

Comparison of linear modularization criteria using the relational formalism, an approach to easily identify resolution limit

Patricia Conde-Céspedes*, Jean-François Marcotorchino**, Emmanuel Viennet*,¹

*L2TI - Institut Galilée - Université Paris 13
99, av. Jean-Baptiste Clément; 93430 Villetaneuse - France
prenom.nom@univ-paris13.fr

**Thales Communications et Sécurité
4 av. des Louvresses; 92230 Gennevilliers - France
jeanfrancois.marcotorchino@thalesgroup.com

Résumé. La modularisation de grands graphes ou recherche de communautés est abordée comme l'optimisation d'un critère de qualité, l'un des plus utilisés étant la modularité de Newman-Girvan. D'autres critères, ayant d'autres propriétés, aboutissent à des solutions différentes. Dans cet article, nous présentons une réécriture relationnelle de six critères linéaires: Zahn-Condorcet, Owsinski-Zadrozny, l'Ecart à l'Uniformité, l'Ecart à l'Indétermination et la Modularité Equilibrée. Nous utilisons une version générique de l'algorithme d'optimisation de Louvain pour approcher la partition optimale pour chaque critère sur des réseaux réels de différentes tailles. Les partitions obtenues présentent des caractéristiques différentes, concernant notamment le nombre de classes. Le formalisme relationnel nous permet de justifier ces différences d'un point de vue théorique. En outre, cette notation permet d'identifier facilement les critères ayant une limite de résolution (phénomène qui empêche en pratique la détection de petites communautés sur de grands graphes). Une étude de la qualité des partitions trouvées dans les graphes synthétiques LFR permet de confirmer ces résultats.

1 Introduction

Networks are studied in numerous contexts such as biology, sociology, online social networks, marketing, etc. Graphs are mathematical representations of networks, where the entities are called nodes and the connections are called edges. Very large graphs are difficult to analyse and it is often beneficial to divide them in smaller homogeneous components easier to handle. The process of decomposing a network has received different names : graph clustering (in data analysis), modularization, community structure identification. The clusters can be called communities or modules ; in this paper we use those words as synonyms.

1. This work is supported by REQUEST project between Thales and Paris 13 University.