



**ACM SIGKDD 2018 Workshop**  
August 20th, London UK

# Outlier Detection De-constructed

Leman Akoglu (Carnegie Mellon University)  
Evgeny Burnaev (Skolkovo Institute of Sci. & Tech.)  
Charu Aggarwal (IBM Research)  
Christos Faloutsos (Carnegie Mellon University)



KDD 2018 IS COMING TO LONDON!

**19 - 23 AUGUST, 2018**



# Anomalies: how do they differ?

glossary definitions of deconstruct :

“reduce (something) to its constituent parts  
in order to reinterpret it”

“analyze (a text or a linguistic or conceptual  
system) by deconstruction, typically in order  
to expose its hidden internal assumptions and  
contradictions and subvert its apparent  
significance or unity”

# ODD questions:

- How can we inspect the constituent parts of a detection model to expose its underlying **reasoning** to flag an outlier ?
- How can we trace the contribution of each input to the output (for a given example) ?
- How can we identify **root causes & generating mechanisms** of outliers for diagnosis and treatment ?
  - What is the extent to which we can draw **causal** (i.e. beyond descriptive) explanations ?
- How can we leverage **human interactions** with experts to explain outliers?
  - How can we incorporate **complex** user feedback ?



# Program: morning

<http://www.andrew.cmu.edu/user/lakoglu/odd/index.html>

8:30 Welcome & Opening remarks

---

8:45 **Keynote Presentation** (abstract)

*'Making sense of unusual suspects - Finding and Characterizing Outliers'*

by **Ira Assent**

---

9:30 **Coffee** 

---

10:00 **Keynote Presentation** (abstract)

*'The Outlier Description Problem - Complexity Results, Declarative Formulations and Applications '*

by **Ian Davidson**

---

10:45 **Contributed papers** (see list) (10+5 minutes each)

---

12:00 **Lunch** 

---

# Program: afternoon

<http://www.andrew.cmu.edu/user/lakoglu/odd/index.html>

13:00 **Keynote Presentation** (abstract)

**'Outlier Description and Interpretation'**

by **Jian Pei**

---

13:45 **Invited papers from KDD track** (see list) (10+5 minutes each)

---

14:30 **Coffee** 

---

15:00 **Keynote Presentation** (abstract)

**'Outlier Detection for Mining Social Misbehavior'**

by **Neil Shah**

---

15:45 **Contributed papers** (see list) (10+5 minutes each)

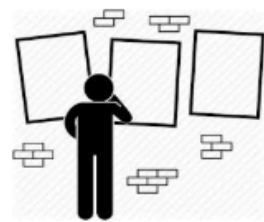
---

17:00 **Spotlight talks for poster papers** (see list) (2 minutes each)

---

17:10

**Poster session**



(for all contributed and poster papers)

---

18:00 **Closing**

---

# Ira Assent

Associate Professor  
Aarhus U.



8:45 **Keynote Presentation (abstract)**

**'Making sense of unusual suspects - Finding and Characterizing Outliers'**

by **Ira Assent**

Outlier detection provides the methodology to identify issues with data quality, to spot unusual data items or to detect rare patterns in data. Interpreting these outliers still remains a challenging task. Data analysts or domain experts trying to understand the nature of outliers cannot be expected to compare outliers to the remaining data to understand the differences. Outlier description therefore aims to provide additional information that supports interpretation by characterizing the differences.

In this talk, I discuss our recent work in outlier detection and description, and highlight some of the current challenges that I consider to be of prime importance for mature outlier description.

# Ian Davidson

Professor  
UC Davis



10:00 **Keynote Presentation (abstract)**

**'The Outlier Description Problem - Complexity Results, Declarative Formulations and Applications '**

by **Ian Davidson**

Outlier *detection* has been studied extensively and employed in diverse applications in the past decades. Yet the related and important outlier *description* problem has been an understudied problem. This problem often arises in practice when we have a small number of data instances that had been identified to be outliers and we wish to explain why they are outliers. Consider the example of automobile recalls where we already have examples of the outliers (recalled cars) and many examples of the non-outliers (non-recalled cars), with the aim to find a description of what made the instances anomalous.

We formalize the outlier description problem, present complexity results and declarative formulations which offers great flexibility in incorporating diverse scenarios arising in practice such as multiple explanations and human in the loop extensions.



# Jian Pei

Professor  
JD.com, Simon Fraser



13:00 **Keynote Presentation (abstract)**  
**'Outlier Description and Interpretation'**  
by **Jian Pei**

In this talk I will revisit the challenge of outlier description and interpretation. Essentially, can we manifest why an object is an outlier? We consider several possible ways to produce description and interpretation, including at the data level, at the feature/subspace, and at the model level. We discuss some example applications of such description and interpretation and examine their pros and cons. The talk will not come to a decisive conclusion about techniques. Instead, hopefully it will inspire a few challenges for future work.



# Neil Shah

Research Scientist  
Snap Inc.



15:00 **Keynote Presentation (abstract)**  
**'Outlier Detection for Mining Social Misbehavior'**  
by **Neil Shah**

Outlier detection has historically been used in a variety of domains for identifying rare samples in datasets. In this talk, I will discuss the particular application of outlier detection for identifying misbehavior and malicious actors in online social platforms. I will overview several previous works which leverage outlier detection techniques for identifying abusive following behavior and fake viewership on online platforms, and discuss technical takeaways and meta-lessons learned from these projects and more. The purpose of this talk is to communicate the great value that outlier detection approaches add in misbehavior detection, but also to pose a call to action for prioritizing the discovery or identification of specific outlying behaviors as opposed to outliers in general.