

Laboratorio di algoritmi e strutture dati

13 novembre 2013

In questa esercitazione si propone di ragionare su alcune situazioni (pseudo-)reali che possono essere modellate attraverso l'uso di grafi. Lavorate in piccoli gruppi (2/3 persone al massimo) e analizzate le tre situazioni descritte cercando di tradurle nel linguaggio della teoria dei grafi.

1 CSI - feb 2012

Un truculento assassinio è avvenuto in un vicolo vicino a Trafalgar Square nell'ora di punta. Donald Algos, l'ispettore capo di Scotland Yard, ha avviato le indagini in modo minuzioso, raccogliendo le dichiarazioni di numerosissimi testimoni e ricavando informazioni utili da alcuni supporti tecnici disponibili (registrazioni delle videocamere che controllano il traffico, ricevute dei parchimetri, ecc). I dati sono troppi per poterli analizzare a mano, ma l'ispettore Algos non si scoraggia perchè ha a disposizione CSI, il *Calcolatore Speciale per Investigazioni*, un sistema informatico che può aiutarlo a ricostruire quanto è accaduto e scoprire chi è il colpevole.

Obiettivo del progetto è sviluppare un sistema che simuli CSI e che sappia svolgere in modo efficiente le operazioni descritte in seguito, su una dimensione arbitraria dei dati in ingresso.

Testimoni Molte persone risultano presenti a Trafalgar Square il giorno del delitto: sono tutte considerate *testimoni*. Ciascun testimone è identificato da un nome (non ci sono due testimoni con lo stesso nome); indichiamo con \mathcal{N} l'insieme dei nomi dei testimoni. Se due testimoni risultano presenti in Trafalgar Square nello stesso momento, diciamo che si sono *incontrati*. La scena del delitto è considerata *deserta all'orario x* se dalle testimonianze nessun testimone risulta presente all'orario x .

Testimonianze orarie Una *testimonianza oraria* consiste nella dichiarazione di un testimone, il quale afferma di essere stato in Trafalgar Square il giorno del delitto, specificando a che ora è arrivato e a che ora se n'è andato; tali testimonianze sono rappresentate da triple della forma $(y, arrivo_y, partenza_y)$, dove y è il nome del testimone mentre $arrivo_y$ e $partenza_y$ sono i due orari di arrivo e partenza e $arrivo_y < partenza_y$. Un *sistema di testimonianze orarie* è un insieme di testimonianze orarie ciascuna relativa ad un testimone diverso. Si osservi che si sta assumendo che ogni testimone è stato sulla scena del delitto una e una sola volta.

Precedenze orarie Una *precedenza oraria* è invece un'informazione che riguarda una coppia di testimoni in \mathcal{N} che non si sono incontrati: la precedenza oraria (y, z) , dove $y, z \in \mathcal{N}$, indica che il testimone di nome y se n'è andato prima dell'arrivo del testimone di nome z .

Un insieme \mathcal{P} di precedenze orarie può *indurre logicamente* eventuali altre precedenze orarie. In generale, indicheremo con $\bar{\mathcal{P}}$ l'insieme ottenuto aggiungendo a \mathcal{P} tutte le precedenze orarie indotte logicamente da \mathcal{P} .

Un insieme \mathcal{P} di precedenze orarie potrebbe però essere contraddittorio, infatti potrebbero esserci dei testimoni che mentono o errori nella raccolta dei dati. Più precisamente, diremo che \mathcal{P} è *contraddittorio* quando esistono due testimoni n e m tali che sia (m, n) che (n, m) sono precedenze orarie in $\bar{\mathcal{P}}$.

Se \mathcal{P} non è contraddittorio, diremo che \mathcal{P} è un *sistema di precedenze orarie*; in questo caso:

- se $(y, z) \in \bar{\mathcal{P}}$, allora diremo che y è andato via prima dell'arrivo di z secondo \mathcal{P} ;
- se $(z, y) \in \bar{\mathcal{P}}$, allora diremo che z è andato via prima dell'arrivo di y secondo \mathcal{P} ;
- se né (y, z) né (z, y) sono in $\bar{\mathcal{P}}$, allora si intende che *sicuramente* y e z si sono incontrati in qualche momento.

Ricostruzioni Un sistema \mathcal{T} di testimonianze orarie è una *ricostruzione* su \mathcal{N} per un sistema \mathcal{P} di precedenze orarie se, per ogni coppia di testimoni y e z in \mathcal{N} , si verifica una e una soltanto delle tre seguenti possibilità:

- y e z si sono incontrati secondo \mathcal{T} ;
- y è andato via prima dell'arrivo di z secondo \mathcal{P} ;
- z è andato via prima dell'arrivo di y secondo \mathcal{P} .

Naturalmente le informazioni date dalle precedenze orarie sono meno precise di quelle date dalle testimonianze orarie: infatti in genere sono possibili diverse ricostruzioni per un dato insieme di precedenze orarie.

Testimoni chiave, sospettati, complici, bande. Un testimone y è considerato *testimone chiave* se, dalle informazioni raccolte, ogni altro testimone ha incontrato sulla scena del delitto al massimo lo stesso numero di persone incontrate da x .

Un insieme di testimoni $\{y_1, y_2, \dots, y_n\}$ è considerata una potenziale *banda* se, dalle informazioni raccolte, in ogni istante nell'arco di tempo che va dall'arrivo del primo (ad arrivare) degli y_i alla partenza dell'ultimo (a partire) degli y_i , almeno uno degli y_i è stato sicuramente presente mentre non è stato mai presente nessun altro testimone al di fuori degli y_i .

Il testimone y è *sospettato* se, dalle informazioni raccolte, risulta essere rimasto solo sulla scena del delitto, *in ogni possibile ricostruzione* di \mathcal{P} su \mathcal{N} .

I testimoni y e z sono potenziali *complici* se, secondo le informazioni raccolte, sono stati presenti sulla scena del delitto contemporaneamente ma senza altri testimoni, *in ogni possibile ricostruzione* di \mathcal{P} su \mathcal{N} .

2 Compagnie aeree - gennaio 2013

Numerose *compagnie aeree* si spartiscono il traffico aereo internazionale che collega vari *aeroporti* sparsi in giro per il mondo. Le compagnie aeree e gli aeroporti sono identificati da stringhe di lunghezza arbitraria. Ciascuna compagnia offre ai suoi clienti alcune *tratte* specifiche, che collegano un aeroporto di partenza con uno di arrivo. Ogni tratta è caratterizzata da una *lunghezza* in chilometri e dal *prezzo* in dollari con cui è messa in vendita alla clientela. Se la compagnia C offre una tratta t che parte dall'aeroporto a e arriva all'aeroporto b , allora diciamo che a e b sono *serviti* da C . Se $C \in \mathcal{C}$ offre una tratta dall'aeroporto a all'aeroporto b , tale tratta può essere percorsa anche in senso inverso, e dunque collegare b ad a .

Non tutte le coppie di aeroporti sono collegate da una tratta; a volte il *viaggio* tra gli aeroporti di partenza e di arrivo è ottenuto combinando più tratte successive e prevede quindi degli *scali*, ovvero degli aeroporti di passaggio tra quelli di partenza e di arrivo. In questo caso la lunghezza e il prezzo del viaggio sono date dalla somma delle lunghezze e dei prezzi, rispettivamente, delle tratte che definiscono il *viaggio*.

Si dice che un *viaggio* è *offerto* da una certa compagnia aerea C se tutte le tratte che lo compongono sono offerte da C .

Una compagnia aerea $C \in \mathcal{C}$ cerca di predisporre un *piano di voli*, vale a dire un insieme $P(C)$ di tratte tale che

1. $P(C)$ è *sufficiente*: ossia le tratte in $P(C)$ permettono di viaggiare da a a b per ogni coppia di aeroporti serviti da C ;
2. La somma delle lunghezze delle tratte in $P(C)$ è minima rispetto a tutti gli insiemi sufficienti di tratte.

Si osservi che non tutte le compagnie aeree dispongono necessariamente di un piano di voli, e che una compagnia può disporre di diversi piani di voli.

3 Citazioni - giu/lug 2011

Il Ministero dell'Istruzione ha deciso di monitorare in maniera piú rigorosa l'attività di ricerca universitaria per migliorarne ed aumentarne la produttività. L'obiettivo principale è ottenere delle statistiche che permettano di associare ad ogni ricercatore vari indici di produttività, per poter riorganizzare le attività in modo da creare le condizioni in cui anche i ricercatori meno produttivi possano incrementare le loro pubblicazioni. Aspetto fondamentale in tale contesto è rappresentato dalle pubblicazioni di ciascun ricercatore, non solo dal punto di vista quantitativo ma anche qualitativo. Ogni articolo infatti può essere collocato in una conferenza o su rivista (nazionale o internazionale), a seconda della collocazione l'articolo assume una rilevanza diversa. Inoltre, ogni articolo contiene una serie di citazioni per altri articoli e, a sua volta, può essere citato da altre pubblicazioni. Quindi, può essere interessante valutare quale impatto un articolo ha avuto sulla comunità internazionale, considerando il numero di citazioni ricevute o se le citazioni ricevute derivano da articoli di grande rilevanza internazionale. Talvolta può essere utile nella valutazione dell'operato di un ricercatore anche la vicinanza (in termini di citazioni) con articoli di riferimento o di rassegna nella specifica area di ricerca, oppure le sue collaborazioni con autori di fama internazionale.

Ricercatori e pubblicazioni. Un ricercatore è rappresentato dal suo nome, una stringa identificativa di lunghezza arbitraria. Un articolo scientifico è caratterizzato dai seguenti elementi:

- il codice identificativo;
- la collocazione editoriale, che può essere: rivista internazionale, conferenza internazionale, rivista nazionale, conferenza nazionale;
- il titolo;
- i suoi autori;
- la bibliografia, ovvero un insieme di altre pubblicazioni che vengono citate.

Coll.	Id	Titolo	Autori	Art. citati
RI	1	the paper	Alan Turing, Douglas Hofstadter, Seymour Papert	
CI	2	algorithms for everybody	Don Knuth, Edsger W. Dijkstra	1
CI	3	un'eterna ghirlanda brillante per finta	Douglas Hofstadter, Edsger W. Dijkstra	1, 2
RN	4	logo programming language	Seymour Papert	3
RI	5	algorithms on graphs	Don Knuth, Paul Erdos	1, 2
CI	6	are you human?	Alan Turing	3
RI	7	mathematicians on computers	Alan Turing, Seymour Papert	5, 6
RN	8	combinatorics on graphs	Don Knuth, Paul Erdos	2
RI	9	divulgazione o scienza?	Alan Turing, Douglas Hofstadter	1, 5, 8

Tabella 1: Elenco Pubblicazioni

Distanza di Collaborazione tra due ricercatori. Due ricercatori si dicono *co-autori* se hanno pubblicato un articolo di cui sono entrambi autori. La *distanza di collaborazione* tra due ricercatori¹ $r_1 \neq r_2$ è il minimo intero n per cui esiste una sequenza a_0, a_1, \dots, a_n di autori a_i tali che $a_0 = r_1$, $a_n = r_2$ e a_i è coautore di a_{i+1} per ogni $i = 0, 1, \dots, n-1$. Se non esiste alcuna sequenza di questo tipo, si dice che r_1 e r_2 hanno *distanza di collaborazione indefinita*. Chiaramente si assume che un autore sia coautore di se stesso.

Inoltre ci interessa poter calcolare la distanza di collaborazione tra due autori, considerando solo le pubblicazioni che hanno una data collocazione.

Distanza editoriale di un articolo da un altro. Siano p_1 e p_2 due pubblicazioni distinte; diciamo che p_1 ha *distanza editoriale* n da p_2 , e scriviamo $n = d(p_1, p_2)$, se n è il minimo intero per cui esiste una sequenza a_0, a_1, \dots, a_n di articoli a_i tali che $a_0 = p_1$, $a_n = p_2$ e a_i cita l'articolo a_{i-1} per ogni $i = n, n-1, \dots, 1$.

Se non esiste alcuna sequenza di questo tipo, si dice che p_1 ha *distanza indefinita da* p_2 . La distanza di un articolo a da un altro articolo di riferimento (ad esempio una rassegna recente A) può essere considerata un indice di rilevanza: piú bassa

¹Prendiamo ispirazione dal *Numero di Erdos*, si veda <http://www.oakland.edu/enp/>.

è la distanza, piú rilevante è l'articolo; ad esempio, gli articoli citati direttamente da A sono probabilmente i piú significativi e si trovano infatti a distanza 1 da A .

Inoltre, ci interessa calcolare la distanza editoriale tra due pubblicazioni che hanno una data collocazione.