

# Laboratorio di Algoritmi e Strutture Dati

Docente: V. Lonati

Progetto “Malattie infettive”

valido per gli appelli di giugno e luglio 2014

## 1 Il problema

L'epidemiologa Esantema de Malatis si occupa della diffusione delle malattie infettive in età prescolare e sta monitorando la popolazione del suo distretto socio-sanitario, di cui fanno parte *migliaia di soggetti*, ciascuno identificato in maniera unica dal suo codice fiscale. La diffusione delle malattie infettive avviene per contatto con i soggetti infetti.

**Esempio 1.** *Anna, Barbara, Carlo, Davide, Elisa, Federico e Giulio hanno vari contatti tra loro: Anna, Barbara e Carlo frequentano la stessa scuola d'infanzia; Anna è sorella di Federico; Elisa e Davide fanno nuoto assieme una volta ogni due settimane; Giulio si incontra spesso con Federico al parco giochi.*

### Catena di contatti

Anche se due soggetti  $x$  e  $y$  non sono in contatto diretto, l'infezione può passare da  $x$  a  $y$  tramite una *catena di contatti* che coinvolge altri soggetti  $t_1 = x, t_2, \dots, t_{n-1}, t_n = y$  tali che  $t_i$  è in contatto con  $t_{i+1}$  per ogni  $i = 1, 2, \dots, n-1$ . L'intero  $n-1$  è detto *lunghezza* della catena.

**Esempio 2.** *Consideriamo i soggetti dell'esempio 1. Se Barbara si ammala, allora anche Federico potrebbe ammalarsi tramite questa catena di contatti: Barbara è in contatto con Carlo (frequentano la stessa scuola d'infanzia); Carlo è in contatto con Anna (frequentano la stessa scuola d'infanzia); Anna è in contatto con Federico (sono fratello e sorella).*

*Si noti che questa non è la catena di contatti più corta tra Barbara e Federico; un'altra catena di lunghezza 2 si ottiene tramite il contatto di Barbara con Anna e di Anna con Federico.*

### Area di contagio e untori

Chiamiamo *area di contagio* di un soggetto l'insieme di tutte le persone che sono in contatto con quel soggetto, di tutte quelle che sono in contatto con quest'ultime, e così via.

Chiamiamo *untore* tra due soggetti  $x$  e  $y$  un soggetto  $u$  tale che:

- $u$  è nell'area di contagio sia di  $x$  che di  $y$ ;
- se non ci fosse  $u$ , allora  $x$  non sarebbe nell'area di contagio di  $y$ .

**Esempio 3.** *Consideriamo i soggetti dell'esempio 1. Se non ci sono altri contatti oltre a quelli elencati, allora l'area di contagio di Carlo è formata da: Anna, Barbara, Giulio e Federico, mentre Davide e Elisa non ne fanno parte.*

*Inoltre Anna è untore tra Barbara e e Giulio, o tra Carlo e Federico.*

### Valore caratteristico di contatto

Il rischio di contagio varia a seconda della frequenza e del tipo di contatto. Ad esempio nei nidi e nelle scuole d'infanzia le occasioni di contagio sono molto frequenti; altre occasioni di contagio sono date dalla

frequentazione di spazi comuni come piscine, palestre, ludoteche, o dal fatto di abitare con bambini in età prescolare.

La Dottoressa Esantema raccoglie informazioni sui contatti tra i soggetti che abitano nel suo distretto e in particolare utilizza un indicatore, il cosiddetto *valore caratteristico di contatto*, per indicare qual'è il rischio di contagio tra due soggetti in contatto tra loro. Più precisamente, il valore caratteristico  $v(x, y)$  del contatto tra due soggetti  $x$  e  $y$  è dato da un numero intero non negativo: i contatti di valore caratteristico 1 sono quelli più a rischio, mentre valori caratteristici maggiori indicano che il rischio è inferiore.

Se esiste una catena di contatti tra  $x$  e  $y$ , allora definiamo anche il *valore caratteristico di contagio*  $c(x, y)$  tra  $x$  e  $y$  come l'intero minimo  $c$  per cui esistono  $n$  soggetti  $t_1 = x, t_2, \dots, t_{n-1}, t_n = y$  tali che

$$c = v(t_1, t_2) + v(t_2, t_3) + \dots + v(t_{n-1}, t_n)$$

**Esempio 4.** Consideriamo i soggetti dell'esempio 1. Il valore caratteristico dei loro contatti può essere stimato come segue:  $v(\text{Anna}, \text{Barbara}) = v(\text{Anna}, \text{Carlo}) = v(\text{Barbara}, \text{Carlo}) = 1$ ;  $v(\text{Anna}, \text{Federico}) = 1$ ;  $v(\text{Elisa}, \text{Davide}) = 10$ ;  $v(\text{Federico}, \text{Giulio}) = 3$ .

Allora il valore caratteristico di contagio tra Barbara e Giulio è pari a 5, dato dalla somma

$$v(\text{Barbara}, \text{Anna}) + v(\text{Anna}, \text{Federico}) + v(\text{Federico}, \text{Giulio}).$$

## Soggetti a maggior rischio

Il *rischio di contagio* per un soggetto  $x$  è funzione del valore caratteristico di contatto con tutti i suoi contatti e, più precisamente, è definito come

$$r(x) = \sum_{y \text{ in contatto con } x} \frac{1}{v(x, y)}$$

I soggetti a *maggior rischio* nella popolazione sono quelli aventi il rischio di contagio massimo.

**Esempio 5.** Consideriamo i soggetti e i contatti dell'esempio 1. Allora il soggetto a maggior rischio è Anna che ha rischio pari a

$$\frac{1}{v(\text{Anna}, \text{Barbara})} + \frac{1}{v(\text{Anna}, \text{Carlo})} + \frac{1}{v(\text{Anna}, \text{Federico})} = 1 + 1 + 1 = 3.$$

Federico ha invece rischio di contagio pari a  $4/3$ .

## Immunizzazione e bacino di una malattia

Quando l'organismo entra in contatto con gli agenti patogeni responsabili di una certa malattia infettiva, sviluppa la capacità di difendersi prontamente in caso di nuovo futuro attacco dello stesso agente patogeno: si parla in questo caso di *immunizzazione* alla malattia.

L'immunizzazione può anche avvenire mediante *vaccinazione*, in questo caso il soggetto non sviluppa la malattia vera e propria.

Se un soggetto è immune ad una malattia  $M$ , allora non può trasmettere il contagio, ovvero la malattia non può trasmettersi lungo una catena di contatti che contiene un soggetto immune.

Chiamiamo *bacino* per una malattia  $M$  rispetto ad un soggetto  $x$  l'insieme di tutte le persone che sono in contatto con quel soggetto e non sono immuni a  $M$ , di tutte quelle che sono in contatto con quest'ultime e non sono immuni a  $M$ , e così via.

**Esempio 6.** Consideriamo i soggetti dell'esempio 1. Se Federico è immune alla varicella, allora il bacino della varicella rispetto a Carlo si riduce rispetto alla sua area di contagio perché restano esclusi Federico e Giulio.

## Tempo di contagio

Ogni malattia ha un *tempo di incubazione* specifico, che può variare leggermente da soggetto a soggetto. Il *tempo stimato di incubazione*  $t(M)$  di una malattia  $M$  indica il numero di giorni che passano solitamente dal momento del contagio al momento in cui compaiono i sintomi della malattia  $M$ .

Il *tempo minimo di contagio*  $t_M(x, y)$  tra due soggetti  $x$  e  $y$  per la malattia  $M$  è dato dal prodotto tra il tempo stimato di incubazione  $t(M)$  per la malattia  $M$  e la lunghezza minima delle catene di contatti (che coinvolgono solo soggetti non immuni a  $M$ ) tra i soggetti  $x$  e  $y$ . Per convenzione, se tra  $x$  e  $y$  non esistono tali catene, allora  $t_M(x, y)$  si definisce pari a infinito.

**Esempio 7.** *Consideriamo i soggetti dell'esempio 1. Se la varicella ha tempo stimato di incubazione di 15 giorni, allora il tempo minimo di contagio della varicella tra Barbara e Federico è 30 giorni. Se Anna è immune, allora tale tempo minimo diventa infinito.*

## Campagna di vaccinazione

La Dottoressa Esantema de Medicis è responsabile di una campagna di vaccinazione per prevenire la diffusione nel suo distretto di alcune malattie infettive particolarmente contagiose; la campagna prevede che tutti i bambini del distretto debbano recarsi in un centro vaccinale al compimento del primo anno d'età. I centri vaccinali devono pertanto essere insediati in modo da favorirne l'accesso a tutta la popolazione.

C'è una strada che collega tutti i paesi del distretto; la strada può essere pensata come un asse a coordinate intere, e la *posizione* di ciascun paese è pertanto identificata da una singola coordinata intera. Non ci sono mai due paesi nella stessa posizione, e possiamo considerare come *distanza* fra due posizioni il valore assoluto della differenza delle loro coordinate.

I centri vaccinali si devono collocare in alcuni paesi, ma non necessariamente in tutti (naturalmente, se un centro vaccinale viene costruito in un paese, il centro vaccinale e il paese si trovano nella stessa posizione). Per collocare i centri vaccinali, occorre sceglierne le posizioni in modo tale che la somma totale di tutte le distanze fra ciascun paese e il più vicino centro vaccinale sia minima. Una volta fissate le posizioni dei paesi e il numero  $n$  di centri vaccinali, una *collocazione* degli  $n$  centri vaccinali si considera *ottima* se realizza la minima somma totale di tali distanze.

**Esempio 8.** *Se il distretto è formato da 10 paesi, posizionati alle coordinate 1 2 3 6 7 9 11 22 44 50, e si vogliono collocare 5 centri vaccinali, allora la somma minima delle distanze dai centri vaccinali è pari a 9 e si ottiene ad esempio collocando i centri vaccinali nei paesi alle coordinate 2 7 22 44 50.*

## 2 Specifiche di progettazione

Si deve progettare un programma che sia in grado di gestire i dati relativi alla diffusione delle malattie infettive.

Non potete fare assunzioni sulla dimensione della popolazione osservata dalla Dott.ssa Esantema, né sul numero delle malattie esaminate.

La progettazione deve prevedere la scelta delle strutture dati da usare per rappresentare i dati e gli algoritmi da applicare per poter eseguire in maniera efficiente le operazioni descritte in seguito. Non basta limitarsi a riferimenti generici alle tecniche algoritmiche utilizzate (es: "l'operazione X si risolve con un algoritmo greedy") ma è necessario dettagliare le procedure da utilizzare, tramite pseudocodice o direttamente in linguaggio C, eventualmente facendo riferimento alla letteratura sugli algoritmi citati.

In particolare si richiede di analizzare, in funzione delle scelte di progettazione fatte, quale risulta essere il costo delle diverse operazioni richieste dalla specifica.

Si richiede inoltre di fornire una rassegna *esaustiva* di esempi che potrebbero essere usati per testare il programma e che mettono in evidenza particolari caratteristiche del suo funzionamento (non solo casi

tipici di input, ma anche casi limite e/o situazioni patologiche; input che evidenzino la differenza di prestazioni tra le soluzioni progettuali scelte e altre più semplicistiche).

Non si richiede un'implementazione completa del progetto; è necessario però fornire l'ossatura del programma (possibilmente suddiviso su più file) contenente in particolare: le definizioni dei tipi fondamentali, i prototipi delle funzioni che realizzano le operazioni specificate nella traccia, e tutte le porzioni di codice utili ad illustrarne il loro funzionamento e uso.

Facoltativamente, è possibile consegnare un'implementazione completa e funzionante del progetto o di alcune sue parti. In questo caso è necessario commentare il codice e fornire nella relazione indicazioni precise sul formato dell'input atteso e sul formato dell'output prodotto, assieme ad alcune *coppie di file* di input e relativo output da usare per testare il codice.

## 2.1 Operazioni da eseguire

- **soggetto** ( $x$ )

Se il soggetto  $x$  (identificato dal codice fiscale) non è già monitorato, viene messo sotto monitoraggio.

- **contatto** ( $x, y, v$ )

Se i soggetti  $x$  e  $y$  sono sotto monitoraggio, annota che  $x$  e  $y$  sono in contatto, e che tale contatto ha valore caratteristico  $v$ .

- **area**( $x$ )

Se il soggetto  $x$  è sotto monitoraggio, stampa tutti i soggetti sotto monitoraggio che si trovano nell'area di contagio di  $x$ .

- **untore**( $x, y$ )

Se i soggetti  $x$  e  $y$  sono sotto monitoraggio, stabilisce se esiste un untore tra  $x$  e  $y$  e ne stampa il codice fiscale.

- **valore** ( $x, y$ )

Se i soggetti  $x$  e  $y$  sono sotto monitoraggio ed esiste una catena di contatti fra loro, calcola il valore di contagio tra  $x$  e  $y$ .

- **maggiore**()

Determina i soggetti, tra quelli sotto monitoraggio, che risultano a maggior rischio.

- **malattia**( $M$ )

Se la malattia  $M$  (identificata da una stringa univoca) non è ancora monitorata, viene messa sotto monitoraggio.

- **incubazione**( $M, n$ )

Se  $M$  è una malattia sotto monitoraggio, annota che il tempo di incubazione per la malattia  $M$  è pari a  $n$  giorni.

- **tempo**( $x, y, M$ )

Se la malattia  $M$  e i soggetti  $x$  e  $y$  sono sotto monitoraggio, ed è noto il tempo di incubazione di  $M$ , calcola il tempo minimo di contagio tra i soggetti  $x$  e  $y$  per la malattia  $M$ .

- **immune**( $x, M$ )

Se la malattia  $M$  e il soggetto  $x$  sono sotto monitoraggio, annota che  $x$  è immune alla malattia  $M$ .

- **contagio** ( $x, y, M$ )

Se la malattia  $M$  e i soggetti  $x$  e  $y$  sono sotto monitoraggio, verifica se esiste una catena di contatti tra  $x$  e  $y$  che coinvolge solo soggetti non immuni a  $M$  e, in caso affermativo, stampa una delle catene di lunghezza minima che coinvolgono solo soggetti non immuni a  $M$ .

- **bacino**( $M, x$ )

Se il soggetto  $x$  e la malattia  $M$  sono sotto monitoraggio, stampa tutti i soggetti sotto monitoraggio che si trovano nel bacino di  $M$  rispetto a  $x$ .

- **vaccinazione**( $n, m, posizioni$ )

Dato un numero  $n$  di paesi e il vettore delle loro posizioni, calcola una collocazione ottima di  $m$  centri vaccinali e stampa la somma minima di tutte le distanze tra ogni paese e il centro vaccinale a lui più vicino, e l'elenco delle posizioni (in ordine crescente) dei paesi in cui vanno collocati i centri vaccinali.

Si noti che le operazioni richieste sono liberamente implementabili; in particolare, non vanno necessariamente intese come prototipi di funzioni.

### 3 Modalità di consegna

La presente traccia è valida per gli appelli di giugno e luglio 2014.

La relazione (non meno di 3, non più di 10 pagine in formato pdf o rtf) va inviata per posta elettronica all'indirizzo `lonati@di.unimi.it` entro lunedì 14 luglio.

**Per coloro che intendono sostenere la prova orale nel mese di giugno, la scadenza è anticipata a martedì 24 giugno 2014.**

La relazione e gli altri file aggiuntivi (file sorgenti C, esempi di input, ecc) devono essere contenuti in un unico archivio `.zip` il cui nome dovrà essere della forma `cognome.matricola.zip`. Tutti i file nell'archivio, compresa la relazione, devono riportare nome, cognome e matricola dell'autore.

In generale non è prevista una discussione orale dei progetti, ma in alcuni casi potranno essere richiesti dei chiarimenti via mail o dal vivo all'autore delle relazione.

La realizzazione del progetto è una prova d'esame da svolgersi **individualmente**. I progetti giudicati frutto di **copiatura** saranno **estromessi** d'ufficio dalla valutazione.

La versione aggiornata del progetto è pubblicata in `.pdf` sul sito: <http://lonati.di.unimi.it/algo/>. Si consiglia di consultare periodicamente questo sito per eventuali correzioni e/o precisazioni relative al testo del progetto. Per ogni ulteriore chiarimento potete chiedere un appuntamento scrivendo una mail all'indirizzo `lonati@di.unimi.it`.