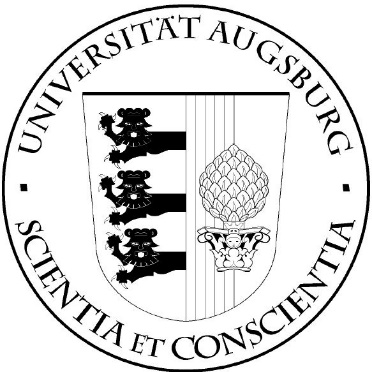
UNIVERSITÄTAUGSBURG

INSTITUT FÜR INFORMATIK



Bachelorarbeit im Studiengang Wirtschaftsinformatik

**Temporale Aggregate auf existierenden Graphberechnungs- systemen**

vorgelegt von

**Marc Aumüller**

Gutachter: Prof. Dr. Fischer

Zweitgutachter: Markus Endres

Betreuer: Prof. Dr. Fischer

Datum: 12. August 2018

**Zusammenfassung**

Ziel einer Zusammenfassung (engl. „Abstract“) ist es, dem Leser einen möglichst schnellen Überblick über den Inhalt der vorliegenden Arbeit zu verschaffen. Zusammenfassungen helfen dem Leser, Informationen aus einer Arbeit zu ziehen, ohne sie komplett lesen zu müssen. Der Leser kann so entscheiden, ob die gesamte Arbeit bedeutsam für ihn ist. Die Zusammenfassung sollte in der Regel **nicht länger als ½ Seite** sein. Halten Sie sich daran, kurze, prägnante Sätze zu schreiben. Aus der Zusammenfassung sollen dieselben Punkte hervorgehen, die auch in der Arbeit behandelt werden (Motivation, Methodik, Ergebnisse etc.) Als Richtlinie empfiehlt es sich, **1-2 kurze Sätze pro inhaltlichem Punkt** zu verwenden. Da Sie die Ergebnisse erst nach Fertigstellung Ihrer Arbeit kennen, wird dringend empfohlen, die Zusammenfassung **erst nach Beendigung Ihrer wissenschaftlichen Arbeit zu schreiben**. In der Zusammenfassung werden keine Zitate verwendet. Sämtliche Aussagen, die Sie in der Zusammenfassung von anderen Autoren übernehmen, müssen in der Arbeit belegt werden.

Inhaltsverzeichnis

[1 Einleitung und Zielsetzung der Arbeit 1](#_Toc518379608)

[2 Temporale Graphen 1](#_Toc518379609)

[2.1 Zeitaspekt 2](#_Toc518379610)

[2.2 Attribute 2](#_Toc518379611)

[3 Datenverarbeitung 2](#_Toc518379612)

[3.1 Snapshot Technik 2](#_Toc518379613)

[3.2 Streaming 3](#_Toc518379614)

[3.3 Sliding Window Aggregation 3](#_Toc518379615)

[4 Berechnungsmodelle temporaler Graphen 3](#_Toc518379616)

[4.1 Inkrementelle Berechnung 3](#_Toc518379617)

[4.1.1 Incremental View Maintenance 4](#_Toc518379618)

[4.2 Parallele Berechnungen 4](#_Toc518379619)

[5 Aggregate 4](#_Toc518379620)

[5.1 Hierarchische Aggregate 4](#_Toc518379621)

[5.2 Aggregatsprobleme auf Graphen 4](#_Toc518379622)

[5.3 Aggregate des temporalen Graphen 4](#_Toc518379623)

[6 Zusammenfassung Anforderungsprofil 5](#_Toc518379624)

[7 Existierende Systeme 5](#_Toc518379625)

[7.1 Kineograph 5](#_Toc518379626)

[7.2 GraphTau 5](#_Toc518379627)

[7.3 TGraph 5](#_Toc518379628)

[7.4 Chronos 5](#_Toc518379629)

[7.5 Naiad und Tornado 5](#_Toc518379630)

[7.6 Vergleich aller Systeme 6](#_Toc518379631)

[8 Diskussion 6](#_Toc518379632)

[Literaturverzeichnis 7](#_Toc518379633)

# Einleitung und Zielsetzung der Arbeit

In diesem Abschnitt: Thema der Arbeit ist temporale Aggregate auf existierenden Graphberechnungssystemen. Ziel ist es den Markt und den aktuellen Stand der Forschung zu analysieren, welche Systeme vorhanden sind, um temporale Aggregate auf dynamischen Graphen zu berechnen. Hierfür werden ausgewählte Systeme und Ansätze miteinander verglichen und anschließend bewertet. Da es für einen Vergleich der Systeme notwendig ist, einheitliche Anforderungen als Basis zu legen, liegt ein weiterer Schwerpunkt der Ausarbeitung darin, einen geeigneten Anforderungskatalog festzulegen, um anschließend die Systeme einordnen zu können. Daher wird die Arbeit in folgende Kapitel unterteilt. Kapitel 1 beschäftigt allgemein mit der Definition temporaler Graphen und dynamischer Netzwerke und legt fest, wie diese in der vorliegenden Arbeit definiert sind. Kapitel 2 wird sich mit der inkrementellen Berechnung beschäftigen, während Kapitel 3 auf temporale Aggregate eingeht und festlegen wird, welche von besonderem Interesse sind. Nachdem Kapitel 4 anschließend nochmals das Anforderungsprofil zusammenfasst, beginnt mit Kapitel 5 der zweite große Teil der Arbeit, in welchem die einzelnen Systeme vorgestellt und auf die vordefinierten Anforderungen hin geprüft werden. Kapitel 6 wird einen abschließenden Ausblick auf zukünftige Arbeiten geben.

# Temporale Graphen

Wie in [1] beschrieben, war eine Herausforderung der vorliegenden Ausarbeitung, dass die Forschung im Bereich temporaler Graphen auf der einen Seite schon sehr viele Erkenntnisse gewonnen hat und eine Vielzahl von Theorien hierzu entwickelt wurden, auf der anderen Seite aber keine einheitliche Formalisierung über die verschiedenen Forschungsbereiche und Disziplinen hinweg vorliegt. Als Beispiel wird je nach Quelle unter den Begriffen *temporal distance*, *reachability* *time* und *information* *latency* der gleiche Sachverhalt untersucht. (Schon allein für den Oberbegriff Temporale Graphen lassen sich in der Literatur zahlreiche Ausdrucksformen wie *temporal* *graphs*, *time*-*varying* *graphs*, *time-evolving graphs*, *dynamic networks* und *real-world graph* finden.) Alle Formalisierungen sind in dieser Arbeit zulässig und werden unter dem gemeinsamen Aspekt einer sich im Zeitverlauf verändernden Struktur des Graphen oder des Netzwerkes betrachtet. Was unter einer Veränderung der Struktur zu verstehen ist, wird in Abschnitt 5 genauer beschrieben. Zunächst wird in diesem Kapitel definiert, was unter einem temporalen Graphen genau zu verstehen ist.

## Dynamische Kanten und Knoten

Graph G mit (V,E) und Kanten und Knoten können hinzugefügt werden und wieder entfernt.

## Zeitaspekt

Da es in der Literatur eine Vielzahl von Überlegungen gibt, wie der Zeitaspekt sowohl im Zusammenhang mit temporalen Daten im Allgemeinen als auch im Kontext temporaler Graphen zu behandeln ist, werden in diesem Abschnitt die häufigsten Definitionen beschrieben und eingeordnet. Abschließend legt dieser Abschnitt fest, welche Definitionen für diese Ausarbeitung gelten.

Zum einen etwas zu Presence und Latency sagen und Fokus auf Presence richten.

Dann etwas zu valid time und transaction time sagen, siehe III. in [2].

Unterschied zwischen einem Zeitpunkt und Intervall und sagen, dass Zeit diskret ist und nicht stetig, dazu auch etwas in [2].

## Attribute

Text: statische und dynamische Attribute sowohl bei Knoten als auch bei Kanten möglich. Definieren.

# Datenverarbeitung und -speicherung

Fragen beantworten, wie: Wie liegen Daten vor? Daten liegen vor und werden rückwirkend untersucht? Werden alle Daten gespeichert? Stream?

## Snapshot Technik

Hier Snapshot Technik vorstellen.

## Streaming

Während bei der Snapshot-Technik davon ausgegangen wird, dass die benötigten Daten irgendwo gespeichert sind und jederzeit verfügbar, wann immer man sie benötigt. Sie können z.B. in einer Datenbank liegen. (Ist das so bei Snapshot Technik?) Im Gegensatz dazu können die ankommenden Daten beim Streaming nicht erfasst und gespeichert werden, um zum Zeitpunkt der Wahl auf sie zurückzugreifen. Wird der ankommende Datenstrom nicht sofort verarbeitet ist er für immer verloren. [3]

Bsp. AVG (= SUM/COUNT): wenn immer nur Daten hinzukommen. Dann SUM um den jeweiligen Wert erhöhen und COUNT + 1 und schon kann der neue AVG berechnet werden. [3]

## Sliding Window Aggregation

Ist eine Unterform von Streaming? Wie in Kapitel 2.3 beschrieben, kann man bei einem Stream nicht alle eintreffenden Daten speichern, weshalb die Form „Sliding Window“ entstanden ist. Ein Sliding Window können dann z.B. die letzte n Elemente oder die innerhalb der letzten t Zeiteinheiten eingetroffenen Daten sein. In [3] Kapitel 4.1.3.

# Berechnungsmodelle temporaler Graphen

Grundsätzlich lassen sich die Berechnungsmodelle nach zwei Hauptansätzen gruppieren. Zum einen können gewünschte Ergebnisse inkrementell und zum anderen parallel berechnet werden. In diesem Kapitel werden die zwei Herangehensweisen kurz beschrieben und herausgearbeitet, wieso es wünschenswert ist, dass die untersuchten Systeme ihre Berechnungen inkrementell durchführen.

## Inkrementelle Berechnung

Als erstes wird der Ansatz der inkrementellen Berechnung genauer betrachtet. Da im Bereich Verwaltung von Sichten in relationalen Datenbanken unter dem Konzept der View Maintenance eine ähnliche Idee umgesetzt wird, wird dies im Folgenden kurz analysiert, um mögliche Use Cases für temporale Graphen zu finden.

### Incremental View Maintenance

Da bei MV ähnliche Idee, kurze Analyse, was hierunter zu verstehen ist, und herausfiltern, welche Use Cases hier wichtig sind.

Da das Konzept und das Forschungsgebiet, wie beschrieben, dem Problem der vorliegenden Arbeit sehr ähnlich sind, werden zwei Use Cases der Incremental View Maintenance aufgegriffen und mit in die Ausarbeitung aufgenommen: zum einen Einfügen und Löschen einer Kante zwischen 2 Objekten (in Bezug auf den temporalen Graphen Knoten) und zum anderen Update eines Werts eines Objekts mit id1 von alterWert zu neuerWert. [4]

## Parallele Berechnungen

Text

# Aggregate

Dieses Kapitel dient dazu herauszufinden, für welche Aggregate auf temporalen Graphen es wünschenswert ist, dass sie die in Kapitel 6 analysierten Systeme berechnen können. In einem ersten Schritt werden 3 Formen des Graphen beschrieben, auf denen jeweils unterschiedliche Aggregate angewandt werden. Diese sind ein einzelner Knoten, eine Community und der ganze Graph.

Auf hierarchische Aggregate eingehen und verschiedene Aggregate in den Ebenen Knoten, Community und ganzer Graph eingehen.

## Hierarchische Aggregate

Text

## Aggregatsprobleme auf Graphen

Text Text.

Das finden von Communities und dabei auf die Erkenntnisse von Sune Lehman eingehen.

## Aggregate des temporalen Graphen

Nach der Betrachtung allgemeiner Aggregatsprobleme soll in diesem Abschnitt nun festgehalten werden, welche Aggregate die untersuchten Systeme als Mindestanforderung auf temporalen Graphen berechnen können sollen.

Das sind X und Y.

# Zusammenfassung Anforderungsprofil

Damit ein System als ein den temporalen Graphen verarbeitendes System klassifiziert werden kann, muss es die Mindestanforderung „temporales Graph-Berechnungssystem“ erfüllen. Unter dieser Anforderung wird verstanden, dass Kanten und Knoten in einen Graphen eingefügt und wieder gelöscht werden können und gleichzeitig modelliert werden kann, in welchen Zeiträumen die jeweiligen Knoten und Kanten gültig sind. Außerdem soll es möglich sein, dass Knoten und Kanten Attribute, wie in Kapitel 2.3 beschrieben, besitzen.

Außerdem wurde in Kapitel 4 deutlich, dass eine inkrementelle Datenverarbeitung wünschenswert ist. Daher stellt die Arbeit an die untersuchten Systeme die Anforderung, dass zumindest auf einer Ebene eine inkrementelle Berechnung möglich sein muss.

# Existierende Systeme

Im zweiten der Ausarbeitung sollen nun verschiedene Systeme vorgestellt und analysiert werden. Wie kam es zu dieser Auswahl der Systeme?

## Kineograph

Stelle mir eine einheitliche Tabelle vor, die schon aufzeigt, was ein System kann und was nicht

Im anschließenden Text wird das dann kurz verbal wiedergegeben und spannende Dinge, wie besondere Berechnungsmodelle kurz erklärt

## GraphTau

Text

## TGraph

Text

## Chronos

Text

## Naiad und Tornado

Text

## Systeme auf Flink und Spark

Tegra und Tink

## Vergleich aller Systeme

Am ende eine große Tabelle, in der alles zusammengefasst ist und in der alle Systeme zusammen aufgelistet.

# Diskussion

Dann abschließende Worte, dass unterschiedliche Anwendungsgebiete, kein einheitliches Konzept, einzelne Systeme und Anwendungen schwer wieder zu verwenden. Daher kann am Ende dieser Ausarbeitung nicht das eine System genannt werden, welches universell für jeden denkbaren Use Case eingesetzt werden kann. Was wäre hilfreich, wenn es das geben würde?

Text

Literaturverzeichnis

Chen, Hsinchun; Chiang, Roger H. L.; Storey, Veda C. (2012): Business Intelligence and Analytics: From Big Data to Big Impact. In: MIS Quarterly 36 (4), S. 1165–1188.

Fettke, Peter (2006): State-of-the-Art des State-of-the-Art. Eine Untersuchung der Forschungsmethode “Review” innerhalb der Wirtschaftsinformatik. In: Wirtschaftsinformatik 48 (4), S. 257–266.

Gregor, Shirley (2006): The Nature of Theory in Information Systems. In: MIS Quarterly 30 (3), S. p 611–642. Online verfügbar unter <http://www.jstor.org/stable/25148742>.

Gregor, Shirley; Hevner, Alan R. (2013): Positioning and Presenting Design Science Research for Maximum Impact. In: MIS Quarterly 37 (2), S. 337–356. Online verfügbar unter <http://dl.acm.org/citation.cfm?id=2535658.2535660>.

Heinrich, Lutz J. (2005): Forschungsmethodik einer Integrationsdisziplin: Ein Beitrag zur Geschichte der Wirtschaftsinformatik. In: NTM International Journal of History & Ethics of Natural Sciences, Technology & Medicine 13 (2), S. 104–117.

Hevner, Alan R.; March, Salvatore T.; Park, Jinsoo; Ram, Sudha (2004): Design Science in Information Systems Research. In: MIS Quarterly 28 (1), S. 75–105.

Sonnenberg, Christian; Vom Brocke, Jan (2012): Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research. In: Design Science Research in Information Systems. Advances in Theory and Practice, Bd. 7286. Berlin, Heidelberg: Springer Berlin Heidelberg (Lecture Notes in Computer Science), S. 381–397.

Webster, Jane; Watson, Richard T. (2002): Analyzing the past to prepare for the future: Writing a literature review. In: MIS Quarterly 26 (2), S. xiii–xxiii.

Wilde, Thomas; Hess, Thomas (2007): Forschungsmethoden der Wirtschaftsinformatik. In: Wirtschaftsinformatik 49 (4), S. 280–287. DOI: 10.1007/s11576-007-0064-z.

Ich versichere, dass ich die

<Bachelorarbeit/Hausarbeit/Seminararbeit/Masterarbeit>

selbständig und ohne unzulässige fremde Hilfe (siehe nächste Seite) angefertigt und keine anderen als die angegebenen Hilfsmittel benutzt habe.

Ich bin damit einverstanden, dass die elektronische Version meiner Arbeit mit Hilfe einer Software zur Plagiatserkennung überprüft wird. Ich bin einverstanden, dass zum Zweck der Überprüfung von Prüfungstexten auf Plagiate mein übermittelter Text dauerhaft in der von Turnitin Netherlands BV ausschließlich für die Wirtschaftswissenschaftliche Fakultät der Universität Augsburg geführten Datenbank gespeichert wird.

Ich bestätige, dass ich die folgenden Richtlinien für den Umgang mit Literaturquellen und Daten verstanden und eingehalten habe:

"Die benutzte Literatur und sonstige Hilfsquellen sind vollständig anzugeben; wörtlich, nahezu wörtlich oder sinngemäß dem Schrifttum entnommene Stellen sind kenntlich zu machen. Es muss durchgängig und unmissverständlich erkennbar sein, was an fremdem geistigem Eigentum übernommen wurde. Zitierten Autoren dürfen keine Aussagen zugeschrieben werden, die diese nicht oder nicht in der wiedergegebenen Form gemacht haben. Internetquellen sind mit vollständiger Adresse und dem Tag des Zugriffs zu versehen. Zugrunde liegende Daten müssen in überprüfbarer Weise dokumentiert werden. Bei gemeinschaftlichen Arbeiten ist der eigene Anteil des jeweiligen Autors deutlich zu machen. Darüber hinaus gelten die Vorgaben des vom betreuenden Lehrstuhl herausgegebenen 'Leitfaden zur Anfertigung von wissenschaftlichen Arbeiten'."

Ich habe den 'Leitfaden zur Anfertigung von wissenschaftlichen Arbeiten' vom betreuenden Lehrstuhl erhalten, ihn gelesen und verstanden.

Ich nehme zur Kenntnis, dass die Verletzung der hier genannten Richtlinien als versuchte Täuschung bzw. als Plagiat gewertet und mit Maßnahmen bis hin zur Exmatrikulation geahndet werden kann.

Ich versichere, dass die in Papierform abgegebene Arbeit identisch ist mit der elektronischen Fassung, die ich in die Plagiatserkennungssoftware hochgeladen bzw. meinem Betreuer übermittelt habe.

Ich stimme zu, dass die Ergebnisse dieser Arbeit von der betreuenden Professur anonymisiert zu Ausbildungszwecken verwendet werden dürfen.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ort, Datum Unterschrift Ort, Datum Unterschrift

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ort, Datum Unterschrift Ort, Datum Unterschrift

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ort, Datum Unterschrift Ort, Datum Unterschrift

**Erläuterungen**

Als unzulässige fremde Hilfe im Sinne dieser Plagiatserklärung gilt jede fremde Hilfe, die über die folgenden Hilfestellungen hinausgeht:

1. Beratung und Unterstützung durch den zuständigen Mitarbeiter am betreuenden Lehrstuhl.
2. Diskussion der Arbeit im Rahmen von Seminarveranstaltungen des betreuenden Lehrstuhls.
3. Ein Probelesen durch Kommilitonen oder andere Personen, die dem Autor ein mündliches Feedback geben und seine Thesen mit ihm diskutieren.
4. Ein Korrekturlesen durch Kommilitonen oder andere Personen oder auch Software, das sich auf Orthographie, Interpunktion und Grammatik beschränkt.
5. Hilfe und Unterstützung bei der technischen Erstellung der Arbeit (Satz, Layout, Druck und Bindung der Arbeit).