

Bundestagswahl 2017: Einblicke in das Tweetverhalten von Politikern mit Netzwerkanalyse und Topic Modelling

Philipp Hartl

Lehrstuhl für Medieninformatik
Universität Regensburg, Germany
Philipp1.Hartl@stud.uni-regensburg.de

Dominik Ramsauer

Lehrstuhl für Informationwissenschaft
Universität Regensburg, Germany
Dominik.Ramsauer@stud.uni-regensburg.de

ABSTRACT

Die sozialen Medien prägen die moderne Medienlandschaft - so hat auch die Politik erkannt, dass sich die Bürger mit diesen Massenmedien in Echtzeit über Meinungen und Neuigkeiten informieren lassen. In dieser Studie untersuchen wir die politische Diskussion und die innerparteilichen Strukturen der Politiker bei der Bundestagswahl 2017 in Deutschland. Als Basis hierfür dient uns ein Twitter-Datensatz, der etwa 400.000 Tweets von über 1000 Bundespolitikern umfasst. Wir möchten mit einem computergestützten Ansatz hierarchische Strukturen, das sog. Kleine-Welt-Phänomen sowie die Robustheit der Parteinetzwerke von außen untersuchen. Darüber hinaus wollen wir aufzeigen, welche Themen die politische Diskussion bewegt haben und ob diese Themen von den Parteien unterschiedlich stark debattiert worden sind. Wir konnten wichtige Diskussionspunkte der Wahl identifizieren und den Parteien Themen zuordnen, die sie überwiegend selbst besetzen.

KEYWORDS

Politik, Twitter, Text Mining, Topic Modelling

1 EINLEITUNG

Social-Media-Plattformen wie Twitter haben sich über die Jahre zu öffentlichen Massen-Kommunikationsmedien entwickelt, welches von einem breiten Teil der Gesellschaft genutzt wird [28]. Durch die Möglichkeit über Twitter ganze Bevölkerungsteile aktuell mit Informationen und Statements zu erreichen, hat auch die Politik einen Nutzen in den sozialen Medien gefunden. So haben politische Bewegungen wie die 15M Bewegung in Spanien, Occupy Wall Street, #YoSoy132 in Mexico, Occupy Gezi in der Türkei und weitere, ähnliche Plattformen genutzt um sich zu organisieren [3]. Neben solch teils eher losen politischen Bewegungen werden soziale Netzwerke auch vermehrt von der etablierten Politik im Wahlkampf eingesetzt [2, 19, 20]. In Deutschland kam es zum ersten mal vermehrt zur strategischen Benutzung der Plattform Twitter im Zuge der Bundestagswahl 2013 [19]. In den letzten vier Jahren hat sich die politische Landschaft in Deutschland außerordentlich stark verändert. Hauptgrund hierfür ist die Etablierung einer neuer politischen Kraft, der AfD (Alternative für Deutschland). Die AfD ist 2013 unter den Nachwirkungen der Wirtschaftskrise im Euro-raum ab 2007 als zuerst Europa-kritische Partei entstanden. Im Zuge der Flüchtlingskrise 2015 in Deutschland und ganz Europa hat sich das politische Profil der AfD nach rechts, hin zum Rechtspopulismus verschoben und damit eine neue politische Position rechts neben den etablierten konservativen Parteien, der Union aus CDU (Christlich Demokratische Union) und CSU (Christlich-Soziale Union) geschaffen [14]. Nun ist die Partei, die besonders

durch ihre aggressive, populistische Kommunikation auffiel [24], mit 12,6% bei der Wahl des deutschen Bundestages 2017 erstmalig in das höchste politische Parlament Deutschlands eingezogen. Außerdem ist die FDP (Freie Demokratische Partei) nach ihrem Ausscheiden im Jahr 2013 mit 10,7% nun erneut in den Bundestag gewählt worden. Bemerkenswert an den Ergebnissen der Wahl ist, dass es für die großen, etablierten Parteien CDU/CSU und die SPD (Sozialdemokratische Partei Deutschlands) die schlechtesten Ergebnisse seit der Gründung der Bundesrepublik Deutschland sind, wohingegen die kleinen Parteien deutlich an Zuspruch gewinnen konnten [9]. Die inter- sowie innerparteiliche Kommunikation ist ein wichtiger Bestandteil des politischen Wahlkampfes. Auf Twitter sind mehrere Arten der Kommunikation möglich. Einerseits das sogenannte "retweeten" eines Beitrags eines anderen Nutzer, sowie das Antworten (Reply/Mention) auf einen Post eines anderen Nutzers. Entstehen also, abhängig von der Parteizugehörigkeit & der Art der Kommunikation, unterschiedliche soziale Strukturen innerhalb der Parteien und welche Verbindungen gibt es zwischen den Parteien?

In jedem Wahlkampf gibt es vorherrschende Themen, welche die Debatten dominieren. Es stellt sich die Frage, ob es Unterschiede bei den vorherrschenden Themen abhängig von der Parteizugehörigkeit gibt und welche Themen unabhängig von der Parteizugehörigkeit den Wahlkampf dominiert haben. Tritt sogenannte *cyber balkanization* [3] auf? Als Grundlage um diese Fragen zu beantworten dient ein Datensatz der über die Twitter REST API bereit gestellt worden ist. Hierbei wurde allen relevanten Politikern auf Twitter gefolgt und deren Tweetverhalten festgehalten.

2 RELATED WORK

2.1 Twitter Netzwerke

Durch die große Anzahl von Nutzern, der Menge an Tweets, die jeden Tag verfasst werden und der leichte Zugang zu den Tweets macht Twitter sehr attraktiv für die Forschung und dient daher oft als Datenquelle in verschiedenen Forschungsgebieten (Java et al. [16]). Neben Forschungsfragen zum Informationsverhalten der Nutzer beschäftigen sich einige Studien mit politischen Strukturen. So untersuchte zum Beispiel Aragón et al. [3] mit einem Netzwerk aus Retweets hierarchische Strukturen von Parteien, das Kleine-Welt-Phänomen und die Dezentralisierung der Parteien, die 2015 in Barcelona zur Stadtratswahl angetreten sind. Besonderes Augenmerk lag hierbei auf der Bewegung "Barcelona en Comú", die sich zu einer Partei zusammenschloss um die Wahl zu gewinnen. Lietz et al. [19] untersuchte, wie Parteien Twitter bei der Bundestagswahl in Deutschland 2013 genutzt haben und wie Online-Konversationspraktiken quantifiziert werden können. Conover et al.

[10] ging der Frage nach, wie die sozialen Medien die Öffentlichkeit beeinflussen und die Kommunikation zwischen gesellschaftlichen Gruppierungen erleichtert. Aus einem Datensatz von 250.000 politisch relevanten Tweets von über 45.000 Usern baute er Retweet- und Mention-Netzwerke und trennte die Beiträge in Gruppen mit Cluster Analysis.

2.2 Topic Modelling

Das Thema von Dokumenten zu bestimmen ist im Bereich des Text Mining ein wichtiges Teilgebiet. Solche Verfahren der automatischen inhaltlichen Erschließung werden als Topic Models bezeichnet. Stier et al. [24] sammelte eineinhalb Jahre lang Facebook-Posts deutscher Parteien, deren Parteimitglieder auf Bundesebene und Posts von Konten, die in Verbindung mit der Pegida-Bewegung stehen. Um Themen aus dem Datensatz zu identifizieren verwendete er die *Latent Dirichlet Allocation* (LDA). Entwickelt von Blei et al. [5] ist LDA ein unüberwachtes Lernverfahren zur latenten semantischen Analyse. Seine Ergebnisse waren, dass Pegida und die AfD eigene Themen hervorheben und damit ähnliche Zielgruppen ansprechen. Des Weiteren untersuchte er, inwieweit die angesprochenen Themen als populistisch einzuordnen sind mit dem Resultat, dass die Posts von AfD und Pegida im Vergleich zu den anderen Parteien wesentlich populistischer sind.

Als in den USA der dunkelhäutige Teenager Michael Brown von einem hellhäutigen Polizisten niedergeschossen wurde, brannte in den sozialen Medien eine Debatte los. Unter dem Hashtag #BlackLivesMatter beschwerten sich Leute über die Ungleichbehandlung von Weißen und Schwarzen seitens der Polizei. Gallagher et al. [11] untersuchte zu diesem Protest und zum Gegenprotest #AllLivesMatter 860.000 Tweets auf mehreren Ebenen. Anstelle der Kullback-Leibler Divergenz (Statistik zur Beurteilung von Verteilungsunterschieden zwischen zweier Texten) nutzte er die Jensen-Shannon Divergenz [17], da die Divergenz ansonsten unendlich groß werden würde, sollte ein Wort in einem Text aber nicht im anderen vorkommen. Er stellte fest, dass die beiden Protestgruppen stark unterschieden.

3 METHODIK

3.1 Daten

Mit Hilfe der Twitter REST API wurden etwa 400.000 Tweets vom 1.1.2017 (37 Wochen vor der Wahl) bis zum 19.10.2017 (3 Wochen nach der Wahl) von über 1000 Politikern gesammelt. Davon sind 76 AfD, 223 CDU, 45 CSU, 56 FDP, 216 "Bündnis90/Die Grünen", 139 "Die Linke" sowie 313 SPD Politiker. Im Datensatz werden zwischen Retweets, Quotes, Replies sowie @Mentions unterschieden. Bis zur Wahl am 24.09.17 wurden davon etwa 350.000 Tweets abgesetzt.

3.2 Parteistrukturen

Basierend auf Wasserman and Faust [26] charakterisieren wir sowohl RT- also auch Reply Netzwerke nach fünf Punkten:

- (1) **Anzahl der Knoten**
Hierbei wird die absolute Zahl an Knoten im jeweiligen Netzwerk angegeben.
- (2) **Kantendichte**
Die Kantendichte

$$\Delta = \frac{L}{g * (g - 1)}$$

ist definiert als die Anzahl der beobachteten Kanten L durch die Anzahl an möglichen Kanten, welche als Paar zwischen zwei Knoten g definiert sind. Es gibt also $g(g-1)$ maximale Möglichkeiten.

- (3) **Anzahl der Komponenten**

Knoten innerhalb eines Graphen können eigene, verbundene Subgraphen bilden, welche keinerlei Verbindungen in andere Subgraphen aufweisen. Diese Subgraphen werden *Komponenten* genannt

- (4) **Durchmesser**

Der Durchmesser eines Graphen ist definiert als längster aller kürzesten Pfade zwischen zwei Knoten A und B .

- (5) **Gewichteter Clusteringkoeffizient**

Als lokales Transitivitätsmaß verwenden wir den gewichteten Clusteringkoeffizient nach Barrat et al. [4]

$$c_i^w = \frac{1}{s_i(k_i - 1)} \sum_{i,j} \frac{(w_{ij} + w_{ih})}{2} a_{ij} a_{ih} a_{jh}$$

eines Knotens i mit dessen k_i Nachbarn, dessen Stärke s_i , den Gewichten w sowie Elementen aus der korrespondierenden Adjazenzmatrix a . Als globalen Clusteringkoeffizienten

$$C_i^w = \frac{1}{N} \sum_i c_i^w$$

betrachten wir das Mittel aller lokalen Clusterkoeffizienten [27].

Inspiriert von Aragón et al. [3] verwenden wir weitere Maße um die Topologie, speziell auf innerparteilicher Ebene, zu evaluieren.

Hierarchische Struktur: Um die hierarchische Struktur innerhalb einer Partei zu quantifizieren benutzen wir den Gini Koeffizienten [12]

$$Gin = \frac{A}{A + B}$$

ein empirisches Maß zur Ungleichheit in einer Verteilung dargestellt als Lorentz-Kurve, wobei A die Fläche zwischen der idealen Verteilung und der Lorentz-Kurve darstellt und B die Fläche unterhalb der Lorentz-Kurve. Im Kontext von Netzwerken wird die Ungleichheit des In-Degrees als Maß für die hierarchische Struktur innerhalb eines Netzwerkes verwendet [3]

Kleine-Welt-Phänomen: Das *Kleine-Welt-Phänomen* stellt die Hypothese auf, dass in Netzwerken die meisten Knoten mit sehr wenigen Schritten erreichbar sind [25]. Wie Aragón et al. [3] verwenden wir die durchschnittliche Pfadlänge $l = \sum_{i \neq j} \frac{d_{ij}}{n(n-1)}$ und den Clusteringkoeffizienten um dieses Phänomen zu untersuchen. Allerdings verwenden wir den Clusteringkoeffizienten nach [4] aus bereits genannten Gründen.

Coreness: Die Coreness ist ein Widerstandmaß eines Netzwerkes von äußeren Einflüssen. Um diesen Widerstand zu messen berechnen wir die K-Cores der einzelnen Netzwerke, wobei wir jeweils den maximalen K-Core und den durchschnittlichen K-Core

angeben. Ein Node gehört zu einem K-Core, wenn er mit mindestens k anderen Knoten innerhalb des Kerns verbunden ist [3].

3.3 Wahlkampfthemen

Twitter bieten seinen Nutzern die Möglichkeit an Hashtags zu benutzen um ihre Tweets zu kategorisieren [13, 18]. Einschlägige Fachliteratur hat Hashtags bereits als Proxies für Themen verwendet [7, 8, 10, 13]. Um wichtige Wahlkampfthemen zu identifizieren ist es notwendig irrelevante Hashtags möglichst zu vermeiden. Wir haben im Zuge unserer Vorbereitungen mit einer Stoppwortliste gearbeitet um die möglichen Hashtags einzudämmen. Hierbei wurden primär deutsche, aber auch einige englischen Stoppwörter verwendet. Ebenfalls haben wir die parteieigenen Hashtags (#afd, #cdu, #csu, #fdp, #gruene, #gruenen, #linke, #dielinke, #spd) entfernt, da diese sehr häufig Vorkommen, allerdings keinerlei Inhalt, sondern nur Parteizugehörigkeit, transferieren. Ebenfalls entfernt haben wir die Hashtags #btw2017 und #btw17 da diese schlicht zu häufig auch außerhalb des Kontextes der Wahl benutzt worden sind. Um relevante Themencluster zu finden haben wir aus den Tweettexten alle verbleibenden Hashtags extrahiert. Des Weiteren haben wir uns nur auf Ko-Okkurrenzen zwischen mindestens zwei Hashtags beschränkt und somit isolierte Hashtags nicht betrachtet. Schlussendlich haben wir ein Netzwerk mit etwa 44.000 Knoten sowie 100.000 Kanten erhalten. Wegen der Größe des Netzwerkes wird es notwendig wichtige Themengebiete zu erkennen, also zusammengehörige Hashtags zu einem Thema zu gruppieren. Um diese Hashtagcluster (bzw. Subnetzwerke) zu erkennen benutzen wir *Community detection*. Damit wir diese durchführen können, müssen wir zuerst einen Clusteringalgorithmus anwenden. Der Louvain-Algorithmus (Blondel et al. [6]) ist einer dieser, er besticht durch hohe Performance, sowie guten Ergebnissen. Ersteres ist vor allem bei großen Netzwerken, wie dem unseren, entscheidend. Er wurde bereits in ähnlichen Studien [3] erfolgreich im Twitterkontext verwendet. Ziel des Berechnungsverfahrens ist es, die *Modularity* [22]

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta(c_i c_j)$$

zu optimieren. Wobei A_{ij} das Kantengewicht der Knoten i und j , k_i und k_j die Summe der Gewichte der Kanten (Gradzentralität -Degree) an i und j , m die komplette Anzahl an Kanten im Graphen, c_i und c_j die Community der Knoten sowie δ das Kronecker Delta darstellt.

Der Algorithmus teilt zuerst jedem einzelnen Knoten sein eigenes Cluster (bzw. eigene Community) zu. Anschließend wird für jeden Knoten i gemessen, wie sich die Modularity ändern, wenn er in die Community des Nachbarn j eingeteilt wird.

$$\Delta Q = [\frac{S_{in} + k_{i,in}}{2m} - (\frac{S_{tot} + k_i}{2m})^2] - [\frac{S_{in}}{2m} - (\frac{S_{tot}}{2m})^2 - (\frac{k_i}{2m})^2]$$

Wobei S_{in} die Summe der Gewichte innerhalb der neuen Community in die der Knoten i geschoben wird, S_{tot} die Summe aller Kantengewichte zu Knoten in der Community, k_i ist die Gradzentralität des Knotens (Degree), $k_{i,in}$ ist die Summe aller Kantengewichte zwischen i und anderen Knoten innerhalb der Community, m ist die

Summe aller Gewichte im kompletten Netzwerk. Dies wird für alle Communities wiederholt und i nun der Community zugewiesen, die die höchsten Anstieg an Modularity aufweist. Dieses Verfahren wird iterativ wiederholt, bis keinerlei Zunahme an Modularity mehr möglich ist. Als zweiten Schritt erzeugt der Algorithmus nun ein neues Netzwerk aus den Communities, das heißt das nun jeder Knoten eine eigene Community darstellt. Nun wird wieder iterativ die Modularity erhöht, bis schließlich ein Maximum erreicht worden ist. Das sind die abschließenden Clustereinteilungen des Algorithmus.

4 ERGEBNISSE

Im Folgenden wollen wir die Ergebnisse unseres rechnerischen Ansatzes zur Erkennung von Parteistrukturen und zur Identifizierung von Wahlkampfthemen vorstellen und einordnen. Es fiel vorab auf, dass Politiker vor allem auf Parteikollegen reagierten (siehe Figure 1). Abgesehen von den Politikern der Unionsparteien CDU/CSU, die als Schwesterparteien auch vermehrt untereinander retweeteten, retweeteten die Parteien zu über 90% sich selbst - bei der AfD sind es sogar 99,1%. Bei den Unionsparteien belaufen sich die Zahlen addiert ebenfalls auf jeweils über 90% (CDU: 95,9%, CSU: 93,4%). Bemerkenswert ist darüber hinaus auch, dass keine der anderen Parteien die AfD geretweetet hat. Die Werte hierfür belaufen sich auf exakt 0 Retweets. Bei den Reply-Werten sind die Werte nicht so extrem, aber auch hier zeigt sich ein Trend dafür ab, dass die Politiker auf andere Politiker der eigenen Partei antworten (AfD: 70,8%, Die Linke: 60,5%, Grüne: 68,6% und SPD: 49,5%).

4.1 Parteistrukturen

Mithilfe des graphopt-Algorithmus[23] haben wir zwei Netzwerke gebildet: ein Retweet-Netzwerk und ein Reply-Netzwerk (siehe Figure 2 2). Die Graphknoten stellen die Accounts der Politiker dar. Es wurden einige Ausreißer aus dem Retweet-Netzwerk entfernt, da einige Politiker ausschließlich sich selbst geretweetet haben.

5-Punkte-Zusammenfassung (siehe Table 1 und 2) Zu den kleineren Parteien zählen die FDP und CSU bei den Retweets (46 und 29), als auch bei den Replies (31 und 24). Besonders dicht sind bei den Retweets vor allem die AfD (0,09462) und Die Linke(0,08015), was darauf schließen lässt, dass diese Parteien eine höhere Kohäsion besitzen. Im Reply-Netzwerk sind die Dichten wesentlich geringer: überwiegend eine Dezimale niedriger. Bei der Berechnung der Komponenten im Retweet-Netzwerk fällt auf, dass parteiintern keine Subgruppen vorhanden sind. Im Reply-Netzwerk sind die Werte höher: sie belaufen sich zwischen 4 (AfD, SPD) und 10 (CDU, CSU). Beim Durchmesser variieren die Zahlen nicht sehr stark. Beim RT-Netzwerk liegen sie zwischen 5 (CSU) und 8 (CDU und SPD), im Reply-Netzwerk zwischen 7 (Die Linke) und 11 (AfD) mit einem deutlichen Ausreißer: CSU mit einem Durchmesser von 3. Da die CSU mit 24 Knoten die kleinste Partei ist könnte eine mögliche Ursache hierfür sein.

Hierarchische Struktur (siehe Figure 3). Aus der Lorenzverteilung der In-Degree-Werte im RT-Netzwerk lässt sich erkennen, dass sich zwei Kurventypen herausgebildet haben. Die Gini-Koeffizienten der etablierten Parteien (CDU/CSU, SPD) und die der FDP sind annähernd gleich: höchster Wert: FDP (0.707), niedrigster Wert: CDU (0.6804). Die übrigen Parteien AfD, Die Linke und die Grünen

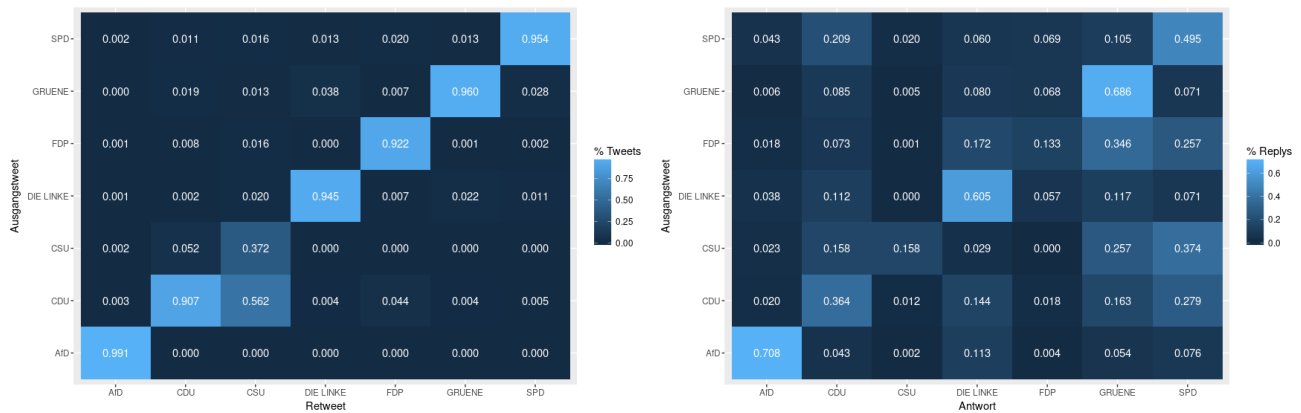


Figure 1

Matrizen mit kumulativen Häufigkeiten, wie oft Politiker einer Partei auf Politiker anderer Parteien reagiert haben. Links: Reaktionen über Retweets. Rechts Reaktionen über Replies.

	Knoten	Dichte	Komponenten	Durchmesser	Clusteringkoeffizient	durch. Pfadlänge	max. K-Core	durch. K-Cores
AfD	70	0.09462	1	7	0.38171	2.51419	11	7.55714
CDU	175	0.03192	1	8	0.31003	3.04798	14	6.10286
CSU	29	0.05788	1	5	0.18834	2.11034	3	1.96552
SPD	252	0.02697	1	8	0.26403	3.08092	17	7.51190
FDP	46	0.06957	1	7	0.29039	2.67820	6	3.76087
GRUENE	200	0.05495	1	6	0.33033	2.61239	21	11.95
DIE LINKE	128	0.08015	1	7	0.35310	2.56801	18	11.78906
Alle Parteien	1068	0.00904	149	11	0.26343	4.00642	21	9.09239

Table 1: Kennzahlen der Retweetnetzwerke

	Knoten	Dichte	Komponenten	Durchmesser	Clusteringkoeffizient	durch. Pfadlänge	max. K-Core	durch. K-Cores
AfD	54	0.01622	4	11	0.13091	3.65191	6	2.41045
CDU	144	0.00672	10	10	0.10105	3.56641	7	2.70541
CSU	24	0.01401	10	3	0.00550	1.71717	2	1.28571
SPD	200	0.00686	4	9	0.11289	3.80851	10	3.21395
FDP	31	0.00965	5	9	0.07684	3.85301	4	1.75152
GRUENE	180	0.00894	8	9	0.16544	3.38812	12	3.85464
DIE LINKE	108	0.01188	5	7	0.17845	3.23074	11	3.54212
Alle Parteien	803	0.00823	19	10	0.15837	3.68733	18	6.99873

Table 2: Kennzahlen der Replynetzwerke

sind mit mind. 10 Prozentpunkte deutlich darunter. Daraus lässt sich schließen, dass die etablierten Parteien und die FDP wesentlich hierarchischer strukturiert sind als die anderen.

Kleine-Welt-Phänomen Wir beobachten in Table 1, dass die AfD den höchsten Clusteringkoeffizienten besitzt ($Cl = 0.382$), gefolgt von den Grünen und der Linken. Die CSU als kleinste Partei ist hingegen mit einem Wert von $Cl = 0.189$ die Partei mit dem niedrigsten Wert. Das lässt sich dadurch erklären, dass die CSU nur wenige Politiker auf Bundesebene hat, die viel Tweets absetzen. Bei der durchschnittlichen Pfadlänge besitzt die CSU ebenfalls den niedrigsten Wert ($l = 2.11$). Das bestätigt sich auch in Table 2: die CSU ist

mit Abstand am niedrigsten. Während die anderen Parteien durchschnittliche Pfadlängen von über 3 haben, so hat die CSU einen Wert von $l = 1.72$.

Coreness In Table 1 und 2 sehen Sie die Zahlen für den maximalen und den durchschnittlichen K-Core. Wir beobachten wie auch beim Kleine-Welt-Phänomen, dass die CSU die mit Abstand niedrigsten Werte aufweist. Die maximalen K-Cores liegen bei 3 (RT-Netzwerk) und 2 (Reply-Netzwerk). Auch die durchschnittlichen K-Core-Werte sind weitaus geringer als bei den anderen Parteien. Die Grünen und Die Linke haben hierbei die höchsten Werte. Daraus lässt sich

	Communityname	Tophashtags
1	berlin	r2g, agh, mahe, tegel, ber
2	merkel	schulz, csu, tvduell, wahlkampf, fedidwgugl
3	g20	hamburg, polizei, nsu, nog20, demaiziere
4	nrw	ltwnrw, dusseldorf, koln, ltnrw, ltwnrw17
5	groko	klimaschutz, energiewende, bundesregierung, dobrindt, zukunft

Table 3: Fünf größte Hashtagcluster aller Parteien (Communityname ist der häufigste Hashtag/ falls mehr als fünf Einträge, waren Hashtags gleich häufig)

schließen, dass die CSU sich offenbar hauptsächlich auf ihre zentralen Leute stützt.

4.2 Wahlkampfthemen

Allgemein dominierende Themen

In Table 3 sieht man die dominierenden Themenblöcke, aller Parteien, samt deren häufigsten Hashtags. In dem häufigsten Cluster *Berlin* geht es primär um die Stadt Berlin selbst, deren Regierung sowie wichtige Orte innerhalb Berlins. Der Hashtag *r2g* ist kurz für "rot-rot-grün". Dies ist eine mögliche Regierungskoalition in Deutschland aus der SPD, den Linken sowie den Grünen. Im Berliner Kontext ist seit Ende 2016 eben eine solche Regierung an der Macht [21]. Der Hashtag *agh* verweist auf das Abgeordnetenhaus in Berlin, *mahe* steht im Berlinerkontext wahrscheinlich für den Bezirk Marzahn-Hellersdorf. Die Kürzel *tegel* und *ber* verweisen auf die beiden Flughäfen Berlins, wobei letzterer unter massiver Kritik steht. Unter dem Hashtag *merkel* geht es sowohl um die Verbindung der amtierenden Bundeskanzlerin Angela Merkel zu ihrem Herausforderer Martin Schulz, diese Konfrontation hatte ihren Höhepunkt im *tvduell* zwischen beiden. Die letzten beiden Hashtags sind wohl im Kontext des CDU-Wahlkampfes zusammen mit *merkel* verwendet worden (*fedidwgugl* ist der Wahlspruch der CDU: "Für ein Deutschland in dem wir gut und gerne Leben"). Ein weiterer großer Punkt in der deutschen Politik 2017 spiegelt sich auch in den Themen wieder. Der *g20*-Gipfel in *Hamburg* ist von heftigen Protesten überschattet worden. Im Zuge dessen hat sich die *Polizei* mit Demonstranten Kämpfe geliefert. Der *nsu* Hashtag ist ebenfalls teil dieses Clusters, was auf einen Vergleich zu den NSU-Morden hindeutet, wie auch immer geartet. Schlussendlich sind noch *#nog20*, welcher den Protest gegen den Gipfel darstellt, sowie der deutsche Innenminister Thomas *demaiziere* Teil dieses Clusters. Letzterer ist für die innere Sicherheit in Deutschland zuständig. Das vierte Topthema ist die Landtagswahl in Nordrhein Westfalen (*nrw*). Mit den passenden Hashtags zur Landtagswahl *ltwnrw*, *ltnrw* und *ltwnrw17* wird dies sehr stark deutlich. Diese Landtagswahl war ein wichtiger "Stimmungstest" in der deutschen Politik vor der Bundestagswahl und ein Indikator für mögliche Ergebnisse, bei der die regierende Koalition aus SPD und Grünen klar abgewählt worden ist. Als letztes Thema hat der Louvain-Clustering Algorithmus die sogenannte *Groko* (große Koalition - Koalition aus CDU/CSU und SPD) bestimmt. Sie ist die aktuelle *bundesregierung*, allerdings scheinen hier Hashtags zu dominieren die wenig inhaltlich zusammenhängend sind. Interessant hierbei ist, das sonst Hashtags dominieren die man

eher bei den Grünen 6 vorfindet. Themen wie Klimaschutz oder auch Energiewende, in diesem Kontext kann der Namen Alexander Dobrindt (wie in 6) mit dem *Dieselgate* in Verbindung gebracht werden. Wobei *Dieselgate* für die absichtliche Manipulation der Abgaswerte von Dieselmotoren der VW-Gruppe steht.

Parteispezifische Unterschiede

Neben den parteiunabhängigen Themen treten deutliche Fixierungen der Parteien auf einzelne Themen oder eine andere Interpretation von den bereits genannten Themengebieten. So ist zum Beispiel *Bayern* das dominierende Thema der CSU (Table 4) Bayern.

	Communityname	Tophashtags
1	bayern	klartext, seehofer, sicherheit, bundeswehr, heimat
2	g20	demaiziere, bmi, hamburg, polizei, wochenende
3	merkel	schulz, tvduell, fedwgugl, fakenews, klarfuerunserland
4	eu	tuerkei, europa, erdogan, brexit, bayernlb
5	unibayreuth	ubt, bayreuth, bgi, che, rassismus

Table 4: Top fünf Hashtagcluster der CSU

Die CSU ist die einzige Partei, welche nur in Bayern wählbar ist aber auch gleichzeitig Teil des Bundestages ist. Dementsprechend konzentriert sich die CSU auf ihr Bundesland Bayern und ihren Ministerpräsidenten Horst *Seehofer*. Des weiteren hat die CSU im Wahlkampf auf Themen wie die innere *Sicherheit* sowie das *Heimat*-Verständnis gesetzt, Begriffe die sich im Cluster der CSU wiederfinden lassen. Im SPD eigenen Hashtag *zeitfuermartin* aus Table 5 wird der Wahlkampf der Sozialdemokraten mit deren Spitzenkandidat Martin *Schulz* deutlich. Im Gegensatz zur Spitzenkandidatin der Union, Angela Merkel, ist Schulz allerdings nur innerhalb der SPD ein Thema. Dies legt den Verdacht nahe, dass innerhalb der SPD, was dieses Thema betrifft, sog. *echo chambers* auftreten [15, S. 76]. Also dass Personen primär nur ihre eigene Meinung innerhalb eines Netzwerkes bestätigt wird - ein "zurück schallen" der eigenen Aussage, durch andere ähnlich gesinnte Personen. Ein weiteres Beispiel für parteispezifische Themen ist der *Klimaschutz* im Cluster der Grünen (Table 6). Die wichtigste Parteipolitik der Grünen ist, seit ihrer Gründung, der Klimaschutz. Es ist also wenig überraschend, dass der Klimaschutz das dominierende Thema innerhalb der Grünen zu sein scheint. Mit Begriffen wie der *Energiewende*, dem *Kohleausstieg*, dem *Klima* oder auch dem

	Communityname	Tophashtags
1	berlin	pflege, r2g, agh, neukolln, schoeneberg
2	zeitfuermartin	schulz, schulzzug, jetztistschulz, martinschulz, zeitfuermehrgerechtigkeit
3	nrw	nrrw, ltnrw, duisburg, machen_wir, dgb
4	rlp	dreyer, mainz, inklusion, landtag, sicherheit, tde2017
5	hamburg	g20, hhhue, polizei, pulseufeurope, g20ham17

Table 5: Top fünf Hashtagcluster der SPD

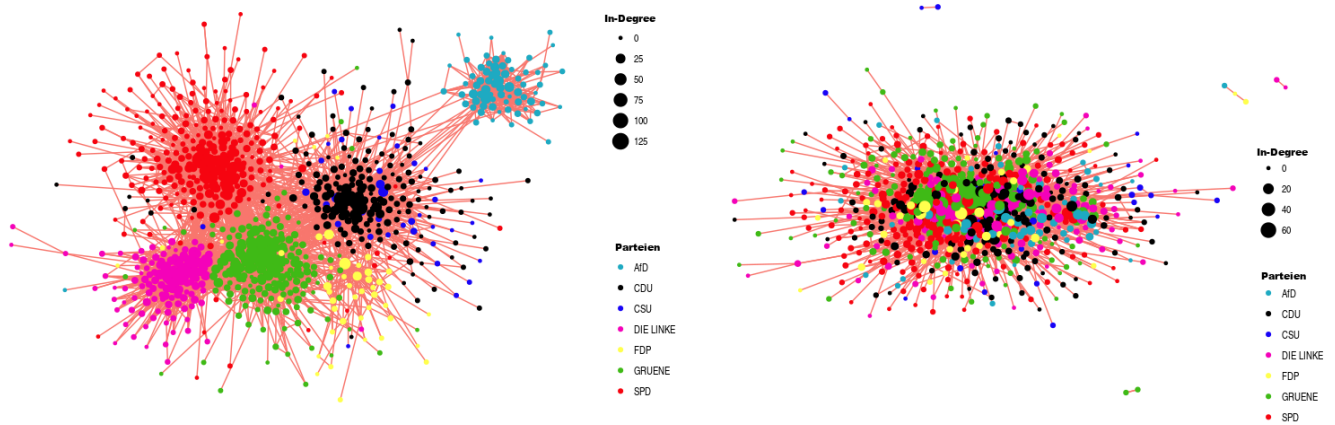
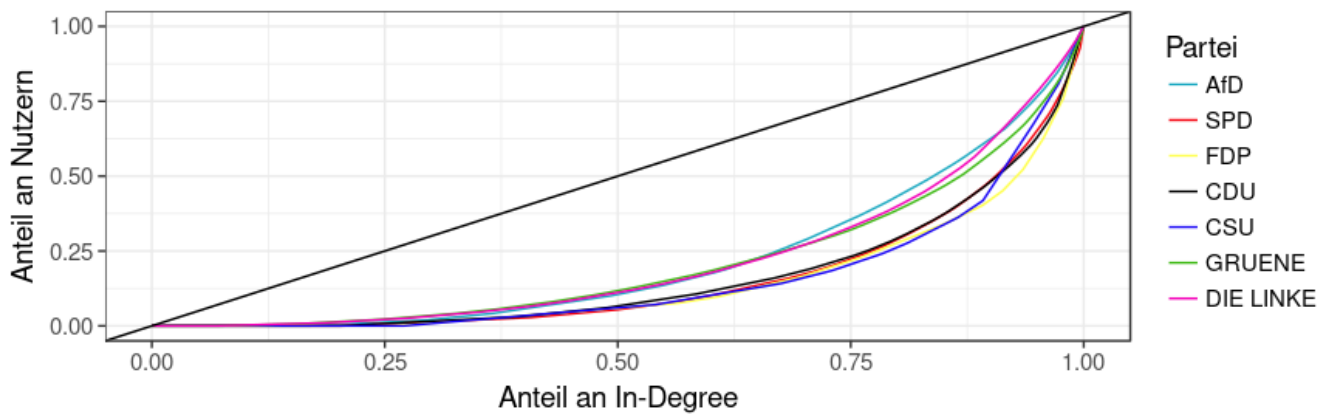


Figure 2

Das Retweet-Netzwerk (links) und das Reply-Netzwerk (rechts), gestaltet mit dem graphopt-Algorithmus[23]. Die Knoten repräsentieren die Twitter-Accounts der Politiker der einzelnen Parteien.



	Partei	Gini
1	FDP	0.7067580
2	CSU	0.6938854
3	SPD	0.6887992
4	CDU	0.6803777
5	GRUENE	0.5832635
6	DIE LINKE	0.5715053
7	AfD	0.5688045

Figure 3

Die Lorenz-Kurve der In-Degree-Verteilung jeder Partei (oben) und die zugehörigen Gini-Koeffizient (unten).

	Communityname	Tophashtags
1	klimaschutz	energiewende, hessen, kohleausstieg, klima, klimawandel
2	nrw	duesseldorf, piraten, ltnrw, noafd, wahlkampf
3	niedersachsen	gruen, ltwnds, jamaika, agrarwende, bienen
4	berlin	neukoelln, bezirkstag, armut, mahe, kinder
5	merkel	tvduell, dobrindt, dieselgate, umwelt, zukunft

Table 6: Top fünf Hashtagcluster der Grünen

Klimawandel wird deutlich worauf die Grünen ihren Hauptaugenmerk lenken. Ähnliche speziellen Themen lassen sich auch bei der FDP (Table 9) erkennen. Im Cluster *Pforzheim* lässt sich das Themengebiet Baden-Württemberg für die FDP erkennen. Traditionell gilt *Badenwuerttemberg* als "Stammland" der FDP guten Ergebnissen. Außerdem ist das Thema *Digitalisierung* zu erkennen, was ein Hauptwahlkampfthema der FDP gewesen ist, dies wird z.B. durch Wahlsprüche wie "Digitalsierung first. Bedenken second." deutlich. Auch die AfD (Table 7) hat ihre eigenen Themenbereiche. Ein wichtiger Punkt ist das Netzdurchsetzungsgesetz (*netzdg*). Hintergrund ist ein Gesetz, welches die Verteilung von sog. *fake news*, Hetze und ähnlichem innerhalb von sozialen Netzwerken eindämmen soll [1]. Die AfD kommuniziert sehr viel über soziale Netzwerke und sieht sich dementsprechend wahrscheinlich in ihrer *Meinungsfreiheit* eingeschränkt, welche Teil einer *Demokratie* sein soll. Konkret angesprochen wird außerdem noch Justizminister Heiko *Maas*, der für das Gesetz verantwortlich ist. Neben dem parteieigenen Hashtag *traudichdeutschland*, der vor allem zum Wahlkampf für die Spitzenkandidaten der AfD dient, verwendet die AfD den Hashtag *Merkel* anders als die Mehrheit der Parteien. Das wird durch den Ausdruck *merkelwegpfeifen* extrem deutlich. Die AfD scheint Angela Merkel sehr stark abzulehnen und sich einen Regierungswechsel zu wünschen. Auch bei den Linken (Table 10) lässt sich ein ähnliches Phänomen beobachten. Der Hashtag der großen Koalition (*groko*) wird hier in Kombination mit eigenen Themengebieten wie *Armut*, *Rente* oder *Hartziv* kombiniert. Es scheint möglich, dass in solchen Fällen sogenanntes "hijacking" von Hashtags verwendet wird (vgl. [11]). Dies lässt sich aber mit unserem Forschungsansatz, aber nicht klären. Neben dem bereits bei der SPD angesprochenen Effekte der Echokammer ist der Effekt der *cyber balkanization* ein oft beobachtetes Phänomen im Kontext sozialer Netzwerke [2, 10]. Auch in unserem Fall lässt sich ein solcher Effekt, zumindest in gewissem Maße, feststellen, wenn man die verschiedenen Themen der einzelnen Parteien betrachtet.

5 SCHLUSSFOLGERUNGEN

Mit dieser Studie wollen wir einen Einblick in die Social-Network-Analysis und das Text-Mining mithilfe computergestützter Verfahren schaffen. Wir haben uns auf einen Twitter-Datensatz von Politikern zur Bundestagswahl 2017 in Deutschland fokussiert. Natürlich sei an dieser Stelle angemerkt, dass die sozialen Medien nur einen Teil der politischen Kommunikation zwischen Parteien ausmachen, und noch einige andere Kurznachrichtendienste neben

	Communityname	Tophashtags
1	thueringen	r2g, btw, gera, ramelueringen, erfurt
2	nrw	ltnrw, petry, ltw nrw, landtag, ltw nrw17
3	netzdg	demokratie, maas, altparteien, bundestag, meinungsfreiheit
4	traudichdeutschland	afdwaehlen, gauland, wahlkampf, weidel, holdirdeinlandzurueck
5	merkel	merkelwegpfeifen, merkels, soros, ,diesel ,heidelberg ,orban, seehofer

Table 7: Top fünf Hashtagcluster der AfD

	Communityname	Tophashtags
1	bundestag	heimat, demokratie, inklusion, viersen, lammert
2	nrw	tvduell, ltw nrw, sicherheit, wahlarena, nrwgehtvor
3	sachsen	deutschland, saxit, europa, bildung, grossenhain
4	berlin	brandenburg, kreisreform, vibb, polizei, feuerwehr
5	r2g	thueringen, ilmenau, gebietsreform, gotha, ramelow

Table 8: Top fünf Hashtagcluster der CDU

	Communityname	Tophashtags
1	pforzheim	stuttgart, politik, badenwuerttemberg, enzkreis, workworkwork
2	nrw	esgehtumunserland, nrwkoalition, ltw nrw17, landtag, ltnrw, politikwechsel
3	bildung	merkel, ltsh, groko, jamaika, schulz
4	berlin	agh, tegel, r2g, txl, ber
5	digitalisierung	ehefueralle, zukunft, startup, bundestag, diskussion

Table 9: Top fünf Hashtagcluster der FDP

	Communityname	Tophashtags
1	sachsen	leipzig, afghanistan, dresden, saxit, gefluechtete
2	berlin	wahlkampf, wedding, mieten, muenchen, neukoelln
3	r2g	agh, tegel, tegelschliessen, txl, amri, digitalisierung
4	groko	armut, rente, schulz, hartziv, arbeit
5	thueringen	noafd, themar, nazis, neonazis, nazi

Table 10: Top fünf Hashtagcluster der Linken

Twitter die Medienlandschaft prägen. Nichtsdestotrotz sind die sozialen Medien ein wichtiger Teil moderner Massenmedien und sind durch die Transparenz und Einfachheit der Kommunikation sowohl beim Bürger als auch beim Politiker beliebt. Die Tatsache, dass die User dadurch geneigt sind oft zu posten, entsteht eine

schiere Menge an Inhalt, die der Wissenschaft als gute Quelle dient um Fragestellungen nachzukommen. Wir haben uns ein Bild davon machen können, welche innerparteilichen Strukturen sich um die Bundestagswahl auf Twitter aufgetan haben. Die etablierten Parteien als auch die FDP, die sich mit ihrem Kanzlerkandidaten Christian Lindner fast gänzlich auf eine Person konzentriert haben, weisen deutlich hierarchischere Strukturen auf als der Rest der Parteien. Außerdem haben wir mithilfe von Topic Modelling Kernthemen der Wahlkampfdiskussion identifizieren können. Daraus ist auch deutlich geworden, dass Parteien einige Themen vorherrschend selbst besetzen. Es liegt sogar der Verdacht nahe, dass innerhalb der SPD mit Hinblick auf Martin Schulz *echo chambers* auftreten. Der AfD haben wir zudem eine aggressive Anti-Merkel-Politik nachweisen können. Für die weitere zukünftige Arbeit bedarf es weiterhin der Forschungsarbeit um zu wissen, welche Schlüsse wir für Politiker und Parteien anhand dieser Erkenntnisse für zukünftige Wahlen ziehen können. Mit der anstehenden Landtagswahl 2018 in Bayern ist das nächste politische Großereignis in naher Ferne und es wird interessant, inwieweit unsere Beobachtungen drauf ebenso zutreffen werden.

REFERENCES

- [1] [n. d.]. NetzDG Netzwerkdurchsetzungsgesetz. ([n. d.]). <https://www.buzer.de/s1.htm?g=Netzwerkdurchsetzungsgesetz++NetzDG>
- [2] Lada A Adamic and Natalie Glance. 2005. The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. (2005), 36–43.
- [3] Pablo Aragón, Yana Volkovich, David Laniado, and Andreas Kaltenbrunner. 2016. When a Movement Becomes a Party: Computational Assessment of New Forms of Political Organization in Social Media. *Icwsm* (2016), 12–21.
- [4] Alain Barrat, Marc Barthélémy, Romualdo Pastor-Satorras, and Alessandro Vespignani. 2003. The architecture of complex weighted networks. (2003). [arXiv:arXiv:cond-mat/0311416v1](https://arxiv.org/pdf/cond-mat/0311416v1) <https://arxiv.org/pdf/cond-mat/0311416.pdf>
- [5] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.
- [6] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008, 10 (2008), P10008. <http://stacks.iop.org/1742-5468/2008/i=10/a=P10008>
- [7] Leticia Bode, Alexander Hanna, Junghwan Yang, and Dhavan V Shah. 2015. Candidate Networks, Citizen Clusters, and Political Expression: Strategic Hashtag Use in the 2010 Midterms. *The ANNALS of the American Academy of Political and Social Science* 659, 1 (2015), 149–165. <https://doi.org/10.1177/0002716214563923>
- [8] Javier Borge-Holthoefer, Walid Magdy, Kareem Darwish, and Ingmar Weber. 2014. Content and Network Dynamics Behind Egyptian Political Polarization on Twitter. *CoRR abs/1410.3097* (2014). [arXiv:1410.3097](https://arxiv.org/abs/1410.3097) <http://arxiv.org/abs/1410.3097>
- [9] Bundeswahlleiter. 2017. Wahl zum 19. Deutschen Bundestag am 24. September 2017. Heft 3 (2017). <https://www.bundeswahlleiter.de/dam/jcr/3f3d42ab-faef-4553-bdf8-ac089b7de86a/btw17>
- [10] M D Conover, J Ratkiewicz, M Francisco, B Gonc, A Flammini, and F Menczer. 2010. Political Polarization on Twitter. (2010), 89–96.
- [11] Ryan J. Gallagher, Andrew J. Reagan, Christopher M. Danforth, and Peter Sheridan Dodds. 2016. Divergent discourse between protests and counter-protests: #BlackLivesMatter and #AllLivesMatter. (2016). [arXiv:1606.06820](https://arxiv.org/abs/1606.06820) <http://arxiv.org/abs/1606.06820>
- [12] C. Gini. 1912. *Variabilità e mutabilità*.
- [13] Frédéric Godin, Viktor Slavkovikj, Wesley De Neve, Benjamin Schrauwen, and Rik Van de Walle. 2013. Using Topic Models for Twitter Hashtag Recommendation. In *Proceedings of the 22Nd International Conference on World Wide Web (WWW '13 Companion)*. ACM, New York, NY, USA, 593–596. <https://doi.org/10.1145/2487788.2488002>
- [14] Alexander Häusler. 2016. *Die Alternative für Deutschland: Programmatik, Entwicklung und politische Verortung*. Springer-Verlag.
- [15] Kathleen Hall Jamieson and Joseph N Cappella. 2008. *Echo chamber: Rush Limbaugh and the conservative media establishment*. Oxford University Press.
- [16] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. 2007. Why We Twitter: Understanding Microblogging Usage and Communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis (WebKDD/SNA-KDD '07)*. ACM, New York, NY, USA, 56–65. <https://doi.org/10.1145/1348549.1348556>
- [17] Lou Jost. 2006. Entropy and diversity. *Oikos* 113, 2 (2006), 363–375.
- [18] Andreas Jungherr. 2016. Twitter use in election campaigns: A systematic literature review. 1681, March (2016). <https://doi.org/10.1080/19331681.2015.1132401>
- [19] Haiko Lietz, Claudia Wagner, Arnim Bleier, and Markus Strohmaier. 2014. When Politicians Talk: Assessing Online Conversational Practices of Political Parties on Twitter. *ICWSM* (2014).
- [20] Avishay Livne, Matthew P Simmons, Eytan Adar, and Lada a Adamic. 2010. The Party is Over Here : Structure and Content in the 2010 Election. *October* 161, 3 (2010), 201–208. <https://doi.org/10.1007/s00024-003-2459-0>
- [21] Senat Müller. [n. d.]. (ab 2014) - Berlin.de. ([n. d.]).
- [22] M. E. J. Newman and M. Girvan. 2004. Finding and evaluating community structure in networks. *Phys. Rev. E* 69 (Feb 2004), 026113. Issue 2. <https://doi.org/10.1103/PhysRevE.69.026113>
- [23] Michael Schmuhl. 2003. <http://www.schmuhl.org/graphopt/>
- [24] Sebastian Stier, Lisa Posch, Arnim Bleier, and Markus Strohmaier. 2017. When populists become popular: comparing Facebook use by the right-wing movement Pegida and German political parties. *Information Communication and Society* 20, 9 (2017), 1365–1388. <https://doi.org/10.1080/1369118X.2017.1328519>
- [25] Jeffrey Travers and Stanley Milgram. 1967. The small world problem. *Psychology Today* 1 (1967), 61–67.
- [26] Stanley Wasserman and Katherine Faust. 1994. *Social network analysis: Methods and applications*. Vol. 8. Cambridge university press.
- [27] Duncan J. Watts and Steven H. Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature* 393, 6684 (jun 1998), 440–442. <https://doi.org/10.1038/30918>
- [28] Shaomei Wu, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. 2011. Who Says What to Whom on Twitter. In *Proceedings of the 20th International Conference on World Wide Web (WWW '11)*. ACM, New York, NY, USA, 705–714. <https://doi.org/10.1145/1963405.1963504>