

Masterarbeit

Klassifizierung von Algorithmen des Graph-Mining zur Beantwortung logistischer Fragestellungen in Supply Chains

Christian Matthias Fahrenholz
Immatrikulationsnummer: 158611
Datum der Abgabe: 19.01.2021

Gutachter:

Name der Erstgutachterin:

Dr.-Ing. Dipl.-Inform. Anne Antonia Scheidler

Name des Zweitgutachters:

M.Sc. Joachim Hunker

Technische Universität Dortmund
Fakultät für Maschinenbau
Fachgebiet IT in Produktion und Logistik
<http://www.itpl.mb.tu-dortmund.de>

Inhaltsverzeichnis

1	Einleitung	1
2	Logistik.....	3
2.1	Grundlagen	3
2.2	Daten- und Informationsmanagement in der Logistik.....	7
2.3	Supply Chain Management	10
2.4	Logistische Fragestellungen des Supply Chain Managements.....	13
2.4.1	Aufgaben und Modelle	13
2.4.2	Fragestellungen des Supply Chain Design	16
2.4.3	Fragestellungen des Supply Chain Planning	18
3	Graphbasierte Verfahren	23
3.1	Grundlagen der Graphentheorie	23
3.2	Graphbasierte Modelle im Supply Chain Management	27
3.3	Graphalgorithmik	33
3.3.1	Graph Management	34
3.3.2	Graph Mining	36
4	Entwicklung einer Klassifizierung von Graph Mining Algorithmen	44
4.1	Vorbereitungen für Klassifizierungsverfahren	44
4.1.1	Auswahl des Bewertungsverfahrens.....	44
4.1.2	Differenzierung von Graph Management und Graph Mining	45
4.1.3	Auswahl der zu klassifizierenden Algorithmen.....	49
4.1.4	Auswahl der logistischen Fragestellungen	50
4.1.5	Definition von Bewertungskriterien	51
4.1.6	Definition von Klassen	52
4.2	Klassifizierung von Graph Mining Algorithmen.....	53
4.2.1	Strategische Netzwerkgestaltung.....	54
4.2.2	Bedarfsplanung	57
4.2.3	Netzwerkplanung.....	59
4.2.4	Beschaffungsplanung	61
4.2.5	Distributionsplanung	63
4.2.6	Verfügbarkeits- und Machbarkeitsprüfung	65
4.2.7	Beschaffungsfineplanung	67
4.2.8	Distributionsfineplanung	68

4.2.9	Zusammenfassung der Klassifizierungen	72
5	Evaluierung der Methode	78
5.1	Fallbeispiel	78
5.2	Fazit	83
6	Zusammenfassung und Ausblick	85
	Abkürzungsverzeichnis	88
	Abbildungsverzeichnis	89
	Tabellenverzeichnis	90
	Literaturverzeichnis	91
	Eidesstattliche Versicherung	99

1 Einleitung

In der Logistik resultieren viele Megatrends in einer zunehmenden Komplexität der Lieferkette. Daraus entstehen zunehmend komplexe Anforderungen, auf die die Unternehmen eingehen müssen, um wettbewerbsfähig und flexibel zu bleiben. (vgl. Wehberg 2017, S. 377) Das Supply Chain Management beschäftigt sich mit der Koordination der logistischen Vorgänge und Funktionen innerhalb dieser Lieferkette und zielt außerdem darauf ab, die Bestände der Supply Chain zu reduzieren und den Service für den Endkunden zu steigern (vgl. ten Hompel und Heidenblut 2011, S. 301).

Ein reales Logistiknetzwerk kann mit seinen Informationen und Daten als ein Graph modelliert werden, dessen Knoten und Kanten je nach Anwendungsfall unterschiedliche Bedeutungen besitzen. (vgl. Mattfeld und Vahrenkamp 2014, S. 5) Durch eine Analyse dieser großen Datenmengen kann ein Unternehmen neben der Prozessoptimierung ebenfalls die Wettbewerbsfähigkeit steigern. (vgl. Runkler 2010, S. 1). Eine erfolgskritische Rolle spielt dabei die Extraktion von entscheidungs- und steuerungsrelevanten Informationen aus diesen Datenmengen, denn mit ihnen kann nach Wehberg (vgl. 2017, S. 377) das Verhalten von Logistiksystemen vorhergesagt werden.

Es existieren unterschiedliche Berechnungsmethoden zur Lösung von logistischen Problemstellungen in Graphen, die sich nach Aggarwal und Wang (vgl. 2010a, S. 1 ff.) in die Algorithmen des Graph Management und des Graph Mining unterteilen lassen. Es existiert jedoch keine klare Abgrenzung dieser graphbasierten Verfahren gegeneinander. Die Algorithmen des Graph Mining verfolgen mit der Extraktion von interessanten Informationen aus großen Datenmengen denselben Zweck wie das Data Mining, lassen sich jedoch auf Graphen anwenden. (vgl. Atastina, et al. 2017, S. 3). Allerdings wurde noch nicht erforscht, wie sich die Verfahren des Graph Mining in Hinblick auf die Beantwortung logistischer Fragestellungen einsetzen lassen.

Das primäre Ziel dieser Arbeit ist deshalb die Klassifizierung von Algorithmen des Graph Mining zur Beantwortung logistischer Fragestellungen in Supply Chains, um eine Entscheidungsgrundlage zu schaffen, welcher Algorithmus für welche Problemstellung effizient genutzt werden kann. Dabei wird der Ansatz verfolgt, die Klassifizierung nach den unterschiedlichen Teilgebieten des Aufgabenmodells des Supply Chain Managements zu gliedern. Aufgrund der fehlenden Abgrenzung der Graphalgorithmen ist jedoch unklar, welche Algorithmen bei der Klassifizierung zu berücksichtigen sind. Deshalb wird das Nebenziel gesetzt, einen detaillierten Vergleich und eine Abgrenzung der Algorithmen des Graph Mining und des Graph Management zu erarbeiten.

Zur Erreichung der genannten Ziele ist die Arbeit in mehrere Forschungsschritte unterteilt. In den Kapiteln 2 und 3 werden die erforderlichen theoretischen Grundlagen in den Bereichen der Logistik und der graphbasierten Verfahren durch eine ausführliche Literaturrecherche erarbeitet. Bevor die für die logistischen Fragestellungen relevanten Teilgebiete des Aufgabenmodells thematisiert werden (vgl. Abschnitt 2.4.2 und Abschnitt 2.4.3), werden in einem ersten Schritt die Begriffsbestimmungen der Logistik, der Supply Chain und des Supply Chain Managements aus einem umfassenden Vergleich verschiedener Definitionen herausgearbeitet (vgl. Abschnitt 2.1 und Abschnitt 2.3). Darüber hinaus werden die Aufgaben und Ziele des Daten- und Informationsmanagements beschrieben (vgl. Abschnitt 2.2). Die theoretischen Grundlagen aus Kapitel 2 bilden unter anderem die Basis für die Auswahl der logistischen Fragestellungen in Abschnitt 4.1.4 und der Definition von Bewertungskriterien in 4.1.5, die beide für das Klassifizierungsverfahren benötigt werden.

Einleitung

Zu Anfang von Kapitel 3 werden zum allgemeinen Verständnis die Grundlagen der Graphentheorie beschrieben (vgl. Abschnitt 3.1). Da eine reale Problemstellung zur Bearbeitung durch einen Algorithmus zunächst abstrahiert werden muss, werden in Abschnitt 3.2 die graphbasierten Modelle im Supply Chain Management beschrieben (vgl. Tarjan 2005, S. 18). Durch die Kenntnis über diese Modelle können in Abschnitt 4.2 geeignete Anwendungsbeispiele gebildet werden. Abschließend werden in den Abschnitten 3.3.1 und 3.3.2 die verschiedenen Algorithmen des Graph Management und des Graph Mining beschrieben. Diese beiden Bereiche werden in Abschnitt 4.1.2 gegeneinander abgegrenzt und es werden die Unterschiede der beiden Verfahren aufgezeigt. Die Erkenntnisse dieser Differenzierung fließen in die Auswahl der zu klassifizierenden Algorithmen (vgl. Abschnitt 4.1.3) ein.

Die eigentliche Methode der Klassifizierung von Graph Mining Algorithmen zur Beantwortung logistischer Fragestellungen wird in Abschnitt 4.1.1 vorgestellt. Die Bewertungskriterien für diese Methode werden in Abschnitt 4.1.5 identifiziert und die Klassen, in die die Algorithmen eingeteilt werden, werden in Abschnitt 4.1.6 definiert. In Abschnitt 4.2.1 bis 4.2.8 werden die Algorithmen hinsichtlich der Beantwortung verschiedener Fragestellungen klassifiziert und die Ergebnisse werden in Abschnitt 4.2.9 zusammengetragen.

Die Ergebnisse der Klassifizierung (vgl. Abschnitt 4.2) werden in Kapitel 5 durch ein Fallbeispiel evaluiert. Dabei soll überprüft werden, ob die Klassifizierung der Algorithmen auf eine ähnliche Fragestellung übertragen werden kann (vgl. Abschnitt 5.1). Abschließend werden in Abschnitt 5.2 einige generelle Erkenntnisse zu der Klassifizierung getroffen, die sich unter anderem auf deren Nützlichkeit beziehen. Danach werden die in dieser Arbeit erreichten Ergebnisse in Kapitel 6 zusammengefasst und es wird ein Ausblick auf zukünftige Forschungsfragen gegeben, die an das Thema anknüpfen.

2 Logistik

In diesem Kapitel werden die Grundlagen der betriebswirtschaftlichen Logistik beschrieben. Nach einer Diskussion des Logistikbegriffs und der Erläuterung einer Supply Chain (Abschnitt 2.1) wird auf den informationstechnischen Teil von Supply Chains eingegangen (Abschnitt 2.2). Daran schließt sich eine Charakterisierung des Supply Chain Managements (Abschnitt 2.3) und dessen logistischen Fragestellungen an (Abschnitt 2.4).

2.1 Grundlagen

Die Logistik erweist sich unter den betriebswirtschaftlichen Disziplinen mit als diffuseste, da sie historisch gesehen zunächst lediglich ein praktisches Anliegen war und erst weitaus später als wissenschaftliche Disziplin wahrgenommen wurde (vgl. Hausladen 2020, S. 2). Im Bereich der Wirtschaft wurde der Begriff der Logistik in den USA seit ungefähr 1950 gebraucht und in Deutschland seit circa 1970 (vgl. Fleischmann, Begriffliche Grundlagen der Logistik 2018, S. 1). Für den Begriff der Logistik existieren zahlreiche Bedeutungsansätze, weshalb es notwendig erscheint, im Folgenden einige dieser Definitionen vorzustellen (vgl. Pfohl, Logistiksysteme: Betriebswirtschaftliche Grundlagen 2018, S. 11).

Die Herkunft des Begriffs ist nicht eindeutig geklärt, doch ein Ursprung im militärischen Bereich scheint plausibel (vgl. Huber und Laverentz 2019, S. 1). Dort beschrieb er bereits im 19. Jahrhundert beispielsweise die Bewegungen und Versorgungen von Truppen oder die Planung des Nachschubs. Vor diesem Hintergrund leitet sich das Wort „Logistik“ vom französischen Wort für die Truppenunterkunft „Logis“ ab. (vgl. Fleischmann, Begriffliche Grundlagen der Logistik 2018, S. 1)

Göpfert (vgl. 2016, S. 41) stellt einen Lösungsversuch für das logistische Identitätsproblem an, dessen Ergebnis eine Zusammenfassung der entscheidenden inhaltlichen Aussagen über die betriebswirtschaftliche Logistik ist, die einem relativen Konsens zugeführt wird. Bestehende Erklärungsansätze für den Begriff der Logistik werden je nach ihrer Betrachtungsebene einer von zweien Gattungen zugeordnet. Die empirisch-induktiven Ansätze beobachten konkrete Probleme aus der Unternehmenspraxis und definieren den Objektbereich der Logistik durch die Zusammenfassung der Probleme zu einer Problemfamilie und das Ziehen von Folgerungen auf allgemeine Sachverhalte. Dabei wird allerdings weder auf die Differenzierung der Logistik von anderen Disziplinen noch auf eine Eingliederung der Logistik in das System der Betriebswirtschaftslehre ausführlich eingegangen. Dies wird von der zweiten Gattung, den logisch-deduktiven Ansätzen, berücksichtigt. Diese schließen vom Allgemeinen auf das Spezielle, indem von einem bestehenden, akzeptierten System der Betriebswirtschaftslehre neue Felder herausgearbeitet werden.

Beim empirisch-induktiven Ansatz können die zahlreichen Auffassungen ebenfalls in zwei Gruppen unterteilt werden (vgl. Göpfert 2016, S. 44). Obwohl beide Definitionen eine flussorientierte Sichtweise besitzen, unterscheiden sich die Logistikverständnisse grundlegend. Die erste Gruppe setzt die Logistik mit den drei Kernaktivitäten (Transportieren, Umschlagen, Lagern) gleich, wohingegen die zweite Gruppe eine weitere Auffassung vertritt und die Logistik nicht auf spezifische Aktivitäten reduziert. Vielmehr bezieht sich die Logistik auf wirtschaftliche Phänomene und Zusammenhänge, wodurch die Logistik auch als Managementphilosophie verstanden werden kann. (vgl. Göpfert 2016, S. 48) Ein Beispiel für das Logistikverständnis nach der ersten Gruppe ist die Definition nach Pfohl (2018, S. 12): „Zur Logistik gehören alle Tätigkeiten, durch die die raumzeitliche Gütertransformation und die damit zusammenhängenden Transformationen hinsichtlich der Gütermengen und -sorten, der Güterhandhabungseigenschaften sowie der logistischen Determiniertheit der Güter geplant, gesteuert, realisiert oder

Logistik

kontrolliert werden. Durch das Zusammenwirken dieser Tätigkeiten soll ein Güterfluss in Gang gesetzt werden, der einen Lieferpunkt mit einem Empfangspunkt möglichst effizient verbindet.“

Die logisch-deduktiven Ansätze beschreiben die Logistik als eine spezielle Führungskonzeption, die auf Objektflüsse ausgelegt ist (vgl. Göpfert 2016, S. 50). Durch die Definition der Logistik als eine Koordinationsfunktion im Führungssystem gehört Weber (vgl. 1994, S. 18) zu den Vertretern dieses Ansatzes. Zur Klassifizierung von existierenden und neuartigen betriebswirtschaftlichen Fachgebieten entwickelte er ein Ordnungsraster, welches Unternehmen in ein Ausführungs-, Führungs- und Metaführungssystem trennt. Im Ausführungssystem werden alle Aufgaben zusammengefasst, die unmittelbar oder mittelbar mit der Generierung von Marktleistungen zusammenhängen. Die Koordination und Strukturierung dieser Aufgaben gehören zum Inhalt des Führungssystem, beispielsweise der Produktionsplanung und -steuerung. (vgl. Weber 1994, S. 3, 18) Bei der Betrachtung aller Aufgaben eines Unternehmens ist es zwecks Übersichtlichkeit sinnvoll, mehrere Führungsteilsysteme zu bilden. Alle struktur- und prozessgestaltenden Führungsfunktionen, die sich auf ebenjene Subsysteme beziehen, werden im Metaführungssystem gebündelt, zum Beispiel die Abstimmung der Produktions- mit der Absatzplanung. (vgl. Weber 1994, S. 5, 18) Aufgrund der sukzessiven Erweiterung des Tätigkeitsfeldes der Logistik um bereichsübergreifende Steuerungsaufgaben des Material- und Warenflusses, ordnet Weber (vgl. 1994, S. 18) die Logistik den Ebenen der Führung und Metaführung zu.

Göpfert (vgl. 2016, S. 51 ff.) formuliert nach einem Vergleich der unterschiedlichen Ansätze und einer Diskussion der Konsens- und Dissensfelder eine eigene Definition des Logistikbegriffs:

„Die Logistik ist eine moderne Führungskonzeption zur Entwicklung, Gestaltung, Lenkung und Realisation effektiver und effizienter Flüsse von Objekten (Güter-, Informations-, Geld- und Finanzflüsse) in unternehmensweiten und unternehmensübergreifenden Wertschöpfungssystemen.“ (Göpfert 2016, S. 59)

Diese Definition als Führungslehre stellt die dritte Entwicklungsphase im Lebenslauf der Logistik dar. In der ersten Phase wurde die Logistik noch als Funktionenlehre bezeichnet, da sie auf die Aktivitäten der räumlichen und zeitlichen Gütertransformation und deren Verknüpfungen spezialisiert war. Problematisch dabei war zum einen die organisatorische Zersplitterung der material- und warenflussbezogenen Bereiche und zum anderen der niedrige Stand der materialfluss- und informationstechnischen Entwicklung. Aus diesen Gründen wurde die Logistik in der zweiten Phase um Führungsaufgaben erweitert, um den Material- und Warenfluss im Leistungssystem ganzheitlich zu koordinieren. Durch die zweite Stufe entwickelt sich die Logistik von einer Funktionen- zu einer Führungslehre. (vgl. Göpfert 2016, S. 59 f.)

Nach Göpfert (vgl. 2016, S. 62) kommt das Supply Chain Management in die dritte Entwicklungsphase. Bevor in Abschnitt 2.3 auf das Verhältnis zwischen dem Supply Chain Management und der Logistik eingegangen wird, wird die Supply Chain als primärer Betrachtungsgegenstand des Supply Chain Management vorgestellt (vgl. Muchna, et al. 2018, S. 16).

Der Begriff der Supply Chain hatte seinen Ursprung in der Arbeit von Houlihan in den 1980er Jahren und beschreibt ein logistisches Netzwerk (vgl. Vahrenkamp und Kotzab 2012, S. 25). Synonym für diese Bezeichnung werden in der deutschen Literatur auch die Begriffe der Versorgungskette oder der Wertschöpfungskette genutzt (vgl. Klaus, Krieger und Krupp 2012, S. 549). Unter ihnen wird eine Verbindungskette von vor- und nachgelagerten Unternehmen, wie den Herstellern und ihren Zulieferern, verstanden. Weitere Partner dieser Kette sind Groß- und Einzelhändler als Absatzmittler, Logistikunternehmen als Absatzhelfer, aber auch der Verbraucher als Endabnehmer. (vgl. Pfohl 2018, S. 337) Die

Logistik

Supply Chain beinhaltet jegliche „Fertigungs- und Absatzstufen von der Rohstoffgewinnung über die Produktion bis hin zum Absatz an den Konsumenten“ (Klaus, Krieger und Krupp 2012, S. 549) und im umfassenden Sinne zusätzlich die Netzwerke zur Entsorgung und zum Recycling (vgl. Schönsleben 2016, S. 11).

Anders als der Name vermuten mag, ist die Supply Chain nicht zwingend eine einfache Kette, sondern in der Realität vielmehr ein Netzwerk (vgl. Muchna, et al. 2018, S. 16). Muchna, et al. (vgl. 2018, S. 17) unterscheiden zwischen einer Basic Supply Chain, die aus der Sicht eines Unternehmens lediglich das eigene Unternehmen, die unmittelbaren Lieferanten und den direkten Kunden beinhaltet, und einer Extended Supply Chain, welche die Erweiterung der Basic Supply Chain in Richtung der Lieferanten und der Kunden darstellt. In **Abbildung 2.1** ist eine Supply Chain mit ihren Partnern und der Unterscheidung nach Basic Supply Chain und Extended Supply Chain nach Muchna, et al. (vgl. 2018, S. 17) beispielhaft dargestellt. Durch das Szenario, dass ein Konsument Teil einer anderen Supply Chain sein kann, welche wiederum ein Investitionsgut fertigt, ergibt sich die Dimension der Mehrstufigkeit. Ein Beispiel dafür ist ein Unternehmen, das mit einer Werkzeugmaschine ein Produkt herstellt, welches zur Fertigung von anderen Maschinen genutzt wird. (vgl. Schönsleben 2016, S. 11) In dieser Arbeit wird die Mehrstufigkeit nicht berücksichtigt.

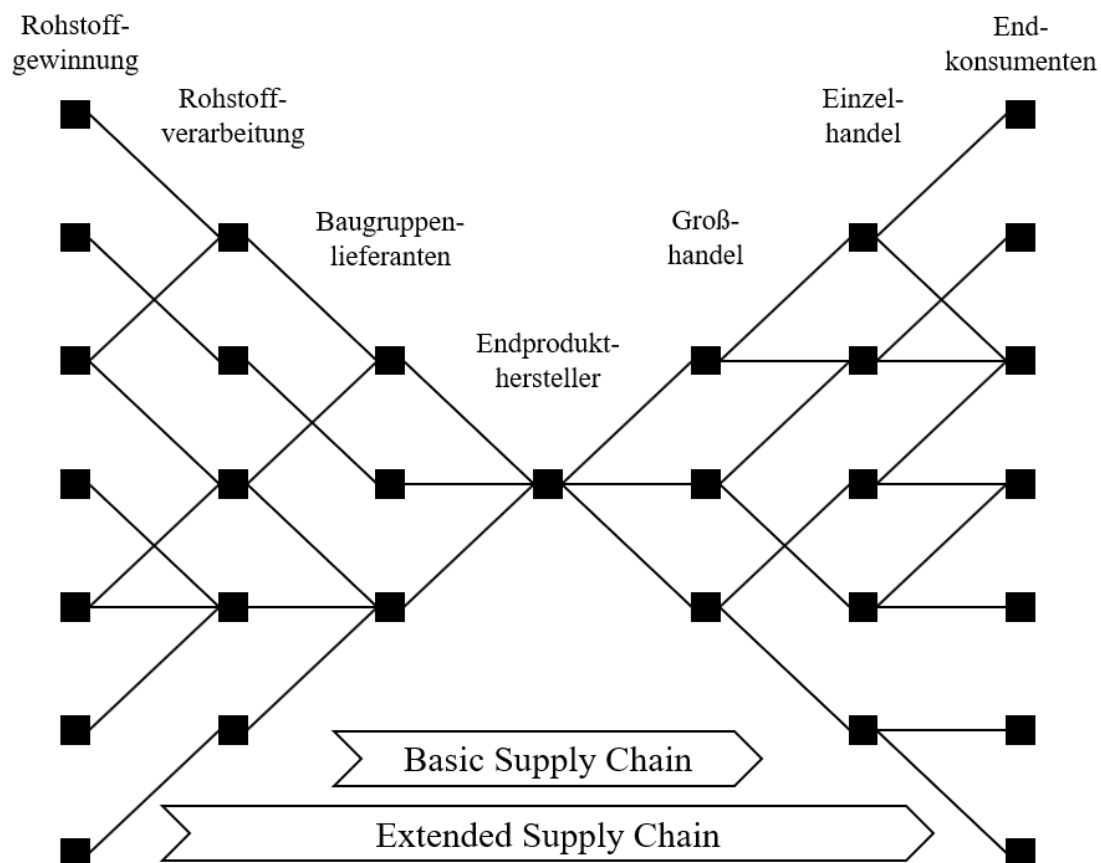


Abbildung 2.1: Beispielhafte Darstellung einer Supply Chain

Ein weiterer Begriff, der oft in Zusammenhang mit einer Supply Chain genannt wird, ist der der Logistikkette. Fleischmann (vgl. 2008, S. 4) beschreibt die Logistikkette als logistisches System eines Industrieunternehmens, die den kompletten Güterfluss von den Lieferanten zum Unternehmen, im Unternehmen selbst und schlussendlich zum Kunden einschließt. Für jeden zwischenbetrieblichen Logistikprozess entsteht somit eine eigene Logistikkette, was nicht dem Regelfall der Logistik entspricht, da sie

Logistik

sich mit mehreren Logistikketten, bzw. -netzwerken beschäftigt. (vgl. Muchna, et al. 2018, S. 17) Werner (vgl. 2017, S. 18) vergleicht zudem die enthaltenen Tätigkeiten beider Begriffe und konstatiert, dass die Supply Chain das umfassendere Konzept darstellt. So bezieht sich die Logistikkette primär auf physische Tätigkeiten, die der Raum- und Zeitüberbrückung dienen, wohingegen die Supply Chain komplette Wertschöpfungsnetzwerke umfasst, die neben den physischen Tätigkeiten auch beispielsweise die Geldflüsse berücksichtigen. (vgl. Werner 2017, S. 18) Aus diesen Gründen wird für diese Arbeit der Begriff der Supply Chains nach Chopra und Meindl (vgl. 2016, S. 13) definiert:

“A supply chain consists of all parties involved, directly or indirectly, in fulfilling a customer request. The supply chain includes not only the manufacturer and suppliers, but also transporters, warehouses, retailers, and even customers themselves. Within each organization, such as a manufacturer, the supply chain includes all functions involved in receiving and filling a customer request. These functions include, but are not limited to, new product development, marketing, operations, distribution, finance, and customer service.” (vgl. Chopra und Meindl 2016, S. 13)

In einem ersten Schritt des inhaltlichen Ablaufs einer Supply Chain meldet ein Konsument seinen Bedarf gegenüber einem an dem Wertschöpfungskettenende befindlichen Unternehmen. Diese Bedarfsinformation löst die Versorgungskette aus und wird von dem am Ende der Wertschöpfungskette agierenden Unternehmen an alle anderen Unternehmen weitergeleitet, die an der Leistungserstellung beteiligt sind. Im Prozess der Leistungserstellung ist die enge Zusammenarbeit der Unternehmen erfolgsentscheidend und erst durch einen funktionierenden Fluss von Informationen entstehen Leistungsbeziehungen zwischen den Unternehmen, welche sich als Material- und Warenflüsse mit den dazugehörigen Geld- und Finanzflüssen dokumentieren. Diese Flüsse gelten zusammen mit dem Informationsfluss als verbindende Elemente der unternehmensübergreifenden Wertschöpfungskette. (vgl. Göpfert 2016, S. 66 f.)

Werden die Logistik mehrerer organisatorischer Einheiten, wie etwa von Unternehmen oder Teilen eines Unternehmens, zu einer umfassenden Logistik zusammengefasst, ist von einem Logistiknetzwerk die Rede. Demnach können auch Produktions- und Beschaffungsnetzwerke als Synonyme für ein Logistiknetzwerk bezeichnet werden. (vgl. Schönsleben 2016, S. 10) Eine weitere Definition für den in der Literatur recht unterschiedlich verwendeten Begriff des logistischen Netzwerkes befasst sich mit den Relationen zwischen Logistikdienstleistungsunternehmen in sogenannten Logistikservice-Netzwerken. Dabei kümmern sich die Logistikdienstleister um die Entwicklung und Umsetzung von komplexen logistischen Leistungssystemen im Wertschöpfungsnetzwerk der Kunden. (vgl. Stich, Quick und Cuber 2013, S. 36 f.) In dieser Arbeit wird ein Logistiknetzwerk jedoch als ein solches bezeichnet, wenn es „durch die unternehmensübergreifende Entwicklung, Beschaffung oder Produktion von Gütern und den damit zusammenhängenden logistischen Prozessen“ (Stich, Quick und Cuber 2013, S. 37) entstanden ist und auf die Organisation eines effizienten und durchgängigen Wertschöpfungsprozesses abzielt. Sowohl die Transport-, Umschlag- und Lagerprozesse zur Überbrückung von Raum und Zeit als auch die damit verbundenen Informationsprozesse können grafisch als Geflecht von Quelle-Senke-Beziehungen dargestellt werden. (vgl. Scheel 2012, S. 399) Die Modellierung dieser Netzwerke wird in Abschnitt 3.2 thematisiert.

Nach den Begriffserläuterungen der Logistik und der Supply Chain werden in diesem Abschnitt abschließend die Ziele der Logistik benannt. Die Gemeinsamkeit aller zuvor genannten Logistikdefinitionen ist das Ziel der Optimierung des Logistikerfolgs, zu dessen Bestandteilen die Logistikleistung und die Logistikkosten gehören (vgl. Schulte 2017, S. 11). Im Wesentlichen zählen zu der Logistikleistung die Lieferzeit, die Lieferqualität, die Lieferzuverlässigkeit sowie die Lieferflexibilität (vgl. Pfohl 2018,

Logistik

S. 38). Die Logistikkosten setzen sich unter anderem aus den Auftragsabwicklungskosten, den Transportkosten, den Lagerbestandskosten, aber auch den Lagerhauskosten zusammen (vgl. Pfohl 2018, S. 34). Grundsätzlich ist es nicht möglich, die Maximierung der Logistikleistung und die Minimierung der Logistikkosten gleichermaßen zu erfüllen, es existiert also ein Zielkonflikt (vgl. Pfohl, Logistiksysteme: Betriebswirtschaftliche Grundlagen 2018, S. 42). Ein konkretes Beispiel stellt der Konflikt zwischen der Lieferzeit und den Lieferkosten dar: Für viele Kunden mag eine garantiert pünktliche Lieferung ein kaufentscheidender Faktor sein. Allerdings dürften sich die Transportkosten, die aus einer entsprechenden Strategie resultieren, aus der betriebswirtschaftlichen Sicht auf Dauer nicht lohnen. (vgl. Hausladen 2020, S. 8)

2.2 Daten- und Informationsmanagement in der Logistik

Die Generierung von Wissen aus Informationen, welches in nachhaltige Wettbewerbsvorteile überführt werden kann, ist das Ziel von wissensorientierter Unternehmensführung (vgl. North 2011, S. 35). Nach North (vgl. 2011, S. 39) müssen zur Erreichung dieses Ziels alle Stufen der Wissenstreppe, die in *Abbildung 2.2* in Anlehnung an North (vgl. 2011, S. 36) dargestellt ist, gestaltet werden, denn die Vernachlässigung einer Stufe hätte negative Folgen auf das operative Geschäft oder die Verwirklichung von Geschäftsstrategien.

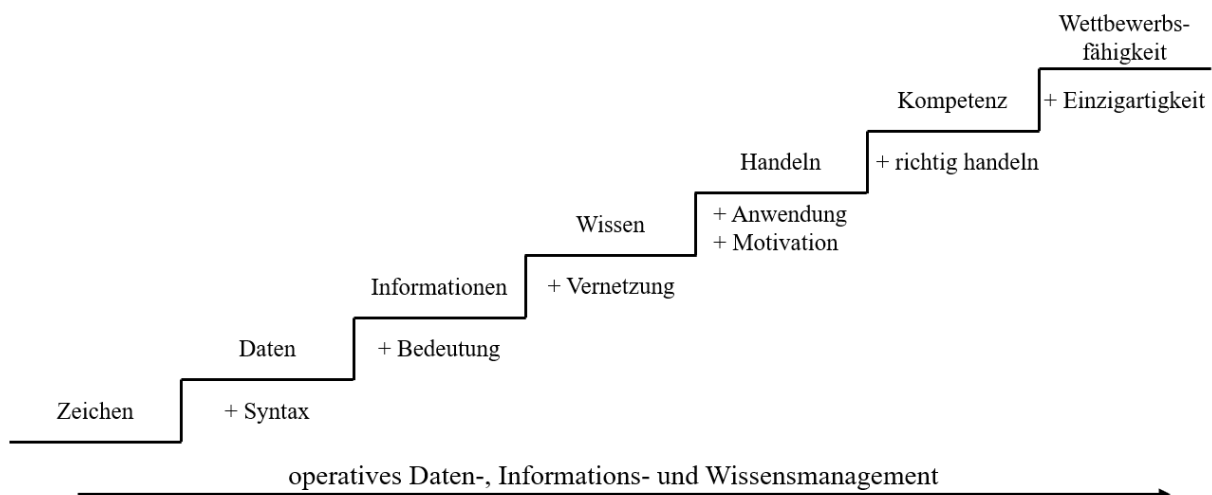


Abbildung 2.2: Die Wissenstreppe

Im Folgenden wird das Modell der Wissenstreppe zusammengefasst. Zuerst werden Zeichen, die als Buchstaben, Ziffern oder Sonderzeichen vorliegen, durch einen Code oder eine Syntax zu Daten. Bevor diese Daten (z.B. 7 %) interpretiert werden können, muss ihnen ein Bedeutungskontext (z.B. Absatzsteigerung von Produkt A im ersten Quartal des aktuellen Jahres) zugewiesen werden. Durch diese Zuweisung entstehen Informationen, die jedoch ohne eine Vernetzung mit anderen Informationen für einen Betrachter wertlos sind. Erst durch eine zweckdienliche Vernetzung kann Wissen generiert werden (z.B. Vergleich mit Absatzentwicklungen des ersten Quartals der Vorjahre). Dabei ist es stets kontextspezifisch und an Personen gebunden, sodass es nicht in einer Datenbank abgespeichert werden kann. Die nächste Stufe des Handelns wird erreicht, wenn eine Person motiviert ist und den Freiraum besitzt, das erworbene Wissen zu einer Fertigkeit umzusetzen. Die erfolgreiche Anwendung von Wissen zur Lösung von Problemen wird auch Kompetenz genannt. Erst durch den Besitz von einzigartigen Kompetenzen, die bei Mitbewerbern nicht vorhanden sind, wird eine Organisation wettbewerbsfähig. (vgl. North 2011, S. 36 ff.)

Logistik

Da Informationen für einen effizienten Wissensaufbau und -transfer bereitgestellt, gespeichert und verteilt werden müssen, ist das Daten- und Informationsmanagement ein wichtiger Grundstein für ein erfolgreiches Wissensmanagement und darüber hinaus für die gewinnbringende Implementierung von Informations- und Kommunikationstechnologie (vgl. North 2011, S. 39 f.).

In der Logistik fällt eine große Menge zweckneutraler Daten an, die mit den daraus generierten zweckgerichteten Informationen die Entscheidungsgrundlage für Planungs-, Steuerungs- und Kontrollprozesse bilden. Das Datenmanagement beschäftigt sich mit Konzepten zur Datenorganisation, -integration und -analyse, damit diese Daten effektiv und effizient genutzt werden können. (vgl. Schulte 2017, S. 104) Die anfallenden Daten lassen sich nach Mertens, et al. (vgl. 2012, S. 38) nach verschiedenen Kriterien klassifizieren. Als erstes Kriterium ist das der Zeichenart, bzw. des Datentyps zu nennen, wonach Daten aus Ziffern, Buchstaben des Alphabets oder einer Mischung der beiden Ausprägungsformen bestehen können. Das Kriterium der Erscheinungsform differenziert zwischen akustischen, bildlichen und schriftlichen Daten. Wird nach der Formatierung unterschieden, können formatierte Daten oder unformatierte Daten vorliegen, etwa als Tabelle oder formfreiem Text. Eine weitere Trennung kann abhängig von der Position im Verarbeitungsprozess in Ein- und Ausgabedaten vorgenommen werden. (vgl. Mertens, et al. 2012, S. 38) Das letzte von Mertens, et al. (vgl. 2012, S. 38) genannte Kriterium ist das des Verwendungszweckes. Die ersten beiden Kategorien sind die der Stamm- und Bewegungsdaten (vgl. Mertens, et al. 2012, S. 38). Zu den Stammdaten zählen längerfristige Informationen wie Material-, Lieferanten- oder Kundenstammdaten, auf die im Sukzessivplanungsprozess ständig zurückgegriffen wird. Beispiele für Bewegungsdaten sind Lagerbestandsdaten, welche unter anderem Informationen zu den Lagerbeständen sowie den Lagerzugängen und -abgängen enthalten. Diese Art von Daten beziehen sich auf gespeicherte Stammdaten und haben im Gegensatz zu den Stammdaten einen kurzfristigen Charakter. Außerdem werden sie für die Verwaltung verschiedener Statuszustände genutzt und besitzen stets einen Zeitbezug. (vgl. Stich, Oedekoven und Brosze 2013, S. 262) Zwei weitere von Mertens, et al. (vgl. 2012, S. 38) genannte Kategorien sind die Transfer- und Vormerkdaten. Transferdaten werden von einem zum anderen Programm transferiert, wobei das erste Programm die Daten generiert und das zweite diese erzeugten Daten weiterverarbeitet. Vormerkdaten existieren nur vorübergehend und werden nach ihrer Nutzung, bzw. nach Eintritt eines definierten Ereignisses, gelöscht. (vgl. Mertens, et al. 2012, S. 38)

Nach Schönsleben (vgl. 2016, S. 789) befasst sich das Informationsmanagement mit jenen Bereichen, die mit der Führung von Informationen im Unternehmen korrelieren. Dazu gehören „das strategische und operationelle Management 1.) der Informationen selbst, 2.) des Lebenszyklus von Informationssystemen und 3.) der Informationstechnologien (das IT-Management)“ (2016, S. 789).

Ein Informationssystem wird aus organisatorisch-technischer Sicht von Schulz-Spathelf (2012, S. 236) als „Sammelbegriff für alle Elemente der geregelten Informationsgewinnung, und -verarbeitung“ definiert. Allen Informationssystemen gemeinsam ist die Erhebung und Pflege von Stamm- und Bewegungsdaten (vgl. Stich, Oedekoven und Brosze 2013, S. 262). Nach Schulte befasst sich ein IT-gestütztes logistisches Informationssystem mit der Unterstützung sämtlicher logistischer Planungs-, Abwicklungs-, Kontroll- und Steuerungsaufgaben und greift dabei auf die Gesamtheit der Hard- und Software, Daten, Netzwerke und Personen zurück (vgl. Schulte 2017, S. 86). Beispiele für Informationssysteme in der Logistik sind Transport-Management-Systeme (TMS), die zum Planen, Steuern und Verwalten von Transportvorgängen und Touren genutzt werden und Enterprise-Resource-Planning-Systeme (ERP), welche bei zahlreichen Geschäftsprozessen eines Unternehmens der Planung und Organisation

Logistik

vorhandener Ressourcen dienen. Warehouse-Management-Systeme (WMS) werden für die Steuerung, Kontrolle und Optimierung von Lager- und Distributionssystemen verwendet. (vgl. Stich, Oedekoven und Brosze 2013, S. 260)

Nach Krupp und Wolf (vgl. 2010, S. 17) gehören flexible IT-gestützte Wertschöpfungsnetzwerke zu den aktuellen Entwicklungen bei Informationssystemen in der Logistik (vgl. Schulte 2017, S. 88). Durch die Fokussierung des Wertschöpfungsnetzwerkes auf logistische Prozesse im Rahmen von Industrie 4.0 und Smart Factory werden neue Begriffe wie Logistik 4.0 oder Smart Factory hervorgebracht. Dabei wird die intelligente IT-gestützte Logistik als Querschnittsfunktion genannt, die unterschiedliche Wertschöpfungsbereiche oder Wirtschaftssektoren beispielsweise durch den Einsatz des Internet of Things verbindet. (vgl. Hausladen 2020, S. 19)

Bei der IT-gestützten Logistik werden laut Hausladen (vgl. 2020, S. 27) Material-, Personen-, Energie- und Informationsflüsse in ihrer Ausgestaltung als E-Business-Systeme geplant, gesteuert und überwacht. Als Electronic Business wird die Anbahnung, Vereinbarung und Abwicklung elektronischer Geschäftsprozesse bezeichnet, um eine Wertschöpfung zu erreichen. (vgl. Meier und Stormer 2012, S. 2) Zu einem erheblichen Nutzungspotenzial der IT-gestützten Logistik zählt die Realisierung eines möglichst reibungslosen und durchgehenden Informationsflusses ohne Medienbrüche und einer geringeren Anzahl von Kommunikationsfehlern, was zu einer größeren Transparenz führt. Dadurch erfolgen die Planungen synchroner und die Prognosen präziser. Außerdem kann auf Ergebnisse und Engpässe umgehend eingegangen werden. (vgl. Hausladen 2020, S. 22 f.)

Inzwischen sind zahlreiche Methoden und Tools der IT-gestützten Logistik in Betrieb, die jedoch auf eine zumeist standardisierte technologische Basis zurückgreifen (vgl. Hausladen 2020, S. 53). Deshalb wird in dem folgenden Abschnitt ein Überblick über diese Technologien gegeben, welche dafür in funktionsbezogene und funktionsübergreifende Technologien unterteilt werden. Aufgrund vielfältiger technischer Besonderheiten kann eine vollständige Überschneidungsfreiheit der Zuordnung nicht gewährleistet werden. (vgl. Hausladen 2020, S. 55)

Die funktionsorientierten Basistechnologien lassen sich wiederum in die Unterbereiche der Lokalisierung mit ihren Satelliten- und Ordnungstechnologien (vgl. Hausladen 2020, S. 54), der mobilen Kommunikation mit beispielsweise dem Mobilfunkstandard 5G (vgl. Hausladen 2020, S. 64), des elektronischen Datenaustauschs mit dem Verfahren des Electronic Data Interchange (EDI) und der Identifikation klassifizieren. Als Beispiel für die Identifikationstechnologien ist die Radio Frequenz Identifikation (RFID) (vgl. Hausladen 2020, S. 55) oder die intelligente Sensorik zu nennen, welche durch ihre Verknüpfung von Maschinen und dem Internet über intelligente Sensoren eine Enabler-Technologie für die Industrie 4.0 darstellt und nach Fürstenberg und Kirsch (vgl. 2017, S. 296) die notwendige Autonomie für eine wandelbare Logistik und Produktion bereitstellen wird. Bei der funktionsübergreifenden Technologie erfolgt in einem zweiten Schritt die Einteilung in die Untersegmente der Architekturkonzepte, der Technologieressourcen sowie der Analytikmethoden (vgl. Hausladen 2020, S. 55). Architekturkonzepte schlagen den Aufbau der IT-Infrastruktur vor, sodass Datenverluste, -redundanzen und -inkonsistenzen möglichst auf ein Minimum reduziert werden (vgl. Hausladen 2020, S. 73). Diese Konzepte werden durch Technologienressourcen als Gestaltungsmerkmale ergänzt. Mithilfe dieser Ressourcen sollen unter anderem Datenredundanzen oder auch Medienbrüche aufgrund abweichender Datenformate verhindert werden. (vgl. Hausladen 2020, S. 79) Die Analytikmethoden des dritten Segments werden zur Steuerung der logistischen Daten- und Informationsflüsse und für eine effiziente Planung genutzt. Als Beispiel eines ganzheitlichen Instruments des Datenmanagements können hier Data Warehouse-

Logistik

Lösungen (DWH) genannt werden, die für die administrative Kontrolle logistischer Informationen zuständig sind. (vgl. Hausladen 2020, S. 86 f.)

Die Kostensenkungspotentiale und Wettbewerbsvorteile gegenüber Mitbewerbern auf den Absatzmärkten sind ohne Informations- und Kommunikationstechnologien kaum zu erreichen, weshalb es erstrebenswert ist, dass alle Unternehmen einer Supply Chain die identischen, bzw. kompatible Technologien und Programme verwenden, um die Integrationsfähigkeit sicherzustellen (vgl. Arnolds, et al. 2016, S. 4).

Einige der vorgestellten Technologien, wie beispielsweise die RFID-Tags, werden bereits seit mehreren Jahren eingesetzt und besitzen eine hohe Variantenvielfalt, sind zuverlässig und wirtschaftlich in der Anschaffung. Neuartige Technologien sind anfangs noch nicht massenhaft verbreitet und durch diese Tatsache preislich nicht so attraktiv wie die bewährten Technologien. Allerdings benötigen neue Technologien gerade diese Verbreitung, um einen neuen Entwicklungsschub zu erhalten. Dieser Widerspruch resultiert in einem Dilemma, nach dem bei neuen Technologien zwischen der Einführung und der massenhaften Verbreitung ein großer Zeitraum liegt und es sogar vorkommen kann, dass alle Aktivitäten und Investitionen rund um den neuen Trend einbrechen. (vgl. Bousonville 2017, S. 15) Nichtsdestotrotz haben Informations- und Kommunikationstechnologien in den letzten Jahrzehnten eine rapide Weiterentwicklung erfahren, was zu vielen technischen Innovationen geführt hat, deren Entwicklung, wie im Falle des robotergestützten Kommissionierens, entweder unmittelbar für die Logistik angedacht waren oder die in der Logistik ein potenzialträchtiges Anwendungsfeld besitzen, z.B. Blockchain und RFID-Lösungen. (vgl. Hausladen 2020, S. 14)

Die umfassende Informatisierung der Logistikbranche, also die digitale Verfügbarkeit von Informationen über die Akteure und Objekte der Branche, wird auch mit dem Begriff der „Logistik 4.0“ bezeichnet. Objekte gelten danach als vernetzt mit ihrer Umgebung, wenn das digitale Abbild der Objekte, mit ihren Eigenschaften und Zustandsgrößen, an die Umgebung kommuniziert werden oder von dieser ausgelesen werden kann. (vgl. Bousonville 2017, S. 5) Die Logistik 4.0 ist dabei nicht mit der Industrie 4.0 gleichzusetzen, da sie sich von Anwendungen abgrenzt, die irrelevant für das sie umgebende logistische System sind und einen ausschließlichen Bezug zur Fertigungstechnologie haben, wie beispielsweise die Fernwartung von Maschinen (vgl. Bousonville 2017, S. 12). Wird der Optimierungsgedanke der Vernetzung intelligenter Lösungen dahingehend weitergeführt, dass entsprechende Technologie über die gesamte Supply Chain eingesetzt wird, impliziert dies nach Hausladen (vgl. 2020, S. 21) auf mittel- bis langfristige Sicht das Konzept eines SCM 4.0.

2.3 Supply Chain Management

Ein umfassendes Redesign von Unternehmen und grundsätzliches Überdenken von dessen Prozessen führten unter anderem neben einer verringerten Fertigungstiefe durch Outsourcing Anfang der 90er Jahre zu einer größeren Anzahl von Stufen in der Supply Chain. Des Weiteren stieg der Vernetzungsgrad durch Global Sourcing und die Schnittstellen zwischen den in einer Supply Chain befindlichen Unternehmen nahm zu. Der daraus resultierende Grad der Komplexität ist schwierig zu beherrschen und kann zum Bullwhip-Effekt führen, sollten die Unternehmen nicht in die prozessorientierte Planung und Steuerung integriert werden. (vgl. Alicke 2005, S. 3) Nach Göpfert (vgl. 2016, S. 67) besteht in nahezu allen Publikationen Konsens darüber, dass der primäre Beweggrund für den Übergang zum Supply Chain Management die Lösung des Bullwhip-Effekts ist, um die Flüsse von Gütern, Informationen, Geld und Finanzen zwischen Unternehmen besser beherrschen zu können.

Logistik

Bei dem durch den amerikanischen Wissenschaftler Jay Forrester nachgewiesenen Phänomen des Bullwhip-Effekts, oder auch Peitscheneffekt, führen bereits kleine Nachfrageveränderungen zu großvolumigen Schwankungen auf den nachgelagerten Versorgungsstufen. Seine Untersuchungen haben beispielsweise ergeben, dass eine 40-prozentige Zunahme des Produktionsausstoßes am Ende der Supply Chain auf eine Zunahme der Bestellungen im Einzelhandel von lediglich 10 Prozent zurückzuführen sei. Das Resultat dieser Betrachtung ist die Tatsache, dass sich die Amplitude der Schwankung aufschaukelt, je weiter eine Stufe in der Supply Chain vom letzten Glied, dem Einzelhandel, entfernt ist. (vgl. Klaus, Krieger und Krupp 2012, S. 186) Zurückführen lässt sich unter anderem auf einen mangelhaften und zeitverzögerten Informationsfluss oder auch auf ein überhöhtes Sicherheitsdenken der Akteure entlang der Supply Chain (vgl. ten Hompel und Heidenblut 2011, S. 224). Nach Simchi-Levi, Kaminsky und Simchi-Levi (vgl. 2000, S. 91) ist einer der häufigsten Vorschläge, dem Bullwhip-Effekt entgegenzuwirken, die Informationen über die aktuelle Nachfrage zu zentralisieren und somit jeder Stufe der Supply Chain transparent bereitzustellen.

Als Absicherung gegen die Nachfrageschwankungen sicherten sich die Unternehmen Anfang der 90er Jahre mit Überbeständen, bzw. Überkapazitäten ab, was allerdings den Anforderungen einer Supply Chain, auf kurze Lebenszyklen und wechselhafter Nachfrage passend zu reagieren, widerspricht. Die Tatsache, dass die meisten Unternehmen ihre eigenen Prozesse wie die Minimierung der Rüstzeiten oder die Ausreizung der Maschinenleistung weitestgehend optimiert haben, führt dazu, dass auch angrenzende Unternehmen in die Planung integriert werden mussten, um die Durchlaufzeit eines Produkts weiter zu verringern. Die Beherrschbarkeit der dynamischen Prozesse am Markt und der Komplexität durch die gestiegene Vernetzung der Unternehmen gehört zu den vorrangigen Aufgaben des Supply Chain Managements, da der Erfolg eines Produktes unmittelbar von der Flexibilität und Reaktionsfähigkeit der Supply Chain abhängt. (vgl. Alicke 2005, S. 3 f.)

Wie auch schon bei der Logistik besitzt das Supply Chain Management eine Vielfalt an Definitionen (vgl. Hausladen 2020, S. 10). Auch zu der Beziehung zwischen Logistik und Supply Chain Management herrscht kein Konsens. So werden die beiden Begriffe teilweise als Synonyme verwendet oder die Definition des Supply Chain Managements geht über den Begriff der Logistik hinaus. (vgl. Schulte 2017, S. 21) Laut Göpfert (vgl. 2016, S. 62) reduzieren sich die zahlreichen Definitionsvorschläge der Literaturanalysen zu der Interpretation des Supply Chain Managements und dessen Inhalt auf zwei Gruppen von Standpunkten.

Von der ersten Autorengruppe wird das Supply Chain Management als neue Entwicklungsstufe der betriebswirtschaftlichen Logistik gesehen (vgl. Göpfert 2016, S. 62). Vertreter dieser Auffassung ist Weber (vgl. 2012, S. 19), der im Supply Chain Management eine Ergänzung der Logistik um die Komponente der flussorientierten Verknüpfung aller in der Wertschöpfungskette befindlichen Unternehmen sieht. Die zweite Definitionsgruppe sieht die Logistik nicht als Führungslehre (vgl. Abschnitt 2.1), sondern als Funktionenlehre. Die Logistik wird von dieser Gruppe entweder gar nicht mit dem Supply Chain Management in Beziehung gesetzt oder aber als Teilmenge von dessen gesehen. So ist die Logistik für Christopher (vgl. 2011, S. 3) nichts weiter als eine Planungsorientierung und Rahmenwerk, welches auf die Erstellung eines einheitlichen Plans für den unternehmensinternen Produkt- und Informationsfluss abzielt. Larson, Poist und Halldórsson (vgl. 2007, S. 5) identifizieren einen weiteren Definitionsansatz, nach dem nur einzelne Überschneidungen zwischen der Logistik und dem Supply Chain Management existieren.

Logistik

Grundsätzlich ist sich die Literatur unter anderem einig, dass das Supply Chain Management analog zur Logistik die Fluss- und Prozessorientierung zur Grundlage hat und die Logistikkonzeption für das Verständnis des Supply Chain Managements herangezogen werden kann (vgl. Schulte 2017, S. 21). Göpfert (vgl. 2016, S. 67) untersucht die inhaltlichen Abläufe einer Supply Chain, um die beiden Definitionsgruppen gegeneinander abzuwägen. Wie bereits in Abschnitt 2.1 beschrieben, entstehen Leistungsbeziehungen zwischen Wertschöpfungspartnern erst durch einen funktionierenden Fluss von Informationen. Außerdem gelten Objektflüsse in Form von Gütern, Informationen, Geld oder Finanzen als Verbindungselemente der unternehmensübergreifenden Wertschöpfungskette. Die Komplexität der Leistungsbeziehungen lässt sich im Falle dieser flussorientierten Betrachtung der Wertschöpfungssysteme auf die Objektflüsse reduzieren, welche traditionell in den Objektbereich der Logistik gehören. Somit entspringt das Supply Chain Management einer zentralen logistischen Problemstellung und ist mit der Entwicklung der Logistik verbunden. (vgl. Göpfert 2016, S. 67 ff.) Die Interpretation der zweiten Gruppe entfernt sich „zu weit vom Kerninhalt der mit Supply Chain Management herausgebildeten neuen Managementqualität“ (Göpfert 2016, S. 69).

Als Ergebnis dieser Abwägung konstatiert Göpfert (vgl. 2016, S. 71), dass das Supply Chain Management eine qualitativ neue Phase der Logistikentwicklung annimmt. Auch Weber (vgl. 2012, S. 20 f.) sieht die Einführung des Supply Chain Management als eigene Stufe der Logistikentwicklung und begründet dies in der Tatsache, dass die Verwirklichung der in Abschnitt 2.1 beschriebenen Supply Chains mit den bisherigen Methoden zu vielschichtigen Führungsproblemen führte. In der vorliegenden Arbeit gilt die folgende Definition:

„Das Supply Chain Management bildet eine moderne Konzeption für Unternehmensnetzwerke zur Erschließung unternehmensübergreifender Erfolgspotenziale mittels der Entwicklung, Gestaltung, Lenkung und Realisation effektiver und effizienter Güter-, Informations-, Geld- und Finanzflüsse.“ (Göpfert 2016, S. 71)

Nach Schulte (vgl. 2017, S. 25) zielt das Supply Chain Management auf „die Erreichung eines Gesamtoptimums in der unternehmensübergreifenden Wertschöpfungskette anstatt lokaler, unternehmensindividueller Optimierung“ ab. Das Gesamtoptimum bezieht sich dabei auf Vorteile in den Bereichen Kosten, Zeit und Qualität. (vgl. Pfohl, Logistiksysteme: Betriebswirtschaftliche Grundlagen 2018, S. 338) Konkrete Vorhaben können die schnellere Anpassung an Änderungen des Marktes oder die Reduktion der Bestandskosten durch die Reduzierung der Bestände in der Supply Chain sein. Weitere Beispiele sind die Verringerung der Auftragsdurchlaufzeiten oder auch die Erreichung einer effizienteren unternehmensübergreifenden Produktionssteuerung und Kapazitätsplanung. (vgl. Vahrenkamp und Kotzab 2012, S. 26) Des Weiteren sind Angebot und Nachfrage über alle Stufen der Supply Chain aufeinander abzustimmen, um die Wirkung des Bullwhip-Effekts zu minimieren (vgl. Göpfert 2016, S. 67).

Ein Konzept, das sich dem Supply Chain Management ähnelt, ist das Logistikmanagement. Pfohl (vgl. 2016, S. 22 f.) definiert den Begriff als „die Gesamtheit der Managementaktivitäten, die sowohl zur Realisation einzelner Logistikprozesse als auch zur Realisation eines umfassenden, ggf. mehrere Unternehmen überspannenden Gesamtprozesses der Logistik erforderlich ist“. Nach Muchna, et al. (vgl. 2018, S. 38) korrespondiert diese Definition gut mit der Entwicklungsstufe der Logistik als Führungslehre, ist aber dennoch gegenüber dem Supply Chain Management abzugrenzen. Allerdings existiert keine klare Abgrenzung der Definitionen zum Logistikmanagement und zum Supply Chain Management, da sie überwiegend davon abhängt, wie umfassend die Definition der Logistik ist. Greift diese beispielsweise die unternehmensübergreifende Flussorientierung auf, wird das Supply Chain Management als die

Logistik

umfangreichere Disziplin zum Managen von Liefer- und Wertschöpfungsketten angesehen, da sich die unternehmensübergreifende Kommunikation und die Kooperationsfelder beim Supply Chain Management im Gegensatz zum Logistikmanagement nicht nur auf die logistischen Bereiche beschränkt. Es stellt sich heraus, dass das Logistikmanagement durch die Konzepte des Supply Chain Managements durch kooperative Ansätze ergänzt wird und sich in die Richtung einer unternehmensübergreifenden Flussorientierung weiterentwickelt. (vgl. Muchna, et al. 2018, S. 19 f.)

2.4 Logistische Fragestellungen des Supply Chain Managements

Das Supply Chain Management zielt nicht auf eine lokale, unternehmensindividuelle Optimierung ab, sondern auf die Erfüllung eines Gesamtoptimums der unternehmensübergreifenden Wertschöpfungskette (vgl. Schulte 2017, S. 25). Eine Vielzahl von logistischen Aufgaben ist mit der Umsetzung der in den Abschnitten 2.1 und 2.3 angesprochenen Ziele der Logistik, bzw. des Supply Chain Managements, beschäftigt. Dabei fallen diese Aufgaben in verschiedene Aufgabenbereiche, die im folgenden Abschnitt thematisiert werden. Darüber hinaus wird auf zwei Supply Chain Management-Modelle.

2.4.1 Aufgaben und Modelle

Die Logistik beschäftigt sich mit der material- und informationsseitigen Koordination einzelner Wertschöpfungsstufen, wie der Gewinnung von Rohstoffen, deren Verarbeitung, dem Transport von Fertigprodukten an weiterverarbeitende Unternehmen oder den Endverbrauchern. Beispiele für die dabei anfallenden logistischen Aufgaben stellen das Transportieren, das Kommissionieren, die Bedarfsermittlung von Material und die Planung sowie Freigabe von Verträgen zur Beschaffung, Produktion und zum Versand dar. Zur besseren Planung werden die verschiedenen und vielfältigen logistischen Einzelaufgaben in Aufgabenschwerpunkte eingeteilt. (vgl. Hausladen 2020, S. 6)

Zsifkovits (vgl. 2018, S. 46) sieht die wesentlichen Aufgaben des Managements in der Zielfindung, der Planung und der Steuerung. Diese Aufgaben erfolgen auf verschiedenen, untereinander abhängigen zeitlichen Ebenen und betreffen unterschiedliche Stufen der Unternehmenshierarchie. Es bietet sich eine Aufteilung ins strategische, taktische und operative Management an. Das strategische Management beinhaltet dabei die Definition und Schaffung von langfristigen Rahmenbedingungen, um eine effektive Unternehmensentwicklung zu gewähren. (vgl. Zsifkovits 2018, S. 46) Die Planung und der Aufbau von logistischen Systemen gehören beispielsweise zu den Aufgaben dieser Ebene (vgl. Huber und Laverentz 2019, S. 11). Die strategischen Ziele werden im taktischen Management sukzessiv in einem mittelfristigen Zeithorizont realisiert, indem die Produktionsinfrastruktur gestaltet und weiterentwickelt wird. Als eine charakteristische taktische Herausforderung ist die Dimensionierung von Produktions- und Lagerkapazitäten zu nennen. Das operative Management beschäftigt sich mit der Festlegung von Elementarzielen und -aufgaben und liegt im kurzfristigen Bereich. Aufträge und Anlieferungen zu disponieren oder kurzfristige Produktionspläne festzulegen und zu steuern gehören zu den typischen operativen Aufgaben. (vgl. Zsifkovits 2018, S. 46 f.) Die Aufgaben dieser Ebene betreffen regelmäßig wiederkehrende Prozesse (vgl. Huber und Laverentz 2019, S. 19).

Um die Eigenschaften und Prozesse einer vorgegebenen Aufgabenstellung systematisch und allgemeingültig zu beschreiben, werden Referenzmodelle genutzt. Diese Modelle besitzen einen allgemeingültigen Charakter für einen definierten Anwendungsbereich. (vgl. ten Hompel und Heidenblut 2011, S. 251) Logistische Referenzmodelle haben die Funktion, die Zusammenhänge zur branchen- und unternehmensneutralen Optimierung von logistischen Ketten zu betrachten (vgl. Kuhn und Hellingrath 2002, S. 29). Das im Jahr 1996 gegründete Supply Chain Council führte das Supply Chain Operations Reference-Model (SCOR) zur Vereinheitlichung von logistischen Prozessen und Umfängen ein (vgl. Kuhn und

Logistik

Hellingrath 2002, S. 105). Mithilfe dieses Modells werden Supply Chains sowohl branchen- als auch unternehmensübergreifend einheitlich beschrieben, bewertet und analysiert (vgl. Kuhn und Hellingrath 2002, S. 29). Die Grundidee des SCOR-Modells ist es, die logistischen Prozesse eines Wertschöpfungsnetzwerkes in ihre sechs Kernprozesse „plan“ (planen), „source“ (beschaffen), „make“ (herstellen), „deliver“ (liefern), „return“ (entsorgen) und „enable“ (unterstützen) zu trennen (vgl. Kuhn, Prozessmodelle, -ketten und -netze 2008, S. 227 f.). Durch die Zusammenfassung der Prozesse, wird das Gesamtmodell des Produktions- und Logistiknetzes gebildet (vgl. Kuhn und Hellingrath 2002, S. 29). Darüber hinaus bietet das SCOR-Modell ein generisches Vorgehensmodell, welches „Konzepte des Business Process Reengineering, Benchmarking und der Prozessbewertung in ein cross-funktionales Rahmenwerk integriert“ (vgl. Kuhn, Prozessmodelle, -ketten und -netze 2008, S. 228).

Wie bereits in Abschnitt 2.2 beschrieben, gehören zur erfolgreichen Umsetzung des Supply Chain Managements die Nutzung von modernen IT-Systemen, mit denen sich einerseits die Prozesse der Wertschöpfungskette planen und steuern lassen und die den Partnern der Wertschöpfungskette zum Informationsaustausch dienen (vgl. Schulte 2017, S. 791). Allerdings ist der Vergleich von Software-Lösungen mehrerer Anbieter und die Zuordnung von ihnen zu einer Aufgabe sehr schwierig, da eine SCM-Software im Allgemeinen verschiedene Aufgabenschwerpunkte unterstützt und unterschiedliche Funktionalitäten bietet. Mit dem SCOR-Modell lassen sich nicht die Prozesse direkt identifizieren, die durch SCM-Software sinnvoll und ausreichend unterstützt werden. Aus diesem Grund wurde das SCM-Aufgabenmodell von den Fraunhofer-Instituten IML und IPA und des Betriebswissenschaftlichen Instituts der ETH Zürich entwickelt, welches die Elemente des SCOR-Modells zerlegt und diese bezüglich der möglichen Software-Funktionalität detailliert. Aus diesem Grund wird das Aufgabenmodell eingesetzt, um SCM-Software zu untersuchen, zu analysieren und auszuwählen. (vgl. Kuhn und Hellingrath 2002, S. 142) Es gliedert sich in die drei Hauptaufgabenbereiche Supply Network Design (Gestaltung), Supply Chain Planning (Planung) und Supply Chain Execution (Ausführung) (vgl. Schulte 2017, S. 790 f.). In *Abbildung 2.3* ist das Aufgabenmodell in Anlehnung an Kuhn und Hellingrath (vgl. 2002, S. 143) dargestellt.

Die strategische Ebene des Supply Network Design beziehungsweise Supply Chain Design beschäftigt sich mit der Gestaltung von Supply Chains (vgl. Hellingrath, Hegmanns, et al. 2008, S. 463). Die langfristigen Planungsaufgaben unterstützen wesentliche Überlegungen, die sich auf eine kostenoptimierte Struktur und die Gestaltung der Logistikkette beziehen, und erstrecken sich dabei über einen Zeitraum von mehreren Jahren (vgl. Kuhn und Hellingrath 2002, S. 143). Softwarelösungen dieser Ebene dienen beispielsweise der Simulation und der kostenseitigen Evaluation von verschiedenen Szenarien (vgl. Hausladen 2020, S. 242).

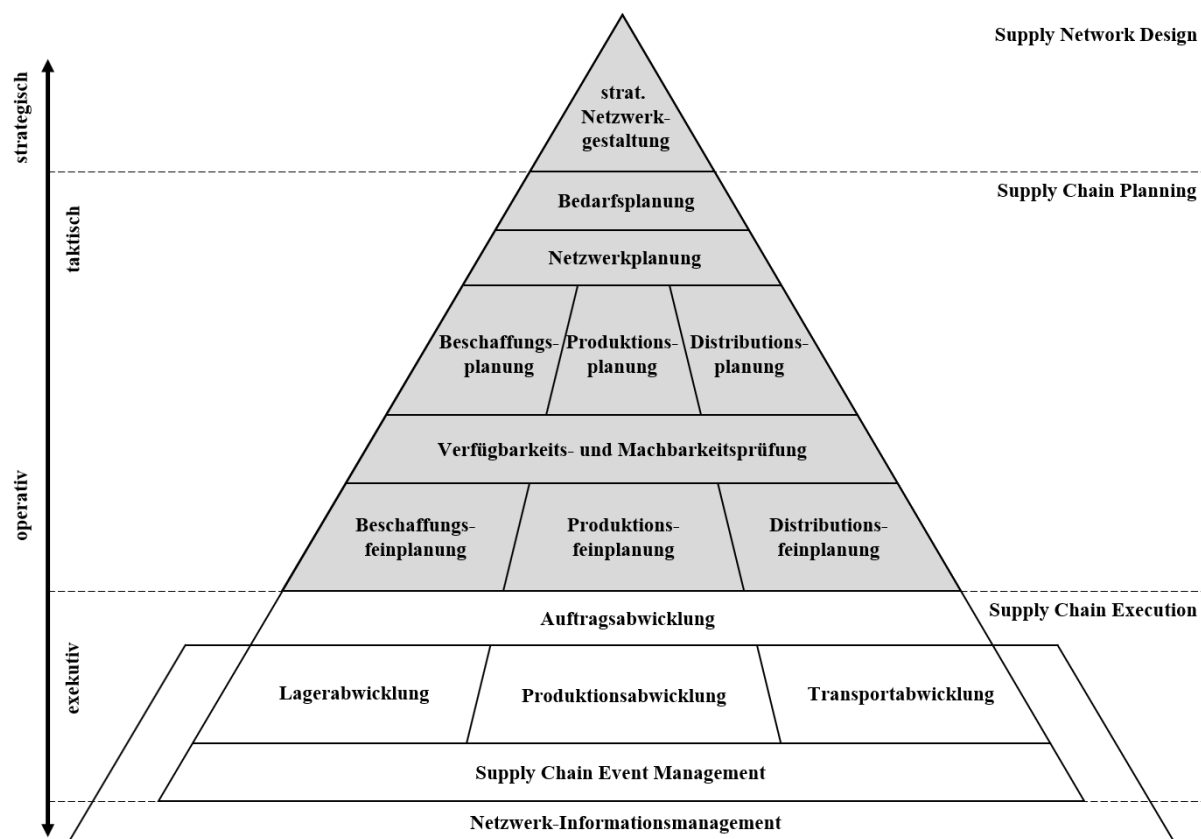


Abbildung 2.3: Aufgabenmodell des Supply Chain Managements

Unter dem Begriff des Supply Chain Planning werden alle taktischen und operativen Planungsaufgaben zusammengefasst, die die für ein erfolgreiches Supply Chain Networking erforderlichen Ressourcen spezifizieren (vgl. Hausladen 2020, S. 242). Außerdem werden hier „für immer detaillierter werdende Bereiche der Wertschöpfungskette (angefangen vom Netzwerk bis herunter auf die Produktionslinie) Aufträge für entsprechend kürzer werdende Zeitbereiche eingeplant“ (Kuhn und Hellingrath 2002, S. 144). Ein Beispiel für einen Bereich ist die Netzwerkplanung, welche die einzelnen Partner der Wertschöpfungskette übergreifend koordiniert. Neben der Netzwerkplanung sei die Distributionsplanung genannt, zu dessen Aufgaben die optimierte Planung der Lagerbestände oder der Transport von Produkten zum Kunden zählen. (vgl. Kuhn und Hellingrath 2002, S. 145 f.) Die übrigen Bereiche lassen sich **Abbildung 2.3** entnehmen. Ziel des Supply Chain Planning ist es, die einzelnen unternehmensspezifischen Planungen entlang der gesamten Supply Chain abzustimmen (vgl. Hausladen 2020, S. 242).

Die Betriebsebene der Supply Chain Execution vereinigt alle Funktionalitäten, die auf die Realisierung einer unternehmensübergreifenden Steuerung der Supply Chain, die Auskunftsfähigkeit und die operative Prozessabwicklung abzielen (vgl. Kuhn und Hellingrath 2002, S. 152). Sie sollen ein möglichst schnelles Einleiten von adäquaten Gegenmaßnahmen auf eintretende Ereignisse begünstigen (vgl. Hausladen 2020, S. 242). Bestandteile der operativen Ausführungsebene sind unter anderem das Controlling oder das Transportmanagement (vgl. Hellingrath, Hegmanns, et al. 2008, S. 463). Entsprechende Software wird aufgrund der vorherrschenden dynamischen Umweltbedingungen immer bedeutsamer (vgl. Hausladen 2020, S. 242). Diese Arbeit beschränkt sich auf die Teilgebiete des Supply Chain Design und des Supply Chain Planning.

2.4.2 Fragestellungen des Supply Chain Design

Der einzige Aufgabenbereich des Supply Chain Designs ist die **strategische Netzwerkgestaltung**, in der das gesamte Logistiknetzwerk kosteneffektiv ausgelegt und gestaltet wird (vgl. Schulte 2017, S. 792). Nach Schulte (2017, S. 792) erfolgt dabei eine „Ausrichtung an den SCM-Strategien eines Unternehmens(-verbundes) sowie den daraus abgeleiteten Zielsetzungen des Supply Netzes“. Diese Strategien unterstützen nach Werner (vgl. 2017, S. 133) die Bereiche der Versorgung, Entsorgung und des Recyclings und werden abhängig von den Besonderheiten eines Unternehmens bestimmt. Die strategischen Aufgaben der Netzwerkgestaltung unterscheiden sich unter anderem deshalb von den mehr operativ orientierten Planungsaufgaben des Supply Chain Planning, da der Planungshorizont einen Zeitraum von mehreren Monaten bis hin zu Jahren darstellt (vgl. Kuhn und Hellgrath 2002, S. 143 f.).

Die Beurteilung von Investitionsvorhaben, beispielsweise der Aufbau von neuen Produktionskapazitäten oder Distributionszentren, gehört dabei zu den Hauptfragestellungen. Veränderungen im Logistiknetzwerk, die sich etwa auf die Anzahl und den Standort von Werken oder Lagern beziehen, haben Auswirkungen, die durch den Einsatz von IT-gestützten Planungswerkzeugen kostenmäßig beurteilt werden können. Als Ergebnis dieser Beurteilung werden die Flexibilität und Sensitivität der potenziellen Planungsszenarien, die sich hinsichtlich der Größe, Anzahl und Standort der unterschiedlichen Elemente in einem Logistiknetzwerk unterscheiden, aufgezeigt. (vgl. Schulte 2017, S. 793) Einige Beispiele für die Fragestellungen, die in der strategischen Netzwerkgestaltung aufkommen, sind:

- „Welches Produkt soll hergestellt werden?
- In welchem Werk soll ein Produkt hergestellt werden?
- Durch welchen Lieferanten soll die Materialbeschaffung erfolgen?
- Welche Distributionskanäle sollen genutzt werden?“ (Schulte 2017, S. 793)

Die Wahl des Distributionskanals wirkt sich auf die Struktur des zu planenden Systems aus. Als Beispiele für verschiedene Distributionskanäle sind der direkte und indirekte Vertrieb zu nennen. Beim direkten Vertrieb (z.B. Onlineshop) tritt der Kunde beispielsweise über das Internet direkt mit dem Hersteller in Kontakt und lässt sich das Produkt bis nach Hause liefern, sodass nur eine Distributionsstufe nötig ist. Darüber hinaus findet der Service beim Kunden statt. Der indirekte Vertrieb kann über die eigene oder eine fremde Handelsstufe geschehen. Beiden gemeinsam ist die Tatsache, dass die Kunden nicht in direktem Kontakt mit dem Händler stehen, sondern die Produkte über einen zwischengeschalteten Händler beziehen, und dass die Belieferung über mehrere Distributionsstufen erfolgt, etwa einem zentralen Lager, mehreren lokalen Lagern und schlussendlich den Händlern. Außerdem übernimmt die Handelsstufe den Service, zu dem die Wartung oder Reparatur zu zählen ist. Eine eigene Handelsstufe (z.B. Autohaus) ist ein Händler, der ausschließlich die Produkte eines bestimmten Herstellers führt. Im Gegensatz dazu steht die fremde Handelsstufe (z.B. Nahrungsmittel), bei der sich das Sortiment eines Händlers aus Produkten von verschiedenen Herstellern zusammensetzt. (vgl. Alicke 2005, S. 77 f.)

Die zu berücksichtigenden Einflussfaktoren bei der Planung des Distributionskanals nach Alicke (vgl. 2005, S. 78) lauten:

- „Die Erwartungshaltung der Kunden an die Reaktionszeit.“ (Alicke 2005, S. 78) Sie prägt die Anzahl der Distributionsstufen zwischen Hersteller und Kunde (vgl. Alicke 2005, S. 77).
- „Die räumliche Ausdehnung des bestehenden Netzwerkes.
- Die Eigenschaft des Produktes wie Wert, Prognostizierbarkeit, Umschlagshäufigkeit, Gewinnerwartung [...].
- Die Organisation des Service-Bereichs (Wartung, Reparatur).

Logistik

- Standorte (Produktion, Distribution etc.) müssen evtl. neu bestimmt werden.“ (Alicke 2005, S. 78)

Zur Netzwerkgestaltung gehört darüber hinaus die betriebliche Standortplanung. Diese beschäftigt sich mit Fragen der Standortwahl für einzelne Unternehmen, Lager, Distributionszentren oder Händler. Bei der Auswahl der Standorte wird auf verschiedene Faktoren geachtet, die sich auf die Konkurrenzfähigkeit eines Unternehmens auswirken können und standortbedingte Wettbewerbsvorteile bieten. Sollte ein Wettbewerber günstigere Standorte besitzen, sind zusätzliche Leistungen nötig, um dessen Vorteil auszugleichen. (vgl. Domschke, Drexl und Mayer, et al. 2018, S. 1) Steglich, Feige und Klaus (vgl. 2016, S. 374) nennen einige qualitative und quantitative Faktoren für die Standortwahl:

- Einsatzbezogen (z.B. Verfügbarkeit von Grund und Boden, von Produktionsfaktoren, Staatsleistungen)
- Produktionsbezogen (z.B. technologische Verhältnisse, meteorologische, geografische und geologische Gegebenheiten, politische und soziale Verhältnisse)
- Absatzbezogen (z.B. Absatzmärkte und -kontakte, Konkurrenzsituation, Entsorgung)

Nicht alle dieser Faktoren werden gleichzeitig für ein Modell genutzt, sondern schrittweise in aufeinander folgenden Modellen unterschiedlicher Aggregation einbezogen. Die Berechnung unterschiedlicher Szenarien auf der Basis unterschiedlicher Annahmen ist sinnvoll (vgl. Steglich, Feige und Klaus 2016, S. 374). Die Entscheidungsfindung für eine Anzahl von Standorten wird nach Chopra und Meindl (vgl. 2016, S. 126) in vier Phasen eingeteilt, die im Folgenden kurz erläutert werden.

Basierend auf der Wettbewerbsstrategie eines Unternehmens wird in der ersten Phase das umfassende Design und die Strategie der Supply Chain definiert. Die Strategie legt fest, welche Fähigkeiten das Netzwerk haben muss, um die Wettbewerbsstrategie zu erfüllen und das Design beinhaltet die Stufenanzahl einer Supply Chain und die Entscheidung, welche Aufgaben betriebsintern durchgeführt und welche ausgelagert werden. Ziel der zweiten Phase ist die Identifikation von Regionen, in denen die Standorte eingerichtet werden, die Definition der ungefähren Anzahl an Standorten sowie die Entscheidung der Art und Anzahl der zu fertigenden Produkte dieser Standorte. Um diese Aussagen zu treffen, werden die möglichen Regionen hinsichtlich der technologischen, makroökonomischen, politischen, und wettbewerblichen Faktoren analysiert. Aus diesen identifizierten Regionen werden in der dritten Phase auf Grundlage einer infrastrukturellen Analyse potenzielle Standorte innerhalb jener Regionen ausgewählt, an denen etwa neue Produktionsstätten oder Lager errichtet werden. In der vierten und letzten Phase werden aus den potenziellen Standorten die exakte Lage und zugehörige Kapazität jeder zu errichtenden Einrichtungen ausgewählt, sodass deren Gesamtgewinn unter Berücksichtigung der Logistik- und Gebäudekosten maximiert wird. (vgl. Chopra und Meindl 2016, S. 126 ff.)

Domschke und Drexl (vgl. 1996, S. 5) unterteilt die Vorschläge zur Lösung der betrieblichen Standortplanung in deskriptive und normative Ansätze. Die deskriptiven Ansätze werden vor allem dazu angewandt, jene Standorteigenschaften zu analysieren und systematisieren, die für die Auswahl eines Standortes von Bedeutung ist. Somit beschäftigen sich diese Ansätze nicht direkt mit der direkten Bestimmung von Standorten, sondern sie liefern Aussagen, die einen Erwägungscharakter besitzen. Es werden die Anforderungen eines Unternehmens an einen potenziellen Standort und dessen Gegebenheiten, die von dem Unternehmen genutzt werden können, gegenübergestellt. (vgl. Domschke und Drexl, Logistik: Standorte 1996, S. 6 ff.) Eine Entscheidung für einen Standort bei konkreten Planungsproblemen kann mithilfe nachprüfbarer Kriterien in Form von Modellen und Lösungsverfahren, die die Gruppe der normativen Ansätze darstellen, getroffen werden. Beispiele dafür sind Modelle zur Standortbestimmung in

Ebenen oder in Netzen und Modelle zur Bestimmung von Zentren in Netzen, auf die in Abschnitt 3.2 weiterführend eingegangen werden. Dabei müssen die Modelle nicht zwingend alle denkbaren Standortfaktoren unterstützen, da oftmals auch vereinfachte Modelle eine hinreichend genaue Abbildung der Realität darstellen. (vgl. Domschke und Drexl, Logistik: Standorte 1996, S. 9 ff.)

2.4.3 Fragestellungen des Supply Chain Planning

Das Aufgabenmodell in *Abbildung 2.3* unterteilt das Supply Chain Planning in die neun Teilgebiete Bedarfsplanung, Netzwerkplanung, Beschaffungs-, Produktions- und Distributionsplanung, Verfügbarkeits- und Machbarkeitsprüfung und Feinplanung für die Beschaffung, Produktion und Distribution. Diese Bereiche werden im Folgenden erläutert. (vgl. Kuhn und Hellingrath 2002, S. 143)

Die primäre Aufgabe der **Bedarfsplanung** ist die Prognose des lang-, mittel- und kurzfristigen Bedarfs zur Erstellung eines Absatzplanes, der die Basis für die Teilgebiete der Beschaffungs-, Produktions- und Distributionsplanung darstellt (vgl. Hellingrath und Witthaut 2004, S. 26). Dabei liegt der Fokus auf der Prognose des mittel- und langfristigen Bedarfs. Zum langfristigen Bedarf zählt die Vorhersage des Absatzes einer Produktgruppe, bzw. -familie über einen Zeitraum von mehreren Jahren und beim mittelfristigen Bedarf wird für einen Zeitraum von mehreren Monaten der Marktbedarf auf Produkttypenebene prognostiziert (vgl. Kuhn und Hellingrath 2002, S. 145). Sind zu unterschiedlichen Betrachtungsebenen eines Produkts oder Zeiträumen bereits Vergangenheitsdaten vorhanden, eignen sich statistische Verfahren für die Prognose. Soll der aus Werbe- oder Preismaßnahmen resultierende Bedarf oder der Absatz eines neuen Produktes prognostiziert werden, so kann auf Prognose- und Simulationsinstrumente zurückgegriffen werden. (vgl. Schulte 2017, S. 793) Im Gegensatz zum lang- und mittelfristigen Bedarf soll der kurzfristige Bedarf nicht prognostiziert, sondern transparent gestaltet werden. Dafür werden alle Bestellungen, die in ERP-Systemen gespeichert sind, über die jeweiligen Supply-Chain-Stufen ausgewertet (vgl. Schulte 2017, S. 793). Nach Schönsleben (vgl. 2016, S. 406) können Vorhersageverfahren in vergangenheitsbasierende und zukunftsbasierende Verfahren gegliedert werden. Die erste Gruppe prognostiziert die zukünftige Nachfrage aufgrund von Verbräuchen aus der Vergangenheit, beispielsweise Verbrauchsstatistiken. Sie lässt sich weiterhin in mathematische Verfahren, wie der verbreiteten Extrapolation von Zeitreihen, oder in graphische Verfahren zur Vorhersage unterteilen, bei denen ein mittlerer Verlauf einer graphisch aufgetragenen Zeitreihe zur Prognose herangezogen wird. Bei den zukunftsbasierten Verfahren werden bereits vorhandene Informationen über die zukünftige Nachfrage, wie feste Bestellungen, berücksichtigt. Diese können in einem weiteren Schritt wieder einerseits in mathematische Verfahren, die für das Beispiel einer Hochrechnung aus bereits getätigten Bestellungen und einem Erfahrungswert den Absatz prognostizieren, und intuitive Vorhersageverfahren geteilt werden. Zu diesen zählen die Befragung oder die Schätzung die auf Informationen durch den direkten Kundenkontakt, über Verkäufer oder spezielle Marktforschungsinstitute basieren. (vgl. Schönsleben 2016, S. 406 f.) Je genauer der Bedarf vorhergesagt wird, desto besser können Kundenbedarfe befriedigt, Kapazitäten in der Supply Chain geplant und darin enthaltenen Bestände optimiert werden. Um negativen Trends entgegenzuwirken, können seitens des Marketings zusätzliche Werbe- oder Preismaßnahmen ergriffen werden. (vgl. Kuhn und Hellingrath 2002, S. 144 f.)

Bei der **Netzwerkplanung** geht es darum, die einzelnen Partner einer Supply Chain oder eines Netzwerkes über einen Planungshorizont von mehreren Monaten bis hin zu Jahren übergreifend zu koordinieren (vgl. Kuhn und Hellingrath 2002, S. 145 f.). Diese Planung kann sich einerseits auf die unternehmensinternen Produktions- oder Logistikstandorte eines Geschäftsbereichs oder andererseits auf unternehmensübergreifende Netzwerke entlang der gesamten logistischen Kette, inklusive der gegenseitigen

Logistik

Abstimmung der darin beteiligten Partner, beziehen (vgl. Werner 2017, S. 89). Die Netzwerkplanung wird nach Kuhn und Hellingrath (vgl. 2002, S. 145) meist von dem Partner des Produktions- und Logistiknetzes vorgenommen, der „neben einem großen Wertschöpfungsanteil auch die größte Nähe zu den Endkunden besitzt“, da dieser die Gesamtsituation mithilfe der ihm zur Verfügung stehenden Informationen am besten beurteilen kann und die generellen Zusammenhänge der Netzwerksteuerung und -planung kennt (vgl. Kuhn und Hellingrath 2002, S. 145). Allgemeines Ziel dieses Teilgebiets ist es, hinsichtlich verschiedener Kriterien das Optimum des Gesamtnetzes zu ermitteln. Beispiele für Optimierungskriterien sind:

- Kapazitätsauslastung
- Nähe zum Absatzmarkt
- Produktionsquoten
- Materialverfügbarkeit

Um dieses Ziel zu erreichen, werden die aus den Absatzplänen der Bedarfsplanung resultierende Bedarfe mit den Beständen, bzw. Kapazitäten der einzelnen Orte abgestimmt. Dabei müssen zusätzlich die grundlegenden Strukturen der Produktions- und Logistikstandorte des logistischen Netzes und ihre gegenseitigen Lieferbeziehungen berücksichtigt werden. Durch diese Abstimmung erhalten die Werke das für sie geltende Produktionsprogramm. (vgl. Hellingrath und Witthaut 2004, S. 27 f.) Es wird auch als Mengengerüst der geplanten Produktion bezeichnet und legt die Arten und Mengen der Produkte fest, die innerhalb des Planungshorizontes zu fertigen sind (vgl. Schulte 2017, S. 581 f.). Mit diesen Informationen liefert das Produktionsprogramm den Input für die Aufgaben der Beschaffungs-, Produktions- und Distributionsplanung (vgl. Schulte 2017, S. 795).

Die Aufgabe der **Beschaffungsplanung** ist es, die Teileversorgung bzw. Bestände einer mehrstufigen Lagerstruktur so zu optimieren, dass die Bereitstellung von genügend Materialien bzw. Teilen am exakten Ort sichergestellt ist, wobei die Bestände möglichst niedrig gehalten werden. Der Planungshorizont erstreckt sich dabei von Tagen bis hin zu Wochen und liegt damit im taktischen Bereich. (vgl. Kuhn und Hellingrath 2002, S. 146) Ein wichtiger Planungstool ist neben den Ergebnissen der Bedarfs- und Netzwerkplanung die Stückliste, da aus ihrer Auflösung die Primär-, Sekundär- und Tertiärbedarfe ermittelt werden können (vgl. Werner 2017, S. 90). Ausgehend von diesen Bedarfen und den zu erwartenden Prognoseabweichungen werden zusätzlich die Minimal-, Maximal- und optimalen Sicherheitsbestände bestimmt (vgl. Kuhn und Hellingrath 2002, S. 146). Weitere Kriterien stellen die Kapazitäten der Lieferanten, der Transporte und der Lager sowie Anlieferhythmen und Wiederbeschaffungszeiten dar (vgl. Hellingrath und Witthaut 2004, S. 30). Zur Identifikation einer optimierten Bedarfserfüllung werden darüber hinaus Simulationen verschiedener Versorgungsszenarien verglichen (vgl. Kuhn und Hellingrath 2002, S. 146). Diese Szenarien berücksichtigen verschiedene Beschaffungsstrategien, wie den Bezug aus nur einer Einkaufsquelle (Single Sourcing) (vgl. ten Hompel und Heidenblut 2011, S. 285), den Bezug aus mehreren Einkaufsquellen (Multiple Sourcing) (vgl. ten Hompel und Heidenblut 2011, S. 202) oder der Entscheidung, eine Produktion nicht selbst auszuführen, sondern an Dritte zu vergeben (Make or Buy) (vgl. ten Hompel und Heidenblut 2011, S. 189). Aus diesem Vorgehen resultiert der Beschaffungsplan, welcher die Basis für die Beschaffungsfineplanung bildet. (vgl. Kuhn und Hellingrath 2002, S. 146)

Bei der **Produktionsplanung** wird aus den Vorgaben der Absatz- und Netzwerkplanung ein optimierter Produktionsplan für ein Werk erstellt, der die Auslastung optimiert, die Bestandskosten minimiert und gleichzeitig die Lieferbereitschaft und Termintreue sicherstellt. Der Zeitraum, für den geplant werden

Logistik

soll, bewegt sich im Monats- bis Wochenbereich. (vgl. Schulte 2017, S. 795) Als Ergebnis wird ein grober Produktionsplan für ein spezifisches Werk erstellt, der auf dem Produktionsprogramm der Netzwerkplanung basiert und sich aus prognostizierten Kapazitätsbedarfen, den für die Beschaffung bedeutenden Mengenbedarfen, und einem Produktionsplan zusammensetzt, der den Fertigungsaufträgen einer Planungsperiode die Kapazitäten sowie das benötigte Material zuordnet. Dafür wird auf Methoden der integrierten Mengen-, Termin- und Kapazitätsplanung zurückgegriffen, auf die im Folgenden nicht näher eingegangen wird. (vgl. Kuhn und Hellingrath 2002, S. 147)

Das Aufgabengebiet der **Distributionsplanung** beschäftigt sich auf Grundlage der in der Bedarfsplanung prognostizierten Absatzpläne und der Produktionsprogramme der Netzwerkplanung mit der optimierten Planung der Produktverteilung zum Kunden und der Lagerbestände in der Distributionskette, sodass die Nachfrage befriedigt werden kann. Dabei werden meist mehrstufige Distributionsstrukturen betrachtet und der Planungshorizont reicht von Tagen bis zu Monaten. (vgl. Kuhn und Hellingrath 2002, S. 147) Die grobe Planung von Transportkapazitäten wird von den Bedarfen und Beständen sowie dem Typ und den Standorten der jeweiligen Lager beeinflusst und die Transportkosten lassen sich durch die Bildung von Kombinationen einzelner Lieferungen optimieren. Darüber hinaus können im mittelfristigen Planungszeitraum unterschiedliche Versorgungsszenarien, wie verschiedene Distributionskanäle oder Transportmittel, bewertet und miteinander verglichen werden. (vgl. Hellingrath und Witthaut 2004, S. 28) Der daraus entstehende Distributionsplan liefert den Input für die Distributionsfeinplanung (vgl. Schulte 2017, S. 795). Vastag (vgl. 2008, S. 425 ff.) unterteilt die Distributionsplanung in die strategische, taktische und operative Planungsebene mit einem eigenen Planungshorizont und abgegrenzten Inhalten. Einen Überblick über diese Gliederung liefert **Tabelle 2.1** (vgl. Vastag 2008, S. 425 ff.).

Tabelle 2.1: Planungsebenen der Distributionsplanung

Planungsebene	Planungshorizont	Inhalt
Strategisch	Langfristig	<ul style="list-style-type: none"> • Festlegung der Strategie • Planung der Absatzwege
Taktisch	Vorwiegend mittelfristig	<ul style="list-style-type: none"> • Entscheidung über Distributionssystemstruktur • Auswahl Verkehrsträgerwahl • Kundenzuordnung der Lager
Operativ	Kurzfristig	<ul style="list-style-type: none"> • Lagerdisposition • Transport- und Lagertechniken • Tourenplanung

Die **Verfügbarkeits- und Machbarkeitsprüfung** wird auch als **Order Promising** bezeichnet (vgl. Werner 2017, S. 90). Innerhalb des Aufgabenmodells nimmt sie die Position zwischen den Planungsaufgaben, die keinen Bezug zu den Kundenaufträgen haben, und den kundenauftragsbezogenen Planungsaufgaben ein. Dabei wird geprüft, ob Kundenanfragen, bzw. -aufträge erfüllbar sind, was in einer Steigerung der Kundenorientierung und dem Kundennutzen resultiert. (vgl. Hellingrath und Witthaut 2004, S. 31) Das Teilgebiet des Order Promising kann verschiedene Ausprägungsformen annehmen (vgl. Schulte 2017, S. 795). Unter Available-to-Promise (ATP) wird die Möglichkeit verstanden, die Liefermengen und -termine bei Kundenanfragen direkt zuzusichern. Dabei wird anhand des freien Bestands und der eingeplanten Produktionsmenge überprüft, ob das gewünschte Produkt verfügbar ist. (vgl. Beckmann 2012, S. 482) ATP zeichnet sich somit dadurch aus, dass es in Richtung des Kunden gerichtet ist. Im Gegensatz dazu wirkt die Ausprägungsform Capable-to-Promise (CTP) in Richtung des

Logistik

Unternehmens, da sie die Fähigkeit eines Unternehmens beschreibt, das dem Kunden gegebene Lieferversprechen auch einzuhalten. (vgl. Werner 2017, S. 91) Dies geschieht durch die Prüfung der Produktions- und Distributionspläne unter Berücksichtigung der verfügbaren Materialien und Kapazitäten. Die letzte Ausprägungsform des Order Promising ist Configure-to-Promise (CoTP) und sie umschreibt neben der Produktkonfiguration nach den Wünschen des Kunden zusätzlich die Prüfung der dazu erforderlichen Kapazitäten und Materialien, um einen Liefertermin zuzusichern. (vgl. Beckmann 2012, S. 482 f.) Unabhängig des Anwendungsfalls ist für eine Kundenanfrage kritisch zu prüfen, wie qualifiziert sie beantwortet werden muss. Während für einfache Anfragen beispielsweise lediglich überprüft werden muss, ob das angefragte Produkt an einem Lagerort zu einem gewissen Zeitpunkt erhältlich ist, erfordern einzelne kundenindividuelle Produktkonfigurationen eine exakte Ermittlung des Produktionsstandortes, der Durchlaufzeit und Distributionszeit. Bei solchen Anfragen werden dann Simulationen oder Optimierungen eingesetzt. (vgl. Kuhn und Hellingrath 2002, S. 148) Die Planungsergebnisse des Order Promising unterscheiden sich ebenfalls je nach Kundenanforderung. So kann für eine angefragte Produktmenge etwa der frühestmögliche Liefertermin identifiziert oder im Falle einer speziellen Produktkonfiguration der gewünschte Lieferzeitpunkt und die Liefermenge bestätigt werden. Sollte der Kundenwunsch mengenmäßig oder zeitlich nicht umsetzbar sein, werden auch alternative, aber lieferbare Produktvarianten vorgeschlagen. (vgl. Schulte 2017, S. 795)

Aufbauend auf den Ergebnissen der Beschaffungs- und Produktionsfeinplanung zählt die optimierte Planung der Anlieferungen zum Aufgabengebiet der **Beschaffungsefinplanung**. Der Planungshorizont dieses Teilgebietes ist kurzfristig und liegt damit genauer gesagt zwischen Tagen bis Stunden. Zur Optimierung der Anlieferungsmengen wird in einem ersten Schritt unter Beachtung der Lagerbestände aus dem Bruttosekundärbedarf der Nettosekundärbedarf bemessen. (vgl. Hellingrath und Witthaut 2004, S. 31) Danach werden die optimalen Anlieferungsmengen für den jeweiligen Planungshorizont errechnet, wobei die in der Beschaffungsplanung bestimmten Minimal-, Maximal- und Sicherheitsbestände in die Berechnung miteinbezogen werden (vgl. Schulte 2017, S. 796). Bei diesen Schritten „müssen alle internen, aber auch externen Kapazitäten und Restriktionen wie Anlieferhythmen, Wiederbeschaffungszeiten und -schwankungen, Wareneingang, etc. berücksichtigt werden“ (Hellingrath und Witthaut 2004, S. 31).

Bei der **Produktionsfeinplanung** geht es darum, den Produktionsbereich für einen mittel- und mehr noch für einen kurzfristigen Zeitraum zu planen. Auf Grundlage des Produktionsplans lassen sich konkrete Fertigungs- und/oder Montageaufträge festlegen, terminieren und freigeben. Zusätzlich zum Produktionsplan sind für dieses Teilgebiet die tatsächlichen Verfügbarkeiten von Materialien und Produktionskapazitäten von Bedeutung. (vgl. Kuhn und Hellingrath 2002, S. 151) Sollten dabei kurzfristig Änderungen oder Störungen auftreten, müssen die Produktionsaufträge nach Rücksprache mit der Beschaffungs- und Distributionsseite flexibel angepasst werden (vgl. Schulte 2017, S. 796). Während diese Produktionsaufträge in der Regel für einen Zeitraum von ungefähr ein bis acht Wochen geplant werden, wird die Produktionsfeinplanung meist täglich durchgeführt, da so auf den aktuellen Produktionsfortschritt eingegangen werden kann. Beispiele für weitere Optimierungskriterien sind die Durchlaufzeit, der Werkstattbestand und die Priorität des Auftrags oder des Kunden. (vgl. Hellingrath und Witthaut 2004, S. 30) Der Unterschied zur Produktionsplanung, die für gewöhnlich für ein Werk durchgeführt wird und die Umfänge der Produktion maximal auf der Ebene der Kapazitätsgruppe plant, ist der, dass eher dezentral für separate Produktionsbereiche geplant wird. Darüber hinaus werden in der Produktionsfeinplanung die Sequenz von unterschiedlichen Produktionsaufträgen und die tatsächlichen Kapazitätseinheiten definiert. (vgl. Kuhn und Hellingrath 2002, S. 151)

Logistik

Die **Distributionsfeinplanung** befasst sich mit der Aufgabe, die Touren, Transportmittel und deren Beladung in Hinblick auf eine termingerechte Belieferung optimal zu planen. Neben der hohen Liefertreue zielt dieses Teilgebiet auf eine geringe Lieferzeit bei gleichzeitig niedrigen Kosten ab. Dafür werden die Ergebnisse der Distributionsplanung für einen kurzfristigen Zeithorizont verfeinert. (vgl. Schulte 2017, S. 796) Dieser liegt nach (vgl. Hellingrath und Witthaut 2004, S. 28) im Bereich von wenigen Wochen, kann aber auch den Horizont von Tagen oder sogar Stunden einnehmen. Diese kurzen Planungszeiträume führen dazu, dass eine Vielzahl von Transportaufträgen zu berücksichtigen sind, um mit möglichst günstigem Einsatz der vorhandenen Ressourcen die vorgegebenen Aufträge zu erfüllen. So werden bei der Fahrzeugdisposition beispielsweise mehrere Aufträge zu kompletten Transportladungen vereint, da ein einzelner Auftrag selten ein Transportmittel auslastet. (vgl. Hellingrath und Witthaut 2004, S. 28) Im Vorfeld einer auftragsbezogenen, endgültigen Festlegung eines Transportmittels für die Auslieferung werden die Kosten verschiedener Transportalternativen errechnet und gegenübergestellt. Weitere zu berücksichtigende Restriktionen stellen etwa verfügbare Kapazitäten, gesetzliche Vorschriften oder Zeiten für die Anlieferung, bzw. Abholung dar. Wird ein Transport nicht durch die eigene Transportflotte abgewickelt, sondern durch einen Spediteur übernommen, kann dieser durch computergestützte Transportbörsen gefunden werden, die eine Übersicht über freie Frachtkapazitäten geben. (vgl. Kuhn und Hellingrath 2002, S. 151 f.) Laut Hellingrath und Witthaut (vgl. 2004, S. 28) werden die Teilgebiete der Distributionsplanung und der Distributionsfeinplanung in anderen deutschsprachigen Quellen oft zusammengefasst und als Touren- und Routenplanung bezeichnet. Weitere Beispiele für Planungsaufgaben dieses Teilgebietes sind neben der Fahrzeugdisposition auch die Bestimmung einer effizienten Reihenfolge der anzufahrenden Standorte einer Tour oder die Zusammenfassung der zu beliefernden Kunden zu einer möglichst kostengünstigen Tour (vgl. Hellingrath und Witthaut 2004, S. 28).

3 Graphbasierte Verfahren

Abschnitt 3.1 vermittelt die graphentheoretischen Grundlagen, die benötigt werden, um die Thematik zu verstehen. Abschnitt 3.2 thematisiert die Modelle, also die Abstraktion von realen Anwendungsfällen. Abschnitt 3.3 geht dann auf die Algorithmen ein, die die Berechnungsmethoden zur Durchführung einer Aufgabe darstellen.

3.1 Grundlagen der Graphentheorie

In diesem Abschnitt wird auf die wissenstheoretischen Grundlagen der Graphentheorie eingegangen. Dabei werden grundsätzliche Begriffe erläutert, auf die Eigenschaften von Graphen eingegangen und beschrieben, wie Graphen repräsentiert werden.

Zuerst wird der Begriff des Graphen definiert:

„Ein Graph G ist ein Tupel (V, E) , wobei V eine (endliche) nicht-leere Menge von Knoten (engl. vertices) ist. Die Menge E ist eine Teilmenge der zweielementigen Teilmengen von V , also $E \subseteq \binom{V}{2} := \{\{x, y\} | x, y \in V, x \neq y\}$. Die Elemente der Menge E bezeichnet man als Kanten (engl. edges).“ (Steger 2002, S. 52)

Zur Darstellung eines Graphen wird jeder Knoten durch je einen Punkt und jede Kante durch je einen Strich repräsentiert. Bezogen auf die Definition eines Graphen werden die Endpunkte x und y durch einen Strich (x, y) verbunden, der Graph ist somit ungerichtet. (vgl. Steger 2002, S. 52) Wird bei den Kanten die Reihenfolge der Punkte berücksichtigt, handelt es sich um einen gerichteten Graph. Bei einem gerichteten Graphen wird der Strich zu einem Pfeil und die beiden Punkte werden je nach der Pfeilrichtung zu einem Anfangs- und einem Endpunkt. (vgl. Turau und Weyer 2015, S. 19)

Mehrere Kanten werden als inzident bezeichnet, wenn sie über einen gemeinsamen Knoten verbunden sind. Analog werden verschiedene Knoten als benachbart oder adjazent definiert, wenn sie durch eine Kante verbunden sind. Handelt es bei den beiden Punkten x und y um ein und denselben Knoten, so ist die verbindende Kante eine Schlinge. Zwei oder mehr Kanten sind parallel bzw. werden als Mehrfachkanten bezeichnet, wenn sie die gleichen Punkte x und y verbinden. (vgl. Volkmann 1991, S. 1 f.) Der Grad $d(x)$ eines Knoten x ergibt sich aus der Anzahl der Kanten, die mit dem Knoten verbunden sind. Ist ein Knoten beispielsweise ohne Nachbar, so besitzt er den Grad $d(x) = 0$ und wird auch als isolierter Knoten bezeichnet. (vgl. Dörfler und Mühlbacher 1973, S. 8)

Graphen ohne Schlingen und Mehrfachkanten heißen auch schlichte Graphen (vgl. Domschke 2007, S. 1). Zu den weiteren Ausprägungsformen zählen Multigraphen, bei denen Schlingen und/oder Mehrfachkanten vorliegen, und vollständige Graphen, deren Knoten allesamt paarweise miteinander verbunden sind (vgl. Steger 2002, S. 52 f.). Lässt sich die Menge V der Knoten in zwei Teilmengen V_1 und V_2 unterteilen, sodass V_1 und V_2 kein gemeinsames Element besitzen und weder Knoten aus V_1 , noch Knoten aus V_2 untereinander benachbart sind, heißt der Graph bipartit (vgl. Turau und Weyer 2015, S. 25). Ein Subgraph beziehungsweise Teilgraph $G_1 = (V_1, E_1)$ liegt vor, wenn $V_1 \subseteq V$ und $E_1 \subseteq E$ (vgl. Dörfler und Mühlbacher 1973, S. 8). Setzt sich die Kantenmenge eines Graphen genau aus den zu einem Knoten inzidenten Kanten zusammen, liegt ein Sterngraph vor, der in **Abbildung 3.1** nach Steger (vgl. 2002, S. 81) dargestellt ist. Durch die Zuordnung von Kantengewichten zu allen Kanten der Menge E können alle genannten Graphentypen auch als gewichteter Graph betrachtet werden (vgl. Mahlmann und

Graphbasierte Verfahren

Schindelhauer 2007, S. 265). Das Kantengewicht ist dabei eine Zahl, die beispielsweise für Kosten oder den Abstand zweier Knoten steht (vgl. Krumke und Noltemeier 2012, S. 169).

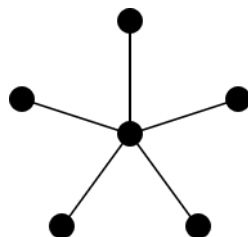


Abbildung 3.1: Beispiel für einen Sterngraphen

Eine Folge von Kanten eines Graphen wird als Kantenzug bezeichnet, wenn die beiden aufeinanderfolgenden Knoten der jeweiligen Kanten benachbart sind. Die Länge des Kantenzugs entspricht der Anzahl der durchlaufenen Kanten. Sind alle Kanten unterschiedlich, wird der Kantenzug als Weg bezeichnet. Der Weg ist offen/geschlossen, wenn der Anfangs- und Endknoten verschieden/identisch sind. (vgl. Turau und Weyer 2015, S. 22) Besteht ein geschlossener Weg mit Ausnahme des Anfangs-/Endknoten aus verschiedenen Knoten, handelt es sich um einen Kreis. Existiert zwischen jedem Knotenpaar eines gerichteten oder ungerichteten Graphen ein Kantenzug, so ist dieser Graph zusammenhängend. Bei einem gerichteten Graphen kann es trotz dieser Verbindung vorkommen, dass ein Knoten aufgrund der Pfeilrichtungen einen zweiten Knoten nicht erreichen kann. Für den Fall, dass in einem gerichteten Graphen jeder Knoten von jedem Knoten erreichbar ist, wird dieser als stark zusammenhängend bezeichnet. (vgl. Domschke 2007, S. 2 f.)

Ein ungerichteter Graph wird als Wald bezeichnet, wenn er keinen Kreis enthält. Ist dieser Graph zusätzlich zusammenhängend, handelt es sich um einen Baum. (vgl. Krumke und Noltemeier 2012, S. 101) Ein aufspannender Baum liegt vor, wenn ein Teilgraph eines Graphen dieselbe Knotenmenge wie der Graph selbst hat und gleichzeitig ein Baum ist. Bäume finden häufig Verwendung, wenn hierarchische Beziehungen dargestellt werden sollen (vgl. Turau und Weyer 2015, S. 57 f.).

Ein zur Beschreibung der Problemstruktur genutzter Graph ist bei vielen Anwendungen von graphentheoretischen Methoden zu mächtig, als dass das zugehörige Problem manuell gelöst werden könnte. Aus diesem Grund werden elektronische Rechenanlagen genutzt, für die ein Graph in einer geeigneten Speicherungsform vorliegt. (vgl. Noltemeier 1976, S. 38) Die simpelste Speicherungsform eines Graphen stellt die Adjazenzmatrix dar (vgl. Krumke und Noltemeier 2012, S. 18). Für n Knoten beträgt die Größe der Matrix $n \times n$ und der Eintrag a_{ij} ist 1, wenn eine Kante zwischen Knoten i und Knoten j existiert, ansonsten gleich 0. Dadurch lassen sowohl ungerichtete Graphen, für die die Matrix symmetrisch ist, als auch gerichtete Graphen darstellen. Nimmt ein Eintrag auf der Hauptdiagonalen den Wert 1 an, handelt es sich dabei um eine Schlinge. (vgl. Turau und Weyer 2015, S. 30) Eine zweite Möglichkeit für die Speicherung eines Graphen mit n Knoten und m Kanten ist die Inzidenzmatrix, die eine Größe von $n \times m$ besitzt. Ist ein Knoten i in einem ungerichteten Graphen mit einer Kante j verbunden, nimmt der Eintrag b_{ij} den Wert $b_{ij} = 1$ an, andernfalls den Wert $b_{ij} = 0$. Für einen gerichteten Graphen beträgt der Wert $b_{ij} = 1$, wenn der Knoten i ein Anfangsknoten der Kante j ist, $b_{ij} = -1$, wenn der Knoten i ein Endknoten der Kante j ist und $b_{ij} = 0$ in den übrigen Fällen. (vgl. Dörfler und Mühlbacher 1973, S. 55 f.) Die Darstellung einer Schlinge ist in der Inzidenzmatrix nicht möglich. Die üblichste und in den meisten Fällen auch geeignetste Speicherung eines Graphen ist die Adjazenzlisten-Repräsentation. Für n Knoten besteht diese aus einem Array ADJ aus n Adjazenzlisten. Jedem Knoten v ist somit

Graphbasierte Verfahren

eine Liste $ADJ[v]$ zugewiesen, die im ungerichteten Fall Zeiger auf alle verbundenen Knoten w und im gerichteten Fall Zeiger auf alle Knoten w , die Nachfolger von v sind, enthält. Diese Repräsentation eignet sich ebenfalls für die Speicherung von Schlingen und Mehrfachkanten. Vorteile dieser Technik sind neben dem geringen Speicherbedarf die einfache Iteration über die Nachbarn eines Knoten. (vgl. Krumke und Noltemeier 2012, S. 20 f.) In **Abbildung 3.2** sind die Speicherungsformen eines kleinen Graphen als Adjazenzmatrix und Adjazenzliste angelehnt an Steger (vgl. 2002, S. 64) illustriert.

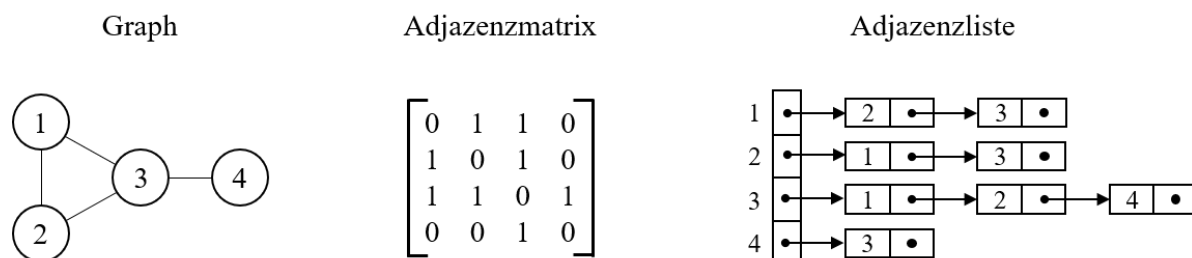


Abbildung 3.2: Speicherungsformen eines Graphen

Würde ein Graph in einem relationalen Datenmodell abgespeichert werden, würden die Tabellen, in denen Datensätze abgespeichert sind, schnell sehr groß werden, da die Einträge mit der Anzahl der verbundenen Elemente exponentiell zunehmen. Außerdem sind die Abfragen aus Sicht der Datenbank sehr teuer. (vgl. Fasel 2016, S. 122) Für die Speicherung von Graphen wird auf Graphdatenbanken zurückgegriffen, die eine spezielle Klasse von NoSQL-Datenbanken darstellen (vgl. Saake, Sattler und Heuer 2018, S. 689). NoSQL-Datenbanken beschreiben Datenbanktechnologien, die nicht-relationale Datenmanagement-Ansätze verwenden, also keine Daten in Tabellen abspeichern und SQL nicht als Datenbanksprache verwenden (vgl. Meier und Kaufmann 2016, S. 18). Graphdatenbanken besitzen im Gegensatz zu relationalen Datenbanken ein Datenmodell, mit dem sich komplexe Beziehungen besser verarbeiten lassen (vgl. Sadalage und Fowler 2013, S. 10). Außerdem werden Abfragen des Graphen durch das Hinzufügen von neuen Knoten und Kanten nicht kostenintensiver (vgl. Fasel 2016, S. 122). Deshalb werden Graphdatenbanken überall dort eingesetzt, wo die Verknüpfung der Datensätze wichtiger sind als die einzelnen Datensätze an sich, also wo Daten in Netzwerken organisiert sind (vgl. Meier und Kaufmann 2016, S. 238).

Zwei konkrete Graph-Datenmodelle für Graphdatenbanken sind das Resource Description Framework (RDF) Modell und das Property-Graph-Modell (vgl. Saake, Sattler und Heuer 2018, S. 694). Das RDF ist eine spezifizierte Sprache, die ursprünglich zur Repräsentation von Informationen über Web-Ressourcen entwickelt wurde und inzwischen ein primäres Element verschiedener Semantic-Web-Technologien darstellt. Das Semantic-Web erweitert das klassische Web, indem alle Web-Informationen mit eindeutigen semantischen Metadaten versehen werden, damit die Inhalte nicht nur von Menschen verarbeitet werden können, sondern auch eine maschinelle Verarbeitung möglich wird. (vgl. Saake, Sattler und Heuer 2018, S. 692 f.) Ein RDF-Graph setzt sich aus einer Menge von Tripeln zusammen, die der Repräsentation von Informationen über Ressourcen dienen und jeweils aus den Komponenten Subjekt, Prädikat und Objekt bestehen. Die Ressource, auf die sich eine Aussage bezieht, wird durch das Subjekt dargestellt. Das Prädikat steht für eine Eigenschaft und das Objekt für den Wert einer Eigenschaft. (vgl. Saake, Sattler und Heuer 2018, S. 694 f.) Somit stellt jedes Tripel eine logische Aussage einer Beziehung zwischen Subjekt und Objekt dar (vgl. Angles und Gutierrez 2018, S. 13). Da Knoten und Kanten beim RDF-Modell keine eigene Struktur haben, können Eigenschaften von Knoten nur über Prädikate zugewiesen werden.

Graphbasierte Verfahren

Für die direkte Modellierung von Knoten und Kanten mit Eigenschaften, wird deshalb auf das Property-Graph-Modell zurückgegriffen. (vgl. Saake, Sattler und Heuer 2018, S. 717 f.) Nach Meier und Kaufmann (vgl. 2016, S. 14) wird ein Property-Graph auch als Eigenschaftsgraph bezeichnet, der aus Knoten und gerichteten Kanten besteht, die jeweils einen Namen (engl. label) tragen und Eigenschaften (engl. properties) aufweisen können. „Die Eigenschaften werden als Attribut-Werte-Paare der Form (attribute: value) mit Attributnamen und zugehörigen Werten charakterisiert.“ (Meier und Kaufmann 2016, S. 14) Auch wenn das Property-Graph-Modell eine flexible Datenspeicherung und Datenzugriff mit vielen Beziehungen ermöglicht, sind für einige Anwendungen erweiterte Graphmodelle erforderlich. Im Folgenden wird auf Hypergraphen und verschachtelte Graphen (engl. nested graphs) eingegangen, die die grundlegenden Strukturen eines Graphen durch die Generalisierung von Knoten bzw. (binären) Kanten erreichen. (vgl. Wiese 2015, S. 58) Bei Hypergraphen wird das Konzept einer Kante auf eine Hyperkante (engl. hyperedge) erweitert, die eine beliebige Menge von Knoten in Verbindung setzt. Die Knoten eines verschachtelten Graphen können selbst Graphen sein und werden auch als Hyperknoten (engl. hypernodes) bezeichnet. (vgl. Angles und Gutierrez 2018, S. 9)

Abschließend wird in diesem Abschnitt die Sichtweise von Tarjan (vgl. 2005, S. 17) über den Prozess der Forschung, Entdeckung und Entwicklung beschrieben, um das Zusammenspiel von Anwendungsfällen aus der realen Welt, deren Modellierung und Algorithmen zur Lösung einer Aufgabe zu verstehen. Der Prozess nach Tarjan (vgl. 2005, S. 18) wird in **Abbildung 3.3** dargestellt.

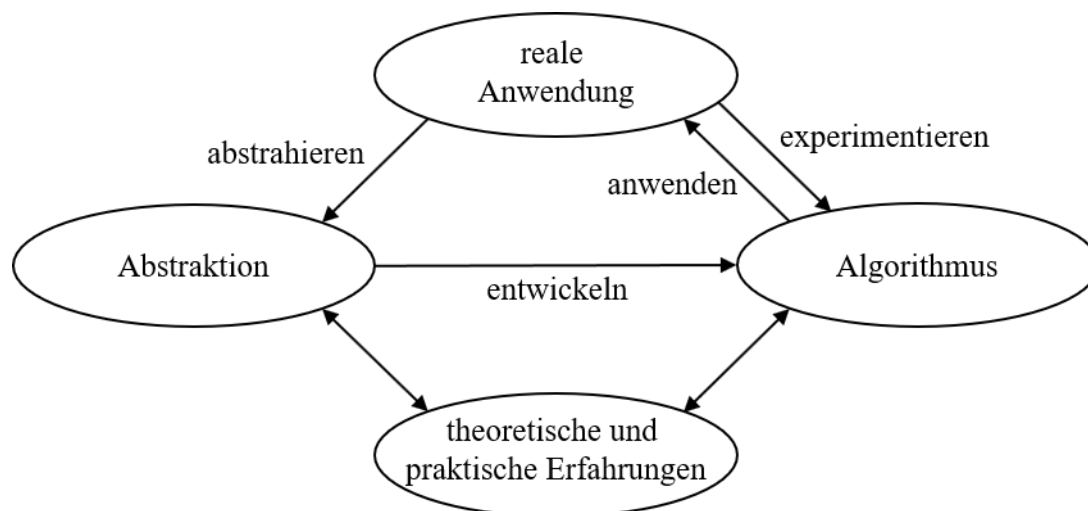


Abbildung 3.3: Prozess der Forschung, Entdeckung und Entwicklung

Tarjan betrachtet in einem ersten Schritt die Anwendung oder das Einsatzgebiet aus der realen Welt, zu der eine Aufgabe existiert, die gelöst werden soll. Aufgrund der Tatsache, dass die reale Welt sehr chaotisch ist, wird diese Anwendung zunächst abstrahiert und als Problem modelliert, welches mit dem passenden theoretischen Hintergrund angegangen werden kann. Im Bereich der Informatik wird basierend auf dieser Abstraktion versucht, einen Algorithmus zu entwickeln, der eine Berechnungsmethode zur Durchführung einer Aufgabe darstellt. Dieser wird anschließend auf die Anwendung aus der realen Welt angewendet und es wird überprüft, ob das reale Problem gelöst werden kann. Sollte der Algorithmus nicht die richtigen Teile des Problems erfassen, kann er die Aufgabe in der realen Welt nicht lösen, sodass eine erneute Iteration durchgeführt wird. Die reale Anwendung und der entwickelte Algorithmus, welcher auf der Abstraktion basiert, stehen also in einem Wechselspiel zueinander. Bei einer neuen Iteration wird die Abstraktion angepasst, dieses neue abstrakte Modell wird gelöst und es wird versucht, die Erkenntnisse in der Praxis anzuwenden. Während des gesamten Prozesses werden neue theoretische

Graphbasierte Verfahren

und praktische Erfahrungen gewonnen, die auf andere Bereiche übertragen werden können. Tarjan merkt an, dass das Verständnis des Problems, die angemessene Darstellung von Problemdaten und die Betrachtung der Operationen, die mit den Daten durchgeführt werden, essenziell wichtig für eine effiziente Durchführung von Berechnungen ist. Diese Punkte führen dazu, dass viele algorithmische Probleme zu Datenmanipulationsproblemen werden und der primäre Schritt zu dessen Lösung die Entwicklung der richtigen Datenstruktur ist. (vgl. Tarjan 2005, S. 17 f.) Für den Entwurf eines effizienten Algorithmus für Graphen gibt es also keine Lösung, da fast jede Problemstellung eine andere Vorgehensweise erfordert (vgl. Turau und Weyer 2015, S. 28).

3.2 Graphbasierte Modelle im Supply Chain Management

Nach der in Abschnitt 2.1 beschriebenen Definition, managt die Logistik Fließsysteme, „denen die Modellvorstellung von Unternehmen und deren Agieren in der Umwelt in Gestalt eines Netzwerks zugrunde liegt“ (Steglich, Feige und Klaus 2016, S. 2). Eine umfassende Betrachtung einer Unternehmenslogistik wird ermöglicht, wenn das Netzwerk als Strukturmodell dargestellt wird und dessen Flüsse als Modell der ablaufenden Prozesse. Zu den Bestandteilen dieses Modells zählen zeitlich ablaufende Prozesse und die Verkettung bzw. die Vernetzung von Aktivitäten. Auf diese Weise können komplexe Probleme mithilfe eines Modells behandelt werden. (vgl. Steglich, Feige und Klaus 2016, S. 2) Die Definition eines Modells lautet:

„Vereinfachte Nachbildung eines geplanten oder existierenden Systems mit seinen Prozessen in einem anderen begrifflichen oder gegenständlichen System“, wobei es „sich hinsichtlich der untersuchungsrelevanten Eigenschaften nur innerhalb eines vom Untersuchungsziel abhängigen Toleranzrahmens vom Vorbild“ unterscheidet (VDI-Fachbereich Fabrikplanung und -betrieb 2014, S. 3).

Ein isomorphes oder strukturgleiches Modell liegt vor, wenn jedes Element bzw. jede Beziehung dieser Elemente sowohl im realen System als auch im Modell vorliegen. Allerdings werden für die Planung die realen Systeme meist abstrahiert, indem Elemente zusammengefasst oder vernachlässigt werden. Die aus der Vereinfachung resultierenden Modelle werden auch als homomorphe bzw. strukturähnliche Modelle bezeichnet. Diese Art der Modellbildung hat auf der einen Seite den Vorteil, dass es im Gegensatz zum isomorphen Modell sehr viel einfacher ist, die planungsrelevanten Aspekte und Zusammenhänge zu begreifen, wenn die für die Planung irrelevanten Merkmale vernachlässigt werden und nur die entscheidenden Merkmale in das Modell übertragen werden. Auf der anderen Seite werden für den Fall der Vernachlässigung von wesentlichen Systemkomponenten lediglich ungünstige Planungsergebnisse erreicht. Aus diesem Grund ist eine Evaluation der Ergebnisse anhand des realen Systems oder eines Modells, das nicht so stark vereinfacht wurde, zwingend notwendig. (vgl. Scholl 2008, S. 36)

Für die Modellierung eines realen Netzwerks können die graphentheoretischen Elemente, je nachdem ob physische, logische, hierarchische oder zeitliche Relationen modelliert werden, völlig verschiedene Bedeutungen haben. Diese möglichen Bedeutungen der Knoten, Kanten und der optionalen Kantengewichte sind in **Tabelle 3.1** aufgeteilt nach den verschiedenen Relationstypen aufgelistet. (vgl. Mattfeld und Vahrenkamp 2014, S. 5)

Tabelle 3.1: Mögliche Bedeutungen der Objekte in der Netzwerkmodellierung

Relation	Knoten	Kanten	Gewichte
Physisch	Orte	Verkehrswege	Distanzen
Logisch	Partner	Güterflüsse	Flüsse
Hierarchisch	Objekte	Ordnung	Mengen

Zeitlich	Zustände	Übergänge	Zeit
----------	----------	-----------	------

Dieser Abschnitt ist an die Sichtweise von Tarjan angelehnt, nach der ein Modell auf ein Problem der realen Welt angepasst ist. Im Folgenden werden also zuerst grundlegende logistische Probleme genannt, bevor auf die zugehörigen Modelle eingegangen wird. Die in den Abschnitten 2.4.2 und 2.4.3 beschriebenen Aufgabenbereiche können auf unterschiedlichste Planungsprobleme reduziert werden. Dazu zählen etwa die Standortbestimmung, die Gestaltung von Lieferbeziehungen, die Auswahl von Routen und Touren oder die Zuordnung von Bedarfs- zu Angebotspunkten. (vgl. Steglich, Feige und Klaus 2016, S. 4)

Transportprobleme können sowohl als mittelfristige Optimierungsaufgaben von Transportwegen und -mitteln in vorhandenen Netzen als auch im Rahmen der langfristigen Gestaltung von Transportnetzen auftreten (vgl. Fleischmann und Kopfer 2018, S. 73). Sie beschäftigen sich im klassischen Fall mit dem Transport eines homogenen Gutes von m Versendern zu n Empfängern, wobei die Versender und Empfänger vollständig und direkt miteinander verbunden sind. Weiterhin sind die Transportkosten für jeden Transportweg gegeben, für die Transportwege gelten keine Beschränkungen bezüglich ihrer Kapazität und es steht nur ein Transportmitteltyp zur Verfügung. Zusammen mit diesen Gegebenheiten und der Vorgabe des individuellen Bedarfs des Empfängers und der Angebotsmengen des Versenders, soll ein kostenminimaler Transportplan bestimmt werden. (vgl. Steglich, Feige und Klaus 2016, S. 73) Weitere Probleme, die sich durch ein Transportmodell abbilden lassen, lauten:

- „Ermittlung optimaler Lieferbeziehungen zwischen Werken und Lagern sowie Distributionszentren und Kundengebieten“
- „Gestaltung der Distribution von Industrieunternehmen zur Minimierung von Transport- und Produktionskosten durch optimale Anpassung der Produktionskapazitäten und Güterflüsse“
- „Minimierung der Beschaffungskosten durch Optimierung der Lieferbeziehungen“
- Bestimmung der optimalen „Verteilung von Ressourcen an Engpasskapazitäten“ (Steglich, Feige und Klaus 2016, S. 5)

Das Transportmodell für den klassischen Fall ist ein bipartites Netzwerk, bei dem die Knoten entweder die Anbieter (S) oder die Nachfrager (T) repräsentieren. Es ist analog zu der Illustration von Mattfeld und Vahrenkamp (vgl. 2014, S. 148) in **Abbildung 3.4** dargestellt. Dabei ist jeder Anbieter mit jedem Nachfrager durch eine gerichtete Kante verbunden. Die zu transportierende Menge von Anbieter k zu Nachfrager i wird als x_{ki} bezeichnet und die für diesen Transport anfallenden Kosten sind c_{ki} . Sowohl die Transportmenge als auch die Transportkosten werden als positiv angenommen und es wird von identischen Angeboten und Bedarfen ausgegangen. (vgl. Mattfeld und Vahrenkamp 2014, S. 147 f.) Darüber hinaus werden Unter- und Obergrenzen der Liefermengen auf den Kanten vorgesehen (vgl. Steglich, Feige und Klaus 2016, S. 69). Ziel des Modells ist es, die Summe der einzelnen Transportkosten, die sich aus der Multiplikation der Transportmenge mit den Transportkosten pro Mengeneinheit zusammensetzen, zu minimieren. Um dieses Ziel zu erreichen, wird also eine Verteilung der Mengenströme x_{ki} gesucht, die die Summe der Transportkosten minimiert. Die Lösung teilt sich in zwei Stufen auf, bei der die erste eine Startlösung und die zweite ein Verfahren bestimmt, welches die Startlösung verbessert und das Optimum identifiziert. (vgl. Mattfeld und Vahrenkamp 2014, S. 148 f.)

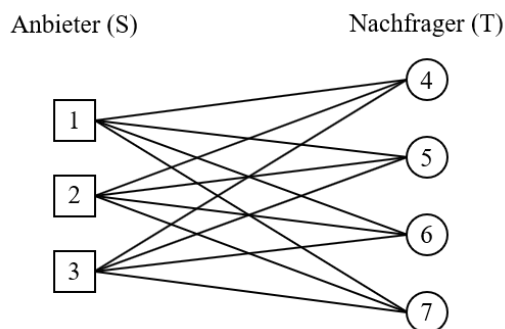


Abbildung 3.4: Ein einfaches Transportnetzwerk

In der Praxis tritt das Transportproblem in den verschiedensten Varianten auf, die sich hinsichtlich ihrer Angebots- und Bedarfsstrukturen, der Zielfunktion und den Lieferbeziehungen unterscheiden (vgl. Steglich, Feige und Klaus 2016, S. 70). Bei Transportproblemen mit ungleichen Angeboten und Bedarfen können die Fälle berücksichtigt werden, dass ein Überschuss der Angebote oder der Bedarfe vorliegt, Mindestangebote und -bedarfe existieren oder Unter- und Obergrenzen für die Angebote und/oder Bedarfe auftreten. In Transportproblemen mit nicht-klassischen Zielfunktionen werden Fälle beachtet, bei denen die Zielfunktion nicht die Summe der Transportkosten minimiert, sondern beispielsweise auf eine Maximierung des Gesamtdeckungsbeitrags aller Transporte oder die Belieferung der Kunden in kürzester Zeit abzielt. Eine weitere Variation des klassischen Transportproblems bezieht nicht-klassische Lieferbeziehungen ein. Bei diesen Problemen ist nicht zwingend jeder Anbieter mit jedem Nachfrager verbunden und es können Kapazitätsprobleme auftreten, wenn die Werte der optimalen Lieferbeziehungen aufgrund der Obergrenzen für die Liefermengen einer Kante nicht realisiert werden können. In diesem Fall nennt man das Transportproblem kapazitiert. Weiterhin kann das einstufige Transportnetzwerk aus Anbieter und Nachfrager durch das Hinzufügen von weiteren Schichten in Form von Umladeknoten zu einem mehrstufigen Netzwerk erweitert werden, wodurch ebenfalls neue Probleme entstehen. Ein weiteres Beispiel für nicht-klassische Lieferbeziehungen ist die Berücksichtigung von Gütern, die miteinander in Beziehung stehen, da sie etwa das gemeinsame Transportmittel benutzen. All diese Kriterien können auch miteinander kombiniert werden, sodass eine Vielzahl von Transportproblemen auftreten kann. (vgl. Steglich, Feige und Klaus 2016, S. 70 ff.)

Bei **logistischen Zuordnungsproblemen** findet eine Zuordnung von einer Anzahl Objekten zu Paaren in der Regel derart statt, dass die Zuordnungskosten minimiert werden (vgl. Mattfeld und Vahrenkamp 2014, S. 168). Zuordnungsprobleme treten neben der Planung von Standorten, von Transporten und des Verkehrs auch in der Einsatzplanung von Fahrzeugen und des Personals auf. So können beispielsweise verschiedene Standorte zu einzelnen Kundengebieten, Fahrzeuge zu separaten Aufträgen oder Mitarbeiter zu einzelnen Touren zugeordnet werden. (vgl. Steglich, Feige und Klaus 2016, S. 5) Diese Probleme können unabhängig von ihrer Ausprägung in der Regel durch das Matching in Graphen gelöst werden. Wenn sich die Menge der Zuordnungsobjekte, die als Knoten in einem Graphen repräsentiert werden, in zwei voneinander verschiedene Mengen, z.B. in Arbeiter und Aufträge, aufteilen lässt, entsteht ein bipartites Matching. (vgl. Steglich, Feige und Klaus 2016, S. 221) Jedoch ist diese Trennung nicht für alle Probleme möglich, da es logistische Aufgaben gibt, bei denen sich prinzipiell alle Zuordnungsobjekte untereinander kombinieren lassen. (vgl. Steglich, Feige und Klaus 2016, S. 261) Ein Beispiel für das sogenannte nicht-bipartite Matching ist die Linienzusammenlegung im Nahverkehr, um die mittleren Zeit- und Streckenkosten zu minimieren (vgl. Steglich, Feige und Klaus 2016, S. 264). Neben der Teilbarkeit der Menge der Zuordnungsobjekte können Zuordnungsprobleme auch in Bezug auf ihre

Graphbasierte Verfahren

Zielfunktion unterschieden werden. So soll beispielsweise beim Bottleneck-Zuordnungsproblem das Maximum der einzelnen Zuordnungskosten minimiert werden. (vgl. Steglich, Feige und Klaus 2016, S. 233)

Zur Erläuterung eines zugehörigen Modells wird im Folgenden auf ein simples Beispiel eines klassischen linearen Zuordnungsproblems eingegangen. Dabei handelt es sich um ein bipartites, gewichtetes Matching-Problem, bei dem die Zielfunktion zu minimieren ist. Die Menge der Zuordnungsobjekte teilt sich bei diesem Beispiel in eine gleich große Menge an Arbeitern und Aufträgen auf, die einander kostenminimal zuzuordnen sind. Allerdings ist gefordert, dass ein Arbeiter nur einem Auftrag zugeordnet und analog dazu ein Auftrag nur von einem Arbeiter angenommen werden kann. Zu minimieren ist bei diesem Problem die Summe der Zuordnungskosten, die sich aus der Multiplikation der Kosten für eine einzelne Zuordnung c_{ij} und der Zuordnungsvariable x_{ij} ergeben, die für den Fall einer bestehenden Zuordnung zwischen Arbeiter i und Auftrag j den Wert $x_{ij} = 1$ annimmt. Der Parameter c_{ij} ist die zu minimierende Kennzahl und könnte alternativ zu den Kosten auch für die Distanz oder das Gewicht stehen. (vgl. Steglich, Feige und Klaus 2016, S. 222 f.)

Eine der wichtigsten Aufgaben der Logistik ist das Sammeln und Verteilen von Gütern oder Personen, da dabei oft ein direkter Kundenkontakt entsteht. Die daraus resultierenden **Routen- und Tourenprobleme** treten in Netzwerken von Transportdienstleistern beispielsweise dann auf, wenn Ladungen auf Nahverkehrsfahrzeuge aufgeteilt werden oder zum Kunden zugestellt werden. Doch nicht nur die Auslieferung zählt zu den Problemen, sondern auch unter anderem die Übergabe von Kundensendungen an den Fernverkehr in den Depots. Die Planung von Routen und Touren bedient sich einer Vielzahl an Modellen, die den verschiedenen Problemstellungen des Kürzeste-Wege-Problems, des Rundreiseproblems, des Briefträgerproblems und des Tourenplanungsproblems zugeteilt werden. (vgl. Steglich, Feige und Klaus 2016, S. 5) Nach Domschke und Scholl (2010, S. 198) werden die Begriffe Tour und Route wie folgt definiert:

„Eine Tour beschreiben (definieren) wir durch die Angabe der Menge aller Kunden, Kanten bzw. Pfeile, die auf ein und derselben, in einem Depot beginnenden und in einem Depot endenden Fahrt „bedient“ werden. Die Reihenfolge, in der die Kunden einer Tour zu bedienen sind, bezeichnen wir als Route.“

Dabei ist ein Depot ein Ort, an dem die Fahrten zur Auslieferung, Sammlung, Ver- und Entsorgung beginnen und enden. Ein Tourenplan löst das jeweilige Tourenproblem und beinhaltet die Touren und zugehörigen Routen, die sämtliche Bedingungen erfüllen. (vgl. Domschke und Scholl 2010, S. 198 f.)

Zur effizienten Gestaltung von Transportsystemen ist es erforderlich, die kürzesten, schnellsten oder günstigsten Wege zu kennen. Die Bestimmung von kürzesten Wegen und Entfernungen kann sich entweder auf Wege zwischen einem bestimmten und allen anderen Knoten beziehen oder aber auf Wege zwischen allen Paaren verschiedener Knoten. Bei dieser Problemstellung wird von einem (gerichteten) Graphen ausgegangen, dessen Knoten etwa die Ortschaften eines realen Straßennetzes repräsentieren und dessen Kanten für die Straßenverbindungen stehen. (vgl. Domschke und Scholl 2005, S. 76) Jene Kanten können durch eine Gewichtung mit ihrer Länge oder anderen Qualitätsmerkmalen bewertet werden. Es existieren bereits zahlreiche spezialisierte Algorithmen für die verschiedenen Anwendungsfälle, wie der Dijkstra-Algorithmus, auf die in diesem Abschnitt jedoch nicht eingegangen wird. Das Problem kann ebenfalls durch eine Variante des Modells, welches für die Transportprobleme vorgestellt wurde, gelöst werden. (vgl. Steglich, Feige und Klaus 2016, S. 271 f.)

Graphbasierte Verfahren

Rundreisenprobleme oder auch Travelling-Salesman-Probleme beschäftigen sich mit der Identifikation der kürzesten geschlossenen Reiseroute über eine Anzahl von Orten, wobei die Entfernungen zwischen den Orten bekannt sind. Dabei muss jeder Ort im besten Fall exakt einmal besucht werden und das Ende der Reise hat am Ursprungsort zu liegen (geschlossene Tour). (vgl. Steglich, Feige und Klaus 2016, S. 278) Aus diesem Grund werden diese Reisen auch als knotenorientierte Rundreisen bezeichnet (vgl. Mattfeld und Vahrenkamp 2014, S. 229). Allerdings setzen praktische Probleme oft nur voraus, dass der geschlossene Weg ausgewählte Knoten des Graphen mindestens einmal enthält und nicht alle Knoten genau einmal enthalten sind (vgl. Domschke und Scholl 2010, S. 96). In weiteren Varianten des Rundreiseproblems wird keine geschlossene Tour, sondern eine offene Rundreise gesucht, die unterschiedliche Start- und Endpunkte hat. Analog zu den Transportproblemen treten auch kapazitierte oder zeitabhängige Probleme auf. (vgl. Steglich, Feige und Klaus 2016, S. 280) Bei der Modellierung von Rundreiseproblemen müssen nur schlichte Graphen betrachtet werden, weil Schlingen irrelevant für das Problem sind und auf Mehrfachkanten aus dem alleinigen Interesse an der am niedrigsten gewichteten Kante verzichtet werden kann. (vgl. Domschke und Scholl 2010, S. 96) Weiterhin kann je nach Abhängigkeit der Kantenrichtungen zwischen symmetrischen und asymmetrischen Problemen differenziert werden. Bei einem symmetrischen Rundreiseproblem sind die Kantengewichte zwischen allen Knotenpaaren richtungsunabhängig (ungerichteter Graph) und bei asymmetrischen Problemen sind sie richtungsabhängig (gerichteter Graph). In einem Modell stehen die Knoten erneut für die einzubeziehenden Orte und die gewichteten Kanten für die Verbindungen dieser Knoten. (vgl. Steglich, Feige und Klaus 2016, S. 280 f.)

Im Gegensatz zu den Rundreiseproblemen sollen bei den Briefträgerproblemen nicht alle Knoten, sondern ausgehend von einem Startknoten jede gerichtete oder ungerichtete Kante mindestens einmal durchlaufen und anschließend zum Ursprungsknoten zurückgekehrt werden. Diese kantenorientierte Rundreise soll die gesamte Entfernung minimieren. (vgl. Steglich, Feige und Klaus 2016, S. 321) Entsprechende Probleme treten beispielsweise bei der Müllabfuhr oder der Post auf, die beide darauf bedacht sind, die einzelnen Straßen eines Bezirks in einer möglichst optimalen Reihenfolge abzufahren (vgl. Mattfeld und Vahrenkamp 2014, S. 221). Der Idealfall, dass jede Kante eines Graphen genau einmal durchlaufen wird, wird in der Mathematik auch als Euler-Tour bezeichnet. Enthält ein Netzwerk eine Euler-Tour, handelt es sich um ein Euler-Netzwerk. Dafür muss der Knotengrad des Netzwerks im ungerichteten Fall gerade und mindestens zwei sein und für ein gerichtetes Netzwerk müssen die Anzahl der Kanten, die in einen Knoten hineingehen und aus diesem Knoten herausgehen gleich sein. (vgl. Steglich, Feige und Klaus 2016, S. 321) Kann für ein Briefträgerproblem keine Euler-Tour bestimmt werden, muss ein Arbeitsschritt durchlaufen werden, der das Netzwerk durch die kosten- oder distanzminimale Erweiterung zu einem Euler-Netzwerk erweitert (vgl. Steglich, Feige und Klaus 2016, S. 324).

Bei den zuvor thematisierten Rundreiseproblemen fanden die Kapazitäten des eingesetzten Transportmittels keine Betrachtung, sondern es wurde nur die kürzeste Route identifiziert. Werden nun die Bedarfe der einzelnen Knoten miteinbezogen, kann nicht davon ausgegangen werden, dass die Fahrzeugkapazität immer größer oder gleich dem Gesamtbedarf aller Knoten ist. Falls die Ladekapazität zu klein für die zu liefernde oder abzuholende Menge ist, muss das Fahrzeug zwischenzeitlich zum Depot zurückkehren und kann die Knoten deshalb nur auf mehreren Routen beliefern. (vgl. Steglich, Feige und Klaus 2016, S. 332) Angelehnt an das Beispiel von Steglich, Feige und Klaus (vgl. 2016, S. 332) ist in **Abbildung 3.5** eine Rundreise und der aus ihr resultierende Tourenplan zu erkennen, der sich ergibt, wenn die Kapazitäten des Fahrzeuges nicht für alle Güter reichen.

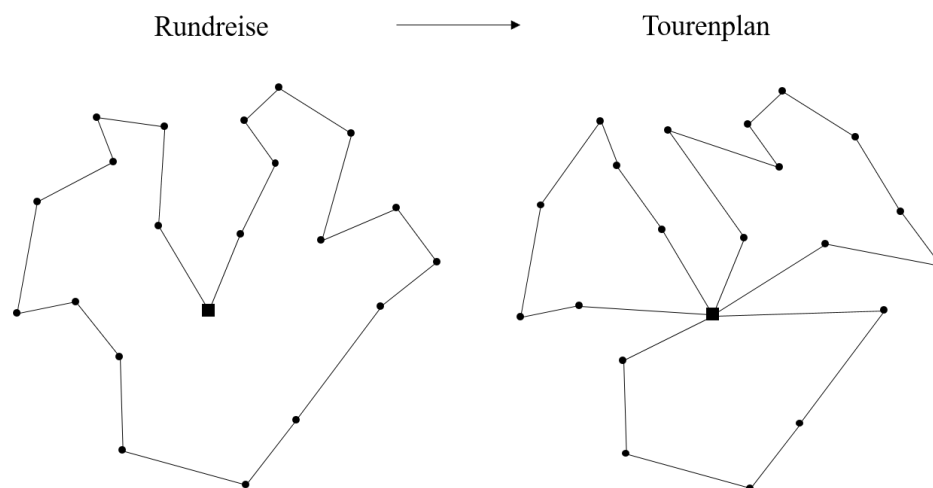


Abbildung 3.5: Rundreise und Tourenplan

Die Knoten des modellierten Graphen sind entweder das Depot oder die Bedarfsknoten, deren einzelne Bedarfswerte bekannt sind. Der Bedarf eines Knotens kann höchstens so groß sein, wie die Kapazität eines Fahrzeugs. Neben den einzelnen Bedarfswerten sind auch die Kantengewichte in Form von Entfernungen oder Kosten bekannt. Aus diesem Modell ergibt sich die Aufgabe der Tourenplanung, die Bedarfe der Bedarfsknoten unter Berücksichtigung der Fahrzeugkapazitäten auf verschiedene Routen zu verteilen. Der Start und das Ende der Routen müssen in einem Depot liegen. Die Aufgabe zielt auf eine Minimierung der Gesamtkosten oder der Gesamtstrecke aller Touren ab. (vgl. Steglich, Feige und Klaus 2016, S. 332)

Die **Planung von Standorten für Logistikknöten** wurde bereits in Abschnitt 2.4.2 beschrieben. Diese allgemeine Problemstellung setzt sich aus mehreren Einzelaufgaben zusammen, die miteinander verknüpft sind und oft gemeinsam gelöst werden müssen. Bei den Einzelaufgaben werden die optimale Knotenanzahl, die optimalen Knotenstandorte und die Zuordnung der Bedarfsorte zu den Knoten ermittelt. (vgl. Steglich, Feige und Klaus 2016, S. 6) Die Probleme der Standortplanung können in zwei unterschiedliche Modelle unterschieden werden (vgl. Steglich, Feige und Klaus 2016, S. 375). Bei den diskreten Modellen bzw. Netzwerkmodellen wird das Entscheidungsproblem als Graph modelliert. (vgl. Mattfeld und Vahrenkamp 2014, S. 101) Die zu beliefernden Bedarfsorte, deren Bedarfe bekannt sind, und die potenziellen Standorte, die auf alle oder einen Teil der Knoten beschränkt sind, werden im Graphen durch die Knoten repräsentiert und der Transport findet auf den Kanten zwischen den Knoten statt. Unter Berücksichtigung dieser Annahmen können insbesondere Median- und Zentren-Probleme in Netzwerken, Überdeckungsprobleme und Warehouse-Location-Probleme formuliert werden. (vgl. Steglich, Feige und Klaus 2016, S. 375) Bei den kontinuierlichen Modellen sind die Bedarfsorte auf der euklidischen Ebene verteilt. Da sie nicht als Graph modelliert werden, werden diese Modelle nicht weiter erörtert. (vgl. Mattfeld und Vahrenkamp 2014, S. 101)

Für die Lösung von Median- und Zentren-Probleme in Netzwerken wird ein Graph verwendet, dessen Knoten entweder Angebots- oder Bedarfsknoten sind. Dabei ist die Nachfrage eines jeden Bedarfsknoten nach einem homogenen Gut bekannt und jede Kombination zweier Knoten besitzt eine Bewertung, die für gewöhnlich für die Distanz steht. Die Aufgabe besteht darin, den optimalen Standort für eine Anzahl von Angebotsknoten zu finden, wobei dafür jeder Knoten des Netzwerkes infrage kommt. Er kann also auch mit einem Bedarfsknoten übereinstimmen. Die Median- und Zentren-Probleme werden zur weiteren Abgrenzung hinsichtlich ihrer Zielkriterien unterschieden. (vgl. Steglich, Feige und Klaus

Graphbasierte Verfahren

2016, S. 375 f.) Bei Median-Problemen wird die Ansiedlung eines Zentral-Lagers, das mehrere Outlets in einer Region versorgen soll, in einem Modell abgebildet und das Ziel der Minimierung der gesamten gewichteten Entfernungen des Zentral-Lagers zu den Outlets verfolgt. (vgl. Mattfeld und Vahrenkamp 2014, S. 122) Bei Zentrenproblem wird zuerst zu jedem Knoten eines Netzwerks die maximale Distanz zu den restlichen Knoten berechnet, bevor ein Zentrum identifiziert wird, welches die maximal auftretenden gewichteten Entfernungen minimiert. Nach dieser Problemstellung sind beispielsweise die Standorte von Krankenhäusern oder Feuerwachen bestimmt worden, damit sie von jedem Punkt des Graphen schnellstmöglich zu erreichen sind. (vgl. Mattfeld und Vahrenkamp 2014, S. 103)

Problemfälle, bei denen die Einhaltung eines vorgegebenen Service-Niveaus wichtiger ist als die Erfüllung einer Zielgröße, können als Überdeckungsprobleme bezeichnet werden. Dabei sollen in einem Graphen, dessen Knoten die Kunden darstellen, Servicezentren oder Outlets derart aufgebaut werden, so dass alle Kunden des Netzwerkes mit möglichst wenigen Standorten abgedeckt werden. Eine Bedingung, die bei den Überdeckungsproblemen eingehalten werden muss, besagt, dass jeder Kunde innerhalb einer maximalen Distanz bzw. Fahrzeit den nächsten Standort erreichen muss. Es wird zwischen zwei Arten von Überdeckungsproblemen unterschieden. (vgl. Mattfeld und Vahrenkamp 2014, S. 108) Im Rahmen des Set-Covering-Location-Problems soll ausgehend von einem vorgegebenen Maximalabstand zwischen den Bedarfsknoten und Standorten, die minimale Anzahl an Standorten identifiziert werden, die zu keiner Überschreitung des Maximalabstands führt und dennoch alle Bedarfsknoten abdecken. Bei einer Vorgabe der Anzahl der Standorte sind diese so zu positionieren, dass möglichst viele Bedarfsorte innerhalb der Gebiete liegen, die durch den Maximalabstand der einzelnen Standorte entstehen. (vgl. Steglich, Feige und Klaus 2016, S. 416)

Sollen in die Standortentscheidung nicht nur die fixen Standortkosten, sondern auch die variablen Transportkosten mit einbezogen werden, handelt es sich um ein typisches Warehouse-Location-Problem (vgl. Domschke, Drexl und Mayer, et al. 2018, S. 4). Um diese Problemstellung zu lösen sind neben den potenziellen Standorten, deren Fixkosten, den Bedarfen für sämtliche Bedarfsorte auch die variablen Kosten für den Transport zwischen den potenziellen Standorten und den Bedarfsorten gegeben. Ziel ist die Identifizierung der Standorte aus der Menge der potenziellen Standorte unter Berücksichtigung der Nachfragebefriedigung der Bedarfsorte und der minimalen Summe der fixen Standortkosten und variablen Transportkosten. Je nach Ausprägung der folgenden Merkmale bilden sich zahlreiche Aufgabenstellungen, die sich in der Komplexität unterscheiden (vgl. Steglich, Feige und Klaus 2016, S. 429):

- „Anzahl der Stufen (ein- oder mehrstufige Probleme),
- Kapazitätsbeschränkungen (unkapazitierte oder kapazitierte Probleme),
- Anzahl einzubeziehender Güter (Ein- oder Mehrgüterprobleme),
- Art der Kostenfunktionen (lineare oder nichtlineare Probleme).“ (Steglich, Feige und Klaus 2016, S. 429)

3.3 Graphalgorithmik

Der inhaltliche Fokus dieses Abschnitts liegt auf der Beschreibung der unterschiedlichen Graphalgorithmen zur Durchführung einer Aufgabe, die nach Aggarwal und Wang (vgl. 2010a, S. 2) in Algorithmen des Graph Management (vgl. Abschnitt 3.3.1) und des Graph Mining (vgl. Abschnitt 3.3.2) unterteilt werden. Sie merken jedoch auch an, dass die Algorithmen nicht immer klar voneinander abzugrenzen sind, da einige Algorithmen zu beiden Bereichen zählen (vgl. Aggarwal und Wang 2010b, S. 15). Neben der Beschreibung der Algorithmen wird in Abschnitt 3.3.2 eine kurze Einführung in das Data Mining gegeben. Im Folgenden werden die englischen Begriffe der Graphalgorithmen übernommen.

3.3.1 Graph Management

Die Datenverwaltung ist für Graphen wesentlich anspruchsvoller als für multidimensionale Daten. Auf der einen Seite besitzt die strukturelle Repräsentation von Graphen zwar eine größere Ausdruckskraft als die von mehrdimensionalen Daten, ist jedoch in Hinblick auf die Datendarstellung, den Zugriff und die Verarbeitung komplizierter, da Operationen wie die Mittelwertbildung oder Abstandsberechnung nicht so intuitiv definiert werden können (vgl. Aggarwal und Wang 2010b, S. 16). Im Bereich des Graph Management werden leistungsfähige Technologien zum Speichern, Verarbeiten und Analysieren einer Vielzahl von Graphdaten erforscht und entwickelt. (vgl. Angles und Gutierrez 2018, S. 13) klassifiziert die in der Graphentheorie umfangreich untersuchten Abfragen in vier Hauptklassen, die im Folgenden beschrieben werden.

Die erste Kategorie der Abfragen stellen die **Adjacency Queries** dar, bei denen die Nachbarschaft (Adjazenz) von Knoten und Kanten abgefragt wird. Eine Nachbarschaft zweier Knoten liegt vor, wenn sie durch eine gemeinsame Kante verbunden sind. Zwei Kanten sind dementsprechend benachbart, wenn sie sich einen gemeinsamen Knoten teilen. Einfache Abfragen, überprüfen ebendiese Nachbarschaft oder geben alle Nachbarn eines Knoten zurück. (vgl. Angles und Gutierrez 2018, S. 14) Auch wenn diese Abfragen simpel erscheinen, sind sie bei großen mageren Graphen aufgrund des Speicherbedarfs im Vergleich zur tatsächlichen Größe sehr anspruchsvoll. (vgl. Kowalik 2007, S. 1) Komplexere Anfragen geben etwa die Menge aller Knoten aus, die von einem zuvor festgelegten Knoten über einen Pfad mit einer bestimmten Anzahl an Kanten, die ebenfalls vorher festgelegt wurde, erreichbar sind. Zum Einsatz kommen Adjacency Queries überall dort, wo der Einfluss eine wichtige Rolle spielt. Beispielsweise können in einem Empfehlungssystems eines Onlineshops die Nutzer ermittelt werden, die ähnliche Interessen haben. (vgl. Angles und Gutierrez 2018, S. 14)

Ein weiterer Typ von Abfragen sind die **Pattern Matching Queries**. Dabei werden all jene Teilgraphen eines Graphen gesucht, die mit einem gegebenen Muster beziehungsweise Pattern übereinstimmen. Ein Muster ist in diesem Fall ein kleiner Graph, bei dem manche Knoten und Kanten mit Variablen beschriftet sind, um die Ausgabe der Abfrage zu definieren. Generell kann sich die Überprüfung der strukturellen Ähnlichkeit zum gegebenen Muster entweder auf Teilgraphen oder komplette Graphen beziehen. (vgl. Angles und Gutierrez 2018, S. 14) Folgende strukturelle Eigenschaften werden nach (vgl. Aggarwal und Wang 2010b, S. 21 f.) beim Vergleich von zwei Graphen betrachtet:

- Attribute der Knoten
- Existenz von Kanten zwischen den entsprechenden Knoten
- Attribute der Kanten

Ein allgemeiner Begriff für den Prozess der Überprüfung der strukturellen Gleichheit von Graphen lautet „Graph Matching“. Bei der Überprüfung unterscheiden sich die Methoden dahingehend, dass sie die Struktur entweder auf exakte oder inexakte Weise abgleichen, also einige Fehler tolerieren. (vgl. Bunke und Neuhaus 2007, S. 17) Das exakte Graph Matching versucht zu bestimmen, ob die Struktur und die Attribute zweier Graphen identisch ist oder nicht. Das heißt, dass eine Kante zwischen einem Knotenpaar zwingend in beiden Graphen existieren muss, wenn die gleiche Kante in einem der beiden Graphen existiert. (vgl. Aggarwal und Wang 2010b, S. 22) Dieses Problem ist rechnerisch sehr anspruchsvoll, da die Elemente eines Graphen, im Gegensatz etwa zu den Symbolen einer Zeichenkette, nicht geordnet werden können (vgl. Bunke und Neuhaus 2007, S. 19). In realen Anwendungen ist eine exakte Prüfung der Übereinstimmung zweier Graphen nicht praktisch, da eher ungefähre statt exakter Übereinstimmungen existieren. (vgl. Aggarwal und Wang 2010b, S. 22) Um die natürlichen Fehler zu berücksichtigen,

Graphbasierte Verfahren

die beim Matching-Prozess auftreten können, wird beim inexakten Graph Matching eine Funktion wie die Graph Edit Distance verwendet. Durch die Messung der Kosten für die Transformation des einen in den anderen Graphen, wird der Abstand beider Graphen bestimmt. Je mehr Knoten oder Kanten für die Transformation eingefügt, gelöscht oder ersetzt werden müssen, desto größer ist der Abstand. (vgl. Aggarwal und Wang 2010b, S. 23) Das Graph Matching besitzt zahlreiche Anwendungsfälle, beispielsweise zur Identifikation von Gemeinschaften oder sozialen Positionen in sozialen Netzwerken oder zur Identifikation von Mustern in biologischen Netzwerken, die Proteine mit ähnlichen Funktionen bestimmen. (vgl. Angles und Gutierrez 2018, S. 14).

Angles und Gutierrez (vgl. 2018, S. 14) nennen die Berechnung der Erreichbarkeit (Reachability) von Informationen als eines der charakteristischsten Probleme in Graphdatenbanken. **Reachability Queries** prüfen, ob es in einem großen gerichteten Graphen einen Pfad zwischen zwei bestimmten Knoten gibt (vgl. Yu und Cheng 2010, S. 182). Sie werden im Zusammenhang mit Graphdatenbanken in der Regel als **Pfad- oder Traversierungsprobleme** modelliert, wobei Pfadprobleme so komplex sind, dass ihre Berechnungen oft eine Suche über einen großen Datenraum erfordert (vgl. Angles und Gutierrez 2018, S. 15 f.). Zur Beantwortung der Frage, ob ein Pfad zwischen zwei Knoten existiert, gibt es zwei offensichtliche Methoden. Bei der ersten wird der Graph ausgehend vom ersten Knoten durchlaufen, um zu sehen, ob der zweite Knoten überhaupt erreicht werden kann. Diese Methode kann je nach Größe des Graphen sehr zeitintensiv sein. Das zweite Verfahren berechnet die transitive Hülle des Graphen und speichert diese ab. (vgl. Aggarwal und Wang 2010b, S. 20) Die transitive Hülle enthält sämtliche Knoten, die durch einen Pfad verbunden werden können (Even und Even 2012, S. 26). Zur Beantwortung der Abfrage wird lediglich in der transitiven Hülle geprüft, ob ein Pfad zwischen den Knoten existiert. Allerdings benötigt dieses Verfahren für große Graphen viel Speicherplatz. Damit ist weder die erste noch die zweite Methode für große Graphen geeignet, weshalb sich die Forschung in diesem Bereich darauf konzentriert, den besten Kompromiss zwischen Abfragezeit und benötigtem Speicherplatz zu finden. (vgl. Aggarwal und Wang 2010b, S. 20) Reachability Queries finden in sehr vielen Bereichen Anwendung. Beispielsweise soll der kürzeste Pfad zwischen zwei Knoten ermittelt werden oder es wird für den Fall eines gewichteten Graphen abgefragt, ob zwei Knoten durch einen Pfad verbunden sind, dessen Kanten durch Kantengewichte oder Attribute eingeschränkt sind. (vgl. Aggarwal und Wang 2010b, S. 21) Konkrete Anwendungsfälle sind das Auffinden von spezifischen biochemischen Pfaden zwischen entfernten Knoten in biologischen Netzwerken oder die Berechnung der minimalen Entfernung zweier Orte in einem Straßennetz (vgl. Angles und Gutierrez 2018, S. 16).

Die vierte und letzte Klasse von Abfragen nehmen nach Angles und Gutierrez (vgl. 2018, S. 16) die **Analytical Queries** ein. Sie berücksichtigen nicht die Struktur des Graphen, sondern sind darauf ausgerichtet, die topologischen Eigenschaften des Graphen quantitativ und meist in konzentrierter Form zu messen. Eine Besonderheit der Analytical Queries ist es, dass sie durch Funktionen, die komplexe Algorithmen umfassen, oder spezielle Operatoren, die Abfrageergebnisse zusammenfassen, unterstützt werden können. Ein solcher Operator kann beispielsweise den Knotengrad, den Durchschnitt, die Anzahl oder ein Maximum bestimmen und wird meist bei einfachen analytischen Abfragen verwendet. Darüber hinaus stehen komplexe analytische Abfragen in Verbindung mit wesentlichen Algorithmen der Graphenanalyse sowie des Graph Mining. Beispiele für diese Algorithmen sind die Bestimmung der durchschnittlich kürzesten Pfadlänge in einem Netzwerk (Charakteristische Pfadlänge), die Berechnung der Wahrscheinlichkeit, dass die Nachbarn eines Knoten ebenfalls miteinander verbunden sind (Clustering-Koeffizient), oder der PageRank-Algorithmus, der zur Messung der relativen Wichtigkeit von

Graphbasierte Verfahren

Knoten in einem Graphen genutzt wird. (vgl. Angles und Gutierrez 2018, S. 16) Der folgende Abschnitt 3.3.2 widmet sich den unterschiedlichen Algorithmen des Graph Mining.

3.3.2 Graph Mining

Bevor das Graph Mining erläutert wird, wird auf das übergeordnete Fachgebiet des Data Mining eingegangen. Data Mining ist wiederum der Knowledge Discovery in Databases (KDD) zuzuordnen. Im KDD-Prozess werden Informationen aus Daten nicht-trivial extrahiert, die nicht implizit, bisher unbekannt und potenziell nützlich sind. (vgl. Bramer 2007, S. 2 f.) Mit einer nicht-trivialen Extraktion ist nicht die einfache Berechnung von vordefinierten Größen, wie die Ermittlung des Durchschnittswertes, gemeint, sondern die Tatsache, dass bei diesem Schritt eine Suche oder Schlussfolgerung durchgeführt wird (vgl. Fayyad, Piatetsky-Shapiro und Smyth 1996, S. 41). Der Prozess der Knowledge Discovery nach Fayyad, Piatetsky-Shapiro und Smyth (vgl. 1996, S. 41) ist in **Abbildung 3.6** verdeutlicht. Dabei werden die verfügbaren Daten während dem ersten Schritt der Selektion identifiziert und auf einem gemeinsamen Speicher abgelegt. In einem zweiten Schritt der Vorverarbeitung werden fehlende Werte behandelt, indem sie ersetzt werden, und problembehaftete Daten korrigiert. Da die eigentlichen Verfahren zur Extraktion von Mustern in der Regel ein bestimmtes Datenformat erfordern, werden die Daten bei der Transformation in adäquate Datenformate umgewandelt. (vgl. Cleve und Lämmel 2020, S. 8 f.) Im zentralen Schritt des Data Mining erzeugt ein Algorithmus aus den aufbereiteten Daten Muster, beispielsweise in Form von Regeln. Diese Muster werden im letzten Prozessschritt durch die Interpretation und Evaluation in neues, potenziell brauchbares Wissen umgewandelt. (vgl. Bramer 2007, S. 2 f.)

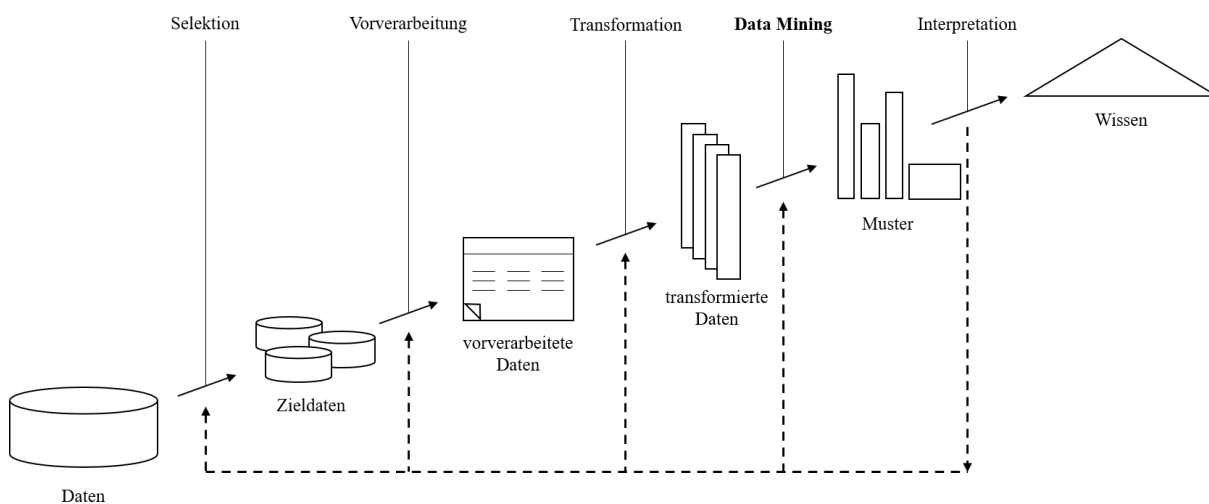


Abbildung 3.6: Knowledge Discovery in Databases-Prozess

In dieser Arbeit wird die Sicht von Fayyad, Piatetsky-Shapiro und Smyth (vgl. 1996, S. 39) vertreten, nach der das Data Mining spezifische Algorithmen anwendet, um Muster aus Daten zu extrahieren. Data Mining ist ein einzelner Schritt im KDD-Prozess. Dieser Gesamtprozess zur Entdeckung von nützlichem Wissen beinhaltet weitere Schritte, wie die Datenaufbereitung, die Datenbereinigung oder die richtige Ergebnisinterpretation des Mining mithilfe von geeignetem Vorwissen. Werden Methoden des Data Mining ohne diese zusätzlichen Schritte des KDD ausgeführt, ist die Wahrscheinlichkeit hoch, dass bedeutungslose und/oder ungültige Muster entdeckt werden. (vgl. Fayyad, Piatetsky-Shapiro und Smyth 1996, S. 39)

Hand, Mannila und Smyth (vgl. 2001, S. 15) gliedern jeden Algorithmus des Data Mining grundsätzlich in vier Komponenten. Bei der ersten Komponente der Modell- oder Musterstruktur wird die zugrunde

Graphbasierte Verfahren

liegende Struktur oder die funktionale Form bestimmt, die in den Daten gesucht wird. (vgl. Hand, Mannila und Smyth 2001, S. 15) Diese Repräsentationen von Daten lassen sich auf zwei verschiedene Weisen charakterisieren. Auf der einen Seite stehen globale Modelle, die den Datensatz zusammenfassen und Aussagen über jeden Punkt des Raums machen. Im Gegensatz dazu stehen die lokalen Muster, die eine Struktur beschreiben, welche sich nicht auf den kompletten Raum bezieht, sondern auf einen relativ kleinen Teil der Daten. (vgl. Hand, Mannila und Smyth 2001, S. 9 f.) Die Score-Funktion stellt die zweite Komponente dar und beurteilt die Qualität eines angepassten Modells, also wie gut der gegebene Datensatz und das Modell bzw. die Parameterstruktur zusammenpassen. Dabei kann es sein, dass die Parameterwerte noch unbekannt sind. (vgl. Hand, Mannila und Smyth 2001, S. 15 f.) Die dritte Komponente besteht aus der Optimierung der Score-Funktion und einer Suchmethode. Die Optimierung zielt darauf ab, jene Struktur bzw. Parameterwerte zu bestimmen, die einen abhängig vom Kontext optimalen Wert der Score-Funktion erreichen. Mithilfe einer Suchmethode wird die Aufgabe gelöst, interessante Muster aus einer großen Familie potenzieller Muster zu finden. (vgl. Hand, Mannila und Smyth 2001, S. 16 f.) Die Datenmanagement-Strategie legt fest, wie der Datenzugriff während der Suche/Optimierung effizient gehandhabt wird und ist die letzte Komponente eines Algorithmus. Neben der Datenspeicherung bestimmt diese Strategie beispielsweise auch die Art und Weise der Indizierung oder des Zugriffs. (vgl. Hand, Mannila und Smyth 2001, S. 17 f.)

Muster, die aus Daten extrahiert werden, sollten nach Fayyad, Piatetsky-Shapiro und Smyth (vgl. 1996, S. 41) einerseits mit gewisser Sicherheit auch für neue Daten gelten und andererseits neuartig, verständlich sowie potenziell nützlich für den Benutzer oder die Aufgabe sein. Der Wert eines Musters wird mithilfe des Begriffes der Interessantheit gemessen. Mit dem Begriff der Interessantheit wird der Wert eines Musters für einen Benutzer beschrieben. Dabei werden Faktoren wie die Neuheit, der Nutzen, die Relevanz oder die statistische Signifikanz berücksichtigt. (vgl. Frawley, Piatetsky-Shapiro und Matheus 1992, S. 59), (vgl. Piatetsky-Shapiro und Matheus 1994, S. 25)) Sollte ein Muster eine gewisse Interessantheitsschwelle übersteigen, kann es nach Fayyad, Piatetsky-Shapiro und Smyth (vgl. 1996, S. 41) als domänenspezifisches und benutzerorientiertes Wissen betrachtet werden.

Angewendet werden können die Methoden des Data Mining auf unterschiedlichste Bereiche. Sei es zur Identifikation von Kaufmustern oder Warenkorbanalyse im Marketing, zur Ermittlung der Ausgaben von verschiedenen Kundengruppen im Kreditwesen, zur Identifikation von betrügerischem Verhalten im Versicherungswesen oder auch zur Analyse von Belademustern im Transportwesen (vgl. Prabhu und Venkatesan 2007, S. 4). Die zur Verfügung stehenden Daten für den Data Mining Prozess können grob in zwei Gattungen unterteilt werden. Die erste Art machen Daten aus, die nicht abhängigkeitsorientiert sind. Dazu zählen in der Regel Datentypen wie Textdaten oder multidimensionale Daten, die einfach sind und am häufigsten vorkommen. Untereinander haben die Datenelemente oder Attribute keine festgelegte Abhängigkeit. Anders sieht das bei der zweiten Gattung, den abhängigkeitsorientierten Daten, aus. Hier können die Datenelemente implizit oder explizit in Beziehung zueinanderstehen. Beispiele dafür sind soziale Netzwerke, bei denen eine Vielzahl von Punkten durch eine Reihe von Kanten verbunden sind, oder etwa Zeitreihen, bei denen zwei aufeinanderfolgende Werte für gewöhnlich in Beziehung stehen. Die vorhandenen Abhängigkeiten resultieren in einer höheren Komplexität der Daten und müssen direkt in den Analyseprozess miteinbezogen werden, um damit kontextuell sinnvolle Ergebnisse zu erhalten. (vgl. Aggarwal 2015, S. 6)

Daten, die in Graphen abgespeichert werden, unterscheiden sich besonders in Hinblick auf ihre Darstellung und wie sie in Beziehung zueinander gesetzt werden von Daten, die in Tabellen gespeichert sind.

Graphbasierte Verfahren

(vgl. Nettleton 2013, S. 6) Viele Anwendungen des Data Mining gelten auch für den Fall von Graphen, jedoch sind die Verfahren aufgrund der Struktur des Graphen weitaus schwieriger zu implementieren. Aus diesem Grund wurden für neue Verfahren, wie das Frequent Pattern Mining, Clustering oder die Classification, für bekannte Mining-Probleme entwickelt. Diese Algorithmen sind Bestandteil des Graph Mining. (vgl. Aggarwal und Wang 2010b, S. 29) Somit verfolgen das Data Mining und Graph Mining denselben Zweck der Extraktion von Informationen oder Wissen in Form von Mustern aus zahlreichen Daten mit dem Unterschied, dass die Algorithmen des Graph Mining auf Graphen angewendet werden (vgl. Atastina, et al. 2017, S. 3). Nach (vgl. Holder und Cook 2007, S. 2) liegt das aus Graphen extrahierte Wissen ebenfalls als Graph vor und besitzt analog zum Data Mining die Eigenschaften der Neuartigkeit und der Nützlichkeit. Dabei kann der durch einen Algorithmus gewonnene Graph entweder aus mehreren kleinen Teilgraphen, sogenannten Transaktionsgraphen, oder einem einzelnen großen Graphen bestehen. (vgl. Atastina, et al. 2017, S. 3) Chakrabarti (vgl. 2017, S. 582) definiert für die Techniken des Graph Mining drei unterschiedliche Einsatzgebiete:

- Die Analyse der Eigenschaften eines realen Graphen
- Die Vorhersage der Auswirkung von der Struktur und den Eigenschaften eines gegebenen Graphen auf eine Anwendung
- Die Entwicklung von Modellen, die realistische Graphen erzeugen können, die mit den Mustern von realen Graphen übereinstimmen (vgl. Chakrabarti 2017, S. 582)

Die Algorithmen des Graph Mining unterscheiden sich nach (vgl. Aggarwal und Wang 2010b, S. 15) dahingehend von den Algorithmen des Graph Management, dass sie Muster, Trends, Klassen und Cluster aus Graphen extrahieren und keine große Menge von Graphdaten verwalten und indizieren. Anwendung findet das Graph Mining unter anderem in Informationsnetzwerken, biologischen oder auch sozialen Netzwerken. Darüber hinaus werden die Algorithmen etwa in der Bio- oder Chemieinformatik angewandt. (vgl. Han, Kamber und Pei 2012, S. 591) Im Folgenden wird ein ausgewählter Überblick über einige strukturelle Algorithmen des Graph-Mining gegeben.

Eines der Hauptverfahren ist das des **Graph Pattern Mining** (vgl. Atastina, et al. 2017, S. 2). Es behandelt das Problem der Aufzählung aller Teilgraphen bzw. Subgraphen, die in den gegebenen Graphen häufig vorkommen und wird auch als Frequent Subgraph Mining bezeichnet (vgl. Takigawa und Mamitsuka 2013, S. 53). Genauer gesagt ist das Frequent Subgraph Mining der Prozess, bei dem jene Subgraphen aus einer Menge von Graphen oder aus einem einzelnen großen Graphen identifiziert werden, die häufiger vorkommen als ein zuvor festgelegter Schwellenwert (vgl. Ramraj und Prabhkar 2015, S. 197 f.). Angles und Gutierrez (vgl. 2018, S. 15) merken an, dass das Graph Pattern Matching dem Graph Pattern Mining sehr verwandt ist, da es sich ebenfalls um häufige Muster handelt. Jedoch werden beim Graph Pattern Matching die Teilgraphen gesucht, die zu einem vorgegebenen Muster passen (matching) und beim Graph Pattern Mining sind diese Muster zuerst unbekannt und werden während des Verfahrens entdeckt (mining) (vgl. Angles und Gutierrez 2018, S. 14 f.). In **Abbildung 3.7** wird ein einfaches Beispiel dieses Prozesses dargestellt, das von Takigawa und Mamitsuka (vgl. 2013, S. 53) beschrieben wurde. Auf der linken Seite sind fünf Molekulargraphen gegeben und auf der rechten Seite befinden sich einige Subgraphen dieser fünf Graphen. Der „Support“ eines Subgraphen wird definiert als die Anzahl der gegebenen Graphen, die diesen Subgraphen enthalten. Beträgt der Schwellenwert, der auch als „minimaler Support“ bezeichnet wird, nun $5/5$, erfüllen nur drei Subgraphen diesen Parameter. (vgl. Takigawa und Mamitsuka 2013, S. 53)

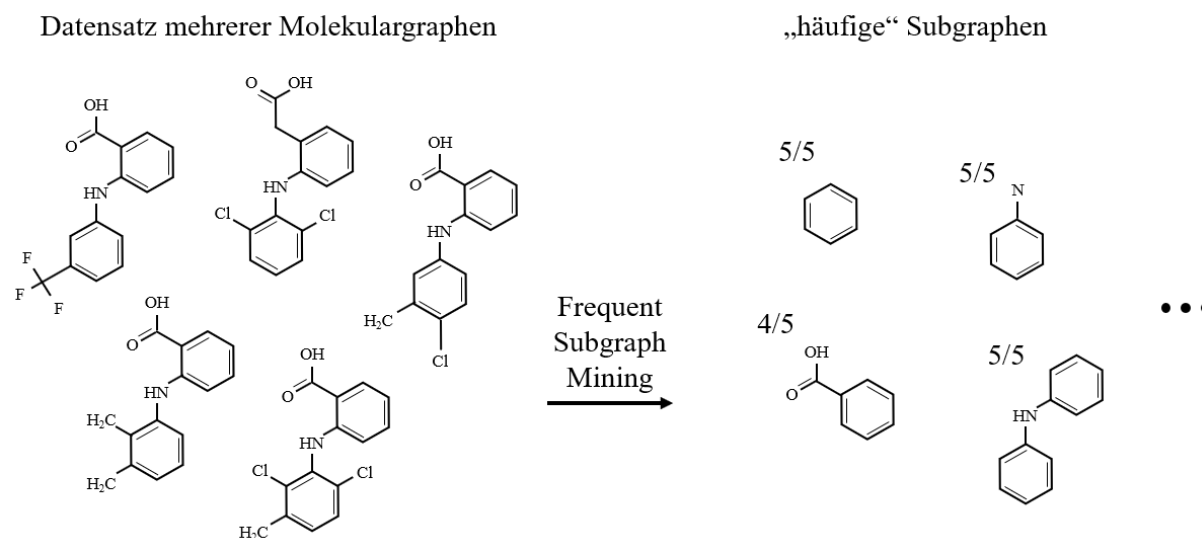


Abbildung 3.7: Beispiel für das Extrahieren häufiger Subgraphen

Grundsätzlich beinhaltet das Frequent Subgraph Mining die beiden Schritte der Generierung von häufigen Subgraphen und der Überprüfung der Häufigkeit der einzelnen Subgraphen in den gegebenen Graphen (Support) (vgl. Yan und Han 2007, S. 100). Allerdings kann die Anzahl von Subgraphen eines Graphen sehr groß werden, sodass im zweiten Schritt für alle möglichen Paare von Subgraphen die strukturelle Übereinstimmung bestimmt werden muss. Da dieses grundsätzliche Vorgehen auch die Subgraphen berücksichtigt, bei denen der Support unter dem minimalen Support liegt, ist es praktisch nicht durchführbar. Nach Cheng, Yan und Han (vgl. 2010, S. 366) resultieren aus dem exponentiell mit der **Patterngröße** wachsenden **Graph-Pattern-Suchraum** zwei gravierende Probleme. Zum einen kommt es zu einem rechnerischen Engpass, der in einer sehr langen Dauer des Mining-Prozesses resultiert und zum anderen behindert die riesige Ergebnismenge des Mining-Prozesses die potenzielle Nutzung von Graphmustern in vielen realen Anwendungen. (vgl. Cheng, Yan und Han 2010, S. 366 ff.) Für einen effizienten Frequent Subgraph Mining Algorithmus werden daher die beiden Bedingungen aufgestellt, dass erstens nicht alle Subgraphen, sondern möglichst nur die häufigen untersucht werden und zweitens jeder Subgraph nur einmal untersucht wird. Diese Bedingungen werden von den gebräuchlichen Algorithmen erfüllt. (vgl. Takigawa und Mamitsuka 2013, S. 53) Um den oben genannten Problemen entgegenzuwirken, wurden Algorithmen zur Extraktion von geschlossenen (closed) und maximalen (maximal) Frequent Subgraphs vorgeschlagen. (vgl. Cheng, Yan und Han 2010, S. 369) Die Methoden zur Extraktion von Mustern aus Graphen können in zwei grundlegende Ansätze unterteilt werden, von denen der erste ein Apriori-basierter und der zweite ein Pattern Growth-basierter Ansatz ist (vgl. Han, Kamber und Pei 2012, S. 591). Außerdem kann nach (vgl. Aggarwal und Wang 2010b, S. 29) das Problem der Suche nach Frequent Patterns je nach Anwendungsbereich auf unterschiedliche Weise definiert werden. Im ersten Fall sollen aus einer Gruppe von Graphen jene Muster bestimmt werden, die in mehreren der gegebenen Graphen vorkommen. Für den zweiten Fall werden in einem einzigen Graphen alle Muster bestimmt, die häufiger als ein gewisser Schwellenwert vorkommen. (vgl. Aggarwal und Wang 2010b, S. 29) Angewendet können diese Algorithmen beispielsweise zur Entdeckung chemischer Strukturen, zur Untersuchung von Proteinstrukturfamilien oder zur Erkennung von Frequent Subgraphs in biologischen Netzwerken. Darüber hinaus können Frequent Subgraphs für eine schnellere Suche in Graphen als Indexierungsmerkmale verwendet werden, wodurch der traditionelle pfadbasierte Indizierungsansatz deutlich übertroffen wird. (vgl. Yan und Han 2007, S. 99) Zusätzlich zur Erstellung von Graphenindizes sind Frequent Pattern für die Charakterisierung von Graphenmengen, für die

Graphbasierte Verfahren

Unterscheidung von unterschiedlichen Graphengruppen und für die Klassifizierung von Graphen nützlich (vgl. Cheng, Yan und Han 2010, S. 366). Da die Suche nach häufigen Teilstrukturen für viele Verfahren des Clusterings oder der Klassifizierung einen Zwischenschritt darstellt, ist das Frequent Subgraph Mining ein grundlegendes Element für Algorithmen des Graph Mining (vgl. Aggarwal 2015, S. 575).

Eine weitere bedeutende Technik des Graph Mining stellt das **Graph Clustering** dar (vgl. Atastina, et al. 2017, S. 2). Dabei wird eine gegebene Menge von Objekten in Gruppen von ähnlichen Objekten unterteilt, wobei die Ähnlichkeit der Objekte normalerweise durch eine mathematische Zielfunktion definiert wird (vgl. Aggarwal und Wang 2010c, S. 276). Auch wenn nicht alle Graphen eine Struktur mit natürlichen Clustern aufweisen, bestimmen Clustering-Algorithmen für jeden vorliegenden Graphen ein Clustering. Im Falle einer gleichmäßigen Verteilung der Kanten über die Menge der Knoten, ist das berechnete Clustering ziemlich willkürlich. Zur Prüfung, ob in einem Graph signifikante Cluster vorhanden sind und ob diese durch ein bestimmtes Clustering aufgedeckt werden, werden Qualitätsmaße und falls möglich auch Visualisierungen des Graphen herangezogen. (vgl. Schaeffer 2007, S. 31) Zheng, et al. (vgl. 2020, S. 7) stellen fest, dass die Methoden zum Clustering ein aktiver Forschungsbereich sind und zahlreiche Clustering-Algorithmen existieren. In dieser Arbeit wird die Ansicht von (vgl. Aggarwal und Wang 2010c, S. 276 f.) vertreten, nach der die Herausforderungen des Graph Clustering in zwei Gruppen abgegrenzt werden – den Node Clustering Algorithmen und Graph Clustering Algorithmen. Für die Anwendung der **Node Clustering Algorithmen** liegt ein großer Graph vor, dessen Knoten mithilfe von numerischen Abstands- oder Ähnlichkeitswerten, die die Kanten des Graphen beschriften, geclustert werden. (vgl. Aggarwal und Wang 2010b, S. 32 f.) Je nachdem welchen Wert die Kanten besitzen, muss die Zielfunktion minimiert (Ähnlichkeitswert) oder maximiert (Abstandswert) werden (vgl. Aggarwal 2017, S. 570). Soll beispielsweise die Ähnlichkeit zwischen den Clustern für eine bestimmte Anzahl an Clustern minimiert werden, reduziert sich dieses Problem im Wesentlichen auf das Problem der Graph-Partitionierung oder das Problem des minimalen Schnitts (vgl. Aggarwal und Wang 2010c, S. 276 f.). Die Graph-Partitionierung zielt etwa auf die Minimierung der Anzahl der Kanten ab, deren Endknoten in unterschiedlichen Clustern liegen (vgl. Schaeffer 2007, S. 31). Die Entdeckung von Quasi-Cliquen stellt eine weitere Möglichkeit zum Clustering dar. Diese Technik zielt nicht darauf ab, die Anzahl der Kanten zwischen den Clustern zu minimieren, sondern die Dichte der Kanten innerhalb eines Clusters zu maximieren. (vgl. Aggarwal 2017, S. 575) Schaeffer (vgl. 2007, S. 27) unterteilt die Algorithmen des Node Clusterings abhängig von deren Input zusätzlich in globale und lokale Clustering Algorithmen. Beim globalen Clustering wird der gesamte Graph geclustert, sodass jeder Knoten einem Cluster zugewiesen wird. Beim lokalen Clustering hingegen wird lediglich eine Teilmenge der Knoten eines Graphs einer Anzahl von Clustern zugeteilt. (vgl. Schaeffer 2007, S. 38) Schaeffer (vgl. 2007, S. 27) stellt außerdem fest, dass viele Algorithmen des Node Clusterings mit denen des Graph Clusterings verwandt sind. Für die **Graph Clustering Algorithmen** liegen mehrere ganze Graphen vor, die nicht allzu groß sind und abhängig von ihrer Struktur geclustert werden müssen. Diese Entscheidung basiert entweder auf einer strukturellen Ähnlichkeitsfunktion oder auf anderen Merkmalen, wie das Vorhandensein von häufigen Mustern in den Graphen. (vgl. Aggarwal und Wang 2010a, S. 6) Diese Art des Clusterings ist anspruchsvoll, da die zugrundeliegenden Graphen strukturell verglichen werden müssen und diese Strukturen anschließend für das Clustering verwendet werden. Zum Einsatz kommen diese Algorithmen einerseits bei klassischen Graphdaten und andererseits bei semistrukturierten Daten. (vgl. Aggarwal und Wang 2010c, S. 277) Die Algorithmen des Graph Clustering besitzen ein breites Anwendungsfeld. Das Clustering von ganzen Graphen wird beispielsweise in der Molekularbiologie, bei

Graphbasierte Verfahren

chemischen Graphen oder XML-Daten verwendet, kann aber generell in jedem Bereich eingesetzt werden, bei dem die zugrundeliegenden Datensätze als kompakte Graphen dargestellt werden können. Das Node Clustering ist besonders dort nützlich, wo die dichten Regionen eines Graphen bestimmt werden müssen. Die Bestimmung der dichten Regionen hängt mit dem Problem der Zusammenfassung von Graphen sowie der Reduktion der Dimensionalität zusammen, durch die Graphen auf einem kleinen Raum dargestellt werden, wodurch sie effektiv für die Indizierung und das Abrufen von Informationen verwendet werden können. (vgl. Aggarwal und Wang 2010c, S. 295 f.) Beispiele für reale Anwendungen sind die Aufdeckung von Gemeinschaften in sozialen Netzwerken und Webanwendungen oder die Erfassung und Extrahierung von nützlichen Informationen aus Bilddateien (vgl. Salem, Moneim und Hassan 2019, S. 12 f.).

Die dritte große Technik des Graph Mining nach (vgl. Atastina, et al. 2017, S. 2) stellt die **Graph Classification** dar. Im Data Mining ist die Klassifizierung generell eng mit dem Clustering verwandt mit dem Unterschied, dass keine ähnlichen Gruppen von Datenpunkten bestimmt werden, sondern mithilfe von Beispielen die Struktur eines Datensatzes gelernt wird. Dabei sind die Beispiele bereits in Gruppen, die auch als Kategorien oder Klassen bezeichnet werden, unterteilt. Zum Lernen wird meist ein Modell verwendet, das für ein oder mehrere ungesehene Datenbeispiele mit zuvor unbekanntem Labels die jeweils zugehörigen Klassen schätzt. (vgl. Aggarwal 2015, S. 285) Ein Label ist definiert als ein Zielwert, der mit jedem Objekt in einem Trainingsdatensatz verbunden ist, und bei der Klassifizierung mit einer Klasse gleichzusetzen ist (vgl. Sammut und Webb 2017, S. 701). Ein Trainingsdatensatz beinhaltet Beispiele, die bereits in verschiedene Klassen partitioniert sind, und ist eine der Eingaben für ein Klassifizierungsproblem. Das durch einen Algorithmus gelernte Modell wird auch als Trainingsmodell bezeichnet und klassifiziert die zuvor ungesehenen Datenpunkte (Testdatensatz). Aufgrund der Tatsache, dass ein Beispieldatensatz zum Lernen der Struktur der Gruppen verwendet wird, wird die Klassifizierung als überwachtes Lernen bezeichnet. (vgl. Aggarwal 2015, S. 285) Auch in der Graphdatenverarbeitung spielt das überwachte Lernen auf Graphen eine zentrale Rolle. Bei den Verfahren der Graph Classification wird wie bereits beim Data Mining davon ausgegangen, dass mit dem Trainingsdatensatz die Zielwerte zur Verfügung stehen, und damit die Zielwerte einer bestimmten Anzahl an ungesehenen Graphen oder eines bestimmten ungesehenen Teils des Graphen abgeleitet wird. (vgl. Tsuda und Saigo 2010, S. 337) Ähnlich zum Graph Clustering existieren für die Graph Classification zwei verschiedene Aufgabenbereiche. Beim **Label Propagation** ist bereits eine Teilmenge der Knoten eines Graphen gelabelt und die Aufgabe besteht darin, auf Grundlage der gelabelten Knoten ein Modell zu lernen, das zur Klassifizierung der ungelabelten Knoten verwendet werden kann. Da angenommen wird, dass ähnliche Knoten auch ähnliche Labels haben sollten, ist die Entwicklung einer Distanzfunktion zur Messung der Ähnlichkeit zwischen zwei Knoten eine zentrale Herausforderung der Label Propagation. Den zweiten Aufgabenbereich stellt die **Graph Classification** dar, bei dem eine Teilmenge der Graphen in einem Graphdatensatz gelabelt ist und versucht wird, daraus ein Modell zu lernen, um die ungelabelten Graphen des Datensatzes zu klassifizieren. (vgl. Aggarwal und Wang 2010b, S. 37 f.) **Abbildung 3.8** ist an die Illustration von Tsuda und Saigo (vgl. 2010, S. 338) angelehnt, bei der einerseits ungelabelte Knoten ("?"") auf Basis der gelabelten Knoten („+1/-1“) mittels Label Propagation klassifiziert werden sollen und andererseits ein einfacher Graph aufgrund seiner Struktur in eine bekannte Klasse durch die Graph Classification eingeteilt wird.

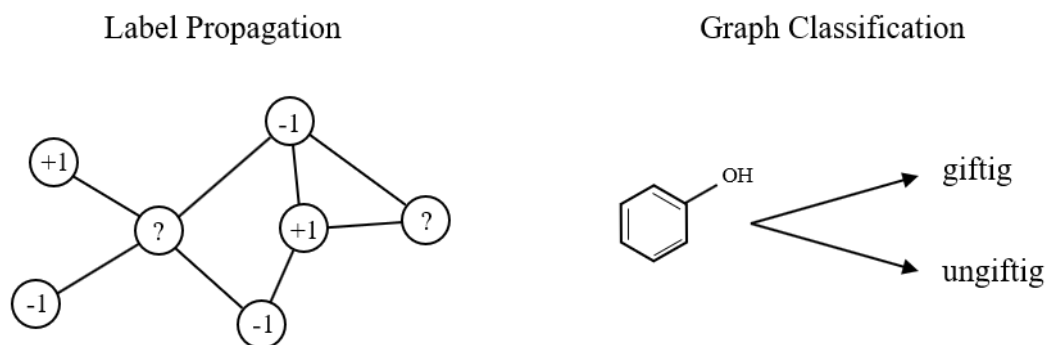


Abbildung 3.8: Die Aufgabenbereiche der Label Propagation und Graph Classification

Anwendung finden die Algorithmen der Label Propagation etwa in der umfangreichen Analyse von sozialen Netzwerken und biologischen Netzwerken (vgl. Tsuda und Saigo 2010, S. 358). Ein konkretes Beispiel ist die Analyse eines sozialen Netzwerks, um das Marketing von Einzelhändlern gezielter auszurichten. Basierend darauf, ob eine Anzahl von Kunden auf eine Werbeaktion reagiert, kann ein Modell aus Kunden gelernt werden, welches die Reaktionen von potenziellen Kunden aus dem sozialen Netzwerk der bekannten Kunden vorhersagt. (vgl. Aggarwal und Wang 2010b, S. 37 f.) Für die Anwendung der Algorithmen des zweiten Aufgabenbereichs gibt es viele Szenarien, wie etwa die Klassifizierung chemischer und biologischer Verbindungen (vgl. Aggarwal und Wang 2010a, S. 6).

Ein weiteres Gebiet ist das der **Link Prediction**. In einem heterogenen Netzwerk existieren zwischen verschiedenen Knoten viele versteckte Rollen oder Beziehungen. Eine Möglichkeit, diese versteckten Beziehungen zu entdecken, ist die Angabe von Einschränkungen basierend auf dem Hintergrundwissen von Experten. Werden diese Einschränkungen beispielsweise durch die Informationsredundanz eines Netzwerks nicht erfüllt, können Knoten oder deren Verlinkungen aussortiert werden. Gleichmaßen können Links ausgehend von einer Bewertung der potenziellen Beziehungen zwischen den Knoten vorhergesagt werden. (vgl. Han, Kamber und Pei 2012, S. 593 f.) Bei der Link Prediction wird angenommen, dass fehlende Links zwischen Knoten existieren oder dass es Links gibt, die erst in Zukunft erscheinen werden. (vgl. Lü und Zhou 2011, S. 1151) Auch wenn die Link Prediction umgangssprachlich mit dem Link Mining gleichgesetzt wird, gehören dem Link Mining zusätzliche Aufgaben wie die Vorhersage des Linktyps, die Erkennung von Gruppen oder der Abgleich von Objekten an (vgl. Han, Kamber und Pei 2012, S. 594). Somit ist die Link Prediction ein wichtiger Teil des Link Mining, der anhand der Informationen der Attribute und der vorhandenen Links vorhersagt, ob es Verbindungen zwischen zwei Knoten geben wird. Diese Aufgabe kann je nachdem ob der neue Link in Zukunft erscheinen wird oder ob es sich um einen versteckten, bisher unbekanntem Link handelt, in zwei verschiedene Vorhersage-Typen unterteilt werden. (vgl. Dong, et al. 2013, S. 1 f.) Die meisten Link Prediction-Algorithmen sind allerdings für statische Netzwerke konzipiert und berücksichtigen nicht die dynamischen Eigenschaften eines Netzwerks, die sich durch die Erweiterung und das Entfernen von Knoten sowie Kanten auszeichnet (vgl. Yao, et al. 2016, S. 82 f.). Yao, et al. (vgl. 2016, S. 84) stellen eine Link Prediction-Methode für dynamische soziale Netzwerke vor. Die simpelste Methode zur Vorhersage von Links ist der ähnlichkeitsbasierte Algorithmus. Dabei wird jedem Knotenpaar ein Wert zugewiesen, der die Ähnlichkeit zwischen den beiden Knoten definiert. Anschließend werden diese Knotenpaare entsprechend ihres Wertes vom größten zum kleinsten sortiert. Je größer der Wert der Ähnlichkeit ist, desto höher ist die Wahrscheinlichkeit, dass eine Verbindung (Link) zwischen den Knoten existiert. (vgl. Lü und Zhou 2011, S. 1153) Dieser Algorithmus ignoriert die zeitlichen Eigenschaften eines Netzwerkes, da er nur einen bestimmten Zeitpunkt angewendet werden kann. Soll auch die Dynamik eines Netzwerkes

Graphbasierte Verfahren

berücksichtigt werden, kann der Graph in unterschiedliche Graph-Sequenzen mit einem gewissen Zeitschritt unterteilt werden. Die Wahl des Zeitschritts ist abhängig von der Anwendung und kann etwa einen Monat oder ein Jahr betragen. (vgl. Dong, et al. 2013, S. 4) Die Methoden der Link Prediction besitzen ein breites Anwendungsfeld, wie etwa das Vorschlagen von Empfehlungen im Onlinehandel, um den Umsatz zu steigern, oder der Identifikation von Untergrundbeziehungen in terroristischen Netzwerken (vgl. Lü und Zhou 2011, S. 1163). Erneut kann diese Technik in biologischen Netzwerken eingesetzt werden, bei der die Existenz einer Verbindung zweier Knoten durch kostspielige Feld- und/oder Laborexperimente nachgewiesen wird. Die experimentellen Kosten könnten durch den Fokus auf die wahrscheinlichsten Verbindungen der Knoten stark reduziert werden. (vgl. Lü und Zhou 2011, S. 1151)

Durch die Entwicklung neuer Technologien und der Entdeckung von neuem Wissen, werden die Graphen immer größer. Umfangreiche Graphen, die unter Umständen mehrere Milliarden Knoten und Kanten enthalten können, stellt das Graph Mining vor neue Herausforderungen. So können beispielsweise die Algorithmen des Pattern Mining stark eingeschränkt werden und auch die Visualisierung solch eines Graphen ist komplex. Ein komprimierter Graph, der die Eigenschaften des ursprünglichen Graphen annimmt, könnte diese Probleme lösen. Die Methoden der **Graph Compression** zielen darauf ab, die High-Level-Struktur eines Graphen leicht verständlich zu machen. (vgl. Zhou 2010, S. 1 f.) Diese Algorithmen identifizieren konkret unterschiedliche Muster in den Graphdaten und ersetzen die größeren Muster durch einen Marker, der die ursprünglichen Daten repräsentiert. Der resultierende Graph kann durch die umgekehrte Anwendung der Algorithmen wieder entkomprimiert werden. Doch auch die komprimierten Daten können für Analysen oder Berechnungen genutzt werden. (vgl. Shah 2018, S. 1) Nach Garg und Jaakkola (vgl. 2019, S. 1) werden die Methoden der Graph Compression je nachdem ob sie Kanten entfernen (Sparsification) oder Knoten zusammenführen (Coarsening) in zwei Kategorien unterteilt. Da die Mitglieder eines Clusters ähnliche Eigenschaften haben und durch Hypernodes dargestellt werden können, adressieren die Graph Clustering-Algorithmen in gewisser Weise einen Teil des Problems der Graph Compression. Allerdings basieren diese Algorithmen meist auf der Distanzmatrix und konzentrieren sich nicht auf die Struktur des ursprünglichen Graphen. (vgl. Zhou 2010, S. 1 f.) Neben der Anwendung in sozialen oder chemischen Netzwerken, können die Methoden der Graph Compression für die Entschlüsselung der Sequenzzusammensetzung von genetischem Material in biologischen Netzwerken eingesetzt werden. (vgl. Besta und Hoefler 2018, S. 15) Darüber hinaus können sie zur Reduktion der benötigten Zeit von Graphabfragen oder zum effizienteren Einsatz von Verfahren des Data Mining verwendet werden (vgl. Besta und Hoefler 2018, S. 31).

4 Entwicklung einer Klassifizierung von Graph Mining Algorithmen

Das eigentliche Ziel dieser Arbeit, die Klassifizierung von Verfahren des Graph Mining hinsichtlich der Beantwortung logistischer Fragestellungen in Supply Chains, wird in diesem Kapitel umgesetzt. Bevor eine Methode zur Klassifizierung erarbeitet wird, müssen in Abschnitt 4.1 Vorbereitungen getroffen werden. Diese betreffen unter anderem die Auswahl des Bewertungsverfahrens für die Klassifizierung (vgl. Abschnitt 4.1.1), die Auswahl der zu klassifizierenden Verfahren des Graph Mining (vgl. Abschnitt 4.1.3) und die Auswahl der logistischen Fragestellungen (vgl. Abschnitt 4.1.4). In Abschnitt 4.2 werden die Algorithmen des Graph Mining anhand der zuvor aufgestellten Bewertungskriterien (vgl. Abschnitt 4.1.5) unterschiedlichen Klassen (vgl. Abschnitt 4.1.6) zugeteilt, die Aufschlüsse darüber geben, ob ein Algorithmus zur Lösung oder Vereinfachung dieser Fragestellung verwendet werden kann. Die Ergebnisse der Klassifizierungen werden in Abschnitt 4.2.9 zusammengefasst.

4.1 Vorbereitungen für Klassifizierungsverfahren

Als Vorbereitung für die Klassifizierung von verschiedenen Algorithmen des Graph Mining wird das Graph Mining dem Graph Management gegenüber abgegrenzt und somit abgesteckt welcher Bereich bei der Klassifizierung berücksichtigt wird. Das Klassifizierungsverfahren teilt sich in dieser Arbeit in zwei Schritte. Im ersten Schritt wird die Eignung von Algorithmen zur Beantwortung verschiedener Fragestellungen anhand von definierten Kriterien bewertet. Bevor dies geschehen kann, müssen die zu klassifizierenden Algorithmen ausgewählt und logistische Fragestellungen formuliert werden. Außerdem werden unterschiedliche Bewertungsverfahren beschrieben, von denen eines ausgewählt wird. In einem zweiten Schritt werden aus den verschiedenen Möglichkeiten der Kriterienerfüllung Klassen gebildet, in die die Algorithmen abhängig von der Bewertungsergebnisse eingeteilt werden.

4.1.1 Auswahl des Bewertungsverfahrens

Um das passende Bewertungsverfahren auszuwählen, werden unterschiedliche Methoden zur Bewertung vorgestellt. Allen Verfahren gemein ist, dass sie eine Auswahl an Kriterien benötigen. In dem vorliegenden Fall sind die Merkmalsausprägungen qualitativ und nicht quantitativ, was die Auswahl der Methoden einschränkt. Aus diesem Grund wird für die Bewertung ein qualitatives Verfahren eingesetzt. Solch ein Verfahren ist im Allgemeinen der Technologiebewertung zuzuordnen. Dieser Bereich des Technologiemanagements ermittelt für ein bestimmtes Bewertungsobjekt, inwieweit es definierte Zielzustände definiert. (vgl. Schuh, Klappert und Moll 2011, S. 17) Qualitative Bewertungsmethoden unterscheiden sich von den quantitativen Methoden in der Art der Kriterien zur Bewertung des untersuchten Objekts, wobei die zusammenfassende Bewertung von artfremden Kriterien zu Problemen führen kann. Die beiden Möglichkeiten zum Ausdruck einer Merkmalsausprägung sind entweder in Zahlen (quantitativ) oder in Worten (qualitativ). (vgl. Hoffmeister 2008, S. 277) Im Folgenden werden eine Auswahl von unterschiedlichen Bewertungsmethoden beschrieben und auf ihre Eignung für die vorliegende Problemstellung überprüft.

Beim simpelsten und bekanntesten Verfahren der **Argumentenbilanz** werden die Vorteile und Nachteile von alternativen Technologien in einer Liste dargestellt. Es kommt vor allem zum Einsatz, wenn keine monetäre oder quantitative Beschreibung möglich ist und schafft einen groben Überblick über die untersuchten Alternativen. Allerdings sind die Argumente nicht präzisierbar, wodurch deren Wichtigkeit nicht gegeneinander erkennbar ist. (vgl. Haag, et al. 2011, S. 324 ff.) Da die Bewertungskriterien für die Algorithmen qualitativ sind, würde sich diese Methode theoretisch zur Bewertung eignen. Jedoch können aus den Ergebnissen für den zweiten Schritt des Klassifizierungsverfahrens nur schlecht

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

unterschiedliche Klassen gegeneinander abgrenzen. Dadurch ist eine Klassifizierung problematisch und die Argumentenbilanz somit nicht als Bewertungsverfahren geeignet.

Bei der Multikriterienbewertung diverser komplexer Alternativen nach der **Nutzwertanalyse** werden überwiegend qualitative Kriterien entsprechend der subjektiven Einschätzung des Methodenanwenders quantifiziert und somit untereinander vergleichbar gemacht. Bei der Einschätzung bestimmt der Anwender den Erfüllungsgrad, also den Wert, wie sehr ein Kriterium von dem Bewertungsobjekt erfüllt wird. Je nach Ausprägung dieses Wertes wird dem Objekt ein Skalenwert für das Kriterium zugeteilt. Darüber hinaus kann durch den Einsatz eines Gewichtungsfaktors die Relevanz einzelner Kriterien berücksichtigt werden. (vgl. Haag, et al. 2011, S. 327 ff.) Durch die Verbindung von Erfüllungsgrad eines Kriteriums und Gewichtungsfaktor entsteht für jedes untersuchte Element ein dimensionsloser Nutzwert, der zur Ordnung der Alternativen dient (vgl. Zangemeister 1976, S. 45). Eingesetzt wird dieses Verfahren hauptsächlich bei Entscheidungssituationen, die aus gleichartigen Alternativen bestehen, bei denen keine monetäre Bewertung möglich ist oder bei denen viele entscheidungsrelevante Merkmale zu berücksichtigen sind (vgl. Hoffmeister 2008, S. 279). Soll die Nutzwertanalyse eingesetzt werden ist es wichtig, dass die Bewertungskriterien unabhängig voneinander gewählt werden müssen, damit Mehrfachbewertungen vermieden werden (vgl. Haag, et al. 2011, S. 327). Bei der Bewertung von Algorithmen des Graph Mining sind die Kriterien nicht unabhängig voneinander und es existieren nicht so viele entscheidungsrelevante Merkmale. Außerdem ist bei den identifizierten Kriterien ein Einsatz eines Erfüllungsgrads nicht zielführend und es ist nicht gefordert die Alternativen nach einem Nutzwert zu ordnen, da das Hauptziel eine Klassifizierung und keine Rangliste ist, welcher Algorithmus am besten zur Beantwortung einer logistischen Fragestellung geeignet ist. Aus diesen Gründen eignet sich die Nutzwertanalyse nicht als Bewertungsmethode.

Ein weiteres Bewertungsverfahren ist die Erstellung von **Checklisten**, durch die alle bedeutenden Kriterien eines bestimmten Themas ohne großen Aufwand erfasst werden können. Dabei wird aus einer qualitativen Bestandsaufnahme wichtiger Kriterien eine Handlungsempfehlung herbeigeführt. Die identifizierten Kriterien lassen sich priorisieren, sodass bei Nichterfüllung eines hochpriorisierten Kriteriums beispielsweise eine Alternative ausgeschlossen werden kann. Außerdem lässt sich der Kriterienkatalog bei diesem Verfahren frei auswählen und durch zusätzliche Kriterien ergänzen. Auf der anderen Seite stellt dies allerdings einen Nachteil dar, da die Anzahl der zu berücksichtigenden entscheidungsrelevanten Kriterien ungeklärt ist. Außerdem können die Kriterien umstrittene Wirkungsrichtungen untereinander besitzen. (vgl. Haag, et al. 2011, S. 326 f.) Da es bei der Bewertung verschiedener Algorithmen zur Beantwortung von logistischen Fragestellungen entscheidend ist, ob identifizierte Kriterien erfüllt werden und der Grad der Erfüllung sowie eine Gewichtung dabei nicht interessiert, ist die Checkliste als Methode grundsätzlich geeignet. Weitere Gründe für dieses Verfahren sind die Betrachtung von qualitativen Merkmalsausprägungen und deren Abhängigkeit untereinander.

Von den beschriebenen Bewertungsverfahren eignet sich die Methode der Checkliste am ehesten für den vorliegenden Anwendungsfall und wird deshalb im weiteren Verlauf genutzt. Die Vorgehensweise dieser Methode teilt sich in drei Schritte. Zuerst werden die Entscheidungskriterien festgelegt (Abschnitt 4.1.5), danach werden aus den Entscheidungskriterien Prüffragen zusammengestellt (Abschnitt 4.1.5) und im letzten Schritt wird die Checkliste aufgestellt (Abschnitt 4.2). (vgl. Haag, et al. 2011, S. 326)

4.1.2 Differenzierung von Graph Management und Graph Mining

In dieser Arbeit sind die Algorithmen des Graph Mining zur Beantwortung von logistischen Fragestellungen im Bereich von Supply Chains zu klassifizieren. Um diese Klassifikation vorzunehmen, muss

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

zuerst Klarheit darüber bestehen, welche Algorithmen dafür betrachtet werden. Die in den Abschnitten 3.3.1 und 3.3.2 vorgestellten Algorithmen sind den Bereichen des Graph Management und des Graph Mining zuzuordnen, jedoch wurde in der Literatur keine klare Abgrenzung dieser beiden Bereiche zueinander gefunden und somit auch keine Klassifizierung dessen Algorithmen. Lediglich Aggarwal und Wang führt unterschiedliche Beschreibungen des Graph Management und Graph Mining auf. Nach ihnen zielen die Algorithmen des Graph Management auf die Verwaltung und Indizierung großer Mengen an Graphdaten ab und die Algorithmen des Graph Mining auf die Extraktion von Mustern, Trends, Klassen und Clustern aus Graphen (vgl. Abschnitt 3.3.2). Dieser Abschnitt stellt einen Versuch dar, die beiden Bereiche aufbauend auf der Auffassung von Aggarwal und Wang gegeneinander abzugrenzen (vgl. Abschnitt 3.3.2).

Wie bereits in Abschnitt 3.1 erwähnt, ist der Algorithmus für Tarjan eine Berechnungsmethode zur Durchführung einer Aufgabe, die basierend auf dem abstrahierten Modell einer realen Aufgabenstellung entwickelt wird. Die Attribute im Prozess der Forschung, Entdeckung und Entwicklung, der in *Abbildung 3.3* dargestellt ist, können für die in dieser Arbeit behandelte Thematik angepasst werden. Die Aufgabenstellung bezieht sich auf ein Einsatzgebiet in der realen Welt und kann mit einer logistischen Fragestellung einer Aufgabe der Teilgebiete aus dem Aufgabenmodell (vgl. Abschnitt 2.4) gleichgesetzt werden. Die Abstrahierung dieser Aufgabenstellung entspricht dabei einem graphbasierten Modell (vgl. Abschnitt 3.2) und die Algorithmen stehen für die unterschiedlichen Verfahren des Graph Management und/oder des Graph Mining. Der angepasste Prozess wird in *Abbildung 4.1* dargestellt und der Fokus dieses Abschnitts liegt auf dem Knoten, der den Algorithmus repräsentiert.

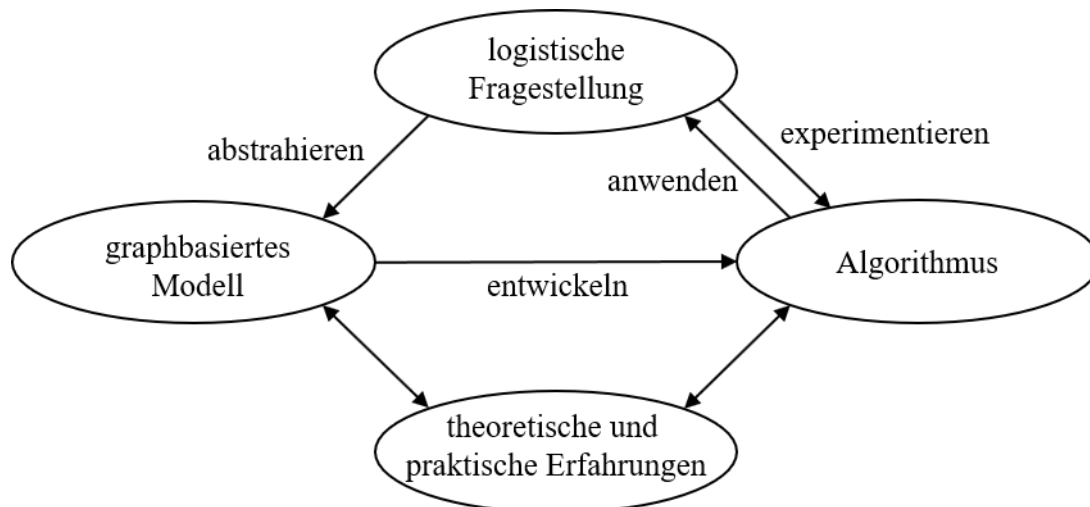


Abbildung 4.1: Angepasster Prozess zur Forschung, Entdeckung und Entwicklung

Um eine begründete Aussage zur Differenzierung der Algorithmen treffen zu können, werden zwei neue Kriterien identifiziert. Das erste Kriterium bezieht sich darauf, ob der Algorithmus neue Muster extrahiert. Wie bereits in Abschnitt 3.3.2 beschrieben, liegt das aus einem Graphen extrahierte Muster ebenfalls als Graph vor und teilt sich mit der Musterdefinition des Data Mining die Eigenschaften der Neuartigkeit und Nützlichkeit. Nach dieser Definition sollten Muster ebenfalls mit gewisser Sicherheit für neue Daten gelten und darüber hinaus verständlich für den Benutzer sein. Das zweite Kriterium steht für den Informationsgehalt des Outputs der Algorithmen, dessen Einteilung sich nach den Stufen der Wissenstreppe nach North (s. Abschnitt 2.2) richtet. Die Ausprägungen dieser Kriterien für die Algorithmen aus Abschnitt 3.3.1 und 3.3.2 werden zusammen mit deren prägnanten Zielbeschreibungen in

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

der folgenden *Tabelle 4.1* zusammengefasst, wobei für eine ausführliche Beschreibung der Zielsetzungen auf Abschnitt 3.3.1 und Abschnitt 3.3.2 zu verweisen ist.

Tabelle 4.1: Algorithmen und ihre Eigenschaften

Algorithmus	Ziel	Extraktion von Mustern	Zuteilung Wissenstreppe
Adjacency Queries	Überprüfung der Nachbarschaft von Knoten oder Kanten	Nein	Informationen
Pattern Matching Queries	Suche aller Teilgraphen, die mit einem gegebenen Muster übereinstimmen	Nein	Informationen
Reachability Queries	Überprüfung der Existenz eines Pfades, der zwei bestimmte Knoten verbindet	Nein	Informationen
Analytical Queries	Messung von topologischen Eigenschaften des Graphen	Eventuell bei komplexen Abfragen	Informationen
Graph Pattern Mining	Extraktion aller Muster, die häufiger als ein bestimmter Schwellenwert aufkommen	Ja	Daten
Graph Clustering	Unterteilung von einzelnen Knoten oder kleinen Graphen in mehrere Gruppen ähnlicher Objekte, wobei nicht bekannt ist, welche Objekte sich ähneln	Ja	Daten
Graph Classification	Erlernen der allgemeinen Struktur eines Graphen/mehrerer Graphen anhand vorgegebener gelabelter Knoten/Graphen und anschließende Klassifizierung der ungelabelten Knoten/Graphen	Ja	Daten
Link Prediction	Vorhersage der potenziellen Beziehung zweier Knoten	Ja	Daten
Graph Compression	Komprimierung durch Identifizierung von häufigen Mustern, die dann durch einen Marker ersetzt werden	Ja	Daten

Die von Aggarwal und Wang genannten Algorithmen des Graph Mining gehören zum Bereich des Data Mining, welcher nach Fayyad, Piatetsky-Shapiro und Smyth einen Teilprozess des KDD-Prozesses darstellt (vgl. Abschnitt 3.3.2). Bei den Algorithmen, die dem Graph Mining zuzuordnen sind, werden ausnahmslos neue Muster extrahiert. Diese Ergebnisse müssen nach dem KDD-Prozess in einem anschließenden Schritt interpretiert werden, um Wissen zu erlangen. Diese Interpretation kann mit der Vernetzung von Informationen mit Kontext, Erfahrungen und Erwartungen, die bei der Wissenstreppe nach North zur Generierung von Wissen aus Informationen nötig ist, verglichen werden. Dieser Schritt fällt bei den Algorithmen, die dem Graph Management zuzuordnen sind, meist weg, da angenommen wird, dass bereits bei der Formulierung der Fragestellung auf Erfahrungen zurückgegriffen wurden, ein Bezug zum Kontext hergestellt wurde und an das Ergebnis gewisse Erwartungen gestellt werden. Somit stellt das Ergebnis der Graph Management Algorithmen nach den hier aufgestellten Annahmen bereits Wissen dar.

Die Ergebnisse des Vergleichs bekräftigen die Aussage von Aggarwal und Wang, dass das Graph Management und das Graph Mining zwei unterschiedliche Bereiche darstellen (vgl. Abschnitt 3.3.2). Aus diesem Grund werden für die Klassifizierung nur die in Abschnitt 3.3.2 vorgestellten Algorithmen herangezogen. Dies soll allerdings nicht bedeuten, dass das Graph Mining das Graph Management ersetzen soll oder umgekehrt. Denn auch wenn die Algorithmen der beiden Bereiche für dasselbe

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

Anwendungsgebiet der Fragestellungen im Supply Chain Management genutzt werden, zielen sie nicht auf das gleiche Ergebnis ab. Wie bereits erwähnt gehört die Verwaltung und Indizierung großer Mengen an Graphdaten zu den Zielen des Graph Management und die Extraktion von Mustern, Trends, Klassen und Clustern zu denen des Graph Mining.

Vielmehr könnten die Algorithmen des Graph Mining möglicherweise dafür verwendet werden, das Graph Management zu unterstützen. Dies wurde bereits von Angles und Gutierrez bei den komplexen Formen der Analytical Queries beschrieben, die sich auch den Algorithmen des Graph Mining bedienen können (vgl. Abschnitt 3.3.1). Allerdings haben zielen diese Abfragen darauf ab, einen Graphen zu analysieren und stellen kein bestimmtes Verfahren dar, um neue Muster in Graphen aufzudecken. Aus diesem Grund werden sie auch weiterhin dem Graph Management zugeordnet. Sie sind ein Beispiel dafür, dass die Algorithmen des Graph Mining zusammen mit denen des Graph Management zur Lösung einer gemeinsamen Problemstellung genutzt werden können. Dieser Gedanke wird auch in den Abschnitten 4.1.5 und 4.1.6 aufgegriffen. Daraus ergibt sich die Forschungsfrage, wie die Algorithmen des Graph Management und es Graph Mining effizient für die gemeinsame Bearbeitung einer Aufgabe verwendet werden können bzw. wie die Algorithmen des Graph Mining unterstützend eingesetzt werden können.

Tarjan vertritt die Auffassung, dass die Abstraktion bzw. das Modell angepasst werden muss, wenn der Algorithmus das reale Problem nicht lösen kann. Da die Algorithmen nicht das gleiche Ziel verfolgen, ist ebenfalls kritisch zu hinterfragen, ob die in Abschnitt 3.2 beschriebenen Modelle nicht eventuell für die Algorithmen des Graph Mining angepasst werden müssten und wie diese Anpassung aussehen könnte. Es werden hier nur die Modelle für das Graph Mining genannt, da die Modelle für das Graph Management bereits einen hohen Optimierungsgrad besitzen. Durch die Unterschiede in den Algorithmen und auch eventuell in den Modellen ergibt sich die potenzielle Folge für den in *Abbildung 4.1* dargestellten Prozess nach Tarjan, dass jeweils eine Abbildung für das Graph Management und das Graph Mining gestaltet werden müsste. Folglich dürften auch die bei diesen Prozessen gewonnenen Erkenntnisse, welche auf andere Bereiche übertragen werden könnten, unterschiedlich sein. Es wäre folglich ebenfalls interessant zu wissen, in welchen Punkten sich diese Erkenntnisse unterscheiden.

Basierend auf den Resultaten des Vergleichs der beiden Algorithmentypen kann die These aufgestellt werden, dass die Ergebnisse der Graph Management Algorithmen auf der Wissenstreppe nach North über den Ergebnissen der Graph Mining Algorithmen liegen, siehe *Abbildung 4.2*. Diese These ist ebenfalls kritisch zu prüfen.

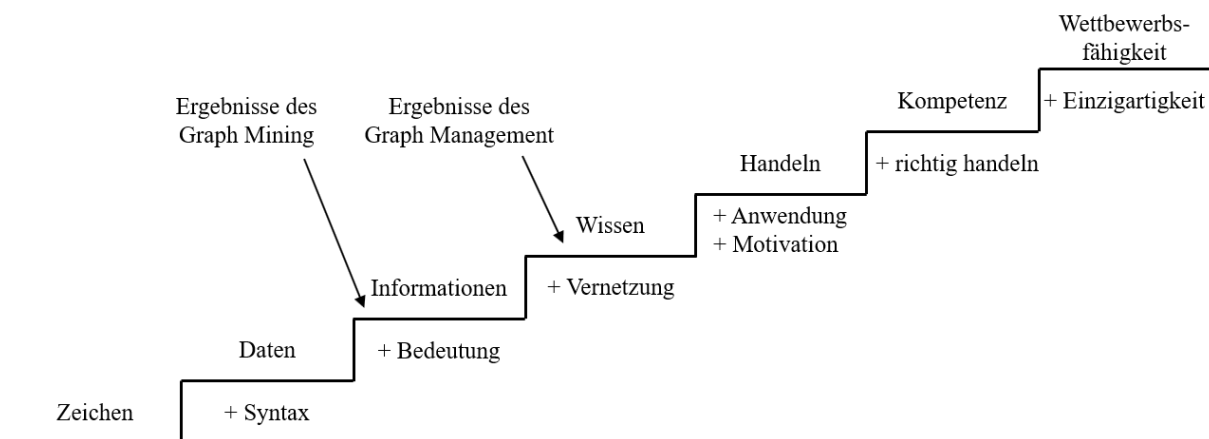


Abbildung 4.2: Einordnung der Ergebnisse des Graph Mining und Graph Management

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

4.1.3 Auswahl der zu klassifizierenden Algorithmen

Basierend auf den Ergebnissen aus Abschnitt 4.1.2 werden lediglich die von Aggarwal und Wang beschriebenen Algorithmen des Graph Mining (vgl. Abschnitt 3.3.2) und nicht zusätzlich die Graph Management Algorithmen für die Klassifizierung verwendet. Die Einsatzmöglichkeiten der einzelnen Algorithmen leiten sich aus der ausführlichen Beschreibung in Abschnitt 3.3.2 ab und werden im Folgenden für die Klassifizierung kurz zusammengefasst.

Graph Pattern Mining Der Algorithmus des Graph Pattern Mining zielt darauf ab, häufige Subgraphen in einem Graphen zu entdecken. Dafür wird zuerst ein Schwellenwert festgelegt und anschließend werden alle Subgraphen identifiziert die häufiger vorkommen als dieser Schwellenwert.

Node-Clustering Der Algorithmus des Node-Clustering ist dem Graph Clustering zuzuordnen. In diesem Fall liegt ein großer Graph vor, für den die Knoten mithilfe einer Zielfunktion unterschiedlichen Clustern zugeteilt werden. Diese Zielfunktion muss entweder im Falle von Ähnlichkeitswerten minimiert oder maximiert werden, wenn Abstandswerte vorliegen.

Graph-Clustering Der Algorithmus des Graph-Clustering ist ebenfalls dem Graph Clustering zuzuordnen. Im Gegensatz zum Node-Clustering liegt dabei nicht nur ein Graph, sondern mehrere Graphen vor, die abhängig von ihrer gesamten Struktur in Cluster eingeteilt werden. Jedoch dürfen diese Graphen nicht allzu groß sein, da es sonst zu einem rechnerischen Engpass kommt.

Label-Propagation Der Algorithmus der Label-Propagation ist der Graph Classification zuzuordnen. Dabei liegt ein Graph vor, für den bereits eine Teilmenge der Knoten gelabelt ist. Auf Grundlage der gelabelten Knoten wird bei diesem Algorithmus ein Modell mit der Struktur des Graphen gelernt, das dafür eingesetzt werden kann, die restlichen ungelabelten Knoten zu klassifizieren.

Graph-Classification Der Algorithmus der Graph-Classification ist ebenfalls der Graph Classification zuzuordnen. Analog zum Graph-Clustering bezieht sich dieser Algorithmus nicht auf einzelne Knoten, sondern auf komplette Graphen. Aus einer zur Verfügung stehenden Menge an Graphen ist eine Teilmenge dieser Graphen gelabelt und der Algorithmus versucht, ein Modell mit der Struktur der Graphen zu lernen, das zur Klassifizierung der restlichen ungelabelten Graphen eingesetzt werden kann.

Link Prediction Der Algorithmus der Link Prediction wird eingesetzt, um in einem Graphen Verbindungen zwischen Knoten zu entdecken. Es wird dabei angenommen, dass es entweder fehlende Verbindungen gibt oder es Verbindungen gibt, die erst in Zukunft erscheinen werden. Diese Extraktion neuer Links geschieht anhand der vorliegenden Attribute und Links. Allerdings berücksichtigen die meisten Algorithmen der Link Prediction keine dynamischen Netzwerke, sondern sind für statische Netzwerke konzipiert. Zur Berücksichtigung der zeitlichen Eigenschaften kann der Graph abhängig eines bestimmten Zeitschritts in verschiedene Graph-Sequenzen unterteilt werden.

Graph Compression Der Algorithmus der Graph Compression wird dafür verwendet, sehr umfangreiche Graphen zu komprimieren und somit die High-Level-Struktur eines Graphen einfach verständlich zu machen. Dies geschieht durch die Identifikation von Mustern in einem Graphen und dem anschließenden Austausch der größeren Muster mit einem Marker, der die Daten dieser Muster repräsentiert. Es können außerdem Kanten entfernt oder Knoten zusammengeführt werden. Bei erfolgreicher Komprimierung wird die Durchführung der Graph Pattern Mining Algorithmen erheblich erleichtert und die Komplexität zur Visualisierung des Graphen wird vermindert.

4.1.4 Auswahl der logistischen Fragestellungen

In diesem Abschnitt werden jene Teilgebiete des SCM-Aufgabenmodells (vgl. Abschnitt 2.4.2 und 2.4.3) identifiziert, die einen direkten Bezug zur Supply Chain besitzen. Im Falle eines direkten Bezugs werden aus dem jeweiligen Teilgebiet eine oder mehrere logistische Fragestellungen und deren Lösung abgeleitet, damit die Eignung der ausgewählten Algorithmen (vgl. Abschnitt 4.1.3) zur Beantwortung der Fragestellung anhand der identifizierten Kriterien (vgl. Abschnitt 4.1.5) überprüft werden kann. Es ist anzumerken, dass die beispielhaften Szenarien aus den Beschreibungen der Teilgebiete in den Abschnitten 2.4.2 und 2.4.3 abgeleitet werden und nicht der Literatur entstammen. Ein Graph Mining Algorithmus trägt zur Lösung der Fragestellung bei, wenn sein Ergebnis nach der Interpretation zur Lösung genutzt werden kann. Er wird als vereinfachend beschrieben, wenn eine zusätzliche Weiterverarbeitung des Ergebnisses durch einen Algorithmus des Graph Management zur Lösung beiträgt.

Das Teilgebiet der **strategischen Netzwerkgestaltung** (vgl. Abschnitt 2.4.2) besitzt einen unmittelbaren Zusammenhang mit der Supply Chain, da es sich mit der kosteneffektiven Auslegung und Gestaltung des gesamten Logistiknetzwerkes beschäftigt. Aus diesem Grund werden zwei logistische Fragestellungen abgeleitet. Die erste lautet: „Wo soll ein neues Werk für Produkt X errichtet werden?“. Ein Ergebnis, das nachvollziehbar begründet, warum an einem bestimmten Ort ein neues Werk für das besagte Produkt errichtet werden sollte, wird als Lösung der Fragestellung angesehen. Aus dem Ergebnis zur zweiten Frage „Wie viele Standorte zur Distribution sind zu errichten?“ sollten Aussagen zur gesuchten Anzahl der zu errichtenden Distributionsstandorten sein.

Die **Bedarfsplanung** (vgl. Abschnitt 2.4.3) hat ebenfalls einen Bezug zur Supply Chain, da sie zur Prognose des lang-, mittel- und kurzfristigen Bedarfs die Bestellungen, die in ERP-Systemen gespeichert sind, über die jeweiligen Supply-Chain-Stufen auswertet. Deshalb wird aus diesem Gebiet die Frage „Wie wird sich der Bedarf an Produkt X das kommende Quartal in Region Y entwickeln?“ abgeleitet. Für die Antwort werden nützliche Ergebnisse erwartet, die bei der Beantwortung der Frage unterstützen.

Die **Netzwerkplanung** (vgl. Abschnitt 2.4.3) beschäftigt sich mit der Koordinierung der einzelnen Partner der Supply Chain oder des Netzwerkes über einen gewissen Planungshorizont und besitzt somit einen Bezug zur Supply Chain. Die aus diesem Gebiet abgeleitete Frage lautet „Welches Werk muss wie viele Produkte eines Typs produzieren, damit der gesamte Bedarf bestmöglich abgedeckt wird?“. Ziel ist es, dass die Ergebnisse dabei helfen, die Menge an zu produzierenden Produkten zu bestimmen.

Bei der **Beschaffungsplanung** (vgl. Abschnitt 2.4.3) wird die Teileversorgung einer mehrstufigen Lagerstruktur so optimiert, dass stets genügend Materialien am exakten Ort bereitgestellt werden. Auch bei diesem Teilgebiet ist ein Bezug zur Supply Chain gegeben, weshalb die Fragestellung „Für welche Werke sind welche Anlieferhythmen zu wählen, damit die Materialbereitstellung sichergestellt wird?“ abgeleitet wird. Ein Ergebnis, aus dem ersichtlich wird, warum ein bestimmter Anlieferungsrhythmus zu bevorzugen ist, würde die Fragestellung beantworten.

Die **Produktionsplanung** (vgl. Abschnitt 2.4.3) beschäftigt sich mit der Erstellung eines Produktionsplans für ein Werk, der unter anderem die Auslastung optimiert und die Termintreue sicherstellt. Bei diesem Gebiet wird die Verbindung zur Supply Chain nicht so deutlich wie bei den anderen Teilgebieten, weshalb keine logistische Fragestellung aus der Produktionsplanung bei der Bewertungsmethode betrachtet wird.

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

Damit die Nachfrage befriedigt werden kann, beschäftigt sich die **Distributionsplanung** (vgl. Abschnitt 2.4.3) unter anderem mit der optimierten Planung der Produktverteilung zum Kunden und hat somit einen direkten Bezug zur Supply Chain. Aus diesem Grund wird die Frage „Auf welchen Verkehrsträger fällt die Wahl vor dem Hintergrund der Transportkostenminimierung?“ in dem weiteren Vorgehen berücksichtigt. Ein Ergebnis, das für einen Verkehrsträger spricht und gleichzeitig nachvollziehbar ist, würde die Fragestellung lösen.

Das Teilgebiet der **Verfügbarkeits- und Machbarkeitsprüfung** (vgl. Abschnitt 2.4.3) prüft, ob Kundenanfragen erfüllbar sind, und hat dadurch einen indirekten Kontakt zur Supply Chain, da für diese Prüfung die verfügbaren Materialien und Kapazitäten berücksichtigt werden. Es lässt sich die Frage „Kann die Lieferung bis zu einem definierten Datum erfolgen?“ ableiten und gesucht wird eine nachvollziehbare Zu- oder Absage dieser Anfrage.

Die **Beschaffungsfeinplanung** (vgl. Abschnitt 2.4.3) beschäftigt sich mit der optimierten Planung der Anlieferungen und besitzt somit einen Bezug zur Supply Chain. Deshalb wird aus diesem Gebiet die Frage „Bei welchen Transportfahrzeugen der Zulieferer ist eine termingerechte Anlieferung zu erwarten?“ gebildet. Ein Ergebnis, welches bei der Beantwortung der Frage unterstützt, wird als Lösung gewertet.

Zur Aufgabe der **Produktionsfeinplanung** (vgl. Abschnitt 2.4.3) gehört die mittel-, bzw. kurzfristige Planung des Produktionsbereichs. Die Erstellung eines passenden Modells für beispielsweise die Festlegung von Fertigungs- oder Montageaufträgen ist nicht zielführend, weshalb auch für dieses Teilgebiet keine logistische Fragestellung betrachtet wird.

Einen direkten Bezug zu Supply Chains besitzt die **Distributionsfeinplanung** (vgl. Abschnitt 2.4.3), da sie sich mit der optimalen Planung von Touren, Transportmitteln und deren Beladung für einen kurzen Zeithorizont, damit die Belieferung termingerecht erfolgt. Das Ergebnis der aus diesem Teilgebiet abgeleiteten Fragestellung „Wie sollten die Verkehrsträger beladen werden, damit sie möglichst effizient genutzt werden?“ sollte Erkenntnisse bringen, mit denen die Beladung der Transportmittel bestimmt werden kann. Für die zweite Frage ist davon auszugehen, dass bei einer großen Anzahl an Kunden ein dringender Bedarf an einem bestimmten Produkt entstanden ist. Die daraus resultierende Frage lautet: „Welche Unternehmen sind als erstes zu beliefern?“. Das Ergebnis der Berechnung sollte bei der Beantwortung dieser Frage unterstützen können.

4.1.5 Definition von Bewertungskriterien

Sollen die Algorithmen des Graph Mining zur Beantwortung von logistischen Fragestellungen durch eine Checkliste (vgl. Abschnitt 4.1.1) bewertet werden, müssen in dem ersten Schritt einige Kriterien definiert werden. Das erste und wichtigste Kriterium ist die grundsätzliche Eignung des Algorithmus zum Einsatz bei der jeweiligen Fragestellung und auf das Modell. Ist diese Eignung nicht gegeben, können auch die folgenden Kriterien nicht erfüllt sein. Ein zweites Kriterium überprüft, ob die Ergebnisse des jeweiligen Algorithmus grundsätzlich zur Lösung der logistischen Fragestellung genutzt werden kann. Es ist anzumerken, dass die Ergebnisse eines Graph Mining Algorithmus erst interpretiert werden müssen, bevor sie Wissen darstellen (vgl. Abschnitt 3.3.2). Sollte ein Algorithmus eine Fragestellung lösen, ist dies von besonderem Interesse. Für das dritte Kriterium wird die in Abschnitt 4.1.2 beschriebene Möglichkeit miteinbezogen, dass die Algorithmen des Graph Mining und des Graph Management gemeinsam für eine Lösung einer logistischen Fragestellung genutzt werden können. Es ist erfüllt, wenn der Graph Mining Algorithmus den Graphen eventuell derart vereinfacht, dass er im

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

weiteren Verlauf effizienter von einem Algorithmus des Graph Management bearbeitet werden kann. Das letzte Kriterium beschäftigt sich mit Anwendungsfällen beziehungsweise Use Cases der Algorithmen zur Beantwortung der jeweiligen logistischen Fragestellung. Dabei bezieht sich der Use Case auf das zweite und dritte Kriterium. Ist eines von beiden oder sind beide erfüllt, so ist auch ein passender potenzieller Use Case anzugeben. Dabei wird der Use Case auf Basis der ausgearbeiteten theoretischen Grundlagen in Kapitel 2 sowie Kapitel 3 beschrieben und wird nicht evaluiert. Auch wenn nur ein Use Case angegeben wird, soll dies nicht heißen, dass damit die einzige oder sogar bestmögliche Verwendung des Algorithmus darstellt wird.

Nach der Auffassung von Tarjan (vgl. Abschnitt 3.1) gibt es kein Universalmodell für jeden Algorithmus. Vielmehr muss das abstrahierte Modell der realen Aufgabenstellung angepasst werden, wenn der Algorithmus das reale Problem nicht löst. Aus diesem Grund wird bei der Beschreibung der Use Cases zwar ein rudimentäres Modell mit den Inhalten der Knoten und Kanten angegeben, welches allerdings noch für die einzelnen Algorithmen präzisiert werden kann.

Zusammengefasst lauten die Bewertungskriterien:

1. Eignung des Algorithmus zum Einsatz in der logistischen Fragestellung
2. Lösung der logistischen Fragestellung durch den Algorithmus
3. Vereinfachung des Modells durch den Algorithmus zur Weiterverarbeitung
4. Existenz eines potenziellen Use Cases

Es ist eine Abhängigkeit der Kriterien zu beobachten. Beispielsweise wird Kriterium zwei oder drei nicht erfüllt sein, wenn Kriterium eins nicht auch erfüllt ist. Falls umgekehrt die Kriterien zwei und drei nicht erfüllt sein sollten und folglich auch kein Use Case definiert werden kann, so wird angenommen, dass das erste Kriterium nicht erfüllt sein kann. Sollten das zweite oder dritte Kriterium erfüllt sein, hat das unweigerlich zur Folge, dass auch das vierte Kriterium erfüllt ist. Nach der Definition der Kriterien werden in einem nächsten Schritt die Prüffragen aus den Kriterien hergeleitet. Diese lauten:

1. Ist der Algorithmus generell für den Einsatz in dieser logistischen Fragestellung geeignet?
2. Können die Ergebnisse des Algorithmus zur Lösung der logistischen Fragestellung genutzt werden?
3. Kann der Algorithmus das abstrahierte Modell eventuell so vereinfachen, dass das Ergebnis effizient von einem Algorithmus des Graph Management weiterverarbeitet werden kann?
4. Kann ein potenzieller Use Case für die Verwendung des Algorithmus für diese logistische Fragestellung identifiziert werden?

Nachdem diese Prüffragen aus den Bewertungskriterien hergeleitet wurden, kann die Methode der Checkliste angewendet werden. Aus Platzgründen werden die Prüffragen in der Liste wie folgt abgekürzt:

1. Eignung
2. Lösung
3. Vereinfachung
4. Use Case

4.1.6 Definition von Klassen

In Abschnitt 4.1.5 wurde durch die Definition der Bewertungskriterien der letzte Schritt für die Methode der Checkliste durchgeführt (vgl. Abschnitt 4.1.1). Bei der Eignungsüberprüfung der verschiedenen Algorithmen des Graph Mining zur Beantwortung von logistischen Fragestellungen, lassen sich die Ergebnisse der Methode je nach Erfüllungsausprägung der verschiedenen Kriterien in vier verschiedene

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

Klassen einteilen. Die unterschiedlichen Klassen lassen sich zusammen mit ihren erfüllten Kriterien *Tabelle 4.2* entnehmen:

Tabelle 4.2: Klassen und deren Kriterien zur Einteilung von Algorithmen des Graph Mining

Klasse	Bezeichnung	Eignung	Lösung	Vereinfachung	Use Case
1	Zur Lösung und Vereinfachung geeignet	✓	✓	✓	✓
2	Nur zur Lösung geeignet	✓	✓	✗	✓
3	Nur zur Vereinfachung geeignet	✓	✗	✓	✓
4	Ungeeignet	✗	✗	✗	✗

Die Algorithmen der ersten Klasse sind sowohl für die Lösung als auch für die Vereinfachung der logistischen Fragestellung geeignet. Für beide Optionen kann ein passender Use Case genannt werden. In die zweite Klasse werden die Algorithmen eingeteilt, dessen Ergebnisse nur zur Lösung der logistischen Fragestellung und nicht zur Vereinfachung geeignet sind. Ein Anwendungsfall für die Lösung kann genannt werden. Die Algorithmen der dritten Klasse eignen sich nur zur Vereinfachung der logistischen Fragestellung. Analog zu den bereits genannten Klassen, kann hier ein geeigneter Use Case beschrieben werden. Die ersten drei Klassen haben gemeinsam, dass sie das erste Kriterium erfüllen, also für die Beantwortung der jeweiligen logistischen Fragestellung geeignet sind. Anders sieht das bei der vierten Klasse aus. Die dort eingeteilten Algorithmen besitzen keine grundsätzliche Eignung für die Beantwortung der jeweiligen logistischen Fragestellung. Es wurden keine Möglichkeiten gefunden, die Ergebnisse der Algorithmen zur Lösung oder Vereinfachung der logistischen Fragestellung zu nutzen. Folglich kann es auch keinen Anwendungsfall geben. Aus Gründen der Lesbarkeit werden den Algorithmen Abkürzungen zugewiesen, die der *Tabelle 4.3* zu entnehmen sind.

Tabelle 4.3: Abkürzungen der Graph Mining Algorithmen

Algorithmus	Abkürzung
Graph Pattern Mining	GPM
Node-Clustering	N-Clu
Graph-Clustering	G-Clu
Label-Propagation	L-Pro
Graph-Classification	G-Clu
Link Prediction	LP
Graph Compression	GC

4.2 Klassifizierung von Graph Mining Algorithmen

In den folgenden Abschnitten wird die Eignung der verschiedenen Graph Mining Algorithmen (vgl. Abschnitt 3.3.2 und 4.1.3) zur Beantwortung der in Abschnitt 4.1.4 definierten logistischen Fragestellungen überprüft. Je nach Ausprägung der in Abschnitt 4.1.5 definierten Bewertungskriterien kann ein Algorithmus einer der vier Klassen (vgl. Abschnitt 4.1.6) zugeteilt werden. Am Ende die jeweiligen Klassifizierungen der Abschnitte 4.2.1 bis 4.2.8 zusammengetragen und analysiert (vgl. Abschnitt 4.2.9). Es sei an dieser Stelle angemerkt, dass die Eignungsüberprüfungen und die Definition simpler Anwendungsfälle auf den theoretischen Grundlagen aus Kapitel 2 und Kapitel 3 basieren und der Übersichtlichkeit halber nicht für jede Aussage darauf verwiesen wird.

4.2.1 Strategische Netzwerkgestaltung

Für die strategische Netzwerkgestaltung werden in diesem Abschnitt zwei Fragen thematisiert, wovon die erste wie folgt lautet: „Wo soll ein neues Werk für Produkt X errichtet werden?“. Die Knoten des zugrundeliegenden Netzwerkmodells können die potenziellen Standorte des neuen Werkes oder die Händler darstellen, die das in den Werken zu produzierende Produkt abnehmen. Die Entfernungen zwischen diesen Orten werden als Kanten repräsentiert.

Durch die Entdeckung von häufigen Subgraphen kann der Algorithmus des Graph Pattern Mining die Fragestellung sowohl lösen als auch vereinfachen. Für die Lösung wird zuerst eine Menge an potenziellen Standorten (Knoten) identifiziert. Diese werden durch Kanten, die die Entfernungen darstellen, mit den bestehenden Händlern (Knoten) verbunden. Allerdings werden nur jene Standorte und Händler verbunden, deren Entfernungen unter 70 Kilometern liegen. Für den Fall, dass ein Werk mindestens 30 Händler versorgen soll, können durch den Algorithmus jene Sterngraphen identifiziert werden, die diese Bedingung erfüllen. Dieses Ergebnis kann nicht nur zur Lösung beitragen, sondern auch zur Vereinfachung, da die identifizierten Sterngraphen über den Algorithmus des Graph Matching mit Mustern verglichen werden können, die zu bestehenden Werken gehören. Liegt eine ähnliche Struktur zwischen den beiden Mustern vor, so ist es naheliegend, dass der potenzielle Standort für den Bau eines neuen Werkes geeignet ist.

Für das Node-Clustering werden im Modell keine potenziellen Standorte, sondern nur die Händler (Knoten) betrachtet werden, die das Produkt abnehmen. Die Entfernungen der Händler zueinander werden durch Kanten repräsentiert. Durch die Minimierung der Zielfunktion werden nah beieinanderliegende Händler geclustert und abhängig von der Anzahl der in den Clustern befindlichen Händler werden Standorte für Werke identifiziert, die sich durch Kundennähe auszeichnen und somit einen schnellen Transport gewährleisten. Das Ergebnis des Node-Clustering kann zusätzlich durch die Analytical Queries des Graph Management analysiert werden, um die Händler in den einzelnen Clustern zu analysieren und somit die Wahl zwischen den Clustern einzugrenzen.

Da die Größe der Graphen, die für den Graph-Clustering Algorithmus eingesetzt werden können, dürften die Graphen nach dem vorgegebenen Netzwerkmodell zu umfangreich sein. Aus diesem Grund ist das erste Kriterium der Eignung nicht erfüllt und es können folglich keine Use Cases identifiziert werden, deren Ergebnisse zur Lösung oder zur Weiterverarbeitung von Graph Management Algorithmen genutzt werden.

Für den Algorithmus der Label-Propagation ist anzunehmen, dass sich die Knoten des Graphen einerseits aus bereits bestehenden Werken (gelabelte Knoten) und andererseits aus potenziellen Standorten für ein Werk (ungelabelte Knoten) zusammensetzen. Diese Knoten beinhalten durch deren Knotengewichte Informationen zu ihrer Lage und den umgebenen Händlern. Da das Unternehmen, welches ein neues Werk errichten möchte, Kenntnisse über die Effizienz der bereits bestehenden Werke besitzt, können die ungelabelten Knoten, die durch den Algorithmus der Label-Propagation in die Klasse der effizienten Werke fallen, als erfolgsversprechende Standorte gesehen werden. Somit kann der Algorithmus zur Lösung der logistischen Problemstellung beitragen. Ebenfalls können diese Ergebnisse dafür genutzt werden, weiterführende Analysen mit Graph Management Algorithmen durchzuführen. Ist für die Standortwahl beispielsweise ein bestimmter Faktor zu berücksichtigen, könnten die Kantengewichte der potenziellen Standorte dahingehend ausgewertet werden. Es ist also auch das dritte Kriterium erfüllt.

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

Analog zum Graph-Clustering sind die Graphen für den Algorithmus der Graph-Classification in ihrer Größe begrenzt. Da sich das Netzwerkmodell nicht in der geforderten Größe abbilden ließe, ist auch das erste Kriterium nicht erfüllt und der Graph-Classification Algorithmus kann nicht zur Lösung oder zur Weiterverarbeitung mit Verfahren des Graph Management verwendet werden.

Für die Frage nach einem neuen Standort für ein Werk sind für diesen Fall die Eigenschaften von potenziellen Standorten, den Händlern (Knoten) und deren bestehenden Entfernung (Kanten) relevant. Die Entdeckung von neuen, bisher unbekanntem Kanten wird also nicht zielführend erachtet, sodass der Algorithmus der Link Prediction nicht für die logistische Fragestellung geeignet ist und somit auch keine Anwendungsfälle identifiziert werden können.

Der Graph Compression Algorithmus ist für diese Fragestellung geeignet, da es sich bei dem vorgegebenen Modell um einen umfangreichen Graphen handelt. Auch wenn kein Anwendungsfall identifiziert werden konnte, dessen Ergebnisse zur unmittelbaren Lösung des Problems genutzt werden können, so kann ein komprimierter Graph dennoch zur leichteren Analyse und Auswertung durch Graph Management Algorithmen genutzt werden. Für einen gegebenen Graphen aus Händlern (Knoten), die das in den Werken hergestellte Produkt abnehmen, und deren Entfernungen (Kanten) können beispielsweise alle Händler entfernt werden, deren Bedarfe bereits durch ein nahegelegenes Werk befriedigt werden. Folglich ist eine Vereinfachung des Graphen gegeben.

Die Ergebnisse der Eignungsüberprüfung der Algorithmen in Bezug auf die erste Frage zur strategischen Netzwerkgestaltung lassen sich in **Tabelle 4.4** zusammenfassen.

Tabelle 4.4: Kriterienausprägungen der Algorithmen für die erste Frage der strategischen Netzwerkgestaltung

Algorithmus	Eignung	Lösung	Vereinfachung	Use Case
Graph Pattern Mining	✓	✓	✓	✓
Node-Clustering	✓	✓	✓	✓
Graph-Clustering	✗	✗	✗	✗
Label-Propagation	✓	✓	✓	✓
Graph-Classification	✗	✗	✗	✗
Link Prediction	✗	✗	✗	✗
Graph Compression	✓	✗	✓	✓

Die Einteilung der untersuchten Algorithmen des Graph Mining in die Klassen kann für die erste Fragestellung der strategischen Netzwerkgestaltung **Tabelle 4.5** entnommen werden.

Tabelle 4.5: Klassifizierung der Algorithmen in Bezug auf die erste Frage zur strategischen Netzwerkgestaltung

Fragestellung	Klasse 1	Klasse 2	Klasse 3	Klasse 4
Wo soll ein neues Werk für Produkt X errichtet werden?	GPM, N-Clu, L-Pro		GC	G-Clu, G-Cla, LP

Die zweite Frage der strategischen Netzwerkgestaltung lautet: „Wie viele Standorte zur Distribution sind zu errichten?“. In einem zugehörigen Graph können die Knoten die Rollen von bestehenden oder potenziellen Distributionsstandorten oder vorhandenen Händlern annehmen, zu denen die Warenströme laufen. Die Kanten stehen etwa für die Entfernungen oder die Objektflüsse zwischen Distributionsstandort und Abnehmer.

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

Der Algorithmus des Graph Pattern Mining ist grundsätzlich zur Anwendung auf diese Problemstellung geeignet. Jedoch konnte kein Use Case identifiziert werden, bei dem die Ergebnisse nach einer Interpretation ohne eine weitere Bearbeitung zur Lösung des Problems verwendet werden können. Dementgegen ist es möglich, das Graph Pattern Mining in Verbindung mit den Analytical Queries des Graph Management zielführend zu nutzen. In dem entsprechenden Use Case verbinden die Kanten als Entfernungen die Distributionsstandorte mit den vorhandenen Händlern, wobei nur solche Händler berücksichtigt werden, die in einem gewissen Umkreis des Distributionsstandortes liegen. Erneut entsteht ein Sterngraph, dessen zentraler Knoten ein potenzieller oder bereits vorhandener Distributionsstandort ist. Von den identifizierten Mustern werden aus wirtschaftlichen Gründen nur solche Distributionsstandorte weiter untersucht, die eine gewisse Mindestanzahl an Händlern beliefern können. Bei der anschließenden Analyse ist es wichtig darauf zu achten, dass ein Händler nicht von mehr als zwei Distributionsstandorten beliefert wird. Ist diese Bedingung erfüllt, lässt sich daraus die gesuchte Anzahl an Distributionsstandorten ableiten.

Bei dem Verfahren des Node-Clustering wird von einem Modell ausgegangen, dessen Knoten für die Händler stehen und die Kanten für deren Entfernungen zueinander. Der Algorithmus teilt die Knoten abhängig von der Zielfunktion in Cluster ein, die im Falle von Ähnlichkeitswerten jeweils Händler beinhalten, die in einem Gebiet liegen. Aus der Anzahl der Cluster ließe sich die Anzahl der Distributionsstandorte ableiten und das zweite Kriterium gilt als erfüllt. Ein Use Case, bei dem der Algorithmus den Graphen derart vereinfacht, dass er danach von Verfahren des Graph Management weiterverarbeitet werden kann, wurde nicht identifiziert.

Da für die Bestimmung der Anzahl von Distributionsstandorten eine Betrachtung des gesamten Netzes erforderlich erscheint und dieses zu umfangreich für die Verarbeitung durch Graph-Clustering Algorithmen ist, gilt das erste Kriterium der Eignung als nicht erfüllt. Deshalb kann dieses Verfahren nicht für das Problem genutzt werden.

Die Ergebnisse des Algorithmus der Label Propagation können entweder nach einer Interpretation oder aber durch die zusätzliche Anwendung von Graph Management Algorithmen die logistische Fragestellung beantworten. Für den ersten Fall werden ungelabelte potenzielle Standorte auf Basis der bereits bestehenden und als wirtschaftlich erwiesenen (gelabelt) Distributionsstandorten klassifiziert. Werden die potenziellen Standorte bereits so gesetzt werden, dass deren verbundene Händler möglichst noch nicht durch bestehende Standorte versorgt werden, so kann daraus die Anzahl der Standorte bestimmt werden. Falls sich bei der ursprünglichen Auswahl der potenziellen Standorte nicht derartige Gedanken gemacht worden sind, ließe sich durch den zusätzlichen Einsatz von Analytical Queries ebenfalls die Anzahl der zu errichtenden Distributionsstandorte bestimmen. Es sind also alle Kriterien für das Verfahren der Label Propagation erfüllt.

Aufgrund der Tatsache, dass die Graphen für die Algorithmen der Graph-Classification nicht zu groß sein sollten, kommt dieses Verfahren nicht für die Beantwortung dieser Fragestellung infrage. Ein Modell mit potenziellen sowie bereits bestehenden Distributionsstandorten und Händlern als Knoten und deren Verbindungen als Kanten wäre zu umfangreich. Eine Unterteilung dieses Graphen in mehrere Subgraphen würde nicht weiterhelfen, da eine Gewährleistung der Netzabdeckung durch die Distributionsstandorte schwierig zu realisieren wäre.

Das Verfahren der Link Prediction ist zur Bestimmung der Anzahl von Distributionsstandorten nicht geeignet, da die Distributionsstandorte bereits mit den umliegenden Händlern verbunden sind und es in

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

diesem Fall sinnlos wäre, mehrere Händler untereinander durch Kanten zu verbinden. Demnach können auch keine Use Cases zur Lösung oder zur Vereinfachung identifiziert werden.

Grundsätzlich kann der Algorithmus der Graph Compression auf das vorgegebene Modell bestehend aus Distributionsstandorten und Händlern angewandt werden. Allerdings konnte kein Use Case ermittelt werden, bei dem das interpretierte Ergebnis zu einer Lösung des Problems führt. Werden aus dem komplexen Graphen die Kanten zwischen den bestehenden Distributionsstandorten und den zugehörigen Händlern im direkten Umfeld entfernt, bleibt ein Graph übrig, dessen Knoten sich auf potenzielle Distributionsstandorte und die restlichen Händler beschränken. Der komprimierte Graph kann anschließend mithilfe von Graph Management Algorithmen zur Bestimmung der nötigen Distributionsstandorte genutzt werden.

Die Ergebnisse der Eignungsüberprüfung der Algorithmen in Bezug auf die zweite Frage zur strategischen Netzwerkgestaltung lassen sich in **Tabelle 4.6** zusammenfassen.

Tabelle 4.6: Kriterienausprägungen der Algorithmen für die zweite Frage der strategischen Netzwerkgestaltung

Algorithmus	Eignung	Lösung	Vereinfachung	Use Case
Graph Pattern Mining	✓	✗	✓	✓
Node-Clustering	✓	✓	✗	✓
Graph-Clustering	✗	✗	✗	✗
Label-Propagation	✓	✓	✓	✓
Graph-Classification	✗	✗	✗	✗
Link Prediction	✗	✗	✗	✗
Graph Compression	✓	✗	✓	✓

Die Einteilung der untersuchten Algorithmen des Graph Mining in die Klassen kann für die zweite Fragestellung der strategischen Netzwerkgestaltung **Tabelle 4.7** entnommen werden.

Tabelle 4.7: Klassifizierung der Algorithmen in Bezug auf die zweite Frage zur strategischen Netzwerkgestaltung

Fragestellung	Klasse 1	Klasse 2	Klasse 3	Klasse 4
Wie viele Standorte zur Distribution sind zu errichten?	L-Pro	N-Clu	GPM, GC	G-Clu, G-Cla, LP

4.2.2 Bedarfsplanung

Die Frage, die sich auf das Teilgebiet der Bedarfsplanung bezieht, lautet: „Wie wird sich der Bedarf an Produkt X das kommende Quartal in Region Y entwickeln?“. Als ein konkretes Anwendungsgebiet wird eine Region betrachtet, in der zahlreiche Autohändler (Knoten). Es wird davon ausgegangen, dass sich der Bedarf an einem bestimmten Modell aus den angefragten Probefahrten mit diesem Modell ableiten lässt. Aus diesem Grund ist zu jedem Händler die Anzahl der angefragten Probefahrten bekannt und darüber hinaus, wie sich diese zum letzten Betrachtungszeitpunkt verändert haben. Die Standorte der Autohändler sind durch Kanten verbunden, dessen Werte für die Entfernung stehen. Für die folgenden Use Cases wird meist davon ausgegangen, dass sich eine Änderung des Bedarfs zuerst bei den größeren Händlern bemerkbar macht und diese Änderung bei den kleineren Händlern mit einer gewissen Verzögerung ankommt.

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

Der Algorithmus des Graph Pattern Mining lässt sich zur Prognose, wie sich der Bedarf entwickeln wird, sowohl eigenständig als auch in Verbindung mit Graph Management Algorithmen verwenden. Die Ergebnisse könnten sich beispielsweise nach einer Interpretation direkt für die Beantwortung verwenden lassen, wenn in dem Graph mehrere Autohändler identifiziert werden, bei denen der Bedarfswert entweder zu- oder abnimmt. Auch wenn so keine genaue Zahl prognostiziert werden könnte, ließe sich eine Tendenz bestimmen, was auch eine Lösung der Fragestellung darstellt. Um diese identifizierten Muster mit Vergangenheitsdaten zu vergleichen, könnten sie mithilfe eines Graph Matching Algorithmus mit vergangenen Mustern verglichen werden. Es könnte etwa die Regelmäßigkeit identifiziert werden, dass der Bedarf nach einem Fahrzeug im letzten Quartal eines jeden Jahres ansteigt.

Das Node Clustering kann zur Lösung der logistischen Problemstellung eingesetzt werden. Wenn durch das Clustering deutlich wird, dass sich in einem Bereich viele Händler befinden, deren Anfragen nach Probefahrten allesamt zunehmen, so ist für dieses Beispiel auch mit einem allgemeinen Bedarfsanstieg in dieser Region zu rechnen. Ein Anwendungsfall zur weiterführenden Verarbeitung mit Graph Management Algorithmen kann ausgemacht werden.

Durch die überschaubare Größe der vorhandenen Graphen ist der Algorithmus des Graph-Clustering zwar zum Einsatz geeignet, doch es wurden keine Use Cases identifiziert, bei denen die Ergebnisse zur Problemlösung beitragen könnten. Wird der Graph der Region in mehrere Abschnitte unterteilt, könnten durch das Graph-Clustering etwa ähnliche Autohändler gefunden werden. Allerdings unterstützen diese Informationen nicht bei der Prognose der Bedarfsentwicklung.

Für das Verfahren der Label-Propagation wurde kein Use Case identifiziert, bei dem die Ergebnisse unmittelbar zur Prognose der Bedarfsentwicklung eingesetzt werden könnte, wohingegen das dritte Kriterium erfüllt ist. Wenn es bekannt sein sollte, dass sich eine Zunahme des Bedarfs bei manchen Autohäusern (gelabelte Knoten) früher bemerkbar macht als bei anderen, könnten durch die Klassifizierung der ungelabelten Knoten ebenfalls Händler identifiziert werden, die eine gewisse Vorreiterrolle innehaben. Diese Informationen können nach einer zusätzlichen Bearbeitung durch Algorithmen des Graph Management zur Beantwortung des logistischen Problems gebraucht werden.

Die Eignung für den Einsatz von dem Graph-Classification ist durch die Tatsache gegeben, dass der Graph mit den Händlern als Knoten nicht allzu groß ist. Der Use Case, mit dem das zweite Kriterium erfüllt wird, geht ebenfalls davon aus, dass sich die Änderung der Bedarfswerte zuerst bei den umsatzstärkeren Autohäusern bekannt macht und erst später bei den kleineren. Der Graph wird weiterhin in mehrere Subgraphen unterteilt, wobei jeder Subgraph ein umsatzstarkes Autohaus enthält und den Input für den Algorithmus darstellt. Aus Vergangenheitswerten sind unterschiedliche Muster für verschiedene Stufen der Bedarfsentwicklung bekannt. So entsteht in einem Bereich ein neuer Bedarf, wenn nur die umsatzstarken Autohäuser einen Anstieg vermerken, die umliegenden kleineren jedoch noch nicht. Bei einem gleichbleibenden Bedarf über einen längeren Zeitraum, besitzen alle Autohäuser beispielsweise einen ähnlichen Bedarfswert. Der Algorithmus der Graph-Classification teilt die Subgraphen in diese Klassen ein und nach einer Interpretation der Ergebnisse kann von der Zuteilung der Subgraphen auf die Entwicklung des Bedarfes der kompletten Region geschlossen werden. Ein Anwendungsfall, bei dem die Ergebnisse nach einer weiterführenden Bearbeitung durch Graph Management Algorithmen zu einer Lösung führen, wurde nicht identifiziert.

Für das Verfahren der Link Prediction wird das eingangs vorgestellte Modell etwas abgewandelt. Die Kanten repräsentieren nun nicht mehr die Entfernungen zwischen den Händlern, sondern stehen für

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

einen speziellen Abhängigkeitswert. Dieser leitet sich aus den Umsätzen der beiden Händler ab und soll verdeutlichen, wie sehr sich die beiden Händler gegenseitig beeinflussen. Durch Veränderungen am Markt kann auch dieser Wert variieren und mit dem Algorithmus der Link Prediction können neue Abhängigkeiten aufgedeckt werden. Da diese Abhängigkeiten nicht unmittelbar zur Problemlösung genutzt werden können, stehen zusätzliche Analysen durch Algorithmen des Graph Management aus. Somit sind alle Kriterien bis auf das zweite erfüllt.

Durch das Verfahren der Graph Compression lässt sich der Graph mit den Autohändlern zwar komprimieren, jedoch wurde kein Use Case gefunden, dessen Ergebnisse zur Lösung der Problemstellung beitragen. Für den Fall, dass die Knoten der kleinen Händler mit niedrigen Bedarfen entfernt werden, wird die Größe des Modells auf die relevanten Händler der betrachteten Region reduziert. Dadurch werden Änderungen des Bedarfes schneller sichtbar und die Ergebnisse können zusammen mit Analytical Queries zur Problemlösung genutzt werden.

Die Ergebnisse der Eignungsüberprüfung der Algorithmen in Bezug auf die Frage zur Bedarfsplanung lassen sich in **Tabelle 4.8** zusammenfassen.

Tabelle 4.8: Kriterienausprägungen der Algorithmen für die Frage der Bedarfsplanung

Algorithmus	Eignung	Lösung	Vereinfachung	Use Case
Graph Pattern Mining	✓	✓	✓	✓
Node-Clustering	✓	✓	✗	✓
Graph-Clustering	✗	✗	✗	✗
Label-Propagation	✓	✗	✓	✓
Graph-Classification	✓	✓	✗	✓
Link Prediction	✓	✗	✓	✓
Graph Compression	✓	✗	✓	✓

Die Einteilung der untersuchten Algorithmen des Graph Mining in die Klassen kann für die Fragestellung der Bedarfsplanung **Tabelle 4.9** entnommen werden.

Tabelle 4.9: Klassifizierung der Algorithmen in Bezug auf die Frage der Bedarfsplanung

Fragestellung	Klasse 1	Klasse 2	Klasse 3	Klasse 4
Wie wird sich der Bedarf an Produkt X das kommende Quartal in Region Y entwickeln?	GPM	N-Clu, G-Cla,	L-Pro, LP, GC	G-Clu

4.2.3 Netzwerkplanung

Für das Teilgebiet der Netzwerkplanung lautet die identifizierte logistische Frage: „Welches Werk muss wie viele Produkte eines Typs produzieren, damit der gesamte Bedarf bestmöglich abgedeckt wird?“. Ihr liegt ein Graph zugrunde, dessen Knoten für die Kunden, die einen bestimmten Bedarf an einem Produkt äußern, oder zusätzlich dazu für die produzierenden Werke stehen können. Die Verbindung zwischen den Knoten kann entweder den Wert einer Entfernung oder eines Objektflusses annehmen.

Grundsätzlich ist das Verfahren des Graph Pattern Mining für die logistische Fragestellung geeignet. Ein Use Case, bei dem die Ergebnisse nach einer Interpretation zur unmittelbaren Problemlösung genutzt werden können, wurde nicht identifiziert. Dafür ließen sich auch einem Graphen, bei dem die

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

Knoten für die Werke sowie die Kunden mit ihren Bedarfen stehen und diese Parteien durch Objektflüsse verbunden werden, die ausgehend von den Werken zu den Kunden führen. Die durch den Algorithmus extrahierten Bedarfsmuster können eventuell nach einer weitergehenden Analyse durch Graph Management Algorithmen zur Lösung des Problems genutzt werden. Somit sind alle bis auf das zweite Kriterium erfüllt.

Durch den Algorithmus des Node-Clustering können die Knoten, die die Kunden repräsentieren, aufgrund ihrer Entfernung geclustert werden. Diese geclusterten Bedarfsknoten können durch zusätzliche Algorithmen analysiert werden, wodurch Regionen bestimmt werden, in denen beispielsweise besonders hohe Bedarfe zu erwarten sind. Ein Werk in diesem Cluster müsste folglich mehr produzieren als ein Werk, in dessen Umgebung nicht so viele Kundenstandorte liegen. Allerdings konnte für das Node-Clustering kein Fall gefunden werden, für den das zweite Kriterium erfüllt wäre.

Aufgrund der Komplexität des Graphen und der Tatsache, dass das Geflecht aus Werken und verbundenen Kunden nur schlecht unterteilt werden kann, ist der Algorithmus des Graph-Clustering nicht für das vorgegebene Modell dieser logistischen Fragestellung geeignet. Es werden keine Kriterien erfüllt.

Die Label-Propagation eignet sich für den großen Graphen, dessen Knoten sich aus Werken und Händlern mit ihren Bedarfen zusammensetzen. Die Kanten stehen in diesem Fall für die Entfernungen der Knoten zueinander. Für dieses Verfahren wird angenommen, dass der Umfang an herzustellenden Produkten, der zur Bedarfsbefriedigung der umliegenden Kunden nötig ist, bereits für einige Werke bekannt ist (gelabelte Knoten). Daraus resultieren Klassen, in die die ungelabelten Werke eingeteilt werden können. Diese Zuteilung könnte zur Beantwortung des Problems verwendet werden. Sollte dieses Ergebnis jedoch nicht unmittelbar zur Lösung der logistischen Fragestellung eingesetzt werden können, ließen sich darüber hinaus die Verfahren des Graph Management nutzen. Somit sind alle Kriterien erfüllt.

Der Algorithmus der Graph-Classification eignet sich nicht unbedingt für den Einsatz bei dem vorliegenden Modell, da der Graph zu umfangreich ist. Eine sinnvolle Aufteilung des Graphs ist ebenfalls sehr schwierig, da sich das komplexe Geflecht aus den Knoten kaum sinnvoll aufteilen ließe. Aus diesem Grund sind keine der Kriterien erfüllt.

Auch das Verfahren der Link Prediction scheint ungeeignet zu sein, da es sich bei der Fragestellung nicht um die Verlinkungen der Knoten, sondern die Bedarfswerte der Knoten an sich geht. Deshalb können auch für die Link Prediction keine Use Cases identifiziert werden und die Kriterien sind folglich nicht erfüllt.

Die Graph Compression ist für den Einsatz bei dieser logistischen Fragestellung zwar geeignet, doch es konnte kein Anwendungsfall gefunden werden, bei dem der Algorithmus zur unmittelbaren Lösung verwendet wird. Dafür kann der umfangreiche Graph zur weiteren Verarbeitung durch Graph Management Algorithmen vereinfacht werden, indem Kunden, die lediglich von einem Werk beliefert werden, zusammengefasst werden. Selbstverständlich werden die einzelnen Bedarfe ebenfalls addiert, sodass der Graph im Endeffekt die gleichen Informationen enthält und trotzdem an Komplexität verloren hat.

Die Ergebnisse der Eignungsüberprüfung der Algorithmen in Bezug auf die Frage zur Netzwerkplanung lassen sich in *Tabelle 4.10* zusammenfassen.

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

Tabelle 4.10: Kriterienausprägungen der Algorithmen für die Frage der Netzwerkplanung

Algorithmus	Eignung	Lösung	Vereinfachung	Use Case
Graph Pattern Mining	✓	✗	✓	✓
Node-Clustering	✓	✗	✓	✓
Graph-Clustering	✗	✗	✗	✗
Label-Propagation	✓	✓	✓	✓
Graph-Classification	✗	✗	✗	✗
Link Prediction	✗	✗	✗	✗
Graph Compression	✓	✗	✓	✓

Die Einteilung der untersuchten Algorithmen des Graph Mining in die Klassen kann für die Fragestellung der Netzwerkplanung *Tabelle 4.11* entnommen werden.

Tabelle 4.11: Klassifizierung der Algorithmen in Bezug auf die Frage der Netzwerkplanung

Fragestellung	Klasse 1	Klasse 2	Klasse 3	Klasse 4
Welches Werk muss wie viele Produkte eines Typs produzieren, damit der gesamte Bedarf bestmöglich abgedeckt wird?	L-Pro		GPM, N-Clu, GC	G-Clu, G-Cla, LP

4.2.4 Beschaffungsplanung

Für das Teilgebiet der Beschaffungsplanung wurde die folgende logistische Fragestellung definiert: „Für welche Werke sind welche Anlieferhythmen zu wählen, damit die Materialbereitstellung sichergestellt wird?“. Als Modell wird ein Graph vorgegeben, dessen Knoten für die Werke sowie einen Wert für die Materialbedarfe inklusive der Sicherheitsbestände steht und die Entfernungen zwischen den Graphen durch die Kanten berücksichtigt werden.

Für den Algorithmus des Graph Pattern Mining wurde kein Fall identifiziert, bei dem die Ergebnisse ohne weitere Bearbeitung durch andere Algorithmen zur Lösung der logistischen Fragestellung genutzt werden kann. Jedoch könnten durch das Graph Pattern Mining ähnliche Bedarfsmuster der Werke aufgedeckt werden und durch diese ließen sich nach einer Analyse mit Graph Management Algorithmen nützliche Erkenntnisse erlangen, die zur Problemlösung verwendet werden können. Somit sind alle Kriterien bis auf das zweite erfüllt.

Das Node-Clustering ist grundsätzlich für den Einsatz im vorgegebenen Graphen geeignet. Mit dem Algorithmus lassen sich Werke mit ähnlichen Materialbedarfen clustern. In einem nächsten Schritt können mit zusätzlichen Algorithmen des Graph Management die Anlieferhythmen für die unterschiedlichen Klassen bestimmt werden. Dieser Anwendungsfall stellt eine Vereinfachung dar, da die Anlieferhythmen nicht mehr einzeln für alle Werke bestimmt werden müssen. Ohne eine Verwendung von weiteren Algorithmen ist die das Node-Clustering nicht zur Beantwortung der Fragestellung geeignet, sodass alle Kriterien bis auf das zweite erfüllt sind.

Auch bei dem Verfahren der Graph-Classification ist die Problemstellung nur durch die Verwendung von weiteren Algorithmen des Graph Management lösbar. Um mehrere Graphen zu erzeugen und die Komplexität gleichzeitig zu reduzieren, wird der große Graph in die einzelnen Werke unterteilt. Damit

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

die resultierenden Graphen nicht nur aus dem Knoten des jeweiligen Werks bestehen, werden sie durch ihre Zulieferer ergänzt, die für die Anlieferung der für die Produktion benötigten Materialien zuständig sind. Die Kanten stehen in diesem Use Case nicht für die Entfernungen, sondern für die Objektflüsse der angelieferten Materialien. Es wird angenommen, dass die Graphen, die durch ihre Anforderungen und Bedingungen den gleichen Klassen zugeordnet werden, auch ähnliche Anlieferhythmen besitzen müssen. Der alleinige Einsatz des Graph-Clustering führt nicht zur Bestimmung der zugehörigen Anlieferhythmen, sodass alle Kriterien bis auf das zweite erfüllt sind.

Das vorgegebene Modell ist für die Verarbeitung mit Label-Propagation Verfahren sehr gut geeignet und kann auch zur Lösung eingesetzt werden. In einem derartigen Use Case könnten Werke, für die bereits ein optimaler Anlieferhythmus bestimmt wurde (gelabelte Knoten), als Grundlage der Klassifizierung gelten. Werke mit ähnlichen Anforderungen, für die die Bestimmung der Anlieferhythmen noch aussteht (ungelabelte Knoten), werden anschließend diesen Klassen zugeordnet. Durch die Ähnlichkeit zu den bereits gelabelten Knoten könnten eventuell auch die zugehörigen Anlieferhythmen für die ungelabelten Knoten gelten. Eine derartige Vereinfachung, dass der Graph anschließend durch eine tiefgehende Analyse mithilfe von Graph Management Algorithmen zur Problemlösung eingesetzt werden kann, wurde nicht identifiziert. Somit ist lediglich das dritte Kriterium nicht erfüllt.

Ebenfalls zur Lösung ist die Graph-Classification geeignet. Wird der Graph wie auch schon für die Graph-Classification in je einen Graphen für jedes Werk unterteilt, reduziert sich die Komplexität auf ein vertretbares Maß. Die Kanten nehmen erneut den Wert von Objektflüssen an. In diesem Use Case wurden bereits für einige Werke die optimalen Anlieferhythmen berechnet (gelabelte Graphen). Werden die ungelabelten Werke aufgrund ihrer ähnlichen Anforderungen in dieselbe Gruppe mit einem gelabelten Knoten zugeteilt, so ist die Vermutung naheliegend, dass auch für diese zugeteilten Werke ähnliche Anlieferhythmen gelten müssen. Analog zum Algorithmus der Label-Propagation konnte für das Verfahren der Graph-Classification zur Bestimmung geeigneter Anlieferhythmen kein passender Use Case gefunden werden, weshalb alle Kriterien bis auf das dritte erfüllt sind.

Der Link Prediction Algorithmus ist zur Bestimmung der Anlieferhythmen nicht geeignet, da das Entdecken neuer Kanten für diese logistische Fragestellung als nicht zielführend erachtet wird. Vielmehr geht es dabei um die Werke mit ihren Bedarfen und somit ist keines der Kriterien erfüllt.

Da für jedes einzelne Werk des Netzes der Anlieferhythmus bestimmt werden soll, macht es keinen Sinn, mehrere Werke zusammenzufassen oder den Graphen auf eine andere Art zu komprimieren. Aus diesem Grund ist der Algorithmus der Graph Compression weder zur unmittelbaren Lösung noch zur zusätzlichen Verwendung mit Verfahren des Graph Management geeignet.

Die Ergebnisse der Eignungsüberprüfung der Algorithmen in Bezug auf die Frage zur Beschaffungsplanung lassen sich in **Tabelle 4.12** zusammenfassen.

Tabelle 4.12: Kriterienausprägungen der Algorithmen für die Frage der Beschaffungsplanung

Algorithmus	Eignung	Lösung	Vereinfachung	Use Case
Graph Pattern Mining	✓	✗	✓	✓
Node-Clustering	✓	✗	✓	✓
Graph-Clustering	✓	✗	✓	✓
Label-Propagation	✓	✓	✗	✓
Graph-Classification	✓	✓	✗	✓
Link Prediction	✗	✗	✗	✗

Graph Compression	X	X	X	X
-------------------	---	---	---	---

Die Einteilung der untersuchten Algorithmen des Graph Mining in die Klassen kann für die Fragestellung der Beschaffungsplanung *Tabelle 4.13* entnommen werden.

Tabelle 4.13: Klassifizierung der Algorithmen in Bezug auf die Frage der Beschaffungsplanung

Fragestellung	Klasse 1	Klasse 2	Klasse 3	Klasse 4
Für welche Werke sind welche Anlieferhythmen zu wählen, damit die Materialbereitstellung sichergestellt wird?		L-Pro, G-Cla	GPM, N-Clu, G-Clu	LP, GC

4.2.5 Distributionsplanung

Die Frage, die sich auf das Teilgebiet der Distributionsplanung bezieht, lautet: „Auf welchen Verkehrsträger fällt die Wahl vor dem Hintergrund der Transportkostenminimierung?“. In dem zugrundeliegenden Modell werden dabei nicht zwingend die verschiedenen Verkehrsträger berücksichtigt. Vielmehr stehen die Knoten für Angebots- und Bedarfsknoten, deren verbindende Kanten für die Entfernungen oder Objektflüsse stehen.

Das Verfahren des Graph Pattern Mining ist für diesen Anwendungsfall nicht zur unmittelbaren Lösung der Problemstellung geeignet. Da der Algorithmus jedoch grundsätzlich auf den Graphen angewandt werden kann, könnten die extrahierten Muster beispielsweise nach einer Bearbeitung durch Analytical Queries zur Beantwortung der Fragestellung genutzt werden.

Für das Node-Clustering wird angenommen, dass der Graph nur aus Bedarfsknoten besteht, deren Verbindungen die Entfernungen darstellen (Kanten). Die Standorte werden derart geclustert, dass sich dicht beieinander gelegene Bedarfsknoten im gleichen Cluster befinden. Beispielsweise könnte durch eine anschließende Untersuchung der Cluster mithilfe von Analytical Queries der durchschnittliche Abstand der Knoten zueinander bestimmt werden. Daraus ließe sich etwa schlussfolgern, dass für nah beieinander liegende Kunden eine Distribution über einen LKW effizient ist, weil mit wenigen Routen viele Bedarfe befriedigt werden können. Eine weitere Erkenntnis könnte lauten, dass Knoten, die keiner Gruppe zuzuordnen sind und abseits von den anderen Punkten liegen, ihre Produkte am besten durch einen Transporter geliefert bekommen, da dieser Fahrzeugtyp nicht so viel Treibstoff wie ein LKW verbraucht. Eine Option, diesen Algorithmus zur direkten Bestimmung von kostengünstigen Verkehrsträgern zu verwenden, wurde nicht identifiziert. Somit sind bis auf das zweite Kriterium alle erfüllt.

Ein Algorithmus, dessen Ergebnisse nach einer Interpretation zur Lösung der logistischen Fragestellung verwendet werden können, ist das Graph-Clustering. Dabei setzen sich die zu clusternden Graphen aus einem Graphen zusammen, der ein Netz aus Angebots- und Bedarfsknoten darstellt, und mehreren anderen, die unterschiedliche Verkehrsträger beschreiben. Für alle Verkehrsträger, die sich mit dem Modell aus Angebots- und Bedarfsknoten in einem Cluster befinden, kann in diesem Use Case davon ausgegangen werden, dass sie einen kostengünstigen Transport gewährleisten. Ein Anwendungsbeispiel für eine zielführende Verwendung des Graph-Clustering Algorithmus zusammen mit Graph Management Verfahren konnte nicht bestimmt werden.

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

Bei dem Verfahren der Label-Propagation wird basierend auf einzelnen gelabelten Knoten ein Modell gelernt, das zur Klassifizierung der restlichen ungelabelten Knoten genutzt werden kann. Es konnten keine Use Cases bestimmt werden, bei denen die Ergebnisse ohne oder mit einer zusätzlichen Bearbeitung durch weitere Algorithmen zur Problemlösung genutzt werden können. Somit sind keine Kriterien erfüllt.

Damit der Algorithmus der Graph-Classification Anwendung finden kann, muss der umfangreiche Graph zuerst in mehrere Subgraphen unterteilt werden. Eine Art der Aufteilung ist die Betrachtung von einem Werk (Angebotsknoten) und den Kunden (Bedarfsknoten) zusammen mit deren Verbindungen als Entfernungen. Konnte bereits für einige der Subgraphen der passende Verkehrsträger (gelabelter Graph) bestimmt werden, wird basierend darauf ein Modell erstellt, das andere noch nicht gelabelte Graphen aufgrund von Ähnlichkeiten in der Struktur einem gewissen Cluster zuordnet. Es wird angenommen, dass für die Werke innerhalb eines Clusters ähnliche Verkehrsträger für den kostengünstigen Transport genutzt werden können. Diese Information kann zur Lösung der logistischen Problemstellung genutzt werden. Auch wenn dadurch das zweite Kriterium erfüllt ist, wurde kein Use Case gefunden, bei dem die Ergebnisse der Graph-Classification von zusätzlichen Algorithmen zielführend weiterverarbeitet werden können.

Das Verfahren der Link Prediction ist nicht für die Frage nach dem besten Verkehrsträger geeignet, da die Auswahl auf den bereits bestehenden Verbindungen beschlossen werden soll. Eine Entdeckung einer neuen Verbindung wäre somit nicht sinnvoll und damit ist keines der Kriterien erfüllt.

Eine Komprimierung des zugrundeliegenden Graphen wäre für eine weiterführende Analyse durch Algorithmen des Graph Management erstrebenswert. Die Kanten stehen dabei nicht für die Entfernungen, sondern für den finanziellen Aufwand, den ein Transport mit sich bringt. Im Use Case könnten Kanten entfernt werden, für die die Transportkosten einen zu hohen Wert annehmen. Damit wurde der Umfang reduziert und das dritte Kriterium gilt als erfüllt. Ein Fall, bei dem die Ergebnisse nach einer Interpretation für die unmittelbare Bestimmung der geeigneten Verkehrsträger genutzt werden könnten, wurde nicht identifiziert.

Die Ergebnisse der Eignungsüberprüfung der Algorithmen in Bezug auf die Frage zur Distributionsplanung lassen sich in **Tabelle 4.14** zusammenfassen.

Tabelle 4.14: Kriterienausprägungen der Algorithmen für die Frage der Distributionsplanung

Algorithmus	Eignung	Lösung	Vereinfachung	Use Case
Graph Pattern Mining	✓	✗	✓	✓
Node-Clustering	✓	✗	✓	✓
Graph-Clustering	✓	✓	✗	✓
Label-Propagation	✗	✗	✗	✗
Graph-Classification	✓	✓	✗	✓
Link Prediction	✗	✗	✗	✗
Graph Compression	✓	✗	✓	✓

Die Einteilung der untersuchten Algorithmen des Graph Mining in die Klassen kann für die Fragestellung der Distributionsplanung **Tabelle 4.15** entnommen werden.

Tabelle 4.15: Klassifizierung der Algorithmen in Bezug auf die Frage der Distributionsplanung

Fragestellung	Klasse 1	Klasse 2	Klasse 3	Klasse 4
Auf welchen Verkehrsträger fällt die Wahl vor dem Hintergrund der Transportkostenminimierung?		G-Clu, G-Cla	GPM, N-Clu, GC	L-Pro, LP

4.2.6 Verfügbarkeits- und Machbarkeitsprüfung

Die Frage „Kann die Lieferung bis zu einem definierten Datum erfolgen?“ ist dem Teilgebiet der Verfügbarkeits- und Machbarkeitsprüfung zuzuordnen. Der zugrundeliegende Graph besteht aus Herstellern und Kunden in Form von Angebots- und Bedarfsknoten. Die Verbindungen sind von den Herstellern zu den Kunden gerichtet und stehen in diesem Fall nicht für die Entfernungen, sondern für die jeweiligen Lieferungen, wobei jede Lieferung eine eigene Kante besitzt. Fragestellungen dieses Teilgebiets werden überwiegend durch Simulationen oder Optimierungen und weniger durch Graphalgorithmen gelöst. Sie haben darüber hinaus nicht so einen direkten Kontakt zur Supply Chain, wie etwa die strategische Netzwerkgestaltung. Nichtsdestotrotz wird das Problem in diesem Abschnitt aufgeführt, um zu verdeutlichen, welche Anforderungen eventuell an die Verfahren des Graph Mining gestellt werden.

Die Entdeckung von häufigen Subgraphen durch den Algorithmus des Graph Pattern Mining bringt keinen Mehrwert, da von potenziellen Mustern keine Rückschlüsse auf den einen Lieferauftrag, den die Frage adressiert, gezogen werden könnten. Es ist folglich kein Kriterium erfüllt.

Beim Verfahren des Node-Clustering werden die Knoten eines Graphen ausgehend von einer vorher bestimmten Zielfunktion in mehrere Cluster unterteilt. Es ist unklar, wie diese Erkenntnisse genutzt werden könnten, das eingangs erwähnt Problem zu lösen. Deshalb ist dieser Algorithmus ungeeignet und alle Kriterien gelten als nicht erfüllt.

Für das Graph-Clustering wird das Modell in mehrere Subgraphen unterteilt, wobei die Subgraphen die Form eines Sterngraphen mit einem Hersteller im Zentrum annehmen. und je nach Anzahl der Lieferungen Kanten zu den umliegenden Kunden besitzen. Durch den Algorithmus des Graph Clustering könnten die Werke nach der Anzahl ihrer zu erfüllenden Lieferungen eingeteilt werden. Sobald ein Werk eine bestimmte Anzahl an Lieferungen überschreitet, könnte aufgrund der bekannten Kapazitäten des Werkes angenommen werden, dass diese Lieferungen als gefährdet gelten. Allerdings wird die logistische Fragestellung bei dieser Annahme nicht zielführend beantwortet werden, weshalb für das Graph Clustering keine Kriterien erfüllt sind.

Der Algorithmus der Label-Propagation erlernt auf Grundlage von bereits gelabelten Angebots- und Bedarfsknoten ein Modell, mit welchem sich noch ungelabelte Knoten klassifizieren ließen. Allerdings wurde für die Frage nach der Machbarkeit einer Lieferung kein Use Case gefunden, bei dem die Ergebnisse des Algorithmus zur Problemlösung genutzt werden könnten. Aus diesem Grund sind die Kriterien nicht erfüllt.

Bei der Graph-Classification werden ungelabelte Graphen klassifiziert, wobei die Graphen eine gewisse Größe nicht übersteigen dürfen. Das Geflecht aus Händlern und Kunden könnte zwar wie für das Graph-Clustering in Sterngraphen unterteilt werden, jedoch wäre eine Klassifizierung unsinnig, da jeder Händler seine eigenen Anforderungen sowie Umgebungseinflüsse hat und die Bedarfe, die an ihn gestellt

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

werden, sehr dynamisch sein können. Zudem könnte eine solche Klassifizierung nicht die Frage nach einer einzelnen Lieferung beantworten. Somit sind keine Kriterien erfüllt.

Der Algorithmus der Link Prediction wird verwendet, um neue Kanten, die derzeit noch fehlen oder in Zukunft erscheinen werden, zu entdecken. Die Anzahl der Lieferungen, die durch Kanten repräsentiert werden, ist vorgegeben und deshalb müssen keine neuen Kanten entdeckt werden. Die Link Prediction ist somit nicht geeignet zur Beantwortung des Problems.

Wie auch schon bei den anderen Algorithmen konnte für die Graph Compression kein Anwendungsfall identifiziert werden, bei dem die Ergebnisse entweder direkt oder nach einer zusätzlichen Analyse durch Graph Management Verfahren zur Problemlösung beitragen können. Aus der Zusammenführung von Knoten oder dem Streichen von Kanten können keine Rückschlüsse auf einzelne Lieferungen gezogen werden.

Diese logistische Fragestellung wurde absichtlich untersucht, obwohl es abzusehen war, dass die Algorithmen des Graph Mining keine sinnvolle Anwendung finden werden. Damit sollte gezeigt werden, wie wichtig die Fragestellung und ein passendes Modell für die Anwendung auf eine reale Problemstellung sind. Daraus ergibt sich unter anderem die Frage, ob und wie ein Anwendungsgebiet zu abstrahieren ist, sodass auch andere logistische Fragestellungen, die keinen direkten Bezug zur Supply Chain haben, beantwortet werden können. Jedoch existieren für viele Fragestellungen eine Auswahl an optimierten und bewährten Verfahren zur effizienten Beantwortung der Frage, die sich nicht durch Graph Mining Verfahren ersetzen ließen.

Die Ergebnisse der Eignungsüberprüfung der Algorithmen in Bezug auf die Frage zur Verfügbarkeits- und Machbarkeitsprüfung lassen sich in **Tabelle 4.16** zusammenfassen.

Tabelle 4.16: Kriterienausprägungen der Algorithmen für die Frage der Verfügbarkeits- und Machbarkeitsprüfung

Algorithmus	Eignung	Lösung	Vereinfachung	Use Case
Graph Pattern Mining	X	X	X	X
Node-Clustering	X	X	X	X
Graph-Clustering	X	X	X	X
Label-Propagation	X	X	X	X
Graph-Classification	X	X	X	X
Link Prediction	X	X	X	X
Graph Compression	X	X	X	X

Die Einteilung der untersuchten Algorithmen des Graph Mining in die Klassen kann für die Fragestellung der Verfügbarkeits- und Machbarkeitsprüfung **Tabelle 4.17** entnommen werden.

Tabelle 4.17: Klassifizierung der Algorithmen in Bezug auf die Frage der Verfügbarkeits- und Machbarkeitsprüfung

Fragestellung	Klasse 1	Klasse 2	Klasse 3	Klasse 4
Kann die Lieferung bis zu einem definierten Datum erfolgen?				GPM, N-Clu, G-Clu, L-Pro, G-Cla, LP, GC

4.2.7 Beschaffungsfineplanung

Für das Teilgebiet der Beschaffungsfineplanung wurde die folgende Frage bestimmt: „Bei welchen Transportfahrzeugen der Zulieferer ist eine termingerechte Anlieferung zu erwarten?“. Für dieses Szenario ist davon auszugehen, dass ein Automobilhersteller täglich eine große Anzahl von Transportfahrzeugen erwartet, damit der Materialbedarf sichergestellt werden kann. Die Frage hat ihren Ursprung in der Problemstellung, dass das Verkehrsaufkommen durch die zahlreichen Transportfahrzeuge in einem Umkreis von mehreren Kilometern um das Werk sehr hoch ist, da sie sich beispielsweise alle denselben Weg zur Anlieferung teilen. Die Knoten des Graphen repräsentieren die Straßenkreuzungen und die Kanten stellen die Straßen dar, die je nach Auslastung durch die Transportfahrzeuge mit einem hohen oder niedrigen Kantengewicht versehen sind.

Der Algorithmus des Graph Pattern Mining ist für den Einsatz bei dieser logistischen Fragestellung, jedoch nicht zur unmittelbaren Lösung ebendieser geeignet. Jedoch könnte das Verfahren unbekannte Anlieferungsmuster wie die hochfrequentierte Nutzung eines Straßenabschnittes identifizieren. Solche Informationen könnten nach einer weitergehenden Analyse zur Bestimmung von pünktlichen Anlieferungen verwendet werden. Das erste, dritte und vierte Kriterium ist deshalb erfüllt.

Beim Node-Clustering können die Straßenkreuzungen (Knoten) mithilfe einer Zielfunktion so geclustert werden, dass nah beieinander liegende Kreuzungen im gleichen Cluster liegen. Wenn auch jede Kreuzung eine Ampel besitzt, kann es durch die hohe Anzahl an Ampeln in einem Gebiet zu Verzögerungen kommen. Damit diese Erkenntnisse die Beantwortung der Fragestellung unterstützen können, wird mit Pattern Matching Queries überprüft, ob ein Transportfahrzeug ein solches Gebiet durchquert. Ein Use Case bei dem der Algorithmus zur Lösung genutzt wird, konnte nicht identifiziert werden.

Damit der Algorithmus des Graph-Clustering angewendet werden kann, wird für jede Lieferung ein eigener Graph generiert, aus dem sich die letzten Kilometer einer Route des Transportfahrzeuges ableiten lassen. Dafür wird den Straßenabschnitten (Kanten), die befahren werden, ein anderer Wert zugewiesen als den unbefahrenen Abschnitten. Um die Komplexität des Graphen gering zu halten, wird beispielsweise nur das Straßennetz in einem Umkreis von wenigen Kilometern um das Werk berücksichtigt. In einem ersten Use Case werden sich ähnelnde Routen zu einem Cluster zusammengefasst. Sobald einem Cluster mehr als eine gewisse Menge an Graphen zugeteilt werden, kann es auf diesen Routen zu Verzögerungen kommen und eine termingerechte Anlieferung kann nicht mehr erfüllt werden. Das Ergebnis könnte in einem nächsten Schritt mithilfe von Analytical Queries dahingehend verarbeitet werden, um zu überprüfen, ob die Fahrzeuge auch zeitgleich die gleichen Routen befahren. Verteilen sich die Graphen eines Clusters etwa über einen ganzen Tag, muss es nicht zwingend zu Verspätungen kommen. Diese Weiterverarbeitung kann theoretisch umgangen werden, indem lediglich die Routen eines gewissen Zeitfensters geclustert werden. Für ein konkretes Beispiel bedeutet dies, dass es wahrscheinlich zu verzögerten Anlieferungen kommt, wenn am Mittwochmorgen über 30 Transportfahrzeuge die gleiche Zufahrt zum Werk nehmen. Somit sind für das Graph-Clustering alle Kriterien erfüllt.

Das Verfahren der Label-Propagation ist für den Einsatz bei dieser Fragestellung geeignet, doch wurde kein Anwendungsfall gefunden, bei dem die Ergebnisse zur Lösung der Problemstellung eingesetzt werden können. Dafür kann der Label-Propagation Algorithmus beispielsweise zusammen mit den Pattern Matching Queries zur Beantwortung genutzt werden. Für diesen Fall sind bereits verschiedene Straßenkreuzungen als Engpässe bekannt (gelabelte Knoten) und es wird ein Modell mit diesen Informationen gelernt. Dadurch können etwa Kreuzungen mit ähnlichen Merkmalen ebenfalls als Engpässe gelabelt werden. In einem nächsten Schritt wird der Graph mit den identifizierten Kreuzungen, die zu

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

Verzögerungen führen können, mit dem Graph eines bestimmten Transportfahrzeuges gematcht. Wenn zwischen diesen Graphen Ähnlichkeiten existieren sollten, könnte sich die Ankunft dieses Verkehrsträgers verzögern.

Für den Algorithmus der Graph-Classification wird analog zum Graph-Clustering davon ausgegangen, dass für jedes Transportfahrzeug ein eigener Graph existiert. Auf Grundlage der gelabelten Graphen, bei denen bereits bekannt ist, ob es bei der Anlieferung zu Verzögerungen kam, lassen sich die ungelabelten Graphen klassifizieren. Wird ein Graph beispielsweise einer Gruppe zugeordnet, deren Lieferungen planmäßig angekommen sind, so ist auch für dieses Transportfahrzeug mit einer pünktlichen Lieferung zu rechnen. Dieses Ergebnis beantwortet die logistische Fragestellung. Für eine gemeinsame Anwendung mit Graph Management Verfahren wurde kein Use Case identifiziert. Somit sind alle außer das dritte Kriterium erfüllt.

Die Entdeckung von neuen Kanten durch den Algorithmus der Link Prediction ergibt bei diesem Anwendungsfall keinen Sinn, da das Straßennetz vorgegeben und nicht erweiterbar ist. Folglich gelten alle Kriterien als nicht erfüllt.

Der Einsatz des Verfahrens der Graph Compression wird ebenfalls als nutzlos gewertet, da eine Zusammenführung von mehreren Knoten oder das Entfernen von Kanten nicht auf das aus der Fragestellung resultierende Szenario übertragbar wäre. Somit ist kein Kriterium erfüllt.

Die Ergebnisse der Eignungsüberprüfung der Algorithmen in Bezug auf die Frage zur Beschaffungseinplanung lassen sich in *Tabelle 4.18* zusammenfassen.

Tabelle 4.18: Kriterienausprägungen der Algorithmen für die Frage der Beschaffungseinplanung

Algorithmus	Eignung	Lösung	Vereinfachung	Use Case
Graph Pattern Mining	✓	✗	✓	✓
Node-Clustering	✓	✗	✓	✓
Graph-Clustering	✓	✓	✓	✓
Label-Propagation	✓	✗	✓	✓
Graph-Classification	✓	✗	✓	✓
Link Prediction	✗	✗	✗	✗
Graph Compression	✗	✗	✗	✗

Die Einteilung der untersuchten Algorithmen des Graph Mining in die Klassen kann für Fragestellung der Beschaffungseinplanung *Tabelle 4.19* entnommen werden.

Tabelle 4.19: Klassifizierung der Algorithmen in Bezug auf die Frage der Beschaffungseinplanung

Fragestellung	Klasse 1	Klasse 2	Klasse 3	Klasse 4
Bei welchen Transportfahrzeugen der Zulieferer ist eine termingerechte Anlieferung zu erwarten?	G-Clu	G-Cla	GPM, N-Clu, L-Pro	LP, GC

4.2.8 Distributionsfeinplanung

Das Teilgebiet der Distributionsfeinplanung wird in diesem Abschnitt durch zwei Fragen thematisiert, wovon die erste wie folgt lautet: „Wie sollten die Verkehrsträger beladen werden, damit sie möglichst

effizient genutzt werden?“ Die Knoten des für diese Fragestellung genutzten Modells stehen für die Distributionszentren, von denen die Waren transportiert werden, und die Kunden mit ihren Bedarfen. Die Verbindungen zwischen diesen Punkten werden dabei durch Kanten dargestellt und müssen sich nicht wie in **Abschnitt 4.2.7** an konkreten Straßenzügen orientieren. Sie können mit einem Kantengewicht versehen sein, wie beispielsweise der Entfernung. Es wird davon ausgegangen, dass ein Verkehrsträger nicht nur die Ware für einen einzelnen Kunden ausliefert, sondern auf seiner Route mehrere Kunden anfährt.

Der Algorithmus des Graph Pattern Mining ist sowohl ohne weitere als auch in Verbindung mit weiteren Verfahren des Graph Management für den Einsatz im vorliegenden Anwendungsgebiet geeignet. So könnte etwa ein Muster Aufschluss darüber geben, dass gewisse Kunden besonders häufig angefahren werden. Statt die Lieferungen auf mehrere Routen aufzuteilen, könnte dieser Kunde beispielsweise von einigen wenigen Transportfahrzeugen beliefert werden. Dadurch ließen sich die Transportkosten reduzieren und die Beladung der einzelnen Verkehrsträger optimieren. In einem weiteren Use Case stehen die häufig entdeckten Muster für Routen. Diese könnten in einem weiteren Schritt durch ein Verfahren des Graph Management dahingehend analysiert werden, ob die zugehörigen Transportfahrzeuge voll beladen sind. Wäre dies nicht der Fall, ließen sich häufige Routen so zusammenfassen, dass die Kapazität der Transportfahrzeuge ausgenutzt wird.

Bei den Use Cases für das Node-Clustering wird davon ausgegangen, dass die Transportkosten einiger Verkehrsträger sehr hoch sind, da sie Kunden beliefern, die nicht nah beieinanderliegen. In dem Graph werden nicht die Routen der Fahrzeuge betrachtet, sondern lediglich die Standorte der Distributionszentren und Kunden (Knoten) und deren Entfernung zueinander (Kanten). Durch den Algorithmus des Node-Clustering können nun die Standorte identifiziert werden, die nah beieinanderliegen. Aus diesen Informationen könnten die Routen dahingehend geplant werden, dass die Verkehrsträger möglichst nur noch mit Waren beladen werden, die an Standorte aus einem Cluster geliefert werden müssen. Allein diese Ergebnisse könnten dafür genutzt werden, die Transportkosten zu minimieren und die Beladung zu optimieren. Jedoch könnten die Standorte eines Clusters in einem zusätzlichen Schritt durch die Verfahren des Graph Management analysiert werden, um weitere Erkenntnisse über die Standorte zu erlangen.

Beim Graph-Clustering können die Ergebnisse ebenfalls zur Lösung und Vereinfachung des Problems genutzt werden. Um die Komplexität des Graphen zu reduzieren, wird für jede Transportroute ein eigener Graph erstellt. In diesem Graph werden nur jene Verbindungen (Kanten) zwischen zwei Standorten (Knoten) eingetragen, die auch auf der Route befahren werden. Durch den Algorithmus des Graph-Clustering werden sich ähnelnde Routen zu Clustern zusammengefasst. Eine große Anzahl an Graphen in einem Cluster könnte ein Anzeichen dafür sein, dass die Verkehrsträger nicht effizient beladen sind. Bestätigt sich diese Information, könnten diese Routen zusammengefasst werden. In einem weiteren Anwendungsfall wird eine neue Route eingeführt. Unter Kenntnis der identifizierten Cluster ließe sich mithilfe von Graph Pattern Matching Queries überprüfen, ob nicht bereits eine ähnliche Route besteht und die zu transportierenden Waren auf diesen Typ aufgeteilt werden können.

Das Verfahren der Label-Propagation ist nicht für das Anwendungsgebiet der logistischen Fragestellung geeignet, da kein Use Case identifiziert werden konnte, bei dem eine Klassifizierung von ungelabelten Knoten auf Basis von bereits gelabelten Knoten hilfreich wäre.

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

Ähnlich zu dem Verfahren des Graph-Clustering würde bei der Graph-Classification in diesem Fallbeispiel jeder Routenplan eines Transportfahrzeuges durch einen Graphen repräsentiert werden. Da die Routen bereits bekannt sind, wurde kein Anwendungsfall identifiziert, für den eine Klassifizierung von ungelabelten Graphen auf der Basis von bereits gelabelten Graphen zielführend wäre. Somit ist keines der Kriterien erfüllt.

Die Entdeckung von fehlenden oder zukünftigen Verbindungen hilft bei der Beantwortung der Fragestellung nicht weiter. Die meisten Routen sind bereits vorgegeben, sodass eine neue Verbindung keine Auswirkungen auf die bestehenden Routen hätte. Die Kriterien gelten als nicht erfüllt.

Auch beim Algorithmus der Graph Compression konnte kein Use Case gefunden werden, bei dem das Verfahren bei der Problemlösung unterstützen könnte. Die Zusammenführung von mehreren Knoten oder das Streichen von bestehenden Kanten ist nicht zielführend, da für die Beantwortung der Frage das komplette Modell betrachtet werden muss. Aus diesem Grund gilt kein einzelnes Kriterium als erfüllt.

Die Ergebnisse der Eignungsüberprüfung der Algorithmen in Bezug auf die erste Frage zur Distributionsfeinplanung lassen sich in **Tabelle 4.20** zusammenfassen.

Tabelle 4.20: Kriterienausprägungen der Algorithmen für die erste Frage der Distributionsfeinplanung

Algorithmus	Eignung	Lösung	Vereinfachung	Use Case
Graph Pattern Mining	✓	✓	✓	✓
Node-Clustering	✓	✓	✓	✓
Graph-Clustering	✓	✓	✓	✓
Label-Propagation	✗	✗	✗	✗
Graph-Classification	✗	✗	✗	✗
Link Prediction	✗	✗	✗	✗
Graph Compression	✗	✗	✗	✗

Die Einteilung der untersuchten Algorithmen des Graph Mining in die Klassen kann für die erste Fragestellung der Distributionsfeinplanung **Tabelle 4.21** entnommen werden.

Tabelle 4.21: Klassifizierung der Algorithmen in Bezug auf die erste Frage der Distributionsfeinplanung

Fragestellung	Klasse 1	Klasse 2	Klasse 3	Klasse 4
Wie sollten die Verkehrsträger beladen werden, damit sie möglichst effizient genutzt werden?	GPM, N-Clu, G-Clu			L-Pro, G-Cla, LP, GC

Für die zweite Frage der Distributionsfeinplanung wird von dem konkreten Fall ausgegangen, dass ein Hersteller von Schrauben viele Unternehmen mit seinen Schrauben beliefert. Die Unternehmen benötigen die Schrauben zur Fertigung ihrer Produkte und bestellen diese erst nach, wenn der Bestand einen vorher festgelegten Sicherheitswert unterschreitet. Aufgrund der Tatsache, dass es zwischen dem Schraubenhersteller und den Unternehmen keinen Rahmenvertrag über die Frequenz und Liefermenge der Bestellungen gibt, kann es vorkommen, dass der Schraubenhersteller spontan auf dringende Bestellungen eingehen muss. Die damit verbundene Frage lautet: „Welche Unternehmen sind als erstes zu beliefern?“. Die Knoten des zugehörigen Graphs stehen für die Unternehmen, die die Schrauben vom

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

Hersteller beziehen, und die Kanten verbinden die Unternehmen untereinander. Dabei besitzen Kanten einen Entfernungswert und die Knoten einen Bedarfswert, der umso höher ist, je niedriger die Restbestände an Schrauben eines Unternehmens sind. Die Unternehmen mit einem hohen Bedarfswert sind bei der Distributionsfeinplanung priorisiert zu berücksichtigen.

Der Algorithmus des Graph Pattern Mining ist für die Anwendung in diesem Szenario geeignet. Es wird davon ausgegangen, dass der Entfernungswert der Kanten, die die Unternehmen untereinander verbinden, nicht länger als zehn Kilometer sein darf. Werden beim Graph Pattern Mining nun eine Anzahl von Sterngraphen extrahiert, deren zentraler Knoten beispielsweise mit fünf weiteren Unternehmen verbunden ist, so ließen sich bei der Belieferung dieses Unternehmens (zentraler Knoten) ohne großen Mehraufwand die Bedarfe der weiteren fünf Unternehmen befriedigen. Somit sind die Ergebnisse des Graph Pattern Mining Algorithmus zur Beantwortung der Fragestellung geeignet, da durch die Entdeckung dicht beieinanderliegender Unternehmen viele Aufträge mit einer einzelnen Route erfüllt werden können. Jedoch wurde kein Use Case identifiziert, bei dem die Ergebnisse sinnvoll durch Verfahren des Graph Management bearbeitet werden könnten. Aus diesem Grund sind alle Kriterien bis auf das dritte erfüllt.

Beim Node-Clustering konnte kein Anwendungsfall identifiziert werden, bei dem die Ergebnisse zur Lösung des Problems beitragen. Stattdessen gibt es einen Use Case, bei dem die Unternehmen (Knoten) nach ihrer Entfernung geclustert werden, sodass sich nah beieinanderliegende Unternehmen im gleichen Cluster befinden. Es könnte sich dafür entschieden werden, die Unternehmen eines dichten Clusters zuerst mit Schrauben zu beliefern, da auf diese Weise durch die Reduzierung der Transportzeiten zahlreiche Aufträge erfüllt werden könnten. Allerdings müsste durch den Einsatz von zusätzlichen Algorithmen entschieden werden, welche Knoten eines Clusters zuerst beliefert werden.

Für die Anwendung des Graph-Clustering Algorithmus ist der zugrundeliegende Graph zu umfangreich und müsste unterteilt werden. Allerdings wurde keine Möglichkeit identifiziert, den Graphen sinnvoll und zielführend in mehrere Subgraphen zu unterteilen. Aus diesem Grund ist die Eignung nicht erfüllt und folglich gilt keines der Kriterien.

Der Algorithmus der Label-Propagation lässt sich sowohl zur Lösung als auch zur Vereinfachung der Problemstellung nutzen. Es wird von dem Szenario ausgegangen, dass der Schraubenhersteller die Bedarfswerte der Unternehmen in unterschiedliche Bedarfsklassen eingeteilt werden. Beispielsweise würde ein Unternehmen der ersten Bedarfsklasse zugeordnet werden, wenn der Restbestand der Schrauben kleiner als 30% ist. Diese Klasse müsste schneller beliefert werden als die zweite Bedarfsklasse, bei der der Restbestand etwa zwischen 30% und 45% liegt. Nun sind jedoch nur manche Unternehmen des Graphen in diese Klassen eingeteilt (gelabelte Knoten). Durch den Algorithmus der Label-Propagation könnten auch die anderen ungelabelten Unternehmen den jeweiligen Bedarfsklassen zugeordnet werden und das Ergebnis wäre eine Aufteilung, welche Aufträge priorisiert behandelt werden müssen. Sollte die Fragestellung durch diese Aufteilung noch nicht beantwortet sein, könnten die Unternehmen einer Klasse durch die Analytical Queries des Graph Management in einem weiteren Schritt untersucht werden. Somit sind alle Kriterien erfüllt.

Bei dem Verfahren der Graph-Classification besteht das gleiche Problem wie auch schon bei dem Graph-Clustering. Für eine Anwendung müsste der komplexe Graph unterteilt werden, was jedoch für diesen Fall als nicht sinnvoll erachtet wird. Da das Kriterium der Eignung nicht erfüllt ist, ist keines der Kriterien erfüllt.

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

Das Verfahren der Link Prediction ist ebenfalls nicht für das vorgestellte Szenario geeignet, da dabei vor allem die Bedarfswerte der einzelnen Unternehmen interessieren und die Unternehmen untereinander keine Abhängigkeiten besitzen. Eine Entdeckung von neuen Verbindungen würde also keinen Sinn machen, weshalb keines der Kriterien erfüllt ist.

Der Algorithmus der Graph Compression ließe sich zwar auf den komplexen Graphen anwenden, jedoch wurde kein Use Case identifiziert, bei dem ein komprimierter Graph unmittelbar zur Lösung der logistischen Fragestellung verwendet werden kann. Dafür könnten in einem anderen Anwendungsfall die Unternehmen (Knoten) aus dem Graph entfernt werden, deren Bedarfswert niedriger als ein vorher festgelegter Wert ist. Sollte ein Unternehmen beispielsweise eine Bestellung aufgegeben haben, obwohl lediglich 40% der Bestände verarbeitet wurden, könnte dieser Knoten aus dem Graph entfernt werden. Das Resultat ist ein komprimierter Graph, dessen Knoten dringend beliefert werden müssen. Die verbliebenen Knoten könnten durch die Algorithmen des Graph Management in einem nächsten Schritt analysiert werden, um die Fragestellung zu beantworten.

Die Ergebnisse der Eignungsüberprüfung der Algorithmen in Bezug auf die zweite Frage zur Distributionsfeinplanung lassen sich in **Tabelle 4.22** zusammenfassen.

Tabelle 4.22: Kriterienausprägungen der Algorithmen für die zweite Frage der Distributionsfeinplanung

Algorithmus	Eignung	Lösung	Vereinfachung	Use Case
Graph Pattern Mining	✓	✓	✗	✓
Node-Clustering	✓	✗	✓	✓
Graph-Clustering	✗	✗	✗	✗
Label-Propagation	✓	✓	✓	✓
Graph-Classification	✗	✗	✗	✗
Link Prediction	✗	✗	✗	✗
Graph Compression	✓	✗	✓	✓

Die Einteilung der untersuchten Algorithmen des Graph Mining in die Klassen kann für die zweite Fragestellung der Distributionsfeinplanung **Tabelle 4.23** entnommen werden.

Tabelle 4.23: Klassifizierung der Algorithmen in Bezug auf die zweite Frage der Distributionsfeinplanung

Fragestellung	Klasse 1	Klasse 2	Klasse 3	Klasse 4
Welche Unternehmen sind als erstes zu beliefern?	L-Pro	GPM	N-Clu, GC	G-Clu, G-Cla, LP

4.2.9 Zusammenfassung der Klassifizierungen

Die Klassifizierungen der Algorithmen aus den Abschnitten 4.2.1 bis 4.2.8 lassen sich in **Tabelle 4.24** zusammenfassen. Diese Tabelle ist aufgeteilt nach den in den Abschnitten 2.4.2 und 2.4.3 definierten Teilgebieten des Aufgabenmodells und ihren zugehörigen logistischen Fragestellungen (vgl. Abschnitt 4.1.4). Für jede logistische Fragestellung werden die Graph Mining Algorithmen (vgl. Abschnitt 3.3.2) basierend auf den Ergebnissen der Eignungsüberprüfungen (vgl. Abschnitt 4.2) in die in Abschnitt 4.1.6 definierten Klassen eingeordnet. Aus Gründen der Leserlichkeit werden in der Tabelle die in Abschnitt 4.1.6 eingeführten Abkürzungen genutzt. Es wird versucht, aus den Ergebnissen Rückschlüsse auf die

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

allgemeine Verwendung der verschiedenen Graph Mining Algorithmen bei logistischen Fragestellungen zu ziehen.

Die Tabelle stellt eine Klassifizierung von Algorithmen des Graph Mining für eine bestimmte logistische Fragestellung dar und könnte beispielsweise zur Lösung einer realen Problemstellung genutzt werden. Dafür müsste der Anwender überprüfen, ob sich eine der vorgestellten Fragestellung auf den realen Anwendungsfall anwenden ließe. Ist dies der Fall, könnte den Eignungsüberprüfungen in Abschnitt 4.2 entnommen werden, ob sich ein Algorithmus zur Lösung, zur Vereinfachung und weiteren Bearbeitung durch einen Algorithmus des Graph Management oder sogar zu beidem einsetzen ließe. Bei der Auswahl eines bestimmten Verfahrens ist jedoch immer noch zu beachten, welches Ziel bei der logistischen Fragestellung eigentlich verfolgt werden soll. So besitzen auch die Algorithmen unterschiedliche Verwendungszwecke, die in Abschnitt 4.1.3 kurz zusammengefasst werden. Aus diesen Gründen lässt sich in der Regel nicht sagen, dass es für einen bestimmten Anwendungsfall nur einen zweckmäßigen Algorithmus gibt. Welches Verfahren letztendlich Anwendung findet, muss für jeden Fall individuell betrachtet werden.

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

Tabelle 4.24: Überblick der Klassifizierungen von Graph Mining Algorithmen in Bezug auf logistische Fragestellungen des Aufgabenmodells

Teilgebiet des Aufgabenmodells	Logistische Fragestellung	Zur Lösung und Vereinfachung geeignet	Nur zur Lösung geeignet	Nur zur Vereinfachung geeignet	Ungeeignet
Strategische Netzwerkgestaltung	Wo soll ein neues Werk für Produkt X errichtet werden?	GPM, N-Clu, L-Pro		GC	G-Clu, G-Cla, LP
	Wie viele Standorte zur Distribution sind zu errichten?	L-Pro	N-Clu	GPM, GC	G-Clu, G-Cla, LP
Bedarfsplanung	Wie wird sich der Bedarf an Produkt X das kommende Quartal in Region Y entwickeln?	GPM	N-Clu, G-Cla, LP		G-Clu, L-Pro, GC
Netzwerkplanung	Welches Werk muss wie viele Produkte eines Typs produzieren, damit der gesamte Bedarf bestmöglich abgedeckt wird?	L-Pro		GPM, N-Clu, GC	G-Clu, G-Cla, LP
Beschaffungsplanung	Für welche Werke sind welche Anlieferhythmen zu wählen, damit die Materialbereitstellung sichergestellt wird?		L-Pro, G-Cla	GPM, N-Clu, G-Clu	LP, GC

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

Distributionsplanung	Auf welchen Verkehrsträger fällt die Wahl vor dem Hintergrund der Transportkostenminimierung?	GPM, N-Clu	G-Clu	GC	L-Pro, G-Cla, LP
Verfügbarkeits- und Machbarkeitsprüfung	Kann die Lieferung bis zu einem definierten Datum erfolgen?				GPM, N-Clu, G-Clu, L-Pro, G-Cla, LP, GC
Beschaffungsfeinplanung	Bei welchen Transportfahrzeugen der Zulieferer ist eine termingerechte Anlieferung zu erwarten?	G-Clu		GPM, N-Clu, L-Pro, G-Cla	LP, GC
Distributionsfeinplanung	Wie sollten die Verkehrsträger beladen werden, damit sie möglichst effizient genutzt werden?	GPM, N-Clu, G-Clu			L-Pro, G-Cla, LP, GC
	Welche Unternehmen sind als erstes zu beliefern?	L-Pro	GPM	N-Clu, GC	G-Clu, G-Cla, LP

In **Tabelle 4.25** wird die Häufigkeit eingetragen, wie oft welcher Graph Mining Algorithmus (vgl. Abschnitt 3.3.2) in welche Klasse (vgl. Abschnitt 4.1.6) eingeteilt wurde. Anschließend wird versucht, aus dieser Zusammenfassung Regelmäßigkeiten zu erkennen und wenn möglich Rückschlüsse zu ziehen. Dieser Schritt ist mit dem Gewinn von neuen theoretischen Erkenntnissen zu vergleichen, die nach Tarjan auf andere Bereiche übertragen werden können (vgl. Abschnitt 3.1). Dabei werden die Algorithmen in der Reihenfolge betrachtet, wie sie in Abschnitt 3.3.2 beschrieben wurden.

Tabelle 4.25: Häufigkeit der Klassenzuteilungen je Graph Mining Algorithmus

Algorithmus	Klasse 1	Klasse 2	Klasse 3	Klasse 4
Graph Pattern Mining	4	1	4	1
Node-Clustering	3	2	4	1
Graph-Clustering	2	1	1	6
Label-Propagation	4	1	1	4
Graph-Classification	0	2	1	7
Link Prediction	0	1	0	9
Graph Compression	0	0	5	5

Der Algorithmus des Graph Pattern Mining hatte in den definierten Fallbeispielen das breiteste Anwendungsgebiet. Er wurde am häufigsten von den betrachteten Verfahren der Algorithmen-Klasse zugeteilt, deren Ergebnisse sowohl für die Lösung als auch für die Vereinfachung der Problemstellung genutzt werden können. Falls das Graph Pattern Mining nicht zur unmittelbaren Lösung beitragen sollte, können die Ergebnisse in den meisten Fällen immer noch von Verfahren des Graph Management verarbeitet werden, um durch diesen zusätzlichen Schritt zu einer Lösung zu kommen.

Das Clustering von Knoten ist für die Anwendung in den meisten logistischen Fragestellungen geeignet, da es abgesehen von einer Ausnahme stets zur Lösung, zur Vereinfachung oder zu beidem genutzt werden kann. Somit gehört der Node-Clustering Algorithmus zusammen mit dem Graph Pattern Mining zu den vielfältigsten Verfahren.

Da der Graph-Clustering Algorithmus nicht geeignet für einen umfangreichen Graphen ist, sondern mehrere überschaubare Graphen voraussetzt, wurde er für einen Großteil der Fragestellungen der vierten Klasse zugeteilt, deren Verfahren weder für die Lösung noch für die Vereinfachung und anschließenden Bearbeitung geeignet sind. Wenn sich jedoch der komplexe Graph sinnvoll unterteilen lässt, dann kann das Graph-Clustering auch zur Lösung oder Vereinfachung der Problemstellung eingesetzt werden.

Bei dem Algorithmus der Label-Propagation lässt sich das Muster erkennen, dass er entweder sowohl für die Lösung als auch für die Vereinfachung durch weitere Verfahren gedacht ist oder für gar nichts von beiden. Entscheidend für Eignung dieses Verfahrens ist die Tatsache, ob dem Anwender bereits gelabelte Knoten zur Verfügung stehen und diese zur Einteilung der ungelabelten Knoten verwendet werden sollen. Trifft diese Voraussetzung zu, ist das Verfahren meist gleichzeitig für Lösung und Vereinfachung geeignet. Da es nicht in jedem Fall sinnvoll ist, Rückschlüsse von einigen wenigen Knoten auf den Rest zu ziehen, wird der Algorithmus in der Regel entweder in die erste oder die vierte Klasse eingeteilt.

Für den Einsatz des Graph-Classification Algorithmus werden ebenfalls mehrere kleinere Graphen, statt ein komplexer Graph benötigt. Dieser Punkt wäre ebenfalls erfüllt, wenn sich der umfangreiche Graph in mehrere kleine Graphen unterteilen ließe, was jedoch in der Praxis oft nicht möglich ist. Aus diesem Grund ist das Verfahren in den meisten Fällen der vierten Kategorie zuzuordnen. Es wurden lediglich

Entwicklung einer Klassifizierung von Graph Mining Algorithmen

vereinzelt Fälle identifiziert, bei denen die Ergebnisse der Graph-Classification entweder zur Problemlösung oder zur Weiterverarbeitung durch ein Graph Management Verfahren genutzt werden können.

Der vom Link Prediction Algorithmus verfolgte Zweck der Entdeckung von versteckten oder zukünftig erscheinenden Kanten konnte bei fast keiner der vorgestellten logistischen Fragestellung zu einer Lösung oder Vereinfachung führen. Ein Grund dafür mag sein, dass die verwendeten Modelle oftmals ein Straßennetz oder ähnliches darstellten, bei dem die relevanten Kanten (z.B. Entfernungen) bereits existieren und eine Ermittlung von neuen Kanten keinen Sinn macht. Bei dem Einsatz solcher Modelle ist die Link Prediction in der Regel ungeeignet. Sollten die Kantengewichte nicht fest definiert, sondern variabel sein (z.B. Objekt- oder Informationsflüsse), kann das Verfahren der Link Prediction je nach Fragestellung zu einer Lösung führen.

Die Ergebnisse des Graph Compression Algorithmus können kein einziges Mal für die Lösung der Problemstellung eingesetzt werden. Doch auch wenn die Komprimierung eines Graphen nicht direkt zur Lösung führt, können die Ergebnisse häufig für die Weiterverarbeitung durch einen Graph Management Algorithmus verwendet werden. Die Graph Compression ist meistens dann zur Vereinfachung geeignet, wenn durch die logistische Fragestellung keine spezifischen Knoten interessieren. Sollte dies allerdings zutreffen, ist das Verfahren für gewöhnlich der vierten Klasse zuzuordnen.

Die abgeleiteten Erkenntnisse beziehen sich nur auf die in Abschnitt 4.1.4 vorgestellten Fragestellungen und es wird kein Anspruch erhoben, dass die Erkenntnisse allgemeingültig sind. Da die Eignung für den Einsatz von gewissen Algorithmen stark von der Fragestellung abhängt, hätten sich die abgeleiteten Erkenntnisse und die Zuteilung zu den Klassen für andere Fragestellungen deutlich unterschieden. Aus diesem Grund wäre es sinnvoll, die Algorithmen anhand des Typs der Fragestellung und worauf diese abzielt zu klassifizieren. Dafür müssten die unterschiedlichen Fragestellungen im Supply Chain Management untersucht werden.

5 Evaluierung der Methode

In diesem Kapitel soll die Gültigkeit der entwickelten Klassifizierung von Graph Mining Algorithmen zur Beantwortung logistischer Fragestellungen in Supply Chains überprüft werden (vgl. Abschnitt 4.2). Dafür werden die Ergebnisse in Abschnitt 5.1 auf ein geeignetes Fallbeispiel angewendet. Das Fallbeispiel bezieht sich auf eine aktuelle und reale Problemstellung, wobei das Szenario geringfügig angepasst wurde. Bei der Evaluierung wird überprüft, ob sich die in Abschnitt 4.2 gewonnenen Erkenntnisse auf den aktuellen Anwendungsfall übertragen lassen. Die Ergebnisse werden im Anschluss in Abschnitt 5.2 kritisch hinterfragt.

5.1 Fallbeispiel

Das Fallbeispiel bezieht sich auf die derzeitige COVID-19-Pandemie. Dabei bezeichnet COVID-19 die Erkrankung, die durch eine Infektion mit dem neuartigen Coronavirus SARS-CoV-2 ausgelöst wird. Seit seinem erstmaligen Nachweis Anfang 2020 hat sich der Virus pandemisch ausgebreitet, sodass bereits zu Beginn des Jahres 2021 allein in Deutschland über 1,7 Millionen Menschen an COVID-19 erkrankt und von denen mehr als 35.000 Menschen gestorben sind. Die Ausbreitung wird begünstigt durch die leichte Übertragbarkeit und die Infektiosität in der Krankheitsphase, in der ein infizierter Mensch entweder noch keine Symptome entwickelt hat (präsymptomatisch) oder gar keine Symptome entwickeln wird (asymptomatisch). (vgl. Robert Koch-Institut 2021, S. 7) Zur Eindämmung der Pandemie wurde aus diesem Grund Mitte März 2020 das öffentliche Leben weitgehend eingeschränkt. Auch wenn durch diesen Lockdown die Infektionsfallzahlen signifikant zurückgegangen sind, nehmen sie seit Anfang September wieder deutlich zu. Durch diese zweite, weit intensivere Infektionswelle wurde Anfang November ein Beschluss gefasst, einen zweiten „Teil-Lockdown“ einzuführen, der ab Mitte Dezember verschärft wurde. (vgl. Robert Koch-Institut 2021, S. 12 f.) Die Kontaktbeschränkungen könnten mittelfristig gelockert werden, wenn ein Großteil der Bevölkerung eine Immunität gegen das Virus entwickelt. Diese Immunität kann durch die Verabreichung eines effektiven und sicheren Impfstoffes entwickelt werden, der auf diese Weise einen wesentlichen Beitrag bei der Bekämpfung der Pandemie leisten kann. (vgl. Robert Koch-Institut 2021, S. 7) Zum Zeitpunkt des 29. Dezember 2020 wird aus diesem Grund an mehr als 230 Impfstoffkandidaten geforscht (vgl. Robert Koch-Institut 2021, S. 9). Die erste Zulassung innerhalb der EU wurde dem Impfstoff der Firma BioNTech/Pfizer am 21. Dezember 2020 erteilt (vgl. Robert Koch-Institut 2021, S. 18).

Der neue Impfstoff wird anfangs aufgrund der begrenzten Verfügbarkeit nicht jedem Menschen zur Verfügung gestellt werden können. Darum sollen Personengruppen, bei denen mit einem schweren oder sogar tödlichen Krankheitsverlauf zu rechnen ist, die der Gefahr einer Infektion aus beruflichen Gründen ausgesetzt sind oder die einen engen Kontakt zu vulnerablen Personengruppen haben, zuerst geimpft werden. Die Priorisierungsempfehlung der Ständigen Impfkommission am Robert Koch-Institut zählt vor allem die Bewohner und Bewohnerinnen von Senioren- und Altenpflegeheimen und Personen, die 80 Jahre alt oder älter sind, zu der am höchsten zu priorisierenden Personengruppe. (vgl. Robert Koch-Institut 2021, S. 3)

An diesem realen Anwendungsgebiet knüpft das Fallbeispiel an, bei dem es um die Organisation der Impfungen gehen soll. Es wird davon ausgegangen, dass eine begrenzte Anzahl von Impfdosen zur Verfügung stehen und diese zuerst den Bewohnern und Bewohnerinnen in Senioren- und Altenpflegeheimen verabreicht werden sollen. Die zuständigen Impfteams müssen dafür in die Heime fahren, da die Bewohner oft nicht so mobil sind, als dass sie zu einem Impfzentrum kämen. Damit eine Impfung

Evaluierung der Methode

durchgeführt werden kann, müssen die Räumlichkeiten des Heims zu Beginn durch die Impfteams vorbereitet und nach der letzten Impfung wieder aufgeräumt werden. Zusätzlich ist vor jeder Impfung ein kurzes Arztgespräch eingeplant, sodass ein Impfteam an einem Tag nur ein Heim anfahren kann. Die daraus resultierende logistische Fragestellung lautet: „In welchen Senioren- oder Altenpflegeheimen soll eine Impfung zuerst angeboten werden?“. Diese Fragestellung lässt sich dem Teilgebiet der Distributionsfeinplanung aus dem Aufgabenmodell zuordnen, da es im übertragenen Sinne um eine kurzfristige Planung der Verteilung eines bestimmten Produkts geht (vgl. Abschnitt 2.4.3).

Das reale Anwendungsgebiet lässt sich zu dem Graphen in **Abbildung 5.1** abstrahieren, dessen Knoten für die Senioren- oder Altenpflegeheime in einer Region stehen (vgl. Abschnitt 3.2). Jedem Knoten des Graphen ist ein Knotengewicht zugeordnet, das für den Anteil Personen in dem Heim steht, die 80 Jahre alt oder älter sind. Die Kanten, die die Knoten verbinden, stehen für die Entfernungen der jeweils verbundenen Heime zueinander, wobei eine Kante nur dann eingezeichnet wird, wenn die Entfernung der Heime höchstens fünf Kilometer beträgt. Der Übersichtlichkeit halber werden die Kanten im Graph nicht mit den Entfernungswerten beschriftet und die Knoten beinhalten keine allgemeinen Informationen zum Heim (Name, Personengröße, etc.), sondern lediglich den Prozentsatz der Personen, die 80 Jahre alt oder älter sind. Bei dem Graphen handelt es sich außerdem um eine qualitative und keine maßstabsgetreue Darstellung des Netzes aus Heimen.

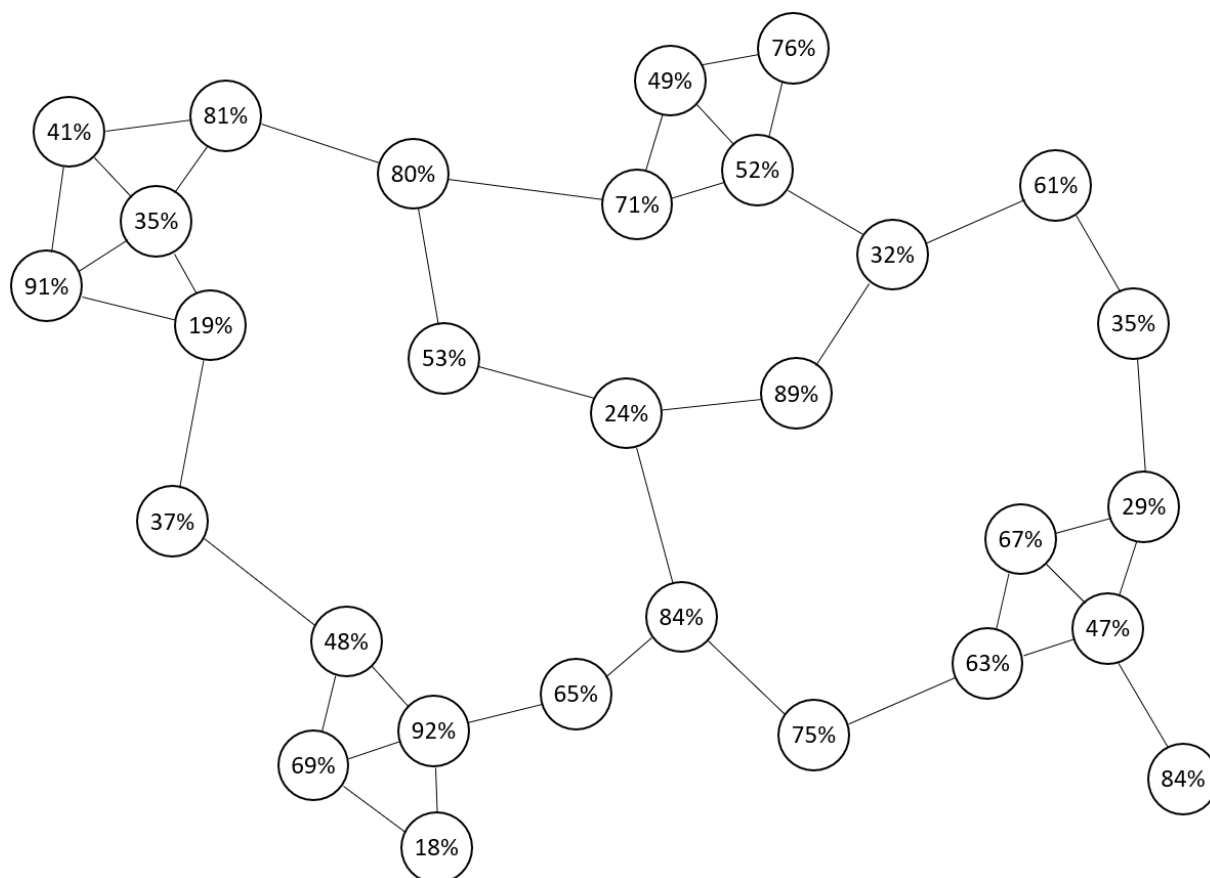


Abbildung 5.1: Ursprungsgraph des Fallbeispiels

In einem nächsten Schritt wird die **Tabelle 4.24** genutzt, um deren beiden logistischen Fragestellungen der Distributionsfeinplanung mit der Fragestellung des Fallbeispiels zu vergleichen. Aus dem Vergleich geht hervor, dass die Frage „Welche Unternehmen sind als erstes mit einer Lieferung zu versorgen?“ der Problemstellung des Fallbeispiels „In welchen Senioren- oder Altenpflegeheimen soll eine Impfung

Evaluierung der Methode

zuerst angeboten werden?“ am nächsten kommt. In beiden Szenarien haben die Knoten einen bestimmten Bedarf an einem Produkt. Ein Dringlichkeitswert ergibt sich bei den Unternehmen aus Abschnitt 4.2.8 aus dem Restbestand an Schrauben und bei den Heimen aus dem Prozentsatz der Personen, die 80 Jahre alt sind oder älter. *Tabelle 4.24* ist zu entnehmen, dass der Algorithmus der Label-Propagation sowohl zur Lösung als auch zur Vereinfachung und anschließenden Weiterverarbeitung durch Graph Management Verfahren geeignet ist. Das Graph Pattern Mining soll die Fragestellung lösen und ist deshalb der zweiten Klasse zugeordnet. Die dritte Klasse enthält Verfahren, dessen Ergebnisse den Graphen so vereinfachen, dass die Problemstellung durch eine zusätzliche Anwendung eines Graph Management Algorithmen lösen lässt. Dieser Klasse sind in der Tabelle das Node-Clustering und die Graph Compression zugeteilt. Lediglich die Algorithmen des Graph-Clustering, der Graph-Classification und der Link Prediction sind nach der Tabelle weder zur Lösung noch zur Vereinfachung geeignet (vgl. Abschnitt 4.2.8). Diese Klassifizierung wird im Folgenden überprüft.

Der Algorithmus der Label-Propagation (vgl. Abschnitt 3.3.2) ist auch für die Fragestellung des Fallbeispiels sowohl für die Lösung als auch für die Vereinfachung und Weiterverarbeitung durch Graph Management Verfahren gedacht. Vor der Ausführung der Label-Propagation sind einige der Knoten bereits zu vier verschiedenen Klassen zugeordnet, deren Zuordnungskriterium der Prozentanteil der Bewohner und Bewohnerinnen ist, die 80 Jahre alt oder älter sind. Es ist bekannt, dass die erste Klasse als erstes beliefert werden soll, da in sie die Knoten mit den höchsten Prozentzahlen eingeordnet werden. Die übrigen gelabelten Knoten gehören einer umso höheren Klasse an, je niedriger ihr Prozentanteil der Bewohner und Bewohnerinnen über 80 Jahren ist. Durch den Algorithmus werden die restlichen ungelabelten Knoten auf Grundlage der bereits gelabelten Knoten in die vier Klassen eingeteilt. Dabei wird die erste Klasse mit dem höchsten Prozentsatz rot, die zweite orange, die dritte gelb und die Klasse mit dem niedrigsten Prozentsatz blau dargestellt, um das Ergebnis visuell hervorzuheben (siehe *Abbildung 5.2*). Aus dem Ergebnis lässt sich deutlich erkennen, welche Senioren- oder Altenpflegeheime den größten Anteil an Personen über 80 Jahren haben. Die Frage kann dadurch beantwortet werden, dass die Impfungen zuerst in den Heimen der ersten Klasse (rot) angeboten werden. Sollten weniger Impfteams verfügbar sein, als es Knoten in der ersten Klasse gibt, so könnten diese Knoten zusätzlich durch die Analytical Queries des Graph Management untersucht werden, um die Frage dennoch beantworten zu können (vgl. Abschnitt 3.3.1). Eine Vereinfachung des Graphen durch die Label-Propagation ist demnach auch erfüllt. Die Klassifizierung der *Tabelle 4.24* konnte in Bezug auf diesen Algorithmus bestätigt werden.

Evaluierung der Methode

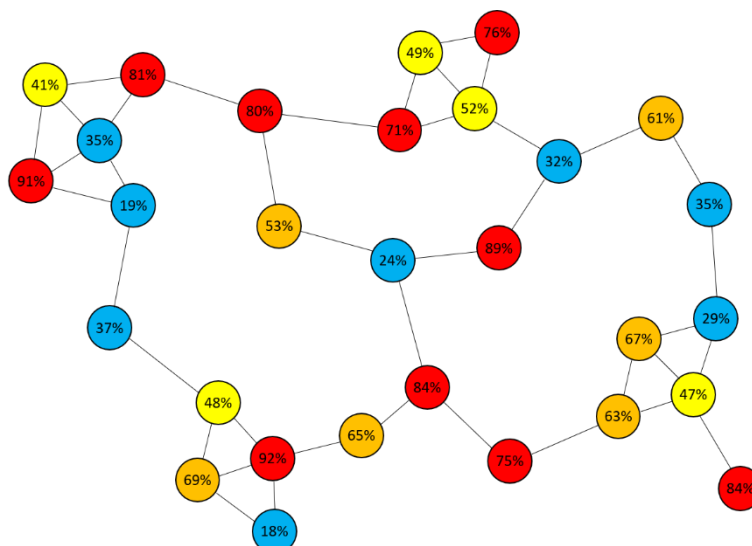


Abbildung 5.2: Graph nach Anwendung des Algorithmus der Label Propagation

Wird beispielsweise nicht die Strategie verfolgt, dass die Altenheime nach dem Prozentsatz priorisiert werden sollen, sondern sollen durch den Transport der Bewohner und Bewohnerinnen aus den umliegenden in ein zentrales Senioren- oder Altenpflegeheim möglichst viele Personen durch ein Impfteam geimpft werden, kann der Algorithmus des Graph Pattern Mining (vgl. Abschnitt 3.3.2) zur Lösung der Problemstellung verwendet werden. Durch die in **Abbildung 5.3** dargestellte Identifizierung von Sterngraphen (vgl. Abschnitt 3.1), bei dem der zentrale Knoten mit vier weiteren Knoten verbunden ist (rot markiert), werden solche zentralen Heime gefunden, in die möglichst viele Bewohner und Bewohnerinnen zum Zweck der Impfung transportiert werden können. Die Frage ist mit der Kenntnis über die zentralen Heime beantwortet. Jedoch wurde kein Fall identifiziert, bei dem die Ergebnisse des Graph Pattern Mining derart von Verfahren des Graph Management weiterverarbeitet werden können, dass die Problemstellung gelöst werden kann. Die Klassifizierung nach der **Tabelle 4.24** in Bezug auf diesen Algorithmus kann somit bestätigt werden.

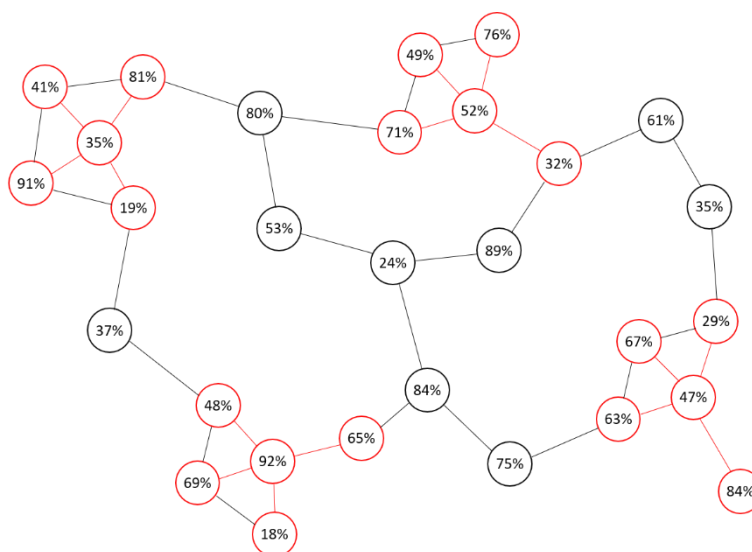


Abbildung 5.3: Graph nach Anwendung des Algorithmus des Graph Pattern Mining

Für das Node-Clustering (vgl. Abschnitt 3.3.2) wurde kein Use Case identifiziert, bei dem die Ergebnisse nach einer Interpretation unmittelbar zur Lösung der Fragestellung genutzt werden könnten. Bei

Evaluierung der Methode

dem Verfahren werden die Knoten mithilfe einer Zielfunktion in unterschiedliche Cluster unterteilt. Diese Unterteilung könnte etwa so aussehen wie in *Abbildung 5.4*, bei der die Cluster der Übersichtlichkeit halber farblich markiert wurden. Geht man von dieser Aufteilung aus, konnten fünf Cluster identifiziert werden. Sind etwa fünf Impfteams verfügbar, wäre es möglich, dass jedes Team einem Cluster zugeteilt wird. Das Senioren- oder Altenpflegeheim, welches keinem Cluster zugeordnet werden konnte, könnte von dem Team angefahren werden, das dem gelben Cluster zugeteilt ist, da es im Gegensatz zu den anderen Teams nicht so viele Heime anzufahren hat. Nichtsdestotrotz stellt dieses Ergebnis nur eine Vereinfachung dar und es muss mit zusätzlichen Verfahren des Graph Management entschieden werden, welche Heime jedes Clusters zuerst angefahren werden (vgl. Abschnitt 3.3.1). Die in der *Tabelle 4.24* aufgeführte Klassifizierung für das Node-Clustering kann somit bestätigt werden.

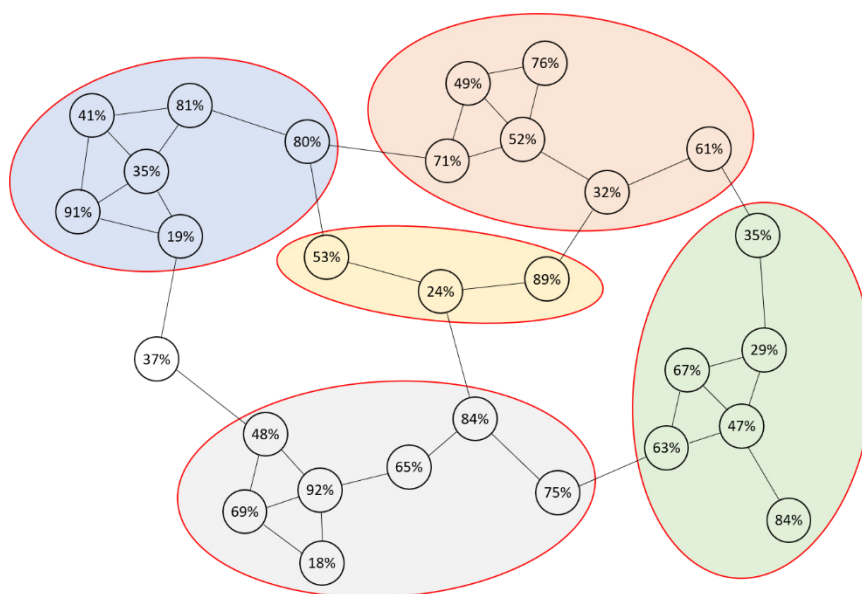


Abbildung 5.4: Graph nach Anwendung des Algorithmus des Node-Clustering

Eine Komprimierung des Graphen durch den Algorithmus der Graph Compression (vgl. Abschnitt 3.3.2) kann die Fragestellung nicht unmittelbar lösen, da einerseits das Entfernen von Kanten unsinnig wäre und sich andererseits durch die Zusammenführung oder das Streichen von Knoten keine Erkenntnisse ableiten ließen, in welchen Senioren- oder Altenpflegeheimen eine Impfung zuerst angeboten werden sollte. Wird jedoch bei der Impfreihenfolge etwa die Strategie verfolgt, dass Heime mit einem hohen Prozentsatz an Personen über 80 Jahren von den Impfteams als erstes angefahren werden sollen, so kann der Graph durch die Graph Compression vereinfacht werden, indem beispielsweise alle Knoten entfernt werden, die einen Prozentsatz von unter 45 Prozent haben. Um aus dem daraus resultierenden Geflecht, das in *Abbildung 5.5* dargestellt wird, Erkenntnisse zu erlangen, mit denen sich die Fragestellung lösen ließe, müssten vorher zusätzliche Analysen durch Algorithmen des Graph Management durchgeführt werden (vgl. Abschnitt 3.3.1). Die aus *Tabelle 4.24* entnommene Klassifizierung für die Graph Compression kann deshalb bestätigt werden.

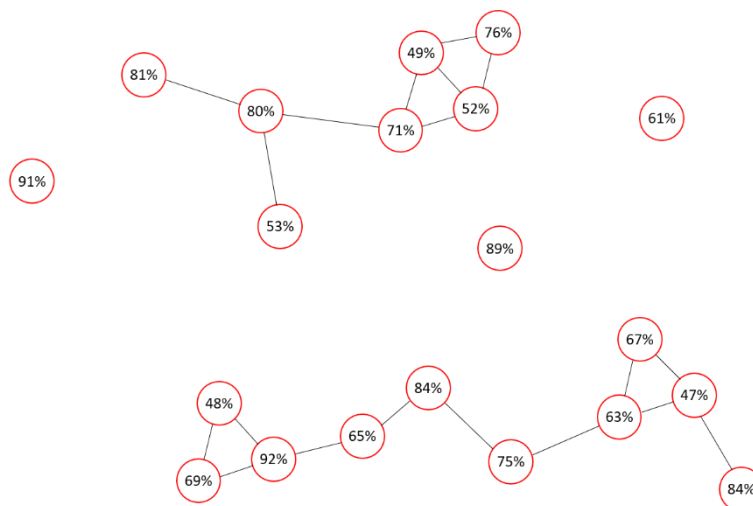


Abbildung 5.5: Graph nach Anwendung des Algorithmus der Graph Compression

Die Einteilung der Algorithmen des Graph-Clustering, der Graph-Classification sowie der Link Prediction in die vierte Klasse ist auf dieses Fallbeispiel ebenfalls übertragbar, da keine Möglichkeit identifiziert wurde, wie die Ergebnisse dieser Verfahren entweder zur direkten Beantwortung führen oder der Graph derart vereinfacht wird, dass er von Graph Management Algorithmen weiterverarbeitet werden kann (vgl. Abschnitt 3.3.2). Da es bei diesem Szenario nur einen Graphen gibt und eine Unterteilung von diesem nicht sinnvoll ist, eignen sich die Verfahren des Graph-Clustering und der Graph-Classification nicht für den Einsatz bei diesem Szenario. Der Algorithmus der Link Prediction ist ebenfalls ungeeignet, da sich eine Entdeckung von neuen Kanten aufgrund der bestehenden Netzstruktur der Heime nicht auf die Realität übertragen ließe. Außerdem ist die Entscheidung über die Reihenfolge der Impfung in den Senioren- und Altenpflegeheimen nicht auf die Kanten, sondern primär auf die Knotengewichte zurückzuführen.

5.2 Fazit

In Abschnitt 5.1 wurde erfolgreich nachgewiesen, dass die Klassifizierung der Graph Mining Algorithmen nach *Tabelle 4.24* auch auf eine ähnliche Fragestellung, deren reales Anwendungsgebiet sich von dem der Fragestellung in Abschnitt 4.2.8 unterscheidet, übertragbar ist. Die Evaluierung bezog sich dabei auf das Teilgebiet der Distributionsfeinplanung. Die Klassifizierungen der Algorithmen in Bezug auf die übrigen Fragestellungen, die auch anderen Teilgebieten des Aufgabenmodells zuzuordnen sind, könnte in zukünftigen Arbeiten anhand weiterer Fallbeispiele evaluiert werden.

Darüber hinaus sei gesagt, dass die Einteilung der Algorithmen in die unterschiedlichen Klassen (vgl. Abschnitt 4.1.6) nur für die beschriebenen Fragestellungen (vgl. Abschnitt 4.1.4) und nicht für das zugehörige Teilgebiet des Aufgabenmodells gilt (vgl. Abschnitte 2.4.2 und 2.4.3). Dies ist in der Tatsache begründet, dass die möglichen Aufgaben eines Teilgebietes so vielfältig sind, dass eine Entscheidung, welcher Algorithmus generell geeignet ist, unmöglich ist. Nach Tarjan liegt der Auswahl eines Algorithmus stets eine konkrete Aufgabenstellung zugrunde (vgl. Abschnitt 3.1).

Aus diesem Grund wird der Ansatz, die Klassifizierung der Graph Mining Algorithmen nach den unterschiedlichen Teilgebieten des Aufgabenmodells (vgl. Abschnitte 2.4.2 und 2.4.3) zu gliedern, aus verschiedenen Gründen, die sich aus der Bearbeitung der Aufgabenstellung ergeben haben, als nicht zielführend erachtet. Vielmehr sollte eine Klassifizierung nach den verschiedenen Typen einer logistischen Fragestellung gegliedert werden, die in einer zukünftigen Arbeit definiert werden sollten. Diese

Evaluierung der Methode

Fragestellungen ließen sich beispielsweise anhand der Kriterien unterscheiden, worauf sie abzielen oder ob eine zeitliche Komponente berücksichtigt wird.

Ein weiterer Faktor, der bei der Auswahl von einem Algorithmus berücksichtigt werden muss, ist das aus dem realen Anwendungsfall abstrahierte Modell. Die für jede Fragestellung definierten Modelle in Abschnitt 4.2 werden nur oberflächlich beschrieben und besitzen nicht den Grad der Komplexität von tatsächlichen Modellen. Außerdem kann mit *Tabelle 4.24* nur ein Verfahren des Graph Mining auf die Eignung überprüft werden und nicht der konkrete Algorithmus. Die Betrachtung der zahlreichen Algorithmen, die einem grundlegenden Verfahren des Graph Mining (vgl. Abschnitt 3.3.2) zuzuordnen sind, würden den Umfang dieser Arbeit übersteigen und wären gegebenenfalls in einer tiefergehenden Analyse zu berücksichtigen.

Abschließend wird auf vier wichtige, generelle Punkte beziehungsweise Hinweise eingegangen, die die logistischen Fragestellungen betreffen und während der Klassifizierung aufgefallen sind. Der erste Punkt betrifft den Graphen, der eine Grundvoraussetzung für die Anwendung von Graph Mining Algorithmen darstellt. Es ist wichtig, dass die Fragestellung einen direkten Bezug zur Supply Chain mit ihren Partnern hat, da sonst kein sinnvolles Modell abstrahiert werden kann und folglich auch keine Verfahren des Graph Mining eingesetzt werden können. Ein Negativbeispiel stellt etwa die Fragestellung für das Teilgebiet der Verfügbarkeits- und Machbarkeitsprüfung dar, bei der jeder Algorithmus in die vierte Klasse eingeteilt wurde (vgl. Abschnitt 4.2.6). Der zweite Punkt bezieht sich auf das Ziel der Fragestellung. Je detaillierter und zielgerichteter eine Frage gestellt ist, desto schwieriger lässt sich ein passender Graph abstrahieren, mit dem sich die Frage lösen lässt. Ein Beispiel für diesen Punkt ist die Frage nach der optimalen Beschaffungsmenge für ein bestimmtes Produkt unter der Berücksichtigung von Sicherheitsbeständen (vgl. Abschnitt 4.2.4). Der dritte Punkt betrifft die Tatsache, dass in einem Graphen nur sehr schwer die Zeit berücksichtigt werden kann. Realisierbar wäre dies über die Unterteilung eines Graphs in unterschiedliche Graph-Sequenzen mit gewissen Zeitschritten (vgl. Abschnitt 3.3.2). Aufgrund der resultierenden Komplexität gilt es die zeitliche Komponente bei der Formulierung von logistischen Fragestellungen möglichst außen vor zu lassen. Der letzte Punkt bezieht sich darauf, dass es nicht nur einen Weg beziehungsweise Algorithmus zur Beantwortung einer Fragestellung gibt. Vielmehr lässt sich das Problem abhängig von seiner Formulierung auch durch unterschiedliche Algorithmen lösen. Sollte dies der Fall sein, ist jedoch anzumerken, dass verschiedene Algorithmen die gleiche Fragestellung im Regelfall mit einer unterschiedlichen Effizienz lösen.

6 Zusammenfassung und Ausblick

In dieser Arbeit wurde eine Klassifizierung von Algorithmen des Graph Mining hinsichtlich der Beantwortung von konkreten logistischen Fragestellungen in Supply Chains entwickelt. Soll eine reale Problemstellung von einem Graph Mining Verfahren entweder gelöst oder so vereinfacht werden, dass sie in einem weiteren Schritt durch einen Algorithmus des Graph Management verarbeitet werden, so gibt *Tabelle 4.24* Aufschlüsse darüber, welche Verfahren dafür genutzt werden könnten. Bedingung dafür ist, dass die reale Problemstellung vergleichbar mit einer der in Abschnitt 4.1.4 beschriebenen Fragestellungen ist.

Zunächst wurden in Kapitel 2 die theoretischen Grundlagen zum Supply Chain Management erarbeitet. In den Abschnitten 2.1 und 2.3 wurde besonders auf die unterschiedlichen Definitionen der Logistik, der Supply Chain und des Supply Chain Managements eingegangen und diese miteinander verglichen, um eine saubere Begriffsdefinition zu schaffen. Die im Supply Chain Management aufkommenden Aufgaben und Fragestellungen können den unterschiedlichen Teilgebieten des in Abschnitt 2.4.1 beschriebenen Aufgabenmodells zugeordnet werden. Diese Aufgabenbereiche stützen sich auf ein funktionierendes Daten- und Informationsmanagement, welches in Abschnitt 2.2 beschrieben wurde. Abschließend wurden die Bereiche des Supply Chain Design und des Supply Chain Planning des Aufgabenmodells detailliert in den Abschnitten 2.4.2 und 2.4.3 vertieft.

Zu Beginn von Kapitel 3 wurden zum allgemeinen Verständnis die Grundlagen der Graphentheorie erläutert. In den folgenden Abschnitten 3.2 und 3.3 wurden die im Supply Chain Management geläufigen graphbasierten Modelle und die zur Problemlösung eingesetzten Graphalgorithmen beschrieben, wobei diese in das Graph Management (vgl. Abschnitt 3.3.1) und das Graph Mining (vgl. Abschnitt 3.3.2) unterteilt wurden. Da der Literatur keine klare Differenzierung dieser beiden Bereiche zu entnehmen ist, wurde sich in Abschnitt 4.1.2 mit ebendieser Abtrennung befasst. Die in diesem Abschnitt aufgestellten Thesen gilt es zu evaluieren.

Die Methode zur Klassifizierung von Graph Mining Algorithmen wurde in Kapitel 4 erarbeitet. Dafür wurde in Abschnitt 4.1.1 zunächst eine Methode der Technologiebewertung identifiziert, mit der sich die Eignung von Algorithmen zur Beantwortung verschiedener Fragestellungen überprüfen lässt. Für diese Methode wurden in Abschnitt 4.1.3 und 4.1.4 die zu klassifizierenden Graph Mining Algorithmen ausgewählt und prägnant beschrieben und die logistischen Fragestellungen formuliert. Darüber hinaus wurden in Abschnitt 4.1.5 Bewertungskriterien identifiziert und in Abschnitt 4.1.6 die Klassen beschrieben, in die die Algorithmen eingeteilt werden. In Abschnitt 4.2.1 bis 4.2.8 wurden die Algorithmen hinsichtlich der Beantwortung verschiedener Fragestellungen klassifiziert und die Ergebnisse wurden in Abschnitt 4.2.9 zusammengetragen.

Die Ergebnisse der Klassifizierung (vgl. Abschnitt 4.2) wurden in Kapitel 5 anhand eines Fallbeispiels evaluiert, das sich mit der zweiten logistischen Fragestellung der Distributionsfeinplanung aus 4.1.4 vergleichen lässt. Für die spezifische Fragestellung konnten die Ergebnisse in Abschnitt 5.1 erfolgreich validiert werden. Abschließend wurden in Abschnitt 5.2 einige generelle Erkenntnisse zu der Klassifizierung getroffen, die unter anderem ihren Nutzen beschreiben.

Es sei gesagt, dass in dieser Arbeit keine allgemeingültige Klassifizierung von Graph Mining Algorithmen in Bezug auf die Beantwortung von logistischen Fragestellungen bestimmt werden konnte. Da die Eignung eines Algorithmus zur Beantwortung einer Fragestellung stets von deren Formulierung abhängt, kann eine Klassifizierung eines Algorithmus in Bezug auf eine logistische Fragestellung nicht

Zusammenfassung und Ausblick

auf das gesamte Teilgebiet dieser Fragestellung verallgemeinert werden. So kann etwa keine Aussage darüber getroffen werden, dass der Algorithmus der Label Propagation jede Fragestellung der Distributionsplanung lösen kann. Bei weiteren Ausarbeitungen ist ein alternativer Ansatz einer Klassifizierung zu verfolgen, der sich etwa daran orientiert, was mit einer Fragestellung erfragt werden soll. Neben diesem Anknüpfungspunkt wurden bei der Ausarbeitung weitere Forschungsfragen identifiziert, die in Zukunft untersucht werden sollten.

In Abschnitt 5.1 wurde ein Teil der Klassifizierung anhand einer logistischen Fragestellung aus einem unterschiedlichen Anwendungsgebiet evaluiert. Es könnten weitere solcher Evaluierungen durchgeführt werden, damit die Ergebnisse der *Tabelle 4.24* bestätigt werden können. Soll weiterhin eine Klassifizierung verfolgt werden, die sich in die unterschiedlichen Fragestellungen der Teilgebiete unterteilt, wird es als sinnvoll erachtet, weitere logistische Fragestellungen aus den Teilgebieten des Aufgabenmodells zu betrachten und die Eignung der Graph Mining Algorithmen in Bezug auf den Einsatz in dieser Problemstellung zu untersuchen. Dadurch wird einem Anwender der Tabelle eine größere Auswahl an Fragestellungen gegeben und es ließen sich analog zu Abschnitt 4.2.9 eventuell weitere allgemeingültige Punkte zu den Algorithmen ableiten.

In Abschnitt 5.2 wurde angesprochen, dass es je nach Formulierung der logistischen Fragestellung sehr schwierig ist, die Algorithmen gewinnbringend einzusetzen. Aus diesem Grund sollte sich in einer zukünftigen Arbeit Gedanken darüber gemacht werden, wie die logistische Fragestellung so zu wählen ist, dass sie die Eigenschaften der Graph Mining Algorithmen bestmöglich berücksichtigt. Es wird empfohlen, eine Klassifizierung nach den verschiedenen Typen der Fragestellung zu unterteilen. Zur Differenzierung der verschiedenen Fragestellungen könnte etwa untersucht werden, auf was sie abzielen oder ob eine zeitliche Komponente berücksichtigt wird.

Nach dem in Abschnitt 3.1 vorgestellten Modell nach Tarjan werden die realen Anwendungsfälle zu einem graphbasierten Modell abstrahiert, auf das die Graphalgorithmen angewendet werden. Sollte ein Algorithmus ein reales Problem nicht lösen können, muss nach der in Abschnitt 3.1 beschriebenen Ansicht die Abstraktion oder das Modell angepasst werden. Aus diesem Grund sollten sich zukünftig Gedanken darüber gemacht werden, ob die in Abschnitt 3.2 vorgestellten Modelle des Supply Chain Managements für die Algorithmen des Graph Mining geeignet sind oder gegebenenfalls angepasst werden sollten. Bereits in Abschnitt 5.2 wurde konstatiert, dass einige Algorithmen zur weiteren Anwendung direkt auszuschließen sind, da sie nicht auf den vorliegenden Graphen oder das Modell angewendet werden können.

Eine weitere Forschungsfrage lautet, wie die Algorithmen des Graph Mining gegenüber anderen Verfahren wie dem Graph Management oder dem Operations Research in Hinblick auf die Beantwortung von logistischen Problemstellungen einzuordnen sind. Dabei wäre zu klären, ob es verschiedene Typen von realen Anwendungsfällen gibt, bei denen eine Methode bevorzugt eingesetzt wird und wie diese definiert sind.

Zu guter Letzt sei angemerkt, dass in dieser Arbeit keine konkreten, sondern nur die generellen Algorithmen des Graph Mining thematisiert wurden. Für jedes der in Abschnitt 3.3.2 vorgestellten Verfahren existieren zahlreiche konkrete Algorithmen, die etwa für unterschiedliche Modelle verwendet werden. Es wäre von Interesse, diese Algorithmen genauer zu untersuchen und ebenfalls zu klassifizieren. Für eine Klassifizierung von konkreten Algorithmen könnte der in Abschnitt 4.1.5 definierte

Zusammenfassung und Ausblick

Kriterienkatalog etwa um die Kriterien der geforderten Rechenkapazität oder der zur Berechnung benötigten Zeit erweitert werden.

Abkürzungsverzeichnis

ATP	Available-to-Promise
CoTP	Configure-to-Promise
CTP	Capable-to-Promise
DWH	Data Warehouse
EDI	Electronic Data Interchange
ERP	Enterprise-Resource-Planning
ETH	Eidgenössische Technische Hochschule
GC	Graph Compression
G-Cla	Graph-Classification
G-Clu	Graph-Clustering
GPM	Graph Pattern Mining
IML	Institut für Materialfluss und Logistik
IPA	Institut für Produktionstechnik und Automatisierung
KDD	Knowledge Discovery in Databases
LP	Link Prediction
L-Pro	Label-Propagation
N-Clu	Node-Clustering
RDF	Resource Description Framework
RFID	Radio Frequenz Identifikation
SCM	Supply Chain Management
SCOR	Supply Chain Operations Reference
SQL	Structured Query Language
TMS	Transport-Management-Systeme
WMS	Warehouse-Management-Systeme

Abbildungsverzeichnis

Abbildung 2.1: Beispielhafte Darstellung einer Supply Chain	5
Abbildung 2.2: Die Wissenstreppe	7
Abbildung 2.3: Aufgabenmodell des Supply Chain Managements	15
Abbildung 3.1: Beispiel für einen Sterngraphen.....	24
Abbildung 3.2: Speicherungsformen eines Graphen.....	25
Abbildung 3.3: Prozess der Forschung, Entdeckung und Entwicklung	26
Abbildung 3.4: Ein einfaches Transportnetzwerk.....	29
Abbildung 3.5: Rundreise und Tourenplan.....	32
Abbildung 3.6: Knowledge Discovery in Databases-Prozess.....	36
Abbildung 3.7: Beispiel für das Extrahieren häufiger Subgraphen	39
Abbildung 3.8: Die Aufgabenbereiche der Label Propagation und Graph Classification	42
Abbildung 4.1: Angepasster Prozess zur Forschung, Entdeckung und Entwicklung	46
Abbildung 4.2: Einordnung der Ergebnisse des Graph Mining und Graph Management	48
Abbildung 5.1: Ursprungsgraph des Fallbeispiels	79
Abbildung 5.2: Graph nach Anwendung des Algorithmus der Label Propagation.....	81
Abbildung 5.3: Graph nach Anwendung des Algorithmus des Graph Pattern Mining.....	81
Abbildung 5.4: Graph nach Anwendung des Algorithmus des Node-Clustering	82
Abbildung 5.5: Graph nach Anwendung des Algorithmus der Graph Compression.....	83

Tabellenverzeichnis

Tabelle 2.1: Planungsebenen der Distributionsplanung	20
Tabelle 3.1: Mögliche Bedeutungen der Objekte in der Netzwerkmodellierung	27
Tabelle 4.1: Algorithmen und ihre Eigenschaften.....	47
Tabelle 4.2: Klassen und deren Kriterien zur Einteilung von Algorithmen des Graph Mining	53
Tabelle 4.3: Abkürzungen der Graph Mining Algorithmen	53
Tabelle 4.4: Kriterienausprägungen der Algorithmen für die erste Frage der strategischen Netzwerkgestaltung	55
Tabelle 4.5: Klassifizierung der Algorithmen in Bezug auf die erste Frage zur strategischen Netzwerkgestaltung	55
Tabelle 4.6: Kriterienausprägungen der Algorithmen für die zweite Frage der strategischen Netzwerkgestaltung	57
Tabelle 4.7: Klassifizierung der Algorithmen in Bezug auf die zweite Frage zur strategischen Netzwerkgestaltung	57
Tabelle 4.8: Kriterienausprägungen der Algorithmen für die Frage der Bedarfsplanung	59
Tabelle 4.9: Klassifizierung der Algorithmen in Bezug auf die Frage der Bedarfsplanung.....	59
Tabelle 4.10: Kriterienausprägungen der Algorithmen für die Frage der Netzwerkplanung	61
Tabelle 4.11: Klassifizierung der Algorithmen in Bezug auf die Frage der Netzwerkplanung	61
Tabelle 4.12: Kriterienausprägungen der Algorithmen für die Frage der Beschaffungsplanung.....	62
Tabelle 4.13: Klassifizierung der Algorithmen in Bezug auf die Frage der Beschaffungsplanung	63
Tabelle 4.14: Kriterienausprägungen der Algorithmen für die Frage der Distributionsplanung.....	64
Tabelle 4.15: Klassifizierung der Algorithmen in Bezug auf die Frage der Distributionsplanung	65
Tabelle 4.16: Kriterienausprägungen der Algorithmen für die Frage der Verfügbarkeits- und Machbarkeitsprüfung.....	66
Tabelle 4.17: Klassifizierung der Algorithmen in Bezug auf die Frage der Verfügbarkeits- und Machbarkeitsprüfung.....	66
Tabelle 4.18: Kriterienausprägungen der Algorithmen für die Frage der Beschaffungsfineplanung... ..	68
Tabelle 4.19: Klassifizierung der Algorithmen in Bezug auf die Frage der Beschaffungsfineplanung.....	68
Tabelle 4.20: Kriterienausprägungen der Algorithmen für die erste Frage der Distributionsfineplanung	70
Tabelle 4.21: Klassifizierung der Algorithmen in Bezug auf die erste Frage der Distributionsfineplanung	70
Tabelle 4.22: Kriterienausprägungen der Algorithmen für die zweite Frage der Distributionsfineplanung	72
Tabelle 4.23: Klassifizierung der Algorithmen in Bezug auf die zweite Frage der Distributionsfineplanung	72
Tabelle 4.24: Überblick der Klassifizierungen von Graph Mining Algorithmen in Bezug auf logistische Fragestellungen des Aufgabenmodells.....	74
Tabelle 4.25: Häufigkeit der Klassenzuteilungen je Graph Mining Algorithmus.....	76

Literaturverzeichnis

- Aggarwal, Charu C. *Data Mining: The Textbook*. Basel: Springer International Publishing Switzerland, 2015.
- Aggarwal, Charu C. „Graph Clustering.“ In *Encyclopedia of Machine Learning and Data Mining*, von Claude Sammut und Geoffrey I. Webb, S. 570-579. New York, New York: Springer Science+Business Media, 2017.
- Aggarwal, Charu C., und Haixun Wang. „A Survey of Clustering Algorithms for Graph Data.“ In *Managing and Mining Graph Data*, von Charu C. Aggarwal und Haixun Wang, S. 275-302. New York, New York: Springer Science+Business Media, 2010c.
- Aggarwal, Charu C., und Haixun Wang. „An Introduction to Graph Data.“ In *Managing and Mining Graph Data*, von Charu C. Aggarwal und Haixun Wang, S. 1-12. New York, New York: Springer Science+Business Media, LLC, 2010a.
- Aggarwal, Charu C., und Haixun Wang. „Graph Data Management and Mining: A Survey of Algorithms and Applications.“ In *Managing and Mining Graph Data*, von Charu C. Aggarwal und Haixun Wang, S. 13-68. New York, New York: Springer Science+Business Media, LLC, 2010b.
- Alicke, Knut. *Planung und Betrieb von Logistiknetzwerken: Unternehmensübergreifendes Supply Chain Management*. 2. Auflage. Heidelberg: Springer-Verlag Berlin Heidelberg, 2005.
- Angles, Renzo, und Claudio Gutierrez. „An Introduction to Graph Data Management.“ In *Graph Data Management: Fundamental Issues and Recent Developments*, von George Fletcher, Jan Hidders und Josep Lluís Larriba-Pey, S. 1-32. Cham: Springer International Publishing AG, 2018.
- Arnolds, Hans, Franz Heege, Carsten Röh, und Werner Tussing. *Materialwirtschaft und Einkauf: Grundlagen - Spezialthemen - Übungen*. 13. Auflage. Wiesbaden: Springer Fachmedien Wiesbaden, 2016.
- Atastina, I., B. Sitohang, G. A. P. Saptawati, und V. S. Moertini. „A Review of Big Graph Mining Research.“ *IOP Conference Series: Materials Science and Engineering 180*. Bandung: IOP Publishing Ltd, 2017. S. 1-10.
- Beckmann, Holger. *Prozessorientiertes Supply Chain Engineering: Strategien, Konzepte und Methoden zur modellbasierten Gestaltung*. Wiesbaden: Springer Fachmedien Wiesbaden, 2012.
- Besta, Maciej, und Torsten Hoefler. *Survey and Taxonomy of Lossless Graph Compression and Space-Efficient Graph Representations*. Preprint, arXiv, 2018.
- Bousonville, Thomas. *Logistik 4.0: Die digitale Transformation der Wertschöpfungskette*. Wiesbaden: Springer Fachmedien Wiesbaden GmbH, 2017.
- Bramer, Max. *Principles of Data Mining*. London: Springer-Verlag London Limited, 2007.
- Bunke, Horst, und Michael Neuhäus. „Graph Matching - exact and error-tolerant methods and the automatic learning of edit costs.“ In *Mining Graph Data*, von Diane J. Cook und Lawrence B. Holder, S. 17-34. Hoboken, New Jersey: John Wiley & Sons, Inc., 2007.

Zusammenfassung und Ausblick

- Chakrabarti, Deepayan. „Graph Mining.“ In *Encyclopedia of Machine Learning and Data Mining*, von Claude Sammut und Geoffrey I. Webb, S. 581-584. New York, New York: Springer Science+Business Media, 2017.
- Cheng, Hong, Xifeng Yan, und Jiawei Han. „Mining Graph Patterns.“ In *Managing and Mining Graph Data*, von Charu C. Aggarwal und Haixun Wang, S. 365-392. New York, New York: Springer Science+Business Media, 2010.
- Chopra, Sunil, und Peter Meindl. *Supply Chain Management: Strategy, Planning, and Operation*. 6. Auflage. Harlow, Essex: Pearson Education Limited, 2016.
- Christopher, Martin. *Logistics & Supply Chain Management*. 4. Auflage. Harlow, Essex: Pearson Education Limited, 2011.
- Cleve, Jürgen, und Uwe Lämmel. *Data Mining*. 3. Auflage. Berlin/Boston: Walter de Gruyter GmbH, 2020.
- Domschke, Wolfgang. *Logistik: Transport*. 5. Auflage. München: Oldenbourg Wissenschaftsverlag GmbH, 2007.
- Domschke, Wolfgang, Andreas Drexl, Gabriela Mayer, und Giorgi Tadumadze. „Betriebliche Standortplanung.“ In *Planung logistischer Systeme*, von Horst Tempelmeier, S. 1-28. Berlin: Springer-Verlag GmbH Deutschland, 2018.
- Domschke, Wolfgang, und Andreas Drexl. *Logistik: Standorte*. 4. Ausgabe. München: R. Oldenbourg Verlag GmbH, 1996.
- Domschke, Wolfgang, und Armin Scholl. *Grundlagen der Betriebswirtschaftslehre: Eine Einführung aus entscheidungsorientierter Sicht*. 3. Auflage. Heidelberg: Springer-Verlag Berlin Heidelberg, 2005.
- . *Logistik: Rundreisen und Touren*. 5. Auflage. München: Oldenbourg Wissenschaftsverlag GmbH, 2010.
- Dong, Liyan, Yongli Li, Han Yin, Huang Le, und Mao Rui. „The Algorithm of Link Prediction on Social Network.“ *Mathematical Problems in Engineering*, 17. September 2013: S. 1-7.
- Dörfler, Willibald, und Jörg Mühlbacher. *Graphentheorie für Informatiker*. Berlin: Walter de Gruyter & Co., 1973.
- Even, Shimon, und Guy Even. *Graph Algorithms*. 2. Auflage. Herausgeber: Guy Even. New York, New York: Cambridge University Press, 2012.
- Fasel, Daniel. „Übersicht über NoSQL-Technologien und -Datenbanken.“ In *Big Data: Grundlagen, Systeme und Nutzungspotenziale*, von Daniel Fasel und Andreas Meier, Herausgeber: Andreas Meier, S. 109-138. Wiesbaden: Springer Fachmedien Wiesbaden, 2016.
- Fayyad, Usama, Gregory Piatetsky-Shapiro, und Padhraic Smyth. „From Data Mining to Knowledge Discovery in Databases.“ *AI Magazine*, 15. März 1996: S. 37-54.

Zusammenfassung und Ausblick

- Fleischmann, Bernhard. „Begriffliche Grundlagen.“ In *Handbuch Logistik*, von Dieter Arnold, Heinz Isermann, Axel Kuhn, Horst Tempelmeier und Kai Furmans, Herausgeber: Kai Furmans, S. 3-12. Heidelberg: Springer-Verlag Berlin Heidelberg, 2008.
- Fleischmann, Bernhard. „Begriffliche Grundlagen der Logistik.“ In *Begriffe der Logistik, logistische Systeme und Prozesse*, von Horst Tempelmeier, Herausgeber: Horst Tempelmeier, S. 1-16. Berlin: Springer-Verlag GmbH Deutschland, 2018.
- Fleischmann, Bernhard, und Herbert Kopfer. „Transport- und Tourenplanung.“ In *Planung logistischer Systeme*, von Horst Tempelmeier, S. 71-98. Berlin: Springer-Verlag GmbH Deutschland, 2018.
- Frawley, William J., Gregory Piatetsky-Shapiro, und Christopher J. Matheus. „Knowledge Discovery in Databases: An Overview.“ *AI Magazine*, 15. September 1992: S. 57-70.
- Fürstenberg, Kay, und Christopher Kirsch. „Intelligente Sensorik als Grundbaustein für cyber-physische Systeme in der Logistik.“ In *Handbuch Industrie 4.0 Bd.3: Logistik*, von Birgit Vogel-Heuser, Thomas Bauernhansel und Michael ten Hompel, Herausgeber: Michael ten Hompel, S. 271-298. Berlin: Springer-Verlag GmbH Deutschland, 2017.
- Garg, Vikas K., und Tommi Jaakkola. „Solving graph compression via optimal transport.“ *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*. Vancouver, 2019. S. 1-17.
- Göpfert, Ingrid. „Die Anwendung der Zukunftsforschung für die Logistik.“ In *Logistik der Zukunft - Logistics for the Future*, von Ingrid Göpfert, S. 39-100. Wiesbaden: Springer Fachmedien Wiesbaden, 2016.
- Haag, Christoph, Günther Schuh, Jennifer Kreysa, und Kristin Schmelter. „Technologiebewertung.“ In *Technologiemangement*, von Günther Schuh und Sascha Klappert, S. 309-366. Heidelberg: Springer-Verlag Berlin Heidelberg, 2011.
- Han, Jiawei, Micheline Kamber, und Jian Pei. *Data Mining: Concepts and Techniques*. 3. Auflage. Waltham, Massachusetts: Morgan Kaufmann Publishers, 2012.
- Hand, David, Heikki Mannila, und Padhraic Smyth. *Principles of Data Mining*. Cambridge, Massachusetts: The MIT Press, 2001.
- Hausladen, Iris. *IT-gestützte Logistik: Systeme - Prozesse - Anwendungen*. 4. Auflage. Wiesbaden: Springer Fachmedien Wiesbaden GmbH, 2020.
- Hellingrath, Bernd, Tobias Hegmanns, Jan-Christoph Maaß, und Michael Toth. „Prozesse in Logistiknetzwerken - Supply Chain Management.“ In *Handbuch Logistik*, von Dieter Arnold, Heinz Isermann, Axel Kuhn, Horst Tempelmeier und Kai Furmans, Herausgeber: Kai Furmans, S. 458-486. Heidelberg: Springer-Verlag Berlin Heidelberg, 2008.
- Hellingrath, Bernd, und Markus Witthaut. *Modellierungsrahmen für Dispositionsstrategien in GNL*. Fachbericht, Dortmund: TU Dortmund, 2004.
- Hoffmeister, Wolfgang. *Investitionsrechnung und Nutzwertanalyse: Eine entscheidungsorientierte Darstellung mit vielen Beispielen und Übungen*. 2. Auflage. Berlin: Berliner Wissenschafts-Verlag GmbH, 2008.

Zusammenfassung und Ausblick

- Holder, Lawrence B., und Diane J. Cook. „Introduction.“ In *Mining Graph Data*, von Diane J. Cook und Lawrence B. Holder, S. 1-14. Hoboken, New Jersey: John Wiley & Sons, Inc., 2007.
- Huber, Andreas, und Klaus Laverentz. *Logistik*. 2. Auflage. München: Verlag Franz Vahlen GmbH, 2019.
- Klaus, Peter, Winfried Krieger, und Michael Krupp. *Gabler Lexikon Logistik: Management logistischer Netzwerke und Flüsse*. 5. Auflage. Herausgeber: Michael Krupp. Wiesbaden: Gabler Verlag | Springer Fachmedien Wiesbaden, 2012.
- Kowalik, Łukasz. „Adjacency queries in dynamic sparse graphs.“ *Information Processing Letters*, 31. Mai 2007: S. 191-195.
- Krumke, Sven Oliver, und Hartmut Noltemeier. *Graphentheoretische Konzepte und Algorithmen*. 3. Auflage. Wiesbaden: Vieweg+Teubner Verlag | Springer Fachmedien Wiesbaden, 2012.
- Krupp, Thomas, und Johannes Wolf. „Grundlagen und Bedeutung der Informationssysteme in der Logistik.“ In *Praxishandbuch IT-Systeme in der Logistik: Von der Speditionssoftware bis zur integrierten Supply-Chain-Planung*, von Thomas Krupp, Rainer Paffrath und Johannes Wolf, S. 15-27. Hamburg: DVV Media Group, 2010.
- Kuhn, Axel. „Prozessmodelle, -ketten und -netze.“ In *Handbuch Logistik*, von Dieter Arnold, Heinz Isermann, Axel Kuhn, Horst Tempelmeier und Kai Furmans, Herausgeber: Kai Furmans, S. 224-228. Heidelberg: Springer-Verlag Berlin Heidelberg, 2008.
- Kuhn, Axel, und Bernd Hellingrath. *Supply Chain Management: Optimierte Zusammenarbeit in der Wertschöpfungskette*. Heidelberg: Springer-Verlag Berlin Heidelberg, 2002.
- Larson, Paul D., Richard F. Poist, und Árni Halldórsson. „Perspectives on Logistics vs. SCM: A Survey of SCM Professionals.“ *Journal of Business Logistics*, März 2007: S. 1-24.
- Lü, Linyuan, und Tao Zhou. „Link Prediction in complex networks: A survey.“ *Physica A*, 15. März 2011: S. 1150-1170.
- Mahlmann, Peter, und Christian Schindelbauer. *Peer-to-Peer-Netzwerke*. Heidelberg: Springer-Verlag Berlin Heidelberg, 2007.
- Mattfeld, Dirk, und Richard Vahrenkamp. *Logistiknetzwerke: Modelle für Standortwahl und Tourenplanung*. 2. Auflage. Wiesbaden: Springer Fachmedien Wiesbaden, 2014.
- Meier, Andreas, und Henrik Stormer. *eBusiness & eCommerce: Management der digitalen Wertschöpfungskette*. 3. Auflage. Heidelberg: Springer-Verlag Berlin Heidelberg, 2012.
- Meier, Andreas, und Michael Kaufmann. *SQL- & NoSQL-Datenbanken*. 8. Auflage. Heidelberg: Springer-Verlag Berlin Heidelberg, 2016.
- Mertens, Peter, Freimut Bodendorf, Wolfgang König, Arnold Picot, Matthias Schumann, und Thomas Hess. *Grundzüge der Wirtschaftsinformatik*. 11. Auflage. Heidelberg: Springer-Verlag Berlin Heidelberg, 2012.
- Muchna, Claus, Hans Brandenburg, Johannes Fottner, und Jens Gutermuth. *Grundlagen der Logistik: Begriffe, Strukturen und Prozesse*. Wiesbaden: Springer Fachmedien Wiesbaden GmbH, 2018.

Zusammenfassung und Ausblick

- Nettleton, David F. „Survey: Data mining of social networks represented as graphs.“ *Computer Science Review*, 12. Februar 2013: S. 1-34.
- Noltemeier, Hartmut. *Graphentheorie: mit Algorithmen und Anwendungen*. Berlin: Walter de Gruyter & Co., 1976.
- North, Klaus. *Wissensorientierte Unternehmensführung: Wertschöpfung durch Wissen*. 5. Auflage. Wiesbaden: Gabler Verlag | Springer Fachmedien Wiesbaden GmbH, 2011.
- Pfohl, Hans-Christian. *Logistikmanagement: Konzeption und Funktionen*. 3. Auflage. Heidelberg: Springer-Verlag Berlin Heidelberg, 2016.
- . *Logistiksysteme: Betriebswirtschaftliche Grundlagen*. 9. Auflage. Berlin: Springer-Verlag GmbH, 2018.
- Piatetsky-Shapiro, Gregory, und Christopher J. Matheus. „The Interestingness of Deviations.“ *AAAIWS'94: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*. Seattle: AAAI Press, 1994. S. 25-36.
- Prabhu, S., und N. Venkatesan. *Data Mining and Warehousing*. New Delhi: New Age International (P) Ltd., Publishers, 2007.
- Ramraj, T., und R. Prabhkar. „Frequent Subgraph Mining Algorithms - A Survey.“ *Procedia Computer Science* 47, 17. Mai 2015: S. 197-204.
- Robert Koch-Institut. „Beschluss der STIKO zur 1. Aktualisierung der COVID-19-Impfempfehlung und die dazugehörige wissenschaftliche Begründung.“ *Epidemiologisches Bulletin*, 14. Januar 2021.
- Runkler, Thomas A. *Data Mining: Methoden und Algorithmen intelligenter Datenanalyse*. Wiesbaden: Vieweg+Teubner | GWV Fachverlage GmbH, 2010.
- Saake, Gunter, Kai-Uwe Sattler, und Andreas Heuer. *Datenbanken: Konzepte und Sprachen*. 6. Auflage. Frechen: mitp Verlags GmbH & Co. KG, 2018.
- Sadalage, Pramod J., und Martin Fowler. *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*. Upper Saddle River, New Jersey: Pearson Education, Inc., 2013.
- Salem, Rashed, Wafaa Abdel Moneim, und Mohamed Hassan. „Graph Mining Techniques for Graph Clustering: Starting Point.“ *Journal of Theoretical and Applied Information Technology*, 15. August 2019: S. 4075-4092.
- Sammut, Claude, und Geoffrey I. Webb. *Encyclopedia of Machine Learning and Data Mining*. 2. Auflage. New York, New York: Springer Science+Business Media, 2017.
- Schaeffer, Satu Elisa. „Survey: Graph clustering.“ *Computer Science Review*, August 2007: S. 27-64.
- Scheel, Joachim. „Logistiknetzwerk.“ In *Gabler Lexikon Logistik: Management logistischer Netzwerke und Flüsse*, von Peter Klaus, Winfried Krieger und Michael Krupp, Herausgeber: Michael Krupp, S. 399. Wiesbaden: Gabler Verlag | Springer Fachmedien Wiesbaden, 2012.
- Scholl, Armin. „Grundlagen der modellgestützten Planung.“ In *Handbuch Logistik*, von Dieter Arnold, Heinz Isermann, Axel Kuhn, Horst Tempelmeier und Kai Furmans, Herausgeber: Kai Furmans, S. 35-43. Heidelberg: Springer-Verlag Berlin Heidelberg, 2008.

Zusammenfassung und Ausblick

- Schönsleben, Paul. *Integriertes Logistikmanagement: Operations und Supply Chain Management innerhalb des Unternehmens und unternehmensübergreifend*. 7. Auflage. Heidelberg: Springer Verlag Berlin Heidelberg, 2016.
- Schuh, Günther, Sascha Klappert, und Torsten Moll. „Ordnungsrahmen Technologiemanagement.“ In *Technologiemanagement*, von Günther Schuh und Sascha Klappert, S. 11-32. Heidelberg: Springer-Verlag Berlin Heidelberg, 2011.
- Schulte, Christof. *Logistik: Wege zur Optimierung der Supply Chain*. 7. Auflage. München: Verlag Franz Vahlen GmbH, 2017.
- Schulz-Spathelf, Johannes. „Informationssysteme, integrierte logistische.“ In *Gabler Lexikon Logistik: Management logistischer Netzwerke und Flüsse*, von Peter Klaus, Winfried Krieger und Michael Krupp, Herausgeber: Michael Krupp, S. 236-243. Wiesbaden: Gabler Verlag | Springer Fachmedien Wiesbaden, 2012.
- Shah, Rushabh Jitendrakumar. *Graph Compression Using Pattern Mining Techniques*. Fachbericht, Carson: Department of Computer Science California State University Dominguez Hills, 2018.
- Simchi-Levi, David, Philip Kaminsky, und Simchi-Levi, Edith. *Designing and Managing the Supply Chain: Concepts, Strategies, and Case Studies*. New York, New York: The McGraw-Hill Companies, Inc., 2000.
- Steger, Angelika. *Diskrete Strukturen 1: Kombinatorik, Graphentheorie, Algebra*. 1. korrigierter Nachdruck. Heidelberg: Springer-Verlag Berlin Heidelberg, 2002.
- Steglich, Mike, Dieter Feige, und Peter Klaus. *Logistik-Entscheidungen: Modellbasierte Entscheidungsunterstützung in der Logistik mit LogisticsLab*. 2. Auflage. Berlin/Boston: Walter de Gruyter GmbH, 2016.
- Stich, Volker, Dirk Oedekoven, und Tobias Brosze. „Informationssysteme für das Logistikmanagement.“ In *Logistikmanagement*, von Günther Schuh und Volker Stich, S. 257-304. Heidelberg: Springer-Verlag Berlin Heidelberg, 2013.
- Stich, Volker, Jerome Quick, und Stefan Cuber. „Konfiguration logistischer Netzwerke.“ In *Logistikmanagement*, von Günther Schuh und Volker Stich, S. 35-76. Heidelberg: Springer-Verlag Berlin Heidelberg, 2013.
- Takigawa, Ichigaku, und Hiroshi Mamitsuka. „Graph mining: procedure, application to drug discovery and recent advances.“ *Drug Discovery Today*, Januar 2013: S. 50-57.
- Tarjan, Robert. „Problems in Data Structures and Algorithms.“ In *Graph Theory, Combinatorics and Algorithms: Interdisciplinary Applications*, von Martin Charles Golumbic und Irith Ben-Arroyo Hartmann, S. 17-40. New York, New York: Springer Science + Business Media, Inc., 2005.
- ten Hompel, Michael, und Volker Heidenblut. *Taschenlexikon Logistik: Abkürzungen, Definitionen und Erläuterungen der wichtigsten Begriffe aus Materialfluss und Logistik*. 3. Auflage. Herausgeber: Michael ten Hompel. Heidelberg: Springer-Verlag Berlin Heidelberg, 2011.

Zusammenfassung und Ausblick

- Tsuda, Koji, und Hiroto Saigo. „Graph Classification.“ In *Managing and Mining Graph Data*, von Charu C. Aggarwal und Haixun Wang, S. 337-364. New York, New York: Springer Science+Business Media, 2010.
- Turau, Volker, und Christoph Weyer. *Algorithmische Graphentheorie*. 4. Auflage. Berlin/Boston: Walter de Gruyter GmbH, 2015.
- Vahrenkamp, Richard, und Herbert Kotzab. *Logistik: Management und Strategien*. 7. Auflage. München: Oldenbourg Wissenschaftsverlag GmbH, 2012.
- Vastag, Alex. „Planung der Distribution.“ In *Handbuch Logistik*, von Dieter Arnold, Heinz Isermann, Axel Kuhn, Horst Tempelmeier und Kai Furmans, Herausgeber: Kai Furmans, S. 423-438. Heidelberg: Springer-Verlag Berlin Heidelberg, 2008.
- VDI-Fachbereich Fabrikplanung und -betrieb. *VDI 3633 Blatt 1: Simulation von Logistik-, Materialfluss- und Produktionssystemen - Grundlagen*. Technische Regel, Düsseldorf: VDI-Gesellschaft Produktion und Logistik, 2014.
- Volkman, Lutz. *Graphen und Digraphen: Eine Einführung in die Graphentheorie*. Wien/New York: Springer-Verlag, 1991.
- Weber, Jürgen. *Logistikkostenrechnung: Kosten-, Leistungs- und Erlösinformationen zur erfolgsorientierten Steuerung der Logistik*. 3. Auflage. Heidelberg: Springer-Verlag Berlin Heidelberg, 2012.
- Weber, Jürgen. *Zur Bildung und Strukturierung spezieller Betriebswirtschaftslehren*. WHU-Forschungspapier Nr. 27, Vallendar: Wissenschaftliche Hochschule für Unternehmensführung (WHU), 1994.
- Wehberg, Götz. „Big Data - Mustererkennung als Erfolgsfaktor der Logistik 4.0.“ In *Handbuch Industrie 4.0 Bd. 3: Logistik*, von Birgit Vogel-Heuser, Thomas Bauernhansel und Michael ten Hompel, Herausgeber: Michael ten Hompel, S. 377-392. Berlin: Springer-Verlag GmbH Deutschland, 2017.
- Werner, Hartmut. *Supply Chain Management: Grundlagen, Strategien, Instrumente und Controlling*. 6. Auflage. Wiesbaden: Springer Fachmedien Wiesbaden GmbH, 2017.
- Wiese, Lena. *Advanced Data Management*. Berlin/Boston: Walter de Gruyter GmbH, 2015.
- Yan, Xifeng, und Jiawei Han. „Discovery of Frequent Substructures.“ In *Mining Graph Data*, von Diane J. Cook und Lawrence B. Holder, S. 99-116. Hoboken, New Jersey: John Wiley & Sons, Inc., 2007.
- Yao, Lin, Luning Wang, Lv Pan, und Kai Yao. „Link Prediction Based on Common-Neighbors for Dynamical Social Network.“ *Procedia Computer Science*, 12. Mai 2016: S. 82-89.
- Yu, Jeffrey Xu, und Jiefeng Cheng. „Graph Reachability Queries: A Survey.“ In *Managing and Mining Graph Data*, von Charu C. Aggarwal und Haixun Wang, S. 181-216. New York, New York: Springer Science+Business Media, LLC, 2010.

Zusammenfassung und Ausblick

Zangemeister, Christof. *Nutzwertanalyse in der Systemtechnik: Eine Methodik zur multidimensionalen Bewertung und Auswahl von Projektalternativen*. 2. Auflage. München: Wittemannsche Buchhandlung, 1976.

Zheng, Xuebin, Bingxin Zhou, Yu Guang Wang, und Xiaosheng Zhuang. *Decimated Framelet System on Graphs and Fast G-Framelet Transforms*. Preprint, ArXiv, 2020.

Zhou, Fang. *Graph Compression*. Fachbericht, Helsinki: Department of Computer Science and Helsinki Institute for Information Technology HIIT, 2010, S. 1-12.

Zsifkovits, Helmut E. *Logistik*. 2. Auflage. Konstanz und München: UVK Verlagsgesellschaft mbH, 2018.