

Un outil pour la visualisation de relations entre gènes

Marie Agier*, Jean-Marc Petit**

*LIMOS, UMR 6158 CNRS, Univ. Clermont-Ferrand II

**LIRIS, UMR 5205 CNRS, INSA Lyon

agier@isima.fr, jean-marc.petit@insa-lyon.fr

1 Introduction

La **reconstruction de réseaux de gènes** à partir de données d'expression a suscité depuis quelques années un intérêt croissant. L'objectif est de proposer des interactions probables entre les gènes, qui pourront être ensuite plus profondément validées avec des expérimentations plus poussées. Plusieurs approches ont été proposées pour inférer ces réseaux de gènes comme les réseaux booléens ou l'analyse des corrélations. La signification de la relation entre les gènes est très différente d'une méthode à l'autre et la comparaison entre les différentes méthodes existantes est difficile. Dans des travaux antérieurs [2], nous avons proposé une approche ayant pour but de découvrir **différents types de règles entre attributs**. L'originalité de notre travail réside dans le fait que nous proposons aux utilisateurs une approche globale pouvant inclure plusieurs sémantiques pour les règles, adaptées à plusieurs types de données. Diverses sémantiques spécialement adaptées aux données d'expression de gènes ont été présentées dans [1]. Pour faciliter l'interprétation des règles par les experts, nous souhaitons proposer une visualisation conviviale des règles générées.

2 Approche proposée

Nous souhaitons avant tout réaliser un outil convivial et proposer ainsi une méthode de visualisation intuitive pour les experts. D'autre part, nous proposons d'appliquer un filtre sur les règles générées en fonction de cinq indices de qualité (support, confiance, lift, leverage et conviction). Ne seront donc visualisées que les règles les plus intéressantes pour les experts, il est donc suffisant de pouvoir visualiser les indices pour la règle ou l'attribut sélectionnés par un simple clic. L'interprétation des règles est une étape particulièrement délicate et très difficile, puisqu'une règle entre deux gènes implique également divers produits associés (protéines, métabolites...). C'est pourquoi les biologistes sont rarement intéressés par les règles avec plus de 3 ou 4 attributs en partie gauche, l'interprétation devenant vite très difficile. Enfin, nous souhaitons offrir la possibilité aux utilisateurs de choisir avant la génération des règles, quelques gènes d'intérêts que nous appelons **gènes centraux**, qu'ils souhaitent voir apparaître en partie gauches ou droites des règles. Les biologistes pourront ainsi visualiser l'ensemble des règles associées à ces quelques gènes qui les intéressent plus particulièrement.

Nous avons donc opté pour une visualisation des règles par graphe qui nous semble être la méthode la plus facile à interpréter et la plus intuitive pour les experts, notre objectif étant aussi de nous positionner dans le cadre de la reconstruction de réseaux de gènes. L'originalité de notre approche réside dans le fait que nous proposons des graphes avec **plusieurs sé-**

Un outil pour la visualisation de relations entre gènes

mantiques, que nous appelons **réseaux globaux**. Chaque sémantique est caractérisée par une couleur particulière. De plus, nous avons choisi de représenter les différents gènes composant les parties droites et gauches des règles, sous forme d'un cercle pour une meilleure lisibilité. Enfin, les gènes centraux sont colorés en rouge, permettant ainsi de les repérer plus facilement.

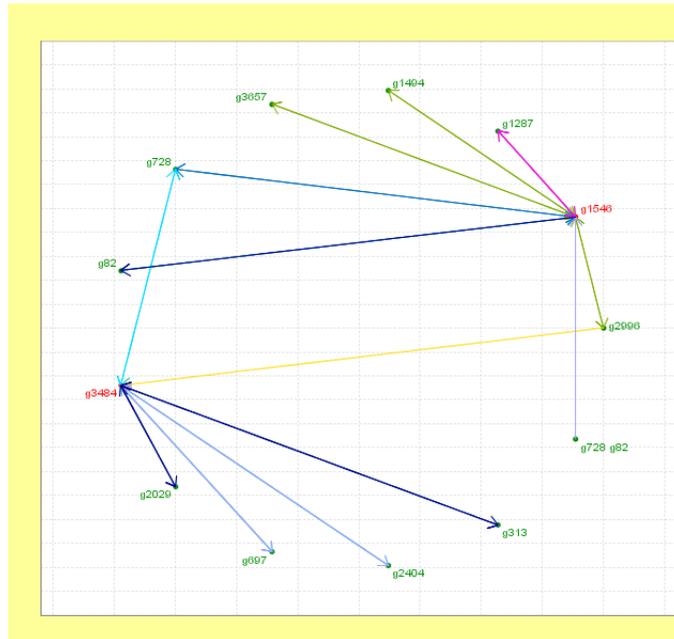


FIG. 1 – Réseau global avec plusieurs sémantiques et deux gènes centraux.

L'approche proposée a été développée spécifiquement pour les données d'expression de gènes. Un nouveau module appelé RG (Rule Generation) a été intégré à un logiciel existant consacré à l'analyse de données d'expression de gènes. Les points-clés de ce module sont tout d'abord une interface conviviale permettant de choisir parmi **plusieurs sémantiques** et de spécifier les **gènes centraux**, le calcul de **5 indices de qualité** pour chaque règle générée, la visualisation de **réseaux globaux** incluant diverses sémantiques, enfin la possibilité de visualiser sous forme textuelle les règles impliquant un gène en particulier avec les différents indices de qualité associés. Le module RG est disponible sur le site : <http://www.isima.fr/agier/GeneRules>.

Références

- [1] M. Agier, J.-M. Petit, V. Chabaud, C. Pradeyrol, Y.-J. Bignon, and V. Vidal. Vers différents types de règles pour les données d'expression de gènes-Application à des données de tumeurs mammaires. In *INFORSID*, pages 351–367, Biarritz, France, 2004.
- [2] M. Agier, J.-M. Petit, and E. Suzuki. Towards ad-hoc rule semantics for gene expression data. In *Proc. of the ISMIS*, pages 494–503, New-York, USA, 2005. Springer-Verlag.