

Abgrenzung von Wirtschaftsräumen auf der Grundlage von Pendlerverflechtungen

Ein Methodenvergleich

Per Kropp, Barbara Schwengler

- Entwurf 2008-08-27 -

Mit der Reihe „IAB-Discussion Paper“ will das Forschungsinstitut der Bundesagentur für Arbeit den Dialog mit der externen Wissenschaft intensivieren. Durch die rasche Verbreitung von Forschungsergebnissen über das Internet soll noch vor Drucklegung Kritik angeregt und Qualität gesichert werden.

The “IAB Discussion Paper” is published by the research institute of the German Federal Employment Agency in order to intensify the dialogue with the scientific community. The prompt publication of the latest research results via the internet intends to stimulate criticism and to ensure research quality at an early stage before printing.

Inhaltsverzeichnis

/Abstract.....	4
Abstract.....	4
1 Einleitung.....	5
2 Forschungsstand.....	7
3 Daten.....	10
4 Abgrenzungskriterien.....	16
5 Abgrenzungsverfahren.....	19
5.1 Schwellenwertmodelle.....	20
5.2 Clusteranalyse.....	24
5.3 Alternative Cluster-Verfahren.....	32
5.4 Faktorenanalyse.....	37
6 Vergleich der Abgrenzungsergebnisse.....	41
6.1 Visueller Vergleich.....	41
6.2 Quantitativer Vergleich.....	42
6.2.1 Modularitätswerte.....	42
6.2.2 Pendlerquoten.....	44
6.2.3 Balance von Arbeitsangebot und –nachfrage (ESC, HSC, HWR).....	44
6.2.4 Die räumliche Struktur.....	48
6.3 Vergleich der Bewertungsmaße.....	49
7 Zusammenfassung.....	52
Literaturverzeichnis.....	56
Anhang.....	60

Abstract

Funktionale Regionen, sogenannte regionale Arbeitsmärkte, spielen in der nationalen und internationalen Regionalforschung eine besondere Rolle. Im Gegensatz zu administrativen Regionen ermöglichen sie es, wirtschaftliche Aktivitäten und Pendlerverflechtungen in geeigneter Weise abzubilden. Für die Abgrenzung dieser Wirtschaftsräume können verschiedene Verfahren herangezogen werden. In diesem Discussion Paper werden Abgrenzungen auf Basis der Kennziffernmethode, der Cluster- und der Faktorenanalyse anhand von Pendlerdaten auf Kreisebene des Jahres 2005 miteinander verglichen und beurteilt. Als Gütekriterien werden dabei die Modularität Q , die von van der Laan und Schalke (2001) verwendeten Kennziffern ESC und HSC für die Balance von Arbeitsangebot und -nachfrage sowie Maße zur Größenhomogenität herangezogen.

Clusterverfahren und auf der Kennziffernmethode basierenden graphentheoretischen Verfahren gelingt es am Besten, die Pendlerverflechtungen zwischen den Kreisregionen zu erfassen. Faktorenanalytische Verfahren produzieren dagegen eher größenhomogene Abgrenzungen, die die Pendlerverflechtungen allerdings weniger gut abbilden. Mit allen hier vorgestellten Verfahren können Wirtschaftsräume so bestimmt werden, dass sie die Pendlerbewegungen in Deutschland deutlich besser wiedergeben als dies für regionalpolitische Zwecke bisher genutzten Abgrenzungen erlauben. Anhand der in dieser Studie vorgestellten Bewertungsmöglichkeit der Güte von Abgrenzungen ist es weiterhin möglich, die optimale Anzahl von Region für eine adäquate Abbildung der Pendlerverflechtungen zu bestimmen.

JEL Klassifikation: C21, J40, R12, R58

Keywords: Abgrenzung, Arbeitsmarktregion, Pendlerverflechtung, funktionale Region, Modularität

1 Einleitung

Ziel der vorliegenden Arbeit ist ein Vergleich von Methoden zur Abgrenzung von Wirtschaftsräumen bzw. regionalen Arbeitsmärkten oder Arbeitsmarktregionen auf der Grundlage von Pendlerverflechtungen. Diese eignen sich besonders gut für Fragestellungen der Regionalforschung, weil sie "relativ eigenständige wirtschaftliche Teilräume" (Keilbach 2000: 120 nach Eckey u.a. 2006: 299) darstellen und die wirtschaftlichen Aktivitäten im Raum in geeigneter Weise abbilden. Der Zuschnitt von administrativen Gebietseinheiten wie Kreisen und Bundesländern orientiert sich hingegen häufig an historisch gewachsenen Verwaltungsstrukturen auf der Landes- und Kommunalebene. Aus diesem Grund eignen sich administrative Regionen weniger für regionalökonomische Analysen, denn sie durchschneiden häufig zusammenhängende Wirtschaftsregionen.

Die praktische Relevanz von funktionalen Gliederungen zeigt sich in der Regionalpolitik oder im Standortmarketing – z.B. bei der Beantwortung von Fragen wie:

- Welche Administrationen sollten zusammenarbeiten, weil ihre Maßnahmen einen gemeinsamen Arbeitsmarkt betreffen? Die Bildung von Metropolregionen in Deutschland in den letzten Jahren zeugt von Versuchen, Wirtschafts- und Arbeitsmarktpolitik entsprechend zu koordinieren, auch wenn das Selbstverständnis der Metropolregionen eher von deren überregionalen Funktion als vom Selbstbild eines regionalen Wirtschaftsraums geprägt ist.¹
- Welche Regionen profitieren von Maßnahmen in Region X? Diese Frage ähnelt der obigen, ist aber vor allem für die Evaluation arbeitsmarktpolitischer Maßnahmen relevant.
- Welches Humankapital finden Unternehmen in einer Region vor? Die Kenntnis von Arbeitsmarktregionen ist unabdingbar für die Bestimmung von Arbeitsangebot und – nachfrage an einem Standort.

Van der Laan und Schalke (2001) resümieren, dass Regionalpolitik für homogene und zusammenhängende Regionen erfolgreicher ist. Genau solche Eigenschaften sollten funktional abgegrenzte Regionen eher aufweisen als administrativ bestimmte.

Die wissenschaftliche Relevanz der Abgrenzung von funktionalen Regionen liegt darin, dass sie die relevanten Analyseeinheiten für die Regionalforschung liefern. Effekte unterschiedlicher Regionalisierungen auf Arbeitsmarktmodelle zeigten schon Barkley u.a. (1995) und Baumann u.a. (1983). Auch Eckey u.a. (2006) verweisen auf die Probleme von Analysen auf Grundlage von Verwaltungseinheiten z.B. bei der

¹ Vgl. z.B. Initiativkreis Metropolregionen in Deutschland (2003) oder die Selbstdarstellung des Initiativkreis Metropolregionen in Deutschland im Internet (<http://www.deutsche-metropolregionen.org> [28.7.2008]).

Berechnung der Arbeitsproduktivität oder der Analyse von Wissens-Spillovern. Vielfach diskutiert sind in der Literatur darüber hinaus die Probleme, die sich aus der unterschiedlichen Aggregation von kleinräumig vorhandenen Daten ergeben – das "modifiable area unit problem" (Openshaw/Taylor 1981).

Die vorliegende Studie ist wie folgt gegliedert: Zunächst wird ein Überblick über den Forschungsstand zu bisherigen Abgrenzungsmethoden auf deutscher und internationaler Ebene gegeben. Anschließend werden die den Analysen zugrunde liegenden Daten vorgestellt und wesentliche Parameter der Pendlerbeziehungen zwischen Regionen in Deutschland beschrieben. In Kapitel 4 folgt eine Diskussion grundlegender Eigenschaften von regionalen Arbeitsmärkten, bevor in Kapitel 5 die verschiedenen Abgrenzungsverfahren und die damit gebildeten Arbeitsmarktregionen vorgestellt werden. Die Abgrenzungen erfolgen in der Regel für 150 Regionen, um die Vergleichbarkeit mit der z.Z. aktuellen Gliederung von Eckey u.a. (2006) zu gewährleisten. Die Ergebnisse der verschiedenen Verfahren werden im Anschluss daran in Kapitel 6 anhand unterschiedlicher Gütekriterien verglichen. In Kapitel 7 wird abschließend diskutiert, für welche Zwecke welche Abgrenzungen am besten geeignet erscheinen und ein Ausblick für weitere Forschungen gegeben.

2 Forschungsstand

Das Konzept funktionaler Regionen versucht die Realität räumlicher ökonomischer Prozesse so gut wie möglich zu berücksichtigen. Funktionale ökonomische Regionen sind definiert als Regionen, die einen Großteil der ökonomischen Aktivitäten innerhalb ihrer Grenzen konzentrieren (Bongaerts u.a. 2004: 2). Für viele Wirtschaftsprozesse lässt sich zwar nur sehr eingeschränkt von begrenzten Wirtschaftsräumen sprechen, weil sie auf globalen Märkten stattfinden; andererseits formt gerade die eingeschränkte Mobilität von Personen regionale Märkte, die sich deutlich voneinander unterscheiden können. Weil die Mobilität begrenzt ist, dürften Waren- oder Immobilienmärkte regionale Besonderheiten aufweisen, auch wenn diese ggf. durch ein Stadt-Umland-Gefälle überdeckt werden. In diesem Sinne lassen sich über Pendlerbewegungen mehr als nur regionale Arbeitsmärkte identifizieren. Auch für eine globale Wirtschaft behält die Kenntnis über die Abgrenzung von Arbeitsmärkten ihre Relevanz – sie ist wichtig für wirtschaftspolitische und wirtschaftliche Entscheidungen, denn sie ermöglicht die Einschätzung des in einer Region zur Verfügung stehenden Humankapitals.

Das Pendeln zwischen Wohn- und Arbeitsort stellt typischerweise eine regional begrenzte Aktivität dar, die wiederum relativ homogene Immobilienmärkte und andere Märkte schafft. Arbeitsmarktregionen, die Pendleraktivitäten gut abbilden, werden darum relativ geschlossene Märkte sowohl für die Arbeit als auch für das Wohnen erfassen. Diese Kriterien können z.B. als ESC- und HSC-Maße (**self-containment ratios for employment and housing**; siehe van der Laan/Schalke 2001: 209 ff.) quantifiziert werden und werden auch in dieser Arbeit für den Vergleich der Güte von Arbeitsmarkt-Abgrenzungen angewandt.

International bekannt sind Travel-to-Work Areas (TTWA), (Local) Labour Market Areas (LLMA) oder Standard Metropolitan Statistical Areas (SMSA). Travel-to-Work Areas sind beispielsweise relevant für Arbeitsmarktanalysen und Industriepolitik, als Basis für die Statistik und auch zur Reorganisation der Regionalverwaltung in Großbritannien (ONS/Coombes 1998). Die Local Labour Market Areas finden Anwendung in den Niederlanden (van der Laan/Schalke 2001, Hensen/Cörvers 2003), Belgien (van Nuffel 2007), Deutschland (Eckey u.a. 2006) und Italien, während Standard Metropolitan Statistical Areas als relevante Einheiten für die Wirtschaftspolitik in den USA herangezogen werden.

In Deutschland liegen seit Anfang der 1970er Jahre zahlreiche wissenschaftliche Studien zur Abgrenzung von regionalen Arbeitsmärkten vor (siehe u.a. Klemmer/Kraemer 1975, Eckey 1988, Budde u.a. 1993, Budde u.a. 1995, Eckey 2001). Die aktuellste Abgrenzung in 150 Arbeitsmarktregionen stammt von Eckey u.a. (2006). Regional begrenzte Untersuchungen jüngerer Datums existieren für die Bundesländer Thüringen und Sachsen-Anhalt (Granato/Farhauer 2007) und Berlin und Brandenburg (Binder/Schwengler 2006).

Neben regionalökonomischen Analysen im theoretisch-wissenschaftlichen Bereich werden funktionale Gliederungen häufig auch im politisch-administrativen Bereich genutzt. Hier werden zum einen „Arbeitsmarktregionen“ als Diagnoseeinheit für die Identifikation von deutschen Fördergebieten verwendet und zum anderen „Raumordnungsregionen“ für die Raumordnungsberichterstattung des Bundes genutzt. Die unterschiedlichen Abgrenzungen dieser funktionalen Regionen im politisch-administrativen Bereich unterliegen dabei häufig bestimmten Zwängen und Vorgaben und unterscheiden sich daher von rein wissenschaftlichen Abgrenzungen, die solche Einschränkungen nicht berücksichtigen müssen.

So mussten beispielsweise im Rahmen der Neuabgrenzung der Fördergebiete der Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur“ im Jahr 2007 die bis dahin gültigen Arbeitsmarktregionen an die förderrechtlichen Vorgaben der Europäischen Kommission angepasst werden. Das bedeutete, dass die ehemals länderübergreifende und nicht kreisscharfe, aber durch starke Pendlerbewegungen zwischen der Stadt Berlin und dem Brandenburger Umland verflochtene Arbeitsmarktregion Berlin neu und länderscharf abzugrenzen war. Vor diesem Hintergrund wurden alle bestehenden Arbeitsmarktregionen in den Ländern Berlin und Brandenburg anhand aktueller Pendlerdaten eingehend untersucht und ein neuer Gebietszuschnitt für diese beiden Bundesländer erarbeitet. Dies führte dazu, dass sich in der aktuell gültigen Abgrenzung der Fördergebiete die Gesamtzahl der Arbeitsmarktregionen von 271 auf 270 reduzierte (Binder/Schwengler 2006; vgl. Abbildung A1.b im Anhang).

Zum anderen gibt es die für administrative Zwecke genutzten normativen Planungs- und Verwaltungsregionen für die Raumordnungsberichterstattung des Bundes. Hierbei handelt es sich um 97 funktionale Raumordnungsregionen des Bundesamts für Bauwesen und Raumordnung (BBR), die alle kreis- und landesscharf geschnitten sind und auch Regierungsbezirke nicht durchschneiden. Die jüngste Abgrenzung dieser großräumigen Einheiten wurde im Jahr 1996 anhand von Pendlerverflechtungen des Jahres 1993 auf Kreisbasis durchgeführt (Böltken u.a. 1996; vgl. Abbildung A1.a im Anhang).

Die folgende Tabelle 1 stellt die vorgestellten administrativen und funktionalen Regionen mit ihren wesentlichen Merkmalen einander gegenüber. Sowohl hinsichtlich der Anzahl als auch der Größenunterschiede der abgegrenzten Regionen (siehe Standardabweichung) zeigen sich deutliche Unterschiede.

Tabelle 1: Administrative und funktionale Gliederungen zum 01.01.2007

Gliederung	Zweck	Einwohner 2005 (in Tausend)			Regionen	Beschränkungen
		Minimum	Maximum	Standardabweichung		
Kreise	administrative Gliederung	35	3.395	219	439	kreis- und bundeslandscharf
Raumordnungsregionen	funktional, Raumplanung	227	3.395	612	97	kreis- und bundeslandscharf
Arbeitsmarktregionen der Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschafts-struktur“	Wirtschaftspolitik	64	3.395	399	270	kreis- und bundeslandscharf
Regionale Arbeitsmärkte (Eckey u.a. 2006)	regional-ökonomische Analysen	64	4.443	680	150	kreisscharf, Minimal- und Maximalgröße

3 Daten

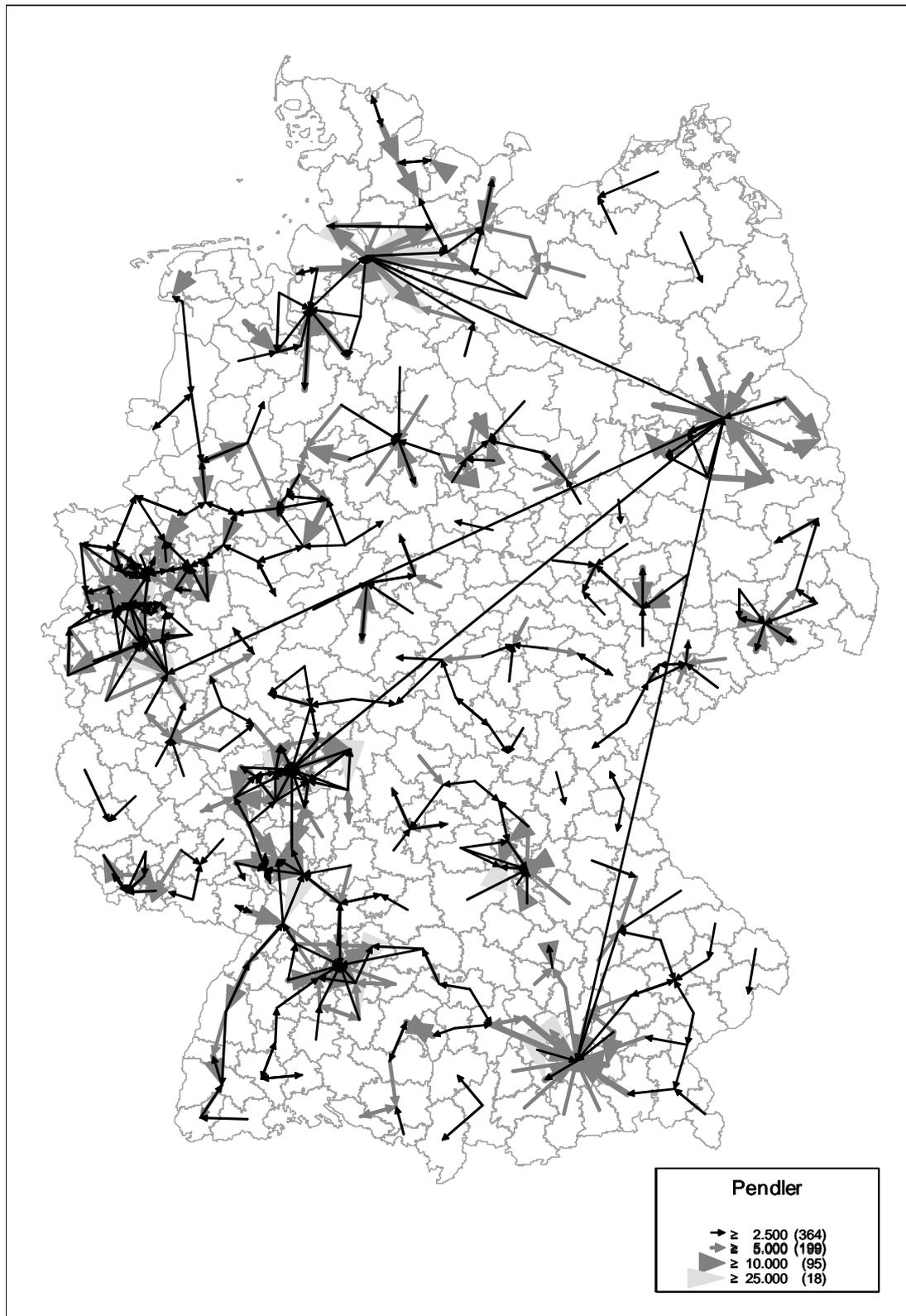
Die Abgrenzung von Arbeitsmarktregionen, sowohl in Deutschland als auch auf internationaler Ebene, basiert größtenteils auf Analysen zu Pendlerverflechtungen.² Um die verschiedenen Verfahren zur Abgrenzung von regionalen Arbeitsmärkten miteinander vergleichen zu können, wurde als einheitliche Datenbasis die Zahl der **sozialversicherungspflichtig Beschäftigten zum 30.06.2005** aus der Beschäftigtenstatistik der Bundesagentur für Arbeit auf der Ebene der Stadt- und Landkreise gewählt. Mit dieser Datenquelle werden 26 Millionen Beschäftigte, d.h. 75-80% aller Erwerbstätigen, erfasst. Die Meldungen der Arbeitgeber an die Sozialversicherungsträger gewährleisten eine eindeutige Verortung jedes Beschäftigten zu seinem Wohn- und Arbeitsort. Mit diesen Beschäftigtenangaben ist es möglich, die Pendlerverflechtungen zwischen den Regionen bis auf Gemeindeebene exakt abzubilden. Aus diesen Angaben lassen sich zwar keine direkten Schlüsse auf die Art des Pendeln ziehen, denn ob es sich um Tagespendler oder Wochenpendler handelt, und inwiefern der gemeldete Wohnsitz wirklich relevant ist, lässt sich aus den Daten nicht ableiten. Die weiter unten (unter Abbildung 3) präsentierten Pendelentfernungen lassen jedoch den Schluss zu, dass der deutlich größte Teil der Pendler Entfernungen im Tagespendelbereich zurücklegt.

Für die vorliegende Analyse wurden in Anlehnung an die Bundesverkehrswegeplanung (BVU/ifo/ITP/PLNCO 2001) die 439 Stadt- und Landkreise zu **377 Kreisregionen** zusammengefasst. Eine Kreisregion wurde immer dann gebildet, wenn eine kreisfreie Stadt nur von einem Landkreis umgeben war, ansonsten bildeten sie wie alle übrigen Kreise eigene Kreisregionen. Aus dieser Gliederung Deutschlands in 377 Regionen ergeben sich 141.752 ($=377 \cdot 377 - 377$) mögliche Pendlerbeziehungen, von denen 105.450 tatsächlich existieren, davon etwa die Hälfte mit weniger als 10 Pendlern und über 20.000 Pendlerbeziehungen mit nur einem Pendler.

Die bedeutendsten Pendlerströme, beginnend mit mehr als 2.500 Pendlern, sind in Abbildung 1 dargestellt. Am deutlichsten sind die Pendlerbewegungen zu den Ballungszentren Hamburg (über 38.000 aus Pinneberg und Harburg), Berlin (über 24.000 aus dem Kreis Oberhavel und rund 22.000 aus Barnim) und München (über 36.000 aus Fürstfeldbruck und über 26.000 aus Dachau) erkennbar. Sie üben offensichtlich eine hohe Attraktivität auf ihr Umland aus. Die stärksten Pendlerströme gibt es jedoch nach Köln (über 52.000 aus dem Erftkreis und über 36.000 aus Bonn), nach Stuttgart (ca. 44.000 aus Ludwigsburg und 32.000 aus dem Rems-Murr-Kreis) sowie nach Frankfurt (über 43.000 aus Offenbach und je rund 30.000 aus dem Main-Taunus- bzw. Main-Kinzig-Kreis).

² Einen anderen Ansatz verfolgt Bode (2008), der die Verflechtungen von Regionen anhand von Landpreisen bestimmt und damit einen Vorschlag zur Abgrenzung von Metropolregionen in West-Deutschland liefert.

Abbildung 1: Bedeutende Pendlerströme zwischen Kreisen 2005



Insgesamt gibt es aus 18 einzelnen Kreisregionen Pendlerströme von über 25.000 Arbeitskräften zu Arbeitsmarktzentren³. In Abbildung 1 sind diese stärksten

³ Die Zahl der Kreisregionen ist in der Legende in den Klammern wiedergegeben.

Pendlerströme als helle, breite Pfeile dargestellt. Darüber hinaus werden 95-mal Grenzen von Kreisregionen von mehr als 10.000 Personen auf dem Weg von ihrem Wohnort zu ihrem Arbeitsplatz überschritten. Auffällig sind ebenfalls die als schmale, dunkle Pfeile dargestellte weite Pendelstrecken, wie zwischen Hamburg, Bonn, Frankfurt, München und Berlin, die immerhin noch von mehr als 2.500 Pendlern zurückgelegt werden. Dabei handelt es sich jedoch vermutlich eher um Wochen- als Tagespendler.

Für die folgenden Analysen wurden die Pendlerdaten als **Pendlermatrix** aufbereitet. In der Matrix bilden die Zeilen die Wohnort- und die Spalten die Arbeitsortregionen. D.h., dass in jeder Zeile die Arbeitsorte aller Beschäftigten eines Wohnorts aufgelistet werden und in den Spalten die Wohnorte der Beschäftigten eines Arbeitsortes. In der nachfolgenden Tabelle 2 wird die Pendlermatrix exemplarisch für sechs Kreise in Schleswig-Holstein dargestellt.

Tabelle 2: Pendlermatrix

Arbeitsort Wohnort	Flensburg	Nordfries- land	Schleswig- Flensburg	Steinburg	Dithmar- schen	Kiel	...	Zeilen- summe
Flensburg	16.955	439	3.881	33	30	404		23.370
Nordfriesland	1.795	38.955	2.041	137	817	566		47.339
Schleswig- Flensburg	13.613	1.668	30.593	90	257	1348		54.648
Steinburg	46	89	65	23.261	1.324	449		40.246
Dithmarschen	94	1132	249	2.274	27.863	496		37.257
Kiel	277	107	259	117	98	50513		68.279
...								
Spaltensumme:	34.802	45.146	40.961	31.263	32.486	98.859		

Aus Flensburg pendelten im Jahr 2005 demnach 404 Personen nach Kiel, während es in die andere Richtung nur 277 waren. Die Zeilensumme ergibt die Anzahl aller in einer Region wohnenden Personen – sowohl der "Inländer" wie der Auspendler. Analog gibt die Spaltensumme alle in der Region arbeitenden Personen wieder, ungeachtet ihres Wohnortes. Deutlich wird schon bei den hier gezeigten Fällen die typische Funktion der Städte, auch das Umland mit Arbeitsplätzen zu versorgen (bzw. die Funktion des Umlandes als bevorzugter Wohnort). So wohnen in Kiel nur reichlich 68.000 Beschäftigte, während die Stadt Arbeitsplätze für knapp 100.000 bietet. Eine besondere Bedeutung hat die Diagonale der Matrix. Sie gibt an, wie viele Personen in der gleichen Region arbeiten und wohnen (Inländer).

Teilt man die Zellenwerte durch die Zeilensumme, so ergeben sich die Anteilswerte mit Ausprägungen zwischen 0 und 1 bzw. zwischen 0 und 100%. Die Anteilswerte – dargestellt in Tabelle 3 - zeigen, welcher Anteil aller in einer Region wohnenden Beschäftigten in jeder Region tätig ist.

Tabelle 3: Zeilenweise normalisierte Pendlermatrix (Auspendleranteile bzw. Anbindungskoeffizienten)

Arbeitsort Wohnort	Flensburg	Nordfriesland	Schleswig-Flensburg	Steinburg	Dithmarschen	Kiel	...	Zeilen- summe
Flensburg	73%	2%	17%	0%	0%	2%		100%
Nordfriesland	4%	82%	4%	0%	2%	1%		100%
Schleswig-Flensburg	25%	3%	56%	0%	0%	2%		100%
Steinburg	0%	0%	0%	58%	3%	1%		100%
Dithmarschen	0%	3%	1%	6%	75%	1%		100%
Kiel	0%	0%	0%	0%	0%	74%		100%

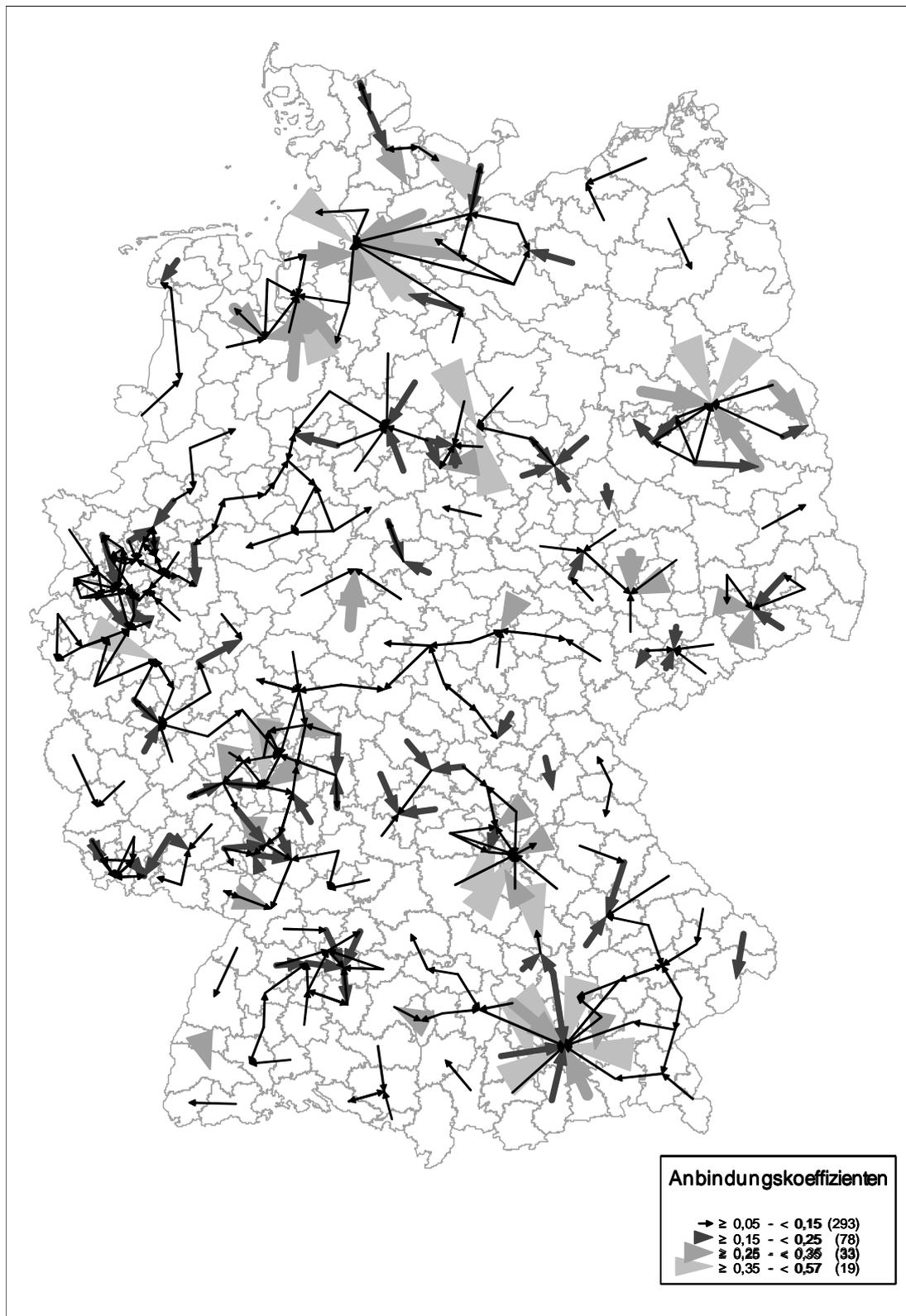
Die Werte auf der Diagonalen geben nunmehr die Autonomie einer Region wieder, d.h. inwiefern sie sich selbst mit Arbeitsplätzen versorgt. Nordfriesland zeigt eine hohe Autonomie, wie es für periphere Gebiete ebenso kennzeichnend ist wie für Arbeitsmarktzentren. Bei der vorliegenden zeilenweisen Normalisierung sind die Zeilenwerte außerhalb der Diagonale die Auspendleranteile. Weil sie angeben, wie stark die Wohnortregion über die Pendler mit der Arbeitsortregion verbunden ist, werden diese Werte auch als **Anbindungskoeffizienten** bezeichnet. Eine spaltenweise Normalisierung würde außerhalb der Diagonalen die Einpendleranteile ergeben und auf der Diagonale die Eigenversorgung mit Wohnungen.

Abbildung 2 zeigt die Auspendleranteile bzw. Anbindungskoeffizienten über 0,05 für die Kreisregionen in Deutschland. Diese Abbildung ist analog zu Abbildung 1 aufgebaut, wobei die Pfeile umso breiter und heller dargestellt sind je höher die Anbindungskoeffizienten sind. Die stärksten Anbindungskoeffizienten entsprechen durchaus nicht zwangsläufig den stärksten Pendlerströmen. Die rund 52.000 Pendler des Erftkreises nach Köln entsprechen eben "nur" einem reichlichen Drittel der im Erftkreis wohnenden Beschäftigten.⁴ Die stärksten Anbindungskoeffizienten weisen mit über 50% Ebersberg, Fürstfeldbruck und Dachau nach München sowie Osterholz nach Bremen auf. Die Analyse der Anbindungskoeffizienten rückt im Vergleich zur Analyse der absoluten Pendlerzahlen die Funktion von Regionen als überregionale

⁴ Das ist der 18.größte Anbindungskoeffizient. Siehe Anhang 1 für die 20 größten Pendlerströme und Anbindungskoeffizienten.

Arbeitsmarktzentren in den Mittelpunkt. Neben den bisher identifizierten bedeutendsten Pendlerströmen erscheinen nunmehr mit Anbindungskoeffizienten von über 0,35 Arbeitsorten wie Bremen, Wolfsburg, Ingolstadt, Nürnberg und Kiel als solche Zentren.

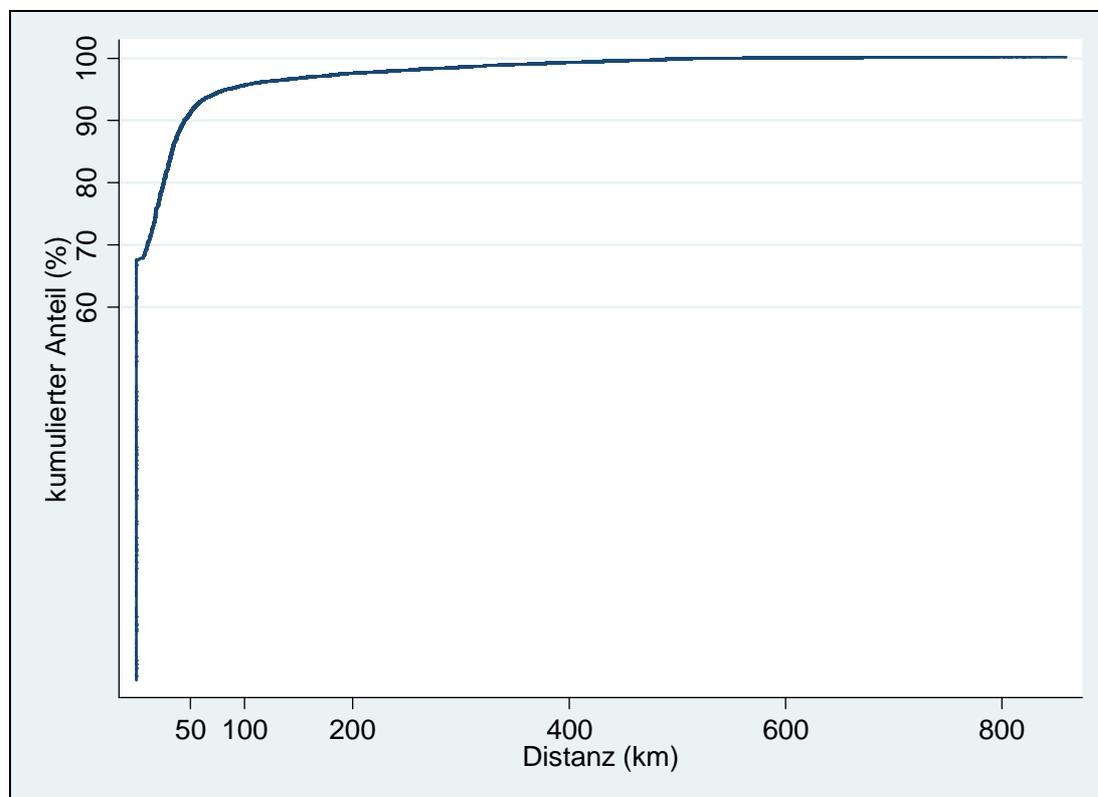
Abbildung 2: Anbindungskoeffizienten



Bereits Abbildung 1 und Abbildung 2 lassen die Identifizierung wesentlicher Merkmale der Pendlerbewegungen zu, die jede zuverlässige Abgrenzungsmethode widerspiegeln muss: München, Hamburg, Berlin, Frankfurt, Düsseldorf und Köln sind wichtige Zentren, während bei zahlreichen Regionen zwischen ihnen die Zuordnung nicht so deutlich ist. Ob es sich bei Regionen, die bei den ersten Analysen nicht mit anderen verflochten erscheinen, eher um autonome oder isolierte Regionen handelt, ist ebenfalls eine Frage, auf die vertiefende Analysen Antwort geben müssen.

Für eine Reihe von Berechnungen ist nicht nur das Pendleraufkommen zwischen Regionen, sondern auch deren **Entfernung** relevant. Verwendet wurden für die Entfernungsberechnung die Koordinaten⁵ der einwohnerreichsten Gemeinde jeder Kreisregion. Betrachtet man den Zusammenhang zwischen Pendleraufkommen und Pendelentfernung, so relativiert sich das eingangs angesprochene Problem, dass wir keine Informationen über das konkrete Pendlerverhalten haben. Über zwei Drittel der Beschäftigten arbeiten in der gleichen Kreisregion in der sie wohnen, für über 90% liegt die Pendelentfernung unter 50 km und nur reichlich 4 % der Pendler pendeln mehr als 100 km, so dass man vermuten kann, dass es sich bei den vorhandenen Daten zum weitaus größten Teil um Tagespendler handeln dürfte, wie es für die Abgrenzung von Arbeitsmarktregionen am sinnvollsten ist. Abbildung 3 verdeutlicht diesen Zusammenhang.

Abbildung 3: kumulierte Häufigkeiten der Pendlerentfernungen



4 Abgrenzungskriterien

Bevor im nächsten Abschnitt näher auf die verschiedenen Abgrenzungsverfahren eingegangen wird und diese beurteilt werden können, sollen zunächst einige grundlegende und häufig verwendete Kriterien für die Abgrenzung von Arbeitsmarktregionen diskutiert werden:

- **Kontingenz:** Die Arbeitsmarktregionen sollen räumlich zusammenhängend sein, müssen in ihrer Summe ein Gebiet (z. B. das Bundesgebiet) flächendeckend abdecken und dürfen sich nicht überschneiden.
- **Geschlossenheit:** Die meisten Arbeitsplätze innerhalb der Arbeitsmarktregion sollen von der dort wohnenden Bevölkerung aufgesucht werden bzw. die meisten Beschäftigten innerhalb der Region arbeiten. D.h. es sollten starke Pendlerverflechtungen innerhalb der Arbeitsmarktregion bestehen und schwache Pendlerverflechtungen zwischen den Regionen. Die Geschlossenheit einer Region wird durch entsprechende Schwellenwerte definiert, die zwischen 65% und 75% liegen. Beispielsweise mussten bei Binder und Schwengler (2006) mindestens 65% der Bevölkerung in der gleichen Region arbeiten und 65% der Beschäftigten in der gleichen Region wohnen, um eine Arbeitsmarktregion zu bilden. Höhere Schwellenwerte von 70% bzw. 75% verwenden hingegen Hensen und Cörvers (2003) bzw. van der Laan und Schalke (2001) für die Niederlande oder Casado-Díaz (2000) für Spanien.
- **Mindestgröße:** In der Regel wird von einer bestimmten Mindestgröße ausgegangen, die sich an der Zahl der Einwohner orientiert. Lag diese Mindestgröße Anfang der 1970er Jahre noch bei 30.000 Einwohnern (Klemmer/Kraemer 1975), erhöhte sich diese bis auf 100.000 (Eckey 2001, Binder/Schwengler 2006). Granato und Farhauer (2007) argumentieren mit Moseley (1974), Hansen (1975) und Boudeville (1966), dass Regionen eine bestimmte Mindestgröße aufweisen müssen, um als potentieller Wachstumspool zu fungieren. Nach Eckey/Klemmer (1991: 574) sei dafür eine Mindestgröße von 50.000 Einwohnern nötig, während überdurchschnittliche Wachstumsraten erst ab einer Größe von 250.000 Einwohnern zu erwarten seien (Junckernheinrich/Skopp 1995: 226).
- **Maximale Ausdehnung:** Die Schwellenwerte für die maximale Ausdehnung von Arbeitsmarktregionen wurden im Laufe der Zeit an die veränderten Pendelgewohnheiten und –entfernungen angepasst. Die Pendelentfernungen werden dabei in der Regel durch Pendelzeiten anstelle der räumlichen Entfernung festgelegt. Wurde Anfang der 1970er Jahre in Deutschland noch eine maximale Pendelzeit von 45 Minuten und eine Mindesteinwohnerzahl von 30.000 bis 40.000 für die Abgrenzung regionaler Arbeitsmärkte zugrunde gelegt (Klemmer/Kraemer 1975), erhöhten sich diese Zahlen in der jüngsten Abgrenzung auf eine einfache Pendelzeit bis zu 60 Minuten und eine Mindesteinwohnerzahl von 50.000 (Eckey u.a. 2006) .

⁵ Die Gemeindekoordinaten (Stand 31.12.2005) wurden uns freundlicherweise vom Bundesamt für Bauwesen und Raumordnung zur Verfügung gestellt.

- **Korrespondenz zu administrativen Gliederungen:** Die kreis- und ggf. bundslandscharfe Abgrenzung der Regionen ist der in der Regel notwendig, um einen Bezug zu den politischen Akteuren herzustellen oder um für regionalwissenschaftliche Analysen notwendige Daten heranziehen zu können, die häufig nur auf Kreisebene vorliegen.

Für die vorliegende Studie sollen die letzten Kriterien nicht a priori gesetzt werden. Gegen die Festsetzung einer Mindestgröße spricht, dass es ggf. gerade die mangelnde Größe einer Region ist, die ihre wirtschaftliche Entwicklung behindert. Die Identifizierung solcher "zu kleinen" Wirtschaftsräume wäre dann ein wichtiges Ergebnis raumökonomischer Forschung und ein wichtiges Argument für die bessere infrastrukturelle Anbindung an ein benachbartes Wirtschaftszentrum.

Auch die Setzung von maximalen Pendelzeiten ist aus inhaltlichen Gründen problematisch. Zum einen variieren die Pendelzeiten stark in Abhängigkeit von regional unterschiedlichen Immobilienpreisen (insbesondere ihrem Zentrum-Umland-Gefälle), der vorhandenen Infrastruktur, den zu erzielenden Löhnen und der lokalen Arbeitsmarktsituation. So zieht zum Beispiel der florierende Münchner Arbeitsmarkt mit hohen Löhnen Beschäftigte aus weit entfernten Regionen an⁶, während die hohen Immobilienpreise verhindern, dass der Arbeitsmobilität eine entsprechende Wohnmobilität folgt. Andererseits zwingt die schwierige Arbeitsmarktlage in vielen Regionen Ostdeutschlands Arbeitskräfte zur Aufnahme einer Beschäftigung in weit entfernten Regionen (vgl. Bundesministerium für Raumordnung, Bauwesen und Städtebau 2005: 78 ff). All diese Faktoren fließen zweifellos in die Entscheidungen über Arbeits- und Wohnort der Beschäftigten ein und sollen darum nicht a priori eingeschränkt werden.⁷

Schließlich soll auch die Korrespondenz zu administrativen Gliederungen nicht vorausgesetzt werden. Die bundslandscharfe Abgrenzung ist insbesondere für die Stadtstaaten problematisch. So erstreckt sich beispielsweise der Hamburger Wirtschaftsraum weit nach Schleswig-Holstein und Niedersachsen. In der Metropolregion Hamburg lebten allein 2005 44% bzw. 16% der Bevölkerung Schleswig-Holsteins und Niedersachsens und es wurden hier 43% bzw. 13% des Bruttoinlandsproduktes der Länder erwirtschaftet (Kropp/Niebuhr 2006: 14, 16). Für Bremen und Berlin gilt, wenngleich weniger ausgeprägt, ähnliches. Problematisch sind bundslandscharfe Abgrenzungen auch hinsichtlich einiger grenzüberschreitender Wirtschaftsräume wie dem Rhein-Main- und dem Rhein-Neckar-Dreieck. Darum werden die Abgrenzungen in der vorliegenden Studie bundeslandübergreifende

⁶ Zur innerdeutschen Pendlerstruktur unterschiedlicher Qualifikationsgruppen siehe z.B. Haas/Hamann (2006).

⁷ Nach dem Dritten Buch des Sozialgesetzbuches III beträgt die zumutbare einfache Pendelzeit für Vollzeitbeschäftigte 75 Minuten (§ 121 Abs. 4 SGB III). Eckey u.a. 2006 wählen je nach Attraktivität einer Arbeitsmarktregion Pendelzeiten von 45-60 Minuten zur räumlichen Begrenzung ihrer Arbeitsmarktregionen.

Regionalisierungen zulassen. Um den Vergleich mit vorhandenen Regionalisierungen zu gewährleisten, werden die Abgrenzungen jedoch kreisscharf erfolgen.

Anhand der Ergebnisse der vorliegenden Studie werden die oben genannten Schwierigkeiten besser einschätzbar und es wird möglich anzugeben, für welche Regionen die Anwendung von Restriktionen wie Mindestgröße, maximale Ausdehnung oder Korrespondenz zu administrativen Gliederungen problematisch ist.

5 Abgrenzungsverfahren

Neben den beschriebenen Abgrenzungskriterien spielt die Wahl für ein bestimmtes Abgrenzungsverfahren eine entscheidende Rolle für den Zuschnitt von Arbeitsmarktregionen. Das Methodenspektrum reicht vom Einsatz relativ einfacher Schwellenwertmodelle (ONS/Coombes 1998; Casado-Díaz 2000; Karlsson/Olsson 2006) über Clusteranalyse (Tolbert/Killian 1987; Bongaerts u.a. 2004) bis hin zur Faktorenanalyse (Eckey u.a. 2006). Einen weitergehenden Überblick geben Goetz (1999), OECD (2002), und van Nuffel (2007). Van Nuffel bietet zudem ebenso wie van der Laan und Schalke (2001) eine sehr informative Systematisierung der Methoden. Die vorliegende Arbeit wird die Varianten dieser Methoden anhand der Daten über Pendlerverflechtungen in Deutschland untersuchen.

Die Bewertung der zentralen Eigenschaften der so erfolgten unterschiedlichen Abgrenzungen wird über die verschiedenen Kriterien erfolgen, die vor allem die Kontingenz, d.h. die Geschlossenheit der Arbeitsmarktregionen prüfen. Maximale Geschlossenheit wäre erreicht, wenn ein Maximum an Pendleraktivitäten innerhalb der Arbeitsmarktregionen erfolgt und ein Minimum zwischen ihnen. In dieser generellen Form ist dieses Kriterium jedoch nicht sinnvoll anwendbar, wie weiter unten (S. 46) im Zusammenhang mit der Gesamtpendlerquote detailliert ausgeführt wird. Newmans Modularitätsansatz (Newman/Girvan 2004) bietet ein leistungsfähigeres Kriterium zur Einschätzung, ob eine Arbeitsmarktregionen-Abgrenzung die tatsächlichen Pendlerbewegungen gut "einfängt". Dieser Ansatz hat in den letzten Jahren eine große Popularität in Disziplinen erlangt, die sich mit der Identifizierung von Gruppierungen in Netzwerken beschäftigen (Brandes u.a. 2008). Die Grundidee besteht darin zu messen, um wie viel besser eine Abgrenzungsmethode beim "Einfangen" existierender Gruppierungen ist als eine zufällig generierte Einteilung mit derselben Anzahl von Gruppen (in unserem Fall Arbeitsmarktregionen) und der gleichen Pendlerstruktur. Die Berechnung basiert auf einer symmetrischen Matrix, die (ähnlich wie die Pendleranteilematrix) in den Clustern enthaltenen Einheiten auf der Diagonale und die Verflechtungen zwischen den Clustern außerhalb der Diagonale als Anteil aller Einheiten einer Zeile bzw. Spalte enthält. Die Differenzen der Einheit in den Clustern i (e_{ii}) zum Quadrat der Anzahl der Einheiten in den Spalten (oder Zeilen) $a_i = \sum_j e_{ij}$ werden dann zur Modularität $Q = \sum_i (e_{ii} - a_i^2)$ aufsummiert. Wenn die Einteilung in Cluster nicht besser als eine Zufallseinteilung ist, ist $Q=0$. Werte, die sich dem Maximum von 1 nähern, weisen auf eine stark modulare Struktur hin, wie durch das Clusterverfahren korrekt erfasst wurde. Üblich sind Werte zwischen 0,3 und 0,7. Höhere Werte sind selten, aber gute Abgrenzungsverfahren in der vorliegenden Studie werden Ergebnisse

mit Werten über 0,8 erzeugen.⁸ Für die in Tabelle 1 aufgeführten administrativen und funktionalen Gliederungen gibt Tabelle 4 die entsprechenden Modularitätswerte für die Pendlerverflechtungen vom 30.06.2005 wieder. Danach erreicht die Abgrenzung der regionalen Arbeitsmärkte von Eckey u.a. (2006) mit 0,81 den höchsten Modularitätswert im Vergleich.

Tabelle 4: Modularitätswerte für administrative und funktionale Gliederungen

Gliederung	Modularität Q
Kreise	0,67
Raumordnungsregionen	0,79
Arbeitsmarktregionen der Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur“	0,75
Regionale Arbeitsmärkte (Eckey u.a. 2006)	0,81

5.1 Schwellenwertmodelle

Eine Gruppe von Verfahren fusioniert Regionen sobald die Pendlerverbindung einen bestimmten Schwellenwert (Kennziffer) übersteigt. Grundlage dafür sind in der Regel die Anbindungskoeffizienten (siehe Abschnitt 3). Im Folgenden wird ein solches Verfahren näher vorgestellt.

Konzept der dominanten Flüsse/Graphentheorie

Das Konzept der dominanten Flüsse (Nystuen/Dacey 1961) geht von einer Anzahl von Knoten, den regionalen Einheiten, aus, die durch unterschiedlich starke Zu- und Abflüsse, die Pendlerströme, verbunden sind. Jeder Knoten ist mit einer Hauptvernetzung, dem dominanten Pendlerstrom, mit je einem weiteren Knoten verbunden. Die Zielregion jedes maximalen Pendlerstroms wird der jeweils höheren Hierarchieebene zugerechnet, wobei die dominanten Pendlerströme anhand eines bestimmten Schwellenwerts festgelegt werden.

Auf Basis der **377** Kreisaggregate wurden so Agglomerationen mit Hilfe der Graphentheorie gebildet. Dafür wurden entsprechend dem Konzept der dominanten Flüsse sowohl die Pendlernetzungen als auch die Struktur der Kreise berücksichtigt. Die Identifizierung der räumlichen Struktur der Pendlerströme erfolgte anhand eines hierarchisch geordneten Graphen über die maximalen Pendlerströme aus einem Kreis in einen anderen Kreis. Die so gebildeten Einzugsgebiete, die Bassins, setzen sich aus einer Anzahl von Kreisen zusammen, den Knoten, die wiederum über eine Baumstruktur, den Graphen, hierarchisch miteinander verbunden sind. Basis für die Bildung der Bassins ist die Festlegung eines Schwellenwerts für die eine Mindestanzahl an Beschäftigten als auch an dominanten Flüssen.

⁸ Newman nutzt den Modularitätsansatz um Algorithmen zu entwickeln, die Partitionierungen nach diesem Kriterium optimieren (Newman 2006). Diese sind jedoch noch nicht für gewichtete Graphen wie die Pendlermatrix anwendbar.

Um einen Vergleich mit der aktuellen Abgrenzung von Eckey u.a. (2006) zu ermöglichen, wurde versucht, ebenfalls 150 Arbeitsmarktregionen mit dem Konzept der dominanten Flüsse zu bilden. Durch die Festlegung verschiedener Schwellenwerte und Hierarchiestufen konnte so die Zahl der Arbeitsmarktregionen annähernd erreicht werden.

In einem Fall ließen sich 151 Bassins nach der ersten Hierarchiestufe über einen Schwellenwert von 5% ermitteln (H: 1; SW: 5). In dem zweiten Fall konnten 148 Arbeitsmarktregionen nach zwei Hierarchiestufen bei einem Schwellenwert von 7,5% gebildet werden (H: 2; SW: 7,5).

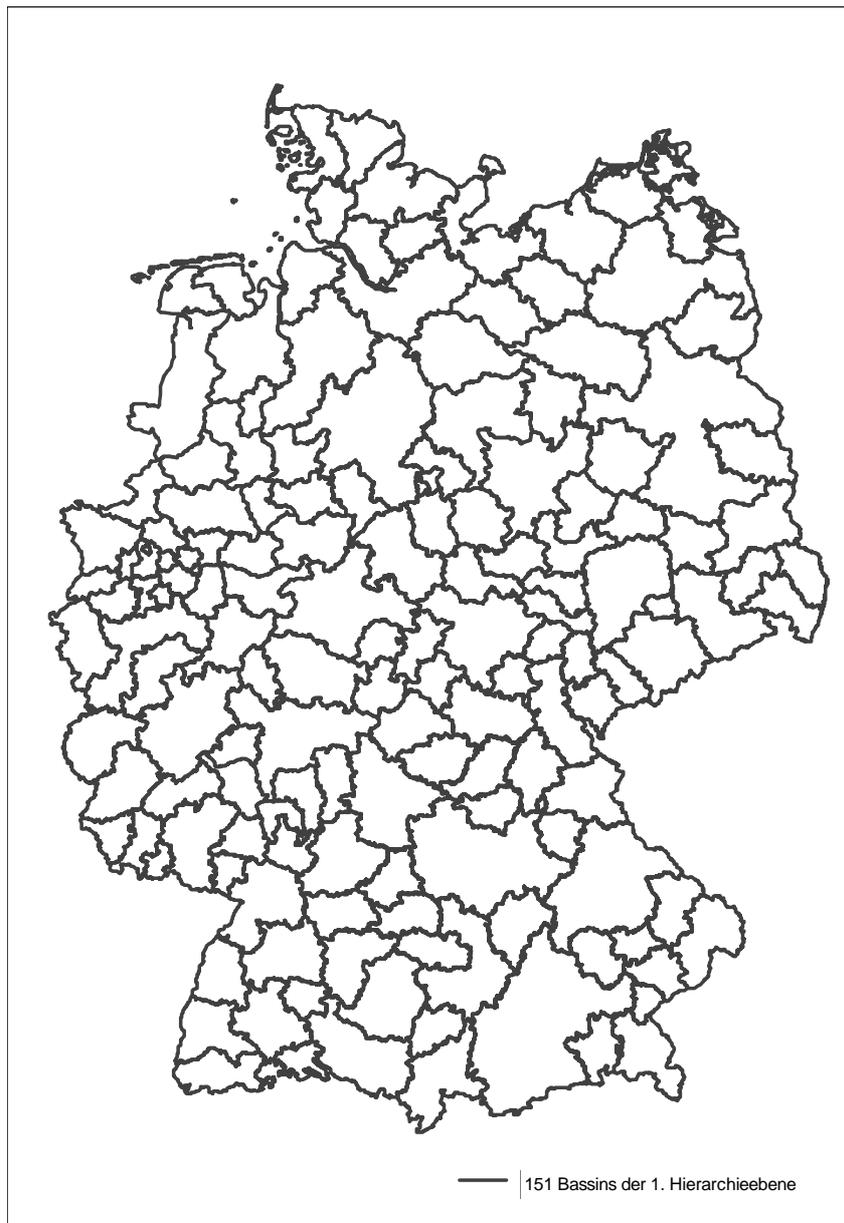
Die Vorgehensweise zur Identifizierung der Bassins verlief in folgenden Schritten:

1. Fall: Bildung von 151 Bassins bei einem Schwellenwert von 5% nach der ersten Hierarchiestufe

1. Zunächst wurde ein Kreis (Wohnort) einem anderen Kreis (Arbeitsort) zugeordnet, wenn der maximale Pendlerstrom aus diesem Kreis (Wohnort) in den entsprechenden Kreis (Arbeitsort) gerichtet war und mehr als 5% der Beschäftigten aus dem Wohnort in diesen Arbeitsort pendelten. Es ließen sich so insgesamt 278 dominante Pendlerströme identifizieren.
2. Dominierte ein Kreis (Arbeitsort) mehrere Kreise (Wohnorte), so wurden all diese Kreise zu einem Bassin zusammengefasst. Insgesamt waren dies 115 Bassins, die sich aus 2 bis 14 Kreisen zusammensetzen und durch signifikante Pendlerströme miteinander verbunden sind.
3. 36 Kreise bildeten eigene geschlossene Einheiten ohne dominante Pendlerströme zu benachbarten Kreisen.

Die so entstandenen 151 Bassins haben eine Mindestgröße von 23.478 Beschäftigten bzw. 72.545 Einwohnern. Die Modularität Q beträgt bei dieser Variante 0,783.

Abbildung 4: 151 Bassins bei einem Schwellenwert von 5% nach dem Konzept der dominanten Flüsse

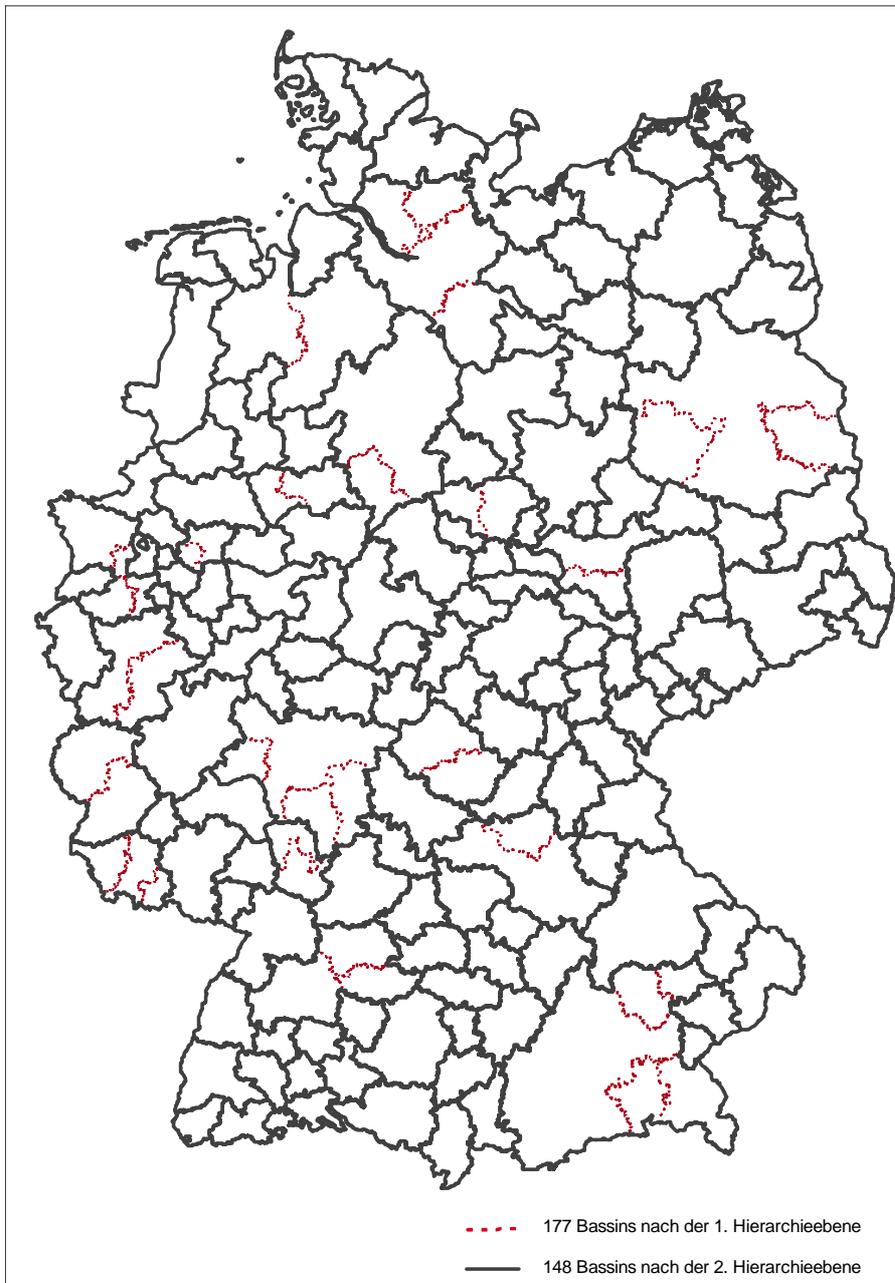


2. Fall: Bildung von 148 Bassins bei einem Schwellenwert von 7,5% nach zwei Hierarchiestufen

1. Auch in diesem Fall wurde ein Kreis (Wohnort) dann einem anderen Kreis (Arbeitsort) zugeordnet, wenn der maximale Pendlerstrom, der aber hier **mindestens 7,5% der Beschäftigten** betragen musste, aus dem Wohnortkreis in den Arbeitsortkreis gerichtet war. Mit diesem höheren Schwellenwert ergaben sich insgesamt **232 dominante Pendlerströme**.
2. Es ließen sich so **101 Bassins** aus 2 bis 15 Kreisen bilden, indem alle dominierenden Arbeitsortkreise mit den entsprechenden Wohnortkreisen zusammengefasst wurden.

3. Gab es in einer zweiten Hierarchieebene weitere dominante Ströme, d.h. der abhängige, dominierte Kreis dominierte selber auch einen anderen Kreis, so erfolgte hier ebenfalls eine Zuordnung dieses Kreises zu dem ursprünglichen Bassin. Insgesamt entstanden so **71 Bassins** unter Einbeziehung von zwei Hierarchieebenen.
4. Ohne dominante Pendlerströme von 7,5% der Beschäftigten oder mehr zu anderen Kreisen verblieben **77 Kreise** als eigenständige Regionen.

Abbildung 5: 148 Bassins bei einem Schwellenwert von 7,5% nach dem Konzept der dominanten Flüsse



Da in dieser Abgrenzung mehr als doppelt so viele Kreise eigenständige Bassins bilden, verringert sich hier auch die Mindestgröße der so entstandenen 148 Bassins deutlich und beträgt nunmehr 8.911 Beschäftigte und 51.352 Einwohner. Die Modularität steigt in diesem Fall sogar auf 0,814 und liegt damit höher als die erste Variante mit einem Schwellenwert von 5% bei Einbeziehung nur einer Hierarchiestufe. Jedoch unterscheidet sich bei diesem höheren Schwellenwert die Größe der so entstandenen Arbeitsmarktregionen deutlich stärker voneinander als bei der vorangegangenen Abgrenzung (vgl. Abbildung 4 mit Abbildung 5). Insbesondere die Arbeitsmarktregionen um Ballungszentren mit einem hohen Pendlereinzugsbereich wie Berlin, München, Hamburg, Hannover und dem Rhein-Main-Gebiet haben sich beträchtlich vergrößert.

5.2 Clusteranalyse

Die Clusteranalyse ist eines der klassischen Verfahren zur Gruppenbildung von Beobachtungen. Es geht "immer um die Analyse einer *heterogenen Gesamtheit von Objekten* ... mit dem Ziel, *homogene Teilmengen von Objekten* aus der Objektgesamtheit zu ermitteln" (Backhaus u.a. 2000: 329, Hervorhebung im Original). Bei der Beurteilung der Zusammengehörigkeit kann dabei mehr als ein Merkmal berücksichtigt werden. Für die Clusteranalyse sind zwei wesentliche Entscheidungen zu treffen – nämlich die Wahl des Proximitätsmaßes, und des Fusionierungsalgorithmus.

Das Proximitätsmaß entspricht einer Ähnlichkeits- oder Distanzmatrix hinsichtlich der Beobachtungen, für deren Berechnung eine Reihe von Verfahren zur Verfügung steht. Die Distanz- oder Ähnlichkeitsmatrix dient im zweiten Schritt als Basis für die Fusionierungsalgorithmen, die die eigentliche Zusammenfassung der Beobachtungen – in unserem Fall der Kreisregionen – zu Gruppen vornehmen. Unter den Clusterverfahren sind die partitionierenden und die hierarchischen die wichtigsten.

Bei **partitionierenden** Verfahren wird versucht, eine vorgegebene Ausgangsgruppierung (Partition) durch Umgruppierung einzelner Objekte zu verbessern (vgl. Backhaus u.a. 2000: 348ff). Da für die vorliegende Problemstellung keine sinnvolle apriori-Partition besteht und die Algorithmen nur lokale, aber keine globale Optima berechnen können, soll auf diese Verfahren nicht zurückgegriffen werden.⁹

Hierarchische Algorithmen haben stets den folgenden Ablauf (vgl. Backhaus u.a. 2000: 353 ff.):

- a) die Wahl einer Ähnlichkeits- oder Distanzmatrix,
- b) die Fusion der ähnlichsten Beobachtungen,
- c) die Berechnung einer reduzierten Ähnlichkeits- oder Distanzmatrix und

⁹ Prinzipiell wäre es möglich, partitionierende Verfahren einzusetzen, um eine vorhandene hierarchische Clusterlösung zu verbessern. Ein Beispiel dafür zeigen Blien/Hirschenauer 2005. Ziel der vorliegenden Arbeit ist jedoch die Identifizierung optimaler Abgrenzungsverfahren, die auch ohne weitere Optimierung gute Ergebnisse liefern können.

d) die Entscheidung über den Abbruch des Cluster-Verfahrens bzw. die optimale Anzahl von Clustern.

Diese Schritte werden im Folgenden ausführlicher erläutert:

a. Ähnlichkeits- oder Distanzmatrix

Wie die Ähnlichkeit von Eigenschaften gemessen wird, hängt in starkem Maße von den Ausgangsdaten und der Forschungsfrage ab. Im Falle der **Pendlerdaten** kann man bereits von einer Ähnlichkeitsmatrix sprechen, wenn man davon ausgeht, dass Regionen mit ähnlichem Pendlermuster zu einer Gruppe gehören. Häufig verwendet werden in Clusteranalysen **Ähnlichkeitsmaße** wie die **Minkowski-Metriken** (oder L-Normen, vgl. Backhaus u.a. 2000: 340ff) oder **Q-Korrelationskoeffizienten**¹⁰. Prinzipiell sind aber auch andere Maße denkbar, wenn sie Daten und Forschungsfrage besser entsprechen.

Um geeignete Matrizen für die Clusteranalyse zu erhalten, werden in Untersuchungen zur Identifizierung von Arbeitsmarktregionen zum Teil sehr komplexe Matrixoperationen angewandt. Das Ziel dabei ist, homogene Randverteilungen zu erreichen (z.B. Zeilen- und Spaltensummen mit dem Wert 1 für alle Zeilen oder Spalten), und dabei die strukturellen Eigenschaften der Matrix zu erhalten. Bongaerts u.a. (2004) und Baumann u.a. (1983) verwenden dafür den **MFPT-Algorithmus** (**M**ean **F**irst **P**assage **T**ime). Baumann u.a. (1983) vergleichen die Resultate explizit mit dem **IPF-Verfahren** (**I**terative **P**roportional **F**itting), das auch in der vorliegenden Studie zur Anwendung kommt.

Im Folgenden werden die Ausgangsmatrizen für die Clusteranalysen beschrieben und anschließend die Ergebnisse in Tabelle 7 zusammengefasst dargestellt. Unterschieden werden dabei die vier Blöcke A bis D.

Im **Block A** sind die Ausgangsdaten als Matrizen zusammengefasst, die die originalen Pendlerhäufigkeiten wiedergeben, wobei in separaten Spalten verschiedene Varianten aufgeführt sind. In der ersten Variante wurde die Matrix symmetrisiert, in einer weiteren die Diagonalwerte gleich Null gesetzt und die dritte Variante enthält beide Manipulationen. Die Werte dieser Matrizen bilden die Grundlage für die Berechnung der folgenden symmetrischen Distanzmatrizen, sind aber selbst nicht für die Clusteranalyse geeignet, weil das die Clusteranalyse mit dem Statistikprogramm Stata spezifische Anforderungen die zu verwendenden Matrizen stellt.¹¹

Die Werte in den Matrizen von **Block B** ergeben sich aus der Division der Zellenwerte der Ausgangsmatrizen durch den größten dort vorkommenden Wert (matrizenweise Normalisierung), durch die Zeilensumme oder durch die Spaltensumme. Bei der

¹⁰ Der Q-Korrelationskoeffizient misst im Gegensatz zum "normalen" Korrelationskoeffizienten nicht den Zusammenhang zwischen Merkmalen (Variablen) sondern zwischen Objekten (Beobachtungen). Voraussetzung für die Berechnung ist die Standardisierung der Merkmale (Eckey u.a. 2002: 215).

¹¹ Die Clusteranalysen wurden mit der clustermat-Prozedur in Stata durchgeführt. Sie erfordert die Verwendung symmetrischer Distanzmatrizen. Die Diagonalwerte werden bei der Clusteranalyse ignoriert (vgl. StataCorp 2005).

zeilenweisen Normalisierung ergibt sich als neue Zeilensumme der Wert 1, bei der Spaltenweisen beträgt die Spaltensumme 1. Die Werte für die symmetrischen Varianten sind zwangsläufig gleich. Über ein iteratives Verfahren wurde eine zeilen- und spaltenweise Normalisierung durchgeführt, nach der sowohl die Zeilen- wie auch die Spaltensummen jeweils 1 ergeben.¹² Diese verschiedenen Formen der Normalisierung reduzieren die Unähnlichkeit zwischen Beobachtungen, die allein dem Größeneffekt der Regionen geschuldet ist. Zugleich gewährleisten sie, dass alle Zellenwerte zwischen 0 und 1 liegen, so dass durch die Subtraktion von 1 Distanzmatrizen gebildet werden können. Tabelle 5 zeigt eine Auswahl der Matrix für die zeilenweise Normalisierung.

Tabelle 5: Die aus der zeilenweise normalisierten Pendlermatrix gebildete Distanzmatrix für die Clusteranalyse

Arbeitsort Wohnort	Flensburg	Nordfries- land	Schleswig- Flensburg	Steinburg	Dithmar- schen	Kiel
Flensburg	0	0,97	0,79	1	1	0,99
Nordfriesland	0,97	0	0,96	1	0,98	0,99
Schleswig- Flensburg	0,79	0,96	0	1	0,99	0,99
Steinburg	1	1	1	0	0,95	0,99
Dithmarschen	1	0,98	0,99	0,95	0	0,99
Kiel	0,99	0,99	0,99	0,99	0,99	0

Der Vergleich mit Tabelle 3 verdeutlicht die für die Clusteranalyse notwendigen Matrixoperationen (siehe auch Fußnote 11). Am Auffälligsten ist sicherlich, dass hohe Werte jetzt nicht mehr eine starke Pendlerverflechtung sondern gerade deren Abwesenheit signalisieren. Zum anderen sind die Diagonalwerte gleich Null gesetzt, weil sie im Clusterprozess nicht berücksichtigt werden. Auch die Symmetrisierung der Matrix wird deutlich: Die Pendlerbeziehung zwischen Schleswig und Flensburg ist nunmehr nicht mehr durch zwei Werte gekennzeichnet (25% bzw. 17%) sondern durch einen Wert ($0,79 [=1-((0,25+0,17)/2)]$).

Im **Block C** folgen verschiedene Zusammenhangsmaße. Der Korrelationskoeffizient misst die Ähnlichkeit zwischen Spalten (üblicherweise Variablen), d.h. im vorliegenden Fall der Auspendlerstruktur. Weil die Clusteranalyse positive Werte erfordert, im vorliegenden Fall jedoch auch geringfügig negative Korrelationen auftreten, wurden die

¹² Hierfür kamen die entsprechenden Prozeduren im Softwarepaket ucinet (Borgatti et al. 2002) zur Anwendung.

Korrelationskoeffizienten auf Werte zwischen Null und ihrem Maximum normalisiert.¹³ Um zu prüfen, ob die Berücksichtigung der Diagonalwerte nachteilig ist, wurde für die Korrelationskoeffizienten auch für eine Pendlermatrix ohne Diagonalwerte berechnet. Weitere Zusammenhangsmaße sind der Q-Korrelationskoeffizient und die Winkeldistanz (angular separation). Der Q-Korrelationskoeffizient gibt – analog zum Korrelationskoeffizienten – die Ähnlichkeit zwischen den Zeilen und somit der Einpendlerstruktur der Regionen wieder. Aus der Perspektive einer Winkelbetrachtung entspricht der (Q-)Korrelationskoeffizient dem Kosinus des Winkels zweier Vektoren, der vom Mittelwert gemessen wurde (Gordon 1999; vgl. StataCorp 2005: 361f). Wird dieser Winkel stattdessen von Null aus gemessen, ergibt sich die Winkeldistanz. Auch zu den Zusammenhangsmaßen wurde durch Subtraktion der Werte von Eins und Symmetrisierung ein entsprechendes Distanzmaß gebildet.

Schließlich wurden auch die bei Clusterverfahren üblichen Distanzmaße der Minowski-Metriken gebildet (**Block D**).

Ausgangspunkt der eigentlichen Clusteranalyse ist die die *Ähnlichkeits- oder Distanzmatrix* mit der feinsten Partition: Jedes Objekt (jede Kreisregion) stellt ein Cluster dar (erster Schritt der Clusteranalyse).

b. Fusion

Beim nächsten Schritt der Clusteranalyse erfolgt die Auswahl und *Fusion* der beiden Objekte mit der geringsten Distanz bzw. größten Ähnlichkeit. Beim in Tabelle 5 wiedergegebenen Beispiel haben Flensburg und Schleswig-Flensburg den geringsten Distanzwert und würden im ersten Clusterschritt fusioniert.

c. Neuberechnung der reduzierten Ähnlichkeits- oder Distanzmatrix

Nach der Fusion zweier Regionen ist die Neuberechnung der reduzierten Ähnlichkeits- oder Distanzmatrix notwendig. Hierfür gibt es eine Reihe von Verfahren. Die bekanntesten sind **Single Linkage, Complete Linkage, Average Linkage (gewichtet und ungewichtet), Zentroid, Median und Ward** (siehe Tabelle 6). Als Illustration kann das Beispiel einer Entfernungstabelle zwischen Städten dienen. Schrittweise werden die zwei am Nächsten liegenden Städte zusammengefasst. Die Berechnung der Distanzen zwischen den jeweils neuen "Mega-Städten" und den verbleibenden Städten erfolgt beim Single-Linkage-Verfahren durch die Auswahl der Entfernung zur nächst gelegenen der fusionierten Städte (darum auch Nearest-Neighbour-Verfahren genannt), beim Complete-Linkage-Verfahren wird die größere Entfernung gewählt (darum auch Furthest-Neighbour-Verfahren) und beim Average-Linkage-Verfahren die gemittelte

¹³ Berechnung: $C^* = \frac{C + |C_{MIN}|}{1 + |C_{MIN}|}$, mit C - alter Korrelationskoeffizient, C* - neuer Korrelationskoeffizient, C_{MIN} –kleinster Wert für C. Während der Zähler für Werte >= 0 sorgt, reduziert der Nenner den Koeffizienten auf Werte <=1.

Entfernung. Tabelle 6 zeigt in Anlehnung an Backhaus u.a. (2000: 366) die wichtigsten Eigenschaften der Neuberechnungsverfahren:

Tabelle 6: Eigenschaften der Linkage-Verfahren

Verfahren	Eigenschaft	Proximitätsmaße
Single Linkage	kontrahierend*; neigt zur Kettenbildung	alle
Complete Linkage	dilatierend**; neigt zu kleinen Gruppen	alle
Average Linkage (gewichtet/ungewichtet)	konservativ***	ungewichtet: alle gewichtet: Distanz. +
Zentroid	konservativ***	Distanzmaße+
Median (Zentroid, gewichtet)	konservativ***	Distanzmaße+
Ward	konservativ***; bildete etwa gleich große Gruppen; sehr valide bei unkorrelierten Merkmalen und bei der Erwartung etwa gleich großer Gruppen (vgl. Bergs 1981, zit. nach Backhaus 2000: 366)	Distanzmaße+

* bildet wenige große Gruppen, denen viele kleine ("Ausreißer") gegenüberstehen;

** Objekte werden tendenziell zu gleich großen Gruppen zusammengefasst;

*** keine systematischen Tendenzen zur Dilatation oder Kontraktion

+ Zahlreiche Autoren empfehlen, für diese Verfahren nur die quadrierten euklidischen Distanzen (L2squared in Tabelle 7) zu nutzen (StataCorp 2005: 79).

d. Abbruch

Der Abbruch des Clusterprozesses erfolgt, wenn alle Objekte in einer Gruppe zusammengefasst sind. Weil die Zusammenfassung in eine Gruppe in der Regel keine sinnvolle analytische Lösung des Forschungsproblems darstellt, muss nun eine Entscheidung darüber getroffen werden, was die **optimale Gruppierung** ist. Für die vorliegende Studie werden aus Gründen der Vergleichbarkeit Abgrenzungen gewählt, die in 150 Regionen resultieren.

Ergebnisse:

Tabelle 7 zeigt die in Abschnitt a.) beschriebenen Distanzmatrizen und gibt für die 29 Clusterlösungen mit einer Modularität $Q > 0,8$ das entsprechende Linkage-Verfahren und den Rangplatz an. Offensichtlich lassen sich auf Grundlage der Ausgangsmatrix und ihrer symmetrischen Variante die besten Ergebnisse erzielen, wenn die Daten zeilenweise oder matrixweise normalisiert werden oder aber auf Zusammenhangsmaße zurückgegriffen wird und wenn das Single- oder Average-Linkage-Verfahren genutzt wird.

Tabelle 7: Ergebnisse der verschiedenen Verfahren der Clusteranalyse für 150 Regionen*

Matrix bzw. Matrixtransformation / Verfahren	Ausgangs-Pendlermatrix	Symmetrisierung	Diagonalwert = 0	Symmetrisierung & Diagonalwert = 0
A. Ausgangsdaten** Pendlermatrix				
B. Normalisierung Matrixweise: Zeilenweise: Spaltenweise: Zeilen- & Spaltenweise:	av: 0,819 (6), wav: 0,806 (24) si: 0,823 (1) si: 0,815 (13) si: 0,816 (12)	av: 0,819 (6), wav: 0,806 (24) si: 0,823 (2) (= Zeilenweise) si: 0,805 (28)	av: 0,819 (6), wav: 0,806 (24) si: 0,812 (14) / /	av: 0,819 (6), wav: 0,806 (24) si: 0,809 (19) (= Zeilenweise) /
C. Korrelationen Korrelationskoeffizient: Q-Korrelationskoeffizient: Winkeldistanz/Angular-Maß:	si: 0,820 (5) av: 0,800 (29) si: 0,816 (10), av: 0,810 (18), wav: 0,807 (22) si: 0,816 (10), av: 0,810 (17), wav: 0,807 (22)	si: 0,820 (3), av: 0,811 (15), wav: 0,808 (21) (= Korrelationskoeffizient.) si: 0,820 (3), av: 0,811 (16), wav: 0,809 (20)	/	/
D. Minowski-Metriken L2: L2squared: L1: Linfinity: Canberra:	/	/	/	/

* Ausgewiesen sind die Rangplätze hinsichtlich der Modularität Q (**fett: Top 10**), wobei: si = single linkage; av = average linkage und wav = weighted average linkage

** Weil die Ausgangsdaten Ähnlichkeitsmatrizen und nur z.T. symmetrisch sind, sind sie selbst nicht als Ausgangsdaten für die Clusteranalyse geeignet, sondern es werden die folgenden Transformationen genutzt.

Die Ergebnisse überraschen insofern, als das Single-Linkage-Verfahren sehr gute Ergebnisse liefert, obwohl das Verfahren in anderen Clusteranalysen zur Kettenbildung neigt, wodurch einige sehr große Cluster sehr kleinen gegenüberstehen. Diese Eigenschaft ist sinnvoll z.B. für die Identifizierung (und Eliminierung) von Ausreißern in Daten. Stattdessen wären für das Complete-Linkage-Verfahren bessere Ergebnisse zu erwarten gewesen, denn es neigt dazu, räumlich kompakte Cluster zu generieren (StataCorp 2005: 76), wie dies für die vorliegende Zielstellung wünschenswert erscheint. Dieses Linkage-Verfahren hat jedoch zu keinen zufrieden stellenden Ergebnissen geführt. Dem Average Linkage-Verfahren wird bescheinigt, in vielen Situationen robuste Ergebnisse zu liefern (vgl. Kaufman/Rousseeuw 1990, zit. nach ebenda). In der Tat finden wir für diese Verfahren und seine anders gewichtende Variante, das weighted Average-Linkage, z.T. gute Ergebnisse.

Neben den genannten Verfahren wurden auch die Ergebnisse des Zentroid-Verfahrens, das als eine weitere Variante von Average Linkage betrachtet werden kann, sowie das Median-Verfahren (eine anders gewichtende Variante des Zentroid-Verfahrens) untersucht, doch sie lieferten keine zufriedenstellenden Ergebnisse. Dasselbe gilt für das Ward-Verfahren. Dieses ist jedoch ohnehin an bestimmte Voraussetzungen geknüpft

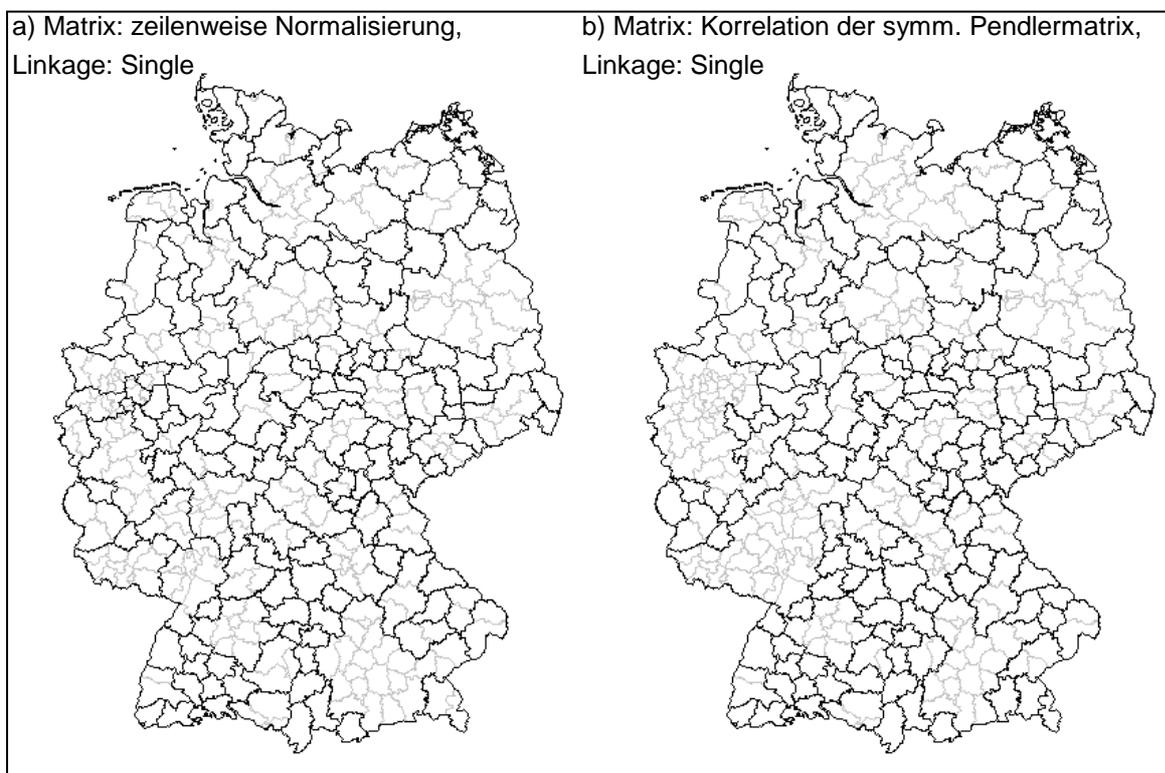
(z.B. sollten die zu bildenden Gruppen ähnlich groß sein und ähnlich viele Beobachtungen aufweisen), die im vorhandenen Beispiel nicht gegeben sind.

Das insgesamt beste Ergebnis lieferte die auf der zeilenweisen Normalisierung basierende Single-Linkage-Clusterlösung mit $Q=0.8234$ ($Q=0.8228$ für die symmetrische Ausgangsmatrix). Diese Ergebnis ist wird auch in Abbildung 6 gezeigt. Gute Ergebnisse zeigt auch das Average-Linkage-Verfahren für die matrixweise normalisierten Pendlermatrizen – und zwar identische Werte für alle Varianten (vgl. Abbildung A1.c im Anhang).

Sehr ähnliche Ergebnisse liefern die im Block D zusammengefassten Verfahren. Dies liegt darin begründet, dass diese Matrizen eine hohe Ähnlichkeit aufweisen. Vor allem die Werte des Q-Korrelationskoeffizienten und des Winkeldistanz- oder Angular-Maßes korrelieren sehr hoch miteinander. Für die symmetrische Pendlermatrix ist die variablen- (=Spalten) oder beobachtungsweise (=Zeilen) Korrelationsanalyse identisch.

Andere Normalisierungen als die zeilen- oder matrixweise sind offensichtlich nicht oder weniger erfolgversprechend. Auch die Minowski-Metriken und auf Matrixtransformationen ohne Diagonalwerte beruhende Clusteranalysen führten nicht zu zufriedenstellenden Ergebnissen und wurden darum nicht in Tabelle 7 aufgeführt.

Abbildung 6: Clusterlösungen mit der höchsten Modularität

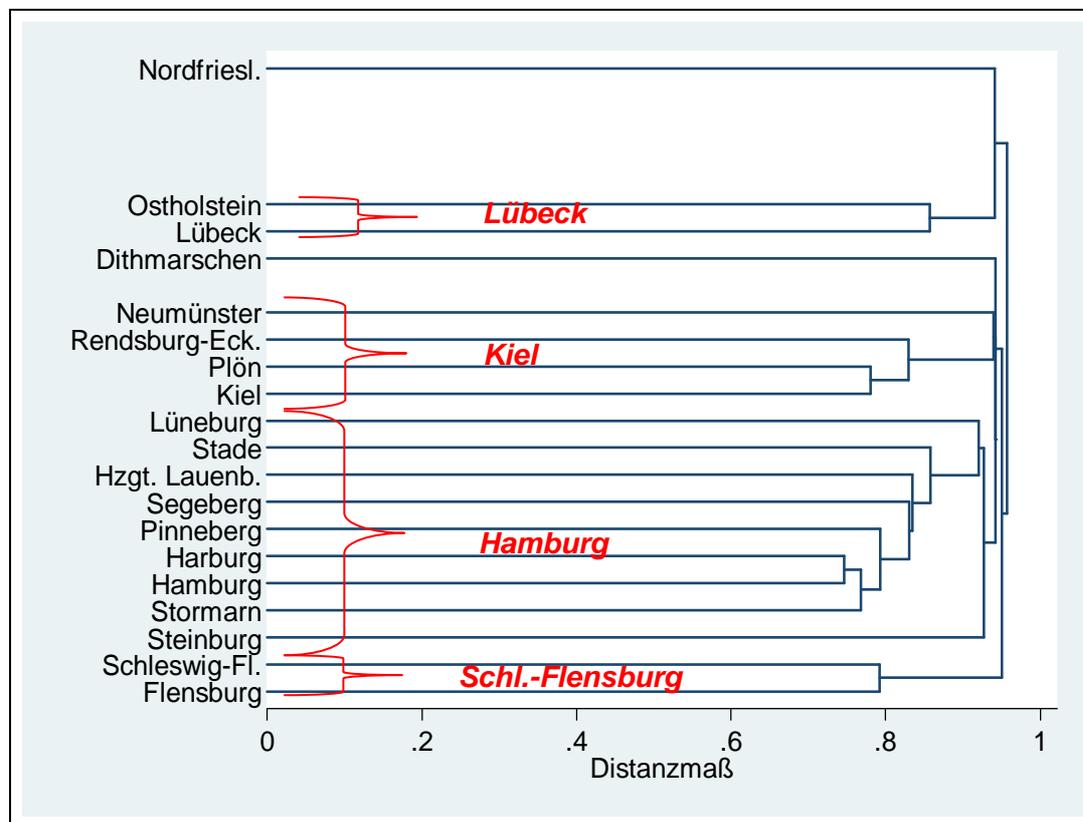


Obwohl sich die Modularitätswerte der Abgrenzungen nur geringfügig unterscheiden, offenbaren die entsprechenden Karten wichtige Unterschiede hinsichtlich der Arbeitsmarktregionen im Rhein-Ruhr-Gebiet sowie für die Frankfurter und Hamburger

Arbeitsmarktregion, die bei der auf der Korrelationsmatrix der Pendlerbeziehungen basierenden Abgrenzung deutlich vergrößert sind.

Zur Veranschaulichung des Clusterprozesses sind Dendrogramme (Baumdiagramme) ein beliebtes Mittel. Abbildung 7 zeigt ein Beispiel für die Norddeutschland.

Abbildung 7: Dendrogramm für die Clusterlösung mit der höchsten Modularität (Ausschnitt für Regionen Norddeutschlands)



Die x-Achse zeigt die Distanzwerte an, bei der die einzelnen Regionen vereinigt wurden. Im vorliegenden Ausschnitt erfolgte die erste Fusion zwischen Harburg und Hamburg. Dieser Region war der Kreis Stormarn am ähnlichsten, der als nächstes mit der Region Hamburg/Harburg fusioniert wurde. Die letzte für die Gliederung in insgesamt 150 Regionen relevante Fusion war die zwischen Rendsburg-Eckernförde/Plön/Kiel und Neumünster. Würde der Clusterprozess nicht an dieser Stelle abgebrochen, so würden alle Regionen Schritt um Schritt zu einer großen Arbeitsmarktregion zusammengefasst. Je nachdem, an welcher Stelle in einem Dendrogramm ein senkrechter Schnitt gesetzt wird, ergeben sich also eine unterschiedlich Zahl von Arbeitsmarktregionen mit entsprechend unterschiedlichen großen Regionen. Auf diese Weise können Dendrogramme auch zu Analyse von Substrukturen oder von Superstrukturen genutzt werden. Betrachtet man den linken Teil des Diagramms, so lässt sich die Region Hamburg/Harburg/Stormarn als Kern des Hamburger Arbeitsmarktes identifizieren, zu dem schrittweise andere, einzelne Regionen zugefügt werden. Fokussiert man auf den rechten Teil, so zeigt sich die enge Verbindung des Hamburger und des Kieler Arbeitsmarktes, denn beide Regionen würden als nächstes fusioniert.

Die Visualisierung des Clusterprozesses anhand des Dendrogrammes gehört sicherlich zu den Vorzügen der Clusteranalyse. Nach praktischen oder analytischen Gesichtspunkten kann der Clusterprozess an verschiedenen Stellen abgebrochen werden mit der entsprechenden Anzahl von verbleibenden Regionen. Eine Reihe von prinzipiellen Problemen der hierarchischen Clusterverfahren werden im Abschnitt zur Faktorenanalyse diskutiert. Ein Problem wurde jedoch bereits weiter oben angeschnitten: Die Clusterprozedur erfordert eine Symmetrisierung der Daten und damit einen gewissen Informationsverlust. Auch die Neuberechnung der Distanzen (über die Linkage-Verfahren) erfolgt, wie im folgenden Abschnitt gezeigt wird, nicht unbedingt den inhaltlichen Erfordernissen von Pendlerverflechtungen zwischen Regionen.

5.3 Alternative Cluster-Verfahren

Schließlich ist zu überlegen, ob die Verfahren der inhaltlichen Aufgabe der Fusion von Regionen aufgrund der Pendlerverflechtungen entsprechen. Wenn zwei Regionen fusioniert werden, sollte die Anzahl der Pendler aus der bzw. in die fusionierte Region nicht zwangsläufig durch die Wahl des kleineren oder größeren oder des gemittelten Wertes erfolgen, sondern durch die Addition der Pendlerströme. Tabelle 8 illustriert dieses Vorgehen.

Tabelle 8: Pendlermatrix vor und nach der Fusion der Kreise Flensburg und Schleswig-Flensburg

Kreisregion	<i>Flensburg</i>	<i>Schleswig-Flensburg</i>	Nordfriesland	Steinburg	Dithmarschen	Kiel	...	Zeilen-summe
a) Ausgangsdaten								
<i>Flensburg</i>	16955	3881	439	33	30	404		23370
<i>Schleswig-Flensburg</i>	13613	2041	1668	90	257	1348		54648
Nordfriesland	1795	30593	38955	137	817	566		47339
Steinburg	46	65	89	23261	1324	449		40246
Dithmarschen	94	249	1132	2274	27863	496		37257
Kiel	277	259	107	117	98	50513		68279
b) Fusion von Flensburg und Schleswig-Flensburg:								
<i>Flensburg/Schleswig-Flensburg</i>		36490	2107	123	287	1752		78018
Nordfriesland		32388	38955	137	817	566		47339
Steinburg		111	89	23261	1324	449		40246
Dithmarschen		343	1132	2274	27863	496		37257
Kiel		536	107	117	98	50513		68279

Tabelle 3 hatte als stärkste Pendlerbeziehung die zwischen Flensburg und Schleswig-Flensburg ausgewiesen. Im oberen Teil der Tabelle 8 sind die Pendlerzahlen wiedergegeben. Fett und kursiv sind die zu fusionierenden Regionen markiert und nur kursiv andere von dieser Fusion betroffene Pendlerzahlen, die demzufolge neu berechnet werden müssen. Die Fusion dieser Regionen ergibt die Werte im unteren Teil

der Tabelle, die wiederum als Grundlage für die Neuberechnung der Anbindungskoeffizienten und damit die Auswahl der nächsten zu fusionierenden Regionen dienen.

Bleibt die Frage nach der Auswahl der zu fusionierenden Regionen. Im obigen Beispiel wurde die stärkste relative Auspendlerbeziehung (der größte Anbindungskoeffizient der Matrix) als Grundlage genommen. Als Alternativen sind jedoch auch Einpendlerbeziehungen oder die symmetrische Pendlermatrix denkbar, für die die Unterscheidung nach Ein- und Auspendlern nicht nötig ist. Schließlich sind unterschiedliche Verfahrensweisen hinsichtlich der Berücksichtigung der Pendlerentfernungen möglich. Die einfachste Variante besteht sicherlich darin, Pendlerbeziehungen ab einer bestimmten Entfernung nicht zu berücksichtigen. Für die vorliegende Studie wurde eine Variante berechnet, in der die Entfernung als Gewicht eingeht. Die Gewichtung wurde vorgenommen, indem der Pendlerzahlen durch die Pendeldistanz geteilt wurde. Dadurch werden die Pendlerzahlen zu entfernten Regionen stärker verringert als zu nahe gelegenen.

Der Clusteralgorithmus lässt sich demnach wie folgt zusammenfassen.

Schritt 1: Auswahl der Pendlermatrix

(Einpendler / Auspendler / Ein & Auspendler (symmetrisch) / Gewichtung der Pendlerzahlen)

Schritt 2: Berechnung der Pendleranteile

Schritt 3: Fusion der Regionen

Es werden die Regionen zusammengefasst, die durch den höchsten Pendleranteil verbunden sind, und die reduzierte Pendlerverflechtung bestimmt, aus der die reduzierte Matrix der Anbindungskoeffizienten berechnet wird (vgl. Tabelle 8). Dabei ist die Pendlerzahl zwischen einer beliebigen Region und der neuen fusionierten die Summe der Pendler aus/in die fusionierten Regionen. Aus diesen Werten können nun erneut die Anbindungskoeffizienten berechnet werden, die die Grundlage für die Auswahl der nächsten zu fusionierenden Regionen bildet (Schritt 2).

Schritt 4: Auswahl der Clusterlösung

Im vorliegenden Fall wird die Clusterlösung mit 150 Regionen gewählt. Denkbar wäre auch die Auswahl von Regionen nach bestimmten Kennziffern, z.B. nach der maximalen Modularität, dem insignifikant werden von räumlichen Autokorrelationseffekten (Fertig u.a. 2006) oder anhand des Verlaufes der Anbindungskoeffizienten (Abbildung 4).

Ergebnisse:

Ähnlich wie bei den Ergebnissen der "klassischen" Clusteranalysen ergeben sich auch hier die besten Ergebnisse für die Auspendlermatrix, in diesem Fall für die symmetrische Variante. Das Ergebnis für die Einpendlermatrix liegt deutlich darunter, aber mit Werten für die Modularität Q von über 0.8 immer noch im Bereich sehr guter Abgrenzungen.

Auch die Ergebnisse der entfernungsgewichteten Matrizen liegen noch in diesem Bereich.

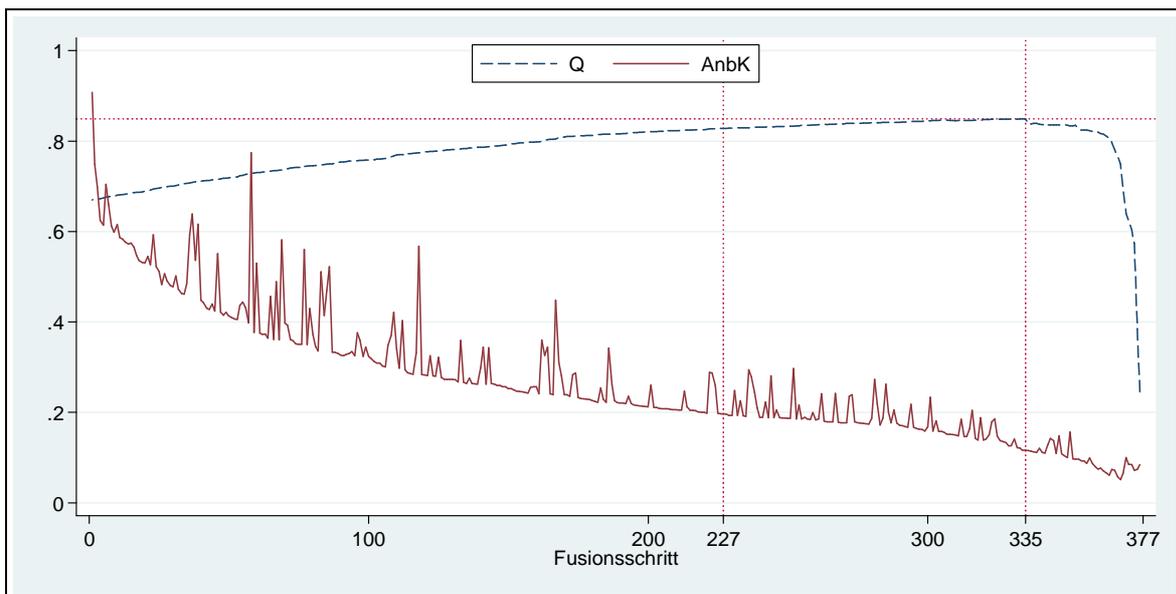
Tabelle 9: Modularitäten für Ein- und Auspendler

Ausgangsmatrix	Modularität Q
Auspendler (zeilenweise Normalisierung*)	0.824
Auspendler, entfernungsgewichtet	0.809
Ein- und Auspendler (zeilenweise Normalisierung*, symmetrisch)	0.828
Ein- und Auspendler, entfernungsgewichtet	0.807
Einpendler (spaltenweise Normalisierung*)	0.807

* jeweils bei der Neuberechnung der Anbindungskoeffizienten

Die folgende Abbildung zeigt den Verlauf des Clusterprozesses für die symmetrische Pendlermatrix.

Abbildung 8: Verlauf der Modularität Q und des Anbindungskoeffizienten über die Fusionsschritte



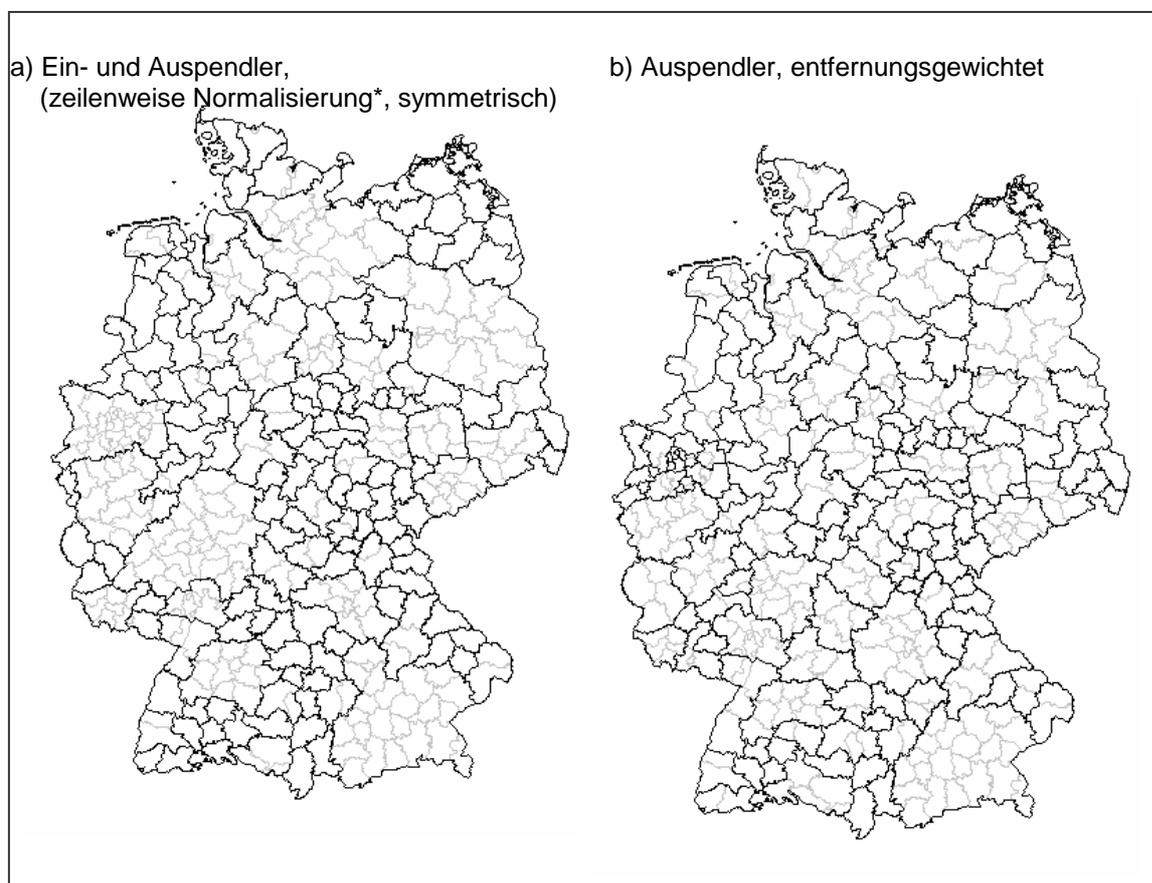
Die Abbildung zeigt deutlich die fallenden Werte des Anbindungskoeffizienten. Nach der Fusion von zwei größeren Regionen können die Werte allerdings wieder deutlich ansteigen. Die Werte für die Modularität starten bei ca. 0,67. Dieser relativ hohe Wert ist durch die zahlreichen Beschäftigten bedingt, die in der gleichen Kreisregion arbeiten und wohnen und die darum in jedem Fall richtig zugeordnet werden. Deutlich ist im Folgenden der stetig zunehmende Wert der Modularität Q, der nach 335 Iterationen (resultierend in 43 Regionen) sein Maximum von 0,85 erreicht und damit deutlich über dem Wert für 150 Regionen liegt, der nach 227 Iterationen erreicht ist (377-150).

Der Verlauf der Modularitätswerte legt nahe, die optimale Abgrenzung von Wirtschaftsräumen bei einer 43-Regionen-Clusterlösung zu suchen. Die Möglichkeit zur

genauen Analyse der einzelnen Fusionsschritte ist sicherlich einer der Vorzüge des vorgestellten Verfahrens, auf die weiter unten auch zurückgegriffen wird.

Zwei ausgewählte Ergebnisse werden in Abbildung 9 als Karten dargestellt. Der in Hinblick auf die Modularität Q besten alternativen Clusterlösung ist die beste entfernungsgewichtete gegenüber gestellt. Wie zu erwarten, kommen durch die Gewichtung weniger sehr große Regionen zustande. Wegen der Bedingung, eine Clusterlösung mit 150 Regionen zu finden, wachsen in diesem Fall allerdings einige kleine Regionen stärker und es gibt weniger Einzelregionen. Insgesamt führt die Entfernungsgewichtung deutlich zu einer vergleichsweise großenhomogenen Abgrenzung.

Abbildung 9: Clusterlösungen mit der höchsten Modularität



Ähnlich wie im Dendrogramm lässt sich auch für die alternativen Clusterverfahren der Clusterprozess nachzeichnen. Tabelle 10 zeigt den Ablauf des Clusterprozesses wiederum für einen Ausschnitt in Norddeutschland.

Tabelle 10: Ablauf des Clusterprozesses (Ausschnitt für Norddeutschland)

Region	Fusionsschritt	Anbindungs- koeffizient	Fusionsziel	Endgültige Arbeitsmarktregion
Flensburg		2	0,75 Schleswig-Fl.	Schleswig-Fl.
Stormarn		4	0,63 Hamburg	Hamburg
Harburg		13	0,58 Hamburg	Hamburg
Hzgt. Lauenburg		19	0,53 Hamburg	Hamburg
Pinneberg		22	0,53 Hamburg	Hamburg
Segeberg		23	0,59 Hamburg	Hamburg
Plön		29	0,48 Kiel	Kiel
Steinburg		62	0,37 Hamburg	Hamburg
Stade		63	0,37 Hamburg	Hamburg
Rendsburg-Eck.		64	0,36 Kiel	Kiel
Neumünster		65	0,46 Kiel	Kiel
Lüneburg		100	0,32 Hamburg	Hamburg
Lübeck		111	0,30 Hamburg	Hamburg
Ostholstein		112	0,40 Hamburg	Hamburg
Schwerin		121	0,28 Ludwigslust	Hamburg
Parchim		122	0,33 Ludwigslust	Hamburg
Wismar		192	0,22 Hamburg	Hamburg
Ludwigslust		193	0,24 Hamburg	Hamburg
Uelzen		200	0,21 Hamburg	Hamburg
Lüchow-Danneberg		201	0,26 Hamburg	Hamburg
Abbruch nach 227 Fusionsschritten = bei 150 Regionen, weitere Fusionen wären:				
<i>Dithmarschen</i>		<i>247</i>	<i>0,19 Hamburg</i>	
<i>Kiel</i>		<i>265</i>	<i>0,18 Hamburg</i>	
<i>Schleswig-Fl.</i>		<i>266</i>	<i>0,18 Hamburg</i>	
<i>Nordfriesl.</i>		<i>267</i>	<i>0,24 Hamburg</i>	
<i>Hamburg</i>		<i>369</i>	<i>0,05 Düsseldorf</i>	

Die Zuweisung von Flensburg zu Schleswig-Flensburg und von Stormarn, Harburg, Hzgt. Lauenburg, Pinneberg und Segeberg zu Hamburg erfolgen bereits sehr früh im Clusterprozess. Dies und die Anbindungskoeffizienten von über 50% zeigen die enge Verflechtung der jeweiligen Regionen. Das zeitweilige Ansteigen des Anbindungskoeffizienten resultiert aus der Tatsache, dass nach der Fusion zweier Regionen, mit denen eine dritte Region eng verflochten ist, die Summe der Pendlerzahlen in jede der fusionierten Regionen die Grundlage für die Neuberechnung der Anbindungskoeffizienten ist. Dieser inhaltlich plausible Vorgang erschwert jedoch die Darstellung des Prozesses als Dendrogramm. Dennoch lässt auch die tabellarische Darstellung ähnliche Interpretationen wie das Dendrogramm in Abbildung 7 zu. Im oberen Teil lassen sich wieder Stormarn und Harburg, aber auch das Hzgt. Lauenburg, Pinneberg sowie Segeberg als Kernregion des Hamburger Arbeitsmarktes identifizieren.

Im unteren Teil der Tabelle sind die Fusionen aufgelistet, die stattfinden würden, wenn der Clusterprozess nicht nach 227 Fusionsschritten abgebrochen worden wäre. Im

vorliegenden Beispiel wäre eine "Superregion Hamburg" entstanden, bevor diese Region selbst einer "Superregion" Düsseldorf, die zu diesem Zeitpunkt nahezu den gesamten westdeutschen Raum umfasst, zugeschlagen würde.¹⁴

5.4 Faktorenanalyse

Ausgangspunkt der Faktorenanalyse ist die Annahme, dass gemessenen Ausprägungen ein oder mehrere dahinterliegende und nicht gemessene (latente) Faktoren zugrunde liegen (StataCorp 2005: 257). Dies entspricht der Vermutung der vorliegenden Studie, dass den Pendlerverflechtungen zwischen administrativen Regionen eine latente Struktur von Arbeitsmarktregionen zugrunde liegt. Das Hauptziel der Faktorenanalyse besteht darin, die in einem Set von Variablen enthaltene Information mit Hilfe weniger Faktoren zu reproduzieren. Angewandt auf die Pendlermatrix bedeutet dies, die Regionen mit ähnlichen Pendlerstrukturen zusammen zu fassen. Die so gebildeten Faktoren können nunmehr als Arbeitsmarktregionen interpretiert werden.

Für die Verwendung der Faktorenanalyse zur Analyse von Pendlerdaten sprechen vor allem zwei Argumente (siehe dazu Eckey u.a. 2006: 302):

- Zum einen ist die Fusion von Regionen in hierarchischen Clusterverfahren wie den oben besprochenen irreversibel. Wird beispielsweise die periphere Region P zunächst einem benachbarten Arbeitsmarktzentrum Z1 zugeschlagen, so ist es durchaus möglich, dass stärkere Pendlerbeziehungen zu einem erst später gewachsenen anderen benachbarten Zentrum Z2 bestehen, wenn in diesem Regionen vereinigt wurden, zu denen P ebenfalls umfangreiche Pendlerbeziehungen hat¹⁵. Beim Clusterverfahren bleibt jedoch die Verbindung zwischen der Region P und dem Arbeitsmarktzentrum Z1 bestehen und kann nicht aufgelöst werden, um nachträglich dem Arbeitsmarktzentrum Z2 zugeordnet zu werden. Beim iterierenden Verfahren der Faktorenanalyse tritt dieses Problem nicht auf, da es flexibel auf die jeweils aktuelle Situation reagieren kann.
- Der zweite Vorzug der Faktorenanalyse besteht darin, dass Regionen nicht auf Basis des Vergleichs von Werten zwischen Dyaden (d.h. jeweils zwei Regionen) fusioniert werden, sondern nach der Ähnlichkeit der gesamten Pendlerstruktur. Dieses Argument gilt allerdings nur bedingt: Soweit die oben beschriebenen Verfahren auf Zusammenhangsmaßen beruhen, berücksichtigen sie genau diese Gesamtstruktur der Pendlerbeziehungen zwischen Regionen.

Die Vorgehensweise für die hier durchgeführte Faktorenanalyse folgt im Wesentlichen Eckey u.a. 2006. Als Grundlage für die Faktorenanalyse dienen dieselben Ähnlichkeitsmatrizen wie für die Clusteranalyse, nur dass diesmal die Transformation zur Distanzmatrix überflüssig ist. Die Faktorenanalyse wurde mit der Vorgabe durchgeführt, alle Faktoren mit einem Eigenwert von mindestens Eins zu extrahieren (Kaiser Kriterium). Das Ergebnis dieses Prozesses sind Faktoren, die in einem bestimmten

¹⁴ Die Bezeichnungen der Arbeitsmarktregionen werden bei jedem Fusionsschritt von der Region mit den meisten Beschäftigten übernommen.

¹⁵ Diesem Problem kann nachträglich durch entsprechende Korrekturverfahren begegnet werden.

Zusammenhang mit den Ausgangsvariablen stehen (Faktorladungen bzw. factor loadings). Die Interpretierbarkeit der Faktoren kann durch Rotation der Faktorenladungen optimiert werden. Dabei werden die Ladungen für wenige Variablen so hoch wie möglich gesetzt und die der verbleibenden verringert (StataCorp 2005: 234). Unter den möglichen Rotationsverfahren wurde die schiefwinklige oblimin Rotation mit Kaiser-Normalisierung genutzt. Die Kaiser-Normalisierung gewährleistet, dass alle Zeilen der Matrix mit demselben Gewicht in die Rotation eingehen (StataCorp 2005: 239). Abschließend wurde jede Region dem Faktor zugeordnet, für den die Faktorladung den höchsten Wert aufwies. Für die Zuweisung der Regionen zu Faktoren ist es dabei unerheblich, ob die Werte der Faktorladung positiv oder negativ sind. Um eine optimale Diskriminierung zu ermöglichen, wurden die Vorzeichen der Faktorladungen für Faktoren gewechselt, deren größte Beträge im negativen Wertebereich lagen.

Im Unterschied zur Clusteranalyse lässt sich die Anzahl der Arbeitsmarktregionen (also Faktoren) nicht exogen vorgeben sondern nur indirekt über die Wahl der akzeptierten Eigenwerte. Mit dem Eigenwert von 1 erhält man je nach verwendeter Ähnlichkeitsmatrix zwischen 130 bis 158 Regionen; bei der Akzeptanz niedrigerer Eigenwerte werden mehr Faktoren berücksichtigt und entsprechend mehr Regionen identifiziert. Untersucht wurden ähnliche Matrizen wie für die Clusteranalyse (Tabelle 11). Der einzige Unterschied zu den dort verwendeten Matrizen liegt darin, dass es sich nunmehr um Ähnlichkeitsmatrizen handelt. Wie bei der Clusteranalyse mussten die Matrizen ggf. symmetrisiert werden. Die Diagonalwerte dürfen jedoch auch andere Werte als Null aufweisen. Aufgrund der unzureichenden Ergebnisse wurden die Matrizen der Minowski-Metrik und die Matrizen, die auf der Pendlermatrix mit Null-Werten auf der Diagonale beruhten, für die aufwändigeren Faktorenanalysen nicht berücksichtigt. Tabelle 11 zeigt die verwendeten Matrizen und die Ergebnisse.

Anstatt der matrixweise normalisierten Varianten können nunmehr die Ausgangsdaten selbst für die Faktorenanalyse genutzt werden. Die Ergebnisse sind mit $Q=0,76$ zwar akzeptabel, liegen aber weit hinter den besten der Clusteranalyse. Alle normalisierten Matrizen weisen ähnliche Werte auf.

Tabelle 11: Ergebnisse der Faktorenanalysen für annähernd 150 Regionen

Matrix	Ausgangs-Pendlermatrix	symmetrisierte Variante
A. Ausgangsdaten	Q=0,762 (N=158; Rang: 10)	Q=0,762 (N=158; Rang: 10)
B. Normalisierung		
Zeilenweise:	Q=0,758 (N=158; Rang: 14)	Q=0,758 (N=158; Rang: 12)
Spaltenweise:	Q=0,762 (N=158; Rang: 8)	(= Zeilenweise)
Zeilen- & Spaltenweise:	Q=0,758 (N=158; Rang: 13)	Q=0,762 (N=158; Rang: 9)
C. Korrelationen		
Korrelationskoeffizient*:	Q=0,776 (N=144; Rang: 7)	Q=0,788 (N=137; Rang: 6)
Q-Koorelationskoeffizient*:	Q=0,796 (N=130; Rang: 3)	(= Korrelationskoeffizient)
Winkeldistanz/Angular*:	Q=0,796 (N=131; Rang: 2)	Q=0,788 (N=138; Rang: 5)
D. Eckey u.a.	Q=0,807 (N=150; Rang: 1)	/
Replikation:	Q=0,794 (N=132; Rang: 4)	/

* auf Werte zwischen 0 und 1 normiert, siehe Fußnote 13

fett: die besten 3 Werte

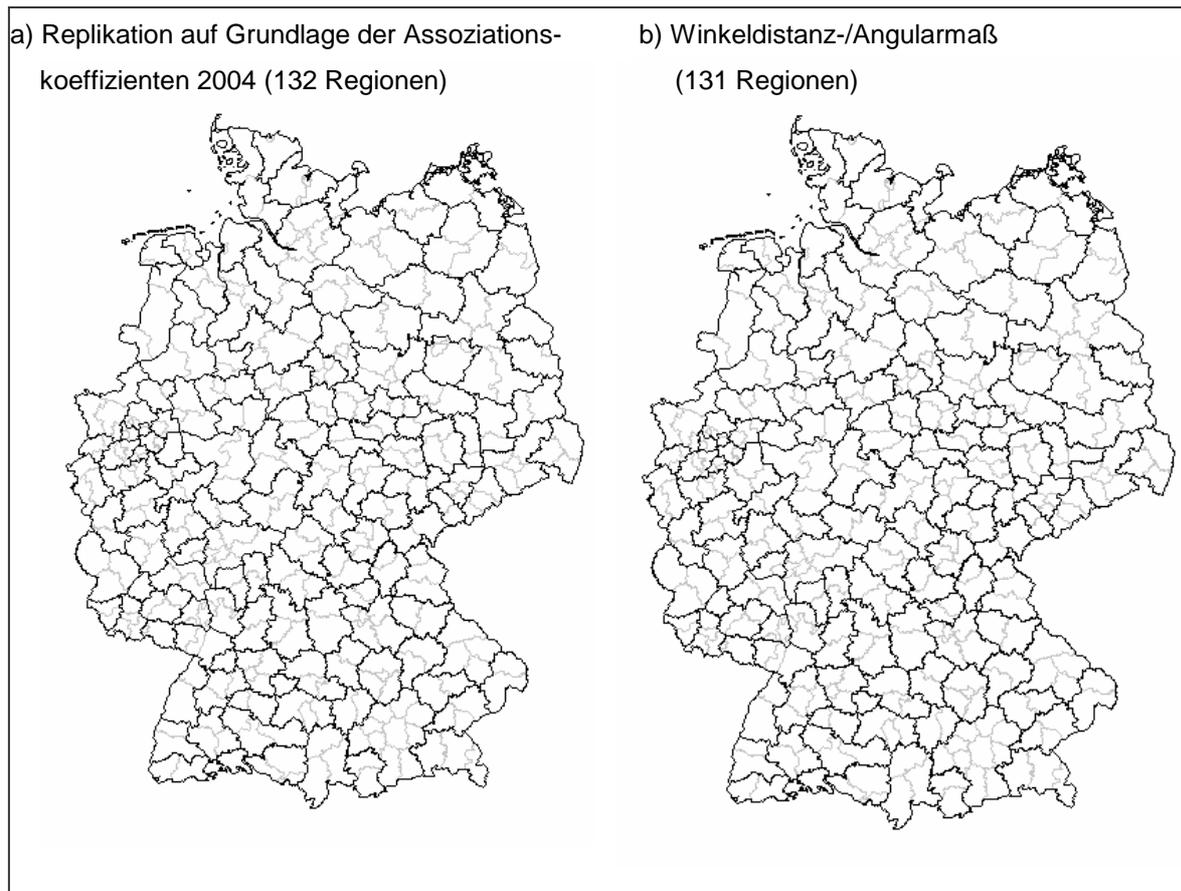
Dem faktoranalytischen Verfahren angemessener sind Korrelationsmatrizen. Sie erzielen in der Tat deutlich bessere Modularitätswerte. Als besonders geeignet erscheinen dabei Korrelationsmaße, die die Ähnlichkeit von Beobachtungen (Zeilen) anstatt von Variablen (Spalten) messen – nämlich die Matrix der Q-Korrelationskoeffizienten und die Angularkoeffizienten. Diese Koeffizienten sind im Übrigen nahezu identisch mit den bei Eckey u.a. (2006) verwendeten Assoziationskoeffizienten. Übertragen auf die hier untersuchten Daten bedeutet dies, dass die Ähnlichkeit der Auspendlerbeziehungen relevanter für eine adäquate Abgrenzung von Arbeitsmarktregionen ist als die Ähnlichkeit der Einpendlerbeziehungen.

Abschließend sind die Ergebnisse für die Abgrenzung von Eckey u.a. (2006) dargestellt. Diese Abgrenzung wurde auf Grundlage der Assoziationskoeffizienten der Pendlermatrix mit Pendlerdaten von 2004 gebildet. Die Replikation der Faktorenlösung für diese Matrix der Assoziationskoeffizienten von 2004 resultiert in der Abgrenzung von 132 Regionen mit einer nur geringfügig geringeren Modularität als für das vergleichbare Verfahren (Q-Korrelationskoeffizient) für 2005. Eckey u.a. (2006) nahmen allerdings eine Reihe von Umgruppierungen sowie Teilungen großer Regionen vor, um die Pendelzeiten auf 45 bis 60 Minuten zu reduzieren. In Tabelle 11 ist diese Abgrenzung enthalten und weist mit Q=0,807 sogar einen besseren Modularitätswert auf als die ursprüngliche 132-Regionen-Lösung für die Daten von 2004 (Q=0,794).

Die folgende Abbildung zeigt die Abgrenzung auf Grundlage der Assoziationskoeffizienten für 2004¹⁶ und das beste Ergebnis der Faktorenanalyse auf Grundlage der Daten von 2005. Beide Ergebnisse unterscheiden sich nur für 7 Kreisregionen. Insgesamt sind die Arbeitsmarktregionen im Vergleich zu anderen hier vorgestellten Abgrenzungen als größenhomogen zu beschreiben.

¹⁶ Wir danken Prof. Eckey für die Bereitstellung dieses Zwischenergebnisses.

Abbildung 10: Ergebnisse der Faktorenanalyse



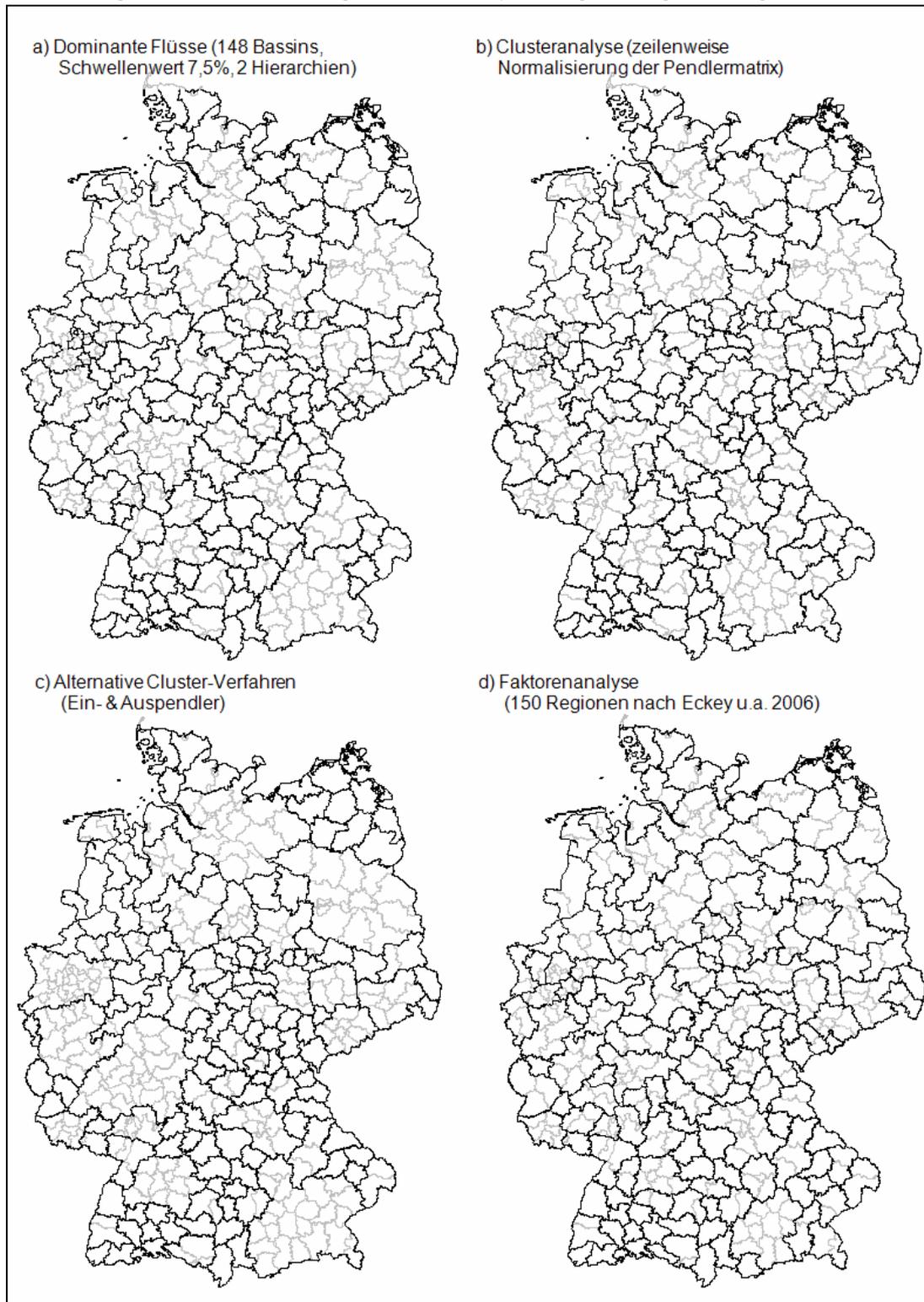
Trotz der starken eingangs diskutierten theoretischen Argumente für die Faktorenanalyse liegen die Modularitätswerte unter denen anderer bereits präsentierter Verfahren. Für die praktische Anwendbarkeit kann sich einer der Vorzüge der Faktorenanalyse als Nachteil erweisen. Die Anzahl der Arbeitsmarktregionen wird endogen, d.h. durch das Verfahren selbst bestimmt. Dadurch entfällt zwar die Notwendigkeit, für diesen Zweck auf externe Kriterien zurückgreifen zu müssen, aber auch die Möglichkeit Teilarbeitsmärkten oder "Superstrukturen" zu identifizieren, wie dies anhand der Dendrogramme bei hierarchischen Verfahren gezeigt wurde. Für viele regionalpolitische Aufgabenstellungen ist es sicherlich von Vorteil, dass bei diesem Verfahren keine extrem großen Regionen entstehen.

6 Vergleich der Abgrenzungsergebnisse

6.1 Visueller Vergleich

In den vorangegangenen Abschnitten wurden die Ergebnisse der verschiedenen Verfahren anhand der Modularitätswerte bewertet und ausgewählte Abgrenzungen als Karte präsentiert. Abbildung 11 fasst die besten Abgrenzungen der Verfahren zusammen.

Abbildung 11: Die besten Ergebnisse der jeweiligen Abgrenzungsverfahren



Alle hier gezeigten Abgrenzungen identifizieren wie in Abschnitt 3 gefordert die wichtigen Arbeitsmarktzentren als Zentren von Arbeitsmarktregionen. Die Identifizierung kleiner Zentren und die Zuweisung der Kreisregionen zu den Arbeitsmärkten erfolgt jedoch durchaus unterschiedlich. Während das beste Ergebnis des alternativen Cluster-Verfahrens Arbeitsmarktregionen mit sehr unterschiedlicher Größe erzeugt, sind die Regionalen Arbeitsmärkte nach Eckey u.a. (2006) deutlich flächenhomogener. Eine Zwischenstellung nehmen hinsichtlich der Größenheterogenität die Ergebnisse des Konzepts der dominanten Flüsse und der Clusteranalyse ein.

6.2 Quantitativer Vergleich

Im folgenden Abschnitt soll die Bewertung der Ergebnisse auf eine breitere quantitative Basis gestellt und mit den bisher existierenden administrativen und funktionalen Abgrenzungen verglichen werden.

Für die Bewertung der Abgrenzungen kommen zwei Gruppen von Maßen zum Einsatz. Zum einen gilt es zu erfassen, wie gut es gelingt, die Pendlerströme "einzufangen", zum anderen kommt der räumlichen Struktur eine besondere Bedeutung zu. Tabelle 12 zeigt diese Maße für die in den obigen Abschnitten behandelten Verfahren, sowie für die Abgrenzung der Gemeinschaftsaufgabe, die Kreisregionen selbst und für eine zufällige Zuweisung von Kreisregionen zu Arbeitsmarktregionen. Grau unterlegt sind Ergebnisse, die mindestens um eine halbe Standardabweichung schlechter sind als der Durchschnitt eines Maßes, und fett hervorgehoben sind Werte, die um mindestens diesen Betrag besser sind als der Durchschnitt. Für diese Bewertung wurden die Vergleichsregionen nicht einbezogen.

6.2.1 Modularitätswerte

Die erste Spalte gibt die Modularität Q und in Klammern den entsprechenden Rangplatz der hier verglichenen 34 Verfahren, den 3 in Abschnitt 2 besprochenen administrativen und funktionalen Gliederungen und einer Zufallsabgrenzung wieder. Alle hier präsentierten Verfahren weisen gute Werte auf. Dagegen unterscheiden sich die Werte der administrativen Gliederung der Kreisregionen nur geringfügig von der Zufallsgliederung für 150 Arbeitsmarktregionen. Auch die für die Arbeitsmarkt- und Wirtschaftsförderung so relevante Abgrenzung der Gemeinschaftsaufgabe gibt die Realität der regionalen Arbeitsmärkte nur sehr begrenzt wieder. Einzig die Raumordnungsregionen, die aber deutlich weniger Regionen aufweisen, können hinsichtlich der hier verwendeten Kriterien mit den erläuterten Verfahren konkurrieren.

Tabelle 12: Maße für die räumlichen Eigenschaften*

Verfahren	Erfassen der Pendlerströme						Räumliche Struktur					
	Q		Pendlerquoten		Mittelwerte für:		Größe					
	Wert (1)	(Rang) (1)	gesamt (2)	ø AMR (3)	ø KR (4)	ESC (5)	HSC (6)	HWR (7)	AMR (8)	StAbw. (9)	Einzel (10)	nicht 1. (11)
Konzept der dominanten Flüsse												
Bassin(H:1;SW:5)	0.783	(33)	0.25	0.61	0.65	76	80	-4.0	151	189	46	92
Bassin(H:2;SW:7,5)	0.814	(11)	0.20	0.54	0.56	77	82	-4.6	148	263	77	47
Clusteranalyse												
Matrix norm.(=D0;=sym)+av	0.819	(7)	0.19	0.58	0.58	76	82	-7.2	150	247	54	92
Matrix norm.(=D0;=sym)+wav	0.806	(25)	0.21	0.61	0.61	75	81	-7.5	150	218	46	101
Zeile norm.+si	0.823	(3)	0.18	0.53	0.50	77	82	-6.3	150	284	90	39
Zeile norm.(sym)+si	0.823	(4)	0.18	0.53	0.50	77	83	-7.0	150	296	89	46
Zeile norm.(D0)+si	0.812	(12)	0.20	0.66	0.57	74	80	-7.2	150	251	78	44
Zeile norm.(sym/D0)+si	0.809	(17)	0.21	0.66	0.59	75	80	-6.5	150	228	68	62
Spalte norm.+si	0.815	(10)	0.19	0.56	0.52	76	82	-7.6	150	310	89	56
Zeile/Spalte norm.+si	0.816	(9)	0.18	0.56	0.51	76	82	-7.9	150	334	89	54
Zeile/Spalte norm.(sym)+si	0.805	(26)	0.21	0.58	0.56	76	82	-7.5	148	244	72	69
Korr./Angular(sym)+si	0.820	(5)	0.14	0.54	0.43	77	82	-7.1	150	465	103	32
Korr.+si	0.820	(6)	0.15	0.55	0.44	76	83	-7.7	150	437	100	37
Angular/Q-Korr.+si	0.816	(8)	0.15	0.54	0.44	77	82	-6.4	150	453	107	26
Korr.(sym)+av	0.811	(13)	0.21	0.50	0.56	78	83	-6.2	150	217	44	101
Angular(sym)+av	0.811	(14)	0.21	0.50	0.56	78	83	-6.1	150	217	45	100
Angular+av	0.810	(15)	0.21	0.50	0.55	78	83	-5.4	150	225	55	87
Q-Korr.+av	0.810	(16)	0.21	0.50	0.56	78	83	-5.3	150	218	53	90
Angular(sym)+wav	0.809	(19)	0.21	0.50	0.57	78	83	-6.0	150	211	46	105
Korr.(sym)+wav	0.808	(20)	0.21	0.51	0.57	78	83	-5.9	150	210	44	108
Angular/Q-Korr+wav	0.807	(23)	0.21	0.50	0.57	78	83	-5.3	150	208	51	90
Alternative Cluster-Verfahren												
Auspendler	0.824	(2)	0.16	0.55	0.48	77	81	-4.8	150	364	103	19
Auspendler, entfernungsgewichtet	0.809	(18)	0.20	0.66	0.57	75	79	-5.7	150	260	80	43
Einpendler	0.807	(24)	0.15	0.56	0.47	76	82	-7.9	150	492	98	51
Ein- und Auspendler	0.828	(1)	0.15	0.55	0.46	77	82	-6.0	150	374	105	19
Ein- und Auspendler, entf.gew.	0.807	(22)	0.21	0.66	0.59	74	80	-7.3	150	236	63	66
Faktoranalyse												
Ausgangsdaten (=sym)	0.762	(35)	0.29	0.64	0.74	75	79	-5.3	158	133	3	200
Korr.	0.776	(34)	0.27	0.60	0.70	76	81	-5.4	144	152	3	171
Korr.(sym)	0.788	(32)	0.25	0.54	0.64	78	82	-4.3	137	171	3	154
Q-Korr.	0.796	(29)	0.23	0.51	0.61	79	82	-4.1	130	186	4	142
Angular	0.796	(28)	0.23	0.51	0.61	79	82	-4.2	131	186	5	140
Angular(sym)	0.788	(31)	0.25	0.54	0.64	78	82	-4.4	138	171	4	154
Replikation	0.794	(30)	0.24	0.52	0.62	78	82	-4.3	132	189	4	146
Eckey u.a.2006	0.807	(21)	0.21	0.51	0.58	78	82	-4.7	150	214	60	78
Vergleichsabgrenzungen												
Raumordungsregion	0.799	(27)	0.23	0.48	0.58	80	83	-4.2	97	194	4	120
Gemeinschaftsaufgabe	0.747	(36)	0.32	0.77	0.81	71	77	-7.4	270	126	208	62
Kreisregion	0.670	(37)	0.48	1.11	/	/	/	/	377	73	377	13
Zufall	0.666	(38)	0.48	1.06	1.10	65	70	-5.2	150	139	41	335

* fett: gute Ergebnisse (mind. eine halbe Standardabweichung besser); grau unterlegt: schlechte Ergebnisse (mind. eine halbe Standardabweichung schlechter)

6.2.2 Pendlerquoten

Das Ausgangsziel der Studie war es, Abgrenzungen von Arbeitsmarktregionen so vorzunehmen, dass die Anzahl der Pendler minimiert wird. Eine Reihe von Kennziffern neben dem Modularitätswert Q erlaubt es, dieses Ziel zu evaluieren. Wenn das Maximum der in den Arbeitsmarktregionen wohnenden Beschäftigten dort auch tätig ist, sollte es möglichst wenige Pendler zwischen den Arbeitsmarktregionen geben. Entsprechende Pendlerquoten lassen sich für die Gesamtzahl der zwischen Arbeitsmarktregionen Pendelnden wie auch für die einzelne Arbeitsmarktregionen und die Kreisregionen ermitteln. Das Verhältnis aller Pendler zu allen Nichtpendlern bildet die *Pendlerquote insgesamt* (dargestellt in Spalte 2 von Tabelle 12). Die Pendlerquote der Arbeitsmarktregionen wird aus der Summe der Ein- und Auspendler im Verhältnis zu den in der Arbeitsmarktregion Wohnenden und Beschäftigten (Nichtpendler) berechnet. Diese Maßzahl liegt für alle Arbeitsmarktregionen vor und ihr Durchschnitt, die *durchschnittliche Pendlerquote der Arbeitsmarktregionen* bilden eine weitere Maßzahl für die Güte der Abgrenzungsverfahren. Analog wird die *durchschnittliche Pendlerquote der Kreisregionen* berechnet.

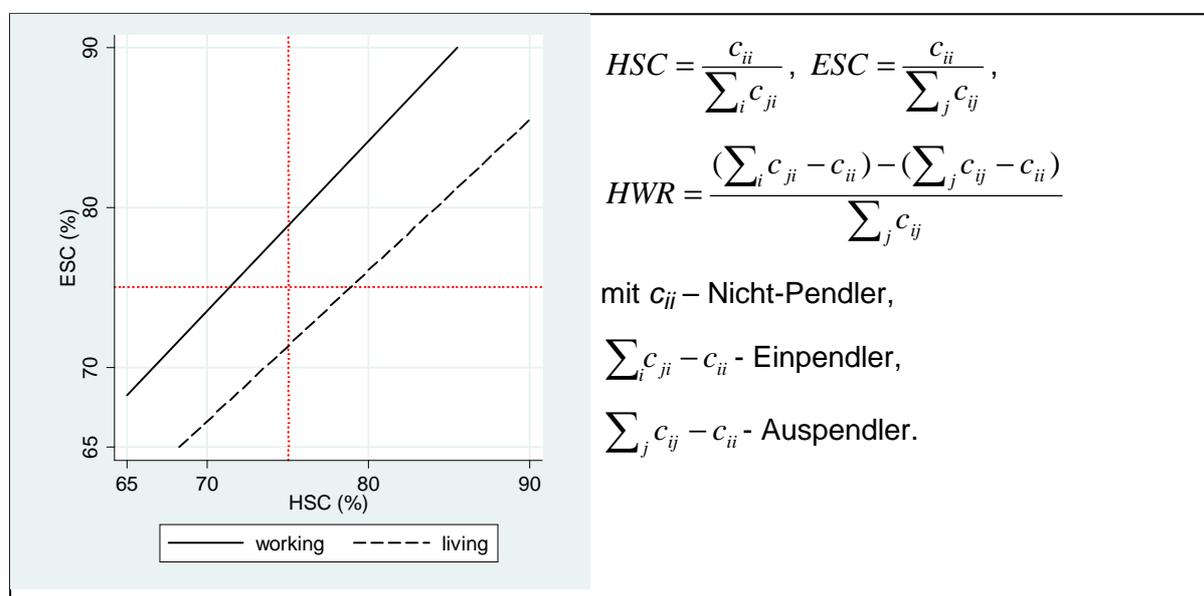
Die Güte dieser Maßzahlen korreliert im Wesentlichen mit der Modularität. Die Top 10 der Modularität haben in der Regel deutlich weniger als 20% Pendler zwischen den Regionen und sehr günstige Werte hinsichtlich der *durchschnittlichen Pendlerquoten der Kreisregionen*. Darum ergibt sich die Frage, ob Verfahren nicht grundsätzlich so gewählt werden sollten, dass die *Gesamtpendlerquote* minimiert wird. Das Problem eines solchen Vorgehens liegt darin, dass die beste analytische Lösung für diese Zielstellung in der Fusion zu einer einzigen Region besteht, die damit zwangsläufig nur noch "Binnenpendler" aufweist, aber sicherlich keine sinnvolle Arbeitsmarkt-Abgrenzung ist. Auch wenn man Verfahren so einschränkt, dass sie eine bestimmte Anzahl von Regionen erzeugen, ist keine sinnvolle Lösung sichergestellt. Die geringsten Gesamtpendlerquoten von unter 10% erzielten Clusterverfahren, die nahezu alle Kreisregionen mit großen Beschäftigtenzahlen zusammenfassten, im Wesentlichen eine große Region von München über Westdeutschland bis Hamburg bilden und eine Vielzahl einzelner Kreisregionen als Arbeitsmarktregionen "übrig" ließen (Anhang: Abbildung A1.d). Auch in diesem Falle kann nicht von einer sinnvollen Abgrenzung gesprochen werden (vgl. Kropp 2009).

6.2.3 Balance von Arbeitsangebot und –nachfrage (ESC, HSC, HWR)

Van der Laan und Schalke (2001: 210f) schlagen weitere Maße vor, die zeigen inwiefern Arbeitsangebot und –nachfrage in einer Region in Balance sind (dargestellt in Spalte 5-7 von Tabelle 12). Die Work-Home-Relation (**HWR**) berechnet sich aus der Differenz von Ein- und Auspendlern im Verhältnis zu allen in einer Region wohnenden Beschäftigten (siehe Abbildung 12). Die HWR ist negativ, wenn die Anzahl der Auspendler überwiegt (dann handelt es sich um eine Region mit dominanter Wohnfunktion) und positiv, wenn es mehr Einpendler gibt (dann handelt es sich um eine Region mit dominanter Arbeitsfunktion). Für gut abgegrenzte regionale Arbeitsmärkte liegt dieser Wert um Null.

In diesem Zusammenhang diskutieren van der Laan und Schalke (2001) weitere Maßzahlen, die die Geschlossenheit der Arbeitsmarktregionen widerspiegeln. Die housing self-containment Relation (**HSC**) gibt an, inwiefern Personen, die in einer Region arbeiten, auch dort wohnen. Wenn zu den in einer Region Wohnenden und Arbeitenden keine weiteren Beschäftigten hinzukommen, beträgt der HSC 100%; je weiter er darunter liegt, desto weniger stellt die Region den dort Arbeitenden auch Wohnungen zur Verfügung. Analog gibt die employment self-containment Relation¹⁷ (**ESC**) wieder, inwiefern eine Region genug Arbeitsplätze für die dort Wohnenden bereitstellt. Hohe Werte für beide Indikatoren geben an, inwiefern eine Region als geschlossen anzusehen ist, was auf eine gelungene Arbeitsmarkt-Abgrenzung hinweist. Van der Laan und Schalke (2001) kombinieren die beschriebenen Maße in einer Grafik, die auch in der vorliegenden Studie genutzt werden soll.

Abbildung 12: Relation zwischen ESC, HSC, HWR



Im ESC-HSC-Diagramm würden sich die Regionen mit ausgeglichener HWR genau auf der Diagonalen befinden. Erlaubt man geringfügige Abweichungen von 5% des ESC bzw. HSC, so ergeben sich entsprechende Grenzen, ab denen man von Regionen mit ausgesprochenem Wohncharakter sprechen kann (unterhalb der gestrichelten Linie), oder bei denen es sich eher um Arbeitsregionen handelt (oberhalb der durchgezogenen Linie). Das Merkmal der Geschlossenheit von Arbeitsmärkten wird über die 75%-Marke jeweils für ESC und HSC gekennzeichnet (gepunktete Linien). Das 75%-Kriterium wurde auch schon von Smart (1974) verwendet, während andere Autoren eher 70% wählen (vgl. van der Laan/Schalke 2001: 211). Die 75% Linie liegt genau zwischen 50 und 100%. Erst ab 50% kann man überhaupt von separaten Arbeitsmarktregionen sprechen,

¹⁷ Eine allgemeine Fassung des self containment nehmen Granato/Farhauer 2007 vor. Sie berechnen ihn aus dem Quotienten aus den Einpendlern durch die Auspendler. Für die Bewertung der Abgrenzungsverfahren liefert diese Maßzahl dieselben Ergebnisse wie der HWR.

da ansonsten mehr als die Hälfte der Beschäftigten bzw. Einwohner der Arbeitsmarktregion Pendler wären. Bei 100% würde eine vollständig geschlossene Arbeitsmarktregion vorliegen.

Da nach den oben genannten Argumenten dann von einer guten Arbeitsmarktabgrenzung gesprochen werden kann, wenn relativ geschlossene und "balancierte" Regionen abgegrenzt werden, müssten sich bei guten Abgrenzungen die Arbeitsmarktregionen vor allem im oberen rechten Teil der Grafik befinden, und zwar genau zwischen der durchgehenden und gestrichelten Linie. Die Visualisierung nach van der Laan und Schalke bietet einen guten Anhaltspunkt für eine Bewertung der qualitativen Eigenschaften der Abgrenzungsverfahren.

Abbildung 13 zeigt die Diagramme für die besten Abgrenzungen der vorgestellten Methoden. In der Titelzeile der einzelnen Diagramme sind auch Angaben dazu enthalten, wie viel Prozent der Regionen im rechten oberen Zielkorridor liegen. Als Vergleich können auch die Diagramme für administrative bzw. regionalpolitische Abgrenzungen in Abbildung 14 herangezogen werden. Zu beachten sind zum Teil unterschiedliche Skalierungen. Für die hier vorgestellten Algorithmen sind nur sehr geringfügige Unterschiede zu erkennen, z.B. der Ausreißer Emden für das Bassin-Modell oder die "Kette" von überregionalen Arbeitsmarktzentren bei faktorenanalytischen Abgrenzungen (Eckey u.a. sowie Angular-Methode). Bei allen Abgrenzungen liegt ein gewisser Bias in Richtung Wohnregionen vor (Regionen unterhalb der gestrichelten Linie), der sich auch in den durchweg negativen HWR-Durchschnittswerten (Spalte 7 in Tabelle 12) widerspiegelt. Dies erscheint durchaus plausibel, stehen doch in der Regel jedem einzelnen Arbeitsmarktzentrum eine Reihe von Regionen mit überwiegender Wohnfunktion gegenüber. Auch der Anteil der korrekt im Zielkorridor platzierten Regionen unterscheidet sich für die Abgrenzungen in Abbildung 13 nur graduell. Den besten Wert erzielt die Abgrenzung des alternativen Cluster-Verfahrens, das auf der Ein- und Auspendlermatrix beruht. Allerdings sind die ESC/HSC/HWR-Mittelwerte für dieses Verfahren nur durchschnittlich, so dass man dieses einzelne Ergebnis nicht überbewerten sollte. Deutlich bessere ESC/HSC/HWR-Mittelwerte weisen im Vergleich die faktorenanalytischen Verfahren auf.

Deutliche Unterschiede zeigen sich dagegen im Vergleich mit Abbildung 14. Zwar weisen die Raumordnungsregionen eine recht ausgewogene HWR und ein knappes Drittel aller Regionen innerhalb des Zielkorridors auf, generell zeigen sich aber wie der Abgrenzung für die Gemeinschaftsaufgabe viele Regionen mit ausgesprochenem Wohnort- oder Arbeitsortcharakter, wie es eher ungünstig für eine valide Arbeitsmarktregionen-Abgrenzung ist. Das zeigt sich auch im Vergleich mit den Ergebnissen für eine Zufallszuweisung von Kreisregionen zu 150 Arbeitsmarktregionen, für die die ESC/HSC/HWR-Mittelwerte nur unwesentliche schlechter sind als für die Abgrenzung für die Gemeinschaftsaufgabe.

Abbildung 13: ESC-HSC-Diagramme für ausgewählte Verfahren

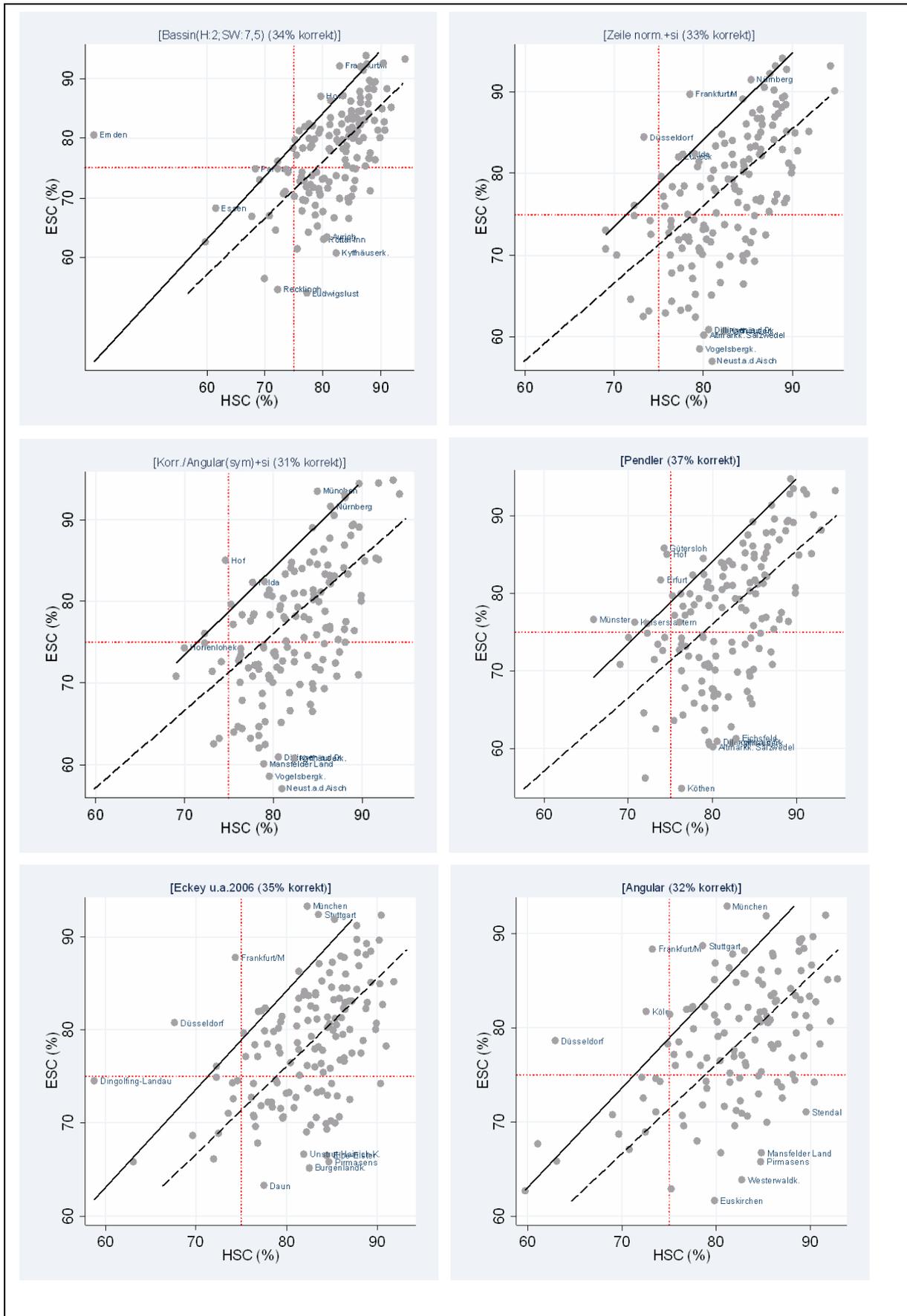
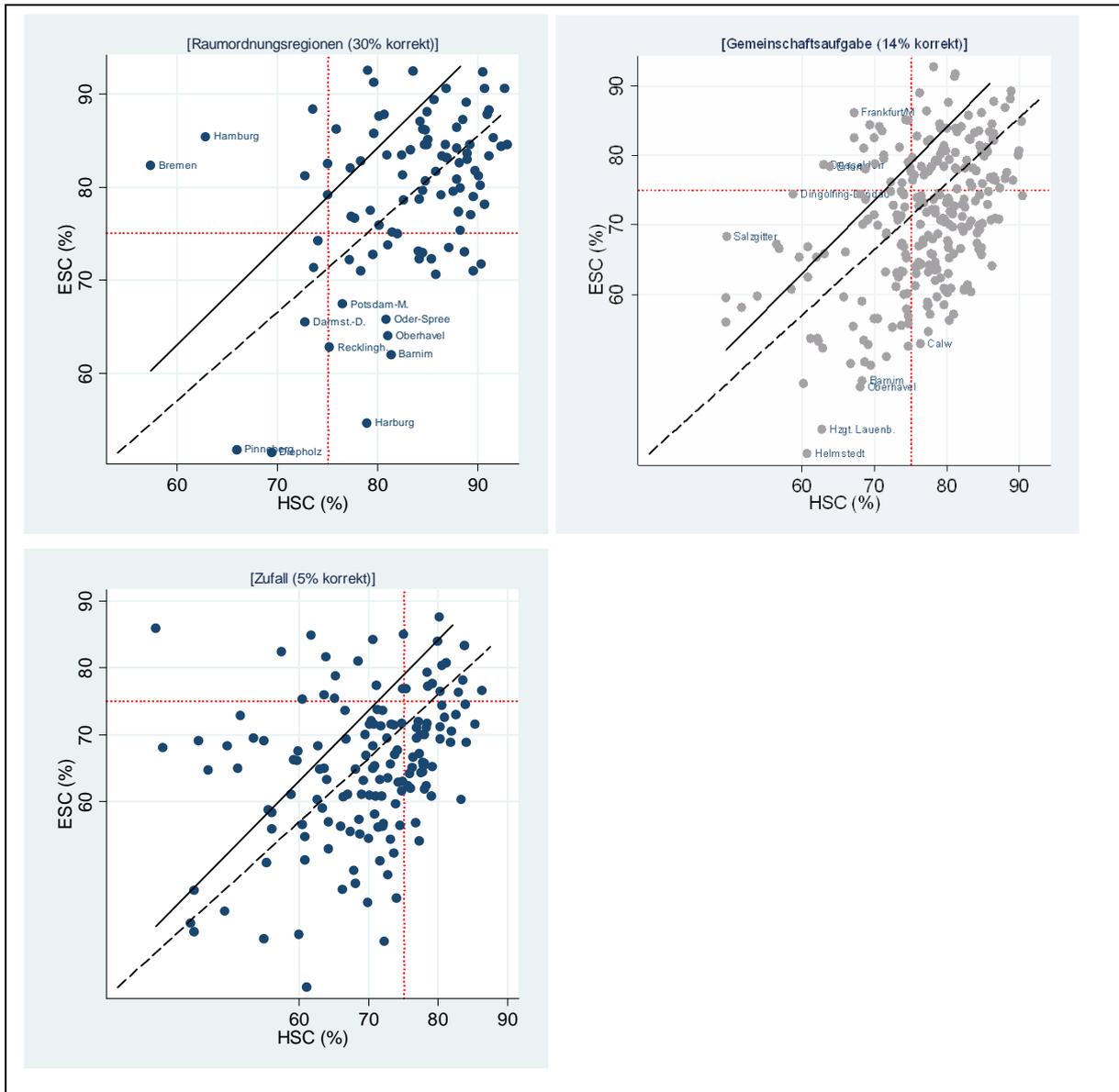


Abbildung 14: ESC-HSC-Diagramme für Vergleichsabgrenzungen



6.2.4 Die räumliche Struktur

Die letzten vier Spalten in Tabelle 12 erfassen quantitative Aspekte der räumlichen Struktur der Abgrenzungen. Ausgewiesen werden zunächst die Anzahl der Arbeitsmärkte in Spalte 8. Als Zielgröße waren 150 Regionen gesetzt, um den Vergleich mit Eckey u.a. 2006 zu ermöglichen. Bei Verfahren, die die Zahl der Regionen endogen bestimmen (Faktorenanalyse, graphentheoretisches Vorgehen), war dies jedoch nur bedingt möglich. Hinsichtlich der Größe der Regionen sind die Streuungsmaße interessant, denn sie geben an, ob ein Verfahren eher gleich große oder sehr unterschiedlich große Regionen generiert. Für möglichst größenhomogene Abgrenzungen sprechen große Minima, kleine Maxima sowie eine geringe Standardabweichung. Die Standardabweichung ist in Spalte 9 wiedergegeben. Es zeigt sich, dass Abgrenzungen, die die Pendlerströme gut erfassen, eher heterogene Arbeitsmarktregionen aufweisen (Spalte 1-4 in Tabelle 12). Für hohe Heterogenität

spricht auch eine große Zahl von Arbeitsmarktregionen, die ausschließlich aus einer Kreisregion bestehen (dargestellt in Spalte 10: *Einzel*).

Die letzte Spalte "*nicht 1.*" gibt die Anzahl der Kreisregionen an, die einer Arbeitsmarktregion zugeordnet wurden, zu der sie nicht die höchste Auspendlerzahl haben (bzw. den maximalen Auspendleranteil). Dieses Kriterium erscheint als ein naheliegendes Kriterium für die Güte einer Abgrenzung. Tatsächlich zeigen die unterschiedlichen Verfahren eine hohe Streuung hinsichtlich dieses Merkmals. Die nähere Betrachtung zeigt jedoch, dass dieses Merkmal bestimmte Verfahren "benachteiligt". Zunächst haben natürlich alle Kreisregionen keine Auspendler in den eigenen Arbeitsmarkt, wenn dieser nur aus genau dieser einen Kreisregion besteht. Für diese "Einzel"-Regionen (oder "isolates" im graphentheoretischen Verfahren) gibt es also immer mehr Auspendler in andere Arbeitsmarktregionen als in die eigene. Darum wurden diese Fälle nicht für den Indikator berücksichtigt. Dennoch bleibt ein Bias. Bildet ein Verfahren nämlich zahlreiche kleinere Arbeitsmärkte, ohne dass diese "Einzel"-Regionen wären, so kommt es relativ häufig vor, dass die Zentren dieser kleinen Arbeitsmärkte stärker mit benachbarten größeren Arbeitsmarktzentren verbunden sind als mit dem eigenen Umland. So bilden beispielsweise zahlreiche Verfahren die Arbeitsmarktregion Lübeck aus den Kreisen Ostholstein und Lübeck. Lübeck hat allerdings mehr Auspendler nach Hamburg als nach Ostholstein. Allerdings ist der Auspendleranteil nicht so hoch, dass für die entsprechenden Verfahren bereits eine Fusion der Arbeitsmarktregion Lübeck mit der Arbeitsmarktregion Hamburg zu rechtfertigen wäre. Dieses Problem kommt kaum vor bei Verfahren, die viele "Einzel"- oder kleine Arbeitsmarktregionen und einige wenige große Arbeitsmarktregionen bilden. Wegen dieses Bias ist die Anzahl der Kreisregionen, die einer Arbeitsmarktregion zugeordnet wurden, zu der sie nicht die höchste Auspendlerzahl haben, eher ein Indikator für die Größenstreuung als die Güte der Abgrenzung.¹⁸

6.3 Vergleich der Bewertungsmaße

Abschließend werden die vorgestellten Bewertungsmaße in Tabelle 13 verglichen. Im Vergleich zu Tabelle 12 wurde das Kriterium der korrekt im Zielkorridor der ESC-HSC-Diagramme platzierten Regionen hinzugefügt. Es zeigt sich, dass die Modularität Q, die Pendlerquoten gesamt und für den Durchschnitt der Kreise in der Tat einen sehr hohen Zusammenhang aufweisen. Zu Berücksichtigen ist, dass für die Modularität Q hohe Werte für eine gute Abgrenzung sprechen, während bei den Pendlerquoten niedrige Werte günstig sind.

Die durchschnittlichen Pendlerquoten für die Arbeitsmarktregionen erweisen sich dagegen eher als ein Maße für die Autonomie bzw. Eigenständigkeit der Regionen, denn

¹⁸ Alle betrachteten Abgrenzungsverfahren weisen die Kreisregionen Arbeitsmarktregionen zu, in denen auch die meisten der in der Kreisregion wohnenden Beschäftigten tätig sind. So hat Lübeck zwar mehr Auspendler nach Hamburg als nach Ostholstein, aber es sind immer noch deutlich mehr Lübecker in der Lübecker Arbeitsmarktregionen (bzw. schon allein in der Stadt Lübeck) tätig, als in Hamburg.

sie korrelieren deutlich höher mit den ESC- und HSC-Durchschnitten als mit den anderen Pendlermaßen. Auch hier ist das negative Vorzeichen der unterschiedlichen Orientierung der Maße geschuldet. Unter den Maßen für die Autonomie bzw. Eigenständigkeit fällt der HWR-Durchschnitt auf. Wie die Diagramme in Abschnitt 6.2.3 gezeigt haben, ist dieses Maß weniger ein Kriterium für die Güte einer Abgrenzung sondern eher ein Indikator für einen Bias in Hinblick auf den Wohn- bzw. Arbeitscharakter von Regionen. Interessanter Weise korreliert er moderat mit Indikatoren aus allen drei Indikatorengruppen, weist aber keine spezielle Trennschärfe auf. Ähnliches gilt für das neue Kriterium der korrekt im Zielkorridor der ESC-HSC-Diagramme platzierten Regionen. Die für die Konstruktion dieses Indikators verwendeten ESC und HSC-Daten sind für sich aussagekräftiger.

Die Anzahl der Regionen einer Abgrenzung sollte nicht zwangsläufig in einem Zusammenhang mit den Qualitätskriterien stehen. Wegen der geringen Varianz dieses Maßes sollten die Korrelationskoeffizienten nur mit Vorsicht interpretiert werden. Die zwar nur moderaten, aber deutlichen Zusammenhänge dürften der Tatsache geschuldet sein, dass kleinere Regionenanzahlen nur bei den faktorenanalytischen Abgrenzungen vorkamen, die die Pendlerströme in der Regel nicht so gut erfassen wie andere in Tabelle 12 präsentierte Verfahren.

Tabelle 13: Vergleich der Bewertungsmaße (Korrelationstabelle zu Tabelle 12, ohne Vergleichsregionen)

	Erfassen der Pendlerströme				Mittelwerte für:				Räumliche Struktur			
	Q Wert	Pendlerquoten gesamt	Ø AMR	Ø KR	ESC	HSC	HWR	% Zielkorridor	AMR	StAbw.	Einzel	nicht 1.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)		(8)	(9)	(10)	(11)
Erfassen der Pendlerströme												
Modularität Q	1,00	-0,88	-0,08	-0,84	-0,11	0,30	-0,54	0,35	0,58	0,64	0,82	-0,83
Pendler ges.	-0,88	1,00	0,03	0,98	0,20	-0,25	0,61	-0,24	-0,53	-0,93	-0,90	0,87
Ø AMR	-0,08	0,03	1,00	0,16	-0,95	-0,90	-0,37	-0,53	0,25	0,05	0,21	-0,26
Ø KR	-0,84	0,98	0,16	1,00	0,08	-0,35	0,53	-0,29	-0,49	-0,92	-0,88	0,84
Autonomie/Eigenständigkeit												
Ø ESC	-0,11	0,20	-0,95	0,08	1,00	0,73	0,64	0,60	-0,41	-0,27	-0,42	0,42
Ø HSC	0,30	-0,25	-0,90	-0,35	0,73	1,00	-0,05	0,32	-0,03	0,13	-0,01	0,10
Ø HWR	-0,54	0,61	-0,37	0,53	0,64	-0,05	1,00	0,50	-0,59	-0,56	-0,62	0,54
% Zielkorridor	0,35	-0,24	-0,53	-0,29	0,60	0,32	0,50	1,00	0,03	0,10	0,14	-0,21
Räumliche Struktur												
AMR	0,58	-0,53	0,25	-0,49	-0,41	-0,03	-0,59	0,03	1,00	0,41	0,74	-0,70
StAbw.	0,64	-0,93	0,05	-0,92	-0,27	0,13	-0,56	0,10	0,41	1,00	0,84	-0,77
Einzel	0,82	-0,90	0,21	-0,88	-0,42	-0,01	-0,62	0,14	0,74	0,84	1,00	-0,98
nicht 1.	-0,83	0,87	-0,26	0,84	0,42	0,10	0,54	-0,21	-0,70	-0,77	-0,98	1,00

Die Schlussfolgerung aus den vorangegangenen Abschnitten, dass größenhomogene Abgrenzungen die Pendlerverflechtungen nicht so gut erfassen können wie größenheterogene, bestätigt sich in Tabelle 13. Eine hohe Standardabweichung hinsichtlich der sozialversicherungspflichtig Beschäftigten korreliert moderat mit einem

hohen Modularitätswert und sehr stark mit einer niedrigen Gesamtpendlerquote und dem Durchschnitt der Pendlerquoten auf der Ebene der Kreisregionen.

Schließlich erlauben die Resultate der Korrelationsanalyse auch eine Präzisierung der Schlussfolgerungen aus Abschnitt 6.2.4. Der letzte Indikator "*nicht 1.*", der die Anzahl der Kreisregionen angibt, die einer Arbeitsmarktregion zugeordnet wurden, zu der sie nicht die höchste Auspendlerzahl haben, ist *sowohl* ein Indikator für die Güte der Abgrenzung in Hinblick auf die Erfassung der Pendlerverflechtungen als auch ein Indikator für die Heterogenität.

Zusammenfassend bestätigen die Ergebnisse, dass Abgrenzungsverfahren auf der bestehenden Datengrundlage nicht zugleich die Pendlerstruktur adäquat erfassen und relativ größenhomogene Regionen abgrenzen können. Diese Realität müssen sowohl die Regionalforschung wie auch die Regionalpolitik Rechnung tragen.

7 Zusammenfassung

In der vorliegenden Arbeit wurden verschiedene Verfahren zur Abgrenzung von funktionalen Regionen, den Arbeitsmarktregionen, vorgestellt und miteinander verglichen. Dabei handelte es sich sowohl um Schwellenwertmodelle der Graphentheorie als auch um verschiedene Verfahren der Clusteranalyse und Faktorenanalyse. Datenbasis hierfür waren die Pendlerverflechtungen der sozialversicherungspflichtigen Beschäftigten zwischen den Kreisen vom 30.06.2005 aus der Beschäftigtenstatistik der Bundesagentur für Arbeit. Um einen Vergleich mit dem aktuellsten gesamtdeutschen Abgrenzungsvorschlag von Eckey et al. (2006) zu ermöglichen, wurde versucht, mit den verschiedenen Verfahren ebenfalls 150 Arbeitsmarktregionen zu generieren. Bei den Verfahren der Clusteranalyse war dies exakt möglich, bei der Schwellenwertmethode und der Faktorenanalyse konnte hingegen die Zahl der Arbeitsmarktregionen nur annähernd erreicht werden, da sie endogen bedingt waren.

Die Güte der einzelnen Abgrenzungen wurde anhand verschiedener Koeffizienten bewertet, wie z.B. der Modularität Q und der employment self-containment Relation (ESC) und der housing self-employment Relation (HSC). Während die Modularität Q ein Indikator dafür ist, wie viel besser die Abgrenzung der Arbeitsmarktregionen gegenüber einer Zufallseinteilung ist, geben die Verhältniszahlen ESC und HSC Auskunft darüber, wie gut die neu gebildeten Arbeitsmarktregionen die Arbeitsort- und Wohnfunktion einfangen oder ob in den Regionen eher der Charakter des Arbeitsorts oder des Wohnorts dominiert. Mit weiteren Maßzahlen lassen sich ergänzend die Pendlerströme und die räumliche Struktur bewerten.

Mit allen hier vorgestellten Verfahren ließen sich Abgrenzungen von Wirtschaftsräumen vornehmen, die die Realität der Pendlerverflechtungen zwischen Kreise deutlich besser einfangen als administrative Abgrenzungen und auf spezifische Sachanforderungen zurechtgeschnittene funktionale Gliederungen. Besonders die Gliederung der Gemeinschaftsaufgabe schneidet im Vergleich der funktionalen Gliederungen ungünstig ab, während die Raumordnungsregionen hinsichtlich zahlreicher Kriterien mit hier präsentierten Ergebnissen mithalten können.¹⁹

Die im letzten Abschnitt gezeigten Vergleichskriterien erlauben die Bewertung der Abgrenzungsverfahren nach verschiedenen Gesichtspunkten. Am erfolgreichsten beim "Einfangen" von Pendlern in Arbeitsmarktregionen waren die Cluster-Verfahren. Bei den besten Verfahren blieben nur noch ca. 15% aller sozialversicherungspflichtig Beschäftigten als Pendler zwischen den Arbeitsmarktregionen. Hinsichtlich der

¹⁹ Zu beachten ist bei diesem Vergleich allerdings die Anzahl der Regionen. Wenn die Kreise - wie bei den 97 Raumordnungsregionen - in vergleichsweise wenige Regionen zusammengefasst werden, bleiben deutlich weniger Pendler zwischen den Regionen übrig, als bei einer feingliedrigern Abgrenzungen wie der für die Gemeinschaftsaufgabe (270), Vergleichbare Ergebnisse liefert darum noch am ehesten die Modularität Q.

Modularitätswerte lieferten sowohl die Clusterverfahren als auch die graphentheoretische Methode sehr gute Ergebnisse. Ein wichtiger Grund für die guten Ergebnisse der alternativen Cluster-Verfahren und der graphentheoretische Methode liegt sicherlich darin, dass sie die Informationen der Pendlermatrix direkt verarbeiten können. Für die Cluster- und Faktorenanalyse sind dagegen verschiedene Matrixoperationen erforderlich. Insbesondere die Beschränkung dieser Verfahren, asymmetrische Pendlerdaten nur indirekt (z.B. über symmetrische Zusammenhangsmaße) verarbeiten zu können, bringt einen Informationsverlust mit sich.

Ein wichtiges Ergebnis der Studie ist es, dass gute analytische Lösungen des Problems, die Pendlerverflechtungen möglichst gut "einzufangen", in Abgrenzungen mit einigen sehr großen Arbeitsmarktregionen resultieren, die für regionalpolitische Zwecke eher ungünstig sein dürften. Scheinbar widersprechen solche großen Räume der Vorstellung, Pendlerverflechtungen adäquat widerzuspiegeln, denn Pendlerbewegungen über die gesamte Fläche erscheinen sehr unplausibel. Andererseits bestätigen empirische Befunde durchaus, dass um die großen Metropolen der Bundesrepublik tatsächlich ein großer Anteil von Beschäftigten weit überdurchschnittliche Pendelentfernungen auf sich nimmt. Zudem resultieren die weiträumigen Abgrenzungen nicht nur aus den direkten Pendlerbeziehungen zwischen Regionen, sondern auch aus indirekten. Wenn z.B. nicht nur aus den München umgebenden Kreisen nach München gependelt wird, sondern auch in starkem Maße aus weiter entfernten Regionen in den Münchner Umkreis, dann liegen solche indirekten Verflechtungen vor, die die gezeigten Ergebnisse durchaus inhaltlich bestätigen.

Für die praktische Anwendbarkeit der Verfahren spielen zusätzliche Kriterien eine Rolle. Die alternativen Cluster-Verfahren und die graphentheoretische Methode sind hinsichtlich ihrer Methodik relativ einfach und darum gut an bestimmte Anforderungen anpassbar. Durch die Gewichtung von Pendlerzahlen mit den Pendlerentfernungen konnten sehr einfach größenhomogenere Abgrenzungen erreicht werden (wenn auch um den Preis höherer Pendlerzahlen zwischen den Regionen). Die hierarchischen Clusterverfahren - und mit Einschränkungen auch der graphentheoretische Ansatz - weisen zu dem die Möglichkeit auf, das Verfahren zu unterschiedlichen Zeitpunkten abzurechnen, was in feineren oder gröberen Gliederungen resultiert, je nachdem, welche praktischen Anforderungen an eine Gliederung bestehen. Ein weiterer Vorzug der einfacheren Verfahren liegt zudem darin, dass ihr Vorgehen auch Laien gut erklärt werden kann. In dieser Hinsicht bieten die Flussdiagramme des Konzepts der dominanten Flüsse oder die Dendrogramme der Clusterverfahren gut Möglichkeiten.

Die letztendliche Entscheidung für ein bestimmtes Verfahren hängt jedoch stark davon ab, für welchen Zweck die Abgrenzung der Arbeitsmarktregionen erfolgen soll. Soll strikt methodisch vorgegangen und an dem Kriterium der Pendlerminimierung festgehalten werden und damit sehr unterschiedliche große Regionen in Kauf genommen werden oder sind eher für die Praxis verwendbare größenhomogene Regionen gewünscht? Danach entscheidet sich auch, welchem Gütemaß der Vorzug gegeben wird.

Gerade die im politischen Bereich verwendeten Abgrenzungen von funktionalen Regionen, wie z.B. die Arbeitsmarktregionen der Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur“ oder die Raumordnungsregionen für die Raumordnungsberichterstattung des Bundes, unterliegen häufig bestimmten Vorgaben und sind daher nicht frei wie rein wissenschaftliche Abgrenzungen. Aus diesem Grund stellt sich auch immer die Frage der praktischen Anwendung.

Im Fokus der vorliegenden Studie stand der Methodenvergleich. Darum wurden in Anlehnung an die derzeit prominenteste Arbeitsmarktgliederung von Eckey u.a. 2006 die Pendlerdaten auf der Grundlage von Kreisregionen als Basis und die Anzahl von 150 Regionen als Referenz gewählt. Für eine optimale Abgrenzung von Arbeitsmarktregionen sind diese Voraussetzungen jedoch nicht geeignet. Bisherige Abgrenzungen regionaler Arbeitsmärkte erfolgten auf der Kreisebene. Da es zwischenzeitlich einen Neuzuschnitt von Kreisen in zwei Bundesländern gegeben hat, die die ursprünglichen Kreisgrenzen durchschneiden, und weitere Kreisgebietsreformen²⁰ geplant sind, ist eine Aktualisierung bisher vorliegender Abgrenzungen von funktionalen Arbeitsmarktregionen unausweichlich. Dabei ist abzuwägen, ob bei einer neuen Abgrenzung nicht besser die Pendlerverflechtungen auf der Gemeindeebene untersucht werden sollten, zumal gerade die Kreisgebietsreformen zu weniger und damit auch größeren Einheiten geführt haben, die sich nicht immer an Pendlerbeziehungen orientieren. Wichtig bei einer gemeindebezogenen Abgrenzung ist jedoch, dass auch die später genutzten Daten auf der Gemeindeebene vorliegen. Auf der Grundlage von Arbeitsmarkt-Abgrenzungen auf Gemeindeebene ist es auch immer möglich, kreisscharfe Abgrenzungen oder Abgrenzungen auf anderen Aggregationsebenen vorzunehmen und zugleich deren Probleme zu diagnostizieren.

Des Weiteren ist nach der optimalen Anzahl von Regionen zu suchen. Die Darstellung des Verlaufes für die Modularität Q in Abbildung 8 weist darauf hin, dass die optimale Anzahl für eine gute Arbeitsmarkt-Abgrenzung deutlich unter 150 liegen dürfte. Auch Varianten des graphentheoretischen Ansatzes weisen bei kleineren Regionenanzahlen bessere Modularitätswerte auf (z.B. $Q=0,829$ bei $N=96$). Die optimale Anzahl von Regionen lässt sich z.B. anhand der Verlaufswerte von Q bestimmen. Aber auch andere Verfahren, wie die Reduzierung der räumlichen Autokorrelation von wirtschaftlichen Zusammenhängen, können hier angewandt werden.²¹ Falls feingliedrigere

²⁰ Zum 1. Juli 2007 fand eine Kreisgebietsreform in Sachsen-Anhalt und am 1. August 2008 in Sachsen statt. Dabei wurden in Sachsen-Anhalt neben den 3 kreisfreien Städten aus ehemals 21 Landkreisen 11 neue Landkreise gebildet. Auch in Sachsen reduzierte sich die Zahl der Landkreise von 22 auf 10, wobei auch hier die 3 kreisfreien Städte erhalten blieben. Für 2009 plant NRW die Bildung der Städteregion Aachen aus der Stadt Aachen und dem Kreis Aachen, 2011 könnten in Mecklenburg-Vorpommern die bisher 12 Kreise zu 5 zusammenzufassen werden. Auch in Schleswig-Holstein gibt es Überlegungen, 2013 die Zahl der Kreise auf vier bis sechs und die Zahl der kreisfreien Städte auf zwei oder keine zu reduzieren (vgl. <http://de.wikipedia.org>, Sichtworte: Kreisreform Mecklenburg-Vorpommern 2011, Schleswig-Holstein, Kreisreformen in der Bundesrepublik Deutschland; abgerufen: 26. August 2008, 13:37).

²¹ Siehe Fertig u.a. 2006 für ein Beispiel.

Abgrenzungen nötig sind, lassen sich für große Arbeitsmarktregionen auch lokale Teilarbeitsmärkte bestimmen, ohne dass der Gesamtzusammenhang verloren geht.

Zur Optimierung der Abgrenzung von Wirtschaftsräumen können auch direkte Optimierungsverfahren eingesetzt werden. Partitionierende Clusterverfahren sind hierfür eine Option oder Korrekturverfahren, die Fehlzusammenhänge, wie sie in hierarchischen Verfahren auftreten können, korrigieren. Für die Auswahl einer optimalen Abgrenzung spielt schließlich auch ihre Robustheit gegenüber kleineren Schwankungen in den Pendlerzahlen eine Rolle.

Van der Laan und Schalke bemerkten bereits 2001 kritisch, dass jedes Land oder Region seine AMR nach anderen Kriterien abgrenzt, und verwiesen auf das Problem, dass dadurch der Vergleich auf europäischer Ebene nur schwer möglich ist (2001: 219). Würde es einheitliche Kriterien zur Abgrenzung von AMR geben, wäre z.B. auch die Vergabe von Fördergeldern für Regionen anhand von Kennziffern (wie AloQ, BIP etc.) objektiver. Die derzeitige Vergabe orientiert sich an der NUTS-2-Ebene, also an sehr großen und unfunktional geschnittenen Regionen. Zudem erschwert die uneinheitliche Klassifikation auch EU-weite Evaluationen.

Es bleibt zu hoffen, dass die vorliegende Arbeit einen Beitrag für eine valide Abgrenzung von Wirtschaftsräumen in Deutschland und möglicherweise auch in Europa leisten kann.

Literaturverzeichnis

- Backhaus, K./Erichson, B./Plinke, W./Weiber, R. (2000): *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* Berlin: Springer.
- Barkley, D.L./Henry, M.S./Bao, S./Brooks, K.R. (1995): How Functional are Economic Areas? Tests for Intra-Regional Spatial Association Using Spatial Data Analysis. *Papers in Regional Science* 74: 297-316.
- Baumann, J.H./Fischer, M.M./Schubert, U. (1983): A multiregional labour supply model for Austria: The effects of different regionalisations in multiregional labour market modelling. *Papers in Regional Science*, 52: 53-83.
- Bergs, S. (1981): *Optimalität bei Clusteranalysen. Experimente zur Bewertung numerischer Klassifikationsverfahren*. Münster (Westfalen), Diss.
- Binder, J./Schwengler, B. (2006): *Arbeitsmarktregionen im Raum Berlin und Brandenburg - Kritische Überprüfung der bisher gültigen Arbeitsmarktregionen und Vorschläge für einen Neuzuschnitt*. IAB Forschungsbericht Nr. 4 / 2006.
- Blien, U./Hirschenauer, F. (2005): *Vergleichstypen 2005. Neufassung der Regionaltypisierung für Vergleiche zwischen Agenturbezirken*. IAB-Forschungsbericht 24, Nürnberg.
- Bode, E. (2008): Delineating Metropolitan Areas using Land Prices. *Journal of Regional Science*, 48 (1): 131-163.
- Böltken, F. (1996): *Neuabgrenzung von Raumordnungsregionen nach den Gebietsreformen in den neuen Bundesländern*. Arbeitspapier 5/1996 der Bundesforschungsanstalt für Landeskunde und Raumordnung, Bonn.
- Bongaerts, D./Cörvers, F./Hensen, M. (2004): *The Delimitation and Coherence of Functional and Administrative Regions*, Research Series 04O19, Ministry of Economic Affairs, The Hague.
- Boudeville, J.-R. (1966): *Problems of Regional Economic Planning*. Edinburgh. University Press.
- Brandes, U./Delling, D./Gaertler, M./Görke, R./Hofer, M./Nikoloski, Z./Wagner, D. (2008): On Modularity Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 20: 172-188.
- Budde, R./Eckey, H.-F./Klemmer, P. (1993): *Vorschlag für die Abgrenzung von Arbeitsmarktregionen in den neuen Bundesländern*. *Gutachten des Rheinisch-Westfälischen Instituts für Wirtschaftsforschung*. Essen. 78 S.
- Budde, R./Eckey, H.-F./Schrumpf, H. (1995): *Vorschlag für die Abgrenzung von Arbeitsmärkten in den neuen Bundesländern auf der Grundlage der durchgeführten Regionalreformen*. *Gutachten des Rheinisch-Westfälischen Instituts für Wirtschaftsforschung*. Essen. 90 S.

- Bundesministerium für Raumordnung, Bauwesen und Städtebau (2005): Raumordnungsbericht 2005, Bonn.
- Borgatti, S.P./Everett, M.G./Freeman, L.C. (2002): Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.
- BVU, ifo, ITP, PLANCO (2001): Verkehrsprognose 2015 für die Bundesverkehrswegeplanung im Auftrag des Bundesministeriums für Verkehr, Bau und Wohnungswesen. München, Freiburg, Essen.
- Casado-Díaz, J.M. (2000): Local labour market areas in Spain: a case study. *Regional Studies* 34 (9): 843-856.
- Eckey, H.-F. (1988): Abgrenzung regionaler Arbeitsmärkte. *Raumforschung und Raumordnung*, H. 1-2: 24-33.
- Eckey, H.-F. (2001): Der wirtschaftliche Entwicklungsstand in den Regionen des vereinigten Deutschlands. *Volkswirtschaftliche Diskussionsbeiträge 20/01 der Universität Kassel*. 63 S.
- Eckey H.-F./ Klemmer, P (1991): Neuabgrenzung von Arbeitsmarktregionen für die Zwecke der regionalen Wirtschaftspolitik, *Informationen zur Raumentwicklung* H. 9/10: 569-577.
- Eckey, H.-F./Kosfeld, R./Rengers, M. (2002): *Multivariate Statistik* Wiesbaden: Gabler-Verlag.
- Eckey, H.-F./Kosfeld, R./Türck, M. (2006): Abgrenzung deutscher Arbeitsmarktregionen, *Raumforschung und Raumordnung* Jg. 64, H. 4: 299-309.
- Fertig, M./Kluve, J./Schmidt, Chr.M. (2006): Die makroökonomische Wirkung aktiver Arbeitsmarktpolitik. Eine Panelanalyse auf Ebene regionaler Arbeitsmärkte. In: *Zeitschrift für ArbeitsmarktForschung*, Jg. 39, H. 3/4, S. 575-601.
- Goetz, S.J. (1999): Migration and Local Labor Markets. In *The Web Book of Regional Science* 2nd 2001 ed.: Scott Loveridge. Morgantown, WV: Regional Research Institute, West Virginia University.
- Granato, N./Farhauer, O. (2007): Die Abgrenzung von Arbeitsmarktregionen: Gütekriterien und –maßzahlen. *Wirtschaftswissenschaftliche Dokumentation. Diskussionspapier 2007/02*. Technische Universität Berlin.
- Gordon, A. D. (1999): *Classification*. Boca Raton, 2nd edition.
- Haas, Anette; Hamann, Silke (2008): Pendeln - ein zunehmender Trend, vor allem bei Hochqualifizierten * Ost-West-Vergleich. (IAB-Kurzbericht, 06/2008), Nürnberg.
- Hansen, N. M.: An Evaluation of Growth-Centre Theory and Practice. In: *Environment and Planning* 7, 1975, S. 821-832.
- Hensen, M./Cörvers, F. (2003): The regionalization of labour markets by modelling commuting behaviour. Maastricht.

- Initiativkreis Metropolregionen in Deutschland (2003): Strategiepapier. Technische Universität Berlin.
- Junkernheinrich M., Skopp R.: Wirtschaftliche Konvergenz und räumliche Wachstumspole. Zur regionalökonomischen Lage in Ostdeutschland. In: Pohl R. (Hrsg.), Herausforderung Ostdeutschland. Fünf Jahre Währungs-, Wirtschafts- und Sozialunion. Berlin 1995, S. 217-238.
- Kaufman, L./Rousseeuw, P.J. (1990): Finding Groups in Data: An Introduction to Cluster Analysis. New York: John Wiley and Sons.
- Karlsson, Ch./Olsson, M. (2006): The identification of functional regions: theory, methods, and applications. *Annals of Regional Science* 40 (1): 1-18.
- Keilbach M. (2000): Spatial Knowledge Spillovers and the Dynamics of Agglomeration and Regional Growth, Diss., Heidelberg, New York 2000.
- Klemmer, P./Kraemer, D. (1975): Regionale Arbeitsmärkte. Ein Abgrenzungsvorschlag für die Bundesrepublik Deutschland. *Beiträge zur Struktur- und Konjunkturforschung*. Band 1. Brockmeyer-Verlag. Bochum. 282 S.
- Kropp, P. (2009): Regional Labour Markets in Germany - Delimitation and Structural Properties. *Journal of Economic & Social Geography*. (im Erscheinen).
- Kropp, P./Niebuhr, A. (2006): Vergleichende Analyse von Länderarbeitsmärkten. Hamburg und die Metropolregion. Nürnberg. IAB regional. Berichte und Analysen.
- Moseley, M. J. (1974): Growth Centres in Spatial Planning, Oxford.
- Newman, M.E.J. (2006): Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103: 8577-8582.
- Newman, M.E.J./Girvan, M. (2004): Finding and evaluating community structure in networks. *Phys. Rev. E* 69 (2004) 026113.
- Nystuen, J.D./Dacey, M.F. (1961): A Graph Theory Interpretation of Nodal Regions. *Papers and Proceedings of the Regional Science Association* 7 (1): 29-42.
- OECD (2002): Redefining territories, the functional regions, Paris.
- ONS (Office for National Statistics) und M. G. Coombes (1998): 1991-based Travel-to-Work Areas. London. Office for National Statistics.
- Openshaw, S./Taylor, P. (1981): The modifiable area unit problem. *Quantitative Geography: A British View*. N. Wrigley. ed., (London: Routledge and Regan Paul).
- Smart, M. W. (1974): Labour market areas: uses and definition, *Progress in Planning*, 2 (4): 239-353.
- StataCorp (2005): Multivariate Statistics Reference Manual. College Station, TX : Stata Press.
- Tolbert, C.M./Killian, M.S. (1987): Labor Market Areas for the United States. Staff Report. Washington, D.C.: USDA, ERS, Agriculture and Rural Economy Division.

- van der Laan, L. /R. Schalke (2001): Reality versus Policy: The Delineation and Testing of Local Labour Market and Spatial Policy Areas, *European Planning Studies*, Vol. 9: 201-221.
- van Nuffel, N. (2007): Determination of the Number of Significant Flows in Origin – Destination Specific Analysis: The Case of Commuting in Flanders. *Regional Studies* 41 (4): 509-524.

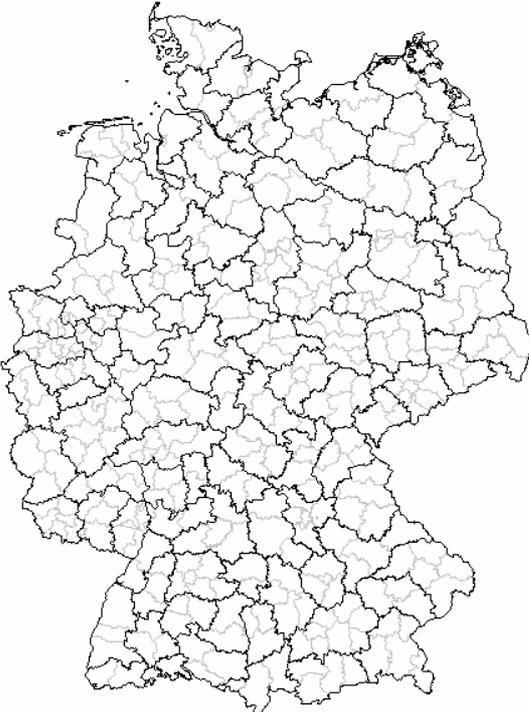
Anhang

Tabelle A1: Die 20 stärksten Pendlerströme bzw. Anbindungskoeffizienten

von	nach	Pendler (Rang)		Anbindungskoeffizient (Rang)	
Erftkreis	Köln	52.015	(1)	36,1%	(18)
Ludwigsburg.	Stuttgart	44.101	(2)	23,8%	(61)
Offenbach	Frankfurt/M	43.317	(3)	28,2%	(41)
Mettmann	Düsseldorf	39.284	(4)	24,1%	(59)
Esslingen	Stuttgart	38.658	(5)	21,0%	(81)
Pinneberg	Hamburg	38.433	(6)	39,3%	(11)
Harburg	Hamburg	38.406	(7)	49,7%	(5)
Bonn	Köln	36.465	(8)	13,4%	(152)
Fürstfeldbruck	München	36.294	(9)	53,2%	(2)
Fürth	Nürnberg	34.618	(10)	42,5%	(9)
Neuss	Düsseldorf	32.926	(11)	23,1%	(65)
Stormarn	Hamburg	32.533	(12)	43,9%	(7)
Rems-Murr-Kreis	Stuttgart	32.097	(13)	22,2%	(69)
Main-Taunus-Kreis	Frankfurt/M	30.965	(14)	38,5%	(13)
Rhein-Neckar-Kreis	Mannheim	29.350	(15)	13,5%	(148)
Main-Kinzig-Kreis	Frankfurt/M	29.238	(16)	21,3%	(76)
Segeberg	Hamburg	27.996	(17)	31,8%	(28)
Dachau	München	26.361	(18)	53,0%	(3)
Wetteraukreis	Frankfurt/M	24.505	(19)	25,3%	(50)
Oberhavel	Berlin	24.437	(20)	37,4%	(14)
Ebersberg	München	24.193	(21)	56,4%	(1)
Gifhorn	Wolfsburg	23.315	(24)	39,5%	(10)
Freising	München	21.532	(28)	35,3%	(19)
Barnim	Berlin	21.502	(29)	37,2%	(15)
Eichstätt	Ingolstadt	20.366	(31)	46,3%	(6)
Osterholz	Bremen	18.129	(40)	50,8%	(4)
Starnberg	München	16.704	(43)	43,8%	(8)
Aichach-Friedberg	Augsburg	15.138	(48)	35,0%	(20)
Plön	Kiel	14.505	(58)	38,9%	(12)
Helmstedt	Wolfsburg	11.262	(92)	37,0%	(16)
Schwabach	Nürnberg	4.784	(326)	36,8%	(17)

Abbildung A1: Weitere Abgrenzungsergebnisse

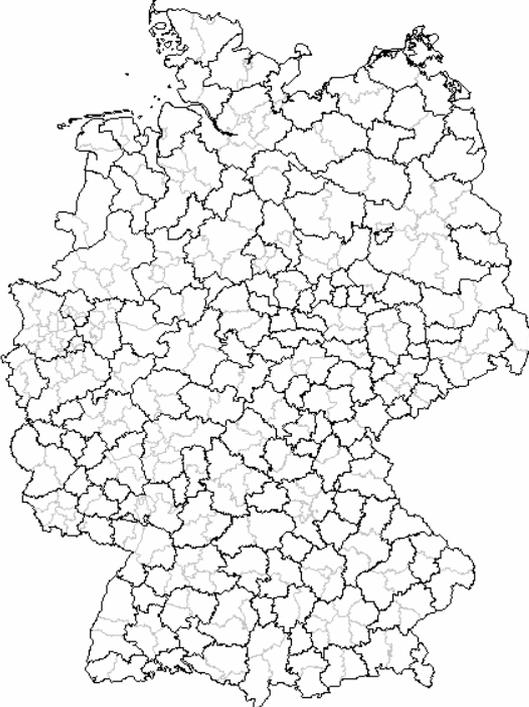
a) Raumordnungsregionen des BBR



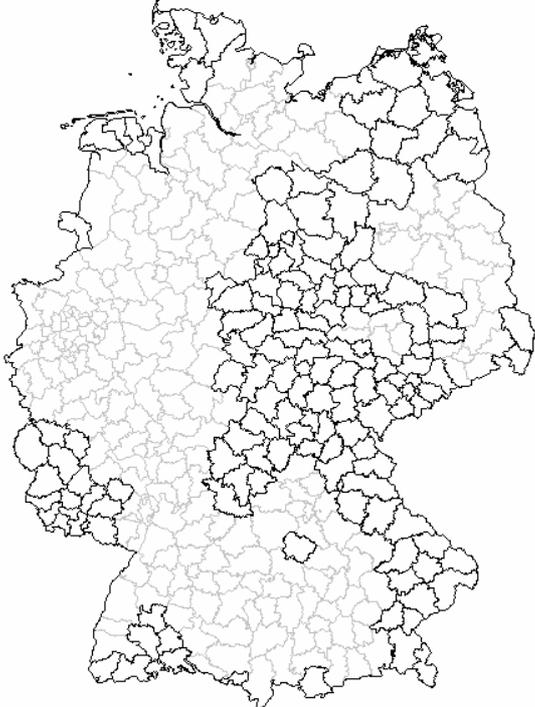
b) Abgrenzung der Gemeinschaftsaufgabe



c) Matrix: Matrixweise Normalisiert; Linkage: Average



d) Matrix: Matrixweise Normalisiert; Linkage: Centroid



In dieser Reihe sind zuletzt erschienen

Diese Seite erstellt das Internet-Büro!

Stand: XXXXX

Eine vollständige Liste aller erschienenen IAB-Discussion Paper finden Sie unter
<http://www.iab.de/de/publikationen/discussionpaper.aspx>