PREVIOUS RESEARCH

Undergraduate Research

As an undergradate at Pomona College, I participated in the Summer Undergraduate Research Program each of my three summers. During my freshman year, Professor Alfred Kwok invited me to investigate the long-range capabilities of "laser tweezers," the technique of trapping small particles approximately 10 microns in diameter in the focal point of an infrared laser beam (3). Working in his optics laboratory alongside fellow undergraduate student, Perry Schiro, I prepared and calibrated equipment, carried out experiments of my design, and performed data analysis of the particles' trajectories. I always liked to describe the project to my friends as a scaled-down version of the Death Star tractor beam, and I later had the opportunity to present my findings to fellow undergraduate researchers (1). I love discussing my research: what went right, what went wrong, and why it was worth getting excited about.

My enthusiasm for research grew stronger when I first began cancer research after my mother was diagnosed with breast cancer and my sister was diagnosed with spindle-cell sarcoma. I received a summer research grant in 2004 from the Howard Hughes Medical Institute to help Professor Ami Radunskaya develop a cellular automata model for melanoma growth. The model considered local concentrations of oxygen and glucose and local pH, all of which can affect the growth rates of both cancerous cells and noncancerous cells. After researching the proposed causes of hypoxia and its role in the development of a necrotic core, the inner section of dead tissue often found within tumors, I helped Dr. Radunskaya create a novel model of melanoma growth which produces the correct large-scale behavior while considering only microenvironmental conditions. It was immensely rewarding to be able to see the visualization of our model in action as chemical concentrations diffused throughout the simulation space, causing tumor cells to either proliferate or die. We shared our work with a group of oncologists at UCLA who were eager for mathematical tools that could inform their work, and both sides gained valuable feedback for future directions.

Eager to continue cancer research, in my undergraduate thesis in applied mathematics I studied the structure of the blood vessels that surround tumors. Tumors often initiate the creation of new blood vessels in order to fuel their rapid growth. These new vessels are more tortuous, clustered, and irregular than vasculature found in normal tissue. Working with Professor Vin de Silva, I developed a protocol that we hoped would distinguish between images of tumor vessels and normal vessels. Employing algorithms from computational topology, the method was able to highlight areas of the images that appeared more tangled. This experience gave me valuable practice with new material such as algebraic topology and gave me plenty of time working independently, programming in the lab and documenting my project in the form of a thesis. The method was of minimal practical use, however, because it required scanning electron microscope images of vasculature, which are both time-intensive and expensive to obtain as they require biopsies. This work provided an early lesson that data-driven methods are highly dependent on the kind of data currently available to the research community. I am confident, however, that computational topology techniques such as these will play a role in the future of medical imaging and diagnostics.

The culmination of this research came in the winter of 2007 when I participated in the Workshop on Applications of Mathematics to Biomedicine at the University of Otago in Dunedin, New Zealand. As part of a team of six, I worked closely with other mathematicians to solve problems in pharmaceutical research, such as predicting drug release kinetics of common tablets.

Graduate Research

Upon beginning my graduate work during the summer of 2008 at the University of California, Irvine in the School of Information and Computer Sciences, my previous interests influenced a new direction of research. My current work explores the hypothesis that massive networks of online collaboration and communication (such as Wikipedia and the blogosphere) can be explained by an accumulation of simple rules. From large amounts of relational data, can one learn time-dependent models of user behavior that are responsible for the complex topologies that we observe? This work, in collaboration with my advisor Padhraic Smyth, is especially exciting because only recently have computational resources advanced to allow one to collect, store, and visualize such massive, complex data sets. Further work, however, is required to produce the algorithms and methods that quickly synthesize this heterogeneous data into a single model and allow analysis at arbitrary levels of abstraction. As an important preliminary step, I am becoming familiar with the software packages currently used for quantitative network analysis. I am also preparing a data set of interactions between more than 10 million blogs over the course of 6 months. This dynamic graph will consist of events for each blog post and the link structure between them. This will provide ample opportunity for the testing of our models against real-world data.

I recently created the UCI Network Data Repository (2) in order to facilitate the use of real data in modeling and prediction efforts as well as encourage the widespread study of network methods for application to a variety of scientific fields. The repository provides the research community with a place to share data sets, methods, tools, and insights, and several major researchers in the field have already agreed to share their data via the site.

In August 2008, I attended the Knowledge Discovery in Data Conference in Las Vegas, NV, where I was able to converse with leaders in the field such as Jon Kleinberg, Jure Leskovec, Christos Faloutsos, Dave Jensen, and many of their coauthors. Whereas the current progress has been in modeling and categorization of various network qualities, they mentioned that prediction and inference are lacking, yet sorely needed, elements in the analysis of large-scale dynamic networks.

I plan to explore the theoretical and computational foundations of dynamic network models and perform prediction in real-world situations. With the NSF Graduate Research Fellowship, I will be able to work directly on this fundamental research. The multi-year support will allow me to integrate related ideas and modeling efforts from a variety of fields, with a focus on bridging the network modeling viewpoints of computer science and statistics. My progress in this burgeoning field could simultaneously advance aspects of economics (by improving models of market contagion), social science (by increasing the scale at which we can perform social network analysis), and systems biology (by allowing for inference on large networks of interacting proteins and other objects). The NSF Fellowship will allow give me the autonomy and flexibility to perform research that spans departments, disciplines, and institutions.

References

[1] DuBois, C. L. Large capture-range of a single-beam gradient optical trap. Symposium on Undergraduate Research, American Physical Society - Division of Laser Science, October 2003.

[2] DuBois, C. L. UCI Network Data Repository, 2008.

[3] Schiro, P. G., DuBois, C. L., and Kwok, A. S. Large capture-range of a single-beam gradient optical trap. *Optics Express 11*, 25 (2003), 3485.