

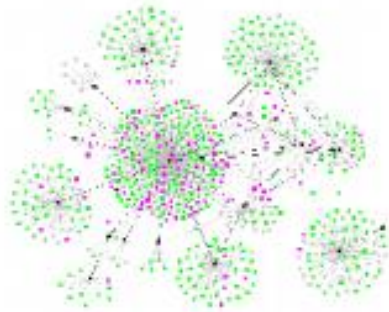


Ανάλυση Πολύπλοκων Βιολογικών Δικτύων

ΠΑΝΑΓΙΩΤΑ ΚΟΝΤΟΥ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

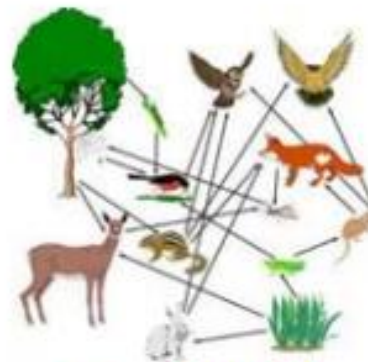
Network Data



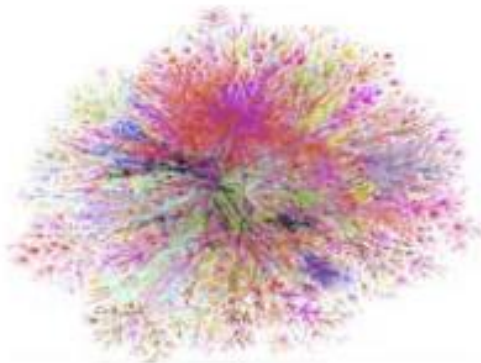
Disease
Spread



Electronic
Circuit



Food Web



Internet



Social Network

Βιολογικά δίκτυα

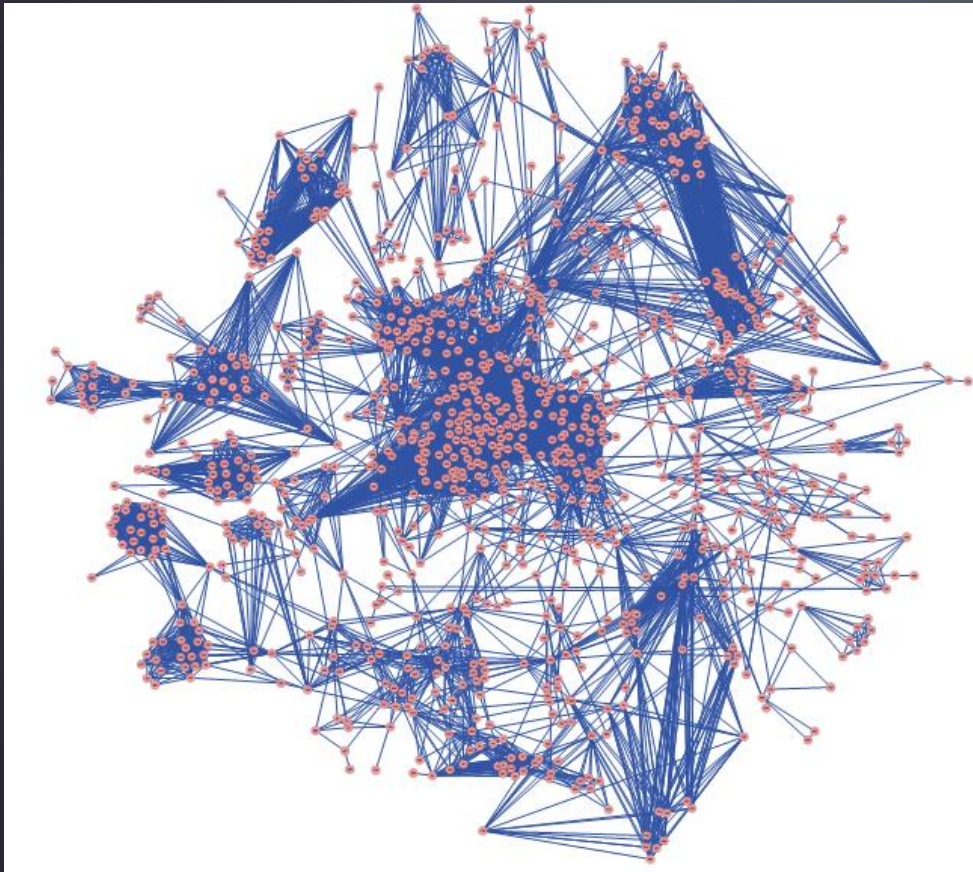
- ▶ Η μελέτη της πολυπλοκότητας των βιολογικών συστημάτων περνάει υποχρεωτικά από την επισκόπηση ενός τεράστιου αριθμού διεργασιών που αλληλοδιαπλέκονται δημιουργώντας πυκνά, σύνθετα δίκτυα.
- ▶ Βιολογικά δίκτυα συναντούμε σε όλα τα επίπεδα μελέτης των επιστημών της ζωής από το πιο μικροσκοπικό (μοριακό) ως το πιο μακροσκοπικό (οικοσυστήματα).
- ▶ 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.α.

Είδη βιολογικών δικτύων

1. Δίκτυα Πρωτεϊνικών αλληλεπιδράσεων

(Protein-protein interactions PPIs)

Οι πρωτεΐνες είναι οι κόμβοι του δικτύου και οι αλληλεπιδράσεις τους οι ακμές



Saccharomyces cerevisiae (yeast)
PPI network 1,004 nodes and
8,323 edges (17)

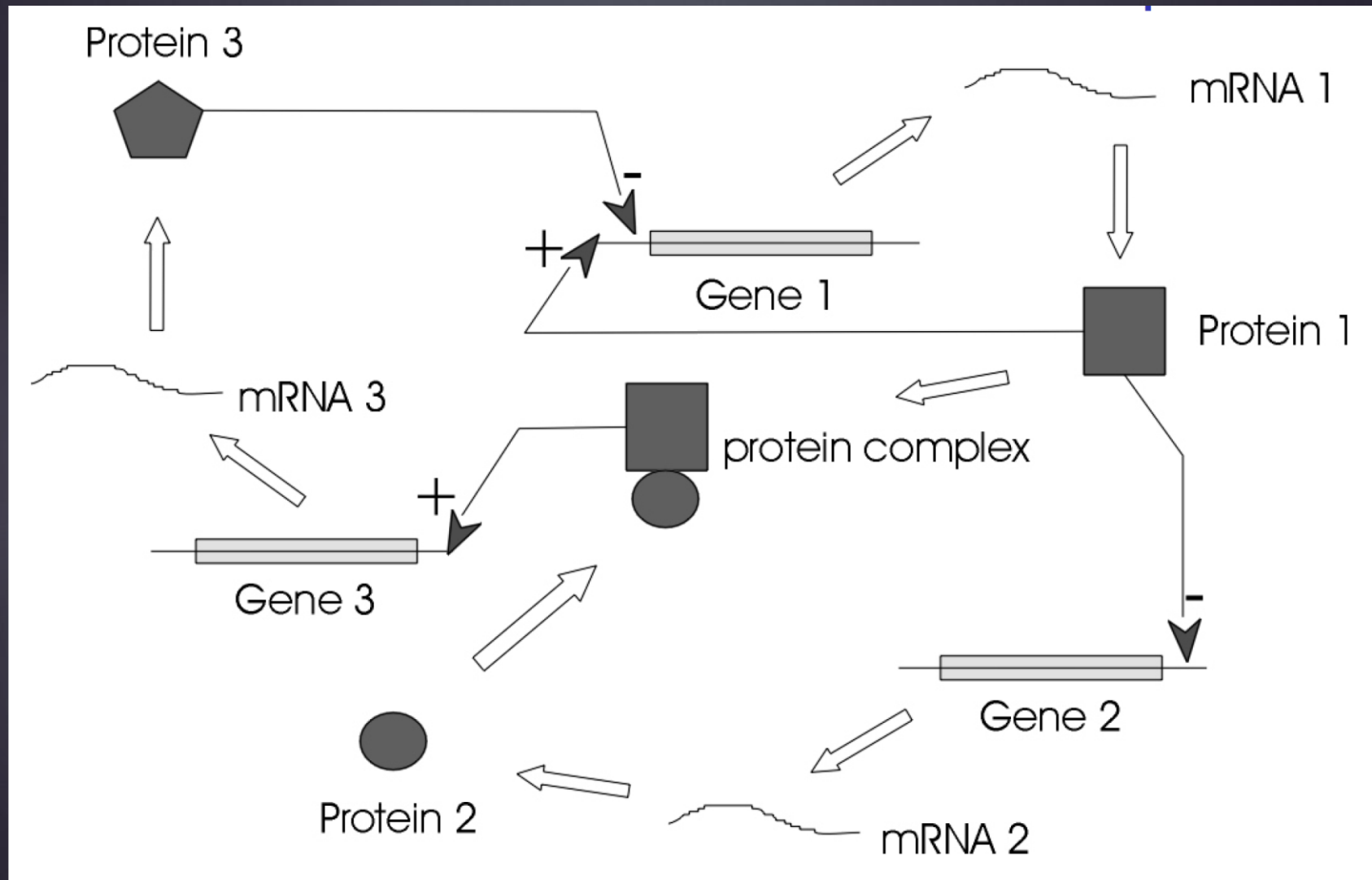
1. Δίκτυα Πρωτεϊνικών αλληλεπιδράσεων

(Protein-protein interactions PPIs)

- ▶ 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).
- ▶ 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.

2. Μεταγραφικά ρυθμιστικά δίκτυα (Transcriptional-Regulatory networks (GRNs))

Μοντελοποιείται ο τρόπος που οι πρωτεΐνες και άλλα βιομόρια εμπλέκονται στην διαδικασία της έκφρασης των γονιδίων

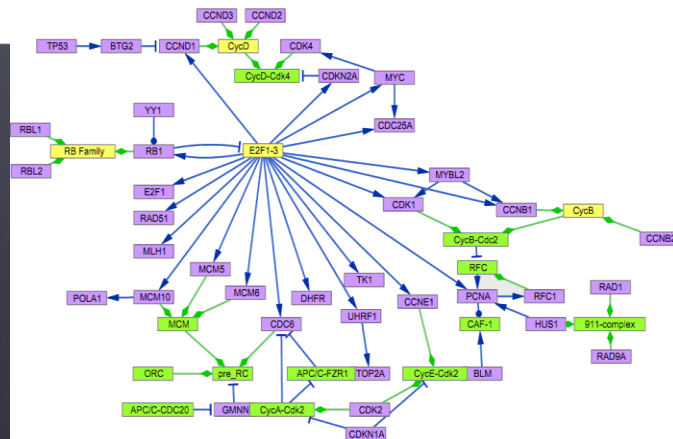
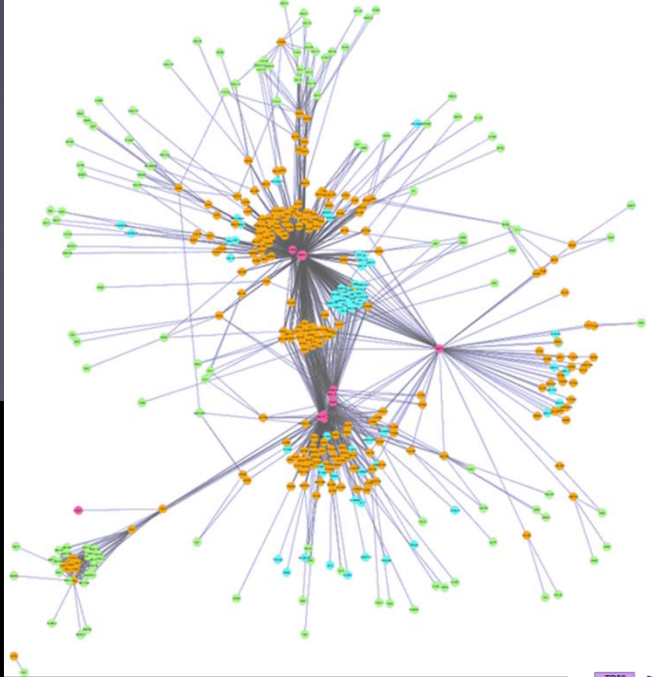
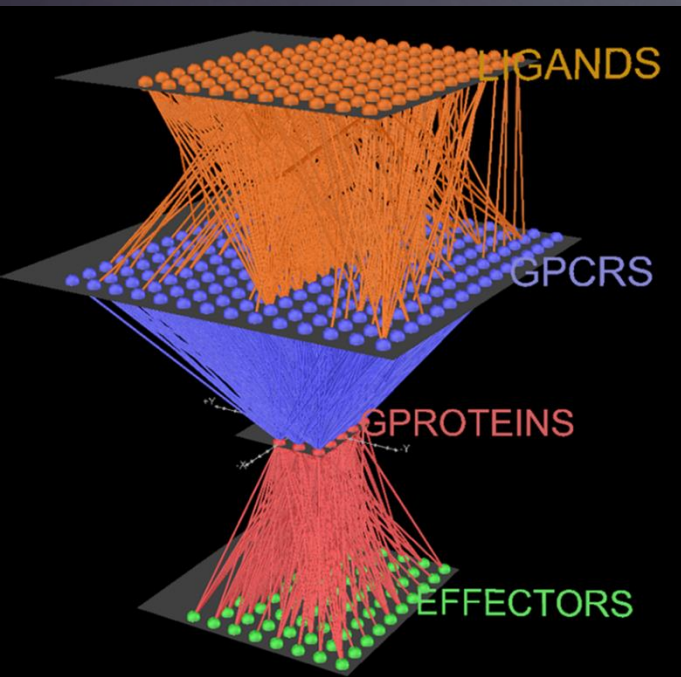


2.Μεταγραφικά ρυθμιστικά δίκτυα (Transcriptional- Regulatory networks (GRNs))

- ▶ 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
- ▶ Είναι ίσως το πιο σημαντικό είδος βιολογικού δικτύου.
- ▶ 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.
- ▶ 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.

3. Δίκτυα Μεταγωγής Σήματος (Signal trasduction networks)

Αναπαριστάται ο τρόπος μετάδοσης του σήματος από τον έξωκυττάριο στον ενδοκυττάριο χώρο, είτε στο εσωτερικό του κυττάρου.



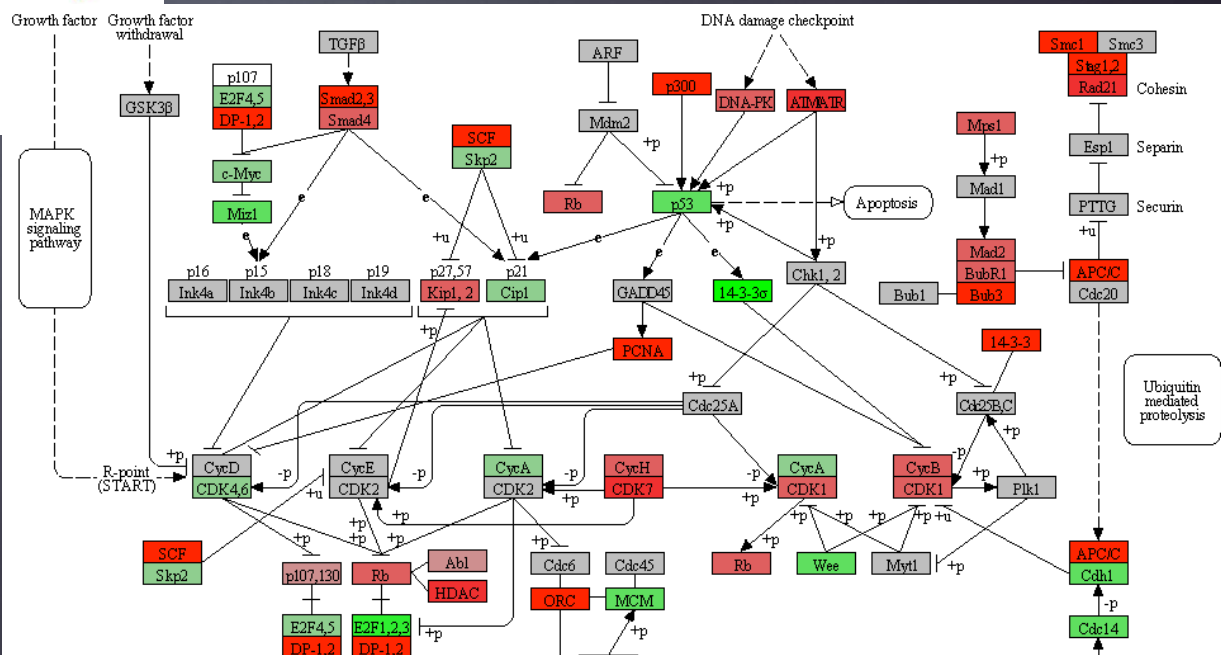
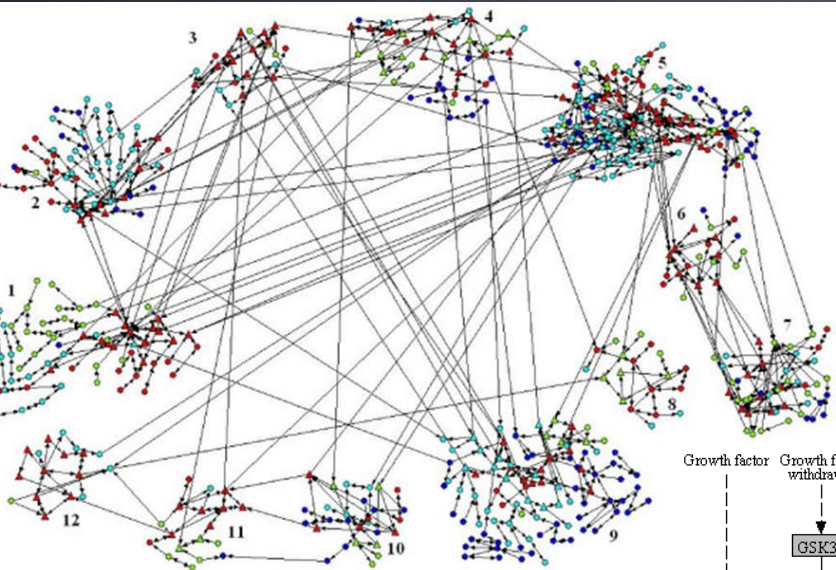
3. Δίκτυα Μεταγωγής Σήματος (Signal trasduction networks)

- ▶ Μαζί με τα μεταβολικά δίκτυα, τα σηματοδοτικά δίκτυα ανήκουν σε μια ευρύτερη κατηγορία που θα μπορούσαμε να ονομάσουμε “δίκτυα αντιδράσεων”.
- ▶ Τα μελετούμε ξεχωριστά γιατί η λειτουργία που επιτελούν δε σχετίζεται με την παραγωγή ή την κατανάλωση ενέργειας αλλά με τη μετάδοση πληροφορίας μέσω μοριακών σημάτων.
- ▶ 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.
- ▶ 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.

4. Μεταβολικά – Βιοχημικά δίκτυα

(Metabolic and Biochemical networks)

Μεταβολικό μονοπάτι θεωρείται μια σειρά από χημικές αντιδράσεις μέσα στο κύτταρο σε διαφορετικές χρονικές καταστάσεις

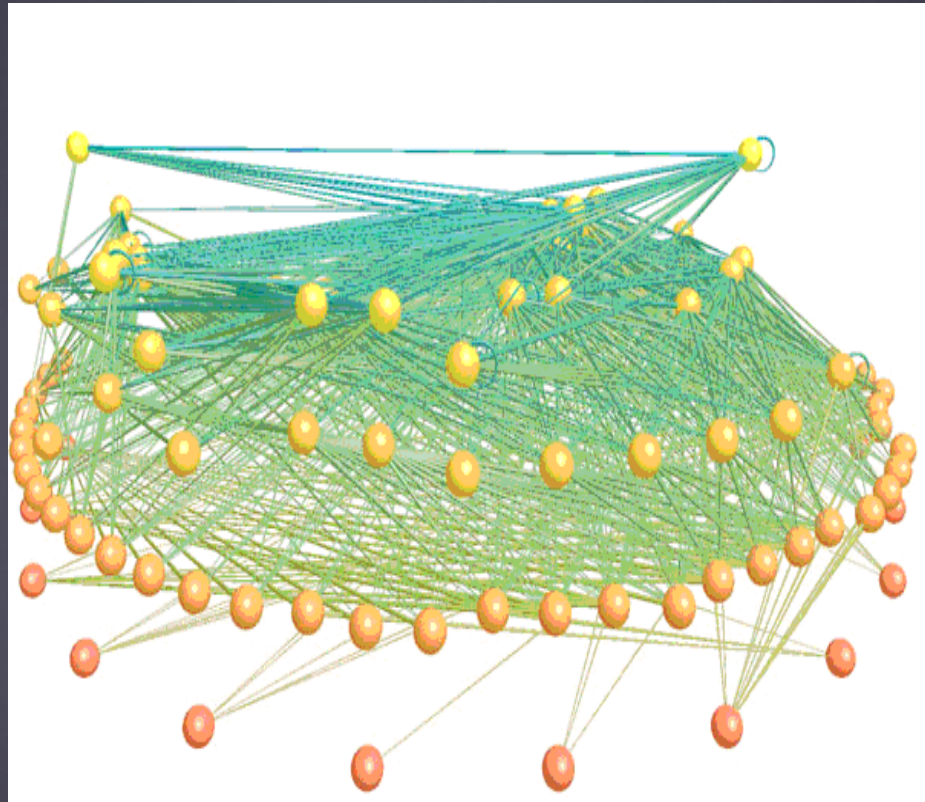


4. Μεταβολικά – Βιοχημικά δίκτυα (Metabolic and Biochemical networks)

- ▶ 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.
- ▶ Δύο στοιχεία ενώνονται αν συμμετέχουν στην ίδια αντίδραση.
- ▶ 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.

5. Οικολογικά δίκτυα-δίκτυα διατροφικών αλυσίδων (Food Webs)

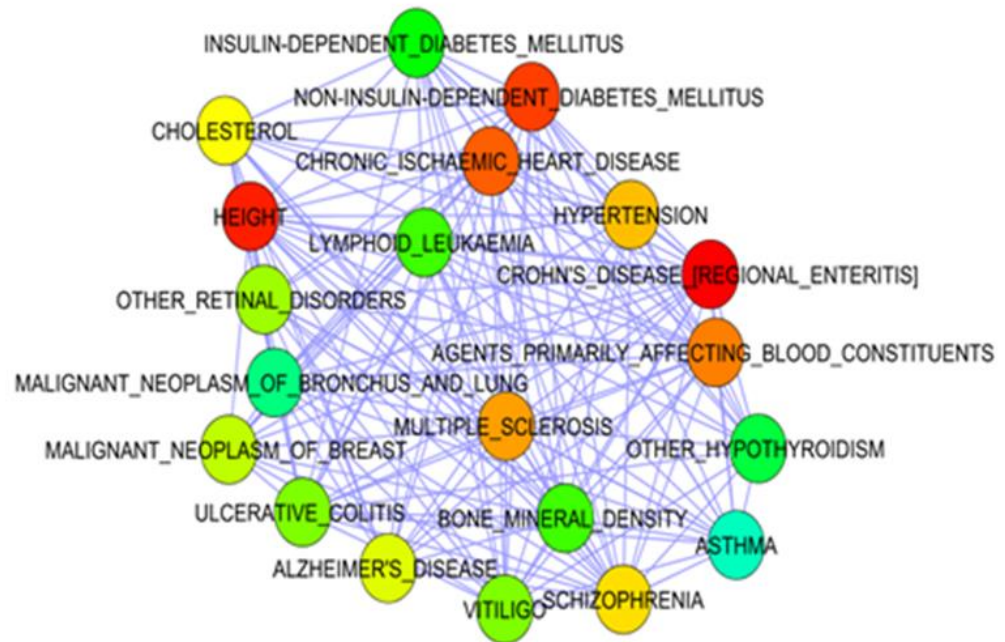
Αναπαριστώνται οι βιοτικές αλληλεπιδράσεις σε ένα οικοσύστημα. Τα είδη των οργανισμών που βρίσκονται σε ένα οικοσύστημα συνδέονται με αλληλεπιδράσεις κατά ζεύγη και μπορεί να είναι είτε τροφικές είτε συμβιωτικές



Food Web, el Verde (www.foodwebs.org, Yoon et al. 2004)

6. Δίκτυα ασθενειών (Diseases networks)

Δίνουν πληροφορίες για την προέλευση ασθενειών



Άλλα βιολογικά δίκτυα

7. Δίκτυα θεραπείας ασθενειών (treatment networks)

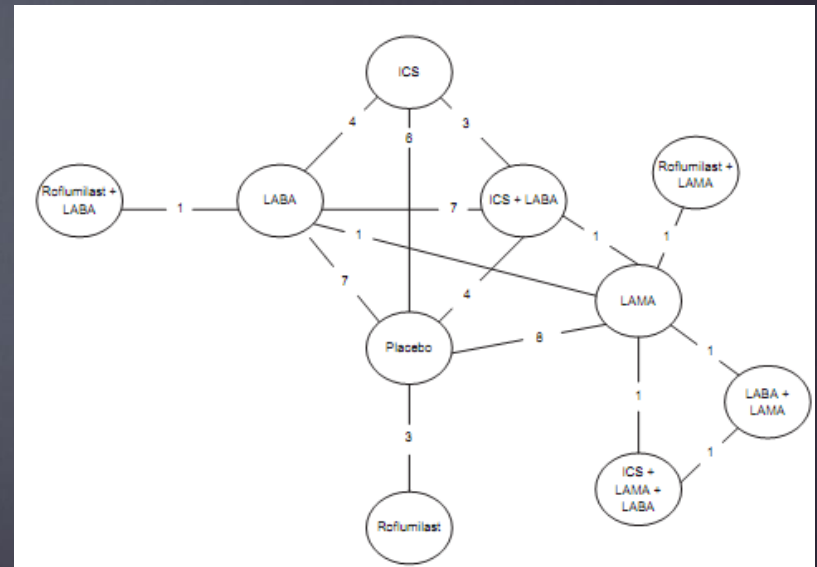
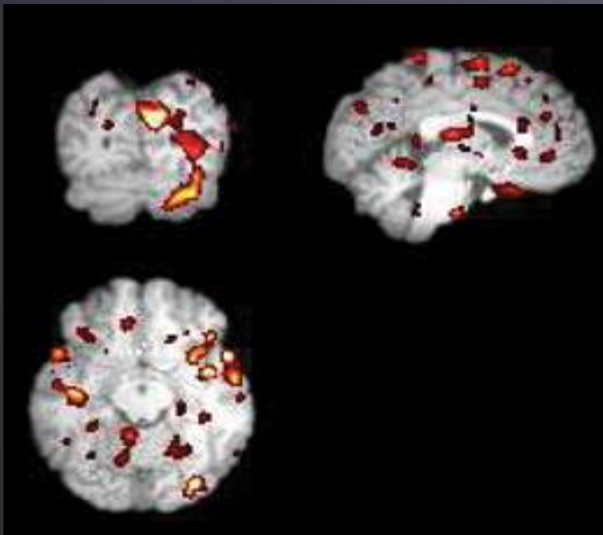
- ▶ Πληροφορίες για την επικοινωνία φαρμάκων και θεραπειών κάποιας ασθένειας

8. Νευρωνικά δίκτυα (neural networks)

- ▶ Πληροφορίες για τον τρόπο μετάδοσης σημάτων στο νευρικό σύστημα

9. Δίκτυα εγκεφάλου (Brain networks)

- ▶ Ο τρόπος που τα διάφορα σημεία του εγκεφάλου ενεργοποιούνται- επικοινωνούν



Brain network: directed links in large scale functional networks
G.A. Cecchi, A.R. Rao, M.V. Centeno, M. Baliki, A.V. Apkarian
& D.R. Chialvo, BMC Cell Biology 8(Suppl 1):S5 (2007)

network of 10 treatments involved in the MTC analyses of the COPD
data: chronic obstructive pulmonary disease;

Αναπαράσταση Δικτύων – Θεωρία Γράφων

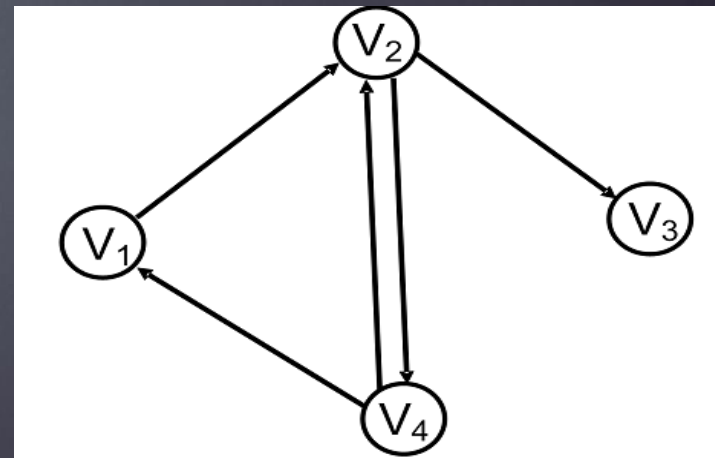
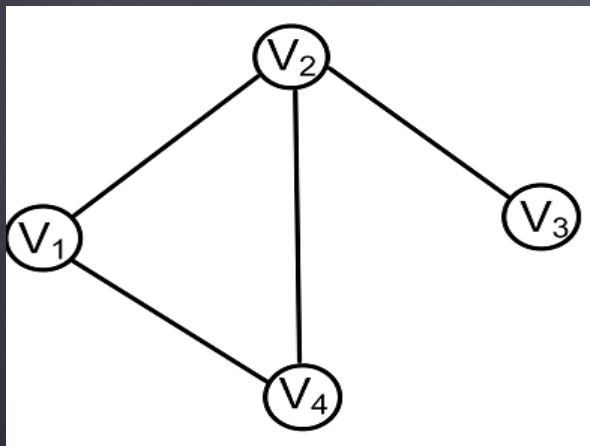
- ▶ Γράφος: μια δομή που αποτελείται από ένα διατεταγμένο ζεύγος $G = (V, E)$

G : ο Γράφος

V : σύνολο κόμβων (*vertices*)

E : σύνολο ακμών (*edges*), οι οποίες ενώνουν τους κόμβους μεταξύ τους

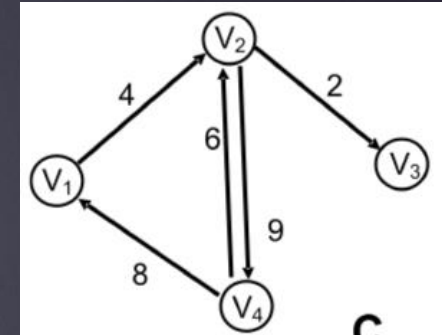
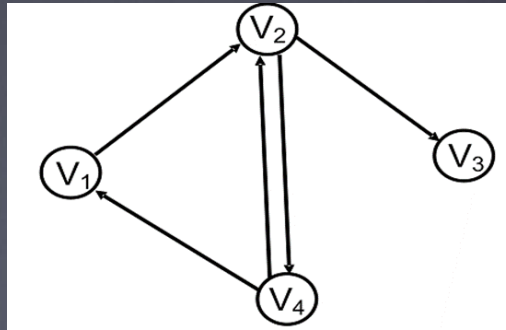
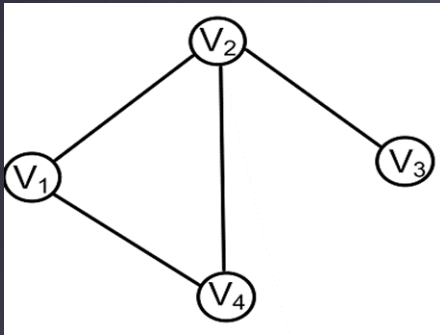
- ▶ Κατευθυνόμενοι (*directed*), μη κατευθυνόμενοι (*undirected*)



ΕΙΔΗ ΓΡΑΦΩΝ

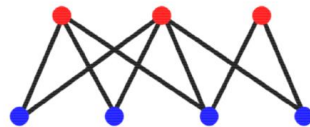
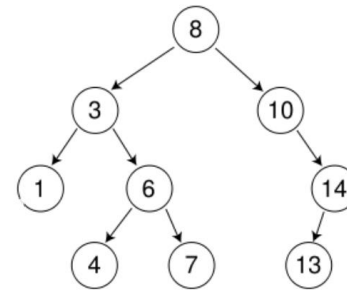
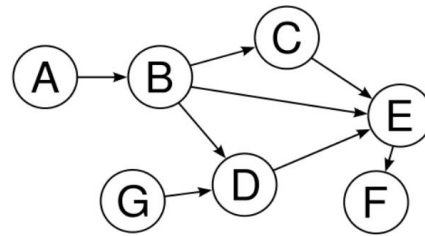
Με βάση τις ακμές:

- Κατευθυνόμενος/μη κατευθυνόμενος
- Σταθμισμένος.

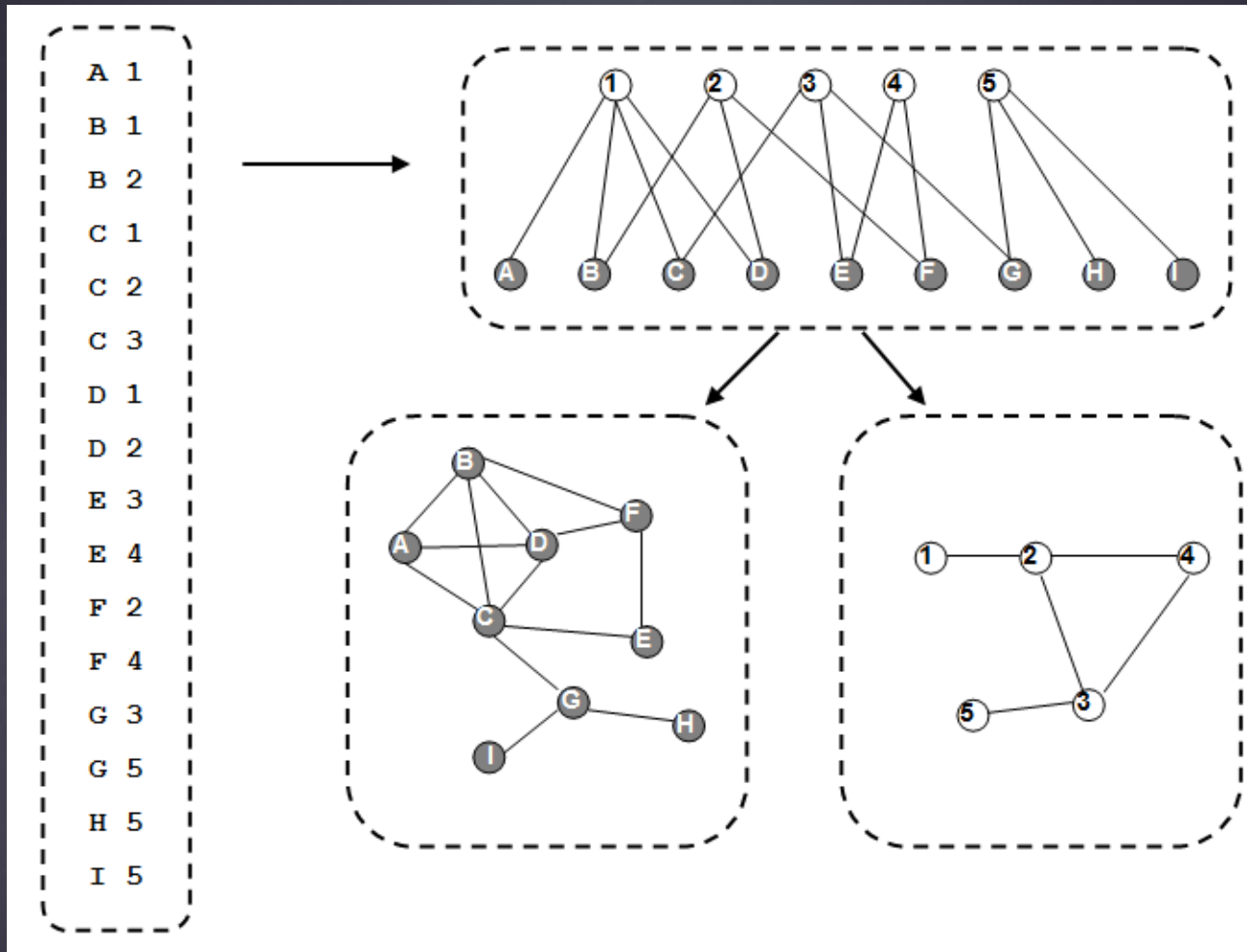


Με βάση τη δομή:

- Κατευθυνόμενοι ακυκλικοί γράφοι (DAG)
- Διαδικά δέντρα.
- Διμερείς γράφοι.

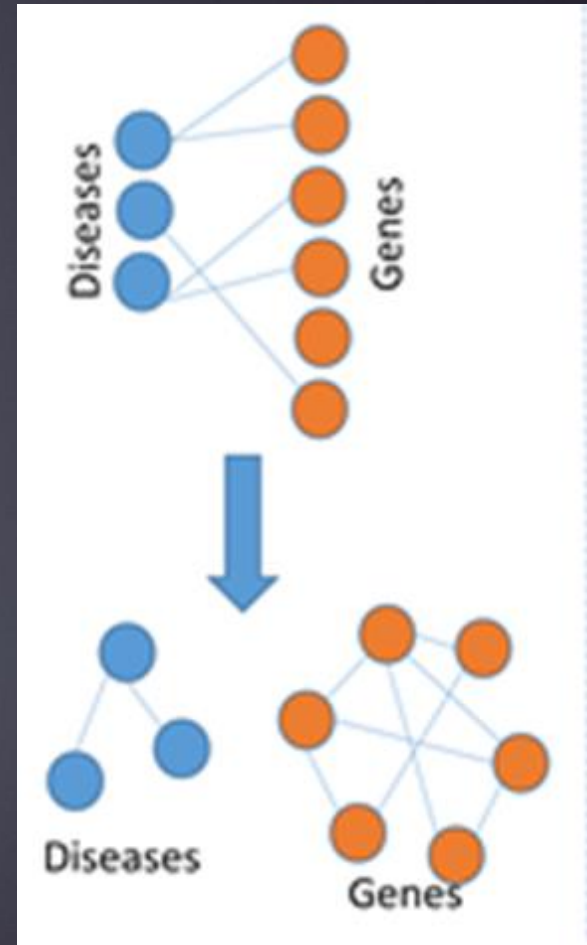


Διμερείς γράφοι (Bipartite graphs)



Παραδείγματα διμερών γράφων

- ▶ Γονίδια/Ασθένειες
- ▶ Ασθένειες/Συμπτώματα
- ▶ Ασθένειες/Φάρμακα

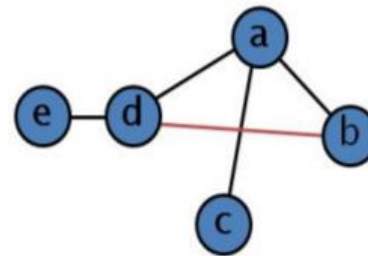


Αναπαράσταση δεδομένων

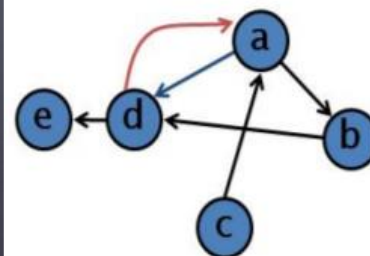
Πίνακας γειτνίασης (adjacency matrix)

- ▶ Η πλήρης μαθηματική αναπαράσταση ενός δικτύου.
- ▶ Παρουσιάζει τις σχέσεις μεταξύ κόμβων με ποσοτικό τρόπο.
- ▶ Περιέχει 0 και 1 για μη σταθμισμένα δίκτυα.
- ▶ Τα μη μηδενικά του στοιχεία μπορεί να είναι οποιοσδήποτε αριθμός (weight) σε σταθμισμένα δίκτυα.

Για έναν γράφο $G = (V, E)$, ο πίνακας γειτνίασης αποτελείται από έναν $|V| \times |V| = n \times n$ πίνακα, $A = (a_{ij})$ έτσι ώστε $a_{ij} = 1$ αν $(i, j) \in E$ ή $a_{ij} = 0$



| | a | b | c | d | e |
|---|---|---|---|---|---|
| a | 0 | 1 | 1 | 1 | 0 |
| b | 1 | 0 | 0 | 1 | 0 |
| c | 1 | 0 | 0 | 0 | 0 |
| d | 1 | 1 | 0 | 0 | 1 |
| e | 0 | 0 | 0 | 1 | 0 |



| | a | b | c | d | e |
|---|---|---|---|---|---|
| a | 0 | 1 | 0 | 1 | 0 |
| b | 0 | 0 | 0 | 1 | 0 |
| c | 1 | 0 | 0 | 0 | 0 |
| d | 1 | 0 | 0 | 0 | 1 |
| e | 0 | 0 | 0 | 0 | 0 |

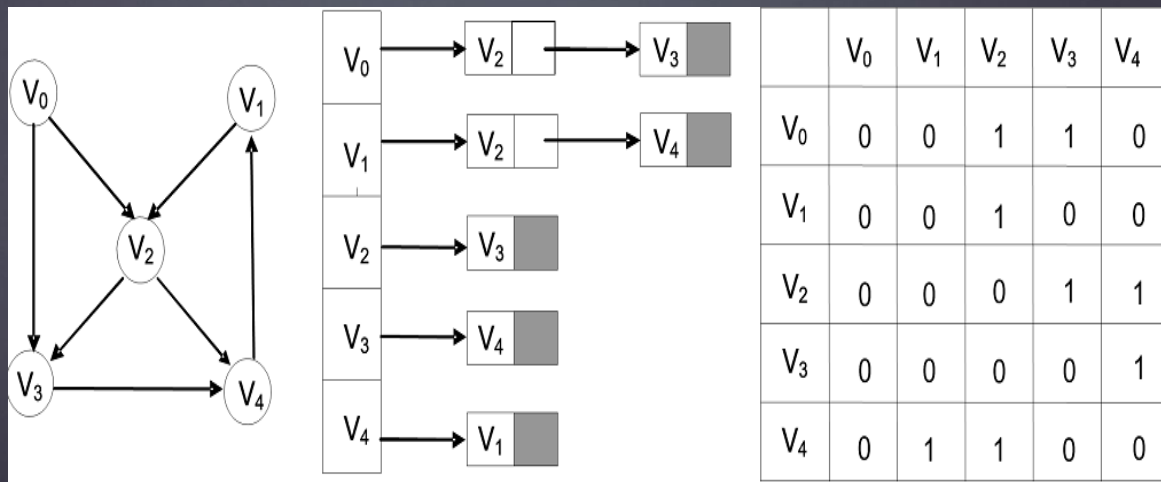
Αναπαράσταση δεδομένων

Λίστα γειτνίασης (adjacency list)

- ▶ Πιο “πυκνή” αναπαράσταση του δικτύου.
- ▶ Κάθε κόμβος γράφεται στη λίστα μαζί με τους κόμβους που είναι συνδεδεμένος.

Ένας γράφος $G=(V,E)$, αναπαριστάται ως ένας μονοδιάστατος πίνακας, όπου

κάθε κόμβος i , είναι δείκτης σε μία συνδεδεμένη λίστα, στην οποία αποθηκεύονται οι κόμβοι που γειτνιάζουν με τον i κόμβο



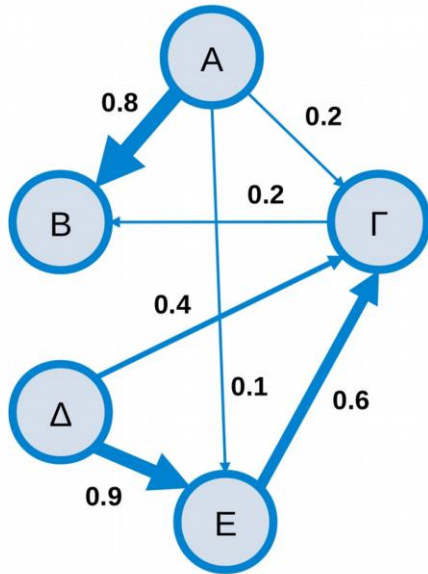
Αναπαράσταση δεδομένων

Λίστα Ακμών (edge list)

- ▶ Κοιτάζει μόνο τις ακμές.
- ▶ Είναι ισοδύναμη με το σύνολο των ακμών.
- ▶ Για μη-κατευθυνόμενους γράφους $(v_i, v_j) = (v_j, v_i)$

Αναπαράσταση ενός κατευθυνόμενου δικτύου με βάρη

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 |

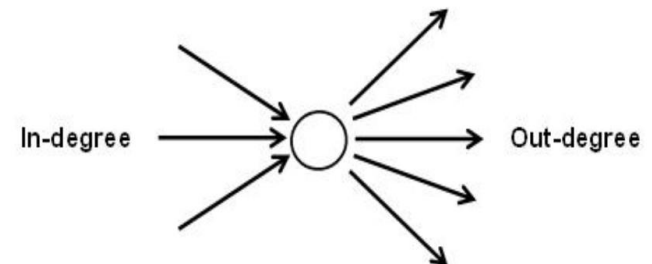
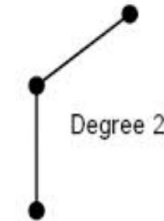
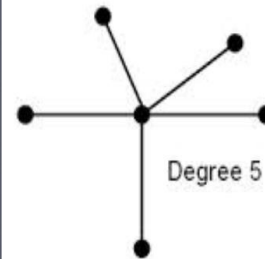
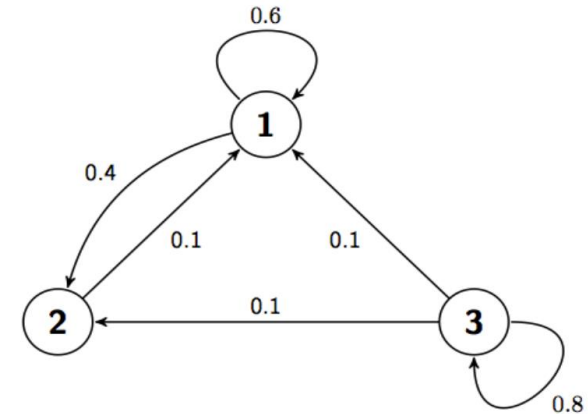
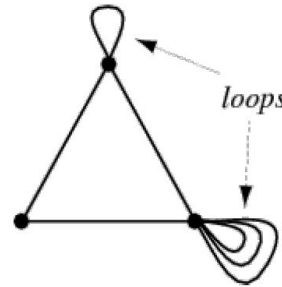
α) Με γράφο όπου οι κατευθύνσεις αποδίδονται με βέλη και τα βάρη αντιστοιχούν στα πάχη των γραμμών

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 $\Gamma \rightarrow A: -0.2$)

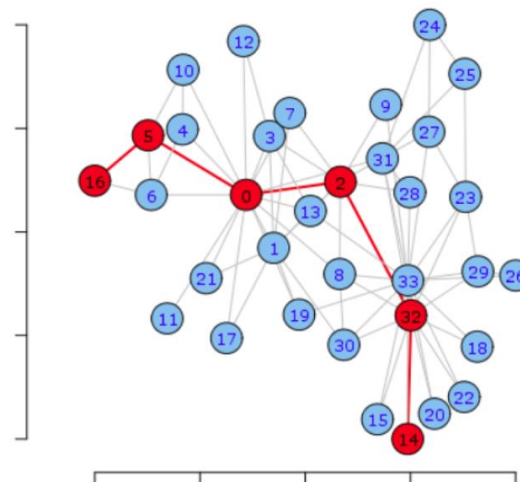
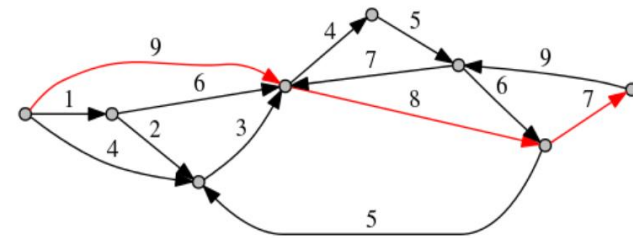
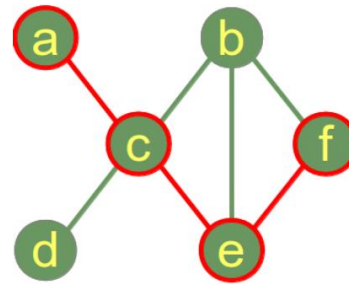
Βασικοί Ορισμοί

- ▶ **Θηλειά/Loop:** Μια ακμή που έχει σαν αρχή και τέλος τον ίδιο κόμβο.
- ▶ **Βαθμός/Degree:** Ο αριθμός των ακμών που συνδέονται σε έναν κόμβο.
 - Μη κατευθυνόμενο: Ο βαθμός είναι το άθροισμα των ακμών.
 - Κατευθυνόμενο: Ο βαθμός χωρίζεται σε εισερχόμενο και εξερχόμενο.



Βασικοί Ορισμοί

- ▶ Μονοπάτι/Path: Μια αλληλουχία ακμών που συνδέει δύο κόμβους. (αν το δίκτυο είναι κατευθυνόμενο ακολουθείται η φορά των ακμών)
- ▶ Κοντινότερο Μονοπάτι/Shortest Path: Το πιο κοντό από όλα τα μονοπάτια που συνδέουν δύο κόμβους. (σε σταθμισμένα δίκτυα αθροίζουμε τα επιμέρους βάρη)
- ▶ Διάμετρος/Diameter: Το μακρύτερο από όλα τα κοντινότερα μονοπάτια ενός δικτύου.



Εισαγωγή στην Ανάλυση Δικτύων

- ▶ Είναι χρήσιμο να αναπτύξουμε τρόπους με τους οποίους θα συγκρίνουμε δίκτυα για να:
 - Αναδείξουμε κοινές αρχές σχεδιασμού (design principles)
 - Ανακαλύψουμε κοινές (ή παρεμφερείς) τοπολογίες.
- ▶ Θα αναπτύξουμε τα κατάλληλα τοπολογικά μέτρα δικτύων τέτοια ώστε:
 - Να είναι απλά (στη χρήση και τον υπολογισμό).
 - Να καταγράφουν την συνολική οργάνωση του δικτύου.
 - Να είναι πιθανώς “σημαντικά”.

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 χιλιάδες 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)

- ▶ 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.

$$\frac{2 | E |}{| V | (| V | - 1)}$$

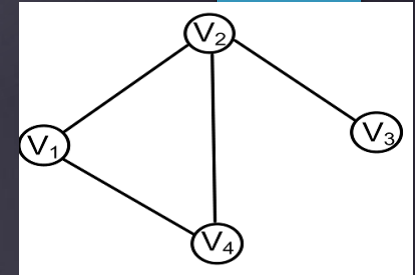
Βαθμός-Node degree

- Ο συνολικός αριθμός των ακμών που προσπίπτουν σε έναν κόμβο
- Κόμβοι με ισχυρή συνδεσιμότητα (high degree) ονομάζονται “hubs”

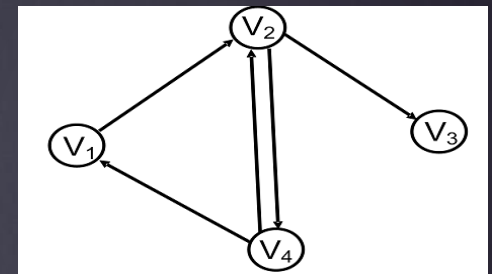
• Undirected: $C_d(i) = \text{deg}(i)$

• Directed: $C_{d_{in}}(i) = \text{deg}_{in}(i)$

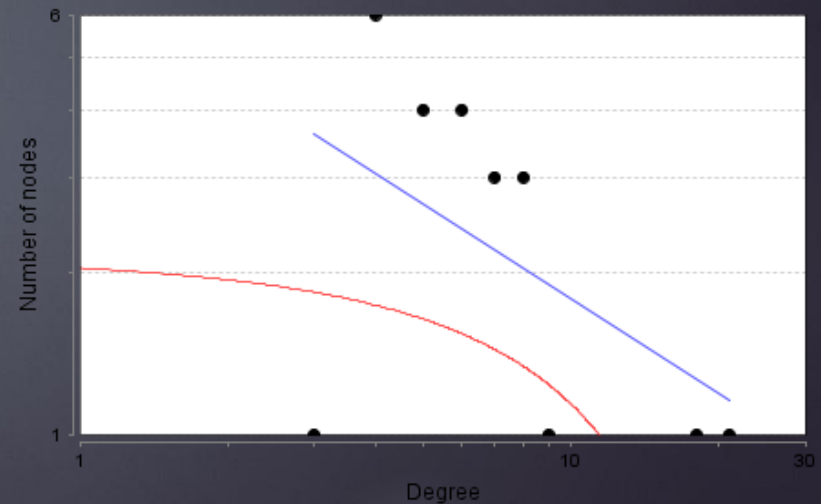
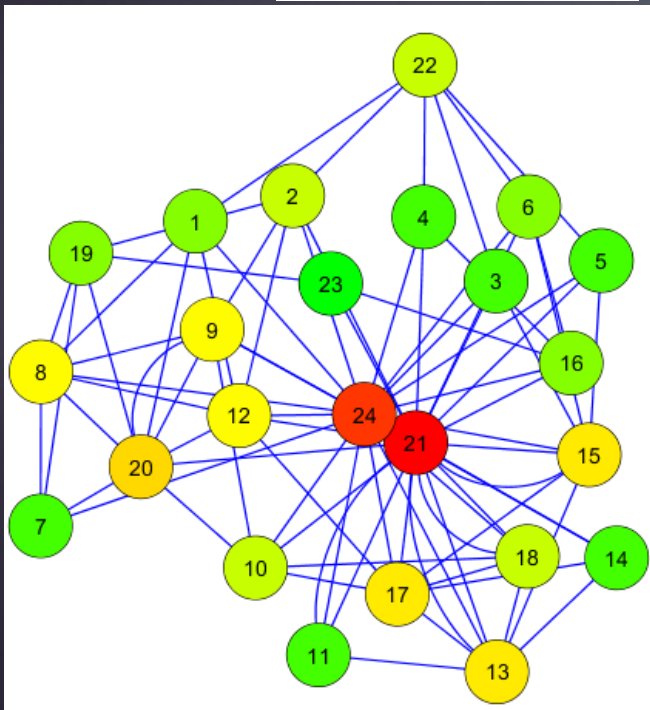
$C_{d_{out}}(i) = \text{deg}_{out}(i)$



Degree (v2): 3



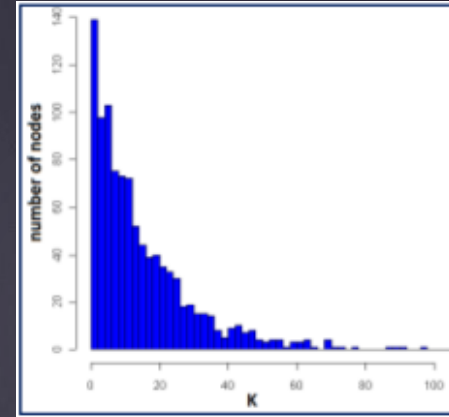
Indegree (v2):2 Outdegree(v2):2



Κατανομές Βαθμών

- ▶ Κατανομή Βαθμών: Η πιθανότητα $P(k)$ ένας κόμβος να έχει βαθμό ακριβώς k .

- ▶ Οι πιο διαδεδομένες κατανομές βαθμών:



Poisson

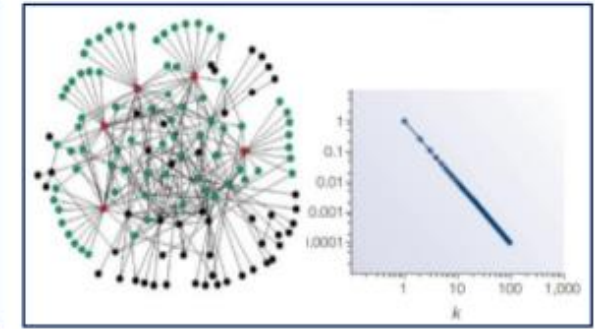
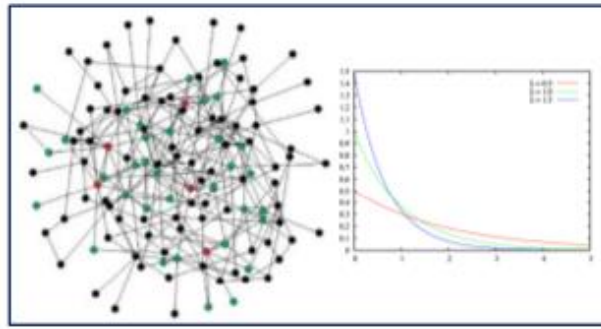
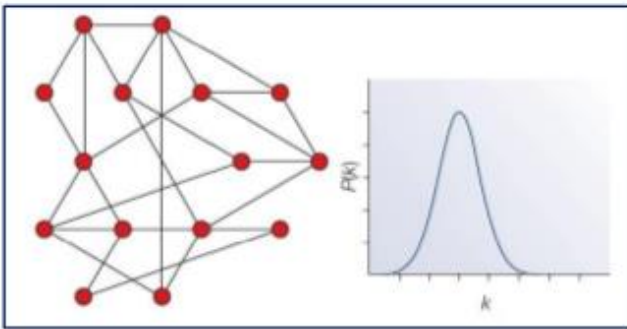
Εκθετική

Νόμου δύναμης (Power law)

$$P(k) = \frac{e^{-d} d^k}{k!}$$

$$P(k) \propto e^{-k/d}$$

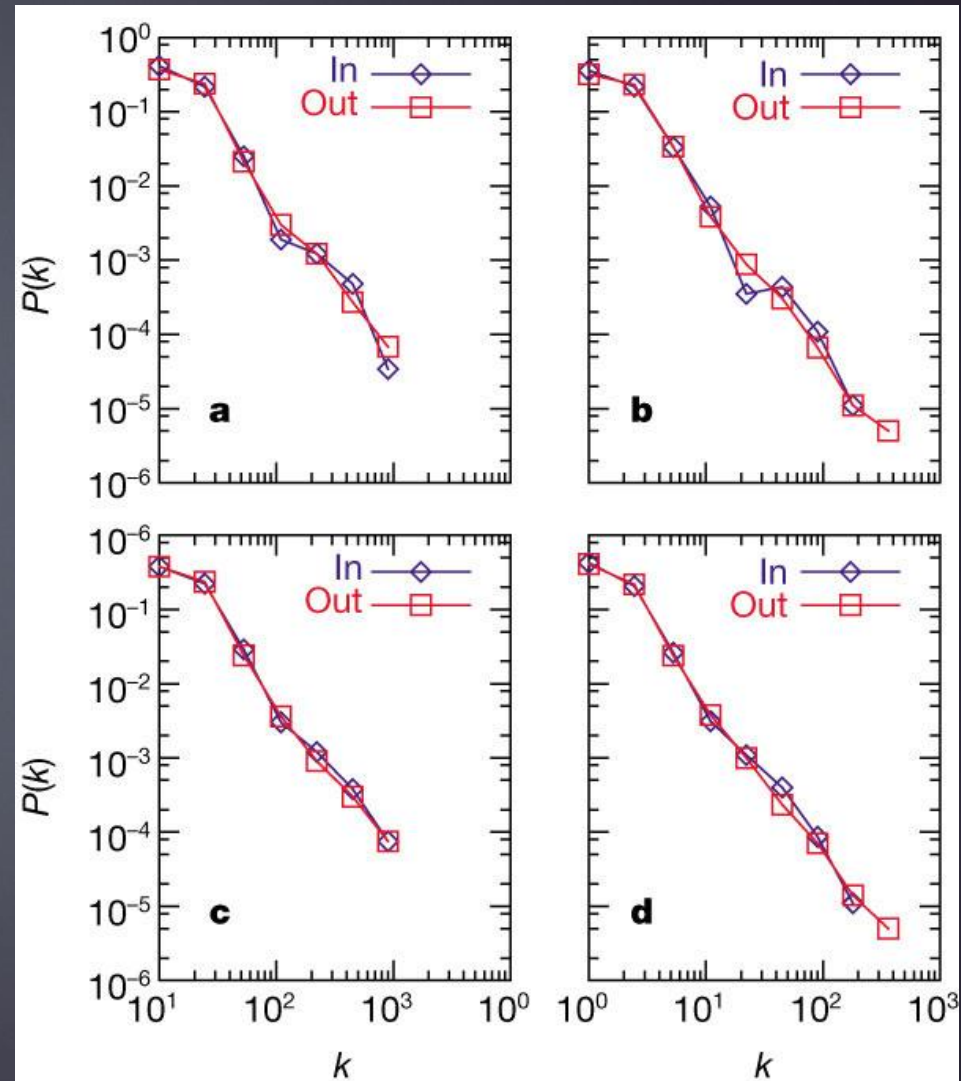
$$P(k) \propto k^{-c}, k \neq 0, c > 1$$



Κατανομές Βαθμών Γνωστών Δικτύων

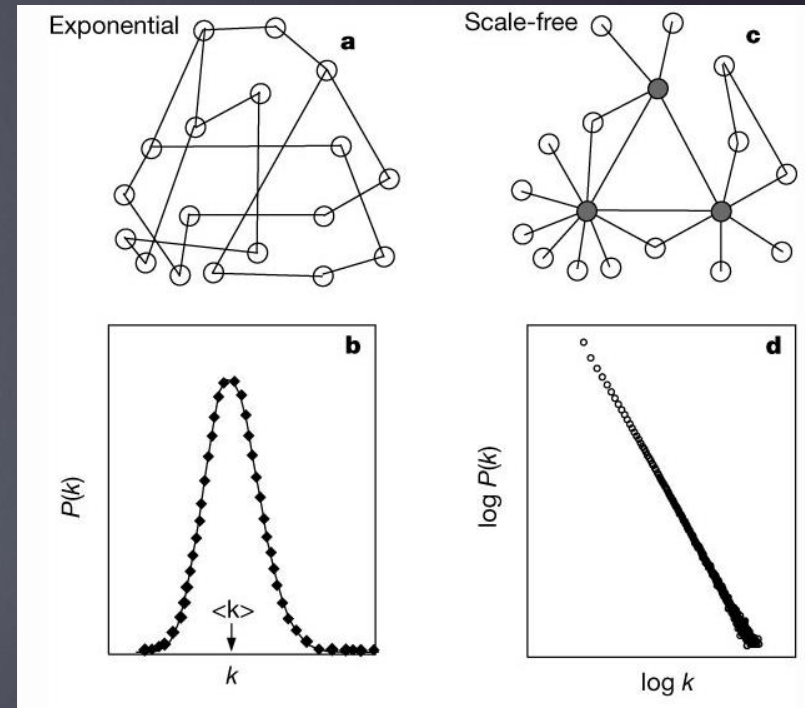
Μεταβολικά Δίκτυα

- ▶ Κόμβοι: Μεταβολίτες
- ▶ Ακμές: (χημικές αντιδράσεις).
- ▶ $P(k) \sim k^{2.2 \pm 2}$
- ▶ Τα μεταβολικά δίκτυα σε όλα τα φύλα είναι ελεύθερα κλίμακας (scale-free) A: *A.fulgidus*, B: *E.coli*, C: *C.elegans*, D: average



Κατανομές Νόμου Δύναμης και Δίκτυα Ελεύθερης Κλίμακας (Power law distribution and Scale-free Networks)

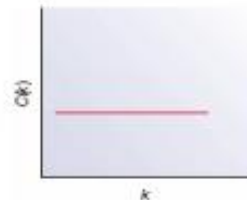
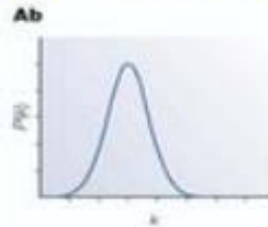
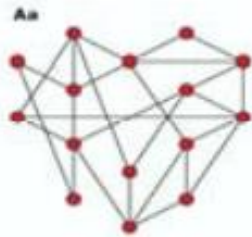
- ▶ Οι κατανομές νόμου δύναμης έχουν “παχιά ουρά”.
- ▶ Χαρακτηρίζονται από μεγάλο αριθμό κόμβων με πολύ λίγες συνδέσεις και μικρό αριθμό με πάρα πολλές συνδέσεις (hubs ή κεντρικούς κόμβους).
- ▶ Ονομάζονται ελεύθερης-κλίμακας, γιατί ανεξάρτητα από την πυκνότητα των συνδέσεων:
- ▶ Δεν έχουν κάποια “χαρακτηριστική” τιμή (μέσο όρο).
- ▶ Οι βαθμοί των κόμβων τους ακολουθούν την κατανομή: $P(k) \sim k^{-\gamma}$



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).

A. Random Networks [Erdos and Rényi (1959, 1960)]



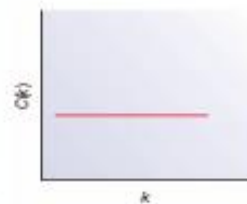
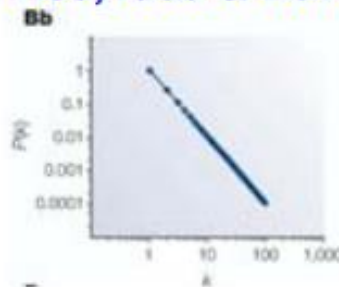
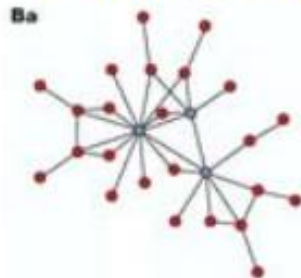
$$P(k) = \frac{e^{-\bar{k}} \bar{k}^k}{k!}$$

Mean path length $\sim \ln(k)$

Phase transition:

Connected if: $p \geq \ln(k) / k$

B. Scale Free [Price, 1965 & Barabasi, 1999]



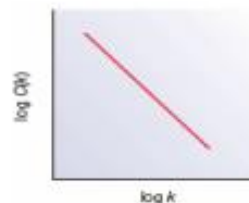
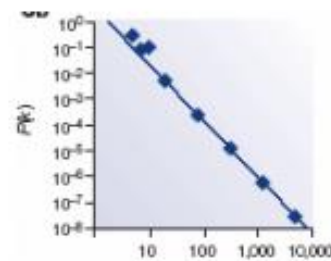
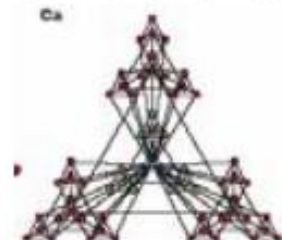
$$P(k) \sim k^{-\gamma}, \quad k \gg 1, \quad 2 < \gamma$$

Mean path length $\sim \ln \ln(k)$

Preferential attachment. Add proportionally to connectedness



C. Hierarchical

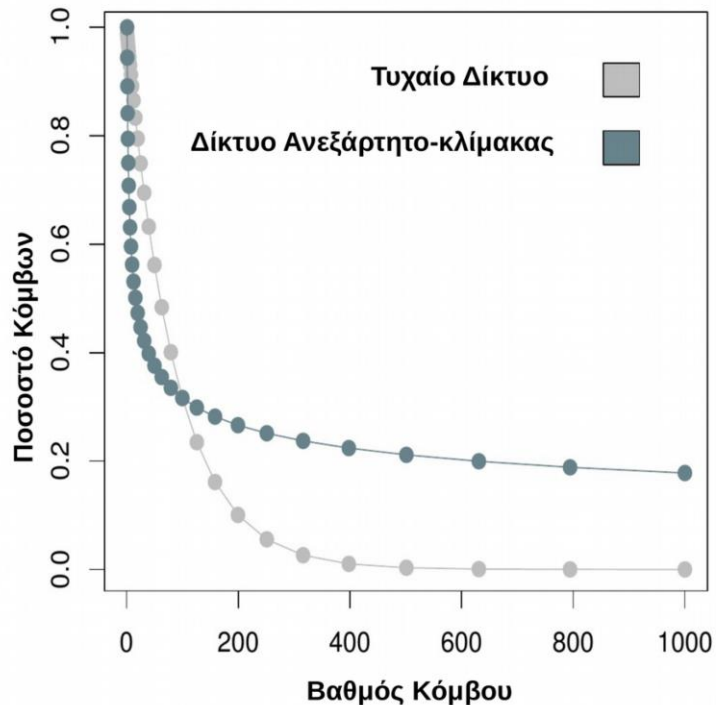


Copy smaller graphs and let them keep their connections.

Τυχαία δίκτυα και Δίκτυα άνευ-κλίμακας

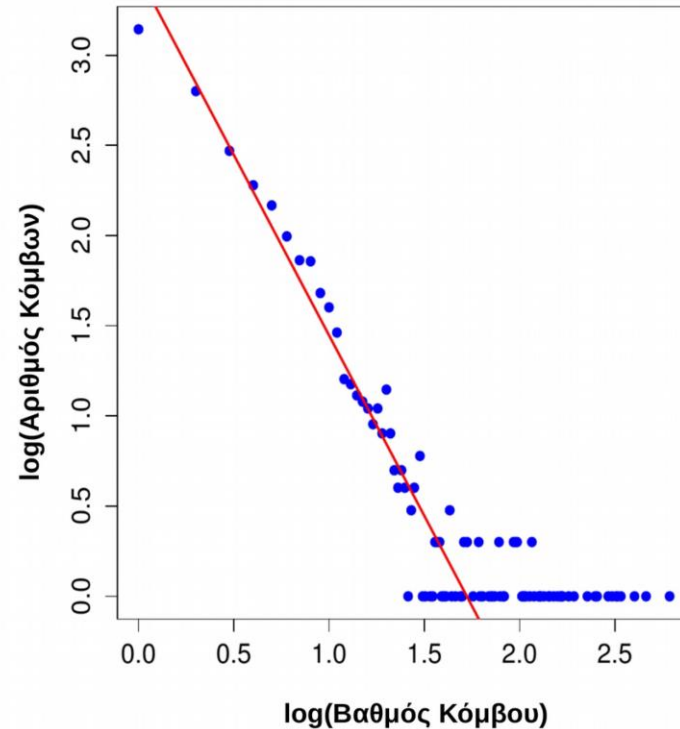
α

Κατανομή Βαθμού Κόμβων



β

Δίκτυο Πρωτεϊνικών Αλληλεπιδράσεων



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, β) όπου αναπαρίσταται γραφικά η κατανομή βαθμού κόμβων για ένα δίκτυο πρωτεϊνικών αλληλεπιδράσεων

Απλές μετρήσεις

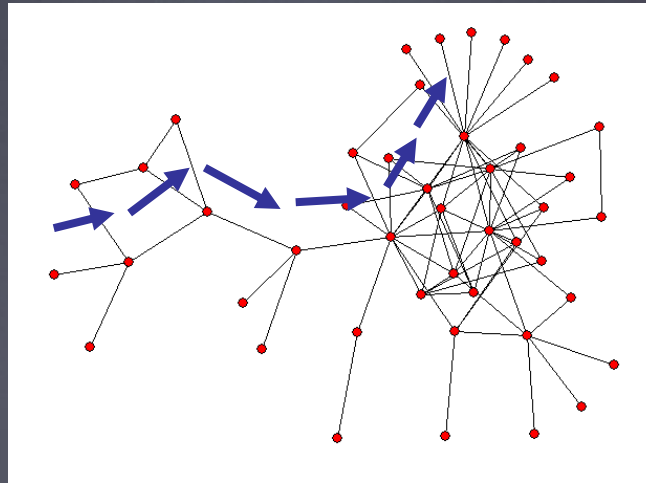
- ▶ Network density \longrightarrow Πόσο αραιός ή πυκνός είναι ένας γράφος

$$\frac{2|E|}{|V|(|V|-1)}$$

$$|E| \approx |V|$$

$$|E| \approx |V|^2$$

- ▶ Shortest paths \longrightarrow ονομάζεται απόσταση $\delta(i, j)$ από τον i κόμβο στον j



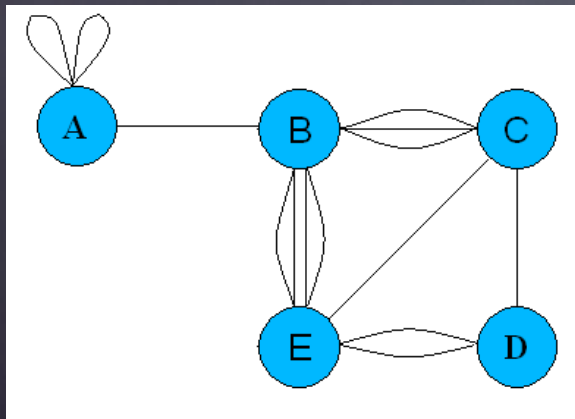
Η συντομότερη διαδρομή (shortest path) μεταξύ δύο δεδομένων κόμβων, αντιστοιχεί στη διαδρομή με την οποία μπορεί κανείς να βρεθεί από τον έναν στον άλλον, ακολουθώντας τις ακμές του δικτύου και διατρέχοντας τον ελάχιστο αριθμό κόμβων. Είναι δηλαδή το συντομότερο μονοπάτι μεταξύ δύο κόμβων και ορίζεται ως η διαδοχή των κόμβων μεταξύ των σημείων έναρξης και τερματισμού.

Απλές μετρήσεις

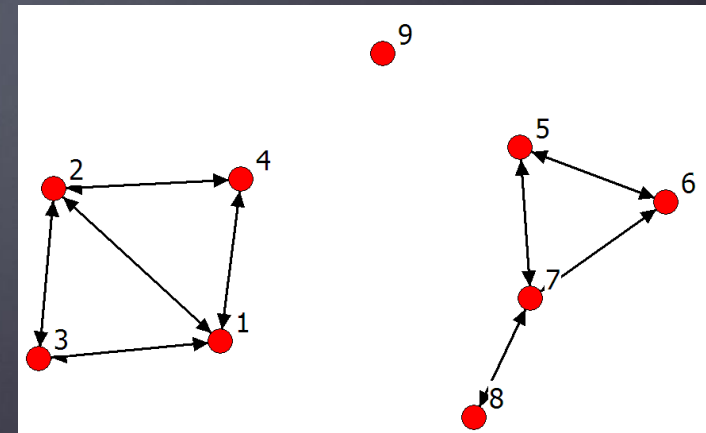
- ▶ Network diameter \longrightarrow η μέγιστη τιμή της απόστασης $D = \max$
- ▶ Network radius \longrightarrow η ελάχιστη τιμή της απόστασης $D = \min \delta_{\min}(i, j)$
- ▶ Characteristic path length \longrightarrow η μέση τιμή της απόστασης $D = \text{avg}$

Απλές μετρήσεις

- ▶ Connected components \implies όλοι οι κόμβοι που ενώνονται ανά ζεύγη
- ▶ Isolated nodes \implies κόμβοι απομονωμένοι χωρίς σύνδεση
- ▶ Number of self loops \implies αριθμός βρόγχων
- ▶ Multi-edge node pairs \implies παραπάνω από μια σύνδεση σε δύο κόμβους
- ▶ Avg. number of neighbors \implies μέσος όρος σύνδεσης του κόμβου στο δίκτυο



Multi-graph



3 components (1,2,3,4), (5,6,7,8),(9) 1 isolate (9).

Συντελεστής Ομαδοποίησης - Clustering Coefficient (C_i)

- ▶ Πόσοι από τους “φίλους” μου γνωρίζονται και φτιάχνουν μεταξύ τους μια ομάδα.
- ▶ Μέτρο της πυκνότητας των τριάδων μέσα σε ένα δίκτυο.

$$C_i = \frac{\# \text{ ακμών μεταξύ γειτόνων του } i}{\text{Μέγιστος } \# \text{ ακμών μεταξύ γειτόνων του } i}$$

Τοπικό μέτρο

$$C_i = \frac{2E_i}{d_i(d_i - 1)}$$

Συνολικό μέτρο

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i$$

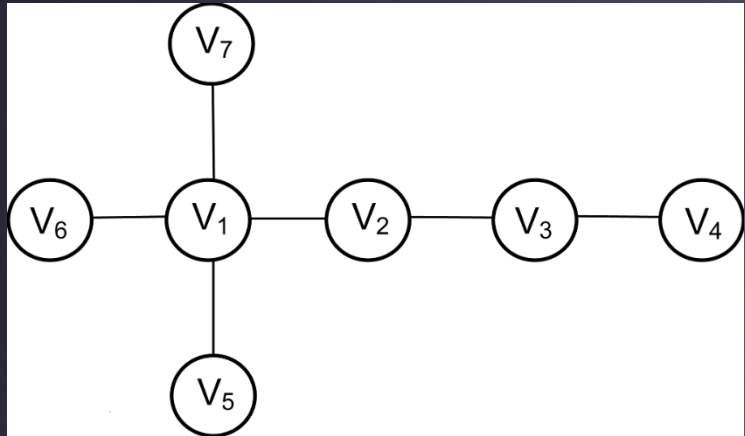
Συντελεστής Ομαδοποίησης - Clustering Coefficient (C_i)

- ▶ Ο συντελεστής ομαδοποίησης 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 ομαδοποίησης 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.
- ▶ Ο συντελεστής ομαδοποίησης 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.

Συντελεστής Ομαδοποίησης - Clustering Coefficient (C_i)

- ▶ Ο Συντελεστής Ομαδοποίησης παίρνει τιμές μεταξύ 0 και 1 και έχει χαρακτηριστικά πιθανότητας.
- ▶ 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.

Ενδιάμεση κεντρικότητα (Betweenness centrality)



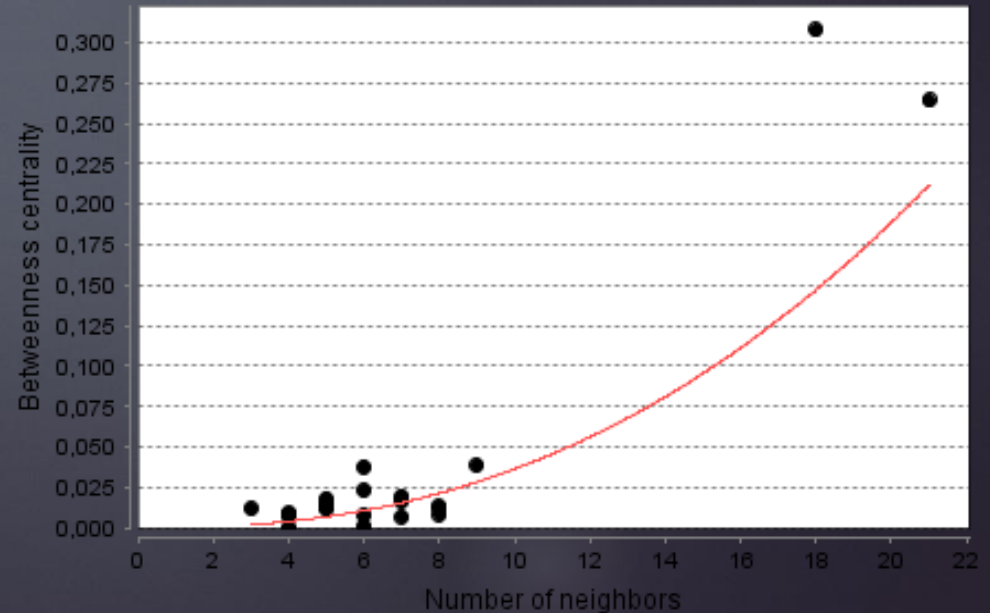
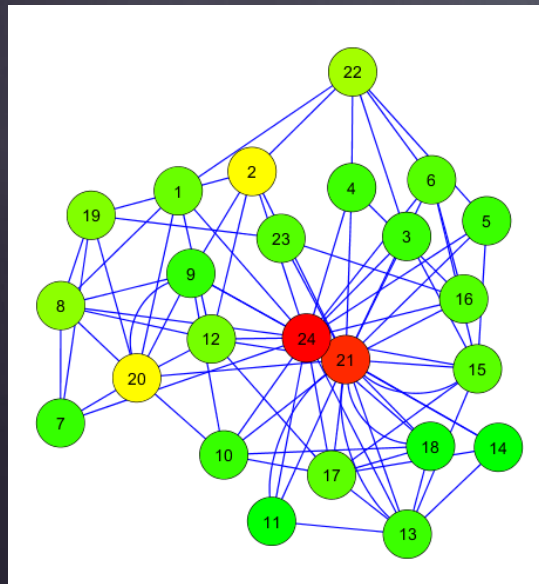
- Δείχνει τους σημαντικούς κόμβους που βρίσκονται σε υψηλό ποσοστό στα μονοπάτια άλλων κόμβων σε ένα δίκτυο

$$C_b(w) = \sum_{(i,j) \in V(w)} \frac{\sigma_{ij}(w)}{\sigma_{ij}}$$

$N_p(1) = 12$, $N_p(2) = 8$, $N_p(3) = 5$, $N_p(4) = N_p(5) = N_p(6) = N_p(7) = 0$

$N_p = N_p(1) + N_p(2) + N_p(3) + N_p(4) + N_p(5) + N_p(6) + N_p(7) = 25$

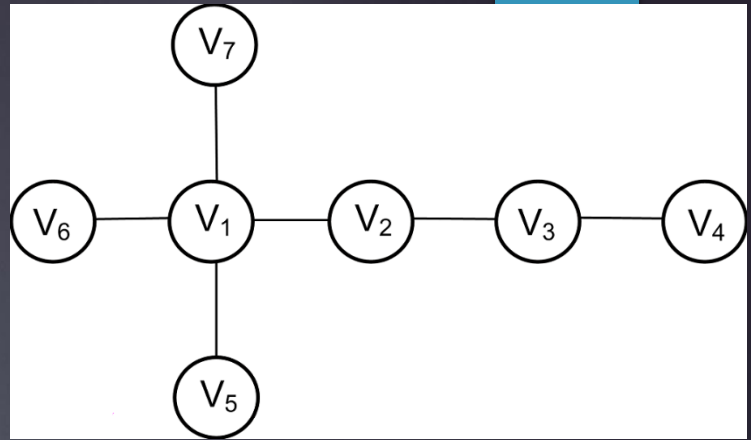
$C_b(1) = 12/25$, $C_b(2) = 8/25$, $C_b(3) = 5/25$, $C_b(4) = C_b(5) = C_b(6) = C_b(7) = 0$



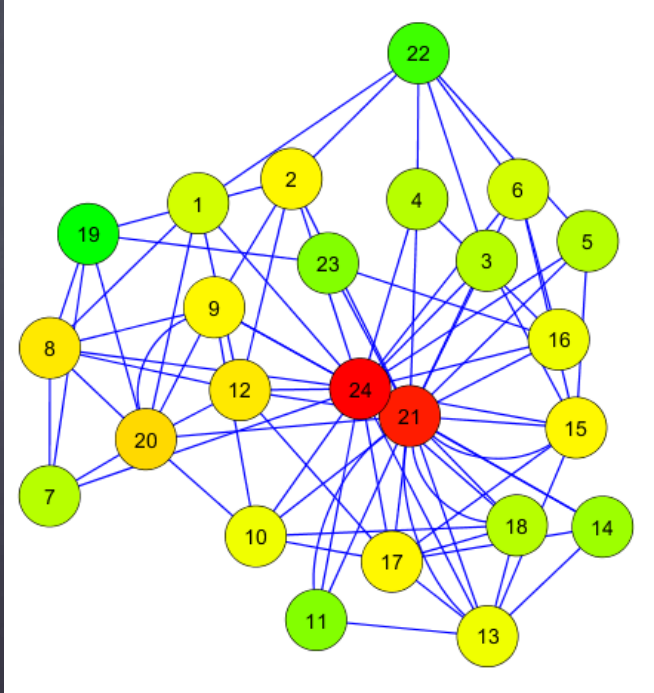
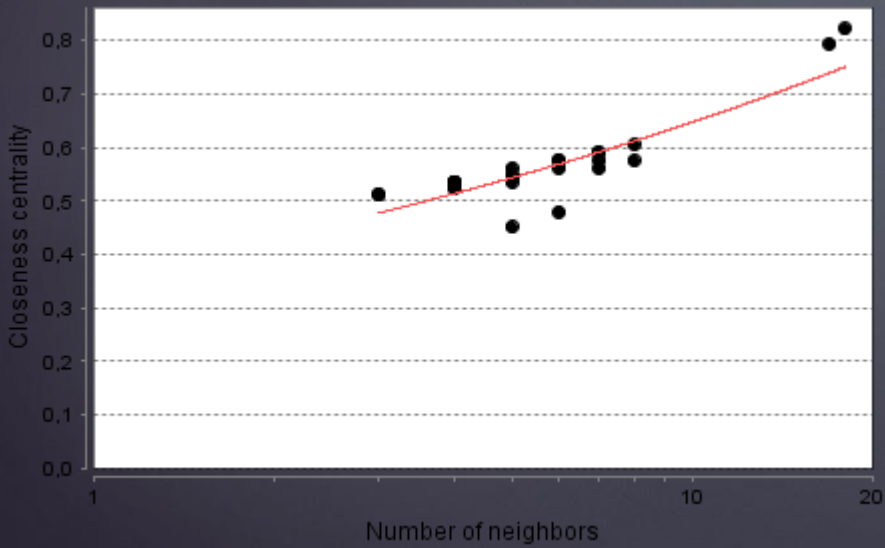
Closeness centrality

• Η μέτρηση αυτή υποδεικνύει τους σημαντικούς κόμβους οι οποίοι μπορούν να επικοινωνήσουν γρήγορα με άλλους κόμβους του δικτύου

$$C_{clo}(i) = \frac{1}{\sum_{t \in V} dist(i, j)}$$



• $d_1 = 4 \times 1 + 1 \times 2 + 1 \times 3 = 9$ $C_{clo}(1) = 6/9$
 • $d_2 = 2 \times 1 + 4 \times 2 = 10 > d_1$ και $C_{clo}(2) = 6/10$
 ο V1 είναι περισσότερο σημαντικός και κεντρικός από τον V2 διότι $d_1 > d_2$



Συμπεράσματα

- ▶ Η μελέτη των δικτύων αποτελεί ένα ιδιαίτερα ενεργό επιστημονικό πεδίο.
- ▶ Τροφοδοτείται από την αύξηση του όγκου των διαθέσιμων δεδομένων σε πολλούς ερευνητικούς τομείς.
- ▶ 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

Βιβλιογραφία

- ▶ Κεφάλαιο 9 Βιολογικά Δίκτυα, Νικολάου, Χ., Χουβαρδάς, Π., 2015. Υπολογιστική βιολογία. [ηλεκτρ. βιβλ.] Αθήνα:Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: <http://hdl.handle.net/11419/1577>
- ▶ <http://barabasi.com/networksciencebook/>

