PROJECT DESCRIPTION

1 Introduction

The principal investigators (David Aldous, Steven Evans, James Pitman), with the cooperation of other faculty (Venkat Anantharam, Elchanan Mossel, Fraydoun Rezakhaniou, Alistair Sinclair, and Allan Sly), propose to establish a research training group of undergraduates, graduate students, postdoctoral fellows, and faculty in the area of mathematical probability and its applications. Probability has long been recognized as a field with broad links within theoretical mathematics (e.g. measure theory, functional analysis, combinatorics) and broader applications (classically to e.g. statistics, operations research, genetics, statistical physics). Over the last generation new links to pure mathematics have been explored (and recognized by several recent Fields medals) and new disciplines involving probability have emerged and grown (finance, Markov chain Monte Carlo in Bayesian statistics and the theory of algorithms, ad hoc networks, etc).

U.C. Berkeley has a long tradition of strength in mathematical probability, at the faculty and Ph.D. program level. Many ingredients and structures for a strong RTG program, as intended in the EMSW21 framework, are already in place, as they have been for many years. These include advanced topics courses and seminars, and mentoring and advising procedures for graduate students. The proposed RTG will work with these existing activities as well as adding new ones. Though centered in the Statistics Department it is intended to have substantial interaction with the Mathematics, Computer Science and Electrical Engineering Departments.

Over the last few years a range of undergraduate research and training activities, outside of the traditional textbook-based course format, have been undertaken on a one-off basis under the initiative of individual faculty. Under the RTG we will restructure all of these activities into a coherent package, to increase visibility and make it clear these are “activities that involve undergraduates in discovery and generate within them appreciation and excitement for the mathematical sciences”, as the program solicitation phrases it.

Regarding the graduate program, we are satisfied with the quality but the actual size of the program (and the number of U.S. students in particular) is currently rather smaller than we wish. The RTG will provide a major opportunity to increase the number of graduate students in the target group that we can train. Several matters specific to Berkeley are worth noting here. We seek to exploit the recent huge increase in size of the undergraduate Statistics major (to around 100 graduates per year, almost all in the target group of citizens or permanent residents) as a pool of strong U.S. students who might be attracted to careers in the mathematical sciences. During 1992-2012 we have been able to offer graduate fellowships in probability funded by a bequest from Madame Loeve, but the terms of the bequest required the money to be completely spent over this 20-year period. In view of the well-publicized budget problems of the State and the University of California, the existence of secure NSF funding might be reassuring to prospective graduate students from outside the State.

In a typical semester there may be several postdoctoral-level researchers in prob-
ability present in Berkeley, but with divergent statuses and often not well integrated into Departmental activities beyond a specific small research group. As another innovation planned under the RTG, we will recruit postdoctoral associates specifically for RTG activities, in particular the mentoring of graduate and undergraduate students, and teaching graduate courses on current topics outside the expertise of permanent faculty.

Graduate students studying probability are typically supported as teaching assistants for introductory statistics or calculus courses. We propose to use RTG support both to increase the number of graduate students and to partially replace such traditional teaching activities by more innovative ones, in particular assisting instructors in the two new upper division courses described later. Currently budget considerations mean that no probability course has a TA, except the first year graduate course where the TA is mostly occupied with grading homework, so the innovation of having the graduate students teach their own subject should be beneficial to both sides.

The new formal activities of the group will be as follows. For undergraduates, a seminar intended to show what Ph.D. level research is like and what subsequent careers are available; two upper-division activity-based courses in probability (one focussing on exploring the breadth of applications, the other focussing on engaging students in a stochastic processes course by having them lecture to the class), and a more organized structure for undergraduate research projects. For graduate students, a graduate expository seminar, specialized seminars, additional graduate topics courses, and a small summer workshop to be followed by team research.

The bulk of our budget request consists of support for 10 graduate students (trainees) and 2 postdoctoral associates, as well as a small amount of support for undergraduate researchers. We are also requesting smaller amounts to complement that support, primarily by providing travel funds for trainees and postdocs to attend conferences, for outside visitors to speak in the summer workshops and the seminars, for students and postdocs from other institutions to attend the summer workshops, for very limited salary for faculty and/or administrative assistance in connection with organization of the program (the summer workshop in particular), and for technical support of content management systems associated with the expository aspects of the project.

The Statistics Dept hosted a VIGRE grant during 2003-2011. A small part of the past activities mentioned in this proposal were supported by that VIGRE grant, though its focus on Statistics was substantially different from the focus of this proposal, and there is no overlap of PIs. However the present PIs will be able to call on institutional experience of the mechanics of running such a program.

2 Proposed project

According to the program solicitation, “The RTG program is not meant to establish a group, but to enhance the training activities of a well established one”. In this section we describe the new and restructured training activities proposed. Section 8 will record more specifically the ongoing research activities of the group.
Undergraduate seminars. Their purpose is to give students a view of possibilities that lie beyond undergraduate experience. Some talks will be from graduate students and postdoctoral associates giving fairly non-technical accounts of the content and the experience of Ph.D. level research. Other talks will describe career possibilities, and for these we can call on recently finished Ph.D.s to talk about careers such as finance, bioinformatics, Google, NSA, or teaching in a small college.

Activity-based upper division courses in probability. The Statistics undergraduate major consists of traditional textbook/homework/final courses on theory and methodology, plus courses on statistical computing and courses emphasizing statistical analysis of real data. We propose to extend the latter style of “activity-based” course by establishing as regularly-taught courses the following two courses emphasizing probability rather than statistics. (over the last 10 years, the first has been taught 3 times and the latter 1 time; we propose now to teach each for 1 semester each year). One (taught by Aldous) is designed to illustrate the breadth of applicability of probability in the real world via lectures on 20-25 very different topics. Students are expected to read further on one of these topics, and to do a substantial course project on a topic of their choice involving real data. This demands substantial individual supervision by the instructor. Because there is a limit to the number of students that one faculty instructor can effectively engage with, we propose to use a RTG-supported graduate trainee as assistant to enable more students to effectively take this course. The other course (taught by Pitman) seeks to engage students more actively in a mathematical stochastic processes course by assigning the students themselves to read sections of a text and to give a lecture. We intend to experiment with ways of involving graduate trainees in this and other innovative styles of undergraduate course.

Undergraduate research projects. Individual faculty have supervised undergraduate research under the VIGRE program and also under campus programs such as URAP Undergraduate Research Apprentice Program or SURF (Summer Undergraduate Research Fellowship). Such one-on-one interaction with faculty is surely one of the most enriching parts of an undergraduate experience. But there is a limit to the amount of time faculty can devote to it. We propose to involve postdoctoral associates in this activity. We also propose to use a graduate student supported by the RTG to hold regular meetings of all undergraduates involved in such research. Having the undergraduates talk about what they are doing to third parties is a good way to check their understanding and progress. Introducing a single web page and a graduate student coordinator will facilitate organizing students to work in small groups. This will help train students both to learn from their peers (many have different second Majors) and to learn how to collaborate. As intended in the program solicitation, “research” here is interpreted very broadly. Past projects have featured data collection and computer simulation and graphics, as well as more explicitly mathematical work, and we will also encourage expository writing in Wikipedia style, emphasizing conciseness and explicit citation. These activities also address the “communication skills” aspect of the solicitation. Finally, some projects are featured in the department’s offering on Cal Day, the campuswide “open house” for the community and prospective
undergraduates.

**Web site improvement.** With minor effort we can improve the current web site to spotlight the RTG activities: (for existing Berkeley students) the undergraduate level activities and (for prospective graduate students) the Ph.D. program in Probability.

**Graduate expository seminar.** We propose a new seminar with the following format. A small group (2 or 3) of graduate trainees are assigned to read a set of papers on a topic, to decide what the most important points are. Then each gives a 90-minute talk, and they collectively write up an extended version of the talks. This is informative for the audience and instructive for early graduate students to learn to write comprehensible mathematics. We will encourage the best write-ups, or those on topics related to someone’s thesis, to be further expanded into publishable survey papers.

We will designate this seminar as the core part of the (graduate and postdoctoral) RTG program, providing refreshments and requiring attendance. The talks will be at a level accessible to all second-year graduate students and to the majority of first-year students. The material is intended to slowly cycle through the different interests of all RTG members. This seminar will also provide a venue for organizational matters.

**Overall seminar activities.** At present there are 3 weekly probability-focussed seminars operating independently. The Department Probability Seminar features outside speakers in the traditional research-talk format with little active audience involvement. Graduate students on their own initiative started a “Combinatorial Stochastic Processes seminar” on topics related to Pitman’s work. Rezakhanlou runs a “Probability and PDE” seminar on topics related to his work. And of course numerous other seminars in Statistics, Mathematics and EECS are somewhat related to probability.

Under the RTG, we will continue the Department Probability Seminar, institute the new Graduate Expository Seminar above, and run more specialized or advanced seminar series (typically two each semester), partly in the spirit of the existing ones and partly to exploit the different research interests of postdoctoral associates.

**Graduate topics courses.** The Statistics Dept offers a first year graduate measure-theoretic probability course each year. This is the only fixed course in the probability Ph.D. program. Otherwise we teach topics courses, according to the current research interests of faculty, at least one each semester, and seldom repeating the same topic. Students can therefore take 6 or 7 different topics courses during their sojourn in Berkeley, representing one of the unique features of the current Berkeley program. The existence of RTG-supported postdoctoral associate will allow the department to increase and diversify such topics course offerings.

**Summer graduate workshop.** Every year 2005–2012 there has been a scheduled Cornell summer school in Probability. And in 5 of those 8 years there has been a PIMS Summer School in Probability at Vancouver or University of Washington. These have been 2 or 3 week courses, aimed at advanced graduate students, postdoctoral students
and young faculty, with 3 mini-courses on current research topics plus talks by individual invited speakers and attendees. Anticipating continuance of the above, we propose a smaller scale and somewhat differently focussed activity. We will advertise and offer a 1 week course on some topic of current research interest, intended both for our RTG-supported graduate students and post docs, and for target group members from other institutions, who will be offered travel support (airfare and a modest per diem). While some attendees will merely listen, the intention is for some of our RTG-supported participants, and any from other institutions who wish, to remain for several more weeks to work in groups on research problems specified in the course. This fits the “supportive environment” and “research training” aspect of the program solicitation. Pitman has informally done such activity, for example recently on convex minorants with post-docs Nathan Ross and Geronimo Bravo and graduate student Josh Abramson, resulting in a 4-author overview paper and 3 joint technical papers. As a related activity, Pitman and Aldous (who, with Marc Yor, founded the journal Probability Surveys) maintain an interest in advanced expository writing, and so will encourage graduate trainees and post-docs to participate in such writing. This benefits both the trainee (via learning material and being trained to communicate) and the broader probability community. We regard this as a small instance of the kind of “cultural change” sought in the solicitation.

**Communication of knowledge.** Pitman’s work on online open access to mathematical knowledge gives RTG participants the opportunity for various activities which have both training value for participants and intrinsic value for the broad probability community. At levels from undergraduate to postdoctoral, these opportunities include (repeating some mentioned previously)
- glossaries of terms found in textbooks
- non-technical accounts of topics, as Wikipedia or StatProb.com entries
- summaries of individual research articles
- expository articles aimed at the research community.

To amplify the third item: as a routine part of training of graduate students, a thesis advisor will often tell an advisee to read a research paper and then explain it to the advisor on the board. But the tangible result is then erased with the chalkboard. Having the student write out a summary, which can then be posted publicly (e.g. at akawiki.org), should both reinforce the individual student’s understanding and be helpful to the community.

### 3 Recruitment and Retention

**Graduate students.** In general, we are satisfied with our current level of retention. As part of the previous VIGRE program the Statistics Department (which will be responsible for admissions for the majority of RTG-supported graduate trainees) revamped the recruitment process, to include the following steps.

1. Meet with Dr. Collette Patt, Director, Science Student Diversity Programs, to brainstorm ideas for increasing the applicant pool and for convincing admits to come to Berkeley.
2. Early in the Fall semester, we send emails to colleagues and targeted institutions around the country requesting that they identify potential applicants for our PhD program. The email includes a short description of RTG opportunities for domestic students. This email is sent to:
   a) Graduates from our department who are now in academic positions in the U.S.
   b) Chairs of math departments at colleges that had produced a high number of female B.A.s who have gone on to obtain Ph.D.s in the mathematical sciences.
3. Admission Letter, emphasizing
   a) Enthusiasm about the prospect of having them come to Berkeley
   b) A description of the exciting research environment here, and how we think that they would make a valuable contribution to it.
   c) The financial package, with sample years.
4. Contact each admitted student by phone to answer any questions that they might have and invite them for a visit to the department.
5. We will use RTG funding to invite admitted students to an special Visit Day, which includes meetings with faculty, lunch with graduate students, and an evening party with graduate students, postdocs, and faculty. We recently reorganized our traditional visit day to include:
   a) More one-on-one interviews with faculty and graduate students;
   b) Shorter and more varied presentations about the department;
   c) Open house with a broad spectrum of graduate students at various stages in the program and in various research areas, where the graduate students were in their offices over a two hour period and the candidates visited those students that they were interested in talking to;
   d) Opportunities to sit in on classes and seminars.
6. After the visit, we ask faculty to contact particular admits who were interested in their research areas.
7. Follow-up phone call and several emails to see if there were any obstacles that were keeping them from deciding to come to Berkeley. We attempt to respond to these issues.

**Under-represented minorities.** In Fall 2010 the department instituted an MA-PhD track for students from under-represented minority backgrounds. Students from minority backgrounds who have the potential to succeed in the Ph.D. program but lack adequate course preparation are admitted with full support into the M.A. program, and, with appropriate advising and course selection, begin taking Ph.D. level courses in the second year of the M.A. program. Satisfactory progress in these courses leads to transfer into the Ph.D. program after the M.A. The program was inaugurated with three students who are doing well and are expected to move to the Ph.D. program in Fall 2012; another student was recruited in 2011. The program is supported in part by Graduate Division and the College of Letters and Sciences, and is viewed by campus leadership as a pilot for more such programs to be developed in campus STEM departments in the future.
Undergraduates. The RTG-supported undergraduate opportunities will be advertised to Berkeley students in relevant courses, to those applying for the Major, and on the web site.

Postdoctoral associates. The Department advertises postdoctoral positions widely. The RTG-supported positions would be coupled with a three-year assistant professor position. Applicants would be selected by the PIs with the assistance of the department Faculty Equity Advisor, charged to take special care that women and minorities are well represented in candidates, to suggest names, seek out candidates, and urge them to apply.

4 Performance Assessment Plan

Our goals are to inspire more undergraduates to continue to graduate training (either specifically in probability or within the wider mathematical sciences), to increase recruitment of the target group (citizens or permanent residents) into the graduate program in probability, and to give postdoctoral associates more substantial engagement with graduate and undergraduate students (beyond the typical participation in a small research group and undergraduate course teaching). The performance of individual graduate trainees and postdoctoral associates during the program will be reviewed by the senior faculty. The researchers will also write yearly reports summarizing the activities and accomplishments funded by the grant. (See supplement for Postdoctoral Researcher Mentoring Plan). We will record, on the RTG website, the post-RTG placement of participants; this will allow us to monitor improvements in training as measured by jobs obtained. We will also work with graduate admissions to monitor our success rate in recruiting graduate students, especially members of underrepresented groups, that might become part of the RTG. Finally, to the extent possible, we will keep track of undergraduates that are channeled into graduate school through RTG activities.

5 Organization and Management Plan

The three PIs (Aldous, Pitman, Evans) will take primary responsibility for seeing that the group’s activities are carried out as set forth in this proposal, and keeping track of these activities for monitoring and reporting purposes. The various seminars will usually be co-organized by a faculty member and a postdoctoral associate. RTG supported graduate trainees will be assigned to (i) maintain the web site (ii) run the coordinating meetings for undergraduates involved in research projects The summer workshop organizing will rotate among the senior personnel with the help of a trainee assistant. Other responsibilities, especially for reviewing applications and making hiring recommendations for postdoctoral associates, will fall to the senior personnel collectively, who will meet as a group to undertake them. Each postdoctoral associate will be assigned one of the senior personnel as a mentor. Incoming graduate students
are assigned mentors by the Department, and later acquire thesis advisors, and for RTG-supported students these will invariably be faculty in the group.

6 Dissemination

As previously mentioned, we will improve the current pages on the Statistics Dept web site to spotlight and coordinate the RTG activities for all interested parties at Berkeley, and as an aid to recruitment. Pitman has expertise at obtaining experts able to set up a professional-quality site initially; we will then assign one RTG graduate student each semester to maintain the site. We are not particularly impressed by the quality of web sites of existing RTG groups, so will aim to make a model for future RTGs at other institutions.

The newly planned lecture write-ups from the Graduate Expository Seminar serve the goal of making material developed in the RTG at Berkeley available elsewhere. The summer workshops provide an interchange of ideas between the Berkeley group and other U.S. institutions. Travel support for graduate student researchers and postdoctoral associates will enable them to disseminate their work through talks at conferences and other universities. Results of undergraduate research can be disseminated in several journals devoted to undergraduate-only authors, and a newer journal (involvemath.org) devoted to research jointly by undergraduates and faculty.

7 Post-RTG plan

We will seek to maintain the structures put in place during the RTG – as much as possible, and to the extent they are deemed successful. The new structures (except for the summer workshop program) should be easy to continue. Without the RTG funding we will need to seek other sources of support for graduate student researchers, postdoctoral associates and summer activities, or to continue on a smaller scale. Demonstrated existence of a visible program in Probability involving undergraduates, graduate students, and post-doctoral fellows will help subsequent recruiting of target group members.

8 Research activities of senior personnel

David Aldous. Aldous is currently focused on three areas.

(i) Random spatial networks. One aspect is the study of different kinds of optimality property for networks over $n$ random points in two-dimensional space. Another aspect concerns the approximately scale-free property of routes in real road networks. We are studying existence and properties of continuum limit road networks, required to have unique routes between almost all point-pairs in the plane. Both aspects lead to many open questions suitable for future Ph.D. theses.

(ii) The notion of interacting particle systems as social dynamics can be axiomatized as follows. There are $n$ “agents”. Each agent has some “information” at time $t$. Each pair $(i,j)$ meets at random times of rate $\nu_{ij}$, at which times they can update their information according to some rule. So it’s a two-level modeling framework. The
bottom level is the “geometry”, specified by the rate matrix \((\nu_{ij})\), of how agents meet. The top level is the “content” of whatever phenomenon we are trying to model. Of course thousands of papers fit this framework. We are both studying new models of information updating, and seeking to study more systematically the effect of geometry, in the spirit of the study of Markov chain mixing times. Again, both aspects lead to many open questions suitable for future Ph.D. theses.

(iii) In connection with developing material for his “probability in the real world” course, Aldous is making efforts to involve undergraduates in research, resulting in papers with undergraduates being recently published in *Statistical Science* and *The American Statistician* and *Journal of Physics A*. He has also been faculty sponsor for undergraduate-initiated-and-taught courses on blackjack and on investment.

**Venkat Anantharam.** Anantharam’s current research interests are in the areas of information theory, communication networks, and coding theory.

1. **Information theory.** Specific areas of current interest include: (i) the role of common information in the creation of correlations across multiple agents in distributed systems. The extent of available correlations has a significant impact on the strategic decision making ability of systems of distributed agents. The cost of communicating randomness to create correlations is of particular interest. (ii) Also related to (i) is the design of schemes to provide information-theoretic security, in effect by extracting common randomness that provides a rate of one-time pads. The structure of “secure information” is being elucidated in this work. (iii) Progress on more traditional problems of multiuser information theory is also being made at a steady rate. The difficulty of these long-standing open problems lies in understanding the structure of “distributed information”. Much of this work is asymptotic in nature and has close ties with mainstream probabilistic developments related to critical values in interacting particle system models, concentration phenomena, correlation inequalities, etc.

2. **Communication networks.** Specific areas of current interest include: (i) the role of long-range-dependence in the performance and design of network algorithms to manage communication networks. (ii) economic aspects of network, in particular incentive issues. Probabilistic methods play a central role in much of this work.

3. **Coding theory.** Work on message passing algorithms in the context of low-density parity check codes is a significant current research interest over the last several years. This has very close ties to mainstream probability developments such as graph limits, Gibbs fields, random walks on trees, etc. This work has the happy property of being both intellectually challenging and bringing direct benefits to industry. For instance recent work of Anantharam and collaborators in this area led to the demonstration of a decoding chip for magnetic recording applications (with over ten million transistors) performing two orders of magnitude better than the existing state of the art, measured by metrics such as power consumption and throughput.

**Steven N. Evans.** Among the areas that Evans works in, the following are particularly relevant to the proposal because of the opportunities they have provided for collaboration with undergraduate students, graduate students and post-docs.

1. **Random matrices and probability on algebraic structures.** Two of Evans’ PhD
students graduated in 2010 with dissertations that contained joint work in this area. Partha Dey (now at Courant) and Evans used Stein’s method to show that the empirical distributions of the eigenvalues of classes of random matrices naturally associated with the representation theory of Coxeter groups converged to Gaussian when the dimension of the matrices went to infinity. Arnab Sen (now at Cambridge), Shankar Bhamidi (at that time a student of another PI, Aldous, and now at North Carolina) and Evans obtained very general results on the asymptotic empirical eigenvalue distributions for the adjacency matrices of a broad class of random trees. Evans is currently working with a PhD student, Ben Tsou, on the asymptotics of matrix entries of representations of the symmetric group with respect to a “uniformly chosen” basis. Evans also has a long-standing interest in probability on local fields and published a paper on notions of expectation and conditional expectation in this setting with an undergraduate, Tye Lidman (now at UCLA), in 2007. A reading course that Evans taught on the general theory of Markov processes resulted in a 2008 paper with Bhamidi, Sen, and two of his PhD students, Peter Ralph (now at UC Davis) and Ron Peled (now at Tel Aviv) on \( q \)-analogues of Brownian motion that is closely related to local fields.

(ii) Phylogenetics and phylogenetics based methods in metagenomics. Evans continues to work with a former post-doc, Erick Matsen (now at the Fred Hutchinson Cancer Research Center), on using ideas from optimal transportation and the theory of metrics on spaces of measures to analyze metagenomic data sets by encoding the collection of species in a metagenomic sample as a probability measure on a background phylogenetic tree. They have two papers on this topic under submission. Evans recently lectured on this work at a school for graduate students and post-docs at the National University of Singapore and also taught a graduate topics course on optimal transport at Berkeley in Spring 2011.

(iii) Random trees and tree-valued stochastic processes. Evans’ recent interest in this area grew out of joint projects with a post-doc, Anita Winter, one of the other PIs, Jim Pitman, and an undergraduate, Tye Lidman. A PhD student, Peter Ralph, and Evans co-authored a 2010 paper on most recent common ancestry in a class of conditioned continuous state branching models and a post-doc, Erick Matsen, and Evans co-authored a 2008 paper on the connection between genetic and genealogical ancestry in diploid Wright-Fisher models.

(iv) Other applications to biology. In 2011, Evans published work with a PhD student, Peter Ralph, and Alistair Boettiger, a PhD student in Molecular and Cell Biology whom Evans supervises as part of Berkeley’s Designated Emphasis in Computational and Genomic Biology, on models for transcriptional regulation and promoter proximal pausing. He also published two papers in 2010 and 2011 with a post-doc, Valerie Hower, and a Berkeley colleague, Lior Pachter, on shape-based peak identification for ChIP-Seq data. Evans has recently submitted a paper with two former students, Peter Ralph and Arnab Sen, and one of Ralph’s post-doc advisors, Sebastian Schreiber at the UC Davis Department of Evolution and Ecology, that uses stochastic differential equation models for population growth in spatially heterogeneous environments to shed light on a number of questions of current interest in conservation biology.
Elchanan Mossel. Mossel studies new combinatorial, probabilistic and analytic concepts, theorems and algorithms in order to analyze and establish new phenomena in stochastic models coming from social choice, biological networks and processes, and theoretical computer science. The interdisciplinary nature of his research exploits contributions from students and post-docs from different backgrounds as well as sophisticated ideas from probability, mathematics and computer science.

In combinatorial statistics, Mossel studies a rigorous approach to studying inference problems. This approach provides new algorithms for solving inference problems in the applied sciences and will further enhance the understanding of the applicability and performance of standard approaches for such problems. Further, new theoretical understanding of the underlying stochastic models are often derived.

Mossel is also currently interested in mathematical problems in the context of social choice. This research includes exciting mathematical directions with connections to analysis and Gaussian Geometry. Further, given the negative nature of classical results such as Arrow’s and the Gibbard-Satterthwaite theorem, the research provides exact classification of the best ways to deal with paradoxes and manipulation by finding the most rational or least manipulable functions. As such the work is closely related to current discussion regarding the optimal voting method.

James Pitman. Pitman’s interests range from mathematical problems at the interface of stochastic processes and random combinatorial structures, to practical problems of academic information management, such as author name disambiguation and design of bibliographic information services. A bridge between these interests is provided by probabilistic models for random partitions and related structures such as random graphs which serve as the basis of machine-learning algorithms for automated classification and entity recognition in the bibliographic universe. The importance of connecting the combinatorics of large graphs to organizing the world’s information is amply demonstrated by the development and industrial application of the page rank algorithm by Google. Much remains to be discovered about how best to structure and manage the deluge of academic information now available to researchers. It is to be expected that ideas from combinatorics, probability and machine learning will continue to find important applications in this endeavor. As an illustration, models for random discrete distributions and random partitions developed by Pitman and Yor in the 1990’s in the context of interval partitions generated by the zeros of Brownian motion and Bessel processes, are now finding extensive applications in machine learning algorithms for dealing with partitions found empirically to exhibit power law behavior, such as arise in citation and collaboration graphs in bibliometric analysis. Pitman has working on these topics with students of Mike Jordan from CS and of Tom Griffiths (Cognitive Science Lab) as well as students in mathematics and statistics. Pitman also maintains an interest in applications of partition-valued and tree-valued processes in biology.

Fraydoun Rezakhanlou. Rezakhanlou is currently focused on two areas.

(i) An important step towards understanding a dynamical system is the construction of its invariant measures because they describe the long time behavior of typical
orbits. Hamiltonian systems are those encountered in celestial mechanics and are used to model fluids and gases microscopically. An important class of invariant measures are constructed by solving associated Hamilton-Jacobi PDE. A classical example of a Hamiltonian is given by $H(x, t, p) = \frac{|p|^2}{2} + V(x, t)$ where $V$ is a spatially and temporarily stationary stochastic process. When $V$ is deterministic and periodic, it is possible to find solutions to the associated Hamilton-Jacobi PDE of the form $u(x, t) = p \cdot x + t\overline{H}(p) + w(x, t)$ with $w$ periodic in $(x, t)$. Such solutions can be used to obtain invariant measures for the corresponding Hamiltonian system. In the case of a stochastic potential, we wish to construct such solutions but now for $w$ which has ergodic and stationary partial derivatives. It is known that such solutions may not exist for every $p$ in the random setting if for example $V$ is independent of $t$. However if $V$ has time dependence and is “sufficiently” random, then we expect to have such solutions for every $p$. Rezakhanlou’s Ph.D student Dave Kaspar is currently investigating the existence and various properties of such solutions for stochastic Hamilton-Jacobi PDEs. The graph of the function $u$ can be regarded as a random interface separating different phases and for certain potential $V$ should belong to Kardar-Parisi-Zhang universality class. This motivates various open questions that are suitable for Ph.D projects.

(ii) Rezakhanlou along with Alan Hammond defined and investigated a stochastic model of coagulating Brownian particles. Microscopically, particles have positions, masses and radii. Each particle travels as a Brownian motion with diffusion coefficient $d(m)$ where $m$ denotes the mass of the particle. Particles may coagulate only when they are sufficiently close. For example the coagulation occurs between particles of positions $x$ and $x'$ only when $|x - x'|$ is of order $\epsilon(r + r')$ where $r$ and $r'$ are the radii of particles and $\epsilon$ is a small parameter. When the dimension $d$ is 3 or more, the initial number of particles is of order $O(N)$ with $N = \epsilon^{2-d}$. An important feature of this model is that giant particles may emerge in finite time i.e. gelation occurs. Indeed, if the relationship between the mass $m$ of a particle and its radius $r$ is given by $r = m^\chi$, then for a gelation we need a condition of the form $\chi > (d-2)^{-1}$ where $d$ is the spatial dimension. If this condition is satisfied we conjecture that an instantaneous gelation would occur. This conjecture and some related problems are suitable for future Ph.D dissertations.

**Alistair Sinclair.** Alistair Sinclair is a theoretical Computer Scientist with a strong interest in interdisciplinary research that explores connections between probability and statistical physics on the one hand, and algorithms and computational complexity on the other. His recent and current research, in many cases in collaboration with graduate students and postdocs, includes the following: random graph models for mobile wireless networks (with connections to continuum percolation); Markov chain Monte Carlo algorithms in combinatorics, statistical physics and optimization; computational aspects of the Boltzmann equation; applications of randomized algorithms in algebra, including computation of permanents, design of efficient error-correcting codes, and matrix balancing.

More broadly, Sinclair has been involved for several years in initiatives that seek to apply the “computational lens” to other sciences, including biology, physics, nan-
otechnology, economics, and neuroscience. He has co-organized three workshops on this general theme (Berkeley, 2001; Princeton, 2006; Caltech, 2007), been the main PI on a major NSF-funded grant on this topic at Berkeley (2001-06), and co-wrote an NSF-sponsored report on the computational lens based on the second and third of the above workshops.

Allan Sly. (i) Mixing times of Markov chains. To better understanding phase transitions in Markov chains, particularly for spin-systems. Another aspect relates to developing new techniques for proving cutoff on Markov chains of weakly interacting particles and other models without a group structure. The area is closely related to questions in computational complexity, particularly approximate counting problems which in turn are related to phase transitions from statistical mechanics. This area contains many interesting problems which could be the basis of PhD theses.

(ii) Phylogenetic reconstruction - the study of reconstructing the ancestral tree of a collection species from their genetic information. Biologists use a range of algorithms, particularly maximum likelihood, Bayesian MCMC and parsimony, which often perform well in practice but lack a firm mathematical understanding. We wish to give sharp quantitative probabilistic estimates on how much data is required for successful reconstruction by these estimators. This involves numerous possible projects suitable for PhD students and postdocs across a spectrum of more theoretical and to more applied work.

9 Recap: intellectual merit and broader impacts

Being a RTG proposal we will not explicitly address the intellectual merits of the specific mathematical research projects being conducted by involved faculty and outlined in the previous section – this is demonstrated by the continuing record of publication in major journals. We expect most postdoctoral associates, and a small majority of graduate trainees, to continue into academic careers, and contribute to the vibrant and broad research activities surrounding mathematical probability. We expect the other graduate trainees to segue into careers which make substantial use of their mathematical training. The effect on undergraduate participants is perhaps less measurable, but the overall intention is to encourage them toward more advanced training than they might otherwise have sought.

To recap some of the broader impacts associated with the project:
Graduate trainees and postdoctoral associates will be more involved in teaching their own subject, and engaging in supervising undergraduate research projects.
The undergraduate seminar, activity-based courses and undergraduate research projects will combine to encourage students to consider more advanced training.
A minor emphasis on training participants at all levels to do expository writing, beneficial to both participants and the academic community.