

# Laboratorio di Algoritmi

## Progetto “Contagio” (giugno 2021)

**Nota:** La scadenza del progetto è fissata per venerdì 18 giugno **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. Si consiglia anche di consultare l’altro documento che riporta avvertenze utili per lo svolgimento del progetto.

**Il problema** È una vecchia tradizione accademica citare i fatti di attualità nei temi d’esame, un po’ per aggiungere una patina di realismo a esercizi virtuali, un po’ per distrarsi facendo ironia spicciola. Dopo un anno di pandemia, intravedendosi una luce di speranza, ecco l’inevitabile progetto a tema. Mi auguro non infastidisca nessuno: un tema simile sarebbe uscito comunque, prima o poi, dato che l’uso di modelli matematici e strumenti informatici è comune in questo campo. Ovviamente, il problema qui proposto è solo un pretesto, estremamente semplificato, per applicare i concetti descritti nel corso e si basa su ipotesi scorrette e irrealistiche.

Si suppone di essere in grado di descrivere la popolazione interessata a livello dei singoli individui. Di ognuno è nota l’età  $e_i$  e un indice  $s_i$  dello stato di salute compreso fra 0 e 1 (dove 1 descrive una condizione ottimale e 0 una pessima). Di ogni individuo si conoscono poi i contatti, ovvero l’insieme degli altri individui con i quali trascorre del tempo a distanza ravvicinata (per lavoro, svago, amicizia, ecc...). Tale relazione è considerata simmetrica, e 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). Vedremo poi che la frequenza è correlata alla potenzialità di trasmissione del contagio dall’uno all’altro individuo.

Si considera un orizzonte temporale costituito da una sequenza di  $t$  unità di tempo discrete (per esempio, 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).

Al principio di ogni giorno, gli individui rilevati entrano in coda all’ospedale (dal loro punto di vista siamo nel giorno  $d + g + 1$ ). Per semplicità, rappresentiamo l’intero sistema sanitario con un solo ospedale, ignorando l’aspetto della distribuzione spaziale. Se c’è posto per tutti, vengono immediatamente ricoverati. L’ospedale ha però una capacità massima di ricovero di  $r$  individui. Se non c’è posto per tutti, si ricoverano gli individui con refrattarietà  $\rho_i$  più bassa, assumendo che tale coefficiente sia correlato anche alla gravità del contagio. 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, a 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  $\rho_i$  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. Gli individui dimessi risultano immuni.<sup>1</sup>

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 gruppi di individui particolarmente legati fra loro e 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), 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  $f_{i,j}$  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

<sup>1</sup>Ho cercato di essere quanto più chiaro possibile, ma ci possono essere sottigliezze legate alla sequenza delle azioni. 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 solo dal giorno seguente. Gli esempi dovrebbero aiutare, ma scrivetemi in caso di dubbi.

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

```
Ordine per refrattarieta':
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 determinare i gruppi di individui più strettamente legati fra loro, trascureremo temporaneamente i contatti meno frequenti, cioè considereremo solo le coppie di individui per le quali la frequenza è  $f_{ij} \geq \max(\rho_i, \rho_j)$ , vale a dire alta rispetto alla refrattarietà di entrambi gli individui. Indicheremo come appartenenti allo stesso gruppo gli individui che sono direttamente o indirettamente legati dalle relazioni rimanenti. Il programma dovrà stampare una riga contenente il numero di tali gruppi seguito dalla parola chiave **gruppi** e i cognomi degli individui appartenenti a ciascun gruppo, una riga per gruppo, ordinando alfabeticamente ciascun gruppo al proprio interno e l'insieme dei gruppi rispetto al primo individuo di ciascun gruppo. Per esempio:

```
5 gruppi
Bazhanova Fiore Galli Pracucci
Bernacci Nurul
Linzi
Pal
Piersanti Vaccaro
```

indica che nel problema di esempio si ottengono 5 gruppi, composti dagli individui elencati, rispettando l'ordine richiesto.

Per descrivere invece la potenziale velocità di diffusione del contagio, il programma determinerà per ogni coppia di individui il numero minimo di passaggi necessari a trasmetterlo dall'uno all'altro (una specie di "distanza", uguale al numero di giorni richiesto dalla trasmissione nelle condizioni più favorevoli), e quindi la coppia di

individui per i quali tale numero minimo è il più alto (una specie di “diametro” della popolazione). Per rendere univoco il risultato, fra le coppie di individui a distanza massima 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’incolonnamento<sup>2</sup>, 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 .....
Fiore .IIII000.....
Galli I000.....
Linzi ....IIII000...
Nurul .....
Pal .IIII000.....
Piersanti .....
Pracucci ...IIIIQ00....
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`<sup>3</sup>, 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
```

---

<sup>2</sup>Basta la specifica `%-15s`.

<sup>3</sup>Per semplicità tutte le parole chiave restano plurali, anche se il numero è 1 o 0.

0 rimasti  
1 individui in attesa  
17 giorni di degenza  
2 giorni di quarantena  
2.000 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 uno ha trascorso almeno un giorno in quarantena domestica. Complessivamente, hanno trascorso 17 giorni in ospedale e 2 in quarantena domestica, con una media di 2.000 giorni in quarantena prima di entrare in ospedale per l'unico individuo che non è stato ricoverato immediatamente.

## Chiarimenti

- Quando si rileva un nuovo contagiato, ma non c'è posto in ospedale, se c'è una persona ricoverata con indice  $\rho_i$  maggiore, il nuovo paziente la a sostituirla o rimane in coda?

I ricoverati restano in ospedale fino a che superano la soglia di dimissione. Un nuovo contagiato per cui non 'è posto va in quarantena domestica e rimane in coda finché non si libera un posto, anche se ci sono ricoverati meno gravi.

- Possiamo supporre che nel giorno 1 l'ospedale sia vuoto, cioè nel giorno 1 nessuno sia in ospedale?

L'ospedale è *inizialmente* vuoto, ma se un individuo inizialmente è contagiato da  $g$  giorni, viene rilevato nel giorno 1, si mette in coda ed entra in ospedale la mattina del giorno 1. Quindi nel giorno 1 l'ospedale può non essere vuoto.