In recent years, many database researchers have become fascinated by graphs, of which I am one. Like many others, I have spent years working with the relational model, including its data access methods, optimizing techniques, and query languages. The rise of the graph model has been somewhat disruptive, and our natural tendency is to ask, “Can we address new challenges using the existing relational model?” Unfortunately, many efforts along this direction do not seem to work well.

The advent of the Web, and in particular large social networks on the Web, highlights the urgency of developing a native graph database. The embracing of the graph model is anything but surprising. After all, data management is about modeling the data and the relationships between the data. Can there be a more natural model than the graph itself? The question, however, was not new. About three or four decades ago, network and hierarchical models, which are also graph based, fought and lost the battle against the relational model. Today, however, we are facing many new challenges that the relational model is not designed for. For example, as graphs are becoming increasingly larger, storing a graph in a relational table and using relational self-joins to traverse the graph are simply too costly.

Over the last decade, much research has been conducted on graphs. In particular, the study of large-scale social networks has been made possible, and many interesting and even surprising results have been published. However, the majority of research focuses on graph analytics using specific graph algorithms (for example, graph reachability, sub-graph homomorphism and matching), and not enough effort has been devoted to developing a new data model to better support graph analytics and applications on graphs.

This book is the first that approaches the challenges associated with graphs from a data management point of view; it connects the dots. As I am currently involved in building a native graph database engine, I encounter problems that arise from every possible aspect: data representation, indexing, transaction support, parallel query processing, and many others. All of them sound familiar to a database researcher, but the inherent change is fundamental as they originate from a new foundation. I found that this book contains a lot of timely information, aiding my efforts. To be clear, it does not offer the blueprint for building a graph database system, but it contains a bag of diamonds, enlightening the readers as they start exploring a field that may fundamentally change data management in the future.

Haixun Wang
Microsoft Research Asia
Haixun Wang, who earned a PhD in Computer Science from UCLA in 2000, joined Microsoft Research Asia in 2009. Before then, he spent nine years as a research staff member at IBM Research Center, where he was a technical assistant to Stuart Feldman, then-vice president of Computer Science, and to Mark Wegman, head of Computer Science. Wang’s interest lies in data management and mining, and he is working on building a large-scale knowledge base for advanced applications, including search. Wang is on the editorial board of IEEE Transactions on Knowledge and Data Engineering, Knowledge and Information Systems, and the Journal of Computer Science and Technology. He has held key leadership roles in various top conferences in his field.