

Spectral Decomposition of the Laplacian Matrix Applied to RNA Folding Prediction

Danny Barash

Genome Diversity Center, Institute of Evolution

University of Haifa

Mount Carmel, Haifa 31905, Israel

dbarash@ucdavis-alumni.com

Abstract

RNA secondary structure consists of elements such as stems, bulges, loops. The most obvious and important scalar number that can be attached to an RNA structure is its free energy, with a landscape that governs the folding pathway. However, because of the unique geometry of RNA secondary structure, another interesting single-signed scalar number based on geometrical scales exists that can assist in RNA structure computations. This scalar number is the second eigenvalue of the Laplacian matrix corresponding to a tree-graph representation of the RNA secondary structure. Because of the mathematical properties of the Laplacian matrix, the first eigenvalue is always zero, and the second eigenvalue (often denoted as the Fiedler eigenvalue) is a measure of the compactness of the associated tree-graph. The concept of using the Fiedler eigenvalue/eigenvector is borrowed from domain decomposition in parallel computing. Thus, along with the free energy, the Fiedler eigenvalue can be used as a signature in a clever search among a collection of structures by providing a similarity measure between RNA secondary structures. This can also be used for mutation predictions, classification of RNA secondary folds, filtering and clustering. Furthermore, the Fiedler eigenvector may be used to chop large RNAs into smaller fragments by using spectral graph partitioning, based on the geometry of the secondary structure. Each fragment may then be treated differently for the folding prediction of the entire domain.

1. Background

The idea of using spectral decomposition in the context of searching and analyzing RNA structures was first proposed in [1], with analogy to computer vision scales and the digital total-variation image processing filter [2]. The

concept is borrowed from the field of domain decomposition in parallel computing, in which spectral-graph partitioning methods are used to subdivide a large domain and assign different processors to each subdomain in order to achieve load-balancing. Let us examine the notion of scales in RNA structures, with the various scales that are illustrated in Figure 1. The *tertiary structure* can be reduced to a *secondary structure*, for which sophisticated computational structure prediction methods based on energy minimization exist (e.g., Zuker's *mfold* and the Vienna RNA package). The secondary structure can be further downscaled to a tree-graph [12, 6, 11] and other graphs [4]. Furthermore, for predicting clever nucleotide mutations that will perturb a given secondary structure [7], the idea of spectral decomposition is to represent these graphs by a Laplacian matrix, and seek the second eigenvalue of the Laplacian (often denoted as the Fiedler eigenvalue [3] that corresponds to the connectivity of the graph). Thus, at the coarsest scale, a single real positive number – the second eigenvalue of the Laplacian matrix – exists that roughly indicates how the folding occurred as opposed to other folding possibilities. This reduced spatial information coexists with the energy of the folded structure, both being real scalar numbers that are calculated independently. Together, they form an effective signature for measuring similarities and discriminating unwanted structures, given a target folded structure. The second eigenvector of the Laplacian matrix can be used to chop the structure of large RNAs such as the ribosome into smaller fragments by using a spectral-graph partitioning method [10] that minimizes dependencies between fragments.

2. Biological Relevance

An example of predicting clever nucleotide mutations by the eigenvalue method on a riboswitch structure [9] to suppress transcription termination in bacteria is depicted in Figure 2. Details on the method will be given elsewhere.

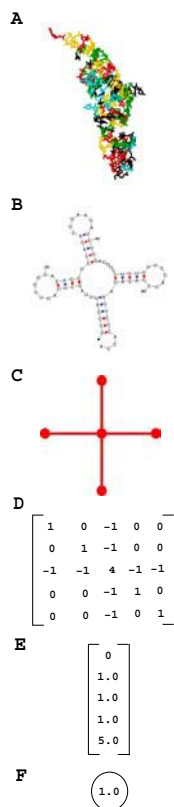


Figure 1. Scales in the RNA biomolecule: (A) Tertiary structure of yeast phenylalanine tRNA. (B) Secondary structure, suboptimal prediction corresponding to experiment. (C) Tree-graph representation. (D) Laplacian matrix. (E) Spectrum of the Laplacian matrix. (F) Second eigenvalue of the Laplacian matrix.

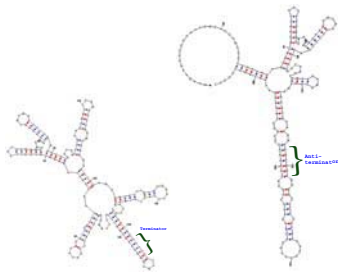


Figure 2. Thiamin pyrophosphate induced riboswitch with a terminator hairpin (left). As a consequence of introducing a predicted mutation found by the eigenvalue method, an antiterminator structure (right) emerges that destroys the terminator hairpin responsible for the transcription termination in *Bacillus subtilis* and other Gram-positive bacteria.

References

- [1] D. Barash and D. Comaniciu, "A Common Viewpoint on Broad Kernel Filtering and Nonlinear Diffusion," *Proceedings of the 4th International Conference on Scale-Space Theories in Computer Vision*, June 10-12, Scotland, UK, 2003.
- [2] T.F. Chan, S. Osher, and J. Shen, "The Digital TV Filter and Nonlinear Denoising," *IEEE Transactions on Image Processing*, Vol. 10, No. 2, p.231, 2001.
- [3] M. Fiedler, "Algebraic Connectivity of Graphs," *Czechoslovak Mathematical Journal*, Vol.23, p.298, 1973.
- [4] H.H. Gan, S. Pasquali, and T. Schlick, "Exploring The Repertoire of RNA Secondary Motifs Using Graph Theory with Implications for RNA Design," *Nucleic Acid Research*, Vol.31, p.2926, 2003.
- [5] I.L. Hofacker, W. Fontana, P.F. Stadler, L.S. Bonhoeffer, M. Tacker, and P. Schuster, "Fast Folding and Comparison of RNA Secondary Structures," *Monatshfte fur Chemie*, Vol. 125, p. 167, 1994.
- [6] S. Y. Le, R. Nussinov, and J.V. Maizel, "Tree Graphs of RNA Secondary Structures and Their Comparisons", *Computers and Biomedical Research*, Vol. 22, p.461, 1989.
- [7] H. Margalit, B.A. Shapiro, A.B. Oppenheim, and J.V. Maizel, "Detection of Common Motifs in RNA Secondary Structure," *Nucleic Acids Research*, Vol. 17, No. 12, p. 4829, 1989.
- [8] D.H. Mathews, J. Sabina, M. Zuker, and D.H. Turner, "Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure," *Journal of Molecular Biology*, Vol. 288, p.911, 1999.
- [9] A.S. Mironov, I. Gusarov, R. Rafikov, L.E. Lopez, K. Shatalin, R.A. Kreneva, D.A. Perumov, and E. Nudler, "Sensing Small Molecules by Nascent RNA: A Mechanism to Control Transcription in Bacteria," *Cell*, Vol.111, p.747, 2002.
- [10] A. Pothen, H. Simon, and K.P. Liou, "Partitioning Sparse Matrices with Eigenvectors of Graphs," *SIAM Journal on Matrix Analysis and Applications*, Vol.11, p.430, 1990.
- [11] C.M. Reidys, P.F. Stadler, and P. Schuster, "Generic Properties of Combinatory Maps: Neural Networks of RNA Secondary Structures," *Bulletin of Mathematical Biology*, Vol.59, No.2, p.339, 1997.
- [12] B.A. Shapiro, "An Algorithm for Comparing Multiple RNA Secondary Structures," *Computer Applications in the Biosciences*, Vol.4, No.3, p.387, 1988.
- [13] M. Zuker, "On Finding All Suboptimal Foldings of an RNA Molecule," *Science*, Vol. 244, p.48, 1989.
- [14] Zuker M., D.H. Mathews, and D.H. Turner, "Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and Biotechnology", *NATO ASI Series*, J.J. Barciszewski and B.F.C. Clark, Eds., Kluwer Academic Publishers, 1999.