

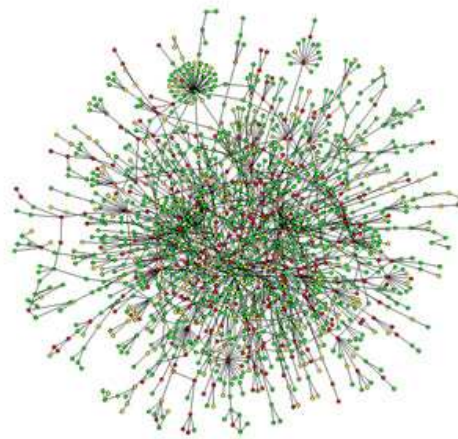
Sujet de stage M1 – 2009-2010

Détection de communautés dans les réseaux d'interactions protéine-protéine

Fabien Tarissan et Jean-Loup Guillaume

Prenom.Nom@lip6.fr

<http://www.complexnetworks.fr>



Contexte. La biologie moléculaire décrit aujourd'hui avec précision les composants des réseaux qui forment l'architecture des systèmes intracellulaires. La question se pose d'en déterminer la structure et de développer des outils capables d'organiser l'information contenue dans ces réseaux afin d'en faire ressortir des propriétés pertinentes. Cette question devient cruciale dans le cas des données recueillies au niveau moléculaire (ADN, ARN, protéines) car la masse d'information disponible rend impossible une exploration exhaustive de ces systèmes sans l'aide d'outils automatisés.

Un ensemble de techniques venant de la théorie des graphes permet d'ores et déjà de représenter de tels systèmes, de modéliser leur structure et d'analyser leur comportement en faisant ressortir des propriétés particulières. Il est par exemple tout à fait naturel de représenter un réseau d'interactions entre protéines par un graphe, dans lequel un nœud est associé à une protéine et un arc entre deux nœuds du graphe indique l'existence d'une interaction entre les deux protéines correspondantes.

Une des approches visant à l'analyse d'un tel réseau consiste à former des groupes de nœuds fortement connectés entre eux et très peu liés avec les nœuds des autres groupes. De tels groupes sont appelés *communautés*. La mise au point d'algorithmes pour détecter ces communautés est un sujet en plein essor et l'un d'eux a été mis au point par l'équipe COMPLEXNETWORKS. Cet algorithme consiste à créer dans un premier temps de petites communautés très denses puis à fusionner ces petites communautés de manière itérative jusqu'à l'obtention d'une structure stable. Cette technique fournit donc de plus une hiérarchie de communautés.

Objectifs du stage. Le but du stage est d'étudier le comportement de cet algorithme de détection de communautés sur des réseaux d'interactions protéine-protéine. Pour cela, plusieurs sujets peuvent être choisis :.

Sujet 1 (1 étudiant) : Tout d'abord, l'algorithme étant non déterministe, un premier travail consisterait à regarder les différents regroupements trouvés lors d'exécutions successives sur des réseaux connus (on pense notamment aux réseaux d'interactions disponibles sur la base de données MINT¹) et de confronter ces résultats avec la connaissance, en terme fonctionnel, qu'on a des éléments de ces réseaux.

Sujet 2 (1 étudiant) : Une seconde perspective de recherche serait d'étudier les résultats de ce même algorithme sur les « graphes de liens » de ces réseaux, c'est à dire des graphes dans lesquels les nœuds identifient les liens entre les protéines. Cela conduirait à une vision complémentaire de la précédente dans laquelle les protéines pourraient être associées à plusieurs communautés.

Sujet 3 (2 étudiants) : Enfin, une troisième approche consisterait à s'intéresser plus spécifiquement à la hiérarchie générée par l'algorithme, c'est-à-dire à s'intéresser aux communautés, sous-communautés, sous-sous-communautés, etc. . . et à regarder leur signification, en terme biologique. Là encore, la recherche de structures stables et leur pertinence biologique seront mises en avant.

¹<http://mint.bio.uniroma2.it/mint/Welcome.do>