

# Analisi della piattaforma di rete sociale Facebook

Massimiliano Leone

[massimiliano.leone@studio.unibo.it](mailto:massimiliano.leone@studio.unibo.it)

[maximilianus@gmail.com](mailto:maximilianus@gmail.com)

Facoltà di Informatica Magistrale  
Sistemi Complessi, 2013

**Abstract.** Il presente documento illustrerà uno studio condotto su sul social network Facebook: a partire da un corpus di messaggi pubblici raccolti dal profilo di circa 400 utenti, e da alcune informazioni pubbliche sugli utenti medesimi, verrà mostrato il confronto tra la visione superficiale dell'ecosistema Facebook, ovvero la rete delle “amicizie”, e il reale comportamento di questo ecosistema, ovvero la rete delle interazioni tra gli utenti del campione. L'analisi è stata effettuata confrontando numericamente (con opportune metriche) e graficamente 2 grafi, generati a partire dalle due differenti e complementari “immagini”.

**Keywords:** facebook, graph theory, complex systems, java

## 1 Introduzione: il modello

### 1.2 La raccolta dei dati

Facebook è una piattaforma web/mobile che offre un servizio di rete sociale, permettendo lo scambio di messaggi/contenuti tra gli utenti affiliati, in particolar modo tra coloro che sono parte mutuamente della lista di “amici”. Regole opportunamente configurate dall'utente (ma sempre all'interno dei vincoli globali) permettono l'interazione e la visibilità dei propri contenuti anche a terze parti, che possono essere “amici di amici” o contatti con cui non si ha alcun legame intermedio (dunque, una visibilità pubblica).

La raccolta dei dati è stata effettuata a partire da un utente centro-stella, poiché Facebook non permette l'accesso ai contenuti se non si è iscritti al sistema; e, poiché

questo studio è focalizzato su un ecosistema di utenti sufficientemente stabile, il centro-stella preso come esempio è l'account dell'autore, e la lista dei suoi contatti “amici” ha fornito un numero sufficiente (~400) di utenti per ottenere un modello ragionevole.

Il client testuale da console `fbcmd` (disponibile su <http://fbcmd.dtompkins.com/>) è il tool che ha permesso di accedere alle informazioni necessarie, in particolare:

- informazioni generiche sugli utenti (sesso, numero globale di amici [se disponibile], numero di amici in comune con l'utente centro-stella)
- informazioni sui messaggi pubblici (da qui in poi “post”) che ogni utente pubblica nel proprio spazio (da qui in poi “feed” o “bacheca” o “wall”) o nello spazio dei propri contatti/amici: di questi messaggi sono state collezionate altre ulteriori informazioni, ovvero:
  - i commenti ricevuti (ogni post ammette commenti e/o apprezzamenti, generando un thread di messaggi)
  - gli apprezzamenti ricevuti (da qui in poi “like”)
  - il numero (se presente) di ripubblicazioni di quel messaggio da parte di altri utenti (da qui in poi “resharing”)
  - le menzioni ad altri utenti all'interno del corpo del messaggio (da qui in poi “tag”), che fan sì che arrivi una notifica all'utente “taggato”, in modo che sappia che esiste un messaggio per lui
  - il “luogo” del messaggio, ovvero la propria “bacheca” o quella di un contatto/amico

Il tool `fbcmd` è scritto in PHP e fornisce output direttamente in console: è stata necessaria quindi qualche piccola modifica per poter ottenere gli output in file JSON, in modo da poterli trattare successivamente.

### 1.3 La costruzione del modello

A partire dal corpus di output JSON, è stato scritto un opportuno parser in Java che ne estrapolasse i contenuti in maniera opportuna, e da questi è stato costruito un modello di dominio, le cui classi principali sono sotto brevemente dai relativi

diagrammi di classe illustrato con gli attributi di classe, le cui classi partecipanti sono: User, Interactions, Post, e Comment.

La classe User rappresenta un utente all'interno di Facebook, e i suoi attributi membro mantengono alcune informazioni “statiche” (sesso, età, userid, ecc) così come alcune relative alla sua attività, ovvero ai propri posts, all'apprezzamento/ricondivisione di questi ultimi da parte degli altri utenti, e alle interazioni che invece l'utente ha con i propri amici. Esistono due specializzazioni di User: Ego e FriendOrAlike: la prima classe rappresenta l'utente centro-stella (“my user”), mentre l'altra rappresenta gli amici e/o gli amici degli amici; la distinzione è necessaria perché Facebook non permette di ottenere la lista degli amici dei propri amici ma, al più, la lista degli amici in comune (“mutual friends”): quindi, l'utente Ego mantiene un campo istanza con la lista dei propri amici, mentre FriendOrAlike manterrà una differente lista con i mutual friends.

Le informazioni statiche sono mutate dalla reference ufficiale per la classe “user”:

<https://developers.facebook.com/docs/reference/fql/user>

La classe Interactions modella le attività tra gli utenti, incapsulando il numero dei commenti, dei *likes* e dei *tags*, nonché la lista dei *posts* verso uno specifico utente.

Infine, la classe Post racchiude, com'è intuibile, tutti i dati correlati al singolo post, come l'id dell'autore del post (actorId) e l'id del destinatario (sourceId, con “source” inteso come la bacheca [wall] che ospita il post), le informazioni sui commenti al post, i tag che dell'autore verso altri utenti, i likes al post (e quindi gli id degli utenti che hanno apposto il like, etc.

La reference ufficiale è disponibile qui:

<https://developers.facebook.com/docs/reference/fql/stream/>.

Analizzando un post è quindi possibile sapere se l'autore è il proprietario medesimo della bacheca, o se un suo amico: e, ovviamente, se questo secondo soggetto non è parte della lista di amici dell'utente centro-stella, è ovviamente un “amico dell'amico”.

Infine, la classe Comment contiene informazioni sul post, tra le quali, di rilievo spicca sempre l'id dell'autore del commento, utile per la medesima distinzione “amico” vs “amico dell'amico” di cui già sopra. [Reference:

<https://developers.facebook.com/docs/reference/fql/comment/>]

Alla conclusione della fase di parsing, effettuata su circa 150 posts ottenuti per utente (è il limite massimo ammesso da Facebook) per circa 400 (esattamente: 417) utenti, è popolata una classe “holder” globale, denominata “World”, che contiene il riferimento all'utente centro-stella, Ego, e due liste, l'una di amici, l'altra di “amici di amici” (una distinzione possibile sfruttando il confronto tra `sourceId` e `actorId`, come già pure accennato nel paragrafo precedente).

#### 1.4 La generazione dei grafi

Il tool utilizzato per la generazione dei grafi, nonché per la loro analisi, è GraphStream: <http://graphstream-project.org>.

Permette la creazione di grafi da codice, così come di importarli da file, di salvarli in diversi formati più o meno standard (in questo caso è stato adottato `graphml`), di visualizzarli con un'estetica anche gradevole (la personalizzazione degli stili è permessa da fogli `css`), di manipolarli durante l'utilizzo, mantenendo un riferimento in tempo reale tra la rappresentazione in memoria del grafo e quella nel frame nel quale visualizzato; infine, fornisce un certo numero di algoritmi per effettuare metriche note.

Come già introdotto nell'abstract, si vuole confrontare il “mondo” delle amicizie (ovvero dei legami tra gli utenti) con quello delle interazioni (ovvero la reale attività che si svolge tra di essi).

Il “mondo” delle amicizie può ben essere modellato con un grafo indiretto, dagli archi non pesati: senza addentrarsi in aspetti sociologici sul “peso” dell'amicizia, e attenendosi invece al concetto che si ha nell'ecosistema di Facebook, il legame di amicizia è bidirezionale e non ha peso (inteso come valore di `legame/arco`): per coerenza rispetto a quanto si vedrà successivamente, è stato tuttavia assegnato un peso unitario ad ogni legame.

La generazione di questo grafo è piuttosto banale: iterando sulla lista di amici, è creato un nodo per ogni utente, un legame con l'utente centro-stella Ego, e un legame con ognuno dei suoi amici in comune con quest'ultimo (i “mutual friends”). Avendo a disposizione anche il numero totale degli amici che ogni utente ha, è stato utilizzato questo valore come attributo di nodo, che graficamente si manifesta nella dimensione del nodo suddetto (ovvero nel raggio della circonferenza rappresentante il nodo).

Il grafo delle interazioni richiede un algoritmo più elaborato: sempre a partire dalla lista degli “amici”, per ognuno di essi si ottiene il riferimento agli amici con cui esistono interazioni, e con essi viene creato un arco rappresentante il legame. L'arco è pesato, e questo peso è la somma delle diverse interazioni, ognuna delle quali moltiplicata per un fattore che rappresenta l'importanza dell'interazione.

La tabella 1, in basso, riassume sinteticamente i pesi assegnati:

<i><b>Tipo legame</b></i>	<i><b>Peso</b></i>
<i>Amicizia</i>	<i>1</i>
<i>Like</i>	<i>2</i>
<i>Resharing</i>	<i>3</i>
<i>Commento</i>	<i>4</i>
<i>Tag</i>	<i>5</i>
<i>Post su bacheca altrui</i>	<i>6</i>

Lo snippet di pseudo-codice che segue illustra il metodo che restituisce il suddetto valore di legame:

```
--  
private float getWeightedInteractions(Interactions  
toOtherUserInteractions) { return  
toOtherUserInteractions.getLikesCount()*InteractionsWeights.LIKE  
+toOtherUserInteractions.getCommentsCount()*InteractionsWeights.  
.COMMENT  
+toOtherUserInteractions.getTagsCount()*InteractionsWeights.TAG  
+toOtherUserInteractions.getPostsCount()*InteractionsWeights.POS  
T; }  
--
```

Utilizzando pesi diversi per diversi tipi di interazione, si vuole quindi dare priorità a certi tipi di attività dell'utente rispetto ad altre: è indubbio che, rispetto ad un “like” apposto a un post, un commento ha un peso maggiore, perché rappresentante una maggiore intenzione di voler interagire con l'autore del post – e, ovviamente, un post sulla bacheca di un amico ha il peso maggiore di tutti, perché sussiste la volontà dell'autore del post di interagire con l'altro utente tale di visualizzare la di lui bacheca e lasciare un messaggio (mentre, per i commenti / like / resharing / tags, non è necessario aprire la pagina personale dell'utente amico, ma è sufficiente visualizzare il flusso globale dei messaggi di tutti, la cosiddetta “Facebook Home” per poter leggere i suddetti messaggi, ed eventualmente interagire).

È da sottolineare che tra 2 utenti possono non sussistere lo stesso numero di interazioni, e, a loro volta, queste possono avere un diverso peso: non può dunque esistere un solo arco per rappresentare l'attività tra due utenti, ma ne sono necessari due (qualora, ovviamente, esistano attività da entrambe le parti, verso entrambi). Il grafo è quindi, oltre ad essere di tipo diretto e pesato, un multigrafo.

Le figure 1.1, 1.2, 1.3, 1.4 illustrano i grafi suddetti: in particolare 1.1 e 1.2 sono, rispettivamente, il grafo delle amicizie e quello delle interazioni; in 1.3 e 1.4 sono invece mostrati i medesimi grafi dopo la rimozione dell'utente Ego - un'operazione necessaria per meglio comprendere quanto quest'ultimo influisca sulla topologia della rete nel suo complesso.

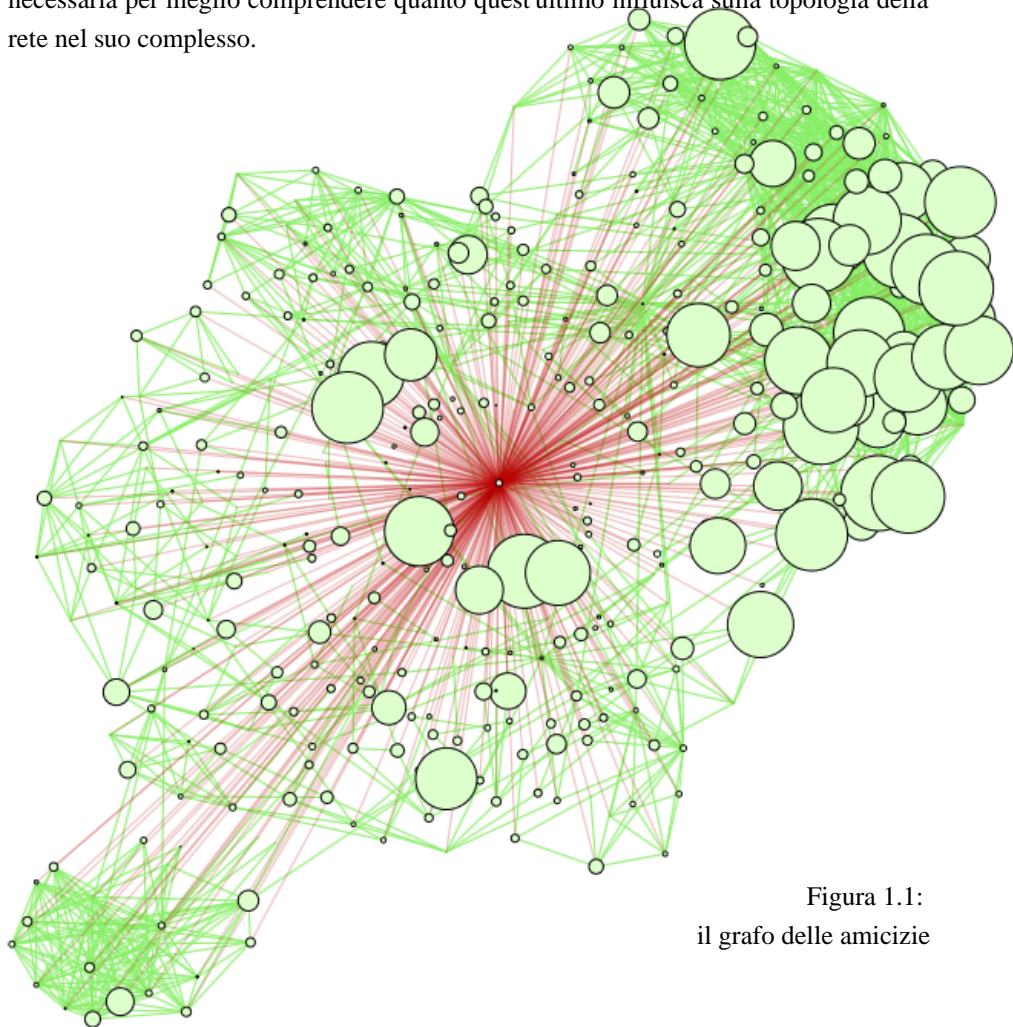


Figura 1.1:  
il grafo delle amicizie

In figura 1.1, il grafo (indiretto, non pesato) della rete di amicizie: il nodo da cui si diramano gli archi in rosso è il centro-stella, l'utente Ego; la dimensione dei nodi è data dal numero di amici di ogni utente.

In figura 1.2, il grafo delle interazioni: al centro, il nodo rappresentante l'utente Ego; gli archi di colore giallo sono entranti nel nodo Ego, e rappresentano le interazioni verso questo utente, mentre gli archi rossi, uscenti dal nodo Ego, sono le interazioni originatesi dall'utente corrispettivo, verso gli amici; gli archi verdi, infine, sono le interazioni tra gli amici medesimi (per semplicità, non sono stati utilizzati diversi colori per contraddistinguere il verso).

È notevole la presenza di molti nodi isolati, senza alcun legame con altri nodi, e di qualche cluster di rilevanza minima, composto da 2-3 nodi. Il cluster di interesse è ovviamente quello comprendente l'utente Ego.

Gia da questa prima analisi grafica si evince una notevole differenza tra i due “mondi”, a dimostrazione quindi della ben farraginosità del legame di “amicizia” che il social network propone.

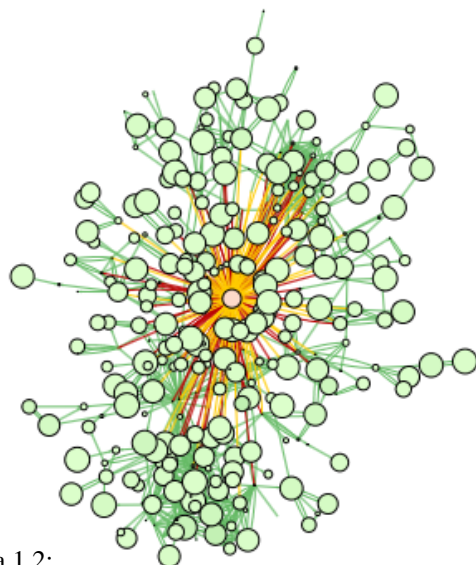


Figura 1.2:  
Il grafo delle interazioni

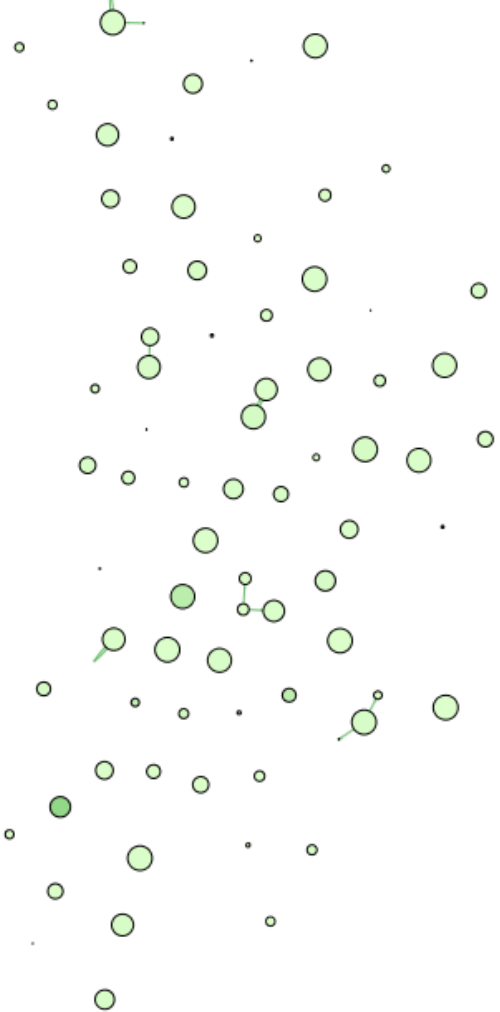


Figura 1.3:  
il grafo “friendships” dopo la  
rimozione del nodo Ego centro-stella.

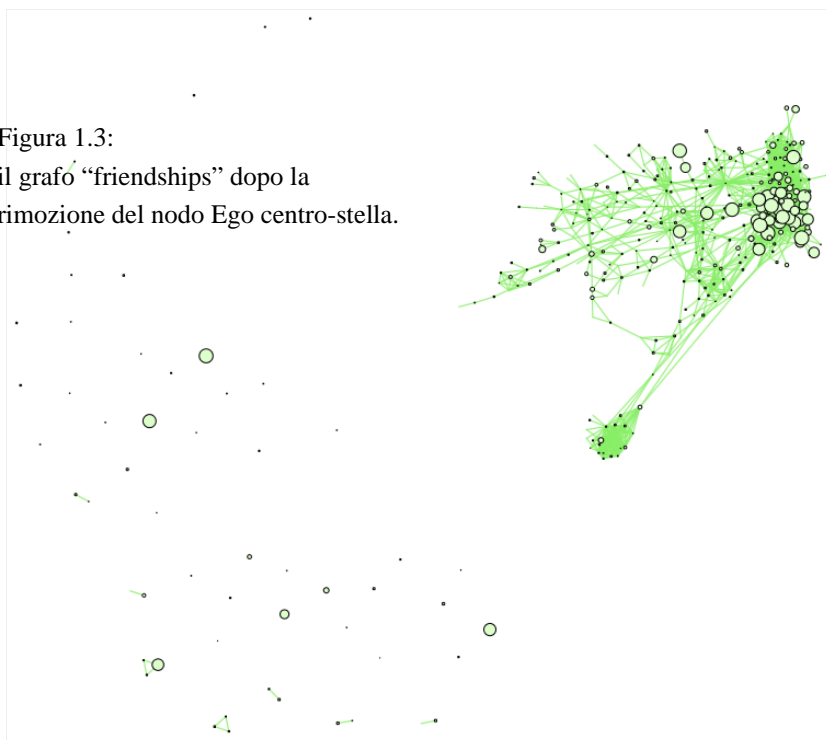
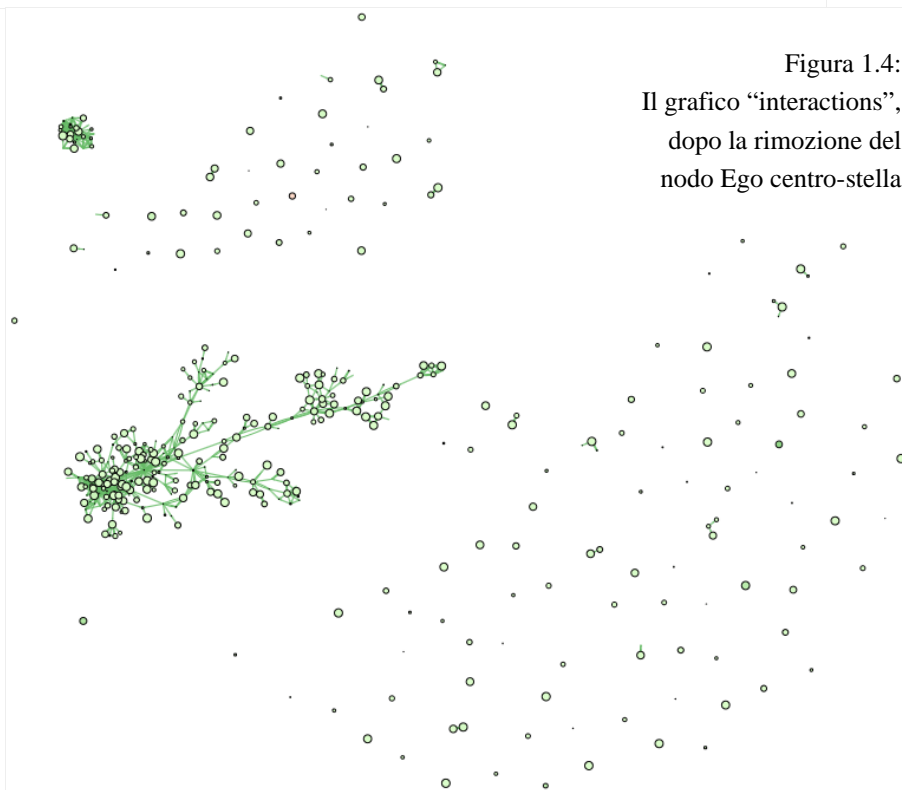


Figura 1.4:  
Il grafico “interactions”,  
dopo la rimozione del  
nodo Ego centro-stella.





## 2 I test

Saranno ora mostrati i risultati di alcuni test di metriche, eseguiti sui 4 grafi precedentemente illustrati, tra cui:

- grado, diametro, densità
- misure di centralità
- misure di connettività

### 2.2 Degree distribution

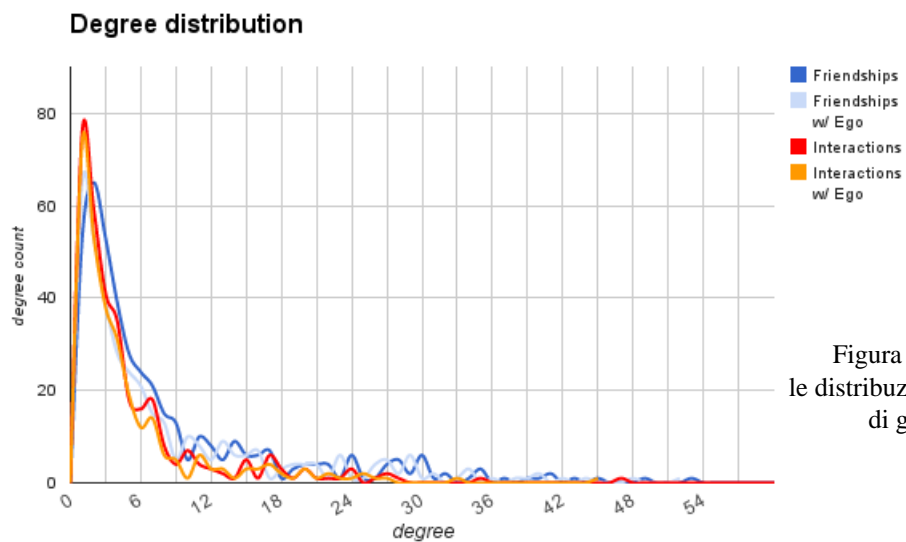


Figura 2.1:  
le distribuzioni  
di grado

Le distribuzioni di grado, in figura 2.1, si presentano nella forma di una curva power-law asimmetriche, coerente con il fatto che le reti sono scale-free. La conferma è ulteriormente data dalla presenza della lunga coda nella visualizzazione dei medesimi grafici su scala logaritmica, in figura 2.2. Le curve hanno ovviamente una discesa più ripide nel caso della rete “interactions”, come pure è intuibile osservando le immagini dei relativi grafi, nelle pagina precedenti (figure 1.2 e 1.4), che presentano un numero non trascurabile di nodi isolati.

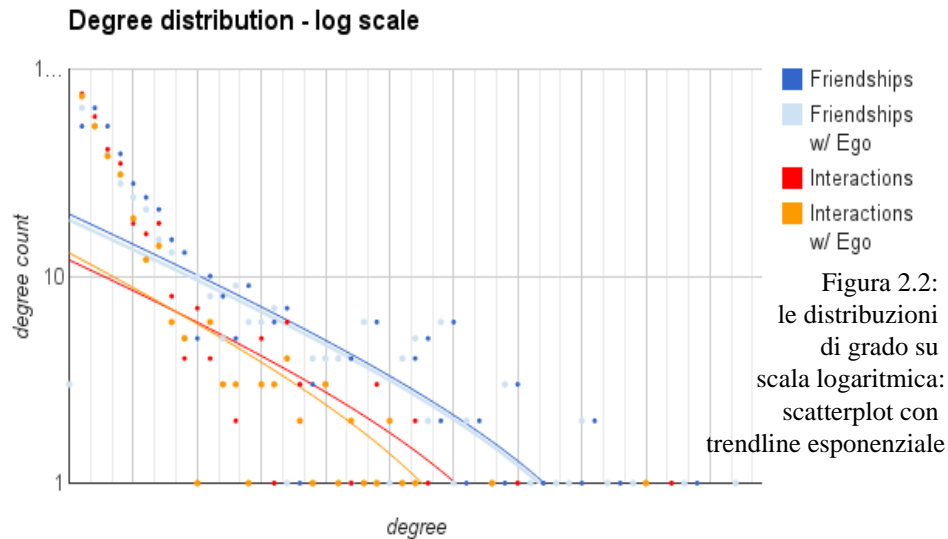


Tabella 2: I risultati di alcune metriche e di algoritmi di “communities detection”

	<i>Friendships (w Ego)</i>	<i>Interactions (w Ego)</i>	<i>Friendships (w/ Ego)</i>	<i>Interactions (w/ Ego)</i>
<i>Density</i>	0.220	0.0183	0.2279	0.0176
<i>Diameter</i>	2	12	12	14
<i>Degree Mean</i>	9.38	5.913	8.456	5.153
<i>Maximum Clique (including Ego)</i>	12	11	8	7
<i>Maximum degree (excluding Ego)</i>	53	47	48	33
<i>Connected Components</i>		9		17
<i>Giant Connected Component Nodes</i>		354		234

In figura 2.3 il grafico delle distribuzioni del coefficiente di clustering, per le 4 casistiche. È interessante notare come, nel grafico “friendships” mancante dell'utente Ego, la componente più vicina allo zero abbia valore quasi del doppio, rispetto al grafico originario – a dimostrazione di quanto il nodo Ego sia (il solo?) responsabile dei legami tra gli utenti. Mentre, per le reti “interactions”, si riscontrano distribuzioni simili – a dimostrazione, in questi diversi casi, che la capacità di clustering di reti modellanti le “volontà” di interazione prescindono dai contatti “hub”.

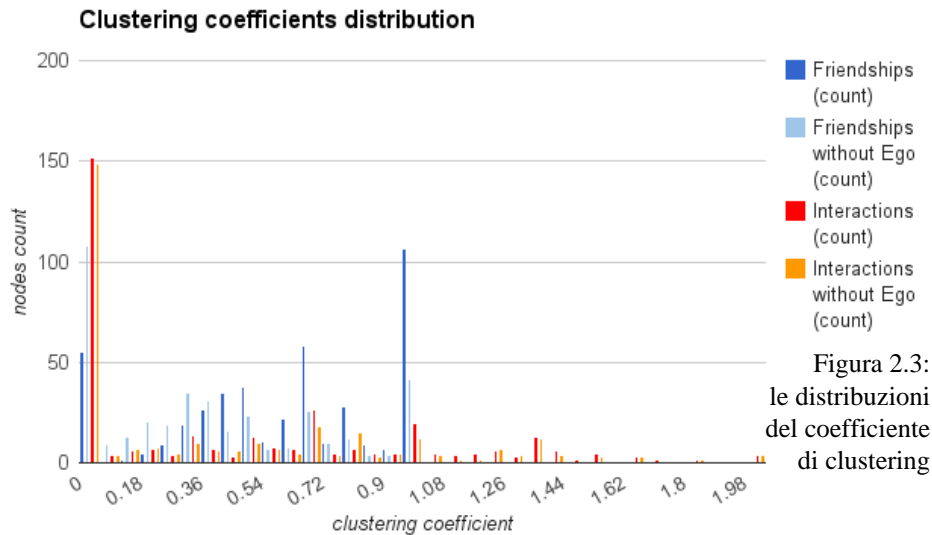


Figura 2.3:  
le distribuzioni  
del coefficiente  
di clustering

### 2.3 Misure di centralità

La centralità di grado, per le 4 casistiche, in figura 2.4: l'utente Ego è stato non incluso nella generazione dell'istogramma, perché lo avrebbe reso poco visibile, con i rispettivi valori di 425 (per “friendships”) e 203 (per “interactions”). Per quanto indubbio sia il fatto che l'utente centro-stella contribuisca al valore di degree-centrality degli altri utenti, si è voluto analizzare l'andamento dei soli valori di questi ultimi: e un risultato interessante (oltre al fatto che le curve si approssimano come power-law) è la poca differenza tra i 2 picchi massimi di colore azzurro e arancione, corrispondenti ai grafi in cui è stato rimosso il nodo relativo a Ego, rispettivamente “friendships” e “interactions”.

Questo mette in evidenza il seguente fenomeno: persone che hanno pochi legami (o meglio: hanno pochi legami in questa rete), tendono a “coltivarli” con un numero elevato di interazioni – mentre si noti la differenza dei valori nelle aree centrali del grafico, per valori di  $20 \leq \text{degree} \leq 36$ .

In figura 2.5, l'istogramma della centralità di betweenness. È evidente come la maggior parte dei nodi (più dell'80%: 351/426 per “friendships” e 284/326 per “interactions”) abbia un valore di betweenness pari quasi a zero: questo, insieme all'analisi della centralità di grado di poc'anzi, dimostra una delle principali ipotesi alla base di questo studio, ovvero che, nonostante Facebook abbia aumentato la “capacità” sociale delle persone (che ne fanno uso), queste ultime tendono sempre a

concentrarsi in “isole” sufficientemente coese e vieppiù isolate dal resto del sistema, in un microcosmo dove ci si conosce tutti (e si vuole conoscere quei tutti, o quasi). È sufficiente anche guardare il grafico “interactions” privato nel nodo Ego, e di come, esclusi i nodi isolati e qualche gruppo di 2 o 3 nodi, rimangano due cluster, l'uno minuscolo, l'altro massivo. Fermo restando che 425 utenti è forse un numero troppo basso per un'analisi del genere (ma la lista degli “amici” dell'autore/utente centro-stella/Ego è così composta), e che sarebbe molto più interessante avere risultati su una rete di >4000 utenti (il limite massimo di “amici” in Facebook è 4999).

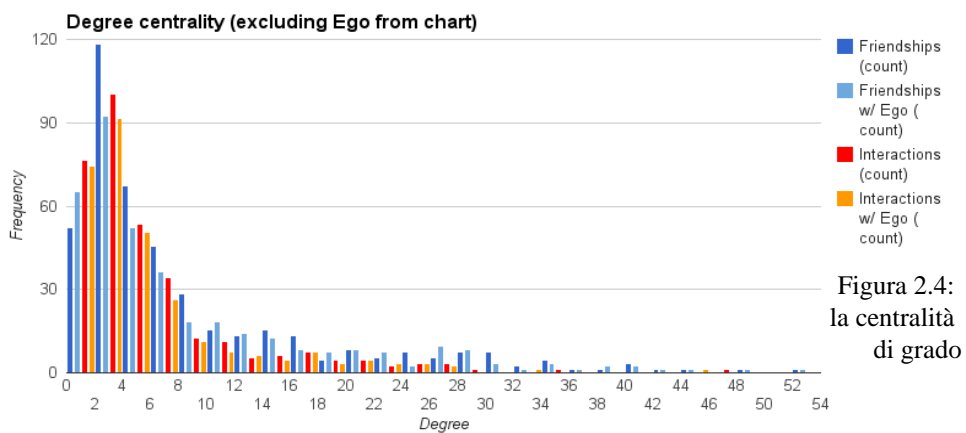


Figura 2.4:  
la centralità  
di grado

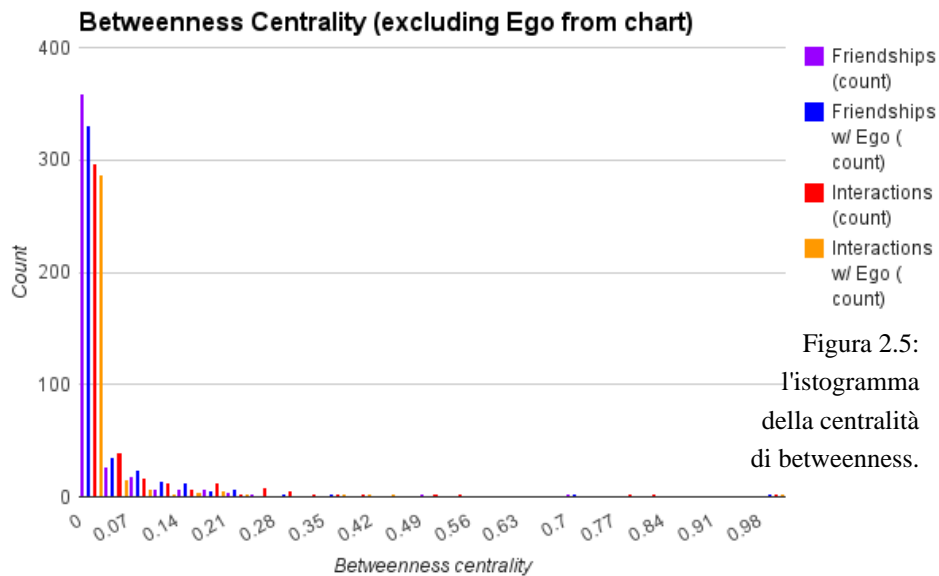


Figura 2.5:  
l'istogramma  
della centralità  
di betweenness.

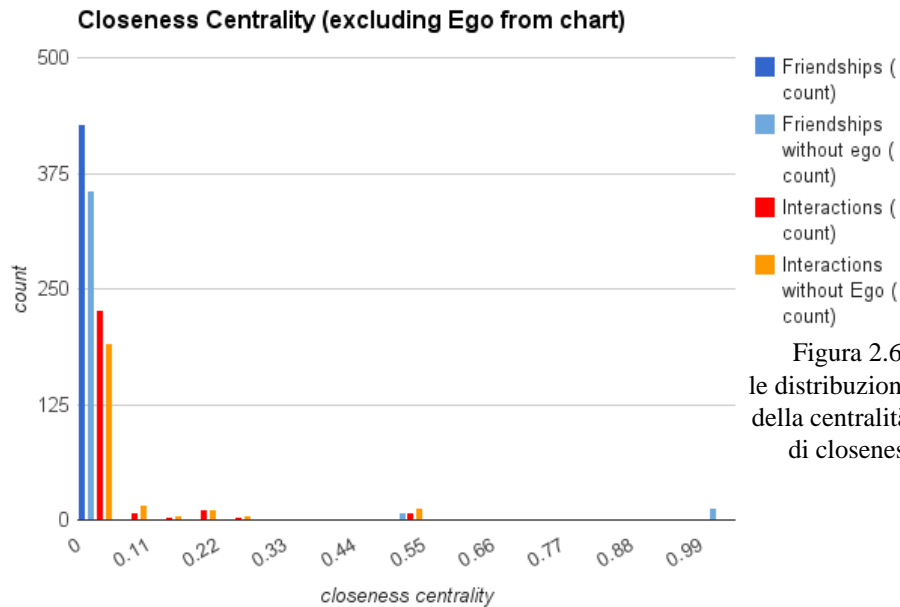


Figura 2.6:  
le distribuzioni  
della centralità  
di closeness

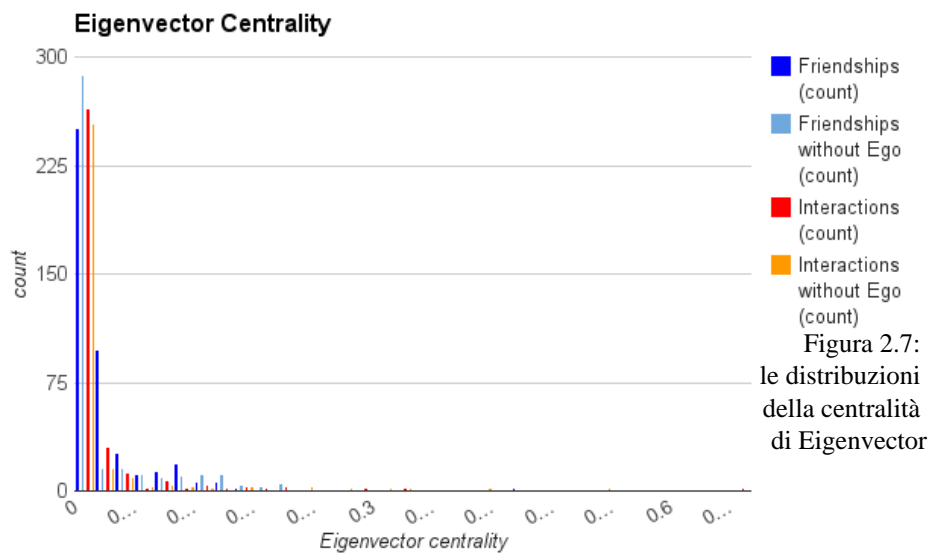


Figura 2.7:  
le distribuzioni  
della centralità  
di Eigenvector

In figura 2.6 l'istogramma della closeness centrality mostra un'alta concentrazione di valori tendenti allo zero: ciò è dovuto alla presenza di moltissimi nodi isolati, e di molti nodi “foglia” - come pure è visibile dalle immagini dei 4 grafi, nelle pagine precedenti.

In figura 2.7, la eigenvector centrality che, si ricorderà, esprime quanto un nodo sia influente rispetto agli altri, considerando la sua connettività nonché quella dei suoi vicini, e iterativamente dei vicini dei vicini, eccetera.

I valori ottenuti sono sempre bassi, per la maggior parte dei nodi, e ciò è sempre dovuto alla maggiore presenza dei nodi foglie, nonché dei molti isolati.

Inoltre, c'era da aspettarsi il fatto che la rete contraddistinta dalla maggioranza dei nodi caratterizzati da un basso valore di questa centralità fosse proprio la “friendships without ego”: mancando il nodo Ego, gran parte degli archi vengono meno, e la connettività generale si abbassa di molto; inoltre, rispetto alle reti “interactions”, la “friendships”, benchè presenti un unico grande cluster, è molto meno densa dell'altra.

## 2.4 Communities Detection

In ultimo, alcuni test sulla communities detection:

- sull'esistenza di sottografi con coppie di vertici connessi da cammini:
  - Friendships rivela avere 9 componenti, con la massima (giant) composta da 354 nodi
  - Interactions rivela avere 17 componenti, con la giant composta da 234 nodi
- sull'esistenza di “cricche” (*cliques*), ovvero di sottografi completi: il termine è mutuato dalla sociologia, e riferisce ad un gruppo di persone che si conoscono tutte tra loro.

L'analisi dei i grafi riporta i seguenti risultati per la cricca massima:

- Friendships: 11 nodi VS Interactions: 7 nodi

È stato escluso Ego, ovviamente (il quale risulterebbe certamente all'interno della cricca, aumentandone di 1 unità il valore, ma sarebbe poco significativo).

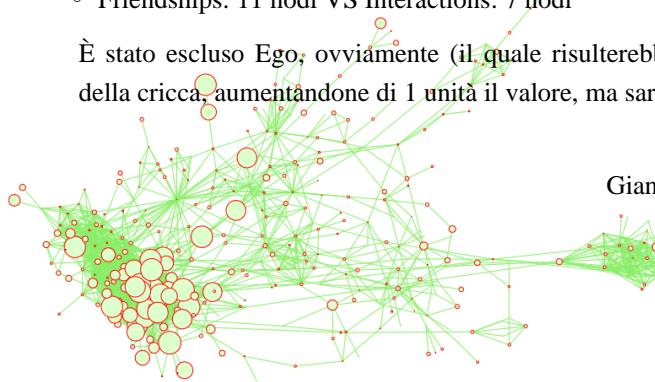


Figura 2.8:  
Giant Component Connected  
per Friendships

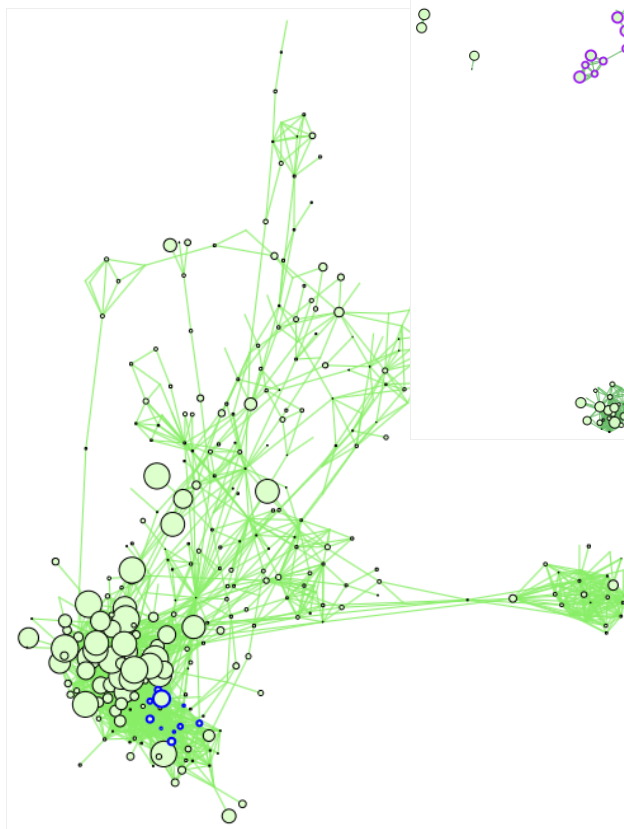


Figura 2.10:  
maximum clique (in blu)  
per Friendships

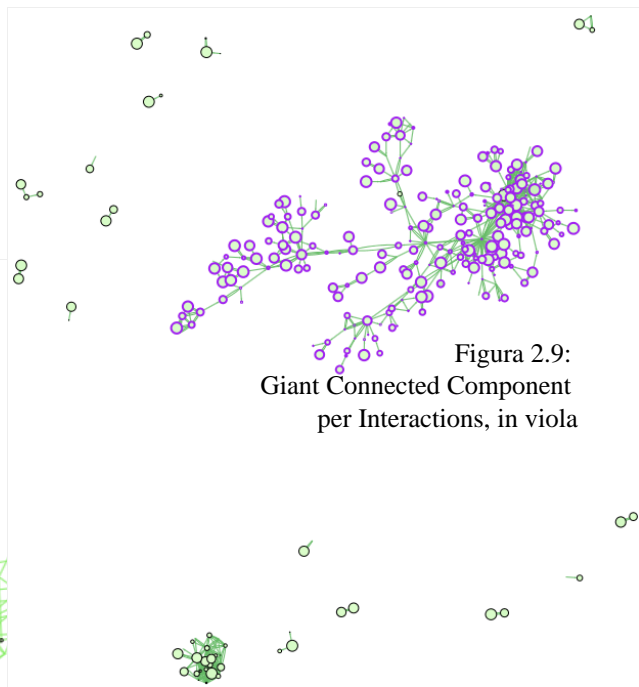


Figura 2.9:  
Giant Connected Component  
per Interactions, in viola

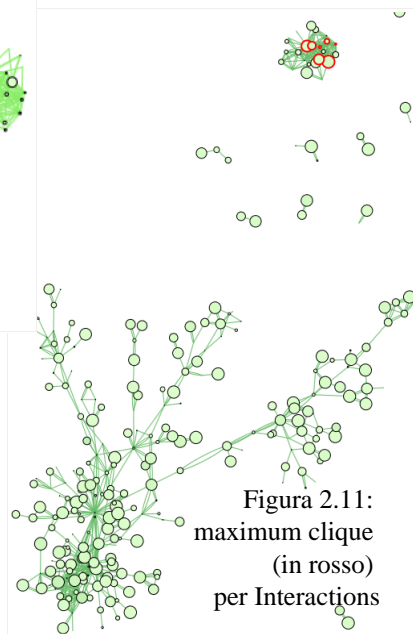


Figura 2.11:  
maximum clique  
(in rosso)  
per Interactions

È interessante notare, nei due grafici riportanti le cricche massime (figure 2.10 e 2.11), come la cricca massima in Friendships sia, dopo la rimozione di Ego, ancora un componente connesso al gruppo principale, mentre questo non vale per il grafo Interactions: è evidente come Ego sia l'unico nodo hub per quel gruppo rispetto al resto del grafo (in effetti, dai risultati numerici alla base degli istogrammi, questo nodo era risultato avere un valore di un ordine superiore per la betweenness centrality, rispetto al massimo di tutti gli altri – in entrambe le reti).

### 3 Conclusione

Il presente documento ha illustrato un tentativo di studiare il social network “Facebook” e certuni fenomeni sociologici ad esso correlati. A partire da cospicua raccolta di dati, sono stati elaborati 2 grafi che modellassero la rete delle “amicizie” e la complementare rete delle “interazioni”, ovvero ciò che effettivamente avviene tra gli utenti. Sono state confermate alcune ipotesi, tra cui il fatto che i legami di “amicizia” (modellati nella prima rete) non corrispondono affatto nella volontà di interagire, tra medesimi utenti (la seconda rete); inoltre, si sono rivelate altre interessanti caratteristiche, ipotizzate empiricamente dall'autore durante l'uso quotidiano del social network, ovvero il fatto che, nonostante quest'ultimo abbia aumentato la capacità/volontà di relazioni sociali (abbassando anche la “distanza” già nota da Milgram con la teoria dei “sei gradi di separazione”, da 6 a 4.5), gli utenti tendono a formare delle “isole” coese, e mantenere contatti giusto all'interno di esse, andando così a creare reti dai legami ancora più forti.

### Bibliografia

<http://sferrett.web.cs.unibo.it/CAS/SLIDES/lucidi.html>

<http://www.scienze.unibo.it/it/corsi/insegnamenti/insegnamento/2014/384800>

<http://graphstream-project.org/>

### Disclaimer

Il presente documento è rilasciato con licenza Creative Commons By-Nc-Sa:

<http://creativecommons.org/licenses/by-nc-sa/2.0/it/>



Il codice, rilasciato con licenza GPL, è disponibile al seguente URL:

<https://github.com/k0smik0/FaCRI>

Parti di rilievo sono i generatori dei grafi, nonché gli analizzatori (entrambi basati su graphstream); link al codice dei suddetti componenti sono facilmente raggiungibili dal post-scriptum del seguente readme:

<https://github.com/k0smik0/FaCRI/blob/master/README.md>

Infine, un foglio di calcolo con i risultati numerici, e gli istogrammi interattivi presenti nelle precedenti pagine è disponibile qui:

[https://docs.google.com/spreadsheets/d/1BspRfHXXd83dZzTLOPvAN\\_BeJsZx\\_L5u5MR2KT1\\_hQE/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1BspRfHXXd83dZzTLOPvAN_BeJsZx_L5u5MR2KT1_hQE/edit?usp=sharing)