

Research Statement

Pegah Kamousi
pegah@cs.ucsb.edu

Department of Computer Science
UC Santa Barbara, CA 93106

My research is inspired by uncertainty, graphs, and geometry. From the tiny neurons in our brains to massive earthquakes, randomness is an essential part of our world's unfolding. Most natural and social systems are highly unpredictable, and man-made systems are often hard to manage: devices fail and bonds break unexpectedly, noise is inevitable, and a myriad of confounding variables can only be accounted for through statistical means. Without taking such uncertainties into account, the scientific quest to fully understand, explain, and predict our world is doomed to fail.

In computer science and mathematical optimization, a number of models and approaches have been considered to account for and analyze the effect of uncertainty: random graphs, stochastic geometry, Bayesian analysis, and multi-stage stochastic optimization being some prominent examples. Most of the classical random models focus on *asymptotic* results on random point sets or random graphs, while much of the current research on stochastic optimization is focused on two-stage models, where one trades the current uncertainty with future price inflation. In short, most of these models and techniques are either too abstract or too application-specific, and there remains a need for basic, general models and algorithms that can be adapted to a broader class of applications. My research has strived to bring us closer to this goal.

1 Uncertainty in Graph theory and Computational Geometry

Many systems, whether physical or abstract, natural or man-made, are described in terms of graphical or geometric models. There is a plethora of well-studied algorithms defined over those models, most of which assume *deterministic*, well defined inputs. This assumption, however, is not realistic for many real-world applications, where uncertainty is either inevitable due to the unreliability of the elements, or needs to be introduced to account for the confounding variables.

Consider a simple uncertainty model on a graph $G = (V, E)$, where each node $v_i \in V$, is only present with an *independent* probability p_i . When a node is absent, all the edges incident to that node disappear as well. The independent node probabilities induce a sample space Ω of 2^n outcomes, corresponding to all the subsets of the nodes, where each subset appears with a certain probability. This model could simply describe a network whose elements may fail individually, and become detached from the rest of the network. Similarly, imagine a set of *stochastic* points in a geometric space, each only present with an independent probability. How can we think about the value of various geometric functions over such a stochastic set?

In the absence of such uncertainties, there are numerous functions and structures defined on a graph: shortest paths, minimum spanning trees, matchings, etc. However, the moment uncertainties come into play, those structures are no longer well-defined: how could we build a minimum spanning tree for a graph, when we do not know which nodes are present? In my research, I considered a problem which *is* well defined in such settings: computing the *expected* value of the functions, or the *expected* weight of different structures, where the expectation is over all the

exponential number of subsets.

As appears so, the introduction of probabilities brings along a combinatorial explosion in the complexity of problems. This introduces a new twist in classical graph theory and computational geometry: the challenge to deal with uncertainties, in spite of the overt explosion in the complexity. My research has taken up this challenge, leading to surprising results, open questions, and new ways to think about many classical problems. Bellow I mention some of my results.

Given a set M of points in a Euclidean space, consider drawing a random subset, which includes each point $m_i \in M$ with probability p_i . We studied the expected weight of the minimum spanning tree defined on the random subsets. While showing that this problem is intractable, we were able to devise randomized and deterministic algorithms to approximate the expected weight, using techniques from probability theory, computational geometry and graph theory. This problem has applications in sensor networks, vehicle routing, robotics, social networks, etc.

One of the fundamental geometric problems, which arises as the building block of numerous algorithms and has numerous applications in biological systems, data mining, computer vision, etc., is to determine the *closest pair* between a set of points. We studied the stochastic version of this problem, namely, when each point is only present with a given probability. We established the hardness of problem in general, but were able to draw conditions on the metric and configurations of the points, under which the problem becomes polynomially solvable. In the same setting, we also considered the stochastic version of the Nearest Neighbor Search problem, and designed a linear space data structure to preprocess the points, such that given a query point, we can compute its *expected* distance to the closest neighbor in logarithmic time.

As problem with a slightly different flavor, I considered the *capacity* of random strings, as part of a research on genome sequencing, in collaboration with the data mining group at UCSB. Given a random string, each of whose indices probabilistically assumes a value from an ordered alphabet (A, C, T, or G, as in the DNA nucleotides), what is the distribution of the (canonically) smallest substring of a given size? I developed and implemented a dynamic programming algorithm to compute this distribution.

Currently I am working on a stochastic version of the Traveling Salesman problem, inspired by routing and navigation in noisy environments, which is a major line of research in robotics. Given a set of (potentially intersecting) disks, whose radii are randomly drawn from a Gaussian distribution, the goal is to find an a priori order on the disk that minimizes the expected cost of the tour visiting all the disks. This problem combines the essence of online algorithms with a natural uncertainty model.

One of the interesting results that leaves behind a fundamental questions, is the surprisingly different behavior of seemingly similar structures under uncertainty. We were able to devise polynomial algorithms to compute the expected weight of many structures such as Delaunay triangulation, and several graphs such as the Gabriel graphs. On the other hand, a similar structure such as the minimum spanning tree proves to be intractable. Is it possible to characterize the structural properties that lead to such different behavior? Could such results be used in the design of systems that are more amenable to efficient calculations in the presence of uncertainties?

2 Research Plan

The purpose of my research has been to develop stochasticity models and algorithms that

1. are sufficiently basic and general that can be applied to a plethora of applications such

as robotics, biological systems, machine learning, social networks, computer vision, pattern recognition and many other fields,

2. can address the uncertainty whether it arises from the structure, behavior or inter-connectivity of the system, i.e., capture the unreliability of individual elements or inputs, as well as the stochastic nature of the relations between them.

Inspired by real world applications, a myriad of problems in graph theory and computational geometry need to be revisited under simple stochasticity models. This motivates a new line of research and demands new techniques to deal with uncertainty. My belief is that my algorithms and techniques can serve as building blocks for more complex, stochasticity-aware algorithms.

Apart from the theoretical side of my research, I am also very interested in real world systems where uncertainty is inherently present, or needs to be introduced. Fortunately, discussions and contributions with researchers from more applied fields such as computational biology, data mining, network modeling and robotics, has revealed great potentials for fruitful application of my research in such fields. I am specially interested in applying my techniques and algorithms in natural and biological systems, cognitive sciences and artificial intelligence.