Εργασία για το μάθημα Θεωρία Δικτύων - Δεκέμβριος 2022

Η εργασία αφορά την **εύρεση θεματικών περιοχών** για περίπου 20Κ επιστημονικά άρθρα των οποίων δίνονται ο τίτλος και η περίληψη. Τα δεδομένα είναι στο αρχείο train.csv όπου έχει την μορφή

ID,TITLE,ABSTRACT,Computer Science,Physics,Mathematics,Statistics,Quantitative Biology,Quantitative Finance

όπου ένα άρθρο θα έχει 1 ή 0 αναλόγως εάν ανήκει στην αντίστοιχη θεματική περιοχή ή όχι.

Θα πρέπει να σχηματίσετε ένα δίκτυο όπου κάθε κορυφή θα αντιστοιχεί σε ένα άρθρο και κάθε ακμή θα περιέχει μια τιμή s(i,j) στο [0,1] αναλόγως πόσο *σχετικά* είναι δύο άρθρα i και j μεταξύ τους, όπου με 1 συμβολίζουμε την μεγαλύτερη συσχέτιση. Θέλουμε να εξετάσουμε το κατά πόσο οι κοινότητες σε αυτό το δίκτυο αντιστοιχουν στις θεματικές ενότητες που δίνονται.

Για τον υπολογισμό της συσχέτισης s(i,j) μπορείτε να χρησιμοποιήσετε το παρακάτω μοντέλο γλώσσας που θα σας δώσει ενβυθίσεις προτάσεων

https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

όπως και οτιδήποτε άλλο μοντέλο που θα κάνει χρήση και τον τίτλο ή/και το κείμενο της περίληψης (ή κάποια τροποποίηση).

Για τα αποτελέσματα σας θα πρέπει να υπολογίσετε και την ακρίβεια σε σχέση με τις δοσμένες θεματικές κατηγορίες.

Θα πρέπει να χρησιμοποιήσετε τα παρακάτω:

- **networkx** για τον χειρισμού του δικτύου
- **gephi** για την οπτική απεικόνιση των αποτελεσμάτων
- pandas