

Clustering multi-niveaux de graphes : hiérarchique et topologique

Nhat-Quang Doan*, Hanane Azzag*, Mustapha Lebbah *

* Université Paris 13, LIPN UMR 7030
99, avenue Jean-Baptiste Clément
93430 Villetaneuse, France
hanane.azzag, lebbah@lipn.univ-paris13.fr

1 Introduction

Dans ce travail, nous souhaitons fournir une décomposition multi-niveaux du graphe d'origine : hiérarchique et topologique. Les modèles topologiques sont souvent utilisés pour la visualisation et la classification non supervisée. Des extensions et des reformulations du modèle SOM ont été décrites dans Bishop et al. (1998) Barbara Hammer (2009). Dans la littérature il existe peu d'algorithmes offrant la possibilité d'avoir en une seule passe une classification multi-niveaux. Les seuls travaux similaires sont ceux de SOM hiérarchique de Dittenbach et al. (2001) où les auteurs présentent une organisation multi-niveaux sur des référents de la carte en proposant ainsi plusieurs niveaux de cartes.

L'idée principale de notre modèle est de reconstruire le graphe d'origine $G(V, E)$ en le décomposant de manière hiérarchique en plusieurs sous-arbres auto-organisés, formant ainsi une forêt d'arbres projetée sur une carte 2D (voir figure ci-dessous). En utilisant les modèles topologiques, un graphe G représenté par X Luxburg (2007) peut être regroupé et visualisé dans une grille régulière en 2D ou 1D (de taille K) en utilisant le processus d'auto-organisation. La grille \mathcal{C} représente une forêt d'arbres organisés sur la grille en K sous-graphes représentés en arbre. Chaque cellule c de la grille est appelée par la suite «support (root)» d'un arbre noté $tree_c$ et chaque nœud de l'arbre représente un nœud $v_i \in V$. Les arbres sont construits en utilisant les principes d'un algorithme de classification non supervisé hiérarchique Azzag et al. (2006). Ainsi, nous associons à chaque $tree_c$ un prototype noté $leader(c)$. Ainsi, la fonction objective de l'auto-organisation des arbres s'écrit comme suit :

$$\mathcal{R}(\phi, \mathcal{L}) = \sum_{c=1}^k \sum_{r=1}^k \sum_{i \in tree_c} \mathcal{K}(\delta(\phi(\mathbf{x}_i), r)) \|\mathbf{x}_i - \mathbf{x}_{leader(r)}\|^2 \quad (1)$$

où $\mathcal{L} = \cup_{r=1}^k leader(r)$ et ϕ est la fonction d'affectation d'un sous-arbre (d'un groupe de nœuds à la fois).

Dans ces travaux nous avons présenté une nouvelle approche pour le clustering de graphes. Nous avons introduit deux nouvelles notions importantes : la notion de leader en tant que prototype et l'utilisation de la structure d'arbre pour définir une fonction d'affectation d'un ensemble