

Diffusione di una notizia nei Social Network

Gianluca Iselli

Dept. of Engineering and Computer Science
University of Bologna
Bologna, Italy
gianluca.iselli@studio.unibo.it

Abstract. Questo lavoro è stato svolto come progetto e relazione per l'esame di Sistemi Complessi.

Il lavoro qui presentato è stato ispirato da una notizia della NBC¹ del 30-12-2014. Nell'articolo si racconta di una maestra dell'Oklahoma che ha voluto dimostrare ai suoi studenti, ragazzi di dodicianni, la velocità di diffusione delle fotografie sui social network. La maestra, Melissa Bour, ha condiviso su Facebook un'immagine², che cita “Caro Facebook, [...] i miei alunni di 12 anni pensano “non è nulla di che” pubblicare immagini di loro stessi[...]. Per favore aiutatemi condividendo e commentando con la vostre città di origine. In questa maniera [...] capiranno quanto velocemente le loro immagini possano fare il giro del mondo. [...]”.

Successivamente Fox 23 ha ricondiviso, sulla pagina Facebook del giornale, una copia della fotografia della maestra. In un'ora e mezza l'immagine ha raggiunto le 1300 condivisioni ed ottenuto quasi 1000 “mi piace”.

Riuscire ad interpretare le proprietà della struttura e degli algoritmi di una rete complessa gioca un ruolo fondamentale in molti campi dei *sistemi complessi*.

Lo studio di queste caratteristiche è spesso richiesto per argomenti importanti come i seguenti:

- Propagazione di una epidemia o pandemie tra la popolazione;
- Diffusione di un virus informatico tra dispositivi mobili e non;
- Divulgazione di una notizia in un piccolo paese, o più in grande tra massmedia e pubblico o, ancora più estesa, in un Social Network

In questo periodo storico i Social Network hanno un successo globale e non è difficile credere che una notizia, su servizi così diffusi, possa fare il giro del mondo.

L'argomento di studio discosta leggermente dalle intenzioni della maestra, infatti, si tratta di un'analisi e di una misurazione di più parametri legati alle reti sociali.

Nell'Introduzione (Capitolo 1) verranno elencati i punti trattati all'interno di questo abstract.

Nel Capitolo “Implementazione” (Capitolo ??) si approfondiranno, poi, le modalità di sviluppo del progetto.

Mentre invece, nel Capitolo “Analisi dei Risultati” (Capitolo 3) sarà presentato quanto elaborato, ovvero il prodotto del lavoro svolto.

¹ Notizia sul sito della NBC: <http://www.nbcbayarea.com/news/national-international/Teachers-Social-Media-Lesson-Goes-Viral-on-Facebook-287159381.html>

² Immagine condivisa dalla maestra, Melissa Bour: http://i.dailymail.co.uk/i/pix/2015/01/04/2466352100000578-2895903-image-a-74_1420339978179.jpg

1 Introduzione

Il capitolo corrente fa da introduzione al lavoro svolto e orienta il lettore ad alcuni aspetti principali che interessano il campo del “Spreading Rumors”.

I primi studi riguardanti la diffusione di una notizia risalgono ai primi anni '60 [1], ma quelli presi in considerazione in questo lavoro sono più attuali e riguardano il campo dei Social Networks.

Lo studio della propagazione di una “voce”, definita più semplicemente notizia o informazione, serve ad analizzare alcuni comportamenti, parametri e modelli di una rete sociale. Le caratteristiche principali sono la topologia della rete e gli utenti che ne fanno parte.

Possiamo quindi iniziare a dire che le simulazioni prevederanno una rete di utenti che “a contatto” con la notizia decideranno a loro volta se ricondividerla o solamente prendere atto della sua esistenza. Al passo 0 della simulazione verrà scelto un utente a cui verrà “insegnata” l’informazione che dovrà essere propagata. Facendo un riferimento al tema della medicina, durante un’indagine epidemiologica, il primo paziente ad aver contratto la malattia viene chiamato paziente zero. Nel nostro caso non si tratta di una malattia ma di una notizia.

Durante le simulazioni verranno inoltre utilizzati alcuni termini che definirò qua di seguito:

- “Ignorants” : al momento della creazione della rete sociale tutti gli utenti vengono definiti in tal modo perchè non sono al corrente della notizia;
- “Spreaders” : tutti gli utenti che decidono di condividere la notizia. Il “paziente zero” fa parte di questo gruppo;
- “Uninterested” : tutti quegli utenti che dopo essere diventati consapevoli dell’esistenza della notizia decidono comunque di non condividerla;
- “Viewers” o visualizzatori : tutti coloro facenti parte del gruppo degli Spreaders e degli Uninterested.

Facendo una veloce panoramica, l’argomento di questo progetto è certamente un ottimo ambiente di studio ed il numero esorbitante di utilizzatori di Social Networks in circolazione crea sicuramente un terreno fertile per la diffusione di notizie ed informazione.

1.1 Obiettivi

Per questo lavoro possiamo definire tre obiettivi principali.

Il primo test permetterà di decidere il modello topologico più consono, tra quelli presi in considerazione, per i successivi test.

Il secondo servirà per mettere in luce come una notizia, con un topic adatto maggiormente ad una certa fascia d’età, venga diffusa nei differenti Social Networks. E’ presente una statistica online piuttosto recente (3°trimestre 2014) che mostra la distribuzione per età di come è divisa l’utenza nei social networks più popolari. Il grafico in figura 1 ne mostra la distribuzione.

L’ultimo obiettivo, invece, servirà ad analizzare un’interazione tra 2 diversi gruppi di utenti, che permetterà di studiare il numero di visualizzazioni della

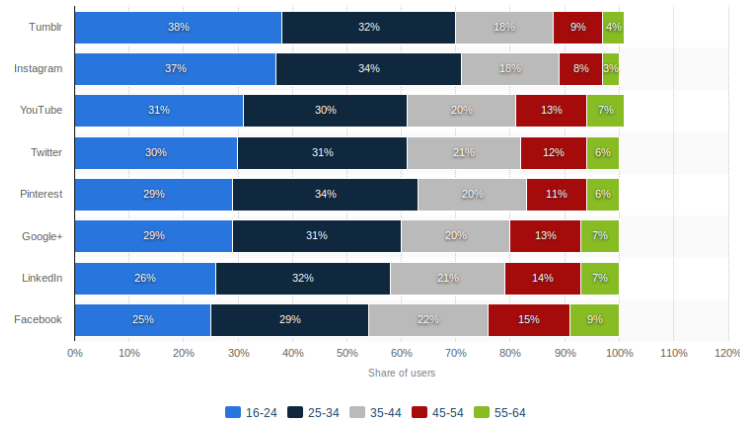


Fig. 1: Distribuzione delle età divisa per Social Network [4]

notizia in casi più complessi. Un esempio potrebbe essere quello di voler dividere il numero totale delle persone in due sottoinsiemi così formati:

- Il primo gruppo è formato da pochi utenti, tipo il 20% del totale, ma ogni componente ha un'ottima probabilità di condivisione.
- Il secondo gruppo, viceversa, è formato dall'80% degli utenti, ma ogni componente ha una possibilità minore di condivisione.

2 Progettazione e Implementazione

Per favorire un modello da studiare che permetta di risolvere gli obiettivi descritti pocanzi sono state prese le decisioni qui di seguito elencate.

Iniziamo con il dire che il social network verrà astratto ad un grafo scale-free dove ogni nodo è un utente che possiede alcune caratteristiche.

Per il primo obiettivo verrà posta l'attenzione su due algoritmi, impiegati per la creazione di grafi, che formano modelli di rete differenti.

Inizialmente verrà fatto vedere come una notizia viene propagata in un grafo di tipo "Preferential Attachment" suggerito da Barabási e Albert [2].

Le simulazioni poi proseguiranno con un'altra topologia di grafo, sempre scale-free con Power Law Degree, descritta però dall'algoritmo di Dorogovtsev e Mendes [3].

Il lavoro non parte da dati reali e la scelta di queste topologie di grafi è data dalla peculiarità di alcune loro caratteristiche che elencherò qui di seguito:

- La topologia di grafo Preferential Attachment:
 - è un modello ampiamente utilizzato per la sua semplicità;
 - è impiegato da buona parte di studi che trattano argomenti somiglianti lo "Spreading Rumors";

- non è un grafo di tipo frattale [5] ma ha una caratteristica simile. Non possedendo cricche di almeno 3 nodi, se un nodo non condivide la notizia, tutti i suoi nodi “figli” non riceveranno mai l’informazione.
- La topologia di grafo definita da Dorogovtsev e Mendes invece:
 - ha un modello decisamente più complesso;
 - può avere all’interno del grafo cricche da 3 o più nodi e questo permette una probabilità maggiore di condivisione della notizia;
 - più somigliante³ ad una struttura reale di rete sociale.

Cercando di migliorare il modello totale della simulazione è stato deciso di fornire alcune proprietà alla notizia e agli utenti(Nodi) che compaiono nella rete sociale.

Osservando il secondo obiettivo dobbiamo fornire alla notizia un “argomento” e per fare ciò si è optato per l’inserimento di N valori che definiranno quanto è adatta la notizia per ogni fascia d’età.

Avendo a disposizione la distribuzione delle età, precedentemente mostrata nel grafico di figura 1, possiamo notare 5 intervalli di anni e quindi stabilire il valore di $N = 5$.

Dobbiamo perciò anche definire 5 gruppi di utenti per distinguere le differenti età. Il numero di persone appartenenti ad ogni gruppo verrà semplicemente definito dalla semplice proporzione:

$$N_NODI_TOTALI \ / \ 100 \ * \ \%_UTENTI_GRUPPO$$

dove $\%_UTENTI_GRUPPO$ è la percentuale presente nel grafico di figura 1.

A questo punto per la condivisione dell’informazione tra nodo e nodo manca solamente la formula per definire la probabilità della propagazione. L’idea è quella di dare la “possibilità” all’utente, come nella realtà, di decidere se la notizia gli interessa oppure no. Questo ci obbliga ad aggiungere un nuovo parametro all’utente, l’astensione alla notizia.

Così facendo la notizia verrà condivisa dall’utente ‘ i ’ con una certa età ‘ e ’ se:

$$\Rightarrow NOTIZIA_e \ > \ UTENTE_i.astensione$$

dove $NOTIZIA_e$ è la forza di condivisione della notizia su una data fascia d’età ‘ e ’, mentre $UTENTE_i.astensione$ è l’astensione alla notizia dell’utente i -esimo.

L’ultimo obiettivo agisce proprio su quest’ultima parte appena definita. Quando si parla di “forza della notizia” e di “forza di astensione” vogliamo considerare un numero tra 0 e 1. Per avere una probabilità di condivisione più o meno alta andremo ad agire solo sulla “forza di astensione”. Per come abbiamo definito la formula della condivisione basterà che la distribuzione dell’astensione sia più tendente a 0 per avere una miglior condivisione, o viceversa, più tendente a 1 per avere una peggior condivisione.

³ Non esiste un modello virtuale di una rete sociale reale. Tutte le ottimizzazioni che vengono apportate sono per rendere i risultati di queste simulazioni più attinenti alla realtà.

2.1 Strumenti

Dopo aver discusso di tutte le fasi della progettazione possiamo passare all'implementazione.

Le simulazioni previste saranno sviluppate in NetLogo che fornisce un ambiente semplice e piuttosto personalizzabile dove si possono condurre studi in parecchi campi. NetLogo è uno strumento di sviluppo per una programmazione basata ad agenti. Nel lavoro qui presentato gli agenti in questione saranno gli utenti della rete sociale.

Esistono molti altri simulatori di reti che prevedono una programmazione ad agenti, ma NetLogo ha risposto brillantemente alle esigenze implementative avendo una sintassi semplice e una moltitudine di API [6] fornite dal linguaggio stesso.

2.2 Topologie dei grafi

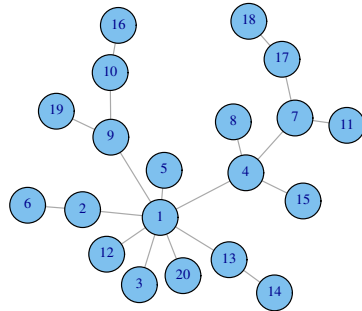
Come già annunciato all'inizio di questo capitolo i modelli di grafo che verranno studiati in questo lavoro sono due.

Preferential Attachment Il modello Preferential Attachment, come già comunicato a inizio capitolo, è altamente diffuso e conosciuto. Tanto da avere una propria API per la creazione nel dizionario di NetLogo.

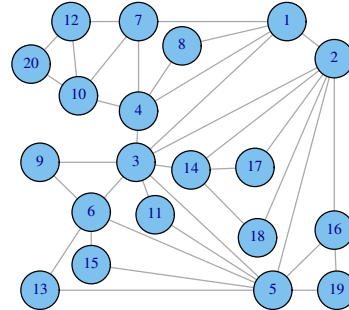
Invocando la funzione:

```
| generate-preferential-attachment turtles links num-nodes
```

che risiede all'interno dell'estensione "network" verrà creato un grafo di *num-nodes* nodi con le proprietà del grafo Preferential Attachment. L'algoritmo di PA implementato in NetLogo crea per ogni nuovo nodo un solo collegamento. In figura 2a è presente una rappresentazione grafica del grafo.



(a) Preferential Attachment



(b) Dorogovtsev e Mendes

Fig. 2: Rappresentazione grafica delle due topologie di grafo studiate.

Modello di Dorogovtsev e Mendes Il modello definito da Dorogovtsev e Mendes viene implementato dalla libreria Java chiamata GraphStream [7]. La libreria è formata essenzialmente da tre parti:

- core: Il package principale di GraphStream;
- algo: Il package dove sono implementati tutti gli algoritmi e i generatori della libreria;
- ui: Il package che permette di visualizzare e dare un layout al grafo.

I grafi che verranno utilizzati nelle simulazioni sono stati creati da un'applicazione implementata appositamente. Una volta creati, questi grafi vengono salvati su file in formato GraphML [8] per dar modo di essere caricati dalla simulazione di NetLogo.

In figura 2b è presente una rappresentazione grafica del grafo.

2.3 Interfaccia grafica per la configurazione della rete

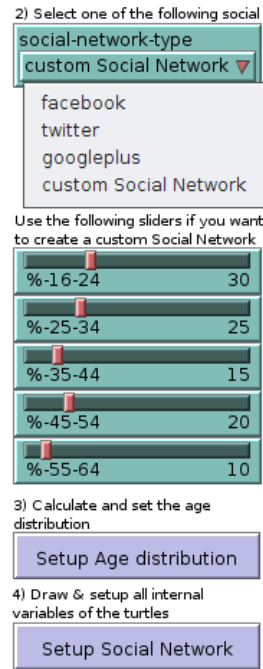


Fig. 4: GUI: scelta del Social Networks; pulsante per il calcolo della distribuzione delle età; pulsante per l'inizializzazione dei nodi.

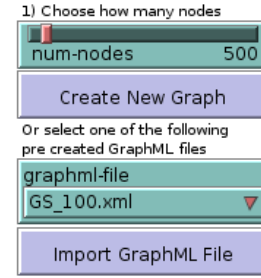


Fig. 3: GUI: Creazione o importazione del grafo

Al fine di portare a termine gli obiettivi in precedenza descritti è stata creata una interfaccia grafica per permettere all'utilizzatore di interagire con il progetto. Nella figura 3 e nella figura 4 vengono mostrate due porzioni di GUI che servono all'inizializzazione delle simulazioni. Nella maggior parte delle dimostrazioni di NetLogo è presente solo un pulsante "Setup" e un pulsante "go". Data la particolarità del progetto non è stato possibile semplificare a questo livello l'interfaccia grafica perciò, qui di seguito, verranno illustrate le caratteristiche principali.

In figura 3 si possono notare due pulsanti:

- il primo serve per creare un grafo di tipo Preferential Attachment (Sezione 2.2) con un numero *num-nodes* di nodi.
- il secondo invece serve per importare un file GraphML (Sezione 2.2) tra quelli già creati in precedenza.

Dopo aver creato o importato il grafo, viene cercato l'hub della rete, ovvero il nodo con il *degree* più alto.

Un'altra fase necessaria è stata quella di importare nel progetto i dati di alcuni Social Networks presenti nel grafico di figura 1.

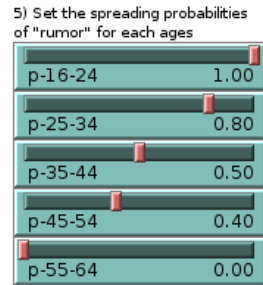


Fig. 5: GUI: cursori per la definizione di una notizia.

Sono stati inclusi solo i tre che vengono mostrati in figura 4 per via della loro popolarità. In alcune simulazioni esiste la necessità di creare condizioni particolari che non possono essere risolte con i dati di questi Social Networks. Per rendere l'applicazione più generale possibile sono stati implementati anche 5 cursori che permettono di far decidere all'utilizzatore dei parametri personalizzati. Ognuno di questi cursori serve per impostare la percentuale di utenti facente parte della suddetta fascia d'età. Per semplicità sono stati chiamati: “%-<fascia età inizio>-<fascia età fine>”.

Si può vedere la loro implementazione nell'esempio

in figura 4 dove si nota che:

- il gruppo formato dalle persone che hanno tra i 16 e i 24 anni sono il 30%;
- quelle che hanno tra i 25 e i 34 anni sono il 25%;
- quelle che hanno tra i 35 e i 44 anni sono il 15%;
- quelle che hanno tra i 45 e i 54 anni sono il 20% gruppo;
- quelle che hanno tra i 55 e i 64 anni sono il 10%;

Per calcolare quanti utenti devono essere presenti per ogni intervallo di età bisogna premere il pulsante “Setup Age Distribution”. Il pulsante “Setup Social Networks”, invece, serve a resettare i dati base degli utenti, riportando la simulazione allo stato iniziale, senza eliminare ne il grafo ne la distribuzione dell'età. Per stato iniziale della simulazione si intende posizionare nell'hub la notizia, facendolo diventare uno “Spreader”. In tutti gli altri utenti del grafo verrà, invece, impostato lo stato “Ignorant” e sarà calcolato un nuovo valore di astensione alla notizia.

Sono presenti altri 5 cursori, mostrati in figura 5, che identificano l'intensità con cui la notizia è forte verso un certo gruppo di utenti. Viene definita come “p-<fascia età inizio>-<fascia età fine>” per indicare la probabilità di condivisione che la notizia ha sulla particolare fascia d'età. Nell'esempio in figura 5 la probabilità nel primo gruppo vale 1.0, ovvero la condivisione avrà *sempre* esito positivo. In caso contrario nel quinto gruppo la probabilità è 0 perciò la condivisione avrà sempre esito negativo. Nei casi intermedi viene confrontato il valore di astensione dell'utente al valore del cursore dell'età dell'utente come è stato definito in fase di progettazione (Sezione 2).

I pulsanti presenti in figura 6 danno la possibilità di far partire la simulazione per il primo e secondo obiettivo. Più precisamente per la prima simulazione viene automaticamente impostato il tipo di Social Network a “custom” per poter personalizzare la percentuale di utenti nei vari gruppi, impostando il 100% nel primo gruppo e 0% negli altri gruppi. A questo punto il test eseguirà la procedura

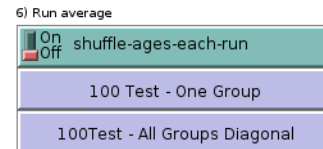


Fig. 6: GUI: pulsanti per dare il via ai primi due test

di propagazione della notizia facendo variare la probabilità della notizia da 0 a 1 con step di 0.05 per il primo gruppo. Si arriverà ad avere un grafico che mostrerà sull'asse X i ticks, ovvero l'avanzamento della simulazione. Mentre sull'asse Y ci sarà la media di 100 test del numero di "Viewers".

La seconda simulazione utilizza invece le % di utenti definite in precedenza e fa variare la probabilità della notizia da 0 a 1 con step di 0.05 per tutti i gruppi. Così facendo si arriverà ad avere un grafico dove sull'asse X ci saranno nuovamente i ticks, ovvero l'avanzamento della simulazione. Mentre sull'asse Y ci saranno tutte le medie di 100 test del numero di "Viewers" di ogni gruppo.

3 Analisi dei Risultati

In quest'ultima analisi, invece, verrà mostrata un'interazione tra 2 diversi gruppi di utenti (Nodi). Il primo gruppo è formato da persone con un'alta probabilità di condividere la notizia mentre nel secondo, al contrario, da persone con una bassa probabilità. Questo studio punta ad analizzare quante visualizzazioni vengono fatte per una singola informazione condivisa. Verrà anche condotto uno studio in cui il totale degli utenti si divide in due gruppi costituiti da, ad esempio, pochi utenti con alte probabilità di condividere l'informazione e, viceversa, molti utenti con basse probabilità di condivisione. La possibilità di condivisione è data nuovamente dal confronto tra la "forza della notizia" e la "forza di astensione"???. In questo caso, però, l'astensione viene calcolata tramite una funzione di distribuzione di probabilità non lineare, quella scelta è la funzione Weibull. La funzione "random-weibull"?? fa variare, per mezzo dei due parametri α e β , la pendenza della distribuzione e, di conseguenza, anche la sua densità.

Per ottenere una curva di distribuzione più o meno inclinata, in seguito ai tentativi attuati, è stato ritenuto opportuno mantenere il valore di β costante a 1.0 e variare, invece, quello di α .

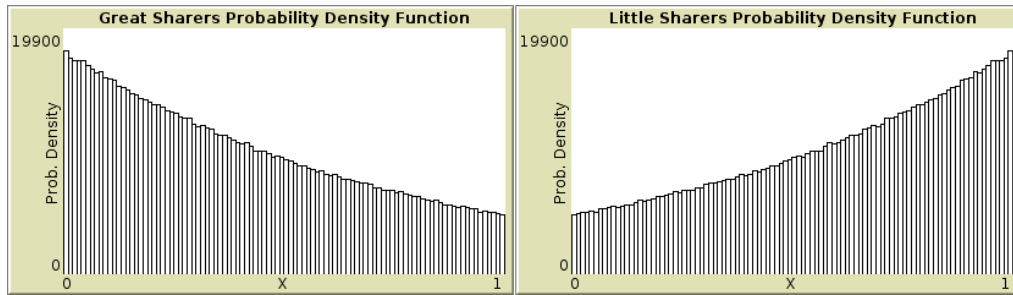


Fig. 7: I due istogrammi mostrano la rappresentazione di un vettore X di 1 milione di cifre calcolato mediante la funzione di densità di probabilità Weibull, con $\alpha = 0.75$. Il grafico di sinistra rappresenta la distribuzione originale, mentre quello di destra la sua speculare calcolata grazie: $\forall x \in X, \quad 1 - x$.

Nella figura 7 il dislivello non è troppo alto e questo permette comunque al gruppo 2 un buon grado di condivisione della notizia.

Questa analisi è stata sviluppata mediante tre cicli annidati che hanno considerato tutte le possibilità che verranno di seguito esposte.

Il primo ciclo più esterno itera su alcuni valori di forza della notizia, ovvero i seguenti: 0.50 0.60 0.70 0.75 0.80 0.90.

Il secondo ciclo, quello centrale, modifica il numero di individui per ogni gruppo. Ad ogni passo viene incrementato il numero di persone con un'alta possibilità di condivisione passando dal 10% al 90%. Operazione inversa viene eseguita per le persone con bassa possibilità di condivisione passando dal 90% al 10% della

popolazione totale. Il terzo ciclo, più interno, invece itera sul parametro α modificando quindi la pendenza della curva per gli individui con alta probabilità di condivisione.

Il test viene composto da i parametri dinamici appena citati e dai seguenti parametri statici:

- La dimensione della popolazione non cambia mai e resta sempre di 500 Nodi;
- Il parametro α del gruppo con bassa probabilità di condivisione rimane 0.75, ottenendo una densità di probabilità come in figura 7;
- La topologia del grafo è la stessa per ogni esecuzione del test;

Il risultato di ogni test inoltre viene mediato sull'esito di 500 test.

La figura 8 mostra il risultato atteso, ovvero la crescita del numero di visualizzazioni sia al variare di α che all'aumentare dei nodi con alta probabilità di condivisione.

Si vuole inoltre porre l'attenzione su come con un valore di α molto basso (0.2) ed un numero di nodi con alta probabilità di condivisione pari a 150 corrisponda a 300 visualizzazioni totali, quasi le stesse ottenute da un valore di $\alpha = 1.0$ e 325 nodi con alta probabilità di condivisione.

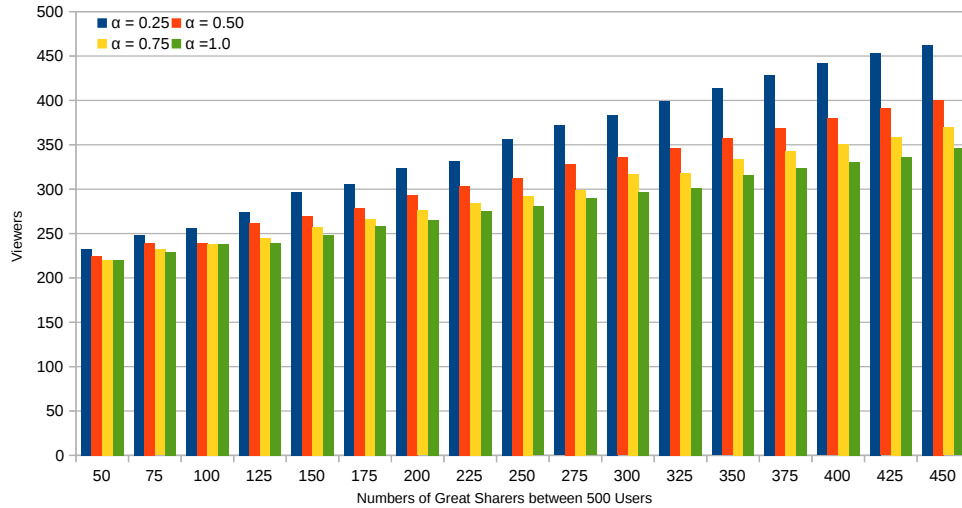


Fig. 8: Grafico del risultato dell'ultimo test con forza della notizia pari a 0.50, 500 nodi totali e α del gruppo con bassa probabilità di condivisione = 0.75

4 Conclusioni

In conclusione i risultati posti come obiettivo sono stati raggiunti.

Come abbiamo potuto osservare sono pienamente in linea con

4.1 Sviluppi Futuri

Prendere i risultati del 3° test e confrontarli con quelli dei test base e vedere come si comportano, minore? maggiore?

References

1. D. J. Daley, D. G. Kendall.
Stochastic Rumours.
<http://imamat.oxfordjournals.org/content/1/1/42.abstract>
2. A.-L. Barabási, R. Albert.
Emergence of scaling in random networks
http://www.barabasilab.com/pubs/CCNR-ALB_Publications/199910-15_Science-Emergence/199910-15_Science-Emergence.pdf
3. S. N. Dorogovtsev, J. F. F. Mendes.
Evolution of networks
<http://arxiv.org/pdf/cond-mat/0106144v2.pdf>
4. Statista 2015.
Age distribution of active social media users worldwide as of 3rd quarter 2014, by platform .
<http://www.statista.com/statistics/274829/age-distribution-of-active-social-media-users-worldwide-by-platform/>
5. Zhongzhi Zhang, Shuigeng Zhou, Tao Zou, Guisheng Chen.
Fractal scale-free networks resistant to disease spread.
<http://arxiv.org/pdf/0804.3186.pdf>
6. <http://ccl.northwestern.edu>.
NetLogo Dictionary.
<http://ccl.northwestern.edu/netlogo/docs/dictionary.html>
7. GraphStream Team
GraphStream Project.
<http://graphstream-project.org/>
8. GraphML Team
The GraphML File Format.
<http://graphml.graphdrawing.org/>
9. Nick Fedewa, Emily Krause, Alexandra Sisson and Advisor: James Angelos.
Spread of A Rumor.
<http://www.siam.org/students/siuro/vol6/S01182.pdf>