Research Statement

My research aims to answer the most pressing problems of “Big data” by combining physical laws and machine learning.

Today, there are two pressing issues yet unresolved in Big-data i.e.
- How to deal with the large amounts of noise in the data set?
- How to deal with feedback within the processes that generate the data?

I have developed new techniques to address both of these issues that deliver faster and more reliable algorithms to uncover the underlying patterns in the data. The ‘secret sauce’ in these techniques is inspired by physical laws and processes like diffusion, Gauss’ law, laws of conversation and heat flow applied to graphs.

To test the robustness, I have implemented these techniques to a wide variety of domains by representing data points and their pairwise relationships as a graph structure without rebuilding them for each domain. These techniques provide superior results in pattern recognition objectives such as image segmentation in computer vision, accurately retrieve similar shapes to a query from a large database of geometric mesh models, prevent false positives in recommender systems and identify the hierarchical and overlapping organization of networks.

I developed these techniques over the course of my broad educational experience
- I initially studied complex interconnected modules in a multiscale system [1]. I investigated the resilience of the multiscale system’s output to different kinds of uncertainty in the input parameters, modeling errors in simulation modules and asynchronous interactions between distributed designers [2, 3]. I developed a framework to optimize systems’ target under uncertainty and resource constraints, using ideas from reinforcement learning [4]. This framework was used to design an underwater submersible at the I.I.T. Kharagpur.
- Subsequently, I developed techniques for shape matching, segmentation and search. I realized an intricate connection between the Laplace-Beltrami operator governing the diffusion equation and the transition matrix of a random walk that went well beyond what was known before it [5]. This connection was used to construct multiscale shape signatures with several desirable properties over the heat kernel signature.
- Next, I established the connection between the Laplacian of a general graph and the flux of a potential field over a boundary. This insight is exactly an extension of Gauss’ law for charge distributions applied to discrete graphs. Consequently, I was able to design a novel method for data clustering using this law in conjunction with laws of conservation and spectral graph theory. My method to identify community boundaries is significantly more robust to noise than traditional community detection of strongly knit modules [6].
- I designed a new method to separate feedback effects in recommender systems (RS) such as Netflix from intrinsic preferences of users. This method was based on a simple deconvolution of an iterative feedback loop common in physical systems that admit feedback. I was able to identify items recommended to a user by the RS by simply inputting the user-item ratings matrix to my algorithm [7].
- Further, I along with my colleagues used a collaborative filtering algorithm ubiquitous in recommender systems, but lesser known to the computer vision community, to develop a hand tracking system. We leveraged connections between point cloud streams from a depth camera and personalized recommendations from a RS to develop a real time algorithm that resulted in a transformative leap above the current state of the art [9].
I am currently focusing on real-time analysis of data using machine learning models inspired by biological neural networks, or deep learning. I constructed a deep model which uses the parametric knowledge of point clouds from a large database to parametrize unknown and incomplete point clouds. Preliminary experiments on depth data from 3D sensors have shown great promise [8]. Furthermore, I am developing a technique to create a geometric image from a point cloud, so that standard convolutional neural networks can be used for 3D mesh analysis [10]. This bridges the gap between traditional geometric mesh processing and image processing techniques.

My goal during my post-doctoral research is to delve deeper into physically motivated algorithms for data mining and pattern recognition in graphs. I aim to couple machine learning with physical laws to reveal patterns that govern communities and society at large. As the underlying motivation of my methods stem from the physical world around us, the resulting algorithms will naturally be simple to implement and easy to understand.

I wish to discover knowledge inherent in big data by using physical and geometric constructs such as:

- Poisson’s equation for network link prediction in analogy with Poisson surface reconstruction and mesh processing.
- Quantum random walks for overlapping cluster detection, inspired by mixed quantum states.
- Diffusion on dynamic networks to reveal patterns over time. I will examine the connections between the time to equilibrium for a random walk and the temporal evolution of a network.
- Resistor-capacitor-inductor circuits for network analysis using Laplace transforms. I will extend known relationships between the Kirchhoff/resistance matrix and the Laplacian to general electrical networks using concepts of circuit theory.
- Partial differential equations beyond the diffusion equation for understanding big data. I intend to use my knowledge of fluid mechanics and thermodynamics gained during my undergraduate and graduate studies in mechanical engineering in the domain of data mining.
- Additionally, I will continue my current work on Gauss’ law for networks by exploring more advanced concepts related to divergence, gradient and Laplacian operators on networks.

In the long-term, it is my strongest desire to enter academia and pursue my research objective of melding physical laws and machine learning. Over time, I am confident my unique approach to data mining will reveal new insights from big data.

References
[6] Author of article “Finding Communities by Averaging Node Similarity in Networks”, submitted to Nature Physics. Reviewed for publication at Science magazine (~15% of all papers)