Specific crossover and mutation operators for a grouping problem based on interaction data in a Regional Science context

Francisco Flórez-Revuelta, Member, IEEE, José M. Casado-Díaz, and Lucas Martínez-Bernabeu

Abstract—This paper proposes a set of specific crossover and mutation operators for the delineation of functional regions through evolutionary computation. We consider a problem of dividing a given territory into local labor market areas based on spatial interaction data. Such areas are defined so that a high degree of inter-regional separation and intra-regional integration—in both cases in terms of commuting flows—exist. A genetic algorithm has been designed based on the maximization of a fitness function that measures aggregate intra-region interaction under constraints of inter-region separation and minimum size. Additional requirements, typical of any functional regionalization, include the absence of overlapping between delineated regions and an exhaustive coverage of the whole territory (so all basic spatial units must be allocated to one—and only one—region). The complex set of restrictions results in conventional operators often generating invalid solutions, impeding or delaying the evolutionary process. This is the reason why an extensive set of operators has been designed that incorporates knowledge about the problem, allowing the evolution of the set of solutions towards the final result.

I. INTRODUCTION

In order to design, implement and monitor effective labor market policies developed countries have in the last decades embarked in the delineation of local labor market areas (LLMAs) [1]. These sets of functional areas are also seen as an alternative to the use of local and regional administrative areas as the relevant geography for statistical purposes and in other public policy areas such education and housing markets. The reason for this is that administrative areas are defined by boundaries that very frequently derive from historical reasons, and so it is not assured that they provide a meaningful insight of the territorial functional reality. Although some countries have delineated homogeneous LLMAs through the aggregation of basic territorial units such as counties, wards or municipalities of similar features (a review of this procedures can be found in [2]), the most popular strand consists in the aggregation of building blocks based on the observed interaction among them in terms of travel-to-work LLMAs.

The ideal LLLMA should be defined by boundaries rarely crossed by people in their journeys to work so as to embrace the majority of the interaction between workers seeking jobs and employers recruiting labor. This quality refers to what Goodman [3] called external perfection (the boundary of the area is rarely crossed in daily journeys to work) and together with a high degree of intra-market interaction (so that the defined market is internally active and so as unified as possible) forms the basis of the ideal LLLMA. In 1992 Eurostat [4] established a code of good practices to guide the selection of a specific procedure: (a) the ideal map of LLMAs should be based on statistical criteria, thus defined in a consistent way to allow comparison for statistical and policy purposes, (b) the procedure should allow the delineation of boundaries between areas within which most people both live and work, (c) each basic spatial unit should be in one, and only one LLLMA, (d) contiguity should be respected, (e) a certain degree of self-containment should be reached, so that most of the LLLMA’s workers live in that area and most of the area’s employed residents should work locally, (f) the map should consist of homogeneous units whose size should overpass a minimum threshold, (g) the areas defined should not be unnecessarily complex from a topographic point of view, (h) the map of LLMAs should respect where possible the standard administrative top tier boundaries, this being considered advantageous from both statistical and policy points of view and finally (i) the procedure should be flexible enough to allow evaluation and adjustment, although the possibility of varying the statistical criteria between regions must be excluded. As pointed out in [5] the preference for detail (delineating as many criteria-meeting LLMAs as possible) is also frequently included as one additional criterion.

Despite sharing a common basic view about the ideal features of such an area, current official methods have a very diverse nature and are mostly based in sets of rules whose sophistication substantially varies nationally and, to
a certain degree, temporally. In [5] several classifications of these official procedures are presented. One of the procedures that has been more successfully applied is that of Coombes et al. [6] which has been used in the United Kingdom for the delineation of LLMAs (so-called Travel-to-Work Areas, TTWAs) since the 80s. This sophisticated procedure was also used, with minor changes, to define LLMAs in Italy [7]- [9], Spain [10], New Zealand [11] and Australia [12], among other countries. This is the procedure that inspires the one proposed in the article. In our proposal the regionalization problem is presented as the maximization of markets’ internal cohesion in terms of travel-to-work subject to a number of restrictions among which stands meeting certain self-containment and minimum size (in terms of occupied population) thresholds, with the aim of identifying as many independent markets as possible, and without making use of contiguity constrictions or distance measures. Unlike most current procedures, the method proposed here is able to meet the criteria listed above and allows a significant improvement in measurable indicators such as the number of LLMAs identified compared with alternative methods.

Given the complexity of the problem, an exhaustive search of the solution is not possible in many cases, and this is the reason why an evolutionary approach is proposed. A deep knowledge of the problem has allowed the design of an extensive set of crossover and mutation operators, some of which have similarities with those used in other grouping and clustering problems to which they could be immediately applied, whilst other are much more related to the very specific nature of the problem, i.e. the delineation of LLMAs. This is a strategy that has proved to allow reaching the final solution, and to do so much more rapidly [19]. We illustrate our approach using the latest Census data available for Spain [13].

II. PROBLEM FORMULATION

Let $A = \{A_1, A_2, \ldots, A_s\}$ be the territory (set of areas). The objective is to divide it into the set of disjoint regions $R = \{R_1, R_2, \ldots, R_r\}$ so as $\bigcup R_i = A$ and $R_i \cap R_j = \emptyset, \forall i, j \in [1,r], i \neq j$, that maximizes a fitness function based on the interaction index II between an area and the rest of the region to which it belongs

$$II(k) = \frac{W^2_{[k],R\setminus k}}{W_{[k],A} \cdot W_{A,R(k)\setminus k}} + \frac{W^2_{R(k)\setminus k,k}}{W_{R(k)\setminus k,A} \cdot W_{A,k}}, k \in A$$  

being $R(k)$ the region containing area $k$, and

$$W_{R_i,R_j} = \sum_{\forall i \in R_i} \sum_{\forall j \in R_j} W_{ij}$$  

where $W_{ij}$ is the number of commuters from area $i$ to area $j$, that is the number of employed residents in area $i$ that work in area $j$.

Among the possible fitness function based on the interaction index (1) we can find:

$$f_1 = \sum_{\forall i \in A} II(i) \quad \text{or} \quad f_2 = \text{card}(R) \times \sum_{\forall i \in A} II(i).$$

Besides, every region $R_i \in R$ must fulfill two constraints in terms of self-containment ($\beta_1$), and minimum size ($\beta_2$):

$$\min \left\{ \frac{W_{R_i,R_i}}{W_{R_i,A}} : \frac{W_{R_i,R_i}}{W_{R_i,A}} \geq \beta_1 \right\}$$  

(3)

$$W_{R_i,A} \geq \beta_2$$  

(4)

A trade-off between both constraints has been introduced similarly to [6], but in the formulation proposed by Casado [10]. According to this trade off, the self-containment requisite is relaxed for regions which are sufficiently large following a linear relationship, and therefore even if conditions (3) and (4) are not met a set of municipalities (a region), qualifies as LLMA if:

$$\min \left\{ \frac{W_{R_i,R_i}}{W_{R_i,A}} : \frac{W_{R_i,R_i}}{W_{R_i,A}} \geq a + b \cdot W_{R_i,A} \right\}$$  

(5)

where $a = \beta_2 - b \beta_4$, $b = \frac{\beta_2 - \beta_1}{\beta_4 - \beta_1}, \beta_4 \geq \beta_1$ and $\beta_1 \leq \beta_1$.

We have also included a requisite to guarantee some degree of contiguity by employing only commuting data: an area can only belong to a region if some of the $\gamma$ areas to/from it has more output/input commuting flows is also part of that region.

III. EVOLUTIONARY PROPOSAL

Our approach is based on a genetic algorithm with hard selection (the best individuals from parent and offspring populations are selected). The structure of the evolutionary algorithm is as follows:

Step 1. Produce an initial population of size $n$. The whole set of areas (the whole territory $A$) is taken as individual one (this is an individual meeting all the requirements). Complete the initial population with $n-1$ randomly generated individuals (experimentation showed that most of these are invalid solutions).

Step 2. Evaluate fitness of all individuals and sort them accordingly.

Step 3. Generate $nr$ new individuals by recombination, selecting the parents from the population by fitness-proportional probability, and selecting the operator to apply according to a pre-established probability.

Step 4. Generate $nm$ new individuals by mutation, selecting the original individual from the population by fitness probability, and selecting the operator to apply according to a pre-established probability.
according to a pre-established probability.

Step 5. Evaluate fitness of all new individuals.

Step 6. Sort the whole population, composed of \( n+n+r+n+m \) individuals, by their fitness value.

Step 7. Generate a new population choosing the n best individuals.

Step 8. Stop condition: if the best individual remains without changes for \( g \) generations, finish. Otherwise, return to step 3.

A. Individual Representation

The individuals of the population represent feasible solutions, that is, the aggregation of all the geographical basic areas composing territory \( A \) into no overlapping local labor markets (regions). There are different alternatives for the encoding of a grouping of a data set [14]-[16]. In our case, we have chosen a group-number encoding [17] where each individual is represented by a vector of \( n \) components, each of which corresponds to an area of \( A \), and takes the value of the identifier of the region the area belongs to (Fig. 1).

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
</table>

Region \( R_i = \{A_1, A_2, A_3\} \)
Region \( R_j = \{A_2, A_3, A_4\} \)
Region \( R_k = \{A_1, A_3\} \)
Region \( R_r = \{A_2\} \)

Fig. 1. Individual representation

B. Selection

Selection of the individuals to be affected by recombination and mutation operations is performed following a ranking method [18], according to which those individuals scoring higher in the fitness function have a larger probability of being selected.

C. Recombination Operators

Due to the large number of constraints that the individuals must fulfill, and very notably to the fact that in a regionalization exercise it is important to guarantee the exhaustive coverage of the territory and the avoidance of overlapping between regions, the usual operator of recombination does not in many cases lead to valid solutions. This is the reason why we have designed a wide group of specific operators that has proved to allow a more rapid evolution of the population towards acceptable solutions:

1) Recombination1: A crossover point is randomly selected. Offspring is generated by taking the initial part of one of the parents and the final part of the other one. This is the standard 1-point crossover operator. However, in this specific case this operator frequently results in invalid offspring due to the lack of a compatible correspondence between the region identifiers of both parents (see Figure 2 – region 3 in parent #1 and region 4 in parent #2 are identical but they are codified with a different region identifier, resulting in a fragmentation of the region in the offspring).

![Fig. 2. Operator Recombination1](image)

2) Recombination2: A region identifier belonging to parent #1 is randomly chosen. The areas with identifiers lower or equal to the chosen one are inherited by the offspring. Each of the rest of the areas takes the identifier from parent #2, except in those cases when it belongs to a region from which one or more of its constituting areas were already in the offspring. In such cases, the areas take the identifiers from parent #1 (see Figure 3 – areas in parent #2 belonging to region \( R_j \) must be assigned to \( R_i \) in the offspring).

![Fig. 3. Operator Recombination2](image)

3) Recombination3: a crossover point is randomly selected. For the areas previous to that point, the offspring takes the values of parent #1. From that crossover point, values from parent #2 are inherited, unless this involves a region with an area already set in the offspring; in such cases the identifier of parent #1 is used (see Figure 4 – areas in parent #2 belonging to \( R_j \) and \( R_k \) must be assigned to \( R_i \) and \( R_k \) respectively).

![Fig. 4. Operator Recombination3](image)

4) Since the areas characterized by lower identifiers are also assigned to regions with lower identifiers, their probability of being taken from parent #1 is greater than that of areas with high identifiers. To cope with this we have added two recombination operators (Recombination4 and Recombination5), as variations of Recombination2 and Recombination3 respectively. In them a random recoding of the regions in the representation of both parents is performed previously to the recombination.
D. Mutation Operators.

We have designed an extensive set of mutation operators, some of them specifically intended for the delineation of local labour market areas, with the aim to accelerate the obtaining of individuals with adequate fitness. These operators have four main functions: division of regions, fusion of regions, reassignment of single areas, and reassignment of group of areas. Fig. 5 shows the results of the application of several of these specific operators.

1) Division1: This operator divides a region into two. The splitting process is as follows:
   a) A region $R_i$ is randomly selected. This region must fulfill two constraints: (a) $W_{R_i,A} > 2\beta_4$ and (b) $W_{R_i,A} - W_{\text{focus}(R_i),A} > \beta_4$

   $$\text{focus}(R_i) = \arg \max_{R \in \text{children}(R_i)} \left( W_{[i],A} + W_{A,[i]} \right)$$
   (that is, the region is large enough).

   b) An area belonging to $R_i$ is randomly chosen. It is then assigned to the new region $R'_i$.

   c) Another area belonging to $R_i$ is randomly chosen. It is then assigned to the new region $R''_i$.

   d) The rest of the areas belonging to $R_i$ are taken at random, being assigned to the region ($R'_i$ or $R''_i$) to which each of them has a stronger link according to the interaction index.

2) Division2: This operator creates a new region from another one by removing from the latter a number of areas sufficiently large so as to form a valid market:
   a) As in step a) of Division1.

b) An area belonging to $R_i$ is chosen at random, being assigned to the new region $R'_i$.

c) If region $R'_i$ does not fulfill the size constraint (4), it takes the area belonging to $R_i$ with which it has a higher interaction index. This process is repeated until $R'_i$ is large enough.

3) Division3: This operator divides a region to form two regions with a similar number of areas:
   a) As in steps a) to c) of Division1.

b) $R'_i$ y $R''_i$, alternately, take the area of $R_i$ to whom they have a higher interaction index, until $R_i$ is empty.

4) Fusion1: Two randomly selected regions are merged.

5) Fusion2: A region is randomly chosen. Each of its constituting areas is then merged with its optimal region, that is, the region with a higher index interaction with it:

$$R'(i) = \arg \max_{R \in \text{children}(R(i))} \left( \frac{W^2_{[i],R_i}}{W_{[i],A} \cdot W_{A,[i]}} + \frac{W^2_{R_i,[i]}}{W_{R_i,A} \cdot W_{A,[i]}} \right)$$

So, as a result of this operator, the number of regions in the offspring is one less compared to its parent.

6) Reassignment1: Similar to the standard mutation operator in EC, randomly reallocating to any region up to 1% of the areas in the territory.

7) Reassignment2: This operator is analogous to Reassignment1. However, the destination region for each mutated area is its optimal region according to (6) (as in Fusion2).

8) Reassignment3: An exchange of areas between regions is performed. One area is randomly chosen and assigned to

Fig. 5. Examples of the application of different mutation operators.
its optimal region. One random area of that optimal region is then transferred to the source region.

9) **GlobalReassignment1**: This operation removes from a region the areas that score lower in the interaction index when measured with regards to the rest of the region. Such areas are then assigned to their optimal regions:
   a) As in step a) of **Division1**.
   b) The area to remove is selected as:
   
   $$ s = \arg \min_{\forall j \in R_i} \left( \frac{W(R_i, -i, A_{R_i} \cdot W(A_{R_i}, -i, A_{R_i} \cdot W(A_{R_i}, -i, A_{i}))}{W_{|i|, A_{R_i} \cdot W(A_{R_i}, -i, A_{R_i} \cdot W(A_{R_i}, -i, A_{i}))}} \right) $$

   c) If $W_{R_i, -i, A} > \beta_4$ (R_i is large enough), the area s is assigned to its optimal region, and step b) is repeated. If that condition is not fulfilled, mutation is finished.

10) **GlobalReassignment2**: This operator is similar to **Reassignment2** in the sense that areas are assigned to their optimal regions. In this case, however, instead of a single area a group of them is transferred. Such a group is chosen so that the relationships among its component areas are high. The process is as follows:
   a) An area i is randomly selected.
   b) The k areas belonging to $R_i$ with which area i has more interaction are also selected. k is chosen at random.
   c) All the selected areas are assigned to the optimal region for area i.

11) **GlobalReassignment3**: As, in some cases, there is a great interaction between regions, this operator tries to redistribute areas in such regions. The procedure is:
   a) A number $k \geq 2$ of regions to be mutated is randomly chosen.
   b) A region $R_i$ is selected at random.
   c) The k-1 regions that have a higher degree of interaction with $R_i$ are selected.
   d) These regions are then disintegrated into their constituting areas.
   e) k areas from this new group are selected at random. These areas act as seeds for the new regions.
   f) The rest of unassigned areas are individually taken at random and merged with their optimal region among those k new regions.

12) **GlobalReassignment4**: This operator is very similar to the previous one. Only the way in which new regions are created has been modified:
   a) As in steps a) to d) of **GlobalReassignment3**.
   b) For each one of the k new regions:
      i. A random area is selected as its seed.
      ii. This new region takes the areas from $R_i$ with whom it has more interaction, until the size of the region is greater than $\beta_4$ (fulfilling equation 5).
   c) Unassigned areas are merged with their optimal region among those k new regions.

IV. **EXPERIMENTATION**

To demonstrate the use of our approach we have delineated the set of local labor markets in the Region of Valencia with size constraints applying Coombes traditional method (top) and our evolutionary approach (bottom).
TABLE I
NUMERICAL RESULTS OF THE DELINEATION

<table>
<thead>
<tr>
<th>Coombes method</th>
<th>Our evolutionary approach</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best individual</td>
</tr>
<tr>
<td>Number of labor markets</td>
<td>44</td>
</tr>
<tr>
<td>Fitness function</td>
<td>120.23</td>
</tr>
</tbody>
</table>

Spanish Census of Population [13] allowed us to build a 541x541 origin-destination commuting matrix (where 541 is the number \( a \) of municipalities that integrate the territory), where each cell represents \( W_{ij} \).

Diverse goal functions based on the interaction index (1) have been tested. Among them in this paper we consider one of the simplest versions including one of the most easy to quantify criteria of those listed in Section I: detail, i.e. reaching the highest possible number of independent LLMAs,

\[
f = \text{card}(R) \times \sum_{i \in A} II(i)
\]  

(8)

Parameters in the following examples were set in these values: size population \( n=100 \), offspring size=170 \((nr=50\text{ and }mn=120)\) where each recombination or mutation operator generates 10 new individuals, generations \( g \) without changes in the best individual to stop the process=250. Parameter \( \gamma \) of “functional neighborhood” is set to 5. Regarding the self-containment and minimum size conditions (equations 3 to 5) values were set at: \( \beta_1 = 0.7 \), \( \beta_2 = 0.75 \), \( \beta_3 = 20,000 \) and \( \beta_4 = 3,500 \); that is, the levels used for the official delineation of Travel-to-Work Areas in the United Kingdom [6].

As it was previously explained, in step 1 of the evolutionary process, one of the individuals of the original population consists on the whole set of areas taken as a single region. Therefore the process starts with at least one individual that fulfils all the constraints.

Our algorithm has been executed 50 times. Best and average results depicted in Table I are quite straightforward. Although the maps are roughly similar, and many LLMAs remain unchanged, our proposal divides the territory with more detail than the traditional method (Fig. 6). Numerically, we get an improvement of 39 percent in the number of labour markets and, less significantly since it was not explicitly considered in the Coombes et al. [6] procedure, 54 per cent increase in the fitness function.

V. BEHAVIOR OF THE OPERATORS

The purpose of the introduction of specific operators was to improve the search for a good solution, in terms of fitness function, but also in consumed time. To assess this we have monitored their trajectory within the final solution (best individual of the last generation), i.e., the sequence of operators that led from the original individual of the first generation to that solution.

Extensive experimentation showed that the degree of “success” of the diverse operators varies notably. The frequency of their use depends on the specific form of the fitness function used and on the algorithm parameters, and we have also tested that their relative success within one solution is also affected by the territory over which the procedure is applied. In the case of the demonstration included in this paper, results (Table II) reveal that the operators of fusion and Reassignment1 almost never contributed to the final best individual, a fact that has been confirmed in many other of the rest of exercises conducted using different versions of the fitness function, parameter values and territorial contexts. The number of operations that on average allows to reach the final solution departing from the original individual is 267.16, where most of them are recombination, division and some types of local and global reassignment operators.

The contribution of each operator is measured in terms of the average improvement in the fitness function when the best individual of the population changes (Table III) as a result of its action. Division operators produce high increases in the evaluation because the increase in the

<table>
<thead>
<tr>
<th>Operator</th>
<th>Average appearances</th>
<th>Percentage of appearances depending on the evolutionary phase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>First quarter</td>
</tr>
<tr>
<td>Recombination1</td>
<td>27.26</td>
<td>3%</td>
</tr>
<tr>
<td>Recombination2</td>
<td>23.24</td>
<td>5%</td>
</tr>
<tr>
<td>Recombination3</td>
<td>10.82</td>
<td>2%</td>
</tr>
<tr>
<td>Recombination4</td>
<td>33.42</td>
<td>5%</td>
</tr>
<tr>
<td>Recombination5</td>
<td>36.90</td>
<td>5%</td>
</tr>
<tr>
<td>Division1</td>
<td>28.58</td>
<td>75%</td>
</tr>
<tr>
<td>Division2</td>
<td>14.84</td>
<td>79%</td>
</tr>
<tr>
<td>Division3</td>
<td>9.98</td>
<td>80%</td>
</tr>
<tr>
<td>Fusion1</td>
<td>1.84</td>
<td>33%</td>
</tr>
<tr>
<td>Fusion2</td>
<td>2.38</td>
<td>60%</td>
</tr>
<tr>
<td>Reassignment1</td>
<td>0.44</td>
<td>9%</td>
</tr>
<tr>
<td>Reassignment2</td>
<td>22.76</td>
<td>12%</td>
</tr>
<tr>
<td>Reassignment3</td>
<td>6.64</td>
<td>21%</td>
</tr>
<tr>
<td>GlobalReassignment1</td>
<td>27.84</td>
<td>26%</td>
</tr>
<tr>
<td>GlobalReassignment2</td>
<td>2.50</td>
<td>58%</td>
</tr>
<tr>
<td>GlobalReassignment3</td>
<td>4.18</td>
<td>49%</td>
</tr>
<tr>
<td>GlobalReassignment4</td>
<td>13.54</td>
<td>20%</td>
</tr>
</tbody>
</table>

Total 267.16
number of markets is directly reflected in the proposed fitness function \((8)\).

Table IV shows the time cost of applying each of the operators. Recombination operators as well as the rarely successful Reassignment1 are the fastest ones. Introduction of knowledge in the rest of the other mutation operators makes them expensive in terms of temporal complexity. However, their specific contribution so as to reach the best solution, by significantly improving the fitness function, balances this higher temporal cost.

Division operators allow reaching the final number of markets from the original individual where all the areas formed a single region. This process of division is mainly performed at the beginning of the evolution (Table II and Fig. 7).

Once the territory has been broadly divided, experimentation stage shows that the operators of recombination perform a local search in the space of solutions, contributing with small improvements in the value of the fitness function. On the other hand, mutations for the reassignment of individual areas -or groups of them- result in local searches as well as in diversification which allow to departure from local minima. These are the reasons why we consider all of them necessary, despite the fact that some of them hardly ever yield valid solutions.

The heterogeneous behavior of the different specific operators throughout the evolutionary process is mainly related to the value of the fitness function. First phase of divisions produces a fast convergence of the evolution, something that allows to improve the result obtained with the traditional method in less than one hundred generations, almost reaching the final value (Fig. 8).

VI. CONCLUSIONS

The degree of success in the delineation, implementation and monitoring of public policies in different contexts (Statistics, labor markets, housing markets, transportation, urban planning…) heavily depends on the adequateness of the geographical reference. Official methods for the delineation of functional areas which serve as a reference for these purposes have until now rely on procedures that very frequently were designed some decades ago and that can now be improved through the use of new procedures such as evolutionary computation. The use of these techniques has allowed us to model the regionalization problem as one of optimization which is then solved through a genetic algorithm that builds upon one of the most widely and successfully applied procedures, the TTWAs one.

![Fig. 7. Number of individuals in the population generated from a specific set of operators in the beginning of the learning process.](image1)

![Fig. 8. Values of the number of markets and fitness function throughout the evolutionary process.](image2)
The delineation of functional regions consists on the aggregation of the basic spatial units constituting a territory into regions according to certain criteria. Very frequently this aggregation is based on information about the spatial interaction between such units. Examples of such datasets are commuting or migration origin-destination matrices between pairs of municipalities. One way of dealing with this problem is to maximize intra-region interaction under constraints of inter-region separation and (eventually) minimum size. The use of functional areas for policy making and statistical purposes makes it necessary to consider at least two additional requirements: absence of overlapping between regions and exhaustive coverage of the territory (every spatial basic unit must be allocated to one region). Given the complexity of these requirements conventional operators hardly ever generate valid solutions. Therefore our genetic algorithm approach includes developing and testing a new set of operators which has been specifically designed to cope with the complex requirements typical of any functional regionalization, and that could be easily generalized for their use in other grouping problems based on interaction data.

Several improvements of this piece of work are currently being developed. First, the changing success of the operators during the process of generating good individuals is leading us to consider the use of a self-adaptive variation of the method, i.e. the proportion of application of each operator will depend on previous results. Therefore each operator would be applied depending on several criteria, such as success, temporal cost or improvement in the best individual, among others. Second, other representations that have been used in grouping or clustering problems such as the Grouping Genetic Algorithms [14] are currently being explored in order to measure their convergence time. Third, we are designing the parallel implementation of the algorithm to improve convergence in big problems. Finally, as the problem has different components that could eventually be evaluated, a re-formulation of the problem based on multi-objective optimization is also being considered.

REFERENCES


