

1^η Εργαστηριακή άσκηση

Ανάλυση Κοινωνικών Δικτύων (Social Network Analysis)

Συμεών Παπαβασιλείου (papavass@mail.ntua.gr)

Ειρήνη Κοιλανιώτη (eirinikoilanioti@mail.ntua.gr)

Ιωάννης Τζανεττής (gtzane@gmail.com)

Βασίλειος Καρυώτης (vassilis@netmode.ntua.gr)

Κωνσταντίνα Σακκά (nsakka@cn.ntua.gr)

Γεώργιος Καλλίτσης (giorgoskallitsis99@gmail.com)

Επισκόπηση

- Κατασκευή και οπτικοποίηση σύνθετων τύπων δικτύων
- Μετρικές δικτύου
 - Συντελεστής ομαδοποίησης (clustering coefficient)
 - Μήκος ελάχιστου μονοπατιού
 - Εκκεντρότητα (eccentricity) κόμβου
 - Διάμετρος, ακτίνα, περιφέρεια, κέντρο
- Μετρικές κεντρικότητας κόμβων
 - Κεντρικότητα βαθμού (degree)
 - Κεντρικότητα εγγύτητας (closeness)
 - Ενδιαμεσική κεντρικότητα (betweenness)
 - Κεντρικότητα Katz
 - Εφαρμογή PageRank σε πραγματικό δίκτυο
- Μελέτη συνεκτικότητας (connectivity) και ευρωστίας (robustness) δικτύων
- Μελέτη εξελικτικής μετατροπής δικτύου
- Μελέτη πραγματικών δικτύων

Τύποι σύνθετων δικτύων

- **Πλέγμα (REG)**
 - Πεπερασμένος γράφος $G(n,d)$
- **Τυχαίος γράφος**
 - Erdos-Renyi (RG-ER) $G(n,M)$
 - Gilbert (RG-G) $G(n,p)$
- **Τυχαίος γεωμετρικός γράφος (RGG)**
 - Επίπεδος $G(n,R,l)$
- **Scale-free (SF)**
 - Barabasi-Albert $G(n,d)$
- **Small-world (SW)**
 - Watts-Strogatz $G(n,d,g_p)$

Μετρικές δικτύου

- Συντελεστής ομαδοποίησης (Clustering coefficient)

τριγώνων που συμμετέχει ο κόμβος u / # τριπλετών με κέντρο τον κόμβο u

- Μήκος ελάχιστου μονοπατιού

- Στον γράφο $G(V,E)$ με $u,v \in V$, η απόσταση $\text{dist}(u,v)$ είναι το μήκος του ελάχιστου (u,v) -μονοπατιού στο G .
- Εκκεντρότητα κόμβου $u \in V$: $\text{ecc}(u) = \max_v \text{dist}(u,v)$
- Διάμετρος γράφου G : $\text{diam}(G) = \max_u \text{ecc}(u)$
- Ακτίνα γράφου G : $\text{rad}(G) = \min_u \text{ecc}(u)$
- Περιφέρεια γράφου G : $P = \{u \in V : \text{ecc}(u) = \text{diam}(G)\}$
- Κέντρο γράφου G : $C = \{u \in V : \text{ecc}(u) = \text{rad}(G)\}$

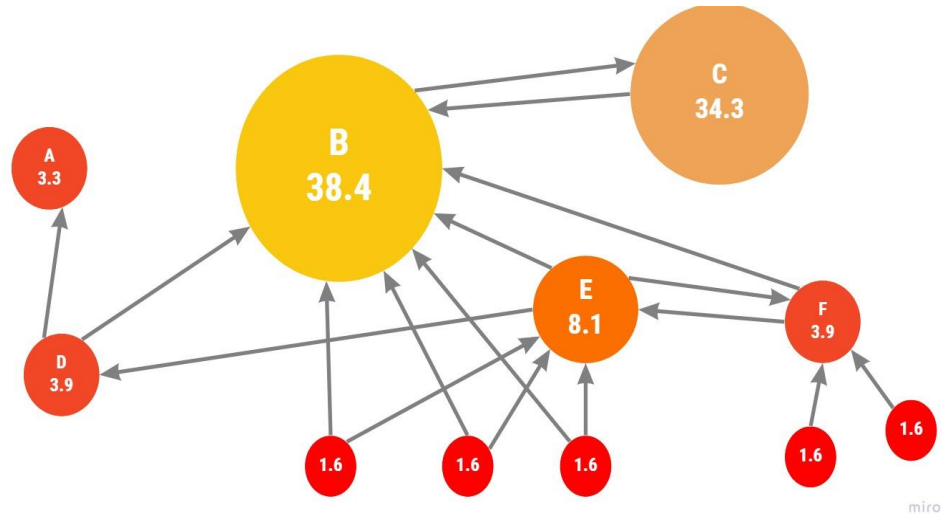
Μετρικές κεντρικότητας κόμβων

- Κεντρικότητα βαθμού
- Κεντρικότητα εγγύτητας
- Ενδιαμεσική κεντρικότητα
 - λαμβάνει υπόψη τα ελάχιστα μονοπάτια μεταξύ των κόμβων.
- Κεντρικότητα Katz
 - λαμβάνει υπόψη όλους τους δυνατούς περίπατους μεταξύ των κόμβων.
 - Για τον κόμβο i , η **κεντρικότητα Katz** δίνεται από τη σχέση:
$$x_i = \alpha \sum_j a_{ij} x_j + \beta, \text{ όπου } \mathbf{A}=(a_{ij}) \text{ ο πίνακας γειτνίασης του γράφου.}$$
Θεωρήστε: $\alpha=(\lambda_{\max})^{-1}-0.01$, $\beta=1$ όπου λ_{\max} η μεγαλύτερη ιδιοτιμή του \mathbf{A} .
- PageRank
 - αλγόριθμος της Google για την κατάταξη ιστοσελίδων: *πιο σημαντικές είναι οι ιστοσελίδες στις οποίες καταλήγει ένας χρήστης μέσω άλλων ιστοσελίδων.*

PageRank

Τυχαίος περίπατος στον κατευθυνόμενο γράφο $G(V, E)$, με V το σύνολο των ιστοσελίδων και E το σύνολο των υπερσυνδέσμων.

Ο χρήστης πλοηγείται σε ιστοσελίδες επιλέγοντας με πιθανότητα ϵ έναν από τους υπερσυνδέσμους της ιστοσελίδας που βρίσκεται και με πιθανότητα $1-\epsilon$ οποιαδήποτε διαθέσιμη ιστοσελίδα.



PageRank κόμβου u , $PR(u,k)$: η πιθανότητα, ξεκινώντας από οποιοδήποτε κόμβο, ο τυχαίος περίπατος να σταματήσει, μετά από k βήματα, στον κόμβο u .

Επαναληπτικός τρόπος υπολογισμού PageRank

1. Όλοι οι κόμβοι ξεκινούν με PageRank ίσο με $1/n$, όπου $n=|V|$.
2. Update Rule: Ο PageRank κάθε κόμβου επαναυπολογίζεται k φορές ως εξής:

$$PR(u,t) = (1 - \epsilon)/|V| + \epsilon \sum_{v:(v,u) \in E} PR(v,t-1) / d_{out}(v),$$

όπου $\epsilon \in [0,1]$, συνήθως λαμβάνει την τιμή 0.85 και $d_{out}(v)$ είναι ο έξω-βαθμός του κόμβου v , $t=1,...,k$.

Στα περισσότερα δίκτυα, ο PageRank συγκλίνει για $k \rightarrow \infty$ σε μια τιμή που εξαρτάται από το ϵ .

Μελέτη συνεκτικότητας και ευρωστίας δικτύων

- Ποσοστό συνεκτικότητας

$$\frac{\# \text{ συνδεδεμένων τοπολογιών}}{\# \text{ παραγόμενων τοπολογιών}}$$

- Μελέτη συνεκτικότητας

1. Κατασκευή τοπολογίας για διάφορες τιμές παραμέτρων
2. Έλεγχος συνεκτικότητας
3. Επανάληψη βημάτων (1), (2) ($k=100$)
4. Υπολογισμός ποσοστού συνεκτικότητας

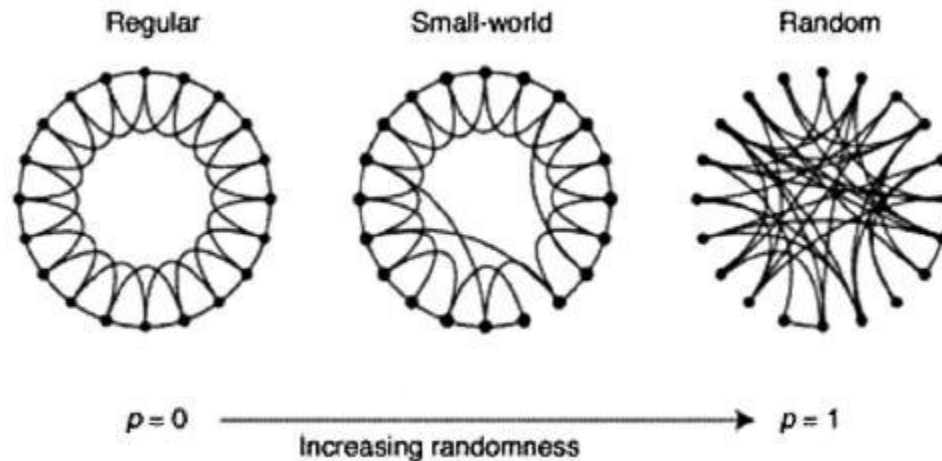
- Μελέτη ευρωστίας δικτύου

Υπολογισμός του ελάχιστου αριθμού κόμβων/ακμών σε ένα συνδεδεμένο δίκτυο, η αφαίρεση των οποίων έχει ως αποτέλεσμα μη συνδεδεμένο δίκτυο .

Μελέτη συνεκτικότητας

Τοπολογία	Εύρος Παραμέτρων
REG	$d \in [2, 10]$ με βήμα 2
RGER	$M \in [100, 800]$ με βήμα 100
RG-Gilbert	$p \in [0.01, 0.1]$ με βήμα 0.01
RGG	$R \in [0.025, 0.25]$ με βήμα 0.025
SF	$d \in [2, 10]$ με βήμα 2
SW	$d \in [2, 10]$ με βήμα 2 και $g_p \in [0.1, 0.7]$ με βήμα 0.1

Εξελικτική μετατροπή δικτύου



Μελέτη του εξελικτικού χαρακτήρα του μοντέλου *Watts – Strogatz* για διάφορες τιμές της πιθανότητας ανασύνδεσης των ακμών.

Πραγματικά δίκτυα

- Εκτέλεση **PageRank** σε μια συνδεδεμένη συνιστώσα του κατευθυνόμενου δικτύου web-Stanford: Οι κόμβοι αναπαριστούν ιστοσελίδες από τον ιστότοπο του [Stanford University](https://stanford.edu) (stanford.edu). Οι ακμές αναπαριστούν υπερσυνδέσμους μεταξύ τους.
(<https://snap.stanford.edu/data/web-Stanford.html>)
- Ανάλυση του δικτύου που σχηματίζεται από τις αλληλεπιδράσεις των χαρακτήρων της σειράς Game of Thrones στον πρώτο κύκλο επεισοδίων (<https://github.com/mathbeveridge/gameofthrones>).
Αναγνώριση του τύπου του υπό εξέταση δικτύου με βάση:
 - την κεντρικότητα του βαθμού κόμβων
 - τον συντελεστή ομαδοποίησης
 - το μέσο μήκος μονοπατιού