Research Statement

Sal Aguinaga
University of Notre Dame
Department of Computer Science and Engineering
saguinag@nd.edu

1 Overview

My research interests are data mining with a focus on large-scale networks. I want to study and understand the patterns that make up the underlying structure of the worldâĂŹs data. My long-term research goals include (1) understanding the major factors that influence network formation; (2) examining the mechanism that contributes to the emergence of interesting and useful patterns in such networks, and (3) developing useful algorithmic solutions to problems in neuroscience and medicine.

Academic Preparation. A wide range of experiences has brought me to this path in computer science. First, I was an electrical engineer for many years working in the electronics industry and transitioned to neuroscience research where I could apply my knowledge of engineering in a biological model. This led to working on problems that could be explored through the lens of electrical engineering. For example, I helped automate electrical recordings of cell membrane capacitance measurements. Later, I went on to work on a research project where I led a team of students and software engineers on the development of the hardware and software for audiology research and its deployment in clinical settings as well. This experience sparked my interest in formal training in computer science (CS). During my CS Ph.D. program, I focused on courses that such as data mining, information retrieval, network science, and others that prepared me with a wealth of knowledge to be able to tackle problems from a variety of disciplines. Using applied mathematics and computer science to explore the structure and dynamical behavior of networks is an exciting way to engage the world's data and its associated challenges.

Research Achievements. My work and ideas have contributed to several publications across diverse disciplines such as chemistry, neuroscience, mobile computing, along with patents in telecommunication system design. My current focus is at the intersection of natural language processing, complex network analysis, graph mining, and machine learning. Discussing this work at local workshop (MBDOC) led to a collaboration with an Argonne National Lab scientist, and we have submitted a project proposal to The Office of Science Graduate Student Research Program for funding to automatically find neural network architectures for deep learning tasks and speed up their training phase. While conducting research in mobile computing, I established a collaboration with Rehabilitation Institute of Chicago scientists to use mobile phone sensors to study early signs of Parkinson's disease to study longitudinal signatures of tremor. This collaboration brought new projects to our lab such as a study of sensor data monitoring to determine the conditions around patient falls after being released from the hospital.

Vision for Future Research. The trajectory of my work is to investigate information networks along three dimensions: (1) study the computational behavior of neural network architectures for deep learning tasks; (2) explore algorithmic solutions to describe the forces or the mechanism underlying dynamic behavior in networks; (3) apply graph mining and machine learning methods to problems in sensory neuroscience.

2 Recent Research

My dissertation research explores hyperedge replacement graphs, a formalism in natural language processing and compiler design, combined with graph theory to build a new way to infer models of the topological structure of networks. My major works are summarized below, which have already been published and presented in top conferences or are slated to be published in top journals such as IEEE Transactions on Pattern Analysis and Machine Intelligence.

Language Theory Meets Graph Theory. This work extracts a hyperedge replacement grammar from a network's clique tree. These grammars guarantee reconstruction of isomorphic copies of the original graph, stochastic approximations, and extrapolations above and below the original network's order and size. Generated graphs have many of same properties as the original. We validate our approach using network analysis and compare our results to those of other popular and well documented graph generation models and find that our approach can outperform existing methods such as Chung-Lu, Kronecker product, and exponential random graph models along various networks metrics [2].

Infinity Mirror test for graph generators This test assesses how robust artificially created graphs, generated using models inferred from real-world networks, are at holding sufficient structure. This test aims to relearn model parameters to generate new synthetic networks and examine whether or not a graph generator endows the generated graphs with the essential micro-structure needed to stand-in in place of the original graph. This test was developed to further validate or stress test generative network graph models such as hyperedge replacement grammar showing that topological structure of the generated networks holds well outperforming other generators (including the Block Two-Level Erdős-Rényi model) [3].

Human Navigation in Complex Knowledge Networks A proxy for how humans leverage knowledge to navigate complex information networks is the way we organize knowledge. For example, Wikipedia is described as a repository of the world's knowledge. This knowledge is organized by concepts such as science, society, mathematics, but also categorizes more concrete objects like Neil deGrasse Tyson (person, American, cosmologist, etc.). How do we find our way from one page to another, when nodes are unrelated, using only the links on the starting page? We draw on our knowledge of the world around us for connections that allow us to hop conceptually closer to a given target. We explored this idea around way-finding problems in knowledge networks to gain insights into how we can develop better data systems [1].

3 Future Research

Data is ubiquitous; we are practically drowning in it. Making sense of this data and, hence, our world is the challenge of the day. I intend to be on the forefront of this task.

Engineering Deep Neural Networks. In collaboration with computer scientists at the Argonne National Laboratory, I am actively working on applying hyperedge (or hypernode) replacement grammars to address an open problem in deep neural netowrks. Training deep neural networks is difficult and continues to be an open problem in machine learning. We are interested in methods that can automatically discover high-performing network architectures that are as good or better than hand-crafted by engineering various hyperparameters such as edge-weights. The objective is to speed up the training phase for deep learning tasks in high-performance, scientific computing applications at Argonne.

Evolving Patterns in Large-Scale Data Systems. What "forces", natural or artificial, influence the way networks grow? Real-world networks are dynamic and the mechanism that influences the process of how nodes join and contribute to the shape of the network may lay along dimensions different than popularity or similarity that are yet to be discovered. I want to study the role time plays on the local and global properties of a network. An application example is the study of Web repositories, such as GitHub, where we can examine the evolution of concepts like scientific equations in coding routines or even ideas of social movements that can have profound political effects.

The Connected Transcriptome of Inner and Outer Hair Cells. Sensory receptor cells critical for hearing in the mammalian inner-ear have different morphology and function. Gaps in our knowledge of the genetic mechanism that controls or defines their morphology continue. Genes that are being actively expressed in a cell might hold the key to understanding that mechanism and network of biological properties of a cell. I would like to build on my previous neuroscience research of the mammalian cochlear hair cells by exploring the array of mRNA transcripts (transcriptome) produced in these cells. From a computer science stand point, I will be using network analysis to build network representations of the expression values of about 40,000 genes in four different cell types in the organ of Corti. We will then use network alignment to examine cell similarity (i.e., compare cells) in terms of their trasncriptome signatures. This network of molecules is also a great biological system to study network dynamics, because the transcriptome actively changes.

References

[1] Salvador Aguinaga, Aditya Nambiar, Zuozhu Liu, and Tim Weninger. "Concept hierarchies and human navigation". In: *Big Data (Big Data), 2015 IEEE International Conference on.* IEEE. 2015, 38–45.

[2] Salvador Aguinaga, Rodrigo Palacios, David Chiang, and Tim Weninger. "Growing Graphs from Hyperedge Replacement Grammars". In: *CIKM.* ACM. 2016.

[3] Salvador Aguinaga and Tim Weninger. "The Infinity Mirror Test for Analyzing the Robustness of Graph Generators". In: *KDD Workshop on Mining and Learning with Graphs.* ACM. 2016.