

Social Network Analysis

Final Project

Giulio Rossetti, Chiara Boldrini, Erica Cau

Introduzione

Il progetto finale del corso di Social Network Analysis ha l'obiettivo di applicare le conoscenze e le competenze acquisite durante il corso per analizzare una rete sociale reale.

Questo progetto permetterà di comprendere meglio le dinamiche delle relazioni sociali e le strutture di rete in un contesto specifico. Il contesto scelto può variare in base agli interessi degli studenti e può includere settori come il business, la comunicazione, la salute, l'educazione, la politica, o le reti online.

Il progetto si compone di quattro parti, di cui tre vincolate ed una a scelta degli studenti.

Struttura del progetto

1. **Identificare il dataset di partenza**

Selezionare una rete di interesse dal sito [networkrepository](#), definendone il contesto latipologia dei nodi (individui, organizzazioni, ecc.) e delle connessioni (relazioni, interazioni, ecc.).

La rete scelta deve essere composta da almeno 15-20.000 nodi.

2. **Analizzare la struttura della rete**

Utilizzare le misure di rete introdotte nella prima parte del corso (distribuzione del grado, densità, clustering coefficient, centralità, assortatività...) per caratterizzare il dataset selezionato.

3. **Comparazione della rete con modelli nulli**

Confrontare la statistiche calcolate sulla rete selezionata con i valori attesi nei principali modelli nulli visti a lezione (ER, BA). Discutere ed interpretare le differenze/similarità osservate.

4. **Task di analisi** (a scelta degli studenti)

Selezionare N task di analisi tra quelli proposti (dove N è il numero dei componenti del gruppo meno 1), applicarli alla rete selezionata, e discutere i risultati ottenuti.

Task di analisi

1. **Community detection:**

Selezionare 3 algoritmi tra quelli visti a lezione e, sfruttando le implementazioni fornite da CDlib, estrarre le relative partizioni del grafo selezionato. Fornire una comparazione degli algoritmi selezionati ed una valutazione/caratterizzazione delle comunità identificate.

2. **Community detection 2:**

In caso di disponibilità di informazioni temporali legate agli archi, creare una rappresentazione dinamica della rete selezionata tramite snapshot graphs, estrarre le comunità da ciascuno snapshot utilizzando un algoritmo di CD a scelta e studiare gli eventi di comunità. Discutere i risultati ottenuti.

3. **Epidemic Modeling:**

Simulare i modelli diffusivi SI/SIS/SIR al grafo selezionato (utilizzando NDlib) e studiarne i trend al variare del valore dei parametri caratterizzanti ciascun modello. Discutere i risultati ottenuti e fornire, ove possibile, la descrizione di un potenziale fenomeno - relativo ai dati selezionati - che possa essere catturato da ciascun modello.

4. **Opinion Dynamics:**

Simulare i modelli di opinion dynamics (a classi discrete) Voter/Majority Rule/Snajdz/Q-Voter al grafo selezionato (utilizzando NDlib) e studiarne i trend al variare del valore dei parametri caratterizzanti ciascun modello. Discutere i risultati ottenuti e fornire, ove possibile, la descrizione di un potenziale fenomeno - relativo ai dati selezionati - che possa essere catturato da ciascun modello.

5. **Opinion Dynamics 2:**

Simulare il modello di opinion dynamics (a classi continue) Algorithmic Bias su un sample (connesso) di 1000 nodi del grafo selezionato (utilizzando NDlib) e studiarne i trend al variare del valore dei parametri epsilon e gamma. Discutere i risultati ottenuti e fornire, ove possibile, la descrizione di un potenziale fenomeno - relativo ai dati selezionati - che possa essere catturato da diverse configurazioni del modello.

6. **Robustezza:**

Realizzare (almeno) 5 simulazioni tra random failures e attacchi a nod/archi finalizzati alla distruzione del grafo. Ogni strategia deve essere basata su un diverso criterio di selezione e ranking dei nodi/archi da rimuovere. Discutere i risultati ottenuti e valutare la resilienza del grafo in termini di dimensione della componente gigante.

Esame

L'esame orale verterà sulla presentazione dei risultati ottenuti in ciascuna delle analisi svolte dal gruppo. Ciascun membro del gruppo dovrà presentare e commentare - in modo succinto - i risultati di uno specifico task.

La presentazione dovrà essere accompagnata da slide contenenti statistiche e visualizzazioni necessarie alla discussione. La visione dei notebook contenenti il codice sviluppato per svolgere i task di progetto potrà essere richiesta in sede di esame per chiarimenti su punti specifici dell'analisi.

A ciascun gruppo sarà allocato uno slot di 10-15 minuti per presentare il lavoro svolto.