

Regionale Charakteristiken politischer Teilhabe in Onlinemedien

Eine Vergleichende Betrachtung der Korrelation von
Wikipediaartikelbearbeitungen mit Breitbandausbau und
Wahlergebnissen

Bachelorarbeit

zur Erlangung des Grades eines Bachelor of Science (B.Sc.)
im Studiengang Informatik

vorgelegt von
Jan Rehfisch

Universität Koblenz-Landau

Erstgutachter: Prof. Dr. Steffen Staab
Institute for Web Science and Technologies

Zweitgutachter: Dr. Oul Han
Institute for Web Science and Technologies

Koblenz, im November 2019

Erklärung

Hiermit bestätige ich, dass die vorliegende Arbeit von mir selbstständig verfasst wurde und ich keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe und die Arbeit von mir vorher nicht in einem anderen Prüfungsverfahren eingereicht wurde.

Ja Nein

Mit der Einstellung dieser Arbeit in die Bibliothek bin ich einverstanden.

Der Veröffentlichung dieser Arbeit im Internet stimme ich zu.

Der Text dieser Arbeit ist unter einer Creative Commons Lizenz (CC BY-SA 4.0) verfügbar.

Der Quellcode ist unter einer GNU General Public License (GPLv3) verfügbar.

Die erhobenen Daten sind unter einer Creative Commons Lizenz (CC BY-SA 4.0) verfügbar.

.....
(Ort, Datum)

(Unterschrift)

Zusammenfassung

In dieser Arbeit wird der Zusammenhang zwischen der Beteiligung an Wikipediaartikeln über aktive Politiker des Deutschen Bundestages und dem Breitbandausbau innerhalb der Bundesländer untersucht. Außerdem wird die Korrelation zwischen Artikelbearbeitungen und Wahlergebnissen der Bundestagswahl 2017 in den einzelnen Ländern betrachtet. Dazu wird der Umstand genutzt, dass unangemeldete Nutzer bei der Bearbeitung eines Wikipediaartikels ihre IP-Adresse hinterlassen. Aus diesen IP-Adressen lassen sich Informationen über den Ursprungsort der Bearbeitung gewinnen. Es konnte im Gegensatz zu ähnlichen Arbeiten kein signifikanter Zusammenhang zwischen politischer Partizipation in Form von Artikelbearbeitungen und Breitbandausbau festgestellt werden. Es konnte jedoch gezeigt werden, dass es eine positive Korrelation zwischen der Anzahl von Artikelbearbeitungen und dem Wahlergebnis der Partei die Grünen gibt.

Abstract

This paper examines the relationship between the participation in Wikipedia articles about current politicians in the German Parliament and broadband expansion within the federal states. In addition, the correlation between article editing and election results of the 2017 German parliamentary elections in the individual states will be examined. In order to make this possible, the fact that unregistered users leave their IP address when editing a Wikipedia article is used. From these IP addresses, information about the place of origin of the changes can be obtained. In contrast to similar work, no significant connection between political participation, in the form of article editing, and broadband expansion could be shown. However, it could be shown that there is a positive correlation between the number of edits of articles and the election results of the Green Party.

Inhaltsverzeichnis

1	Motivation.....	1
2	Literatur.....	2
3	Vorgehen	3
4	Grundlagen	4
4.1	Wikipedia.....	4
4.2	Wiki.....	5
4.3	Wikimedia-API	5
4.4	IP-Adressen.....	7
4.5	Geo-Daten.....	8
5	Daten	9
5.1	Daten zu Politikern	9
5.2	Daten zur Partizipation	10
6	Statistische Methoden.....	13
6.1	Lineare Regression	13
6.2	R^2	14
6.3	Pearson Korrelation.....	14
7	Hypothesen.....	15
8	Analyse	16
8.1	Breitbandausbau	16
8.2	Wahlergebnis.....	23
9	Diskussion.....	27
	Quellen:	29

1 Motivation

Wikipedia ist das am häufigsten genutzte digitale Nachschlagewerk [1]. Dies macht Wikipedia auch zu einem relevanten Medium der Meinungsbildung. Allerdings sind nicht nur die auf Wikipedia gespeicherten Informationen frei zugänglich, auch das Erstellen und Bearbeiten von Artikeln ist jeder Person, die über einen Internetzugang verfügt, möglich. Diese Form der Partizipation an Onlineaktivitäten ist in der Bevölkerung nicht gleichverteilt. Verschiedene Faktoren wie zum Beispiel die lokale Internetgeschwindigkeit (Breitbandausbau) und soziodemokratische Merkmale haben Einfluss auf das Nutzungsverhalten der Bevölkerung [2]. Somit besteht die Möglichkeit, dass der Inhalt Wikipedias in erster Linie die Ansichten derer widerspiegelt, die maßgeblich an der Erstellung von Artikeln beteiligt sind. Dies bietet die Grundlage für die Fragestellung, ob die Kombination aus freiem Zugang und ungleichverteilter Nutzung sich auch in der Autorenschaft von Wikipedia und in den von ihr verfassten Artikeln widerspiegelt. Dies hat besondere Relevanz für Artikel, die sich mit politischen Themen auseinandersetzen. Denn wenn sich herausstellt, dass maßgeblich externe Faktoren wie der Ausbau von Breitbandanschlüssen für die Partizipation an Wikipediaartikeln verantwortlich sind, wäre damit aufgezeigt, dass es im Interesse eines demokratischen Systems sein muss, allen Menschen einen leistungsfähigen Internetzugang zu ermöglichen, da angenommen werden kann, dass politische Informationen eine wichtige Entscheidungsbasis für Wahlen bilden. Des Weiteren könnte eine Korrelation von Artikelbearbeitungen mit den Wahlergebnissen einer bestimmten Partei, darauf hindeuten, dass die Beteiligung an der Erstellung von Wikipediaartikeln in der Bevölkerung nicht gleichverteilt ist und somit nicht alle Politiker gleichermaßen repräsentiert werden.

2 Literatur

Da in dieser Arbeit untersucht werden soll ob und in wieweit es eine ungleich verteilte Partizipation an Wikipediaartikeln zu Abgeordneten des Deutschen Bundestages gibt, ist es zunächst nötig, eine Methodik zu entwickeln, die es erlaubt von Wikipediaartikelbearbeitungen Rückschlüsse auf die Herkunft der Autoren zu ziehen. Eine vergleichbare Fragestellung verfolgten 2015 Mitarbeiter der Universität Oxford [3]. Auf der Basis von IP-Adressen von Wikipedia wurde der Zusammenhang von Partizipation an Onlineinhalten und dem Ausbau von Breitbandschlüssen untersucht und eine positive Korrelation festgestellt. Es wurden die IP-Adressen von verschiedenen Sprachversionen Wikipedias gesammelt und mithilfe einer Adressdatenbank verschiedenen Ländern zugeordnet. Diese Daten wurden anschließend mit dem jeweiligen Breitbandausbau einer Region verglichen.

Dari E. Sylvester und Adam J. McGlynn haben den Zusammenhang von Breitbandausbau und politischer Partizipation, in Form von Kontaktaufnahme zu Regierungsvertretern, untersucht und ebenfalls eine positive Korrelation festgestellt [4]. Ihre Analyse hat außerdem gezeigt, dass Menschen aus ländlichen Regionen deutlich weniger partizipiert haben als Menschen aus städtischen Regionen.

Eine Grundlage für die Fragestellung ob es einen Zusammenhang zwischen dem Wahlergebnis einer Partei bei der Bundestagswahl 2017 in einem Bundesland und der Anzahl an Artikelbearbeitungen aus diesem Land gibt, liefert die Arbeit von Andreas Jungherr, Pascal Jürgens und Harald Schoen [5]. Sie haben anhand von Twitterbeiträgen die Repräsentanz von politischen Parteien untersucht und festgestellt, dass die Anzahl der Beiträge zu einer Partei nicht zwangsläufig mit deren Wahlergebnis korreliert. Sie konnten zeigen, dass die 2009 nicht in den Bundestag eingezogene Piratenpartei auf Twitter deutlich überrepräsentiert war.

3 Vorgehen

Es ist jedem Besucher eines Wikipediaartikels möglich, dessen Inhalt zu bearbeiten. Damit die Übersicht darüber gewahrt wird wie und wann ein Artikel geändert wurde, werden alle Änderungen in der sogenannten Revisionsgeschichte gespeichert. Um diese Änderungen einem spezifischen Nutzer zuweisen zu können, wird auch eine Nutzerkennung gespeichert. Jeder Editor eines Wikipediaartikels hinterlässt entweder seinen Nutzernamen oder seine IP-Adresse. Wenn der Nutzer ein Nutzerkonto bei Wikipedia hat, wird dessen Nutzernamen gespeichert, andernfalls wird dessen IP-Adresse anstelle eines Nutzernamens in der Revisionsgeschichte hinterlegt. In dieser Arbeit werden nur solche Beiträge berücksichtigt, bei denen die Nutzer ihre IP-Adresse hinterlassen haben. Denn aus der IP-Adresse lassen sich mithilfe verschiedener Datenbanken Informationen über die Herkunft des Nutzers gewinnen. Mithilfe der Programmierschnittstelle von Wikipedia ist es möglich alle IP-Adressen, die sich in der Revisionsgeschichte eines bestimmten Artikels befinden auszulesen. Für diese Arbeit werden die IP-Adressen von allen Artikeln zu aktiven Politikern des Deutschen Bundestages ausgelesen. Von jeder IP-Adresse wird anschließend ermittelt aus welchem Bundesland sie stammt. Die gewonnenen Daten werden dann in einer Struktur gespeichert, die eine vergleichende Betrachtung ermöglicht.

Die grundlegende Annahme dieser Arbeit ist, dass es einen signifikanten Zusammenhang zwischen der Verbreitung von Breitbandinternetanschlüssen in einem bestimmten Bundesland und der Beteiligung an Artikeln von Wikipedia gibt. Um dies zu überprüfen, wird die Beteiligung an Wikipediaartikeln in einem Bundesland ermittelt und mit dem entsprechenden Breitbandausbau verglichen. Außerdem wird untersucht, ob die Präferenz einer bestimmten Partei, die Beteiligung der Bevölkerung besser erklären kann. Um diese Frage beantworten zu können werden neben statistischen Methoden wie der linearen Regression und entsprechender Analyse mit R^2 und dem Pearson Korrelationskoeffizienten auch verschiedene Diagramme und Karten genutzt.

4 Grundlagen

In diesem Abschnitt werden die nötigen Grundlagen zu Wikipedia und IP-Adressen erläutert.

4.1 Wikipedia

Ein großer Teil, der in dieser Arbeit verwendeten Daten stammt von Wikipedia. Um nachvollziehbar zu machen, warum Wikipedia als Untersuchungsgegenstand gewählt wurde, soll der folgende Abschnitt die Relevanz von Wikipedia als Informationsplattform verdeutlichen.

Das 2001 ins Leben gerufene Onlinelexikon Wikipedia ist das am häufigsten genutzte digitale Nachschlagewerk. Mittlerweile zählt allein die deutschsprachige Version über 2 Millionen Artikel [6]. Verfasst wurden diese Artikel von freiwilligen Autoren. Jeder Person mit einem Internetzugang ist es möglich, bei Wikipedia als Autor mit zu wirken oder als Leser Information zu sammeln. Dies macht Wikipedia zu einem relevanten Informationsmedium und somit zu einem interessanten Untersuchungsgegenstand für diese Arbeit. Wikipedia verwendet die im nächsten Kapitel erläuterte Wiki Technologie, um es Nutzern möglich zu machen ohne technische Vorkenntnisse an der Erstellung von Artikeln mit zu wirken. Es ist jedem Nutzer möglich einen neuen Artikel zu erstellen oder den Inhalt eines bestehenden Artikels zu verändern. Die Bearbeitung oder Erstellung von Wikipediartikeln erfordert nur ein geringes technisches Verständnis und ist jeder Person mit grundlegenden Computerkenntnissen möglich. Da das technische Verständnis nur eine geringe Rolle bei der Frage spielt, wovon es abhängig ist, ob Personen einen Artikel bearbeiten, soll untersucht werden, welche externen Faktoren maßgeblich Einfluss nehmen.

4.2 Wiki

Um besser nachvollziehen zu können woher die gewonnen Daten stammen und wie sie gesammelt wurden, werden im Folgenden die grundlegenden technischen Eigenschaften der hinter Wikipedia liegenden Wiki Technologie erläutert.

Die freie online Enzyklopädie Wikipedia baut, wie viele andere Informationsplattformen auch auf der Wiki Technologie auf. Als Wiki (Hawaiianisch für schnell) werden einfache online Content-Management-Systeme bezeichnet, die es Nutzern erlauben den Inhalt einer Seite direkt im Browser zu bearbeiten. Dabei wird zugunsten einer einfachen Anwendbarkeit auf tiefergehende Gestaltungsmöglichkeiten verzichtet. Das Design der Seite kann also von den Nutzern nur in geringem Maße beeinflusst werden, das Bearbeiten des Inhaltes ist allerdings mit nur wenig Aufwand möglich. Auf diese Weise ist es einem Kollektiv von Nutzern ohne tiefergehende technische Kenntnisse möglich, gemeinschaftlich den Inhalt einer Seite zu generieren und zu verändern. Um den Inhalt sowohl vor Vandalismus zu schützen als auch einen Verlauf über die Entstehung eines Artikels zu haben, ist jedem Artikel eine Versionsgeschichte zugeordnet. Darin werden nach jeder Änderung der gesamte Artikel sowie einige Metadaten, wie etwa der jeweilige Autor, das Datum und auch die Größe der Änderung in Bytes gespeichert.¹

4.3 MediaWiki-API

Im folgenden Abschnitt soll gezeigt werden, wie es möglich ist, Daten von Wikipediaservern zu beziehen.

Eine API, kurz für application programming interface, wird von einem System bereitgestellt um es anderen Programmen zu ermöglichen, Anfragen zu stellen. Dieser Anfragen werden von dem Server, an den sich die Anfrage richtet, gelesen und

¹ Die Revisionsgeschichte von Angela Merkel etwa erreicht man über https://de.wikipedia.org/w/index.php?title=Angela_Merkel&action=history.

bearbeitet. Die Form der Anfrage muss einem strikten Protokoll folgen, damit der Server sie entsprechend auswerten kann. Mit einer Anfrage an einen Wikipediaserver zu einem bestimmten Titel, ist es beispielsweise möglich, den entsprechenden Artikel im HTML-Format als Antwort zu erhalten. Es lassen sich aber auch konkretere Anfragen formulieren.

Das folgende Python-Programm in Algorithmus 1 ermöglicht die Abfrage aller Nutzer die Änderungen an dem Artikel „Universität Koblenz-Landau“ vorgenommen haben. Die Daten lassen sich mithilfe einer spezifischen URL gezielt abfragen. Die Bibliothek `requests` vereinfacht das Aufrufen von Internetseiten aus dem Programm heraus. Dadurch ist es leicht möglich den Inhalt einer Seite in einem gewünschten Datentyp zu speichern.

In Zeile 4 bis 13 des Programms wird folgende URL generiert:

```
https://de.wikipedia.org/w/api.php?action=query&format=json&titles=Universit%C3%A4t_Koblenz-Landau&prop=revisions&rvprop=user|userid&rvlimit=max
```

Ruft man die URL in einem Browser auf, werden die ersten 500 Änderungen an dem Wikipediaartikel „Universität Koblenz-Landau“ angezeigt.

Mit `'prop': 'revisions'` in Zeile 11 wird festgelegt, dass man auf die Revisionsgeschichte zugreifen möchte. Also die Sammlung, aller an einem bestimmten Artikel gemachter Änderungen. Mit `'rvprop': 'user|userid'` in Zeile 12 wird festgelegt, dass man von jeder gespeicherten Änderung den Namen und die Id des jeweiligen Nutzers haben möchte. Wenn der Nutzer in einem Wikipediakonto angemeldet ist, ist unter `'user'` dessen Benutzername gespeichert, andernfalls dessen IP-Adresse. Die `'userid'` ist bei unangemeldeten Nutzern immer 0. Mithilfe der Id ist es also möglich nach unangemeldeten Nutzern zu filtern. Das Folgende Beispiel zeigt einen Ausschnitt aus dem Ergebnis der Abfrage im json-Format.

```
1. {,user': ,Firefox13', ,userid': 66337}
2. {,user': ,89.55.29.211', ,anon': ,', ,userid': 0}
3. {,user': ,84.188.250.173', ,anon': ,', ,userid': 0}
4. {,user': ,SieBot', ,userid': 384519}
5. {,user': ,Swann', ,userid': 172997}
```

Da eine Abfrage von Daten über die API auf 500 Einträge begrenzt ist, ist es nötig, die Abfrage so lange zu wiederholen bis keine weiteren Einträge mehr vorhanden sind. In Zeile 26 des Programms wird der Startwert der neuen Anfrage auf den Endwert der

Vorherigen gesetzt. Es wird also bei der letzten ausgelesenen Änderung mit dem Auslesen fortgefahren, bis keine weiteren Einträge mehr vorhanden sind. In der letzten Zeile werden anschließend alle IP-Adressen von unangemeldeten Nutzern in einer Liste gespeichert und alle mehrfach vorkommenden Einträge werden entfernt.

```
1. import requests
2.
3.
4. title = 'Universität Koblenz-Landau'
5.
6. BASE_URL = 'http://de.wikipedia.org/w/api.php'
7.
8. parameters = { 'action': 'query',
9.                'format': 'json',
10.               'titles': title,
11.               'prop': 'revisions',
12.               'rvprop': 'user|userid',
13.               'rvlimit': 'max'}
14.
15. resultList = []
16.
17. while True:
18.
19.     wp_call = requests.get(BASE_URL, params=parameters)
20.     response = wp_call.json()
21.
22.     for page_id in response['query']['pages']:
23.         resultList = resultList + response['query']['pages'][page_id]['revisions']
24.
25.     if 'continue' in response:
26.         parameters['rvcontinue'] = response['continue']['rvcontinue']
27.     else:
28.         break
29.
30. ipList = list(dict.fromkeys([x['user'] for x in resultList if x['userid']==0]))
```

Algorithmus 1: Alle IP-Adressen, von Änderungen an einem Wikipediaartikel

4.4 IP-Adressen

In diesem Abschnitt wird neben dem grundlegenden Aufbau von IP-Adressen auch auf die hierarchische Gliederung, der bei der Vergabe von IP-Adressen beteiligten Institutionen eingegangen.

Bei IP-Adressen handelt es sich um eine Nummer, deren Zweck es ist, dass an ein Netzwerk angeschlossene Computer innerhalb dieses Netzes eindeutig identifizierbar sind. IP-Adressen sind in diesem Sinne sehr analog zu Post-Adressen. Jeder Internetzugangspunkt besitzt eine eindeutige Adresse. Diese Adresse kann sich ändern. In der Regel wird einem Computer jedes Mal, wenn er versucht, sich mit dem Internet zu verbinden, von dem übergeordneten Netzwerk, in dem er sich befindet, eine IP-Adresse zugewiesen. Diese Zuweisung erfolgt nicht willkürlich. Netzwerke haben eine bestimmte Reichweite von IP-Adressen, die sie vergeben können. Eines der gängigsten IP-Adressenformate IPv4, besteht beispielsweise aus 4 Bytes, die zur besseren Lesbarkeit jeweils durch einen Punkt getrennt werden. Eine vollständige IPv4 Adresse kann zum Beispiel so aussehen 141.26.67.87. In dieser Form ist die Darstellung von bis zu 2^{32} verschiedenen Adressen möglich. Denn mit jedem Byte lassen sich 2^8 Zahlen darstellen. Da die Adresse aus 4 Bytes besteht gilt, $(2^8)^4 = 2^{32}$.

Die Vergabe von IP-Adressen erfolgt nicht willkürlich. Die übergeordnete Verantwortung für die Verteilung von IP-Adressen obliegt der „Internet Assigned Numbers Authority (IANA)“. Verschiedenen Regionen stehen verschiedene Adressbereiche zur Verfügung. Verantwortlich für die regionale Verteilung von IP-Adressen sind die jeweiligen „Regional Internet Registries“. Für die Verteilung im europäischen Raum ist beispielsweise das „Réseaux IP Européens Network Coordination Centre (RIPE NCC)“ verantwortlich. RIPE NCC sind wiederum lokale Organisationen untergeordnet, die „Local Internet Registry“. Bei diesen handelt es sich in der Regel um Internetdienstleister. [7]

4.5 Geo-Daten

Im Folgenden soll erläutert werden, wie es möglich ist mithilfe der IP-Adressen Information über den Standort eines Internetzugangspunktes zu gewinnen.

Der hierarchischen und streng dokumentierten Vergabemethode von IP-Adressen ist es zu verdanken, dass aus der IP-Adresse Rückschlüsse auf die Ursprungsregion der Adresse gezogen werden können. Denn die Local Internet Registry muss angeben für

welchen Zweck und welche Region sie einen Adressblock, also eine bestimmte Anzahl an IP-Adressen benötigt. Diese Informationen können mit weiteren Daten verknüpft werden. Falls ein Internetnutzer sich nicht aktiv darum bemüht seine IP-Adresse zu verbergen, kann jede Internetseite, die besucht wird, die IP-Adresse lesen. Wenn der Besucher dann Informationen hinterlässt, die Rückschlüsse auf dessen Wohnort erlauben, kann die IP-Adresse mit einer Region verbunden werden. Es gibt Unternehmen, die sich auf das Sammeln und Handeln von IP-Adressen spezialisiert haben. Die für diese Arbeit verwendete Datenbank GeoLite2 City von MaxMind enthält nach eigenen Angaben in 74% der Fälle korrekte geografische Informationen zu einer bestimmten IP-Adresse in Deutschland. [8, 9]

5 Daten

In diesem Abschnitt werden die für die Arbeit relevanten gewonnenen Daten erläutert.

5.1 Daten zu Politikern

Für diese Arbeit wurden Daten von allen aktiven Mitgliedern des aktuellen Deutschen Bundestages gesammelt. Derzeit gibt es 709 Abgeordnete. Zu jedem dieser Abgeordneten konnten durchschnittlich 19,6 Änderungen von unangemeldeten Nutzern gesammelt werden. Zu insgesamt 50 Politikern konnten keine IP-Adressen gewonnen werden. Darunter fallen sowohl Politiker, deren Wikipediaartikel nicht von unangemeldeten Nutzern bearbeitet wurden, als auch Politiker deren Namen nicht mit dem Titel ihres Wikipediaartikels übereinstimmen. Wikipedia fügt dem Titel eine weitere Beschreibung der Person hinzu, falls es weitere Artikel zu gleichnamigen Personen gibt. Der Titel für die SPD Politikerin Dagmar Schmidt lautet beispielsweise „Dagmar Schmidt (Politikerin, 1973)“. Diese Problematik wurde erst in einem weit fortgeschrittenen Stadium der Arbeit festgestellt. Unter der Annahme, dass nicht eindeutige Namen unter allen Parteien und Regionen gleichverteilt sind, sollten die fehlenden Daten die Resultate nicht maßgeblich beeinflussen.

Tabelle 1 zeigt die Anzahl der Abgeordneten jeder Partei in den einzelnen Bundesländern. Parteilose abgeordnete und Mitglieder der Partei die Blauen wurden aufgrund ihrer geringen Anzahl, von jeweils 2, ignoriert. Sie werden allerdings als Politiker aus ihrem jeweiligen Bundesland in betreffenden Berechnungen berücksichtigt, wenn die Parteizugehörigkeit keine Rolle spielt.

State\Party	CDU/CSU	SPD	Grüne	Linke	FDP	AfD
Bayern	46	18	11	7	12	14
Nordrhein-Westfalen	42	40	12	12	20	13
Baden-Württemberg	38	16	13	6	12	11
Niedersachsen	21	20	6	5	7	7
Hessen	17	12	5	4	6	6
Rheinland-Pfalz	14	9	3	3	4	4
Sachsen	12	4	2	6	3	10
Schleswig-Holstein	10	6	3	2	3	2
Sachsen-Anhalt	9	3	1	4	2	4
Brandenburg	9	4	1	4	2	5
Thüringen	8	3	1	3	2	5
Mecklenburg-Vorpommern	6	2	1	3	1	3
Berlin	6	5	4	6	3	4
Hamburg	4	5	2	2	2	1
Saarland	3	3	1	1	1	1
Bremen	1	2	1	1	0	1

Tabelle 1: Anzahl der Abgeordneten nach Bundesland

Da die CSU ausschließlich in Bayern gewählt werden kann, und somit CSU Politiker auch nur in bayrischen Wahlkreisen gewählt werden können, spielen Partei und Bundesland eine gewisse Sonderrolle. Um diese Problematik zu umgehen, wird die CSU in Bayern als Teil der CDU betrachtet werden.

5.2 Daten zur Partizipation

Mit der zuvor erläuterten Methodik konnten insgesamt 13937 verschiedene Artikelbearbeitungen von unangemeldeten Wikipedianutzern gesammelt werden. Da nicht jeder Nutzer nur genau eine Änderung verfasst, ergibt sich nach dem

Aussortieren von doppelten Einträgen eine Anzahl von insgesamt 3684 Nutzern, die Artikel bearbeitet haben. Jeder unangemeldete Nutzer hat also im Durchschnitt ca. 3,8 Beiträge zu Artikeln von Politikern des Aktuellen Deutschen Bundestages verfasst. Von diesen 3684 Editoren stammen 3112 aus Deutschland und konnten einem Bundesland zugeordnet werden. Dabei ergibt sich folgende Verteilung.

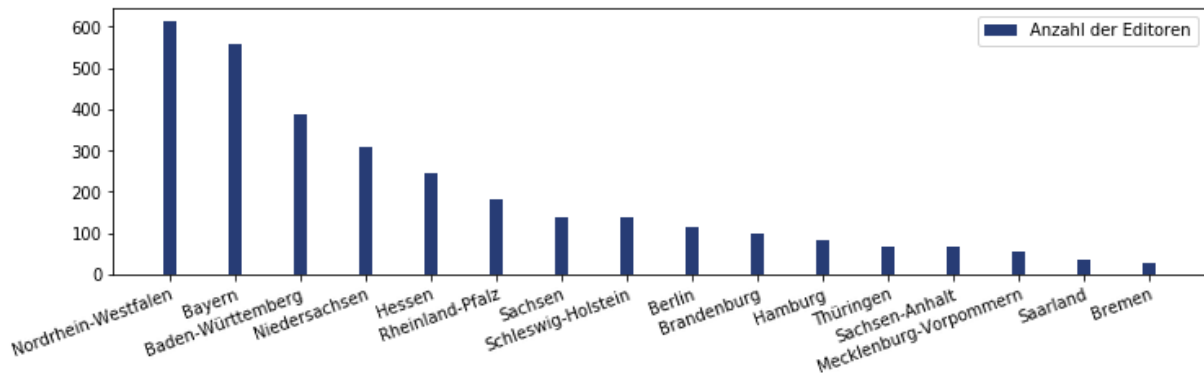


Abb.1: Anzahl der Editoren nach Bundesland

Das obige Balkendiagramm zeigt die ungleiche Verteilung von Editoren in den Bundesländern. Die Anzahl schwankt von 615 in Nordrhein-Westfalen zu lediglich 26 in Bremen. Aufgrund der relativ geringen Menge an Daten wird von einer tieferen Gliederung, als auf Bundesländerebene abgesehen werden müssen. Eine tiefere Aufteilung wie beispielsweise in einzelne Wahlkreise, hätte zufolge, dass es viele Gebiete gäbe, zu denen keine Daten existieren.

Um die Anzahl der Editoren mit der Anzahl an Einwohnern jedes Bundeslandes vergleichen zu können, wurden die anteiligen Werte in Prozent ermittelt. In Nordrhein-Westfalen leben beispielsweise ca. 21% der Einwohner Deutschlands und knapp 20% aller Personen die einen Artikel bearbeitet haben taten dies von einem Internetzugangspunkt, der sich ebenfalls in Nordrhein-Westfalen befindet. Abbildung 2 zeigt, dass es wie zu erwarten einen starken Zusammenhang zwischen der Einwohnerzahl eines Bundeslandes und der Anzahl an Editoren gibt.

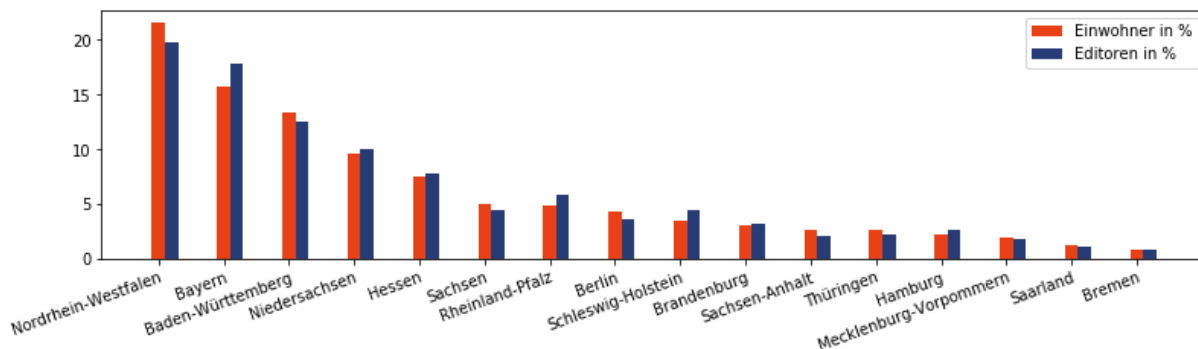


Abb. 2: Vgl. Einwohner - Editoren

Die Differenz zwischen der Anzahl der Einwohner und der Anzahl der Editoren gibt bereits Aufschluss darüber wie aktiv die Bewohner eines bestimmten Bundeslandes an der Gestaltung von Wikipediaartikeln mitgewirkt haben. In Nordrhein-Westfalen ist beispielsweise der Anteil an Einwohnern größer als der Anteil an Editoren. Dem gegenüber steht Bayern, wo der Anteil an Editoren größer ist als der Anteil an der Gesamtbevölkerung. In Bayern wurden folglich pro Einwohner mehr Artikel bearbeitet als in Nordrhein-Westfalen. Das Verhältnis der Anzahl von Editoren eines Bundeslandes zu der Anzahl der Einwohner, also die Anzahl der Editoren pro Einwohner, wird eine wichtige Grundlage für die weiteren Betrachtungen bilden.

In Abbildung 3 sind die Bundesländer in absteigender Reihenfolge nach dem Anteil von Editoren an der jeweiligen Bevölkerung dargestellt. Durchschnittlich kommen auf einhunderttausend Einwohner ca. 3,7 Editoren. In Schleswig-Holstein sind es mit ca. 4,7 die meisten und Sachsen-Anhalt mit lediglich ca. 2,9 die wenigsten. An Abbildung 3 lässt sich auch erkennen, dass bis auf Brandenburg alle neuen Bundesländer sich deutlich in der unteren Hälfte befinden.

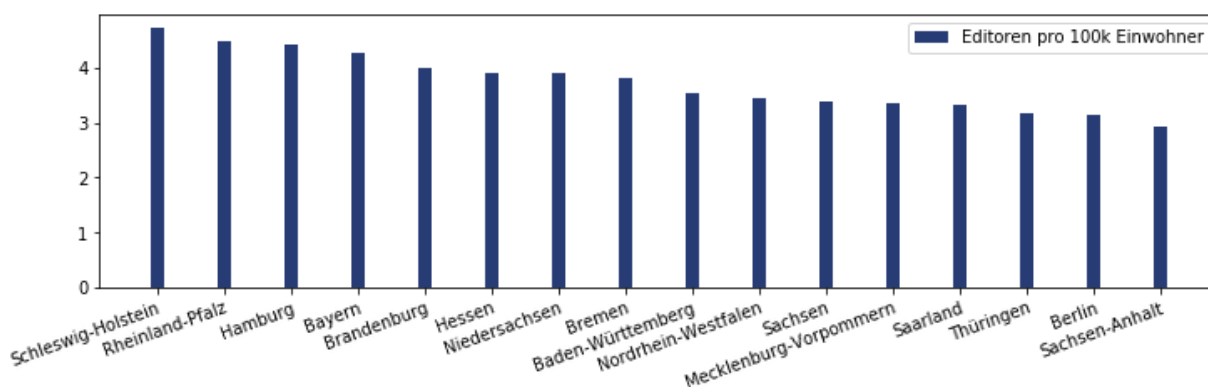


Abb. 3: Editoren pro Einwohner

6 Statistische Methoden

In diesem Kapitel werden die Methoden erläutert, mit deren Hilfe die Daten analysiert werden.

6.1 Lineare Regression

Regression ist ein statistisches Verfahren, welches genutzt wird, um den Zusammenhang zwischen verschiedenen Variablen untersuchen zu können. Ziel der linearen Regression ist es, eine Gerade zu finden, die möglichst genau das Verhältnis verschiedener Messwerte beschreibt. Den einfachsten Fall stellt dabei die einfache lineare Regression mit einer Einflussgröße X und einer Zielgröße Y dar. Dabei ist das Ziel, herauszufinden ob die gemessenen Werte der abhängigen Variable Y mit der unabhängigen Variable X erklärt werden können.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (1)$$

Die obige Formel stellt die Gleichung, zur Bestimmung eines Punktes, auf der Regressionsgeraden, mit den entsprechenden Regressionsparametern da. Das Resultat der Regressionsberechnung sind die entsprechenden β_0 und β_1 für die gilt, dass die Summe aller ε_i , minimal ist. ε_i ist der Abstand der Regressionsgeraden an der stelle x_i zu y_i . Also werden β_0 und β_1 , der Schnittpunkt mit der Y -Achse und die Steigung der Geraden, so gewählt dass gilt, dass die Summe der Abstände zwischen Regressionsgeraden und Datenpunkten in Y -Richtung minimal ist.

Ein möglicher beobachtbarer Zusammenhang bei der Analyse zweier Variablen wäre, dass bei steigendem X auch Y steigt. Ein Beispiel dafür ist das Verhältnis von Körpergröße und Körpergewicht. Im Allgemeinen gilt, je größer eine Person ist desto schwerer ist sie. [10]

6.2 R^2

Mithilfe linearer Regression lässt sich untersuchen ob ein Zusammenhang zwischen verschiedenen Variablen besteht. Ein Mittel um zu beurteilen wie stark dieser Zusammenhang ist, ist R^2 .

$$R^2 := \frac{SQE}{SQT} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

Dabei steht *SQE* für Sums of Squares Explained. Und meint die quadrierte und aufsummierte Abweichung, die die Regressionsgrade an der Stelle x zu dem Mittelwert von y hat. *SQT* steht für Sum of Squares Total und steht für die Summe der quadrierten Differenz von dem beobachteten Wert y_i zu dem Mittelwert \bar{y} . R^2 steht also für das Verhältnis der Abweichung, die die Regression zu dem Mittelwert von y , zu dem Abstand der einzelnen y Werte zu ihrem Mittelwert hat. Oder etwas vereinfacht formuliert, je geringer der Abstand der gemessenen Daten zum Regressionsgraphen ist, desto besser eignet sich die Regression als Modell um die Daten zu erklären. R^2 kann Werte zwischen 1 und 0 annehmen. Die Gleichung ergibt 1, wenn für jeden Punkt \hat{y}_i der Regressionsgeraden gilt, dass der Abstand zum Mittelwert \bar{y} identisch ist mit dem Abstand von y_i zu \bar{y} . Also, wenn gilt: $\sum_{i=1}^n (y_i - \hat{y}_i)^2 = 0$. Ist dies der Fall, so liegen alle Punkte exakt auf der Regressionsgeraden. Ein Wert für R^2 nahe 1 deutet auf einen starken Zusammenhang der Variablen hin. [11]

6.3 Pearson Korrelation

Mithilfe der Pearson Korrelation (PK) misst man die lineare Abhängigkeit zwischen zwei Variablen. Die PK für 2 Variablen X, Y mit der Samplegröße n ist definiert als:

$$r_{xy} := \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3)$$

Für $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ gilt, dass die Größe der Summe abhängig von der gemeinsamen Verteilung von X und Y um ihren jeweiligen Mittelwert ist. Die Summe wird umso größer je ähnlicher sich die beiden Variablen um ihren jeweiligen Mittelwert verteilen. In anderen Worten je größer die positive lineare Abhängigkeit der beiden Variablen ist desto größer wird die Summe.

Wenn für alle X, Y gilt: $(x_i - \bar{x}) = (y_i - \bar{y})$.

Dann gilt auch: $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}$.

In diesem Fall ist $r = 1$. Ein Wert von 1 bedeutet, die beiden Variablen vollständig voneinander abhängig sind und positiv Korrelieren. Je unabhängiger die Verteilung der beiden Variablen ist, desto näher ist der Wert für r an 0. Je näher r an -1 liegt desto größer ist der negative lineare Zusammenhang [12]. Im weiteren Verlauf der Arbeit werden die Ergebnisse der Berechnung der PK in folgender Form angegeben: ($r=0.4946$, $p=0.0257$). Dabei ist r das Resultat der PK und p die Wahrscheinlichkeit, dass die beobachtete Korrelation oder eine stärkere auftritt, unter der Bedingung, dass die Nullhypothese, es existiert keine positive Korrelation, wahr ist.

7 Hypothesen

Die primäre Frage, die mit dieser Arbeit beantworten soll ist, ob es einen Zusammenhang zwischen der Verfügbarkeit von schnellem Internet und der Beteiligung an Wikipediaartikeln zu Abgeordneten des Deutschen Bundestages gibt.

Eine weiter Frage ist, ob die Wahlergebnisse der einzelnen Bundesländer, bei der letzten Bundestagswahl 2017 besser geeignet sind die Beteiligung an Wikipediaartikeln in den jeweiligen Bundesländern zu erklären als der jeweilige Breitbandausbau.

Dies führt zu folgender Nullhypothese H_0 :

Es gibt keine positive Korrelation zwischen externen Faktoren und der Beteiligung an Wikipediaartikeln zu Abgeordneten des Deutschen Bundestages.

Und den jeweiligen Alternativhypothesen:

H1:

Es gibt eine positive Korrelation zwischen dem Breitbandausbau in einem Bundesland und der Beteiligung an Wikipediaartikeln zu Abgeordneten des Deutschen Bundestages.

H2:

Es gibt eine positive Korrelation zwischen den Wahlergebnissen der einzelnen Bundesländer zu Bundestagswahl 2017 und der Beteiligung an Wikipediaartikeln zu Abgeordneten des Deutschen Bundestages.

Die Nullhypothese soll abgelehnt werden, wenn die beobachtete Korrelation, unter der Bedingung, dass die Nullhypothese wahr ist, mit einer Wahrscheinlichkeit von weniger als 5% erreicht wird. Also wenn gilt $p < 0,05$. Dies entspricht einem Signifikanzniveau von $\alpha = 0,05$.

8 Analyse

In diesem Kapitel werden mithilfe der gewonnen Daten die aufgestellten Hypothesen untersucht.

8.1 Breitbandausbau

Um einen ersten Eindruck von der geografischen Verteilung der Artikelbearbeitungen zu bekommen, bietet sich ein Blick auf die Heat Map in Abbildung 4 an. Auf dieser Karte sind die Koordinaten aller aus Deutschland stammenden IP-Adressen eingetragen, die für diese Arbeit gesammelt wurden. Mit Ausnahme von Berlin, dem als Stadtstaat ohnehin eine Sonderrolle zukommt, ist das Ost-West Gefälle auf der Karte gut zu erkennen. In den Stadtstaaten ist trotz einer relativ geringen Anzahl von Daten, die Datendichte durch die hohe Bevölkerungsdichte relativ hoch. In

Mecklenburg-Vorpommern und Sachsen-Anhalt sind nur sehr wenige Editoren zu sehen.

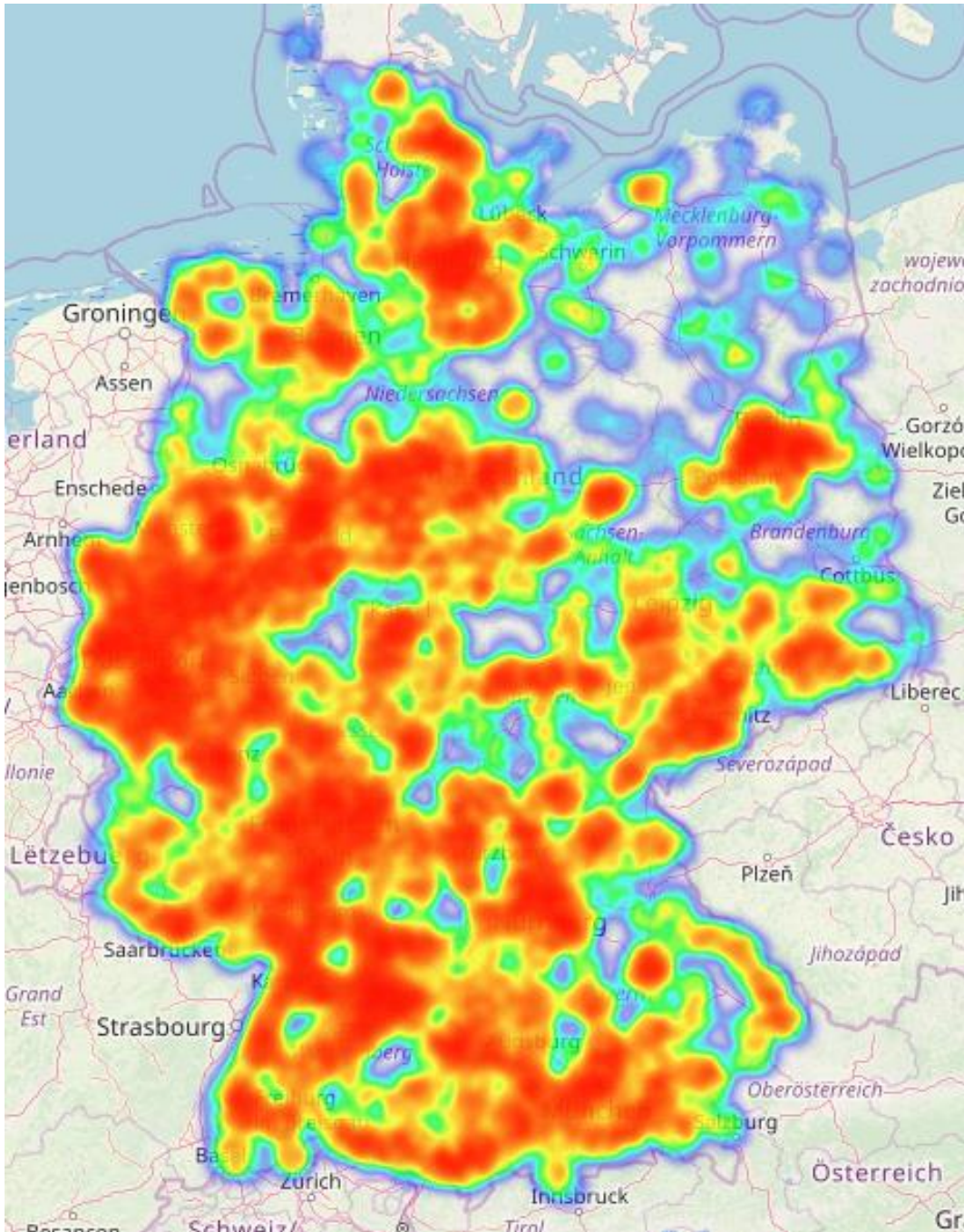


Abb. 4: Heat Map mit allen Editoren

Der Vergleich von Abbildung 4 mit Abbildung 5 zeigt bereits, dass die Verteilung von Artikelbearbeitungen Ähnlichkeiten mit der Verteilung des Breitbandausbaus in Deutschland hat. Auch bei der Verteilung der Breitbandverfügbarkeit ist der Unterschied zwischen West und Ost gut zu erkennen. Auch die hohe Verfügbarkeit von Breitbandanschlüssen in den 3 Stadtstaaten lässt sich erkennen.

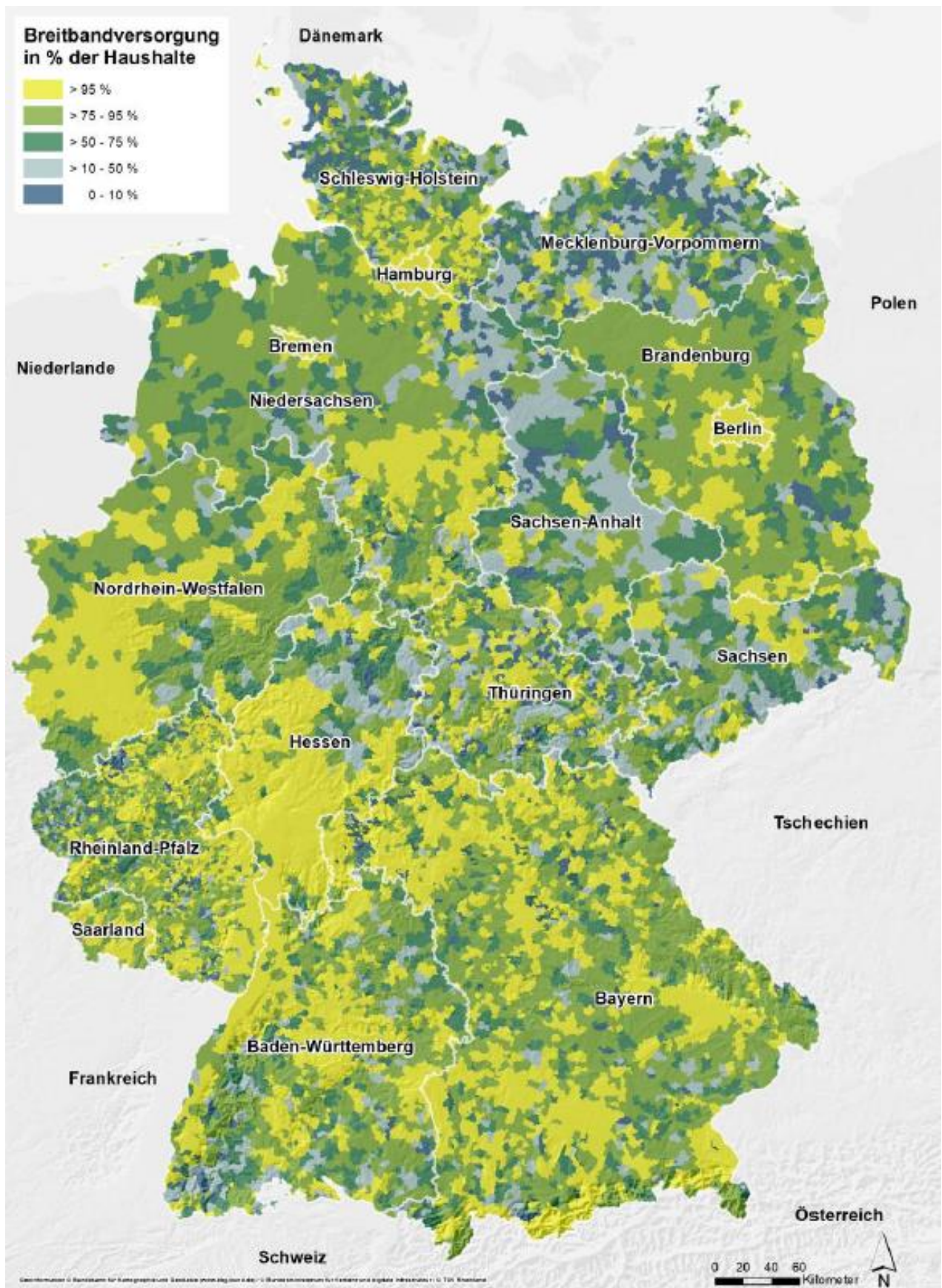


Abb. 5: „Breitbandausbau in Deutschland“ Quelle: https://www.bmvi.de/SharedDocs/DE/Anlage/Digitales/bericht-zum-breitbandatlas-ende-2017-ergebnisse.pdf?__blob=publicationFile

Das Übereinanderlegen beider Karten in Abbildung 6 macht den Zusammenhang besonders deutlich.

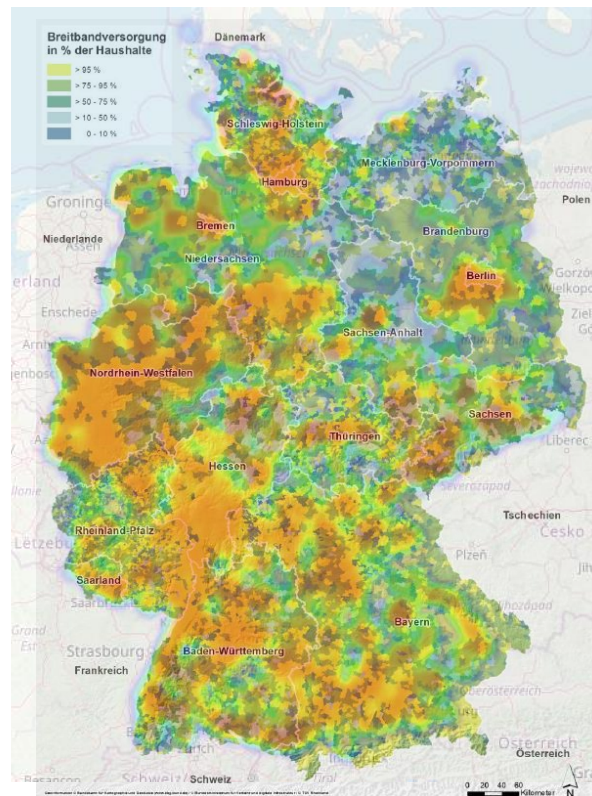


Abb. 6: Kombination Heat Map mit Breitbandausbau

Da sowohl Breitbandausbau [13] als auch die Dichte von Artikelbearbeitungen von der jeweiligen Bevölkerungsdichte abhängig sind, ist dieser starke Zusammenhang zu erwarten. Um zu überprüfen, ob es einen von der Bevölkerungsdichte unabhängigen Zusammenhang zwischen Breitbandausbau und der Anzahl von Artikelbearbeitungen in einem Bundesland gibt, werden im Folgenden die Anzahl der Editoren pro einhunderttausend Einwohner mit dem prozentualen Anteil von Internetanschlüssen, die mindestens 30Mb/s (Breitbandanschluss) schnell sind, verglichen [14]. Abbildung 7 zeigt das Verhältnis von Breitbandausbau und der Anzahl von Editoren. Die einzelnen Datenpunkte entsprechen dabei den Bundesländern. Es lässt sich ein grober Zusammenhang erkennen, aber das Resultat des R^2 -Tests zeigt, dass lediglich ca. 24% der Verteilung der Editoren mit der Breitbandabdeckung erklärt werden können. Der Zusammenhang wird etwas deutlicher, wenn man die Stadtstaaten in den Vergleich nicht mit einbezieht. Da die Stadtstaaten aufgrund ihrer Bevölkerungsdichten geringen Fläche, die 3 Bundesländer mit dem höchsten Anteil an

Breitbandanschlüssen darstellen, kommt ihnen eine gewisse Sonderrolle zu. Abbildung 8 zeigt, dass ohne die Stadtstaaten über 37% der Varianz der Editoren erklärt werden können.

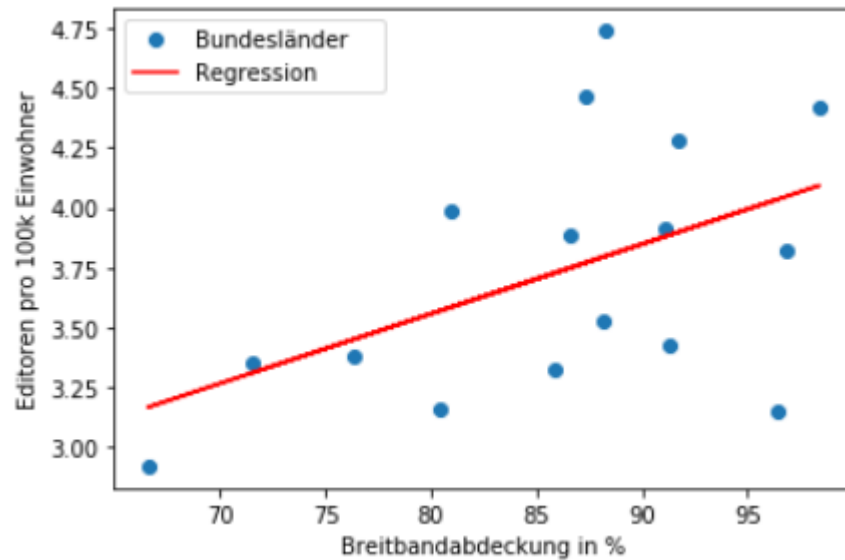


Abb. 7: Vgl. Editoren-Breitbandausbau Regression, mit R-Squared = 0.236548

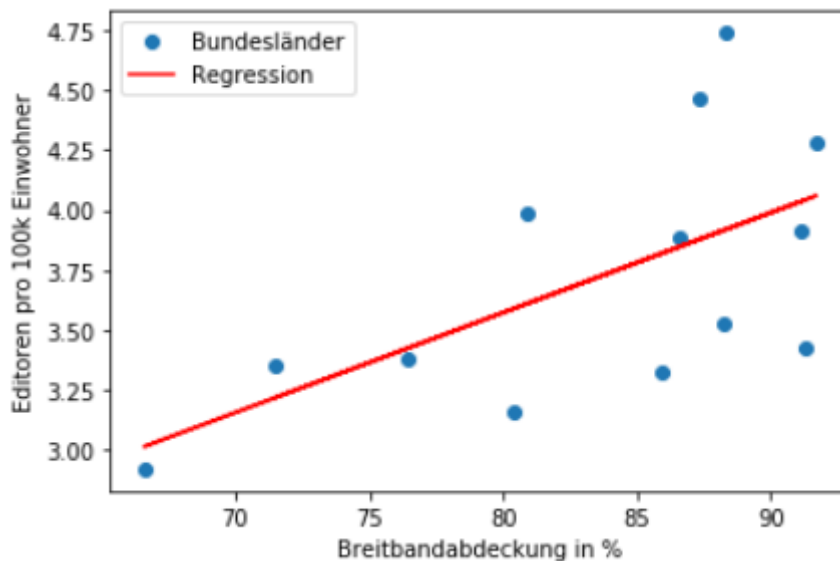


Abb. 8: Vgl. Editoren-Breitbandausbau Regression, ohne Stadtstaaten mit R-Squared = 0.372524

Die PK ergibt für den Vergleich zwischen der Anzahl der Editoren und der Breitbandabdeckung in den einzelnen Bundesländern ($r=0.4863$, $p=0.0280$). Dies bedeutet, dass die Nullhypothese, dass es keinen signifikanten positiven Zusammenhang zwischen der Breitbandabdeckung und der Beteiligung an Wikipediaartikeln gibt, bei einem Signifikanzniveau von $\alpha = 0,05$ abgelehnt werden

kann. Ohne Berücksichtigung der Stadtstaaten liefert die PK mit ($r=0.6103$, $p=0.0134$) eine deutlich höhere Korrelation.

Betrachtet man allerdings alte und neue Bundesländer getrennt voneinander, hebt sich die Korrelation bei den alten Bundesländern völlig auf. Abbildung 9 zeigt, dass die beiden Variablen unabhängig voneinander sind, wenn die neuen Bundesländer nicht mit einbezogen werden. Auch der PK zeigt mit ($r=0.0735$, $p=0.4313$), dass innerhalb der alten Bundesländer kein Zusammenhang zwischen den beiden Variablen besteht.

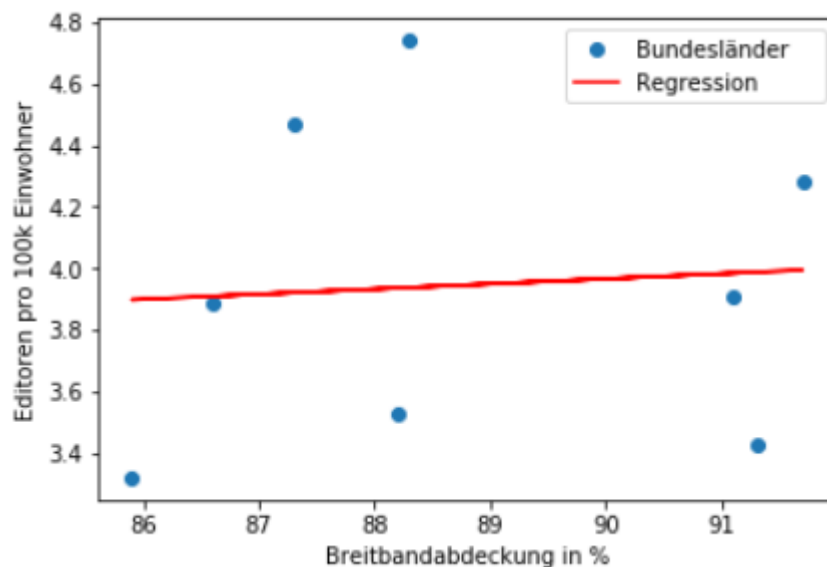


Abb. 9: Vgl. Editoren-Breitbandausbau ohne neue Bundesländer mit R-squared = 0.005408

In Abbildung 10 werden die Daten der neuen Bundesländer miteinander verglichen. Hier ist zwar ein Zusammenhang zwischen den beiden Variablen zu erkennen, aber die PK zeigt mit ($r=0.6604$, $p=0.1125$) keine hohe Signifikanz. Da es keine signifikante Abhängigkeit gibt, wenn man alte und neue Bundesländer getrennt voneinander betrachtet, lässt sich Schlussfolgern, dass der in Abbildung 8 beobachtete Zusammenhang primär durch die Unterschiede zwischen den beiden Gruppen, alte und neue Bundesländer, besteht.

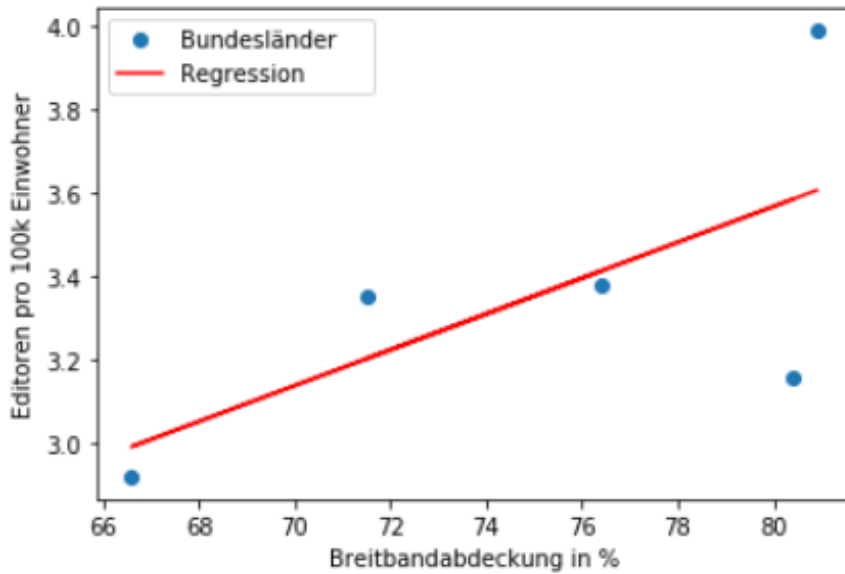


Abb. 10: Vgl. Editoren-Breitbandausbau neue Bundesländer R-squared = 0.436248

Abbildung 11 zeigt den deutlichen Unterschied, der zwischen den neuen und alten Bundesländern, insbesondere hinsichtlich des Breitbandausbaus existiert. Kein westliches Bundesland hat einen Anteil von Breitbandanschlüssen der unter 86% (im Mittel 88.8%) liegt. Im Osten hingegen, hat kein Land einen Anteil von über 80% (im Mittel 75.16%).

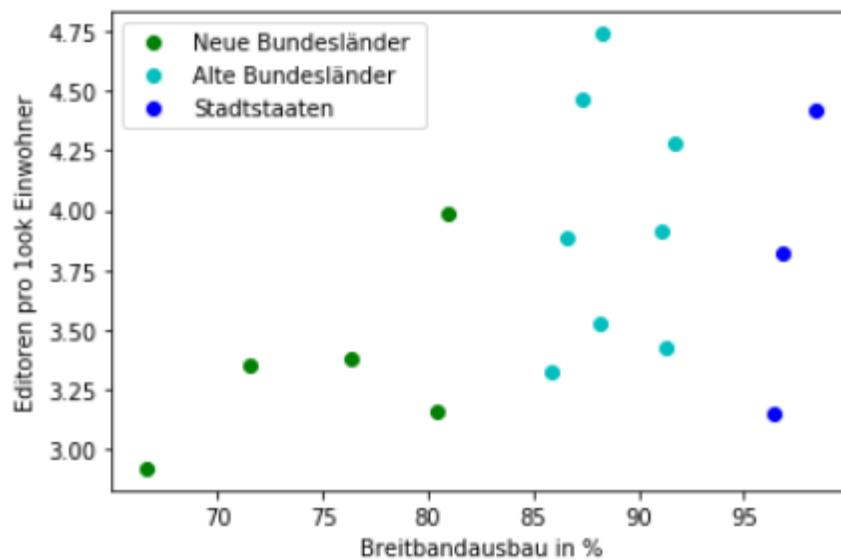


Abb. 11: Trennung nach neuen Ländern, alten Ländern und Stadtstaaten.

8.2 Wahlergebnis

Der Vergleich zwischen dem Wahlergebnis und dem prozentualen Anteil von Editoren jeder Partei in Abbildung 12 zeigt, ein Ungleichgewicht zugunsten der Editoren, sowohl bei der Partei Die Linke als auch bei den Grünen. Artikel zu Politikern dieser Parteien haben folglich mehr Artikelbearbeitungen, als es bei Anzahl der Politiker der jeweiligen Parteien zu erwarten wäre.

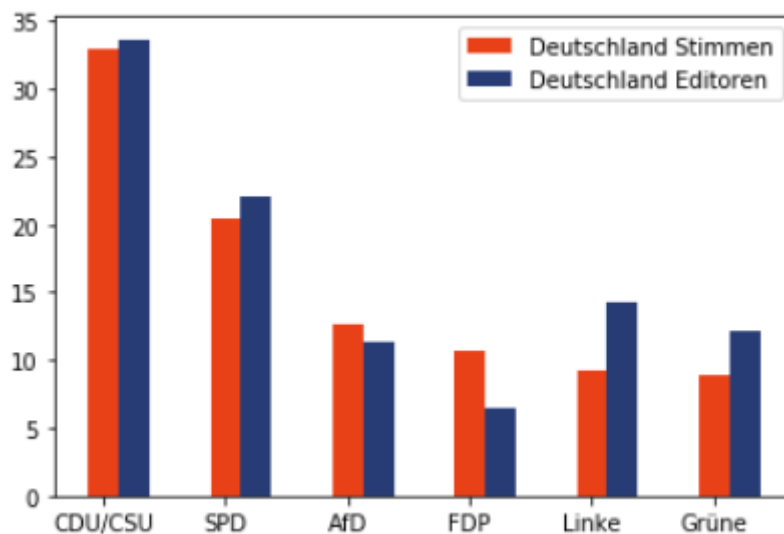


Abb. 12: Anteil der Stimmen und Editoren in %. (Aufgrund nicht in den Bundestag gelangter Parteien, summiert sich das Wahlergebnis nicht auf 100%)

Da Politiker von Linken und Grünen überproportional häufig bearbeitet werden, soll im Folgenden untersucht werden, ob sich anhand der Bundestagswahlergebnisse dieser Parteien auf Landesebene, Rückschlüsse auf die Anzahl der Editoren ziehen lassen.

Abbildung 13 zeigt den Vergleich der Editorenanzahl mit dem Wahlergebnis der Linken. Die Regression suggeriert zwar eine negative lineare Abhängigkeit, diese besteht jedoch nur aufgrund der recht klar in alte und neue Länder trennbaren Wahlergebnisse. Insbesondere in den alten Ländern lässt sich keine Abhängigkeit erkennen.

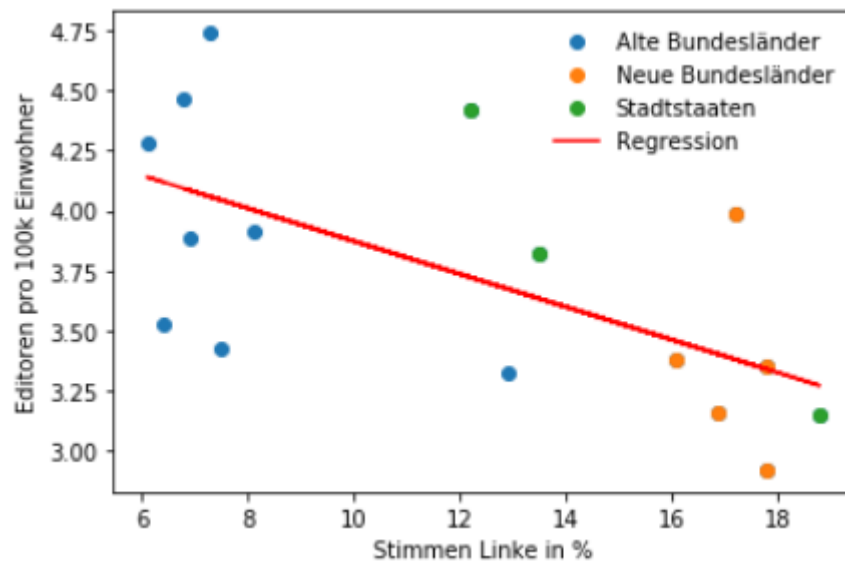


Abb. 13: Vgl. Stimmen Linke-Edits mit R-squared: 0.387410

Dem gegenüber steht der positive Zusammenhang, in Abbildung 14. Hier werden die Artikelbearbeitungen mit dem Wahlergebnis der Grünen verglichen.

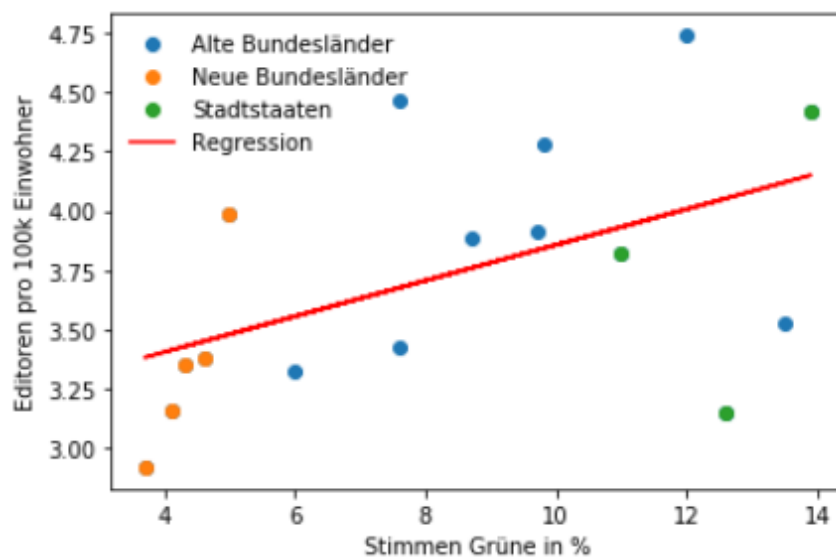


Abb. 14: Vgl. Stimmen Grüne-Edits mit R-squared: 0.244662

Das Wahlergebnis der Grünen ist mit PK ($r=0.4946$, $p=0.0257$) geeignet die Nullhypothese abzulehnen. Insbesondere wenn man die Sonderrolle der Grünen in Baden-Württemberg berücksichtigt. Mit Winfried Kretschmann stellen sie seit 2011 den ersten grünen Ministerpräsidenten. Bei der letzten Landtagswahl 2016 waren die Grünen die stärkste Partei im Land. Mit 30,3% der Stimmen hatten sie in Baden-Württemberg über 10% mehr Stimmen als bei jeder anderen Landtagswahl. Wenn

man Baden-Württemberg aufgrund dieser besonderen Stellung der Grünen für die Berechnung nicht berücksichtigt, ergibt die PK ($r=0.5815$, $p=0.0115$). Es besteht ohne Baden-Württemberg also eine hohe Korrelation zwischen dem Wahlergebnis und der Anzahl der Editoren. Wenn man zusätzlich die Stadtstaaten nicht betrachtet können, wie in Abbildung 15 zu sehen, über 66% der Varianz erklärt werden. Die PK berechnet mit ($r=0.8134$, $p=0.0006$) eine sehr hohe Korrelation.

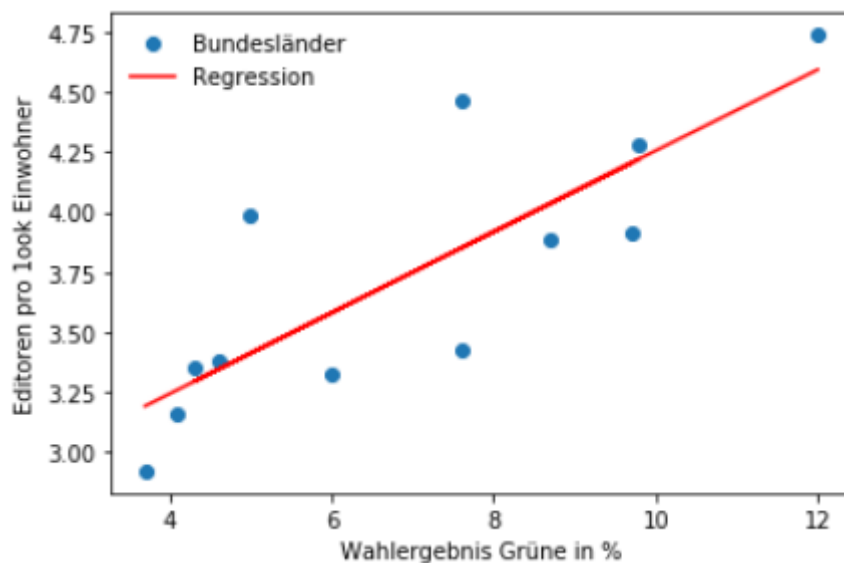


Abb. 15: Vgl. Wahlergebnis Grüne – Edits ohne BW und Stadtstaaten mit R-squared: 0.661623

Auch wenn man alte und neue Bundesländer getrennt voneinander betrachtet, bleibt der Zusammenhang bestehen. Abbildung 16 zeigt, den starken linearen Zusammenhang des Wahlergebnisses der Grünen mit der Anzahl an Artikelbearbeitungen in den neuen Bundesländern. Über 91% der beobachteten Varianz der Artikelbearbeitung lassen sich mit dem Wahlergebnis erklären. Die PK zeigt mit ($r=0.9589$, $p=0.0050$) für die neuen Länder eine sehr hohe Korrelation.

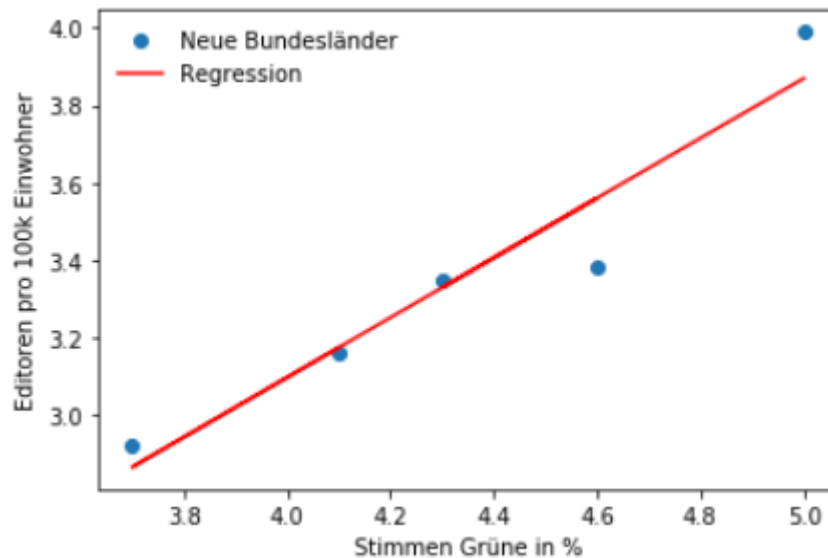


Abb. 16: Vgl. Wahlergebnis grüne-Edits in den neuen Bundesländern mit R-squared: 0.919569

In Abbildung 17 ist zu sehen, dass es auch bei den alten Bundesländern eine positive Korrelation gibt, wenn man der obigen Argumentation folgend, Baden-Württemberg nicht mit einbezieht und auch die Stadtstaaten aufgrund ihrer Sonderstellung nicht mitberücksichtigt. 56% der Varianz lassen sich mit diesem Modell erklären und die PK zeigt mit ($r=0.7410$, $p=0.0261$) eine hohe Korrelation.

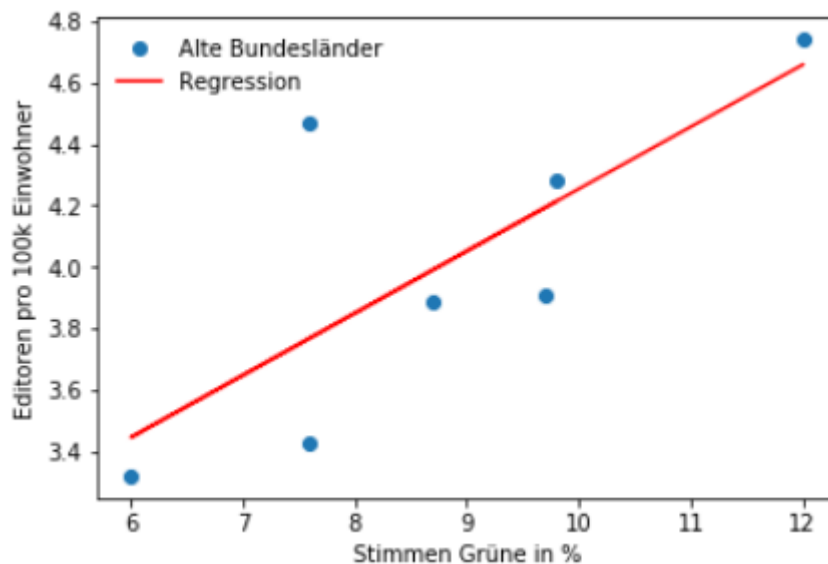


Abb. 17: Vgl. Wahlergebnis grüne-Edits in den alten Bundesländern ohne BW und Stadtstaaten mit R-squared: 0.562450

Der Zusammenhang zwischen dem Wahlergebnis der Grünen und der Anzahl von Artikelbearbeitungen besteht, anders als beim Breitbandausbau, nicht nur zwischen alten und neuen Bundesländern, sondern auch jeweils innerhalb dieser beiden Gruppen.

9 Diskussion

Die ursprüngliche Annahme war, je weiter der Breitbandausbau eines Bundeslandes vorangeschritten ist, desto ausgeprägter ist die Beteiligung der Bevölkerung an der Erstellung von Wikipediaartikeln. Ausschlaggebend für diese Hypothese war unter anderem der Zusammenhang von schnellem Internet und politischer Partizipation [3, 4]. Es konnte zwar eine Korrelation zwischen Breitbandausbau und der Anzahl der Editoren festgestellt werden, aber es hat sich gezeigt, dass sie primär in den Unterschieden zwischen alten und neuen Bundesländern zu finden ist. Insbesondere innerhalb der alten Bundesländer hat sich keine Korrelation zwischen den beiden Variablen feststellen lassen. Die eingangs formulierte Hypothese, dass ein flächendeckenderer Breitbandausbau einen positiven Effekt auf die politische Partizipation und somit auf den demokratischen Prozess hat, konnte mit dieser Arbeit nicht belegt werden. Dies steht im Kontrast zu den Resultaten vorausgegangener Arbeiten. Allerdings ist es diskutabel, ob die Erstellung von Wikipediaartikeln zu Politikern tatsächlich aus politischer Motivation geschieht oder ob nicht andere Gründe ausschlaggebender sind. Dies zu überprüfen war mit den zur Verfügung stehenden Daten allerdings nicht möglich. Es ist auch möglich, dass eine Betrachtung auf Länderebene der heterogenen Verteilung von Breitbandanschlüssen nicht gerecht wird. Um dies zu überprüfen, müssten die Daten von einer weit größeren Anzahl an Wikipediaartikeln gesammelt werden, damit eine feingliedrigere Untersuchung möglich wird. Generell wäre eine größere Datenmenge empfehlenswert, denn lediglich 8 Artikelbearbeitungen mehr hätten Bremen, in dem es nur 26 Editoren gab, von einer mittleren Position der Artikelbearbeitungen pro Einwohner, auf den vordersten Platz gebracht. Diese Arbeit sollte jedoch zum einen auf aktive Politiker des aktuellen Deutschen Bundestages beschränkt werden und zum anderen haben die genutzten kostenfreien Services zur Bestimmung der Ortszugehörigkeit von Längen- und Breitengraden aufgrund zahlreicher Verbindungsprobleme, viel manuellen Aufwand erfordert.

Es hat sich gezeigt, dass das Wahlergebnis der Partei die Grünen weit besser geeignet ist, die Anzahl der Editoren zu erklären als der Breitbandausbau. Insbesondere in den neuen Bundesländern besteht eine sehr hohe Abhängigkeit der beiden Variablen zueinander. Ein Teil dieses Ergebnis lässt sich mit generellen Unterschieden im

Wahlverhalten zwischen alten und neuen Bundesländern erklären. In den neuen Bundesländern wird seit der Wiedervereinigung anders gewählt als in den alten. Die Grünen bekamen mit 8,7% in den neuen Bundesländern deutlich mehr Stimmen als in den alten mit 4,6%. Da die Korrelation aber auch innerhalb alter und neuer Bundesländer beobachtet werden konnte muss es eine weitere Ursache geben. Von allen Parteien haben die Grünen den größten prozentualen Anteil an unter 35-jährigen Wählern [15]. Da über 50% der Wikipediaautoren unter 30 Jahre alt sind [16], sollten sich unter den Editoren vergleichsweise viele Grünenwähler befinden. Nach einer Erhebung aus dem Jahr 2005 hatten die Grünen von allen Parteien den größten Anteil aktiver Internetnutzer [17]. Diese Daten würden zu dem beobachteten Ungleichgewicht von Stimmen und der Anzahl an Editoren in Abbildung 12 passen. Allerdings gibt es dieses Ungleichgewicht auch bei der Partei Die Linke. Zwischen Linke und Editorenanzahl gibt es aber aufgrund der hohen Differenz der Stimmanteile in alten und neuen Ländern eine negative Abhängigkeit. Es wäre denkbar, dass die Wertvorstellungen und Lebensbedingungen, die eine Person motivieren Wikipediaartikel zu verfassen, denen ähneln, die eine Person dazu bewegen die Grünen zu wählen. Dass der Anteil an Artikelbearbeitungen zu Politikern der Linken überproportional groß ist, sich das Wahlergebnis aber nicht eignet die Bearbeitungen anderer Parteien zu erklären, könnte darauf hindeuten, dass es unter Wählern der Linken zwar eine erhöhte Bereitschaft gibt die Repräsentanz von Abgeordneten der Linken zu erhöhen, nicht aber die von Politikern anderer Parteien. Um zu überprüfen in wieweit sich die überproportionale Bearbeitung von Politikern der Linken und der Grünen auch in anderen Artikeln findet, könnte eine ähnliche Methodik auf eine größere Anzahl von Politikern angewendet werden. Sollte sich beispielsweise ergeben, dass sich dieses Verhältnis auch auf kommunaler Ebene zeigt, könnte man versuchen Methoden zu entwickeln, die diesem Ungleichgewicht entgegensteuern.

Quellen:

- [1] „Top Sites in Germany,“ 2019. [Online]. Available: <https://www.alex.com/topsites/countries/DE>. [Zugriff am 15 Mai 2019].
- [2] A. Jungherr und H. Schoen, „Das Internet in Wahlkämpfen,“ Bamberg, Springer VS, Wiesbaden, 2013, pp. 43-47.
- [3] M. Graham, R. K. Straumann und B. Hogan, „Digital Divisions of Labor and Informational Magnetism: Mapping Participation in Wikipedia,“ *Annals of the Association of American Geographers*. 105, pp. 1158-1178, September 2015.
- [4] D. E. Sylvester und A. J. McGlynn, „The Digital Divide, Political Participation, and Place,“ *Social Science Computer Review* 28, no. 1, pp. 64-74, Februar 2010.
- [5] J. Andreas, J. Pascal und S. Harald, „Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan, A., Sprenger, TO, Sander, PG, & Welp, IM “Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment”,“ *Social Science Computer Review*, 30(2), p. 229–234, 25 April 2011.
- [6] „Wikipedia,“ Wikipedia, 09 August 2019. [Online]. Available: <https://de.wikipedia.org/wiki/Wikipedia>. [Zugriff am 21 August 2019].
- [7] „Number Resources,“ Iana, [Online]. Available: <https://www.iana.org/numbers>. [Zugriff am 5 August 2018].
- [8] „Geolite2,“ MaxMind, [Online]. Available: <https://dev.maxmind.com/geoip/geoip2/geolite2/>. [Zugriff am 18 Juli 2019].
- [9] „GeoIP2 City Accuracy,“ MaxMind, [Online]. Available: <https://www.maxmind.com/en/geoip2-city-accuracy-comparison?country=Germany&resolution=50>. [Zugriff am 18 Juli 2019].

- [10] „Lineare Regression,“ Wikipedia, 2019. [Online]. Available:
https://de.wikipedia.org/wiki/Lineare_Regression. [Zugriff am 20 Oktober 2019].
- [11] „Bestimmtheitsmaß,“ Wikipedia, 2019. [Online]. Available:
<https://de.wikipedia.org/wiki/Bestimmtheitsma%C3%9F>. [Zugriff am 20 Oktober 2019].
- [12] „Korrelationskoeffizient,“ Wikipedia, 2019. [Online]. Available:
<https://de.wikipedia.org/wiki/Korrelationskoeffizient>. [Zugriff am 20 Oktober 2019].
- [13] B. Williger und A. Wojtech, *Digitalisierung im ländlichen Raum, Status Quo & Chancen für Gemeinden*, Frauenhofer SCS, 2018, pp. 7-8.
- [14] „Aktuelle Breitbandverfügbarkeit in Deutschland (Stand Mitte 2018),“
Bundesministerium für Verkehr und digitale Infrastruktur (BMVI), Berlin, 2018.
- [15] K. Kobold und S. Schmiedel, „Wahlverhalten bei der Bundestagswahl 2017 nach Geschlecht und Alter,“ Statistisches Bundesamt, Wiesbaden, 2018.
- [16] M. Pande, „Wikipedia editors do it for fun: First results of our 2011 editor survey,“ 10 Juni 2011. [Online]. Available:
<https://blog.wikimedia.org/2011/06/10/wikipedia-editors-do-it-for-fun-first-results-of-our-2011-editor-survey/>. [Zugriff am 5 November 2019].
- [17] N. Czernich, „Broadband Internet and Political Participation: Evidence for Germany,“ in *Kyklos*, 65: 31-52. doi:10.1111/j.1467-6435.2011.00526.x, 2012, p. 21.