

Le partitionnement de graphes

Application à la construction et à la caractérisation d'un réseau de villes

Auteurs :

Pascal Eusebio

Psar Analyse territoriales

Jean-Michel Floch

Département de l'action régionale

David Levy

Psar Analyses territoriales

Résumé

Ces dix dernières années, les méthodes appliquées aux réseaux ont connu un rapide essor. Soutenue par une montée en puissance des calculateurs, la théorie des graphes est appliquée dans tous les domaines scientifiques (mathématique, informatique ou biologie), dans le domaine sociologique (les réseaux sociaux est sans doute l'exemple plus célèbre) ou encore en géographie avec l'étude des réseaux de villes. Dans ce cas, les villes constituent les nœuds du réseau et les flux (domicile-travail, migrations résidentielles ou relation établissement siège) les arêtes.

La première partie de la communication sera consacrée à la présentation des concepts et des méthodes qui servent à partitionner les graphes. (graph clustering ou community detection dans la littérature. Ce domaine est actuellement en pleine expansion et de nombreux algorithmes sont actuellement proposés. On introduira notamment le concept de modularité, introduit par Girvan et Newman, qui reste au centre de nombreuses analyses. La modularité sert à mesurer la qualité du partitionnement, en faisant la différence entre la somme des flux internes à une communauté et la somme des flux d'un graphe aléatoire ayant le même nombre de sommets et la même distribution des degrés. On présentera aussi un bref panorama des méthodes qui permettent le partitionnement : méthodes hiérarchiques agrégatives ou divisives, méthodes de type k-means, méthodes spectrales. Quelques-uns des principaux algorithmes permettant de les implémenter seront mentionnés.

Les méthodes de partitionnement de graphes ont été employées pour regrouper des villes⁽¹⁾ tenant compte de l'ensemble des liens entre les villes regroupées mais aussi du caractère privilégié de ces liens. Elles évitent ainsi les écueils des regroupements itératifs fondés uniquement sur les flux majeurs entre territoires. On identifie donc ; à partir de l'analyse des ensembles de villes (sous-réseaux, communautés..) qui fonctionnent de façon privilégiée. A l'intérieur de ces communautés, les relations sont particulièrement fortes, tandis qu'elles sont limitées entre les communautés. Dans une approche agglomérative, on fusionne de manière récursive de petites communautés en de plus grandes - avec comme point de départ les villes- en effectuant les choix qui maximisent la « modularité ».

Ce partitionnement des villes est accompagné d'indicateurs permettant de caractériser les communautés mises en évidence. A côté d'indicateurs statistiques classiques, on en a construit d'autres issus de l'analyse des réseaux (modularité, degré de centralisation, « pagerank », rang d'autorité et de « hub ».

Ces méthodes ont permis de répondre à des demandes nationales (CGET,ex DATAR) et de produire des analyses régionales, qui fourniront la matière des exemples présentés dans la communication. Les mailles géographiques sont modulables et plusieurs types de flux peuvent être mobilisés.

Les prochaines étapes des travaux concernent la façon de combiner les analyses issues des différents flux, et l'examen des nouveaux algorithmes implémentés dans les logiciels.

(1) Communes, aires urbaines ...