

Chevauchement de séquences

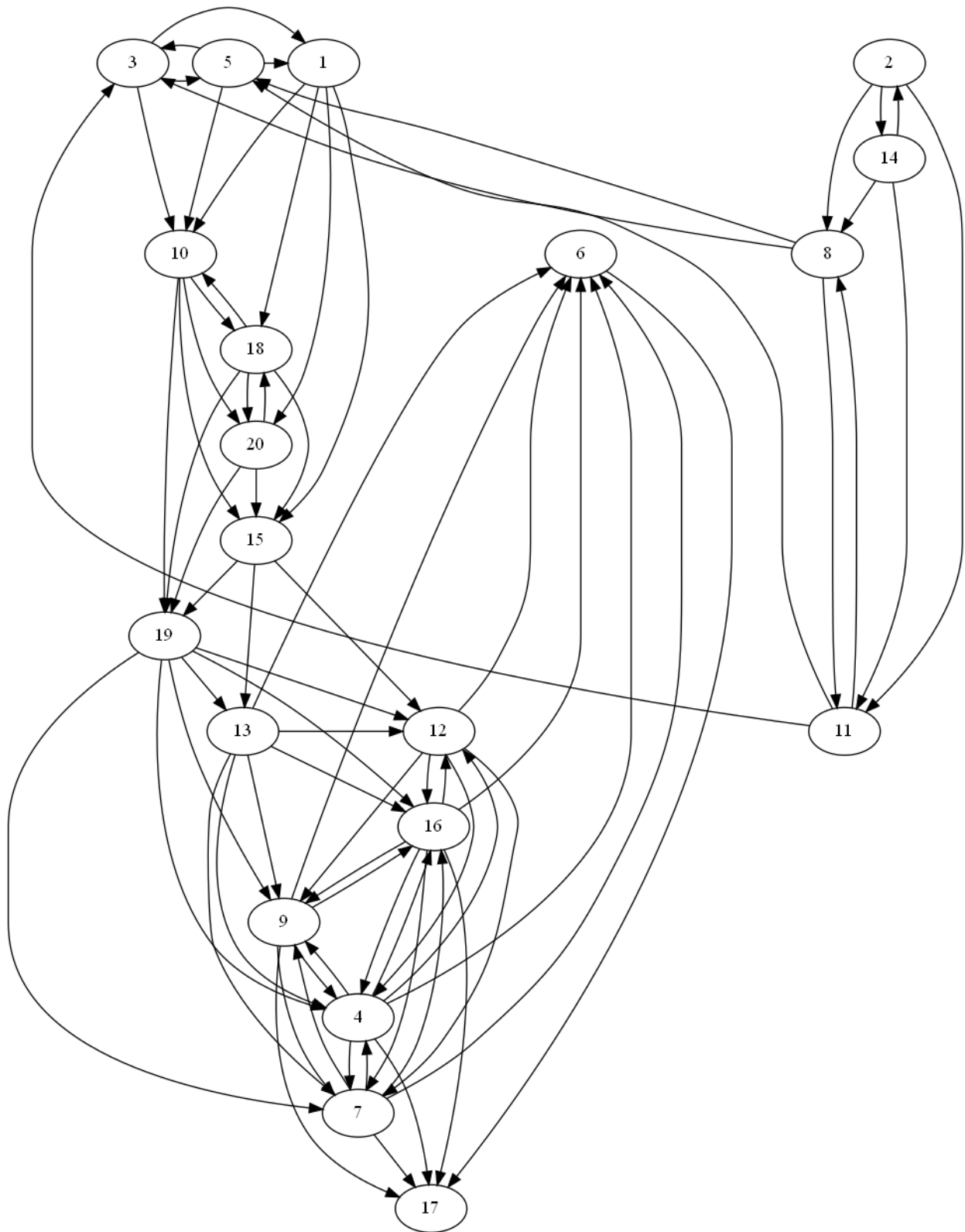
1) La différence entre cet alignement et l'alignement global est que au lieu d'essayer d'avoir les deux séquences une par dessus l'autre avec le moins de décalage on essaie d'avoir le suffixe de S_1 aligné avec le préfix de S_2 . Donc S_1 et S_2 seront décalées dans l'alignement.

2) $V(0, j) = -j$ $V(i, 0) = 0$. La 1^{re} colonne contient des zéros car on peut ne pas s'arrêter au début du mot avec le suffixe. La 1^{re} ligne contient $-j$ car on ne veut pas s'arrêter avant d'avoir eu le préfix complet du mot avec le préfixe.

$$3) V(i, j) = \max \begin{cases} V(i-1, j) = 8 \\ V(i, j-1) = 8 \\ V(i-1, j-1) - 4 \text{ si } v_i \neq w_j \\ V(i-1, j-1) + 4 \text{ sinon} \end{cases}$$

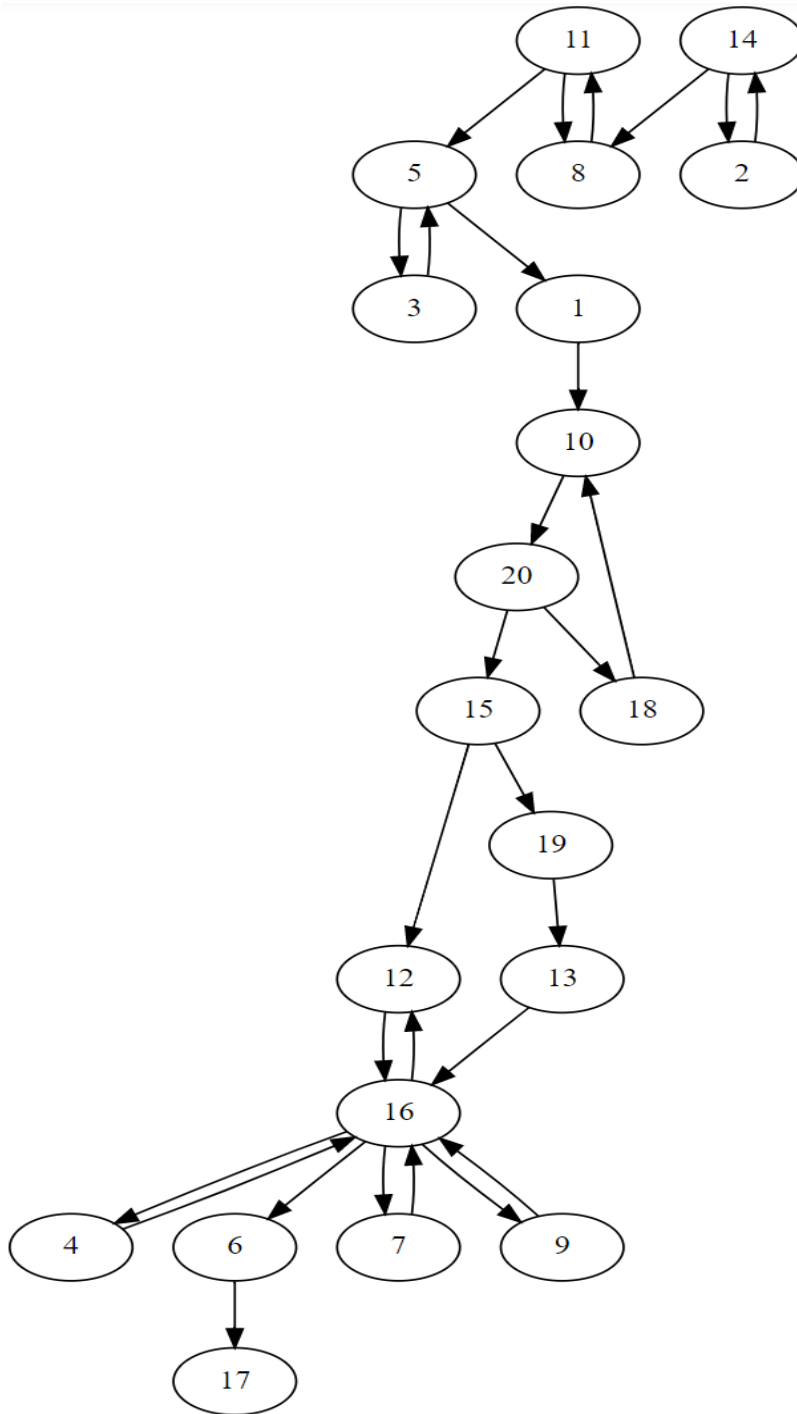
4) on regarde la dernière ligne et on trouve la valeur maximale de cette ligne. Ensuite, on suit les pointeurs jusqu'à temps qu'on retourne à la 1^{re} colonne.

Question 2a :

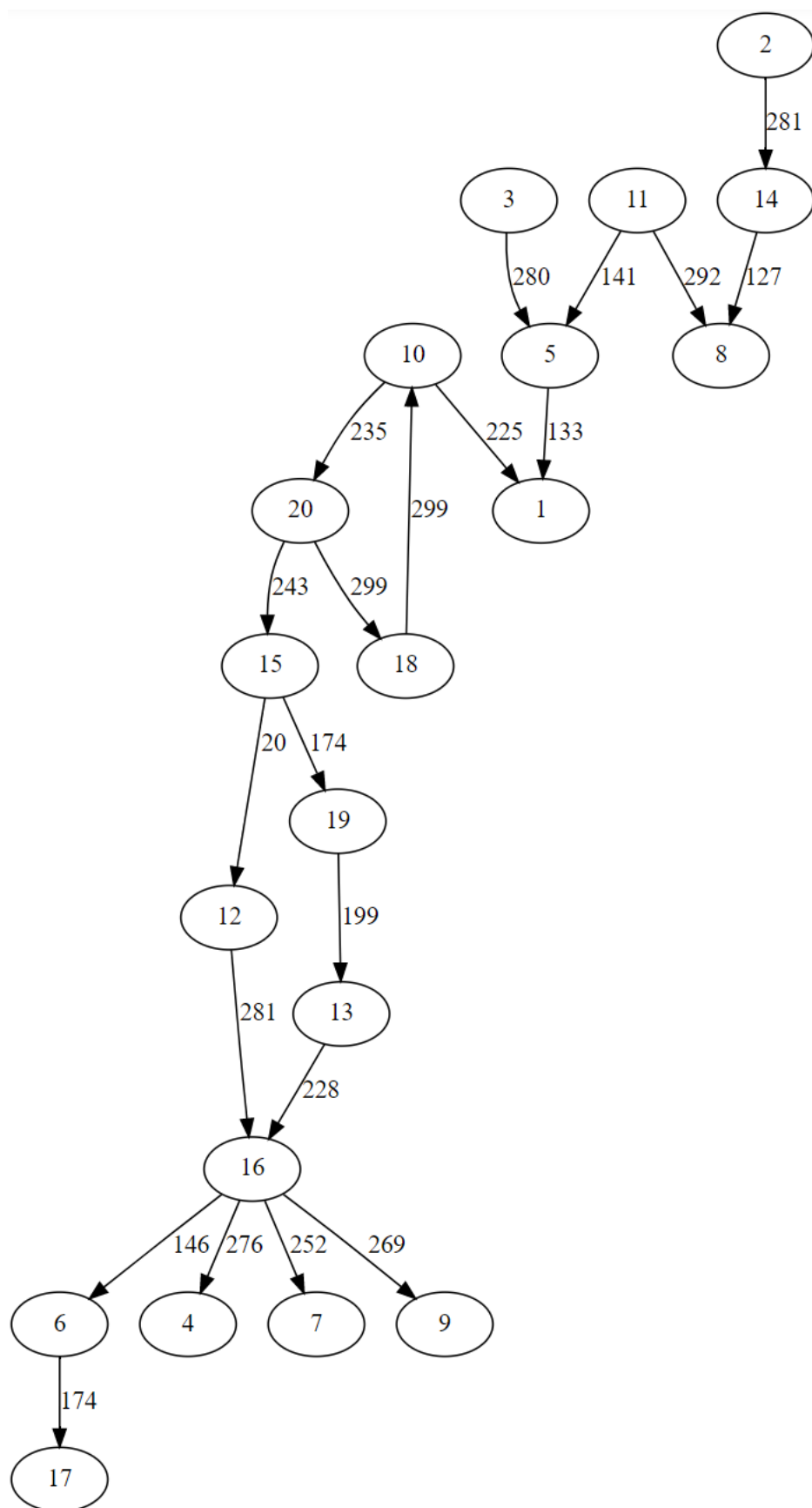


Question 2c :

Voici le graphe après avoir appliqué le principe de réduction transitive. Toute fois il est avec les conditions énumérées au numéro 1, on peut enlever les arêtes dans le mauvais sens à l'aide des scores obtenus.



Voici le graphe final après avoir enlevé les arêtes du mauvais sens et les chiffres sur les arêtes correspondent à la longueur de chevauchement.



Question 2c:

La séquence du fragment génomique séquencé est :

```
CTTTGTTTCAGTTGGCTAGAGATTTACTACATCCGACCTTGAATATGAAAAGAAAAACATAAAAAGAAACGCCTA
GTACAAAGTCCAACCTTCTTACTTTATGGATGTAAAATGTCCAGGTAAAATTTGAAATCTTAATTCCTTTACTAAAGA
AAATTTCTGTAGGGATTGCTAGTGTGGTGTGTATAGTTAAGATACATTAGAATCCTCTGTTGAGTAGAAGTGGGAT
TACAGAATTGGACATGTCAGGGACAATTCATAATAAATGTACTGAAACCATTGTAGAAACAT
```

```
CTTTGTTTCAGTTGGCTAGAGCTTTACTACATCCGTCCTTGAAGAGGAAAAGAAAAACATAAAAAGAAACGCCTA
GTACAAAGTCCAAATCTTACTTTATGGATGTAAAATGTCCCGGTCAAATTTGAAATCTTAATTCCTTTACTAAAGA
AAATTTCTGAAGGGATTGCTAGTGTGGTGTGTATAGTTAAGATACATTCTGAATCCTCTGTTGAGTAGAAGTGGGAT
TACAGAATTGGAAATGTGAGGGACATTTTCATAATAAATGTACTGAAACCATTGTAGAAACAT
```

La longueur de chevauchement est 292 et son score est 1072.

Question 3a :

Le codon start (M) se trouve dans chaque cadre de lecture. (Faire rouler le coder pour voir notre résultat des cadres)

Question 3b:

b) On va appliquer l'algorithme de Nussinov, donc voici nos conditions initiales :

$D(i,i) = 0$ pour $1 \leq i \leq n$

$D(i, i-1) = 0$ pour $2 \leq i \leq n$

La relation de récurrence est :

$D(i,j) = \max \{D(i+1,j), D(i,j-1), D(i+1,j-1) + p(S[i], S[j]), \max_{i+1 \leq k \leq j-2} [D(i,k) + D(k+1,j)]\}$

$p(S[i], S[j]) = 1$ si $S[i] - S[j]$ forme une paire de base sinon c'est égale à 0.

Pour l'alignement optimal, on commence par la valeur se situant en haut à droite dans la matrice et on backtracking vers «le bas» de la matrice en fonction des flèches qu'on a fait pour calculer la matrice.