Mathematics in the Public Sector

(Org: **Megan Dewar** (Tutte Institute for Mathematics and Computing) and/et **Kseniya Garaschuk** (Canadian Centre for Cyber Security))

BENOIT HAMELIN, Tutte Institute for Mathematics and Computing
Telemetry representation and interactive labeling to facilitate cyber defense

Cyber defense processes involve the processing and analysis of large volumes of telemetry data. Many phenomena of interest are described over tangled subsequences of telemetry streams, and cyber analysts experience difficulty making out the data patterns that drive their investigations. Behaviours over telemetry streams form a heavy-tailed distribution, precluding the proxy approach of anomaly detection; and there exists no labeling information to discern patterns of interest. This work presents tools for the interactive annotation of such collective telemetry phenomena by characterizing them from mutual similarity and time persistence. The annotation process and outcome is demonstrated through the example of host-based sensor telemetry.

KORAY KARABINA, National Research Council of Canada
Cryptography Meets Topological Data Analysis

Topological Data Analysis (TDA) offers a suite of computational tools that provide quantified shape features in high-dimensional data, which can be utilized by modern statistical and predictive machine learning models. In particular, persistent homology (PH) takes in data and derives compact representations of latent topological structures, known as persistence diagrams. PH has been widely adopted for model development on sensitive data, motivating the computation of PH on encrypted data. In this presentation, I will provide brief introductions to TDA and secure computing and then demonstrate how to modify the boundary matrix reduction algorithm to compute PH on encrypted data using homomorphic encryption.

MASOUD M NASARI, Bank of Canada
Disaggregating low-frequency economic measures

Having access to high-frequency economic measures is essential for making informed decisions and understanding economic trends in real-time. These measures, such as the Gross Domestic Product (GDP), are not always available at a desired high-frequency. For example, the GDP for Canadian regions is available only annually. A method of disaggregating the annual GDP of the regions to estimate their quarterly and monthly values will be presented in this talk. The results will, in turn, be used for nowcasting and forecasting future values of the high-frequency regional GDP

VALÉRIE POULIN, Tutte Institute for Mathematics and Computing (CSE)
Hypergraph exploration via vectorization

Representing data in a relational manner has become a very common decision for analyzing data sets. In some situations, hypergraph structures are preferred over graphs as they allow for capturing more complex relationships amongst objects. In this work, we present a joint vertex and hyperedge vectorization strategy. The joint vectorization we propose is performed in two steps. We first create vertex vectors based on co-occurrences of vertices in hyperedges. Once vertices are embedded in a space, we consider each hyperedge as a distribution over the vertex space and define a hyperedge distance using a distribution metric. The distance between hyperedges is therefore not limited to the intersection size: two non-overlapping hyperedges can end up being similar if the vertices they contain are.

Through a concrete example we demonstrate how this vectorization allows for visual exploration, cluster interpretation and much more. This analytical framework perfectly illustrates what guides unsupervised data science research at the Institute.

QUESTION AND ANSWER PERIOD,

QUESTION AND ANSWER PERIOD,

MARK REMPEL, Defence Research and Development Canada
Practical applications of reinforcement learning for decision support in defence and security

Sequences of decisions that occur under uncertainty arise in a variety of settings, including transportation, communication networks, finance, defence, etc. The classic approach to find an optimal decision policy for a sequential decision problem is dynamic programming; however its usefulness is limited due to the curse of dimensionality and the curse of modelling, and thus many real-world applications require an alternative approach. Given its success in recent years, Reinforcement Learning (RL) has gained popularity as an approach to solve these types of problems. In addition, in the field of operations research, Powell's recently published unified framework for sequential decisions provides a methodology that links mathematical modelling, stochastic optimization, approximate dynamic programming, RL, simulation, as well as other related fields with the aim to model and solve sequential decision problems. In this talk we discuss two recent decision support applications—mass evacuation in the Arctic, and capital investment planning—that focus on using Powell's framework and RL concepts within defence and security. Lastly, ongoing activities within a newly formed NATO research task group that is focused on RL-based decision support will be highlighted.

BENJAMIN SANTOS, Statistics Canada
Multi-Party Privacy Preserving Record Linkage based on Circuit Private Set Intersection

Record Linkage (RL) is the process of combining information about entities in multiple data sources into a single linked dataset. In some linkages, the desired output is not the linked data itself, but a set of aggregates based on the cross-linked dataset, such as, aggregated tables. In our previous work [1], we designed and implemented a protocol for Privacy-Preserving RL (PPRL) with aggregation based on Oblivious Programmable Pseudo-Random Functions (OPPRFs) and Secure Multi-Party Computation (SMPC). This protocol allows two parties with datasets, e.g., a National Statistical Office (NSO) and a Government Agency (GA), to obtain weighted aggregates based on values present in the intersection of both datasets while ensuring privacy in a semi-honest scenario. The goal is to extend it to more than two parties, i.e., Multi-Party PPRL (MP-PPRL). This is a natural extension since parties could be playing the role of an NSO, GAs, regional and/or private partners. We based our work on Chandran et al. [2], that implements Relaxed Batch OPPRFs and SMPC to build a protocol for Circuit Private Set Intersection, which we extended to MP-PPRL. We found that the multi-party extension to PPRL is more complex and stiffer, meaning the solution must be tailored to the problem of study: datasets and aggregations.

 References

 [1] Dugdale, et al. Practical Privacy-Aware Data Linkage and Statistical Aggregation based on Privacy Enhancing Techniques. CROSS-NTTS 2023.

 [2] Chandran, et al. Efficient Linear Multiparty PSI and Extensions to Circuit/Quorum PSI. Proceedings of the 2021 ACM SIGSAC CCCS.

SICHUN WANG, Defence Research and Development Canada
Miscellaneous Applications of Mathematics and Statistics in Statistical Signal Processing and White-Box Cryptography

Mathematics and statistics are not only essential in natural and social sciences, computer science, medicine and finance but also indispensable in a vast array of practical industrial engineering applications, such as error control coding in wireless communications, data encryption and user authentication in cybersecurity and spacecraft orbit determination for global navigation

satellite systems (GNSS). Resolution of mathematical problems arising from various engineering applications presents unique challenges and often requires a combination of engineering insights and sophisticated techniques and tools from mathematics, statistics, numerical analysis and computer science. In this talk, we use real examples to illustrate how engineering problems can be solved by finding solutions to their mathematical/statistical models. More specifically, we shall touch upon the following three topics:

(1) Numerical computation of the normalized detection threshold for FFT filter bank-based signal detection schemes in civilian spectrum monitoring and military radio surveillance. (2) Construction of permutation polynomials on the ring of integers modulo n and their applications in turbo codes, software obfuscation and protection, and white-box cryptography. (3) Geolocation of COSPAS-SARSAT emergency search and rescue beacons.

During the talk, open problems motivated by these three applications will also be briefly discussed.