

# Detección de comunidades en redes complejas

Rodrigo Aldecoa García

31 de mayo de 2013

Networks have become a widely used tool for modeling complex systems in many different fields. This approach is extremely useful for representing interactions among genes, social relationships, Internet communications or correlations of prices within a stock market, to name just a few examples. By analyzing the structure of these networks and understanding how their different elements interact, we could improve our knowledge of the whole system. Usually, nodes that compose these networks tend to create tightly knit groups. This property, of high interest in many scientific fields, is called *community structure* and improving its detection and characterization is what this thesis is all about. The first objective of this work is the generation of efficient methods able to characterize the communities of a network and to understand its structure. Second, we will try to create a set of tests where such methods can be studied. Finally, we will suggest a statistical measure in order to be able to properly assess the quality of the community structure of a network. To accomplish these objectives, first, we generate a set of algorithms that can transform a network into a hierarchical tree and, from there, to determine their most relevant communities. Furthermore, we have developed a new type of benchmarks for effectively testing these and other community detection algorithms. Finally, and as the most important contribution of this work, it is shown that the community structure of a network can be accurately evaluated using a hypergeometric distribution-based index. Thus, the maximization of this measure, called *Surprise*, appears as the best proposed strategy for detecting the optimal partition into communities of a network. *Surprise* exhibits an excellent behavior in all networks analyzed, qualitatively outperforming any previous method. Thus, it appears as the best measure proposed to this end and the data suggests that it could be an optimal strategy to determine the quality of the community structure of complex networks.