

Topological classification of graphs with genus trace

Sebastian Zajac

Decision Analysis and Support Unit
SGH Warsaw School of Economics, Poland

13.05.2022

Outline of talk

1 About me

- Education
- Scientific research and publications
- Private sector work

2 Topological classification of the proteins and RNA

- Why?
- Definitions
- Technical problems
- Total genus
- Genus trace
- Community detection

3 Other directions

Table of Contents

1 About me

- Education
- Scientific research and publications
- Private sector work

2 Topological classification of the proteins and RNA

- Why?
- Definitions
- Technical problems
- Total genus
- Genus trace
- Community detection

3 Other directions

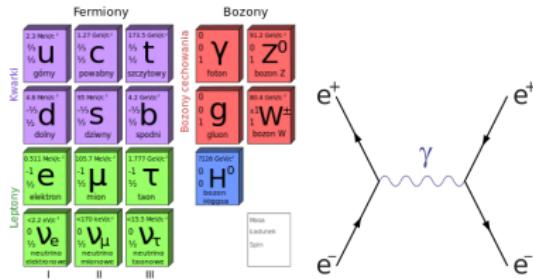
About me

- Bachelor of Science:,,Time Series analysis with ARMA and ARIMA processes. Application in SAS”
- Master of Science:,,Geometrical and topological methods in a classical and quantum field theory”

Ph.D

Accelerator neutrino oscillations and their non-standard interactions.

We can't solve problem by using the same kind of thinking we used we create them.



Interesting things from QM and its (No) connection with statistics.

- A. Einstein, B. Podolsky, N. Rosen: „Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?” *Physical Review* 47 (1935) 777–780
- G. Birkhoff, J von Neumann: „The logic of QM” *J. Ann.of Math.* 37 (1936)
- J. Bell: „On the Einstein Podolsky Rosen Paradox” *Physics* 1.3 (1964)
- A. Aspect: „Proposed experiment to test the nonseparability of QM” **Physical Review D**. 14 (8) 1944–1951 (1976)
- P. Billingsley: „Probability and Measure” J Wiley & Sons Inc (1979)
- B. Feintzeig: Hidden Variables and Commutativity in QM (2013)

How to create a real random number generator using a computer?

Scientific research

Universitates

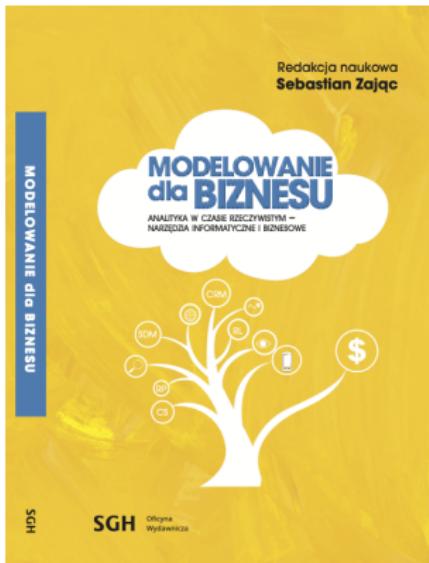
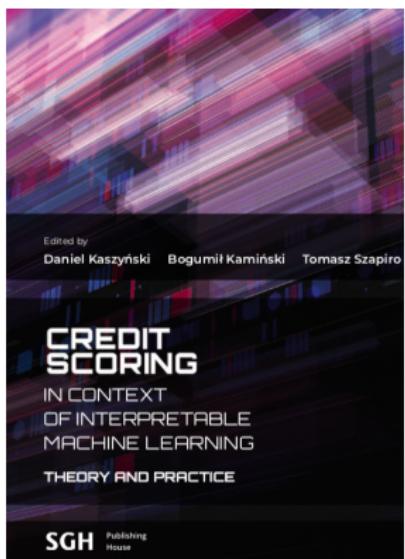
- University of Silesia in Katowice (cooperation 2014–2018)
- University of Warsaw (2013–2014, 2019)
- Cardinal Stefan Wyszynski University in Warsaw (2015–2019)
- SGH Warsaw School of Economics (2019)

Publications

- ① 8 publications on neutrino interactions and discrete, family symmetries for lepton masses and mixing in SM and Beyond.
Physical Review D, Symmetry, Phys.of Atomic Nuclei, Acta Physica Polonica B
- ② publication on connections between statistics and (classical) physical models. *Physica Status Solidi B*

Publications

- ④ 3 monographies on AI and Machine Learning methods (feature selection, modeling) in credit scoring and Real time data analysis. (In prep. *Quantum Machine Learning for Business*)



Private sector work, Achievements

Work

- Data Analysis Expert (2017–2019), Data Science Engineer (2019–2021), MLOps Expert in Risk Department (2021)...
- free Advanced Scorecard Builder Python library, stocking optimization, influence of new data on model quality, model implementation in the cloud environment...

QPoland 2019

QPoland is organized by quantum computing researchers and educators from Poland.

- Workshops and First in Poland Qchallenge with IBM and Paribas
- Moderator on meetup with Washington DC, Toronto QC...

<https://sebastianzajac.pl>



Table of Contents

1 About me

- Education
- Scientific research and publications
- Private sector work

2 Topological classification of the proteins and RNA

- Why?
- Definitions
- Technical problems
- Total genus
- Genus trace
- Community detection

3 Other directions

Topological classification of the proteins

scientific reports

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [scientific reports](#) > [articles](#) > [article](#)

Article | [Open Access](#) | Published: 03 December 2018

Genus trace reveals the topological complexity and domain structure of biomolecules

[Sebastian Zajac](#), [Cody Geary](#), [Ebbe Sloth Andersen](#), [Pawel Dabrowski-Tumanski](#), [Joanna I. Sulkowska](#) & [Piotr Sulkowski](#) 

[Scientific Reports](#) **8**, Article number: 17537 (2018) | [Cite this article](#)

1738 Accesses | 3 Citations | 6 Altmetric | [Metrics](#)

Abstract

The structure of bonds in biomolecules, such as base pairs in RNA chains or native interactions in proteins, can be presented in the form of a chord diagram. A given biomolecule is then characterized by the genus of an auxiliary two-dimensional surface

Why it could be interesting?

Topological Classification of RNA Structures Bon, Vernizzi, Orland, Zee

Journal of Molecular Biology. 379, 4, (2008)

Topological recursion for chord diagrams, RNA complexes, and cells in moduli spaces

Andersen, Chekhov, Penner, Reidys, **Sulkowski**; Nucl. Phys. B866 (2013) 414

standard protein representation

Protein primary structure: the linear sequence of amino acids in a peptide or protein. GCGCGGAUA...

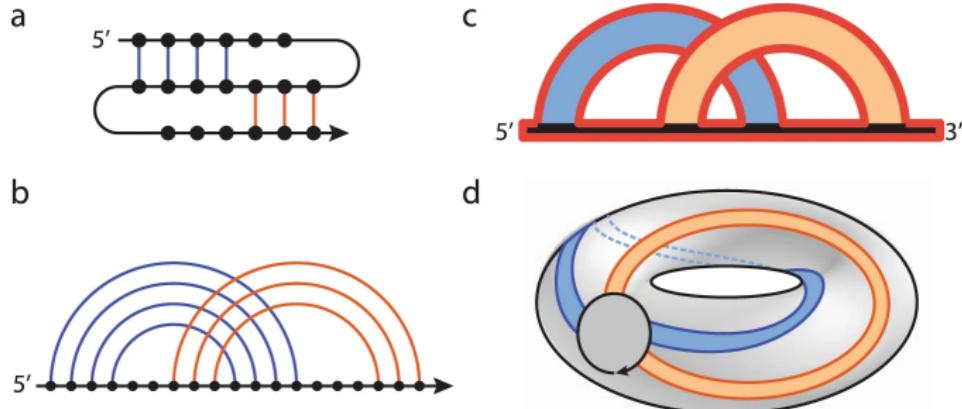
Protein secondary structure:

- linear: the base sequence on an oriented straight line, starting from the 5' to 3' end.
- circular: replace straight line by a close circle

Protein fold = biological function—how to describe?

Energy-based methods, entropy of loops, partition function of molecule.

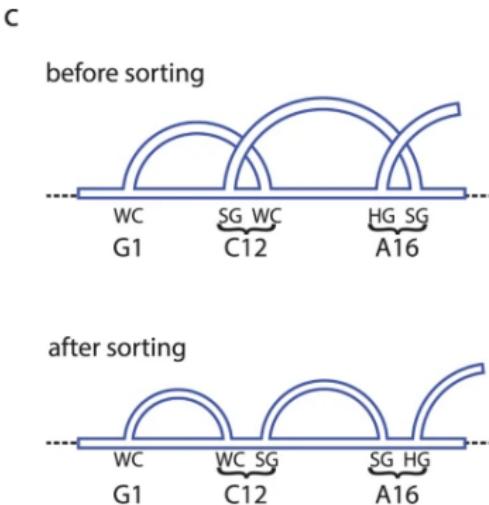
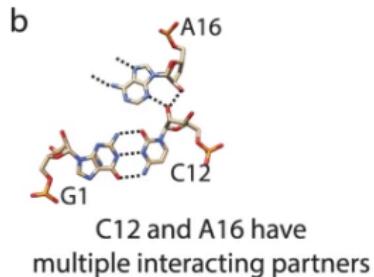
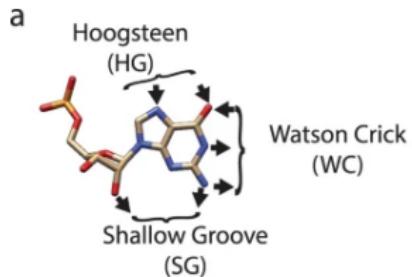
Proteins secondary structure



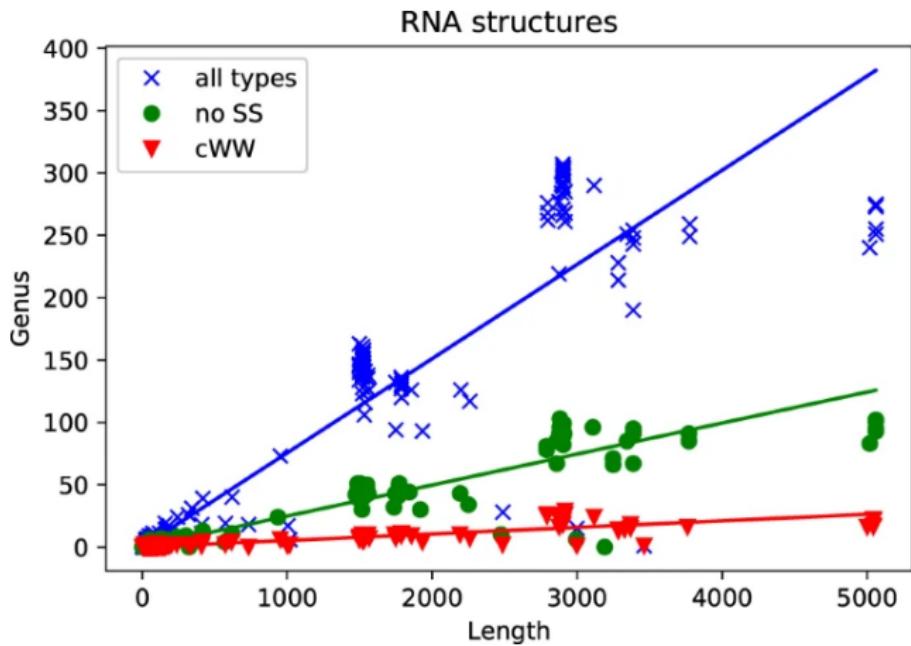
Chord diagram: b horizontal intervals (backbones), n arcs (chords)

$$\text{Genus (Euler): } b - n = 2 - 2g - r$$

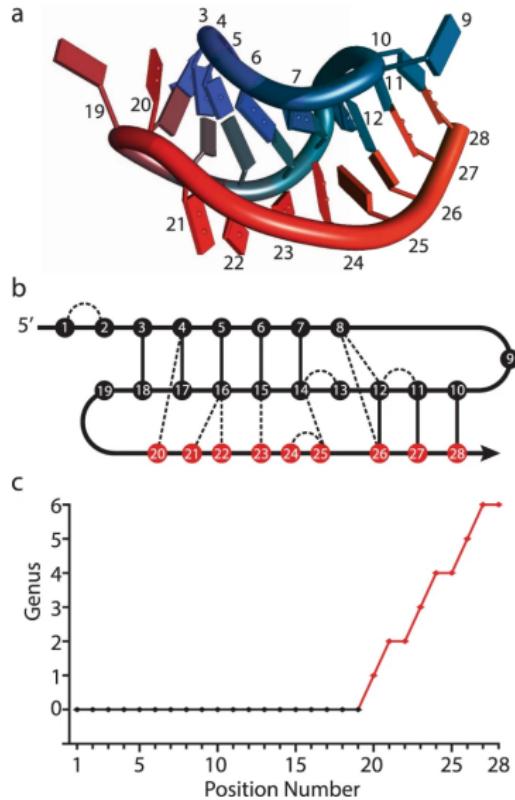
Bifurcation problem



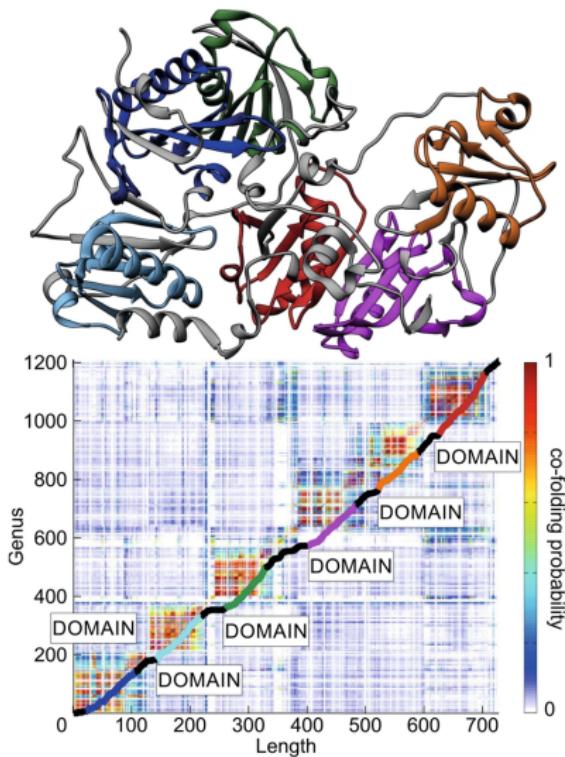
Total genus for structures



Trace genus for the one structure



Domain searching





Volume 48, Issue D1

08 January 2020

Article Contents

Abstract

INTRODUCTION

DISCUSSION

FUNDING

REFERENCES

Comments (0)

< Previous Next >

Genus for biomolecules

Paweł Rubach, Sebastian Zajac, Borys Jastrzebski, Joanna I Sulkowska,
Piotr Sułkowski 

Nucleic Acids Research, Volume 48, Issue D1, 08 January 2020, Pages D1129–D1135,
<https://doi.org/10.1093/nar/gkz845>

Published: 04 October 2019 Article history ▾

 PDF  Split View  Cite  Permissions  Share ▾

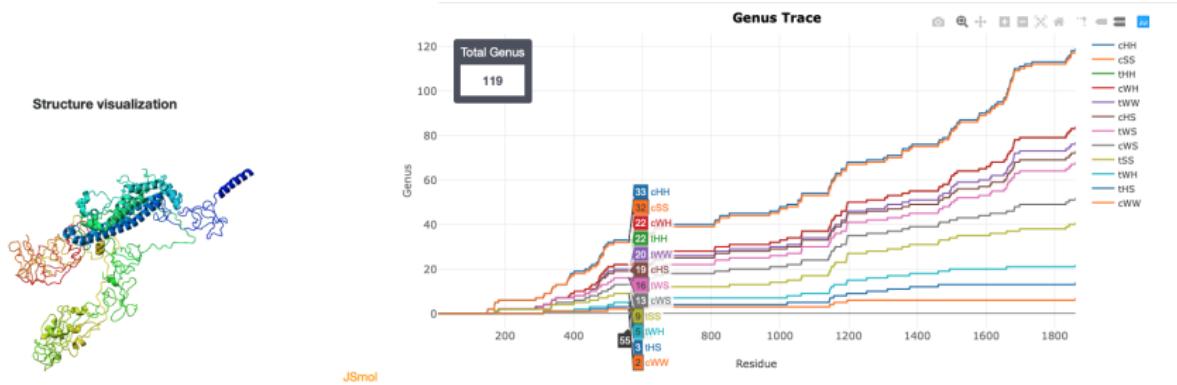
Abstract

The 'Genus for biomolecules' database (<http://genus.fuw.edu.pl>) collects information about topological structure and complexity of proteins and RNA chains, which is captured by the genus of a given chain and its subchains. For each biomolecule, this information is shown in the form of a genus trace plot, as well as a genus matrix diagram. We assemble such information for all and RNA structures deposited in the Protein Data Bank (PDB). This database presents also various statistics and extensive information about the biological function of the analyzed biomolecules. The database is regularly self-updating, once new structures are deposited in the PDB. Moreover, users can analyze their own structures.

Genus web service

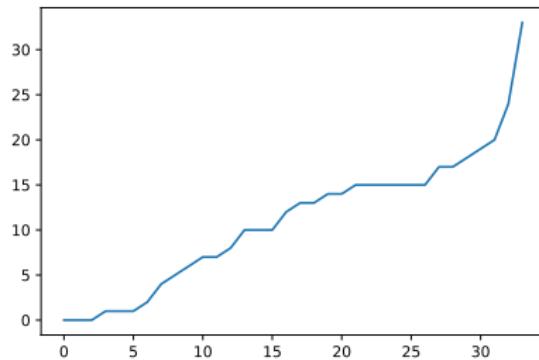
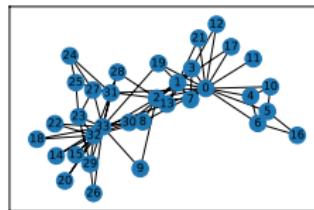
<https://genus.fuw.edu.pl>

Web Service (API) with 309851 proteins and 1575 RNA Structures.



Genus trace for Zachary graph

$N = 34$, $E = 78$, Community = 4



Total genus: 33, 4 Domain detected.

Next work

- Classical community detection algorithms vs. Domain analysis for biological structures.
- Undirected graphs case
- Quantum annealing and quantum approximate optimization algorithm for community detection.
- ...

Table of Contents

1 About me

- Education
- Scientific research and publications
- Private sector work

2 Topological classification of the proteins and RNA

- Why?
- Definitions
- Technical problems
- Total genus
- Genus trace
- Community detection

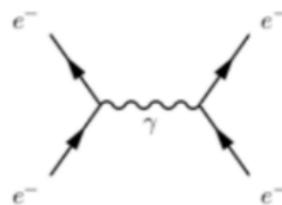
3 Other directions

Diagrams in physics



“Physics is like sex: sure, it may give some practical results, but that's not why we do it.”

— Richard Feynman



„Category Theory is like sex: it may give some practical results, but that's not why we do it. „ - Sebastian Zajac

The category of chord diagrams

We fix three positive integers k , ℓ , and m and a multifunction

$$\varphi: \ell \times \ell \rightarrow m$$

The objects of $\mathfrak{C}_{k,\varphi}$ are structures of the form

$$S = \langle S, <^S, \{B_i^S\}_{i < k}, \{N_i^S\}_{i < \ell}, \{E_i\}_{i < m} \rangle,$$

where:

- (D1) $\langle S, <^S \rangle$ is a finite linearly ordered set.
- (D2) $\{B_i^S\}_{i < k}$ and $\{N_i^S\}_{i < \ell}$ are partitions of S .
- (D3) $B_{i_0} < B_{i_1}$ whenever $i_0 < i_1 < k$.
- (D4) $\langle S, E_i \rangle$ is a graph for every $i < m$.
- (D5) $E_{i_0} \cap E_{i_1} = \emptyset$ whenever $i_0 \neq i_1$.
- (D6) If $x \in N_{i_0}$, $y \in N_{i_1}$ and $\langle x, y \rangle \in E_j$, then $j \in \varphi(i_0, i_1)$.

The sets B_i are called *backbones*, while the sets N_i are *types of nodes* and E_i are *types of edges*.

Category theory

A $\mathfrak{C}_{k,\varphi}$ -morphism from \mathbb{S} to $\mathbb{T} = \langle T, <^T, \{B_i^T\}_{i < k}, \{N_i^T\}_{i < \ell}, \{E_i\}_{i < m} \rangle$ is a mapping $f: S \rightarrow T$ that preserves the linear orderings (that is, $x <^S y \implies f(x) <^T f(y)$) and satisfies for every $x, y \in S$:

- (M1) $f(x) \in B_i^T \iff x \in B_i^S$
- (M2) $f(x) \in N_i^T \iff x \in N_i^S$
- (M3) $\langle f(x), f(y) \rangle \in E_i^T \iff \langle x, y \rangle \in E_i^S.$

Informally, a $\mathfrak{C}_{k,\varphi}$ -arrow is a mapping that preserves the structure of \mathbb{S} , "adding" new vertices and new edges of various types.

In the language of model theory, $\mathfrak{C}_{k,\varphi}$ -arrows are called *embeddings*.

It is clear that $\mathfrak{C}_{k,\varphi}$ forms a category.

Random number

```
import numpy as np
from qiskit import QuantumRegister, ClassicalRegister,
                  QuantumCircuit, execute, Aer

q = QuantumRegister(8, 'q')
c = ClassicalRegister(8, 'c')

circuit = QuantumCircuit(q,c)
circuit.h(q)
circuit.measure(q,c)

simulator = Aer.get_backend('qasm_simulator')
job = execute(circuit, simulator, shots=2)
result = job.result()
counts = result.get_counts(circuit)
print(counts)

{'01000000': 1, '11111011': 1}
```

Summary

Thanks for Your Attention!
sebastian.zajac@sgh.waw.pl