

Ludwig-Maximilians-Universität München

Institut für Statistik

Projekt im Rahmen des statistischen Consultings

---

# **Internationaler Waffenhandel: Die Anwendung neuer Verfahren der statistischen Netzwerkanalyse**

Eine Netzwerkanalyse des internationalen Kleinwaffenhandels 1992 - 2011  
Kooperation mit dem Lehrstuhl für empirische Politikforschung

---

*Autor:*

Felix Loewe  
loewe.felix@gmail.com

*Projektpartner:*

Prof. Dr. Paul W. Thurner

*Betreuer:*

Prof. Dr. Göran Kauermann

## **Abstract**

hier kommt noch das abstract rein

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>4</b>
<b>2</b>	<b>Zusammenfassung der Graphentheorie</b>	<b>5</b>
2.1	Graph - Grundlegende Begriffe . . . . .	5
2.2	Degree-Verteilung . . . . .	6
2.3	Nachbarschaftsmatrix . . . . .	6
2.4	Zentralität . . . . .	7
2.5	Dichte . . . . .	7
<b>3</b>	<b>Datengrundlage</b>	<b>9</b>
<b>4</b>	<b>Deskriptive Analyse</b>	<b>10</b>
4.1	Deskriptive Netzwerkmaßzahlen . . . . .	10
4.2	Die Degree-Verteilung . . . . .	11
4.3	Handelswerte . . . . .	13
4.4	Top-Akteure . . . . .	14
4.5	Visualisierungen . . . . .	18
<b>5</b>	<b>Exponential Random Graph Model</b>	<b>20</b>
5.1	Definition des Modells . . . . .	20
5.1.1	Simulation von Zufallsgraphen . . . . .	21
5.1.2	Schätzung der Modell-Parameter . . . . .	22
5.2	Endogene Statistiken . . . . .	24
5.3	Degeneration und alternative Modellierung . . . . .	25
5.4	Exogene Kovariablen . . . . .	26
5.4.1	Polity Score . . . . .	27
5.4.2	Composite Index of National Capability . . . . .	27
5.4.3	Intra-State Conflicts . . . . .	27
5.4.4	GDP . . . . .	27
5.5	geschätzte Modelle . . . . .	27
5.5.1	Interpretation der Parameter . . . . .	28
5.5.2	Überprüfung der Ergebnisse . . . . .	31
	<b>Literaturverzeichnis</b>	<b>34</b>

# 1 Einleitung

Die statistischen Methoden der Netzwerkanalyse finden ein breites Anwendungsfeld. Neben klassischen Anwendungen wie in den Sozialwissenschaften (Freundesnetzwerk, Kollegenkreis) und Politikwissenschaften (internationale Beziehungen) ist sie auch in hoch modernen Bereichen der Informatik (Computernetzwerk, soziale Netzwerke im Internet) und der Biologie (Proteinverbindungen der menschlichen DNA) von großem Nutzen.

Was ist jedoch das Besondere an der statistischen Analyse von Netzwerken? Durch ihre Abhängigkeitsstruktur stellen sie relationale Daten dar. Gewöhnliche Datensätze werden in der Statistik zumeist mit der Annahme analysiert, dass die Beobachtungen als unabhängig voneinander betrachtet werden können. Bei Netzwerkdaten ist das nicht der Fall. Hier stehen die Beobachtungen, oft genannt *Akteure* des Netzwerkes, in einer Beziehung zueinander die bei der Analyse der Daten selbst von zentralen Interesse ist. Besteht jedoch eine Beziehung zwischen den Beobachtungen, können diese nicht mehr als unabhängig angesehen werden, was alternative Methoden der statistischen Modellierung erfordert.

Das Bestehen oder Nicht-Bestehen einer Beziehung bezeichnet man als die Netzwerkstruktur (Abhängigkeitsstruktur), die zusätzlich zu den Daten eines gewöhnlichen Datensatzes besteht. Die Abhängigkeitsstruktur wird durch die *Adjazenzmatrix*  $A_{ij} \in N \times N$  ausgedrückt.

Gegenstand der Analyse in dieser Arbeit ist ein Datensatz zum internationalen Handel mit kleinen und leichten Waffen. Bei der Analyse des Waffenhandels wird aus den Akteuren die exportierenden und importierenden Länder und aus deren Abhängigkeit ein Handel.

Die Arbeit ist wie folgt aufgebaut: Im ersten Kapitel erfolgt eine kurze Einführung in die Grundlagen der Graphentheorie. Netzwerkspezifische Begriffe sowie deskriptive Maßzahlen werden definiert. Darauf folgt eine Erläuterung der Datengrundlage der NISAT Datenbank mit den daraus resultierenden Möglichkeiten und Einschränkungen. Im dritten Abschnitt wird der Datensatz deskriptiv analysiert. Darauf aufbauend folgt im vierten Abschnitt die Modellierung per ERGMs. Im Anschluss werden die Ergebnisse der vorhergehenden Analyse mit denen einer ähnlichen Arbeit über den Handel mit Großwaffen verglichen. Im letzten Abschnitt werden die erarbeiteten Ergebnisse zusammengefasst und diskutiert.

## 2 Zusammenfassung der Graphentheorie

Um Netzwerkdaten statistisch analysieren zu können, muss die Abhängigkeitsstruktur der Daten adäquat modelliert werden. Eine grundlegende mathematische Theorie, die verwendet wird, um *relationale Daten* zu beschreiben, ist die Graphentheorie. Im Rahmen dieses Kapitels wird nur auf die wichtigsten Aspekte der Graphentheorie eingegangen. Es handelt sich im Wesentlichen um eine Zusammenfassung der Kapitel 2.1 und 4.2 aus Kolaczyk (2009).

### 2.1 Graph - Grundlegende Begriffe

Ein *Graph*  $G = (V, E)$  ist die mathematische Beschreibung eines Netzwerkes.

Er besteht aus einer Knotenmenge  $V$  und einer Kantenmenge  $E$ . Ein *Knoten*  $v$  repräsentiert einen Akteur des Netzwerkes. Eine *Kante*  $e$  verbindet zwei Akteure und kennzeichnet eine Beziehung zwischen ihnen. Die Anzahl der Knoten  $N_V = |V|$  wird üblicherweise als kleiner unendlich vorausgesetzt. Häufig benennt man die Knoten eines Netzwerkes einfach nach ihrem Index  $i = 1, \dots, N_V$ . Eine Kante  $\{i, j\}$  ist ein Element der Menge  $E$  und beschreibt die Verbindung zwischen Knoten  $i$  und  $j$ . Man unterscheidet zwischen *gerichteten* und *ungerichteten* Graphen. Ein ungerichteter Graph setzt eine symmetrische Beziehung zwischen den Akteuren voraus. Akteur  $i$  steht also zu Akteur  $j$  in der gleichen Beziehung wie  $j$  zu  $i$  (z.B. Facebook Freundschaft). Bei einem gerichteten Graphen hingegen ist die Richtung der Beziehung entscheidend. Dies ist in unserem Datensatz der Fall. Wir unterscheiden zwischen Exporteur und Importeur eines Handels. Die Kante  $\{i, j\}$  ist hier also von der Kante  $\{j, i\}$  zu unterscheiden. Einen gerichteten Graphen nennt man auch *Digraph*.

Per Definition enthält ein Graph weder Schleifen noch multiple Kanten. Von einer *Schleife* spricht man, wenn eine Kante  $\{i, j\}$  den gleichen Anfangs- und Endpunkt besitzt ( $i = j$ ). Von einer *multiplen Kante* spricht man, falls zwischen zwei Knoten mehrere Verbindungen bestehen. Enthält ein Netzwerk solche Eigenschaften spricht man von einem *Multigraphen* ansonsten von einem *einfachen Graphen*.

Betrachtet man die Anzahl der möglichen Kanten eines Graphen, die später als Benchmark dafür auftaucht, wie dicht ein Graph sein kann, so wird ersichtlich, dass diese Anzahl für einen einfachen ungerichteten Graphen kleiner ist als die Anzahl der möglichen Kanten eines gerichteten Graphen. Für einen einfachen ungerichteten Graphen ist sie gegeben

## 2 Zusammenfassung der Graphentheorie

durch  $V_H(V_H - 1)/2$ . Ein gerichteter Graph kann logischerweise maximal die doppelte Anzahl an Kanten enthalten.

Notwendigerweise gibt es einige Begriffe um über die Konnektivität von Graphen zu reden. Am gebräuchlichsten ist das Prinzip der *Nachbarschaft*. Zwei Knoten  $i, j \in V$  gelten als benachbart wenn sie von einer Kante  $\{i, j\} \in E$  verbunden werden. Zwei Kanten wiederum gelten als benachbart, wenn beide einen gemeinsamen Knoten aus  $V$  enthalten.

Ein sogenannter *Weg* auf einem Graphen von  $v_0$  nach  $v_l$  bezeichnet die alternierende Sequenz  $\{v_0, e_1, v_1, e_2, \dots, v_{l-1}, e_l, v_l\}$ .  $l$  bezeichnet hierbei die Länge des Weges. Weiter beschreibt ein *Pfad* einen Weg ohne wiederholte Knoten oder Kanten. Als *Distanz* zwischen zwei Knoten eines Graphen definiert man die Länge des kürzesten Pfades der sie verbindet. Die Länge der längsten Distanz innerhalb eines Graphen nennt man den *Durchmesser* oder *Diameter* eines Graphen.

Soviel zu den grundlegenden Begrifflichkeiten der Graphentheorie. Mit ihrer Hilfe werden nun in den folgenden Unterpunkten einige für die folgenden Analysen besonders wichtige Definitionen erläutert.

### 2.2 Degree-Verteilung

Ein weiterer Begriff ist der des *Degree* eines Knotens  $d_v$ , definiert als die Anzahl der Kanten, die den Knoten  $v$  enthalten. Die *Degree-Sequenz* eines Graphen  $G$  ist die Sequenz die entsteht, wenn man alle Knoten-Degrees in aufsteigender Reihenfolge sortiert. Nun definiert man  $f_d$  als den Anteil der Knoten  $v \in V$  mit Degree  $d_v = d$ . Die Reihe  $\{f_d\}_{d \geq 0}$  heißt *Degree-Verteilung* von  $G$ .

Bei einem gerichteten Graphen unterscheidet man zwischen *In-Degree* ( $d_v^{in}$ ) und *Out-Degree* ( $d_v^{out}$ ) und zählt nur die Kanten die zu einem Knoten hin, beziehungsweise von ihm weg laufen. Die eben genannten Begriffe lassen sich hierfür trivial erweitern.

### 2.3 Nachbarschaftsmatrix

Die Struktur eines Graphen ist vollständig bestimmt durch seine binäre und symmetrische *Nachbarschaftsmatrix* (auch *Adjazenzmatrix*)  $A \in |V| \times |V|$ , wobei  $a_{ij} = 1$  falls zwischen Knoten  $i$  und Knoten  $j$  eine Kante besteht und  $a_{ij} = 0$  sonst. Diese Matrix hat einige nützliche Eigenschaften. Zum Beispiel ergibt die Zeilensumme  $A_{i,} = \sum_j A_{ij}$  den Degree  $d_i$  von Knoten  $i$ . Bei gerichteten Graphen besteht ein Unterschied zwischen der Kante  $(i, j)$  und der Kante  $(j, i)$ . Die Adjazenzmatrix ist hier also nicht symmetrisch. Allerdings enthält sie immer noch ähnliche Informationen wie zum Beispiel  $A_{i,} = d_i^{out}$  und  $A_{,j} = d_j^{in}$ .

## 2.4 Zentralität

Viele Fragestellungen der Netzwerkanalyse drehen sich darum welche Akteure auf eine bestimmte Weise besonders wichtig für ihr Netzwerk sind. Maßzahlen der Zentralität sind dafür gedacht um diese Wichtigkeit der Knoten zu quantifizieren und dadurch die Beantwortungen solcher Fragestellungen zu vereinfachen. Das wichtigste Konzept in diesem Zusammenhang, den Degree haben wir bereits kennen gelernt. Zusätzlich werden nun kurz die Begriffe Closeness- und Betweenness Zentralität eingeführt.

Bei der *Closeness* Zentralität wird ein Knoten als zentral angesehen, falls er vielen anderen Knoten „nahe“ ist. Man verwendet für dieses Maß die Inverse der Summe der Distanzen eines Knotens zu allen anderen,

$$c_{Cl}(v) = \frac{1}{\sum_{u \in V} dist(v, u)},$$

wobei  $dist(v, u)$  die Distanz zwischen den Knoten  $v$  und  $u$  bezeichnet. Um eine Vergleichbarkeit mit anderen Graphen und Zentralitätsmaßen herzustellen normalisiert man die Maßzahl in das Intervall  $[0, 1]$  indem man mit dem Faktor  $N_v - 1$  multipliziert.

Ein anderes Konzept basiert auf dem Gedanken, dass die Wichtigkeit eines Knotens darauf beruht, wie oft sich ein Knoten auf der Verbindung zwischen anderen Paaren von Knoten befindet. Knoten die auf vielen Pfaden eines Netzwerkes sitzen, werden als wichtig für die Kommunikation innerhalb des Netzwerkes angesehen. Die *Betweenness* Zentralität wird deswegen üblicherweise definiert als

$$c_B(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma(s, t|v)}{\sigma(s, t)},$$

wobei  $\sigma(s, t|v)$  die Anzahl der kürzesten Pfade zwischen  $s$  und  $t$  ist, die durch  $v$  verlaufen und  $\sigma(s, t) = \sum_v \sigma(s, t|v)$ .

## 2.5 Dichte

Die *Dichte* eines Netzwerkes ist die Anzahl der Kanten des Netzwerkes, geteilt durch die mögliche Anzahl an Kanten des Netzwerkes. Es ergibt sich der Term

$$den(G) = \frac{|E_G|}{|V_G|(|V_G| - 1)/2}$$

für die Dichte eines ungerichteten Netzwerkes und

$$den(G) = \frac{|E_G|}{|V_G|(|V_G| - 1)}$$

## 2 Zusammenfassung der Graphentheorie

für ein gerichtetes Netzwerk. Die Dichte liegt zwischen null und eins. Es sei angemerkt, dass die Dichte auch als Skalierung des durchschnittlichen Degrees  $\bar{d}(G)$  angesehen werden kann, denn  $den(G) = (|V_G| - 1)\bar{d}(G)$ .



### 3 Datengrundlage

Die Datengrundlage für das Kleinwaffenhandelsnetzwerk ist die *NISAT* (Norwegian Initiative on Small Arms Transfers) Datenbank. Das Peace Research Institute Oslo (PRIO) ist der Auftraggeber dieser Datenbank. Das Institut hat es sich zur Aufgabe gemacht „[...] die Bedingungen die ein friedliches Zusammenleben zwischen Staaten, Gruppen und Menschen [...]“<sup>1</sup> ermöglichen zu erforschen. Die Datenbank enthält Daten über den Handel von kleinen und leichten Waffen. Der abgedeckte Zeitraum beträgt die Jahre 1992 bis 2011. Berichtet wird hierin von insgesamt 239 Ländern und 109522 Waffentransaktionen.

Die Daten liegen in der Form einer gerichteten Kanten-Liste vor. Das bedeutet jede Zeile im Datensatz entspricht einem Handel zwischen einem exportierenden und einem importierenden Land. Zusätzliche Attribute ist der Correlates of War Codes des jeweiligen Landes, der monetäre Wert des Handels in US Dollar, der gehandelte Waffentyp, die berichtende Datenquelle sowie das Jahr in dem der Handel stattgefunden hat. Bei der Betrachtung dieser Daten fallen folgende Probleme auf:

Da nach Waffentypen unterschieden wird, existieren in den einzelnen Jahren multiple Kanten. Das bedeutet zwischen zwei Ländern werden im gleichen Jahr mehrere Handel in der gleichen Richtung aufgeführt. Diese wurden für die weitere Analyse zusammengefasst indem der Wert der Lieferungen addiert wurde.

Der Datensatz enthält Schleifen. Das er enthält Zeilen in denen Exporteur und Importeur eines Handels übereinstimmen. Da es hierfür keine sinnvolle inhaltliche Erklärung gibt wurden die entsprechenden Beobachtungen gelöscht. Eine Liste der gelöschten Kanten befindet sich im Anhang.

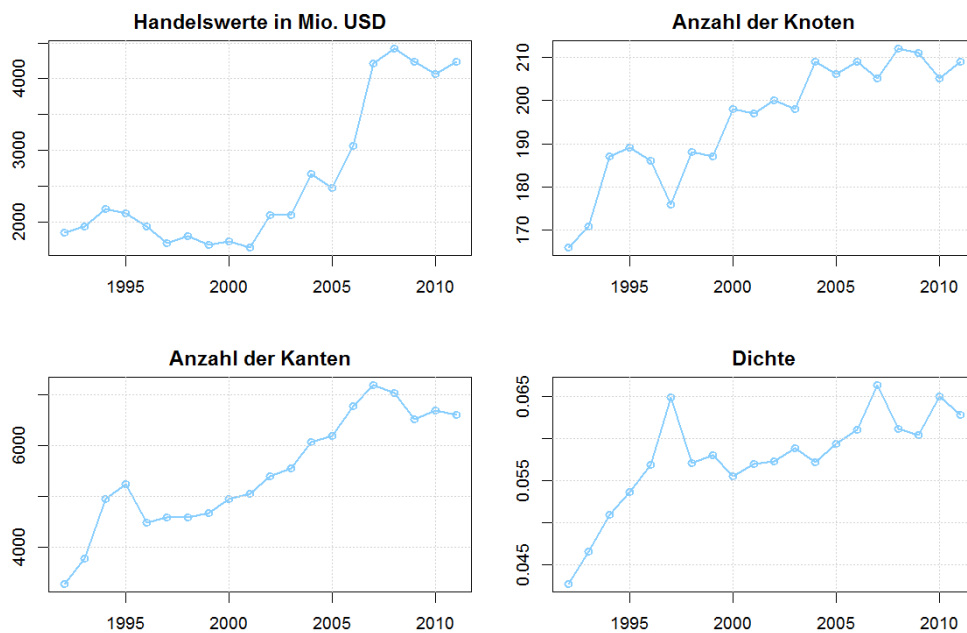
---

<sup>1</sup><https://www.prio.org/>

## 4 Deskriptive Analyse

Durch eine ausführliche deskriptive Analyse sollen Strukturen und Besonderheiten des vorliegenden Datensatzes herausgearbeitet werden. Vom besonderen Interesse ist hierbei die zeitliche Komponente. Hierzu werden deskriptive Maßzahlen, Handelswerte und die Degree- Verteilungen für die 20 im Datensatz enthaltenen Jahre verglichen, um etwaige strukturelle Veränderungen des Netzwerkes zu erkennen. Ein weiterer Teil befasst sich damit, die wichtigsten Akteure des Netzwerkes zu identifizieren. Zuletzt wird versucht das Netzwerk zu visualisieren.

### 4.1 Deskriptive Netzwerkmaßzahlen



**Abbildung 4.1:** Deskriptive Maßzahlen des Kleinwaffenhandelsnetzwerkes 1992-2011

Zunächst erfolgt eine Darstellung grundlegender deskriptiver Netzwerkmaßzahlen, die das Kleinwaffenhandelsnetzwerk beschreiben. Die Maßzahlen werden für jedes Jahr berechnet und als Zeitreihe dargestellt, um die zeitliche Entwicklung des Netzwerkes zu visualisieren. Abbildung 4.1 zeigt die zeitliche Entwicklung von Handelswert, Knotenanzahl, Kantenanzahl und Dichte des Netzwerkes.

## 4 Deskriptive Analyse

Betrachtet man zuerst die Zeitreihe der Handelswerte so stellt man fest, dass diese in den Jahren 1992 bis 2001 relativ konstant zwischen 1.5 und 2.5 Milliarden US Dollar verweilt, nach 2001 jedoch bis auf ca 4.5 Milliarden im Jahr 2008 ansteigt und anschließend auf diesem Level konstant bleibt. Die Zeitreihe der Anzahl der am Waffenhandel beteiligten Nationen steigt recht gleichmäßig zwischen Jahren 1992 und 2011. Lediglich zwischen 1993 und 1994 ist ein außergewöhnlich starker Anstieg von 170 auf 187 zu beobachten. Im Jahr 1997 ist die Anzahl der am Waffenhandel beteiligten Nationen auffallend von 186 auf 176 gesunken. Allerdings stellt sich gleich im Folgejahr wieder die ursprüngliche Anzahl ein. Das Maximum der Zeitreihe liegt mit 212 Nationen im Jahr 2008. Auch die Anzahl der Netzwerkkanten die der Anzahl der vollzogenen Waffentransaktionen zeigt einen regelmäßig steigenden Trend von ca 3000 im Jahr 1992 bis ca 7000 im Jahr 2011. Auch hier erkennen wir einen sprunghaften Anstieg zwischen 1993 und 1994 sowie ein Einknicken im Jahr 1996. Die Dichte des Netzwerkes steigt in den Jahren 1992-1997 rasch von ca 0.04 auf 0.065 und schwankt anschließend auf einem Level zwischen 0.055 und 0.065.

### 4.2 Die Degree-Verteilung

Eine erste nützliche Analyse um die Struktur des Netzwerkes zu erfassen, ist die Betrachtung der Knoten-Degrees. Im Falle eines gerichteten Netzwerkes unterscheidet man zwischen In-Degree und Out-Degree. Inhaltlich interpretiert entspricht dies den Anzahlen der Importe und Exporte eines Landes. Hierzu betrachten wir Abbildung 4.2. Sie zeigt zu jedem im Datensatz enthaltenen Jahr je einen Boxplot der In-Degrees und Out-Degrees aller Länder. Innerhalb der farblich gekennzeichneten Box liegen jeweils die mittleren 50 Prozent der entsprechenden Daten. Der mittlere schwarze Strich in jeder Box kennzeichnet den Median, und damit den Wert, unter dem genau die Hälfte aller Werte liegt.

Betrachtet man zuerst die Boxplots zum In-Degree, so stellt man fest, dass die Breite der Box über die Jahre zunimmt. Im Jahr 1992 reicht sie lediglich von zwei bis acht, während sie sich im Jahr 2011 von drei bis 21 erstreckt. Auch der Median steigt über diesen Zeitraum von drei auf zehn. Daraus lässt sich schließen, dass die mittlere Anzahl Importpartner eines einzelnen Landes über die Zeit größer geworden ist. Auffallend sind in allen Jahren einige mit Kreisen markierte Ausreißer, die aus bis zu 67 verschiedenen Ländern im gleichen Jahr Waffen beziehen.

Bei den Boxplots zum Out-Degree fallen sofort die eher kleinen Boxen auf. Übereinstimmend in allen Jahren exportieren mindestens 25 Prozent der Länder überhaupt keine Waffen, und 50 Prozent der Länder an höchstens 2 andere Staaten. Allerdings gibt auch in allen Jahren eine recht große Anzahl von Ausreißern mit hohem Out-Degree mit bis zu 150 belieferten Staaten.

Die Betrachtung der Degrees legt nahe, dass der Kleinwaffenhandel von einigen wenigen

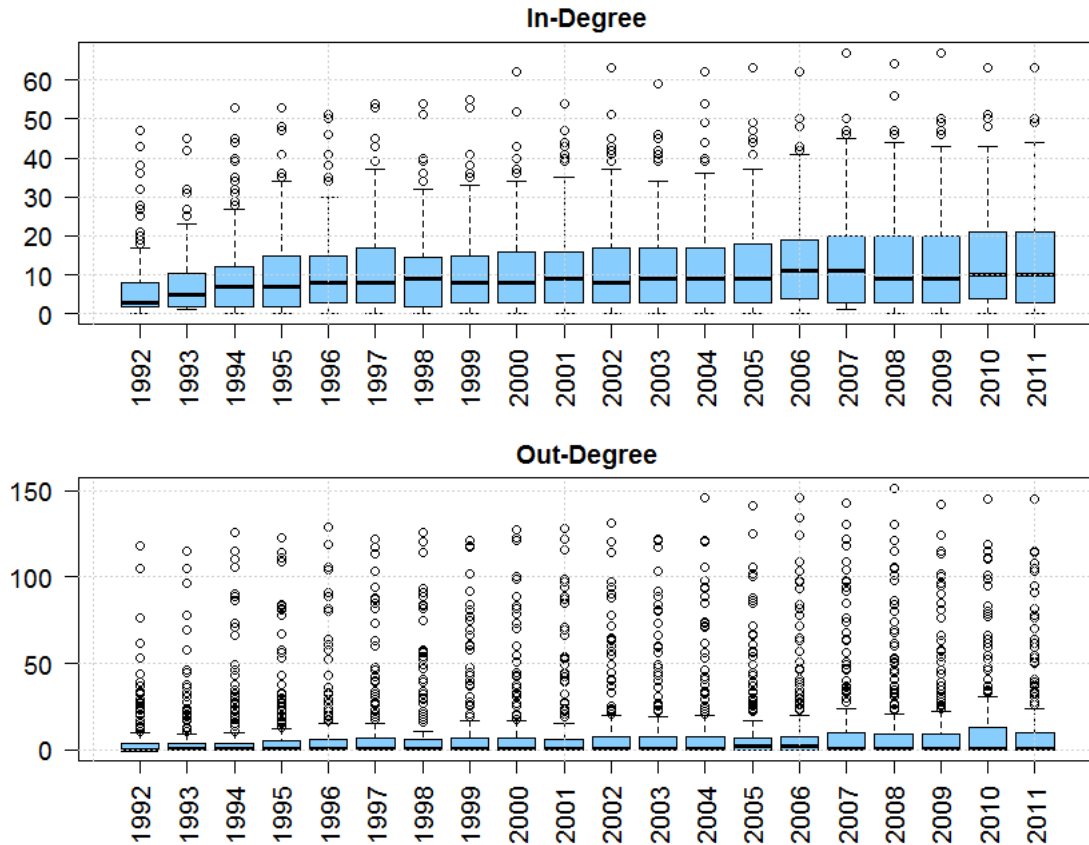


Abbildung 4.2: Boxplot für In- und Out-Degree in den Jahren 1992-2011

Akteuren dominiert wird, während die große Masse der restlichen Staaten eher einen geringen Einfluss auf die Geschehnisse hat. Der Identifikation dieser wichtigen Akteure wenden wir uns in Abschnitt 4.4 zu.

Interessante Einblicke in die Zusammenhänge innerhalb eines Netzwerkes lassen sich auch generieren, indem man den Degree eines Knoten mit dem seiner Nachbarn vergleicht. Zwei Netzwerke mit gleicher Degree-Verteilung können dennoch unterschiedliche Strukturen besitzen, falls sich die Knoten der Netzwerke darin unterscheiden mit welchen anderen Knoten sie sich bevorzugt verbinden. Hierzu betrachten wir Abbildung 4.3. Hier ist das Jahr 1998 exemplarisch ausgewählt. Dem logarithmierten Knoten Degree auf der X-Achse ist der logarithmierte durchschnittliche Degree der Nachbar auf der Y-Achse gegenübergestellt. Man erkennt, dass sich Knoten mit geringerem Degree tendenziell eher mit Knoten mit hohem Degree verbinden als Knoten, die selbst einen hohen Degree besitzen. Zentrale Akteure des Netzwerkes handeln also eher mit kleineren Akteuren des Netzwerkes als mit anderen zentralen Akteuren.

## 4 Deskriptive Analyse

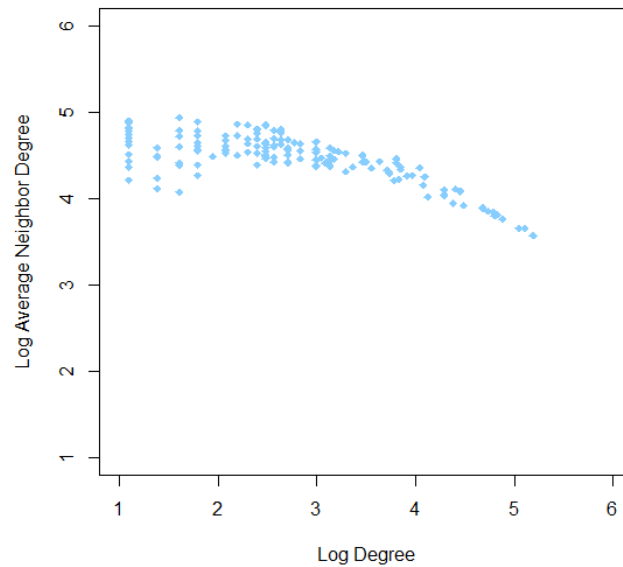


Abbildung 4.3: Average nearest Neighbor Degree am Beispiel von 1998

### 4.3 Handelswerte

Betrachtet man die Gewichte der Kanten, in diesem Fall die monetären Handelswerte, so fällt einem ein deutliches Ungleichgewicht auf. Wie Abbildung 4.4 zeigt ist die Summe der 1% teuersten Waffenhandel über alle Jahre hinweg ähnlich der Summe der 99% billigsten. Einige wenige große Waffentransaktionen wiegen also alle restlichen in ihrem monetären Gewicht auf.

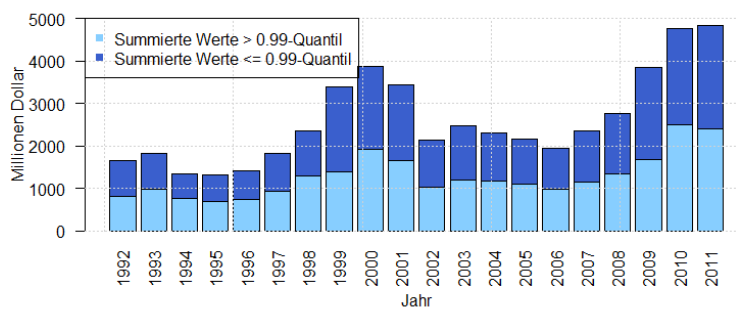


Abbildung 4.4: Vergleich der 1% teuersten Waffenkäufe mit den 99% billigsten in den Jahren 1992-2011

## 4.4 Top-Akteure

In Abschnitt 4.2 wurde anhand der Degree-Verteilungen gezeigt, dass der Kleinwaffenhandel von einigen wenigen Nationen dominiert wird, die sich deutlich von der großen Masse der restlichen Akteure abheben. Diese sollen in diesem Abschnitt ermittelt werden. Hierzu werden in Tabelle 4.1 diejenigen Nationen aufgelistet, die über den Zeitraum von 20 Jahren das größte Import- und Exportvolumen, gemessen am Geldwert der gehandelten Waffen, aufweisen.

Platz	Land	Exportvol. [Mrd.]	Platz	Land	Importvol. [Mrd.]
1	USA	9.2	1	USA	16
2	Italy	7.9	2	Germany	2.3
3	Germany	4.6	3	France	2.3
4	Brazil	3.7	4	Canada	1.9
5	Austria	2.7	5	United Kingdom	1.8
6	United Kingdom	2	6	Saudi Arabia	1.7
7	Belgium	1.8	7	Belgium	1.2
8	Switzerland	1.5	8	Spain	1.2
9	Russia	1.4	9	Australia	1.2
10	Czech Republic	1.4	10	Turkey	1

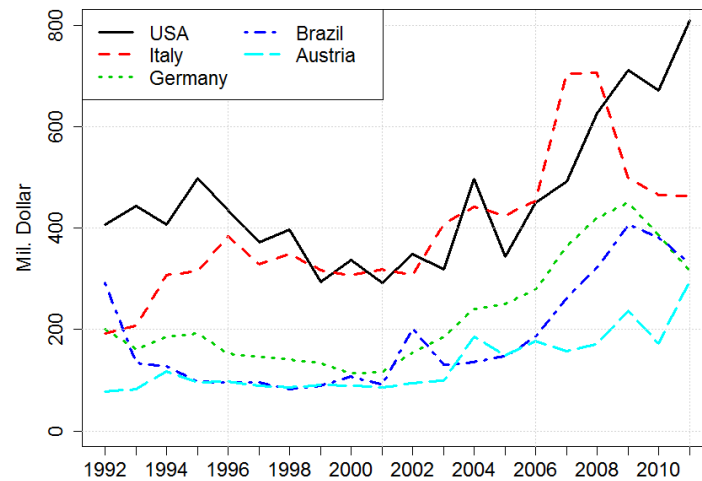
**Tabelle 4.1:** Summierte Handelswerte der Top-Importeure und Top-Exporteure des Netzwerkes von 1992 bis 2011

Es ist ersichtlich, dass die Vereinigten Staaten von Amerika mit 9.2 Milliarden Dollar am meisten Waffen exportiert. Italien steht mit 7.9 Milliarden Dollar Exportvolumen an zweiter Stelle. Deutschland exportiert mit 4.6 Milliarden Dollar gehandelten Waffen am drittmeisten. Auf dem vierten und fünften Platz folgen die Länder Brasilien und Österreich. Ab dem sechsten Platz erfolgen nur noch unwesentliche Verringerungen des Exportvolumens im Bereich von 2 bis 1 Milliarde Dollar. Hierin befinden sich Nationen wie Großbritannien, Belgien, die Schweiz, Russland und die tschechische Republik.

Auch bei den Importen steht die USA an erster Stelle. Die Nation gibt mit 16 Milliarden Dollar mehr für den Import von Kleinwaffen aus als die 9 nachfolgenden Länder zusammen. Deutschland und Frankreich teilen sich mit 2.3 Milliarden Dollar Importvolumen den zweiten Platz. Auf der vierten Stelle befindet sich Kanada mit einem Importvolumen von circa 2 Milliarden Dollar. Großbritannien verwendet 1.8 Milliarden Dollar, um Waffen zu exportieren, und Saudi Arabien 1.7 Milliarden Dollar. Auf dem siebten, achten, neunten und zehnten Platz sehen wir ähnliche Exportausgaben von circa 1.2 bis 1 Milliarde Dollar. Dies sind die Länder Belgien, Spanien, Australien und Türkei.

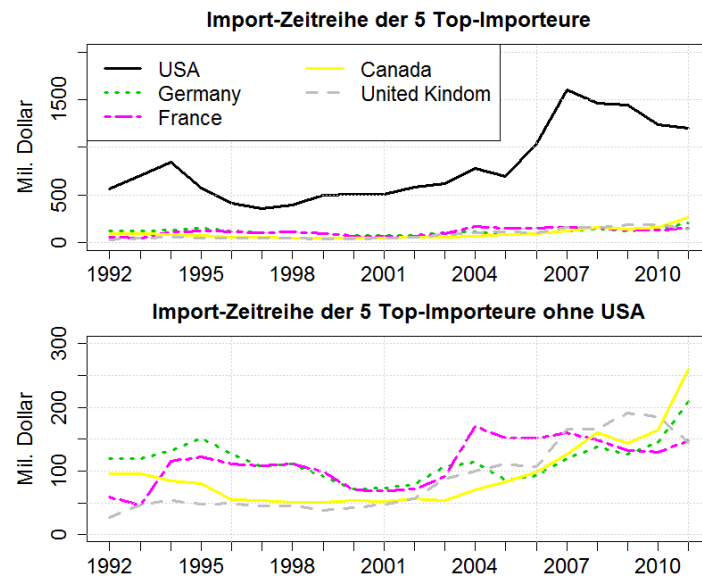
Anschließend interessiert, ob sich die Handelsvolumen der Top-Exporteure und Importeure über die Jahre verändert. Hierfür betrachten wir für die Exporte die Abbildung 4.5. In dieser Grafik sind die Handelsvolumen in Millionen US Dollar der fünf Top-Exporteure über den Zeitraum 1992-2011 dargestellt. Die USA ist in fast allen Jahren der Waffenexporteur mit den höchsten monetären Volumen. Sie wird nur in wenigen Jahren von Italien übertroffen. Deutschland, Brasilien und Österreich exportieren in allen Jahren

#### 4 Deskriptive Analyse



**Abbildung 4.5:** Zeitreihen der jährlichen Handelswerte der Top-Exporteure von 1992 bis 2011

deutlich weniger Waffenwert als die USA. Auffällig ist ein relativ konstanter Verlauf der Zeitreihen im Zeitraum von 1992 bis ca 2001, während danach bei allen Ländern ein kräftiger Anstieg der Handelswerte feststellbar ist. Deutschland, Brasilien und vor allem Italien zeigen allerdings ab circa 2008 wiederum einen abfallenden Trend.



**Abbildung 4.6:** Zeitreihen der jährlichen Handelswerte der Top-Importeure von 1992 bis 2011

Die beiden Grafiken aus Abbildung 4.6 zeigt die Handelsvolumen der Top-Importeure. Die USA ist hier unangefochten an der Spitze. Sie importiert Kleinwaffen im Wert zwischen circa 400 und 1600 Millionen US Dollar pro Jahr während die restlichen Akteure höchstens Kleinwaffen im Wert von circa 250 Millionen Dollar pro Jahr importieren. Ähnlich wie bei den Exportzeitreihen ist auch hier ein relativ konstanter Verlauf bis circa 2001 zu beobachten während die Ausgaben in den nachfolgenden Jahren kontinuierlich ansteigen.

#### 4 Deskriptive Analyse

Die USA verringerte ihre Importausgaben ab dem Jahr 2007 jedoch wieder deutlich.

Bei den Top Akteuren des Waffenhandels handelt es sich also größtenteils um westliche Industrienationen. Diese bestimmen durch Wirtschaftskraft und großen Handelsvolumen sowohl den Export als auch den Import von Klein- und Leichtwaffen. Interessanter Weise bekommt man ein anderes Bild, wenn man die Handelswerte mit einem Maß der Wirtschaftskraft der Länder, hier dem Bruttoinlandsprodukt pro Kopf, skaliert. In Tabelle 4.2 sind diejenigen Länder aufgeführt, die relativ zu ihrem Bruttoinlandsprodukt pro Kopf im Zeitraum von 1992 - 2011 am meisten Geld für den Import beziehungsweise den Export von kleinen und leichten Waffen ausgegeben haben.

Platz	Land	Exportvol. / BIP pro Kopf	Platz	Land	Importvol. / BIP pro Kopf
1	China	114735	1	Tanzania	54562
2	Brazil	53225	2	Thailand	49636
3	Italy	48862	3	India	32416
4	Spain	40822	4	Pakistan	30290
5	Germany	38039	5	South Korea	27208
6	Turkey	36174	6	China	25402
7	South Korea	29131	7	Indonesia	24268
8	United States	26539	8	Kenya	22907
9	India	24615	9	Malaysia	22330
10	Austria	23149	10	Bukina Faso	22183

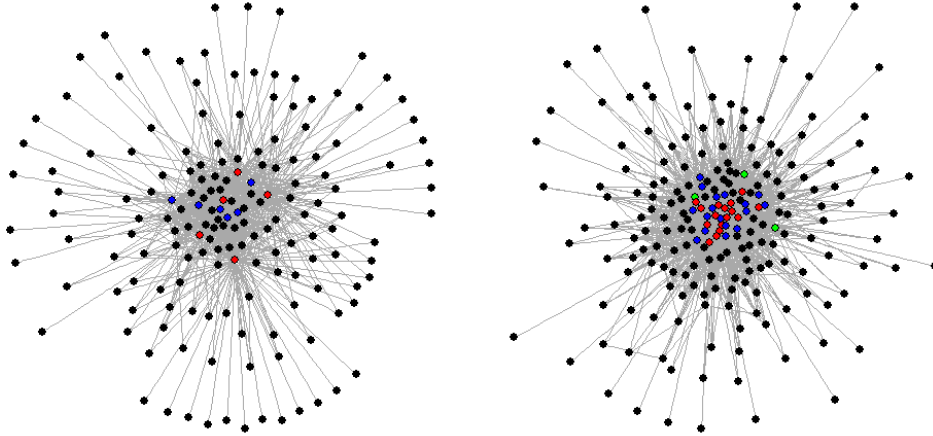
**Tabelle 4.2:** Summierte Handelswerte der Top-Exporteure und Top-Importeure relativ zum BIP pro Kopf des Netzwerkes von 1992 bis 2011

Bei den Exporten sind immer noch westliche Industrienationen unter den zehn führenden Ländern am häufigsten vertreten. Auffallend ist, dass die zuvor deutlich dominierenden Vereinigten Staaten von Amerika nur noch auf dem achten Platz liegen. Gemessen an der eigenen Wirtschaftskraft exportiert China nun am meisten Waffen. Neu unter den zehn Führenden sind mit Indien und Süd Korea auch zwei wirtschaftsstarke asiatische Länder. Im Gegensatz hierzu steht die Tabelle mit den zehn stärksten Importeuren von kleinen und leichten Waffen im Vergleich zu ihrer Wirtschaftskraft. Hier nehmen nun vor allem asiatische und afrikanische Entwicklungsländer die vorderen Plätze ein. Alle westlichen Industrienationen hingegen tauchen in den zehn vorderen Plätzen nicht mehr auf. Führend sind hier die Länder Tanzania, Thailand, Indien, Pakistan und Süd Korea.

Eine andere Methode, um zentrale Akteure des Netzwerkes zu identifizieren ist sich die Degree-Sequenz der Netzwerkknoten zu betrachten. Welche Knoten (Länder) besitzen sowohl einen hohen In-Degree als auch einen hohen Out-Degree und können demnach als zentrale Akteure des Handelsnetzwerkes identifiziert werden?

Abbildung 4.7 zeigt hierzu das gesamte Netzwerk in den Jahren 1992 und 2011. Jeder Punkt stellt ein am Waffenhandel beteiligtes Land dar. Ein Pfeil zwischen den beiden Ländern symbolisiert einen Handel. Länder mit einem In-Degree von größer als 30, die also aus mindestens 30 anderen Ländern Waffen beziehen sind grün. Länder mit einem Out-Degree von größer als 30, die also in mindestens 30 andere Länder Waffen liefern sind blau. Akteure die beide Bedingungen erfüllen sind rot eingefärbt. Übereinstimmend





**Abbildung 4.7:** Visualisierung des Netzwerkes 1992 (l.) und 2011(r)

mit unseren einleitenden deskriptiven Analysen scheint das Netzwerk aus dem Jahr 2011 auf den ersten Blick größer und dichter zu sein als das Netzwerk im Jahr 1992. Auch die Anzahl der farblich markierten Knoten ist im Jahr 2011 deutlich größer als im Jahr 1992. Man erkennt also, dass die Anzahl der „großen Akteure“ auf dem Kleinwaffenmarkt gemessen an der Anzahl ihrer Importe und Exporte über die Zeit deutlich zugenommen hat. In Tabelle 4.3 werden die rot markierten Knoten für beide Jahre aufgelistet.

## 4 Deskriptive Analyse

1992			2011		
Land	In-Degree	Out-Degree	Land	In-Degree	Out-Degree
Switzerland	32	76	Switzerland	41	103
USA	43	105	USA	63	145
Germany	47	118	Finland	34	70
Spain	36	62	Italy	39	114
Sweden	38	33	France	41	82
			Poland	32	34
			Czech Rep.	37	105
			Germany	50	115
			UK	44	95
			Norway	32	39
			Spain	37	91
			Canada	49	78
			Austria	41	108
			South Africa	34	32
			Belgium	31	64
			Australia	39	51

**Tabelle 4.3:** Zentrale Akteure des Netzwerkes 1992 und 2011

### 4.5 Visualisierungen

In diesem Abschnitt wird versucht durch verschiedene Visualisierungen des Netzwerkes einen Überblick über mögliche Strukturen und Zusammenhänge des Kleinwaffenhandels zu erhalten. Da das Netzwerk recht groß ist erzielt eine einfache Abbildung aller Knoten und Kanten, wie in Abbildung 4.7 geschehen ein nicht zufriedenstellendes Ergebnis. Ein möglicher Ansatz um die Übersichtlichkeit der Visualisierung des Netzwerkes zu verbessern ist es daher, die Knoten in Gruppen einzuteilen. Eine naheliegende Kategorisierung von Ländern sind die Kontinente auf denen sie sich befinden. Die Einteilung geschieht mit Hilfe des R-Pakets *countrycode* (Arel-Bundock; 2014). Mit Hilfe der im Datensatz gegebenen Correlates of War Country Codes und der kontinentalen Zuordnungen der Vereinten Nationen (UNO) weist dieses Software Paket jedem Land einen Kontinent zu. 34 der im Datensatz vorkommenden Länder sind sogenannte Sonderverwaltungszone und werden daher keinem Kontinent zugeordnet. Beispiele hierfür sind Französisch Guyana, Hong Kong und die Falkland Inseln. Eine vollständige Liste aller Länder und ihrer Zuweisung eines Kontinentes findet sich im elektronischem Anhang dieser Arbeit. Aus dieser Kategorisierung der Knoten resultiert die Grafik ??.

In der Grafiken ist die Größe der Knoten jeweils proportional zum Degree (In-Degree + Out-Degree) also dem Anzahl der Handelsaktivitäten eines Knotens gewählt. Die Breite der Kanten wiederum ist proportional zum monetären Wert der Handelsströme zwischen zwei Kontinenten gewählt. Die Positionierung der Knoten wurde zur besseren Vergleichbarkeit der Jahre untereinander fixiert. In Abbildung 4.8 erkennt man, dass Europa durchgängig mit dem größten Kreis markiert ist, also an mehr Handelsaktionen beteiligt ist, als die anderen Kontinente. Dies stimmt mit den vorhergehenden deskriptiven Analysen überein. Die Dicke der Kanten und damit der zwischen den Kontinenten gehandelte monetäre Wert,

#### *4 Deskriptive Analyse*

**Abbildung 4.8:** Handelsströme zwischen den Kontinenten von 1992-2011

variiert stark zwischen den Jahren. Hier ist kein gleichbleibendes Muster zu erkennen.

## 5 Exponential Random Graph Model

Wie bereits in der Einleitung erläutert stellt die Abhängigkeitsstruktur der Beobachtungen in einem Netzwerk besondere Ansprüche an die statistische Modellierung eines Netzwerkes. Aus der deskriptiven Analyse unseres Datensatzes geht außerdem hervor, dass die Kantenbildung zwischen zwei Knoten in unserem Netzwerk nicht nur von der Struktur des restlichen Netzwerkes, sondern auch von exogenen Kovariablen wie zum Beispiel der Wirtschaftskraft und der geografischen Lage eines Landes abhängt.

Üblicherweise verwendet man zur Modellierung von Netzwerken das *Exponential Random Graph Model (ERGM)*, denn

„The ERGM is a statistical model that can be used to estimate the effects of covariates on the ties in a network while simultaneously estimating parameters that provide a precise and parsimonious description of the forms of dependence that can exist in relational data.“ (Cranmer and Desmarais; 2011)

Das ERGM kann also sowohl den Einfluss exogener Kovariablen, als auch den strukturellen endogenen Einfluss eines Netzwerkes auf sich selbst modellieren. In den folgenden Abschnitten wird zunächst das Modell definiert um es danach auf den Datensatz zum Handel mit kleinen und leichten Waffen anzuwenden. Die Definition des Modells in Abschnitt 5.1 folgt im wesentlichen dem Artikel von Hunter et al. (2008). Die darauf folgenden Abschnitte 5.1.1 und 5.1.2 wurden aus den Kapiteln 12.2 und 12.3 des Buches von Lusher et al. (2012) erarbeitet.

### 5.1 Definition des Modells

Vorab einige benötigte Notation:

Sei  $X$  eine matrixwertige Zufallsvariable und repräsentiere die Nachbarschaftsmatrix eines Netzwerkes.  $\mathcal{X}$  soll die Menge aller möglichen Nachbarschaftsmatrizen von binären Netzwerken mit fester Knotenanzahl  $n$  darstellen, die keine Schleifen oder multiple Kanten enthalten. Daher ist  $\mathcal{X}$  die Menge aller  $n \times n$  Matrizen deren Einträge entweder Eins oder Null sind und die auf ihrer Diagonalen nur Nuller besitzen. Das  $ij$ -te Element von  $X$  repräsentiert eine Kante zwischen Knoten  $i$  und  $j$ . Bei einem ungerichteten Netzwerk ist  $X$  eine symmetrische Matrix. Bei einem gerichteten Netzwerk kann sich hingegen der Eintrag an der Stelle  $ij$  von dem an der Stelle  $ji$  unterscheiden.

## 5 Exponential Random Graph Model

Die Wahrscheinlichkeitsverteilung der matrixwertigen Zufallsvariable  $X$  lässt sich nun in folgender Form ausdrücken:

$$P_{\theta, \mathcal{X}}(X = x) = \frac{\exp\{\theta^T g(x)\}}{\kappa(\theta, \mathcal{X}, \mathcal{Y})} \quad (5.1)$$

mit

- $x \in \mathcal{X}$
- $\theta \in \Omega \subset \mathbb{R}^q$  ... Vektor der Modellparameter
- $g(x)$  ...  $q$ -Vektor aus Statistiken basierend auf der Nachbarschaftsmatrix  $X$

Ersetzt man in Gleichung 5.1  $g(x)$  durch  $g(x, Y)$  so ermöglicht man die Aufnahme von zusätzlichen exogenen Informationen  $Y$  über das Netzwerk (siehe Kapitel 5.4).  $\kappa(\theta, \mathcal{X}) = \sum_{x \in \mathcal{X}} \exp\{\theta^T g(x)\}$  ist ein Normalisierungsfaktor, der sicherstellt, dass sich bei der Gleichung 5.1 eine Wahrscheinlichkeitsverteilung mit Wertebereich  $[0, 1]$  ergibt. Die genaue Berechnung von  $\kappa(\theta, \mathcal{X})$  ist problematisch, da die Summe über alle möglichen Ausprägungen der Zufallsvariable  $\mathcal{X}$  summiert. Da aber bereits bei  $n$  Knoten  $2^{n(n-1)}$  verschiedene Netzwerke denkbar sind, erhält man schnell unvorstellbar große Anzahlen von möglichen Netzwerken die den Umgang mit  $\kappa(\theta, \mathcal{X})$  zum entscheidenden Problem dieses Modells werden lassen. Eine analytische Lösung des Modells ist hierdurch oft unmöglich. Es werden Methoden benötigt, um durch simulationgestützte Verfahren geeignete Schätzer entwickeln zu können. Hiermit beschäftigen sich die nächsten beiden Kapitel.

### 5.1.1 Simulation von Zufallsgraphen

Die Simulation von Zufallsgraphen aus einer Zielverteilung  $P_\theta(x)$  basiert auf dem Makro Chain Monte Carlo (MCMC) Algorithmus. Hierbei beginnt man mit einem beliebigen Netzwerk mit fester Knotenzahl  $n$ . Nun sucht man sich ein zufälliges Knotenpaar und fügt eine Verbindung hinzu, beziehungsweise entfernt eine bestehende Verbindung. Besitzt der so erhaltene Graph nun eine höhere Wahrscheinlichkeit als der vorherige wird er akzeptiert. Besitzt der Graph eine geringere Wahrscheinlichkeit, so wird er mit einer Wahrscheinlichkeit akzeptiert, die proportional zum Verhältnis aus den Wahrscheinlichkeiten von altem und neuem Graph ist. Nach ausreichend großer Anzahl von Iterationen erhält man ein zufälliges Netzwerk, das als Ziehung aus der Zielverteilung  $P_\theta(x)$  gezogen wurde. Alle vorherigen Ziehungen nennt man *Burn In*. Der Burn In wird benötigt, damit der Algorithmus den beliebig gewählten Anfangszustand „vergessen“ kann. Bei ausreichend großem Burn In ist das gezogene Netzwerk unabhängig von Startpunkt und jedes der nach ihm gezogenen Netzwerke entspringt ebenfalls der Zielverteilung. Genauer gesagt benutzen wir einen Metropolis Algorithmus. Wir erstellen eine Sequenz von Gra-

## 5 Exponential Random Graph Model

phen  $X^{(0)}, X^{(1)}, \dots, X^{(M-1)}, X^{(M)}$ . Im  $m$ -ten Iterationsschritt wird hierbei folgendermaßen vorgegangen:

1. Aus dem aktuellen Graphen  $x^{(m-1)}$  wird ein zufälliges Knotenpaar  $i, j$  ( $i, j \in 1, \dots, N$ ) ausgewählt.
2. Der vorgeschlagene Graph  $x^* = x^{(m-1)}$  bis auf  $x_{ij}^{(m-1)} = 1 - x_{ij}^{(m-1)}$ .
3. Der vorgeschlagene Graph wird mit der Wahrscheinlichkeit  $\min\{1, \frac{P_\theta(x^*)}{P_\theta(x^{m-1})}\}$  akzeptiert.
4. Bei Akzeptanz entspricht der neue Graph  $x^m = x^*$  und sonst  $x^m = x^{m-1}$ .

Hierbei ist für das Verhältnis von  $\frac{P_\theta(x^*)}{P_\theta(x^{m-1})}$  die Berechnung der Change-Statistiken ausreichend, denn

$$\begin{aligned} \log\left\{\frac{P_\theta(x^*)}{P_\theta(x^{m-1})}\right\} &= \log\{P(X_{ij} = 1 - x_{ij}^{m-1} | X_{-ij} = x_{ij}^{m-1})\} \\ &= \theta_1(z_1(x^*) - z_1(x^{m-1})) + \theta_2(z_2(x^*) - z_2(x^{m-1})) \\ &\quad + \dots + \theta_p(z_p(x^*) - z_p(x^{m-1})) \end{aligned}$$

.

Diese Methode kann computationally optimiert werden, indem man zum Beispiel größere Update-Schritte oder Update-Schritte mit ungleichen Gewichten zulässt.

Will man mehrere Graphen aus der selben Verteilung ziehen, so kann die gleiche Kette verwendet werden. Allerdings sind zwei direkt nacheinander beobachtete Graphen einer Kette natürlich stark korreliert. Daher lässt man zwischen den beobachteten Graphen jeweils  $k$  Iterationen aus. Den Wert von  $k$  nennt man *thinning*. Die Autokorrelation zwischen den gezogenen Graphen sollte durch ein möglichst großes  $k$  gering gehalten werden.

### 5.1.2 Schätzung der Modell-Parameter

Ziel der Schätzung der Parameter ist es die Verteilung der Statistiken der Simulierten Netzwerke über denen des beobachteten Netzwerkes zu zentrieren, so dass

$$E_\theta(z(X)) - z(x_{obs}) = 0 \tag{5.2}$$

Das Auflösen von Gleichung 5.2 liefert die Parameterwerte  $\theta$  die die beobachteten Daten  $x_{obs}$  am besten beschreiben. Analog hierzu funktioniert die Maximum Likelihood Theorie. Hier wird derjenige Parametervektor  $\theta$  gesucht, der die Wahrscheinlichkeit  $P_\theta(x_{obs})$  optimiert. Dieses Vorgehen liefert den gleichen  $\theta$ -Vektor wie die Lösung von Gleichung 5.2,

## 5 Exponential Random Graph Model

denn die partielle Ableitung nach  $\theta$  liefert

$$\begin{aligned}\frac{\partial}{\partial \theta} \log(P_\theta(x_{obs})) &= z(x_{obs}) - \frac{\partial}{\partial \theta} \log\left\{ \sum_{x \in X} \exp(\theta_1 z_1(x) + \dots + \theta_p z_p(x)) \right\} \\ &= z(x_{obs}) - \sum_{x \in X} z(x) P_\theta(x)\end{aligned}$$

was Gleichung 5.2 entspricht.

Eine analytische Lösung dieser Gleichung ist auf Grund der großen Anzahl an möglichen Netzwerken nicht möglich. Im Prinzip funktioniert die Lösung der Gleichung durch einfaches Ausprobieren. Das bedeutet, man wählt einen beliebigen Parametervektor, simuliert mit diesem auf oben erklärte Art und Weise eine große Anzahl an Graphen und checkt im Anschluss ob Gleichung 5.2 erfüllt ist. Dieses Vorgehen ist aber natürlich computational ineffizient. Eine Möglichkeit zur effizienteren Implementierung dieses Vorgehens bietet das *Importance Sampling* nach Geyer-Thompson was im **statnet**-Paket (Handcock et al.; 2014) in R (R Development Core Team; 2008) standardmäßig verwendet wird. Der Geyer-Thompson Algorithmus zieht eine große Stichprobe von Graphen, für einen vorläufigen Parametervektor  $\tilde{\theta}$ . Diese Stichprobe wird nun als repräsentativ für alle Graphen angesehen. Die Stichprobe wird nun immer weiter verwendet, selbst wenn sich der Parametervektor im Laufe des Algorithmus verändert. Um zu berücksichtigen, dass es sich nur um eine Stichprobe und keine vollständige Sammlung aller möglichen Graphen handelt, muss ein gewichteter Durchschnitt der Statistiken verwendet werden um  $\bar{z}_\theta$  zu berechnen. Genauer gesagt, falls die Stichprobe aus der Verteilung  $P_\theta(x)$  generiert wurde ist der Stichprobendurchschnitt  $\bar{f}_\theta = w^1 f(x^1) + w^2 f(x^2) + \dots + w^M f(x^M)$  der Funktion  $f$  mit den Gewichten  $w^{(m)} = \frac{e^{(\theta_1 - \tilde{\theta}_1)z_1(x^{(m)}) + \dots + (\theta_p - \tilde{\theta}_p)z_p(x^{(m)})}}{\sum_{k=1}^M e^{(\theta_1 - \tilde{\theta}_1)z_1(x^{(k)}) + \dots + (\theta_p - \tilde{\theta}_p)z_p(x^{(k)})}}$  eine gute Approximation für den echten Erwartungswert  $E_\theta(f(X))$  wenn  $M$  groß wird und  $\tilde{\theta}$  nahe am wahren  $\theta$  liegt. Umso näher  $\tilde{\theta}$  und  $\theta$  zusammen liegen, desto näher kommen die Gewichte dem Wert  $\frac{1}{M}$ . Falls  $\theta$  und  $\tilde{\theta}$  jedoch weiter auseinander liegen besitzen die Gewichte eine hohe Streuung und somit die Schätzung eine große Standardabweichung. Um die Likelihood nun zu lösen, erzeugt man eine Sequenz von Parametern  $\tilde{\theta}, \theta^1, \theta^2, \dots, \theta^G$  mit Hilfe eines Verfahrens wie *Newton-Raphson* oder *Fisher Scoring*. Eine Aktualisierung der Sequenz erfolgt durch  $\theta^{(g)} = \theta^{(g-1)} - D(\theta^{(g-1)})^{-1} \left\{ \sum_{m=1}^M w^{(m)} z(x^{(m)}) - z(x_{obs}) \right\}$ . Da  $\sum_{m=1}^M w^{(m)} z(x^{(m)})$  eine Approximation von  $E_\theta(z(X))$  ist, falls  $\theta^{(g-1)}$  der wahre Parameter ist, ergibt sich dann  $E_\theta(z(X)) - z(x_{obs}) = 0$  und  $\theta^{(g)}$  bleibt unverändert zu  $\theta^{(g-1)}$ . Die Skalierungsmatrix  $D(\theta)$  skaliert die Unterschiede zwischen den beobachteten und simulierten erwarteten Werten der Statistiken, da die Statistiken sich in ihrer Sensitivität gegenüber Parameteränderungen unterscheiden können und die Parameterwerte nicht nur ihre eigenen sondern auch fremde Statistiken beeinflussen können.  $D$  ist die gewichtete Stichproben Kovarianzmatrix  $\Sigma_m w^{(m)} z(x^{(m)}) z(x^{(m)})^T - \left[ \sum_m w^{(m)} z(x^{(m)}) \right] \left[ \sum_m w^{(m)} z(x^{(m)}) \right]^T$ . Typischer Weise startet man diesen Algorithmus einige Male und setzt den vorgeschlagenen Parameter  $\tilde{\theta}$  gleich

## 5 Exponential Random Graph Model

$\theta^{(G)}$ . Für diesen Algorithmus ist es sehr wichtig, dass man den Startpunkt  $\tilde{\theta}$  nicht zu weit entfernt vom richtigen ML-Schätzer wählt, was in der Praxis schwierig sein kann. Zusammengefasst lässt sich das Prinzip des Importance-Sampling schematisch folgendermaßen darstellen:

1. Ziehung einer großen Stichprobe von Graphen auf Basis eines vorläufigen Parametervektors  $\tilde{\theta}$ .
2. Benutzung gewichteter Stichprobendurchschnitte der Statistiken.
3. Erzeugen einer Sequenz von Parametern  $\tilde{\theta}, \theta^{(1)}, \theta^{(2)}, \dots, \theta^{(G)}$  durch *Fisher Scoring*.
4. Neustart mit  $\theta^{(G)}$  als  $\tilde{\theta}$ .

### 5.2 Endogene Statistiken

Netzwerkdaten sind in der Regel relationale Daten. Das bedeutet sie beschränken sich nicht lediglich auf Variablen die einzelne Akteure betreffen, sondern zeichnen sich viel mehr dadurch aus, dass sie Beziehungen zwischen ihnen abbilden. Die Knoten(Akteure) eines Netzwerkes können daher genauso wenig wie die sie verbindenden Kanten als unabhängig voneinander betrachtet werden. Sie hängen explizit vom Zustand des restlichen Netzwerkes ab. Diese Eigenschaft erschwert die richtige Spezifizierung und Schätzung eines Modells im Vergleich zur gewöhnlichen statistischen Inferenz (Handcock et al.; 2008). Die Zielvariable in einem Exponential Random Graph Model ist die Existenz beziehungsweise Abwesenheit einer Kante. Als erklärende Kovariablen werden hierbei Statistiken verwendet, die aus Funktionen eben genau dieser Kanten hervorgehen. Man kann das Modell also in gewisser Weise als autoregressives Modell betrachten (Morris et al.; 2008). Welche Statistiken aufgenommen werden sollten um ein gutes Modell zu erhalten lässt sich nicht allgemeingültig beantworten. Abhängig von Struktur und Art des Netzwerkes sollten aus den im Folgenden vorgestellten Gruppen die jeweils passenden gewählt werden. Für eine (vollständige) Auflistung der im **ergm**-Paket (Handcock et al.; 2015) implementierten Statistiken verweise ich auf (Morris et al.; 2008) und die entsprechende Hilfe Seite in R.

Die Basis eines jeden Modells sollte eine Statistik bilden, die die Grundwahrscheinlichkeit der Bildung einer Kante im Netzwerk wiedergibt. Üblicherweise verwendet man hierfür die Anzahl an Kanten, die im **ergm**-Paket durch die Statistiken *edge* und *density* implementiert sind und im Falle eines gerichteten Netzwerkes äquivalent sind. Zusätzlich können die Statistiken *mutual* und *asymmetric* verwendet werden, die die Anzahl an Knotenpaaren  $i, j$  zählen für die sowohl die Kante  $(i, j)$  als auch die Kante  $(j, i)$  besteht, beziehungsweise eben nur genau eine von beiden.

Die Degree-Verteilung wie in Abschnitt 4.2 beschrieben, lässt sich durch den Term  $degree(d)$  im Modell berücksichtigen. Diese Statistik zählt alle Knoten im Netzwerk die



genau den Degree  $d$  aufweisen. In einem gerichteten Netzwerk verwendet man  $idegree(d)$  und  $odegree(d)$  analog für In- und Out-Degree. Eine ähnliche Statistik ist  $kstar(k)$ . Sie zählt die Anzahl an  $k$ -stars im Netzwerk. Ein  $k$ -star ist definiert als ein Knoten der zu  $k$  anderen Knoten eine Kante besitzt. Anders als bei  $degree$  kann hier ein Knoten mehr als einmal gezählt werden. Bei gerichteten Netzwerken verwendet man alternativ  $istar(k)$  und  $ostar(k)$ .

Gruppenbildungen und Transitivitätseigenschaften lassen sich bei gerichteten Netzwerken durch Statistiken wie *ctriple*, *ttriple* oder *transitive* erfassen. Diese führen aber bekannter Weise zur Degenerierung des modellierten Netzwerkes (siehe zum Beispiel (Morris et al.; 2008; Hunter et al.; 2008; Handcock et al.; 2008)) und werden deshalb gerne durch alternative neuere Statistiken ersetzt die im folgenden Abschnitt vorgestellt werden.

### 5.3 Degeneration und alternative Modellierung

In diesem Absatz gebe ich die Erklärungen zur Degeneration aus (Handcock et al.; 2008) wieder, falls nicht anderst ausgewiesen.

Ein häufig auftretendes Problem des Exponential Random Graph Modells ist, dass das simulierte Modell degeneriert. Man spricht von einem degenerierten Modell, falls es einen Großteil seiner Wahrscheinlichkeitsmasse auf wenige extreme und unrealistische Netzwerkkonfigurationen wie zum Beispiel ein volles oder komplett leeres Netzwerk legt. Für eine genauere mathematische Erklärung verweise ich auf (Handcock et al.; 2003). Ein degeneriertes Modell führt oft zur Divergenz des Schätzalgorithmus, wodurch der ML-Schätzer der Parameter im Modell nicht bestimmt werden kann. Teilweise kann mit einem degenerierten Modell zwar eine Likelihood-Schätzung durchgeführt werden, allerdings ist ihr Fit in der Regel schlecht .

„Degeneracy is an implication of network mis-spezifikation - not a short coming of the MCMC estimation procedure.“(Handcock et al.; 2008)

Der Grund, dass einfache Zählstatistiken wie *ttriple* oder *ctriple* zur Degeneration eines Netzwerkes führen ist, dass die Hinzunahme einer einzelnen Kante ihren Wert so stark verändern kann, dass der Einfluss des Parameterwerts auf das Netzwerk unrealistisch verstärkt wird. Das wird an einem einfachen Beispiel deutlich. Ein Knoten mit Degree vier besitzt sechs Two-Stars. Verbindet man ihn jedoch mit nur einem einzigen weiteren Knoten, so kommen bereits auf einem Schlag 4 weitere Two-Stars hinzu. Hierfür gibt es in der neueren Netzwerkforschung eine empfohlene Lösungsstrategie: die Aufnahme von nicht linearen Einflüssen durch sogenannte *Curved Exponential Family Models(CEFM)* (Hunter; 2007).

Auch die Beschränkung auf lineare Effekte der Statistiken scheint intuitiv unrealistisch zu sein. Man denke sich ein Freundesnetzwerk. Zwei Akteure, mit null oder einem guten

## 5 Exponential Random Graph Model

Freund werden sich deutlicher voneinander unterscheiden, als zwei Akteure mit 20 oder 21 guten Freunden. Der Einfluss der Veränderung der Degree-Statistik um eins sollte also mit der Höhe der Statistik abnehmen. Dies gewährleistet zum Beispiel die *Geometrically Weighted Degree* (GWD) Statistik:

$$u(x, \phi_s) = e^{\phi_s} \sum_{i=1}^{n-1} \left\{ 1 - (1 - e^{-\phi_s})^i \right\} D_i(x) \quad (5.3)$$

$u(x, \phi_s)$  ist eine ist hierbei eine Funktion der Degreestatistiken  $D_i(x)$  abhängig vom Parameter  $\phi_s$ .

Ähnliche Statistiken, die Stern und Dreiecks-Statistiken ersetzen können sind die *Geometrically Weighted Edgewise Shared Partners* (GWESP) und *Geometrically Weighted Dyadic Shared Partners* (GWDSP) Statistiken  $v, w$ :

$$v(x, \phi_t) = e^{\phi_t} \sum_{i=1}^{n-2} \left\{ 1 - (1 - e^{-\phi_t})^i \right\} EP_i(x) \quad (5.4)$$

$$w(x, \phi_p) = e^{\phi_p} \sum_{i=1}^{n-2} \left\{ 1 - (1 - e^{-\phi_p})^i \right\} DP_i(x) \quad (5.5)$$

$EP_i(x)$  zählt die Anzahl der Kanten in  $x$  die genau  $i$  gemeinsame Nachbarn haben.  $EP_i(x)$  zählt die Anzahl der Knotenpaare  $(k, l)$  in  $x$  die genau  $i$  gemeinsame Nachbarn haben unabhängig von  $X_{kl}$ .

Die Parameter  $\phi_s, \phi_t$  und  $\phi_p$  sind sogenannte Decay-Parameter und bestimmen den Grad, in dem der Effekt der Statistiken für größere Werte abnimmt. Umso größer  $\phi$ , desto langsamer ist die Abnahme des Effekts.

### 5.4 Exogene Kovariablen

Ein weiterer möglicher Grund für die Degeneration eines Modells ist das Fehlen von exogenen Unterscheidungsmerkmalen. In einem ERGM das nur mit endogenen Statistiken arbeitet werden zunächst alle Knoten als gleich angenommen und dadurch bekommen alle strukturgleichen Netzwerke die gleiche Wahrscheinlichkeit zugewiesen. Eigenschaften einzelner Knoten oder Kanten werden hier also nicht weiter berücksichtigt. Im deskriptiven Teil dieser Arbeit wurde aber gezeigt, dass die Teilnahme am Waffenhandelsnetzwerk durchaus von Eigenschaften wie der Wirtschaftskraft oder der geographischen Lage eines Landes abhängig ist. Um also ein gutes Modell anzupassen werden exogene Kovariablen benötigt, die Informationen über diese Eigenschaften liefern. Bei der Analyse des SALW Datensatzes standen folgenden zusätzlichen exogenen Kovariablen zur Verfügung:

### 5.4.1 Polity Score

Der nächste verwendete Datensatz entspringt dem *Polity IV Project* welches vom *Center for Systemic Peace* (CSP) (Marshall; n.d.a) betrieben wird. Dieser Datensatz weißt jedem Land abhängig von seinem Demokratiestatus einen Wert zwischen  $-10$  und  $10$  zu.  $-10$  kennzeichnet hierbei den schlechtest möglichen Wert, wobei  $10$  für bestmögliche demokratische Standards steht. Verwendet wird eine gewichtete Nachbarschaftsmatrix deren  $ij$ -ter Eintrag dem Unterschied im Demokratiescore zwischen Land  $i$  und Land  $j$  entspricht.

### 5.4.2 Composite Index of National Capability

Der *Composite Index of National Capability* aus der neusten Version des *National Material Capabilities Dataset*(Version 4.1) (Singer; 1988) ist ein Maß der nationalen Macht, dass vom Correlates of War Projekt erstellt wurde. Der Index berücksichtigt die Gesamtbevölkerungsanzahl, die Anzahl urbaner Bevölkerung, die Eisen- und Stahlproduktion, den Energieverbrauch sowie die Militärausgaben und Militärgröße eines Landes. Diese Daten liegen nur bis zum Jahr 2007 vor und dienen hier als Knotenattribut.

### 5.4.3 Intra-State Conflicts

Eine zusätzliche Kovariate in unserem Modell sammelt Informationen über internationale, civile, ethnische und (?genocidal) Gewalt und Kriege. Die Daten kommen vom *Major Episodes of Political Violence* Projekt (Marshall; n.d.b), dass wie auch das Polity IV Projekt von CSP bereitgestellt wird. Die Konflikte sind abhängig von der Stärke ihrer Auswirkung auf die betroffene Bevölkerung auf einer Skala von  $1 - 10$  bewertet. Diese Informationen werden als Knotenattribut verwendet.

### 5.4.4 GDP

Zu guter Letzt verwenden wir auch das Bruttoinlandsprodukt der Länder als Knotenattribut. Die Daten entstammen dem *Maddison Project*(The Maddison-Project; n.d.).

## 5.5 geschätzte Modelle

Nach den vorhergehenden theoretischen Überlegungen und ergibt sich nach einigem Ausprobieren ein Modell dessen Output in Tabelle 5.1 wiedergegeben ist. Aufgenommen wurden die beiden einfachen Statistiken *edges* und *mutual* und die in Abschnitt 5.3 eingeführten Statistiken *gwdegree*, *gwesp* und *gwdsp*. Alle zur Verfügung stehenden exogenen Kovariablen werden in das Modell mit aufgenommen. Zusätzlich beinhaltet das Modell die Kontinente der Länder als faktorielle Einflussgröße.

## 5 Exponential Random Graph Model

**Tabelle 5.1:** Summary von Modell 1 im Jahr 1996

ergm-term	Estimate	Std.Error	p-Value
edges	-6.111e+00	2.281e-01	< 1e-04 ***
mutual	2.120e+00	9.507e-02	< 1e-04 ***
gwidegree	1.895e+00	4.818e-01	< 1e-04 ***
gwodegree	-1.311e+00	3.307e-01	< 1e-04 ***
gwestp.fixed.0.2	2.641e+00	1.778e-01	< 1e-04 ***
gwdsp.fixed.0.2	-5.686e-02	6.008e-03	< 1e-04 ***
nodeicov.ext_cinc	3.071e+00	1.291e+00	0.01740 *
nodeocov.ext_cinc	-5.967e+00	1.361e+00	< 1e-0 ***
nodeicov.ext_gdp	3.479e-06	2.099e-06	0.09749 .
nodeocov.ext_gdp	4.392e-06	1.586e-06	0.00562 **
nodeicov.ext_conflict	2.310e-02	1.704e-02	0.17530
nodeocov.ext_conflict	-1.398e-01	2.763e-02	< 1e-04 ***
nodeifactor.Continent.America	6.645e-02	6.799e-02	0.32839
nodeifactor.Continent.Asien	9.525e-02	6.473e-02	0.14116
nodeifactor.Continent.Europe	5.353e-03	7.312e-02	0.94164
nodeifactor.Continent.Oceania	-2.323e-02	1.136e-01	0.83795
nodeofactor.Continent.America	2.055e-01	6.589e-02	0.00182 **
nodeofactor.Continent.Asien	1.494e-01	6.452e-02	0.02055 *
nodeofactor.Continent.Europe	8.579e-01	7.226e-02	< 1e-04 ***
nodeofactor.Continent.Oceania	2.311e-01	9.602e-02	0.01611 *
absdiff.ext_polity	-9.360e-03	3.249e-03	0.00397 **

In Tabelle 5.1 sind die Ergebnisse der Modellschätzung von Modell 1 im Jahr 1996 angegeben. Nahezu alle aufgenommenen Statistiken erweisen sich als Signifikant. Lediglich der Faktor Kontinent für den Export und die Kovariable Conflict für den Import haben keinen signifikanten Einfluss auf den Handel mit kleinen und leichten Waffen (p-Wert > 0).

### 5.5.1 Interpretation der Parameter

Zur anschaulichen Interpretation der Parameterwerte stelle man sich zwei Netzwerke  $X^{(1)}$  und  $X^{(2)}$  vor, die in allen Statistiken gleich sind, bis auf in der Statistik  $g_i(X)$  mit

$$\delta_i(X) = g_i(X^{(1)}) - g_i(X^{(2)}).$$

Für eine einfache Interpretation der Modellparameter lässt sich Modellgleichung 5.1 nun umschreiben zu

$$\frac{P(X^{(1)})}{P(X^{(2)})} = \exp(\theta_i \delta_i(X)) \quad (5.6)$$

Das Verhältnis der Wahrscheinlichkeiten der beiden Netzwerke auf der linken Seite von Gleichung 5.6 hängt nun also nur von dem Unterschied der beiden Netzwerke in der Statistik  $g_i$  und dem dazugehörigen Parameter  $\theta_i$  ab. Nehmen wir nun an die Statistik  $g_i$

## 5 Exponential Random Graph Model

sei im Netzwerk  $X^{(1)}$  größer als im Netzwerk  $X^{(2)}$ , somit ist  $\delta_i(X) > 0$  und es ergibt sich eine einfache Interpretation des Parameters  $\theta_i$ :

- Ist  $\theta_i > 0$ , so ist  $X^{(1)}$  plausibler als  $X^{(2)}$ .
- Ist  $\theta_i = 0$ , so sind sie gleich plausibel.
- Ist  $\theta_i < 0$ , so ist  $X^{(2)}$  plausibler als  $X^{(1)}$ .

Der negative Wert der des zur Statistik *edges* gehörenden Parameters lässt sich also so interpretieren, dass unser Modell Netzwerken mit einer eher geringen Dichte eine höhere Wahrscheinlichkeit zuordnet als Netzwerken mit einer eher hohen Dichte. Dies stimmt mit den Ergebnissen der deskriptiven Analyse überein.

Die folgenden Aufzählung gibt zu jedem signifikanten Parameterwert eine kurze Interpretation die auf analogen Überlegungen basiert:

- **Edges** -6.111: Tendenz zu wenigen Kanten.
- **Mutual** 2.120: Tendenz zu gegenseitigen Handel.
- **nodeicov(CINC)** 3.071: „Mächtige“ Länder als Importland wahrscheinlich.
- **nodeocov(CINC)** -5.967: „Mächtige“ Länder als Exportland unwahrscheinlich.
- **nodeicov(GDP) & nodeocov(GDP)**  $> 0$ : Wirtschaftsstarke Länder als Handelspartner wahrscheinlich.
- **nodeocov(Conflict)** -0.1398: In Konflikte verwickelte Länder als Exporteure unwahrscheinlich.
- **nodeofactor(Continent)**  $> 0$ : Europäische Länder als Exporteure am wahrscheinlichsten, Afrikanische Länder als Exporteure am unwahrscheinlichsten.
- **absdiff(Polity)** -0.00936: Handel zwischen Ländern mit geringem Unterschied im Demokratiescore wahrscheinlich.

Die Ergebnisse erscheinen einleuchtend bis auf der negative Parameterwert der Statistik *nodeocov(CINC)*. Dieser bedeutet, dass das Modell 1 einem Knoten mit einem kleinen CINC-Wert eher als Exportpartner auswählt, als einen Knoten mit großem CINC-Wert. Das würde bedeuten, dass ein mächtiges Land eine geringere Wahrscheinlichkeit besitzt als Exporteur aufzutreten als ein Land mit geringer Macht, was in gewisser Weise kontraintuitiv wirkt. Möglicherweise liegt dies an Eigenschaften der CINC-Variable selbst. Durch eine deskriptive Betrachtung der Variable könnte eine alternative Modellierung (z.B quadratische Modellierung) gerechtfertigt werden. Dies ist im Rahmen dieser Arbeit allerdings nicht geschehen.

## 5 Exponential Random Graph Model

Eine etwas andere Herangehensweise erfordert die Interpretation der gewichteten Statistiken. Dies wird nun beispielhaft anhand der Statistik *gwidegree* erläutert: Dazu nimmt man an, dass das Hinzufügen einer Kante nur den Degree  $k$  eines Knotens des Netzwerkes verändert. Daraus folgt eine Veränderung der Degreesequenz eines Netzwerkes von  $(D_k, D_{k+1})$  zu  $(D_k - 1, D_{k+1} + 1)$ . Nun lässt sich das Verhältnis der Wahrscheinlichkeit des Netzwerkes vor Änderung der Sequenz zu der des Netzwerkes nach Änderung der Sequenz mithilfe von Gleichung 5.3 ausdrücken als:

$$\frac{P_{After}}{P_{Before}} = \exp(\theta \rho^k), \rho = 1 - e^{-\phi} \quad (5.7)$$

Der Parameter  $\theta$  lässt sich nun so Interpretieren, dass ein positiver Parameter für eine Tendenz zum Hinzufügen von Kanten spricht, da  $\rho^k > 0$  ist und dadurch  $\exp(\theta \rho^k) > 1$ . Ein negatives  $\theta$  hingegen spricht gegen die Hinzunahme von zusätzlichen Kanten. Auch die Interpretation des Decay-Parameters  $\phi$  wird aus Gleichung 5.7 für den Wertebereich  $]0, \infty[$  anschaulich. Ein  $\phi$  nahe Null resultiert in einem kleinen Wert von  $\rho$ . Hier verschwindet bei zunehmendem Degree  $k$  der Einfluss des Parameters  $\theta$  schnell, während ein großer Wert von  $\phi$ , also ein  $\rho$  nahe Eins nur einen geringen Verfall des Einflusses des  $\theta$  Parameters für wachsendes  $k$  zufolge hat.

Demnach lassen sich die Parameterwerte für die gewichteten Statistiken aus Modell 1 interpretieren. Der positive Wert des zur Statistik *gwidegree* gehörenden Parameters begünstigt die Hinzunahme von Kanten die zu einem Knoten hinführen. Das Modell tendiert also dazu einem Land zusätzliche Importpartner zuzuweisen. Der Decay-Parameter von 0.2 sorgt dafür, dass diese Tendenz bei steigendem In-Degree der Knoten rasch abnimmt. Die restlichen Parameterwerte lassen sich wie folgt interpretieren:

- **GWIDEGREE** 1.895: Tendenz zur Hinzunahme von Importpartnern.
- **GWODEEGREE** -1.211: Tendenz gegen Hinzunahme von Exportpartnern.
- **GWESP** 2.641: Tendenz zur Schließung von Dreiecken.
- **GWDSP** -0.05686: Tendenz gegen Schließung von offenen Dreiecken.
  
- **DECAY** 0.2: Tendenz verschwindet schnell.

Die Wahl des Decay-Parameters erfolgte im Rahmen dieser Arbeit durch ausprobieren. Eine freie Schätzung des Parameters ist zwar theoretisch möglich, führt aber in der Praxis häufig zu sehr langen Rechenzeiten und zur Divergenz des Schätzalgorithmus (Goodreau et al.; 2008).

### 5.5.2 Überprüfung der Ergebnisse

Das Modell 1 wurde nun geschätzt und interpretiert. Anschließend soll die Validität dieser Ergebnisse anhand zweier Kriterien überprüft werden. Zum einen wird Anhand einer MCMC-Diagnose überprüft ob der in Abschnitt 5.1.1 eingeführte Algorithmus zur Simulation von Zufallsgraphen erfolgreich konvergiert ist. Zum anderen werden durch sogenannte Goodness of Fit Plots aus dem Modell gezogene Netzwerke hinsichtlich einiger Statistiken mit dem beobachteten Netzwerk verglichen, um sicherzustellen, dass das Modell das beobachtete Netzwerk realistisch abbildet.

#### MCMC-Diagnose

Anhand einer MCMC-Diagnose kann überprüft werden ob ein Makrov Chain Monte Carlo Algorithmus wie in Abschnitt 5.1.1 konvergiert ist. Uns diene der Algorithmus um zufällige Graphen aus einer Zielverteilung  $P_\theta$  zu ziehen. Bei gewünschter Funktionsweise und Konvergenz des Algorithmus erhält man nach dem Verwerfen des Burn-Ins und ausreichend großem Thinning zufällige Ziehungen. Die Konvergenz kann nun überprüft werden, indem man sich die Werte der Statistiken der gezogenen Graphen betrachtet. Stammen die Graphen tatsächlich aus der selben Zielverteilung so sollten die Statistiken um einen Mittelwert schwanken und keine Trends mehr aufweisen.

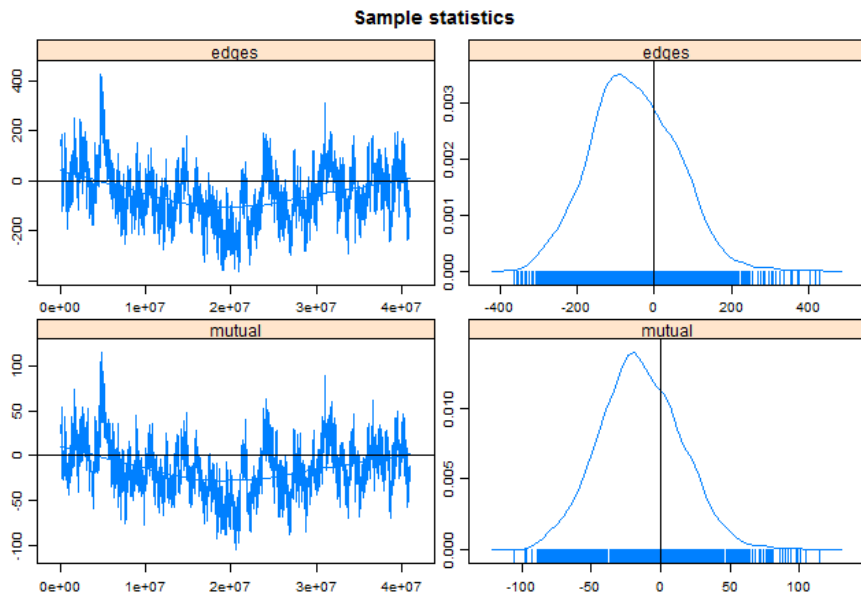
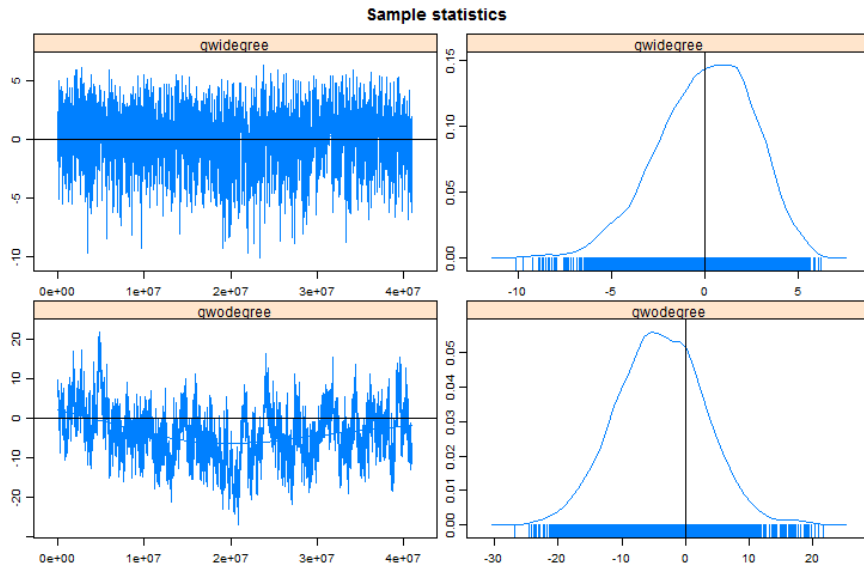


Abbildung 5.1: MCMC Diagnose von Modell 1 (1996) - edges und mutual

Beispielhaft aufgeführt sind hier die MCMC-Diagnosen der Statistiken *edges*, *mutual*, *gwidegree* und *gwodegree*. Bei allen vier Statistiken fällt die Diagnose zufriedenstellend aus. Es ist kein klarer Trend über die Ziehungen zu erkennen und der Werte der Statistiken schwanken, wenn auch nicht perfekt, symmetrisch um ihre Mittelwerte. Die MCMC-

## 5 Exponential Random Graph Model



**Abbildung 5.2:** MCMC Diagnose von Modell 1 (1996) - qwidegree und gwdegree

Diagnosen der restlichen im Modell 1 enthaltenen Statistiken geben ein ähnliches Ergebnis und sind hier deswegen nicht komplett abgebildet.

### Goodness of Fit

Nachdem die Funktion des Simulationsalgorithmus überprüft wurde stellt sich nun die Frage ob das Modell auch die richtige Zielverteilung spezifiziert. Zu überprüfen gilt also: Stimmen die gezogenen zufälligen Graphen mit dem beobachteten Graphen der die Grundlage der Modellierung bildet überein. Hierzu verwendet man sogenannte Goodness of Fit Plots. Den Statistiken aus vielen zufällig gezogenen Graphen wird jene des beobachteten Graphen gegenübergestellt. Idealerweise sollte die Statistiken der gezogenen Graphen im Mittelwert mit dem beobachteten Statistiken übereinstimmen oder diesen zumindest überdecken.

Bei den Goodness of Fit Plots in Abbildung 5.3 fällt auf: out degree und minimum geodesic distance sind sehr gut getroffen, aber die esp- Statistiken werden bei kleinen Werten stark überschätzt und danach unterschätzt. Der In-Degree wird vor allem bei den Werten 1 und 2 nicht getroffen. Mögliche Lösung: Variation des decay-Parameters? Leider tendiert das Modell für größere decay Werte zur Degeneration wie auch in Goodreau et al. (2008) berichtet.



## 5 Exponential Random Graph Model

Goodness-of-fit diagnostics

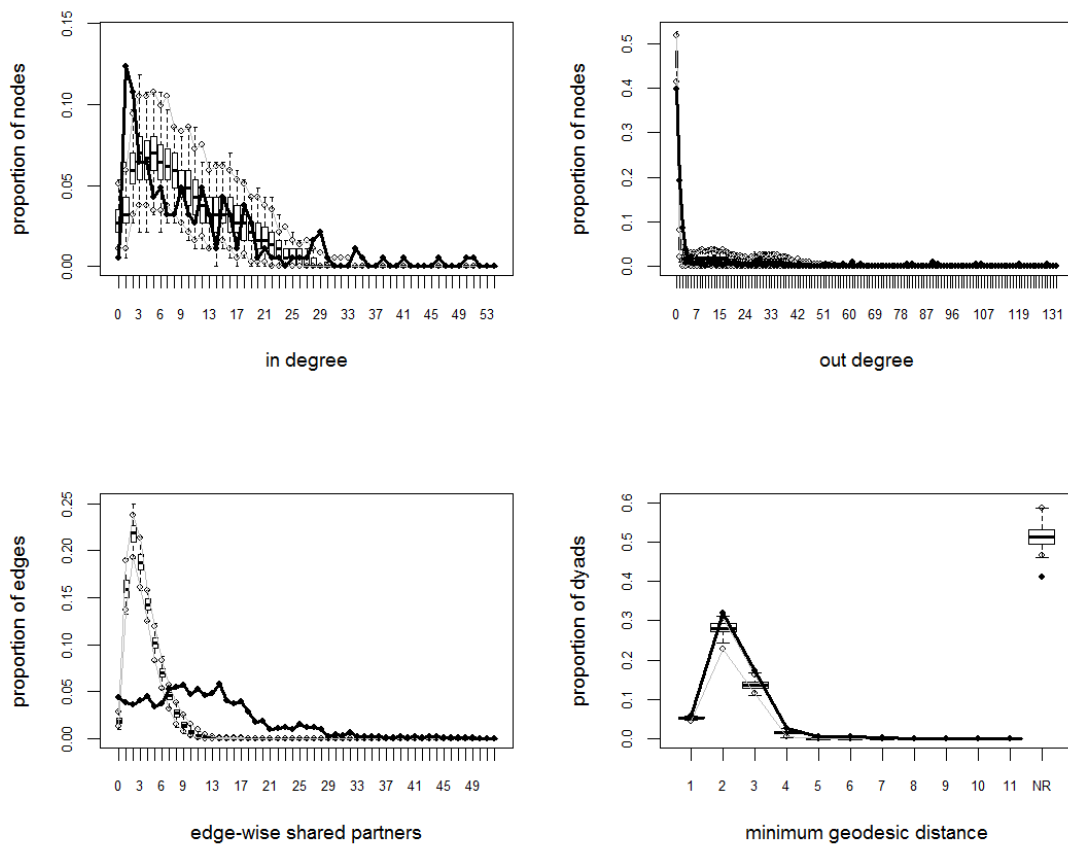


Abbildung 5.3: Goodness of Fit von Modell 1

# Literaturverzeichnis

Arel-Bundock, V. (2014). *countrycode: Convert Country Names and Country Codes*. R package version 0.18.

**URL:** <http://CRAN.R-project.org/package=countrycode>

Cranmer, S. J. and Desmarais, B. A. (2011). Inferential network analysis with exponential random graph models, *Political Analysis* **19**(1): 66–86.

Goodreau, S. M., Handcock, M. S., Hunter, D. R., Butts, C. T. and Morris, M. (2008). A statnet tutorial, *Journal of statistical software* **24**(9): 1.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., Krivitsky, P. N., Bender-deMoll, S. and Morris, M. (2014). *statnet: Software tools for the Statistical Analysis of Network Data*, The Statnet Project (<http://www.statnet.org>). R package version 2014.2.0.

**URL:** [CRAN.R-project.org/package=statnet](http://CRAN.R-project.org/package=statnet)

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., Krivitsky, P. N. and Morris, M. (2015). *ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks*, The Statnet Project (<http://www.statnet.org>). R package version 3.4.0.

**URL:** <http://CRAN.R-project.org/package=ergm>

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M. and Morris, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data, *Journal of statistical software* **24**(1).

Handcock, M. S., Robins, G., Snijders, T. A., Moody, J. and Besag, J. (2003). Assessing degeneracy in statistical models of social networks, *Technical report*, Citeseer.

Hunter, D. R. (2007). Curved exponential family models for social networks, *Social networks* **29**(2): 216–230.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M. and Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks, *Journal of statistical software* **24**(3).

Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data*, Springer New York.

**URL:** <http://ebooks.ub.uni-muenchen.de/40006/>

## Literaturverzeichnis

- Lusher, D., Koskinen, J. and Robins, G. (2012). *Exponential random graph models for social networks: Theory, methods, and applications*, Cambridge University Press.
- Marshall, M. G. (n.d.a). International military alliances , 1648-2008, Polity IV Project, Center for Systematic Peace. Version 4.1.
- Marshall, M. G. (n.d.b). Major episodes of political violence , 1946-20013, Center for Systematic Peace. Version 4.1.
- Morris, M., Handcock, M. S. and Hunter, D. R. (2008). Specification of exponential-family random graph models: terms and computational aspects, *Journal of statistical software* **24**(4).
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.  
**URL:** <http://www.R-project.org>
- Singer, J. D. (1988). Reconstructing the correlates of war dataset on material capabilities of states, 1816–1985, *International Interactions* **14**(2): 115–132.
- The Maddison-Project, . (n.d.). Growth domestic product, <http://www.ggd.net/maddison/maddison-project/home.htm>. Version 4.1.

# Abbildungsverzeichnis

4.1	Deskriptive Maßzahlen des Kleinwaffenhandelsnetzwerkes 1992-2011 . . .	10
4.2	Boxplot für In- und Out-Degree in den Jahren 1992-2011 . . . . .	12
4.3	Average nearest Neighbor Degree am Beispiel von 1998 . . . . .	13
4.4	Vergleich der 1% teuersten Waffenkäufe mit den 99% billigsten in den Jahren 1992-2011 . . . . .	13
4.5	Zeitreihen der jährlichen Handelswerte der Top-Exporteure von 1992 bis 2011	15
4.6	Zeitreihen der jährlichen Handelswerte der Top-Importeure von 1992 bis 2011	15
4.7	Visualisierung des Netzwerkes 1992 (l.) und 2011(r) . . . . .	17
4.8	Handelsströme zwischen den Kontinenten von 1992-2011 . . . . .	19
5.1	MCMC Diagnose von Modell 1 (1996) - edges und mutual . . . . .	31
5.2	MCMC Diagnose von Modell 1 (1996) - gwidegree und gwodegree . . . . .	32
5.3	Goodness of Fit von Modell 1 . . . . .	33

# Tabellenverzeichnis

4.1	Summierte Handelswerte der Top-Importeure und Top-Exporteure des Netzwerkes von 1992 bis 2011 . . . . .	14
4.2	Summierte Handelswerte der Top-Exporteure und Top-Importeure relativ zum BIP pro Kopf des Netzwerkes von 1992 bis 2011 . . . . .	16
4.3	Zentrale Akteure des Netzwerkes 1992 und 2011 . . . . .	18
5.1	Summary von Modell 1 im Jahr 1996 . . . . .	28
5.2	Auflistung der gelöschten Schleifen im Datensatz . . . . .	41

# Anhang

row	Reporter_Name	Partner_Name	Year	Value	PRIO_Weapons_Code
328	Albania	Albania	2005	3366.00	223
346	Albania	Albania	2006	13355.00	223
4143	Australia	Australia	2004	586143.00	210
4161	Australia	Australia	2004	13241.00	223
4183	Australia	Australia	2004	721.00	227
4211	Australia	Australia	2004	4693884.00	260
4224	Australia	Australia	2004	48697.00	417
4257	Australia	Australia	2005	15054.00	210
4318	Australia	Australia	2005	3180066.00	260
4329	Australia	Australia	2005	639.00	417
4362	Australia	Australia	2006	42325.00	210
4396	Australia	Australia	2006	31458.00	227
4422	Australia	Australia	2006	2473903.00	260
4435	Australia	Australia	2006	682288.00	417
4466	Australia	Australia	2007	1431868.00	210
4528	Australia	Australia	2007	4132772.00	260
4579	Australia	Australia	2008	869812.00	210
4641	Australia	Australia	2008	2870285.00	260
4653	Australia	Australia	2008	6652.00	417
4692	Australia	Australia	2009	37710.00	210
4731	Australia	Australia	2009	8857.00	227
4758	Australia	Australia	2009	857590.00	260
4804	Australia	Australia	2010	19084.00	210
4839	Australia	Australia	2010	13162.00	227
4866	Australia	Australia	2010	580277.00	260
4876	Australia	Australia	2010	1836.00	417
4915	Australia	Australia	2011	283630.00	210
4955	Australia	Australia	2011	33770.00	227
4985	Australia	Australia	2011	4529308.00	260
9934	Belgium	Belgium	2010	1056327.00	223
9951	Belgium	Belgium	2010	27087.00	227
11189	Bosnia-Herzegovina	Bosnia-Herzegovina	2003	94087.00	417
13542	Bulgaria	Bulgaria	2011	472000.00	200
15412	Canada	Canada	2002	2397.00	227
15434	Canada	Canada	2002	823.00	260
15455	Canada	Canada	2002	91030.00	417
15567	Canada	Canada	2003	13484.00	260
15586	Canada	Canada	2003	194230.00	417
15626	Canada	Canada	2004	11122.00	210

# *Tabellenverzeichnis*

15698	Canada	Canada	2004	7179.00	260
15721	Canada	Canada	2004	221463.00	417
15848	Canada	Canada	2005	32756.00	417
15937	Canada	Canada	2006	7181.00	227
15967	Canada	Canada	2006	1616.00	260
15997	Canada	Canada	2006	355640.00	417
16085	Canada	Canada	2007	13454.00	227
16146	Canada	Canada	2007	208218.00	417
16178	Canada	Canada	2007	12.00	418
16201	Canada	Canada	2008	1000.00	210
16242	Canada	Canada	2008	60664.00	227
16268	Canada	Canada	2008	250.00	260
16292	Canada	Canada	2008	35525.00	417
16383	Canada	Canada	2009	23435.00	227
16408	Canada	Canada	2009	45762.00	260
16428	Canada	Canada	2009	24582.00	417
16509	Canada	Canada	2010	34898.00	223
16532	Canada	Canada	2010	78136.00	227
16566	Canada	Canada	2010	9716.00	260
16593	Canada	Canada	2010	938822.00	417
16691	Canada	Canada	2011	19734.00	227
16722	Canada	Canada	2011	6694.00	260
16745	Canada	Canada	2011	11456.00	417
16780	Canada	Canada	2011	10.00	418
21855	Cyprus	Cyprus	2001	157920.00	418
21957	Cyprus	Cyprus	2003	606.00	418
24043	Czech Republic	Czech Republic	2004	133060.00	200
28760	Estonia	Estonia	2007	5085.00	417
32299	France	France	2000	43412.00	223
32350	France	France	2000	361290.00	417
32375	France	France	2000	619970.00	418
32398	France	France	2001	89525.00	223
32424	France	France	2001	6714.00	227
32453	France	France	2001	82811.00	417
32473	France	France	2001	125337.00	418
32501	France	France	2002	16244.00	223
32525	France	France	2002	11601.00	227
32557	France	France	2002	11601.00	417
32577	France	France	2002	475712.00	418
32606	France	France	2003	2723.00	223
32632	France	France	2003	2723.00	227
32662	France	France	2003	1078653.00	417
32683	France	France	2003	1070481.00	418
32703	France	France	2004	990720.00	223
32728	France	France	2004	1937735.00	227
32759	France	France	2004	748867.00	417
32784	France	France	2004	1264627.00	418
32898	France	France	2006	79966.00	223
32925	France	France	2006	82722.00	227

# *Tabellenverzeichnis*

32957	France	France	2006	24817.00	417
33016	France	France	2007	55657.00	223
33039	France	France	2007	131821.00	227
33080	France	France	2007	90809.00	417
33106	France	France	2007	29292.00	418
33130	France	France	2008	15382.00	223
33154	France	France	2008	153832.00	227
33183	France	France	2008	21535.00	417
33208	France	France	2008	24613.00	418
33237	France	France	2009	8619.00	223
33261	France	France	2009	20111.00	227
33293	France	France	2009	31603.00	417
33320	France	France	2009	301662.00	418
33406	France	France	2010	8121.00	417
33425	France	France	2010	2706.00	418
33472	France	France	2011	2782.00	227
33501	France	France	2011	11132.00	417
60737	Malaysia	Malaysia	2010	67114.00	418
68357	New Zealand	New Zealand	2008	5021.00	223
68378	New Zealand	New Zealand	2008	4422.00	227
68708	New Zealand	New Zealand	2011	648.00	227
84157	Slovakia	Slovakia	2002	115.00	260
84174	Slovakia	Slovakia	2002	28790.00	417
84230	Slovakia	Slovakia	2003	11114.00	260
84249	Slovakia	Slovakia	2003	37417.00	417
95974	Thailand	Thailand	2007	15078.00	210
96004	Thailand	Thailand	2007	320.00	227
96375	Thailand	Thailand	2011	11096.00	227
101434	United Kingdom	United Kingdom	2000	2663869.00	223
101462	United Kingdom	United Kingdom	2000	1450872.00	227
101507	United Kingdom	United Kingdom	2000	216810.00	418
101542	United Kingdom	United Kingdom	2001	3715829.00	223
101569	United Kingdom	United Kingdom	2001	503356.00	227
101646	United Kingdom	United Kingdom	2002	1602392.00	223
101671	United Kingdom	United Kingdom	2002	1051572.00	227
101702	United Kingdom	United Kingdom	2002	177831.00	417
101718	United Kingdom	United Kingdom	2002	374054.00	418
101752	United Kingdom	United Kingdom	2003	6230061.00	223
101782	United Kingdom	United Kingdom	2003	1625580.00	227
101811	United Kingdom	United Kingdom	2003	15105.00	417
101859	United Kingdom	United Kingdom	2004	8316031.00	223
101883	United Kingdom	United Kingdom	2004	2539407.00	227
101932	United Kingdom	United Kingdom	2004	6603.00	418
101967	United Kingdom	United Kingdom	2005	7487298.00	223
101996	United Kingdom	United Kingdom	2005	2504286.00	227
102027	United Kingdom	United Kingdom	2005	53655.00	417
102045	United Kingdom	United Kingdom	2005	6252.00	418
102085	United Kingdom	United Kingdom	2006	7520540.00	223
102115	United Kingdom	United Kingdom	2006	3570588.00	227



### *Tabellenverzeichnis*

102148	United Kingdom	United Kingdom	2006	49884.00	417
102193	United Kingdom	United Kingdom	2007	7235351.00	223
102222	United Kingdom	United Kingdom	2007	3207893.00	227
102255	United Kingdom	United Kingdom	2007	22582.00	417
102305	United Kingdom	United Kingdom	2008	3432365.00	223
102333	United Kingdom	United Kingdom	2008	983561.00	227
102417	United Kingdom	United Kingdom	2009	3660097.00	223
102442	United Kingdom	United Kingdom	2009	1124347.00	227
102483	United Kingdom	United Kingdom	2009	22303.00	417
102540	United Kingdom	United Kingdom	2010	4121609.00	223
102564	United Kingdom	United Kingdom	2010	732804.00	227
102647	United Kingdom	United Kingdom	2011	4910676.00	223
102670	United Kingdom	United Kingdom	2011	1131532.00	227
102699	United Kingdom	United Kingdom	2011	25154.00	417

---

**Tabelle 5.2:** Auflistung der gelöschten Schleifen im Datensatz

# Eidesstattliche Erklärung

Ich erkläre hiermit, dass ich diese Arbeit ohne fremde Hilfe angefertigt und nur die im Literaturverzeichnis aufgeführten Quellen und Hilfsmittel benutzt habe. Diese Arbeit wurde noch nicht zu anderen prüfungsrelevanten Zwecken vorgelegt.

.....

*Ort, Datum*

.....

*Felix Loewe*