Σύνθετα Δίκτυα

com+plex: with+ -fold (having parts)

Διδάσκων –
Δημήτριος Κατσαρός

Διάλεξη 4η

Δυναμο-νόμοι

Power-laws

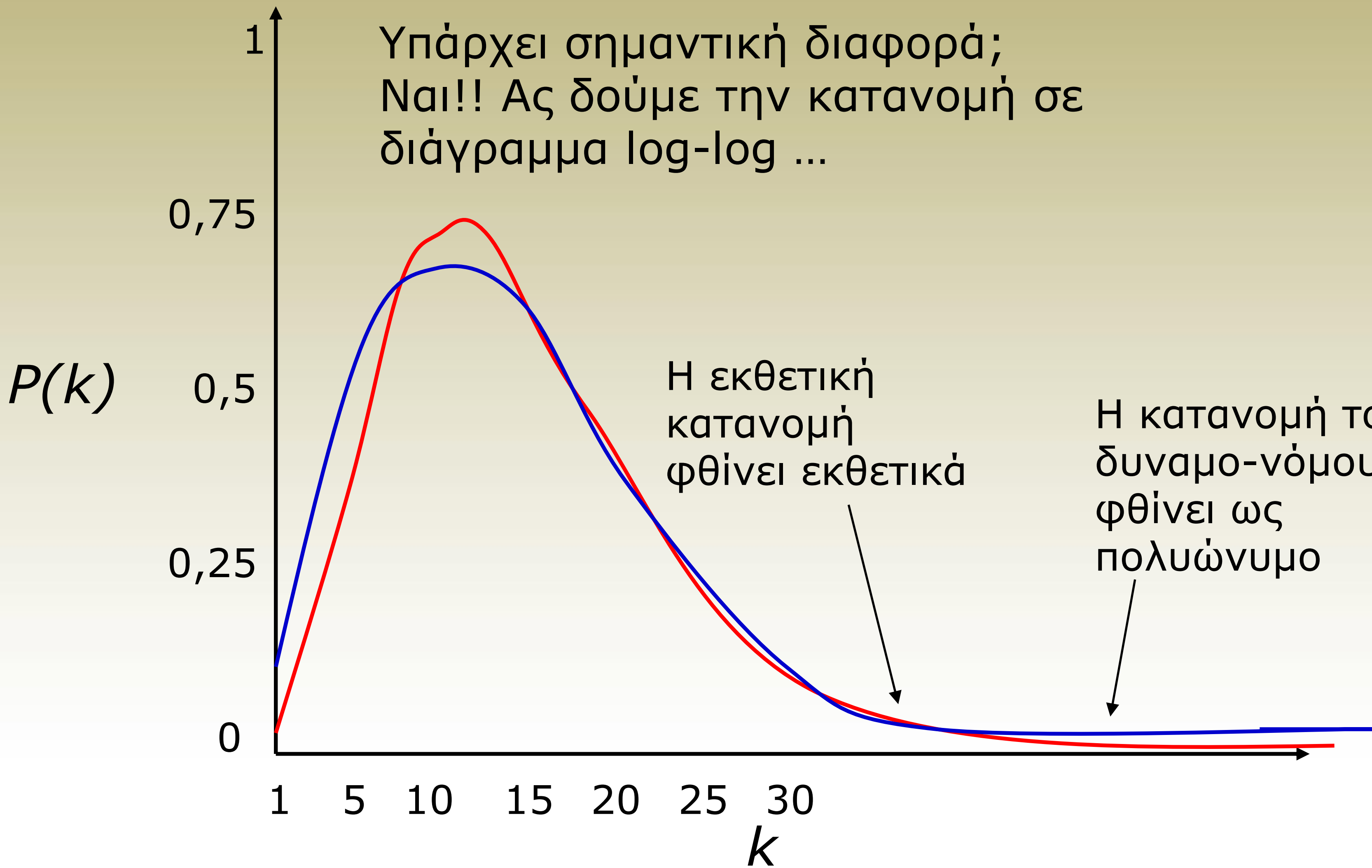
Η κατανομή του δυναμο-νόμου

- Τα περισσότερα δίκτυα του πραγματικού κόσμου υπακούουν σε κάποιον “δυναμο-νόμο” για τη συνδεσμικότητα των κόμβων
- Όπως είδαμε, μια κατανομή πιθανότητας είναι “δυναμο-νόμος” εάν
  - Η πιθανότητα $P(k)$ ότι μια δεδομένη μεταβλητή k έχει συγκεκριμένη τιμή
  - Φθίνει ανάλογα προς το k εις την $-\gamma$, όπου γ είναι κάποια σταθερά
- Για τα δίκτυα, αυτό υπονοεί ότι
  - Η πιθανότητα κάποιου κόμβου να έχει k συνδέσεις
  - Είναι ανάλογη του $\alpha k^{-\gamma}$

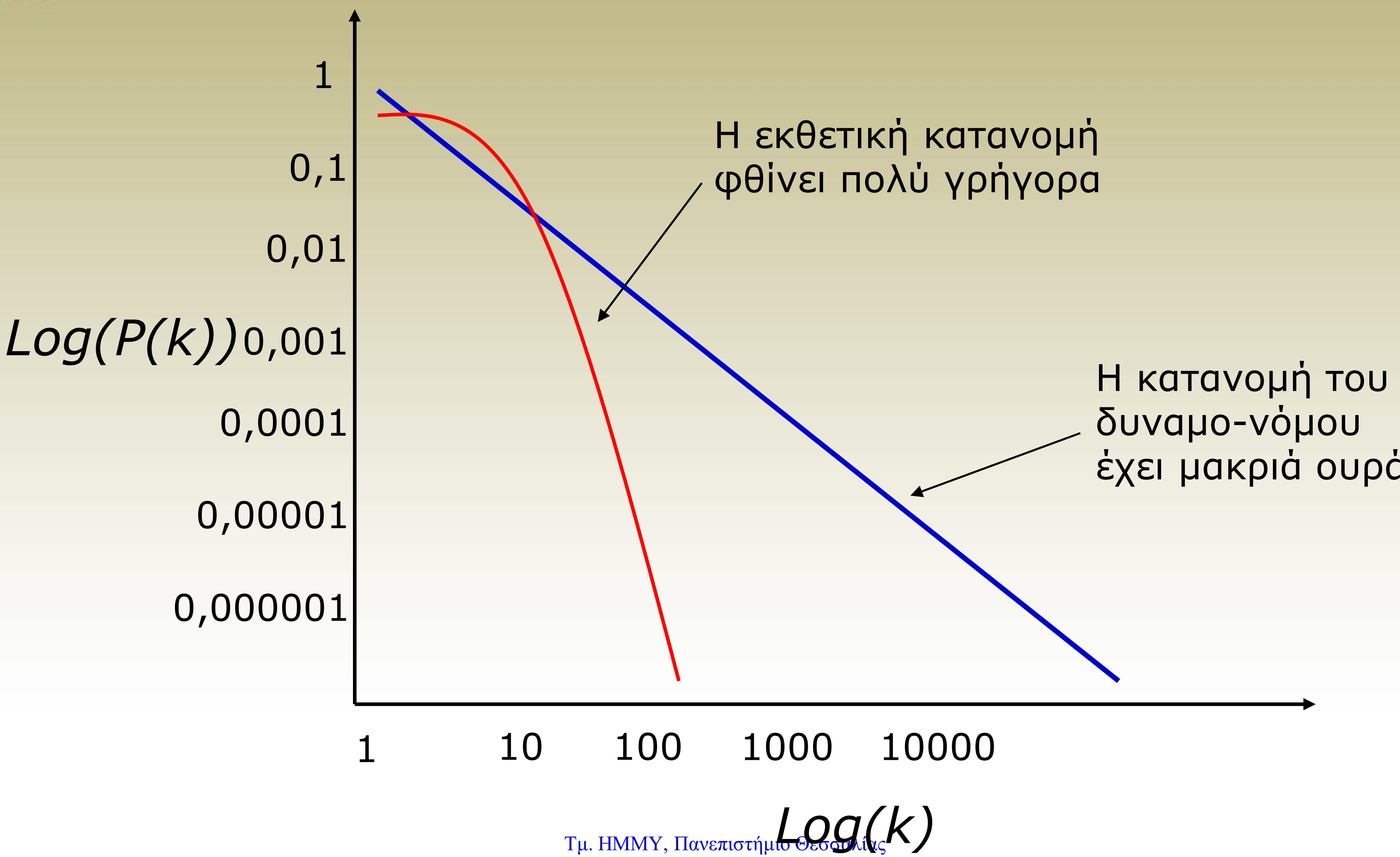
$$P(k) = \alpha k^{-\gamma}$$

Δυναμο-νόμος vs. εκθετική κατανομή

Υπάρχει σημαντική διαφορά;
Ναι!! Ας δούμε την κατανομή σε διάγραμμα log-log …



Δυναμο-νόμος vs. εκθετική κατανομή



Η παχιά ουρά

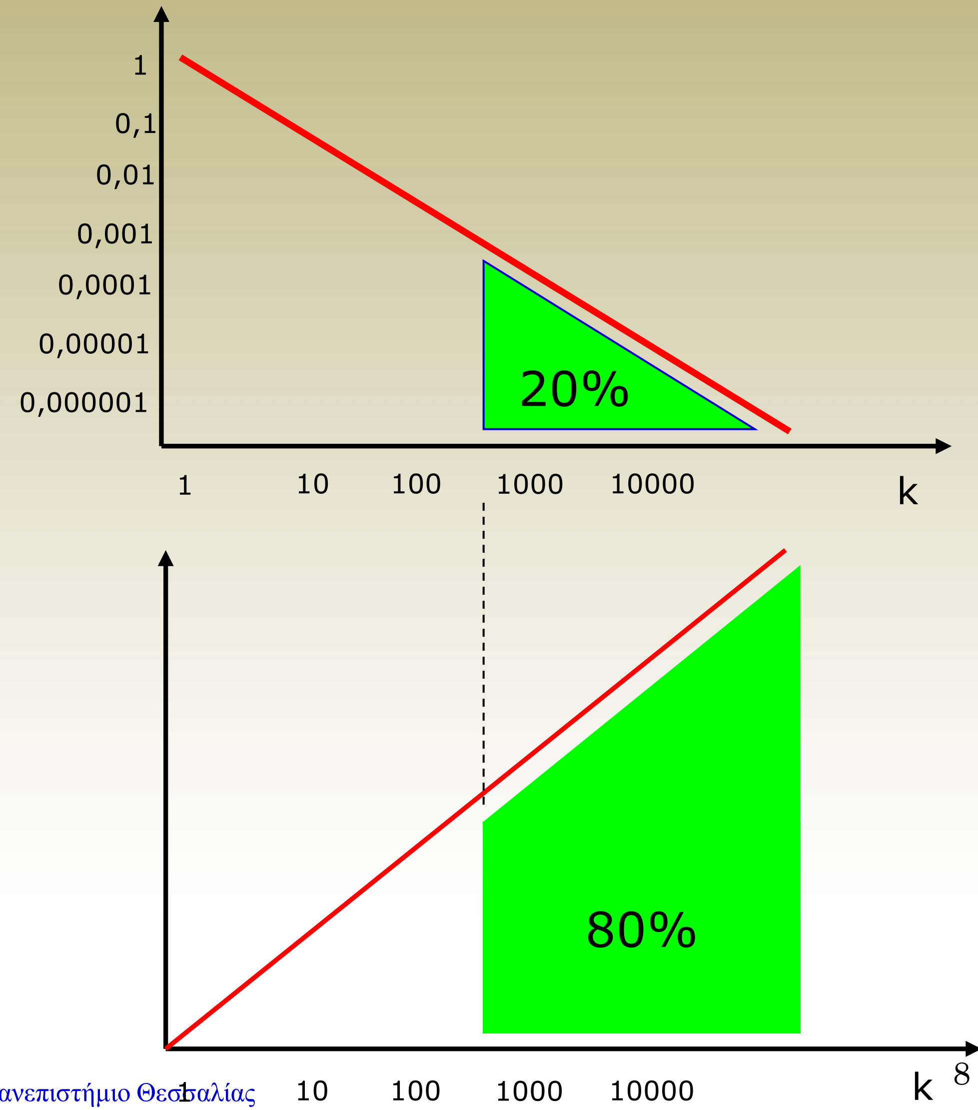
- Η κατανομή του δυναμο-νόμου υπονοεί “infinite variance”
  - Το εμβαδόν των “μεγάλων *ks*” σε μια εκθετική κατανομή τείνει στο μηδέν όταν k→∞
  - Αυτό δεν ισχύει για την κατανομή ενός δυναμο-νόμου, υπονοώντας infinite variance
  - Η ουρά της κατανομής έχει αξία!!!
- Με άλλα λόγια, ο δυναμο-νόμος υπονοεί ότι
  - Η πιθανότητα να έχουμε στοιχεία πολύ μακριά από τον μέσο όρο δεν είναι αμελητέα
  - Όπως λέμε, “ … οι μεγάλοι αριθμοί μετράνε”
- Χρησιμοποιώντας μια εκθετική κατανομή
  - Η πιθανότητα μια Web page να έχει περισσότερους από 100 εισερχόμενους συνδέσμους, εξετάζοντας τον μέσο αριθμό συνδέσμων ανά σελίδα, θα ήταν μικρότερη από 1^{-20}
  - Το οποίο αντικρούει το γεγονός ότι γνωρίζουμε πολλά “well linked” sites …

Ο κανόνας του 20-80

- Examples
  - The 20% of the Web sites gests the 80% of the visits (actual data: 15%-85%)
  - The 20% of the Internet routers handles the 80% of the total Internet traffic
  - The 20% of world industries hold the 80% of the world's income
  - The 20% of the world population consumes the 80% of the world's resources
  - The 20% of the Italian population holds the 80% of the lands (that was true before the Mussolini fascist regime, when lands re-distribution occurred)
  - The 20% of the earthquakes caused the 80% of the victims
  - The 20% of the rivers in the world carry the 80% of the total sweet water
  - The 20% of the proteins handles the 80% of the most critical metabolic processes
- Does this derive from the power law distribution? YES!

Τα μυστικά του 20-80

- Το 20% του πληθυσμού
  - Θυμηθείτε ότι το εμβαδόν αναπαριστά τον πληθυσμό της κατανομής
- Παίρνει το 80% των πόρων
  - Στην πραγματικότητα, μπορεί να βρεθεί ότι το ποσό των πόρων (δηλ., το ποσό των συνδέσεων του δικτύου) είναι το ολοκλήρωμα του $P(k)*k$, το οποίο είναι σχεδόν γραμμικό

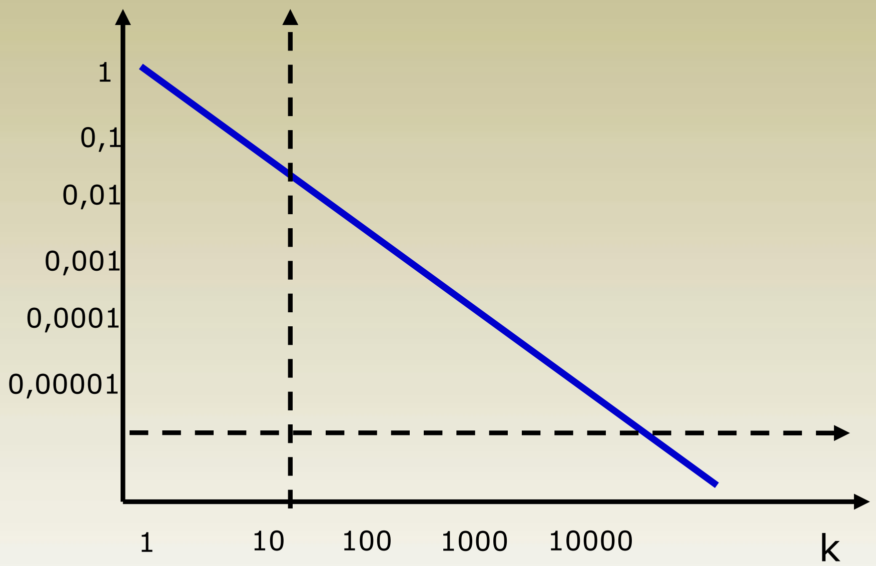


Hubs (Πλήμνες) και connectors

- Τα δίκτυα άνευ κλίμακας περιέχουν κόμβους, οι οποίοι
  - Λειτουργούν ως **hubs**, δηλ., ως σημεία στα οποία συνδέονται οι περισσότεροι υπόλοιποι κόμβοι
  - Λειτουργούν ως **connectors**, δηλ., κόμβοι που συνεισφέρουν τα μέγιστα στο να κρατούν ένα μεγάλο μέρος του δικτύου συνδεδεμένο
  - Υπάρχουν "μικρότεροι κόμβοι" που λειτουργούν ως hubs ή connectors για τοπικά μέρη-τμήματα του δικτύου
- Αυτό μπορεί να έχει σοβαρές επιπτώσεις, όπως θα δούμε στη συνέχεια των διαλέξεων

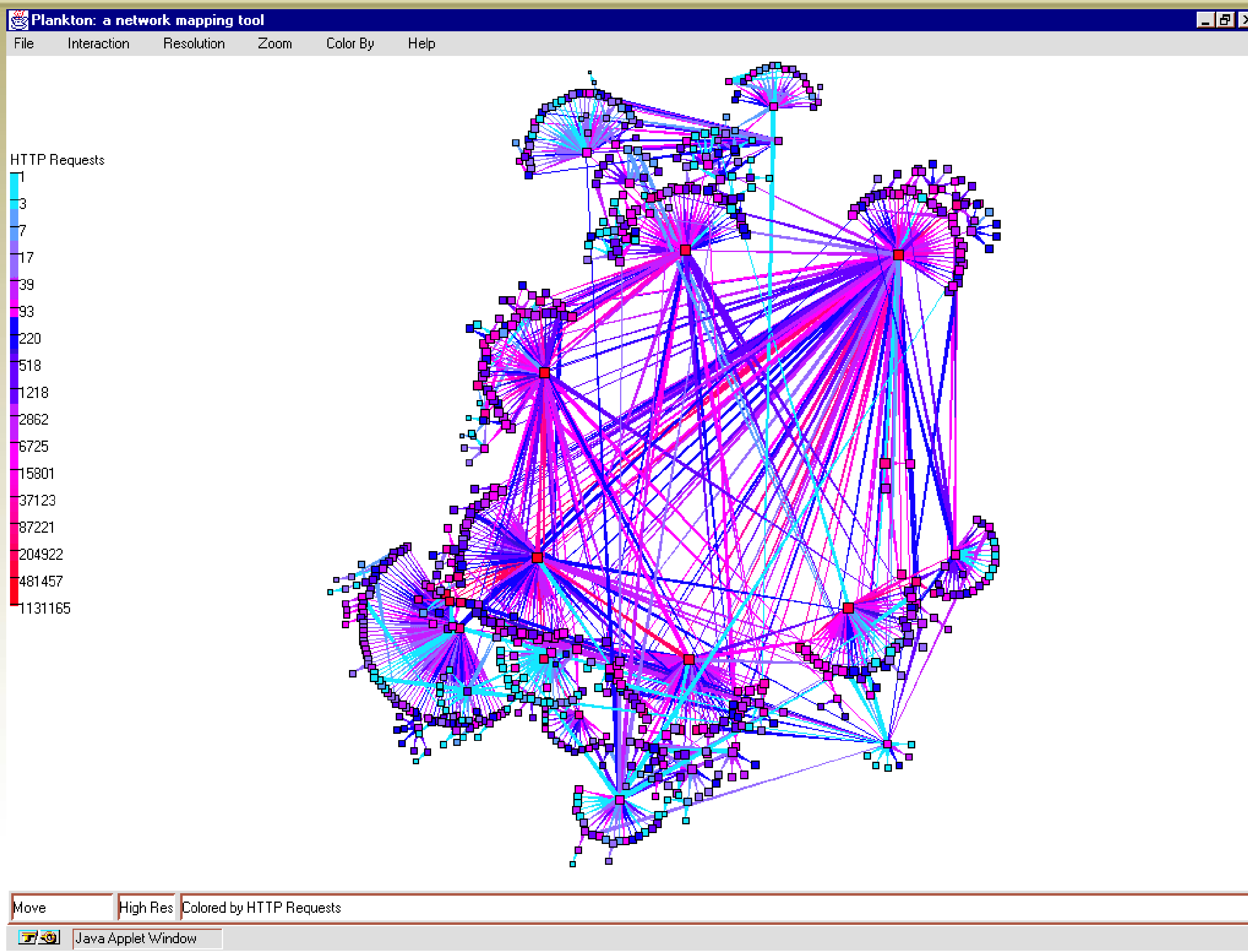
Τα “scale-free” δίκτυα (άνευ κλίμακας)

- Γιατί τα δίκτυα των οποίων οι συνδέσεις υπακούουν σε κατανομή δυναμο-νόμου αποκαλούνται **άνευ κλίμακας** (“**scale free**”)?
  - Σε όποια κλίμακα και εάν τα παρατηρήσουμε …
  - Το δίκτυο μοιάζει το ίδιο, δηλ., μοιάζει ίδιο με τον εαυτό του
- Όλες οι ιδιότητες του δικτύου διατηρούνται ανεξάρτητα της κλίμακας
- Ειδικότερα:
  - Εάν “κόψουμε” κάποιες λεπτομέρειες του δικτύου – διαγράφοντας του κόμβους με λίγες συνδέσεις – το δίκτυο θα διατηρήσει τη δομή του ψς δυναμο-νόμος
  - Εάν θεωρήσουμε ένα υπο-τμήμα του δικτύου, θα έχει την ίδια δομή με το συνολικό δίκτυο



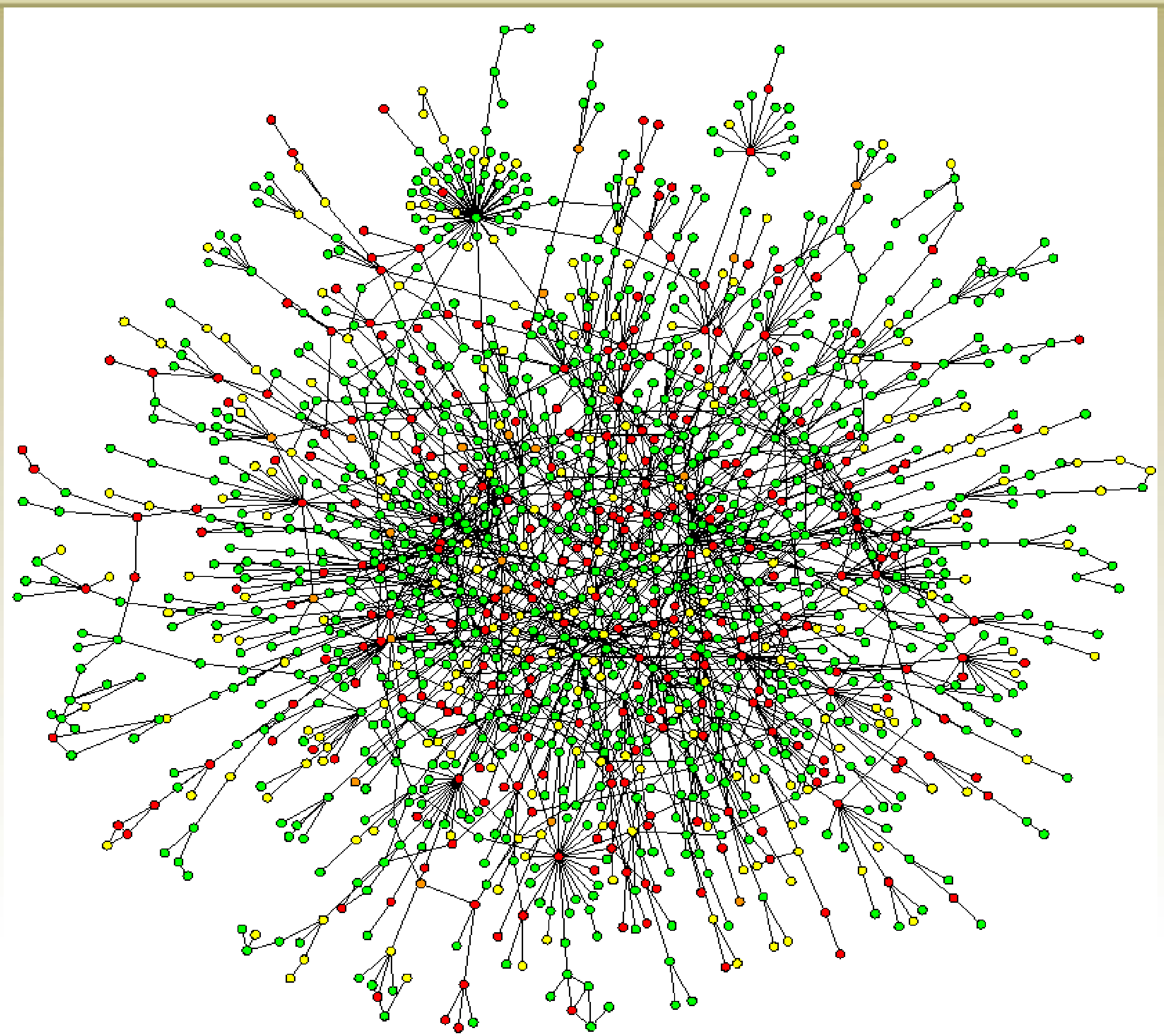
Πώς μοιάζουν τα scale-free δίκτυα;

Δίκτυο caches του Παγκοσμίου Ιστού



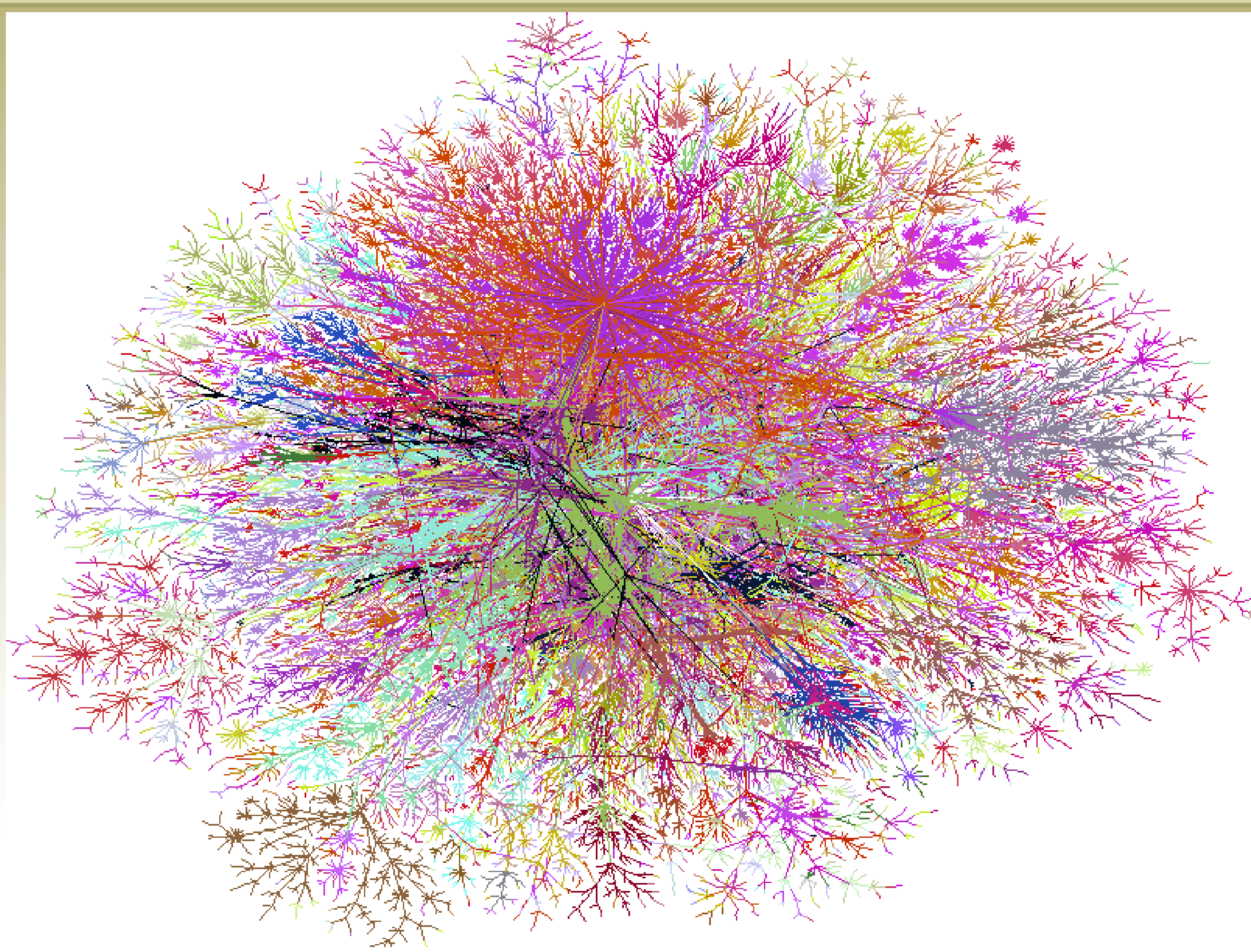
Πώς μοιάζουν τα scale-free δίκτυα;

Δίκτυο
Πρωτεϊνών



Πώς μοιάζουν τα scale-free δίκτυα;

Οι δρομολογητές του Διαδικτύου

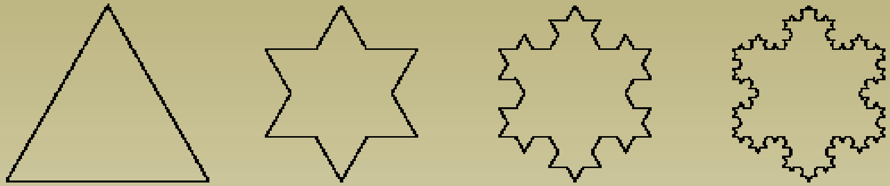


Fractals και scale-free δίκτυα

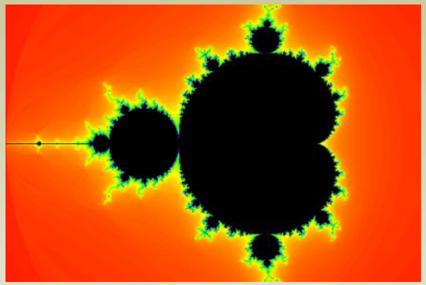
- Η φύση βρίθει “fractal αντικειμένων”
- Ο όρος fractal (αυτο-όμοιος) πηγάζει από το γεγονός ότι δεν έχουν ακέραια διάσταση
  - Τα 2-d αντικείμενα έχουν κάποιο μέγεθος “size” (δηλ., μια επιφάνεια) η οποία κλιμακώνεται με το τετράγωνο του γραμμικού μεγέθους $A=kL^2$
  - Τα 3-d αντικείμενα έχουν κάποιο μέγεθος “size” (δηλ., κάποιον όγκο) ο οποίος κλιμακώνεται με τον κύβο του γραμμικού μεγέθους $V=kL^3$
  - Τα fractal αντικείμενα έχουν κάποιο μέγεθος “size” το οποίο κλιμακώνεται με κάποιο κλάσμα/τα του γραμμικού μεγέθους $S=kL^{a/b}$
- Τα fractal αντικείμενα έχουν την ιδιότητα ότι είναι “self-similar” ή “scale-free”
  - Η “εμφάνισή” τους είναι ανεξάρτητη από την κλίμακα της παρατήρησης
  - Είναι όμοια προς τον εαυτό τους ανεξάρτητα από το εάν τα παρατηρούμε από κοντά ή μακριά
  - Δηλαδή, είναι άνευ κλίμακας (scale-free)

Παραδείγματα fractals

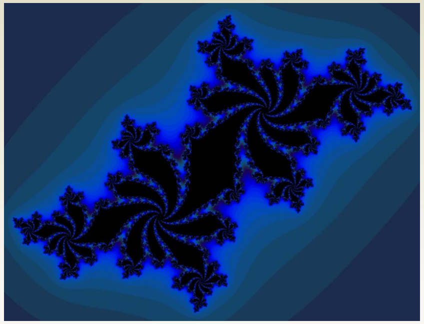
- Koch snowflake



- Mandelbrot set



- Julia set



- Τα δίκτυα άνευ κλίμακας είναι fractals?
  - ❖ Ναι, στην πραγματικότητα:
    - ❑ Είναι τα ίδια, σε όποια διάσταση και αν τα παρατηρούμε
    - ❑ Επίσης, το γεγονός ότι μεγαλώνουν σύμφωνα με κάποιον δυναμο-νόμο μπορούν να θεωρηθούν ως μια μορφή fractal διάστασης του δικτύου …

Δίκτυα μικρού κόσμου

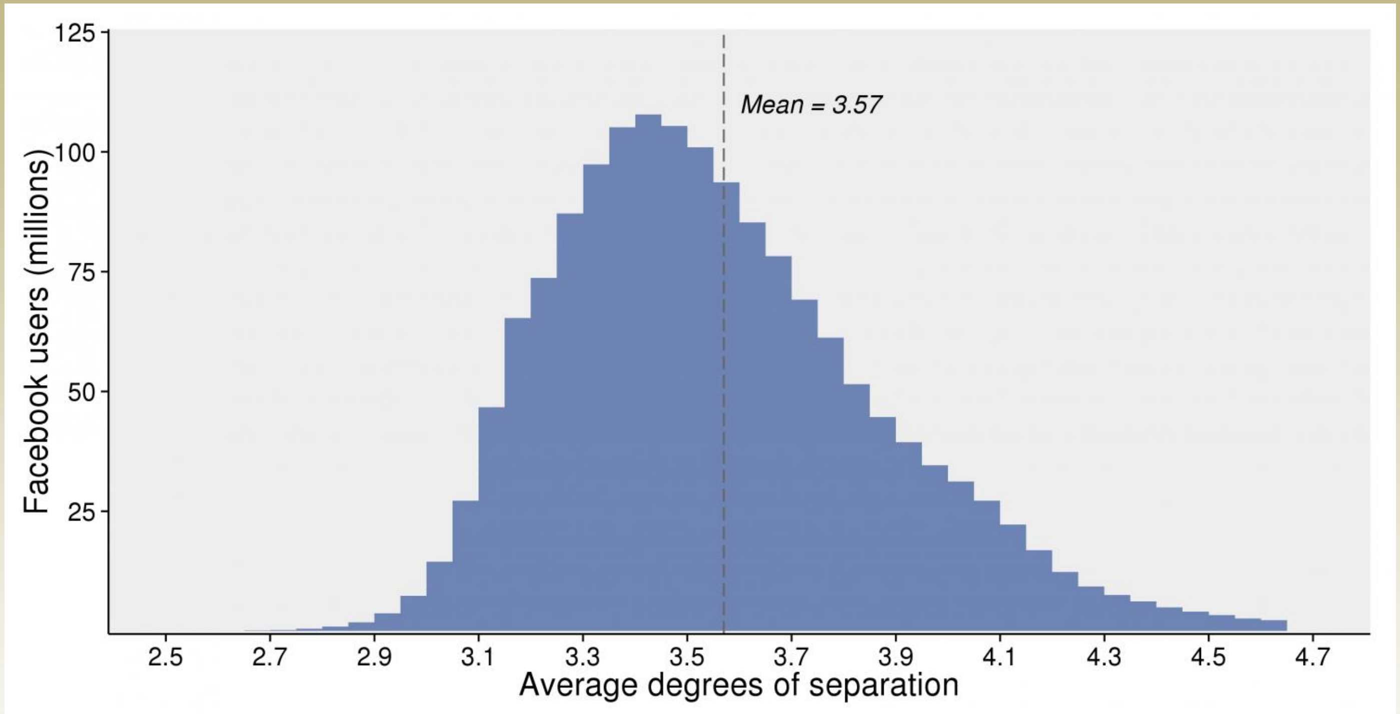
Small-world networks

Το πείραμα του Milgram

- Εντόπισε μικρές (short) αλυσίδες γνωριμιών μεταξύ ζευγών ανθρώπων στις Η.Π.Α., συνδέοντας ανθρώπους που ο ένας δεν γνώριζε τον άλλο;
  - Πρόσωπο-πηγή στη Nebraska
  - Πρόσωπο-προορισμός στην Massachusetts.
  - Αποστολή μηνύματος με προώθηση προς ανθρώπους που τους γνώριζαν προσωπικά (που εικάζανε ότι ήταν “κοντινοί ”στο πρόσωπο-προορισμός
- Μέσο μήκος αλυσίδας (για όσες τερματίστηκαν) ήταν μεταξύ 5 και 6 βήματα
- Η αρχή “***six degrees of separation***”

In Facebook …

- http://techxplore.com/news/2016-02-facebook-blog-degrees.html



- Το 2011, ερευνητές από το Cornell, το Università degli Studi di Milano, και το Facebook υπολόγισαν την μέση απόσταση για 721 εκατομμύρια χρήστες του Facebook, και την βρήκαν ίση με 3.74
- Το 2016, με διπλάσιο αριθμό χρηστών, το δίκτυο είναι καλύτερα διασυνδεδεμένο, και η απόσταση μίκρυνε. Βρέθηκε ίση με 3.57

Η ορθή ερώτηση

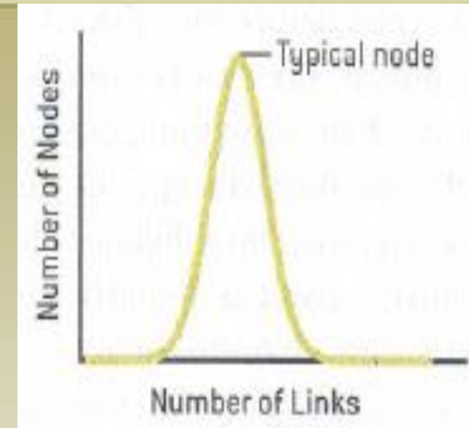
- ΓΙΑΤΙ υπάρχουν μικρές αλυσίδες γνωριμιών συνδέοντας τυχαία ζεύγη αγνώστων μεταξύ τους προσώπων;;;
  - ή, καλύτερα
- Γιατί μας εκπλήσσει αυτή η παρατήρηση;

Τυχαία (random) δίκτυα

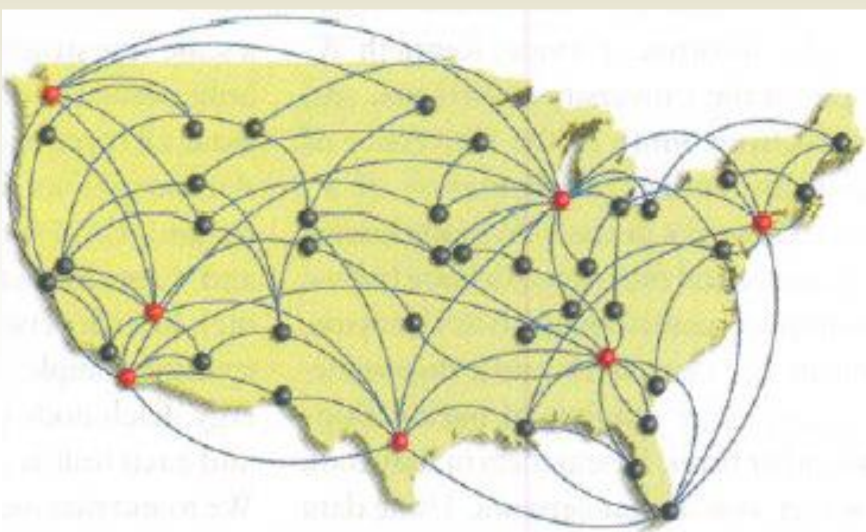
- Σε ένα τυχαίο (*random*) δίκτυο, εάν καθένας μας έχει 100 φίλους κατανεμημένους τυχαία στον παγκόσμιο πληθυσμό, αυτό δεν είναι παράξενο
- Σε 6 hops, θα φτάναμε σε 100^6 ($=10^{12}$) ανθρώπους - ένα million million >> 6,000 million (παγκόσμιος πληθυσμός)

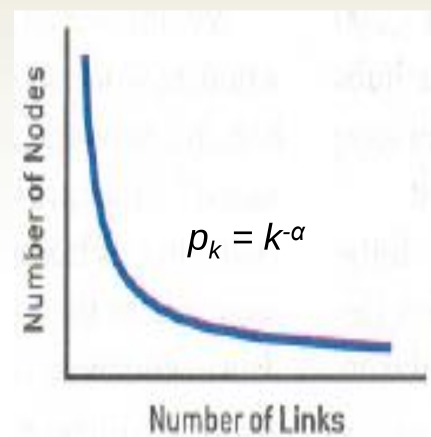
- ΑΛΛΑ: τα κοινωνικά μας δίκτυα τείνουν να είναι ομαδοποιημένα (clustered)

Τυχαία vs. δίκτυα δυναμο-νόμου



$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k} = \frac{z^k e^{-z}}{k!}$$





Τα κοινωνικά δίκτυα

- Όχι τυχαία
- Αλλά **clustered**
- Οι περισσότεροι των φίλων μας προέρχονται από την γεωγραφική ή επαγγελματική μας γειτονιά
- *Οι φίλοι μας τείνουν να έχουν τους ίδιους φίλους*

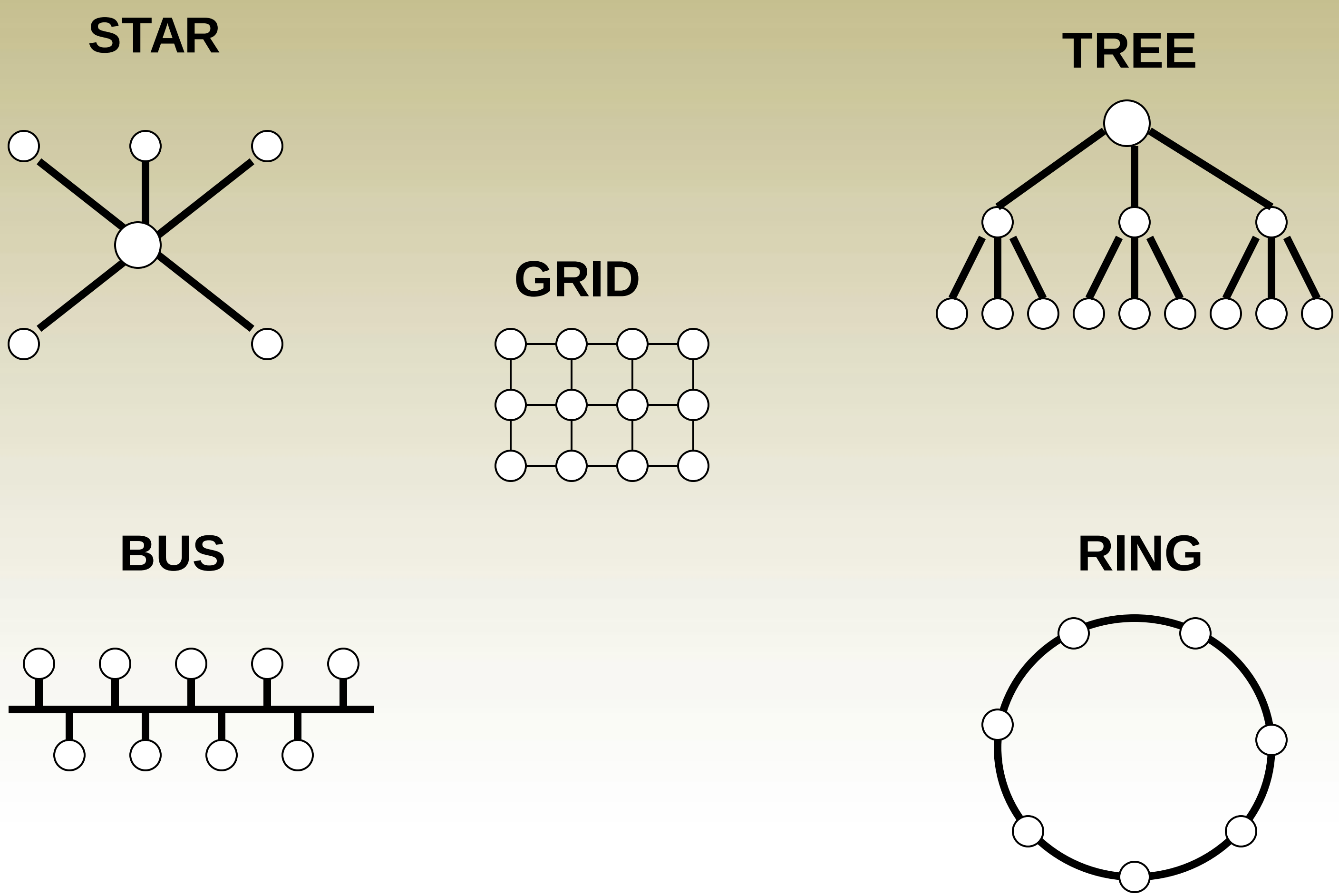
ΑΛΛΑ

- Παρά το ότι τα κοινωνικά μας δίκτυα τείνουν να είναι clustered, φαίνεται ότι υπάρχουν σύντομα (μικρά, short) μονοπάτια μεταξύ τυχαίων κόμβων

Παράμετροι ενός δικτύου

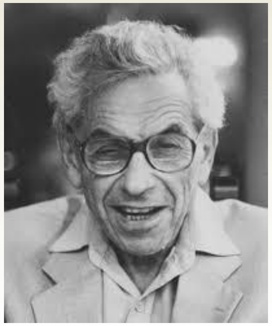
- Τύποι δικτύων
  - Κανονικά (Regular)
  - Τυχαία (Random)
  - Πραγματικά (Natural)
- Μέγεθος: # κόμβων
- Αριθμός συνδέσεων:
  - Μέρος όρος & κατανομή
- Επιλογή των γειτόνων

Τοπολογίες κανονικών (regular) δικτύων



Συνδεσιμότητα σε τυχαία δίκτυα

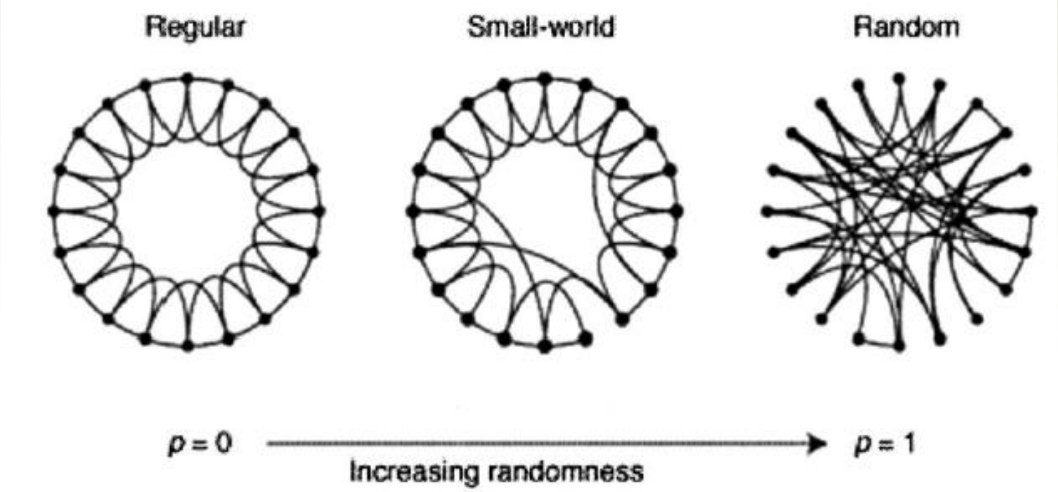
- Οι κόμβοι συνδέονται με συνδέσεις με εντελώς τυχαίο τρόπο
- Πόσο μεγάλη είναι η μεγαλύτερη συνιστώσα(component)? (ως ποσοστό του αριθμού όλων των κόμβων)
  - Εξαρτάται από τον αριθμό των συνδέσεων ανά κόμβο



- (Erdös, Rényi 1959)

Small-World δίκτυα

- Τυχαία ανα-καλωδίωση (random rewiring) του κανονικού γραφήματος (είναι το small-world μοντέλο των Watts and Strogatz)
  - Με πιθανότητα *p (ή β)* rewire κάθε σύνδεσμο του regular graph σε έναν τυχαία επιλεγμένο κόμβο
  - Το προκύπτον γράφημα έχει ιδιότητες και των regular και των random γραφημάτων
    - Υψηλό clustering και μικρό path length

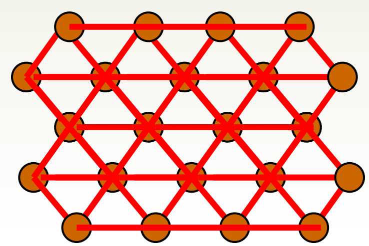


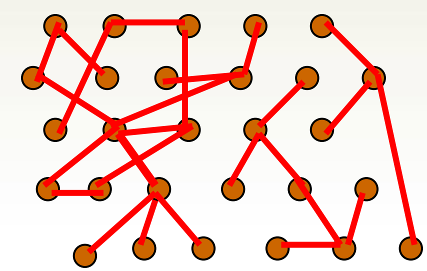
Μετρικές δικτύου

- Η **Συνδεσιμότητα (connectivity)** δεν είναι η κύρια μετρική
- **Characteristic Path Length** **(L)** :
  - Το μέσο μήκος από το σύνολο των συντομοτέρων μονοπατιών που συνδέει κάθε ζεύγος κόμβων
- **Clustering Coefficient** **(C)** είναι μια μετρική της τοπικής αλληλοσύνδεσης
  - Εάν ο κόμβος i έχει k_i άμεσους γείτονες, C_i, είναι το κλάσμα των συνολικών δυνατών $k_i*(k_i-1) / 2$ συνδέσεων που μπορεί να υπάρξουν μεταξύ των γειτόνων του i. Το C είναι απλά ο μέσος όρος των C_i.
- **Διάμετρος**: το μήκος του μακρύτερου από τα συντομότερα μονοπάτια

Κανονικά vs. Τυχαία δίκτυα

	Regular	Random
Μ.Ο. συνδέσεων ανά κόμβο	**μερικές, clustered**	**λιγότερες, spread**
Αριθμός συνδέσεων που απαιτούνται για πλήρη συνδεσμολογία	**πολλές**	**λιγότερες (<2/3)**
Διάμετρος	**μεγάλη**	**μέτρια**





Clustering

- Το clustering μετρά το ποσοστό των γειτόνων ενός κόμβου που είναι συνδεδεμένοι μεταξύ τους
- Τα regular γραφήματα έχουν υψηλό clustering coefficient
  - Αλλά και μεγάλη διάμετρο
- Τα random γραφήματα έχουν χαμηλό clustering coefficient
  - Αλλά μικρή διάμετρο
- Και τα δυο μοντέλα ικανοποιούν τις ιδιότητες που αναμένονται από τα πραγματικά δίκτυα!

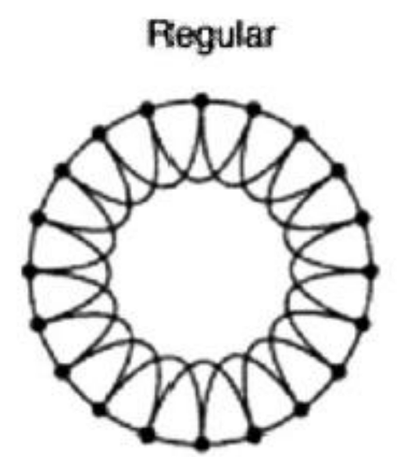
Regular Graph (k=4)

Μακριά μονοπάτια

- $L \sim n/(2k)$

Υψηλά clustered

- $C \sim 3/4$



Random Graph (k=4)

Μικρό μήκος μονοπατιού

- $L \sim \log_k N$

Σχεδόν χωρίς clustering

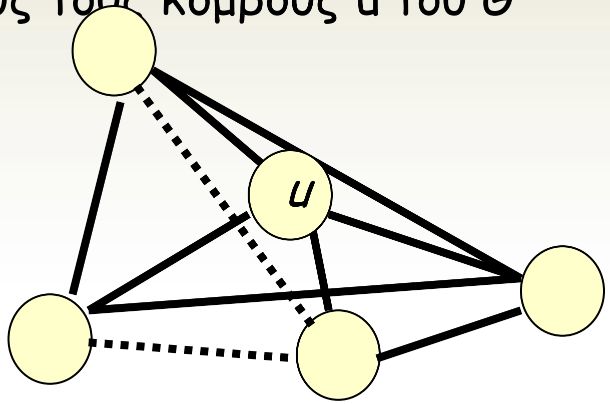
- $C \sim k/n$

Το Base δίκτυο είναι ο κύκλος

O clustering coefficient ενός δικτύου

- Διαίσθηση: μια μετρική του κατά πόσο οι ακμές είναι "bunched up"
- Ο συντελεστής ομαδοποίησης (clustering coefficient) ενός κόμβου u:
  - Έστω k = βαθμός του u = αριθμός γειτόνων του u
  - k(k-1)/2 = *max possible # ακμών μεταξύ των γειτόνων του* u
  - c(u) = (*πραγματικός* # ακμών μεταξύ των γειτόνων του u)/[k(k-1)/2]
  - Ποσοστό ζευγών γειτόνων που είναι επίσης "φίλοι"
  - 0 <= c(u) <= 1; Μετρική της *cliquishness* της γειτονιάς του u
- <u>Clustering coefficient ενός γραφήματος *G*</u>:
  - *CC(G)* = μέσος όρος των c(u) πάνω σε όλους τους κόμβους u του *G*

k = 4
k(k-1)/2 = 6
c(u) = 4/6 = 0.666…



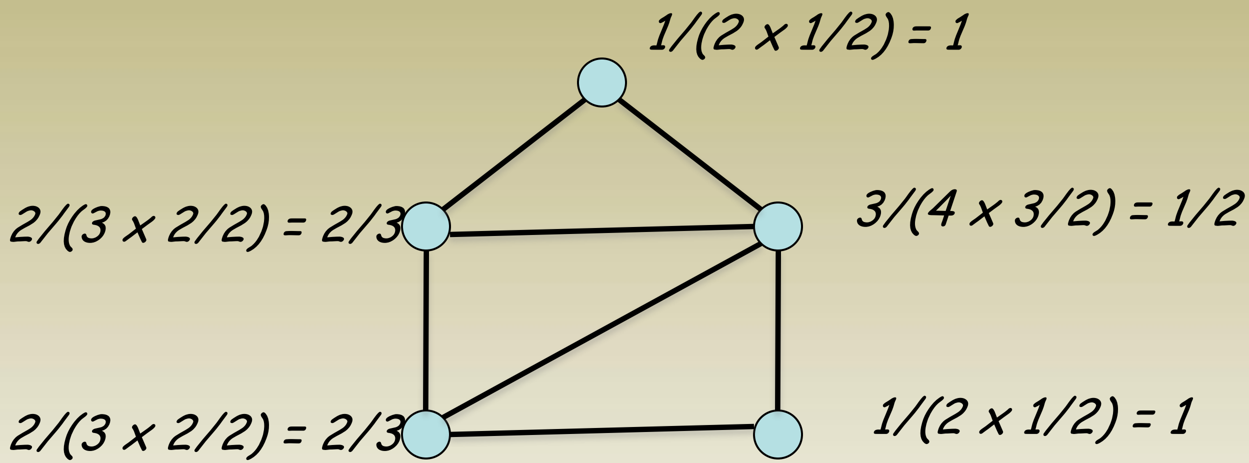
Τι εννοούμε με τον όρο “high clustering”

- CC(G) μετρά πόσο πιθανόν είναι οι κορυφές που έχουν έναν κοινό γείτονα να είναι και γείτονες μεταξύ τους
- Πρέπει να συγκρίνεται με το πόσο πιθανό είναι να έχουμε τυχαία ζεύγη κόμβων που είναι γείτονες μεταξύ τους
- Έστω p η πυκνότητα ακμών ενός γραφήματος G:

$$p = E/(N(N-1)/2)$$

- Εδώ E = συνολικός αριθμός ακμών του G
- Εάν επιλέγαμε στην τύχη ένα ζεύγος κόμβων του G, τότε η πιθανότητα ότι είναι συνδεδεμένοι είναι ακριβώς p
- Έτσι θα λέμε ότι η ομαδοποίηση (clustering) είναι υψηλή εάν CC(G) >> p

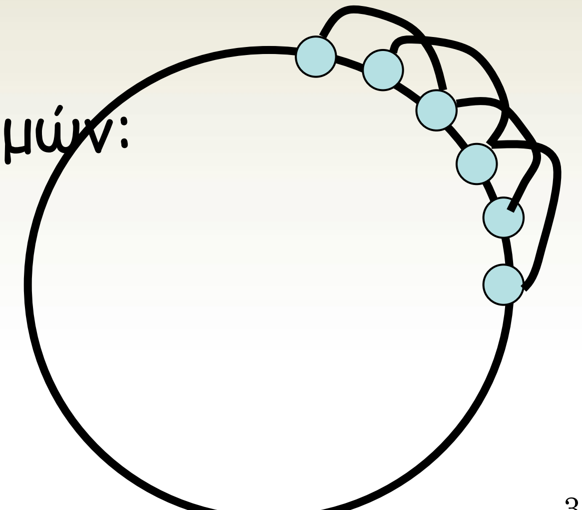
Clustering coefficient: Παράδειγμα 1



C.C. = $(1 + \frac{1}{2} + 1 + 2/3 + 2/3)/5 = 0.7666\ldots$
$p = 7/(5 \times 4/2) = 0.7$
Όχι υψηλά ομαδοποιημένο

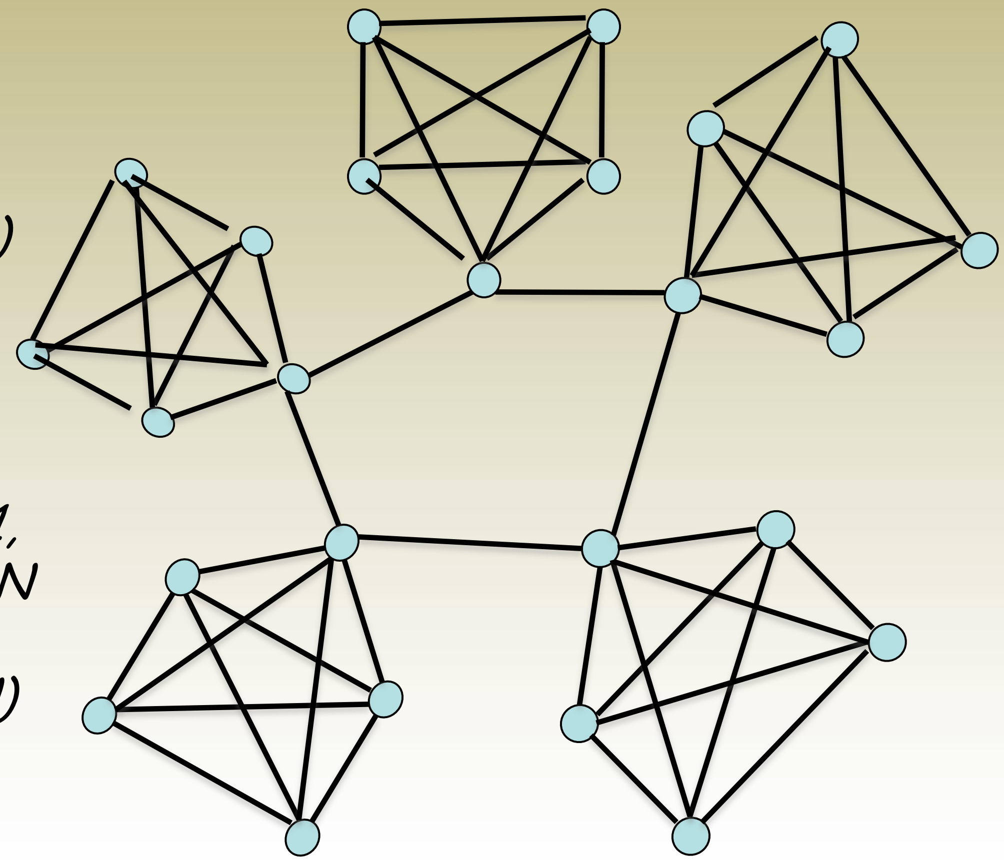
Clustering coefficient: Παράδειγμα 2

- Δίκτυο: απλός κύκλος + ακμές σε κόμβους 2 hops μακριά στον κύκλο
- Λόγω συμμετρίας, όλοι οι κόμβοι έχουν ίσο clustering coefficient
- Ο clustering coefficient ενός κόμβου v:
  - Βαθμός της v είναι 4, έτσι ο αριθμός των *δυνατών ακμών* μεταξύ ζευγών γειτόνων του v είναι 4 x 3/2 = 6
  - Πόσα ζεύγη γειτόνων του v είναι πραγματικά *συνδεδεμένα;* 3 --- οι δυο κατά τη φορά των δεικτών του ρολογιού, οι δυο αντίστροφα, και οι άμεσοι "κυκλικοί" γείτονες
  - Έτσι ο c.c. του v είναι 3/6 = ½
- Συγκρίνετε με τη συνολική πυκνότητα ακμών:
  - Συνολικός αριθμός ακμών = 2N
  - Πυκνότητα ακμής p = 2N/(N(N-1)/2) ~ 4/N
  - Καθώς το N γίνεται μεγάλο, ½ >> 4/N
  - Συνεπώς αυτό το κυκλικό γράφημα δεν είναι υψηλά ομαδοποιημένο



Clustering coefficient: Παράδειγμα 3

Διαιρέστε τους N κόμβους σε sqrt(N) ομάδες μεγέθους sqrt(N) (εδώ N = 25) η κάθε μιά
Προσθέστε όλες τις συνδέσεις μέσα σε κάθε ομάδα (cliques), συνδέστε τους "leaders" σε έναν κύκλο
N – sqrt(N) non-leaders έχουν C.C. = 1, έτσι το δίκτυο έχει C.C. → 1 καθώς το N γίνεται μεγάλο
Η πυκνότητα ακμών είναι p ~ 1/sqrt(N)



Συντελεστής ομαδοποίησης

- Ο συντελεστής ομαδοποίησης (clustering coefficient) ενός κόμβου u τον οποίο ορίσαμε νωρίτερα ονομάζεται και *τοπικός συντελεστής ομαδοποίησης* (local clustering coefficient)

- Clustering Coefficient για ολόκληρο ο δίκτυο
  - Είδαμε ήδη τον το **average των cc κάθε κόμβου**
  - ΑΛΛΑ, εξίσου δημοφιλές είναι και το **network transitivity τ**:

$$\tau = \frac{3 \times \text{ number of triangles}}{\text{number of connected triplets}}$$

Τι συμβαίνει με τα κοινωνικά δίκτυα;

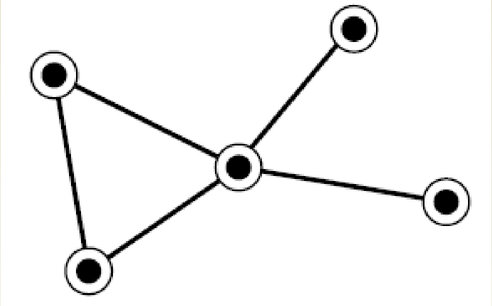
Θα μπορούσε ένα δίκτυο που είναι τόσο ισχυρά ομαδοποιημένο σε τοπικό επίπεδο να είναι ταυτόχρονα και small world?

ΝΑΙ. Δεν χρειαζόμαστε παρά μερικές τυχαίες συνδέσεις

Ασκήσεις

- **Άσκηση 1.**

  Για το δίκτυο παρακάτω βρείτε τον local clustering coefficient κάθε κόμβου και το μέσο όρο τους, καθώς και την network transitivity και συγκρίνετε.

  Λύση. avg(cc)=0.433, netTran=0.375

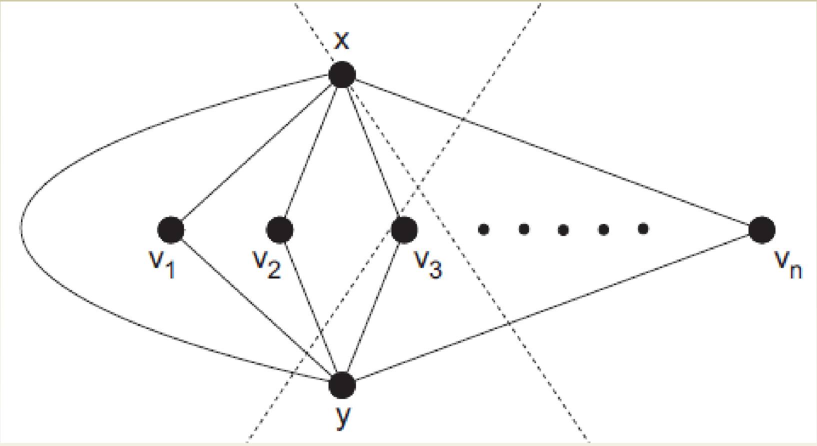
- **Άσκηση 2.**

  Υπολογίστε τον clustering coefficient του δικτύου (δηλ., τον average clustering coefficient κάθε κόμβου) ενός *(N,m) circle graph*, δηλαδή δικτύου αποτελούμενο από *N* κόμβους που καθένας τους έχει *m* 1-hop γείτονες εκατέρωθέν του.

  Αυτό το δίκτυο συμβολίζεται και ως WS(n,m,p)=WS(n,m,0).

Ασκήσεις

- **Άσκηση 3.**
  Υπολογίστε τον clustering coefficient του δικτύου G_k που εικονίζεται παρακάτω, καθώς και το όριό του όταν $k \to \infty$.



- **Απάντηση 2**. $cc = \dfrac{3 \times (m-1)}{2 \times (2 \times m - 1)}$ Εάν k=2m: $cc = \dfrac{3 \times (k-2)}{4 \times (k-1)}$
- **Απάντηση 3.** $cc = \dfrac{1}{k+2} \times \left(2 \times \dfrac{2}{k+1} + k \times 1\right)$

Ασκήσεις

- **Άσκηση 4.**

Βρείτε την αναλυτική σχέση μεταξύ average (localized) clustering coefficient και network transitivity ενός δικτύου.

Λύση.

Εύκολα διαπιστώνεται ότι ο αριθμός των ακμών μεταξύ των 1-hop γειτόνων ενός κόμβου i ταυτίζεται με τον αριθμό των τριγώνων (n_i^{Δ}) που έχουν ως κορυφή το i.

Ο αριθμός των τριπλετών (n_i^{Λ}) στις οποίες συμμετέχει ο i (δηλ., βρίσκεται πάνω στην γωνία) στο μικρό δίκτυο που περιλαμβάνει τους 1-hop γείτονες του κόμβου i και τις συνδέσεις είναι: $k_i(k_i-1)/2$, όπου k_i είναι ο βαθμός του κόμβου.

Συνεπώς ο clustering coefficient του i μπορεί να οριστεί ως: $\mathbf{cc_i = n_i^{\Delta}/n_i^{\Lambda}}$

Από τον ορισμό του network transitivity έχουμε: $\boldsymbol{\tau = 3n^{\Delta}/n^{\Lambda}}$

Όμως ισχύουν τα εξής: $\mathbf{n^{\Delta} = \Sigma n_i^{\Delta}/3}$ (επειδή κάθε τρίγωνο προσμετράται 3 φορές), δηλ.,

$\mathbf{n^{\Delta} = (1/3)\Sigma n_i^{\Delta} = (1/3)\Sigma n_i^{\Lambda}\, cc_i}$ και $\mathbf{n^{\Lambda} = \Sigma n_i^{\Lambda}}$

Συνεπώς: $$\tau = \frac{\sum_i n_i^{\Lambda} cc_i}{\sum_i n_i^{\Lambda}}$$

Δηλαδή, ο cc είναι ο **αριθμητικός μέσος** των cc_i, ενώ ο τ είναι ο **βεβαρημένος** μέσος των cc_i, όπου το βάρος για κάθε κόμβο είναι ίσο με τον αριθμό των τριπλετών με κορυφή τον κόμβο αυτόν.

Ασκήσεις

- **Άσκηση 5.**

Εκκινώντας από τον ορισμό μιας scale free συνάρτησης, δηλ., $p(bx) = g(b)*p(x)$ δείξτε ότι η $p(x)$ υπακούει σε power-law.

Λύση.

Για $x=1$, έχουμε: $p(b)=g(b)p(1) \Rightarrow g(b)=p(b)/p(1)$.

Άρα, $p(bx)= p(b)*p(x)/p(1)$

Αφού αληθεύει για κάθε b, παραγωγίζοντας ως προς b, έχω:

$x*p'(bx)=[p'(b)*p(x)]/p(1)$, όπου p' η παράγωγος ως προς το αντίστοιχο όρισμα

Θέτοντας $b=1$, έχουμε:

$x*[dp/dx] = [p'(1)/p(1)] * p(x)$,

First-order, first degree (not exact) ordinary D.E. Use of *integrating factor* $1/(x*p(x))$ to turn it into Variables Separable D.E.

Η λύσης της είναι: $\ln[p(x)] = [p'(1)/p(1)]*\ln(x) + \text{constant}$

Θέτοντας $x=1$, βρίσκουμε την constant ίση με $\ln[p(1)]$ και συνεπώς:

$p(x) = p(1)\, x^{-\alpha}$, όπου $\alpha= -p'(1)/p(1)$

Ασκήσεις

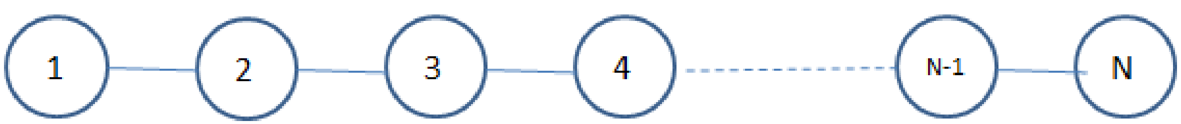
- **Άσκηση 6.**

  Να δείξετε ότι η αναμενόμενη τιμή του βαθμού των κόμβων ενός Erdos-Renyi δικτύου ER(n,p), δηλ., αποτελούμενο από n κόμβους και με πυκνότητα συνδεσμολογίας p, είναι p(n-1).

Άσκηση 7.

Να βρείτε το Characteristic Path Length (CPL) για ένα (μη κατευθυνόμενο, χωρίς βάρη στις ακμές) “δίκτυο γραμμή” (linear graph), το οποίο αποτελείται από N κόμβους

[Υπόδειξη. Βρείτε το CPL για μικρά δίκτυα, π.χ., με 2, 3, και 4 κόμβους, ώστε να αντιληφθείτε τον μαθηματικό τύπο που το υπολογίζει, και κατόπιν αποδείξτε την ισχύ του τύπου αυτού με μαθηματική επαγωγή.]



Ασκήσεις

- **Άσκηση 8.**

  Να δείξετε ότι ο clustering coefficient οποιουδήποτε $ER(n,p)$ δικτύου είναι ίσος με p.

- **Άσκηση 9.**

  Να δείξετε ότι σε ένα Watts-Strogatz δίκτυο $WS(n,k,0)$ (k άρτιο, $k/2$ γείτονες αριστερά και $k/2$ γείτονες δεξιά) το μέσο μήκος των συντομοτέρων μονοπατιών avg.d(u) από μια δεδομένη κορυφή u προς οποιαδήποτε άλλη κορυφή στο δίκτυο προσεγγίζεται από τον κάτωθι τύπο:

$$avg.d(u) \approx \frac{(n-1)(n+k-1)}{2kn}$$

Λύση Άσκησης 6

$$ER(n,p) = \sum_{k=1}^{n-1} k \times P(\delta = k) = \sum_{k=1}^{n-1} \binom{n-1}{k} \times k \times p^k \times (1-p)^{n-1-k} =$$

$$= \sum_{k=1}^{n-1} \frac{(n-1)!}{k!(n-1-k)!} \times k \times p^k \times (1-p)^{n-1-k} =$$

$$= \sum_{k=1}^{n-1} \frac{n-1}{k} \times \frac{(n-2)!}{(k-1)!(n-1-k)!} \times k \times p \times p^{k-1} \times (1-p)^{n-1-k} =$$

$$= p(n-1) \times \sum_{k=1}^{n-1} \frac{(n-2)!}{(k-1)!(n-1-k)!} \times p^{k-1} \times (1-p)^{n-1-k} =$$

$$= p(n-1) \times \sum_{l=0}^{n-2} \frac{(n-2)!}{l!(n-l-2)!} \times p^l \times (1-p)^{n-l-2} =$$

$$= p(n-1) \times \sum_{l=0}^{n-2} \binom{n-2}{l} \times p^l \times (1-p)^{n-l-2} =$$

$$= p(n-1) \times \sum_{l=0}^{m} \binom{m}{l} \times p^l \times (1-p)^{m-l} = p(n-1) \times 1 \implies$$

$$ER(n,p) = p(n-1)$$

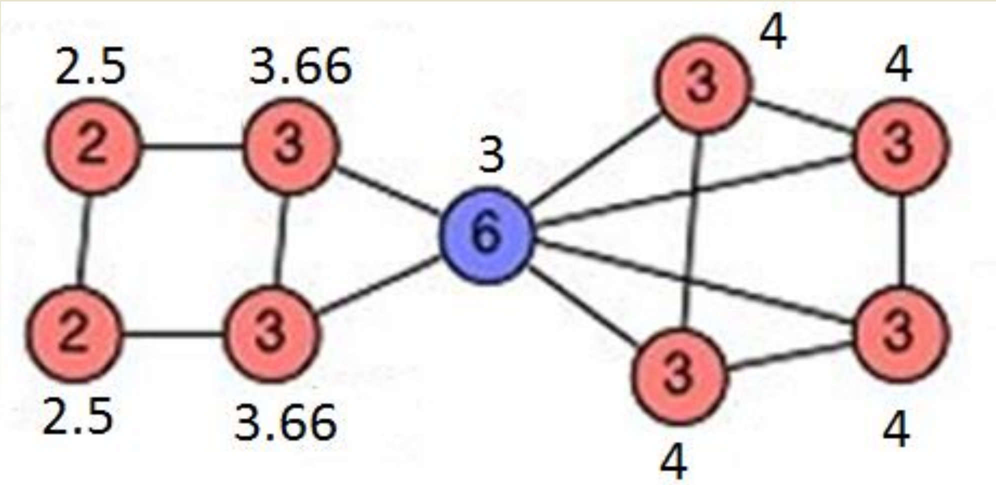
Ασκήσεις: The Friendship paradox

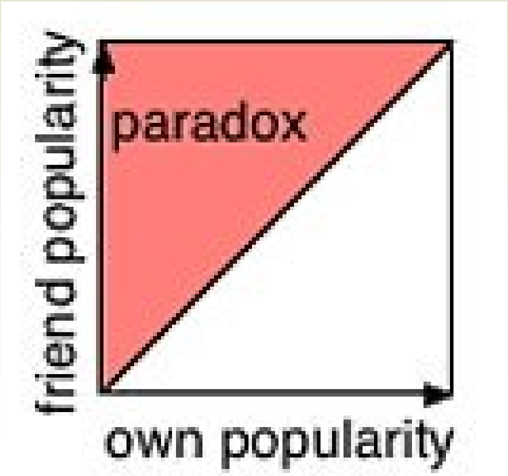
- **Άσκηση 8.**

  Αποδείξτε ότι οι φίλοι σας (δηλ., 1-hop γείτονές σας) έχουν περισσότερους φίλους από εσάς κατά την μέση τιμή, δηλαδή αυτοί είναι πιο δημοφιλείς από όσο είστε εσείς!

  Αυτό είναι γνωστόν ως **Friendship paradox**.

  Λύση.





Ασκήσεις: The Friendship paradox

Πρέπει να υπολογίσουμε δυο ποσότητες:

- Τον μέσο αριθμό φίλων στο συνολικό γράφημα
- Τον μέσο αριθμό φίλων ενός τυχαίου κόμβου του γραφήματος

Συμβολισμοί:

- Ένας κόμβος i έχει έναν αριθμό x_i φίλων
- Υπάρχουν συνολικά n κόμβοι

Ασκήσεις: The Friendship paradox

- Ποιος είναι ο μέσος αριθμός φίλων στο συνολικό γράφημα;
  - Απλά, προσθέτουμε τον συνολικό αριθμό φίλων, και κατόπιν διαιρούμε με τον συνολικό αριθμό κόμβων:

$$\mu = \frac{\sum x_i}{n}$$

- Ποιος είναι ο μέσος αριθμός φίλων των φίλων;
  - Τρικ: Για έναν συγκεκριμένο κόμβο i, πόσες φορές θα εμφανιστεί ο όρος x_i στην τελική άθροιση;
  - Η μόνη φορά που απαιτείται να μετρήσουμε τους φίλους του κόμβου i είναι για τοους φίλους του κόμβου i
  - Αυτό σημαίνει ότι οι φίλοι του κόμβου i (υπάρχουν x_i τέτοιοι φίλοι) θα συνεισφέρουν τον όρο x_i στο τελικό άθροισμα. Επομένως, η τελική άθροιση έχει τον όρο (x_i) x $(x_i) = x_i^2$
  - Επομένως: μέσος αριθμός των φίλων των φίλων= $\dfrac{\sum (x_i)^2}{\sum x_i}$

Ασκήσεις: The Friendship paradox

- Θυμηθείτε την variance μιας τυχαίας μεταβλητής:

$$\sigma^2 = \frac{\sum (x_i)^2}{n} - \mu^2$$

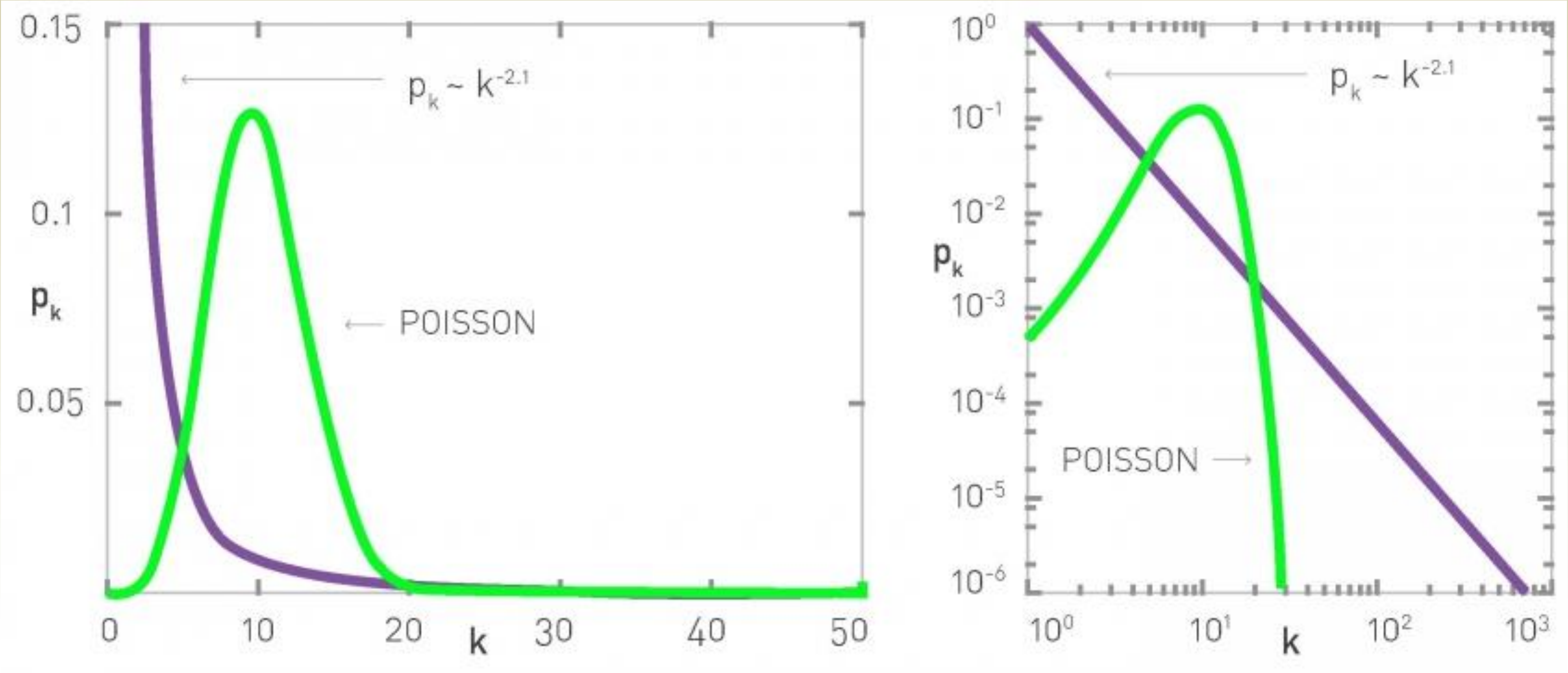
- Άρα: $$\sum (x_i)^2 = (\mu^2 + \sigma^2)n$$

- Επομένως: $$\frac{\sum (x_i)^2}{\sum x_i} = \frac{n(\mu^2 + \sigma^2)}{n\mu} = \mu + \frac{\sigma^2}{\mu}$$

- Δηλαδή: ο μέσος αριθμός φίλων των φίλων σας είναι ίσος με τον μέσο αριθμό φίλων σας συν κάποιον θετικό όρο!

Ασκήσεις: The Friendship paradox

Στα social networks, η κατανομή των βαθμών ακολουθεί κάποιον δυναμονόμο, δηλ., υπάρχουν πλήμνες (hub nodes) οι οποίες αυξάνουν την variance



Από το βιβλίο του Barabasi

Κατασκευή ER(n,p) δικτύου

1. Ξεκινούμε με n απομονωμένους κόμβους
2. Επιλέγουμε τυχαία ένα ζεύγος κόμβων, και παράγουμε έναν τυχαίο αριθμό U(0,1)
3. Εάν ο τυχαίος αριθμός ξεπερνά το p, τότε δημιουργούμε την ακμή μεταξύ του ζεύγους, αλλιώς τούς αφήνουμε ασύνδετους
4. Επαναλαμβάνουμε το Βήμα 2 για κάθε ένα από τα n(n-1)/2 ζεύγh κόμβων

Κατασκευή WS δικτύου

Definition 4.7 (The WS model) *Start with a (N, m) circle graph. A (N, m, p) WS graph is constructed by considering each of the links of the circle graph and independently rewiring each of them with some probability p $(0 \leq p \leq 1)$ as follows:*

1. *Visit each node along the ring one after the other, moving clockwise.*
2. *Let i be the current node. Each edge connecting node i to one of its m neighbours in a clockwise sense is rewired with a probability p, or left in place with a probability $1 - p$.*
3. *Rewiring means shifting the end of the edge, other than that in node i, to a new vertex chosen uniformly at random from the whole lattice, with the constraint than no two vertices can have more than one edge running between them, and no vertex can be connected by an edge to itself.*