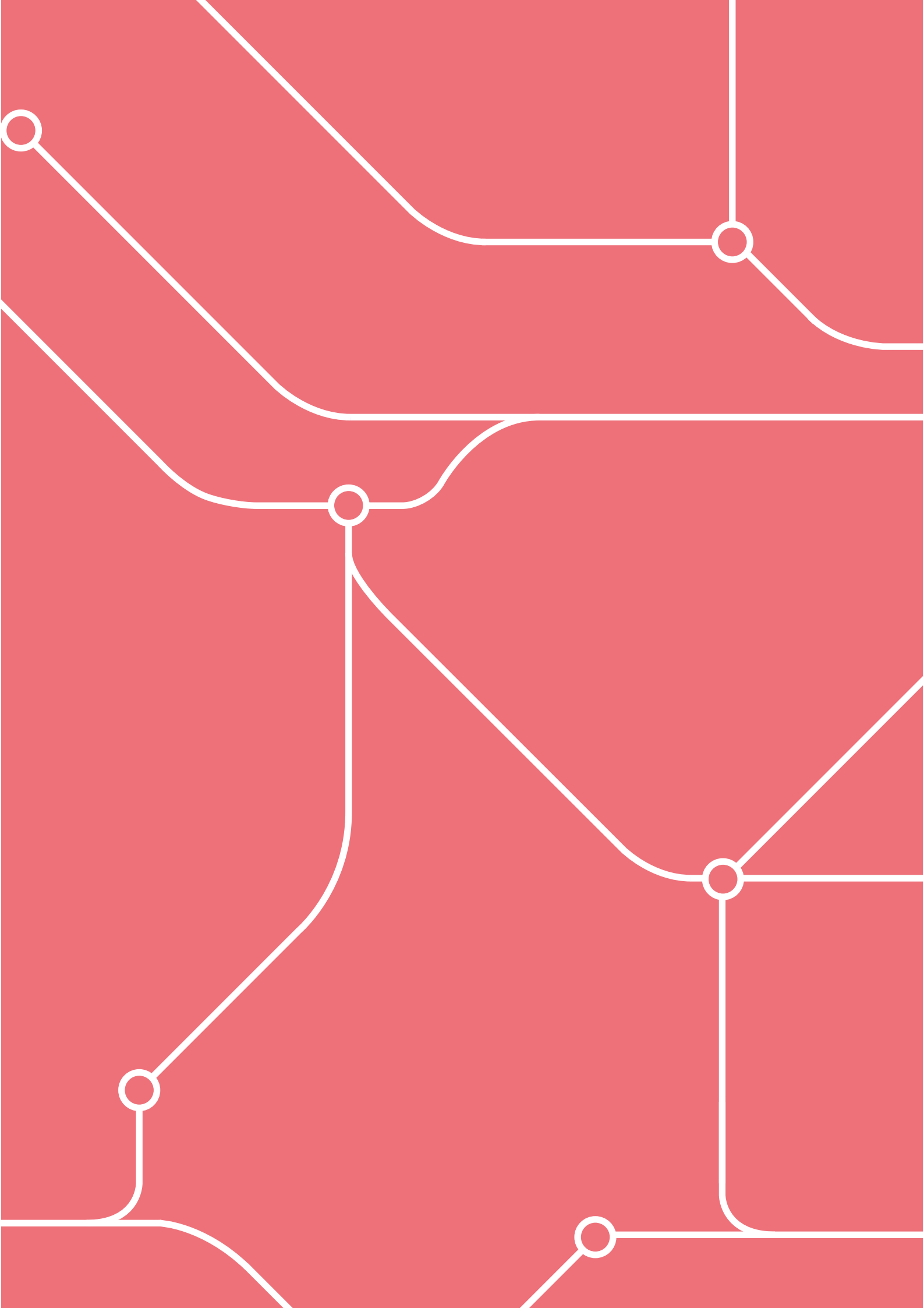




WPCCS  
2019



It is my pleasure to welcome you to the 17th Warwick Postgraduate Colloquium in Computer Science. This year, the event is in the new Mathematical Sciences Building, which we hope will be the home of many colloquia to come.

The past year has seen many great developments in the department, including the founding of our Centre for Doctoral Training and Research in Computer Science. Thanks to the CS CDT, we are now able to provide multiple scholarships for international, EU, and UK students. This has already helped us with the continued growth and depth of our research activities. We have welcomed many new PhD students to our PhD programme, and established a new multi-day PhD Welcome Event which proved to be a highly productive introduction which we look forward to organising again in future years. This year was special for the Urban Science CDT, which has seen its first graduates. We wish them all the very best in their future careers.

Today, we reflect on our research over the past year. Each of you are exploring exciting avenues in your own respective areas and have already started to produce high quality work. These early years are among the most valuable times to build expertise and experience which will be useful for many years to come.

WPCCS provides a great environment to learn from each other. Please talk with your peers, after the presentations, and during the poster sessions and breaks. I hope you will have fruitful and creative discussions about your work, and the work of those around you.

Welcome again to WPCCS 2019!

- Hakan Ferhatosmanoglu

Welcome to  
**WPCCS 2019**

---

Mathematical  
Sciences Building  
Mon 9th December

**TRACK 1 · MB0.01**Computational  
Biology

John Pocock  
Rawan Albusayli  
Yijun Quan  
Ruqayya Awan

**TRACK 2 · MB0.07**Theory, Foundations &  
Discrete Mathematics

Jacques Dark  
Eleanor Davies  
David Purser  
Alex Dixon

**TRACK 3 · MB0.08**Computer Security  
& Networks

David Richardson  
Abdullah Al Hajri  
Shin Wang  
Sara Alhajaili

**BREAK****TRACK 1 continued**Computational  
Biology

Hammam Alghamdi  
Josh Cavie  
Ruiqing Feng  
Abdullah Algamdi

**TRACK 2 continued**Theory, Foundations &  
Discrete Mathematics

Grzegorz Lisowski  
Thejaswini Raghavan  
Marcel De Sena Dall'Agnol  
Christopher Hickey

**TRACK 4 · MB0.08**Machine Learning &  
Artificial Intelligence

Yujue Zhou  
Abeer Almowallad  
Bowen Du  
Sarah Alshamrani

**LUNCH****TRACK 5 · MB0.01**HPC &  
Databases

Dean Chester  
Kamalavasan Kamalakkannan  
Dom Brown  
Richard Kirk  
Edward Chuah

**TRACK 6 · MB0.07**Urban  
Science

Natalie Rothwell  
Ivana Tosheva  
Alexander Noll  
Jonathan Davies  
Nicole Hengesbach

**TRACK 4 continued**Machine Learning &  
Artificial Intelligence

Naif Alotaibi  
Huda Alrashidi  
Gabriele Pergola  
Aparajita Haldar  
Fitri Nurinsani Rachbini

**BREAK****TRACK 5 continued**HPC &  
Databases

Ali Meree  
Qingzhi Ma  
Shuang Wang  
Ali Mohammadi  
Shanghooshabad

**TRACK 5 continued**Urban  
Science

Teddy Cunningham  
Andra Sonea  
Matteo Mazzamurro  
Chris Conlan

**TRACK 4 continued**Machine Learning &  
Artificial Intelligence

Wentai Wu  
Man Luo  
Helen McKay  
Amir Shirian

# When & Where

9:00 am Welcome to WPCCS 2019 in MB0.07  
9:10 am Guest talk from Torsten Mutze in MB0.07

---

## 9:45 am Tracks 1, 2 & 3

10:50 am Break with refreshments in foyer

---

## 11:15 am Tracks 1, 2 & 4

12:20 pm Lunch and posters in foyer

---

## 1:20 pm Tracks 4, 5 & 6

2:40 pm Break with refreshments in foyer

3:05 pm Guest talk from Feng Hao in MB0.07

---

## 3:40 pm Tracks 4, 5 & 6

4:50 pm Prize-giving in MB0.07

---

## 5:00 pm Networking in foyer

With festive drinks and snacks provided by the CDTs in Computer Science and Urban Science

---

## SCHEDULE

---

### Computational Biology

Zhuo Wan  
Muhammad Shaban  
Navid Alemi Koochbanani

### Comp. Sec. & Networks

Jasmine Grosso  
Alwaleed Alharbi  
Mohammed Mufareh A Maray  
Mansour Aldawood Aldawood

### POSTERS

---

Presented in the ground floor foyer at lunch

### Machine Learning & AI

Tom Wood  
Junyu Li  
Aseel Alturki  
Nouf Almujaally  
Hao Wu  
Haoyi Wang  
Nicole Peinelt  
Ebtehal Quqandi  
Debaleena Roy  
Jiashu Liao

### Urban Science

Timothy Sit  
Philipp Ulbrich  
Shanaka Perera  
Katherine Ascott  
Melissa Kenny  
Oba Waiyaki  
Ciaran Devlin  
Christian Caton  
Elisa Baioni  
Bradley Sheridan

# Guest Talks

## Origami and the Mathematics Behind It

**Torsten Mutze · Presentation at 9:10am in MB0.07**

In this talk we explore the mathematical and algorithmic theory behind the Japanese art of paper folding, illustrated with some beautiful models.

## Tackling Real-World Security Problems

**Feng Hao · Presentation at 3:05pm in MB0.07**

Cryptography is a useful tool that often allows people to solve intuitively impossible problems with real-world impact. In this talk, I will present three examples from our previous research work. In the first example, I will show a technique that allows a user to anonymously send a veto message with untraceable origin even when all communication is under the strict surveillance, say by a big brother. In the second example, I will show a protocol that allows two remote users to establish a “high”-entropy session key from a “low”-entropy shared secret (password). The high-entropy session key can then be used for secure communication. In the third example, I will show a voting system that allows every voter to tally votes by themselves without needing any trusted tallying authority, while preserving the voter’s privacy. In each of these examples, cryptography is a key enabling technology that transforms the impossible to the possible.

# Prizes

WPCCS aims to engage attendees with their colleagues' research and with research in the wider community. For this purpose, the Department of Computer Science and the Research Development Programme have kindly agreed to sponsor the event.

For today's posters and presentations, over 20 prizes will be awarded by the Programme Committee. Amazon vouchers will be awarded for posters and presentations in each track that best inform and educate with engaging delivery. Two further prizes will also be awarded, 'Best Poster in Colloquium' and 'Best Presentation in Colloquium'.

Prizes will be awarded at 4:50pm in MB0.07.

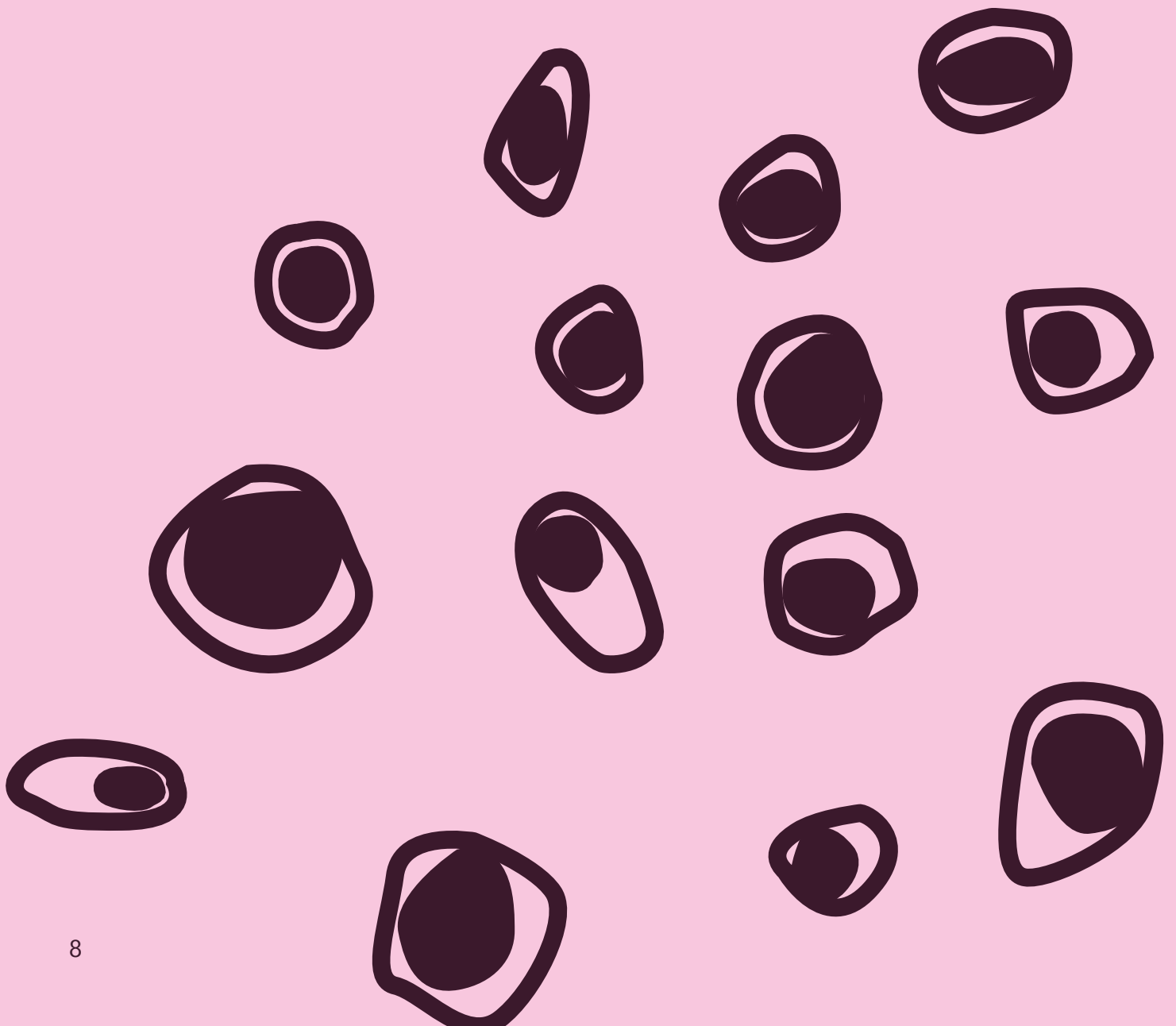
# Computational Biology

Track 1

**MB0.01 · 9:45am**

---

Session Chairs:  
Chris Conlan &  
Alexander Noll





# Challenges of Working with Large Biological Images

**John Pocock · Presentation at 9:50am**

Digital pathology images are very large in both x and y dimension, typically over 50,000px in each axis. This presents many challenges such as how to store, transfer, and process the images. Digital pathology images are also an excellent candidate for the application of machine learning methods. However, the large dimensions of the input data introduces additional complexities when applying machine learning methods due to GPU memory constraints and data availability requirements for running training algorithms for convolutional neural networks (CNNs) efficiently on GPU hardware. Many approaches exist such as splitting the problem into many smaller sub-problems, compressing the input data, or reducing redundant input data. This presentation will explore such methods currently in use in addition to promising emerging methods.

---

## Automatic Classification of Cancer Tissues in Breast Cancer

**Rawan Albusayli · Presentation at 10:05am**

Histological images hold valuable information which is useful for assessing cancer diseases. In computational pathology, the use of deep learning will be helpful to make and support the decision about cancer grades, survival outcomes, and potential treatments in most minimum time. My study will focus on creating a module that inspects whole-slide images (WSIs) and produce useful information that supports clinical decisions. The research concentrates on triple-negative breast carcinoma (TNBC) as 12% to 24% of breast cancer cases are categorized under this type; it acquires so many lives due to the late discovery of the disease and the lack of effective targeted therapies. This work hopefully will be beneficial to the pathology field as the goal is to produce an early cancer detection system which helpful for treatment responding, in addition to tackling the problem of pathologists shortage.

## Camera Sensitivity Dependent PRNU-Based Image Forensics

**Yijun Quan · Presentation at 10:20am**

Photo Response Non-Uniformity (PRNU) left in the images by imaging sensors is a powerful device fingerprint for image forgery detection. The correlation between an image's noise residual with the device's reference PRNU is often compared with a decision threshold to check the existence of PRNU. A PRNU correlation predictor is a natural approach to determine the threshold. Camera sensitivity, commonly known by the name of ISO speed, allows photos to achieve desired exposure under various conditions. Meanwhile, such freedom also brings challenges to PRNU based image forgery detection due to PRNU correlation's dependency on ISO speed. Thus, we propose that correlation predictor is ISO-specific: reliable correlation predictions can only be made when a correlation predictor is trained with images of similar ISO speeds to the image in question. To overcome the problem that an image's ISO speed may be absent, a block-matching method is proposed.

---

## Identification of Low and High Risk Urine Cytology Images using Cell-Based Classification

**Ruqayya Awan · Presentation at 10:35am**

Urine cytology is a gold standard for the screening of bladder cancer and is mainly used for detecting high-grade cancer. Accurate identification of atypical and malignant cells is a challenging task and is an essential part of identifying different diagnosis with low-risk and high-risk malignancy. The computer-assisted identification of high-risk cancers can be complementary to the clinicians for treatment management and in providing advice for carrying out further tests on an urgent basis. In this study, we employed two different deep learning-based approaches for the detection and classification of different types of cells. Based on network predictions at the cell level, we identified low-risk and high-risk cases using count of atypical cells and the total count of atypical and malignant cells. The area under the curve using a total count of atypical and malignant cells is 13% higher than that obtained with the count of malignant cells only.

# A Hybrid Pipeline to Assess Oestrogen Receptor Stained Nuclei in Invasive Breast Cancer

**Hammam Alghamdi · Presentation at 11:20am**

Oestrogen Receptor (ER) expression status in invasive breast cancer not only determines the use of endocrine therapy but its level of expression also provides critical prognostic and predictive information. Digital pathology opens new avenues for computational algorithms to provide objective and accurate assessment of ER status. In this study, we propose a novel hybrid pipeline that combines deep learning (DL) and relatively inexpensive colour histogram features in order to recognise and assess different cell types, including ER positive (ER+) and negative (ER-) tumour cells. Our pipeline consists of a deep neural network for simultaneous detection and classification (SimNuc-Net) of nuclei, followed by unsupervised hierarchical clustering. First, the SimNuc-Net classifies ER+ and ER- invasive tumour nuclei and nuclei of other cell types. We then classify all ER+ nuclei into four categories based on staining intensity. We show that the proposed pipeline outperforms the DL only pipeline and other existing techniques.

---

# Normative Modelling in Depression

**Josh Cavie · Presentation at 11:35am**

Depression has been ranked by the World Health Organization as the leading cause of years-of-life lived with disability, with the number of people who suffer from the disease at some point in their lives existing within the 8-12% range in most countries. Research into the theory of depression is therefore of great importance, both in terms of the health of the rising number of individuals affected, and the wider societal effects it can have. Due to their all-encompassing nature, traditional group-level analyses ignore the inter-individual variability across patients with depression. Normative modelling is a new approach for parsing heterogeneity in clinical conditions, aiming to model biological variation within clinical cohorts. This research has used normative modelling to analyse changes in functional connectivities of patients with depression and the symptoms associated with them, and used these to subtype the patients studied.

# Association between Brain Functional Connectivity and Hypertension

**Ruiqing Feng** · Presentation at 11:50am

There is less study has examined how hypertension affect brain functional connectivity. The UK biobank dataset is helpful to find out the association between high blood pressure and brain functional connectivity and which brain area is the most influenced by hypertension on a larger scale. So, we report an analysis of hypertension based on the UK Biobank data. Finally, these results suggest that hypertension will reduce the strength of functional connectivity among whole brain especially the hippocampus.

---

# Drone Action Recognition using Capsule Network

**Abdullah Algamdi** · Presentation at 12:05pm

Videos captured by drones are challenging to work with due to the unfamiliar viewpoints of pedestrians in addition to their size variation with differing camera altitudes and angles. Understanding pedestrian actions from these frames is a challenging task in computer vision, especially if an individual is expected to perform multiple actions in the same set of frames. We propose DroneCaps, a matrix capsule architecture to recognise actions from videos captured by drones. The DroneCaps allows the network to understand the different viewpoints and poses of the person. Our results show that DroneCaps outperforms many existing methods based on the Okutama-Action dataset in both single and multi-label classes with 34.14 % and 52.44 %, respectively.

## Prediction of Sensation-Seeking From Functional Connectivities of the Medial Orbitofrontal Vortex with the Anterior Cingulate Cortex

**Zhuo Wan** · Poster presented in foyer

Sensation seeking is a multifaceted personality trait with components that include experience seeking, thrill and adventure seeking, disinhibition, and susceptibility to boredom, and is an aspect of impulsiveness. We analyzed brain regions involved in sensation-seeking in a large-scale study with 414 participants, and showed that the sensation-seeking score could be predicted from the functional connectivity with correlation  $r=0.34$  ( $p=7.3 \times 10^{-13}$ ) between the predicted and actual sensation-seeking score across all participants. Interestingly, most of these links were between the medial orbitofrontal cortex and the anterior cingulate cortex. We propose that this important aspect of personality, sensation-seeking, reflects a strong effect of reward (in which the medial orbitofrontal cortex is implicated) on promoting actions to obtain rewards (in which the anterior cingulate cortex is implicated). This discovery helps to show how these two brain regions influence behaviour and personality, and shows that risk-taking involving these systems plays a role in sensation-seeking.

---

## A Novel Digital Score for Abundance of Tumour Infiltrating Lymphocytes Predicts Disease Free Survival in Oral Squamous Cell Carcinoma

**Muhammad Shaban** · Poster presented in foyer

Oral squamous cell carcinoma (OSCC) is the most common type of head and neck (H&N) cancers with an increasing worldwide incidence and a worsening prognosis. The abundance of tumour infiltrating lymphocytes (TILs) has been shown to be a key prognostic indicator in a range of cancers with emerging evidence of its role in OSCC progression and treatment response. However, the current methods of TIL analysis are subjective and open to variability in interpretation. We propose a novel method for objective quantification of TIL abundance in OSCC histology images. The proposed TIL abundance (TILAb) score is calculated by first segmenting the whole slide images (WSIs) into underlying tissue types and then quantifying the co-localization of lymphocytes and tumour areas in a novel fashion. We show that the TILAb score is a strong prognostic indicator ( $p=0.0006$ ) of disease-free survival (DFS) on our OSCC test cohort.

---

# Nuclear Instance Segmentation using a Proposal-Free Spatially Aware Deep Learning Framework

**Navid Alemi Koohbanani** · Poster presented in foyer

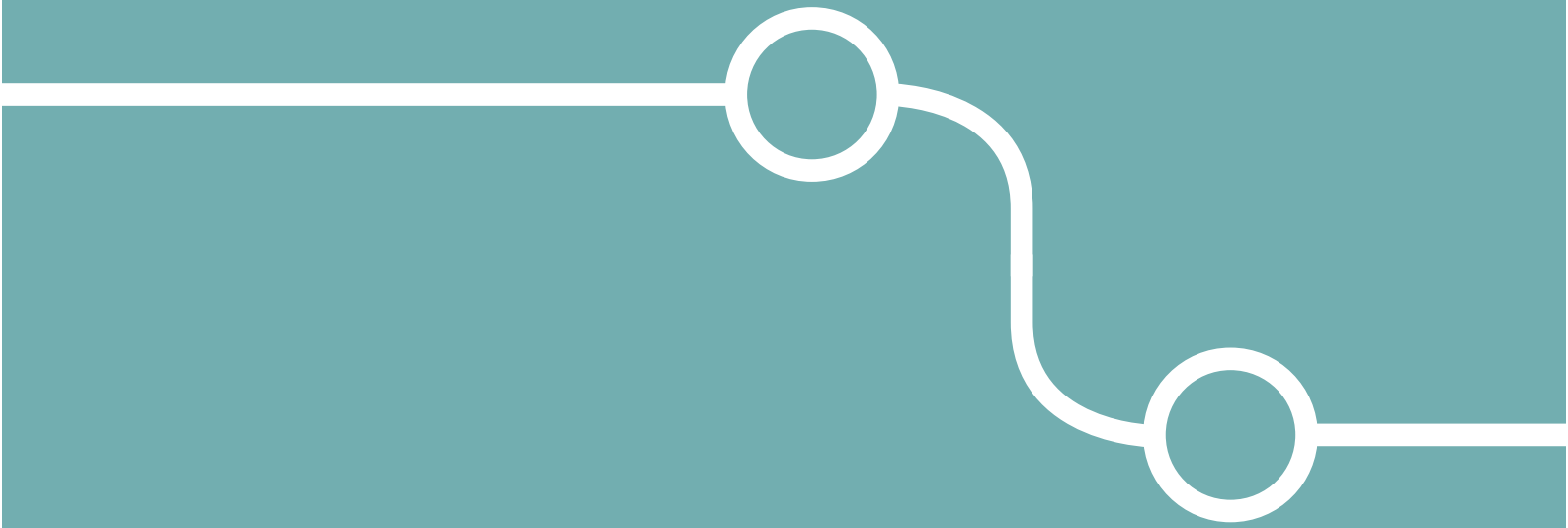
One of the main hurdles in nuclear instance segmentation is overlapping nuclei where a smart algorithm is needed to separate each nucleus. In this paper, we introduce a proposal-free deep learning based framework to address these challenges. To this end, we propose a spatially-aware network (SpaNet) to capture spatial information in a multi-scale manner. A dual-head variation of the SpaNet is first utilized to predict the pixel-wise segmentation and centroid detection maps of nuclei. Based on these outputs, a single-head SpaNet predicts the positional information related to each nucleus instance. Spectral clustering method is applied on the output of the last SpaNet, which utilizes the nuclear mask and the Gaussian-like detection map for determining the connected components and associated cluster identifiers, respectively. The output of the clustering method is the final nuclear instance segmentation mask.

# Theory, Foundations & Discrete Mathematics

Track 2  
**MB0.07 · 9:45am**

---

Session Chairs:  
Richard Kirk &  
Mary Scott



## Two-Party Maximum Matching

**Jacques Dark · Presentation at 9:50am**

Consider the following two-party setting: Alice knows  $A$  - the edges of some graph, while Bob knows  $B$  - a subset of the edges which are deleted. Alice can send a single message to Bob, who must then output a  $(1/c)$ -approximate maximum matching in  $A \setminus B$  (i.e. a subset of edges not sharing any vertices which is at least a  $1/c$  fraction of the largest possible). We extend an existing communication problem to exhibit two-dimensional structure and use it to show that this problem requires a message of size  $\Omega(n^2/c^3)$ . This matches the existing tight bound for dynamic edge streams but gives a new proof that does not require the stream to allow multi-edges.

---

## Having It All: Modularity and Performance and Correctness in Compiler Tree Transformations

**Eleanor Davies · Presentation at 10:05am**

Compilers often make use of abstract syntax tree transformations in translating code from one language to another. Automatically fusing these tree transformations promotes modularity whilst mitigating the performance impact of a resulting increase in the number of transformations. However, existing approaches to fusing tree transformations tend to take an informal approach to soundness, or be too restrictive to consider the kind of transformations needed in a compiler. We use postconditions to define a more useful formal notion of successful fusion, and present criteria that allow modular reasoning about the effects of fusion on transformation behaviour.



# The Complexity of Verifying Circuits as Differentially Private

**David Purser · Presentation at 10:20am**

We study the problem of verifying differential privacy for straight line programs with probabilistic choice. Programs in this class can be seen as randomized Boolean circuits. We focus on two different questions: first, deciding whether a program satisfies a prescribed level of privacy; second, approximating the privacy parameters a program realizes. We show that the problem of deciding whether a program satisfies  $\epsilon$ -differential privacy is  $\text{coNP}^{\#P}$ -complete. In fact, this is the case when either the input domain or the output range of the program is large. Further, we show that deciding whether a program is  $(\epsilon, \delta)$ -differentially private is  $\text{coNP}^{\#P}$ -hard for  $(\epsilon, \delta)$ -differential privacy, and in  $\text{coNP}^{\#P}$  for small output domains, but always in  $\text{coNP}^{\#P^{\#P}}$ . Finally, we show that the problem of approximating the level of differential privacy is both NP-hard and  $\text{coNP}$ -hard. Joint work with Marco Gaboardi and Kobbi Nissim.

---

# Exploring Space with the power of Space Exploration!

**Alex Dixon · Presentation at 10:35am**

We will take a whistlestop tour through the exciting, expressive power of the Petri Net model and its classical decision problems. We will then investigate ten ways in which this formalism can be used to power projects of all kinds, from logistics to actual space exploration. Existing applications include business process modelling (workflow management); traffic planning and modelling; railway signalling; hardware and software verification; game design and development; and many more. Further ideas for applications of this exciting and flexible formalism are invited.

# Coalitional Hotelling-Downs Games

**Grzegorz Lisowski · Presentation at 11:20am**

We study a variation of the Hotelling-Downs model, which is aimed at capturing coalitional strategic candidacy. We consider a number of political parties selecting a single representative among members with predefined positions on the political spectrum. Subsequently, we assume that each voter is attracted to the closest candidate. Then, the goal of each party is to attract as many voters as possible. Within this framework we explore algorithmic properties of classical game-theoretic solution concepts, focusing on Nash equilibria (NE) and dominant strategy equilibria (DSE). We show that while finding a DSE is tractable regardless of the number of competing parties, finding a NE is NP-complete when the number of parties is large. However, if there are only two competing groups, checking if a NE exists in a given game is possible in linear time. This holds as well for DSE.

---

# Parity Games, Strahler Number and Register Index

**Thejaswini Raghavan · Presentation at 11:35am**

We introduce the Strahler number of a parity game as a measure of its structural complexity and argue that the progress-measure lifting algorithm can be adapted to solve parity games of Strahler number  $k$  in polynomial time when  $k$  is fixed. We also discuss the relationship between its Strahler number and Register index defined by Lehtinen. Furthermore, we will also discuss the improvement to the running time and the space needed to solve parity games of a given register index.

# Sunflowers, Daisies and Local Codes

**Marcel De Sena Dall’Agnol · Presentation at 11:50am**

Error-correcting codes are procedures that map messages (strings) to codewords (larger strings) in a way that makes the message recoverable even when the codeword is significantly corrupted. Codes that are locally decodable have additional structure: in order to recover a single character of the message, one need only read a small number of characters of the codeword. Combinatorial sunflowers are set systems where the pairwise intersection between any two sets coincides with the intersection among all of them. They have become objects of intense study, largely driven by the famous “sunflower conjecture” of Erdős and Rado. This presentation will discuss two mathematical objects - daisies (a relaxation of combinatorial sunflowers) and relaxed locally decodable codes - and the relationship between them. We highlight recent lower bounds for the size of such codes, obtained via a translation of algorithms for local codes into the language of set systems.

---

# Streaming Zero Knowledge

**Christopher Hickey · Presentation at 12:05am**

An interactive proof is a conversation between a powerful machine, the ‘helper’, and a ‘verifier’ with low resources. The aim of the conversation is for the helper to convince the verifier about the output of a (likely computationally or space intensive) function over some shared data set. The concept of streaming interactive proofs (SIPs) considers a verifier with very small space, who streams the shared data, as well as the messages sent by the prover. We introduce the new concept of Streaming Zero Knowledge for interactive proofs, which is the extension of regular SIPs to having the verifier learn no additional information about the data set besides the fact the prover is trying to prove. We show several examples and generalisations in order to showcase the strengths of this new model.

# Computational Security & Networks

Track 3  
**MB0.08 · 9:45am**

---

Session Chairs:  
Teddy Cunningham &  
Aparajita Haldar



# Protocol Transformation for Transiently Powered Wireless Sensor Networks

**David Richardson · Presentation at 9:50am**

Transiently powered wireless sensor networks (TPWSNs) introduce computing problems that are not faced by traditional wireless sensor networks. Nodes in a TPWSN may crash and later recover at an arbitrary time due to their power cell being drained and subsequently recharging to a point where the node can resume its task. Any state lost during the crash may have to be re-acquired upon restart. The overhead of recomputing lost state can be partially mitigated using techniques such as checkpointing, however, it is an expensive process itself. One further way to reduce this overhead is to (re-)design protocols and applications to be naturally resilient to transient power failures. In this presentation we demonstrate our transformation process by introducing TP-failure resilience to a broadcast protocol that performs poorly in a TPWSN.

---

# Real-time Forensic-Aware Logging System for Hypervisors in Cloud Environment

**Abdullah Al Hajri · Presentation at 10:05am**

Cloud computing helps organizations and individuals benefit from large scale computing power and capabilities in a fast and convenient manner by provisioning elastic and scalable services closer to them. Despite its global acceptance, there is some reluctance among organizations (i.e. government entities) that hold critical and confidential information from using this technology. Most of this hesitation originates from security and privacy concerns where any breach to their cloud-hosted systems could result in a devastating outcome. Event logging helps system administrators in addition to forensic investigators when looking for malicious behaviours and to build a consistent timeline of these events. Accordingly, log forensics in cloud systems and applications are becoming important. We want to explore existing solutions in logging hypervisor events and propose a real-time approach for logging systems in order to help investigators in tracking digital evidence in a more effective and reliable manner to be admissible in courts.

# Using Fingerprint for Banknote's Protection

**Shin Wang · Presentation at 10:20am**

Despite the wide adoption of electronic payment methods, banknote still is the most important payment instrument in the world. However, with the development of printing techniques for companies, the tendency of counterfeiting is increasing in recent years. Counterfeiting on banknotes could be traced back to the emerging of the banknote. Different from the legitimate banknotes, which need to be durable, distinguishable, even environment-friendly, the counterfeits are only required to simulate the product that is good enough for at least one transaction. The traditional approaches to distinguish counterfeit mainly includes visually checking and touchable feeling, which is not precise, even unpractical in some environments. To solve this problem, we propose a new method to prevent banknote's counterfeiting. The computer biometric and cryptography techniques are adopted to compensate for the traditional scrutiny approaches. The combination achieves cost-efficiency with high precision.

---

# Toward the Design of Auditable Systems

**Sara Alhajaili · Presentation at 10:35am**

Auditing is a process during which a systematic investigation of a system is performed to ascertain that the system satisfies some requirements. Auditing is a critical part of building dependable systems such as secure systems or reliable systems by creating recovery points or by allowing an analysis of the system behaviour. However, in many cases, the audit process does not result in any definite answer in cases of failures. For example, it is challenging to accurately detect errors in large scale cluster systems or open distributed systems such as IoT when systems logs are analysed. What is lacking is a framework that can guide the development of auditable systems. We define the auditability problem, identify conditions for a program to be auditable, and provide examples of auditable programs. Finally, we show an impossibility result for auditability.

## Towards the Selection of Monitors for Predicate Detection in Wireless Sensor Networks

**Alwaleed Alharbi** · Poster presented in foyer

Due to the nature of the environment in which wireless sensor networks (WSNs) are deployed, failures of the network are expected to occur. For the network to adapt following these failures, the network needs to be monitored. Monitoring of the network, including the applications running on top, usually consists of (i) providing the conditions to be monitored and (ii) choosing the monitors. We make the following contributions: (i) we formalise the monitors selection problem and show it to be NP-complete, (ii) we develop a heuristic for selecting monitors, (iii) we then show the viability of the solution by monitoring the correctness conditions for an implementation of a TDMA protocol. Results show the technique helps in achieving a high monitoring accuracy. When compared to a previous technique called TREE, which is geared towards the monitoring of quickly changing conditions, our technique uses more messages but has a higher monitoring accuracy.

---

## Reliable Many-to-Many Routing in Wireless Sensor Networks Using Ant Colony Optimisation

**Jasmine Grosso** · Poster presented in foyer

Wireless sensor networks (WSNs) have been widely studied in the context of many-to-one communication, in which multiple data sources send messages to a dedicated sink. However, there has been little research in the area of many-to-many communication. Many-to-many communication in WSNs is a growing application area, with examples including fire detection in both natural and urban areas, and the monitoring of heating and air conditioning within buildings. We propose a scalable many-to-many routing protocol that makes use of Ant Colony Optimisation (ACO). The protocol aggregates data sent from multiple sources into a single, shared *backbone* of nodes to reduce the number of packets sent, increasing network lifetime. Results from simulations using the COOJA Network simulator show that the protocol is able to achieve packet delivery ratios above 95%, with the algorithm becoming more efficient with larger networks, sending fewer packets relative to network size.

# Scheduling Dependent Tasks in Edge Networks

**Mohammed Mufareh A Maray · Poster presented in foyer**

A MEN consists of a number of base stations or cloudlets that execute tasks on behalf of users, i.e., users offload tasks to the MEN nodes which then execute them before returning the results to the users. We formalize the problem as a constraint satisfaction problem and we provide a heuristic for scheduling the dependent tasks. We conduct simulation experiments in the NS3 network simulator to study the performance of the proposed heuristic. Our results show that (i) it is impossible to obtain a schedule when the speed is very high, (ii) when the speed is low, it is better to allow the tasks to run locally on the mobile device and (iii) when a MEN schedule is obtained, the reduction in completion time is proportional to the size of a job. For instance, in our simulations, we have obtained a completion time reduction of ~45%.

---

# Secure Virtual Machines Allocation That Minimizes The Number Of Migrations To Mitigate The Risk Of Malicious Co-Residency In Cloud Computing

**Mansour Aldawood Aldawood · Poster presented in foyer**

The allocation of virtual machines in cloud computing aims to optimise the resources available by meeting certain service level agreements (SLAs). Keeping the workload of physical machines (PMs) stable and balanced among virtual machines (VM) is one of the main objectives. However, a secure allocation of the VMs is a vital aspect to avoid a malicious co-residency between VMs. Malicious co-residency meaning that a normal VM is allocated on the same PM with a malicious VM. This malicious co-residency could lead to a side-channel attack (SCA) where the malicious VM uses a side channel to attack the target VM. This work proposes an allocation that aims to mitigate the possibility of co-residency at initial allocation while reducing the number of virtual migrations in future stages.



# Machine Learning & Artificial Intelligence

Track 4  
**MB0.08 · 11:15am**

---

Session Chairs:  
David Purser &  
Harry Wilde

## Using GAN to Implement Data Augmentation in Human Activity Recognition

**Yujue Zhou · Presentation at 11:20am**

Human activity recognition (HAR) is learning profound high-level knowledge about human activity from raw sensor input. HAR is an unsupervised classification task in machine learning, which means the categories of data need to be manually labeled. However, the raw data of HAR is the sequential data from wearable sensors, and the labeling work of such data is very difficult and time-consuming. So collecting a large number of labeled activity recognition data has a huge challenge. Data augmentation is a strategy that enables practitioners to significantly increase the diversity of data available for training models, without actually collecting new data. The traditional data augmentation technologies are mainly designed for image data in the field of computer vision, which are not suitable for processing sequential data. Here we propose a method to use Generative Adversarial Network (GAN) to implement data augmentation in HAR tasks.

---

## Deep Emotion Distribution from Images

**Abeer Almowallad · Presentation at 11:35am**

Human Emotions Recognition (HER) in Human-Computer Interaction applications has interested many researchers in machine learning. One challenge of HER is to measure the intensity of each emotion since there is always a mixture of emotions present. Previous studies have dealt with this challenge by using Label Distribution Learning (LDL), whose goal is to measure the degree of each emotion. This has proven effective in these studies, however, it has done this by focusing on maximising entropy for a non-linear conditional probability mass function, which is weak in generalisation. We could avoid this by adapting the deep CNN to be an LDL problem, as we know that CNNs deal very well with generalisation. We use the minimisation of Kullback-Leibler divergence to act as the loss function, as it measures the distance between two distributions. Then we evaluate the performance by using the cosine coefficient to measure the similarity between two distributions.

## EV-Gait: Event-based Robust Gait Recognition using Dynamic Vision Sensors

**Bowen Du · Presentation at 11:50am**

In this paper, we introduce a new type of sensing modality, the Dynamic Vision Sensors (Event Cameras), for the task of gait recognition. Compared with the traditional RGB sensors, the event cameras have many unique advantages such as ultra low resources consumption, high temporal resolution and much larger dynamic range. However, those cameras only produce noisy and asynchronous events of intensity changes rather than frames, where conventional vision-based gait recognition algorithms can't be directly applied. To address this, we propose a new Event-based Gait Recognition approach, which exploits motion consistency to effectively remove noise, and uses a deep neural network to recognise gait from the event streams. To evaluate the performance of EV-Gait, we collect two event-based gait datasets, one from real-world experiments and the other by converting the publicly available RGB gait recognition benchmark CASIA-B.

---

## Applying Augmented Reality (AR) Technology in Computing Higher Education

**Sarah Alshamrani · Presentation at 12:05pm**

Augmented Reality is defined as using augmented video by covering the image with generated data in order to achieve a high performance three-dimensional image (Koll-Schretzenmayr and Casaulta-Meyer, 2016) Augmented reality will be applied to Computer science labs in Saudi Arabia as it's a new experience which may help the education field to be improved by deploying it in most subjects. The new user experience will be tested in Saudi Arabia and the ability to apply this tools in Labs will be tested too. The methodology followed in this research will be a design science methodology approach for information system and software engineering which define as a supportive in behavioural research its divided in three main phases: 1- Exploratory study: focus group & interviews 2- Development phase: develop the AR application and implementation with hands on experience will be done. 3- Evaluation phase: using analysis tool e.g. SPSS.

---

# English-Arabic Cross-language Plagiarism Detection

**Naif Alotaibi · Presentation at 1:25pm**

Development of the world wide web and information technology have contributed to the growth of digital libraries and automatic machine translation tools, which easily translate texts from language to another. Also, these have increased contents in different languages, it results in increasing cases of translated plagiarism. When textual content is converted into another language, words would be changed and replaced by their synonyms. Consequence, recognition plagiarism among texts in different languages will be more challenging, since each language has its own structure. There are numerous researches have been done on detection monolingual plagiarism in natural language, however, cross language plagiarism such as Arabic-English and Indian-English has received little attention. Few studies have been conducted to improve methods to detect cross language plagiarism in form of English-Arabic, whereas, these have not given inaccurate results.

---

# Reflective Writing Analysis Approach based on Semantic Concepts: An Evaluation of WordNet Affect Efficiency

**Huda Alrashidi · Presentation at 1:40pm**

Reflective Writing (RW) can be beneficial in increasing students' awareness of how they are learning to gain insight into their understandings. However, there is a lack of literature on an assessment criterion for reflective writing about computer science education. The goal of this study was to develop learning analytics (LA) tool for automated RW detection based on the RW framework to assess students' reflective writing in relation to computer science (CS). The RW framework was developed based on 12 semi-structured interviews. Findings from the interviews identified (a) three reflection levels and, (b) seven indicators relating to reflective writing in CS.

# TDAM: a Topic-Dependent Attention Model for Sentiment Analysis

**Gabriele Pergola · Presentation at 1:55pm**

We propose a topic-dependent attention model for sentiment classification and topic extraction. Our model assumes that a global topic embedding is shared across documents and employs an attention mechanism to derive local topic embedding for words and sentences. These are subsequently incorporated in a modified Gated Recurrent Unit (GRU) for sentiment classification and extraction of topics bearing different sentiment polarities. Those topics emerge from the words' local topic embeddings learned by the internal attention of the GRU cells in the context of a multi-task learning framework. In this paper, we present the hierarchical architecture, the new GRU unit and the experiments conducted on users' reviews which demonstrate classification performance on a par with the state-of-the-art methodologies for sentiment classification and topic coherence outperforming the current approaches for supervised topic extraction. In addition, our model is able to extract coherent aspect-sentiment clusters despite using no aspect-level annotations for training.

---

# Knowledge-Enhanced Search Graphs for Approximate Nearest Neighbour Search

**Aparajita Haldar · Presentation at 2:10pm**

In artificial intelligence, features learned from the data must be encoded in some meaningful way. To represent multimodal data, we consider a shared feature space representing a mixture of source modalities. The use of multimodal proximity graphs is proposed to make semantically-grounded, explainable comparisons between different modalities. Moreover, machine learning systems often benefit from world knowledge to incorporate common-sense reasoning about entities and their relationships. This information is leveraged from knowledge graph structures in this work to generate novel “knowledge-enhanced multimodal proximity graphs”. This new search graph is compared with state-of-the-art graph-based retrieval systems on the approximate nearest neighbour search task, in terms of efficiency and accuracy.

## Incorporating Common Sense to Identify Mental State in Affective Event

**Fitri Nurinsani Rachbini · Presentation at 2:25pm**

Affective events are activities that might affect people's mental states. When a sentence contains affective event, we can plausibly infer emotional states of the agent, for example "I got engaged" might infer happiness while "I got fired" can infer "frustration". While previous studies predict the polarity of affective event based on human-needs assessment, this study proposes to use human common sense to infer emotional states of affective event in daily dialog or/and narrative stories. Incorporating existing knowledge base of human common sense to neural network-based model, this study aims to: 1) Identifying the emotional states of affective event sentence in dialogue and/or narrative text. 2) Generating response or action of post-affective event taken by agents.

---

## Local Trend Inconsistency: A Prediction-driven Approach to Unsupervised Anomaly Detection in Multi-seasonal Time Series

**Wentai Wu · Presentation at 3:45pm**

On-line detection of anomalies in time series is a key technique in various event-sensitive scenarios such as robotic system monitoring, smart sensor networks and data center security. However, the increasing diversity of data sources and demands are making this task more challenging than ever. First, the rapid increase of unlabeled data makes supervised learning no longer suitable in many cases. Second, a great portion of time series have complex seasonality features. Third, on-line anomaly detection needs to be fast and reliable. In view of this, we adopt an unsupervised prediction-driven approach on the basis of a backbone model combining a series decomposition part and an inference part. We then propose a novel metric, Local Trend Inconsistency (LTI), along with a detection algorithm that efficiently computes LTI chronologically along the series and marks each data point with a score indicating its probability of being anomalous.

## Dynamic Bike Reposition:A Spatial-Temporal Reinforcement learning Approach

**Man Luo · Presentation at 4:00pm**

Bike-sharing systems are widely deployed in many major cities, while the jammed and empty stations in them lead to severe customer loss. Currently, operators try to constantly reposition bikes among stations when the system is operating. However, how to efficiently reposition to minimize the customer loss in a long period remains unsolved. We propose a spatio-temporal reinforcement learning based bike reposition model to deal with this problem. An inter-independent inner-balance clustering algorithm is proposed to cluster stations into groups. Clusters obtained have two properties, i.e. each cluster is inner-balanced and independent from the others. As there are many trikes repositioning in a very large system simultaneously, clustering is necessary to reduce the problem complexity.

---

## Online Transfer Learning for Concept Drifting Data Streams

**Helen McKay · Presentation at 4:15pm**

Transfer learning uses knowledge learnt in a source domain to aid predictions in a target domain. When both source and target domains are online, each are susceptible to concept drift, which may alter the mapping of knowledge between them. Drifts in online domains can make additional information available, necessitating knowledge transfer both from the source to the target and vice versa. To address this we introduce the Bi-directional Online Transfer Learning framework (BOTL), which uses knowledge learnt in each online domain to aid predictions in others. We also introduce two variants of BOTL that incorporate model culling to minimise negative transfer in frameworks with large numbers of domains. Empirical results are presented on real-world data predicting Time To Collision (TTC) from vehicle telemetry. The evaluation shows BOTL and its variants outperform the existing state-of-the-art technique.

## Graph Convolutional Network

**Amir Shirian · Presentation at 4:30pm**

The significance of emotion recognition for humans is reflected by the variety of applications such as presence of deception in interviews, human-machine conversation systems, driver safety, and automatic pain assessment. In general, facial expression can be recognized from multiple modalities such as image, speech, action units, and facial landmarks. Among these modalities, dynamic modeling of facial landmarks has received relatively less attention than others. Earlier methods employ different static image data and corresponding audio to form feature vector and then try to classify emotions based on them. Although many papers used static information to investigate emotions, we know each emotion is not necessarily a static process and has expansion, peak, and reduction phases as a temporal process. In our work, we investigate the importance of dynamic facial landmark modeling for emotion recognition task by using graph representation of data and processing with Graph convolutional network.

---

## Automatic Music Transcription and Correcting Errors in Existing Transcriptions

**Tom Wood · Poster presented in foyer**

Automatic Music Transcription (AMT) is the process of converting musical audio into some form of musical notation. We give an overview of the problem of AMT and we are going to look at how errors could be corrected in existing transcriptions using Recurrent Neural Networks (RNNs). More specifically, we use Long Short Term Memory (LSTM) Recurrent Neural Networks. Due to their ability to capture long term time dependencies, these are especially useful for musical applications as music is full of patterns that keep reappearing over time.

---

## Loss Prediction Based Asynchronous Stochastic Gradient Descent for Distributed Training of Deep Neural Networks

**Junyu Li · Poster presented in foyer**

Training Deep Neural Network is a computationally intensive and time-consuming task. Asynchronous Stochastic Gradient Descent (ASGD) is an effective solution to accelerate the training process since it enables the network to be trained in a distributed fashion, but with a main issue of the delayed update. We propose a novel algorithm, called LC-ASGD, to compensate for the delay based on Loss Prediction, which can effectively extend the delay duration for compensation. The algorithm is evaluated on the widely used networks and benchmark dataset. The experimental results show that LC-ASGD significantly improves over existing methods, especially when the network is trained with a large number of workers.

---



## The Abandon and Parallelization GAN

**Hao Wu · Poster presented in foyer**

We proposed a framework that depends on many GANs trained simultaneously, abandon some discriminators in each iteration to increase the diversity of the modes proposed to solve the collapse of the generators. We will also exchange the generators and discriminators to compose new GAN pairs. By using the operation of abandon and parallelization, which can urge generators to learn more features than before and save lots of time. We are still doing the fine-tuning in order to get better results.

---

## Complex Networks Community Structure - Is the UK Railway Network Well Structured?

**Aseel Alturki · Poster presented in foyer**

A study of the UK's Railway Network to investigate its community structure using a number of community detection algorithms considering weights (number of travellers) and without, then compare their results, hopefully by the end of the study their will be recommended changes in some of the routes. In addition, analysing the results of these algorithms will show the suitability of different types of algorithms to our data. And this might be applicable to any data with similar features.

---

## Fusion Network For Face-Based Age Estimation

**Haoyi Wang · Poster presented in foyer**

Convolutional Neural Networks (CNN) have been applied to age-related research as the core framework. Although faces are composed of numerous facial attributes, most works with CNNs still consider a face as a typical object and do not pay enough attention to facial regions that carry age-specific feature for this particular task. In this paper, we propose a novel CNN architecture called Fusion Network (FusionNet) to tackle the age estimation problem. Apart from the whole face image, the FusionNet successively takes several age-specific facial patches as part of the input to emphasize the age-specific features. Through experiments, we show that the FusionNet significantly outperforms other state-of-the-art models on the MORPH II benchmark.

---

## Aiming Beyond the Obvious: Identifying Non-Obvious Cases in Semantic Similarity Datasets

**Nicole Peinelt** · Poster presented in foyer

Existing datasets for scoring text pairs in terms of semantic similarity contain instances whose resolution differs according to the degree of difficulty. We propose to distinguish obvious from non-obvious text pairs. We characterise existing datasets in terms of containing difficult cases and find that recently proposed models struggle to capture the non-obvious cases of semantic similarity. We describe metrics that emphasise difficult cases and propose that these are used for evaluating systems for semantic similarity.

---

## The Role of AR Technology in Supporting Nursing Acquiring Clinical Skills Independently

**Ebtehal Quqandi** · Poster presented in foyer

The current learning approach of acquiring nursing clinical skills has limitations in supporting independent learning. The students use the manikin that does not represent real patient symptoms and they do not have enough experience to create a clear mental image needed for the simulated scenario. The study aims to overcome the limitations and introduces a new learning strategy by utilizing Mobile Augmented reality to support independent learning when acquiring clinical skills. An investigative study has been conducted to understand how the current learning approach supports independent learning. The outcomes were intended to help to understand the limitations and the possibility of integrating MAR technology. The research proposed a framework that will overcome the limitations, then used that framework to develop an application with AR. The result shows that students prefer using the proposed approach than the current approach in terms of supporting their independent learning

## Pose-Invariant Face Recognition

**Jiashu Liao · Poster presented in foyer**

The face recognition problem is one of the most widely studied topics in computer vision due to its wide potential applications and benefits to human beings in law enforcement, bio-metrics, marketing, and so forth. Recently, significant progress has been achieved in face recognition with deep learning-based methods. A recent study by Ding, Changxing has observed a significant drop, over 10%, in the performance of most algorithms from near-frontal face recognition (NFFR) to pose-invariant face recognition (PIFR) face recognition, while the human performance degraded slightly. This indicates that the PIFR remains to be a significant challenge in face recognition and warrants future study while recognizing identities under varied views is an inherent ability for human beings and other living beings.

---

## One Graph-One Signal

**Debaleena Roy · Poster presented in foyer**

This work introduces the GBT-L, a novel class of Graph-based Transform. The GBT-L is constructed using a 2D graph with unit edge weights and weighted self-loops in every vertex. The weighted self-loops are selected based on the residual values to be transformed. We also introduce a coding framework that uses a template-based strategy to predict residual blocks in the pixel and residual domains to avoid additional signalling to the decoder for reconstruction. Evaluation results, in terms of the percentage of preserved energy and mean square error, show that the GBT-L can outperform the DST, DCT and the Graph-based Separable Transform.

---

## A Framework for Enhancing the Capturing and Sharing of Best Teaching Practices amongst Universities' Instructors

**Nouf Almujaally · Poster presented in foyer**

Many universities still struggle in documenting, sharing and applying the teaching practices gained by instructors due to the absence of Knowledge Management Systems (KMSs). A major limitation of the previous KMSs is that they disregard the actual knowledge sharing behaviour of end-users. This shortcoming has been solved in the proposed Best Teaching Practices Management Framework (BTPMF) based on intensive literature review and academics' behaviour obtained from interviews. The BTPMF can be used as a guide when designing and implementing a new KMS for knowledge sharing activities in academic institutions.

---

# High Performance Computing & Databases

Track 5

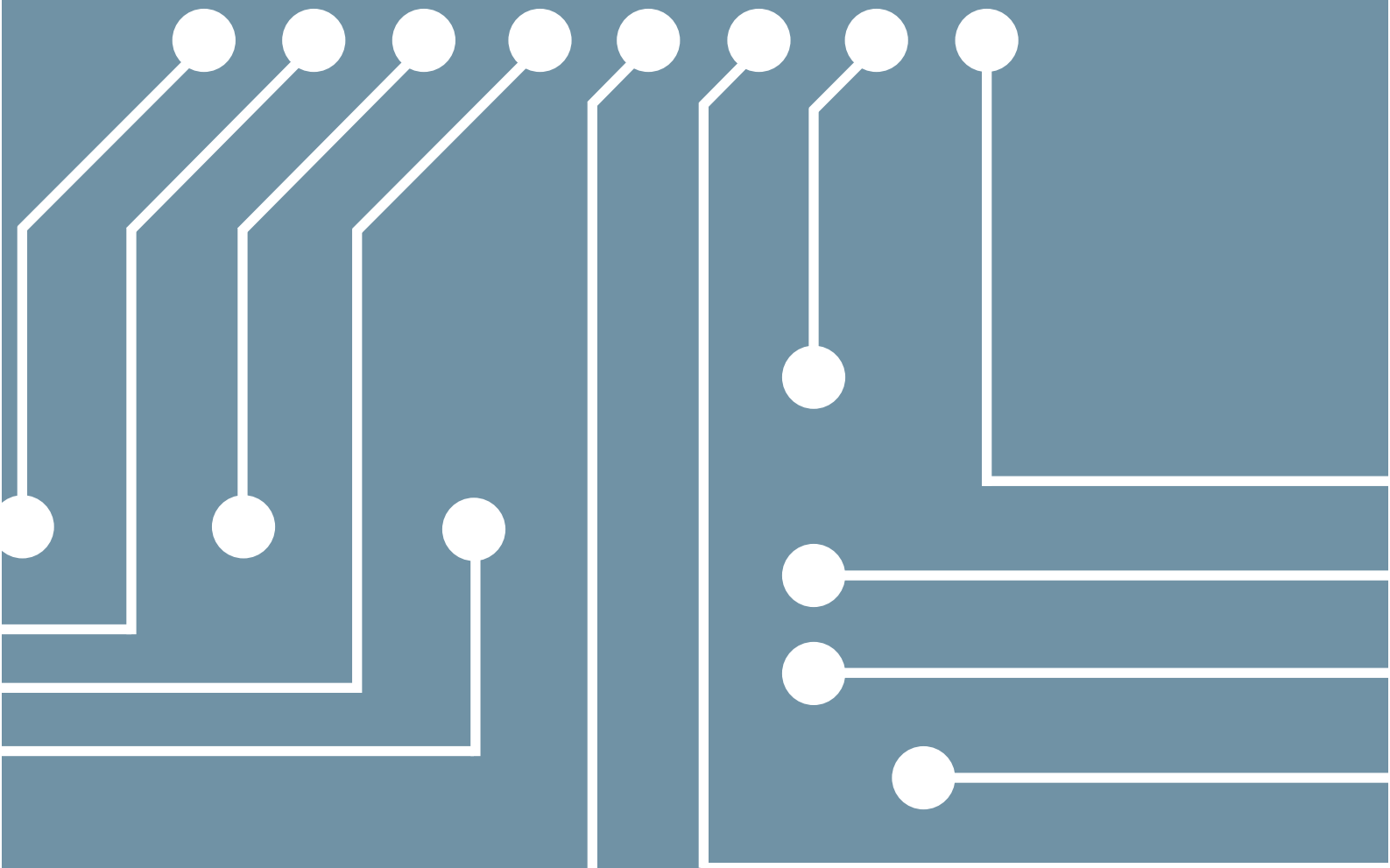
**MB0.01** · 1:20pm

---

Session Chairs:

Alex Dixon &

David Richardson



## Modeling HPC Networks

**Dean Chester · Presentation at 1:25pm**

Production machine performance has large variability. On the UK National Supercomputing Service, the time a job takes to complete can vary by as much as 53%. Load imbalance and shared resource contention are largely responsible, but we find that previous efforts to model application/architecture performance do not typically take these into account. In this research we model and simulate network contention, which allows us to explore the impact of multiple interacting jobs and approaches to alleviate these effects, including network re-design and communication-staging within applications. We show the utility of this work on a variety of systems and interacting applications.

---

## High Level FPGA Accelerator Design for Numerical Algorithms

**Kamalavasan Kamalakkannan · Presentation at 1:40pm**

FPGAs are becoming increasingly attractive as accelerator devices due to their high performance for certain classes of applications and low power consumption. However, a key limitation is the time and effort needed to utilize/program FPGAs, requiring skill in data-flow programming to achieve good performance. While latest synthesis tools allow the use of high-level languages such as C/C++/OpenCL, low-level customization is still required to gain performance benefits. In this work, we use the OPS domain specific language to generate optimized code targeting FPGAs. OPS provide a language/API with which a structured-mesh application can be written using domain specific constructs and then automatically generate parallel code for different architectures. This research aims to target Xilinx FPGA devices through OPS. In this presentation, I will detail initial performance of a hand-coded OpenCL application on an FPGA testbed, its OPS implementation and compare techniques explored to bridge its performance gap compared to traditional CPUs/GPUs.

## Higher-Order Particle Representation for Particle-in-Cell Simulations

**Dom Brown · Presentation at 1:55pm**

Particle-in-Cell (PIC) methods are commonly used to simulate the behaviour of plasmas under various physical conditions, enabling researchers to use simulation to carry out experiments that would be both time consuming and expensive. Additionally, the emergence of modern computational architectures that offer extreme levels of parallelism is causing a resurgence of higher-order schemes due to the increased arithmetic intensity that such methods provide. However, these PIC methods also require a smoother particle representation to achieve their theoretical convergence. In this talk we present an alternative approach to the representation of simulation particles for unstructured electrostatic/electromagnetic PIC simulations. In our modified PIC algorithm we represent particles as having a smooth shape. A unique feature of our approach is the representation of this shape by surrounding simulation particles with a set of virtual particles. The algorithm is implemented within Sandia National Laboratories' EMPIRE-PIC code, for electrostatic and electromagnetic simulations.

---

## Warwick Data Store: A HPC Library for Flexible Data Storage in Multi-Physics Applications

**Richard Kirk · Presentation at 2:10pm**

With the increasing complexity of memory architectures and multi-physics applications, developing data structures that are performant, portable, scalable, and support developer productivity, is difficult. In order to manage these complexities and allow rapid prototyping of different approaches we are building a lightweight and extensible C++ template library called the Warwick Data Store (WDS). WDS is designed to abstract details of the data structure away from the user, thus easing application development and optimisation. We show that WDS generates a minimal performance overhead, via a variety of different scientific benchmarks and proxy-applications.

## Using Resource Use Data and System Logs for HPC System Error Propagation and Recovery Diagnosis

**Edward Chuah · Presentation at 2:25pm**

Failure analysis plays an important role in the reliability of HPC systems and failure diagnosis using only system logs is known to be incomplete. Resource data provides another potential useful source of data for failure analysis. Recent work combining analysis of system logs with resource data show promising results. In this presentation, we describe EXERMEST for identifying patterns of significant errors and resource metrics and correlating these patterns by time with system failure and recovery. Application of EXERMEST on log-data on Ranger has yielded improved diagnosis over previous research. EXERMEST: (i) showed that more resource metrics and errors can only be identified by applying different feature extractors, (ii) identified CPU I/O bottlenecks and Lustre client eviction, (iii) identified packet drops and Lustre I/O errors, (iv) identified virtual memory and harddisk I/O errors, (v) showed that time-bins of different granularities are required for identifying the errors.

---

## Computational Offloading in Trusted Nodes of Edge Computing

**Ali Meree · Presentation at 3:45pm**

Trust reputation in the edge computing is a system that allows edge users to rate the edge nodes based on the previous history of the operations between them to build trust through reputation. We are focusing to build a robust trust model in the edge computing to minimise the delay time during the offloading process. Our model is a fully distributed reputation system and decentralised overlay network. The reputation machines we use is getting the recommendations from other peers by examining any direct interactions between the edge user and the edge node in a specific time. We are planning to use NS-3 to simulate our work and get the results that explain how is the malicious nodes in the edge increase the delay in the edge during the offloading process.

## DBEst: Revisiting Approximate Query Processing Engines with Machine Learning Models

**Qingzhi Ma · Presentation at 4:00pm**

In the era of big data, computing exact answers to analytical queries becomes prohibitively expensive. This greatly increases the value of approaches that can compute efficiently approximate, but highly-accurate, answers to analytical queries. Alas, the state of the art still suffers from many shortcomings: Errors are still high unless large memory investments are made. Many important analytics tasks are not supported. Hence, the following questions are crucial: Can we develop AQP engines that reduce response times by orders of magnitude, ensure high accuracy, and support most aggregate functions? With this study, we show that the answers to all the questions above can be positive. This study presents DBEst, a system based on Machine Learning models. It will discuss its limitations, promises, and how it can complement existing systems. It will substantiate its advantages using queries and data from the TPC-DS benchmark and real-life datasets.

---

## Query-Conscious Predictive Spatio-Temporal Quantization

**Shuang Wang · Presentation at 4:15pm**

PQ-trajectory, a predictive quantization based framework for query friendly compression and querying of dynamic trajectory data, is presented. PQ-trajectory includes a predictive spatio-temporal quantizer (PQ) that captures correlations of multiple consecutive trajectory points and quantizes the prediction errors within a narrow range into a compact codebook. It serves as a compact and dynamic structure for efficient query processing. A coordinate quadtree is introduced to reduce the information loss and processing time. A time-series density-based index is developed to avoid reconstructing the full trajectories to answer queries.

---

## Uniform Sampling over big Joins

**Ali Mohammadi Shanghoosabad · Presentation at 4:30pm**

Many-to-many join over large scale relations is one of the most expensive operations in databases, especially, when the number of involved relations increases. It becomes more critical when relations are distributed over different systems or tables are in different locations. Sampling is one of the efficient approaches to meet the needs in query processing, but there is a barrier when we use sampling to join relations: In 1999, Chaudhuri et al showed that the resulted relation from joining uniform samples of relations is not a uniform sample of joined relations. Since the number of distinct values (NDV) in many-to-many joins is not equal to the size of tables, here, we could introduce an approach based on generative modes that 1) does not care the size of tables, but NDVs 2) needs low amount of memory 3) can be used over distributed data 4) can be used on forgotten data.

---



# Urban Science

Track 6  
**MB0.07 · 1:20pm**

---

Session Chairs:  
Isabella Slattery &  
Liam Steadman



## How is Gender being Addressed in Low-Income Housing Indicators?

**Natalie Rothwell · Presentation at 1:25pm**

Gender equality is a core part of the United Nation's Sustainable Development Goals (SDGs) to end poverty by 2030. To achieve global goals such as the SDGs, we require data that can demonstrate current opportunities and challenges for women and men in the Global South, as well as to show progress towards desired improvements in their lives. Housing forms a core part of this because many of the world's poor live in inadequate shelter. Organisations including non-profits and housing associations respond to this via low-income housing programmes and use data-driven processes like indicators to ensure their work is meeting the needs of vulnerable groups including women, the elderly and disabled. This research combines these areas by examining how gender is being addressed in low-income housing indicators. An evaluation is carried out to identify, synthesise and assess indicator programmes before providing recommendations for scholarly and practitioner communities.

---

## Design Fiction for the Future of Learning Spaces

**Ivana Tosheva · Presentation at 1:40pm**

As technology increasingly infiltrates our everyday lives, it affects the way we communicate, acquire knowledge and create things. Higher education institutions are experiencing and will continue to experience great challenges in providing appropriate, up to date and innovative learning environments for their students. This research will aim to develop an approach to designing future learning spaces through exploring the current state of university learning facilities—how, when and why are these used. Design fiction, as a type of speculative design influenced by science fiction, will be used to encourage critical thinking for the future of learning environments. Combining the research findings with investigation into upcoming technologies influencing behaviour and space, the project will use design fiction to develop scenarios for the future outlook of learning facilities. Ultimately, the thesis will aspire to develop a new method for analysing architectural space and propose design thinking solutions for the future of learning.

# Smart City-Regions: On the Emerging Practice of Data-Informed Urbanism in Regional Governance

**Alexander Noll · Presentation at 1:55pm**

The concept of ‘smart cities’ has emerged in the past two decades as a new paradigm for city management and planning. Although no unanimous definition of the term exists, a common denominator is the belief that harnessing digital technology and its associated urban data could lead to improved service provision and a better understanding of the city. Meanwhile, the United Kingdom is passing through a process of devolution in which powers are decentralised from Whitehall to regions and city-regions while regional planning is subject to constant change. A new sub-national level of governance is emerging in the form of Combined Authorities which are granted certain decision-making and funding powers. This on-going research is placed at the interface of these two phenomena: a data-informed urbanism agenda and the re-emergence of strategic planning. It explores on a case-study basis which contributions a data-centric approach can have in supporting combined authorities.

---

## Online Platforms of Public Participation

**Jonathan Davies · Presentation at 2:10pm**

Trust and confidence in democratic institutions is at an all-time low. Meanwhile the complex issues faced by our cities remain unresolved. To confront these concerns, many argue that citizens should, through the use of digital tools, be better included within decision-making processes. This study evaluates one such tool, city-level online platforms, set up by local governments to enable citizens to submit ideas, rank priorities and allocate public resources. Are these platforms able to embody a form of a deliberative democracy, allowing for greater participation, knowledge co-production and the continual emergence of a collective intelligence? If not, is it possible to adapt the traditional models of offline participation and the enabling factors of collective intelligence to create a new model of online participation which may be used as a guideline for future platforms?

# Visualising the Limitations of Air Quality Sensor Data

**Nicole Hengesbach · Presentation at 2:25pm**

Air quality data visualisations currently largely fail to take into account how air quality sensor data may be contingent, incorrect, incomplete, relative, and locally and socially situated. In other words, the limitations of these data, despite their importance, are not represented within visualisations. Visualising air quality data through common maps and graphs omits much of the design space and, with that, the opportunities that may enable empowering interaction with the data and the matter itself. Drawing on (interdisciplinary) work from social sciences and digital humanities allows us to take fresh perspectives on how data are currently (not) visualised and on the conventions that have emerged. Rethinking how we visualise these data will further our understanding of urban data and involves exploring techniques that enable representing a range of contextual information as well as revisiting the choice of graphical primitives. I propose starting points for re-questioning visualisation designs.

---

# Source-Destination Balanced Facility Location

**Teddy Cunningham · Presentation at 3:45pm**

Facility location problems typically consider that users only interact with a facility network once per trip, such as when one visits a service station. However, the rise of new forms of mobility requires multiple interactions with the facility network -- for example, the use of bikesharing stations. Moreover, each trip involves the movement of a resource (e.g. a bike) that needs to be rebalanced across the facility network. Recently, there has been extensive research on the rebalancing problem, with post-processing solutions that cause additional traffic in the network. This research incorporates the rebalancing objective into the design of the facility network and efficiently optimises the facility locations using trajectory information. The source-destination balanced facility location problem is introduced and a solution that obtains near-optimal results for this problem in an accurate and efficient manner is also presented.

# Exploratory Spatial Analysis of Access to Physical and Digital Banking Channels in the UK

**Andra Sonea · Presentation at 4:00pm**

We measured the distance from the centroids small statistical areas of the UK to the first and the second closest point of access to physical banking channels (ATM, Post Office, branch). Access to digital banking was approximated from geographic customer segmentation based on the distances to the nearest mobile base station and local telephone exchanges. Exploratory spatial data analysis at both UK and regional level showed strong spatial patterns; significant rural/urban clusters could be identified as well as a North/South divide. No significant association was found between distance metrics and income and employment. The indicators used in this study can be used to identify areas vulnerable to the closure of the last points of access. Retail banking access should be treated as a joined-up system so that territorial coverage can be ensured and entire communities are not accidentally excluded from participation in the economy.

---

# The Demographic Evolution of Systems of Cities, a Graph-Theoretical Approach

**Matteo Mazzamurro · Presentation at 4:15pm**

Despite their huge variety in form and structure, systems of cities present some remarkable statistical regularities. Among the most impressive examples, we have Zipf's law, evidencing the inverse proportionality of a city's population to a power of its position in the national ranking, and Christaller's Central Place Theory, proposing a rigid but surprisingly successful service-provision-based hierarchisation of cities to explain their size, number, and spatial distribution. While surprising, some of these regularities actually arise naturally from purely geometrical considerations about the system. In this talk, I will present my first efforts in studying an example of statistical regularity linked to the time evolution of networks representing systems of several cities. More precisely, I will introduce the concept of graph entropy, compute it for several systems of cities, highlight common patterns, and compare the results to random models to establish whether they indeed provide us with insights about the systems.

# Collective Shortest Paths for Minimising Congestion on Networks

**Chris Conlan · Presentation at 4:30pm**

Congestion in cities remains a major problem. As urban populations continue to grow and the construction of new roads in built-up spaces is difficult it may only get worse. The impact of congestion is not just felt by drivers, but also wider society as road congestion contributes significantly to air pollution and can negatively impact the quality of life for residents. Navigational services typically seek to optimise routes for the individual user. In this work, we ask: can we optimise navigation queries at the system level? The goal is to reduce congestion and to better utilise the road networks. We make use of capacity-aware networks to develop two new algorithms which we have tested on both synthetic and real-world data. The algorithms are successfully able to target and reduce congestion, particularly on synthetic data, and have shown greater utilisation of the road networks across a range of metrics we've developed.

---

# Recalibrating urban monitoring frameworks in the Global South

**Philipp Ulbrich · Poster presented in foyer**

Marginalised urban neighbourhoods are more likely to be severely affected by disasters. This has led to a number of initiatives aimed at managing risk more effectively. Yet, despite such programmes being designed to reduce inequitable access to basic services, there is a disconnect between official risk management and the differential needs of marginalised urban residents. Urban monitoring faces a similar challenge: SDG11 aims to “make cities inclusive, safe, resilient and sustainable”. Concern has been expressed that some of its monitoring framework operates within rigid parameters that do not account for intra-urban variation (Ulbrich et al., 2019). The reliance on conventional methods for measuring urban progress may therefore perpetuate inequitable processes. This research addresses this challenge by analysing the implementation of urban risk and SDG11 monitoring frameworks, and comparing them to communities' experiences in Medellín, and thus provides a conceptual basis for recalibrating these frameworks in cities in the Global South.

# Planning in Practice for Resilience and Climate Risk in Extreme and 'Extreme-ing' Urban Environments

**Melissa Kenny** · Poster presented in foyer

Resilience has evolved to become a key priority within planning policy and practice in many urban contexts, especially in relation to climate change and extreme weather; manifesting itself within urban planning practice. In light of the increasing vulnerability to long and short-term climate related challenges, it is crucial to understand how resilience can be fully established within planning practice; at what scales should planners be involved and how can planning be integrated into other resilience related endeavours? It can be argued that currently, planning does not possess the impetus required to drive a fundamental and transformational shift in practice that is required to implement resilience. Whilst the seeds of resilience have been planted within planning practice, growth is slow and complex. By mainstreaming resilience approaches into everyday practices, planners, in combination with a full range of stakeholders, can help to ensure that urban systems remain functioning in the face of climate change or the event of an extreme weather shock.

---

## Smart Urbanism in a Post-Colonial Setting

**Oba Waiyaki** · Poster presented in foyer

The current master planning of cities, especially in Africa, is criticised for planning strategies that don't advocate community consultations, thus a lack of dynamism. These cities aspire to be "world class', 'smart' and 'global' as there's a need to emulate corporate office parks filled with placeless gleaming towers, showcasing a rejection of indigenous knowledge, craftsmanship, morphology and commonly used local materials." This positions global corporate culture as the modern future with the aim of achieving a borderless world. The rhetoric of leapfrogging onto a tech savvy utopia is the premise on which policy shifts are legitimised through technology seduction which result in urban transformations i.e. smart cities that are not fit for purpose.

## Digital Planning in the Smart City

**Ciaran Devlin · Poster presented in foyer**

Urban planning has traditionally been the method used to assess and manage how cities develop and incorporate new infrastructure and technological advances. The smart city has emerged as a concept which can profoundly enhance urban data collection and systems, providing an opportunity to improve urban planning and governance. This study aims to investigate these opportunities and implications by answering the following research questions; How can urban planning utilise the emerging technologies? How will the use of technology affect urban governance? What are the implications for the planning profession? The relationship between urban planning and smart city technology will be reciprocal, affecting the decisions and plans made in relation to urban development and smart technology. Therefore, it is vital to understand the interactions to ensure the most effective strategies for future urban planning and smart city implementation are adopted.

---

## Characterising Geotechnical Parameters Across The London Basin

**Bradley Sheridan · Poster presented in foyer**

A brief insight to the mechanical behaviour of reconstituted London Clay samples in an experimental study that concentrates on a series of one-dimensional (1D) compressibility tests that use reconstituted saturated clay samples obtained from the £600m project upgrade at Bank Station. For the experimental study, a series of 1D consolidation tests were conducted using a multi-stage loading (MSL) technique, with the maximum vertical stress of 1280kPa applied and on two different water contents (39 and 43%), in order to understand the stress and rate dependency behaviours. For these MSL (Multi-stage loading) tests, a conventional oedometer cell used in a room temperature environment. The tests results displayed a stress dependence of primary and secondary consolidation responses of each sample at different water contents



## Using the Weather Research and Forecasting Model to validate and improve high resolution weather forecasts around cities

**Timothy Sit** · Poster presented in foyer

Simulations of city-sized mesoscale meteorology at the sub-kilometre scale rarely succeeds due to the incompatibility of the numerics used at global scales with the requirements of doing simulations with local resolution below 1 kilometre. Numerical schemes used in the Weather Research and Forecasting (WRF) model are capable of simulating conditions as fine as 100 metres and successful in generating sufficient resolved scale turbulence, something which current Met Office Unified Model (MetUM) calculations fail at when resolving at these sub-kilometre scales. My plans will be to demonstrate that WRF can indeed predict high rainfall events at these finer resolutions. I will then compare WRF simulations with MetUM simulations of the same data. This will eventually lead into providing a set of reliable high-resolution reference simulations over a range of urban environments so that city councils can better prepare for hazardous weather events such as flooding.

---

## Modelling Business Rates in England with Big Spatial Data

**Shanaka Perera** · Poster presented in foyer

The continuing growth of the e-commerce industry has increasingly put pressure on traditional retail businesses. Additionally, the traditional retail industry carries a higher tax burden, as they require prime locations to attract a larger customer base, which reflects in higher business rates (non-domestic property tax). Current business rate revaluation has been criticized for misrepresenting true market prices. A better approach to model rateable values is hence needed. We introduce a large-scale, geospatial data set of UK nondomestic rateable values at the most granular level. We propose a state-of-the-art Fixed Rank Kriging model to cope with high-dimensionality and learn rateable values from spatial context and property characteristics. By accounting for spatial effects, our model improves on current business rates valuation practice and helps with making the process more fair and transparent.

## Solute Mixing in the Hyporheic Region

**Elisa Baioni · Poster presented in foyer**

We investigate the solute mixing within a porous medium during the presence of turbulent flow conditions. Our setup is representative of mixing processes taking place within the hyporheic zone and considers transport processes of dissolved chemicals close to the interface between a free fluid system and a porous medium. Our purpose is to investigate solute transport and mixing within the hyporheic zone. Turbulence is assumed as boundary condition between the free fluid and the porous column in order to simulate the effect of the river flow on the exchange interface in addition to its propagation within the porous medium. This latter effect corresponds to the critical area below and adjacent to a river or a stream where the interaction and mixing between the surface water and groundwater occurs.

---

## How Can Geo-Resources Build Sustainable and Resilient Cities?

**Katherine Ascott · Poster presented in foyer**

UK cities have seen the largest phase of growth since the Industrial Revolution. This growth is placing significant pressure on the resources that cities depend upon to meet their every-day needs. The subsurface offers geological resources (geo-resources) essential to support sustenance and urban expansion including; geo-materials, sub-surface space, groundwater and geothermal energy. This study evaluates the potential use of geo-resources in the context of sustainable and resilient urban design. Using a multiple case study framework, three methodologies have been implemented: 1) The creation of a geo-resource potential mapping model to explore site-specific geo-resources. 2) Interviews with a range of experts and representatives from urban stakeholder organisations to gain an understanding of the perception of geo-resources. 3) A document analysis of the inclusion of geo-resources across England's planning system. Combined, the results contribute towards the development of an urban geo-resource tool for sustainable and resilient cities.

# The role of public engagement in design discourse and urban planning

**Christian Caton · Poster presented in foyer**

In the past, urban design has struggled to find an equilibrium between creative innovation and purpose. The concept of deviation from designing forms we are accustomed to seeing is a struggle for designers through fear of negative feedback; deterring architects from taking risks or being creative in their field. This research aims to examine high-profile design solutions being proposed to deal with city-faced challenges and exploring how the public responds, playing a major role in design outcomes. This research will consider how capabilities of public engagement on social media might contribute in shaping people's behaviour and perceptions. With the capabilities of big data and social networking, this research seeks to implement methodologies that can prove useful for designers or planners in their future work; becoming more interconnected with the public and offering more freedom to experiment radically with their creativity in designing solutions for the city.

“The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them”

**Sir William Bragg**

---

“Nearly every example of faulty reasoning that has been published is accompanied by the phrase ‘of course’ or its equivalence”

**Donald Knuth**

Warwick Postgraduate Colloquium in Computer Science 2019

Sponsored by the Department of Computer Science and the Research Development Programme

[warwick.ac.uk/wpccs19](http://warwick.ac.uk/wpccs19)