HoangNguyen
Full name: Nguyen Thai Hoang

gear.github.io/aboutme/
github.com/gear/
hoangnt.titech@gmail.com
Room 816, 2-2-B Aomi, Koto-ku, Tokyo 135-0064, Japan

Statement of Objectives

To Ecolé Polytechnique Fédérale De Lausanne Ph.D. Admission Committee,

My name is Hoang Nguyen. Currently, I am a second year Master student at Tokyo Institute of Technology. My life passion is to pursue a research career in the Computer Science field, specialized in Discrete Optimization and Machine Learning. Because EPFL's Ph.D. Program appeals to me in many aspects, I have decided to apply. I believe having a Ph.D. degree from EPFL will be a significant boost for my academic career.

Throughout my academic journey until now, I have always been regarded as gifted in technological subjects among peers and teachers. Prior to higher education, I was admitted to the High School for Gifted Students (Hanoi National University of Education) specializing in Physics. I took part in multiple high school level Physics competitions. After high school, I passed the entrance examination to Hanoi University of Science and Technology, which is one of the best, if not the best, technological university in Vietnam. In university, I found computer architecture and integrated circuit design quite similar to playing LEGO. I enjoyed working on a 50MHz FPGA prototype board on which I can run a cloned version of Bomber-man game smoothly. I was (and still am) excited about hardware implementation of various algorithms and software. Besides playing with circuits in my lab, I taught science subjects to high school students of Hanoi International School as a part-time job. During this time period, online courses provided by MIT or Coursera became popular in my circle of friends. Among all those amazing free online courses, I was deeply inspired by the Linear Algebra course taught by professor Gilbert Strang at MIT. His online lectures not only reinforced my linear algebra knowledge but also showed me how to teach. Professor Strang has become my personal hero ever since. I am pursuing the academic career because one day I want to be able to teach and inspire students as he does.

For the last two years studying as a Master student in Japan, I have always wondered about how we can make sense of network-structured data and learn patterns from it. Furthermore, if we have a large network and collecting data from each individual node is not feasible, how can we approximate the network status and to what extent is our approximation reliable? I think the problem above can be split into two sub-problems: learning patterns of a network-structured data and summarizing a network-structured data. In my Master studies, I study about practical techniques to address the first problem. I have been designing an algorithm that returns a network nodes classifier given the network structure and partial ground-truth labelings. Analyzing the existing methods, I noticed some labeling errors come from the fact that existing algorithms failed to capture structural equivalency between different group of nodes. Based on this observation and an intuition about the importance of network motifs, I proposed two methods, in which network motif representations are injected into the classifier training process.

My motif injection method (also called Motif-Aware Graph Embedding) improves state-of-the-art methods by 2-3 percentage points. Through this project, I have obtained some practical understanding of complex networks, information processing on graphs, and deep neural networks. However, my current methods lack theoretical explanation, which will be one of the subjects for my future work.

As mentioned in the previous paragraph, the second sub-problem is to summarize a network with respect to some constraints. Regarding summarization problems, one promising approach is to employ diversity selection models such as the Determinantal Point Processes or Probabilistic Submodularity models. I am particularly interested in the submodular approach because of the fact that it addresses the diminishing return property and it has solid theoretical guarantees. While the current literature of submodularity focuses on greedy processes with Knapsack constraints or matroid constraints, it is insufficient to address our problem here while we need to summarize a large network with respect to both computational resources and network structural constraints. With regard to submodular maximization on graphs, Jan Vondrák has addressed the optimization problems on random subgraphs. Towards my Ph.D., I plan to develop a unified method for submodularity maximization on real-world networks (or graphs) with respect to a certain information propagation process defined on the networks.

I decided to apply to the LIONS research group at EPFL because their research aligns very well with my interests. In addition to the work on streaming algorithms for submodular coverage and Bayesian optimization methods, LIONS also has collaborations with many leading research groups working on submodular maximization such as LAS at ETHZ. Furthermore, the Ph.D. program at EPFL encourages doctoral students to be involved in teaching, which I am looking forward to. I believe that pursuing a Ph.D. at EPFL will not only expand my knowledge but also bring out my potential as an educator in the future. On the other hand, if I have the opportunity of being a Ph.D. student at EPFL, I will not only be dedicated to my studies, but I will also do my best to promote collaboration between my laboratory at Tokyo Institute of Technology and my new research group at EPFL.

Lastly, I would like to thank the Admission Committee for your time and your consideration.

Best Regards,

Hoang Nguyen.