

FLICKR

Metodi e Applicazioni per social network

Bassani Elisa 855403

Indice

- Introduzione
- Punti di forza
- Somiglianze con altri siti
- Dataset
- Grafo
- Scale free?
- Osservazioni
- Domande di ricerca
 - Rete creata dagli utenti
 - Rete creata dai gruppi
 - Componenti connesse
 - Assortatività
 - Community
- Conclusioni
- Sviluppi futuri

Introduzione

- Flickr è un sito web multilingua, che ha anche un'applicazione per dispositivi mobili; è di proprietà del gruppo Yahoo!.
- Permette agli utenti che si sono iscritti di condividere fotografie personali, ricercare fotografie secondo categorie, geotag, recenti, popolari.
- Flickr è stato sviluppato dalla Ludicorp, una compagnia canadese (di Vancouver) che venne fondata nel 2002 da Stewart Butterfield e Caterina Fake.
- Il termine usato come nome del sito, viene preso dall'inglese; flicker, significa "tremolare", o "sfavillare".
- Nel marzo del 2005, la Ludicorp, e quindi Flickr, sono stati comprati da Yahoo!, con sede negli Stati Uniti.
- Il 16 maggio 2006 Flickr concluse la fase di sviluppo beta, quindi, non fu più in fase di prova ma in stato di perpetua evoluzione.

Punti di forza

- Un utente registrato a Flickr, ha la possibilità di tenere salvate le sue foto personali su uno **spazio cloud** di 1 Tb, che viene 'consegnato' al momento della creazione dell'account; così da poter essere utilizzato per raccogliere la digigrafia delle persone, è un servizio privato, che se l'utente vuole può scegliere di mettere in condivisione con gli altri.
- L'**organizzazione delle foto** è molto semplice, anche con grosse quantità di foto scattate con diversi strumenti fotografici; tramite l'applicazione online, o nel web browser, o via MMS o via e-mail, subito dopo aver scattato la foto. Le fotografie vengono catalogate ed indicizzate attraverso delle parole chiave e dei tag, anche per luogo di scatto.
- Flickr permette di controllare chi ha il **diritto di vedere le foto**, tramite l'impostazione pubblica o privata; l'impostazione privata permette di distinguere anche le cerchie di contatti ai quali far vedere determinate foto.

- Il sito ha implementato un sistema per **filtrare** le fotografie potenzialmente controverse e volgari. Questo sistema permette agli utenti di giudicare e segnalare un'immagine come sicura, ristretta o da censura; gli utenti possono scegliere se visualizzare tutte le immagini oppure solo quelle appartenenti a una delle 3 categorie
- Non esistono le persone **famose**; chi ha molti followers è perché se li è guadagnati

Somiglianze con altri siti

- Meccanismo di follower/following (Twitter, Instagram, LinkedIn, Pinterest..).
- Utilizzo di tag sulle foto, che è l'equivalente degli hashtag.
- Utilizzo del geo tagging per le foto.
- Ricerca di utenti, persone e gruppi.
- Condivisione di foto, video e immagini

Dataset

- I dati sono stati presi tramite le api messe a disposizione da Flickr.
- Le api utilizzate sono disponibili al seguente indirizzo:
<https://www.flickr.com/services/api/>
- Avendo a disposizione le api, non ci sono limiti sulla quantità di dati scaricabili, ma solo limiti di tempo impiegato nel download.
- Per fare l'analisi ho fatto due sotto studi:
 1. Per il primo, partendo dagli id delle foto più recenti e cercando gli utenti che le avevano postate, da questi ho ricavato i gruppi a cui questi utenti sono iscritti.
Ho scaricato circa 290 mila id di foto prese dalla sezione 'foto recenti', da questi id delle foto, ho ricercato gli id delle persone che le avevano pubblicate, ho trovato circa 140 mila utenti.
Di questi ho cercato di raccogliere gli username dei profili e il luogo (quelli che lo hanno inserito) di abitazione.

2. Il **secondo** invece, prendendo dei gruppi randomicamente tramite alcune parole chiave (come photo, travel, nature, animal,...) e da questi si sono presi gli utenti iscritti a questi.

Ho scaricato circa 470 mila id di gruppi, togliendo i doppioni ne ho ricavati circa 3 mila, da questi ho preso gli utenti iscritti.

Grafo

- Creazione grafo:

1. Per il **primo** grafo, ho creato un grafo non orientato, dove ho creato dei nodi utente e gruppo, i collegamenti sono presenti tra il nodo utente e il gruppo al quale questo è iscritto.

In questo modo ho ottenuto un grafo con :

141 068 nodi

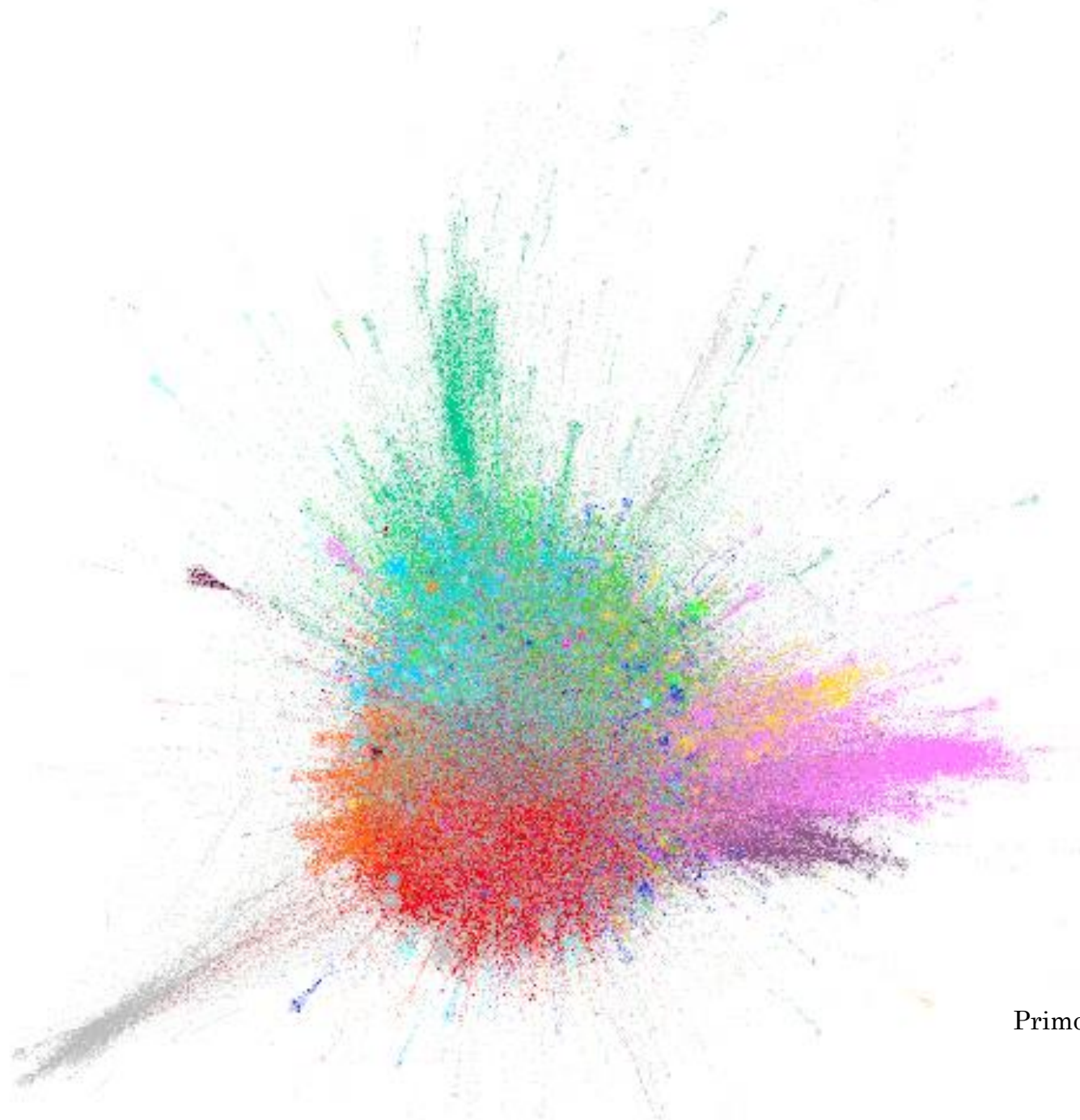
557 140 archi

2. Per il **secondo** grafo, ho creato un grafo non orientato, dove, i nodi sono di tipo utente o gruppo, i collegamenti li ho creati tra il gruppo e gli utenti iscritti a questo gruppo.

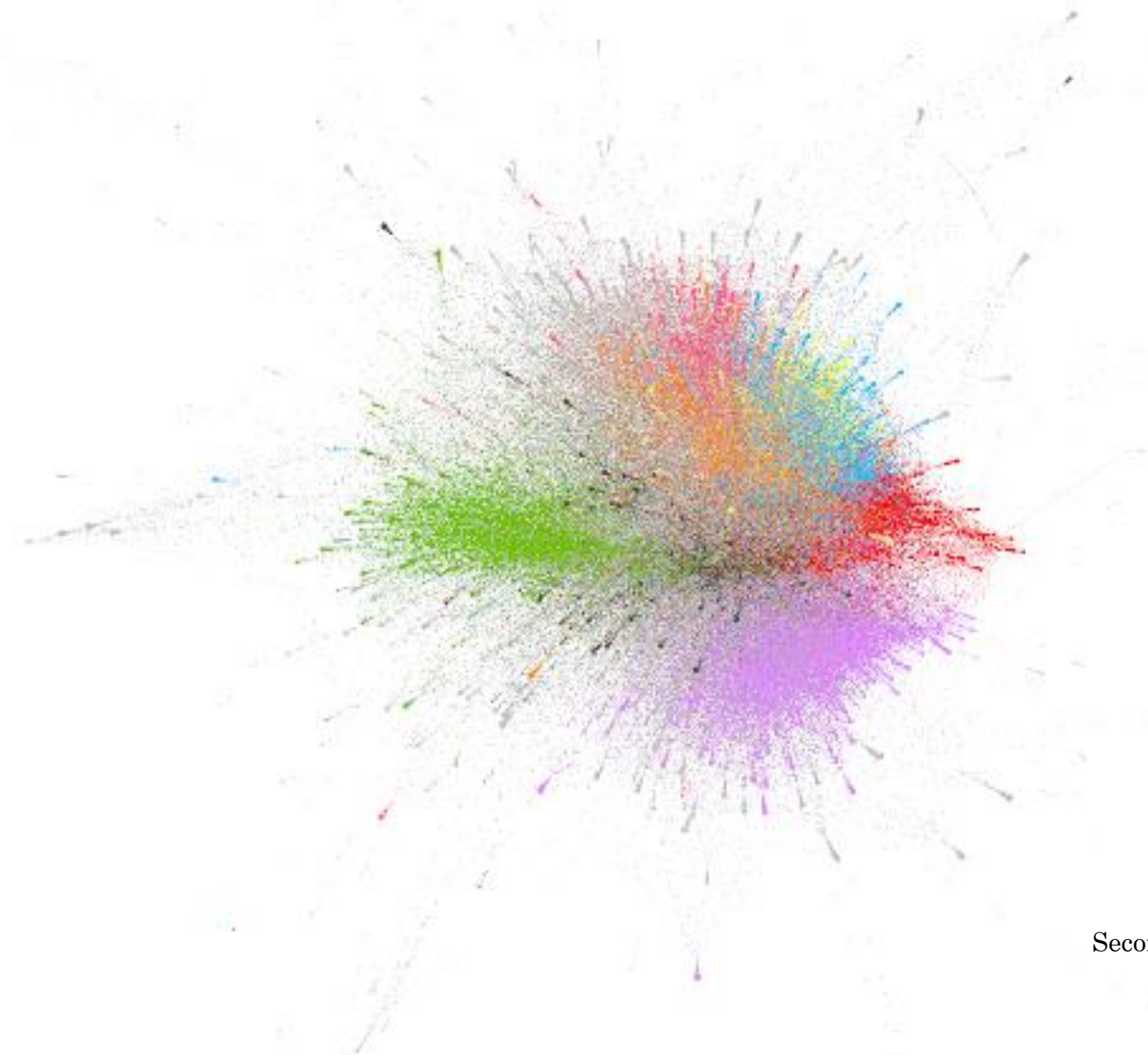
In questo modo ho ottenuto un grafo con:

158 197 nodi

281 763 archi



Primo grafo

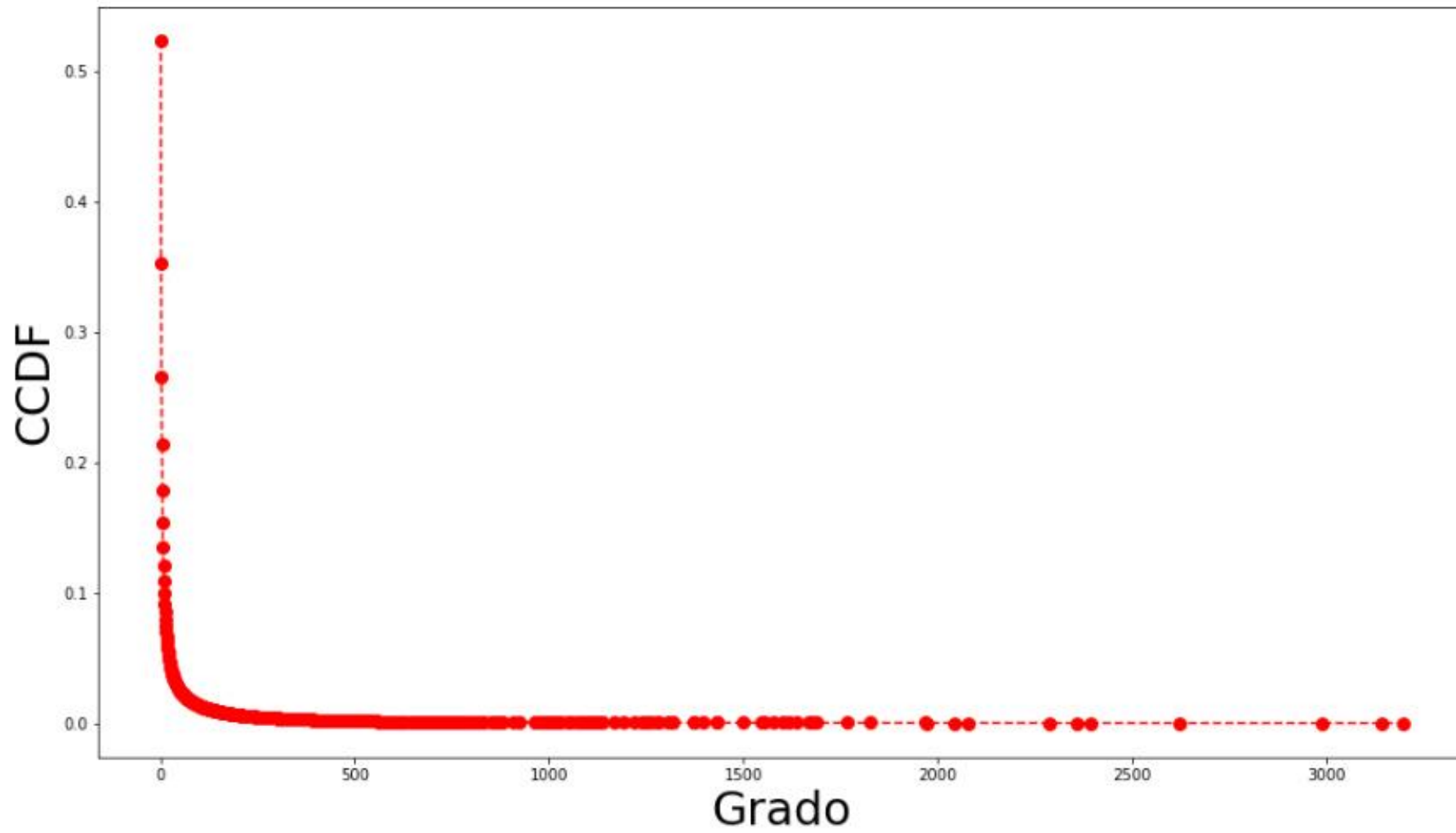


Secondo grafo

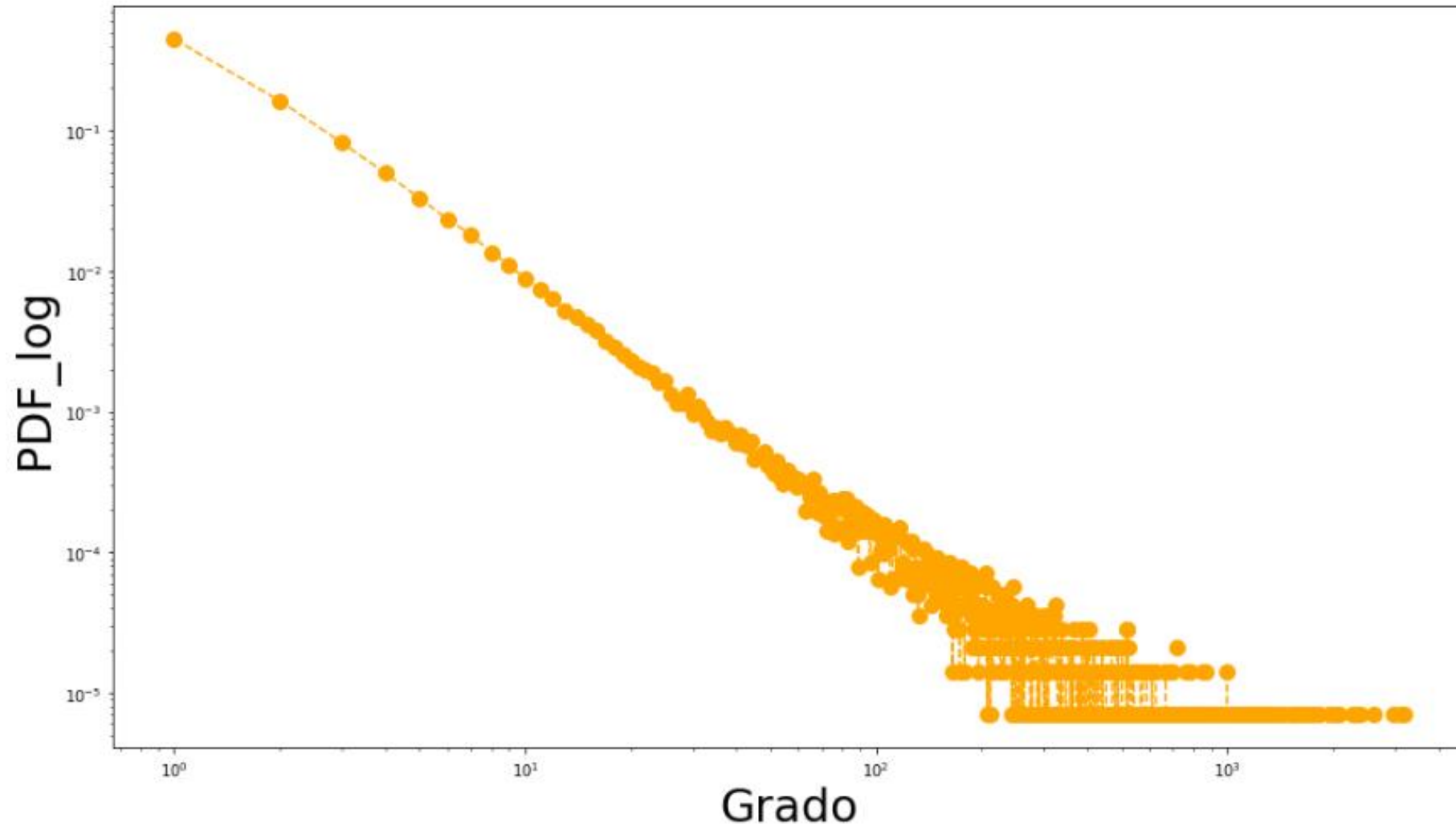
Scale free?

- In una rete scale free ci sono alcuni nodi che hanno un alto grado, che vengono chiamati hub, sono i nodi più grandi della rete a cui vengono passate le informazioni quando si vogliono diffondere a un grande numero di persone.
- Una rete power law si riconosce dalla media alta e dalla mediana bassa, come succede nella mia analisi, la rappresentazione mostrata è quella in scala loglog; in questo modo si può vedere meglio l'andamento, più si avvicina a una retta e più la rete è considerata power law.
- Le reti reali tendono a seguire il modello della power law.
- Per svolgere le operazioni ho tolto dalla rete trovata, i nodi con grado 0.

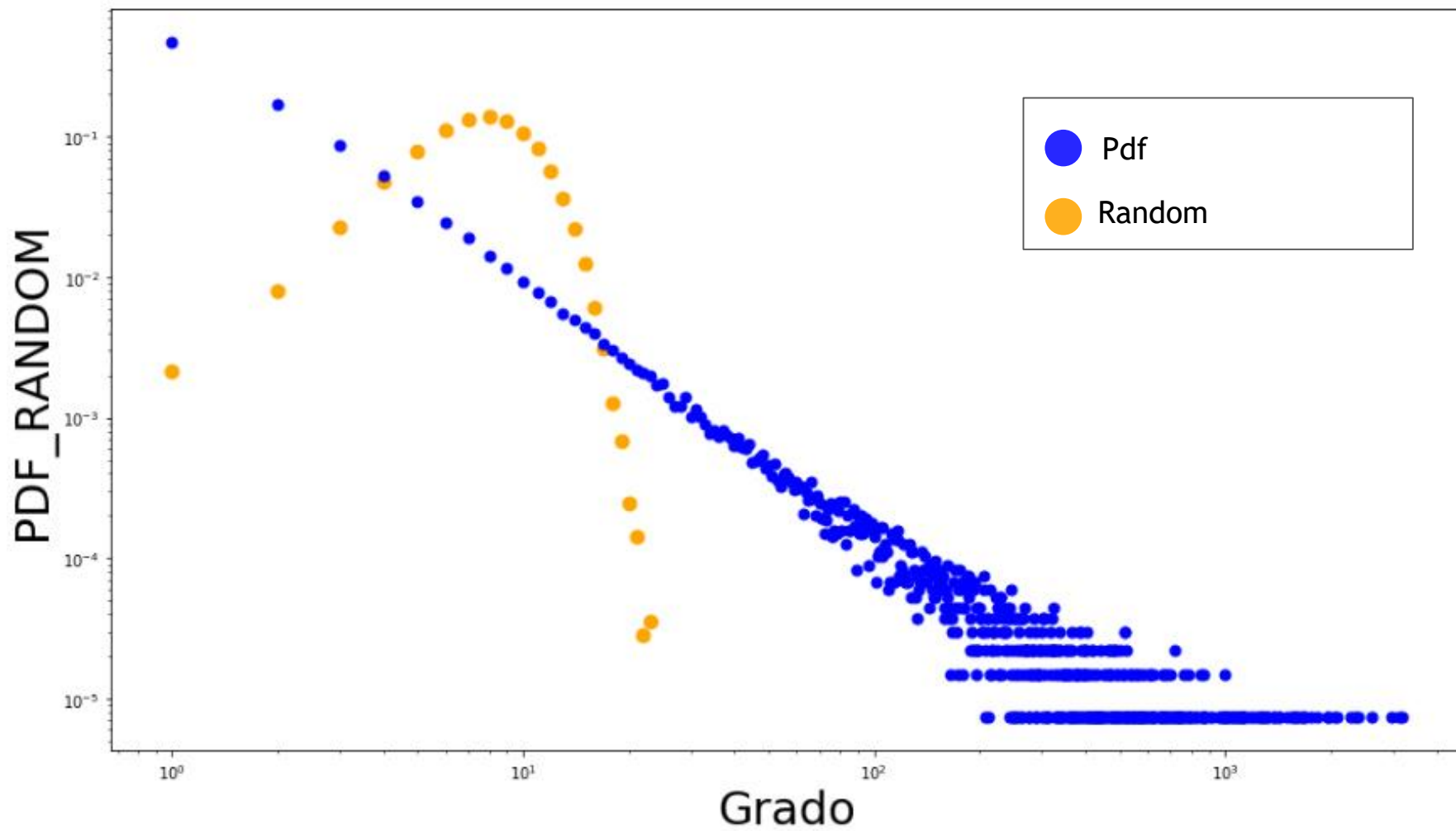
CCDF primo grafo



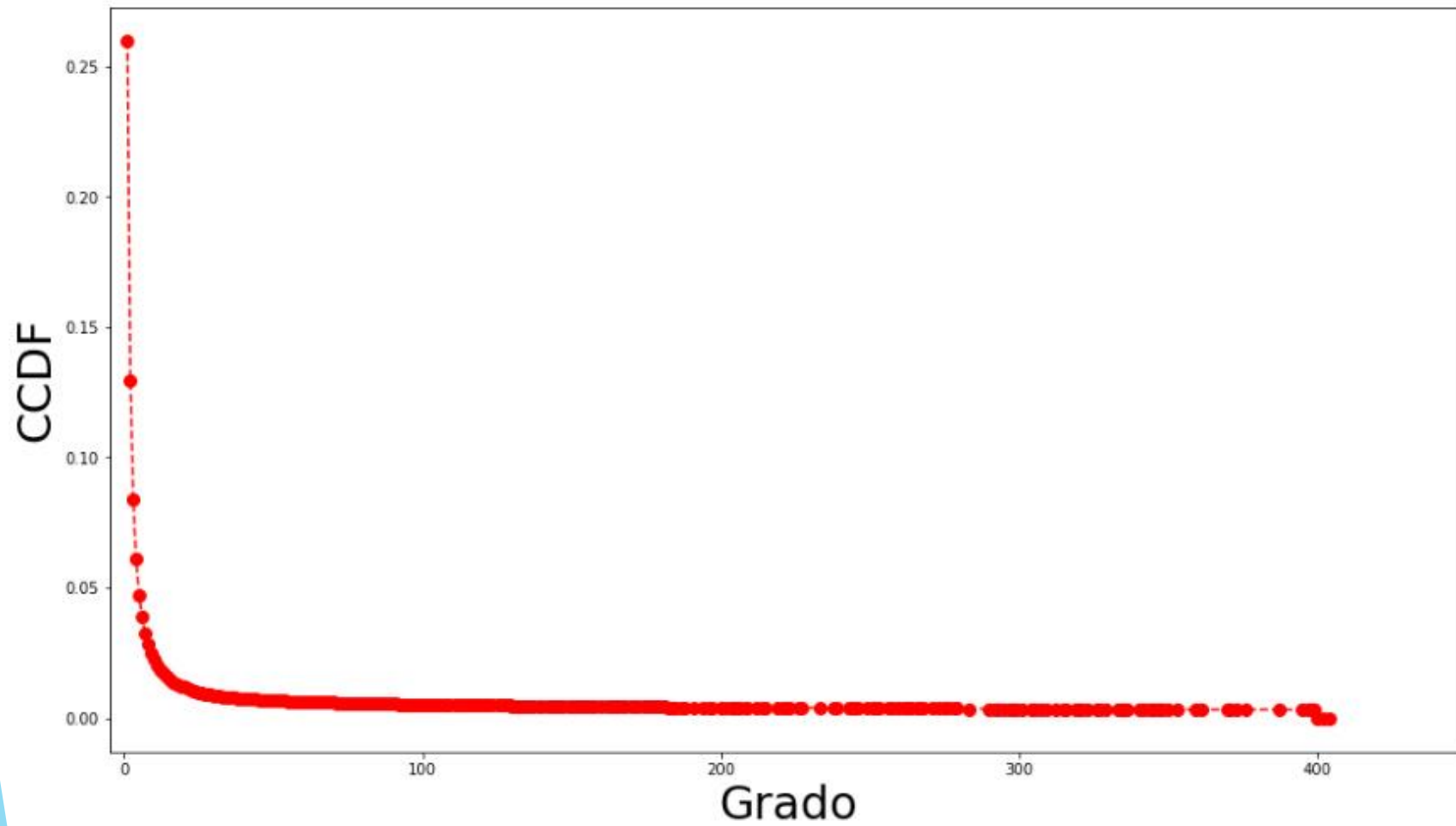
PDF primo grafo



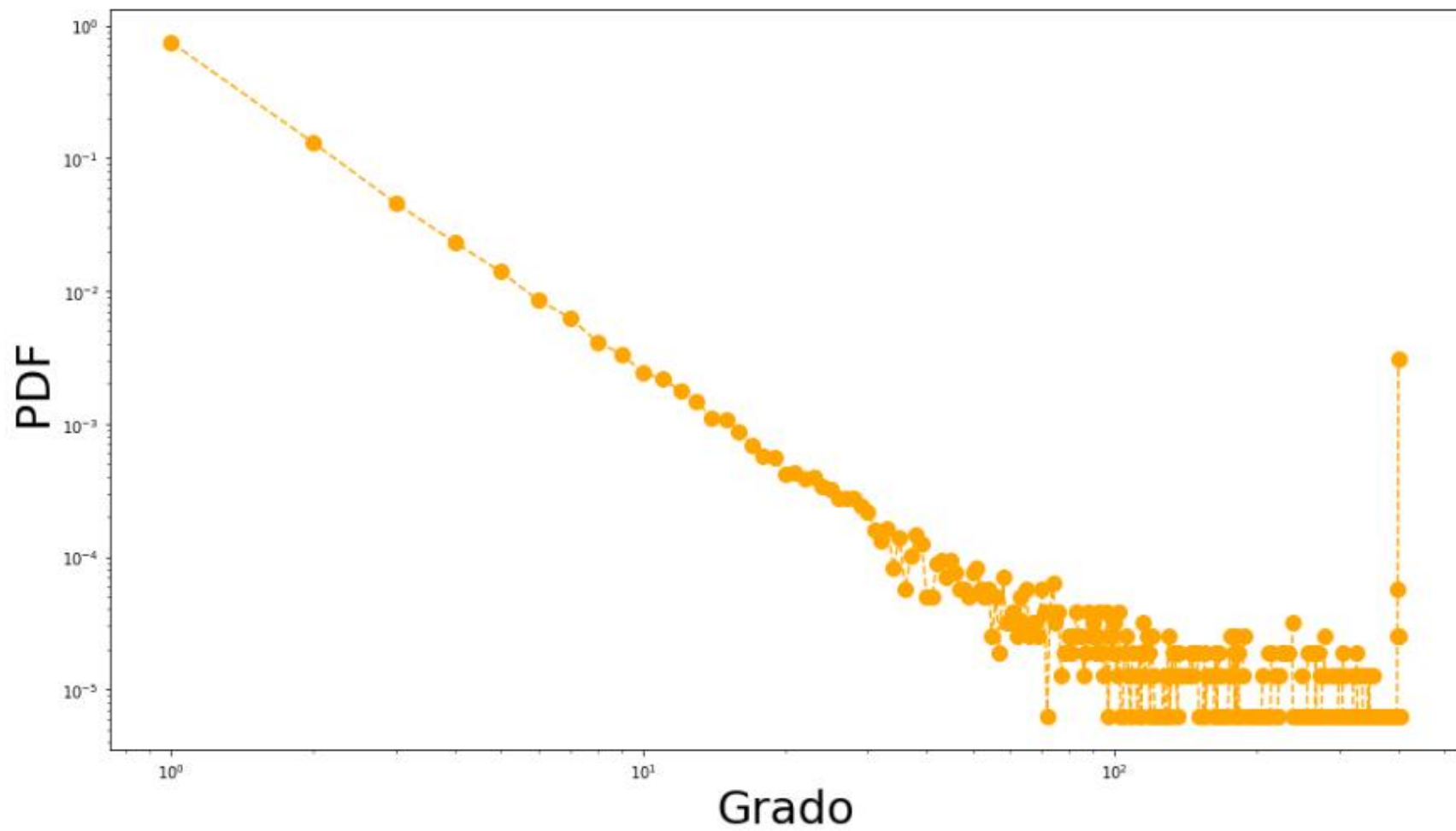
Confronto con random



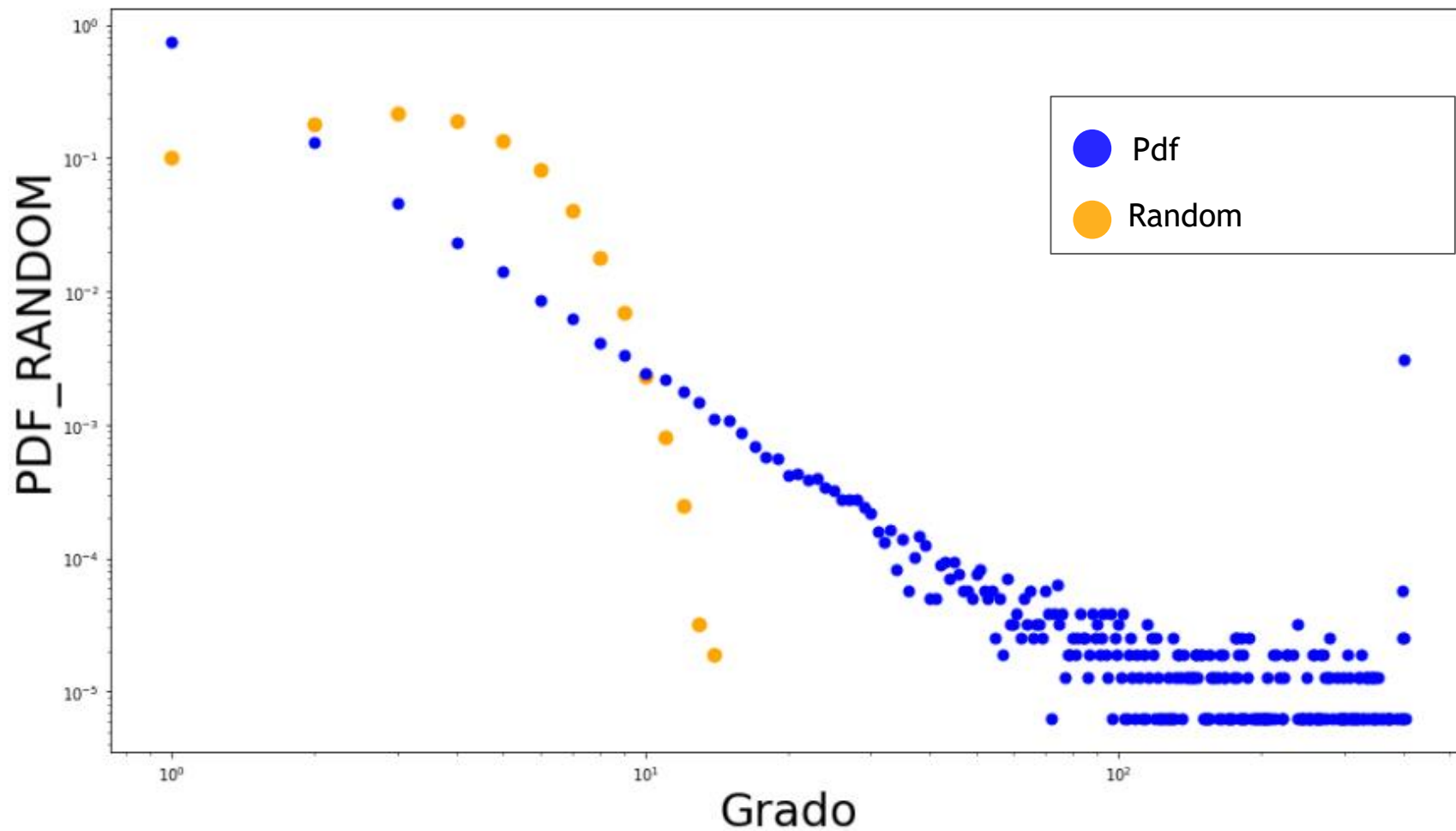
CCDF secondo grafo



PDF secondo grafo



Confronto con random



Osservazioni

- La rete che ho analizzato risulta essere una *scale free*, dove la distribuzione del grado segue l'andamento di una *power law*. Una scale free è una rete senza una scala interna.
- Primo grafo:
 - $\langle d \rangle = 8.3002845522399173$
 - mediana = 2.0
- Secondo grafo:
 - $\langle d \rangle = 3.562201319881666$
 - mediana = 1.0

Domande di ricerca

- Come è fatta la rete creata dagli utenti
 - % di utenti
 - % di gruppi
 - modularità
- Come è fatta la rete creata tramite i gruppi
 - % di utenti
 - % di gruppi
 - modularità
- Numero di componenti connesse presenti nelle due reti
- Studio della componente connessa più grande nelle due reti
- Assortatività per nazione
- Community detection

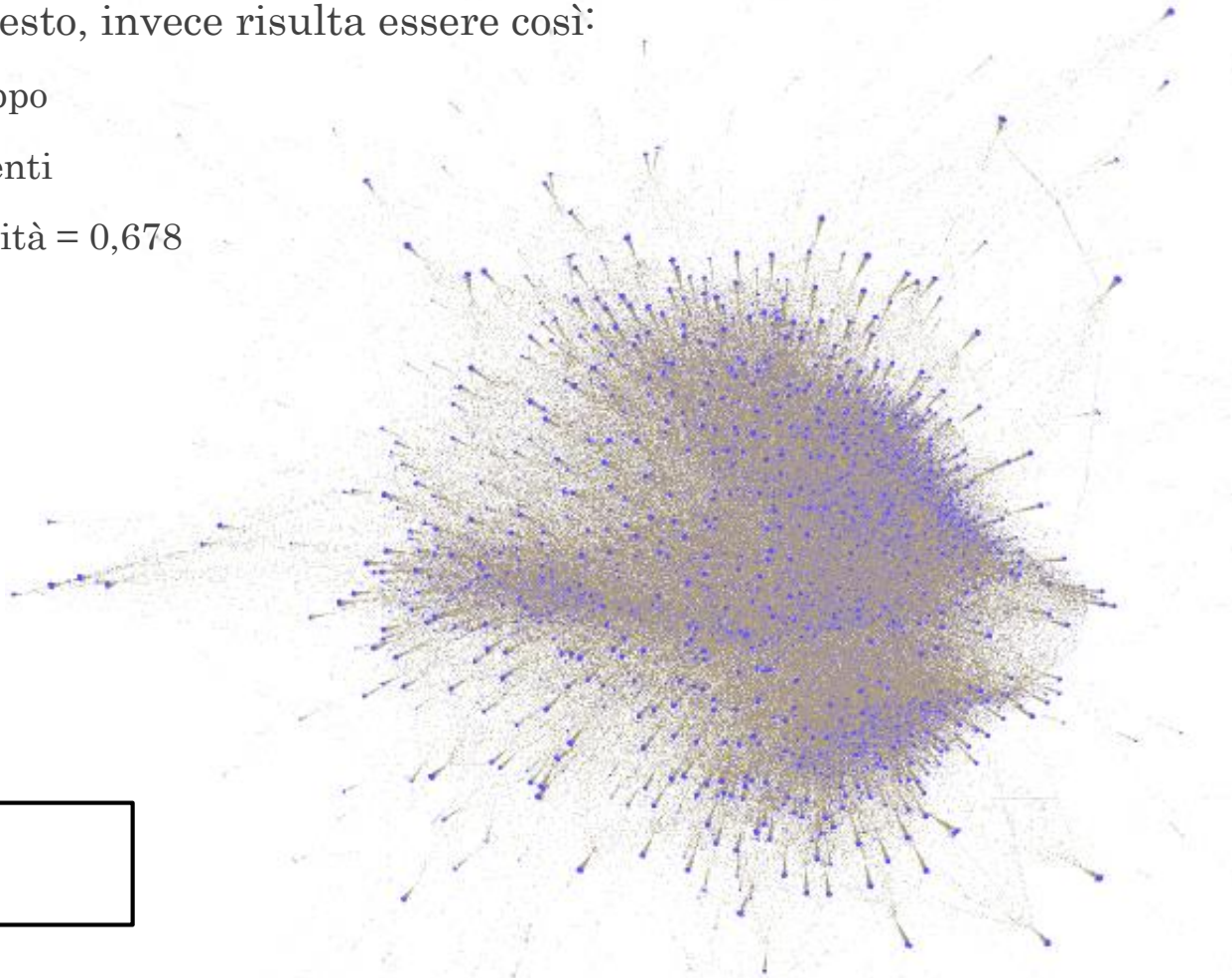
Rete creata dagli utenti

- La rete creata partendo dallo studio degli utenti e guardando i gruppi a cui questo è iscritto risulta essere formata così:
 - 90 % gruppo
 - 10 % utenti
 - modularità = 0,407
 - modularità_python
0.42638059947379964

Utente
Gruppo

Rete creata dai gruppi

- La rete creata partendo dallo studio dei gruppi e cercando poi gli utenti che sono iscritti a questo, invece risulta essere così:
 - 1 % gruppo
 - 99 % utenti
 - modularità = 0,678



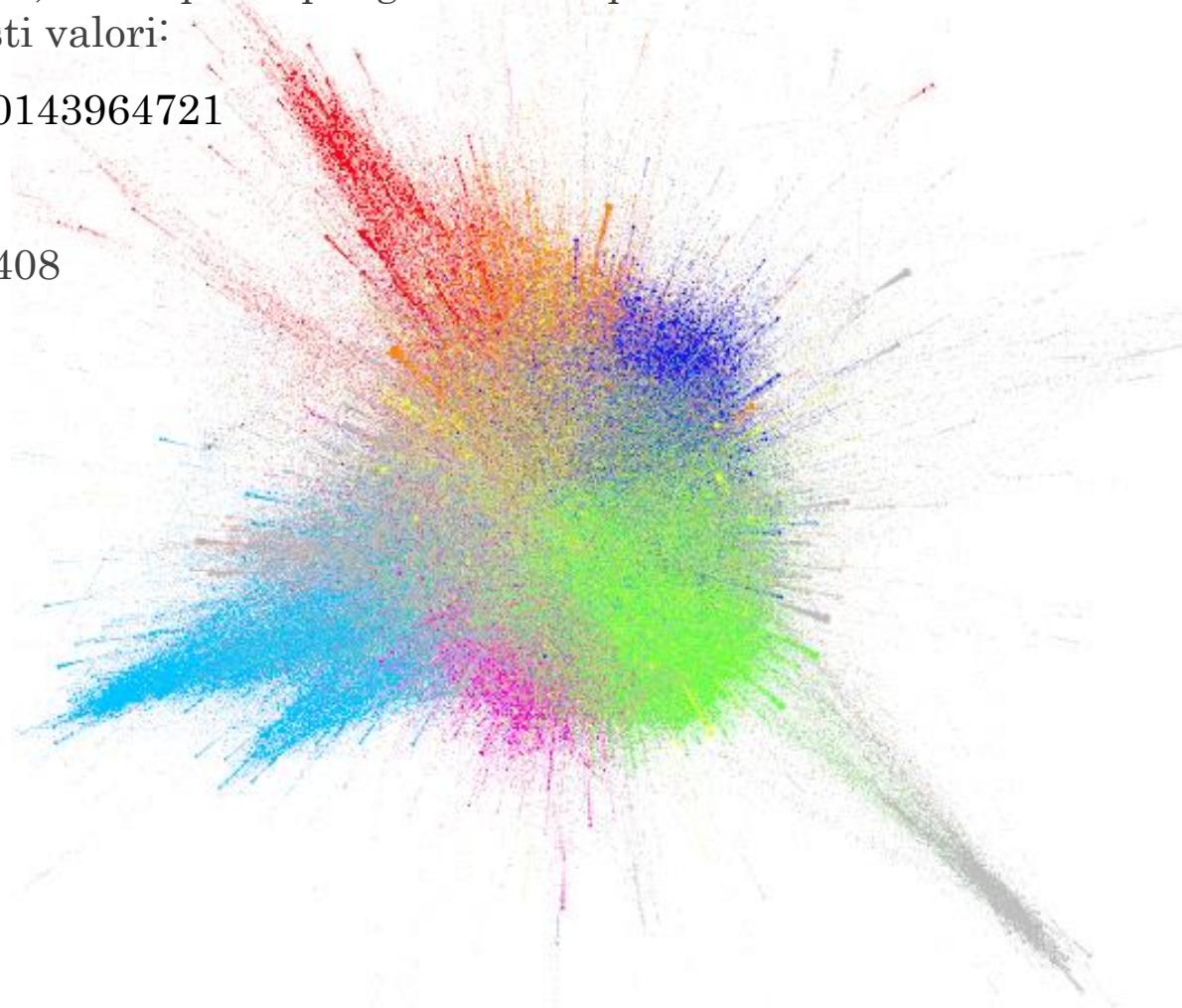
Utente
Gruppo

Componenti connesse

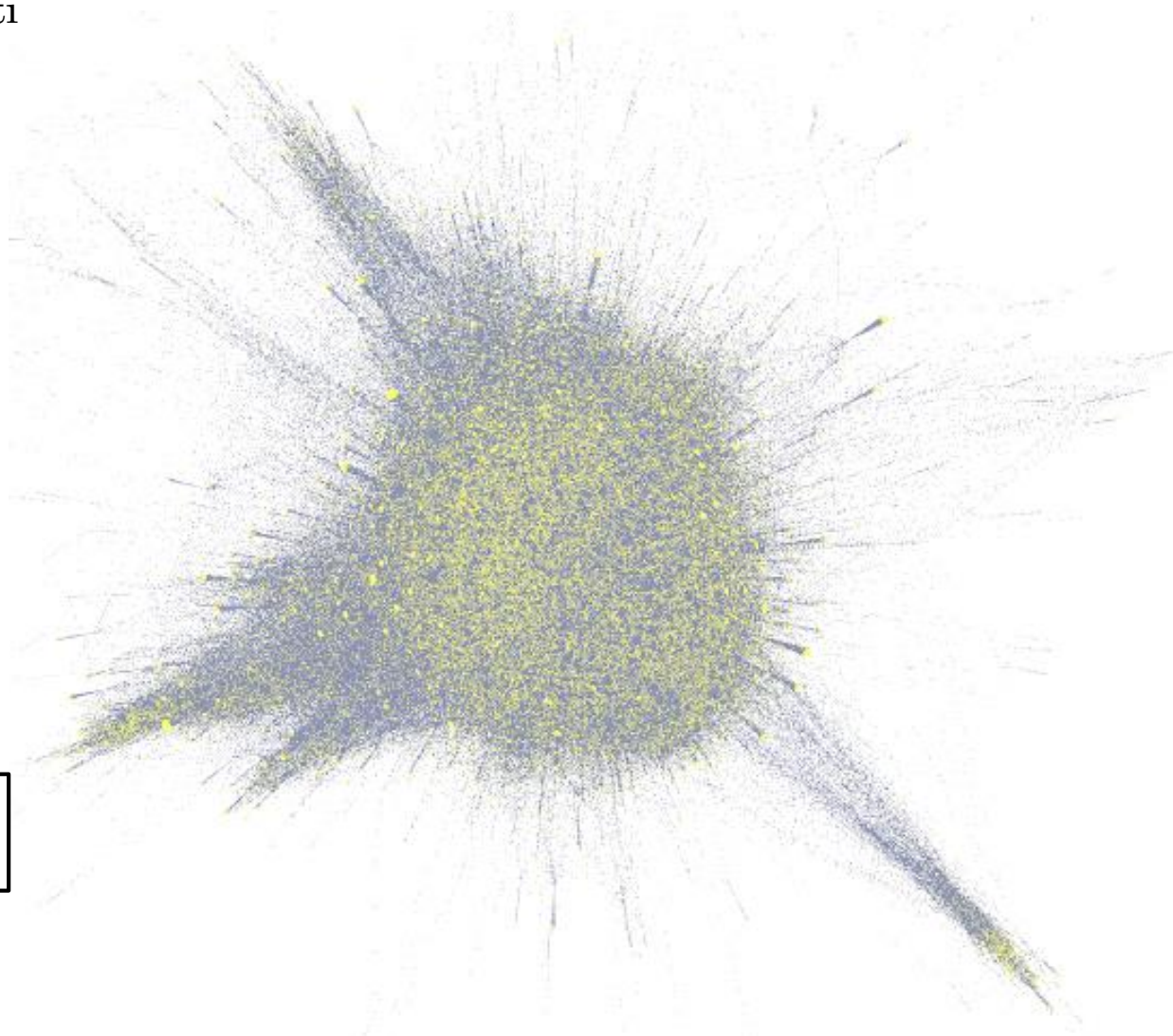
- Le componenti connesse che ho trovato nelle due reti sono:
 - Nella prima, ossia quella dove si è partiti dall'id dell'utente, si sono formate 336 componenti connesse
 - Nella seconda, ovvero, quella creata partendo dai gruppi, le componenti connesse che sono state generate sono 154

Componente connessa primo

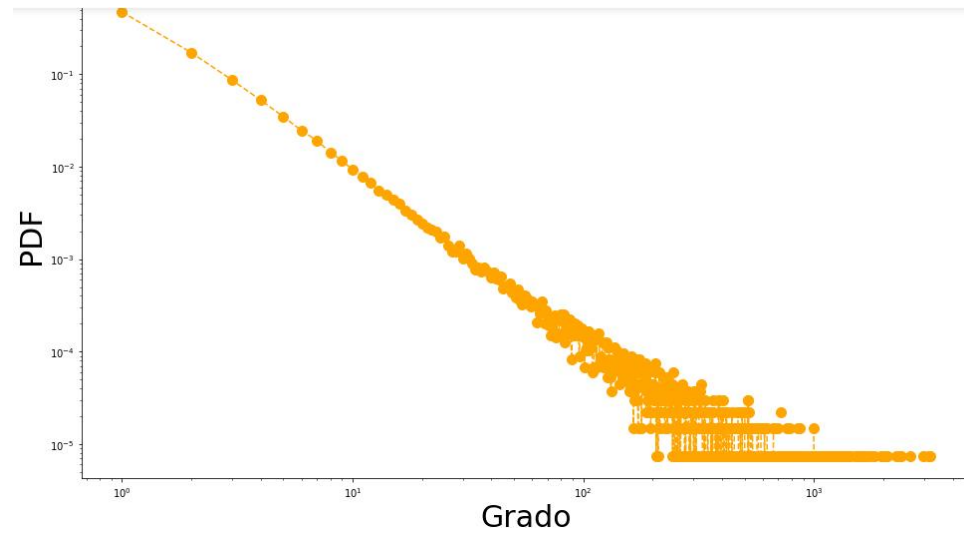
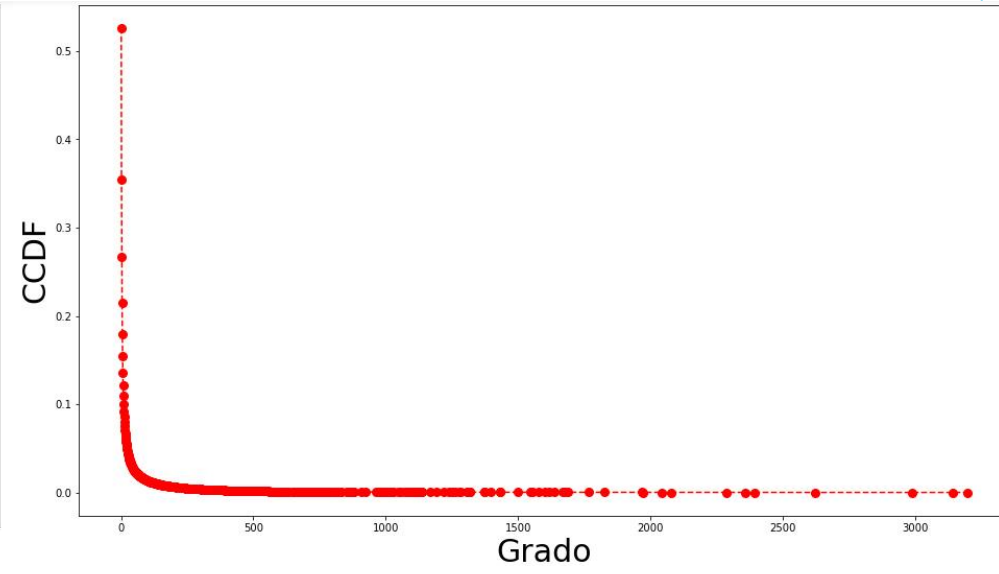
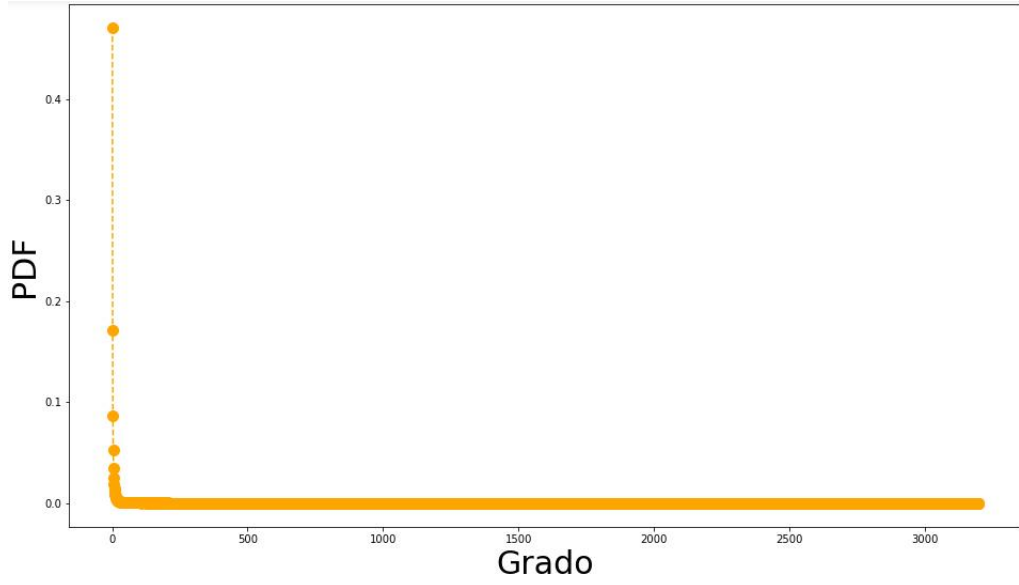
- In questa componente, che è quella più grande tra quelle che si sono formate nella rete, otteniamo questi valori:
 - $\langle d \rangle = 8.3468800143964721$
 - mediana = 2.0
 - modularità = 0,408



- Questa componente è formata dal
 - 5 % di gruppi
 - 95 % di utenti

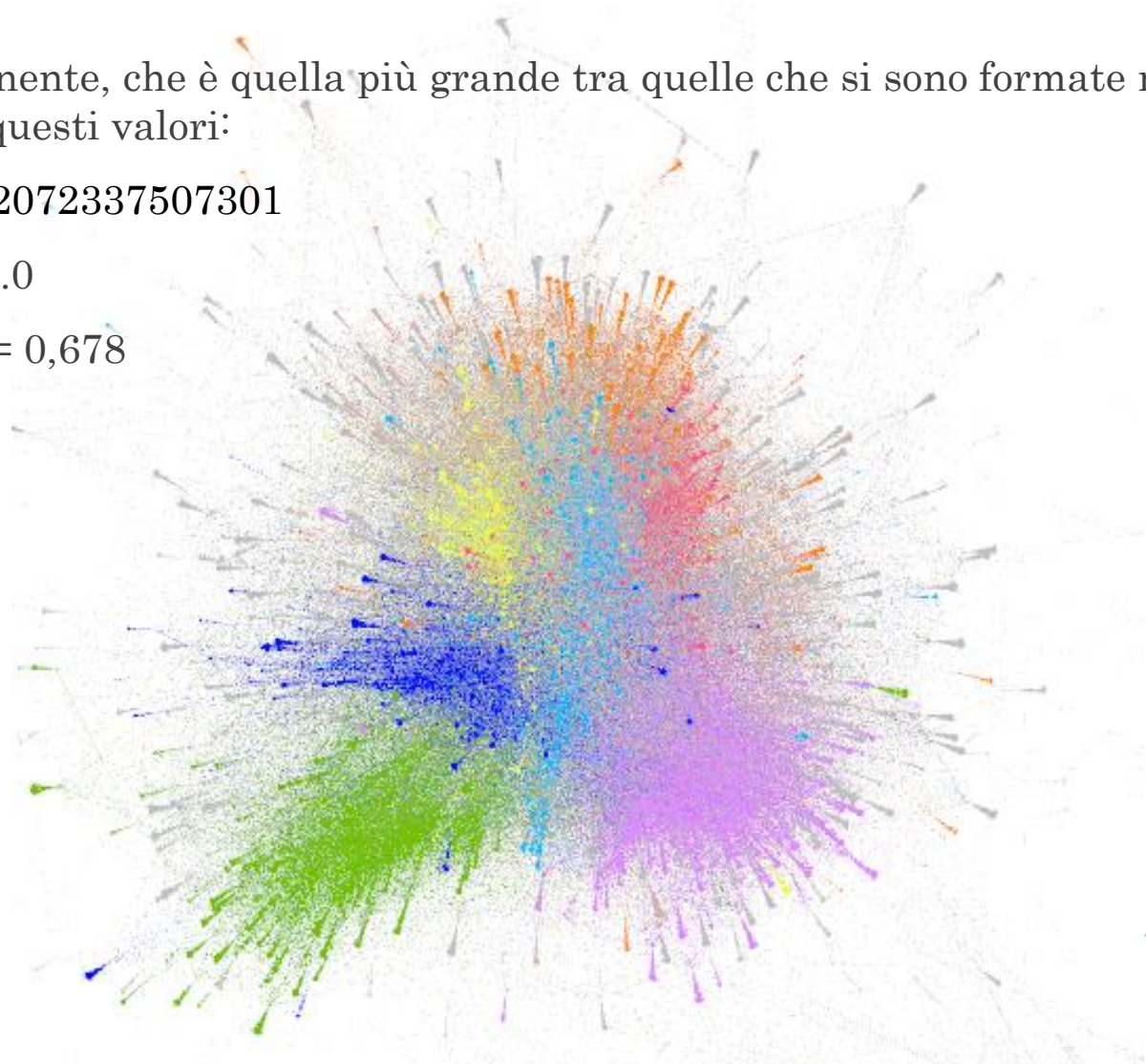


Utente
Gruppo

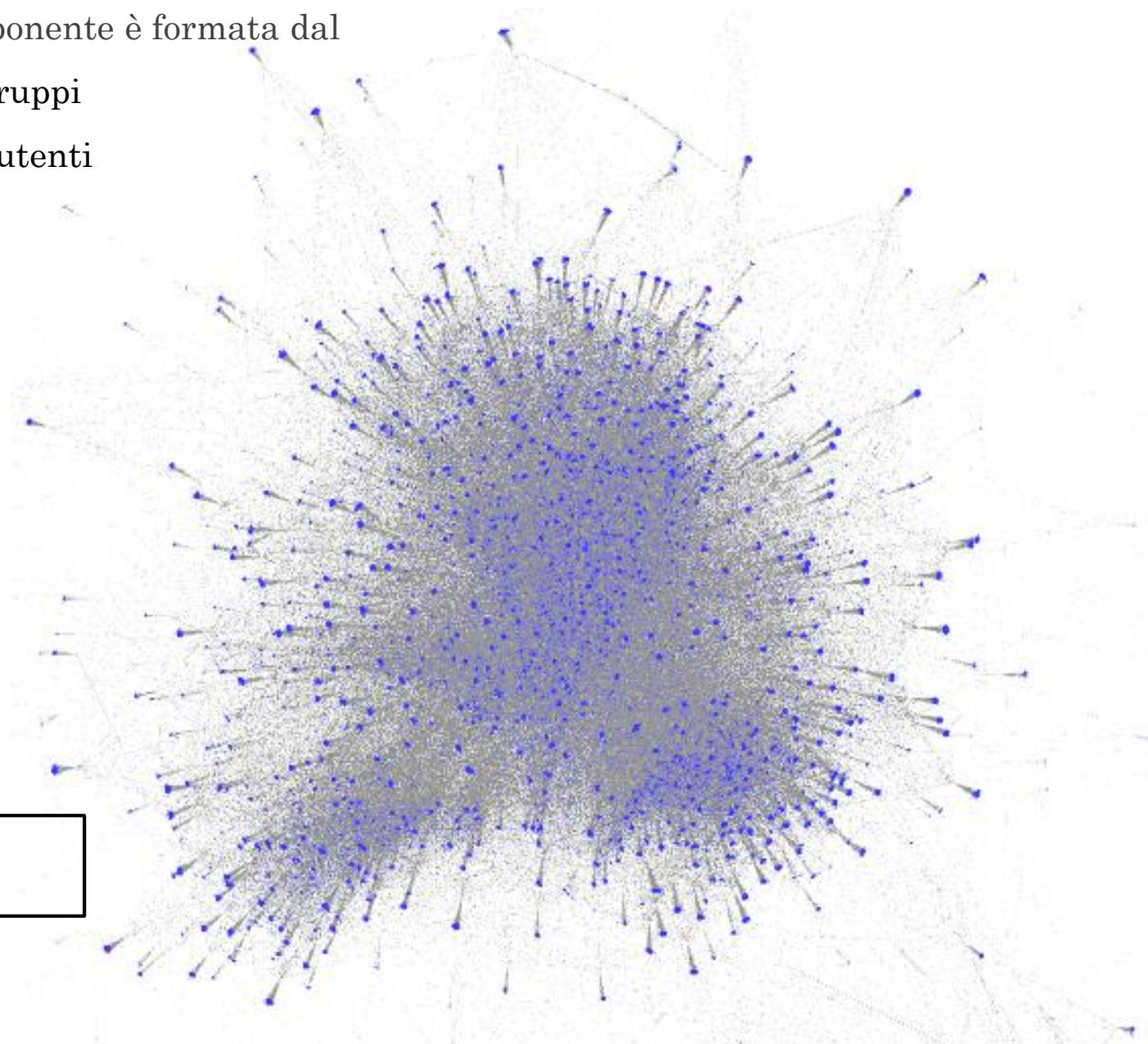


Componente connessa secondo

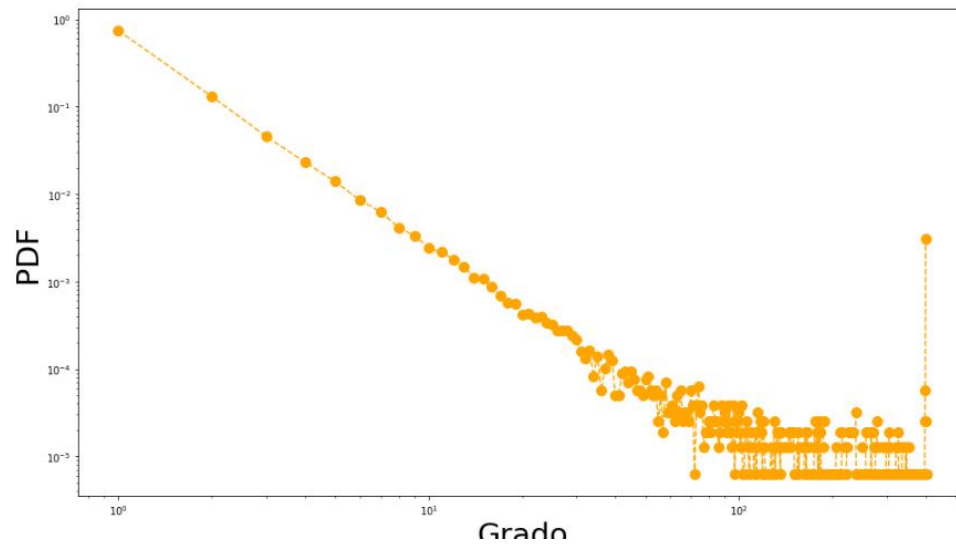
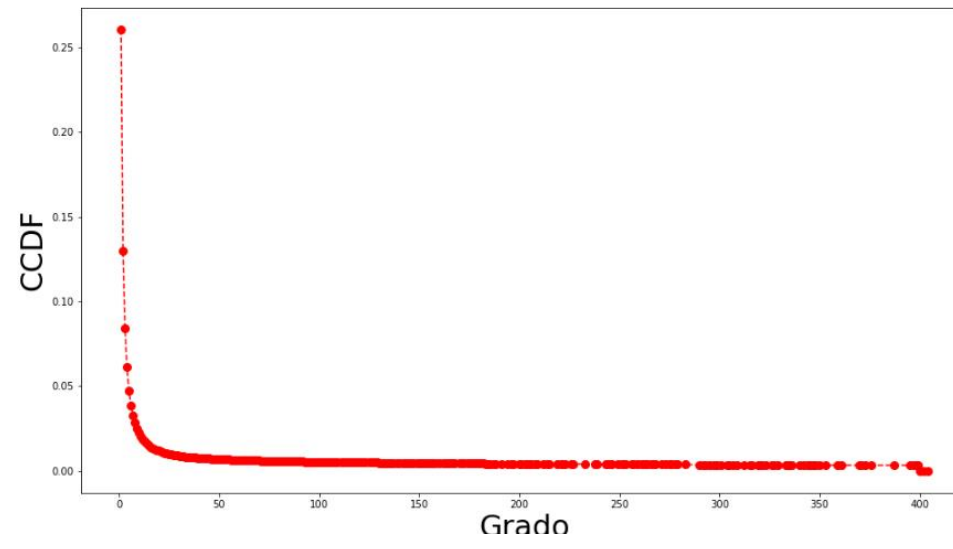
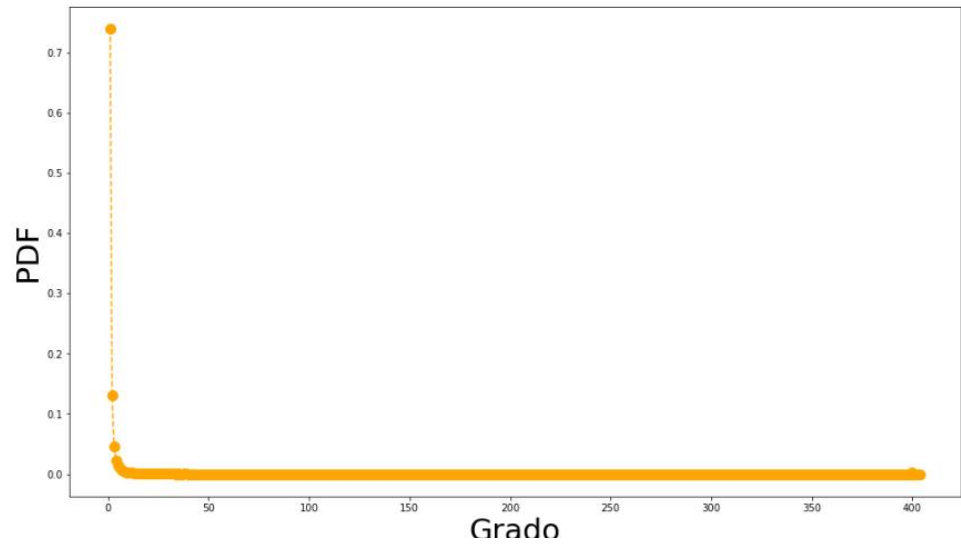
- In questa componente, che è quella più grande tra quelle che si sono formate nella rete, otteniamo questi valori:
 - $\langle d \rangle = 3.5712072337507301$
 - mediana = 1.0
 - modularità = 0,678



- Questa componente è formata dal
 - 1 % di gruppi
 - 99 % di utenti



Utente
Gruppo



Assortatività

- I luoghi più presenti sono i seguenti:
 - United Kingdom : 12)
 - London: 3
 - Houston, USA': 3
 - UK': 3
 - Scotland', 3
 - Australia', 3
 - USA', 3
 - Toronto, Canada 3
 - Canada', 2

Community

- Le community più presenti sono composte da questi membri:
- 4, 19833
- 5, 18252
- 2, 15954
- 7, 13941
- 6, 8743
- 11, 8124
- 8, 7936
- 14, 6311
- 12, 6179
- 1, 3749
- 9, 3396
- 20, 3390
- 13, 3148
- 0, 2873
- 10, 2390
- 15, 2304
- 18, 2159
- 3, 2144

Conclusioni

- E' scale free
- Nel grafo creato partendo dagli utenti, c'è la presenza di molti gruppi e pochi utenti
- Nel grafo creato partendo dai gruppi, c'è la presenza di molti utenti e pochi gruppi
- Entrambe le reti sono molto omogenee, per utenti o per gruppi, c'è una prevalenza di uno dei due sull'altro, questo è dovuto al fatto che si parte da uno delle due tipologie presenti nel grafo
- Pochi utenti inseriscono la localizzazione, quelli che la inseriscono, sono maggiormente localizzati nell'Europa del Nord e negli Stati Uniti.
- Le community sono presenti con molti utenti, sono molte e ben distribuite
- La componente maggiore ha una scala molto simile a quella dell'intera rete

Sviluppi futuri

- Vorrei studiare la rete per vedere se è presente un modello Small World.
- Calcolare l'assortatività per altri attributi
- Osservare il meccanismo di chiusura delle triadi