Σύνθετα Δίκτυα

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

Διδάσκων –
Δημήτριος Κατσαρός

Διάλεξη 5η

Μεγέθυνση δικτύου με προνομιακή προσκόλληση

Network growth with preferential attachment

Κατανομή βαθμών για τυχαία δίκτυα

- Σε ένα τυχαίο γράφημα με πιθανότητα συνδέσεων ίση με p, ο βαθμός k_i ενός κόμβου i ακολουθεί διωνυμική κατανομή με παραμέτρους N-1 και p:

$$P(k_i = k) = C_{N-1}^{k} p^{k} (1-p)^{N-1-k}$$

- Αυτή η πιθανότητα αντιπροσωπεύει τον αριθμό τρόπων με τους οποίους k ακμές μπορούν να επιλεγούν από έναν συγκεκριμένο κόμβο: η πιθανότητα των k ακμών είναι p^k, η πιθανότητα της απουσίας επιπλέον ακμών είναι $(1-p)^{N-1-k}$, και υπάρχουν C_{N-1}^{k} ίδιοι τρόποι επιλογής των k άκρων αυτών των ακμών

Κατανομή βαθμών για τυχαία δίκτυα

- Επιπλέον, εάν i και j είναι διαφορετικοί κόμβοι, $P(k_i=k)$ και $P(k_j=k)$ είναι σχεδόν ανεξάρτητες τυχαίες μεταβλητές. Για να βρούμε την κατανομή βαθμών για το γράφημα, πρέπει να μελετήσουμε τον αριθμό των κόμβων με βαθμό k, δηλαδή X_k
- Ο κύριος στόχος είναι να προσδιορίσουμε την πιθανότητα ότι η X_k παίρνει μια δεδομένη τιμή, $P(X_k=r)$
- Σύμφωνα με την προηγούμενη σχέση, η expectation value του αριθμού των κόμβων με βαθμό k είναι

$E(X_k) = N \times P(k_i=k) = \lambda_k$

όπου $\lambda_k = N \times C^k_{N-1}\, p^k(1-p)^{N-1-k}$

Κατανομή βαθμών για τυχαία δίκτυα

- Η κατανομή Poisson φθίνει γρήγορα για μεγάλες τιμές του r, η τυπική απόκλιση της κατανομής είναι $\sigma_k = sqrt(\lambda_k)$
- Απλοποιώντας, θα μπορούσαμε να πούμε ότι $P(X_k=r)$ υπονοεί ότι η X_k δεν αποκλίνει πολύ από το προσεγγιστικό αποτέλεσμα $X_k=N \times P(k_i=k)$, που ισχύει μόνο εάν οι κόμβοι είναι ανεξάρτητοι
- Επομένως, προσεγγιστικά η κατανομή βαθμών ενός τυχαίου γραφήματος είναι μια διωνυμική κατανομή,

$$P(k)=C_{N-1}^{k}p^{k}(1-p)^{N-1-k}$$

που όμως για μεγάλα N πρέπει ν' αντικατασταθεί από μια Poisson

Κατανομή βαθμών για τυχαία δίκτυα

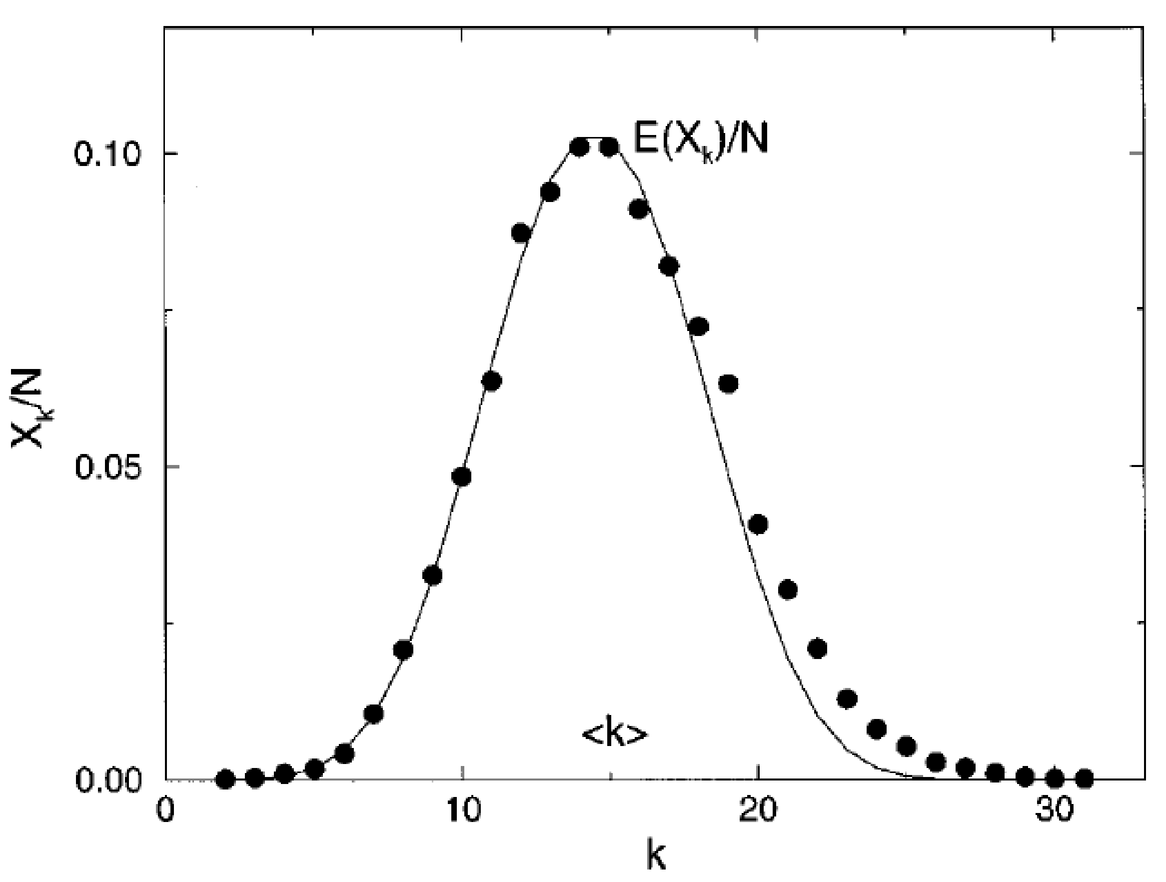
- Η κατανομή των τιμών του $P(X_k=r)$, πλησιάζει μια κατανομή Poisson

$$P(X_k=r)=e^{-\lambda_k}\frac{\lambda_k^r}{r!}$$

- Συνεπώς ο αριθμός των κόμβων με βαθμό k ακολουθεί κατανομή Poisson με μέση τιμή λ_k
- Να σημειωθεί ότι η expectation value της παραπάνω κατανομής είναι η συνάρτηση λ_k που δείξαμε νωρίτερα και δεν είναι σταθερά

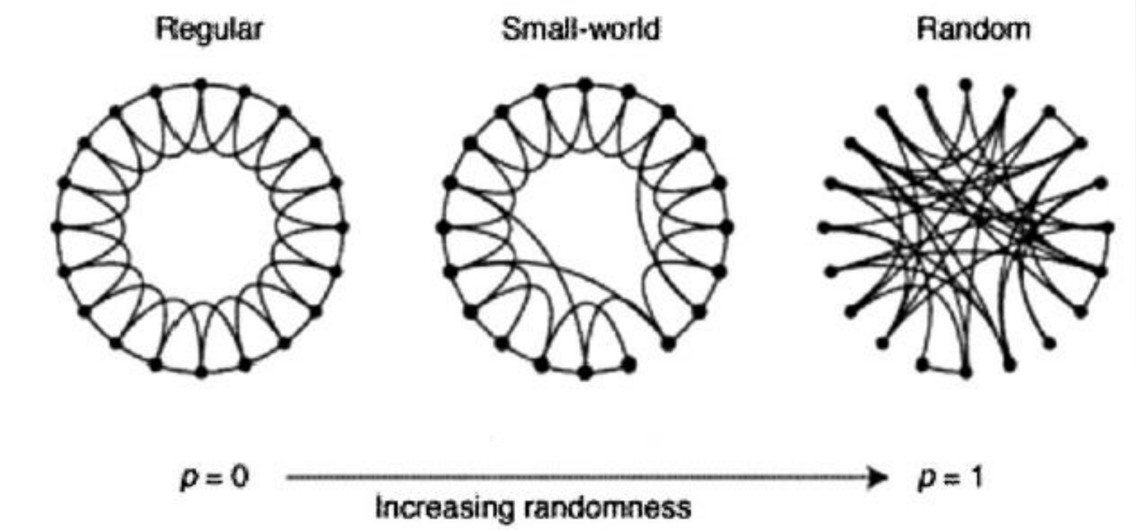
Κατανομή βαθμών για τυχαία δίκτυα

- Η κατανομή βαθμών ενός τυχαίου γραφήματος που δημιουργείται με προσομοίωση
- Δημιουργήθηκε ένα τυχαίο γράφημα με N=10 000 κόμβους και πιθανότητα σύνδεσης p=0.0015
- Υπολογίστηκε ο αριθμός των κόμβων με βαθμό k, X_k



Θυμηθείτε το small-world ή WS μοντέλο…

- Το μοντέλο Duncan Watts και Steven Strogatz
- 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

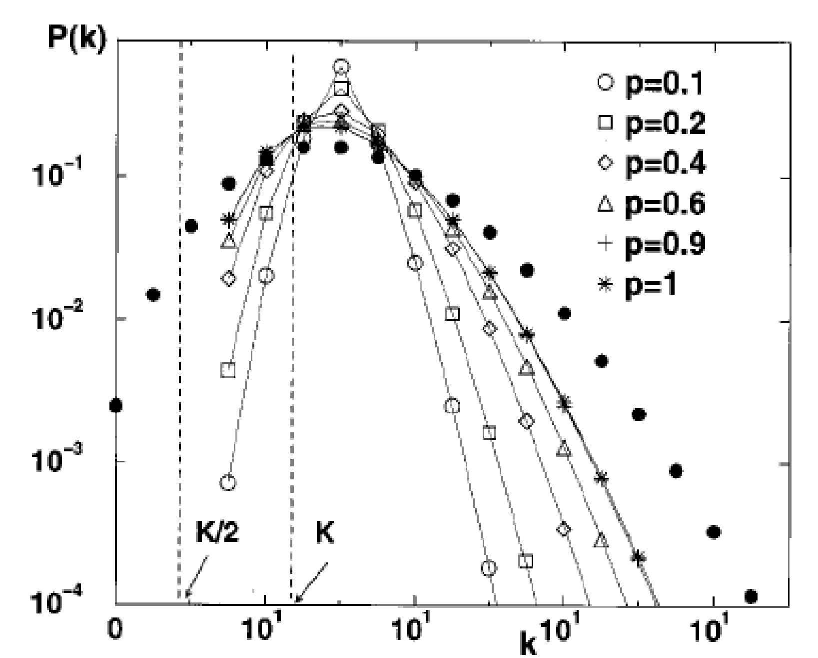


Κατανομή βαθμών για WS δίκτυα

- Μπορεί να δειχτεί ότι η κατανομή βαθμών για WS δίκτυα είναι η ακόλουθη:

$$P(k) = \sum_{n=0}^{f(k,K)} C_{K/2}^{n} (1-p)^n p^{K/2-n} \frac{(pK/2)^{k-K/2-n}}{(k-K/2-n)!} e^{-pK/2}$$

Κατανομή βαθμών για k=3 και διάφορα p



Θυμηθείτε πάλι τα δίκτυα άνευ κλίμακας …

- Τα πραγματικά αποτελέσματα δείχνουν ότι πολλά μεγάλα δίκτυα είναι scale-free, δηλαδή, η κατανομή βαθμών τους ακολουθεί έναν δυναμο-νόμο
- Επιπρόσθετα, ακόμα και για εκείνα τα δίκτυα για τα οποία ο P(k) έχει εκθετική ουρά, η κατανομή βαθμών αποκλίνει σημαντικά από μια κατανομή Poisson
- Έχουμε ήδη δει ότι τα μοντέλα των τυχαίων γραφημάτων και το WS μοντέλο ΔΕΝ μπορούν να αναπαράξουν αυτό το χαρακτηριστικό

Θυμηθείτε πάλι τα δίκτυα άνευ κλίμακας …

- Φυσικά, θα μπορούσαμε να φτιάξουμε μια διαδικασία γέννησης ενός τυχαίου γραφήματος που η κατανομή βαθμών του υπακούει σε κάποιον δυναμο-νόμο, όμως, απλά θ' αναβάλλαμε το εξής σημαντικό ερώτημα:
- **Ποιος είναι εκείνος ο μηχανισμός που δημιουργεί scale-free δίκτυα;**
- Για ν' απαντήσουμε στο ερώτημα αυτό πρέπει να φύγουμε από τη διαδικασία μοντελοποίησης της τοπολογίας ενός δικτύου και να πάμε στις διαδικασίες μοντελοποίησης της συνάθροισης (assembly) και εξέλιξης (evolution) του δικτύου
- Θα μιλήσουμε για τη **δυναμικότητα του δικτύου (network dynamics)**

Το μοντέλο Barabasi-Albert

- Οι Barabasi και Albert, πρώτοι απ' όλους, πρότειναν ότι οι άνευ κλίμακας μορφή των πραγματικών δικτύων οφείλονται σε δυο γενικούς μηχανισμούς:
  - Τα μοντέλα που παρουσιάστηκαν μέχρι στιγμής (RG και WS) υποθέτουν ότι ξεκινάμε με έναν σταθερό αριθμό N κόμβων και τυχαία ανακαλωδιώνουμε ή συνδέουμε κόμβους χωρίς να μεταβάλλουμε το N
  - [**1$^{\text{ος}}$ ΜΗΧΑΝΙΣΜΟΣ**] Όμως, τα πραγματικά δίκτυα είναι *ανοιχτά συστήματα (open systems)* που μεγαλώνουν με τη συνεχή προσθήκη νέων κόμβων
  - Ξεκινώντας από έναν μικρό πυρήνα κόμβων, ο αριθμός των κόμβων αυξάνει σε όλη τη διάρκεια ζωής του δικτύου με συνεχή προσθήκη νέων κόμβων
  - Για παράδειγμα:
    - Ο Παγκόσμιος Ιστός αυξάνει εκθετικά με το χρόνο
    - Η επιστημονική βιβλιογραφία αυξάνει σταθερά με τη δημοσίευση νέων εργασιών (papers)

Το μοντέλο Barabasi-Albert

- [**2ος ΜΗΧΑΝΙΣΜΟΣ**] Τα μοντέλα που παρουσιάστηκαν μέχρι στιγμής, υποθέτουν ότι η πιθανότητα σύνδεσης (εξ υπαρχής ή κατά την ανα-καλωδίωση) δυο κόμβων είναι ανεξάρτητη από τον βαθμό των κόμβων, δηλαδή ότι οι ακμές τοποθετούνται τυχαία
- Όμως, τα περισσότερα πραγματικά δίκτυα, παρουσιάζουν *προνομιακή προσκόλληση* (*preferential attachment*), κατά την οποία η πιθανότητα σύνδεσης σε έναν κόμβο εξαρτάται από το βαθμό του κόμβου.
- Για παράδειγμα
  - Μια ιστοσελίδα είναι πιο πιθανό ότι θα έχει υπερ-συνδέσμους προς δημοφιλή URLs που έχουν ήδη μεγάλο βαθμό, επειδή τέτοια υψηλά διασυνδεδεμένα URLs είναι εύκολο να εντοπιστούν και έτσι είναι γνωστά
  - Ένα νέο άρθρο είναι πιο πιθανό να κάνει αναφορά (cite) γνωστά και συνεπώς much-cited άρθρα, παρά less-cited και συνεπώς λιγότερο γνωστά άρθρα

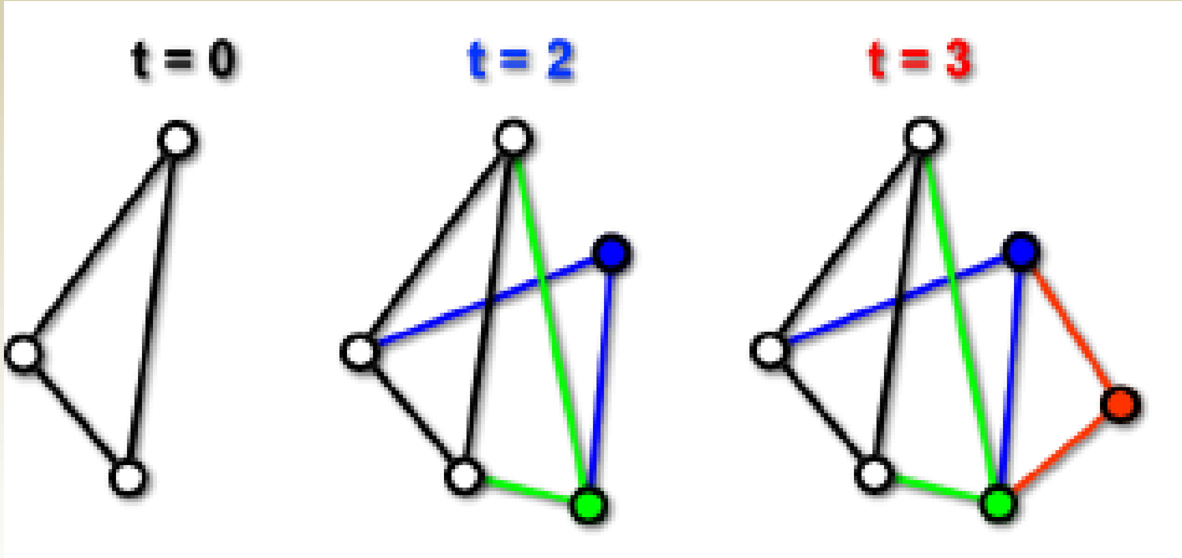
Το μοντέλο Barabasi-Albert

- Αυτά τα δυο συστατικά, αύξηση (growth) και προνομιακή προσκόλληση (preferential attachment), ενέπνευσαν την εισαγωγή του μοντέλου των Barabasi-Albert, το οποίο έδωσε για πρώτη φορά ένα δίκτυο με κατανομή δυναμο-νόμου στους βαθμούς των κόμβων
- Ο αλγόριθμος των Barabasi-Albert είναι ο εξής:
  - <u>Αύξηση</u>: Ξεκινώντας με έναν μικρό αριθμό (m_0) κόμβων, σε κάθε χρονικό βήμα, προσθέτουμε έναν νέο κόμβο με $m(<m_0)$ ακμές που συνδέουν τον νέο κόμβο με m διαφορετικούς κόμβους ήδη υπάρχοντες στο δίκτυο
  - <u>Προνομιακή προσκόλληση</u>: Όταν επιλέγονται οι κόμβοι στους οποίους θα συνδεθεί ένας νέος, υποθέτουμε ότι η πιθανότητα Π με την οποία ο νέος κόμβος θα συνδεθεί σε κάποιον κόμβο i εξαρτάται από το βαθμό k_i του κόμβου i, ώστε:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

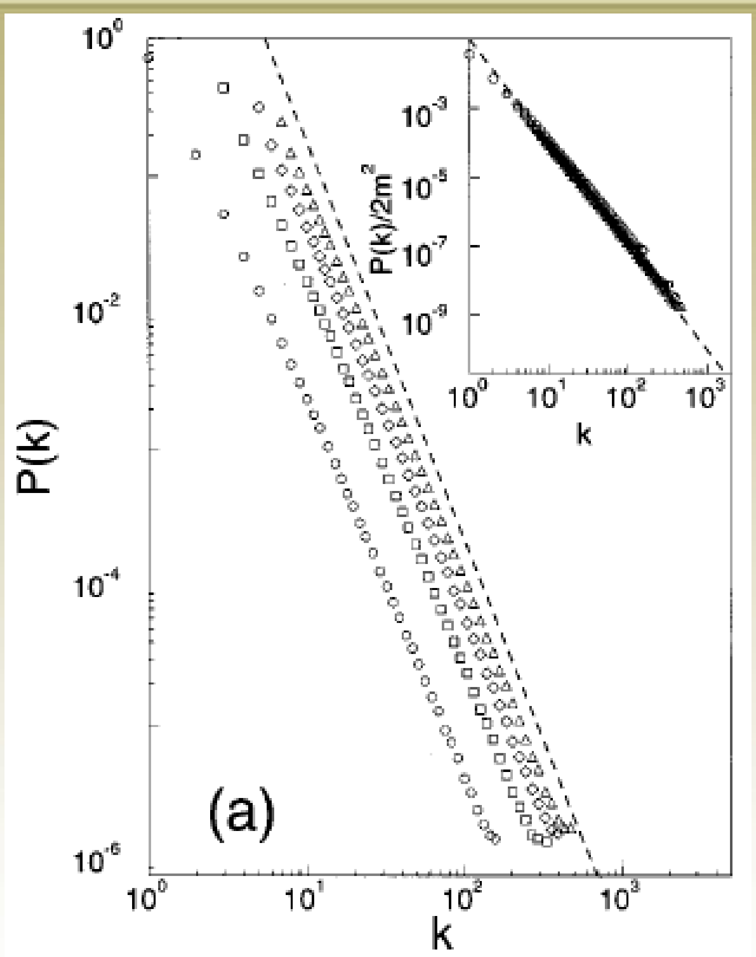
Το μοντέλο Barabasi-Albert

- Ξεκινάμε με ένα μικρό σύνολο m_0 κόμβων και $m_0(m_0-1)/2$ ακμές



Το μοντέλο Barabasi-Albert

- Μετά από t χρονικές στιγμές, αυτή η διαδικασία δημιουργεί ένα δίκτυο με $N=t+m_0$ κόμβους και με ακμές $m_0(m_0-1)/2+mt$
- Προσομοιώσεις δείχνουν ότι αυτό το δίκτυο εξελίσσεται σε scale invariant κατάσταση με την πιθανότητα κάποιος κόμβος να έχει k ακμές να ακολουθεί δυναμο-νόμο με εκθέτη $\gamma_{BA}=3$
- O scaling εκθέτης είναι ανεξάρτητος από το m, τη μοναδική παράμετρο του μοντέλου



Κατανομή βαθμών για το μοντέλο Barabasi-Albert, με $N=m_0+t=300\,000$ s,

○ $m_0=m=1$; □ $m_0=m=3$ ◇ $m_0=m=5$ △ $m_0=m=7$

Η κλίση της στικτής γραμμής είναι γ=2.9

Το μοντέλο Barabasi-Albert

Ιδιότητες του μοντέλου:

- Αρ. κόμβων: $m_0 + t$
- Αρ. ακμών: $m_0(m_0-1)/2+mt$

- Μέσος βαθμός: $$\langle k \rangle = \frac{2E}{N} \rightarrow 2m$$

- Κατανομή βαθμού: $$P(k) \xrightarrow[t\rightarrow\infty]{} Ak^{-3}$$

Κατανομή βαθμών για BA δίκτυα

- Υποθέτοντας ότι το k_i είναι μια συνεχής πραγματική μεταβλητή, ο ρυθμός με τον οποίο μεταβάλλεται ο k_i αναμένεται να είναι ανάλογος της $\Pi(k_i)$
- Συνεπώς, το k_i ικανοποιει την εξίσωση

$$\frac{\partial k_i}{\partial t} = m\Pi(k_i) = m\frac{k_i}{\sum_{j=1}^{N-1} k_j}$$

- Το άθροισμα στον παρονομαστή περιλαμβάνει όλους τους κόμβους του δικτύου, εκτός από τον νεοεισαχθέντα.

Επομένως, η τιμή του είναι $\sum(k_j)$=2mt-m, που οδηγεί στην:

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

Κατανομή βαθμών για BA δίκτυα

- Η λύση αυτής της εξίσωσης, με την αρχική συνθήκη ότι κάθε κόμβος i κατά την εισαγωγή του στο δίκτυο έχει $k_i(t_i)=m$, είναι:

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta}$$

όπου

$$\beta = \frac{1}{2}$$

- Η εξίσωση δείχνει ότι ο βαθμός όλων των κόμβων εξελίσσεται με τον ίδιο τρόπο, ακολουθώντας δυναμο-νόμο, με μόνη διαφορά τη βάση του νόμου
- Με χρήση αυτής της σχέσης, μπορούμε να υπολογίσουμε την πιθανότητα ότι κάποιος κόμβος έχει βαθμό $k_i(t)$ μικρότερο από k, δηλαδή:

$$P[k_i(t) < k] = P\left(t_i > \frac{m^{1/\beta} t}{k^{1/\beta}}\right)$$

Κατανομή βαθμών για BA δίκτυα

- Υποθέτοντας ότι η προσθήκη κόμβων γίνεται σε ίσα διαστήματα στο χρόνο, οι τιμές t_i έχουν σταθερή probability density, ίση με:

$$P(t_i) = \frac{1}{m_0 + t}$$

- Αντικαθιστώντας στην τελευταία σχέση της προηγούμενης διαφάνειας

$$P\left(t_i > \frac{m^{1/\beta} t}{k^{1/\beta}}\right) = 1 - \frac{m^{1/\beta} t}{k^{1/\beta}(t + m_0)}$$

- Η κατανομή του βαθμού P(k) μπορεί να βρεθεί, ως εξής:

$$P(k) = \frac{\partial P[k_i(t) < k]}{\partial k} = \frac{2m^{1/\beta} t}{m_0 + t} \frac{1}{k^{1/\beta + 1}}$$

Κατανομή βαθμών για BA δίκτυα

- Δείχνοντας ότι ασυμπτωτικά (t→∞)

$$P(k) \sim 2m^{1/\beta} k^{-\gamma}$$

όπου $\gamma = \frac{1}{\beta} + 1 = 3$

το οποίο είναι ανεξάρτητο του m, όπως και στα αποτελέσματα της προσομοίωσης

Ασκήσεις (Εξέλιξη δικτύου)

- **Άσκηση 1.**

  Θυμηθείτε το μοντέλο Barabasi-Albert ("the rich gets richer" ή "Matthew effect")

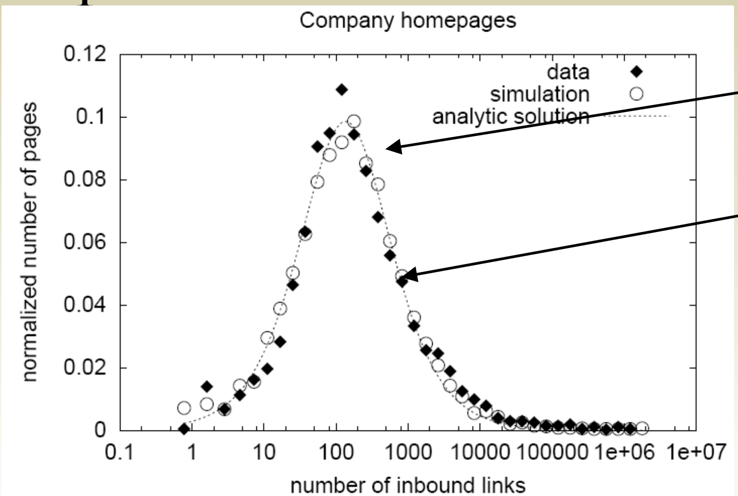
  - Ναι, αλλά η Google, αν και μεταγενέστερη της Yahoo, την ξεπέρασε σε δημοφιλία!
  - Κάποια άρθρα που γράφονται "σήμερα" θα ξεπεράσουν σε αριθμό αναφορών (citations) πολλά από τα σημερινά που έχουν μεγάλους αριθμούς αναφορών!

  Άρα, υπάρχει κάποιος παράγοντας που δεν περιγράφεται καλά από τους BA, και τελικά "οι νικητές δεν τα παίρνουν όλα"

Ασκήσεις (Εξέλιξη δικτύου)

- **Άσκηση 1 (συνέχεια).**

  Πράγματι, μελετώντας τον in-degree ιστοσελίδων του ίδιου τύπου (π.χ., πανεπιστημιακών, δημοσίων εταιρειών, κ.τ.λ.) παρατηρήθηκε ότι για μικρά και μεσαία *k* η κατανομή αποκλίνει από τον δυναμο-νόμο:



t=4923, m=1356, α=0.95

$$\widetilde{\Pr}(k) = \frac{\ln 10}{6} \cdot [2m(1-\alpha)]^{\frac{1}{\alpha}} \cdot k \cdot [\alpha k + 2m(1-\alpha)]^{-1-\frac{1}{\alpha}}$$

Φτιάξτ… δεν μοιράζε… κάθε κόμβου

$$\Pi(k_i) = \alpha \frac{k_i}{2mt} + (1-\alpha)\frac{1}{m_0 + t}$$

Ασκήσεις (Εξέλιξη δικτύου)

- **Άσκηση 2.**

  Για το προηγούμενο μοντέλο, να δείξετε ότι to degree distribution υπακούει σε δυναμο-νόμο.

  Είναι ένα μοντέλο με μείξη προνομιακής και ομοιόμορφης προσκόλλησης, γνωστό ως μοντέλο του David Pennock

Degree distribution για το Pennock et al.

Λύση άσκησης 2.

- Υποθέτουμε ότι το k_i είναι μια συνεχής πραγματική μεταβλητή, και ότι ο $\Pi(k_i)$ είναι ο ρυθμός αύξησης του k_i. Τότε:

$$\frac{\partial k_i}{\partial t} = A\Pi(k_i) = A\alpha\frac{k_i}{2mt} + A(1-\alpha)\frac{1}{m_0+t}$$

- Αφού το $\Pi(k_i)$ αθροίζει στο ένα, και η συνολική συνδεσιμότητα αυξάνει κάθε χρονικό quantum κατά 2m, το A πρέπει να ισούται με 2m. Για το υπόλοιπο των πράξεων υποθέτουμε ότι $t >> m_0$. Αντικαθιστώντας το A, βρίσκουμε ότι:

$$\frac{\partial k_i}{\partial t} = \alpha\frac{k_i}{t} + (1-\alpha)\frac{2m}{t}$$

Κατανομή βαθμών για το Pennock et al.

- Με την αρχική συνθήκη ότι ο κόμβος i ξεκινά τη στιγμή t_i χωρίς ακμές, δηλ., $k_i(t_i)= 0$:

$$k_i(t) = 2m\left(\frac{1-\alpha}{\alpha}\right)\left(\frac{t^\alpha - t_i^\alpha}{t_i^\alpha}\right)$$

- Τότε η πιθανότητα ότι η συνδεσιμότητα ki(t) του κόμβου i, είναι μικρότερη από k είναι:

$$\Pr(k_i(t) < k) = \Pr\left(t_i > t\left[\frac{2m(1-\alpha)}{\alpha k + 2m(1-\alpha)}\right]^{\frac{1}{\alpha}}\right) = 1 - \left[\frac{2m(1-\alpha)}{\alpha k + 2m(1-\alpha)}\right]^{\frac{1}{\alpha}}$$

Κατανομή βαθμών για το Pennock et al.

- όπου το τελευταίο βήμα υποθέτει ότι οι κόμβοι προστίθενται σε ίσα διαστήματα στο χρόνο, και έτσι η pdf του t_i είναι ομοιόμορφη, δηλ., $Pr(t_i)=1/t$. Η pdf του k είναι η μερική παράγωγος της cdf (δηλ., της προηγούμενης σχέσης) σε σχέση με το k:

$$\Pr(k) = \frac{\partial \Pr(k_i(t) < k)}{\partial k} = [2m(1-\alpha)]^{\frac{1}{\alpha}} [\alpha k + 2m(1-\alpha)]^{-1-\frac{1}{\alpha}}$$

- Καθώς $k \to \infty$, η density Pr(k) είναι ανάλογη του $k^{-(1+1/\alpha)}$, δηλαδή δυναμο-νόμος με εκθέτη $\gamma = 1 + 1/\alpha$
- Για παράδειγμα, εάν $\alpha=1/2$, τότε $\gamma=3$, δηλαδή ο ίδιος με αυτόν του μοντέλου BA

Ασκήσεις (Εξέλιξη δικτύου)

- **Άσκηση 3.**

Αναπτύξτε ένα μοντέλο μεγέθυνσης δικτύου, στο οποίο η πιθανότητα "προσκόλλησης" ενός νεοσειρχομένου κόμβου σε έναν ήδη υπάρχοντα κόμβο είναι συνάρτηση όχι μόνο του βαθμού του, αλλά και ενός παράγοντα fitness. Δηλαδή, ένας τέτοιος νόμος θα ενσάρκωνε την έννοια του "the fittest survives", δηλ., ένας late comer θα μπορούσε να κυριαρχήσει εάν είχε ικανή δυναμική. Π.χ., η Google κυριάρχησε στην αγορά μηχανών αναζήτησης, παρόλο που η Yahoo εισήλθε πρώτη στην αγορά, και άρα είχε πλεονέκτημα.

- **Λύση.**

$$p_i = \frac{k_i \phi_i}{\sum_{j \in N} k_j \phi_j}$$

Είναι το γνωστό Bianconi-Barabasi μοντέλο.