

A memetic algorithm for the delineation of local labour markets

Francisco Flórez-Revuelta¹, José Manuel Casado-Díaz², Lucas Martínez-Bernabeu¹, and Raúl Gómez-Hernández¹

¹ Research Unit on Industrial Computing and Computer Networks, University of Alicante, P.O. Box 99, E-03080, Alicante, Spain

florez@dtic.ua.es, Lucas.Martinez@ua.es, rgomez@dtic.ua.es
<http://www.dtic.ua.es/i2rc>

² Institute of International Economics, University of Alicante, P.O. Box 99, E-03080, Alicante, Spain

jmcasado@ua.es
<http://iei.ua.es>

Abstract. Given a territory composed of basic geographical units, the delineation of local labour market areas (LLMAs) can be seen as a problem in which those units are grouped subject to multiple constraints. In previous research, standard genetic algorithms were not able to find valid solutions, and a specific evolutionary algorithm was developed. The inclusion of multiple ad hoc operators allowed the algorithm to find better solutions than those of a widely-used greedy method. However, the percentage of invalid solutions was still very high. In this paper we improve that evolutionary algorithm through the inclusion of (i) a reparation process, that allows every invalid individual to fulfil the constraints and contribute to the evolution, and (ii) a hillclimbing optimisation procedure for each generated individual by means of an appropriate reassignment of some of its constituent units. We compare the results of both techniques against the previous results and a greedy method.

Key words: memetic algorithm, evolutionary computation, regionalization, zone design, combinatorial optimization

1 Introduction

Local labour market areas (LLMAs) are geographical entities defined to serve as a territorial framework to design, implement and monitor effective labour market policies and statistical operations at sub-national levels. The success of these key policies crucially depends on the adequacy of the LLMAs delimitation. According to the code of good practices established by Eurostat [1] to guide the selection of a specific procedure, the resulting LLMAs geography must be conformed by disjoint areas exhaustively covering a given territory, characterised by a high degree of self-containment in terms of travel-to-work trips (i.e. most workers in a specific LLMA must live in that area and most of the LLMA's employed

residents should also work locally), and relatively homogeneous in population size (exceeding a minimum size constraint, for instance). The problem is therefore the grouping of basic spatial units (BSU) -such as districts, municipalities or counties- into functional areas so that the proportion of workers that cross their boundaries in their travel to work is low, while the number of defined areas is maximized. This problem is analogous to a Graph Partitioning Problem (GPP) where the optimal number k of partitions is unknown and the requisite of size homogeneity is relaxed or removed, so it is expected to be at least as hard as the standard GPP (that is NP-hard). Thus, an exhaustive resolution of the problem is not possible.

One of the more widely and successfully used official procedures is that of Travel-to-Work Areas (TTWAs) in the UK (it is fully described in [2], and has been applied with minor changes in other countries: [3], [4], [5], and [6]). This regionalization method can be defined as a greedy algorithm that iteratively aggregates a given set of BSUs based on the relative attraction (in terms of commuting flows) between them until all the defined functional areas meet both self-containment and size constraints (in terms of employed population). The method allows reaching adequate solutions with little CPU time.

In order to get solutions closer to the optimal, an evolutionary approach [7] was designed. The multiple constraints which are part of the problem cause the number of valid solutions –those that meet the constraints– to be extraordinarily small with regards to the search space, so standard genetic operators didn't lead to valid solutions in a reasonable lapse of time. This is the reason why an extensive set of specific crossover and mutation operators was proposed [8]. Whilst some of which have similarities with those used in other grouping and clustering problems, others are much more related to the very specific nature of the problem. However, and despite the design of *ad hoc* genetic operators, a high percentage of the individuals generated were not valid. In this paper we propose the inclusion of an intermediate stage in the evolutionary process aimed at 'repairing' every invalid individual. Additionally, we also include a stage of local optimization through optimal reassignment of BSUs chosen at random. The application of both techniques should accelerate the evolutionary process in terms of generations and help to reach solutions closer to the optimal one.

2 Problem formulation

Let $S = \{S_1, S_2, \dots, S_n\}$ be a set of BSUs (the territory to be divided into LLMAAs) and W_{S_i, S_j} the number of commuters from BSU S_i to BSU S_j , that is, the number of residents in S_i that work in S_j (thus, W_{S_i, S_i} is the amount of people who simultaneously live and work within the boundaries of BSU S_i). The objective is to obtain the set of markets (LLMAAs) $M = \{M_1, M_2, \dots, M_m\}$, where m is unknown *a priori*, so as $M_i \neq \emptyset, \forall M_i \in M$; $\bigcup_{i=1}^m M_i = S$ and $M_i \cap M_j = \emptyset, \forall i, j \in [1, m], i \neq j, 1 \leq m \leq n$, that maximizes fitness function f . Let II be the interaction index between two markets:

$$II(M_i, M_j) = \underbrace{\frac{W_{M_i, M_j}}{R_i}}_{PE_{M_i, M_j}} \times \underbrace{\frac{W_{M_i, M_j}}{J_j}}_{PJ_{M_i, M_j}} + \underbrace{\frac{W_{M_j, M_i}}{R_j}}_{PE_{M_j, M_i}} \times \underbrace{\frac{W_{M_j, M_i}}{J_i}}_{PJ_{M_j, M_i}} \quad (1)$$

where

$$W_{M_s, M_t} = \sum_{\forall S_i \in M_s} \sum_{\forall S_j \in M_t} W_{S_i, S_j} \quad (2)$$

is the total number of commuters residing in the set of BSUs of M_s that works in any of the BSUs of M_t ; $R_k = W_{\{M_k\}, S}$ the total number of workers residing in M_k ; and $J_k = W_{S, \{M_k\}}$ the total number of jobs in M_k .

Factor PE_{M_i, M_j} is the fraction of the employed population residing in M_i and working in M_j ; and PJ_{M_i, M_j} is the portion of jobs in M_j that are held by workers residing in (coming from) M_i .

This interaction index can be the base for different fitness functions. Among them in this exercise we have decided to test our method with

$$f(M) = \text{card}(M) \times \sum_{\forall S_i \in S} II(\{S_i\}, M_{S_i} - \{S_i\}) \quad (3)$$

where M_{S_i} is the market S_i belongs to. In our case we calculate the interaction index between a BSU S_k –which is considered as a mono-BSU market– and the market that would result if that BSU S_k is subtracted from the market M_k it belongs to. This interaction index between a BSU and the market it belongs to is a generalization of the interaction index used in [2]. The inclusion of the number of LLMA as a factor allows to reach the highest possible number of independent LLMA –this is one of the criteria usually applied in practical exercises [9].

Besides, each market $M_i \in M$ must fulfil two requirements in terms of minimum self-containment percentages ($\beta_1, \beta_2, 0 \leq \beta_1 \leq \beta_2 \leq 1$) –i.e. both the proportion of the occupied working locally, and the proportion of jobs filled by local workers must exceed a given threshold–, and minimum size in terms of employed population ($\beta_3, \beta_4, 1 \leq \beta_4 \leq \beta_3$):

$$\min \left(\frac{W_{M_i, M_i}}{W_{M_i, S}}, \frac{W_{M_i, M_i}}{W_{S, M_i}} \right) \geq \beta_1 \quad (4)$$

$$W_{M_i, S} \geq \beta_4 \quad (5)$$

Very urbanized environments are in real world characterised by the intensity and complexity of the network of commuting flows, something which makes it difficult to identify isolated groups of BSU. To facilitate the identification of a larger number of separate LLMA in such environments a trade-off between both constraints (self-containment and minimum size) has been introduced similarly to [2], but using the formulation proposed by Casado-Díaz[9]. According to this proposal, the minimum self-containment requirement is linearly relaxed from β_2

to β_1 for populations sizes from β_4 to β_3 . For each market in a given solution, this trade-off is evaluated as follows:

$$\frac{\min\left(\frac{W_{M_i, M_i}}{W_{M_i, S}}, \frac{W_{M_i, M_i}}{W_{S, M_i}}\right)}{\beta_2 - m\beta_4 + mW_{M_i, S}} \geq 1 \quad (6)$$

$$m = \frac{\beta_2 - \beta_1}{\beta_4 - \beta_3} \quad (7)$$

We have also included a minimum connectivity requisite to guarantee some degree of territorial contiguity without employing spatial data: a BSU can only belong to a market if it is reachable from any other BSU of that market through the γ largest outgoing/incoming commuting flows of each BSU in the market (we call this functional neighbourhood).

3 Evolutionary proposal

The structure of the initial evolutionary algorithm for the regionalization of a given territory follows the next steps:

Step 1. Produce an initial population consisting of n_p individuals. At least one of the individuals in the population must be valid -i.e. it must meet all the constraints. To assure this the first individual generated consists of a single market covering all the territory. Complete the initial population with $n_p - 1$ randomly generated individuals (in practice, all of these are invalid solutions).

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

Step 3. Repeat n_r times: select two valid individuals from the current population by fitness-proportional probability, select with uniform probability one crossover operator, apply it to the two selected parents generating a new individual, and evaluate its fitness.

Step 4. Sort the whole population, composed of $n_p + n_r$ individuals, by their fitness value.

Step 5. Repeat n_m times: select a valid individual from the current population (including the new offspring from the recombination stage) by fitness-proportional probability, select with uniform probability one mutation operator, apply it to the selected individual generating a new individual, and evaluate its fitness.

Step 6. Sort the whole population, composed of $n_p + n_r + n_m$ individuals, by their fitness value.

Step 7. Select the population for the next generation choosing the n_p best individuals (truncation scheme).

Step 8. Stop condition: if the best individual has changed in the last g generations, return to step 3. Otherwise, finish.

3.1 Genetic representation

The individuals which constitute the population represent feasible solutions, that is, the aggregation of all the BSUs composing territory S , into non over-lapping

LLMAs. We have used a group-number encoding [10] where each individual is represented by a vector of n components, each of which corresponds to a BSU of S , and takes the value of the identifier of the market the BSU belongs to (Figure 1). This representations ensures the non-overlapping constraint is fulfilled.

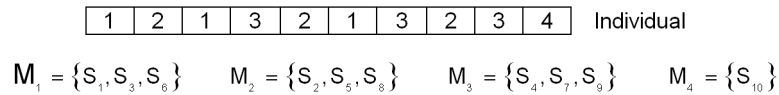


Fig. 1. Representation of individuals

3.2 Selection

The selection of the individuals to be affected by recombination and mutation operations is performed following a ranking method, according to which those individuals scoring higher in the fitness function have a larger probability of being selected. We use truncation for the selection of the survival individuals that compose the population in the next generation. So for every generation, the population is composed by the n_p better solutions in the previous generation.

3.3 Genetic operators

Due to the large number of constraints that the individuals must meet, the usual operators of recombination and mutation seldom lead to valid solutions. This makes the evolution difficult or even unable to progress. For this reason we designed four *ad hoc* crossover operators and eleven mutation operators (see [8] for a detailed description). Specialized crossover operators consider the codification of both parents when the offspring is generated avoiding discrepancies between all of them. On the other hand, mutation operators have four main functions: division of markets, fusion of markets, reassignment of single BSUs, and reassignment of groups of BSUs. The goal of division operators is to increase the number of markets –i.e. $card(M)$ – in the regionalization, so as to improve the detail of the result. Fusion operators eliminate markets to go back in the process of division. Reassignment operators try to improve the solution by reassigning specific BSUs between markets in a local search procedure.

3.4 Summary of the evolutionary process

To summarize, from an individual in which all the BSUs are merged to conform a single LLMA, successive applications of division and aggregation of markets,

reassignment of single BSUs or groups of BSUs between markets, and recombinations, allow increasing the number of LLMA, assigning the basic geographical units to the relevant LLMA so that the fitness function is maximised.

4 Memetic proposal

One of the main problems of the application of this evolutionary algorithm, caused by restrictions that must be fulfilled, is the high percentage of invalid individuals resulting from the use of stochastical operators, where little or none information about the problem is used.

To achieve a higher success rate for those operators without complicating their algorithms, we implement a repair stage after each genetic operator. In this step invalid markets are disaggregated and their BSUs reassigned –among the rest of markets that form the individual being repaired–, until all the markets in the individual meet the constraints. As a side effect, this technique makes all randomly generated individuals in the initial population to be valid, and therefore the inclusion of an specific valid individual in step 1 to start the evolution is unnecessary.

Moreover, just after its evaluation, every valid individual goes on to a local search process, by means of the reassignment of individual BSUs among the existing markets, in order to optimize the generated individuals.

4.1 Repair of invalid individuals

The process of repair of an individual is based on the technique used by Coombes et al. [2] when a market does not fulfil the established restrictions. These invalid markets are successively disintegrated into their constituent BSUs which are then reassigned to the market with which they have the larger mutual interaction.

In an algorithmic way the process of repair would be as follows:

1. Verify the fulfilment of the constraints of each market $M_i \in M$ (see 3). If there is no invalid market, repair is finished.
2. Chose at random an invalid market M_i and disintegrate it.
3. Assign each BSU $S_k \in M_i$ (chosen in random order) to the market M' with which it possesses more interaction. That is:

$$M' = \underset{\forall M_j \in M, M_j \neq M_i}{\arg \max} II(\{S_k\}, M_j) \quad (8)$$

4. Once all the BSUs in M_i are reassigned, return to step 1.

As a result of this process, the individual will fulfil the constraints (unless there is no valid solutions in the search space due to excessive restriction), but the total number of LLMA will decrease at least in one (if the individual was invalid). This worsens the fitness, since the specific function that we use in this exercise (3) tends to overvalue the solutions with larger number of markets (there is a preference for detail). Note that the final number of markets disaggregated

in a repaired individual can be lower than the number of invalid markets before repair, because the reassigned BSUs of the first disintegrated markets can turn some other invalid markets into valid ones before the repair process choose them for disaggregation.

4.2 Improvement of valid individuals

We have also included a process of improvement of the generated individuals that fulfil all the restrictions or those that have been repaired. The goal is to carry out a local search to improve the fitness of the individual by reassigning single BSUs between markets. The process starts by randomly selecting a market (LLMA) to be optimised. Next, the BSU belonging to that market and having the lowest interaction with the rest of its constituent BSUs, is reassigned to the market with which it has more interaction. This process continues while it increases the value of the fitness function. Likewise, we have also included a tolerance parameter that allows a number of unsuccessful reassignments before stopping the optimization process.

The process of optimization is as follows:

1. Set counter of unsuccessful attempts to $c = 0$.
2. A market $M_i \in M$ of the individual I is selected at random.
3. The BSU to be removed is selected as:

$$S_r = \arg \min_{\forall S_j \in M_i} II(\{S_j\}, M_i - \{S_j\}) \quad (9)$$

4. In a new individual I' , copy of I , BSU S_r is assigned to its optimal market following equation (8).
5. If the new individual I' is invalid, it is repaired.
6. If $f(I') > f(I)$ (the new individual is better than the original), $I = I'$. Else increment the counter of failed attempts, $c = c + 1$
7. If $c < \xi$ return to step 2.

4.3 Results

To test our proposal we use a case study: the delineation of a set of LLMA's in the Region of Valencia, Spain. Travel-to-work data derived from the Spanish Census of Population [11] allowed us to build a 541×541 origin-destination commuting matrix (where 541 is the number of municipalities that integrate the territory), where each cell represents W_{S_i, S_j} . Parameters were set in these values: size population $n_p = 100$, recombination offspring $n_r = 10$ and mutations $n_m = 24$. The condition of termination, i.e. generations without changes in the best individual is set to $g = 250$. Parameter γ of minimum flow connectivity or functional neighbourhood is set to 5. In the memetic algorithm, parameter ξ of allowed failed attempts in the optimization process is set to 5.

These evolutionary proposals substantially improve the results obtained by the traditional methods in both number of markets and fitness function (Table 1). The memetic algorithm obtains better mean results than the original

evolutionary proposal and their dispersion is smaller. In an ANOVA test with confidence level 5%, we have obtained an F value equal to 16.04, greater than the critical value 3.86. So, the improvement of the memetic proposal is significant, even circumstantially the best result was obtained with the original EA. However, although the solutions are obtained in less iterations, evolution time is around four times higher due to the great percentage of individuals that must be evaluated (Table 2) in the course of the reparation and optimization steps.

Table 1. Comparison of results

		Fitness value	Number of LLMA's	Generations	Time consumed (s.)
TTWAs method		120.23	44	-	1
Original EA [7]	Best	190.01	62	3620	1068
	Mean	180.54	59.52	3088.1	853.43
	σ	4.19	1.24	588.09	166.76
Memetic algorithm	Best	189.64	62	2580	4478
	Mean	182.04	59.43	2074.7	3332.58
	σ	3.21	1.14	566.84	894.44

Table 2. Percentage of individuals repaired and improved

	Percentage
Repaired by failing to fulfil contiguity constraint	53.44%
Repaired by failing to fulfil eq. (4) to (6)	35.50%
Improved individuals	7.58%

One of the consequences of the application of this processes of repair and optimization is that the success rate of the different operators –measured in terms of individuals generated by that individual that remain in the population at the beginning of each generatio– differs from the original algorithm (Figure 2). For instance, in the evolutionary proposal division operators were successful in the beginning of the evolution. Once the number of markets reached its maximum, division of markets led to invalid individuals because they did not fulfil the size constraint. However, with the memetic approach this markets can be repaired and become successful.

5 Conclusions and current works

We have presented a memetic version of our previous evolutionary algorithm for delimitation of functional areas (an unsupervised multi-constrained graph partitioning problem), and compared both results. Given the complexity of the

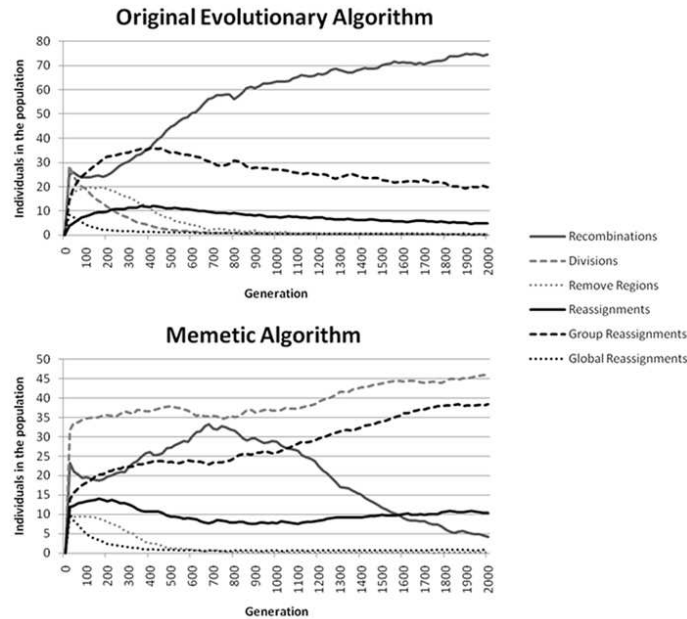


Fig. 2. Number of individuals in the population generated with the different operators

problem, when the requirements associated to such a procedure are applied in real case studies (where the number of base spatial units is frequently very high), conventional genetic operators hardly ever lead to valid solutions. We tried to avoid this problem by designing ad-hoc operators. These specialized operators allowed to obtain good final delineations. However, the percentage of invalid generated individuals continued being very high. Our memetic extension of the EA includes a repair process to turn these solutions into valid ones that can be evaluated and contribute to the evolutionary process. Both techniques result in a faster evolutionary process, in terms of generations. Besides, some of the operators that in previous versions were of low usefulness in the course of the evolution, are now much more significant in the whole process. The time consumed by the whole memetic approach is however comparatively very high. The repair and optimization procedures, and the numerous extra evaluations associated to them, makes the process longer. And opposite to our first estimations, the improvement in quality of the solutions is small, although statistically significant. This can be consequence of the high value that the objective function chosen in this exercise assigns to the number of delimited markets, since the repair process tends to reduce that value. Moreover the optimization process uses itself the repair process in its iterations (so it is time consuming) and it is based in one of the mutation operators already working in the mutation stage (so its effectiveness would be improved if that mutation operator is disabled). So now we are

studying how this MA performs with other fitness functions and representations, and how we could optimize the repair and optimization processes.

We are also considering other ways of improvement, like the use of an adaptive scheme for the probability of application of the genetic operators in order to take advantage of the changing efficiency of the operators during the evolution, an island model for the parallel implementation of the algorithm, a reformulation of the problem based on multiobjective optimization, and the application of the method to other GPPs.

Acknowledgments. José M. Casado-Díaz has received financial support from the Spanish Ministry of Transports (Ministerio de Fomento) for the project "New methods for the analysis of the determinants and effects of commuting flows (MOVICO)" under Grant T 75/2006 (Spanish National Plan of A+D+i). Lucas Martínez-Bernabeu acknowledges financial support from the Spanish Ministry of Education and Science, the European Social Fund (ESF) and the University of Alicante. Raúl Gómez-Hernández acknowledges financial support from the University of Alicante.

References

1. EUROSTAT, M. G. Coombes: Étude sur les zones d'emploi, Document E/LOC/20, Office for Official Publications of the European Communities, Luxembourg (1992)
2. M. G. Coombes, A. E. Green, S. 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)
3. ISTAT, I sistemi locali del lavoro 2001. Available: <http://www.istat.it>(2005)
4. J. M. Casado-Díaz, Local Labour Market Areas in Spain: A Case Study, *Regional Studies*, 34(9), 843–856 (2000)
5. K. Papps, J. O. Newell, Identifying Functional Labour Market Areas in New Zealand: A Reconnaissance Study Using Travel-to-Work Data, Institute for the Study of Labor (IZA), Bonn, Discussion Paper n. 443 (2002)
6. M. Watts, Local Labour Markets in New South Wales: Fact or Fiction?, Centre of Full Employment and Equity, The University of Newcastle, Australia, WP 04-12 (2004)
7. F. Flórez-Revuelta, J.M. Casado-Díaz, L. Martínez-Bernabeu: An Evolutionary Approach to the Delineation of Functional Areas Based on Travel-to-work Flows. *International Journal of Automation and Computing*, 05(1), 10–21 (2008)
8. F. Flórez-Revuelta, J.M. Casado-Díaz, L. Martínez-Bernabeu: Specific crossover and mutation operators for a grouping problem based on interaction data in a regional science context. In: *IEEE Congress on Evolutionary Computation*, Singapore, pp. 378–385. IEEE (2007)
9. J. M. Casado-Díaz, M. G. Coombes. The Delineation of 21st Century Local Labour Market Areas (LLMAs). In *Proceedings of the 8th Nectar Conference*, Las Palmas de Gran Canaria, Spain, 2005, unpublished.
10. D. A. Jones, M. A. Beltramo: Solving partitioning problems with genetic algorithms, In: *Proceedings of the Fourth International Conference on Genetic Algorithms*, Morgan Kaufmann Publishers, pp. 442–449 (1991)
11. Instituto Nacional de Estadística. Censo de población, <http://www.ine.es> (2001)