# Proceedings of the
# ACM SIGKDD Workshop on
# Interactive Data Exploration and Analytics



These proceedings are not included in the ACM Digital Library.

*IDEA'18*, August 20, 2018, London, United Kingdom.

# ACM SIGKDD Workshop on Interactive Data Exploration and Analytics

## General Chairs

Jefrey Lijffijt (Ghent University)
Minsuk (Brian) Kahng (Georgia Tech)
Remco Chang (Tufts University)
Duen Horng (Polo) Chau (Georgia Tech)
Dafna Shahaf (Hebrew University of Jerusalem)
Christos Faloutsos (Carnegie Mellon University)

## Program Committee

Acar Tamersoy (Symantec, USA)
Adam Perer (IBM Research, USA)
Ahmad Mel (Ghent U, Belgium)
Andreas Henelius (Aalto U, Finland)
Arjun Srinivasan (Georgia Tech, USA)
Arnab Nandi (Ohio State, USA)
Bart Goethals (U of Antwerp, Belgium)
Bo Kang (Ghent U, Belgium)
Danai Koutra (U of Michigan, USA)
Daniel Keim (U of Konstanz, Germany)
Di Jin (U of Michigan, USA)
Emilia Oikarinen (FIOH, Finland)
Esther Galbrun (INRIA Nancy, France)
Fred Hohman (Georgia Tech, USA)
Geoff Webb (Monash U, Australia)
George Forman (Amazon, USA)
Hanghang Tong (Arizona State, USA)
Hannah Kim (Georgia Tech, USA)
Jaegul Choo (Korea U, South Korea)
James Abello (Rutgers U, USA)
Jia-Yu (Tim) Pan (Google, USA)
Jilles Vreeken (MPI-INF, Germany)
Josua Krause (NYU, USA)
Kai Puolamäki (Aalto U, Finland)
Kevin Roundy (Symantec, USA)
Kijung Shin (Carnegie Mellon U, USA)
Marco Cavallo (IBM Research, USA)
Mario Boley (MPI-INF, Germany)
Nan Cao (NYU Shanghai, China)
Robert Pienta (Symantec, USA)
Robin Vandaele (Ghent U, Belgium)
Saleema Amershi (Microsoft Research, USA)
Siegfried Nijssen (Catholic U of Louvain, Belgium)
Stefan Kramer (U of Mainz, Germany)
Steffen Koch (U of Stuttgart, Germany)
Stephan Günnemann (TU Munich, Germany)
Sucheta Soundarajam (Syracuse U, USA)
Thibault Sellam (Columbia U, USA)
Thomas Gärtner(U of Nottingham, UK)
Thomas Seidl (LMU Munich, Germany)
Tijl De Bie (Ghent U, Belgium)
Xu Chu (Georgia Tech, USA)
Zhiyuan (Jerry) Lin (Stanford U, USA)

# Preface

It has been premised by many that interactive data mining techniques that aptly integrate human intuition, visual pattern recognition, and background knowledge with the number crunching capabilities of modern computers—by means of visualization and feedback loops—could help researchers, businesses, journalists, and other data users gain insights unattainable otherwise. Yet, the development of methods and techniques that can function as building blocks for such interactive systems is still in its infancy.

The goal of the IDEA workshop is to encourage the development of methods and techniques to (1) more effectively and efficiently **communicate results** from analysis algorithms, e.g., visualization of patterns and models, (2) enable rich, efficient, and accurate **user feedback** to select and manipulate models or provide information about their relevance or interestingness, (3) provide **data analysis at interactive speed**. We aim to raise awareness of the need for data analysis algorithms that enable interactive use (take feedback into account), so the algorithms and the user work together to make discoveries in complex data. This *hybrid approach*—combining powerful techniques from data mining, visualization, and HCI—forms a new class of techniques that enable truly inter-active data analysis. The focus and emphasis is on interactivity and communication—information transfer to and from users, i.e., **techniques at the interface of data mining, visualization, and HCI**. In other words, we intend to explore how the best of these different but related domains can be combined such that *the sum is greater than the parts*.

Following the great success of IDEA at KDD 2013, 2014, 2015, 2016, and 2017 the main program of IDEA'18 consists of thirteen papers that cover various aspects of interactive data exploration and analytics. In addition there were three keynotes. Six papers were presented orally, and seven were presented during the interactive poster and demo session. These papers were selected from a total of 20 submissions after a thorough reviewing process. We sincerely thank the authors of the submissions and the attendees of the workshop. We wish to thank the members of our program committee for their help in selecting a set of high-quality papers. Furthermore, we are very grateful to Min Chen, Tijl De Bie, and Daniel Keim for engaging keynote presentations on the fundamental aspects of interactive data exploration, analysis, and visualization.

Jefrey Lijffijt &
Minsuk Kahng &
Remco Chang &
Polo Chau &
Dafna Shahaf &
Christos Faloutsos

Ghent, August 2018

## Invited Talk

# Building a Theoretical Foundation for Visual Analytics

Min Chen
Oxford University
min.chen@oerc.ox.ac.uk

## Abstract

In this talk, the speaker will first provide an overview of the theoretical discourses in the field of visualization and visual analytics, discussing the aspects of the Theoretical Foundation of Visual Analytics, including taxonomies and ontologies, principles and guidelines, conceptual models and theoretic frameworks, and quantitative laws and theoretic systems. The speaker will then make an argument that information theory can potentially bring the four main components of visual analytics (i.e., statistics, algorithms, visualization, and interaction) together under the same theoretic framework.

## Bio

Min Chen developed his academic career in Wales between 1984 and 2011. He is currently the professor of scientific visualization at Oxford University and a fellow of Pembroke College. His research interests include visualization, computer graphics and human-computer interaction. He has co-authored some 200 publications, including his recent contributions in areas such as theory of visualization, video visualization, visual analytics, and perception and cognition in visualization. He has worked on a broad spectrum of interdisciplinary research topics, ranging from the sciences to sports, and from digital humanities to cybersecurity. His services to the research community include papers co-chair of IEEE Visualization 2007 and 2008, Eurographics 2011, IEEE VAST 2014 and 2015; co-chair of Volume Graphics 1999 and 2006, EuroVis 2014; associate editor-in-chief of IEEE Transactions on Visualization and Computer Graphics; and co-director of Wales Research Institute of Visual Computing. He is currently an editor-in-chief of Computer Graphics Forum. He is a fellow of British Computer Society, European Computer Graphics Association, and Learned Society of Wales.

# Invited Talk

# An information-theoretic framework for data exploration: From itemsets to embeddings, from interestingness to privacy

Tijl De Bie

Ghent University

tijl.debie@ugent.be

## Abstract

Appropriately quantifying the interestingness (aka 'quality') of patterns (aka 'relations', 'aspects', 'views', 'data mining results'...) extracted from data is the crux of many data exploration tasks. For example, the interestingness of frequent itemsets in a binary data matrix, of dense communities in a network, of associations in relational a database, and of low-dimensional embeddings of a high-dimensional data set, all have been quantified in numerous ways, reflecting the intrinsic difficulty and ambiguity of this challenge. In this talk I will present an overview of the main results of my ongoing ERC project FORSIED ("Formalizing Subjective Interestingness in Exploratory Data Mining"), which aims to address this challenge by developing an information-theoretic unifying perspective on it. I will end by discussing some extensions of these results towards privacy-preserving data mining, fairness, and beyond.

## Bio

Tijl De Bie is currently Full Professor at the University of Ghent. Before moving to Ghent, he was a Reader at the University of Bristol, where he was appointed Lecturer (Assistant Professor) in January 2007. Before that, he was a postdoctoral researcher at the KU Leuven (Belgium) and the University of Southampton. He completed his PhD on machine learning and advanced optimization techniques in 2005 at the KU Leuven. During his PhD he also spent a combined total of about 1 year as a visiting research scholar in U.C. Berkeley and U.C. Davis. He is currently most actively interested in the formalization of subjective interestingness in exploratory data mining, and in the use of machine learning and data mining for music informatics as well as for web and social media mining. He currently holds a grant portfolio of around EUR 4M, including a prestigious ERC Consolidator Grant titled "Formalizing Subjective Interestingness in Exploratory Data Mining" (FORSIED), as well as an FWO Odysseus grant titled "Exploring Data: Theoretical Foundations and Applications to Web, multimedia, and Omics Data".

# Invited Talk

# Solving Problems with Visual Analytics: The Role of Visualization and Analytics in Exploring Big Data

Daniel Keim
University of Konstanz

keim@uni-konstanz.de

## Abstract

Never before in history data is generated and collected at such high volumes as it is today. As the volumes of data available to scientists increase, their effective use becomes more challenging. Keeping up to date with the flood of data, using standard tools for data analysis and exploration, is fraught with difficulty. Visual analytics seeks to provide people with better and more effective ways to understand and analyze large datasets, while also enabling them to immediately act upon their findings. Visual analytics integrates the analytic capabilities of the computer and the abilities of the human analyst, allowing novel discoveries and empowering individuals to take control of the analytical process. In the visual analysis process, it is not obvious what can be done by automated analysis and what should be done by interactive visual methods. In dealing with massive data, the use of automated methods is mandatory, but there is also a wide range of problems where the use of interactive visual methods is necessary. The talk discusses the different roles of visualization and analysis techniques and exemplifies them with several application examples, illustrating the exiting potential of current visual analysis techniques but also their limitations.

## Bio

Daniel Keim is professor and head of the Information Visualization and Data Analysis Research Group at the Computer Science Department of the University of Konstanz, Germany. He has been actively involved in data analysis and information visualization research for more than 30 years and developed a number of novel visual analysis techniques for very large data sets. He has been program co-chair of the IEEE InfoVis and IEEE VAST as well as the ACM SIGKDD conference, and he is chair of the IEEE VAST steering committee. He has been coordinator of the German Science Foundation funded Strategic Research Initiative “Scalable Visual Analytics” and scientific coordinator of the European Commission funded Coordination Action “Visual Analytics – Mastering the Information Age (VisMaster)”. Dr. Keim got his Ph.D. and habilitation degrees in computer science from the

University of Munich. Before joining the University of Konstanz, Dr. Keim was associate professor at the University of Halle, Germany and Senior Technology Consultant at AT&T Shannon Research Labs, NJ, USA.