

# Eine relationale Strategie zur Einteilung von Gruppen auf Basis flüchtiger Kontakte



Diplomarbeit  
von  
Gabi Dorf Müller

Universität Konstanz  
Fachbereich Mathematik und Statistik

Februar 2005

Betreuer: Prof. Dr. Ulrik Brandes,  
Fachbereich Informatik und Informationswissenschaft



# Zusammenfassung

Die Ermittlung von Gruppen zusammengehöriger Objekte findet in vielen Bereichen des Lebens Anwendung. So möchte man auch Personen in „gute“ Arbeitsgruppen einteilen. Dabei bedeutet die Bezeichnung „gut“, dass die Zusammenarbeit der jeweiligen Personen erfolgreich ist. Eine mögliche Voraussetzung dafür, dass Personen gut zusammen arbeiten, ist Sympathie. Aus diesem Grund wird eine Strategie zur Gruppeneinteilung entwickelt, welche derartige Sympathien berücksichtigt. Der Knödel-Graph erweist sich dabei als günstiger Graph, um flüchtige Kontakte zwischen den Teilnehmern auszudrücken. Anschließend ausgesprochene Sympathien und Antipathien werden durch entsprechende Kantengewichte in diesem Graphen dargestellt. Zur Einteilung in gleich große Gruppen eignen sich schließlich verschiedene Sektionierungsalgorithmen auf Basis des Kernighan-Lin-Algorithmus.



# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>7</b>
1.1	Ursprung . . . . .	7
1.2	Problemstellung . . . . .	8
1.3	Überblick . . . . .	9
<b>2</b>	<b>Grundlagen der Graphentheorie</b>	<b>11</b>
2.1	Definitionen . . . . .	11
2.2	„Gossiping“ und „Broadcasting“ . . . . .	15
2.3	Partitionierung und Clusterung . . . . .	17
<b>3</b>	<b>Kontaktgraphen</b>	<b>19</b>
3.1	Anforderungen an den Kontaktgraphen . . . . .	21
3.2	Wahl eines Kontaktgraphen . . . . .	25
3.3	Anwendungsbeispiel . . . . .	41
<b>4</b>	<b>Bewertungsformen</b>	<b>43</b>
4.1	Erzeugung des Sympathie- und Antipathiegraphen . . . . .	43
4.2	Vereinigung von Sym- und Antipathiegraph? . . . . .	45
4.3	Erfüllung aller Antipathien? . . . . .	47
4.4	Wahl der Bewertungsform . . . . .	48
<b>5</b>	<b>Einteilung der Gruppen</b>	<b>49</b>
5.1	Problemstellung . . . . .	49
5.2	Optimierungsverfahren . . . . .	52
5.3	$\mathcal{NP}$ -Vollständigkeiten . . . . .	53
5.4	Minimale Bisektionierung . . . . .	61
5.5	Minimale Sektionierung . . . . .	69
5.6	Gruppeneinteilung durch MinMax-Sektionierung . . . . .	73
<b>6</b>	<b>Experimente</b>	<b>77</b>
6.1	Generierung der Daten . . . . .	78

6.2	Verschiedene Parameter . . . . .	79
6.3	Darstellungsform der Ergebnisse . . . . .	82
6.4	Experimentelle Ergebnisse . . . . .	85
6.5	Auswertung . . . . .	141
<b>7</b>	<b>Zusammenfassung und Ausblick</b>	<b>145</b>
	<b>Anhang</b>	<b>149</b>
<b>A</b>	<b>„Gossiping“-Zeit des <math>K_n</math></b>	<b>149</b>
<b>B</b>	<b>Bestimmte <math>(r, t)</math>-Cages</b>	<b>151</b>
<b>C</b>	<b>Inhalt der beigefügten CD</b>	<b>153</b>
	<b>Bibliographie</b>	<b>155</b>
	<b>Danksagung</b>	<b>157</b>
	<b>Eidesstattliche Erklärung</b>	<b>159</b>

# Kapitel 1

## Einführung

### 1.1 Ursprung

Die Idee für die Entwicklung einer Strategie zur Einteilung von Gruppen auf der Basis flüchtiger Kontakte entstammt einer Art Spiel in der Partnervermittlung. Die temporeiche Partnersuche wird *Speed-Dating* genannt und kommt aus den USA. Erfunden wurde es 1998 von einer jüdischen Glaubensgemeinde in Los Angeles. Man wollte damit Ehen zwischen gleich gesinnten Gläubigen fördern. Bei diesem Spiel sitzen je ein Mann und eine Frau beisammen und haben 3 bis 10 Minuten Zeit, sich kennen zu lernen. Beim Gongschlag rücken die Männer einen Tisch weiter zum nächsten „Turbo-Flirt“, so lange, bis sie mit allen Frauen gesprochen haben. Danach gibt jeder Teilnehmer auf einer Art Stimmzettel an, welche seiner Gegenüber er sympathisch gefunden hat, welche Frau bzw. welchen Mann er/sie gerne wieder treffen möchte. Schließlich gibt der Speed-Dating-Veranstalter nur bei gegenseitigem Interesse die Adressen der jeweiligen Teilnehmer weiter.

Mittlerweile wird dieses „Kuppelspiel“ in vielen deutschen Großstädten oder im Internet<sup>1</sup> veranstaltet. Auch in Konstanz wurde im Sommer 2004 ein solches Event organisiert (siehe [Sta04]). Daran nahmen 16 Singles aus zwei verschiedenen Altersstufen teil, von denen jeder vier Personen des anderen Geschlechts kennen lernte. Die Auswertung der Stimmzettel ergab bei dieser Veranstaltung keinerlei gegenseitige Sympathie.

Solche flüchtigen Kontakte werden nun verwendet, um beispielsweise für Seminare oder ähnliche Veranstaltungen Gruppen einzuteilen. Man vermutet nämlich, dass eine Gruppe von Personen, die sich nicht kennen, besser als Team zusammen arbeitet, wenn möglichst viele der Personen Sympathie füreinander empfinden. Ähnlich wie beim Speed-Dating reden die Teilnehmer

---

<sup>1</sup> siehe <http://www.speeddating.de>

in Zwiegesprächen miteinander und geben anschließend an, mit welchen ihrer Gesprächspartner sie in einer Gruppe zusammen arbeiten möchten. Von diesen Sympathien, ob erwidert oder nicht, sollen nun möglichst wenige, durch eine Einteilung in verschiedene Gruppen, getrennt werden. Wird zudem die Aussprache von Antipathien erlaubt, so sollen Personen, die sich nicht sympathisch finden, nicht in dieselbe Gruppe eingeteilt werden.

## 1.2 Problemstellung

Eine Gruppe von  $n$  Personen wird in  $g$  Gruppen, deren Größen sich jeweils höchstens um eine Person unterscheiden, eingeteilt. Ziel hierbei ist, möglichst viele der nach den flüchtigen Kontakten ausgesprochenen Sympathien und Antipathien zu *erfüllen*. Dabei bedeutet die *Erfüllung* der Sym- und Antipathien, dass Personen, die sich sympathisch finden, in dieselbe Gruppe eingeteilt werden, und Personen, die sich nicht mögen, in verschiedene.

Für die flüchtigen Kontakte ist zu entscheiden, wer sich mit wem in einem Zwiegespräch unterhält und in welcher Reihenfolge dies geschehen soll, damit keine großen Wartezeiten für die Teilnehmer entstehen. Gibt es weitere Kriterien, die die Art und Weise der Gesprächsführung erfüllen sollte? Es ist zu überlegen, welche Formen der gegenseitigen Bewertung möglich sind. Sie soll die empfundene Sym- bzw. Antipathie ausdrücken und für die Lösung des betrachteten Problems sinnvoll eingesetzt werden können. Ist es zum Beispiel möglich, Gruppen zu bilden, die alle ausgesprochenen Antipathien zugleich erfüllen? Mit Hilfe welcher Algorithmen können nun Einteilungen in möglichst gleich große Gruppen gefunden werden, die möglichst wenige Sympathien und viele Antipathien trennen? Anschließend ist zu prüfen, ob die entwickelte Strategie nach gewissen Kriterien „gute“ Ergebnisse liefert.

Zur Darstellung und Untersuchung der Gesprächsführung und der Bewertungen, sowie zur Gruppeneinteilung durch Anwendung eines Algorithmus, werden Graphen als zugrunde liegende Datenstruktur verwendet. Betrachtet man die Gesprächsführung, so sind diese ungerichtet und ungewichtet. Werden daraufhin Bewertungen ausgesprochen, so arbeitet man auf gerichteten und eventuell gewichteten Graphen. Mit Hilfe graphentheoretischer Sätze und Methoden, sowie mathematischer Überlegungen wird schließlich eine relationale Strategie zur Einteilung von Gruppen entwickelt. Die Verwendung flüchtiger Kontakte zur Vergabe von Sympathie- und Antipathiebewertungen basiert auf Versuchen von Kommunikationswissenschaftlern, die belegen, dass die erste kurze Begegnung schon über die zwischenmenschliche Beziehung auf lange Sicht entscheidet (siehe [SR04]).

Die Strategie soll letztendlich zur Gruppeneinteilung von 20 bis 100 Per-



sonen benutzt werden. Der Einsatz bei Veranstaltungen mit mehr Teilnehmern scheint wenig sinnvoll, da hier zu viele Gespräche notwendig wären, um signifikante Gruppen zu erhalten. Da in der Anwendung häufig Gruppen bestimmter Größe gewünscht werden, und erfahrungsgemäß Teams bestehend aus 6 Personen gut zusammenarbeiten, wird jeweils eine Einteilung in entsprechend viele Gruppen vorgenommen.

## 1.3 Überblick

Zu Beginn wird eine kleine Einführung in die Grundlagen der Graphentheorie gegeben. Hier werden die in dieser Arbeit verwendeten graphentheoretischen Begriffe und Bezeichnungen definiert. Darauf folgen drei Kapitel, in denen die drei Arbeitsschritte zur Entwicklung der relationalen Strategie erläutert werden:

- Schritt 1: Wahl eines Kontaktgraphen
- Schritt 2: Wahl einer Bewertungsform
- Schritt 3: Einteilung der Gruppen

Für die Wahl eines Kontaktgraphen, der die flüchtigen Kontakte zwischen den Personen ausdrückt, werden in Kapitel 3 Kriterien ausgearbeitet, die der Graph erfüllen sollte. Daraufhin werden verschiedene Graphklassen auf ihre Einsetzbarkeit als Kontaktgraph untersucht. In Kapitel 4 werden unterschiedliche Formen der gegenseitigen Bewertung betrachtet, die die empfundenen Sym- und Antipathien ausdrücken. Anhand dieser Bewertungen soll schließlich die Gruppeneinteilung vorgenommen werden. Dazu wird in Kapitel 5 ein Algorithmus entwickelt, der die Personen so in Gruppen einteilt, dass möglichst viele der ausgesprochenen Sym- und Antipathien erfüllt sind. Kapitel 6 beschreibt abschließend Experimente, die untersuchen, ob die entwickelte Strategie „gute“ Gruppeneinteilungen liefert. Dazu werden Variablen der verschiedenen Algorithmen verändert und die Erfüllung verschiedener Kriterien beobachtet.



# Kapitel 2

## Grundlagen der Graphentheorie

Dieses Kapitel fasst einige grundlegende Definitionen aus der Graphentheorie zusammen. Die eingeführten Bezeichnungen werden in den folgenden Kapiteln verwendet.

### 2.1 Definitionen

#### 2.1.1 Graphen

Ein *Graph*  $G = (V, E)$  setzt sich aus der endlichen, nicht leeren Menge  $V$  von *Knoten* und der Menge  $E$  von *Kanten* zusammen. Dabei ist die Menge  $E$  eine binäre Relation auf  $V$ . Gilt  $e = (v_1, v_2) \in E$  mit  $v_1, v_2 \in V$ , so bezeichnet man  $v_1$  und  $v_2$  allgemein als *Endknoten* der Kante  $e$ . Ein Graph wird *gerichtet* genannt, wenn sich die Kanten aus geordneten Paaren von Knoten zusammensetzen, und *ungerichtet*, wenn diese ungeordnet sind.

Zur Veranschaulichung eines Graphen werden seine Knoten durch Kreise und seine Kanten durch Verbindungslinien oder -pfeile zwischen den Kreisen dargestellt. Dabei werden Pfeile benutzt, um die Richtung einer Kante anzugeben.

Ein Graph  $G' = (V', E')$  heißt *Subgraph* von  $G = (V, E)$ , wenn  $V'$  nichtleere Teilmenge von  $V$  und  $E'$  Teilmenge von  $E$  ist. Der Ausdruck  $E|V^* \subseteq E$  bezeichnet die Menge aller Kanten aus  $E$ , deren Endknoten Element von  $V^* \subseteq V$  sind. Der mit Hilfe dieser Bezeichnung definierte Graph  $G[V^*] = (V^*, E|V^*)$  wird der *von  $V^*$  induzierte Subgraph* von  $G$  genannt. Analog ist  $V|E^* \subseteq V$  die Menge aller Knoten aus  $V$ , die Endknoten eines Elementes von  $E^* \subseteq E$  sind. Dann bezeichnet  $G[E^*] = (V|E^*, E^*)$  den *von*

$E^*$  induzierten Subgraphen von  $G$ .

Der Komplement  $\bar{G} := (V, \bar{E})$  eines Graphen  $G = (V, E)$  ist der Graph mit Kantenmenge  $\bar{E} := \{(v_1, v_2) \mid (v_1, v_2) \notin E\}$ . Der Graph  $\bar{G}$  hat also die gleiche Knotenmenge wie  $G$  und enthält alle Kanten, die nicht Element von  $G$  sind.

Weitere Informationen, die die Struktur eines Graphen beinhalten können, sind *Knoten-* bzw. *Kantengewichte*. Auf den Knoten bzw. Kanten ist dann eine Kostenfunktion  $w : V \rightarrow \mathbb{R}$  bzw.  $w : E \rightarrow \mathbb{R}$  definiert. In beiden Fällen spricht man von einem *gewichteten Graphen* mit *Knoten-* bzw. *Kantengewichtsfunktion*  $w$ . Das Gewicht von Kanten, welche nicht zu  $E$  gehören, beträgt stets 0. Gilt für alle Knoten  $v \in V$   $w(v) = 1$  bzw. für alle Kanten  $e \in E$   $w(e) = 1$ , so nennt man den Graphen *ungewichtet*.

Die folgenden Definitionen und Bezeichnungen werden oftmals nur für ungerichtete Graphen angegeben, da im Verlauf dieser Arbeit die Richtung der Kanten nur selten eine Rolle spielt. Es sind ungerichtete, ungewichtete Graphen gemeint, wenn von Graphen die Rede ist. Nur in Fällen, in denen die Differenzierung von Interesse ist, werden die Bezeichnungen ‚gerichtet‘, ‚ungerichtet‘, ‚gewichtet‘ und ‚ungewichtet‘ eingesetzt. Ist allerdings aus dem Kontext klar ersichtlich, dass auf dem jeweiligen Graphen eine Gewichtsfunktion definiert ist, wird das Wort ‚gewichtet‘ weggelassen.

## 2.1.2 Knoten und Kanten

Sei  $G = (V, E)$  ein gerichteter Graph. Eine Kante  $e \in E$  repräsentiert dann ein geordnetes Paar  $(v_1, v_2)_g$  von Knoten  $v_1, v_2 \in V$ . Es heißt,  $e$  *verbindet*  $v_1$  und  $v_2$ , und  $v_2$  ist *adjazent* zu  $v_1$ . Die Kante  $e$  ist *inzident* zu  $v_1$  und  $v_2$ . Der Knoten  $v_1$  wird *Startknoten* der Kante  $e$  und  $v_2$  ihr *Endknoten* genannt. Kanten, die  $v$  als Start- und Endknoten haben, heißen *Schleifen*. In einem ungerichteten Graphen  $G = (V, E)$  heißen  $v_1$  sowie  $v_2$  *Endknoten* der Kante  $e = (v_1, v_2)$ . Außerdem ist in diesem Graph sowohl  $v_1$  *adjazent* zu  $v_2$ , als auch  $v_2$  *adjazent* zu  $v_1$ . Besitzen zwei Kanten  $e_1, e_2 \in E$  einen gemeinsamen Endknoten, so bezeichnet man sie ebenfalls als *inzident*. Die Graphen, die hier betrachtet werden, sind *einfach*, enthalten also keine Schleifen oder mehr als eine Kante zwischen zwei bestimmten Knoten (*Mehrfachkanten* genannt). Gilt  $|V| = n$ , so nennt man den Graphen von der *Ordnung*  $n$ .

Die *Nachbarschaft*  $N(v)$  eines Knotens  $v \in V$  ist die Menge aller Knoten, die mit  $v$  verbunden sind. Die Elemente von  $N(v)$  werden *Nachbarn* von  $v$  genannt. Ein Knoten heißt *isoliert*, wenn er keine Nachbarn besitzt. Die *Nachbarschaft*  $N(V')$  einer echten Teilmenge  $V' \subset V$  ist definiert als die Vereinigung aller Nachbarschaften  $N(v)$  der Knoten  $v \in V'$ .

Der *Grad*  $\deg(v)$  eines Knotens  $v \in V$  ist die Anzahl Kanten, die inzident

zu  $v$  sind. Der *Minimalgrad*  $\delta(G) := \min_{v \in V} \deg(v)$  bezeichnet den kleinsten Grad und der *Maximalgrad*  $\Delta(G) := \max_{v \in V} \deg(v)$  den größten Grad der Knoten eines Graphen  $G$ . Ein Graph heißt *regulär*, wenn  $\delta = \Delta$  gilt, also alle Knoten den gleiche Grad haben. Setzt man  $r := \delta = \Delta$ , so nennt man  $G$  auch  $r$ -regulär.

Ein *Weg*  $W$  ist eine Sequenz  $(v_0, v_1, \dots, v_k)$  von Knoten  $v_i \in V$ , so dass  $e_i = (v_{i-1}, v_i) \in E$  für alle  $i \in \{1, 2, \dots, k\}$  gilt. Manchmal wird auch die Schreibweise  $W = (e_1, e_2, \dots, e_k)$  verwendet. Es bezeichnet  $V(W) = \{v_0, v_1, \dots, v_k\}$  die *Knotenmenge* des Weges und  $E(W) = \{e_1, e_2, \dots, e_k\}$  seine *Kantenmenge*. Ein solcher Weg *verbindet* die *Endknoten*  $v_0$  und  $v_k$  des Weges und hat die *Länge*  $k$ . Man bezeichnet  $W$  als *Zykel*, falls der Startgleich dem Endknoten ist, also  $v_0 = v_k$  gilt. Ein Zykel heißt speziell *Kreis*, falls nur Start- und Endknoten identisch sind, das heißt, falls  $v_0 = v_k$  und für alle  $i, j \in \{0, \dots, k-1\}$ ,  $i \neq j$  gilt, dass  $v_i \neq v_j$ .

Mit diesen Bezeichnungen definiert man dann den *Kreisgraphen*  $C_n$ , den Graphen der Ordnung  $n$ , der einen Kreis der Länge  $n$  enthält. Der  $C_3$  heißt auch *Dreieck*. Ein Graph ohne Dreiecke als Subgraphen bezeichnet man als *dreiecksfrei*. Zwei Knoten heißen *zusammenhängend*, wenn im Graphen ein Weg existiert, der die beiden Knoten verbindet. Jeder Subgraph, dessen Knoten zusammenhängen, heißt *Zusammenhangskomponente* des Graphen. Besteht ein Graph nur aus einer einzigen Zusammenhangskomponente, so nennt man ihn *zusammenhängend*.

Ein zusammenhängender Graph ohne Kreise wird *Baum* genannt. Die Bäume bilden somit eine besondere Klasse von Graphen. In diesen wird jeder Knoten mit Grad 1 *Blatt* genannt.

Der *Abstand*  $d(v_1, v_2)$  zweier Knoten  $v_1, v_2 \in V$  ist die Länge eines kürzesten Weges, der  $v_1$  und  $v_2$  verbindet. Falls ein solcher nicht existiert, so wird der Abstand unendlich gesetzt. Der *Durchmesser*  $D(G)$  eines Graphen bezeichnet den größten Abstand zwischen zwei Knoten des Graphen, also  $D(G) = \max_{v_1, v_2 \in V} d(v_1, v_2)$ . Die Länge eines kürzesten nicht trivialen Kreises heißt *Tailenweite*  $T(G)$ . Die Tailenweite eines Baumes wird unendlich gesetzt.

Es bezeichne  $N_{=j}(v^*)$  die Menge aller Knoten  $v \in V$ , für die  $d(v^*, v) = j$ , und  $N_{\leq j}(v^*)$  die Menge aller Knoten  $v \in V$ , für die  $d(v^*, v) \leq j$  gilt. Sie sind weitere Nachbarschaften von Knoten. Dann ist  $n_j := \max_{v \in V} |N_{=j}(v)|$  die Anzahl Knoten der größten Nachbarschaft  $N_{=j}(v^*)$  eines Knotens  $v^* \in V$ . In einem einfachen Graphen  $G = (V, E)$  gelten folglich:  $N(v) = N_{=1}(v)$  für jeden Knoten  $v \in V$  und  $\Delta(G) = n_1$ .

### 2.1.3 Chromatische Zahl

Die *Chromatische Zahl*  $Ch(G)$  eines Graphen  $G$  ist die kleinste Anzahl an Farben, die benötigt werden, um die Knoten des Graphen so einzufärben, dass keine zwei adjazente Knoten dieselbe Farbe haben. Ein Graph  $G$  mit Chromatischer Zahl  $c := Ch(G)$  wird auch *c-färbbar* genannt. Die 2-färbbaren Graphen sind äquivalent zu den *bipartiten Graphen*. Deren Knotenmenge lässt sich so in zwei disjunkte Mengen  $V_1, V_2 \subsetneq V$  teilen, dass  $V = V_1 \cup V_2$  und keine zwei Knoten aus derselben Menge  $V_i$  adjazent sind. Die bipartiten Graphen enthalten keine Kreise ungerader Länge, da deren Knoten abwechselnd in  $V_1$  und  $V_2$  liegen. Ist bei der Betrachtung eines Graphen  $G = (V, E)$  seine Bipartitheit entscheidend, so wird er als  $G = (V_1 \cup V_2, E)$  notiert, wobei  $V_1$  und  $V_2$  die disjunkten Knotenmengen mit  $V = V_1 \cup V_2$  sind. Gilt dabei  $|V_1| = |V_2|$ , so werden die Knoten des bipartiten Graphen der Ordnung  $|V_1| + |V_2|$  meist so nummeriert, dass  $V_1 = \{0, 2, 4, \dots, |V_1| + |V_2| - 2\}$  und  $V_2 = \{1, 3, 5, \dots, |V_1| + |V_2| - 1\}$ . Die Menge  $V_1$  enthält somit die Knoten *gerader Nummerierung*, wobei die 0 zu den geraden Zahlen gerechnet wird, und  $V_2$  besteht aus den Knoten *ungerader Nummerierung*.

### 2.1.4 Cliquenzahl

Ein Graph  $G$  wird *vollständig* genannt, wenn seine Kantenmenge maximal ist, also jedes Knotenpaar durch eine Kante verbunden ist. Der vollständige Graph der Ordnung  $n$  wird mit  $K_n$  bezeichnet und besitzt  $\binom{n}{2} = (n^2 - n)/2$  Kanten. Ein bipartiter Graph  $G = (V_1 \cup V_2, E)$  heißt *vollständig*, wenn jeder Knoten  $v_1 \in V_1$  mit jedem Knoten  $v_2 \in V_2$  verbunden ist. Gilt zusätzlich  $|V_1| = m$  und  $|V_2| = n$ , so wird dieser mit  $K_{m,n}$  bezeichnet.

Eine *Clique* ist eine echte Teilmenge  $V' \subsetneq V$ , so dass der von  $V'$  induzierte Subgraph vollständig ist. Die Anzahl Knoten der größten Clique eines Graphen heißt *Cliquenzahl*  $Cl(G)$ . Dabei ist die Cliquenzahl kleiner oder gleich der Chromatischen Zahl, da zur Färbung eines vollständigen Graphen so viele Farben benötigt werden, wie er Knoten hat.

### 2.1.5 Matching

Ein *Matching*  $M \subseteq E$  von  $G = (V, E)$  bezeichnet eine Kantenmenge, so dass keine zwei Kanten  $e_1, e_2 \in M$  inzident sind. Das größtmögliche Matching eines Graphen gerader Ordnung  $n$  besteht aus  $|M| = n/2$  Kanten. Es wird *perfektes Matching* genannt. Nicht alle Graphen besitzen ein solches Matching. Es gibt jedoch zu jedem Graphen ein *maximales Matching*  $M^*$ , so dass für jedes Matching  $M$  des Graphen  $|M| \leq |M^*|$  gilt. Zwei Matchings

heißen *kantendisjunkt*, wenn die Kantenmengen disjunkt sind.

### 2.1.6 Knotenpartition

Eine *Knotenpartition*, kurz *Partition*  $P(G) = (P_1, P_2, \dots, P_k)$  eines Graphen  $G = (V, E)$  ist eine Aufteilung der Knoten aus  $V$  in  $k$  disjunkte nichtleere Teilmengen  $P_i, i = 1, 2, \dots, k$ , so dass gilt:  $\bigcup_{i=1}^k P_i = V$ . Dabei sind  $P_i$  die *Elemente* und  $k$  die *Größe* der Partition. Bei einer  $k$ -*Partition* ist die Größe der Partition auf  $k$  festgelegt. Eine 2-Partition  $P(G) = (P_1, P_2)$  wird auch *Bipartition* genannt. Eine *Knotensektion*, kurz *Sektion*  $S(G)$  eines Graphen  $G$  der Ordnung  $n$  ist eine Partition  $P(G) = (P_1, P_2, \dots, P_k)$  für deren Elemente  $P_i, i = 1, 2, \dots, k$ , entweder  $|P_i| = \lfloor n/k \rfloor$  oder  $|P_i| = \lceil n/k \rceil$  gilt. Die Größen der Elemente  $P_i$  unterscheiden sich demnach höchstens um 1. Die Sektion wird auch *balancierte Partition* genannt. Analog zu oben ist  $k$  die *Größe* der Sektion und eine  $k$ -*Sektion* eine Sektion mit gegebener Anzahl  $k$  von Elementen. Eine 2-Sektion bezeichnet man auch als *Bisektion*.

### 2.1.7 Isomorphismus

Zwei Graphen  $G_1 = (V_1, E_1)$  und  $G_2 = (V_2, E_2)$  sind *isomorph*, wenn eine bijektive Abbildung  $\pi : V_1 \rightarrow V_2$  existiert, so dass gilt:

$$\forall u, v \in V : (u, v) \in E_1 \Leftrightarrow (\pi(u), \pi(v)) \in E_2.$$

Das heißt, zwei isomorphe Graphen haben die gleiche Anzahl Knoten, welche auf die gleiche Weise miteinander verbunden sind. Die Abbildung  $\pi$  heißt dann *Isomorphismus*.

Ein *Automorphismus* eines Graphen  $G = (V, E)$  ist ein Isomorphismus des Graphen auf sich selbst. Die bijektive Abbildung  $\pi : V \rightarrow V$  ist also eine Permutation der Knotenmenge  $V$ , wobei die permutierten Knoten durch dieselben Kanten verbunden sind wie die ursprünglichen. Gibt es zu jedem Knotenpaar  $v_1, v_2 \in V$  eines Graphen  $G = (V, E)$  einen Automorphismus  $\pi$  mit  $\pi(v_1) = v_2$ , so bezeichnet man  $G$  als *knotentransitiv*. Es ist trivial, dass ein solcher Graph regulär ist.

## 2.2 ‚Gossiping‘ und ‚Broadcasting‘

*Kommunikationsnetzwerke* sind Graphen, deren Knoten Personen und deren Kanten eine Art der Kommunikation zwischen den Personen darstellen. Ist ein solches Netzwerk ungerichtet, so wird bei der Kommunikation stets Information in beide Richtungen übertragen, vergleichbar einem Gespräch. Hier

soll Kommunikation zu einem bestimmten Zeitpunkt immer nur zwischen zwei Personen möglich sein, so wie es bei Telefonaten der Fall ist. Ein Kommunikationsnetzwerk, das diese Bedingung erfüllt, wird als *1-Kanal-Netzwerk* bezeichnet, alle anderen sind *Mehr-Kanal-Netzwerke*. Zudem werden hier nur Netzwerke mit *Einheitskosten* betrachtet. Bei diesen dauert eine Kommunikation zwischen zwei Personen stets eine Zeiteinheit lang, unabhängig davon wie viel Information ausgetauscht wird.

In Kommunikationsnetzwerken bezeichnen ‚*Gossiping*‘<sup>1</sup> und ‚*Broadcasting*‘<sup>2</sup> zwei Arten der Informationsverbreitung. Beim ‚*Gossiping*‘ kennt jede Person des Netzwerkes eine Information, von der die anderen nichts wissen, und soll diese durch Kommunikation an alle anderen verbreiten. Beim ‚*Broadcasting*‘ hat nur eine Person eine Information, die sie an alle anderen weiterleiten soll. Ein sehr bekanntes Beispiel für das ‚*Gossiping*‘-Problem, von dem es seinen Namen hat, ist das Folgende: Mehrere Frauen wissen jeweils eine Klatschgeschichte, die alle übrigen Frauen nicht kennen. Es telefonieren je zwei Frauen miteinander und erzählen sich gegenseitig alle Geschichten, von denen sie zum Zeitpunkt des Gesprächs wissen. Die Frage ist nun, wie viele Telefonate nötig sind bis alle Frauen alle Klatschgeschichten kennen. Ein Beispiel für ‚*Broadcasting*‘ in einem gerichteten Mehr-Kanal-Netzwerk ist die Aussendung von Information über Radio oder Fernsehen. Hier wird Information von einer zentralen Stelle aus verbreitet, und zwar an alle Haushalte zugleich.

Sei nun  $G = (V, E)$  ein Kommunikationsnetzwerk mit der Knotenmenge  $V$  von Personen und der Kantenmenge  $E$ , der Kommunikation zwischen diesen Personen. Die Bezeichnung  $g(G)$  steht für die ‚*Gossiping*‘-Zeit, die für die Kommunikationen benötigte Zeit bis jede Person die Information jeder anderen Person kennt. Wie in Anhang A gezeigt, beweist W. Knödel für vollständige Graphen  $K_n$  (siehe [Knö74]):

$$g(K_n) = \begin{cases} \lceil \log_2 n \rceil & \text{für } n \text{ gerade,} \\ \lceil \log_2 n \rceil + 1 & \text{für } n \text{ ungerade.} \end{cases}$$

Es ist trivial, dass die ‚*Gossiping*‘-Zeit  $g(K_n)$  über allen Graphen der Ordnung  $n$  minimal ist. Jeder Graph  $G$  der Ordnung  $n$ , für den  $g(G) = g(K_n)$  gilt, wird schließlich ‚*Gossip*‘-Graph genannt. Diese Graphen benötigen für die Informationsverbreitung also genauso wenig Zeit, wie der vollständige Graph  $K_n$ . Ein *minimaler* ‚*Gossip*‘-Graph  $MGG_n$  der Ordnung  $n$  ist dann ein ‚*Gossip*‘-Graph mit minimaler Anzahl an Kanten. Die Anzahl Kanten dieses Graphen wird mit  $G(n)$  bezeichnet. Die ‚*Broadcasting*‘-Zeit  $b(v)$  einer

<sup>1</sup> ‚gossip‘ (engl.) heißt Altweibergeschätz, Klatsch, Trasch

<sup>2</sup> ‚broadcast‘ (engl.) heißt Aussendung, Ausstrahlung (Radio,TV)



Person  $v$  ist die Zeit, die für die Kommunikationen benötigt wird bis die Person ihre Information an jede andere Person des jeweiligen Graphen verbreitet hat. Die ‚Broadcasting‘-Zeit  $b(G)$  eines Graphen  $G$  ist damit definiert als  $b(G) = \max\{b(v) | v \in V\}$ . Anhand des Beweises in Anhang A ist die Gültigkeit von  $b(K_n) = \lceil \log_2 n \rceil$  für den vollständigen Graphen leicht zu sehen. Analog zum ‚Gossiping‘ werden folgende Begriffe für das ‚Broadcasting‘ definiert: Ein ‚Broadcast‘-Graph ist ein Graph, für den  $b(G) = b(K_n) = \lceil \log_2 n \rceil$  gilt. Ein *minimaler* ‚Broadcast‘-Graph  $MBG_n$  der Ordnung  $n$  bezeichnet einen ‚Broadcast‘-Graph mit minimaler Anzahl an Kanten und  $B(n)$  deren Kantenzahl.

## 2.3 Partitionierung und Clusterung

Sei nun  $G = (V, E)$  ein ungerichteter, gewichteter Graph mit Kantengewichtsfunktion  $w : E \rightarrow \mathbb{R}$ , und  $P(G)$  eine  $k$ -Partition des Graphen. Die Kanten von  $G$ , deren Endknoten in verschiedenen Elementen der Partition  $P$  liegen, heißen dann *Partitionskanten* und werden in  $E(P)$  zusammengefasst. Es gilt:

$$E(P) := \{e = (v_1, v_2) \in E \mid \exists i, j \text{ mit } 1 \leq i \neq j \leq k : v_1 \in P_i \wedge v_2 \in P_j\}.$$

Dann ist der *Schnitt*  $cut(P)$  der Partition  $P$  wie folgt definiert:

$$cut(P) := \sum_{e \in E(P)} w(e).$$

Der Wert  $cut(P)$  ist somit die Summe der Gewichte von Kanten, die die Partition kreuzen. Sind alle Kantengewichte 1 so entspricht  $cut(P)$  der *Größe* eines Schnittes, welche die Anzahl an Partitionskanten  $|E(P)|$  bezeichnet. Eine *maximale Partition* ist eine Partition, deren Schnitt maximal, eine *minimale Partition* eine, deren Schnitt minimal ist, und zwar unter den Schnitten aller möglichen Partitionen. Letztere kann zum Beispiel durch wiederholte Anwendung des Algorithmus von M. Stoer und F. Wagner (siehe [SW97]) ermittelt werden. Dieser bestimmt eine minimale 2-Partition. Unter der *minimalen* bzw. *maximalen Partitionierung* eines Graphen  $G = (V, E)$  versteht man dann die Aufgabe, eine minimale bzw. maximale Partition des Graphen  $G$  zu finden. Möchte man klar machen, dass eine Partition einer bestimmten Größe  $k$  gesucht wird, so verwendet man auch die Bezeichnungen *k-Partitionierung* und *Bipartitionierung*, falls  $k = 2$ . Für die Suche nach minimalen oder maximalen Sektionen sind die entsprechenden *Sektionierungen*, *k-Sektionierungen*, wie auch *Bisektionierungen* analog definiert..

Unter der *Clusteranalyse*<sup>3</sup> versteht man verschiedene Arten der Datenanalyse zur Ermittlung von Gruppen (so genannte *Cluster*) zusammengehöriger Objekte aus einer Grundmenge numerisch beschriebener Objekte. In der Graphentheorie wird je nach Definition der „Zusammengehörigkeit“ eine *Kostenfunktion* über den Partitionen festgelegt. Das Ergebnis der *Clusterung* eines Graphen ist dann eine  $k$ -Partition mit  $k$  beliebig und minimalen Kosten unter allen möglichen Partitionen. Cluster können beispielsweise Anhäufungen von Knoten sein, deren Abstand zu Knoten des gleichen Clusters geringer ist als der zu Nachbarn anderer Cluster. Die Kostenfunktion ist hier die Summe der Abstände aller Knoten, die einem Partitionselement angehören.

---

<sup>3</sup> ‚cluster‘ (*engl.*) heißt Anhäufung

# Kapitel 3

## Kontaktgraphen

Der erste Schritt zur Entwicklung einer Gruppeneinteilungsstrategie ist die Wahl eines Kommunikationsnetzwerkes, das die flüchtigen Kontakte in Form von kurzen Gesprächen zwischen den teilnehmenden Personen darstellt.

Nach Betrachtung eines in der Praxis schon einmal getesteten Netzwerkes, werden verschiedene Kriterien erarbeitet, die ein so genannter *Kontaktgraph* erfüllen sollte. Deshalb werden unterschiedliche Graphen untersucht, von denen ein Graph ausgewählt wird, der zur Verwendung als Kontaktgraph sinnvoll erscheint.

Im August 2003 fand an der Universität in Tilburg (Niederlande) die dritte POLNET Summer School<sup>1</sup> statt. Hier wurde eine erste einfache Strategie zur Einteilung der Teilnehmer in Arbeitsgruppen angewendet. Es diente hier ein *k-zyklischer Graph* als Kontaktgraph.

**Definition 3.1** *Ein zusammenhängender Graph  $G = (V, E)$  heißt k-zyklisch, wenn eine Partition  $V_0, V_1, \dots, V_{k-1}$  von  $G$  existiert, so dass für jede Kante  $(u, v) \in E$  gilt:*

$$\exists i \in \{0, 1, \dots, k-1\} : u \in V_i \wedge v \in V_{(i+1) \bmod k}.$$

Die an der Veranstaltung teilnehmenden Personen wurden in  $k$  Mengen  $V_0, V_1, \dots, V_{k-1}$  eingeteilt, und die Personen jeder Menge  $V_i$  führten Gespräche mit den Personen der zwei „Nachbar“-Mengen  $V_{(i-1) \bmod k}$  und  $V_{(i+1) \bmod k}$ . Die zugrunde liegende Idee hinter der Wahl des *k-zyklischen Graphen* ist die folgende: Zur Einteilung der Gruppen wird im Sympathiegraphen nach *möglichst zusammenhängenden Pseudowegen* der Länge  $\geq k$  gesucht. Der Begriff „möglichst zusammenhängender Pseudoweg“ steht für

---

<sup>1</sup> International Summer School on the Analysis of Political and Managerial Networks, siehe <http://www.polnet-school.info>

einen Weg, der auch aus disjunkten Teilwegen bestehen kann. Dabei soll die Kantenzahl aller Pseudowege, die je eine der gesuchten Personengruppe bilden, zusammen maximal sein. Zudem soll jeder der Pseudowege mindestens einen Knoten jeder Menge  $V_i$  enthalten. Entspricht die Länge eines Pseudoweges der Partitionsgröße des zugrunde liegenden zyklischen Graphen, so wird dieser auch *Pseudokreis* genannt. Um auf die beschriebene Weise  $n = k \cdot g$  Personen in  $g$  gleich große Gruppen einzuteilen, wird  $k = n/g$  gewählt, also auf die Größe der gesuchten Gruppen gesetzt, und der  $k$ -zyklische Graph auf einer  $k$ -Sektion gebildet. So ist der  $k$ -zyklische Graph zusätzlich  $(2 \cdot g)$ -regulär. Ist der Quotient  $(n/g)$  nicht ganzzahlig, so wird  $k = \lfloor n/g \rfloor$  gewählt. Dann haben  $(n \bmod g)$  Elemente der  $k$ -Sektion Größe  $(g + 1)$ , die übrigen Größe  $g$ . In diesem Graphen werden schließlich  $g_1 := (n \bmod g)$  Pseudowege der Länge  $\lceil n/g \rceil$  und  $g_2 := g - g_1$  Pseudokreise gesucht. Ist  $n$  durch  $g$  teilbar, so sind  $g$  Pseudokreise der Länge  $n/g$  zu finden.

Ein weiterer Grund für die Wahl des  $k$ -zyklischen Graphen als Kontaktgraph ist die günstige Gesprächsführung. Ist die Anzahl  $k$  der Partitionselemente  $V_i$  des Kontaktgraphen gerade, so können in einer ersten Runde die Gespräche zwischen den „Nachbar“-Mengen  $V_i$  und  $V_{(i+1) \bmod k}$ , sowie in einer zweiten Runde die Gespräche zwischen  $V_i$  und  $V_{(i-1) \bmod k}$  für alle  $i \in \{0, 2, \dots, k-2\}$  geführt werden. Ist  $k$  allerdings ungerade, so muß je eine der zwei benachbarten Mengen  $V_{k-1}$  und  $V_{k-2}$  in den zwei eben beschriebenen Runden der Gesprächsführung aussetzen. In einer weiteren Runde müssen schließlich die Gespräche zwischen den Personen dieser beiden Mengen nachgeholt werden, in der die Personen aller übrigen Mengen, also unter Umständen sehr viele Personen aussetzen müssen. Die Anzahl, wie oft die teilnehmenden Personen bei einer Gesprächsführung im Kontaktgraphen  $G$  zusammen aussetzen müssen, berechnet sich wie folgt:

$$|V_{k-1}| \cdot \max_{i \in \{0, \dots, k-2\}} |V_i| \quad (3.1)$$

$$+ |V_{k-2}| \cdot \max_{i \in \{0, \dots, k-3, k-1\}} |V_i| \quad (3.2)$$

$$+ (|V_0| + \dots + |V_{k-3}|) \cdot \max_{i=k-1, k-2} |V_i| \quad (3.3)$$

$$+ (g - 1 - (n \bmod g)) \cdot 2 \cdot \min_{i=0, \dots, k-1} |V_i| + 2 \cdot \max_{i=0, \dots, k-1} |V_i|. \quad (3.4)$$

Dabei geben die Summanden 3.1, 3.2 bzw. 3.3 die durch ungerades  $k$  verursachten Aussetzer in Runde 1, Runde 2 bzw. Runde 3 an, und der Summand 3.4 umfasst alle übrigen Aussetzer aufgrund ungleicher Größen der Partitionselemente. Letzterer ist für den Fall angegeben, dass alle Partitionselemente einer Größe „benachbart“ sind. Alternativ wählt man  $k = \lceil n/g \rceil$ , wenn  $k = \lfloor n/g \rfloor$  ungerade ist und sucht  $g_1 := (n \bmod k)$  Pseudowege der

Länge  $\lfloor n/g \rfloor$  und  $g_2 := g - g_1$  Pseudokreisen. Die Anzahl Aussetzer umfasst bei geradem  $k$  nur den oben gegebenen Summanden 3.4, wobei wieder die Reihenfolge der Partitionselemente festgelegt ist:

$$(g - 1 - (n \bmod g)) \cdot 2 \cdot \min_{i=0, \dots, k-1} |V_i| + 2 \cdot \max_{i=0, \dots, k-1} |V_i|.$$

Somit verringert sich die Anzahl Aussetzer durch eine bessere Wahl von  $k$ . Jedoch sind hier  $g_1$  Pseudowege zu suchen, deren Länge um 1 kleiner als die Anzahl Partitionselemente des Graphen ist. Zu jedem solchen Pseudoweg existiert also ein  $V_i$ , aus dem kein Knoten Teil des Pseudoweges ist, und somit keine Verbindung zwischen den zwei Knoten aus  $V_{(i-1) \bmod k}$  und  $V_{(i+1) \bmod k}$  bestehen kann.

Weiterer Nachteil dieses Verfahrens ist, dass die Anzahl Gespräche, die eine Person aus der Menge  $V_i$  führt auf maximal  $|V_{(i-1) \bmod k}| + |V_{(i+1) \bmod k}|$  festgelegt ist. Natürlich können weitere Gesprächsrunden abgehalten werden. Es ist jedoch dabei darauf zu achten, dass möglichst wenige Personen aussetzen und kein Personenpaar doppelt miteinander spricht. Eine Regelung für eine solche Runde kann demnach nicht allgemein für alle Personen- und Gruppengrößen angegeben werden.

Zudem haben in diesem Kontaktgraphen die Personen eines Partitionselementes alle dieselben Gesprächspartner, und bei Personen der zwei Mengen  $V_i$  und  $V_{(i+2) \bmod k}$  stimmen je die Hälfte der Gesprächspartner überein. Dies kann auf die Teilnehmer allerdings so wirken, als würde ihnen die Bekanntschaft der Personen bestimmter Partitionselemente vorenthalten. Außerdem verursacht dies, dass im Sympathiegraphen ein starker Zusammenhang zwischen benachbarten Partitionselementen des Graphen besteht, was jedoch bei dem oben beschriebenen Verfahren zur Einteilung der Gruppen nicht genutzt wird. Stattdessen findet jeder Teilnehmer nur höchstens drei seiner Gesprächspartner in seiner Gruppe wieder, die er unter Umständen noch nicht einmal sympathisch findet.

Durch die vielen Bedingungen, die der  $k$ -zyklische Graph erfüllen muß, um eine günstige Gesprächsführung auszudrücken, ist dieser Graph recht unflexibel als Kontaktgraph. Zudem hängt er von der Anzahl einzuteilender Gruppen ab. Es stellt sich daher die Frage, welche Graphen als Kontaktgraphen besser geeignet sind und welche graphentheoretischen Kriterien diese erfüllen sollten.

### 3.1 Anforderungen an den Kontaktgraphen

Anhand der Beobachtungen am  $k$ -zyklischen Graphen können Kriterien festgelegt werden, die die Qualität eines Graphen als Kontaktgraphen ausma-

chen. Dabei werden neben den gefundenen Vor- und Nachteilen des  $k$ -zyklischen Graphen noch weitere Überlegungen angestellt, wie der Kontaktgraph aussehen sollte, damit er eine unter verschiedenen Gesichtspunkten gute Gesprächsführung ausdrückt.

Der günstigste Graph für die Entstehung von Sym- und Antipathien zwischen allen Teilnehmern, ist der vollständige Graph. Denn wenn eine Person mit jeder anderen Person Kontakt aufnimmt, kann sie zwischen allen entscheiden, mit wem sie am liebsten in einer Gruppe zusammenarbeiten möchte.

Bei einer geraden Anzahl  $n$  von Personen besteht außerdem die Möglichkeit in  $(n - 1)$  Runden jeden mit jedem sprechen zu lassen. Stelle man sich dazu den vollständigen Graphen  $K_{n-1}$  vor, dessen Ecken je einen Teilnehmer darstellen und wie bei einem regelmäßigen  $(n - 1)$ -Eck angeordnet sind. Markiert man in diesem eine Seite und alle dazu parallelen Seiten, so bleibt ein Knoten ohne adjazente markierte Kante übrig. Dieser wird mit dem nicht zum  $K_{n-1}$  gehörenden Knoten verbunden und die entsprechende Kante ebenfalls markiert. Die markierten Kanten bilden nun eine Gesprächsrunde, an der jede teilnehmende Person beteiligt ist. Fährt man auf die gleiche Art und Weise fort und beginnt jeweils mit einer noch nicht markierten Kante des  $(n - 1)$ -Eck, so erhält man schließlich  $(n - 1)$  Gesprächsrunden, in denen alle zu führenden Gespräche geführt werden. Diese Gesprächsführung ist in Abbildung 3.1 anhand des  $K_6$  als Kontaktgraph gezeigt.

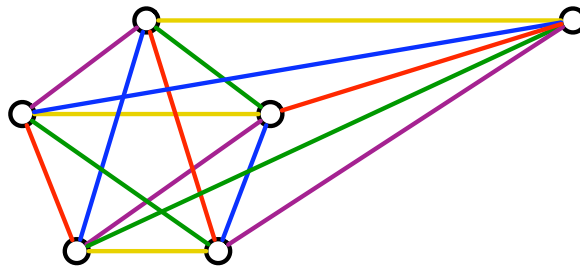


Abbildung 3.1:  $K_6$  als Kontaktgraph – die Kanten einer Farbe bilden die Gespräche einer Runde.

Im entstehenden Graphen mit den Sympathien als Knotenmenge könnte man dann nach möglichst zusammenhängenden Knotenmengen suchen. Denn man vermutet, dass Personengruppen, bei denen zwischen möglichst vielen Personen Sympathien vorliegen, am besten zusammenarbeiten. Lässt man zudem die Aussprache von Antipathien zu, so ist zusätzliches Ziel der

Gruppeneinteilung, möglichst viele Knoten, die durch eine Antipathiekante miteinander verbunden sind, voneinander zu trennen. Denn Personen, die sich nicht mögen, können die Teamfähigkeit der Gruppe verringern.

### 3.1.1 Dreiecke

Die Methode, jeden mit jedem reden zu lassen, ist allerdings sehr zeitaufwendig und anstrengend für die Teilnehmer. Außerdem ist es bei kurzer Kontaktaufnahme mit zu vielen Personen schwierig, sich eine Meinung zu bilden mit wem man lieber in einer Gruppe zusammenarbeiten möchte. Daher sollte man bei zwischenmenschlicher Sympathie darauf bauen, dass Sympathieträger von Personen, die man sympathisch findet, einem selber ebenfalls sympathisch sind. Es soll also angenommen werden, dass Sympathie eine transitive Relation zwischen Personen ist. Diese Spekulation soll dem Teilnehmer allerdings verborgen bleiben. Aus diesem Grund wird gefordert, dass der Kontaktgraph keine Dreiecke enthalten darf. Ein Dreieck in der Gesprächsführung bedeutet nämlich, dass sich drei Personen, namens  $A$ ,  $B$  und  $C$  je miteinander unterhalten. Man nehme an, dass  $A$  und  $B$ , sowie  $B$  und  $C$  sich sympathisch sind, und schließt daraus, dass sich auch  $A$  und  $C$  sympathisch finden. Ist letzteres nun aber nicht der Fall und sind die drei Personen trotzdem in derselben Gruppe, so können sie leicht erkennen, dass hier mit „Sympathie-Transitivität“ gearbeitet wurde. Die Gefahr, dass in den anhand der Sympathien gebildeten Gruppen Antipathien auftreten, besteht zwar bei jeder Art der Gesprächsführung. Allerdings ist der Schluß auf Sympathie aufgrund der Transitivität dieser Relation in einem Kontaktgraphen, der nur Vierecke oder größere Kreise enthält, für den Teilnehmer weniger gut durchschaubar. Angenommen man spekuliert bei einem Weg aus Sympathien, der mindestens Länge 3 hat, dass sich auch die zu den beiden Endknoten des Weges gehörenden Personen sympathisch finden, dann ist dies für die beiden Personen nur schwer erkennbar, da sie nicht unbedingt wissen, mit wem sich ihre Gesprächspartner unterhalten haben.

### 3.1.2 Gesprächsführung

Die Untersuchungen des  $k$ -zyklischen, wie auch des vollständigen Graphen umfassten unter anderem die Berechnung der Anzahl Aussetzer bei der Gesprächsführung. Der Kontaktgraph soll eine Ausführung der Gespräche ermöglichen, bei der möglichst wenige Personen aussetzen müssen. Es sollen also möglichst wenig Gesprächsrunden nötig sein, in denen jeweils möglichst viele Gespräche parallel durchführbar sind. Das heißt, dass in  $r$  Gesprächsrunden möglichst jede Person mit  $r$  verschiedenen Partnern spricht. Graphen-

theoretisch ausgedrückt bedeutet dies, dass der Graph  $r$ -regulär sein und  $r$  kantendisjunkte perfekte Matchings enthalten soll.

### 3.1.3 Streuung

Verwendet man einen  $k$ -zyklischen Graphen als Kontaktgraph, so haben viele Teilnehmer dieselben Gesprächspartner. Stellen die Teilnehmer dies fest, fühlen sie sich ungerecht behandelt, da ihnen die Bekanntschaft der Personen bestimmter Partitionselemente vorenthalten wurde. Natürlich gibt es abgesehen vom vollständigen Graphen in jedem Graphen eine Menge von Knoten zu denen ein bestimmter Knoten nicht adjazent ist. Jedoch ist die schlechte Verteilung der Gesprächspartner auf alle Teilnehmer der Gesprächsführung sehr gut ersichtlich. Deshalb sollte der Kontaktgraph realisieren, dass je zwei Personen mit möglichst vielen verschiedenen Personen sprechen. Dieses Kriterium wird als „große Streuung“ im Kontaktgraphen bezeichnet. Graphentheoretisch ausgedrückt bedeutet große Streuung, dass die Nachbarschaft jeglicher zweier Knoten des Graphen maximal ist. Als Maß für die Streuung betrachtet man nun die maximale Anzahl an gemeinsamen Nachbarn zweier beliebiger Knoten geteilt durch den maximalen Grad über beide Knoten. Je näher dieser Wert bei 0 liegt, um so größer ist die Streuung. Die kleinstmögliche Streuung hat ein  $r$ -regulärer Graph, wenn zwei seiner Knoten  $r$  gemeinsame Nachbarn haben. Die Streuung beträgt dann 1. Dies gilt es beim Kontaktgraphen zu verhindern.

### 3.1.4 Flexibilität

Die Anzahl Personen  $n$ , die Zahl zu bildender Gruppen  $g$ , sowie auch die gewünschte Anzahl  $r$  flüchtiger Kontakte pro Person kann bei Veranstaltungen wie der POLNET, jede beliebige Größe annehmen. Ziel bei der Entwicklung einer Strategie zur Gruppeneinteilung ist es, sie in Bezug auf diese Werte so flexibel wie möglich zu halten. Es sollen so gut wie keine Einschränkungen nötig sein. Der Kontaktgraph soll somit unabhängig von der Anzahl Personen eine Gesprächsführung ausdrücken, deren Anzahl Aussetzer über allen Graphen möglichst klein ist. Zudem soll der Kontaktgraph so konzipiert sein, dass er nicht von der Anzahl zu bildender Gruppen abhängt. Dies ist zum Beispiel bei der Verwendung des  $k$ -zyklischen Graphen notwendig, wo für das Verfahren der Gruppeneinteilung  $k = \lfloor n/g \rfloor$  gefordert wird. Außerdem ist unabhängig von der Anzahl zu führender Gespräche die Erfüllung obiger Anforderungen unerlässlich.



## 3.2 Wahl eines Kontaktgraphen

In Anbetracht oben genannter Kriterien, die ein Graph zur Verwendung als Kontaktgraph erfüllen sollte, werden nun verschiedene Graphklassen untersucht. Deren Reihenfolge entspricht weitgehend der Reihenfolge, die zur Entdeckung des für obige Zwecke günstigsten Graphen durch die Autorin führte. Manche Idee zur Betrachtung einer Graphenklasse entstammt den Kriterien der Graphen der vorherigen Abschnitte.

### 3.2.1 Bipartite Graphen

Betrachtet man den  $k$ -zyklischen Graphen und setzt voraus, dass  $k$  gerade ist, so stellt man fest, dass dieser bipartit ist. Da die vielen Anforderungen an die Personen- und Gruppengröße zur Bildung dieses speziellen bipartiten Graphen daher rühren, dass der Graph zusätzlich  $k$ -zyklisch ist, wird nun der bipartite Graph im Allgemeinen betrachtet.

Stelle also der bipartite Graph  $G = (V_1 \cup V_2, E)$  mit den zwei disjunkten Knotenmengen  $V_1$  und  $V_2$  den Kontaktgraphen dar. Das bedeutet für die Gesprächsführung, dass die Menge der Teilnehmer in zwei Gruppen geteilt wird und die Personen je nur mit Personen der anderen Gruppe sprechen. Damit jeder Teilnehmer gleich viele Gesprächspartner hat, wählt man den Graphen zusätzlich  $r$ -regulär, wobei  $r > 0$  die gewünschte Anzahl an Gesprächen ist. Daraus ergibt sich dann  $r \cdot |V_1| = |E| = r \cdot |V_2|$ , also  $|V_1| = |V_2|$ . Für die  $r$ -Regularität eines bipartiten Graphen müssen demnach die Mengen  $V_1$  und  $V_2$  gleich groß gewählt werden, also die Anzahl Personen gerade sein. Dann gelten folgende zwei Sätze (Satz 3.1 und 3.3) von D. König (siehe [Vol91], Seite 92):

**Satz 3.1 (König)** *Ein  $r$ -regulärer, bipartiter Graph  $G = (V_1 \cup V_2, E)$  besitzt ein perfektes Matching.*

**Beweis:** Ist  $S \subseteq V_1$  und ist  $e \in E$  eine Kante, die inzident zu einem Knoten aus  $S$  ist, so ist  $e$  auch inzident zu einem Knoten aus  $N(S) \subseteq V_2$ . Ist also  $E_1 \subseteq E$  die Kantenmenge, deren Elemente zu den Knoten aus  $S$ , und  $E_2 \subseteq E$  die Kantenmenge, deren Elemente zu den Knoten aus  $N(S)$  inzident sind, so folgt:  $E_1 \subseteq E_2$ . Somit gilt:

$$r \cdot |S| = |E_1| \leq |E_2| = r \cdot |N(S)|,$$

woraus  $|S| \leq |N(S)|$  für alle  $S \subseteq V_1$  folgt. Nach dem folgenden Satz 3.2 existiert dann ein Matching  $M \subseteq E$ , deren Elemente zu allen Knoten aus  $V_1$

inzident sind. Dieses muss dann wegen  $|V_1| = |V_2|$  perfekt sein.  $\square$

Bevor der zweite Satz von König folgt, wird der oben verwendete Satz von P. Hall notiert:

**Satz 3.2 (Hall's Theorem)** *Es sei  $G = (A \cup B, E)$  ein bipartiter Graph mit der Knotenmenge  $V := A \cup B$ . Es gibt genau dann ein Matching  $M$  von  $G$  mit  $(V|M) \cap A = A$ , wenn für alle  $S \subseteq A$  gilt:*

$$|S| \leq |N(S)|.$$

Über die Jahre sind viele Beweise von Hall's Theorem entstanden. Hier wird ein Beweis von P. R. Halmos und H. E. Vaughn präsentiert (siehe [LP86], Seite 5):

**Beweis:** Sei  $M = \{m_1, \dots, m_p\} \subseteq E$  ein Matching von  $G$  mit  $(V|M) \cap A = A$ . Ist  $m_i = (a_i, b_i)$  mit  $a_i \in A$ , so gilt notwendig  $b_i \in B$  für jedes  $i = 1, 2, \dots, p$ , und die Knoten  $b_1, b_2, \dots, b_p$  sind paarweise verschieden. Daraus folgt für alle  $S \subseteq A$  mit  $S = \{a_{j_1}, a_{j_2}, \dots, a_{j_q}\}$

$$|S| = q = |\{b_{j_1}, b_{j_2}, \dots, b_{j_q}\}| \leq |N(S)|.$$

Gelte nun umgekehrt  $|S| \leq |N(S)|$  für alle  $S \subseteq A$ . Wie im Folgenden beschrieben, führt man Induktion über  $|A|$ : Der Induktionsanfang mit  $|A| = 0$  oder  $|A| = 1$  ist klar. Für den Induktionsschritt betrachtet man die zwei Fälle:

1. für alle  $S \subset A, S \neq \emptyset$ :  $|S| < |N(S)|$ ,
2. es gibt ein  $A^* \subset A, A^* \neq \emptyset$ :  $|A^*| = |N(A^*)|$  und für alle übrigen  $S \subset A, S \neq \emptyset$ :  $|S| < |N(S)|$ .

Man betrachte Fall 1: Seien  $a$  und  $b$  adjazente Knoten mit  $a \in A$ . Bezeichne  $G' := G - \{a, b\}$  und  $S$  eine beliebige Teilmenge von  $A - \{a\}$ . Gilt  $S = \emptyset$ , so folgt  $|S| = 0 = |N(S)|$ . Man nehme also  $S \neq \emptyset$  an. Wegen  $S \neq A$ , gilt nach Voraussetzung  $|S| < |N(S)|$  und somit  $|N_{G'}(S)| \geq |N_G(S)| - 1 \geq |S|$ , wobei  $N_{G'}(S)$  die Nachbarschaft der Knoten aus  $S$  im Graphen  $G'$  beschreibt. Aufgrund der Induktionsannahme gibt es dann ein Matching  $M'$  in  $G'$ , für welches  $(V|M') \cap (A - \{a\}) = (A - \{a\})$  gilt. Setzt man dann  $M := M' \cup \{(a, b)\}$ , so erfüllt diese  $(V|M) \cap A = A$ .

Gelte in Fall 2 nun  $|A^*| = |N(A^*)|$  für ein  $A^* \subset A$  mit  $A^* \neq \emptyset$ . Dann spaltet man  $G$ , wie folgt, in zwei kleinere Subgraphen: Sei  $G_1$  der von  $A^* \cup N(A^*)$  induzierte Subgraph und  $G_2 := G - (A^* \cup N(A^*))$ . Dann zeigt man, dass

$G_1$ , sowie  $G_2$  die Induktionsannahme erfüllen. In  $G_1$  sein für den Beweis  $S \subseteq A^*$ . Dann gilt  $N_G(S) \subseteq N_G(A^*)$ , außerdem  $N_{G_1}(S) = N_G(S)$  und folglich  $|N_{G_1}(S)| = |N_G(S)| \geq |S|$  nach Voraussetzung. Für  $G_2$  sei nun  $S \subseteq A - A^*$ . Dann gilt  $N_G(S \cup A^*) = N_{G_2}(S) \cup N_G(A^*)$ , weil  $N_{G_2}(S) \cap N_G(A^*) = \emptyset$ . Und daraus folgt  $|N_{G_2}(X)| = |N_G(S \cup A^*)| - |N_G(A^*)| \geq |S \cup A^*| - |N_G(A^*)| = |S \cup A^*| - |A^*| = |S|$  aufgrund der Voraussetzung und  $S \cap A^* = \emptyset$ . Nach Anwendung der Induktionsvoraussetzung auf  $G_1$  und  $G_2$  existieren somit Matchings  $M_1$  und  $M_2$  mit  $(V|M_1) \cap A^* = A$  und  $(V|M_2) \cap (A - A^*) = (A - A^*)$ . Für  $M_1$  gilt zusätzlich  $(V|M_1) \cap N_G(A^*) = N_G(A^*)$ . Das Matching  $M = M_1 \cup M_2$  erfüllt dann  $(V|M) \cap A = A$ .  $\square$

**Satz 3.3 (König)** *Ein  $r$ -regulärer, bipartiter Graph  $G = (V_1 \cup V_2, E)$  besitzt  $r$  kantendisjunkte perfekte Matchings.*

**Beweis:** Nach Satz 3.1 besitzt  $G$  ein perfektes Matching  $M$ . Dann sei  $G' := (V_1 \cup V_2, E')$  mit  $E' := E - M$  der Graph, der aus  $G$  durch Löschen aller Kanten aus  $M$  entsteht.  $G'$  ist  $(r - 1)$ -regulär und ebenfalls bipartit. Folgt man rekursiv weiter mit Satz 3.1, so ist die Existenz von  $r$  kantendisjunkten perfekten Matchings gezeigt.  $\square$

Werden also die Kontakte zwischen einer geraden Anzahl von Personen durch einen  $r$ -regulären, bipartiten Graphen ausgedrückt, so ist es in einer Gesprächsführung mit  $r$  Runden möglich, dass jede teilnehmende Person mit  $r$  verschiedenen Partnern spricht. Als Beispiel für eine solche Gesprächsführung betrachte man den Graphen  $G = (V, E)$  mit  $V = \{0, 1, \dots, n\}$ ,  $V_1 \subset V$  die Menge der geraden Zahlen und  $V_2 \subset V$  die der ungeraden. Sei dabei die Personenzahl  $n = |V|$  gerade und die Anzahl Gespräche  $r < n/2$ . Nun ist jeder Knoten  $2 \cdot i \in V_1$  adjazent zu  $2 \cdot i + 2 \cdot j + 1 \in V_2$  mit  $i \in \{0, 1, \dots, n/2 - 1\}$  und  $j \in \{0, 1, \dots, r - 1\}$ . Dabei sind die  $(n/2)$  Gespräche zu festem  $j$  gleichzeitig führbar, da ihre Kanten jeweils ein perfektes Matching in  $G$  bilden. Dabei legt  $(j + 1)$  die Runde fest, in der die jeweiligen Gespräche geführt werden. Da im Fall von ungerader Personenzahl kein  $r$ -regulärer, bipartiter Graph existiert, wird einfach ein ‚Dummy‘-Knoten hinzugefügt. Dann bedeutet Adjazenz zum ‚Dummy‘-Knoten, dass die Person in der jeweiligen Runde aussetzt. Somit sind bei einem bipartiten Kontaktgraphen bei ungerader Personenzahl insgesamt  $r$  Aussetzer notwendig.

Durch die Bipartitheit ist in diesem Kontaktgraphen die Anzahl führbarer Gespräche beschränkt. Die obere Schranke liegt jedoch bei  $\lceil n/2 \rceil$  und bietet im Vergleich zum  $k$ -zyklischen Kontaktgraphen einen größeren Spielraum. Betrachtet man nun zwei beliebige Knoten, einen aus der Menge  $V_1$  und den anderen aus  $V_2$ , so ist deren gemeinsame Nachbarschaft mit der Größe

$(2 \cdot r)$ , die größtmögliche unter allen Nachbarschaften zweier Knoten in  $r$ -regulären Graphen. Bei Personen, die derselben Knotenmenge  $V_i$ ,  $i = 1, 2$ , angehören, hängt die Größe der Streuung von der Art des bipartiten Graphen ab. In oben beschriebener Gesprächsführung haben je zwei Knoten  $x, y \in V$ , für die  $x - y \equiv 2 \cdot d \pmod{n}$  mit  $1 \leq d \leq r$  gilt, beispielsweise  $(r - d)$  viele gemeinsame Gesprächspartner. Alle übrigen Knotenpaare haben keine gemeinsame Nachbarschaft. Die Streuung ist demnach mit  $(r - 1)/r$  kleiner als die des  $k$ -zyklischen Graphen, dessen Streuung 1 beträgt. Es sei hier der Vollständigkeit halber angemerkt, dass der hier betrachtete Graph isomorph zur Klasse der später definierten bipartiten ‚Circulant‘-Graphen ist (siehe Abschnitt 3.2.4).

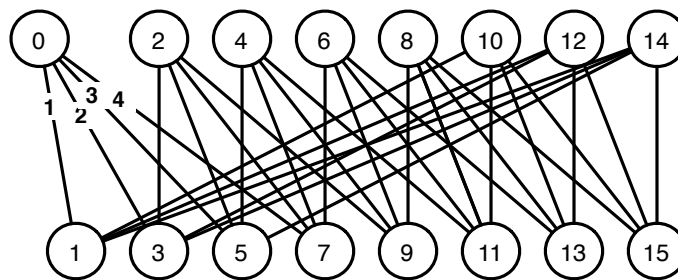


Abbildung 3.2: Bipartiter Graph der Ordnung 16 – Die Kanten-Etikette bezeichnet die Runden.

Aber auch in Anbetracht des Verbotes von Dreiecken ist der bipartite Graph eine gute Wahl als Kontaktgraph. Er enthält nur Kreise gerader Länge, somit also keine Dreiecke. Dies ist unabhängig von der Anzahl an Knoten und Kanten der Fall. Daher besteht selbst bei nachträglicher Erhöhung der Gesprächszahl auf höchstens  $\lceil n/2 \rceil$  keine Gefahr, dass im Kontaktgraphen Dreiecke entstehen.

Bipartite Graphen zeichnen sich also durch eine hohe Flexibilität aus. Sie stellen für jede Zahl von Teilnehmern und Gesprächen eine Gesprächsführung ohne Dreiecke und mit höchstens  $r$  Aussetzern dar. Die Größe der Streuung variiert. Außerdem ist dieser Graph unabhängig von der Anzahl zu bildender Gruppen.

### 3.2.2 Minimale $(r, t)$ -,Cages‘

Mit dem  $(r, t)$ -,Cage‘<sup>2</sup> wird ein ebenfalls  $r$ -regulärer, jedoch nicht unbedingt bipartiter Graph in Betracht gezogen. Er besitzt eine vorgegebene Tailleweite, die Länge seines kürzesten nicht trivialen Kreises.

**Definition 3.2** Ein  $(r, t)$ -,Cage‘ ist ein  $r$ -regulärer Graph  $G = (V, E)$  mit Tailleweite  $T(G) = t$ . Er heißt *minimal*, wenn er unter allen  $(r, t)$ -,Cages‘ minimale Anzahl  $n^*(r, t) := |V|$  von Knoten hat.

Als Beispiele für  $(r, t)$ -,Cages‘ seien der  $(2, t)$ -,Cage‘  $C_t$ , welcher als Kreisgraph der Ordnung  $t$  definiert wurde, und der  $(r, 3)$ -,Cage‘  $K_{r+1}$ , der vollständige Graph der Ordnung  $r+1$  genannt. Abbildung B.1 im Anhang zeigt weitere  $(r, t)$ -,Cages‘, welche sogar minimal sind. Es ist leicht zu sehen, dass diese Graphen die vorgegebene Regularität und Tailleweite haben. Doch existiert zu je zwei natürlichen Zahlen  $r \geq 2$ ,  $t \geq 2$  überhaupt ein  $(r, t)$ -,Cage‘? Die Forderung, dass im Graphen jeder Knoten Grad  $r$  haben soll, bedeutet anschaulich gesprochen, dass er „recht dicht“ ist. Will man zudem, dass jeder Kreis Länge  $\geq t$  hat, so bedeutet diese Forderung, dass der Graph nicht „zu dicht“ sein darf. Dass obige Frage trotzdem bejaht werden kann, zeigt ein Satz von H. Sachs (siehe [WV74], Seite 105 f). Dieser Beweis ermöglicht zwar eine effektive Konstruktion von  $(r, t)$ -,Cages‘, wird hier jedoch ausgelassen, da im Folgenden speziell die minimalen Graphen dieser Klasse betrachtet werden, für die diese allgemeine Konstruktion nicht gilt.

Für den Kontaktgraphen ist nun ein  $(r, t)$ -,Cage‘ mit  $r$  die Anzahl Gespräche und  $t > 3$  nötig, da dieser keine Dreiecke enthält. Und je größer die Tailleweite ist, um so weniger leicht durchschaubar für den Teilnehmer wird die Annahme von Transitivität in Bezug auf die Sympathie. Dabei stellt sich jedoch die Frage, welche Tailleweite bei einer festen Anzahl von Knoten und Gesprächen möglich ist. Der folgende Satz von P. Erdős und H. Sachs (siehe [WV74], Seite 116) liefert eine allgemeine untere Schranke für die Knotenzahl  $n(r, t)$  eines  $(r, t)$ -,Cages‘.

**Satz 3.4** Es sei  $G = (V, E)$  ein  $(r, t)$ -,Cage‘ mit  $r \geq 3$  und  $t \geq 2$ . Für die Anzahl Knoten  $n(r, t)$  von  $G$  gilt:

$$n(r, t) \geq n^*(r, t) \geq \begin{cases} 1 + \frac{r}{r-2} \cdot ((r-1)^{\lfloor (t-1)/2 \rfloor} - 1) & \text{für ungerades } t, \\ 2 \cdot \left(1 + \frac{r-1}{r-2} \cdot ((r-1)^{\lfloor (t-2)/2 \rfloor} - 1)\right) & \text{für gerades } t. \end{cases}$$

Gleichheit tritt ein, wenn gilt:  $r = 3 \wedge t \in \{3, 4, 5, 6, 8\}$ .

---

<sup>2</sup> ‚cage‘ (engl.) heißt Käfig

**Beweis:** Die obigen Ungleichungen folgen aus den Hilfssätzen 3.1 und 3.2. Die Gültigkeit der Gleichheit wird durch die Existenz von Graphen entsprechender Ordnung gezeigt. Minimale Graphen zu  $r = 3$  und  $t \in \{3, 4, 5, 6\}$  sind in Anhang B abgebildet.  $\square$

Hier die zwei für den Beweis von Satz 3.4 nötigen Hilfssätze:

**Hilfssatz 3.1** *Es seien  $r, s, t$  natürliche Zahlen mit  $r \geq 3, t \geq 3$  und  $s \leq \lfloor \frac{t-1}{2} \rfloor$ . Es sein  $G = (V, E)$  ein Graph mit Tailenweite  $T(G) = t$ , in dem jeder Knoten Grad  $r \geq r$  besitzt. Für die Anzahl Knoten  $n(r, t)$  von  $G$  gilt:*

$$n(r, t) \geq n^*(r, t) \geq r_s \cdot \frac{(r-1)^{\lfloor (t-1)/2 \rfloor - (s-1)} - 1}{r-2} + 1,$$

wobei  $r_s := n_s$ , die in Abschnitt 2.1.2 definierte Knotenzahl der größten Nachbarschaft  $N_{=s}(v)$  eines Knotens  $v \in V$ .

**Beweis:** Bezeichne  $a \in V$  einen Knoten, für den  $|N_{=s}(a)| = r_s$  mit  $s \leq \lfloor \frac{t-1}{2} \rfloor$  gilt. Dann kann jeder Knoten aus  $N_{\leq \lfloor \frac{t-1}{2} \rfloor}(a)$  von  $a$  aus durch genau einen Weg der Länge  $\leq \lfloor \frac{t-1}{2} \rfloor$  erreicht werden. Gäbe es mehrere Wege dieser Länge, so würde ein Kreis der Länge  $< t$  existieren, was wegen der Tailenweite  $T(G) = t$  nicht möglich ist. Da der Grad eines jeden Knoten  $\geq 3$  ist und wegen  $r_j \geq r_s \cdot (r-1)^{j-s}$  für  $s \leq j \leq \lfloor \frac{t-1}{2} \rfloor$ , gilt nach einfacher Induktion:

$$\begin{aligned} |N_{\leq j}(a)| &\geq |N_{s-1}(a)| + r_s + r_s \cdot (r-1) + \cdots + r_s \cdot (r-1)^{j-s} \\ &= |N_{s-1}(a)| + r_s \cdot \frac{(r-1)^{j-(s-1)} - 1}{r-2} \end{aligned}$$

für  $s \leq j \leq \lfloor \frac{t-1}{2} \rfloor$ .

Man setzt nun  $j = \lfloor \frac{t-1}{2} \rfloor$  und schätzt, wie folgt, ab:

$$\begin{aligned} n(r, t) &\geq |N_{\leq \lfloor \frac{t-1}{2} \rfloor}(a)| \\ &\geq 1 + r_s \cdot \frac{(r-1)^{\lfloor \frac{t-1}{2} \rfloor - (s-1)} - 1}{r-2}, \end{aligned}$$

womit obige Ungleichung gezeigt wäre.  $\square$

Mit  $s = 1$  folgt aus Hilfssatz 3.1 für die Knotenzahl  $n(r, t)$  eines  $r$ -regulären Graphen  $G = (V, E)$ :

$$\begin{aligned} n(r, t) &\geq 1 + r \cdot \frac{(r-1)^{\lfloor \frac{t-1}{2} \rfloor} - 1}{r-2} \\ &= 1 + r \cdot \sum_{i=1}^{\lfloor \frac{t-1}{2} \rfloor} (r-1)^{i-1}. \end{aligned} \tag{3.5}$$

Dann enthält  $G$  einen Baum von dem Typ, wie er in Abbildung 3.3 dargestellt ist. In diesem besitzen alle Knoten, die nicht Endknoten des Baumes sind, den gleichen Grad  $r$ . Es folgt daraus:

Gilt für einen  $r$ -regulären Graphen in Ungleichung 3.5 Gleichheit, so ist offenbar  $t$  ungerade.

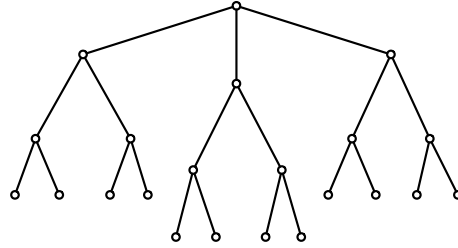


Abbildung 3.3: Baum, den jeder reguläre Baum enthält.

Ist  $t$  gerade, so lässt sich für  $s = 1$  die Abschätzung in Hilfssatz 3.1 nach einer von F. Kerteszi angegebenen Methode verbessern. Dies wird im folgenden Hilfssatz gezeigt:

**Hilfssatz 3.2** *Es seien  $r$  und  $t$  natürliche Zahlen mit  $r \geq 3$  und  $t \geq 4$ . Ferner sei  $G = (V, E)$  ein Graph mit der Tailenweite  $T(G) = t$ , in dem jeder Knoten Grad  $r \geq 3$  besitzt. Dann gilt für die Anzahl Knoten  $n(r, t)$  von  $G$ :*

$$n(r, t) \geq n^*(r, t) \geq 2 + 2 \cdot (r - 1) \frac{(r - 1)^{\lfloor (t-2)/2 \rfloor} - 1}{r - 2}.$$

**Beweis:** Seien  $a, b \in V$  zwei verschiedene Knoten des Graphen. Dann ist  $N_{\leq s}(\{a, b\})$  die Menge der Knoten, die von  $\{a, b\}$  einen Abstand  $\leq s$  haben. Jeder Knoten von  $N_{\leq \lfloor \frac{t-2}{2} \rfloor}(\{a, b\})$  kann von  $\{a, b\}$  aus durch genau einen Weg, der  $a$  bzw.  $b$  nicht enthält, mit einer Länge  $\leq \lfloor \frac{t-2}{2} \rfloor$  erreicht werden. Siehe dazu Abbildung 3.4.

Da der Grad eines jeden Knoten  $\geq 3$  ist, zeigt eine einfache Induktion:

$$|N_{\leq s}(\{a, b\})| \geq 2 + 2 \cdot (r - 1) \cdot \sum_{i=1}^{s-1} (r - 1)^i$$

für  $1 \leq s \leq \lfloor \frac{t-2}{2} \rfloor$ . Man setzt  $s = \lfloor \frac{t-2}{2} \rfloor$  und erhält

$$\begin{aligned} n(r, t) &\geq |N_{\leq \lfloor \frac{t-2}{2} \rfloor}(\{a, b\})| \\ &\geq 2 + 2 \cdot (r - 1) \cdot \frac{(r - 1)^{\lfloor \frac{t-2}{2} \rfloor} - 1}{r - 2}, \end{aligned}$$

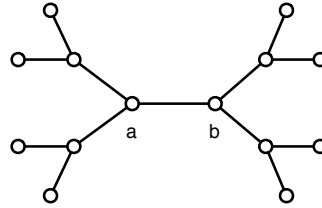


Abbildung 3.4:  $N_{\leq \lfloor \frac{t-2}{2} \rfloor}(\{a, b\})$  für  $r = 3$  und  $t = 6$ .

womit Hilfssatz 3.2 gezeigt wäre.  $\square$

Wie bei Hilfssatz 3.1 überlegt man sich, dass  $t$  gerade sein muß, wenn für einen  $r$ -regulären Graphen in Hilfssatz 3.2 Gleichheit gilt. Und auch hier enthält jeder Graph, der den Voraussetzungen genügt, einen Baum, wie er in Abbildung 3.4 dargestellt ist.

In anderer Literatur, wie [Jør03] sind die unteren Schranken aus Satz 3.4 auch als ‚*Moore-Bound*‘ bekannt.

Anhand dieser unteren Schranke für die Anzahl Knoten eines  $(r, t)$ -Cages‘ stellt man nun fest, dass das obige Beispiel eines  $(r, 3)$ -Cages‘ schon der minimale ‚Cage‘ für  $t = 3$  ist. Somit gilt  $n^*(r, 3) = r + 1$ . Dass auch der  $C_t$  bis auf Isomorphie der minimale  $(2, t)$ -Cage‘ ist und demnach  $n^*(2, t) = t$  gilt, ist offensichtlich.

Wähle man nun die Tailenweite  $t = 4$ , so hat L.Posa mit Hilfe von Satz 3.4 und Hilfssatz 3.2 folgendes gezeigt (siehe [WV74], Seite 121):

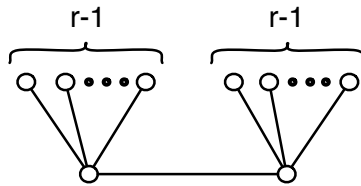
**Satz 3.5** *Es gilt  $n^*(r, 4) = 2 \cdot r$ , und die Menge der  $(r, 4)$ -Cages‘ der Ordnung  $2 \cdot r$  enthält bis auf Isomorphie genau einen Graphen, den vollständigen bipartiten Graphen  $K_{r,r}$ .*

**Beweis:** Aus Satz 3.4 folgt unmittelbar  $n(r, 4) \geq n^*(r, 4) \geq 2 \cdot r$ . Der vollständige bipartite Graph  $K_{r,r}$  ist  $r$ -regulär und hat offensichtlich Tailenweite  $T(K_{r,r}) = 4$ . Folglich ist  $n^*(r, 4) = 2 \cdot r$ .

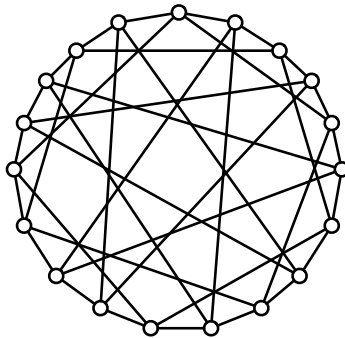
Aus dem Beweis von Hilfssatz 3.2 folgt, dass jeder  $(r, 4)$ -Cage‘  $G$  einen Baum enthält, wie er in Abbildung 3.5 dargestellt ist. Man überzeugt sich nun leicht, dass  $G$  isomorph zu  $K_{r,r}$  ist.  $\square$

Der in Abschnitt 3.2.1 schon erwähnte bipartitete ‚Circulant‘-Graph gerader Ordnung  $n$  ist also ein Untergraph des minimalen  $(n/2, 4)$ -Cage‘. Er besteht ebenfalls aus  $n$  Knoten, hat jedoch eine kleinere Kantenmenge als dieser.



Abbildung 3.5: Baum, den jeder  $(r, 4)$ -Cage enthält.

Erhöht man die Tailenweite auf 5, so beträgt die untere Schranke aus Satz 3.4 für die Knotenzahl des minimalen  $(r, 5)$ -Cage  $r^2 + 1$ . Will man also einen Graphen mit Tailenweite 5, so steigt die Knotenzahl mindestens polynomial mit dem Knotengrad. Am Beispiel des  $(4, 5)$ -Cage sieht man jedoch, dass die Knotenzahl des minimalen ‚Cage‘ von dieser Schranke abweichen kann. Der Robertson (siehe Abbildung 3.6) ist bis auf Isomorphie der einzige minimale  $(4, 5)$ -Cage und besteht aus  $19 > 17 = 4^2 + 1$  Knoten.

Abbildung 3.6: Robertson-Graph,  $(4, 5)$ -Cage.

Weitere minimale ‚Cages‘ mit  $t = 5$  und deren Knotenzahlen sind in [Wei01] zu finden. So ist der bisher einzige bekannte minimale  $(8, 5)$ -Cage der Cayley-Graph der Ordnung 80. Hier ist die Differenz zwischen der tatsächlichen Anzahl und deren untere Schranke von 65 Knoten also recht groß. In Bezug auf den Kontaktgraphen bedeutet dies, dass durch eine Erhöhung der Tailenweite auf 5 zum einen nur bestimmte Teilnehmerzahlen  $n$  zu vorgegebener Gesprächszahl und Tailenweite möglich sind und zum anderen dann auch nicht mehr als  $r \leq \sqrt{n-1}$  Gespräche pro Person möglich sind. Bleibt man aus diesem Grund bei Tailenweite 4, erhält man wieder die bipartiten

Graphen, die schon im vorherigen Abschnitt als geeignete Kontaktgraphen erarbeitet wurden.

### 3.2.3 Perfekte Graphen

Nun betrachte man Graphen bestimmter Taillenweite ohne die Regularität zu berücksichtigen. Hat ein Graph eine hohe Taillenweite, so sieht er lokal um jeden Knoten herum aus wie ein Baum. Er ist dort also jeweils 2-färbbar. Es gibt nach einem Satz von Erdős (siehe [Die96], S.235) allerdings auch Graphen beliebig hoher Taillenweite und gleichzeitig beliebig hoher Chromatischer Zahl. Eine hohe Chromatische Zahl ist demnach nicht notwendig eine lokale Dichteigenschaft. Die Klasse der *perfekten Graphen* beinhaltet nun jene Graphen  $G$ , bei denen die triviale untere Schranke von  $Cl(G)$  Farben zur Knotenfärbung ausreicht, wobei  $Cl(G)$  als die Cliquenzahl eines Graphen  $G$  definiert wurde.

**Definition 3.3** *Ein Graph  $G$  heißt perfekt, wenn in all seinen Subgraphen  $G' \subseteq G$  gilt, dass die Chromatische Zahl  $Ch(G')$  der Cliquenzahl  $Cl(G')$  entspricht.*

Enthält ein Graph eine Clique der Ordnung 3 oder größer, so beinhaltet er Dreiecke. Aus diesem Grund wählt man für den Kontaktgraphen  $G$  die Cliquenzahl  $Cl(G) = 2$ . Dass dann die Chromatische Zahl ebenfalls 2 beträgt, bedeutet, dass der Graph bipartit ist. Die bipartiten Graphen zählen demnach zur Klasse der perfekten Graphen. Also auch in Anbetracht der perfekten Graphen stellen sich die 2-färbbaren Graphen als optimal für die zu realisierenden Anforderungen heraus.

### 3.2.4 ,Circulant'-Graphen

Bei Betrachtung der ,*Circulant'-Graphen*<sup>3</sup> als Kontaktgraphen spielt wieder die Regularität eine Rolle. Diese Graphen sind nämlich knotentransitiv und somit regulär.

**Definition 3.4** *Sei die Menge  $S \subseteq \{1, 2, \dots, \lfloor n/2 \rfloor\}$ . Der ,Circulant'-Graph  $Ci(n, S)$  der Ordnung  $n$  ist der Graph mit Knotenmenge  $V = \{0, 1, \dots, n-1\}$  und Kantenmenge  $E = \{(i, j) \mid \exists a \in S : i + a \equiv j \pmod{n}\}$ .*

Der ,Circulant'-Graph  $Ci(n, S)$  mit  $S = \{1, 2, \dots, \lfloor n/2 \rfloor\}$  gibt den vollständigen Graphen  $K_n$  an, und der Graph  $Ci(n, S)$  mit  $S = \{1\}$  ist der

<sup>3</sup> von ,circulating' (engl.) heißt zirkulierend, kursierend

Kreisgraph  $C_n$ . Einen weiteren ‚Circulant‘-Graphen zeigt Abbildung 3.7. Der  $K_{3,3}$  aus Abbildung B.2(b) im Anhang ist ebenfalls ein Graph dieser Klasse mit  $n = 6$  und  $S = \{1, 3\}$ . An diesem Beispiel sieht man, dass wenn  $n$  gerade und  $n/2 \in S$  gilt, der Graph  $(2 \cdot |S| - 1)$ -regulär ist. In allen anderen Fällen sind ‚Circulant‘-Graphen  $(2 \cdot |S|)$ -regulär.

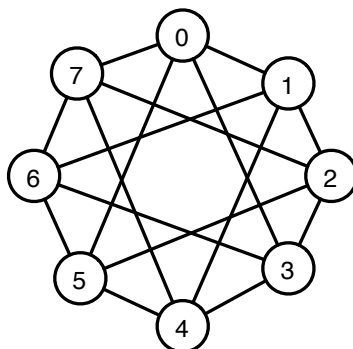


Abbildung 3.7: ‚Circulant‘-Graph  $Ci(8, S)$  mit  $S = \{1, 3\}$ .

Durch Wahl spezieller Mengen  $S$  entstehen Teilmengen der Klasse der ‚Circulant‘-Graphen. Im Folgenden werden zwei solche in Bezug auf ihre Tauglichkeit als Kontaktgraphen betrachtet.

### Bipartite ‚Circulant‘-Graphen

Wie oben angegeben, ist der ‚Circulant‘-Graph  $Ci(6, S)$  mit  $S = \{1, 3\}$  der vollständige bipartite Graph  $K_{3,3}$ . Folgender Satz zeigt, dass die Schnittmenge der bipartiten Graphen und der ‚Circulant‘-Graphen gerader Ordnung durch die Wahl bestimmter Mengen  $S$  gegeben ist.

**Satz 3.6** ‚Circulant‘-Graphen  $Ci(n, S)$  gerader Ordnung  $n$  sind genau dann bipartit, wenn  $S$  nur aus ungeraden Zahlen besteht.

**Beweis:** Sei  $Ci(n, S) = (V_1 \cup V_2, E)$  ein bipartiter ‚Circulant‘-Graph gerader Ordnung  $n$ . Dann gilt aufgrund der Regularität, wie schon in Abschnitt 3.2.1 gezeigt,  $|V_1| = |V_2|$ . Nummeriert man nun die Knoten des Graphen so, dass  $V_1$  die Knoten gerader Nummerierung und  $V_2$  die Knoten ungerader Nummerierung sind, so kann  $S$  nur aus ungeraden Zahlen bestehen. Sei nun umgekehrt  $Ci(n, S)$  ein ‚Circulant‘-Graph gerader Ordnung  $n$  mit  $S$  bestehend aus ungeraden Zahlen, dann verbinden die Kanten des Graphen

ausschließlich Knoten gerader Nummerierung mit Knoten ungerader Nummerierung. Der Graph ist somit bipartit, wobei die Knoten gerader Nummerierung und die ungerader Nummerierung die disjunkten Knotenmengen bilden.  $\square$

Mit Hilfe dieses Satzes kann man folgende Definition der ‚Circulant‘-Graphen, die bipartit sind, formulieren:

**Definition 3.5** *Der bipartite ‚Circulant‘-Graph  $BCi(n, S)$  mit gerader Knotenzahl  $n$  ist ein ‚Circulant‘-Graph, dessen Menge  $S$  nur aus ungeraden Zahlen besteht.*

Diese Graphenklasse fasst somit die schon in Abschnitt 3.2.1 betrachteten  $r$ -regulären, bipartiten Graphen zusammen. Wie dort erläutert, wird bei ungerader Anzahl von Teilnehmern an der Gruppeneinteilung ein ‚Dummy‘-Knoten hinzugefügt. Wegen der Bipartitheit und der Regularität ist dann eine Gesprächsführung ohne Aussetzer möglich. Anhand der Definition der ‚Circulant‘-Graphen können sehr einfach deren kantendisjunkte perfekte Matchings ausgemacht werden. Die Kanten zu je einem Element von  $S - \{n/2\}$  bilden zwei dieser perfekten Matchings und die Kanten zu  $n/2$  eines.

Die in Abschnitt 3.2.1 beschriebene Gesprächsführung ist nun isomorph zum bipartiten ‚Circulant‘-Graphen  $BCi(n, S)$ . Dabei besteht  $S$  aus aufeinander folgenden ungeraden Zahlen. Es gilt genauer  $S \subsetneq \{1, 3, 5, \dots, s\}$  mit  $s := n/2 - 1$ , falls  $n$  durch 4 teilbar, bzw.  $s := n/2$ , falls nicht, und  $S$  entweder 1 oder  $s$  enthält. In diesem Graphen haben ebenfalls je zwei Knoten  $x, y \in V$ , für die  $x - y \equiv 2 \cdot d \pmod{n}$  mit  $1 \leq d \leq r$  gilt,  $(r - d)$  gemeinsame Nachbarn. Die Streuung dieses Graphen liegt damit bei  $(r - 1)/r$ , also nahezu bei 1. Diese Wahl der Menge  $S$  scheint demnach eine sehr große Streuung im bipartiten Graphen zu verursachen. Lässt man sogar  $S = \{1, 3, 5, \dots, n - 1\}$  zu, so erhöht sich auch hier die Streuung auf 1.

Was passiert, wenn  $S \subsetneq \{1, 3, 5, \dots, n - 1\}$  nicht aus aufeinander folgenden ungeraden Zahlen besteht und  $1, s \notin S$  gilt? Besteht  $S$  aus mehreren einzelnen Folgen  $S = \{3, 5\} \cup \{9, 11\} \cup \dots$ , so kann bei bestimmter Knotenzahl die Streuung trotzdem 1 oder zumindest nahezu 1 sein. Als Beispiel dazu betrachte man den Graphen  $BCi(28, S)$  mit  $S = \{3, 7, 11\}$ . Die Knoten mit der Nummerierung 0 und 14 beispielsweise haben die gemeinsame Nachbarschaft  $N(0) \cap N(14) = \{3, 7, 11, 17, 21, 25\}$ . Dieser Graph hat demnach Streuung 1.

Die Flexibilität dieser Graphenklasse in Bezug auf Knotenzahl wurde schon in Abschnitt 3.2.1 erläutert. Die Möglichkeiten für die Höhe der Regularität hängt, wie oben beschrieben, von der Wahl der Menge  $S$ , also Art des jeweiligen bipartiten ‚Circulant‘-Graphen ab.

### Rekursive ‚Circulant‘-Graphen

Eine weitere Teilmenge von ‚Circulant‘-Graphen bildet die Klasse der *rekursiven ‚Circulant‘-Graphen*. Sie heißen „rekursiv“, weil J.-H. Park und K.-Y. Chwa gezeigt haben, dass sie bei bestimmter Ordnung  $n$  rekursiv aus kleineren Graphen ihrer Art konstruiert werden können (siehe [FR98], Seite 66).

**Definition 3.6** *Der rekursive ‚Circulant‘-Graphen  $RCi(n, d)$  ist ein ‚Circulant‘-Graph, dessen Menge  $S$  aus den Potenzen  $\{d^0, d^1, d^2, \dots, d^{\lceil \log_d n \rceil - 1}\}$  besteht.*

*Wählt man  $S' \subset S$  so, dass  $S' = \{d^0, d^1, d^2, \dots, d^{r-1}\}$  mit  $r < \lceil \log_d n \rceil$ , dann entsteht der rekursive ‚Circulant‘-Graph  $RCi(n, d, r)$ , ein Subgraph des  $RCi(n, d)$ .*

Die Graphen  $RCi(n, d)$  sind nach Definition  $(2 \cdot \lceil \log_d n \rceil - 1)$ -regulär, falls  $n/2 = d^k$  ( $k \in \mathbb{N}$ ), und  $(2 \cdot \lceil \log_d n \rceil)$ -regulär, sonst. In  $RCi(n, d, r)$  wird der Grad eines jeden Knoten auf  $2 \cdot r$  reduziert.

Jedoch sind die rekursiven ‚Circulant‘-Graphen nicht unbedingt bipartit und können daher Dreiecke enthalten. Man betrachte beispielsweise den Graphen  $RCi(12, 4)$ ; die Knoten mit den Nummerierungen 0, 4 und 8 bilden hier ein Dreieck.

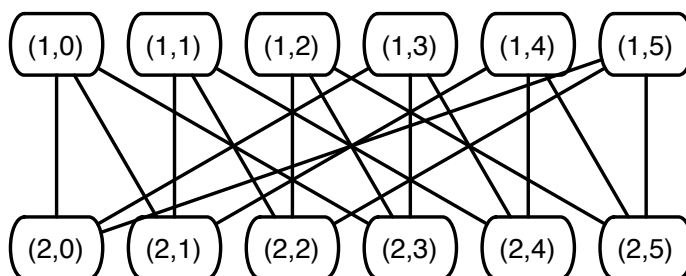
### 3.2.5 Knödel-Graphen

Die *Knödel-Graphen* sind vom Prinzip her ähnlich aufgebaut wie die rekursiven ‚Circulant‘-Graphen. Sie sind ebenfalls knotentransitiv und demnach regulär. Es werden zwei alternative Definitionen genannt: die Erste (Definition 3.7) macht die Bipartitheit des Knödel-Graphen ersichtlich, die Zweite (Definition 3.8) ähnelt der Definition der ‚Circulant‘-Graphen.

**Definition 3.7** *Der vollständige Knödel-Graph  $Knö(n)$  der Ordnung  $n \geq 4$ ,  $n$  gerade, wird gebildet aus Knoten  $(i, j)$ , wobei  $i = 1, 2$  und  $0 \leq j \leq n/2 - 1$ , und aus Kanten zwischen jeglichen zwei Knoten  $(1, j)$  und  $(2, (j + 2^l - 1) \bmod n/2)$  für jedes  $j, 0 \leq j \leq n/2 - 1$  und jedes  $l, 0 \leq l \leq \lceil \log_2 n \rceil - 1$ . Alle Kanten zu einem bestimmten  $l$  werden von der *Dimension  $l$*  genannt.*

*Gilt  $0 \leq l \leq r - 1$  mit  $r < \lceil \log_2 n \rceil$ , so entsteht der *r-reguläre Knödel-Graph  $Knö(n, r)$ , ein Subgraph des  $Knö(n)$ .**

Im Graphen  $Knö(n)$  liegt nach Definition  $\lceil \log_2 n \rceil$ -Regularität vor. In  $Knö(n, r)$  kann durch Wahl einer kleineren oberen Schranke  $r < \lceil \log_2 n \rceil$  für die Dimension  $l$  der Grad eines jeden Knoten auf  $r$  reduziert werden. Anhand dieser Definition erkennt man leicht, dass der Knödel-Graph  $G = (V_1 \cup V_2, E)$

Abbildung 3.8: Knödel-Graph  $Kn\ddot{o}(12)$  nach Definition 3.7.

bipartit ist mit  $V_1 = \{(1, j) \mid 0 \leq j \leq n/2 - 1\}$  und  $V_2 = \{(2, j) \mid 0 \leq j \leq n/2 - 1\}$ . Abbildung 3.8 zeigt einen Beispielgraphen der Ordnung 12.

In der zweiten, im Weiteren verwendeten Definition der Knödel-Graphen werden die Knoten wieder von 0 bis  $n - 1$  nummeriert. Im Vergleich zur vorherigen Definition gilt für die Bezeichnungen der Knoten:  $(1, j) \hat{=} 2 \cdot j$  und  $(2, j) \hat{=} 2 \cdot j + 1$  mit  $0 \leq j \leq n/2 - 1$ . Die Knotenmengen des bipartiten Graphen  $G = (V_1 \cup V_2, E)$  werden somit von den Knoten gerader und ungerader Nummerierung gebildet. Um diese Definition besser nachvollziehen zu können, betrachte man parallel den Beispielgraphen in Abbildung 3.9.

**Definition 3.8** *Der vollständige Knödel-Graph  $Kn\ddot{o}(n)$  der Ordnung  $n \geq 4$ ,  $n$  gerade, ist der Graph mit Knotenmenge  $V = \{0, 1, \dots, n - 1\}$  und Kantenmenge  $E = \{(i, j) \mid i \text{ gerade, } j \text{ ungerade, } i + 2^l - 1 \equiv j \pmod{n} \text{ f\"ur jedes } l, 1 \leq l \leq \lfloor \log_2 n \rfloor\}$ .*

*Gilt  $1 \leq l \leq r$  mit  $r < \lfloor \log_2 n \rfloor$ , so entsteht der  $r$ -reguläre Knödel-Graph  $Kn\ddot{o}(n, r)$ , ein Subgraph des  $Kn\ddot{o}(n)$ .*

Die Knödel-Graphen  $Kn\ddot{o}(n)$  bzw.  $Kn\ddot{o}(n, r)$  sind also bipartit, enthalten demnach keine Dreiecke. Desweiteren sind sie  $\lfloor \log_2 n \rfloor$ -regulär bzw.  $r$ -regulär mit  $r < \lfloor \log_2 n \rfloor$  und haben gerade Knotenzahl  $n$ . Verwendet man sie somit als Kontaktgraphen, ist eine Gesprächsführung ohne Aussetzer möglich. Die Kanten der Dimension  $l$  bilden dabei je eines der entsprechenden perfekten Matchings. Bei ungerader Anzahl von Teilnehmern wird wieder ein ‚Dummy‘-Knoten eingefügt, und  $\lfloor \log_2 n \rfloor$  bzw.  $r$  Personen müssen dann je einmal aussetzen.

Bestimmte Knödel-Graphen, sowie rekursive ‚Circulant‘-Graphen sind in der Informationsverbreitung von besonderem Interesse. So gehören  $Kn\ddot{o}(n)$  und  $RCi(n, 4)$  mit  $n = 2^k, k \in \mathbb{N}$  zu den drei Familien nicht isomorpher minimaler ‚Gossip‘- und ‚Broadcast‘-Graphen. Aber auch für andere Knotenzahlen zeigen G. Fertin und A. Raspaud in [FR98], dass viele minimale ‚Gossip‘-

und ‚Broadcast‘-Graphen isomorph zu entsprechenden Knödel-Graphen sind. Der Durchmesser ist eine der Haupt-Invarianten in Kommunikationsnetzwerken, da er die Entfernung angibt, die eine Information auf ihrem kürzesten Weg von der Quelle zum Ziel im schlechtesten Fall zurückzulegen hat. Obwohl der Knödel-Graph schon vor etwa 30 Jahren von W. Knödel eingeführt wurde, konnte der Durchmesser dieses Graphen im Allgemeinen noch nicht ermittelt werden. Nur für den Fall  $n = 2^k$  mit  $k \in \mathbb{N}$  und  $k \geq 2$  zeigen obige Autoren in [FR<sup>+</sup>00], dass

$$D(\text{Knö}(2^k)) = \left\lceil \frac{k+2}{2} \right\rceil.$$

Die Streuung des Knödel-Graphen kann hingegen für den allgemeinen Fall ( $n \in \mathbb{N}$ ) gezeigt werden. Es gilt folgender Satz:

**Satz 3.7** *Sei  $\text{Knö}(n) = (V, E)$  ein Knödel-Graph mit  $n \geq 4$ . Für zwei Knoten  $v_1, v_2 \in V$ ,  $v_1 \neq v_2$  gilt:*

$$|N(v_1) \cap N(v_2)| \leq 2.$$

**Beweis:** Wählt man  $v_1 \in V_1$  und  $v_2 \in V_2$ , wobei  $\text{Knö}(n) = (V_1 \cup V_2, E)$  gilt, so folgt  $N(v_1) \cap N(v_2) = \emptyset$ . Seien also o.B.d.A.  $v_1, v_2 \in V_1$ , mit  $v_1 \neq v_2$  und  $V_1$  die Menge der Knoten gerader Nummerierung. Mit  $k := \lfloor \log_2 n \rfloor$  können deren Nachbarschaften wie folgt angegeben werden:

$$\begin{aligned} N(v_1) &= \{v_1 + 2^1 - 1, v_1 + 2^2 - 1, \dots, v_1 + 2^k - 1\} \\ N(v_2) &= \{v_2 + 2^1 - 1, v_2 + 2^2 - 1, \dots, v_2 + 2^k - 1\} \end{aligned}$$

Dann gilt für  $x \in N(v_1)$ ,  $y \in N(v_2)$ :

$$\begin{aligned} x = y &\quad \text{mit} \quad x = v_1 + 2^{l_x} - 1, \quad y = v_2 + 2^{l_y} - 1 \quad \text{und} \quad l_x, l_y \in \{1, 2, \dots, k\} \\ \Leftrightarrow &\quad v_1 + 2^{l_x} - 1 \equiv v_2 + 2^{l_y} - 1 \pmod{n}. \end{aligned}$$

Aufgrund der Knotentransitivität, sei nun o.B.d.A.  $v_1 = 0$ . Dann folgt:

$$\begin{aligned} x = y, \quad x \in N(0), y \in N(v_2) \\ \Leftrightarrow &\quad 2^{l_x} \equiv v_2 + 2^{l_y} \pmod{n} \\ \Leftrightarrow &\quad v_2 \equiv 2^{l_x} - 2^{l_y} \pmod{n} \quad \text{für} \quad l_x, l_y \in \{1, 2, \dots, k\} \end{aligned} \quad (3.6)$$

Der Knoten 0 hat also genau dann gemeinsame Nachbarn mit dem Knoten der Nummerierung  $v_2$ , wenn  $v_2$  äquivalent zur Differenz zweier Zweierpotenzen  $2^{l_x}$  und  $2^{l_y}$  ist, wobei  $l_x, l_y \in \{1, 2, \dots, k\}$ . Da  $0 \neq v_2 \in V_1$ , also  $0 < v_2 < n$

und wegen  $l_x, l_y \in \{1, 2, \dots, k\}$ , kann es dann nur folgende zwei Möglichkeiten geben:

$$\begin{aligned} v_2 &= 2^{l_x} - 2^{l_y} && \text{mit } l_x > l_y \\ v_2 &= 2^{l_x} - 2^{l_y} + n && \text{mit } l_x < l_y. \end{aligned}$$

Damit ist gezeigt, dass für je zwei Knoten  $v_1, v_2 \in V$ ,  $v_1 \neq v_2$

$$|N(v_1) \cap N(v_2)| \leq 2$$

gilt.

Betrachte man speziell den Fall, dass  $n = 2^k$  gilt. Dann gibt es in Äquivalenz 3.6 für den Knoten  $v_2$ , die folgenden zwei Möglichkeiten, falls  $v_2$  ebenfalls eine Zweierpotenz ist:

$$\begin{aligned} l_y = \log_2(v_2) \quad \wedge \quad l_x = l_y + 1 & \quad \text{oder} \\ l_x = \log_2(v_2) \quad \wedge \quad l_y = k. & \end{aligned}$$

Somit haben in  $Kn\ddot{o}(2^k)$  ( $k \in \mathbb{N}, k \geq 2$ ) zwei Knoten derselben Knotenmenge, deren Nummerierungen sich um eine Zweierpotenz unterscheiden, je genau zwei gemeinsame Nachbarn. Bei den übrigen Knoten hängt es von der Anzahl Knoten ab, welche von ihnen zwei, einen oder auch keine gemeinsamen Nachbarn mit dem Knoten der Nummerierung 0 haben.  $\square$

Betrachte man als Beispiel für die Nachbarschaften der Knoten, deren Nummerierung keine Zweierpotenz ist, den Graphen  $Kn\ddot{o}(16)$ . Für die Knotennummerierungen 6, 10, 12 und 14 gelten folgende Gleichungen:

$$\begin{aligned} 6 &= 2^3 - 2^1, \\ 10 &= 2^1 - 2^3 + 16, \\ 12 &= 2^4 - 2^2 = 2^2 - 2^3 + 16, \\ 14 &= 2^4 - 2^1 = 2^1 - 2^2 + 16. \end{aligned}$$

Die Knoten 6 und 10 haben somit je einen gemeinsamen Nachbarn mit Knoten 0, die Knoten 12 und 14 hingegen zwei. Abbildung 3.9 zeigt die gemeinsamen Nachbarn der Knoten 0 und 12 in  $Kn\ddot{o}(16)$ .

Die Streuung des Knödel-Graphen  $Kn\ddot{o}(n)$  mit  $n \geq 4$  beträgt demnach  $\frac{2}{\lceil \log_2 n \rceil}$ . Betrachtet man den Graphen  $Kn\ddot{o}(n, r)$ , so vergrößert sich die Streuung entsprechend auf  $\frac{2}{r}$ .

Der Knödel-Graph  $Kn\ddot{o}(n)$  erfüllt unter allen betrachteten Graphen am besten die genannten Anforderungen. Er wird daher als Kontaktgraph der hier entwickelten Strategie verwendet.



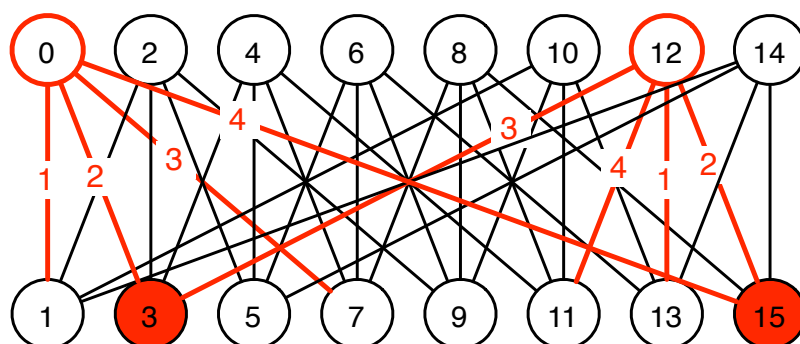


Abbildung 3.9: Nachbarschaft im Knödel-Graphen  $Kn\ddot{o}(16)$  – Die Zahlen an den Kanten sind die jeweilige Dimension.

Wenn bei einer Veranstaltung allerdings mehr als  $\lfloor \log_2 n \rfloor$  Gespräche pro Person geführt werden sollen, wird die Bipartitheit des Knödel-Graphen ausgenutzt: Es werden zusätzliche perfekte Matchings für weitere Gesprächsrunden eingefügt. Dabei vergrößert sich zwar die Streuung, die geforderten Kriterien der Dreiecksfreiheit und der geringen Aussetzerzahl bleiben jedoch erfüllt. Aus diesem Grund wird auch hingenommen, dass die Größe der Streuung eines abgewandelten Knödel-Graphen im Allgemeinen nicht bekannt ist. Im folgenden Abschnitt wird die Streuung eines Beispielgraphen angegeben, dessen Regularität  $\lfloor \log_2 n \rfloor + 3$  beträgt.

### 3.3 Anwendungsbeispiel

Zum ersten Mal angewendet wurde eine Abwandlung des Knödel-Graphen als Kontaktgraph für eine Gruppeneinteilung bei der POLNET Summer School im Sommer 2004. An dieser nahmen 37 Personen teil. Der zu verwendende Kontaktgraph hat demnach Ordnung 38. Aufgrund ungerader Teilnehmerzahl wurde ein ‚Dummy‘-Knoten hinzugefügt. Da  $\lfloor \log_2 n \rfloor = 5$  gilt, ist der vollständige Knödel-Graph  $Kn\ddot{o}(38)$  5-regulär. Es sind im Prinzip also nur 5 Gesprächsrunden möglich. Um die vom Veranstalter erwünschten 8 Gespräche pro Person zu erhalten, spricht jede Person gerader Nummerierung  $x$  zusätzlich mit den Personen  $(x + 2^l - 1) \bmod n$  für jedes  $3 \leq l \leq \Delta$ . Dieser spezielle Kontaktgraph ist dann 8-regulär. Seine Streuung beträgt  $5/8 = 0,625$ , da je zwei Knoten maximal 5 gemeinsame Nachbarn haben. Es mußten bei der Gesprächsführung demnach insgesamt 8 verschiedene Personen je in einer Gesprächsrunde aussetzen. Anschließend hat jeder Teilnehmer

auf einem Stimmzettel vier Personen angekreuzt, mit denen er gerne in eine Gruppe eingeteilt werden möchte. Anhand dieser Bewertungen wurden dann die Gruppen eingeteilt. Da zu diesem Zeitpunkt der Algorithmus zur Gruppeneinteilung zwar schon fest stand, aber noch nicht implementiert war, wurden die Gruppen per Hand ermittelt. Um dabei die Arbeit etwas geringer zu halten, wurden für die Einteilung nur die 47 erwiderten Sympathien berücksichtigt.

# Kapitel 4

## Bewertungsformen

Ziel des zweiten Arbeitsschrittes ist, eine Form zu finden, in der jeder Teilnehmer seine Gesprächspartner bewertet. Durch die flüchtigen Kontakte tun sich erste Sym- und Antipathien zwischen den teilnehmenden Personen auf, die erwidert sein können oder einseitig sind. Anhand von *Bewertungen*, die diese Sym- und Antipathien ausdrücken, soll im letzten Arbeitsschritt die Einteilung in Gruppen vorgenommen werden, die möglichst viele Sym- und Antipathien erfüllt. Sie wird durch eine Knotenpartition dargestellt.

### 4.1 Erzeugung des Sympathie- und Antipathiegraphen

#### 4.1.1 Unskalierte Graphen

Betrachtet man nach der Gesprächsführung eine Kante  $(u, v) \in E$  des Kontaktgraphen  $G = (V, E)$ , so steht diese für einen stattgefundenen flüchtigen Kontakt zwischen den zwei Personen  $u, v \in V$ . Ist  $v$  nun einer der Gesprächspartner von  $u$ , die  $u$  sympathisch findet, so stellt man diese sogenannte *Sympathiebewertung* im gerichteten *Sympathiegraphen* mit Knotenmenge  $V$  durch eine Kante  $(u, v)_g$  dar. Analog zum Sympathiegraphen entsteht der gerichtete *Antipathiegraph* aus den *Antipathiebewertungen*.

Gibt es auch die Möglichkeit einen Gesprächspartner *neutral* zu bewerten, das heißt keine Sympathie und keine Antipathie auszusprechen, so enthalten der Sympathie- und der Antipathiegraph an der entsprechenden Stelle keine gerichtete Kante. Allerdings spielt die Richtung einer Kante bei der Gruppeneinteilung keine Rolle; Denn ist die  $(u, v)_g$  Partitions-kante einer Knotenpartition, so auch  $(v, u)_g$ . Aus diesem Grund werden gerichtete Sympathie- und Antipathiegraphen zuerst in ungerichtete Graphen *konvertiert*. Dies kann

auf unterschiedliche Art und Weise geschehen, wie unten beschrieben. Allen Möglichkeiten gemein ist, dass ein Graph konstruiert wird, der zu jeder gerichteten Kante  $(u, v)_g$  die ungerichtete Kante  $(u, v)$  enthält. Die aus den oben beschriebenen gerichteten Graphen durch Konvertierung entstehenden Graphen werden im Folgenden wieder als Sympathie- und Antipathiegraphen bezeichnet. Ist die gerichtete Form dieser Graphen gemeint, so wird dies im Folgenden explizit angegeben.

Existiert zwischen zwei Knoten  $u$  und  $v$  nur die gerichtete Kante  $(u, v)_g$ , so enthält der ungerichtete Graph wie oben beschrieben die Kante  $(u, v)$ . Der Existenz beider Kanten  $(u, v)_g$  und  $(v, u)_g$  kommt im Fall der hier betrachteten Bewertungen allerdings eine besondere Bedeutung zu. Sie drücken eine erwiderte Sym- bzw. Antipathie aus, welche für die Einteilung der Gruppen eine stärkere Rolle spielen, als wären diese nur einseitig. Um diese Information bei der Konvertierung nicht zu verlieren, werden Kantengewichte eingesetzt. Die Kanten  $(u, v)_g$  und  $(v, u)_g$  im gerichteten Graphen werden zur Kante  $e = (u, v)$  mit Kantengewicht  $w(e) = 2$  konvertiert. Die Kante  $(u, v)_g$  zu einer einseitigen Sym- bzw. Antipathie wird im ungerichteten Sympathie- bzw. Antipathiegraphen zur Kante  $e = (u, v)$  mit  $w(e) = 1$ . Diese Kante fällt demnach im Schnitt einer Knotenpartition nur halb so stark ins Gewicht wie die Kante, die eine erwiderte Sympathie oder Antipathie darstellt. Die durch Konvertierung entstehenden Sympathie- und Antipathiegraphen, werden in Anbetracht des folgenden Abschnittes auch als *unskaliert* bezeichnet.

### 4.1.2 Erweiterung durch Skalierung

Eine erweiterte Form der Bewertung bildet die *Skalierung* der Sym- und Antipathien, das heißt es werden Sympathie- und Antipathiebewertungen verschiedener *Stärken* verwendet. Beispielsweise auf einer Sympathieskala von  $1, 2, \dots$  bis  $p$  bedeutet Stärke 1, dass die Sympathie gering und Stärke  $p$ , dass diese stark ist. Eine Bewertung mit 0 gibt die Möglichkeit einen Gesprächspartner neutral einzustufen, also keinen Wunsch zu äußern, ob man mit dieser Person in eine Gruppe möchte. Diese Stärken werden schließlich als Kantengewichte in den gerichteten Sympathie- bzw. Antipathiegraphen verwendet.

Um nun bei der Konvertierung in die sogenannten *skalierten* Graphen, wie oben beschrieben, keine Information zu verlieren, entsteht aus zwei entgegengesetzten Kanten  $e_1 = (u, v)_g$  und  $e_2 = (v, u)_g$  die Kante  $e = (u, v)$  mit  $w(e) = w(e_1) + w(e_2)$ . Statt die Werte zu addieren, kann auch deren arithmetisches Mittel genommen werden. In beiden Fällen ist jedoch zu berücksichtigen, dass eine nicht existierende Kante Kantengewicht 0 hat. Das Kantengewicht einer einseitigen Sym- bzw. Antipathie ist somit auch hier kleiner

als das Kantengewicht einer erwiderten. Werden Bewertungen verschiedener Stärken verwendet, so ist das Kantengewicht der nicht erwiderten Kante allerdings nicht unbedingt halb so groß wie das der erwiderten.

## 4.2 Vereinigung von Sym- und Antipathiegraph?

Die Informationen aus Sympathie- und Antipathiegraph in einem Graphen zusammenzufassen, ist eine nahe liegende Idee. Im Folgenden wird untersucht, wie diese Vereinigung der Graphen prinzipiell aussehen könnte und ob sie in diesem speziellen Fall sinnvoll ist.

Für die Einteilung der Gruppen wird allerdings ein einfacher Graph erwartet. Enthalten also sowohl der Sympathie- als auch der Antipathiegraph die Kante  $(u, v)$ , so muss eine Methode gefunden werden, wie der neue Graph diese Tatsache wiedergibt. Zudem besitzen die ursprünglichen Graphen Kantengewichte. Will man diese miteinander verrechnen und als Kantengewichte des sogenannten *Vereinigungsgraphen* verwenden, ist es günstig, die Kantengewichte des Antipathiegraphen mit  $(-1)$  zu multiplizieren. Damit wird gewährleistet, dass beispielsweise die Summe der Gewichte zweier Kanten  $(u, v)$  widerspiegelt, ob eine stärkere Sympathie oder eine stärkere Antipathie zwischen den Personen  $u$  und  $v$  vorliegt.

Von den vielen verschiedenen Möglichkeiten, die es zur Verrechnung gibt, werden hier das arithmetische Mittel und die Addition untersucht: Das arithmetische Mittel und die Addition der Kantengewichte ergeben ein gutes Maß für eine Bewertung im Vereinigungsgraphen. Beide liefern eine neutrale Bewertung, wenn man Sym- und Antipathie gleicher Stärke miteinander verrechnet, und beispielsweise eine negative Bewertung, wenn die Stärke der Antipathie größer ist als die Stärke der Sympathie. Hierbei muss beachtet werden, dass eine nicht existierende Kante das Gewicht 0 hat.

Durch eine stärkere Gewichtung des Antipathiegraphen vor der Verrechnung kann gewährleistet werden, dass Antipathien bei der Gruppeneinteilung „Vorrang“ vor Sympathien haben. Dazu werden alle Kantengewichte des Graphen mit einem entsprechenden Wert multipliziert. Ist dann eine Kante sowohl im Sympathie- als auch im Antipathiegraphen enthalten, und wird der Antipathiegraph doppelt gewichtet, so muss beispielsweise bei der Addition die Sympathiebewertung doppelte Stärke wie die Antipathiebewertung haben, um im Vereinigungsgraphen eine neutrale Bewertung zu erhalten.

Es stellt sich jedoch die Frage, ob bei der Vereinigung von Sympathie- und Antipathiegraphen wichtige Informationen verloren gehen. Man betrachte

dazu folgenden Fall bei Verrechnung durch Addition:

Sei  $G_s = (V, E_s)$  der Sympathiegraph mit Gewichtsfunktion  $s : E_s \rightarrow \mathbb{R}$ ,  
 $G_a = (V, E_a)$  der Antipathiegraph mit Gewichtsfunktion  $a : E_a \rightarrow \mathbb{R}$  und  
 $G_v = (V, E_v)$  der Vereinigungsgraph mit Gewichtsfunktion  $v : E_v \rightarrow \mathbb{R}$ .  
 Gelten dann, wie in Abbildung 4.1 dargestellt,

$$\begin{aligned} (u_1, v_1) \in E_s : s(u_1, v_1) &= 1 \quad \wedge \quad (u_2, v_2) \in E_s : s(u_2, v_2) = 0 \quad \text{und} \\ (u_1, v_1) \in E_a : a(u_1, v_1) &= -3 \quad \wedge \quad (u_2, v_2) \in E_a : a(u_2, v_2) = -2, \end{aligned}$$

so folgt:

$$(u_1, v_1) \in E_v : v(u_1, v_1) = -2 \quad \text{und} \quad (u_2, v_2) \in E_v : v(u_2, v_2) = -2.$$

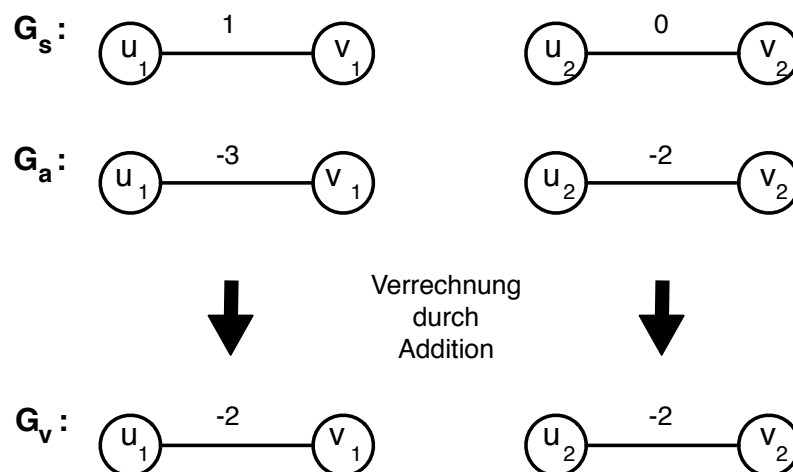


Abbildung 4.1: Ungünstiger Fall für die Vereinigung von Sympathiegraph  $G_s$  und Antipathiegraph  $G_a$ .

Aus der Verrechnung der Gewichte beider Kanten  $(u_1, v_1)$  und  $(u_2, v_2)$  resultieren also Kanten im Vereinigungsgraphen, deren Gewichte  $-2$  sind. Steht der Algorithmus für die Gruppeneinteilung nun vor der „Entscheidung“,  $(u_1, v_1)$  oder  $(u_2, v_2)$  zu schneiden, so hat er keinerlei Information darüber, dass eine von beiden Kanten aus einer Sympathie- und einer Antipathiebewertung zustande gekommen ist und die andere nur aus einer Antipathiebewertung. Möchte man diese Information also erhalten, so ist die Vereinigung von Sympathie- und Antipathiegraph nicht sinnvoll.

### 4.3 Erfüllung aller Antipathien?

Wie oben beschrieben ist eine Überlegung bei der Hinzunahme von Antipathien in die Bewertung, die Kanten des Antipathiegraphen für die Gruppeneinteilung stärker zu gewichten. Denn es wird angenommen, dass die Unzufriedenheit der Teilnehmer mit ihrer Gruppe bei Nichterfüllung der Antipathien größer ist als bei Nichterfüllung der Sympathien. Diese Idee geht sogar so weit, dass der Schnitt im Sympathiegraphen minimiert werden soll, unter der Nebenbedingung, dass im Antipathiegraphen alle Kanten Schnittkanten der Knotenpartition sind. Doch was bewirkt diese Erfüllung aller Antipathien bei den hier verwendeten bipartiten Knödel-Graphen?

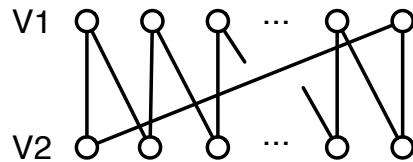


Abbildung 4.2: Ungünstiger Fall für einen Antipathiegraphen bei einer ausgesprochenen Antipathie pro Person.

Selbst, wenn man nur pro Person eine Antipathie aussprechen lässt, kann beim gewählten Kontaktgraphen im ungünstigen Fall der Antipathiegraph aus Abbildung 4.2 entstehen. Der Kreisgraph hat Chromatische Zahl 2, und die einzig mögliche 2-Färbung ist die Bipartition  $(V1, V2)$ , wobei die Knoten der Mengen  $V1$  und  $V2$  je verschiedene Farben haben.  $V1$  und  $V2$  bilden demnach die einzig mögliche Einteilung in zwei Gruppen, und es ist keine Verkleinerung des Schnittes im Sympathiegraphen mehr möglich. Jedoch enthält dann jede der beiden Gruppen ausschließlich Personen, die während der Gesprächsführung nicht miteinander gesprochen haben. So werden zwar alle Antipathien, jedoch keine einzige ausgesprochene Sympathie der Teilnehmer erfüllt. Sucht man mehr als zwei Gruppen oder lässt mehr als eine Antipathie zu, so wird eine vollständige Erfüllung der Antipathien in bestimmten Fällen unmöglich.

Unabhängig von der Anzahl Gruppen, der Anzahl ausgesprochener Antipathien und zudem von der Art des Kontaktgraphen besteht die Möglichkeit, dass ein Teilnehmer von all seinen Gesprächspartnern unsympathisch gefunden wird. Gibt es eine Gruppeneinteilung, die all diese Antipathien erfüllt, so befindet sich dieser Teilnehmer in einer Gruppe mit lauter Teilnehmern, die

ihm aus der Gesprächsführung fremd sind. Aus seiner Sicht war die Gruppeneinteilung auf Basis flüchtiger Kontakte somit erfolglos, und er kann sogar auf die ihm allseits entgegengebrachte Antipathie rückschließen.

## 4.4 Wahl der Bewertungsform

Um zu garantieren, dass pro Person eine gewisse Anzahl an ausgesprochenen Sympathie- und Antipathiebewertungen vorhanden ist, die dann bei der Gruppeneinteilung erfüllt werden können, sollten für diese eine Mindestzahlen vorgeschrieben werden. Setzt man diese Mindestzahlen für die Sympathien als auch für die Antipathien allerdings auf 1, so gibt es für die jeweilige Erfüllung keine „Alternativen“. Damit ist gemeint, dass dann entweder diese eine ausgesprochene Sym- bzw. Antipathie erfüllt wird, oder keine.

Betragen die Mindestzahlen 2, so besteht die Möglichkeit, dass entweder die eine oder die andere von beiden Sym- bzw. Antipathien erfüllt wird, wenn nicht beide zugleich erfüllt werden können. Je mehr „Alternativen“ vorhanden sind, um so größer ist die Wahrscheinlichkeit, dass bei der Gruppeneinteilung pro Person je eine der ausgesprochenen Sympathie- und Antipathiebewertungen erfüllt wird. Schreibt man feste Anzahlen  $p$  und  $n$  für die Aussprache von Sympathie- und Antipathiebewertungen vor, so legen diese eine untere Schranke für die Minimalgrade des Sympathie- und Antipathiegraphen fest. Klar ist, dass der gerichtete Sympathie- bzw. Antipathiegraph dann  $p$ - bzw.  $n$ -regulär ist. Durch oben beschriebene Konvertierung in die ungerichteten Graphen vergrößert sich der Grad mancher Knoten. Somit beträgt der Minimalgrad der ungerichteten Graphen mindestens  $p$  bzw.  $n$ .

Es besteht schließlich, wie oben beschrieben, die Möglichkeit, eine Skalierung der Sym- und Antipathien einzusetzen. Aber auch bei dieser Bewertungsform sollten Mindestzahlen für die Aussprache von Sympathie- und Antipathiebewertungen gewisser Stärken festgelegt werden. Denn spricht eine Person ausschließlich Bewertungen kleiner Stärken aus, alle anderen aber viele Bewertungen großer Stärke, so kann eine Gruppeneinteilung günstig sein, die keine dieser Bewertungen kleiner Stärken erfüllt. Da es allerdings recht viele Möglichkeiten gibt, Skalierungen und somit auch Mindestzahlen festzulegen, wird in dieser Arbeit nur die einfache Bewertungsform ohne Skalierung untersucht. Mit den dabei erzielten Ergebnissen können eventuell erste Vermutungen aufgestellt werden, welche Skalierungen und Mindestzahlen sinnvoll zu testen wären.



# Kapitel 5

## Einteilung der Gruppen

Das Verfahren, das bei der POLNET - Summer School 2003 zur Gruppeneinteilung verwendet wurde, hängt stark mit der Beschaffenheit des gewählten Kontaktgraphen zusammen: Gilt für die Anzahl Knoten  $n = k \cdot g$ , so sollen bei diesem Verfahren im Subgraphen des  $k$ -zyklischen Kontaktgraphen, dessen Basis eine  $g$ -Sektion ist,  $g$  „möglichst zusammenhängende“ Kreise der Länge  $k$  gefunden werden (siehe Kapitel 3). Hier liegt nun der bipartite Knödel-Graph als Kontaktgraph vor. Nach Aussprache der Sym- und Antipathien entstehen aus diesem zwei Subgraphen des Knödel-Graphen, der Sym- und der Antipathiegraph. Anhand dieser sollen die Teilnehmer so in  $g$  gleich große Gruppen eingeteilt werden, dass möglichst viele Sym- und Antipathien erfüllt sind. Wie in Abschnitt 1.2 definiert, bedeutet dabei die Erfüllung der Sym- und Antipathien, dass Personen, die sich sympathisch finden, in dieselbe Gruppe eingeteilt werden, und Personen, die sich nicht mögen, in verschiedene.

### 5.1 Problemstellung

Um eine günstige Einteilung der Personen in Gruppen zu finden, verwendet man folgende Idee: Man sucht nach einer Knotensektion, deren Schnitt im Sympathiegraphen minimal ist. Denn dann werden in jedem Sektionselement eine Menge von Sympathien erfüllt, deren Kantengewichtssumme auf dem gesamten Sympathiegraphen maximal ist. Wenn man die Gesamtheit der teilnehmenden Personen betrachtet, sollte dies für die Teilnehmer eine möglichst zufrieden stellende Gruppeneinteilung ergeben. Das Problem, eine minimale  $k$ -Sektion zu finden, ist jedoch  $\mathcal{NP}$ -vollständig (siehe Beweis in Abschnitt 5.3.2). Im Gegensatz zur Suche nach einer  $k$ -Partition minimalen Schnittes mit dem Algorithmus von M. Stoer und F. Wagner (siehe

[SW97]) ist hier nämlich die Größe der einzelnen Elemente fest vorgegeben. Es können also nur heuristische Methoden zur annähernden Lösung dieses Problems verwendet werden.

Berücksichtigt man bei der Gruppeneinteilung zudem die ausgesprochenen Antipathien, so ist Ziel der Sektionierung, dass die  $k$ -Sektion im Antipathiegraphen zusätzlich maximal sein soll. Dieses Problem liegt allerdings im Allgemeinen ebenfalls in  $\mathcal{NP}\text{-}C^1$ . Hier wird jedoch durch die Verwendung des Knödel-Graphen als Kontaktgraph ausschließlich auf bipartiten Graphen gearbeitet. Wie in Abschnitt 4.3 beschrieben, ist die dem Graphen zugrunde liegende Bisektion eine maximale Bisektion. Eine maximale  $k$ -Sektion mit geradem  $k$  kann folglich in Linearzeit gefunden werden, indem die beiden disjunkten Knotenmengen des bipartiten und  $r$ -regulären Graphen je in  $k/2$  gleich große Mengen geteilt werden. Bei der Suche nach einer  $k$ -Sektion mit ungeradem  $k$  greift die  $\mathcal{NP}$ -Vollständigkeit. Denn in diesem Fall ist ein Sektionselement zu finden, das je zur Hälfte aus Knoten der beiden Mengen des bipartiten Graphen besteht.

Die beschriebenen maximalen Sektionen im Antipathiegraphen sind jedoch nach Konstruktion auch maximale  $k$ -Sektionen im Sympathiegraphen. Umgekehrt gilt genauso, dass eine minimale Sektion im Sympathiegraphen keine maximale Sektion im Antipathiegraphen ist. Eine Sektion, die beide Kriterien zugleich erfüllt, gibt es folglich nicht. Demnach kann nur nach einer Sektion gesucht werden, die beide Kriterien annähernd erfüllt. Bezeichnet man den Schnitt einer Sektion im Sympathiegraphen als *Sympathieschnitt* und den im Antipathiegraphen als *Antipathieschnitt*, so hat die für die Gruppeneinteilung gesuchte Sektion minimalen Sympathie- und maximalen Antipathieschnitt zugleich.

### 5.1.1 Gewichtung

Eine derart zweigeteilte Optimierung schürt die Idee, darüber hinaus eine Gewichtung der beiden zu optimierenden Kriterien zu ermöglichen. Legt man mehr Wert auf die Minimierung der Sympathieschnitte als auf die Maximierung der Antipathieschnitte, so sollte das gewählte Verfahren zur Gruppeneinteilung dies ermöglichen. Das heißt, die Zielfunktion der Optimierungsverfahren sollte Variablen enthalten, mit denen entweder das Minimierungs- oder das Maximierungskriterium stärker gewichtet werden kann.

---

<sup>1</sup> (*Abk.*) Menge der  $\mathcal{NP}$ -vollständigen Probleme, von ‚ $\mathcal{NP}$ -complete‘ (*engl.*) heißt  $\mathcal{NP}$ -vollständig

### 5.1.2 Gruppenbalanciertheit

Betrachte man die Bipartitheit des verwendeten Kontaktgraphen: Die Teilnehmer werden für die Gesprächsführung beliebig in zwei gleich große Mengen  $A$  und  $B$  geteilt und unterhalten sich nur mit Personen aus der jeweils anderen Menge. Die aus den Sym- und Antipathien konstruierten Subgraphen des Kontaktgraphen sind demnach ebenfalls bipartit. Wie man eine Sektion maximalen Antipathieschnittes findet, wurde schon in Abschnitt 5.1 beschrieben. Betrachte man umgekehrt die Beschaffenheit einer Sektion mit minimalem Sympathieschnitt. Die Kanten, die nicht zu den Sektionskanten gehören, sind Kanten der von den Sektionselementen induzierten Graphen. Bestünden die Sektionselemente nur aus Knoten einer Menge des bipartiten Graphen, sei es  $A$  oder  $B$ , so enthalten die entsprechenden Subgraphen keine Kanten. Der Schnitt dieser Sektion entspricht dann also dem maximal Möglichen. Angenommen die Knoten eines Sektionselementes gehören zum einen Teil der Menge  $A$  und zum anderen der Menge  $B$  an, so induzieren sie in jedem Fall einen bipartiten Graphen. Dieser enthält im Allgemeinen mehr Kanten, wenn die Schnittmengen seiner Knotenmenge mit  $A$  und  $B$  gleich groß sind. Unter allen Teilmengen der Knotenmenge  $A \cup B$  des Kontaktgraphen scheint demnach eine Teilmenge mit gleich vielen Knoten aus  $A$  und  $B$  eine günstige Wahl für ein Sektionselement darzustellen, wenn der Schnitt der Sektion minimal sein soll. Beträgt nun das Verhältnis von Knoten der Menge  $A$  zu Knoten der Menge  $B$  in allen Elementen einer Sektion 1, so bezeichnet man die Sektion, sowie die einzelnen Sektionselemente als *gruppenbalanciert*.

Es wird nach obigen Erläuterungen vermutet, dass die Gruppenbalanciertheit der Sektionen bei der Minimierung des Schnittes im Sympathiegraphen erreicht wird. Dies wird in Kapitel 6 durch Experimente überprüft. Was passiert allerdings, wenn man diese Minimierung mit der Maximierung des Schnittes im Antipathiegraphen kombiniert? Die Elemente einer maximalen Sektion sind schließlich völlig unbalanciert. Auch hierzu werden Experimente gemacht.

### 5.1.3 Erfüllung der Sym- und Antipathien

Ziel der Gruppeneinteilung ist es, möglichst viele Sym- und Antipathien zu erfüllen. Damit aber auch jeder einzelne Teilnehmer mit seiner Gruppe zufrieden ist, sollte er möglichst viele Personen in seiner Gruppe sympathisch und möglichst wenige unsympathisch finden. Daher wird in Kapitel 6 untersucht, wie viele Sym- und Antipathien bei der Gruppeneinteilung pro Person erfüllt werden. Wird bei der Gruppeneinteilung keine einzige Bewertung eines Teilnehmers verwirklicht, so ist die Strategie aus dessen Sicht fehlgeschlagen.

Wie oben beschrieben, gibt es auf bipartiten Graphen zwar Sektionen, die zumindest je Teilnehmer eine Antipathie erfüllen, diese schneiden dann jedoch auch alle Kanten des Sympathiegraphen. Fraglich ist demnach, wie vielen Personen bei der gesuchten Gruppeneinteilung je mindestens eine Sympathie oder eine Antipathie erfüllt wird. Dieser Frage wird ebenfalls anhand von Experimenten nachgegangen.

## 5.2 Optimierungsverfahren

Will man in einer Grundmenge von Objekten Gruppen bilden, deren Elemente zusammengehören, denkt man als Erstes an Clusterung. Hierbei entstehen Partitionen, deren Kosten minimal unter denen aller möglichen Partitionen sind. Eine solche Kostenfunktion bei der Clusterung von Graphen könnte beispielsweise der Abstand sein, den Knoten voneinander haben. Clusterungsverfahren versuchen dann, die von den Knoten der einzelnen Partitionselemente induzierten Subgraphen in Bezug auf ihren Zusammenhang zu optimieren. Der Zusammenhang innerhalb der gebildeten Cluster soll also so groß wie möglich sein. Diese Eigenschaft wäre auch für die oben beschriebene Gruppeneinteilung sinnvoll. Die Clusterungsverfahren bilden allerdings meist beliebig viele Cluster, deren Größen unausgeglichen sein können. Die Forderung, dass eine bestimmte Anzahl gleich großer Gruppen gesucht ist, ist demnach für diese Verfahren zu stark.

Sind also Nebenbedingungen gegeben, wie Größe der Partition oder, dass die Partition balanciert sein soll, so sind gewöhnliche Partitionierungsverfahren besser geeignet. Sie legen den Fokus der Optimierung auf den entstehenden Schnitt der Partition und nicht unbedingt auf die induzierten Subgraphen. In bestimmten Fällen verursachen sie allerdings auch eine Einteilung nach Zusammengehörigkeit. Hat ein Graph signifikante Gruppen, so werden deren Kerne isoliert und die restlichen Knoten entsprechend verteilt. Das heißt, dass Knoten, die nach gewählter Kostenfunktion stärker zusammengehören als andere und deren Anzahl kleiner oder gleich der Größe der gesuchten Partitionselemente ist, zusammengefasst werden. Diese Anhäufungen von Knoten werden schließlich so vereint und alle übrigen Knoten so verteilt, dass die gesuchte Anzahl Partitionselemente entsprechender Größen entstehen und der Schnitt der Partition kostenoptimal ist. Würden die Knoten einer solchen signifikanten Gruppe auf verschiedene Partitionselemente verteilt, so hat die Partition einen Schnitt, der sehr wahrscheinlich durch entsprechenden Knotentausch optimiert werden kann.

### 5.3 $\mathcal{NP}$ -Vollständigkeiten

Wie schon erwähnt sind die Probleme, eine Sektion minimalen Schnittes und eine Partition maximalen Schnittes zu finden,  $\mathcal{NP}$ -vollständig. Dies gilt, da die entsprechenden Entscheidungsprobleme, ob es in einem Graphen eine Partition mit Schnitt  $\geq c$  oder eine Sektion mit Schnitt  $\leq c$  gibt, ebenfalls in  $\mathcal{NP}$ -C liegen. Beides wird im Folgenden für gewichtete, wie ungewichtete und eine entsprechende Teilung in zwei oder mehr Mengen gezeigt. Abbildung 5.1 gibt einen Überblick über die dafür geführten Reduktionen.

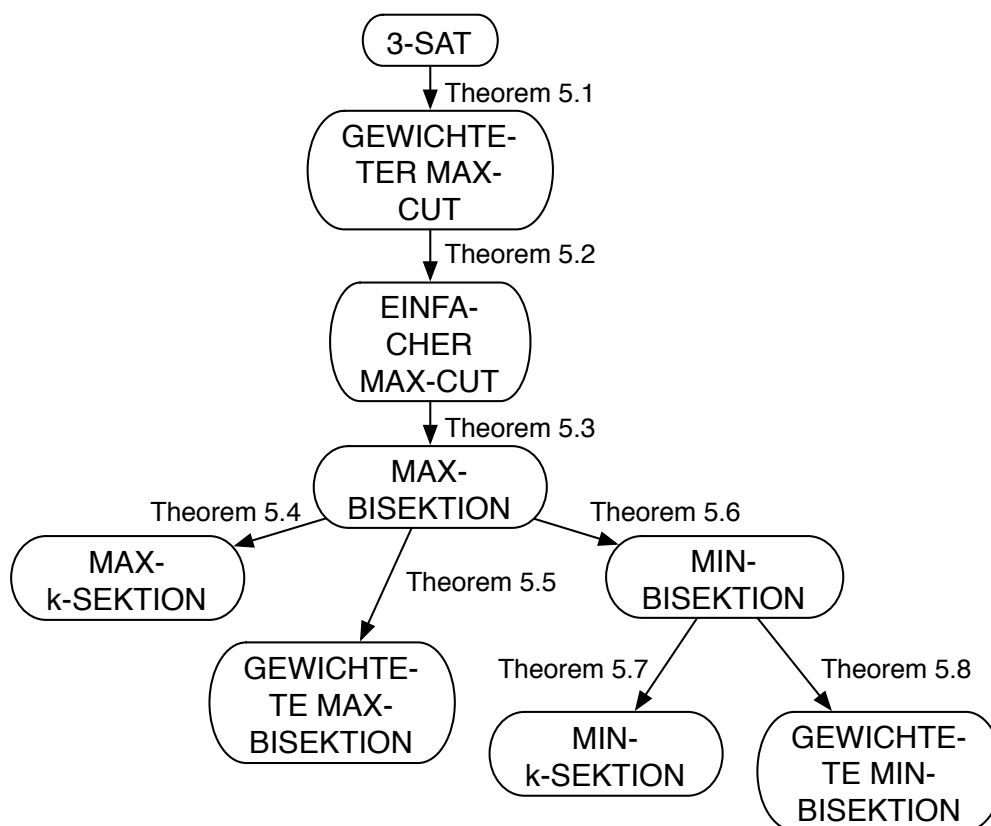


Abbildung 5.1: Überblick über die Reduktionen der  $\mathcal{NP}$ -Vollständigkeitsbeweise

Zum Beweis der  $\mathcal{NP}$ -Vollständigkeit muß abgesehen von der Reduktion eines geeigneten Problems aus  $\mathcal{NP}$ -C auf das betrachtete Problem noch gezeigt werden, dass es in  $\mathcal{NP}$  liegt. Dieser Beweis sieht allerdings bei den folgenden Entscheidungsproblemen immer gleich aus: Es kann je eine Par-

tition bzw. Sektion geraten, dessen Schnitt in polynomialer Zeit berechnet und mit der jeweiligen oberen bzw. unteren Schranke des Minimierungs- bzw. Maximierungsproblems verglichen wird. Bei den im Folgenden betrachteten Entscheidungsproblemen wird demnach nur noch die jeweilige Reduktion ausgeführt.

### 5.3.1 Maximale Schnitte

Das GEWICHTETER MAX-CUT-Problem ist wie folgt definiert:

**Problem 5.1** *GEWICHTETER MAX-CUT*

*Gegeben:* Ein gewichteter Graph  $G = (V, E)$

und eine positive ganze Zahl  $c$ .

*Gesucht:* Eine Bipartition  $(S, V - S)$ , so dass gilt:

$$\text{cut}(S, V - S) \geq c.$$

Wie im Folgenden gezeigt wird, liegt dieses Problem in  $\mathcal{NP}$ -C (siehe [Eve79], Seite 226ff).

**Satz 5.1** *Das GEWICHTETER MAX-CUT-Problem ist  $\mathcal{NP}$ -vollständig.*

**Beweis:** Es wird gezeigt, dass  $3\text{SAT} \propto \text{GEWICHTETER MAX-CUT}$ , also das 3SAT-Problem auf das GEWICHTETER MAX-CUT-Problem reduziert.

Seien  $C_1, C_2, \dots, C_m$  die Klauseln von 3SAT, jedes bestehend aus genau 3 Literalen. Die Menge der Literale sei  $L = \{x_1, x_2, \dots, x_n, \bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ . Man konstruiere die Eingabe des MAXIMUM-WEIGHTED-CUT-Problems bestehend aus  $G' = (V', E')$ , einer Gewichtsfunktion  $w'(e)$  und einer positiven Zahl  $c'$ , wie folgt:

$$V' = \{v_i | 0 \leq i \leq m\} \cup L.$$

Für jedes  $1 \leq i \leq m$  sei

$$A_i = \{v_0\} \cup \{v_i\} \cup C_i.$$

Damit sei

$$E' = \{(u, v) | u \neq v \text{ und } \exists i : u, v \in A_i\} \cup \{(x_j, \bar{x}_j) | 1 \leq j \leq n\}.$$

Die Gewichtsfunktion  $w$  wird definiert durch

$$\begin{aligned} w'(v_0, \xi) &= \sum_{i=1}^m |C_i \cap \{\xi\}|, \text{ wobei } \xi \in L, \\ w'(\xi', \xi'') &= \sum_{i=1}^m |C_i \cap \{\xi'\}| \cdot |C_i \cap \{\xi''\}|, \text{ wobei } \xi', \xi'' \in L \text{ und } \xi' \neq \xi'', \\ w'(x_j, \bar{x}_j) &= 10 \cdot m + 1, \\ w'(v_i, u) &= 1, \text{ wobei } i > 0 \text{ und } u \in A_i, \end{aligned}$$

und für  $c'$  gilt

$$c' = (10 \cdot m + 1)n + 6 \cdot m.$$

In dieser Konstruktion gilt  $|V'| = 2n + m + 1$ . Die Anzahl an Kanten,  $|E'|$ , ist begrenzt durch  $3n + 7m$ . Jede Klausel  $C_i$  wird also durch eine Clique  $A_i$  repräsentiert. Zusätzlich gibt es Kanten zwischen  $x_j$  und  $\bar{x}_j$ , deren Gewicht  $10 \cdot m + 1$  beträgt. (Dieses Gewicht ist so gewählt, dass es die Summe aller Kanten anderen Typs übersteigt.) Das Gewicht einer Kante  $(v_0, \xi)$  entspricht der Anzahl Vorkommen des Literals  $\xi$  in den Klauseln. Das Gewicht einer Kante  $(\xi', \xi'')$  entspricht der Anzahl Klauseln, in denen beide Literale auftauchen. Das Gewicht aller Kanten, die inzident zu  $v_i$ ,  $i > 0$ , sind, ist 1. Die Summe der Gewichte aller Kanten dieser drei Klassen ist genau  $10 \cdot m$ .

Es wird nun behauptet, dass die Lösung des 3SAT-Problems dieselbe ist, wie die Antwort auf die Frage: Gibt es eine Teilmenge  $S' \subseteq V'$ , so dass

$$\sum_{e \in (S', V-S')} w(e) \geq c'?$$

Angenommen das 3SAT-Problem ist lösbar, und  $\tau$  sei die Menge der Literale mit Wert ‚true‘ in einer widerspruchsfreien Belegung, welche alle Klauseln erfüllt. Sei  $\tau \subset S'$  und  $L - \tau \subset (V - S')$ . Natürlich gehört für jedes  $1 \leq j \leq n$ ,  $(x_j, \bar{x}_j)$  zum Schnitt und trägt  $10 \cdot m + 1$  zu dessen Gewicht bei. Hiermit beträgt das Gewicht des Schnittes schon  $(10 \cdot m + 1) \cdot n$ . Man füge nun  $v_0$  zu  $V - S'$  hinzu und betrachte die übrigen Kanten und deren Gewichte wie folgt: Jedes  $A_i$  ist eine Clique, und jede Kante taucht so oft auf, wie sie zu solchen Cliquen gehört, was genau ihrem definierten Gewicht entspricht. In jeder dieser  $m$  Cliquen ist mindestens ein Literal-Knoten, der zu  $S'$  gehört.  $v_0$  ist in  $V - S'$  und  $v_i$  kann so zu  $S'$  oder  $V - S'$  hinzugefügt werden, dass zwei der Knoten der  $A_i$ 's auf der einen Seite der Teilung sind und die restlichen drei auf der anderen. Folglich trägt die Clique 6 Kanten mit Gewicht 1 zum Schnitt bei, die  $m$  Cliquen zusammen also  $6 \cdot m$  zum Gewicht des Schnittes. Obige Argumentation zeigt auch, dass das gesamte Gewicht eines Schnittes  $c'$

nicht übersteigen kann. Wenn die Frage des GEWICHTETEN MAX-CUT-Problems bejaht werden kann, dann müssen alle Kanten vom Typ  $(x_j, \bar{x}_j)$  im Schnitt sein. Jede der  $m$  Cliques trägt dann 6 zum Gewicht des Schnittes bei, welches der maximale Wert ist, den eine solche Clique betragen kann. Folglich sind zwei Knoten einer jeden Clique auf der einen Seite des Schnittes und die übrigen drei auf der anderen Seite. Nennt man nun die Seite auf der  $v_0$  ist die ‚*false*‘-Seite und die andere die ‚*true*‘-Seite, dann folgt, dass höchstens zwei Literal-Knoten auf der ‚*false*‘-Seite sein können. Damit ist die so definierte Belegung widerspruchsfrei und erfüllt alle Klauseln.

Das 3SAT-Problem hat also genau dann eine widerspruchsfreie Belegung, welche alle Klauseln erfüllt, wenn  $G'$  einen Schnitt mit Gewicht größer oder gleich  $c'$  hat. Damit ist die Reduktion bewiesen.  $\square$

Im Folgenden wird auch der ungewichtete Fall benötigt. Daher wird gezeigt, dass das GEWICHTETER MAX-CUT-Problem  $\mathcal{NP}$ -vollständig bleibt, wenn alle Kantengewichte 1 sind. Das entsprechende Problem auf ungewichteten Graphen kann im Prinzip recht einfach bewiesen werden, wenn man in obiger Reduktion jede Kante mit Gewicht  $w$  durch  $w$  Kanten des Gewichtes 1 ersetzt. Sind dabei alle Kantengewichte polynomial beschränkt durch die Länge der Eingabe des 3SAT-Problems, so ist auch die erhöhte Kantenzahl polynomial beschränkt. Die Reduktion ist folglich in polynomialer Zeit möglich. Da hier jedoch auf einfachen Graphen gearbeitet wird, ist ein anderer  $\mathcal{NP}$ -Vollständigkeitsbeweis erforderlich.

Das Problem des maximalen Schnittes auf ungewichteten einfachen Graphen ist wie folgt definiert:

**Problem 5.2** *EINFACHER MAX-CUT*

*Gegeben:* Ein ungewichteter einfacher Graph  $G = (V, E)$   
und eine positive ganze Zahl  $c$ .

*Gesucht:* Eine Bipartition  $(S, V - S)$ , so dass gilt:  
 $cut(S, V - S) \geq c$ .

Es gilt also auch hier (siehe [Eve79], Seite 228ff):

**Satz 5.2** *Das EINFACHER MAX-CUT-Problem ist  $\mathcal{NP}$ -vollständig.*

**Beweis:** Es wird gezeigt, dass GEWICHTETES MAX-CUT  $\propto$  EINFACHER MAX-CUT, also das GEWICHTETES MAX-CUT-Problem auf das EINFACHER MAX-CUT-Problem reduziert.

Man kann ein Problem in einem gewichteten Graphen ganz einfach auf ein Problem in einem ungewichteten Graphen reduzieren, indem man jede Kante



mit Gewicht  $w$  durch  $w$  parallele Kanten mit Gewicht 1 ersetzt. Denn wenn alle Gewichte polynomial durch die Eingabelänge begrenzt sind, ist dies auch für die Erhöhung der Anzahl Kanten der Fall. Allerdings erhält man auf diese Art und Weise einen Graphen mit Mehrfachkanten. Da allerdings der Graph im EINFACHER MAX-CUT-Problem einfach sein muß, ist eine Reduktion auf einen Graphen ohne parallele Kanten erforderlich.

Angenommen die Kantengewichte eines Graphen  $G = (V, E)$  sind polynomial begrenzt durch die Länge der Eingabe des GEWICHTETEN MAXIMUM-CUT-Problems, dann ersetze man jede Kante  $e = (u, v) \in E$  mit Gewicht  $w(e)$  durch den Subgraphen aus Abbildung 5.2. Dabei sind die Knoten  $a_1, a_2, \dots, a_{w(e)}$  und  $b_1, b_2, \dots, b_{w(e)}$  neu und der entstandene Graph wird mit  $G'$  bezeichnet.

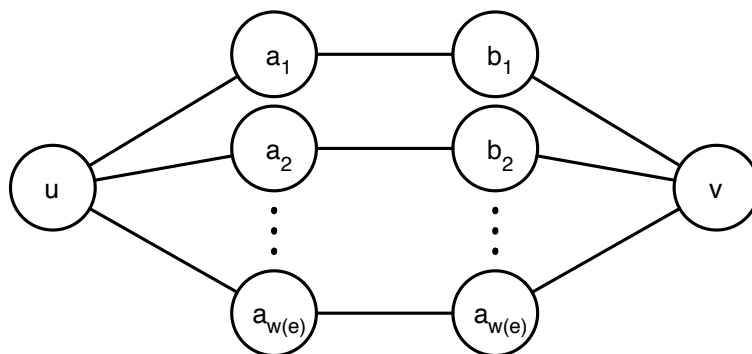


Abbildung 5.2: Subgraph, der die Kante  $e = (u, v)$  mit Gewicht  $w(e)$  ersetzt.

Es wird nun behauptet, dass  $G'$  genau dann einen Schnitt der Größe  $k' = 2 \cdot \sum_{e \in E} w(e) + k$  besitzt, wenn  $G$  einen gewichteten Schnitt der Größe  $k$  besitzt.

Wenn  $u$  und  $v$  auf derselben Seite des Schnittes liegen, können aus jedem Weg  $(u, a_i, b_i, v)$  maximal zwei Kanten Schnittkanten sein. Sind die beiden Knoten auf verschiedenen Seiten, und gibt man  $b_i$  auf die Seite von  $u$  und  $a_i$  auf die Seite von  $v$ , so trägt der Weg drei Kanten zum Schnitt bei. Somit kann jede Kante  $e$  des Graphen  $G$  eine Erhöhung des Schnittes im Graphen  $G'$  von mindestens  $2 \cdot w(e)$  Kanten hervorrufen, selbst wenn die Endknoten durch diesen Schnitt nicht getrennt werden. Insgesamt für alle Kanten des Graphen  $G$  ergibt dies einen Beitrag von  $2 \cdot \sum_{e \in E} w(e)$  Kanten zum Schnitt in  $G'$ .

Angenommen das Gewicht des Schnittes  $S, V - S$  in  $G$  beträgt  $k$ . Seien dann  $S \subset S'$  und  $V - S \subset (V - S')$ . Sind  $u$  und  $v$  auf derselben Seite des Schnittes,

so werden die Knoten  $a_i$  und  $b_i$  so verteilt, dass der Weg  $(u, a_i, b_i, v)$  2 zum Schnitt beiträgt. Sind sie auf verschiedenen Seiten, so werden die Knoten dieses Weges so zugeordnet, dass der Weg einen Beitrag von 3 Kanten leistet. Damit berechnet sich das Gewicht des Schnittes in  $G'$  wie folgt:

$$\text{cut}(S', V - S') = 2 \cdot \sum_{e \in E} w(e) + k = k'$$

Umgekehrt wird angenommen, dass der Schnitt  $(S', V - S')$  im Graphen  $G'$  mindestens aus  $k'$  Schnittkanten besteht. Von diesen sind höchstens  $2 \cdot w(e)$  Kanten der Beitrag einer Kante  $e = (u, v)$  in  $G$  mit  $u$  und  $v$  auf der gleichen Seite des Schnittes. Ohne Beschränkung der Allgemeinheit oder durch Anpassung von  $S'$  kann angenommen werden, dass dies genau der Beitrag des Subgraphen einer jeden Kante ist. Ebenso gilt, dass der Beitrag, den der Subgraph einer Kante  $(u, v)$  mit  $u$  und  $v$  auf verschiedenen Seiten des Schnittes leistet, genau  $3 \cdot w(e)$  beträgt. Somit gilt:

$$|(S', V - S')| = 2 \cdot \sum_{e \in E} w(e) + \sum_{e \in (S, V - S)} w(e),$$

wobei  $S = S' \cap V$ . Da  $|(S', V - S')| = k'$  gilt, erhält man:

$$k = \sum_{e \in (S, V - S)} w(e).$$

Somit hat  $G$  genau dann einen Schnitt der Größe  $k$ , wenn  $G'$  einen Schnitt der Größe  $k'$  besitzt. Damit ist die Reduktion bewiesen.  $\square$

Wird auf einfachen Graphen statt einer Bipartition eine Bisektion gesucht, so ergibt sich folgendes Problem, das ebenfalls  $\mathcal{NP}$ -vollständig ist.

### Problem 5.3 MAX-BISEKTION

*Gegeben:* Ein ungewichteter Graph  $G = (V, E)$   
und eine positive ganze Zahl  $c$ .

*Gesucht:* Eine Bisektion  $(S, V - S)$ , so dass gilt:  
 $\text{cut}(S, V - S) \geq c$ .

**Satz 5.3** Das MAX-BISEKTION-Problem ist  $\mathcal{NP}$ -vollständig.

**Beweis:** Es wird EINFACHER MAX-CUT  $\propto$  MAX-BISEKTION gezeigt.

Dazu werden zum Graph  $G$   $|V|$  isolierter Knoten hinzugefügt. Der erhaltene Graph  $G'$  enthält also  $2 \cdot |V|$  Knoten. Da die neuen Knoten keine adjazenten

Kanten besitzen, verändert sich die Größe eines Schnittes in  $G$  nicht, egal wie sie den Mengen der zugehörigen Partition zugeteilt werden. Es kann also jeder Schnitt  $(S, V - S)$  in  $G$  in eine Bisektion in  $G'$  umgewandelt werden, indem die hinzugefügten Knoten entsprechend auf  $S$  und  $V - S$  verteilt werden. Somit hat  $G$  einen Schnitt der Größe  $c$  genau dann, wenn  $G'$  eine Bisektion der Größe  $c$  besitzt.  $\square$

Sucht man nun eine  $k$ -Sektion mit  $k > 2$  statt einer Bisektion, d.h. eine Einteilung der Knoten in mehr als zwei gleich große Gruppen, so ist das entsprechende Problem wie folgt definiert:

**Problem 5.4** *MAX- $k$ -SEKTION*

*Gegeben:* Ein ungewichteter Graph  $G = (V, E)$   
und eine positive ganze Zahl  $c$ .

*Gesucht:* Eine  $k$ -Sektion  $S(G) = \{S_1, S_2, \dots, S_k\}$ , so dass gilt:  
 $cut(S) \geq c$ .

**Satz 5.4** *Das MAX- $k$ -SEKTION-Problem ist  $\mathcal{NP}$ -vollständig.*

**Beweis:** Hier ist es nicht notwendig eine Reduktion auszuführen. Der Satz kann durch einen Widerspruchsbeweis gezeigt werden:

Wäre das MAX- $k$ -SEKTION-Problem in  $\mathcal{P}$ , so würde dies auch für das MAX-BISEKTION-Problem als dessen Spezialfall mit  $k = 2$  gelten. Die  $\mathcal{NP}$ -Vollständigkeit des letzten Problems wurde aber in Satz 5.3 bewiesen. Somit muss das MAX- $k$ -SEKTION-Problem ebenfalls  $\mathcal{NP}$ -vollständig sein.  $\square$

Im Folgenden kann  $\mathcal{NP}$ -Vollständigkeit noch öfter durch Widerspruch gezeigt werden. Das Prinzip ist immer wie im Beweis zu Satz 5.4. Es wird daher in den entsprechenden Beweisen nur noch der Spezialfall des betrachteten Problems angegeben, deren  $\mathcal{NP}$ -Vollständigkeit bereits gezeigt wurde. In obigem Beweis der  $\mathcal{NP}$ -Vollständigkeit des MAX- $k$ -SEKTION-Problems ist der Spezialfall das MAX-BISEKTION-Problem.

Mit einem solchen Widerspruchsbeweis kann man auch die  $\mathcal{NP}$ -Vollständigkeit der Suche nach einer Bisektion maximalen Schnittes in einem gewichteten Graphen zeigen. Das Problem ist folgendermaßen definiert:

**Problem 5.5** *GEWICHTETE MAX-BISEKTION*

*Gegeben:* Ein gewichteter Graph  $G = (V, E)$   
und eine positive ganze Zahl  $c$ .

*Gesucht:* Eine Bisektion  $(S, V - S)$ , so dass gilt:  
 $cut(S, V - S) \geq c$ .

**Satz 5.5** *Das GEWICHTETE MAX-BISEKTION-Problem ist  $\mathcal{NP}$ -vollständig.*

**Beweis:** Das MAX-BISEKTION-Problem ist ein Spezialfall des GEWICHTETE MAX-BISEKTION-Problems mit Gewichtsfunktion  $w : E \rightarrow 1$ .  $\square$

Analog lässt sich auch das GEWICHTETE MAX- $k$ -SEKTION-Problem definieren und deren  $\mathcal{NP}$ -Vollständigkeit beweisen. Da dieses Problem allerdings im Folgenden nicht explizit betrachtet wird, wird es hier nicht notiert.

### 5.3.2 Minimale Schnitte

Das MIN-BISEKTION-Problem ist wie folgt definiert:

**Problem 5.6** *MIN-BISEKTION*

*Gegeben: Ein ungewichteter Graph  $G = (V, E)$   
und eine positive ganze Zahl  $c$ .*

*Gesucht: Eine Bisektion  $(S, V - S)$  von  $G$ , so dass gilt:  
 $cut(s, V - S) \leq c$ .*

**Satz 5.6** *Das MIN-BISEKTION-Problem ist  $\mathcal{NP}$ -vollständig.*

**Beweis:** Es wird  $\text{MAX-BISEKTION} \propto \text{MIN-BISEKTION}$  gezeigt.  
Ein Graph  $G = (V, E)$  mit  $|V| = 2 \cdot n$  besitzt genau dann eine Bisektion  $(S, V - S)$  mit  $cut(S, V - S) = c$ , wenn sein Komplement  $\bar{G} = (V, \bar{E})$  eine Bisektion  $(S, V - S)$  mit  $cut(S, V - S) = n^2 - c$  besitzt. Dann gibt es in  $G$  aber auch genau dann eine Bisektion mit Schnitt  $\geq c$ , wenn es in  $\bar{G}$  eine Bisektion mit Schnitt  $\leq n^2 - c$  gibt.  $\square$

Sucht man nun eine  $k$ -Sektion statt einer Bisektion, d.h. eine Einteilung der Knoten in mehr als zwei gleich große Gruppen, so ist das entsprechende Problem wie folgt definiert:

**Problem 5.7** *MIN- $k$ -SEKTION*

*Gegeben: Ein ungewichteter Graph  $G = (V, E)$   
und eine positive ganze Zahl  $c$ .*

*Gesucht: Eine  $k$ -Sektion  $S(G) = \{S_1, S_2, \dots, S_k\}$ , so dass gilt:  
 $cut(S) \leq c$ .*

**Satz 5.7** *Das MIN- $k$ -SEKTION-Problem ist  $\mathcal{NP}$ -vollständig.*

**Beweis:** Das MIN-BISEKTION-Problem ist ein Spezialfall des MIN- $k$ -SEKTION-Problems mit  $k = 2$ .  $\square$

Das GEWICHTETE MIN-BISEKTION-Problem ist wie folgt definiert:

**Problem 5.8** *GEWICHTETE MIN-BISEKTION*

*Gegeben:* Ein gewichteter Graph  $G = (V, E)$   
und eine positive ganze Zahl  $c$ .

*Gesucht:* Eine Bisektion  $(S, V - S)$  von  $G$ , so dass gilt:  
 $cut(S, V - S) \leq c$ .

**Satz 5.8** *Das GEWICHTETE MIN-BISEKTION-Problem ist  $\mathcal{NP}$ -vollständig.*

**Beweis:** Das MIN-BISEKTION-Problem ist ein Spezialfall des GEWICHTETE MIN-BISEKTION-Problems mit Gewichtsfunktion  $w : E \rightarrow 1$ .  $\square$

Analog lässt sich auch wieder das GEWICHTETE MIN- $k$ -SEKTION-Problem definieren und deren  $\mathcal{NP}$ -Vollständigkeit beweisen. Da dieses Problem ebenfalls im Folgenden nicht verwendet wird, wird es hier nicht notiert.

## 5.4 Minimale Bisektionierung

Zur annähernden Lösung des MIN-BISEKTION-Problems gibt es verschiedene Heuristiken, die eine Bisektion möglichst minimalen Schnittes liefern. Da für die hier zu entwickelnde Strategie allerdings neben der Optimierung des minimalen Schnittes der Sympathien gleichzeitig der Schnitt der Antipathien maximiert werden soll, fallen Näherungsverfahren, die beispielsweise auf der Spektralmethode oder der Lösung anderer graphentheoretischer Probleme basieren, weg (siehe Abschnitt 5.4.2). Nur die bewegungsbasierten oder die daraus hergeleiteten clusterbasierten Methoden sind unabhängig von der Art des Optimierungskriteriums und können somit so verändert werden, dass sie auch den maximalen Schnitt optimieren. Aus diesem Grund wird hier der Schwerpunkt auf die bewegungsbasierten Methoden gelegt und die alternativen Heuristiken nur kurz beschrieben.

### 5.4.1 Bewegungsbasierte Methoden

Eine Methode zur Partitionierung wird *bewegungsbasiert* genannt, wenn unter Sortierung der möglichen *lokalen Veränderungen* und Speicherung vorheriger

Operationen iterativ neue Lösungskandidaten konstruiert werden. Alle bewegungsbasierten Methoden starten dabei mit einer gegebenen Partition, der sogenannten *Startpartition* und konstruieren aus ihr bessere Partitionen. Die Iteration stoppt, wenn ein lokales Optimum erreicht ist; das heißt, wenn alle möglichen lokalen Veränderungen einer Partition schlechter sind als diese.

Die lokalen Veränderungen sind all diejenigen Partitionen, die vom sogenannten *Nachbarschaftsoperator* gebildet werden können. Für die Partitionierung typisch sind Operatoren, die zwei Knoten verschiedener Partitionselemente miteinander tauschen oder einen einzelnen Knoten von einem in ein anderes Element verschieben. Der Lösungsraum der lokalen Veränderungen wird schließlich danach sortiert, um wieviel die Partitionen besser sind als die Ausgangspartition. Die Partition, die jeweils den maximalen Gewinn bringt, wird weiter bearbeitet. Bei einfachen Greedy-Methoden wird die nächst bessere Lösung je nur aus der aktuellen gebildet. Im Gegensatz dazu werden bei bewegungsbasierten Methoden mehrere vorherige Partitionen gespeichert und die jeweils beste von ihnen weiter bearbeitet. Beim im Folgenden beschriebenen *Kernighan-Lin-Algorithmus* werden beispielsweise die Ergebnisse eines Iterationsschrittes gespeichert und die beste Bisektion von allen zur weiteren Iteration verwendet.

Bewegungsbasierte Methoden sind unabhängig von der Art der Zielfunktion, die benutzt wird, um die Qualität der Lösung zu messen. Es können somit beliebige Optimierungskriterien in einer Zielfunktion zusammengefasst werden, wie für eine kombinierte Optimierung des Sympathie- und Antipathieschnittes notwendig. Aber auch aus weiteren Gründen dominieren die bewegungsbasierten Methoden in Literatur und Praxis: Es ist zum einen sehr intuitiv, eine gegebene Lösung durch wiederholte kleine Veränderungen durch den Nachbarschaftsoperator zu verbessern, und zum anderen sind die Operationen der verwendeten Operatoren leicht zu beschreiben und implementieren.

Der hier vorgestellte bewegungsbasierte Algorithmus wurde 1970 von B. W. Kernighan und S. Lin eingeführt (siehe [KL70]) und wird häufig als die erste „gute“ Heuristik zur minimalen Bisektionierung bezeichnet. Da die zu entwickelnde Strategie ausschließlich auf gewichteten Graphen arbeitet, wird hier direkt der Kernighan-Lin-Algorithmus zur gewichtete Bisektionierung vorgestellt. Im ungewichteten Fall werden wieder alle Kantengewichte gleich 1 gesetzt. Der zu bearbeitende Graph wird hier durch seine Adjazenzmatrix  $W = (w_{ij})_{0 \leq i, j \leq n-1}$  gegeben. Es gilt:

$$w_{ij} = \begin{cases} w(e) & \text{mit } e = (v_i, v_j) \in E, \\ 0 & \text{sonst.} \end{cases}$$

### Kernighan-Lin-Algorithmus

Als Nachbarschaftsoperator dieser bewegungsbasierten Methode wird der Tausch zweier Knoten aus verschiedenen Elementen der Bisektion verwendet, wobei allerdings in jedem Iterationsschritt jeder Knoten nur ein einziges Mal bewegt wird. Um dies zu gewährleisten, werden die Knoten markiert, nachdem sie an einer Tauschoperation beteiligt waren. Im Kernighan-Lin-Algorithmus werden jeweils die Knoten getauscht, deren Tausch den größten *Gewinn* erzielt. Dabei bezeichnet der Gewinn  $G = \text{cut}(P) - \text{cut}(P')$  den Wert, um den sich der Schnitt beim Übergang von Bisektion  $P$  zu  $P'$  verkleinert. Um diesen Gewinn zu ermitteln, werden die sogenannten ‚dValues‘<sup>2</sup> der Knoten berechnet, welche die Differenz zwischen den nachstehend definierten *externen* und *internen Kosten* bezeichnen.

Betrachtet man eine gegebene Bisektion  $(A, B)$ , so beschreiben die externen Kosten  $E_a$  des Knotens  $a \in A$  die Summe der Gewichte seiner inzidenten Kanten, die ihn mit Knoten der Menge  $B$  verbinden. Die internen Kosten  $I_a$  sind entsprechend die Summe aller Gewichte von Kanten, die im von  $A$  induzierten Subgraphen liegen. Es gilt somit:

$$E_a = \sum_{y \in B} w_{ay} \quad \text{und} \quad I_a = \sum_{x \in A} w_{ax}.$$

$E_b$  und  $I_b$  für  $b \in B$  sind analog definiert. Aus diesen Werten berechnet man die ‚dValues‘  $D_a := E_a - I_a$  der Knoten  $a \in A$ .

Es gilt dann folgender Hilfssatz (siehe [KL70], Seite 296):

**Hilfssatz 5.1** *Seien  $a \in A$  und  $b \in B$  zwei Knoten aus verschiedenen Elementen einer Knotenbisektion  $P = (A, B)$  des Graphen  $G = (V, E)$ . Wenn man  $a$  und  $b$  tauscht, das heißt die Bisektion  $P' = (A - \{a\} \cup \{b\}, B - \{b\} \cup \{a\})$  bildet, gilt:*

$$\text{cut}(P) - \text{cut}(P') = D_a + D_b - 2 \cdot w_{ab}.$$

**Beweis:** Sei  $z$  die Summe der Gewichte aller Schnittkanten der Bisektion  $P$  ohne diejenigen Kanten, die adjazent zu  $a$  oder  $b$  sind. Dann gilt:

$$\text{cut}(P) = z + E_a + E_b - w_{ab}.$$

Tauscht man nun  $a$  und  $b$  und erhält dadurch die Bisektion  $P'$ , so gilt:

$$\text{cut}(P') = z + I_a + I_b + w_{ab}.$$

---

<sup>2</sup> ‚difference value‘ (engl.) Differenzwert

Daraus folgt:

$$\text{cut}(P) - \text{cut}(P') = D_a + D_b - 2 \cdot w_{ab}.$$

□

Ein solcher Gewinn  $\text{cut}(P) - \text{cut}(P')$ , den die Bipartition  $P'$  gegenüber  $P$  bringt, wird auch als  $\text{gain}(a, b)$  des Knotenpaares  $a$  und  $b$  bezeichnet. Dieser lässt sich somit einfach aus den ‚dValues‘ berechnen.

Sei nun für einen Iterationsschritt des Kernighan-Lin-Algorithmus eine Knotenbisektion  $P = (A, B)$  des zu bearbeitenden Graphen gegeben. Dann werden zuerst die ‚dValues‘ seiner Knoten berechnet und Knoten  $a_i \in A$  und  $b_i \in B$  ermittelt, für die der Wert  $G_i = \text{gain}(a_i, b_i) = D_{a_i} + D_{b_i} - 2 \cdot w_{a_i b_i}$  maximal ist. Das entsprechende Knotenpaar  $a, b$  wird schließlich markiert, aber noch nicht getauscht. Um nun auf der Bisektion  $(A - \{a_i\}, B - \{b_i\})$  aller nicht markierten Knoten weiterarbeiten zu können, werden die ‚dValues‘ so aktualisiert, als ob  $a_i$  und  $b_i$  vertauscht worden wären. Die neuen ‚dValues‘  $D'_x$  und  $D'_y$  lassen sich wie folgt aus den alten Werten  $D_x$  und  $D_y$  der Knoten  $x \in A - \{a_i\}$  und  $y \in B - \{b_i\}$  berechnen:

$$\begin{aligned} D'_x &:= D_x + 2 \cdot w_{xa_i} - 2 \cdot w_{xb_i} \quad \text{und} \\ D'_y &:= D_y + 2 \cdot w_{yb_i} - 2 \cdot w_{ya_i}. \end{aligned}$$

Dies ist leicht nachzuvollziehen: Da die Kante  $(x, a_i)$  bei  $D_x$  zu den internen Kosten gezählt wird und in  $D'_x$  den externen Kosten angerechnet werden soll, wird bei der Aktualisierung der Wert  $2 \cdot w_{xa_i}$  addiert. Analog wird der Wert  $2 \cdot w_{xb_i}$  subtrahiert, um das Gewicht der Kante  $(x, b_i)$  von den externen zu den internen Kosten zu überführen.

Die Wahl zweier Knoten  $a_{i+1}$  und  $b_{i+1}$  mit maximalem Wert  $G_{i+1} = \text{gain}(a_{i+1}, b_{i+1})$  und die Aktualisierung der ‚dValues‘ werden nun auf der Bisektion  $(A - \{a_i\}, B - \{b_i\})$  aller nicht markierten Knoten wiederholt. Sind schließlich alle Knoten markiert, so wird die Zahl  $j$  so bestimmt, dass die Summe  $G := \sum_{i=1, \dots, j} G_i$  maximal ist. Gilt  $G > 0$ , so liefert der Tausch aller Knoten  $a_i$  und  $b_i$  für  $0 \leq i \leq j$  die Bisektion  $P' = (A', B')$  mit  $A' := A - \{a_1, \dots, a_j\} \cup \{b_1, \dots, b_j\}$  und  $B' := B - \{b_1, \dots, b_j\} \cup \{a_1, \dots, a_j\}$ , für deren Schnitt  $\text{cut}(P') = \text{cut}(P) - G$  gilt. Diese Bisektion  $P'$  wird schließlich als Eingabe für den nächsten Iterationsschritt verwendet. Dies geschieht so lange, bis für eine Bisektion ein Gewinn  $G \leq 0$  erzielt wird. Dann ist ein lokales Minimum erreicht, das heißt alle möglichen lokalen Veränderungen bringen eine Verschlechterung der Bisektion. Diese *lokal minimale* Bisektion ist schließlich die Ausgabe des Algorithmus. Abbildung 5.3 zeigt den Kernighan-Lin-Algorithmus in Pseudocode.



**Algorithmus:**


---

 Minimale Bisektionierung (nach B. W. Kernighan und S. Lin)
 

---

Eingabe: Startbisektion  $(A, B)$ .Ausgabe: Lokal minimale Bisektion  $(A, B)$ .

- 1: Wiederhole...
  - 2: Berechne die *dValues* der Knoten.
  - 3: Setze  $i=0$ .
  - 4: Solange nicht alle Knoten markiert...
  - 5: Suche ein unmarkiertes Knotenpaar  $a_i \in A$  und  $b_i \in B$ , deren Gewinn  $G_i := \text{gain}(a_i, b_i)$  maximal ist.
  - 6: Markiere  $a_i$  und  $b_i$ .
  - 7: Aktualisiere die *dValues* der nicht markierten Knoten, als ob  $a_i$  und  $b_i$  vertauscht worden wären.
  - 8: Setze  $i=i+1$ .
  - 9: Bestimme  $j$  so, dass  $G = \sum_{i=1, \dots, j} G_i$  maximal ist.
  - 10: Wenn  $G > 0$ , dann...
  - 11: Aktualisiere  $A = A - \{a_1, \dots, a_j\} \cup \{b_1, \dots, b_j\}$  und  $B = B - \{b_1, \dots, b_j\} \cup \{a_1, \dots, a_j\}$ .
  - 12: Solange bis  $G \leq 0$ .
- 

Abbildung 5.3: Kernighan-Lin-Algorithmus

**Laufzeitanalyse**

Für die Laufzeitanalyse beim Kernighan-Lin-Algorithmus wird zuerst der einzelne Iterationsschritt betrachtet. Für die Berechnung der ‚dValues‘ von  $n$  Knoten (siehe Abbildung 5.3, Schritt 2) werden  $n^2$  Arbeitsschritte benötigt, da hier für jeden Knoten alle anderen Knoten in Betracht gezogen werden müssen. Die in Schritt 4 beginnende Schleife wird  $n/2$  mal durchlaufen, da bei jedem Durchlauf zwei Knoten markiert werden. Die Laufzeit der Schleife liegt in  $\mathcal{O}(n^2)$ . Bezeichne  $m \leq n$  die Anzahl nicht markierter Knoten, so werden nämlich  $m^2$  Arbeitsschritte für die Suche nach dem nächsten Knotenpaar (Schritt 5) und  $m$  für die Aktualisierung der ‚dValues‘ (Schritt 7) benötigt. Da die Laufzeiten aller übrigen Schritte des Algorithmus in  $\mathcal{O}(n)$  oder  $\mathcal{O}(1)$  liegen, berechnet sich die Gesamtlaufzeit eines Iterationsschrittes auf  $n^3$ .

Die nötige Zahl an Iterationsschritten wurde von B. W. Kernighan und

S. Lin experimentell bestimmt (siehe [KL70], S. 300). Sie machten empirische Versuche auf Graphen der Ordnung  $\leq 360$ . Fast immer wurde nach 2 bis 4 Iterationen ein lokales Minimum erreicht. Aufgrund dieses Ergebnisses scheint keine strenge Abhängigkeit von der Knotenzahl vorzuliegen. Da hier jedoch auf Subgraphen des Knödel-Graphen gearbeitet wird, ist eine spezifische Untersuchung der Iteration notwendig. Die Beschreibung eines entsprechenden Experimentes ist in Abschnitt 6.4.3 zu finden. Dieses ergibt, dass auf Graphen der Ordnung  $\leq 100$  bei Verwendung bestimmter Startbisektionen sogar weniger Iterationsschritte notwendig sind.

Auf den hier zu bearbeitenden Graphen der Ordnung 20 bis 100 ist ein  $\mathcal{O}(n^3)$ -Algorithmus annehmbar. Für größere Graphen sind zur Bisektionierung, wie auch zur  $k$ -Sektionierung die *METIS-Algorithmen*<sup>3</sup> besser geeignet. Sie bilden eine Familie von Partitionierungsalgorithmen für unstrukturierte Graphen und Hypergraphen. Beim sogenannten ‚multilevel graph partitioning scheme‘ werden die Graphen zum Beispiel verkleinert, in diesem Zustand partitioniert und die erhaltene Partition dann wieder in den ursprünglichen Graphen übertragen. Sollen die Partitionselemente gleich groß sein, so ist eine spezielle Verkleinerung und Übertragung in den ursprünglichen Graphen notwendig. Der entsprechende Algorithmus zur Bisektionierung ist in [KK95] beschrieben.

### Startbisektionen

Der Kernighan-Lin-Algorithmus fordert die Eingabe einer Startbisektion. Diese kann zufällig gewählt werden, doch nach B. W. Kernighan und S. Lin reduziert eine gezielte Wahl die Laufzeit. Es können beispielsweise Clusterungsverfahren (siehe Abschnitt 5.4.2) eingesetzt werden, um Startbisektionen zu erhalten, die zumindest keine stark zusammenhängenden Cluster trennen. Wird der Algorithmus allerdings wie hier auf bipartite Graphen angewendet, so können gezielt gewissermaßen beliebige, gute Startbisektionen angegeben werden: Wählt man nämlich eine gruppenbalancierte Bisektion, so ist deren Schnitt auf allen bipartiten Graphen  $G = (V_1 \cup V_2, E)$  kleiner als der Schnitt der Bisektion  $(V_1, V_2)$  (siehe Abschnitt 5.1.2). Zum Vergleich werden in Kapitel 6 zusätzlich auch die Ergebnisse unter Verwendung der Bisektion  $(V_1, V_2)$  betrachtet.

### ‚Hill Climb‘

Am Beispiel des Kernighan-Lin-Algorithmus kann nun demonstriert werden, dass im Gegensatz zu herkömmlichen Greedy-Verfahren bei bewegungsba-

<sup>3</sup> siehe <http://www-users.cs.umn.edu/~karypis/metis>

sierten Methoden eine Art ‚*Hill Climb*‘<sup>4</sup> aus einem lokalen Minimum möglich ist. Dabei bezeichnet ein ‚Hill Climb‘ eine Verschlechterung der Partition, die letztendlich eine Verbesserung ermöglicht.

Betrachtet man einen Iterationsschritt, so werden nacheinander die den größten Gewinn bringenden Tauschoperationen der noch nicht markierten Knoten betrachtet. Dabei werden das jeweils  $i$ -te zu tauschende Knotenpaar und deren Gewinn  $G_i$  gespeichert, die Knoten markiert, aber noch nicht getauscht. Sind alle Knoten markiert, wird der über alle betrachteten Tauschoperationen mögliche Gewinn  $G = \sum_{i=1, \dots, j} G_i$  maximiert. Einige der Werte  $G_i$  sind negativ, falls nicht  $G_i = 0$  für alle  $0 \leq i \leq n/2 - 1$  gilt. Dies folgt daraus, dass  $\sum_{i=1, \dots, n/2} G_i = 0$  gilt. Denn tauscht man alle  $j$  Knotenpaare, so erhält man die ursprüngliche Bipartition  $P = (A, B)$ . Sind nun einer oder mehrere Summanden  $G_i$  für  $i < j$  gleich 0 oder negativ, so findet in diesem Iterationsschritt ein ‚Hill Climb‘ statt. Bei einfachen Greedy-Verfahren, bei denen zur Ermittlung einer besseren Lösung allein die aktuelle Bisektion in Betracht gezogen wird, können solche ‚Hill Climbs‘ nicht vorkommen. Hier wird nach der  $i$ -ten Tauschoperation gestoppt, wenn für dessen Gewinn  $G_i \leq 0$  gilt.

Abbildung 5.4 zeigt einen ‚Hill Climb‘ bei der Bisektionierung eines bipartiten Graphen der Ordnung 16. Der betrachtete Knotentausch in der obersten Bisektion, der unter allen Knotenpaaren maximalen Gewinn bringt, ist der Tausch des Knotenpaares 8 und 10. Der Gewinn für den Schnitt beträgt allerdings 0. Dieser Tausch verändert jedoch unter anderem die ‚dValues‘ der Knoten 11 und 15, wie in der mittleren Bisektion ersichtlich. In Folge des vorherigen Knotentausches ist nun eine Tauschoperation möglich, deren Gewinn 2 und demnach größer 0 ist. Man sieht, dass der erste Tausch ohne Gewinn notwendig ist, bevor die Bisektion durch einen weiteren Tausch verbessert wird.

### 5.4.2 Alternative Methoden

Im Allgemeinen können auch andere Methoden zur minimalen Bisektionierung verwendet werden. Sie sind allerdings abhängig von der Art des Optimierungskriteriums oder können nur in Kombination mit den bewegungsbasierten Methoden eingesetzt werden. Ausführliche Beschreibungen dieser Methoden sind in [AK95] zu finden.

---

<sup>4</sup> ‚hill climb‘ (*engl*) heißt Bergrennen

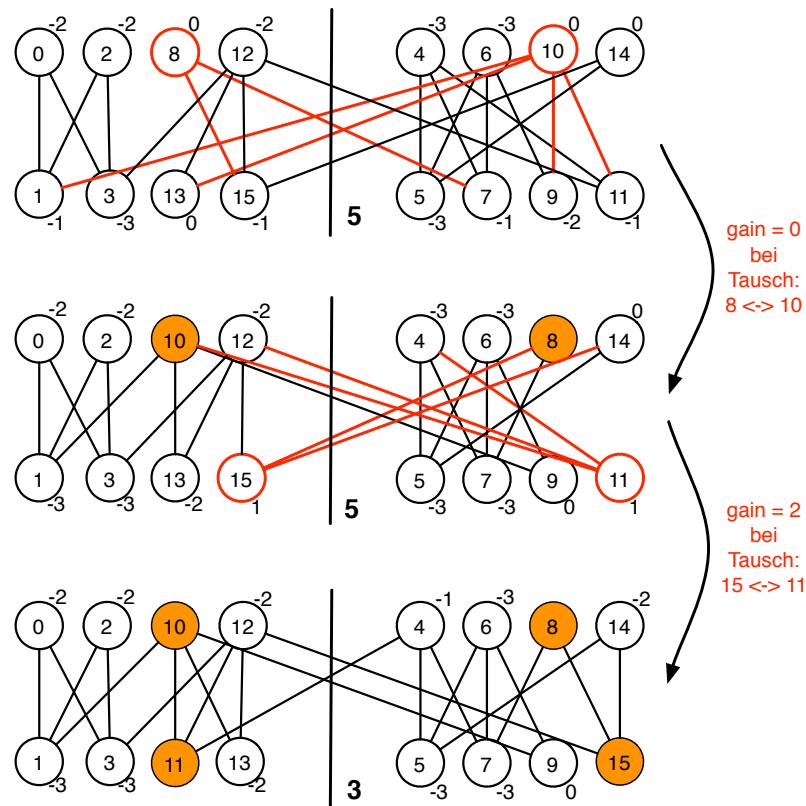


Abbildung 5.4: Beispiel für einen ‚Hill Climb‘ – Die Zahlen außen an den Knoten sind die ‚dValues‘; die vertikale Linie ist die Trennlinie der Bisektion des Graphen und die Zahl an der Linie deren Schnitt; rot umrandete Knoten sind die im nächsten Schritt zu tauschenden Knotenpaare; orangene Knoten sind markiert.

### Spektralbisektion

Bei der *Spektralbisektion* wird die *Spektralmethode* verwendet um eine geometrische Repräsentation des Bisektionierungsproblems zu konstruieren. Es wird der Eigenvektor  $y$  zum zweitgrößten Eigenwert der Laplace-Matrix  $Q = D - W$  bestimmt, wobei  $W = (w_{ij})_{0 \leq i, j \leq n-1}$  die Adjazenzmatrix des Graphen und  $D$  eine Diagonalmatrix mit  $d_{i,i} = \sum_{e=(v_i, v_j) \in E} w(e)$  ist. Mit dem sogenannten Fiedler-Vektor  $y = (y_0, y_1, \dots, y_{n-1})^T$  wird die Knotenmenge  $V$  wie folgt in zwei gleich große Mengen  $V_1$  und  $V_2$  geteilt: Sei  $r$  das Mittel über alle Werte  $y_i$ ,  $i = 0, 1, \dots, n-1$ . Dann gehören alle Knoten  $j$  mit  $y_j \leq r$  der Menge  $V_1$  an und alle übrigen Knoten der Menge  $V_2$ .

### Kombinatorische Methoden

*Kombinatorische Methoden* transformieren das Bisektionierungsproblem in andere „klassische“ Optimierungsprobleme, wie beispielsweise das der maximalen Flüsse. Diese Transformation basiert auf dem ‚max-flow-min-cut‘-Theorem von L. R. Ford und D. R. Fulkerson (siehe z. B. [Die96], S. 129).

### Clusterungsbasierte Methoden

Basis der *clusterungsbasierten Methoden* ist das ‚bottom-up‘-Prinzip. Es wird mit  $|V|$  Clustern gestartet, die je einen Knoten enthalten, und von diesen werden dann Schritt für Schritt je zwei so zu größeren Clustern zusammengefügt, dass eine Bisektion entsteht. Die erhaltenen Cluster werden dann beispielsweise als Eingabe für andere Bisektionierungsalgorithmen verwendet.

## 5.5 Minimale Sektionierung

Der Kernighan-Lin-Algorithmus beschränkt sich lediglich auf die Bisektionierung von Graphen. Der verwendete Knotentausch findet zwischen zwei Knotenmengen statt. Kann man nach demselben Prinzip auch eine minimale  $k$ -Sektionierung mit  $k > 2$  vornehmen? Bei einer gegebenen  $k$ -Sektion kann es jedoch notwendig sein, drei Knoten verschiedener Sektionselemente zyklisch miteinander zu vertauschen, um eine Sektion mit kleinerem Schnitt zu erhalten. Ein entsprechender Nachbarschaftsoperator hat allerdings sehr hohe Laufzeit, möchte man Tauschoperationen zwischen beliebig vielen Sektionselementen ermöglichen. Aus diesem Grund wird bei der *Zwei-Wege-* und der *Mehr-Wege-Methode* der Kernighan-Lin-Algorithmus mehrfach angewendet, um  $k$ -Sektionen mit annähernd minimalem Schnitt zu finden.

### 5.5.1 Zwei-Wege-Methode

Bei der Zwei-Wege-Methode wird der Kernighan-Lin-Algorithmus rekursiv angewendet. Das heißt, eine Startbisektion wird mit dem Kernighan-Lin-Algorithmus minimiert, und anschließend Bisektionen der beiden Elemente der erhaltenen Bisektion je für sich wieder minimiert. Dies geschieht so oft, bis die gewünschte Anzahl an Sektionselementen erreicht wird. Dabei ist als Sektionsgröße allerdings nur eine Zweierpotenz möglich. Führt man  $r$  mal rekursiv den Kernighan-Lin-Algorithmus aus, so erhält man eine  $2^r$ -Sektion. Abbildung 5.5 zeigt den Algorithmus in Pseudocode. Darin stellt *Kernighan – Lin*( $A, B$ ) in Schritt 6 den Aufruf des Kernighan-Lin-Algorithmus unter Eingabe der Startbisektion ( $A, B$ ) dar.

**Algorithmus:**


---

 Minimale  $2^r$ -Sektionierung mit der Zwei-Wege-Methode
 

---

Eingabe: Startbisektion  $(A, B)$ ,  
ganze Zahl  $r$ .

Ausgabe: Lokal minimale  $2^r$ -Sektion  $P = (P_1, P_2, \dots, P_{2^r})$ .

- 1: Bilde leere  $2^r$ -Sektion  $P := (P_1, P_2, \dots, P_{2^r})$ .
  - 2: Setze  $P_1 = V$ .
  - 3: Für  $i = 1, 2, \dots, r$  führe aus:
    - 4: Für  $j = 1, 3, 5, \dots, 2^i - 1$  führe aus:
      - 5: Bilde Startbisektion  $(A, B)$  der Knoten  $P_j$ .
      - 6:  $(P_j, P_{j+1}) = \text{Kernighan-Lin}(A, B)$ .
    - 7: Wenn  $i \neq r$  gilt:
      - 8: Für  $j = 1, 2, \dots, 2^i$  führe aus:
        - 9: Verschiebe  $P_{2 \cdot j - 1} = P_j$ .
- 

Abbildung 5.5: Zwei-Wege-Methode

Die Laufzeit der Zwei-Wege-Methode auf Graphen der Ordnung  $n$  beträgt bei  $2^r$ -Sektionierung:

$$\sum_{i=1}^r 2^{i-1} \cdot \left(\frac{n}{2^i}\right)^3 \in \mathcal{O}(r \cdot n^3).$$

Die Zwei-Wege-Methode liefert jedoch  $2^r$ -Sektionen  $P = (P_1, P_2, \dots, P_{2^r})$ , in denen nur  $2^{r-1}$  Elementpaare lokal minimal sind, wobei jedes Element genau in einem dieser Paare enthalten ist. Im letzten Durchlauf von Schritt 3 (siehe Abbildung 5.5) entstehen nämlich die minimalen Bisektionen  $(P_1, P_2)$ ,  $(P_3, P_4)$  bis  $(P_{2^{r-1}}, P_{2^r})$ . Die im vorherigen Durchlauf minimierten Bisektionen  $(P_1 \cup P_2, P_3 \cup P_4)$  bis  $(P_{2^{r-3}} \cup P_{2^{r-2}}, P_{2^{r-1}} \cup P_{2^r})$  garantieren nicht, dass beispielsweise auch die Elementpaare  $(P_1, P_3)$  oder  $(P_1, P_4)$  lokal minimal sind. Somit können, trotz der Minimierungen in den ersten  $r - 1$  Schleifendurchläufen, in der ausgegebenen  $k$ -Sektion die Schnitte der meisten Elementpaare sehr groß sein.

Wie geht man allerdings vor, wenn man in einem Arbeitsschritt mehr als  $2^{r-1}$  Elementpaare minimieren möchte? Die Anwendung des Kernighan-Lin-Algorithmus auf das Elementpaar  $(A, B)$  kann nämlich den Schnitt des

zuvor lokal minimierten Paares  $(A, C)$  erneut verschlechtern. Es stellt sich außerdem die Frage, ob eine derartige Minimierung wirklich bessere Sektionen liefert. Um dies untersuchen zu können, wird im folgenden Abschnitt eine weitere Methode eingeführt.

### 5.5.2 Mehr-Wege-Methode

Die Mehr-Wege-Methode liefert lokal minimierte  $k$ -Sektionen für beliebige Werte  $k$ , indem eine gegebene Start- $k$ -Sektion weitestgehend paarweise minimiert wird. Dazu werden je zwei Elemente einer  $k$ -Sektion als Bisektion betrachtet und diese mit dem Kernighan-Lin-Algorithmus lokal minimiert. Man kann sich vorstellen, dass sich die Laufzeit des Algorithmus enorm erhöht, wenn man dabei alle  $\binom{n}{2} = (n^2 - n)/2$  Elementpaare berücksichtigt. Denn tauscht man zwischen zwei Elementen der  $k$ -Sektion Knoten aus, so verändert dies deren Schnitte in Kombination mit den übrigen Elementen. Es wären somit mehrere Durchläufe über die Elementpaare notwendig, bis die  $k$ -Sektion paarweise minimal ist. Gerät man dabei unter Umständen in eine Endlosschleife von Knotentauschoperationen, so kann gar keine lokal minimale Sektion gefunden werden. Aufgrund empirischer Tests von B. W. Kernighan und S. Lin (siehe [KL70], S. 304) reicht es aus, pro Durchlauf je zwei Elemente der  $k$ -Sektion zu paaren, von denen maximal eines bereits im entsprechenden Durchlauf verändert wurde. Dazu wird eine Permutation  $P_{i_1}, P_{i_2}, \dots, P_{i_k}$  der Sektionselemente bestimmt, und nacheinander für  $1 \leq j \leq k - 1$  das Elementpaar  $(P_{i_j}, P_{i_{j+1}})$  minimiert. Es wird somit jedes Element in einem Durchlauf höchstens zwei Mal bearbeitet. Die Permutation für die sogenannte *Durchlaufreihenfolge* kann entweder für alle Durchläufe fest vorgegeben oder in jedem Durchlauf neu gewählt werden. In jedem Fall wird die Sektionierung gestoppt, wenn ein Durchlauf keinen weiteren Gewinn bringt. Dabei ist der Gewinn eines Durchlaufs die Summe aller Gewinne  $G_{\text{Kernighan-Lin}}$ , die durch die Aufrufe des Kernighan-Lin-Algorithmus erzielt werden. Abbildung 5.6 zeigt den Algorithmus in Pseudocode.

Die Laufzeit der Schritte 4 bis 7 auf Graphen der Ordnung  $n$  beträgt bei  $k$ -Sektionierung:

$$(k - 1) \cdot \left( \left\lceil \frac{n}{k} \right\rceil \right)^3 \in \mathcal{O}(n^3).$$

Hier werden drei verschiedene Durchlaufreihenfolgen (siehe Abbildung 5.6, Schritt 4) vorgestellt: die *feste*, die *zufällige* und die *sektionsoptimierte*. Bei der festen Reihenfolge wird als Permutation der Sektionselemente die Reihenfolge der Elemente nach ihrer Nummerierung verwendet. Das heißt, es gilt  $i_j = j$  für alle  $1 \leq j \leq k$ . In diesem Fall werden bei jedem Durchlauf die gleichen Elementpaare minimiert. Die einzelnen Elemente verändern sich

**Algorithmus:**


---

 Minimale  $k$ -Sektionierung mit der Mehr-Wege-Methode
 

---

Eingabe: Start- $k$ -Sektion  $P = (P_1, P_2, \dots, P_k)$ .Ausgabe: Lokal minimale  $k$ -Sektion  $P = (P_1, P_2, \dots, P_k)$ .

- 1: Bilde Start- $k$ -Sektion  $P := (P_1, P_2, \dots, P_k)$  der Knoten  $V$ .
  - 2: Wiederhole...
  - 3: Setze  $G=0$ .
  - 4: Bilde eine Permutation  $(P_{i_r})_{1 \leq r \leq k}$ .
  - 5: Für  $j = 1, 2, \dots, k - 1$  führe aus:
  - 6:  $(P_{i_j}, P_{i_{(j+1)}}) = \text{Kernighan-Lin}(P_{i_j}, P_{i_{(j+1)}})$ .
  - 7: Setze  $G = G + G_{\text{Kernighan-Lin}}$ .
  - 8: Solange bis  $G \leq 0$ .
- 

Abbildung 5.6: Mehr-Wege-Methode

zwar bei jeder Anwendung des Kernighan-Lin-Algorithmus, doch trotzdem beschränkt sich die paarweise Minimalität hier auf gewisse Elementpaare. Um dies zu verhindern, wird bei der zufälligen Durchlaufreihenfolge in jedem Durchlauf eine neue zufällige Permutation der Sektionselemente gewählt. In diesem Fall werden in mehreren Durchläufen je verschiedene Elementpaare minimiert und somit insgesamt eine stärkere paarweise Minimalität erzielt. Dabei kann es allerdings passieren, dass Elementpaare ausgelassen werden, deren Schnitt sehr groß ist. Dies umgeht man weitestgehend, indem man die jeweilige Permutation für die Durchlaufreihenfolge erst während der Schleifendurchläufe in Schritt 5 bildet. Dabei wird im  $j$ -ten Schleifendurchlauf das nächste Permutationselement  $P_{i_{(j+1)}}$  danach auswählt, für welches Sektionselement das entsprechende Elementpaar  $(P_{i_j}, P_{i_{(j+1)}})$  maximalsten Schnitt hat.

Die Anzahl Durchläufe beträgt laut B. W. Kernighan und S. Lin (siehe [KL70], S. 304) bei Graphen der Ordnung  $\leq 100$  und  $k < 6$  weniger als 5. Es ist allerdings nicht ersichtlich, was für eine Durchlaufreihenfolge sie bei ihren empirischen Versuchen verwendet haben.

Bei dieser Methode ist die Eingabe einer Start- $k$ -Sektion erforderlich. Hier reduziert die Verwendung besserer  $k$ -Sektionen, wie von B. W. Kernighan und S. Lin bestätigt wird, die Laufzeit. Die Bearbeitung von bipartiten Graphen ermöglicht wieder eine gezielte, aber doch gewissermaßen zufällige Wahl der



Start- $k$ -Sektion. Wie in Abschnitt 5.1.2 erläutert, haben hier  $k$ -Sektionen, die gruppenbalanciert sind, fast immer einen kleineren Schnitt als andere. Sie werden demnach vorzugsweise als Start- $k$ -Sektionen ausgewählt. Dass die Gruppenbalanciertheit der  $k$ -Sektionen allerdings nach der Sektionierung mit der Mehr-Wege-Methode nie ganz erhalten bleibt, wird anhand entsprechender Experimente in Kapitel 6 herausgefunden.

## 5.6 Gruppeneinteilung durch MinMax-Sektionierung

Aufgabe bei der Gruppeneinteilung ist es, eine  $k$ -Sektion der Knotenmenge  $V$  zu finden, deren Schnitt im Sympathiegraph  $G = (V, E)$  mit Gewichtsfunktion  $w : E \rightarrow \mathbb{R}$  minimal und im Antipathiegraph  $G' = (V, E')$  mit Gewichtsfunktion  $w' : E' \rightarrow \mathbb{R}$  maximal ist.

Wie bei der Einführung bewegungsbasierter Methoden erläutert, sind diese unabhängig von der Art ihrer Zielfunktion. Die Qualität einer Bisektion kann demnach im Kernighan-Lin-Algorithmus nach beliebigen Kriterien gemessen werden. Somit kann der Algorithmus aus Abbildung 5.3 statt zur minimalen auch zur maximalen Bisektionierung eingesetzt werden. Dafür wird die Berechnung des Gewinnes  $gain(a, b)$  durch den Tausch der Knoten  $a$  und  $b$  entsprechend abgeändert. Die bei der minimalen Bisektionierung verwendeten ‚dValues‘ werden nun zu sogenannten ‚nValues‘  $N_v$  der Knoten  $v \in V$ , die den negativen Werten der ‚dValues‘  $D_v$  entsprechen. Es gilt demnach:

$$N_v := I_v - E_v = -D_v \quad \text{für alle } v \in V.$$

Analog zu Hilfssatz 5.1 kann folgender Hilfssatz bewiesen werden:

**Hilfssatz 5.2** *Seien  $a \in A$  und  $b \in B$  zwei Knoten aus verschiedenen Elementen einer Bisektion  $P = (A, B)$  des Graphen  $G = (V, E)$ . Wenn man  $a$  und  $b$  tauscht, das heißt die Bisektion  $P' = (A - \{a\} \cup \{b\}, B - \{b\} \cup \{a\})$  bildet, gilt:*

$$cut(P') - cut(P) = N_a + N_b + 2 \cdot w_{ab}.$$

Verwendet man also die ‚nValues‘ statt der ‚dValues‘ und maximiert in Schritt 5 des Kernighan-Lin-Algorithmus entsprechend den Gewinn  $gain(a_i, b_i) = N_{a_i} + N_{b_i} + 2 \cdot w_{a_i b_i}$  über alle unmarkierten Knoten  $a_i \in A$  und  $b_i \in B$ , so erhält man eine lokal maximale Bisektion. Bei der Aktualisierung der ‚nValues‘ in Schritt 7 werden die neuen Werte  $N'_x$  und  $N'_y$  wie folgt aus den alten

‚nValues‘  $N_x$  und  $N_y$  der Knoten  $x \in A - \{a_i\}$  und  $y \in B - \{b_i\}$  berechnet:

$$\begin{aligned} N'_x &= N_x - 2 \cdot w_{xa_i} + 2 \cdot w_{xb_i}, \\ N'_y &= N_y - 2 \cdot w_{yb_i} + 2 \cdot w_{ya_i}. \end{aligned}$$

Die Begründung hierfür ist analog zu der bei den ‚dValues‘.

Will man nun die Zwei-Wege- oder auch die Mehr-Wege-Methode zur Gruppeneinteilung verwenden, so muss eine geeignete Zielfunktion des Kernighan-Lin-Algorithmus gefunden werden:

Für die zu lösende Aufgabe misst man die Qualität einer Bisektion  $P = (A, B)$  der Knotenmenge  $V$  anhand des Wertes  $cut_{G'}(P) - cut_G(P)$ , wobei  $cut_G(P)$  den Schnitt der Bisektion im Graphen  $G$  beschreibt. Um das oben beschriebene Ziel zu erreichen, gilt es dann, diesen Wert zu maximieren. Dazu werden die ‚dValues‘  $D_v$  der Knoten  $v$  in  $G$  und die ‚nValues‘  $N_v$  der Knoten  $v$  in  $G'$  berechnet. Mit diesen Werten wird dann in Schritt 5 wie folgt der Gewinn berechnet:

$$\begin{aligned} gain(a_i, b_i) &:= gain_G(a_i, b_i) + gain_{G'}(a_i, b_i) \\ &= D_{a_i} + D_{b_i} - 2 \cdot w_{a_i b_i} + N_{a_i} + N_{b_i} + 2 \cdot w'_{a_i b_i} \end{aligned}$$

Dabei beschreibt  $gain_G(a, b)$  analog wieder den Gewinn, der im Graphen  $G$  durch den Tausch der Knoten  $a$  und  $b$  erzielt wird. Vertauscht man in der Bisektion  $P$  nun zwei Knoten  $a_i$  und  $b_i$ , für die  $gain(a_i, b_i) > 0$  gilt, dann verändern sich der Schnitt  $cut_G(P)$  im Sympathiegraphen und der Schnitt  $cut_{G'}(P)$  im Antipathiegraphen so, dass  $cut_{G'}(P) - cut_G(P)$  nach dem Tausch größer ist als davor.

Da der ursprüngliche Kernighan-Lin-Algorithmus erhalten bleibt, der einzige Zusatz darin besteht, dass hier doppelt so viele Werte, die ‚dValues‘ und die ‚nValues‘ der Knoten, bestimmt und aktualisiert werden, ist die Laufzeit des hier entwickelten *MinMax-Bisektionierungsalgorithmus* ebenfalls in  $\mathcal{O}(n^3)$ .

Um die gesuchten  $k$ -Sektionen zu erhalten, werden wieder die Zwei-Wege- und die Mehr-Wege-Methode eingesetzt. Sie werden im Folgenden als *MinMax-Sektionierungsalgorithmen* bezeichnet. Die in Abschnitt 5.5.2 vorgestellte sektionsoptimierte Durchlaufreihenfolge für die Mehr-Wege-Methode wird hier entsprechend angepaßt: Das nächste zu optimierende Elementpaar  $P = (A, B)$  wird danach ausgesucht, wessen Differenz  $cut_{G'}(P) - cut_G(P)$  am kleinsten ist.

Die Berechnung des Gewinnes  $gain(a_i, b_i)$  als Summe der einzelnen Gewinne im Sympathie- und Antipathiegraphen bietet die Möglichkeit eine Gewichtung vorzunehmen. Verwendet man als Gewinn

$$gain(a_i, b_i) := g^+ \cdot gain_G(a_i, b_i) + g^- \cdot gain_{G'}(a_i, b_i) \quad \text{mit } g^+, g^- \in \mathbb{R}$$

und wählt  $g^+ > g^-$ , so wird bei der zugehörigen Sektionierung die Minimierung der Sympathieschnitte entsprechend stärker gewichtet als die Maximierung der Antipathieschnitte. Wünscht man keine Gewichtung, so setzt man  $g^+ = g^- = 1$ . Dass durch stärkere Gewichtung der Sympathiegewinne eine „Verschiebung“ der Sympathie- und Antipathieerfüllungen verursacht wird, geht aus den entsprechenden Experimenten in Abschnitt 6.4.10 hervor.



# Kapitel 6

## Experimente

Bei Experimenten zum Kernighan-Lin-Algorithmus in [KL70] wurde beispielsweise untersucht, welche Startbisektion sich zur Optimierung am besten eignet. B. W. Kernighan und S. Lin testeten außerdem, wie viele Iterationsschritte nötig sind, bis ein lokales Minimum erreicht wird. Sie beobachteten bei der Mehr-Wege-Methode, wie stark die Durchlaufzahl steigt, wenn man die Anzahl Knoten und Gruppen erhöht. Allerdings wurden bei diesen Experimenten beliebige zufällige Graphen verwendet.

Um die hier entwickelte Strategie zur Gruppeneinteilung zu untersuchen, werden aus der Kontaktaufnahme resultierende Bewertungen konstruiert, die nah an denen der Realität sind. Auf diesen speziellen Sympathie- und Antipathiegraphen werden die in Kapitel 5.6 definierten MinMax-Sektionierungsalgorithmen getestet. Unter anderem werden Experimente zur Verwendung unterschiedlicher Startsektionen durchgeführt und die Anzahl der Schleifendurchläufe in den unterschiedlichen Sektionierungsalgorithmen beobachtet. Zudem wird bei den MinMax-Sektionierungen, wie auch bei den Sektionierungen mit Hilfe des ursprünglichen Kernighan-Lin-Algorithmus untersucht, in wieweit Gruppenbalanciertheit erreicht wird. Weitere Experimente beschäftigen sich mit der Erfüllung der Sym- und Antipathien. Die Frage ist, wie viele der ausgesprochenen Sym- und Antipathien bei der ermittelten Gruppeneinteilung pro Person realisiert werden. Und wie verändern sich diese Ergebnisse, wenn man beispielsweise mehr Sym- als Antipathien aussprechen lässt oder die Gewichtung innerhalb der Zielfunktion der MinMax-Sektionierungen variiert?

## 6.1 Generierung der Daten

Unter Angabe der geraden Personenzahl  $n$  und der Anzahl pro Person zu führender Gespräche  $g \leq \lceil \log_2 n \rceil$  kann die Adjazenzmatrix des Knödel-Graphen  $Knö(n, g)$  ermittelt werden. Nun soll eine realistische Aussprache von  $p$  Sympathien und  $n$  Antipathien nachgebildet werden, wobei  $p + n \leq g$  gilt. In der Realität ist zu beobachten, dass es Personen gibt, die den übrigen Teilnehmern sympathischer sind als alle anderen. Wird nun angenommen, dass jedem Gesprächspartner einer Person mit Wahrscheinlichkeit  $(1/g)$  eine bestimmte Bewertung ausgesprochen wird, so tritt der Fall, dass eine Person von all seinen Gesprächspartnern als „sympathisch“ bewertet wird, mit Wahrscheinlichkeit  $(1/g)^p$  ein. Aufgrund der Tatsache, dass die Teilnehmer nur flüchtigen Kontakt zueinander aufnehmen und dann recht spontan Bewertungen vergeben, scheint obige Wahrscheinlichkeitsverteilung der Bewertung für die Generierung der Sympathie- und Antipathiegraphen geeignet zu sein. Die Adjazenzmatrizen dieser Graphen werden daher aus der Adjazenzmatrix des Kontaktgraphen gebildet, indem pro Zeile  $p$  und  $n$  unterschiedliche Einträge  $\neq 0$  ausgewählt werden. Alle übrigen  $(g - p - n)$  Einträge einer Zeile werden 0 gesetzt, stellen somit die neutral bewerteten Personen dar. Die beiden entstehenden Adjazenzmatrizen entsprechen nun also den gerichteten Sympathie- und Antipathiegraphen. Diese werden, wie in Kapitel 4.2 beschrieben, in ungerichtete Graphen konvertiert und im Folgenden zur experimentellen Untersuchung der Gruppeneinteilung verwendet.

Für die Experimente werden je 100 Graphenpaare aus Sympathie- und Antipathiegraphen mit 20, 30, 40,  $\dots$ , 100 Knoten generiert. Der vollständige Knödel-Graph ist in diesen Fällen 4-, 5- oder 6-regulär. Um bei Graphen unterschiedlicher Ordnungen sowohl die Schnitte, als auch die Erfüllungen von Sym- und Antipathien der gefundenen Sektionen besser miteinander vergleichen zu können, wird für alle Graphen die Gesprächszahl  $g = 4$  gewählt. Der jeweilige Kontaktgraph der Ordnung  $n$  ist demnach der Knödel-Graph  $Knö(n, 4)$ .

Bei der Generierung der Daten kann außerdem gewählt werden, wie viele Sym- und Antipathien pro Person ausgesprochen werden. Für einen großen Teil der durchgeführten Experimente werden bei 4 flüchtigen Kontakten pro Person je zwei Sym- und Antipathien vergeben. Wird in Abschnitt 6.4.9 untersucht, wie sich die Ergebnisse verändern, wenn die Anzahl Sympathien die Anzahl Antipathien überwiegt, so werden Daten generiert, die die Aussprache von drei Sympathien und einer Antipathie nachbilden.

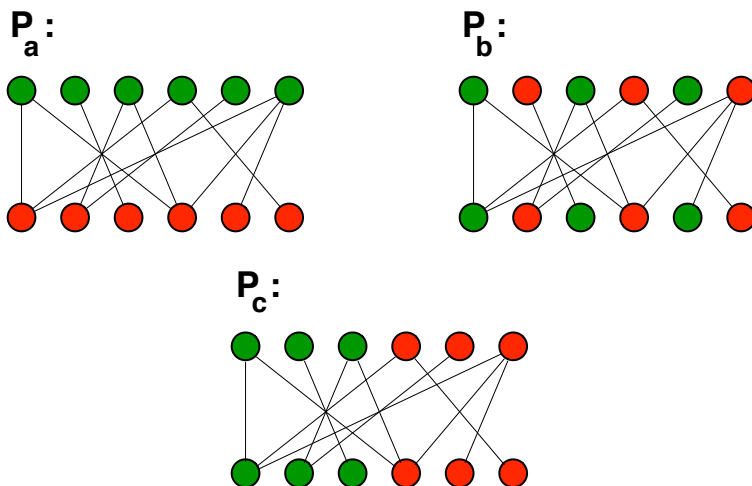
## 6.2 Verschiedene Parameter

### 6.2.1 Startbisektionen

Wie in Abschnitt 5.4.1 beschrieben, können auf bipartiten Graphen recht gute und doch gewissermaßen beliebige Bisektionen als Eingabe für den ursprünglichen Kernighan-Lin-Algorithmus angegeben werden. Die Elemente dieser Startbisektionen bestehen je zur Hälfte aus Knoten der Menge  $A$  bzw.  $B$  des bipartiten Graphen  $G = (A \cup B, E)$ . Sie werden nun auch für die MinMax-Sektionierung bei Verwendung der Zwei-Wege-Methode getestet.

(Zu den folgenden Definitionen betrachte man untere Abbildung.)

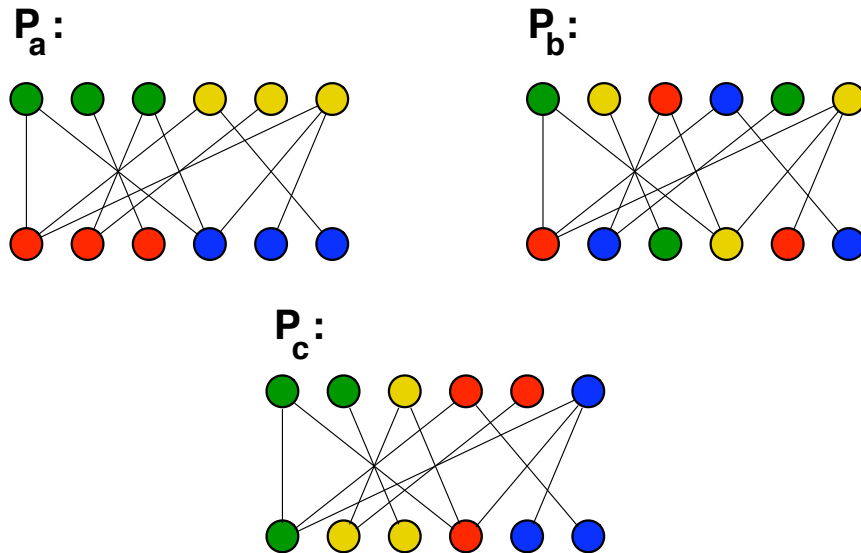
Zur Untersuchung, ob die Eingabe anderer Bisektionen schlechtere lokal optimale Sektionen liefern, wird auch eine Startbisektion verwendet, deren Schnitt maximal ist. Diese ist die Bisektion  $P_a := (P_1, P_2)$  mit  $P_1 = A$  und  $P_2 = B$ . Eine weitere, diesmal gruppenbalancierte Startbisektion  $P_b := (P_1, P_2)$  ist die Knotenteilung, bei der  $P_1$  und  $P_2$  je genau zwei Knoten aufeinanderfolgender Nummerierung  $i$  und  $i + 1$  enthalten. Das heißt es gilt  $P_1 := \{0, 1, 4, 5, 8, \dots\}$  und  $P_2 := \{2, 3, 6, 7, 10, \dots\}$ . Die dritte untersuchte Startbisektion  $P_c := (P_1, P_2)$  setzt sich zusammen aus  $P_1 := \{0, 1, 2, \dots, n/2 - 1\}$  und  $P_2 := \{n/2, n/2 + 1, \dots, n - 1\}$ . Sie ist ebenfalls gruppenbalanciert. Beim jeweiligen Experiment wird angegeben, welche der drei Startbisektionen  $P_a, P_b$  und  $P_c$  verwendet wird.



### 6.2.2 Start- $k$ -Sektionen

(Zu den folgenden Definitionen betrachte man untere Abbildung.)

Auf ähnliche Weise werden mehr oder weniger günstige  $k$ -Sektionen als Start- $k$ -Sektionen der Mehr-Wege-Methode festgelegt. Die erste  $k$ -Sektion  $P_a := (P_1, P_2, \dots, P_k)$  teilt die Knotenmengen  $A$  und  $B$  in  $k/2$  annähernd gleich große Teile. Ist  $k$  allerdings ungerade, so entsteht neben  $(k-1)/2$  Gruppen, die nur aus Knoten der Menge  $A$  bzw.  $B$  des bipartiten Graphen  $G = (A \cup B, E)$  bestehen, auch eine Gruppe, die gruppenbalanciert ist. Der Schnitt dieser  $k$ -Sektion ist folglich auf allen bipartiten Graphen relativ groß. Bei der  $k$ -Sektion  $P_b := (P_1, P_2, \dots, P_k)$  mit  $P_1 := \{„jedes  $k$ -te Element der nach Nummerierung sortierten Menge  $A$  und fortlaufend der Menge  $B$ “\}$ ,  $P_2 := \{„jedes  $k$ -te Element beginnend bei dem zweiten Element der Menge  $A$  und fortlaufend der Menge  $B$ “\}$  usw. sind hingegen alle Elemente gruppenbalanciert. Eine weitere gruppenbalancierte  $k$ -Sektion  $P_c := (P_1, P_2, \dots, P_k)$  setzt sich aus  $P_1 := \{0, \dots, n/k - 1\}$ ,  $P_2 := \{n/k, \dots, 2 \cdot n/k - 1\}$  bis  $P_k := \{(k-1) \cdot n/k, \dots, n - 1\}$  zusammen. Es wird jeweils angegeben, welche der verschiedenen Start- $k$ -Sektionen  $P_a, P_b$  und  $P_c$  bei den Experimenten getestet werden. Zur besseren Lesbarkeit werden sie hier auch als *Startsektionen* bezeichnet.





### 6.2.3 Fokusse der Sektionierung

Wie beschrieben, kann der eigentliche Kernighan-Lin-Algorithmus zur Suche einer lokal minimalen Sektion auch verwendet werden, um eine lokal maximale Sektion oder eine Sektion zu finden, die in einem von zwei verschiedenen Graphen lokal minimal und im anderen zugleich lokal maximal ist. Die entstehenden Sektionierungsalgorithmen auf Basis dieser drei Varianten des Kernighan-Lin-Algorithmus werden hier *Min-*, *Max-* und *MinMax-Algorithmus* genannt. Es wird in den Experimenten angegeben, die Ergebnisse welcher dieser Algorithmen jeweils betrachtet werden.

### 6.2.4 Sektionierungsmethoden

Es werden die Zwei-Wege- und die Mehr-Wege-Methode untersucht. Allerdings können mit der Zwei-Wege-Methode nur  $k$ -Sektionen mit  $k = 2^r$  ermittelt werden. Da erfahrungsgemäß 6 Personen sehr gut im Team zusammenarbeiten, wird entsprechend der Graphgrößen eine 4-, 8- oder 16-Sektionierung vorgenommen. Die einzelnen Werte  $k$  zu den Ordnungen  $n$  der Graphen sind Tabelle 6.1 zu entnehmen. Ist  $n$  nicht durch  $k$  teilbar, so entstehen Sektionselemente, deren Größen  $g$  sich um 1 unterscheiden. Diese Größen werden durch „ $\lfloor n/k \rfloor - \lceil n/k \rceil$ “ notiert.

$n$	20	30	40	50	60	70	80	90	100
$k$	4	4	8	8	8	16	16	16	16
$g$	5	7 – 8	5	6 – 7	7 – 8	4 – 5	5	5 – 6	6 – 7

Tabelle 6.1: Zwei-Wege-Methode – Anzahl Sektionselemente  $k$  und deren Größen  $g$  bei Graphen der Ordnung  $n$

Werden in den Experimenten die Ergebnisse der Zwei- und die Mehr-Wege-Methode miteinander verglichen (siehe Abschnitt 6.4), so werden bei der Mehr-Wege-Methode dieselben  $k$ -Sektionierungen vorgenommen. Betrachtet man jedoch einzig Ergebnisse der Mehr-Wege-Methode, so können beliebige große Sektionen ermittelt werden. Die Größe  $k$  der Sektionen richtet sich dann alleine nach der Ordnung  $n$  der Graphen und der hier gewählten, idealen Gruppengröße  $g \approx 6$ . Tabelle 6.2 zeigt die entsprechend gewählten Werte für  $k$ .

$n$	20	30	40	50	60	70	80	90	100
$k$	3	5	7	8	10	11	13	15	17
$g$	6 – 7	6	5 – 6	6 – 7	6	6 – 7	6 – 7	6	5 – 6

Tabelle 6.2: Mehr-Wege-Methode – Anzahl Sektionselemente  $k$  und deren Größen  $g$  bei Graphen der Ordnung  $n$

### 6.2.5 Durchlaufreihenfolgen der Mehr-Wege-Methode

Für die Mehr-Wege-Methode wurden in Abschnitt 5.5.2 drei verschiedene Durchlaufreihenfolgen beschrieben, die feste, zufällige und sektionsoptimierte. In den Experimenten wird jeweils angegeben, welche Reihenfolgen verwendet werden. Beim Vergleich der Zwei- mit der Mehr-Wege-Methode wird beispielsweise ausschließlich die zufällige Reihenfolge eingesetzt, da diese meist die besseren Ergebnisse liefert. Um deutlich zu machen, dass es sich hierbei um einen Parameter der Mehr-Wege-Methode handelt, wird im Folgenden die Bezeichnung „MW-Durchlaufreihenfolge“ verwendet.

### 6.2.6 Gewichtungen

Dieser Parameter betrifft ausschließlich die MinMax-Sektionierung. Wie in Abschnitt 5.6 beschrieben, kann bei der Berechnung des Gewinnes  $gain(a_i, b_i)$  eine Gewichtung der einzelnen Gewinne im Sympathie- und im Antipathiegraphen vorgenommen werden. In fast allen der hier durchgeführten Experimente wird  $g^+ = g^- = 1$  gewählt. Wird in Abschnitt 6.4.10 die Abhängigkeit der Ergebnisse von einer stärkeren Gewichtung der Sympathiegewinne untersucht, so gilt  $g^+ = 2$  und  $g^- = 1$ . Die Gewinne im Sympathiegraphen werden hier also doppelt so stark gewichtet wie die im Antipathiegraphen. Die in diesen Experimenten angegebenen Ergebnisse des Min-Algorithmus dienen dem Vergleich, die Gewichtung hat aber keinerlei Auswirkung auf diese.

## 6.3 Darstellungsform der Ergebnisse

In den beschriebenen Experimenten werden die vorgestellten Sektionierungsalgorithmen untersucht. Die Darstellung der erzielten Ergebnisse erfolgt durch Graphiken, die mit dem Programm Gnuplot erstellt wurden. In diesem Abschnitt sollen die dazu gewählte Darstellungsformen erläutert werden.

### 6.3.1 Anzahl Iterationsschritte und Durchläufe

Untersucht man die Anzahl der Iterationsschritte bzw. Durchläufe der verschiedenen Algorithmen, so werden die Ergebnisse durch sogenannte *Fehlerbalken* dargestellt. Dabei ist der Kreis auf einem Fehlerbalken der Mittelwert der Anzahlen über alle 100 bearbeiteten Graphen. Das obere bzw. untere Ende des Balken ist der Mittelwert aller Werte, die oberhalb bzw. unterhalb des gesamten Mittelwertes liegen. So kann abgeschätzt werden, wie viele Iterationen bzw. Durchläufe im Mittel notwendig sind, bis ein lokales Optimum erreicht wird, und wie stark durchschnittlich die Abweichung von diesem Mittelwert ist.

### 6.3.2 Abweichung des Schnittes vom Optimum

In bestimmten Experimenten werden über alle Sympathie- und Antipathieschnitte eines Graphen je das Minimum und das Maximum ermittelt. Sie werden *absolute Minimum* und *absolute Maximum* genannt und dienen dazu die prozentualen Abweichungen der größeren Sympathie- bzw. kleineren Antipathieschnitte vom Minimum bzw. Maximum berechnen zu können. Die Abweichungen multipliziert mit dem absoluten Minimum bzw. Maximum ergibt also die jeweils erzielten Sympathie- bzw. Antipathieschnitte. In den entsprechenden Graphiken werden die absoluten Minima und Maxima als Linie auf Höhe des Wertes 1 dargestellt. Die Abweichungen werden schließlich je für den Min-, Max- und den MinMax-Algorithmus über 100 Graphen einer Ordnung gemittelt und durch Impulse zur Linie „1“ dargestellt. In den Graphiken wird die Abweichung der Sympathieschnitte vom absoluten Minimum *Min-Abweichung* und die der Antipathieschnitte vom absoluten Maximum *Max-Abweichung* genannt.

### 6.3.3 Schnitt

Eine weitere Möglichkeit, die Ergebnisse verschiedener Algorithmen oder eines Algorithmus unter Verwendung verschiedener Startsektionen zu untersuchen, ist, die Schnitte der jeweils erhaltenen lokalen Optima in Graphiken zu präsentieren. Dazu werden die Schnitte im Sympathie- und Antipathiegraphen, die vom MinMax-Algorithmus Bestimmt werden, durch oben eingeführte Fehlerbalken dargestellt. Um sehen zu können, wie nah die jeweiligen Schnitte am vom Min-Algorithmus ermittelten Sympathieschnitt und am vom Max-Algorithmus ermittelten Antipathieschnitt liegen, werden Geraden eingefügt, die dies ermöglichen. Eine Gerade als obere Schranke für die Maxima der Antipathieschnitte ist leicht zu bestimmen: Im bipartiten

Antipathiegraphen ist der maximal mögliche Schnitt das Produkt aus der Anzahl pro Person ausgesprochener Antipathien und der Personenzahl. Für die Gerade zur Darstellung der Minima kann nur eine Näherung anhand der Ergebnisse des Min-Algorithmus angegeben werden: Dafür werden die für die Fehlerbalken notwendigen Werte ermittelt, an die dann mit dem Levenberg-Marquardt-Verfahren (siehe [Ahn04], S. 58 ff) eine Gerade genähert wird. Anhand dieser Gerade können letztendlich auch die Ergebnisse des Min-Algorithmus miteinander verglichen werden.

### 6.3.4 Gruppenbalanciertheit

Die in Abhängigkeit von den Mengen  $A$  und  $B$  des bipartiten Graphen  $G = (A \cup B, E)$  definierte Gruppenbalanciertheit einer Sektion (siehe Abschnitt 5.1.2) wird in den Experimenten wie folgt untersucht:

Weicht das Verhältnis von Knoten der Menge  $A$  zu Knoten der Menge  $B$  in einem Sektionselement von 1 ab, so befinden sich in diesem ohne Beschränkung der Allgemeinheit mindestens zwei Knoten mehr aus  $A$  als aus  $B$ . Tauscht man in einem minimal unbalancierten Element folglich einen Knoten aus  $A$  gegen einen Knoten aus  $B$  aus, so beträgt die Balanciertheit 1. Den Wert auszutauschender Knoten kann man als Maß für die Unbalanciertheit verwenden. Ist die Unbalanciertheit also groß, so sind viele Knoten auszutauschen, um Gruppenbalanciertheit herzustellen.

Betrachtet man Sektionselemente mit ungerader Anzahl von Knoten, so kann das Verhältnis der Knoten aus  $A$  zu denen aus  $B$  nie 1 betragen. Folglich wird hier als Maß die Anzahl Knoten verwendet, die auszutauschen ist, bis die maximal mögliche Balanciertheit erreicht ist.

Die über 100 bearbeiteten Graphen ermittelten Unbalanciertheiten werden in Graphiken ebenfalls mit Fehlerbalken abgebildet. Im Falle der Sektionierung mit der Mehr-Wege-Methode kann fast maximale Unbalanciertheit eintreten. Es wird in den entsprechenden Graphiken zur Veranschaulichung eine Gerade eingefügt, die eine obere Schranke für die Unbalanciertheiten darstellt. Diese erfüllt  $f(x) = \frac{x}{2}$ . Die genauen Größen der maximal möglichen Unbalanciertheiten hier betrachteter Graphen sind in Tabelle 6.3 zu finden.

$n$	20	30	40	50	60	70	80	90	100
$u$	8	15	18	23	30	31	38	45	48

Tabelle 6.3: Maximale Unbalanciertheit  $u$  von Graphen der Ordnung  $n$  bei Sektionierung mit der Mehr-Wege-Methode

### 6.3.5 Erfüllung der Sym- und Antipathien

Wie in Abschnitt 5.1.3 beschrieben, wird untersucht, wie viele der pro Person ausgesprochenen Sym- und Antipathien bei der Gruppeneinteilung erfüllt werden. Dazu wird über alle Personen jedes Graphen einer Ordnung die Anzahl erfüllter Sym- und Antipathien gemittelt. Punkte in den Graphiken geben dieses Mittel prozentual zu den jeweils ausgesprochenen Bewertungen an. Die Prozentzahl wird als *Erfüllung pro Person* bezeichnet. Bei Betrachtung der Ergebnisse ist zu berücksichtigen, dass in den Experimenten von Abschnitt 6.4.9 drei Sympathien und eine Antipathie ausgesprochen werden und deren Anzahlen in den übrigen Abschnitten jeweils 2 betragen.

Um genauere Informationen zur Erfüllung der Sym- und antipathien zu erhalten, wird in einem anderen Experiment beobachtet, wie viele der Personen bei der Gruppeneinteilung keine Sympathie, keine Antipathie oder gleichzeitig keine Sympathie und keine Antipathie erfüllt bekommen. Diese Personenzahl wird prozentual zur gesamten Anzahl Personen angegeben und in den Graphiken durch Punkte dargestellt. Die Prozentzahlen werden als *Nichterfüllung* der jeweiligen Bewertungen bezeichnet. Da meist die Nichterfüllung der Sympathien um einiges größer ist als die Nichterfüllung der Antipathien und der Sym- und Antipathien zugleich, werden sie in getrennten Graphiken präsentiert.

## 6.4 Experimentelle Ergebnisse

Im Folgenden werden die ausgeführten Experimente und ihre Ergebnisse beschrieben. Die zugehörigen Graphiken sind jeweils in den Abbildungen gleicher Nummerierung wie die Experimente zusammengestellt.

### 6.4.1 Anzahl Iterationsschritte des Kernighan-Lin-Algorithmus

B. W. Kernighan und S. Lin geben in [KL70] an, dass auf Graphen der Ordnung  $\leq 360$  im Mittel 2 bis 4 Iterationen notwendig sind bis der Min-Algorithmus ein lokales Minimum erreicht. Es ist nun zu überprüfen, ob dies auch auf den hier zu sektionierenden bipartiten Graphen gilt. Außerdem stellt sich die Frage, wie viele Iterationsschritte beim Max- und MinMax-Algorithmus benötigt werden.

**Experiment 6.1** *Nach wie vielen Iterationsschritten wird bei der Bisektion mit dem Kernighan-Lin-Algorithmus ein lokales Optimum erreicht?*

*Startbisektionen:  $P_a$ ,  $P_b$  und  $P_c$ ;*

*Algorithmen: Min-, Max- und MinMax-Algorithmus.*

Klar ersichtlich ist, dass im Schnitt nicht mehr als 4 Iterationsschritte benötigt werden. Die von B. W. Kernighan und S. Lin empirisch ermittelten 2 bis 4 Iterationen bis ein lokales Minimum erreicht wird, können in diesem Experiment also bestätigt werden. Bei Verwendung der Startbisektionen  $P_a$  und  $P_b$  wird die Iteration im Min-Algorithmus auf größeren Graphen häufig auch mehr als 4 mal durchlaufen. Die Bisektion  $P_c$  scheint hingegen „näher“ an einem lokalen Minimum zu liegen. Dieses wird in den meisten Fällen schon nach 1 bis 2 Iterationsschritten erreicht. Dabei scheint die Zahl an Iterationen sogar unabhängig von der Größe der bearbeiteten Graphen zu sein, was bei den anderen Startbisektionen nicht der Fall war. Dort steigt mit der Ordnung des Graphen auch die Anzahl notwendiger Iterationsschritte.

Betrachtet man die Ergebnisse des Max-Algorithmus so erkennt man sofort, dass mit der Bisektion  $P_a$  ein lokales Maximum als Startbisektion verwendet wird. Es wird kein einziger Iterationsschritt ausgeführt, da der Schnitt dieser Bisektion nach Definition maximal über alle Bisektionen dieser Graphen ist. Bei den beiden anderen Bisektionen  $P_b$  und  $P_c$  sind durchschnittlich 1 bis 3 bzw. 4 Iterationen notwendig. Auch hier wird bei größeren Graphen erst nach längerer Iteration ein lokales Maximum erreicht.

Beim MinMax-Algorithmus ist kaum ein Unterschied zwischen den verschiedenen Startbisektionen zu erkennen. Alle scheinen gleich „nah“ an einem lokalen Optimum zu liegen. Im Schnitt werden 1 bis 3 Iterationsschritte benötigt.

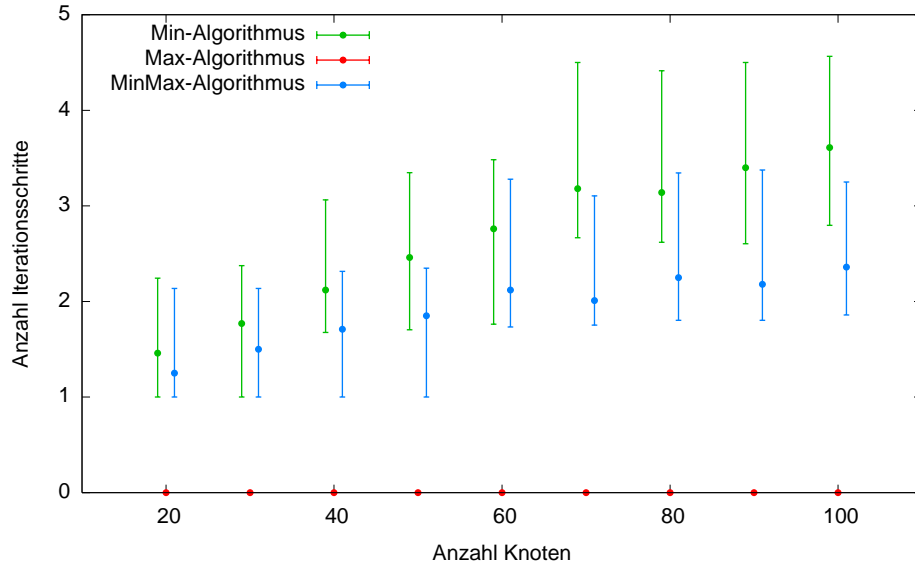
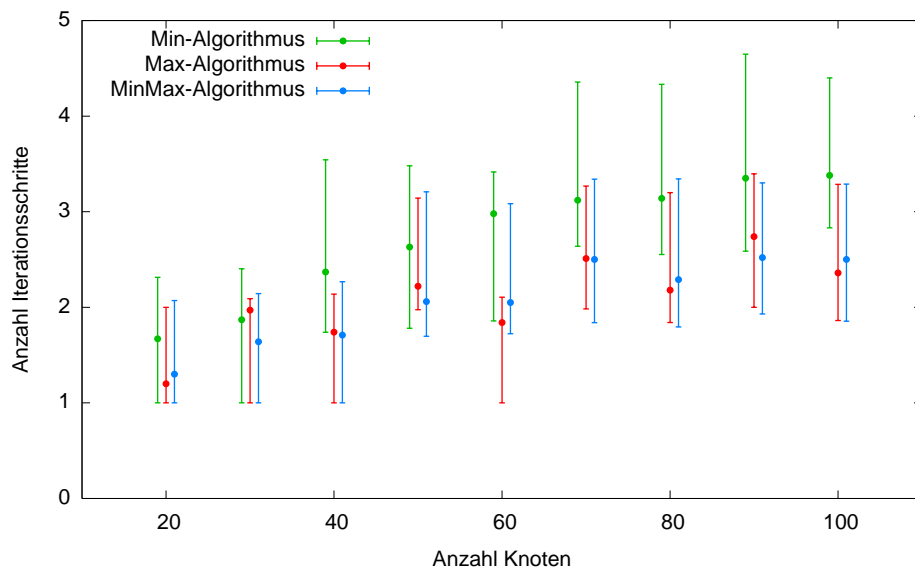
Bilden die lokalen Optima, die hier nach unterschiedlich vielen Iterationsschritten ausgegeben werden, aber auch „gute“ Bisektionen? Der Frage, ob sich die eine oder die andere Bisektion besser als Startbisektion des Kernighan-Lin-Algorithmus eignet, wird in Experiment 6.4 in Abschnitt 6.4.3 nachgegangen.

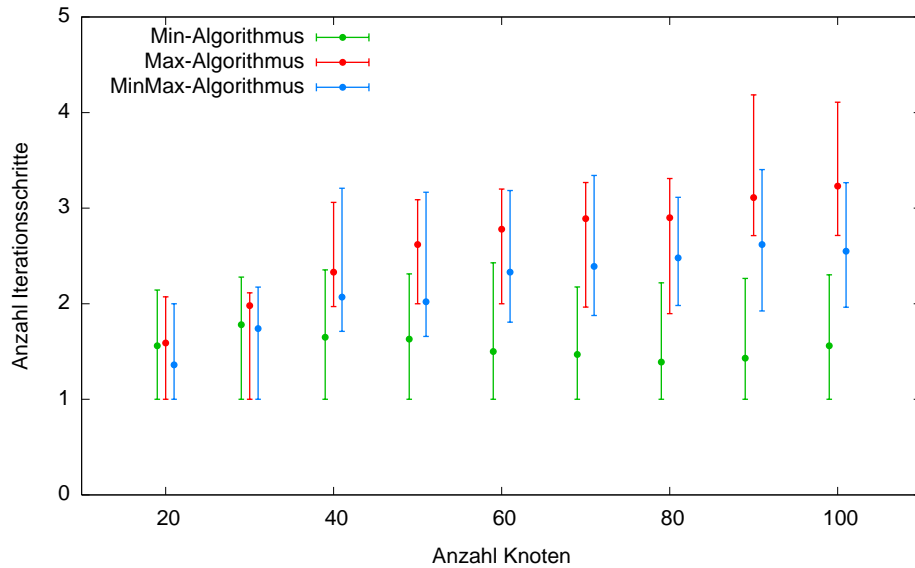
## 6.4.2 Anzahl Durchläufe der Mehr-Wege-Methode

Zur Untersuchung, nach wie vielen Durchläufen die Mehr-Wege-Methode ein lokales Optimum erreicht, werden zwei Experimente durchgeführt. Das Erste bildet für alle Graphen eine 4-Sektionierung. B. W. Kernighan und S. Lin schreiben in [KL70], dass bei der  $k$ -Sektionierung von Graphen der Ordnung  $\leq 100$  mit  $k < 6$  weniger als 5 Durchläufe benötigt werden.

**Experiment 6.2** *Nach wie vielen Durchläufen der Mehr-Wege-Methode zur 4-Sektionierung wird ein lokales Optimum erreicht?*

Abbildung 6.1: Experiment 6.1

(a) Startbisektion  $P_a$ (b) Startbisektion  $P_b$

(c) Startbisektion  $P_c$ 

*Start-k-Sektionen:  $P_a, P_b$  und  $P_c$ ;*

*MW-Durchlaufreihenfolgen: fest, zufällig, sektionsoptimiert;*

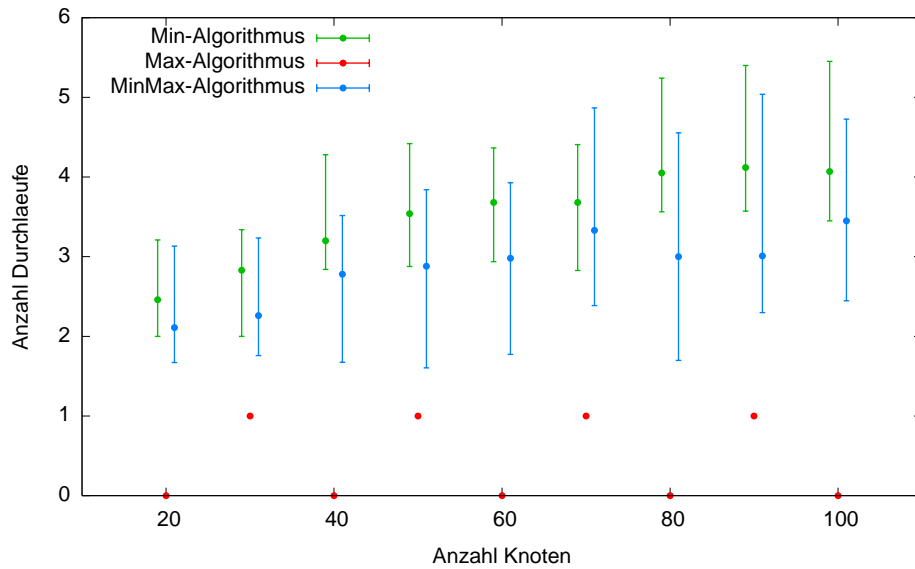
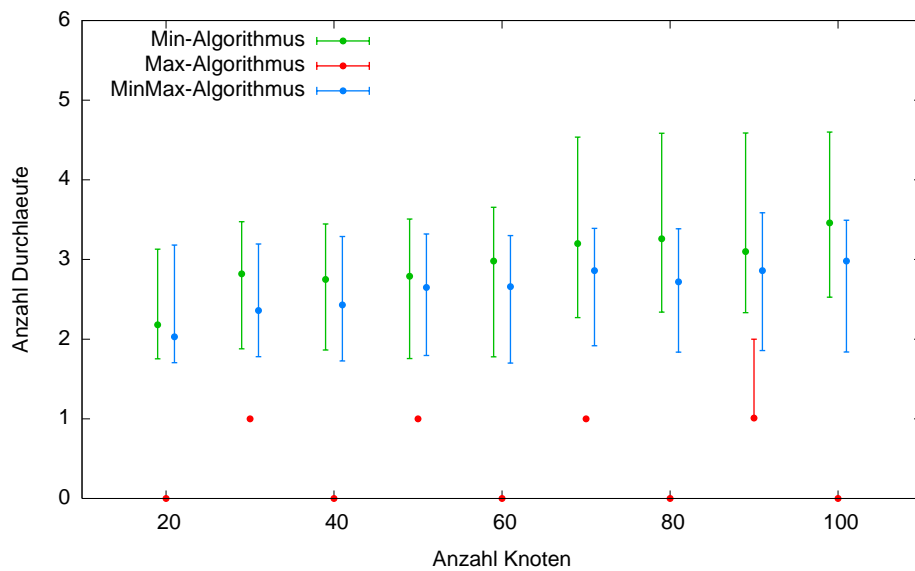
*Algorithmen: Min-, Max- und MinMax-Algorithmus.*

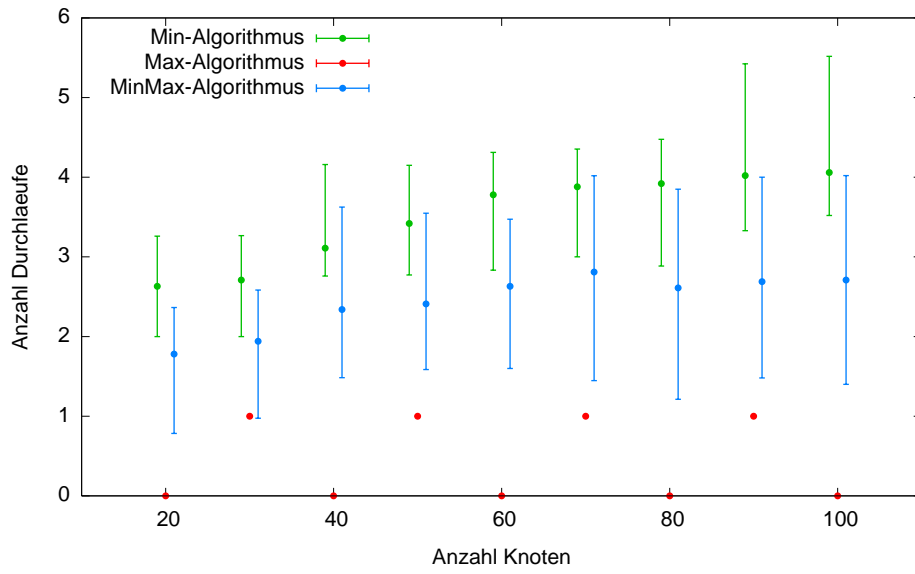
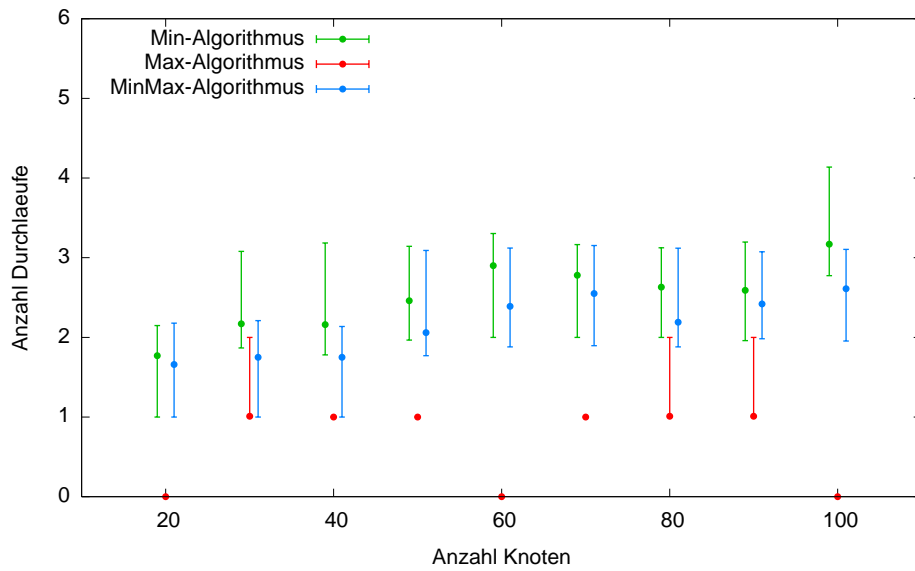
Es fällt als Erstes auf, dass der Max-Algorithmus generell schneller ein lokales Optimum erreicht als die anderen beiden Algorithmen. Man erkennt, dass die Startsektion  $P_a$  sogar selber ein lokales Maximum ist, falls die Ordnung des Graphen durch 4 teilbar ist. Nach Definition bestehen die einzelnen Sektionselemente dann nämlich ausschließlich aus Knoten der einen Menge des bipartiten Graphen. Falls die Knotenzahl  $n$  kein Vielfaches von 4 ist, ist stets genau ein Durchlauf notwendig. Dies hängt wohl damit zusammen, dass die Startsektion in diesem Fall ein gruppenbalanciertes Element enthält. Die Startsektion scheint folglich nach einer Anwendung des Kernighan-Lin-Algorithmus auf die Kombination dieses Elementes mit einem beliebigen anderen Sektionselement lokal maximal zu sein. Auch bei Verwendung der Startsektion  $P_b$  kommen lokale Maxima zum Vorschein. Bei den Knotenzahlen 20, 60 und 100 ist ebenfalls keine Verbesserung der Sektion durch die Anwendung der Mehr-Wege-Methode möglich. Bei den übrigen Graphen ist häufig nur ein Durchlauf zum Erreichen eines lokalen Maximums notwendig.

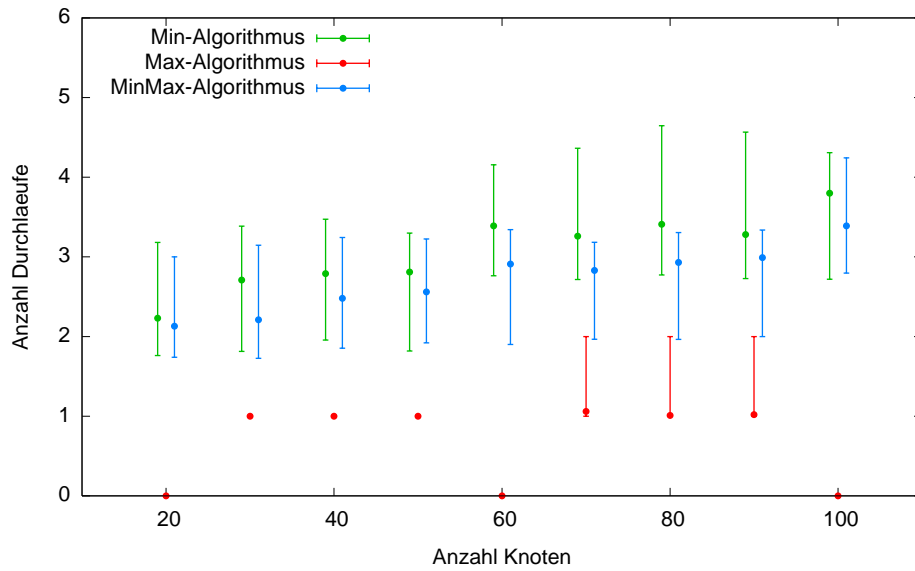
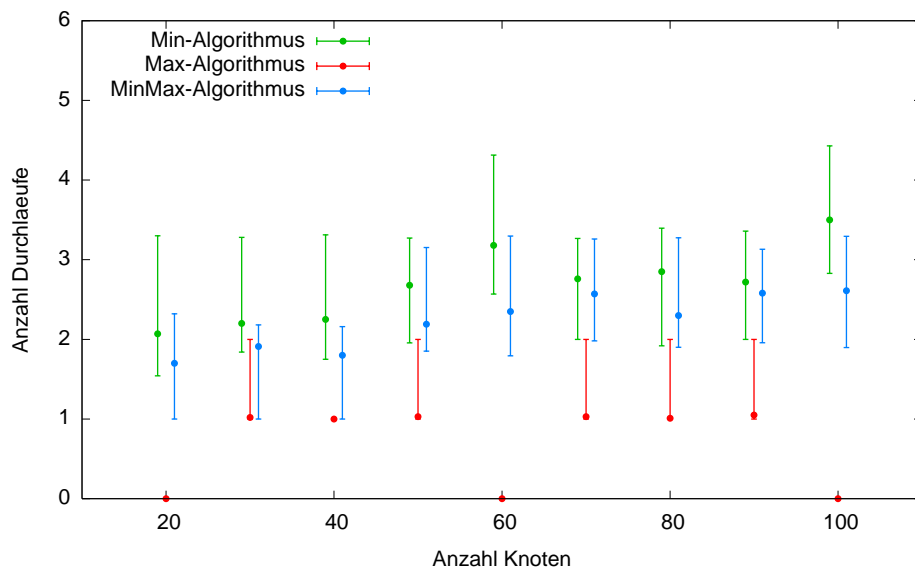
Für den Min- und den MinMax-Algorithmus werden mehr Durchläufe benötigt, meist 2 bis 5 bzw. 4. Betrachtet man die Startsektionen  $P_a$  und

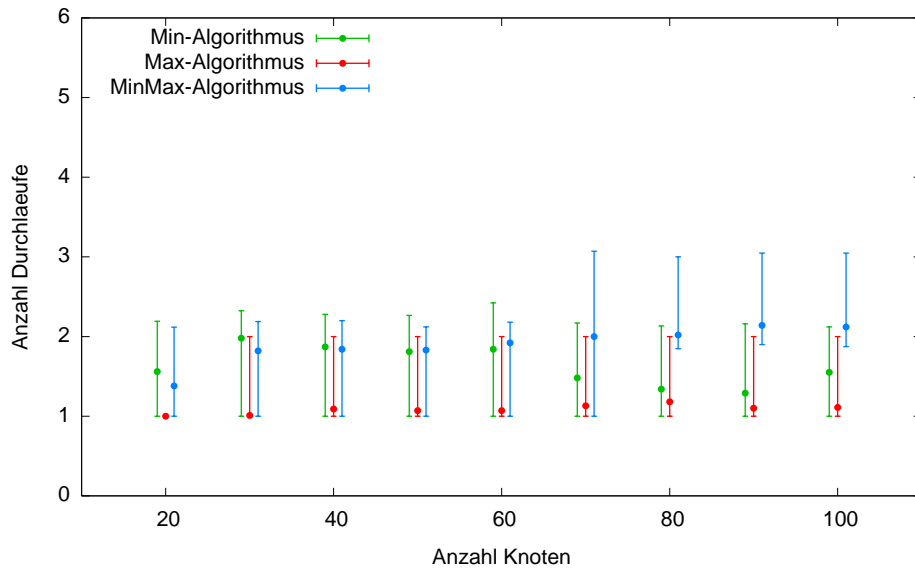
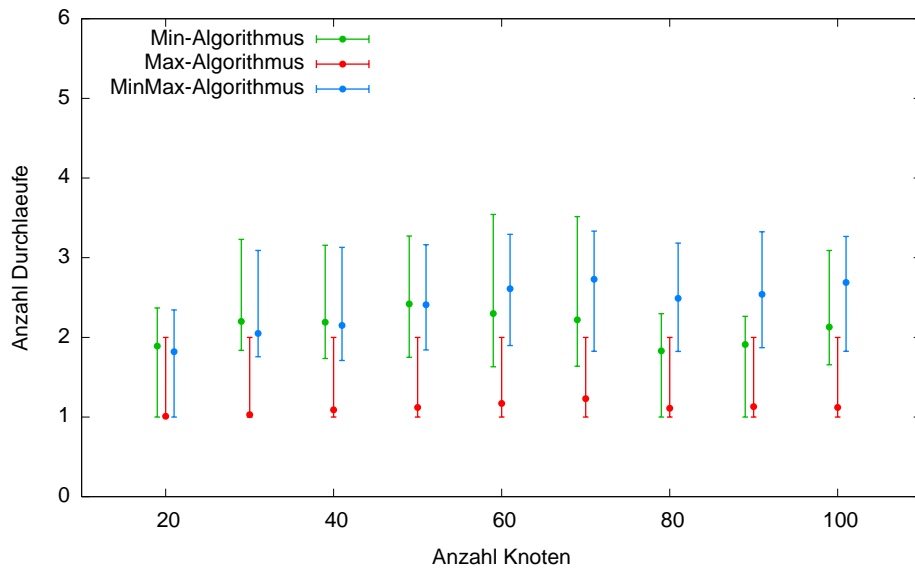


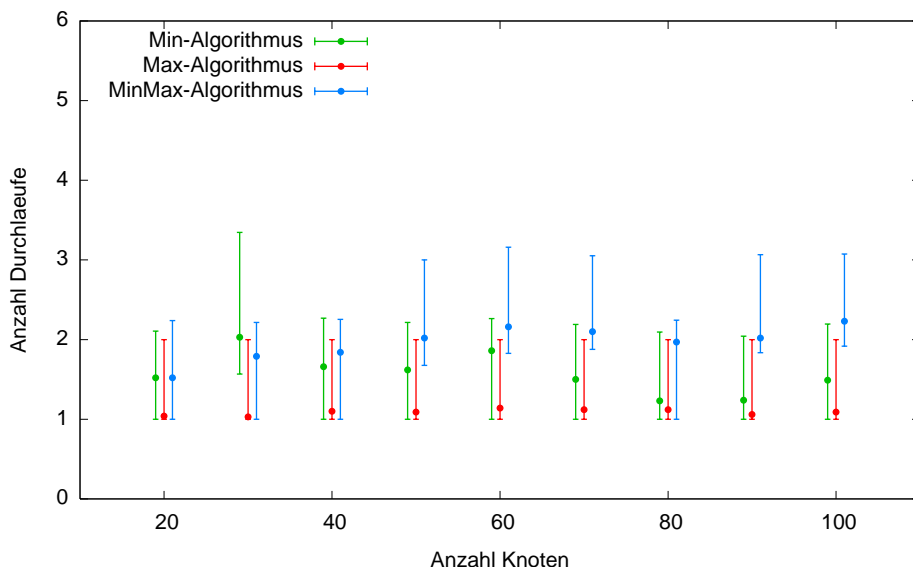
Abbildung 6.2: Experiment 6.2

(a) Start- $k$ -Sektion  $P_a$ , feste Reihenfolge(b) Start- $k$ -Sektion  $P_a$ , zufällige Reihenfolge

(c) Start- $k$ -Sektion  $P_a$ , sektionsoptimierte Reihenfolge(d) Start- $k$ -Sektion  $P_b$ , feste Reihenfolge

(e) Start- $k$ -Sektion  $P_b$ , zufällige Reihenfolge(f) Start- $k$ -Sektion  $P_b$ , sektionsoptimierte Reihenfolge

(g) Start- $k$ -Sektion  $P_c$ , feste Reihenfolge(h) Start- $k$ -Sektion  $P_c$ , zufällige Reihenfolge

(i) Start- $k$ -Sektion  $P_c$ , sektionsoptimierte Reihenfolge

$P_b$ , so braucht der Min-Algorithmus gegenüber dem MinMax-Algorithmus häufig mehr Schritte. Während bei Verwendung von  $P_b$  schon nach weniger Durchläufen als bei  $P_a$  ein lokales Optimum erreicht wird, ist die Anzahl Durchläufe bei  $P_c$  noch geringer; sie beträgt in den meisten Fällen weniger als 2. Bei  $P_c$  und  $P_c$  fällt zusätzlich auf, dass beide Algorithmen, der Min- und der MinMax-Algorithmus, bei zufälliger Durchlaufreihenfolge mehr Durchläufe benötigen. Dies hängt wohl damit zusammen, dass in jedem Durchlauf unterschiedliche Elemente miteinander kombiniert werden und jeweils eine Optimierung mit dem Kernighan-Lin-Algorithmus möglich ist.

Das zweite Experiment ermittelt die Anzahl Durchläufe bei gleichmäßiger Sektionierung, das heißt, bei  $k$ -Sektionierungen mit unterschiedlich großen Werten von  $k$ .

**Experiment 6.3** *Nach wie vielen Durchläufen der Mehr-Wege-Methode zur  $k$ -Sektionierung ( $k$  nach Tabelle 6.2) wird ein lokales Optimum erreicht?*

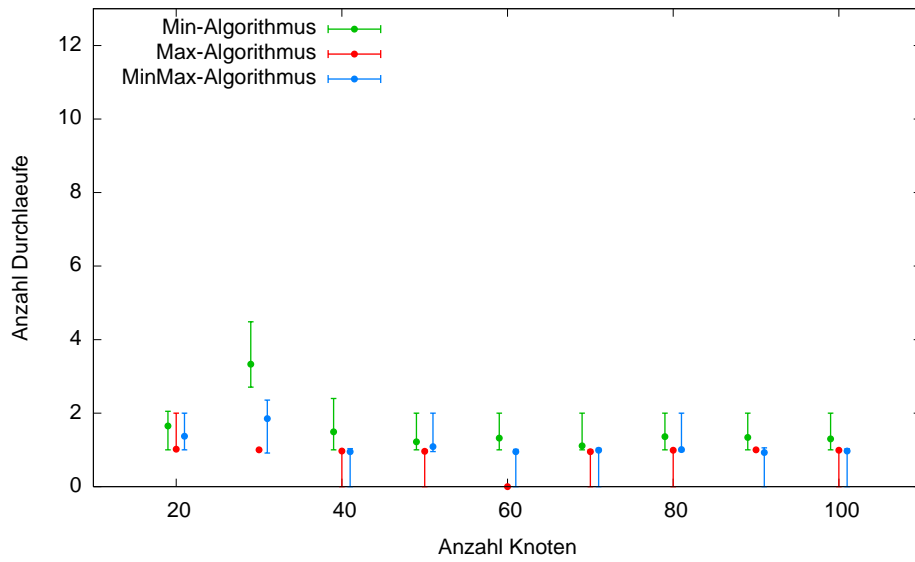
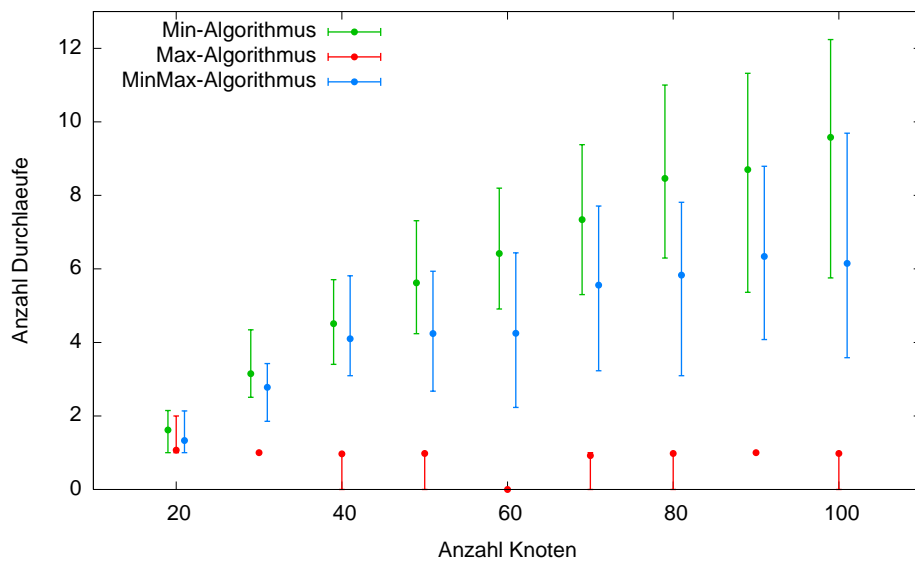
*Start- $k$ -Sektionen:  $P_a, P_b$  und  $P_c$ ;*

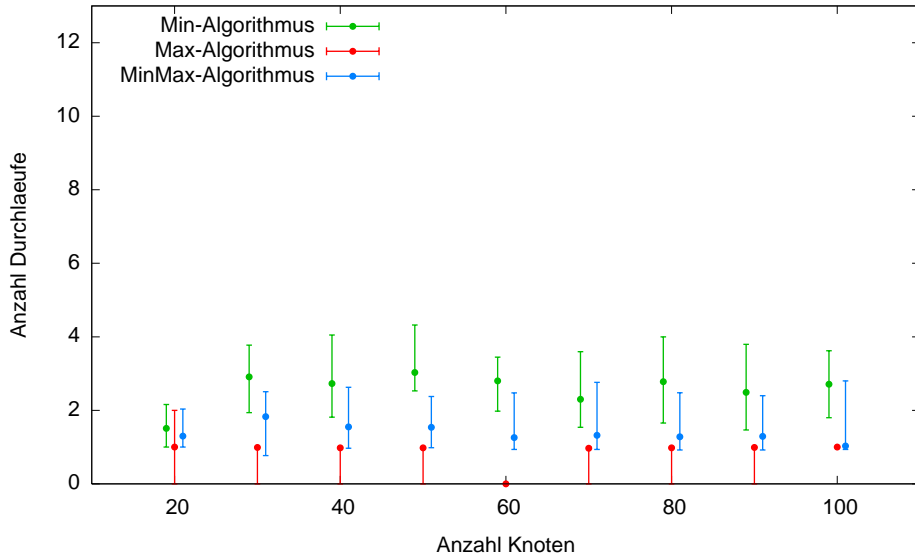
*MW-Durchlaufreihenfolgen: fest, zufällig, sektionsoptimiert;*

*Algorithmen: Min-, Max- und MinMax-Algorithmus.*

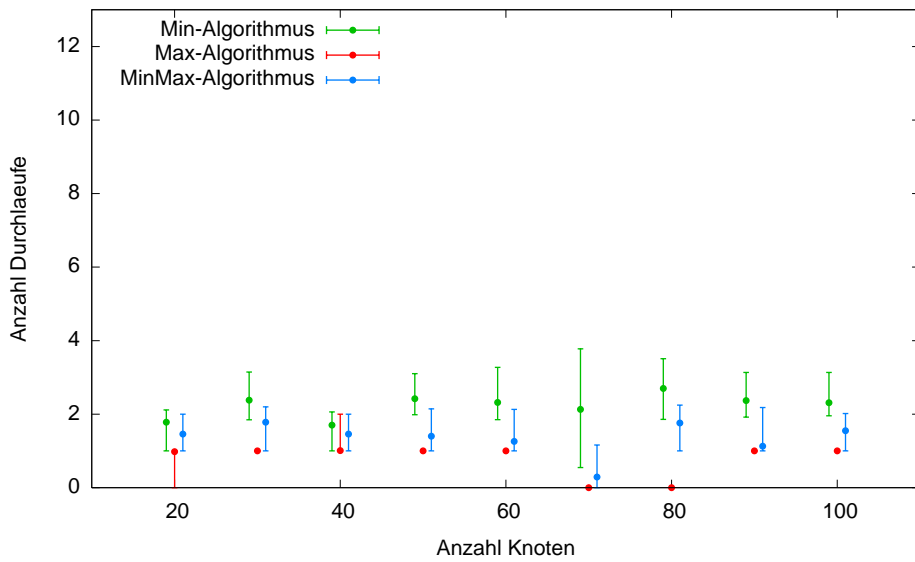
Wird die Gruppeneinteilung also derart vorgenommen, dass möglichst Gruppen mit 6 Personen entstehen, so fallen starke Unterschiede bei der

Abbildung 6.3: Experiment 6.3

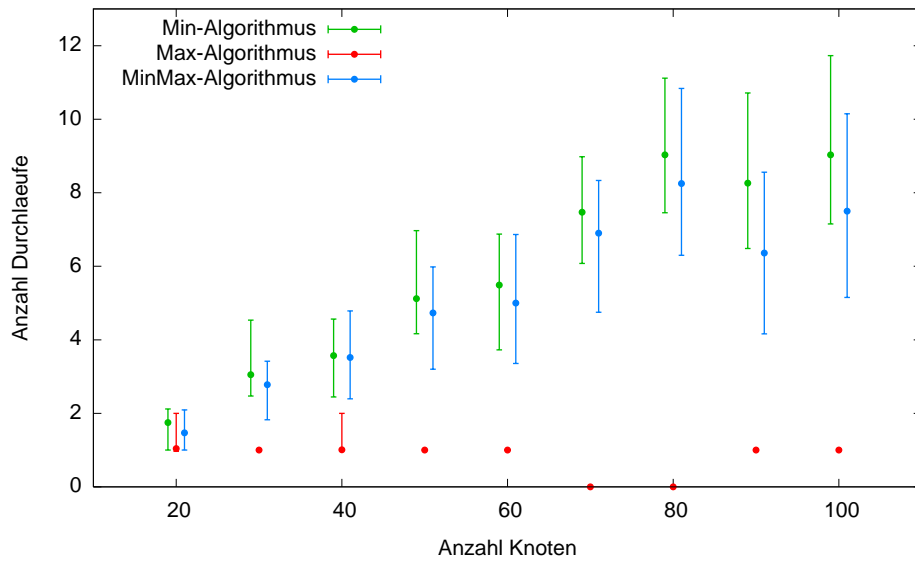
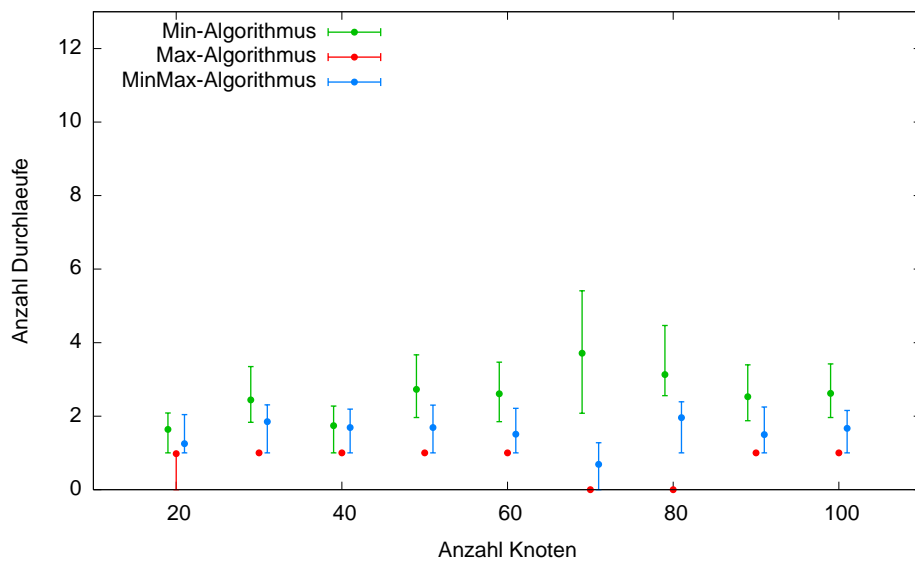
(a) Start- $k$ -Sektion  $P_a$ , feste Reihenfolge(b) Start- $k$ -Sektion  $P_a$ , zufällige Reihenfolge



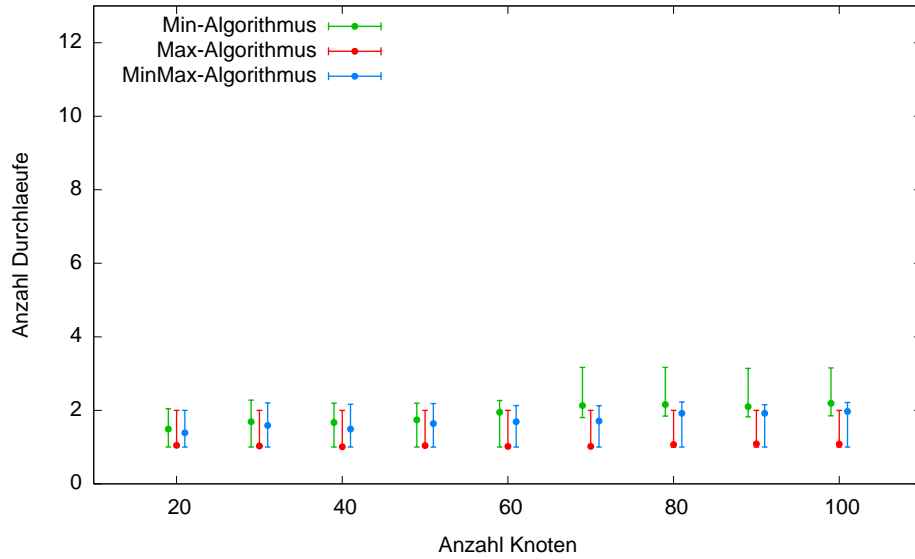
(c) Start- $k$ -Sektion  $P_a$ , sektionsoptimierte Reihenfolge



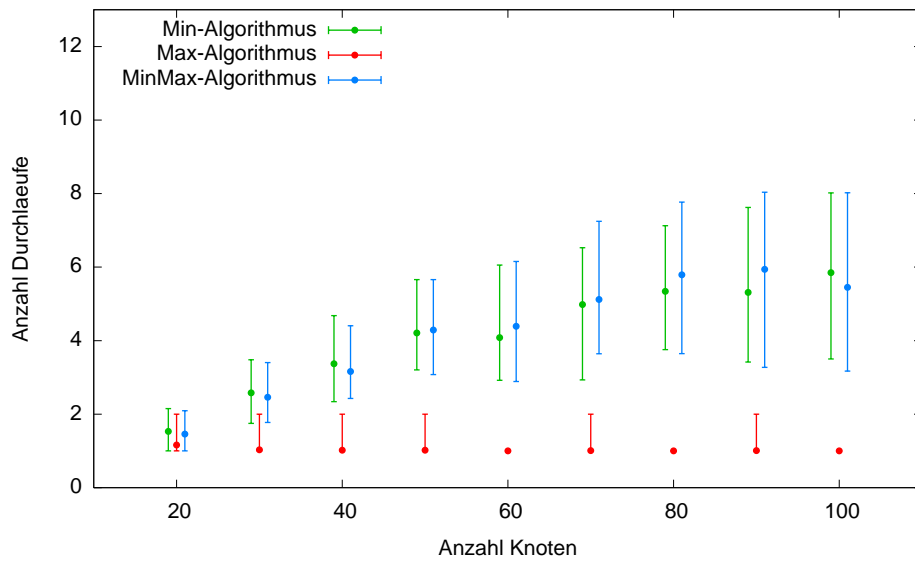
(d) Start- $k$ -Sektion  $P_b$ , feste Reihenfolge

(e) Start- $k$ -Sektion  $P_b$ , zufällige Reihenfolge(f) Start- $k$ -Sektion  $P_b$ , sektionsoptimierte Reihenfolge

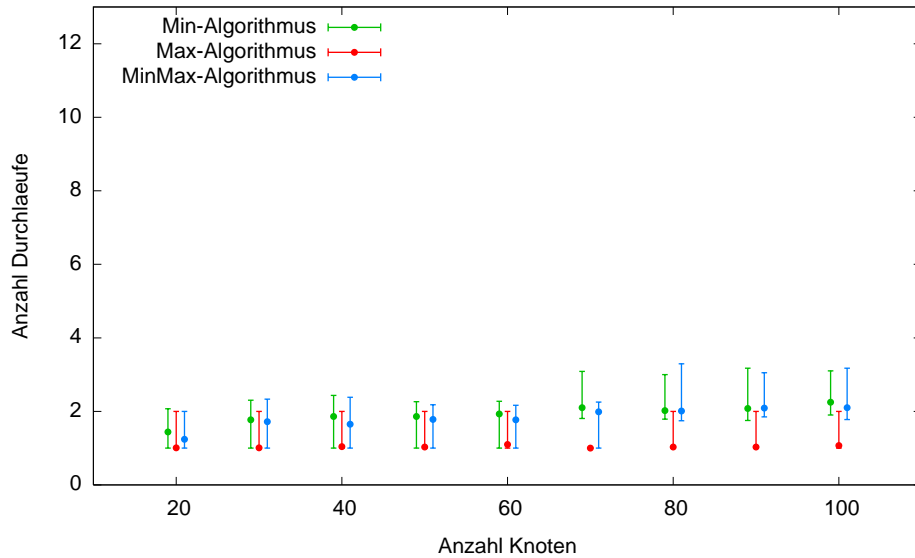




(g) Start- $k$ -Sektion  $P_c$ , feste Reihenfolge



(h) Start- $k$ -Sektion  $P_c$ , zufällige Reihenfolge

(i) Start- $k$ -Sektion  $P_c$ , sektionsoptimierte Reihenfolge

Verwendung verschiedener Durchlaufreihenfolgen auf. Bei der Mehr-Wege-Sektionierung mit zufälliger Reihenfolge werden auf großen Graphen bis zu 10 bzw. 8 Durchläufe im Min- bzw. MinMax-Algorithmus benötigt. Wie im vorherigen Experiment erläutert, hängt dies mit den unterschiedlichen Elementkombinationen zusammen, für welche es natürlich mehr Möglichkeiten gibt, wenn mehr Elemente zur Verfügung stehen. Dabei erreicht der MinMax-Algorithmus stets etwas schneller ein lokales Optimum als der Min-Algorithmus. Die Startsektionen scheinen demnach „näher“ an einem lokalen Optimum der MinMax-Sektionierung zu liegen als an einem lokalen Minimum.

Der Max-Algorithmus stoppt meist nach nur einem Durchlauf, egal welche Startsektion oder Durchlaufreihenfolge man wählt. Interessant ist, dass bei den Graphen der Ordnung 70 und 80 die Sektion  $P_b$  ein lokales Maximum ist. Da hier jedoch die Ergebnisse des MinMax-Algorithmus stärker in Betracht gezogen werden sollen, wird im Weiteren nicht näher untersucht, ob diese lokalen Maxima „gute“ maximale Antipathieschnitte sind.

### 6.4.3 Startbisektionen des Kernighan-Lin-Algorithmus

Bei diesem Experiment soll die Frage beantwortet werden, wie stark der Einfluß der Startbisektion auf die Ergebnisse des Kernighan-Lin-Algorithmus ist. Um die Ergebnisse der Algorithmen unter Verwendung der verschiedenen Startbisektionen  $P_a$ ,  $P_b$  und  $P_c$  miteinander zu vergleichen, wird zuerst untersucht, wie stark die jeweils erzielten Schnitte vom absoluten Optimum abweichen.

**Experiment 6.4** *Wie groß ist die Abweichungen vom absoluten Optimum bei Verwendung verschiedener Startbisektion im Kernighan-Lin-Algorithmus? Startbisektionen:  $P_a$ ,  $P_b$  und  $P_c$ ; Algorithmen: Min-, Max- und MinMax-Algorithmus.*

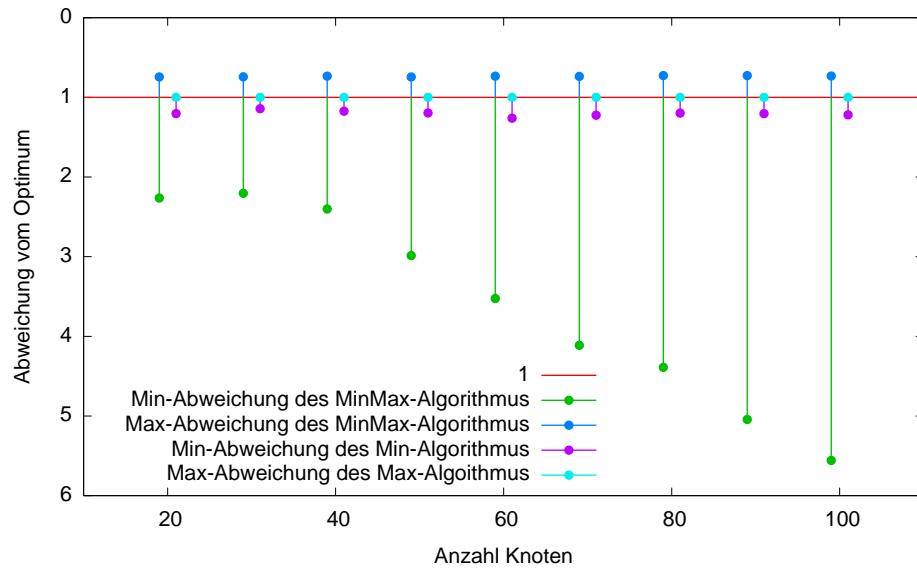
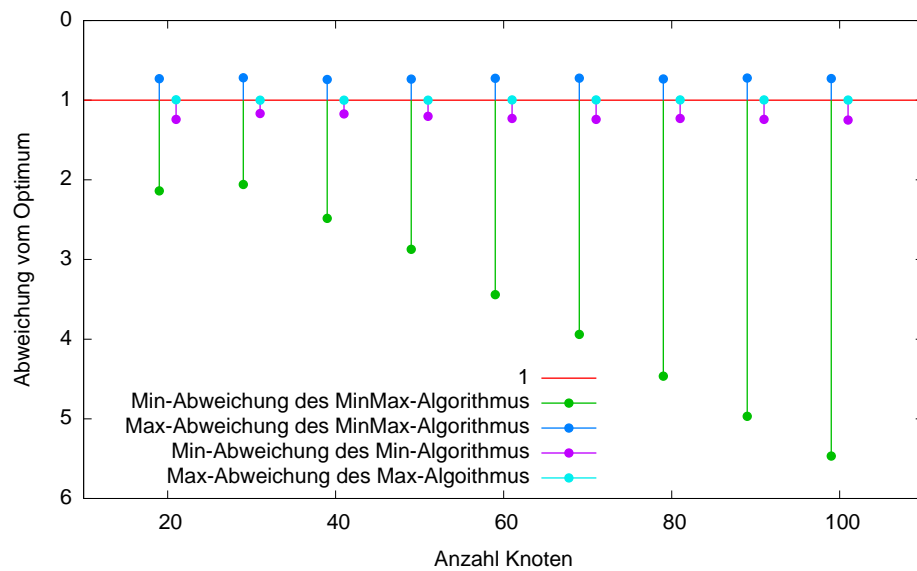
Man erkennt, dass sich die Graphiken zu den drei Startbisektionen kaum unterscheiden. Die Startbisektionen scheinen demnach keinen Einfluß auf die Ergebnisse des Kernighan-Lin-Algorithmus unter Verwendung verschiedener Fokusse zu haben.

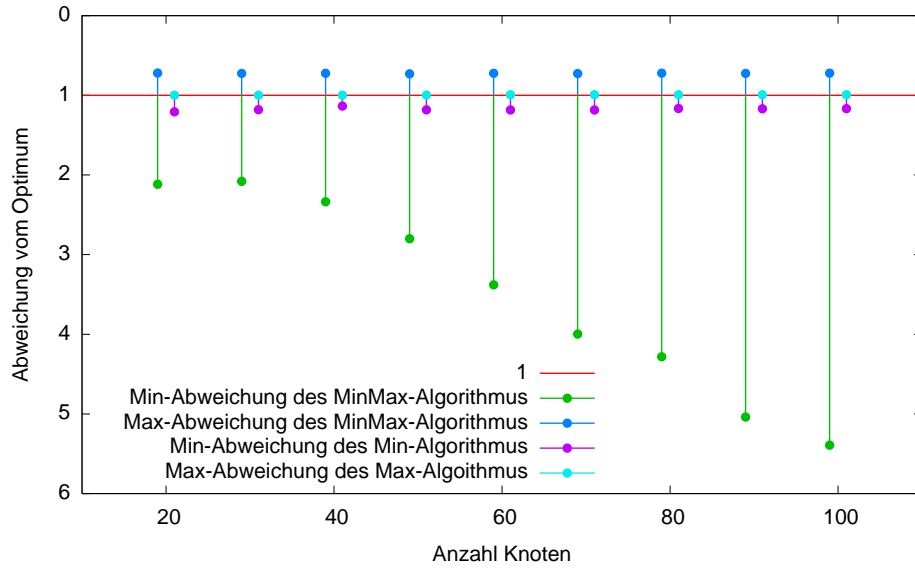
Das Experiment zeigt im Einzelnen, dass die Ergebnisse des Min-Algorithmus proportional zum absoluten Minimum nur geringfügig größer sind als das absolute Minimum. Die Größe dieser Abweichung ist unabhängig von der Ordnung des bearbeiteten Graphen. Bei den Ergebnissen des Max-Algorithmus stimmen hingegen alle mit dem absoluten Maximum überein. Das heißt, die Erfolgsrate des Kernighan-Lin-Algorithmus mit Fokus „Maximierung“ scheint 1 zu sein. Man beobachte nun die Ergebnisse des MinMax-Algorithmus: Mit der Größe der Graphen wächst die proportionale Abweichung des Sympathieschnittes vom absoluten Minimum. Bei Graphen der Ordnung 100 ist der lokal optimale Sympathieschnitt durchschnittlich sogar 6 mal so groß wie das absolute Minimum. Der lokal maximale Antipathieschnitt weicht hingegen proportional nur wenig vom absoluten Maximum ab. Das bedeutet, die Differenz zwischen dem lokal optimalen Antipathieschnitt und dem absoluten Maximum wächst proportional zum absoluten Maximum.

Man beobachte nun die Schnitte der Ergebnissektionen, die der Min- und der MinMax-Algorithmus liefern. Da obiges Experiment ergeben hat, dass die Verwendung der drei verschiedenen Startbisektionen keine stark unterschiedlichen Lösungen liefern, wird das Folgende nur für  $P_c$  ausgeführt.

**Experiment 6.5** *Welche Lösungen liefert die Bisektion mit dem Kernighan-Lin-Algorithmus? Startbisektion:  $P_c$ ; Algorithmen: Min- und MinMax-Algorithmus.*

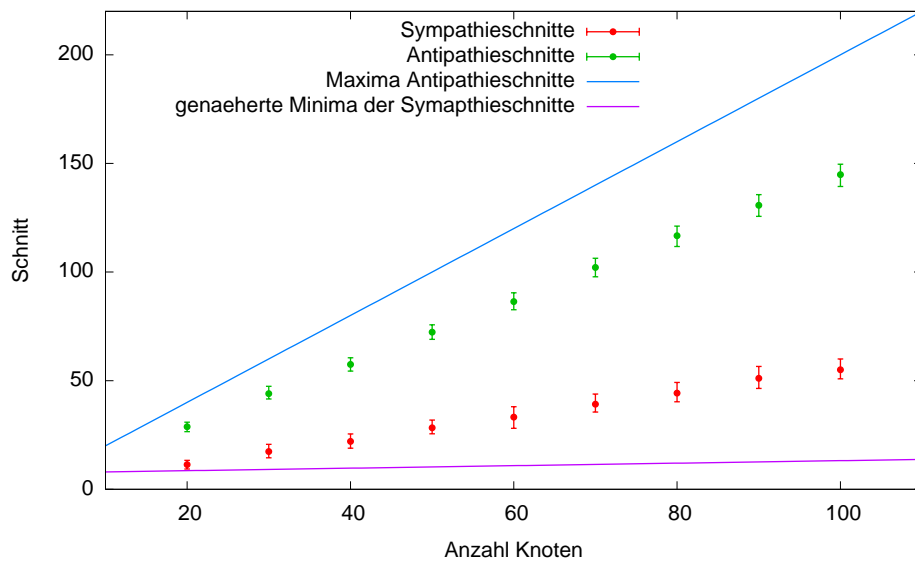
Abbildung 6.4: Experiment 6.4

(a) Startbisektion  $P_a$ (b) Startbisektion  $P_b$



(c) Startbisektion  $P_c$

Abbildung 6.5: Experiment 6.5



Der Min-Algorithmus liefert bei der Bisektionierung fast unabhängig von der Größe der Graphen Sympathieschnitte der Größe 10. Die an die absoluten Minima genäherte Gerade hat nur eine sehr geringe Steigung. Betrachtet man die Ergebnisse des MinMax-Algorithmus in Abhängigkeit der Größe der Graphen, so erkennt man auch in Graphik 6.5, dass die Abweichung des Sympathieschnittes vom absoluten Minimum proportional zur Größe des Minimums stark wächst. Die Differenz zwischen den Schnitten des lokalen Optimum und dem jeweiligen absolutem Optimum scheint allerdings beim Sympathie- und Antipathieschnitt gleich stark zu wachsen.

#### 6.4.4 Startbisektionen der Zwei-Wege-Methode

Bei der Zwei-Wege-Methode werden durch  $r$  rekursive Aufrufe des Kernighan-Lin-Algorithmus lokal optimale  $2^r$ -Sektionen gebildet. Auch wenn die Ausgaben der Bisektionierung mit diesem Algorithmus, wie im Experiment 6.4 gezeigt, unabhängig von den verwendeten Startbisektionen gleich sind, sollte für die Zwei-Wege-Methode die Verwendung der verschiedenen Startbisektionen untersucht werden.

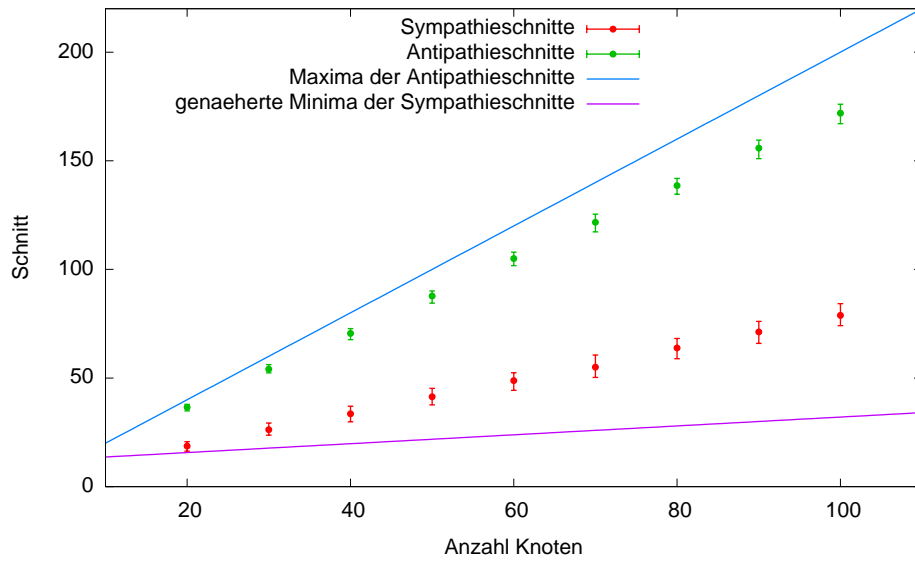
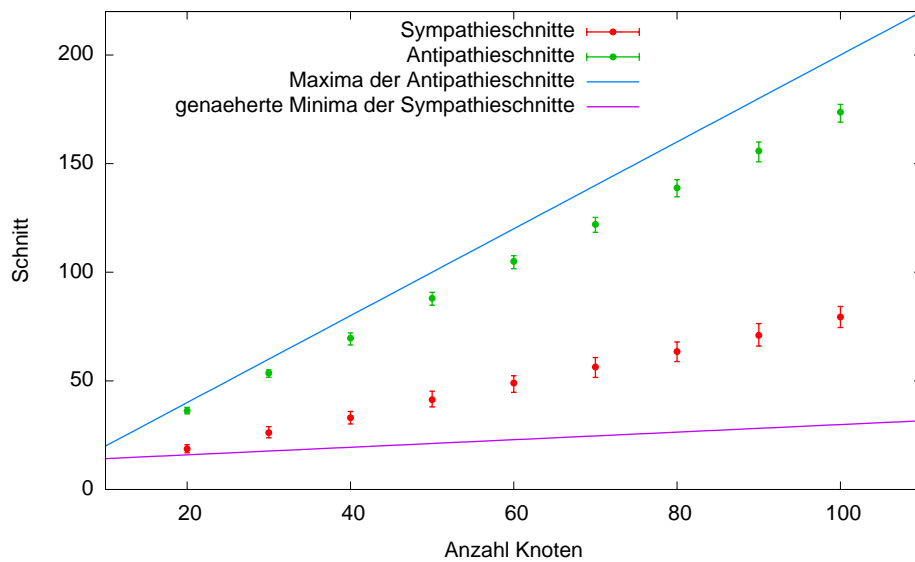
**Experiment 6.6** *Welche Lösungen liefern die verschiedenen Startbisektion bei der Zwei-Wege-Methode?*

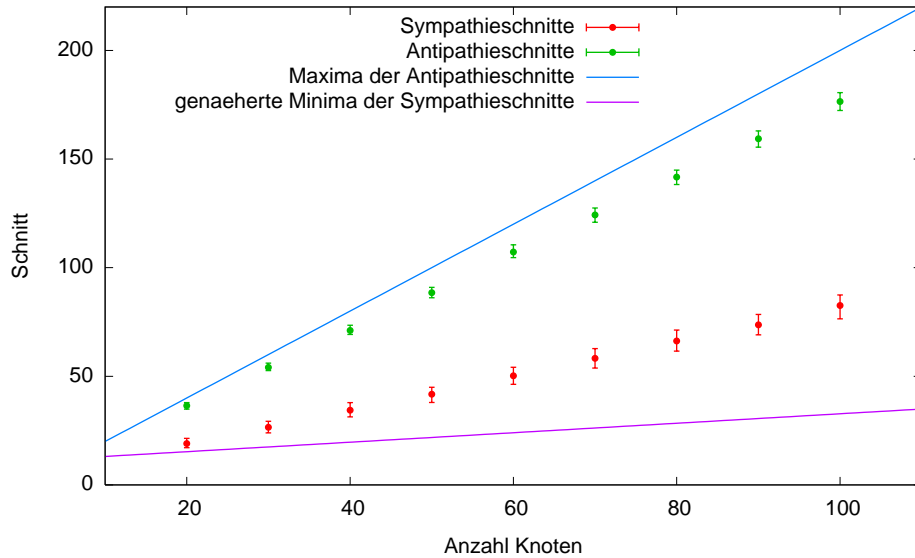
*Startbisektionen:  $P_a$ ,  $P_b$  und  $P_c$ ;*

*Algorithmen: Min- und MinMax-Algorithmus.*

Auch hier kann man keine Abhängigkeit von der verwendeten Startbisektion feststellen. Die an die absoluten Minima genäherte Gerade hat hier allerdings eine größere Steigung als bei der Bisektionierung des vorherigen Experimentes. Dies kommt daher, dass eine Sektionierung in mehr als zwei Elemente, wie sie bei allen Graphen ausgeführt wird, auch einen größeren Schnitt verursacht. Der MinMax-Algorithmus liefert ebenfalls bei allen Startbisektionen  $P_a$ ,  $P_b$  und  $P_c$  Sympathie- und Antipathieschnitte gleicher Größen. Die Antipathieschnitte weichen hier allerdings deutlich weniger von den absoluten Maxima ab, als die Sympathieschnitte von den absoluten Minima. Die Sympathieschnitte sind auf Graphen der Ordnung 100 sogar mehr als doppelt so groß wie das absolute Minimum. Der mit dem MinMax-Algorithmus erzielten Sektionen scheinen demnach bei der  $k$ -Sektionierung mit  $k > 2$  deutlich näher am absoluten Minimum zu sein, als dies bei der Bisektionierung der Fall war.

Abbildung 6.6: Experiment 6.6

(a) Startbisektion  $P_a$ (b) Startbisektion  $P_c$

(c) Startbisektion  $P_b$ 

### 6.4.5 Startsektionen der Mehr-Wege-Methode

Nun werden die Ergebnisse der Mehr-Wege-Methoden unter Eingabe der drei verschiedenen Start- $k$ -Sektionen miteinander verglichen. Dabei werden wieder alle drei Durchlaufreihenfolgen eingesetzt.

**Experiment 6.7** Welche Lösungen liefern die verschiedenen Start- $k$ -Sektion bei der Mehr-Wege-Methode?

Start- $k$ -Sektionen:  $P_a, P_b$  und  $P_c$ ;

MW-Durchlaufreihenfolgen: fest, zufällig, sektionsoptimiert;

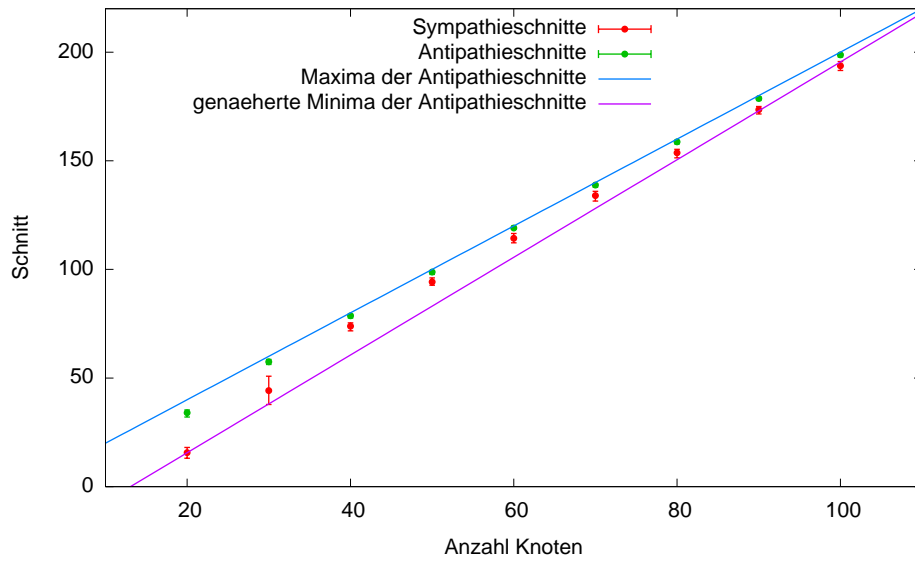
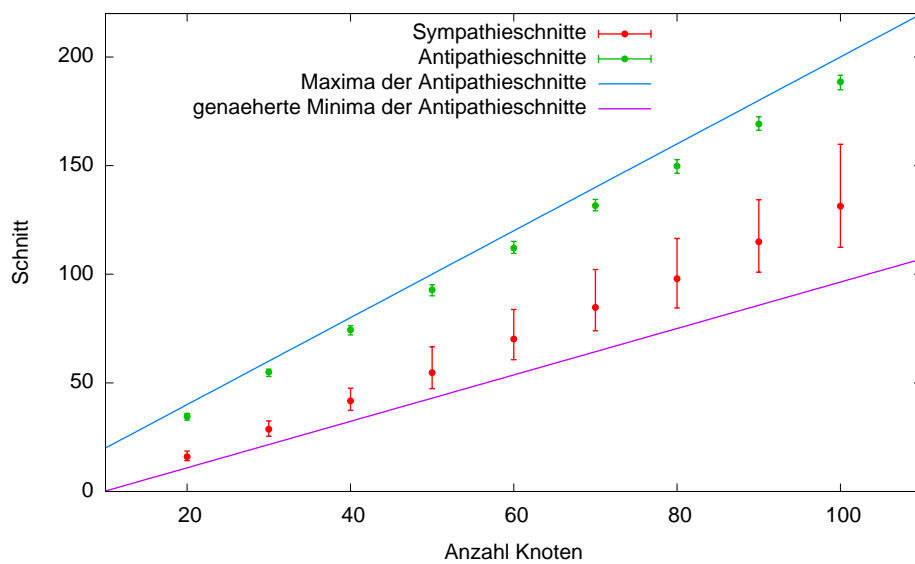
Algorithmen: Min- und MinMax-Algorithmus.

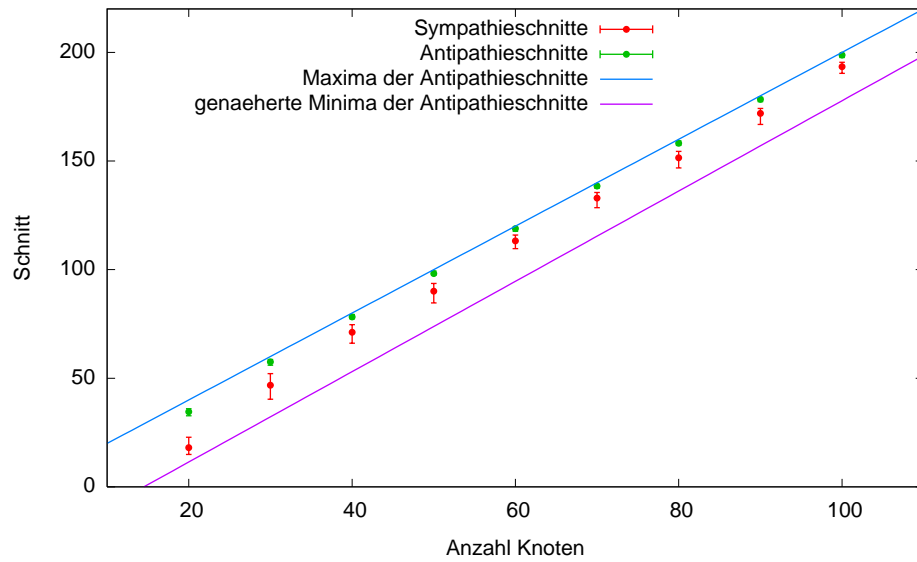
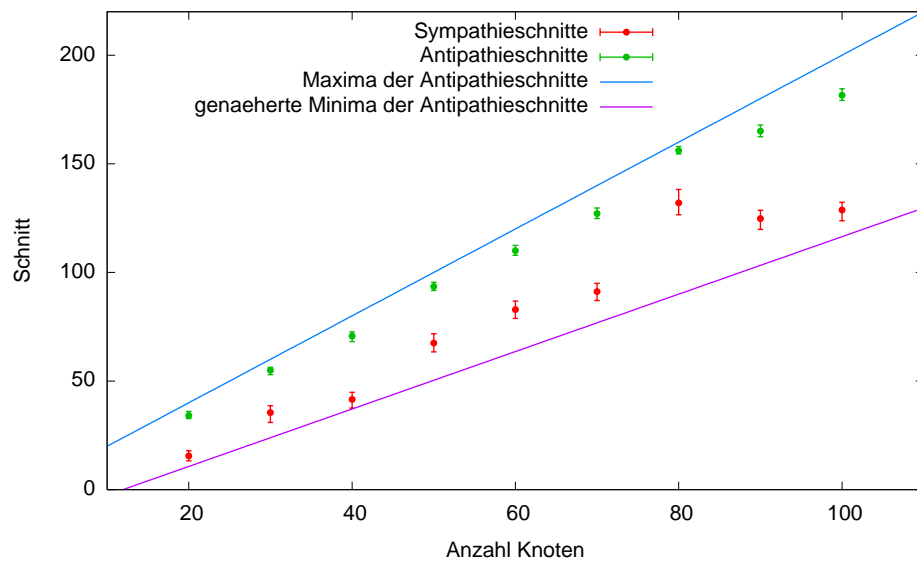
Vergleicht man als Erstes die Graphiken der Abbildungen 6.6 und 6.7 miteinander, so fällt auf, dass die an die absoluten Minima genäherte Gerade auch im besten Fall bei der Mehr-Wege-Methode eine deutlich größere Steigung als die bei der Zwei-Wege-Methode hat. Obwohl bei der Mehr-Wege-Methode bei Graphen der Ordnung 100 eine nur um eins größere Sektion gebildet wird, ist das absolute Minimum hier mehr als doppelt so groß. Welche der beiden Methoden bei gleich großer Sektionierung bessere Ergebnisse liefert, wird in Abschnitt 6.4.6 untersucht.

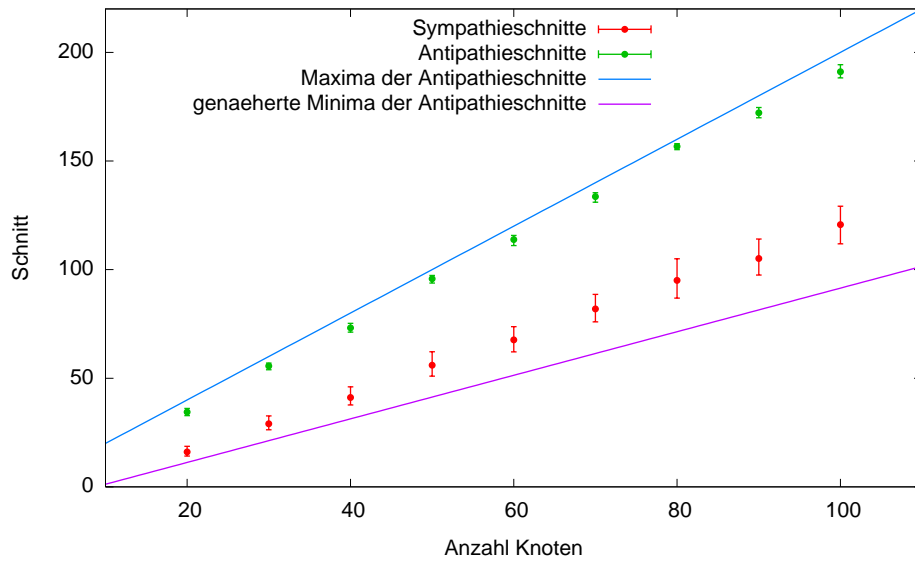
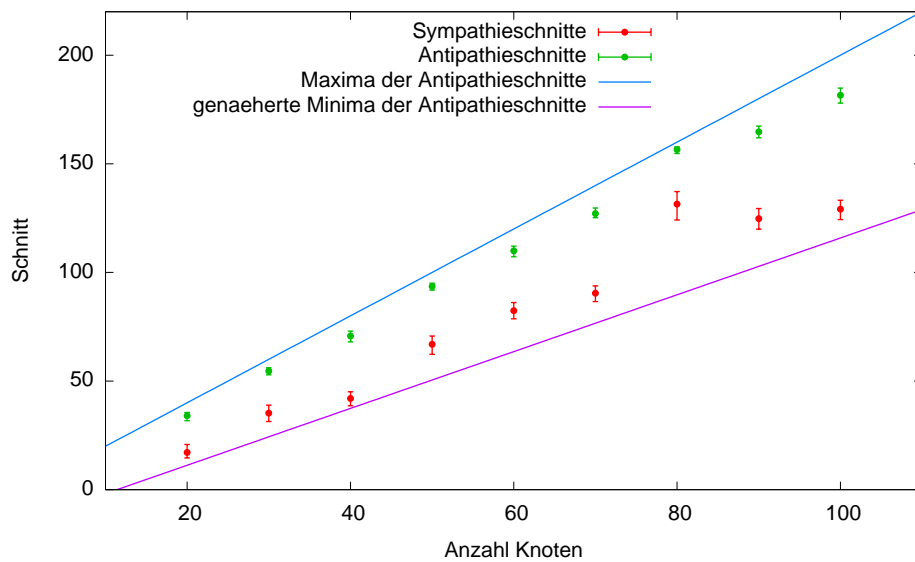
Die  $k$ -Sektion  $P_a$  scheint aufgrund ihres großen Schnittes auf beliebigen bipartiten Graphen zur Verwendung als Startsektion eher ungeeignet.

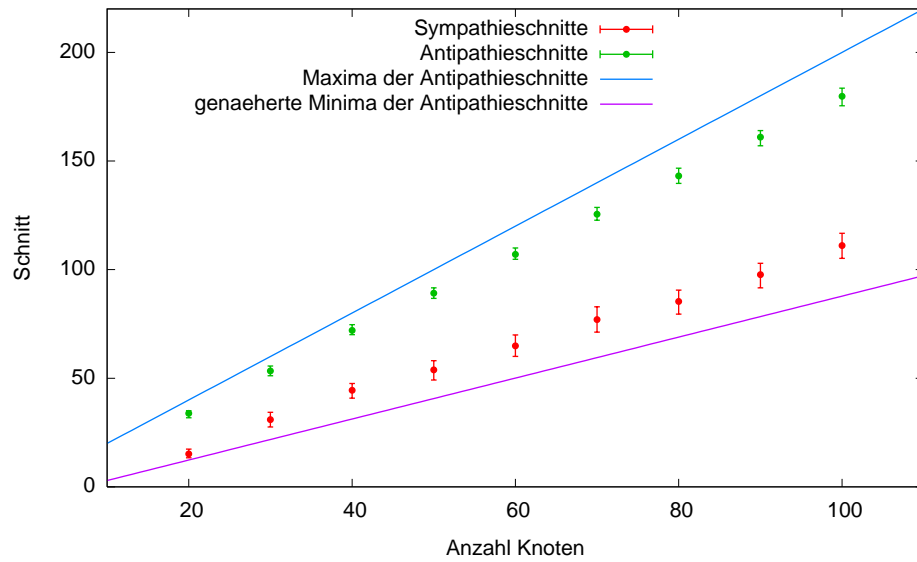
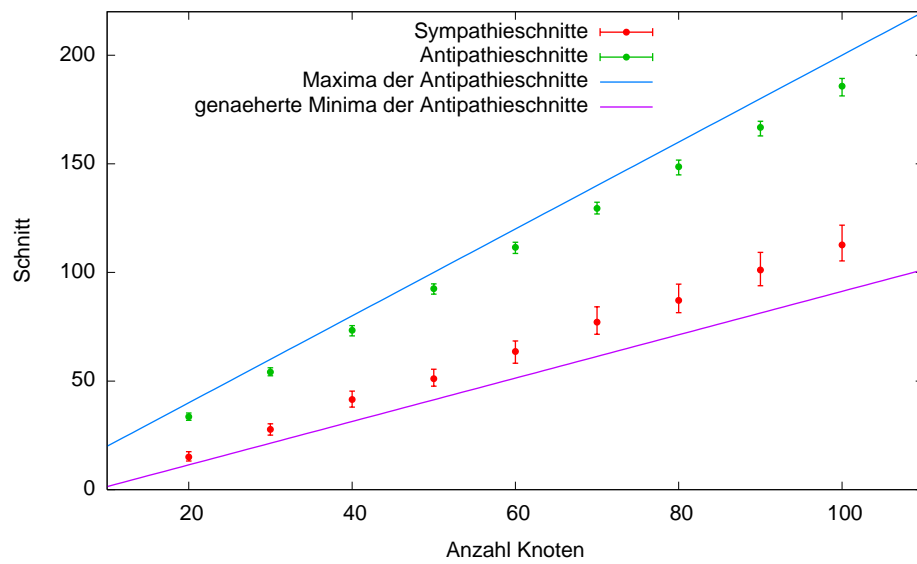


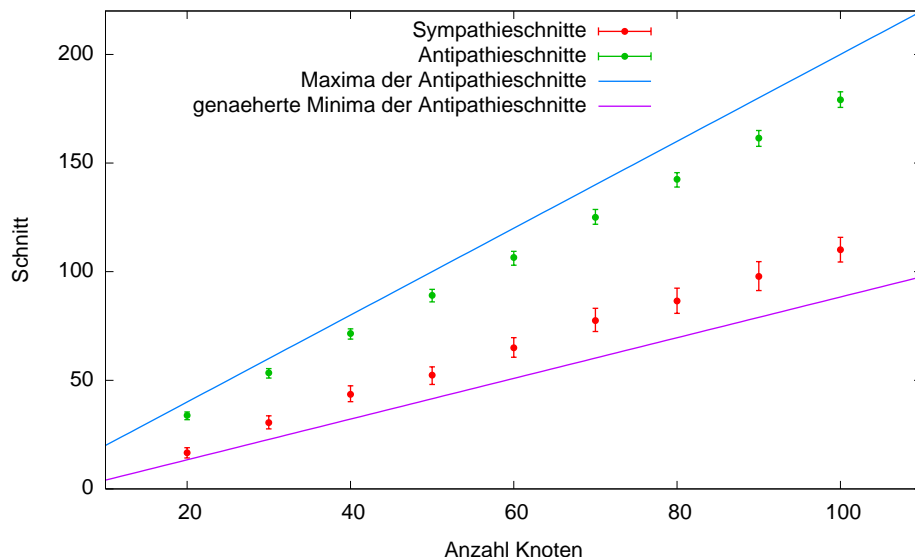
Abbildung 6.7: Experiment 6.7

(a) Start- $k$ -Sektion  $P_a$ , feste Reihenfolge(b) Start- $k$ -Sektion  $P_a$ , zufällige Reihenfolge

(c) Start- $k$ -Sektion  $P_a$ , sektionsoptimierte Reihenfolge(d) Start- $k$ -Sektion  $P_b$ , feste Reihenfolge

(e) Start- $k$ -Sektion  $P_b$ , zufällige Reihenfolge(f) Start- $k$ -Sektion  $P_b$ , sektionsoptimierte Reihenfolge

(g) Start- $k$ -Sektion  $P_c$ , feste Reihenfolge(h) Start- $k$ -Sektion  $P_c$ , zufällige Reihenfolge

(i) Start- $k$ -Sektion  $P_c$ , sektionsoptimierte Reihenfolge

Die Ergebnisse des Experimentes bestätigen diese Vermutung. Der MinMax-Algorithmus, als auch der Min-Algorithmus liefern hier lokale Minima, deren Schnitte sehr nah am absoluten Maximum sind. Nur bei Verwendung der zufälligen Durchlaufreihenfolge hat die an die absoluten Minima genäherte Gerade eine deutlich geringere Steigung als bei Verwendung der beiden anderen Reihenfolgen. Hier variieren jedoch die Größen der Sympathieschnitte des MinMax-Algorithmus stärker als in den anderen Fällen. Bei Eingabe der Startsektion  $P_b$  sind die Sympathieschnitte bei Verwendung der zufälligen Durchlaufreihenfolge ebenfalls leicht kleiner als die bei Verwendung der alternativen Reihenfolgen. Startsektion  $P_c$  scheint hingegen unabhängig von der Durchlaufreihenfolge zu sein. Die zugehörigen Graphiken unterscheiden sich kaum voneinander, und die Ergebnisse gehören zu den Besten der mit der Mehr-Wege-Methode in diesem Experiment ermittelten.

#### 6.4.6 Vergleich von Zwei- und Mehr-Wege-Methode

Zum Vergleich von Zwei-Wege- und Mehr-Wege-Methode werden die Sympathie- und Antipathieschnitte der einzelnen Algorithmen wieder in Abhängigkeit des jeweiligen absoluten Optimum betrachtet. Dazu wird hier bei beiden Methoden eine  $k$ -Sektionierung mit  $k = 2^r$  nach Tabelle 6.1 durchgeführt.

**Experiment 6.8** *Wie groß sind die Abweichungen vom Optimum bei der Zwei- und der Mehr-Wege-Methode?*

*Startbisektion bzw.  $k$ -Sektion:  $P_c$ ;*

*MW-Durchlaufreihenfolge: zufällig;*

*Algorithmen: Min- und MinMax-Algorithmus.*

Es ist offensichtlich, dass alle Werte der Mehr-Wege-Methode stärker vom absoluten Optimum abweichen als die Werte der Zwei-Wege-Methode. Aufgrund der Ergebnisse des Min-Algorithmus wird vermutet, dass die Zwei-Wege-Methode eine höhere Erfolgsrate als die Mehr-Wege-Methode hat. Weitere hier nicht aufgeführte Experimente belegen dies: In 45% aller Fälle liefert die Zwei-Wege-Methode bei Verwendung der Startbisektion  $P_c$  auf Graphen der Ordnungen 20 bis 100 den minimal möglichen Sympathieschnitt. Bei den beiden anderen Startbisektionen beträgt die Erfolgsrate ca. 33%. Währenddessen liegen die Erfolgsraten der Mehr-Wege-Methode bei Verwendung der verschiedenen Start- $k$ -Sektionen alle unter 5%.

Vergleicht man die Graphiken mit denen der Abbildung 6.4, so stellt man fest, dass die Abweichungen in diesem Experiment alle wesentlich kleiner sind als die bei der Bisektion mit dem Kernighan-Lin-Algorithmus. Die Sympathieschnitte des MinMax-Algorithmus unter Verwendung der Zwei-Wege-Methode sind hier nur 1,3 bis 1,6 mal so groß wie das absolute Minimum. Die Ursache hierfür sind die höheren Sektionsgrößen. Bildet man Sektionen mit mehr Elementen, so vergrößert sich auch der minimal mögliche Sympathieschnitt. Proportional zu absoluten Minimum verkleinert sich dann die Abweichung der Sympathieschnitte vom Minimum.

Dass auch hier die Antipathieschnitte des MinMax-Algorithmus proportional näher am absoluten Maximum liegen als die Sympathieschnitte am absoluten Minimum, lässt vermuten, dass mehr Anti- als Sympathien erfüllt werden. Experiment 6.11 in Abschnitt 6.4.8 untersucht diese Spekulation.

### 6.4.7 Gruppenbalanciertheit

Hier wird für den Min- und den MinMax-Algorithmus anhand des in Abschnitt 6.3.4 definierten Maßes untersucht, in wie weit ihre Ergebnisregionen gruppenbalanciert sind. Dazu werden zuerst die Verwendung der Zwei-Wege-Methode betrachtet.

**Experiment 6.9** *Wie unbalanciert sind die Lösungen der Zwei-Wege-Methode?*

*Startbisektionen:  $P_a$ ,  $P_b$  und  $P_c$ ;*

*Algorithmen: Min- und MinMax-Algorithmus.*

Abbildung 6.8: Experiment 6.8

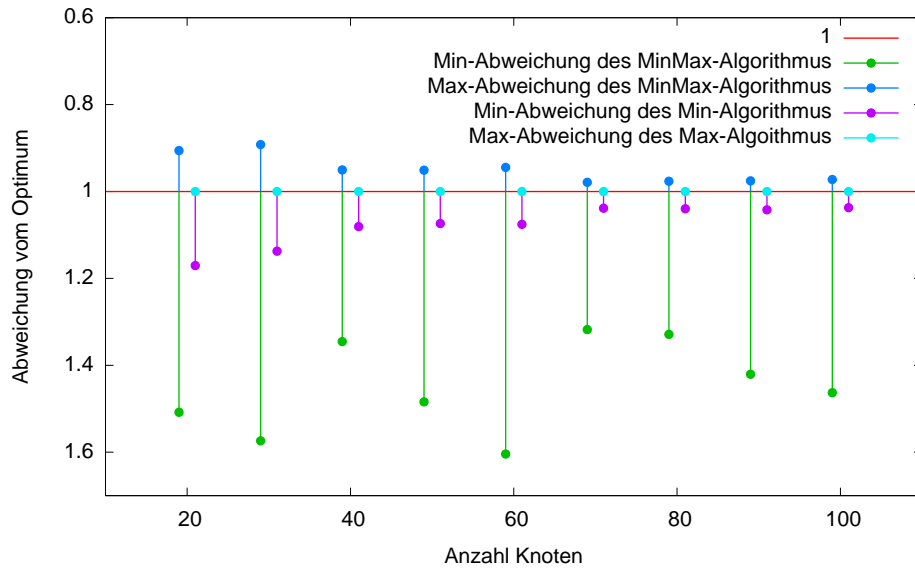
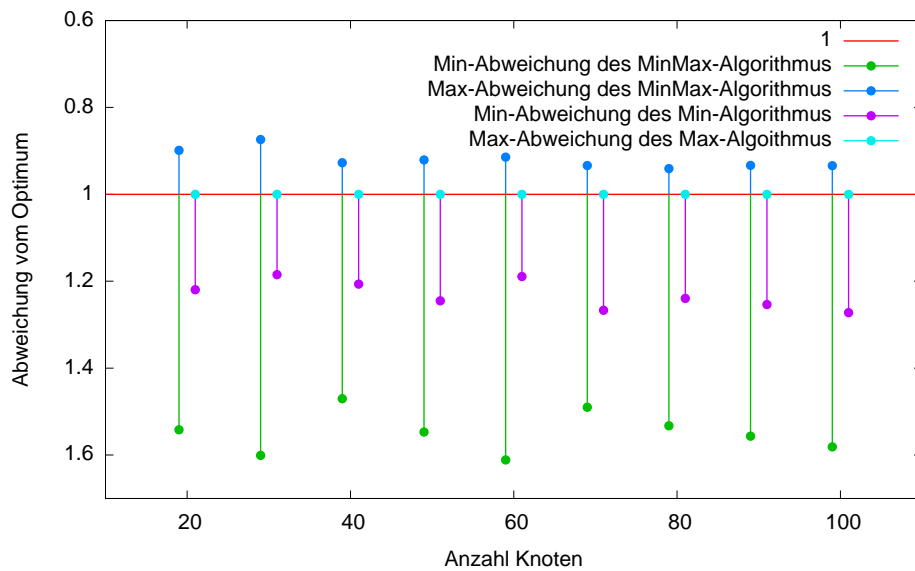
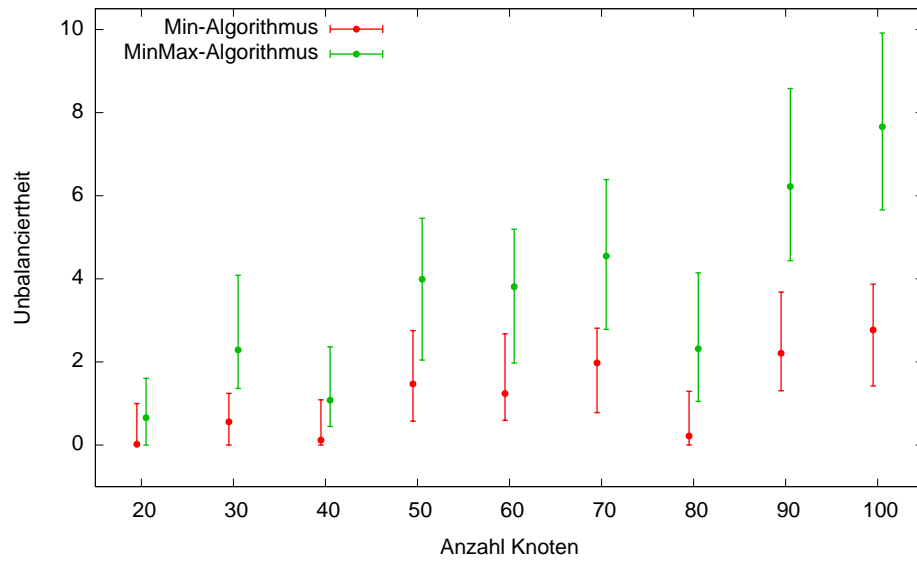
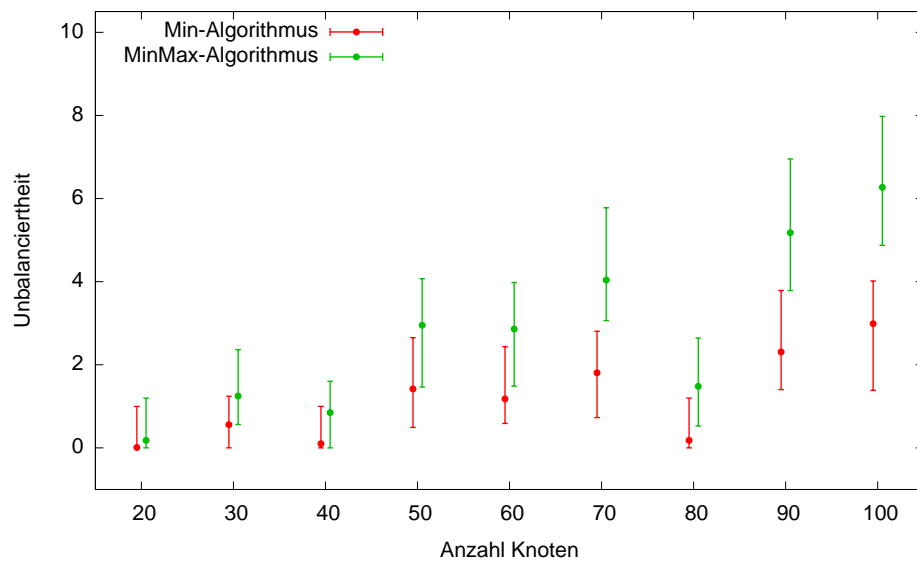
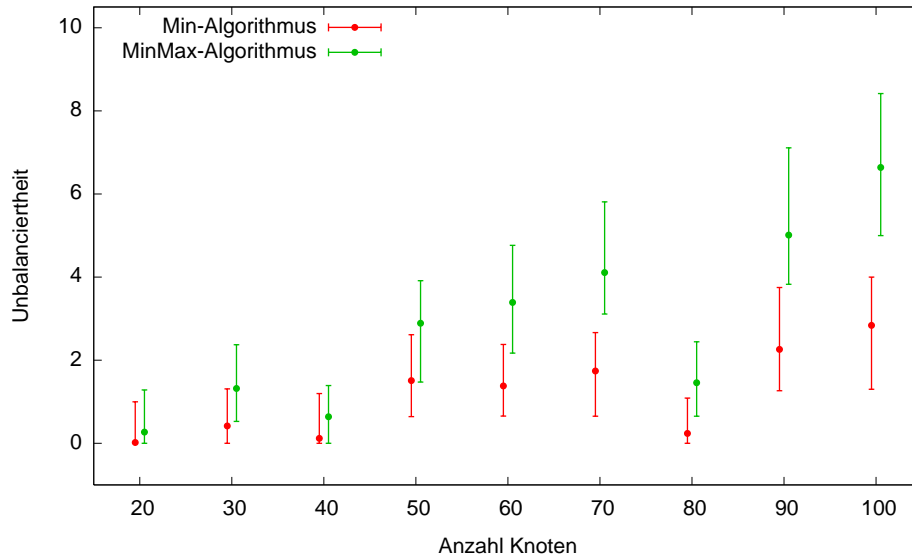
(a) Zwei-Wege-Methode, Startbisektion  $P_c$ (b) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_c$

Abbildung 6.9: Experiment 6.9

(a) Zwei-Wege-Methode, Startbisektion  $P_a$ (b) Zwei-Wege-Methode, Startbisektion  $P_b$



(c) Zwei-Wege-Methode, Startbisektion  $P_c$ 

In den Graphiken von Abbildung 6.9 ist zu sehen, dass die Sektionierung mit der Zwei-Wege-Methode unabhängig von der verwendeten Startbisektion eine  $k$ -Sektion liefert, deren Gruppenbalanciertheit sich bei größeren Graphen verschlechtert. Dabei ist die Verschlechterung beim MinMax-Algorithmus größer als beim Min-Algorithmus. Eine Unbalanciertheit der Größe 10 bei Graphen der Ordnung 100 ist bei einer Einteilung der Knoten in 16 Gruppen allerdings nicht viel. Im Schnitt sind in diesem Fall nur 63% der Gruppen minimal unbalanciert, das heißt, das Verhältnis der Knoten gerader zu den Knoten ungerader Nummerierung weicht nur entsprechend wenig von 1 ab. Was zusätzlich auffällt ist, dass die Ergebnissektionen der Graphen mit 20, 40 oder 80 Knoten, deren Elemente alle Größe 5 haben, die Gruppenbalanciertheit größer ist als bei den übrigen Graphen. Im folgenden Experiment wird beobachtet, ob dies allein mit der Teilbarkeit der Knotenzahl durch die verwendete Gruppenzahl zusammenhängt, oder nur bei der Sektionierung mit der Zwei-Wege-Methode auftritt.

**Experiment 6.10** *Wie unbalanciert sind die Lösungen der Mehr-Wege-Methode?*

*Start- $k$ -Sektionen:  $P_a, P_b$  und  $P_c$ ;*

*MW-Durchlaufreihenfolgen: fest, zufällig und sektionsoptimiert;*

*Algorithmen: Min- und MinMax-Algorithmus.*

Bevor Beobachtungen der Gruppenbalanciertheit zu Graphen bestimmter Ordnungen angestellt werden, stellt man in Anbetracht der Graphiken in Abbildung 6.10 fest, dass sich auch in diesem Experiment zur Mehr-Wege-Methode die Startsektion  $P_a$  als ungeeignet erweist. Die Unbalanciertheit erreicht bei Verwendung der festen und der sektionsoptimierten Durchlaufreihenfolge in vielen Fällen sogar die maximal mögliche Unbalanciertheit. Dies lässt darauf schließen, dass sowohl die erzielten lokalen Optima des MinMax-Algorithmus, als auch die lokalen Minima des Min-Algorithmus alle Antipathien und nur wenige Sympathien erfüllen. Nähere Untersuchungen zur Erfüllung folgen im nächsten Abschnitt. Die Verwendung der zufälligen Reihenfolge in der Mehr-Wege-Methode liefert also in Bezug auf die Gruppenbalanciertheit auch bei unbalancierteren Startsektionen bessere Ergebnisse.

Wie bei der Zwei-Wege-Methode sieht man auch hier, dass die Unbalanciertheit mit der Größe der Graphen bis maximal 10 wächst. Außerdem sind die vom Min-Algorithmus ausgegebenen Sektionen auch bei Verwendung der Mehr-Wege-Methode balancierter als die des MinMax-Algorithmus.

Verwendet man als Eingabe die Startsektion  $P_b$ , so fällt auf, dass die erzielten Sektionen bei Graphen der Ordnung 70 besser balanciert sind als man dies in Anbetracht der Ergebnisse der übrigen Graphen erwartet. Dieses Phänomen kann hier aber nicht plausibel begründet werden.

#### 6.4.8 Erfüllung der Sym- und Antipathien

Nun soll in mehreren Experimenten die in Abschnitt 5.1.3 beschriebene Erfüllung der Sym- und Antipathien untersucht werden. Neben der Beantwortung der Frage, wie viele der ausgesprochenen Sym- bzw. Antipathien pro Person durchschnittlich erfüllt werden, interessiert man sich in Experimenten 6.12 und 6.12 dafür, wie viel Prozent der Personen bei der Gruppeneinteilung keine Sympathie und/oder Antipathie erfüllt bekommen.

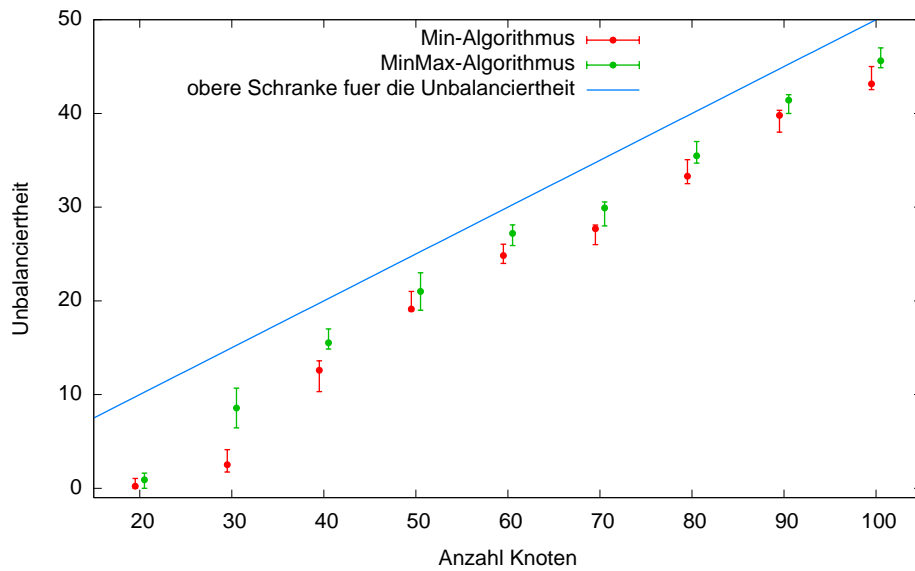
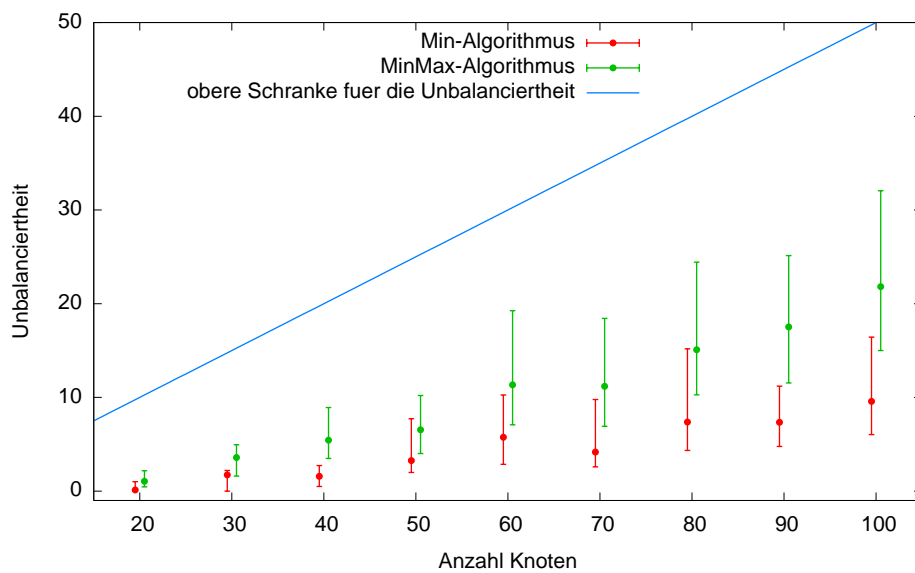
**Experiment 6.11** *Wie viel Prozent der ausgesprochenen Sym- bzw. Antipathien werden bei der Zwei-Wege-Methode pro Person erfüllt?*

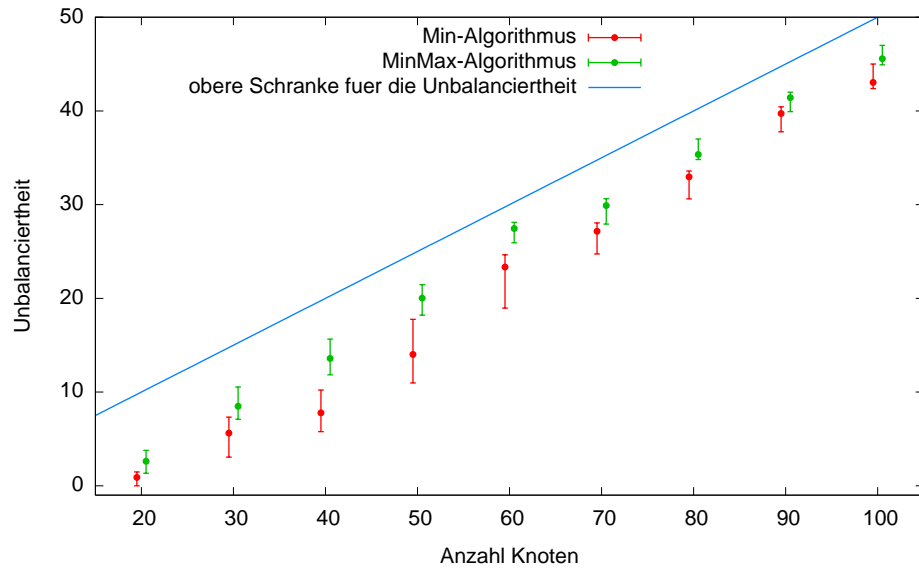
*Startbisektion:  $P_a, P_b$  und  $P_c$ ;*

*Algorithmen: MinMax-Algorithmus.*

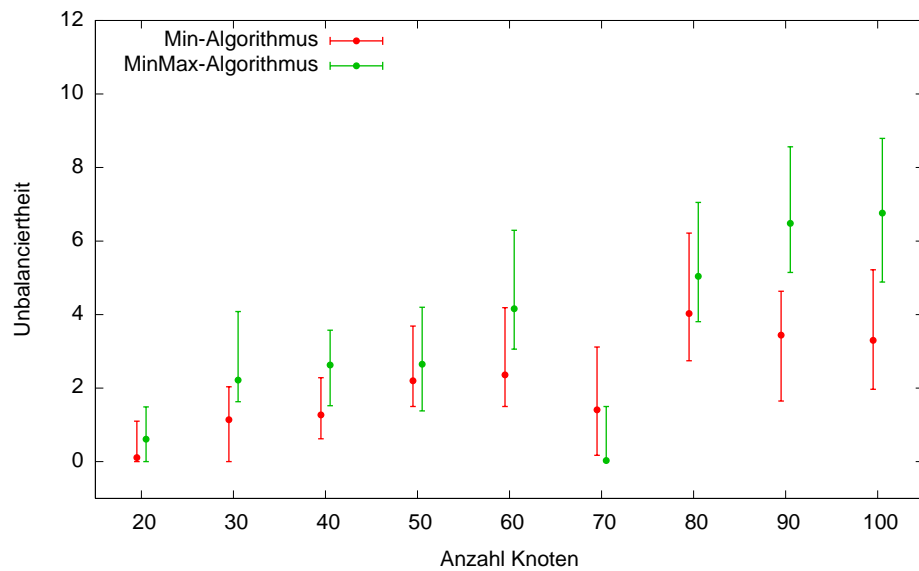
Bei der Zwei-Wege-Methode werden pro Person durchschnittlich ungefähr 50% der ausgesprochenen Sympathien und knapp 100% der Antipathien erfüllt. Somit hat bei dieser Gruppeneinteilung jede Person im Schnitt eine der beiden als „sympathisch“ bewerteten Person in ihrer Gruppe und keine der Personen, die sie unsympathisch findet. Es scheinen also stets pro Person mehr Antipathien als Sympathien erfüllt zu werden. Dies hängt damit zusammen, dass durch eine Einteilung in mehr als zwei Sektionselemente jeder

Abbildung 6.10: Experiment 6.10

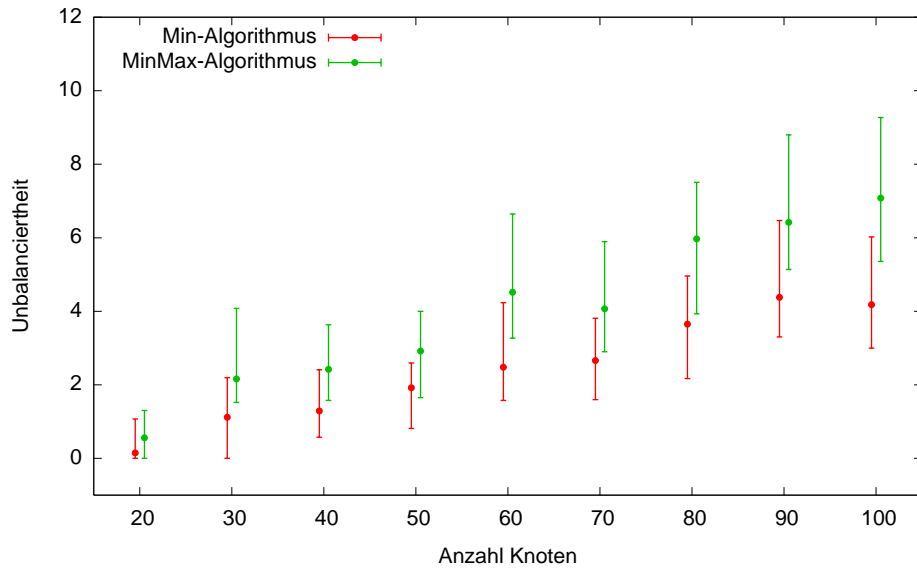
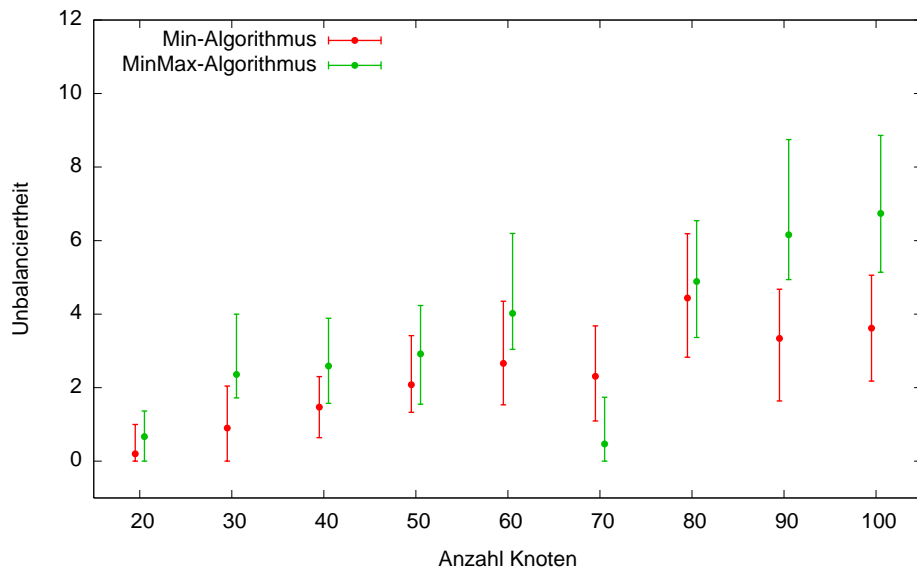
(a) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_a$ , feste Reihenfolge(b) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_a$ , zufällige Reihenfolge

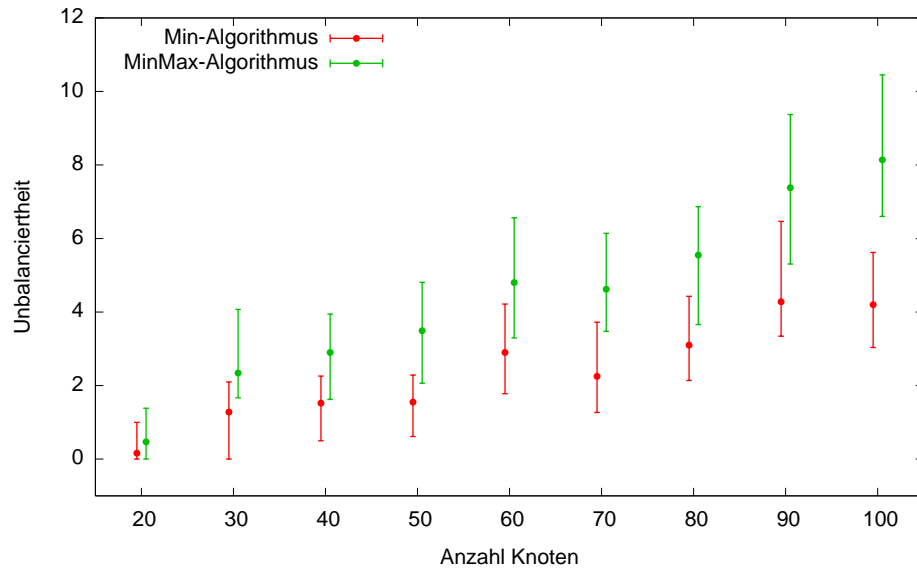
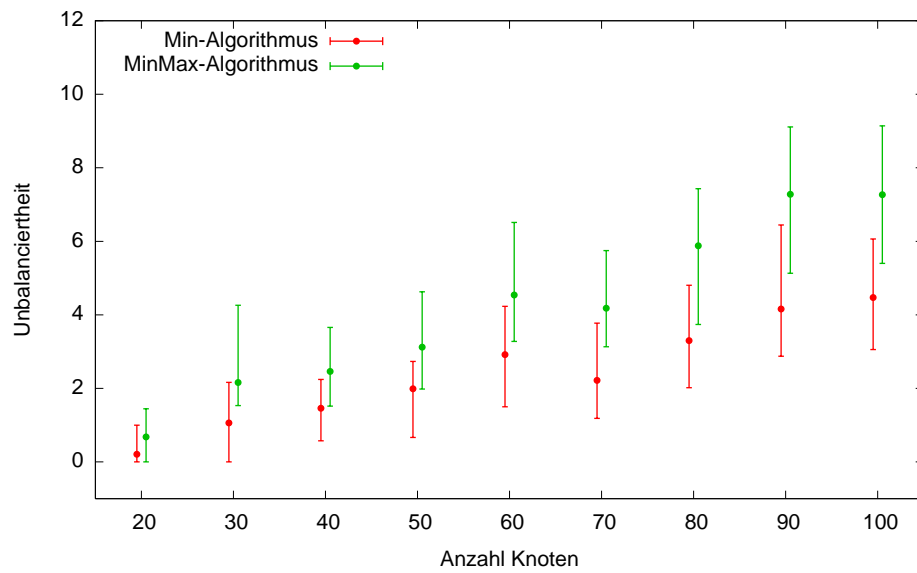


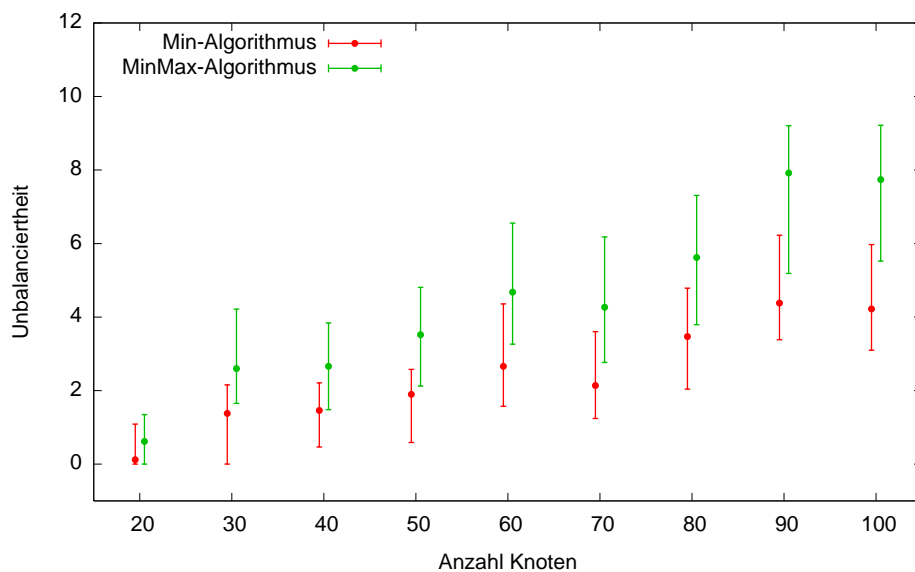
(c) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_a$ , sektionsoptimierte Reihenfolge



(d) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_b$ , feste Reihenfolge

(e) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_b$ , zufällige Reihenfolge(f) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_b$ , sektionsoptimierte Reihenfolge

(g) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_c$ , feste Reihenfolge(h) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_c$ , zufällige Reihenfolge



(i) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_c$ , sektionsoptimierte Reihenfolge

Knoten von mehr Knoten getrennt wird, als in jeder Sektion Knoten enthalten sind. Da die in diesem Experiment gebildeten Sektionselemente aller Graphen ungefähr Größe 6 haben, ist die Erfüllung der Sympathien beinahe unabhängig von der Ordnung der Graphen. Dadurch, dass mit der Zwei-Wege-Methode nur 4-, 8- oder 16-Sektionen ermittelt werden können (siehe Tabelle 6.1 in Abschnitt 6.2.4), kann man beobachten, dass, wie zu erwarten ist, bei der Einteilung in größere Gruppen mehr Sympathien erfüllt werden als bei kleineren. Auch dieses Experiment ergibt in Anbetracht der Graphiken in Abbildung 6.11, dass die verwendete Startsektion keinen Einfluß auf die Ergebnisse der Mehr-Wege-Methode hat.

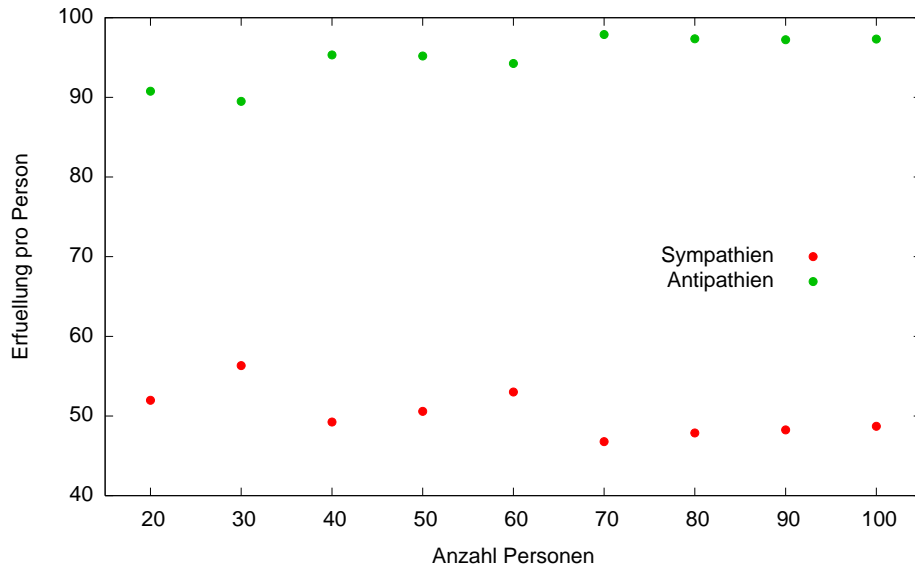
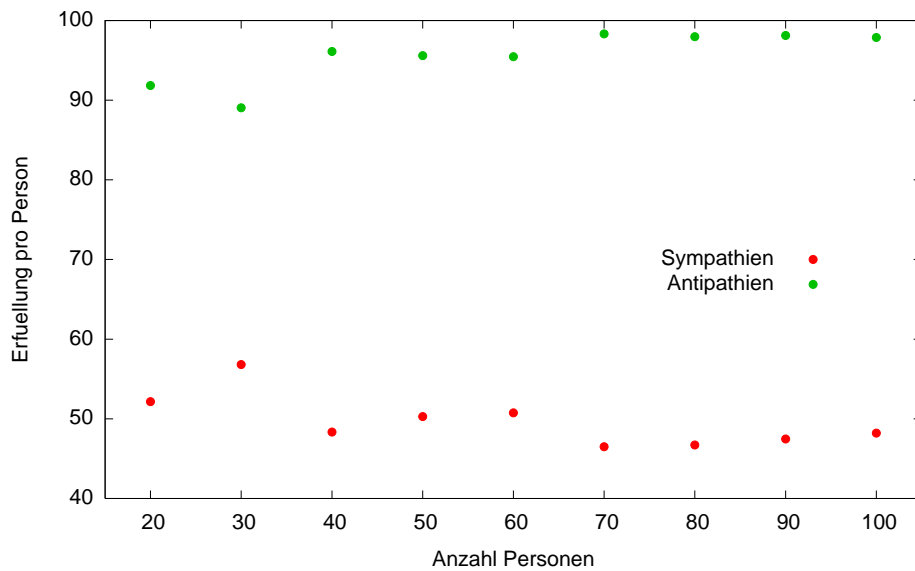
Die Tatsache, dass im Schnitt 50% der zwei ausgesprochenen Sympathien erfüllt werden, garantiert jedoch nicht die Erfüllung einer Sympathie pro Person. Das folgende Experiment gibt Aufschluss über die Anzahl Personen, denen bei der Gruppeneinteilung keine einzige Sympathie erfüllt wird.

**Experiment 6.12** *Wie viel Prozent der Personen bekommen bei der Zwei-Wege-Methode keine der ausgesprochenen Sympathien und/oder Antipathien erfüllt?*

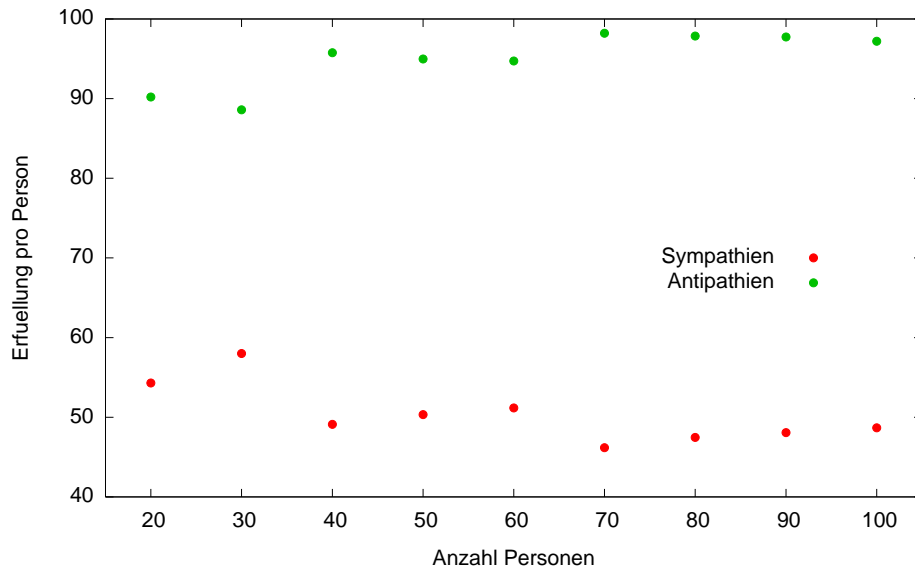
*Startbisektion:  $P_a, P_b$  und  $P_c$ ;*

*Algorithmen: MinMax-Algorithmus.*

Abbildung 6.11: Experiment 6.11

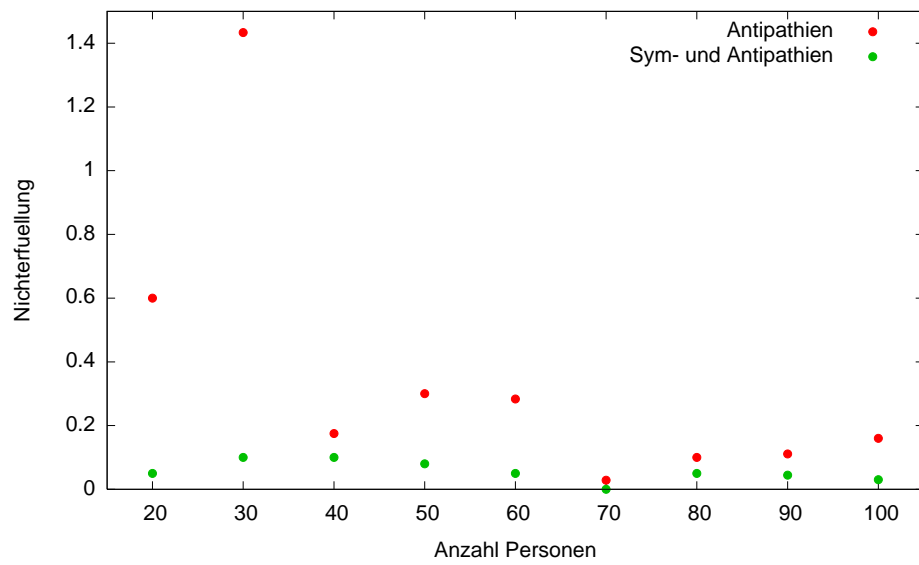
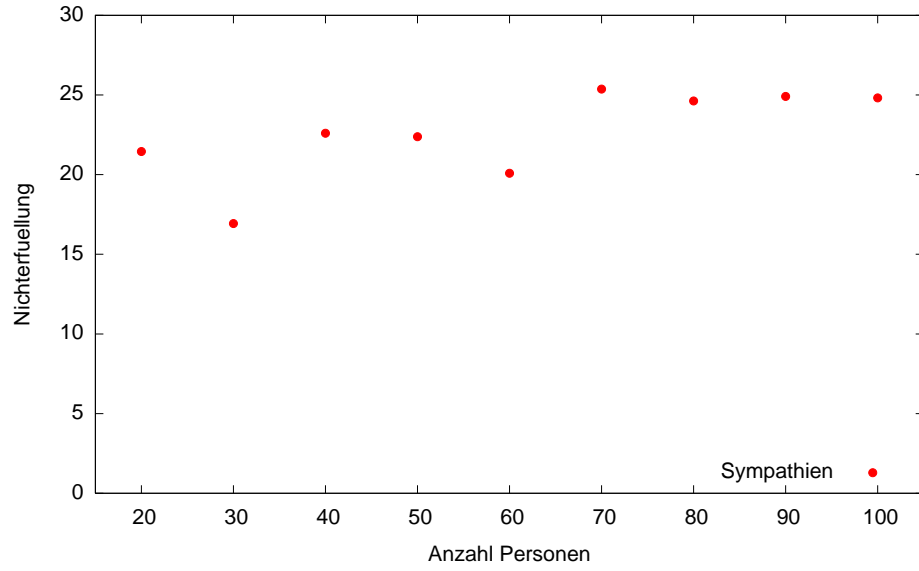
(a) Zwei-Wege-Methode, Startbisektion  $P_a$ (b) Zwei-Wege-Methode, Startbisektion  $P_b$

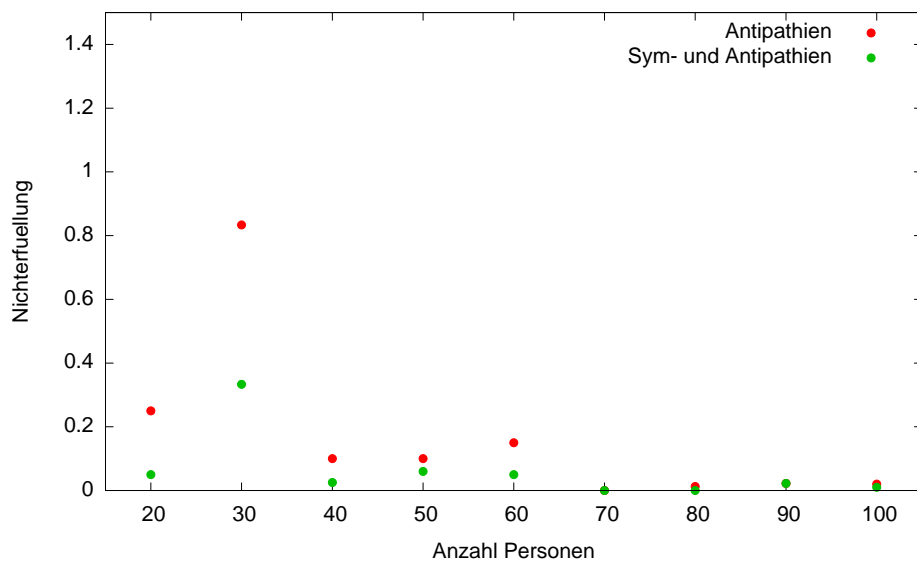
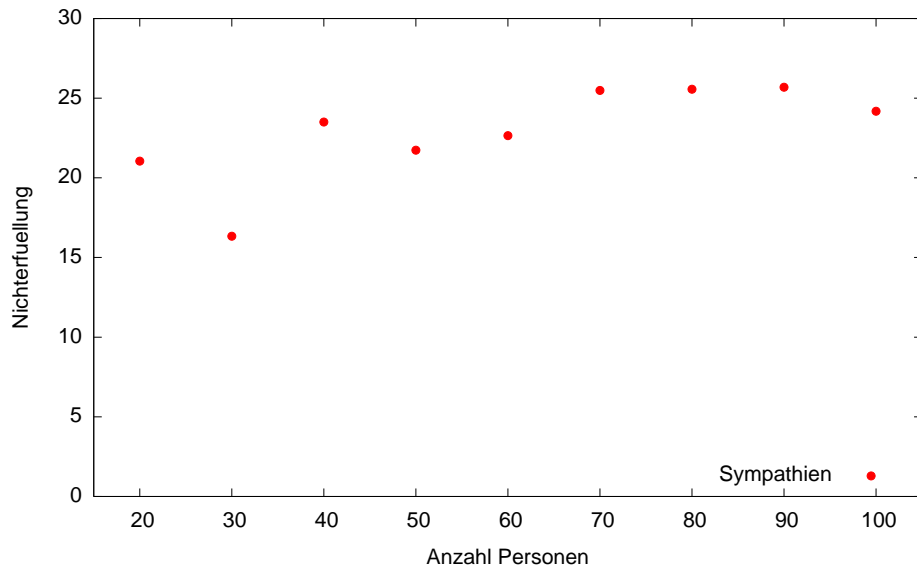


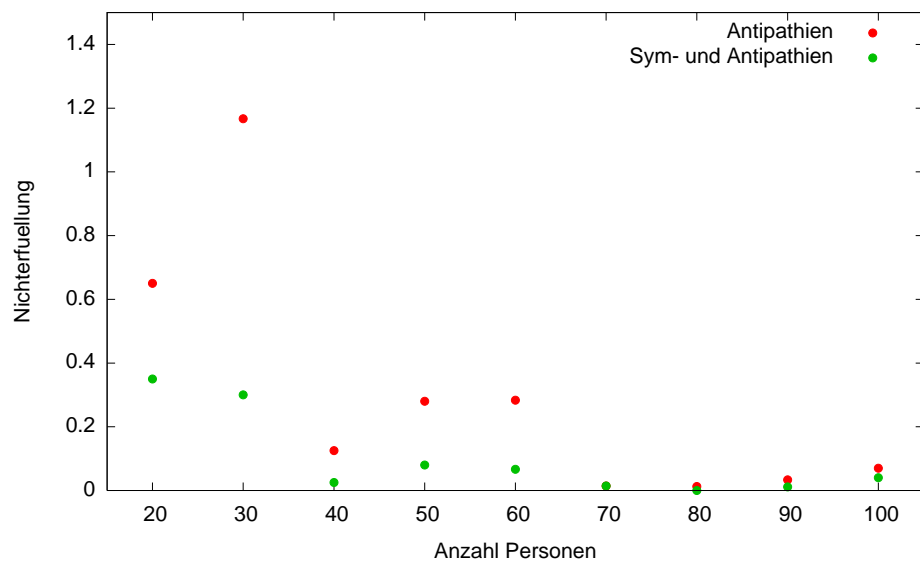
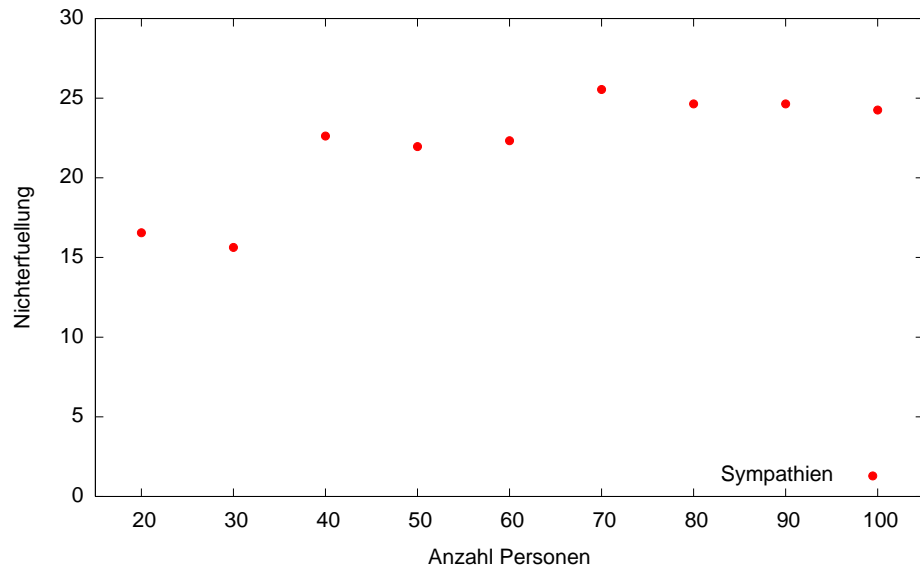
(c) Zwei-Wege-Methode, Startbisektion  $P_c$ 

Auch in den Graphiken der Abbildung 6.12 sind keine wesentlichen Unterschiede bei Verwendung der drei verschiedenen Startbisektionen erkennbar. Nachdem das vorherige Experiment schon ergeben hat, dass pro Person mehr Antipathien als Sympathien erfüllt werden, zeigen nun die jeweils oberen Graphiken, dass im Durchschnitt über 20% der teilnehmenden Personen mit keiner der von ihnen sympathisch gefundenen Personen in einer Gruppe ist. Betrachtet man weiter die unteren Graphiken, so erkennt man, dass nur bei der 4-Sektionierung der kleineren Graphen bis zu 1,4% Nichterfüllungen von Antipathien auftreten. Bei allen übrigen Sektionierungen wird höchstens eine Nichterfüllung von ca. 0,3% erreicht. Es haben demnach nur sehr selten Personen einen Gesprächspartner in ihrer Gruppe, dem sie eine Antipathiebewertung ausgesprochen haben. Um den Erfolg der Gruppeneinteilung besser beurteilen zu können, sind nun die Personen interessant, denen keine einzige ausgesprochene Bewertung erfüllt wurde, sei es eine Sympathie- oder eine Antipathiebewertung. Abgesehen von einzelnen „Ausreißern“ sieht man in den jeweils unteren Graphiken der Abbildung 6.12, dass für weniger als 0,1% der Teilnehmer die durchgeführte Gruppeneinteilung erfolglos zu sein scheint. Das bedeutet, dass bei Anwendung der beschriebenen Strategie unter Einsatz der Zwei-Wege-Methode nur mit geringer Wahrscheinlichkeit Personen bei der Erfüllung von Sympathie- und Antipathiebewertungen leer ausgehen.

Abbildung 6.12: Experiment 6.12

(a) Zwei-Wege-Methode, Startbisektion  $P_a$

(b) Zwei-Wege-Methode, Startbisektion  $P_b$

(c) Zwei-Wege-Methode, Startbisektion  $P_c$

Betrachte man nun dieselben Experimente für die Mehr-Wege-Methode:

**Experiment 6.13** *Wie viel Prozent der ausgesprochenen Sym- bzw. Antipathien werden bei der Mehr-Wege-Methode pro Person erfüllt?*

*Start-k-Sektion:  $P_a, P_b$  und  $P_c$ ;*

*MW-Durchlaufreihenfolge: fest, zufällig, sektionsoptimiert;*

*Algorithmen: MinMax-Algorithmus.*

Die Ergebnisse dieses Experimentes bestätigen die in Abschnitt 6.4.7 aufgestellte Vermutung, dass bei Verwendung der Startsektion  $P_a$  und der festen oder sektionsoptimierten Reihenfolge auf größeren Graphen pro Person nur wenige (weniger als 5%) der ausgesprochenen Sympathien erfüllt werden. Dafür beträgt die Erfüllung der Antipathien pro Person bei Graphen der Ordnung  $\geq 40$  fast 100%. Auch hier erhält man bessere Ergebnisse bei Verwendung der zufälligen Reihenfolge, wobei allerdings bei Einsatz der Startsektion  $P_c$  prozentual mehr Sympathien erfüllt werden, als wenn die beiden anderen Startsektionen eingegeben werden. Die in diesem Experiment maximal mögliche Erfüllung von Sympathien pro Person beträgt bei den kleinsten Graphen 60% und sinkt mit größer werdender Ordnung der Graphen auf ca. 43% bei Ordnung 100. Im Zuge dessen vergrößert sich die Erfüllung der Antipathien von ca. 90% auf 100%. Bei Verwendung von  $P_c$  als Startsektion fällt wieder der „Ausreißer“ bei Graphen der Ordnung 70 auf. Wie schon aus der Gruppenbalanciertheit dieser Graphen geschlossen werden kann, werden hier beinahe keine Sympathien und alle Antipathien erfüllt. Verwendet man in diesem Fall allerdings die zufällige Durchlaufreihenfolge, so verschwindet dieser „Ausreißer“.

Da hier also für jede Startsektion die Verwendung der zufälligen Reihenfolge die besten Ergebnisse für die Erfüllung der Sym- und Antipathien pro Person liefert, wird für das folgende Experiment nur Ergebnisgraphiken zur zufälligen Durchlaufreihenfolge dargestellt.

**Experiment 6.14** *Wie viel Prozent der Personen bekommen bei der Mehr-Wege-Methode keine der ausgesprochenen Sympathie und/oder Antipathien erfüllt?*

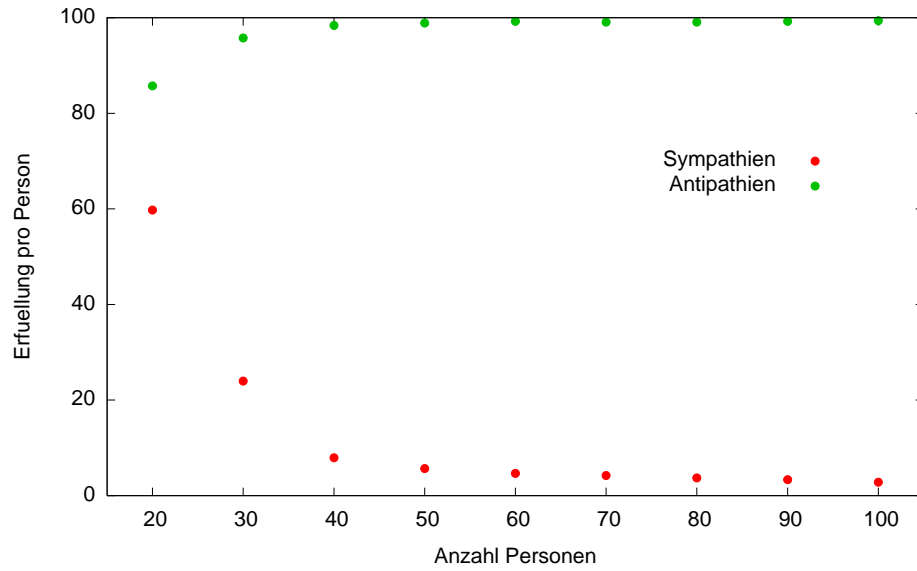
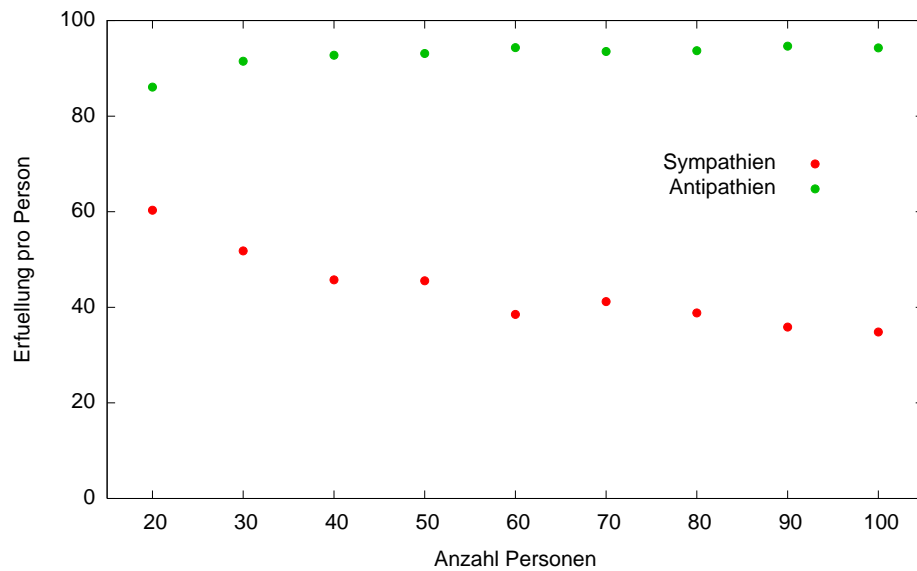
*Start-k-Sektion:  $P_a, P_b$  und  $P_c$ ;*

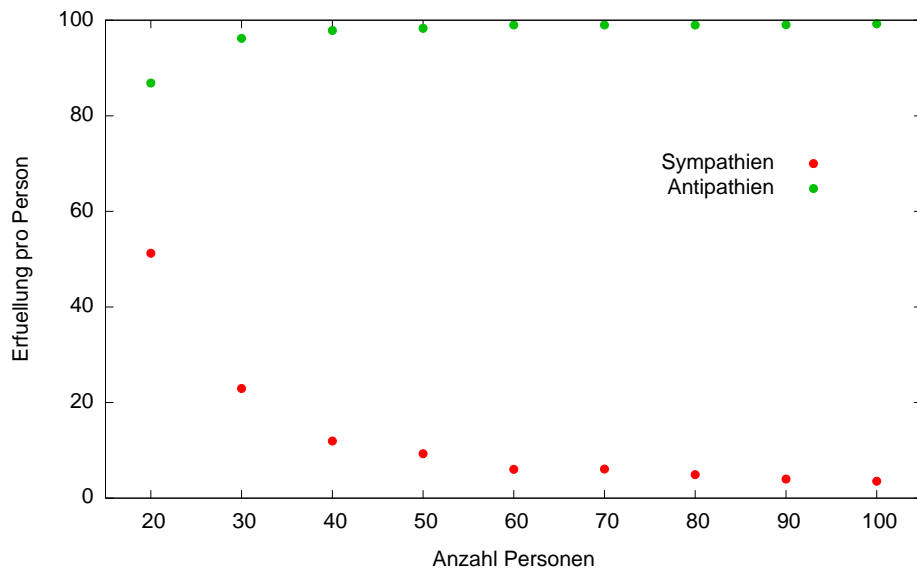
*MW-Durchlaufreihenfolge: zufällig;*

*Algorithmen: MinMax-Algorithmus.*

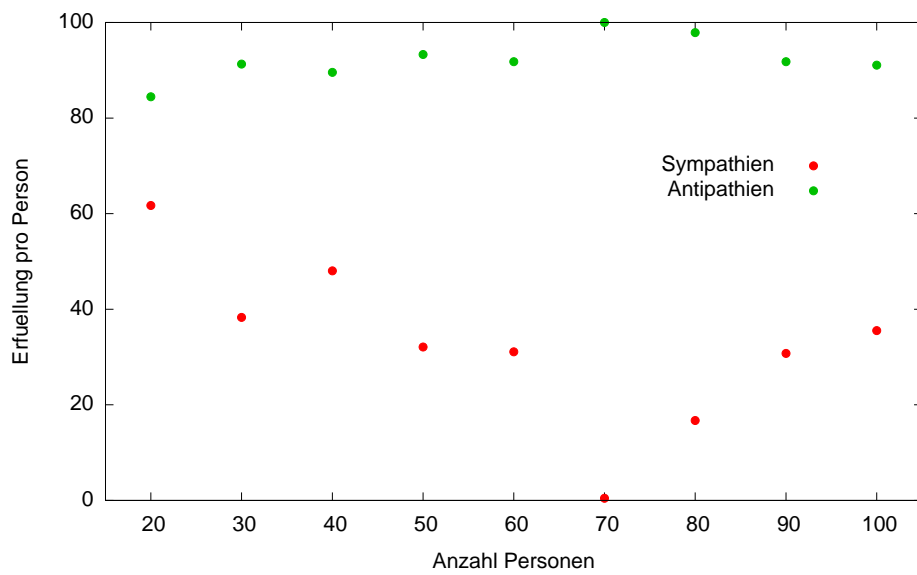
Die jeweils oberen Graphiken in Abbildung 6.14 zeigen, dass die Eingabe der Startsektion  $P_a$  auch hier schlechtere Ergebnisse liefert: Die Nichterfüllung der Sympathien liegt auf den größeren Graphen bei 40 – 50%

Abbildung 6.13: Experiment 6.13

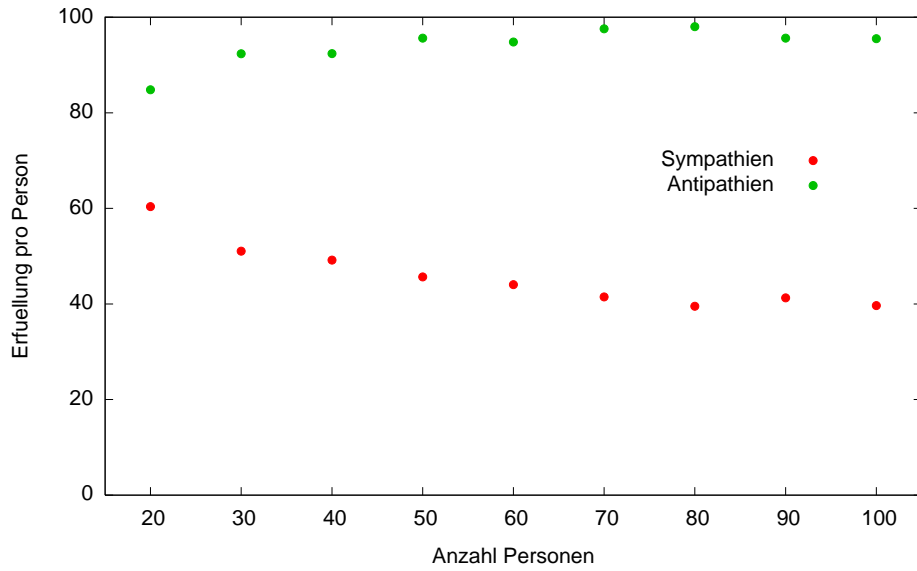
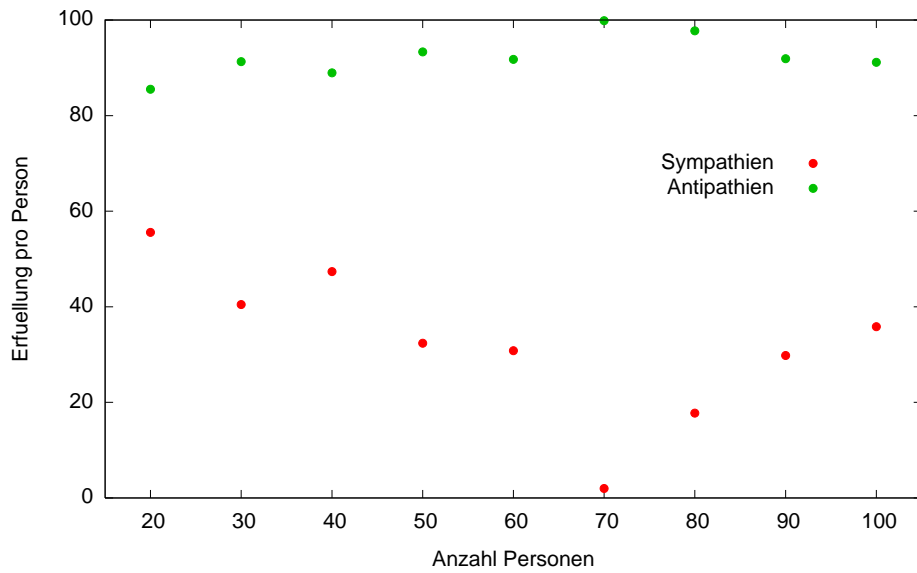
(a) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_a$ , feste Reihenfolge(b) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_a$ , zufällige Reihenfolge



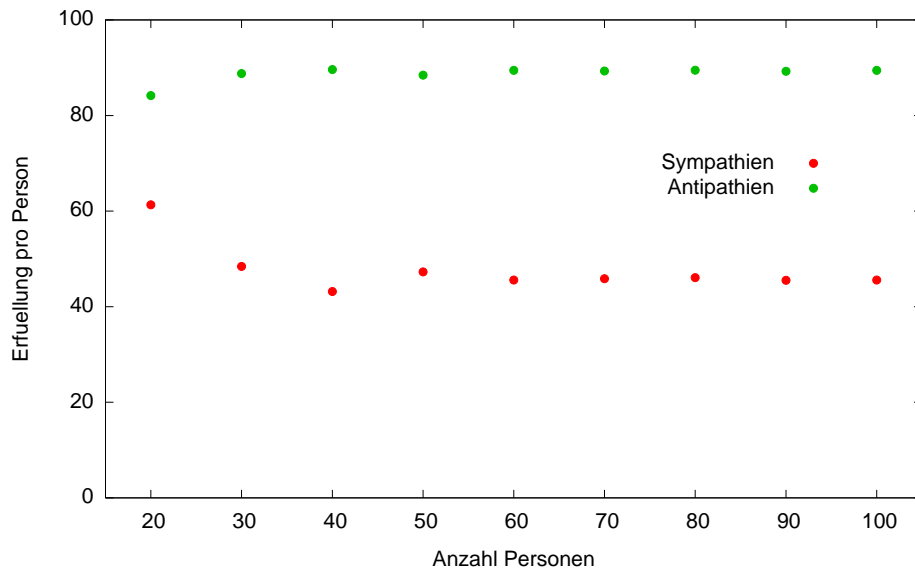
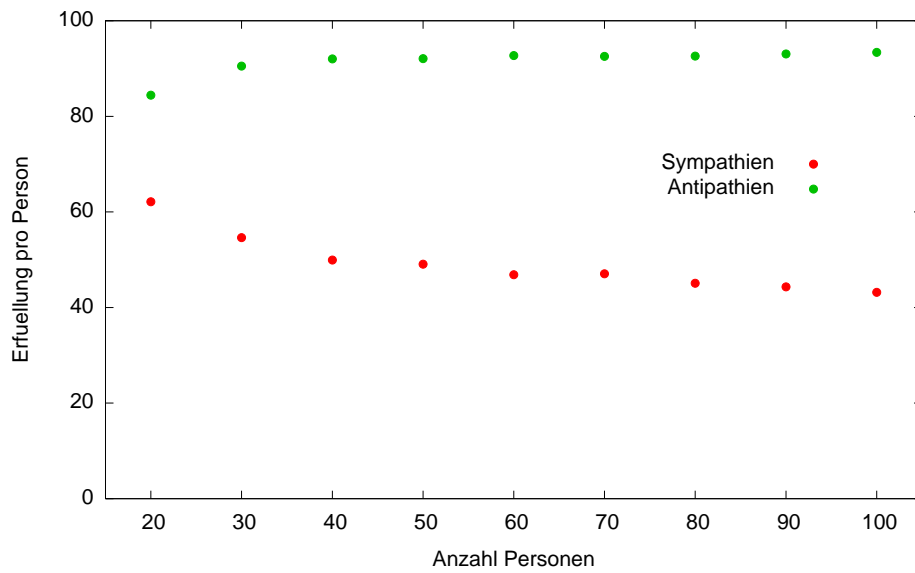
(c) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_a$ , sektionsoptimierte Reihenfolge

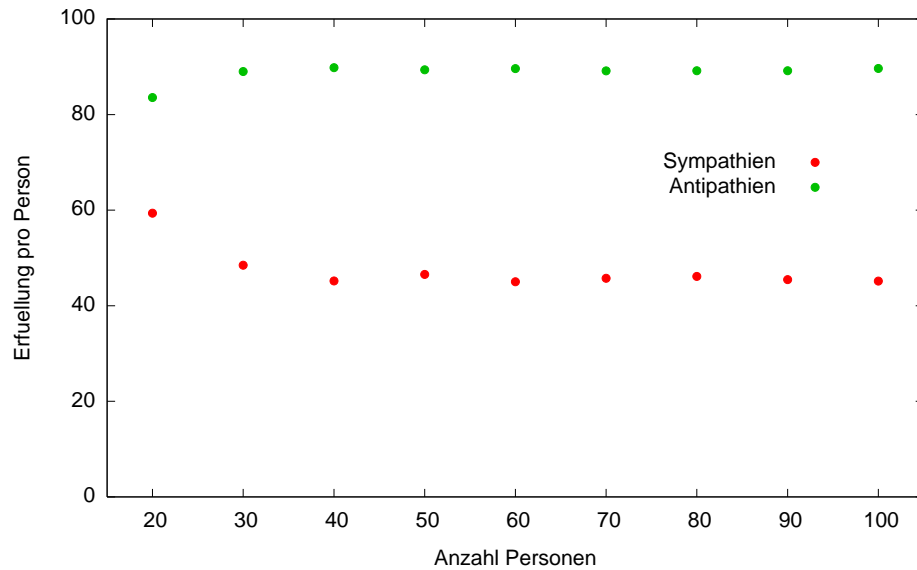


(d) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_b$ , feste Reihenfolge

(e) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_b$ , zufällige Reihenfolge(f) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_b$ , sektionsoptimierte Reihenfolge



(g) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_c$ , feste Reihenfolge(h) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_c$ , zufällige Reihenfolge



(i) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_c$ , sektionsoptimierte Reihenfolge

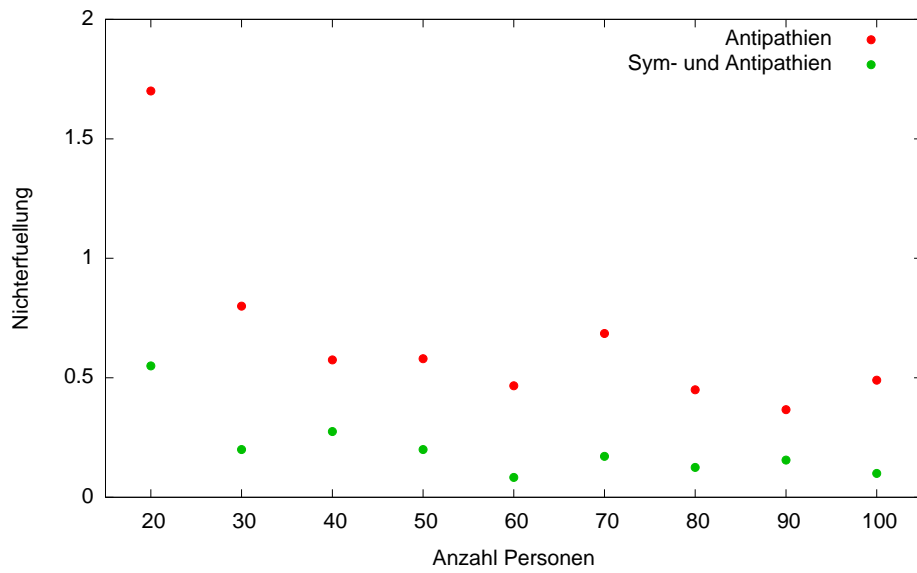
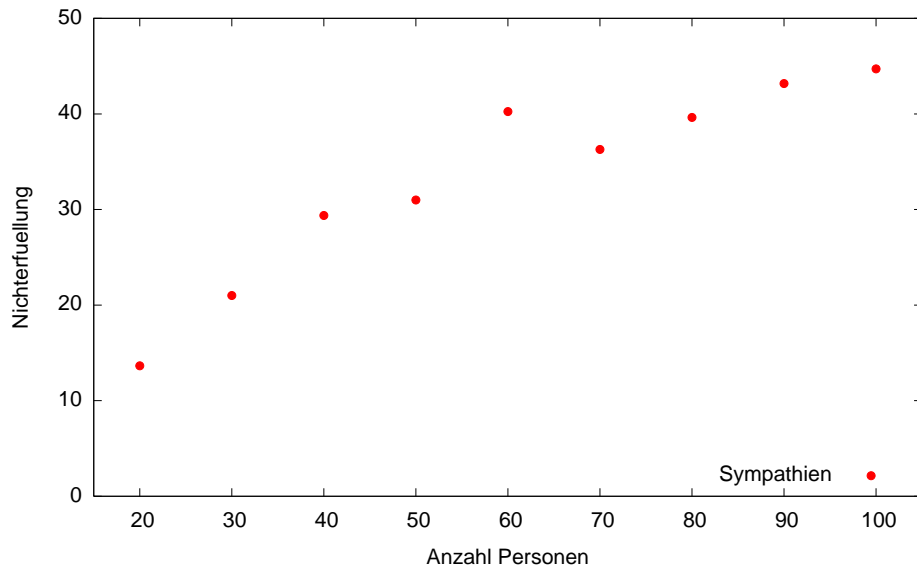
der Personen. Lässt man die unterschiedlich großen Sektionierungen außer Acht, dann sieht man im Vergleich zu entsprechenden Graphik in Abbildung 6.12 (c), dass die Verwendung von Startsektion  $P_c$  für die Sympathien annähernd so gute Ergebnisse wie die Zwei-Wege-Methode liefert.

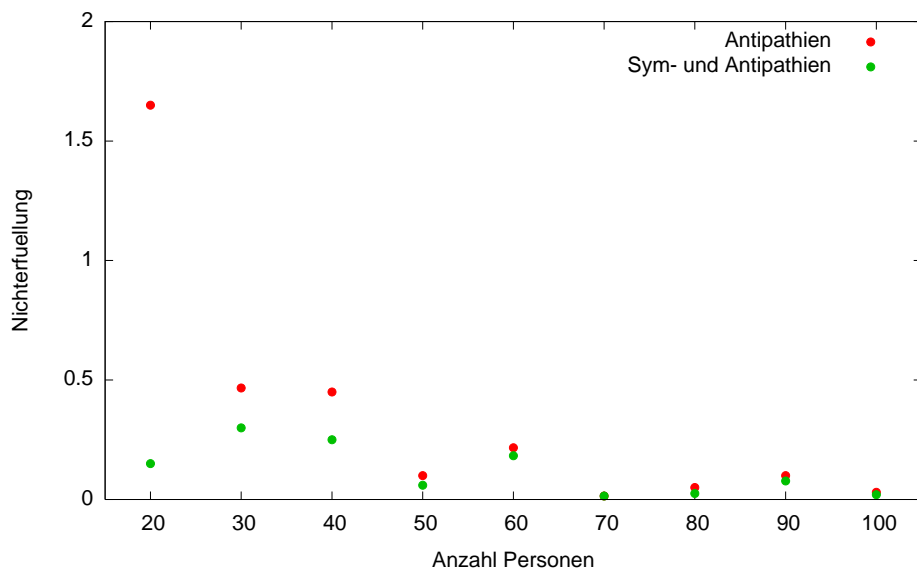
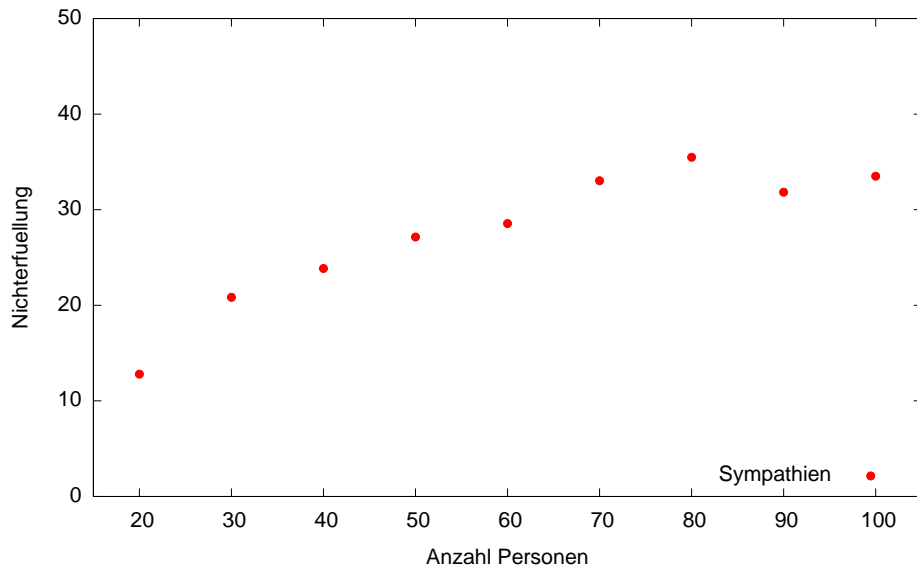
Betrachtet man die unteren Graphiken, so erkennt man, dass die Nichterfüllung der Sym- und Antipathien hier im Vergleich zur Zwei-Wege-Methode (siehe Abbildung 6.12, obere Graphiken) öfter eintritt. Da aber die durchschnittlich ermittelten 0,3% von der jeweiligen Anzahl Personen nicht viel sind, ist auch hier die Wahrscheinlichkeit sehr klein, dass einer Person keine einzige der ausgesprochenen Sym- und Antipathien erfüllt bekommt. Die Verwendung von Startsektion  $P_c$  liefert auf größeren Graphen sogar noch bessere Ergebnisse, wie in Anbetracht der entsprechenden Graphik in Abbildung 6.13 zu erwarten ist.

### 6.4.9 Drei Sympathien und eine Antipathie

Im Gegensatz zu den vorherigen Experimenten werden in diesem Abschnitt Daten von Sympathie- und Antipathiegraphen verwendet, bei denen pro Person drei Sympathien und eine Antipathie ausgesprochen werden. Verglichen

Abbildung 6.14: Experiment 6.14

(a) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_a$

(b) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_b$

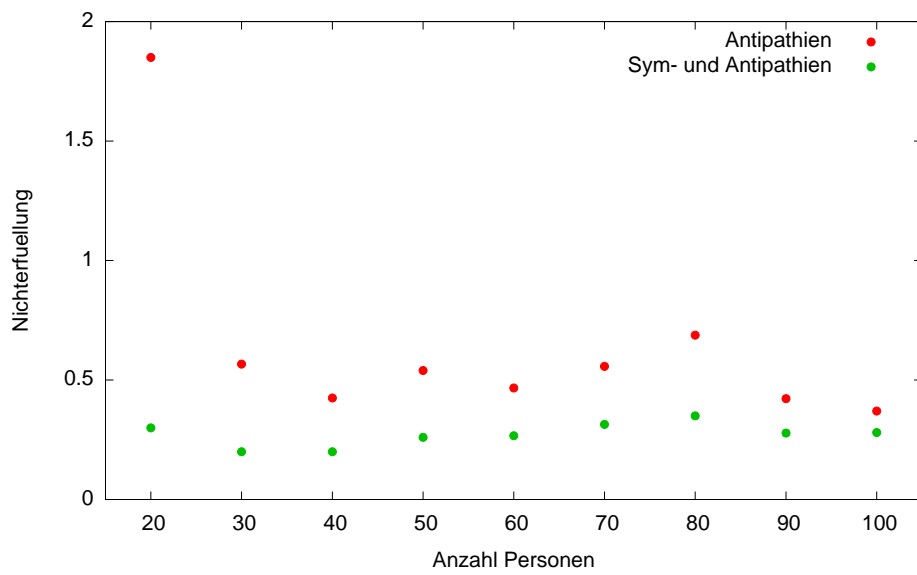
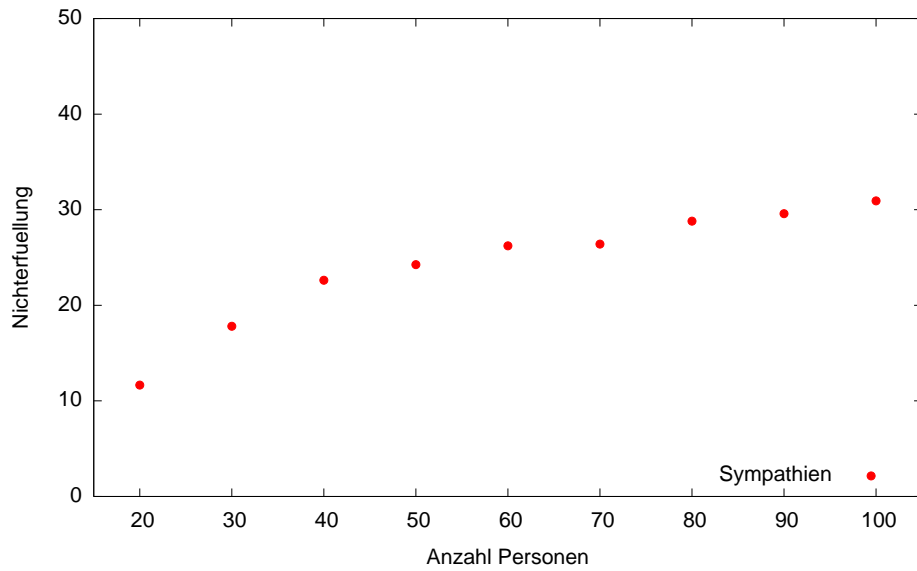
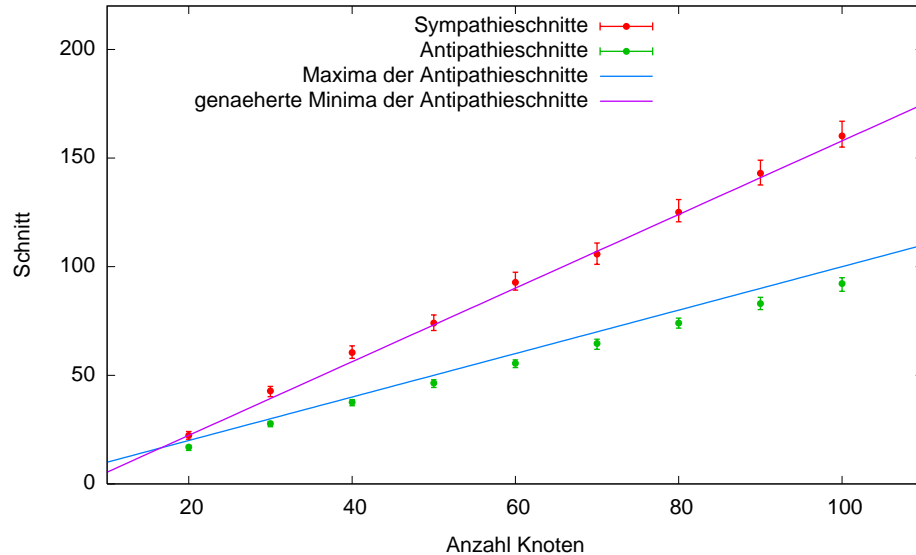
(c) Mehr-Wege-Methode, Start- $k$ -Sektion  $P_c$

Abbildung 6.15: Experiment 6.15



werden die Ergebnisse dieser Experimente mit den entsprechenden Ergebnissen unter Verwendung von Daten mit je zwei ausgesprochenen Sym- und Antipathien. Es soll beobachtet werden, in wie weit sich die lokal optimalen  $k$ -Sektionen verändern, wenn man mehr Sympathien als Antipathien aussprechen lässt.

Die folgenden Experimente werden ausschließlich für die Mehr-Wege-Methode beschrieben. Betrachtet man schließlich die Größen der möglichen Sektionierungen, so ist diese vielseitiger zur Gruppeneinteilung einsetzbar. Es sei aber der Vollständigkeit halber erwähnt, dass die Ausführungen mit der Zwei-Wege-Methode entsprechende Ergebnisse liefern.

**Experiment 6.15** *Welche Lösungen liefert die Mehr-Wege-Methode?*

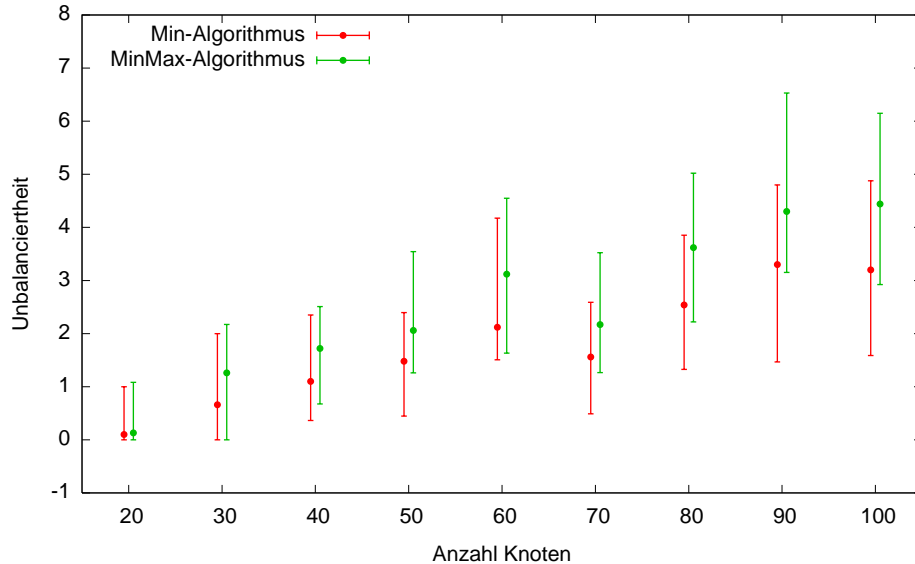
*Start- $k$ -Sektionen:  $P_c$ ;*

*MW-Durchlaufreihenfolgen: zufällig;*

*Algorithmen: Min- und MinMax-Algorithmus.*

Dadurch, dass hier pro Person drei Sympathien und eine Antipathie ausgesprochen werden, liegt in Abbildung 6.15 die an die absoluten Minima genäherte Gerade oberhalb der Gerade der absoluten Maxima. Man erkennt, dass die vom Min-Max-Algorithmus erzielten Sympathieschnitte sehr nahe an denen des Min-Algorithmus liegen. Zudem sind die erzielten Antipathieschnitte nur geringfügig kleiner als die absoluten Maxima. Vergleiche man

Abbildung 6.16: Experiment 6.16



diese Graphik nun mit der in Abbildung 6.7 (h), wo im zugehörigen Experiment pro Person zwei Sym- und zwei Antipathien ausgesprochen wurden: Die Differenz zwischen den Schnitten des vom MinMax-Algorithmus erzielten lokalen Optimum und den (genäherten) absoluten Optima ist hier im Allgemeinen kleiner.

Betrachte man im Folgenden die Gruppenbalanciertheit und die Erfüllung der Sym- und Antipathien auf den Ergebnissen der hier betrachteten Daten:

**Experiment 6.16** *Wie unbalanciert sind die Lösungen der Mehr-Wege-Methode?*

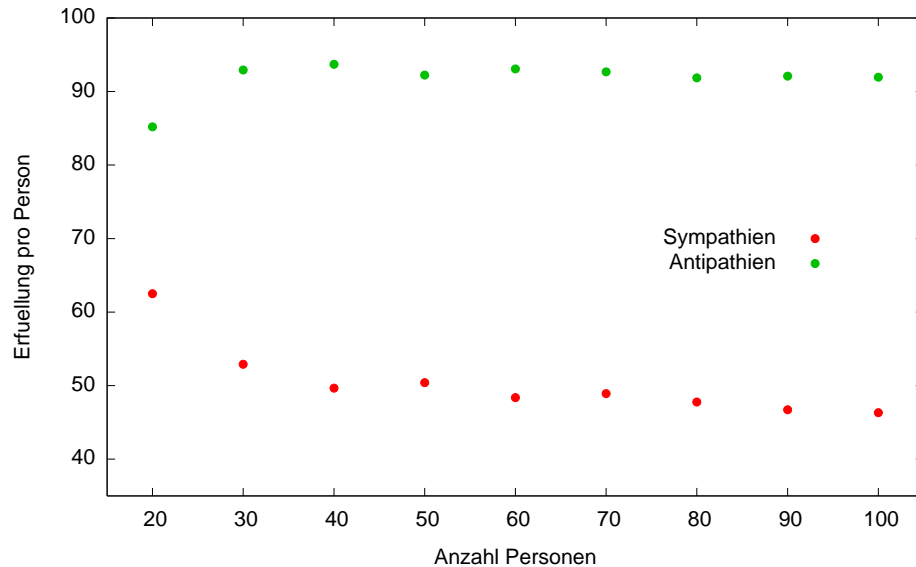
*Start-k-Sektionen:  $P_c$ ;*

*MW-Durchlaufreihenfolgen: zufällig;*

*Algorithmen: Min- und MinMax-Algorithmus.*

Die Ergebnisse der Gruppenbalanciertheit ähneln auf den ersten Blick denen in Abbildung 6.10 (h). Auch hier liefert der Min-Algorithmus balanciertere Sektionen als der MinMax-Algorithmus, und die Unbalanciertheit der Sektionselemente wächst mit der Ordnung der Graphen. Die Größen der Unbalanciertheiten sind in diesem Experiment allerdings ca. um 1 kleiner. Das bedeutet, dass die Aussprache von drei Sympathien und einer Antipathie im Gegensatz zu je zwei eine Verbesserung der Gruppeneinteilung in Bezug auf die Gruppenbalanciertheit bringt.

Abbildung 6.17: Experiment 6.17



**Experiment 6.17** *Wie viel Prozent der ausgesprochenen Sym- bzw. Antipathien bei der Mehr-Wege-Methode werden pro Person erfüllt?*

*Start-k-Sektion:  $P_c$ ;*

*MW-Durchlaufreihenfolge: zufällig;*

*Algorithmen: MinMax-Algorithmus.*

Wird die zugehörige Graphik in Abbildung 6.17 wieder mit der entsprechenden Graphik in Abbildung 6.13 (h) verglichen, so kann auch hierbei auf den ersten Blick kein Unterschied festgestellt werden. Es werden pro Person jeweils der gleiche Prozentteil der ausgesprochenen Sym- bzw. Antipathien erfüllt. Allerdings stellen 50% erfüllte Sympathien bei Aussprache von Dreien 1,5 Sympathien dar. Das heißt, es werden hier pro Person in Etwa ein bis zwei Sympathien und eine Antipathie erfüllt. Zur genaueren Untersuchung betrachte man zusätzlich wieder die Ergebnisse des folgenden Experimentes:

**Experiment 6.18** *Wie viel Prozent der Personen bekommen bei der Mehr-Wege-Methode keine der ausgesprochenen Sympathie und/oder Antipathien erfüllt?*

*Start-k-Sektion:  $P_c$ ;*

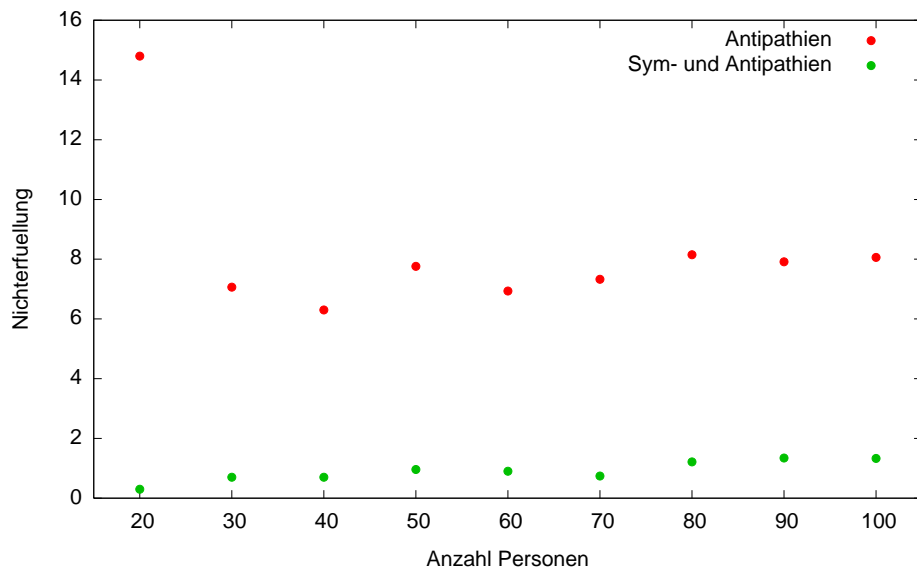
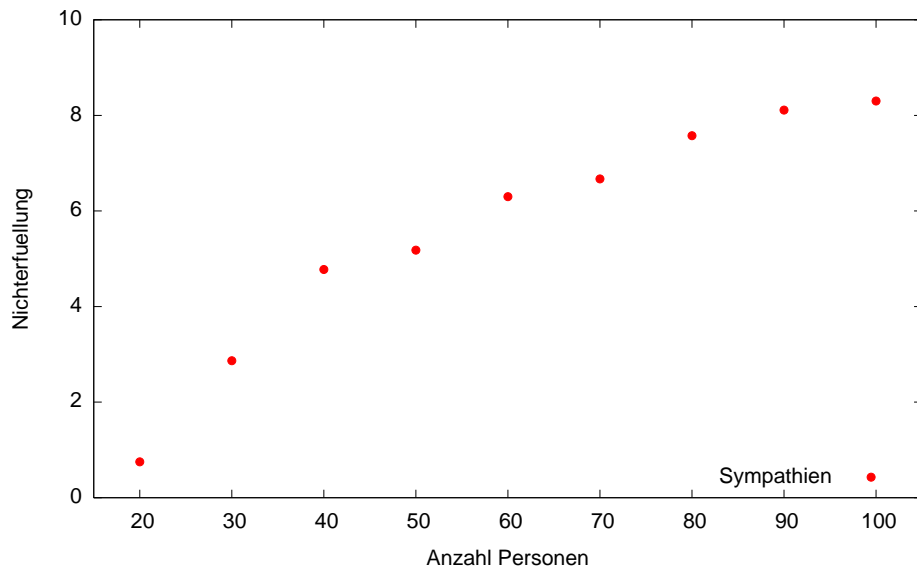
*MW-Durchlaufreihenfolge: zufällig;*

*Algorithmen: MinMax-Algorithmus.*

Hier ist eine deutliche Veränderung der Ergebnisse dieses Experimentes im Gegensatz zu den Ergebnissen des entsprechenden Experimentes 6.14



Abbildung 6.18: Experiment 6.18



in Abschnitt 6.4.8 zu erkennen. Betrachtet man die jeweils oberen Graphiken der zugehörigen Abbildungen (6.14 (c) und 6.18), so sieht man, dass bei der Aussprache von mehr Sympathien wesentlich weniger Teilnehmer keine Sympathie erfüllt bekommen. Im Gegenzug steigt allerdings die Anzahl Personen, denen keine Antipathie erfüllt wird, wie die unteren Graphiken zeigen. Während im Experiment 6.14 bei Graphen der Ordnung 100 ca. 30 Personen ohne Sympathieerfüllung sind und ungefähr eine Person ohne Antipathieerfüllung, sind es bei diesem Experiment je 8 Personen. Die Anzahl Personen ohne irgendeine Erfüllung steigt bei mehr Sympathien allerdings nur gering. Die Nichterfüllung der Sym- und Antipathie liegt hier bei ca. 1,5%. Es findet demnach bei den Nichterfüllungen eine Art „Verschiebung“ statt: Während die Erhöhung der Anzahl von Sympathien eine Verringerung der Nichterfüllung der Sympathien verursacht, erhöht sie gleichzeitig die Zahl an Personen ohne Antipathieerfüllung. Der betrachtete Fall von drei ausgesprochenen Sympathien und einer Antipathie hat demnach ergeben, dass eine „Mehrauswahl“ an Sympathien für den verwendeten Sektionierungsalgorithmus eine Verringerung der Nichterfüllung der Sympathien bewirkt.

#### 6.4.10 Doppelte Gewichtung der Sympathiegewinne

Die Experimente dieses Kapitels werden ebenfalls nur für die Mehr-Wege-Methode beschrieben. Hier werden, wie in Abschnitt 6.2.6 geschildert, in der Zielfunktion des MinMax-Algorithmus die Sympathiegewinne doppelt so stark gewichtet wie die Antipathiegewinne. Die jeweiligen Ergebnisse werden mit den entsprechenden Ergebnissen bei gleicher Gewichtung (1 : 1) verglichen.

**Experiment 6.19** *Welche Lösungen liefert die Mehr-Wege-Methode?*

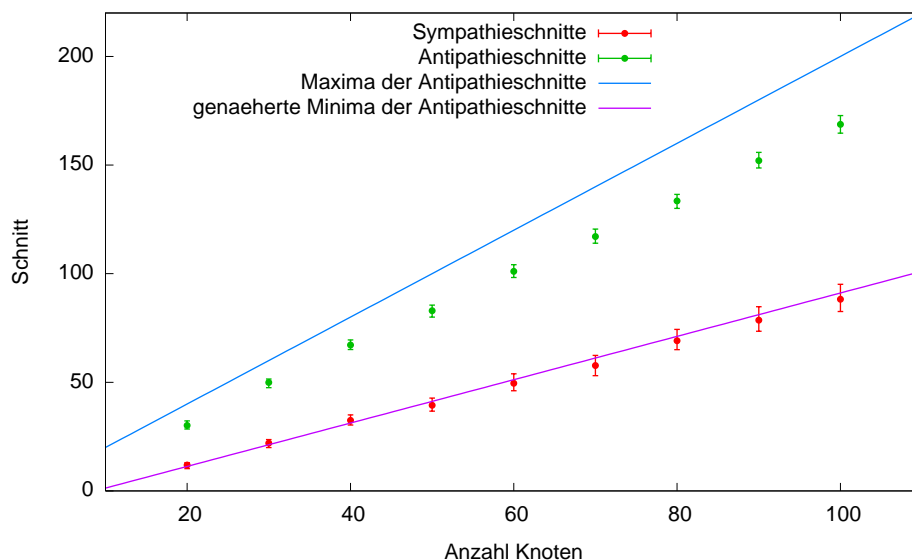
*Start-k-Sektionen:  $P_c$ ;*

*MW-Durchlaufreihenfolgen: zufällig;*

*Algorithmen: Min- und MinMax-Algorithmus.*

Vergleichen mit Abbildung 6.7 (h) „verschieben“ sich hier die Sympathie- und Antipathieschnitte des MinMax-Algorithmus nach unten. Durch diese „Verschiebung“ liegen die Sympathieschnitte sogar „auf“ der an die Minima genäherten Gerade. Das bedeutet, dass durch die doppelte Gewichtung der Sympathiegewinne bei der Sektionierung mit dem MinMax-Algorithmus annähernd genauso kleine Sympathieschnitte erzielt werden wie mit dem Min-Algorithmus. Dafür weichen allerdings die vom MinMax-Algorithmus erzielten Antipathieschnitte stärker von den absoluten Maxima ab, als dies bei gleicher Gewichtung der Fall ist. Es scheinen hier also mehr Sympathien

Abbildung 6.19: Experiment 6.19



und weniger Antipathien erfüllt zu werden als zuvor. Experiment 6.21 geht dieser Vermutung nach. Zuerst wird jedoch die Balanciertheit der Ergebnis-sektionen untersucht:

**Experiment 6.20** *Wie unbalanciert sind die Lösungen der Mehr-Wege-Methode?*

*MW-Durchlaufreihenfolgen: zufällig;*

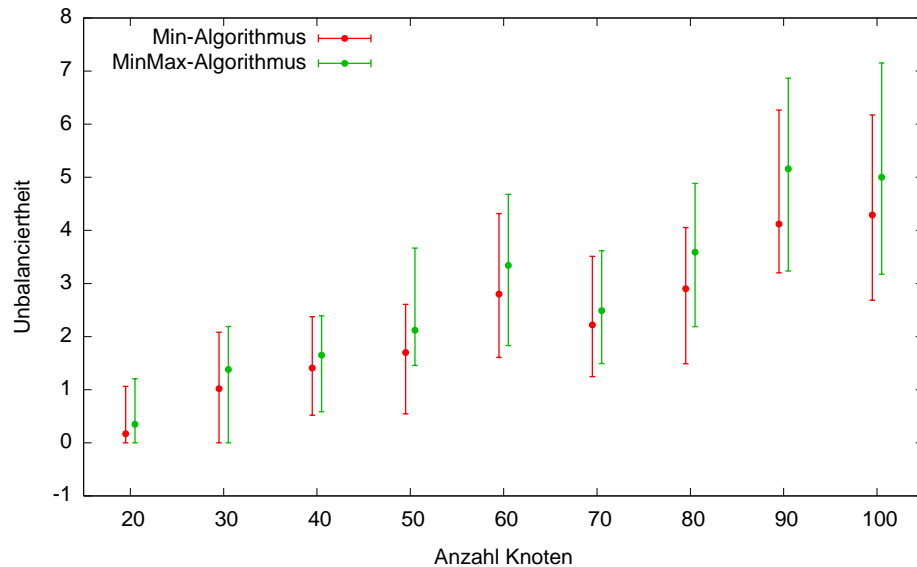
*Algorithmen: Min- und MinMax-Algorithmus.*

Die Ergebnisse des Min-Algorithmus verändern sich, wie in Abschnitt 6.2.6 erläutert, durch die doppelte Gewichtung der Sympathiegewinne nicht. Verwendet man deren Unbalanciertheiten aber zum Vergleich, so stellt man in Anbetracht der Abbildungen 6.20 und 6.10 (h) fest, dass sich die Unbalanciertheit der vom MinMax-Algorithmus erzielten Sektionen durch die Gewichtung verringert. Statt der Unbalanciertheit der Größe 10 beträgt diese in diesem Experiment bei Graphen der Ordnung 100 nur ungefähr 7. Die Balanciertheit der ausgegebenen Sektionen scheint sich somit bei doppelter Gewichtung der Sympathiegewinne an die Balanciertheit der vom Min-Algorithmus erzielten Sektionen „anzunähern“.

**Experiment 6.21** *Wie viel Prozent der ausgesprochenen Sym- bzw. Antipathien bei der Mehr-Wege-Methode werden pro Person erfüllt?*

*Start-k-Sektion:  $P_c$ ;*

Abbildung 6.20: Experiment 6.20



*MW-Durchlaufreihenfolge: zufällig;*  
*Algorithmen: MinMax-Algorithmus.*

Im Vergleich zur Graphik in Abbildung 6.13 (h) liegen die Punkte der Sympathien und die der Antipathien hier näher beisammen. Die Mitte zwischen ihnen scheint sich jedoch kaum zu verschieben. Das bedeutet, dass durch die doppelte Gewichtung der Sympathiegewinne mehr Sympathien und dafür weniger Antipathien pro Person erfüllt werden. Betrachte man beispielsweise wieder die Graphen der Ordnung 100: Während bei gleicher Gewichtung pro Person ca. 45% der ausgesprochenen Sympathien und 95% der Antipathien erfüllt werden, sind es hier 55% und 85%. Bedeutet diese Tatsache, dass auch hier der Prozentteil der Personen ohne Sympathieerfüllung geringer ist? Das folgende Experiment soll Auskunft darüber geben:

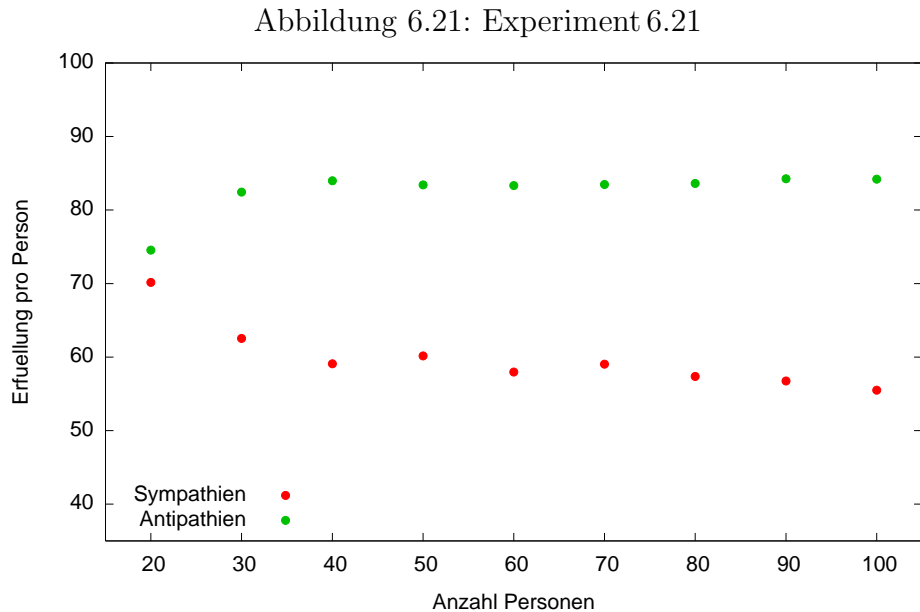
**Experiment 6.22** *Wie viel Prozent der Personen bekommen bei der Mehr-Wege-Methode keine der ausgesprochenen Sympathie und/oder Antipathien erfüllt?*

*Start-k-Sektion:  $P_c$ ;*

*MW-Durchlaufreihenfolge: zufällig;*

*Algorithmen: MinMax-Algorithmus.*

Auch hier zeichnet sich eine deutliche Veränderung gegenüber der entsprechenden Abbildung 6.14 (c) in Abschnitt 6.4.8 ab: Der Prozentteil der Personen ohne Sympathieerfüllung verringert sich bei allen Graphen ungefähr auf



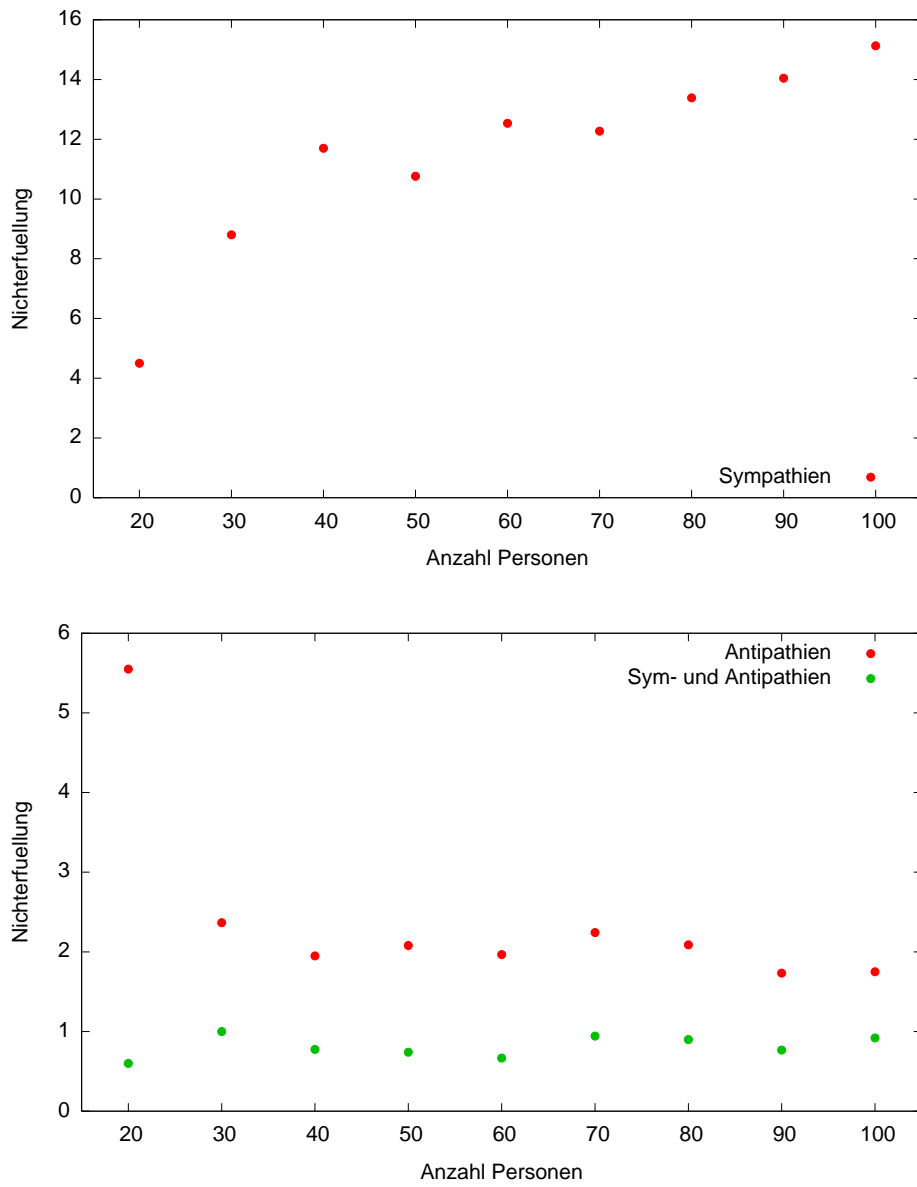
die Hälfte. Dabei vergrößert sich jedoch der Prozentteil Personen ohne Antipathieerfüllung im Schnitt von 0,5% auf ca. 2%, wie die jeweils unteren Graphiken zeigen. Es bleibt dabei aber in allen Graphen die Anzahl Personen ohne irgendeine Erfüllung unter einem Prozent der Personenzahl. Das bedeutet also, dass die doppelte Gewichtung der Sympathien gegenüber der Antipathien eine Verbesserung der Nichterfüllung der Sympathien hervorruft, ohne dabei eine starke Vergrößerung der Nichterfüllung der Antipathien mit sich zu ziehen. Gewichtet man demnach in der Zielfunktion des Sektionierungsalgorithmus die Sympathiegewinne doppelt so stark wie die Antipathiegewinne, so wird bei der Gruppeneinteilung einer höheren Anzahl an Personen mindestens eine ausgesprochene Sympathie erfüllt.

## 6.5 Auswertung

Bei allen betrachteten Algorithmen wachsen die jeweiligen Anzahlen an Durchläufen mit größer werdender Ordnung der Graphen nur geringfügig. In Kapitel 5 notierte Laufzeiten der Algorithmen vergrößern sich daher in diesen Fällen nur um einen konstanten Faktor.

Unabhängig von der verwendeten Startbisektion liefert die Bisektionierung mit dem Kernighan-Lin-Algorithmus, wie auch die  $k$ -Sektionierung mit der Zwei-Wege-Methode Sektionen mit gleichem Schnitt. Die Untersuchung

Abbildung 6.22: Experiment 6.22



der Gruppenbalanciertheit und der Erfüllung der Sym- und Antipathien für die Mehr-Wege-Methode ergibt ebenfalls keine Unterschiede bei der Verwendung von  $P_a$ ,  $P_b$  und  $P_c$ . Bisektion  $P_c$  scheint lediglich „näher“ an einer lokal minimalen Bisektion zu liegen als die anderen.

Die Mehr-Wege-Methode ist allerdings wesentlich vielseitiger zur Gruppeneinteilung einsetzbar: Mit ihr können beliebige lokal optimale  $k$ -Sektionen ermittelt werden, während die Sektionierung mit der Zwei-Wege-Methode auf  $k$ -Sektionen mit  $k = 2^r$ ,  $r \in \mathbb{N}$  beschränkt ist. Vergleicht man allerdings die Sympathieschnitte der erzielten lokal minimalen Sektionen miteinander, so ist die Erfolgsrate des Min-Algorithmus unter Einsatz der Zwei-Wege-Methode wesentlich größer. Dasselbe gilt für die Sympathieschnitte der vom MinMax-Algorithmus erzielten Sektionen: Diese sind bei Verwendung der Mehr-Wege-Methode ebenfalls schlechter. Wünscht man also eine  $k$ -Sektionierung mit  $k = 2^r$ ,  $r \in \mathbb{N}$ , so erhält man bessere Sektionen, wenn man die Zwei-Wege-Methode unter Eingabe einer beliebigen Startbisektion verwendet.

Bei der Mehr-Wege-Methode liegt die Start- $k$ -Sektion  $P_c$  für die Graphen aller betrachteten Ordnungen „näher“ an einem lokalen Minimum als die Übrigen. Für die hier erzielten Sektionen spielt neben der Startsektion zusätzlich die verwendete Durchlaufreihenfolge eine Rolle: Während  $P_a$  in jedem Fall schlechtere lokal optimale Sektionen liefert, verbessert bei den Startsektion  $P_a$  und  $P_b$  die Verwendung der zufälligen Reihenfolge die Ergebnisse. Startsektion  $P_c$  scheint relativ unabhängig von der verwendeten Reihenfolge zu sein. Auch in Anbetracht der Gruppenbalanciertheit und der Erfüllung der Sym- und Antipathien zeigt sich, dass  $P_a$  sich eher nicht als Startsektion eignet,  $P_b$  nur unter Verwendung der zufälligen Reihenfolge, und  $P_c$  kann bei jeder verwendeten Reihenfolge gut zur Sektionierung eingesetzt werden.

In Anbetracht der Erfüllung der Sym- und Antipathien kann allgemein festgehalten werden, dass pro Person stets mehr Anti- als Sympathien erfüllt werden. Der Anteil an Personen, die bei der Gruppeneinteilung keine der ausgesprochenen Sym- sowie Antipathien erfüllt bekommen, beträgt immer weniger als 1%. Wünscht man, dass bei der Gruppeneinteilung mehr Sympathien erfüllt werden, so hat man zwei Möglichkeiten: Zum einen kann man pro Person mehr Sympathien und dafür weniger Antipathien aussprechen lassen, und zum anderen kann man in der Zielfunktion des MinMax-Algorithmus die Sympathiegewinne stärker gewichten. Bei drei ausgesprochenen Sympathien und einer Antipathie geht die Verringerung der Personenzahl ohne Sympathieerfüllung dann aber mit einer stärkeren Vergrößerung der Personenzahl ohne Antipathieerfüllung einher. Diese Auswirkung fällt bei doppelter Gewichtung der Sympathieschnitte geringer aus. Der Prozentteil Personen ohne irgendeine Erfüllung bleibt bei beiden Möglichkeiten niedrig (unter 2%).





# Kapitel 7

## Zusammenfassung und Ausblick

Der Knödel-Graph stellt sich in Anbetracht der angegebenen Kriterien als ein günstiger Kontaktgraph heraus. Er ermöglicht unter anderem eine Gesprächsführung ohne viele Aussetzer und ist flexibel, was die Anzahl teilnehmender Personen und einzuteilender Gruppen betrifft. Die im Anschluß an die flüchtigen Kontakte ausgesprochenen Sym- und Antipathien erzeugen zwei verschiedene Untergraphen des Knödel-Graphen, den Sympathie- und den Antipathiegraphen. Die beschriebene Graphkonvertierung garantiert dabei, dass Sym- bzw. Antipathien, die erwidert sind, bei der Gruppeneinteilung stärker berücksichtigt werden, als die einseitigen Sym- und Antipathien. Diese Form der Bewertung macht außerdem eine Erweiterung durch Skalierung der Sym- und Antipathien möglich. Der Kernighan-Lin-Algorithmus mit entsprechend gewählter Zielfunktion liefert schließlich eine optimierte Knotenbisektion, dessen Schnitt im Sympathiegraphen minimiert und zugleich im Antipathiegraphen maximiert ist. Sowohl mit Hilfe der Zwei-, als auch mit Hilfe der Mehr-Wege-Methode können auf Basis dieser Bisektionierung optimierte  $k$ -Sektionen für  $k > 2$  gebildet werden, wobei die Zwei-Wege-Methode nur  $k = 2^r$  ( $r \in \mathbb{N}$ ) zulässt. Die Elemente dieser optimierten  $k$ -Sektionen stellen letztendlich die Einteilungen der Personen in gleich große Gruppen dar.

Die Experimente zeigen, dass bei einer  $k$ -Sektionierung mit  $k = 2^r$  ( $r \in \mathbb{N}$ ) die Zwei-Wege-Methode bessere Ergebnisse liefert. Sie ist außerdem im Gegensatz zur Mehr-Wege-Methode unabhängig von der verwendeten Startsektion. Benötigt man allerdings andere  $k$ -Sektionen, so eignen sich am besten die Start- $k$ -Sektion  $P_c$  und die zufällige Durchlaufreihenfolge. Legt man bei der Sektionierung mehr Wert auf Gruppenbalanciertheit und eine geringere Nichterfüllung der Sympathien, so kann man entweder mehr Sympathien als Antipathien aussprechen lassen oder die Gewichtung der Sympathiegewinne in der Zielfunktion des Kernighan-Lin-Algorithmus erhöhen.

Dadurch, dass im bewegungsbasierten Kernighan-Lin-Algorithmus immer

zwei Knoten ausgetauscht werden, erreicht man Ergebnis-sektionen, die nur lokal optimal sind. Wie in den Experimenten gesehen, liefert die Mehr-Wege-Methode unter Verwendung der unterschiedlichen Startsektionen und Durchlaufreihenfolgen mal bessere und mal schlechtere  $k$ -Sektionen. Es stellt sich die Frage, ob man die im Algorithmus verwendete Zielfunktion „aufweichen“ kann, um temporär eventuell schlechtere Sektionen zu erzeugen, schließlich aber bessere lokale Optima zu treffen. Ein solches „Aufweichen“ könnte beispielsweise der zirkuläre Austausch zwischen drei oder mehr Sektionselementen sein.

Sozialwissenschaftler der Universität Tilburg (Niederlande) haben bereits ein erstes Experiment zur Untersuchung von Gruppen durchgeführt, die auf Basis flüchtiger Kontakte eingeteilt werden. Diese empirische Studie legt nahe, dass auf diese Weise gebildete Arbeitsgruppen in mehrfacher Hinsicht signifikant effektiver sind, als zufällig oder attributbasiert zusammengestellte (siehe [BC<sup>+</sup>05]). Eine Möglichkeit für eine eventuell weitere Verbesserung der Gruppen wäre, die hier entwickelte Strategie mit einer Einteilung der Personen nach Interessengebieten zu kombinieren.

# Anhang



# Anhang A

## ‚Gossiping‘-Zeit des $K_n$

W. Knödel hat in [Knö74] wie folgt die ‚Gossiping‘-Zeit vollständiger Graphen  $K_n$  bestimmt:

**Satz A.1** Für vollständige Graphen  $K_n$  gilt:

$$g(K_n) = \begin{cases} \lceil \log_2 n \rceil & \text{für } n \text{ gerade,} \\ \lceil \log_2 n \rceil + 1 & \text{für } n \text{ ungerade.} \end{cases}$$

**Beweis:** Dass die gegebenen Werte die besten oberen Schranken sind, folgt aus der Tatsache, dass es  $k$  Telefonate braucht, um eine Information von einer Person an  $2^k$  Personen zu verbreiten. Somit sind bei geradem  $n$  mindestens  $\lceil \log_2 n \rceil$  Telefonate nötig, bei ungeradem  $n$  mindestens ein Telefonat mehr, da eine Person in der ersten Runde keinen Partner zum Telefonieren hat.

Betrachte man zuerst den Fall  $n = 2^k$  und nummeriere die Personen von 0 bis  $n-1$ . Führen in der ersten Runde die Personen  $2j$  und  $2j+1$ ,  $j = 0, \dots, 2^{k-1}-1$ , ein Telefonat, so können in den folgenden Runden die Personen mit gerader bzw. ungerader Nummerierung rekursiv dieselbe Prozedur mit  $2^{k-1}$  Personen durchführen. Nach  $k$  Runden weiß schließlich jede Person die Information jeder anderen Person. Es gilt demnach:  $g(K_{2^k}) = k = \lceil \log_2 (2^k) \rceil$ .

Bei beliebigem ungeradem  $n$  wählt man  $2^{\lceil \log_2 n \rceil}$  Personen aus und lässt alle übrigen Personen in der ersten und letzten Runde mit je einer von ihnen telefonieren. Mit den ausgewählten Knoten verfährt man zwischen diesen beiden Runden wie im oben beschriebenen Fall. So erfahren die  $2^{\lceil \log_2 n \rceil}$  ausgewählten Personen zuerst die Informationen der übrigen Personen, verbreiten alle Informationen unter sich und tragen sie anschließend den übrigen Personen zu. Wegen  $\lceil \log_2 n \rceil + 2 = \lceil \log_2 n \rceil + 1$  ist die obere Schranke von  $g(K_n)$  somit gleich der unteren Schranke.

Im generellen Fall von geradem  $n$  telefonieren bei gleicher Nummerierung

wie oben in der  $r$ -ten Runde,  $r = 1, \dots, \lfloor \log_2 n \rfloor$ , die Personen  $2j$  und  $2j + 2^r - 1 \pmod n$ ,  $j = 0, \dots, n/2 - 1$ , miteinander. Durch Induktion über  $r$  kann gezeigt werden, dass in Runde  $r$  die Personen  $0$  und  $2^r - 1$  die Information aller Personen mit dazwischen liegender Nummerierung kennen. Dies gilt auch für alle Personen  $2j$  und  $2j + 2^r - 1 \pmod n$ ,  $j = 0, \dots, n/2 - 1$ . Folglich weiß nach  $r$  Runden jede Person die Information jeder anderen Person.  $\square$

# Anhang B

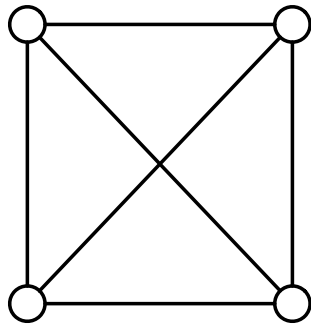
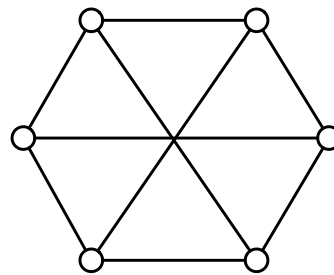
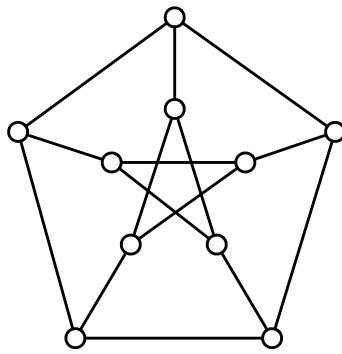
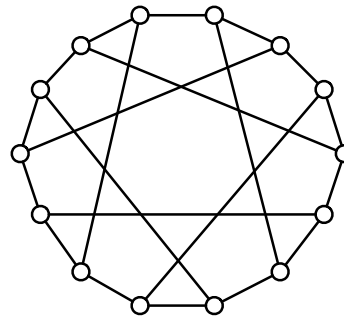
## Bestimmte $(r, t)$ -Cages

Für  $(r, t)$ -Cages mit  $r = 3$  und  $t \in \{3, 4, 5, 6\}$  gilt aufgrund der Existenz der Graphen in Abbildung B.1 folgende Gleichheit aus Satz 3.4:

$$n^*(r, t) = \begin{cases} 1 + \frac{r}{r-2} \cdot ((r-1)^{(t-1)/2} - 1) & \text{für ungerades } t, \\ 2 \cdot (1 + \frac{r-1}{r-2} \cdot ((r-1)^{(t-2)/2} - 1)) & \text{für gerades } t. \end{cases}$$

Man erhält folglich (siehe Abbildung B.1 (a)–(d)):

- $n^*(3, 3) = 4,$
- $n^*(3, 4) = 6,$
- $n^*(3, 5) = 10,$
- $n^*(3, 6) = 14.$

Abbildung B.1: Minimale  $(r, t)$ -Cages für  $r = 3 \wedge t \in \{3, 4, 5, 6\}$ (a) Graph  $K_4$ ,  
minimaler  $(3, 3)$ -Cage(b) Graph  $K_{3,3}$ ,  
minimaler  $(3, 4)$ -Cage(c) Petersen-Graph,  
minimaler  $(3, 5)$ -Cage(d) Heawood-Graph,  
minimaler  $(3, 6)$ -Cage



# Anhang C

## Inhalt der beigefügten CD

Auf der beigefügten CD befinden sich die für diese Arbeit verwendeten Java-Programme. Diese setzen sich zusammen aus dem Generator für den Knödel-Graphen, die beschriebenen Algorithmen zur Knotensektionierung der Graphen und dem Generator für die Daten, die in den Experimenten verwendet wurden.



# Literaturverzeichnis

- [Ahn04] S. J. Ahn. *Least Squares Orthogonal Distance Fitting of Curves and Surfaces in Space*. Lecture Notes in Computer Science, Vol. 3151, 2004.
- [AK95] C. J. Alpert, A. B. Kahng. *Recent Directions in Netlist Partitioning: A Survey*. In: Integration - The VLSI Journal, Vol. 19 (1-2), 1995, S. 1-81.
- [BC<sup>+</sup>05] U. Brandes, P. Curseu, G. Dorfmüller, M. Gaertler, P. Kenis, J. Raab. *Teamdating: Does Speed Dating Generalize To Effective Teams?* Konferenzbeitrag bei der International Sunbelt Social Network Conference (Sunbelt XXV). Redondo Beach, Kalifornien, USA, 16.-20. Februar 2005.
- [Die96] R. Diestel. *Graphentheorie*. Springer Verlag, Berlin, 1996.
- [Eve79] S. Even. *Graph Algorithms*. Computer Science Press, USA, 1979.
- [FR98] G. Fertin, A. Raspaud. *Families of Graphs Having Broadcasting and Gossiping Properties*. In: J. Hromkovič, O. Sýkora. *Graph-Theoretic Concepts in Computer Science: 24th International Workshop, WG '98*. Lecture Notes in Computer Science, Vol. 1517, 1998, S. 63-77.
- [FR<sup>+</sup>00] G. Fertin, A. Raspaud, H. Schröder, O. Sýkora, I. Vrto. *Diameter of the Knödel Graph*. In: U. Brandes, D. Wagner. *Graph-Theoretic Concepts in Computer Science: 26th International Workshop, WG 2000*. Lecture Notes in Computer Science, Vol. 1928, 2000, S. 149-160.
- [GJS76] M. R. Garey, D. S. Johnson, L. Stockmeyer. *Some Simplified NP-Complete Graph Problems*. In: Theoretical Computer Science, Vol. 1, 1976, S. 237-267.

- [Jør03] L. K. Jørgensen. *Girth 5 Graphs from Relative Difference Sets*. <http://www.math.auc.dk/~leif/research/girth5/girth5.pdf>.
- [KL70] B. W. Kernighan, S. Lin. *An Efficient Heuristic Procedure for Partitioning Graphs*. In: The Bell System Technical Journal, Vol. 49, 1970, S. 291-307.
- [Knö74] W. Knödel. *New Gossips and Telephones*. In: Discrete Mathematics, Vol. 13, 1975, S. 95.
- [KK95] G. Karypis, V. Kumar. *A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs*. <http://www-users.cs.umn.edu/~karypis/publications/partitioning.html>.
- [LP86] L. Lovász, M. D. Plummer. *Matching Theory*. North Holland Mathematics Studies, Vol. 121, North Holland, 1986.
- [SR04] M. Sunnafrank, A. Ramirez (Jr.). *At First Sight: Persistent Relational Effects of Get-Acquainted Conversations*. In: Journal of Social and Personal Relationships, Vol. 21 (3), 2004, S. 361-379.
- [Sta04] C. Stadtmüller. *Temporeiches Liebesspielchen*. In: Südkurier, Ausgabe vom 06.07.2004.  
C. Stadtmüller. *Anbandeln in zehn Minuten*. In: Südkurier, Ausgabe vom 24.07.2004.
- [SW97] M. Stoer, F. Wagner. *A Simple Min-Cut Algorithm*. In: Journal of the ACM, Vol. 44 (4), 1997, S. 585-591.
- [Vol91] L. Volkmann. *Graphen und Digraphen: eine Einführung in die Graphentheorie*. Springer Verlag, Wien, 1991.
- [Wei01] E. W. Weisstein. „Cage Graph“. From MathWorld – A Wolfram Web Resource. <http://mathworld.wolfram.com/CageGraph.html>.
- [WV74] H. Walther, H.-J. Voss. *Über Kreise in Graphen*. VEB Deutscher Verlag der Wissenschaften, Berlin, 1974.

# Danksagung

An dieser Stelle möchte ich meinem Betreuer Prof. Dr. U. Brandes für seine wertvollen Ratschläge und die fortwährende Unterstützung herzlich danken. Ebenfalls danke ich meinen Kommilitonen, die Interesse an meiner Arbeit zeigten und mir durch anregende Gespräche hilfreich zur Seite standen. Ich danke außerdem den Mitarbeiterinnen des Südkurier Konstanz für die Möglichkeit, den Ablauf eines Speed-Dating beobachten zu können. Nicht zuletzt geht mein Dank an meine Familie und meine Freunde, für ihre Geduld, ihr Verständnis und ihre Unterstützung während der Erstellung der Diplomarbeit.



# Eidesstattliche Erklärung

Hiermit versichere ich, dass diese Arbeit von mir selbstständig verfasst wurde, und ich keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

---

Ort, Datum

---

Unterschrift