

Report delle attività svolte

- Weighted Network Analysis

Ho svolto un'analisi (superficiale) del network pesato contenuto nel file *weighted_academic_graph.txt*. Nel network, ogni nodo rappresenta uno stato, mentre ogni arco rappresenta uno spostamento da uno stato ad un altro. Inoltre, ogni arco contiene informazioni riguardo all'anno in cui lo spostamento è avvenuto e a quante volte in quell'anno è stato condotto uno spostamento tra i due stati. Quest'ultima informazione rappresenta la variabile peso caratterizzante il network. Eseguendo il notebook *weighted_network_analysis.ipynb* è possibile osservare la mia analisi, nel quale conduco uno studio sulla distribuzione dei gradi generali, in entrata e in uscita dei nodi componenti il network, sulla connectedness del network, sulla distribuzione del clustering coefficient tra i nodi componenti il network e sulla lunghezza dei cammini.

- Estrazione dati dai dump di Zenodo

Ho estratto e processato i dati contenuti nei dump di Zenodo (reperibili [qui](#)). Tali dati sono principalmente di due tipi, uno riguardante le organizzazioni coinvolte nel finanziamento di attività inerenti alla ricerca, e l'altro riguardante i progetti presentati all'interno di un determinato framework. Le organizzazioni possono essere di varia natura, come, per esempio, università o finanziatori privati. Il procedimento di estrazione consiste in uno script Python (*zenodo_parser.py*) che decompime, decripta e converte i record contenuti nei dump in un unico file XML. Successivamente, tramite un secondo script Python (*zenodo_xml_miner.py*), i dati più interessanti contenuti nel file XML vengono selezionati e salvati in un file JSONL. I dati relativi alle organizzazioni consistono in una divisione per Paesi degli ID caratterizzanti ogni organizzazione, mentre quelli relativi ai framework sono organizzati in modo tale che ogni riga contiene le informazioni relative a un determinato ricercatore comparso all'interno del framework come, per esempio, nome, cognome, id MAG, ORCID o URL (se presenti), e una lista dei progetti al quale egli ha partecipato. Ogni progetto è individuato da un nome, ossia il titolo che è stato dato alla sua pubblicazione, e una serie di keywords. Per motivi di spazio, i file JSONL risultanti dall'intero processo vengono salvati su un'apposita macchina remota, ospitata dal CNR.

- Analisi del Coauthors Network

Per ognuno dei network pesati dei coautori, che consistono in network non diretti nei quali i pesi rappresentano il numero di volte che un coppia di ricercatori hanno collaborato durante un particolare anno, ho calcolato le statistiche di base per quanto riguarda le analisi delle reti. Ho riportato in dei CSV le informazioni per quanto riguarda il numero di nodi, il numero di archi, il coefficiente di clustering medio, la transitivity, il diametro e il raggio. Oltre a queste statistiche di

base, ho fornito anche delle visualizzazioni riguardo la distribuzione dei gradi dei nodi, la distribuzione dei pesi degli archi, la degree centrality dei nodi, la closeness centrality dei nodi e la betweenness dei nodi.

- Costruzione dei Networks a partire dai Dump di Zenodo

work in progress