K-Nearest Neighbor (Java)

Description

The K-Nearest-Neighbor is a measure of the *correlation* between the degree of a node and that of its neighbors. In many systems strong correlations are observed, and one distinguishes networks into *assortative* and *disassortative*. A network is assortative if large (small) degree nodes tend to be linked with large (small) degree nodes. Social networks are typical examples of assortative networks. A network is disassortative if large (small) degree nodes tend to be linked with small (large) degree nodes. Biological and technological networks (like the Internet) are examples of disassortative networks.
The undirected K-Nearest-Neighbor is defined as follows: a node is selected and the average degree of all its neighbors is calculated. By repeating the procedure for all nodes of the network one derives a pair for each node, where is the degree of the node. By averaging over nodes with equal degree one derives the function , which allows to study the correlation.
If grows with , the network is assortative; if decreases with , the network is disassortative. A flat curve would indicate the absence of correlation.
In the absence of correlations, the function would be the same as for a random network with equal degree distribution . In this case it is possible to prove that

where is the second moment of the distribution (i.e. the expected value of the degree squared) and is the average degree of the network. For this reason, the function calculated by the algorithm is normalized in that we divide it by the constant .

Pros & Cons

The network to analyze must be undirected, otherwise there are no special constraints.

Applications

The algorithm is used to disclose affinities/diversities between neighboring nodes. Many properties of networks and of processes that take place on networks are affected by the presence of degree-degree correlations.

Implementation Details

The algorithm requires two inputs, the file where the edges of the network are listed and the number of points one wishes to have in the binned correlation function described below. A first read-in of the inputfile will set the values of the number of nodes and edges of the network. In the second read-in the degrees of all nodes will be calculated and the edges are stored in an array. Then the of all nodes are calculated.
The program generates two output files, corresponding to two different ways of partitioning the interval spanned by the values of degree. In the first output, the is averaged among all nodes with equal degree; the output displays all degree values with their average s (normalized as described in the Description section).
The second output gives the *binned* correlation function, i.e. the interval spanned by the values of degree is divided into bins whose size grows while going to higher values of the variable. The size of each bin is obtained by multiplying by a fixed number the size of the previous bin. The program calculates the average for all nodes whose degree falls within each bin.
This technique is particularly suitable to study skewed correlation functions: the fact that the size of the bins grows large for large degree values compensates for the fact that not many nodes have high degree values, so it suppresses the fluctuations that one would observe by using bins of equal size. On a double logarithmic scale the points of will appear equally spaced on the x-axis.
The program runs in a time O(m), m being the number of edges of the network.

Usage Hints

A simple application of this algorithm could be to calculate the for networks created by the modeling algorithms of the NWB. For instance, the inputfile can be created through the Barabasi-Albert model.

Links

- Source Code

Acknowledgements

The algorithm was implemented and documented by S. Fortunato, integrated by S. Fortunato and W. Huang.

References

Pastor-Satorras, R., Vazquez, A., Vespignani, A. (2001) Dynamical and Correlation Properties of the Internet.
Physical Review Letters 87:258701.

See Also

⊘ The license could not be verified: License Certificate has expired! Generate a Free license now.