BioGeometry NEWS
January 2005
biogeometry.cs.duke.edu

## Educational Outreach

First year seminars at UNC Chapel Hill give professors a chance to teach something new and interdisciplinary, and give students a chance to have a small class academic experience in their first year of university. Courses cover the gamut of disciplines, and make many wish they were freshmen again (see the current selection at http://www.unc.edu/fys/current_courses.html).

Jack Snoeyink developed "Folding from paper to proteins" in Fall 2002, and taught it for the second time in Fall 2004. This is a course about shape and structure, explored through origami, robotics, and molecular biology. It considers many puzzle-like questions about folding shapes and structures, including what is one of the biggest puzzles in science, "How does the sequence of amino acids coded by a gene reliably fold into the three-dimensional structure to be a functioning protein?"

The last assignment was to develop a lesson plan to teach a middle school class about some aspect of math or science through folding. These opportunities were advertized through the UNC international center, and a couple of class visits have already been made.

Nikki Malatin, a science teacher at West Caldwell High School in Lenoir, North Carolina, was one of the first to reply because she was excited by the idea of using origami to present lessons in chemistry. Thus, Jack Snoeyink went with graduate student Andrea Mantler and first-year student Amy Jensen to present lessons on chemical bonding in two science classes.

Each student made an atom out of two squares of paper, using a module designed by Yoshihide Momotani. This let them talk about the geometry of bonds: carbons with four bonds in tetrahedral conformation (shown by Laura Thrift and Trista Lee), zwitterions in a benzene ring (held up by Chris Hollar), and how the geometric properties of the oxygen bonding leads to the charged nature of water (oxygen with two bound hydrogens).


Trista Lee with carbon bond


Chris Hollar with benzene ring


Laura Thrift with carbon atom

## Student Profile: Dmitriy Morozov

Dmitriy Morozov is in the second year of his graduate program at Duke University. He is originally from Snezhinsk, Russia, and came to the United States of America in 1998, completing the last year of high-school. He did his undergraduate work at the North Carolina State University in Raleigh, with a double major in Computer Science and Mathematics. He worked with Erich Kaltofen on memory management questions for LinBox, a C++ library for exact linear algebra computation with sparse and structured matrices, and with Larry Norris on a Maple system to grade homework solutions in calculus.

At Duke, Dmitriy first looked at questions about independent complexes, which form an interesting combinatorial relaxation of Delaunay triangulations or, more precisely, of dual complexes of disks in the plane. To define these concepts, we recall that the union of a finite set of disks can be decomposed into closed convex cells using the Voronoi diagram or, more precisely, the power diagram of the disks. The *dual complex* is then the nerve of the collection of closed cells. As it turns out, every simplex (vertex, edge, or triangle) in the dual complex corresponds to a combinatorially unique configuration of (one, two, or three) disks. Specifically, the collection is *independent* by which we mean that for every possible assignment of $k$ boolean variable there exists a point that is contained in the $i$-th disk if and only if the $i$-th variable is **true**. Besides the dual complex, there are other complexes that cover the same subset of the plane and whose simplices correspond to independent collections of disks. We refer to them as

*independent complexes*. Our interest in these complexes is motivated by the fact that they correspond to correct inclusion-exclusion formulas for the union of disks [1]. Dmitriy looked at the question whether or not the *flip-graph* of the independent complexes is connected. In other words, is it always possible to go from one independent complex to another in a sequence of edge flips without leaving the set of independent complexes? An affirmative answer to this question would suggest a particular approach to maintaining an independent complex under continuous motion of the disks (or to the same problem for balls in space used to model a protein or other molecule). Unfortunately, the flip-graph is not connected, as shown by a simple counterexample that Dmitriy found. More generally, he showed that even if we consider combinatorial moves that are more global than edge flips (they affect some constant number of disks at a time), the corresponding graph of independent complexes (whose arcs represent the moves) is not necessarily connected; see Figure 1.

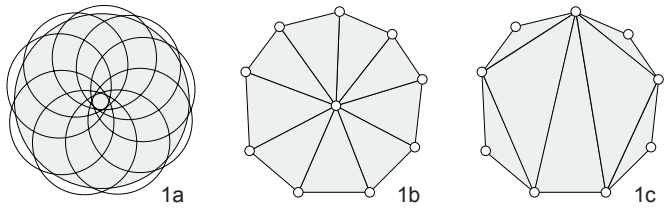


Figure 1. 1b and 1c show two independent complexes for the set of disks in 1a. The complex in 1b forms an isolated vertex in the flip-graph since no flip that affects fewer than all 10 disks will produce a valid independent complex.

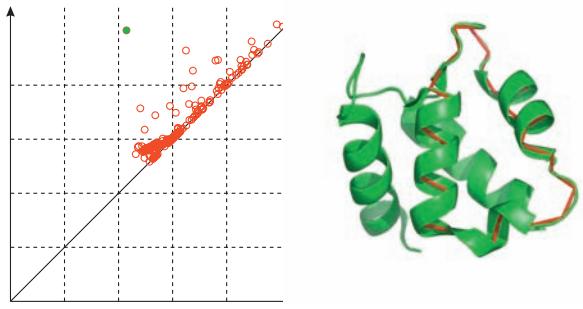

Figure 2. Right: the protein 1IG5 and its most persistent one-dimensional cycle. Left: the persistence diagram of the distance function defined by the protein. The (marked) point furthest away from the diagonal corresponds to the most persistent cycle in the protein.

Dmitriy officially joined the Bio-Geometry group at Duke when he embarked on his second year project with Herbert Edelsbrunner. The topic is the study of persistence diagrams and their use in bio-geometry. The motivation for this work was the stability of these diagrams under perturbations of the input function proved in [2]. To get started, Dmitriy re-implemented the original persistence algorithm [3] and used it to construct the diagram from filtrations of simplicial complexes. Second, he adjusted the geometric matching algorithm of Efrat, Itai and Katz [4] to compute the *Fréchet bottleneck distance* between persistence diagrams, which is the infimum, over all bijections between the points of the two diagrams, of the supremum of the $L_\infty$ distance between a point and its image. This concept figures prominently in the stability result, which states that the Fréchet bottleneck distance between the diagrams is bounded from above by the maximum difference between the two functions.

Using these implementations, Dmitriy can now easily experiment with data from application areas. Take, for example, a protein structure represented by its sequence of $\alpha$-carbon atoms in three-dimensional space. The *distance function* maps every point of space to its Euclidean distance from the nearest atom in this sequence. The persistence diagram records how the topology of the sublevelset of this function changes as we increase the threshold, and it pairs off the changes in topology and maps them to points whose coordinates mark the birth and death of the corresponding homological feature. We have a diagram for each homology group, and the one for the first homology group (representing one-dimensional cycles) is shown in Figure 2.

Dmitriy Morozov

The stability result implies that if we perturb the protein a little bit, we change the diagram only by a little bit. It is therefore tempting to conjecture that proteins in the same structural family have persistence diagrams with smaller Fréchet bottleneck distance than proteins in different families. While this is true in quite a few cases, there are also counterexamples. Dmitriy is now looking deeper into this question, using the distance in space but also other functions. He also studies the folding trajectory of a protein, which maps to a 1-parameter family (a stack) of persistence diagrams. The hope is that the continuous changes in the diagram provide or suggest a qualitative as well as quantitative language to describe and talk about such trajectories.


[1] D. Attali and H. Edelsbrunner. Inclusion-exclusion formulas from independent complexes. In “Proc. 21st Ann. Sympos. Comput. Geom., 2005”, submitted.

[2] D. Cohen-Steiner, H. Edelsbrunner and J. Harer. Stability of persistence diagrams. In “Proc. 21st Ann. Sympos. Comput. Geom., 2005”, submitted.

[3] H. Edelsbrunner, D. Letscher and A. Zomorodian. Topological persistence and simplification. *Discrete Comput. Geom.* **28** (2002), 511-533.

[4] A. Efrat, A. Itai and M.J. Katz. Geometry helps in bottleneck matching and related problems. *Algorithmica* **31** (2001), 1-28.


*- Profile by Herbert Edelsbrunner*


The BioGeometry project is funded by the National Science Foundation under grant CCR-00-86013. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.