Faculty of Engineering
Te Kura Pāngarau | School of Mathematics and Statistics

Dr. Alex Gavryushkin
Associate Professor of Mathematics and Data Science
RSNZ Rutherford Discovery Fellow
Associate Dean (Research) of Engineering (incoming)

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Examiner's report on thesis submitted to the University of Warsaw in fulfilment of the requirements of Doctor of Philosophy in Computer Science by Senbai Kang

The thesis on "Probabilistic graphical models for inferring tumor phylogeny and genomic variants from single cell DNA sequencing data" by Senbai Kang is a significant contribution to the field of cancer phylogenetics. Understanding molecular causes of cancer initiation, development, progression, and metastasis has long been a problem of paramount importance for personalized medicine as well as fundamental genomics and biomedicine. A widely accepted hypothesis today is that the molecular mechanisms responsible for various crucial stages of tumor development can be learned from genomic and epigenomic information carried by tumor cells. This, however, presents highly complex technical challenges in statistics, data science, and computer science, many of which are yet to be solved to a satisfactory degree. These challenges are, in particular, due to so-called tumor heterogeneity, which effectively means significant amount of genomic variability between tumor cells even when they come from the same patient, collected at the same time, from proximal locations, and sequenced using the same technology. An approach to this problem that received significant attention in the literature was enabled by single-cell sequencing technologies, which allow approximating the genomic make up of individual cells. Phylogenetic methodology has naturally become the primary tool for reconstructing somatic evolution of these individual cells, because tumor growth is fundamentally a tree-like process driven by cell division and proliferation. A number of tools have been developed to perform phylogenetic reconstruction from single-cell sequencing data, but a reliable and robust method is yet to be developed due to the complexities that Kang accurately summarized in the presented work. The thesis concentrates on two of such complexities — (1) a variant calling method that is reliable enough for traditional phylogenetic inference algorithms, which effectively assume no sequencing errors; (2) a model that accounts for single nucleotide variants as well as other sequence alterations that are very common in cancer such as allele dropout, small insertion, deletions, and copy number changes.

Kang developed two methods to address problems (1) and (2) above — SIEVE, which enables phylogeny-aware variant calling, and DelSIEVE, which extends SIEVE to model small deletions explicitly and as a result to model allele dropouts more accurately. The presented work is novel, original, and it advances the field of cancer phylogenetics. The thesis provides a critical insight into how single-cell DNA-seq data from tumor cells should be modelled and analysed. The main limitations of this work, as acknowledged in the thesis, are the inability to account for small insertions as well as larger copy number changes, which are known to be very important in development of certain cancers.

To summarize, I believe that the results of the thesis constitute a significant step forward in the field of computational cancer phylogenetics, and the thesis is sufficient to grant the degree sought by the candidate, PhD in Computer Science.

Alex Gavryushkin

Feb. 19, 2024
Christchurch
New Zealand