Editorial

Hannes Voigt¹ · Marcus Paradies² · Theo Härder³

© Springer-Verlag GmbH Deutschland 2017

1 Schwerpunktthema: Big Graph Data Management

Graphen sind eine intuitive mathematische Abstraktion, um zu erfassen, wie die Dinge miteinander verbunden sind. In den vergangenen zehn Jahren hat sich der Schwerpunkt in vielen Datenmanagement-Anwendungen von einzelnen Entitäten und Summendaten auf die Verbindungen und Beziehungsnetzwerke zwischen Entitäten verschoben. Daher dient die Graph-Abstraktion heute als natürliches Datenmodell für eine zunehmende Zahl von Anwendungen. Diese reichen dabei von der interaktiven Datenverarbeitung bis zu analytischen Systemen. Graphenspezifische Anwendungsfälle finden sich in einer Vielzahl sehr verschiedener Domänen wie zum Beispiel die Analyse sozialer Netzwerke, Produktempfehlungen, Betrugserkennung, die Analyse von Geschäftsprozessen, die Optimierung von Lieferketten und Logistikprozessen und die Wissensrepräsentation. Viele dieser Anwendungsgebiete einer graphorientierten Datenrepräsentation finden sich in traditionellen Unternehmens-Szenarien

und unterstreichen die umfassende Bedeutung von Graphdaten, aktuell und in Zukunft.

Graphdatenverwaltung stellt eine neue Herausforderung für Datenmanagementsysteme dar. Anfragen und Auswertungen von Graphdaten beinhalten häufig aufwändige Fixpunkt-Berechnungen über die Graphstruktur. Auf der operativen Seite beinhalten typische interaktive Abfragen eine transitive Verschlussberechnung entlang von Pfaden. Übliche Auswertungen wie Pagerank oder Kürzester-Weg-Suche sind auch wesentlich komplexer als herkömmliche Aggregationsanfragen. Mit Blick auf Cache-effiziente Zugriffe und eine effiziente Parallelverarbeitung ist die unregelmäßige und verzerrte Struktur von Graphen eine besondere Herausforderung, insbesondere wenn der Graph groß ist und Änderungen unterworfen ist. Weitere Herausforderungen sind unter anderem deklarative Anfragesprachen für Graphanalysen, Topologieindizierung, Graphstrukturkompression und die gezielte Verwendung moderner Hardware.

Graphdatenverwaltung ist daher ein spannendes Forschungsfeld, jetzt und für die kommenden Jahre. Das Schwerpunktthema dieses Heftes möchte auf dieses Gebiet aufmerksam machen und zu weiterer Forschung motivieren. Das Heft bietet dafür einen Überblicksartikel und zwei anwendungsorientierte Beiträge. Der Überblicksartikel *Big Graph Data Analytics on Single Machines – An Overview* führt in das Gebiet Graphdatenverarbeitung auf leistungsstarken Einzelrechnern ein, erläutert den aktuellen Stand der Technik und zeigt weiterführende Forschungsthemen auf. Graphdatenverarbeitung auf Einzelrechnern ist vor allem in traditionellen Geschäftsanwendungen interessant, wo Graphen häufig aus existierenden Datenbeständen extrahiert werden und dynamische Skalierung eine zweitrangige Anforderung ist. Wer sich in dieses Gebiet einarbeiten möchte, für den bietet dieser Artikel einen kompakten Einstieg.

Hannes Voigt
hannes.voigt@tu-dresden.de

Marcus Paradies
m.paradies@sap.com

✉ Theo Härder
haerder@cs.uni-kl.de

¹ Fakultät Informatik, Professur für Datenbanken, TU Dresden, Nöthnitzer Str. 46, 01062 Dresden, Deutschland

² SAP SE, Dietmar-Hopp-Allee, 69190 Walldorf, Deutschland

³ AG Datenbanken und Informationssysteme, TU Kaiserslautern, 67663 Kaiserslautern, Deutschland

Der Beitrag *Dynamic Event-Activity Networks in Public Transportation* beschreibt die Verwendung von Graphdatenanalyse zur Unterstützung der Disponentenentscheidung hinsichtlich Anschlussgewährung bei Verspätungen im Bahnverkehr. Dieses Anwendungsszenario zeigt sehr schön, wie aufwendig Graphdatenverarbeitung sein kann und wie speziell die Lösungen zu deren Bewältigung ausfallen können. Damit ist der Beitrag als eine Aufforderung an die Datenbankgemeinde zu verstehen, das Leistungsspektrum von Graphdatenbanktechnologie deutlich zu erweitern.

Schließlich beschreibt der Beitrag *The Hydra.PowerGraph System – Building digital archives with typed, directed hypergraphs* die Nutzung der Graphabstraktion im Kontext von Wissensrepräsentation auf dem Feld digitaler Geisteswissenschaften. Hierbei wird die Modellierungsflexibilität, die Graphdatenmodelle bieten, deutlich. Gleichzeitig zeigen beide Anwendungsbeispiele aber auch die Datenmodellvielfalt – von Hypergraphen bis hin zu Event-Activity Netzwerken; von Propertygraphen bis hin zu RDF – auf, von der das Feld der Graphdaten geprägt ist. Diese Modellvielfalt stellt gleichfalls eine besondere, bisher noch ungelöste Herausforderung für die Universalität und die Interoperabilität von Graphdatenbanksystemen dar.

Wir hoffen so, mit dem Schwerpunktthema einige Anregungen und Anknüpfungspunkte für weitere Forschungsaktivitäten auf dem Feld der Graphdatenverwaltung und -verarbeitung geben zu können.

2 Fachbeiträge

Auf der BTW 2017 wurden drei „Best Papers“ ausgewählt, die wegen ihrer Qualität ins Datenbank-Spektrum aufgenommen werden sollten. Diese Beiträge wurden von den eingeladenen Autoren im Vergleich zur BTW-Version erweitert und verbessert, bevor sie nach einer erneuten Begutachtung zur Publikation angenommen wurden.

Im ersten dieser Beiträge *Efficient Storage and Analysis of Genome Data in Databases* untersuchen Sebastian Dorok (1), Sebastian Breß (2, 4), Jens Teubner (3), Horstfried Läßle (5), Gunter Saake (1) und Volker Markl (4, 2) ((1) Uni Magdeburg, (2) DFKI GmbH, (3) TU Dortmund, (4) TU Berlin, (5) Bayer HealthCare AG), wie Datenbanksysteme besser an die Anforderungen des Genom-Analyse-Prozesses angepasst werden können. Insbesondere integrieren Sie mit Hilfe eines spezialisierten Datenbankschemas Genom-spezifische Kompressionsverfahren, wodurch der Speicherplatzbedarf solcher Anwendungen um bis zu 35 % reduziert werden kann. Außerdem zeigen Sie, wie sich durch Nutzung spezieller Charakteristika der Genom-Daten der Analyse-Prozess beträchtlich beschleunigen lässt.

Effiziente Suchverfahren nach ähnlichen Objekten in unstrukturierten Daten wie z. B. Genom-Sequenzen, Text-Korpora oder Multimedia-Inhalten werden in den letzten Jahren immer dringender benötigt. Während beim Entwurf klassischer Indexverfahren die Reduktion der Externspeicherzugriffe im Mittelpunkt steht, ist in solchen Domänen die Reduzierung oder Minimierung von oft sehr rechenintensiven Distanz-Bestimmungen die wichtigste Optimierungsgröße. Der Beitrag *Reducing the distance calculations when searching an M-Tree* von Steffen Guhlemann (TU Dresden), Uwe Petersohn (TU Dresden) und Klaus Meyer-Wegener (Uni Erlangen-Nürnberg) schlägt Optimierungen für gewisse Indexstrukturen zur Ähnlichkeitssuche in metrischen Räumen vor. Insbesondere präsentieren die Autoren neue Algorithmen, die deutliche Leistungsverbesserungen bei allgemeinen Suchaufgaben (Suche nach k-nächsten Nachbar, Bereichssuche) in M-Baum-kompatiblen Datenstrukturen erzielen.

Distanz- und Zentralitätsberechnungen sind wichtige Bausteine für Graph-Datenbanken und für Graph-Analytik-Systeme. Im Beitrag *Efficient Batched Distance, Closeness and Betweenness Centrality Computation in Unweighted and Weighted Graphs* zeigen Manuel Then, Stephan Günemann, Alfons Kemper und Thomas Neumann (TU München), wie sich durch stapelweise Verarbeitung für solche Aufgaben beträchtliche Leistungssteigerungen erreichen lassen. Ihre empirischen Auswertungen bestätigen, dass die Laufzeit zur Bestimmung solcher allgemeiner Metriken um über eine Größenordnung reduziert werden kann.

Ein weiterer Fachbeitrag skizziert eine *Big Data*-Anwendung mit dem Titel *Daten wie Sand am Meer – Datenerhebung, -strukturierung, -management und Data Provenance für die Ostseeforschung*. Das Autorenkollektiv – Ilvio Bruder, Meike Klettke, Mark Lukas Möller, Frank Meyer und Andreas Heuer (Uni Rostock) sowie Susanne Jürgensmann und Susanne Feistel (Leibniz-Institut für Ostseeforschung, Warnemünde) – beschreibt am Beispiel verschiedener Projekte aus dem maritimen Umfeld spezielle Probleme beim Datenmanagement für heterogene Umweltdaten. Dabei werden vor allem die Nachvollziehbarkeit der Daten (Data Provenance) und die Berücksichtigung temporaler Aspekte bei der Erhebung, Speicherung und Auswertung der Daten betont.

3 Community-Beiträge in diesem Heft

Über die *BTW 2017 in Stuttgart*, die als 17. GI-Fachtagung „Datenbanksysteme für Business, Technologie und Web“ Anfang März 2017 stattfand, geben Holger Schwarz und Bernhard Mitschang einen ausführlichen Bericht mit vielen Bildern. Dabei heben Sie vor allem die drei Hauptvorträge

und die *Data Science Challenge* heraus, die erstmalig bei einer BTW stattfand.

Schließlich berichtet die Rubrik „Community“ unter *News* über weitere aktuelle Informationen, welche die DBIS-Gemeinde betreffen.

4 Künftige Schwerpunktthemen

4.1 Best Workshop Papers of BTW 2017

This special issue of the “Datenbank-Spektrum” is dedicated to the Best Papers of the Workshops running at the BTW 2017 at the University of Stuttgart. The selected Workshop contributions should be extended to match the format of regular DASP papers.

Paper format: 8–10 pages, double-column

Selection of the Best Papers by the Workshop chairs and the guest editor: April 15th, 2017

Deadline for submissions: June 1st, 2017

Issue delivery: DASP-3-2017 (November 2017)

Guest editor:

Theo Härder, University of Kaiserslautern, haerder@cs.uni-kl.de

4.2 Data Processing in Industrie 4.0

The Data Processing paradigm undergoes several changes recently. In the vision of Industrie 4.0 assets become smart and more autonomous. This leads to new application areas for analytics and data processing in general. We invite submissions on original research as well as overview articles covering topics from the following non-exclusive list:

- Industrie 4.0 Reference Architectures
- Sensor Data Streaming
- Sensor Data Management
- Digital Twin Technology
- Analytics in Industrie 4.0
 - Edge Analytics/Fog Computing
 - Sensor Data Analytics
 - Advanced Analytics

Expected size of the paper: 8–10 pages (double-column)
Contributions either in German or in English are welcome.

Deadline for submissions: Oct. 1st, 2017

Issue delivery: DASP-1-2018 (March 2018)

Guest editors:

Bernhard Mitschang, University of Stuttgart,
Bernhard.Mitschang@ipvs.uni-stuttgart.de
NN

4.3 Big Data Analytics and Exploration

We are looking for contributions from researchers and practitioners in the above described context, which may be submitted in German or in English.

Important dates:

- Deadline for submissions: February 1st, 2018
- Issue delivery: DASP-2-2018 (July 2018)

Paper format: 8–10 pages, double column (cf. the author guidelines at www.datenbank-spektrum.de)

Guest editors:

Sebastian Michel, TU Kaiserslautern, smichel@cs.uni-kl.de
Rainer Gemulla, Universität Mannheim, rgemulla@uni-mannheim.de