Visualisation d'Information Projet – Visualisation de réseau métabolique –

Objectif: L'objectif de ce projet est de produire une représentation (lisible) d'un réseau métabolique complet. Pour cela, vous trouverez ci-dessous plusieurs sous-objectifs à atteindre (le barème est donné à titre indicatifs). Le résultat sera remis sous la forme d'un fichier contenant le code python (compatible avec la vue Tulip « Python Script View ») permettant à partir d'un fichier au format Tulip (cf ci-après) d'atteindre ces différents sous-objectifs. Vous devrez aussi fournir un rapport d'une dizaine de pages expliquant les algorithmes et choix que vous aurez faits.

Modalités : Projet à réaliser en binôme

Date de remise : mardi 14 février à minuit

Fichiers fournis: Vous trouverez à l'adresse :

www.labri.fr/perso/bourqui/downloads/cours/Master/2016/Projet/

deux fichiers contenant les réseaux métaboliques de Buchnera Aphidicola et d'Escherichia coli K12 qui vous servirons dans ce projet.

Partie 1 : Pré-traitement & première visualisation (/5 points)

Comme vous pouvez le voir après le chargement de l'un des réseaux fournis, le graphe tel qu'il est affiché n'apporte que peu d'information. Écrire le code python permettant d'affecter

- des étiquettes aux sommets du réseaux (le numéro EC seront utilisés pour les réactions)
- une taille (non-nulle) à chaque sommet afin de pouvoir visualiser correctement les étiquettes
- des couleurs différentes aux métabolites et réactions (on distinguera aussi les réaction ayant un numéro EC des autres)
- des positions aux sommets du graphe (le choix de l'algorithme de dessin est libre)

Partie 2 : Extraction des voies & des structures topologiques (/3 points)

Question 2.1: Écrire un algorithme permettant d'extraire un sous-graphe par voie métabolique.

On considère dans un premier temps que les réactions du réseau ne sont pas réversibles (i.e. l'orientation des arcs est prise en compte)

Question 2.2 : Écrire un algorithme permettant de détecter et d'extraire (sous-graphe) le(s) plus long(s) cycle(s) de réactions d'une voie métabolique.

Si l'on considère maintenant que certaines réactions sont réversibles (i.e. seule l'orientation les arcs adjacents à une réaction non-réversible sera prise en compte).

Question 2.3 : Écrire l'algorithme permettant de détecter et d'extraire (sous-graphe) les cycles de réactions (**attention**, les cycles devront correspondre à des cycles « biologiques »).

Partie 3 : Dessin d'une voie métabolique (/4 points)

Écrire un algorithme permettant de dessiner une voie métabolique. L'approche considérée ici consiste à :

- 1. détecter/extraire les cycles de réactions
- 2. dessiner de manière circulaire chaque cycle
- 3. grouper chaque cycle (création d'un métanoeud)
- 4. dessiner le graphe quotient résultant
- 5. ouvrir les métanoeuds

Attention, lors de l'ajout d'un métanoeud dans un graphe, tous les graphes « parents » se voient aussi ajouter ce métanoeud (afin de respecter la notion de sous-graphe).

A l'issue de cette (ces) fonction(s), le graphe doit rester inchangé (les métanoeuds éventuellement ajoutés ou les sous-graphe éventuellement créés devront être supprimés).

Partie 4 : Dessin des voies métaboliques (/4 points)

Dans cette partie du projet, nous souhaitons produire une vue contenant un dessin de chaque voie métabolique (ensemble de vues locales sur le réseau).

Question 4.1: Écrire un algorithme permettant de dupliquer les réactions et métabolites appartenant à plusieurs voies métaboliques. Cet algorithme devra créer un nouveau sous-graphe contenant le « nouveau » réseau ainsi qu'un sous-graphe par voie métabolique.

Question 4.2 : Écrire un algorithme permettant de dessiner chacune des voies métaboliques (créées par la question 4.1) et de les afficher dans un seul dessin.

Partie 5: Pour aller plus loin (/4 points)

Comme vu en cours, la méthode décrite ci-dessus ne permet pas de comprendre les processus biologiques ayant lieu sur deux (ou plus) voies métaboliques puisque celles-ci sont représentées indépendamment les unes des autres.

Proposez un algorithme de dessin permettant de respecter la topologie initiale du réseau ou qui permette de faire un compromis entre la contrainte de proximité et la contrainte de duplication.

Expliquez vos choix et les avantages/inconvénients de votre méthode.

NB: Vous pouvez implémenter une méthode de l'état de l'art.

Description du format de graphe :

Dans les fichiers fournis, un certain nombre de propriétés existent (en plus des propriétés visuelles, commençant par « view* ») :

Nom	Туре	Description
Pathways	StringVectorProperty	Pour chaque sommet, contient la liste des voies métaboliques auxquelles il appartient
ecNumber	StringProperty	Spécifie pour chaque réaction son <i>EC number</i> (Enzyme Commission number)
name	StringProperty	Spécifie pour chaque sommet un nom.
reaction	BooleanProperty	Spécifie pour chaque sommet s'il est une réaction ou non
reversible	BooleanProperty	Spécifie pour chaque réaction si elle est réversible ou non
listOfReactants	IntegerVectorProperty	Spécifie pour chaque réaction la liste de ses réactants sous la forme d'un vecteur d'identifiants
listOfProducts	IntegerVectorProperty	Spécifie pour chaque réaction la liste de ses produits sous la forme d'un vecteur d'identifiants