

Tomasz Gubiec

Between a complete graph and its minimal spanning tree

The most commonly used indicator of dependencies between stock is the asset correlation. If we consider a set of n stocks we obtain $n(n-1)/2$ correlation coefficients, which may be considered as a graph with nodes representing stocks and edges with weight related to correlation. A natural question may arise: which of these edges are the most important? Simple thresholding is not satisfactory as the graph quickly becomes disconnected. Since Rosario Mantegna [1] applied a minimal spanning tree in analysis of correlation-based networks, many other ideas were analyzed (e.g. planar graphs). Here, I propose a new method of removing edges with more adequate criterion than simple thresholding. The main advantage of this method is that it will never result in a disconnected graph and as the final result of removing all allowed edges we obtain a minimal spanning tree. Graph obtained with this algorithm may be considered a generalization of the minimal spanning tree carrying more informations.

[1] Mantegna, Rosario N. "Hierarchical structure in financial markets." *The European Physical Journal B-Condensed Matter and Complex Systems* 11.1 (1999): 193-197.