BioGeometry NEWS
June 2004
biogeometry.cs.duke.edu

BioGeometry Meeting

We held the first BioGeometry Workshop on June 12, 2004 at Polytechnic University, Brooklyn, as one of the satellite events of the 20th Annual Symposium on Computational Geometry, held at the same location on June 9-11, 2004. The goal of the workshop was to bring researchers from various fields together so that the problem solvers could begin to understand the favorite molecules of the biochemists and the problem posers could begin to understand the favorite methods of the computational geometers.

Over 60 participants registered for the workshop. We thank John Iacono and Hervé Brönniman for taking care of local arrangements. The workshop consisted of four technical sessions with sufficiently long coffee-breaks so that participants could interact with each other.

Each session began with a longer talk by a biochemist on a focused problem – protein-protein docking – and disciplines and problems enriched as the day went on. The first session, chaired by Jack Snoeyink, featured two fifty-minute talks: Johannes Rudolph (Duke) focusing on protein-protein interaction and protein interface surfaces, and Jeff Skolnick (SUNY, Albany) speaking more broadly about protein structure prediction and interaction. Both speakers identified the challenges that lie ahead in understanding interaction between proteins.

The second session, chaired by Herbert Edelsbrunner, focused on shape similarity. In the first talk, Rong Chen (Scripps Institute) gave a summary of the ZDOCK software, one of the most successful procedures for docking two proteins. In the second talk, Yusu Wang (Duke) described an algorithm for protein-protein docking. Then Xueyi Wang (UNC) described a dynamic programming algorithm for placing hydrogen bonds. Finally, Rachel Kolodny (Stanford) gave an in-depth comparison of the well-known methods for computing similarity between two protein structures, discussing the relative performance of these methods.

The first session in the afternoon, chaired by Leo Guibas, was on computational approaches to conformations. In the first talk, Deok-Soo Kim (Hanyang University) described an algorithm for computing the Euclidean Voronoi diagram of a set of atoms represented as spheres. This is the additively weighted Voronoi diagram in computational geometry terminology. Next, Itay Lotan (Stanford) discussed the use of inverse kinematics and Monte Carlo to search for missing fragments in electron density maps. In the last talk of the session, Ioannis Emiris (National University of Athens, Greece) talked about a matrix perturbation solution of the distance reconstruction problem from NMR.

The topic of the final session, chaired by Pankaj Agarwal, was algorithmic problems of various levels of the biological hierarchy of an organism. In the first talk of the session, Daniel Russel (Stanford), described an approach for representing a protein using a spanner. Instead of storing all pairwise distances, this method stores a small subset without losing much information. An advantage of this over competing algorithms is that a spanner can be updated efficiently as the protein deforms. Next, Craig Falls (UNC) presented an efficient algorithm using Turan graphs to compute COGS (Clusters of Orthologous Groups). Finally, David Cardoze (CMU) discussed a method for simulating elastic membranes. The workshop was rounded off by an insightful and entertaining talk by Bud Mishra (NYU) on algebraic models that are useful for systems biology.

The audience was as diverse as the selection of talks, as could easily be seen by the questions from the audience to the speakers. Those who commented said that the talks were stimulating and appreciated, but some computer scientists (who dominated since the workshop followed the ACM Symposium on Computational Geometry) wished that the talks that were intended to be tutorial were at a more basic level. In retrospect, we should have given at least one computer scientist the task of introducing the spectrum of biological problems in the terminology of computer science, rather than beginning with two focussed talks by biochemists. On the whole, however, the interdisciplinary participation was impressive.

Postdoc Profiles: Vicky Choi, Alper Üngör, David Cohen-Steiner

During the academic year 2003/04, three postdoctoral fellows were with the biogeometry group at Duke University. Only one was paid directly from the NSF ITR project, while funds for the other two came from different sources, the BGT Postdoc Program and the Dean's Office of the School of Arts and Sciences at Duke. All three spent a productive few years at Duke and are now about to start faculty and research careers elsewhere. We all wish them the very best in their careers. This short contribution to the newsletter is an attempt to let the wider biogeometry community know a bit about their background and some highlights of their technical work in the project.

Vicky Choi joined our group two years ago, in Summer 2002, after a pre-doctoral fellowship at the NIH and a short postdoctoral fellowship in the

Department of Genetics at Case Western Reserve University. After finishing her Master in Computer Science at the Hong Kong University of Science and Technology in 1995, she ventured abroad and did her PhD in Computer Science at Rutgers University. Working with Martin Farach-Colton, she developed an assembly algorithm for clone-based sequences of whole genomes. She graduated in 2001.

After arriving at Duke, Vicky took over the rigid docking project from Sergei

Bespamyatnikh. At that time the focus was on producing a reliable prediction algorithm for the docking configurations of protein pairs given in their rigid conformations ready for interaction. The major findings were:

1. such reliable prediction is possible based on shape complementarity alone,
2. but it requires a fine sampling of the six-dimensional space of rigid motions in three-dimensional space.

The fine sampling requires a lot of computing time, so Vicky engaged in a new project: the local improvement of approximate docking configurations. With such a tool, we could afford sampling the space of rigid motions considerably coarser. We could also connect the local improvement algorithm with methods that predict hopeful approximate docking configuration. The latter option is particularly attractive since it holds the promise for dramatic time-savings through intelligent sampling of the space of rigid motions. With the completion of a local improvement algorithm and its successful implementation, Vicky has added a major piece to our broad computation attack of the vexing problem of protein-protein docking and interaction.

Alper Üngör joined our group two years ago, in May 2002. Alper grew up in Turkey and did his undergraduate work in Computer Engineering at the Middle East Technical University in Ankara. He then left his country to study in the U.S., graduating with a Master in Computer Science from the Rensselaer Polytechnic University in 1996 and a Ph.D. in Computer Science from the University of Illinois at Urbana-Champaign in 2002. Working with Shang-Hua Teng and Jeff Erickson, he developed algorithms for parallel Delaunay refinement and for space-time meshing.

At Duke, Alper further pursued his research interest in mesh generation, now with a bend towards applications to structural biology. His first piece of work was a relaxed scheduling mechanism that solved an open problem in our work maintaining a high-quality mesh of a molecular skin surface that deforms over time. Alper also worked on the construction of Jacobi curves through adapting a mesh to the local requirements that permit the construction of a clean curve. The problem is made difficult by the inherent instability of the curve. The underlying algorithm due to Edelsbrunner and Harer is combinatorial in nature implying it constructs structurally correct curves, always, and that the instability of the problem manifests itself only in the layout of the curve, which can be zigzag and touch itself at occasions. Alper developed a mesh refinement technique that leads to a smooth curve construction in the majority case in which the local neighborhood varies quadratically. There are, however, cubic neighborhoods (related to birth-death points in a time-varying view of the problem) for which we still lack effective mesh refinement techniques.

Besides pushing the envelope of meshing algorithms, Alper spent his time teaching and working with students on course related projects. He taught Mesh Generation in the Fall of 2003 and co-taught Computational Geometry in the Spring of 2004. In both cases he did very well, earning high grades for his teaching skills from his students who obviously liked his teaching style.

David Cohen-Steiner joined our group a little less than a year ago, in the Fall of 2003. He did his education in France at the Ecole Polytechnique and the Ecole Supérieur des Télécommunications, with specializations in Mathematics and Signal Processing. In 2000 he completed his Master in Computer Vision and Machine Learning with highest honors at ENS Cachan. He then pursued a Ph.D. on topics in surface discretization with Jean-Daniel Boissonnat at INRIA Sophia-Antipolis and he graduated in early 2004.

Even though David's research career is very young, he is well known for landmark results on discrete curvature and isotopic surface meshing. There is another landmark result in the making, namely the proof of the stability of persistence diagrams and its application to shape analysis and characterization. The main result of this work is a central theorem on the stability of geometric structure, and we can expect that it simplifies current work in this field. Consider a continuous function on a topological space and use this function to sweep out the space in the direction of increasing function value. The evolution of the swept portion forms a nested sequence of spaces and we may characterize the connectivity of each using its homology groups. In the evolution of homology groups along this sequence, we see the birth and death of homology classes (topological features of the swept space). The persistence algorithm due to Edelsbrunner, Letscher and Zomorodian is a formal way to pair up these events. In the persistence diagram, we replace each feature by the pair of function values that mark its birth and death, which is a point in the extended real plane. In a nutshell, the theorem says that the persistence diagram (a set of points in the plane) is stable. In particular, small perturbations of the function imply small changes in the diagram. To put this result in perspective, we note that alternative notions that may seem more natural candidates for characterizing the topological features of the function are not stable. These include the critical values that mark births and deaths, the critical points at which the function assumes critical values, and the pairs of critical points identified by the persistence algorithm. Perhaps it is not immediately obvious how this result relates to the biogeometry project, but we should keep in mind that much of our understanding of the world around us is in terms of functions and potentials and that we tend to see meaning where we can detect stable structures. This is particularly true for the world of proteins, which is accessible to us only through indirect means, such as reconstruction from x-ray diffraction patterns.

- Profiles 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.