A STOCHASTIC PROCESS WITH A LAPLACIAN CONNECTION

IQRA ALTAF GILLANI

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI

MARCH 2021

A STOCHASTIC PROCESS WITH A LAPLACIAN CONNECTION

by

IQRA ALTAF GILLANI

Department of Computer Science and Engineering

Submitted

in fulfillment of the requirements of the degree of

Doctor of Philosophy

to the

Indian Institute of Technology Delhi

March 2021

'It's your road and yours alone.
Others may walk it with you, but no one can walk it for you.'
- Rumi

To
Papa and Mama
Thanks to whose endless support, patience, and faith
I could walk this road.

Certificate

This is to certify that the thesis titled **A Stochastic Process with a Laplacian Connection** being submitted by **Iqra Altaf Gillani** for the award of **Doctor of Philosophy** in **Computer Science and Engineering** is a record of bona fide work carried out by her under my guidance and supervision at the Department of Computer Science and Engineering, Indian Institute of Technology Delhi. The work presented in this thesis has not been submitted elsewhere, either in part or full, for the award of any other degree or diploma.

Amitabha Bagchi
Professor
Department of Computer Science and Engineering
Indian Institute of Technology Delhi
New Delhi - 110016

Acknowledgements

All praises are due to Allah, the Almighty, on whom ultimately we depend for sustenance and guidance. All thanks are due to him for letting me hold onto *Tawakkul* (trust in the Almighty's plan) that gave me the strength to persevere till the very end.

Every Ph.D. is a different ride with its unique track of highs and lows. However, one thing that is common in all is that it can never be a comfortable one without co-passengers. Those who may not directly help you in your drive, but their mere presence gives you a sense of security and motivation to continue. I want to acknowledge all those who made my Ph.D. ride not only comfortable but a memorable one too.

First of all, I would like to thank my advisor, Prof. Amitabha Bagchi, for his invaluable guidance and support. Right from the beginning, he whole-heartedly supported me even when I decided to switch to the part-time mode despite being reluctant about it initially, and time and again, he helped me to make this mode successful. Be it our online discussions when I was in Srinagar or our rigorous offline discussions on the whiteboard when I was back in Delhi, he made it a point that the continuity and quality of research were not compromised because of our long-distance communication. His words of encouragement and unwavering faith have helped me during the most testing times of my Ph.D. There are lots of things for which I owe him thanks. To summarize them all, thank you Sir, for being a *great mentor*.

The initial and the final years of the Ph.D. are the most challenging when you start discovering your problem domain and then finally when you wind up your findings. I have been fortunate enough to have the guidance of Dr. Pooja Vyavahare during the first year of my Ph.D. and Prof. Sayan Ranu during the final year of this journey. I thank both of them for their feedback and valuable discussions. I would also like to thank the staff of the CSE department IIT Delhi for their help in several administrative matters.

My winter and summer trips to IIT Delhi have been the most productive phases of my Ph.D. and these would have been impossible without my friends' support. I thank all my labmates over these years for creating a wonderful lab environment and for helping me out whenever I needed their guidance. I would also like to thank the entire Ijtema group for those soul-refreshing after-dinner sessions that drained away all kinds of stress. Thank you for also hosting me whole-heartedly every time I came to Delhi and had nowhere to go. Special thanks to my fellow Trainee teachers. It would have been impossible for me to survive this scheme until the end without your company.

Even imagining the thought of embarking on a journey like a Ph.D. without family's support is difficult. I thank my entire extended family for their support and patience over these years. Every time I left for IIT Delhi, my grandmothers would ask me, 'When will your studies end? Will this be your last trip?'. Thank you for your prayers and blessings. Finally, I'm near the end of it. I would also like to thank my siblings for always being there, just a call away whenever I needed to confide in them. Their role in this journey is more than they can realize. Special thanks to my Mamus (maternal uncles) for understanding me and bearing with my sudden schedule changes. I have spoiled a lot of our pre-planned trips because of my work. I owe those trips to my siblings and you.

Finally, two people without whom this work would never materialize, to whom I dedicate this thesis: My father and my mother. Thank you for the unconditional support, encouragement, and freedom you gave me at every phase of my life. I cannot possibly finish the list of things I want to thank you for. Everything I know or I have achieved in my life, it is all because of your efforts. Thank you for everything.

Iqra Altaf Gillani

Abstract

In this thesis we study a discrete-time stochastic process modeling the collection of data in a multi-hop sensor network that we call the *Data Collection Process*. We assume data is generated at the sensor nodes as a discrete-time Bernoulli process. All nodes in the network maintain a queue and relay data, which is finally collected by a designated sink. We prove that the resulting multi-dimensional Markov chain representing the queue size of nodes has two behavior regimes depending on the data generation rate. In particular, we show a non-trivial critical value of data rate below which the chain is ergodic and converges geometrically to a stationary distribution and above which it is non-ergodic, i.e., the queues at the nodes grow in an unbounded manner.

We then extend the Data Collection Process to a multi-commodity setting to provide a family of bounds for the rate at which the functions of many inputs can be computed in-network on general topologies. Going beyond simple symmetric functions where the output is invariant to the permutation of the operands, e.g., average, parity, we describe an algorithm that is analyzed to provide throughput lower bounds for the general functions, particularly the ones with a binary tree schema. Our lower bounds depend on the underlying graph and schema parameters, and we show that these are optimal for the complete network topology, the star topology, and the hypercube topology.

We also present an interesting connection between the Data Collection Process and Laplacian systems of equations. In particular, we show that the steady-state equation of the stochastic process is equivalent to the special class of *one-sink Laplacian systems* of the

form $L\boldsymbol{x} = \boldsymbol{b}$ where exactly one of the coordinates of $\boldsymbol{b}$ is negative. As an application of this connection, we first present a completely new approach to solve Laplacian systems using queueing networks, which marks a significant departure from the existing techniques mostly based on graph-theoretic constructions and sampling. Our Laplacian solver can be used to adapt the approach by Kelner and Mądry (2009) to give the first distributed algorithm to compute approximate random spanning trees efficiently. We also extend our algorithm for solving the Laplacian system of equations on strongly connected directed graphs. This is the first distributed algorithm for this problem that can be analyzed in terms of the underlying graph's parameters.

As a second application of the Laplacian connection of the Data Collection Process, we propose a group-to-group version of random walk betweenness centrality. In particular, we relate our proposed metric to the steady-state equation of the Data Collection Process, which allows us to use a Laplacian solver to compute it efficiently. Empirical evaluation on real-world networks proves that our proposed metric is more effective than all natural baselines for controlling the spread of infection from an infected group to a vulnerable group.

सार

इस थीसिस में हम एक मल्टी-हॉप सेंसर नेटवर्क में डेटा के संग्रह को मॉडलिंग करने वाले एक असतत समय स्टोकेस्टिक प्रक्रिया का अध्ययन करते हैं जिसे हम डेटा संग्रह प्रक्रिया कहते हैं। हम मानते हैं कि डेटा संवेदक नोड्स में असतत-समय बर्नौली प्रक्रिया के रूप में उत्पन्न होता है। नेटवर्क में सभी नोड्स एक कतार और रिले डेटा बनाए रखते हैं, जो अंत में एक निर्दिष्ट सिंक द्वारा एकत्र किया जाता है। हम यह साबित करते हैं कि परिणामी बहु-आयामी मार्कोव श्रृंखला जो नोड्स की कतार के आकार का प्रतिनिधित्व करती है, डेटा उत्पादन दर के आधार पर दो व्यवहार नियम हैं। विशेष रूप से, हम डेटा दर का एक गैर-तुच्छ महत्वपूर्ण मान दिखाते हैं जिसके नीचे श्रृंखला एर्गोडिक है और ज्यामितीय रूप से एक स्थिर वितरण में परिवर्तित होती है और जिसके ऊपर यह गैर-एर्गोडिक है, अर्थात, नोड्स पर कतार एक अनबिके तरीके से बढ़ती हैं।

हम तब डेटा संग्रह प्रक्रिया को एक बहु-वस्तु सेटिंग के लिए विस्तारित करते हैं, जिसमें दर के लिए एक सीमा प्रदान की जाती है, जिस पर कई इनपुट के कार्यों को सामान्य टोपोलॉजी पर नेटवर्क में गणना की जा सकती है। साधारण सममितीय कार्यों से परे जाकर जहां आउटपुट ऑपरेंड्स के क्रमचय के लिए अपरिवर्तनीय है, उदाहरण के लिए, औसत, समता, हम एक एल्गोरिथ्म का वर्णन करते हैं जो सामान्य कार्यों के लिए थ्रूपुट निचले सीमा प्रदान करने के लिए विश्लेषण किया जाता है, विशेष रूप से एक द्विआधारी पेड़ स्कीमा के साथ। हमारे निचले सीमा अंतर्निहित ग्राफ और स्कीमा मापदंडों पर निर्भर करते हैं, और हम बताते हैं कि ये पूर्ण नेटवर्क टोपोलॉजी, स्टार टोपोलॉजी और हाइपरक्यूब टोपोलॉजी के लिए इष्टतम हैं।

हम समीकरणों के डेटा संग्रह प्रक्रिया और लाप्लासियन प्रणालियों के बीच एक दिलचस्प संबंध भी प्रस्तुत करते हैं। विशेष रूप से, हम दिखाते हैं कि स्टोकेस्टिक प्रक्रिया का स्थिर-राज्य समीकरण $Lx = b$ फॉर्म के वन-सिंक लाप्लासियन सिस्टम के विशेष वर्ग के बराबर है, जहां ठीक एक निर्देशांक है b नकारात्मक है। इस संबंध के एक अनुप्रयोग के रूप में, हम पहले कतारबद्ध नेटवर्क का उपयोग करके लाप्लासियन सिस्टम को हल करने के लिए एक पूरी तरह से नया दृष्टिकोण पेश करते हैं, जो मौजूदा तकनीकों से महत्वपूर्ण रूप से प्रस्थान करता है जो ज्यादातर ग्राफ-सिद्धांत निर्माण और नमूनाकरण पर आधारित है। हमारे लाप्लासियन सॉल्वर का उपयोग केल्नर और मैड्री (2009) द्वारा दृष्टिकोण को अनुकूलित करने के लिए किया जा सकता है, जो पहले वितरित एल्गोरिथम को अनुमानित यादृच्छिक फैले हुए पेड़ों की कुशलता से गणना करने के लिए देता है। हम अपने एल्गोरिथ्म का विस्तार दृढ़ता से जुड़े निर्देशित ग्राफ़ पर समीकरणों के लाप्लासियन सिस्टम को हल करने के लिए भी करते हैं। इस समस्या के लिए यह पहला वितरित एल्गोरिथम है जिसका विश्लेषण अंतर्निहित ग्राफ़ के मापदंडों के संदर्भ में किया जा सकता है।

डेटा संग्रह प्रक्रिया के लाप्लासियन कनेक्शन के एक दूसरे अनुप्रयोग के रूप में, हम यादृच्छिक वॉक केंद्रीयता के एक समूह-

से-समूह संस्करण का प्रस्ताव करते हैं। विशेष रूप से, हम अपने प्रस्तावित मीट्रिक को डेटा संग्रह प्रक्रिया के स्थिर-राज्य समीकरण से संबंधित करते हैं, जो हमें कुशलतापूर्वक गणना करने के लिए एक लाप्लासियन सॉल्वर का उपयोग करने की अनुमति देता है। वास्तविक दुनिया के नेटवर्क पर अनुभवजन्य मूल्यांकन यह साबित करता है कि हमारे प्रस्तावित मीट्रिक एक संक्रमित समूह से एक कमजोर समूह में संक्रमण के प्रसार को नियंत्रित करने के लिए सभी प्राकृतिक आधारभूत से अधिक प्रभावी है।

Contents

List of Figures

List of Tables