Fenomenologia del gossip

Problema di novembre 2008

1. Regole del gioco 1 “MWoWT”.

a) Ci sono n persone “equivalenti” ed “equicomunicanti”. Ciascuna persona accede al sistema di comunicazione con le stesse modalità delle altre;
b) Ogni persona conosce 1/n della notizia oggetto del gossip; per ricostruire l’informazione completa ogni persona deve ricevere le altre n-1 informazioni in possesso degli altri;
c) Ogni persona manda un messaggio alla volta contenente tutte le informazioni in suo possesso in quel momento ad una persona sola per ogni unità di tempo (*clock*) con modalità unidirezionale; nessuno “salta un turno” senza comunicare le informazioni in suo possesso a qualcuno;
d) Ogni persona conosce il numero e l’identità delle altre $n-1$ persone con cui può (e deve) comunicare;
e) Ognuno può compilarsi l’elenco alfabetico delle n persone coinvolte nello scambio di informazioni ottenendo un elenco ordinato. Al posto del criterio alfabetico potrebbe essere usato un qualsiasi altro criterio perché noto a tutti e basato su un dato oggettivo in modo che tutti siano in grado di prodursi l’elenco che sarà identico a quello ottenuto dagli altri;
f) Le ipotesi contenute in questo elenco non sono da considerarsi arbitrarie in quanto sono deducibili direttamente dalle regole date da Rudy e comunque non ne violano gli assunti.

2. Svolgimento del gioco 1.

- All’inizio, t = 0, ogni persona (indicata con a_1, a_2, a_3, ... a_n) possiede solo la sua informazione cioè 1/n del totale.
- Dopo un’unità di tempo t (in unità arbitrarie) ogni persona ha spedito un messaggio alla persona che la segue nell’elenco alfabetico delle persone coinvolte nella diffusione del pettegolezzo (l’ultimo, cioè a_n , la spedisce ad a_1, come avviene nelle formule cicliche). Di conseguenza ogni persona ha anche ricevuto un messaggio contenente un’informazione di cui non era in possesso. *Quindi, per t =1, ognuno conosce 2 parti della notizia complessiva (2/n).*
- Dopo un’ulteriore unità di tempo ogni persona comunica con un messaggio le due notizie in suo possesso alla persona che la segue due posti dopo, sempre secondo l’elenco alfabetico (regola “e”).
  Quindi per t =2 ognuno conosce 4 parti della notizia complessiva (4/n).
- Passa un’altra unità di tempo t e ogni persona manda un messaggio contenente le quattro informazioni in suo possesso alla persona che la segue 4 posti dopo; se si raggiunge la fine dell’elenco si deve continuare dall’inizio, come già detto.
  Quindi per t =3 ognuno conosce 8 parti della notizia complessiva (8/n).
- Dopo un’ulteriore tempo t la comunicazione continua tenendo presente che bisogna spedire il messaggio contenente le notizie in proprio possesso più quelle appena ricevute alla persona che segue dopo 2^t posti, dove t è il tempo in cui ci si trova misurato in “messaggi” spediti. Questo algoritmo assicura la spedizione (e quindi la ricezione) di notizie che non sono doppioni. La persona

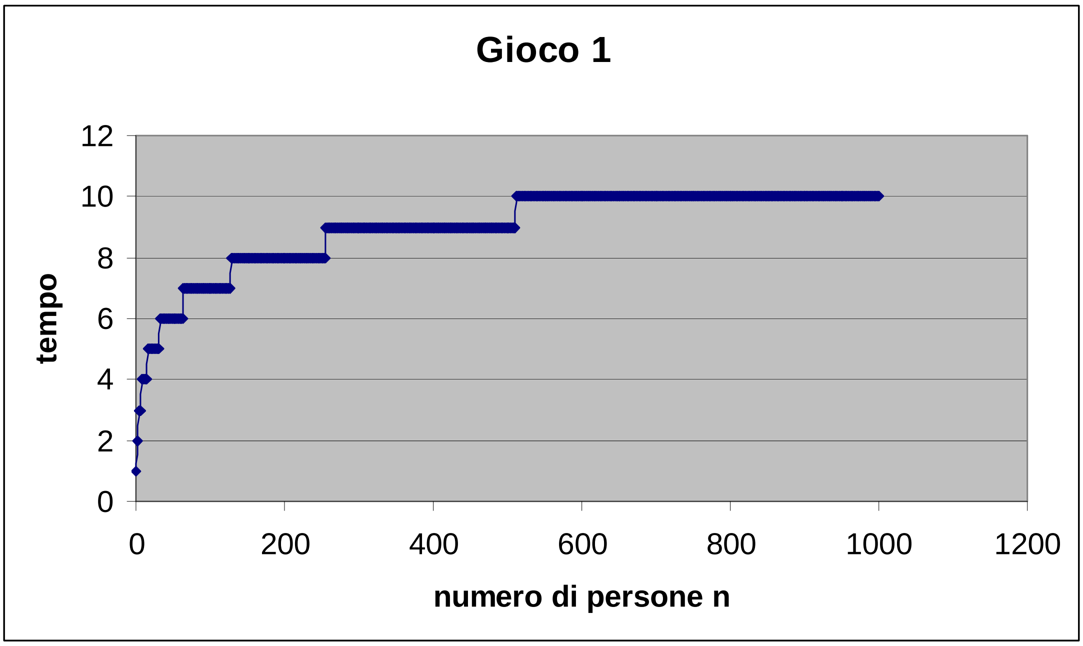
cessa di spedire messaggi quando ha raggiunto la conoscenza di n parti della notizia complessiva, cioè tutto. La rigida osservanza dell'algoritmo precedente assicura che:

a) *Se io ho n parti del messaggio, anche tutti gli altri ne hanno n parti;*
b) *Nessuno può avere doppioni di parti del messaggio;*
c) *Se io ho ricostruito tutto il messaggio, anche tutti gli altri hanno raggiunto lo stesso risultato e quindi la trasmissione di informazioni può cessare.*

3. Conclusione, ovvero *Tiriamo le somme del gioco 1*.

Vediamo adesso di valutare la bontà dell'efficienza della strategia seguita nella diffusione del gossip secondo le regole del gioco 1.
Ad ogni *clock* che impone a tutti di trasmettere un messaggio, l'informazione posseduta dalle persone è uguale a 2^t. Il gioco si conclude quando $2^t = n$.
Questa è un'equazione esponenziale in *t* che ha come soluzione $t = \log_2 n$ equivalente a $t = \dfrac{\log_{10} n}{\log_{10} 2}$. Potendo essere *t* un numero non intero (anzi,sicuramente sarà non intero) bisognerà assumere per *t* il valore intero immediatamente maggiore del valore calcolato: quindi $t = \text{int}\left[\dfrac{\log_{10} n}{\log_{10} 2}\right] + 1$.

Dall'esame dei valori ottenuti possiamo sottolineare che, innanzitutto, la diffusione del pettegolezzo segue la legge esponenziale 2^t, dove *t* è il tempo trascorso in unità arbitrarie; in secondo luogo è da evidenziare che il valore del tempo necessario affinché tutti sappiano tutto è una funzione logaritmica. Ciò significa che al crescere di *n*, numero delle persone cui far conoscere la notizia nella sua interezza ma anche indice della frammentazione dell'informazione, il tempo necessario perché *lo sappiano tutti*, cresce molto lentamente, secondo la curva logaritmica a scalini del grafico seguente:



Dal grafico apprendiamo che, ad esempio, in un paesino di 1000 abitanti, dopo 10 giorni tutti conosceranno una certa notizia che è stata diffusa, ad esempio, con una e-mail al giorno. Se ognuno inviasse, invece, 5 e-mail al giorno, in due giorni l'intera notizia, composta di 1000 frammenti diversi, sarebbe di pubblico dominio.

4. **Valutazione dell'efficienza dell'algoritmo.**

Per valutare l'efficienza dell'algoritmo trovato dobbiamo confrontarlo con il sistema di base della diffusione del gossip (algoritmo "*ingenuo*"). Per ottenere la funzione di quest'ultimo possiamo ipotizzare il seguente comportamento. Ogni persona si incarica di trasmettere il messaggio in suo possesso ad una sola persona alla volta, senza seguire un ordine preciso ma anche evitando doppioni. In questo modo occorreranno *n-1* unità di tempo.

L'efficienza *e* sarà allora data dal rapporto

$$e_1 = \frac{n-1}{\mathrm{int}\left[\frac{\log_{10} n}{\log_{10} 2}\right]+1}$$

Si deduce che l'efficienza è una funzione crescente di *n* e che per $n \to \infty$ implica $e = \infty$. Ad esempio, per n = 1000, ci vorrebbero 999 unità di tempo con il sistema di base mentre con l'algoritmo seguito nel gioco 1 ne bastano 10: l'efficienza è uguale a 99,9, quindi quasi 100 volte più efficiente. Non male!

5. **Gioco 2**

In questo "gioco" si ipotizza che il contatto tra due delle *n* persone coinvolte produca un reciproco scambio di informazioni. Ad ogni contatto quindi si produce un raddoppio delle parti del gossip conosciute da ciascuno, sempre che si curi di evitare contatti con persone che siano portatrici di "doppioni", ossia notizie già note.

a) *t* = 1: Ogni persona telefona alla persona che la segue (dopo un posto) nell'elenco per trasmettergli la notizia in suo possesso e ricevere contemporaneamente quella dell'altro. Ognuno anche riceve una telefonata. Alla fine ognuno possiede 3 informazioni (3/*n*);
b) *t* = 2: Ogni persona telefona alla persona che la segue dopo 3 posti nell'elenco e si scambia le notizie; riceve anche analoga telefonata dalla persona che la precede di tre posti. Alla fine di questo punto ciascuno possiede 9 informazioni (3/*n*);
c) *t* = 3: Ogni persona telefona alla persona che la segue dopo 9 posti, ecc.
d) *t* = … : Ogni persona telefona alla persona che la segue dopo 3^t posizioni; le informazioni possedute dopo un tempo *t* sono $\frac{3^t}{n}$. Valgono le solite convenzioni sulla ciclicità.
e) Il gossip diviene totalmente conosciuto quando ognuno conosce *n* parti (diverse) della notizia, cioè tutto il gossip. Questo accade quando $3^t = n$; $t = \log_3 n$; $t = \frac{\log_{10} n}{\log_{10} 3}$. Anche in questo caso dovremo tenere conto che *t* deve essere un numero intero e quindi la formula diventa $t = \mathrm{int}\left[\frac{\log_{10} n}{\log_{10} 3}\right]+1$.

Il corrispondente dell'algoritmo "*ingenuo*" del punto 4 del gioco 1 questa volta prevede che ogni persona telefoni a quella che la segue nell'elenco scambiando la propria informazione, allo stesso tempo riceverà una telefonata da chi la precede. La telefonata successiva la fa a chi la segue dopo 2 posti, poi dopo 3, ecc. L'informazione cresce secondo $1+2\cdot t$ e quindi per avere tutte le n parti del gossip occorrerebbe un tempo *t* dato dalla soluzione dell'equazione $1+2\cdot t=n$, cioè $t=\frac{n-1}{2}$ cicli di clock, la metà del caso corrispondente del gioco 1.

L'efficienza "*e*" misurata rispetto all'algoritmo "ingenuo" è quindi data da:

$$e_2=\frac{\frac{n-1}{2}}{\mathrm{int}\left[\frac{\log_{10} n}{\log_{10} 3}\right]+1}.$$

Invece l'efficienza dell'algoritmo utilizzato nel gioco 2 per diffondere la notizia (la bidirezionalità della comunicazione) rispetto al gioco 1 (unidirezionalità della comunicazione) è data dal rapporto

$$e=\frac{\log_2 n}{\log_3 n}=\frac{\frac{\log_{10} n}{\log_{10} 2}}{\frac{\log_{10} n}{\log_{10} 3}}=\frac{\log_{10} n}{\log_{10} 2}\cdot\frac{\log_{10} 3}{\log_{10} n}=\frac{\log_{10} 3}{\log_{10} 2}=\frac{0{,}477121}{0{,}301029}=1{,}584962$$

dove è stata omessa la funzione "parte intera" per semplificare l'espressione.

Il calcolo con la parte intera, come propriamente deve essere, ammette come limite il valore sopra indicato. Ad esempio con n = 10000:

n = 10000					
	caso *ingenuo* gioco 1	**gioco 1**	**caso ingenuo gioco 2**	**gioco 2**	**efficienza gioco 2/gioco 1**
tempo necessario (in unità arbitrarie)	**9999**	**14**	**4999,5**	**9**	**1,5556**

6. Epilogo.

A conclusione dei "sudati calcoli", come non ricordare la locuzione latina *Verba volant scripta manent* il cui significato non è, come i più credono, che le cose scritte valgono e si impongono mentre le parole "parlate" sono effimere ed inutili, bensì esattamente il contrario. Basterà ricordare l'etimologia di *manére* (it. *fermarsi, sostare, essere immobile*) e quella di *volare* (*slanciarsi, staccarsi, prendere il volo*).

Le parole si diffondono avvalendosi dell'esponenziale, i latini lo avevano capito; se poi anziché usare l'e-mail incontriamo il nostro interlocutore, la base 3 scaccia la quasi immobile base 2 e il gossip prende davvero… *il volo!*

Decollatura, 16 dicembre 2008.