

## Εργασία 5: Node Embeddings

Για την εργασία σας, θα χρησιμοποιήσετε το Books about US Politics Dataset (<http://networkdata.ics.uci.edu/data/polbooks/>).

Εφαρμόστε όλες τις μεθόδους στην μεγαλύτερη συνεκτική συνιστώσα του γράφου.

Παράγετε node embeddings (διανυσματικές αναπαραστάσεις) χρησιμοποιώντας:

- Node2Vec με  $q=2$  και  $p=1$
- Node2Vec με  $q=0.5$  και  $p=1$
- Node2Vec με  $q=1$  και  $p=1$

Σε όλα τα πειράματα δώστε διάσταση embedding 64 (ή 32 αν το θεωρείτε καλύτερο).

Για τις υπόλοιπες παραμέτρους χρησιμοποιήστε είτε τις default τιμές, είτε προσαρμόστε τις όπως σας εξυπηρετεί για καλύτερη ταχύτητα ή καλύτερα αποτελέσματα.

Σημείωση: Χρησιμοποιήστε τις ίδιες παραμέτρους σε όλες τις εκτελέσεις της μεθόδου.

Για να αποτιμήσετε την απόδοση των embeddings θα τα χρησιμοποιήσετε στα εξής προβλήματα:

1. **Link Prediction:** Κάντε όποια μοντελοποίηση θέλετε και χρησιμοποιείτε όποιον κατηγοριοποιητή θέλετε. Μετρήστε accuracy/precision/recall.
2. **Clustering:** Εφαρμόστε K-Means στα embeddings. Θεωρήστε  $K=3$  και μετρήστε modularity, και purity. Χρησιμοποιήστε την μέθοδο t-SNE (<https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html>) για την οπτικοποίηση των διανυσματικών αναπαραστάσεων των κόμβων στις δύο διαστάσεις. Χρησιμοποιήστε την οπτικοποίηση για να συγκρίνετε τα αποτελέσματα του clustering με την πραγματική κατηγορία των κόμβων.

Παρουσιάστε και σχολιάστε τα αποτελέσματά σας.