

Laboratorio di Algoritmi

Progetto “Epidemia” (settembre 2023)

Nota: La scadenza del progetto è fissata per venerdì 15 settembre **compreso**.

Nota: Si consiglia di controllare gli eventuali aggiornamenti di questo documento, che potrebbero riportare le risposte a dubbi degli studenti e la correzione di eventuali errori, di consultare il documento con le indicazioni tassative e le informazioni utili per svolgere e consegnare correttamente il progetto.

Il problema Il progetto si propone di simulare il diffondersi di un’epidemia in una popolazione al livello dei singoli individui. Per ogni individuo è nota l’età e_i e un indice s_i compreso fra 0 e 1 che ne descrive lo stato di salute (1 indica una condizione ottimale e 0 una pessima). Si conoscono anche i contatti di ogni individuo, cioè l’insieme degli altri individui con i quali questi trascorre del tempo a distanza ravvicinata per qualsiasi motivo. I contatti danno luogo a una relazione simmetrica; ad ogni coppia di individui i e j in contatto è associato un indice f_{ij} di frequenza compreso fra 0 e 1 (dove 0 esprime brevi incontri saltuari e 1 una coabitazione continuativa). La frequenza sarà correlata alla potenzialità di trasmissione del contagio dall’uno all’altro individuo.

Si considera un orizzonte temporale costituito da una sequenza di t giorni. Il meccanismo ipotizzato per l’andamento del contagio è il seguente. Ogni giorno gli individui che non sono mai stati contagiati frequentano i propri contatti. Se la somma dei coefficienti di frequenza per i contatti con altri individui infettivi, ma non ancora rilevati, supera strettamente un’opportuna soglia di refrattarietà, un individuo risulta infetto (e non rilevato) a partire dal giorno seguente. La soglia è diversa per ciascun individuo i e pari all’espressione $\rho_i = s_i(1 - e_i/100)^3$, a indicare che individui più sani e giovani sono più refrattari.

Mentre sono in questo stato, gli individui si muovono liberamente e possono diffondere il contagio ai propri contatti. Dopo un’incubazione che si ipotizza di durata pari a g giorni per tutti gli individui, l’infezione viene rilevata. In altre parole, un individuo contagiato nel giorno d è infettivo e non rilevato dal giorno $d + 1$ compreso al giorno $d + g$ compreso. Alcuni individui sono già infettivi al principio dell’orizzonte temporale, per cui possono essere rilevati in anticipo (ad esempio, gli individui infettivi da g giorni vengono rilevati nel primo giorno), ma nessuno è rilevato (men che meno dimesso) prima dell’orizzonte temporale: le prime rilevazioni possono aver luogo solo il primo giorno.

Per semplicità, rappresentiamo l’intero sistema sanitario con un solo ospedale, ignorando la distribuzione spaziale di contagi e cure. Al principio di ogni giorno, gli individui rilevati entrano in coda all’ospedale (dal loro punto di vista siamo nel giorno $d + g + 1$). L’ospedale ha la capacità di ricoverare r individui. Se c’è posto per tutti, i nuovi arrivati vengono immediatamente ricoverati. Altrimenti, si ricoverano gli individui con refrattarietà ρ_i più bassa, assumendo che tale coefficiente sia correlato alla gravità del contagio. In caso di pari refrattarietà, vale l’ordine alfabetico. Gli individui per i quali non c’è posto vanno in quarantena domestica, ma rimangono in coda, cioè al principio di ciascun giorno seguente possono essere ricoverati se si liberano posti in ospedale e non si presentano altri casi di refrattarietà inferiore.

Sia i ricoverati in ospedale sia quelli in quarantena domestica smettono di contagiare i loro contatti, e rimangono nel loro stato oppure passano da quarantena a ospedale finché non vengono dimessi. Al principio di ogni giorno, per ogni individuo ricoverato o in quarantena si valuta il numero di giorni trascorsi dal rilevamento

(compreso quello del rilevamento stesso, il famoso $d+g+1$), assegnando peso doppio ai giorni trascorsi in ospedale, per indicare che sono più efficaci. Se il numero di giorni trascorsi in quarantena più due volte quello dei giorni trascorsi in ospedale supera strettamente la soglia $g_{\max}(1 - \rho_i)$ (da cui risulta che il coefficiente ρ_i è legato anche alla durata del contagio), un individuo viene dimesso o liberato dalla quarantena, fin dalla mattina stessa. Se era ricoverato in ospedale, quindi, lascia un posto libero disponibile il giorno stesso. Chi entra in ospedale ci rimane fino alla dimissione. Gli individui dimessi risultano immuni. In sintesi, ogni giorno prima si dimettono i guariti dall'ospedale e dalla quarantena, poi si mettono in coda i nuovi rilevati, poi si dividono nuovi e vecchi rilevati fra ospedale e quarantena e infine si valutano i nuovi contagi, che cominceranno ad essere infettivi dal giorno seguente.

Lo scopo del progetto è per prima cosa costruire elenchi di individui ordinati in base alle caratteristiche utili per il progetto (descritte più in dettaglio nella sezione seguente). Quindi, occorrerà analizzare la struttura della relazione che descrive i contatti, determinando metriche correlate alla potenziale velocità di diffusione del contagio. Infine, sarà richiesto di simulare l'andamento dettagliato del contagio e di fornire indici statistici aggregati sul numero di contagiati e ricoverati, sul carico e sul livello di qualità del sistema sanitario.

Il progetto Il programma da realizzare carica i dati da un file di testo il cui nome va fornito nella linea di comando. Il file riporta nella prima riga il numero totale di individui, la lunghezza dell'orizzonte temporale, il ritardo g fra infezione e rilevazione del contagio, il parametro g_{\max} che influenza la durata della degenza, la capacità r dell'ospedale. Sono tutti numeri interi positivi. Per esempio:

```
10 15 4 6 3
```

indica che vi sono 10 individui, che l'orizzonte temporale è di 15 giorni consecutivi, che la durata dell'incubazione e del decorso della malattia sono, rispettivamente, di 4 e 6 giorni e che l'ospedale ospita al massimo 3 individui contemporaneamente.

Ciascuna delle righe seguenti è dedicata a un individuo e riporta il cognome (una singola parola di 15 caratteri al massimo), l'età e_i (numero intero fra 1 e 100), lo stato di salute s_i (numero reale fra 0 e 1), il numero di giorni già trascorsi dall'eventuale contagio al principio dell'orizzonte temporale (0 se l'individuo non è contagiato, altrimenti un numero intero positivo $\leq g$), il numero e l'elenco degli individui con cui ha contatto. Questo elenco è formato da coppie, composte dal cognome dell'altro individuo j e dalla frequenza f_{ij} del contatto (numero reale fra 0 e 1). Per esempio:

```
Bernacci 30 0.383 0 2 Nurul 0.406 Vaccaro 0.002
```

indica che Bernacci ha 30 anni, stato di salute 0.383 (mediocre), non è contagiato al principio dell'orizzonte temporale (0) e ha contatti con 2 individui, che sono Nurul (con frequenza 0.406, media) e Vaccaro (con frequenza 0.002, molto bassa).

Il programma deve per prima cosa creare due elenchi ordinati degli individui. Il primo elenco è ordinato per età e_i decrescente (in caso di parità, per stato di salute s_i crescente; in caso di ulteriore parità, in ordine alfabetico). L'elenco è introdotto dalla riga `Ordine per eta'`: a cui segue una riga per ogni individuo, che ne riporta cognome, età e stato di salute (con 3 cifre decimali). Per esempio:

```
Ordine per eta':
Pal 87 0.113
Fiore 77 0.881
Linzi 71 0.056
Galli 65 0.161
```

Nurul 58 0.238
 Bazhanova 56 0.331
 Vaccaro 52 0.862
 Bernacci 30 0.383
 Pracucci 17 0.338
 Piersanti 15 0.629

Il secondo elenco è ordinato per indice ρ_i crescente (e, in caso di parità, ordine alfabetico) in un formato analogo al precedente. L'elenco è introdotto dalla riga **Ordine per refrattarietà'**: a cui segue una riga per ogni individuo, che riporta il cognome e l'indice di refrattarietà con 3 cifre decimali:

Ordine per refrattarietà':
 Pal 0.000
 Linzi 0.001
 Galli 0.007
 Fiore 0.011
 Nurul 0.018
 Bazhanova 0.028
 Vaccaro 0.095
 Bernacci 0.131
 Pracucci 0.193
 Piersanti 0.386

Per descrivere la potenziale velocità di diffusione del contagio, il programma determinerà per ogni coppia di individui il numero minimo di passaggi da individuo a individuo necessari a trasmetterlo dall'uno all'altro. Si tratta di una specie di distanza e coincide col minimo numero di giorni richiesto dalla trasmissione nell'ipotesi pessima che ogni contatto fra individui porti a un contagio immediato. La coppia di individui per i quali tale numero minimo è il più alto è una specie di "diametro" della popolazione ed è legato alla velocità di diffusione dell'epidemia. Per ottenere un risultato univoco, fra le coppie di individui a distanza massima si riporterà la prima in ordine alfabetico, ponendo anche i due individui in ordine alfabetico. Il programma deve stampare il numero di passaggi, seguito dalla parola chiave **giorni:** e dalla coppia di individui corrispondente. Nell'esempio considerato:

6 giorni: Linzi Nurul

indica che la massima distanza tra due individui è 6 e la prima coppia in ordine alfabetico posta a tale distanza è formata da Linzi e Nurul (in questo caso il percorso è unico, ma la soluzione non cambierebbe se ve ne fosse più d'uno).

La simulazione dell'andamento del contagio verrà riportata stampando una riga per ogni individuo, in ordine alfabetico. La riga deve cominciare con il cognome, esteso esattamente su 15 caratteri e allineato a sinistra, eventualmente con spazi bianchi terminali per mantenere l'incolonnamento¹, e proseguire con un carattere al giorno, che indichi lo stato dell'individuo: un punto ('.') per individui non infettivi (cioè sani, guariti o appena contagiati il giorno stesso), una 'I' per individui infettivi e non rilevati, una 'O' per individui ricoverati in ospedale e una 'Q' per individui in quarantena domestica. Si ricordi che gli individui in coda per il rilevamento vengono valutati e assegnati all'ospedale o alla quarantena domestica il giorno stesso. Per esempio:

Bazhanova .IIII000.....
 Bernacci

¹Basta la specifica %-15s.

Fiore	.IIII000.....
Galli	I000.....
Linzi	...IIIIQ000....
Nurul
Pal	.IIII000.....
Piersanti
Pracucci	..IIIIQ000....
Vaccaro

indica che il primo giorno Galli è infettivo, contagia Bazhanova, Fiore e Pal, ecc...

Al termine, occorre stampare le seguenti informazioni statistiche, una per riga: il numero totale dei contagiati seguito dalla parola chiave **contagiati**², quello dei ricoverati in ospedale seguito dalla parola chiave **ricoverati**, quello degli individui che sono ancora contagiati l'ultimo giorno dell'orizzonte temporale seguito dalla parola chiave **rimasti** e il numero di individui che sono stati contagiati e non subito ricoverati seguito dalle parole chiave **individui in attesa**, il numero totale dei giorni di degenza in ospedale seguito dalle parole chiave **giorni di degenza**, il numero totale dei giorni di quarantena domestica seguito dalle parole chiave **giorni di quarantena**, e infine la durata media dell'attesa del ricovero per i soli individui che l'hanno subita (quindi, il rapporto fra il numero totale dei giorni in quarantena e il numero di individui che sono stati contagiati e non subito ricoverati, compresi quelli che non sono stati mai ricoverati). Quest'ultima informazione va espressa come numero reale con 3 cifre decimali seguito dalle parole chiave **giorni in attesa**. Se nessun individuo ricade in questa categoria, il valore non va riportato perché è indefinito. Nell'esempio considerato:

```
6 contagiati
6 ricoverati
0 rimasti
2 individui in attesa
17 giorni di degenza
3 giorni di quarantena
1.500 giorni in attesa
```

indica che 6 individui sono stati contagiati e tutti sono stati ricoverati. Nessuno è ancora ammalato al termine dell'orizzonte temporale, e 2 hanno trascorso almeno un giorno in quarantena domestica. Complessivamente, hanno trascorso 17 giorni in ospedale e 3 in quarantena domestica, con una media di 1.500 giorni in quarantena prima di entrare in ospedale per i 2 individui che non sono stati ricoverato immediatamente.

Chiarimenti

²Per semplicità tutte le parole chiave restano plurali, anche se il numero è 1 o 0.