

Primo Progetto Social Computing

A.A. 2021/22

M. Giunta¹ S. Bibalo² G. Agrate³ G. De Nardi⁴

Dicembre 2021

¹Marco Giunta 147852 giunta.marco@spes.uniud.it

²Simone Bibalo 143184 bibalo.simone@spes.uniud.it

³Giulio Agrate 137617 agrate.giulio@spes.uniud.it

⁴Gianluca Giuseppe Maria De Nardi 142733@spes.uniud.it

Indice

| | | |
|----------|---|----------|
| 1 | Introduzione | 2 |
| 2 | Costruzione dei grafi | 2 |
| 2.1 | Estrazione e serializzazione dati | 2 |
| 2.2 | Costruzione grafo principale | 2 |
| 2.3 | Creazione sottografi | 2 |
| 2.4 | Interpretazione grafica | 3 |
| 3 | Analisi dei dati ottenuti | 3 |
| 3.1 | Proprietà grafi | 3 |
| 3.2 | Clustering | 3 |
| 3.3 | Misure di centralità | 3 |
| 3.4 | Small-world-ness | 5 |
| 3.5 | Analisi delle correlazioni di Pearson e Kendall | 5 |
| 4 | Conclusioni | 6 |

1 Introduzione

In questo progetto abbiamo costruito un grafo e due sottografi partendo da due account Twitter :

- @Mizzarro
- @Miccighel_

Il linguaggio di programmazione utilizzato per questo progetto è **Python** che, grazie alla libreria **Tweepy**, rende possibile manipolare in modo semplice le informazioni recuperate attraverso l'API di **Twitter**. Abbiamo deciso di utilizzare gli endpoint V2, in quanto mettono a disposizione tutte le funzionalità necessarie per il recupero dei dati richiesti.

2 Costruzione dei grafi

2.1 Estrazione e serializzazione dati

Come da specifiche, il primo passo è stato quello di scaricare 5 follower e 5 following dei due account principali (@Mizzarro e @Miccighel_). Grazie al fatto che non si volevano account con più di 1000 *follower/following*, non è stato necessario utilizzare la paginazione, visto che la quantità massima di dati scaricabili con un'unica interrogazione è proprio 1000. Questo ha reso possibile la semplificazione del codice all'interno del Notebook.

Successivamente abbiamo scaricato, per per ognuno dei *follower/following*, altri 5 follower e 5 following. Per cercare di ottenere il numero più alto possibile di account diversi, abbiamo evitato di scaricare *follower/following* già presenti, creando di fatto una lista univoca di account.

Infine, come indicato nelle specifiche del progetto, abbiamo selezionato **solo** i *follower*, ottenendo un totale di 112 account.

Per evitare di dover riscaricare ad ogni esecuzione del Notebook tutti i dati, viste le limitazioni degli account gratuiti usati per collegarsi alle API di Twitter, tutte le informazioni ottenute sono state serializzate e salvate all'interno di file *json*, separati per tipologia (*follower*, *following*, *tweets*).

2.2 Costruzione grafo principale

Per costruire il grafo principale abbiamo utilizzato il modulo **NetworkX** usando come nodi gli account. Con i dati in nostro possesso, abbiamo controllato l'esistenza di una relazione di *follower/following* tra tutti i nodi e creato gli archi corrispondenti all'interno del grafo ottenendo un totale di 391 archi.

2.3 Creazione sottografi

Dal grafo principale abbiamo creato due sottografi di tipo diretto, uno per i *follower* e uno per i *following*. Per ottenerli è stato sufficiente filtrare gli archi sul parametro **key**, valorizzato opportunamente durante la creazione del grafo principale.

2.4 Interpretazione grafica

Abbiamo utilizzato Pyvis per la parte grafica in modo da avere grafi interattivi e facilmente comprensibili. Per ottenere una migliore visualizzazione del grafo, in particolar modo per poter osservare separatamente le coppie di archi entranti ed uscenti da un singolo nodo, abbiamo usato la classe `MultiDiGraph`.

Purtroppo non tutti i metodi del modulo `NetworkX` funzionano con questa classe: diverse volte siamo stati costretti a convertire i grafi in `DiGraph`: entrambe le classi permettono la creazione di grafi di tipo diretto, ma la seconda non permette archi multipli tra due nodi.

3 Analisi dei dati ottenuti

3.1 Proprietà grafi

Il grafo `Completo` non è né connesso né bipartito, invece i grafi `Follower` e `Following` sono connessi ma non bipartiti. In tutti i grafi la massima distanza è 7 mentre la media è circa 3.850 per tutti.

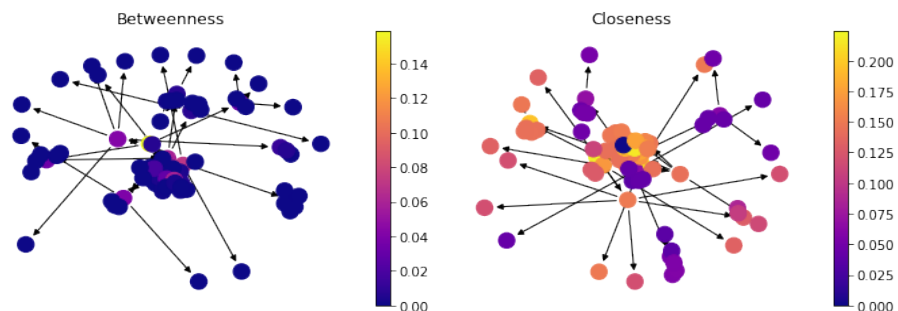
3.2 Clustering

I coefficienti di clustering ottenuti sono circa: 0.192 sul grafo `Completo`, 0.283 sul grafo dei `Follower`, 0.276 sul grafo dei `Following`. Possiamo dire che nei due sottografi si può rilevare una tendenza a formare delle *terzine* che, socialmente parlando, rappresentano i pattern di relazione di amicizia.

3.3 Misure di centralità

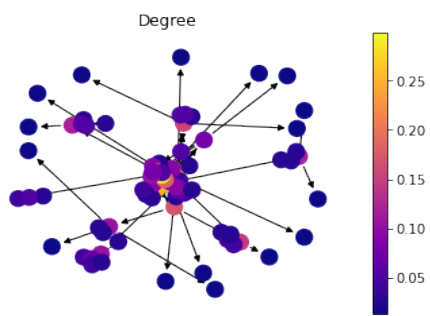
- *Betweenness e Closeness*

| Grafo | Betweenness | Nodo | Closeness | Nodo |
|-----------|-------------|---------|-----------|------------|
| Completo | 0.319 | mizzaro | 0.306 | mizzaro |
| Follower | 0.158 | mizzaro | 0.224 | OlegZendel |
| Following | 0.158 | mizzaro | 0.384 | Miccighel_ |



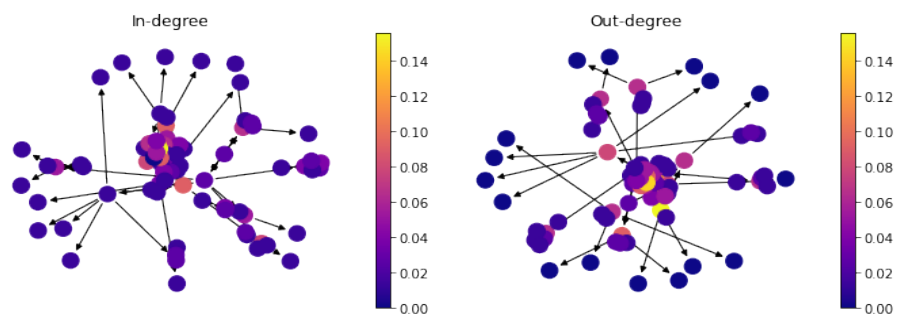
• Degree

| Grafo | Degree | Nodo |
|-----------|--------|-----------------|
| Completo | 0.234 | sophiaalthammer |
| Follower | 0.298 | sophiaalthammer |
| Following | 0.298 | sophiaalthammer |



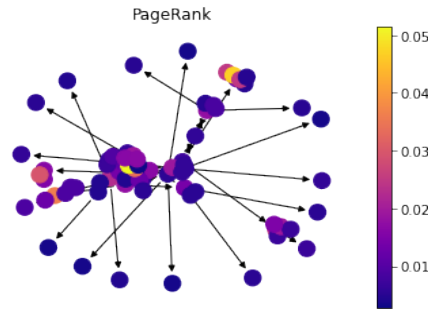
• In-degree e Out-degree

| Grafo | In-degree | Nodo | Out-degree | Nodo |
|-----------|-----------|-----------------|------------|-----------------|
| Completo | 0.117 | sophiaalthammer | 0.117 | sophiaalthammer |
| Follower | 0.155 | sophiaalthammer | 0.155 | rodger_benham |
| Following | 0.155 | rodger_benham | 0.155 | sophiaalthammer |



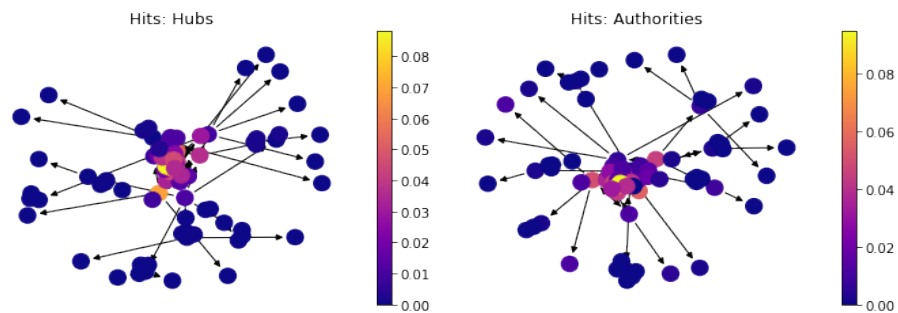
- *Pagerank*

| Grafo | Page Rank | Nodo |
|-----------|-----------|-----------------|
| Completo | 0.338 | rodger_benham |
| Follower | 0.051 | sophiaalthammer |
| Following | 0.084 | Outsider289 |



- *HITS*

In tutti i grafi i valori più alti di *authorities* e di *hub* sono di **sophiaalthammer**. Nello specifico nel grafo dei **Follower** i valori più alti di ciascuno sono 0.088 e 0.094.



3.4 Small-world-ness

Essendo Σ maggiore di 1 possiamo affermare che tutti i grafi si comportano come uno "Small-World" inoltre, grazie al valore Ω compreso tra 0 e 1, possiamo affermare che è uno "Small-World" con tendenza ad un grafo randomico (se fosse stato compreso tra 0 e -1 avremo una tendenza a un grafo reticolare).

| Grafo | Ω | Σ |
|-----------|----------|----------|
| Completo | 0.297 | 2.263 |
| Follower | 0.293 | 2.485 |
| Following | 0.295 | 2.518 |

3.5 Analisi delle correlazioni di Pearson e Kendall

Tra le tabelle di Pearson e Kendall non vi sono differenze tali da dare risultati diversi, quindi per l'osservazione dei dati abbiamo scelto la correlazione lineare

di Pearson. Guardando la tabella del grafo **Completo** si nota subito che la correlazione tra **Degree**, **In-Degree** e **Out-Degree** è circa la stessa ed è tendente a 1, a dimostrazione che i due sottografi sono uno speculare dell'altro. Guardando la tabella del grafo dei **Follower** si nota una leggera correlazione tra la **Betweenness** e i vari **Degree**, mentre una forte correlazione si può notare tra **Degree** e **In/Out-Degree**.

| | Betweenness | Closeness | Degree | In-Degree | Out-Degree |
|-------------|-------------|-----------|--------|-----------|------------|
| Betweenness | 1.000 | 0.411 | 0.574 | 0.573 | 0.574 |
| Closeness | 0.411 | 1.000 | 0.682 | 0.681 | 0.682 |
| Degree | 0.574 | 0.682 | 1.000 | 0.999 | 0.999 |
| In-Degree | 0.573 | 0.681 | 0.999 | 1.000 | 0.998 |
| Out-Degree | 0.574 | 0.682 | 0.999 | 0.998 | 1.000 |

Tabella 1: Pearson grafo completo.

| | Betweenness | Closeness | Degree | In-Degree | Out-Degree |
|-------------|-------------|-----------|--------|-----------|------------|
| Betweenness | 1.000 | 0.426 | 0.722 | 0.658 | 0.708 |
| Closeness | 0.426 | 1.000 | 0.508 | 0.550 | 0.419 |
| Degree | 0.722 | 0.508 | 1.000 | 0.941 | 0.952 |
| In-Degree | 0.658 | 0.550 | 0.941 | 1.000 | 0.795 |
| Out-Degree | 0.708 | 0.419 | 0.952 | 0.795 | 1.000 |

Tabella 2: Pearson Follower

4 Conclusioni

Per l'osservazione dei dati abbiamo scelto il grafo dei **Follower**, principalmente perché il grafo completo ha un numero elevato di nodi non connessi che alterano troppo i risultati, e il grafo dei **Following** non è niente altro che lo speculare di quello dei **Follower**. Osservando il grafo dei **Follower** ottenuto graficamente (*follower_network_graphs.html*) si vede che:

- Solo una delle due radici segue l'altra e l'unico follower in comune è **ChriShot90**.
- Nei follower di **Miccighel_** si formano dei piccoli gruppi che in parte tendono a seguirsi a vicenda formando anche dei *triangoli*, ricordando il modello e pattern di amico di Facebook.
- Nei follower di **mizzaro** vi sono 3 gruppi di cui uno molto grande e con la maggior parte dei nodi che si seguono a vicenda.