

Primo Progetto per il corso di Social Computing

A.A. 2021/22

10 Novembre 2021

Cosa fare

Costruite un sottografo del grafo di Twitter nel modo seguente. Per ciascuno dei seguenti due account (i dati numerici potrebbero subire minime variazioni):

- @mizzaro - 162 Follower - 356 Following
 - @Miccighel_ - 330 Follower - 214 Following
1. Usando Tweepy scaricate 5 utenti follower (*follower di primo livello*) e 5 utenti following (*following di primo livello*), ciascuno con meno di 1000 following/follower
 2. Scaricate poi 5 utenti follower e 5 utenti following, sempre con meno di 1000 following/follower, per ciascuno di quelli scaricati durante il punto precedente (*followers e following di secondo livello*)
 3. Scaricare i dettagli del profilo dei followers di primo e secondo livello
 4. Scaricate gli ultimi 5 tweet di ciascun follower di primo e secondo livello
 5. Costruite la rete sociale (grafo):
 - Inserite l'*id* di ciascun follower di primo e secondo livello come identificatore del nodo
 - Costruite le relazioni di *follower* e *following* rispetto a ciascun utente; per fare ciò dovrete verificare l'esistenza delle relazioni scaricando ed esplorando la lista di followers e following di ciascun utente (terzo livello)
 - Inserite un attributo con valore "*following*" o "*follower*" per ciascun arco
 - Inserite i dettagli del profilo di ciascun utente come *attributi del nodo*
 - Inserite la collezione di tweet scaricata per ciascun utente come *attributo* del nodo
 - Inserite i membri del vostro gruppo come *attributi del grafo*
 - Per ogni nodo, aggiungete le *metriche pubbliche* dell'utente come attributi
 6. Costruite:
 - Un sottografo delle relazioni *follower*
 - Un sottografo delle relazioni *following*
 7. Producete una visualizzazione interattiva del grafo e dei due sottografi usando pyvis
 8. Verificate se i grafi:
 - sono fortemente/debolmente connessi
 - sono bipartiti
 9. Calcolate e disegnate la distribuzione dei gradi dei grafi
 10. Calcolate il coefficiente di clustering dei grafi
 11. Misurate le seguenti distanze sui grafi:
 - Centro
 - Raggio
 - Distanza Media
 - Distanza Massima
 12. Calcolate le seguenti misure di centralità sui grafi:
 - Betweenness centrality
 - Closeness centrality
 - Degree centrality
 - In-degree centrality
 - Out-degree centrality
 - Page Rank
 - HITS
 13. Calcolate i seguenti coefficienti per stimare la "small-world-ness" dei grafi:
 - Coefficiente omega

- Coefficiente sigma
- 14. Calcolare le correlazioni di Pearson e Kendall fra le misure di centralità; riportare il risultato in due tabelle

Come consegnare

1. Dovete costituire dei gruppi che **devono** essere formati da **quattro** persone (i gruppi più o meno numerosi verranno penalizzati)
2. Si devono consegnare i seguenti elementi:
 - Relazione di al massimo 5 pagine (con anche i vostri nomi cognomi e numeri di matricola) che descrive tutto il lavoro svolto
 - Una serializzazione dei dati scaricati mediante API di Twitter con le funzioni spiegate a lezione (in un unico file)
 - Una cartella graphs/ contenente una serializzazione di ciascun grafo prodotta con le funzioni spiegate a lezione (tre file)
 - Il codice prodotto (in un unico notebook)
 - Una cartella html/ contenente le visualizzazioni interattive prodotte mediante pyvis
3. Consegnate via mail a entrambi i docenti (un unico messaggio indirizzato a entrambi)
 - mizzaro@uniud.it
 - michael.soprano@uniud.it
 - oggetto della mail nel formato:
[Progetto SocCom 1] cognome1_cognome2_cognome3_cognome4
 - in allegato alla mail un unico file zippato che quando scompattato produce una singola cartella con nome cognome1_cognome2_cognome3_cognome4
4. **Scadenza: Lunedì 6 Dicembre 2021 AoE Timezone**
5. Punteggio:
 - 5 punti in trentesimi per i migliori 20%,
 - 4 punti per i seguenti 20%,
 - 3 punti per i seguenti 20%,
 - 2 punti seguenti 20%,
 - 1 punto per i seguenti 20%,
 - 0 punti a discrezione dei docenti per progetti non adeguati o per chi non consegna

Informazioni aggiuntive

- Viste le limitazioni poste sugli endpoint da Twitter, durante la prima fase potete parallelizzare il lavoro (ogni componente del gruppo scarica dati da un account alla volta), serializzate e deserializzate a posteriori
- Nello scaricare following, followers, tweet o quant'altro fate attenzione a non utilizzare account con profilo protetto (field *protected* dell'oggetto *User*)
- Sfruttate quando possibile e/o sensato le funzionalità degli endpoint V2 di Twitter spiegate a lezione (fields, expansions, paginazione, etc.)
- Tutte le misure, proprietà e verifiche richieste su grafi sono definite nella documentazione di NetworkX; consultatela per capire come rispondere alle varie richieste
- Conservate codice e dati prodotti in vista del secondo progetto