

Projet AGGP - Dossier d'init

Anthony Tschirhard, Marie Paturel, Marion Poirel, Balthazar Rouberol

1 Introduction

2 Contexte et rappel du problème

2.1 Contexte

L'étude des systèmes biologiques complexes par modélisation sous forme de réseau est un moyen efficace de comprendre le fonctionnement intrinsèque à ces systèmes. En effet, il est possible de relier la structure du système aux fonctions de ses composantes. Par exemple, en étudiant la structure d'un réseau social, on peut comprendre comment l'information se diffuse ainsi que le rôle tenu par les utilisateurs dans son relai.

Les réseaux biologiques sont caractérisés par une architecture "sans-échelle", dans laquelle certaines composantes jouent un rôle plus important que d'autres : les hubs. On citera les exemples de Google, Facebook et Twitter sur le réseau Internet ainsi que l'ATP dans le réseau métabolique de la cellule. Cette architecture implique une distribution des degrés des noeuds selon une loi de puissance de paramètre γ , compris entre 2 et 3 pour la grande majorité des réseaux biologiques ($\gamma_{moy} \simeq 2.1$). Une autre caractéristique d'un tel réseau est la formation de cliques au sein de sa structure.

2.2 Objectifs

Le but de ce projet est de générer un réseau biologique en utilisant un algorithme génétique, développé en scéance d'Optimisation. Il faudra donc pour cela décider d'une fonction de fitness adaptée à ce type de réseau, afin que les réseaux générés se rapprochent de génération en génération d'un réseau biologique "idéal".

Les propriétés d'architecture "sans-échelle" et de modularité du réseau sont relativement contradictoires : la présence de hubs rend relativement improbable l'isolation de certains noeuds du réseau, ce qui est pourtant suggéré par les cliques dans contenues dans la structure. Il faudra donc construire la fonction de fitness autour d'un compromis acceptable entre ces deux conditions.

3 Documents de référence

Les connaissances générales sur la structure, l'architecture et le comportement des réseaux biologiques nous viennent de deux articles :

- "Exploring complex networks", *Steven H. Strogatz*, Nature, Vol. 410, March 2001
- "Network biology : understanding the cell's functional organization", *Albert-Lázlo Barabási, Zoltán N. Oltvai*, Nature reviews, Genetics, Volume 5, February 2001

L'algorithme génétique est celui contenu dans le fichier correction `pyAG.py`, fourni par M. Hédi Soula.

4 Contraintes générales

4.1 Contraintes

4.2 Risques

5 Organisation du travail

5.1 Rôles distribués

5.2 Règles de suivi

5.3 Organigramme des tâches

5.4 Outils utilisés

Afin de gérer au mieux ce projet, il convenait de mettre toutes les chances de notre côté en utilisant des outils adaptés à ce genre de travaux. Deux outils nous paraissaient particulièrement nécessaires :

- Un outil permettant de partager du code, à la manière de *subversion*, afin que nous puissions tous travailler en collaboration de manière efficace ;
- un outil de gestion de projet, afin que nous puissions tous définir précisément nos objectifs, nos contraintes temporelles et avancés.

Git

Pour le logiciel de gestion de versions, nous avons décidé d'utiliser **Git**, un logiciel libre simple et efficace, dont la principale tâche est de gérer l'évolution du contenu d'une arborescence. Comme tous les logiciel de gestion de versions, Git permet de créer de nouveaux dépôts, de faire des *commit*, de pousser, de récupérer les données distantes et de les fusionner avec celles de nos machines.

Pour héberger ce projet, nous avons décider d'utiliser le site de stockage de projet **GitHub**. Ce service propose l'hébergement de projets sous Git, mais dispose également de fonctionnalités de type réseaux sociaux – suivi de personnes ou de projets, graphes de réseau pour les dépôts etc. Après avoir ouvert des comptes à nos noms et avoir

6 Conclusion