

Panorama de quelques approches récentes pour la classification non supervisée de graphes

Pascale Kuntz, Fabien Picarougne

Equipe COD - Laboratoire d'Informatique de Nantes Atlantique
Ecole Polytechnique de l'Université de Nantes
rue Christian Pauc, BP 50609
44306 Nantes Cedex3

{pascale.kuntz ; fabien.picarougne}@univ-nantes.fr,
<http://www.polytech.univ-nantes.fr/COD/>

Résumé. Les avancées technologiques récentes ont permis d'acquérir dans de nombreux domaines des corpus de graphes. Une problématique en plein essor consiste à classer ces données complexes pour établir des typologies. Différentes approches développées en fouille de données sont présentées dans cet article: la visualisation de graphes dans une perspective exploratoire, la caractérisation des graphes par des descripteurs structurels et fonctionnels, par des sous-structures et par des décompositions spectrales, et les méthodes à noyaux.

1 Introduction

Pour les pionniers de l'utilisation intensive de l'informatique dans l'analyse des données il y a une trentaine d'années, la tâche ne manquait pas ; que ce soit en Sciences Humaines ou en Sciences de la Vie par exemple, les données s'offraient à l'analyse sous un angle différent avec des technologies prometteuses. Au début des années quatre-vingt, I.C. Lerman notait : « on peut à présent prétendre dégager, derrière une grande masse d'informations, des structures d'organisation entre ces dernières, et cela au grand bonheur d'investigation dans les nombreuses disciplines (...) où la part due à l'observation et donc à la collecte des données est si importante » Lerman (1981). Si les corpus examinés à cette époque avaient été savamment, et souvent laborieusement, constitués par des experts de divers champs disciplinaires, les technologies d'aujourd'hui permettent de recueillir non seulement des volumes de données très importants sur des échelles de temps de plus en plus réduites, mais également des données de plus en plus complexes qui ne se réduisent plus simplement à des tableaux *Individus* \times *Variables*.

Les graphes, qui modélisent des systèmes de relations, représentent une illustration paradigmatique de l'accessibilité de ces nouvelles données. A titre d'exemple, T. Abello et al.

Cet article est la transcription d'une communication invitée présentée aux 3^{èmes} journées thématiques Apprentissage Artificielle et Fouille de Données. Il n'est donc pas un état de l'art exhaustif mais une introduction à une problématique en plein essor dans de nombreux domaines.