**CT 1.3.1**

# Introduction to Network complexity

Jens Christian Claussen[1] - Thomas Wilhelm[2]

[1]Theoretical Physics and Astrophysics, University Kiel, Leibnizstr. 15, 24098 Kiel, Germany
[2] Institute of Food Research, Norwich, Norwich Research Park, NR4 7UA, UK

## Introduction

A vast variety of biological, social, and economical networks shows topologies drastically differing from random graphs; yet the quantitative characterization remains unsatisfactory from a conceptual point of view. Motivated from the discussion of small scale-free networks, in (Claussen 2007a) a biased link distribution entropy is defined, which takes an extremum for a power law distribution. This approach is extended to the node-node link cross-distribution, whose nondiagonal elements characterize the graph structure beyond link distribution, cluster coefficient and average path length. From here a simple (and computationally cheap) complexity measure can be defined. This Offdiagonal Complexity (OdC) is proposed as a novel measure to characterize the complexity of an undirected graph, or network (Claussen 2007a). While both for regular lattices and fully connected networks OdC is zero, it takes a moderately low value for a random graph and shows high values for apparently complex structures as scale-free networks and hierarchical trees. The Offdiagonal Complexity apporach is applied to the Helicobacter pylori protein interaction network and randomly rewired surrogates. In addition, OdC is used to characterize the spatial complexity of cell aggregates (Claussen 2007b).

## Complexity measures in biology

In biological sciences, the evolution of life is studied in detail and at large; and it is observed qualitatively that evolution creates, on average, organisms of increasing complexity. If one wants to quantify an increase of complexity, one has to define suitable complexity measures. In some sense, the number of cells may be an indicator, but quantifies rather body size than complexity. Instead one may observe the number of organelles, the size of the metabolic network, the behavioural complexity of social organisms, or similar properties. To have a time series of the complexity distribution of all organisms during evolution on earth, would be highly interesting for the test of models of evolution, speciation and extinctions. But apart from such academic questions, there are many areas of practical use of complexity measures in biology and medicine, as the complexity of morphological structures, cell aggregates, metabolic or genetic networks, or neural connectivities.

## Complexity measures

For text strings (as computer programs, or DNA) there are common complexity measures in theoretical computer science, such as Kolmogorov complexity (and Lempel-Ziv complexity) and algorithmic information content (AIC). For example, AIC is defined by the length of the shortest program generating the string. For random structures, thus also for random graphs, these measures indicate high complexity.
A distinction of complex structured (but still partly random) structures from completely random ones usually is prohibitive for this class of measures.

## Definition

Let g_{ij} be the adjacency matric of a graph with N nodes. For each node i of the graph, let l(i) be its node degree. Let c_{nm} be the number of edges between all pairs of nodes i and j with node degrees m=l(i), n=l(j), l(j) >= l(i) (ordered pairs). (This gives a degree correlation matrix with ordered pairs of nodes.) Then,
b_k = sum_{i=0}^{k_max -k}, A=sum_{k=0}{k_max}, a_k=b_k/A.
Then OdC is defined as: OdC = - sum_{k=0}^{k_max} a_k ln a_k.

## A Variant: FullOdC

From the same coefficients c_{ij} as above, it is also possible to define an entropy of all degree correlation indices, i.e., without summation over the offdiagonals. This has been proposed in (Claussen2007b) as
FOdC = - sum_{i=0}^{k_max} sum_{j=i}^{k_max} c_{ij} ln c_{ij}
and behaves similar to OdC. FOdC is less self-averaging than OdC.

## Results

For orthogonal lattices, OdC is zero. For a fully connected graph, OdC also vanishes. For a random graph OdC is nonzero, but takes small values compared with the maximal possible value. Real networks (as a protein interaction network and a cell adjacency network) show higher values than their randomly reshuffled counterparts.

## Discussion

It seems to remain a challenge to estimate complexity from a simple or local properties. The transfer of complexity approaches for strings is difficult as no unique mapping of a given graph to a string is established yet (and if it exists, it is probably of nonpolynomial computational complexity). OdC itself has the advantage that it requires only costs quadratic in network size.

## Relation to degree correlations

Obviously, OdC quantifies disorder in degree correlations. Therefore, perfectly degree-correlated nodes result in a vanishing OdC, as all entries are on the main diagonal. (Note that FullOdC needs not to vanish in this case.) If a graph has statistically no degree correlations (as obtained by the stubs model or configuration model) OdC assumes the value of a random graph with the same degree distribution. While a scalar degree correlation coefficient captures only a part of the degree correlation structure, the vectorial distribution a_k that underlies OdC contains additional information, and, while itself again being a scalar, provides additional information.

## Acknowledgement

Stimulating discussions with Yu. Pis'mak, K. Sneppen, S. Bornholdt, H. Horner, J. Kurths, P. Grassberger, J. Nagler and J. Reichardt are gratefully acknowledged.

## References

J. C. Claussen, Physica A 375, 365-373 (2007a)
J. C. Claussen, in: Mathematical Modeling of Biological Systems II. Ed. A.Deutsch et al., Birkhaeuser Boston 291-299 (2007b)