Preferential duplication of intermodular hub genes: an evolutionary signature in genome networks

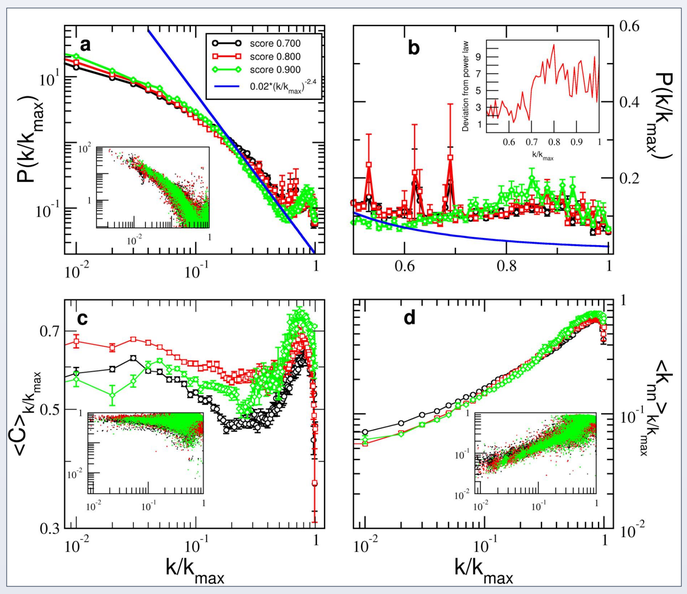
RMFerreira, JL Rybarczyk-Filho,RJS Dalmolin,MAA Castro,JCF Moreira,LG Brunnet, RMC de Almeida

Instituto de Fisica e Departamento de Bioquímica

UFRGS

Introduction

We considered all 268 core organisms in STRING database, version 8.3 with confidence scores 0.700, 0.800, and 0.900 using "experimental" and "database" (95% of these interactions) added with "neighborhood", "fusion", "co-expression", and "co-occurrence" evidence. This information renders possible to build a network, where each node corresponds to a protein with at least one known protein-protein association, and links correspond to these associations. To each network node we assign a degree , which is the number of links arriving at that node. For each organism and score we produce a network and calculate the probability P(k/kmax) that a protein has k/kmax links.



There is a local maximum in P(k/kmax), <C>, and <knn>.

Model

Simulations start with 5 nodes, each linked to two others, forming a ring. To acquire a new gene we first choose either *de novo* mechanism, with probability q , or duplication, with probability 1-q . If the *de novo* mechanism is chosen, each existing node is linked to the new one with probability proportinoal to its degree , and the procedure is repeated until the new node presents at least one link. In case of duplication, the node to be duplicated is chosen by using the probability defined as

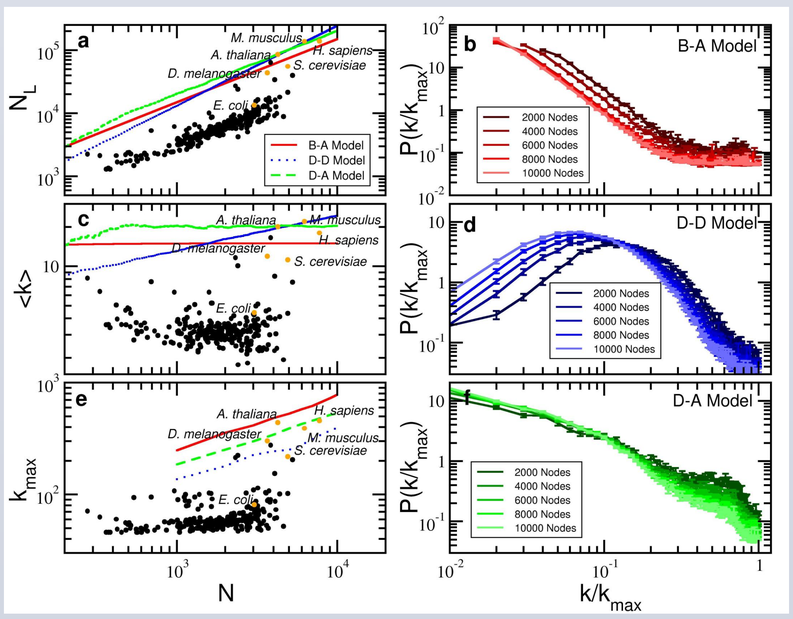
$$p_i^D = \frac{k_i(1-C_i)}{\sum_j k_j(1-C_j)}$$

with

$$C_i = \frac{2}{k_i(k_i-1)}\sum_{j=1}^{N}\sum_{l=1}^{N} M_{ij}M_{jl}M_{li}$$

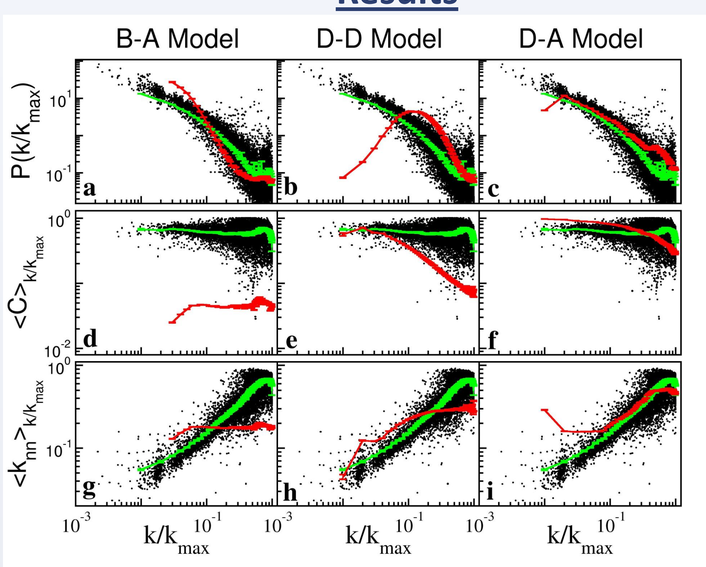
where Mij is the association matrix for genes i and j.

We compare with Barabasi-Albert and Duplication-Divergence models.



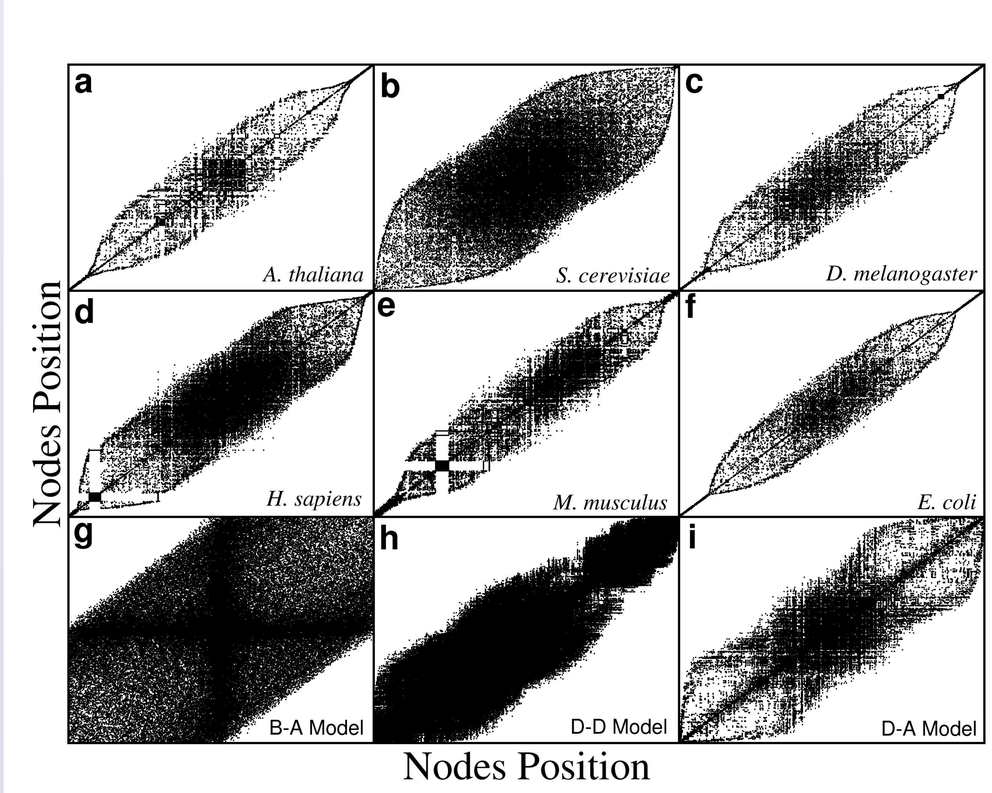
Evolution of simulated models. Barabási-Albert, duplication-divergence and duplication-acquisiton networks (red, blue and green lines, respectively). The black dots represent all core organisms from STRING database, where six well studied organisms are highlighted in orange. (**a**) Number of links, (**c**) mean degree and (**e**) maximum degree are shown as functions of the total number of nodes in the network. The degree distribution was calculated in five snapshots of the evolution of (**b**) Barabási-Albert, (**d**) duplication-divergence, and (**f**) duplication-acquisition models, in intervals of 2000 nodes.

Results



Comparison of topological measures for simulated networks. The black dots represent the superposed networks for all core organisms from string database with confidence score 0.800, the green lines are averages taken in intervals of , and the red lines are weighted averages of simulated networks. The upper, central, and lower rows show, respectively, degree distribution, clustering coefficient, and nearest neighbor mean degree. Each column refers to a simulated model: Barabási-Albert on the left, duplication-divergence on the center and duplication-acquisition on the right.

Topology



Ordered association matrices. This figure presents the association matrices for *Homo sapiens*, *Mus musculus*, *Arabidopsis thaliana*, *Drosophila melanogaster, Saccharomyces cerevisiae*, *Escherichia coli*, Barbási-Albert model, duplication-divergence model and duplication-acquisition model after running the ordering algorithm. The black dots represent interactions between two nodes.

References

Ferreira RM, Rybarczyk-Filho JL, Dalmolin RJS, Castro MAA, Moreira JCF, Brunnet LG, and deAlmeida RMC, Preferential duplication of intermodular hub genes: an evolutionary signature in genome networks, (2012) submitted

Barabasi AL, Albert R (1999) Emergence of scaling in random networks. Science 286: 509-512. 7898

Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL (2000) The large-scale organization of metabolic networks. Nature 407: 651-654. 10.1038/35036627

Vázquez A, Flammini A, Maritan A, Vespignani A (2003) Modeling of Protein Interaction Networks. Complexus 1: 38-44.

Vázquez A (2003) Growing network with local rules: preferential attachment, clustering hierarchy, and degree correlations. Phys Rev E Stat Nonlin Soft Matter Phys 67: 056104.

Dalmolin R, Castro M, Rybarczyk Filho J, Souza L, de Almeida R, Moreira J (2011) Evolutionary plasticity determination by orthologous groups distribution. Biology Direct 6: 22.