theta_+ r n^{-1/3} + \lambda r n^{-1/3} \right) && \text{για κατάλληλα μεγάλο } n \\ &\geq \left[K_6 e^{-(\theta-r)^2 r} - K_4 \left(\lambda^2 \theta_+ r^{3/2} + \lambda r^{3/2} \right) \right] \frac{1}{\sqrt{k}} \\ &= \frac{c_1(r, \theta)}{\sqrt{k}} \end{aligned}$$

όπου επισημαίνεται πως το K_6 είναι μικρότερο από το $\frac{1}{\sqrt{2\pi}}$ και άρα και το $c_1(r, \theta) > 0$ είναι μικρότερό του. Μας μένει να δείξουμε ότι το $c_1(r, \theta) > 0$ και ότι το $K_5 > c_1(r, \theta)$. Από τις “επισημάνσεις” μας το δεύτερο έπεται. Για το πρώτο θα πρέπει να έχουμε πως

$$K_6 e^{-(\theta-r)^2 r} - K_4 \left(\lambda^2 \theta_+ r^{3/2} + \lambda r^{3/2} \right) > 0$$

με τα K_4, K_6 να είναι σταθερές θετικές και ανεξάρτητες από τα r, θ . Καθώς πηγαίνουμε το r κοντά στο 0 το $K_6 e^{-(\theta-r)^2 r} \rightarrow K_6 > 0$ και το $K_4 \left(\lambda^2 \theta_+ r^{3/2} + \lambda r^{3/2} \right) \rightarrow 0$. Αυτό σημαίνει πως για κατάλληλα μικρό r το $c_1(r, \theta) > 0$ όπως θέλουμε. □

Επιπλέον μας μένει να μιλήσουμε για μια ακόμα χρήσιμη πρόταση.

Πρόταση 4.20. Έστω ότι στο $\lambda = 1 + \theta n^{-\frac{1}{3}}$ πάρουμε το $\theta < 0$. Τότε για όλα τα $n \in \mathbb{N}$ ισχύει ότι

$$\mathbb{E}(|C_i|) \leq \frac{n^{\frac{1}{3}}}{|\theta|} \quad (4.100)$$

Απόδειξη. Από το θεώρημα 4.1 βλέπουμε πως το T_{bin} κυριαρχεί επί του $|C_i|$, αφού δίνεται μεγαλύτερη πιθανότητα στο ενδεχόμενο $\{T_{\text{bin}} \geq k\}$ απ'ότι στο $\{|C_i| \geq k\}$ για κάθε $k \in \mathbb{N}$. Εφόσον συμβαίνει αυτό, η αναμενόμενη τιμή του $|C_i|$ θα είναι μικρότερη της αντίστοιχης του T_{bin} και άρα από το θεώρημα 3.11 έχουμε τελικά ότι

$$\mathbb{E}_\lambda (|C_i|) \leq \mathbb{E}_{n, \frac{\lambda}{n}} (T_{\text{bin}}) = \frac{1}{1 - n \frac{\lambda}{n}} = \frac{n^{\frac{1}{3}}}{|\theta|} \quad (4.101)$$

□

Είμαστε έτοιμοι πλέον να αποδείξουμε το κεντρικό θεώρημα αυτής της ενότητας.

Θεώρημα 4.21. Έστω ότι το $\lambda = 1 + \theta n^{-\frac{1}{3}}$ με το $\theta \in \mathbb{R}$. Τότε υπάρχει μια σταθερά $b = b(\theta)$ τέτοια ώστε για κάθε $\omega > 1$ να ισχύει

$$\mathbb{P}_{1+\theta n^{-\frac{1}{3}}} \left(\frac{n^{\frac{2}{3}}}{\omega} \leq |C_{\text{max}}| \leq \omega n^{\frac{2}{3}} \right) \geq 1 - \frac{b}{\omega} \quad (4.102)$$

Απόδειξη. Στην περίπτωση όπου το $\omega \leq b$ η 4.102 ισχύει τετριμμένα αφού κάθε πιθανότητα είναι εξ'ορισμού μη-αρνητική. Επίσης, η ίδια σχέση δεν έχει νόημα όταν το $\frac{n^{\frac{2}{3}}}{\omega} \leq 1$ και το $\omega n^{\frac{2}{3}} \geq n$ αφού η μεγαλύτερη συνιστώσα (όπως και κάθε συνιστώσα) θα έχει μέγεθος ανάμεσα σε αυτές τις τιμές. Βλέπουμε λοιπόν από αυτές τις παρατηρήσεις ότι για $n \leq \omega^{\frac{3}{2}}$ ισχύει και πάλι τετριμμένα το θεώρημα.

Θα αρχίσουμε αντιμετωπίζοντας την πιθανότητα η μεγαλύτερη συνιστώσα να έχει μέγεθος μεγαλύτερο από $\omega n^{\frac{2}{3}}$. Έστω

$$Z_{\geq k} = \sum_{i \in [n]} \mathbb{1}_{\{|C_i| \geq k\}}$$

να είναι ο αριθμός των κορυφών που ανήκουν σε συνιστώσες μεγέθους μεγαλύτερου ή ίσου από k . Τότε, τα

$$\{|C_{\text{max}}| \geq k\} = \{Z_{\geq k} \geq k\}$$

καθώς όταν η μεγαλύτερη συνιστώσα έχει τουλάχιστον k κορυφές, το ίδιο θα ισχύει και για τον αριθμό των κορυφών που ανήκουν σε συνιστώσες μεγέθους μεγαλύτερου ή ίσου από k και το αντίστροφο.

Σε συνδυασμό με την ανισότητα του Markov γράφουμε

$$\begin{aligned} \mathbb{P}_{1+\theta n^{-\frac{1}{3}}} \left(|C_{\text{max}}| \geq \omega n^{\frac{2}{3}} \right) &= \mathbb{P}_{1+\theta n^{-\frac{1}{3}}} \left(Z_{\geq \omega n^{\frac{2}{3}}} \geq \omega n^{\frac{2}{3}} \right) \\ &\leq \frac{1}{\omega n^{\frac{2}{3}}} \mathbb{E}_{1+\theta n^{-\frac{1}{3}}} \left(Z_{\geq \omega n^{\frac{2}{3}}} \right) \end{aligned} \quad (4.103)$$

Με τη βοήθεια της πρότασης 4.19 μπορούμε να γράψουμε

$$\begin{aligned} \mathbb{E}_{1+\theta n^{-\frac{1}{3}}} \left(Z_{\geq \omega n^{\frac{2}{3}}} \right) &= n \mathbb{P}_{1+\theta n^{-\frac{1}{3}}} \left(|C_v| \geq \omega n^{\frac{2}{3}} \right) \\ &\leq n c_2 \frac{f(r, \theta)}{\sqrt{\omega n^{\frac{2}{3}}}} \\ &= \frac{K n^{2/3}}{\sqrt{\omega}} \end{aligned} \quad (4.104)$$

και άρα η 4.103 γίνεται

$$\mathbb{P}_{1+\theta n^{-\frac{1}{3}}} \left(|C_{\text{max}}| \geq \omega n^{\frac{2}{3}} \right) \leq \frac{K}{\omega \sqrt{\omega}} \quad (4.105)$$

Για την πιθανότητα n μεγαλύτερη συνιστώσα να είναι μικρότερη από το $\frac{n^{\frac{2}{3}}}{\omega}$ θα κάνουμε και πάλι χρήση της μονοτονίας της $|C_{\max}|$ ως προς το λ . Δηλαδή πως όσο μεγαλώνει το λ , η πιθανότητα $n |C_{\max}|$ να είναι μικρότερη από κάποια ποσότητα μικραίνει. Αυτό σημαίνει πως

$$\mathbb{P}_{1+\theta n^{-\frac{1}{3}}}\left(|C_{\max}| \leq \frac{n^{\frac{2}{3}}}{\omega}\right) \leq \mathbb{P}_{1-\hat{\theta} n^{-\frac{1}{3}}}\left(|C_{\max}| \leq \frac{n^{\frac{2}{3}}}{\omega}\right) \quad (4.106)$$

με το $\hat{\theta} = \max\{|\theta|, 1\}$. Δηλαδή από εδώ και πέρα μπορούμε να θεωρούμε πως δουλεύουμε με ένα $\theta < -1$. Για να προχωρήσουμε χρειάζεται να δούμε πως

$$\{|C_{\max}| < k\} = \{Z_{\geq k} = 0\}$$

και άρα πως χρησιμοποιώντας την ανισότητα του Chebyshev έχουμε

$$\begin{aligned} \mathbb{P}_{1+\theta n^{-\frac{1}{3}}}\left(|C_{\max}| \leq \frac{n^{\frac{2}{3}}}{\omega}\right) &= \mathbb{P}_{1+\theta n^{-\frac{1}{3}}}\left(Z_{\geq \omega^{-1}n^{\frac{2}{3}}} = 0\right) \\ &\leq \frac{\text{Var}_{1+\theta n^{-\frac{1}{3}}}\left(Z_{\geq \omega^{-1}n^{\frac{2}{3}}}\right)}{\mathbb{E}_{1+\theta n^{-\frac{1}{3}}}\left(Z_{\geq \omega^{-1}n^{\frac{2}{3}}}\right)^2} \end{aligned} \quad (4.107)$$

Από την πρόταση 4.19 έχουμε ότι

$$\mathbb{E}_{1+\theta n^{-\frac{1}{3}}}\left(Z_{\geq \omega^{-1}n^{\frac{2}{3}}}\right) = n \mathbb{P}_{1+\theta n^{-\frac{1}{3}}}\left(|C_v| \geq \omega^{-1}n^{\frac{2}{3}}\right) \geq c_1(r, \theta) n^{\frac{2}{3}} \sqrt{\omega} \quad (4.108)$$

Για τη διασπορά θα χρησιμοποιήσουμε το αποτέλεσμα της πρότασης 4.5 για να πάρουμε μαζί με την 4.100 πως

$$\begin{aligned} \text{Var}_{1+\theta n^{-\frac{1}{3}}}\left(Z_{\geq \omega^{-1}n^{\frac{2}{3}}}\right) &\leq n \mathbb{E}_{1+\theta n^{-\frac{1}{3}}}\left(|C_v| \mathbb{1}_{\{|C_v| \geq \omega^{-1}n^{\frac{2}{3}}\}}\right) \\ &\leq n \mathbb{E}_{1+\theta n^{-\frac{1}{3}}}\left(|C_v|\right) \\ &\leq \frac{n^{\frac{4}{3}}}{|\theta|} \\ &\leq n^{\frac{4}{3}} \end{aligned}$$

Τελικά η 4.107 γίνεται

$$\mathbb{P}_{1+\theta n^{-\frac{1}{3}}}\left(|C_{\max}| \leq \frac{n^{\frac{2}{3}}}{\omega}\right) \leq \frac{n^{\frac{4}{3}}}{\left(c_1(r, \theta) n^{\frac{2}{3}} \sqrt{\omega}\right)^2} = \frac{1}{(c_1(r, \theta))^2 \omega} \quad (4.109)$$

Μαζεύοντας όλα τα αποτελέσματα μαζί παίρνουμε ότι

$$\begin{aligned} \mathbb{P}_{1+\theta n^{-\frac{1}{3}}}\left(\frac{n^{\frac{2}{3}}}{\omega} \leq |C_{\max}| \leq \omega n^{\frac{2}{3}}\right) &= 1 - \mathbb{P}_{1+\theta n^{-\frac{1}{3}}}\left(|C_{\max}| < \omega^{-1}n^{\frac{2}{3}}\right) - \mathbb{P}_{1+\theta n^{-\frac{1}{3}}}\left(|C_{\max}| > \omega n^{\frac{2}{3}}\right) \\ &\geq 1 - \frac{1}{(c_1(r, \theta))^2 \omega} - \frac{K}{\omega \sqrt{\omega}} \\ &\geq 1 - \frac{b}{\omega} \end{aligned}$$

για το $b = \frac{1}{(c_1(r, \theta))^2} + K$, αφού το ω (και άρα και το $\sqrt{\omega}$) είναι μεγαλύτερο του 1. \square

4.7 Το τυχαίο γράφημα ακολουθεί την κατανομή Poisson ασυμπτωτικά

Όπως είδαμε και στη σχέση 2.3, ο βαθμός μιας κορυφής του τυχαίου γραφήματος $G(n, p)$ ακολουθεί την κατανομή $\text{Bin}(n-1, p)$. Είναι ευρέως γνωστό ότι παραμετροποιώντας το p ως $p = \frac{\lambda}{n}$, καθώς το $n \rightarrow \infty$ ο βαθμός μιας κορυφής θα ακολουθεί την κατανομή Poisson με μέσο όρο $\frac{\lambda(n-1)}{n} \rightarrow \lambda$. Σε αυτήν την ενότητα θα δείξουμε κάτι ακόμα πιο ισχυρό, συγκεκριμένα πως το ποσοστό των κορυφών που έχουν βαθμό k τείνει στη συνάρτηση μάζας πιθανότητας της Poisson: $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$, δηλαδή πως η ακολουθία βαθμών (degree sequence) του τυχαίου γραφήματος τείνει να ακολουθεί την κατανομή Poisson.

Με

$$P_k^{(n)} = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{D_i = k\}} \quad (4.110)$$

θα συμβολίζουμε το ποσοστό των κορυφών στο $G(n, \frac{\lambda}{n})$ που έχουν βαθμό ίσο με k . Επίσης θα θεωρούμε πως το $\lambda > 0$ είναι σταθερό. Τότε το θεώρημα που θα αποδείξουμε σε αυτήν την ενότητα είναι το ακόλουθο.

Θεώρημα 4.22. Για κάθε ε_n τέτοια ώστε $n\varepsilon_n^2 \rightarrow \infty$ ισχύει πως

$$\mathbb{P} \left(\max_k |P_k^{(n)} - p_k| \geq \varepsilon_n \right) \rightarrow 0 \quad (4.111)$$

Απόδειξη. Όπως είπαμε, ο βαθμός μιας οποιαδήποτε κορυφής θα ακολουθεί τη $\text{Bin}(n-1, \frac{\lambda}{n})$ και άρα έχουμε ότι

$$\mathbb{E} \left(P_k^{(n)} \right) = \mathbb{P}(D_1 = k) = \binom{n-1}{k} \cdot \left(\frac{\lambda}{n} \right)^k \cdot \left(1 - \frac{\lambda}{n} \right)^{n-1-k} \quad (4.112)$$

όπου με D_1 συμβολίζεται ο βαθμός του κόμβου με δείκτη 1, αφού εκ των προτέρων όλοι οι κόμβοι έχουν την ίδια αντιμετώπιση.

Επίσης, μπορούμε να γράψουμε

$$\sum_{k=0}^{\infty} |p_k - \mathbb{P}(D_1 = k)| = \sum_{k=0}^{\infty} |\mathbb{P}(X^* = k) - \mathbb{P}(X_n = k)| \quad (4.113)$$

όπου με X^* συμβολίζουμε μια τυχαία μεταβλητή που ακολουθεί την κατανομή Poisson με παράμετρο το λ και με X_n μια τυχαία μεταβλητή που ακολουθεί τη διωνυμική κατανομή με παραμέτρους $n-1$ και $\frac{\lambda}{n}$. Θα φράξουμε αυτό το άθροισμα χρησιμοποιώντας σύζευξη. Παίρνοντας την $Y_n = X_n + I_n$ με τη I_n να είναι μια δοκιμή Bernoulli με πιθανότητα επιτυχίας $p = \frac{\lambda}{n}$ η οποία είναι ανεξάρτητη της X_n , μπορούμε να γράψουμε

$$\begin{aligned} \sum_{k=0}^{\infty} |\mathbb{P}(X_n = k) - \mathbb{P}(Y_n = k)| &= \sum_{k=0}^{\infty} |\mathbb{P}(X_n = k) - \mathbb{P}(X_n = k, I_n = 0) - \mathbb{P}(X_n = k-1, I_n = 1)| \\ &= \sum_{k=0}^{\infty} |\mathbb{P}(X_n = k) - \mathbb{P}(X_n = k)\mathbb{P}(I_n = 0) - \mathbb{P}(X_n = k-1)\mathbb{P}(I_n = 1)| \\ &= \sum_{k=0}^{\infty} |\mathbb{P}(X_n = k) - \mathbb{P}(X_n = k)(1-p) - \mathbb{P}(X_n = k-1)p| \\ &= \frac{\lambda}{n} \sum_{k=0}^{\infty} |\mathbb{P}(X_n = k) - \mathbb{P}(X_n = k-1)| \\ &\leq \frac{\lambda}{n} \left(\sum_{k=0}^{\infty} |\mathbb{P}(X_n = k)| + \sum_{k=0}^{\infty} |\mathbb{P}(X_n = k-1)| \right) \\ &= \frac{2\lambda}{n} \end{aligned} \quad (4.114)$$

Άρα έχουμε ότι

$$\begin{aligned}
 \sum_{k=0}^{\infty} |\mathbb{P}(X^* = k) - \mathbb{P}(X_n = k)| &= \sum_{k=0}^{\infty} |\mathbb{P}(X^* = k) - \mathbb{P}(Y_n = k) + \mathbb{P}(Y_n = k) - \mathbb{P}(X_n = k)| \\
 &\leq \sum_{k=0}^{\infty} |\mathbb{P}(X^* = k) - \mathbb{P}(Y_n = k)| + \sum_{k=0}^{\infty} |\mathbb{P}(Y_n = k) - \mathbb{P}(X_n = k)| \\
 &\leq \frac{2\lambda + \lambda^2}{n}
 \end{aligned}
 \tag{4.115}$$

από 4.114 και το 3.8

όπου από το θεώρημα 3.8 παίρνουμε πως εάν οι X^*, Y_n είναι αυτές του θεωρήματος, τότε $n \mathbb{P}(X^* \neq Y_n) \leq \frac{\lambda^2}{n}$ και άρα σε συνδυασμό με το γεγονός πως

$$\begin{aligned}
 \mathbb{P}(X^* \neq Y_n) &= \sum_{k=0}^{\infty} (\mathbb{P}(X^* = k) \cdot (1 - \mathbb{P}(Y_n = k))) \\
 &= \sum_{k=0}^{\infty} (\mathbb{P}(X^* = k) - \mathbb{P}(X^* = k)\mathbb{P}(Y_n = k)) \\
 &\geq \sum_{k=0}^{\infty} (\mathbb{P}(X^* = k) - \mathbb{P}(Y_n = k))
 \end{aligned}$$

Επίσης εξαιτίας της συνθήκης $n\varepsilon_n^2 \rightarrow \infty$, θα πρέπει το $\varepsilon_n^2 = \Omega\left(\frac{1}{n}\right)$ και άρα και το $\varepsilon_n = \Omega\left(\frac{1}{n}\right)$, αφού εάν το $\varepsilon_n > 1$ η σχέση ισχύει τετριμμένα και εάν το $\varepsilon_n < 1$ έχουμε πως το $\varepsilon_n > \varepsilon_n^2$.

Καταλήγουμε στο συμπέρασμα πως η σχέση 4.115 δηλώνει πως

$$\sum_{k=0}^{\infty} |\mathbb{P}(X^* = k) - \mathbb{P}(X_n = k)| \leq \frac{\varepsilon_n}{2}
 \tag{4.116}$$

και άρα οι σχέσεις 4.112 και 4.113 δίνουν ότι

$$\sum_{k=0}^{\infty} |p_k - \mathbb{E}(P_k^{(n)})| \leq \frac{\varepsilon_n}{2}
 \tag{4.117}$$

για κατάλληλα μεγάλο n . Οπότε μένει να δείξουμε πως

$$\mathbb{P}\left(\max_k |P_k^{(n)} - \mathbb{E}(P_k^{(n)})| \geq \frac{\varepsilon_n}{2}\right) = o(1)
 \tag{4.118}$$

ώστε

$$\mathbb{P}\left(\max_k |P_k^{(n)} - p_k| \geq \varepsilon_n\right) \rightarrow 0
 \tag{4.119}$$

Από την ανισότητα του Boole έχουμε ότι

$$\mathbb{P}\left(\max_k |P_k^{(n)} - \mathbb{E}(P_k^{(n)})| \geq \frac{\varepsilon_n}{2}\right) \leq \sum_{k=0}^{\infty} \mathbb{P}\left(|P_k^{(n)} - \mathbb{E}(P_k^{(n)})| \geq \frac{\varepsilon_n}{2}\right)
 \tag{4.120}$$

ενώ η ανισότητα του Chebyshev μας δίνει ότι

$$\mathbb{P}\left(|P_k^{(n)} - \mathbb{E}(P_k^{(n)})| \geq \frac{\varepsilon_n}{2}\right) \leq \frac{4}{\varepsilon_n^2} \text{Var}(P_k^{(n)})
 \tag{4.121}$$

Υπολογίζουμε

$$\begin{aligned}
\text{Var}\left(P_k^{(n)}\right) &= \mathbb{E}\left[\left(P_k^{(n)}\right)^2\right] - \left[\mathbb{E}\left(P_k^{(n)}\right)\right]^2 \\
&= \mathbb{E}\left[\frac{\sum_{i,j \in [n]} \mathbb{1}_{\{D_i=D_j=k\}}}{n^2}\right] - \left[\frac{\sum_{i \in [n]} \mathbb{P}(D_i=k)}{n}\right]^2 \\
&= \frac{\mathbb{P}(D_1=k)}{n} + \frac{n(n-1)\mathbb{P}(D_1=D_2=k)}{n^2} - [\mathbb{P}(D_1=k)]^2 \\
&= \frac{1}{n} [\mathbb{P}(D_1=k) - [\mathbb{P}(D_1=k)]^2] + \frac{n-1}{n} [\mathbb{P}(D_1=D_2=k) - [\mathbb{P}(D_1=k)]^2] \tag{4.122}
\end{aligned}$$

όπου με D_1 και D_2 συμβολίζουμε τους βαθμούς με δείκτες 1 και 2 αντίστοιχα χωρίς να υπάρχει κάποια διαφορά στην κατανομή τους σε σχέση με τις υπόλοιπες κορυφές, απλά όταν μιλάμε για τα D_1, D_2 θα εννοούμε πως μιλάμε για τους βαθμούς δύο διαφορετικών κορυφών του γραφήματος.

Θα χρησιμοποιήσουμε πάλι μια τεχνική σύζευξης. Έστω ότι οι X_1, X_2 είναι δύο ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν κατανομή $\text{Bin}(n-2, \frac{\lambda}{n})$ και οι I_1, I_2 είναι δύο ανεξάρτητες δοκιμές Bernoulli με πιθανότητα επιτυχίας $p = \frac{\lambda}{n}$. Τότε η κατανομή του ζεύγους (D_1, D_2) θα είναι η ίδια με αυτήν του $(X_1 + I_1, X_2 + I_1)$, αφού οι X_1, X_2 συμβολίζουν το πλήθος των γειτόνων των κορυφών 1 και 2 ανάμεσα σε όλες τις υπόλοιπες $n-2$ κορυφές και η I_1 είναι η τυχαία μεταβλητή που δηλώνει εάν οι κορυφές 1 και 2 συνδέονται και μεταξύ τους. Επίσης το ζεύγος $(X_1 + I_1, X_2 + I_2)$ μπορεί να ερμηνευτεί ως δύο ανεξάρτητα αντίγραφα των γειτόνων της κορυφής 1, δηλαδή του D_1 . Άρα θα έχουμε ότι

$$\mathbb{P}(D_1 = D_2 = k) = \mathbb{P}[(X_1 + I_1, X_2 + I_1) = (k, k)] \tag{4.123}$$

$$[\mathbb{P}(D_1 = k)]^2 = \mathbb{P}[(X_1 + I_1, X_2 + I_2) = (k, k)] \tag{4.124}$$

και άρα με επιχείρημα παρόμοιο με πριν έχουμε πως το

$$\mathbb{P}(D_1 = D_2 = k) - [\mathbb{P}(D_1 = k)]^2 \leq \mathbb{P}[(X_1 + I_1, X_2 + I_1) = (k, k), (X_1 + I_1, X_2 + I_2) \neq (k, k)] \tag{4.125}$$

Όταν το $(X_1 + I_1, X_2 + I_1) = (k, k)$ και το $(X_1 + I_1, X_2 + I_2) \neq (k, k)$, θα πρέπει αναγκαστικά να ισχύει πως $I_1 \neq I_2$. Αυτό σημαίνει πως όταν το $I_1 = 0$, το $I_2 = 1$ και αντίστροφα όταν το $I_2 = 0$, το $I_1 = 1$. Στην πρώτη περίπτωση το $X_1 = k$ και στη δεύτερη το $X_2 = k - 1$.

$$\begin{aligned}
\mathbb{P}(D_1 = D_2 = k) - [\mathbb{P}(D_1 = k)]^2 &\leq p\mathbb{P}(X_1 = k) + p\mathbb{P}(X_2 = k - 1) \\
&= \frac{\lambda}{n} [\mathbb{P}(X_1 = k) + \mathbb{P}(X_2 = k - 1)] \tag{4.126}
\end{aligned}$$

Άρα από τις σχέσεις 4.122 και 4.126 παίρνουμε πως

$$\text{Var}\left(P_k^{(n)}\right) \leq \frac{1}{n}\mathbb{P}(D_1 = k) + \frac{\lambda}{n} [\mathbb{P}(X_1 = k) + \mathbb{P}(X_2 = k - 1)] \tag{4.127}$$

Δηλαδή η σχέση 4.120 με τη βοήθεια των σχέσεων 4.121 και 4.127 θα μας δώσει

$$\begin{aligned}
\mathbb{P}\left(\max_k |P_k^{(n)} - \mathbb{E}(P_k^{(n)})| \geq \frac{\varepsilon_n}{2}\right) &\leq \frac{4}{\varepsilon_n^2} \sum_{k=0}^{\infty} \left[\frac{1}{n}\mathbb{P}(D_1 = k) + \frac{\lambda}{n} [\mathbb{P}(X_1 = k) + \mathbb{P}(X_2 = k - 1)]\right] \\
&= \frac{4(2\lambda + 1)}{n\varepsilon_n^2} \rightarrow 0 \tag{καθώς το } n\varepsilon_n^2 \rightarrow \infty
\end{aligned}$$

□

4.8 Μονοτονία στο τυχαίο γράφημα

Σε αυτήν την ενότητα θα δώσουμε έναν πιο τυπικό ορισμό της μονοτονίας και θα δείξουμε πως ορισμένες τυχαίες μεταβλητές όπως το μέγεθος της μεγαλύτερης συνεκτικής συνιστώσας στο τυχαίο γράφημα είναι αύξουσες. Υπενθυμίζεται πως κάθε δυνατή ακμή εμφανίζεται στο τυχαίο γράφημα με πιθανότητα p ανεξάρτητα από την εμφάνιση των υπολοίπων.

Ορισμός 4.23. Ένα ενδεχόμενο είναι αύξον όταν εάν το ενδεχόμενο πραγματοποιείται για κάποιες ακμές που εμφανίζονται στο γράφημα, τότε πραγματοποιείται και με την εμφάνιση κάποιων ακόμα ακμών στο γράφημα.

Ορισμός 4.24. Μια τυχαία μεταβλητή X με τιμές στους πραγματικούς αριθμούς αποκαλείται αύξουσα εάν τα ενδεχόμενα $\{X \geq x\}$ είναι αύξοντα για κάθε $x \in \mathbb{R}$.

Τώρα κάνουμε το εξής. Για καθεμία από τις δυνατές ακμές του γραφήματος, θέτουμε μια τυχαία μεταβλητή U_e η οποία ακολουθεί την ομοιόμορφη κατανομή στο $[0,1]$. Έτσι, σύμφωνα με τις U_e βρίσκουμε μια τιμή για κάθε δυνατή ακμή. Στη συνέχεια σταθεροποιούμε ένα p . Τις ακμές των οποίων η τιμή είναι μικρότερη ή ίση από το p τις θεωρούμε να εμφανίζονται στο γράφημα. Για καθεμία ακμή η πιθανότητα να συμβεί αυτό είναι ίση με $\mathbb{P}(U_e \leq p) = p$ και άρα το γράφημα που προκύπτει είναι το $G(n, p)$, αφού οι $U_e, e \in E$ είναι ανεξάρτητες μεταξύ τους. Αν πάρουμε ένα $p' > p$ και αφήσουμε τις τιμές των $U_e, e \in E$ όπως είχαν, τότε το νέο γράφημα θα περιέχει τις ακμές που περιείχε προηγουμένως, αλλά επιπλέον μπορεί και κάποιες ακόμα ακμές να εμφανίστηκαν. Με την παραπάνω σύζευξη φαίνεται πως αυξάνοντας την πιθανότητα p αυξάνεται και ο αριθμός των ακμών στο τυχαίο γράφημα.

Ας πάρουμε για παράδειγμα τον αριθμό των κορυφών της μεγαλύτερης συνιστώσας, $|C_{\max}|$, για να δείξουμε πως είναι μια αύξουσα τυχαία μεταβλητή. Θα πρέπει δηλαδή να δείξουμε πως τα ενδεχόμενα $\{|C_{\max}| \geq x\}$ είναι αύξοντα για κάθε $x \in \{1, 2, \dots, n\}$. Έστω πως το ενδεχόμενο $\{|C_{\max}| \geq x\}$ πραγματοποιείται για κάποιο x . Εάν εμφανιστούν περισσότερες ακμές στο γράφημα, τότε είτε η μεγαλύτερη συνεκτική συνιστώσα θα μεγαλώσει όντας συνδεδεμένη με καινούργιες κορυφές είτε κάποια άλλη συνιστώσα θα έχει μεγαλύτερο μέγεθος πλέον ή οι νέες ακμές δε θα επιφέρουν καμία αλλαγή στο μέγεθος της μεγαλύτερης συνιστώσας. Πάντως όπως και να έχει, το ενδεχόμενο $\{|C_{\max}| \geq x\}$ θα συνεχίσει να πραγματοποιείται και άρα το $|C_{\max}|$ είναι μια αύξουσα τυχαία μεταβλητή.

Το τελευταίο δικαιολογεί και το γεγονός πως

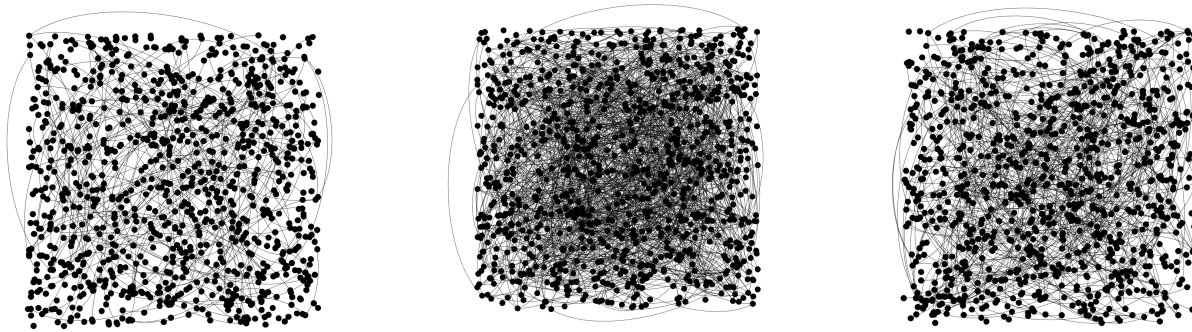
$$\mathbb{P}_p(|C_{\max}| \geq x) \leq \mathbb{P}_{p'}(|C_{\max}| \geq x) \quad (4.128)$$

για $p' > p$, αφού τότε εμφανίζονται περισσότερες ακμές και άρα η πιθανότητα η $\{|C_{\max}| \geq x\}$ αυξάνεται.

4.9 Παράδειγμα τυχαίου γραφήματος

Σε αυτήν την ενότητα θα δώσουμε ένα παράδειγμα του τυχαίου γραφήματος με τη βοήθεια του προγράμματος *Gephi* με το οποίο μπορεί κανείς να δημιουργεί γραφήματα και να παίρνει αποτελέσματα για τις ιδιότητές τους.

Δημιουργήσαμε τρία τυχαία γραφήματα με 1000 κορυφές το καθένα. Στο πρώτο βάλαμε την πιθανότητα εμφάνισης ακμής να είναι $p = \frac{0.5}{1000}$ και άρα πήραμε το $\lambda = 0.5$, στο δεύτερο βάλαμε το $\lambda = 2$ και στο τρίτο οπτικοποιήσαμε την περίπτωση όπου το $\lambda = 1$.



(α) $\lambda=0.5$

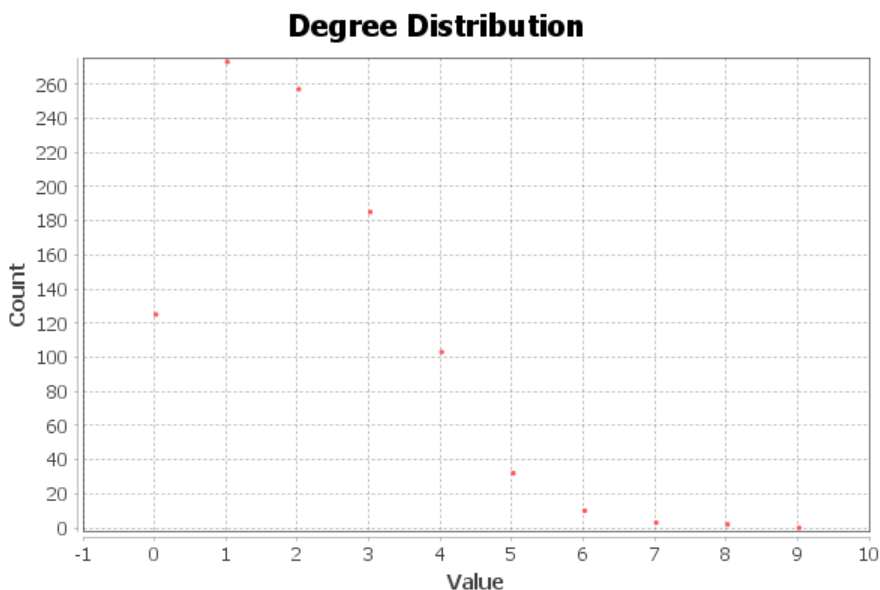
(β) $\lambda=2$

(γ) $\lambda=1$

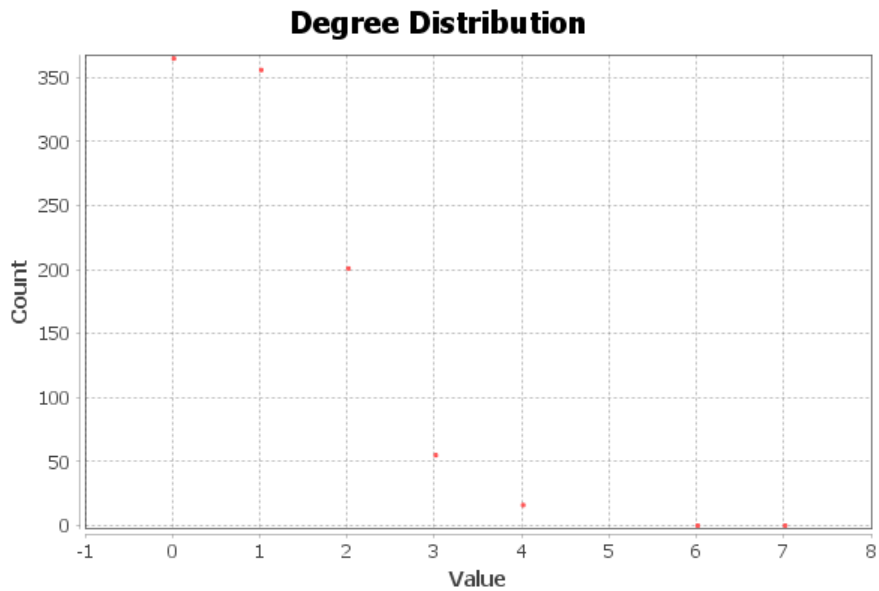
Σχήμα 4.3: Τρία τυχαία γραφήματα με 1000 κόμβους και πιθανότητα $p=\lambda/n$

Παρόλο που τα γραφήματα είναι πολύ μεγάλα για να καταλάβουμε πιθανές ιδιότητες που αφορούν τους βαθμούς και τις συνεκτικές συνιστώσες τους, είναι εμφανές πως το δεύτερο γράφημα είναι αρκετά πιο πυκνό (έχει περισσότερες ακμές) σε σχέση με τα άλλα δύο, γεγονός που περιμέναμε φυσικά αφού το λ είναι αρκετά μεγαλύτερο σε αυτήν την περίπτωση. Επίσης, για τον ίδιο λόγο και το γράφημα με $\lambda = 1$ είναι πιο πυκνό σε σχέση με αυτό όπου το $\lambda = 0.5$.

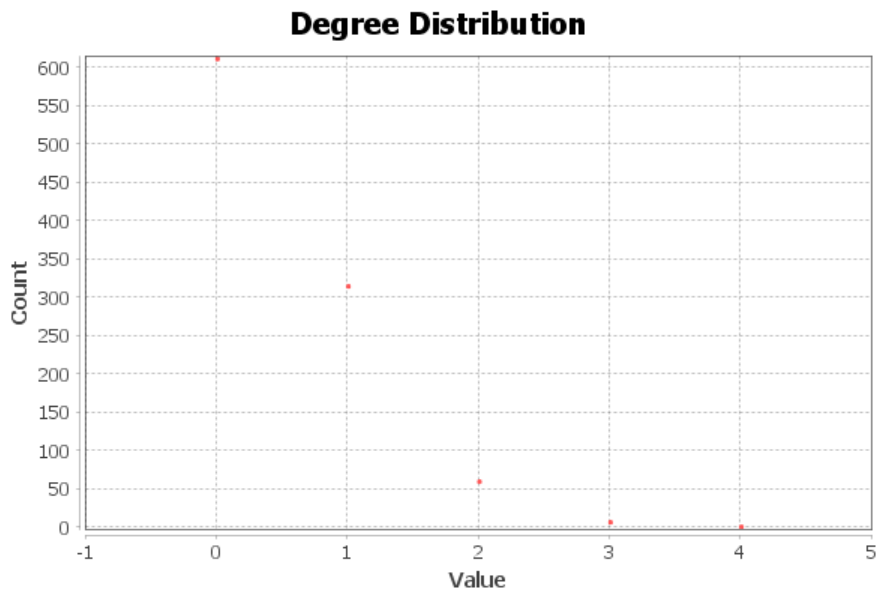
Για την ακρίβεια, το *Gephi* μας δίνει πως οι ακμές στα τρία γραφήματα από το μεγαλύτερο έως το μικρότερο λ είναι 1028, 505 και 235 αντίστοιχα, αριθμοί που συνάδουν με τον αναμενόμενο αριθμό ακμών στο τυχαίο γράφημα (σχέση 2.2) που είναι $\frac{\lambda(n-1)}{2}$. Στα επόμενα σχήματα φαίνονται οι κατανομές των βαθμών στα τρία αυτά γραφήματα.



Σχήμα 4.4: Η κατανομή των βαθμών για ένα τυχαίο γράφημα με $\lambda=2$



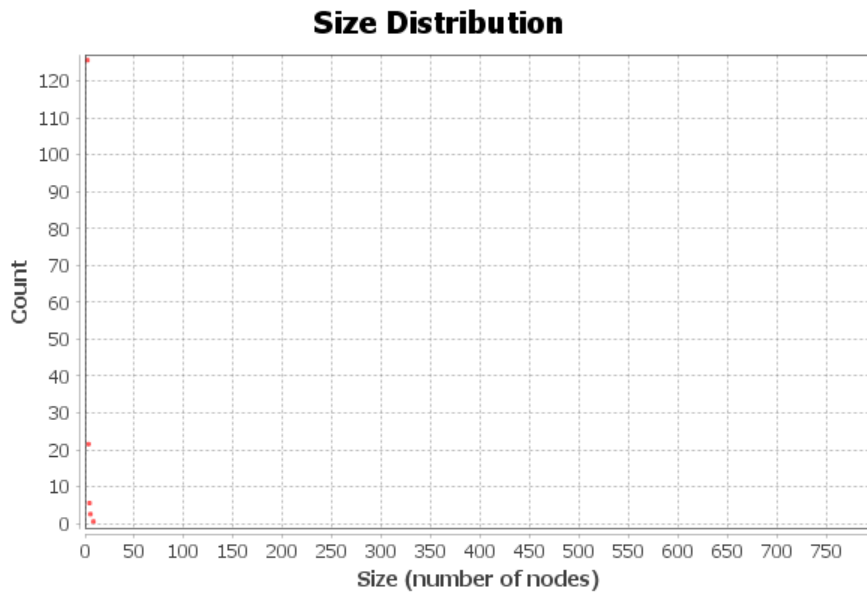
Σχήμα 4.5: Η κατανομή των βαθμών για ένα τυχαίο γράφημα με $\lambda=1$



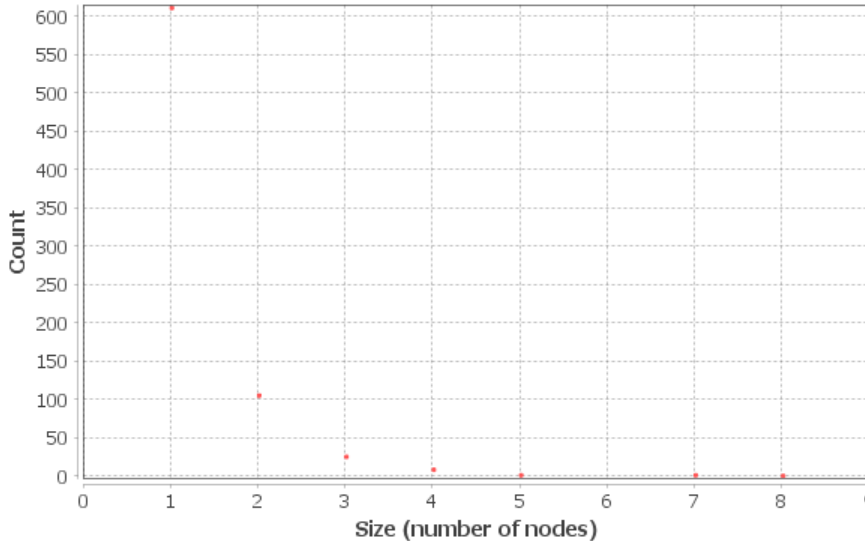
Σχήμα 4.6: Η κατανομή των βαθμών για ένα τυχαίο γράφημα με $\lambda=0.5$

Βλέπουμε πως το εύρος των βαθμών των κορυφών μειώνεται καθώς μειώνεται το λ , ενώ ο αριθμός των κορυφών με βαθμό 0 αυξάνεται.

Μέσω του προγράμματος *Gerhi* μπορούμε να δούμε και τις συνεκτικές συνιστώσες που συναντώνται σε αυτά τα γραφήματα.



Size Distribution



Σχήμα 4.9: Η κατανομή των συνεκτικών συνιστώσων για ένα τυχαίο γράφημα με $\lambda=0.5$

Σε αυτά τα τελευταία σχήματα βλέπουμε ουσιαστικά και τα αποτελέσματα τα οποία αποδείξαμε σε αυτήν την εργασία. Όταν το $\lambda=2$, υπάρχει μια μεγάλη συνιστώσα, η γιγαντιαία, η οποία αποτελείται από περίπου 800 κορυφές. Ταυτόχρονα, η δεύτερη σε μέγεθος συνιστώσα έχει λιγότερες από 20 κορυφές. Επίσης, το ποσοστό των κορυφών που ανήκουν στη γιγαντιαία συνιστώσα είναι 0.8. Αν κοιτάξουμε και το σχήμα 4.1 όπου είχε χρησιμοποιηθεί η γεννήτρια συνάρτηση πιθανότητας της κατανομής Poisson για $\lambda = 2$ θα δούμε πως η πιθανότητα εξαφάνισης είναι περίπου 0.2. Αυτό σημαίνει για την πιθανότητα επιβίωσης πως $\zeta_\lambda = 1 - \eta_\lambda = 0.8$. Άρα και τα πειραματικά αποτελέσματα συμφωνούν με το θεώρημα 4.15 σύμφωνα με το οποίο για μεγάλο n το ποσοστό των κορυφών που ανήκουν στη γιγαντιαία συνιστώσα είναι περίπου ίσο με το ζ_λ .

Στο σχήμα 4.8 βλέπουμε πως η μεγαλύτερη συνιστώσα, η κυρίαρχη, έχει μέγεθος 51 κορυφών, καθώς και πως υπάρχουν άλλες δύο συνιστώσες με το ίδιο περίπου μέγεθος. Παρατηρούμε πάντως πως παρότι το μέγεθος έχει μειωθεί σημαντικά σε σχέση με την περίπτωση όπου το $\lambda = 2$, υπάρχουν ακόμη συνιστώσες σχετικά μεγάλου μεγέθους. Επίσης η τάξη μεγέθους συνάδει με τη θεωρητική που βρήκαμε στην ενότητα 4.6. Εκεί είχαμε δείξει πως η μεγαλύτερη συνιστώσα έχει μέγεθος $Kn^{2/3}$. Το $1000^{2/3} = 100$ και άρα βλέποντας πως η $|C_{\max}| \simeq 0.5 \cdot 100$ επαληθεύεται και αυτό το αποτέλεσμα.

Το μέγεθος της μεγαλύτερης συνιστώσας για $\lambda = 0.5$ βλέπουμε πως έχει συρρικνωθεί και είναι μόλις 8 κορυφές. Με τη σειρά του συνάδει και αυτό το αποτέλεσμα με τα θεωρητικά αποτελέσματα της ενότητας 4.3, αφού είχαμε δείξει πως το μέγεθος της μεγαλύτερης συνιστώσας θα είναι τάξης μεγέθους $\log n$ με το \log να είναι ο φυσικός λογάριθμος και το $\log 1000 \simeq 6.9$.

Επαληθεύονται δηλαδή όλα τα θεωρητικά αποτελέσματα με αυτά τα παραδείγματα. Φυσικά δεν πρέπει να ξεχνάμε πως αυτά τα αποτελέσματα είναι ασυμπτωτικά και ισχύουν για το τυχαίο γράφημα ως ένα χώρο πιθανότητας. Εμείς εδώ πήραμε τρία στιγμιότυπα από τους δειγματικούς χώρους των $G(1000, \frac{2}{1000})$, $G(1000, \frac{1}{1000})$ και $G(1000, \frac{0.5}{1000})$. Γι'αυτό το λόγο περιμένουμε το πείραμα να μας δώσει τιμές κοντά στις θεωρητικές, ωστόσο χωρίς να ανησυχούμε εάν δεν επαληθευτούν αυτές ακριβώς. Εάν τα αποτελέσματά μας ήταν τελείως εκτός των αναμενόμενων, απλά θα ξανατρέχαμε το πείραμα ή/και θα μεγαλώναμε το πλήθος των συνολικών κορυφών.

Κεφάλαιο 5

Επιλεκτική Τοποθέτηση

5.1 Το μοντέλο των Barabási και Albert

Για να ξεπεραστεί το πρόβλημα της κατανομής βαθμών που δημιουργείται στο τυχαίο γράφημα $G(n, p)$ έχουν προταθεί διάφορα μοντέλα, εκ των οποίων ένα είναι το μοντέλο των Barabási-Albert με το οποίο ασχολήθηκαν το 1999 ([3]). Όπως έχουμε πει, τα εμπειρικά δεδομένα από πραγματικά δίκτυα δείχνουν πως τα περισσότερα έχουν κατανομή βαθμών n οποία είναι δεξιά στρεβλή (right-skewed) ή τουλάχιστον n ουρά της κατανομής των βαθμών τους συμπεριφέρεται ως τέτοια. Σε αντίθεση με αυτά, στα γραφήματα που προκύπτουν από το $G(n, p)$ η κατανομή των βαθμών προσεγγίζει την κατανομή Poisson καθώς το μέγεθος του γραφήματος μεγαλώνει. Αυτό σημαίνει πως, όπως είδαμε στην ενότητα 4.7, στο τυχαίο γράφημα το ποσοστό των κορυφών που έχουν ένα συγκεκριμένο βαθμό πέφτει με εκθετικό ρυθμό ως προς το βαθμό, ενώ στα περισσότερα δίκτυα της πραγματικότητας αυτό φαίνεται να συμβαίνει με πολυωνυμικό ρυθμό.

Το μοντέλο των Barabási-Albert καταφέρνει και περιλαμβάνει το φαινόμενο της κατανομής των βαθμών που παρατηρείται πολύ συχνά στα πραγματικά δίκτυα. Ως μοντέλο, βασίζεται ουσιαστικά στο σκεπτικό της **επιλεκτικής σύνδεσης (preferential attachment)**. Σύμφωνα με αυτό, βρισκόμαστε σε ένα γράφημα το οποίο αρχικά έχει ένα συνεκτικό σύνολο r κορυφών και στη συνέχεια μεγαλώνει κάθε χρονική στιγμή κατά μία κορυφή. Τη στιγμή που προσθέτουμε αυτήν την κορυφή, τη συνδέουμε με m από τις προηγούμενες κορυφές. Όμως αντί να επιλέγονται ομοιόμορφα οι m γείτονες της νέας κορυφής ανάμεσα σε όλους τους υποψηφίους, υπάρχει μια μεροληψία υπέρ των καλύτερα συνδεδεμένων από αυτούς. Δηλαδή, οι κορυφές οι οποίες έχουν μεγαλύτερο βαθμό έχουν μεγαλύτερη πιθανότητα να επιλεγούν ως γείτονες της νέας κορυφής απ'ότι οι υπόλοιπες. Μάλιστα, το "μέγεθος" αυτής της μεροληψίας είναι ανάλογο με το βαθμό που έχουν έως τότε οι κορυφές. Εάν μια κορυφή i έχει βαθμό d_i , τότε η πιθανότητα να επιλεγεί για να γειτονεύσει με τη νέα κορυφή του γραφήματος είναι

$$\frac{d_i}{\sum_{v \in V} d_v} \quad (5.1)$$

Όλη η σημασία και η ιδιαιτερότητα αυτού του μοντέλου έγκειται στον τρόπο με τον επιλέγονται οι γείτονες της νέας κορυφής. Δίνοντας ένα προβάδισμα στις κορυφές οι οποίες είναι ήδη αρκετά συνδεδεμένες σχηματίζεται μια "ελίτ" κορυφών οι οποίες θα έχουν σημαντικά μεγαλύτερο βαθμό σε σχέση με την πλειοψηφία των κορυφών που απαρτίζουν το γράφημα. Αυτή η συμπεριφορά ενδεχομένως να είναι και η αλήθεια για αρκετά από τα δίκτυα στην πραγματική ζωή. Υπάρχουν κάποιοι άνθρωποι π.χ. οι οποίοι είναι πολύ πιο δημοφιλείς από τους υπολοίπους και έχουν περισσότερες διασυνδέσεις από ότι η πλειοψηφία. Έτσι, για παράδειγμα, όταν ένας νέος φοιτητής επισκέπτεται το πανεπιστήμιο στο οποίο θα φοιτήσει, η πιθανότητα να γνωριστεί με το "διάσημο" πρόσωπο της σχολής είναι μεγαλύτερη από την πιθανότητα να γνωριστεί με κάποιο άτομο χωρίς πολλές διασυνδέσεις.

Φυσικά, αυτό το μοντέλο παραβλέπει άλλα γνωρίσματα των δικτύων όπως το ότι μια νέα κορυφή στο δίκτυο μπορεί να συνδέεται με μεγαλύτερη πιθανότητα με τις κορυφές στις οποίες με κάποιον τρόπο μοιάζει (ομοφιλία). Για αυτό το λόγο, πρέπει να θυμόμαστε πως το γεγονός πως τα γραφήματα που παράγονται από αυτό το μοντέλο μεγαλώνουν (κάθε φορά κατά μια κορυφή και m ακμές) σε συνδυασμό με το ότι η κατανομή των βαθμών τους είναι δεξιά στρεβλή καθώς τα γραφήματα γίνονται πολύ μεγάλα, είναι απλά μια εξήγηση για τον ενδεχόμενο μηχανισμό από τον οποίο προκύπτει η κατανομή βαθμών που παρατηρείται συχνά στα πραγματικά δίκτυα και όχι απόδειξη πως αυτή είναι η συμπεριφορά που προκαλεί την προκειμένη κατανομή.

5.2 Μια νέα θεώρηση της επιλεκτικής σύνδεσης

Ωστόσο το μοντέλο που χρησιμοποίησαν ο Barabási και η Albert έχει κάποια προβλήματα τα οποία για πρώτη φορά λύθηκαν με αποτελεσματικό τρόπο το 2001 από τους Bollobás κ.ά. [6].

Συγκεκριμένα, οι Barabási-Albert [3] διατυπώνουν πως:

...ξεκινώντας από ένα μικρό αριθμό (m_0) κορυφών, σε κάθε χρονική στιγμή προσθέτουμε μια κορυφή με $m(\leq m_0)$ ακμές προς διαφορετικές προϋπάρχουσες κορυφές στο σύστημα. Για να ενσωματώσουμε την επιλεκτική σύνδεση, υποθέτουμε πως η πιθανότητα Π ώστε η νέα κορυφή να συνδεθεί με μια κορυφή i εξαρτάται από το βαθμό της k_i , έτσι ώστε $\Pi(k_i) = k_i / \sum_j k_j$. Ύστερα από t στιγμές το μοντέλο οδηγεί σε ένα τυχαίο δίκτυο με $t + m_0$ κορυφές και mt ακμές...

Σύμφωνα με πειράματα που διεξήγαν σε H/Y, για το ποσοστό των κορυφών που έχουν βαθμό d ισχύει πως $P(d) \propto d^{-\gamma}$ με το $\gamma = 2.9 \pm 0.1$. Στη συνέχεια, χρησιμοποιώντας ένα ευρετικό επιχείρημα, οι Barabási-Albert προτείνουν πως το $\gamma = 3$.

Παρόλα αυτά, υπάρχουν τρία προβλήματα στο μοντέλο των Barabási-Albert τα οποία διατυπώνουν οι Bollobás κ.ά. [6]:

- Εφόσον οι m_0 αρχικές κορυφές δεν έχουν ακμές, δεν είναι ξεκάθαρο πώς ξεκινάει το μοντέλο.
- Ο αναμενόμενος αριθμός των ακμών που συνδέουν την καινούργια κορυφή v σε προηγούμενες κορυφές είναι $\sum_i \Pi(k_i) = 1$, όχι m .
- Διαλέγοντας το σύνολο S με τις m κορυφές για τη σύνδεση με την καινούργια κορυφή μονομιάς, δεν προσδιορίζεται η κατανομή του S , αφού γνωρίζουμε μόνο τις περιθώριες πιθανότητες για την επιλογή κάθε κορυφής ξεχωριστά.

Για αυτούς τους λόγους προτάθηκε στην ίδια δημοσίευση ([6]) ένα μοντέλο λίγο διαφορετικό το οποίο όμως αντιμετωπίζει αυτά τα εμπόδια με επιτυχία, ενώ καταφέρνει να δείξει πως για το μοντέλο ασυμπτωτικά ισχύει ότι $P(d) \propto d^{-3}$.

Στις επόμενες υποενοότητες δίνεται μια εισαγωγή του μοντέλου και του τρόπου λειτουργίας του, καθώς και το ζητούμενο θεώρημα για την ασυμπτωτική κατανομή των βαθμών του. Όλα τα αποτελέσματα βρίσκονται στο άρθρο [6] και επιπλέον έχει δοθεί μια έμφαση στην όσο το δυνατό διασαφήνιση κάποιων εννοιών που χρησιμοποιούνται σε αυτό. Υπάρχει επίσης μια διαισθητική, αλλά πειστική μεταχείριση του ποσοστού των κορυφών που έχουν βαθμό d στο κεφάλαιο 13 στο [12].

5.2.1 Το εναλλακτικό μοντέλο

Για να ξεπεραστούν τα παραπάνω προβλήματα, η επιλογή των m κορυφών που συνδέονται με τη νέα κορυφή θα γίνεται μία-μία. Έτσι, σε αυτό το μοντέλο είναι πιθανόν να υπάρχουν και πολλαπλές ακμές μεταξύ κορυφών καθώς και αυτοπαθείς βρόχοι. Σημειώνουμε εδώ πως αυτή η συμπεριφορά μπορεί να παρατηρηθεί συχνά σε δίκτυα όπως το διαδίκτυο όπου κάποια σελίδα μπορεί να έχει έναν σύνδεσμο προς τον εαυτό της ή πολλούς συνδέσμους προς μια άλλη ιστοσελίδα. Πάντως, όπως θα δούμε, όσο περνάει ο χρόνος θα γίνεται ακόμη πιο σπάνια η εμφάνιση πολλαπλών ακμών και αυτοπαθών βρόχων, οπότε η ύπαρξή τους δεν επηρεάζει σημαντικά το μοντέλο. Επίσης, μπορούμε να θεωρούμε πως το μοντέλο παράγει ένα κατευθυνόμενο γράφημα του οποίου κάθε κορυφή έχει εξωτερικό βαθμό ίσο με m , ενώ με $d_G(v)$ θα συμβολίζουμε το συνολικό βαθμό της κορυφής v .

Αρχικά θα μιλήσουμε για την περίπτωση όπου $n = m = 1$. Με v_i θα συμβολίζουμε την κορυφή που δημιουργήθηκε από το μοντέλο τη στιγμή i . Η διαδικασία από τα τυχαία γραφήματα $(G_1^t)_{t \geq 0}$ (με το t να συμβολίζει τη χρονική στιγμή στην οποία βρίσκεται το μοντέλο και το $1 = m$ τις ακμές που δημιουργούνται με την προσθήκη κάθε κορυφής) που παράγονται από αυτό το μοντέλο ορίζεται ως εξής: αρχικά έχουμε το γράφημα G_1^1 το οποίο αποτελείται από 1 κορυφή και έναν αυτοπαθή βρόχο. Δεδομένου του G_1^{t-1} , το G_1^t φτιάχνεται προσθέτοντας μια κορυφή v_t στο G_1^{t-1} και συνδέοντας μια ακμή με κατεύθυνση από την v_t στη v_i , με την v_i να επιλέγεται τυχαία με πιθανότητα

$$\mathbb{P}(i = s) = \begin{cases} \frac{d_{G_1^{t-1}}(v_s)}{2t-1} & , 1 \leq s \leq t-1 \\ \frac{1}{2t-1} & , s=t \end{cases} \quad (5.2)$$

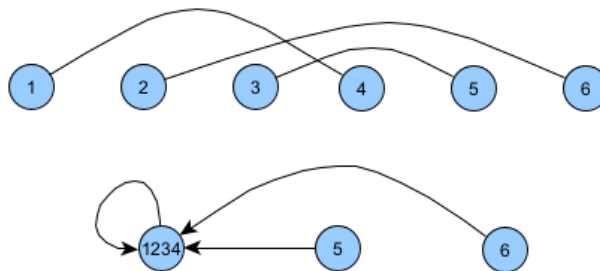
Δηλαδή δίνουμε σε κάθε κορυφή πιθανότητα ανάλογη ως προς το συνολικό βαθμό της να επιλεγεί, καθώς και πιθανότητα που είναι ανάλογη με τον εξωτερικό βαθμό της κορυφής v_t ώστε να είναι βρόχος n νέα ακμή.

Για τη γενική περίπτωση όπου οι νέες ακμές $m > 1$ το μοντέλο ορίζεται μέσω του $(G_1^t)_{t \geq 0}$. Τρέχουμε το $(G_1^{t'})_{t' \geq 0}$, για $t' = mt$ από το οποίο παράγονται οι κορυφές $v'_1, v'_2, \dots, v'_m, v'_{m+1}, \dots, v'_{tm}$ και ακμές μεταξύ τους. Για να βγάλουμε το G_m^t συγχωνεύουμε κατά σειρά τις κορυφές ανά m -άδες σε μία, δηλαδή τις v'_1, v'_2, \dots, v'_m τις κάνουμε μία, την v_1 , τις v'_{m+1}, \dots, v'_{2m} τις κάνουμε την v_2 , κ.ο.κ. Για τις ακμές, εάν υπήρχε ακμή ανάμεσα στις κορυφές v'_i και v'_j , ενώνουμε τώρα τις κορυφές στις οποίες συγχωνεύθηκαν οι v'_i και v'_j . Έτσι, και στο G_m^t κάθε κορυφή v_i , $i < t$ έχει πιθανότητα να επιλεγεί ως γείτονας της νέας κορυφής v_t μέσω της ακμής e , $1 \leq e \leq m$, n οποία είναι ανάλογη με το βαθμό της έως εκείνο το σημείο. Όταν μιλάμε για ιδιότητες του G_m^n , θα μιλάμε για ένα γράφημα το οποίο έχει επιλεγεί τυχαία με ομοιόμορφο τρόπο από το χώρο πιθανότητας \mathcal{G}_m^n ο οποίος περιέχει όλα τα γραφήματα τα οποία γίνεται να δημιουργηθούν μέσω της διαδικασίας $(G_m^t)_{t \geq 0}$.

Εφόσον το G_m^n έχει οριστεί μέσω του G_1^{mn} , θα ασχολούμαστε κυρίως για την περίπτωση όπου το $m = 1$. Ένας έξυπνος και χρήσιμος τρόπος να σκεφτόμαστε το G_1^n είναι με τη βοήθεια των **ζευγαρώματων** (pairings).

Ένα n -ζευγάριωμα \mathcal{P} είναι ένας διαχωρισμός του συνόλου $\{1, 2, \dots, 2n\}$ σε ζευγάρια. Υπάρχουν $(2n)!/(n!2^n)$ διαφορετικά n -ζευγαρώματα. Τα $2n$ στοιχεία μπορούμε να τα ταξινομήσουμε με $(2n)!$ τρόπους. Τα πρώτα δύο στοιχεία αποτελούν ένα ζευγάρι, τα επόμενα δύο άλλο κ.ο.κ. Έχουμε δηλαδή n ζευγάρια. Αυτά τα συγκεκριμένα ζευγάρια έχουμε $n!$ τρόπους να τα πάρουμε. Επίσης για κάθε ένα από τα n ζευγάρια δε μας ενδιαφέρει η σειρά των δύο στοιχείων του, δηλαδή εάν ένα ζεύγος είναι το (i, j) δε θέλουμε να το διπλομετρήσουμε μετρώντας και το (j, i) και άρα για κάθε τέτοιο ζευγάρι πρέπει να διαιρέσουμε με 2. Για αυτό το λόγο πρέπει να διαιρέσουμε το $(2n)!$ με $n!2^n$ ώστε να πάρουμε το πλήθος των διαφορετικών ζευγαριών.

Αν σκεφτούμε τα στοιχεία $1, 2, \dots, 2n$ ως σημεία πάνω σε μια οριζόντια γραμμή και κάθε ζευγάρι να είναι ενωμένο με μια ακμή, τότε μπορούμε να μιλάμε για το αριστερό και το δεξί σημείο κάθε ζευγαριού. Φτιάχνουμε τότε ένα κατευθυνόμενο γράφημα $\varphi(\mathcal{P})$ από ένα n -ζευγάριωμα \mathcal{P} με τον ακόλουθο τρόπο: αρχίζοντας από αριστερά στην οριζόντια γραμμή, συγχωνεύουμε όλα τα σημεία έως και το πρώτο δεξί σημείο που θα βρούμε σε μια κορυφή v_1 . Κάνουμε το ίδιο για τα υπόλοιπα σημεία μέχρι και το δεύτερο δεξί σημείο και δημιουργείται έτσι n κορυφή v_2 . Συνεχίζουμε αυτή τη διαδικασία μέχρι να δημιουργηθεί και n κορυφή v_n . Για κάθε ακμή που υπήρχε ανάμεσα σε δύο σημεία, φτιάχνουμε μια νέα n οποία ενώνει την κορυφή n οποία αντιστοιχεί στο δεξί σημείο με την κορυφή που αντιστοιχεί στο αριστερό σημείο, με κατεύθυνση από τα δεξιά προς τα αριστερά.



Σχήμα 5.1: Ένα παράδειγμα ενός 3-ζευγαρώματος και το γράφημα που παράγεται από αυτό

Θα δείξουμε τώρα πως εάν επιλεγεί ένα n -ζευγάριωμα \mathcal{P} τυχαία με ομοιόμορφο τρόπο από όλα τα $(2n)!/(n!2^n)$ διαφορετικά n -ζευγαρώματα, τότε το γράφημα $\varphi(\mathcal{P})$ που παράγεται με τον τρόπο που περιγράψαμε παραπάνω έχει την ίδια κατανομή με ένα ομοιόμορφα τυχαία επιλεγμένο $G_1^n \in \mathcal{G}_1^n$.

Ένα ζευγάριωμα 2 στοιχείων οδηγεί στο γράφημα με μια κορυφή και έναν αυτοπαθή βρόχο, όπως είναι και η αρχή της διαδικασίας $(G_1^n)_{n \geq 0}$. Εάν έχουμε, ένα τυχαίο $(n-1)$ -ζευγάριωμα και του βάλουμε ένα δεξί

σημείο δεξιότερα από όλα τα έως τότε σημεία και επιλέξουμε στην τύχη τη θέση από τις $2n-1$ δυνατές στην οποία θα βάλουμε το αριστερό σημείο, φτιάχνουμε ένα τυχαίο n -ζευγάρι. Τώρα μια κορυφή βαθμού d στο $\varphi(\mathcal{P})$ αντιστοιχεί σε $d-1$ διαθέσιμα διαστήματα για το νέο αριστερό σημείο, καθώς και ένα ακριβώς στα αριστερά του αριστερότερου σημείου της, δηλαδή συνολικά το νέο αριστερό σημείο έχει $d-1+1=d$ θέσεις όπου εάν τοποθετηθεί, θα δημιουργηθεί μια ακμή ανάμεσα στη νέα κορυφή και στην κορυφή βαθμού d . Άρα η πιθανότητα να επιλεγεί η ακμή της καινούργιας κορυφής στο $\varphi(\mathcal{P})$ να ενώνεται με μια από τις $(n-1)$ προηγούμενες κορυφές είναι η ίδια με αυτή στη σχέση 5.2.

Με τη βοήθεια των n -ζευγαρωμάτων θα γίνεται ο υπολογισμός ιδιοτήτων του G_m^n πιο εύκολος. Είμαστε πλέον έτοιμοι να μιλήσουμε για τους βαθμούς του G_m^n .

5.2.2 Η κατανομή βαθμών στο G_m^n

Συμβολίζοντας με $N_m^n(d)$ τον αριθμό των κορυφών στο G_m^n οι οποίες έχουν εσωτερικό βαθμό d , δηλαδή συνολικό βαθμό $d+m$, σκοπός μας είναι να δείξουμε το ακόλουθο θεώρημα.

Θεώρημα 5.1. Έστω ένα σταθερό $m \geq 1$, και η διαδικασία $(G_m^n)_{n \geq 0}$ όπως ορίστηκε παραπάνω. Εάν

$$\alpha_{m,d} = \frac{2m(m+1)}{(d+m)(d+m+1)(d+m+2)}$$

και πάρουμε οποιοδήποτε $\varepsilon > 0$, τότε με πιθανότητα που τείνει στο 1 καθώς το γράφημα μεγαλώνει ($n \rightarrow \infty$) ισχύει πως

$$(1-\varepsilon)\alpha_{m,d} \leq \frac{N_m^n(d)}{n} \leq (1+\varepsilon)\alpha_{m,d} \quad (5.3)$$

για κάθε d με $0 \leq d \leq n^{1/15}$.

Το θεώρημα αυτό μας δείχνει πως όντως το $P(d) \propto d^{-3}$ καθώς το $n \rightarrow \infty$. Η στρατηγική της απόδειξης του θεωρήματος έχει ως εξής: τα αποτελέσματα για το γενικό μοντέλο με το $m > 1$ θα ακολουθήσουν μέσω των n -ζευγαρωμάτων από την απόδειξη για την περίπτωση όπου το $m = 1$. Πρώτα θα βρούμε επακριβώς την κατανομή του $D_k = \sum_{i < k} d_i$, δηλαδή του αθροίσματος των βαθμών των πρώτων k κορυφών που δημιουργήθηκαν στο μοντέλο. Μετά θα μιλήσουμε για την κατανομή του βαθμού της επόμενης κορυφής v_{k+1} δεδομένου του D_k . Δείχνοντας πως το D_k είναι συγκεντρωμένο γύρω από μια συγκεκριμένη τιμή, βρίσκουμε προσεγγιστικά την πιθανότητα ο $d_{G_1^n}(v_{k+1}) = d+1$ και άρα και την αναμενόμενη τιμή του $N_m^n(d)$. Με ένα λήμμα για martingales, τέλος, αποδεικνύεται και η συγκέντρωση την οποία θέλουμε να δείξουμε στην 5.3.

Γυρίζουμε στη μελέτη της διαδικασίας $(G_1^n)_{n \geq 0}$ όπου με $d_i = d_{G_1^n}(v_i)$ θα συμβολίζουμε το συνολικό βαθμό της i -οστής, κατά σειρά δημιουργίας, κορυφής του μοντέλου. Από ότι θα φανεί και στη συνέχεια είναι χρήσιμο να μιλήσουμε για τα αθροίσματα $D_k = \sum_{i < k} d_i$.

Έστω το ενδεχόμενο $\{D_k - 2k = s\}$, με το $0 \leq s \leq n - k$. Αυτό είναι το ενδεχόμενο σύμφωνα με το οποίο οι τελευταίες $n - k$ κορυφές του G_1^n στέλνουν s ακμές προς τις πρώτες k κορυφές. Είναι δηλαδή σαν να έχουμε ένα n -ζευγάρι \mathcal{P} στο οποίο το k -οστό δεξί σημείο είναι το στοιχείο $2k + s$. Ένα τέτοιο \mathcal{P} μπορούμε να το σπάσουμε σε δύο μερικά ζευγαρώματα, το αριστερό μερικό ζευγάρι \mathcal{L} και το δεξί μερικό ζευγάρι \mathcal{R} με τον ακόλουθο τρόπο. Το \mathcal{L} περιέχει τα στοιχεία $\{1, 2, \dots, 2k + s\}$, ενώ το \mathcal{R} τα $\{2k + s + 1, \dots, 2n\}$. Κάθε ένα από αυτά διατηρεί κάποιον αριθμό από ζεύγη ανάμεσα στα στοιχεία του όπως είχαν καθοριστεί στο \mathcal{P} , ενώ κάποια άλλα στοιχεία (s το πλήθος) είναι πλέον αζευγάριωτα καθώς το αριστερό στοιχείο του ζεύγους ανήκει στο \mathcal{L} και το δεξί στο \mathcal{R} .

Εφόσον, όπως είπαμε, το $2k + s$ είναι δεξί σημείο, θα ζευγαρώνει με ένα από τα στοιχεία $\{1, 2, \dots, 2k + s - 1\}$ και ακριβώς s από τα υπόλοιπα $2k + s - 2$ στοιχεία του \mathcal{L} είναι αζευγάριωτα και άρα τα $2(k-1)$ εναπομείναντα σημεία ζευγαρώνουν μεταξύ τους. Αυτό σημαίνει πως το \mathcal{L} σχηματίζεται με

$$(2k + s - 1) \binom{2k + s - 2}{s} \frac{(2k - 2)!}{2^{k-1}(k-1)!} \quad (5.4)$$

διαφορετικούς τρόπους, με τον τελευταίο όρο να προκύπτει από τον υπολογισμό των συνολικών $2(k-1)$ -ζευγαρωμάτων όπως τον εξηγήσαμε προηγουμένως.

Με παρόμοιο επιχείρημα βρίσκουμε πως το \mathcal{R} έχει

$$\binom{2n-2k-s}{s} \frac{(2n-2k-s)!}{2^{n-k-s}(n-k-s)!} \quad (5.5)$$

διαφορετικούς τρόπους να σχηματιστεί.

Κάθε ένα από τα δυνατά \mathcal{L} μπορεί να συνδυαστεί με ένα από τα \mathcal{R} , με τα αζευγάρωτα σημεία να έχουν $s!$ τρόπους να συνδυαστούν. Άρα πολλαπλασιάζοντας αυτές τις τρεις ποσότητες και διαιρώντας με τον αριθμό όλων των δυνατών $2n$ -ζευγαρωμάτων $\mathcal{P} ((2n)!/(n!2^n))$ παίρνουμε πως

$$\mathbb{P}(D_k - 2k = s) = \frac{(2k+s-1)!(2n-2k-s)!n!2^{s+1}}{s!(k-1)!(n-k-s)!(2n)!} \quad (5.6)$$

για $1 \leq k \leq n$ και $0 \leq s \leq n-k$.

Εάν αυτήν την πιθανότητα τη συμβολίσουμε με $p_s = p_{s,k}$, και επιπλέον ορίσουμε το

$$r_s = \frac{p_{s+1}}{p_s} = 2 \frac{(2k+s)(n-k-s)}{(s+1)(2n-2k-s)} \quad (5.7)$$

μπορούμε να δούμε τα εξής. Αρχικά, το r_s , το πηλίκο δύο “συνεχόμενων” πιθανοτήτων, είναι φθίνουσα συνάρτηση ως προς το s . Επίσης, η μοναδική θετική ρίζα της εξίσωσης $r_s = 1$ δίνεται από τη σχέση

$$s = -2k + \sqrt{4kn - 2n + 1/4} + 1/2$$

Επειδή το s είναι ακέραιος, αυτό σημαίνει πως η τιμή $s_0 = \lceil -2k + \sqrt{4kn - 2n + 1/4} + 1/2 \rceil$ είναι η μία από τις το πολύ δύο τιμές του s με τις οποίες το $\{D_k - 2k = s\}$ συγκεντρώνει τη μεγαλύτερη πιθανότητα να ικανοποιηθεί. Επίσης, για κατάλληλα μεγάλο n ισχύει πως

$$\begin{aligned} \frac{r_{s+1}}{r_s} &= \left(1 - \frac{2k-1}{(s+2)(2k+s)}\right) \left(1 - \frac{n-k}{(2n-2k-s-1)(n-k-s)}\right) \\ &\leq \left(1 - \frac{2k-1}{2n^2}\right) \left(1 - \frac{n-k}{2n^2}\right) \\ &\leq \exp\left(-\frac{2k-1}{2n^2}\right) \exp\left(-\frac{n-k}{2n^2}\right) \\ &\leq \exp\left(-\frac{1}{2n^2}\right) \end{aligned}$$

Αφού το $r_{s_0} \leq 1$, κάνοντας τις ίδιες πράξεις με παραπάνω καταλήγουμε στο ότι $r_{s_0+x} \leq \exp(-x/(2n))$ για $x \in \mathbb{N}$, και άρα από τον ορισμό του r_s και με επαγωγή βρίσκουμε πως $p_{s_0+x} \leq \exp(-x(x-1)/(4n))$. Με μια παρόμοια ανισότητα για το p_{s_0-x} και παίρνοντας τις ακέραιες τιμές κάποιων ποσοτήτων όπου κρίνεται αναγκαίο έχουμε πως

$$\mathbb{P}\left(|D_k - (2k + s_0)| \geq 3\sqrt{n \log n}\right) = o(n^{-1})$$

Μάλιστα επειδή το $|s_0 - (2\sqrt{kn} - 2k)| \leq 2\sqrt{n}$, έχουμε ότι

$$\mathbb{P}\left(|D_k - 2\sqrt{kn}| \geq 4\sqrt{n \log n}\right) = o(n^{-1}) \quad (5.8)$$

Θα δούμε τώρα την πιθανότητα ο συνολικός βαθμός της $(k+1)$ -οστής κορυφής $d_{k+1} = d + 1$, δεδομένου του D_k . Έστω πως το $1 \leq k \leq n-1$ και το $0 \leq s \leq n-k$ και ας θεωρήσουμε ένα αριστερό μερικό ζευγάρι \mathcal{L} όπως πριν. Είδαμε πως ένα τέτοιο αριστερό μερικό ζευγάρι μπορεί να συνδυαστεί με

$$s! \binom{2n-2k-s}{s} \frac{(2n-2k-s)!}{2^{n-k-s}(n-k-s)!}$$

διαφορετικούς τρόπους με δεξιά μερικά ζευγαρώματα ώστε να παραχθεί ένα n -ζευγάρωμα για το οποίο το $D_k - 2k = s$. Για το γράφημα που αντιστοιχεί με την επιλογή ενός από αυτά τα δεξιά μερικά ζευγαρώματα θα ισχύει πως το $d_{k+1} = d + 1$ εάν και μόνο εάν το στοιχείο $2k + s + d + 1$ είναι δεξί σημείο και κάθε ένα από τα στοιχεία $2k + s + 1, \dots, 2k + s + d$ είναι αριστερό σημείο. Αφού επίσης το αριστερό σημείο που ενώνεται με το $2k + s + d + 1$ είναι είτε ένα από τα s αζευγάρωτα στοιχεία στο \mathcal{L} είτε ένα από τα $2k + s + 1, \dots, 2k + s + d$, καθώς και πως ακριβώς $s - 1 + d$ ζευγάρια ξεκινούν πριν το $2k + s + d + 1$ και τελειώνουν στα δεξιά του, κάθε \mathcal{L} έχει

$$(s+d)(s+d-1)! \binom{2n-2k-s-d-1}{s+d-1} \frac{(2(n-k-s-d))!}{(n-k-s-d)!2^{n-k-s-d}}$$

δυνατές κατάλληλες επιλογές για να συνδυαστεί, με τον πρώτο όρο να υποδηλώνει τις επιλογές του αριστερού σημείου του στοιχείου $2k + s + d + 1$, το δεύτερο όρο τις μεταθέσεις των υπολοίπων αριστερών σημείων πριν το $2k + s + d + 1$ τα οποία ενώνονται με δεξί σημείο μετά το $2k + s + d + 1$, τον τρίτο όρο τις επιλογές των δεξιών σημείων μετά το $2k + s + d + 1$ που θα ενωθούν με τα ακριβώς $s + d - 1$ αριστερά σημεία πριν το $2k + s + d + 1$ που έχουν μείνει αζευγάρωτα και τον τέταρτο όρο τα δυνατά ζεύγη που σχηματίζονται ανάμεσα στα σημεία που βρίσκονται μετά το $2k + s + d + 1$ και θα ενωθούν με σημεία που βρίσκονται μετά το $2k + s + d + 1$.

Θυμούμενοι πως το $0 \leq d \leq n - k - s$ και γράφοντας $(a)_b = a!/((a-b)!)$, έχουμε πως

$$\mathbb{P}(d_{k+1} = d + 1 | D_k - 2k = s) = (s+d)2^d \frac{(n-k-s)_d}{(2n-2k-s)_{d+1}} \quad (5.9)$$

Σε αυτό το σημείο να παρατηρήσουμε πως θα μπορούσαμε να συνδυάσουμε τις σχέσεις 5.6 και 5.9, όμως η κατανομή της d_{k+1} που θα παίρναμε δε θα ήταν διαφωτιστική. Χρησιμοποιώντας όμως τις 5.8 και 5.9 μπορούμε να εκτιμήσουμε την αναμενόμενη τιμή των κορυφών με βαθμό $d + 1$, $N_1^d(d)$. Από εδώ και στο εξής οι συμβολισμοί $o()$ και \sim χρησιμοποιούνται για την ασυμπτωτική συμπεριφορά καθώς το $n \rightarrow \infty$. Επίσης, η σταθερά στο συμβολισμό $\mathcal{O}()$ είναι απόλυτη, δηλαδή ανεξάρτητη και από τα k, d .

Έστω το $M = \lfloor n^{4/5}/\log n \rfloor$, το $k = k(n)$ να είναι οποιαδήποτε συνάρτηση που ικανοποιεί τη συνθήκη $M \leq k \leq n - M$ και το $d = d(n)$ να είναι οποιαδήποτε συνάρτηση για την οποία $0 \leq d \leq n^{1/5}$. Τότε, για κάθε D για το οποίο $|D - 2\sqrt{kn}| \leq 4\sqrt{n \log n}$ μπορούμε μέσω της σχέσης 5.9 να γράψουμε την $\mathbb{P}(d_{k+1} = d + 1 | D_k = D)$ ως

$$(2\sqrt{kn} - 2k + \mathcal{O}(\sqrt{n \log n}))2^d \frac{(n+k-2\sqrt{kn} + \mathcal{O}(\sqrt{n \log n}))^d}{(2n-2\sqrt{kn} + \mathcal{O}(\sqrt{n \log n}))^{d+1}}$$

Χρησιμοποιώντας τα όρια $M \leq k \leq n - M$, $0 \leq d \leq n - k - s$ και $0 \leq d \leq n^{1/5}$ μπορούμε να δούμε ότι το κλάσμα του $n+k-2\sqrt{kn} = (\sqrt{n}-\sqrt{k})^2$ προς το $d\sqrt{n \log n}$ τείνει στο άπειρο καθώς το $n \rightarrow \infty$. Το ίδιο ισχύει και για το $(2n-2\sqrt{kn})/(d\sqrt{n \log n})$. Επίσης, το $\sqrt{n \log n} = o(2\sqrt{kn} - 2k)$ και άρα η παραπάνω πιθανότητα γίνεται

$$(1+o(1)) \frac{2\sqrt{kn} - 2k}{2n - 2\sqrt{kn}} \left(\frac{2(\sqrt{n} - \sqrt{k})^2}{2(n - \sqrt{kn})} \right)^d \sim \sqrt{k}(1 - \sqrt{k})^d$$

με το $\kappa = k/n$. Εφόσον αυτή η εκτίμηση εφαρμόζεται στην $\mathbb{P}(d_{k+1} = d + 1 | D_k = D)$ για όλα τα D με $|D - 2\sqrt{kn}| \leq 4\sqrt{n \log n}$, παίρνουμε από τη σχέση 5.8 πως

$$\mathbb{P}(d_{k+1} = o(n^{-1}) + (1+o(1))\sqrt{k}(1 - \sqrt{k})^d$$

Σταθεροποιώντας τώρα το n και αφήνοντας το k να βρίσκεται στο $[M, n - M]$, βρίσκουμε πως ο αναμενόμενος αριθμός των κορυφών v_{k+1} , με το $M \leq k \leq n - M$, οι οποίες έχουν βαθμό $d + 1$ είναι

$$o(1) + \sum_{k=M}^{n-M} (1 + o(1)) \sqrt{k/n} (1 - \sqrt{k/n})^d$$

Αθροίζοντας έτσι για όλες τις περιπτώσεις (μαζί με το $k < M$ και το $k > n - M$) παίρνουμε πως

$$\mathbb{E}(N_1^n(d)) = \mathcal{O}(M) + o(1) + (1 + o(1)) \sum_{k=M}^{n-M} \sqrt{k/n} (1 - \sqrt{k/n})^d \quad (5.10)$$

Γράφοντας $f = \sqrt{\kappa}(1 - \sqrt{\kappa})^d$, έχουμε

$$\frac{1}{f} \frac{df}{d\kappa} = \frac{\kappa^{-1}}{2} - \frac{d}{2} \cdot \frac{\kappa^{-1/2}}{1 - \kappa^{1/2}}$$

Εφόσον τα $n\kappa = k$ και $n(1 - \kappa) = n - k$ τείνουν στο άπειρο η αλλαγή στην f καθώς το κ αλλάζει κατά $1/n$ είναι $o(1)$. Αυτό σημαίνει πως το άθροισμα στη σχέση 5.10 μπορεί να γραφτεί ως

$$(1 + o(1))n \int_{(M+1)/n}^{1-M/n} \sqrt{\kappa}(1 - \sqrt{\kappa})^d d\kappa \sim \int_0^1 \sqrt{\kappa}(1 - \sqrt{\kappa})^d d\kappa \quad (5.11)$$

Χρησιμοποιώντας την αντικατάσταση $\kappa = (1 - u)^2$ και υπολογίζοντας το ολοκλήρωμα, παίρνουμε

$$\mathbb{E}(N_1^n(d)) = \mathcal{O}(M) + o(1) + (1 + o(1)) \frac{4n}{(d+1)(d+2)(d+3)} \sim \frac{4n}{(d+1)(d+2)(d+3)} \quad (5.12)$$

η οποία είναι και η επιθυμητή μορφή της κατανομής του $\mathbb{P}(d)$ (δεν έχουμε δείξει όμως ακόμα πως το $\mathbb{P}(d) \propto d^{-3}$, μόνο ότι για την αναμενόμενη τιμή ισχύει αυτό).

Μπορούμε τώρα να στρέψουμε το βλέμμα μας στη γενική περίπτωση όπου ο αριθμός των ακμών που “φέρνει” μαζί της μια νέα κορυφή μπορεί να είναι μεγαλύτερος του 1 ($m \geq 1$). Θα συμβολίζουμε με d'_k το συνολικό βαθμό της κορυφής v_k σε ένα τυχαίο γράφημα $G_m^n \in \mathcal{G}_m^n$. Θα χρησιμοποιούμε το συμβολισμό d_k για το συνολικό βαθμό της κορυφής v_k στο G_1^{mn} . Πάλι θέλουμε να εκτιμήσουμε την $\mathbb{P}(d'_{k+1} = d + m)$, δηλαδή την πιθανότητα μια κορυφή στο G_m^n να έχει εσωτερικό βαθμό ίσο με d . Για αυτό το λόγο θα μιλήσουμε για τις κατανομές των βαθμών $d_{Y+1}, d_{Y+2}, \dots, d_{Y+m}$ στο G_1^N , όπου το $Y = mk$ και το $N = mn$. Το επιχείρημα που οδήγησε στη δεσμευμένη πιθανότητα της σχέσης 5.9, μπορεί να χρησιμοποιηθεί και δεδομένης ολόκληρης της ακολουθίας βαθμών των κορυφών που προϋπήρχαν της εξεταζόμενης κορυφής. Έτσι, για $M \leq k \leq n - M$ και $0 \leq d \leq n^{1/15}$, αποδεικνύεται πως δεδομένου πως κανένα από τα $|D_{Y+j} - 2\sqrt{(Y+j)N}|$ δεν είναι αρκετά μεγάλο ισχύει ότι

$$\mathbb{P}(d_{Y+j+1} = d + 1 | d_1, d_2, \dots, d_{Y+j}) \sim \sqrt{(Y+j)/N} \left(1 - \sqrt{(Y+j)/N}\right)^d \sim \sqrt{\kappa}(1 - \sqrt{\kappa})^d \quad (5.13)$$

όπου το $j = 0, \dots, m - 1$ και το $\kappa = Y/N$. Άρα χρησιμοποιώντας τη σχέση 5.8 παίρνουμε πως

$$\begin{aligned} \mathbb{P}(d'_{k+1}) &= o(n^{-1}) + (1 + o(1)) \sum_{a_1 + a_2 + \dots + a_m = d} \prod_{j=1}^m \sqrt{\kappa}(1 - \sqrt{\kappa})^{a_j} \\ &= o(n^{-1}) + (1 + o(1)) \binom{d+m-1}{m-1} \kappa^{m/2} (1 - \sqrt{\kappa})^d \end{aligned}$$

με τα a_1, a_2, \dots, a_m να συμβολίζουν τον εσωτερικό βαθμό των κορυφών $Y+1, Y+2, \dots, Y+m$ στο G_1^{mn} . Προχωρώντας όπως στις σχέσεις 5.10 και 5.11, υπολογίζουμε το ολοκλήρωμα

$$\int_0^1 \kappa^{m/2} (1 - \sqrt{\kappa})^d d\kappa = 2 \frac{(m+1)!d!}{(d+m+2)!}$$

χρησιμοποιώντας την αντικατάσταση $\kappa = (1 - u)^2$. Εν τέλει καταλήγουμε στο συμπέρασμα πως

$$\mathbb{E}(N_m^n(d)) \sim \frac{2m(m+1)n}{(d+m)(d+m+1)(d+m+2)} \quad (5.14)$$

Πριν προχωρήσουμε στην απόδειξη του θεωρήματος 5.1, θα χρειαστούμε το ακόλουθο λήμμα ([2],[9]).

Λήμμα 5.2. (Ανισότητα Azuma-Hoeffding) Έστω $(X_t)_{t=0}^n$ να είναι ένα martingale με $|X_{t+1} - X_t| \leq c$ για κάθε $t = 0, \dots, n-1$. Τότε ισχύει πως

$$\mathbb{P}(|X_n - X_0| \geq x) \leq \exp\left(-\frac{x^2}{2c^2n}\right) \quad (5.15)$$

Απόδειξη. Θεωρήματος 5.1.

Είμαστε πάλι στην περίπτωση όπου ένα τυχαίο γράφημα του μοντέλου παράγεται μέσα της διαδικασίας $(G_m^t)_{t \geq 0}$. Θεωρούμε πως έχουμε σταθεροποιήσει τα m, n, d και θεωρούμε τη διαδικασία $X_t = \mathbb{E}(N_m^n(d)|G_m^t)$ για κάθε $0 \leq t \leq n$. Η X_t είναι μια Doob-martingale. Επίσης, ισχύει πως το $X_n = N_m^n(d)$, ενώ το $X_0 = \mathbb{E}(N_m^n(d))$. Τέλος, οι διαφορές $|X_{t+1} - X_t|$ δεν μπορούν να ξεπερνούν το $2m$, καθώς το εάν τη χρονική στιγμή t ενώσουμε την κορυφή v_t με μέσω της πρώτης ακμής με την κορυφή v_i ή την v_j , μέσω της δεύτερης ακμής με την κορυφή v_i ή την v_m , κ.ο.κ., δεν επηρεάζει σε μεταγενέστερες στιγμές το βαθμό των κορυφών v_α με $\alpha \notin \{i, j\}$, το βαθμό των κορυφών v_β με $\beta \notin \{l, m\}, \dots$, και άρα ο $\mathbb{E}(N_m^n(d))$ αλλάζει το πολύ κατά $2m$ σε μια χρονική στιγμή. Μια πιο αυστηρή απόδειξη αυτού του αποτελέσματος βρίσκεται στο Λήμμα 8.6, στο [13].

Παίρνοντας έτσι το λήμμα 5.2, έχουμε ότι

$$\mathbb{P}(|N_m^n(d) - \mathbb{E}(N_m^n(d))| \geq \sqrt{n \log n}) \leq \exp(-\log n/8) = o(n^{-1/15}) \quad (5.16)$$

Από τη σχέση 5.14 βρήκαμε όμως ότι $\mathbb{E}(N_m^n(d)) \sim \frac{2m(m+1)n}{(d+m)(d+m+1)(d+m+2)}$, ποσότητα n οποία είναι κατά πολύ μεγαλύτερη του $\sqrt{n \log n}$, και άρα το ζητούμενο αποδεικνύεται. \square

Βιβλιογραφία

- [1] Alon, N. & Spencer, J., *The Probabilistic Method*, Fourth Edition, John Wiley and Sons, (2016)
- [2] Azuma, K., *Weighted sums of certain dependent variables*, Tôhoku Math J 3, 357-367, (1967)
- [3] Barabási, A.-S. & Albert, R., *Emergence of scaling in random networks*, Science Magazine, **286**, 509-512 (1999)
- [4] Bollobás, B., *The Evolution of Random Graphs*, Transactions of the American Mathematical Society, (1984)
- [5] Bollobás, B., *Random Graphs*, Second Edition, Cambridge University Press, (2001)
- [6] Bollobás, B., Riordan, O., Spencer, J. & Tusnády, G., *The Degree Sequence of a Scale-Free Random Graph Process*, Random Structures & Algorithms, **18**, 279-290, (2001)
- [7] Erdős, P. & Rényi, A., *On the Evolution of Random Graphs*, Publications of the Mathematical Institute of the Hungarian Academy of Science, (1958-1960)
- [8] Gilbert, E. N., *Random Graphs*, Bell Laboratories, Inc., (1959)
- [9] Hoeffding, W., *Probability inequalities for sums of bounded random variables*, J Amer. Stat. Assoc. **58**, 13-30, (1963)
- [10] Janson, S., Luczak, T., Rucinski, A., *Random Graphs*, John Wiley and Sons, (2000)
- [11] Loulakis, M., *Στοχαστικές Διαδικασίες*, Επικαιροποιημένη Έκδοση, Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, (2019)
- [12] Newman, M., *Networks*, Second Edition, Oxford University Press, (2018)
- [13] van der Hofstad, R., *Random Graphs and Complex Networks*, Volume 1, Cambridge University Press, (2016)