Editor's Notes

Welcome to the December 2017 issue of the ACM SIGMOD Record!

This issue starts with the Database Principles column featuring an article by Pierre Senellart on "Provenance and Probabilities in Relational Databases". This article describes different provenance formalisms, from Boolean provenance to provenance semirings and beyond, which can be used to answer a variety of questions regarding the output of a query. It also discusses representation systems for data provenance, circuits in particular, with a focus on its implementation, and how provenance is practically used for query evaluation in probabilistic databases. The article concludes that practical implementation of provenance management is very much possible because it introduces a relatively low overhead.

The Vision column features an article by Pitoura et al. on "Measuring Bias in Online Information". The work is motivated by the observation that as we live in an information age today, the majority of our diverse information needs are satisfied online by search engines, social networks and media, e-shops, and other online information providers. Bias in online information has recently become a pressing issue, with search engines, social networks and recommendation services being accused of exhibiting some form of bias. In this vision paper, the authors make the case for a systematic approach towards measuring bias, with a focus on formal measures for quantifying the various types of bias and the system components necessary for realizing them. The article closes by highlighting the related research challenges and open problems.

The Systems and Prototypes column features an article by Spyropoulos and Kotidis on the Digree system for scalable graph analytics, an issue critical to many application domains such as social networks and electronic commerce. The Digree system enables distributed execution of graph pattern matching queries in a cloud of interconnected graph databases. The system decomposes a graph query into independent sub-patterns, processes them in parallel on distributed graph databases, and finally synthesizes the results at a master node. By comparing to Graphframes, a package for Apache Spark on 18 VMs, Digree is shown to provide superior performance on real world datasets.

The Distinguished Profiles column features Dan Suciu, Professor at the University of Washington. Dan has two Test of Time Awards from PODS as well as Best Paper Awards from SIGMOD and ICDT. In this interview, Dan talked about his research projects and major research results, from XML to privacy to probabilistic data to data markets finally to scalable query processing. During the interview, Dan also discussed his favorite style of research, "a good combination of both theory and practice," because he believes that "the most difficult theory questions are those that are grounded in practice and that the most interesting systems are those that have a strong theoretical component."

The Open Forum column features an article by Sadiq et al. to call to action for promoting empiricism in data quality research. The authors identify two inter-related dimensions of empiricism that help locate the sweet-spot for empiricism in advancing data quality research and practice. These are the type of metric and the scope of method. A third aspect, namely, the nature of the data, exposes a data continuum that defines the setting in which the data quality metrics and methods can be evaluated. The article further presents the various ways in which the dimensions of empiricism can be positioned, thus providing a lens through which the role of empiricism in data quality re-

search can be studied. In order to gain a deeper insight into each of these positions, the authors reached out to thought leaders in data quality research to help elaborate on the motivation and rationale, key approaches, and possible challenges against each position. The viewpoints presented in this article are extracted from a series of interviews conducted with the experts and are supplemented with a review of relevant literature.

The Reports Column features two articles. The first report summarizes the presentations and discussions of the fourth workshop on Algorithms and Systems for MapReduce and Beyond (BeyondMR'17) held in conjunction with the 2017 SIGMOD/PODS conference. The goal of the workshop was to bring together researchers and practitioners to explore algorithms, computational models, languages and interfaces for systems that provide large-scale parallelization and fault tolerance. The program featured two invited talks, the first on current and future development in big data processing by Matei Zaharia from Databricks and Stanford University, and the second on computational models for big data analytics algorithms by Ke Yi from the Hong Kong University of Science and Technology. The second report summarizes a tutorial on "Commonsense Knowledge in Machine Intelligence" presented at the ACM Conference on Information and Knowledge Management (CIKM) in November 2017. This article is motivated by the viewpoint that the future of computing depends on our ability to exploit big data on the Web to enhance intelligent systems, that is, to endow machines with commonsense knowledge. The overview of the state of the art on Commonsense Knowledge (CSK) in Machine Intelligence provides insights into CSK acquisition, CSK in natural language processing, and applications of CSK, as well as the set of open issues.

On behalf of the SIGMOD Record Editorial board, I hope that you enjoy reading the December 2017 issue of the SIGMOD Record!

Your submissions to the SIGMOD Record are welcome via the submission site:
http://sigmod.hosting.acm.org/record

Prior to submission, please read the Editorial Policy on the SIGMOD Record's website:
https://sigmodrecord.org

Yanlei Diao

December 2017

Past SIGMOD Record Editors:

Ioana Manolescu (2009-2013)	Alexandros Labrinidis (2007–2009)	Mario Nascimento (2005–2007)
Ling Liu (2000–2004)	Michael Franklin (1996–2000)	Jennifer Widom (1995–1996)
Arie Segev (1989–1995)	Margaret H. Dunham (1986–1988)	Jon D. Clark (1984–1985)
Thomas J. Cook (1981–1983)	Douglas S. Kerr (1976-1978)	Randall Rustin (1974-1975)
Daniel O'Connell (1971–1973)	Harrison R. Morse (1969)	