

Ανάλυση Κοινωνικών Δικτύων

2η Εργαστηριακή Άσκηση

Συμεών Παπαβασιλείου (papavass@mail.ntua.gr)

Βασίλειος Καρυώτης (vassilis@netmode.ntua.gr)

Γιώργος Μήτσης (gmitsis@netmode.ntua.gr)

Μαργαρίτα Βιτοροπούλου (mvitoropoulou@netmode.ntua.gr)

Κωνσταντίνος Τσιτσεκλής (ktsitseklis@netmode.ntua.gr)

Κωνσταντίνα Σακκά (nsakka@cn.ntua.gr)

Ανάλυση Κοινοτήτων σε Τεχνητές και Πραγματικές τοπολογίες

Στην άσκηση θα μελετηθούν:

- Χαρακτηριστικά κόμβων πραγματικών τοπολογιών με μετρικές που χρησιμοποιήθηκαν στην 1^η άσκηση.
- Εντοπισμός κοινοτήτων με τρεις αλγορίθμους:
 - Spectral Clustering
 - Newman-Girvan
 - Modularity Maximization
- Αξιολόγηση διαμέρισης με την μετρική modularity.

Πραγματικά δίκτυα

Τρία διαφορετικά πραγματικά δίκτυα:

- American College Football
- Les Miserables
- Dolphin Social Network

Πληροφορίες για τα δίκτυα:

<http://www-personal.umich.edu/~mejn/netdata/>

Μετρικές που θα χρησιμοποιηθούν:

- **Βαθμός** κόμβου - μέσος βαθμός τοπολογίας
- **Clustering coefficient** κόμβου - μέσο clustering coefficient τοπολογίας
- **Ego Betweenness Centrality** κόμβου - μέσο ego betweenness τοπολογίας

Τα αποτελέσματα θα πρέπει να συγκριθούν με τα αντίστοιχα για τις συνθετικές τοπολογίες της προηγούμενης άσκησης.

Community detection (Ανίχνευση Κοινοτήτων)

- Ο τρόπος με τον οποίο εξελίσσονται τα σύνθετα δίκτυα οδηγεί στην δημιουργία κοινοτήτων εντός τους.
- *Κοινότητες είναι ομάδες κόμβων που έχουν κοινές ιδιότητες ή/και έχουν κοινό ρόλο στο δίκτυο (π.χ. κόμβοι που αλληλεπιδρούν μεταξύ τους πιο συχνά σε σχέση με άλλους κόμβους, συνιστούν μια κοινότητα)*
- Πολλές μέθοδοι για την ανίχνευση κοινοτήτων: Θα εξετάσουμε τρεις.
- Αξιολόγηση διαμέρισης με την μετρική της αρθρωτότητας (modularity).

Modularity (Αρθρωτότητα)

- Χρήσιμη μετρική για την αξιολόγηση της διαμέρισης ενός δικτύου σε κοινότητες (όχι η μοναδική όμως!).
- Έστω δίκτυο με m ακμές και διαμέρισή του σε k κοινότητες. Ο πίνακας γειτνίασης είναι ο A και ο αναμενόμενος αριθμός ακμών μεταξύ δυο κόμβων: $d_i d_j / 2m$

$$Q = \frac{1}{2m} \sum_{\ell=1}^k \sum_{i \in C_\ell, j \in C_\ell} (A_{ij} - d_i d_j / 2m)$$

- Το modularity υπολογίζει το πόσες ακμές υπάρχουν σε μια κοινότητα σε σχέση με το πόσες θα υπήρχαν αν είχαν κατανεμηθεί τυχαία μεταξύ των κόμβων της. Τιμή μεταξύ $[-1, 1]$.

Spectral Clustering

- Χρησιμοποιεί τις ιδιοτιμές και τα ιδιοδιανύσματα του Laplacian πίνακα:

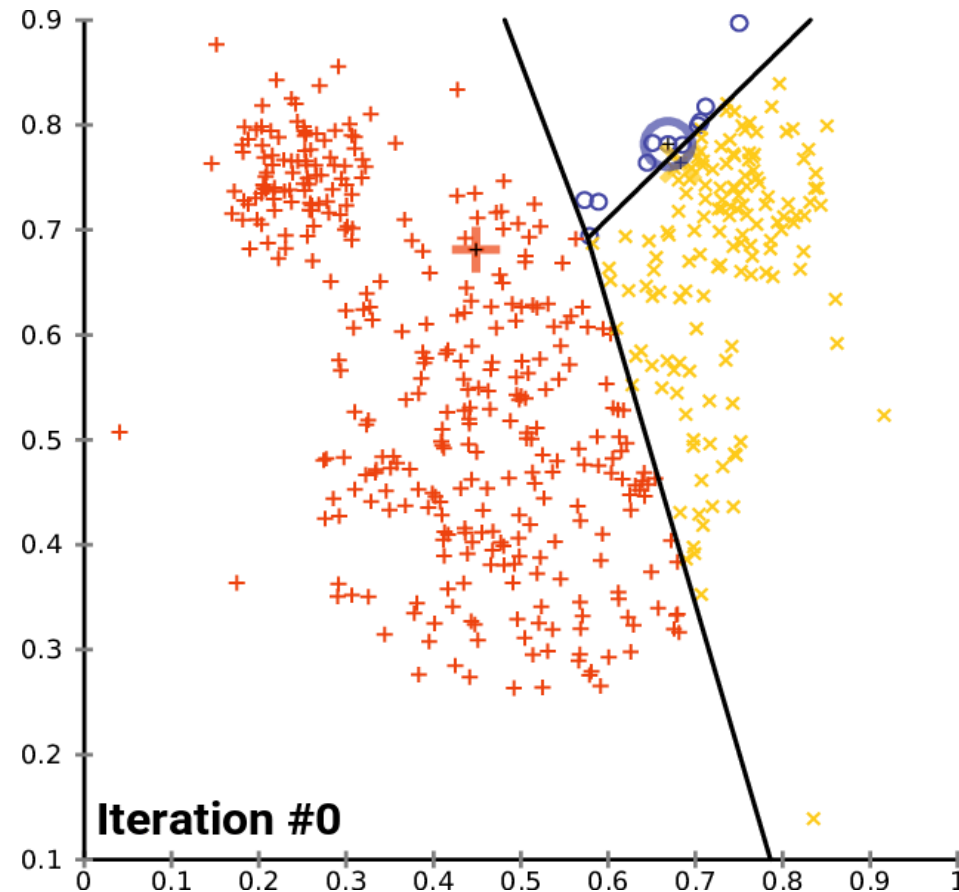
$$\mathbf{L} = \mathbf{D} - \mathbf{A}$$

- Ιδιοτιμές χρήσιμες για ανακάλυψη ιδιοτήτων που αφορούν στην ομαδοποίηση (πλήθος συνδεδεμένων συνιστωσών, πυκνότητα δικτύου κλπ).
- Ιδιοδιανύσματα χρησιμοποιούνται ως συντεταγμένες για τους κόμβους. Το πρόβλημα μεταφέρεται σε ανακάλυψη ομάδων σε σημεία ενός χώρου διάστασης ίσης με το πλήθος των κοινοτήτων.
- Χρήση αλγορίθμου ομαδοποίησης (k-means) και αντιστοίχιση σε κοινότητες.

K-means

- Ομαδοποίηση σημείων σε χώρο n διαστάσεων.
- Τοποθέτησε k αυθαίρετα σημεία (centroids) μέσα στον χώρο (αντιπρ. k κοινότητες).
- Τα σημεία ανατίθενται στην κοινότητα με το centroid της οποίας έχουν την ελάχιστη απόσταση.
- Υπολόγισε τα νέα centroids: ο αριθμητικός μέσος των σημείων της κοινότητας.
- Επανάλαβε μέχρις ότου να μην γίνονται αλλαγές στις κοινότητες.

K-means Παράδειγμα



Newman-Girvan

- Χρησιμοποιεί την μετρική Edge Betweenness Centrality (EBC). Είναι ανάλογη της μετρικής Betweenness Centrality για τους κόμβους. Πιο κεντρική ακμή αυτή που έχει το μεγαλύτερο ποσοστό συντομότερων μονοπατιών που την χρησιμοποιούν.
- Τέτοιες ακμές είναι συχνά "γέφυρες", δηλ, συνδέουν κόμβους που ανήκουν σε διαφορετικές κοινότητες. Η αφαίρεσή τους αποσυνδέει τον γράφο.
- Ο αλγόριθμος βρίσκει τέτοιες ακμές και τις αφαιρεί έως ότου προκύψουν είτε όλες οι διαμερίσεις (κάθε κόμβος αποτελεί κοινότητα) ή μέχρι ενός προκαθορισμένου σημείου.

Modularity Maximization

- Ζητούμενη η διαμέριση σε κοινότητες που θα παρήγαγε την μεγαλύτερη τιμή modularity.
- Το πρόβλημα αυτό είναι NP-hard. Άρα, χρησιμοποιούνται ευριστικές/προσεγγιστικές μέθοδοι.
- Μέθοδος Clauset-Newman-Moore: Όλοι οι κόμβοι ξεκινούν σαν ξεχωριστές κοινότητες. Σε κάθε ζεύγος κόμβων εξετάζεται το αν η τοποθέτηση τους στην ίδια κοινότητα αυξάνει το modularity. Αν ναι, τοποθετούνται στην ίδια κοινότητα. Όταν πάνω από ένας κόμβος υπάρχει σε μια κοινότητα, αντικαθίσταται από "υπερκόμβο" (supernode).
- Μετά από έναν αριθμό επαναλήψεων, δεν αλλάζει πλέον το modularity.

Ζητούμενο

- Να εκτελέσετε τους 3 αλγορίθμους community detection για όλες τις τοπολογίες (πραγματικές και συνθετικές).
- Να οπτικοποιήσετε τις διαμερίσεις.
- Να σχολιάσετε τα αποτελέσματά σας.

Όλοι οι αλγόριθμοι υπάρχουν υλοποιημένοι στην βιβλιοθήκη networkx!

The end...

Απορίες;