Comparing regional delimitations by Newman’s modularity approach

EU-TTWA workshop
Roma, 27.09.2016

Per Kropp*
Barbara Schwengler
Institute for Employment Research
*Regional Research Network, Halle
1. Problem: Delimit Labor Market Regions (LMR)
2. Methods:
   - Established methods
   - Own methods
   - Newman-Girvan’s Modularity measure $Q$
3. Data: German commuting data
4. Comparison with other delimitations
5. Discussion
Advantages of **functional regions** (LMR) vs. **administrative regions** (like districts):

- more reliable / comparable regional statistics
- more reliable basis
  - labour market policy
  - regional planning
- many forms of interaction are captured
  - fewer spatial interdependencies to deal with
1. Problem: Delimit Labour Markets Regions

- What is a good LMR delineation?
  - Capture as many commuters as possible within LMR.
  - TTWA: Provide the **smallest possible regional unit**.

- Methodological issues
  - Best procedure to **identify and merge strong interacting units**.
    (usually represented in an interaction matrix)
  - **Define** interaction between units.
  - Find an **optimal number of final (merged) units** (≠ 1).
  - **Compare** different results.
  - How robust (e.g. stable over time) are results?

- Practical issues
  - How many LMR?
  - Sizes of LMR?
2. Previous approaches

- **Threshold methods: Coombes’ Travel-to-Work-Areas**
  - + simple
  - - arbitrary thresholds

- **Graph theoretical approach: Concept of dominant flows** (Nystuen/Dacey 1961; Rabion/Occelli 1997)
  - + simple; efficient
  - - not well known, merges to one region

- **Cluster analysis** (Cörvers et al. 2009 / Baumann et al. 1983)
  - + established
  - - re-allocation problem of hierarchical clustering

- **Factor analysis** (Eckey/Kosfeld/Türck 2006)
  - + theoretically best method (using indirect relations as well)
  - - no top result
2. Our approaches (Kropp/Schwengler)

- **Three Step Method** (Kropp/Schwengler 2014) combines Graph theoretical approach with a threshold model
  - + best results
  - – very heterogeneous regions

- **Method of Decreasing Thresholds**
  - + good results, very simple
  - – heterogeneous regions

< Both methods use a quality criterion (**Modularity Q**; Newman / Girvan 2004) to select the best delineation out of many.

- **Robust Merging** (of different delineations)
  - + good & and very robust results
  - – heterogeneous regions
3. Data: German commuting data

- Data of all employees covered by the social security system in Germany (~ 70% of all 40 Mill. employees)
  - cross sections, June 30th each year (here: pooled data 2006-2008)
  - information about place of work and place of living on municipality level
  - 12,000 municipalities
4. Quality of delineations

Measures for comparison

- Modularity Q (Newman/Girvan 2004)
- Commuting ratio: Ratio commuters / no-commuters
- Relation between labour demand and supply (van der Laan/Schalke 2001):
  - Home-Work-Ratio (HWR)
  - Housing Self-Containment Ratio (HSC)
  - Employment Self-Containment Ratio (ESC)

- (Usefulness for practical application?)
4. Quality of delineations

**Modularity** $Q$ (Newman/Girvan 2004)

- Start: symmetric matrix $e$ (LMR x LMR)
- cells $e_{ij}$: commuters $i>j$ (commuters) as share of all units
- diagonal $e_{ii}$: units within clusters (not-commuters) as share of all units
- **comparison with random clustering** (null model) with same row totals and column totals as matrix $e$:

$$Q = \sum_i (e_{ii} - a_i^2) \quad \text{with} \quad a_i = \sum_j e_{ij}$$

$0 < Q < 1$, $Q=0 \ldots$ random, often: $.3 < Q < .7$
4. Quality of delineations

Relation between ESC, HSC & HWR

Employment Self-Containment Ratio (%)

Housing Self-Containment (%)

Home-Work-Relation
Difference in- & out-commuter in relation to all employees in a region
< 0: more out-commuting (region with living function)
> more in-commuting (region with living function)

ESC/HSC-optimum
4. Quality of delineations

Three Step Method

(78% of 50 LMR in ESC/HSC-optimum)
4. Quality of delineations

Decreasing Thresholds
(72% of 58 LMR in ESC/HSC-optimum)
4. Quality of delineations

Robust Merging (first approach) 
(41 % of 124 LMR in ESC/HSC-optimum)
4. Quality of delineations

Eckey/Kosfeld/Türck 2006
150 homogenous LMR (2006)
→ 35% LMR in ESC/HSC-optimum

- **Data**: employees covered by the social security system 2004, commuting between districts
- **Method**: Factor analysis
- **Restrictions**:
  - Min.size: 50,000 inhabitants
  - Max.commuting distance: 45 to 60 minutes
4. Quality of delineations

- **Joint Task** of the Federal Government and the federal states dedicated to the "Improvement of Regional Economic Structure"

270 Joint Task Regions (2007)

→ 17% LMR in ESC/HSC-optimum

**Data:** employees covered by the social security system, district level

**Method:** accessibility and commuting + factor analysis

**Restrictions:**
- Min. size 100.000 inh.
- Max. commuting distance: 45 minutes
## 5. Comparison – methods

<table>
<thead>
<tr>
<th>Delineation</th>
<th>Regions</th>
<th>Modularity Q</th>
<th>Commuter ratio</th>
<th>ESC/HSC Optimum</th>
<th>Empl. Self-Cont. of regions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Three Step Method</td>
<td>50</td>
<td>.848</td>
<td>10.2%</td>
<td>78.0%</td>
<td>86.06 ± 4.8, Min. 72.3, Max. 94.6</td>
</tr>
<tr>
<td>Decreasing Thresholds</td>
<td>58</td>
<td>.846</td>
<td>10.6%</td>
<td>72.4%</td>
<td>86.19 ± 4.2, Min. 75.4, Max. 94.8</td>
</tr>
<tr>
<td>Robust Merging</td>
<td>124</td>
<td>.837</td>
<td>12.4%</td>
<td>40.7%</td>
<td>65.45 ± 27.4, Min. 3.1, Max. 94.4</td>
</tr>
<tr>
<td>Eckey e.a. 2006</td>
<td>150</td>
<td>.793</td>
<td>18.9%</td>
<td>36.0%</td>
<td>76.79 ± 7.3, Min. 43.3, Max. 91.8</td>
</tr>
<tr>
<td>- recent approach</td>
<td>141</td>
<td>.785</td>
<td>19.9%</td>
<td>31.9%</td>
<td>75.96 ± 9.4, Min. 42.0, Max. 91.8</td>
</tr>
<tr>
<td>&quot;Joint Task&quot;-delineation</td>
<td>270</td>
<td>.730</td>
<td>25.9%</td>
<td>13.3%</td>
<td>68.92 ± 10.3, Min. 35.0, Max. 91.2</td>
</tr>
<tr>
<td>- recent approach</td>
<td>258</td>
<td>.733</td>
<td>25.5%</td>
<td>14.0%</td>
<td>69.33 ± 10.2, Min. 35.0, Max. 91.2</td>
</tr>
<tr>
<td>Employment Agencies</td>
<td>174</td>
<td>.726</td>
<td>26.4%</td>
<td>21.8%</td>
<td>71.82 ± 10.3, Min. 43.5, Max. 91.6</td>
</tr>
<tr>
<td>- recent approach</td>
<td>154</td>
<td>.718</td>
<td>27.3%</td>
<td>19.5%</td>
<td>71.48 ± 10.8, Min. 45.2, Max. 91.6</td>
</tr>
<tr>
<td>&quot;Raumordnungsregionen&quot;</td>
<td>96</td>
<td>.787</td>
<td>19.6%</td>
<td>32.3%</td>
<td>78.50 ± 8.6, Min. 50.3, Max. 91.6</td>
</tr>
<tr>
<td>Districts</td>
<td>402</td>
<td>.613</td>
<td>38.1%</td>
<td>5.0%</td>
<td>58.92 ± 13.6, Min. 20.5, Max. 86.7</td>
</tr>
</tbody>
</table>

Mean ESC of all 50 LMR
## 5. Comparison – Our method vs. TTWA

<table>
<thead>
<tr>
<th>Delineation</th>
<th>Reg. N</th>
<th>Modularity Q</th>
<th>Commuter ratio</th>
<th>ESC/HSC Optimum</th>
<th>Empl. Self-Cont. of regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three Step Method</td>
<td>50</td>
<td>.848</td>
<td>10.2%</td>
<td>78.0%</td>
<td>86.06  4.8  72.3  94.6</td>
</tr>
<tr>
<td>Decreasing Thresholds</td>
<td>58</td>
<td>.846</td>
<td>10.6%</td>
<td>72.4%</td>
<td>86.19  4.2  75.4  94.8</td>
</tr>
<tr>
<td>Robust Merging</td>
<td>124</td>
<td>.837</td>
<td>12.4%</td>
<td>40.7%</td>
<td>65.45  27.4  3.1  94.4</td>
</tr>
</tbody>
</table>

**TTWA results:**

<table>
<thead>
<tr>
<th>reg minSZ minSC tarSZ tarSC</th>
<th>GR 10000 .75 100000 .90</th>
<th>70</th>
<th>.834</th>
<th>15.4</th>
<th>67.1</th>
<th>83.4</th>
<th>4.9</th>
<th>73.5</th>
<th>92.8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GR 1000 .75 10000 .90</td>
<td>78</td>
<td>.832</td>
<td>15.2</td>
<td>56.4</td>
<td>82.1</td>
<td>5.5</td>
<td>71.0</td>
<td>92.6</td>
</tr>
<tr>
<td></td>
<td>GR 10000 .667 100000 .75</td>
<td>148</td>
<td>.789</td>
<td>23.8</td>
<td>36.5</td>
<td>77.4</td>
<td>5.6</td>
<td>65.7</td>
<td>91.9</td>
</tr>
<tr>
<td></td>
<td>GR 1000 .667 10000 .75</td>
<td>191</td>
<td>.782</td>
<td>25.2</td>
<td>27.2</td>
<td>75.3</td>
<td>6.2</td>
<td>63.5</td>
<td>91.5</td>
</tr>
<tr>
<td>Distr. 10000 .75 100000 .90</td>
<td>85</td>
<td>.830</td>
<td>16.2</td>
<td>61.2</td>
<td>82.8</td>
<td>4.4</td>
<td>75.2</td>
<td>92.7</td>
<td></td>
</tr>
<tr>
<td>Distr. 1000 .75 10000 .90</td>
<td>93</td>
<td>.829</td>
<td>16.5</td>
<td>60.2</td>
<td>82.1</td>
<td>4.5</td>
<td>75.1</td>
<td>92.6</td>
<td></td>
</tr>
<tr>
<td>Distr. 10000 .667 100000 .75</td>
<td>142</td>
<td>.792</td>
<td>23.4</td>
<td>38.0</td>
<td>77.9</td>
<td>5.7</td>
<td>66.0</td>
<td>91.8</td>
<td></td>
</tr>
<tr>
<td>Distr. 1000 .667 10000 .75</td>
<td>163</td>
<td>.791</td>
<td>23.4</td>
<td>30.7</td>
<td>76.4</td>
<td>6.3</td>
<td>65.0</td>
<td>91.8</td>
<td></td>
</tr>
</tbody>
</table>
5. Comparison – TTWA Germany

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Distr. 1000 .667 10000 .75</td>
<td>163</td>
<td>.791</td>
<td>23.4</td>
<td>30.7</td>
</tr>
<tr>
<td>Distr. 10000 .75 100000 .90</td>
<td>85</td>
<td>.830</td>
<td>16.2</td>
<td>61.2</td>
</tr>
<tr>
<td>GR 1000 .667 10000 .75</td>
<td>191</td>
<td>.782</td>
<td>25.2</td>
<td>27.2</td>
</tr>
</tbody>
</table>
5. Comparison – TTWA Sardinia

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>reg minSZ minSC tarSZ tarSC</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sard. 1000 .667 10000 .75</td>
<td>36</td>
<td>.752</td>
<td>55.6</td>
<td>11.4</td>
<td></td>
</tr>
<tr>
<td>Sard. 10000 .667 100000 .75</td>
<td>10</td>
<td>.747</td>
<td>70.0</td>
<td>6.3</td>
<td></td>
</tr>
<tr>
<td>Sard. 1000 .75 10000 .90</td>
<td>17</td>
<td>.764</td>
<td>76.5</td>
<td>7.7</td>
<td></td>
</tr>
<tr>
<td>Sard. 10000 .75 100000 .90</td>
<td>7</td>
<td>.723</td>
<td>85.7</td>
<td>3.8</td>
<td></td>
</tr>
</tbody>
</table>
## 5. Comparison – DT in Germany & Italy

<table>
<thead>
<tr>
<th>Delineation</th>
<th>Reg.</th>
<th>Modularity Q</th>
<th>Commuter ratio</th>
<th>ESC/HSC Optimum</th>
<th>Empl. Self-Cont. of regions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D Three Step Method</td>
<td>50</td>
<td>.848</td>
<td>10.2%</td>
<td>78.0%</td>
<td>86.06, 4.8, 72.3, 94.6</td>
</tr>
<tr>
<td>D Decreasing Thresholds (Optimum T=6.9)</td>
<td>58</td>
<td>.846</td>
<td>10.6%</td>
<td>72.4%</td>
<td>86.19, 4.2, 75.4, 94.8</td>
</tr>
<tr>
<td>D Robust Merging</td>
<td>124</td>
<td>.837</td>
<td>12.4%</td>
<td>40.7%</td>
<td>65.45, 27.4, 3.1, 94.4</td>
</tr>
<tr>
<td>I Decreasing Thresholds (Optimum T=5.1)</td>
<td>118</td>
<td>.912</td>
<td>3.1%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6. Conclusions

- **Methods:**
  - The Modularity approach
    - makes a comparison between different delineations possible
    - helps to develop new methods to identify functional regions
    - \( Q \) does not dependent on the number of regions

- **New delineations for the German labor market:**
  - fewer LMR than in other delineations
  - stronger heterogeneity of sizes
  - stable structures

- **Further research:**
  - Are price levels for labor, housing & goods more homogeneous within LMR?
Thanks for your attention!

Dr. Per Kropp
Institute for Employment Research
Regional Research Network
Frau-von-Selmnitz-Str. 6, D-06110 Halle
Germany

Phone: +49/345-1332-236 (secretary: -255; fax: -555)
mail: per.kropp@iab.de
www.iab.de/iab-sachsen-anhalt-thuringen


2. Common restrictions

Usual restrictions for practical application:

- follow district or „Länder“-borders (NUTS 1 & 3)
- spatial contingency & complete coverage
- max. interaction within LMR
  - most employees live within LMR
  - most inhabitants work within LMR
- low interaction with other LMR
  - thresholds for **Self-Containment** between 65% & 75% (Binder/Schwengler 2006; Hensen/Cörvers 2003; Casado-Díaz 2000)
- min. size (inhabitants or employees)
  - 30.000 inh. (Klemmer/Kraemer 1975), 50.000 (Eckey/Klemmer 1991, Eckey/Kosfeld/Türck 2006); 100.000 (Eckey 2001, Binder/Schwengler 2006); 200.000 (van der Laan/Schalke 2001: 205)
- max. size (spatial extension within daily commuting range)
  - 45, max. 60 minutes (Eckey/Kosfeld/Türck 2006)

„closed“ LMR = high **Self-Containment**
2. Method of Decreasing Thresholds

- start with all municipalities and their interaction matrix

- **merge** region A with region B when more than 50% of all employed inhabitants of A work in B

- recalculate interaction matrix

- repete above steps with **lower thresholds**

- select the result with the **highest modularity Q**: threshold = 6.9
2. Method of Robust Merging

- choose good delimitations
- classify all municipalities as:
  - core (same LMR)
  - part (same LMR | own LMR)
  - overlap (different LMR)
- merge municipalities of overlapping regions with the core or part to which they are strongest connected
- (optional:) merge municipalities of very small cores or parts with other cores or parts to which they are strongest connected
4. Quality of delineations

- **Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR)**
  - **Data**: employees covered by the social security system, district level
  - **Method**: accessibility and commuting
  - **Restrictions**: "Planing regions", Länder borders
  - **Result**: Compromise between planing traditions and functional regions
  - **96 „Raumordnungsregionen“ (2008) ➔ 31% LMR in ESC/HSC-optimum
4. Quality of delineations

Administrative delineations:


413 Districts (2008) → 5 % LMR in ESC/HSC-optimum