

Laboratorio di Algoritmi

Progetto “Oligomeri” (febbraio 2022)

Nota: La scadenza del progetto è fissata per lunedì 14 febbraio **compreso**.

Nota: Si consiglia di consultare sulla pagina web il documento che riporta le avvertenze utili per lo svolgimento del progetto. Si consiglia anche di verificare di tanto in tanto gli aggiornamenti a questo documento, che potranno riportare risposte ai dubbi degli studenti e correzioni di eventuali errori.

Il problema Una delle tecniche più fruttuose nello studio dei farmaci è la simulazione di dinamica molecolare, che riproduce al calcolatore il comportamento delle molecole investigate in base alle leggi della Meccanica e dell'Elettrodinamica. La simulazione parte dalle posizioni iniziali degli atomi (modellati come punti nello spazio a tre dimensioni) che costituiscono un insieme dato di molecole e genera le posizioni da loro assunte negli istanti successivi. Siccome le interazioni descritte sono fisiche, non chimiche o nucleari, gli insiemi degli atomi e delle molecole e la relazione di appartenenza degli atomi alle molecole non cambiano. Nei limiti entro i quali la simulazione è fedele, essa equivale a disporre di un microscopio ad altissima risoluzione che riprenda le molecole, per così dire, nel loro “ambiente naturale”.

Uno degli aspetti interessanti di tale analisi è il continuo gioco di rottura e formazione dei *legami a idrogeno*, che svolgono ad esempio un ruolo estremamente importante nel definire le proprietà di sostanze come gli oligomeri (brevi sequenze di aminoacidi) immersi in soluzione acquosa. Ogni atomo di idrogeno del sistema appartiene a una molecola ed è in essa legato con un legame chimico covalente a un atomo *donore* di un altro elemento. L'atomo di idrogeno, però, può anche stabilire un debole legame temporaneo con un atomo *accettore*, anch'esso di un elemento chimico diverso, appartenente ad un'altra molecola. Nei casi più comuni, donore e accettore sono atomi di ossigeno, ma possono essere anche atomi di azoto, zolfo, ecc. . . La causa del legame è che l'unico elettrone dell'idrogeno tende a non trovarsi più esclusivamente intorno al proprio nucleo, ma spostato verso quello dell'atomo donore, il che tende a creare una debole (e fluttuante) carica positiva nella direzione opposta, la quale a sua volta attira gli elettroni dell'atomo accettore. A questo punto, il nucleo dell'idrogeno risulta legato sia al donore (dal forte legame covalente, di natura chimica) sia all'accettore (dal debole legame a idrogeno, di natura elettrostatica). I legami a idrogeno sono molto più deboli rispetto a quelli chimici, e si rompono e riformano di continuo. Nonostante ciò, determinano molte proprietà fisiche importanti dell'acqua, degli oligomeri, delle proteine e degli acidi nucleici. Per convenzione, si ritiene che si stabilisca un legame a idrogeno quando

1. donore e accettore sono vicini, cioè la distanza fra loro è inferiore a una data soglia $d_m = 3.5 \text{ \AA}$;
2. la terna donore-idrogeno-accettore forma una linea quasi retta, cioè l'angolo formato da essa supera una data soglia $\alpha_m = 150^\circ$.

La descrizione fine del sistema attraverso le coordinate di tutti gli atomi racchiude tutte le informazioni utili sulle interazioni fra le molecole del soluto e quelle del solvente, ma in modo troppo implicito. Si vogliono rendere esplicite alcune di tali informazioni estraendo proprietà di livello più alto dai risultati grezzi della simulazione. In particolare, la topologia dei legami a idrogeno, cioè le coppie di molecole interagenti fornisce informazioni più utili e leggibili.

Il progetto Il progetto richiede la stesura di un programma che legga da un file l'immagine di una soluzione acquosa in un determinato istante di tempo. Il file rispetta il formato PDB (Protein Data Bank), secondo il quale ogni riga descrive le caratteristiche di un atomo. Ai fini del progetto, semplificheremo il formato supponendo che ogni riga del file sia costituita da:

- la parola chiave `ATOM`;
- un indice intero progressivo;
- una stringa di caratteri (al massimo 4) che identifica l'elemento chimico dell'atomo;
- una stringa di caratteri (al massimo 8) che identifica la sostanza chimica della molecola cui l'atomo appartiene;
- l'indice intero progressivo della molecola nel file;
- le coordinate rispetto agli assi x , y e z dell'atomo, in Ångström.

Per esempio:

```
...
ATOM 32 H URE 4 13.800 14.055 14.470
ATOM 33 O WAT 5 11.596 7.495 20.457 ...
```

indica che il trentaduesimo atomo è di idrogeno, appartiene alla una molecola di urea, che è la quarta, e ha coordinate (13.800, 14.055, 14.470), mentre il trentatreesimo atomo è di ossigeno, appartiene a una molecola di acqua, che è la quinta e ha coordinate (11.596, 7.495, 20.457).

Il programma deve per prima cosa ricostruire i legami a idrogeno, applicando le convenzioni indicate più sopra e stampare a video il numero l dei legami a idrogeno e il numero n degli atomi di idrogeno, nel formato:

l legami su n atomi

Quindi, deve calcolare per ogni molecola il numero di altre molecole alle quali essa è legata da legami a idrogeno (definito *grado* della molecola). Una molecola può essere legata a un'altra da molti legami a idrogeno, ma il loro contributo al grado è unitario, perché l'altra molecola è una. Si deve quindi stampare il numero di molecole che hanno ciascun grado, in ordine crescente da zero al grado massimo, secondo il formato:

```
Grado 0:  $n_0$  molecole
Grado 1:  $n_1$  molecole
...
Grado  $g_{\max}$ :  $n_{g_{\max}}$  molecole
```

Definito *cluster* un insieme di molecole legate da legami a idrogeno, direttamente o indirettamente, si devono determinare i *cluster* e stamparli in ordine di cardinalità $n^{(i)}$ decrescente. In caso di pari cardinalità, si consideri l'indice minimo fra le molecole componenti un *cluster* e si stampino i *cluster* per indice minimo crescente. Il formato deve essere il seguente:

```
Cluster 1:  $n^{(1)}$  molecole -  $m_1^{(1)}$   $m_2^{(1)}$   $m_3^{(1)}$  ...
Cluster 2:  $n^{(2)}$  molecole -  $m_1^{(2)}$   $m_2^{(2)}$   $m_3^{(2)}$  ...
...
```

dove $n^{(i)} \geq n^{(i-1)}$ per ogni i , le molecole $m_j^{(i)}$ di ogni *cluster* i occupano la stessa riga e sono ordinate per indici crescenti: $m_j^{(i)} < m_{j+1}^{(i)}$ per ogni j .

Definiamo poi *distanza* fra due molecole il numero minimo di legami a idrogeno necessario a passare da una molecola all'altra: 1 se sono legate direttamente, 2 se sono entrambe legate a una terza (ma non fra loro), $+\infty$ se non sono legate nemmeno indirettamente. Ogni *cluster* ha un *diametro* che è la distanza massima fra due molecole in esso contenute; se il *cluster* consiste di una sola molecola, ha diametro nullo. Ovviamente il diametro è sempre un numero intero non negativo. Il programma deve stampare il diametro del primo *cluster* (quello di cardinalità massima, ed eventualmente indice minimo), nel formato seguente:

Diametro: δ

Infine, si attribuisca ad ogni legame a idrogeno una *lunghezza*, data dalla distanza in Å fra donore e accettore e si calcoli l'albero ricoprente di lunghezza minima per il primo *cluster*, considerando per ogni coppia di molecole del *cluster* fra cui esista almeno un legame a idrogeno la lunghezza minima di tali legami. Si stampi il risultato nel formato seguente, indicando anche il numero di lati che formano l'albero ricoprente stesso:

Costo MST: ℓ (m legami)

riportando il costo ℓ con 3 cifre decimali.

Esempio Si consideri il seguente esempio, costituito da cinque molecole di acqua, cioè 15 atomi:

```

ATOM 1 O WAT 1 20.807 4.147 17.051
ATOM 2 H WAT 1 21.623 3.888 16.622
ATOM 3 H WAT 1 20.190 4.281 16.331
ATOM 4 O WAT 2 20.382 0.571 16.989
ATOM 5 H WAT 2 20.902 -0.232 16.955
ATOM 6 H WAT 2 20.841 1.127 17.619
ATOM 7 O WAT 3 21.115 2.271 19.201
ATOM 8 H WAT 3 21.528 2.692 19.955
ATOM 9 H WAT 3 21.095 2.951 18.528
ATOM 10 O WAT 4 24.474 22.754 20.464
ATOM 11 H WAT 4 25.093 22.226 19.962
ATOM 12 H WAT 4 25.024 23.296 21.030
ATOM 13 O WAT 5 23.357 21.041 22.507
ATOM 14 H WAT 5 22.483 21.426 22.567
ATOM 15 H WAT 5 23.741 21.441 21.726

```

L'analisi delle posizioni spaziali rivela che i dieci atomi di idrogeno sono impegnati in tre legami a idrogeno. Delle cinque molecole, una ha due legami, mentre le altre ne hanno uno solo a testa. Le molecole formano due *cluster*, rispettivamente di cardinalità 3 e 2. Il primo *cluster* ha diametro 2 e l'albero minimo che lo ricopre è formato da due lati, con un costo complessivo di 5.754 Å. Il risultato è quindi il seguente:

```

3 legami su 10 atomi
Grado 0: 0 molecole
Grado 1: 4 molecole
Grado 2: 1 molecole
Cluster 1: 3 molecole - 1 2 3
Cluster 2: 2 molecole - 4 5

```

Diametro: 2
Costo MST: 5.754 (2 legami)

Chiarimenti

In questa sezione saranno riportate le risposte a domande e dubbi.