Optimization Letters manuscript No.
(will be inserted by the editor)

A Min-Cut Approach to Functional Regionalization, with a Case Study of the Italian Local Labour Market Areas

Gianpiero Bianchi · Renato Bruni · Alessandra Reale · Fabio Sforzi

Received: date / Accepted: date

Abstract In several economical, statistical and geographical applications, a territory must be subdivided into functional regions. Such regions are not fixed and politically delimited, but should be identified by analyzing the interactions among all its constituent localities. This is a very delicate and important task, that often turns out to be computationally difficult. In this work we propose an innovative approach to this problem based on the solution of minimum cut problems over an undirected graph called here transitions graph. The proposed procedure guarantees that the obtained regions satisfy all the statistical conditions required when considering this type of problems. Results on real-world instances show the effectiveness of the proposed approach.

Keywords Functional Regionalization · Min-Cut problem · Local Labour Market Areas · Territorial Districting

1 Introduction

Unlike administrative regions, that have been politically delimited and are, more or less, fixed, *functional regions* are defined over a territory or a geographical area by a system of interactions. The identification of functional regions within a given territory is the basis of several important analysis in many economical, statistical and geographical applications. However, this identification, also called *functional regionalization*, is often a difficult task, especially from the computational point of view, mainly because of the size of the problems. Indeed, the pattern of functional

G.Bianchi, R. Bruni, A.Reale
Department of Computer, Control, and Management Engineering,
Sapienza University of Roma
Via Ariosto 25, 00185 Roma, Italy
E-mail: gianpiero.bianchi@uniroma1.it, bruni@dis.uniroma1.it, alessandra.reale@uniroma1.it

F. Sforzi
Department of Economics, University of Parma
Via J.F. Kennedy 6, Parma 43125, Italy
E-mail: fabio.sforzi@unipr.it

regions is generally not stable over the years, because of changes in geographical organization of production-system. The general problem of partitioning a territory, also known as Territorial Districting or Territory Design, is central to several applicative tasks, see, e.g., [12,33].

One very important case of regionalization is the identification of the so-called Local Labour Markets Areas (LLMAs). A LLMA is "a geographical region where the majority of the local population seeks employment and from which the majority of local employers recruit labour" [19]. In other words, it is an area where demand and supply for labour meet at a high degree, and commuters entering the area or going out from it are only a minority. LLMAs are widely accepted as the most appropriate units for the analysis of spatial patterns of employment and unemployment, for the detection of industrial districts, and for the evaluation of productivity advantages of local economies (see, e.g., [11]). LLMAs often do not fit administrative boundaries, since these boundaries usually remain unchanged for decades and so become unrelated to the changing economic reality [38].

Numerous alternative approaches to the identification of LLMAs have been developed during the past decades. They have been based on a variety of different methods using different terminology. Fischer provided in [14] a summary of the earlier purely statistical approaches to the regionalization problem, and classifies them as either hierarchic or non-hierarchic. A particularly well-known procedure is the intra-regional interaction maximization, or INTRAMAX, initially developed by Masser and Brown [29], also implemented in the software FLOWMAP developed at the University of Utrecht [41]. This procedure is a stepwise hierarchical algorithm based on an interaction matrix, which contains the interactions (e.g., journey-to-work flows) between all the localities constituent a given territory, including the interactions of each locality with itself. In each step, two localities are selected and merged together, producing a new locality (a region) that is the union of the two, and all the interactions involving this new locality are updated accordingly. To avoid possible cases of fusions between nonadjacent localities, a contiguity constraint has also been incorporated into the grouping procedure [29]. However, this constraint causes some loss of information and affects the empirical findings as showed in [34].

According to [8], the methods for the identification of LLMAs can be subdivided into three types of approaches. The first one consists of methods reliant on the manual intervention of experts of the field, and typically involves designating a number of focal points, usually the city centres, before assigning the remaining localities to these points (e.g., [9]). An alternative approach, that is more based on statistical methods, is to employ numerical taxonomy principles. These methods typically use a single procedure that seeks to maximize a statistical criterion representing the objective. They include cluster analyses and specific regionalization algorithms (e.g., [29]) The third approach, initially proposed by Coombes and Openshaw to define a set of Travel-To-Work Areas (TTWA) in Britain [6, 7], is a possible evolution of the other two alternatives, in the sense that it relies on multiple-step statistical methods and principles, in order to ensure that the final boundaries meet predefined criteria. The assignment of a group of localities to a TTWA (a region) is guided by the maximization of the interaction between those localities. Thus, indirectly, the interaction that crosses boundaries between TTWAs is minimized. Besides, every TTWA must reach a minimum level of self-containment and a minimum size in terms of resident occupied population. The

trade-off between size and self-containment is the peculiarity of the TTWA algorithm. A more general version of the TTWA-based algorithm was devised by Sforzi, Openshaw and Wymer to regionalize Italy in LLMAs with support from Istat [35]. The presence/absence of the trade-off between size and self-containment is the main difference between the TTWA-based algorithm and the LLMA-based Italian one. The choice to remove the trade-off was justified by the discretionality of the size of employment and to preserve the established threshold of self-containment, recognized as the key criterion. This algorithm was officially applied by Istat to process journey-to-work data for the functional regionalization of Italy in 1981, 1991 and 2001. To process 2011 data, Istat has modified the algorithm used in the previous decades by restoring the trade-off between size and self-containment. This decision is explained in the methodological note [23] with the need to conform to a procedure recommended by Eurostat to explore the possible definition of a European algorithm for statistical reporting areas.

Van der Laan and Schalke [26] also develop a multi-level classification of the LLMAs identification methods. They basically distinguish between methods allowing for heterogeneity among LLMAs and methods which provide homogeneity. Then, they subdivide the homogeneous category into deductive methods, which identify at first urban centres around which the LLMAs are constructed, and inductive methods, which do not use such pre-conceived structures. In their review, Casado-Díaz and Coombes [5] present another taxonomy of methods, and suggests that identifying best practice has so far been held back by a lack of clear criteria for objectively evaluating different methods. Flórez-Revuelta *et al.* present in [15, 28] a new approach to the identification of LLMAs based on evolutionary computation. The procedure is based on the maximization of a fitness function that measures the aggregate intra-region interaction under constraints of inter-region separation and minimum size.

Another interesting approach to the problem is based on measures of *modularity* (see, e.g., [17]), since there should be dense connections within each region, but only sparse connections between regions. Girvan and Newman [18] propose an algorithm that uses betweenness centrality to find community boundaries. Modularity is used as quality index of a partition of a network into communities. It measures internal (and not external) connectivity, but it does so with reference to a randomized null model. This approach should have specific advantages over existing regionalization procedures, particularly in the context of disaggregate commuting patterns of socio-economics subgroups [13]. Following this line of research, modularity has been very influential in recent community detection literature, and one can use spectral techniques to approximate it [18, 42]. However, Guimerá *et al.* [20] and Fortunato and Barthélemy [16] show that random graphs have high-modularity subsets and that exists a size scale below which modularity cannot identify communities. Finally, Kim et al. [25] propose an exact approach to the case of regionalization problem with a predetermined number of regions.

We propose here a new approach to regionalization problems, with specific reference to the identification of LLMAs. The problem is converted into a graph partitioning problem, and the solution is obtained by solving a sequence of minimum cut problems over an undirected graph obtained from the interactions among the localities. This graph is here called transitions graph. The procedure has been implemented in c++ and tested on real data from the Italian Census of Population 2001. The results of the proposed approach are compared to those of the procedure

officially used by the Italian National Institute of Statistics (Istat) in 1981,1991 and 2001 to define the Italian LLMAs ([32,36,37]). After the choice of a method highly specialized to determine the Italian LLMAs, another comparison is with the state-of-the-art procedure for general purpose graph partitioning METIS [24].

The main contribution of this work is therefore an innovative and effective approach based on Combinatorial Optimization for solving an economically important and challenging real-world problem. The rest of this work is organized as follows. Section 2 describes in detail the problem of the identification of LLMAs. Section 3 explains the proposed procedure, based on the solution of minimum cuts problems. Section 4 outlines the algorithms used for the comparison. Section 5 reports the experiments and the comparison on data from the Italian Census of Population 2001, along with a discussion of the empirical results.

2 The Regionalization Problem

The approach proposed for the problem of the identification of functional regions will be hereinafter explained by referring to the specific case of the identification of Local Labour Markets Areas (LLMAs). This is indeed one of the most important cases of regionalization, because it has a great economic relevance and it requires to deal with a very large set of data. In the described case, we have the set

$$A = \{a_1, \ldots, a_n\}$$

of all the localities a_i situated in a territory T, that is the geographical area under analysis. Set A is such that $\bigcup_{i=1}^{n} a_i = T$ and $a_i \cap a_j = \phi$ for $i \neq j$. Moreover, we have an $n \times n$ matrix

$$F = \begin{pmatrix} f_{11} & \ldots & f_{1n} \\ & \ldots & \\ f_{n1} & \ldots & f_{nn} \end{pmatrix}$$

of the interactions existing between all the pairs of localities. In particular, value $f_{ij} \geq 0$ is a measure of the flow of workers that reside in locality a_i and work in locality a_j, and is called *commuting flow*, or also daily journey-to-work flow. Clearly, F is not necessarily symmetric. The identification of the LLMAs consists of a partition of the set A into subsets $R_1, \ldots, R_m$ (the functional regions) such that $R_p \cap R_q = \phi$ for $p \neq q$ and $\bigcup_{p=1}^{m} R_p = T$. The goals of this partition may be viewed from different perspectives, but basically consist of maximizing the number of LLMA such that the obtained regions remain statistically and economically meaningful. This means that: (*i*) each LLMA must be sufficiently self-contained; (*ii*) each LLMA must have a sufficient number of workers; (*iii*) each LLMA must be composed by a set of localities; (*iv*) each LLMA must be internally contiguous.

To impose condition (*i*) one needs to evaluate *self-containment*. The total occupied population working in locality a_i (i.e., for short, *workers* in a_i) is $w(a_i) = \sum_{k=1}^{n} f_{ki}$. Consequently, the total number of workers in region R_p is

$$w(R_p) = \sum_{a_i \in R_p} \sum_{k=1}^{n} f_{ki}.$$

Specularly, the total occupied population residing in a_i (i.e., for short, *residents* in a_i) is $r(a_i) = \sum_{j=1}^{n} f_{ij}$. Consequently, the total number of residents in R_p is

$$r(R_p) = \sum_{a_i \in R_p} \sum_{j=1}^{n} f_{ij}.$$

Also, the total number of workers in locality a_i that reside outside of a_i (the incoming commuters) is $c^-(a_i) = \sum_{k=1, k\neq i}^{n} f_{ki}$. Consequently, the total number of incoming commuters in region R_p is

$$c^-(R_p) = \sum_{(ki): a_k \notin R_p, a_i \in R_p} f_{ki}.$$

Conversely, the total number of residents in locality a_i that work outside of a_i (the outgoing commuters) is $c^+(a_i) = \sum_{j=1, j\neq i}^{n} f_{ij}$. Consequently, the total number of outgoing commuters from region R_p is

$$c^+(R_p) = \sum_{(ij): a_i \in R_p, a_j \notin R_p} f_{ij}.$$

Finally, the total number of residents in region R_p that also work in R_p (the internal flow) is

$$l(R_p) = \sum_{(i,j):\ a_i, a_j \in R_p} f_{ij}.$$

Hence, value f_{ii} is also called the internal flow of locality a_i. The *Supply-side self-containment* function for a_i is defined to evaluate the portion of people residing and working in locality a_i within the total number of workers in a_i, as follows:

$$sc^w(a_i) = \frac{f_{ii}}{w(a_i)}.$$

Consequently, the Supply-side self-containment for a region R_p is

$$sc^w(R_p) = \frac{l(R_p)}{w(R_p)}. \tag{1}$$

The *Demand-side self-containment* function for a_i is defined to evaluate the portion of people residing and working in a_i within the total residents in a_i, as follows:

$$sc^r(a_i) = \frac{f_{ii}}{r(a_i)}.$$

Consequently, the Demand-side self-containment for a region R_p is

$$sc^r(R_p) = \frac{l(R_p)}{r(R_p)}. \tag{2}$$

Finally, we define the *Overall self-containment* function for a_i to evaluate the portion of people residing and working in a_i within the total number of persons interacting with a_i (that is, working and/or residing), as follows:

$$sc(a_i) = \frac{f_{ii}}{f_{ii} + c^-(a_i) + c^+(a_i)}$$

Consequently, the Overall self-containment for a region R_p is

$$sc(R_p) = \frac{l(R_p)}{l(R_p) + c^-(R_p) + c^+(R_p)}. \tag{3}$$

Self-containment functions have been used in the literature in different manners. Clearly, the first two only consider partial aspects. We select the Overall self-containment one, because in our opinion it is more coherent with the definition of LLMA. Hence, a region R_p is sufficiently self-contained, i.e., respects condition (*i*) above, when $sc(R_p) \geq c_1$, where c_1 is a threshold defined for the specific analysis. Since in any case $sc(R_p) \in [0, 1]$, possible thresholds range in $(0.5, 1]$. We mention, however, that several authors make a different choice, and consider simultaneously the Supply-side self-containment and the Demand-side self-containment for their analysis [8]. Nevertheless, note that $sc(R_p) \geq c_1$ implies both $sc^w(R_p) \geq c_1$ and $sc^r(R_p) \geq c_1$, since all the values involved in their computation are nonnegative.

A region R_p has a sufficient number of workers, i.e., it respects condition (*ii*) above, when $w(R_p) \geq c_2$, where c_2 is a natural number, again depending on the specific analysis and on the size of the a_i (as an example, for the present case-study $c_2 = 1000$). A region R_p respects condition (*iii*) above when the number of localities composing it is such that $|R_p| \geq c_3$, where c_3 is another natural number, again depending on the specific analysis. For example, $c_3 = 2$, see also [35]. However, the proposed approach is flexible and allows setting c_3 to different natural numbers whenever the application would require it.

Condition (*iv*), finally, is of difficult mathematical formalization. However, several authors in the literature state that this condition should not be imposed during the generation of the LLMAs, because otherwise, at each step of a generic regionalization procedure, it would limit the possible choices, and this would very likely lead to the determination of worse LLMAs boundaries, as explained in, e.g., [5]. So, the possibility of working with non-contiguous proto-regions before producing the final (contiguous) LLMAs should be allowed during the generation process. On the contrary, contiguity should be checked on the final LLMAs, and, if one of them does not respect it, one has to disassemble that LLMA and possibly merge some of its localities with other LLMAs contiguous to them.

Conditions (*i*) and (*ii*) can be slightly relaxed, in the sense that the (almost) full satisfaction of one condition is considered enough to compensate a little unsatisfaction of the other. Hence, the following condition (4) can subsume the two conditions (*i*) and (*ii*), where c is a threshold with a value generally ≥ 0.75 [39].

$$\left(\min\left\{\frac{sc(R_p)}{c_1}, 1\right\}\right) \cdot \left(\min\left\{\frac{w(R_p)}{c_2}, 1\right\}\right) \geq c \tag{4}$$

Evidently, all the above conditions are more easily respected by large LLMAs, and, indeed, one unique LLMA over the whole territory under analysis would fully satisfy them (assuming of course that the values of c_2 and c_3 are feasible for that territory). In order to avoid such kind of solutions, useless from the practical point of view, and to maximize the number of LLMAs, one usually wants that the above conditions (*i*), (*ii*) and (*iii*) are satisfied with the minimum values of self-containment, workers and areas that are able to do that, possibly relaxing conditions (*i*) and (*ii*) with condition (4). When this happens, and also condition (*iv*) is met, the regionalization task has been successfully performed. Note that

the practical cases of this problem are generally very computationally demanding, and all the approaches used for this problem are actually approximate procedures (unless simplifications of the problem are considered, for example by pre-assigning the number m of LLMAs that should be generated).

3 The Proposed Min-Cut Approach

By viewing the set A as the set of the vertices V of a graph, and the set of the values of F as the weights of a set of edges E' connecting all the pairs of vertices in V, the problem evidently becomes a type of *graph partitioning* problem over a complete graph $G' = (V, E')$. However, differently from the standard cases, G' contains also loops, i.e., arcs of the type (i, i), going from i to i itself and corresponding to the mentioned internal flows. Since F is not symmetric, G' will be a directed graph. Graph partitioning problems have been extensively studied (see [2] for references), since they have applications in many areas, e.g., clustering, detection of cliques in social, pathological and biological networks, programs mapping onto parallel architectures, image segmentation, numerical analysis, VLSI design. Typically, graph partitioning problems fall under the category of NP-hard problems, and practical solutions algorithms are based on heuristics (see, e.g., [31]). One widely used approach is the so-called multilevel one. Multilevel algorithms iteratively reduce the size of the graph by collapsing vertices and edges, partition the smaller graph, then map back and refine this partition on the original graph. A good example of this approach is implemented in the software METIS [24]. We analyze the usability of similar approaches for solving the regionalization problem in the following Section 4.

However, we note that expression (3) for the computation of *sc*, has the following property: both numerator and denominator are constituted of sums of commuting flows, and f_{ij} is contained in one of those sums if and only if also f_{ji} is contained in it. Therefore, we consider an undirected complete graph $G = (V, E)$ having for each arc $(i, j) \in E$ a weight

$$w(i, j) = f_{ij} + f_{ji}$$

Graph G can be used instead of G' for the computation of (3) or (4) without any loss of information. The advantage is that G has a considerably smaller number of arcs. We also observe that a partition solving our regionalization problem disconnects G in such a way that the arcs that are removed constitute a set that should have *small* total sum of the $w(i, j)$ flows and that certainly would not include loops. Therefore, when searching for such a set of arcs, we can remove all loops from G. We obtain in this manner an undirected complete graph G, that we call *transitions graph*, having $\frac{n^2-n}{2}$ arcs, instead of the directed graph G' having n^2 arcs. Graph G does not contain the whole information of the commuting flows. Nevertheless, each solution to the original problem is obtainable as a partition of G, since loops would never be cut.

On the other hand, the internal flows are needed for the evaluation of self-containment. Therefore, our algorithm includes a validation step, during which the partitions obtained on G are checked for the satisfaction of condition (4) by computing *sc* as in (3), that is, by considering also the internal flows. As showed

in the previous section, given a generic LLMA, an Overall self-containment $\geq c_1$ implies that both its Supply- and Demand-side self-containment are $\geq c_1$. As a consequence, we can guarantee that the partitions of G produced by our algorithm respect the conditions described in the previous section. Note also that the internal flows generally represent the larger values in matrix F, and some of them may be order of magnitude larger than all non-internal flows. So, the restriction of the partitioning problem to G also improves the numerical condition of the problem, that may indeed be originally ill-conditioned.

We now describe the recursive partitioning procedure that we apply. We denote the set of vertices of a generic graph G by $V(G)$ or simply by V when there is no ambiguity. A *cut* in G is a partition $(S, \bar{S})$ of V, a *cutset* is the set of arcs connecting S and $\bar{S}$ in G. We define the weight of cut $(S, \bar{S})$ as

$$W(S, \bar{S}) = \sum_{i \in S,\ j \in \bar{S}} w(i, j).$$

Similarly, we define for any two sets A, B of vertices of G, the weight

$$W(A : B) = \sum_{i \in A, j \in B} w(i, j)$$

The transitions graph G is partitioned in order to obtain the subsets of vertices $(S_1, \ldots, S_m)$ corresponding to the LLMAs $(R_1, \ldots, R_m)$ by finding cuts with minimum weight. We also need to define an operation, called *contraction*, or equivalently *join* or *merge*, of two or more vertices of a graph G, that produces a graph with less vertices, as follows. Given a graph G, the contraction of two vertices ν and μ produces a new graph $G/\nu \sim \mu$, where ν and μ are replaced by a new vertex $[\nu] = [\mu]$, and the weights of the edges (v, ν) and (v, μ), for any generic vertex $v \neq \nu, \mu$, are summed, i.e., $w(v, [\nu]) = w(v, \nu) + w(v, \mu)$. The contraction of a set of vertices is the repeated contraction of its pairs of vertices. The contraction of a (sub-)graph is the contraction of the set of all its vertices.

We say that a (sub-)graph G is *feasible* when it satisfies conditions (4) and $|V(G)| \geq c_3$. We say that we *split* a (sub-)graph G when we remove from it all the arcs of a cutset. We also say that a (sub-)graph G is *unsplittable* when it is feasible but it has values of self-containment, workers and localities such that any further splitting of G will produce subgraphs that are not feasible. Since our aim is to partition the transitions graph at the maximum extent, we try to obtain a partition corresponding to subgraphs that are all unsplittable.

In the procedure below we use a list of *open problems* L to store all the subgraphs of G that have not yet been identified as unsplittable, and so will undergo the cutting operation. We also use a list of *closed problems* T to store the subgraphs G_p that have been recognized as unsplittable.

Procedure for the generation of LLMAs

Input An undirected complete graph $G(V, E)$ with n vertices associated with the n localities a_i, edge weights $w(i, j) = (f_{ij} + f_{ji}) \geq 0$, vertex weights $f(i, i) \geq 0$.

Output A partition of V into $(S_1, \ldots, S_m)$ such that $|S_p| \geq c_3$ for $p = 1, \ldots, m$, and that the corresponding $(R_1, \ldots, R_m)$ respect condition (4) and are contiguous. Value m is not fixed in advance.

Initialization:
Remove from E all the edges with $w(i,j) = 0$
Identify the connected components of G, call them $G_0, \ldots, G_c$ and insert them into the list of open problems L
The list of closed problems T is empty

Iteration:
Cut If L is empty: break the iteration and goto **Contiguity_enforcement**
If L is not empty: extract (sub-)graph G_h from L
Apply procedure **MinCut** to G_h to obtain the minimum weight cut $(S, \bar{S})_h$
Remove the corresponding cutset obtaining subgraphs G_{h+1} and G_{h+2}
Valid Check if G_{h+1}, G_{h+2} satisfy condition (4) and have at least c_3 localities each
Case 1 If both G_{h+1} and G_{h+2} satisfy these conditions: insert G_{h+1} and G_{h+2} in L and repeat the **Iteration**
Case 2 If neither G_{h+1} nor G_{h+2} satisfy these conditions: G_h is unsplittable
Insert G_h in T and repeat the **Iteration**
Case 3 If only one of them, say w.l.o.g. G_{h+1}, satisfies these conditions:
Contract G_{h+2} into a single vertex ν,
Contract ν with the vertex $\mu = \underset{v \in V(G_{h+1})}{\operatorname{argmax}} \{w(\nu, v)\}$
Add the obtained $G_{h+3} = G_{h+1}/\nu \sim \mu$ to L and repeat the **Iteration**

Contiguity_enforcement:
Split For each $G_p \in T$, consider the set $S_p = V(G_p)$. If S_p is non-contiguous, split it in its contiguous parts $S_{p1} \ldots S_{pk}$
Join Join each S_{pi} with the set $S_q = \underset{S:\ \text{co.}\wedge\text{ne.}}{\operatorname{argmax}} \{W(S : S_{pi})\}$, where
co. = contiguous and ne. = neighboring to S_{pi}
Return $(S_1, \ldots, S_m)$

During the initialization step, we remove all zero edges from G in order to further reduce the size of the problem. In case this operation disconnects G, we simply work independently on each of its connected components. Quite often, however, G remains connected. We use a depth-first search algorithm for finding the connected components in linear time (see, e.g., [22]). After the initialization step, we repeatedly apply the iteration step, that means we split each (sub-)graph G_h contained in L by removing the edges corresponding to the minimum weight cut. The procedure for finding such cut is described below. If the subgraphs G_{h+1} and G_{h+2} obtained in this way are still feasible, they may be even further splittable. Therefore, we insert them in L so that they will undergo a new cutting operation. On the contrary, if G_{h+1} and G_{h+2} are not feasible, this means that G_h is unsplittable, ans so G_h is inserted in T. Finally, when exactly one of G_{h+1} and G_{h+2} is feasible, the infeasible one represents a set of localities that cannot remain alone but are strongly interconnected. Therefore, we contract them into a single vertex, and join it to the vertex of the feasible subgraph that maximizes the interaction. Since the graph obtained by this operation may be even further splittable, we insert it in L. The procedure stops when L becomes empty, that means that all the generated subgraphs are unsplittable. When this happens, T contains those m unsplittable subgraphs whose sets of vertices $(S_1, \ldots, S_m)$ constitute the wanted partition. The LLMAs $(R_1, \ldots, R_m)$ corresponding to such sets must now be checked for geographical contiguity. Each non-contiguous R_p is split into its contiguous parts

$R_{p1} \dots R_{pk}$. Then, each part R_{pi} is joined with a neighboring LLMA R_q such that the total weights of the arcs connecting R_{pi} and R_q is maximum. The described procedure guarantees that the generated LLMAs satisfy conditions (4), (*iii*) and (*iv*), (or other conditions that may be imposed in the validation step). The number m of generated LLMAs is not guaranteed to be maximum, thought it is generally large enough.

A minimum weight cut, also called for brevity minimum cut or min-cut, of an undirected graph with edge weights, is a set of edges with minimum sum of weights, such that its removal would cause the graph to become disconnected. The total weight of the edges in a minimum cut of G is denoted by $\lambda(G)$ and called edge-connectivity of G (see, e.g., [10]). To compute such a cut, there exist in the literature many methods. One group of algorithms is based on the well-known result of Ford and Fulkerson regarding the duality of maximum s-t-flows and minimum s-t-cuts for arbitrary vertices s and t. Following this approach, Hao and Orlin [21] showed an algorithm to solve all the necessary max-flow problems in time asymptotically equal to one max-flow computation, requiring $O(|V| \times |E| \log(|V|^2/|E|))$ steps. Nagamochi and Ibaraki [30] described an algorithm without using maximum flows. Instead, they construct spanning forests and iteratively contract edges with high weights. This leads to an asymptotic runtime of $O(|V| \times |E| + |V|^2 \log |V|)$ on undirected graphs with nonnegative real edge weights. Their approach was refined in [40] by Stoer and Wagner, by replacing the construction of spanning forests with the construction of Maximum Adjacency, and by Brinkmeier [3] by contracting more than one pair of vertices if possible and by introducing an alternative data structure called priority queues with threshold. Incorporating these improvements, the algorithm obtains an asymptotic runtime of $O(|V|^2 \lambda(G))$ for undirected graphs with nonnegative integer weights, as it is the case of our transitions graph, and this is the algorithm that we apply. A Maximum Adjacency (MA) order is defined as follows. Given a graph $G = (V, E)$, its vertices are arranged in an MA order, if, for each v_i with $i > 1$, the sum of the weights from v_i to all preceding vertices $v_1, \dots, v_{i-1}$ is maximal among all vertices v_k with $k \geq i$ [40]. We now describe the procedure for finding the minimum weight cut.

Procedure MinCut

Input An undirected connected $G(V, E)$ with edge weights $w(i, j) \geq 0$
Output A cut $(S, \bar{S})$ in G with minimum weight

Initialization:
Chose any vertex $\in V$ and call it v_1
Let $n = |V|$. Let $S = \{v_1\}$

for i = 2 to n:
Let v_i be the vertex corresponding to $\underset{v \in V \setminus S}{\operatorname{argmax}} W(S : \{v\})$
Let $S := S \cup v_i$
endfor

if n = 2: return the cut $(\{v_1\}, \{v_n\})$
else: return the minimum cut between $(\{v_1, \dots, v_{n-1}\}, \{v_n\})$
and **MinCut**$(G/v_{n-1} \sim v_n)$

The algorithm starts with any vertex, say v_1, and builds an ordering of the vertices by always adding to the set S of the selected vertices the vertex whose total weight to S is maximized. This provides an MA ordering. After this, the cut induced by the last vertex in the ordering is considered, as well as the cuts obtained by recursively applying the procedure to the graph obtained by contracting the last two vertices. The minimum among the cuts obtained during these recursions is the global minimum weight cut of the graph.

4 Comparison with Other Techniques

In this section we describe two existing solution approaches to functional regionalization that we selected for a comparison with our procedure. First we describe the procedure officially used by the Italian National Institute of Statistics (Istat) in 1981,1991 and 2001 to define the Italian LLMAs ([37]). This procedure has been selected as our reference algorithm because, among the ones proposed in literature, it appears to be the most appropriate to provide a meaningful functional regionalization of Italy (see, e.g., [11,4]). After the choice of a method highly specialized in the individuation of LLMAs, we consider a general purpose graph partitioning method. A very effective among such methods is METIS [24], that is therefore selected for the second comparison.

4.1 Established Algorithm for the definition of Italian LLMAs

We describe here the procedure that has been officially used by Istat for the definition the Italian LLMAs. It is an agglomerative multi-stage heuristic constituted by 5 phases: (1) Identification of potential LLMAs focal points; (2) Amalgamation of potential LLMAs focal points; (3) Expansion of focal points into proto-LLMAs; (4) Identification of potential LLMAs; (5) Optimization of LLMAs boundaries. This procedure, described in detail in [37], uses the supply-side self-containment $sc^w(a_i)$ and the demand-side self-containment $sc^r(a_i)$ defined in Section 2 and the Job ratio function defined as: $jr(a_i) = c^-(a_i)/c^+(a_i)$.

In phase (1) the algorithm receives a set of localities (that is, the Italian municipalities) $M \subset A$ and selects those that could be used as focal points for building LLMAs. For each candidate locality $a_i \in M$, it evaluates $sc^w(a_i)$ and $jr(a_i)$ and chooses the localities that have values in the top 20% for either of the two measures. In phase (2) the algorithm amalgamates focal points that exhibit a high degree of interaction. For focus a_i, having a high degree of interaction with another focus a_j means that: either supply-side or demand-side self-containment must be less than 0.5; and that a_i receives at least 10% of the flows coming out of a_j, that is $f_{ji} \geq 0.1(c^+(a_j) + f_{jj})$; and that a_j receives at least 1% of flow coming out of a_i, that is $f_{ij} \geq 0.01(c^+(a_i) + f_{ii})$. Therefore, all focal points are sorted in descending order by their value of incoming flows $c^-(a_i)$. In this order, if some a_i has $\min\{sc^w(a_i), sc^r(a_i)\} < 0.5$, then it is merged with the a_j that maximizes the following weighted interaction index, provided that this exceeds 0.002.

$$\frac{f_{ji}^2}{(f_{jj} + c^+(a_j)(f_{ii} + c^-(a_i))} + \frac{f_{ij}^2}{(f_{ii} + c^+(a_i)(f_{jj} + c^-(a_j))}. \tag{5}$$

The new combined locality replaces both a_i and a_j and is considered as a focus. This process continues until no more of such amalgamations can be done. At the end of this phase we obtain a small number of localities containing a single municipality or a few of those. In phase (3) the algorithm expands focal points to proto-LLMAs by allocating them localities with which they have a high degree of interaction. Hence, each proto-LLMA will be a set P of localities. This phase considers links between the focal points themselves and also between focal and non focal localities. All focal points are sorted in descending order of the following function, where t_1=t_2=0.75, t_3=1000.

$$F(a_i) = \left(\min\left\{\frac{sc^w(a_i)}{t_1}, \frac{sc^r(a_i)}{t_2}, 1\right\}\right) \cdot \left(\min\left\{\frac{w(a_i)}{t_3}, 1\right\}\right) \tag{6}$$

If some a_i has $F(a_i) < 0.75$, then it is merged in a P_k with another locality a_j such that: $F(a_j) < 0.75$ and a_i receives at least 10% of the flows coming out of a_j and a_j is the one that, when merged to a_i, maximizes the above $F(P_k)$. Any locality a_i such that $F(a_i) \geq 0.75$ is a proto-LLMA by itself.

In phase (4) the algorithm allocates remaining non-focal localities to existing proto-LLMAs. At first it iteratively dismembers groups of areas with $F(P_k) < 0.75$, in order to reallocate their localities. After this, the set of localities not yet allocated to proto-LLMAs is sorted in decreasing order of in-flow $c^-(a_i)$, and each of them is joined with the proto-LLMA with which it show the strongest connection, i.e., the one maximizing (5). Only localities without in-flows and out-flows are left isolated. Each of such localities will be joined with the nearer contiguous LLMA at the end of the procedure. Then, iteratively, proto-LLMAs are again checked and those having $F(P_k) < 0.75$ are dismembered and its constituent localities are once again joined to the remaining proto-LLMA just like at the beginning of phase 4. The phase continues until there are only proto-LLMAs with $F(P_k) \geq 0.75$. In phase (5), finally, the algorithm checks if each proto-LLMA is contiguous. Those that are not contiguous are dismembered and each of its constituent localities is joined to the contiguous proto-LLMA that maximizes (5), similarly to phase (4).

4.2 Functional Regionalization based on Graph Partitioning

Approaches based on graph partitioning have been used for functional regionalization since pioneering works. Recently, similar techniques are used to find *network communities* in real-world networks. Due to some similarity of the two problems, methods used for network community detection provide a possible approach for the identification of LLMAs. An analysis of the main heuristic used for network community detection is in Leskovec *et al.* [27]. They explore a range of such methods in order to understand their relative performance and the systematic biases in the networks community they identify. In particular, they compare structural properties of network community extracted by two completely different computational paradigms: a spectral based graph partitioning method Local Spectral [1] and the multilevel partitioning algorithm METIS [24]. Analyzing the network community profile (NCP) of a large number of communities, they highlight that METIS is generally better than Local Spectral at the nominal task of finding cuts with low conductance, although some of METIS clusters may be internally disconnected.

We now briefly describe METIS algorithm. It is a general purpose graph partitioner, not specifically designed for the identification of LLMAs, but aiming at

finding high quality partitions in short times for large problems. METIS initially coarsens the graph by generating a sequence of graphs $G_0, \ldots, G_m$ such that G_0 is the original graph, and, for each $j \geq i$, the number of vertices in G_j is smaller than the number of vertices in G_i. Each coarser graph is obtained by searching for a maximal matching and by collapsing the couples of adjacent vertices corresponding to the edges of the matching. Coarsening a graph using matchings preserves many properties of the original graph. After this, METIS computes a recursive bisection, or directly a k-way partition, of the smaller graph G_m by minimizing the sum of the weights of the edges crossing the partitions. Since the size of the coarser graph G_m is small (i.e., $|V_m| < 100$), this step takes a small amount of time. Finally, METIS projects the partition back through the sequence of $G_m, \ldots, G_0$, using also, in each step, local refinement heuristics to improve the projected partition of each graph, until producing the partition of the original graph G_0.

5 Experimental Results

The procedure described in Section 3 was implemented in C++ and tested by generating the LLMAs for all the Italian administrative regions. The commuting flows considered for this test were gathered via the 2001 Italian Population Census. This Census collects data about inter-municipal commuting flows. Thus, municipalities are the localities constituting our basic units of data. The target was to meet conditions (4), (*iii*) and (*iv*), with $c = 0.75, c_1 = 0.75, c_2 = 1000, c_3 = 2$. The experiments were conducted on a PC Intel Pentium CPU 3.10 Ghz with 4Gb of RAM under MS windows7 64 bit Operating System.

We first compare our algorithm with the reference algorithm based on [37] and described in Section 4.1. For each Italian administrative region, Table 1 reports the number of LLMAs generated by our algorithm based on minimum cuts; their average value of Overall self-containment (mean sc); their minimum value of Demand- and Supply-side self-containment (min sc^r, min sc^w); time required for processing the whole administrative region; the number of LLMAs generated by the reference algorithm; all the values already provided for the first algorithm. Conditions (4), (*iii*) and (*iv*) are always satisfied by the two approaches. Times are in minutes and are computed on the same machine. Although both methods were successful in generating feasible LLMAs, our approach showed to be able to generate a number of LLMAs that is (often considerably) larger than the reference algorithm in all cases. In addition, the values of self-containment of the LLMAs generated by our approach are closer (from above) to the self-containment threshold than those provided by the reference algorithm. These are positive features for a regionalization algorithm, as explained in Section 2. Hence, the proposed method provides LLMAs with better statistical quality. Times needed by the proposed algorithm are generally much shorter than those of the reference algorithm. While the total running time of our procedure, for all Italian administrative regions, is 8 hours and 22 minutes, the total running time of the reference procedure is 16 hours.

Then, we compare our algorithm with the state-of-the-art graph partitioner METIS [24]. METIS takes in input the number of clusters (the LLMAs in our case) that should be produced. This number was assigned to the same number of LLMAs produced by our algorithm in order to make a fair comparison of the regionalization quality. For each Italian administrative region, Table 2 reports the

number of vertices (localities) and of edges (linkages) of the transitions graph, and an analysis of the self-containment of the LLMAs produced by the two algorithms, by giving their minimum and maximum values, their average and their variance. METIS produces LLMAs that often are not enough self-contained (they often do not satisfy conditions (4) and (*iii*)), and the self-containement values of the LLMAs generated by METIS are much more variable than those of the LLMAs generated by our algorithm. Indeed, some lack of internal cluster connection is quite intrinsic in METIS approach. Therefore, the proposed algorithm appears a better option for performing a similar functional regionalization, although METIS is a faster graph partitioner.

	Our Algorithm					Reference Algorithm				
Region	LLMAs	mean sc	min sc^r	min sc^w	time	LLMAs	mean sc	min sc^r	min sc^w	time
Piemonte	36	0.69	0.67	0.71	90	35	0.70	0.71	0.75	239
V. d'Aosta	4	0.71	0.73	0.73	4	3	0.83	0.89	0.86	21
Lombardia	55	0.65	0.63	0.68	320	54	0.67	0.66	0.76	452
Trentino AA	33	0.66	0.61	0.72	4	28	0.74	0.64	0.76	16
Veneto	37	0.66	0.65	0.70	19	35	0.70	0.65	0.76	34
Friuli VG	10	0.68	0.65	0.75	3	6	0.75	0.65	0.78	7
Liguria	17	0.68	0.62	0.72	3	16	0.72	0.64	0.77	13
Emilia Rom.	45	0.68	0.63	0.74	4	41	0.68	0.64	0.76	10
Toscana	47	0.69	0.68	0.69	3	40	0.73	0.73	0.76	7
Umbria	15	0.74	0.66	0.78	1	14	0.73	0.68	0.77	5
Marche	34	0.66	0.67	0.73	2	29	0.69	0.67	0.76	7
Lazio	21	0.70	0.62	0.80	7	18	0.71	0.63	0.80	23
Abruzzo	22	0.68	0.64	0.64	4	20	0.70	0.66	0.76	15
Molise	13	0.69	0.66	0.72	1	9	0.76	0.72	0.80	5
Campania	54	0.66	0.66	0.68	16	49	0.67	0.62	0.75	32
Puglia	48	0.68	0.65	0.67	3	34	0.71	0.68	0.75	8
Basilicata	21	0.71	0.67	0.78	1	18	0.72	0.72	0.76	5
Calabria	60	0.67	0.64	0.69	7	47	0.72	0.66	0.79	25
Sicilia	66	0.69	0.63	0.70	5	48	0.74	0.70	0.76	17
Sardegna	44	0.69	0.65	0.69	5	38	0.72	0.69	0.76	18

Table 1 Comparison of our and reference algorithm on all Italian administrative regions

6 Conclusions

We proposed an innovative approach to the problem of the generation of LLMAs by using techniques of Combinatorial Optimization. This procedure is based on the iterative partitioning of the transitions graph, which represents the interaction among the localities of the territory under analysis. The proposed procedure works at the formal level, hence it can be used for other problems of different origin but sharing the same structure. Since the arising minimum cut problems can be solved to optimality in extremely short times by using state-of-the-art min-cut algorithms, the procedure is able to generate LLMAs in large real-world networks such as the Italian administrative regions in times that are very reasonable and much shorter than the official reference method used for comparison. The statistical quality of the partitions generated by the proposed method is generally better than that obtained by using the reference method, and clearly better than that obtained by using a general purpose graph partitioner not specifically designed for this task.

Region	$\mid V \mid$	$\mid E \mid$	LLMAs	Our Algorithm				METIS Algorithm			
				min *sc*	max *sc*	mean *sc*	var *sc*	min *sc*	max *sc*	mean *sc*	var *sc*
Piemonte	1206	35903	36	0.57	0.89	0.69	0.007	0.26	0.85	0.61	0.033
V. d'Aosta	74	1021	4	0.64	0.89	0.71	0.010	0.58	0.85	0.70	0.017
Lombardia	1546	91625	55	0.56	0.87	0.65	0.006	0.27	0.89	0.55	0.023
Trentino AA	339	5826	33	0.57	0.95	0.66	0.006	0.21	0.89	0.61	0.034
Friuli VG	219	7168	37	0.57	0.82	0.66	0.006	0.31	0.89	0.61	0.020
Veneto	581	24926	10	0.57	0.87	0.68	0.009	0.27	0.90	0.65	0.035
Liguria	235	3523	17	0.57	0.95	0.68	0.017	0.34	0.95	0.64	0.040
Emilia Rom.	341	9529	45	0.57	0.85	0.68	0.006	0.29	0.85	0.58	0.024
Toscana	287	7462	47	0.58	0.91	0.69	0.006	0.30	0.92	0.63	0.025
Umbria	92	1321	15	0.58	0.87	0.74	0.008	0.39	0.90	0.67	0.032
Marche	246	5586	34	0.56	0.84	0.66	0.005	0.30	0.95	0.60	0.023
Lazio	378	8729	21	0.56	0.92	0.70	0.007	0.27	0.89	0.59	0.036
Abruzzo	305	5615	22	0.56	0.87	0.68	0.010	0.32	0.86	0.60	0.033
Molise	136	1785	13	0.59	0.82	0.69	0.006	0.38	0.86	0.61	0.023
Campania	551	20590	54	0.56	0.80	0.66	0.004	0.33	0.84	0.58	0.017
Puglia	258	7372	48	0.56	0.83	0.68	0.005	0.36	0.89	0.60	0.020
Basilicata	131	1958	21	0.61	0.86	0.71	0.004	0.39	0.90	0.68	0.019
Calabria	409	8247	60	0.57	0.87	0.67	0.004	0.30	0.88	0.62	0.015
Sicilia	390	9189	66	0.57	0.93	0.69	0.007	0.33	0.89	0.65	0.019
Sardegna	377	7686	44	0.56	0.88	0.69	0.008	0.31	0.90	0.62	0.028

Table 2 Comparison of our algorithm and METIS on all Italian administrative regions

References

1. Andersen, R., Chung, F., Lang, K.: Local graph partitioning using PageRank vectors. In Proc. 47th Annual IEEE Symposium on Foundations of Computer Science, 475-486 (2006)
2. Bichot, C.E., Siarry, P.: Graph Partitioning. ISTE - Wiley, London (2011)
3. Brinkmeier, M.: A Simple and Fast Min-Cut Algorithm. Theory of Computing Systems 41(2), 369-380 (2007)
4. Canello, J., Pavone, P.: Mapping the Multifaceted Patterns of Industrial Districts: A New Empirical Procedure with Application to Italian Data. Regional Studies http://dx.doi.org/10.1080/00343404.2015.1011611 (2015)
5. Casado-Díaz, J.M., Coombes, M.G.: The delineation of 21st century local labour market areas: a critical review and research agenda. Boletín de la Asociación de Geógrafos Españoles 57, 7-32 (2011)
6. Coombes, M.G., Openshaw, S.: The Use and Definition of Travel-to-Work Areas in Great Britain: Some Comments. Regional Studies 16(2), 141-149 (1982)
7. Coombes, M. G., Green, A.E., Openshaw, S.: An efficient algorithm to generate official statistical reporting areas: the case of the 1984 Travel-to-Work Areas revision in Britain. Journal of the Operational Research Society 37, 943-953 (1986)
8. Coombes, M.G.: Defining boundaries from syntetic data. Environment and Planning 32, 1499-1518, 2000.
9. Dahmann, D.C., Fitzsimmons, J.D. (eds.) Metropolitan and nonmetropolitan areas: New approaches to geographical definition. Bureau of the Census Working Paper 12, Washington: Bureau of the Census (1995)
10. Diestel, R.: Graph Theory, 4th edition, Springer (2010)
11. Di Giacinto, V., Gomellini, M., Micucci, G., Pagnini, M.: Mapping local productivity advantages in Italy: industrial districts, cities or both? Journal of Economic Geography 14 (2), 365-394 (2014)
12. Duque, J.C., Ramos, R., Suriach, J.: Supervised regionalization methods: a survey. International Regional Science Review 30, 195220 (2007)
13. Farmer, C.J.Q., Steward Fotheringam, A.: Network-based functional regions. Environmenl and Planning 43, 2723-2741 (2011)
14. Fischer, M.M.: Regional taxonomy: A comparison of some hierarchic and non-hierarchic strategies. Regional Science and Urban Economics 10, 503-537 (1980)

15. Flórez-Revuelta, F., Casado-Díaz, J.M., Martínez-Bernabeu, L.: An evolutionary approach to the delineation of functional areas based on travel-to-work flows. International Journal of Automation and Computing 5(1), 10-21 (2008)
16. Fortunato, S., Barthélemy, M.: Resolution limit in community detection. Proceedings of the National Academy of Sciences of the United States of America, 104(1), 36-41 (2007)
17. Fusco, G., Caglioni, M.: Hierarchical clustering through spatial interaction data. The case of commuting flows in South-Eastern France. LNCS 6782, 135-151 (2011)
18. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America 99(12), 7821-7826 (2002)
19. Goodman, J.F.B.: The definition and analysis of local labour markets: some empirical problems. British Journal of Industrial Relations 8, 179-186 (1970)
20. Guimerá, R., Sales-Pardo, M., Amaral, L.: Modularity from fluctuations in random graphs and complex networks. Physical Review E, 70:025101, (2004)
21. Hao, J.X., Orlin, J.B.: A Faster Algorithm for Finding the Minimum Cut in a Directed Graph. Journal of Algorithms 17(3), 424-446 (1994)
22. Hopcroft, J., Tarjan, R.: Efficient algorithms for graph manipulation. Communications of the ACM 16(6), 372-378 (1973)
23. ISTAT: I sistemi locali del lavoro 2011. Nota metodologica. Retrieved from http://www.istat.it/it/files/2014/12/nota-metodologica_SLL2011_rev20150205.pdf (2015)
24. Karypis, G., Kumar, V.: A fast and high quality multilevel scheme for partitioning irregular graphs. SIAM Journal on Scientific Computing 20(1), 359-392 (1999)
25. Kim, H., Chun, Y., Kim, K.: Delimitation of Functional Regions Using a p-Regions Problem Approach. International Regional Science Review doi:10.1177/0160017613484929 (2013)
26. van der Laan, L., Schalke, R.: Reality versus Policy: The Delineation and Testing of Local Labour Market and Spatial Policy Areas. European Planning Studies 9(2), (2001)
27. Leskovec, J., Lang, K.J., Mahoney, M.W.: Network Empirical Comparison of Algorithms for Community Detection. In Proceedings of the 19th International Conference on World Wide Web, WWW 2010, April 26-30, AACM New York, NY, USA (2010)
28. Martínez-Bernabeu, L., Flórez-Revuelta, F., Casado-Díaz, J.M.: Grouping genetic operators for the delineation of functional areas based on spatial interaction. Expert Systems with Applications 39, 6754-6766 (2012)
29. Masser, I., Brown, P.J.B.: Hierarchical aggregation procedures for interaction data. Environment and Planning 7, 509-523 (1975)
30. Nagamochi, H., Ibaraki, T.: Computing edge-connectivity in multigraphs and capacitated graphs. SIAM J. Discrete Mathematics 5, 54-66 (1992)
31. Nemhauser, G.L., Wolsey, L.A.: Integer and Combinatorial Optimization. J. Wiley, New York (1988)
32. Orasi, A., Sforzi, F.: I sistemi locali del lavoro. ISTAT research report published online: http://dawinci.istat.it/daWinci/jsp/md/download/sll_comunicato.pdf. (2001)
33. Ricca, F. , Scozzari, A., Simeone B.: Political Districting: from classical models to recent approaches. Annals of Operations Research 204, 271-299 (2013)
34. Sforzi, F., Openshaw, S., Wymer, C.: La delimitazione di sistemi spaziali sub-regionali: scopi, algoritmi, applicazioni, 3rd AISRe Annual Confer., Venezia, 10-12 November (1982)
35. Sforzi, F. (ed.): I mercati locali del lavoro in Italia. ISTAT-IRPET, Seminario su: Identificazione di sistemi territoriali. Analisi della struttura sociale e produttiva in Italia, Roma 3-4 December (1986)
36. Sforzi, F. (ed.) I mercati locali del lavoro in Italia. Franco Angeli, Milan, Italy (1989)
37. Sforzi, F., Openshaw, S., Wymer, C.: La procedura di identificazione dei sistemi locali del lavoro. ISTAT, Argomenti 10, 235-247, Rome, Italy, (1997)
38. Sforzi, F.: From Administrative Spatial Units to Local Labour Market Areas. Some Remarks on the Unit of Investigation of Regional Economics with Particular Reference to the Applied Research in Italy. In: Fernández Vázquez, E., Rubiera Morollón, F. (eds.) Defining the Spatial Scale in Modern Regional Analysis, 3-21, Springer, Berlin (2012)
39. Smart, M.W.: Labour market areas: uses and definition. Progress in Planning 2(4), 239-353 (1974)
40. Stoer, M., Wagner, F.: A simple min-cut algorithm. J. of the ACM, 44(4), 585-591 (1997)
41. van der Zwan, J., van der Wel, R., de Jong, T., Floor, H.: Flowmap 7.2 Manual, Faculty of Geosciences, University of Utrecht (2005)
42. White, S., Smyth, P.A.: Spectral clustering approach to finding communities in graphs. In Proceedings of the 5th SIAM International Conference on Data Mining, 76-84 (2005)