Annex C. LabourMarketAreas – R package

The R package LabourMarketAreas implements the whole labour market area (LMA) delineation process and has a modular structure. It consists of a series of functions; each function addresses a specific stage of the LMA delineation process. In the subsequent section, a basic example of a typical delineation process is summarised. The complete set of functions of the package and the corresponding detailed description is available in Cran.[1]

The only input required for the algorithm to run is the commuting flows matrix, the initial parameters (size and self-containment as presented in Chapter 4) and information on the type of coding system used for the basic territorial units (those for whom the commuting data is available).

An example of an LMA delineation process

The LMA delineation process may comprise of the following stages:

1. Preliminary treatment: the algorithm identifies basic territorial units (communities) presenting anomalies (only incoming or outgoing or internal flows) and provides a report of such features in the data (function *findClusters*).

2. Regionalisation algorithm: this is the core of the package and it is dealt with by the function *findCluster*. This function implements a greedy algorithm to aggregate the basic territorial units into clusters and find the territorial partition representing the LMAs. Starting from the basic territorial units, the algorithm iteratively aggregates them until all clusters satisfy the validity criteria (Annex A) set by the validity function and the parameters. Figure A C.1 presents a schematic workflow of the algorithm. A very basic application of the algorithm is given below:

    library(LabourMarketAreas)

    ##read commuting flows data

    dat=fread("commuting flows.txt")

    ##apply the iterative algorithm

    *out<- **findClusters**(LWCom=dat,minSZ=1000,minSC=0.6667,tarSZ=10 00,tarSC=0.75)*

    ## the object out contains all the information on the set of LMAs found by the algorithm.

    ## To view how the basic territorial units have been grouped type:

    outlmaclusterList.

3. Naming and visualisation of the LMAs: LMAs are named after the community having the maximum incoming flows. When community geographical co-ordinates are available in geospatial vector format (.shp files), tools to deploy this information

at LMA level are included in the package in order to produce LMAs shapefiles and visualise the obtained geography (the function involved are *AssignLmaName* and *CreateLMAShape*).

The figure below illustrates the implementation of the R package LabourMarketAreas.

Figure A C.1. Scheme of the implementation of the R package LabourMarketAreas
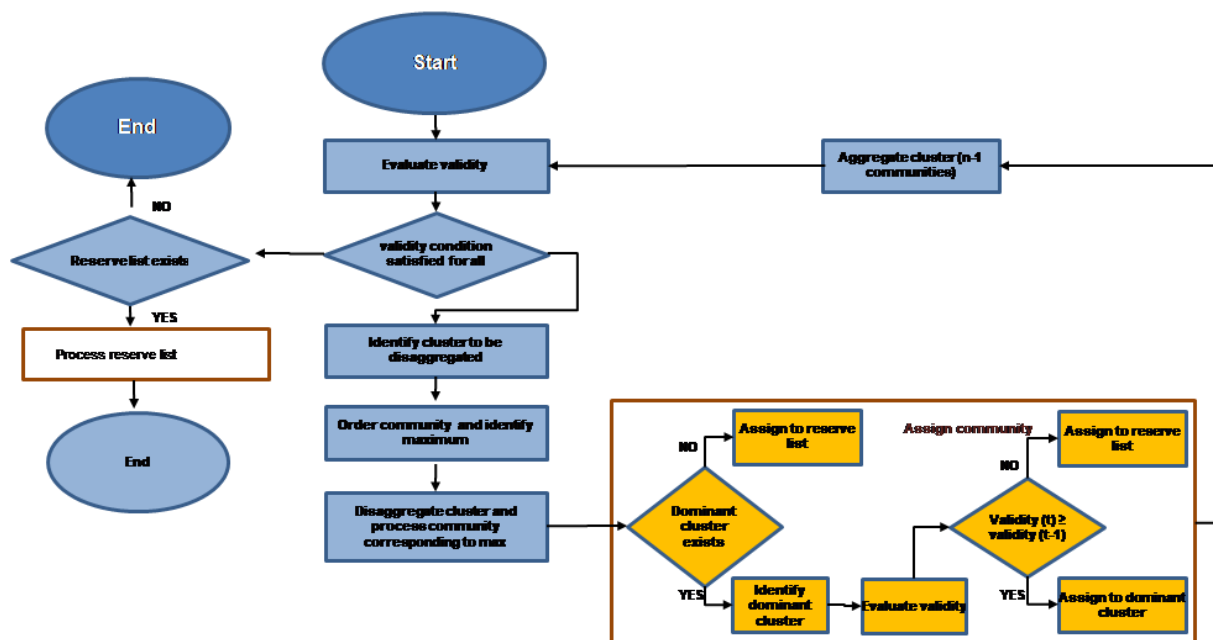
Source: Derived from Istat, 2014: https://www.istat.it/it/archivio/142676.

4. Fine-tuning of the geography: As the algorithm is based exclusively on commuting flows, some areas may include communities not spatially contiguous. Based on geospatial information, the tools implemented in the R package allow complying with the contiguity principle. This stage of the process is performed in an interactive manner, as expert knowledge has to be exploited to assign correctly communities to labour market areas. Four distinct functions implement the fine-tuning process, namely: *CreateLMAShape*, *FindIsolated*, *FindContig* and *AssignSingleComToSin gleLma*. These functions respectively create the LMA geospatial vector, find the isolated territorial units one after the other, propose the possible LMAs that are contiguous to the isolated territorial unit under examination and assign the latter to the one selected by the user. A schematic application of this principle is given below:

> *shape_terr_unit=rgdal::readOGR(dsn = "my_directory" layer = "BasicTerritorial_Units_shape_file")*
>
> *shape_lma=**CreateLMAShape**(lma=out$lma,shp_com=shape_terr_unit , …)*
>
> *iso=**FindIsolated**(lma=out$lma,   lma_shp=shape_lma$shp_lma, com_shp=shape_terr_unit, …)*

> *conti.lma=**FindContig**(type="lma",lma=out$lma,contig.matrix=iso$iso lated.lma$contig.matrix.lma,*
> *isolated=iso$isolated.lma$lma.unique$lma.unique.ID)*
>
> *out_1=**AssignSingleComToSingleLma**(out$lma,names(conti.lma)[1],co nti.lma[[1]],dat).*

5. Comparison of possible alternative LMAs in a given area of the country; to analyse the coherence, consistency and appropriateness of individual allocations of basic territorial units, the function *PlotLmaCommunity* compares two candidate LMA partitions containing the specified territorial units.

6. Sensitivity analysis: Different sets of the initial parameters imply slightly different LMA configurations; the investigation of such geographies is essential in order to address the issue of finding the most appropriate partition satisfying the many different requirements needed by each country. The sensitivity analysis can be performed by setting different groups of parameters in the function *findClusters* and collecting the results. The functions *CompareLMAsStat* and *StatClusterData* enable the quantitative analysis of the output stemming from a specified set of initial parameters. These functions provide statistics on different dimensions: single LMAs, commuting flows between LMAs and complete set of LMAs (the partition) as a whole. Examples of such statistics pertain:

    a. LMA statistics such as the number of residents or workers, home-work ratio, supply and demand self-containment values, internal cohesion link and flows, etc.

    b. Commuting flows statistics such as the percentage of flows below a given threshold, descriptive statistics on incoming or outgoing flows, identification of the LMAs reaching the minimum or maximum incoming or outgoing flows, etc.

    c. Quality statistics on the partition such as the number of clusters, descriptive statistics on supply and demand self-containment, descriptive statistics on number of residents, workers or resident workers, Q-modularity index, etc.

    An example of the use of the function *StatClusterData* is:

    > *Stats = StatClusterData(outlma,outparam,1000,dat)*

7. Further analysis: Further analysis is possible, such as the analysis of the reserve list i.e. those basic territorial units not assigned during the aggregation process to avoid damaging the already existing clusters. The function *StatReserveList* produces statistics on the components.

8. Dissemination: The release of a geography implies the dissemination of a series of products that enable users to both understand and make use of it. Besides the table of correspondence between the basic territorial units (usually municipalities) and the corresponding LMA (provided by *outlmaclusterList*), the geospatial vectors allowing their cartographic representation and some descriptive statistics are usually released coupled with socioeconomic indicators at the LMA level. The function *AddStatistics* joins directly the LMA structure to the desired statistics to ease their further usage and representation. The guidelines2 provide further directions on possible products to be made available.

9. Updating of the geography: The demography of the basic territorial units (fusion of municipalities, changes in their territories, etc.) in some cases may cause the change of the LMA borders. Treatments of these cases need to be addressed in order to keep the different levels of the geographies coherent.

The software modularity has the advantage that new elements can be added quite easily in the package. Further developments are already foreseen at the time of writing.

Notes

[1] https://cran.r-project.org/web/packages/LabourMarketAreas/LabourMarketAreas.pdf.

[2] https://ec.europa.eu/eurostat/cros/system/files/guidelines_for_lmas_production08082017_rev300817.pdf.
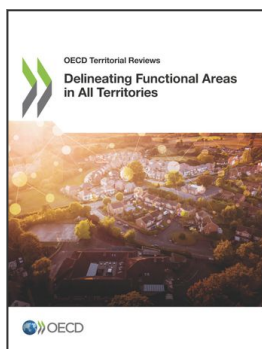
From:

Delineating Functional Areas in All Territories

Access the complete publication at:
https://doi.org/10.1787/07970966-en