



HAL
open science

Coarsened Spectral Clustering

Pierre Borgnat, Rémi Vaudaine, Etienne Lassale, Paulo Gonçalves, Rémi Gribonval, Marton Karsai

► **To cite this version:**

Pierre Borgnat, Rémi Vaudaine, Etienne Lassale, Paulo Gonçalves, Rémi Gribonval, et al.. Coarsened Spectral Clustering. NetSci 2024, Jun 2024, Québec (CA), Canada. hal-04732048

HAL Id: hal-04732048

<https://hal.science/hal-04732048v1>

Submitted on 11 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Coarsened Spectral Clustering

R. Vaudaine¹, E. Lasalle¹, P. Borgnat², P. Gonçalves¹, R. Gribonval¹, and M. Karsai^{3,4}

1. Univ Lyon, EnsL, UCBL, CNRS, Inria, LIP, F-69342, LYON, France
2. CNRS, Univ de Lyon, ENS de Lyon, Laboratoire de Physique, F-69342 Lyon, France
3. Department of Network and Data Science, Central European University, 1100 Wien, Austria
4. Rényi Institute of Mathematics, 1053 Budapest, Hungary

Spectral clustering is a theoretically and practically well-studied method to partition a graph into k communities [1]. It is based on a spectral decomposition of the (normalized) Laplacian of a graph. The main steps are the computation of the Laplacian, the projection on the k -dimensional eigenspace relative to the k smallest eigenvalues and the clustering, usually a k -means. However, for large graphs, the computation of some steps may be prohibitive and reduction techniques are necessary. Thus, graph coarsening consists in the reduction of the size of a graph by aggregation of some nodes. While structured graph coarsening for spectral clustering has been well studied [2], unstructured graph coarsening is only at its beginnings [3].

Let us consider a graph with n nodes and m edges. Standard spectral clustering requires $O(m)$ time and memory to compute the Laplacian matrix, $O(kn^2)$ time and $O(m)$ memory to compute the spectral decomposition of the k smallest eigenvalues and $O(kn)$ average time and $O(n)$ space to compute the clustering. The overall memory complexity is low and is, usually, not an issue to compute spectral clustering. Here, we focus on reducing the time complexity of the algorithm. Thanks to structured coarsening or random hashing of the nodes, a technique of random unstructured graph coarsening, the number of nodes of the graph is reduced to $n_s \ll n$. But, this reduction introduces uncertainty in the graph that is lifted by using several different reductions of the same graph. The choice technique is based on the size of the graph. Random hashing is used for very large graphs since its computation does not depend on the number of nodes or edges. Structured coarsening is used for smaller graphs.

Thus, we propose, first, an acceleration of the current method for structured coarsening and, second, a framework to compute the spectral clustering of a graph using any graph coarsening (see Fig.1). The first step consists in computing, in parallel, several coarsened versions of the input graph. Then, every small graph will be partitioned using standard spectral clustering. Finally, we use a majority vote to uncover the partition of the input graph. The main advantage of our method is that the complexity of the spectral decomposition is lowered from $O(kn^2)$ to $O(kn_s^2)$. Results show that we accelerate the computation of structured coarsening and that the framework improves the clustering performances.

[1] A. Ng, M. Jordan and Y. Weiss, *On Spectral Clustering: Analysis and an algorithm*, NIPS, 2:849–856, 2002.
 [2] N. Tremblay and A. Loukas, *Approximating Spectral Clustering via Sampling: a Review*, Sampling Techniques for Supervised or Unsupervised Tasks, 129–183, 2020.
 [3] R. Vaudaine et al., *Temporal network compression via network hashing*, Applied Network Science, 9, 3, 2024.

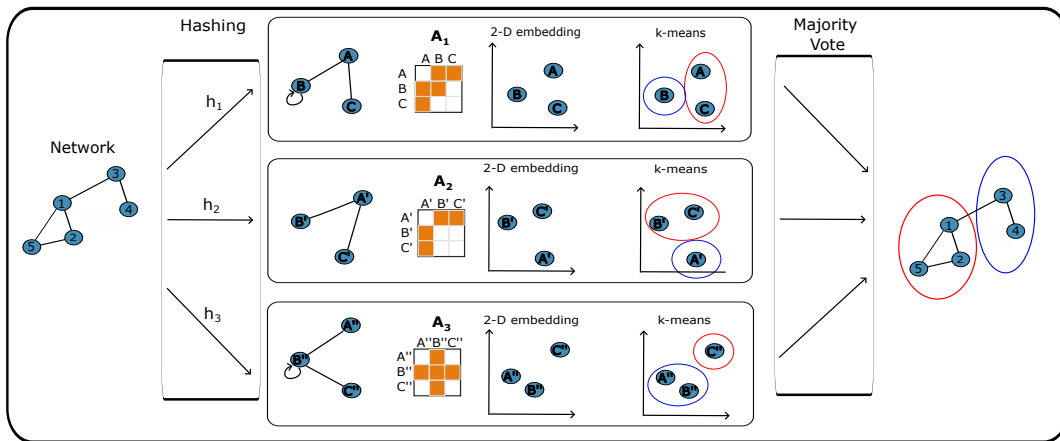


Figure 1: Left to right. A static graph. Several random hash functions are applied to create hashed graphs. Their adjacency matrices A_i can be extracted and used to compute their Laplacian. Then, a $2d$ eigen-projection is performed and a k -means too. Finally, the different partitions are aggregated to uncover the communities of the original graph. In the case of structured coarsening, the hashing functions are replaced by their structured equivalents.