

Complexity and Fitness of Methylated DNA sites and RNAs in the context of Breast Cancer Tissues

January 19, 2023

1 Aim of the study

Highlighting the nestedness structure of the bipartite network representing correlations between methylated DNA sites and RNAs.

2 Methods and Data

2.1 Nestedness

Nested structures appear in the nature such that in ecological and socio-economic systems. In ecological systems, for instance, there is a nested distribution of species habitats. Where as some species only live in particular locations, some others live in a large variety of environments including those particular locations [CITATION]. In the context of network theory, a perfectly nested structure is such that for any pair of node (i, j) with $k_i < k_j$ their respective degree (number of neighbors), the set of nodes i is interacting with is included in the set of nodes j is interacting with. The adjacency matrix of such a network shows a triangular shape when columns and rows are sorted accordingly to the degree ordering of respective nodes. There are many metrics (other than degrees) permitting to infer the nested structure of a given network. Nestedness can appear in unipartite and bipartite networks. In the context of economical systems, especially the world trade network seen as a country-product bipartite network, recent research shows the existence of nestedness [CITATION]. The metric used here is the fitness-complexity metric [DEFINITION] which is based on a non-linear iterative algorithm. Once rows (products) and columns (countries) are sorted in a decreasing ordering of complexity and fitness scores, the biadjacency matrix representing the products exported by the countries shows a triangular shape that is characteristic of nested structure.

2.2 Data

We used an open access cancer dataset from GDC Data Portal. From raw data consisting in methylated DNA and RNA's beta-values for up to 841 tumorous and normal breast tissues, we built a Matrix of Pearson Correlation Coefficients between Methylated DNA and RNA from these beta-values. A network representing the Pearson correlation coefficients between pairs (M-DNA, RNA) consists in a bipartite graph with coefficients as weight of the links. Two possibilities are considered either the correlation is positive or negative. Therefore we consider a signed bipartite graph with undirected links. Two other correlation matrices can be constructed from these beta-values, the DNA-DNA and RNA-RNA correlation matrices.

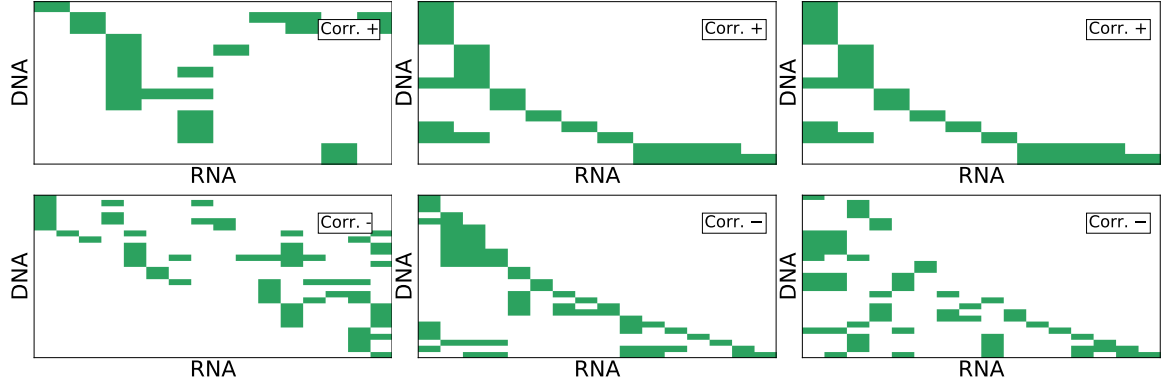


Figure 1: Binary bi-adjacency matrix representing the interactions between DNAs and RNAs, in the context of 841 breast tumorous tissues (left panels). Matrix entries equal 1 (green cells) if there is an interaction between entities, 0 otherwise (white cells). Rows (columns) have been swapped in decreasing ordering of DNAs' fitness (RNAs' complexity) scores (right panels). Positive (top panels) and negative (bottom panels) interactions have been considered independently. Pearson correlation coefficient absolute values are below 0.4.

We are doing an analogy with the fitness-complexity nestedness model proposed in [citation] where DNAs and RNAs are treated as countries and products.

3 Preliminary results

Here is presented some preliminary results obtained by performing fitness-complexity model on a subset of the dataset. We are considering 120 RNAs and 106 DNAs and kept all Pearson correlation coefficients such that $|\rho| < 0.4$ (this operation reduces the number of RNAs and DNAs constituting the bipartite network). The corresponding bi-adjacency matrix is presented in Fig. 1 left panel and the one obtained by column and row swaps in Fig. 1 right panel. We can see that both positive and negative Pearson correlations based interactions between DNAs and RNAs, from this subset, present a nested structure.