

Primo Progetto di Social Computing A.A. 2022/23

11 Novembre 2022

Consegna

Costruire un sottografo del grafo di Twitter nel modo seguente.

1. Usando Tweepy scaricare tutti i follower del profilo Twitter con username *@KevinRoitero* (ad oggi 131 Follower). Per ciascuno di questi follower scaricare:
 - 1.1. Attributi di default
 - 1.2. Descrizione del profilo
 - 1.3. Metriche pubbliche dell'account (followers_count, following_count, tweet_count, listed_count)
 - 1.4. Se l'account è protetto (protected)
 - 1.5. Salvare una unica serializzazione JSON dei risultati recuperati.
2. Per ciascun follower del profilo *@KevinRoitero* utilizzare Tweepy per scaricare il numero di tweet pubblicati da quel profilo durante l'ultima settimana ed aggiungere tale informazione al JSON prodotto al punto 1.
3. Per ciascuno dei follower di *@KevinRoitero* che hanno almeno un follower e che non sono account protected scaricare 1000 follower salvando per ciascuno di essi le stesse informazioni dettagliate al punto 1: descrizione del profilo, metriche pubbliche dell'account e se l'account è protetto. (Se un profilo ha meno di 1000 follower, scaricarli tutti; se ne ha di più, limitarsi a 1000). Serializzare in locale il JSON risultante, che avrà una struttura simile al seguente esempio:

```
{
  id : {
    'name': 'Nome Cognome',
    'username': 'username',
    'description': 'Descrizione del profilo',
    'public_metrics': {
      'followers_count': 0,
      'following_count': 0,
      'tweet_count': 0,
      'listed_count': 0
    },
    'protected': Bool,
    'last_week_tweets_count': 0,
    followers = [{...}]
  },
  id : {...},
  ...
}
```

4. Utilizzando NetworkX costruire la rete sociale diretta (grafo diretto) dove:
 - 4.1. I nodi sono *KevinRoitero* ed i suoi follower con id del nodo uguale all'id del profilo utente e ciascun nodo ha come attributi username, la descrizione, ed il numero di follower di quel profilo.
 - 4.2. Utilizzando il JSON prodotto al punto 3 aggiungere al grafo un arco per ogni coppia di profili per cui esiste una relazione di following tra i due profili nei dati recuperati al punto 3.

Es. Siano A e B due profili follower di *@KevinRoitero* appartenenti al grafo. Aggiungo un arco (A, B) al grafo se e solo se nel JSON prodotto al punto 3 trovo che A è follower di B.

NOTA: In questo modo il grafo prodotto sarà subottimale e non conterrà tutte le relazioni di follow tra i follower di *@KevinRoitero*. Inoltre, gruppi diversi potrebbero ottenere grafi con archi diversi, a seconda dei 1000 profili scaricati per ogni follower al punto 3.

5. Generare un secondo grafo, nel modo seguente. Utilizzando la funzione apposita di NetworkX, trasformare il grafo prodotto al punto 3 in indiretto. Utilizzare poi il metodo del preferential attachment per aggiungere nodi e archi a questo grafo indiretto, fino ad ottenere un grafo con il doppio dei nodi rispetto al grafo di partenza. Ciascuno dei nuovi nodi dovrà avere 2 archi uscenti.
6. Produrre due visualizzazioni per ciascuno dei due grafi prodotti ai punti 3 e 5:
 - 6.1. Una interattiva utilizzando PyVis
 - 6.2. Una statica utilizzando NetworkX e il layout di Fruchterman Reingold in cui:
 - 6.2.1. Per la versione diretta del grafo la dimensione dei nodi dipende dal grado in ingresso del nodo
 - 6.2.2. Per la versione indiretta del grafo la dimensione dipende dal grado del nodo
7. Per ciascuno dei due grafi, identificare la più grande componente fortemente connessa SCC e produrre una visualizzazione statica del grafo con una colorazione rossa dei nodi appartenenti alla SCC, nera per gli altri.
8. Misurare le seguenti distanze sui due grafi (usando le opportune funzioni di NetworkX):
 - 8.1. Centro
 - 8.2. Raggio
 - 8.3. Distanza Media
 - 8.4. Distanza Massima
9. Calcolare le seguenti misure di centralità sui due grafi:
 - 9.1. Betweenness centrality
 - 9.2. Closeness centrality
 - 9.3. Degree centrality
 - 9.4. In-degree centrality
 - 9.5. Out-degree centrality
 - 9.6. Page Rank
 - 9.7. HITS
10. Calcolare i seguenti coefficienti per stimare la “small-world-ness” dei due grafi (usando le opportune funzioni di NetworkX):
 - 10.1. Coefficiente omega
 - 10.2. Coefficiente sigma

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 singola serializzazione** come da esempio dei dati scaricati mediante API di Twitter con le funzioni spiegate a lezione (in un unico file)
 - Una cartella `graphs/` contenente una serializzazione JSON di ciascun grafo prodotta con le funzioni spiegate a lezione (due file)
 - Il codice prodotto (in un **unico notebook** adeguatamente strutturato e commentato)

- Una cartella html/ contenente le visualizzazioni interattive prodotte mediante pyvis
- 3. Consegnate via mail a tutti i docenti (un unico messaggio)
 - david.labarbera@uniud.it
 - 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: Venerdì 9 Dicembre 2022 AoE Timezone**
- 5. Punteggio:
 - 5 punti in trentesimi per i migliori 25%,
 - 4 punti per i seguenti 25%,
 - 3 punti per i seguenti 20%,
 - 2 punti seguenti 15%,
 - 1 punto per i seguenti 15%,
 - 0 punti a discrezione dei docenti per progetti non adeguati o per chi non consegna

Informazioni aggiuntive

- Viste le limitazione poste sugli endpoint da Twitter, durante la prima fase è possibile se necessario parallelizzare il lavoro (ogni componente del gruppo scarica dati da un account alla volta), serializzando e deserializzando a posteriori
- I risultati scaricati vanno serializzati e consegnati in un singolo json formattato come da esempio
- Sfruttare 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, in particolare per i concetti non spiegati a lezione
- Conservate codice e dati prodotti in vista del secondo progetto