Persistent Homology for biomolecular systems

Guowei Wei

Department of Mathematics Center for Mathematical Molecular Biosciences Michigan State University <u>http://www.math.msu.edu/~wei</u>

> SIAM Life Conference Charlotte, August, 2014

> > **Grant support:**



NIH R01GM, NSF DMS (Focus Research Group), CCF, IIS Michigan State University



Classical topological objects

Möbius Strips (1858)

Klein Bottle (1882)

Sphere













Torus

Double Torus

Trefoil Knot

Topological invariants: Betti numbers

 β_0 is the number of connected components. β_1 is the number of tunnels or circles. β_2 is the number of cavities or voids.



Topological invariants: Euler characteristic						
$C = V - E + F - C = \operatorname{a}^{*}(-1)^{k} b_{k}$						
Vertices (V)	4	4	4	8	8	8
Edges (<i>E</i>)	6	6	6	12	12	12
Faces (F)	0	4	4	0	6	6
Cells (<i>C</i>)	0	0	1	0	0	1
Betti-0 (β_0)	1	1	1	1	1	1
Betti-1 (β_1)	3	0	0	5	0	0
Betti-2 (β_2)	0	1	0	0	1	0
Euler-C.(χ)	-2	2	1	-4	2	1

Network topology





Topology in nano-bio systems

Hierarchical Self-assembly Of DNA Into Symmetric Supramolecular Polyhedra









(Zhang et al., Nature, 452, 198 (2008))























Opportunities, challenges and promises

Opportunities from topological methods:

***New approach for big data characterization and classification.**

***Dramatic reduction of dimensionality and data size.**

***Applicable to a variety of fields.**

Challenges with topological methods:

Geometric methods are often inundated with too much structural detail.

>Topological tools incur too much reduction of original geometric information.

>Topology is hardly used for quantitative prediction.



Mug ⇔ Doughnut

Promises from persistent homology: ✓ Embeds geometric information in topological invariants. ✓ Bridges the gap between geometry and topology.

Vietoris-Rips complexes of planar point sets



Homology



Persistent homology is created through a filtration process

(Xia, Feng, Tong & Wei, 2014)

$$0 = K^0 \subseteq K^1 \subseteq K^2 \ldots \subseteq K^m = K$$

Distance filtration:



Matrix filtration:



Persistent homology prediction of fullerence heat formation energies

(Xia, Feng, Tong & Wei, 2014)



Persistent homology prediction of fullerence isomer total strain energies

(Xia, Feng, Tong & Wei, 2014)



 $E \propto 1/L(\beta_2)$



Topological fingerprints of an alpha helix



O















Protein topological fingerprints

All atom representation

Beta sheet

vs Coarse-grained model Alpha helix

2

6



2

6

4

Topological fingerprints of beta barrel

(Xia & Wei, IJNMBE, 2014)



Topology-function relationship of protein flexibility (Xia & Wei, IJNMBE, 2014)





Microtubule analysis with cryo-EM data







EMD_1129

Molecular structure fitted with tubulin 1JFF

Helix backbone

(Xia & Wei, 2014)

Topological fingerprints at different noise levels

(Xia & Wei, 2014)







Twenty-iteration denoising











Persistent homology aided structure prediction

Processed original data





Fitted with onetype of tubulins





Fitted with twotypes of tubulins

(Xia & Wei, 2014)





Concluding remarks

Introduced molecular topological fingerprints.

Introduced topology-function relationship.

Introduced persistent-homology based quantitative predictions of protein flexibility, protein folding and microtubule composition.

