



UNIVERSITÄT  
KOBLENZ · LANDAU

Fachbereich IV  
Institute for Web Science and Technologies

**Masterarbeit**

Zur Erlangung des akademischen Grades Master of Science Informatik

Titel:

**Methoden zur Messung von Polarisierung in einer Liquid  
Democracy**

Vorgelegt am 25.11.2015

von

Manuel Mittler

Matrikel-Nummer 213202876

mail@m-mittler.de

**Betreuer** Dipl.-Inf. Christoph C. Kling

**Prüfer** Prof. Dr. Markus Strohmaier

# Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, dass alle Stellen der Arbeit, die wörtlich oder sinngemäß aus anderen Quellen übernommen wurden, als solche kenntlich gemacht und dass die Arbeit in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegt wurde.

Ort, Datum

Unterschrift

# Zusammenfassung

Politische und gesellschaftliche Polarisierung ist ein interessantes Phänomen, über dessen Auswirkungen viele unterschiedliche, zum Teil auch gegensätzliche, Theorien existieren. Polarisierung wird in der Literatur mit unterschiedlichen Methoden gemessen. Die vorliegende Arbeit gibt einen Überblick über existierende Polarisierungsmaße und es werden zwei neuartige Maße aus dem Gebiet der spektralen Graphentheorie vorgestellt. Anschließend werden die bekannten und die neu entwickelten Maße auf den LiquidFeedback-Datensatz der Piratenpartei Deutschland angewandt. Als Ergebnis lässt sich festhalten, dass die Maße teilweise zu unterschiedlichen Ergebnisse kommen. Dies liegt darin begründet, dass nicht alle Maße das Gleiche messen. Um zu verstehen was die einzelnen Maße aussagen, werden wesentliche Eigenschaften von Polarisierungsmaßen herausgearbeitet und es wird für jedes Maß dargelegt, welche Eigenschaften es erfüllt. Die angesprochenen Polarisierungsmaße beziehen sich auf die Entwicklung der Polarisierung zwischen Usern des LiquidFeedback-Systems. Bei der Betrachtung von einzelnen Personen und Abstimmungen fiel unter anderem auf, dass polarisierende Personen mehr Macht durch Delegationen besitzen als die restlichen Personen und dass polarisierte Vorschläge circa doppelt so häufig umgesetzt werden.

# Inhaltsverzeichnis

|   |    |
|---|----|
| Abbildungsverzeichnis   | V  |
| Tabellenverzeichnis   | VI |
| Symbolverzeichnis   | 1  |
| 1 Einleitung  | 2  |
| 1.1 Motivation . . . . .  | 4  |
| 1.2 Was ist Polarisierung? . . . . .  | 5  |
| 1.3 Liquid Democracy und LiquidFeedback . . . . .                                   | 6  |
| 1.4 Eingrenzung des Untersuchungsgegenstandes . . . . .                             | 9  |
| 2 Related Work  | 10 |
| 3 Methodik  | 12 |
| 3.1 Formen von Polarisierung . . . . .  | 12 |
| 3.2 Algebraische und netzwerktheoretische Grundlagen . . . . .                      | 14 |
| 3.3 Messung von Polarisierung . . . . .   | 18 |
| 3.3.1 Messung von Polarisierung auf Abstimmungsebene . . . . .                      | 18 |
| 3.3.2 Messung von Polarisierung auf Netzwerkebene . . . . .                         | 20 |
| 3.4 Diskussion . . . . .  | 24 |
| 3.4.1 Krippendorffs Alpha als Polarisierungsmaß . . . . .                           | 24 |
| 3.4.2 Polarisierungsmaß gemäß Fläche und Macy . . . . .                             | 24 |
| 3.4.3 Korrelationskoeffizient zwischen Abstimmungen als Polarisierungsmaß . . . . . | 27 |
| 3.4.4 Gruppenähnlichkeit als Polarisierungsmaß . . . . .                            | 28 |
| 3.4.5 Anteil frustrierter Kanten als Polarisierungsmaß . . . . .                    | 29 |

## *Inhaltsverzeichnis*

|       |   |    |
|-------|---|----|
| 3.4.6 | Korrelationen der vorgestellten Maße . . . . .  | 29 |
| 4     | Verwendete Datensätze   | 31 |
| 5     | Polarisierung auf der Mikroebene - polarisierte Initiativen und polarisierende User           | 32 |
| 5.1   | Polarisierende Texte . . . . .  | 33 |
| 5.2   | Polarisierende Personen, Initiativen und Verbesserungsvorschläge . . . . .                    | 34 |
| 5.3   | Polarisierung und Popularität . . . . .   | 34 |
| 5.4   | Polarisierung und Macht . . . . .   | 35 |
| 6     | Polarisierung auf der Makroebene - Polarisation als Prozess über die Zeit                     | 37 |
| 6.1   | Entwicklung von Krippendorffs Alpha im LiquidFeedback-Datensatz . . . . .                     | 37 |
| 6.2   | Entwicklung des Polarisierungsmaßes von Flache und Macy im LiquidFeedback-Datensatz . . . . . | 40 |
| 6.3   | Entwicklung der Abstimmungskorrelationen im LiquidFeedback-Datensatz                          | 42 |
| 6.4   | Entwicklung der Clusterähnlichkeiten im LiquidFeedback-Datensatz . . . . .                    | 44 |
| 6.5   | Entwicklung des Anteils frustrierter Kanten im LiquidFeedback-Datensatz .                     | 47 |
| 7     | Zusammenfassung   | 51 |
| A     | Anhang  | 53 |
| A.1   | Quelltexte . . . . .  | 53 |
|       | Literaturverzeichnis  | 62 |

# Abbildungsverzeichnis

|      |  |    |
|------|--|----|
| 1.1  | LiquidFeedback Antrags-Prozess . . . . .   | 8  |
| 3.1  | Abstimmungsszenarien . . . . .   | 13 |
| 3.2  | Structural Balance . . . . .   | 17 |
| 3.3  | Übereinstimmung und Lagerbildung bei unvollständigen Daten . . . . .             | 26 |
| 3.4  | User-Übereinstimmung und Abstimmungskorrelation . . . . .                        | 27 |
| 3.5  | Gruppengröße und Abstimmungskorrelation . . . . .                                | 28 |
| 5.1  | Twitter-Follower-Netzwerk . . . . .  | 36 |
| 6.1  | Krippendorffs Alpha im Zeitverlauf . . . . .                                     | 38 |
| 6.2  | User-Übereinstimmung und Abstimmungskorrelation . . . . .                        | 39 |
| 6.3  | Perfekte Polarisierung zwischen zwei ungleich großen Lagern . . . . .            | 40 |
| 6.4  | Entwicklung des Polarisierungsmaße nach Fläche und Macy [FM11] . . . . .         | 41 |
| 6.5  | Issue Constraints . . . . .  | 42 |
| 6.6  | Entwicklung der durchschnittlichen positiven und negativen Korrelation . . . . . | 43 |
| 6.7  | Größen der beiden Cluster . . . . .  | 46 |
| 6.8  | Kosinus-Ähnlichkeit innerhalb und zwischen den Clustern . . . . .                | 47 |
| 6.9  | Polarisierungsmaß basierend auf Clusterähnlichkeiten . . . . .                   | 48 |
| 6.10 | Anteil frustrierter Kanten im Graph . . . . .                                    | 50 |

# Tabellenverzeichnis

|     |   |    |
|-----|---|----|
| 3.1 | Koinzidenz-Matrix . . . . .                         | 22 |
| 3.2 | Reliabilität-Matrix . . . . .                       | 22 |
| 3.3 | Eigenschaften der Polarisierungsmaße . . . . .      | 25 |
| 3.4 | Korrelationsmatrix der Polarisierungsmaße . . . . . | 30 |
| 5.1 | Textmerkmale und Regressionskoeffizienten . . . . . | 33 |

# Symbolverzeichnis

## Allgemeine Symbole

| Symbol       | Bedeutung                               |
|--------------|---|
| $a$          | der Skalar $a$                          |
| $\mathbf{A}$ | die Matrix $\mathbf{A}$                 |
| $\alpha$     | Krippendorffs Alpha                     |
| $\lambda$    | Eigenwert einer Matrix                  |
| $\vec{x}$    | der Vektor $\vec{x}$                    |
| $\omega$     | Wölbung                                 |
| $\xi$        | Algebraischer Konflikt                  |
| $P_{Flache}$ | Polarisierungsmaß nach Fläche und Macy  |
| $b_f(G)$     | Anteil frustrierter Kanten im Graph $G$ |



# 1 Einleitung

Polarisierung ist allgegenwärtig. Ob in der aktuellen Flüchtlingsdebatte [Mit] oder im amerikanischen Kongress.

„It’s official: The 112th Congress was the most polarized ever“

So lautete der Titel der Washington Post im Januar 2013. [Mat13] Zu diesem Ergebnis kamen die Politikwissenschaftler Keith Poole und Howard Rosenthal mittels der von ihnen entwickelten DW-NOMINATE-Methode. [Car09] Aufgrund der zunehmenden Polarisierung zwischen Republikanern und Demokraten entstehen im amerikanischen Kongress immer häufiger Blockaden, welche die Gesetzgebung erschweren. [Zei12] Dies wird durch das amerikanische Gesetzgebungsverfahren bedingt, welches vorsieht, dass Gesetzesinitiativen vom Kongress ausgehen und von diesem auch verabschiedet werden. Da seit der letzten Wahl im Jahr 2014 mehrheitlich Republikaner im Kongress sitzen und der Präsident ein Vetorecht besitzt, kann das Staatsoberhaupt vom Kongress beschlossene Gesetzesinitiativen stoppen. Negative Folgen einer solchen Blockade hätten die USA beinahe in Form der sogenannten „fiscal cliff“ zu spüren bekommen. [Sch12] Hinter diesem Begriff verbergen sich unter anderem Ausgabenkürzungen und Steuererhöhungen, welche automatisch in Kraft treten, sobald die Staatsverschuldung der USA einen vorgeschriebenen Höchstwert überschreitet. Diese Maßnahmen und die damit verbundenen negativen Auswirkungen auf die Wirtschaft konnten nur durch einen Kompromiss der beiden Parteien in wortwörtlich letzter Sekunde verhindert werden. Dieses Beispiel zeigt, dass eine starke Polarisierung in einer Demokratie zu einem Problem werden kann, wenn die Gegensätze so groß sind, dass Einigungen kaum noch möglich sind. Das Zweiparteiensystem der USA begünstigt das Zustandekommen von Gegensätzlichkeiten, da Politiker versuchen, sich klar von den Positionen der anderen Partei abzugrenzen.

Politische Polarisierung ist Forschungsgegenstand vieler Studien, die unterschiedliche Aspekte des Phänomens untersuchen. Auffällig bei den Studien ist, dass unterschiedliche Autoren teilweise zu völlig gegensätzlichen Ergebnissen kommen. Zum Beispiel wollen

## 1 Einleitung

Fiorina et al. [Fio10] herausgefunden haben, dass bei steigender Polarisierung die Wahlbeteiligung zurückgeht. Demgegenüber steht eine Untersuchung von Abramowitz und Saunders [Abr08], die zu dem Schluss kommt, dass Polarisierung politische Partizipation stimuliert. Der Dissens über die Auswirkungen von Polarisierung zeigt, dass es nicht trivial ist, valide Aussagen diesbezüglich zu treffen. Dies liegt mitunter daran, dass kein allgemein anerkanntes Maß für Polarisierung existiert. Methoden zur Quantifizierung von Polarisierung werden in Abschnitt 3.3 vorgestellt.

Mit dem Voranschreiten der technischen Entwicklung werden auch politische Vorgänge und Prozesse zunehmend digitalisiert. Die Piratenpartei ist eine der ersten Parteien, welche das Internet zur internen Meinungsbildung und Entscheidungsfindung nutzt. Hierzu ist die LiquidFeedback-Software <sup>1</sup> im Einsatz. Ihr liegt das Liquid-Democracy-Prinzip zugrunde. Eine detaillierte Beschreibung von Liquid Democracy und LiquidFeedback ist in Abschnitt 1.3 zu finden.

Dass neuartige Technologien und Algorithmen zur Implementierung künstlicher Intelligenz Polarisierung verstärken oder sogar hervorrufen können, davor warnt eine Gruppe von Wissenschaftlern in einem „Digitalen Manifest“ [Dir]. Durch personalisierte, gefilterte Informationen, wie zum Beispiel bei Facebook und Google, kann es zu dem sogenannten „Echokammereffekt“ kommen:

„Damit werden lokale Trends durch Wiederholung allmählich verstärkt [...] Am Ende bekommt man nur noch seine eigenen Meinungen widergespiegelt. Das bewirkt eine gesellschaftliche Polarisierung, also die Entstehung separater Gruppen, die sich gegenseitig nicht mehr verstehen und vermehrt miteinander in Konflikt geraten. So kann personalisierte Information den gesellschaftlichen Zusammenhalt unabsichtlich zerstören.“[Dir]

Die vorliegende Arbeit untersucht wie Polarisierung gemessen werden kann und wendet diese Polarisierungsmaße anschließend auf den LiquidFeedback-Datensatz der Piratenpartei Deutschland an.

---

1 <http://liquidfeedback.org/>

### 1.1 Motivation

Politische und gesellschaftliche Polarisierung ist interessant, da sie Ursache von Konflikten sein kann. Unterschiedliche Auffassungen und Meinungen gehören zu einer pluralistischen Gesellschaft dazu und der öffentlicher Diskurs ist essenzieller Bestandteil demokratischer Regierungsformen. Dennoch kann ein Zuviel an Gegensätzlichkeit schädlich für eine Gesellschaft sein, wenn dadurch das Funktionieren als Gemeinschaft nicht mehr möglich ist.

Durch den Einsatz von LiquidFeedback und dem Selbstverständnis der Piratenpartei eine transparente Politik zu betreiben, ist es möglich, neue Erkenntnisse bezüglich politischer Polarisierung zu gewinnen. Für die folgenden Auswertungen wurden die Datenbank-Snapshots des LiquidFeedback-Systems der Piratenpartei verwendet. Dies ist insofern interessant, als es nach dem Kenntnisstand des Autors zuvor keinen so umfangreichen Datensatz gab, der es ermöglicht, das Phänomen Polarisierung so umfassend zu analysieren. Da es trotz zahlreicher Studien umstritten ist, welche Auswirkungen Polarisierung auf Wahlbeteiligung, Abstimmverhalten etc. hat, wird in der vorliegenden Arbeit untersucht, ob Polarisierung zwischen den Mitgliedern der Piratenpartei erkennbar ist, wie sie sich im Laufe der Zeit entwickelt und welche Zusammenhänge bei Polarisierung im Rahmen einer Liquid Democracy auftreten. Eine erste Vermutung des Autors war, dass die zuletzt im System beobachtbare zurückgehende Aktivität mit zunehmender Polarisierung und einer erschwerten Konsensfindung einhergeht, was zu Frustration unter den Mitgliedern führt. Um hierüber eine Aussage machen zu können, mussten zuerst Methoden gefunden werden, um Polarisierung zu messen. Ein wesentlicher Bestandteil der Arbeit war es, geeignete Methoden zu finden, welche unterschiedliche Aspekte von Polarisierung betrachten und diese miteinander zu vergleichen. Kapitel 2 liefert einen Überblick der in der Literatur verwendeten Polarisierungsmaße. Diese dienten als Einstiegspunkt in das Thema und wurden teilweise auf dem LiquidFeedback-Datensatz angewendet. Jedoch waren manche Methoden aufgrund der Eigenschaften des untersuchten Datensatzes nicht direkt anwendbar und mussten modifiziert werden. Des Weiteren konnten mit diesen Maßen nicht alle Aspekte des Phänomens Polarisierung betrachtet werden, so dass noch zusätzliche Methoden entwickelt wurden.

## 1.2 Was ist Polarisierung?

Die meisten haben ein intuitives Verständnis des Begriffs Polarisierung. Da er jedoch in unterschiedlichen Kontexten benutzt wird und in der Literatur kaum Definitionen zu finden sind, soll dieses Kapitel das dieser Arbeit zugrunde liegende Verständnis des Begriffs verdeutlichen. Über die Bedeutung des Wortes steht im Duden Folgendes geschrieben:

„Aufspaltung (in zwei Lager o.Ä.), bei der die Gegensätze deutlich hervortreten.“[Bib]

Der Begriff Polarisierung wird häufig in der Soziologie im Zusammenhang von Einkommensunterschieden und sozialer Ungleichheit sowie in der Politik verwendet. Politische Polarisierung beschreibt den Dissens politischer Akteure zu einem oder mehreren Themen. Charakteristisch für Meinungspolarisierung ist, dass sich beide Lager bei ihren Positionen extrem unterscheiden und diese oftmals unvereinbar miteinander sind. Extreme Positionen alleine sind jedoch keine hinreichende Bedingung für Polarisierung. Wenn zum Beispiel alle Personen extreme Positionen einnehmen, jedoch jedes Mal per Münzwurf entscheiden würden auf welcher Seite sie stehen, kann man nicht von Polarisierung sprechen. Es muss demnach ein Muster bei den Abstimmungen erkennbar sein, so dass Personen desselben Lagers einheitlich und Personen unterschiedlicher Lager entgegengesetzt abstimmen. **Die wesentliche Eigenschaft von Polarisierung ist demnach das Vorhandensein von Personen mit ähnlichem Abstimmverhalten, d.h. ihre Meinungen stimmen bei vielen Fragestellungen überein, bei gleichzeitigem (extremen) Dissens zu den restlichen Abstimmenden. Das bedeutet, dass mindestens 2 Lager existieren, die als Gruppe betrachtet weitgehend konträr zueinander abstimmen.** Dieser Arbeit liegt die letztgenannte Definition von Polarisierung zugrunde.

Gemäß Dimaggio et al. [DiM96] kann Polarisierung sowohl Zustand als auch Prozess sein:

„Polarization is both a state and a process. Polarization as a state refers to the extent to which opinions on an issue are opposed in relation to some theoretical maximum. Polarization as a process refers to the increase in such opposition over time.“ [DiM96]

Aussagen über die Entwicklung von Polarisierung über einen Zeitraum zu treffen, ist trivial, da anhand der Zahlen ersichtlich ist, ob sich das Ausmaß an Polarisierung relativ gesehen verändert. Polarisierung in einem Zustand zu quantifizieren, ist dagegen nicht

## 1 Einleitung

ohne Weiteres möglich. Absolute Werte, ab wann zum Beispiel eine Abstimmung als polarisiert bezeichnet werden kann, existieren nicht. Zudem variieren diese je nach Szenario. Wesentliche Unterschiede sind:

- Anzahl der Personen/Abstimmungsteilnehmer
- Anzahl der Abstimmungen
- Art der Abstimmungsskala
- vollständige oder unvollständige Daten (Stimmt jede Person bei jeder Abstimmung ab?)
- Anzahl der Gruppen/Fraktionen

Untersuchungsgegenstand dieser Arbeit ist die Meinungspolarisierung zwischen einzelnen Personen und Gruppen innerhalb der Piratenpartei Deutschlands, basierend auf den abgegebenen Stimmen in LiquidFeedback. Dazu ist anzumerken, dass nicht alle Mitglieder der Partei das System nutzen.

### 1.3 Liquid Democracy und LiquidFeedback

Liquid Democracy kombiniert Elemente direkter und repräsentativer Demokratie und es überlässt dem Benutzer die Entscheidung, in welchem Umfang er direkt abstimmt, beziehungsweise seine Stimme delegiert. Der Begriff „Liquid“ (deutsch: flüssig) steht in diesem Zusammenhang dafür, dass Begrenzungen, die bei einer repräsentativen Demokratie existieren, „verflüssigt“, d.h. aufgelöst, werden. Der Verein Liquid Democracy sieht in diesem Ansatz die Möglichkeit, Begrenzungen zeitlicher (Wahlen nur alle 4 Jahre), inhaltlicher (Auswahl nur zwischen wenigen „Komplettpaketen“/Parteien) und partizipatorischer (Ausarbeitung von Gesetzen nur für politische Eliten) Art aufzuheben. [Liqc] Ein Liquid-Feedback sehr ähnliches System, mit dem Namen Adhocracy<sup>2</sup>[Liqa], wird u.a. von der SPD und der Linkspartei genutzt [Bun]. Adhocracy wird vom Verein Liquid Democracy e.V. [Liqb] entwickelt und betrieben. Die Vision und Motivation, die zur Gründung des Vereins führte, wird wie folgt beschrieben:

---

<sup>2</sup> <https://adhocracy.de/>

## 1 Einleitung

„Unser Ziel ist die Etablierung eines durchlässigen und demokratischen Prinzips in allen politischen und gesellschaftlichen Bereichen zur Stärkung der gesellschaftlichen Mitwirkung und Mitbestimmung. Wir arbeiten an Ideen und Projekten, die unsere heutige Demokratie flüssiger, transparenter und flexibler gestalten sollen. Dabei setzen wir auf den Einsatz moderner Medien. Denn hier sehen wir neue Formen der Informationsaufbereitung, Kommunikation und Interaktion, die zu einer Stärkung der diskursorientierten und demokratischen Teilhabe genutzt werden können.“ [Liqd]

LiquidFeedback wird von der Piratenpartei für demokratische Abstimmungen zur innerparteilichen Meinungsbildung genutzt. Da die Partei basisdemokratisch ausgerichtet ist, sollen Anträge auf Parteitag diskutiert werden und als Feedback für den Vorstand dienen. Der damalige politische Geschäftsführer Jan Huwald beschreibt die Vision hinter dem Einsatz von LiquidFeedback folgendermaßen:

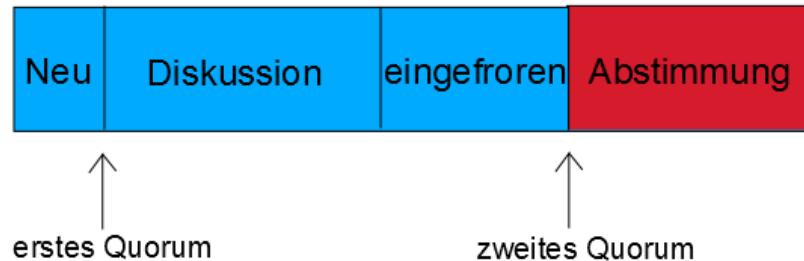
„Das langfristige Ziel, den Bürger ständig an Entscheidungen über die Entwicklung der Gesellschaft teil haben zu lassen, betrachte ich, unabhängig von der konkreten Durchführung, als außerordentlich lohnenswert. Die fortgeschrittene Vernetzung der Menschen und die Freiheit von Informationen machen diesen Traum greifbar.“ [Mü07]

Die wesentlichen Funktionen von LiquidFeedback lassen sich wie folgt zusammenfassen: User können Anträge in das System eingeben, die bei LiquidFeedback „Initiativen“ genannt werden. Jede Initiative ist einem Themenbereich zugeordnet und User können Mitglied von solchen Themenbereichen werden. Die Anzahl der Mitglieder eines Themenbereiches ist insofern relevant, als Initiativen nur zur Abstimmung zugelassen werden, falls diese mindestens von 10 Prozent der Mitglieder unterstützt werden. Hat eine Initiative dieses Quorum erreicht, folgt eine Diskussionsphase, in der Vorschläge zur Änderung oder Verbesserung eingereicht werden können. Anschließend wird die Initiative „eingefroren“, was verhindern soll, dass der Antragsteller den Text ohne Kenntnis der Unterstützer kurz vor der Abstimmung noch ändert. Um zur Abstimmung zugelassen zu werden, muss der Antrag deshalb noch einmal ein zweites Quorum in Höhe von 10 Prozent erreichen, da er in der Zwischenzeit möglicherweise abgeändert wurde. Ist auch diese Hürde genommen, wird die Initiative zur Abstimmung freigegeben. In Abbildung 1.1 ist der Prozess noch einmal graphisch dargestellt.

Stimmen für Initiativen können in LiquidFeedback auf einer Likert-Skala mit den Werten

## 1 Einleitung

Abbildung 1.1: Antrags-Prozess, Quelle: eigene Darstellung



minus zwei, minus eins, null, plus eins und plus zwei abgegeben werden. Bei Änderungsvorschlägen zu laufenden Initiativen entfällt der Null-Wert, da eine neutrale Meinung zu einem Vorschlag von keinem Nutzen ist. Verbesserungsvorschläge können inhaltlicher oder auch formaler Art sein. Wie eingangs erwähnt, ist Liquid Democracy eine Mischform aus direkter und repräsentativer Demokratie. Repräsentation wird in LiquidFeedback mittels Delegationen umgesetzt. Das bedeutet, dass jeder User seine Stimme einem anderen User übertragen kann. Dabei kann er entscheiden, ob er seine Stimme global, für einen Themenbereich oder nur für ein bestimmtes Thema übergibt. Delegationen in LiquidFeedback sind transitiv. Das bedeutet: Gibt ein User, der bereits Stimmen von anderen Usern bekommen hat, diese weiter, erhält der Delegierte das komplette Stimmgewicht.

Dies sind die zum Verständnis dieser Arbeit notwendigen Grundlagen des Systems. Für weitere Informationen zur Funktionsweise von LiquidFeedback wird auf [Jab11] verwiesen.

## 1.4 Eingrenzung des Untersuchungsgegenstandes

In der vorliegenden Arbeit wurden im ersten Schritt wesentliche Eigenschaften von Polarisierung herausgearbeitet. Anschließend wurde eine Literaturrecherche betrieben, um einen Überblick über bereits existierende Methoden zur Messung von Polarisierung zu gewinnen. Zusätzlich wurden noch zwei eigene Methoden entwickelt. Kapitel 3 erläutert die verwendeten Polarisierungsmaße und stellt sie in Abschnitt 3.4 gegenüber. Es wird dabei untersucht, welche Eigenschaften mit welchem Maß gemessen werden können. In Kapitel 5 und Kapitel 6 werden die Maße auf den LiquidFeedback-Datensatz angewendet. Dabei werden ausschließlich die Ergebnisse vorgestellt. Eine inhaltliche Interpretation wird nicht vorgenommen. Es werden lediglich Hypothesen für kausale Zusammenhänge aufgestellt.



## 2 Related Work

Das Phänomen Polarisierung wurde bereits in vielen Studien untersucht. Fiorina und Abrams [Fio08] liefern einen Überblick über verschiedene Kriterien, die in der Literatur betrachtet wurden, um Polarisierung zu untersuchen. Sie kommen zu dem Ergebnis, dass es kaum Hinweise darauf gibt, dass Polarisierung in der amerikanischen Gesellschaft in den letzten zwei Jahrzehnten zugenommen hat. Beispielsweise wurden der Zusammenhang von soziokulturellen Eigenschaften und Parteizugehörigkeit untersucht. DiMaggio et al. [DiM96] sind in ihrer Studie zum General Social Survey [Nat] und dem National Election Survey [Sta] der Frage nachgegangen, ob eine Polarisierung bezüglich der Einstellungen zu den dort gestellten Fragen stattgefunden hat. Um dies zu ermitteln, wurden die Verteilungen der abgegebenen Stimmen betrachtet. Auch hier kam man zu dem Ergebnis, dass eine Zunahme an Polarisierung in den letzten zwei Jahrzehnten nicht beobachtbar ist. Hierfür wurde eine multidimensionale Definition von Polarisierung entwickelt, welche die verschiedenen Eigenschaften von Polarisierung misst. Zum einen wurde die Verteilung der Stimmen betrachtet. Liegen die Stimmen weit auseinander, ist es schwierig, einen Konsens zu erreichen. Dieses Maß, genauer gesagt die Varianz der Verteilung, wurde auch von Flache und Macy [FM11] zur Quantifizierung von Polarisierung verwendet. Allerdings wird die Varianz dort in einem anderen Kontext benutzt. Anstelle der Varianz der abgegebenen Stimmen einer Initiative, wird die Varianz der Meinungsabstände zwischen Personen betrachtet. Diese sagt etwas darüber aus, in welchem Ausmaß eine Population in zwei gleich große Lager mit maximaler Meinungsverschiedenheit aufgeteilt ist. Als eine weitere Dimension von Polarisierung wurde die Korrelation von Abstimmungen untereinander betrachtet. Ist diese hoch, deutet dies darauf hin, dass die abstimmenden Personen häufig bei beiden Abstimmungen eine identische Meinung haben. Baldassarri und Gelman [BG08] bezeichnen diese Korrelation als *issue constraint*. Munzert und Bauer [MB13] nutzen ebenfalls die Korrelation von Einstellungen deutscher Bürger, um die politische Polarisierung in der deutschen Gesellschaft zu messen.

Eine weitere Herangehensweise dieser Arbeit, Polarisierung zu messen, bestand darin, die

## 2 Related Work

Ähnlichkeit von Personen innerhalb einer Gruppe und die Unterschiedlichkeit zwischen Gruppen zu analysieren. Spectral Clustering hat sich zu einer der populärsten Methoden für die Clusteranalyse entwickelt, da es die Leistung von traditionellen Clusterverfahren übertrifft. Eine Anleitung hierfür liefert Luxburg [VL07] in ihrem Tutorial. Methoden der linearen Algebra zur Clusteranalyse eines vorzeichenbehafteten Graphen liefern [Kun14] und [KSL<sup>+</sup>10]. Für dieses Verfahren muss das Abstimmungsnetzwerk als Graph modelliert werden. Ebenfalls aus dem Gebiet der Graphentheorie kommt das Konzept der *balance theory* [Har53]. Diese Theorie besagt, dass soziale Netzwerke dazu neigen, balanciert zu sein. Das bedeutet, dass die Knoten in zwei Mengen aufgeteilt werden können, so dass Knoten innerhalb der beiden Mengen durch positive Kanten und Knoten unterschiedlicher Mengen mit negativen Kanten verbunden sind. Eine Kennzahl, wie stark die Balance eines signierten Graphen ausgeprägt ist, sagt somit etwas darüber aus, wie gut eine Aufteilung in zwei Mengen unter dieser Voraussetzung gelingt. Des Weiteren kann die strukturelle Balance eines Graphen dazu verwendet werden, den Konflikt innerhalb eines sozialen Netzwerkes zu messen. [Kun14] Abschnitt 3.2 und Abschnitt 3.3 erläutern den Zusammenhang zwischen struktureller Balance und Polarisierung.

## 3 Methodik

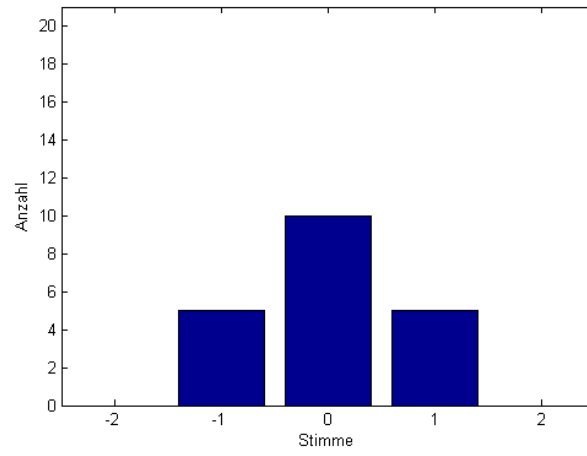
Im folgenden Kapitel werden zunächst unterschiedliche Formen und Merkmale von Polarisierung vorgestellt. Anschließend werden in Abschnitt 3.2 algebraische und netzwerktheoretische Grundlagen eingeführt, welche für das Verständnis von Abschnitt 3.3 (Messung von Polarisierung) notwendig sind.

### 3.1 Formen von Polarisierung

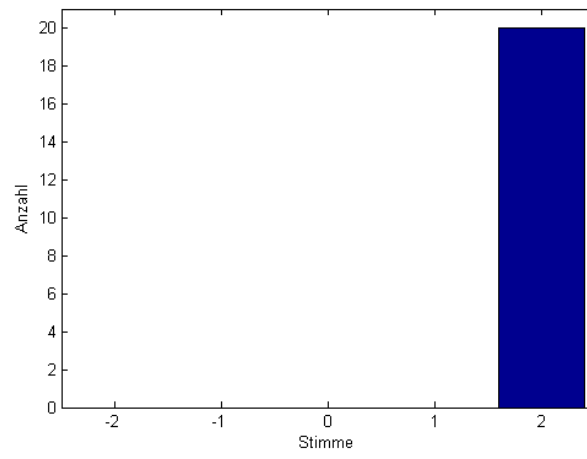
Um einen ganzheitlichen Blick auf Polarisierung zu erhalten, müssen unterschiedliche Dimensionen des Phänomens betrachtet werden. Auf der Mikroebene werden einzelne Abstimmungen untersucht und deren potenzielle Polarisierung analysiert. Diese spiegelt sich in der Verteilung der abgegebenen Stimmen wider. Bei Abstimmungen, in denen Konsens darüber herrscht, ob ein Vorschlag gut oder schlecht ist, haben im Extremfall alle Stimmen den gleichen Wert. Im Gegensatz dazu sind bei extrem polarisierten Abstimmungen die Stimmen gleichmäßig an beiden Enden der Skala angesiedelt. Die Diagramme in Abbildung 3.1 verdeutlichen unterschiedliche Szenarien. Abbildung 3.1a zeigt das Ergebnis normalverteilter Stimmen, Abbildung 3.1b kennzeichnet Konsens/Einstimmigkeit und Abbildung 3.1c bringt Polarisierung zum Ausdruck.

Auf der Makroebene äußert sich Polarisierung dadurch, dass sich bei mehreren Abstimmungen häufig dieselben Personen diametral gegenüberstehen. Ist das der Fall, können Gruppen beziehungsweise Lager innerhalb der Partei identifiziert werden. Charakteristisch für das Abstimmverhalten von Lagern ist, dass Mitglieder desselben Lagers häufig ähnlich und Mitglieder unterschiedlicher Lager oftmals konträr zueinander abstimmen.

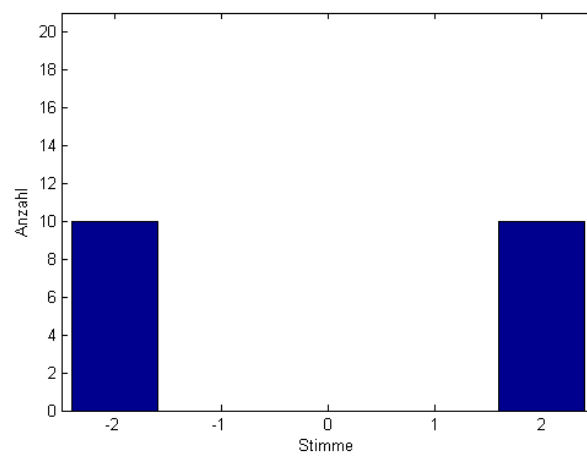
### 3 Methodik



(a)



(b)



(c)

Abbildung 3.1: Abstimmungsszenarien

## 3.2 Algebraische und netzwerktheoretische Grundlagen

Ein Ansatz, die Polarisierung im LiquidFeedback-System der Piratenpartei zu analysieren, bestand darin, Methoden der algebraischen Graphentheorie zu verwenden. Dies liegt nahe, da das Netzwerk aus Mitgliedern, Initiativen und abgegebenen Stimmen als Graph modelliert werden kann. Initiativen und User bilden die Knoten des Graphen und die abgegebenen Stimmen werden als Kanten repräsentiert. Genauer gesagt handelt es sich hierbei um einen vorzeichenbehafteten bipartiten Graph. Man sagt ein Graph ist vorzeichenbehaftet, sofern seine Kanten entweder mit positivem oder negativem Vorzeichen gekennzeichnet sind. In dem Fall des Abstimmungsnetzwerkes kann man an dem Vorzeichen folglich ablesen, ob ein Mitglied für oder gegen eine Initiative gestimmt hat. Da neutrale Stimmen mit dem Wert Null keinen Mehrwert für die Untersuchung der Polarisierung liefern, wurden diese ignoriert. Mathematisch ist ein vorzeichenbehafteter Graph definiert als  $G = (V, E, \sigma)$ . Wobei  $V$  die Menge der Knoten,  $E$  die Menge der Kanten und  $\sigma : E \rightarrow \{-1, +1\}$  die Vorzeichenfunktion repräsentiert. [Zas13] Die Vorzeichenfunktion  $\sigma$  ordnet jeder Kante ein positives oder negatives Vorzeichen zu. Sind zwei Knoten  $u$  und  $v$  benachbart, wird dies durch  $u \sim v$  gekennzeichnet. Ein Graph wird bipartit genannt, falls seine Knoten in zwei disjunkte Mengen aufgeteilt werden, so dass jede Kante einen Knoten der einen Menge mit einem Knoten der anderen Menge verbindet. Oder anders formuliert: Keine Kante verbindet Knoten derselben Menge. [Eas10] Formal ist ein bipartiter Graph ein Tripel  $G = (U, V, E)$  mit den beiden Knotenmengen  $U = \{u_1, \dots, u_n\}$  und  $V = \{v_1, \dots, v_n\}$  und der Menge der Kanten  $E = \{\{u_1, v_1\} \mid u_1 \in U, v_1 \in V\}$ . In unserem Fall ist  $U$  die Menge der Initiativen und  $V$  die Menge der Mitglieder. Um algebraische Methoden auf einen Graphen anwenden zu können, werden die darin enthaltenen Informationen in einer Matrix gespeichert. Diese wird auch Adjazenzmatrix genannt, da aus ihr ersichtlich ist, welche Knoten über eine Kante verbunden sind. Sie ist definiert als

$$A_{ij} = \begin{cases} \sigma(\{u, v\}) & \text{falls } \{u, v\} \in E, \\ 0 & \text{sonst.} \end{cases} \quad (3.1)$$

Viele Eigenschaften eines Graphen lassen sich mit Hilfe der Laplace-Matrix berechnen. Diese ist definiert als

$$L = D - A. \quad (3.2)$$

### 3 Methodik

Bei  $D$  handelt es sich um eine diagonale Matrix, welche den Grad eines jeden Knoten speichert:

$$D_{ij} = \begin{cases} \text{Grad des Knoten } i & \text{falls } i = j \in E, \\ 0 & \text{sonst.} \end{cases} \quad (3.3)$$

Der Grad eines Knoten  $v$  ist definiert als die Anzahl von Kanten, die den Knoten  $v$  mit anderen Knoten verbinden:

$$d(u) = \{v \mid u \sim v\} \quad (3.4)$$

Häufig wird die *Degree Matrix*  $D$  auch als

$$D_{ii} = \sum_{j \neq i} A_{ij} \quad (3.5)$$

definiert. Bei Verwendung dieser Definition muss Gleichung 3.2 für vorzeichenbehaftete Graphen dahingehend angepasst werden, dass man für  $D$  die Beträge von  $A_{ij}$  aufsummiert. Die vorzeichenbehaftete Laplace-Matrix eines Graphen  $G$  mit Adjazenzmatrix  $A$  ist definiert als

$$\bar{L} = \bar{D} - A \quad (3.6)$$

wobei die *Degree Matrix*  $\bar{D} \in \mathbb{R}^{V \times V}$  mit

$$\bar{D}_{ii} = \sum_{j \neq i} |A_{ij}| \quad (3.7)$$

gegeben ist. Die symmetrische Adjazenzmatrix eines bipartiten Graphen  $\mathbf{A}$  mit  $m$  Initiativen und  $n$  Mitgliedern, kann als

$$\mathbf{A} = \begin{bmatrix} \mathbf{0}_{m,m} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0}_{n,n} \end{bmatrix} \quad (3.8)$$

dargestellt werden.  $\mathbf{0}$  repräsentiert die Nullmatrix. Die Laplace-Matrix eignet sich gut für reguläre Graphen. Ein Graph heißt regulär, falls alle seine Knoten gleich viele Nachbarn haben, also den gleichen Grad besitzen. Liegt jedoch ein irregulärer Graph vor, wie es bei dem LiquidFeedback-Datensatz der Fall ist, eignet sich die normalisierte Form der Laplace-Matrix besser [GMZ03]. Die normalisierte Laplace-Matrix ist definiert als:

$$L_{sym} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \quad (3.9)$$

### 3 Methodik

Die normalisierte Laplace-Matrix eines Graphen  $G$  enthält somit folgende Einträge:

$$L_{sym}(G) = \begin{cases} 1 & \text{falls } u = v \text{ und } \text{Grad}(v) \neq 0, \\ -\frac{1}{\sqrt{\text{Grad}(u)\text{Grad}(v)}} & \text{falls } u \sim v, \\ 0 & \text{sonst.} \end{cases} \quad (3.10)$$

Die Eigenwerte und die dazugehörigen Eigenvektoren einer Adjazenz- oder Laplace-Matrix werden auch das Spektrum des dazugehörigen Graphen genannt. Da dieses zur Untersuchung vieler Eigenschaften eines Graphen genutzt werden kann, hat sich ein eigenes Forschungsgebiet zu diesem Thema etabliert. Anwendung finden die Erkenntnisse der Spektralen Graphentheorie zum Beispiel in der Clusteranalyse und bei der Visualisierung von Graphen [KSL<sup>+</sup>10]. Das Partitionieren eines Graphen erfolgt meistens über das Formulieren und Lösen des Minimum-Cut-Problems. Hierbei wird eine Partition der Menge der Knoten  $V$  in zwei disjunkte Mengen  $U$  und  $W$  gesucht, so dass die Anzahl an Kanten  $e = \{u, w\}, u \in U$  und  $w \in W$  minimal ist. Dieser minimale Schnitt  $cut(U, W)$  misst, wie stark zwei Cluster verbunden sind. Da man zwei voneinander getrennte Cluster finden will, muss der Schnitt minimiert werden. Dies führt meistens jedoch zu keinem befriedigenden Ergebnis, da es zu einer Partition führt, die sehr wenige Knoten vom Rest des Graphen separiert. Angewendet auf Abbildung 5.1 würde dies beispielsweise zu einer Partition führen, bei der ein Knoten mit Grad von 1, wie zum Beispiel einer der beiden Knoten unten rechts in der Ecke, vom Rest der Knoten separiert würde. Um dem entgegenzuwirken, wurde das Konzept des Ratio-Cuts von Wei und Cheng [WC89] entwickelt, welches sich als sehr erfolgreich erwiesen hat.

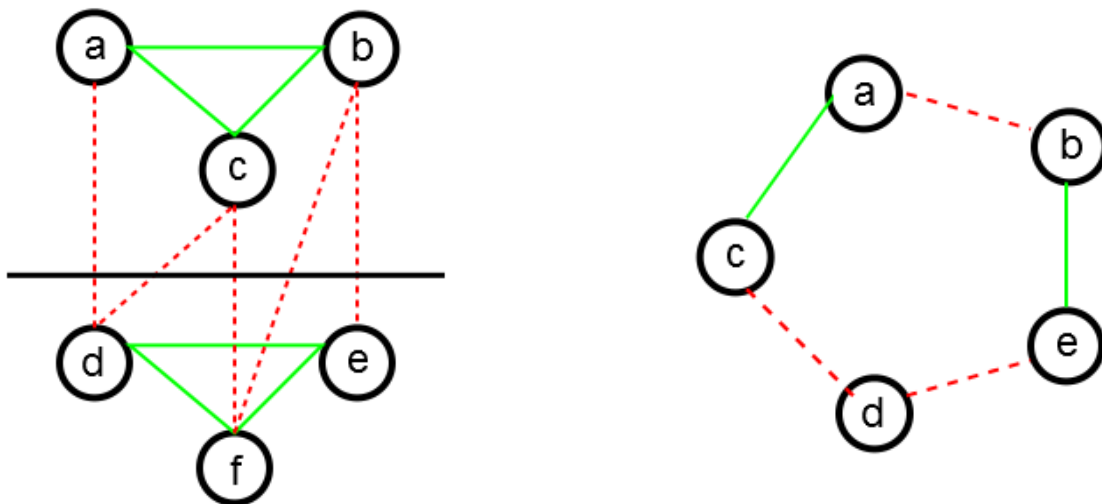
$$RatioCut(U, W) = cut(U, W) \left( \frac{1}{|U|} + \frac{1}{|W|} \right) \quad (3.11)$$

Der Ratio-Cut findet somit „natürliche“ Partitionierungen, da nicht nur die Kosten des Cuts, sondern auch die Größe der beiden Partitionen in die Berechnung miteinbezogen werden. Bei der spektralen Clusteranalyse wird das Minimum-Ratio-Cut-Problem für zwei Cluster mit Hilfe des kleinsten Eigenvektors der Laplace-Matrix gelöst. Dieser teilt, basierend auf dem Vorzeichen des zugehörigen Vektoreintrags, die Knoten in zwei Gruppen auf. [KSL<sup>+</sup>10]

Aussagen über die Struktur eines Graphen können anhand der kleinsten Eigenwerte, der zu dem Graph gehörenden Laplace-Matrix, getroffen werden. Zum Beispiel ist bekannt,

### 3 Methodik

dass der kleinste Eigenwert  $\lambda_1$  bei einem vorzeichenbehafteten Graphen größer Null ist, sofern der Graph unbalanciert ist. [Hou05] Man spricht auch von algebraischem Konflikt. Ein Graph ist balanciert, wenn seine Knoten in zwei Gruppen aufgeteilt werden können, so dass alle positiven Kanten Knoten innerhalb derselben Gruppe verbinden und alle negativen Kanten zwischen Knoten unterschiedlicher Gruppen verlaufen. [Har53] Abbildung 3.2a zeigt einen balancierten Graph, der mit Hilfe der schwarzen Linie in zwei Gruppen partitioniert werden kann. Harary [Har53] liefert noch eine weitere, äquivalente Definition: Ein Graph ist balanciert, wenn alle seine Zyklen positiv sind. Ein Zyklus ist positiv, sofern er eine gerade Anzahl von negativen Kanten besitzt, beziehungsweise das Produkt der Kanten positiv ist. Wenn man die strukturelle Balance eines Graphen aus dem



(a) Ein Graph ohne negative Zyklen kann in zwei Gruppen aufgeteilt werden, so dass alle Kanten innerhalb jeder Gruppe positiv und alle Kanten zwischen den beiden Gruppen negativ sind. Der Graph ist balanciert.

(b) Ein unbalancierter Graph enthält mindestens einen negativen Zyklus. Ein solcher Graph kann nicht in zwei Gruppen aufgeteilt werden, so dass alle Kanten innerhalb jeder Gruppe positiv und alle Kanten zwischen den beiden Gruppen negativ sind.

Abbildung 3.2: Structural Balance

Blickwinkel von Gruppenbildung betrachtet, wird klar, dass ein balancierter Graph wie in Abbildung 3.2a mit starker Polarisierung gleichzusetzen ist. Je größer  $\lambda_1$ , desto mehr Kanten existieren, die zur Unbalanciertheit beitragen.



### 3.3 Messung von Polarisierung

Wie in Abschnitt 3.1 beschrieben, unterscheidet der Autor bei Polarisierung zwischen Mikro- und Makroebene. Die folgenden Kapitel erläutern, wie diese beiden unterschiedlichen Arten identifiziert und gemessen wurden.

#### 3.3.1 Messung von Polarisierung auf Abstimmungsebene

DiMaggio et al. [DiM96] sind in ihrer Studie zum General Social Survey (GSS) [Nat] und dem National Election Survey (NES) [Sta] der Frage nachgegangen, ob eine Polarisierung bezüglich der Einstellungen zu den dort gestellten Fragen stattgefunden hat. Um dies zu ermitteln, wurden die Verteilungen der abgegebenen Stimmen betrachtet. Zwei der vier dort genannten Maße wurden für die Analyse der LiquidFeedback-Datensätze übernommen:

- Varianz als Streuungsmaß: Je verteilter die Stimmen sind, desto schwieriger ist es, einen Konsens zu finden. Die Varianz spiegelt die durchschnittliche Abweichung von zwei zufällig ausgewählten Stimmen wider. Polariserte Abstimmungen weisen daher eine größere Varianz auf als harmonische.

$$\text{Var}(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3.12)$$

- Wölbung, oder auch Kurtosis genannt, als Maß für die Bimodalität einer Verteilung [Dar70] [Hil71]: Eine bimodale Verteilung weist im Gegensatz zu einer unimodalen zwei Modi auf. Meinungen zu einem Thema gelten als polarisiert, falls sich die abgegebenen Stimmen bei zwei gegensätzlichen Positionen anhäufen und Positionen dazwischen dünn besetzt sind. Um die Kurtosis besser einschätzen zu können, wird sie mit der Wölbung einer Normalverteilung verglichen. Da eine Normalverteilung eine Kurtosis von drei besitzt, wird durch das Subtrahieren von drei sichergestellt, dass ein normalverteiltes Abstimmungsergebnis den Wert null annimmt. Ist eine Verteilung steil, was auf ein hohes Maß an Konsens hindeutet, ist der Wert der Kurtosis positiv. Bei einer Verteilung, welche flacher als die Normalverteilung ist, wird der Wert negativ. Der kleinst mögliche Wert liegt bei -2, was perfekte Bimodalität

### 3 Methodik

widerspiegelt.

$$\omega = \left( \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s} \right)^4 \right) - 3 \quad (3.13)$$

Gleichung 3.12 und Gleichung 3.13 wurden dazu genutzt, um den Grad der Polarisierung einzelner Initiativen zu ermitteln. Das dritte Maß, welches von Dimaggio et al. verwendet wurde, ist Cronbachs Alpha [Cro51]. Es ist definiert als die durchschnittliche Korrelation zwischen zwei Items und wird in [DiM96] dazu verwendet, die *opinion constraints* innerhalb des Fragebogens zu berechnen. Damit beziehen sich die Autoren auf das Ausmaß, in dem Meinungen zu einer Fragestellung von Meinungen zu anderen Fragen abhängig sind. Als Beispiel wird die Einstellung gegenüber vorehelichem Geschlechtsverkehr und Abtreibung genannt. Ist die Einstellung zu einem der beiden Themen bekannt, lässt sich die Einstellung zum anderen mit hoher Wahrscheinlichkeit voraussagen. Daher werden die Korrelationen zwischen einzelnen Fragestellungen als Indikator für ideologischen Zusammenhang betrachtet. Cronbachs Alpha konnte auf dem LiquidFeedback-Datensatz nicht angewandt werden, da es nicht mit fehlenden Werten umgehen kann. Um Cronbachs Alpha berechnen zu können, hätte jede Initiative von jedem User bewertet werden müssen. Des Weiteren ist es nur bei intervall- und verhältnisskalierten Daten anwendbar. Einen ähnlichen Ansatz zur Quantifizierung der Abhängigkeiten zwischen Fragen/Themen/Sachverhalten verfolgen Baldassarri und Gelman [BG08]. Sie untersuchten ebenfalls politische Polarisierung im NES-Datensatz anhand der Korrelation zwischen Paaren von Fragen. Sie nennen es *issue alignment* und berechnen hierfür den Korrelationskoeffizienten für Paare von Fragen. Da die 47 Fragen aus dem Fragebogen 1081 potenzielle Paare ergeben, wurden nur die sieben Fragen betrachtet, welche am stärksten mit Parteizugehörigkeit und Ideologie korrelieren. Baldassarri und Gelman kamen zu dem Ergebnis, dass sich die Korrelationen zwischen den sieben Fragen in den betrachteten Jahren kaum verändert haben. Um die Entwicklung der Abhängigkeiten zwischen den Initiativen des LiquidFeedback-Datensatzes zu analysieren, wurde die Kosinus-Ähnlichkeit verwendet. Sie kann Werte im Bereich von -1 bis 1 annehmen. Ein Wert von 1 bedeutet exakte Gleichheit und ein Wert von -1 maximale Ungleichheit. Der Korrelationskoeffizient und die Kosinus-Ähnlichkeit können beide als Varianten des Skalarprodukts angesehen werden. Der Unterschied der beiden Maße wird anhand der Definitionen klar:

$$\text{Kosinus-Ähnlichkeit} = \frac{\langle a, b \rangle}{\|a\| \times \|b\|} = \frac{\sum a_i b_i}{\sqrt{\sum a_i^2} \sqrt{\sum b_i^2}} \quad (3.14)$$

### 3 Methodik

$$\rho = \frac{\sum(a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum(a_i - \bar{a})^2} \sqrt{\sum(b_i - \bar{b})^2}} = \text{Kosinus-Ähnlichkeit}(a - \bar{a}, b - \bar{b}) \quad (3.15)$$

Der Korrelationskoeffizient ist invariant gegenüber Verschiebungen, die Kosinus-Ähnlichkeit jedoch nicht. Wenn ein Vektor  $x$  um eins verschoben würde ( $x + 1$ ), bekäme man einen anderen Wert für die Kosinus-Ähnlichkeit. Für den untersuchten LiquidFeedback-Datensatz, wo Zustimmung mit +1 und Ablehnung mit -1 modelliert wurde, ist die Verwendung der Kosinus-Ähnlichkeit angemessen. Die Wahl für das Maß der Abhängigkeiten zwischen Initiativen fiel auf die Kosinus-Ähnlichkeit, da sie mit den vorliegenden Matrizen der Form *Initiative*  $\times$  *Mitglied* durch Matrixmultiplikation leicht zu berechnen war.

#### 3.3.2 Messung von Polarisierung auf Netzwerkebene

Um die Entwicklung der globalen Polarisierung über die Zeit zu untersuchen, wurde zunächst die Distanz der abgegebenen Stimmen für jedes Paar von Usern ermittelt:

$$d_{i,j} = \frac{1}{N} \sum_{n=1}^N |s_{j,n} - s_{i,n}| \quad (3.16)$$

Der Grad an globaler Polarisierung nach Flache et al. [Fla11] ergibt sich aus der Varianz der Verteilung der  $d_{i,j}$ .

$$P_{Flache} = \frac{1}{N(N-1)} \sum_{\substack{i=Nj=N \\ i \neq j}} (d_{i,j} - \bar{d})^2 \quad (3.17)$$

Als zusätzliches Maß für die globale Polarisierung wurde Krippendorffs Alpha verwendet. Hierbei handelt es sich um ein statistisches Maß, das den Grad der Übereinstimmung der abgegebenen Stimmen zwischen allen Usern misst. Es existieren verschiedene Methoden, diese sogenannte Interrater-Reliabilität zu messen:

- Scotts (Pi)  $\pi$
- Cohens (Kappa)  $\kappa$
- Fleiss' (Kappa)  $\kappa$
- Krippendorffs (Alpha)  $\alpha$

### 3 Methodik

Scotts Pi [Sco55] und Cohens Kappa [Coh60] eignen sich, um die Übereinstimmung von zwei Beurteilern zu messen, mit der Einschränkung, dass ein nominales Skalenniveau vorliegen muss. Fleiss [Fle71] verallgemeinerte Scotts Pi für mehrere Beurteiler. Krippendorff [Kri11] ging noch einen Schritt weiter und entwickelte ein Reliabilitätsmaß, welches auf Daten mit folgenden Eigenschaften anwendbar ist:

- jede Anzahl von Beurteilern, nicht nur zwei
- jedes Skalenniveau (Nominalskala, Ordinalskala, Intervallskala, Verhältnisskala)
- unvollständige oder fehlende Daten

Krippendorffs Alpha eignet sich somit für die Analyse der Interrater-Reliabilität des vorliegenden LiquidFeedback-Datensatzes, da die Urteilsübereinstimmung aller User analysiert werden soll, ein ordinales Skalenniveau vorliegt und viele Daten fehlen (Nicht jeder User nimmt an jeder Abstimmung teil). Alphas allgemeine Form lautet:

$$\alpha = 1 - \frac{D_o}{D_e} \quad (3.18)$$

Bei  $D_o$  handelt es sich um die beobachtete Meinungsverschiedenheit der abgegebenen Stimmen

$$D_o = \frac{1}{n} \sum_c \sum_k o_{ck} \text{ metric} \delta_{ck}^2 \quad (3.19)$$

und bei  $D_e$  um die erwartete Meinungsverschiedenheit, wenn die Stimmen zufällig abgegeben worden wären.

$$D_e = \frac{1}{n * (n - 1)} \sum_c \sum_k n_c * n_k \text{ metric} \delta_{ck}^2 \quad (3.20)$$

Die Variablen  $o_{ck}$ ,  $n_c$ ,  $n_k$  und  $n$  beziehen sich auf Werte der Koinzidenz-Matrix, welche im Folgenden erläutert wird. Bei der Berechnung von Krippendorffs Alpha hat sich der Autor an [Kri70] orientiert. Um Krippendorffs Alpha berechnen zu können, benötigt man eine sogenannte Koinzidenz-Matrix. Hierbei handelt es sich um eine zweidimensionale Matrix, in der die Häufigkeiten von paarweise vorkommenden Variablenkombinationen enthalten sind. Im untersuchten Datensatz konnten die User, wie in Abschnitt 1.3 beschrieben, zwischen fünf beziehungsweise vier unterschiedlichen Variablenausprägungen wählen. Somit ergibt sich eine 5x5-Koinzidenz-Matrix für die *votes* und eine 4x4-Koinzidenz-Matrix für die *opinions*.

### 3 Methodik

**Tabelle 3.1:** Koinzidenz-Matrix

|   |          |   |          |   |          |
|---|----------|---|----------|---|----------|
|   | 1        | . | k        | . | 5        |
| 1 | $o_{11}$ | . | $o_{1k}$ | . | .        |
| . | .        | . | .        | . | .        |
| . | .        | . | .        | . | .        |
| c | $o_{c1}$ | . | $o_{ck}$ | . | .        |
| 5 | .        | . | .        | . | $o_{55}$ |
|   | $n_1$    | . | .        | . | $n_5$    |

Zur Erstellung der Koinzidenz-Matrix wird zuerst eine  $m \times n$  Reliabilität-Matrix für die  $m$  User und die  $n$  Abstimmungen benötigt, welche die Abstimmungswerte aller User zu allen Initiativen/Verbesserungsvorschlägen enthält.

**Tabelle 3.2:** Reliabilität-Matrix

|   |          |   |          |   |          |
|---|----------|---|----------|---|----------|
|   | 1        | . | k        | . | n        |
| 1 | $v_{11}$ | . | $v_{1k}$ | . | .        |
| . | .        | . | .        | . | .        |
| c | $v_{c1}$ | . | $v_{ck}$ | . | .        |
| . | .        | . | .        | . | .        |
| m | .        | . | .        | . | $v_{mn}$ |
|   | $m_1$    | . | .        | . | $m_n$    |

Der Wert  $m_i$  enthält die Anzahl von Werten, die zur Abstimmung  $i$  abgegeben wurden. Der Wert  $o_{ck}$  für die Koinzidenz-Matrix errechnet sich wie folgt:

$$o_{ck} = \sum_u \frac{\text{Anzahl von c-k Paaren in Abstimmung } u}{m_u - 1} \quad (3.21)$$

Für die Berechnung von  $\alpha$  fehlt noch eine Metrik zur Bestimmung der Differenz zwischen verschiedenen Abstimmungswerten. Für ordinal skalierte Daten ergibt sich der Wert für  $ordinal \delta_{ck}^2$  folgendermaßen:

$$ordinal \delta_{ck}^2 = \left( \sum_{g=c}^{g=k} n_g - \frac{n_c + n_k}{2} \right)^2 \quad (3.22)$$

Zusammengefasst ergibt dies folgende Berechnungsformel für  $\alpha$ :

$$\alpha = 1 - \frac{D_o}{D_e} = 1 - (n - 1) \frac{\sum_c \sum_{k>c} o_{ck} \text{ ordinal } \delta_{ck}^2}{\sum_c \sum_{k>c} n_c \cdot n_k \text{ ordinal } \delta_{ck}^2} \quad (3.23)$$

### 3 Methodik

Für den Fall, dass die User perfekt übereinstimmen, ist die beobachtete Meinungsverschiedenheit  $D_o = 0$  und  $\alpha = 1$ . Wenn die Übereinstimmung nicht über das Maß einer zufälligen Übereinstimmung hinausgeht, ist  $D_o = D_e$  und  $\alpha = 0$ . Ein Alphawert  $< 0$  deutet auf systematische Meinungsverschiedenheit hin.

Eine weitere Methode, die Polarisierung im gesamten Abstimmungsnetzwerk zu bestimmen, war die Ähnlichkeiten innerhalb von Gruppen und die Unterschiedlichkeit zwischen den Gruppen zu betrachten. Wie in Abschnitt 3.2 erwähnt, kann die Matrixrepräsentation des Abstimmungsnetzwerkes dazu verwendet werden, ein Clustering zu erstellen. Ein Clustering ist eine vielversprechende Methode, um die Polarisierung einer Menge von Personen zu bestimmen, da die wesentlichen Merkmale wie Grad der Übereinstimmung innerhalb einer Gruppe und Nicht-Übereinstimmung zwischen Gruppen anhand der Einteilung in Gruppen berechnet werden kann. Eine Herausforderung hierbei besteht darin, vorab eine geeignete Anzahl von Clustern zu finden.

Ein völlig anderer Ansatz, aus der Netzwerktheorie, die globale Polarisierung zu messen, ist die in [Kun14] definierte Kennzahl  $\zeta$ . Hierbei handelt es sich um den kleinsten Eigenwert der Laplace-Matrix. Bei einem vorzeichenbehafteten Graphen ist dieser Null, sofern der Graph balanciert ist. Ist ein Graph balanciert, d.h. alle positiven Kanten verlaufen innerhalb einer Gruppe und alle negativen Kanten verlaufen zwischen unterschiedlichen Gruppen, bedeutet dies, dass er maximal polarisiert ist. Je größer  $\zeta$ , desto größer ist der algebraische Konflikt und desto mehr Kanten existieren, die zur Unbalanciertheit des Graphen beitragen.  $\zeta$  gibt somit einen Hinweis darauf, wie stark der Graph polarisiert ist. Um den algebraischen Konflikt von Graphen mit unterschiedlichen Anzahlen von Kanten vergleichen zu können, wurde sich des Konzepts der „frustrierten Kanten“ [Kun15] bedient. Das Maß der Polarisierung  $b_f$  ergibt sich aus dem Verhältnis von frustrierten Kanten zur Gesamtanzahl der Kanten. Eine Kante wird frustriert genannt, sofern sie negativ ist und innerhalb eines Clusters verläuft oder wenn sie positiv ist und zwischen zwei Clustern verläuft. Kunegis zeigt in seiner Arbeit [Kun15], dass  $\zeta$  dazu verwendet werden kann, die Anzahl der frustrierten Kanten in einem Graphen zu approximieren. Eine detaillierte Beschreibung des Konzepts wird in Kapitel 6 gegeben.

## 3.4 Diskussion

Die vorgestellten Maße zur Messung von Polarisierung beleuchten unterschiedliche Aspekte von Polarisierung. Um sie miteinander vergleichen zu können, sind in Tabelle 3.3 verschiedene Eigenschaften der Maße aufgeführt.

### 3.4.1 Krippendorffs Alpha als Polarisierungsmaß

Beginnend mit Krippendorffs Alpha werden im Folgenden die Eigenschaften der Polarisierungsmaße für jedes Maß erörtert. Da es sich bei Krippendorffs Alpha um die Interrater-Reliabilität handelt und folglich nur das Ausmaß an Übereinstimmungen der abgegebenen Stimmen betrachtet wird, ist dieses Maß nicht geeignet, um Gruppen zu identifizieren. Abschnitt 6.1 zeigt anhand eines Beispiels, warum dies so ist. Krippendorffs Alpha ist user-zentriert und gibt einen ersten Anhaltspunkt, ob ein allgemeiner Konsens oder Uneinigkeit im Abstimmungsnetzwerk vorliegt. Liegt bereits eine Aufteilung in Gruppen vor, kann Krippendorffs Alpha dazu verwendet werden, den Grad der Polarisierung anhand der Inter- und Intra-Gruppenübereinstimmung zu bestimmen. Ein kombinierter Wert aus  $\alpha$  innerhalb und zwischen Gruppen ist ein Indiz dafür, wie stark die Polarisierung ausgeprägt ist. Je größer  $\alpha$  innerhalb und je kleiner zwischen Gruppen ist, desto stärker ist die Polarisierung. Die Anwendung ist hierbei nicht auf den Zwei-Gruppen-Fall beschränkt. Da das Maß aus dem Verhältnis von erwarteter und beobachteter Übereinstimmung resultiert, ist es invariant gegenüber unterschiedlich großen Datensätzen und kann auch bei fehlenden Daten angewendet werden.

### 3.4.2 Polarisierungsmaß gemäß Fläche und Macy

Das Polarisierungsmaß nach Fläche und Macy, bei dem es sich um die Varianz der Meinungsdistanzen handelt, zeigt die Spanne der abgegebenen Stimmen auf. Sie ist maximal, wenn die sich abgegebenen Stimmen gleichmäßig an beiden Enden der Skala verteilen. Somit wird perfekte Polarisierung erkannt, jedoch nur im Zwei-Gruppen-Fall und bei gleicher Größe der beiden Gruppen. Liegt perfekte Polarisierung zwischen unterschiedlich großen Gruppen vor, wird  $P_{Fläche}$  kleiner. Daraus folgt, dass man anhand eines kleineren Wertes nicht erkennen kann, ob der Wert aus unterschiedlich großen Gruppen oder aus geringer

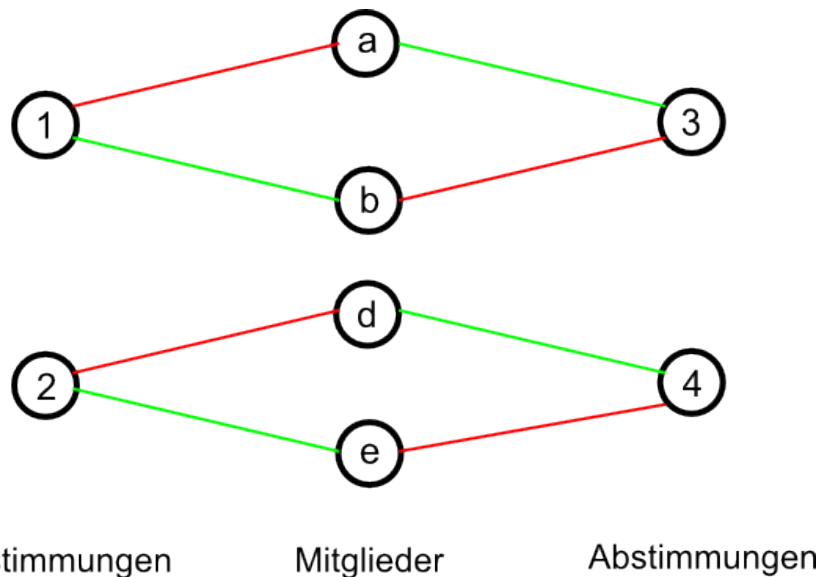
Tabelle 3.3: Eigenschaften der Polarisierungsmaße

|  | Misst Grad der Lagerbildung | User- oder abstim-mungs-zentriert | Misst User-Übereinstimmung | Misst Inter-/Intra-Gruppen-Übereinstimmung | Größen-invarianz | Für mehr als zwei Gruppen geeignet | Für unvollständige Daten geeignet |
|--|-----------------------------|-----------------------------------|----------------------------|--|------------------|------------------------------------|-----------------------------------|
| Krippendorffs Alpha                    | ✗                           | user-zentriert                    | ✓                          | ✗  | ✓                | ✓                                  | ✓                                 |
| Polarisierung nach Flake und Macy      | (✓)                         | user-zentriert                    | (✓)                        | ✗  | ✗                | ✓                                  | ✗                                 |
| Korrelationen zwischen Fragestellungen | (✓)                         | abstim-mungs-zentriert            | ✗                          | ✗  | ✓                | ✓                                  | (✓)                               |
| Cluster-Ähnlichkeit                    | ✓                           | user-zentriert                    | ✓                          | ✓  | ✓                | ✓                                  | ✓                                 |
| Anteil frustrierter Kanten             | ✓                           | user-zentriert                    | ✗                          | ✗  | ✓                | ✗                                  | ✓                                 |



### 3 Methodik

polarisierten Abstimmungen resultiert. Die Varianz der Meinungsdistanzen eignet sich somit nur bedingt zur Identifizierung von Gruppenbildung. Im Gegensatz zu Krippendorffs Alpha kann keine Aussage über den Grad der User-Übereinstimmung getroffen werden. Abbildung 3.3 zeigt ein Minimalbeispiel, aus dem ersichtlich wird, warum dies der Fall ist. Die User sind bei zwei Fragestellungen entgegengesetzter Meinung und  $P_{Flache} = 0$ , da  $\bar{d} = 2$  und  $d_{i,j} = 2 \forall i, j, i \neq j$ , was als das minimale Level an Polarisierung interpretiert wird. Stimmen die zwei Personen bei beiden Abstimmungen überein, ist  $P_{Flache}$  ebenfalls 0. Lediglich im Falle von zwei Gruppen und vollständigen Daten (jeder User hat bei jeder Frage abgestimmt) kann die Varianz der Meinungsdistanzen als Maß der User-Übereinstimmung herangezogen werden. Abbildung 3.3 verdeutlicht ein weiteres Problem von  $P_{Flache}$  als Maß für Polarisierung. Liegen unvollständige Daten vor, was bei Datensätzen, insbesondere bei LiquidFeedback, oft der Fall ist, versagt  $P_{Flache}$  als Kennzahl für Polarisierung. Wie bereits erwähnt, liegt der Wert von  $P_{Flache}$  bei 0, obwohl Polarisierung sowohl zwischen a und b, als auch zwischen d und e erkennbar ist. Des Weiteren ist  $P_{Flache}$



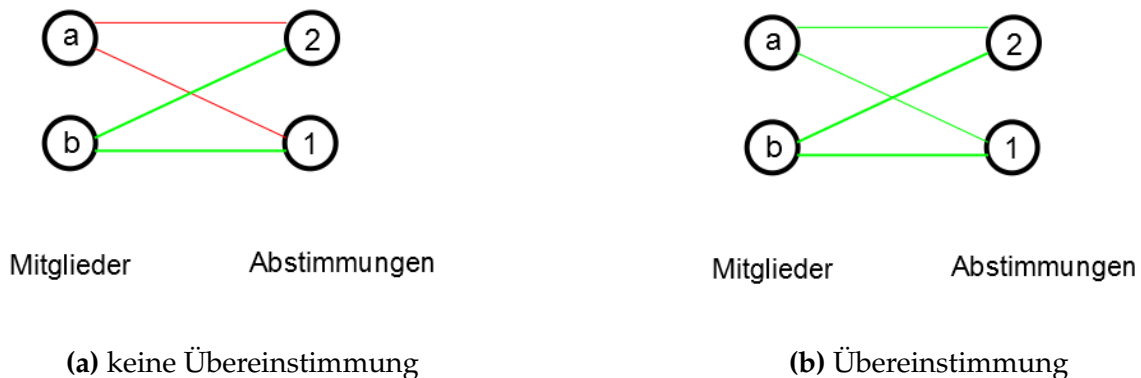
**Abbildung 3.3:** Übereinstimmung und Lagerbildung bei unvollständigen Daten

nicht invariant gegenüber unterschiedlich großer Datensätze. Der Wert für maximale Polarisierung ist abhängig von der Anzahl der User [FM11]:

$$P_{Flache} = 1 - \frac{1}{(N - 1)^2} \quad (3.24)$$

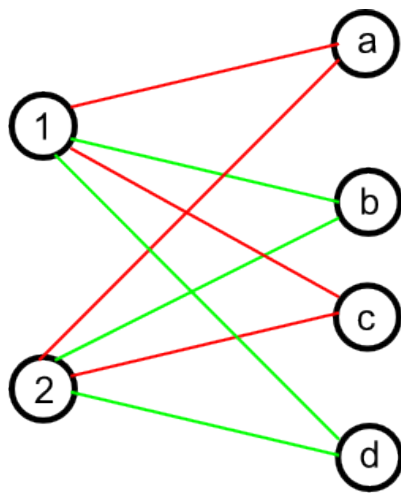
### 3.4.3 Korrelationskoeffizient zwischen Abstimmungen als Polarisierungsmaß

Der Korrelationskoeffizient als Maß für Polarisierung betrachtet im Gegensatz zu den anderen Maßen nicht die User, sondern die Abhängigkeiten von Fragen zueinander. Was allerdings bei vollständigen Daten indirekt auch etwas über das Abstimmverhalten der User aussagt. Eine hohe Korrelation zwischen zwei Fragestellungen bedeutet, dass Personen, unabhängig von ihrer Gruppenzugehörigkeit, bei beiden Fragestellungen identisch abgestimmt haben. Hieraus kann man jedoch keine Aussage über die Übereinstimmung der Personen treffen, da eine Korrelation von 1 aus völliger Übereinstimmung als auch aus Meinungsverschiedenheiten resultieren kann. Im Normalfall sagt der Korrelations-



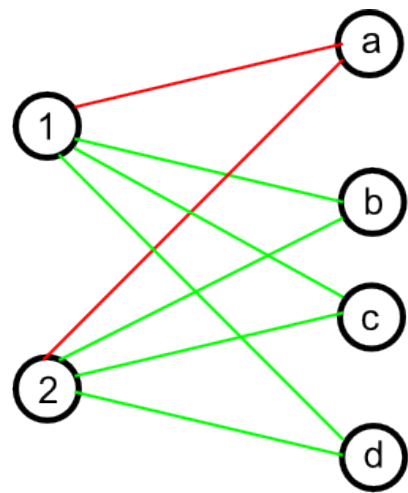
**Abbildung 3.4:** In beiden Szenarien haben die Abstimmungen einen Korrelationskoeffizienten von 1, jedoch stimmen die beiden Personen in Abbildung 3.4b überein und in Abbildung 3.4a nicht.

koeffizient etwas über den Grad der Lagerbildung aus, sofern Abstimmungen sowohl positive als auch negative Stimmen enthalten. Ein hoher Korrelationskoeffizient bedeutet, dass viele Personen zu unterschiedlichen Fragestellungen gleich abstimmen. Liegt zum Beispiel perfekte Korrelation mit  $\rho = 1$  vor und bei beiden Abstimmungen gibt es sowohl positive als auch negative Stimmen, spiegelt dies starke Polarisierung wider. Dabei spielt die Größe der Gruppen keine Rolle. Eine Fünfzig-Fünfzig-Aufteilung kann ebenso wie eine sehr ungleiche Aufteilung der Gruppen eine Korrelation von 1 aufweisen. Abbildung 3.5 verdeutlicht dies grafisch. Der Ausnahmefall, dass alle Abstimmungen ausschließlich positive oder ausschließlich negative Stimmen erhalten, welcher vollständige Übereinstimmung widerspiegelt, muss natürlich ausgeschlossen werden. Da der Korrelationskoeffizient normalisiert ist, ist er invariant gegenüber unterschiedlich großer



Abstimmungen Mitglieder

(a) keine Übereinstimmung



Abstimmungen Mitglieder

(b) corrUngleich

**Abbildung 3.5:** In beiden Szenarien korrelieren die Abstimmungen perfekt, bei unterschiedlicher Gruppengröße

Datensätze. Verwendet man den Korrelationskoeffizienten, um eine Aussage darüber zu treffen, ob sich die Abhängigkeiten zwischen Fragestellungen verändern, kann er auch bei mehr als zwei Gruppen verwendet werden.

### 3.4.4 Gruppenähnlichkeit als Polarisierungsmaß

Bei Verwendung der Gruppen-Ähnlichkeit als Polarisierungsmaß wird im ersten Schritt eine Clusteranalyse vorgenommen, um Gruppen zu identifizieren. Zur Identifikation von homogenen Gruppen wird ein Ähnlichkeitsmaß verwendet, welches das Abstimmverhalten der Personen miteinander vergleicht. Personen mit ähnlichem Abstimmverhalten werden in Gruppen zusammengefasst. Ziel der Clusteranalyse ist es, Gruppen zu identifizieren, so dass Personen einer Gruppe ein ähnliches Abstimmverhalten und Personen unterschiedlicher Gruppen ein eher konträres Abstimmverhalten aufweisen. Die Ähnlichkeiten zwischen den Gruppen sagt somit etwas über die Übereinstimmung der Personen aus. Da die Vektoren bei der Berechnung der Kosinus-Ähnlichkeit normiert werden, ist das Maß auch invariant gegenüber unterschiedlich großer Vektoren.

### 3.4.5 Anteil frustrierter Kanten als Polarisierungsmaß

Will man den Anteil frustrierter Kanten  $b_f$  berechnen, muss hierfür ebenfalls zuerst eine Clusteranalyse durchgeführt werden. Schneller und einfacher gelingt dies mit der in Unterabschnitt 3.3.2 eingeführten Kennzahl  $\zeta$ . Abschnitt 6.5 zeigt wie  $\zeta$  zur Approximation von  $b_f$  genutzt werden kann. Da  $b_f$  nur etwas über die Anzahl von Kanten aussagt, die entfernt werden müssten, um eine perfekte Aufteilung in Gruppen zu erreichen, sind keine Rückschlüsse auf Gruppenzusammensetzung und Konsens beziehungsweise Meinungsverschiedenheit möglich. Das Maß variiert nicht bei mengenmäßiger Veränderung der Abstimmung (Zum Beispiel bei Verdopplung der abgegeben Stimmen), da sich die Art der Kante (positiv oder negativ) hierdurch nicht verändert. Das in dieser Arbeit vorgestellte Verfahren zur Approximation von  $b_f$  ist nur für den Zwei-Gruppen-Fall geeignet.  $b_f$  kann auch bei unvollständigen Daten angewendet werden, da der kleinste Eigenwert der Laplace-Matrix auch für im Falle unvollständiger Daten berechnet werden kann.

### 3.4.6 Korrelationen der vorgestellten Maße

Um das Verhältnis der Maße untereinander näher zu betrachten, wurde eine Korrelationsmatrix erstellt (Tabelle 3.4).  $\alpha$  steht für Krippendorffs Alpha,  $\rho_{Initiativen}$  für den Korrelationskoeffizienten zwischen Initiativen,  $P_{Cluster}$  für das in Abschnitt 6.4 eingeführte Maß der Clusterähnlichkeiten und  $\zeta$  für den Anteil frustrierter Kanten (siehe Abschnitt 6.5). Da man nicht annehmen kann, dass die Beziehungen zwischen den Maßen linear sind, wurde der Rangkorrelationskoeffizient Kendalls Tau verwendet. Er ist wie folgt definiert:

$$\tau = \frac{n_k - n_d}{\frac{1}{2}n(n-1)} \quad (3.25)$$

Wobei  $n_k$  die Anzahl konkordanter und  $n_d$  die Anzahl diskordanter Paare kennzeichnet. Um  $\tau$  zu berechnen, werden sortierte Paare  $(x_i, y_i)$  und  $(x_j, y_j)$  mit  $i = 1, \dots, n-1$  und  $j = i+1, \dots, n$  betrachtet. Es gilt demnach:  $x_1 \leq x_2 \leq \dots \leq x_n$ . Anschließend werden die Paare paarweise miteinander verglichen. Ein Paar heißt konkordant, falls  $x_i < x_j$  und  $y_i < y_j$  und diskordant, falls  $x_i < x_j$  und  $y_i > y_j$ .

Signifikante Korrelationen mit einem p-Wert kleiner als 0,05 sind fett gedruckt. Demnach besteht ein relativ starker Zusammenhang zwischen Krippendorffs Alpha und der durchschnittlichen Korrelation zwischen Initiativen. Dies leuchtet auch ein, da zumindest

### 3 Methodik

**Tabelle 3.4:** Korrelationsmatrix der Polarisierungsmaße

|                      | $\alpha$      | Flache und Macy | $\rho_{Initiativen}$ | $P_{Cluster}$  | $\xi$          |
|----------------------|---------------|-----------------|----------------------|----------------|----------------|
| $\alpha$             |               | -0.2623         | <b>0.6660</b>        | 0.0986         | -0.1918        |
| Flache und Macy      | -0.2623       |                 | -0.1338              | <b>0.4013</b>  | 0.0573         |
| $\rho_{Initiativen}$ | <b>0.6660</b> | -0.1338         |                      | 0.1045         | <b>-0.5448</b> |
| $P_{Cluster}$        | 0.0986        | <b>0.4013</b>   | 0.1045               |                | <b>-0.3534</b> |
| $\xi$                | -0.1918       | 0.0573          | <b>-0.5448</b>       | <b>-0.3534</b> |                |

bei vollständigen Daten per Definition eine höhere Übereinstimmungsrate gleichzeitig mit einer höheren Korrelation zwischen den Fragestellungen einhergeht. Auch die etwas schwächere negative Korrelation zwischen dem Anteil frustrierter Kanten und der durchschnittlichen Korrelation zwischen Initiativen beziehungsweise der Clusterähnlichkeit erscheint plausibel. Ein niedriger Anteil frustrierter Kanten bedeutet starke Polarisierung, genauso wie eine hohe Korrelation zwischen Initiativen oder ein hoher Wert für das Polarisierungsmaß auf Basis der Clusterähnlichkeiten.

## 4 Verwendete Datensätze

Für die Untersuchungen wurde ein Datensatz, bestehend aus Datenbankauszügen, welche täglich seit Einführung des LiquidFeedback-Systems erstellt wurden, verwendet. Die Auszüge umfassen die Tage vom 13. August 2010 bis zum 10. März 2015. In diesem Zeitraum wurden 501707 Stimmen und 66574 Verbesserungsvorschläge zu den insgesamt 6655 Initiativen abgegeben. Die Nutzung des System schwankt sehr stark im Zusammenhang mit dem Auftreten von politischen Ereignissen. Zum Beispiel ist ein starker Anstieg aktiver User, sowohl nach dem Wahlerfolg der Berliner Piratenpartei in 2011, als auch vor den Bundestagswahlen in 2012 erkennbar. [KKH<sup>+</sup>15] Gegen Ende des Beobachtungszeitraums fällt die Aktivität immer stärker ab. Aktuell (September 2015) ist das System außer Betrieb genommen worden. Dies wurde damit begründet, dass die

...aktuelle Instanz von LiquidFeedback [...] nach Recherchen unseres derzeitigen Datenschutzbeauftragten (DSB) ohne jegliche Vorabprüfung des damaligen DSB in Betrieb gegangen [ist]. Wir sind somit gehalten, dieses System erst mal offline zu nehmen. [Car15]

Über die Art und Weise des weiteren Einsatzes sowie die generelle Nutzung von LiquidFeedback wird derzeit innerhalb der Partei diskutiert.

## 5 Polarisierung auf der Mikroebene - polarisierte Initiativen und polarisierende User

Wie in Abschnitt 3.3 beschrieben, kann anhand der Wölbung einer Häufigkeitsverteilung etwas über deren Bimodalität ausgesagt werden. Um einen genauen Wert der Wölbung zu definieren, ab dem man sagen kann, dass es sich um eine bimodale Verteilung handelt, wurden zunächst statistisch signifikante Werte für die Wölbung mit Hilfe einer Monte-Carlo-Simulation ermittelt. Hierbei handelt es sich um Werte, deren Wahrscheinlichkeit bei zufälliger Verteilung der Stimmen äußerst gering ist. Als Signifikanzniveau wurden fünf Prozent gewählt. Um diese Werte zu bestimmen, wurden für jede mögliche Anzahl von abgegebenen Stimmen eine Million zufällige Abstimmungsergebnisse erzeugt und deren Wölbung berechnet. Anschließend wurden die so errechneten Werte aufsteigend sortiert. Werte unterhalb der Fünf-Prozent-Schwelle wurden als signifikant betrachtet, sofern diese sich von dem kleinsten Wert oberhalb dieser Schwelle unterschieden. Diese Einschränkung stellt sicher, dass ein statistisch signifikanter Wert existiert. Die Auswertung ergab, dass es für Abstimmungen mit weniger als fünf abgegeben Stimmen keine Kurtosis mit einem p-Wert kleiner als 0,05 existiert. Dies liegt an der begrenzten Anzahl an Möglichkeiten, wie die fünf Stimmen verteilt sein können. Um polarisierte Initiativen zu identifizieren, wurde zunächst die Kurtosis für jede Abstimmung berechnet und anschließend mit dem errechneten signifikanten Wert verglichen. Im Anschluss wurden noch alle Abstimmungen aussortiert, welche zwar einen signifikanten Wert bei der Kurtosis aufweisen, deren Modi jedoch beide bei einem positiven oder beide bei einem negativen Wert liegen. Durch dieses Vorgehen wurden 303 Initiativen als polarisiert eingestuft. Diesen konnten 212 Initiatoren zugeordnet werden.

## 5.1 Polarisierende Texte

Neben den Initiativen wurden auch die Verbesserungs-/Änderungsvorschläge zu Initiativen auf Polarisierung untersucht. Es wurden 52 solcher *suggestions* als polarisiert klassifiziert. Um herauszufinden, welche charakteristischen Merkmale bei einem polarisierenden Text auftreten, wurde jeder Text daraufhin untersucht, ob er das Merkmal aufweist oder nicht. Im Anschluss wurde analog zu [NGKA11] eine logistische Regressionsanalyse durchgeführt, in der ein Mapping der Textmerkmale zu dem binären Wert „polarisiert“  $\in \{0, 1\}$  gelernt wurde. Wenn  $f_{i,j}$  das  $i$ -te Merkmal des  $j$ -ten Textes ist und  $polarisiert_j$  der binäre Wert, welcher angibt, ob der Text  $j$  polarisiert ist, dann lernt die logistische Regression unter folgendem Modell:

$$P(polarisiert_j|f) = \frac{1}{1 + \exp^{-(w_0 + \sum_i w_i f_{ij})}} \quad (5.1)$$

Die gelernten Koeffizienten  $w_i$  können als die Log-Odds des Merkmals  $j$  interpretiert werden. Die Odds sind der Quotient aus der Wahrscheinlichkeit, dass ein Text polarisiert und der Wahrscheinlichkeit, dass er nicht polarisiert.

$$w_i = \ln \left[ \frac{P(polarisiert_j|f_{ij} = 1)}{P(polarisiert_j|f_{ij} = 0)} \right] \quad (5.2)$$

Positive Koeffizienten kennzeichnen somit eine größere Wahrscheinlichkeit, dass ein Text polarisiert, sofern er dieses Merkmal aufweist. Tabelle 5.1 zeigt die verwendeten Merkmale und berechneten Regressionskoeffizienten. Anhand der gelernten Regressionskoeffizienten

| <b>Merkmal</b>     | <b>Regressionskoeffizienten</b> |
|--------------------|---------------------------------|
| Fragezeichen       | -0.2197                         |
| Ausrufezeichen     | -0.2763                         |
| Hyperlink          | 0.1191                          |
| Positiver Emoticon | -0.0193                         |
| Negativer Emoticon | 0.0414                          |
| Hashtag            | 0.0159                          |
| Paragraph          | 0.2174                          |
| Zitat              | -0.3979                         |
| Nummer             | -0.3349                         |

**Tabelle 5.1:** Textmerkmale und Regressionskoeffizienten

können interessante Feststellungen gemacht werden:



## 5 Polarisierung auf der Mikroebene - polarisierte Initiativen und polarisierende User

- Verbesserungsvorschläge, welche ein Zitat, eine Nummer oder ein Frage- bzw. Ausrufezeichen enthalten, sind mit einer höheren Wahrscheinlichkeit nicht polarisierend.
- Enthält ein Verbesserungsvorschlag dagegen ein Paragraphenzeichen oder einen Hyperlink, polarisiert er statistisch gesehen mit einer größeren Wahrscheinlichkeit.

## 5.2 Polarisierende Personen, Initiativen und Verbesserungsvorschläge

Bei der Analyse der polarisierten Initiativen fiel auf, dass diese eine höhere Beteiligung als die restlichen Initiativen aufweisen. Bei den polarisierten Initiativen stimmten durchschnittlich 205 Personen und bei den restlichen nur durchschnittlich 150 Personen ab. Auch die Autoren der polarisierten Initiativen sind erheblich aktiver als die restlichen User des Systems. Im Mittel gaben polarisierende Personen 368 Stimmen ab. Dem gegenüber stehen durchschnittlich 61 Stimmen der übrigen User. Dies stützt die von Abramowitz und Saunders aufgestellte Hypothese, dass Polarisierung für politische Partizipation förderlich ist. [Abr08]

Kling et al. [KKH<sup>+</sup>15] fanden in ihrer Studie zu Abstimmverhalten und Macht im LiquidFeedback-System der Piratenpartei heraus, dass mächtige Personen, sogenannte „Super-voter“, eine höhere Zustimmungsrate aufwiesen als Personen ohne Delegationen. Die Zustimmungsrate von Autoren polarisierter Initiativen liegt bei 71 Prozent. Bei allen anderen Personen liegt sie bei 61 Prozent. Es ist somit auch bei polarisierenden Personen eine höhere Zustimmungsrate beobachtbar, was auf die Macht von polarisierenden Usern (siehe Abschnitt 5.4) zurückzuführen ist.

Im Kontext von polarisierten Verbesserungsvorschlägen fiel auf, dass diese circa doppelt so häufig umgesetzt wurden (34 Prozent) wie alle anderen Verbesserungsvorschläge (16 Prozent).

## 5.3 Polarisierung und Popularität

Für den erfolgreichen Ausgang einer Abstimmung ist es für den Initiator wichtig, potenzielle „Dafür-Stimmer“ zu mobilisieren. Viele Piraten nutzen hierfür Twitter als Plattform, um auf neue Initiativen aufmerksam zu machen. Geht man davon aus, dass Personen sich

## 5 Polarisierung auf der Mikroebene - polarisierte Initiativen und polarisierende User

auf Twitter folgen, da sie ähnliche Interessen haben, stellt sich die Frage, ob diese auch ein ähnliches Abstimmverhalten aufweisen. Um dies herauszufinden, wurde das Wiki der Piratenpartei <https://wiki.piratenpartei.de/> bemüht. Viele Mitglieder haben dort ein öffentliches Profil, das bei manchen unter anderem auch Twitter-ID und LiquidFeedback-ID beinhaltet. Somit konnte ein Mapping von Twitter-ID zu LiquidFeedback-ID für 168 Personen erstellt werden. Anschließend wurden mit Hilfe der Twitter-API alle „Follower“ dieser Twitter-User abgefragt und ein Netzwerk von Piraten erstellt, welche sich untereinander „folgen“. Von den 168 Personen, die Twitter- und LiquidFeedback-ID angegeben haben, folgen sich 91 auf Twitter untereinander. Abbildung 5.1 zeigt das „Follower-Netzwerk“. Um die Übereinstimmung der Twitter-User zu messen, wurde aus allen von diesen Personen abgegebenen Stimmen (17026) Krippendorffs Alpha berechnet. Das Ergebnis  $\alpha = 0.18$  lässt nicht darauf schließen, dass Personen, welche auf Twitter „befreundet“ sind, häufiger übereinstimmen als der Durchschnitt (Vgl. Abbildung 6.1). Demnach ist anhand dieser recht kleinen Stichprobe nicht erkennbar, dass „Follower“ auf Twitter ähnlich abstimmen wie die Personen, denen sie folgen.

### 5.4 Polarisierung und Macht

Kling et al. haben in ihrer Studie [KKH<sup>+</sup>15] zu Abstimmverhalten und Macht, ebenfalls auf dem LiquidFeedback-Datensatz, einen Power-Index entwickelt, der die Macht einer einzelnen Person quantifiziert. Bei dem sogenannten „Beta2-Index“ handelt es sich um eine modifizierte Version des Shapley-Power-Index [SS54]. In dieser Arbeit wurde der Index dazu verwendet, die mächtigsten Personen zu identifizieren und ihre Übereinstimmung untereinander sowie die Rolle bei polarisierten Abstimmungen zu untersuchen. Krippendorffs Alpha unter den zehn Prozent der mächtigsten Personen beträgt 0,29. Dieser Wert ist höher als der  $\alpha$ -Wert unter den Twitter-Usern und auch etwas über dem Durchschnitt des Gesamtnetzwerkes. Die unteren 90 Prozent weisen ein Alpha von 0,23 auf. Dass große Einigkeit unter den Mächtigsten vorherrscht, legt ein Wert von 0,29 allerdings nicht nahe. Eine interessante Feststellung wurde bei der Betrachtung der Power-Indizes von Initiatoren polarisierender Initiativen gemacht. Die Auswertung ergab, dass 85 Prozent der Initiatoren von polarisierenden Initiativen einen höheren Power-Index als den Mittelwert (Median) aufweisen.

5 Polarisierung auf der Mikroebene - polarisierte Initiativen und polarisierende User

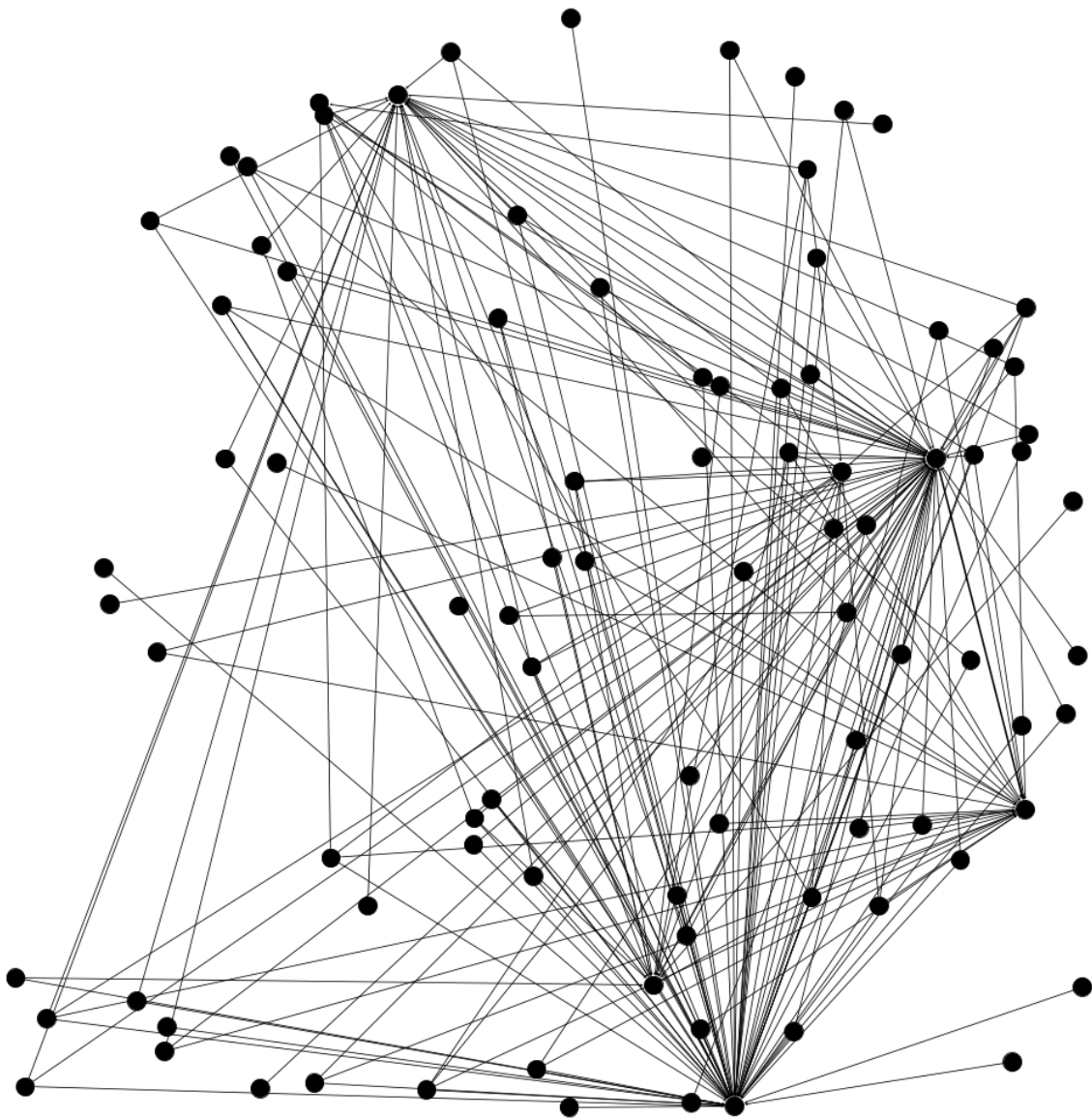


Abbildung 5.1: Twitter-Follower-Netzwerk

# 6 Polarisierung auf der Makroebene - Polarisierung als Prozess über die Zeit

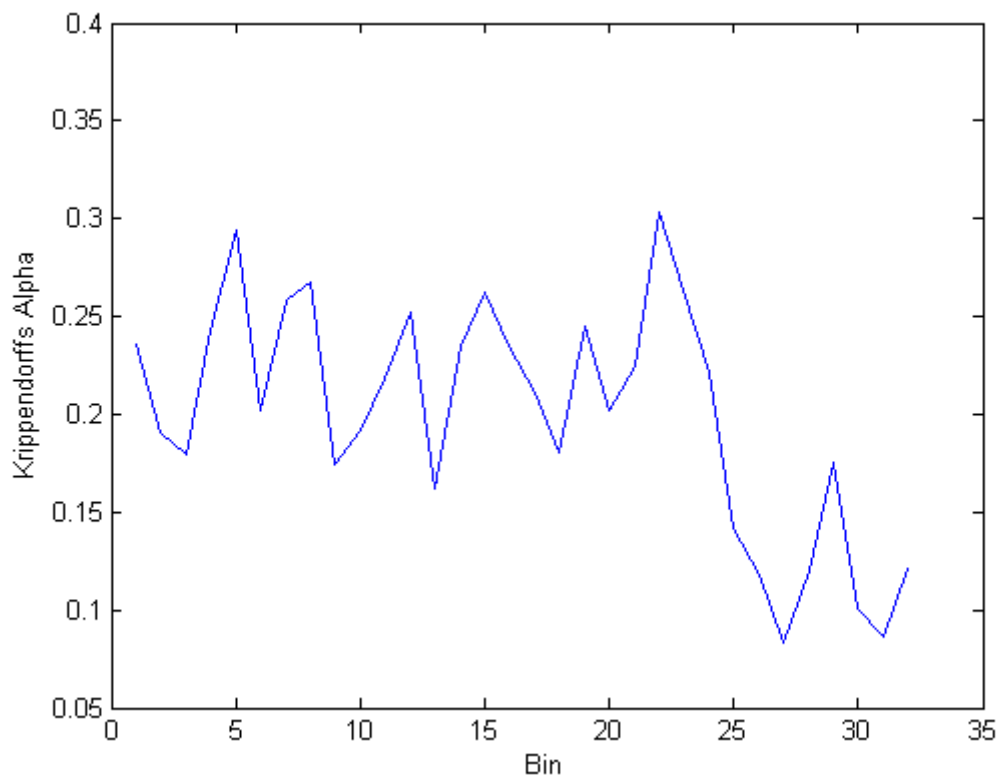
Nachdem im vorherigen Kapitel die Polarisierung einzelner Initiativen und User untersucht wurde, soll nun die Polarisierung im gesamten Netzwerk analysiert werden. Als Grundlage hierfür dienen alle abgegebenen Stimmen über alle Initiativen hinweg. Um die Abstimmungsdaten analysieren zu können, wurden sie aus der Datenbank extrahiert und in CSV-Dateien geschrieben. Das Format sieht folgendermaßen aus: Initiative,User,Stimme. Die Polarisierungskennzahlen der einzelnen Bins mit jeweils einhundert Initiativen wurden mit verschiedenen MATLAB- und Perl-Skripten (siehe Anhang A) berechnet und anschließend grafisch dargestellt.

## 6.1 Entwicklung von Krippendorffs Alpha im LiquidFeedback-Datensatz

Wie in Abschnitt 3.3 dargelegt, kann Krippendorffs Alpha als erstes Indiz für Polarisierung auf globaler Ebene herangezogen werden. Demnach wäre ein hoher positiver Alphawert ein Indiz dafür, dass sich die User weitgehend einig und nur selten gespalten sind. Negative und Werte nahe Null können auf eine inhärente Uneinigkeit hindeuten. Da es wenig sinnvoll ist, einen einzigen absoluten Alphawert als Grad der Polarisierung zu betrachten, wurden mehrere solcher Werte zu unterschiedlichen Zeitpunkten berechnet. Anhand der relativen Änderung von Krippendorffs Alpha über die Zeit kann eine Aussage über die Entwicklung der Polarisierung im System gemacht werden. Da der vorliegende Datensatz Datenbank-Snapshots von jedem Tag beinhaltet, war der erste Ansatz, Krippendorffs Alpha für jeden einzelnen Tag zu berechnen. Dazu wurde jeweils die Tabelle *vote* jedes Snapshots unter Verwendung eines Shellscripts in eine CSV-Datei geschrieben. Mit einem

## 6 Polarisierung auf der Makroebene - Polarisierung als Prozess über die Zeit

Perlsript wurde daraufhin Krippendorffs Alpha für jede Datei berechnet. Der Nachteil dieses Ansatzes liegt darin, dass es sich bei diesem Vorgehen um kumulierte Werte handelt und Änderungen gegen Ende kaum noch erkennbar sind. Um dieses Problem zu umgehen, wurde ein Daten-Binning von jeweils einhundert Initiativen vorgenommen und Krippendorffs Alpha für jeden einzelnen Bin berechnet. Dies ist möglich, da die Initiativen fortlaufend nummeriert werden und der Alpha-Wert der Bins somit annähernd die Entwicklung des Wertes über die Zeit widerspiegelt. Abbildung 6.1 zeigt die Entwick-

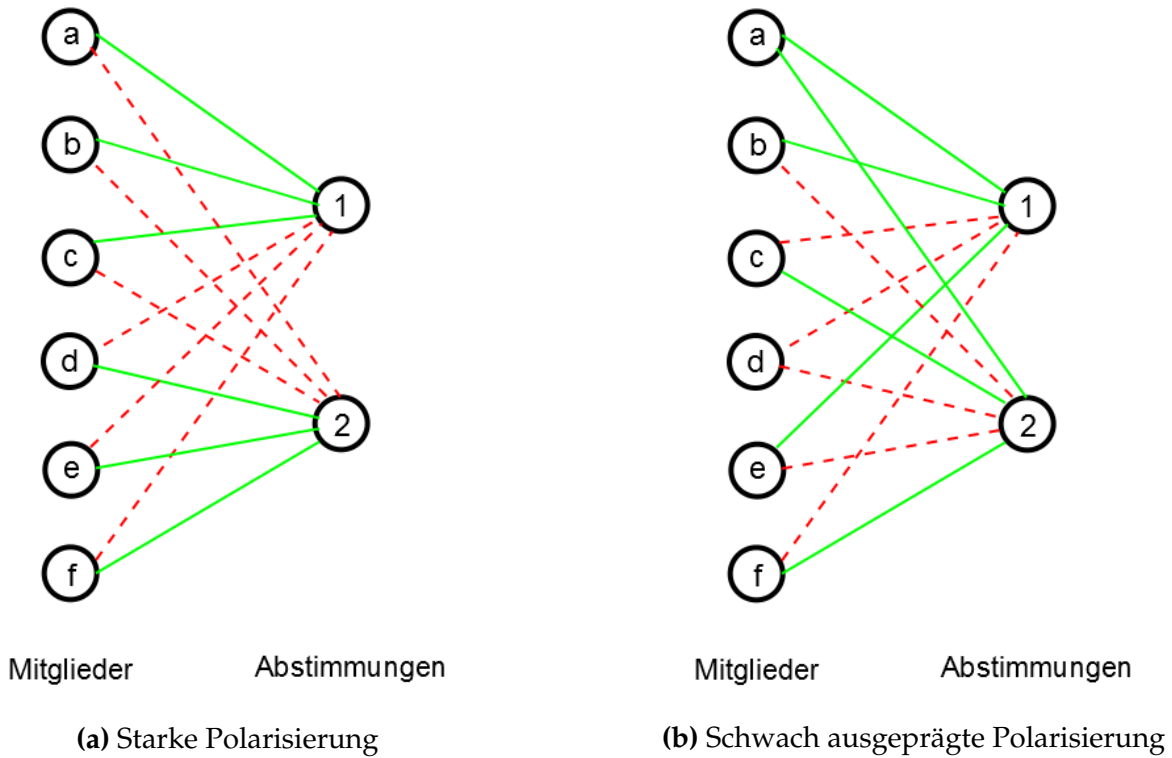


**Abbildung 6.1:** Krippendorffs Alpha im Zeitverlauf

lung von Krippendorffs Alpha. Es ist zu erkennen, dass die Zustimmungsrates zwischen den abstimmenden Personen gegen Ende deutlich abnimmt. Dies könnte ein Indiz dafür sein, dass die Polarisierung zugenommen hat. Krippendorffs Alpha macht allerdings nur eine Aussage darüber, in welchem Ausmaß die User übereinstimmen. Es lässt keine Rückschlüsse über Lagerbildung und wiederkehrende Meinungskonflikte zwischen denselben Usern zu. Abbildung 6.2 zeigt zwei Szenarien mit identischem Alpha. Beide weisen die gleiche Anzahl an Übereinstimmungen und Nichtübereinstimmungen auf, jedoch ist in Abbildung 6.2a perfekte Polarisierung zwischen den beiden Lagern ( $\{a, b, c\}$  und

## 6 Polarisierung auf der Makroebene - Polarisierung als Prozess über die Zeit

$\{d, e, f\}$ ) erkennbar. Dagegen ist in Abbildung 6.2b keine Struktur erkennbar und Übereinstimmungen beziehungsweise Nichtübereinstimmungen sind eher zufällig. Ein weiteres



**Abbildung 6.2:** Beide Szenarien weisen ein identisches Alpha auf  $(-0,1)$ . Abbildung 6.2a lässt im Gegensatz zu Abbildung 6.2b eine klare Aufteilung in zwei Lager erkennen.

Problem bei der Verwendung von Krippendorffs Alpha als Maß für Polarisierung ist, dass es unterschiedliche Ergebnisse für unterschiedlich große Lager liefert. Da in Wirklichkeit keine gleichgroßen Lager zu erwarten sind und auch Szenarien wie in Abbildung 6.3 nach Meinung des Autors als polarisiert angesehen werden, liegt hier eine Schwachstelle von Krippendorffs Alpha.

## 6 Polarisierung auf der Makroebene - Polarisierung als Prozess über die Zeit

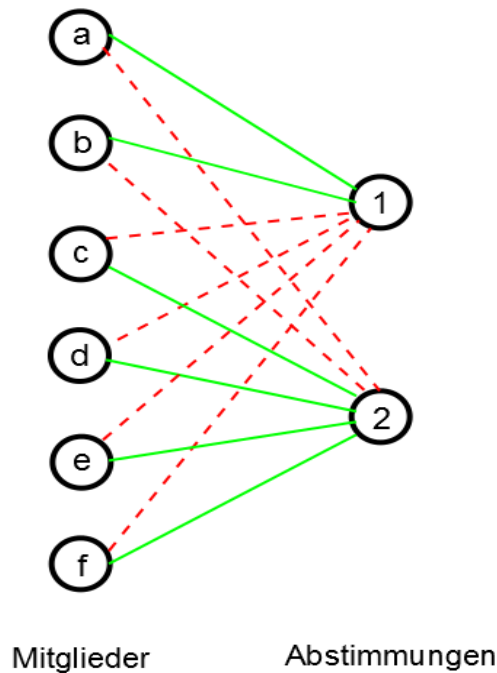


Abbildung 6.3: Perfekte Polarisierung zwischen zwei ungleich großen Lagern

### 6.2 Entwicklung des Polarisierungsmaßes von Fläche und Macy im LiquidFeedback-Datensatz

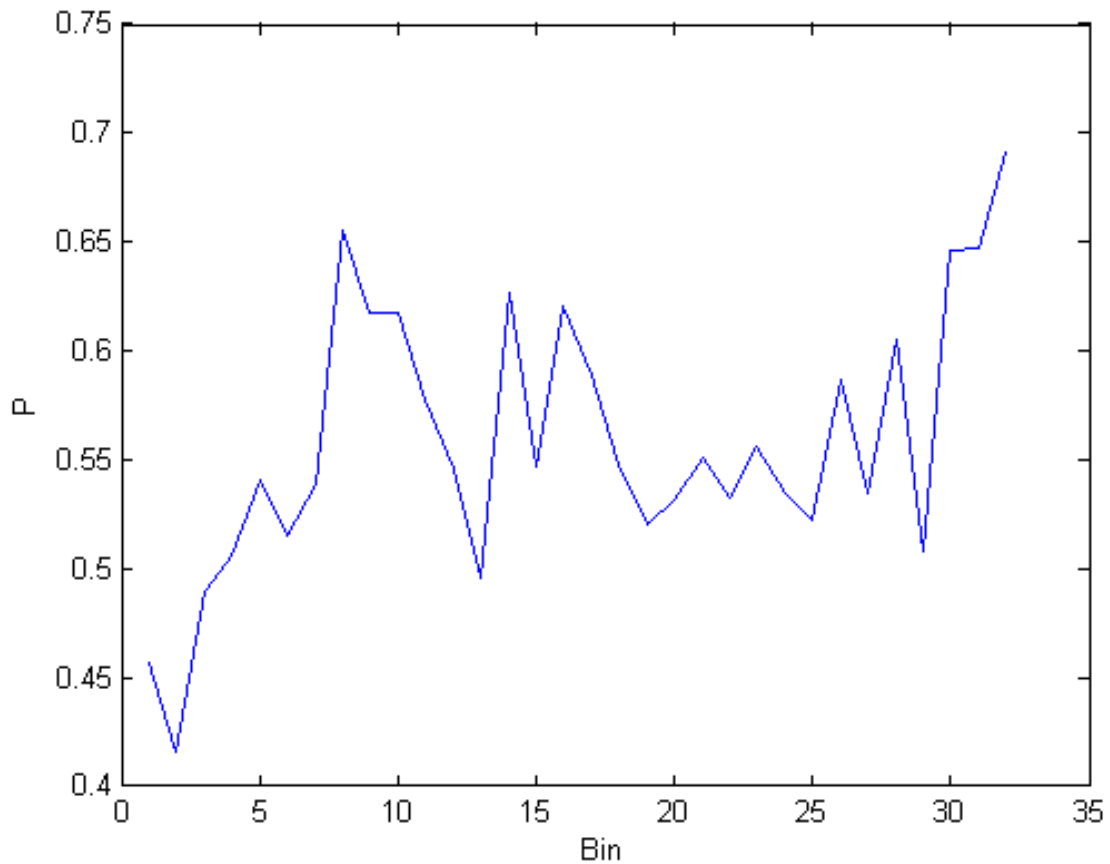
Das gleiche Problem hat man bei Verwendung des Maßes für Polarisierung von Fläche und Macy [FM11]:

$$P_t = \frac{1}{N(N-1)} \sum_{\substack{i=N, j=N \\ i \neq j}} (d_{ij,t} - \bar{d}_t)^2 \quad (6.1)$$

Hierbei handelt es sich um die Varianz der Meinungsabstände von jeweils zwei Personen. Die Entwicklung von  $P_t$  des LiquidFeedback-Datensatzes ist in Abbildung 6.4 dargestellt. Um  $P_t$  auf dem LiquidFeedback-Datensatz anwenden zu können, muss die Berechnung leicht modifiziert werden, so dass die Varianz bei unvollständigen Daten berechnet werden kann. Der Faktor  $\frac{1}{N(N-1)}$ , welcher die Anzahl von Personen-Paaren widerspiegelt, muss bei unvollständigen Daten die Anzahl von Paaren im Nenner enthalten, die mindestens bei einer gemeinsamen Abstimmung teilgenommen haben.

Das Maß eignet sich insofern besser zur Messung von Polarisierung als Krippendorffs Alpha, als es Lagerbildung grundsätzlich erkennt. Maximale Polarisierung  $P \rightarrow 1$  wird

## 6 Polarisierung auf der Makroebene - Polarisierung als Prozess über die Zeit



**Abbildung 6.4:** Entwicklung des Polarisierungsmaße nach Fläche und Macy [FM11]

erreicht, wenn die Grundgesamtheit in zwei gleichgroße Lager aufgeteilt werden kann, die sich bei allen Abstimmungen gegenüberstehen.  $P_t$  wird umso kleiner, je ungleicher die Grundgesamtheit auf die beiden Lager verteilt ist. Somit ist anhand von  $P_t$  nicht ersichtlich, ob sich die Distanz zwischen den Lagern beziehungsweise innerhalb der Lager verändert hat oder ob sich lediglich die relative Größe der Lager verändert hat. Um  $P_t$  korrekt interpretieren zu können, muss demnach die Größe der Lager bekannt sein.



### 6.3 Entwicklung der Abstimmungskorrelationen im LiquidFeedback-Datensatz

Um die *Constraints* [DiM96] beziehungsweise das *Issue Alignment* [BG08] zu bestimmen, wurde, wie in Abschnitt 3.3 beschrieben, die Kosinus-Ähnlichkeit zwischen den Initiativen verwendet. Abbildung 6.5 zeigt die durchschnittliche Abhängigkeit von zwei Initiativen im Zeitverlauf. Das Absinken des Wertes gegen Ende besagt, dass zwei Abstimmungen im

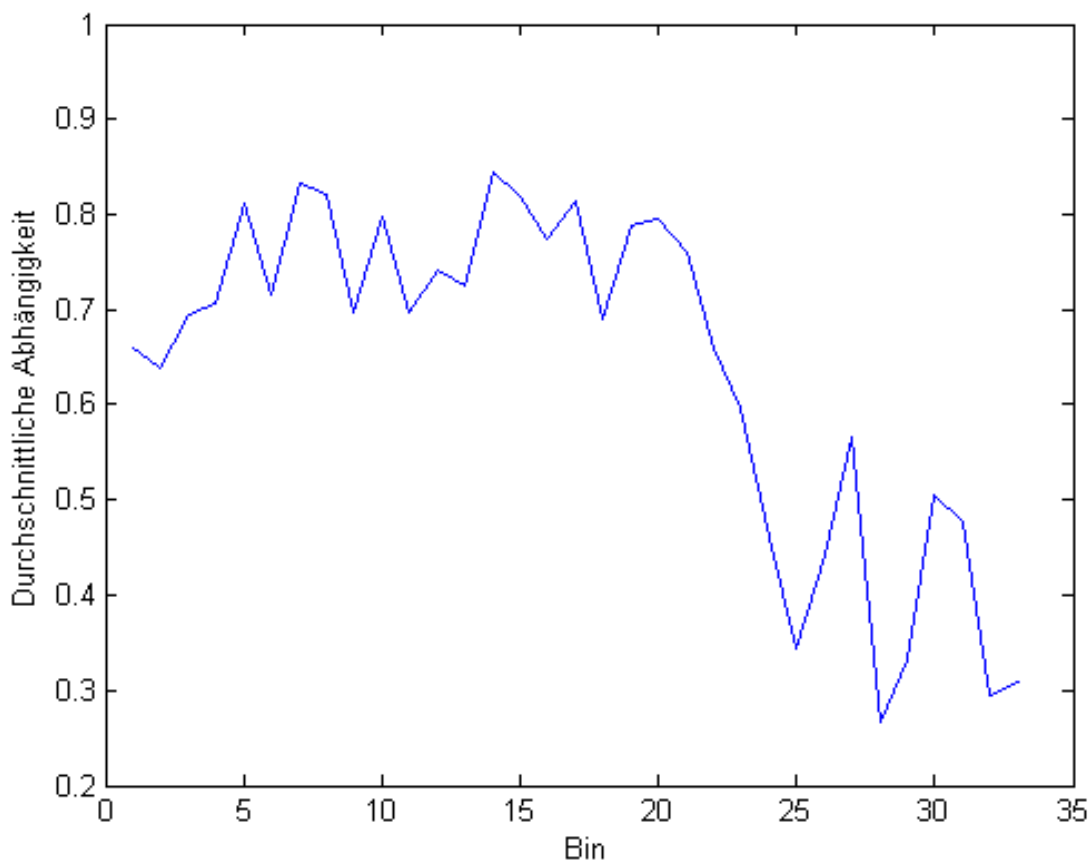
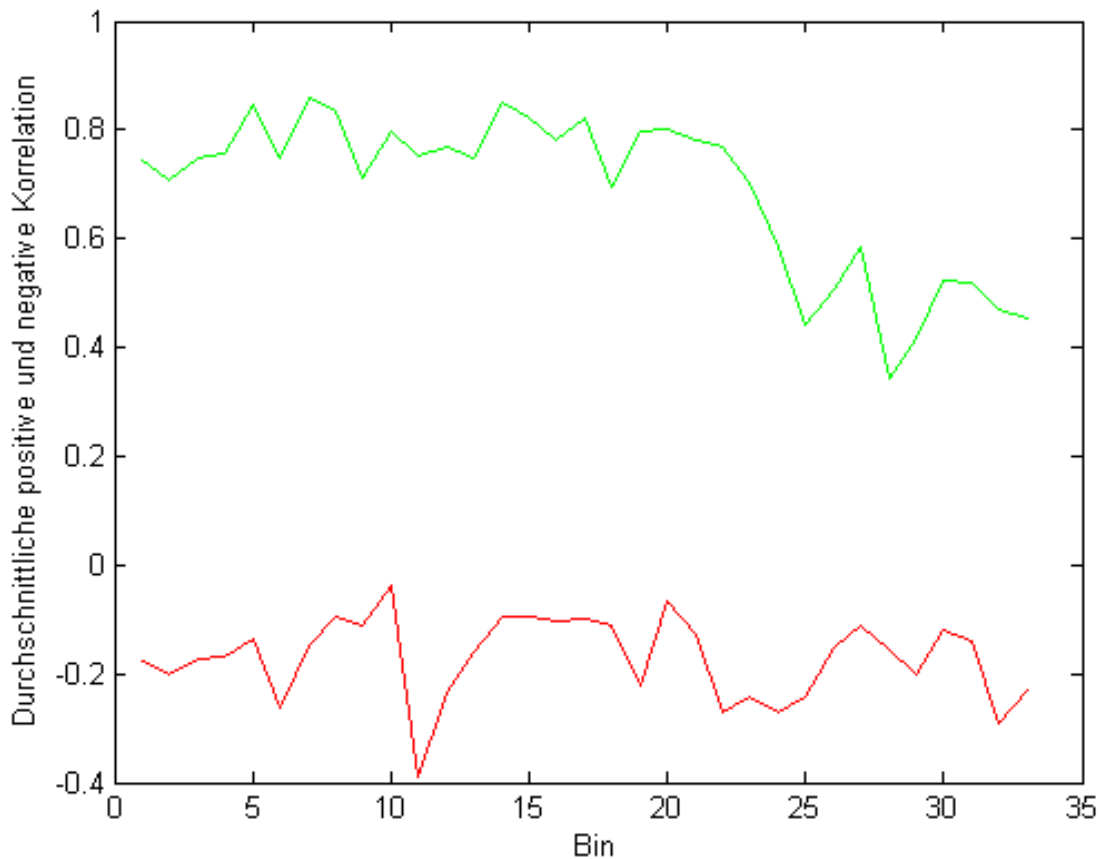


Abbildung 6.5: Issue Constraints

Durchschnitt nur noch schwach miteinander korreliert sind. Wie in Abschnitt 3.4 beschrieben, muss das jedoch nicht zwangsläufig bedeuten, dass die Polarisierung abgenommen hat. Folgendes Szenario soll den Sachverhalt verdeutlichen: Angenommen es gäbe zwei Gruppen von Abstimmungen. Abstimmungen innerhalb einer Gruppe weisen eine positive Korrelation und Abstimmungen unterschiedlicher Gruppen weisen eine negative

## 6 Polarisierung auf der Makroebene - Polarisierung als Prozess über die Zeit

Korrelation auf. Sinkt nun die durchschnittliche Korrelation zwischen allen Abstimmungen, kann dies unter anderem daraus resultieren, dass die negativen Korrelationen stärker geworden sind. Der Wert der negativen Korrelation ist demnach kleiner und die durchschnittliche Korrelation folglich auch. Da starke Polarisierung sowohl bei einem sehr hohen positiven, als auch einem sehr niedrigen negativen Wert vorliegen kann, macht es Sinn, die Entwicklung von negativ und positiv korrelierten Abstimmungen getrennt voneinander zu betrachten. Wie Abbildung 6.6 zeigt, resultiert das Absinken der durchschnittlichen Korrelation aus der schwächer werdenden Korrelation unter den positiv korrelierten Abstimmungen und annähernd gleichbleibender Korrelation unter den negativ korrelierten Abstimmungen. So gesehen deutet die Entwicklung der Korrelationen zwischen den Ab-



**Abbildung 6.6:** Entwicklung der durchschnittlichen positiven und negativen Korrelation

stimmungen darauf hin, dass die Polarisierung gegen Ende des Beobachtungszeitraums abgenommen hat.

## 6.4 Entwicklung der Clusterähnlichkeiten im LiquidFeedback-Datensatz

Um festzustellen, ob es zu Lagerbildung unter den Mitgliedern der Piratenpartei kommt, wurde eine Clusteranalyse durchgeführt. Hierzu wurde eine Ähnlichkeitsmatrix auf Basis der abgegebenen Stimmen erstellt, mit deren Hilfe ein Clustering vorgenommen werden kann. Als Ähnlichkeitsmaß wurde die Kosinus-Ähnlichkeit gewählt. Sie ist wie folgt definiert:

$$\text{Kosinus-Ähnlichkeit} = \frac{\langle a, b \rangle}{\|a\| \times \|b\|} \quad (6.2)$$

Wie in Abschnitt 3.2 beschrieben, werden beim Spectral Clustering die Eigenvektoren der Laplace-Matrix zum Auffinden von Clustern genutzt. [KSL<sup>+</sup>10][VL07][NJW<sup>+</sup>02] Da die Anzahl identifizierbarer Cluster im LiquidFeedback-Datensatz nicht gegeben war, musste diese zuerst herausgefunden werden. Um die am besten geeignete Anzahl von Clustern für den Algorithmus zu bestimmen, wurde der Silhouettenkoeffizient verwendet.

„Der Silhouetten-Koeffizient ist ein Maß für die Güte eines Clustering, das unabhängig von der Anzahl  $k$  der Cluster ist. Je größer der Wert  $s_C$ , desto besser ist das Clustering.“ [ES13]

Gehört das Objekt  $o$  zum Cluster  $A$ , so ist die Silhouette von  $o$  definiert als:

$$S(o) = \begin{cases} 0 & \text{falls } dist(A, o) = 0, \\ \frac{dist(B, o) - dist(A, o)}{\max\{dist(A, o), dist(B, o)\}} & \text{sonst.} \end{cases} \quad (6.3)$$

$dist(A, o)$  bezeichnet den durchschnittlichen Abstand des Objekts  $o$  zu „seinem“ Cluster  $C$  und  $dist(B, o)$  bezeichnet die durchschnittliche Distanz eines Objekts  $o$  zum „nächstgelegenen“ Cluster  $B$ . Die Werte der Silhouette eines Objekts sind normiert und liegen zwischen  $-1$  und  $1$ . Ein Wert von  $0$  kann so interpretiert werden, dass das Objekt ungefähr zwischen seinem zugeordneten und dem nächstgelegenen Nachbarcluster liegt.  $S(o) \approx 1$  bedeutet, dass der Abstand zu seinem zugeordneten Cluster wesentlich kleiner ist als der zu seinem nächstgelegenen Cluster. Folglich besagt  $S(o) \approx -1$ , dass der Abstand des

## 6 Polarisierung auf der Makroebene - Polarisierung als Prozess über die Zeit

Objekts zu seinem Nachbarcluster wesentlich kleiner ist als zu dem Cluster, zu welchem es zugeteilt wurde. Der Silhouettenkoeffizient  $S(o)$  ist definiert als das arithmetische Mittel aller Silhouetten [ES13]:

$$S(o) = \frac{1}{n_c} \sum_{o \in C} S(o) \quad (6.4)$$

Wie in Abschnitt 3.2 erläutert, kann im Fall von zwei Clustern die Partitionierung der Elemente anhand des Vorzeichens vorgenommen werden. Möchte man eine Clusteranalyse mit mehr als zwei Clustern durchführen, kann dies ebenfalls mit Hilfe der kleinsten Eigenvektoren erfolgen. Die Partitionierung ist jedoch nicht auf Anhieb anhand der Vorzeichen ablesbar. Der im Folgenden vorgestellte von NG, Jordan und Weiss entwickelte Algorithmus für Spectral Clustering ist für jede Anzahl  $k$  von Clustern geeignet. [NJW<sup>+</sup>02]

Input: Ähnlichkeits-Matrix  $S \in \mathbb{R}^{n \times n}$ , Anzahl Cluster  $k$

- Berechne die normalisierte Laplace-Matrix  $L_{sym}$ .
- Berechne die ersten  $k$  Eigenvektoren  $u_1, \dots, u_k$  von  $L_{sym}$ .
- Sei  $U \in \mathbb{R}^{n \times k}$  die Matrix, welche die Vektoren  $u_1, \dots, u_k$  als Zeilen enthält.
- Erstelle die Matrix  $T \in \mathbb{R}^{n \times k}$  von  $U$  und normalisiere die Zeilen zu Norm 1. Setze hierzu  $t_{ij} = U_{ij} / (\sum u_{ik}^2)^{1/2}$ .
- Für  $i = 1, \dots, n$ , sei  $y_i \in \mathbb{R}^k$  der korrespondierende Vektor der  $i$ -ten Zeile aus  $T$ .
- Partitioniere die Punkte  $(y_i)_{i=1, \dots, n}$  mit dem k-Means-Algorithmus in Cluster  $C_1, \dots, C_k$

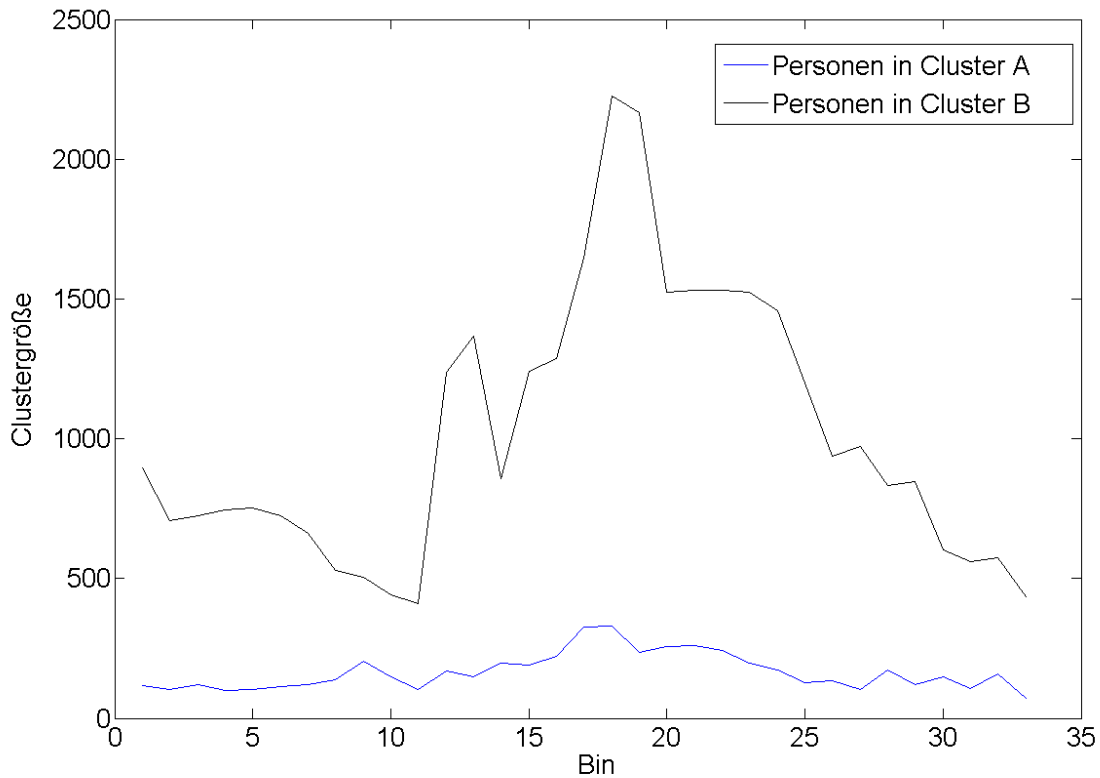
Output: Cluster  $A_1, \dots, A_k$  mit  $A_i = \{j | y_j \in C_i\}$ .

Die Implementierung in Matlab ist in Abschnitt 7 unter „Matlab-Skript SilhouetteTesting“ und „Matlab-Skript Spectral Clustering für  $k > 2$ “ zu finden.

Die Analyse der Silhouettenkoeffizienten für verschiedene Anzahlen von Clustern ergab, dass ein bestmögliches Ergebnis bei zwei Clustern erreicht wird. Bei der Analyse fiel außerdem auf, dass der Koeffizient im Laufe der Zeit abnimmt und nie einen Wert größer als

## 6 Polarisierung auf der Makroebene - Polarisierung als Prozess über die Zeit

0,5 annimmt, was als schwache Struktur interpretiert werden kann. [KR09] Abbildung 6.7 stellt die Entwicklung der Größen der beiden Cluster dar.



**Abbildung 6.7:** Größen der beiden Cluster

Aus der Grafik ist ersichtlich, dass die Größen der beiden Gruppen über den gesamten Zeitraum sehr unterschiedlich sind und eine kleine Randgruppe existiert, die einer großen Mehrheit gegenübersteht.

Zunehmende Polarisierung würde man bei einer Clusteranalyse daran erkennen, dass die (Meinungs-)Distanz von Personen innerhalb eines Clusters im Laufe der Zeit abnimmt und die (Meinungs-)Distanz von Personen unterschiedlicher Cluster zunimmt. Dies ist im LiquidFeedback-Datensatz nicht zu erkennen. Abbildung 6.8 zeigt, dass sich die Kosinus-Ähnlichkeit innerhalb der Cluster sowie zwischen den Clustern nur marginal verändert. Es ist kein Trend in die eine oder andere Richtung erkennbar. Um das Ergebnis mit den anderen Polarisierungsmaßen vergleichen zu können, wurde  $P_{Cluster}$  wie folgt definiert:

$$P_{Cluster,t} = ((sim_{A,t} + sim_{B,t})/2) - sim_{between,t} \quad (6.5)$$

## 6 Polarisierung auf der Makroebene - Polarisierung als Prozess über die Zeit

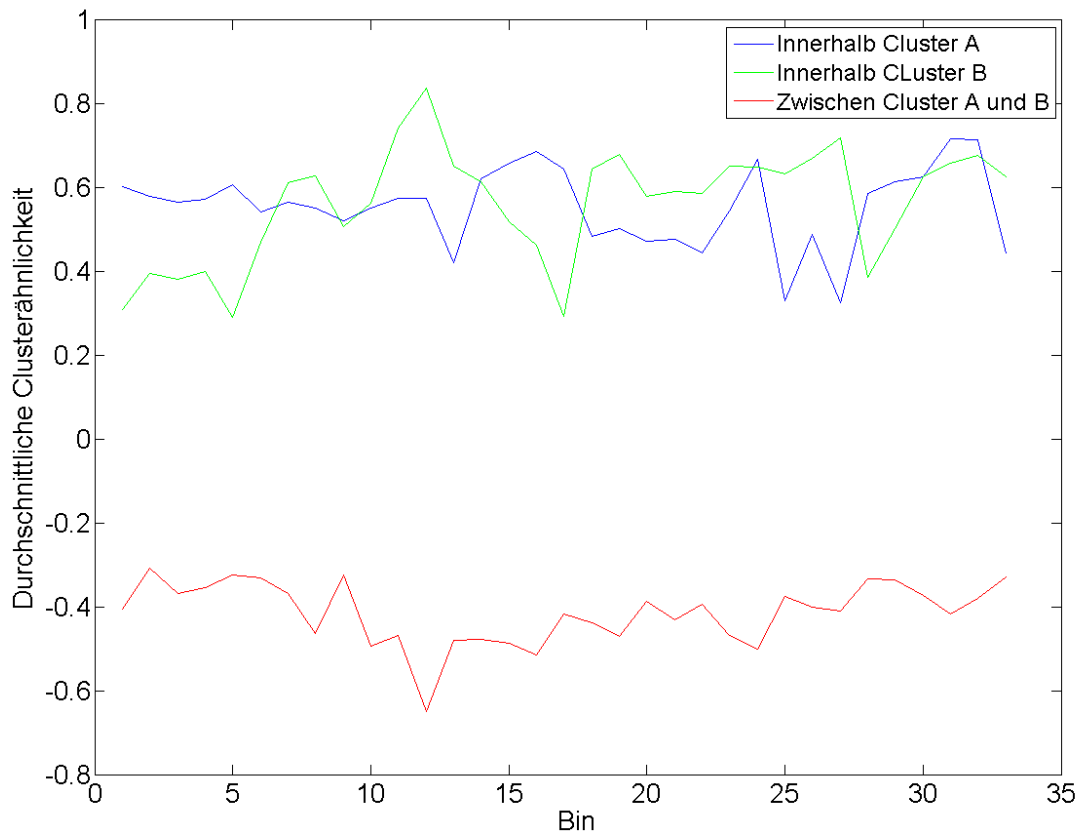


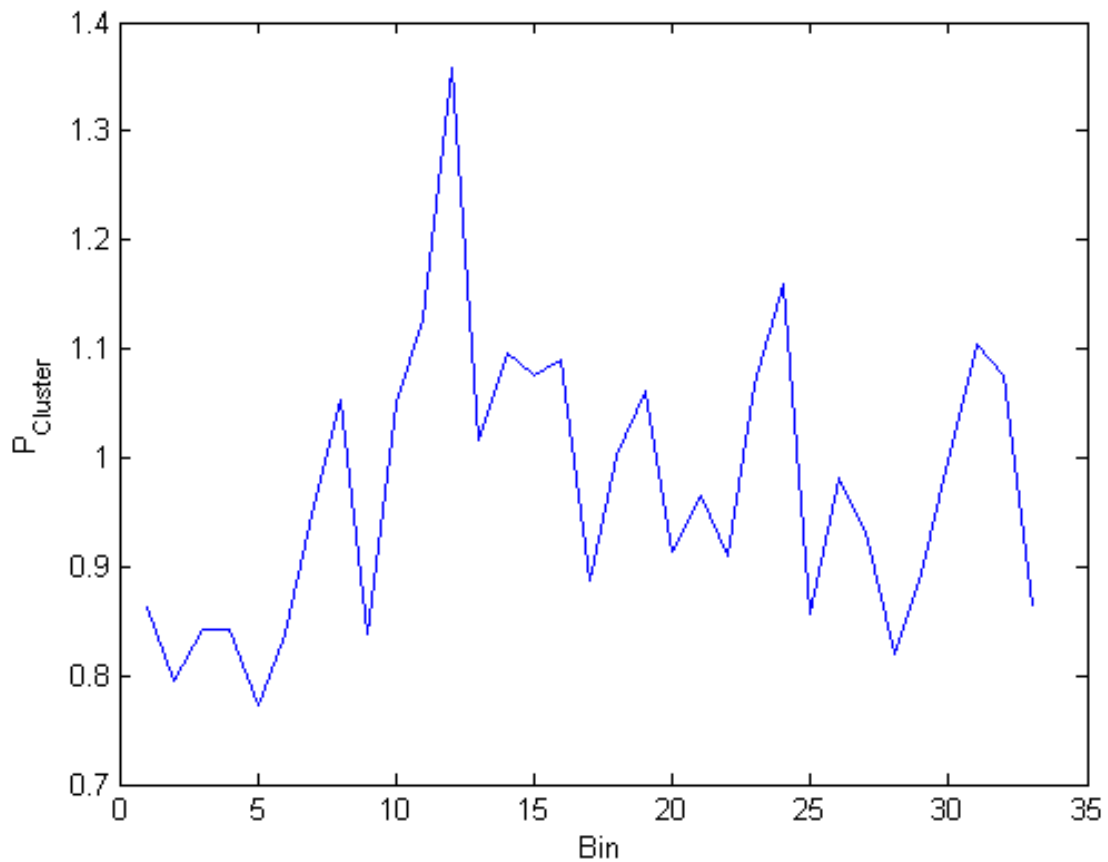
Abbildung 6.8: Kosinus-Ähnlichkeit innerhalb und zwischen den Clustern

Abbildung 6.9 zeigt die Entwicklung von  $P_{Cluster}$  über die Zeit.

### 6.5 Entwicklung des Anteils frustrierter Kanten im LiquidFeedback-Datensatz

Ein weiterer Ansatz die Polarisierung zu messen, war, den algebraischen Konflikt zu berechnen. Wie in Abschnitt 3.2 beschrieben, ist der kleinste Eigenwert  $\lambda_1 = 0$ , wenn der Graph balanciert ist. [Hou05] Die Balance äußert sich darin, dass die Knoten in zwei Gruppen aufgeteilt werden können, so dass alle positiven Kanten Knoten innerhalb derselben Gruppe verbinden und alle negativen Kanten zwischen Knoten unterschiedlicher Gruppen verlaufen (siehe Abbildung 3.2a). [Har53] Kunegis bezeichnet den kleinsten

## 6 Polarisierung auf der Makroebene - Polarisierung als Prozess über die Zeit



**Abbildung 6.9:** Polarisierungsmaß basierend auf Clusterähnlichkeiten

Eigenwert  $\lambda_1$  als  $\zeta$  und definiert es als die Menge des Konflikts in einem Graphen. [Kun14] Bezogen auf den Grad der Polarisierung, ist ein balancierter Graph mit  $\zeta = 0$  maximal polarisiert und je größer  $\zeta$  ist, desto mehr Konflikte existieren und desto schwächer ist die Polarisierung ausgeprägt. Als Eigenschaft von  $\zeta$  wurde darüber hinaus festgestellt, dass das Hinzufügen einer Kante, egal welchen Gewichts, den algebraischen Konflikt nur erhöhen kann und nicht verringern. Da die Bins des Datensatzes unterschiedlich viele Kanten aufweisen (nicht jedes Mitglied stimmt bei jeder Initiative ab), ist  $\zeta$  als erster Anhaltspunkt ganz gut geeignet. Es muss jedoch, um vergleichbar zu sein, mit der Anzahl der Kanten normalisiert werden. Zu diesem Zweck wurde der *frustration index* bemüht:

„The frustration index of a signed graph  $\Sigma$ ,  $f(\Sigma)$ , is the minimum number of edges whose deletion yields a balanced signed graph. If  $\Sigma$  is balanced then

## 6 Polarisierung auf der Makroebene - Polarisierung als Prozess über die Zeit

$f(\Sigma) = 0.$  [Bow12]

Kunegis hat in [Kun15] ein Maß entwickelt, welches den Anteil der frustrierten Kanten an der Gesamtzahl misst. Der Hintergrund hierbei ist jedoch ein anderer. Es wurden unsignierte Graphen betrachtet und der Anteil der frustrierten Kanten wurde als Maß der *non-bipartivity* definiert. Folglich ist die Definition einer frustrierten Kante in [Kun15] auch eine andere. Hat man eine Bipartition  $V = V_1 \cup V_2$  gegeben, ist eine frustrierte Kante eine Kante, welche zwei Knoten in  $V_1$  oder zwei Knoten in  $V_2$  verbindet. Bei Betrachtung der Polarisierung handelt es sich um eine frustrierte Kante, falls sie negativ ist und innerhalb eines Clusters verläuft oder falls sie positiv ist und zwischen zwei Clustern verläuft. Das Maß

$$b_f(G) = \frac{f}{|E|} \quad (6.6)$$

kann, leicht modifiziert, dazu verwendet werden, den *frustration index* eines signierten Graphen zu bestimmen. Ausgehend von unserer Aufteilung in zwei Cluster  $V = V_1 \cup V_2$ , definieren wir den charakteristischen Vektor  $x \in \mathbb{R}^{|V|}$  als

$$x_u = \begin{cases} +1/2 & \text{falls } u \in V_1 \\ -1/2 & \text{falls } u \in V_2 \end{cases} \quad (6.7)$$

Die Frustration  $f$  des signierten Graphen ist somit gegeben durch:

$$f = \frac{1}{2} \sum_{(u,v) \in E} (x_u - w(u,v)x_v)^2 = \frac{1}{2} x^T \bar{L} x \quad (6.8)$$

$\bar{L} = \bar{D} - A$  ist die *signed Laplacian matrix* des zugrunde liegenden Graphen und  $w(u,v) \in \{-1, +1\}$  ist das Gewicht der Kante. Die minimale Anzahl an frustrierten Kanten  $f$  ist damit gegeben durch:

$$f = \min_{x \in \{\pm 1/2\}^{|V|}} \frac{1}{2} x^T \bar{L} x \quad (6.9)$$

Durch Lockerung der Bedingung  $x \in \{\pm 1/2\}^{|V|}$  kann  $f$  als Funktion von  $\bar{L}$ s kleinstem Eigenwert ausgedrückt werden. In Anlehnung an [Kun15] ergibt sich  $f$ , basierend auf der Tatsache, dass die Norm aller Vektoren  $x \in \{\pm 1/2\}^{|V|}$  gleich  $\sqrt{|V|/4}$  ist und dass der kleinste Eigenwert dem kleinsten Rayleigh-Quotient entspricht, wie folgt:

$$\frac{2f}{|V|/4} \approx \min_{x \neq 0} \frac{x^T \bar{L} x}{\|x\|^2} = \lambda_{\min}[\bar{L}] \quad (6.10)$$

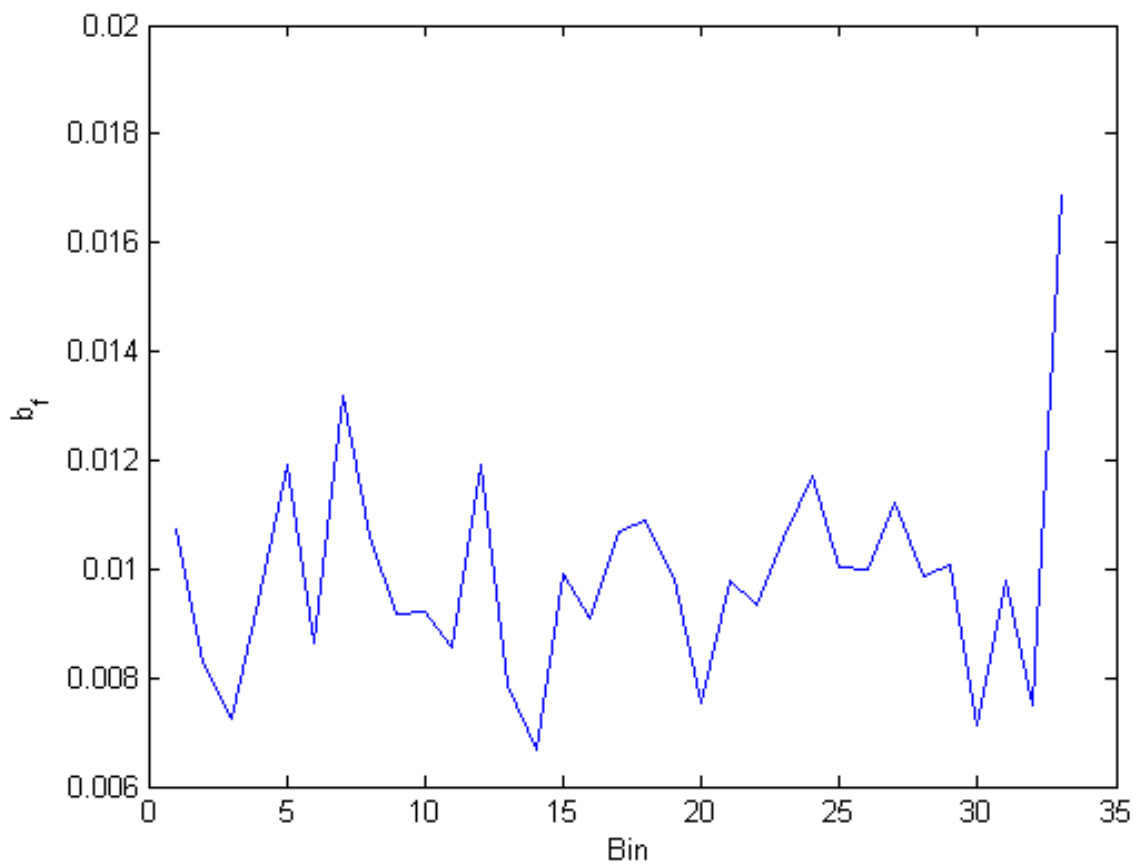


## 6 Polarisierung auf der Makroebene - Polarisierung als Prozess über die Zeit

Somit kann der Frustrationsindex  $b_f$  mit

$$b_f \approx \frac{|V|}{8|E|} \lambda_{\min}[\bar{L}] \quad (6.11)$$

approximiert werden. Wie auch bei den vorherigen Polarisierungsmaßen, ist bei  $b_f$  kein Trend in die eine oder die andere Richtung erkennbar. (siehe Abbildung 6.10)



**Abbildung 6.10:** Anteil frustrierter Kanten im Graph

## 7 Zusammenfassung

Die vorliegende Arbeit hat gezeigt, dass es nicht trivial ist, den Grad der Polarisierung in einer Gesellschaft zu quantifizieren. Generell lassen sich Aussagen über die Entwicklung eines Wertes zur Messung von Polarisierung einfacher treffen als Aussagen über absolute Werte. Um Polarisierung ganzheitlich zu untersuchen, müssen ihre unterschiedlichen Facetten betrachtet werden. Die in der Literatur gefundenen Maße eignen sich gut für die dort verwendeten Datensätze der jeweils durchgeführten Umfragen. Bei Anwendung auf den LiquidFeedback-Datensatz traten jedoch einige Probleme zu Tage, die auf die Unvollständigkeit der Daten zurückzuführen sind. Deshalb wurden zwei, nach Kenntnisstand des Autors, neuartige Methoden entwickelt, diesem Problem zu begegnen. Die erfolgversprechendste der vorgestellten Methoden zur Messung von Polarisierung einer Menge von Personen ist die in Abschnitt 6.4 erläuterte Methode der Clusterähnlichkeiten. Mit ihr lassen sich alle notwendigen Eigenschaften messen und sie ist für eine beliebige Anzahl von Gruppen anwendbar. Sie ist invariant gegenüber unterschiedlich großer Datensätze und funktioniert auch bei unvollständigen Daten. Zusammenfassend lässt sich in Bezug auf die Polarisierung im LiquidFeedback-Datensatz sagen, dass es bei den Polarisierungsmaßen aus der spektralen Graphentheorie (Abschnitt 6.4 und Abschnitt 6.5) keine wesentlichen Veränderungen im Zeitverlauf gab. Somit konnte die eingangs aufgestellte These, dass die zurückgehende Aktivität im System möglicherweise in einer zunehmenden Polarisierung begründet liegt, nicht verifiziert werden. Betrachtet man das Polarisierungsmaß nach Fläche und Macy, sieht es so aus, als ob die Polarisierung anfangs ansteigt und dann weitgehend konstant bleibt. Wie in Abschnitt 6.2 und Unterabschnitt 3.4.2 erläutert, kann dies jedoch aus unterschiedlich großen Gruppen resultieren. Definiert man Polarisierung in der Art, dass sie maximal für zwei gleich große Gruppen sein soll, macht dies Sinn. In der vorliegenden Arbeit wurde Polarisierung jedoch so verstanden, dass sie auch bei unterschiedlich großen Gruppen maximal sein kann.

Um die Polarisierung einzelner Abstimmungen zu bestimmen, erwiesen sich die in [DiM96] vorgestellten Maße Varianz und Wölbung der Stimmenverteilung als hilfreich.

## 7 Zusammenfassung

Mittels einer Monte-Carlo-Simulation konnten signifikante Werte der beiden Kennzahlen für unterschiedliche Grundgesamtheiten bestimmt werden. Die so als polarisiert klassifizierten Initiativen, Verbesserungsvorschläge und Autoren wurden daraufhin genauer untersucht. Interessante Feststellungen in diesem Zusammenhang waren, dass die Autoren polarisierender Initiativen in 85 Prozent der Fälle einen höheren Power-Index aufweisen als den Median und dass Personen, welche sich auf Twitter folgen, im Durchschnitt nicht häufiger übereinstimmen als der Durchschnitt aller Personen. Außerdem hat die Untersuchung der Verbesserungsvorschläge zu Initiativen gezeigt, dass polarisierte Vorschläge ungefähr doppelt so häufig umgesetzt werden wie alle anderen.

In zukünftigen Arbeiten zum Thema Polarisierungsmaße könnte man bei delegativen Plattformen überlegen, die Delegationsgewichte in die Polarisierungsmaße miteinzubeziehen. Damit ist gemeint, dass zum Beispiel eine Aufteilung in zwei Lager mit 10 Personen, die jeweils 100 Delegationen bekommen haben und 100 Personen ohne Delegationen denselben Wert haben sollte wie eine Fünfzig-Fünfzig-Aufteilung von Personen ohne Delegationen. Des Weiteren wäre es für die Untersuchung von Polarisierung in Gesellschaften mit mehr als zwei Gruppen interessant, eine Generalisierung des Polarisierungsmaß *Anteil der frustrierten Kanten* zu erarbeiten.

# A Anhang

## A.1 Quelltexte

### Perl-Skript Krippendorffs Alpha

```
1  #!/usr/bin/perl
2  # Datei: parse_Table_opinion.pl
3
4  use strict;
5  use warnings;
6  use IO::File;
7
8  my $initiative;
9  my $suggestion;
10 my $member;
11 my $degree = 0;
12 my $fullfilled;
13 my $previous;
14 my $previous_initiative;
15 my $previous_issue;
16 my $tmp;
17 my $array_length;
18 my $file_extension = ".csv";
19 my $file;
20 my $first = 1;
21 my $sorted = "sorted";
22 my $first_dimension;
23 my $second_dimension;
24 my $i;
25 my $j;
26 my $max_member = 0;
27 my $max_initiative = 0;
28 my $max_prev_member = 0;
```

```
29 my $sum_correlation = 0;
30 my $number_of_edges = 0;
31
32 my $sum = 0;
33 my $count = 0;
34 my $polarized = 0;
35 my @reliabilityMatrix;
36 my @coincidenceMatrix;
37 my @differenceMatrix;
38 my @mu;
39 my @counter;
40 my @n;
41 my $numerator;
42 my $denominator;
43 my $alpha = 0;
44 my $first_element;
45 my $countvariable = 1;
46
47
48 open (KRIALPHABINNED, ">alphaNominalBeta2NoPower$file_extension") or die ">
    Konnte die OUTPUT-Datei nicht oeffnen";
49
50
51 open(TABLE, "<votes_beta2_noPower$file_extension") or die $!;
52
53 while (<TABLE>){
54
55
56     #einzelne Zeile der Datei einlesen
57     my $zeile = TABLE->getline;
58     my $zeile = $_;
59
60     #CSV-Datei splitten mit Trennzeichen ;
61     my ($initiative, $member, $degree) = split(/,/, $zeile);
62     if ($first == 1){
63         $previous = $initiative;
64     }
65     if($initiative eq $previous){
66         if($degree > 0){
67             $reliabilityMatrix[$initiative][$member] = 1;
68         }
69         if($degree < 0){
70             $reliabilityMatrix[$initiative][$member] = 0;
71         }
72         if($degree == 0){
73
74         }
75         if($member > $max_member){ $max_member = $member}
76         if($initiative > $max_initiative){ $max_initiative = $initiative}
77         $count++;
78         $previous_initiative = $initiative;
79         $previous = $initiative;
80         $first = 0;
```

```

81     } else{
82         if($degree > 0){
83             $reliabilityMatrix[$initiative][$member] = 1;
84         }
85         if($degree < 0){
86             $reliabilityMatrix[$initiative][$member] = 0;
87         }
88         if($degree == 0){
89             }
90         }
91         if($member > $max_member){ $max_member = $member}
92         if($initiative > $max_initiative){ $max_initiative = $initiative}
93         $mu[$previous] = $count;
94         $count = 1;
95         $previous = $initiative;
96         $previous_initiative = $initiative;
97         $first = 0;
98     }
99 }
100 $mu[$previous] = $count;
101 close(TABLE);
102
103 $first_dimension = $max_member; #scalar @{$reliabilityMatrix[0]};
104 $second_dimension = $max_initiative; #scalar @{$reliabilityMatrix[1]};
105 $counter[0] = 0;
106 $counter[1] = 0;
107 $coincidenceMatrix[0][0] = 0;
108 $coincidenceMatrix[0][1] = 0;
109 $coincidenceMatrix[1][0] = 0;
110 $coincidenceMatrix[1][1] = 0;
111 for ($i=0; $i <= $second_dimension; $i++) {
112     for($j=0; $j <= $first_dimension; $j++){
113         if(defined $reliabilityMatrix[$i][$j]){
114             $counter[$reliabilityMatrix[$i][$j]]++;
115         }
116     }
117
118     if(defined $mu[$i] && $mu[$i]>1){
119
120         if($counter[0] > 1) {$coincidenceMatrix[0][0] += ($counter[0]*($counter[
121             [0]-1))/(($mu[$i]-1)}
122         $coincidenceMatrix[0][1] += ($counter[0] * $counter[1])/(($mu[$i]-1);
123         $coincidenceMatrix[1][0] += ($counter[1] * $counter[0])/(($mu[$i]-1);
124         if($counter[1] > 1) {$coincidenceMatrix[1][1] += $counter[1]*($counter[
125             [1]-1)/(($mu[$i]-1)}
126     }
127     $counter[0] = 0;
128     $counter[1] = 0;
129 };
130 $n[0] = $coincidenceMatrix[0][0] + $coincidenceMatrix[0][1];
131 $n[1] = $coincidenceMatrix[1][0] + $coincidenceMatrix[1][1];
132 $sum = $n[0] + $n[1];
133 print "$sum\n";

```

```

132
133
134     $numerator = ($sum-1)*($coincidenceMatrix[0][0]+$coincidenceMatrix[1][1]) -
135         ($n[0]*($n[0]-1) + $n[1]*($n[1]-1));
136     $denominator = $sum*($sum-1)-($n[0]*($n[0]-1) + $n[1]*($n[1]-1));
137     if($denominator != 0) {$alpha = $numerator/$denominator}
138     else{$alpha = 5000}
139     print KRIALPHABINNED $alpha;
140     print KRIALPHABINNED "\n";

```

## Perl-Skript Varianz der Meinungsdistanzen

```

1  #!/usr/bin/perl
2  # Datei: parse_Table_opinion.pl
3
4  use strict;
5  use warnings;
6  use IO::File;
7
8  my $zeile2;
9  my $initiative;
10 my $suggestion;
11 my $member;
12 my $degree = 0;
13 my $fulfilled;
14 my $previous;
15 my $previous_initiative;
16 my $previous_issue;
17 my $tmp;
18 my $array_length;
19 my $file_extension = ".csv";
20 my $file;
21 my $first = 1;
22 my $sorted = "sorted";
23 my $first_dimension;
24 my $second_dimension;
25 my $i;
26 my $j;
27 my $k;
28 my $polarization = 0;
29 my $max_member = 0;
30 my $max_initiative = 0;
31 my $max_prev_member = 0;
32 my $sum_correlation = 0;
33 my $number_of_edges = 0;
34 my $count_mem = 1;
35 my $max_pol = 0;
36
37 my $sum = 0;
38 my $count = 0;

```

```

39 my $average_distance = 0;
40 my @distance;
41 my @average_distance;
42 my @reliabilityMatrix;
43 my @coincidenceMatrix;
44 my @differenceMatrix;
45 my @mu;
46 my @count;
47 my @n;
48 my $numerator;
49 my $denominator;
50 my $variance = 0;
51 my $first_element;
52 my $countvariable = 1;
53
54
55 open (KRIALPHABINNED, ">testVar${file_extension}") or die "Konnte die OUTPUT-
    Datei nicht öffnen";
56
57
58
59 open(TABLE, "<test${file_extension}") or die $!;
60
61 while (<TABLE>){
62
63
64 #einzelne Zeile der Datei einlesen
65 #my $zeile = TABLE->getline;
66 my $zeile = $_;
67
68 #CSV-Datei splitten mit Trennzeichen ;
69 my ($initiative, $member, $degree) = split(/./,$zeile);
70 if ($first == 1){
71     $previous = $initiative;
72 }
73 if($initiative eq $previous){
74     if($degree > 0){
75         $reliabilityMatrix[$initiative][$member] = 1;
76     }
77     if($degree < 0){
78         $reliabilityMatrix[$initiative][$member] = -1;
79     }
80     if($degree == 0){
81
82     }
83     if($member > $max_member){ $max_member = $member}
84     if($initiative > $max_initiative){ $max_initiative = $initiative}
85     $count++;
86     $previous_initiative = $initiative;
87     $previous = $initiative;
88     $first = 0;
89     if($member>$count_mem) {$count_mem=$member}
90 } else{
91
92     if($degree > 0){
93         $reliabilityMatrix[$initiative][$member] = 1;
94     }
95     if($degree < 0){
96         $reliabilityMatrix[$initiative][$member] = -1;
97     }
98     if($degree == 0){
99
100    }
101    if($member > $max_member){ $max_member = $member}
102    if($initiative > $max_initiative){ $max_initiative = $initiative}
103    $mu[$previous] = $count;
104    $count = 1;
105    $previous = $initiative;
106    $previous_initiative = $initiative;
107    $first = 0;
108 }
109 $mu[$previous] = $count;
110 close(TABLE);
111
112 $k = 1;
113 $count = 0;
114
115 print $max_initiative, ", max member: ", $max_member, "\n";
116
117 for ($i=1; $i <= $max_member; $i++){
118     for ($j=1; $j <= $max_initiative; $j++){
119         while ($k<=$max_member){
120             if($i == ($i+$k)){
121                 $k++;
122                 if($k==$max_member){last;}
123             }
124             if(defined $reliabilityMatrix[$j][$i]){
125                 if(defined $reliabilityMatrix[$j][$i+$k]){
126                     if(not defined $distance[$i][$i+$k]) {
127                         $distance[$i][$i+$k] = 0;
128                         $count[$i][$i+$k] = 0;
129                     }
130                     $distance[$i][$i+$k]=$distance[$i][$i+$k]+abs($reliabilityMatrix[
131                         $j][$i] - $reliabilityMatrix[$j][$i+$k]);
132                     $count[$i][$i+$k] = $count[$i][$i+$k]+1 ;
133                 }
134                 $k++;
135             }
136             $k = 0;
137             if($i==1){
138                 $k = 1;
139             }
140         }
141     }
142     $k = 0;

```

```

143     }
144     $count = 0;
145     for ($i=1; $i<= $max_member; $i++){
146         for ($j=1; $j<= $max_member; $j++){
147             if (defined $distance[$i][$j]) {
148                 #$distance[$i][$j] = 1 - $distance[$i][$j];
149                 $distance[$i][$j] = $distance[$i][$j] * (1/$count[$i][$j]);
150                 print $i, ",," , $j, ",,";
151                 print $distance[$i][$j], ",,count_u", $count[$i][$j];
152                 print "\n";
153                 $count = $count + $count[$i][$j];
154                 $average_distance = $average_distance + $distance[$i][$j];
155             }
156         }
157     }
158     $average_distance = $average_distance * (1/$count);
159     print "avg_dist=u";
160     print $average_distance;
161     print "\n";
162     $count = 0;
163     for ($i=1; $i<= $max_member; $i++){
164         for ($j=1; $j<= $max_member; $j++){
165             if (defined $distance[$i][$j]) {
166                 $polarization = $polarization + ($distance[$i][$j] - $average_distance)
167                     )**2;
168                 $count = $count+$count[$i][$j];
169             }
170         }
171     }
172     $variance = $polarization * (1/$count);
173     print $count;
174     print "\n";
175     print $polarization;
176     print "\n";
177     print KRIALPHABINNED $variance;
178     print KRIALPHABINNED "\n";
179     print "variance:u";
180     print $variance;
181     print "\n";
182     print "max_pol:u";
183     $max_pol = 1 - 1/($count_mem - 1)**2;
184     print $max_pol;
185     print "\n";
186     print "count_mem:u";
187     print $count_mem;
188     print "\n";
189     undef (@distance);
190     undef (@count);
191     undef (@reliabilityMatrix);
192     $max_initiative=0;
193     $max_member=0;
194     $polarization=0;

```

```

195     $average_distance=0;
196     $count=0;
197     $countvariable++;

```

## Matlab-Skript Spectral Clustering & Issue constraints

```

1  fid = fopen('Cluster.txt', 'w');
2  fid2 = fopen('ConstraintsMember.txt', 'w');
3  withinA=0;
4  withinB=0;
5  counter=0;
6
7  for x = 1:1:33
8      s1 = '100erBinWithoutZeros';
9      s2 = int2str(x);
10     s3 = '.csv';
11     s = strcat(s1,s2,s3);
12     data2 = importdata(s);
13     issue=data2(:,1);
14     member=data2(:,2);
15     grade=data2(:,3);
16     A2=sparse(issue, member, grade);
17     A2_absolute=abs(A2);
18     edges=nnz(A2);
19     [m,n] = size(A2);
20     bipartite=[sparse(m,m) A2; A2' sparse(n,n)];
21     full_bipartite=full(bipartite);
22     A2_full=full(A2);
23     A2_absolute_full=full(A2_absolute);
24     A_squared=full_bipartite*full_bipartite;
25     full_bipartite_absolute=abs(full_bipartite);
26     A_squared_absolute=full_bipartite_absolute*full_bipartite_absolute;
27     constraints_member=A_squared(1:m,1:m);
28     constraints_votes=A_squared(m+1:n+m,m+1:n+m);
29     votes=full_bipartite(m+1:m+n,1:m);
30     common_voters=A_squared_absolute(m+1:n+m,m+1:n+m);
31     common_votes=A_squared_absolute(1:m,1:m);
32     [row,col]=find(common_voters>0);
33     constraints_votes(find(common_voters))=constraints_votes(find(
34         common_voters))./common_voters(find(common_voters));
35     constraints_member(find(common_voters))=constraints_member(find(
36         common_voters))./common_voters(find(common_voters));
37
38     pairs = 0;
39     summe = 0;
40     for i = 1:length(common_votes);
41         for j = i+1:length(common_votes)
42             pairs=pairs+1;

```

```

41         summe=summe+constraints_member(i,j);
42     end
43 end
44 constraint_issues(x)=(summe/pairs);
45 pairs = 0;
46 summe = 0;
47 for i = 1:1:length(common_voters);
48     for j = i+1:1:length(common_voters)
49         pairs=pairs+1;
50         summe=summe+constraints_votes(i,j);
51     end
52 end
53 constraint_member(x)=(summe/pairs);
54
55
56 overall_constraint_issues(x)=summe/pairs;
57 for i = 1:1:length(constraints_member)
58     constraints_member(i,i) = 0;
59     constraints_votes(i,i)=0;
60     common_votes(i,i) = 0;
61 end
62 constraints_votes_abs=abs(constraints_votes);
63
64 %unnormalized Laplacian
65 %degrees = (sum(constraints_votes_abs,1))';
66 %D=diag(degrees);
67 %L=D - constraints_votes;
68 %[v,EIGEN]=eig(L);
69 %constraints_votes(find(constraints_votes>0))=1;
70 %constraints_votes(find(constraints_votes<0))=-1;
71
72 constraints_votes_abs=abs(constraints_votes);
73 degrees = (sum(abs(constraints_votes),1))';
74 %D=diag(degrees);
75 %L=D - constraints_votes;
76
77 degrees(degrees==0) = eps; % avoid dividing by zero
78 Dnorm = diag(1./(degrees.^0.5)); % calculate D^(-1/2)
79 %Dnorm = diag(1./(degrees.^0.5), 0, size(D, 1), size(D, 2));
80 identity = eye(size(constraints_votes,1));
81 Lnorm=identity - Dnorm * constraints_votes * Dnorm; % symmetric normalized
82     Laplacian
83 [v,EIGEN]=eig(Lnorm);
84
85 ClusterA=find(v(:,1)>0);
86 ClusterB=find(v(:,1)<0);
87
88 disp(x);
89 disp(length(ClusterA));
90 disp(length(ClusterB));
91 disp('-----');
92 if (length(ClusterA)>length(ClusterB)) sizeBig(x)=length(ClusterA);

```

```

93 else sizeBig(x)=length(ClusterB)
94 end
95 if (length(ClusterB)<length(ClusterA)) sizeSmall(x)=length(ClusterB);
96 else sizeSmall(x)=length(ClusterA);
97 end
98
99 man=pdist(votes,'cityblock');
100 manMatrix=squareform(man);
101
102 withinA=0;
103 withinB=0;
104 between=0;
105 manDistA=0;
106 manDistB=0;
107 manDistBetween=0;
108 counter=0;
109 for i=1:1:length(ClusterA)
110     for j = i+1:1:length(ClusterA)
111         if (common_voters(i,j)>0)
112             manDistA=manDistA+manMatrix(i,j);
113             withinA=withinA+constraints_votes(ClusterA(i),ClusterA(j));
114             if (constraints_votes(ClusterA(i),ClusterA(j))~=0) counter=
115                 counter+1;
116         end
117     end
118 end
119 withinA=withinA/counter;
120 manDistA=manDistA/counter;
121 counter=0;
122 for i=1:1:length(ClusterB)
123     for j = i+1:1:length(ClusterB)
124         if (common_voters(i,j)>0)
125             manDistB=manDistB+manMatrix(i,j);
126             withinB=withinB+constraints_votes(ClusterB(i),ClusterB(j));
127             if (constraints_votes(ClusterB(i),ClusterB(j))~=0) counter=
128                 counter+1;
129         end
130     end
131 end
132 withinB=withinB/counter;
133 manDistB=manDistB/counter;
134 counter=0;
135 for i=1:1:length(ClusterA)
136     for j = 1:1:length(ClusterB)
137         if (common_voters(i,j)>0)
138             manDistBetween=manDistBetween+manMatrix(i,j);
139             between=between+constraints_votes(ClusterA(i),ClusterB(j));
140             if (constraints_votes(ClusterA(i),ClusterB(j))~=0) counter=
141                 counter+1;
142         end
143     end

```



```

143         end
144     end
145     between=between/counter;
146     manDistBetween=manDistBetween/counter;
147
148     betweenVec(x)=between;
149     withinBVec(x)=withinB;
150     withinAVec(x)=withinA;
151     withinAMan(x)=manDistA;
152     withinBMan(x)=manDistB;
153     betweenMan(x)=manDistBetween;
154
155     if (length(ClusterA)>length(ClusterB)) ratio(x)=length(ClusterA)/(length(ClusterA)+length(ClusterB));
156     else ratio(x)=length(ClusterB)/(length(ClusterA)+length(ClusterB));
157     end
158
159     min(x) = EIGEN(1,1);
160
161     ones(x)=length(find(common_voters==1))/length(find(common_voters>0));
162     sum_pos=sum(constraints_member(find(constraints_member>0)));
163     numPos(x)=length(find(constraints_member>0));
164     sum_neg=sum(constraints_member(find(constraints_member<0)));
165
166     numNeg(x)=length(find(constraints_member<0));
167     posConstraints(x)=sum_pos/numPos(x);
168     negConstraints(x)=sum_neg/numNeg(x);
169
170     averageCommonVote(x)=sum(sum(common_voters))/length(find(common_voters>0));
171     ;
172     ratio(x)=numNeg(x)/numPos(x);
173
174     gplot(constraints_votes, [v(:,1) v(:,2)], '-g');
175     hold on;
176     neg=zeros(size(constraints_votes));
177     %constraints_votes(constraints_votes>0)=0;
178     gplot(constraints_votes, [v(:,1) v(:,2)], '-r');
179     hold off;
180
181     histVec=common_voters(find(common_voters>0));
182     hist(histVec,1000);
183     med(x)=median(histVec);
184 end

```

## Matlab-Skript Silhouette Testing

```

1 for x = 1:1:33
2     s1 = '100erBinWithoutZeros';
3     s2 = int2str(x);
4     s3 = '.csv';
5     s = strcat(s1,s2,s3);
6     data2 = importdata(s);
7     issue=data2(:,1);
8     member=data2(:,2);
9     grade=data2(:,3);
10    A2=sparse(issue, member, grade);
11    A2_absolute=abs(A2);
12    edges=nnz(A2);
13    [m,n] = size(A2);
14    bipartite=[sparse(m,m) A2; A2' sparse(n,n)];
15    full_bipartite=full(bipartite);
16    A2_full=full(A2);
17    A2_absolute_full=full(A2_absolute);
18    A_squared=full_bipartite*full_bipartite;
19    full_bipartite_absolute=abs(full_bipartite);
20    A_squared_absolute=full_bipartite_absolute*full_bipartite_absolute;
21    constraints_member=A_squared(1:m,1:m);
22    constraints_votes=A_squared(m+1:n+m,m+1:n+m);
23    votes=full_bipartite(m+1:n+m,1:m);
24    common_voters=A_squared_absolute(m+1:n+m,m+1:n+m);
25    common_votes=A_squared_absolute(1:m,1:m);
26    [row,col]=find(common_voters>0);
27    constraints_votes(find(common_voters))=constraints_votes(find(common_voters))./common_votes(find(common_voters));
28    constraints_member(find(common_votes))=constraints_member(find(common_votes))./common_votes(find(common_votes));
29
30
31    for k=2:1:5
32        C=SpectralClustering(constraints_votes, k, 3);
33        sil=silhouette(constraints_votes,C);
34        silhouette(constraints_votes,C);
35        silu(x,k)=sum(sil)/length(sil);
36        disp(sum(sil)/length(sil));
37    end
38    disp('-----');
39
40
41 end

```

## Matlab-Skript Spectral Clustering für k>2

```

1 function [C, L, U] = SpectralClustering(W, k)
2 %
3 % 'W' - Adjacency matrix, needs to be square
4 % 'k' - Number of clusters to look for
5
6
7 % calculate degree matrix
8 abs_W=abs(W);
9 degs = sum(abs_W, 2);
10 D = sparse(1:size(W, 1), 1:size(W, 2), degs);
11
12 % compute unnormalized Laplacian
13 L = D - W;
14
15 % avoid dividing by zero
16 degs(degs == 0) = eps;
17 % calculate D^(-1/2)
18 D = spdiags(1./(degs.^0.5), 0, size(D, 1), size(D, 2));
19
20 % calculate normalized Laplacian
21 L = D * L * D;

```

## A Anhang

```
22
23 % compute the eigenvectors corresponding to the k smallest
24 % eigenvalues
25 diff = eps;
26 [U, ~] = eigs(L, k, diff);
27
28 % normalize the eigenvectors row-wise
29
30 U = bsxfun(@divide, U, sqrt(sum(U.^2, 2)));
31
32 % now use the k-means algorithm to cluster U row-wise
33 % C will be a n-by-1 matrix containing the cluster number for
34 % each data point
35 C = kmeans(U(:,1), k, 'start', 'cluster', ...
36           'EmptyAction', 'singleton');
37
38 end
```

## Matlab-Skript Xi

```
1 for x = 1:1:33
2     s1 = '100erBinWithoutZeros';
3     s2 = int2str(x);
4     s3 = '.csv';
5     s = strcat(s1,s2,s3);
6     data2 = importdata(s);
7     issue=data2(:,1);
8     member=data2(:,2);
9     grade=data2(:,3);
10    A2=sparse(issue, member, grade);
11    A2_absolute=abs(A2);
12    edges=nnz(A2);
13    [m,n] = size(A2);
14    bipartite=[sparse(m,m) A2; A2' sparse(n,n)];
15    full_bipartite=full(bipartite);
16    A2_full=full(A2);
17    A2_absolute_full=full(A2_absolute);
18    A_squared=full_bipartite*full_bipartite;
19    full_bipartite_absolute=abs(full_bipartite);
20    A_squared_absolute=full_bipartite_absolute*full_bipartite_absolute;
21    constraints_issues=A_squared(1:m,1:m);
22    constraints_member=A_squared(m+1:n+m,m+1:n+m);
23    %votes=full_bipartite(m+1:m+n,1:m);
24    common_votes=A_squared_absolute(m+1:n+m,m+1:n+m);
25    common_voters=A_squared_absolute(1:m,1:m);
26    [row,col]=find(common_voters>0);
27    constraints_member(find(common_votes))=constraints_member(find(common_votes))./common_votes(find(common_votes));
28    constraints_issues(find(common_voters))=constraints_issues(find(common_voters))./common_voters(find(common_voters));
29
30    A=full_bipartite;
31    A(find(A>0))=1;
32    A(find(A<0))=-1
33
34    abs_A=abs(A);
35    degs = sum(abs_A, 2);
36    D = sparse(1:size(A, 1), 1:size(A, 2), degs);
37    D = full(D);
38
39    L = D - A;
40    [v,EIGEN]=eig(L);
41    min(x) = EIGEN(1,1);
42
43    degs(degs==0) = eps; % avoid dividing by zero
44    Dnorm = diag(1./(degs.^0.5)); % calculate D^(-1/2)
45    %Dnorm = diag(1./(degrees.^0.5), 0, size(D, 1), size(D, 2));
46    identity = eye(size(A,1));
47    Lnorm=identity - Dnorm * A * Dnorm; % symmetric normalized Laplacian
48    [v,EIGEN]=eig(L);
```

## A Anhang

```
49     minNorm(x)=EIGEN(1,1);
50
51
52     format longG;
53     nodes=m+n;
54     denominator=8*edges;
55     b_f=nodes/denominator;
56     b_f_x(x)=b_f*EIGEN(1,1);
57     format LongG;
58
59     end
```

# Literaturverzeichnis

- [Abr08] Abramowitz, Saunders Kyle L. Alan I. I. Alan I.: Is Polarization a Myth? In: *The Journal of Politics* 70 (2008), S. 542–555
- [BG08] Baldassarri, Delia; Gelman, Andrew: Partisans without constraint: Political polarization and trends in American public opinion. In: *AJS; American journal of sociology* 114 (2008), Nr. 2, S. 408
- [Bib] Bibliographisches Institut GmbH Dudenverlag: *Polarisierung, die*. <http://www.duden.de/rechtschreibung/Polarisierung>. – aufgerufen am: 2015-04-08
- [Bow12] Bowlin, Garry S.: Maximum frustration in bipartite signed graphs. In: *the electronic journal of combinatorics* 19 (2012), Nr. 4, S. P10
- [Bun] Bundeszentrale für politische Bildung: *Digitale Politik und Partizipation: Möglichkeiten und Grenzen*. <http://www.bpb.de/apuz/75834/digitale-politik-und-partizipation-moeglichkeiten-und-grenzen?p=8>. – aufgerufen am: 2015-04-03
- [Car09] Carroll, Lewis Jefferey B. Lo James Poole Keith T. Rosenthal H. Royce: Measuring Bias and Uncertainty in DW-NOMINATE Ideal Point Estimates via the Parametric Bootstrap. In: *Political Analysis* (2009), S. 261–275
- [Car15] Carsten Sawosch: *Teamleiter Projektgruppe „Relaunch Liquid-Feedback“*. <https://vorstand.piratenpartei.de/2015/05/18/teamleiter-projektgruppe-relaunch-liquidfeedback/>. Version: Mai 2015
- [Coh60] Cohen, Jacob: A Coefficient of Agreement for Nominal Scales. In: *Educational and Psychological Measurement* 20 (1960), S. 37–46

## Literaturverzeichnis

- [Cro51] Cronbach, Lee J.: Coefficient alpha and the internal structure of tests. In: *psychometrika* 16 (1951), Nr. 3, S. 297–334
- [Dar70] Darlington, Richard B.: Is Kurtosis Really Peakedness? In: *The American Statistician* 24 (1970), S. 19–22
- [DiM96] DiMaggio, Evans Paul Bryson B. Paul: Have American's Social Attitudes Become More Polarized? In: *American Journal of Sociology* 102 (1996), S. 690–755
- [Dir]
- [Eas10] Easley, Kleinberg J. David: *Networks, Crowds and Markets*. Cambridge University Press, 2010
- [ES13] Ester, Martin; Sander, Jörg: *Knowledge discovery in databases: Techniken und Anwendungen*. Springer-Verlag, 2013
- [Fio08] Fiorina, Abrams Samuel J. Morris p. p. Morris p.: Political Polarization in the American Public. In: *The Annual Review of Political Science* 11 (2008), Nr. 1], S. 563–588
- [Fio10] Fiorina, Samuel J. Pope Jeremy C. Morris P. Adams A. Morris P. Adams: *Culture War? The Myth of a Polarized America*. Bd. 3. Longman, 2010
- [Fla11] Flache, Macy Micheal W. Andreas: Small Worlds and Cultural Polarization. In: *The Journal of Mathematical Sociology* 35 (2011), S. 146–176
- [Fle71] Fleiss, Joseph L.: Measuring nominal scale agreement among many raters. In: *Psychological Bulletin* 76 (1971), S. 378–382
- [FM11] Flache, Andreas; Macy, Michael W.: Small worlds and cultural polarization. In: *The Journal of Mathematical Sociology* 35 (2011), Nr. 1-3, S. 146–176
- [GMZ03] Gkantsidis, Christos; Mihail, Milena; Zegura, Ellen: Spectral analysis of Internet topologies. In: *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies* Bd. 1 IEEE, 2003, S. 364–374
- [Har53] Harary, Frank: On the notion of balance of a signed graph. In: *Michigan Math. J.* 2 (1953), Nr. 2, 143–146. <http://dx.doi.org/10.1307/mmj/1028989917>. – DOI 10.1307/mmj/1028989917

## Literaturverzeichnis

- [Hil71] Hildebrand, David K.: Kurtosis Measures Bimodality? In: *The American Statistician* 25 (1971), S. 42–43
- [Hou05] Hou, Yao P.: Bounds for the least Laplacian eigenvalue of a signed graph. In: *Acta Mathematica Sinica* 21 (2005), Nr. 4, S. 955–960
- [Jab11] Jabbusch, Sebastian: *Liquid Democracy in der Piratenpartei: Eine neue Chance für innerparteiliche Demokratie im 21. Jahrhundert?*, Universität Greifswald, Diplomarbeit, 2011
- [KKH<sup>+</sup>15] Kling, Christoph; Kunegis, Jérôme; Hartmann, Heinrich; Strohmaier, Markus; Staab, Steffen: Voting Behaviour and Power in Online Democracy: A Study of LiquidFeedback in Germany's Pirate Party. (2015). <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10566>
- [KR09] Kaufman, Leonard; Rousseeuw, Peter J.: *Finding groups in data: an introduction to cluster analysis*. Bd. 344. John Wiley & Sons, 2009
- [Kri70] Krippendorff, Klaus: Bivariate agreement coefficients for reliability data. In: *Sociological methodology* 1970 (1970), S. 139–150
- [Kri11] Krippendorff, Klaus: *Computing Krippendorff's Alpha Reliability*. [http://repository.upenn.edu/asc\\_papers/43?utm\\_source=repository.upenn.edu%2Fasc\\_papers%2F43&utm\\_medium=PDF&utm\\_campaign=PDFCoverPages](http://repository.upenn.edu/asc_papers/43?utm_source=repository.upenn.edu%2Fasc_papers%2F43&utm_medium=PDF&utm_campaign=PDFCoverPages). Version: 2011. – aufgerufen am: 2015-04-08
- [KSL<sup>+</sup>10] Kunegis, Jérôme; Schmidt, Stephan; Lommatzsch, Andreas; Lerner, Jürgen; De Luca, Ernesto W.; Albayrak, Sahin: Spectral analysis of signed graphs for clustering, prediction and visualization. In: *SDM* Bd. 10 SIAM Int. Conf. on Data Mining, 2010, S. 559–559
- [Kun14] Kunegis, Jérôme: Applications of Structural Balance in Signed Social Networks. (2014). <http://arxiv.org/abs/1402.6865>
- [Kun15] Kunegis, Jérôme: Exploiting the Structure of Bipartite Graphs for Algebraic and Spectral Graph Theory Applications. In: *Internet Mathematics* 11 (2015), Nr. 3, S. 201–321
- [Liq] Liquid Democracy e.V.: *Adhocracy*. <https://adhocracy.de/>. – aufgerufen am: 2015-04-03

## Literaturverzeichnis

- [Liqb] Liquid Democracy e.V.: *Liquid Democracy*. <https://liqd.net/>. – aufgerufen am: 2015-04-03
- [Liqc] Liquid Democracy e.V.: *Liquid Democracy - Theoretische Grundlagen*. <https://liqd.net/schwerpunkte/theoretische-grundlagen/liquid-democracy/>. – aufgerufen am: 2015-04-03
- [Liqd] Liquid Democracy e.V.: *Was ist Adhocracy?* <https://adhocracy.de/static/ueber-uns/was-ist-adhocracy-1.html>. – aufgerufen am: 2015-04-03
- [Mü07] Mühlbauer, Peter: *Warum Partei und nicht Religion?* <http://www.heise.de/tp/artikel/26/26173/1.html>. Version: 2007. – aufgerufen am: 2015-04-06
- [Mat13] Matthews, Dylan: *It's official: The 112th Congress was the most polarized ever*. <http://www.washingtonpost.com/blogs/wonkblog/wp/2013/01/17/its-official-the-112th-congress-was-the-most-polarized-ever/>. Version: Januar 2013. – aufgerufen am: 2015-04-01
- [MB13] Munzert, Simon; Bauer, Paul C.: Political depolarization in German public opinion, 1980–2010. In: *Political Science Research and Methods* 1 (2013), Nr. 01, S. 67–89
- [Mit] Mitteldeutscher Rundfunk: *Wie können wir eine Spaltung der Gesellschaft verhindern?* [http://www.mdr.de/sachsenspiegel/polarisierung-fluechtlingsdebatte100\\_zc-1b6dad2f\\_zs-c5332a56.html](http://www.mdr.de/sachsenspiegel/polarisierung-fluechtlingsdebatte100_zc-1b6dad2f_zs-c5332a56.html). – aufgerufen am: 2015-11-18
- [Nat] National Opinion Research Center: *General Social Survey*. [http://www3.norc.org/GSS+Website/About+GSS/](http://www3.norc.illinois.edu/GSS+Website/About+GSS/). – aufgerufen am: 2015-04-08
- [NGKA11] Naveed, Nasir; Gottron, Thomas; Kunegis, Jérôme; Alhadi, Arifah C.: Bad news travel fast: A content-based analysis of interestingness on twitter. In: *Proceedings of the 3rd International Web Science Conference ACM*, 2011, S. 8
- [NJW<sup>+</sup>02] Ng, Andrew Y.; Jordan, Michael I.; Weiss, Yair; others: On spectral clustering: Analysis and an algorithm. In: *Advances in neural information processing systems* 2 (2002), S. 849–856
- [Sch12] Schultz, Stefan: *Drohender Finanzkollaps der USA: Der Zeitlupen-Crash*. <http://www.spiegel.de/wirtschaft/soziales/>



## Literaturverzeichnis

was-der-sprung-von-der-fiskalklippe-bedeutet-a-874791.html.

Version: Dezember 2012. – aufgerufen am: 2015-04-01

- [Sco55] Scott, William A.: Reliability of content analysis: The case of nominal scale coding. In: *Public Opinion Quarterly* 19 (1955), S. 321–325
- [SS54] Shapley, Lloyd S.; Shubik, Martin: A method for evaluating the distribution of power in a committee system. In: *American Political Science Review* 48 (1954), Nr. 03, S. 787–792
- [Sta] Stanford University, University of Michigan, National Science Foundation: *American National Election Studies*. <http://www.electionstudies.org/>. – aufgerufen am: 2015-04-08
- [VL07] Von Luxburg, Ulrike: A tutorial on spectral clustering. In: *Statistics and computing* 17 (2007), Nr. 4, S. 395–416
- [WC89] Wei, Yen-Chuen; Cheng, Chung-Kuan: Towards efficient hierarchical designs by ratio cut partitioning. In: *Computer-Aided Design, 1989. ICCAD-89. Digest of Technical Papers., 1989 IEEE International Conference on IEEE, 1989*, S. 298–301
- [Zas13] Zaslavsky, Thomas: Matrices in the theory of signed simple graphs. In: *arXiv preprint arXiv:1303.3083* (2013)
- [Zei12] Zeitung, Süddeutsche: *Amerikaner wählen Blockade im Kongress*. <http://www.sueddeutsche.de/politik/senat-und-repraesentantenhaus-amerikaner-waehlen-blockade-im-kongress-1.1516345>. Version: November 2012. – aufgerufen am: 2015-04-01