Σημασιολογικός και Κοινωνικός Ιστός

Διάλεξη 10 – Θεωρία Γράφων

Γεώργιος Δημητρακόπουλος

dimitrakopoulos@ionio.gr

Τμήμα Πληροφορικής

Ιόνιο Πανεπιστήμιο

Γραφήματα ή Γράφοι

Ένα γράφημα $G = (V, E)$ αποτελείται από ένα πεπερασμένο μη κενό σύνολο V με *κορυφές* ή *κόμβους*, και ένα σύνολο $E \subseteq V \times V$ *με ακμές* το οποίο περιλαμβάνει *μη διατεταγμένα* ζεύγη κορυφών.

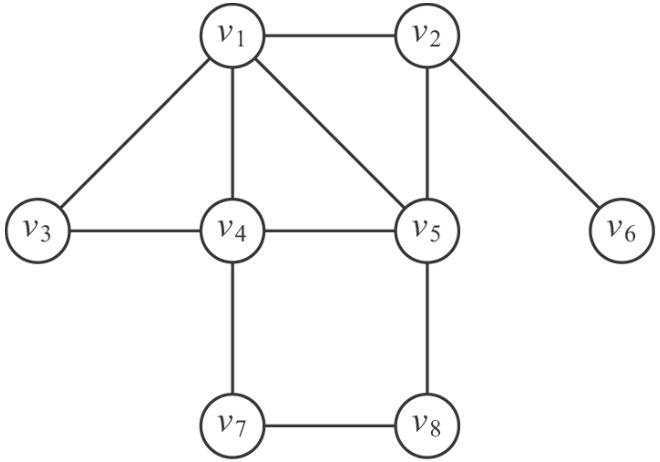
Το πλήθος των κόμβων του γραφήματος G ορίζεται ως $|V| = n$ και ονομάζεται *τάξη* του γραφήματος, ενώ το πλήθος των ακμών ορίζεται ως $|E| = m$ και ονομάζεται *μέγεθος* του G.

Ένα *κατευθυνόμενο γράφημα* (digraph) διαθέτει ένα σύνολο ακμών E το οποίο αποτελείται από *διατεταγμένα* ζεύγη κορυφών.

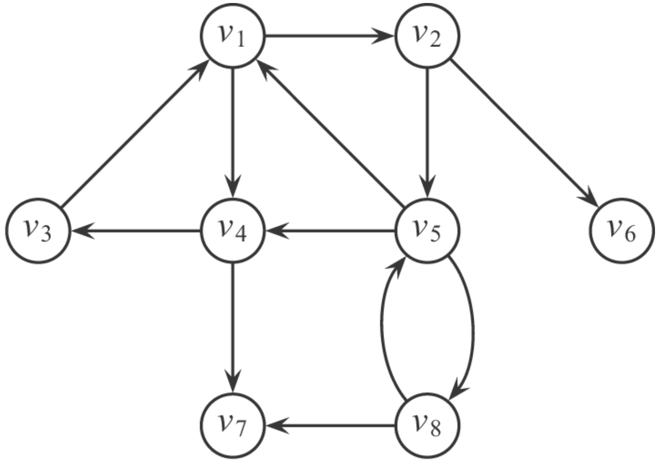
Σε ένα *σταθμισμένο γράφημα*, σε κάθε ακμή $(v_i, v_j) \in E$ έχει αντιστοιχιστεί ένα βάρος (δηλαδή ένας συντελεστής στάθμισης) w_{ij}.

Ένα γράφημα $H = (V_H, E_H)$ ονομάζεται *υπογράφημα* του $G = (V, E)$ αν ισχύει $V_H \subseteq V$ και $E_H \subseteq E$.

Μη κατευθυνόμενα και κατευθυνόμενα γραφήματα



(α)



(β)

Κατανομή βαθμών

Ο *βαθμός* ενός κόμβου $v_i \in V$ είναι το πλήθος των ακμών που προσπίπτουν σε αυτόν, και συμβολίζεται με $d(v_i)$ ή πιο απλά με d_i.

Η *ακολουθία βαθμών* ενός γραφήματος είναι η λίστα με τους βαθμούς των κόμβων ταξινομημένους κατά μη αύξουσα σειρά.

Έστω ότι με N_k συμβολίζουμε το πλήθος των κορυφών με βαθμό k. Η *κατανομή συχνότητας βαθμών* ενός γραφήματος ορίζεται ως

$$(N_0, N_1, \dots, N_t)$$

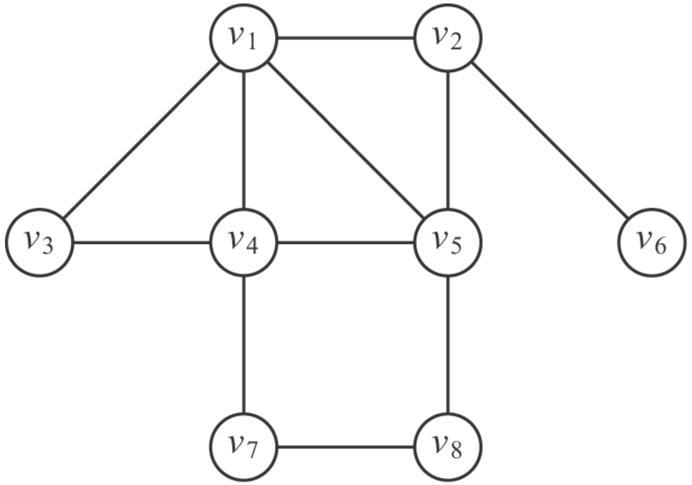
όπου t είναι ο μέγιστος βαθμός για έναν κόμβο του G.

Έστω ότι X είναι μια τυχαία μεταβλητή που αναπαριστά τον βαθμό ενός κόμβου. Η *κατανομή βαθμών* ενός γραφήματος δίνει τη συνάρτηση μάζας πιθανότητας f για τη μεταβλητή X, που ορίζεται ως

$$(f(0), f(1), \dots, f(t))$$

όπου $f(k) = P(X = k) = \frac{N_k}{n}$ είναι η πιθανότητα ενός κόμβου με βαθμό k.

Κατανομή βαθμών



Η ακολουθία βαθμών του γραφήματος είναι

$$(4, 4, 4, 3, 2, 2, 2, 1)$$

Η κατανομή της συχνότητας βαθμών είναι

$$(N_0, N_1, N_2, N_3, N_4) = (0, 1, 3, 1, 3)$$

Η κατανομή βαθμών είναι

$$(f(0), f(1), f(2), f(3), f(4)) = (0, 0.125, 0.375, 0.125, 0.375)$$

Διαδρομή, απόσταση

- Ένας *περίπατος* μεταξύ των κόμβων x και y σε ένα γράφημα G είναι μια διατεταγμένη ακολουθία κορυφών, η οποία ξεκινά από τον κόμβο x και τελειώνει στον κόμβο y,

$$x = v_0, v_1, \ldots, v_{t-1}, v_t = y$$

  τέτοια ώστε να υπάρχει μια ακμή που ενώνει κάθε ζεύγος διαδοχικών κορυφών, δηλαδή ισχύει $(v_{i-1}, v_i) \in E$ για όλα τα $i = 1, 2, \ldots, t$. Το μήκος t του περιπάτου μετριέται σε *άλματα*, δηλαδή είναι το πλήθος των ακμών κατά μήκος του περιπάτου.
- Μια *διαδρομή* είναι ένας περίπατος χωρίς επαναλαμβανόμενες κορυφές (με εξαίρεση την αρχική και την τελική κορυφή). Μια διαδρομή ελάχιστου μήκους μεταξύ των κόμβων x και y ονομάζεται *συντομότερη διαδρομή*· το μήκος της συντομότερης διαδρομής είναι η *απόσταση* των κόμβων x και y και συμβολίζεται με $d(x, y)$. Αν δεν υπάρχει διαδρομή μεταξύ των δύο κόμβων, τότε θεωρούμε ότι η απόσταση είναι $d(x, y) = \infty$.

Συνεκτικότητα

- Αν υπάρχει διαδρομή μεταξύ δύο κόμβων v_i και v_j, λέμε ότι οι κόμβοι είναι *συνδεδεμένοι*.
- Ένα γράφημα είναι *συνεκτικό* αν υπάρχει διαδρομή για όλα τα ζεύγη κορυφών. Μια *συνεκτική συνιστώσα*, ή απλώς *συνιστώσα*, ενός γραφήματος είναι ένα υπογράφημα που έχει μέγιστη συνεκτικότητα.
- Ένα κατευθυνόμενο γράφημα είναι *ισχυρά συνεκτικό* αν υπάρχει μια κατευθυνόμενη διαδρομή για κάθε διατεταγμένο ζεύγος κορυφών. Αντίθετα, είναι *ασθενώς συνεκτικό* αν υπάρχει μια διαδρομή για κάθε ζεύγος κόμβων μόνο όταν λαμβάνονται υπόψη οι μη κατευθυνόμενες ακμές.

Μήτρα γειτνίασης

Ένα γράφημα $G = (V, E)$ με $|V| = n$ κορυφές μπορεί να αναπαρασταθεί βολικά με τη μορφή μιας συμμετρικής δυαδικής *μήτρας γειτνίασης* **A**, διαστάσεων $n \times n$, η οποία ορίζεται ως εξής:

$$\mathbf{A}(i, j) = \begin{cases} 1 & \text{αν η κορυφή } v_i \text{ είναι γειτονική με την κορυφή } v_j \\ 0 & \text{διαφορετικά} \end{cases}$$

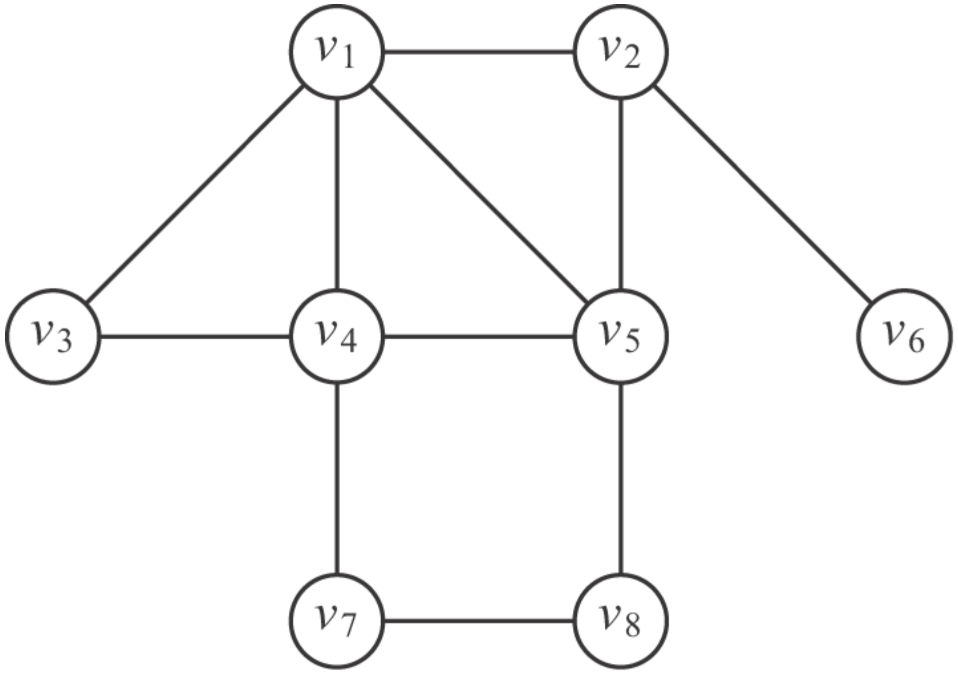
Αν το γράφημα είναι κατευθυνόμενο, τότε η μήτρα γειτνίασης **A** δεν είναι συμμετρική.

Αν το γράφημα είναι σταθμισμένο, τότε μπορούμε να ορίσουμε μια *σταθμισμένη μήτρα γειτνίασης* **A**, διαστάσεων $n \times n$, όπως φαίνεται παρακάτω:

$$\mathbf{A}(i, j) = \begin{cases} w_{ij} & \text{αν η κορυφή } v_i \text{ είναι γειτονική με την κορυφή } v_j \\ 0 & \text{διαφορετικά} \end{cases}$$

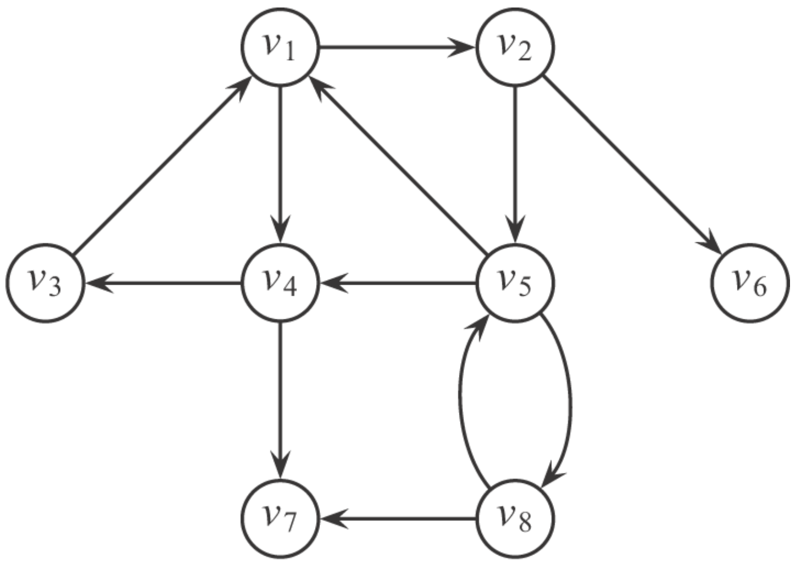
όπου w_{ij} είναι το βάρος της ακμής $(v_i, v_j) \in E$.

Μήτρα γειτνίασης



	v1	v2	v3	v4	v5	v6	v7	v8
v1	0	1	1	1	1	0	0	0
V2	1	0	0	0	1	1	0	0
v3	1	0	0	1	0	0	0	0
v4	1	0	1	0	1	0	1	0
v5	1	1	0	1	0	0	0	1
v6	0	1	0	0	0	0	0	0
v7	0	0	0	1	0	0	0	1
v8	0	0	0	0	1	0	1	0

Μήτρα γειτνίασης



(β)

	v1	v2	v3	v4	v5	v6	v7	v8
v1	0	1	0	1	0	0	0	0
V2	0	0	0	0	1	1	0	0
v3	1	0	0	0	0	0	0	0
v4	0	0	1	0	0	0	1	0
v5	1	0	0	1	0	0	0	1
v6	0	0	0	0	0	0	0	0
v7	0	0	0	0	0	0	0	0
v8	0	0	0	0	1	0	1	0

ασθενώς συνεκτικό γράφημα: πχ από την v6 δεν μπορούμε να πάμε σε άλλη κορυφή, συνδέεται μόνο με μη κατευθυνόμενο τρόπο με τις υπόλοιπες κορυφές

Γραφήματα από τη μήτρα δεδομένων

Πολλά σύνολα δεδομένων δεν έχουν τη μορφή γραφημάτων, αλλά αυτό δεν σημαίνει ότι είναι αδύνατη η σχετική μετατροπή.

Έστω ότι $\mathbf{D} = \{\mathbf{x}_i\}_{i=1}^{n}$ (με $\mathbf{x}_i \in \mathbb{R}^d$) είναι ένα σύνολο δεδομένων. Ορίζουμε ένα σταθμισμένο γράφημα $G = (V, E)$, με βάρος ακμής

$$w_{ij} = sim(\mathbf{x}_i, \mathbf{x}_j)$$

όπου $sim(\mathbf{x}_i, \mathbf{x}_j)$ είναι η ομοιότητα των σημείων $\mathbf{x}_i$ και $\mathbf{x}_j$.

Για παράδειγμα, μπορούμε να χρησιμοποιήσουμε την ομοιότητα κατά Gauss

$$w_{ij} = sim(\mathbf{x}_i, \mathbf{x}_j) = \exp\left\{-\frac{\| \mathbf{x}_i - \mathbf{x}_j \|^2}{2\sigma^2}\right\}$$

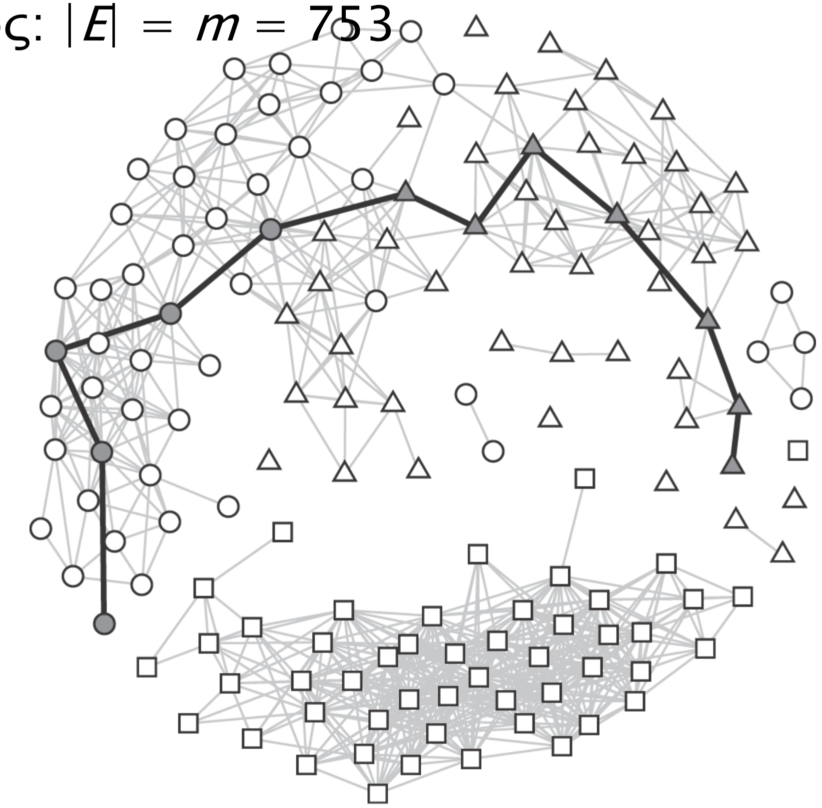
όπου σ είναι η παράμετρος διασποράς.

Εναλλακτικά, μπορούμε να χρησιμοποιήσουμε μια μετρική απόστασης.

Γράφημα ομοιότητας για το σύνολο δεδομένων Iris: Ομοιότητα κατά Gauss

$\sigma = 1/\sqrt{2}$ = 0.777, υπάρχει ακμή αν $w_{ij} \geq 0.777$
(για απόσταση, υπάρχει ακμή αν $w_{ij} < d$)

τάξη: $|V| = n = 150$, μέγεθος: $|E| = m = 753$



Τοπολογικά γνωρίσματα γραφημάτων

Τα γνωρίσματα ενός γραφήματος είναι *τοπικά* αν ισχύουν μόνο για έναν κόμβο (ή μία ακμή), ή *καθολικά* αν αναφέρονται σε ολόκληρο το γράφημα.

Βαθμός: Ο βαθμός ενός κόμβου $v_i \in G$ ορίζεται ως

$$d_i = \sum_j \mathbf{A}(i, j)$$

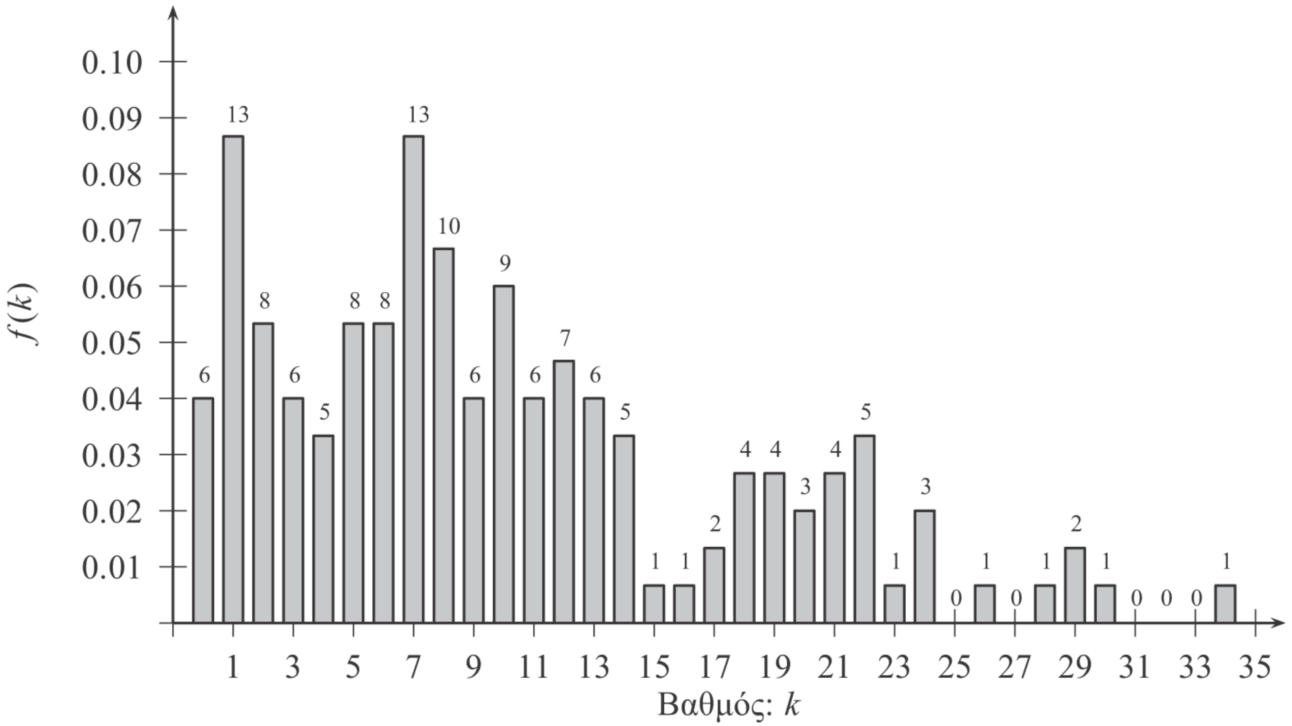
Το αντίστοιχο καθολικό γνώρισμα για ολόκληρο το γράφημα G είναι ο μέσος βαθμός:

$$\mu_d = \frac{\sum_i d_i}{n}$$

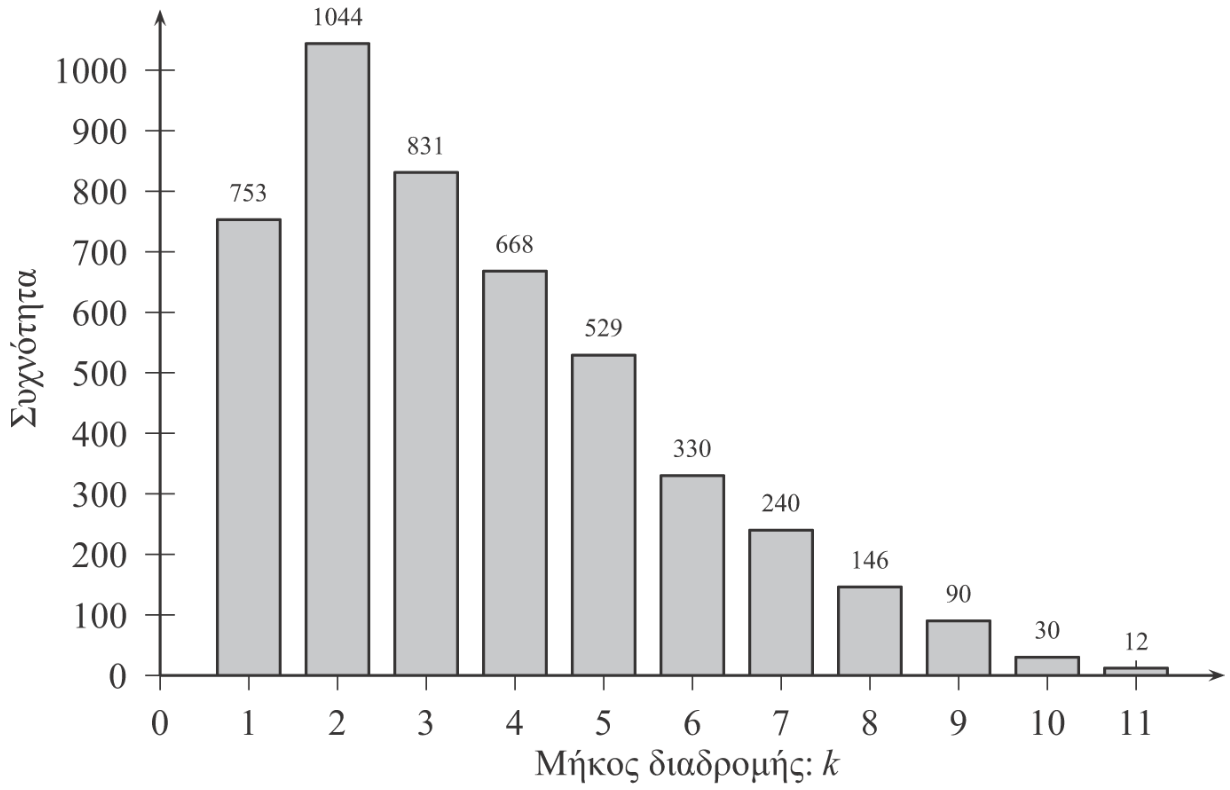
Μέσο μήκος διαδρομής: Το *μέσο μήκος διαδρομής δίνεται από τη σχέση*

$$\mu_L = \frac{\sum_i \sum_{j>i} d(v_i, v_j)}{\binom{n}{2}} = \frac{2}{n(n-1)} \sum_i \sum_{j>i} d(v_i, v_j)$$

Γράφημα του συνόλου δεδομένων Iris: Κατανομή βαθμών



Γράφημα του συνόλου δεδομένων Iris: Ιστόγραμμα μήκους διαδρομής



Εκκεντρότητα, ακτίνα και διάμετρος

Η *εκκεντρότητα* ενός κόμβου v_i είναι η μέγιστη απόσταση του v_i από οποιονδήποτε άλλον κόμβο του γραφήματος:

$$e(v_i) = \max_j \{d(v_i, v_j)\}$$

Η *ακτίνα* ενός συνεκτικού γραφήματος συμβολίζεται με $r(G)$ και είναι η ελάχιστη εκκεντρότητα οποιουδήποτε κόμβου του γραφήματος:

$$r(G) = \min_i \{e(v_i)\} = \min_i \left\{\max_j \{d(v_i, v_j)\}\right\}$$

Η *διάμετρος* συμβολίζεται με $d(G)$ και είναι η μέγιστη εκκεντρότητα οποιουδήποτε κόμβου του γραφήματος:

$$d(G) = \max_i \{e(v_i)\} = \max_{i,j} \{d(v_i, v_j)\}$$

Στην περίπτωση ενός μη συνεκτικού γραφήματος, οι τιμές υπολογίζονται για όλες τις συνεκτικές συνιστώσες του γραφήματος.

Η διάμετρος ενός γραφήματος G επηρεάζεται από τις έκτοπες τιμές (outliers). Μια πιο ανθεκτική έννοια είναι η *πραγματική διάμετρος*, που ορίζεται ως το ελάχιστο πλήθος αλμάτων για το οποίο ένα μεγάλο ποσοστό –συνήθως της τάξης του 90%– των κόμβων όλων των συνδεδεμένων ζευγών είναι προσπελάσιμοι ο ένας από τον άλλο.

Συντελεστής συσταδοποίησης

Ο *συντελεστής συσταδοποίησης* ενός κόμβου v_i είναι ένα μέτρο της πυκνότητας των ακμών στη γειτονιά του v_i.

Έστω ότι $G_i = (V_i, E_i)$ είναι το υπογράφημα που επάγεται από τους γείτονες της κορυφής v_i. Σημειώστε ότι ο κόμβος $v_i \notin V_i$ επειδή έχουμε υποθέσει ότι το υπογράφημα G είναι απλό.

Έστω ότι $|V_i| = n_i$ είναι το πλήθος των γειτόνων του v_i, και $|E_i| = m_i$ είναι το πλήθος των ακμών που υπάρχουν ανάμεσα στους γείτονες του v_i. Ο συντελεστής συσταδοποίησης του v_i ορίζεται ως

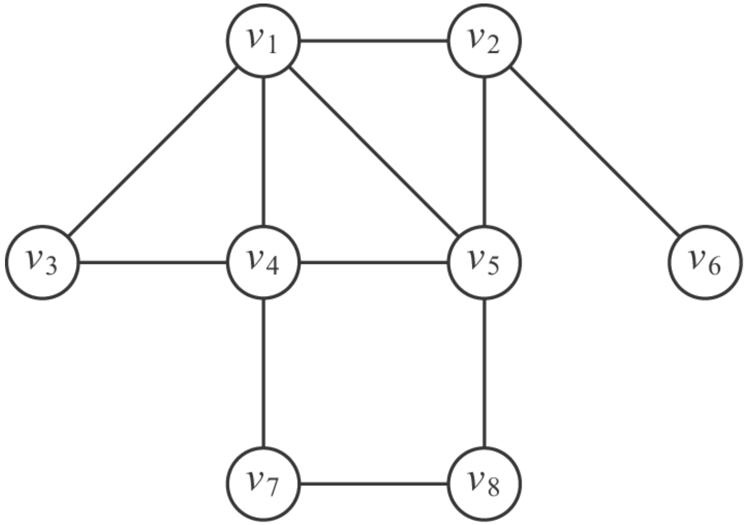
$$C(v_i) = \frac{\text{πλήθος κόμβων στο } G_i}{\text{μέγιστο πλήθος κόμβων στο } G_i} = \frac{m_i}{\binom{n_i}{2}} = \frac{2 \cdot m_i}{n_i(n_i - 1)}$$

Ο συντελεστής συσταδοποίησης ενός γραφήματος G είναι απλώς ο μέσος συντελεστής συσταδοποίησης για όλους τους κόμβους και δίνεται από τη σχέση

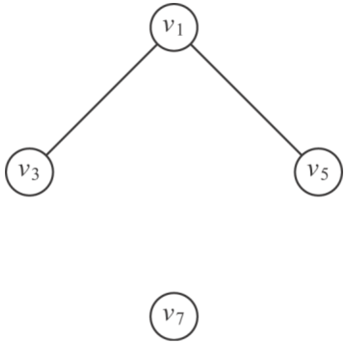
$$C(G) = \frac{1}{n} \sum_i C(v_i)$$

Επειδή το $C(v_i)$ είναι καλώς ορισμένο μόνο για κόμβους με βαθμό $d(v_i) \geq 2$, μπορούμε να ορίσουμε ότι θα ισχύει $C(v_i) = 0$ αν $d_i < 2$.

Συντελεστής συσταδοποίησης



Το γράφημα που επάγεται από τον κόμβο v_4:



Ο συντελεστής συσταδοποίησης του κόμβου v_4 είναι

$$C(v_4) = \frac{2}{\binom{4}{2}} = \frac{2}{6} = 0.33$$

Ο συντελεστής συσταδοποίησης για ολόκληρο το γράφημα G είναι

$$C(G) = \frac{1}{8}\left(\frac{1}{2} + \frac{1}{3} + 1 + \frac{1}{3} + \frac{1}{3}\right) = \frac{2.5}{8} = 0.3125$$

Ανάλυση κεντρικότητας

Κεντρικότητα του βαθμού: Η απλούστερη έννοια κεντρικότητας είναι ο βαθμός d_i μιας κορυφής v_i –όσο μεγαλύτερος είναι ο βαθμός, τόσο πιο σημαντική ή κεντρική είναι η κορυφή.

Κεντρικότητα της εκκεντρότητας: Η κεντρικότητα της εκκεντρότητας ορίζεται ως εξής:

$$c(v_i) = \frac{1}{e(v_i)} = \frac{1}{\max_j \left\{ d(v_i, v_j) \right\}}$$

Όσο μικρότερη είναι η εκκεντρότητα ενός κόμβου e (ή πιο μεγάλη η κεντρικότητα της εκκεντρότητας), τόσο πιο κεντρικός είναι ο κόμβος.

Κεντρικότητα της εγγύτητας: Η κεντρικότητα της εγγύτητας χρησιμοποιεί το άθροισμα όλων των αποστάσεων για να υπολογίσει πόσο κεντρικός είναι ένας κόμβος:

$$c(v_i) = \frac{1}{\sum_j d(v_i, v_j)}$$

Κεντρικότητα της ενδιαμεσότητας

Η κεντρικότητα της ενδιαμεσότητας μετρά πόσες συντομότερες διαδρομές μεταξύ όλων των ζευγών κορυφών περιλαμβάνουν την κορυφή v_i. Αυτό παρέχει μια ένδειξη για το πόσο κεντρικός είναι ο ρόλος «παρακολούθησης» που παίζει η v_i για διάφορα ζεύγη κόμβων.

Έστω ότι με η_{jk} συμβολίζουμε το πλήθος των συντομότερων διαδρομών μεταξύ των κορυφών v_j και v_k, και έστω ότι με $\eta_{jk}(v_i)$ συμβολίζουμε το πλήθος τέτοιων διαδρομών στις οποίες περιλαμβάνεται ή περιέχεται η κορυφή v_i.

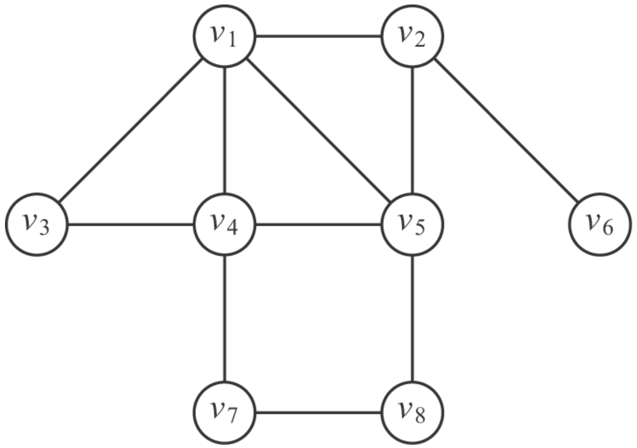
Το ποσοστό των διαδρομών που διέρχονται από την κορυφή v_i συμβολίζεται με

$$\gamma_{jk}(v_i) = \frac{\eta_{jk}(v_i)}{\eta_{jk}}$$

Η κεντρικότητα της ενδιαμεσότητας για έναν κόμβο v_i ορίζεται ως

$$c(v_i) = \sum_{j \neq i} \sum_{\substack{k \neq i \\ k > j}} \gamma_{jk}(v_i) = \sum_{j \neq i} \sum_{\substack{k \neq i \\ k > j}} \frac{\eta_{jk}(v_i)}{\eta_{jk}}$$

Τιμές κεντρικότητας



Κεντρικότητα	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
Βαθμού	**4**	3	2	**4**	**4**	1	2	2
Εκκεντρότητας	**0.5**	0.33	0.33	0.33	**0.5**	0.25	0.25	0.33
$e(v_i)$	2	3	3	3	2	4	4	3
Εγγύτητας	**0.100**	0.083	0.071	0.091	**0.100**	0.056	0.067	0.071
$\sum_j d(v_i, v_j)$	10	12	14	11	10	18	15	14
Ενδιαμεσότητας	4.5	6	0	5	**6.5**	0	0.83	1.17

Με έντονα ο πιο κεντρικός κόμβος ανάλογα με τη μετρική

Ιδιότητα του μικρού κόσμου

Έχει παρατηρηθεί ότι πολλά πραγματικά γραφήματα διαθέτουν την ιδιότητα του αποκαλούμενου *μικρού κόσμου*, η οποία ορίζει ότι υπάρχει μια σύντομη διαδρομή μεταξύ οποιουδήποτε ζεύγους κόμβων. Λέμε ότι ένα γράφημα G εμφανίζει συμπεριφορά μικρού κόσμου αν το μέσο μήκος διαδρομής μ_L αλλάζει κλίμακα λογαριθμικά ως προς το πλήθος των κόμβων του γραφήματος, δηλαδή αν ισχύει

$$\mu_L \propto \log n$$

όπου n είναι το πλήθος των κόμβων του γραφήματος.

Λέμε ότι ένα γράφημα έχει την ιδιότητα του υπερ-μικρού κόσμου (ultra-small-world property) αν το μέσο μήκος διαδρομής είναι πολύ μικρότερο από το $\log n$, δηλαδή αν ισχύει $\mu_L \ll \log n$.

Η ιδιότητα της ανεξαρτησίας από την κλίμακα

Σε πολλά πραγματικά γραφήματα, έχει παρατηρηθεί ότι η εμπειρική κατανομή βαθμών *f*(*k*) εμφανίζει μια συμπεριφορά *ανεξαρτησίας από την κλίμακα* (scale-free) η οποία αποτυπώνεται από μια σχέση δυνάμεων του *k*· δηλαδή, η πιθανότητα να έχει ένας κόμβος βαθμό *k* ικανοποιεί τη συνθήκη

$$f(k) \propto k^{-\gamma}$$

Υπολογίζοντας τον λογάριθμο και για τα δύο μέλη της Εξίσωσης, παίρνουμε

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

$$\text{ή } \log f(k) = -\gamma \log k + \log \alpha$$

δηλαδή την εξίσωση μιας ευθείας στο διάγραμμα των *k* και *f*(*k*) με άξονες λογαριθμικής κλίμακας (log-log), όπου το $-\gamma$ δίνει την κλίση της ευθείας.

Το φαινόμενο της συσταδοποίησης

Στα πραγματικά γραφήματα παρατηρείται συχνά και το *φαινόμενο της συσταδοποίησης*: Δύο κόμβοι είναι πιο πιθανό να συνδέονται αν έχουν κοινό γείτονα. Το φαινόμενο αυτό αποτυπώνεται στον υψηλό συντελεστή συσταδοποίησης για το γράφημα G.

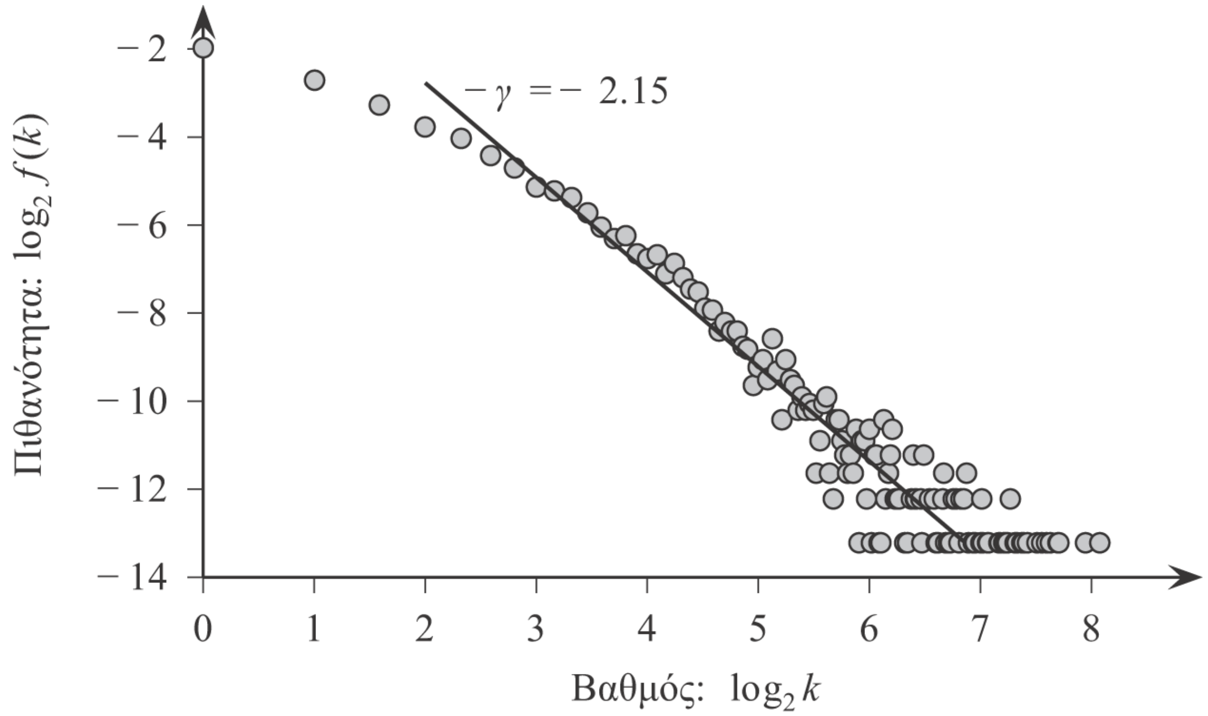
Έστω ότι με $C(k)$ συμβολίζουμε τον μέσο συντελεστή συσταδοποίησης για όλους τους κόμβους με βαθμό k· τότε, το φαινόμενο της συσταδοποίησης εκδηλώνεται επίσης ως μια σχέση δυνάμεων μεταξύ των $C(k)$ και k:

$$C(k) \propto k^{-\gamma}$$

Δηλαδή, το διάγραμμα log-log των k και $C(k)$ εμφανίζει τη συμπεριφορά ευθείας με αρνητική κλίση $-\gamma$.

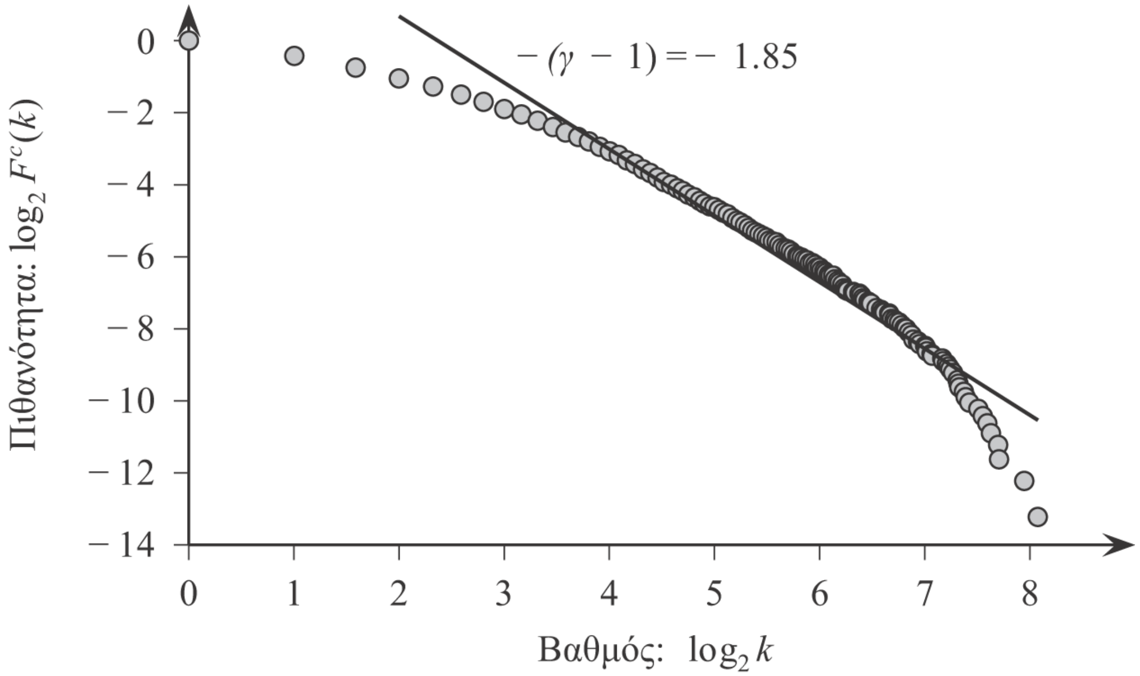
Κατανομή βαθμών: Δίκτυο αλληλεπίδρασης ανθρώπινων πρωτεϊνών

$|V| = n = 9521$, $|E| = m = 37060$

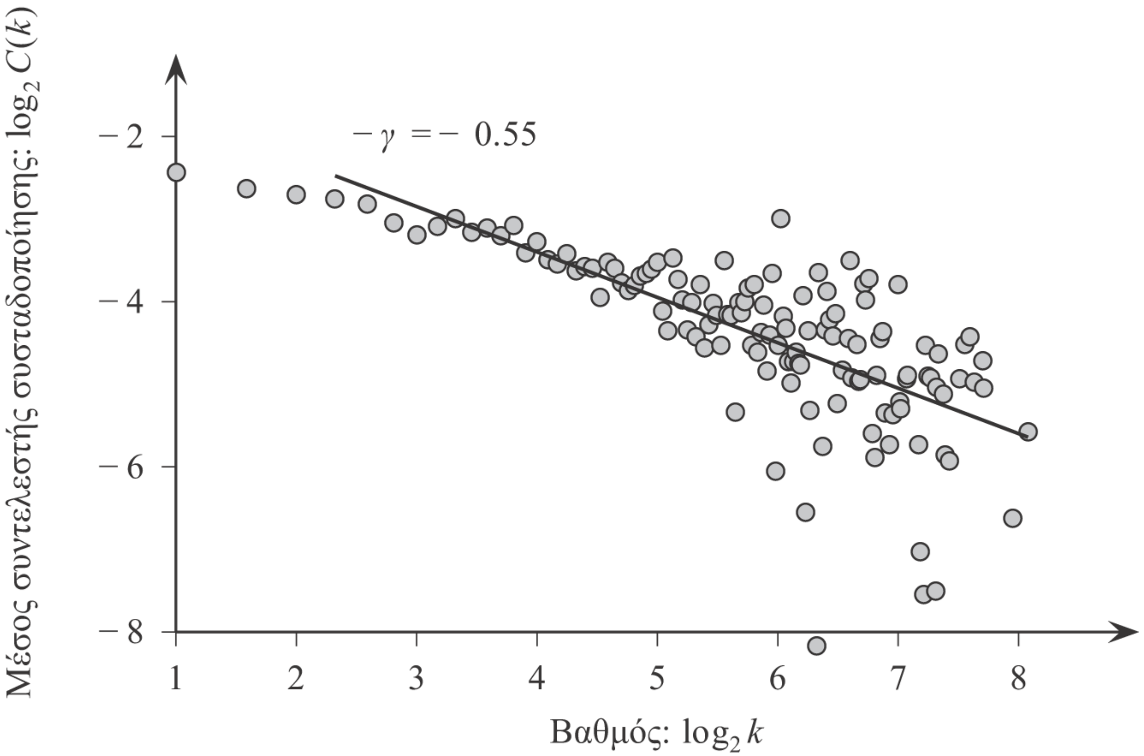


Αθροιστική κατανομή βαθμών

$F^c(k) = 1 - F(k)$, όπου $F(k)$ είναι η συνάρτηση CDF για την $f(k)$



Μέσος συντελεστής συσταδοποίησης



Το μοντέλο του τυχαίου γραφήματος των Erdös-Rényi

Το μοντέλο ER καθορίζει μια συλλογή γραφημάτων $\mathcal{G}(n,m)$ με *n* κόμβους και *m* ακμές, τέτοια ώστε κάθε γράφημα $G \in \mathcal{G}$ να έχει την ίδια πιθανότητα επιλογής:

$$P(G) = \frac{1}{\binom{M}{m}} = \binom{M}{m}^{-1}$$

όπου $M = \binom{n}{2} = \frac{n(n-1)}{2}$ και $\binom{M}{m}$ είναι το πλήθος όλων των γραφημάτων που μπορούν να προκύψουν με *m* ακμές (και *n* κόμβους).

Παραγωγή τυχαίου γραφήματος: Επιλέγουμε τυχαία δύο διαφορετικές κορυφές v_i, $v_j \in V$, και προσθέτουμε την ακμή (v_i, v_j) στο *E*, με την προϋπόθεση ότι η ακμή δεν υπάρχει ήδη στο γράφημα *G*. Η διαδικασία επαναλαμβάνεται μέχρι να προστεθούν στο γράφημα ακριβώς *m* ακμές.

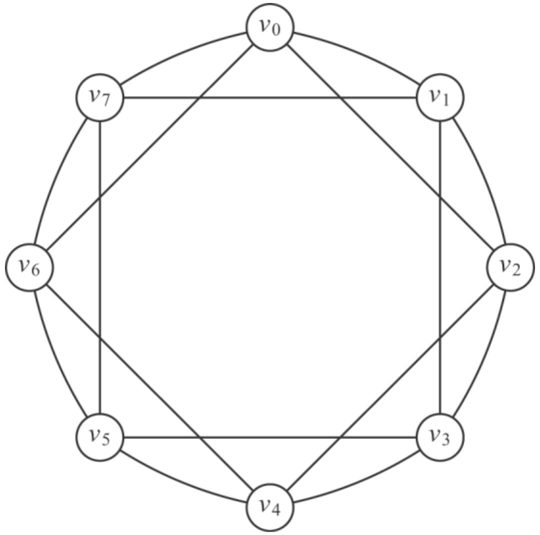
Αποδεικνύεται ότι τα τυχαία γραφήματα ER εμφανίζουν συμπεριφορά μικρού κόσμου.

Κανονικό γράφημα

Κανονικό (regular) γράφημα βαθμού 2k: έχει n κόμβους σε κυκλική διάταξη, όπου κάθε κόμβος είναι συνδεδεμένος με k γείτονες στα δεξιά και k γείτονες στα αριστερά.

Το κανονικό γράφημα εμφανίζει υψηλή τοπική συσταδοποίηση.

Κανονικό γράφημα: n = 8, k = 2

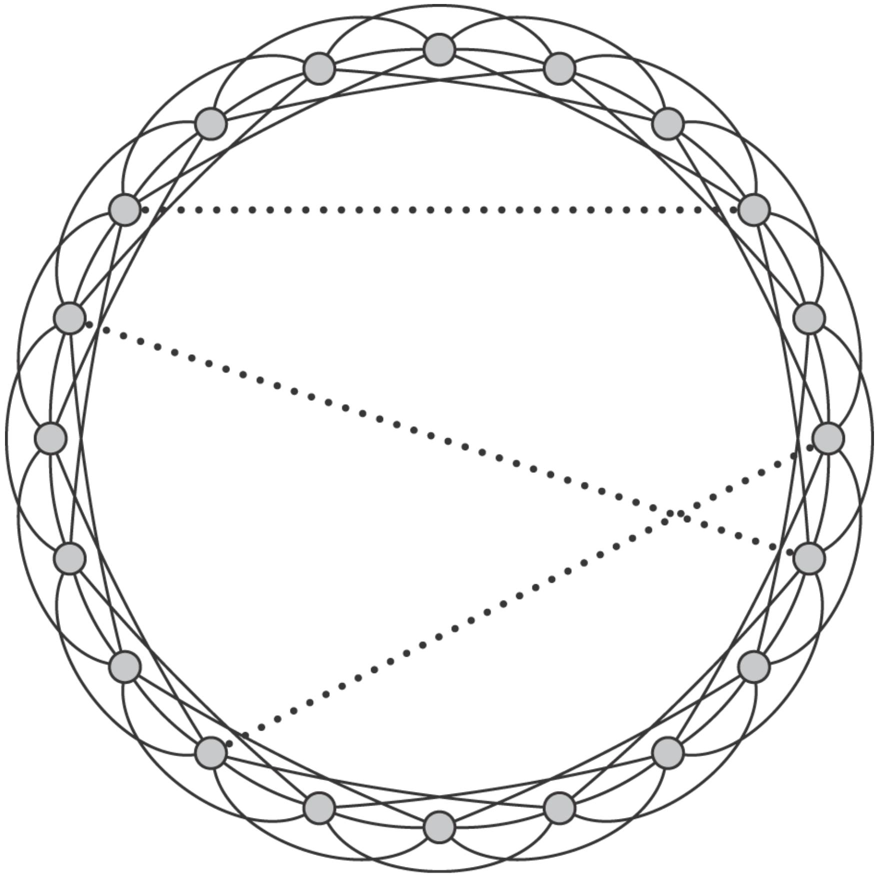


Το μοντέλο του γραφήματος μικρού κόσμου των Watts-Strogatz

Το μοντέλο των Watts-Strogatz (WS) ξεκινά από ένα *κανονικό* (regular) γράφημα βαθμού 2k.

Το κανονικό γράφημα εμφανίζει υψηλή τοπική συσταδοποίηση. Αν προσθέσουμε λίγη τυχαιότητα, αυτό θα έχει ως αποτέλεσμα να εκδηλωθούν φαινόμενα μικρού κόσμου.

$n = 20,\ k = 3$



Τυχαία διαταραχή του κανονικού γραφήματος

Επανασύνδεση ακμών: Για κάθε ακμή (u, v) του γραφήματος, με πιθανότητα r, αντικαθιστούμε τον κόμβο v με κάποιον άλλο τυχαία επιλεγμένο κόμβο αποφεύγοντας τους βρόχους και τις επαναλαμβανόμενες ακμές.

Συντομεύσεις ακμών: Προσθέτουμε λίγες ακμές *συντόμευσης* μεταξύ τυχαίων ζευγών κόμβων, με το r *να* είναι η πιθανότητα, ανά ακμή, της προσθήκης μιας ακμής συντόμευσης.

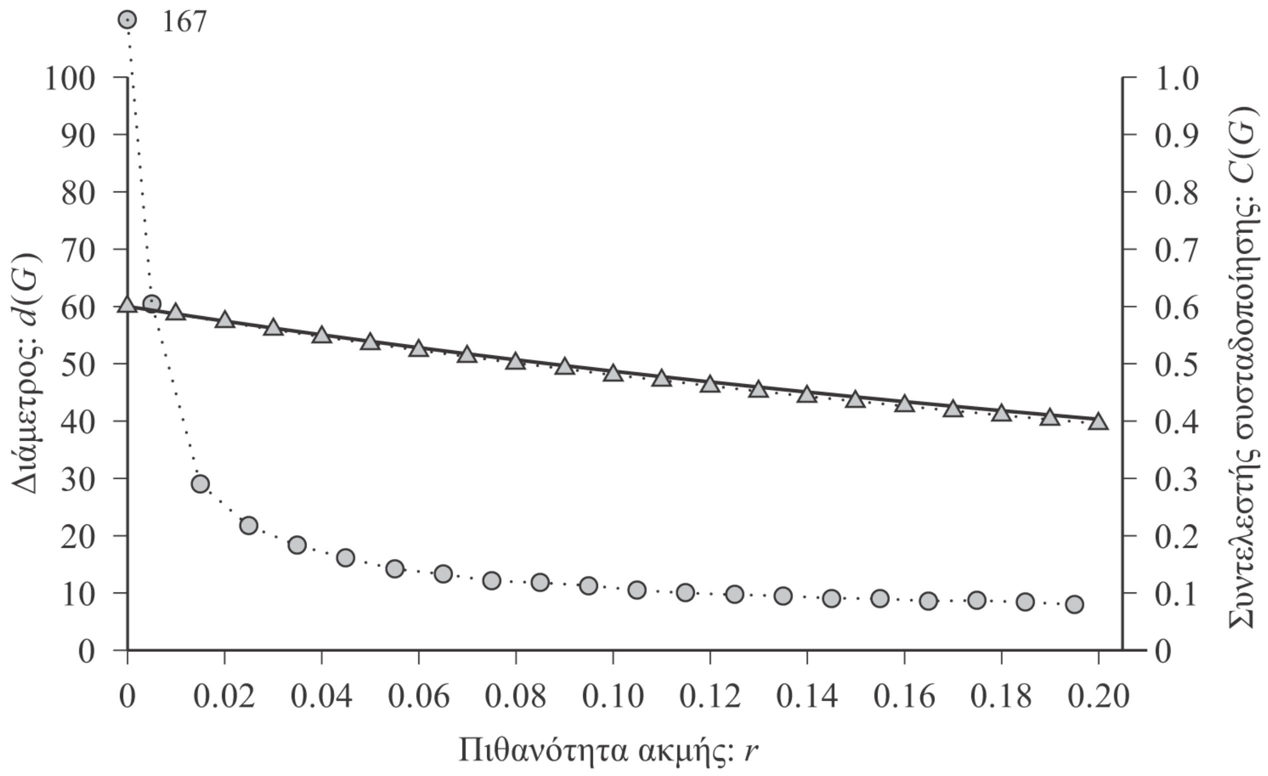
Ιδιότητες των γραφημάτων Watts-Strogatz

Κατανομή βαθμών: Η κατανομή βαθμών δεν συμμορφώνεται με κάποια σχέση δυνάμεων.

Συντελεστής συσταδοποίησης: Για μικρές τιμές του r, ο συντελεστής συσταδοποίησης παραμένει υψηλός.

Διάμετρος: Ακόμα και μικρές τιμές της πιθανότητας r μιας ακμής συντόμευσης είναι αρκετές για να μειωθεί η διάμετρος από $O(n)$ σε $O(\log n)$.

Μοντέλο των Watts-Strogatz: Διάμετρος (κύκλοι) και συντελεστής συσταδοποίησης (τρίγωνα)



Το μοντέλο ανεξαρτήτου κλίμακας των Barabási-Albert

Το μοντέλο των Barabási-Albert (BA) δίνει μια ανεξάρτητη από την κλίμακα κατανομή βαθμών, βασιζόμενο στην έννοια της *προνομιακής προσάρτησης*· δηλαδή, οι ακμές που ξεκινούν από τη νέα κορυφή είναι πιο πιθανό να ενώνονται με κόμβους υψηλότερων βαθμών.

Έστω ότι τη χρονική στιγμή t, το γράφημα συμβολίζεται με G_t, το πλήθος των κόμβων του συμβολίζεται με n_t, και το πλήθος των ακμών του συμβολίζεται με m_t.

Καθορισμός αρχικής κατάστασης: Το μοντέλο BA ξεκινά με ένα αρχικό γράφημα G_0, όπου κάθε κόμβος είναι συνδεδεμένος με τον αριστερό και τον δεξιό γείτονά του σε κυκλική διάταξη. Επομένως, ισχύει $m_0 = n_0$.

Ανάπτυξη και προνομιακή προσάρτηση: Το μοντέλο BA παράγει ένα νέο γράφημα G_{t+1} από το G_t προσθέτοντας ακριβώς έναν νέο κόμβο u, καθώς και $q \leq n_0$ νέες ακμές από τον u προς q διαφορετικούς κόμβους $v_j \in G_t$, όπου η πιθανότητα $\pi_t(v_j)$ να επιλεγεί ο κόμβος v_j είναι ανάλογη του βαθμού του στο G_t και δίνεται από τη σχέση

$$\pi_t(v_j) = \frac{d_j}{\sum_{v_i \in G_t} d_i}$$

Γράφημα των Barabási-Albert

Στη χρονική στιγμή $t = 0$, το γράφημα έχει αρχικά 3 πλήρως συνδεδεμένες κορυφές, v_0, v_1 και v_2 (γκρίζο χρώμα).

Στη χρονική στιγμή $t = 1$, προστίθεται η κορυφή v_3, με ακμές προς τις v_1 και v_2, οι οποίες επιλέγονται σύμφωνα με την κατανομή

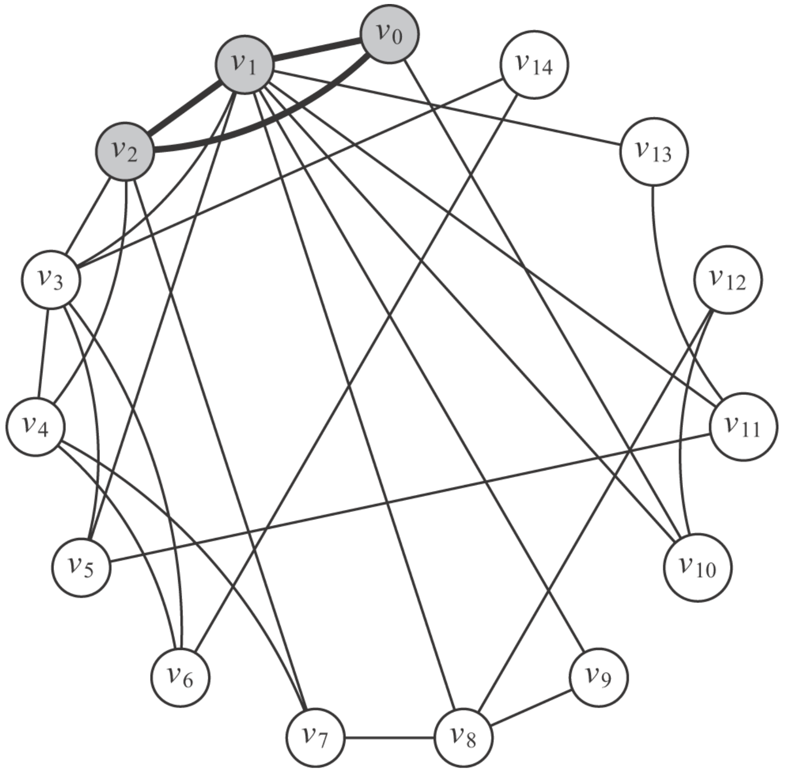
$$\pi_0(v_i) = 1/3 \text{ για } i = 0, 1, 2$$

Στη χρονική στιγμή $t = 2$, προστίθεται η κορυφή v_4. Επιλέγονται κατά προτίμηση οι κορυφές v_2 και v_3 σύμφωνα με την κατανομή πιθανότητας

$$\pi_1(v_0) = \pi_1(v_3) = \frac{2}{10} = 0.2$$

$$\pi_1(v_1) = \pi_1(v_2) = \frac{3}{10} = 0.3$$

$$n_0 = 3,\ q = 2,\ t = 12$$



Ιδιότητες των γραφημάτων BA

Κατανομή βαθμών: Για σταθερό *q* και μεγάλο *k, τ*ο μοντέλο BA παράγει μια κατανομή βαθμών που περιγράφεται από μια σχέση δυνάμεων με $\gamma = 3$, ειδικά για μεγάλους βαθμούς.

Διάμετρος: τα γραφήματα εμφανίζουν συμπεριφορά *πολύ μικρού κόσμου* για $q > 1$.

Συντελεστής συσταδοποίησης: Ο αναμενόμενος συντελεστής συσταδοποίησης των γραφημάτων BA είναι παρόμοιος με τον συντελεστή συσταδοποίησης για τυχαία γραφήματα.

Μοντέλο των Barabási-Albert: Κατανομή βαθμών

$n_0 = 3,\ t = 997,\ q = 3$