

# Κεφάλαιο 9

## Βιολογικά Δίκτυα

### Σύνοψη

Το κεφάλαιο αυτό ενσωματώνει στοιχεία από τα δύο αμέσως προηγούμενα και επικεντρώνεται στη μελέτη των βιολογικών δικτύων. Μετά από μια εισαγωγή των βασικών βιολογικών εννοιών που περιλαμβάνουν τα είδη των δικτύων (ρυθμιστικά, μεταβολικά κ.ά.) περνάει στο πραγματικό βιολογικό πρόβλημα που αφορά την εξαγωγή ενός δικτύου ρυθμιστικών αλληλεπιδράσεων πρωτεϊνών από δεδομένα γονιδιακής έκφρασης. στο πλαίσιο της συζήτησης του προβλήματος θα εισαχθούν βασικές έννοιες από τη θεωρία δικτύων (τοπολογία, βαθμός, μέγεθος δικτύου κλπ) και στη συνέχεια θα αναλυθούν οι στατιστικές ιδιότητες των βιολογικών δικτύων με έμφαση στις δομές αυτο-οργάνωσης, τα ανεξάρτητα κλίμακας (*scale-free*) δίκτυα και τους νόμους δύναμης. Οι πρακτικές ασκήσεις θα περιλαμβάνουν την εξαγωγή δικτύων και την ανάλυση των στατιστικών τους ιδιοτήτων.

### Στο τέλος του Κεφαλαίου θα πρέπει να μπορείτε:

- Να περιγράψετε βασικά είδη βιολογικών δικτύων.
- Να αναγνωρίσετε τα κυριότερα μοτίβα δικτύων και τις ιδιότητές τους.
- Να αναλύσετε ποσοτικά τα βασικότερα χαρακτηριστικά γράφων/δικτύων.
- Να διακρίνετε μεταξύ τυχαίων δικτύων και δικτύων ανεξάρτητων κλίμακας.

## Εισαγωγή

Στα δύο προηγούμενα κεφάλαια μελετήσαμε τρόπους για την ανάλυση της γονιδιακής έκφρασης και για την ερμηνεία της σε επίπεδο λειτουργιών. Σε διάφορα σημεία του προηγούμενου κεφαλαίου αναφερθήκαμε στην έννοια των βιολογικών δικτύων και στην οργάνωση των κυτταρικών λειτουργιών σε αυτά. Ακόμα και οι απλούστεροι οργανισμοί χρειάζεται να ρυθμίσουν χιλιάδες βιοχημικές διεργασίες προκειμένου να προσλάβουν μεταβολίτες και να συνθέσουν τα απαραίτητα μεταβολικά προϊόντα, να εξάγουν από αυτά την ενέργεια που χρειάζονται για να αποκριθούν σε ερεθίσματα και να επεξεργαστούν την πληροφορία που μεταδίδουν από γενιά σε γενιά. Το γονιδίωμα ενός βακτηρίου *E. coli* περιέχει περίπου 4000 γονίδια, όμως ο αριθμός των αντιδράσεων που συμβαίνουν κατά τη διάρκεια του κύκλου ζωής ενός κυττάρου του ξεπερνά τις 100.000 (Almaas, Kovács, Vicsek, Oltvai & Barabási, 2004)! Ο συντονισμός τους μέσω της διαδικασίας της εξέλιξης έχει οργανωθεί με τρόπο που να είναι ταυτόχρονα αποτελεσματικός αλλά και ευέλικτος, ώστε να επιτρέπει την προσαρμογή τους ανάλογα με τις διαρκείς μεταβολές του κυτταρικού περιβάλλοντος. Όλοι οι ζωντανοί οργανισμοί αποτελούν εξαιρετικά πολύπλοκα συστήματα (Barve & Wagner, 2013).

Η μελέτη της πολυπλοκότητας των βιολογικών συστημάτων περνάει έτσι, υποχρεωτικά από την επισκόπηση ενός τεράστιου αριθμού διεργασιών που αλληλοδιαπλέκονται δημιουργώντας πυκνά, σύνθετα δίκτυα. Η ανασύσταση, αρχικά, και η ανάλυση στη συνέχεια αυτών των δικτύων σε ό,τι αφορά το μέγεθος, τις επιμέρους υπομονάδες τους και τις στατιστικές τους ιδιότητες μας επιτρέπει να προσεγγίσουμε σε βάθος πολύ σημαντικά ερωτήματα σχετικά με τη ρύθμιση βιολογικών διεργασιών, την οργάνωση των κυττάρων σε επίπεδο συστημάτων αλλά και την ανάδυση γενικότερων ιδιοτήτων που αντανakλούν τον τρόπο με τον οποίο τα πολύπλοκα βιολογικά συστήματα εξελίσσονται σε διαρκώς μεταβαλλόμενα περιβάλλοντα διατηρώντας ταυτόχρονα μια αξιοθαύμαστη σταθερότητα (Wagner, 2013).

Στο συγκεκριμένο κεφάλαιο θα συζητήσουμε για την έννοια του δικτύου, ως μαθηματικό αντικείμενο και θα αναλύσουμε τις ιδιότητες των δικτύων από θεωρητική σκοπιά, πριν περάσουμε στην περιγραφή των κυριότερων ειδών βιολογικών δικτύων. Στη συνέχεια θα δούμε πώς η ανάλυση στοιχείων τέτοιων βιολογικών δικτύων μπορεί να μας βοηθήσει να κατανοήσουμε βασικούς μηχανισμούς ρύθμισης σε λειτουργικό επίπεδο. Τέλος, θα εξετάσουμε χαρακτηριστικά στατιστικής φύσεως των βιολογικών δικτύων και τον τρόπο με τον οποίο αυτά αντανakλούν βαθύτερες φυσικές αρχές για την οργάνωση των βιολογικών συστημάτων και της ζωής γενικότερα.

## Το βιολογικό πρόβλημα

Τα προβλήματα που σχετίζονται με τα βιολογικά δίκτυα είναι αντίστοιχης πολυπλοκότητας και ποικιλομορφίας με τα ίδια τα δίκτυα. Έτσι δε θα μπορούσαμε να επικεντρωθούμε σ' ένα συγκεκριμένο ερώτημα. Αντίθετα, στο πλαίσιο της μελέτης της βιολογικής πολυπλοκότητας, θα συζητήσουμε μια σειρά από πιο γενικά ερωτήματα που σχετίζονται τόσο με επιμέρους

χαρακτηριστικά όσο και με γενικότερες ιδιότητες των δικτύων.

Ξεκινώντας από τα επιμέρους, φανταστείτε, για παράδειγμα, ότι μελετάτε ένα σύνολο μεταβολικών αντιδράσεων που αποτελούν μέρος ενός βιοχημικού μονοπατιού (π.χ. της Γλυκόλυσης) και σας ενδιαφέρει να προβλέψετε τον τρόπο με τον οποίο θα αυξομειώνονται οι συγκεντρώσεις των μεταβολιτών που συμμετέχουν στις αντιδράσεις. Πέρα από το καθαρά επιστημονικό ενδιαφέρον, ένα τέτοιο ερώτημα μπορεί να έχει και πρακτικές εφαρμογές στον τομέα της βιοτεχνολογίας, όπου συχνά μας ενδιαφέρει να διαμορφώσουμε τη διαδοχή (ή το δίκτυο) των αντιδράσεων ώστε να μεγιστοποιήσουμε την παραγωγή ενός συγκεκριμένου μεταβολίτη, να ελαχιστοποιήσουμε τα παραπροϊόντα κλπ. Η ερώτηση που προκύπτει είναι:

*Πώς θα προβλέψουμε τη διαμόρφωση των συγκεντρώσεων των προϊόντων ενός δικτύου βιοχημικού αντιδράσεων από τη δομή του δικτύου;*

Ο τρόπος με τον οποίον προσεγγίζουμε μια τέτοια ερώτηση δε διαφέρει πρακτικά αν το δίκτυο αποτελείται από βιοχημικές αντιδράσεις με μεταβολίτες ως υποστρώματα και προϊόντα ενζύμων ή από γονίδια που αλληλο-ρυθμίζουν την έκφρασή τους. Στην ενότητα των Μοτίβων Βιολογικών Δικτύων θα δούμε περισσότερα για αυτό το ερώτημα.

Τα κυριότερα ερωτήματα που ανακύπτουν από τη μελέτη γενικότερων χαρακτηριστικών των βιολογικών δικτύων σχετίζονται με τις στατιστικές τους ιδιότητες. Τέτοια ερωτήματα έχουν πρακτικό ενδιαφέρον, όπως για παράδειγμα:

*Πώς μπορούμε να εντοπίσουμε τους πιο σημαντικούς μεταγραφικούς παράγοντες σε ένα δίκτυο μεταγραφικής ρύθμισης γονιδίων;*

ή μπορούν να σχετίζονται με θεωρητικά προβλήματα όπως:

*Κατά πόσο ένα δίκτυο πρωτεϊνικών αλληλεπιδράσεων θα μπορούσε να έχει προκύψει τυχαία;*

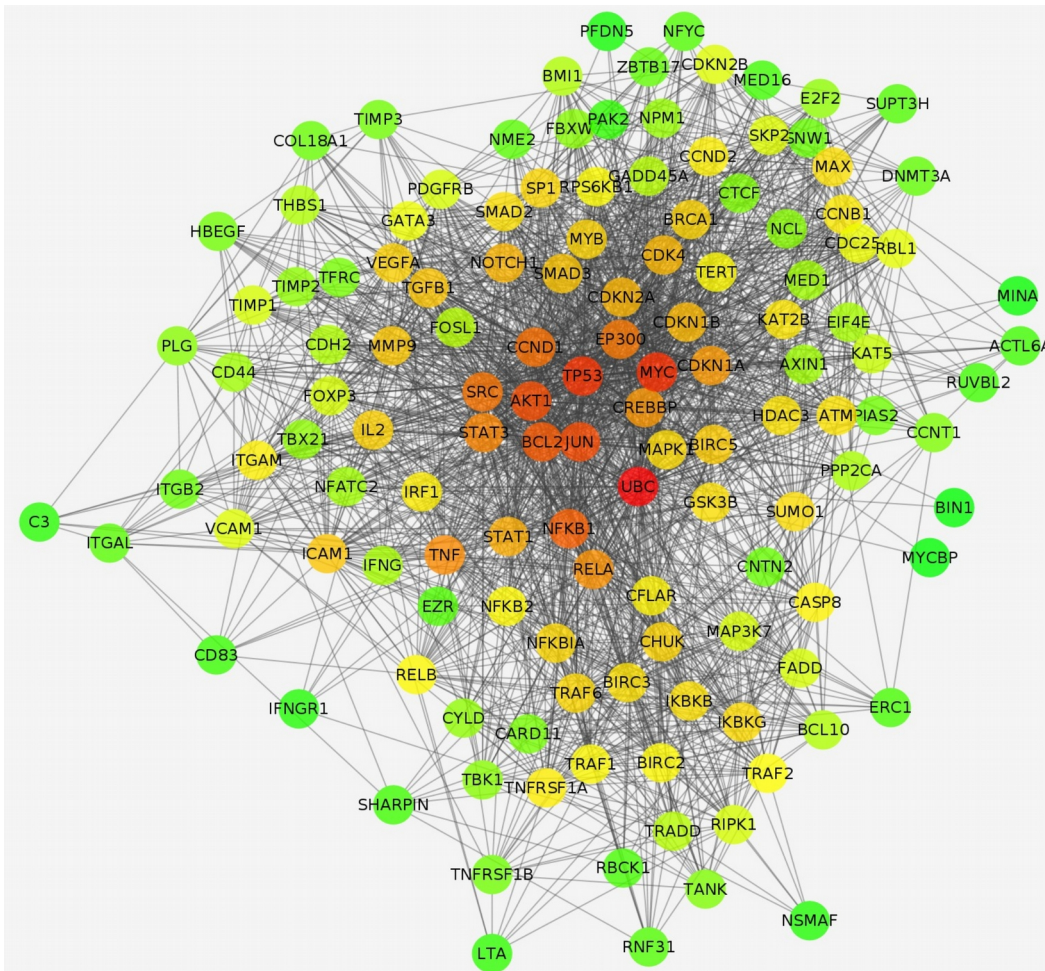
Μολονότι φαινομενικά πολύ διαφορετικά και τα δύο παραπάνω προβλήματα σχετίζονται με μαθηματικές ιδιότητες των δικτύων όπως είναι η κατανομή των στοιχείων τους και των μεταξύ τους συνδέσεων, η γενικότερη τοπολογία τους κλπ. Μετά την εισαγωγική συζήτηση για τα μαθηματικά χαρακτηριστικά των δικτύων θα μπορέσουμε αρχικά να προσεγγίσουμε το πρώτο από τα δύο παραπάνω ερωτήματα με ποσοτικό και αντικειμενικό τρόπο. Για το δεύτερο, θα πρέπει να περιμένουμε ως το τελευταίο μέρος αυτού του κεφαλαίου, οπότε και θα μελετήσουμε τη φύση των βιολογικών δικτύων από στατιστική άποψη, θα ορίσουμε πότε ακριβώς ένα δίκτυο είναι “τυχαίο” και θα δούμε πώς μπορούμε με υπολογιστικές προσεγγίσεις να διακρίνουμε τα δίκτυα που έχουν μια συγκεκριμένη οργάνωση από τα τυχαία.

Πρώτα απ' όλα, όμως, ας δούμε για τι είδους παράξενα αντικείμενα μιλάμε όταν μιλάμε για δίκτυα και με ποιες μορφές αυτά εμφανίζονται στα βιολογικά συστήματα.

## Δίκτυα

Όταν ακούμε τη λέξη “Δίκτυο”, στο μυαλό μας έρχονται πιθανότατα εικόνες από ηλεκτρικά κυκλώματα, καλωδιώσεις τηλεπικοινωνιών ή οδικές αρτηρίες που διασταυρώνονται. Πράγματι, όλα τα παραπάνω έχουν χαρακτηριστικά δικτύων και δεν είναι προφανώς τα μόνα. Αν προσπαθήσουμε να δώσουμε έναν απλό ορισμό στην έννοια του δικτύου, αυτός θα μπορούσε να είναι:

Μια θεωρητική κατασκευή που περιγράφει τις σχέσεις μεταξύ ομοειδών ή ομάδων ομοειδών στοιχείων αναπαριστώντας τις με σχηματικό τρόπο.



**Εικόνα 9.1:** Ένα παράδειγμα δικτύου πρωτεϊνικών αλληλεπιδράσεων. Οι πρωτεΐνες αναπαρίστανται ως κόμβοι και οι μεταξύ τους σχέσεις με ακμές, που αντιστοιχούν σε σχέσεις ενεργοποίησης. Το χρώμα των κόμβων είναι ανάλογο των αριθμητικών τιμών χαρακτηριστικών ιδιοτήτων του δικτύου που θα συζητηθούν στη συνέχεια του κεφαλαίου. Η Εικόνα δημιουργήθηκε με τη χρήση του Cytoscape (Shannon et al., 2003).

Ο παραπάνω ορισμός δεν είναι αυστηρός από μαθηματική άποψη (αφήνουμε κάτι τέτοιο για επόμενη ενότητα αυτού του κεφαλαίου) αλλά είναι ικανοποιητικός για να μας βοηθήσει να καταλάβουμε το είδος των οντοτήτων που θα συζητήσουμε στη συνέχεια. Η βασική έννοια στον ορισμό των δικτύων είναι αυτή των "σχέσεων". Απομονωμένα στοιχεία οποιουδήποτε είδους δε συνιστούν δίκτυο αλλά ένα απλό σύνολο. Αυτό που ορίζει ένα δίκτυο είναι το πλέγμα των σχέσεων ή των "συνδέσεων" μεταξύ τους. Τα δίκτυα είναι στην ουσία γραφικές αναπαραστάσεις πολύπλοκων συστημάτων που μας βοηθούν να τα κατανοήσουμε καλύτερα και πληρέστερα. Έτσι π.χ. το οδικό δίκτυο μιας χώρας δεν αποτελείται μόνο από τις πόλεις και τα χωριά της αλλά κυρίως από τους δρόμους και τις οδικές αρτηρίες που τα συνδέουν. Από τις συγκοινωνίες, τις τηλεπικοινωνίες και τις

κοινωνικές σχέσεις, ως τις οικονομικές συναλλαγές και το διαδίκτυο (εξ ορισμού), δίκτυα υπάρχουν παντού. Στη συνέχεια θα δούμε τα είδη των δικτύων που συναντούμε στη βιολογία.

**Ερώτηση:** Προσπαθείστε να σκεφτείτε όσο περισσότερα δίκτυα μπορείτε από διάφορους επιστημονικούς τομείς; Ποια είναι τα κοινά χαρακτηριστικά τους και σε ποια στοιχεία διαφέρουν;

## Είδη Βιολογικών Δικτύων

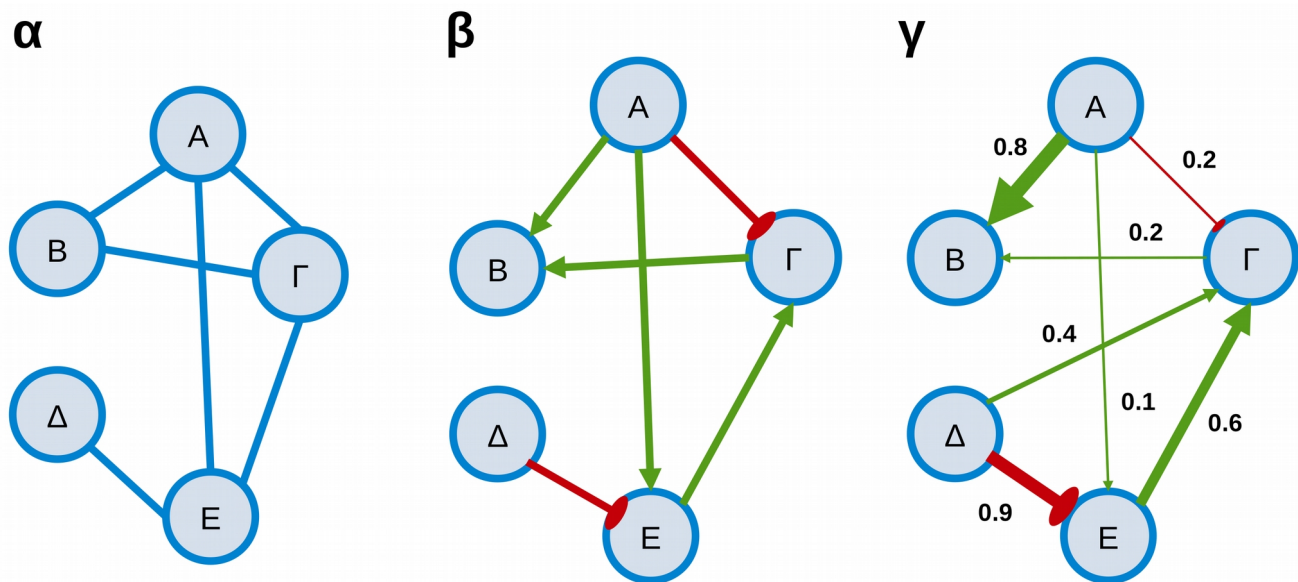
Βιολογικά δίκτυα συναντούμε σε όλα τα επίπεδα μελέτης των επιστημών της ζωής από το πιο μικροσκοπικό (μοριακό) ως το πιο μακροσκοπικό (οικοσυστήματα). Παρά το γεγονός ότι η πολυπλοκότητα των βιολογικών συστημάτων (όπως τη συζητήσαμε στην Εισαγωγή) είναι πιο μεγάλη στο μοριακό επίπεδο (λόγω του αριθμού των αλληλεπιδράσεων), αυτό δε σημαίνει πως δεν υπάρχει μια πληθώρα δικτύων βιολογικού ενδιαφέροντος σε άλλους τομείς των βιολογικών επιστημών όπως της φυσιολογίας (π.χ. κυκλοφοριακό και αναπνευστικό σύστημα), της οικολογίας (π.χ. διατροφικές αλυσίδες, σχέσεις θηρευτή-θηράματος) κ.ά. Για τους σκοπούς αυτού του κεφαλαίου, ωστόσο, θα συζητήσουμε αποκλειστικά δίκτυα συστημάτων της μοριακής βιολογίας, των οποίων τα στοιχεία δηλαδή είναι μόρια (μεταβολίτες, πρωτεΐνες και γονίδια). Στη συνέχεια θα αναφέρουμε περιληπτικά τα πιο σημαντικά και ευρέως μελετημένα δίκτυα μοριακής βιολογίας, με έμφαση στα χαρακτηριστικά που τα κάνουν ενδιαφέροντα από πλευράς μηχανισμών.

## Ρυθμιστικά Δίκτυα

Ρυθμιστικά δίκτυα (regulatory networks) ονομάζονται τα δίκτυα που ορίζουν το πλέγμα των λειτουργικών αλληλεπιδράσεων της ρύθμισης της έκφρασης μεταξύ γονιδίων. Τα δίκτυα αυτά αποτελούνται από (έχουν ως στοιχεία τους) γονίδια (ή εναλλακτικά τις κωδικοποιούμενες από αυτά πρωτεΐνες) και οι μεταξύ τους σχέσεις αντιστοιχούν στις σχέσεις ρύθμισης μεταξύ τους (Davidson, 2010). Έτσι π.χ. αν το γονίδιο A ρυθμίζεται από το B αυτό θα αναπαρίσταται στο δίκτυο με μια σύνδεση μεταξύ των A και B, όπως βλέπουμε στην Εικόνα 9.2α. Το είδος της ρύθμισης, αν δηλαδή πρόκειται για ενεργοποίηση ή για καταστολή, μπορεί να λαμβάνεται υπόψη από το δίκτυο (Εικόνα 9.2β), όπως επίσης και ο βαθμός στον οποίο η ρύθμιση αυτή διαμορφώνει την έκφραση του κάθε στοιχείου. Αν, για παράδειγμα, το B ενεργοποιείται τόσο από το A όσο και από το Γ αλλά η εξάρτησή του από το A είναι μεγαλύτερη, αυτό μπορεί να αποτυπωθεί στον τρόπο, με τον οποίο αναπαρίστανται οι συνδέσεις στο δίκτυο, όπως φαίνεται στην Εικόνα 9.2γ. Περισσότερα για τους τρόπους αναπαράστασης δικτύων θα δούμε πιο αναλυτικά στην ενότητα όπου θα συζητήσουμε τις ιδιότητες των δικτύων ως γράφων (βλ. Μαθηματικό Ιντερμέδιο I).

Τα ρυθμιστικά δίκτυα είναι ίσως το πιο σημαντικό είδος βιολογικού δικτύου για πολλούς λόγους. Αρχικά, διότι αποτελούν αντανάκλαση μιας πολύ σημαντικής κυτταρικής λειτουργίας που αποτελεί τη βάση για την κυτταρική εξειδίκευση, την ανάπτυξη και την απόκριση των οργανισμών σε περιβαλλοντικά ερεθίσματα. Έπειτα γιατί εξαιτίας ακριβώς αυτού του κομβικού ρόλου, τα δίκτυα

ρύθμισης είναι εξαιρετικά εύπλαστα, μπορούν δηλαδή να μετασχηματιστούν δυναμικά, μεταβάλλοντας τον αριθμό των στοιχείων τους, το είδος και τον αριθμό των συνδέσεων μεταξύ τους κλπ. Τέλος, τα ρυθμιστικά δίκτυα και συγκεκριμένα τα δίκτυα μεταγραφικής ρύθμισης (transcription regulatory networks) είναι ευκολότερο να μελετηθούν από άλλα δίκτυα εξαιτίας του άμεσου τρόπου με τον οποίο μπορούμε πλέον να μετρήσουμε τα επίπεδα έκφρασης πολλών γονιδίων σε πολλές συνθήκες (βλ. Κεφάλαιο 7). Ο δυναμικός τους χαρακτήρας και η “πλαστικότητα” τους, μπορούν έτσι να αξιολογηθούν με μεγάλη ακρίβεια και κυρίως μπορούν να μας δώσουν πολύ σημαντικές πληροφορίες για μια βασική κυτταρική διεργασία όπως είναι ο έλεγχος της γονιδιακής έκφρασης (Walhout, 2011). Λόγω ακριβώς αυτής της βασικής λειτουργίας που περιγράφουν, τα ρυθμιστικά δίκτυα συνδέονται άμεσα με όλα τα υπόλοιπα βιολογικά δίκτυα που θα συζητήσουμε στη συνέχεια καθώς η ύπαρξη ή όχι στοιχείων που συμμετέχουν σε αυτά (ένζυμα και άλλες πρωτεΐνες) είναι απόλυτα εξαρτημένη από τη γονιδιακή ρύθμιση.

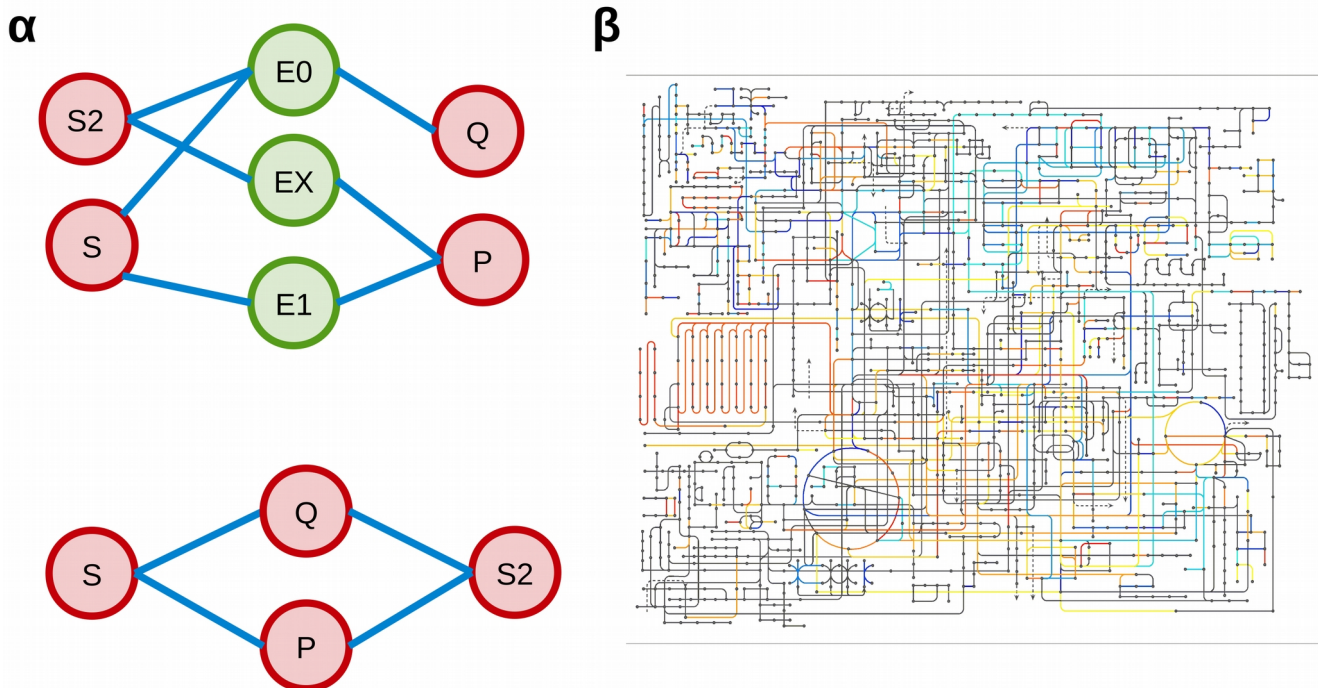


**Εικόνα 9.2:** Τρία παραδείγματα δικτύων μεταγραφικής ρύθμισης με πέντε κόμβους που παριστάνουν σχέσεις με διαφορετικά επίπεδα ενσωμάτωσης πληροφορίας. α) Μη κατευθυνόμενο δίκτυο που αναπαριστά μόνο τους κόμβους που διασυνδέονται μέσω των ακμών. β) Κατευθυνόμενο δίκτυο με ακμές δύο τύπων: θετικές (πράσινο) και αρνητικές (κόκκινο) που σε ένα ρυθμιστικό δίκτυο μπορούν να αντιστοιχούν σε ενεργοποίηση και καταστολή αντίστοιχα. γ) Κατευθυνόμενο δίκτυο με βάρη ακμών τα οποία παριστάνονται με διαφορετικά πάχη των ακμών που συνοδεύονται από τις αντίστοιχες αριθμητικές τιμές τους. Στο συγκεκριμένο σχήμα, η κατασταλτική δράση του Δ στο E είναι 4.5 φορές μεγαλύτερη από αυτήν του A στο Γ.

## Μεταβολικά Δίκτυα

Τα μεταβολικά δίκτυα έχουν συνήθως στοιχεία που ανήκουν σε δύο κατηγορίες και που ονομάζονται “διμερή” (bipartite). Περιλαμβάνουν αφενός τα ένζυμα που καταλύουν τις μεταβολικές αντιδράσεις και αφετέρου τους μεταβολίτες που αποτελούν τα υποστρώματα και τα προϊόντα των

αντιδράσεων αυτών. Οι σχέσεις μεταξύ των στοιχείων είναι εδώ προφανείς. Δύο στοιχεία ενώνονται αν συμμετέχουν στην ίδια αντίδραση (Stelling, Klamt, Bettenbrock, Schuster & Gilles, 2002). Στο παράδειγμα της Εικόνας 9.3α, αν το ένζυμο E1 δρα πάνω στο υπόστρωμα S και παράγει το μεταβολίτη P το E1 θα συνδέεται τόσο με το S όσο και με το P (το είδος της αναπαράστασης γι' αυτές τις συνδέσεις μπορεί να διαφέρει αν θέλουμε να διακρίνουμε υποστρώματα από προϊόντα). Αντίστοιχα αν το S αποτελεί με τη σειρά του προϊόν μιας άλλης αντίδρασης την οποία καταλύει το E0 τότε θα συνδέεται και με αυτό κ.ο.κ. Σε κάποιες περιπτώσεις μπορεί κανείς να θεωρήσει ένα μεταβολικό δίκτυο που να αποτελείται μόνο από τα πρώτα (σχεδόν ποτέ όμως μόνο από τα δεύτερα), οπότε οι συνδέσεις γίνονται μεταξύ στοιχείων που μοιράζονται έναν μεταβολίτη. Τέτοιες αναπαραστάσεις έχουν σκοπό να απλοποιήσουν τα δίκτυα τα οποία ακριβώς λόγω των εξαιρετικά πολύπλοκων μεταβολικών σχέσεων που έχουν αναπτυχθεί μέσω της εξέλιξης (για λόγους οικονομίας και καλύτερου συντονισμού) εμφανίζουν μεγάλη πολυπλοκότητα. Στην Εικόνα 9.3β βλέπουμε μια τέτοια “απλοποιημένη” αναπαράσταση του δικτύου του βασικού μεταβολισμού του ανθρώπου (ή έμφαση στο “βασικός”, οι αντιδράσεις είναι πολλαπλάσιες από τις συνδέσεις αυτού του σχήματος) (Ogata et al., 1999).



**Εικόνα 9.3:** α) Παράδειγμα διμερούς δικτύου μεταβολικής ρύθμισης όπου με κόκκινο εμφανίζονται οι μεταβολίτες (S, S2, P, Q) και με πράσινο (E0, E1, EX) τα ένζυμα που καταλύουν τις μεταξύ τους αντιδράσεις. Πάνω: το πλήρες δίκτυο. Κάτω: το δίκτυο που περιλαμβάνει μόνο τους μεταβολίτες και έχει προκύψει από συνένωση των ακμών του πάνω δικτύου. β) Ένα μεταβολικό δίκτυο που αντιστοιχεί σε μέρος των αντιδράσεων του μεταβολισμού ενός ανθρώπινου κυττάρου. Σε αυτόν τον πολύπλοκο “μεταβολικό χάρτη” διακρίνονται οι κύκλοι του κιτρικού οξέος και της ουρίας. (Εικόνα από J3D3, CC BY-SA 4.0, από Wikimedia Commons).

Η μελέτη των μεταβολικών δικτύων συνδέεται σε μεγάλο βαθμό με τα δίκτυα μεταγραφικής

ρύθμισης, μέσω των σχετικών συγκεντρώσεων των ενζύμων που εξαρτώνται πολύ συχνά από συγκεκριμένα προγράμματα έκφρασης γονιδίων. Ένα βασικό πλεονέκτημα, ωστόσο, των δικτύων αυτών σε ό,τι αφορά τη μελέτη τους είναι οι μονοσήμαντες σχέσεις μεταξύ των στοιχείων τους. Από τη στιγμή που ένας μεταβολίτης ταυτοποιηθεί ως υπόστρωμα (ή προϊόν) μιας αντίδρασης η σχέση αυτή δεν αλλάζει.

## Σηματοδοτικά Δίκτυα

Σηματοδοτικά είναι τα δίκτυα εκείνα που περιγράφουν διεργασίες κυτταρικής σηματοδότησης (cell signaling) μέσω της ενεργοποίησης/καταστολής πρωτεϊνών/ενζύμων. Μαζί με τα μεταβολικά δίκτυα, τα σηματοδοτικά δίκτυα ανήκουν σε μια ευρύτερη κατηγορία που θα μπορούσαμε να ονομάσουμε “δίκτυα αντιδράσεων”, ωστόσο τα μελετούμε ξεχωριστά γιατί οι λειτουργία που επιτελούν δε σχετίζεται με την παραγωγή ή την κατανάλωση ενέργειας αλλά με τη μετάδοση πληροφορίας μέσω μοριακών σημάτων (Bhalla and Ravi Iyengar, 1999). Στα σηματοδοτικά δίκτυα τα στοιχεία είναι σχεδόν αποκλειστικά πρωτεΐνες (και κάποιες φορές μικρός αριθμός μικρών μορίων) και οι σχέσεις μεταξύ τους αντιστοιχούν σε αντιδράσεις ενεργοποίησης που αποτελούν στάδια στη μετάδοση ενός σήματος. Η συχνότερη περίπτωση είναι η σειριακή μετάδοση ενός περιβαλλοντικού ερεθίσματος από τον εξωκυττάριο χώρο μέσω υποδοχέων, σε μια αλυσίδα σηματοδοτικών μορίων μέσα από το κυτταρόπλασμα με τελικούς αποδέκτες μεταγραφικούς παράγοντες εντός του πυρήνα. Πολύ συχνά όμως, περισσότεροι από ένας τέτοιοι “καταρράχτες πληροφορίας” (information cascades) αλληλοδιαπλέκονται μεταξύ τους δημιουργώντας πολύπλοκα δίκτυα (Eissmann et al., 2010).

Τα σηματοδοτικά δίκτυα είναι εξαιρετικά δυναμικά, με χρόνους απόκρισης που μπορεί να είναι τάξεις μεγέθους μικρότεροι από τους αντίστοιχους των ρυθμιστικών και μεταβολικών δικτύων και για το λόγο αυτό αποτελούν ίσως την πιο δύσκολη προς μελέτη κατηγορία. Από την άλλη πλευρά η σημασία τους για τη μελέτη της απόκρισης σε εξωτερικά ερεθίσματα και της κυτταρικής επικοινωνίας, λειτουργιών που σχετίζονται στενά με παθολογικές καταστάσεις τα κάνει ιδιαίτερα ενδιαφέροντα κυρίως για τομείς όπως είναι π.χ. η μοριακή ανοσολογία (Wilson, Dixit & Ashkenazi, 2009).

## Δίκτυα Πρωτεϊνικών Αλληλεπιδράσεων

Στα δίκτυα πρωτεϊνικών αλληλεπιδράσεων αναφερθήκαμε και στο προηγούμενο κεφάλαιο ως ξεχωριστή κατηγορία λειτουργικής ομαδοποίησης. Με την αυστηρή έννοια του όρου όλα τα βιολογικά δίκτυα είναι δίκτυα πρωτεϊνικών αλληλεπιδράσεων εφόσον περιέχουν πρωτεΐνες μεταξύ των στοιχείων τους. Για το λόγο αυτό, διακρίνουμε τα δίκτυα πρωτεϊνικών αλληλεπιδράσεων από τα υπόλοιπα όταν δεν μπορούμε με σαφήνεια να αποδώσουμε τη φύση της αλληλεπίδρασης ή εναλλακτικά όταν οι αλληλεπιδράσεις είναι ποικίλων τύπων (π.χ. και ρυθμιστικές αλλά και σηματοδοτικές). Από αυτήν την άποψη τα δίκτυα αυτού του τύπου είναι συχνά “μικτά” δίκτυα με



πολλές και διαφορετικές λειτουργικές σχέσεις μεταξύ των πρωτεϊνών (Johnson, Mortazavi, Myers & Wold, 2007). Ένα άλλο βασικό χαρακτηριστικό τους είναι ότι πολύ συχνά (και κάποιες φορές αποκλειστικά) αφορούν σχέσεις φυσικής αλληλεπίδρασης, όταν δηλαδή δύο ή περισσότερες πρωτεΐνες συνδέονται μοριακά κατά τη δημιουργία πρωτεϊνικών συμπλόκων. Από την άλλη πλευρά, η ταυτοποίηση τέτοιων αλληλεπιδράσεων είναι εξαιρετικά δύσκολη πειραματικά, γεγονός που οδηγεί συχνά στην αναπαράσταση σε τέτοιου τύπου δίκτυα αρκετά ετερόκλητης πληροφορίας που περιλαμβάνει συνέκφραση γονιδίων, εξελικτικές σχέσεις, πειραματικές ενδείξεις κ.ά.

**Ερώτηση:** Μπορείτε να σκεφτείτε με ποιον τρόπο μελέτες έκφρασης γονιδίων μπορούν να οδηγήσουν στη δημιουργία δικτύων; (Ίσως να σας βοηθούσε αν ανατρέχατε στο Κεφάλαιο 7)

## Μαθηματικό Ιντερμέδιο I. Δίκτυα ως Γράφοι

Έχοντας εξετάσει κάποια ποιοτικά χαρακτηριστικά των βιολογικών δικτύων, θα περάσουμε στη συνέχεια σε μια πιο αυστηρή ποσοτική ανάλυση των δικτύων γενικότερα. Από μαθηματική άποψη, τα δίκτυα είναι γράφοι (graphs) στους οποίους αναπαρίστανται οι σχέσεις μεταξύ των στοιχείων ενός συνόλου αντικειμένων. Τα στοιχεία ενός δικτύου ονομάζονται **κόμβοι** (nodes ή vertices) και οι συνδέσεις μεταξύ τους αναπαρίστανται με τη μορφή γραμμών που συνδέουν τους κόμβους και ονομάζονται **ακμές** (edges). Τυπικά συμβολίζουμε ένα δίκτυο ως το γράφο  $G$  που ορίζεται από το σύνολο των κόμβων  $V$  και των ακμών του  $E$ ,  $G(V, E)$  (Diestel, 2006). Ανάλογα με το είδος του δικτύου (βλ. παρακάτω) η μορφή του μπορεί να ποικίλει σε ότι αφορά τα είδη των ακμών, τους περιορισμούς στη διάρθρωση των κόμβων κλπ. Για παράδειγμα, τα φυλογενετικά δέντρα που συζητήσαμε στο Κεφάλαιο 6 είναι κι αυτά ένα είδος γράφου που θα μπορούσαν να αναπαρασταθούν και ως δίκτυα. Στη συνέχεια θα συζητήσουμε ποιες από τις επιμέρους ιδιότητες των γράφων είναι σημαντικές για τα βιολογικά δίκτυα. Επίσης, θα χρησιμοποιήσουμε στο εξής αποκλειστικά τον όρο “δίκτυο” για να περιγράψουμε τα είδη των γράφων που έχουν σημασία για τους σκοπούς αυτού του κεφαλαίου.

### Αναπαράσταση Δικτύων

Υπάρχουν τρεις βασικοί τρόποι για να αναπαραστήσουμε ένα δίκτυο. Ο γραφικός τρόπος είναι ο πιο συνηθισμένος και αναπαριστά τους κόμβους του δικτύου με σημεία σε ένα επίπεδο και τις ακμές με γραμμές (όχι υποχρεωτικά ευθείες) που ενώνουν τα σημεία αυτά (βλ. Εικόνα 9.1). Η αναπαράσταση αυτή έχει το πλεονέκτημα ότι προσφέρει μια καλύτερη εποπτεία του δικτύου, ωστόσο δεν είναι εύκολο να αναλυθεί περαιτέρω. Για την ανάλυση ποσοτικών χαρακτηριστικών ενός δικτύου αυτό θα πρέπει να πάρει μία από τις δύο τυπικές μορφές που είναι η λίστα γειτνίασης (adjacency list) ή ο πίνακας γειτνίασης (adjacency matrix). Οι λίστες (ή κατάλογοι) γειτνίασης περιέχουν τις σχέσεις μεταξύ των κόμβων με τη μορφή ζευγών κόμβων σε γειτονικές στήλες (βλ.

Εικόνα 9.4β). Ο κατάλογος έχει μήκος ίσο με το άθροισμα των ακμών και των μη συνδεδεμένων (μονήρων) κόμβων και για το λόγο αυτό, οι λίστες γειτνίασης δεν χρησιμοποιούνται για “πυκνά” δίκτυα όπου ο αριθμός των ακμών ξεπερνά το τετράγωνο του αριθμού των κόμβων ( $E > V^2$ ). Στην περίπτωση αυτή προτίμονται οι πίνακες γειτνίασης στους οποίους οι κόμβοι σχηματίζουν έναν τετραγωνικό πίνακα διαστάσεων  $G[V^2]$  και οι ακμές ορίζονται μέσω των αριθμητικών τιμών του πίνακα. Έτσι, στην απλούστερη περίπτωση ενός δικτύου, του οποίου οι ακμές δεν έχουν άλλα ποιοτικά ή ποσοτικά χαρακτηριστικά (βλ. παρακάτω), αυτό θα αναπαρίσταται ως ένας τετραγωνικός πίνακας, όπου δύο κόμβοι  $i, j$  συνδέονται με ακμή αν  $G[i,j]=1$ , ενώ δε συνδέονται αν  $G[i,j]=0$ . Στην περίπτωση που το δίκτυο έχει βάρη οι τιμές του πίνακα θα αντιστοιχούν στα βάρη και εφόσον το δίκτυο είναι κατευθυνόμενο, η κατεύθυνση της ακμής θα δίνεται από το πρόσημο της τιμής  $G[i,j]$ , όπως φαίνεται στην Εικόνα 9.4γ. Στην αμέσως επόμενη ενότητα θα εξηγήσουμε τη σημαίνει “δίκτυο με βάρη” και “κατευθυνόμενο δίκτυο”.

## Είδη Δικτύων

Ανάλογα με το είδος των σχέσεων που περιγράφει ένα δίκτυο μπορεί να είναι:

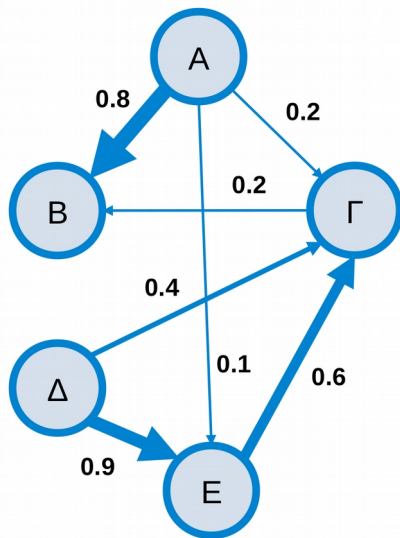
α) μη-κατευθυνόμενο (undirected) (βλ. Εικόνα 9.2α)

β) κατευθυνόμενο (directed) (βλ. Εικόνα 9.2β)

Σ' ένα μη-κατευθυνόμενο δίκτυο οι σχέσεις δεν έχουν κάποια ποιοτικά χαρακτηριστικά και οι ακμές είναι συμμετρικές, δηλαδή η σχέση του A με το B είναι ταυτόσημη με τη σχέση του B με το A. Στην περίπτωση αυτή οι ακμές αναπαρίστανται σ' ένα γράφο ως γραμμές που ενώνουν τα A και B χωρίς άλλα επιμέρους χαρακτηριστικά και ο αντίστοιχος πίνακας γειτνίασης περιέχει μόνο δύο είδη τιμών (0, 1). Αυτό που είναι σημαντικό είναι ότι σ' έναν πίνακα γειτνίασης μη κατευθυνόμενου δικτύου ισχύει  $G[i,j]=G[j,i]$  για κάθε ζεύγος  $i, j$ .

Σ' ένα κατευθυνόμενο δίκτυο, οι σχέσεις γειτνίασης έχουν και ποιοτικό χαρακτήρα που υποδηλώνει την προέλευση μίας λειτουργίας μεταξύ κόμβων. Στο γράφο οι ακμές συμβολίζονται πλέον ως βέλη των οποίων η κατεύθυνση υποδηλώνει μη-συμμετρικές σχέσεις μεταξύ κόμβων, κάτι που σημαίνει ότι στο δίκτυο της Εικόνας 9.4α η σχέση μεταξύ A και B δεν είναι ποιοτικά η ίδια μ' αυτήν μεταξύ B και A. Στον πίνακα γειτνίασης αυτό μεταφράζεται στη σχέση  $G[i,j] \neq G[j,i]$  που σημαίνει ότι αν  $A \rightarrow B$  τότε  $G[A,B]=1$  και  $G[B,A]=-1$  που υποδηλώνει τη διαφορά μεταξύ “εισερχόμενης” (incoming) και εξερχόμενης (outgoing) ακμής. Παράδειγμα κατευθυνόμενου δικτύου στη βιολογία είναι η αναπαράσταση ρυθμιστικών σχέσεων μεταξύ μεταγραφικών παραγόντων και των γονιδίων-στόχων τους. Οι μεταγραφικοί παράγοντες θα ρυθμίζουν τα γονίδια χωρίς να ισχύει το αντίθετο και έτσι οι σχέσεις θα είναι ασύμμετρες.

**Ερώτηση:** Μπορείτε να σκεφτείτε πώς θα διαφέρουν οι πίνακες γειτνίασης μεταξύ μη-κατευθυνόμενων και κατευθυνόμενων δικτύων; Ποια ιδιότητα του πρώτου, δε θα ισχύει στο δεύτερο;

**α****β**

A	B	0.8
	Γ	0.2
	E	0.1
Γ	B	0.2
Δ	Γ	0.4
Δ	E	0.9
E	Γ	0.6

**γ**

	A	B	Γ	Δ	E
A	0	0.8	0.2	0	0.1
B	-0.8	0	-0.2	0	0
Γ	-0.2	0.2	0	-0.4	-0.6
Δ	0	0	0.4	0	0.9
E	0	0	0.6	-0.9	0

**Εικόνα 9.4:** Αναπαράσταση ενός κατευθυνόμενου δικτύου με βάρη α) Με γράφο όπου οι κατευθύνσεις αποδίδονται με βέλη και τα βάρη αντιστοιχούν στα πάχη των γραμμών β) Με λίστα γειτνίασης. Κάθε αλληλεπίδραση που συναντάται πρώτη φορά καταγράφεται με τη φορά της εξερχόμενης ακμής. Όλες οι τιμές βαρών είναι έτσι θετικές γ) Με πίνακα γειτνίασης όπου η ίδια αλληλεπίδραση καταγράφεται δύο φορές με αντίθετη φορά ( $A \rightarrow \Gamma: 0.2$  αλλά  $\Gamma \rightarrow A: -0.2$ )

Όχι σπάνια, οι σχέσεις μεταξύ των στοιχείων ενός δικτύου έχουν και ποσοτικά χαρακτηριστικά. Πέρα δηλαδή από το αν συνδέει ή όχι δύο κόμβους και από την κατεύθυνση με την οποία γίνεται η σύνδεση, μια ακμή μπορεί να είναι “σταθμισμένη” (weighted) να έχει δηλαδή ένα “βάρος” που να είναι ανάλογο μιας ποσότητας σχετικής με μια συγκεκριμένη ιδιότητα. Μένοντας στο παράδειγμα των δικτύων μεταγραφικής ρύθμισης ένα γονίδιο στόχος Γ μπορεί να ελέγχεται από δύο μεταγραφικούς παράγοντες A και B με διαφορετικό βαθμό εξάρτησης (ο A μπορεί να έχει μεγαλύτερη συγγένεια για τον υποκινητή του Γ ή ο B να εκφράζεται λιγότερο κλπ). Αυτό μπορεί να αποτυπωθεί γραφικά με την αναγραφή αριθμητικών τιμών πάνω στην ακμή (βλ. Εικόνες 9.2γ και 9.4α) ή σε κάποιες περιπτώσεις πιο αφαιρετικά με διαφορετικά πάχη μεταξύ ακμών διαφορετικού βάρους. Στην τυπική αναπαράσταση των πινάκων γειτνίασης τα βάρη αντικαθιστούν τις τιμές στα στοιχεία που αντιστοιχούν σε συνδεδεμένους κόμβους. Στο παράδειγμα που εξετάζουμε στην Εικόνα 9.4α θα ισχύει ότι  $G[A,\Gamma]=0.2$  και  $G[B,\Gamma]=-0.2$ .

Πώς όμως γίνεται η απόδοση τιμών βαρών σε δίκτυα με σταθμισμένες ακμές; Αυτό διαφέρει ανάλογα με το είδος του δικτύου και κυρίως την πειραματική προσέγγιση με βάση την οποία προκύπτει το δίκτυο. Έτσι σ' ένα μεταβολικό δίκτυο τα βάρη μπορεί να αντιστοιχούν στη θεωρητική ή πειραματικά προσδιοριζόμενη συγγένεια ενζύμου-υποστρώματος, ενώ κάτι ανάλογο είδαμε στο ρυθμιστικό παράδειγμα που συζητήσαμε παραπάνω. Τα δίκτυα που μελετούμε στη βιολογία είναι πολύ συχνά δίκτυα που προκύπτουν *de novo* από μια πειραματική διαδικασία. Μια προσέγγιση που χρησιμοποιείται συχνά είναι η μελέτη συν-έκφρασης γονιδίων μέσω της οποίας συνδέουμε γονίδια

των οποίων η έκφραση συμμεταβάλλεται σε σημαντικό βαθμό μεταξύ διαφορετικών συνθηκών ή χρονικών σταδίων (η συντονισμένη έκφραση είναι μια πολύ ισχυρή ένδειξη κοινής ρύθμισης, λειτουργίας ή ακόμα και φυσικής αλληλεπίδρασης). Η συμμεταβολή αυτή μπορεί να ποσοτικοποιηθεί με συντελεστές στατιστικής συσχέτισης (βλ. Εξίσωση 7.10, Κεφάλαιο 7) ή με μέτρα πληροφορίας όπως είναι η αμοιβαία πληροφορία (mutual information,  $I$ ) που βασίζεται στην Εντροπία Shannon (βλ. Κεφάλαιο 3). Πιο συγκεκριμένα, η αμοιβαία πληροφορία είναι ένα μέτρο της αμοιβαίας εξάρτησης δύο μεταβλητών. Αν π.χ. για δύο γονίδια  $x$ ,  $y$  έχουμε  $n$  μετρήσεις έκφρασης η αμοιβαία πληροφορία δίνεται από τη σχέση:

$$I(x,y) = \sum_{i=1}^n \sum_{j=1}^n p(x_i, y_j) \log \left( \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \right) \quad 9.1$$

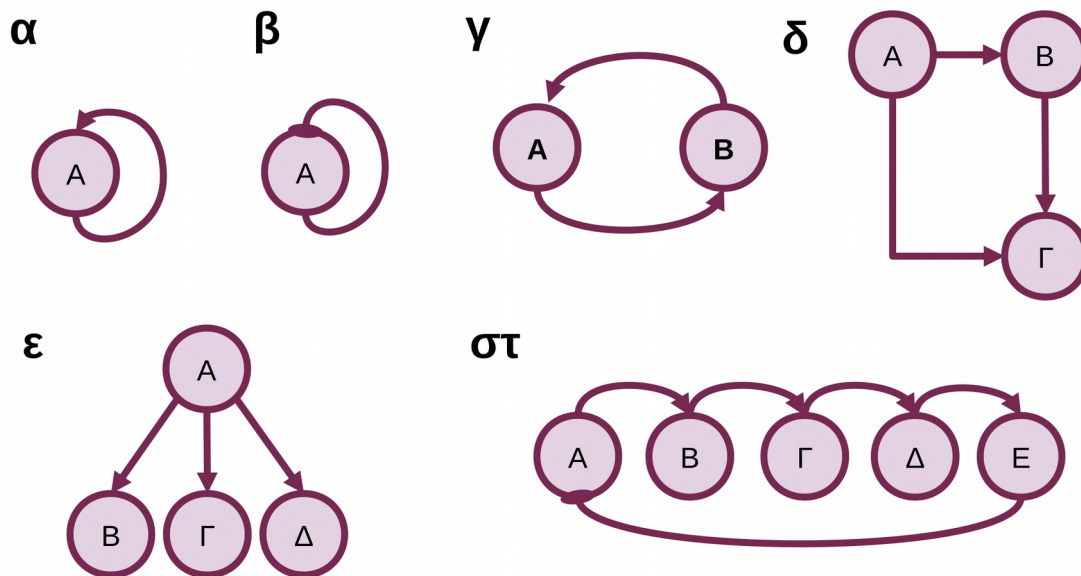
Τόσο συντελεστές συσχέτισης όπως ο Pearson's  $r$  ή μέτρα πληροφορίας όπως η  $I$  μπορούν έτσι να χρησιμοποιηθούν για να δημιουργήσουν έναν πίνακα γειτνίασης και κατά συνέπεια ένα δίκτυο με σταθμισμένες ακμές.

## Δίκτυα στη Βιολογία. Μοτίβα Δικτύων και Μοντελοποίηση

Έχοντας προσεγγίσει σ' ένα πρώτο ποσοτικό επίπεδο τα δίκτυα μπορούμε τώρα να περάσουμε σε πιο πρακτικά ερωτήματα. Σε τι μας χρησιμεύουν; Τι μπορούμε να μάθουμε από αυτά; Με βάση τα όσα έχουμε δει ως τώρα ας προσπαθήσουμε να απαντήσουμε το πρώτο από τα τρία ερωτήματα που θέσαμε στην αρχή του κεφαλαίου. Θυμηθείτε ότι αυτό ήταν:

*Πώς θα προβλέψουμε τη διαμόρφωση των συγκεντρώσεων των προϊόντων ενός δικτύου βιοχημικών αντιδράσεων από τη δομή του δικτύου;*

Η απάντηση προϋποθέτει καταρχάς τη γνώση του μεταβολικού δικτύου ή για την ακρίβεια τη γνώση του μέρους εκείνου του δικτύου που αφορά τους μεταβολίτες που μας ενδιαφέρουν. Τμήματα ενός ευρύτερου δικτύου έχουν συχνά χαρακτηριστική δομή και εμφανίζουν ιδιότητες αυτόνομων υπο-συστημάτων και για το λόγο αυτό τα ονομάζουμε "μοτίβα δικτύων" (network motifs). Τα μοτίβα δικτύων έχουν ιδιότητες μοτίβων, με την έννοια ότι η δομή τους επανεμφανίζεται σε διάφορα μέρη του γενικότερου δικτύου (Segal et al., 2003). Επιπλέον, παρουσιάζουν τυπική, δυναμική συμπεριφορά που είναι εύκολο να μελετηθεί, να αναλυθεί σε συνιστώσες και ως εκ τούτου να μοντελοποιηθεί. Η μοντελοποίηση περιλαμβάνει την περιγραφή των μοτίβων μέσα από συστήματα διαφορικών εξισώσεων και την επίλυσή τους με τη βοήθεια (τις περισσότερες φορές) αριθμητικών προσεγγίσεων μέσω υπολογιστή. Ωστόσο, για τους σκοπούς αυτού του κεφαλαίου θα παραλείψουμε το μέρος των εξισώσεων και θα επικεντρωθούμε στα ποιοτικά χαρακτηριστικά αυτών των συστημάτων. Στην Εικόνα 9.5 παρουσιάζονται σχηματικά μερικά από τα πιο χαρακτηριστικά μοτίβα που συναντώνται τόσο σε ρυθμιστικά όσο και σε μεταβολικά και σηματοδοτικά βιολογικά δίκτυα (Lee et al., 2002). Η πολυπλοκότητά τους προφανώς αυξάνεται με τον αριθμό των εμπλεκόμενων κόμβων, όπως όμως θα δούμε στις αμέσως επόμενες ενότητες μπορούμε να προβλέψουμε αρκετά πολύπλοκες συμπεριφορές ακόμα και από μικρά μοτίβα με δύο ή τρεις κόμβους.



**Εικόνα 9.5:** Βασικά μοτίβα μεταγραφικής ρύθμισης. α) Αυτο-ρύθμιση με ενεργοποίηση β) Αυτο-ρύθμιση με καταστολή (βρόχος ανάδρασης) γ) Βρόχος δύο συνιστωσών δ) Πρόδρομη ρύθμιση ενεργοποίησης (feed-forward stimulation) ε) Μοτίβο μοναδικής εισόδου στ) Αλυσίδα ρύθμισης με ανάδραση (Lee et al., 2002).

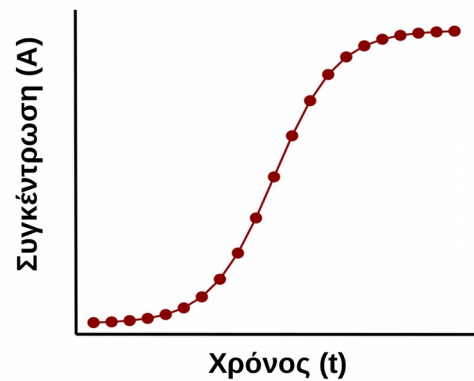
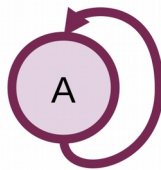
## Ανατροφοδότηση (feedback)

Μοτίβα ανατροφοδότησης προκύπτουν όταν ένας από τους κόμβους μιας αλυσίδας αλληλοσυνδεόμενων κόμβων συνδέεται απευθείας με έναν από τους προηγούμενους κόμβους της αλυσίδας. Πρόκειται για το δικτυακό ανάλογο ενός βραχυκυκλώματος όπως φαίνεται στην Εικόνα 9.6α. Στην απλούστερη περίπτωση, ο ίδιος ο κόμβος μπορεί να αλληλεπιδρά με τον εαυτό του, σ' ένα μοτίβο που ονομάζεται αλλιώς και "αυτο-ρύθμιση" (self-regulation) (Εικόνα 9.5α). Τα μοτίβα ανατροφοδότησης είναι πολύ συχνά στα μεταβολικά δίκτυα όπου συναντούμε ένζυμα των οποίων η ενεργότητα ρυθμίζεται από τα προϊόντα τους ή προϊόντα αντιδράσεων επόμενων σταδίων του μεταβολικού μονοπατιού. Για παράδειγμα, στη γλυκόλυση, η 6-φωσφορική-γλυκόζη (6GP), που είναι το προϊόν της δράσης της εξοκινάσης, αναστέλλει το ένζυμο όταν βρεθεί σε υψηλά επίπεδα. Η ανατροφοδότηση είναι σ' αυτήν την περίπτωση αρνητική ή ανασταλτική (negative feedback). Η σχέση αυτή φαίνεται στην Εικόνα 9.6β όπου συγκρίνονται τα επίπεδα 6GP με ή χωρίς αρνητική ανατροφοδότηση του ενζύμου της εξοκινάσης, με μια προφανή υστέρηση στην περίπτωση της αρνητικής ανατροφοδότησης.

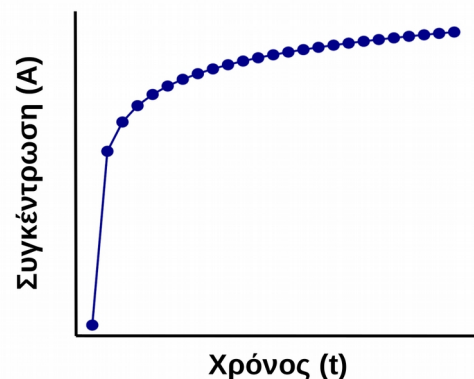
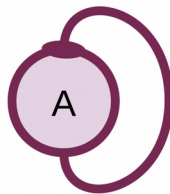
Η θετική ανατροφοδότηση είναι αντίστοιχα το μοτίβο δικτύου στο οποίο η αλληλεπίδραση μεταξύ "θυγατρικού" και "μητρικού" κόμβου είναι θετική. Το μεταβολικό ανάλογο θα ήταν η περαιτέρω ενεργοποίηση ενός ενζύμου από το προϊόν του (ή το προϊόν μιας από τις επόμενες αντιδράσεις του μονοπατιού). Μοτίβα θετικής ανατροφοδότησης είναι σπάνια στα μεταβολικά

δίκτυα (μπορείτε να σκεφτείτε γιατί;) αλλά παρατηρούνται σε κάποιες περιπτώσεις ρυθμιστικών δικτύων αυτο-ενεργοποίησης. Ένα χαρακτηριστικό παράδειγμα σηματοδοτικού-ρυθμιστικού μοτίβου θετικής ανατροφοδότησης είναι η ενεργοποίηση του μεταγραφικού παράγοντα NF-κΒ από την κυττοκίνη TNF-α. Ο NF-κΒ, με τη σειρά του, επάγει στη συνέχεια την περαιτέρω έκφραση του TNF-α. Μοτίβα θετικής ανατροφοδότησης είναι χαρακτηριστικά στις περιπτώσεις που το σύστημα επιτείνει ένα μήνυμα (signal amplification) καθώς είναι ο ταχύτερος τρόπος δημιουργίας ενός βρόχου ανάδρασης (feedback loop) που αυτο-τροφοδοτείται.

α



β



**Εικόνα 9.6:** Μοτίβο ανατροφοδότησης α) Θετική ανατροφοδότηση β) Αρνητική ανατροφοδότηση και οι αντίστοιχες διαφορετικές δυναμικές συμπεριφορές του μοτίβου. Στην πρώτη περίπτωση η θετική ανατροφοδότηση οδηγεί σε μια σιγμοειδή αύξηση της συγκέντρωσης, ενώ στη δεύτερη τα αρχικά ποσοστά αυξάνονται με όλο και μικρότερο ρυθμό λόγω της καταστολής από την αρνητική ανατροφοδότηση.

## Πρόδρομη Επαγωγή (feed-forward loops)

Οι βρόχοι πρόδρομης επαγωγής είναι μοτίβα δικτύων που περιλαμβάνουν τρεις κόμβους και αποτελούν το πιο συχνά απαντώμενο μοτίβο ρυθμιστικών δικτύων. Στην Εικόνα 9.7 φαίνονται τα βασικά χαρακτηριστικά μοτίβων πρόδρομης επαγωγής, η βασική δομή των οποίων είναι η

αλληλεπίδραση ενός μητρικού κόμβου X με ένα θυγατρικό κόμβο Z με δύο τρόπους: α) απευθείας και β) μέσω ενός ενδιάμεσου κόμβου Y. Στη συνεκτική (coherent) πρόδρομη επαγωγή τόσο η άμεση όσο και η έμμεση αλληλεπίδραση έχουν το ίδιο αποτέλεσμα (π.χ. ενεργοποίηση του Z από το X αλλά και από το Y) ενώ στη μη-συνεκτική (incoherent) η άμεση και η έμμεση αλληλεπίδραση έχουν διαφορετικό αποτέλεσμα.

### Συνεκτική πρόδρομη επαγωγή και ρυθμιστική μνήμη

Φαινομενικά και οι δύο τύποι πρόδρομης επαγωγής φαίνονται να μην έχουν ιδιαίτερο ενδιαφέρον. Στην περίπτωση της συνεκτικής ρύθμισης οι δύο συγκλίνουσες αλληλεπιδράσεις φαίνονται πλεονασματικές, καθώς μία από τις δύο θα αρκούσε για την ενεργοποίηση του Z, ενώ στην περίπτωση της μη-συνεκτικής φαίνονται να δημιουργούν πρόβλημα καθώς το σύστημα δεν μπορεί να αποφασίσει αν θα ενεργοποιήσει το Z ή όχι (Εικόνες 9.7α και β, αριστερό μέρος). Στην πραγματικότητα τα πράγματα δεν είναι τόσο απλά εξαιτίας της δυναμικής του συστήματος, που θα πρέπει να το φανταστούμε να εξελίσσεται στο χρόνο και όχι σε μια στατική κατάσταση. Στην Εικόνα 9.7α φαίνεται η μοντελοποίηση ενός μοτίβου συνεκτικής πρόδρομης επαγωγής. Το X αρχικά ενεργοποιεί άμεσα τα Y και Z, ωστόσο στη συνέχεια αποσύρεται από το σύστημα. Μπορείτε να φανταστείτε ένα μεταγραφικό παράγοντα που αποικοδομείται ταχύτατα, έχει μικρό χρόνο ημιζωής ή που πρέπει να κατασταλλεί για κάποιο εξωτερικό λόγο. Το Z δεν ενεργοποιείται πλέον από τον X, η παρουσία του Y ωστόσο εξασφαλίζει ότι το Z θα διατηρηθεί σε υψηλά επίπεδα για δεδομένο χρονικό διάστημα (που εξαρτάται προφανώς από φυσικοχημικές παραμέτρους των αντιδράσεων). Το σύστημα επιδεικνύει έτσι χαρακτηριστικά “μνήμης” και επιτρέπει τη διατήρηση του μηνύματος που εκκίνησε ο παράγοντας X, ακόμα και όταν αυτός έχει πάψει να αποτελεί ενεργό μέρος του (Alon, 2007).

### Μη-συνεκτική πρόδρομη επαγωγή και ρυθμιστικοί διακόπτες

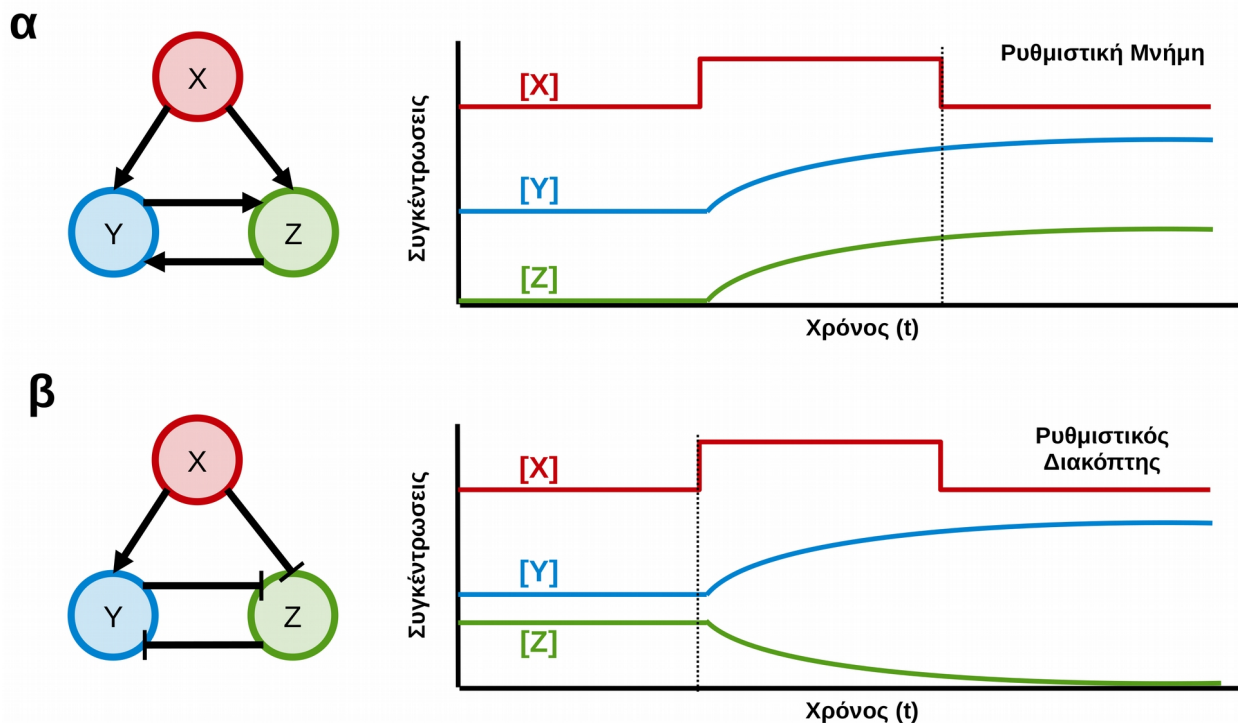
Στην περίπτωση της μη συνεκτικής ρύθμισης, το μοτίβο της πρόδρομης επαγωγής αποδίδει στο σύστημα μια ακόμα πιο ενδιαφέρουσα ιδιότητα. Στην Εικόνα 9.7β, το X ενεργοποιεί το Y και το Z τα οποία όμως μεταξύ τους έχουν σχέσεις αλληλοκαταστολής, δηλαδή το Z καταστέλλει το Y και αντίστροφα. Ανάλογα με τις κινητικές παραμέτρους του συστήματος και το ποιο από τα Z ή Y θα ενεργοποιηθεί ταχύτερα από το X, το σύστημα μπορεί να περάσει σε μια κατάσταση με υψηλό Z ή Y. Αν επίσης στο σύστημα προϋπάρχουν σε κάποια επίπεδα τα Y, Z τότε κάτω από συγκεκριμένες συνθήκες τα σχετικά τους επίπεδα μπορούν να αντιστραφούν με το μοτίβο πρόδρομης επαγωγής ν' αποδίδει στο σύστημα χαρακτηριστικά ενός ρυθμιστικού “διακόπτη” (Alon, 2007).

Τα μοτίβα πρόδρομης επαγωγής ανήκουν σε μια ευρύτερη κατηγορία μοτίβων που ονομάζονται πλήρως διασυνδεδεμένες τριάδες (fully-connected triads, FCT) (Faucon et al., 2014). Οι FCT αποτελούν το πιο διαδεδομένο μοτίβο ρυθμιστικών δικτύων σε όλους τους γνωστούς οργανισμούς, λόγω της εξαιρετικής τους ευελιξίας σε ό,τι αφορά την απόδοση στο σύστημα

ιδιοτήτων όπως η ρυθμιστική μνήμη, οι ρυθμιστικοί διακόπτες, η μετατροπή χρονικών σημάτων σε τοπικά φαινόμενα κλπ (Stergachis et al., 2014).

**Ερώτηση:** Πόσα μοτίβα FCT μπορείτε να σχεδιάσετε; Ποια από αυτά περιμένετε να έχουν τις πιο ενδιαφέρουσες ιδιότητες;

Όλα τα παραπάνω αποτελούν χαρακτηριστικές περιπτώσεις αναδυόμενων ιδιοτήτων (emergent properties), γενικότερων δηλαδή χαρακτηριστικών που εμφανίζονται σ' επίπεδο συστήματος από φαινομενικά απλούστερα στοιχεία (όπως είναι οι αλληλεπιδράσεις μόλις τριών παραγόντων). Στη συνέχεια αυτού του κεφαλαίου θα δούμε πώς αναδυόμενες ιδιότητες μεγαλύτερης τάξης μεγέθους μπορούν να παρατηρηθούν μέσω της μελέτης των δικτύων στο σύνολό τους. Πριν απ' αυτό όμως, θα πρέπει να εξετάσουμε μερικές βασικές μαθηματικές ιδιότητες των δικτύων.



**Εικόνα 9.7:** Δύο μοτίβα πρόδρομης ρύθμισης με πολύ διαφορετικά λειτουργικά χαρακτηριστικά. α) Πρόδρομη συνεκτική επαγωγή σε μια πλήρως συνδεδεμένη τριάδα (X,Y,Z). Το X ενεργοποιεί οποιοδήποτε από τα Y, Z τα οποία λόγω της μεταξύ τους αλληλο-επαγωγικής σχέσης παραμένουν ενεργά ακόμα και μετά την απόσυρση του X από το σύστημα (Ρυθμιστική Μνήμη). β) Σε μια μη-συνεκτική επαγωγή τα Y, Z έχουν αλληλο-κατασταλτική σχέση που οδηγεί στην εναλλαγή της ενεργότητάς τους. Το Z που αρχικά ήταν ενεργό καταστέλλεται τόσο από το X όσο και από το Y που το X έχει ενεργοποιήσει. Απλή στιγμιαία ενεργοποίηση του X αντιστρέφει αυτόματα την ισορροπία Y, Z (Ρυθμιστικός Διακόπτης). Η Εικόνα βασίζεται στις αντίστοιχες απεικονίσεις από το (Alon, 2007).



## Μαθηματικό Ιντερμέδιο II. Εισαγωγή στη Θεωρία Δικτύων

Έχοντας εξετάσει τρόπους μοντελοποίησης στο επίπεδο των μοτίβων δικτύων θα περάσουμε τώρα στην ανάλυση γενικότερων στατιστικών ιδιοτήτων των δικτύων. Μελετούμε τις στατιστικές ιδιότητες ενός δικτύου όχι μόνο επειδή μας επιτρέπουν να το χαρακτηρίσουμε και να διακρίνουμε μεταξύ διαφορετικών ειδών, αλλά κυρίως επειδή συγκεκριμένα στατιστικά χαρακτηριστικά του μπορούν να μας δώσουν πληροφορία για τον τρόπο με τον οποίο το δίκτυο έχει προκύψει μέσα από το συνδυασμό των συστατικών του στοιχείων. Στη βιολογία, η έννοια αυτή της “ανάδυσης” ιδιοτήτων έχει ιδιαίτερη σημασία καθώς τα βιολογικά συστήματα διαμορφώνονται μέσω της διαδικασίας της εξέλιξης. Σ’ αυτήν την ενότητα θα περιγράψουμε μερικά από τα βασικά ποσοτικά χαρακτηριστικά που μας επιτρέπουν να περιγράψουμε στατιστικά ένα δίκτυο. Καθώς το σύνολο των ιδιοτήτων των δικτύων αποτελούν ένα ολόκληρο επιστημονικό πεδίο των μαθηματικών που ονομάζεται θεωρία γράφων (graph theory)<sup>1</sup> (Diestel, 2006), θα επικεντρωθούμε αρχικά στα πιο σημαντικά από αυτά και στη συνέχεια θα δούμε πώς μπορούμε να τα χρησιμοποιήσουμε για τη στατιστική ανάλυση των βιολογικών δικτύων.

### Μέγεθος Δικτύου

Ορίζουμε το μέγεθος ενός δικτύου ως τον ακέραιο αριθμό  $N$  που ισούται με το πλήθος των κόμβων που αυτό περιέχει. Το κατά πόσο ένα δίκτυο θεωρείται “μεγάλο” ή “μικρό” σχετίζεται με το είδος των αντικειμένων που περιγράφει. Βιολογικά δίκτυα αλληλεπίδρασης πρωτεϊνών μπορεί να περιέχουν ακόμα και εκατοντάδες κόμβους. Ένα άλλο μέτρο του μεγέθους του δικτύου είναι η διάμετρος, η οποία λαμβάνει υπόψη όχι μόνο τον αριθμό των κόμβων αλλά και το βαθμό διασύνδεσής τους.

### Πυκνότητα Δικτύου (density)

Πέρα από το συνολικό μέγεθος ενός δικτύου, που αποτελεί μια εκτατική ιδιότητα, μεγάλη σημασία έχει και ο βαθμός, στον οποίο τα στοιχεία του δικτύου είναι διασυνδεδεμένα μεταξύ τους. Ένα δίκτυο μεγάλου μεγέθους (μεγάλου αριθμού κόμβων) μπορεί να έχει ελάχιστες συνδέσεις (ακμές) μεταξύ των στοιχείων του, την ίδια στιγμή που ένα πολύ μικρότερο μπορεί να είναι πολύ πυκνότερο. Την έννοια αυτή της πυκνότητας του δικτύου την ορίζουμε ποσοτικά όσο το πηλίκο των πραγματικών συνδέσεων (ακμών) που υπάρχουν στο δίκτυο προς τον αριθμό των συνδέσεων που θα είχε ένα πλήρως διασυνδεδεμένο (fully-connected) δίκτυο ίσου μεγέθους. Ένα πλήρως διασυνδεδεμένο δίκτυο είναι ένα δίκτυο στο οποίο κάθε στοιχείο συνδέεται άμεσα με καθένα από τα υπόλοιπα. Ο αριθμός των συνδέσεων ενός πλήρως διασυνδεδεμένου δικτύου είναι:

<sup>1</sup> Θυμηθείτε ότι τα δίκτυα από μαθηματική άποψη είναι γράφοι.

$$E_{\text{exp}}(N) = \frac{N(N-1)}{2} \quad 9.2$$

εκτός αν υπάρχει η δυνατότητα αυτο-ρύθμισης (αλληλεπίδρασης δηλαδή του κάθε κόμβου με τον εαυτό του) οπότε ισχύει:

$$E_{\text{exp}}(N) = \frac{N(N+1)}{2} \quad 9.3$$

Ανάλογα με την περίπτωση που αντιστοιχεί στο δίκτυο που μελετούμε, η πυκνότητα  $d$  του δικτύου δίνεται από το λόγο:

$$d = \frac{E_{\text{obs}}}{E_{\text{exp}}} \quad 9.4$$

Όπου  $E_{\text{obs}}$  είναι το συνολικό πλήθος των ακμών του δικτύου. Η τιμή  $d$  κυμαίνεται προφανώς μεταξύ 0 (για ένα δίκτυο χωρίς ακμές) και 1 (για το πλήρως διασυνδεδεμένο) κι είναι ενδεικτική του βαθμού με τον οποίο αλληλεπιδρούν τα στοιχεία του δικτύου.

## Βαθμός Κόμβου (node degree)

Ορίζουμε βαθμό (degree) ενός κόμβου το σύνολο των ακμών με τις οποίες διασυνδέεται με άλλους κόμβους στο δίκτυο. Ανάλογα με το αν το δίκτυο είναι κατευθυνόμενο ή όχι, ο βαθμός ενός δεδομένου κόμβου μπορεί να διακριθεί σε βαθμό εισερχόμενων και εξερχόμενων ακμών. Το άθροισμα των δύο ισούται σε αυτήν την περίπτωση με το συνολικό βαθμό του κόμβου:

$$D(n) = I(n) + O(n) \quad 9.5$$

όπου  $D(n)$  είναι ο βαθμός του κόμβου  $n$  διακρινόμενος σε εισερχόμενες ( $I$ ) και εξερχόμενες ( $O$ ) ακμές. Η μελέτη των βαθμών των κόμβων μπορεί να μας δώσει άμεσα πολλές πληροφορίες για συγκεκριμένα ιδιαίτερα στατιστικά χαρακτηριστικά του δικτύου. Για παράδειγμα, ο απλός υπολογισμός του βαθμού όλων των κόμβων ενός δικτύου μας βοηθάει να εντοπίσουμε τους κόμβους που είναι περισσότερο διασυνδεδεμένοι στο δίκτυο, εκείνα δηλαδή τα στοιχεία που αλληλεπιδρούν πιο έντονα με τα υπόλοιπα. Θυμηθείτε το ερώτημα από την αρχή αυτού του κεφαλαίου:

*Πώς μπορούμε να εντοπίσουμε τους πιο σημαντικούς μεταγραφικούς παράγοντες σ' ένα δίκτυο μεταγραφικής ρύθμισης γονιδίων;*

Η απάντηση σ' αυτήν την ερώτηση μπορεί να προκύψει από μια απλή ανάλυση των βαθμών των κόμβων του δικτύου. Οι μεταγραφικοί παράγοντες με το μεγαλύτερο βαθμό (ή με το μεγαλύτερο βαθμό εξερχόμενων ακμών αν το δίκτυο είναι κατευθυνόμενο) θα είναι αυτοί που θα έχουν και τη μεγαλύτερη επίδραση στο σύστημα. Φυσικά η έννοια της "σημασίας" για μια διαδικασία τόσο πολύπλοκη όσο η μεταγραφική ρύθμιση, πιθανότατα να μην ερμηνεύεται τόσο απλά και να υπάρχουν κι άλλα στοιχεία που θα πρέπει κανείς να αναλογιστεί. Όπως θα δούμε στη συνέχεια, υπάρχουν διάφορα μέτρα ποσοτικοποίησης του κατά πόσο ένας κόμβος είναι "κεντρικός" σε ένα δίκτυο, όμως ο βαθμός των κόμβων είναι μια πολύ καλή πρώτη προσέγγιση.

## Συντομότερη διαδρομή (shortest path)

Η συντομότερη διαδρομή (shortest path) μεταξύ δύο δεδομένων κόμβων, αντιστοιχεί στη διαδρομή με την οποία μπορεί κανείς να βρεθεί από τον έναν στον άλλον, ακολουθώντας τις ακμές του δικτύου και διατρέχοντας τον ελάχιστο αριθμό κόμβων. Είναι δηλαδή το συντομότερο μονοπάτι μεταξύ δύο κόμβων και ορίζεται ως η διαδοχή των κόμβων μεταξύ των σημείων έναρξης και τερματισμού. Απ' αυτήν την άποψη, η συντομότερη διαδρομή δεν είναι μόνο ένα μέγεθος αλλά ένα διάνυσμα με καθορισμένη σειρά στοιχείων. Πολύ συχνά μας ενδιαφέρει περισσότερο το μήκος του διανύσματος της συντομότερης διαδρομής καθώς μας δίνει ένα βαθμό της εγγύτητας μεταξύ των δύο στοιχείων στο δίκτυο. Μια μικρού μήκους συντομότερη διαδρομή είναι ενδεικτική μιας στενότερης σχέσης μεταξύ των κόμβων. Στην Εικόνα 9.8β για παράδειγμα η διαδρομή μεταξύ A και B μπορεί να είναι 5 ή 3 ανάλογα με τους κόμβους από τους οποίους διερχόμαστε. Η συντομότερη (πράσινη) είναι ίση με 3 κι έτσι συμβολίζουμε τη συντομότερη διαδρομή ως  $\sigma_{A,B}=3$ .

Σε μικρά δίκτυα όπως αυτό της Εικόνας 9.8β, η συντομότερη διαδρομή είναι εύκολο να υπολογιστεί εποπτικά. Σε μεγαλύτερα, ωστόσο, δίκτυα με μερικές εκατοντάδες ή χιλιάδες κόμβους χρειάζεται κανείς να καταφύγει σε πιο συστηματικές μεθόδους για τον υπολογισμό της. Μια αποτελεσματική τέτοια μέθοδος είναι η Αναζήτηση μέσω Πλάτους (Breadth-first search) (Zhou & Hansen, 2006), η οποία υπολογίζει την απόσταση μεταξύ δύο κόμβων A και B, ξεκινώντας από τον κόμβο A και προχωρώντας μέσω διαδοχικών ελέγχων των άμεσων γειτόνων ως το B. Πιο αναλυτικά περιγράφεται στον παρακάτω αλγόριθμο:

### Αλγόριθμος :: BreadthFirstSearch

Δεδομένου ενός πίνακα γειτνίασης  $M[N^2]$

Δεδομένων δύο στοιχείων A, B

Αρχικοποίηση distance=1

(1) Για  $i=1$  έως  $i=N$  ( $i \neq A$ ) # το  $i$  να μην είναι A

Έλεγχε Αν  $M[i,A]=1$ :

Τότε:

$i \rightarrow G[A]$  # γείτονες του A

Αλλιώς:

Συνέχισε

Για  $i=1$  έως  $n$  # αριθμός γειτόνων του A

Έλεγχε Αν  $i=B$ : # αν ταυτίζεται με το B

Τότε:

$d(A,B)=distance$

Τερματισμός

Αλλιώς:

$A=G[A]$  # αντικατάστησε το A με τους γείτονες του A

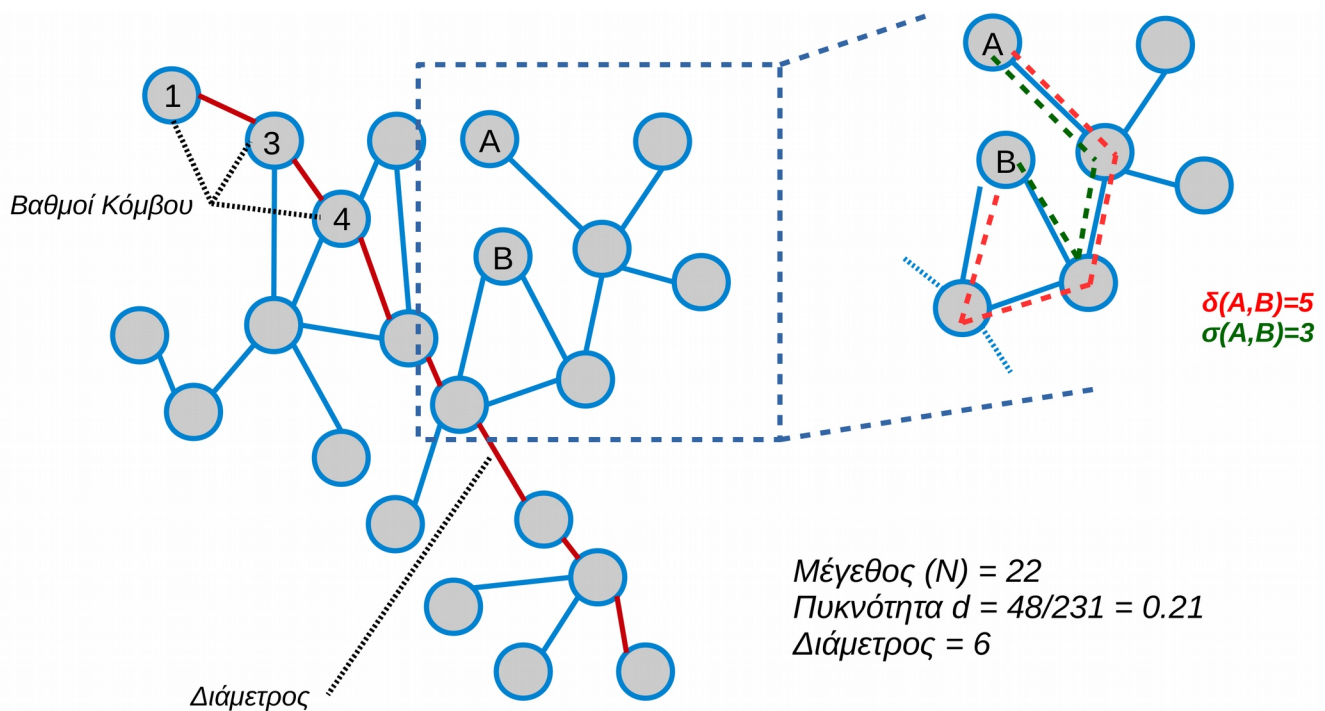
$distance=distance+1$  # αύξησε την απόσταση κατά 1

Επίστρεψε στο (1)

Έλεγχε Αν για κανένα  $i$  δεν ισχύει  $i=B$ .

Τότε distance=[άπειρη]

Ο αλγόριθμος Breadth-first Search ξεκινάει από τους πρώτους γείτονες του A (ελέγχοντας τις τιμές των στοιχείων που βρίσκονται στην ίδια στήλη ή γραμμή με το A στον πίνακα γειτνίασης). Αν ο B είναι μεταξύ αυτών τότε η απόσταση  $\sigma_{A,B}$  είναι 1. Αν ο B δεν είναι πρώτος γείτονας, η αναζήτησή του συνεχίζεται αλλά πλέον μεταξύ των πρώτων γειτόνων του A. Αν ο B είναι γείτονας με κάποιον από αυτούς η απόστασή του από τον A είναι ίση με 2. Αν δεν είναι αναζητούμε το B στους γείτονες των πρώτων γειτόνων των πρώτων γειτόνων του A (αυξάνοντας την απόσταση κατά 1) κ.ο.κ. μέχρι να εντοπιστεί ο B σε κάποια “γενιά” γειτόνων. Ο αριθμός των “κύκλων” που χωρίζουν τους A και B είναι ίσος με την απόστασή τους ενώ αν ο B δεν βρεθεί ποτέ ανάμεσα στους γείτονες του A τότε οι A και B είναι ασύνδετοι κόμβοι με απόσταση  $\sigma_{A,B} = \infty$ .



**Εικόνα 9.8:** Χαρακτηριστικά μεγέθη και ιδιότητες δικτύων. Το δίκτυο της εικόνας αποτελείται από 22 κόμβους των οποίων οι βαθμοί κυμαίνονται μεταξύ 1 και 5. Το σύνολο των ακμών είναι 48 με το μέγιστο δυνατό να είναι  $(22 \times 21)/2 = 231$  κι έτσι η πυκνότητα είναι ίση με 0.21. Στην ένθετη εικόνα φαίνεται η ελάχιστη απόσταση δύο κόμβων A, B. Από τις δύο δυνατές διαδρομές (κόκκινη, πράσινη) μεταξύ A και B η ελάχιστη είναι η πράσινη ( $\sigma(A,B)=3$ ). Η μεγαλύτερη ελάχιστη διαδρομή στο δίκτυο, η οποία αντιστοιχεί στη διάμετρό του, φαίνεται στη μεγαλύτερη εικόνα σκιασμένη με βαθύ κόκκινο και είναι ίση με 6.

Στη γενικότερη μορφή της, που περιλαμβάνει και τα δίκτυα με βάρη, η συντομότερη διαδρομή ορίζεται ως η διαδρομή από το A στο B για την οποία το άθροισμα των βαρών των ακμών διέλευσης είναι το ελάχιστο και το μήκος αυτό αποτελεί το χαρακτηριστικότερο μέγεθος για την απόσταση μεταξύ δύο κόμβων. Η έννοια της εγγύτητας μεταξύ κόμβων του δικτύου έχει σημαντικές

προεκτάσεις, τόσο για χαρακτηριστικά που είναι ειδικά για συγκεκριμένους κόμβους, όσο για γενικότερες ιδιότητες του δικτύου όπως θα δούμε στη συνέχεια.

## Κεντρικότητα Κόμβων (centrality)

Η κεντρικότητα ενός δεδομένου κόμβου είναι ένα πιο εξειδικευμένο μέτρο της σημασίας που έχει ο κόμβος στο δίκτυο. Για τον υπολογισμό της υπάρχουν διάφοροι τρόποι, η έννοια όμως παραμένει η ίδια και σχετίζεται με τη σχετική θέση του κόμβου στο δίκτυο.

Η κεντρικότητα εγγύτητας (closeness centrality) ορίζεται ως το αντίστροφο του αθροίσματος των αποστάσεων του κόμβου απ' όλους τους άλλους κόμβους. Η απόσταση εδώ ορίζεται, όπως είδαμε παραπάνω, ως το μήκος του διανύσματος της συντομότερης διαδρομής. Έτσι η κεντρικότητα εγγύτητας του κόμβου  $A$  σ' ένα δίκτυο με  $N$  κόμβους δίνεται από την εξίσωση:

$$C(A) = \frac{1}{\sum_{i=1}^{N-1} \sigma_{A,i}} \quad 9.6$$

Το συγκεκριμένο μέτρο εγγύτητας είναι μεν δηλωτικό της θέσης του κόμβου στο δίκτυο από άποψη αποστάσεων, αλλά δεν αποδίδει τις διαφορές που μπορεί να υπάρχουν μεταξύ ενός κόμβου, που απλά βρίσκεται κοντά στο κέντρο του δικτύου από γεωμετρική άποψη, και των κόμβων που είναι και κεντρικοί από πλευράς διασυνδεσιμότητας. Ένα μέτρο που επιτρέπει να ποσοτικοποιηθεί αυτή η έννοια είναι η διακεντρικότητα (betweenness centrality). Η διακεντρικότητα ορίζεται ως ο λόγος των συντομότερων διαδρομών που διέρχονται από έναν κόμβο προς το σύνολο των συντομότερων διαδρομών:

$$G(A) = \frac{\sigma_{i,j}(A)}{\sum_{i=1, j=1, i \neq j \neq A}^N \sigma_{i,j}} \quad 9.7$$

Λαμβάνοντας δηλαδή υπόψη μας το σύνολο των συντομότερων διαδρομών που δεν αφορούν την απευθείας σχέση του  $A$  με τους άλλους κόμβους, υπολογίζουμε πόσες από αυτές περιλαμβάνουν τον κόμβο  $A$ . Είναι προφανές ότι η τιμή  $G(A)$  εξαρτάται από τον αριθμό των ζευγών κόμβων (και συνεπώς το μέγεθος του δικτύου) και για το λόγο αυτό η τιμή θα πρέπει να διορθωθεί με βάση το πλήθος των ζευγών. Για μέγεθος δικτύου  $N$  (και εφόσον ο  $A$  δεν συνυπολογίζεται) το πλήθος των ζευγών είναι  $(N-1)(N-2)$  για κατευθυνόμενα και  $(N-1)(N-2)/2$  για μη-κατευθυνόμενα δίκτυα.

Επιστρέφοντας στο ερώτημα των μεταγραφικών παραγόντων με τη μεγαλύτερη σημασία για ένα δίκτυο μεταγραφικής ρυθμίσης θα μπορούσαμε να πούμε πως οι πιο σημαντικοί παράγοντες θα είναι αυτοί με τις μεγαλύτερες τιμές κεντρικότητας. Πράγματι ακόμα κι αν ένας παράγοντας έχει μεγαλύτερο βαθμό από κάποιον με μεγάλη τιμή κεντρικότητας, είναι λογικό να υποθέσουμε πως πιο σημαντικός για το δίκτυο θα είναι ο δεύτερος, ιδιαίτερα αν το μέγεθος του δικτύου είναι αρκετά μεγαλύτερο από το μέγιστο βαθμό κόμβου.

### Διάμετρος Δικτύου (diameter)

Η διάμετρος είναι ένα ακόμα μέτρο του μεγέθους ενός δικτύου, που σχετίζεται όμως όχι μόνο με το πλήθος των κόμβων, αλλά λαμβάνει υπόψη του και τη διασυνδεσιμότητα. Ορίζεται ως η μέγιστη συντομότερη διαδρομή μεταξύ όλων των πιθανών ζευγών του δικτύου και δίνεται από τη σχέση:

$$r(A) = \max(\sigma_{i,j}) \quad 9.8$$

Είναι έτσι το ανάλογο της διαμέτρου ενός γεωμετρικού σχήματος και αποτελεί μια καλή προσέγγιση για το βαθμό “διασποράς” των κόμβων ενός δικτύου. Η ερμηνεία της ωστόσο, χρειάζεται προσοχή καθώς σε περιπτώσεις ακραίων, πολύ απομακρυσμένων κόμβων η τιμή της μπορεί να μην είναι αντιπροσωπευτική.

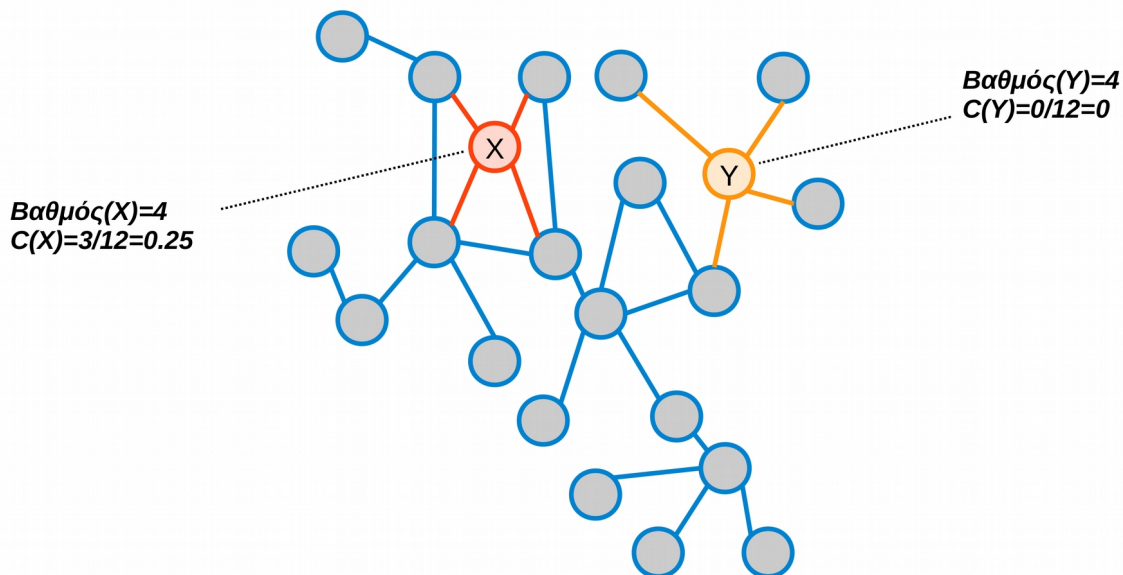
### Συντελεστής συσσωμάτωσης (clustering coefficient) και συνδεσιμότητα (connectedness)

Ο συντελεστής συσσωμάτωσης ενός κόμβου είναι μια πολύ σημαντική στατιστική ιδιότητα που σχετίζεται με το βαθμό στον οποίο οι κόμβοι του δικτύου τείνουν να σχηματίζουν τοπικά υποδίκτυα. Σχηματικά μπορεί κανείς να συλλάβει την έννοια του συντελεστή συσσωμάτωσης για έναν κόμβο A ως τον αριθμό των τριγωνικών σχέσεων που αναπτύσσονται μεταξύ του A και των κόμβων με τους οποίους ο A συνδέεται άμεσα. Στην Εικόνα 9.9 φαίνεται ένα παράδειγμα αυτής της σχέσης. Ο κόμβος A συνδέεται άμεσα με τους B, Γ, Δ και E. Ο συντελεστής συσσωμάτωσης θα είναι ψηλότερος όσο περισσότερο οι B, Γ, Δ και E τείνουν να είναι διασυνδεδεμένοι και μεταξύ τους, δημιουργούν δηλαδή τριγωνικές σχέσεις με το A. Ο συντελεστής συσσωμάτωσης σχετίζεται άμεσα με την έννοια της κλίκας (clique), η οποία περιγράφει ένα δίκτυο (ή ένα τμήμα δικτύου) του οποίου τα στοιχεία είναι πλήρως και άμεσα διασυνδεδεμένα μεταξύ τους. Θυμηθείτε ότι για ένα πλήρως διασυνδεδεμένο δίκτυο ορίσαμε την πυκνότητα ως το πηλίκο του αριθμού παρατηρούμενων συνδέσεων προς το μέγιστο δυνατό αριθμό τους. Ο συντελεστής συσσωμάτωσης για έναν δεδομένο κόμβο A αντιστοιχεί έτσι σε μια τοπική τιμή πυκνότητας για την “γειτονιά” του η οποία ορίζεται ως το σύνολο των άμεσα διασυνδεδεμένων με το A κόμβων. Έτσι αν ο A έχει βαθμό k, τότε η μέγιστη τιμή συνδέσεων στην γειτονιά του A είναι (για μη κατευθυνόμενο δίκτυο χωρίς αυτό-ρύθμιση)  $E = k(k-1)$ . Ο συντελεστής συσσωμάτωσης C(A) τότε δίνεται από την τιμή:

$$C(A) = \frac{\sum_{i=1, j=1, i \neq j}^k \text{edges}_{i,j}}{k(k-1)} \quad 9.9$$

Ο συντελεστής συσσωμάτωσης παίρνει τιμές μεταξύ 0 και 1 και έχει χαρακτηριστικά πιθανότητας. Αν για παράδειγμα ένας κόμβος A με βαθμό k έχει συντελεστή συσσωμάτωσης C(A), τότε η πιθανότητα δύο τυχαία επιλεγμένων γειτόνικων κόμβων του A να είναι συνδεδεμένοι και μεταξύ τους είναι ίση με  $p = C(A)$ . Από την εξίσωση 9.9 είναι προφανές ότι ο συντελεστής

συσσωμάτωσης δεδομένου κόμβου εξαρτάται από το βαθμό του. Η σχέση με την οποία τα δύο μεγέθη συσσωμάτωσης είναι πολύ σημαντική για το είδος του δικτύου. Στα βιολογικά δίκτυα η σχέση του βαθμού  $d(A)$  και του συντελεστή συσσωμάτωσης  $C(A)$  είναι αντίστροφη. Όσο δηλαδή αυξάνεται ο βαθμός των κόμβων τόσο μειώνεται ο συντελεστής συσσωμάτωσης τους. Μια τέτοια σχέση αντανάκλα την τάση των κόμβων μεγάλου βαθμού να "αποκεντρώνουν" τις πυκνές συνδέσεις σε μικρότερου βαθμού κόμβους και είναι ενδεικτική μιας ιεραρχικής δομής. Σε δίκτυα που είναι σχεδιασμένα για να προσδιορίσουν ιεραρχικές σχέσεις (όπως π.χ. η οργάνωση των εργαζόμενων μιας εταιρείας) η αντίστροφη σχέση μεταξύ συντελεστή συσσωμάτωσης και βαθμού κόμβων είναι ακόμα πιο έντονη.



**Εικόνα 9.9:** Συντελεστής συσσωμάτωσης κόμβων. Οι δύο κόμβοι X, Y του σχήματος έχουν τον ίδιο βαθμό (4) όμως διαφέρουν πολύ σε ότι αφορά το συντελεστή συσσωμάτωσης. Τρεις από τους τέσσερις γείτονες του X συνδέονται μεταξύ τους κι έτσι με βάση το μέγιστο αριθμό συνδέσεων  $k(k-1)=12$  ο συντελεστής συσσωμάτωσης για το X είναι ίσος με  $3/12=0.25$ . Η αντίστοιχη τιμή για τον Y είναι 0 καθώς μεταξύ των γειτόνων του δεν υπάρχει καμία σύνδεση.

**Ερώτηση:** Σκεφτείτε λίγο, ποια μπορεί να είναι η φυσική σημασία αυτής της αντίστροφης σχέσης  $d(A)$  και  $C(A)$  σε βιολογικά δίκτυα;

Μια γενικευμένη έκφραση του C είναι ο μέσος συντελεστής συσσωμάτωσης  $\langle C \rangle$  που αντιστοιχεί στη μέση τιμή των τιμών C για όλους τους κόμβους N ενός δικτύου:

$$\langle C \rangle = \frac{\sum_{A=1}^N C(A)}{N} \quad 9.10$$

Ο μέσος συντελεστής συσσωμάτωσης είναι ένα μέτρο της συνδεσιμότητας (connectedness) ενός δικτύου. Ένα δίκτυο για το οποίο ισχύει  $\langle C \rangle = 0$ , πρόκειται για ένα σύνολο πλήρως ασύνδετων

κόμβων, ενώ από την άλλη για  $\langle C \rangle = 1$ , το δίκτυο είναι πλήρως συνδεδεμένο με όλους τους κόμβους να συνδέονται μεταξύ τους με όλους τους δυνατούς τρόπους, να συνθέτουν δηλαδή μια κλίκα.

### Συντελεστής συσχέτισης (correlation coefficient) και καταταξιμότητα (assortativity)

Ο συντελεστής συσχέτισης κόμβων είναι μια ποσότητα που σχετίζεται άμεσα με την καταταξιμότητα (assortativity), μια βασική ιδιότητα του κάθε δικτύου, που αντανάκλα την τάση των στοιχείων του να αλληλεπιδρούν μεταξύ τους με συγκεκριμένο τρόπο. Παρόλο που ως ποσότητα, ο συντελεστής συσχέτισης κόμβων, που περιγράφεται εδώ διαφέρει από αυτόν που συζητήσαμε στο Κεφάλαιο 7, οι δύο έννοιες μοιράζονται ποιοτικά χαρακτηριστικά. Ο συντελεστής συσχέτισης κόμβων ισούται με το συντελεστή συσχέτισης Pearson του βαθμού των κόμβων μεταξύ άμεσα διασυνδεδεμένων κόμβων. Σε αναλογία με την εξίσωση 7.10 δίνεται από τη σχέση:

$$r = \frac{N \sum_{i=1}^N d_{i,1} d_{i,2} - \sum_{i=1}^N d_{i,1} \sum_{i=1}^N d_{i,2}}{\sqrt{\left( N \sum_{i=1}^N d_{i,1}^2 - \left( \sum_{i=1}^N d_{i,1} \right)^2 \right) \left( N \sum_{i=1}^N d_{i,2}^2 - \left( \sum_{i=1}^N d_{i,2} \right)^2 \right)}} \quad 9.11$$

όπου με  $d$  συμβολίζουμε το βαθμό ενός κόμβου. Η πολυπλοκότητα της σχέσης 9.11 δε θα πρέπει να μας ανησυχεί καθώς υπολογιστικά προγράμματα μπορούν να υπολογίσουν αυτήν την ποσότητα με πολύ μεγάλη ταχύτητα. Αυτό που έχει σημασία είναι η θεωρητική της σημασία. Τιμές  $r > 0$  είναι δηλωτικές μιας θετικής σχέσης μεταξύ του βαθμού διασυνδεόμενων κόμβων. Αυτό σημαίνει ότι κόμβοι με παρόμοιο βαθμό τείνουν να συνδέονται μεταξύ τους (θυμηθείτε ότι με ανάλογο τρόπο τιμές  $r > 0$  στα επίπεδα έκφρασης γονιδίων στο Κεφάλαιο 7 ήταν δηλωτικές μιας ομοιότητας μεταξύ των προτύπων έκφρασής τους). Για  $r > 0$  οι κόμβοι με μεγάλο βαθμό θα συνδέονται με άλλους κόμβους με μεγάλο βαθμό περισσότερο από όσο θα περίμενε κανείς να συμβαίνει στην τύχη. Αντίθετα, τιμές  $r < 0$  υποδηλώνουν ότι η τάση είναι αντίστροφη και κόμβοι με μεγάλο βαθμό τείνουν να συνδέονται με κόμβους με μικρό βαθμό, υπάρχει δηλαδή μια αντίθετη προτίμηση κατά τη σύνδεση των κόμβων.

Η έννοια της καταταξιμότητας είναι έτσι στενά συνδεδεμένη με τάσεις προτίμησης που διαμορφώνουν τη διασύνδεση των κόμβων ενός δικτύου και ως εκ τούτου τη μορφή του. Σχετίζεται άμεσα με χαρακτηριστικά που έχουν να κάνουν με την τυχαιότητα και τις αποκλίσεις από αυτήν και που θα εξετάσουμε στη συνέχεια (Mark Newman, Barabási & Watts, 2006).

### Στατιστικές Ιδιότητες Δικτύων

Ως εδώ αναφερθήκαμε στα βιολογικά δίκτυα με ποιοτικούς μόνο όρους. Από την προηγούμενη ενότητα όμως έχουμε περιγράψει αναλυτικά και τα ποσοτικά στοιχεία εκείνα στοιχεία,



που μπορούν να μας βοηθήσουν να αναλύσουμε τις γενικότερες ιδιότητες των βιολογικών δικτύων. Οι στατιστικές ιδιότητες των δικτύων μπορούν να μας οδηγήσουν σε συμπεράσματα για διάφορα σημαντικά στοιχεία, όπως είναι το πώς δημιουργούνται, ποια είναι η δομή τους και ποια πλεονεκτήματα προσφέρει αυτή, καθώς και πώς αντανakλάται στη δομή των δικτύων η βασική ιδιότητα των βιολογικών συστημάτων που είναι η δυναμική τους εξέλιξη στο χρόνο. Με όσα έχουμε συζητήσει ως τώρα, μπορούμε να γυρίσουμε στο δίκτυο των πρωτεϊνικών αλληλεπιδράσεων που είδαμε στην Εικόνα 9.1 και να εξετάσουμε το τελευταίο από τα ερωτήματα που θέσαμε στην αρχή αυτού του κεφαλαίου:

*Κατά πόσο ένα τέτοιο δίκτυο πρωτεϊνικών αλληλεπιδράσεων θα μπορούσε να έχει προκύψει τυχαία;*

Η ερώτηση ακούγεται αρχικά αρκετά φιλοσοφική. Η διαίσθησή μας μας λέει ότι ένα δίκτυο αλληλεπιδράσεων σύνθετων μεγαλομορίων, καθένα από τα οποία έχει εξελιχθεί για εκατομμύρια χρόνια αποκλείεται εξ' ορισμού να είναι τυχαίο. Τι σημαίνει όμως "τυχαίο"; Πώς θα μπορούσε να έχει προκύψει τυχαία ένα δίκτυο εκατοντάδων πρωτεϊνών; Και σε τι θα διέφερε από το δίκτυο της Εικόνας 9.1; Για να απαντήσουμε στο παραπάνω ερώτημα θα χρειαστεί πρώτα να ορίσουμε ένα τυχαίο δίκτυο, να εξετάσουμε τις ιδιότητες που αυτό προβλέπεται να έχει και στη συνέχεια να συγκρίνουμε τη συμπεριφορά του μ' ένα πραγματικό βιολογικό δίκτυο.

## Τυχαία Δίκτυα

Μ' ένα λίγο αυτο-αναφορικό μαθηματικό τρόπο, ονομάζουμε "τυχαίο" οποιοδήποτε δίκτυο, οι συνδέσεις μεταξύ των κόμβων του οποίου ακολουθούν μια τυχαία κατανομή. Η κατανομή αυτή μπορεί να διαφέρει τόσο σε είδος όσο και σε παραμέτρους, ωστόσο τα περισσότερα μοντέλα τυχαίων δικτύων βασίζονται στο "πρωταρχικό" τυχαίο μοντέλο που προτάθηκε από τους Paul Erdős και Alfred Renyi (προς τιμήν των οποίων τα τυχαία δίκτυα ονομάζονται και Erdős-Renyi). Ένα τυχαίο δίκτυο κατά Erdős-Renyi προκύπτει με τον εξής απλό τυχαιοποιημένο αλγόριθμο:

### Αλγόριθμος :: RandNet

Δεδομένου ενός αριθμού στοιχείων  $N$

Δεδομένης μιας πιθανότητας σύνδεσης  $p$  ( $0 \leq p \leq 1$ )

Δημιούργησε έναν πίνακα γειτνίασης  $M[N^2]=0$

Απαρίθμηση : Για  $k=1$  έως  $N(N-1)$

1. Επίλεξε τυχαία ένα ζεύγος στοιχείων  $N(i), N(j)$
2. Επίλεξε τυχαία έναν αριθμό  $r$  στο διάστημα  $[0, 1]$   $\#(0 \leq r \leq 1)$
3. Έλεγχος:     Αν  $r \geq p \rightarrow$  Σύνδεσε τα  $i, j$ ,  $M[i, j]=1$   
                  Αν  $r < p \rightarrow$  Αφησε τη σχέση  $i, j$  ως έχει

4. Επίστρεψε στο 1

Απόδωσε τον πίνακα γειτνίασης  $M$

Τερματισμός

Ο παραπάνω αλγόριθμος παράγει ένα δίκτυο, του οποίου όλες οι συνδέσεις έχουν προκύψει

τυχαία. Το μέγεθος του δικτύου θα είναι  $N$  και η πυκνότητά του θα εξαρτάται από την τιμή της παραμέτρου  $p$  (όσο χαμηλότερη, τόσο πυκνότερο το δίκτυο). Ποια άλλα χαρακτηριστικά μεγέθη μπορούμε να υπολογίσουμε για ένα τέτοιο δίκτυο; Μια βασική στατιστική ιδιότητα που μας ενδιαφέρει είναι η κατανομή των βαθμών των κόμβων του. Με βάση τον τρόπο που σχηματίστηκε το δίκτυο, μπορούμε να υπολογίσουμε την πιθανότητα που έχει ο κάθε κόμβος να έχει βαθμό  $d$ .

Χωρίς να μπορούμε σε λεπτομέρειες, αρκεί να πούμε ότι η πιθανότητα αυτή προσεγγίζεται πολύ ικανοποιητικά από την κατανομή Poisson. Ως εκ τούτου, τα τυχαία δίκτυα έχουν μια πολύ χαρακτηριστική κατανομή βαθμού κόμβων σε σχέση με τη μέση τιμή (όπως διαμορφώνεται από την παράμετρο  $p$ ). Η κατανομή αυτή προβλέπει ότι ο μεγαλύτερος αριθμός των κόμβων θα έχει βαθμό  $d$  κοντά στην μέση τιμή των βαθμών του δικτύου. Ένας τρόπος για να παρουσιάσει κανείς αυτήν την ιδιότητα γραφικά είναι να αναπαραστήσει την αθροιστική κατανομή βαθμών κόμβων σε διπλή λογαριθμική κλίμακα. Η γραφική αυτή αναπαράσταση φαίνεται στην Εικόνα 9.10α και προκύπτει ως εξής:

- Δεδομένων  $N$  κόμβων στο δίκτυο.
- Για κάθε κόμβο  $k$  υπολογίζουμε την ποσότητα  $p(k)=1/d(k)$  όπου  $d(k)$  είναι ο βαθμός του κόμβου.
- Για κάθε κόμβο  $k$  υπολογίζουμε μια τιμή  $n(k)$  που ισούται με το πλήθος των κόμβων που έχουν βαθμό  $d \leq d(k)$ .

Σε διπλή λογαριθμική κλίμακα αναπαριστούμε τα μεγέθη:  $\log(n(k))$  και  $\log(p(k))$ .

Η μορφή της κατανομής αυτής για ένα τυχαίο δίκτυο προσομοιάζει μια φθίνουσα εκθετική κατανομή με μια παράμετρο που είναι ανάλογη του  $p$ .

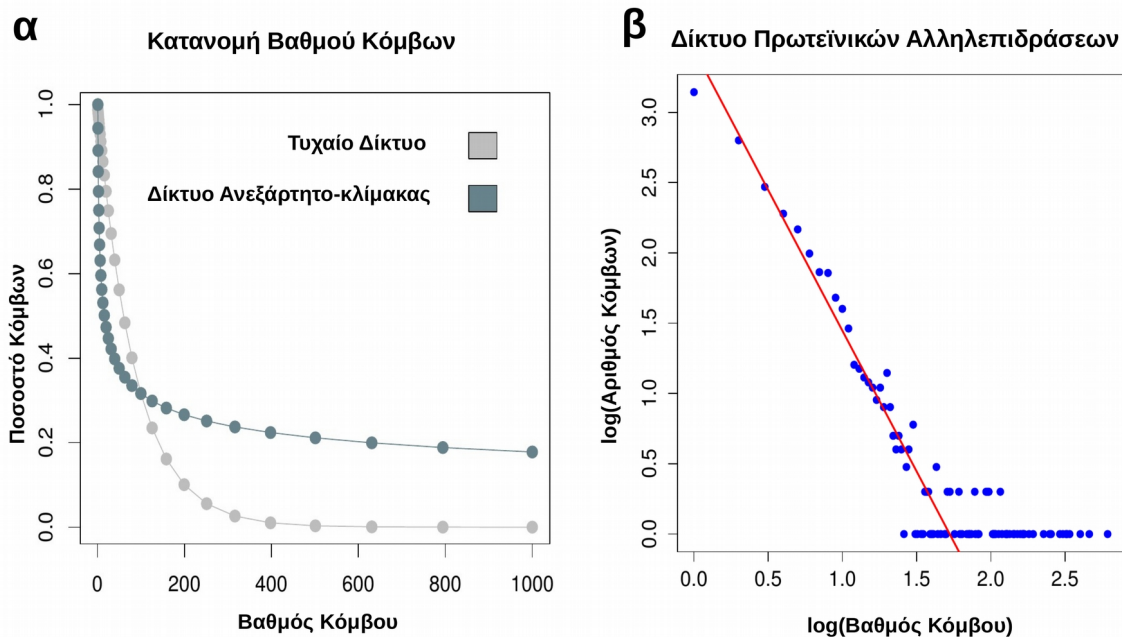
Τι συμβαίνει όμως με το πραγματικό βιολογικό δίκτυο των πρωτεϊνικών αλληλεπιδράσεων;

## Δίκτυα Ανεξάρτητα-κλίμακας (scale-free)

Πριν εφαρμόσουμε την παραπάνω συλλογιστική στο δίκτυο των πρωτεϊνών της Εικόνας 9.1α προσπαθήσουμε να κάνουμε κάποιες προβλέψεις. Αρχικά, όπως αναφέραμε και νωρίτερα, δε θα περιμέναμε το δίκτυο των πρωτεϊνών να συμπεριφέρεται ως τυχαίο κι ως εκ τούτου περιμένουμε μια διαφορετική κατανομή του βαθμού των κόμβων. Τι σχήμα όμως θα έχει αυτή η κατανομή; Σ' ένα τυχαίο δίκτυο δεν υπάρχει κανένας λόγος να περιμένουμε ότι κάποιοι κόμβοι θα έχουν εξαιρετικά μεγάλους ή μικρούς βαθμούς, καθώς από κατασκευής, όλοι οι κόμβοι έχουν την ίδια πιθανότητα διασύνδεσης.

Σ' ένα φυσικό δίκτυο όμως, όπως αυτό των πρωτεϊνικών αλληλεπιδράσεων (και σε πολλά άλλα όπως θα δούμε παρακάτω) κάτι τέτοιο δεν ισχύει. Υπάρχουν πρωτεΐνες που είναι εξαιρετικά σημαντικές για το κύτταρο και που αναμένουμε να αλληλεπιδρούν με μεγάλο αριθμό μορίων. Αντίστοιχα υπάρχουν πρωτεΐνες με πολύ εξειδικευμένη λειτουργία, που εκφράζονται μόνο κάτω από πολύ περιορισμένες συνθήκες και που περιμένουμε να είναι απλώς "περιφερειακές" στο δίκτυο αλληλεπιδράσεων. Αυτό το εύρος των λειτουργιών και η πολυπλοκότητα του συστήματος, μας κάνει να περιμένουμε ότι ένας αριθμός πρωτεϊνών θα έχει πολύ μεγάλο αριθμό αλληλεπιδράσεων και θα αποτελεί έτσι ένα σύνολο υπερ-κόμβων (hubs) στο δίκτυο, συγκεντρώνοντας ένα σημαντικό

ποσοστό των συνολικών ακμών. Πράγματι, ο βαθμός των κόμβων στο πρωτεϊνικό δίκτυο της Εικόνας 9.1 ακολουθεί μια διαφορετική κατανομή, η οποία φαίνεται στην Εικόνα 9.10β και της οποίας το βασικό διαφοροποιητικό χαρακτηριστικό, σε σχέση με την τυχαία κατανομή, είναι η έκταση του δεξιού κάτω μέρους προς σημαντικά μεγαλύτερες τιμές. Η έκταση αυτή αντανακλά ακριβώς την ιδιότητα της ύπαρξης υπερ-κόμβων καθώς αντιστοιχεί σε μια τάση να υπάρχει σημαντικός αριθμός κόμβων με πολύ μεγάλο βαθμό (πολύ μικρό  $p(k)$ ).



**Εικόνα 9.10:** Τυχαία δίκτυα και Δίκτυα ανεξάρτητα-κλίμακας. α) Κατανομή Βαθμού Κόμβων για ένα τυχαίο δίκτυο (γκρι) κι ένα ανεξάρτητο-κλίμακας δίκτυο (γαλάζιο). Η χαρακτηριστική “μακριά ουρά” του δεύτερου είναι ενδεικτική μιας κατανομής νόμου δύναμης, η οποία δίνει ευθεία γραμμή σε διπλή λογαριθμική κλίμακα, όπως φαίνεται στο β) όπου αναπαρίσταται γραφικά η κατανομή βαθμού κόμβων για ένα δίκτυο πρωτεϊνικών αλληλεπιδράσεων (όπως αυτό της Εικόνας 9.1).

Η μορφή αυτής της κατανομής, δείχνει με κάποιον τρόπο να “αδιαφορεί” για το μέγεθος του δικτύου και να μην επιτρέπει τον εύκολο υπολογισμό μιας μέσης τιμής για το βαθμό των κόμβων. Ονομάζουμε τα δίκτυα που ακολουθούν αυτές τις κατανομές βαθμών κόμβων “δίκτυα ανεξάρτητα-κλίμακας” (scale-free networks). Κατανομές που είναι “ανεξάρτητες-κλίμακας” (scale-free) είναι αρκετά γνώριμες σε όσους είναι εξοικειωμένοι με τη μελέτη πολύπλοκων συστημάτων, καθώς προσομοιάζονται από νόμους δύναμης (power-laws) (Clauset, Shalizi & Newman, 2009), ένα είδος κατανομών με εξαιρετικό ενδιαφέρον, για το οποίο αξίζει να κάνουμε μια μικρή (τελευταία) παρέκβαση.

**Ερώτηση:** Με ποια διαδικασία (ανάλογη του RandNet αλγορίθμου) θα μπορούσαμε να δημιουργήσουμε ένα δίκτυο ανεξάρτητο-κλίμακας;

## Μαθηματικό Ιντερμέδιο III. Νόμοι Δύναμης (Power Laws)

Τι κοινό έχουν το διαδίκτυο, το εισόδημα των φορολογούμενων μιας χώρας και ο Moby Dick; Φαινομενικά τελείως άσχετα μεταξύ τους, τα συστήματα αυτά εμφανίζουν μια συγκεκριμένη μορφή κατανομής όταν εξεταστούν κάτω από ένα συγκεκριμένο πρίσμα. Αν κανείς υπολογίσει πόσες φορές εμφανίζεται κάθε λέξη στον Moby Dick, το μνημειώδες μυθιστόρημα του Herman Melville, θα βρεθεί μπροστά στο εντυπωσιακό συμπέρασμα ότι περίπου οι μισές από αυτές (~44%) εμφανίζονται μόνο μία φορά (είναι δηλαδή αυτό που στη στατιστική γλωσσολογία ονομάζεται “*άπαξ λεγόμενον*”). Αν κατατάξει τις λέξεις από τη συχνότερα εμφανιζόμενη (που λογικά είναι το άρθρο “the”) στη σπανιότερη και αναπαραστήσει γραφικά τη σχέση κατάταξης-συχνότητας θα βρεθεί μπροστά σε μια κατανομή που μοιάζει πάρα πολύ με αυτήν της Εικόνας 9.10β<sup>2</sup>. Πρόκειται για μια κατανομή νόμου-δύναμης, σαν κι αυτήν που θα παρατηρήσει κανείς αν αναπαραστήσει γραφικά τον αριθμό των link μεταξύ σελίδων του παγκόσμιου ιστού (www), το εισόδημα των φορολογούμενων των ΗΠΑ αν τους κατατάξουμε από το φτωχότερο στον πλουσιότερο, τον αριθμό των πτήσεων που ενώνουν τα αεροδρόμια της Ευρώπης ή τον πληθυσμό των 1000 μεγαλύτερων πόλεων του πλανήτη (Soo, 2005)!

Οι νόμοι-δύναμης είναι κατανομές που περιγράφουν χαρακτηριστικά ανεξάρτητα-κλίμακας με την έννοια ότι οι μεταβλητές μεταβάλλονται με τρόπο που δεν επηρεάζεται από την κλίμακα. Αυτό σημαίνει ότι μπορούμε να βρούμε 100.000 Αμερικανούς πολίτες με εισόδημα διπλάσιο από το μέσο κατά κεφαλήν εισόδημα, μερικές χιλιάδες με εισόδημα 100 φορές μεγαλύτερο από τη μέση τιμή και ακόμα και μια δεκάδα που απολαμβάνει εισόδημα 320 φορές μεγαλύτερο από το μέσο όρο (Okuyama, Takayasu & Takayasu, 1999)! Με ανάλογο τρόπο, ένα τυχαίο δίκτυο θα έχει πολύ λίγους κόμβους με μεγαλύτερο βαθμό από το 150% της μέση τιμής, όμως σ' ένα βιολογικό δίκτυο, ανεξάρτητο κλίμακας, μπορούμε να βρούμε σημαντικό αριθμό κόμβων με τιμές μεγαλύτερες από το διπλάσιο της, το τριπλάσιο της κ.ο.κ. (ανάλογα με την έκταση της συμπεριφοράς του νόμου-δύναμης). Η πολύ ενδιαφέρουσα αυτή ιδιότητα οφείλεται στην μορφή της κατανομής που ορίζεται από τη σχέση:

$$p(k) \sim k^{-a} \quad 9.12$$

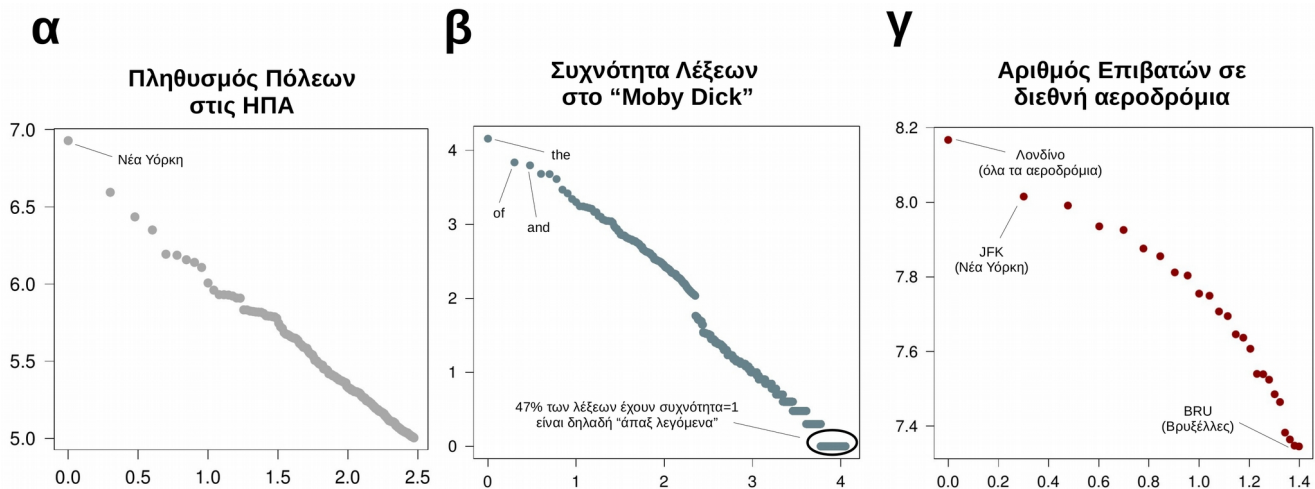
σύμφωνα με την οποία η συχνότητα εμφάνισης μιας τιμής  $p(k)$  εξαρτάται από τον αρνητικό εκθέτη  $-a$ , ο οποίος είναι κλασματικός και συνήθως κυμαίνεται μεταξύ  $2 < a < 3$ . Λογαρίθμηση της 9.12 δίνει την:

$$\log(p(k)) = -a \log(k) + \beta \quad 9.13$$

με  $\beta$  μια σταθερά. Η μορφή της εξίσωσης 9.13, είναι αυτή που περιγράφει τη συμπεριφορά των Εικόνων 9.10β και 9.11 με την παράμετρο  $a$  να ισούται με την κλίση της ευθείας. Βασικές ιδιότητες των κατανομών νόμου-δύναμης, που πηγάζουν από την ανεξαρτησία τους από την κλίμακα, είναι η αδυναμία προσδιορισμού της διασποράς τους και (για συγκεκριμένες τιμές  $a$ ) της μέσης τιμής (Clauset et al., 2009) αλλά κυρίως το γεγονός ότι μπορούν να προκύψουν από

<sup>2</sup> Η κατανομή αυτή ονομάζεται και κατανομή Zipf από τον George Kingsley Zipf που αν και δεν ήταν ο πρώτος που την ανέφερε, ήταν ο πρώτος που προσπάθησε να την ερμηνεύσει μέσω χαρακτηριστικών γλωσσολογικών ιδιοτήτων (MEJ Newman, 2005).

συγκεκριμένα δυναμικά συστήματα με κοινό χαρακτηριστικό την εξέλιξη στο χρόνο. Είναι αυτή η δυναμική ανάπτυξη των κατανομών νόμων-δύναμης μέσα από φαινομενικά απλές διεργασίες που τους δίνει έτσι ιδιότητες καθολικότητας (universality) και κάνει τα δίκτυα βιολογικών αλληλεπιδράσεων να συμπεριφέρονται στατιστικά ανάλογα με τους πληθυσμούς των πόλεων ή τα λογοτεχνικά κείμενα



**Εικόνα 9.11:** Νόμοι Δύναμης σε διαφορετικά συστήματα. Κατανομές κατάταξης από πολύ διαφορετικά συστήματα εμφανίζουν χαρακτηριστικά πρότυπα νόμου-δύναμης που υποδηλώνουν μια κοινή δυναμική ιδιότητα κατά την εξέλιξή τους. α) Πληθυσμός πόλεων στις ΗΠΑ (πηγή: Wikipedia) από τη μεγαλύτερη στη μικρότερη, στον οριζόντιο άξονα αναγράφεται ο λογάριθμος της κατάταξης και στον κάθετο ο λογάριθμος του πληθυσμού της β) Συχνότητα εμφάνισης λέξεων στο Moby Dick του Herman Melville (πηγή κειμένου: Project Gutenberg) στον οριζόντιο άξονα ο λογάριθμος της κατάταξης (από τη συχνότερη στη σπανιότερη) και στον κάθετο ο λογάριθμος των εμφανίσεων της κάθε λέξης στο κείμενο. γ) Αριθμός επιβατών που διακινούνται από τα 25 μεγαλύτερα διεθνή αεροδρόμια (πηγή: CAPA, Centre for Aviation). Η κλίση της κατανομής δεν είναι πλήρως γραμμική λόγω της επιλογής των 25 μεγαλύτερων αεροδρομίων, ακόμα κι έτσι όμως η τάση για γραμμικότητα είναι εμφανής.

Πέρα από την προέλευσή τους, τα δομικά τους χαρακτηριστικά κάνουν τέτοια δίκτυα εξαιρετικά αποτελεσματικά στη μετάδοση πληροφορίας, την ταχεία επικοινωνία μεταξύ των μελών τους αλλά κυρίως τους προσδίδουν τη βασική ιδιότητα της σταθερότητας (robustness) (Wagner, 2013). Τα δίκτυα ανεξάρτητα-κλίμακας είναι εξαιρετικά ανθεκτικά στις διαταραχές. Η ιδιότητα αυτή της σταθερότητας μπορεί να εξηγηθεί μ' ένα απλό νοητικό πείραμα. Φανταστείτε ένα βιολογικό δίκτυο μεταγραφικής ρύθμισης, στο οποίο με κάποιο δραστικό τρόπο (π.χ. μαζική μεταλαξίγνεση) ένα σημαντικό ποσοστό (π.χ. ~10%) των κόμβων του δικτύου καθίστανται μη λειτουργικοί. Αν το δίκτυο ήταν τυχαίο κάτι τέτοιο θα σήμαινε ότι το ποσοστό των ακμών που θα απαλείφονταν θα ήταν περίπου ~10% λόγω της ομοιομορφίας με την οποία οι ακμές κατανέμονται σ' ένα τυχαίο δίκτυο. Σ' ένα δίκτυο ανεξάρτητο-κλίμακας αυτή η απαλοιφή θα αφορούσε ένα σημαντικά μικρότερο αριθμό ακμών, καθώς λόγω ακριβώς της δομής του δικτύου είναι πολύ πιθανότερο το 10% των κόμβων που χάθηκε να είναι κόμβοι με μικρό βαθμό (Wagner, 2000). Η ύπαρξη ενός σημαντικού αριθμού υπερκόμβων εξασφαλίζει πως είναι πολύ δύσκολο να διαταραχθεί σημαντικά η λειτουργία του δικτύου ακόμα κι αν ένας σημαντικός αριθμός κόμβων απαλειφθεί μονομιάς. Η σταθερότητα αυτή των

δικτύων που είναι ανεξάρτητα-κλίμακας είναι η βασική αιτία πίσω από την καθολικότητά τους. Πώς μπορούν όμως να προκύψουν κατανομές νόμου-δύναμης σε βιολογικά συστήματα; Αφήνουμε την τελευταία αυτή πιο ενδιαφέρουσα ερώτηση για την τελευταία ενότητα αυτού του κεφαλαίου.

## Αυθόρμητη ανάπτυξη ιδιοτήτων βιολογικών δικτύων

Στην Εικόνα 9.12 αναπαρίσταται γραφικά το δίκτυο μιας συνωμοτικής, εγκληματικής οργάνωσης της οποίας οι πυρήνες έχουν μέγεθος πέντε ατόμων. Τα μέλη κάθε πυρήνα αναφέρουν στον αρχηγό του κεντρικού πυρήνα ο οποίος ηγείται των πέντε πυρήνων που δραστηριοποιούνται γεωγραφικά στην ίδια περιοχή. Το δίκτυο που προκύπτει από τις σχέσεις μεταξύ των μελών της οργάνωσης είναι έτσι απόλυτα ιεραρχικά οργανωμένο, ως αποτέλεσμα του συνωμοτικού σχεδιασμού. Τα χαρακτηριστικά αυτού του δικτύου και των δικτύων που σχεδιάζονται με αντίστοιχη ιεραρχική και σπονδυλωτή (modular) αρχιτεκτονική, είναι από στατιστική άποψη πολύ παρόμοια με αυτά των ανεξάρτητων-κλίμακας δικτύων. Ακολουθούν έτσι νόμους δύναμης όχι μόνο στην κατανομή των βαθμών των κόμβων αλλά και στην κατανομή των συντελεστών συσσωμάτωσης  $C(k)$ . Τέτοια δίκτυα που εμφανίζονται στην οργάνωση των μελών της μαφίας ή στα οργανογράμματα των πολυεθνικών εταιρειών (μία μόνο από τις πολλές τους ομοιότητες) προκύπτουν ως αποτέλεσμα ενός συγκεκριμένου σχεδιασμού και δημιουργούνται εξ αρχής με αυτόν τον τρόπο για να εκμεταλλευτούν την ιδιότητα της σταθερότητας που αναφέραμε στην προηγούμενη ενότητα.

Τα βιολογικά δίκτυα, από την άλλη πλευρά, είναι αποτέλεσμα δυναμικών διαδικασιών που εξελίσσονται κατά τη διάρκεια εκατομμυρίων χρόνων. Το ερώτημα λοιπόν που προκύπτει είναι με ποιον τρόπο μπορεί να έχουν εξελιχθεί τα βιολογικά δίκτυα, ώστε να προσομοιάζουν σε μεγάλο βαθμό δίκτυα τα οποία είναι αποτέλεσμα ενός “ευφυούς σχεδιασμού”<sup>3</sup>; Για ν' απαντήσουμε σ' αυτό το ερώτημα δε χρειάζεται να καταφύγουμε σε μεταφυσικές θεωρίες αλλά σε μια έξυπνη μοντελοποίηση της βιολογικής διαδικασίας της εξέλιξης. Πράγματι, αν συγκρίνουμε ένα βιολογικό δίκτυο με το δίκτυο της Εικόνας 9.12α θα δούμε ότι ένα βασικό κοινό τους χαρακτηριστικό είναι (σε κάποιο βαθμό) η επανάληψη δομικών υπομονάδων. Θυμηθείτε ότι τα βιολογικά δίκτυα αποτελούνται από συγκεκριμένα μοτίβα που ονομάζονται έτσι επειδή εμφανίζονται με επαναλαμβανόμενη δομή, με τρόπο ανάλογο μ' αυτόν των πεντάδων στο δίκτυο των συνωμοτών. Μια τέτοια οργάνωση δεν χρειάζεται να είναι κατ' ανάγκη σχεδιασμένη εξ αρχής, αλλά μπορεί να προκύψει σ' ένα βιολογικό δίκτυο μέσω μιας διαδικασίας σταδιακής προσθήκης συνδέσεων σε ένα προϋπάρχον δίκτυο. Μπορούμε να καταλάβουμε καλύτερα πώς μπορεί να συμβεί κάτι τέτοιο με το παρακάτω παράδειγμα:

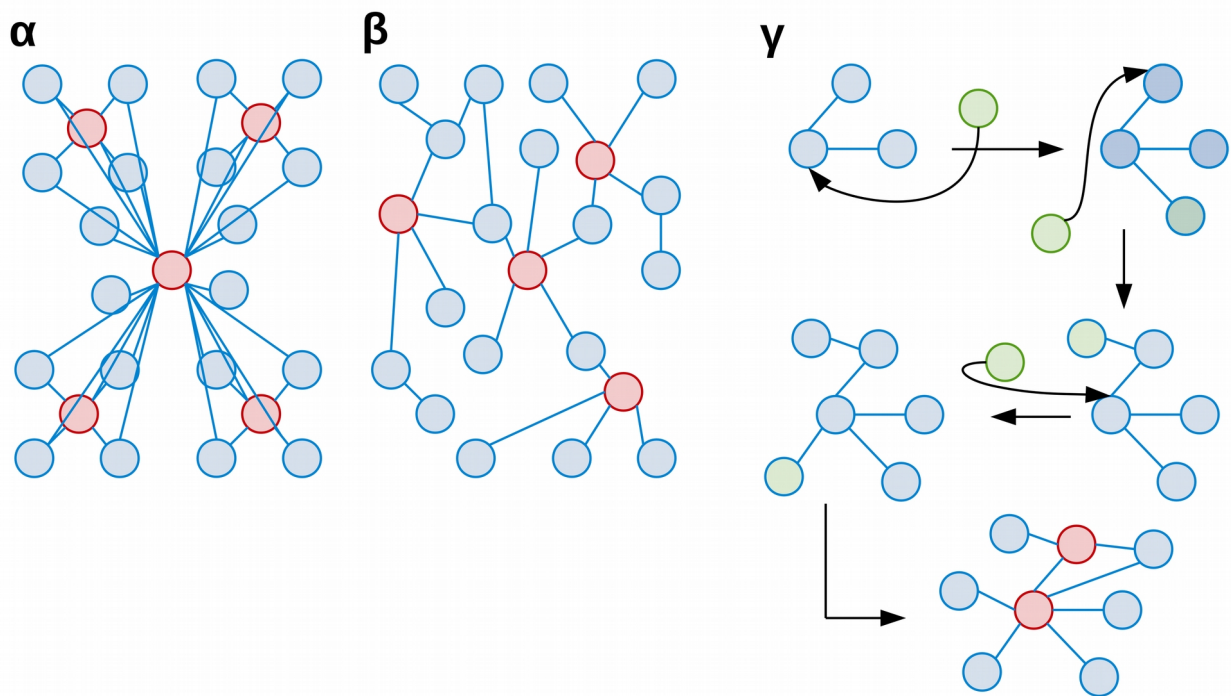
Φανταστείτε ένα πρωτο-δίκτυο μεταγραφικής ρύθμισης στο οποίο ένας μικρός αριθμός μεταγραφικών παραγόντων ρυθμίζει μερικές δεκάδες γονιδίων. Καθώς το σύστημα μεγαλώνει και γίνεται περισσότερο πολύπλοκο, το σύνολο των γονιδίων που υπόκεινται σε ρύθμιση μεγαλώνει. Λόγω της αφθονίας των συγκεκριμένων μεταγραφικών παραγόντων, της διάχυσής τους στον πυρήνα, ώστε να βρίσκονται κοντά στα γονίδια-στόχους τους κλπ, είναι αρκετά πιθανό ένα γονίδιο

<sup>3</sup> Η επιλογή των όρων δεν είναι τυχαία, καθώς οι ιδιότητες των βιολογικών δικτύων έχουν χρησιμοποιηθεί από τους θιασώτες της ψευδο-επιστημονικής θεωρίας του “ευφυούς σχεδιασμού”.

που μέχρι πρότινος δε ρυθμιζόταν να αποτελέσει στόχο ενός από αυτούς<sup>4</sup> και μάλιστα με πιθανότητα ανάλογη του αριθμού των γονιδίων στόχων που ο παράγοντας ήδη έχει. Η πιθανότητα αυτή, της σύνδεσης ενός νέου κόμβου με έναν κόμβο  $i$  που αποτελεί ήδη μέρος του δικτύου μπορεί να εκφραστεί ως εξής:

$$P(i) = \frac{k_i}{\sum k_i} \quad 9.14$$

Όπου  $k_i$  είναι ο βαθμός του κόμβου  $i$ . Ένα τέτοιο μοντέλο προβλέπει ότι οι νέες συνδέσεις στο δίκτυο δε θα προκύπτουν τυχαία αλλά με μια προτίμηση προς τους προϋπάρχοντες κόμβους μεγάλου βαθμού, με μια διαδικασία που σταδιακά οδηγεί σε δίκτυα με υπερ-κόμβους και χαρακτηριστικά ανεξάρτητα κλίμακας.



**Εικόνα 9.12:** α) Ιεραρχικό Δίκτυο. β) Δίκτυο ανεξάρτητο κλίμακας γ) Το μοντέλο προτιμησησικής σύνδεσης.

Το μοντέλο που περιγράψαμε παραπάνω και που σχηματικά φαίνεται στην Εικόνα 9.12β, ονομάζεται μοντέλο της “προτιμησησικής σύνδεσης” (preferential attachment) (Barabási & Albert, 1999) και μπορεί να οδηγήσει στην ανάπτυξη δικτύων ανεξάρτητων-κλίμακας χωρίς την *a priori* ύπαρξη μιας ιεραρχικής αρχιτεκτονική. Ανάλογο της προτιμησησικής σύνδεσης σε άλλα βιολογικά δίκτυα είναι ο διπλασιασμός υπομονάδων που ευθύνονται για αλληλεπιδράσεις πρωτεϊνών σε δίκτυα πρωτεϊνών (Shahzad, Mittenthal & Caetano-Anollés, 2015), ή η χρήση ανάλογων ενεργών κέντρων που συνδέονται με πολύ άφθονους μεταβολίτες σε μεταβολικά δίκτυα (Barve & Wagner,

<sup>4</sup> Η διαδικασία αυτή προφανώς περιλαμβάνει τυχαίες μεταλλαγές καταλοίπων στον υποκινητή του εν λόγω γονιδίου. Η ιδέα είναι ότι τέτοιες μεταλλαγές θα υιοθετούνται με μεγαλύτερη πιθανότητα αν μετατρέπουν το γονίδιο σε στόχο ενός μεταγραφικού παράγοντα.

2013). Σε μη βιολογικά δίκτυα μπορούμε να σκεφτούμε πόσο πιο εύκολο είναι να γίνει κάποιος πολυεκατομμυριούχος αν είναι ήδη εκατομμυριούχος<sup>5</sup>, ή πόσο μια πόλη που είναι ήδη αρκετά μεγάλη μπορεί να προσελκύσει ακόμα μεγαλύτερο αριθμό νέων κατοίκων. Είναι προφανές ότι η δυναμική εξέλιξη βιολογικών δικτύων είναι μια πολύ πιο πολύπλοκη διαδικασία για να εξηγηθεί με μια απλή σχέση προτίμησης, ωστόσο το μοντέλο της προτιμησιακής σύνδεσης είναι επαρκές για να εξηγήσει το μεγαλύτερο μέρος των στατιστικών τους ιδιοτήτων.

## Συμπεράσματα

Τα τελευταία χρόνια, η μελέτη των δικτύων αποτελεί ένα ιδιαίτερα ενεργό επιστημονικό πεδίο, που τροφοδοτείται από την αύξηση του όγκου των διαθέσιμων δεδομένων σε πολλούς ερευνητικούς τομείς, ένα ιδιαίτερο ενδιαφέρον για τα μέσα κοινωνικής δικτύωσης (και τα κοινωνικά δίκτυα που αυτά αντανakλούν) και την προσδοκία της δυνατότητας ανάλυσης συστημάτων που μέχρι πρότινος φάνταζαν απροσπέλαστα από πλευράς πολυπλοκότητας. Μεταξύ των συστημάτων αυτών, τα βιολογικά έχουν προεξάρχουσα θέση.

Σε θεωρητικό επίπεδο η ανάλυση βιολογικών δικτύων μπόρεσε για πρώτη φορά να συγκεράσει δύο χαρακτηριστικές ιδιότητες των βιολογικών συστημάτων, της εξέλιξης και της σταθερότητας, που μέχρι πρότινος φάνταζαν ασύνδετες, αν όχι ασύμβατες μεταξύ τους (Wagner, 2013). Σ' ένα πιο πρακτικό επίπεδο, η μελέτη των δικτύων οργάνωσης σε γονιδιακά, πρωτεϊνικά και μεταβολικά συστήματα αφήνει υποσχέσεις για την κατανόηση πολύπλοκων διαδικασιών που ενδεχομένως να αποτελούν το πρώτο βήμα για την κατανόηση διαδικασιών που σχετίζονται με παθολογικές καταστάσεις (Barabási, Gulbahce & Loscalzo, 2011). Η θεωρία δικτύων έχει ήδη συμβάλει σημαντικά σ' αυτόν τον τομέα με μελέτες που σχετίζονται με την εκδήλωση και εξάπλωση επιδημιών (Ball & Neal, 2008).

Ανάμεσα στα πιο λεπτά σημεία αυτού του κεφαλαίου, βρίσκεται η σύγκριση των βιολογικών δικτύων με άλλα δίκτυα. Είναι σημαντικό να γίνει σαφές ότι ούτε όλα τα βιολογικά δίκτυα είναι ανεξάρτητα-κλίμακας, ούτε το αντίθετο. Πολλά βιολογικά δίκτυα έχουν χαρακτηριστικά τυχαίων δικτύων (μπορείτε να σκεφτείτε μερικά;), ενώ πολλά δίκτυα ανεξάρτητα-κλίμακας έχουν χαρακτηριστικά που δε συναντώνται σε βιολογικά δίκτυα. Μια χαρακτηριστική διαφορά π.χ. των κοινωνικών δικτύων με τα δίκτυα πρωτεϊνικών αλληλεπιδράσεων (με τα οποία συχνά συγκρίνονται σε μια κρίση ψευδο-δαρβινισμού) είναι ότι είναι τα μεν πρώτα είναι κατατάξιμα (assortative) ενώ τα δεύτερα δεν είναι (βλ. Ερωτήσεις/Ασκήσεις). Σε κάθε περίπτωση, η μελέτη των βιολογικών δικτύων θα πρέπει να γίνεται, όπως και κάθε μελέτη, στο πλαίσιο της όσο καλύτερης γνώσης του συστήματος.

<sup>5</sup> Η αρχή αυτή είναι γνωστή ως η “Αρχή του Ματθαίου” (“The Mathew Effect”) από το εδάφιο στο κείμενο του ευαγγελιστή Ματθαίου στο οποίο παρατίθεται η ρήση του Ιησού “τῷ γὰρ ἔχοντι παντὶ δοθήσεται καὶ περισσευθήσεται τοῦ δὲ μὴ ἔχοντος καὶ ὃ ἔχει ἀρθήσεται ἀπ’ αὐτοῦ.”



## Ερωτήσεις/Ασκήσεις

1. Γιατί μοτίβα θετικής ανατροφοδότησης παρατηρούνται σπάνια (έως ποτέ) σε μεταβολικά δίκτυα;
2. Το βασικό ρυθμιστικό ένζυμο της γλυκόλυσης, η φωσφοφρουκτοκινάση (PFK) ενεργοποιείται από την 2,6-διφωσφορική φρουκτόζη (F-2,6-BP) που παράγεται σε ένα βήμα από το ίδιο το προϊόν της αντίδρασης της φωσφοφρουκτοκινάσης κατά το σχήμα. Σε ποιο μοτίβο δικτύου αντιστοιχεί ο μηχανισμός; Τι θα συμβεί αν η συγκέντρωση του υποστρώματος της PFK (F-6P) αυξηθεί;



3. Σ' ένα δίκτυο μεταγραφικής ρύθμισης, αν ένας παράγοντας έχει μεγαλύτερο βαθμό από κάποιον με μεγάλη τιμή κεντρικότητας, είναι λογικό να υποθέσουμε πως πιο σημαντικός για το δίκτυο θα είναι ο δεύτερος, ιδιαίτερα αν το μέγεθος του δικτύου είναι αρκετά μεγαλύτερο από το μέγιστο βαθμό κόμβου. Εξηγήστε γιατί.
4. Ποιος είναι ο ελάχιστος και ποιος ο μέγιστος αριθμός ακμών ενός μη-κατευθυνόμενου δικτύου με 10 κόμβους;
5. Δεδομένου ενός πίνακα γειτνίασης που περιγράφει ένα δίκτυο, πώς θα εντοπίσετε τυχόν αποσυνδεδεμένα τμήματα δικτύου, τμήματα δηλαδή που ορίζουν αυτόνομα υπο-δίκτυα;
6. Περιγράψτε τη σειρά των αναλύσεων που θα κάνατε για να διακρίνετε ένα τυχαίο, από ένα δίκτυο ανεξάρτητο-κλίμακας κι ένα σπονδυλωτό δίκτυο.
7. Σε αντίθεση με πολλά δίκτυα ανεξάρτητο-κλίμακας τα δίκτυα πρωτεϊνικών αλληλεπιδράσεων δεν είναι κατατάξιμα. Μπορείτε να σκεφτείτε γιατί; Ποιο πλεονέκτημα προσφέρει η μη-καταταξιμότητα στα βιολογικά δίκτυα;

## Διαβάστε Περισσότερα

### Για τα βιολογικά δίκτυα γενικά:

Μια πολύ καλή εισαγωγή δίνεται στο Κεφάλαιο 9 του *Systems Biology* (Klipp et al., 2013), όπου περιγράφονται τόσο οι θεωρητικές αρχές όσο και τεχνικές για τη δημιουργία και μελέτη βιολογικών δικτύων.

### Για τα δίκτυα γονιδιακής ρύθμισης:

Για τον αναγνώστη που θέλει να εμβαθύνει, ένα εξαιρετικό βιβλίο είναι το *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution* (Davidson, 2010). Κάθε κεφάλαιο αναφέρεται και σε μια κατηγορία βιολογικής λειτουργίας (μεταβολισμός, ανάπτυξη κλπ) παρουσιάζοντας τη σημασία των δικτύων μεταγραφικής ρύθμισης που τη διέπουν.

### Για τα μοτίβα δικτύων:

Το βιβλίο του Uri Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Alon, 2006) είναι το σύγγραμμα αναφοράς γι' αυτό το εξαιρετικά ενδιαφέρον αλλά και δύσκολο τεχνικά βιολογικό πρόβλημα.

### Για τη Θεωρία Δικτύων γενικότερα:

Το *Graph Theory* (Diestel, 2006) είναι η καλύτερη εισαγωγή για τη μαθηματική θεώρηση των γράφων. Σε ό,τι όμως αφορά την "Επιστήμη των Δικτύων", το *Network Science*, του Albert-Laszlo Barabási, του ανθρώπου που εισήγαγε τη μελέτη των δικτύων ως αυτόνομο επιστημονικό πεδίο, είναι ελεύθερα διαθέσιμο εδώ (<http://barabasi.com/networksciencebook/>) και περιέχει όλα όσα θα θέλατε να μάθετε για τα δίκτυα (κι ακόμα περισσότερα). Μια ελαφρώς πιο τεχνικά απαιτητική εκδοχή του είναι το παλιότερο, *The Structure and Dynamics of Networks* (Newman, Barabási & Watts, 2006).

### Για τις ανεξάρτητες κλίμακας ιδιότητες και τους νόμους δύναμης:

Μια αναλυτική εισαγωγή δίνεται στο Κεφάλαιο 4 του *Critical Phenomena in Natural Sciences* (Sornette, 2000). Περισσότερα για το ιδιαίτερα ενδιαφέρον θέμα της αυτο-οργάνωσης και αυτο-ομοιότητας, όπως ήδη συζητήσαμε στο Κεφάλαιο 5, μπορείτε να βρείτε στο *The Fractal Geometry of Nature* (Mandelbrot, 1983).

### Για το μοντέλο της προτιμησιακής σύνδεσης:

Η πιο σημαντική εργασία στη θεωρία δικτύων από τον καιρό του Erdős είναι το άρθρο του Albert-Laszlo Barabási και της Réka Albert (Barabási & Albert, 1999), στο οποίο εισάγονται για πρώτη φορά τα δυναμικά στοιχεία της εξέλιξης των δικτύων χωρίς κλίμακα και το μοντέλο της προτιμησιακής σύνδεσης.

## Βιβλιογραφία

- Almaas, E., Kovács, B., Vicsek, T., Oltvai, Z. N. & Barabási, A.-L. (2004). Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature*, 427(6977), 839–43. <http://doi.org/10.1038/nature02289>
- Alon, U. (2006). *An Introduction to Systems Biology: Design Principles of Biological Circuits*. CRC Press.
- Alon, U. (2007). Network motifs: theory and experimental approaches. *Nature Reviews. Genetics*, 8(6), 450–61. <http://doi.org/10.1038/nrg2102>
- Ball, F. & Neal, P. (2008). Network epidemic models with two levels of mixing. *Mathematical Biosciences*, 212(1), 69–87. <http://doi.org/10.1016/j.mbs.2008.01.001>
- Barabási, A.-L. & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(October), 509–512.
- Barabási, A.-L., Gulbahce, N. & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nature Reviews. Genetics*, 12(1), 56–68. <http://doi.org/10.1038/nrg2918>
- Barve, A. & Wagner, A. (2013). A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature*, 500(7461), 203–6. <http://doi.org/10.1038/nature12301>
- Bhalla and Ravi Iyengar, U. S. (1999). Emergent Properties of Networks of Biological Signaling Pathways. *Science*, 283(5400), 381–387. <http://doi.org/10.1126/science.283.5400.381>
- Clauset, A., Shalizi, C. R. & Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. *SIAM Review*, 51(4), 661–703. <http://doi.org/10.1137/070710111>
- Davidson, E. H. (2010). *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution*. Academic Press.
- Diestel, R. (2006). *Graph Theory*. Springer Science & Business Media.
- Eissmann, P., Evans, J. H., Mehrabi, M., Rose, E. L., Nedvetzki, S. & Davis, D. M. (2010). Multiple mechanisms downstream of TLR-4 stimulation allow expression of NKG2D ligands to facilitate macrophage/NK cell crosstalk. *Journal of Immunology (Baltimore, Md. : 1950)*, 184(12), 6901–9. <http://doi.org/10.4049/jimmunol.0903985>
- Faucon, P. C., Pardee, K., Kumar, R. M., Li, H., Loh, Y.-H. & Wang, X. (2014). Gene networks of fully connected triads with complete auto-activation enable multistability and stepwise stochastic transitions. *PLoS One*, 9(7), e102873. <http://doi.org/10.1371/journal.pone.0102873>
- Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science (New York, N.Y.)*, 316(5830), 1497–502. <http://doi.org/10.1126/science.1141319>
- Klipp, E., Liebermeister, W., Wierling, C., Kowald, A., Lehrach, H. & Herwig, R. (2013). *Systems Biology*.

John Wiley & Sons.

- Lee, T. I., Rinaldi, N. J., Odom, D. T., Bar-joseph, Z., Gerber, G. K., Hannett, N. M., ... Young, R. a. (2002). Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*, 298(October), 799–804.
- Mandelbrot, B. B. (1983). *The Fractal Geometry of Nature*. Henry Holt and Company.
- Newman, M. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), 323–351. <http://doi.org/10.1080/00107510500052444>
- Newman, M., Barabási, A.-L. & Watts, D. J. (2006). *The Structure and Dynamics of Networks*: Princeton University Press.
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H. & Kanehisa, M. (1999). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 27(1), 29–34. <http://doi.org/10.1093/nar/27.1.29>
- Okuyama, K., Takayasu, M. & Takayasu, H. (1999). Zipf's law in income distribution of companies. *Physica A: Statistical Mechanics and Its Applications*, 269(1), 125–131. [http://doi.org/10.1016/S0378-4371\(99\)00086-2](http://doi.org/10.1016/S0378-4371(99)00086-2)
- Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D. & Friedman, N. (2003). Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature Genetics*, 34(2), 166–76. <http://doi.org/10.1038/ng1165>
- Shahzad, K., Mitterenthal, J. E. & Caetano-Anollés, G. (2015). The organization of domains in proteins obeys Menzerath-Altmann's law of language. *BMC Systems Biology*, 9, 44. <http://doi.org/10.1186/s12918-015-0192-9>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., ... Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498–504. <http://doi.org/10.1101/gr.1239303>
- Soo, K. T. (2005). Zipf's Law for cities: a cross-country investigation. *Regional Science and Urban Economics*, 35(3), 239–263. <http://doi.org/10.1016/j.regsciurbeco.2004.04.004>
- Sornette, D. (2000). *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools*. Springer Science & Business Media.
- Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. & Gilles, E. D. (2002). Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912), 190–193. <http://doi.org/10.1038/nature01166>
- Stergachis, A. B., Neph, S., Sandstrom, R., Haugen, E., Reynolds, A. P., Zhang, M., ... Stamatoyannopoulos, J. a. (2014). Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature*, 515(7527), 365–370. <http://doi.org/10.1038/nature13972>
- Wagner, A. (2000). Robustness against mutations in genetic networks of yeast. *Nature Genetics*, 24(april), 355–361.

- Wagner, A. (2013). *Robustness and Evolvability in Living Systems*. Princeton University Press.
- Walhout, A. J. M. (2011). What does biologically meaningful mean? A perspective on gene regulatory network validation. *Genome Biology*, 12(4), 109. <http://doi.org/10.1186/gb-2011-12-4-109>
- Wilson, N. S., Dixit, V. & Ashkenazi, A. (2009). Death receptor signal transducers: nodes of coordination in immune signaling networks. *Nature Immunology*, 10(4), 348–55. <http://doi.org/10.1038/ni.1714>
- Zhou, R. & Hansen, E. A. (2006). Breadth-first heuristic search. *Artificial Intelligence*, 170(4-5), 385–408. <http://doi.org/10.1016/j.artint.2005.12.002>