Statement of Research Accomplishments and Plans

Jonathan W. Berry

December 17, 2001

My research areas are experimental algorithmics and software systems for discrete mathematics. Specifically, I am interested in using computer tools to aid the process of developing new algorithms and heuristics for discrete optimization problems. I also have experience in the design and development of general-purpose software for discrete mathematics. It is a long-term project of mine to help DIMACS (the Center for Discrete Mathematics and Theoretical Computer Science) increase the influence of discrete mathematics in education at all levels.

My favorite discrete optimization problems are those involving *graph partitioning*. The goal of this type of problem is to split the vertex set of a graph into two or more parts, with approximately the same number of vertices in each part, such that the number of relationships (edges) between the vertices in different parts is maximized or minimized. This turns out to be crucially important in computer design and in parallel computation. It is remarkable that nobody knows how to solve these problems both optimally and efficiently.

I have had success both with algorithm design and with software systems. My highest-visibility research accomplishment to date has been the Path Optimization heuristic, developed jointly with Mark K. Goldberg, which uses a novel definition of "neighborhood" for neighborhood search. The latter is a technique in which the "neighbors" of a solution to a problem are determined and examined. For example, the neighbors of a candidate graph partitioning solution might be those obtained by swapping two vertices between partitions. In Figure 1, a split vertically down the middle would create the partitioning $\{\{1,3\},\{2,4\}\}$ with a cut size of four edges. A neighboring solution could be found by swapping vertices 1 and 2, resulting in the partitioning $\{\{2,3\},\{1,4\}\}$ with a cut size of 3 edges. Our journal paper on this topic in *Discrete Applied Mathematics* (90), 1999 has been named to that publication's "Editor's Choice, 1999" edition.

I was also project leader for the multi-institutional *LINK* software system. The canonical paper describing the system appears in *Software Practice & Experience* (30), 2000. My leadership in this effort resulted in an invitation to be one of the primary lecturers in DIMACS' 1998 "Reconnect" conference, designed to reconnect scholars at non-research institutions with the research enterprise. It was gratifying to be chosen as one of the leaders, considering that since my institution is a teaching institution, presumably I should have been one of those needing "reconnection." I was also invited to give a plenary lecture at DIMACS' "DREI" (DIMACS Research and Education Institute), which was similar to Reconnect, but sought to create bonds between high school teachers and researchers. I had trained some of my students to serve as teaching assis-

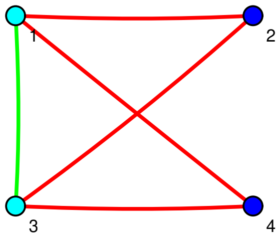

Figure 1: An example graph

tants to help the various faculty learn and use *LINK*, and they had a rewarding experience that summer.

In addition to these largely individual successes in two different areas, I have been able to sponsor undergraduate students to work with me during four of my five summers at Elon. I was also able to sponsor an REU student during my last summer at DIMACS, where I was a postdoctoral fellow. My students and I have worked in both of my areas, and the work has been steady and incremental. I am particularly excited about my work with then-undergraduate John D. Marshall, II, who is currently a graduate student at Florida State University.

John and I spent the summer of 1999 developing extensions to the *LINK* system that were tailored to help designers of graph partitioning heuristics build customized visualizations. This work built upon that of myself and then-undergraduate Damon DeSonier, who had worked with me in the summer of 1997. At the end of the summer with John, we were both invited to Rice University by Dr. Nate Dean. Our task was to help Rice graduate students implement similar extensions to *LINK*. We were successful in this regard, but much more significantly, we used that week to develop a new heuristic for graph partitioning. The new work, as yet unpublished, is quite unrelated to Path Optimization.

Local optimization techniques such as the famous Kernighan-Lin algorithm, or Path Optimization, must be supplied with initial partitionings. Furthermore, there is significant evidence that the choice of initial partitioning is important. My work with John Marshall proposes a new algorithm for creating initial partitionings in a novel way. Our work, which has been ongoing for the past year and a half, meshes nicely with the state of the art in local optimization, "Reactive Randomized Tabu Search," by Battiti and Bertossi. We have combined our ideas with theirs to achieve new world-best partitionings on a familiar testbed of graphs often cited in parallel computation literature.

Although our work is not yet finished, John presented some preliminary results at the 2000 ACM Southeast Regional Conference, and won third prize in their student research contest. A fair question to ask might be, "Why should we hire the mentor of a *third* place effort?" My defense is that the only research projects rated ahead of John's were large-scale, long-term, team research projects (e.g. RoboCup) in which the associated undergraduates played minor roles. I am much more impressed by John's publishable one-on-one work with me than the judges were.

Most recently, I have worked with undergraduate Ruth Brown to evaluate "Boost" template library as a potential software basis for a complete redesign of the *LINK* software system. This template library is a serious candidate for incorporation into C++'s Standard Template Library, yet it makes no allowance for hypergraphs (graphs in which relationships are not limited to those between exactly two vertices – three or more vertices can be related in one "hyperedge"), and seems vastly overcomplicated.

In the near term, I have several definite research plans. In particular, I have been accepted to spend a sabbatical year at Sandia National Laboratories in the massively parallel computing group. Although my upcoming switch in institutions is going to mean that I will not be able to spend the year there as I had planned, I hope to work with the researchers there in shorter blocks of time, perhaps bringing students along. I was to use my skills in both experimental algorithmics and software design to help them design and run massive experiments with discrete algorithms.

I have also been selected to be a site leader for another instance of DIMACS's Reconnect program in the summer of 2004. The proposal has been submitted to the NSF by DIMACS, listing myself as a site director at Elon. My switch in institution may make the new institution one of the Reconnect sites, though I am not sure about the portability of this NSF grant.

In the long term, my plans are somewhat contingent upon the environment in which I am to work. At Elon, I have always had at least six distinct course preparations per year, so I have had to content myself with sporadic work on graph partitioning during the academic year followed by flurries of very focused effort in the summers. In an environment with a lighter teaching load and stronger research expectations, I plan to expand my interests and activities to include:

- considering the new models of partitioning problems recent proposed by Bruce Hendrickson at Sandia,
- addressing other discrete optimization problems,
- using my experiences with software and educational users to design a successor to *LINK*,
- a background study that may lead to original research in computational biology, which was one of the subjects of Reconnect '98, and on which I have worked with Elon Science Fellow students, and
- a background study of quantum computing.

It goes without saying that I intend to involve students in all of my scholarly activities.