

Visualisierungsinstitut der Universität Stuttgart  
Universität Stuttgart  
Universitätsstraße 38  
D-70569 Stuttgart

Bachelor Thesis Nr. 17

# **Visualisierung von unscharfen Gruppenzugehörigkeiten innerhalb von Netzwerken**

Thomas Reinhardt

<b>Studiengang:</b>	Informatik
<b>Prüfer:</b>	Prof. Dr. Daniel Weiskopf
<b>Betreuer:</b>	Dipl. Ing. Corinna Vehlow
<b>begonnen am:</b>	18. Juni 2012
<b>beendet am:</b>	20. November 2012
<b>CR-Klassifikation:</b>	E.1, H.3.3, D.2.2



## **Kurzfassung**

Die Graphvisualisierung ist die fachübergreifende Disziplin, die Beziehungen zwischen Elementen visualisiert. Sie bildet ein effizientes Werkzeug für die Analyse abstrakter Daten. Einzelne Elemente lassen sich dabei unter Verwendung spezieller Metriken zu Mengen gruppieren. Die Gruppen besitzen spezielle Eigenschaften bzw. stehen untereinander sehr in Relation. In den vergangenen Jahrzehnten wurden diese Gruppen meist mit Hilfe einer eindeutigen Zuordnung der Elemente in der Gruppen gebildet. Weiterentwicklungen der Detektionsalgorithmen in den vergangenen Jahren ermöglichen nun aber auch eine Betrachtung unscharfer Zugehörigkeiten. Für diese sind jedoch die verwendeten Visualisierungsansätze nicht ohne Anpassungen anwendbar.

Ziel der vorliegenden Bachelorarbeit ist es, Methoden zur Visualisierung von unscharfen Gruppenzugehörigkeiten innerhalb von Netzwerken zu entwickeln. Es wird dazu zuerst der verwendete Fuzzy-Clustering-Algorithmus vorgestellt. Auf den Ergebnissen des Algorithmus aufbauend, werden neue Ansätze für Visualisierung der unscharfen Gruppenzugehörigkeiten erarbeitet. Sie werden abschließend in zwei Fallstudien angewandt und bewertet.



# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>7</b>
<b>2. Grundlagen</b>	<b>9</b>
2.1. Unscharfe Mengenlehre . . . . .	9
2.2. Graphen . . . . .	10
2.2.1. Definitionen . . . . .	10
2.2.2. Datenstrukturen . . . . .	11
2.2.3. Gruppen . . . . .	13
2.2.4. Soziale Netzwerke . . . . .	14
2.3. Visualisierung . . . . .	14
<b>3. Visualisierung</b>	<b>17</b>
3.1. Ansätze der Visualisierung . . . . .	17
3.1.1. Motivation . . . . .	17
3.1.2. Graphzeichnen . . . . .	19
3.1.3. Visualisierung . . . . .	23
3.1.4. Navigation . . . . .	24
3.2. Auswahl verwandter Arbeiten . . . . .	26
3.3. Algorithmen zur Detektion unscharfer Gruppen . . . . .	29
<b>4. Visualisierung unscharfer Gruppenzugehörigkeiten</b>	<b>33</b>
4.1. Allgemeine Übersicht . . . . .	34
4.2. Kreis-Ansicht . . . . .	38
4.3. Aggregierung . . . . .	39
4.4. Einzel-Ansicht . . . . .	41
4.5. Kombination . . . . .	43
<b>5. Evaluierung</b>	<b>45</b>
5.1. Anwendung auf reale Datensätze . . . . .	46
5.1.1. American Football Network . . . . .	46
5.1.2. Les Misérables . . . . .	49
5.2. Evaluierung der Ansätze . . . . .	52
<b>6. Zusammenfassung und Ausblick</b>	<b>57</b>
<b>A. Anhang</b>	<b>59</b>
<b>Literaturverzeichnis</b>	<b>61</b>

# Abbildungsverzeichnis

---

1.1.	Beispiel – Unübersichtliche Darstellung eines Netzwerkes . . . . .	8
2.1.	Scharfe und unscharfe Definition einer „angenehmen“ Temperatur . . . . .	9
2.2.	Beispiel – Darstellung von Graphen . . . . .	12
2.3.	Gruppenbildung in einem Netzwerk . . . . .	13
2.4.	Visueller Analyseprozess (Keim <i>et al.</i> ) . . . . .	15
3.1.	Hierarchisches- / nicht-hierarchisches Netzwerk . . . . .	18
3.2.	Layoutalgorithmen des Graphzeichnen . . . . .	20
3.3.	Beispiel Visualisierung überlappender Gruppen . . . . .	24
3.4.	Fokus+Kontext – Fisheye Verkrümmung . . . . .	26
3.5.	Prinzip der Kantenbündelung nach Holten . . . . .	27
3.6.	Beispiel – Anwendung der Kantenbündelung . . . . .	27
3.7.	Beispiel – Anwendung der Darstellung per Adjazenzmatrix . . . . .	29
4.1.	Beispiel – Anwendung neuer Distanznorm . . . . .	35
4.2.	Prinzip der Färbung eines Knotens nach dominierender Zugehörigkeit . . . . .	36
4.3.	Prinzip der Richtung der Färbung eines Knotens . . . . .	36
4.4.	Prinzip der Färbung einer Kante . . . . .	37
4.5.	Beispiel – Benutzerinteraktion . . . . .	37
4.6.	Vergleich verschiedener Kreislayoutvariationen . . . . .	38
4.7.	Prinzip der Aggregierung von Knoten . . . . .	40
4.8.	Beispiel – Anwendung der Aggregierung mit verschiedenen Grenzwerten . . . . .	41
4.9.	Beispiel – Anwendung der Einzel-Ansicht mit verschiedenen Grenzwerten . . . . .	42
4.10.	Beispiel – Anwendung des semantischen Zoom als Interaktion mit Benutzer . . . . .	44
5.1.	American College Football – Kreis- & kombinierte-Ansicht . . . . .	47
5.2.	American College Football – Aggregierte-Ansicht . . . . .	48
5.3.	Les Misérables – Widersprüchliche Ergebnisse 1 . . . . .	51
5.4.	Les Misérables – Widersprüchliche Ergebnisse 2 . . . . .	51
A.1.	Vergrößerte Version des Beispieles in Abbildung 3.7 . . . . .	60

# Tabellenverzeichnis

---

5.1.	Zusammenfassung der Evaluation . . . . .	55
------	--	----

# 1. Einleitung

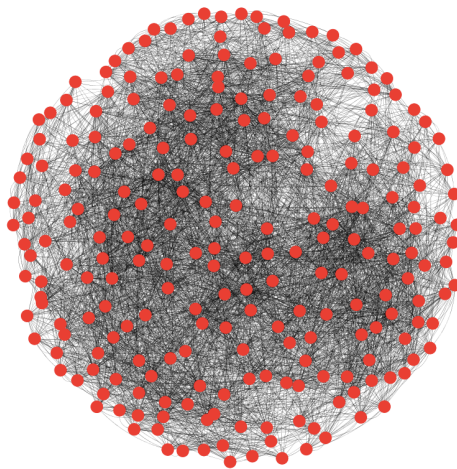
*„Ein Bild ist tausend Worte wert.“*

- Chinesisches Sprichwort.

Die Analyse der Eigenschaften von Netzwerken, ist in den vergangenen zwei Jahrzehnten immer mehr in den Fokus vieler Forschungsfelder gerückt [New10]. Verbesserungen in der Erfassung relationaler Daten, aber auch die stetig wachsenden frei zu Verfügung stehende Netzwerke, wie *Facebook* oder *Twitter* fordern eine immer bessere Aufbereitung bzw. sinnvollere Darstellung der Daten. Eine alleinige Betrachtung der Daten als numerische Werte in Form von Tabellen, ist jedoch keine sinnvolle Methodik, für den Erhalt aussagekräftiger Ergebnisse. Die graphische Darstellung der Netzwerke, sowie die visuelle Abbildung bestimmter Merkmale ist daher ein elementares Werkzeug zur Unterstützung einer Analyse.

Die Eigenschaft der Gruppenbildung von Elementen innerhalb der Netzwerke ist von besonders hohem Interesse. Der Fokus lag hier zunächst primär auf Algorithmen und Methoden zur Detektierung und Darstellung von eindeutigen Zugehörigkeiten. Allerdings stellten die in den letzten Jahren entwickelten Algorithmen zur Detektierung unscharfer Gruppenzugehörigkeiten die reelle Anwendbarkeit der alleinigen Betrachtung von scharfen Zugehörigkeiten in Frage [NPNBo7][ZWZ07]. Viele Darstellungsformen sind nicht direkt auf die Betrachtung unscharfer Zugehörigkeiten übertragbar. Es gilt neue Ansätze sowie Erweiterungen alter Ansätze zu entwickeln, damit die Eigenschaften die Unschärfe der Objekte analysiert werden können.

In dieser Ausarbeitung werden mögliche Visualisierungsansätze für unscharfe Gruppenzugehörigkeiten, am Beispiel von sozialen Netzwerken, vorgestellt. Es werden zusätzlich Interaktionstechniken für den Umgang mit den resultierenden Darstellungen entwickelt und anhand einer Fallstudie mit einem parallel zu dieser Ausarbeitung entwickelten Programm bewertet. Ziel ist es, neue Erkenntnisse über die unscharfen Gruppenstrukturen aus Netzwerken zu gewinnen.



**Abbildung 1.1.:** Beispiel für eine unübersichtliche Darstellung eines Netzwerkes. Es können keine relevanten Informationen gelesen / erkannt werden.

### Gliederung

Der Aufbau dieser Arbeit ist wie folgt gegliedert:

**Kapitel 1 – Einleitung:** Erläutert die Motivation dieser Arbeit und bietet einen Überblick hinsichtlich der Aufgabenstellung und der Ziele dieser Arbeit.

**Kapitel 2 – Grundlagen:** Es werden die notwendigen theoretischen Grundlagen dieser Arbeit beschrieben. Es werden zunächst die Grundlagen der unscharfen Mengenlehre (2.1), anschließend werden wesentliche Begriffe aus dem Gebiet des Graphen (2.2), der Sozialen Netzwerke (2.2.4) geklärt. Zuletzt werden wichtige Erkenntnisse aus dem Bereich der Datenvisualisierung (2.3) angeführt.

**Kapitel 3 – Visualisierung:** Analysiert und präsentiert grundlegende Ansätze aus dem Bereich des Graphzeichnen für die Visualisierung von Netzwerken und die daraus entstehende Hürde der Darstellung von Relationen. Es umfasst Algorithmen zur Detektierung unscharfer Gruppen, die als Grundlage für das nachfolgende Kapitel dienen.

**Kapitel 4 – Visualisierung unscharfer Gruppenzugehörigkeiten:** Stellt neue Ansätze als auch Erweiterungen bereits existierender Ansätze, zur Visualisierung unscharfer (*fuzzy*) Gruppen vor. Es knüpft an die in Kapitel 3 gewonnen Kenntnisse an.

**Kapitel 5 – Evaluierung:** Es werden die in Kapitel 4 erarbeiteten Ansätze anhand einer Fallstudie, mittels einem parallel zu dieser Ausarbeitung entworfenem Programm genutzt, um neue Erkenntnisse aus wohlbekanntem Netzwerken zu gewinnen.

**Kapitel 6 – Zusammenfassung und Ausblick:** Fasst die Ergebnisse der Arbeit zusammen und stellt Anknüpfungspunkte vor.



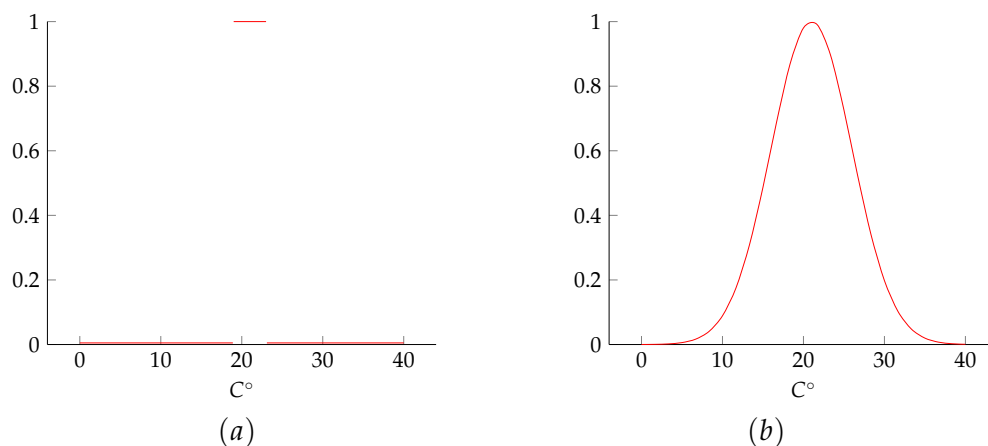
## 2. Grundlagen

In diesem Kapitel werden die erforderlichen Grundlagen gelegt, welche für eine Analyse und darauf folgende Visualisierung notwendig sind.

Es werden die für diese Arbeit relevanten theoretischen Konzepte und Werkzeuge aus der Graphentheorie, dem Gebiet des Graphzeichnens und der Informationsvisualisierung (InfoVis) sowie der Begriff eines Sozialen Netzwerkes zusammengefasst und geklärt.

### 2.1. Unscharfe Mengenlehre

Die von Lotfi Zadeh begründete „unscharfe Mengenlehre“ bzw. „Fuzzy-Logik“ gilt als dominierender Ansatz im Umgang mit Unsicherheiten [Zad65] [Zad78]. Zadeh *et al.* versuchen die Schwierigkeit im Umgang mit nicht eindeutig zuordenbaren Begriffen zu lösen, indem sie den Begriff der Zugehörigkeit zu einer Menge bzw. den zugehörigen Wahrheitswert „fuzzifiziert“, ihn als *unscharf* definiert. Die grundlegende Idee besteht darin, dass neben den Begriffen *wahr* und *falsch* Zwischenwerte eingefügt werden, die als Zugehörigkeitsgrad interpretiert werden können. Je größer dieser Grad ist, umso mehr sollte er unserer Vorstellung der tatsächlichen Zugehörigkeit entsprechen [Timo2].



**Abbildung 2.1.:** (a) Scharfe Definition einer „angenehmen“ Temperatur  
(b) unscharfe Definition

## 2. Grundlagen

---

Eine anschauliche Erläuterung dieses Problems ist die Definition einer angenehmen Raumtemperatur. Die Definition ist subjektiv, sie wird von jedem Menschen unterschiedlich ausgelegt. Eine exakte Definition zu finden erweist sich daher als schwierig. Eine Festlegung zwischen  $19^{\circ}\text{C}$  und  $23^{\circ}\text{C}$  wirft die Frage auf, warum  $19^{\circ}\text{C}$  angenehm sei,  $18,9^{\circ}\text{C}$  aber nicht. Ein stetiger Übergang von einer unangenehmen Temperatur hin zu einer angenehmen, liegt in diesem Beispiel als Definition näher. Fuzzy-Mengen bieten genau eine solche Möglichkeit. (siehe Abbildungen 2.1 (a) und (b))

Formal ist eine Fuzzy-Menge durch eine Funktion  $\mu$  über einer Menge  $U$  definiert. Es gilt:  $\mu : U \rightarrow [0,1]$ . Für ein Wert  $u_1 \in U$  mit  $\mu(u_1) = 0$  gilt, dass das Element  $u_1$  nicht dem definierten Fuzzy-Modell entspricht, für ein Wert  $u_2 \in U$  mit  $\mu(u_2) = 1$  gilt, dass das Element  $u_2$  voll dem definierten Fuzzy-Modell entspricht.

Zunächst vorrangig in Fachgebieten wie z. B. der Regelungstechnik, der Qualitätskontrolle oder in der Bildverarbeitung [CYP96] [Zim99] verbreitet, findet die Idee der unscharfen Zugehörigkeitsgrade in den letzten Jahren immer größere Anwendung in dem Bereich der Clusteranalyse von Netzwerken. Objekte können verschiedenen Clustern mit jeweils unterschiedlichen Zugehörigkeitsgrade zugeordnet werden. Aus diesen unscharfen Zugehörigkeitsgraden kann eine detailliertere Beschreibung der vorliegenden Strukturen abgeleitet werden.

## 2.2. Graphen

Der Begriff der *Graphentheorie* bezeichnet ein Teilgebiet der Mathematik und der Informatik, dass die Eigenschaften von *Graphen* und ihre Beziehungen untereinander untersucht.

In der Graphentheorie ist ein Graph als eine Menge von Objekten (Knoten) und einer Menge an Relationen (Kanten) definiert. Graphen werden häufig benutzt, um relationale Daten zu modellieren und darzustellen.

Graphen die Gegenstand der Graphentheorie sind, sind zu unterscheiden von Funktionsgraphen in der Mathematik.

### 2.2.1. Definitionen

#### **Definition 1. Einfacher Graph:**

Ein Graph  $G$  ist ein Tupel  $G = (V, E)$  wobei:

- $V$  eine endliche Menge von Objekten, oft auch bezeichnet als Ecken oder Knoten (engl. vertices), von  $G$ .
- $E$  eine endliche Menge  $E \subseteq [V]^2$  genannt Relation, auch bezeichnet als Kanten (engl. edges), von  $G$ .
- wobei keine Kante  $e \in E$  mit  $e = (v_i, v_i)$  existiert.
- $e_{i,j} = (v_i, v_j)$  höchstens einmal in  $E$  existiert.

**Informell:** Ein einfacher Graph  $G$  ist eine Menge  $V$  von Objekten, die durch Relationen  $e_i$  verbunden sind. Dabei kann eine Relation  $e_i$  nicht denselben Anfangs- und Endpunkt besitzen, sowie nur einmal in  $G$  definiert sein.

Die *Ordnung*  $|G|$  eines Graphen, ist durch die Anzahl  $|V|$  seiner Objekte definiert. Äquivalent dazu seine *Größe*  $\|G\|$  durch die Anzahl  $|E|$  der Relationen zwischen seinen Objekten. Der *Grad* eines Objektes ist durch die Anzahl der ab- bzw. eingehenden Relationen definiert. Die *Dichte*  $D$  eines Graphen ist durch die Anzahl der Relationen relativ zu der maximal möglichen Anzahl beschrieben:  $D = \frac{2|E|}{|V|(|V|-1)}$ . Gilt  $O(|V|) < |E| \ll O(|V|^2)$  so nennt man ihn *dünn*, ist  $D$  nahe bei 1 *dicht*.

Es existieren verschiedene Varianten und Erweiterungen des Einfachen Graphen. Die genaue Definition wird von den speziellen Eigenschaften der Relationen bestimmt.

Durch zusätzliche Definition einer *Funktion*  $\sigma(e)$ , die jeder Relation einen numerischen Wert zuweist, können *gewichtete Graphen* beschrieben werden. Ist das mehrfache Vorkommen einer Relation bzw. eine Relation mit demselben Objekt als Anfangs- und Endpunkt erlaubt, spricht man von einem *Multigraphen*. Ist eine Relation als Verbindung von mehr als zwei Objekten definiert so handelt es sich um einen *Hypergraphen*.

Stehen zwei Objekte des Graphen zueinander in Relation bezeichnet man sie als *adjacent*. Eine Relation zwischen zwei Objekten kann sowohl *gerichtet* als auch *ungerichtet* sein. Sind alle Relationen eines Graphen gerichtet so bezeichnet man ihn als einen *gerichteten Graphen*. Sind entsprechend alle Relationen ungerichtet so bezeichnet man ihn als einen *ungerichteten Graphen*. Beinhaltet ein Graph sowohl gerichtete als auch ungerichtete Relationen so nennt man ihn *gemischt* [HMM00].

Ein *Sub-Graph*  $G_S$  besteht aus einer Teilmenge  $V_S \subseteq V$  sowie aus allen zugehörigen Relationen  $E_S \subseteq E$ . Ein Graph ist *vollständig*, wenn alle Objekte adjacent zueinander sind. Graphen werden in dieser Ausarbeitung oft mit dem Begriff *Netzwerk* bezeichnet. Entgegen der oft verwendeten Definition eines Netzwerkes als gerichteter Graph [Tam12], werden in den folgenden Kapiteln auch ungerichtete Graphen als Solche bezeichnet.

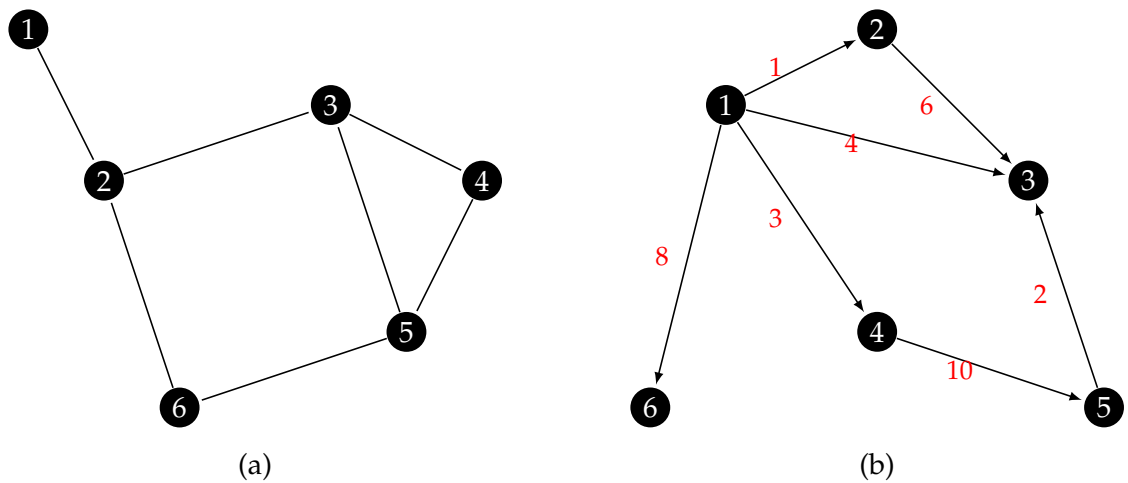
Zur Visualisierung von Graphen werden häufig *Node-Link-Diagramme* genutzt. Sie stellen Objekte als Knoten in einer beliebigen geometrischen Form und die ihnen zugehörigen Relationen verbindende Linie dar.

Die Art der Darstellung dieser Punkte und Linien ist eine Frage der Zweckmäßigkeit. Die formale Definition eines Graphen ist von seiner bildlichen Darstellung unabhängig [Die10]. Abbildung 2.2 (a) stellt einen einfachen Graphen dar, (b) einen gerichteten Graphen mit Kantengewichten.

### 2.2.2. Datenstrukturen

Es existieren unterschiedliche Ansätze Graphen mathematisch zu beschreiben. Betrachten wir ein Netzwerk mit  $n$  Knoten und mit Beschriftungen  $1 \dots n$ , wie z. B. in Abbildung 2.2 gegeben.

Eine Kante  $e_k$  zwischen den Knoten  $v_i$  und  $v_j$  kann als Tupel  $(i, j)$  beschrieben werden. Der



**Abbildung 2.2.: Zwei Graphen.** (a) Ein einfacher Graph, d.h. er besitzt keine mehrfach Kanten und kein Knoten hat eine Kante auf sich selbst. (b) Ein gerichteter Graph mit Kantengewichten (rot)

gesamte Graph lässt sich somit durch eine Liste von  $n$  Tupeln beschreiben. Beispielsweise wäre das Netzwerk aus Abbildung 2.2 (a) durch die 7 Tupel  $(1,2)$ ,  $(2,3)$ ,  $(2,6)$ ,  $(3,4)$ ,  $(3,5)$ ,  $(4,5)$  und  $(6,5)$  vollständig spezifiziert. Diese Form der Beschreibung wird auch *Kantenliste* genannt. Sie wird meist in der Speicherung der Netzwerke auf Computern angewendet, ist jedoch für mathematische Zwecke, wie die Ausführung von Algorithmen, ungeeignet. Eine für diese Zwecke besser geeignete Datenstruktur ist die *Adjazenzmatrix*. Sie repräsentiert einen Graphen mit  $n$  Knoten durch eine  $n \times n$  Matrix. Die Einträge  $A_{ij}$  bilden sich dabei wie folgt:

$$(2.1) \quad A_{ij} = \begin{cases} 1 & \text{es existiert eine Kante zwischen } v_i \text{ und } v_j \\ 0 & \text{sonst} \end{cases}$$

Alternativ für gewichtete Graphen kann der Eintrag  $A_{ij}$  auch zur Speicherung eines *Kantengewichtes*  $g$  genutzt werden. In diesem Fall gilt:

$$(2.2) \quad A_{ij} = \begin{cases} g_{ij} & \text{Es existiert eine Kante zwischen } v_i \text{ und } v_j \\ 0 & \text{sonst} \end{cases}$$

Die Graphen aus Abbildung 2.2 lassen sich demnach folgendermaßen beschreiben:

$$(2.3) \quad A_{(a)} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix} \quad A_{(b)} = \begin{pmatrix} 0 & 1 & 4 & 3 & 0 & 8 \\ 0 & 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Eine spezielle Eigenschaft der Adjazenzmatrix eines ungerichteten Graphen ist die *Symmetrie* zur Hauptdiagonalen. (Erkennbar in Gleichung 2.3)

### 2.2.3. Gruppen

Der Begriff der Gruppenbildung aus einer Menge von Elementen ist in vielen unterschiedlichen Bereichen vorzufinden. Im Folgenden wird jedoch ausschließlich auf die Gruppenbildung innerhalb von Graphen Bezug genommen.

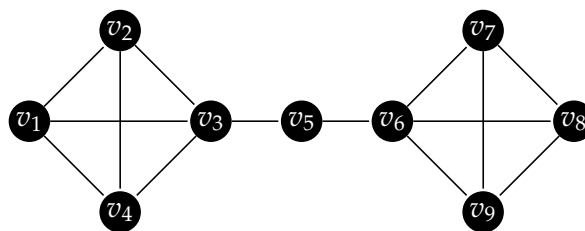
Eine *Gruppe*, auch *Cluster* genannt (engl. *community*) ist ein spezieller Sub-Graph, dessen Objekte stark untereinander, jedoch nur schwach mit Knoten des restlichen Netzwerkes verbunden sind. Eine eindeutige Definition ist nicht bekannt; der Begriff einer Gruppe wird je nach Anwendung unterschiedlich ausgelegt. Verschiedene Metriken versuchen unterschiedliche Aspekte bei der Betrachtung der Gruppen abzudecken. Die jeweilige Wahl ist stark von den bezweckten Zielen der Analyse abhängig.

Gruppen im Allgemeinen, können auch als unscharfe Mengen, wie sie in Abschnitt 2.1 definiert sind, betrachtet werden. Dies trifft vor allem auf Netzwerkanalysen zu. In vielen Netzwerken ist es oft der Fall, dass ein Objekt mehr als einer Gruppe zugeordnet ist. In diesem Fall spricht man von *unscharfer Zugehörigkeit* des Objektes zu einer Gruppe.

Dieser Begriff lässt sich leicht an folgendem Beispiel klären:

Abbildung 2.3 zeigt einen Graphen mit 9 Knoten und 14 Kanten. Es ist leicht zu erkennen, dass sich 2 Gruppen aus dem Graphen hervorheben. Beschreibt man diesen Graphen mittels einer eindeutigen Gruppen Definition, besteht dieser aus der Gruppe  $G_1$  mit den Knoten  $v_1, v_2, v_3$  und  $v_4$ , und der Gruppe  $G_2$  mit den Knoten  $v_6, v_7, v_8$  und  $v_9$ . Knoten  $v_5$  wäre somit keiner der beiden Gruppen zuordenbar, bzw. je nach Algorithmus nur einer der beiden Gruppen.

Nutzt man jedoch die unscharfe Gruppen Definition, so kann  $v_5$  beiden Gruppen zu gleichen Teilen zugeordnet werden. Es können sogar (wie es in diesem Beispiel der Fall ist) spezielle *Brückenknoten* erkannt werden. Von Brückenknoten spricht man sobald ein Knoten eine Verbindung zwischen zwei verschiedenen Gruppen bildet.



**Abbildung 2.3.:** Beispiel für ein Netzwerk mit 9 Knoten. Es ist einteilbar in 2 Gruppen, wobei Knoten 5 nicht eindeutig zuzuordnen ist.

### 2.2.4. Soziale Netzwerke

Die in dieser Arbeit näher betrachteten Datensätze, sind aus dem Bereich der *Soziale Netzwerke*. Darunter versteht man im Allgemeinen, ein Netzwerk von Akteuren, die untereinander mit Beziehungen, auch *Relationen* genannt, verbunden sind.

Sie können sowohl einer hierarchischen Struktur als auch einer nicht-hierarchischen Struktur entsprechen.

### 2.3. Visualisierung

*„Die Visualisierung von Daten ist die bildliche Veranschaulichung ihrer relevanten Aspekte und hat schon zu allen Zeiten eine wichtige Rolle gespielt, um sowohl die Erkenntnis als auch die Kommunikation zu erleichtern. [...] Bildhafte Darstellungen sind ein weit verbreitetes Medium zur Präsentation von Fakten und Informationen [...] um eine effizientere Analyse und Kommunikation zu erreichen.“ [SMoo]*

Ziel einer Visualisierung ist es anhand graphischer Darstellung neue Erkenntnisse über die vorliegenden Daten zu gewinnen. Die Identifikation von Mustern gibt Auskunft über die Einbindung von Akteuren und deren Beziehungen untereinander und basiert auf drei Kernfragen [Kre05]:

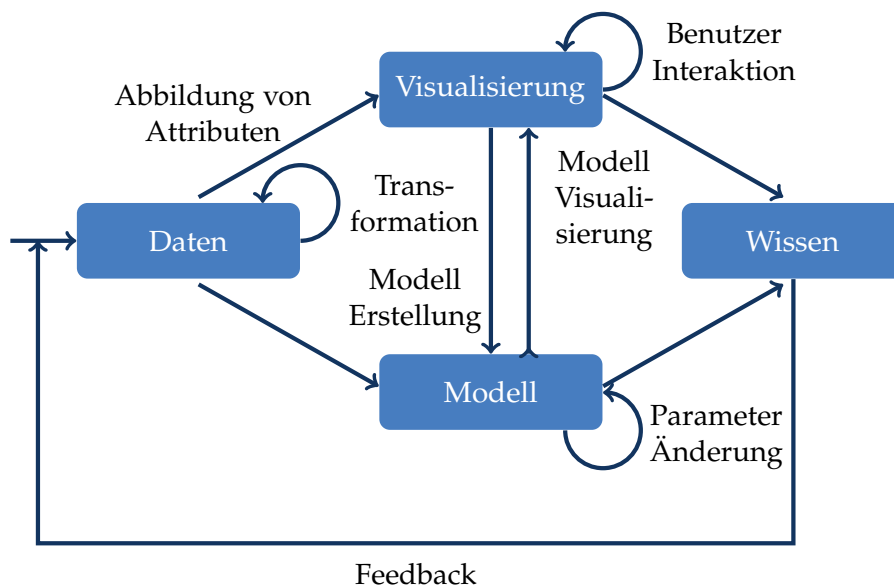
- (1) Wie und mit welchen Informationen müssen die Knoten positioniert werden, um aus dem Netzwerk eine Darstellung zu bilden mit der die wesentlichen Eigenschaften abgelesen werden.
- (2) Welche strukturellen Eigenschaften des Graphen können in die Visualisierung integriert werden, damit Besonderheiten klar zu identifizieren sind?
- (3) Auf welche Art und Weise ist es möglich externe Informationen visuell zu kodieren, um lokale Konzentrationen von Merkmalen sichtbar zu machen?

Krempel [Kre05] unterteilt das Vorgehen bei der Erstellung einer Visualisierung in drei separate Schritte:

- (1) Das **Layout** bildet den grundlegenden Schritt für eine Visualisierung und ist entscheidend für eine überschaubare Zuordnung von Zusammenhängen. Metriken sollen so gewählt werden, dass sie ein Maximum von Eigenschaften des Graphen miteinbeziehen [Kre05]. *„Die entscheidende Aufgabe [...] ist [...] der Abgleich mit der räumlichen Anordnung innerhalb des Diagramms mit den strukturellen Eigenschaften des Graphen. Dieser Schritt wird auch das Layout der Visualisierung genannt.“ [Bra99]* Ein geeignetes Layout zeichnet sich folglich dadurch aus, dass sowohl alle Eigenschaften des Graphen klar identifizierbar als auch spezielle Informationen, wie Rollen, eindeutig zu unterscheiden sind.

- (2) Mit Hilfe der **Abbildung von Attributen** werden zusätzliche Informationen in die Darstellung integriert. Es wird ermittelt, wie visuelle Elemente wie Größen, Farben und Formen verwendet werden können, um Merkmale der Strukturen hervorzuheben.
- (3) Eine **Interaktion** erlaubt es zusätzliche Informationen in das Netzwerk mit einzubringen und es mit diesen weiter zu untersuchen.

Keim beschreibt den Ablauf dieser Schritte als kontinuierlichen Prozess [KAF<sup>+</sup>08] (vgl. Abbildung 2.4). Die vorliegenden Daten können durch Transformationen, wie Filterung, Aggregation oder Untersuchung auf spezielle Eigenschaften wie unscharfe Gruppenzugehörigkeiten, vorverarbeitet werden. Anschließend können die transformierten Daten für die visuelle Repräsentation des Graphen genutzt werden. Einerseits für den Layout-Prozess, andererseits als zusätzliche Abbildung von Attributen. Aus den daraus entstehenden visuellen Modellen der Daten können, durch eine Interaktion mit dem Benutzer, neue Erkenntnisse aus den Daten gewonnen werden. Mit diesem Wissen können Anpassungen der Transformationen unternommen werden. Dieser Ablauf ist auf jede Visualisierung übertragbar, in dieser Arbeit wird er in den in Kapitel 4 vorgestellten Ansätzen verwendet.



**Abbildung 2.4.:** Visueller Analyse Prozess beschrieben bei Keim *et al.* [KAF<sup>+</sup>08]. Sie beschreiben den Vorgang durch vier Schritte: (1) Pre-processing der Daten, (2) Layout erstellen / Abbildung von Attributen, (3) Visuelle Interaktion mit dem Benutzer und (4) Modellbasierte Analyse.





## 3. Visualisierung

Dieses Kapitel behandelt Vorarbeiten sowie Grundlagen aus dem Bereich der Graphvisualisierung, speziell das Teilfeld der Visualisierung und Analyse von Gruppenstrukturen in Netzwerken.

Das Kapitel ist dabei in folgende drei Abschnitte gegliedert:

- (1) Der erste Teil behandelt grundlegende Ansätze für die Visualisierung von Graphen.
- (2) Verwandte Arbeiten, die einige vorgestellte Ideen erweitern, bzw. einen neuen Ansatz für die Darstellung präsentieren.
- (3) Als Vorbereitung für das nachfolgende Kapitel, werden im dritten Abschnitt zwei ausgewählte Algorithmen für die Detektion unscharfer Cluster vorgestellt und ihre Anwendbarkeit in Bezug auf Soziale Netzwerke auf Plausibilität untersucht.

### 3.1. Ansätze der Visualisierung

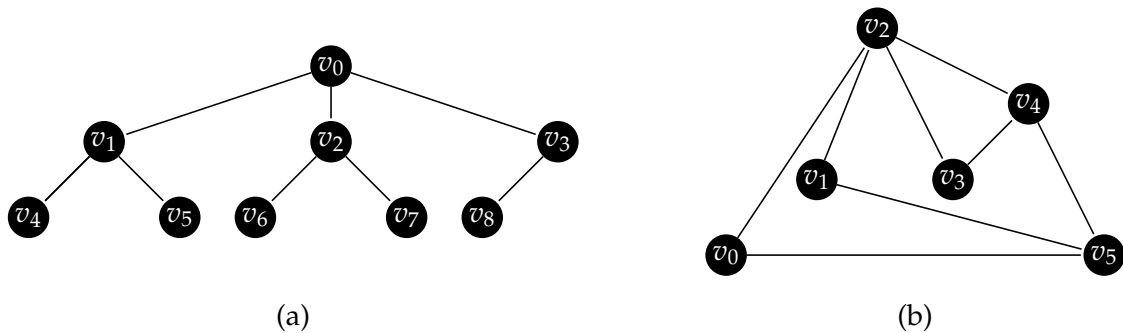
#### 3.1.1. Motivation

Eine universelle Visualisierung, mit der jeder Datensatz ohne Anpassungen darstellbar ist, existiert aus vielerlei Gründen nicht:

Zum einen existiert eine Vielzahl an verschiedenen Arten von Datensätzen, die sich in der vorausgehenden Zielsetzung unterscheiden. Diese bestimmt *was* und welche Eigenschaften mit den Daten erfasst werden sollen. Werden ausschließlich Datensätze die sich zur Clusteranalyse eignen betrachtet, bilden sich mehrere Unterkategorien. Sie werden benutzt, um heterogene Gruppen von Objekten bzw. eine homogene Menge an Elementen zu identifizieren. Das trifft sowohl auf Datensätze für eine Kaufverhaltensanalyse zu, als auch auf relationale Datenätze, wie es z. B. Netzwerke sind. Verschiedene Datensätze besitzen verschiedene Aspekte, welche es zu analysieren gilt.

Zum anderen betrachtet jede Visualisierung die Daten unter verschiedenen Gesichtspunkten. Es gibt verschiedene Ansätze, welche sich für die jeweiligen Daten und Analyseaspekte eignen. Eine Analyse eines Datensatzes benötigt unterschiedliche Hervorhebungsmerkmale, die meist nicht von einer einzigen Visualisierung abgedeckt werden können.

Die Art der verwendeten Visualisierung hängt demnach sehr von dem zu Grunde liegenden Datensatz ab.



**Abbildung 3.1.:** (a) Beispiel für ein hierarchisches Netzwerk mit Knoten  $v_0$  als Wurzel. (b) Beispiel für ein nicht-hierarchisches Netzwerk

In dieser Arbeit wird die Visualisierung von Netzwerken näher betrachtet. Diese Unterkategorie der relationalen Datensätze kann im Groben in *hierarchisch* und *nicht-hierarchisch* unterteilt werden. Hierarchische Netzwerke werden mit einem *Eltern-Kind* Modell beschrieben, diese Beziehungen werden auch *Inklusion* genannt. Nicht-hierarchische Netzwerke lassen sich nicht durch ein striktes Eltern-Kind Modell beschreiben. Ein Objekt kann mit jedem anderen Objekt des Netzwerkes in Beziehung stehen. Beziehungen in dieser Art werden *Adjazent* genannt. (Für ein Beispiel siehe Abbildung 3.1)

In der Netzwerkanalyse bildete sich eine dritte Unterkategorie heraus, welche die Eigenschaften der hierarchischen und nicht-hierarchischen Netzwerke vereinigt. Beispiele für solche Netzwerke, die auch im Hinblick auf die Clusteranalyse von hohem Interesse sind [Holo6]:

- Hierarchisch organisiertes Referenznetzwerk, indem die Publikationen die unterste Hierarchie Ebene darstellen, Institute oder Organisationen die die höheren Ebenen repräsentieren. Adjazente Verbindungen zwischen den Publikationen stellen Referenzen bzw. Zitate zwischen ihnen dar.
- Hierarchisch aufgebautes Software-System, z. B. Code, unterteilt in Klassen, Pakete und Daten. Beziehungen zwischen diesen Elementen, wie z. B. ein Aufruf einer bestimmten Funktion einer Klasse, bilden wiederum adjazente Verbindungen zwischen den einzelnen Elementen.
- Soziale Netzwerke, können durch ein solches Modell beschrieben werden. Personen bzw. Individuen bilden, ähnlich wie bei einem Referenznetzwerk, die unterste Ebene, Gruppierungen von ihnen die höheren Ebenen. Adjazente Verbindungen zwischen ihnen zeigen auf, in welcher Beziehung die Gruppen bzw. Personen stehen.

Es existiert eine Vielzahl an Darstellungstechniken für Bäume als auch Netzwerke. Sie werden in *Space filling*, *Node-Link basiert*, *Matrizen basiert* sowie *Hybride* klassifiziert [LKS<sup>+</sup>11]. Allgemein sind Node-Link basierte Ansätze ein sehr beliebtes Mittel um Netzwerke darzustellen [HRF10]. Sie sind einer breiten Gruppe bekannt. Der Umgang bzw. die Bedeutung verwendeter Objekte ist intuitiv versteh- bzw. erlernbar. Zeit kostende Einführungen für die Frage, *wie* die Visualisierung zu lesen sei, sind meist nicht nötig. Aus dieser Gegebenheit

können Muster in den Daten auch von Personen, welche nicht mit den Daten vertraut sind, erkannt und interpretiert werden [HRF10]. Andere Darstellungsformen haben zum Teil eine größere Mächtigkeit, benötigen jedoch für die Interpretation eine vorausgehende Einführung und sind somit nur bedingt einsetzbar. Auch im Hinblick auf die Betrachtung von Cluster/Gruppenzugehörigkeiten unterscheiden sich die verschiedenen Visualisierungs Ansätze stark in ihrer Einsetzbarkeit. Dies trifft besonders auf die Betrachtung von unscharfen Gruppenzugehörigkeiten zu. Bedingt durch die nicht eindeutige Zuordnung eines Objektes, ist selbst die Verwendung der Visualisierungsansätze, die sich für eindeutige Cluster gut eignen, nicht immer möglich.

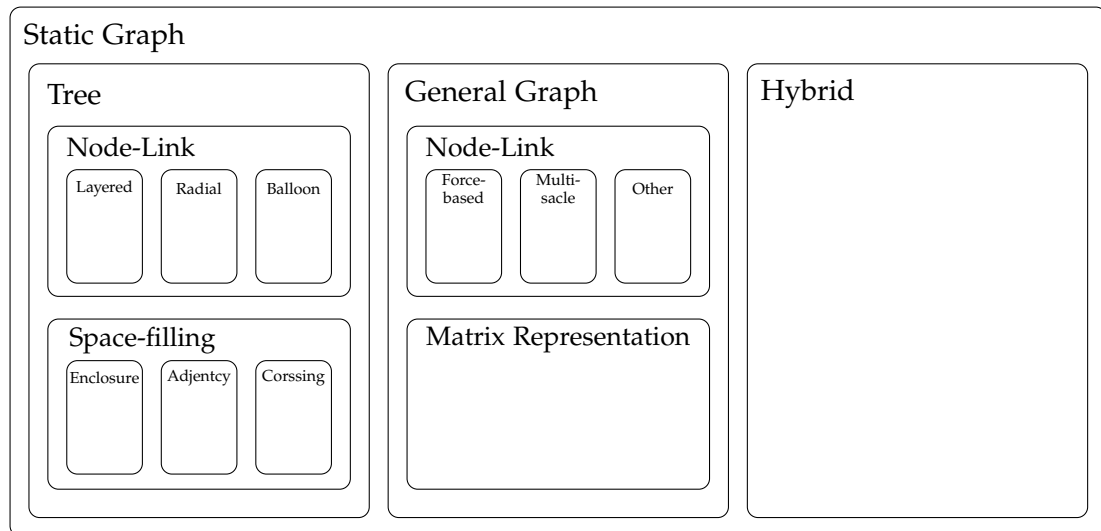
Der Begriff der Graphvisualisierung umfasst zwei Teilgebiete. Das Gebiet des *Graphzeichnen* und das Gebiet der *Visualisierung*. Beide verfolgen unterschiedliche Ziele: das Graphzeichnen das Finden einer geeigneten Position der Objekte, die Visualisierung das Finden einer geeigneten Abbildung verschiedener Eigenschaften der Objekte auf die Zeichnung. Beide sind jedoch elementar für eine gute und vor allem auch nützliche Graphvisualisierung.

#### 3.1.2. Graphzeichnen

Algorithmen aus dem Gebiet des Graphzeichnen versuchen für eine gegebene Menge an Objekten und Relationen, Positionen für die Objekte auf der Bildebene sowie eine repräsentierende Kurve für jede Relation zu errechnen. Dieses Vorgehen wird als das *Layout* eines Graphen bezeichnet. Das Graphzeichnen versucht zu ermitteln, was eine *gute* Darstellungen eines Graphen ist. Für eine wissenschaftliche Betrachtung dieser subjektiv erscheinenden Wertung, müssen Regeln, bezeichnet als *Ästhetiken*, definiert werden. Wagner und Kaufmann [KW01], Battista [BETT99] sowie Bennett [BRSG07] geben eine umfassende Übersicht der meist referenzierten *Ästhetiken* wieder. Ein Ausschnitt dieser Richtlinien sind:

- **Einheitliche Platzierung:** Knoten sollen einheitlich, mit gleichem Abstand zueinander dargestellt werden.
- **Kantenlänge:** Alle Kanten sollen in einer einheitlichen Länge gezeichnet werden, dabei soll sie aber so minimal wie möglich sein.
- **Überlappungen:** Es sollten Überlappungen bzw. Überschneidungen von Knoten als auch Kanten vermieden werden. Zwei sich überlappende Knoten werden in der Regel nicht als zwei unterschiedliche Objekte erkannt. Ähnlich auch die Überschneidung von Knoten und Kanten. Ein Knoten der eine Kante überschneidet, wirkt als sei er mit zwei separaten Kanten verbunden.
- **Platzminimierung:** Ein Graph scheint lesbarer zu wirken, wenn er gleichmäßig auf dem zu Verfügung stehenden Platz verteilt ist.

Diese Richtlinien sind untereinander zum Teil sehr widersprüchlich bzw. können nicht immer gleichzeitig erfüllt werden. Ein gutes Beispiel dafür ist die Platzminimierung: ein sehr komplexes Netzwerk ist bei Einhaltung dieser Richtlinie nicht lesbar (vgl. Abbildung 1.1). Purchase untersuchte in einer experimentellen Studie viele dieser gegebenen Richtlinien



**Abbildung 3.2.:** Überblick über Algorithmen für das Layouten eines Graphen. Abbildung erstellt nach Sammlung von [Tam12] und [HMM00]. Abbildung erfasst nur einen Teil aller existierenden Algorithmen.

[Pur97] und stellte einige von ihnen in Frage. Im Allgemeinen sind sie jedoch elementar für die Lesbarkeit einer Darstellung.

Verschiedene Graphzeichnen Algorithmen versuchen daher, unter Beachtung unterschiedlicher Ästhetiken eine gute Positionierung zu erreichen. Sie werden dabei durch die von ihnen angewandten Ideen bzw. durch das von ihnen erzeugte Layout kategorisiert. Generell unterscheidet man zwischen Algorithmen für Baum-Strukturen und Algorithmen für allgemeine Graphen [LKS<sup>+</sup>11]. Node-Link Layouts, die in beiden dieser Kategorien vorzufinden sind, bilden eine der meist genutzten Unterkategorien [HRF10]. Eine Übersicht über die verschiedenen Kategorien ist in Abbildung 3.2 gegeben.

Analysen verschiedener Graph Darstellungen mittels Eye-Tracking Techniken [Hua07] [PSD09] haben gezeigt, dass besonders Kräftebasierte Ansätze in verschiedenen Anwendungsfällen von Vorteil sind. Bei der Betrachtung von Gruppenstrukturen eignen sich jedoch auch Baum- bzw. Hierarchische Layouts für die visuelle Analyse. Bedingt aus den sich ergebenden hierarchischen Inklusionen zwischen den Gruppen und den ihnen zugehörigen Objekten bzw. den Relationen zwischen den einzelnen Gruppen, kann eine Vielzahl an Datensätzen auch solche mit dieser Kategorie dargestellt werden. Für Hierarchische Layout Verfahren entwickelte sich das *Sugiyama Framework* [STT81] als dominierendes Vorgehen [Tam12]. Das Prinzip dahinter ist, die Knoten in Ebenen aufzuteilen und so unter anderem eine einheitliche Kantenlänge zu generieren. Das Verfahren erzeugt eine gleichmäßige Knotenverteilung, wodurch das visuelle Durcheinander bei sehr komplexen fast hierarchischen Strukturen stark reduziert wird.

Dazu ähnlich können auch Baum basierte Layout Algorithmen als Grundlage für eine

Visualisierung der Gruppenstrukturen genutzt werden. Die dem Sugiyama Framework ähnelnden level-basierten Ansätze eignen sich, wegen ihrer ineffizienten Platzverwendung nur bedingt bzw. nur für Netzwerke niedriger Ordnung. Generell ist  $\mathcal{O}(n^2)$  Platz erforderlich, bei ausbalancierten Bäumen hoher Ordnung mit geringer Dichte, ist der Baum jedoch um ein vielfaches breiter als er hoch ist. So genannte *Balloon-* oder auch *Radiale-*Layouts [HMM00], die die Kinds-knoten kreisförmig um die Eltern-knoten anordnen, bieten in dieser Hinsicht eine bessere Nutzung des zu Verfügung stehenden Platzes. Neben diesen Node-Link Ansätzen eignen sich auch platzfüllende Techniken, wie *Treemaps* für die Darstellung von Gruppenzugehörigkeiten.

Baum- bzw. Hierarchische-Layouts sind eine verbreitete Visualisierung von Netzwerken [Tam12]. Bedingt durch die vorliegenden hierarchischen Strukturen, ist es meist einfach, sie in diese Darstellungen umzuwandeln. Auch bei der Betrachtung von eindeutigen Clustern ist es möglich, das Netzwerk in Form dieser Darstellung als Grundlage für eine Visualisierung zu nutzen. Eine alleinige Betrachtung von Eltern-Kind-Beziehungen ist in den wenigsten Fällen sinnvoll. Jedoch entsteht bei Verwendung zusätzlicher adjazenter Beziehung zwischen den Kinds- bzw. Eltern-Knoten die Gefahr des Visuellen Cluttering. Daher ist eine Verwendung dieser Verfahren für eine Visualisierung stark adjazenter Netzwerke, ohne Anpassungen nur bedingt möglich. Ein ähnliches Problem mit diesen Techniken entsteht bei der Betrachtung von unscharfen Clustern. Durch die mehrfache Zuordnung eines Objektes zu verschiedenen Clustern, ist eine Verwendung einer Eltern-Kind Beziehung, die einzelne Objekte den Gruppen zuordnet, nicht mehr möglich. Diese Eigenschaft lässt sich zwar durch eine Filterung der Zugehörigkeiten zum Teil wieder herstellen, jedoch erscheint dieser Schritt wenig sinnvoll. Es würden wiederum für diese spezielle Analyse wichtige Informationen verloren gehen. Aus dieser Gegebenheit werden für die in dieser Arbeit untersuchten Visualisierungen für unscharfe Gruppenzugehörigkeiten Baum- bzw. Hierarchische-Layout Algorithmen nicht als Grundlage verwendet. Es wird vielmehr auf Layout Verfahren für allgemeine Graphen zurückgegriffen, im Speziellen auf die Unterkategorie der *Kräftebasierten-*Layoutverfahren. In diesen Verfahren wird das Netzwerk als physikalische Einheit modelliert, auf dessen Objekte Kräfte einwirken. Diese Kräfte werden durch die Relationen des Netzwerkes repräsentiert. Sie können mit diesem Hintergrund als elastisches Element angesehen werden. Position der Knoten und Kanten werden durch die Algorithmen so gewählt, dass die insgesamt wirkende Kräfte minimal sind. Die Optimierung wird meist durch einen iterativen Ansatz gelöst. Sie kann als physikalisches Problem angesehen werden, das mit Gesetzen aus der Physik oder Vereinfachungen dieser gelöst werden kann. Algorithmus 3.1 veranschaulicht den generellen iterativen Ablauf eines Kräftebasieren Algorithmus.

Der Algorithmus von Kamada und Kawai [KK89], auch als *KK-Layout* bekannt, interpretiert diese Kraft als Resultat einer Feder die dem Hookesches Gesetz gehorcht. Sie modellierten den Graphen als System aus Stahlringen, die mit Federn verbunden sind. Die Energie  $E$  die in dem gesamten Graphen vorhanden ist, beschreiben sie als:

$$(3.1) \quad E = \sum_{\leq i < j \leq |V|} k_{ij} (|p_i - p_j| - l_{ij})^2$$

Wobei  $p_i$  die Position des gezeichneten Knotens für das jeweils korrespondierende Objekt  $v_i$ ,  $k_{ij}$  eine Feder-Konstante für die wirkende Kraft zwischen  $p_i$  und  $p_j$  und  $l_{ij}$  die optimale

### 3. Visualisierung

---

---

**Algorithmus 3.1** Beispielhafte Implementierung eines Kräftebasierten Algorithmus

---

```
Initiiere alle Knoten  $v_i$  mit Startgeschwindigkeit  $v_i \leftarrow (0, 0)$ 
Initiiere alle Knoten  $v_i$  mit zufälliger Startposition
do
   $E_{kin} \leftarrow 0$ 
  for jeden Knoten  $v_i$  do
     $Kraft_{Netz}(v_i) \leftarrow (0, 0)$ 
    for jede verbundene Feder  $e$  do
       $Kraft_{Netz}(v_i) \leftarrow Kraft_{Netz} + Hook(v, e)$ 
    end for
     $v_i.g \leftarrow (v_i.g + schritt * Kraft_{Netz}(v_i)) * dämpfung$ 
     $v_i.pos \leftarrow v_i.pos + schritt * v_i.g$ 
     $E_{kin} \leftarrow E_{kin} + v_i.m * (v_i.g)^2$ 
  end for
while  $E_{kin} \leq Grenzwert$ 
```

---

Distanz zwischen den Objekten  $v_i$  und  $v_j$  darstellt. Sie minimieren die Energie  $E$  mittels der Lösung partieller Differentialgleichungen, mit welcher für jeden Knoten  $v_i$  eine neue Position  $p_i$  ermittelt wird.

Ein großer Vorteil Kräftebasierter Algorithmen ist, das Verwenden nur jener Informationen, die mit der Struktur des Netzwerkes gegeben sind. Durch die Idee der Anwendung von Naturgesetzen neigen sie dazu relativ ästhetische Darstellungen zu produzieren und sind dadurch ebenso ein leicht zu verstehendes Mittel um schnell zufriedenstellende Darstellungen zu erreichen. Bedingt durch die hohe Laufzeit, sind viele Kräftebasierte Algorithmen schlecht auf große Netzwerke anwendbar. Einer der Hauptgründe für diese Tatsache ist, dass das Physikalische Modell meist mehrere lokale Minima besitzt, die Algorithmen jedoch nur bedingt diese von einem Globalen Minima unterscheiden können [Tam12]. Für eine Anwendung auf die Untersuchung von unscharfen Clusterstrukturen sind sie nichtsdestotrotz ein sehr wichtiges Mittel. Sie bieten eine einfach zu implementierende Klasse an Algorithmen, mit denen durch einfache Anpassungen, wie z. B. einer neuen Distanznorm oder durch eine wohl überlegte Reduktion der Daten wertvolle Visualisierungen erreicht werden. Einige der in Kapitel 4 vorgestellten Ansätze verwenden Kräftebasierte Algorithmen als grundlegendes Darstellungsmittel.

Aus dem Bereich des Graphzeichnen bietet Battista eine umfassende Sammlung vieler Algorithmen [BETT94] [BETT99]. Tamassia evaluiert vieler diese Algorithmen in „Handbook of Graph Drawing and Visualization“ [Tam12]. Ebenso bietet Landesberger [LKS<sup>+</sup>11] eine *state-of-the-art* Untersuchung aus Bereich der visuellen Analyse von großen Graphen.

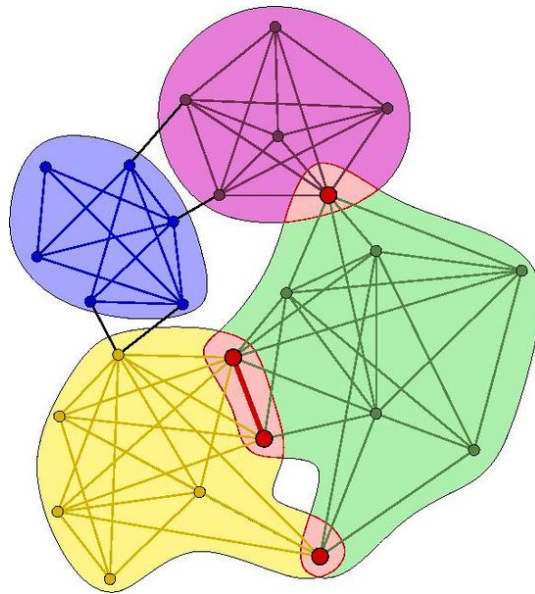
### 3.1.3. Visualisierung

Das Graphzeichnen erstellt die Grundlage, auf die eine „gute“ Visualisierung aufbaut. Diese versucht weitere Informationen in die Visualisierung abzubilden, damit tiefer gehende Erkenntnisse über das vorliegende Netzwerk getroffen werden können. Dieser Vorgang ist Kern der *Informations Visualisierung*, kurz *InfoVis*.

Die Betrachtung von Gruppen, in die sich einzelne Objekte einordnen lassen beschränkt sich nicht alleine auf die relationalen Datensätze, wie es Graphen sind. Eine weit verbreitete Methodik zur Visualisierung von Gruppenstrukturen ist die Verwendung von so genannten *Streudiagrammen* (engl. *scatter plot*). Hierbei wird ein Werte-Paar in einen kartesisches Koordinatensystem eingetragen, wodurch sich eine Punktwolke ergibt. Gruppenstrukturen können als Häufung von Werte-Paaren innerhalb eines bestimmten Bereiches interpretiert werden. Auch *Parallel coordinate plots* [HW10] werden oft für die Analyse von Cluster-Strukturen eingesetzt. Sie sind eine Methode zur Visualisierung von hochdimensionalen Strukturen und multivariater Daten.

In dieser Ausarbeitung werden Netzwerke des Typs der *Sozialen Netzwerke* (siehe Abschnitt 2.2.4) näher betrachtet. In Netzwerken allgemein, lassen sich oft Mengen von Objekten zu Gruppen (siehe Abschnitt 2.2.3) zusammenfügen. In den vergangenen Jahren wurden diese Gruppen oft mit einer eindeutigen Zugehörigkeit der Objekte beschrieben. Durch die eindeutige Zuordnung jedes Objektes zu einer Gruppe wird oft eine hierarchische Struktur als Metrik für Ordnung der Objekte verwendet (siehe Abschnitt 3.1.2). In diesem Layout werden die Gruppen meist durch ein Symbol bzw. einer farblichen Markierung hervorgehoben. Bei der Betrachtung von allgemeinen Node-Link-Diagrammen werden Gruppenzugehörigkeiten oft mit einer farblichen Hervorhebung der einzelnen *konvexen Hüllen* dieser Sub-Graphen visualisiert.(vgl. Abbildung 3.3)

Diese Techniken eignen sich gut für die Darstellung eindeutiger Cluster, jedoch nicht für eine Betrachtung von unscharfen Strukturen. Darstellung durch Bäume lassen keine Zuordnung unscharfer Knoten zu. Bei einer Darstellung unter Verwendung der Konvexen Hüllen (vgl. Abbildung 3.3) ist dies zwar generell möglich, jedoch stößt auch diese Darstellung bei der Annahme, dass Knoten beliebig vielen Clustern zugehörig sind, schnell an ihre Grenzen. Die Darstellung von doppelter Gruppenzugehörigkeit ist weitgehend übersichtlich, jedoch müssten bereits bei Dreifacher Zugehörigkeit Fragen hinsichtlich der Hervorhebung, neu geklärt werden. Kapitel 4 beschäftigt sich näher mit diesen Aspekten.



**Abbildung 3.3.:** Beispiel eines Graphen mit überlappenden Gruppenstrukturen. Probleme der sinnvollen Hervorhebung der sich überlappenden Teile unter Verwendung einer Färbung der konvexen Hüllen [For10].

#### 3.1.4. Navigation

Kein Layout und eine darauf aufbauende Visualisierung alleine, können die Probleme, die mit großen Netzwerken auftreten, kompensieren. Die Aufgabe des Aufdeckens verborgener Strukturen fordert innovative und logische Ansätze in der Benutzerinteraktion. Eine globale Darstellung großer Netzwerke ermöglicht meist, durch die endliche Größe des zur Verfügung stehenden Platzes, keine ausreichende Analyse. Es ist sinnvoller nur einen Ausschnitt des Netzwerkes, das die derzeitigen Interessen widerspiegelt, genauer zu betrachten.

##### „Zoom and Pan“

Ein traditionelles Werkzeug zur Navigation stellt das so genannte „Zoom and Pan“ dar. Pan stellt hier eine Verschiebung des Betrachtungspunktes dar. Bei der Zoom-Operation unterscheidet man in *geometrischem* und *semantischem* Zoom [HMM00]. Ein geometrischer Zoom vergrößert den ausgewählten Bereich, während ein semantischer Zoom eine Änderung des Informationsgrades ermöglicht (sehr gut vergleichbar mit dem Konzept von *Level of Detail (LOD)* Stufen aus der Computergraphik). In diesem Konzept liegt die Schwierigkeit nicht in der Ausführung der Zoom-Operation, sondern in dem Finden einer geeigneten Detailstufe. Ein entscheidender Nachteil dieser Technik ist der Verlust des globalen Kontextes. Bei einer Zoom-Operation wird zur näheren Betrachtung nur ein kleiner Ausschnitt des Netzwerkes dargestellt, dies ermöglicht zwar eine lesbarere Visualisierung, verhindert jedoch gleichzeitig eine Untersuchung im Bezug auf das gesamte Netzwerk.

Eine simple, aber sehr effektive Lösung dieses Problems stellt die Nutzung einer dauerhaften globalen Referenzansicht dar. Sie ermöglicht eine Analyse eines Ausschnittes, im Bezug auf



das gesamte Netzwerk. Ebenso kann sie als Navigationsmittel genutzt werden. Dieses Werkzeug ist gut für Darstellung von Graphen bzw. Netzwerken geeignet. Bedingt durch die Verwendung von vektorbasierten Formen, wie z. B. Kreise für die Darstellung der Knoten und einfache Linien für die Darstellung der Kanten, kann der Zoom durch eine einfache Bildtransformation ausgeführt werden. Im Gegensatz zur Verwendung von Rastergrafiken, sind Aliasing-Probleme mit diesem Ansatz möglich.

### Fokus+Kontext - Fisheye Verzerrung

Einer der größten Nachteile eines Zooms ist, dass bei der Betrachtung eines speziellen Ausschnittes meist die Informationen, die Beschreiben wie sich dieser in dem gesamten Kontext des Netzwerkes einordnet, verloren gehen. Dieser Informationsverlust stellt unter Umständen ein großes Hindernis in der Bedienbarkeit eines Analysetools dar.

Neben der Ergänzung mittels einer globalen Übersicht bieten so genannte „Fokus+Kontext“-Techniken (engl. „focus+context“) eine gute Alternative um dem Benutzer eine Fokussierung auf spezielle Details zu bieten, ohne dabei Information bezüglich der Einordnung in das gesamte Netzwerk zu verlieren.

Eine weit verbreitete Technik dieser Art stellt die *Fisheye Verzerrung* dar. Vergleichbar mit dem Prinzip eines Fisheye-Objektives einer Kamera, vergrößert sie den aktuellen Interessenbereich, während umliegende Teile in einem niedrigeren Detailgrad bzw. kleineren Größe dargestellt werden.

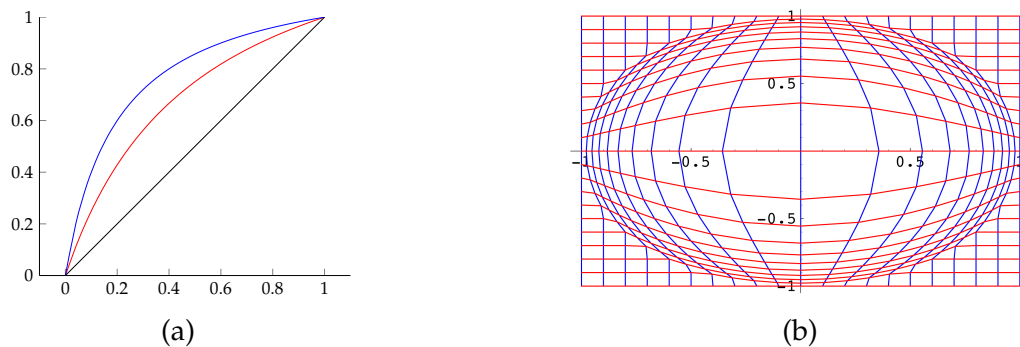
Mathematisch kann diese Eigenschaft als *Verkrümmungsfunktion*  $G(x)$  beschrieben werden. Dabei sollte  $G(x)$  eine konkave, monoton wachsende Funktion, mit  $G(x) : [0, 1] \rightarrow [0, 1]$  sein. Eine korrekte Fisheye Verzerrung sollte Knoten, die nahe 0, im Fokus des Fisheye, liegen, stärker beeinflussen, als Knoten die nahe bei 1 liegen. Sarkar und Brown [SB94] definierten eine einfache Verzerrungsfunktion mit der diese Eigenschaften erzielt werden:

$$(3.2) \quad G(x) = \frac{(d+1)x}{dx+1}$$

Die Konstante  $d$  ist hierbei der so genannte *Verzerrungsfaktor*. Sie kann als mögliche Schnittstelle zu einem Benutzer dienen, sodass eine interaktive Änderung des Effektes möglich ist. Sie sollte jedoch immer größer 0 sein. Ist sie gleich 0 so ergibt sich eine lineare Skalierung, ist sie kleiner 0 so kehrt sich der Effekt um. Je größer diese Konstante ist, desto stärker der Effekt.

Abbildung 3.4 (a) zeigt  $G(x)$  mit  $d = 0$ ,  $d = 1$  und  $d = 4$ . Abbildung 3.4 (b) zeigt eine beispielhafte Fisheye-Verkrümmung auf einem Gitternetz mit  $d = 4$ .

Es muss betont werden, dass Fokus+Kontext-Techniken als ergänzendes Mittel anzusehen sind. Sie ersetzen nicht die Funktionalität eines Zoom and Pens [HMMoo].



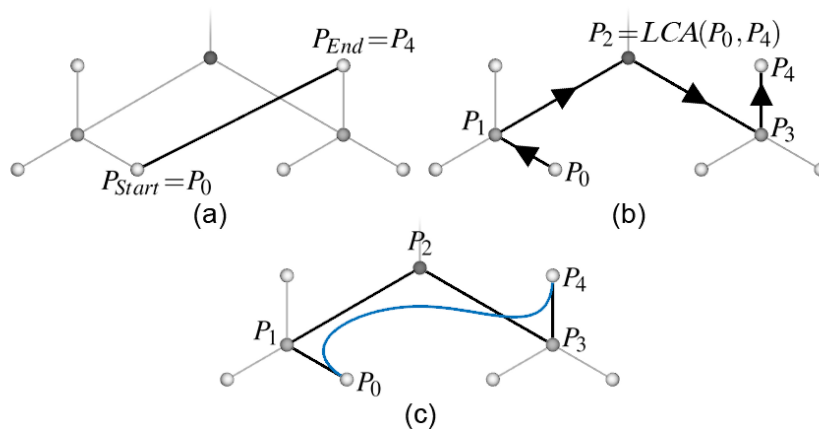
**Abbildung 3.4.:** (a) Verkrümmungsfunktion nach Sarkar-Brown mit Verkrümmungsfaktoren 0 (schwarz), 1 (rot) und 4 (blau). (b) Fisheye Verkrümmung auf einem Gitternetz mit Faktor 4.

## 3.2. Auswahl verwandter Arbeiten

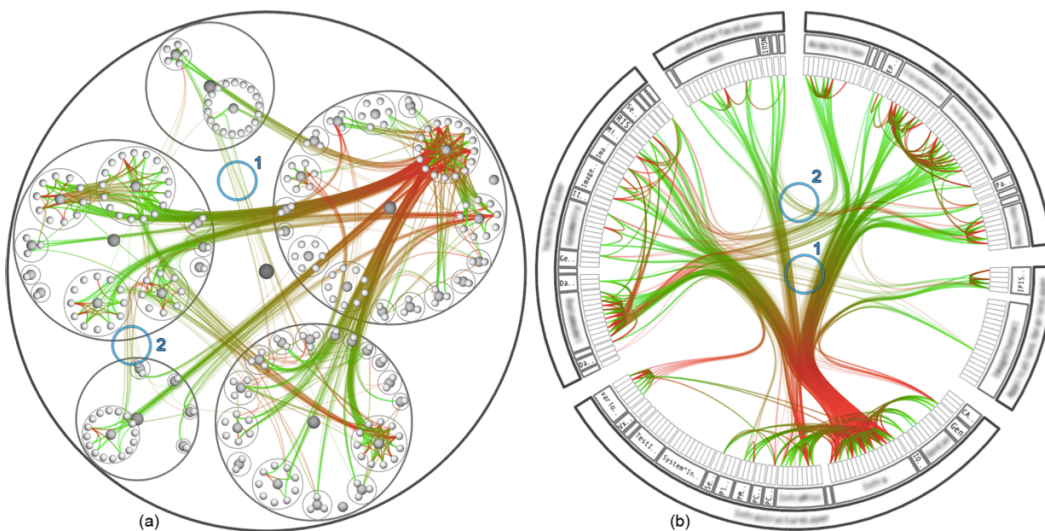
Holten [Holo6] versucht die aufkommenden Probleme des visuellen Clutters mittels einer Bündelung ähnlich laufender Kanten zu lösen (siehe Abbildung 3.6). Sein Ansatz bezieht sich auf die in Abschnitt 3.1.1 vorgestellten Datensätze, die die Eigenschaften hierarchische und nicht-hierarchische Netzwerke vereinen. Er nimmt dabei den Hierarchie-Pfad zwischen zwei Knoten mit einer adjazenten Relation als Kontrollpolygon einer Spline-Kurve (kubische B-Spline). Die Kontrollpunkte  $P_i$  des Kontrollpolygons werden aus den Punkten des Pfades entlang der Hierarchie-Kanten von  $P_{Start}$  und  $P_{Ende}$  gebildet. Dabei wird immer der Pfad mit den wenigsten Punkten gewählt. Das angewandte Prinzip wird in Abbildung 3.5 veranschaulicht.

Sein Ansatz reduziert stark das Durcheinander bei Netzwerken mit sehr vielen Kanten und Knoten. Durch eine Kontrolle der Stärke der Bündelung ist es möglich verschiedene Informationen aus dem Graphen zu extrahieren. Eine geringe Bündelungsstärke bietet detaillierte Übersicht sowie Informationen über die Knoten-Knoten Verbindungen. Eine starke Bündelungsstärke bietet hingegen eine grobe Übersicht des Netzwerkes, sowie implizit eine Visualisierung der adjazenten Kanten der Eltern-Knoten.

Holten erstellt keinen neuen Ansatz für eine visuelle Analyse von Netzwerken. Er stellt vielmehr Verbesserungen vor, mit denen die Probleme des visuellen Durcheinander reduziert werden können. Diese Ansätze lassen sich sehr gut bei der Analyse von Gruppenstrukturen verwenden. Eine starke Verbundenheit zwischen zwei Gruppen, kann so durch eine Bündelung der Kanten, somit einer Reduktion möglicher Überschneidungen von Kanten und Knoten, lesbarer visualisiert werden. Diese Anwendung ist nicht beschränkt auf die Betrachtung eindeutigen Gruppenstrukturen, sie ist auch ideal bei der Betrachtung von unscharfen Gruppenstrukturen. Jedoch ist eine Implementierung nicht immer möglich. Viele Frameworks betrachten Kanten als eigenständige Entität. Eine Bündelung, wie sie von Holten vorgeschlagen ist, ist nicht ohne massive Veränderungen des Framework möglich. Dieses Vorgehen wird daher, trotz der guten Tauglichkeit, nicht in den in Abschnitt 4 vorgestellten Ansätzen in Betracht gezogen.



**Abbildung 3.5.:** Bündelung von adjazenten Kanten unter Ausnutzung des Hierarchiepfades. (a) Einfache Gerade zwischen den zwei Punkten  $P_0$  und  $P_4$  (ohne Verwendung der Hierarchie Informationen). (b) Pfad entlang der Hierarchie von  $P_0$  zu  $P_4$ . (c) Spline-Kurve zur Darstellung der Verbindung zwischen  $P_0$  und  $P_4$  unter der Verwendung des Pfades aus (b) als Kontrollpolygon [Holo6].



**Abbildung 3.6.:** Software System mit damit verbundener Call-Hierarchie als Graph (grün: Caller rot: gecallter). (a) und (b) zeigen das System mit einer Bündelstärke von 0.85 mittels einem Balloonlayout (a) bzw. Radial-Layout (b). Die Bündelung reduziert merklich das Durcheinander, dass durch die Kanten produziert wird und hebt Relationen zwischen schwach verbundenen Systemen sichtbar hervor. Die umkreisten Flächen markieren identische Punkte in (a) und (b) [Holo6].

### 3. Visualisierung

---

Andere Ansätze kritisieren Node-Link Diagramme wegen der bei zunehmenden Größe des Graphen immer stärker werdenden Schwierigkeiten in der idealen Positionierung der einzelnen Knoten und Kanten. Sie versuchen durch alternative Darstellungsformen neue Möglichkeiten zu schaffen, größere Netzwerk ähnlich zu den Node-Link-Diagrammen intuitiv analysierbar zu machen, ohne dabei an Informationen zu verlieren.

Eine Alternative zu Node-Link-Diagrammen, welche besser skaliert, ist die nähere Betrachtung der Adjazenzmatrix (siehe Abschnitt 2.2.2). Anders als Node-Link-Diagramme, die in einem dichten bzw. hoch komplexen Netzwerk sehr viele Überschneidungen der Kanten produzieren, skalieren Adjazenzmatrizen sehr gut mit der Größe des Netzwerkes bzw. dessen Komplexität. Werden zusätzlich numerische Werte durch graphische Indikatoren ersetzt, erhält man eine intuitiv lesbare Darstellung [Ber83]. Es müssen dennoch mehrere Punkte für eine gute Darstellung betrachtet werden [HRF10]:

- (1) Nur durch eine sinnvolle Sortierung der Reihen und Spalten werden sonst verborgene Strukturen sichtbar.
- (2) Bedingt durch die quadratisch zur Anzahl der Knoten wachsende Größe der Matrix, bedarf es einer durchdachten sowie effektiven Navigationstechnik zur Analyse der Daten.

Bertin zeigte in seinem Fachbuch „Semiology of graphics“, dass eine Ersetzung der numerischen Werte durch graphische Indikatoren sowie einer anschließenden Sortierung der Reihen und Spalten, die Lesbarkeit verbessert werden kann, sowie Muster in den Daten erkennbar gemacht werden können [Ber83]. Ein Beispiel für eine Visualisierung der Adjazenzmatrix wird in Abbildung 3.7 dargestellt.

Es existiert eine Vielzahl von Algorithmen, die eine möglichst optimale Neuordnung einer Adjazenzmatrix produzieren. Sie unterscheiden sich je nach Datenkontext, stark in ihrer Komplexität, sowie in ihren Ergebnissen. Ein Vergleich bzw. eine Bewertung verschiedener Algorithmen erweist sich daher als sehr schwierig [HRF10].

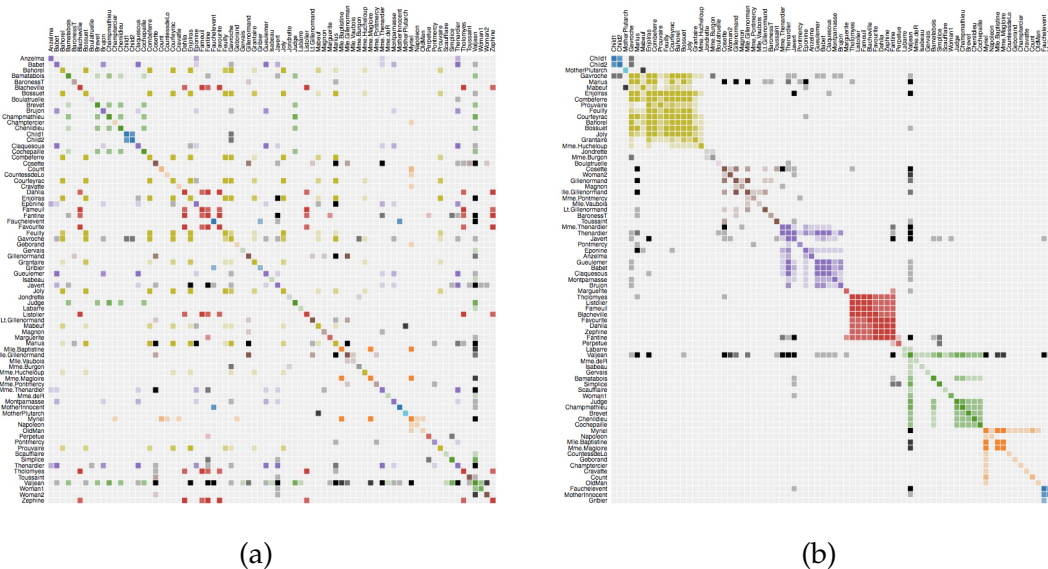
Muster können sich, ähnlich wie in Node-Link-Diagrammen, aus „gut geordneten“ Matrizen besonders hervorheben. Es ist jedoch wichtig zu verstehen, dass die Ordnungen der Matrix starken Einfluss auf die Lesbarkeit, sowie resultierende Interpretation der entstehenden Muster hat [HRF10].

Riche und Fekete befassten sich näher mit Adjazenzmatrix Darstellungen in ihrem Artikel „Novel Visualisation and Interactions for Social Network Implementation“ [HRF10].

Sie betonten, dass eine Matrixdarstellung ein mächtiges Werkzeug ist, um einen Überblick über ein Netzwerk zu erlangen. Sie ist, bedingt durch die niedrige Generierungszeit geeignet für eine schnelle Analyse. Sie heben außerdem hervor, dass es, bedingt durch die Anordnung in einer Matrix, weder zu Knoten- noch zu Kantenüberdeckungen kommen kann. Dies ist vor allem bei Verwendung von Knoten/Kanten Annotationen von Vorteil. Zuletzt betonten sie, dass immer eine globale Sicht auf das Netzwerk gegeben ist. Resultierend daraus, dass alle Knoten-Paare visualisiert werden, können sofort die Relationen zwischen einem Knoten-Paar erkannt werden.

Adjazenzmatrizen eignen sich gut für die Visualisierung von unscharfen Gruppen. Wählt man für die Färbung einzelner Kanten einen Farbgradienten anstatt einer festen Farbe, ist es möglich diese Gruppenstrukturen klar hervorzuheben.

### 3.3. Algorithmen zur Detektion unscharfer Gruppen



**Abbildung 3.7.:** Beispiel einer Darstellung per Adjazenzmatrix. Sie visualisiert gemeinsames Auftreten der Figuren aus Victor Hugos *Les Misérables*. Eingefärbte Zellen zeigen auf, dass zwei Figuren in einem gleichen Kapitel in Erscheinung traten. Dunkel gefärbte Zellen zeigen ein mehrfaches, gemeinsames Erscheinen auf. (a) ohne Sortierung der Reihenfolge. (b) mit Sortierung der Reihenfolge. Abbildung entnommen aus [Bos12], größere Ansicht in Anhang A.1

In dieser Arbeit werden jedoch Node-Link näher betrachtet. Sie haben, wie bereits in Kapitel 3 betont, den Vorteil eines leichten Verständnisses bei einer großen Anzahl an Benutzern.

### 3.3. Algorithmen zur Detektion unscharfer Gruppen

In den vergangenen Jahren wurden einige Algorithmen zur Detektion von Cluster-Strukturen in Netzwerken entwickelt. Fortunato [For09] bietet eine Übersicht und Vergleich verschiedener Algorithmen.

Dabei kristallisierten sich zwei Kernprobleme heraus:

- (1) Die Anzahl der enthaltenen Cluster ist, bis auf wenige Ausnahmen, meist unbekannt. Viele Algorithmen können daher kein allgemein gültiges Kriterium für die Größe der Strukturen bieten. Mehrere Durchläufe produzieren unter Umständen grundverschiedene Ergebnisse.
- (2) In vielen realen Netzwerken ist ein Knoten in mehr als einer Gruppe enthalten. Mit anderen Worten, die Cluster können sich überschneiden. Diese Eigenschaft wird von vielen Algorithmen nicht betrachtet. Jedoch spielen genau diese Knoten eine wichtige Rolle in der Netzwerkanalyse.

### 3. Visualisierung

---

In Hinblick auf das erste Kernproblem führten Newman und Girvan [NG04] eine modularitäts Funktion  $Q$  ein. Diese misst für ein *ungerichtetes* Netzwerk  $G(V, E, W)$  mit Knoten  $V$ , Kanten  $E$  sowie der Gewichtsfunktion  $W = [w_{i,j}]_{n \times n}$ , mit  $w_{i,j} \geq 0$  und  $n = |V|$  die Qualität einer gegebenen Cluster Einteilung. Newman und Girvan zeigten ebenso, dass durch eine Maximierung der Funktion  $Q$  über die Anzahl  $k$  der Cluster, eine optimale Struktur erreicht werden kann [NG04]. Formal ist  $Q$  durch

$$(3.3) \quad Q(P_k) = \sum_{i=0}^n \left[ \frac{L(V_c, V_c)}{L(V, V)} - \left( \frac{L(V_c, V)}{L(V, V)} \right)^2 \right]$$

definiert, wobei  $P_k$  eine Partition der Knoten in  $k$  Cluster und  $L(V', V'') = \sum_{i \in V', j \in V''} w(i, j)$  darstellt.

Lösungsansätze für das zweite Kernproblem wurden lange Zeit nicht erforscht. In den vergangenen Jahren gab es allerdings eine Bewegung weg von der Betrachtung eindeutiger Cluster, hin zu einer Betrachtung unscharfer Clusterzugehörigkeiten. Es wurden einige Algorithmen entwickelt, die es ermöglichen jedem Knoten mehrere Clusterzugehörigkeiten zuzuweisen [ZWZ07] [PDFV05][SGSH11].

Sie haben alle den gemeinsamen Ansatz, den Zugehörigkeitswert  $U_i$  eines Knotens  $i$  zu dem Netzwerk als Fuzzy-Menge über die existierenden Cluster zu beschreiben. 2.1 Dabei gilt für  $U_i$ :

$$(3.4) \quad U_i = \sum_{c \in 0 \dots k} U_i(c) = 1$$

Zwei dieser Algorithmen werden im Folgenden genauer betrachtet. Sie repräsentieren zwei verschiedene Ansätze um unscharfe Zugehörigkeiten zu berechnen.

**Zhang et al.** - *Detecting overlapping Communities with Fuzzy C-Means* [ZWZ07]:

Dieser Algorithmus vereinigt eine abgewandelte modularitäts Funktion  $\tilde{Q}$  mit Methoden aus dem spektralen Clustering.

$$(3.5) \quad \tilde{Q}(U_k) = \sum_{c=1}^k \left[ \frac{A(\tilde{V}_c, \tilde{V}_c)}{A(V, V)} - \left( \frac{A(\tilde{V}_c, V)}{A(V, V)} \right)^2 \right]$$

Sie formen mit den  $k - 1$  obersten Vektoren des Eigenvektorsystems  $Ax = tDx^1$  eine  $k - 1$  Dimensionale Abbildung in den Euklidischen Raum. In diesem Fall gilt für  $k$ :  $2 \leq k \leq K$ , wobei  $K$  eine frei gewählte obere Grenze der maximalen Anzahl an Cluster darstellt. Sie maximieren die abgewandelte Funktion  $\tilde{Q}$ , unter der zu Hilfenahme des Eigenvektorsystems  $E_k$ , indem sie die Reihenvektoren mittels eines fuzzy c-means Algorithmus nach Dunn [Dun73] und Bezdek [Bez81] in Cluster aufteilen, um eine Partition  $U_k$  der Knoten in  $k$  unscharfe Cluster zu erreichen.

<sup>1</sup>A ist dabei die Adjazenzmatrix mit  $A = (a_{ij})_{n \times n}$ , D eine Diagonalmatrix mit  $d_{ij} = \sum_k a_{ik}$

Der Ansatz von Zhang *et al.* ist nur bedingt auf größere Netzwerke anwendbar. Der Algorithmus liefert zwar eine möglichst korrekte Berechnung der unscharfen Gruppenzugehörigkeiten, jedoch ist seine Laufzeit durch die Verwendung von allgemeinen Eigenwerte sehr hoch. Daher wird dieser Algorithmus für die Erstellung der Werte nicht in Betracht gezogen.

**Nepusz *et al.*** - *Fuzzy communities and the concept of bridgeness in complex networks [NPNB07]*: In diesem Ansatz wird nach einer Ähnlichkeitsfunktion für die Knoten gesucht. Die Idee dahinter ist, dass Knoten des selben Clusters in gewisser Weise ähnlich zueinander sein müssen. Dabei nehmen sie an, dass eine Kante zwischen zwei Knoten  $v_i$  und  $v_j$  eine Ähnlichkeit dieser Knoten impliziert und umgekehrt keine Kante zwischen  $v_i$  und  $v_j$  eine Unähnlichkeit der Knoten.

Mit dieser Überlegung formulieren die *Ähnlichkeitsfunktion*  $s(U, i, j)$  oder kurz  $s_{ij}$  als Fuzzy-Menge (2.1) mit:

$$(3.6) \quad s_{ij} = \sum_{k=1}^c u_{ki} u_{kj}$$

Dabei definieren sie  $s_{ij} = 1$  bzw.  $s_{ij} = 0$  wenn die Zugehörigkeitswerte von  $v_i$  und  $v_j$  größt mögliche Ähnlichkeit bzw. Unähnlichkeit aufzeigen.

Sie reduzieren anschließend das Finden einer optimalen Partition der Knoten auf die Optimierung der Funktion:

$$(3.7) \quad D_G(U) = \sum_{i=1}^N \sum_{j=1}^N w_{ij} (\tilde{s}_{ij} - s_{ij})^2$$

Wobei  $c$  die Anzahl der Cluster,  $w_{ij}$  ein optionales Kantengewicht und  $\tilde{s}_{ij}$  ein vordefiniertes Ähnlichkeitskriterium repräsentiert. (Letzteres bildet sich in der Regel aus der Adjazenzmatrix  $A$  des Graphen  $G$ ).

Sie führten zusätzlich, gerade auch im Hinblick auf Soziale Netzwerke, den Begriff des Brückenfaktors (engl. bridgeness) ein. Dieser beschreibt Knoten, die einen signifikanten Anteil in einer speziellen Gruppe besitzen, aber auch als Brücke zu anderen Clustern fungieren. Diese Knoten sind speziell in Sozialen Netzwerken wichtige Knoten, da mit ihnen eine erklärbare Verbindung zwischen einigen Clustern hergestellt werden kann.

Im Gegensatz zu dem Algorithmus von Zhang *et al.* verwendet der Algorithmus von Nepusz *et al.* einfache Mathematische Operationen, seine Laufzeit ist wesentlich kürzer. Auch die verwendete Metrik liegt subjektiv gesehen näher an Menschlichen Vorstellungen eines Netzwerkes.





## 4. Visualisierung unscharfer Gruppenzugehörigkeiten

Die Aufgabe der Visualisierung der Unschärfe erfordert ein Umdenken in Umgang mit dem Attribut der Gruppenzugehörigkeit. In Abschnitt 3.1.2 sowie 3.1.3 wurden Ansätze für die Visualisierung von eindeutigen Gruppenzugehörigkeiten vorgestellt. Es wurde jedoch gleichzeitig begründet, dass sich diese Ansätze nicht bzw. nur schwer auf unscharfe Zugehörigkeiten übertragen lassen. Dieses Kapitel stellt daher neue Ansätze sowie Ideen vor, mit denen eine Visualisierung der unscharfen Gruppenzugehörigkeiten ermöglicht wird. Ziel dabei ist, dass die Gruppen erkennbar werden sowie die einzelnen Zugehörigkeiten jedes Knotens sinnvoll in die Ansicht integriert werden. Besonders unscharfe Knoten sollten zusätzlich schnell identifiziert werden können. Es soll auch mit Hilfe von Interaktionstechniken eine tiefere Analyse ermöglicht werden. Die Visualisierungsansätze sollen sowohl auf kleinen, als auch auf größeren Netzwerken angewandt werden können.

Für diese Zielsetzung wurden mehrere verschiedene Ansichten konzipiert:

- (1) **Allgemeine Übersicht.** Mit ihr soll ein Einstieg in die Analyse der Daten auf unscharfe Gruppenzugehörigkeiten möglich sein. Es werden daher die vorgestellten Kräftebasierte Algorithmen für die Positionierung der Knoten verwendet. Dabei sollten die Knoten jedoch grob in anhand ihrer stärksten Gruppenzugehörigkeiten eingeteilt werden.
- (2) **Kreis-Ansicht.** Eine Anordnung der Knoten in einem zirkulären Layout soll eine detailliertere Betrachtung der Gruppen und den Zusammenhängen zwischen ihnen ermöglichen.
- (3) **Aggregierte-Ansicht.** Mit dieser Ansicht soll eine detaillierte Analyse der Zusammenhänge zwischen den Gruppen ermöglicht werden. Zusätzlich sollen weitere Informationen, wie z. B. die durchschnittliche Gruppenstärke mit abgebildet werden.
- (4) **Einzel-Ansicht.** Soll eine nähere Betrachtung einer einzelnen Gruppe ermöglichen. Dabei sollen im Speziellen die Strukturen, die sich aus den verschiedenen starken Zugehörigkeitswerten bilden, ersichtlich werden.
- (5) **Kombinierte-Ansicht.** Es wird durch Kombination verschiedener Ansichten versucht, die Vorteile, die sich aus den verschiedenen Ansichten bilden, zu vereinen. Eine Kombination der Einzel-Ansicht mit der Allgemeinen-Ansicht, bietet eine Darstellung der Einbindung der einzelnen Gruppen in den Gesamtkontext des Netzwerkes. Daraus soll eine auf große Netzwerke gut skalierbare Darstellung gebildet werden. Sie soll sowohl detailreiche Zusammenhänge auf einer Gruppen-Ebene, als auch globale Zusammenhänge auf der gesamten Netzwerk-Ebene darstellen können.

Alle Ansichten sollen mit einer sinnvollen Abbildung weitere Attribute eine, „optimale“ Analyseplattform bilden. In Abschnitt 3.1.2 wurden verschiedene Darstellungsformen für eine Graphvisualisierung vorgestellt, die nachfolgenden Visualisierungsansätze verwenden Node-Link-Diagramme als Grundlage für die darauf aufbauende Abbildung weiterer Attribute. Das übergreifende Ziel ist es, neue Mittel zur Darstellung der unscharfen Gruppenzugehörigkeiten der Objekte zu erstellen.

### 4.1. Allgemeine Übersicht

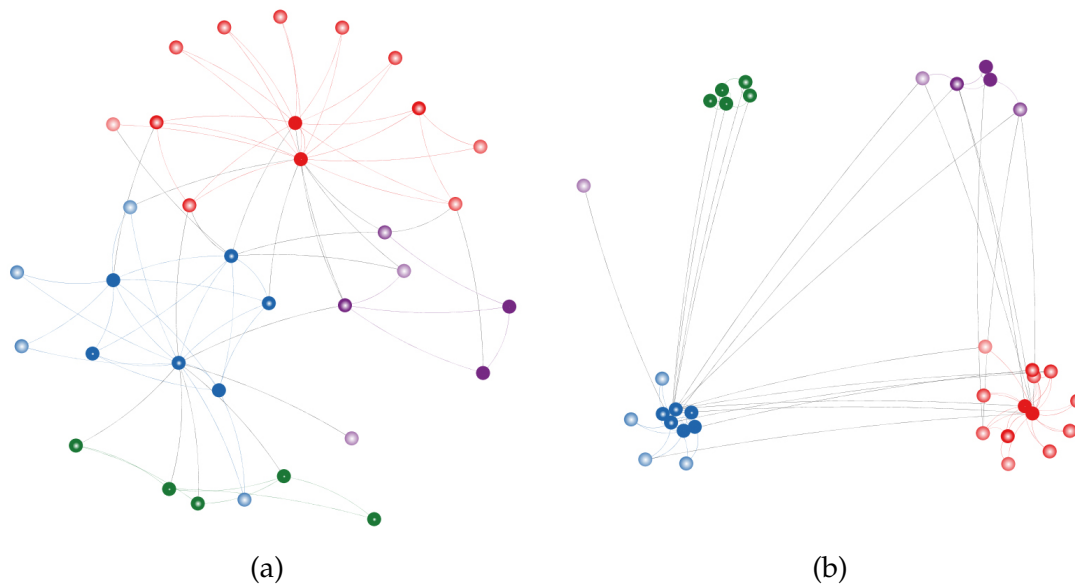
Die Allgemeine Übersicht dient einer ersten Betrachtung des Netzwerkes und als Einstiegspunkt für eine detailliertere Analyse. Für die Übersicht wird ein Kräftebasierter Layout-Algorithmus, wie z. B. der Kamada-Kawai- [KK89] oder der Fruchterman-Reingold- [FR91] oder alternativ der ISOM-Algorithmus [Mey98] vorgeschlagen. Die Positionierung der Knoten durch diese Algorithmen ist ein ausreichendes Mittel, um einen groben Überblick der Einteilung der Knoten in einzelne Gruppen zu erlangen. Zusätzlich besteht durch Einbringung der unscharfen Gruppenzugehörigkeiten die Möglichkeit der Erweiterung der *Distanznorm*. Diese wird in die Algorithmen als Beschreibung der Stärke der Verbindung zwischen zwei Knoten verwendet (siehe Kapitel 3.1.2). Die grundlegende Idee ist dabei, dass zwei Knoten mit derselben Gruppe als stärkste Zugehörigkeit, näher beieinander gezeichnet werden, als Knoten, die nicht dieselbe stärkste Gruppe teilen. Dieses Prinzip kann insoweit verfeinert werden, dass die genaue Differenz der beiden Zugehörigkeitswerte betrachtet wird. Je größer die Differenz zwischen zwei Knoten ist, umso weiter entfernt von einander sollten sie gezeichnet werden.

Die Distanz kann als minimalste Differenz aller gleichen Zugehörigkeiten gebildet werden. So können besonders unscharfe Knoten an den Rand ihrer dominierenden Gruppe angeordnet werden. Relationen mit Knoten einer anderen Gruppe können für die genaue Positionierung beachtet werden. Sind zwei Knoten zueinander adjazent, teilen aber keine gemeinsame Gruppe, so kann die Distanz nicht über die Differenz berechnet werden. Eine einfache Lösung wäre die Annahme, dass die beiden Knoten so weit wie möglich entfernt voneinander gezeichnet werden sollen. Die Distanz muss hierbei auf einen geeigneten großen Wert gesetzt werden. Formal kann die Distanz durch die Funktion  $Dist(e(v_s, v_e))$  über die Kante mit Startknoten  $v_s$  und Endknoten  $v_e$  definiert werden:

$$(4.1) \quad Dist(e(v_s, v_e)) = \begin{cases} \min_{k \in \{1, \dots, n\}} (|membership_k(v_s) - membership_k(v_e)|) & \exists k \\ \text{sehr groß} & \text{sonst} \end{cases}$$

Abbildung 4.1 zeigt die beispielhafte Anwendung der neuen Distanznorm auf Basis eines Kamada-Kawai-Algorithmus generierten Layout.

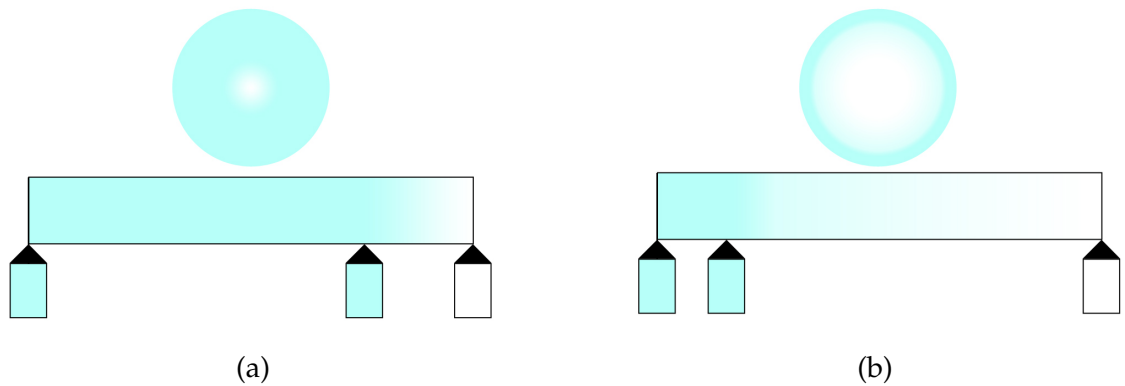
Ein weiterer Schritt hin zu einem „guten“ Einstiegspunkt für eine tiefergehende Analyse, ist die Abbildung von Eigenschaften bzw. Attributen auf die Knoten und Kanten. Darunter ist beispielsweise die Färbung eines Knotens oder die Biegung einer Kante unter Einbeziehung geeigneter Informationen zu verstehen.



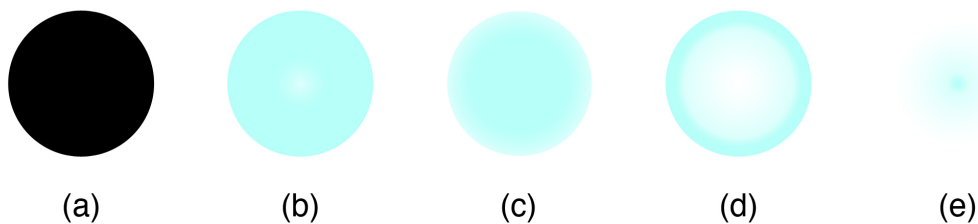
**Abbildung 4.1.:** Vergleich eines Graphen ohne (a) und mit (b) Anwendung der neuen Distanznorm

Die Analyse von Gruppenzugehörigkeiten legt nahe, den Wert der Gruppenzugehörigkeit eines Knotens auf dessen Form oder Farbe abzubilden. Knoten verschiedener Gruppen können somit zusätzlich neben ihrer Position visuell hervorgehoben werden. Eine Abbildung auf die Farbe ist dabei besser geeignet, da das Finden eines geeigneten Farbschemas, eine besser skalierbare Aufgabe darstellt, als das Finden unterschiedlicher, klar unterscheidbarer Formen. Darüber hinaus ist es möglich, das Finden eines passenden Farbschemas durch einen geeigneten Algorithmus zu automatisieren. In Bezug auf unscharfe Gruppenzugehörigkeiten ist es auch möglich, den *Grad der Unschärfe* auf die Färbung des Knotens abzubilden. Ein geeignetes Mittel dafür ist die Verwendung eines Farbverlaufes. Der prozentuale Wert der dominierenden Zugehörigkeit kann hierbei als Parameter für den Verlauf genutzt werden. Eine mögliche detaillierte Umsetzung dieses Ansatzes ist in Abbildung 4.2 beispielhaft erläutert. Es ist aus Gründen der Lesbarkeit zu empfehlen, den Farbgradient von starkgesättigter Farbe zu heller, weniger gesättigter Farbe zu zeichnen. Dies ermöglicht eine bessere Abgrenzung des Knotens vom Hintergrund. Abbildung 4.3 demonstriert dieses Verhalten. Sowohl (d) als auch (e) stellen den gleichen Zugehörigkeitswert dar, Abbildung (e) verfälscht jedoch durch den Eindruck einer kleineren Größe eine mögliche Analyse. Es ist potentiell möglich einen so gefärbten Knoten zu übersehen.

Auf ähnliche Weise können diese Informationen auf die Kanten abgebildet werden. Eine Kante bei der beide Endpunkte ihre stärkste Zugehörigkeit in derselben Gruppe besitzen, kann in der Farbe der Gruppe gezeichnet werden. Eine Kante, deren Endpunkte ihre stärkste Zugehörigkeit in zwei verschiedenen Gruppen besitzen, kann durch einen farblichen Verlauf der beiden Gruppenfarben, mit dem in Abbildung 4.4 vorgeschlagenen Prinzip, hervorgehoben werden.



**Abbildung 4.2.:** Bildung des Farbverlaufes eines Knotens anhand des dominierenden Zugehörigkeitwertes. Der äußere Endpunkt wird mit der Farbe der Gruppe, der innere Endpunkt mit der Hintergrundfarbe belegt. Ein dritter Farbpunkt dient zur Anpassung des Verlaufes. Ist die Zugehörigkeit groß (a), so wird dieser nahe an den inneren Endpunkt gesetzt. Ist die Zugehörigkeit klein (b), so wird dieser nahe an den äußeren Endpunkt gesetzt.



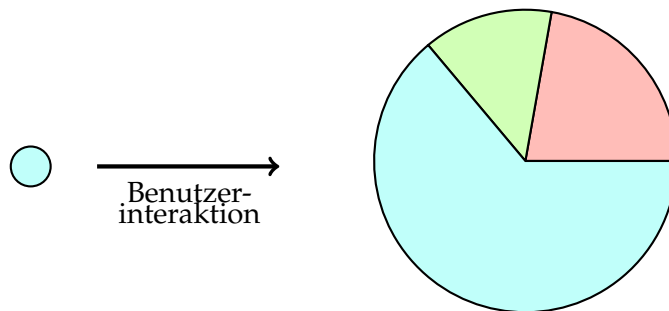
**Abbildung 4.3.:** Beeinträchtigung des Informationswertes bei falscher Anwendung eines Farbverlaufes. Abbildungen (b) und (c) sowie (d) und (e) geben jeweils dieselben Eigenschaften ab. Abbildung (c) als auch (e) verfälschen jedoch eine Analyse, durch den visuellen Eindruck eines kleineren Objektes.

Bedingt durch den nur begrenzt zur Verfügung stehenden Darstellungsraum ist es nicht möglich, beliebig viele Eigenschaften und Attribute in der Visualisierung abzubilden. Daher ist eine Zuhilfenahme von Interaktionstechniken notwendig, die sowohl eine Navigation (vorgestellt in Kapitel 3.1.4) als auch eine *Manipulation des Kontextes* ermöglichen. Hierzu werden Methoden beschrieben, die eine Auswahl bestimmter Heraushebungsmerkmale erlauben.

Neben der farblichen Hervorhebung eines Knotens mit der Farbe der Gruppe seiner stärksten Zugehörigkeit, ist auch die Betrachtung *all* seiner Gruppenzugehörigkeiten von großem Interesse. Eine dauerhafte Abbildung ist jedoch ab einer gewissen Ordnung des Graphen nur schwer möglich. Die Fläche, auf der ein Knoten gezeichnet wird, ist zu klein um alle diese Merkmale *lesbar* darzustellen. Die Interaktion in Form der Auswahl eines Knotens mit



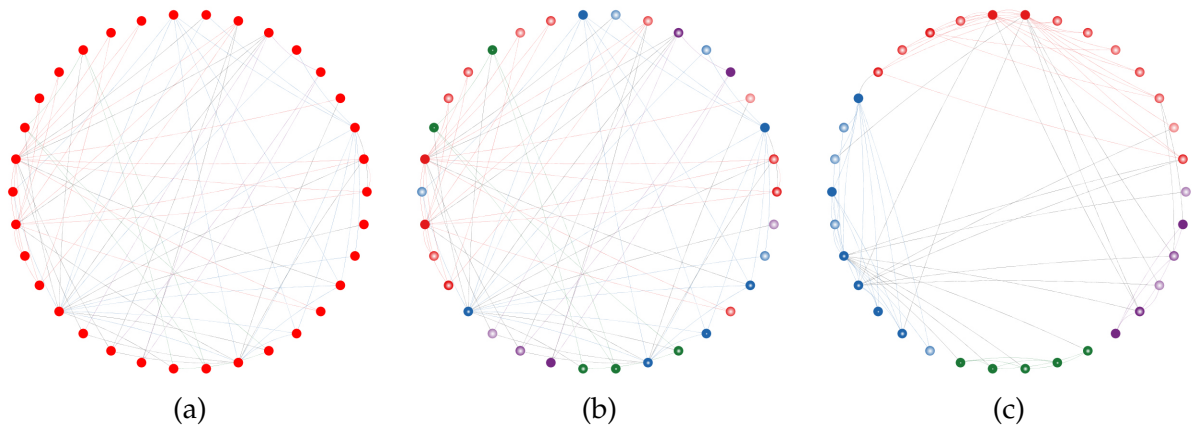
**Abbildung 4.4.:** Färbung eines Knotens als Möglichkeit weitere Informationen für den Benutzer zu Visualisieren. (a) Kante zwischen zwei Knoten desselben Clusters. (b) Kante zwischen zwei Knoten aus unterschiedlichen Clustern. Farblicher Verlauf, mit Umkehrung der Endpunkt-Farben erlaubt einfachere Kontexterkennung.



**Abbildung 4.5.:** Bei Bedarf könne durch Interaktion zusätzliche Informationen eingeblendet werden. Bei Auswahl eines Knotens werden alle Gruppenzugehörigkeiten in einem Kuchendiagramm gezeichnet.

darauf folgenden Einblenden der nötigen Informationen, stellt ein Werkzeug zur Lösung dieses Problems dar. Dieses Vorgehen ist bekannt unter dem Begriff *Details on demand*. Auf Auswahl eines oder einiger Knoten werden diese in einem größeren Maßstab dargestellt, dies ermöglicht das Einbinden genauerer Informationen. Beispielsweise können alle Zugehörigkeiten des Knotens in Form eines Kuchendiagramms dargestellt werden (Abbildung 4.5). Die Veränderung der Größe eines gezeichneten Knotens birgt das Risiko der Verdeckung anderer Knoten. Allerdings stellt dies, bedingt durch die nur temporäre Änderung, kein Risiko des Verlustes wichtiger Informationen dar.

## 4.2. Kreis-Ansicht



**Abbildung 4.6.:** Vergleich verschiedener Kreislayoutvariationen: (a) Layout ohne jegliche Färbung/Sortierung. (b) Layout mit Färbung der Knoten nach ihrer stärksten Zugehörigkeit. (c) Layout mit Färbung der Knoten sowie einer Sortierung nach Gruppen bzw. Zugehörigkeitswert innerhalb der Gruppen.

Eine simple aber für viele Analysezwecke sehr nützliche Anordnung der Knoten, ist die Positionierung in einem Kreis-Layout.

Knoten sollten in dieser Anordnung jedoch keine zufällige Position auf dem Kreis erhalten. Vielmehr kann auch hier die stärkste Gruppenzugehörigkeit für eine genauere Positionierung genutzt werden. Da jeder Knoten mindestens einer Gruppe zuordenbar ist, kann die stärkste Gruppenzugehörigkeit als Vergleichsparameter für eine erste, grobe Sortierung verwendet werden. Diese Sortierung ist eine erste Maßnahme hin zur Reduzierung des Visual Clutter, da nach der Definition von Gruppen (siehe Abschnitt 2.2.3) Knoten innerhalb der selben Gruppe stärker mit einander verbunden sind, als Knoten, die unterschiedlichen Gruppen zugehörig sind. Der im Falle von unscharfen Gruppenzugehörigkeiten ermittelte Grad der Unschärfe kann für eine weitere, feinere Sortierung der Knoten innerhalb der einzelnen Gruppen genutzt werden. Knoten die sehr stark der Gruppe zugehörig sind, haben eine hohe Wahrscheinlichkeit, dass sie mit vielen der restlichen Knoten der Gruppe verbunden sind. Eine mittige Anordnung dieser Knoten reduziert somit ebenfalls Visual Clutter Erscheinungen. Das Ordnen der Knoten erleichtert gleichzeitig das schnelle Erkennen der Gruppen und den unsicheren Knoten.

Die Knoten können mit dem in Abbildung 4.3 beschriebenen Verfahren gefärbt werden. Für eine bessere Unterscheidung von Kanten zwischen Knoten innerhalb einer Gruppe und Knoten verschiedener Gruppen, kann die Form der Kante genutzt werden. Beispielsweise können Kanten innerhalb einer Gruppe mit einer geraden Kante, Knoten unterschiedlicher Gruppen mit einer gebogenen Kante dargestellt werden.

Ähnlich zu der Allgemeinen Ansicht (4.1) ist das Einblenden von zusätzlichen Informationen bei Benutzerinteraktion sehr sinnvoll. Die in Abbildung 4.5 beschriebene Technik

ist jedoch nicht in dieser Art anwendbar. Eine für ein Kuchendiagramm benötigte vergrößerte Zeichnung eines Knotens würde in jedem Fall benachbarte Knoten überdecken. Es eignet sich daher die Abbildung aller Zugehörigkeiten eines Knotens in Form eines *Balken-Diagrammes*. Die Balken können über dem zugehörigen Knoten gezeichnet werden. Um Überdeckungen zu vermeiden, sollten sie dabei radial zum Mittelpunkt des Kreises ausgerichtet werden. Eine Sortierung der einzelnen Balken von großer nach geringer Zugehörigkeit, vereinheitlicht die Darstellung und verbessert somit zusätzlich die Lesbarkeit. Oftmals sind auch die Endpunkte der Kanten bei Betrachtung einer kleinen Menge an Knoten interessant. Hierfür eignet sich die farbliche Hervorhebung der Kanten ausgewählter Knoten. Somit ist es möglich, den Verlauf zu verfolgen und das Netzwerk auf bestimmte Verbindungen zu analysieren.

### 4.3. Aggregierung

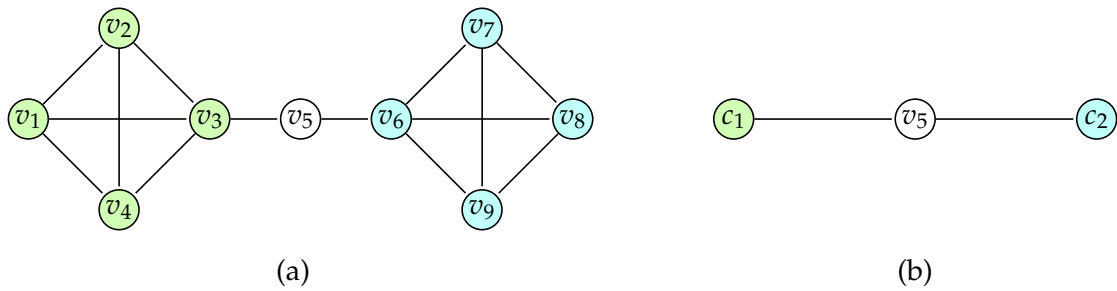
Ein allgegenwärtiges Problem bei der Darstellung von Netzwerken, ist die Gefahr des *Visual Clutter*, „[...] dem Zustand, in welchem überschüssige Elemente, deren Darstellung oder Organisationen zu einer Verschlechterung der Leistung in einigen Aufgaben führt“<sup>1</sup>. In der Visualisierung von Netzwerken tritt dieses Phänomen meist bei sehr dichten Graphen auf. Layout-Algorithmen positionieren zwar die Knoten unter Beachtung möglichst vieler Ästhetiken (3.1.2), jedoch skalieren sie schlecht bei dicht verbundenen Netzwerken. Besonders Kräftebasierte Ansätze produzieren bei zunehmender Dichte bzw. Ordnung, eine nicht optimale Platzierung [LKS<sup>+</sup> 11].

Neben dem vorgestellten Ansatz einer erweiterten Distanznorm, ist eine Aggregierung der Daten ein geeignetes Werkzeug um die Dichte eines Netzwerkes zu reduzieren ohne dabei jegliche Informationen zu verlieren. Besonders bei der Betrachtung von Gruppenzugehörigkeiten eignet sich die Bildung einer Norm für die Filterung, speziell in diesem Ansatz, einer *Aggregation* (siehe Abbildung 4.7) von ähnlichen Knoten. Eine einfache Norm dieser Art ist, einen Grenzwert festzulegen, der bestimmt ab wann ein Knoten „fest“ in eine Gruppe zugeordnet werden kann. Dies sorgt neben der Reduzierung der vorhandenen Knoten auch für eine Reduzierung der Kanten. Mehrfach auftretende Kanten können zu einer einzigen Kante verschmolzen werden. Bei richtiger Anwendung können die groben Zusammenhänge des Netzwerkes erhalten werden. Eine dynamische Anpassung des Grenzwertes ermöglicht das Finden einer optimalen Aggregierung der Knoten.

<sup>1</sup>Clutter is the state in which excess items, or their representation or organization, lead to a degradation of performance at some task [RLMJ05].

#### 4. Visualisierung unscharfer Gruppenzugehörigkeiten

---



**Abbildung 4.7.:** Verschmelzen der Knoten eines Clusters in einen „super-Knoten“. (a) Graph mit zwei Clustern (b) die Knoten der Cluster zu jeweils einem Knoten verschmolzen. Aufbau des Graphen ist weiterhin ersichtlich.

Eine alleinige Aggregation von Knoten und Kanten ist jedoch in vielen Fällen keine dauerhafte Lösung. Die aggregierten Gruppen sind bei Verwendung derselben Markierungen der normalen Knoten nicht von diesen ohne weiteres zu unterscheiden. Zusätzlich werden für eine Analyse interessante Werte, die es sich in die Visualisierung abzubilden lohnt, nicht genützt. Ein erster Ansatz ist es daher, die aggregierten Gruppen von den Knoten hervorzuheben. Idealerweise kann dies durch Verwendung einer unterschiedlichen geometrischen Form geschehen. Werden beispielsweise „normale“ Knoten durch einen Kreis repräsentiert, so bietet es sich an, die aggregierten Gruppen durch ein Rechteck o.ä. darzustellen. Auf dieser neuen Form aufbauend, können weitere Attribute abgebildet werden. Äquivalent zu einem „normalen“ Knoten, kann eine Färbung Aufschluss über die Stärke der Zugehörigkeit geben. Da es sich allerdings um eine Sammlung von Knoten handelt, gilt es hier den durchschnittlichen Wert bzw. den Median aller der aktuell in der Gruppe beinhalteten Zugehörigkeitswerte als Parameter für den Verlauf zu verwenden. Die Färbung kann dann mit dem in Abbildung 4.2 Ansatz umgesetzt werden.

Die Größe eines solchen Knotens kann für die Abbildung der Größe der Gruppe genutzt werden. Unterschiedliche Formgrößen wirken sich jedoch, je nach verwendetem Layout-Algorithmus, negativ auf die Visualisierung aus. Sind die Formen zu groß, so kommt es zu Überschneidungen oder im schlimmsten Fall zu kompletter Überdeckung kleiner Formen oder normalen Knoten. Es ist wichtig, dass dieser Sachverhalt zu jeder Zeit vermieden wird. Dies kann z. B. durch Verwendung einer angepassten Distanznorm, die die gezeichnete Größe der Form beachtet, geschehen. Ebenso erscheint es wenig sinnvoll die Größe direkt aus der Anzahl der beinhalteten Knoten zu berechnen. Ein Graph der sehr viele Knoten, jedoch nur einige Gruppen aufweist wird unlesbar, sobald ein aggregierter Knoten das gesamte Sichtfeld füllt und somit jegliche Informationen über die genau Einbindung in das Netzwerk überdeckt. Eine Berechnung der Größe mit einem fest definierten Maximum stellt die bessere Lösung des Problems dar. Dieser Faktor kann beispielsweise durch eine prozentuale Größe des Clusters zur Gesamtgröße des Graphen definiert werden:

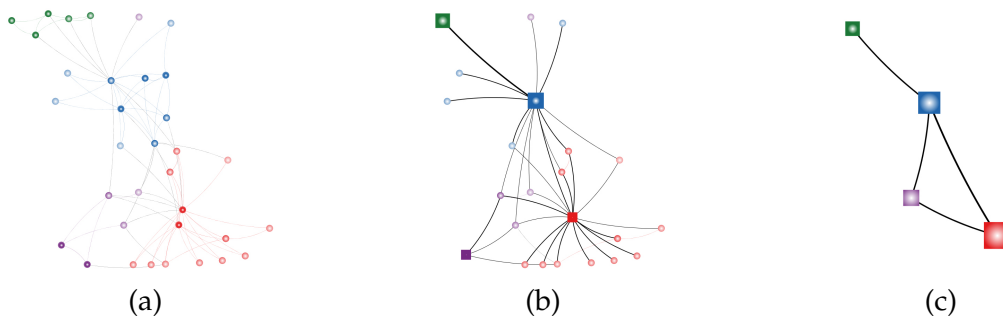
$$(4.2) \quad w = w_{max} * \left( \frac{|Cluster|}{|Graph|} \right)$$



Die direkte Abbildung der Werte auf die Größe, ist meist nicht sinnvoll [SMoo]. Der doppelte Wert eines Objektes würde einen doppelt so großen Durchmesser der Zeichnung, gleichzeitig aber auch den vierfachen Wert der verwendeten Fläche produzieren. Bei der Betrachtung der Darstellung ist bei dem visuellen Vergleich zweier Objekte die Fläche das entscheidende Maß [SMoo]. Eine Verdopplung des Durchmessers produziert somit einen deutlich größeren Wert Anstieg. Schumann schlägt vor, die Abbildung des Wertes mit Hilfe von z. B. der Stevenschen psychologischen Potenzfunktion zu berechnen:

$$(4.3) \quad x = (2 * \pi * r^2)^{0.7} \Leftrightarrow r = \frac{x^{\frac{5}{7}}}{\sqrt{2\pi}}$$

Das Aggregieren identischer Kanten kann dazu führen, dass Informationen über die *Verbundenheit* zwischen Gruppen bzw. zwischen einer Gruppe und einem Knoten verloren gehen. Eine Anpassung der Breite kann ähnlich wie bei der Größe der Knoten dazu genutzt werden, um diese Verbundenheit abzubilden. Existieren viele Kanten zwischen zwei Gruppen, so kann dies durch eine sehr breite Kante dargestellt werden. Äquivalent dazu, bei Existenz weniger Kanten durch eine dünne Kante. Es gilt aber auch hier, dass die Breite nicht beliebig skalierbar ist. Es müssen Maßnahmen identisch zu den der Größe eines Knotens getroffen werden, damit eine Kante nicht breiter gezeichnet wird, als die Knoten die sie verbindet.



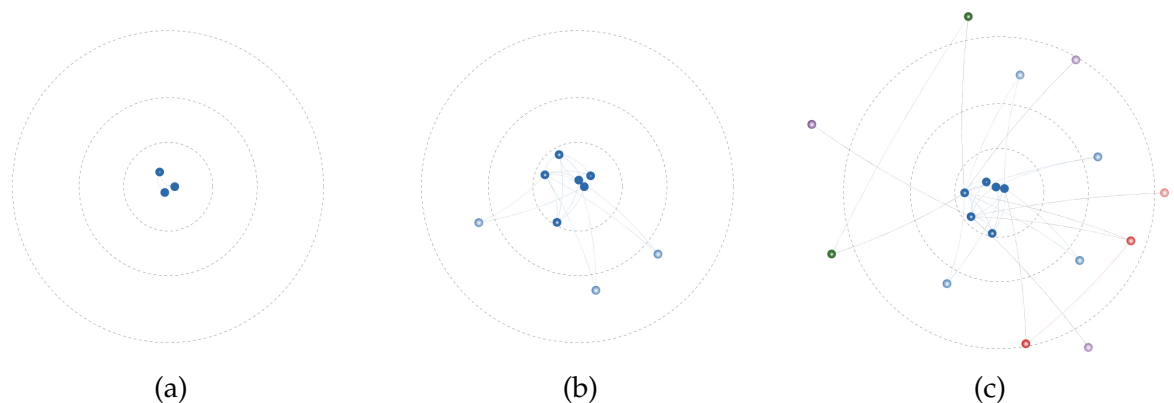
**Abbildung 4.8.:** Aggregiertes Layout mit verschiedenen Grenzwerten. (a)  $> 1$  . (b)  $> 75\%$  . (c)  $> 25\%$ .

## 4.4. Einzel-Ansicht

Die Betrachtung des Sub-Graphen eines einzelnen Clusters als eigenständiger Graph, kann bei großen Netzwerken von besonderem Interesse sein. Neben der *Unschärfe* des gesamten Clusters, ist auch die Struktur, die sich unter Einbeziehung der einzelnen Zugehörigkeiten bildet, von hohem Interesse. Es ist so möglich gruppeninterne Strukturen genauer zu analysieren. Die Platzierung der Knoten kann zwar dabei durch ein beliebiges Layout realisiert werden, es bietet sich aber an, die Interessen der Analyse mit einzubeziehen. Die verschiedenen starken Zugehörigkeitswerte der Knoten bilden eine Hierarchie ähnliche Struktur. Eine dem *Radialen-Baum-Layout* (siehe [HMMoo]) angelehnte Positionierung der

#### 4. Visualisierung unscharfer Gruppenzugehörigkeiten

---



**Abbildung 4.9.:** Einzel-Ansicht einer Gruppe mit verschiedenen Grenzwerten  $G$ . (a) Kern der Gruppe  $G \geq 90$ . (b)  $G \geq 50$ . (c)  $G \geq 10$ . Abbildung (c) zeichnet auch Objekte, deren stärkste Zugehörigkeit nicht in der ausgewählten Gruppe liegt. Die Ringe in der Abbildung dienen für die Orientierung. Sie repräsentieren jeweils fest definierte Grenzwerte: von innen nach außen: 90%, 60% und 20%.

Knoten, kann daher ein gute Grundlage für eine genauere Analyse bieten:

Definiert man den Kern eines Clusters als die Menge an Knoten mit einer Zugehörigkeit nahe bei oder gleich eins und ordnet man die verbleibenden Knoten in Ringe um diesen Kern, so erhält man eine Darstellung in der Cluster-interne Zusammenhänge klar ersichtlich sind. Der Abstand eines Knotens zum Zentrum des Kerns wird durch seine Zugehörigkeit bestimmt. Für einen Knoten  $v_i$  mit der Zugehörigkeit  $z_j$  zu dem gewähltem Cluster  $c_j$  gilt somit:

$$(4.4) \quad r = w - z_j * w$$

Wobei  $r$  der Radius des Ringes auf dem sich  $v_i$  befindet und  $w$  der maximale Abstand zum Zentrum des Kerns darstellt.

In einigen Anwendungsfällen ist es interessant nur Knoten ab einer gewissen Zugehörigkeit mit dieser Ansicht zu betrachten. Daher ist die Änderung des Grenzwertes, ab dem ein Knoten zu einem Cluster als zugehörig angesehen wird, eine empfehlenswerte Interaktionstechnik. Äquivalent zu der Änderung des Zugehörigkeit-Grenzwertes in der Aggregierten-Ansicht (siehe Abschnitt 4.3) kann so eine „dynamische“ Betrachtung ermöglicht werden. Eine Einbindung dieser Ansicht bei der Verwendung eines semantischen Zoom (vorgestellt in Abschnitt 3.1.4), kann in Kombination mit einer Aggregation, eine weitere Interaktionstechnik darstellen. Der Grenzwert muss hierbei nicht mittels eines Schieberegler o.ä. eingestellt werden, sondern wird je nach der Stufe des Zoom automatisch angepasst. Eine mögliche zukünftige Erweiterung ist ein automatischer Übergang aus einer aggregierten Ansicht in eine Einzel-Ansicht ab Erreichen einer definierten Stufe. Diese Technik würde dem Benutzer eine Fokussierung auf die alleinige Betrachtung / Analyse der Visualisierung erlauben. Es müssten keine zusätzlichen Einstellungen über Menüleisten getätigt werden.

## 4.5. Kombination

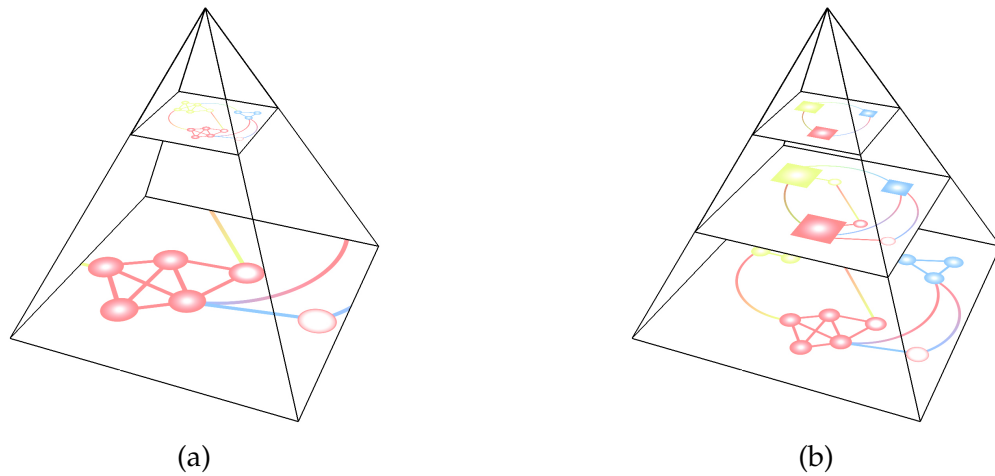
Viele der in den vorherigen Ansätzen und die in Abschnitt 3.1.2 beschriebenen Algorithmen für das Erzeugen eines geeigneten Layouts eines Graphen, sind nicht in der Lage, alle gewünschten ästhetischen Kriterien immer bzw. gleichzeitig zu erfüllen. Es werden Kompromisse eingegangen, die das Anwenden der Ansätze unter bestimmten Voraussetzungen ermöglicht. Es werden aber auch Einschränkungen in Kauf genommen, die z. B. die Skalierbarkeit auf beliebig größere Netzwerke einschränken. Es ist daher meist ein guter Ansatz, die verschiedenen Vorteile unterschiedlicher Layout- sowie Visualisierungs-Ansätze zu kombinieren und so aufkommende Probleme zu vermeiden.

In dieser Ausarbeitung wurde die Kombination der Einzel- bzw. der Kreis-Ansicht mit einem Kräftebasiertem Layout Algorithmus näher betrachtet: Dazu werden die Knoten anhand ihrer dominierenden Zugehörigkeit in Sub-Graphen aufgeteilt. Knoten innerhalb eines Sub-Graphen werden darauf folgend mit dem Ansatz der Einzel- bzw. der Kreis-Ansicht positioniert. Die Sub-Graphen werden anschließend durch ein Kräftebasiertes Verfahren auf der „globalen“ Ebene positioniert. Der Algorithmus für die Anordnung der Knoten lässt sich im Detail in drei Schritte aufteilen.

- (1) **Aufteilung** der Knoten in Sub-Graphen. Der dominierende Zugehörigkeitswert wird als Parameter für die Aufteilung verwendet. Ist der stärkste Zugehörigkeitswert  $z_j$  eines Knotens  $v_i$  zu der Gruppe  $j$  größer oder gleich zu einem definierten Grenzwert  $g$ , so wird er in den Sub-Graphen  $G_j$  eingefügt. Ist  $z_j$  kleiner  $g$ , so wird er zu der Gruppe  $j$  nicht als zugehörig angesehen. Die Aufteilung erfolgt für alle Knoten des Graphen.
- (2) **Lokale Positionierung** der Knoten innerhalb des Sub-Graphen. Jeder Sub-Graph wird in diesem Schritt als eigenständiges System aufgefasst. Die Positionierung der Knoten erfolgt dabei unabhängig zu der Positionierung der Knoten anderer Sub-Graphen.
- (3) **Globale Positionierung**. Die Sub-Graphen und verbleibenden, nicht zugeordneten Knoten, werden durch einen geeigneten Layout-Algorithmus auf einer globalen Ebene positioniert. Dabei werden die verbleibenden Knoten zu ihren Zugehörigkeitswerten passend zwischen den Einzelnen Sub-Graphen platziert. Die Sub-Graphen werden bei der Anordnung als *eine* aggregierte Einheit behandelt.

Dieser Ansatz ist besonders bei der Verwendung eines semantischen Zoom interessant. Er kann als Erweiterung der aggregierten-Ansicht (siehe Abschnitt 4.3) angesehen werden und somit in Kontext des semantischen Zoom als höchste Stufe genutzt werden. Abbildung 4.10 zeigt eine detaillierte Beschreibung dieses Einsatzbereichs.

Im Allgemeinen kann durch Veränderung des Grenzwertes eine sinnvolle Schnittstelle zu dem Benutzer geschaffen werden. Das Prinzip dieser Interaktion verhält sich identisch mit dem der aggregierten-Ansicht (Abschnitt 4.3), die interaktive Veränderung des Grenzwertes erlaubt es, die sich aus den Zugehörigkeitswerten bildenden Strukturen unterschiedlich detailliert darzustellen.



**Abbildung 4.10.:** Vergleich geometrischem zu semantischem Zoom. (a) Zoom vergrößert nur geometrische Figuren. (b) Einsatz verschiedener Ansichten auf verschiedenen Zoom-Stufen. Die kleinste Stufe kann durch eine Aggregation realisiert werden. Je „tiefer“ in die Darstellung hinein gezoomt wird, desto besser können durch eine Anpassung des Grenzwertes die Zugehörigkeiten der Knoten dargestellt werden. Die höheren Stufen können dabei durch die vorgestellte kombinierte Ansicht realisiert werden.

## 5. Evaluierung

In diesem Kapitel werden mit Hilfe von zwei Fallstudien unter Verwendung verschiedener Datensätzen, die in Kapitel 4 vorgestellten Ansätze für die Visualisierung unscharfer Gruppen bewertet und auf ihre Nutzbarkeit hin untersucht. Die Ansätze wurden für diesen Zweck in einem Java basierten Programm prototypisch umgesetzt. Das JUNG2 Framework (Java Universal Network/Graph Framework) dient hierbei als Grundlage für die Implementierung.

Die Datensätze **American Football Network** (5.1.1), sowie **Les Misérables Co-Appearance Network** (5.1.2) werden für die Ausführung der Fallstudie verwendet. Beide Datensätze können der Kategorie der Sozialen Netzwerke (2.2.4) zugeordnet werden. Sowohl die vorgestellten Ansätze, als auch die Implementierung können jedoch auch für Netzwerke anderer Kategorien verwendet werden.

Die unscharfen Gruppenzugehörigkeiten wurden mit dem in Abschnitt 3.3 vorgestellten Algorithmus von Neputz *et al.* [NPNBo7] ermittelt.

Die Fallstudien sind wie folgt aufgebaut: Sie beginnen mit einer kurzen Vorstellung des jeweiligen Datensatzes. Es werden aus der Beschreibung Thesen aufgestellt, die durch die vorgestellten Visualisierungsansätze geprüft werden sollen. Die Prüfung erfolgt mit einer detaillierte Untersuchung unter Zuhilfenahme der Implementierung. Es wird mit einer Evaluierung der vorgestellten Ansätze, die weitere Aspekte wie z. B. die Skalierbarkeit näher betrachtet, abgeschlossen.

### 5.1. Anwendung auf reale Datensätze

#### 5.1.1. American Football Network

Dieses ungerichtete Netzwerk repräsentiert geplante US College Football Spiele, aus der Saison 2001. Die Elemente/Knoten des Netzwerkes repräsentieren hierbei die einzelnen Mannschaften, wobei zwei Mannschaften genau dann in Relation zu einander stehen, wenn eine stattgefundenen Partie zwischen den beiden Mannschaften existiert [GN02][New11]. Das Netzwerk beinhaltet 115 Knoten sowie 615 Kanten.

Von den 115 Mannschaften der Saison 2001 sind bis auf 8 Mannschaften alle in einer *Conference*<sup>1</sup>. Die 8 unabhängigen Mannschaften werden im weiteren als *Independent* bezeichnet und bestehen aus: Notre Dame, Navy, Connecticut, Central Florida, Middle Tennessee State, Louisiana Tech, Louisiana Monroe und Louisiana Lafayette. Die Mannschaften haben jeweils zwischen 7 und 13 Spiele ausgetragen, im Durchschnitt waren es 10.7 Spiele, somit war es nicht möglich, dass jede Mannschaft gegeneinander spielen konnte. Die Mehrzahl der Spiele fand innerhalb einer Conference statt.

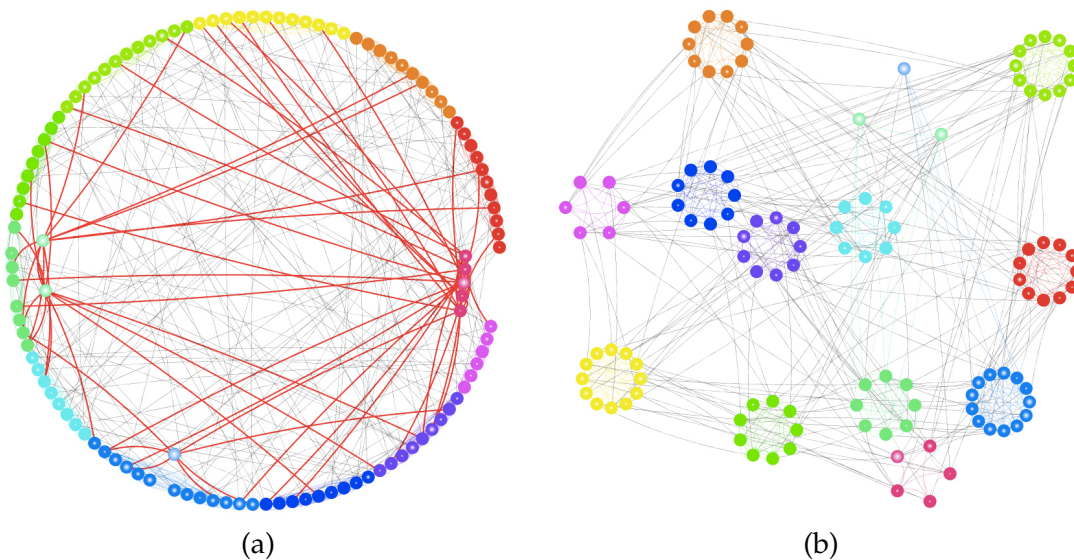
Aus der Beschreibung des Netzwerkes lassen sich verschiedene Thesen ableiten, die in erster Linie durch die entwickelten Visualisierungen bestätigt werden sollten.

- Das Netzwerk teilt sich in verschiedene Gruppen, den Conferences, auf.
- Spiele werden zum größten Teil innerhalb einer Conference abgehalten. Die Knoten sollten durchschnittlich einen starken dominierenden Zugehörigkeitswert besitzen.
- Durch üblicherweise in der Conference stattfindenden Spiele, sollten die einzelnen Gruppen eine hohe Interkonnektivität aufweisen.
- Umgekehrt, sollten die Gruppen durch die Einteilung in Conferences schwach untereinander verbunden sein.

Das Netzwerk teilt sich in 12 unterschiedliche Gruppen auf. Die Betrachtung mit Hilfe der aggregierten- sowie mit der Kreis-Ansicht bestätigen die Einteilung. Die Gruppen sind sich in ihrer Größe ähnlich, ihre durchschnittliche Gruppenzugehörigkeitsstärke ist bei allen Gruppen hoch. Sie bewegen sich in dem Intervall  $[0.8, 1.0]$ . Die meisten Elemente können mit einer relativ großen Sicherheit einer bestimmten Gruppe zugeordnet werden. Ab einem Zugehörigkeits-Grenzwert  $> 41\%$  können bereits alle Elemente einer Gruppen zugeordnet werden.

Anhand der Analyse mittels der Kreis- und kombinierten-Ansicht (vgl. Abbildung 5.1) fällt auf, dass die Independent-Mannschaften nicht als unabhängige, eigenständige Mannschaften erkannt werden. Mit *Louisiana Tech*, *Louisiana Monroe*, *Louisiana Lafayette* und *Middle Tennessee State* werden 5 der 8 Mannschaften einer eigenen (Independent-)Gruppe zugeordnet. Die

<sup>1</sup>Eine Conference bezeichnet die regionale Liga, in die sich die einzelnen Universitäten zusammen geschlossen haben. Die Spiele der Mannschaften werden größtenteils unabhängig voneinander in ihren zugehörigen Conferences ausgetragen. Spiele zwischen Mannschaften unterschiedlicher Conferences sind meist Bestandteil von Meisterschaften [Wik12].

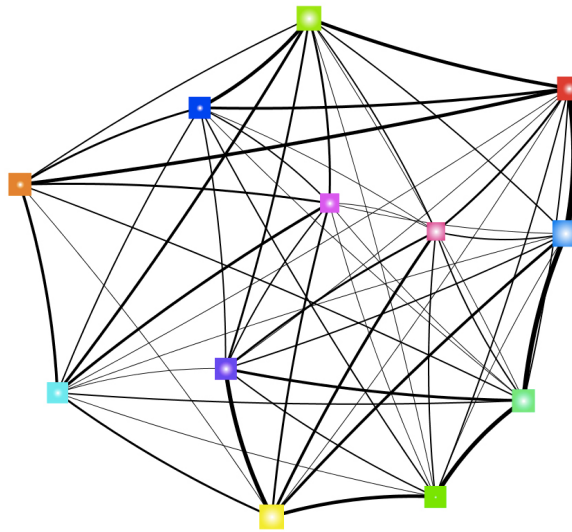


**Abbildung 5.1.:** American College Football – (a) Kreis-Ansicht: Aufteilung der Knoten in die 12 Gruppen. Hervorgehobenen Knoten stellen die „Independent“-Mannschaften dar. Rechts (pink) die „Independent“-Gruppe mit den Mannschaften *Louisiana Tech*, *Louisiana Monroe* und *Louisiana Lafayette* und *Middle Tennessee State*, unten-links (blau) *Connecticut* sowie links (grün) *Notre Dame* (unten) und *Navy* (oben). (b) Kombinierte-Ansicht: Mannschaften sind in Sub-Graphen angeordnet. Die einzelnen Knoten, stellen die Mannschaften mit den unsichersten Gruppenzugehörigkeiten dar. Auf Grund ihrer unscharfen Zugehörigkeitswerte konnten sie keinem Sub-Graphen zugeordnet werden. Es handelt sich dabei um die drei Mannschaften die nicht der „Independent“-Gruppe zugeordnet werden konnten.

verbleibenden Knoten ordnen sich anderen Gruppen zu. *Notre Dame* und *Navy* der grünen Gruppe sowie *Connecticut* der blauen Gruppe. Dieses Verhalten ist auf die verwendete Metrik des Detektionsalgorithmus (vgl. Abschnitt 3.3) zurückzuführen. Der Algorithmus nimmt die vorhandenen Relation zwischen den Elementen als Maß für die Zugehörigkeit. Die Independent-Mannschaften der detektierten Independent-Gruppe sind stark untereinander verbunden, vier der fünf Mannschaften bilden eine Clique. Die Mannschaften *Notre Dame* und *Navy* sind nicht mit der Independent-Gruppe verbunden, die Mannschaft *Connecticut* ist mit einer Relation, nur sehr schwach mit der Independent-Gruppe verbunden. Sie heben sich jedoch, durch ihre schwachen Zugehörigkeitswerte, klar von den ihnen zugewiesenen Gruppen ab. Dies ist in Abbildung 5.1 (b) anhand der Gradientenfärbung leicht zu erkennen.

Erwartungsgemäß sind die verbleibenden Elemente der Conference-Gruppen sehr stark verbunden. Sie bilden, ausnahmslos, vollständige Sub-Graphen.

Die Gruppen sind sich in ihrer Größe, sowie in ihre durchschnittlichen Gruppenzugehörigkeitsstärke ähnlich. Die Zugehörigkeitsstärke ist bei allen Gruppen sehr hoch, d.h.



**Abbildung 5.2.:** American College Football – Aggregierte-Ansicht (Grenzwert entspricht 41%) : Das Netzwerk teilt sich in 12 Gruppen auf. Der durchschnittliche Zugehörigkeitswert jeder Gruppe ist groß, das Netzwerk besteht daraus folgend aus Knoten mit sehr scharfer Gruppenzugehörigkeit.

das Netzwerk besteht aus vielen, sehr scharfen Elementen. Dieser Sachverhalt ist in Abbildung 5.2 anhand der Färbung der Knoten gut erkennbar. Eine nähere Betrachtung der einzelnen Gruppen zeigt eine starke Verbundenheit innerhalb der Gruppen auf. Die bilden in vielen Fällen vollständige Sub-Graphen.



### 5.1.2. Les Misérables

Dieses Netzwerk beschreibt die Relationen aus einer Menge an Figuren aus Victor Hugos Les Misérables. Zwei Figuren stehen in Relation zueinander, genau dann wenn sie in einer Szene gemeinsam auftreten. Das Netzwerk umfasst 77 Elemente sowie 254 Relationen.

Aus der Beschreibung des Netzwerkes lassen sich verschiedene Thesen ableiten:

- (1) Figuren, die in vielen Szenen/Kapiteln auftreten, sollten eine unsichere Gruppenzugehörigkeit ausweisen. Insbesondere die Hauptfiguren, die in vielen Szenen/Kapiteln auftreten sollten deutlich durch viele Gruppenzugehörigkeiten erkennbar sein, wohingegen die Nebenfiguren, mit wenigen Zugehörigkeiten, als sehr sichere Knoten erkennbar sein sollten.
- (2) Die Gruppierungen von Figuren:
  - Sie sollten in gewisser Weise die Kapitel des Romans widerspiegeln.
  - Sie sollten die Figuren in verschiedene Abstufungen anhand ihrer Wichtigkeit einteilen.
  - Sie sollten verschiedene Handlungsstränge wiedergeben.
- (3) Die verschiedenen Gruppen sollten schwach bis mäßig miteinander verbunden sein. Unter der Annahme, dass die Gruppen verschiedene Klassen von Figuren darstellen, sollte die Gruppe der Hauptfiguren stark mit dem restlichen Netzwerk verbunden sein.

Das Netzwerk teilt sich in 12 Gruppen auf. Diese sind dabei unterschiedlich groß und besitzen unterschiedliche durchschnittliche Zugehörigkeitsstärken. Kleinere Gruppen mit zwei bis vier Elementen besitzen dabei relativ hohe, größere Gruppen mit mehr als vier Elementen relativ niedrige Zugehörigkeitsstärken. (vgl. Abbildung 5.3 und 5.4)

Die Annahme der 2. These, die Gruppen repräsentieren einzelne Szenen oder Kapitel kann nicht nachgewiesen werden. Der Roman besteht aus fünf „Büchern“, mit jeweils einer unterschiedlicher Anzahl an Kapitel. Es kann keine Verbindung zwischen der Anzahl der vorhandenen Gruppen und dem Aufbau des Romans erkannt werden. Auch die Annahme, dass die Gruppen verschiedene Einteilungen der Figuren, gemessen an der Häufigkeit ihres Auftretens, repräsentieren, kann nicht bestätigt werden. Wäre dies der Fall, so würde die Gruppe um den Protagonisten des Romans, *Valjean*, weitere Hauptfiguren beinhalten. Zudem sollte sie stark mit den verbleibenden Gruppen des Netzwerkes verbunden sein. Letzteres Kriterium, wie in Abbildung 5.3 zu sehen, wird von der Gruppe um *Valjean* (gelbe Gruppe) erfüllt, sie ist mit 9 der 12 Gruppen des Netzwerkes verbunden. Entscheidender ist jedoch die Zusammensetzung der Gruppe: zwar sind mit den Figuren *Javert*, *M. Thénardier* und *Mme Thénardier* und *Éponine*, 5 der 11 Hauptfiguren in der Gruppe vertreten, jedoch beinhaltet die Gruppe 6 weitere Figuren. Dazu gehören sowohl größere Nebenfiguren, wie z. B. *Azelma*, als auch unbedeutende Figuren, wie z. B. die namenlose *Women 2*. Somit scheint es sehr unwahrscheinlich, dass die Gruppen die Wichtigkeit der Figuren widerspiegeln. Daher sollte die These der Gruppen als Repräsentation der einzelnen Handlungsstränge näher untersucht werden. Da die Gruppen in der Regel sowohl aus Hauptfiguren, als auch aus Nebenfiguren bestehen, liegt diese Annahme sehr nahe. Jedoch ist für weitere Schlussfolgerungen bzw. für einen eindeutigen Beweis dieser These ein tiefergehendes Wissen für den Roman *Les*

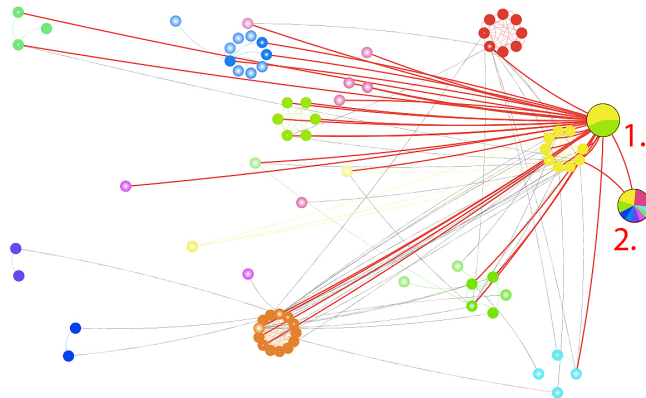
## 5. Evaluierung

---

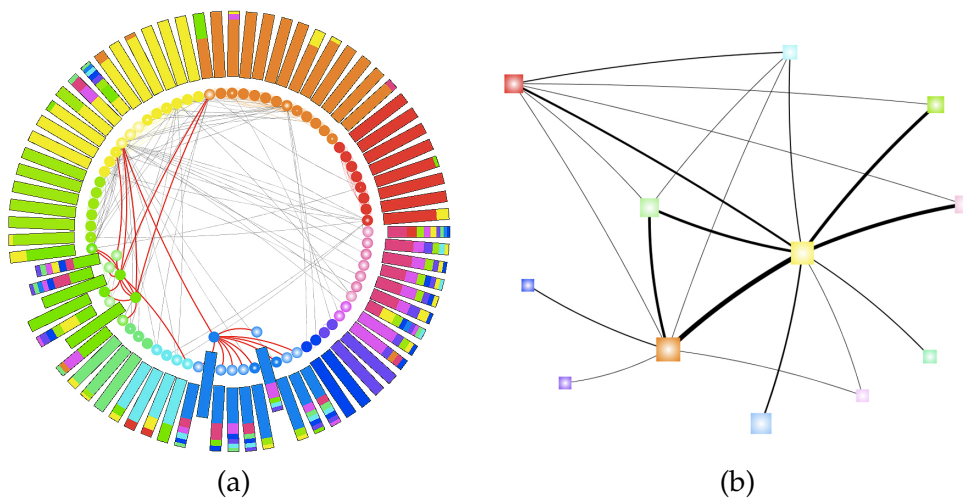
*Misérables* notwendig. Zu dem Zeitpunkt dieser Arbeit ist dieses jedoch nicht vorhanden, daher kann nur auf ein Zutreffen der dritten Annahme spekuliert werden.

Die Betrachtung der Ergebnisse für die These, die Hauptfiguren seien durch ihr häufiges Auftreten Objekte mit einer unsicheren Gruppenzugehörigkeit, erscheint ebenso als nicht nachweisbar. In Abbildung 5.4 ist deutlich zu erkennen, dass Hauptfiguren zum Teil als sichere Objekte mit wenigen Zugehörigkeiten, als auch als unscharfe Objekte mit vielen Zugehörigkeiten ermittelt werden. Vermeintliche Nebenfiguren werden ähnlich zu den Hauptfiguren auch als unscharfe bzw. als scharfe Objekte ermittelt. Auch Abbildung 5.3 verdeutlicht diese Beobachtung.

Zur Verdeutlichung dieses auffälligen Verhaltens, ist die Betrachtung des Protagonisten *Valjean* geeignet: der These entsprechend, sollte er viele Zugehörigkeiten besitzen, bei denen kein Wert einen starken dominierenden Term darstellt. Die Visualisierung zeigt jedoch auf, dass *Valjean* nur 2 Gruppen zugehörig ist. Dies ist insbesondere sehr fragwürdig, da er sehr gut mit dem Netzwerk außerhalb der ihm zugewiesenen Gruppe verbunden ist (vgl. Abbildung 5.3 (1.)). Hingegen ist die namenlose Nebenfigur, *Women 1* 10 der 12 Gruppen zugeordnet. Dieser Sachverhalt erscheint als fragwürdig bzw. als falsch, die Figur besitzt nur zu einer der zehn Gruppen eine Verbindung. Ein zufälliges Auftreten dieses Ergebnisses kann ausgeschlossen werden. Der fragwürdige Sachverhalt, tritt sowohl bei wiederholter Ausführung des Algorithmus, als auch an weiteren Stellen innerhalb des Netzwerkes auf. Abbildung 5.4 (a) zeigt mit Hilfe der Kreis-Ansicht weitere Objekte mit widersprüchlichen Zugehörigkeiten. Insbesondere ist die hellblaue Gruppe ein weiteres Indiz für einen Fehler des Detektionsalgorithmus. Das Zentrale Objekt dieser Gruppe, das Verbindungen zu allen weiten Objekten der Gruppe besitzt, besitzt eine eindeutige Zugehörigkeit. Objekte, die jedoch nur mit diesem Objekt verbunden sind, besitzen viele Gruppenzugehörigkeiten. Diese Widersprüche sind nicht auf einen Fehler der Visualisierung zurück zu führen. Es ist wahrscheinlich, dass es sich hierbei um Fehler des verwendeten Detektionsalgorithmus handelt



**Abbildung 5.3.:** Les Misérables – Widersprüchliche Ergebnisse: In der kombinierten-Ansicht sind die widersprüchlichen Ergebnisse des Fuzzyclustering Algorithmus deutlich erkennbar. Die Hauptfigur *Valjean* (1.), ist trotz vieler Relationen zu weiteren Figuren nur zwei Gruppen zugeordnet. Die unbedeutende Nebenfigur *Women 1* (2.) ist hingegen 10 der 12 Gruppen zugeordnet. Zu 9 dieser Gruppen existieren keinerlei Relationen. Diese fragwürdigen Ergebnisse deuten auf ein Fehler in dem verwendeten Algorithmus hin.



**Abbildung 5.4.:** Abbildung (a) hebt weitere Objekte, deren Zugehörigkeiten auf ein falschen Verhalten des verwendeten Detektionsalgorithmus schließen lassen, hervor. Die eingerückten Objekte besitzen Zugehörigkeiten zu Gruppen, zu denen sie nicht in Relationen stehen. Abbildung (b) stellt das Netzwerk in der Aggregierten-Ansicht dar. Anhand der Farbgradienten der Knoten ist deutlich die hohe Unschärfe der Gruppen zu erkennen. Die Dicke der Kanten verdeutlicht eine starke Verbundenheit der Gruppe um den Protagonisten *Valjean* (gelbe Gruppe), mit der orangen und pinken Gruppe. In diesen Gruppen sind viele der weiteren Hauptfiguren der Romans zugeordnet. Dies weist auf eine Bestätigung der These, die Gruppen repräsentieren verschiedene Handlungsstränge, hin.

### 5.2. Evaluierung der Ansätze

Die Ergebnisse der Falluntersuchung zeigen auf, dass die vorgestellten Ansätze gut für eine Analyse von Netzwerken auf unscharfe Gruppenzugehörigkeiten geeignet sind. Es können Unregelmäßigkeiten in den Cluster Ergebnissen erkannt und untersucht werden. Jedoch kann die Falluntersuchung nicht alle Aspekte, die für eine Beurteilung der Datensätze notwendig sind, abdecken. Aus diesem Grund werden im Folgenden die erstellten Ansätze auf verschiedene ästhetische Kriterien untersucht. Die Kriterien werden dabei in zwei verschiedene Kategorien unterteilt:

- Allgemein ästhetische Aspekte
  - Vermeidung bzw. Reduzierung des *Visual Clutter*
  - Reduzierung von Aliasing Effekten. Unter diesem Begriff werden u.a. Überlappungen von Elementen mit der Folge, dass der Benutzer sie nicht mehr unterscheiden kann, verstanden.
  - Skalierbarkeit der Ansätze auf Netzwerke höherer Ordnung bzw. Dichte.
- Aspekte für die Betrachtung von unscharfen Gruppen
  - Erkennen der unscharfen Gruppenstruktur
  - Erkennen von Elementen mit besonders unscharfen Gruppenzugehörigkeiten.
  - Darstellung aller Zugehörigkeiten der Elemente.

#### **Allgemeine-Ansicht:**

Sinn und Zweck der Allgemeinen-Ansicht, ist der Erhalt eines Überblicks über das Netzwerk. In der Vorstellung dieser Ansicht (siehe Abschnitt 4.1) wurde bereits die Problematik der Skalierbarkeit genannt. Durch die Verwendung der neuen Distanz Norm wird diese Problematik reduziert, jedoch ist eine unbegrenzte Größe der verwendeten Datensätze nicht möglich. Das Problem liegt hierbei nicht in der Skalierbarkeit der Objekte, sondern in der Skalierbarkeit der Relationen. Bei der Betrachtung von sehr dichten Netzwerken, ist die Gefahr des Visual Clutter groß, diese treten oft in Form von so genannten Haarbällen auf. Die Kanten überschneiden sich mit gezeichneten Knoten bzw. anderen Kanten und können nicht mehr als einzelne Relation erkannt werden. Somit sind die Aspekte des Visual Clutters und Aliasing nur für überschaubar groß Netzwerke erfüllt.

Betrachtet man die Eignung auf die Analyse von unscharfen Gruppenzugehörigkeiten, so ist die Allgemeine-Ansicht, unter der Beachtung der allgemeinen ästhetischen Aspekte, gut geeignet. Die verwendete Distanz Norm hebt klar die Strukturen der Gruppen und deren Einbindung in das gesamte Netzwerk hervor. Ebenso können die Eigenschaften der einzelnen Objekte sehr gut erkannt werden. Die vorgestellten Interaktionstechniken ermöglichen eine zusätzliche Einbindung von Informationen. So können auch alle Zugehörigkeiten eines Objektes dargestellt werden. Um Überdeckungen zu vermeiden, kann diese aber nicht dauerhaft und für alle Objekte genutzt werden.

Die Allgemeine-Ansicht erfüllt die Zielsetzung für eine geeignete Darstellung, für einen Einstieg in die Analyse des Netzwerkes. Die aufkommenden Probleme, vor allem die der

Skalierung, können durch andere Ansichten ausgeglichen werden.

### **Kreis-Ansicht:**

Die Kreis-Ansicht weist ähnliche Probleme in der Skalierung auf, wie sie in der Allgemeinen-Ansicht der Fall sind. Generell kann die Kreis-Ansicht aber besser auf Netzwerke eines höheren Grades bzw. mit einer höheren Dichte angewandt werden. Sie birgt jedoch ähnlich zur Allgemeinen-Ansicht die Gefahr der Haarball-Bildung. Bei Netzwerken mit einer sehr hohen Dichte, werden hierbei die Verbindungen auf der Zeichnungsfläche als eine Einzige wahrgenommen. Sie skaliert besser mit der Anzahl der Knoten. Hierbei ist jedoch die Anzahl der möglichen Winkel für die Positionierung der Objekte auf dem Kreis als Obergrenze. Es existiert zwar eine theoretisch unendliche Anzahl an möglichen Winkeln, jedoch ist sie durch Beachtung des ästhetische Kriterium der Vermeidung der Überlappung von Objekten, stark begrenzt. Analysen mit Netzwerken mit ca. 2000 Objekten legte eine Obergrenze, mit der eine Betrachtung des gesamten Netzwerkes möglich ist, fest. Allgemein ist es aber durch Verwenden eines Zoom möglich, Netzwerke mit mehr als 2000 Objekten zu betrachten. Dabei gehen jedoch eventuell wichtige Informationen über die Einbindung der einzelnen Objekte in den Gesamtkontext verloren.

Betrachtet man die Analyse von unscharfen Gruppenzugehörigkeiten, so ist diese Ansicht gut geeignet. Bedingt durch die Anordnung der Knoten auf einem Kreis, kann jedoch keine Information über die Strukturen innerhalb einer Gruppen dargestellt werden. Die Anordnung ermöglicht stattdessen, ein dauerhaftes Einblenden aller Zugehörigkeitswerte der Objekte. Unscharfe Objekte können durch den Farbgradienten leicht erkannt werden.

Die Kreis-Ansicht eignet sich für eine tiefere Analyse des gesamten Netzwerkes. Die Fallstudien zeigten auf, dass es anhand der Anordnung auf einem Kreis möglich ist, alle Objekte sowie die Relationen zwischen ihnen zu betrachten. Dies ermöglicht eine Auswahl von Objekten die für eine weitere Analyse interessant erscheinen. Die Einblendung aller Zugehörigkeitswerte in Form eines Balkendiagramms, ermöglicht die Betrachtung aller Zugehörigkeiten und unterstützt den Prozess der Auswahl der interessanten Objekte. Für die Visualisierung der Gruppen-internen Strukturen ist die Kreis-Ansicht ungeeignet. Es können keine Aussagen über die genaue Struktur der Gruppen getroffen werden.

### **Aggregierte-Ansicht:**

Im Gegensatz zu der Allgemeinen- oder der Kreis-Ansicht, bietet die Aggregierte-Ansicht optimal skalierbare Plattform. Ab einem gewissen Grenzwert ist es möglich, alle Knoten Gruppen zu aggregieren. Die Fallstudien zeigten auf, dass auch sehr große Netzwerke mit mehreren zehntausend Knoten, meist eine geringe Anzahl an Gruppen besitzen. Dies ergibt eine starke Reduktion der zu visualisierenden Datenmenge und ermöglicht eine kompakte Darstellung des Netzwerkes. Zusätzlich reduziert die Aggregation der Kanten Visual Clutter Erscheinungen. Gleichzeitig erhält die Verwendung der Dicke der Kanten für die Abbildung der Anzahl der Relationen zwischen zwei Objekten, die Information über die Verbundenheit des Netzwerkes. Jedoch ist durch die Anpassung der Größe der aggregierten Knoten die Gefahr der Überschneidung von verschiedenen Knoten gegeben. Dieses Problem lässt sich, entgegen der vorgeschlagenen Anpassung der Distanz Norm, nicht für jedes beliebige Netzwerk vermeiden.

Sie eignet sich insbesondere für die Analyse der Strukturen, die sich aus den Verbindungen

zwischen den einzelnen Gruppen ergeben. Allerdings verhindert die Aggregation der Knoten eine Visualisierung der Strukturen innerhalb einer Gruppe. Die Abbildung der durchschnittlichen Gruppenzugehörigkeit auf den Farbgradienten, ermöglicht eine Identifikation der Gruppen die vielen Knoten mit hoher Unsicherheit beinhalten. Jedoch können weitere Informationen, wie die einzelnen Zugehörigkeiten der beinhalteten Knoten, aufgrund der Darstellung mit einem einzelnen Knoten pro Gruppe, nicht abgebildet werden.

### **Einzel-Ansicht:**

Die Einzel-Ansicht kann als eigenständige Ansicht genutzt werden, bei verschiedenen Fallstudien stellte sich jedoch heraus, dass sie besser als ergänzende Erweiterung der Aggregierten-, bzw. der Kombinierten-Ansicht geeignet ist. Der Verlust der Einbindung der Gruppe in den Gesamtkontext des Netzwerkes, ermöglicht nur Aussagen über Strukturen innerhalb der Gruppe.

Bewertet man die Ansicht auf die aufgestellten allgemeinen ästhetischen Aspekte, stellt sich heraus, dass Visual Clutter Erscheinungen nicht oder nur selten auftreten. Bei Betrachtung der einzelnen Gruppen, reduziert sich die zu visualisierende Datenmenge. Auch die Positionierung der Knoten auf Kreisen um einen Kern reduziert zusätzlich diese Gefahr. Bedingt durch die Annahme, dass die Knoten mit starker Zugehörigkeit stark mit weiteren Knoten mit starker Zugehörigkeit verbunden sind, Knoten mit schwacher Zugehörigkeit nur schwach mit weiteren Knoten der Gruppe verbunden sind, konzentrieren sich die Relationen auf den Kern der Ansicht. Überschneidungen von Knoten und von Kanten sind nur in Ausnahmefällen vorhanden.

Betrachtet man die Eignung für die Analyse von unscharfen Gruppenzugehörigkeiten, so ist sie bedingt einsetzbar. Die Reduzierung der zu visualisierenden Elemente auf den Sub-Graphen der jeweiligen Gruppe, ist unter alleiniger Verwendung, keine Betrachtung der Globale Struktur möglich. Im Gegensatz dazu eignet sich die Ansicht sehr gut für die Betrachtung der Struktur innerhalb der Gruppe, die sich unter der Verwendung des Zugehörigkeitswertes bilden. Ebenso sind besonders unscharfe Objekte, durch die Verwendung des Gradienten für die Abbildung der Grad der Unschärfe des Objektes, schnell zu erkennen. Zusätzlich ist es mit Hilfe von Interaktionstechniken möglich, alle Zugehörigkeitswerte der Objekte einzublenden und für eine tiefere Analyse zu nutzen.

### **Kombinierte-Ansicht:**

Ziel der Kombinierten-Ansicht ist es, die Vorteile verschiedener Ansichten zu vereinen, um so eine ausführliche Analyse eines Netzwerkes auf unscharfe Gruppenzugehörigkeiten zu ermöglichen.

Die in dieser Arbeit vorgeschlagenen Kombination der Einzel-Ansicht und der Allgemeinen Ansicht erfüllt diese Zielsetzung leider nur bedingt: Die Einteilung der Knoten durch ihre stärkste Gruppenzugehörigkeit, reduziert zwar die Gefahr des Visual Clutter, jedoch ist bei großen Netzwerken diese Gefahr immer noch präsent. Die Einteilung ermöglicht jedoch auch eine Reduzierung des verwendeten Platzes. Daher ist diese Ansicht ähnlich skalierbar wie die Allgemeine-Ansicht. Die Kombinierte-Ansicht ist, bedingt durch die starke Verbundenheit innerhalb einer Gruppe jedoch besser, sowohl in der Anzahl der Knoten, als auch in der Anzahl der Kanten skalierbar. Es kann jedoch bei großen Netzwerken zu Überschneidungen von den Elementen kommen, sehr dichte Netzwerke führen in vielen

Anwendungsfällen zu den genannten Visual Clutter Erscheinungen.

In Bezug auf die Analyse von unscharfen Gruppenzugehörigkeiten ist die Kombinierte-Ansicht gut geeignet. Sie kann als Fortführung der Aggregierte-Ansicht angesehen werden. Ähnlich wie diese, ist ein dauerhafter globaler Überblick über das Netzwerk möglich. Zusätzlich können jedoch auch die Strukturen innerhalb der Gruppe ohne zusätzlich Verwendung von weiteren Werkzeugen betrachtet werden. Auch die Einbindung dieser Strukturen bzw. einzelner Objekte der Gruppen ist jeder Zeit erkennbar. Ähnlich zu allen Ansichten, ist durch die Verwendung eines Farbgradienten eine Erkennung von besonders unscharfen Elementen möglich. Eine Betrachtung aller Zugehörigkeiten ist jedoch nur bedingt durch die Verwendung von Interaktionstechniken möglich. Bei dem zusätzlichen Einblenden aller Zugehörigkeitswerte nach der beschriebenen Methode, besteht jedoch auch grundsätzlich die Gefahr von möglichen Überdeckungen von Knoten. Diese Methode ist daher generell nur bedingt geeignet.

Kriterien		Ansichten				
		Allgemeine-Übersicht	Kreis-Ansicht	Aggregierte-Ansicht	Einzel-Ansicht	Kombinierte-Ansicht
Allgemeine ästhetische	Visual Clutter	o	o	+	+	o
	Kompakte Darstellung	o	+	+	+	+
	Aliasing	o	+	o	+	o
	Skalierbarkeit der Knoten	+	o	+	+	+
	Skalierbarkeit der Kanten	-	-	+	+	o
(unscharfe-) Gruppen	Erkennung der globalen Struktur des Netzwerkes	+	o	+	-	+
	Erkennung der Gruppen Struktur	+	-	o	+	+
	Erkennung besonders unscharfer Elemente	+	+	-	+	+
	Erkennung aller Zugehörigkeiten der Elemente	o	+	-	o	o

**Tabelle 5.1.:** Wertung der vorgestellten Ansätze – Es wird in zwei verschiedenen Kategorien unterschieden: allgemeine ästhetische Aspekte und Aspekte der Betrachtung unscharfer Gruppenstrukturen.

Die Wertung erfolgt in: - schlecht, o bedingt, + gut





## 6. Zusammenfassung und Ausblick

Dieses Kapitel fasst Erkenntnisse dieser Arbeit zusammen und zeigt Möglichkeiten zur Weiterentwicklung.

### Zusammenfassung

In dieser Arbeit wurde die Visualisierung von Gruppenzugehörigkeiten vorgestellt. Es wurden zuerst konventionelle Ansätze für die Betrachtung eindeutiger Zugehörigkeiten, anschließend neue bzw. Erweiterungen bestehender Ansätze für die Betrachtung von unscharfen Zugehörigkeiten vorgestellt. Die Herausforderung lag dabei in dem Umgang mit dem Begriff der Unschärfe. Mittels den vorgestellten Ansätzen, ist eine ausführliche Analyse der unscharfen Gruppenzugehörigkeiten und den sich daraus bildenden Strukturen möglich. Verschiedene Analyse Interessen werden mit Hilfe einer Aufteilung auf verschiedene Ansichten abgedeckt. Die Allgemeine-Übersicht dient als Einstieg, mit der Kreis-Ansicht wird durch die Einteilung der Elemente in ihre jeweils dominierenden Gruppen eine erste Betrachtung auf die existierenden Gruppen ermöglicht. Die Aggregierte-, Kombinierte- sowie Einzel-Ansicht bieten für eine detaillierte Betrachtung der internen Gruppenstrukturen, sowie für die Zusammenhänge zwischen den einzelnen Gruppen, ein geeignetes Werkzeug. Die konzipierten Werkzeuge für eine Benutzerinteraktion bieten übergreifend auf alle vorgestellten Ansätze die Möglichkeit einer tiefer gehenden Analyse. Es ist möglich bestimmte Aspekte näher zu betrachten bzw. bestimmte Attribute hervorzuheben.

Die Fallstudie und Evaluation zeigten jedoch auch die Schwierigkeiten bei sehr großen bzw. sehr dichten Netzwerken auf. Bedingt durch die bei vielen Ansätzen zu Grunde liegenden Kräftebasierten Layout Algorithmen, entstehen trotz der Anpassungen Visual Clutter Erscheinungen. Daher sind die vorgestellten Ansätze nur bedingt auf Netzwerke mit mehreren tausend Knoten bzw. Kanten skalierbar.

Entgegen dieser Probleme eignen sich die vorgestellten Ansätze gut für eine Analyse von unscharfen Gruppenzugehörigkeiten. Die entdeckten Unregelmäßigkeiten in den Ergebnissen des verwendeten Detektionsalgorithmus, unterstützt diese Aussage. Die Ansätze eignen sich so neben der Analyse von Netzwerken, auch für die Verifikation von Algorithmen zur Detektierung unscharfer Gruppenstrukturen.

### **Ausblick**

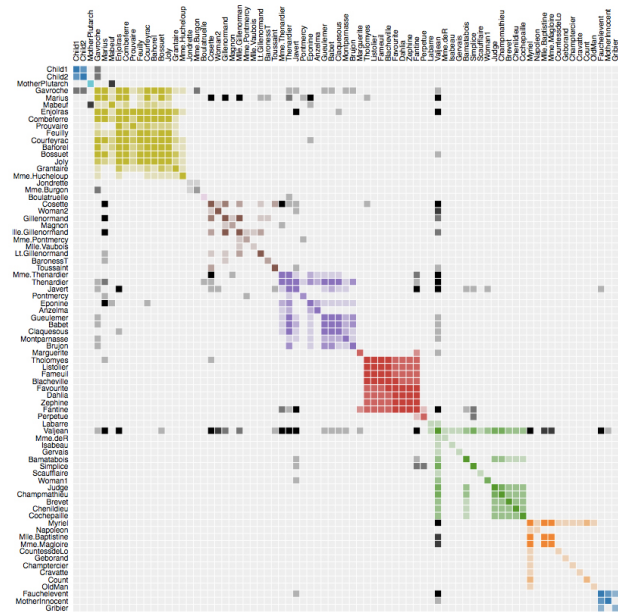
Die in der Fallstudie verwendeten Datensätze sind den kleineren Datensätzen zuzuordnen. Für die Arbeit mit heutigen realen (Sozialen-)Netzwerken mit Millionen an Elementen bzw. mehreren Milliarden an Verbindungen (Facebook besitzt Stand September 2012 mehr als eine Milliarde aktive Benutzer [Fac12]), müssen die vorgestellten Ansätze, speziell die Graphzeichnen Grundlage in der Skalierbarkeit angepasst werden. Eine mögliche Fortführung wäre zusätzlich die Betrachtung von Gruppenzugehörigkeiten in sich zeitlich verändernden Netzwerken. Hierfür müssen jedoch zuerst neue Algorithmen für die Detektierung von unscharfen Gruppen entwickelt werden, mit denen eine Echtzeit nahe Berechnung ermöglicht wird.

## A. Anhang

Abbildung A.1 stellt für die bessere Lesbarkeit eine vergrößerte Version der Abbildung 3.7 dar.



(a)



(b)

Abbildung A.1.: Vergrößerte Version des Beispiels in Abbildung 3.7

# Literaturverzeichnis

- [Ber83] J. Bertin. *Semiology of graphics: diagrams, networks, maps; [tab.]*. University of Wisconsin Press, Madison, Wis. [u.a.], 1983. (Zitiert auf Seite 28)
- [BETT94] G. D. Battista, P. Eades, R. Tamassia, I. G. Tollis. Algorithms for Drawing Graphs: an Annotated Bibliography. *Comput. Geom.*, 4:235–282, 1994. (Zitiert auf Seite 22)
- [BETT99] G. D. Battista, P. Eades, R. Tamassia, I. G. Tollis. *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice-Hall, 1999. (Zitiert auf den Seiten 19 und 22)
- [Bez81] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA, 1981. (Zitiert auf Seite 30)
- [Bos12] M. Bostock. *Les Misrables Co-occurrence*, 2012. URL <http://bost.ocks.org/mike/miserables/>. (Zitiert auf Seite 29)
- [Bra99] U. Brandes. *Layout of Graph Visualizations*. Dissertation, 1999. (Zitiert auf Seite 14)
- [BRSG07] C. Bennett, J. Ryall, L. Spalteholz, A. Gooch. The aesthetics of graph visualization. In *Proceedings of the Third Eurographics conference on Computational Aesthetics in Graphics, Visualization and Imaging*, Computational Aesthetics'07, S. 57–64. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 2007. doi: 10.2312/COMPAESTH/COMPAESTH07/057-064. URL <http://dx.doi.org/10.2312/COMPAESTH/COMPAESTH07/057-064>. (Zitiert auf Seite 19)
- [CYP96] Z. Chi, H. Yan, T. Pham. *Fuzzy Algorithms - With Applications to Image Processing and Pattern Recognition*. World Scientific, Singapur, 1996. (Zitiert auf Seite 10)
- [Die10] R. Diestel. *Graphentheorie*. Springer, Heidelberg [u.a.], 4. Aufl. Auflage, 2010. URL <http://swbplus.bsz-bw.de/bsz329802542inh.htm>. (Zitiert auf Seite 11)
- [Dun73] J. C. Dunn. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 3(3):32–57, 1973. (Zitiert auf Seite 30)
- [Fac12] Facebook. *Key Facts*, 2012. URL <https://newsroom.fb.com/content/default.aspx?NewsAreaId=22>. (Zitiert auf Seite 58)
- [For09] S. Fortunato. Community detection in graphs. *CoRR*, abs/0906.0612, 2009. (Zitiert auf Seite 29)

- [For10] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75 – 174, 2010. doi:DOI:10.1016/j.physrep.2009.11.002. URL <http://www.sciencedirect.com/science/article/B6TVP-4XPYXF1-1/2/99061fac6435db4343b2374d26e64ac1>. (Zitiert auf Seite 24)
- [FR91] T. M. J. Fruchterman, E. M. Reingold. Graph Drawing by Force-directed Placement. *Softw., Pract. Exper.*, 21(11):1129–1164, 1991. (Zitiert auf Seite 34)
- [GN02] M. Girvan, M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002. URL <http://dx.doi.org/10.1073/pnas.122653799>. (Zitiert auf Seite 46)
- [HMM00] I. Herman, G. Melançon, M. S. Marshall. Graph Visualization and Navigation in Information Visualization: A Survey. *IEEE Transactions on Visualization and Computer Graphics*, 6(1):24–43, 2000. doi:10.1109/2945.841119. URL <http://dx.doi.org/10.1109/2945.841119>. (Zitiert auf den Seiten 11, 20, 21, 24, 25 und 41)
- [Holo6] D. Holten. Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data. *IEEE Transactions on Visualization and Computer Graphics*, 12:741–748, 2006. doi:<http://doi.ieeecomputersociety.org/10.1109/TVCG.2006.147>. (Zitiert auf den Seiten 18, 26 und 27)
- [HRF10] N. Henry Riche, J.-D. Fekete. Novel Visualizations and Interactions for Social Networks Exploration. In Springer, Herausgeber, *HANDBOOK OF SOCIAL NETWORK TECHNOLOGIES AND APPLICATIONS*, S. 611–636. Springer, 2010. doi:10.1007/978-1-4419-7142-5\\_28. URL <http://hal.inria.fr/hal-00698894>. (Zitiert auf den Seiten 18, 19, 20 und 28)
- [Hua07] W. Huang. Using eye tracking to investigate graph layout effects. In S.-H. Hong, K.-L. Ma, Herausgeber, *APVIS*, S. 97–100. IEEE, 2007. (Zitiert auf Seite 20)
- [HW10] D. Holten, J. J. van Wijk. Evaluation of Cluster Identification Performance for Different PCP Variants. *Comput. Graph. Forum*, 29(3):793–802, 2010. (Zitiert auf Seite 23)
- [KAF<sup>+</sup>08] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, G. Melançon. Information Visualization. Kapitel Visual Analytics: Definition, Process, and Challenges, S. 154–175. Springer-Verlag, Berlin, Heidelberg, 2008. doi:10.1007/978-3-540-70956-5\_7. URL [http://dx.doi.org/10.1007/978-3-540-70956-5\\_7](http://dx.doi.org/10.1007/978-3-540-70956-5_7). (Zitiert auf Seite 15)
- [KK89] T. Kamada, S. Kawai. An algorithm for drawing general undirected graphs. *Inf. Process. Lett.*, 31(1):7–15, 1989. doi:10.1016/0020-0190(89)90102-6. URL [http://dx.doi.org/10.1016/0020-0190\(89\)90102-6](http://dx.doi.org/10.1016/0020-0190(89)90102-6). (Zitiert auf den Seiten 21 und 34)
- [Kre05] L. Krempel. *Visualisierung komplexer Strukturen: Grundlagen der Darstellung mehrdimensionaler Netzwerke*. Campus-Verl., Frankfurt a. M. [u.a.], 1. Aufl. Auflage, 2005. URL [http://digitool.hbz-nrw.de:1801/webclient/DeliveryManager?pid=1918546&custom\\_att\\_2=simple\\_viewer](http://digitool.hbz-nrw.de:1801/webclient/DeliveryManager?pid=1918546&custom_att_2=simple_viewer). (Zitiert auf Seite 14)

- [KW01] M. Kaufmann, D. Wagner, Herausgeber. *Drawing Graphs, Methods and Models*, Band 2025 von *Lecture Notes in Computer Science*. Springer, 2001. (Zitiert auf Seite 19)
- [LKS<sup>+</sup>11] T. von Landesberger, A. Kuijper, T. Schreck, J. Kohlhammer, J. J. van Wijk, J.-D. Fekete, D. W. Fellner. Visual Analysis of Large Graphs: State-of-the-Art and Future Research Challenges. *Comput. Graph. Forum*, 30(6):1719–1749, 2011. (Zitiert auf den Seiten 18, 20, 22 und 39)
- [Mey98] B. Meyer. Self-Organizing Graphs - A Neural Network Perspective of Graph Layout. In *Proceedings of the 6th International Symposium on Graph Drawing, GD '98*, S. 246–262. Springer-Verlag, London, UK, UK, 1998. URL <http://dl.acm.org/citation.cfm?id=647550.729090>. (Zitiert auf Seite 34)
- [New10] M. Newman. *Networks: An Introduction*. Oxford University Press, Inc., New York, NY, USA, 2010. (Zitiert auf Seite 7)
- [New11] M. E. J. Newman. *Network data*, 2011. URL <http://www-personal.umich.edu/~mejn/netdata/>. (Zitiert auf Seite 46)
- [NG04] M. E. J. Newman, M. Girvan. Finding and evaluating community structure in networks. *Physical Review*, E 69(026113), 2004. (Zitiert auf Seite 30)
- [NPNB07] T. Nepusz, A. Petrczi, L. Ngyessy, F. Bacs. Fuzzy communities and the concept of bridgeness in complex networks. Technischer Bericht arXiv:0707.1646, 2007. Comments: 10 pages, 7 figures. (Zitiert auf den Seiten 7, 31 und 45)
- [PDFV05] G. Palla, I. Dernyi, I. Farkas, T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005. (Zitiert auf Seite 30)
- [PSD09] M. Pohl, M. Schmitt, S. Diehl. Comparing the readability of graph layouts using eyetracking and task-oriented analysis. In *Proceedings of the Fifth Eurographics conference on Computational Aesthetics in Graphics, Visualization and Imaging, Computational Aesthetics'09*, S. 49–56. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 2009. doi:10.2312/COMPAESTH/COMPAESTH09/049-056. URL <http://dx.doi.org/10.2312/COMPAESTH/COMPAESTH09/049-056>. (Zitiert auf Seite 20)
- [Pur97] H. C. Purchase. Which Aesthetic has the Greatest Effect on Human Understanding? In *Proceedings of the 5th International Symposium on Graph Drawing, GD '97*, S. 248–261. Springer-Verlag, London, UK, UK, 1997. URL <http://dl.acm.org/citation.cfm?id=647549.728779>. (Zitiert auf Seite 20)
- [RLM]05] R. Rosenholtz, Y. Li, J. Mansfield, Z. Jin. Feature congestion: a measure of display clutter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '05*, S. 761–770. ACM, New York, NY, USA, 2005. doi:10.1145/1054972.1055078. URL <http://doi.acm.org/10.1145/1054972.1055078>. (Zitiert auf Seite 39)

- [SB94] M. Sarkar, M. H. Brown. Graphical fisheye views. *Commun. ACM*, 37(12):73–83, 1994. doi:10.1145/198366.198384. URL <http://doi.acm.org/10.1145/198366.198384>. (Zitiert auf Seite 25)
- [SGSH11] P. G. Sun, L. Gao, S. Shan Han. Identification of overlapping and non-overlapping community structure by fuzzy clustering in complex networks. *Inf. Sci.*, 181(6):1060–1071, 2011. doi:10.1016/j.ins.2010.11.022. URL <http://dx.doi.org/10.1016/j.ins.2010.11.022>. (Zitiert auf Seite 30)
- [SM00] H. Schumann, W. Müller. *Visualisierung: Grundlagen und allgemeine Methoden*. Springer, Berlin, 2000. URL <http://www.bsz-bw.de/cgi-bin/ekz.cgi?SWB06919688>. (Zitiert auf den Seiten 14 und 41)
- [STT81] K. Sugiyama, S. Tagawa, M. Toda. Methods for Visual Understanding of Hierarchical System Structures. *Systems, Man and Cybernetics, IEEE Transactions on*, 11(2):109–125, 1981. doi:10.1109/TSMC.1981.4308636. (Zitiert auf Seite 20)
- [Tam12] R. Tamassia. *Handbook of Graph Drawing and Visualization*. CRC Press, 2012. URL <http://www.cs.brown.edu/~rt/gdhandbook/>. (Zitiert auf den Seiten 11, 20, 21 und 22)
- [Tim02] H. Timm. *Fuzzy-Clusteranalyse: Methoden zur Exploration von Daten mit fehlenden Werten sowie klassifizierten Daten*. Dissertation, Otto-von-Guericke-Universität Magdeburg, 2002. (Zitiert auf Seite 9)
- [Wik12] Wikipedia. *College Football — Wikipedia, Die freie Enzyklopädie*, 2012. URL [http://de.wikipedia.org/wiki/College\\_Football#Struktur](http://de.wikipedia.org/wiki/College_Football#Struktur). (Zitiert auf Seite 46)
- [Zad65] L. A. Zadeh. Fuzzy Sets. *Information and Control*, 8:338–353, 1965. URL <http://www-bisc.cs.berkeley.edu/Zadeh-1965.pdf>. (Zitiert auf Seite 9)
- [Zad78] L. Zadeh. Fuzzy sets as a basis for theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978. (Zitiert auf Seite 9)
- [Zim99] H.-J. Zimmermann, Herausgeber. *Practical applications of fuzzy technologies*. Kluwer, Boston [u.a.], 1999. (Zitiert auf Seite 10)
- [ZWZ07] S. Zhang, R.-S. Wang, X.-S. Zhang. Identification of Overlapping Community Structure in Complex Networks Using Fuzzy c-means Clustering. *Physica A: Statistical Mechanics and its Applications*, 374(1):483–490, 2007. (Zitiert auf den Seiten 7 und 30)

Alle URLs wurden zuletzt am 18. November 2012 geprüft.



## **Erklärung**

Hiermit versichere ich, diese Arbeit selbständig verfasst und nur die angegebenen Quellen benutzt zu haben.

---

(Thomas Reinhardt)