



# #CISDC18

School of Computing and Information Systems  
6th Annual Doctoral Colloquium  
27 September 2018

# Report

@cis\_dc

fb.me/melbcisdc

CIS-DC@unimelb.edu.au

cis.unimelb.edu.au/colloquium

## WELCOME

Technology has become an integral part of our life and with every passing day, technology is revolutionizing our ways of living and learning. Being aware of its importance, a diverse interdisciplinary research is being pursued at School of Computing and Information Systems (CIS) of the University of Melbourne.

The Doctoral Colloquium (DC) is a key event of CIS, which seeks to provide a friendly, supportive and constructive environment where PhD students can present their research in progress. The DC provides an opportunity to doctoral students to gain valuable feedback on their theses, share research ideas and concepts, develop networks with colleagues, and learn about new, exciting and adjacent research areas.

Now in its sixth year, we have strived to make the DC even more industry oriented by introducing a special industry talk session. On the one hand, it allows our industry partners to touch base with our School's latest research. On the other hand, it can assist the students to bridge the transformation gaps between industry and academia.

To make this event more inclusive, students at all stages of their studies, from pre-confirmation all the way closer to thesis submission, were encouraged to participate in one of three tracks:

1. The paper track was aimed for students in advanced years to report on their significant milestones.
2. The poster track offered a more casual environment for students to socialise and discuss any pertinent aspects of their research.
3. Lastly, the 3 Minutes Research (3M<sup>r</sup>) track was aimed to sharpen students' communication skills; both new and early stage doctoral students could participate. This year, to make the 3 minutes track more challenging and competitive, we included immediate feedback and advice by the judges. Each presenter got reviews on their research pitch in terms of their topic, contents, key terms and making visual contact with the audience. The motive was the students can learn from the immediate feedback.

In total there were forty-one (41) submissions to the CIS-DC 2018: 15 papers, 18 posters and 8 three-minute research (3M<sup>r</sup>). The diversity of the research being undertaken at the CIS School is apparent from the range of projects presented at the event, such as students' emotions to route planning, avatars to brain networks.

This year, there were more than 150 participants at the DC, majority of them were graduate students from CIS. The rest were Academics and staff members of CIS, as well as our industry guests (sponsors and speakers).

The DC program began with a keynote speech by Prof Stan Skafidas. Prof Skafidas is the Managing Director of AMD Australia and cofounder Nitero company that commercialized 60GHz transceiver technology. The talk provides a guidance to the graduate students and briefs potential ways to create impact of their research.

Prof. Dr. Jürg von Känel from **IBM**, Dr. Kendra Vant from **Seek** and Chris Lewin from **Deloitte** were invited to provide their valuable guidance and advice to the students through the special industry talk session. The distinguished industry speakers discussed how to develop transferable skills, how to switch from one industry to another, the do's and don'ts of interview, how to develop decision making skills and lastly, they encouraged the women in computing how to stay actively engaged with industry.

The program concluded with the awards ceremony. Some of our sponsors (Branko Panich from **Grant Thornton Australia**, T'arne Whelan from **Australian Computer Society "ACS"** and Chris Lewin from **Deloitte**) honoured this ceremony and presented prizes to the winners.

We offer hearty thanks to our session chairs and reviewers for their excellent contributions. Sincere thanks also to our extremely generous sponsors. Without you the Colloquium would not have been possible.

Lastly, we express our special appreciation to the dedicated CIS staff for their outstanding support of the organising committee. We also owe a huge debt of gratitude to the small army of volunteers who put up their hands to make this event a reality.

On behalf of everyone involved, we thank you and we hope you had a rewarding and engaging experience at the Colloquium.

**The CIS DC 2018 Organising Committee**

## Table of Contents

Organising Committee.....	5
Sponsors.....	6
Location .....	7
Keynote Address.....	8
Event Schedule .....	10
Session Details.....	11
Paper Sessions .....	11
Three-Minute Research Session .....	17
Poster Session .....	21
Industry Talks Session .....	29
Best Awards Winners .....	32
More Information.....	33

## ORGANISING COMMITTEE



**Conference Chair**  
Ziad Al Bkhetan



**Programme Chair**  
Partha De



**Finance Chair**  
Sachini Jayasekara



**Finance Secretary**  
Tabinda Sarwar



**Venue and Logistics Chair**  
Sadia Nawaz



**Publicity and Communications Chair**  
Sameera Kannangara



**Publications Chair**  
Sadegh Motallebi

## SPONSORS

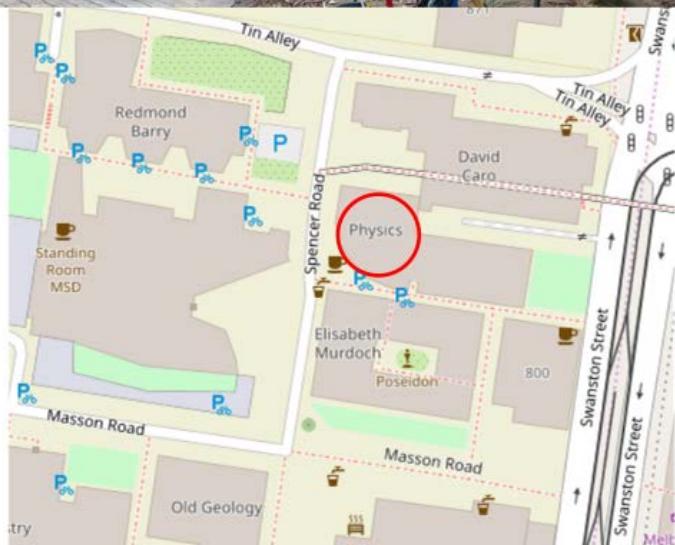


## LOCATION

### Physics South Building

Laby Theatre

Hercus Theatre



## KEYNOTE ADDRESS

**Title:** The many potential ways to impact

**Speaker:** Prof Stan Skafidas, Managing Director of AMD (Wireless) Australia

### Abstract

As a young starry eyed doctoral student, some of your first meetings involve your PhD supervisor describing the great importance of your chosen area of work, all the great things that will flow from it and how your expected work outputs will revolutionise your chosen field of research and, just maybe, even change the world. You are told the path to impact is simple - you will strive to publish your results in the top journals in your field. The journal "nature" has an impact factor of 40+ you are told.

Uncertain of what the term impact factor exactly means, other than the fact that 40 is greater than 4, the impact factor of the latest journal paper that the senior post doc in your lab has just published, you are excited by the prospects of the impact you are going to have. You are eager to race to the lab to begin your journey- your mission is clear.

However, is the path to impact for your research as clear cut as you have been led to believe? Are there other potential avenues to impact you could also be considering and working on simultaneously?

### Bio:

Professor Skafidas is the Managing Director of AMD Australia. He received a PhD from the University of Melbourne in 1997. Before joining NICTA in 2004, he was Chief Technology Officer at Bandspeed (1998–2004), a company he cofounded, based in Austin, Texas, which designed and manufactured semiconductor chipsets for wireless systems and products. At Bandspeed, Professor Skafidas co-invented Adaptive Frequency Hopping – a coexistence technology that allows Bluetooth devices to sense and avoid radio interference. This technology transformed Bluetooth into the robust and easy to use radio communication medium that it is today. AFH has been incorporated in every Bluetooth product since ver 1.2. In 2014 alone, there were over 3 billion devices sold incorporating this technology. The Bandspeed company was acquired by multinational semiconductor provider Broadcom Inc.

In 2008, a team led by Professor Skafidas developed the world's first completely integrated 60GHz transceiver on CMOS. This single chip wireless system was able to deliver 5Gbps data rate at cost of approximately \$1 whilst being only 5mm x 5mm in size. Professor Skafidas co-founded Nitero, the spin out company formed to commercialise 60GHz technology. Nitero received investment from leading venture funds such as Southern Cross Ventures and US based Austin Ventures. The technology and products based on this technology have been the recipient of multiple industry awards including the 2015 CES Innovation award. The CES Innovation Awards is the premier international annual competition honoring outstanding design and engineering in consumer technology products. Today, this technology, that has both

high throughput and low latency, is being used to produce chipsets for wireless virtual reality systems. In March 2017 the Nitero company was acquired by technology multinational Advanced Micro Devices (AMD) and has led to the establishment of AMD's global wireless research and development centre in Australia.



# EVENT SCHEDULE

09:30–09:45	<b>Registration</b> Foyer – Laby Theatre	
09:45–10:30	<b>Welcome and Keynote Address</b> Laby Theatre	
10:30–10:50	<b>Google Morning Tea Break</b> Foyer – Laby Theatre	
10:50–11:50	<b>Paper Session 1A</b> <b>Microsoft Research Centre for Social NUI Theatre</b> (Laby Theatre)	<b>Paper Session 1B</b> <b>Amazon Theatre</b> (Hercus Theatre)
11:50–12:05	<b>CIS-GReS Break</b> Foyer – Laby Theatre	
12:05–13:05	<b>Paper Session 2A</b> <b>Microsoft Research Centre for Social NUI Theatre</b> (Laby Theatre)	<b>Paper Session 2B</b> <b>Amazon Theatre</b> (Hercus Theatre)
13:05–14:05	<b>Lunch Break</b>	
14:05–14:55	<b>Three-Minute Research Session</b> Laby Theatre	
14:55–15.10	<b>CIS-GReS Afternoon Tea Break</b> Foyer – Laby Theatre	
15.10–16:30	<b>Industry Talks</b> Laby Theatre	
16:30–17:30	<b>Poster Session and CIS Evening Refreshments</b> Foyer – Laby Theatre	
17:30–18:00	<b>Award Presentations and Closing</b> Laby Theatre	

## SESSION DETAILS

### Paper Sessions

#### Data and Knowledge & Platforms and Systems

Microsoft Research Centre for Social NUI Theatre

(Laby Theatre)

### Paper 1

**Title:** *SLA-aware Scheduling of Big Data Jobs in Heterogeneous Clouds.*

**Authors:** Muhammed Tawfiqul Islam, Satish Narayana Srirama, Shanika Karunasekera and Rajkumar Buyya.

#### Abstract

Scheduling big data jobs in a cloud-deployed cluster is challenging due to both job and Virtual Machine (VM) heterogeneity. Big data framework schedulers fail to reduce the cost of VM usages in the cloud environment while satisfying the performance constraints (SLA) of each job. In this paper, we propose efficient scheduling algorithms that reduce the cost of resource usage in a cloud-deployed, multi-tenant big data processing cluster while improving job performance.

### Paper 2

**Title:** *Corner-based Building Height Estimation for Complex Street Scenes.*

**Authors:** Yunxiang Zhao, Jianzhong Qi and Rui Zhang.

#### Abstract

Estimating a building's height is important in many applications such as 3D city reconstruction, urban planning, and navigation. In this paper, we propose a corner-based building height estimation method which can better handle complex street view images. Experiments on real-world datasets show that our building height estimation algorithm outperforms the baseline algorithm consistently in terms of estimation accuracy.

### Paper 3

**Title:** *Instance Spaces for Multi-objective Black-box Optimisation.*

**Authors:** Estefania Yap, Andres Munoz Acosta and Kate Smith-Miles.

#### Abstract

The large quantity of existing algorithms introduces difficulty for identifying which algorithms should be run on real world problems that are encountered. Algorithms must be assessed objectively in order to make appropriate selections. However, the conclusions drawn on optimisation algorithm performance are subject to bias from

selection and diversity of the test problems on which they are evaluated. Rather than concluding superior algorithm performance by evaluating an algorithm over chosen test instances, it is more important to acknowledge that each algorithm will have strengths and weaknesses across the portfolio of all possible problems. Our work is the multi-objective, continuous black-box extension on the methodology of Smith-Miles (2014) for objective algorithm performance assessment.

## Paper 4

**Title:** *Multi-Perspective Process Model Discovery for Robotic Process Automation.*

**Authors:** Volodymyr Leno.

### Abstract

Robotic Process Automation (RPA) promises cost reductions and increased operational efficiency. RPA tools can automate repeatable tasks, thus reducing the error rates and increasing overall process performance. Being widely used in organizations, RPA solutions suffer from high time consumption allocated to the training of software robots (bots for short) - programmable software components that are used to automate tasks. Moreover, the models or instructions used for training may be inaccurate given that they are based on users' observations, which leads to an increase of the time spent on testing the bots. We propose to devise novel process mining techniques to extract the underlying user actions from user interface (UI) logs, which can then be used to train the bots automatically.

## Paper 5

**Title:** *Towards an Architecture of Explainable Agency*

**Authors:** Prashan Madumal, Tim Miller, Frank Vetere and Liz Sonenberg.

### Abstract

To generate 'trust' towards an intelligent system, it is essential to explain actions, behaviours and decisions of the system to the users in an understandable way. Decision making processes of these intelligent systems are often shrouded in 'black-box' models, which makes interpreting them by the end-user more difficult. Explanations are much needed and sometimes vital in human-agent collaborative scenarios where the human rely on or make judgement based on the decisions provided by the system. Explainable Artificial intelligence (XAI) aims to answer this problem by building explainable models of the underlying system enabling to effectively generate and communicate explanations. Although XAI field has been explored widely, much research and practise users the researchers' intuition of what amounts to a 'good' explanation. In this paper, we argue that XAI systems should include user centric components that inform explainable models and introduce an architecture for explainable agency.

## Paper 6

**Title:** *A Joint Model for Multimodal Document Quality Assessment.*

**Authors:** Aili Shen.

### Abstract

The quality of a document is affected by various factors, including grammaticality, readability, stylistics, and expertise depth, making the task of document quality assessment a complex one. In this paper, we explore this task in the context of assessing the quality of Wikipedia articles and academic papers. Observing that the visual rendering of a document can capture implicit quality indicators that are not present in the document text --- such as images, font choices, and visual layout --- we propose a joint model that combines the text content with a visual rendering of the document. Experimental results over Wikipedia and academic paper rating datasets reveal that textual and visual features are complementary, achieving state-of-the-art results.

## Paper 7

**Title:** *A Personalised Approach for Mining Top- $k$  Groups of Objects with Subspace Similarity.*

**Authors:** Tahrima Hashem, Lida Rashidi, Lars Kulik and James Bailey.

### Abstract

The identification of people (or objects) who are similar to a given person (or query object) is important for personalisation and more generally to understand the role of a person (or an object) among others. In this paper, we tackle the challenge of personalised pattern mining and propose a query-driven approach to mine objects with subspace similarity in a multi-variate categorical dataset.

## Paper 8

**Title:** *Estimating Task Difficulty Using Physiological Sensors.*

**Authors:** Namrata Srivastava, James Bailey, Eduardo Velloso, Sarah Erfani and Jason Lodge.

### Abstract

Digital learning environments provide rich and engaging experiences for students to develop different knowledge and skills. However, learning systems in these environments generally lack the capacity to assess student difficulties in real-time and also fail to provide automatic, context-specific feedback to learners. The lack of timely assessment and guidance can result in unproductive floundering and associated frustration. Existing studies show that physiological sensors like eye-trackers and thermal cameras have great potential for cognitive state estimation without any associated risks to the participants. To further understand the relationship, we have collected eye-tracking and thermal data of 100 participants while they were watching two-short video lectures. The instructional materials (video-lectures) used in the study

are based on previous research, which have reported significant changes in the cognitive levels of the students. Therefore, we expect to see same variability in physiological data of students while watching these video lectures.

## **People and Organisations**

### **Amazon Theatre**

(Hercus Theatre)

## **Paper 1**

**Title:** *Effect of Distinct Ambient Noise Types on Mobile Interaction.*

**Authors:** Zhanna Sarsenbayeva, Niels van Berkel, Eduardo Velloso, Vassilis Kostakos and Jorge Goncalves.

### **Abstract**

The adverse effect of ambient noise on humans has been extensively studied in fields like cognitive science, indicating a significant impact on cognitive performance, behaviour, and emotional state. Surprisingly, the effect of ambient noise has not been studied in the context of mobile interaction. As smartphones are ubiquitous by design, smartphone users are exposed to a wide variety of ambient noises while interacting with their devices. In this paper, we present a structured analysis of the effect of six distinct ambient noise types on typical smartphone usage tasks. The evaluated ambient noise types include variants of music, urban noise and speech. We analyse task completion time and errors, and find that different ambient noises affect users differently. For example, while speech and urban noise slow down text entry, being exposed to music reduces completion time in target acquisition tasks. Our study contributes to the growing research area on situational impairments by disclosing the effect of ambient noise on mobile interaction.

## **Paper 2**

**Title:** *Crowdsourcing Kinship Verification.*

**Authors:** Danula Hettiachchi, Niels van Berkel, Simo Hosio, Miguel Bordallo, Vassilis Kostakos and Jorge Goncalves.

### **Abstract**

Kinship verification is the process whereby a third party determines whether two people are related. Despite previous research in Psychology and Machine Vision, the factors affecting a person's verification ability are poorly understood. Through an online crowdsourcing study, we investigate the impact of gender, race and media type on kinship verification - taking into account the demographics of both raters and ratees. A total of 303 workers completed over 50,000 kinship verification tasks consisting of pairs of faces shown in images and videos from three widely used

datasets. Our results identify an own-race bias and a higher verification accuracy for same-gender image pairs than opposite-gender image pairs. Furthermore, we show that humans perform better when presented with videos instead of still images. Our findings contribute to future human kinship verification tasks, including time-critical use cases such as identifying missing persons.

## Paper 3

**Title:** *Mobile Device Framework to Address Knowledge Leakage in Organizations.*

**Authors:** Carlos Andres Agudelo.

### Abstract

The use of mobile devices such as laptops, tablets, and smartphones in organizations has surpassed the use of conventional workstations constrained within the corporate boundaries. Although convenient and efficient for mobile workers, such a phenomenon also poses a management risk that threatens the confidentiality of organizational knowledge. To minimize such risk, we propose a framework that aims organizations in safeguarding their knowledge assets against the threats produced by the misuse of mobile devices.

## Paper 4

**Title:** *Haplotype Estimation, Evaluation and Improvement via Consensus Methods.*

**Authors:** Ziad Al Bkhetan, Benjamin Goudey and Karin Verspoor.

### Abstract

Genetic studies can benefit from specific information about the inheritance of sequences of DNA nucleotides, called haplotypes. Experimental haplotypes are not readily available, yet they can be estimated via computational methods. In this study, we aggregate the outputs of different haplotype estimation tools to build a consensus sequence that is more accurate than the output of any single tool used in this study. We report an evaluation of seven haplotype estimation tools and a consensus of three of them on a real dataset.

## Paper 5

**Title:** *ETAS: Energy and Thermal-Aware Dynamic Virtual Machine Consolidation in Cloud Data Center with Proactive Hotspot Mitigation.*

**Authors:** Shashikant Ilager, Ramamohanarao Kotagiri and Rajkumar Buyya.

### Abstract

Data centers consume an enormous amount of energy to meet the ever-increasing demand for cloud resources. Computing and Cooling are the two main subsystems that largely contribute to energy consumption in a data center. Dynamic Virtual Machine (VM) consolidation is a widely adopted technique to reduce the energy consumption of computing systems. However, aggressive consolidation leads to the creation of local hotspots that has adverse effects on energy consumption and reliability of the system. These issues can be addressed through efficient and thermal-

aware consolidation methods. We propose an Energy and Thermal-Aware Scheduling (ETAS) algorithm that dynamically consolidates VMs to minimize the overall energy consumption while proactively preventing hotspots. We perform extensive experiments by using the real world traces with precise power and thermal models. The experimental results and empirical studies demonstrate that ETAS outperforms other state-of-the-art algorithms by reducing overall energy without any hotspot creation.

## Paper 6

**Title:** *SLA-Aware Live Migration in SDN-NFV-enabled Data Centers.*

**Authors:** Tianzhang He, Adel Nadjaran Toosi and Rajkumar Buyya.

### Abstract

With a huge demand for resource management of dynamic end-to-end services, the Software-Defined cloud data centers are introduced by leveraging the software solution, Software-Defined Networking (SDN) and Network Functions Virtualization (NFV). SDN allows the centralized controller of the switches to allocate bandwidth and route the network flows. Meanwhile, Network Function Virtualization (NFV) manages and orchestrates the network functionality which is virtualized and divided into multiple components called Virtual Network Functions (VNFs). Thus, SDN controller can also provide the instantiation and steering of port chaining through different VNFs on demand which is termed Service Function Chaining (SFC). In this research, we studied the overheads of live VM and Virtual Network Function (VNF) migration to minimize the impacts on SLAs. Moreover, we will investigate the SLA-aware algorithm of multiple migration planning meanwhile alleviate the influence on the quality of service.

## Paper 7

**Title:** *Brownout-Oriented Approach for Energy Efficient Cloud Data Centers.*

**Authors:** Minxian Xu.

### Abstract

Energy consumption of data centers has become a major concern for researchers. Energy efficient approaches have been applied to reduce data center power consumption, however, the dominant approaches, like dynamic voltage frequency scaling and VM consolidation cannot work well when the whole data center is overloaded. A self-adaptive approach called brownout, has been proposed and applied to data centers to manage resources. We take advantage of brownout to manage the energy and applications in data centers to improve system performance via dynamically turning on/off applications components/microservices. In this paper, we introduce one taxonomy and four research works that we have conducted based on brownout approach. The results in these papers show that brownout provides a promising direction for managing resources and energy in cloud data centers. The objective of this paper is attracting more attention on brownout-based approach in cloud data center.

## **Three-Minute Research Session**

Laby Theatre

### **Presentation 1**

**Title:** *Social esports spectatorship.*

**Presenter:** David Cumming.

#### **Abstract**

Esports (the professional, competitive practice of video gaming) is a rapidly growing phenomenon and is spectated by millions across the globe. We already know that many people watch esports, but we don't have a comprehensive idea of how people watch esports; most current esports literature assumes that it is just watched individualistically at home via live-streaming services like Twitch.tv. However, public communal spectating environments like stadiums and bars are becoming increasingly popular venues for esports consumption. To have a more comprehensive understanding of esports spectatorship, we must also consider consumption outside of home. My research places myself in the field; I go to these esports stadiums and bars to learn why people are there. So far, my research is revealing a social dimension to esports spectatorship at these public communal environment, defying the traditional "basement dwelling gamer" stereotype.

### **Presentation 2**

**Title:** *Using natural language processing and vetcompass to understand antimicrobial usage patterns in Australia.*

**Presenter:** Brian Hur.

#### **Abstract**

Antimicrobial usage for veterinarians in Australia is currently under scrutiny due to the rise of antimicrobial resistant pathogens. VetCompass Australia is collecting medical records from 181 Veterinary clinics across Australia. This includes over 3 million free text records in an unstructured format which are currently being analysed. To assist analysis, natural language processing techniques using rules based algorithms, word embeddings, neural networks and machine learning models are being developed and tested to structure, classify, and label the data. This should enable the monitoring of current and future antimicrobial usage patterns which will enable the monitoring of educational programs performed as part of the Australian Antimicrobial Stewardship Initiative. Models created through this project will then have performance compared to similar Natural Language Processing tasks in human medicine in an effort to improve clinical informatics through the use of veterinary data which is more readily accessible.

## **Presentation 3**

**Title:** *Style Transfer in Text.*

**Presenter:** Shraey Bhatia.

### **Abstract**

Let us take an image. Now we want to add filters, make the image more bright, change the background or say even add some external objects. We all have the apps now in our phones to do it. So essentially we are keeping the same content of the image but changing the stylistic properties. Now the task is to transpose the same idea from images to text. Hence, style transfer deals with changing the properties or rephrasing the text without changing the purpose or the primary context of the given text. For instance, you may want to change a sentence in Shakespearean English to non-shakespearean or convert a news headline from The Guardian into a more tabloid style of Buzzfeed or say even change the political slant of a statement from left wing into more right wing. This all comprises style transfer.

## **Presentation 4**

**Title:** *Political Trust Analysis using Natural Language Processing.*

**Presenter:** Shivashankar Subramanian.

### **Abstract**

Recent adoption of natural language processing methods has led to significant advances in the field of Computational Social Science, including Political Science. Political text can be generated by various stakeholders in a democracy --- political parties, politicians, people, news media and other organizations. In this thesis, we primarily focus on text generated by political parties and people. The sources of text for political parties include election manifestos, media releases, parliament proceedings, tweets, etc. And for people Twitter is a major source. Using such text we build intelligent systems targeted at two main goals: (a) improve the accountability of political parties thereby increase the trust in any democracy (b) measure citizen's engagement in a democracy which can be seen as a proxy for quantifying trust on political system. Finally, using this system we can relate citizen's trust to political parties' actions, which can be a critical index in any democracy.

## **Presentation 5**

**Title:** *Understanding the Structural Connectivity of Brain.*

**Presenter:** Tabinda Sarwar

### **Abstract**

The nervous system is a complex network of interconnected regions of brain that controls the functioning of the body by sending electronic signals to each other. Diffusion-weighted magnetic resonance imaging (DWI) is a non-invasive imaging tool that measures the microstructural properties of the brain. Tractography is a computational reconstruction method based on DWI that locates the white matter pathways revealing the structural connection of the brain, which can be used to study

neurological disorders such as Schizophrenia, Alzheimer. Many tractography algorithms have been developed, but quantification of their efficiency and accuracy in terms of structural connectivity is still under research due to the absence of ‘gold standard’. The aim of this research is to develop a ‘ground-truth’, that encapsulates the information of fiber bundles connecting different regions of the brain. This information is then used to evaluate the state-of-the-art algorithms and overcome their limitations to regional connectivity by devising new algorithms.

## Presentation 6

**Title:** *Panning for the gold standard in Machine Translation evaluation.*

**Presenter:** Nitika Mathur

### Abstract

It is very difficult to make progress in any field of research if we do not have a reliable way to evaluate it. When your task is subjective, for example translating from one language to another, there are multiple ways for a system to get it right, and multiple ways to get it wrong. It is very difficult to automate this evaluation, and so the most reliable way is to ask humans how they would rate a set of translations. However, if you have ever marked an exam or completed a survey, you know that responses will be very noisy. So we aggregate ratings of multiple annotators. You could use the mean or the majority vote, but some workers are more accurate than others. And different people have different ideas on what's good and what's bad. What if we learn to model the bias and accuracy of each annotator?

## Presentation 7

**Title:** *Explaining to Customers: Towards Ethical Big Data Analytics Services.*

**Presenter:** Sadaf Afrashteh.

### Abstract

Big data analytics has become a prevalent phenomenon in modern society and is a topic of much discussion in the information systems field. It uses algorithms for decision-making and targeting of customers. Algorithms process large-scale data sets and create efficiencies in the decision-making process for organisations, but are incomprehensible for most customers and sometimes opaque in nature. In response to this concern and for providing a greater transparency, the recent European Union regulations require that organisations communicate meaningful information to customers on the use of algorithms and the reasons behind decisions made about customers. In my research progress, I explore the use of explanations in big data analytics services. I rely on discourse ethics to argue that explanations can facilitate a balanced communication between organisations and customers, leading to transparency and trust for customers as well as customer engagement and reduced reputation risks for organisations.

## Presentation 8

**Title:** *Expert Finding in Bibliographic Network via Author-Topic-Citation Distribution.*

**Presenter:** Han Liu.

### Abstract

Citation plays an important role in evaluating an author's expertise. The current indicators such as h-index can only measure an author's expertise in total, while incapable of capturing expertise in a specific domain. The purpose of the study is to provide a new lens for expert finding by analyzing the distribution of citation of an author on each topic. In addition, we construct a topic co-occurrence network and calculate the similarity between academic topics by using network embedding. By applying the Author-Topic-Citation model, we can identify the relationship between academic topics, and rank experts at the topic level.

## **Poster Session**

### **Poster 1**

**Title:** *Towards Laboratory Testing for Mobile Context-Aware Applications.*

**Authors:** Chu Luo.

#### **Abstract**

Mobile context-aware applications (MCAA) can provide adaptive services to users in dynamic situations. As context recognition involves various data sources, large data size and complex algorithms, it is labourious and costly to test MCAA in real situations. Our work aims to enable testers to conduct low-cost and efficient tests by simulating real context in lab settings. In this poster we report a novel concept, a testing tool and the work in progress on unsolved problems.

### **Poster 2**

**Title:** *Interpretable Process Mining.*

**Authors:** Adriano Augusto.

#### **Abstract**

Companies information systems, daily, record business processes execution data (namely, event logs). This information is used to obtain insights into the quality and the performance of the processes, as well as a starting point for their improvements. Process mining is a branch of computer science, whose aim is to provide tools to analyze and monitor processes and their execution data. During the past decades, great focus was given to automated process discovery techniques, whose goal is to automatically generate a graphical representation of a process starting from its event log. Despite many studies have been published in this field, all the state-of-the-art automated process discovery techniques underperform when applied in real-life contexts. In this presentation, we show our latest developments in the field of automated process discovery, and our proposed solutions to tackle the state-of-the-art limitations, bring together their strengths and move forward to an ultimate automated process discovery technique.

### **Poster 3**

**Title:** *Medical Image Security using Spread Spectrum Steganography.*

**Authors:** Peter Eze.

#### **Abstract**

Steganography and Digital watermarking are alternative and Complimentary methods of providing security apart from cryptography. In Teleradiology where the privacy of patients' data as well as the integrity of the medical image scan need to be simultaneously preserved, Spread Spectrum Steganography data hiding method is the ideal solution. The objective of this research is to design a new algorithm that will detect malicious tampering as well as increase the amount of health record hidden in

an MRI image scan for transmission to a remote medical expert for diagnosis in Teleradiology.

## Poster 4

**Title:** *Generative Implication using LSTMs, Siamese Networks and Semantic Representations with Synonym Fuzzing.*

**Authors:** Anirudh Joshi, Timothy Baldwin, Richard Sinnott and Cecile Paris.

### Abstract

This poster describes a warrant classification system for SemEval 2018 Task 12, that attempts to learn semantic representations of reasons, claims and warrants. The system consists of 3 stacked LSTMs: one for the reason, one for the claim, and one shared Siamese Network for the 2 candidate warrants. Our main contribution is to force the embeddings into a shared feature space using vector operations, semantic similarity classification, Siamese networks, and multi-task learning. In doing so, we learn a form of generative implication, in encoding implication interrelationships between reasons, claims, and the associated correct and incorrect warrants. We augment the limited data in the task further by utilizing WordNet synonym "fuzzing". When applied to SemEval 2018 Task 12, our system performs well on the development data, and officially ranked 8th among 21 teams.

## Poster 5

**Title:** *Embedding Warfare Practices in Cyber Security Management.*

**Authors:** Abid Hussain Shah, Atif Ahmad and Sean Maynard.

### Abstract

Today's organizations use control-centred security management systems as a preventative shield against a broad spectrum of attacks. However, these have proven to be less effective against the customized and innovative strategies and operational techniques used by the recently emerging Advanced Persistent Threat (APTs). Aim of this research is to suggest that to combat APTs, organizations need a strategic-level paradigm shift away from prevention to response. Drawing on the concept of Information Operations (IO) from military studies, the research examines the applicability of capabilities of Information Operations in the civilian domain. The research objective is to develop a whole-of-organization's response through a model enhanced with IO capabilities such as deception and psychological operations to explain how the strategic and tactical challenges of APTs can be addressed.

## Poster 6

**Title:** *Using Biometric Sensors to Predict Attentional States of Researchers.*

**Authors:** Ebrahim Babaei.

### Abstract

The emergence of biometric sensors gives researchers the opportunity to have a better understanding of people mental and attentional state. Previous works such as tried to

detect their subjects' mood, perceived difficulty and progress in their tasks by biometric sensors to improve their productivity. In our work, want to capture researchers' biometric signals and their level of engagement in situ to have a model for researchers' engagement estimation.

## Poster 7

**Title:** *Route Assignment for Autonomous Vehicle Navigation.*

**Authors:** Sadegh Motallebi, Egemen Tanin, Jianzhong Qi and Ramamohanarao Kotagiri.

### Abstract

Traffic optimization is an important and challenging problem due to the highly dynamic and unpredictable nature of a traffic system. Existing traffic optimization approaches rely on historical traffic data to improve road planning and real-time traffic data to optimize traffic light scheduling. The upcoming era of autonomous vehicles offers a great opportunity for traffic optimization. The routes followed by autonomous vehicles can be assigned by a centralized system with better predictability. There is also more potential for users to follow exact navigation advice. We study the problem of route assignment for autonomous vehicles, with the aim to optimize traffic effectively. Since routes intersecting each other are a major source of delays, we transform the route assignment problem into a problem of finding road network paths with few intersections.

## Poster 8

**Title:** *Predicting Note-taking activities to students in learning environment using Bio Sensors.*

**Authors:** Anam Ahmad Khan, James Bailey and Eduardo Velloso.

### Abstract

The aim of this project is to analyse the note-taking behaviour of the participant during lectures so that an automated system could be build which during a learning task predicts when the participant should make notes and which note-taking strategy should be adopted by them. Prediction of note-taking activity can be made by capturing different aspects of students' mental states, such as their cognitive load. The content of student notes will be analysed along with their learning performance to recommend the best note-taking strategies for participants. Note-taking positively affects learning by increasing the attention of learners and encoding learning material to be stored in long-term memory. During reading and problem-solving note-taking is seen as a mean to reduce cognitive load by supporting retention of important information in working memory that will be used in elaborating final solution. In this work, we will consider the combination of bio-sensors such as thermal and eye to complement each other, aiming to build a robust system that can monitor the cognitive load of student and predict the need to take notes when a cognitive overload is observed due to the difficulty of the learning material.

## **Poster 9**

**Title:** *Secure Innovations in Organizations.*

**Authors:** Subramaniam Ramasubramanian, Atif Ahmad, Sean Maynard and Kevin Desouza.

### **Abstract**

The impact on organizations from the loss of intellectual property can be substantial. The theft of intellectual property is often referred to as a Knowledge Leakage problem. Knowledge leakage is the accidental or deliberate transfer of explicit and/or tacit organizational knowledge to unauthorized individuals inside or outside organizational boundaries. In knowledge-intensive organizations, where knowledge fuels innovation and the bottom line, leakage can erode competitive advantage and threaten the viability of the business firm. In this research, we take a practice perspective to explore how intellectual assets in the innovation process can be secured from leakage. We propose the Intellectual Asset Security Management (IASM) model to address the lack of guidance in literature on leakage mitigation strategies. The IASM model proposes a comprehensive suite of protection measures to protect intellectual assets (especially tacit knowledge) by mapping a range of information security paradigms to innovation activities

## **Poster 10**

**Title:** *Measuring quality of process models against event logs.*

**Authors:** Daniel Reissner.

### **Abstract**

Business process management is an interdisciplinary discipline that aims at improving organizational performance using a business process' lens. Within BPM, the merging field of process mining aims at inferring business process knowledge from execution data, so called event logs. A typical application of process mining is assessing the quality of a process model with respect to an event log. Measuring model quality is used to assess the quality of underlying automated discovery techniques or to assess the conformance against normative models. Quality assessment is typically done by comparing the model and the event log along four quality dimensions, i.e. fitness, precision, generalization and simplicity. To date, existing measures for these four dimensions suffer from a lack of scalability, orthogonality and comparability. Improving the measures will therefore two benefits. First, it will allow process analysts to better pinpoint differences between process models and event logs. Second, it will provide data scientists with a means to select the best process model among a set of automatically discovered models. In the light of the above, the goal of this poster is to address the following question: "How can the quality of a process model be assessed with respect to a given event log?"

## **Poster 11**

**Title:** *Are you confused, do you need help?*

**Authors:** Sadia Nawaz, James Bailey and Gregor Kennedy.

### **Abstract**

A considerable amount of research on emotions and learning has been undertaken in recent years. Confusion has been noted as a particularly important emotion as it can promote students' engagement in learning tasks. However, unresolved confusion may turn into frustration, boredom and ultimately disengagement. The exploratory part of this research investigated whether learning analytics could be used to successfully determine indicators or patterns of interactions that may be associated with confusion in a simulation-based learning environment. The findings of this study indicated that when taken individually, measures on specific learning tasks only hint at when students are struggling, but when taken together these indicators present a pattern of student interactions or a student profile that could be indicative of confusion. The second part of this research was more hypothesis driven and investigated if the likely moments of learners' struggle can be predicted in accordance with the design of the educational framework. The findings of this study indicated that learners' confidence in prior knowledge can contribute to how they experience confusion. Moreover, in accordance with the learning framework, confusion was mostly detected when learners discovered inconsistencies between their existing knowledge and the new information. The findings of this research may guide future works towards enhanced feedback in digital learning environments.

## **Poster 12**

**Title:** *Towards Robust and Privacy-preserving Text Representations.*

**Authors:** Yitong Li.

### **Abstract**

Written text often provides sufficient clues to identify the author, their gender, age, and other important attributes. Consequently, the authorship of training and evaluation corpora can have unforeseen impacts, including differing model performance for different user groups, as well as privacy implications. In this paper, we propose an approach to explicitly obscure important author characteristics at training time, such that representations learned are invariant to these attributes. Evaluating on two tasks, we show that this leads to increased privacy in the learned representations, as well as more robust models to varying evaluation conditions, including out-of-domain corpora.

## **Poster 13**

**Title:** *Understanding how technologies enable organisations to engage in sustainable supply chain management practices.*

**Authors:** Imairi Eitiveni, Sherah Kurnia and Rajkumar Buyya.

### **Abstract**

Supply chain activities have major environmental and social implications. For example, business processes involved in producing and distributing products often require excessive consumption of water and land and produce gas emissions and other waste that negatively impact the environment threatening biological diversity. Socially irresponsible practices are also prevalent including utilisation of hazardous materials, hiring child labours, violating workers' and communities' rights. Information system (IS) has the potential to help organisations minimise negative environmental and social impacts. However, there are limited studies investigating how IS/IT enables organisations to engage in sustainable supply chain practices. Furthermore, there is no empirical study investigating the relationship between IS/IT resources and IS/IT capability development that leads to the creation of supply chain wide sustainability capabilities.

## Poster 14

**Title:** *A Jointly Learned Context-Aware Place of Interest Embedding for Trip Recommendations.*

**Authors:** Jiayuan He, Jianzhong Qi and Kotagiri Ramamohanarao.

### Abstract

Trip recommendation is an important location-based service that helps relieve users from the time and efforts for trip planning. It aims to recommend a sequence of places of interest (POIs) for a user to visit that maximizes the user's satisfaction. When adding a POI to a recommended trip, it is essential to understand the context of the recommendation, including the POI popularity, other POIs co-occurring in the trip, and the preferences of the user. These contextual factors are learned separately in existing studies, while in reality, they impact jointly on a user's choice of a POI to visit. In this study, we propose a POI embedding model to jointly learn the impact of these contextual factors. We call the learned POI embedding a context-aware POI embedding. To showcase the effectiveness of this embedding, we apply it to generate trip recommendations given a user and a time budget. We propose two trip recommendation algorithms based on our context-aware POI embedding. The first algorithm finds the exact optimal trip by transforming and solving the trip recommendation problem as an integer linear programming problem. To achieve a high computation efficiency, the second algorithm finds a heuristically optimal trip based on adaptive large neighborhood search. We perform extensive experiments on real datasets. The results show that our proposed algorithms consistently outperform state-of-the-art algorithms in trip recommendation quality, with an advantage of up to 43% in F1-score.

## **Poster 15**

**Title:** *A time-motion based evaluation of the impact of information systems in an emergency department.*

**Authors:** Atlarelang Basetana Kesiilwe.

### **Abstract**

This is a time motion study based on a single case study methodology where consultants in an emergency department will be observed and interviewed with the aim of evaluating the impact of the implementation of an information system in an emergency department in a metropolitan hospital in Australia. The study is a 3-stage longitudinal study which is building on data from two previous studies that were conducted in the same setting in 2008 and 2012. The evaluations will be based on predefined task categories to investigate how information systems in health settings affects time spent on tasks to understand in detail how information systems fit in with the structure of work. The outcome of this research can inform guidelines for how information systems for hospitals can be designed and also, serve as some criteria for assessment before information system implementation.

## **Poster 16**

**Title:** *ICT and Sustainable Development Goals (SDGs): Insights from e-Health Initiatives in Indonesia.*

**Authors:** Luthfi Ramadani, Sherah Kurnia and Christoph Breidbach.

### **Abstract**

ICT has been acknowledged as the critical enabler of UN's Sustainable Development Goals (SDGs), perceived as a catalyst for their achievement. However, ICT and development evaluations mostly investigated restrictive indicators, mainly on the investment of internet and broadband as input and the Triple Bottom Line as the outputs, oversimplifying the meaning of development. This study aims to uncover the process of how ICT can be implemented successfully and subsequently contribute significantly to SDGs.

## **Poster 17**

**Title:** *Using a Traffic Simulator for Navigation Service.*

**Authors:** Abdullah Aldwyish, Hairuo Xie, Egemen Tanin, Shanika Karunasekera and Ramamohanarao Kotagiri.

### **Abstract**

Traffic congestion is a serious problem that is only expected to get worse in the future. Statistics shows that half of traffic congestion is caused by temporary disruptions like accidents. These events have dramatic impact on road network availability and cause huge delays for commuters. Also, they are usually unexpected and hard to manage by traffic authorities. State-of-the-art navigation systems started to provide real-time information about traffic conditions to help users make better routing decisions. However, traffic in the road network changes rapidly and the advice calculated now

may not be valid after few minutes. This is especially critical in the presence of traffic incidents, where the impact of the incident could cause traffic to propagate to nearby roads. Thus, it is important for navigation systems to consider the evolution and future impact of traffic events. In this work, we present a navigation system that uses faster than realtime simulations to predict the evolution of traffic events and help drivers proactively avoid congestion caused by events. The system can subscribe to real-time traffic information and forecast the traffic conditions using fast simulations. We evaluate our approach through extensive experiments to test the performance and accuracy of the simulator with real data obtained from TomTom Traffic API. Also, we test the quality of navigation advice in realistic settings and show that our solution is able to help drivers avoid congested areas in cases where even real-time update methods lead drivers to congested routes.

## Poster 18

**Title:** *Older Adults Designing Portrait and Cartoon Avatars for Self-expression.*

**Authors:** Romina Carrasco.

### Abstract

As people grow older, age related changes restrict the physical expression of the self. The physical appearance in later life becomes less relevant as older adults believe it no longer represents who they are but the years they have. Avatars can support expression of identity through the design of a full body depiction not restricted by the offline features of the ageing body. In order to understand if designing avatars could support the expression of identity among older adults, we engaged 18 participants (71-81 years old) in the design of a portrait avatar to replicate the ageing body and a cartoon avatar to support a more playful representation. Our findings suggest that the design of the portrait avatar extends the understanding of what ageing features should be accommodated to portray ageing in avatar customization systems. The design of the cartoon avatar suggests a new way of using online tools to overcome the invisibility of ageing. Finally, the method of designing first with the portrait and then with the cartoon constitutes a new and rich way to explore ageing through avatar design.

## INDUSTRY TALKS SESSION

**Speaker:** Prof. Dr. Jürg von Känel from IBM

**Title:** "Now that I have a shiny new degree - what can I do with it?"

### Abstract

Whether you are going into academic or industry research - being a trail breaker at the forefront of the technology is the job. In this talk I will outline the career path of an industry researcher, what it takes, why it's fun and where the current bleeding edge is.

### Bio:

Dr. Jürg von Känel is the associate director of the "IBM Research – Australia" lab in Melbourne. He studied math and computer science at ETH Zürich and holds a Ph.D. in Computer Science (1991).

Joined IBM in 1985 in Zürich Switzerland. In 1991 he moved to TJ Watson Research center in the US and most recently managed the relationship between Research and the financial services industries.

In 2004 he initiated an Enterprise Risk & Compliance Framework focused primarily on the financial industry. This lead to the Treasury & Risk magazine listing him as one of the 100 most influential people in finance in 2006.

He is a member of the IBM Academy of Technology, an honorary Enterprise Professor at the University of Melbourne in the Melbourne School of Engineering, and a member of the Industry Advisory Board of RMIT University.

Since June 2011, he has moved to Melbourne, Australia to lead the establishment and operation of the IBM Research lab in Australia. He manages the innovation around cognitive and AI industry solutions; as well as building tools to make sense out of structured and unstructured data for human decision makers.

In his scarce spare time, he and his wife invent, design and make mechanical puzzles  
<http://www.woodpuzzles.com>.



**Speaker:** Dr. Kendra Vant from Seek

**Title:** “Never die wondering – navigating a career that lasts a lifetime”

### **Summary**

In this talk, Kendra discussed the challenges of professional journey along with the ways to find a balance between family and career. She demonstrated the importance of being versatile and life learner. Based on her personal experience, she explained her journey from quantum physics to data science and how she switched from academia to industry. The talk highlights: individual must adapt to changing management policies, learning new skills and growing social needs. An active learning process is a key factor to dynamic, successful and rewarding experience. After the talk and some conversation with ex-academic colleagues, Kendra was inspired to write a blog post which expands on the pointes discussed in her talk.

<https://medium.com/@kendra.vant/on-leaving-academia-and-on-how-to-navigate-a-career-that-lasts-a-lifetime-6baed8130c13>

### **Bio:**

A firm conviction not to die wondering ‘what if?’ has given Kendra Vant a rich and varied career working with memorable people, companies and universities across New Zealand, Australia, the US and Malaysia. Through it all, her greatest satisfaction has come from working with smart people to solve difficult problems.

After doctoral research in experimental quantum physics at MIT and postdoctoral work in applied quantum computing at Los Alamos National Laboratory, Kendra was serendipitously placed to ride the tsunami of corporate interest in applying machine learning to create personalised experiences in an increasingly connected and digital world.

She has worked in insurance, banking, telecommunications, government, gaming and the airline industry and is currently Principal Data Scientist with Seek Asia Pacific and Americas, applying emerging techniques in natural language processing and deep learning to the problem of finding other people their dream job.



**Speaker:** Chris Lewin from Deloitte

**Title:** “”

### **Abstract**

#### **Bio:**

Chris is a Partner in the Consulting: Automate Practice in Deloitte Australia. He has extensive business and technology experience across complex business transformation programs, alongside a deep background in integration projects that sit at the intersection of data and process. He has deep experience in complex technology deployments such as Big Data, Analytics and Automation solutions across Telecommunications, FSI and Public Sector clients - generating operational insights and productivity outcomes.



## BEST AWARDS WINNERS

### *Best Paper Award*

Session 1: Sponsored by <b>Grant Thornton Australia</b>	<b>Namrata Srivastava</b>
Session2: Sponsored by <b>Deloitte Australia</b>	<b>Danula Hettiachchi</b>

### *3 Minutes Research Award*

Judges award: Sponsored by <b>Australian Computer Society (ACS)</b>	<b>Tabinda Sarwar</b>
People's choice award: Sponsored by <b>Silverpond</b>	<b>Shraey Bhatia</b>

### *Best Poster Award*

Data and Knowledge - Platforms and Systems: Sponsored by <b>Google</b>	<b>Sadia Nawaz</b> <b>Abdullah Aldwyish</b>
People and Organisations: Sponsored by <b>Melbourne Bioinformatics</b>	<b>Romina Carrasco</b>

## **MORE INFORMATION**

CIS-DC webpage: <https://cis.unimelb.edu.au/colloquium/>

Email: [CIS-DC@unimelb.edu.au](mailto:CIS-DC@unimelb.edu.au)

Twitter: [https://twitter.com/cis\\_dc](https://twitter.com/cis_dc)

Facebook: <https://www.facebook.com/melbcisdc>