

ABSoNeS: Agent Based Social Network Simulation

Simone Ciccolella¹ and Daniele Bellani²

¹*email*

²*email*

12 maggio 2017

TO DO

- Questione **omofilia**
- Definire costati di proporzione
- Definire alta e bassa probabilità e attività

1 Modellazione

Per la creazione del modello deriviamo delle costanti che saranno sempre vere:

Fasi Temporalì (FT) Dividiamo la giornata in 12 fasi che rappresentano 2 ore l'una in cui ogni utente ha un personale valore di **activity** che rappresenta il suo utilizzo di Twitter. Tali fasi sono così suddivise:

- 4 fasi con **activity** pari a 0 corrispondente a 8 ore di sonno, ovvero la quantità consigliata (e quasi mai rispettata).
- 4 fasi con **activity** bassa, corrispondente a 8 ore di lavoro
- 4 fasi con **activity** elevata che rappresenta 8 ore di tempo libero in cui l'utente ha una attività maggiore su Twitter

Al momento ignoramo il weekend, ma cercheremo di definirlo e implementarlo nelle prossime versioni. [•**SC**: definire **BASSA** e **ELEVATA** attività •]

Topic ovvero i possibili argomenti di interesse esistenti nel modello perciò abbiamo una lista **topic** = (t_1, \dots, t_n) definiti a priori e immutabili. Non vengono mai utilizzati direttamente, ma vengono utilizzati come indici di riferimento dunque non è necessario implementarli realmente.

1.1 Utenti

Gli utenti sono rappresentati dai nodi della rete. Vediamo ora dunque in che modo viene generato ogni nodo:

Personal Interest (PI) è una lista $PI = (p_1, \dots, p_n)$ dove $p_i \in [0, 1]$ [●SC: valutare inclusione degli estremi ●] indica la probabilità di interesse del nodo rispetto al topic i generata casualmente. Al momento tale valore è immutabile. [●SC: modellare valore variabile ●]

Timezone (TZ) ovvero una lista $TZ = (tz_1, \dots, tz_{12})$ dove i tz_i sono generati in accordo con le **FT** generati come segue:

- Genere $1 \leq i \leq 12$ casualmente
- Imposto le ore di sonno $tz_j = 0$ per $j = i, \dots, i + 3$
- Imposto le ore di lavoro $tz_j = \text{bassa activity}$ per $j = i + 4, \dots, i + 7$ [●SC: definire ●]
- Imposto le ore di tempo libero $tz_j = \text{alta activity}$ per $j = i + 8, \dots, i + 11$ [●SC: definire ●]

Tutte le precedenti operazioni di indici sono da considerarsi *mod* 12.

Per poter fare operazioni sugli utenti abbiamo inoltre bisogno delle seguenti definizioni:

Followers per ogni utente U definiamo come **followers**(U) l'insieme dei followers di U .

Following similmente **following**(U) indica l'insieme degli utenti seguiti da U .

Direct TAG (dtag) **dtag**(U) è l'insieme dei tweets che contengo un tag all'utente U . Sarà spiegato meglio nella sezione riguardante i tweet [1.3].

Interessi per ogni utente U **interest**(U) è l'insieme dei topic che interessano a U .

Field Of View (FOV) **FOV**(U) rappresenta l'insieme delle notizie visualizzate dall'utente U , con $FOV \subseteq \text{following}(U) \cup \text{dtag}(U)$ Questo insieme è definito sia per una questione computazionale, ma anche per un motivo reale in quanto è difficile che un utente nell'arco della giornata riesca a vedere tutti i tweets e i retweets degli utenti che segue e i tweet che in cui risulta direttamente *taggato*. Questa affermazione diventa sempre più ragionevole al crescere della popolarità di un utente.

Le relazioni tra gli utenti sono descritti dagli archi tra di essi.

Edges un arco tra due utenti U_1 e U_2 , scritto come (U_1, U_2) rappresenta la relazione di following tra il primo ed il secondo. Di conseguenza, ovviamente, risulta che $U_1 \in \text{followers}(U_2)$ e simmetricamente $U_2 \in \text{following}(U_1)$

Attachment ad un arco $e_j \in U \times U$ è associato un valore $\text{attachment}(U_1, U_2) \in [0, 1]$ che rappresenta l' *attaccamento* di U_1 a U_2 , più questo valore si avvicina a 0 più la probabilità che U_1 smetta di seguire U_2 aumenta e viceversa. Quando il nodo viene creato il valore di **attachment** è relativamente alto (~ 0.8) in quanto ci si aspetta che un utente non smetta di seguire un altro utente poco dopo aver iniziato a seguirlo.

1.2 Probabilità

In tutto il modello sono definite globalmente le probabilità di compiere una determinata azione.

Tweet ogni utente U ad ogni tempo T ha la possibilità di creare un tweet secondo la seguente probabilità:

$$P_{tweet} = \alpha PI(T) followers(U)$$

Questo perchè ci si aspetta che un utente popolare sia più attivo di uno sconosciuto, per mantenere il suo livello di popolarità.

Retweet similmente al precedente un utente U al tempo T ha una possibilità di effettuare un retweet come:

$$P_{tweet} = \beta PI(T) followers(U)$$

Risulta da valutare la possibilità che un utente possa fare più retweet rispetto al numero dei tweet, in quanto essa è un'operazione decisamente meno impegnativa e perciò, ,intuitivamente, con una più elevata probabilità di accadere.

Tag ogni tweet ha una possibilità di contenere un tag diretto ad un altro utente V con probabilità

$$P_{tweet} = \gamma follower(U) followers(V)$$

Ovvero ci si aspetta che il **dtag** sia proporzionale alla popolarità di entrambi gli utenti coinvolti.

1.3 Attività sociali

Per prima cosa definiamo gli oggetti che riguardano le attività sociali:

Tweet Il tweet dell'utente U al tempo T è definito come $\text{tweet}(U, T) = (j, \text{likability}, \text{dislikability}, \text{dtag}, U)$ dove:

- j è il **topic** su cui il tweet verte
- $\text{likability} \in [0, 1]$ indica la probabilità di quanto il tweet possa piacere agli utenti a cui interessa il **topic** j .
- dislikability rappresenta la probabilità di quanto il tweet possa non piacere a chi non è interessato all'argomento

- **dtag** indica l'utente V taggato nel tweet. Tale valore può anche essere nullo.
- U è l'utente che ha effettuato il tweet. Necessario per l'implementazione.

Retweet Il retweet effettuato dall'utente U al tempo T del tweet di V al tempo \bar{T} , definito come: $\text{retweet}(U, T) = \text{tweet}(V, \bar{T})$

In base agli oggetti definiti in precedenza possiamo definire le azioni di:

Post Nel momento in cui l'utente U è abilitato alla creazione di un tweet allora viene generato casualmente un **topic** j su cui verterà il tweet in modo tale che un topic di interesse per U sia selezionato con maggiore probabilità. Una volta scelto il topic j :

- se j è di interesse per U ($U.PI(j) \geq 0.5$) allora il **tweet** risultante avrà una **likability** elevata mentre la **dislikability** sarà casuale.
- viceversa il **tweet** generato avrà una **dislikability** alta ed una **likability** casuale.

Unfollow Un utente U può decidere di smettere di seguire un utente V, grazie alla variabile **attachment** che li lega, e all'ultimo **tweet** di V con probabilità

$$P_{unfollow}(U, V) = P(\text{dislikability} | \text{attachment})$$

Nel caso in cui U decida di continuare a seguire V allora $\text{attachment}(U, V) = P_{unfollow}(U, V)$ aggiornando dunque la probabilità di unfollow a quella appena calcolata. Risulta dunque che la probabilità di unfollow dipende solo dallo stato precedente di **attachment**.
 [•SC: Hidden Markov Model?•]

Follow Esistono diversi tipi di possibili modalità di following:

By Retweet (BR) By retweet Ancora da definire. Probabilità simmetrica a **Unfollow**. [•SC: definire•]

Outside Factor (OF) avviene quando un utente U comincia a seguire un nodo V per fattori esterni al Social Network, quali ad esempio nuova amicizia nella vita reale, nuovo follow su altri mezzi di comunicazioni online, ecc. [•SC: prob da definire•]

Active Search (AS) un utente U ricerca un qualunque altro utente V tramite la rete sociale anche se non ha collegamenti con esso. Tale ricerca ha senso secondo il modello di omofilia [•SC: definire•].