



## disP - The Planning Review

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/rdsp20>

### Agglomerationen mit nutzergenerierten Inhalten neu definiert Visualisierung der Nordostschweiz mithilfe von Wikipedia

André Bruggmann<sup>a</sup>, Marco M. Salvini<sup>b</sup> & Sara I. Fabrikant<sup>c</sup>

<sup>a</sup> Doktorand Geographische Informationsvisualisierung & -analyse Geographisches Institut Universität Zürich - Irchel Winterthurerstrasse 190 CH-8057 Zürich

<sup>b</sup> Projektleiter Stadtentwicklung Stadtbauamt Aarau Rathausgasse 1 CH-5000 Aarau

<sup>c</sup> Professorin für Geographische Informationsvisualisierung & -analyse Geographisches Institut Universität Zürich - Irchel Winterthurerstrasse 190 CH-8057 Zürich

Published online: 12 Mar 2014.

To cite this article: André Bruggmann, Marco M. Salvini & Sara I. Fabrikant (2013) Agglomerationen mit nutzergenerierten Inhalten neu definiert Visualisierung der Nordostschweiz mithilfe von Wikipedia, *disP - The Planning Review*, 49:4, 37-45, DOI: [10.1080/02513625.2013.892789](https://doi.org/10.1080/02513625.2013.892789)

To link to this article: <http://dx.doi.org/10.1080/02513625.2013.892789>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

# Agglomerationen mit nutzergenerierten Inhalten neu definiert

*Visualisierung der Nordostschweiz mithilfe von Wikipedia*

André Bruggmann, Marco M. Salvini und Sara I. Fabrikant

**Abstract:** Commuters have had an important role in shaping the spatial organization of Switzerland, as commuter flows have been one of the most significant criteria to delineate urban agglomeration zones. Even though urban areas and respective agglomerations have continuously gained in importance in Switzerland to this day, the Swiss national population census will no longer include commuter data at high spatial resolution. Hence, the definition of the rapidly evolving urban agglomeration concept will have to be modified for future urban research and planning purposes.

We propose a crowdsourcing approach to overcome this data gap, and employ the open and web-based Wikipedia encyclopedia as a new resource to delineate agglomeration areas. Using the North Eastern parts of Switzerland in this case study, we systematically evaluate whether user-generated content can serve as an option to fill the commuter data gap in future Swiss national population censuses to define agglomeration areas. In a second step, we evaluate the influence of potential edge effects on our chosen approach.

We employ the number of hyperlinks in the Wikipedia database to quantify the strength of functional relationships between municipalities appearing in the Wikipedia encyclopedia. Next, we visualize the extracted municipality network structure for the chosen study area. Finally, we cluster the connected municipalities to agglomeration zones, and compare the computed municipality clusters with the agglomeration areas currently defined by the Swiss census.

Our results suggest that the aggregation structure of our crowdsourcing approach is congruent with the officially developed agglomeration areas proposed by the Swiss census. Crowdsourced data thus might be an additional future data resource to complement more traditional census statistics for space districting purposes or socio-economic research in urban geography and planning.

However, our results also suggest that geographic space indeed influences even non-spatially organized, crowdsourced encyclopedic

entries, and this must be systematically studied further in future studies.

**English Title:** Agglomerations newly defined with crowdsourced data: visualizing North Eastern Switzerland based on Wikipedia content.

## 1. Einleitung

Die Agglomerationsdefinition der Schweiz und damit die Einteilung der Schweizer Gemeinden in Agglomerations- und Nichtagglomerationsgemeinden wurde in den letzten Jahren vermehrt in raumplanerischen und politischen Debatten thematisiert (Bundesamt für Statistik 2007: 1). Da mit der Umstellung der schweizerischen Volkszählung von einer Vollerhebung zu einer Auswertung amtlicher Registerdaten mit ergänzenden stichprobenbasierten Erhebungen ab dem Jahre 2010 die Pendlerverflechtungen, welche das wichtigste Kriterium in der bisherigen Agglomerationsdefinition darstellten, nicht mehr in der gleichen räumlichen Auflösung erhoben werden können, hat der Bund im Jahre 2006 das Projekt *Agglosuisse* lanciert (Bundesamt für Statistik 2011: 2–5). Das *Agglosuisse*-Projekt setzt sich als Ziel, die Agglomerationsdefinition neu zu konzipieren, und hat eine Laufzeit bis zum Jahre 2014 (Bundesamt für Statistik 2007: 1–3).

Die vorliegende Arbeit folgt der gleichen Zielsetzung wie das Projekt *Agglosuisse* und testet das Potential einer zur Volkszählung alternativen Datengrundlage. Dabei besteht nicht der Anspruch, einen Ansatz zu präsentieren, der genau in dieser Form umzusetzen ist, sondern vielmehr soll ein Beitrag zum wissenschaftlichen Diskurs geleistet werden, indem eine aktuelle wirtschaftsgeographische Problemstellung mit visuell-analytischen Methoden angegangen wird. Als Datengrundlage dieser Arbeit wird die im Laufe des letzten Jahrzehnts von NutzerInnen generierte, mittlerweile sehr umfangreiche und kostenlos auf dem Internet zugängliche Enzyklopädie *Wikipedia* eingesetzt (Wikipedia 2012). Die räumlich-zeitlichen Analysen mit Wikipedia von Hecht (2007)

*André Bruggmann* ist Doktorand in der Abteilung Geographische Informationsvisualisierung & -analyse am Geographischen Institut der Universität Zürich.

*Marco M. Salvini* hat in der Abteilung Geographische Informationsvisualisierung & -analyse am Geographischen Institut der Universität Zürich doktriert und arbeitet jetzt als Projektleiter Stadtentwicklung beim Stadtbauamt Aarau.

*Sara I. Fabrikant* ist Professorin am Geographischen Institut der Universität Zürich und leitet die Abteilung Geographische Informationsvisualisierung & -analyse.

untermauern die Eignung von Wikipedia als geographische Datengrundlage und dienen als Ausgangspunkt für die vorliegende Untersuchung der Agglomerationsdefinition mit diesem nutzergenerierten Datensatz.

Auf die zwei folgenden Fragestellungen wird dabei vertieft eingegangen.

- Eignen sich die auf Wikipedia verfügbaren nutzergenerierten Inhalte, um mit modernen Methoden und Mitteln der Informationsvisualisierung und -analyse Agglomerationen zu identifizieren?
- Welche Probleme treten in Verbindung mit den Raum- und Randeffekten in den Daten auf und welche Gebiete sind hauptsächlich betroffen davon?

## 2. Theoretischer Hintergrund

Die bisherige Agglomerationsdefinition stützt sich auf Minimalkriterien. Sind diese erfüllt, werden Kernzonen und umliegende Gemeinden zu Agglomerationen zusammengefasst. Kernzonen müssen dabei einen genügend hohen Anteil an Arbeitsplätzen pro Erwerbstätige zur Verfügung stellen. Weitere Gemeinden, die zu einer Agglomeration gehören, müssen einerseits 1/6 ihrer Erwerbstätigen in die Kernzone für die Arbeit entsenden. Andererseits sind für diese Gemeinden weitere Kriterien bezüglich des baulichen Zusammenhangs, der Einwohner- und Arbeitsplatzdichte, des Bevölkerungswachstums und der Ausprägung von Wirtschaftssektoren zu erfüllen (Schuler et al. 2005: 148, 149).

Der Bund schlägt für die neue Agglomerationsdefinition funktionale und morphologische Kriterien vor. Einerseits soll beispielsweise ein rasterbasiertes Dichtekriterium in Bezug auf die Anzahl Beschäftigte und EinwohnerInnen für die Definition von Kernzonen zur Anwendung kommen. Andererseits sollen Agglomerationen je nach Verflechtungsintensität und Bedeutung in Haupt- und Nebenkerne unterschieden werden. Viele Fragen bleiben jedoch bei der Festlegung des Agglomerationsperimeters ausserhalb der Kernzonen offen. Auch ist zum heutigen Zeitpunkt noch nicht klar, ob die Pendlerbewegungen adäquat durch die neuen Kriterien ersetzt werden können und wie funktionale Verflechtungen, wie von Thierstein et al. (2006) gefordert, besser miteinbezogen werden können (Bundesamt für Statistik 2012).

Analyse und Visualisierung von Wikipedia-Daten stellen in diesem Kontext eine Alternative dar, um diese Problemstellung anzugehen.

Wikipedia zeichnet sich als Datenquelle vor allem durch seine Aktualität aus. Neue Entwicklungen und Ereignisse werden sofort dokumentiert. Ausserdem ist es mit einer Plattform wie Wikipedia möglich, eine breite Autorenschaft von Interessierten und FachexpertInnen zu erreichen. Die freie Zugänglichkeit Wikipedias führt hingegen auch dazu, dass die Objektivität der Artikel nicht gewährleistet ist, was zu Verzerrungen in Untersuchungsergebnissen führen kann. Die Objektivität ist jedoch auch bei der Wahl vergleichbarer Datensätze sowie bei der Anwendung der bisherigen Agglomerationsdefinition diskussionswürdig. Aus diesem Grund sind bei der Arbeit mit Wikipedia und vergleichbaren Datenquellen ausführliche Evaluierungen, wie in Kapitel 5 dieses Artikels vorgestellt, von grosser Wichtigkeit.

Das folgende Kapitel legt die dafür entwickelte Methodik dar.

## 3. Methodik

Das exemplarische Untersuchungsgebiet dieser Arbeit umfasst die Nordostschweiz. Dieses Gebiet ist repräsentativ für die Agglomerationsthematik der gesamten Schweiz, da Agglomerationen verschiedener Grössen sowie grenz- und kantonsübergreifende Agglomerationen darin vorkommen. Als Untersuchungsobjekte werden, wie in der offiziellen Definition des Bundesamtes für Statistik, nur Gemeinden berücksichtigt. Zum Untersuchungsgebiet zählen folglich alle Gemeinden der Kantone Appenzell Ausserrhoden, Appenzell Innerrhoden, Schaffhausen, St. Gallen und Thurgau. Ausserdem werden angrenzende ausländische Gemeinden sowie Gemeinden des Kantons Zürich berücksichtigt, welche gemäss Agglomerationsdefinition aus dem Jahre 2000 Teil einer grenz- oder kantonsüberschreitenden Agglomeration der Nordostschweiz sind. Insgesamt werden 247 Gemeinden berücksichtigt. Diese Gemeinden bilden im Untersuchungsgebiet gemäss offizieller Agglomerationsdefinition zehn unterschiedlich grosse Agglomerationen.

In dieser Arbeit wird auf die deutschsprachige Wikipedia zurückgegriffen, da diese im Vergleich zu anderen Sprachversionen die ausführlichsten Artikel und die meisten Hyperlinks zu Gemeinden im ausgewählten Untersuchungsgebiet enthält. Inspirationsquelle für die Vorgehensweise ist die Dissertation von Salvini (2012), welche sich mit der Visualisierung des globalen Städtetetzes mithilfe von Wikipedia-Daten beschäftigt.

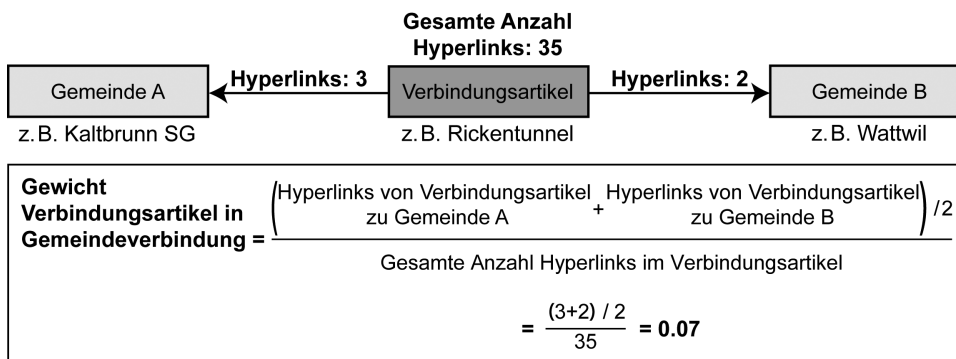


Abb. 1: Quantifizierung der Gemeindeverbindungen.

Die Methodik zur Quantifizierung der Gemeindeverbindungen, welche eine Modifikation der Formel von Hecht und Raubal (2008) darstellt, kann in Abbildung 1 nachvollzogen werden.

Gemeinden (z.B. Kaltbrunn SG und Wattwil) und Verbindungsartikel (z.B. Rickentunnel) bilden das Kernelement des angewandten Ansatzes. Zu einem Verbindungsartikel gehören alle Artikel, welche gleichzeitig mindestens einen Hyperlink zur Gemeinde A und zur Gemeinde B im Untersuchungsgebiet haben. Je höher die Anzahl der Verbindungsartikel zwischen zwei Gemeinden ist, desto stärker wird die Beziehung zwischen diesen zwei Gemeinden gewichtet.

Um das Gewicht eines Verbindungsartikels für eine Gemeindeverbindung zu bestimmen, werden einerseits die Anzahl Hyperlinks zu den Gemeinden A und B und andererseits die gesamte Anzahl Hyperlinks eines Verbindungsartikels verwendet. Je mehr Hyperlinks ein Verbindungsartikel zu den Gemeinden A und B und je weniger Hyperlinks der Verbindungsartikel insgesamt hat, desto stärker ist sein Gewicht in der entsprechenden Gemeindeverbindung. Abbildung 1 zeigt die komplette Formel. Das Gewicht des Verbindungsartikels *Rickentunnel* in der Gemeindeverbindung beträgt folglich 0.07. Mit dieser Methodik werden die funktionalen Verflechtungen zwischen den Agglomerationsgemeinden im Untersuchungsgebiet nachgezeichnet und gleichzeitig wird überprüft, ob die Gemeinden, welche geographisch nahe beieinander liegen und starke funktionale Beziehungen haben, auch auf Wikipedia stark miteinander verknüpft sind.

Die resultierende Matrix, welche quantitative Werte für Beziehungen zwischen den Gemeinden enthält, wird zu einer Netzwerkvisualisierung transformiert, welche die strukturell

wichtigsten Gemeindeverbindungen im Untersuchungsgebiet in Form von Verbindungslinien zwischen den Gemeinden zeigt und die Gemeinden gemäss der Distanz-Ähnlichkeitsmetapher zueinander darstellt (Fabrikant et al. 2004). Dies bedeutet, dass sich Gemeinden im Netzwerk nahe zueinander befinden, wenn sie starke Beziehungen zueinander aufweisen. Diese Darstellungsform setzt methodische Ansätze der Informationsvisualisierung und -analyse um und hilft BetrachterInnen, Strukturen in den Daten explorativ zu erkunden.

In einem letzten Schritt werden die Gemeinden gemäss Blondel-Community-Algorithmus gruppiert (Blondel et al. 2008). Dieser Algorithmus fasst Gemeinden zusammen, die untereinander viele und zu anderen Gemeinden wenige Verbindungen gemäss Abbildung 1 aufweisen.

Um die erste Fragestellung zu beantworten, werden die erhaltenen Gemeindegruppen den Agglomerationen gemäss Agglomerationsdefinition aus dem Jahre 2000 gegenübergestellt. Als statistisches Mass, um die Güte dieser Übereinstimmung zu testen, wird der hypergeometrische Test verwendet (Kos, Psenicka 2000). Dieser verwendet die hypergeometrische Verteilung als Grundlage, um zu überprüfen, ob die Gruppierung von Elementen einer Referenzeinteilung folgt. Als Referenzeinteilung gilt die offizielle Einteilung der Gemeinden in Agglomerationen gemäss Bundesamt für Statistik (2013). Kann die Nullhypothese dieses Tests abgelehnt werden, stimmen die Agglomerationen gemäss offizieller Definition mit den in dieser Arbeit erhaltenen Blondel-Gruppen gut überein. Die Eignung der angewandten Methode, um Agglomerationen nachzubilden, wird somit bestätigt.

Um die zweite Fragestellung zu beantworten, welche den Einfluss des Raumes auf die erhaltenen Resultate beurteilt, wird das Netzwerk mit quantitativen statistischen Methoden

analysiert. Dazu wird einerseits eine Korrelationsanalyse berechnet, welche die Stärke und die Richtung des Zusammenhangs zwischen der Einwohnerzahl und der Zentralität der untersuchten Gemeinden aus gibt (Methodenberatung 2012). Als Mass für die Zentralität der Gemeinden wird der gewichtete Grad benutzt, welcher pro Gemeinde alle Gewichte der Verbindungen zu allen anderen Gemeinden aufsummiert (Wasserman, Faust 1994: 100). Um andererseits räumliche Einflüsse vertieft zu untersuchen und Autokorrelationsmuster in den Daten zu erkennen, wird eine geographisch gewichtete Regression mit der unabhängigen Variable *Einwohnerzahl* und der abhängigen Variable *gewichteter Grad* berechnet (Fotheringham et al. 2002, 2012). Mit diesem Ansatz wird im Vergleich zur globalen linearen Regression für jede einzelne Gemeinde eine eigene Regressionsgleichung ausgegeben. Die Berechnungen der geographisch gewichteten Regression basieren jedoch im Vergleich zu einer globalen linearen Regression nicht auf allen, sondern nur auf den Werten der Gemeinden im näheren Umkreis der betrachteten Gemeinde. Dabei werden Gemeinden, welche in der Nähe einer betrachteten Gemeinde liegen, stärker gewichtet. Als Resultat der Berechnung der geographisch gewichteten Regression werden der erwartete Wert der abhängigen Variablen und die Güte der Anpassung der Regressionsparameter an die tatsächlichen Werte pro Gemeinde ausgegeben. Anhand der räumlichen Verteilung dieser Werte kann beurteilt werden, ob ein Raumeinfluss im Untersuchungsgebiet vorhanden ist. Hat der Raum keinen Einfluss, verteilen sich die Werte für die Parameter zufällig im Raum. Ergeben sich hingegen klare räumliche Muster gleich hoher Parameterwerte in benachbarten Gemeinden, kann davon ausgegangen werden, dass die geographische Position die Stärke der Abhängigkeit der beiden untersuchten Variablen massgebend beeinflusst (Fotheringham et al. 2002).

#### 4. Resultate

Das auf Wikipedia basierende Gemeindefeldnetzwerk der Nordostschweiz ist in Abbildung 2 wiedergegeben. Es sind nur die strukturell wichtigsten Verbindungen dargestellt.

Die Stärke der Gemeindeverbindungen zeigt sich im Helligkeitswert und in der Strichstärke der Linien. Je dunkler und dicker eine Linie zwischen zwei Gemeinden ist, desto stärker ist die Verbindung zwischen den Gemein-

den. Die Gemeinden selbst sind als Punktsymbole abgebildet. Die Grösse der Punkte widerspiegelt die Zentralität der Gemeinden in Bezug auf ihren gewichteten Grad. Je grösser eine Gemeinde dargestellt ist, desto zentraler ist ihre Stellung im Netzwerk. Die Übersichtskarte links oben in Abbildung 2 zeigt das Untersuchungsgebiet in geographischer Perspektive. Die unterschiedliche Schraffur im Netzwerk sowie in der Übersichtskarte repräsentiert die Zugehörigkeit der jeweiligen Gemeinden zu den fünf Blondel-Gruppen. In der Darstellung sind einige Gemeinden beschriftet, auf die im Detail in den Abschnitten *Evaluation* und *Diskussion* eingegangen wird. Im Folgenden werden einige Besonderheiten des abgebildeten Netzwerkes beschrieben. Eine detaillierte Analyse ist in Bruggmann (2012) zu finden.

Das Gemeindefeldnetzwerk weist eine sternförmige Struktur auf. Eine Vielzahl von Verbindungen führen von der im Zentrum liegenden Gemeinde St. Gallen nach aussen. Die Verbindungen sind insbesondere auf den vom Zentrum ausgehenden Hauptästen des Netzwerkes überdurchschnittlich stark. Die Verbindungen, welche sich nicht auf diesen Hauptästen befinden, sind vergleichsweise schwach.

Weiter zeigt sich eine klar ausgeprägte topologische und räumliche Struktur bei der Zugehörigkeit der Gemeinden zu den Blondel-Gruppen. Einerseits befinden sich Gemeinden, welche zur gleichen Blondel-Gruppe gehören, auch im Netzwerk nahe beieinander. Andererseits zeigt der Vergleich mit der Übersichtskarte, dass die gruppierten Gemeinden auch geographisch nahe beieinander liegen. Um alle Regionen von Blondel-Gruppen in der Übersichtskarte könnten somit Grenzen gezogen werden, innerhalb welcher keine Gemeinden einer anderen Blondel-Gruppe vorkämen.

Es zeigt sich, dass die Gemeinden der Agglomeration Buchs (SG)-Vaduz gemäss offizieller Agglomerationsdefinition einheitlich als Blondel-Gruppe abgebildet werden. Die vor allem im Zentrum des Netzwerkes konzentrierte Gruppe *St. Gallen-Appenzell* hingegen enthält neben der Agglomeration St. Gallen noch weitere Agglomerationsgemeinden, unter anderem die Gemeinden der Agglomeration Rapperswil-Jona-Rüti. Diese Gemeinden werden als nur schwach bis mittel in Bezug auf Zentralität und Stellung im Netzwerk ausgewiesen und werden daher vom angewandten Algorithmus nicht als eigenständige Gemeindegruppe identifiziert, was im nächsten Teil dieser Arbeit genauer evaluiert wird.





bei einem Signifikanzniveau von 5% abgelehnt wird. Die Hypothese, dass die Gruppierung der Gemeinden nach Blondel et al. (2008) nur zufällig zustande gekommen ist, wird also deutlich verworfen. Trotz Agglomerationen wie Buchs (SG)-Vaduz, die sehr gut nachgezeichnet werden, liegt das Testresultat nicht bei  $p=0.00$ . Dies hat seine Ursache unter anderem in der Agglomeration Rapperswil-Jona-Rüti. Die Gemeinden dieser Agglomeration sind auf Wikipedia im gewählten Untersuchungsgebiet nur sehr schwach verbunden und werden nicht als eigene Gemeindegruppe ausgewiesen, sondern zur grossen Gruppe mit St.Gallen im Zentrum des Untersuchungsgebietes gezählt, was bereits in Abbildung 2 zu erkennen ist. Der Ursache dieser Auffälligkeit wird im nächsten Teil der Evaluation nachgegangen.

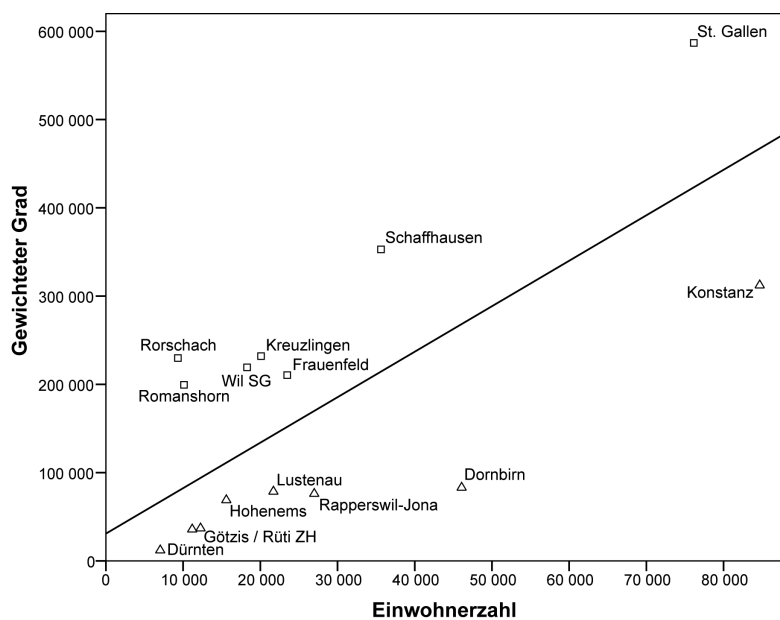
### 5.2 Quantifizierung und Auswertung des räumlichen Einflusses

Um den Einfluss des Raumes auf die präsentierten Resultate zu eruieren, wird in einem ersten Schritt eine Korrelationsanalyse mit den beiden Variablen *Einwohnerzahl* und *gewichteter Grad* durchgeführt. Die Berechnung des Korrelationskoeffizienten nach Spearman (Methodenberatung 2012) ergibt einen Wert von 0.645, der bei einem Signifikanzniveau von 5% auf einen signifikanten positiven Zusammen-

hang der beiden Variablen schliessen lässt. Dies bedeutet, dass eine Gemeinde mit einer hohen Einwohnerzahl tendenziell auch eine höhere Zentralität aufweist als eine Gemeinde mit einer tiefen Einwohnerzahl.

In Abbildung 3 ist auf der x-Achse die Einwohnerzahl und auf der y-Achse der gewichtete Grad aufgetragen. Die schwarze Linie, welche vom unteren linken zum oberen rechten Bildrand verläuft, stellt die lineare Anpassungsgerade an die Gemeinden dar. Die meisten Gemeinden im Untersuchungsgebiet haben eine Einwohnerzahl von weniger als 10 000 und einen gewichteten Grad von kleiner als 100 000. Abgebildet sind jedoch nur die beschrifteten und als Drei- oder Vierecke dargestellten Gemeinden, welche sich durch eine überdurchschnittlich hohe Einwohnerzahl und einen auffällig hohen beziehungsweise tiefen gewichteten Grad auszeichnen, da diese in der folgenden Evaluierung genauer betrachtet werden.

Die Gemeinden mit einem überdurchschnittlich hohen gewichteten Grad (Rechtecke) sind Kernagglomerationsgemeinden im schweizerischen Teil des Untersuchungsgebietes. Ihre überdurchschnittlich hohe Zentralität hat sich bereits in Abbildung 2 manifestiert, indem alle diese Gemeinden entlang der Hauptäste des Netzwerkes liegen. Die Gemeinden, deren Werte für den gewichteten Grad im Vergleich zur Einwohnerzahl sehr tief sind (Dreiecke), lie-



#### Gemeinden im Untersuchungsgebiet

- Gemeinde mit einer hohen Einwohnerzahl und einem verhältnismässig hohen gewichteten Grad
- △ Gemeinde mit einer hohen Einwohnerzahl und einem verhältnismässig tiefen gewichteten Grad

Abb. 3: Zusammenhang der Variablen *Einwohnerzahl* (x-Achse) und *gewichteter Grad* (y-Achse) ausgewählter Gemeinden der Nordostschweiz.

gen im ausländischen Teil des Untersuchungsgebietes oder in der Agglomeration Rapperswil-Jona-Rüti. Diese Gemeinden befinden sich alle am Rande des Untersuchungsgebietes. Dies könnte darauf zurückzuführen sein, dass Gemeinden am Rande vergleichsweise viele im Zentralitätswert nicht erfasste Verbindungen zu Gemeinden, welche ausserhalb des Untersuchungsgebietes liegen, aufweisen. Um diesen postulierten räumlichen Einfluss zu überprüfen, wird die gesamte methodische Vorgehensweise wiederholt, jedoch werden neu auch alle Gemeinden des Kantons Zürich zum Untersuchungsgebiet gerechnet. Dabei zeigt sich, dass der gewichtete Grad der Gemeinden der Agglomeration Rapperswil-Jona-Rüti im Vergleich zu den restlichen Gemeinden des Untersuchungsgebietes überdurchschnittlich stark ansteigt. Diese starke Verbindung der Gemeinden der Agglomeration Rapperswil-Jona-Rüti zu Zürcher Gemeinden stimmt mit den Pendlerverflechtungen gemäss der Volkszählung 2000 überein. Mehr als die Hälfte der zehn wichtigsten Pendlerzielgemeinden dieser Agglomerationsgemeinden befinden sich im Kanton Zürich.

Dieselben Feststellungen hinsichtlich der räumlichen Einflüsse bei Gemeinden der Agglomeration Rapperswil-Jona-Rüti können auch auf andere Gemeinden am Rande des Untersuchungsgebietes übertragen werden. Führt man folglich eine Korrelationsanalyse ohne die Gemeinden im ausländischen Teil des Untersuchungsgebietes und ohne die Agglomerationsgemeinden im Kanton Zürich durch, steigt der Korrelationskoeffizient nach Spearman auf 0.704 an, was den postulierten räumlichen Einfluss bestätigt.

Um den räumlichen Einfluss vertieft zu untersuchen und zu lokalisieren, wird im nächsten Schritt eine geographisch gewichtete Regression durchgeführt. Die Durchführung einer geographisch gewichteten Regression eignet sich, wenn Autokorrelationsmuster im räumlichen Kontext erwartet werden. Dies bedeutet, dass die Werte einer Variablen nicht zufällig im Raum verteilt sind, sondern dass sie von ihrer geographischen Position abhängig sind (Cliff, Ord 1970). Die Konzentration gleicher Werte in einer Region stellt ein typisches Autokorrelationsmuster dar. Die Güte des Modells wird mit dem lokalen  $R^2$  quantifiziert.  $R^2$  zeigt die Höhe der Anpassung der tatsächlichen an die erwarteten Werte der Regressionsparameter (Backhaus et al. 2011). Je höher  $R^2$  liegt, desto besser ist die Anpassung der tatsächlichen an die erwarteten Werte. Die Einwohnerzahl wird als unabhängige Variable und der gewichtete Grad als abhängige Variable festgelegt. Eine Übersicht über die lokalen  $R^2$ -Werte im Untersuchungsgebiet ist in Abbildung 4 zu sehen. Dabei weist ein heller Punkt auf einen tiefen, ein dunkler Punkt auf einen hohen  $R^2$ -Wert hin.

Wie Abbildung 4 zeigt, weisen insbesondere Gemeinden im Osten und im Süden am Rande des Untersuchungsgebietes tiefe lokale  $R^2$ -Werte auf. In der gesamten Agglomeration Rapperswil-Jona-Rüti liegen die Werte für den lokalen  $R^2$ -Wert unter 50%. Im zentralen Bereich sowie im Nordwesten nehmen die Gemeinden hingegen verhältnismässig hohe Werte bis zu 98% an. Der Randeffekt tritt also auch bei der Evaluierung mittels geographisch gewichteter Regression deutlich zum Vorschein.

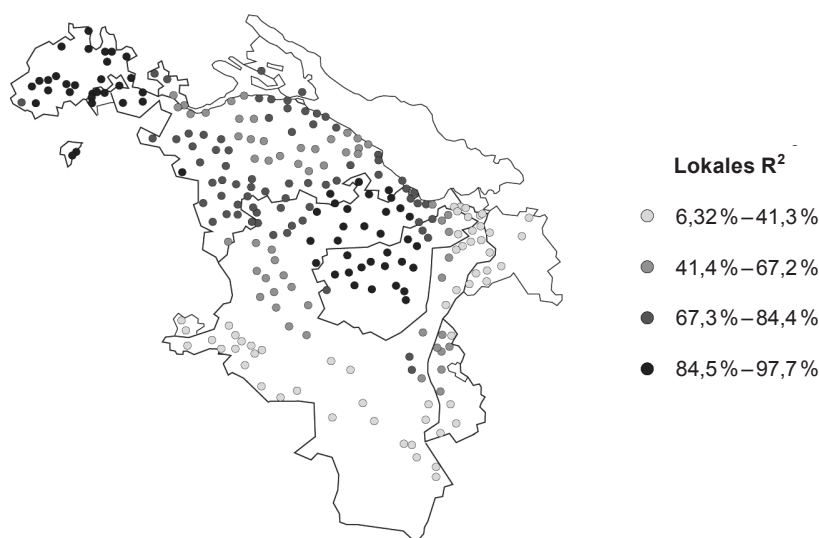


Abb. 4: Lokale  $R^2$  der geographisch gewichteten Regression mit der unabhängigen Variable *Einwohnerzahl* und der abhängigen Variable *gewichteter Grad*.



Es zeigt sich ausserdem, dass der lokale  $R^2$ -Wert keiner zufälligen Verteilung über das gesamte Untersuchungsgebiet folgt. Vielmehr sind gleiche Werte in denselben Regionen konzentriert, was für räumlich autokorrelierte Regressionsparameter spricht. Die restlichen Regressionsparameter zeigen eine ähnliche räumliche Verteilung. Die Abhängigkeit der Variable *gewichteter Grad* von der Variable *Einwohnerzahl* wird also vom Raum beeinflusst und ist je nach Region unterschiedlich stark.

## 6. Schlussfolgerungen

Diese Arbeit evaluiert nutzergenerierte Textdatenbanken, spezifisch die Online-Enzyklopädie Wikipedia, als alternative Datengrundlage zur Definition von Agglomerationen in der Schweiz. Im Vergleich zu traditionellen Datensätzen wie der Volkszählung zeichnet sich Wikipedia vor allem durch die Aktualität und die unentgeltliche Verfügbarkeit ihres Inhaltes im World Wide Web aus. Im Kontext dieser Arbeit hat sich insbesondere auch die Möglichkeit, die aufgezeigte Methodik grenzüberschreitend anzuwenden, als sehr nützlich herausgestellt. Gemeinden verschiedener Länder können auf einfache Art und Weise in die Untersuchung integriert werden. Dies ermöglicht die Beantwortung von Fragestellungen zu grenzüberschreitenden Themen. Die erste Fragestellung, ob sich Wikipedia mit den heute verfügbaren nutzergenerierten Inhalten eignet, um mit modernen Methoden der Informationsvisualisierung und -analyse Agglomerationen zu identifizieren, lässt sich mit *ja* beantworten. Die allgemeinen, im Netzwerk ersichtlichen Strukturen und Gemeindegruppen bilden die geographischen und administrativen Verhältnisse im gewählten Untersuchungsgebiet gut ab. Die Gemeinden vieler Agglomerationen, wie die der Agglomeration Buchs (SG)-Vaduz, werden, wie in der Definition vom Bundesamt für Statistik, einheitlich als Agglomerationen ausgewiesen.

Im Rahmen der Untersuchung der zweiten Fragestellung, welche die Raum- und Randeffekte und besonders davon betroffene Regionen evaluiert, hat sich hingegen gezeigt, dass der Raum einen nicht zu vernachlässigenden Einfluss auf die Resultate dieser Arbeit ausübt. Das Beispiel der Agglomeration Rapperswil-Jona-Rüti versinnbildlicht die angesprochene Problematik. Die Korrelationsanalyse zeigt, dass die Gemeinden dieser Agglomeration im Vergleich zu den anderen Gemeinden des Untersuchungsgebietes viel zu tiefe Werte für den

gewichteten Grad im Verhältnis zu ihren Einwohnerzahlen aufweisen. Gleiches gilt auch für andere Gemeinden am Rande des Untersuchungsgebietes. Dies kann unter anderem darauf zurückgeführt werden, dass Gemeinden am Rande des Untersuchungsgebietes starke Verbindungen zu Gemeinden ausserhalb des Untersuchungsgebietes aufweisen, die im gewichteten Grad nicht berücksichtigt werden.

Die geographisch gewichtete Regressionsanalyse bestätigt diese Ergebnisse und zeigt des Weiteren auf, dass die Abhängigkeit der Variable *gewichteter Grad* von der Variable *Einwohnerzahl* durch die geographische Lage der jeweiligen Gemeinden beeinflusst wird. Benachbarte Gemeinden weisen also grundsätzlich ähnlich hohe Regressionsparameter auf. Neben dem gezeigten Randeffekt ist also auch ein starker Raumeffekt in Form von räumlich autokorrelierten Regressionswerten zu erkennen.

Es zeigt sich, dass räumliche Einflüsse bei der Untersuchung einer an sich nicht-geographischen Datengrundlage wie Wikipedia vorhanden und sorgfältig zu analysieren sind. Um die Randeffekte bei der Agglomeration Rapperswil-Jona-Rüti zu verkleinern, könnte man zum Beispiel den gesamten Kanton Zürich in das Untersuchungsgebiet aufnehmen. Dann würden jedoch erneut räumliche Effekte am Rande des neu gewählten Untersuchungsgebietes auftreten. Deshalb ist es nicht möglich, diese Problematik bei der Wahl eines räumlichen Ausschnittes als Untersuchungsgebiet ganzheitlich zu beseitigen. Vielmehr sollte man sich dieser Problematik bei einer Untersuchung, welche nicht auf eine Datengrundgesamtheit zurückgreift, bewusst sein und die daraus resultierenden Effekte möglichst genau lokalisieren und quantifizieren.

Weitere Untersuchungen mit diesem Datensatz könnten, wie in Bruggmann (2012) gezeigt, zusätzlich die inhaltliche Dimension der Wikipedia-Verbindungen aufgreifen. Dadurch könnten Erkenntnisse darüber gewonnen werden, welche Themen für die einzelnen Gemeindeverbindungen in Agglomerationen prägend sind. Dies hilft, die Forderung nach einer neuen Agglomerationsdefinition in der Schweiz mit möglichen neuen Themen und Variablen auszustatten, um die Variable *Pendlerbewegungen* adäquat zu ersetzen. Die Resultate von Bruggmann (2012) bekräftigen insbesondere die von Thierstein et al. (2006) formulierte Forderung, wirtschaftliche Beziehungen in der Form von unternehmensinternen und -externen Verflechtungen in der Agglomerationsdefinition vermehrt zu berücksichtigen.

Möglich wäre auch eine Untersuchung, welche zusätzlich die zeitliche Dimension berücksichtigt, um Veränderungen über einen längeren Zeitraum zu dokumentieren. Dafür wäre es notwendig, die Gemeindeverbindungen zeitlich differenziert zu betrachten.

Zusätzlich wäre es denkbar, das Untersuchungsgebiet auf die gesamte Schweiz oder den gesamten deutschsprachigen Raum auszuweiten, um mögliche räumliche Einflüsse und Randeffekte einzudämmen. Bei einer Fokussierung auf die gesamte Schweiz müssten jedoch entweder Verbindungen verschiedener Sprachversionen von Wikipedia berücksichtigt oder auf die sprachneutrale englische Version von Wikipedia zurückgegriffen werden, welche jedoch bedeutend weniger Verbindungsartikel zu Gemeinden in der Schweiz enthält als beispielsweise die deutsche Sprachversion.

Ebenfalls könnten die Rollen der Gemeinden innerhalb der Agglomerationen eingehender untersucht werden. Kerngemeinden einer Agglomeration könnten dann von Nicht-Kerngemeinden und von ländlichen Gemeinden unterschieden werden.

Diese Arbeit hat das Potential nutzergenerierter Inhalte im Kontext von wirtschaftsgeographischen Fragestellungen aufgezeigt. Insbesondere die Möglichkeit, administrative Einheiten verschiedener Länder in einer Untersuchung gemeinsam zu integrieren, zeigt sich als grosse Stärke dieser Arbeit und stellt einen Vorteil gegenüber der Volkszählung als traditionelle statistische Datengrundlage zur Definition von Agglomerationen dar.

## Literatur

- BACKHAUS, K.; ERICHSON, B.; PLINKE, W.; WEIBER, R. (2011): Regressionsanalyse. In *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung* (13. Auflage). Berlin/Heidelberg: Springer Verlag.
- BLONDEL, V. D.; GUILLAUME, J.-L.; LAMBIOTTE, R.; LEFEBVRE, E. (2008): Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* P10008.
- BRUGGMANN, A. (2012): Netzwerkvisualisierung der Ostschweiz – Die Raumgliederung der Schweiz mit Wikipedia neu formuliert. *Geographic Information Visualization & Analysis*. Universität Zürich.
- BUNDESAMT FÜR STATISTIK (2007): Überarbeitung der Agglomerationsdefinition – Grundlagenstudie zur Bestandesanalyse und Bedürfnisabklärung. Sektion räumliche Analysen, <http://www.bfs.admin.ch/bfs/portal/de/index/regionen/11/pro/01.Document.105558.pdf> (Zugriff: September 2012).
- BUNDESAMT FÜR STATISTIK (2011): Die neue Volkszählung. Eidgenössisches Departement des Inneren, <http://www.volkszählung.ch> (Zugriff: Januar 2012).
- BUNDESAMT FÜR STATISTIK (2012): Agglosuisse – Eine neue Agglomerationsdefinition für die Schweiz. <http://www.bfs.admin.ch/bfs/portal/de/index/regionen/11/pro/01.html> (Zugriff: Februar 2012).
- BUNDESAMT FÜR STATISTIK (2013): Räumliche Gliederungen der Schweiz - Analyseregionen. [http://www.bfs.admin.ch/bfs/portal/de/index/regionen/11/geo/analyse\\_regionen/01.html](http://www.bfs.admin.ch/bfs/portal/de/index/regionen/11/geo/analyse_regionen/01.html) (Zugriff: Oktober 2013).
- CLIFF, A. D.; ORD, K. (1970): Spatial Autocorrelation: A Review of Existing and New Measures with Applications. *Economic Geography*, 46, S. 269–292.
- FABRIKANT, S. I.; MONTELLO, D.; RUOCCO, M.; MIDDLETON, R. (2004): The distance-similarity metaphor in network display spatializations. *Cartography and Geographic Information Science*, 31, S. 237–252.
- FOTHERINGHAM, A. S.; BRUNSDON, C.; CHARLTON, M. (2002): *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester, England: Wiley.
- FOTHERINGHAM, A. S.; BRUNSDON, C.; CHARLTON, M. (2012): *Geographically Weighted Regression – What is GWR?* <http://ncg.nuim.ie/ncg/GWR/whatis.htm> (Zugriff: März 2012).
- HECHT, B. (2007): *Utilizing Wikipedia as a Spatiotemporal Knowledge Repository*. University of California, Santa Barbara.
- HECHT, B.; RAUBAL, M. (2008): GeoSR: Geographically Explore Semantic Relations in World Knowledge. In BERNARD, L.; FRIIS-CHRISTENSEN, A.; PUNDT, H. (Hrsg.), *11th AGILE International Conference on Geographic Information Science*. Girona, Spain.
- KOS, A. J.; PSENICKA, C. (2000): Measuring cluster similarity across methods. *Psychological Reports*, 86, pp. 858–862.
- METHODENBERATUNG (2012): Rangkorrelation nach Spearman. <http://www.methodenberatung.uzh.ch/datenanalyse/zusammenhaenge/rangkorrelation.html> (Zugriff: März 2012).
- SALVINI, M. M. (2012): Spatialization von nutzergenerierten Inhalten für die explorative Analyse des globalen Städtetetzes. *Geographic Information Visualization & Analysis*. Universität Zürich.
- SCHULER, M.; DESSEMONTET, P.; JOYE, D.; PERLIK, M. (2005): *Die Raumgliederungen der Schweiz*. Neuenburg: Bundesamt für Statistik.
- THIERSTEIN, A.; KRUSE, C.; GLANZMANN, L.; GABI, S.; GRILLON, N. (2006): *Raumentwicklung im Verborgenen die Entwicklung der Metropolregion Nordschweiz*. Zürich: Verlag Neue Zürcher Zeitung.
- WASSERMAN, S.; FAUST, K. (1994): *Social network analysis methods and applications*. Cambridge: Cambridge University Press.
- WIKIPEDIA (2012): Deutsche Wikipedia-Version. <http://de.wikipedia.org> (Zugriff: April 2012).

MSc André Bruggmann  
Doktorand Geographische  
Informationsvisualisierung  
& -analyse  
Geographisches Institut  
Universität Zürich – Irchel  
Winterthurerstrasse 190  
CH-8057 Zürich  
[andre.bruggmann@geo.uzh.ch](mailto:andre.bruggmann@geo.uzh.ch)

Dr. Marco Michele Salvini  
Projektleiter Stadtentwicklung  
Stadtbauamt Aarau  
Rathausgasse 1  
CH-5000 Aarau  
[marco.salvini@aarau.ch](mailto:marco.salvini@aarau.ch)

Prof. Dr. Sara Irina Fabrikant  
Professorin für Geographische  
Informationsvisualisierung  
& -analyse  
Geographisches Institut  
Universität Zürich – Irchel  
Winterthurerstrasse 190  
CH-8057 Zürich  
[sara.fabrikant@geo.uzh.ch](mailto:sara.fabrikant@geo.uzh.ch)