

Lab Session 13: Network Analysis II

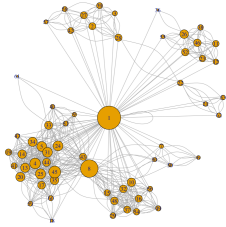
Jesús Antoñanzas, Alex Carrillo

Cerca i Anàlisi de la Informació, Ciència i Enginyeria de Dades (UPC)

December 24, 2019

When trying to create the graph from the edgelist, `igraph::graph_from_edgelist` seemed convinced that some vertex was negative. Updating the package did not fix it. Some digging revealed that no node can be named '0'. So, adding 1 to all nodes before creating the graph fixed the issue.

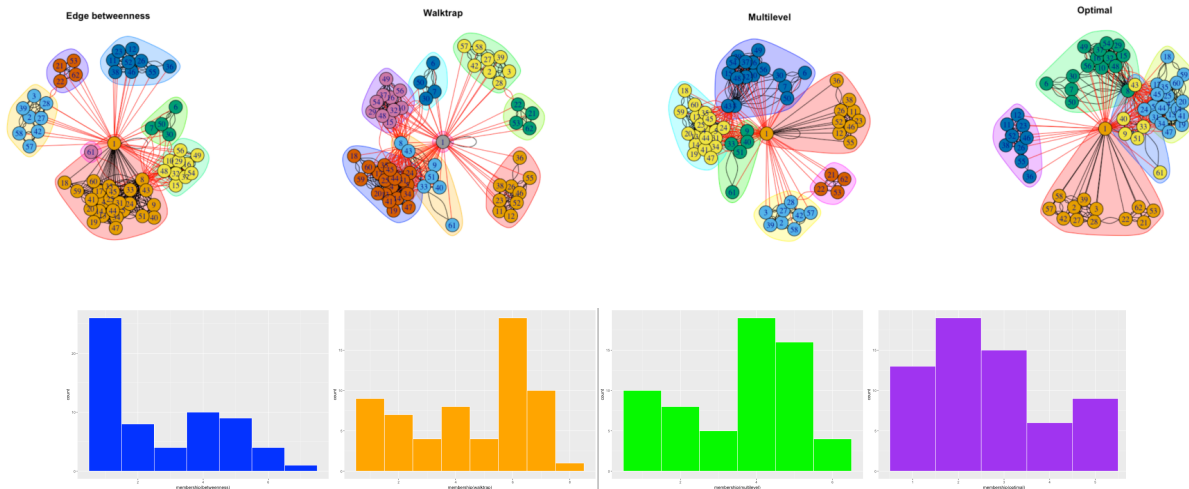
After reading the graph, we can examine it. It is of order 62 (number of vertices) and has size 602 (number of edges), as given by `igraph`'s `gorder` and `gsize` functions, respectively. The diameter, that is the "length of the longest geodesic" is 2, its transitivity (overall probability for the network to have the edges interconnected) is 0.523 and its degree distribution is distributed among 21 of 64 degrees, with degrees 7, 14 and 40 being higher than the others, which look somehow uniform. Because of this degree distribution we can guess with some certainty that the network is not a random *Barabási-Albert* network.



Moreover, plotting the graph with the size of the nodes being their pagerank score (adjusting the `vertex.size` parameter in the plot function) clearly reaffirms the previous guess (in order to do this we had to scale the pagerank scores by a factor of 400 as they were just too small for the `plot.igraph` function).

Now, we try some community detection algorithm of our choice from the list provided. We apply the two previous tested algorithms *edge betweenness* and *walktrap*, and we test two more: *multilevel* and *optimal*.

For each algorithm we found 7, 8, 6 and 5 communities, respectively. Those results indicate, in our opinion, that the first two algorithms are more aggressive in the number of clusters to choose, while the two last are kind of more conservative.



In the images above we show the graph's communities representations and the histograms of their 'memberships'. It should be stated that, in our test we did not notice a consistency in number of nodes that the largest community contains, although it contains in average around 20 nodes.