

# 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`

**Sommario** 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 NBC<sup>1</sup> 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'immagine<sup>2</sup>, 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 2) 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.

---

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

<sup>2</sup> Immagine condivisa dalla maestra, Melissa Bour: [http://i.dailymail.co.uk/i/pix/2015/01/04/2466352100000578-2895903-image-a-74\\_1420339978179.jpg](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 relativi al campo dello “Spreading Rumors”.

I primi studi riguardanti la diffusione di una notizia risalgono ai primi anni '60 [2], 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 studiate sono la topologia della rete e gli utenti che ne fanno parte.

Possiamo quindi dire che le simulazioni prevederanno una rete di utenti che, “a contatto” con la notizia, decideranno se ricondividerla o se solamente prenderne atto. Al passo 0 della simulazione verrà scelto un utente a cui verrà “insegnata” l’informazione da propagare. Facendo un riferimento al tema della medicina, durante un’indagine epidemiologica, il primo paziente ad aver contratto la malattia viene chiamato paziente zero; nella situazione in questione, anziché di una malattia, si tratterà di una notizia.

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

- “Ignorants”[1] : al momento della creazione della rete sociale tutti gli utenti vengono definiti in tal modo, in quanto non al corrente della notizia;
- “Spreaders”[1] : tutti gli utenti che decidono di condividere la notizia. Il “paziente zero” viene incluso in questo gruppo;
- “Uninterested” : tutti quegli utenti che dopo aver appreso la notizia, decidono 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; inoltre il numero esorbitante di utilizzatori di Social Networks in circolazione crea, sicuramente, un terreno fertile per la diffusione di notizie ed informazioni.

### 1.1 Obiettivi

Per questo lavoro si possono inquadrare 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 test 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. On-line è presente una statistica piuttosto recente (3° trimestre 2014) che mostra una distribuzione relativa alla divisione, per età, dell’utenza nei social networks più popolari. Il grafico in figura 1 ne mostra la distribuzione.

Il terzo, ed ultimo, test ha come obiettivo, invece, l’analisi di un’interazione tra 2 diversi gruppi di utenti e permetterà di studiare il numero di visualizzazioni

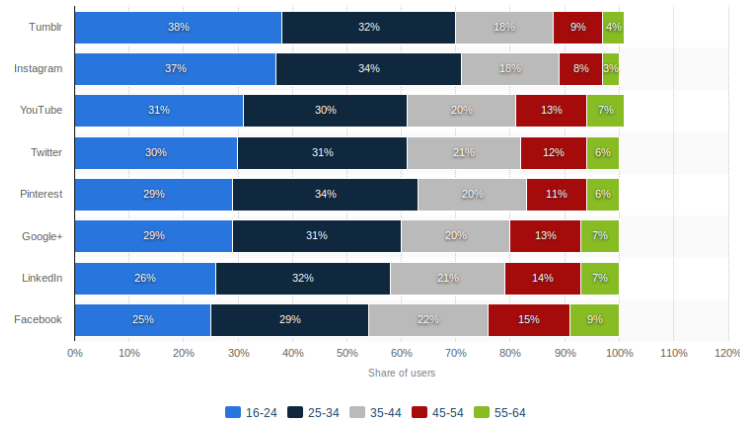


Figura 1: Distribuzione delle età divisa per Social Network [5]

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

- Il primo gruppo è formato da pochi utenti, ad esempio 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 lo studio di un modello che permetta di raggiungere gli obiettivi descritti pocanzi, sono state prese le decisioni di seguito elencate.

Si anticipa che il social network verrà astratto ad un grafo scale-free nel quale ogni nodo corrisponde ad un utente in possesso di alcune caratteristiche.

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

Inizialmente verrà mostrato il propagarsi di una notizia in un grafo di tipo “Preferential Attachment” suggerito da Barabási e Albert [3]. Le simulazioni proseguiranno poi attraverso un’altra topologia di grafo, sempre scale-free con Power Law Degree, descritta però dall’algoritmo di Dorogovtsev e Mendes [4].

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

- La topologia di grafo Preferential Attachment si contraddistingue per le seguenti caratteristiche:
  - è un modello ampiamente utilizzato per la sua semplicità;
  - è impiegato in buona parte di studi che trattano argomenti somiglianti lo “Spreading Rumors”;

- non è un grafo di tipo frattale [6] ma ha una caratteristica simile. Infatti, non possedendo cricche di almeno 3 nodi, nel momento in cui un nodo non condividesse la notizia, tutti i suoi nodi “figli” non riceverebbero mai l’informazione.
- La topologia di grafo definita da Dorogovtsev e Mendes invece viene descritta come segue:
  - è caratterizzata da un modello più complesso;
  - può avere all’interno del grafo cricche da 3 o più nodi, permettendo così una probabilità maggiore di condivisione della notizia;
  - è più somigliante<sup>3</sup> 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. Mirando al secondo obiettivo risulta necessario fornire alla notizia un “argomento” e si è così optato per l’inserimento di  $N$  valori che definiranno quanto la notizia sia adatta ad per ogni fascia d’età. Avendo a disposizione la distribuzione delle età, precedentemente mostrata nel grafico di figura 1, è possibile osservare 5 intervalli di età e quindi è possibile stabilire il valore di  $N = 5$ .

Risulta ora necessario definire 5 gruppi di utenti per distinguere le differenti età. Il numero di persone appartenenti ad ogni gruppo verrà definito dalla semplice proporzione:

$$\frac{N\_NODI\_TOTALI}{100} \cdot \%\_UTENTI\_GRUPPO$$

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

Giunti a questo punto si può constatare come la formula che definisca la probabilità di propagazione sia l’unica mancante alla condivisione dell’informazione tra nodo e nodo. L’idea è quella di dare la “possibilità” all’utente di decidere, come nella realtà, se la notizia gli interessi o meno. Risulta quindi necessario aggiungere all’utente un nuovo parametro, ovvero l’astensione alla notizia.

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

$$NOTIZIA_e > UTENTE_i.astensione$$

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 si rifà su quest’ultima parte appena definita. Quando si parla di “forza della notizia” e di “forza di astensione” si vuole considerare un numero tra 0 e 1. Per ottenere una probabilità di condivisione più o meno alta si agirà solo sulla “forza di astensione”. Per come è stata definita la formula della condivisione, sarà sufficiente che la distribuzione dell’astensione sia più tendente a 0 per avere una miglior condivisione o, viceversa, più tendente a 1 per avere una minor condivisione.

<sup>3</sup> Non esiste un modello virtuale di una rete sociale reale, bensì ogni ottimizzazione apportata permette di rendere i risultati delle simulazioni più vicini alla realtà.

## 2.1 Strumenti

Dopo aver discusso tutte le fasi della progettazione, si passa ora alla fase della così detta implementazione.

Le simulazioni previste verranno sviluppate in NetLogo che fornisce un ambiente semplice e piuttosto personalizzabile nel quale si possano condurre studi in campi differenti. NetLogo è uno strumento di sviluppo per una programmazione basata ad agenti; nel lavoro qui presentato gli agenti in questione saranno rappresentati dagli utenti della rete sociale. Esistono vari simulatori di reti che prevedono una programmazione ad agenti, ma NetLogo, grazie alla sua semplice sintassi ed una moltitudine di API<sup>4</sup> [7] fornite dal linguaggio stesso, ha risposto brillantemente alle esigenze implementative. NetLogo è stato scelto anche per la facilità con la quale permette di modellare un'interfaccia grafica<sup>5</sup>.

## 2.2 Topologie dei grafi

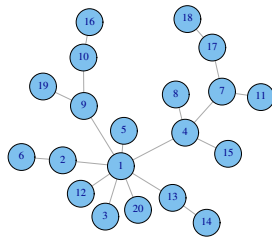
Come già annunciato all'inizio di questo capitolo, i modelli di grafo che verranno studiati in questo lavoro sono due.

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

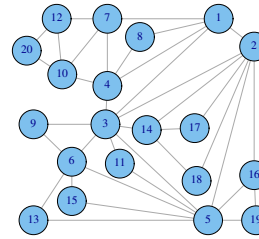
```
| generate-preferential-attachment turtles links num-nodes
```

risiede all'interno dell'estensione “network”, e invocandola, 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 si ha una rappresentazione grafica di un grafo con 20 nodi.



(a) Preferential Attachment



(b) Dorogovtsev e Mendes

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

<sup>4</sup> Application Programming Interface: un insieme di funzionalità utilizzabili dal programmatore

<sup>5</sup> Dall'inglese GUI ovvero Graphical User Interface

**Rete di Dorogovtsev e Mendes** Il modello definito da Dorogovtsev e Mendes viene implementato dalla libreria Java chiamata GraphStream [8]. La libreria è formata da tre parti, le seguenti:

- core: Il package principale di GraphStream;
- algo: Il package dove vengono implementati tutti gli algoritmi e i generatori della libreria;
- ui: Il package che permette la visualizzazione e che da 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 [9] in modo da essere caricati dall'applicativo di NetLogo.

In figura 2b è presente una rappresentazione grafica di un grafo con 20 nodi.

### 2.3 Inizializzazione del Modello

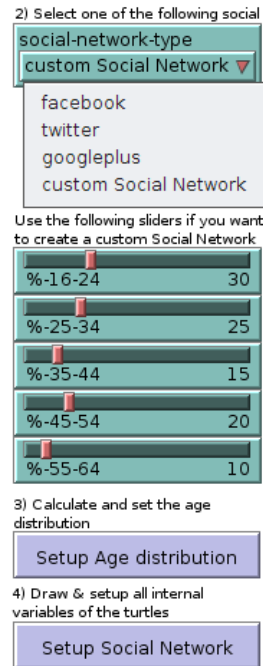


Figura 4: GUI: scelta del Social Networks; pulsante per il calcolo delle età; pulsante per l'inizializzazione dei nodi.

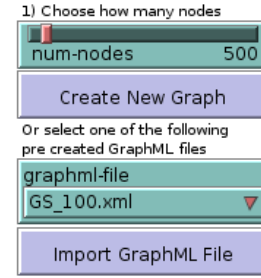


Figura 3: GUI: Creazione o importazione del grafo

Al fine di portare a termine gli obiettivi in precedenza descritti è stata creata un'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 allo stesso modo l'interfaccia grafica; di seguito verranno illustrate le principali caratteristiche grafiche del progetto qui presentato.

In figura 3 si possono notare due pulsanti:

- il primo pulsante permette la creazione di un grafo di tipo Preferential Attachment (Capitolo 2.2) con un numero *num-nodes* di nodi.
- il secondo pulsante, invece, serve per l'importazione di un file GraphML (Capitolo 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 dell'importazione nel progetto dei dati di alcuni Social Networks presenti nel grafico di figura 1.

Nel progetto qui presentato si è scelto di includere i dati dei tre Social Network più popolari, essi

vengono mostrati in figura 4. In alcune simulazioni si manifesta la necessità di creare condizioni particolari che non possono essere risolte con i dati di questi Social Network. Con l'intenzione di rendere l'applicazione più generica possibile, sono stati implementati anche 5 cursori che permettono di far selezionare all'utilizzatore parametri personalizzati. Ognuno dei cinque cursori serve ad impostare la percentuale di utenti facente parte della suddetta fascia d'età e, per semplicità sono stati chiamati “%-<fascia età inizio>-<fascia età fine>”.

Nell'esempio di figura 4 si può osservarne l'implementazione dove si nota quanto segue.

- il gruppo formato di utenti che hanno tra i 16 e i 24 anni rappresenta il 30%;
- il gruppo costituito dalle persone che hanno tra i 25 e i 34 anni sono il 25%;
- il gruppo di persone che hanno tra i 35 e i 44 anni costituisce il 15%;
- le persone che hanno tra i 45 e i 54 anni sono il 20%;
- infine, il gruppo di persone che hanno tra i 55 e i 64 anni sono il 10%;

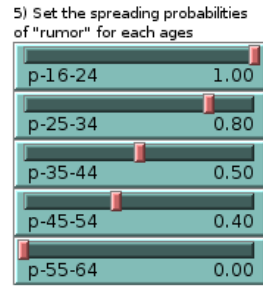


Figura 5: GUI: cursori per personalizzare la notizia.

iniziale della simulazione” si intende l’inserimento della notizia nell’hub, in modo da farlo diventare il primo “Spreader”; intanto gli altri nodi del grafo saranno etichettati come “Ignorant” e, per ciascuno, sarà calcolato un nuovo valore di astensione alla notizia.

Esistono altri 5 cursori, mostrati in figura 5, i quali identificano l'intensità con la quale la notizia è importante verso un certo gruppo di utenti. La definizione

“p-<fascia età inizio>-<fascia età fine>” si utilizza per indicare la probabilità di condivisione della notizia in una particolare fascia d'età. Nell'esempio in figura 5 la probabilità nel primo gruppo vale 1.0, ovvero la condivisione avrà *sempre* esito positivo.

Nel quinto gruppo si ha, al contrario, probabilità nulla, pari a 0 e la condivisione non andrà mai a buon fine. Nei casi intermedi, invece, viene confrontato il valore di astensione dell'utente col valore del cursore dell'età dell'utente come è stato definito in fase di progettazione (Capitolo 2). I pulsanti presenti in figura 6 servono a dare il via al test del primo e secondo obiettivo.

Per calcolare la percentuale necessaria di utenti per ogni intervallo di età è necessario premere il pulsante “Setup Age Distribution”. Verrà creato così un vettore grande come il numero dei nodi del grafo, e al suo interno, saranno inseriti gli indici dei 5 gruppi nella quantità prevista dalle percentuali. Questo procedimento farà sì che ogni nodo del grafo avrà un suo preciso valore nell'array, di modo da far sì che si identifichi il gruppo di appartenenza. Il pulsante “Setup Social Networks”, invece, serve a eseguire il reset dei dati base degli utenti, riportando la simulazione allo stato iniziale, senza eliminare né il grafo né il vettore delle “età”. Col termine “stato

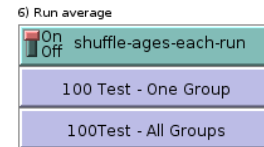


Figura 6: GUI: pulsanti per dare il via ai primi due test

## 2.4 Sviluppo delle Simulazioni

Oltre all'inizializzazione della rete e alla configurazione dei suoi agenti, è stata progettata un'architettura in grado di soddisfare gli obiettivi prefissati.

**Propagazione della Notizia** L'algoritmo 1, data la sua importanza, è stato definito il nucleo delle simulazioni ed è colui che calcola, ad ogni passo, come la notizia si propaghi.

---

**Algorithm 1** Nucleo della propagazione della notizia

---

```

1: procedure CORE(nodes, news)
2:   if (all nodes have seen the news) then
3:     return END_OK
4:   else
5:     for all (node in nodes with [seen_news = false]) do
6:       if (any nodeneighbors with [shared_news = true]) then
7:         ▷ Leggo il valore di probabilità della notizia per l'età del utente corrente
8:          $p\_news \leftarrow news_{prob[node\_age]}$ 
9:         if  $p\_news \geq node.chance\_of\_abstention$  then
10:          nodetmp_shared_news  $\leftarrow$  true
11:        end if
12:        nodetmp_seen_news  $\leftarrow$  true
13:      end if
14:    end for
15:    if (no nodes with [tmp_seen_news = True]) then
16:      return END_KO
17:    end if
18:    for all (node in nodes with [tmp_seen_news = true]) do
19:      nodeseen_news  $\leftarrow$  true
20:      nodetmp_seen_news  $\leftarrow$  false
21:    end for
22:    for all (node in nodes with [tmp_shared_news = true]) do
23:      nodeshared_news  $\leftarrow$  true
24:      nodetmp_shared_news  $\leftarrow$  false
25:    end for
26:    return SIMULATION_NOT_FINISHED_YET
27:  end if
28: end procedure

```

---

La funzione Core prima verifica se ci sia qualche nodo che non ha visualizzato la notizia (Riga 2), poi richiede a tutti i nodi “Ignorants” se hanno un vicino “Spreader” (Riga 5 e 6). Nel caso in cui questa ipotesi sia soddisfatta, viene letta la probabilità della notizia rispetto all'età del nodo “Ignorant” corrente (Riga 8) e viene eseguito il confronto con il valore di astensione alla notizia (Riga 9), come stabilito in fase di progettazione descritta nel capitolo 2. Inoltre esiste un'altra funzione, altrettanto importante, che si occupa di calcolare la media di  $N$  test del numero di “Viewers”. Ad ogni esecuzione, però, la funzione si preoccupa anche



di effettuare il reset di tutti gli agenti presenti nella simulazione. Essa calcola nuovamente un valore di astensione per ogni nodo e, se richiesto, rimescola gli elementi del vettore contenente le differenti età al fine di avere ampia casualità nei test effettuati.

**“100 Test - One Group”** Questo pulsante viene premuto per eseguire la prima simulazione. Come prima azione verrà impostato il Social Network di tipo “custom”, esso permetterà di poter personalizzare la percentuale di utenti nei 5 gruppi, impostando il 100% nel primo gruppo e 0% negli altri gruppi.

Successivamente il test eseguirà la procedura di propagazione della notizia facendo variare la probabilità della notizia per il primo gruppo, quindi il cursore “p-16-24”, da 0 a 1 con passi di 0.05. Il risultato sarà un grafico che mostrerà sull’asse X i ticks, ovvero l’avanzamento della simulazione, quindi la forza della notizia; mentre sull’asse Y ci sarà la media di 100 test del numero di “Viewers”.

**“100 Test - All Groups”** La seconda simulazione utilizza invece le percentuali di utenti definite nel grafico della statistica di figura 1, facendo variare una delle probabilità della notizia da 0 a 1 con step di 0.05. Inizialmente si studierà cosa capita al solo variare della forza della notizia del gruppo più numeroso; successivamente verrà invece analizzato il comportamento della rete al variare della forza della notizia di tutti i gruppi. Così facendo si otterrà un grafico dove sull’asse X ci saranno nuovamente i ticks, ovvero l’avanzamento della simulazione, quindi la forza della notizia, mentre sull’asse Y ci saranno tutte le medie di 500 test del numero di “Viewers” di ogni gruppo.

**“Great Sharers vs. Little Sharers”** L’ultima simulazione è quella che si distingue di più dalle altre considerato il concetto su cui si basa; si vuole infatti studiare il confronto tra due gruppi che, a differenza dei precedenti test, si distinguono per la loro probabilità di astensione alla divulgazione della notizia. Nelle prime due simulazioni questo dato veniva stabilito da una funzione random caratterizzata da una distribuzione di densità lineare. In questo test, invece, l’astensione alla condivisione viene calcolata tramite una funzione di distribuzione di probabilità non lineare; quella scelta è la funzione Weibull[10]. La funzione “random-weibull” fa variare la densità della distribuzione per mezzo di due parametri  $\alpha$  e  $\beta$ . Il valore restituito dalla funzione è sempre un numero tra 0 ed 1 perciò, per il calcolo della propagazione della notizia, può essere riutilizzato l’algoritmo 1 mostrato in precedenza. In figura 7 sono presenti 3 slider per la manipolazione degli  $\alpha$  e di  $\beta$ . Il parametro  $\alpha$  modifica la pendenza della curva mentre il parametro  $\beta$  modifica soprattutto la forma della curva. Per la simulazione ci serviremo di due cursori

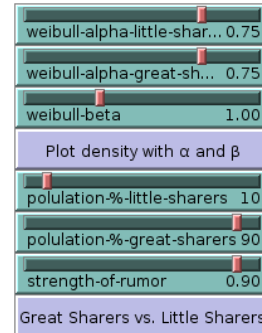


Figura 7: GUI: terzo test

per la manipolazione del parametro  $\alpha$  dei due gruppi. In figura 8 viene mostrato il dislivello a cui si è appena fatto riferimento. Nel primo istogramma, con  $X$  più densa vicino a 0, viene permessa una miglior condivisione mentre nel secondo grafico ne risulta una peggiore.

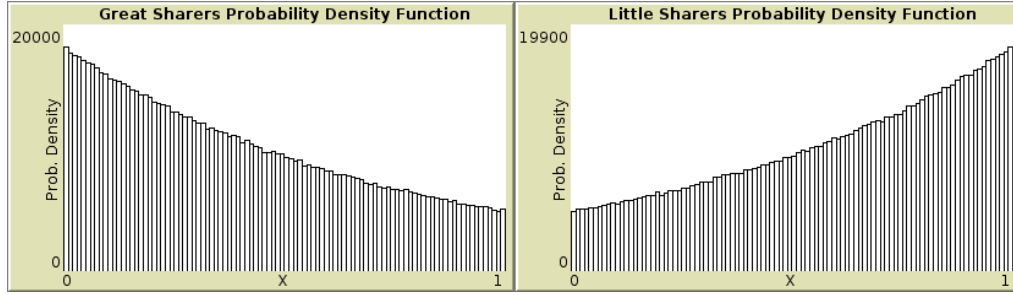


Figura 8: I due istogrammi mostrano la rappresentazione di un vettore  $X$  di 1 milione di cifre calcolato mediante la funzione di densità di probabilità Weibull, con  $\alpha = 0.75$ . Il grafico di sinistra rappresenta la distribuzione originale, mentre quello di destra la sua *quasi*<sup>6</sup> speculare calcolata grazie:  $\forall x \in X, \quad 1 - x$ .

In figura 7 sono presenti altri 2 cursori per definire la percentuale degli utenti appartenenti ad ogni gruppo e il cursore per impostare la “forza” della notizia. Alla pressione del tasto “Great Sharers vs. Little Sharers” viene lanciata una simulazione che riguarda i parametri selezionati. Vista la quantità di test da fare con tutti le combinazioni di parametri possibili è stata scritta una procedura<sup>7</sup> che, grazie ai suoi tre cicli annidati, riesce a considerare tutte le possibilità richieste dal terzo obiettivo. L’algoritmo 2 mostra su quali parametri agiscono i tre cicli.

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 a quella appena descritta 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  $\alpha$  modificando quindi la pendenza della curva per gli individui con alta probabilità di condivisione. Come si può notare il parametro  $\alpha$  del gruppo che condivide meno viene preso in ingresso dalla funzione e mai manipolato.

<sup>6</sup> Quando si parla di numeri random non è possibile pretendere di ottenere lo stesso identico comportamento.

<sup>7</sup> La procedura che permette di elaborare tutti i test impiega un tempo prolungato per la sua esecuzione. Per invocarla è sufficiente inserire il comando “great-sharers-in-comparison-to-little-sharers-all-tests” in modalità *Observer* nel *Command Center* di NetLogo.

---

**Algorithm 2** Tre cicli annidati per considerare tutte le possibilità interessanti del terzo obiettivo

---

```

1: procedure THIRD-TEST-ALL(alpha_little_sharers)
2:   weibull_alpha_little_sharers  $\leftarrow$  alpha_little_sharers
3:   for all (news_strength in [0.50 0.60 0.70 0.75 0.80 0.90]) do
4:     population_%_step  $\leftarrow$  10
5:     while (population_%_step  $\leq$  90) do
6:       %_little_sharers  $\leftarrow$  (100 - population_%_step)
7:       %_great_sharers  $\leftarrow$  population_%_step
8:       for all (weibull_alpha_great_sharers in [0.25 0.50 0.75 1.0]) do
          ▷ Eseguo la simulazione 500 volte con i paramentri correnti per avere un
          risultato mediato più valido possibile
9:       end for
10:      population_%_step  $\leftarrow$  population_%_step + 5
11:    end while
12:  end for
13: end procedure

```

---

### 3 Analisi dei Risultati

Prima di iniziare ad elencare i test effettuati e mostrare i risultati ottenuti, è necessario sottolineare una questione relativamente importante: in seguito all'osservazione del comportamento dell'applicativo con grafi molto grandi ( $> 500$  nodi) è stato ritenuto opportuno abbassare il numero dei nodi affinché l'esecuzione non impiegasse troppo tempo e risorse. In conclusione tutte le analisi sono state effettuate con reti composte da 500 nodi. Un secondo punto da chiarire è che avendo due applicativi, il primo in Java che punta a creare grafi (Capitolo 2.2) ed il secondo in NetLogo che mira ad utilizzarli nelle simulazioni, è stato ritenuto opportuno non cambiare mai il grafo durante i test.

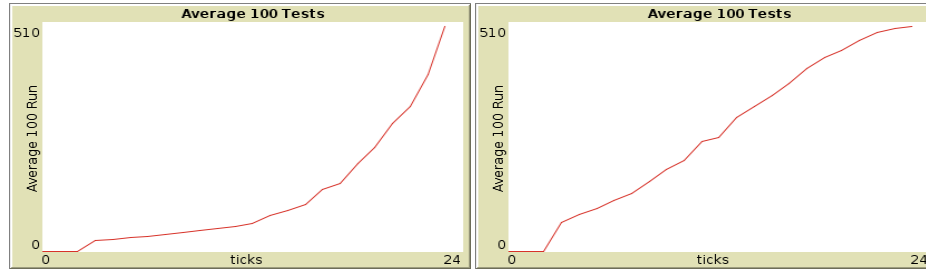
#### 3.1 Primo test

In questa prima analisi verranno confrontati i due tipi di grafo descritti nel capitolo 2.2.

Per questo test abbiamo bisogno di alcuni elementi imprescindibili, tali:

- un gruppo di utenti che forma il 100% dei nodi;
- un grafo di 500 nodi creato grazie all'algoritmo Preferential Attachment (Capitolo 2.2);
- un grafo di 500 nodi creato grazie all'algoritmo di Dorogovtsev e Mendes (Capitolo 2.2);
- un ciclo che vari la “forza” della notizia portandola da 0 a 1 con passi di 0.05.

Grazie al grafo descritto da Barabási e Albert si ottiene il risultato mostrato nel grafico di figura 9a, come è possibile notare, in questo caso, si avrà una crescita molto lenta e solo dopo che la probabilità di condivisione avrà superato il valore di 0.5 si otterrà un aumento tangibile. La topologia descritta da



(a) Con Preferential Attachment

(b) Con Dorogovtsev e Mendes

Figura 9: Risultati del primo test

Dorogovtsev e Mendes, invece, mostra un comportamento differente come da grafico in figura 9b, si evidenzia una crescita quasi lineare e alla fine, quando la forza è attorno l'85% crea una “pancia” verso l'alto, a dimostrare l'altissima condivisione.

Il primo grafo risulta essere svantaggiato per via della sua struttura, come affermato in fase di progettazione (Capitolo 2) questo grafo non presenta cricche al suo interno. Le proprietà appena descritte rappresentano una penalità sulla probabilità di condivisione per le simulazioni basate sul numero di visualizzazioni. Il modello descritto da Dorogovtsev e Mendes invece presenta cricche ed il grafo di questo esempio ne è la prova; esso ha infatti quasi il doppio dei link del Preferential Attachment.

Considerato quanto appena esposto, il confronto tra i due modelli di grafo ci permette di desumere come la seconda topologia affronti in maniera più migliore questo studio. Si è perciò optato di sostenere i prossimi test con la seconda topologia realizzata da Dorogovtsev e Mendes.

Inoltre, nonostante si tratti solo del punto di vista dell'autore, nel seguente modello di grafo si può notare una maggior somiglianza con la topologia più comune di un Social Network.

### 3.2 Secondo test

In seguito al confronto studiato nel primo test, ed aver deciso quale fosse il modello di grafo migliore per questo tipo di analisi, è possibile concentrarsi sulla seconda fase definita degli obiettivi. Questo secondo studio permetterà di analizzare la modalità di condivisione di una notizia, con un argomento, in differenti Social Network<sup>8</sup>. La notizia ha 5 valori di forza di condivisione che si riferiscono ai 5 differenti gruppi di età. La prima prova ha implicato l'inizializzazione dei valori di forza, di tutti i gruppi tranne che per quello con più utenti, a 0.75<sup>9</sup>. Quest'ultimo infatti verrà fatto variare da 0 a 1 con passi di 0.05. Ci permetterà di capire come cambiano i “Viewers” in queste condizioni.

<sup>8</sup> Per Social Network viene inteso lo stesso grafo di 500 nodi utilizzato in precedenza ma con le età degli utenti prese dallo studio statistico di figura 1.

<sup>9</sup> 0.75 è un valore piuttosto alto che permette una buona condivisione.

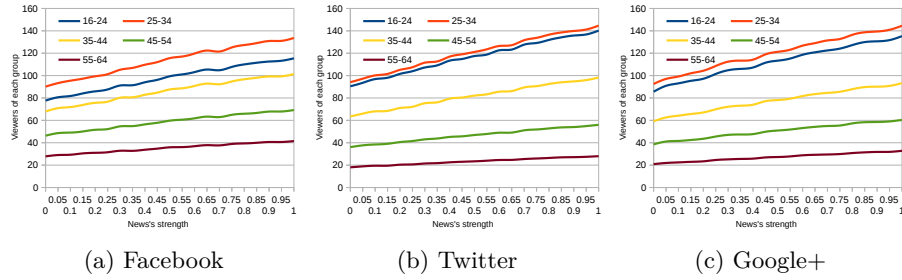


Figura 10: Risultati del secondo test

Nei grafici di figura 10 si nota un'alta linearità con le percentuali di età definite nel social Network, mentre la crescita risulta dovuta dall'aumento progressivo della forza di condivisione della notizia. Le curve nei grafici non partono dall'origine poichè il valore 0.75 fornisce una buona condivisione, sebbene non ottima; sommando questi ultimi infatti, l'ottimo, ovvero le 500 condivisioni, non viene mai raggiunto.

Per testare la linearità del grafo è stato prodotto un secondo test dove, a differenza di quello appena compiuto, verranno variati tutti e 5 i valori di forza da 0 a 1.

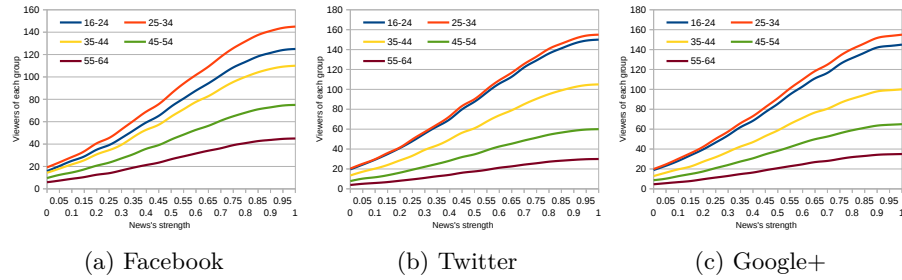


Figura 11: Risultati del secondo test

I grafici di figura 11 confermano la crescita lineare dovuta dall'aumento della forza per tutti i gruppi. Anche in questo caso è possibile osservare la stessa pancia verso l'alto, vista nel grafico di figura 9b. Si vuole inoltre far notare come, al passo 0, le visualizzazioni siano differenti da 1 ovvero quella del paziente zero. Il motivo è semplice, infatti, tutti quei nodi direttamente connessi al primo condivisore diventano "Viewers".

### 3.3 Terzo test

In quest'ultima analisi, invece, verrà mostrata un'interazione tra 2 diversi gruppi di utenti (Nodi): il primo gruppo sarà formato da persone con una maggior probabilità di condividere la notizia mentre nel secondo, al contrario, da persone con una minor probabilità. Questo studio punta ad analizzare quante visualizzazioni

vengono fatte per una singola informazione condivisa. In seguito verrà poi condotto uno studio in cui il totale degli utenti si dividerà 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 sarà data nuovamente dal confronto tra la “forza della notizia” e la “forza di astensione”. In questo caso, però, l'astensione verrà calcolata tramite una funzione di distribuzione di probabilità non lineare, quella scelta è la funzione Weibull già descritta al capitolo 2.4.

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

Il valore di  $\alpha$  per il secondo gruppo invece è stato fissato a 0.75 perchè la distribuzione di figura 8 mostra un dislivello non troppo alto che consente, comunque, un buon grado di condivisione della notizia.

Riassumendo il test viene composto da i parametri dinamici citati nel capitolo 2.4 e dai seguenti parametri statici:

- La dimensione della popolazione non cambia mai e resta sempre di 500 Nodi;
- Il parametro  $\alpha$  del gruppo con bassa probabilità di condivisione rimane 0.75, ottenendo una densità di probabilità come in figura 8;
- La topologia del grafo è la stessa per ogni esecuzione del test;

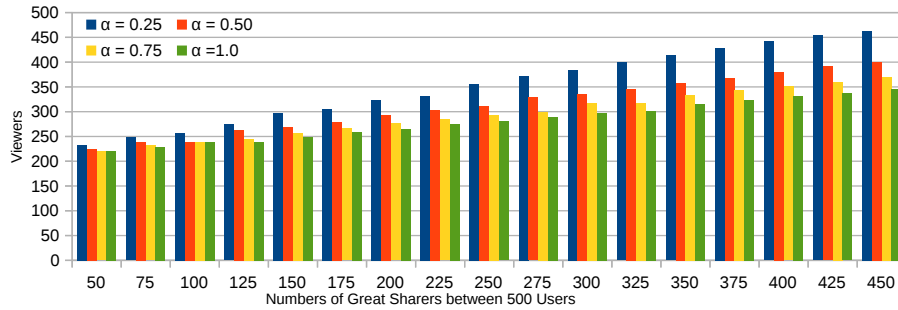


Figura 12: Grafico del risultato dell'ultimo test con forza della notizia pari a 0.50, 500 nodi totali e  $\alpha$  del gruppo con bassa probabilità di condivisione = 0.75

La figura 12 mostra il risultato atteso, ovvero la crescita del numero di visualizzazioni sia al variare di  $\alpha$  che all'aumentare dei nodi con alta probabilità di condivisione. Si vuole inoltre porre l'attenzione su come con un valore di  $\alpha$  molto basso (0.2) ed un numero di nodi con alta probabilità di condivisione pari a 150 corrisponda a  $\approx 300$  visualizzazioni totali, quasi le stesse ottenute da un valore di  $\alpha = 1.0$  e 325 nodi con alta probabilità di condivisione.

## 4 Conclusioni

In conclusione i risultati posti come obiettivo sono stati raggiunti.

Innanzitutto è stato studiato quale modello di grafo fosse più in linea con il tipo di analisi che questo progetto tratta. La topologia di grafo proposta da Dorogovtsev e Mendes ha avuto risultati di condivisione migliori. Il parametro chiave è definito dalla sua struttura particolare che prevede un numero di connessioni quasi doppio rispetto al Preferential Attachment.

Il secondo studio invece ha fornito una risposta concreta su come un'informazione con un certo argomento venga condivisa in un Social Network con differenti utenti. L'argomento viene "deciso" tramite 5 valori che vanno ad incidere sulla probabilità di condivisione della notizia. Si può quindi affermare che, più la notizia ha importanza, più questa viene condivisa. I test effettuati non sono stati esaustivi, ma sono risultati sufficienti per capire la tendenza appena affermata. Effettuare tutti i test sarebbe stato poco produttivo e avrebbe impiegato un'enormità di risorse; si vuol far notare infatti che per affrontare tutte le possibilità, dati i 5 parametri, ci sarebbero volute

$$\left(\frac{1-0}{0.05}\right)^5 * 500_{test} = 1.600.000.000$$

simulazioni complete volendo garantire la stessa precisione.

Per quanto concerne il terzo studio, sono stati confrontati due gruppi di utenti con differenti proprietà di astensione alla notizia. Queste proprietà sono state descritte dalla funzione di probabilità Weibull che, grazie ad un parametro  $\alpha$ , permette di modificare la densità della distribuzione. Così facendo si avrà un gruppo con una probabilità di condivisione maggiore dell'altro. Nel test effettuato è stato mostrato un nuovo risultato: la forza della notizia in questo caso non influisce con la crescita dei "Viewers". Il parametro importante ottenuto risulta essere l'aumentare degli individui con la bassa astensione alla notizia rispetto a quelli con una più alta. Si pone anche l'attenzione su come, in alcuni casi, pochi utenti che condividono molto possono avere un'influenza maggiore rispetto a tanti utenti che condividono meno.

#### 4.1 Sviluppi Futuri

Come già affermato nell'introduzione, considerata la rilevanza dell'argomento nella società moderna, questo tema, nonostante sia già stato affrontato in passato, continuerà ad esserlo anche in futuro.

Vi sono ancora molti test da effettuare ed alcuni temi su cui si potrebbe lavorare sono di seguito elencati.

- Un primo studio potrebbe puntare ad un'estensione dove si inserisca la durata della notizia, intesa come visibilità. Nella maggior parte dei Social Network esiste una "home" dove vengono visualizzate le notizie condivise dagli altri utenti a cui si è connessi. Queste notizie, però, risultano avere una durata massima; essa viene definita in base alla data di condivisione della stessa e dal grado di interazione degli altri utenti, ovvero più una notizia viene condivisa più questa diventa importante e rimane in vista.

- Un secondo studio potrebbe essere quello di rielaborare i test da un punto di vista differente. In questo lavoro si assume che la notizia venga condivisa inizialmente dal nodo con il maggior numero di vicini, quindi, paragonandolo alla realtà si potrebbe definire l’hub come un personaggio noto ad un pubblico vasto. Nell’espansione che si sta definendo il paziente zero potrebbe essere estratto casualmente tra tutti i nodi, o mediante una decisione più complessa.
- Un terzo studio potrebbe riguardare la rete sociale: in questo lavoro, a causa della divisione dei due applicativi, le simulazioni fanno riferimento ad uno stesso grafo di 500 nodi, grafo creato grazie all’algoritmo proposto da Dorogovtsev e Mendes. Considerato quanto appena esplicitato, potrebbe risultare opportuno analizzare nuovamente gli obiettivi con più reti, studiandone perciò le differenze.
- Un ultimo studio potrebbe puntare a confrontare i risultati del 2° e 3° test effettuati in questo lavoro; si potrebbe quindi analizzare le differenti crescite, di modo da osservarne il comportamento in più situazioni.

## Riferimenti bibliografici

1. Nick Fedewa, Emily Krause, Alexandra Sisson and Advisor: James Angelos. *Spread of A Rumor*.  
<http://www.siam.org/students/siuro/vol6/S01182.pdf>
2. D. J. Daley, D. G. Kendall. *Stochastic Rumours*.  
<http://imamat.oxfordjournals.org/content/1/1/42.abstract>
3. 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](http://www.barabasilab.com/pubs/CCNR-ALB_Publications/199910-15_Science-Emergence/199910-15_Science-Emergence.pdf)
4. S. N. Dorogovtsev, J. F. F. Mendes. *Evolution of networks*.  
<http://arxiv.org/pdf/cond-mat/0106144v2.pdf>
5. 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/>
6. Zhongzhi Zhang, Shuigeng Zhou, Tao Zou, Guisheng Chen. *Fractal scale-free networks resistant to disease spread*.  
<http://arxiv.org/pdf/0804.3186.pdf>
7. <http://ccl.northwestern.edu>. *NetLogo Dictionary*.  
<http://ccl.northwestern.edu/netlogo/docs/dictionary.html>
8. GraphStream Team. *GraphStream Project*.  
<http://graphstream-project.org/>
9. GraphML Team. *The GraphML File Format*.  
<http://graphml.graphdrawing.org/>
10. Fritz Scholz. *Inference for the Weibull Distribution*.  
<http://www.stat.washington.edu/fritz/DATAFILES498B2008/WeibullBounds.pdf>