

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

Bachelorarbeit

Statistische Analyse von Netzwerkdaten

Verfasserin: Carolin Nawab

Bachelorbetreuer: Prof. Dr. Göran Kauermann

München, der 19.08.2024



Zusammenfassung

Die Analyse von sozialen und politischen Phänomenen in Politik- und Sozialwissenschaften beruht häufig auf Netzwerkdaten. Quantitative Methoden der Netzwerkdatenanalyse aus der Statistik umfassen unter anderem klassische deskriptive Ansätze sowie statistische Modelle. In dieser Arbeit sollen Grundlagen der Netzwerkdatenanalyse erklärt werden, wobei der Fokus auf der Analyse der Effekte der Homophilie bezüglich der Bindungsbildung mithilfe einer speziellen Art von Modellen liegt, den exponential random graph models. Anhand eines Datensatzes zu Co-Sponsorship-Beziehungen in der italienischen Regierung aus den Jahren 2008 bis 2013 soll anschließend ein exponential random graph model (ERGM) gefittet werden, um mögliche Effekte der Homophilie in der Tendenz zur Bindung durch Co-Sponsorship zwischen den Abgeordneten zu untersuchen. Im ersten Abschnitt sollen Grundbegriffe geklärt werden. Im zweiten Abschnitt wird die Datensituation erklärt, wichtige theoretische Grundlagen der ERGMs eingeführt sowie das Modell spezifiziert. Zuletzt werden die Ergebnisse des ERGMs evaluiert und vorgestellt, sowie ein Resümee gezogen.

Inhaltsverzeichnis

Zusammenfassung.....	
1. Einleitung.....	1
1.1. Netzwerkdaten.....	1
1.2. Homophilie.....	6
1.3. Co-sponsoring	7
1.4. Forschungsfrage	7
2. Methodik.....	8
2.1. Daten.....	8
2.2. Co-Sponsorship in den Daten.....	8
2.3. ERGMs	14
2.4. Modellspezifikation	18
3. Ergebnisse	20
3.1. Modellqualität	20
3.2. Ergebnispräsentation.....	21
4. Resümee.....	24
Literaturverzeichnis	25
Appendix	27
Urheberschaftserklärung	35

1. Einleitung

Bei der Netzwerkanalyse handelt es sich um eine Reihe von Methoden, mit denen untersucht wird, wie verschiedene Akteure miteinander in Beziehung stehen. Dabei werden die Akteure innerhalb eines Netzwerkes als Knoten und die Beziehungen zwischen ihnen als Kanten dargestellt. Netzwerkdatenanalyse wird häufig in Bereichen wie Politik- und Sozialwissenschaften verwendet und benötigt dazu auch quantitative Ansätze aus der Statistik. Die verschiedenen Ansätze der quantitativen Netzwerkanalyse reichen von klassischer deskriptiver Statistik, Visualisierung und Charakterisierung der Struktur der Netzwerkdaten über Stichproben, Modellierung und Inferenz (vgl. Kolaczyk und Csárdi 2020; Preface: S. xii). Traditionelle Modelle gehen jedoch oft von einer Unabhängigkeit der Akteure aus, was zu fehlerhaften Schlussfolgerungen führen kann (vgl. Cranmer und Desmarais 2011: S. 66). Die Netzwerkanalyse mit exponential random graph models (ERGM), wie sie in dieser Arbeit angewendet werden soll, ermöglicht die gleichzeitige Analyse von Einfluss der Eigenschaften der Akteure und inhärenten Strukturen der Netzwerkdaten. Gleichzeitig sind keine Annahmen der Unabhängigkeit der Akteure oder deren Verbindungen innerhalb des Netzwerkes erforderlich (vgl. Cranmer und Desmarais 2011: S. 67).

1.1. Netzwerkdaten

Netzwerke, auch Graphen genannt, bestehen aus zwei Hauptkomponenten, den Knoten und den Kanten. Knoten sind die einzelnen Elemente des Netzwerks, Kanten sind die Verbindungen, die zwischen den Knoten bestehen (vgl. Haim, 2023: S. 281). Im Beispiel des später analysierten Datensatzes ist jeder einzelne italienische Abgeordnete im Datensatz ein Knoten. Die Beziehungen zwischen zwei Abgeordneten, die anhand von Co-Sponsoring der vorgestellten Gesetzesentwürfe zwischen den Abgeordneten fassbar gemacht werden, sind die Kanten.

Durch mehrere dieser Beziehungen, den Kanten, zwischen verschiedenen Netzwerkelementen, den Knoten, wird ein Netzwerk aufgespannt. In der englischsprachigen Fachliteratur werden Knoten als „vertices“ oder „nodes“ bezeichnet, die Kanten als „edges“ (vgl. Kolaczyk und Csárdi, 2020: S. 8 und Hunter, 2019: S. 25).

Formell definieren Kolaczyk und Csárdi das Netzwerk, bzw. den Graphen: „graph $G = (V, E)$ is a mathematical structure consisting of a set V of vertices [...] and a set E of edges [...], where

elements of E are unordered pairs $\{u, v\}$ of distinct vertices $u, v \in V$ “(Kolaczyk und Csárdi, 2020: S. 14). Des weiteren definieren Sie die „number of vertices $N_v = |V|$ and the number of edges $N_e = |E|$ “ (ebd.: S. 14).

Man spricht von gerichteten Kanten, wenn die Richtung der Kanten bekannt ist. Eine gerichtete Kante besitzt einen Knoten als Startpunkt und einen Knoten als Endpunkt. Bei ungerichteten Kanten hingegen ist die Ausrichtung der Kante unbekannt, nur dass eine Kante zwischen zwei Knoten besteht ist gewiss. Ein Netzwerk, dessen Kanten gerichtet sind, wird auch als gerichtetes Netzwerk bezeichnet (vgl. Haim, 2023: S. 282). Im Falle von gerichteten Graphen können zwischen einem Knotenpaar also auch zwei gegensätzlich gerichtete Kanten bestehen, im Englischen wird dies als *mutuality*, also *Mutualität* bezeichnet. In dem Fall nimmt jeder der beiden Knotenpunkte also jeweils einmal den Start bzw. den Endpunkt der Kante ein (vgl. Kolaczyk und Csárdi, 2020: S. 15-16).

Netzwerke können graphisch oder in Matrixform dargestellt werden. In der grafischen Darstellung sind die Knoten als Punkte mit Linien, die die Kanten darstellen, verbunden. In Abbildung 1 sieht man zwischen Knoten 2 und 3 eine *Mutualität* in den Kanten. In dieser Abbildung sind die Knoten nummeriert, man kann die Knoten und Kanten der Netzwerke mit Attributen dekorieren (vgl. Kolaczyk und Csárdi, 2020: S. 18). Zum einen wären das die Namen der Knoten; in den Beispielabbildungen Abbildung 2 und Abbildung 1 die Nummern der Knoten.

Andere Knotenattribute sind diskrete oder stetige Variablen, wie etwa das Geschlecht der Akteure des Netzwerks oder ihre Zugehörigkeit zu Gruppen wie z.B. Klassen oder Parteien. Zu möglichen stetigen Knotenattributen zählen z.B. das Alter der Akteure, Größe oder ihr Einkommen. Einige der Knotenattribute können in der grafischen Darstellung der Netzwerke abgebildet werden. Knoten können wie in den Beispielabbildungen nummeriert bzw. mit Namen versehen werden, Gruppenzugehörigkeiten können durch Farben der Knoten und stetige Variablen wie Alter oder Einkommen durch die Knotengrößen dargestellt werden. Bei Netzwerken mit großer Knotenzahl scheitert das Hinzufügen solcher Metainformationen aber oft an der Unübersichtlichkeit der Darstellung (vgl. Haim, 2023: S. 282).

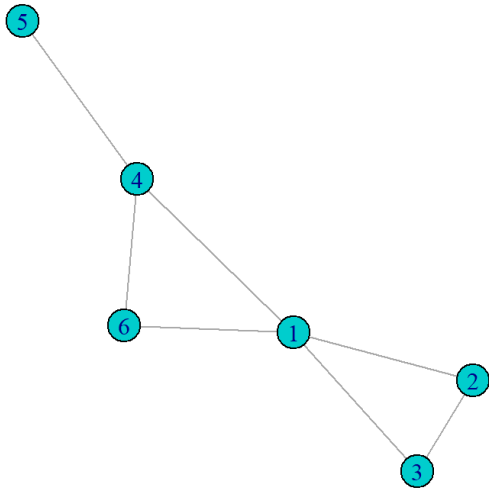


Abbildung 2: Netzwerk Darstellung ungerichtetes Netzwerk (eigene Abbildung)

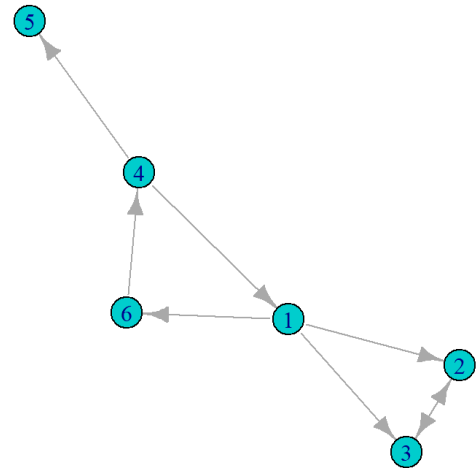


Abbildung 1: Netzwerk Darstellung gerichtetes Netzwerk (eigene Abbildung)

Kolaczyk und Csárdi beschreiben Kantenattribute als Gewichtung der Kanten. „Edge weights generally are non-negative, by convention, and often are scaled to fall between zero and one. A graph for which the edges are equipped with weights is referred to as a weighted graph“ (Kolaczyk und Csárdi, 2020: S. 19). Auch die Gewichte der Kanten können in die graphische Darstellung der

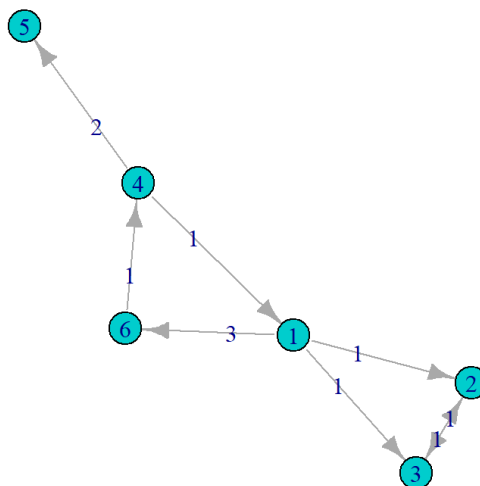


Abbildung 3: Netzwerk Darstellung ungerichtetes Netzwerk (eigene Abbildung)

Netzwerke einfließen. So können Kanten mit höherer Gewichtung zum Beispiel mit stärkerer Linienbreite dargestellt werden.

Es kann also auch zwischen gewichteten und ungewichteten Netzwerken unterschieden werden. So beschreibt Haim, dass „[in] einigen Fällen [...] mit der Beziehung zwischen zwei Knoten auch eine Stärke der Beziehung einher [geht], die sich relativ von den anderen Beziehungen in einem Netzwerk unterscheiden kann“ (Haim, 2023: S. 282). Eine Gewichtung einer Kante, also einer Beziehung zwischen zwei Akteuren in einem Netzwerk, kann so zum Beispiel einfach die Anzahl der Kontakte zwischen ihnen widerspiegeln. So wird die Kante des einmaligen Kontaktes mit 1, die Kante zwischen Akteuren mit zweimaligem Kontakt mit 2 gewichtet usw. (vgl. Kolaczyk und Csárdi, 2020: S. 22). Netzwerke, in denen Kanten nicht dementsprechend gewichtet werden, nennt man ungewichtete Netzwerke.

Die $N_v \times N_v$ Matrixdarstellung der Netzwerke, im Englischen „adjacency matrix“ (Kolaczyk und Csárdi, 2020: S. 17) genannt, ist eine weitere Darstellungsform von Netzwerken. Kolaczyk und Csárdi definieren sie für einen Graphen A folgendermaßen:

$$A_{ij} = \begin{cases} 1, & \text{falls } \{i, j\} \in E, \\ 0, & \text{sonst} \end{cases} \quad (1)$$

So ist die Matrix des Netzwerks also 1 für Einträge der Matrix, deren Zeilen- und Spaltenindizes (i und j) mit Knoten im Netzwerk korrespondieren, die mit einer Kante von i nach j verbunden sind. Die Matrixeinträge sind 0 für die Knoten ohne Verbindung. Bei ungerichteten Netzwerken ist die Matrix symmetrisch (vgl. Kolaczyk und Csárdi, 2020: S. 17).

	1	2	3	4	5	6
1	0	1	1	1	0	1
2	1	0	1	0	0	0
3	1	1	0	0	0	0
4	1	0	0	0	1	1
5	0	0	0	1	0	0
6	1	0	0	1	0	0

Tabelle 1: Adjacency Matrix für das ungerichtete Netzwerk aus Abbildung 2

	1	2	3	4	5	6
1	0	1	1	0	0	1
2	0	0	1	0	0	0
3	0	1	0	0	0	0
4	1	0	0	0	1	0
5	0	0	0	0	0	0
6	0	0	0	1	0	0

Tabelle 2: Adjacency Matrix für das gerichtete Netzwerk aus Abbildung 1

Tabelle 1 und Tabelle 2 zeigen die adjacency Matrix für die Beispiel Netzwerke aus der graphischen Netzwerkdarstellung (Abbildung 2 und Abbildung 1). Man sieht: das gerichtete Netzwerk aus Abbildung 1 hat eine unsymmetrische adjacency Matrix. Im Falle von gewichteten Netzwerken kann statt 1 die Gewichtung der Kanten in die Matrix eingetragen werden (vgl. Kolaczyk und Csárdi, 2020: S. 20 Fußnote).

	1	2	3	4	5	6
1	0	1	1	0	0	3
2	0	0	1	0	0	0
3	0	1	0	0	0	0
4	1	0	0	0	2	0
5	0	0	0	0	0	0
6	0	0	0	1	0	0

Tabelle 3: Adjacency Matrix für das gerichtete, gewichtete Netzwerk aus Abbildung 3 (man beachte z.B. die Gewichtung für den Eintrag (1,6))

Kolaczyk und Csárdi sprechen auch vom Konzept der Konnektivität von Netzwerken. Ein Konzept zur Bemessung der Konnektivität ist der Grad eines Knotens, auf Englisch degree genannt. „A vertex $v \in V$ is incident on an edge $e \in E$ if v is an endpoint of e . From this follows the notion of the degree of a vertex v , say d_v , defined as the number of edges incident on v “ (Kolaczyk und Csárdi, 2020: S. 23). Das bedeutet: ist ein Knoten ein Start oder Endpunkt von einer gewissen Anzahl an Kanten, so ist diese Anzahl der Grad des Knotens. In dem ungerichteten Beispielnetzwerk aus Abbildung 2 mit adjacency Matrix in Tabelle 1 hat der Knoten Nummer 2 den Grad zwei, der Knoten 1 den Grad vier ($d_2 = 2$; $d_1 = 4$).

Bei gerichteten Netzwerken unterscheidet man auch zwischen in- und out-degree, also für wie viele Kanten der Knoten Ziel- bzw. Startpunkt ist (vgl. Kolaczyk und Csárdi, 2020: S. 23). Für das gerichtete Beispielnetzwerk aus Abbildung 1 mit adjacency Matrix in Tabelle 2 beträgt der in-degree des Knotens Nummer 1 eins und der out-degree drei.

Für Netzwerkdaten ist es des Weiteren interessant, Statistiken zu betrachten, die die speziellen Charakteristiken von Netzwerkdaten zum Ausdruck bringen. Dazu gehören z.B. der Anteil an k-Stars oder Dreiecken im Netzwerk. Man möchte zum Beispiel wissen, inwieweit es wahrscheinlich ist, dass zwei Knoten, die beide mit einem dritten Knoten verbunden sind, auch miteinander verbunden werden. Dieser Begriff entspricht dem Konzept der Transitivität in sozialen Netzwerken und kann numerisch durch eine Aufzählung des Anteils der Knotentripel erfasst werden, die

verbundene Dreiecke bilden (d. h. alle drei Knoten sind durch Kanten verbunden) (vgl. Kolaczyk und Csárdi 2020: S. 5). In unserem Beispielnetzwerk aus Abbildung 1 bilden die Tripel (1, 4, 6) sowie (1, 2, 3) jeweils ein verbundenes Dreieck. Dieser Anteil von realisierten verbundenen Dreiecken an möglichen Tripeln im Datensatz kann als globaler Clustering-Koeffizient des Netzwerks betrachtet werden und wird auch Transitivität des Netzwerkes genannt (vgl. Kolaczyk und Csárdi 2020: S. 56).

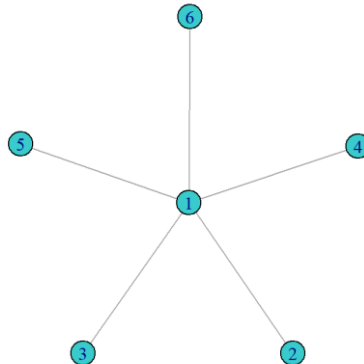


Abbildung 4: Graphische Netzwerkdarstellung k-Star (eigene Abbildung)

K-Stars sind spezielle Fälle von Netzwerkanordnungen, die aus einem mittleren Knoten bestehen, von dem aus genau k Kanten zu k anderen Knoten abgehen, die untereinander keine Kanten haben (keine Loops) (vgl. Kolaczyk und Csárdi 2020: S. 26). Abbildung 4 zeigt einen k -Star mit Knoten 1 als Mittelpunkt und 5 abzweigenden Kanten.

1.2. Homophilie

Homophilie ist das Ergebnis des Prozesses des assortative mixing, der Dominanz der Beziehungsbildung innerhalb der gleichen Kategorie (vgl. Goodreau et al.: S. 105). Homophilie ist also die Tendenz eines Akteurs eher Beziehungen zu Akteuren mit gleichen Charakteristika zu bilden. McPherson et al. merken an: „Homophily limits people’s social worlds in a way that has powerful implications for the information they receive, the attitudes they form, and the interactions they experience“ (McPherson et al. 2001: S. 415).

1.3. Co-sponsoring

Die Politikwissenschaft bietet eine Fülle von Studien, die die Bedeutung der Zusammenarbeit zwischen den Mitgliedern des Parlaments, den Abgeordneten, im Gesetzgebungsprozess betonen. Eine häufig untersuchte Art der Kollaboration zwischen Abgeordneten ist das Co-Sponsoring. Während ihrer Amtszeit haben Abgeordnete die Möglichkeit, sich gegenseitig zu unterstützen, indem sie einen von (einem) anderen Abgeordneten vorgelegten Rechtsakt unterzeichnen, z.B. einen Änderungsantrag, oder einen Gesetzentwurf. Die Möglichkeit des Co-Sponsorings gibt es nicht in jeder Demokratie, ist aber in Ländern mit parlamentarischen und präsidentiellem System verbreitet (vgl. Briatte 2016 Appendix: S. 2). Co-Sponsoring ist für Abgeordnete z.B. in den USA (vgl. Campbell 1982), Argentinien und Chile (vgl. Alemán und Calvo 2013) und einigen Ländern in Europa möglich. In Deutschland wird der Großteil legislativer Unterstützung von Gesetzesvorschlägen auf Ebene von parlamentarischen Gruppen durchgeführt und nicht als individuelle Abgeordnete (vgl. Brunner 2013: S. 16). Innerhalb der EU ist Co-Sponsoring z.B. in Island, Lichtenstein, der Schweiz und Norwegen möglich (vgl. Briatte 2016 Appendix: S. 4).

Frühere Forschung zu Co-Sponsoring fokussiert sich vor allem darauf, zugrundeliegende Motive des Co-Sponsorings aufzudecken, wie zum Beispiel die Rolle des Co-Sponsorings als Mittel zur Signalsetzung innerhalb des Parlaments und als Form der Positionseinnahme gegenüber Wählern (vgl. Schiller 1995; Kessler und Krehbiel 1996; Wilson und Young 1997; Koger 2003).

Co-Sponsoring ist von Natur aus relational, weshalb neuere Studien sich nicht nur damit beschäftigen, welche Gründe dem Co-Sponsoring Verhalten von Abgeordneten unterliegt, sondern damit, mit wem Abgeordneten Co-Sponsoring-Beziehungen eingehen (vgl. Briatte 2016 Appendix: S. 3).

Für diese Untersuchung nutzen neuere Studien Netzwerkdatenanalyse. Bezüglich des Konzepts der Homophilie untersuchen Studien unter anderem den Einfluss des gleichen Geschlechts der Abgeordneten (vgl. Clark und Caro 2013: S. 2), der gleichen Ethnizität (vgl. Bratton und Rouse 2011) oder Zugehörigkeit zur selben Partei oder Wahlkreis (vgl. Alemán und Calvo 2013).

1.4. Forschungsfrage

Das Interesse dieser Arbeit besteht daraus, Homophilie in Co-Sponsoring-Beziehungen im italienischen Parlament zu untersuchen. Dies soll mit Hilfe eines ERGMs geschehen, eines exponential random graph models. Mit dieser Methode wurden in der Fachliteratur schon einige Datensätze zu

Homophilie im Co-Sponsoring Verhalten innerhalb von Parlamenten mehrerer Länder untersucht. Bratton und Rouse beschränken sich auf die USA (vgl. Bratton und Rouse 2011), Alemán auf Chile und Argentinien (vgl. Alemán und Calvo 2013) und Baller auf das Europäische Parlament (vgl. Baller 2017). Gerade im Vergleich zum Zwei Parteien System der USA scheint die Wahl von Italien als Datenquelle interessant.

2. Methodik

Zunächst soll der verwendete Netzwerkdatensatz erklärt und die vorhandenen Variablen im Datensatz deskriptiv beschrieben werden. Anschließend wird die Motivation für das Verwenden eines ERGMs, eines exponential random graph modells, für die weitere Analyse des Datensatzes beschrieben, sowie die entsprechende Theorie eingeführt. Zuletzt wird konkret beschrieben, welche Terme für das berechnete Modell verwendet werden sollen.

2.1. Daten

Die Daten zum Co-Sponsoring in der italienischen Regierung, die in dieser Arbeit analysiert werden, stammen aus der Parlnet Datenbank (Briatte 2016), welche Daten zum Co-Sponsoring von Gesetzentwürfen in Regierungen aus 20 verschiedenen Ländern zur Verfügung stellt. Die Daten werden dabei per web scraping von Regierungswebseiten gewonnen.

Der verwendete Datensatz beinhaltet Information der Abgeordneten und deren Co-Sponsorship-Beziehungen in den Jahren 2008 bis 2013 aus der italienischen unteren Kammer des Zweikammersystems, der „Camera dei deputati“. Dies sind die neuesten vollständig vorhandenen Daten, die von der Parlnet Datenbank zur Verfügung gestellt werden.

2.2. Co-Sponsorship in den Daten

Im Zeitraum von 2008 bis 2013 war im italienischen Parlament zuerst die politisch rechtsgerichtete Koalition aus Forza Italia mit Premierminister Silvio Berlusconi und zwei weiteren Parteien bis November 2011 regierende Koalition. Im Herbst 2011 trat Berlusconi zurück und ein parteiloses Expertenkabinett, angeführt von ehemaligem EU-Kommissar Mario Monti bildete die Regierung (vgl. Grasse 2012: S. 751). Im Datensatz wird der genaue Zeitpunkt innerhalb der Zeitspanne von

2008 bis 2013, an dem das Co-Sponsorings zwischen den Abgeordneten geschehen ist, nicht bestimmt.

Im Netzwerk befinden sich 663 Knoten, bzw. Abgeordnete. Zwischen ihnen spannen sich 25730 gerichtete Kanten auf. Die Kanten sind gewichten und geben die Anzahl der Co-Sponsoring Kontakte zwischen zwei Abgeordneten an.

Von 520 Abgeordneten wurden insgesamt 4007 Gesetzentwürfe eingereicht. Im Durchschnitt reicht jeder Abgeordnete im Datensatz 6 Gesetzentwürfe ein. Jeder Gesetzentwurf wird im Mittel von 12 Abgeordneten unterschrieben. Die durchschnittliche Gewichtung der Kanten (Anzahl der Gesetzentwürfe, die zwischen einander unterschrieben wurden) zwischen zwei Abgeordneten liegt bei 2.

Die Anzahl an Abgeordneten im Datensatz, die mindestens einen Gesetzentwurf unterschrieben haben, also durch Co-Sponsoring unterstützt haben, liegt bei 662. Die durchschnittliche Anzahl an unterstützen Gesetzentwürfen durch Co-Sponsoring pro Abgeordneten liegt bei 83 (Briatte - Summary Statistics. 2016).

	Min	Max	Median	Mean
Sponsoren pro Gesetzentwurf	1	228	3	11.5
Gesetzentwürfe gesponsort pro Abgeordneten	0	78	3	6
Gesetzentwürfe vorgestellt pro Abgeordneten	0	1372	68	83

Tabelle 4: Deskriptive Statistiken des Netzwerks aus Briatte-Summary Statistics (2016)

Im Datensatz sind einige Knoten und Kanteneigenschaften vorhanden. Dazu gehört das Geschlecht der Abgeordneten, hinterlegt als binäre Variable, sowie die Zugehörigkeit zu einer von 7 Parteien. Zusätzlich gibt es im Datensatz 33 parteilose Abgeordnete, die in eine unabhängige Gruppe gegliedert werden. Die Partei mit der höchsten Anzahl an Abgeordneten im Datensatz ist Silvio Berlusconi's Forza Italia - Il Popolo della Libertà, die bis zum Regierungswechsel Ende 2011 die

politisch rechtsgerichtete Regierungskoalition mit den Parteien Lega Nord und Futuro e Libertà per Italia bildete, wobei Futuro e Libertà per Italia Ende 2010 von Popolo e Territorio abgelöst wurde. Die zweitgrößte Partei im Datensatz und wichtigste Oppositionspartei bis zum Regierungswechsel Ende 2011 ist die politisch linke Partito Democratico. Weitere Parteien sind die Mitte-Rechts-Christdemokraten Unione di Centro und die Mitte-Links-Partei Italia dei Valori. Die genaue Aufschlüsselung der Anzahl der Abgeordneten im Datensatz nach Partei ist in der Tabelle 6 dargestellt.

Ein weiteres Knotenattribut im Datensatz ist das Geburtsjahr der Abgeordneten, aus welchem eine Zuordnung zu einer von fünf Altersklassen abgeleitet werden kann. Tabelle 5 zeigt die Aufteilung der Abgeordneten auf die fünf Altersklassen.

Altersgruppe	Anzahl	Anteil
<40	85	0.13
40 - 49	208	0.31
50 - 59	247	0.37
60 - 69	108	0.16
>= 70	15	0.02

Tabelle 5: Aufteilung der Altersgruppen

Zusätzlich ist im Datensatz die Zuordnung der Partei auf eine Skala im Politischen Spektrum mit der Spanne von 0 (links) bis 10 (rechts) vorhanden. Der Wert wird aus einem gewichteten Mittelwert von Expertenbefragungen gebildet und von der Parl Gov Datenbank zur Verfügung gestellt (vgl. Briatte 2016 Appendix: S. 17).

In der untenstehenden Tabelle 6 ist die Aufteilung Abgeordneten auf die Parteien im Datensatz sowie deren politische Skala zu sehen.

Partei	Parteiname	Anzahl Abgeordnete	Politische Skala
FI-PDL	Forza Italia - Il Popolo della Libertà	256	7.1
FLI-TP	Futuro e Libertà per l'Italia	24	7.4
IDV	Italia dei Valori	25	4.9
IND	Freie Abgeordnete	33	-
LN	Lega Nord	69	7.8
PD	Partito Democratico	208	2.6
PT	Popolo e Territorio	6	6.2
UDC-TP	Unione di Centro	42	6.1

Tabelle 6: Parteien und politische Skala

Zu den Kantenattributen gehört die Gewichtung der Kanten mit der Anzahl der Co-Sponsoring-Beziehungen zwischen zwei Abgeordneten.

Für die weitere Analyse und das Modell wird das Netzwerk der Einfachheit halber als ungerichtetes, ungewichtetes Netzwerk betrachtet, wobei jede Kante mit Gewichtung ≥ 1 als eine Co-Sponsoring-Beziehung zwischen zwei Abgeordneten gezählt wird. Eine Kante im Datensatz, die von Abgeordneten x zu Abgeordneten y verläuft, und mit 3 gewichtet ist, was bedeutet, dass Abgeordneter x 3 Gesetzentwürfe von Abgeordneten y gesponsort hat, ist im neuen ungerichteten und ungewichteten Datensatz eine Co-Sponsoring-Beziehung (Kante) zwischen x und y.

	Min	Max	Median	Mean
Anzahl Co-Sponsoring-Beziehungen pro Abgeordneten	1	289	26	40.6

Tabelle 7: Deskriptive Statistiken des Netzwerks nach Umwandlung in ungewichtetes, ungerichtetes Netzwerk

Die Verteilung der Grade ist rechtsschief, die meisten der Knoten im Netzwerk haben einen Grad von unter hundert, d.h. die meisten Abgeordneten im Datensatz haben weniger als 100 Co-Sponsorship Beziehungen. Das arithmetische Mittel des Datensatzes liegt bei ca. 78.

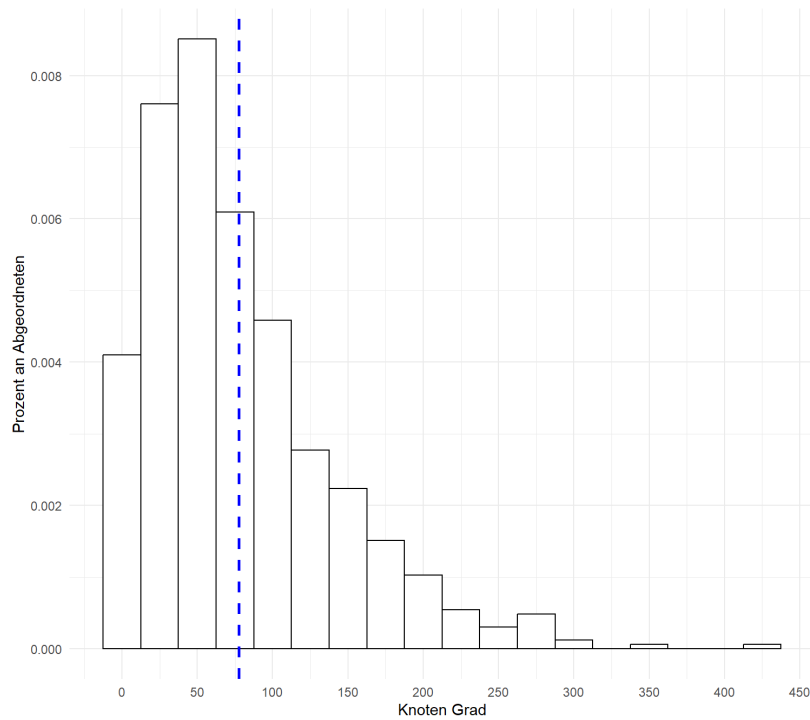


Abbildung 5: Verteilung der Grade der Knoten; arithm. Mittel in blau (eigene Abbildung)

In der graphischen Darstellung des Netzwerkes (Abbildung 6) kann man eine Gruppierung in Parteien erahnen, was für eine mögliche Homophilie bezüglich der Parteizugehörigkeit spricht. Für die Darstellung wurde eine Methode genutzt die nach dem Prinzip der Multidimensionalen Skalierung funktioniert (vgl. Kolaczyk und Csárdi 2020: S. 33).

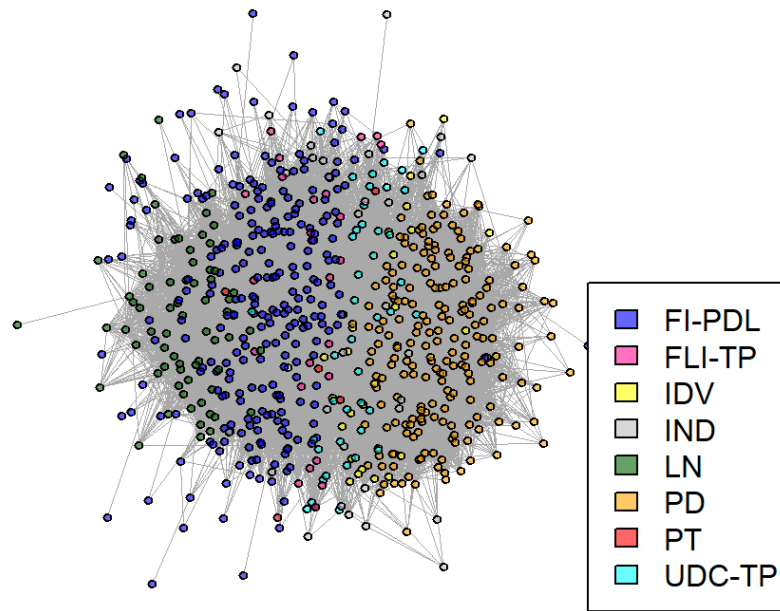


Abbildung 6: Abbildung des Netzwerkes, Färbung der Knoten nach Parteizugehörigkeit (MDS Darstellung)

Werden die Kanten nach Übereinstimmungen zwischen den beiden Knoten, die sie aufspannen, untersucht, stimmen in 73% der Co-Sponsoring-Beziehungen die Parteien der Abgeordneten überein. In 64% der Fälle stimmt das Geschlecht überein und in 31% die Altersgruppe der beiden Abgeordneten. Um diese mögliche Homophilie in der Kantenbildung im Netzwerk zu untersuchen soll im weiteren Verlauf der Arbeit ein ERGM genutzt werden.

Kategorie	Anteil Kanten im Datensatz mit Übereinstimmung
Partei	0.73
Geschlecht	0.64
Altersgruppe	0.31

Tabelle 8: Homophilie Anteile in Kanten

2.3. ERGMs

In relationalen Daten ist es üblich, dass die Bindung zwischen einem bestimmten Paar von Akteuren von einer oder mehreren anderen Bindungen im Netzwerk abhängt. Standardregressionsmodelle sind so konzipiert, dass sie die Wirkung von Kovariablen auf das Ergebnis schätzen, aber nicht den Einfluss der Ergebnisse aufeinander. Das Regressionsframework ist daher nicht nur nicht in der Lage, Netzwerkabhängigkeiten zu modellieren, sondern die Gültigkeit der Regressionsergebnisse bricht zusammen, wenn Netzwerkendogene Abhängigkeiten nicht modelliert werden (vgl. Cranmer und Desmarais 2011: S. 67).

Man nehme unseren Beispieldatensatz zum Co-Sponsoring im italienischen Parlament und ein Standardregressionsmodell als Beispiel, mit der Outcome Variable ‚Co-Sponsoring-Beziehung‘ und der einzigen unabhängigen Einflussvariable ‚gemeinsame Parteizugehörigkeit‘. Es besteht die Wahrscheinlichkeit, dass der Freund meines Freundes auch mein Freund ist, d.h. ein Abgeordneter z, der eine Co-Sponsoring-Beziehung zu Abgeordnetem x pflegt, wahrscheinlich auch eine Beziehung zu y pflegt, falls dieser seinerseits eine Co-Sponsoring-Beziehung zu x hat. Die drei Knoten würden im Netzwerk also ein Dreieck bilden. Wenn es der Fall ist, dass der Freund meines Freundes auch mein Freund ist, ist die Outcome Variable (Co-Sponsoring-Beziehung) aufgrund des endogenen Effekts im Netzwerkgenerierungsprozesses nicht unabhängig von anderen Bindungen im Netzwerk (vgl. Cranmer und Desmarais 2011: S. 67) .

Weiter beschreiben Cranmer und Desmarais:

„A triadic effect such as this cannot be captured by covariates and so, in the context of a traditional regression, must be omitted from the model specification. The effect of this omission will be bias in the coefficients and inconsistency in the SEs because of model misspecification“ (Cranmer und Desmarais 2011: S. 67).

Wird also ein Standardregressionsmodell verwendet, selbst wenn nur der Effekt einer bestimmten Kovariablen von Interesse ist, kann der Bias, der durch das Weglassen der Netzwerkspezifischen Effekte eintritt, die Ergebnisse der Analyse verfälschen (vgl. Cranmer und Desmarais 2011: S.67).

Für die Analyse von Netzwerkdaten wird darum ein ERGM, ein exponential random graph model verwendet. Das ERGM ist ein statistisches Modell, das verwendet werden kann, um die Auswirkungen von Kovariablen auf die Bindungen in einem Netzwerk zu schätzen und gleichzeitig

Parameter zu schätzen, die endogene Effekte im Netzwerk beschreiben (vgl. Cranmer und Desmarais 2011: S.67). ERGMs ermöglichen die Erstellung eines Modelles durch einen Mechanismus, der von der Regression abweicht. Das beobachtete Netzwerk wird beim ERGM als eine einzige Realisierung aus einer multivariaten Verteilung von möglichen Netzwerken behandelt; in Folge dessen sind keine Annahmen über die Unabhängigkeit der Akteure oder Bindungen innerhalb des Netzwerks erforderlich (vgl. Cranmer und Desmarais 2011: S. 67).

„The key difference in the way one must think of inference in an ERGM context versus a regression context is that the ERGM treats the entire network as a single observation. In other words, instead of considering each dyadic tie as a conditionally independent observation, the ERGM considers the whole network of ties to be one observation from a complex multivariate distribution—thus making it completely free of independence assumptions—and then asks what the probability of observing that network is compared to a random network with the same number of nodes“ (Cranmer, Desmarais, und Morgan 2020: S. 292 - 293).

ERGMs können sowohl exogene Effekte (Kovariablen) als auch Effekte, die endogen für das Netzwerk sind (strukturelle Effekte im Netzwerk), modellieren. Abhängigkeitsterme in der Spezifikation des ERGMs, bei denen es sich um endogene Maße im Netzwerk handelt, erfassen Theorien, die Beziehungen zwischen den Bindungen im Netzwerk widerspiegeln (z. B. der Freund meines Freundes ist mein Freund) (vgl. Cranmer und Desmarais 2011: S.67). Cranmer und Desmarais fassen zusammen: „In other words, researchers can proceed with ERGM analysis based on hypotheses similar to those that would produce regression specifications (i.e., covariate x is expected to affect the outcome y), and as much network structure (dependence) as they see fit“ (Cranmer und Desmarais 2011: S. 67).

Für ungewichtete, ungerichtete Netzwerke wurde von Frank and Strauss (1986) das ERGM eingeführt. Unter anderem wurde das ERGM von Frank (1991), Wassermann und Pattison (1996) erweitert, mit dem Fokus auf gerichtete Netzwerke (vgl. Snijders 2002: S. 3).

Man betrachte $G = (V, E)$ als ungerichteten, ungewichteten Graphen. Sei $Y_{ij} = Y_{ji}$ eine binäre Zufallsvariable, die das Vorhandensein oder Fehlen einer Kante $e \in E$ zwischen den beiden Eckpunkten i und j in V anzeigt. Die symmetrische Matrix $\mathbf{Y} = [Y_{ij}]$ ist somit die Adjacency Matrix für G .

Sei $\mathbf{y} = [y_{ij}]$ eine bestimmte Realisierung von \mathbf{Y} . Ein ERGM ist ein Modell, das in der Exponentialfamilienform für die gemeinsame Verteilung der Elemente in \mathbf{Y} angegeben ist (vgl. Kolaczyk und Csárdi 2020: S. 88 - 89).

Wasserman und Pattison formulierten eine Wahrscheinlichkeitsverteilung der Beobachtung des Netzwerks \mathbf{y} über alle möglichen Netzwerke, die beobachtet hätten werden können, die wie folgt gegeben ist:

$$\mathbb{P}_{\theta}(\mathbf{Y} = \mathbf{y}) = \left(\frac{\exp \{\theta s(\mathbf{y})\}}{\kappa(\theta)} \right) \quad (2)$$

Dabei ist θ ein Vektor der Modellparameter, die geschätzt werden und $s(x)$ ein Vektor der Netzwerkstatistiken (vgl. Wasserman und Pattison 1996: S. 406). Aus (2) ergibt sich auch der Name der Modelle als exponential random graph model: „These models are of exponential family form, in which the probability function depends on an exponential function of a linear combination of network statistics“ (Wasserman und Pattison 1996: S. 406). Wasserman und Pattison selbst nennen das Modell p^* Model (vgl. ebd. 1996: S. 406).

Die Normierungskonstante $\kappa = \kappa(\theta)$:

$$\kappa(\theta) = \sum_{\mathbf{y}} \exp \{\theta s(\mathbf{y})\} \quad (3)$$

stellt sicher, dass die Wahrscheinlichkeitsverteilung normiert ist, d.h. sich über den Stichprobenraum von \mathbf{Y} , allen möglichen Netzwerken, zu 1 aufsummiert (vgl. Wasserman und Pattison 1996: S. 407).

In der einfachsten Ausführung von $s(\mathbf{y})$ beinhaltet es nur die Anzahl an Kanten in \mathbf{y} , das resultierende Modell wird auch Erdős-Rényi-Gilbert Model genannt (vgl. Hunter 2019: S. 28) und kann hinsichtlich weiterer endogener Netzwerkstatistiken erweitert werden: „Since the dimension [...] of $s(\mathbf{y})$ can take any value we wish, it is possible to incorporate into the model [...] any set of sufficient statistics that we feel are relevant to the probability of a particular network configuration“ (Hunter 2019: S. 29).

Erdős-Rényi-Gilbert (1959) haben dieses einfache Modell und weitere Modelle zeitgleich mit Gilbert (1959) beschrieben und asymptotische Eigenschaften für N , die Zahl an Knoten, gegen unendlich untersucht (vgl. Hunter 2019: S. 26).

Um die relationale Struktur der Netzwerkdaten bestmöglich auszunutzen, ist es aber gerade von Interesse auch Statistiken für endogene Terme höherer Ordnung in $s(x)$ einzubauen. Eine Vielzahl an Termen sind möglich, zum Beispiel Anzahl an k -Sternen oder Dreiecksbeziehungen (vgl. Kolaczyk und Csárdi 2020: S. 91). Die einfachste Art wäre es, in den $s(x)$ Vektor Terme für die Kantenzahl (Erfassung der Gesamtvernetzung im Netzwerk) und die Anzahl an Dreiecken oder den Clustering-Koeffizienten (Erfassung der Neigung für oder gegen Triadenschließung) aufzunehmen.

Die natürliche Übersetzung des ERGMs wäre ein s -Vektor, der Terme für die Kantenzahl (Erfassung der Gesamtsozialität) und die Dreieckszahl oder den Clustering-Koeffizienten (Erfassung der zusätzlichen Neigung für oder gegen die Triadenschließung) enthält. Jedoch ist es so, dass es bei solchen Modellen zu einem Phänomen, das ‚model degeneracy‘ genannt wird, kommen kann. Durch Simulation können durch das berechnete Modell mehrere Netzwerke erstellt werden, bei denen es zu einem von zwei verschiedenen Mustern kommt: Für die meisten Kombinationen von θ Werten erstellt das Modell Netzwerke, in denen entweder alle Knoten durch Kanten verbunden sind oder gar keine – so genannte leere Netzwerke. Bei Werten für θ , die keine solcher extremen Netzwerke produzieren, ist die Verteilung der simulierten Netzwerke oft Bimodal. Das heißt, entweder haben die Netzwerke viele geschlossene Dreiecke aber geringe Dichte, oder sie haben eine hohe Dichte und geringe Anzahl geschlossener Dreiecke. Die ursprünglich beobachtete Netzwerkrealisierung würde so fast nie vom Modell erstellt werden (vgl. Goodreau, Kitts, und Morris 2009: S. 110).

Viele der möglichen Lösungen für model degeneracy ersetzen starke Homogenitätsannahmen im Modell durch eine Form von Heterogenität. Eine plausible Alternative im Fall, in dem Triadenschließung gezählt werden soll, kann ein Term sein, der von einer Verringerung des marginalen Einflusses der Wirkung von Dreiecken im Netzwerk auf die Bildung von weiteren Kanten ausgeht, wenn schon viele Dreiecke vorhanden sind (vgl. Goodreau, Kitts, und Morris 2009: S. 110). Hunter (vgl. Hunter 2007) und Hunter und Handcock (vgl. Hunter und Handcock 2006) formulieren dazu die Statistik der ‚geometrically weighted edgewise shared partner distribution‘ (GWESP),

die eine Weiterentwicklung einer Statistik von Snijders et al. (vgl. Snijders et al. 2006) darstellt (vgl. Goodreau, Kitts, und Morris 2009: S. 110). Dabei teilen sich zwei Knoten einen Partner, wenn beide eine Kante zu demselben Partner haben und jeder gemeinsame Partner bildet ein Dreieck, wenn das ursprüngliche Paar auch durch eine Kante verbunden ist. Im Gegensatz zum Zählen von Dreiecken oder dem Clustering-Koeffizienten, welche ein einzelnes Maß für das Netzwerk ergeben, wird diese gemeinsame Partnerzählung also an jeder Kante vorgenommen, die dann ein Dreieck bildet (vgl. Goodreau, Kitts, und Morris 2009: S. 110). „The GWESP statistic defines a parametric form of this count distribution that gives each additional shared partner a declining positive impact on the probability of two persons forming a tie“ (Goodreau, Kitts, und Morris 2009: S. 111).

Für ein Set von $s(y)$ Statistiken soll also der θ -Vektor identifiziert werden, der die Modellwahrscheinlichkeit maximiert. Es ist jedoch so, dass die Normierungskonstante κ im Nenner von (2) nur für die kleinsten aller Netzwerke berechnen werden kann, was eine direkte Berechnung der Likelihoodfunktion verhindert. Mithilfe logistischer Regression kann im Spezialfall des Erdős-Rényi-Gilbert Modells, das nur die Anzahl der Kanten beinhaltet, geschätzt werden. Der Ansatz wird maximum pseudolikelihood estimation (MPLE) genannt. Für Modelle mit Abhängigkeitsstruktur zwischen Netzwerkelementen neigt MPLE aber zu schlechter Performance, weshalb die Likelihoodfunktion mit Hilfe von Markov-Ketten-Monte-Carlo-Simulationsmethoden (MCMC) angenähert wird (vgl. Goodreau, Kitts, und Morris 2009: S. 109 - 110). MCMC erzeugt für die Schätzung des θ -Vektor eine Stichprobe aus dem Raum der möglichen Netzwerke. Der MCMC Ansatz bringt den zusätzlichen Vorteil, dass der Algorithmus dazu verwendet werden kann um Netzwerke für einen gegebenen θ -Vektor zu simulieren. Eine Stichprobe solcher Realisierungen kann anschließend zur Untersuchung der Modellanpassung verwendet werden (vgl. Goodreau, Kitts, und Morris 2009: S. 110).

2.4. Modellspezifikation

Aufgrund der oben beschriebenen model degeneracy soll in unserem Modell die Dreiecksbildung mithilfe der GWESP Statistik der Form:

$$w = e^{\tau} \sum_{i=1}^{n-2} \{1 - (1 - e^{-\tau})^i\} p_i \quad (4)$$

inkludiert werden. Hier ist p_i die Anzahl an Abgeordneten, die eine Co-Sponsoring-Beziehung zueinander haben und die beide mit den gleichen i Abgeordneten durch Co-Sponsoring verbunden sind. Die GWESP Statistik enthält auch den Parameter τ , welcher für eine zunehmende Anzahl von gemeinsamen Partnern die Abnahmerate des Effekts des Triadenschlusses auf die Bindungswahrscheinlichkeit kontrolliert (vgl. Goodreau, Kitts, und Morris 2009: S.111). Im Modell wählen wir den Wert 0.2 für τ . Je niedriger der Wert von τ , desto geringer ist die Wahrscheinlichkeit für model degeneracy (vgl. Goodreau et al. 2008: S. 16).

Neben den endogenen Effekten sollen auch exogene Effekte in das Modell inkludiert werden. Zum einen sollen die Haupteffekte der Knoteneigenschaften, nämlich die Parteizugehörigkeit, das Geschlecht und die Zugehörigkeit der Abgeordneten zu einer von 5 Altersgruppen im Modell geschätzt werden. Außerdem soll das Modell Homophilie-Effekte bzgl. Parteizugehörigkeit, Geschlecht und Altersgruppe schätzen. Dabei wird zwischen uniformer und differenzieller Homophilie unterschieden. Uniforme Homophilie beschreibt z.B. die Bereitschaft grundsätzlich eher Co-Sponsoring-Beziehungen zum gleichen Geschlecht aufzubauen, unabhängig vom Geschlecht. Differenzielle Homophilie unterscheidet je nach Ausprägung; untersucht also die Bereitschaft von Frauen eher Co-Sponsoring-Beziehungen zu Frauen aufzubauen oder die Bereitschaft von Männern eher Co-Sponsoring-Beziehungen zu Männern aufzubauen. Bei den Variablen Parteizugehörigkeit und Altersgruppe soll differenzielle Homophilie untersucht werden, bei der Variable Geschlecht uniforme Homophilie. Für die Variable Geschlecht bestehen nur drei mögliche Kombinationsmöglichkeiten (MM, MF, FF), wenn auch die Haupteffekte geschätzt werden, bliebe nur ein Freiheitsgrad übrig (vgl. Goodreau, Kitts, und Morris 2009: S. 111).

Das ERGM Modell wird in der R Statistik Software (Version 4.3.0; R Core Team 2023) mithilfe der Statnet Package (Krivitsky et al. 2003) Umgebung gefittet und analysiert.

3. Ergebnisse

Zuerst soll die Qualität des Modells anhand von Überprüfung auf model degeneracy und der Modellanpassung an die Daten mithilfe von goodness-of-fit Plots bewertet werden und anschließend die geschätzten Modellkoeffizienten vorgestellt werden.

3.1. Modellqualität

Die Qualität der MCMC-Simulation, die die Koeffizienten des Modells geschätzt hat sollte überprüft werden, um sicherzugehen, dass das Modell verlässlich ist. Der MCMC-Algorithmus konvergiert und es kommt nicht zu model degeneracy. Für die Modelldiagnostik kann die Funktion `mcmc.diagnostics` aus der Statnet Umgebung in R (vgl. Krivitsky et al. 2003) genutzt werden. Sie produziert einiges an Output, von Interesse sind die zwei Textoutputs und zwei Diagnoseplots.

Die ‚sample statistics auto correlation‘ (Appendix: Abbildung 10) misst die Korrelation zwischen den Stichprobenstatistiken an unterschiedlichen Punkten im MCMC-Algorithmus. Ein guter Algorithmus sollte niedrige Werte (nahe 0 nach Lag 0) der auto-correlation haben, er mixt zufällig und biegt nicht in eine falsche Richtung ab. Das ist für unsere Statistiken der Fall.

Die ‚sample statistics burn-in diagnostic (Geweke)‘ (Appendix: Abbildung 11) gibt ein Maß der Konvergenz an, indem die Mittelwerte der Stichprobenstatistiken an unterschiedlichen Punkten im MCMC-Algorithmus verglichen werden. Die Mittelwerte sollten gleich sein, was bedeutet, dass wir nach p-Werten nahe 1 suchen; unsere p-Werte sind auch nahe 1.

Die zwei Plots zur Diagnose sind die ‚MCMC trace plots‘ und die ‚MCMC density plots‘ (Appendix: Abbildung 12- Abbildung 16). In den ersteren wird der Unterschied zwischen den Stichprobenstatistiken und des beobachteten Netzwerks für jeden Schritt der Simulation geplottet. Der Plot sollte zufällige Varianz, zentriert um 0 als Zeichen für „mixing“ in jedem Schritt zeigen. Die MCMC density plots zeigen die Verteilung der Stichprobenstatistiken, sie sollten Normalverteilt um 0 sein, d.h. nicht verschieden vom beobachteten Netzwerk (vgl. Stanford Human Evolutionary Ecology and Health 2018).

Um die Anpassung des Modells an die Daten zu überprüfen, werden goodness-of-fit (GOF) Plots genutzt. Durch das Modell wird eine Stichprobe von 100 Netzwerke simuliert. Anschließend werden eine Auswahl an im realisierten Netzwerk beobachteten Netzwerkstatistiken mit den

Mittelwerten (Median) der Stichprobe verglichen. Bei einem gutem Modellfit sollten die Werte in der Nähe des Medians der Stichprobe liegen. Wir betrachten drei Statistiken: die degree distribution (Verteilung der Grade), edgewise shared partner distribution (Verteilung der Kantenweise gemeinsamen Partner) und die geodesic distance distribution (Verteilung eines Distanzmaßes) (vgl. Hunter, Goodreau, und Handcock 2008). Hunter, Goodreau und Handcock beschreiben die drei Statistiken genauer:

„The degree distribution for a network consists of the values $D_0/n, \dots, D_{n-1}/n$, where D_k/n equals the proportion of nodes that share edges with exactly k other nodes. The edgewise shared partner distribution consists of the values $EP_0/E, \dots, EP_{n-2}/E$, where E denotes the total number of edges and EP_k equals the number of edges whose endpoints both share edges with exactly k other nodes. [...] Finally, the geodesic distance distribution consists of the relative frequencies of the possible values of geodesic distance between two nodes, where the geodesic distance between two nodes equals the length of the shortest path joining those two nodes[...]“ (Hunter, Goodreau, und Handcock 2008).

Man sieht, dass das Model die minimum geodesic distance relativ gut, aber degree- und edge-wise shared partners Verteilung etwas schlechter einfängt.

Die edge-wise shared partner Verteilung wird trotz Inklusion des GWESP Terms für lokales Clustering und Dreiecksbildung etwas unterschätzt. Das Modell schätzt, dass die meisten Abgeordneten Co-Sponsoring-Beziehungen zu etwa 5 bis 11 anderen Abgeordneten haben, die auch mit demselben Abgeordneten eine Co-Sponsoring-Beziehung haben (friend of a friend). In den Originaldaten haben die Abgeordnete aber zwischen etwa 11 und 46 Co-Sponsoring-Beziehungen zu friends of friends (Abbildung 9).

3.2. Ergebnispräsentation

Equivalent zu (2), beschreibt das Model die logarithmierte Chance (log-odds), dass eine Kante existiert, gegeben dem Rest des Netzwerks als:

$$\text{logit}(Y_{ij} = 1) = \log \left[\frac{\mathbb{P}_\theta(Y_{ij} = 1 \mid \mathbf{Y}_{(-ij)} = \mathbf{y}_{(-ij)})}{\mathbb{P}_\theta(Y_{ij} = 0 \mid \mathbf{Y}_{(-ij)} = \mathbf{y}_{(-ij)})} \right] = \theta^T \delta s(\mathbf{y}) \quad (5)$$

Y_{ij} zeigt, wie in 2.3 definiert, das Vorhandensein oder Fehlen einer Beziehung zwischen einem Abgeordnetenpaar in \mathbf{Y} . $Y_{(-ij)}$ sind alle Elemente von \mathbf{Y} außer Y_{ij} und $\delta s(\mathbf{y})$ ist die Änderung in $s(\mathbf{y})$, wenn sich der Wert von Y_{ij} von 0 auf 1 ändert (vgl. Kolaczyk und Csárdi 2020: S. 96). Die geschätzten Koeffizienten des Modells sind in Tabelle 9 abgebildet.

Die Interpretation der geschätzten Parameter des ERGMs erfolgt analog zu GLMs. Wenn also ein geschätzter Koeffizient positiv ist, steigt die logarithmierte Chance (log-odds) einer Kante; ist er negativ, sinkt die log-odds einer Kante. Der edge Koeffizient im Modell gibt für ein Knotenpaar, das sich in allen Attributen unterscheidet und keine anderen Knoten als friend of a friend gemeinsam haben, die log-odds von -7.512 dafür, eine gemeinsame Kante zu bilden. Wenn sie irgendeine Anzahl friends of a friend gemein haben, und jeder der beiden Knoten in zumindest in einem Dreieck mit diesen Knoten ist, dann steigen die log-odds, dass die zwei Knoten eine Kante bilden auf -3.43 (-7.512 + 4.082).

In Tabelle 9 sind rechts die exponenzierten Werte der geschätzten Koeffizienten angegeben. Betrachten wir zuerst die exogenen Haupteffekte des Modells, die mit B gekennzeichnet sind. Alle Interpretationen sind als ceteris paribus zu verstehen, d.h. unter sonst gleichen Umständen, also während alle anderen Variablen fest sind. Im Vergleich zu parteilosen Abgeordneten haben Abgeordnete der Parteien Forza Italia – Il Popolo della Libertà, Lega Nord und Partito Democratico eine geringere Chance Co-Sponsorship-Beziehungen zu formen. Für Lega Nord und Partito Democratico ist die Chance fast halb so groß. Alle anderen Parteien haben im Vergleich zu parteilosen Abgeordneten eine höhere Chance auf Co-Sponsorship-Beziehungen.

Männliche Abgeordnete haben eine im Durchschnitt um den Faktor $\exp(-0.417) = 0.659$ geringere Chance als Frauen, Co-Sponsoring-Beziehungen einzugehen. Alle Altersgruppen außer die der 60 bis 69 jährigen haben im Vergleich der unter 40 jährigen eine höhere Chance auf Co-Sponsorship-Beziehungen, wobei die Faktoren für alle Altersgruppen nahe an 1 sind.

	Koeffizienten und (SE)	exp(θ)
A: edges	-7.512 (0.364)***	0.001
A: gwesp.fixed.0.2	4.082 (0.283)***	59.266
B: party FI-PDL	-0.168 (0.042)***	0.846
B: party FLI-TP	0.522 (0.048)***	1.686
B: party IDV	0.472 (0.048)***	1.603
B: party LN	-0.765 (0.049)***	0.466
B: party PD	-0.556 (0.042)***	0.573
B: party PT	0.307 (0.078)***	1.359
B: party UDC-TP	0.171 (0.047)***	1.187
B: sex Male	-0.417 (0.017)***	0.659
B: age group >= 70	0.142 (0.041)***	1.153
B: age group 40 - 49	0.007 (0.024)	1.007
B: age group 50 - 59	0.106 (0.024)***	1.111
B: age group 60 - 69	-0.051 (0.025)*	0.950
C: match party IND	1.111 (0.117)***	3.036
C: match party FI-PDL	1.820 (0.042)***	6.170
C: match party FLI-TP	2.057 (0.140)***	7.822
C: match party IDV	3.403 (0.160)***	30.047
C: match party LN	5.028 (0.079)***	152.632
C: match party PD	3.175 (0.045)***	23.918
C: match party PT	2.227 (0.520)***	9.272
C: match party UDC-TP	2.433 (0.091)***	11.389
C: match sex	0.190 (0.021)***	1.209
C: match age group <40	0.184 (0.067)**	1.202
C: match age group >= 70	0.842 (0.274)**	2.321
C: match age group 40 - 49	0.042 (0.038)	1.043
C: match age group 50 - 59	0.033 (0.034)	1.034
C: match age group 60 - 69	-0.010 (0.062)	0.990
AIC	115872.702	
BIC	116161.071	
Log Likelihood	-57908.351	

*** p < 0.001; ** p < 0.01; * p < 0.05

Tabelle 9: ERGM mit Termen für endogene Netzwerkeigenschaften (A), für Knoteneigenschaften (B) und für Homophilie (C); Das Modell wurde in R mit dem statnet package (Krivitsky et al. 2003) erstellt, die Ergebnistabelle mithilfe des texreg Packages (Leifeld 2013)

Die Terme für Homophilie sind durchgehend positiv, bis auf den Term der differenziellen Homophilie für die Altersgruppe 60 bis 69. Abgeordnete dieser Altersgruppe haben also eine geringere Chance, eine Co-Sponsorship-Beziehung zu Abgeordneten der selben Altersgruppe aufzubauen. Der Koeffizient für uniforme Homophilie für das Geschlecht ist positiv und gibt an, dass Abgeordnete durchschnittlich eine um den Faktor $\exp(0.19) = 1.209$ höhere Chance haben, Co-Sponsorship-Beziehungen zu Abgeordneten des selben Geschlechts zu formen, als zu Abgeordneten anderen Geschlechts.

Die größten Koeffizienten sind die differenziellen Homophilie-Terme der Parteizugehörigkeit. Gerade die Chance für Abgeordnete der Partei Liga Nord, Co-Sponsorship-Beziehungen innerhalb der Partei einzugehen, ist durchschnittlich ca. um den Faktor 152 höher als für Beziehungen zu Abgeordneten anderer Parteien.

Der positive Koeffizient für den netzwerkendogenen Term GWESP gibt eine erhöhte Chance für eine Kante zwischen zwei Abgeordneten, die einen friend of a friend gemeinsam haben gegenüber einem Paar Abgeordneter, die das nicht haben, an.

4. Resümee

Wegen der relationalen Natur von Netzwerken sind ERGMs besser geeignet als klassische Regressionsmodelle um Netzwerkdaten zu modellieren. Das klassische Regressionsframework ist nicht in der Lage, Netzwerkabhängigkeiten zu modellieren. Mit ERGMs können netzwerkexogene sowie netzwerkendogene Eigenschaften geschätzt werden.

Die Forschungsfrage, Homophilie in Co-Sponsoring-Beziehungen im italienischen Parlament zu untersuchen, wurde mithilfe eines ERGMs bearbeitet. Effekte der Homophilie auf die Bildung von Co-Sponsorship-Beziehungen wurden für die Attribute der Altersklasse, des Geschlechts und der Parteizugehörigkeit der Abgeordneten festgestellt. Trotzdem ist es so, dass Bindungen im Netzwerk nicht nur abhängig von Homophilie in Charakteristika der Abgeordneten sind, sondern auch mit netzwerkendogenen Termen erklärt werden sollten, was der geschätzte GWESP Term für friend of a friend Bindungen, oder der Neigung zu Dreiecksbildung, zeigt.

Literaturverzeichnis

- Alemán, Eduardo, und Ernesto Calvo. 2013. „Explaining Policy Ties in Presidential Congresses: A Network Analysis of Bill Initiation Data“. *Political Studies* 61 (2): 356–77. <https://doi.org/10.1111/j.1467-9248.2012.00964.x>.
- Baller, Inger. 2017. „Specialists, Party Members, or National Representatives: Patterns in Co-Sponsorship of Amendments in the European Parliament“. *European Union Politics* 18 (3): 469–90. <https://doi.org/10.1177/1465116517712049>.
- Bratton, Kathleen A., und Stella M. Rouse. 2011. „Networks in the Legislative Arena: How Group Dynamics Affect Cosponsorship“. *Legislative Studies Quarterly* 36 (3): 423–60. <https://www.jstor.org/stable/23053249>.
- „Briatte - Summary Statistics“. 2016. GitHub. 2016. <https://github.com/briatte/parlnet/blob/master/parlnet.csv>.
- Briatte, François. 2016. „Network Patterns of Legislative Collaboration in Twenty Parliaments“. *Network Science* 4 (2): 266–71. <https://doi.org/10.1017/nws.2015.31>.
- Brunner, Martin. 2013. *Parliaments and Legislative Activity: Motivations for Bill Introduction*. Wiesbaden: Springer Fachmedien. <https://doi.org/10.1007/978-3-531-19612-1>.
- Campbell, James E. 1982. „Cosponsoring Legislation in the U. S. Congress“. *Legislative Studies Quarterly* 7 (3): 415–22. <https://doi.org/10.2307/439366>.
- Clark, Jennifer Hayes, und Veronica Caro. 2013. „Multimember Districts and the Substantive Representation of Women: An Analysis of Legislative Cosponsorship Networks“. *Politics & Gender* 9 (01): 1–30. <https://doi.org/10.1017/S1743923X12000670>.
- Cranmer, Skyler J., und Bruce A. Desmarais. 2011. „Inferential Network Analysis with Exponential Random Graph Models“. *Political Analysis* 19 (1): 66–86. <https://doi.org/10.1093/pan/mpq037>.
- Cranmer, Skyler J., Bruce A. Desmarais, und Jason W. Morgan. 2020. *Inferential Network Analysis*. 1. Aufl. Cambridge University Press. <https://doi.org/10.1017/9781316662915>.
- Erdős, P., und A. Rényi. 1959. „On Random Graphs I“. *Publicationes Mathematicae Debrecen*, Nr. 6, 290–97.
- Frank, O. 1991. „Statistical Analysis of Change in Networks“. *Statistica Neerlandica* 45 (3): 283–93. <https://doi.org/10.1111/j.1467-9574.1991.tb01310.x>.
- Frank, Ove, und David Strauss. 1986. „Markov graphs“. *Journal of the American Statistical Association* 81 (395): 832–42.
- Gilbert, E. N. 1959. „Random Graphs“. *The Annals of Mathematical Statistics* 30 (4): 1141–44. <https://doi.org/10.1214/aoms/1177706098>.
- Goodreau, Steven M., Mark S. Handcock, David R. Hunter, Carter T. Butts, und Martina Morris. 2008. „A Statnet Tutorial“. *Journal of Statistical Software* 24 (Mai): 1–26. <https://doi.org/10.18637/jss.v024.i09>.
- Goodreau, Steven M., James A. Kitts, und Martina Morris. 2009. „Birds of a Feather, or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks“. *Demography* 46 (1): 103–25. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2831261/>.
- Grasse, Alexander. 2012. „Die Jagd auf den Zins oder: die Fiskalpolitik der Regierung Monti und ihre Folgen“. *Wirtschaftsdienst* 92 (11): 751–58. <https://doi.org/10.1007/s10273-012-1450-8>.

- Haim, Mario. 2023. *Computational Communication Science: Eine Einführung*. Studienbücher zur Kommunikations- und Medienwissenschaft. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-40171-9>.
- Hunter, David R. 2007. „Curved exponential family models for social networks“. *Social Networks*, Special Section: Advances in Exponential Random Graph (p*) Models, 29 (2): 216–30. <https://doi.org/10.1016/j.socnet.2006.08.005>.
- . 2019. „A Statistician’s View of Network Modeling“. In *Network Science: An Aerial View*, herausgegeben von Francesca Biagini, Göran Kauermann, und Thilo Meyer-Brandis, 23–41. Cham: Springer International Publishing.
- Hunter, David R., Steven M. Goodreau, und Mark S. Handcock. 2008. „Goodness of Fit of Social Network Model“. *Journal of the American Statistical Association* 103 (481): 248–58. <https://www.jstor.org/stable/27640035>.
- Hunter, David R., und Mark S. Handcock. 2006. „Inference in Curved Exponential Family Models for Networks“. *Journal of Computational and Graphical Statistics* 15 (3): 565–83. <https://www.jstor.org/stable/27594198>.
- Kessler, Daniel, und Keith Krehbiel. 1996. „Dynamics of Cosponsorship“. *The American Political Science Review* 90 (3): 555–66. <https://doi.org/10.2307/2082608>.
- Kolaczyk, Eric D., und Gábor Csárdi. 2020. *Statistical Analysis of Network Data with R. Use R!* Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-44129-6>.
- Krivitsky, Pavel N., Mark S. Handcock, David R. Hunter, Carter T. Butts, Michał Bojanowski, Chad Klumb, Steven M. Goodreau, und Martina Morris. 2003. „Statnet: Tools for the Statistical Modeling of Network Data“. Statnet Development Team. <https://statnet.org>.
- Leifeld, Philip. 2013. „texreg: Conversion of Statistical Model Output in R to LATEX and HTML Tables“. <https://doi.org/10.18637/jss.v055.i08>.
- McPherson, Miller, Lynn Smith-Lovin, und James M Cook. 2001. „Birds of a Feather: Homophily in Social Networks“. *Annual Review of Sociology* 27 (1): 415–44. <https://doi.org/10.1146/annurev.soc.27.1.415>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Snijders, Tom A B. 2002. „Markov Chain Monte Carlo Estimation of Exponential Random Graph Models“. *Journal of Social Structure* (3): k.A.
- Snijders, Tom A. B., Philippa E. Pattison, Garry L. Robins, und Mark S. Handcock. 2006. „New specifications for exponential random graph models“. *Sociological Methodology* 36 (1): 99–153. <https://doi.org/10.1111/j.1467-9531.2006.00176.x>.
- Stanford Human Evolutionary Ecology and Health. 2018. „Introduction to ERGMs“. *Social Network Analysis for Anthropologists*. April 2018. <https://eehh-stanford.github.io/SNA-workshop/ergm-intro.html>.
- Wasserman, Stanley, und Philippa Pattison. 1996. „Logit Models and Logistic Regressions for Social Networks: I. An Introduction to Markov Graphs Andp“. *Psychometrika* 61 (3): 401–25. <https://doi.org/10.1007/BF02294547>.

Appendix

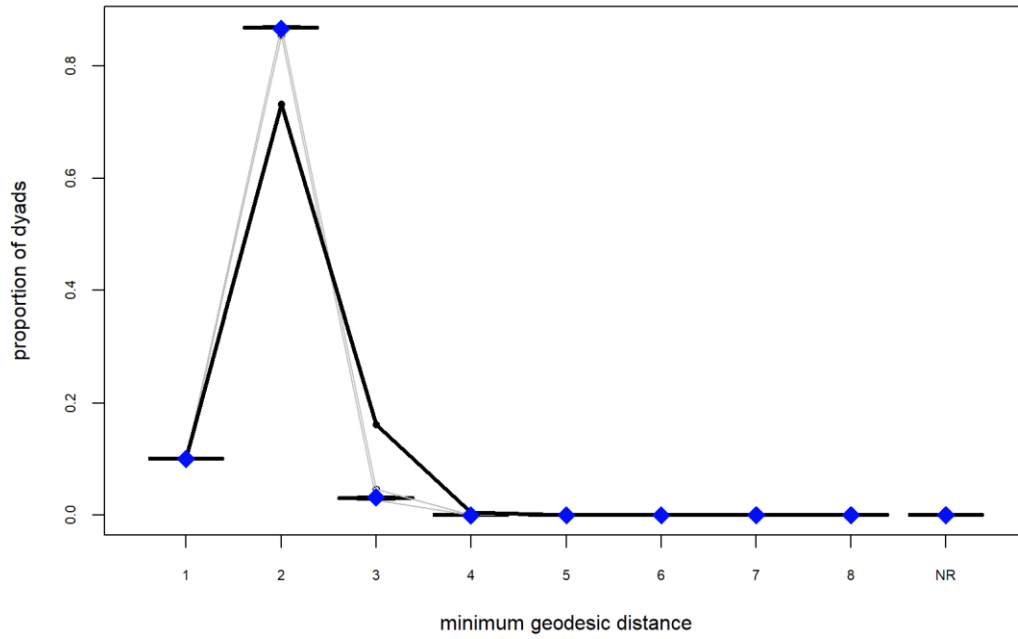


Abbildung 8: GOF Plot minimum geodesic distance distribution (erstellt mit *gof* Funktion aus der *statnet* package Umgebung (Krivitsky et al. 2003))

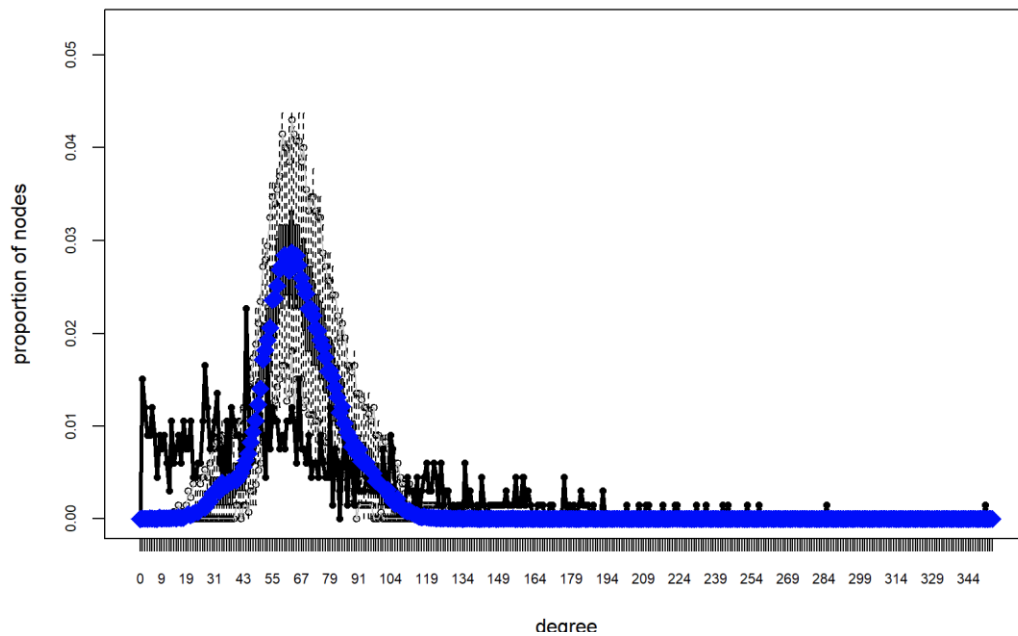


Abbildung 7: GOF Plot der degree distribution (erstellt mit *gof* Funktion aus der *statnet* package Umgebung (Krivitsky et al. 2003))

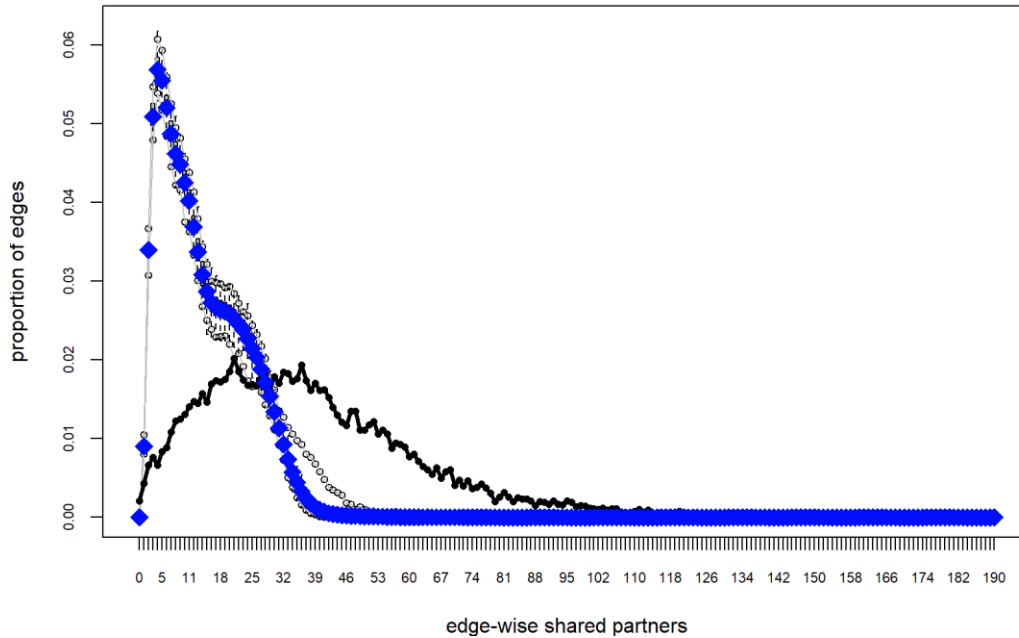


Abbildung 9: GOF Plot der edge-wise shared partners distribution (erstellt mit *gof* Funktion aus der *statnet* package Umgebung (Krivitsky et al. 2003))

```

Sample statistics auto-correlation:
Chain 1
edges gwesp.fixed.0.2 nodefactor.party_relevel.FI-PDL nodefactor.party_relevel.FLI-TP nodefactor.party_relevel.IDV
Lag 0 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000
Lag 131072 0.21308767 0.21205263 0.13235068 0.14225501 0.178370864
Lag 262144 0.05250879 0.05153318 0.01321949 0.06084622 0.099796454
Lag 393216 0.05188754 0.05123173 0.04772191 0.04192211 0.058015565
Lag 524288 0.04319483 0.04273856 0.03176882 0.01538848 0.029519079
Lag 655360 -0.00242964 -0.00201419 -0.01941506 0.01271344 0.002800665

nodefactor.party_relevel.LN nodefactor.party_relevel.PD nodefactor.party_relevel.PT nodefactor.party_relevel.UDC-TP nodefactor.sex.M
Lag 0 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000
Lag 131072 0.44800891 0.342965808 0.08138431 0.21249938 0.211026337
Lag 262144 0.23380249 0.134368849 0.03201743 0.09936711 0.032307190
Lag 393216 0.17667658 0.048590883 0.02535742 0.02177533 0.031730889
Lag 524288 0.08297173 0.022287438 0.01412478 0.04363270 0.036382684
Lag 655360 0.03404547 0.004369169 -0.01150247 0.02338075 -0.006750557

nodefactor.ageGroup.>= 70 nodefactor.ageGroup.40 - 49 nodefactor.ageGroup.50 - 59 nodefactor.ageGroup.60 - 69
Lag 0 1.00000000 1.00000000 1.00000000 1.00000000
Lag 131072 0.21841583 0.24233645 0.2264847512 0.18987882
Lag 262144 0.08111930 0.09619946 0.0677177752 0.06029356
Lag 393216 0.03204682 0.06473335 0.0430194483 0.03609339
Lag 524288 0.04634548 0.03070133 0.0343992555 0.04250244
Lag 655360 0.01801388 0.02394116 -0.0005176009 0.02046584

nodematch.party_relevel.AAIND nodematch.party_relevel.FI-PDL nodematch.party_relevel.FLI-TP nodematch.party_relevel.IDV
Lag 0 1.00000000 1.00000000 1.00000000 1.00000000
Lag 131072 0.390705011 0.1680685570 0.55671122 0.6979414
Lag 262144 0.136381065 0.0334443671 0.31133395 0.4777797
Lag 393216 0.052309101 0.0471251463 0.17300784 0.3084610
Lag 524288 0.007977164 0.0382996560 0.08595966 0.2093497
Lag 655360 -0.006252509 -0.0005056244 0.04803110 0.1511253

nodematch.party_relevel.LN nodematch.party_relevel.PD nodematch.party_relevel.PT nodematch.party_relevel.UDC-TP nodematch.sex
Lag 0 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000
Lag 131072 0.60809018 0.3646355074 0.57794993 0.50319123 0.220613484
Lag 262144 0.36847380 0.1484830877 0.31771494 0.25365920 0.055182596
Lag 393216 0.22475084 0.0378869781 0.18522166 0.13314905 0.018070546
Lag 524288 0.12536776 0.0001045162 0.09285680 0.07950509 0.040458595
Lag 655360 0.06717369 -0.0107694702 0.04295858 0.06236780 0.004565216

nodematch.ageGroup.<40 nodematch.ageGroup.>= 70 nodematch.ageGroup.40 - 49 nodematch.ageGroup.50 - 59 nodematch.ageGroup.60 - 69
Lag 0 1.00000000 1.00000000 1.00000000 1.00000000 1.00000000
Lag 131072 0.29434111 0.298888408 0.26721306 0.24237834 0.186470578
Lag 262144 0.11442715 0.116194130 0.10669067 0.07273947 0.089776684
Lag 393216 0.05307485 0.046917147 0.06445670 0.02596951 0.021670948
Lag 524288 0.03286386 0.009481107 0.02968200 0.01814607 -0.006163310
Lag 655360 -0.01392276 0.007710520 0.02019779 -0.01491231 -0.001150067

```

Abbildung 10: MCMC Diagnostics sample statistics auto-correlation (erstellt mit `mcmc.diagnostics` aus der Statnet package Umgebung (Krivitsky et al. 2003))

```

Sample statistics burn-in diagnostic (Geweke):
Chain 1

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

edges
0.78035103
nodefactor.party_relevel.IDV 0.43935725
nodefactor.party_relevel.UDC-TP 0.96994528
nodefactor.ageGroup.50 - 59 1.37821710
nodematch.party_relevel.FLI-TP -0.38641917
nodematch.party_relevel.PT 0.16386899
nodematch.ageGroup.>= 70 -0.87605889

gwpes.fixed.0.2 0.78014208
nodefactor.party_relevel.LN 0.67117872
nodefactor.sex.M 0.39158619
nodefactor.ageGroup.60 - 69 -0.62918219
nodematch.party_relevel.IDV -0.52075839
nodematch.party_relevel.UDC-TP -0.26438591
nodematch.ageGroup.40 - 49 -0.79914742

nodefactor.party_relevel.FI-PDL -1.05283045
nodefactor.party_relevel.PD 0.94493710
nodefactor.ageGroup.>= 70 -2.11616679
nodematch.party_relevel.AAIND -0.27316921
nodematch.party_relevel.LN -0.16970246
nodematch.sex -0.03155281
nodematch.ageGroup.50 - 59 0.20100948

nodefactor.party_relevel.FLI-TP 0.48244357
nodefactor.party_relevel.PT -1.04186954
nodefactor.ageGroup.40 - 49 -0.33105413
nodematch.party_relevel.FI-PDL -1.77602940
nodematch.party_relevel.PD 0.60921886
nodematch.ageGroup.<40 0.66117279
nodematch.ageGroup.60 - 69 0.77938716

Individual P-values (lower = worse):
edges
0.43518428
nodefactor.party_relevel.IDV 0.66040269
nodefactor.party_relevel.UDC-TP 0.33207377
nodefactor.ageGroup.50 - 59 0.16813627
nodematch.party_relevel.FLI-TP 0.69918626
nodematch.party_relevel.PT 0.86983428
nodematch.ageGroup.>= 70 0.38099802

gwpes.fixed.0.2 0.43530725
nodefactor.party_relevel.LN 0.50210668
nodefactor.sex.M 0.69536400
nodefactor.ageGroup.60 - 69 0.52922978
nodematch.party_relevel.IDV 0.60253509
nodematch.party_relevel.UDC-TP 0.79148257
nodematch.ageGroup.40 - 49 0.42420493

nodefactor.party_relevel.FI-PDL 0.29241871
nodefactor.party_relevel.PD 0.34469099
nodefactor.ageGroup.>= 70 0.03433062
nodematch.party_relevel.AAIND 0.78472315
nodematch.party_relevel.LN 0.86524414
nodematch.sex 0.97482868
nodematch.ageGroup.50 - 59 0.84069116

nodefactor.party_relevel.FLI-TP 0.62949087
nodefactor.party_relevel.PT 0.29747217
nodefactor.ageGroup.40 - 49 0.74060360
nodematch.party_relevel.FI-PDL 0.07572807
nodematch.party_relevel.PD 0.54237938
nodematch.ageGroup.<40 0.50850151
nodematch.ageGroup.60 - 69 0.43575169

Joint P-value (lower = worse): 0.5828349

```

Abbildung 11: MCMC Diagnostics sample statistics burn-in diagnostics (Geweke) (erstellt mit `mcmc.diagnostics` aus der Statnet package Umgebung (Krivitsky et al. 2003))

Sample statistics

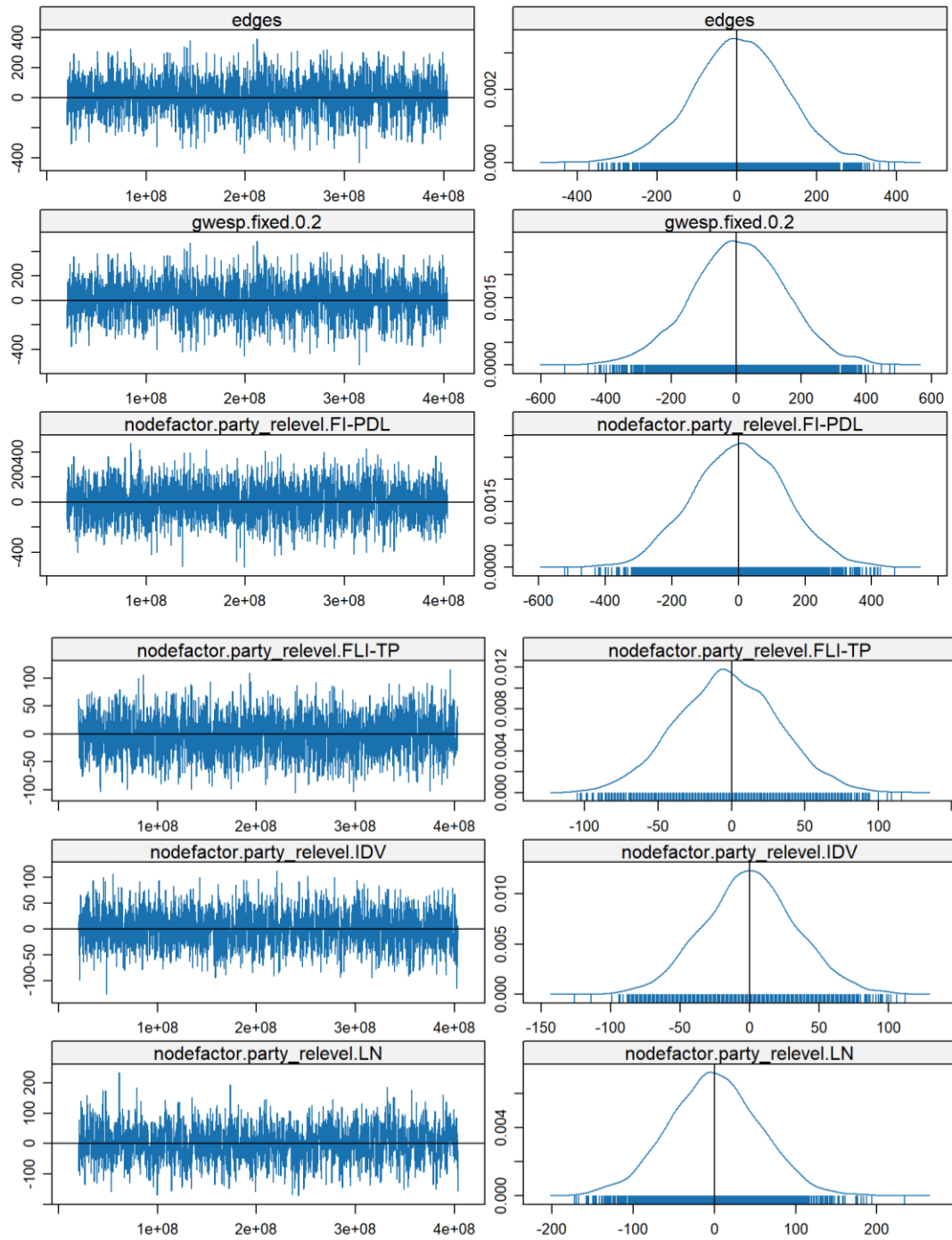


Abbildung 12: MCMC Diagnose für das Modell; erstellt in R mit `mcmc.diagnostics` aus dem `statnet` package (Krivitsky et al. 2003)

Sample statistics

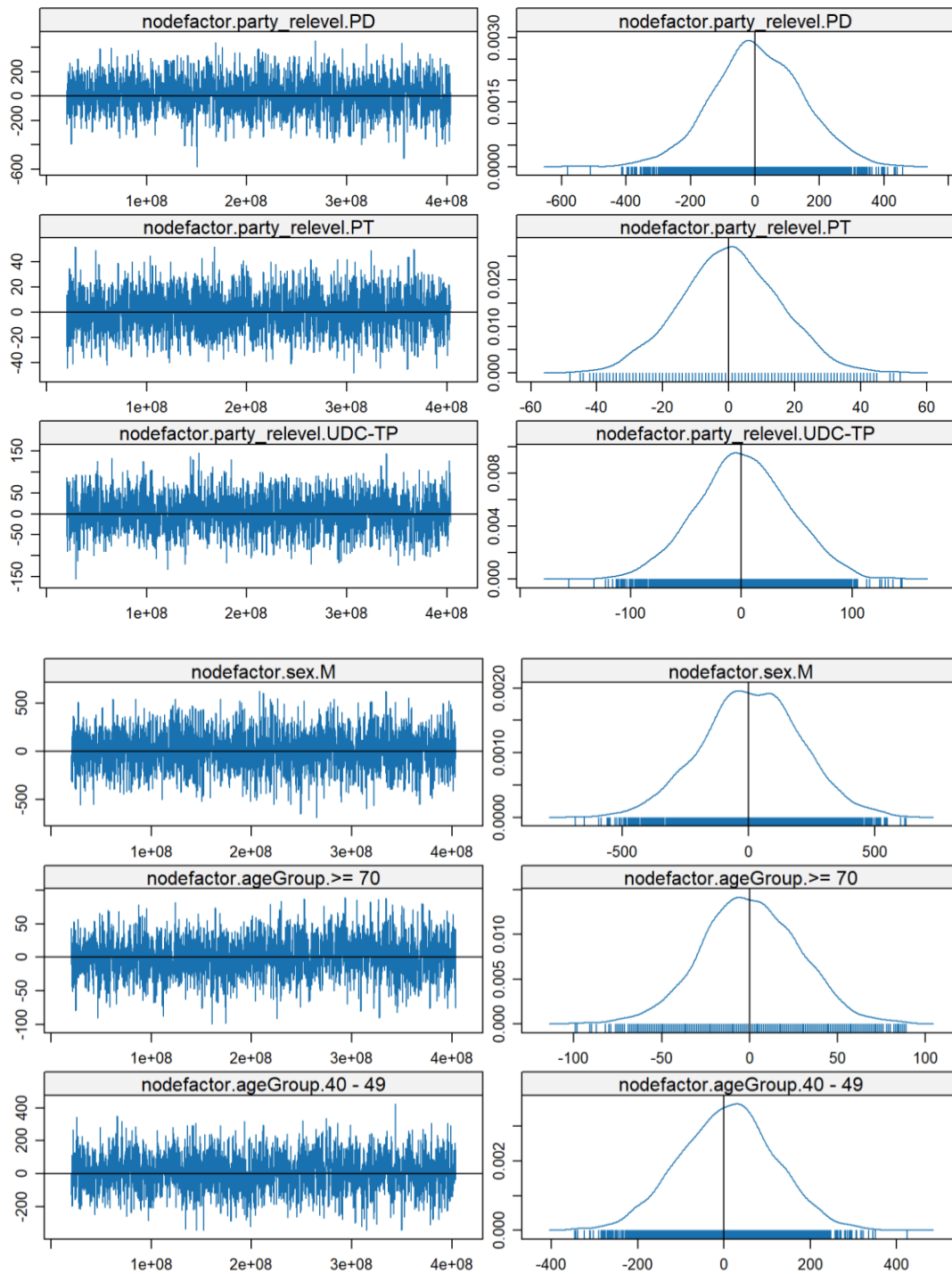


Abbildung 13: MCMC Diagnose für das Modell; erstellt in R mit `mcmc.diagnostics` aus dem `statnet` package (Krivitsky et al. 2003)

Sample statistics

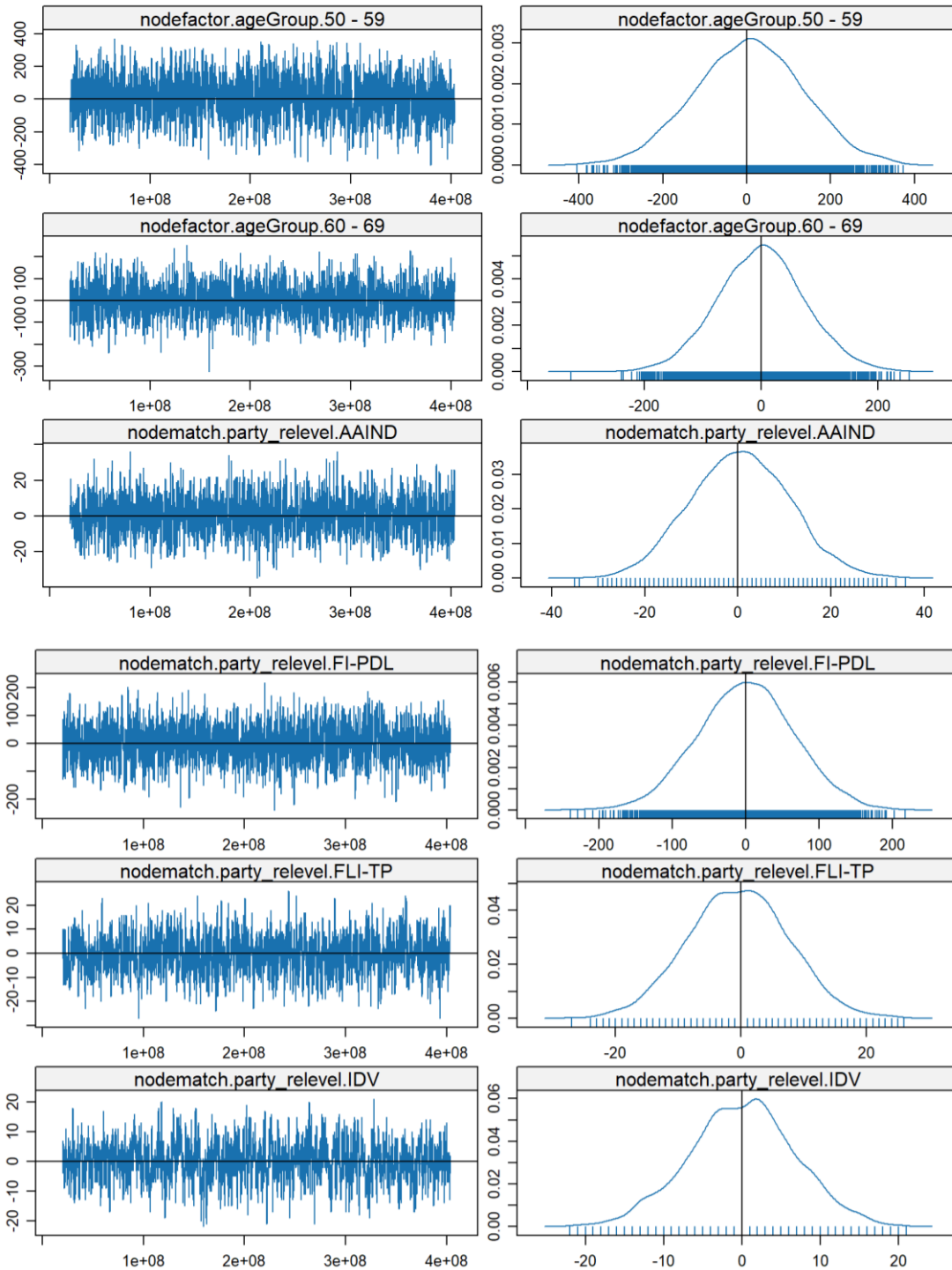
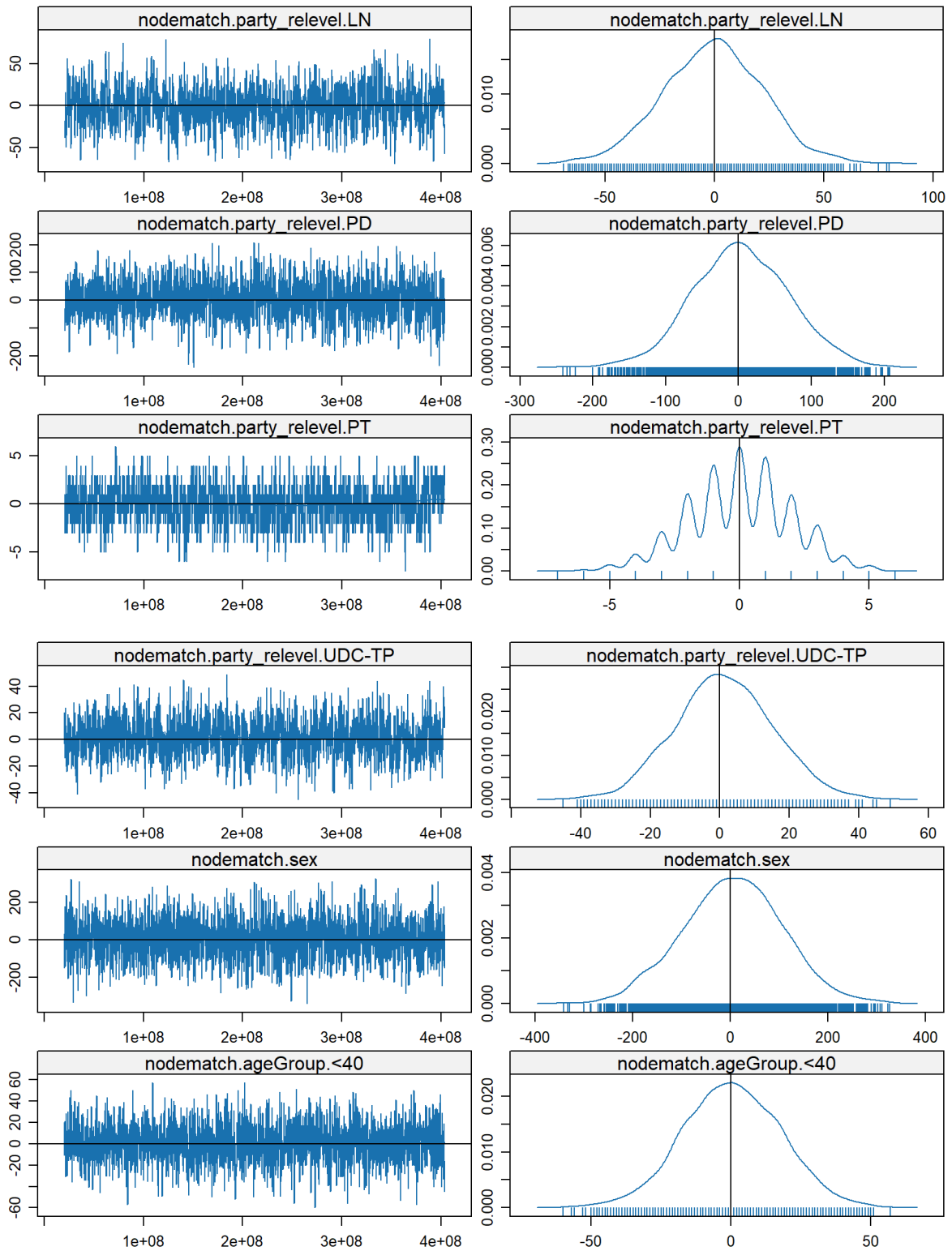


Abbildung 14: MCMC Diagnose für das Modell; erstellt in R mit `mcmc.diagnostics` aus dem `statnet` package (Krivitsky et al. 2003)

Sample statistics



2003)Anmerkung: Man sieht im Dichteplot für `Nodematch.party_relevel.PT` ein Sägeblattmuster, das kann bei diskreten Variablen mit kleiner Spannweite passieren, ist aber kein Problem, solange die Gesamtform weiterhin grob Normalverteilt und um 0 zentriert ist.

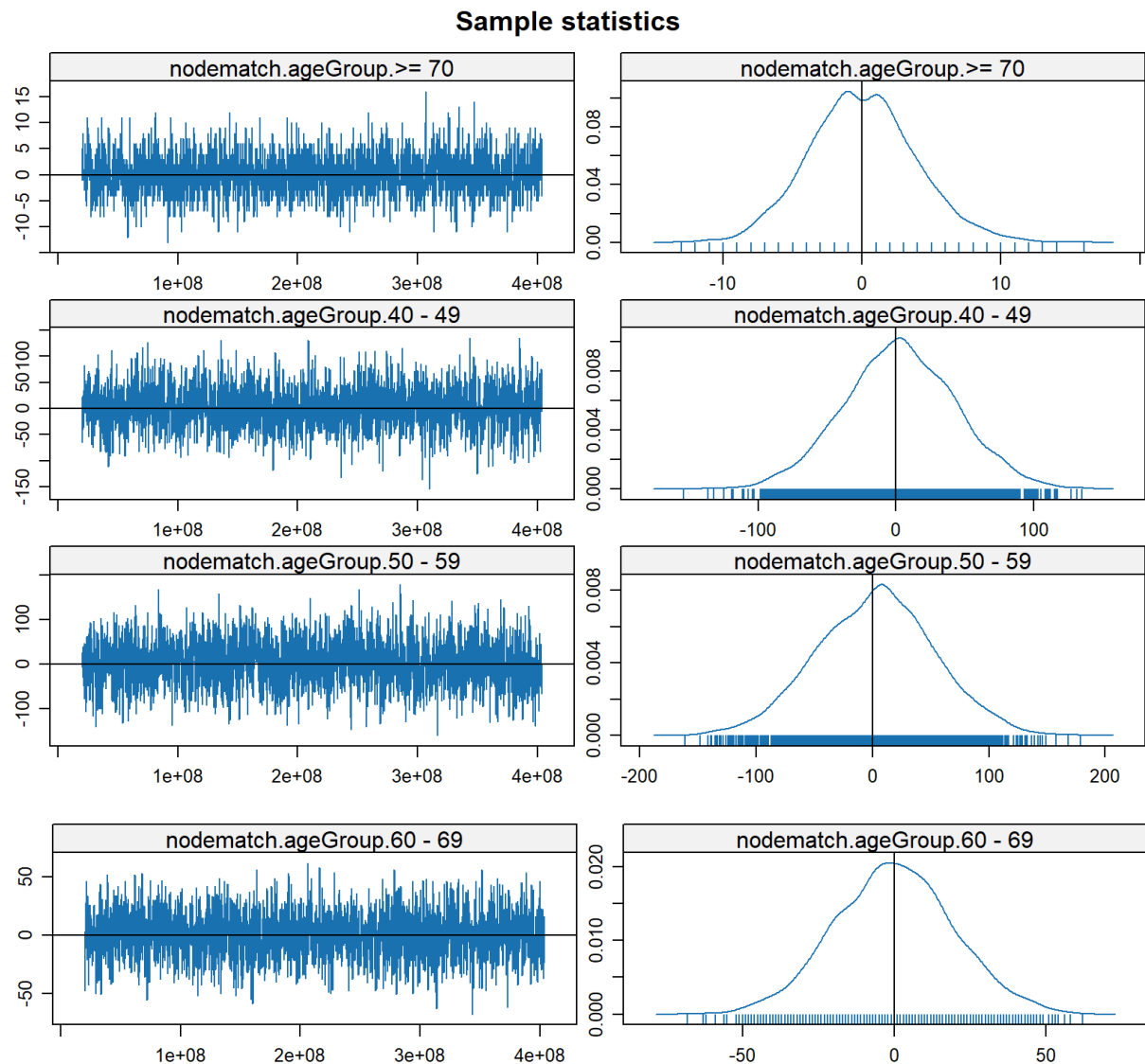


Abbildung 16: MCMC Diagnose für das Modell; erstellt in R mit `mcmc.diagnostics` aus dem `statnet` package (Krivitsky et al. 2003)

Urheberschaftserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit eigenständig und ohne fremde Hilfe angefertigt habe. Textpassagen, die wörtlich oder dem Sinn nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche kenntlich gemacht.

Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

München, Datum

Name (+ Unterschrift)