



Traitement de fichiers et interprétation de graphes

Ingénieur réseau, vous êtes missionné pour un audit sur un réseau de télécommunication national, qui possède près de 10 000 stations éparpillées sur tout le territoire français.

Vous trouverez dans le fichier *donnees.csv*, la liste des villes hébergeant une station du réseau. Vous disposez pour chaque ville de différentes informations dont son nom, ses coordonnées géographiques et son nombre d'habitants.

1. Proposez une fonction en Python qui renvoie un dictionnaire ayant pour clefs les villes hébergeant une station et pour valeurs le nombre d'habitants correspondant.
2. Quel est le nombre exact de stations qui composent ce réseau ?
3. Vous décidez d'isoler un sous-réseau, constitué des 20 villes les plus peuplées. Proposez une fonction en Python qui renvoie la liste correspondante à ce réseau.

D'une ville à l'autre, certaines stations sont reliées entre elles. Vous trouverez dans le fichier *liens.txt*, l'ensemble des liens qui existent entre ces stations.

4. Proposez une fonction en Python qui retourne une liste contenant les liens qui ont pour ville d'origine et pour ville d'arrivée, des villes de votre sous-réseau.
5. Combien de ces liens existent-ils ?

Le package Python *networkx* possède une classe *Graph* qui permet de créer un graphe non orienté. Une fois créé, on peut utiliser sur un objet *Graph* sa méthode *add_egge* qui prend en argument deux nœuds pour les ajouter au graphe et les relier.

6. Installez le package et créez un graphe représentant votre sous-réseau d'étude.

Le package Python *matplotlib.pyplot* possède une fonction *show* qui ne prend pas d'argument mais permet d'afficher un graphe précédemment dessiné avec la fonction *draw_networkx* du package *networkx*.

7. Installez le package et affichez un graphe représentant votre sous-réseau d'étude.

Après avoir isolé votre sous-réseau d'étude, vous vous lancez dans l'analyse de ce dernier.

Pour répondre aux questions qui suivent, vous ne devez utiliser que les fonctions et les attributs listés ci-dessous. Pour savoir comment les utiliser, vous pouvez vous référer à la documentation du package *networkx*, auquel tous ces outils appartiennent.

nodes - *nodes()* - *edges()* - *density()* - *degree()* - *shortest_path_length()* - *shortest_path()* -
betweenness_centrality()

8. Quel est le nombre de stations connectées ?
9. Quel est le nombre de câbles utilisés par le réseau ?
10. Comment expliquez-vous la différence entre le nombre de câbles utilisés par le réseau et le nombre de liens récupérés à la question 5 ?
11. Quel est le degré d'interconnexion du réseau (en %) ?
12. Quelle est la station la plus connectée ?
13. Quelle est la station la moins connectée ?
14. Quelles sont les deux stations les plus éloignées ? Quelle est la distance qui les sépare ? Quel est le chemin à parcourir ?
15. Quelle est la station la plus stratégique du réseau ?
16. Quelle est la station la moins stratégique du réseau ?

La drépanocytose

La drépanocytose est la maladie génétique la plus fréquente en France. Elle résulte d'une mutation dans l'ADN du gène codant pour l'hémoglobine.

L'hémoglobine est une protéine qui se trouve dans les globules rouges. Son rôle est de fixer l'oxygène pour pouvoir la transporter depuis les poumons vers tous les organes du corps.

Dans la drépanocytose, la mutation de l'hémoglobine entraîne une modification de la forme du globule rouge qui adopte une forme caractéristique de faucille.

Le premier symptôme de cette maladie est une anémie, c'est-à-dire la diminution du nombre

d'hémoglobines dans le sang. En effet, à cause de leur forme particulière, les globules rouges qui portent une hémoglobine mutée, deviennent beaucoup plus fragiles. Alors que la durée de vie moyenne d'un globule rouge est de 120 jours chez un patient non atteint, elle dégringole à une vingtaine de jours chez les personnes drépanocytaires. Résultat, il y a beaucoup moins de globules rouges pour transporter l'oxygène dans le corps, entraînant pâleur, fatigue chronique et parfois une jaunisse.

Autres symptôme résultant de la forme particulière des globules rouges : les crises vaso-occlusives. En effet, cette forme en faucille n'est pas adaptée à la circulation dans les vaisseaux sanguins. Il peut donc arriver que ces globules rouges déformés se bloquent et obstruent des petits vaisseaux sanguins, provoquant ces crises vaso-occlusives souvent extrêmement douloureuse. Si une crise survient au niveau du poumon, on parle d'un syndrome thoracique aigu. L'oxygénation de tout l'organisme est alors compromise et le risque de décès, très important.

Si on regarde cette mutation plus en détail, on réalise que la drépanocytose est due à une mutation ponctuelle, unique, faux-sens, du gène bêta globine (HBB) situé sur le chromosome 11. On a, chez les patients atteints, une substitution par transversion d'une Thymine (T) en une Adénine (A), entraînant le remplacement d'un Acide glutamique (E) de codon GAG par une Valine (V) de codon GTG.

Séquences nucléiques

...ACTCCTGAGGAGAAG...

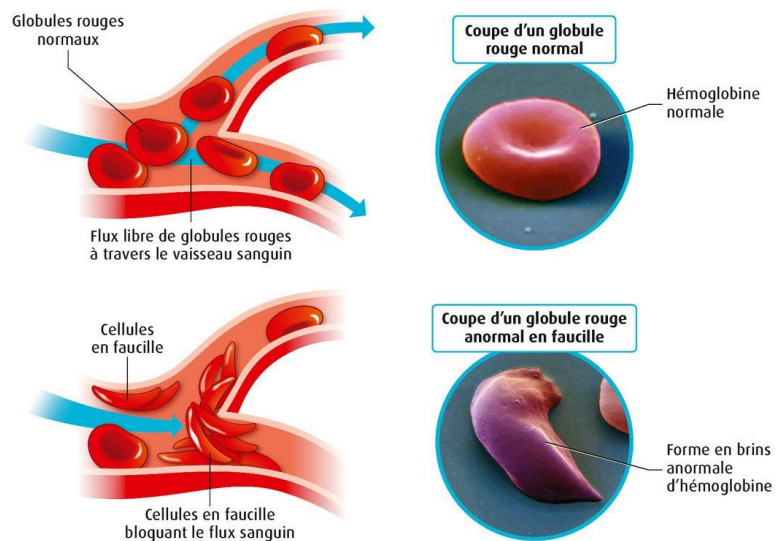
...ACTCCTGTGGAGAAGT...



Séquences protéiques

...TPEEK...

...TPVEK...



© Belin Éducation/Humensis, 2019 SVT 1re enseignement de spécialité
© Amandine Wanert ; Eye of Science/Phanie

A l'aide de vos connaissances et de l'outil *String* (string-db.org), récupérez le réseau d'interaction de la protéine HBB, exportez-le depuis les options du site, puis créez une fonction en Python permettant d'afficher le réseau génétique. Mettez en page et commentez le résultat obtenu.