# The Noise Cluster Model, a Greedy Solution to the Network Community Extraction Problem

Etienne Côme*, Eustache Diemert†

* IFSTTAR - Bâtiment Descartes 2,
2, Rue de la Butte verte,
93166 Noisy le Grand Cedex, France
etienne.come@ifsttar.fr

† BestOfMedia Group,
485 avenue de l'Europe,
F-38330 Montbonnot, France
ediemert@bestofmedia.com

**Abstract**

This paper presents an algorithm designed to extract one community (a collection of vertices that are densely connected amongst themselves) from a graph given some seeds (nodes known to belong to the community). Starting from these seed nodes, new nodes will be added to the community by selecting them among the successors of the current community members. The process used to select the community members among the successors is based on a generative model closely related to Erdös-Rényi mixture [7] called the Noise Cluster Model. An on-line estimation procedure [23] is used to update the model parameters during the community extraction process. This approach is local, the complexity is mainly influenced by the community size and does not depend upon the graph size. This method can therefore be used to deal with huge graphs. Eventually, experiments on real blog communities will show the interest of such an approach.

**Key-words:** graph clustering, community extraction, semi-supervised, noise cluster model