Fifteenth Annual
Student Research Forum

March 27, 2018

COMPUTER SCIENCE AND SCIENTIFIC COMPUTING

Session Chair: *Dr. Yuanzhu Chen*
Head, Computer Science: *Dr. Minglun Gong*
Director, Scientific Computing: *Dr. Ron Haynes*

TUESDAY, MARCH 27, 2018, EN-2022

1000   Opening Remarks

1010   Toward Understanding Hidden Patterns in Human Mobility Using Wi-Fi, **Ali Farrokhtala**

1030   Stereo Matching for 3D Reconstruction, **Wendong Mao**

1050   The Computational Complexity of Simultaneously Designing Homogeneous Swarms of Robots by Library Selection and Environments for Distributed Construction, **Mesam Timmar**

1110   Using Evolutionary Computation methods in Astrophysics for Exoplanet Classification, **Jake Dowzell**

1130   Peer-to-peer Online Social Networks, **Arastoo Bozorgi**

1150   Understanding dynamics of human behaviour based on individual mobility and impact a social network, **Shorouq AL-Eidi**

PIZZA LUNCH

1310   Applying Audit Trails from Access Logs to Achieve Fine-Grained Access in Healthcare System, **Majid Afshar**

1330   Explore meta-learning via inductive transfer learning methods, **Yubin Li**

1350   Meshless and conservative methods for atmospheric and tsunami computations, **Rudiger Brecht**

1410   Drop and Optimize short connections in Deep Networks, **Mingjie Wang**

BREAK

1450   Measuring the Importance of Vertices in Weighted Human Disease Network, **Seyed Mehrzad Almasi**

1510   Designing a walkability measure for St. John's, Canada, **Hoda Rafieipour**

1530   An Evolutionary and networking Approach for Metabolomics using Genetic Programming, **Fatemeh Tavafi**

1550   **CLOSING REMARKS**

10:10   **Ali Farrokhtala**, Ph.D. Computer Science

Title: **Toward Understanding Hidden Patterns in Human Mobility Using Wi-Fi**

Abstract: Identifying spatial-temporal patterns in human mobility through a reliable model is rewarding in many applications such as computer networking and mobile communication. These hidden routines are inherited from our repeating behaviours with respect to three primary contexts of time, space, and social environments. Hence, studying a suitable source of sensor data that is scalable, multidimensional, and social network illustrative can enable us to develop a universal human mobility model and potentially prediction system. We first demonstrate that Wi-Fi events (connectivity logs and scans records) from mobile phone devices share a similar set of characteristics to some of the real-world large-scale networks. One particularly is the long-tailed property of node degree distribution in produced projection networks from the bipartite network of consisting users and access points as the nodes. This feature can be interpreted as the robustness of Wi-Fi system against structural changes of removing a set of nodes or connections. Later, we transform Wi-Fi logs into a tabular data format holding different time granularities and location-tagged information. The new dataset is sparse and difficult to analyze. Thus, we reduce dimensionality by extracting its structural patterns using principal component of new features. Our analysis shows that we can reconstruct original data with more than 90% accuracy using only a set of top eigenvectors with the quarter size of original features, while the outliers with noisy data are filtered out. Our proposed technique help to visualize users similarities and their behaviour dynamics, besides reducing computation complexity of any further analysis.

* * *

10:30   **Wendong Mao**, Ph.D. Computer Science

Title: **Stereo Matching for 3D Reconstruction**

Abstract: 3D Reconstruction of real world scenes are widely required in both military and civilian fields. Many approaches have been used to acquire 3D information, but most work is based on images or/and laser data. Multi-view stereopsis relying on the former mainly focuses on stereo matching for image pairs. Due to its important applications, stereo matching problem has been extensively studied over the past decades, but it is still an ill-posed problem. As the forthcoming artificial intelligence revolution extends, using neural networks to perform 3D reconstruction for multi-view stereopsis has become possible.

* * *

10:50   **Mesam Timmar**, M.Sc. Computer Science

Title: **The Computational Complexity of Simultaneously Designing Homogeneous Swarms of Robots by Library Selection and Environments for Distributed Construction**

Abstract: Using teams of robots in the construction process of a structure is very prominent problem in distributed robotics and lots of effort has been put into studying this. In most practical cases, when robots are used in construction, they are being supervised in one way or other. Studies have been done on unsupervised construction of structures by robots most of which is inspired by natural swarms. Previous work involved in this field is related to algorithms for construction of homogeneous robot swarms with stochastic behavior. Complexity analysis has also been done in this field, older work done is more general in terms of formalizing control mechanism and environments. Recent complexity work has been done on designing robot controllers or environments for construction tasks and on construction, repair and maintenance tasks by selecting robot controllers from a simple library. The missing chunk here is co-design of homogeneous robot swarm by designing controllers (through library selection) and environment simultaneously. We use complexity analysis tools to determine that if an algorithm can both be fast in producing swarm of robots (by the library selection) while simultaneously creating an environment for quick and guaranteed completion of construction-related tasks. We will study an extension of above stated problem which is related to design of robots by library selection but with limited modification. For studying the above stated problems we use a finite state robot controller model with non-continuous deterministic motion in grid-based environments. Our results will show whether there exist algorithms that produce correct and quick results for every input and we will see the restrictions if any under which these algorithms exist.

* * *

11:10   **Jake Dowzell**, M.Sc. Scientific Computing

Title: **Using Evolutionary Computation methods in Astrophysics for Exoplanet Classification**

Abstract: "As discoveries and confirmations of exoplanets increase there comes a greater need for an adaptive classification system, as so few exoplanets to date have been succinctly classified. As such, the focus of my research will be to build and implement an exoplanet classification system, building on the ideas and methodologies of Genetic Algorithms and Evolutionary Programming. By 'training' my system to classify exoplanets into simple categories (size, mass, composition, etc) using known data, I can then run unconfirmed or unknown exoplanet data through it to classify these new exoplanets. Once categorised I can perform further analysis of the exoplanets to identify the presence of organic compounds in atmospheric data, which may help indicate whether or not a given exoplanet is capable of supporting biological life."

* * *

11:30   **Arastoo Bozorgi**, Ph.D. Computer Science

Title: **Peer-to-peer Online Social Networks**

Abstract: The client-server architecture of existing online social networks (OSNs) facing their users with some privacy concerns which is due to the centralized nature of user data, stored on provider's storages and controlled by one single entity. In such platforms, the more the users share private information about themselves, the more their privacy is violated. On the other hand, user devices and machines are provided with a huge amount of storage which can be used as the primary data storage for OSN users. This way of thinking needs a shift from the traditional client-server architectures towards the peer-to-peer (P2P) ones to design an OSN with the main aim of protecting user data.

To do so, we need a P2P file system to handle data storage and retrievals, look-up operations and user communications. In this presentation, we introduce known distributed file systems and discuss their features and explain how we can use such file systems to design a P2P online social network.

* * *

11:50 **Shorouq AL-Eidi**, Ph.D. Computer Science

Title: **Understanding dynamics of human behaviour based on individual mobility and impact a social network**

Abstract: Over the past few years, social and technological networks analyses have attracted attention as technological advances have given us a wealth of empirical data. Previous studies focused on analyzing these networks as static networks which are defined as networks that do not change over time as do biological and social systems. These studies also analysed the networks in aggregate over at intervals of time. Because of collection measurements related to large real network traces, researchers are starting to realise that networks connections change rapidly at time; thus, greater of dimensionality can be captured than with static networks.

We will first, introduce the temporal graph model which will formalise the notion of shortest paths that are widely applied in graph theory and show that static models ignore the time order of contacts and overestimates the available links. Secondly, we focus on finding important nodes in a network structure using the temporal nodes centrality measure, such as betweenness and closeness based on temporal shortest paths utilizing these metrics to explore temporal periodic patterns of human contact networks will improve the ad hoc network. Finally, we study the predictability of links in dynamic social networks to predict both relations between the individuals, such as friendship, and future behaviour such as collaborations.

The motivation of this study is that the use of time aware- metrics for analyzing real dynamic networks opens the door to more precise and effective exploitation of complex network science.

* * *

13:10   **Majid Afshar**, Ph.D. Computer Science

Title: **Applying Audit Trails from Access Logs to Achieve Fine-Grained Access in Healthcare System**

Abstract: To keep Electronic Health Records (EHRs) secure and private, a wide range of studies has been done by healthcare organizations. Common security mechanisms such as secure networks with firewalls, encrypted communication and strong policy for selecting complex passwords have been applied in healthcare organizations to improve the level of patients' privacy. Nevertheless, it still endures a quandary of how to permit the real eligible users to access the patient records while blocking abnormal accesses by authorized users of the system. Health Insurance Portability and Accountability Act (HIPAA) policies say that users allow having access to records for only Treatment, Payment, and healthcare Operations (TPO). In order to perform their duties, authorized users are usually allowed restricted access to those functions. However, a type of privacy breach happens once authorized users of a healthcare organization access protected records without TPO reasons. Of course, sometimes, improper accesses have been associated with the criminal aspect. We have applied a special Attribute-Based Access Control (ABAC) in healthcare environment by understanding the healthcare standards and deriving its security requirements. Moreover, we used access logs auditing of clinical systems to detect malicious users. With well-structured and fine-grained authentication, analysis of access logs could be a highly beneficial tool for learning how to recognize authorized users with abnormal access requests and reduce the demand for exception-based access.

Keyword: EHR, ABAC, Privacy, Healthcare, Access log

* * *

13:30   **Yubin Li**, M.Sc. Computer Science

Title: **Explore meta-learning via inductive transfer learning methods**

Abstract: Meta-learning studies how learning systems can improve themselves through experience, the goal of meta-learning is to figure out how learning itself can become flexible, or how to extract the underlying knowledge in certain domain and use it to improve itself in future learning task. In this project, the plan is to utilized inductive transfer learning method to achieve meta-learning, and test how it works by predicting ferromagnetic behavior of Strongly Correlated Fermions.

* * *

13:50   **Rudiger Brecht**, M.Sc. Scientific Computing

Title: **Meshless and conservative methods for atmospheric and tsunami computations**

Abstract: We develop a variational integrator for the shallow-water equations on a rotating sphere. The variational integrator is built around a discretization of the continuous Euler–Poincaré reduction framework for Eulerian hydrodynamics. This will be applied as a tsunami inundation and propagation model and/or climate model.

* * *

14:10   **Mingjie Wang** , Ph.D. Computer Science

Title: **Drop and Optimize short connections in Deep Networks**

Abstract: Neural networks have resurged recently, most in Convolutional Neural Network(CNN). A variety of novel architectures have been proposed during past few years and achieved the state-of-the-art performance on a wide range of visual tasks, such as AlexNet, VGG, GoogLeNet and ResNet etc. Recent work has demonstrated that shorter connections are benecial for training deeper CNNs effectively. DenseNet(Gao Huang 2017) has taken advantage of direct-connections between each layer of deep network. Based on this design and Dropout regularizer, in this proposal, I evaluate the necessity of fully-dense connections and present a novel regularizer called 'DropShortConnetion', which randomly select subsets of Short Connections to be dropped from the Network according to a given Drop probability. This approach has several compelling advantages: (1) Be implicitly similar to the average of various CNNs. (2) Signicantly alleviate the problem of 'Vanishing gradients'. (3) Reduce the number of parameters. This technique can be viewed as the generation of Dropout(Hinton 2016). In addition, to find an optimal set of short connections which can achieve competitive results and reduce computational costs, I adopt Genetic Algorithm to traverse the large search space and automatically create architectures with high capacity using subset of skip connections. This paper utilizes traditional Stochastic Gradient Descent(SGD) to train networks and defines validation accuracy as the fitness function of Genetic Algorithm. Remaining setting parameters(e.g. network congurations and parameters of training) are kept the same as these in DenseNet. The rest content also analyses obtained architectures through visualization of intermediate feature-maps to help deeply understand the importance of Short Connections in Deep Networks.

* * *

14:50   **Seyed Mehrzad Almasi**, M.Sc. Computer Science

Title: **Measuring the Importance of Vertices in Weighted Human Disease Network**

Abstract: Many human genetic disorders and diseases are known to be related to each other through frequently observed co-occurrences. Studying the correlations among multiple diseases provides an important avenue to better understand the common genetic background of diseases and to help develop new drugs that can treat multiple diseases. Meanwhile, network science has seen increasing applications on modeling complex biological systems, and can be a powerful tool to elucidate the correlations of multiple human diseases. In this article, known disease-gene associations are represented using a weighted bipartite network. We extract a weighted human diseases network from such a bipartite network to show the correlations of diseases. Subsequently, we propose a new centrality measurement designed specifically for the weighted human disease network in order to quantify the importance of diseases. We evaluate our centrality measurement and compare it with the most commonly used centralities in biological networks including degree, closeness, and betweenness. The results show that our new centrality is able to find more important diseases since the removal of the top-ranked diseases leads to a higher decline rate of the network efficiency.

Keywords: Complex networks, bipartite graphs, weighted networks, vertex importance, vertex centrality, human disease network, disease-gene association.

* * *

15:10   **Hoda Rafieipour**, M.Sc. Computer Science

Title: **Designing a walkability measure for St. John's, Canada**

Abstract: Over recent decades, by decreasing physical activity in both developed and developing countries, the rates of obesity and diabetes have considerably increased. Also, research indicates Newfoundland and Labrador has the highest rate compared to other cities of Canada. Since improving the trend of walking is an important step towards a healthier life, the features of urban built environments play a key role in health outcomes. Walkability is a measure which evaluates the level of friendliness of an area for walking. Walkability indices are powerful tools to study the impact of the built environment on physical activity, obesity, and diabetes. A common walkability measure which is available for researchers and also has been used in Canada called Walkscore. Although Walkscore is available for Canada, there are some limitations to calculate this measure. The first limitation is that algorithm and methodology are proprietary to calculate this measure. It means that our findings can be significantly changed by making any change in our algorithm and methodology. Therefore, our goal is to develop a walkability index for St. John's. In the first phase of our research, we conducted a systematic review of walkability measures to

establish which variables should be included in the walkability measure. In the second phase of our project, we developed a methodology to calculate walkability measure. First, we modeled a road network from the Open Street Maps dataset and considered just the areas around the Memorial University and downtown. Then we extended the studied areas to approximately whole St. John's area and determined all connected components of the selected area and computed average of Closeness centrality and PageRank for each component that represent street connectivity. Finally, our proposed Walkscore has been computed for each component and it contains the number of amenities and achieved centralities.

* * *

15:30   **Fatemeh Tavafi**, Ph.D. Computer Science

Title: **An Evolutionary and networking Approach for Metabolomics using Genetic Programming**

Abstract: The emerging omics technologies have empowered researchers to step forward in the study of biological systems. Metabolomics as the youngest member of omics family is the comprehensive study of the small molecule metabolites present in fluids and tissues. Metabolomics is more time-sensitive and dynamic than other omics. Therefore, from biotechnology point of view, it is very difficult to measure multiple metabolite concentration levels at the same time. Clinical research is mainly concentrated on disease diagnosis and the goal of metabolomics is to provide practical and proper diagnostic applications with the highest accuracy to determine the best treatment plan. Feature selection methods can be used to identify the most relevant metabolites associated with a particular disease. The aim of this project is to propose a novel metabolomics framework using linear genetic programming for automatic feature selection combine with classification and network analysis based on Newfoundland Osteoarthritis Study (NFOAS) dataset to identify major metabolites related to Knee Osteoarthritis, and finally develop a new bioinformatic webserver for disease prediction and analysis based on the proposed methodology. The main challenge of this research is high-dimensionality.

* * *

15:50   **Closing Remarks**