

# Une métrique de sélection de variables appliquée à la centralité et à la détection des rôles communautaires

Nicolas Dugué\*, Jean-Charles Lamirel\*\*

\* LIUM - University du Maine, 72000, Le Mans, France

\*\* LORIA - SYNALP, 54000, Nancy, France

**Résumé.** La F-Mesure de trait est une métrique de sélection de variables statistique sans paramètres qui a montré de bonnes performances pour la classification, l'étiquetage de clusters ou encore la mesure de qualité des clusters. Dans cet article, nous proposons d'évaluer son utilisation dans le contexte des graphes de terrain et de leur structure communautaire pour bénéficier de son système sans paramètres et de ses performances bien évaluées. Nous étudions donc sur des graphes synthétiques réalistes les corrélations qui existent entre la F-Mesure de trait et certaines mesures de centralité, mais surtout avec des mesures destinées à caractériser le rôle communautaire des nœuds. Nous montrons ainsi que cette mesure est liée à la centralité des nœuds du réseau, et qu'elle est particulièrement adaptée à la mesure de leur connectivité au regard de la structure de communautés. Nous observons par ailleurs que les mesures usuelles de détection des rôles communautaires sont fortement dépendantes de la taille des communautés alors que celles que nous proposons sont par définition liées à la densité de la communauté, ce qui rend les résultats comparables d'un réseau à un autre. Ceci offre donc la possibilité d'applications comme le suivi temporel de la structure des communautés. Enfin, le processus de sélection appliqué aux nœuds permet de disposer d'un système universel, contrairement aux seuils fixés auparavant empiriquement pour l'établissement des rôles communautaires.

## 1 Introduction

De nombreux systèmes du réel sont modélisés et étudiés sous forme de réseau. L'université de Koblenz (Kunegis (2013)) recense et compile ainsi par exemple des réseaux sociaux, biologiques, routiers, lexicaux, de collaboration, etc. Ces réseaux fournissent une source importante de données pour étudier les systèmes qu'ils modélisent. Pour mieux exploiter ces données, les chercheurs d'un grand nombre de domaines (physiciens, sociologues, informaticiens, etc) ont développé des outils théoriques capables d'explorer et de fouiller ces données structurées en réseaux (Newman (2003)). Par exemple, la notion de *centralité* est étudiée depuis les années cinquante et l'impact de la centralité d'un individu dans une organisation a été discuté entre autres dans le cadre de la perception du leadership, de l'efficacité organisationnelle, de la diffusion de l'innovation technologique (Freeman (1979)). Cette notion reste très utilisée pour l'étude des nœuds d'un réseau, de leur position, de leur importance ou de leur influence, nous en rappelons donc les définitions essentielles.