Espace intrinsèque d'un graphe et recherche de communautés.

Alain Lelu*,†,‡ — Martine Cadot*,††

* LORIA, Nancy - France

† Université de Franche-Comté/LASELDI, Besançon - France
Alain.Lelu@univ-fcomte.fr

†† Université de Nancy/Département Informatique, Nancy - France
Martine.Cadot@loria.fr

‡ Institut des Sciences de la Communication du CNRS, Paris - France

Abstract

Determining the number of relevant dimensions in the eigen-space of a graph Laplacian matrix is a central issue in many spectral graph-mining applications. We tackle here the problem of finding out the "right" dimensionality of Laplacian matrices, especially those often encountered in the domains of social or biological graphs: the ones underlying large, sparse, unoriented and unweighted graphs, often endowed with a power-law degree distribution. We present here the application of a randomization test to this problem. After a small introductive example, we validate our approach first on an artificial sparse and scale-free graph, with two intermingled clusters, then on two real-world social graphs ("Football-league", "Mexican Politician Network"), where the actual, intrinsic dimensions appear to be 10 and 2 respectively ; we illustrate the optimality of the transformed dataspaces both visually and numerically, by means of a density-based clustering technique and a decision tree.

Key-words: graph, graph Laplacian, dimensionality reduction, intrinsic dimension, randomization test, dominant eigen-subspace, graph clustering, density clustering method, scale-free graph, Cattell's scree.