



Universitat d'Alacant  
Universidad de Alicante

Optimización de áreas funcionales espaciales  
Mediante algoritmos evolutivos multioperador

Aplicación a la delimitación  
de mercados locales de trabajo

Lucas Martínez Bernabeu



Tesis

**Doctorales**

[www.eltallerdigital.com](http://www.eltallerdigital.com)

UNIVERSIDAD de ALICANTE

Optimización de áreas funcionales espaciales  
mediante algoritmos evolutivos multioperador

Aplicación a la delimitación  
de mercados locales de trabajo

Lucas Martínez Bernabeu

Universitat d'Alacant  
Universidad de Alicante

Junio 2012



TESIS DOCTORAL

OPTIMIZACIÓN DE ÁREAS FUNCIONALES  
ESPACIALES MEDIANTE ALGORITMOS  
EVOLUTIVOS MULTIOPERADOR.  
APLICACIÓN A LA DELIMITACIÓN DE  
MERCADOS LOCALES DE TRABAJO.

Presentada por:

LUCAS MARTÍNEZ BERNABEU

Dirigida por:

DR. FRANCISCO FLÓREZ REVUELTA

DR. JOSÉ MANUEL CASADO DÍAZ

Departamento de Tecnología Informática y Computación

JUNIO 2012



A mis padres, a mi hermano y a Cristina.

Si no fuerais mi familia os adoptaría, nada tiene sentido sin vosotros.



Universitat d'Alacant  
Universidad de Alicante



Universitat d'Alacant  
Universidad de Alicante

# Agradecimientos

Esta tesis ha sido posible gracias a los fondos recibidos desde la Universidad de Alicante en modalidad de beca y del Plan Nacional de I+D+i en varios proyectos de investigación coordinados: desde el Ministerio de Educación y Ciencia a través del Programa de Personal Técnico de Apoyo en la modalidad de Proyectos de I+D (ref. solicitud PTA-2003-02-00178, 495) para el proyecto “Movilidad cotidiana y mercados de vivienda y trabajo en las áreas urbanas españolas” (ref. BEC2003-02391), que recibió fondos procedentes de los programas FEDER y FSE de la UE; desde el Ministerio de Fomento a través del proyecto “Movilidad cotidiana. Nuevos métodos para el análisis de sus determinantes y efectos (MOVICO)” (ref. T 75/2006); y desde el Ministerio de Ciencia e Innovación y del programa FEDER de la UE a través de los proyectos “Mercados Locales de Trabajo. Nuevos métodos de delineación y análisis (MoviTra-LAB)” (ref. SEJ2007-67767-C04-02) y “La movilidad en entornos urbanos: nuevos desarrollos en la delimitación de áreas funcionales y el análisis de datos individuales y agregados (MOVE-NUR)” (ref. CSO2011-29943-C03-02).

Por supuesto, en un plano diferente, ha sido decisivo el apoyo humano y la guía que he recibido a lo largo de este trabajo, y también antes.

La ciencia siempre me ha atraído, desde la infancia. Hasta los cinco años quise ser mago, luego me di cuenta de que la verdadera magia está en la ciencia, y la ingeniería. Si Paco no me hubiera invitado a sus clases de doctorado cuando yo aún no había terminado la carrera es posible que nunca hubiera leído un artículo de investigación.

Ese no fue el único hito decisivo que me ha conducido hasta este momento, en que estoy escribiendo los agradecimientos en la memoria de una

tesis de doctorado. La semilla de esta investigación fue la de la amistad y colaboración entre mis tutores, los doctores Casado y Flórez. Todo el trabajo que conforma esta tesis parte de ahí, y me siento orgulloso de formar parte de esa colaboración y de esa amistad. Y no solo pusieron la semilla, son coautores de los artículos que la forman.

A todos los profesores del programa de doctorado de Tecnologías de la Sociedad y la Información, en especial a Juanma. Me enseñaron en qué consiste la labor del investigador y me señalaron el camino para llegar a ser uno de ellos.

También agradezco el apoyo y consejos de mis profesores y compañeros del Departamento de Tecnología Informática y Computación y de la Escuela Politécnica Superior en general. El hecho de estar radicados en dependencias alejadas no me ha permitido mantener un estrecho contacto, pero siempre han estado ahí cuando los he necesitado. Debo destacar a Jorge y Pablo, fueron grandes apoyo en nuestras espontáneas conversaciones.

A los miembros de EU/ME y demás colegas del campo de las metaheurísticas, que en diferentes congresos y reuniones han escuchado mis dudas y problemas y me han sabido dar grandes consejos y una fuerte motivación: muchas gracias, me habéis acompañado en este mundo. J.J., Carlos C., Ferguson, Juanlu, Bernabé, Marc S., Kenneth S. y tantos otros, espero que nos volvamos a encontrar pronto en otros eventos para seguir compartiendo conocimiento, risas y canciones.

Merecen también una mención especial los demás miembros del campo de las ciencias regionales con los que he colaborado en los proyectos coordinados que han motivado esta investigación. Su esfuerzo por comprender mis discursos en jerga de computación evolutiva y optimización me ha permitido aprender a explicarme.

A mi familia y mi pareja, que me han soportado y animado constantemente a seguir, o a descansar, durante todos estos años. Mis padres y mi hermano han sido el lugar de retiro donde recuperar fuerzas. Por último, Cristina, no hay palabras para expresar mi agradecimiento por todas las horas en que escuchaste pacientemente mis disquisiciones y angustias y asustaste mis fantasmas, y por tus valiosísimas correcciones. Por mi culpa tú también eres una experta en mercados locales y algoritmos evolutivos. Gracias a ti yo soy mejor humano.

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. <b>Ámbito y motivación</b>	<b>1</b>
1.2. <b>Áreas funcionales espaciales</b>	<b>4</b>
1.2.1. Problema de la unidad de área modificable	6
1.2.2. Mercados locales de trabajo	8
1.2.2.1. Aspectos fundamentales en la delimitación de MLT	10
1.3. <b>Métodos de delimitación de MLT</b>	<b>12</b>
1.3.1. Método de las TTWA o ERA	13
1.3.2. Método Intramax	18
1.3.3. Método de las CZ empleado en EE.UU.	19
1.4. <b>Problemas de agrupación basada en interacción</b>	<b>22</b>
1.4.1. Modularización de <i>software</i>	24
1.4.2. Manufactura celular	27
1.4.3. Planificación de áreas de localización	28
1.4.4. Detección de comunidades	29
1.5. <b>AE en problemas de agrupación</b>	<b>31</b>
1.5.1. Algoritmos evolutivos de agrupación	32
1.5.2. Operadores informados y algoritmos híbridos	36
1.6. <b>Propuesta de solución</b>	<b>37</b>
1.6.1. Formulación como problema de optimización	39
1.6.2. Función objetivo para la delimitación de MLT	40
1.6.2.1. Índice de interacción entre dos áreas	42
1.6.2.2. Funciones objetivo propuestas	44
1.6.3. Restricciones	44
1.6.3.1. Tamaño y autonomía mínimos	44
1.6.3.2. Vecindad funcional y contigüidad geográfica	46
1.6.4. Representación de las soluciones	47

<b>2. Publicaciones</b>	<b>51</b>
2.1. PPSN 2006	52
2.1.1. Estructura del algoritmo evolutivo	53
2.1.2. Operadores genéticos ad hoc	53
2.1.3. Experimentación, resultados y conclusiones	55
2.2. CEC 2007	56
2.2.1. Nuevos operadores	56
2.2.2. Comportamiento de los operadores	57
2.2.3. Experimentación, resultados y conclusiones	57
2.3. IJAC 2008	58
2.3.1. Comportamiento con otras funciones y parámetros	58
2.3.2. AEA como optimizador local	59
2.3.3. Análisis de estabilidad	59
2.3.4. Experimentación, resultados y conclusiones	59
2.4. PPSN 2008	60
2.4.1. Algoritmo memético	60
2.4.2. Experimentación, resultados y conclusiones	61
2.5. CEC 2009	62
2.5.1. Propuestas adaptativas y autoadaptativa	63
2.5.2. Experimentación, resultados y conclusiones	64
2.6. Thomson-Civitas 2010	65
2.6.1. Resolución de discontinuidades	66
2.6.2. Evaluación de calidad	66
2.6.3. Experimentación, resultados y conclusiones	67
2.7. ESWA 2012	68
2.7.1. Aglomeración jerárquica estocástica	70
2.7.2. Operadores genéticos de agrupación	70
2.7.2.1. Operador de cruce basado en grupos	76
2.7.3. Resultados y comparación con propuestas anteriores	77
2.7.3.1. Comunidad Valenciana	77
2.7.3.2. Resolviendo problemas grandes: España	78
2.7.4. Conclusiones	79
<b>3. Conclusiones</b>	<b>81</b>
3.1. Discusión	81
3.2. Objetivos de trabajo futuros	85

<b>A. An evolutive approach for the delineation of local labour markets (PPSN 2006)</b>	<b>101</b>
<b>B. Specific crossover and mutation operators for a grouping problem based on interaction data in a regional science context (CEC 2007)</b>	<b>113</b>
<b>C. A Memetic Algorithm for the Delineation of Local Labour Markets (PPSN 2008)</b>	<b>123</b>
<b>D. An evolutionary approach to the delineation of functional areas based on travel-to-work flows (IJAC 2008)</b>	<b>135</b>
<b>E. Adaptive Evolutionary Algorithms for the Delineation of Local Labour Markets (CEC 2009)</b>	<b>149</b>
<b>F. Los mercados locales de trabajo españoles. Una aplicación del nuevo procedimiento británico (Thomson-Civitas 2010)</b>	<b>159</b>
<b>G. Grouping genetic operators for the delineation of functional areas based on spatial interaction (ESWA 2012)</b>	<b>201</b>

Universitat d'Alacant  
Universidad de Alicante



# Índice de figuras

1.1. Diagrama del método de las TTWA (COOMBES <i>et al.</i> , 1986) .	16
1.2. Representación de individuos (cromosoma) . . . . .	48
2.1. Pseudocódigo del AE (PPSN 2006) . . . . .	53
2.2. Pseudocódigo del AE (PPSN 2006) . . . . .	61



Universitat d'Alacant  
Universidad de Alicante



Universitat d'Alacant  
Universidad de Alicante

# Índice de ecuaciones

1.1.	Índice de interacción proporcional de Coombes <i>et al.</i> . . . . .	17
1.2.	Índice de interacción absoluta de M&B . . . . .	18
1.3.	Índice de interacción proporcional de M&B . . . . .	19
1.4.	Flujos proporcionales entre UT de T&S . . . . .	20
1.5.	Disimilitud entre UT de T&S . . . . .	20
1.6.	Disimilitud entre grupos ( <i>average linkage</i> ) de T&S . . . . .	20
1.7.	Partición de un conjunto, $\mathcal{P}$ . . . . .	22
1.8.	Medida de cohesión y acoplamiento, EVM . . . . .	25
1.9.	Calidad de un <i>cluster</i> en EVM . . . . .	26
1.11.	Calidad de modularización MQ . . . . .	26
1.12.	Calidad de un <i>cluster</i> en MQ . . . . .	26
1.13.	Flujo absoluto entre dos grupos, $W(A, B)$ . . . . .	39
1.14.	Índice de interacción proporcional entre dos grupos, $\Pi_{AB}$ . . . . .	43
1.15.	. . . . .	43
1.16.	Interacción acumulada de cada UT en su mercado por el número de mercados (FO), $f_0(\mathcal{P})$ . . . . .	44
1.17.	Interacción acumulada de cada UT en su mercado por el número de mercados (FO), $f_1(\mathcal{P})$ . . . . .	44
1.18.	Restricción de tamaño mínimo . . . . .	45
1.19.	Restricción de autonomía mínima . . . . .	45
1.20.	Balance entre tamaño y autonomía . . . . .	45



Universitat d'Alacant  
Universidad de Alicante

# Capítulo 1

## Introducción

### 1.1. Ámbito y motivación

En esta tesis se considera la utilización de un algoritmo evolutivo de agrupación (AEA) para la delimitación basada en flujos de interacción de mercados locales de trabajo (MLT), un tipo de área funcional (AF) de interés en análisis económico regional y urbano, así como en el diseño, puesta en marcha y evaluación de políticas de empleo, de desarrollo y de promoción del comercio y la industria entre otros usos (OCDE, 2002; CASADO-DÍAZ y COOMBES, 2005; MA y BANISTER, 2006).

Se trata de ámbitos territoriales en los que determinado grupo de oferentes y demandantes de empleo entran en contacto entre sí. La razón de ser de las AF en este ámbito es que las unidades de carácter administrativo en que se divide el territorio no responden apropiadamente a los fenómenos que se pretende analizar, lo que exige el desarrollo de sistemas territoriales de referencia alternativos (OPENSHAW, 1984). Sin embargo, reorganizar las estructuras de gobierno local de acuerdo a una particular división del territorio diferente de la división administrativa puede requerir un gran esfuerzo. Es por ello que una delimitación funcional ha de presentar claros beneficios sobre la división administrativa para hacerla realmente adecuada para la creación de políticas (CÖRVERS *et al.*, 2009).

Los MLT comparten gran similitud con otras AF basadas en matrices o grafos de interacción —regiones, mercados, comunidades, módulos, celdas o grupos compuestos por elementos densamente relacionados entre sí y escasamente relacionados con elementos de otros grupos—, como son los módulos *software*, las celdas de manufactura, las áreas de localización y las comunidades (sociales y biológicas). Estas AF son de sumo interés en muy diversas técnicas y disciplinas de la ciencia; podemos mencionar ingeniería del *software*, ingeniería del *hardware*, biología (especialmente genómica y proteómica), química, neurociencias, sociología e ingeniería industrial, entre otras.

La delimitación de este tipo de AF forma parte de una clase de problemas de elevada complejidad, al igual que muchos problemas de *clustering* catalogados como NP-difícil o NP-completo (GAREY y JOHNSON, 1979), tal y como se mostrará más adelante.

Aunque la delimitación de algunos de esos tipos de AF ha recibido o está recibiendo considerable atención por parte de la comunidad de las ciencias de la computación, en cuyo ámbito se ha aplicado una gran variedad de técnicas de búsqueda y optimización, el caso de los MLT y otras AF de naturaleza económica o geográfica es un área de estudio en la que únicamente se aplican técnicas voraces de *clustering* —con larga tradición en la investigación operativa—. Al igual que otros métodos de agregación voraces, estos procedimientos permiten llegar a soluciones adecuadas en un corto plazo de tiempo, pero no se utiliza un enfoque de optimización (de hecho, este tipo de procedimientos tienen como su condición de terminación encontrar la primera solución factible). En consecuencia, el resultado de tales procedimientos (el conjunto de MLT delimitados) es sólo una entre muchas otras posibles soluciones, de las que se puede esperar que estén muy lejos de ser la óptima. Además, esos métodos están sujetos a parámetros subjetivos, que no permiten una clara comparación entre casos de estudio en diferentes territorios.

Se pretende por tanto avanzar, con un enfoque multidisciplinar, en la delimitación de MLT mediante el estudio de las técnicas empleadas en otras disciplinas, que se revisan en la sección 1.4, y la presentación de un algoritmo evolutivo de agrupación —generalización de los algoritmos genéticos de

---

agrupación, técnica empleada en otros problemas de delimitación de AF y que se revisará en la sección 1.5—, capaz de resolver casos reales con altos estándares de calidad, además de continuar el estudio de funciones objetivo adecuadas para la evaluación y delimitación de MLT y otras AF afines.

El desarrollo de este trabajo ha sido posible gracias al apoyo de diferentes proyectos financiados por el Plan Nacional de I+D+i (incluidos en los agradecimientos), y ha quedado reflejado en una serie de ponencias en congresos de prestigio (publicadas luego como capítulo de libro en las correspondientes actas), artículos en revistas internacionales y un capítulo de libro. En este documento se incluyen esas publicaciones tal y como se pueden obtener electrónicamente, precedidas de una síntesis en la que se explican los conceptos previos para la mejor comprensión del problema abordado, se plantean los objetivos e hipótesis de partida de esta investigación, y se compendian los resultados y conclusiones presentados en las distintas publicaciones, así como las aportaciones realizadas por el conjunto de este trabajo y las líneas de investigación que abre. A continuación se indica la organización de este documento.

En el resto de este capítulo, la sección 1.2 describe el concepto de MLT dentro del más amplio de las AF (lo que permitirá relacionarlo con los de otras disciplinas en las que se han resuelto problemas similares) y la complejidad inherente de su delimitación. En la sección 1.3 se revisan los métodos actualmente empleados para realizar las delimitaciones de MLT en varios países. En la sección 1.4 se revisan problemas similares en otras cuatro disciplinas y los métodos empleados para resolverlos, que han recibido atención desde diferentes ramas de la ciencia de la computación desde hace más tiempo que el de la delimitación de MLT, y cuyos resultados indican la idoneidad de aplicar alguna de las técnicas más recientes, como es el caso de los algoritmos evolutivos de agrupación, que se revisan en la sección 1.5 junto al concepto más amplio de computación evolutiva y la alternativa de introducir conocimiento en los operadores. La sección 1.6 cierra el capítulo con un resumen de las partes comunes a las diferentes publicaciones que compendia esta tesis.

En el capítulo 2 se incluye una sección por cada una de las publicaciones que se compendian en esta tesis, que incluyen un resumen del planteamiento

realizado y de los resultados y conclusiones obtenidos. Por último, el capítulo 3 presenta la síntesis de las conclusiones finales de este trabajo y resalta las líneas de investigación futura más prometedoras. En los anexos al final del documento pueden consultarse las publicaciones tal y como están disponibles para su descarga en formato electrónico.

## 1.2. Áreas funcionales espaciales

En una definición muy general, un área funcional es un conjunto de elementos altamente relacionados entre sí y poco relacionados con elementos de otras áreas. Aunque existe una relación directa, el término no ha de confundirse con el que comúnmente se utiliza en el ámbito empresarial, donde cada departamento o sección comprende los procedimientos y sistemas relativos a una determinada área de negocio (como producción o *marketing*), de forma que resulta más fácil y eficiente la gestión de los recursos de la empresa.

En el ámbito de la ciencia regional y urbana, un área económica funcional (también llamada área de mercado o región funcional, entre otros) es un espacio geográfico en que opera un determinado mercado económico (ya sea de vivienda, trabajo, ocio, transporte, comunicaciones, servicios públicos, mercancías o una combinación de éstos u otros) (COMMUNITIES y GOVERNMENT, 2010). En otras palabras, es el área en que se encuentran y comercian entre sí un conjunto de oferentes y demandantes (agentes) de un determinado producto (o conjunto de productos).

Casos particulares de áreas funcionales o de mercado utilizados en diversos países para propósitos oficiales, de negocios o académicos son: las *economic areas* (JOHNSON y KORT, 2004) en EE.UU. definidas por la Oficina de Análisis Económico (BEA en sus siglas en inglés), los mercados de vivienda o *housing market areas* (BROWN y HINCKS, 2008; WELSH ASSEMBLY GOVERNMENT, 2006), las áreas de transporte (de pasajeros y/o carga) (KRYGSMAN *et al.*, 2009; LIMBOURG y JOURQUIN, 2010) y las áreas de comercio (FLÖRKEMEIER, 2002).

Todas estas AF se caracterizan por basar su delimitación en datos de interacción espacial (ROY y THILL, 2004), es decir, variables de escala ordinal, de intervalo o de razón, que caracterizan o miden cierta forma de

relación entre pares de unidades espaciales básicas o unidades territoriales (UT). Ejemplos de dichas interacciones serían el flujo de trabajadores que viajan desde su residencia a su lugar de trabajo diariamente (movilidad laboral cotidiana o *commuting*), los desplazamientos cotidianos por ocio, los flujos migratorios y los flujos de bienes, servicios e información entre sus productores y sus consumidores. Las UT serían, a su nivel de detalle más bajo posible, puntos georreferenciados (viviendas, negocios, etc.), aunque en la mayor parte de los casos se trata de agregaciones de éstos: distritos postales, áreas censales, barrios, municipios, comarcas, condados, etc.

Nótese que no estamos considerando un problema de *clustering* por atributos: las áreas no tienen un elemento medio o mediana ni se agrupan por homogeneidad de sus características. No existe un vector de atributos que permita medir una función de distancia o similitud entre dos grupos de áreas. En lugar de eso, los elementos se asocian entre sí por el grado de interacción que comparten, que se puede entender como la fuerza de atracción que los une. Esta interacción puede considerarse como una forma de similitud, y como una forma de repulsión si la invertimos, pero no tiene las propiedades para ser una distancia o pseudodistancia: no es simétrica si se considera la atracción relativa al tamaño de cada punto; la distancia de un punto a sí mismo puede no ser cero (normalmente no lo será, salvo que así se fije y se descarte esa información) y no cumple la desigualdad triangular. Además, la similitud o distancia sólo serían inequívocas entre pares de UT aisladas: el valor del flujo entre ambas. Al medir la relación entre agregaciones de UT (áreas) se hace necesario escoger una forma de medirla: el flujo agregado entre ambas áreas, la media de los flujos entre cada par de UT en distinta área, el flujo simple (entre dos UT) máximo entre ambas áreas o también valores calculados a partir de los anteriores. La medida ha de considerar los tamaños relativos de cada UT. Por lo tanto, métodos de *clustering* bien conocidos en las ciencias de la computación como el *K-means* o el *K-medoids* no son directamente aplicables, ya que no se dispone de un elemento promedio o central que caracterice a cada grupo de elementos.

En la definición ideal de mercado, todo proveedor debe ser capaz de comerciar con cualquier demandante en su mercado y viceversa. Además, no debe haber interacción con agentes situados en áreas de mercado diferentes, para reducir las correlaciones entre regiones separadas en los análisis que se

apliquen. En la práctica, un AF se caracteriza por un alto nivel de cohesión entre sus integrantes y un alto grado de autonomía con respecto a otras áreas. Como resultará obvio tras analizar el problema, ambas condiciones se limitan entre sí en la mayoría de problemas reales: p. ej., la mayoría de los países son muy autónomos en términos de mercados de trabajo o vivienda (pocos ciudadanos trabajan en el extranjero o se comprarán allí una casa), pero un país de tamaño medio difícilmente puede ser considerado como un mercado único (pocas personas trabajan o compran en lugares muy alejados de su residencia, o consideran mudarse a regiones distantes); y viceversa, cuanto más pequeña es una región, más probable es que el grado de relación con alguna región vecina sea alto.

La delimitación de AF es muy relevante en distintos contextos. Es crucial para los analistas de fenómenos sociales y económicos en los niveles urbano y regional, y son utilizados por las autoridades para definir, implementar y supervisar políticas de diversa índole. Los ejemplos incluyen la planificación del suelo dedicado a la vivienda o las diferentes actividades económicas en un determinado lugar, la elección de políticas de empleo más adecuadas para cada región, o el establecimiento de prioridades de inversión y la predicción de necesidades de desarrollo de los pasajeros y el transporte de mercancías. La calidad de los resultados de esta amplia gama de usos, y de los análisis que se derivan de ellos, depende de la precisión en la delimitación de los límites de estas áreas de mercado (COMMUNITIES y GOVERNMENT, 2010).

### 1.2.1. Problema de la unidad de área modificable

No existe un conjunto de reglas fijo para la agregación de áreas, ni estándar, ni convención internacional que guíe el proceso de agregación espacial. Como consecuencia, “las unidades de área usadas en muchos estudios geográficos son arbitrarias, modificables y sujetas a los caprichos y ocurrencias de quien esté haciendo, o hizo, la agregación” (OPENSHAW, 1984, p. 1)<sup>1</sup>. Además, muchos factores locales y regionales influyen en el desempeño de

---

<sup>1</sup>Original: *the areal units (zonal objects) used in many geographical studies are arbitrary, modifiable, and subject to the whims and fancies of whoever is doing, or did, the aggregating* (OPENSHAW, 1984, p. 1).

esas áreas y “los resultados del análisis espacial siempre dependen de las unidades de área que están siendo estudiadas” (OPENSHAW, 1984, p. 37)<sup>2</sup>.

A este problema se le denominó problema de la unidad de área modificable (*Modifiable Areal Unit Problem*, MAUP en adelante), y fue descrito en detalle por Openshaw, y descompuesto en dos subproblemas más concretos: el problema de escala y el problema de agregación. El problema de escala se refiere a la elección de un número de áreas adecuado, mientras que el problema de agregación se refiere a la elección de una regionalización adecuada dado un número concreto de áreas (BAUMANN *et al.*, 1983).

El MAUP es una forma de la falacia ecológica, en la que se hace una interpretación errónea de los datos estadísticos al inferir la naturaleza de los individuos a partir de las estadísticas agregadas del grupo al que dichos individuos pertenecen. La manipulación de los límites de los distritos electorales para dar (o quitar) ventaja a un partido político o candidato mediante la concentración o dispersión de grupos concretos de electores, conocido en inglés como *gerrymandering*, es posible debido al MAUP.

Openshaw propuso en su estudio:

En lugar de intentar ajustar el modelo a un sistema arbitrario de áreas, es necesario considerar qué sistema de áreas permite los mejores resultados y considerar qué propiedades debería tener el sistema o los datos agregados que produzca. Las regionalizaciones producidas por sistemas de agrupación óptimos para un determinado propósito pueden contribuir por sí mismas al proceso de análisis espacial (OPENSHAW, 1984, p. 37)<sup>3</sup>

En lugar de considerar en el análisis espacial los errores estadísticos que pueden estar introduciendo las delimitaciones y tener que aceptar las limitaciones que eso impone en el estudio, se ha de buscar la mejor delimitación posible considerando todas las características (restricciones) que han

---

<sup>2</sup>Original: *the results of spatial study will always depend on the areal units that are being studied* (OPENSHAW, 1984, p. 37).

<sup>3</sup>Original: *Instead of trying to fit a model to an arbitrary zoning system, it is necessary to consider which zoning systems provide the best results and to consider what properties they, or the aggregated data they produce, should have. The map patterns produced by optimal zoning systems for particular purposes may themselves contribute to the spatial analysis process* (OPENSHAW, 1984, p. 37)

de tener los resultados, para minimizar así el error de análisis, algo de vital importancia en los estudios comparativos que tratan diferentes regiones o momentos del tiempo.

En este punto radica la importancia de emplear técnicas de búsqueda que permitan encontrar la mejor solución posible al problema de la delimitación de AF (que en adelante denominaremos PDAF).

### 1.2.2. Mercados locales de trabajo

La familia más grande de áreas funcionales de carácter geográfico está constituida por los diferentes enfoques del concepto de mercado local de trabajo (OCDE, 2002) (en adelante MLT, *local labor market area* en inglés), como son las *travel-to-work areas* en Reino Unido (en adelante TTWA), las *zones d'emploi* francesas, el *sistemi locali del lavoro* italiano o las *commuting zones* y las *economic areas* en EE.UU., entre otros (véase una revisión de estos procedimientos y de su uso a nivel internacional en OCDE (2002) y en CASADO-DÍAZ y COOMBES (2005), y una actualización de este último trabajo en CASADO-DÍAZ y COOMBES (2011)). Este tipo de AF tienen uso en el análisis socioeconómico, la planificación residencial y de mercado, y la creación de políticas dirigidas a mejorar el acceso al empleo y la vivienda o a reducir el exceso de movilidad cotidiana, entre otros. Como se señaló previamente, el éxito de estas políticas cruciales depende ante todo de la calidad de la delimitación de AF empleada.

En estas regionalizaciones la interacción espacial se calcula sobre los datos de movilidad (laboral) cotidiana: los viajes diarios de los trabajadores entre su lugar de residencia y de trabajo.

Diferentes taxonomías se han empleado para categorizar los diversos métodos empleados atendiendo a varios criterios (COOMBES, 2000; CASADO-DÍAZ y COOMBES, 2005), que sintetizamos como sigue:

1. Si son o no jerárquicos: es decir, si las agrupaciones realizadas en un momento dado del procedimiento han de mantenerse unidas en las fases posteriores o pueden ser disgregadas/recombinadas.

2. Si emplean técnicas de análisis *cluster* (*clustering analysis*), que maximizan/minimizan un cierto criterio estadístico hasta alcanzar cierto umbral, o si en vez de eso son los llamados basados en reglas, que aplican sucesivamente una serie de criterios diversos y/o cambiantes.

Los métodos de análisis *cluster*, como el denominado Intramax (MASSER y BROWN, 1975), son en su mayoría jerárquicos. Y viceversa, los métodos basados en reglas, como el TTWA (COOMBES *et al.*, 1986), suelen permitir la modificación de agupaciones realizadas en etapas o pasos anteriores del procedimiento, lo que es visto como una ventaja:

Los procedimientos jerárquicos son inherentemente propensos a proporcionar resultados subóptimos [en problemas de] mayor escala, por la simple razón de que las agrupaciones efectuadas en las etapas iniciales del análisis restringen severamente las opciones disponibles en las etapas posteriores (COOMBES, 2000)<sup>4</sup>.

Los partidarios de métodos de análisis *cluster* resaltan la neutralidad y objetividad de esos métodos, y señalan como subjetivos o arbitrarios los criterios y parámetros empleados en los métodos basados en reglas (CÖRVERS *et al.*, 2009). Los defensores de los métodos basados en reglas aducen que su aproximación es más flexible y adaptable a las características concretas de las unidades básicas de territorio y del tipo de área que se esté delimitando, permitiendo mayor control sobre las características estadísticas de las áreas producidas (COOMBES *et al.*, 1986). Además sostienen que emplean medidas más fácilmente reconocibles e intuitivas —como el tamaño del flujo que une dos grupos o su importancia relativa—, mientras que los métodos de *clustering* usualmente se basan en medidas abstractas de la relativa similitud en las propiedades estadísticas globales de las áreas, como las de una matriz de afinidad (COOMBES, 2000). Las revisiones de estos métodos realizadas por algunos autores resuelven que no existe una ventaja clara por usar uno frente al otro (BAUMANN *et al.*, 1983).

---

<sup>4</sup>Original: *Hierarchical procedures are inherently prone to providing suboptimal results at larger scales, for the simple reason that the area groupings made at the early stages of the analysis severely restrict the options available at the later stages* (COOMBES, 2000).

Por supuesto, el MAUP también afecta a esta familia de áreas funcionales. Así, WATTS (2009) compara los resultados de los métodos Intramax y TTWA para el caso de los MLT en Australia en 2001, y concluye que ninguno de estos métodos parece mostrar propiedades locales robustas con respecto a un criterio de cierre (autonomía), como el definido por una forma revisada imparcial de la función de pertenencia de una UT a un área utilizada por FENG (2008)<sup>5</sup>. Estos resultados llevan al autor a concluir que las técnicas de optimización combinatoria, como por ejemplo los algoritmos evolutivos (véase más adelante, sección 1.5), podrían ser un enfoque prometedor para la delimitación de MLT.

En CÖRVERS *et al.* (2009) los autores afirman que la fase de ajuste de los parámetros de las restricciones, necesaria para obtener resultados “satisfactorios”, es también una fuente de arbitrariedad; y que, independientemente del método que se emplee, las AF delimitadas deberán ser evaluadas frente a la alternativa de utilizar las áreas de carácter administrativo (comarcas, condados, etc.), empleando un conjunto de criterios relevantes para el uso que se vaya a dar a esas áreas. En su estudio concluyen que si las delimitaciones propuestas no superan ampliamente a las fronteras administrativas con respecto a una serie de indicadores relevantes no habrá un beneficio para la creación de políticas que empleen esas AF.

#### 1.2.2.1. Aspectos fundamentales en la delimitación de MLT

Los aspectos fundamentales a los que debe responder la delimitación de MLT son recopilados por EUROSTAT y COOMBES (1992) y citados por varios autores (CASADO-DÍAZ, 2000; COOMBES, 2002; FERIA *et al.*, 2005; CASADO-DÍAZ y COOMBES, 2011):

1. Propósito: deben ser áreas delimitadas en función de criterios estadísticos que resulten adecuadas para la planificación y generación de políticas.

---

<sup>5</sup>Esta función calcula la media entre la proporción de empleos de la UT cubiertos desde el área y la proporción de residentes de la UT empleados en el área

2. Relevancia: cada área delimitada debe ser identificada como un mercado laboral, esto es, como un área donde la mayor parte de la población reside y trabaja y cuyos límites son relativamente impermeables (áreas autocontenidas/autónomas).
3. Partición: cada UT (unidad mínima del territorio sobre la cual se realiza el agrupamiento) debe ser asignada a un único MLT, por lo que no debe existir solapamiento y la cobertura del territorio debe ser exhaustiva.
4. Contigüidad: cada MLT debe conformar un único territorio contiguo.
5. Autonomía: maximización de los flujos intrazonales, lo cual implica autonomía tanto desde el punto de vista de la oferta como de la demanda de la fuerza de trabajo.
6. Homogeneidad: el tamaño de las áreas resultantes, en términos de número de trabajadores, debe encontrarse en un cierto rango, de manera que se no se aceptan mercados demasiado pequeños o demasiado grandes.
7. Coherencia: los límites de las áreas obtenidas deben ser fácilmente reconocibles y reflejar las redes de transporte y los patrones de asentamiento.
8. Conformidad: se dará preferencia a aquellas soluciones que respeten los límites administrativos, dado que éstos son un referente importante en la generación de estadísticas y en la aplicación de políticas.
9. Flexibilidad: la metodología debe permitir cierta modificación de los resultados obtenidos a fin de reflejar las opiniones de las autoridades locales, así como resultar apropiada a la diversidad de patrones de desplazamiento existente en diferentes regiones.

Los puntos 1 y 2 se consideran objetivos del procedimiento, los puntos 3 y 4 son restricciones, los puntos 5 a 8 son los criterios que dirigen la regionalización (en orden descendente de importancia) y el último punto es una característica deseable del método que se emplee.

La preferencia por un mayor nivel de detalle (por soluciones con un número superior de mercados identificados a igualdad de otros criterios) es

también un principio rector común en este tipo de delimitaciones funcionales (ONS *et al.*, 2008). Esto se relaciona con la conveniencia de definir MLT que sean lo suficientemente pequeños como para permitir que sus residentes puedan cambiar de empleo a otro en el mismo mercado sin necesidad de cambiar su lugar de residencia (debe ser posible y razonable viajar diariamente entre cualquier par de puntos residencia-trabajo, según la definición de mercado). Todos estos requisitos son coherentes con los principales requisitos de un AF económica antes indicados: autonomía y cohesión, con la adición de tamaño mínimo y las limitaciones de contigüidad, y el hecho de que ninguna de las UT puede quedar sin asignar a algún MLT al final del proceso.

En la práctica, los criterios estadísticos que se contemplan directamente en los requerimientos del problema en los métodos de delimitación basados en reglas son (COOMBES *et al.*, 1986; CASADO-DÍAZ y COOMBES, 2011):

- Autonomía mínima (*self-containment*): cada mercado de la delimitación ha de alcanzar un cierto umbral de autonomía mínima, siendo calculada de forma más frecuente como el mínimo entre la autonomía de oferta (la proporción de residentes del área del mercado que encuentran trabajo en la misma) y la autonomía de demanda (la proporción de puestos de empleo del área de mercado que están ocupados por residentes de la misma).
- Tamaño mínimo: cada mercado ha de alcanzar un cierto número de residentes (normalmente) o de empleos.

Los objetivos que se persiguen indirectamente, no representados en las restricciones del problema, son:

- Nivel de detalle, número de mercados: es deseable aumentar el número de mercados sin reducir su autonomía. Este indicador está relacionado con el siguiente.
- Cohesión: si la autonomía se refiere a la ausencia de relaciones con el exterior de un mercado, la cohesión es el concepto inverso en el interior de los mercados: las partes que conforman un mercado han de estar relacionadas entre sí, cuanto más, mejor.

Perseguir el último objetivo, cohesión, implica que no se desean mercados formados por zonas con poca relación entre sí, aun cuando sean mercados autónomos, porque separados seguirían siéndolo y se conseguiría mayor nivel de detalle. Este indicador no suele considerarse en la literatura sobre delimitación de AF, debido a la dificultad de su cálculo (no existe una medida concreta que mida directamente la cohesión), aunque se trata el concepto. En los métodos más utilizados, como el TTWA, se aborda indirectamente a través del nivel de detalle, al implicar que aumentar la autonomía de los mercados sin reducir su número, o aumentar el número de mercados sin reducir su autonomía, aumenta también la cohesión, según la definición informal que hemos dado al principio de este capítulo.

### 1.3. Métodos de delimitación de MLT

A título de ejemplo, en esta sección se describen varios métodos usados en distintos lugares para la identificación de MLT. No se trata de una revisión exhaustiva dado que la finalidad es mostrar algunas de las diferentes estrategias que han sido típicamente utilizadas para abordar el problema y que, como se verá, comparten características que, en última instancia, hacen patente la pertinencia de introducir innovaciones metodológicas que permitan mejorar los resultados obtenidos.

Así, en esta sección se describen en detalle tres de los procedimientos de mayor aceptación a nivel institucional: Intramax (MASSER y BROWN, 1975), recientemente utilizada en MITCHEL y WATTS (2010); una variante de ese mismo método, empleada en EE.UU. para la delimitación de las *commuting zones* —una forma de MLT en la que no se impone ningún tamaño mínimo— y las *local market areas* (TOLBERT y SIZER, 1987) y que ha sido utilizado también en Suecia para delimitar los MLT de Suecia, Noruega, Dinamarca y Finlandia (SCB, 1992; KARLSSON y OLSSON, 2006); y la generalización del algoritmo TTWA empleado en Reino Unido denominada ERA por Eurostat (EUROSTAT y COOMBES, 1992).

El algoritmo de las TTWA en Reino Unido (COOMBES *et al.*, 1986) es un algoritmo de agregación no jerárquico, empleado en múltiples países con

variaciones: Nueva Zelanda (PAPPS y NEWELL, 2002; NEWELL y PERRY, 2005); Nueva Gales del Sur en Australia (WATTS, 2004); Comunidad Valenciana (CASADO-DÍAZ, 2000), Andalucía (FERIA *et al.*, 2005) y Aragón (ALONSO LOGROÑO *et al.*, 2008) en España; y con modificaciones más profundas en Italia (ISTAT, 1997, 2005). Asimismo, la influencia del algoritmo británico es señalada por ANDERSEN (2000, 2002) en Dinamarca y sus principios básicos han sido incluso reconocidos por el Eurostat, que recomienda que la generalización de ese método —conocida como *European Regionalisation Algorithm* (ERA)— sea considerada como un estándar para la definición de MLT en los países de Europa (EUROSTAT y COOMBES, 1992).

Todos los procedimientos de delimitación de AF trabajan sobre matrices de flujo origen-destino elaboradas a partir de los datos de movilidad recogidos en censos nacionales, inicialmente sin preprocesamiento, en las que las filas se corresponden con las UT de origen, las columnas con las UT de destino y la celda  $W_{ij}$  contiene el número de trabajadores que viajan diariamente desde la UT  $i$  a la UT  $j$ . En la nomenclatura que emplearemos a lo largo de este trabajo denominaremos a esta matriz  $W$  (de *workers* en inglés).

### 1.3.1. Método de las TTWA o ERA

El método de las TTWA, un método no jerárquico basado en reglas, fue presentado por COOMBES *et al.* (1986).

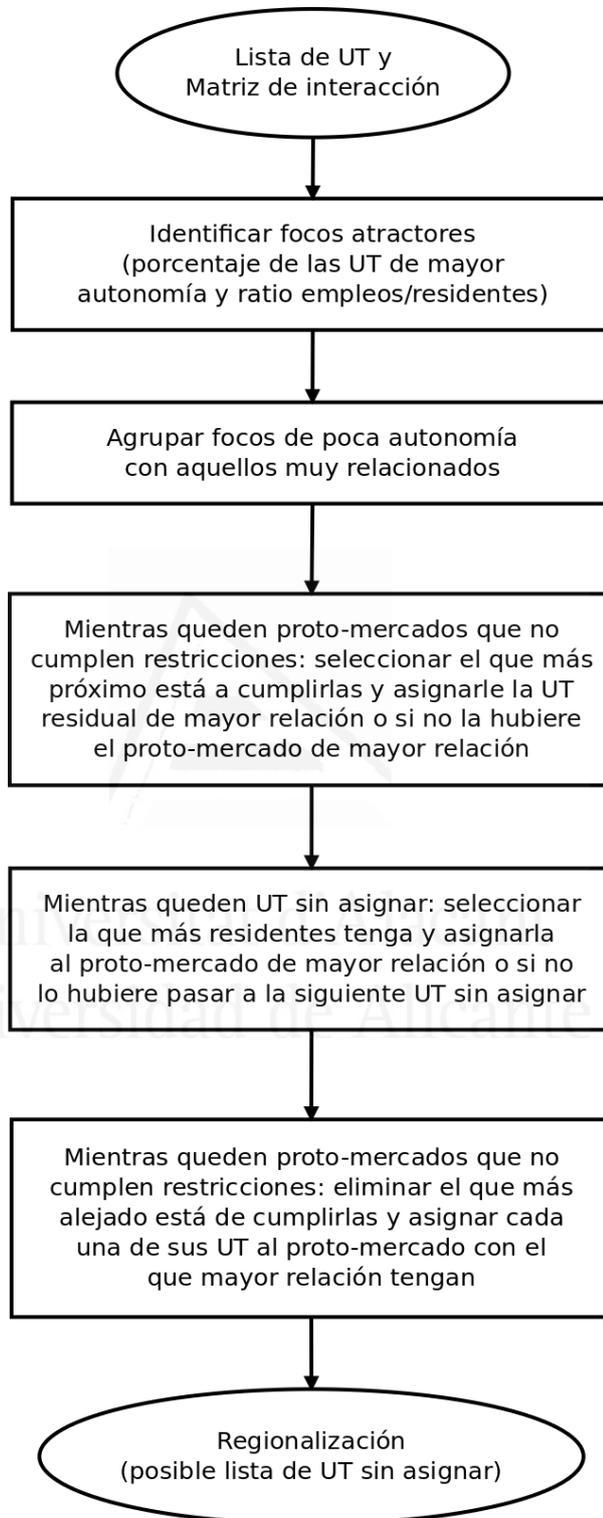
En la definición general del algoritmo TTWA original (COOMBES *et al.*, 1986), recomendada por Eurostat con la denominación de ERA (EUROSTAT y COOMBES, 1992; COOMBES, 2000), se contemplan las siguientes fases, no todas necesarias:

1. identificar entre las UT los focos en torno a los cuales se van a construir las AF;
2. agrupar los focos que estén estrechamente vinculados entre sí (p. ej., UT identificadas como focos como resultado de una mala selección de parámetros en la delimitación o de los caprichos de las fronteras de las UT);

3. asociar progresivamente cada UT no foco a las regiones con las que más vinculación tenga;
4. mientras alguna región no cumpla con los criterios estadísticos: desmembrar regiones empezando por la que más lejos esté de cumplir los criterios y asociar individualmente las UT que la constituían a las regiones de mayor vinculación.
5. considerar una etapa final de optimización de las regiones resultantes, p. ej., un procedimiento formal que trate de mejorar una puntuación general sobre los criterios estadísticos realizando cambios en las fronteras de las regiones mediante la reasignación de una UT en cada paso (OPENSHAW, 1984) o un proceso de consulta a las autoridades locales (en el que se acepten los cambios sólo cuando estén bien argumentados y no causen que cualquier región deje de cumplir los criterios estadísticos).

La primera fase puede seleccionar UT como focos en función de diferentes criterios, como por ejemplo aquellos que alcancen un umbral mínimo de número de residentes o de puestos de trabajo o un porcentaje de las UT totales. Si se toman todas las UT como focos esta etapa puede descartarse. En la segunda etapa se empleará otro criterio para identificar los vínculos fuertes entre focos, normalmente un umbral mínimo que la relación entre ambos focos debe superar, para fusionarlos entre sí por considerarlos como un único mercado. La tercera, que asigna las UT residuales a los protomercados creados anteriormente, suele emplear el mismo criterio que se haya escogido en la etapa anterior, en ocasiones con algún ajuste en su umbral. Es la cuarta etapa, siempre presente, es la que hace que estos métodos no sean jerárquicos, ya que permite romper los mercados que no cumplen los criterios estadísticos y distribuir sus UT entre los demás. El diagrama de este algoritmo, para el caso concreto del método aplicado en décadas anteriores en UK (COOMBES *et al.*, 1986), puede consultarse en la fig. 1.1.

En el último ejercicio oficial de delimitación de TTWA llevado a cabo en Reino Unido (ONS *et al.*, 2008) se emplea una versión simplificada de este algoritmo, en la que sólo se consideran dos etapas: formar una región con cada UT (se ignoran las etapas 2 y 3) y aplicar iterativamente la etapa 4. La justificación que se aduce a la simplificación es que las etapas previas

Figura 1.1: Diagrama del método de las TTWA (COOMBES *et al.*, 1986)

permitían reducir la complejidad del procedimiento, y que el incremento de rendimiento de los sistemas computacionales hace esto innecesario (ONS *et al.*, 2008).

Mucho más destacable es el hecho de que considerar todas las UT iniciales como focos hace posible la aparición de mercados (principalmente aquellos formados por pequeñas UT que individualmente no cumplían los requisitos de foco) que no se podrían dar de lo contrario, a pesar de cumplir con los criterios estadísticos de autonomía y tamaño mínimos, repercutiendo negativamente en la calidad de la solución final. Estos casos se dan en territorios con zonas de distintas características demográficas o geográficas (densidad y distribución de la población, nivel de detalle de las UT empleadas, características geográficas locales como ríos o cordilleras, entre otros), que requerirían un ajuste de parámetros independiente. Un ejemplo ilustrativo de esto se encuentra en el territorio de los EE.UU.: la región de la costa este está dividida en condados de pequeño tamaño y mucha población, situación que se invierte gradualmente según se avanza hacia la costa oeste, donde se pueden encontrar condados con una extensión comparable a la de muchos países pero una población inferior a diez mil habitantes. Exigir en esas regiones que al menos una de las UT que conformen cada mercado tenga un tamaño mínimo superior a cierto umbral puede obligar a la creación de mercados realmente enormes con foco en las grandes ciudades de la costa que difícilmente pueden justificar una cohesión interna aceptable: la gente de los condados de baja densidad del interior no viaja diariamente a la costa recorriendo 300 km. Se debe destacar que este problema no se resuelve completamente con tomar todas las UT como zonas foco, ya que el tamaño mínimo exigido a los mercados (independientemente de sus UT) puede también propiciar la aparición de mercados desmesurados y poco cohesionados. Algunos ejercicios solventan esto aplicando distintos umbrales en regiones diferentes (se divide el problema y se realizan análisis diferentes para parte como se hace en TOLBERT y SIZER, 1996) o empleando restricciones más elaboradas que tienen en cuenta características locales como la superficie de las áreas delimitadas (JOHNSON y KORT, 2004).

En las aplicaciones oficiales del algoritmo TTWA en Reino Unido (COOMBES *et al.*, 1986; ONS *et al.*, 2008), así como en las adaptaciones citadas anteriormente, el grado de vinculación entre una UT  $i$  (residual) y un área  $M$

(agregación de UT) es considerado como una relación simétrica de atracción o índice de interacción  $\Pi_{iM}$ , que toma en cuenta el flujo bidireccional entre ambos extremos y sus respectivos tamaños en número de residentes y de empleos. Esta interacción se calcula como la suma de los productos de las proporciones de residentes en origen y de empleos en destino que supone el flujo en cada sentido:

$$\Pi_{iM} = \Pi_{Mi} = \frac{W_{iM}}{R_i} \times \frac{W_{iM}}{J_M} + \frac{W_{Mi}}{R_M} \times \frac{W_{Mi}}{J_i} \quad (1.1)$$

donde  $W_{iM}$  es el flujo agregado de trabajadores de la UT  $i$  que viajan a cualquiera de las UT del área  $M$ ,  $W_{Mi}$  es el flujo en sentido inverso,  $J_i$  y  $R_i$  se corresponden con el total, respectivamente, de empleos y de residentes en la UT  $i$ , y  $J_M$  y  $R_M$  son los valores correspondientes para el agregado de UT que conforman el área  $M$ .

Este indicador será estudiado en profundidad en la sección 1.6.1, donde se emplea una generalización del mismo para formular el problema de optimización.

Se ha de destacar también que este tipo de métodos no emplea una función objetivo para dirigir su procedimiento, tan solo un indicador global que alcanza su valor máximo (la condición de terminación) cuando todos los mercados delimitados cumplen con los criterios estadísticos exigidos, siendo insensible a ningún otro indicador de calidad de la solución. En consecuencia, el procedimiento no constituye un método de optimización ni siquiera local, sino un algoritmo voraz que termina en cuanto la solución que se está construyendo cumple los criterios.

La sugerencia de una etapa final de optimización local, no definida formalmente, no es suficiente para asegurar una correcta exploración del espacio de soluciones y, por lo tanto, dada la complejidad del problema resulta poco probable que este método nos ofrezca soluciones próximas a las óptimas. Además, la elección de los umbrales en los criterios estadísticos sigue siendo una fuente de arbitrariedad —como el umbral de autonomía mínima típico, 75 %, escogido por ser el punto intermedio entre 50 % y 100 % sin más argumentación (COOMBES *et al.*, 1986)— que requiere de un fuerte conocimiento del problema y del territorio a analizar por parte de los expertos, y que entorpece o impide las labores de análisis comparativo entre diferentes ejercicios.

### 1.3.2. Método Intramax

El método de *clustering* Intramax fue presentado en MASSER y BROWN (1975) —casos de aplicación reciente incluyen NEL *et al.* (2008) en Sudáfrica y MITCHEL y WATTS (2010) en Australia—. El objetivo de este procedimiento es maximizar la proporción de interacción intramercado en cada paso del análisis *cluster* jerárquico.

Como en el caso anterior, este procedimiento toma como datos de entrada una matriz de flujos de movilidad cotidiana o interacción que transforma en la tabla de contingencia correspondiente (se añade la columna de totales por filas, la fila de totales por columnas y el gran total). Inicialmente se considera que cada UT es un AF. Iterativamente, se seleccionan las dos áreas  $A$  y  $B$  que más flujo comparten, aquellas que maximizan la ecuación:

$$\text{IMB1}_{AB} = \text{IMB1}_{BA} = \sum_{i \in A} \sum_{j \in B} (W_{ij} + W_{ji}) \quad (1.2)$$

En cada paso se combinan la pareja de áreas seleccionadas en una única área que agrega los flujos de ambas. El proceso termina cuando el territorio completo forma un único mercado. El algoritmo da como resultados la tabla informe y el dendrograma (diagrama en árbol) del proceso de *clustering* jerárquico. El experto ha de seleccionar entonces el nivel de agregación que tomará como la delimitación final de mercados, y quizás haya de tomar distintos niveles de división en ramas del dendrograma diferentes según criterios diversos, como pueden ser las características locales de las UT implicadas, la consulta a las administraciones locales, el respeto o no de fronteras administrativas, etc.

En una segunda versión del método Intramax los autores tomaron en cuenta las variaciones de los totales de fila y columna de la matriz de movilidad al medir la importancia de la interacción y adaptaron el criterio de agregación del análisis *cluster* como se muestra en la siguiente ecuación:

$$\text{IMB2}_{AB} = \text{IMB2}_{BA} = \sum_{i \in A} \sum_{j \in B} \left( \frac{W_{ij}}{R_i \times J_j} + \frac{W_{ji}}{R_j \times J_i} \right) \quad (1.3)$$

La primera versión daba preferencia a las relaciones en valor absoluto aun cuando supusieran relaciones proporcionalmente insignificantes: p. ej.,

entre Madrid y Barcelona viaja diariamente más gente que entre Madrid y Valverde de Alcalá (municipio de la Comunidad de Madrid de menos de quinientos habitantes), aunque para este último la dependencia de Madrid es mucho mayor que la de Madrid con Barcelona o viceversa. La segunda versión mide proporciones, de forma que se soluciona esa tendencia a que las relaciones entre grandes UT eclipsen a las menores. Esta propiedad la comparte el indicador empleado en el método TTWA ya descrito. De hecho, el indicador empleado en ese método está basado en el que se utiliza en la segunda versión del Intramax.

Los autores incluyeron una restricción de contigüidad, de forma que no se podían combinar dos áreas que no fueran colindantes (y en ese caso se combinaba el siguiente par de áreas más relacionadas). Esto ha sido criticado por algunos autores, que argumentan que el caso de que se produzcan discontinuidades es muy improbable dada la propia naturaleza de los datos de interacción y el algoritmo Intramax, y que la restricción puede influir en el algoritmo produciendo soluciones de inferior calidad (MITCHEL y WATTS, 2010).

### 1.3.3. Método de las CZ empleado en EE.UU.

El método empleado para delimitar las *commuting zones* y *local market areas* de EE.UU. en las últimas décadas (TOLBERT y SIZER, 1996), como se ha indicado previamente, es una forma de agregación jerárquica al igual que el método Intramax (MASSER y BROWN, 1975), aunque éste emplea unos criterios diferentes para dirigir el proceso de combinación de *clusters*.

Como en el caso anterior, este procedimiento toma como datos de entrada una matriz de flujos de movilidad cotidiana y compone con ella la tabla de contingencia correspondiente. A partir de la misma se calcula la matriz  $P$  de flujos proporcionales empleando la siguiente fórmula:

$$P_{ij} = \frac{W_{ij} + W_{ji}}{\min(R_i, R_j)} \quad (1.4)$$

De esta forma se obtiene una matriz simétrica ( $P_{ij} = P_{ji}$ ). A partir de la matriz  $P$  se calcula una matriz de disimilitud o distancia (funcional)  $D$ ,

como aproximación a los coeficientes de distancia necesarios para la técnica de análisis *cluster*:

$$D_{ij} = D_{ji} = 1 - P_{ij} \quad (1.5)$$

Valores de  $D_{ij}$  próximos a cero indicarán relaciones fuertes en ambos sentidos, y valores próximos a uno indicarán ausencia de relación.

Una vez se tiene la matriz de disimilitud se puede aplicar la técnica de análisis de agrupación jerárquica —*hierarchical clustering analysis*—. El algoritmo analiza la estructura presente en la matriz de disimilitud para construir un dendrograma que comienza con todas las UT separadas formando cada una un grupo, combinando en cada paso los dos grupos más cercanos entre sí en uno de nivel superior, y terminando con todas las UT unidas en un único grupo. La distancia entre dos grupos A y B se toma como el promedio de todas las distancias entre pares de objetos  $x \in A$  e  $y \in B$ , es decir, la distancia media entre los elementos de cada grupo (*average linkage*):

$$D(A, B) = \frac{1}{|A| \times |B|} \sum_{i \in A} \sum_{j \in B} D_{ij} \quad (1.6)$$

donde  $|A|$  y  $|B|$  son el número de elementos en cada grupo.

El método requiere de un parámetro, el umbral mínimo de distancia inter-*cluster*, para ofrecer una solución concreta (uno de los niveles del dendrograma). En las aplicaciones del método descrito los autores consideraron que los *clusters* producidos estaban formados por condados unidos entre sí por enlaces relativamente fuertes al fijar 0,98 como la distancia media inter-*cluster* a alcanzar. Al igual que en los métodos anteriores, la selección de este valor es intuitivo (como se dijo, en el caso del TTWA se trabaja sobre una magnitud mas comprensible, la autonomía), dependiente de la experiencia del autor y de las características del problema analizado, y por lo tanto supone una fuente de arbitrariedad y de dificultad para comparar entre diferentes estudios y extrapolar conclusiones aun usando la misma metodología.

Por otra parte, la función objetivo de un proceso de optimización no puede basarse únicamente en este indicador (1.6) (empleando, p. ej., la media o el mínimo de esa función entre cada par de mercados) dado que no considera todos los criterios estadísticos objetivos: maximizando la distancia

media inter-*cluster* obtenemos un territorio sin divisiones, que no captura ninguna información útil de la estructura de los mercados.

A diferencia de este método, en el Intramax no se emplea un indicador global, de toda la partición, para decidir el punto en que se corta el proceso de agregación. Pero si se crease ese indicador basado en la función de interacción que dirige el proceso de combinaciones, con las dos variantes de la función de interacción (1.2) y (1.3), tampoco se podría emplear como función objetivo dado que al maximizar los flujos intra-*cluster* se está maximizando la autonomía. Algo similar ocurre en el algoritmo de las TTWA, en que el indicador global que finaliza el proceso de combinaciones (y desintegraciones) es la autonomía mínima entre todas las áreas.

Una diferencia importante entre el algoritmo de las TTWA y los dos métodos jerárquicos descritos es que en estos últimos no se puede especificar una autonomía o tamaño mínimos deseados directamente. En el caso de las CZ, una solución con distancia media inter-*cluster* de 0,98 puede presentar mercados con una autonomía mínima inferior a 50% (algo que el método de TTWA no permitiría). De este modo, el método TTWA permite mayor control directo sobre los resultados a obtener. Aunque esto se puede conseguir con los métodos de análisis jerárquico fijando diferentes valores o alturas de finalización del *clustering* en ramas distintas del dendrograma, ese comportamiento no está integrado en los métodos descritos.

Los tres métodos se diferencian también en el cálculo de la interacción entre pares de UT o de agregados de UT. Como se explicará en la sección 1.6.1, esta diferencia es crucial a la hora de escoger una función objetivo para un problema de optimización.

## 1.4. Problemas de agrupación basados en interacción

Sintetizando lo expuesto en la sección 1.2, tenemos una familia de problemas en el ámbito de la ciencia regional y urbana que tratan la optimización de agrupaciones de elementos espaciales para maximizar la relación entre

elementos dentro de cada grupo (la cohesión de cada grupo) y minimizar la relación entre elementos de grupos distintos (maximizar la autonomía de cada grupo) sujeta a una serie de restricciones que varían de una aplicación a otra y que suelen incluir contigüidad, tamaño mínimo y número de grupos.

Este tipo de problemas guarda una gran similitud con otros muchos presentes en diferentes contextos y disciplinas.

Muchos problemas naturales consisten en agrupar un conjunto  $\mathbf{U} = \{u_1, u_2, \dots, u_n\}$  de objetos en una partición, es decir, una colección  $\mathcal{P} = \{C_1, C_2, \dots, C_m\}$  ( $m \leq n$ ) de subconjuntos de  $\mathbf{U}$  mutuamente disjuntos y conjuntamente exhaustivos tales que:

$$C_i \neq \emptyset, \quad \bigcup_{i=1}^m C_i = \mathbf{U} \quad \text{and} \quad C_i \cap C_j = \emptyset \quad \forall i, j \in [1, m], i \neq j \quad (1.7)$$

También se puede ver estos problemas como aquellos donde el objetivo es separar los miembros de un conjunto de objetos en uno o más grupos, donde cada objeto pertenece únicamente a un grupo; es decir, encontrar una agrupación de esos objetos. Aunque el término agrupación puede referirse a otro tipo de problemas (como solapamientos y coberturas), en este texto se utilizará agrupación para referirse siempre a problemas de partición. Además de las condiciones expuestas, deberán cumplirse una serie de restricciones y optimizarse una función de coste.

Un ejemplo de este tipo de problemas es el llamado *bin packing* (empaquetamiento), en el que una serie de objetos con cierto peso, coste o espacio se han de empaquetar en una cantidad finita de contenedores con capacidad  $V$  de forma que ningún contenedor exceda su capacidad y se minimice el número de contenedores (GAREY y JOHNSON, 1979).

El problema de delimitación de MLT es similar al de *bin packing* en cuanto que ambos son problemas de partición de un conjunto. Sin embargo, aún guardan una diferencia considerable: el *bin packing* (y muchos otros problemas de partición de un conjunto) no emplea datos de interacción. Ni las restricciones ni la función objetivo dependen de las relaciones entre los elementos que deben ser empaquetados. Si se añade o se sustrae un elemento de o a un contenedor sólo se ha de considerar la información local

del elemento que se está reasignando para actualizar el valor de la función objetivo, pero no es necesario consultar o calcular relaciones entre elementos (el peso de un elemento es el que es independientemente del de los demás).

Otros problemas de partición muy conocidos ya mencionados en este trabajo son los problemas de *clustering*, en los que una serie de elementos se han de agrupar o clasificar en función de la similitud entre el vector de características del elemento y un vector de esas mismas características que represente a la clase, que puede ser el vector promedio (k-medias) o el vector del individuo que mejor represente a la clase, el medioide (k-medioides), minimizando en ambos casos el error cuadrático medio. Como ya se señaló en la sección 1.2.2, los PDAF no disponen de un vector de atributos para cada elemento. La relación (ya sea considerada de atracción/similitud o de distancia/disimilitud) entre un elemento y un conjunto de elementos depende de todos y cada uno de los elementos que conformen ese conjunto.

Estos problemas, que se encuentran dentro del campo más extenso de la optimización combinatoria, son resueltos con frecuencia mediante programación lineal, técnicas de *clustering* y de partición como las mencionadas en este capítulo, además de redes neuronales autoorganizativas, algoritmos genéticos y otras técnicas de optimización estocástica.

Todos estos problemas son de gran complejidad, habiendo sido demostrada para alguno de ellos su pertenencia a la clase de problemas NP-difícil o NP-completo. De hecho, varios problemas de partición de grafos con dominios sustancialmente restringidos son NP-completos (GAREY *et al.*, 1976). El número de posibles particiones de un conjunto en conjuntos no vacíos sigue la secuencia de los números de Bell (ROTA, 1964) (p. ej., para 10 elementos es 115 975, para 20 es mayor que  $5,17E13$  y para 30 es mayor que  $8,47E23$ ). Por lo tanto, soluciones exactas pueden ser calculadas (normalmente con programación lineal) sólo en instancias muy pequeñas de los mismos (decenas o centenas de nodos, aunque depende de la complejidad del problema concreto y de las relajaciones que se empleen), siendo frecuente en los problemas reales encontrar tamaños de miles o millones de elementos. Es por esto por lo que se ha de buscar una aproximación a la solución óptima, siendo las más utilizadas las técnicas de optimización estocástica llamadas metaheurísticas, que incluyen a la clase de los algoritmos evolutivos así como otras técnicas de optimización local como son el *simulated annealing*, los

algoritmos de enjambre y la búsqueda dispersa (*scatter search*) entre otros (BLUM *et al.*, 2005).

En las clases de problemas de partición básica y de *clustering*, la complejidad de la función objetivo depende del número de grupos a evaluar, del número de elementos en cada grupo y del número de capacidades o características considerado. En cambio, en los problemas de agrupación basados en datos de interacción la complejidad depende también del número de aristas (relaciones) de los elementos (el número de características de los problemas de *clustering* se corresponderá al número de tipos de dato de interacción en los problemas basados en interacción). Dado que los problemas de optimización combinatoria calculan la función objetivo una gran cantidad de veces (tantas como soluciones se exploren), su coste computacional depende en gran medida de la misma. Esto implica que, de entre las tres subclases de problemas de partición que se acaban de esbozar, esta última tenderá a tener los problemas de más costosa solución.

En la literatura científica reciente podemos encontrar otras AF basadas en datos de interacción (no necesariamente espacial) que son objeto de interés en otras disciplinas como sociología, neurología, proteómica y genética, investigación operativa e ingeniería del *software*, entre muchas otras. A continuación se describen brevemente algunos casos concretos de esas AF y de los métodos con que se delimitan.

#### 1.4.1. Modularización de *software*

La *Search Based Software Engineering* (SBSE) es una aproximación a la ingeniería del *software* en la que problemas de la misma (diseño de pruebas, diseño de *software*, diseño de requerimientos, gestión de sistemas *software* y refactorización de código) son resueltos empleando algoritmos de optimización basados en búsqueda, principalmente técnicas de búsqueda local (como *hill-climbing*), *simulated annealing*, algoritmos genéticos (los más ampliamente empleados) y programación genética. Para el SBSE, en un problema de optimización basada en búsqueda el espacio de soluciones es explorado en busca de aquellas que pueden ser consideradas óptimas o casi óptimas, guiándose por una función objetivo que distinga entre soluciones mejores

o peores. Las aplicaciones de este tipo de enfoque han sido numerosísimas en los últimos años (véase HARMAN *et al.*, 2009), con mucha frecuencia mediante el uso de algoritmos evolutivos (véase RELA, 2004; MANTERE y ALANDER, 2005; JIANG, 2006).

Las funciones objetivo empleadas con éxito en estos problemas, particularmente en los problemas de rediseño o modularización de *software* (*Search Based Software Modularization* y refactorización de código), en los que se reconstruyen las fronteras entre módulos para facilitar su comprensión, gestión, mantenimiento y escalado, buscan maximizar la cohesión de los módulos del sistema y minimizar el acoplamiento entre módulos (*coupling*), equivalente a la dependencia externa en las áreas funcionales económicas, el concepto inverso al de autonomía (véase RÄIHÄ, 2009).

Centrándonos en el caso concreto de la modularización de *software*, también llamado problema de *clustering* de módulos, varios autores lo definen como similar al problema de encontrar casi-cliques en un grafo, los nodos del cual denotan módulos *software* y las aristas denotan relaciones de dependencia entre módulos. Los primeros en abordar este problema (MANCORIDIS *et al.*, 1998) llaman a este grafo un grafo de dependencia de módulos (*Module Dependency Graph*) (MANCORIDIS *et al.*, 1999). Por su parte, HARMAN *et al.* (2005) exploran la robustez de la función objetivo MQ de MANCORIDIS *et al.* (1999)<sup>6</sup> en comparación con la medida alternativa de cohesión y acoplamiento EVM de TUCKER *et al.* (2001), empleada inicialmente en trabajos de *clustering* de datos de expresión de genes (*gene expression data*) (KELLAM *et al.*, 2002), concluyendo que ambos muestran una alta robustez en presencia de ruido, siendo *EVM* la más robusta de las dos.

La definición formal de la medida  $EVM(\mathcal{P})$  para una partición (o *clustering*)  $\mathcal{P} = \{C_1, C_2, \dots, C_m\}$  es (TUCKER *et al.*, 2001):

$$EVM(\mathcal{P}) = \sum_{i=1}^m h(C_i) \quad (1.8)$$

donde  $h(C_i)$  es la evaluación asignada al *cluster*  $C_i$ , definida como:

$$h(C_i) = \begin{cases} \sum_{a=1}^{k_i-1} \sum_{b=a+1}^{k_i} L(c_{ia}, c_{ib}) & \text{si } k_i > 1 \\ 0 & \text{en otro caso} \end{cases} \quad (1.9)$$

<sup>6</sup>como veremos más adelante, con la terminología empleada en las AF y MLT, es equivalente a la suma de las autonomías de cada uno de los mercados

y donde  $k_i = |C_i|$  es el número de elementos del *cluster*  $C_i$ ,  $c_{xy}$  es el  $y$ -ésimo elemento del *cluster*  $C_x$  y  $L(c_{ia}, c_{ib})$  es la variable de relación entre elementos, que se define como:

$$L(c_{xy}, c_{pq}) = \begin{cases} 1 & \text{si existe una relación entre } c_{xy} \text{ y } c_{pq} \\ -1 & \text{en otro caso} \end{cases} \quad (1.10)$$

La definición formal (MITCHELL, 2002) de la medida  $\text{MQ}(\mathcal{P})$ <sup>7</sup> para una partición  $\mathcal{P} = \{C_1, C_2, \dots, C_n\}$  es:

$$\text{MQ}(\mathcal{P}) = \sum_{i=1}^n \text{CF}(C_i) \quad (1.11)$$

donde  $\text{CF}(C_i)$  es la evaluación asignada a un *cluster*  $C_i$ , definida como:

$$\text{CF}(C_i) = \frac{\mu_i}{\mu_i + \frac{1}{2} \sum_{j=1, j \neq i}^n \epsilon_{ij} + \epsilon_{ji}} \quad (1.12)$$

y donde  $\mu_i$  es el número de relaciones entre elementos del *cluster*  $C_i$  y  $\epsilon_{ij}$  es el número de relaciones entre elementos del *cluster*  $C_i$  y el *cluster*  $C_j$ . Esto es: el numerador incluye los flujos internos del *cluster* considerado y el denominador incluye esos mismos flujos y los que se comparten con otros *clusters*, es decir, todos los flujos con origen o destino en el *cluster* considerado. Estamos por tanto, como se ha dicho, considerando la autonomía del *cluster*, en términos de mercados locales.

Ninguna de las dos medidas es una métrica, no están normalizadas (al ser una suma de valores para cada *cluster* no están acotadas superiormente). Cuando se trata con grafos de dependencia de módulos conectados (en que no existe una “partición perfecta”, formada por *clusters* totalmente conectados internamente e independientes entre sí), el óptimo global de la función EVM se corresponde con la partición en un único *cluster*. Esto no es así en el caso de MQ: supongamos que dos *clusters*  $c_1$  y  $c_2$  comparten una relación entre ellos (una de ida y una de vuelta) y cada uno tiene internamente nueve relaciones ( $\mu_1 = \mu_2 = 9$ ,  $\epsilon_{12} = \epsilon_{21} = 1$ ), con los módulos en un único *cluster* obtenemos  $\text{MQ} = 1$  (100% de autonomía), mientras que en dos *clusters* diferentes obtenemos  $\text{MQ} = 1,8$ , por lo que se asigna mejor puntuación a la

<sup>7</sup>De hecho, en esa tesis se presentan dos versiones de esa medida, BasicMQ y TurboMQ; aquí nos referimos a la segunda

partición que mantiene separados esos dos *clusters*. De esta forma se tienen dos grupos con un 90 % de autonomía en lugar de uno con 100 %. Esto será realmente preferible cuando la cohesión del *cluster* único se deteriore lo suficiente como para justificar una pérdida del 10 % de autonomía.

Dado que estas medidas se basan en los mismos conceptos usados en la delimitación de MLT, será de interés considerar su aplicación a dicho tipo de problemas.

#### 1.4.2. Manufactura celular

El modelo conocido en inglés como *cellular manufacturing* se usa para la organización física de los lugares de trabajo y sus instalaciones, a partir de los principios de tecnología de grupo o *Group Technology* (GT), como se conoce comúnmente (KING y NAKORNCHAI, 1982; KUSIAK, 1987). GT es una filosofía o doctrina de manufactura que busca aumentar la eficiencia de los procesos de diseño y fabricación (reducir costes y/o aumentar el rendimiento), agrupando, por un lado, las partes o piezas que comparten similitudes en cuanto a su uso, geometría o requerimientos de fabricación (creando así familias de partes, *part families*) y, por otro lado, la maquinaria que se empleará para procesar cada familia de partes (formando así las celdas de manufactura, *manufacturing cells*). De este modo se minimiza el trasiego de piezas entre distintas celdas de manufactura (que pueden encontrarse en emplazamientos muy alejados entre sí) y se compartimentan los procesos de producción, lo que permite simplificar el diseño, operación y mantenimiento gracias a la normalización y al procesamiento en común (*common processing*).

En las aplicaciones de este tipo de problemas, conocidos en inglés como *machine-part cell formation problem*, se dispone de las relaciones en forma de dependencias entre las piezas implicadas en un producto y las máquinas requeridas para producirlas (la matriz máquina-parte de ceros y unos) y el objetivo que se persigue es agrupar las máquinas (en celdas o bloques) de forma que se minimicen las dependencias (trasiego de piezas) entre diferentes celdas y el coste de montaje de cada celda (en función de las características de las máquinas que la componen).

Otros problemas relacionados con el concepto de tecnología de grupo son la organización espacial de lugares de trabajo (*facilities/workshop layout*), en que se busca una distribución óptima de equipamiento, materiales, personal, infraestructuras y puntos de introducción de datos para minimizar movimientos y costes.

Este tipo de problemas se estudia desde hace varias décadas (FRANCIS y WHITE, 1974; MELLER y GAU, 1996), en el transcurso de las cuales se han aplicado multitud de metodologías de optimización combinatoria en su resolución (PAPAIOANNOU y WILSON, 2010), que incluyen numerosas técnicas de *clustering* y partición, programación lineal y metaheurísticas como *simulated annealing*, colonias de hormigas y algoritmos genéticos, siendo estos últimos los más utilizados y diversificados. De hecho, una revisión sistemática de la literatura de este campo en las dos últimas décadas centrada en las aplicaciones de algoritmos genéticos y sus variantes (GHOSH *et al.*, 2010) revela una tendencia a la hibridación y mayor complejidad de ese tipo de algoritmos, que incluye técnicas, principios y heurísticas de otras metodologías (búsqueda local, búsqueda tabú, búsqueda dispersa, *simulated annealing* y evolución diferencial, entre otros), ha permitido conseguir mejores resultados que las versiones no híbridas. Sus autores justifican la amplia utilización de este tipo de algoritmos, los algoritmos genéticos, en su capacidad de hibridación y de considerar múltiples objetivos, así como por la creencia generalizada de que es un método de búsqueda estocástica robusto e imparcial para explorar grandes espacios de soluciones. Algunos trabajos recientes que igualan o mejoran los resultados anteriores (JAMES *et al.*, 2007a) emplean algoritmos genéticos de agrupación (AGA), un tipo de algoritmo genético que se describe más adelante en esta introducción.

### 1.4.3. Planificación de áreas de localización

El problema se conoce comúnmente en inglés como *location (or registration) area planning*. Una aplicación real de este problema es el de la delimitación y planificación de áreas de localización (o de registro) en redes de telefonía móvil digital (*personal communication services*, PCS) (DEMIRKOL *et al.*, 2001). Partiendo de los datos agregados de conexiones entre dispositivos móviles —el número medio de conexiones por unidad de tiempo entre

cada posible origen y destino, que son las celdas (antenas) en que estén registrados en el momento de la llamada— y de itinerancia de dispositivos entre celdas —comúnmente referido como *roaming*—, se busca agrupar celdas contiguas en bloques (*location areas*) de forma que se minimice el coste de actualización (mantener la información de en qué área de localización está cada dispositivo) y el de paginación (determinar en qué celda, dentro del bloque, está el dispositivo). Estos objetivos (a minimizar) se oponen de una forma similar a como lo hacen la autonomía y la cohesión en el problema de la delimitación de los MLT, por lo que se ha de encontrar un equilibrio. Como se discute en DEMIRKOL *et al.* (2001), los costes de actualización y de paginación son incomparables y asignarles pesos relativos puede no ser acertado en toda la red analizada. En la literatura dedicada a este problema se ha representado el coste de paginación como una restricción en la formulación del problema fijando un límite superior en cada área de localización, siendo el objetivo a minimizar el coste de actualización (WANG *et al.*, 1998; DEMIRKOL *et al.*, 2001, 2004; VROBLEFSKI y BROWN, 2006; JAMES *et al.*, 2007b). En WANG *et al.* (1998) se presenta una reducción de esa formulación al del problema de partición de grafos, demostrando así que es NP-completo. Para la resolución de estos problemas se ha aplicado diversidad de métodos, que incluye de generación aleatoria, búsqueda tabú, algoritmos voraces como el *hill-climbing*, *simulated annealing* y algoritmos genéticos por nombrar algunos. Como en el caso anterior, es posible encontrar en la literatura reciente aplicaciones de algoritmos genéticos de agrupación con buenos resultados (VROBLEFSKI y BROWN, 2006). En JAMES *et al.* (2007b) se presenta un algoritmo genético de agrupación híbrido que iguala o mejora los mejores resultados obtenidos hasta el momento para una serie de instancias del problema.

#### 1.4.4. Detección de comunidades

Esta familia de problemas consiste en hallar en un grafo una división en grupos de nodos con alta densidad de conexiones internas y escasamente conectados a otros grupos, para encontrar lo que en teoría de redes se conoce como *community structure*, “divisiones naturales de las redes en módulos de alta densidad de conexiones”. El problema es conocido principalmente como detección de comunidades (*community detection*) (FORTUNATO, 2010). Se ha

usado una diversidad de técnicas para su resolución, entre ellas, k-partición de grafos (que no es tanto una técnica como un problema en sí mismo), *clustering* jerárquico (con técnicas tanto agregativas como divisivas), *clustering* particional (*partitional clustering*, la familia de métodos que incluye el k-medias y sus variantes) y *clustering* espectral (que emplea el vector de *eigenvalores* de la matriz de similaridad para realizar la agrupación con un número de dimensiones reducido). No se han encontrado aplicaciones de algoritmos evolutivos directamente relacionadas con la detección de comunidades en la literatura reciente (en LANCICHINETTI y FORTUNATO, 2009 se presenta una exhaustiva revisión hasta 2008), pero sí *simulated annealing* (GUIMERA y NUNES AMARAL, 2005) empleando como función objetivo una medida de modularidad a maximizar (NEWMAN y GIRVAN, 2004). En BRANDES *et al.* (2007) se presenta una prueba de que este problema es NP-completo según la formulación empleada en NEWMAN y GIRVAN (2004). Estos algoritmos se han aplicado especialmente a la detección de comunidades en redes de referencias bibliográficas en artículos académicos (*citation and collaboration networks*) (DE SOLLA PRICE, 1965), componiendo así mapas de ciencia que ayudan a visualizar los patrones de flujo de información entre disciplinas, autores o revistas (ROSVALL y BERGSTROM, 2008), así como a detectar posibles nuevas áreas o subáreas de estudio (SHIBATA *et al.*, 2008).

Se pueden encontrar muchas otras redes sociales que presentan estructura de comunidades. Se han delimitado este tipo de agrupaciones sobre redes de llamadas de teléfono móvil, conexiones entre perfiles públicos en redes sociales virtuales como LinkedIn y redes de colaboración (profesional, artística o estudiantil), entre otras (FORTUNATO, 2010).

Dentro de la biología se encuentran también muchos ejemplos de este tipo de redes (FORTUNATO, 2010), entre ellas las siguientes:

- Áreas funcionales en el córtex cerebral: regiones de neuronas densamente conectadas entre sí. Su detección es crucial para descubrir la organización y dinámica de funcionamiento del cerebro y sus procesos.
- Complejos de proteínas: grupos de proteínas que interaccionan mutuamente en el mismo lugar y momento, formando una suerte de objeto físico.

- Módulos funcionales de proteínas: conjuntos de proteínas implicadas en un proceso concreto, incluso si no lo hacen en el mismo lugar ni momento.
- Redes de interacción gen-proteína: para encontrar grupos de genes relacionados y ayudar en el análisis de la función de los genes.

El caso de la detección de comunidades —un campo de estudio relativamente reciente y en plena expansión— es el único de entre los descritos aquí en que la mayoría de métodos empleados corresponden a la categoría de heurísticas voraces y no se han aplicado algoritmos evolutivos para su resolución. Sin embargo, sí se ha resuelto empleando *simulated annealing*, cuya formulación de función objetivo y requerimientos puede ser empleada directamente en un AE u otras metaheurísticas poblacionales y cuyos operadores de búsqueda pueden ser considerados como mutaciones en algoritmos genéticos, de modo que su adaptación sería casi directa.

## 1.5. Algoritmos evolutivos en problemas de agrupación

Como se ha señalado, la estructura de todos los problemas descritos en la sección anterior es la misma: un problema de partición en que la función objetivo depende de las interacciones entre grupos o entre elementos de un grupo y no de medidas individuales de los elementos. Las diferencias entre los mismos radican en los requerimientos y la función objetivo o de coste a optimizar.

La gran flexibilidad de las técnicas de computación evolutiva —que incluyen algoritmos genéticos, programación genética y estrategias evolutivas (EE) (DE JONG, 2006)— para la hibridación con otras técnicas de búsqueda local y global (HRUSCHKA *et al.*, 2009), junto con la capacidad, común a todas las metaheurísticas poblacionales (aquellas que optimizan simultáneamente varias soluciones), de realizar una búsqueda imparcial de extensos espacios de soluciones (BLUM *et al.*, 2005), convierte a estos métodos en buenos candidatos para abordar problemas de gran complejidad, como son

los de delimitación de AF económicas y geográficas como las descritas en la sección 1.2.2 y otros tipos de AF como las revisadas en la sección 1.4.

Los algoritmos genéticos (AG) (HOLLAND, 1975), una de las clases de algoritmos evolutivos más ampliamente utilizados, son métodos de optimización estocástica poblacional basados en los principios de la evolución y su concepto de selección natural. Esta técnica tiene un largo historial de aplicación en muchas áreas de las ciencias de la computación (véase, p. ej., GOLDBERG, 1989 para una extensa introducción y DE JONG, 2006, donde se presenta un modelo unificado de estudio de algoritmos evolutivos, que engloba a los AG, las estrategias evolutivas [*evolution strategies*] y la programación evolutiva [*evolutionary programming*], así como aspectos teóricos y prácticos sobre el diseño y configuración de este tipo de metaheurísticas basadas en el paradigma de la computación evolutiva [*evolutionary computation*]).

Además de la intrínseca capacidad de exploración de los algoritmos evolutivos, su alta capacidad de paralelización (TOMASSINI, 2005; LUQUE *et al.*, 2005) los hace especialmente útiles en aquellos casos reales en que el tamaño del problema hace necesario un gran esfuerzo de cómputo sin importar la metaheurística empleada (ALBA *et al.*, 2005). En TOMASSINI (2005) se describen en detalle los modelos de estructura espacial, tanto convencionales como no ortodoxos, su análisis y un estudio empírico. La forma en que se estructure las poblaciones de soluciones en los algoritmos evolutivos —es decir, la vecindad entre individuos en la población de soluciones, qué soluciones pueden combinarse entre sí y cuándo— tiene una influencia considerable en el rendimiento general del método. Los modelos de islas (diferentes poblaciones interaccionan entre sí mediante migraciones o cruces de individuos) y celular (los individuos se disponen en una rejilla que define la vecindad de cada uno de ellos) ofrecen de forma sistemática mejores resultados y/o en menor tiempo que los modelos de *panmixia* (todos los individuos en la misma población) y *multistart* (cada población evoluciona de forma independiente y aislada).

### 1.5.1. Algoritmos evolutivos de agrupación

A excepción del caso de la detección de comunidades, en todas las familias de problemas descritos en la sección 1.4 encontramos aplicaciones de

una versión de los algoritmos genéticos denominada algoritmo genético de agrupación (AGA, *grouping genetic algorithm* en inglés), descritos en profundidad por FALKENAUER (1998), o también otros algoritmos evolutivos que integran una parte de los principios de los AGA. Estos algoritmos se diferencian de otros AG en la codificación de las soluciones (genotipo) para tener en cuenta la representación de particiones y en que los operadores (mutaciones y cruces) operan a nivel de grupo, siendo sensibles a su contexto, en lugar de operar a nivel de objeto. Por extensión, llamaremos AEA (término ya mencionado en la introducción) a cualquier técnica de computación evolutiva orientada a problemas de agrupación.

Los AGA buscan resolver o mitigar algunas de las principales desventajas de los AG aplicados a problemas de agrupación (FALKENAUER, 1998):

- Redundancia de codificación: cada posible partición (fenotipo) tiene múltiples representaciones (genotipo). Tanto la representación por numeración de grupos (en la que el cromosoma es un vector de enteros en que cada celda se corresponde con un elemento y aquellos elementos que comparten número están en el mismo grupo) como la de permutaciones (los números asignados a cada celda del vector se corresponden con el orden en que se deben leer los elementos para ser asignados a grupos) (JONES y BELTRAMO, 1991) se ven fuertemente afectadas por este problema (p. ej., en el caso de la numeración de grupos, 1 1 2 2 3 3 representa la misma partición que 2 2 3 3 1 1 y que 4 4 5 5 6 6). En la práctica aumenta innecesariamente el tamaño del espacio de búsqueda (replicando soluciones), reduciendo drásticamente la eficiencia del algoritmo (RADCLIFFE, 1991).
- Insensibilidad al contexto y perturbación de esquemas (*schema disruption*): los operadores de cruce orientados a objetos no consideran la información de grupos completos, permitiendo la destrucción de la información de los “buenos” grupos, bien al unirse con otros con los que no tienen relación porque compartían el mismo identificador de grupo, o bien al dividirse, a pesar de estar presente el mismo grupo (los mismos elementos) en ambos individuos combinados, porque no compartían el mismo identificador. Esto se puede apreciar en un sencillo ejemplo: las particiones 1 2 3 4 1 3 y 1 1 2 3 1 2 se diferencian entre

sí únicamente en que el segundo elemento forma un grupo unario en la primera partición y en la segunda se ha unido al grupo formado por los elementos primero y quinto. Los grupos 3 y 4 de la primera partición corresponden directamente con los grupos 2 y 3 de la segunda (son idénticos a pesar de tener diferente numeración). Al cruzar esos cromosomas entre las posiciones tercera y cuarta obtenemos 1 2 3 3 1 2 y 1 1 2 4 1 3: en el primer caso encontramos que el cuarto elemento, que formaba un grupo unitario en ambos padres, forma ahora un grupo con el tercer elemento, que a su vez formaba el mismo grupo en ambos padres con el sexto elemento. De hecho, el cruce de dos representaciones diferentes para la misma partición también produce soluciones muy diferentes de la original. De esta manera el cruce se comporta más como una mutación agresiva que como una recombinación genética, acabando con uno de los principios básicos de la computación evolutiva: que los descendientes de una solución dada se parezcan a su progenitor y con el principio de conservación de los buenos esquemas, fundamental en los algoritmos evolutivos (FALKENAUER, 1998). De esta forma, se está empleando información fuera de contexto: no es el identificador asignado a un grupo lo que lo caracteriza, sino los elementos que lo forman, y si esa información se ignora en el proceso de combinación, se ignoran también las características que hacían buena a la solución. Por este motivo los operadores de cruce estándar resultan muy destructivos para este tipo de problemas. La solución del reenumerado de grupos, para hacer coincidir el código del grupo que se inserta en un individuo con el código del grupo ya presente en el individuo que más se parece al primero, reduce enormemente esta desventaja, pero aún deja sin tratar varios aspectos como la coevolución de grupos (algunos grupos son “buenos” sólo en presencia de otros).

La representación en los AGA añade a la representación entera básica de numeración de grupos (p. ej., 112312) un vector de identificadores de grupo sobre el que realizar las operaciones de cruce (112312|123). Así se resuelve el problema de la insensibilidad de contexto y se reduce considerablemente la redundancia de la representación.

FALKENAUER (1998) sugiere la utilización de hasta tres tipos generales de operadores de mutación —destrucción de grupo, creación de grupo y modificación de grupo—, complementados con una rutina de reparación para

aquellos casos en que las restricciones del problema puedan hacer inválidos los resultados de estos operadores. En las aplicaciones de AEA a manufactura celular (JAMES *et al.*, 2007a) y de planificación de áreas de localización (JAMES *et al.*, 2007b) —los problemas más similares al PDAF en que se ha encontrado aplicación de AEA— no se emplea ningún operador de mutación y se utiliza un operador de cruce como el descrito en FALKENAUER (1998):

1. Seleccionar dos puntos de cruce en la sección de grupos de uno de los cromosomas padre.
2. Copiar la parte seleccionada de la sección de grupos del padre contribuyente sobre el padre receptor.
3. Modificar la sección de asignación de elementos del cromosoma receptor para reflejar la asignación a grupos del padre contribuyente.
4. Aplicar una heurística local, dependiente de la función objetivo y las restricciones del problema, para reparar individuos inválidos y/o mejorar la asignación.

En dos de los problemas de agrupación resueltos en FALKENAUER (1998) (*bin packing* y *load balancing*) se emplea un único operador de mutación, que elimina un grupo y emplea una heurística para repartir los elementos del grupo eliminado entre los restantes que crea un nuevo grupo para aquellos elementos que no puedan asignarse a ningún grupo (de este modo se combinan las tres estrategias de mutación de grupos en un solo operador). En otros dos problemas (economías de escala y *clustering* conceptual) emplea dos operadores de mutación, uno que crea un nuevo grupo extrayendo elementos de otros y otro que trata de eliminar un grupo repartiendo todos sus elementos entre los demás grupos (sin crear grupo nuevo). Vemos, pues, que no es frecuente que se implementen todas las operaciones propuestas o que se apliquen en operadores separados.

En BROWN y SUMICHRAST (2005) se evalúa el rendimiento de AGA frente a AG para diferentes problemas de optimización de agrupaciones (*assembly line balancing*, *bin packing* y *machine-part cell formation*). Los autores encuentran evidencias de que los primeros son generalmente mejores que los segundos debido a su mayor eficiencia computacional, pero que el aumento en rendimiento o calidad de las soluciones no siempre justifica el coste extra de diseño e implementación de los más personalizados (*tailored*) AGA.

### 1.5.2. Operadores informados y algoritmos híbridos

Los operadores genéticos estándar son aleatorios e insensibles al contexto del problema (el fenotipo), esto es, simplemente cambian alelos del genotipo al azar entre un conjunto o intervalo de posibles valores. Es el proceso de selección influida por la calidad (la selección natural) lo que aporta inteligencia al proceso de búsqueda. Si se sabe o se tienen indicios de que algunos valores posibles para cierto alelo son mejores que otros, se puede esperar que limitar las elecciones del operador genético mejore la tasa de convergencia y/o la calidad de la solución final alcanzada. Sin embargo, limitar demasiado el comportamiento de los operadores de búsqueda puede conducir a la incapacidad de éstos de alcanzar la solución óptima.

En HRUSCHKA *et al.* (2009) se presenta una revisión de la literatura reciente concerniente a la aplicación de algoritmos evolutivos para problemas de *clustering* y partición que incluye una taxonomía de este tipo de métodos en función de (a) si optimizan el número de particiones o es fijado a priori, (b) si la codificación es binaria, entera o real, (c) si la representación empleada presenta redundancia, (d) si los operadores son orientados a elementos o a grupos, (e) si los operadores de cruce son sensitivos al contexto y (f) si los operadores son guiados/informados o no. La bibliografía presentada en esa revisión revela una creciente tendencia a utilizar algoritmos de la clase de los AEA que introducen conocimiento del problema en los operadores.

En problemas con fuertes restricciones, como es el de delimitación de MLT, la proporción de soluciones válidas frente a las inválidas puede llegar a ser muy baja. En estos casos la eficiencia de la búsqueda aumentará, quizás drásticamente, al reducir el número de soluciones inválidas (aquellas que no cumplen las restricciones) generadas, que consumen recursos sin aportar nada al proceso de evolución. De forma que introducir en los operadores heurísticas sencillas que permitan producir tan pocos individuos inválidos como sea posible, sin coartar sus habilidades de explorar de forma robusta el espacio de soluciones, será una buena estrategia para aumentar la eficiencia de una metaheurística aplicada al PDAF. Esa es la estrategia seguida por, por ejemplo, SANDOU *et al.* (2008), quienes emplean operadores informados que reducen en gran medida el número de soluciones inválidas producidas por la mutación.

Si además se incorpora a los operadores información sobre la función objetivo a optimizar, de forma que los cambios más prometedores (en términos de calidad además de viabilidad) tengan más probabilidades de ocurrir, podría esperarse una aceleración de la convergencia al óptimo (como se demuestra en WOLPERT y MACREADY, 1997, con la aplicación del teorema de *no free lunch*). En ese caso se habrá de poner especial cuidado en la forma de selección de esos cambios para evitar aplicar una presión demasiado fuerte que haga demasiado poco probables cambios menos prometedores que propician el estancamiento en óptimos locales. En RASHEED *et al.* (2003) se realizan varias mutaciones aleatorias sobre un modelo simplificado de la función objetivo a optimizar y la mutación que mejor resultado consigue es seleccionada para su aplicación efectiva en el modelo completo, mejorando ampliamente los resultados de la versión con mutación puramente aleatoria.

Los algoritmos meméticos (MOSCATO y COTTA, 2010), en que un AE se combina con técnicas (normalmente heurísticas de búsqueda local) de reparación de individuos inválidos y optimización local, pueden ser vistos como una forma de incorporación de información en el proceso evolutivo, dado que esas heurísticas estarán guiadas por alguna forma de medida de calidad. Por ejemplo, en KRASNOGOR y GUSTAFSON (2004) se estudia el uso de heurísticas de búsqueda local autoadaptativas en un algoritmo memético cuyos resultados supera ampliamente a los del AG equivalente. En ŠTRANCAR *et al.* (2005) se presenta un AG que también incluye un operador de búsqueda local (*downhill simplex*) y un operador informado.

## 1.6. Propuesta de solución mediante AEA

En esta tesis se propone, pues, abordar el problema de la delimitación de MLT basada en datos de interacción, ya introducido en la sección 1.2.2, como un problema de optimización sujeto a una serie de restricciones resuelto con un AEA, en el que el objetivo es identificar MLT cohesionados (alta interacción entre las UT que los constituyen) y relativamente autónomos (poca interacción entre distintos MLT).

Como se ha indicado en la sección 1.5.1, el AEA se instrumenta con un conjunto de operadores genéticos orientados a grupos, es decir, que tienen en

cuenta la información de grupos y el contexto de sus relaciones, en contraposición a los operadores típicos, orientados a objeto, que sólo consideran la información propia de cada objeto. Estos operadores cubren las tres estrategias empleadas comúnmente en los operadores de mutación de los algoritmos genéticos de agrupación —creación de grupos, destrucción de grupos y modificación de grupos— mediante una serie de operaciones lógicas sobre grupos disjuntos, generales y configurables. De esta forma, se persigue asegurar una gran versatilidad de búsqueda, tanto de explotación como de exploración, en cualquier tipo de problema de agrupación, a la vez que se pretende dotar al método de flexibilidad y fácil adaptación para su aplicación en otros problemas similares. Por otra parte, se tiene la posibilidad de emplear el conjunto de operadores diseñados en otras metaheurísticas de búsqueda estocástica como *simulated annealing* (KIRKPATRICK *et al.*, 1983), búsqueda dispersa (GLOVER, 1977) o algoritmos de enjambre (DORIGO, 1992). Estos operadores se presentan aquí adaptados a los objetivos y restricciones específicos del problema tratado específicamente en este trabajo, la delimitación de MLT, aunque el enfoque desarrollado permite la delimitación de AF mediante una estrategia general, flexible y objetiva, de fácil implementación y alto rendimiento, que permite su aplicación en problemas con diferentes requerimientos u objetivos (dentro de la familia de problemas) con pocos ajustes en el método, así como la comparación objetiva entre diferentes casos del mismo problema.

Los datos de interacción con que se opera en este problema son los flujos de trabajadores en sus viajes cotidianos, con origen en la unidad territorial (UT) de residencia y destino en la UT en que se localice el puesto de empleo. Las UT en este caso pueden ser términos municipales, comarcas, condados, áreas censales, zonas catastrales y cualquier otra forma de parcelación existente en el territorio tratado sobre la que se puedan referenciar los orígenes y destinos de los viajes cotidianos de los trabajadores.

En la nomenclatura que emplearemos a lo largo de esta sección, denotaremos la matriz de flujos como  $W$  (de trabajadores en inglés). Las UT se representarán con letras minúsculas ( $a, i, u$ , etc.), los mercados (o grupos de UT en general) con letras mayúsculas ( $A, B, X$ , etc.), las particiones (de un territorio) con letras caligráficas ( $\mathcal{P}, \mathcal{Q}$ ), y al territorio a particionar, el conjunto de UT, con letras mayúsculas en negrita ( $\mathbf{U}$ ). Así mismo emplearemos

$W(A, B)$  para referirnos al flujo agregado total con origen en el mercado  $A$  y destino en el  $B$ , es decir, el total de trabajadores que residen en alguna UT asignada a  $A$  y trabajan en alguna UT asignada a  $B$ :

$$W(A, B) = \sum_{\forall i \in A} \sum_{\forall j \in B} W_{ij} \quad (1.13)$$

Nótese que  $W(A, B)$  no tiene por qué coincidir con  $W(B, A)$  y, de hecho, rara vez lo hará.

En lo que resta de esta sección se sintetizan las cuestiones más directamente relacionadas con la formulación del problema y se describen las partes comunes a la mayoría de publicaciones que forman esta tesis por compendio, entre las que se incluyen las restricciones del problema, el desarrollo de las funciones objetivo, la representación de las soluciones (cromosoma), la estructura general del AEA y los casos de estudio empleados en la experimentación.

### 1.6.1. Formulación como problema de optimización

Como se indicó en la sección 1.2.2, el problema consiste en la agrupación de una serie de elementos espaciales denominados unidades territoriales (UT) —típicamente distritos, municipios, condados o áreas censales—, en áreas funcionales, AF, de modo que la proporción de flujos —trabajadores en su viaje diario al trabajo— que cruzan fronteras de mercado en su viaje al trabajo sea baja, mientras que el número de AF definidas sea elevado y las UT que las conformen estén fuertemente relacionadas entre sí, sujetas a una serie de restricciones que incluyen unos valores mínimos de autonomía y tamaño.

La formulación es la siguiente: Sea  $\mathbf{U} = \{a, b, \dots, n\}$  un conjunto de  $N = |\mathbf{U}|$  UT (el territorio a particionar) y  $W$  la matriz cuadrada de  $N \times N$  celdas que contiene los flujos de trabajadores entre cada par de UT, donde  $W_{ij}$  es el número de trabajadores que residen en  $i$  y trabajan en  $j$  (así,  $W_{ii}$  es la cantidad de gente que simultáneamente reside y trabaja en la UT  $i$ ). El objetivo en este problema de optimización es obtener la partición de

$\mathbf{U}$ ,  $\mathcal{P} = \{A, B, \dots, Q\}$  (para que sea una partición se ha de cumplir que  $X \neq \emptyset \forall X \in \mathcal{P}$ ;  $\bigcup_{X \in \mathcal{P}} X = \mathbf{U}$  y  $A \cap B = \emptyset \quad \forall A, B \in \mathcal{P}, A \neq B$ ) con  $K_{\mathcal{P}} = |\mathcal{P}|$  desconocido a priori, que maximice una función objetivo  $f(\mathcal{P})$  y satisfaga las restricciones del problema, que se describen más adelante en esta sección.

### 1.6.2. Función objetivo para la delimitación de MLT

Como se vió en las secciones 1.2.2 y 1.3, no existe una formulación única para las restricciones o los objetivos a cumplir en una delimitación de MLT. En este trabajo, en que se aborda el problema como uno de optimización, es necesario definir una función objetivo que permita comparar la calidad de diferentes soluciones. La manera de medir la calidad de las soluciones es un debate abierto cuya resolución escapa al alcance de este trabajo. Sin embargo, es interesante hacer un análisis de las distintas magnitudes que se pueden medir en una partición de las celdas de una matriz de interacción (agregados de flujos) y ponerlas en relación con los distintos objetivos y restricciones considerados en la delimitación de MLT (autonomía, detalle, cohesión). En este apartado se presenta ese análisis y la propuesta de formulación de función objetivo y restricciones que se empleará en el resto del trabajo.

Primero se revisarán los diferentes objetivos que se persiguen comúnmente en los PDAF, ya tratados en el capítulo 1:

- Autonomía (*self-containment*): los mercados han de ser lo más autónomos posible, en términos de flujos, esto es, poco relacionados con el resto de mercados. Este atributo se mide de diferentes maneras (como se verá más adelante), según se prefiera la autonomía de demanda (flujos de entrada cuyo origen está en el mismo mercado), la de oferta (flujos de salida cuyo destino está en el mismo mercado) o una forma que combine a ambas, como puede ser la autonomía mínima (el mínimo entre los valores de oferta y de demanda) y el valor medio entre ambos, que en este trabajo denominamos autonomía combinada.
- Nivel de detalle, número de mercados: se prefiere aumentar el número de mercados sin reducir su autonomía (como se indicó en la sección

1.2.2.1). Este indicador se emplea, combinado con el anterior, como aproximación al siguiente.

- **Cohesión:** si la autonomía se refiere a la ausencia de relaciones con el exterior de un mercado, la cohesión es el concepto contrario en el interior de los mercados: las partes que conforman un mercado han de estar relacionadas entre sí, cuanto más, mejor. Y, por lo tanto, no se desean mercados formados por UT con poca relación entre sí, aun cuando sean mercados autónomos. Este indicador no suele considerarse directamente en la literatura sobre delimitación de MLT, aunque se trata el concepto en las definiciones teóricas. En los métodos más utilizados, como el TTWA, se aborda indirectamente a través del nivel de detalle, al implicar que aumentar la autonomía de los mercados sin aumentar su tamaño, o aumentar el número de mercados sin reducir su autonomía, aumenta también la cohesión. Pero no existe en la literatura sobre AF económicas ninguna medida directa de esta magnitud.

Las características de las delimitaciones se pueden medir a tres niveles de agregación: nivel de UT (elemento, nodo), nivel de mercado (grupo, sub-grafo) y nivel de delimitación (partición, grafo). Por ejemplo, la autonomía, definida como la proporción de flujos con destino y origen en el mismo mercado, puede medirse para el total de la delimitación, para cada uno de los mercados independientemente y para cada una de las UT, pudiendo emplear en los dos últimos casos la media u otro estadístico como indicador de calidad de la delimitación. Cada forma implica unas ventajas distintas. Siguiendo con el ejemplo de la autonomía, el valor a nivel de delimitación representa directamente la cantidad de flujo (de trabajadores) que se ha conseguido confinar dentro de mercados, mientras que la media y el mínimo de la autonomía de los mercados es un mejor indicativo de la calidad general de los mismos (dado que un alto valor a nivel de delimitación puede esconder mercados con muy poca autonomía), pero podría enmascarar resultados negativos a nivel de partición: p. ej., gran cantidad de pequeños mercados muy autónomos pero de poca población y unos pocos grandes mercados con baja autonomía pero que representan el grueso de la población en el territorio, de forma que la autonomía de la delimitación es baja; a su vez, guiarse por la autonomía de la delimitación podría ofrecer un resultado contrario pero igualmente indeseable: muchos mercados pequeños de baja autonomía

y unos pocos grandes autónomos. Ambos casos pueden ocurrir también al contabilizar la autonomía a nivel de UT, pero en ese caso las desigualdades entre tamaños de diferentes UT son conocidas a priori y están más limitadas que en el caso de nivel de mercados (las desigualdades entre agregaciones de UT pueden ser mucho mayores).

Como se explica en el resto de esta sección, en este trabajo se ha optado por desarrollar una función objetivo basada en los valores de interacción acumulada a nivel de mercado. Por tanto, queda descartada la evaluación directa de la bondad de asignación de cada UT individual. Esta cuestión se retomará en el trabajo que se resume en la sección 2.6.

### 1.6.2.1. Índice de interacción entre dos áreas

Una diferencia importante entre los tres métodos descritos en la sección 1.3 (TTWA, Intramax y métodos de las CZ) radica precisamente en cómo se cuantifica la relación (de atracción) entre dos UT o agregaciones de UT. En el TTWA y en la segunda versión de Intramax se considera el peso del flujo con respecto a los tamaños de ambos extremos de la relación, mientras que en el método de las CZ sólo se considera (indirectamente) la proporción en el extremo de menor tamaño. En el último caso se asignará la misma relación a flujos del mismo valor absoluto pero extremos de tamaños muy desiguales<sup>8</sup>.

Por lo tanto se preferirá un índice de interacción que, al igual que en la segunda versión del método Intramax (1.3) y en el método de las TTWA, tenga en cuenta el tamaño de ambos extremos, así como el flujo en ambos

<sup>8</sup>Esto se explica mejor con un ejemplo: sea una UT  $a$  con cien residentes y ochenta empleos, una UT  $b$  con diez mil residentes y once mil empleos, y otra UT  $c$  con doscientos residentes y doscientos veinte empleos; treinta residentes de la UT  $a$  trabajan en la UT  $b$  y diez de sus empleos los ocupan residentes de la UT  $b$ ; y la misma cantidad de trabajadores viajan diariamente entre las UT  $a$  y  $c$ , y entre  $c$  y  $a$ , respectivamente. Al ser los flujos absolutos iguales y emplear en ambos casos el mismo denominador, la disimilitud (1.5) de ambas relaciones es la misma,  $D_{ab} = 0,6 = D_{ac}$ . En cambio, el índice de interacción (1.1) sí distingue entre ambos casos:  $\Pi_{ac} = 0,010340909 \gg 0,000943182 = \Pi_{ab}$ . Dado que el flujo compartido entre  $a$  y  $c$  es mucho más significativo para  $c$  de lo que lo es para  $b$ , es razonable que la medida empleada lo refleje. De hecho, en ciertos casos ambas medidas darán resultados contrarios: si el flujo entre  $a$  y  $c$  fuera ligeramente inferior al compartido por  $a$  y  $b$ , la ecuación (1.5) les asignaría mayor disimilitud (menor relación), mientras que (1.1) seguiría dándole mucha más importancia que al otro flujo.

sentidos (práctica común a todos los métodos descritos). Concretamente nos basaremos en el índice de interacción empleado en el método de las TTWA (1.1). La ecuación se describe a continuación:

$$\Pi(A, B) = \Pi(B, A) = \frac{W(A, B)^2}{R_A J_B} + \frac{W(B, A)^2}{R_B J_A} \quad (1.14)$$

donde  $R_X = \sum_{\forall i \in X} \sum_{\forall j \in U} W_{ij}$  es el total de residentes ocupados (salidas) del mercado  $X$ , y  $J_X = \sum_{\forall i \in X} \sum_{\forall j \in U} W_{ji}$  es el total de empleos (entradas). También se puede expresar en términos de proporciones de flujos de entrada y salida en uno y otro sentido de la relación:

$$\Pi(A, B) = \Pi(B, A) = \underbrace{\frac{W(A, B)}{R_A}}_{PR_{AB}} \cdot \underbrace{\frac{W(A, B)}{J_B}}_{PJ_{BA}} + \underbrace{\frac{W(B, A)}{R_B}}_{PR_{BA}} \cdot \underbrace{\frac{W(B, A)}{J_A}}_{PJ_{AB}} \quad (1.15)$$

donde  $PR_{XY}$  representa la proporción de población ocupada residente en  $X$  que está empleada en  $Y$ , y  $PJ_{XY}$  es la proporción de empleos en  $X$  que son ocupados por residentes provenientes de  $Y$ . Al ser la suma de productos de proporciones que no pueden superar nunca la unidad, este índice está acotado en  $[0, 2]^9$ . Para interpretar este índice de forma más intuitiva se puede considerar hallar su raíz cuadrada<sup>10</sup>.

Sintetizando, se tiene un índice de interacción que mide la atracción entre dos agregados de una o más UT en función del flujo bidireccional agregado entre ambos grupos de UT y de los tamaños de cada extremo del flujo (totales de entradas y salidas). Esta función es similar a una medida de atracción gravitacional, de forma que se tiene en cuenta el tamaño de los polos en relación con el valor absoluto del flujo entre ellos, y se asigna así mayor valor a la interacción entre mercados pequeños, relaciones que quedarían eclipsadas por los flujos entre mercados más grandes que frecuentemente tienen valores absolutos superiores, a pesar de que en ambos extremos representan una pequeña porción de los trabajadores y empleos. De esta forma se mitiga la generación de mercados de mayor tamaño y menor cohesión, lo que convier-

<sup>9</sup> Aunque tenderá a dar valores cercanos a cero y típicamente inferiores a 0,01

<sup>10</sup> Un ejemplo: supongamos que dos UT  $c$  y  $d$  de igual tamaño envían cada una el 7% de sus trabajadores a la otra;  $\Pi(c, d) = 0,07^2 + 0,07^2 = 0,0098$ ; luego un valor de 10% en el índice se correspondería con un flujo bidireccional de aproximadamente el 7% de los trabajadores de ambas UT

te a esta medida en una especialmente interesante en aquellos PDAF en los que prime la cohesión o el nivel de detalle.

### 1.6.2.2. Funciones objetivo propuestas

Construimos, entonces, la función objetivo sobre ese indicador de interacción entre pares de áreas (1.14). Diferentes alternativas se pueden plantear, en función, por una parte, de si se consideran las relaciones entre UT individualmente o de forma agregada y, por otra parte, de si se considera únicamente la interacción acumulada por cada mercado o la diferencia de ésta con, p. ej., la interacción externa máxima. En esta investigación se optó por emplear la interacción acumulada de cada UT con su mercado, calculada en cada caso como la interacción entre la UT y el agregado del resto de UT de ese mercado:

$$f_0(\mathcal{P}) = \sum_{\forall A \in \mathcal{P}} \sum_{\forall i \in A} \Pi(\{i\}, A \setminus \{i\}) \quad (1.16)$$

Además, para dar mayor prioridad a la detección de más mercados independientes, se incluye también la versión que multiplica el valor final por el número de mercados identificados (ecuación (1.17)). La función objetivo se puede expresar con las siguientes ecuaciones:

$$f_1(\mathcal{P}) = K_{\mathcal{P}} \sum_{\forall A \in \mathcal{P}} \sum_{\forall i \in A} \Pi(\{i\}, A \setminus \{i\}) \quad (1.17)$$

### 1.6.3. Restricciones

Como ya se comentó en este capítulo, diferentes restricciones pueden imponerse sobre las soluciones de este problema, siendo típicas las de autonomía mínima, tamaño mínimo, superficie máxima o flujo hacia otro mercado máximo.

#### 1.6.3.1. Tamaño y autonomía mínimos

En la formulación que escogemos para este trabajo, siguiendo la que se emplea en los trabajos de Coombes y de Casado-Díaz (véase, p. ej., CASADO-DÍAZ y COOMBES, 2005), se considerará como válida una solución  $\mathcal{P}$  dada,

siempre que cada mercado  $A \in \mathcal{P}$  supere unos umbrales mínimos de tamaño en términos de población ocupada ( $\beta_4$ ) y de autonomía mínima ( $\beta_1$ ). La ecuación que describe la restricción de tamaño mínimo es la siguiente:

$$R_A \geq \beta_4; \beta_4 \geq 1 \quad (1.18)$$

La restricción de autonomía mínima se describe en la siguiente ecuación:

$$AM_A = \frac{W(A, A)}{\max(R_A, J_A)} \geq \beta_1; 0 \leq \beta_1 \leq 1 \quad (1.19)$$

De este modo ambas autonomías unidireccionales, la de oferta y la de demanda, han de superar cierto umbral mínimo.

Entornos muy urbanizados del mundo real se caracterizan por la intensidad y la complejidad de la red de flujos de movilidad, lo que hace difícil identificar a los grupos aislados de UT en estos entornos, lo que resulta en enormes mercados que, a pesar de ser muy autónomos, son demasiado grandes para permitir viajes diarios al trabajo entre cualquier par de UT en el mercado (el mercado tiene una baja cohesión y podría estar muy lejos de la definición del mercado ideal, donde todo proveedor puede negociar con cualquier demandante).

Para facilitar la identificación en esos entornos de más y menores MLT independientes, se introdujo una “compensación” (*trade-off*) entre ambas restricciones (autonomía y tamaño mínimos), similar a la utilizada en el método TTWA original (COOMBES *et al.*, 1986) pero que emplea la formulación propuesta por CASADO-DÍAZ y COOMBES (2005). De acuerdo con esa propuesta, se introducen otros dos umbrales,  $\beta_2$  y  $\beta_3$ , de forma que la autonomía mínima es relajada linealmente de  $\beta_2$  a  $\beta_1$  ( $0 \leq \beta_1 \leq \beta_2 \leq 1$ ) para tamaños de población de  $\beta_4$  a  $\beta_3$  ( $1 \leq \beta_4 \leq \beta_3$ ). Para cada mercado en una delimitación dada, esta compensación se evalúa como sigue:

$$AM_A \geq \beta_2 + \frac{\beta_2 - \beta_1}{\beta_4 - \beta_3} (R_A - \beta_4) \quad (1.20)$$

Si un mercado no satisface la tres restricciones anteriores —ecuaciones (1.18), (1.19) y (1.20)—, la solución que lo contiene es considerada inválida.

Al igual que se hace en el método de las TTWA, en alguna de las rutinas que se presentan en las publicaciones de los apéndices D y G (PPSN 2008 y ESWA 2012) se calcula el grado de validez de los mercados que no cumplen todos los requisitos, para tratar de reparar primero aquellos más próximos a ser válidos. El grado de validez de un mercado  $A$  está en el intervalo  $[0, 1]$ , siendo necesario alcanzar 1,0 para ser válido. Para aquellos mercados que no cumplen las restricciones (1.18), (1.19) y (1.20) se calcula como sigue:

$$\text{validez}(A) = \min\left(\frac{AM_A}{\beta_2}; 1\right) \times \min\left(\frac{R_A}{\beta_4}; 1\right) \quad (1.21)$$

### 1.6.3.2. Vecindad funcional y contigüidad geográfica

La formulación de las restricciones descritas hasta aquí en este apartado es común a todas las publicaciones incluidas en esta tesis. Además de las anteriores, en varias de las publicaciones presentadas se incluyó el requisito de que todas las UT de un mercado válido estén conectadas entre sí a través de los  $\lambda$  flujos de entrada o de salida de mayor valor absoluto. Se llamó a esta condición “vecindad funcional”. El motivo es tratar de reducir la aparición de mercados discontinuos sin introducir más información al problema (como se señaló en la sección 1.3.2, algunos autores consideran perjudicial la inclusión de la restricción de contigüidad geográfica durante el proceso de delimitación al considerar que coartaba el análisis, lo que propicia resultados subóptimos; nótese sin embargo que esas consideraciones eran referentes al empleo de técnicas voraces). Como se verá en el capítulo 2, esta restricción de vecindad funcional da buenos resultados frente a su no utilización, pero no es suficiente para acabar con las discontinuidades, por lo que se hace necesario desarrollar algún método que las solvete a posteriori del análisis. Esto puede deberse en parte a las características de los datos de interacción y a la dificultad de ajustar el parámetro  $\lambda$  independientemente para cada UT o porción del territorio (se utilizó el mismo valor para todo el territorio).

Por esta causa y teniendo en cuenta que, por una parte, no existe un claro consenso a la hora de incluir o no la contigüidad geográfica y que, por otra, no se había abordado nunca el problema mediante una técnica de optimización estocástica, se decidió incluir en una de las versiones del AEA

la restricción de contigüidad geográfica. A tal efecto, se incluyó en los datos de entrada la matriz de adyacencia  $C_{N \times N}$ , donde  $C_{ij}$  toma el valor 1 si las UT  $i$  y  $j$  son geográficamente adyacentes y 0 en caso contrario, por lo que cualquier UT en un MLT válido debe ser accesible desde cualquier otra UT del mismo MLT sin cruzar sus límites geográficos, es decir, para cada par de UT en un mercado dado debe existir al menos un camino de longitud finita en el subgrafo formado por las UT de dicho mercado sobre las aristas contenidas en la matriz de adyacencia (sin pasar por ninguna UT asignada a un MLT diferente).

#### 1.6.4. Representación de las soluciones

Los cromosomas de los individuos de la población representan soluciones viables, es decir, particiones de  $\mathbf{U}$ , que constituyen una regionalización del territorio en MLT. Como se esbozó en la sección 1.5.1, la manera en que se codifiquen esos cromosomas afecta por diferentes vías al rendimiento general del algoritmo.

Contrariamente a la estrategia utilizada en varias aplicaciones de AEA citadas en el primer capítulo, en este trabajo se ha decidido utilizar una variación de la codificación grupo-número (JONES y BELTRAMO, 1991) (*group-number encoding*), donde cada individuo es representado por un vector de  $N$  componentes, cada uno de los cuales corresponde a una UT de  $\mathbf{U}$  y toma el valor del identificador del mercado al que la UT pertenece (fig. 1.2). Esta representación asegura la ausencia de solapamiento y, por tanto, hace innecesario contemplar esa restricción. Como se explica en FALKENAUER (1998), el estándar de codificación de grupo-número sufre de una redundancia muy alta (ver sección 1.5.1), mientras que la representación AEA sufre de baja redundancia (en el grupo de partes, como se describe más adelante en esta sección). Solucionamos el problema de redundancia recodificando cualquier solución nueva para que los identificadores de mercado mantengan una progresión creciente en el vector de UT, de tal manera que el grupo de identificadores de la primera UT en el vector es siempre 1, la siguiente UT en el vector que no pertenecen al mercado 1 recibe el identificador de grupo 2, el mercado de la siguiente UT en el vector que no pertenezca a un mercado ya numerado recibe el siguiente código de grupo (último empleado más 1). Por

	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$
cromosoma:	1	2	1	3	2	2	1	4	4	2

$$M_1=\{s_1, s_3, s_7\} \quad M_2=\{s_2, s_5, s_6, s_{10}\} \quad M_3=\{s_4\} \quad M_4=\{s_8, s_9\}$$

Figura 1.2: Representación de individuos (cromosoma)

lo tanto, los vectores de 12346 (no consecutivos), 13121 o 32123 (no ordenados) serían recodificados a las representaciones equivalentes 12345, 12131 y 12321 respectivamente. Esta estrategia asegura que las soluciones equivalentes (fenotipo) tienen representaciones iguales (cromosoma) y evita toda redundancia siempre que se mantenga fija la secuencia en que se ordenan las UT en el vector. Esto permite identificar nuevas soluciones ya presentes en la población (individuos repetidos) con facilidad (sin tener que evaluarlos) y asegura que el espacio de búsqueda no es más grande que el espacio de soluciones.

Debe tenerse en cuenta que el máximo identificador de grupo en un cromosoma es también el número de mercados en esa solución. Esto implica que esta representación se transforma fácilmente en la de los AEA (p. ej., 12131 sería 12131|123) y que ambas representaciones obtienen los beneficios de la codificación orientada a grupos. Por otra parte, la representación empleada en este trabajo sufre de una redundancia inferior a la de los AEA, ya que lo que correspondería a la parte de grupos (12131|**123**) en nuestra codificación es siempre única para una partición dada, no siendo así necesariamente en el caso de la representación utilizada en FALKENAUER (1998). Además, la codificación empleada en este trabajo es de longitud fija en su forma básica (el número de UT), lo que simplifica la programación de los operadores (de nuevo algo que no ocurre con la propuesta por Falkenauer). Pero, en la implementación, los valores de algunos mercados (como el número de UT que los forman, el total de flujos de entrada y salida, el valor de la función objetivo contribuido, entre otros) se pueden incluir en la estructura de datos que contiene el cromosoma con el fin de acelerar la ejecución del algoritmo (especialmente la evaluación de las soluciones que se han modificado de manera parcial). Dado que el número de mercados en una solución varía a lo largo de la evolución, mantener esa información extra implicaría una repre-

sentación de longitud variable, pero en la práctica se puede implementar de forma que no afecte a la lógica de los operadores.

Esta codificación es la misma que se empleará en los trabajos presentados en las publicaciones dedicadas a AEA (2-5 y 7).



Universitat d'Alacant  
Universidad de Alicante

## Capítulo 2

# Publicaciones

Como se ha indicado previamente, este capítulo recoge un resumen del planteamiento, experimentación, resultados y conclusiones presentados en cada una de las publicaciones que se compendian en esta tesis. Las publicaciones son las siguientes:

1. Flórez-Revuelta, F.; Casado-Díaz, J.M.; Martínez-Bernabeu, L. (2006) “An evolutive approach for the delineation of local labour markets”, Parallel Problem Solving from Nature - PPSN IX, In: Lecture Notes in Computer Science vol. 4193, pp. 342–351, Springer Berlin / Heidelberg. (Indexado en ranking de congresos CORE [2010] A)
2. Flórez-Revuelta, F.; Casado-Díaz, J.M.; Martínez-Bernabeu, L. (2007) “Specific crossover and mutation operators for a grouping problem based on interaction data in a regional science context”. IEEE Congress on Evolutionary Computation, CEC 2007, pp. 378–385, IEEE. (Indexado en ranking de congresos CORE [2010] A)
3. Flórez-Revuelta, F.; Casado-Díaz, JM; Martínez-Bernabeu, L.; Gómez-Hernández, R. (2008) “A Memetic Algorithm for the Delineation of Local Labour Markets”, Parallel Problem Solving from Nature PPSN X. In: Lecture Notes in Computer Science, vol. 5199, pp. 1011–1020, Springer Berlin / Heidelberg. (Indexado en ranking de congresos CORE [2010] A)

4. Flórez-Revuelta, F.; Casado-Díaz, J.M.; Martínez-Bernabeu, L. (2008) “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, Institute of Automation, Chinese Academy of Sciences, co-published with Springer-Verlag GmbH. (Indexado en Scopus, SNIP [2010] 2.032, SJR [2010] 0.064)
5. Flórez-Revuelta, F.; Casado-Díaz, J.M.; Martínez-Bernabeu, L. (2009) “Adaptive Evolutionary Algorithms for the Delineation of Local Labour Markets”. *IEEE Congress on Evolutionary Computation, CEC 2009*, pp. 2354–2360, IEEE. (Indexado en ranking de congresos CORE [2010] A)
6. Casado, J.M.; Martínez, L. y Flórez, F. (2010) “Los mercados locales de trabajo españoles. Una aplicación del nuevo procedimiento británico”, en Albertos, J.M. y Fera, J.M. (ed.). *La ciudad metropolitana en España: procesos urbanos en los inicios del siglo XXI*, pp. 275–313, Madrid: Thomson-Civitas.
7. Martínez-Bernabeu, L.; Flórez-Revuelta, F.; Casado-Díaz, J.M. (2012) “Grouping genetic operators for the delineation of functional areas based on spatial interaction”, *Expert Systems with Applications*, 39 (8): 6754-6766. doi:10.1016/j.eswa.2011.12.026, Elsevier. (Indexado en JCR [2010]: 1.926; pos. 34/108 [CS-AI])

## 2.1. An evolutive approach for the delineation of local labour markets (PPSN 2006)

Este trabajo es el primero del que se tiene constancia en que se emplea una técnica de optimización estocástica, concretamente un AE, para delimitar MLT y uno de los primeros que abordan la delimitación de MLT como un problema de optimización combinatoria.

Como ya se indicó en el capítulo anterior, la formulación del problema y la representación de las soluciones son comunes a todas las publicaciones dedicadas a AEA y están descritas en la sección 1.6. A continuación se describe la estructura del AEA, el desarrollo de los operadores genéticos y los resultados y conclusiones obtenidos en la experimentación.

### 2.1.1. Estructura del algoritmo evolutivo

1. Generar una población inicial de  $N_p - 1$  soluciones aleatorias, además de la solución trivial (todo el territorio en un único mercado).
2. Evaluar la validez y calidad de las soluciones y ordenarlas.
3. Mientras no haya cambio en la mejor solución durante  $N_i$  iteraciones:
  - 3.1. Generar  $N_r$  nuevas soluciones por recombinación
  - 3.2. Evaluar la validez y calidad de las soluciones.
  - 3.3. Generar  $N_m$  nuevas soluciones por mutación
  - 3.4. Evaluar la validez y calidad de las soluciones.
  - 3.5. Ordenar la población por calidad.
  - 3.6. Seleccionar las  $N_p$  mejores soluciones para formar la población de la siguiente generación.

Figura 2.1: Pseudocódigo del AE (PPSN 2006)

El algoritmo presentado en este trabajo sigue la estructura estándar de los algoritmos genéticos (fig. 2.1).

Dentro de los pasos de generación de nuevas soluciones, la selección de los operadores a aplicar se realiza con probabilidad uniforme y la de los individuos a recombinar/mutar con probabilidad proporcional a la evaluación. Como se indica en el pseudocódigo, la selección de población para la siguiente generación se hace por truncamiento.

### 2.1.2. Operadores genéticos ad hoc

Como se comentó en la sección 1.5.1, en FALKENAUER (1998) se propone emplear un cruce y hasta tres formas de mutación de grupos: creación, destrucción y alteración, todos ellos adaptados al problema concreto a resolver y a la codificación empleada. Se señaló, además, que las aplicaciones de esos métodos a problemas de agrupación, análogos en mayor o menor medida al que nos ocupa, utilizaban en su mayoría uno o dos operadores (en algunos

casos integrando en una sola función varias de las operaciones básicas propuestas), siendo suficientes para alcanzar buenos resultados en el problema concreto en comparación con otros métodos del estado del arte respectivo.

En el problema de delimitación de MLT, sin embargo, las fuertes restricciones hacen que el número de posibles asignaciones no válidas de un elemento sea muy elevado, cuando el territorio a delimitar es lo bastante extenso, y provoca que los operadores genéticos estándar no ofrezcan resultados válidos. Esto es fácil de entender con el siguiente ejemplo: supongamos que estamos regionalizando España, con sus más de 8000 municipios; si se escoge un municipio de la provincia de Alicante y se asigna aleatoriamente a cualquier mercado presente en la solución es muy probable que acabe en un mercado con el que no guarda relación alguna (exactamente lo que se pretende evitar en una delimitación funcional), dado que la mayoría de mercados estarán demasiado lejos como para que exista un flujo cotidiano de trabajadores entre ambos. Es por esto por lo que parece conveniente no sólo la diversificación de operaciones posibles, sino la introducción de conocimiento del problema en los operadores.

El desarrollo de los operadores de ese trabajo siguió un esquema de abajo arriba, introduciendo primero pocos operadores, estudiando al detalle la evolución de las soluciones y los resultados finales para identificar puntos de mejora, e introduciendo después nuevos operadores o modificaciones y nuevas versiones de los operadores anteriores. Nos referimos a esos operadores en este trabajo como operadores ad hoc, debido a la forma en que fueron desarrollados. Siguiendo este procedimiento se alcanzó un número relativamente alto de operadores: se incluyeron tres operadores de cruce y diez operaciones de mutación.

Dos de estos operadores están directamente basados en los dos operadores estándar en los AG (*1-point crossover* y mutación aleatoria de alelos) y los demás implementan operaciones más complejas o informadas. En concreto se desarrollan dos formas de cruce que pretenden mitigar los efectos de insensibilidad al contexto y perturbación de esquemas (1.5.1) y nueve operadores de mutación que implementan las siguientes operaciones: reasignación óptima de UT elegidas al azar (óptimo según el índice de interacción), unión al azar de dos mercados, disgregación de un mercado y reasignación óptima

de sus UT, división de un mercado mediante la selección de dos semillas y la asignación óptima del resto de sus UT a esas dos semillas, división de un mercado mediante la extracción al azar de UT hasta formar un mercado válido, extracción de las UT de un mercado que peor índice de interacción presentan y su reasignación óptima a otros mercados, otro operador similar al anterior en que la reasignación se hace por bloques de UT, una operación de intercambio simultáneo de dos UT entre dos mercados y, por último, un operador de fusión y partición de varios grupos simultáneamente (se selecciona al azar un mercado y un número  $k$  de los mercados más relacionados con el primero, se desintegran todos, se seleccionan  $k+1$  semillas de entre esas UT y se reasignan de forma óptima en orden azaroso el resto de UT). La descripción completa de estos operadores se puede consultar en la propia publicación, en el apéndice A.

### 2.1.3. Experimentación, resultados y conclusiones

Para su validación el método se aplicó a la delimitación de los MLT en la Comunidad Valenciana. Este caso de estudio se describe con detalle en la publicación, en el apéndice A.

En resumen, el método presentado en esta publicación consigue mejorar ampliamente, en términos de autonomía alcanzada y número de mercados delimitados, los resultados obtenidos con los métodos más usados hasta el momento, como es el método de las TTWA. Además, y a pesar del carácter no determinista del método, consigue delimitar mercados con muy pocas discontinuidades (una cantidad similar a la que produce el método de las TTWA) sin emplear ninguna información de contigüidad geográfica.

Como desventajas del método, por un lado, queda por resolver la reticencia de los expertos en economía y geografía a aceptar métodos no deterministas, para lo que se hace necesario desarrollar técnicas que reduzcan en lo posible la variabilidad de los resultados, y, por otro lado y especialmente, el alto coste computacional del método, que impide aplicarlo a casos de estudio de un tamaño muy superior, que son los más frecuentes. Esto se conseguirá solventar en posteriores trabajos (véase la sección 2.7).

Como principales aportaciones de este trabajo cabe resaltar la aproximación al problema desde un enfoque multidisciplinar, llevando técnicas de

inteligencia artificial al campo de la delimitación de regiones funcionales, así como el desarrollo de una formulación para abordar el problema como uno de optimización, lo que permitió mejorar la calidad de las delimitaciones producidas y, con ello, la fiabilidad de los análisis económicos que se realicen sobre las mismas.

## **2.2. Specific crossover and mutation operators for a grouping problem based on interaction data in a regional science context (CEC 2007)**

En este trabajo se continúa con lo realizado en el descrito en la sección anterior. En primer lugar, se desarrollan nuevos operadores que se añaden a los anteriores y, en segundo lugar, se analiza su comportamiento a través de una serie de indicadores que puedan ayudar a mejorar la configuración del algoritmo, con el fin de aumentar su rendimiento para poder resolver problemas de mayor dimensión.

La codificación de las soluciones y la formulación del problema son las mismas que ya se han descrito en las secciones 1.6.4 y 1.6.1. La estructura del algoritmo es también la misma que en la publicación anterior (véase sección 2.1.1).

### **2.2.1. Nuevos operadores**

Se presentan dos nuevos operadores de cruce, que reordenan las celdas del cromosoma o los identificadores de grupo antes de realizar la operación de cruce, ideados para descartar la influencia que tenga en la operación la ordenación inicial de UT en el cromosoma (de forma que se reducen más los efectos de la insensibilidad al contexto que presentan los operadores empleados anteriormente).

También se presentan dos nuevos operadores de mutación, uno similar al de fusión y partición simultánea en el que cambia la forma en que se asignan las UT restantes tras elegir las semillas (en este caso, cada semilla

absorbe las UT remanentes en orden de mayor a menor interacción hasta que se forma un mercado válido antes de pasar a completar el mercado de la siguiente semilla) y una nueva operación de división en la que dos semillas alternadamente van escogiendo la UT de mayor interacción de entre las UT restantes .

### 2.2.2. Comportamiento de los operadores

Para caracterizar el comportamiento de los 17 operadores del algoritmo, se contabilizó la cantidad de apariciones en el historial del mejor individuo, el tiempo de cómputo consumido por cada operador y el incremento medio en la evaluación del individuo generado en comparación con el padre (no contando en el último indicador con los casos en que la diferencia es negativa o el resultado inválido).

### 2.2.3. Experimentación, resultados y conclusiones

Para la experimentación en este trabajo también se recurrió al caso de la Comunidad Valenciana. Los resultados fueron ligeramente superiores en calidad a los obtenidos con el primer conjunto de operadores.

Entre las aportaciones de este trabajo se encuentran los nuevos operadores desarrollados y el estudio del comportamiento de todos ellos, que muestra una enorme variabilidad en el grado de éxito de cada operador a lo largo de la evolución (tablas II a IV y fig. 7 del apéndice B, pp. 383 y 384 del artículo original). La fig. 8 en el apéndice muestra la evolución de la calidad media entre ejecuciones y del mejor individuo obtenido de los primeros 20 minutos de ejecución (el tiempo medio de finalización estaba cercano a 40 minutos, aunque variaba entre 25 minutos y 1 hora) y se puede observar como en apenas 2 minutos se alcanza un valor de interacción notablemente superior al del método de las TTWA.

### **2.3. An evolutionary approach to the delineation of functional areas based on travel-to-work flows (IJAC 2008)**

Esta publicación prosigue el trabajo desarrollado en las dos anteriores, continuándolo por otras vías diferentes del desarrollo de operadores, en concreto, se analiza el comportamiento de los operadores a lo largo del proceso de evolución y la estabilidad de las soluciones obtenidas en diferentes ejecuciones. Además, se explora la opción de aplicar el algoritmo evolutivo como una herramienta de optimización local de soluciones producidas por otros métodos.

Además, en esta publicación se comenzó a utilizar la restricción de vecindad funcional descrita al principio de la sección 1.6.3.2. El algoritmo y los operadores son los mismos que aparecen en la anterior publicación y no se describen aquí.

#### **2.3.1. Comportamiento con otras funciones y parámetros**

En primer lugar, se comparan los resultados de una segunda función objetivo, en la que la interacción acumulada no se multiplica por el número de mercados identificados. Como era de esperar, el número de mercados identificados es menor, mientras que la interacción acumulada es mayor. En consecuencia, los mercados son más autónomos, aunque no necesariamente menos cohesionados.

También se analiza el comportamiento del método al cambiar los parámetros estándar del problema de delimitación (los valores mínimos de autonomía y tamaño), eliminando el parámetro de tamaño mínimo y comparando esos resultados con los que se obtienen con el método de las TTWA.

Para caracterizar el comportamiento de los operadores se calculan los indicadores que se emplearon en la publicación anterior.

### 2.3.2. AEA como optimizador local

En este trabajo se propone la utilización del AEA como herramienta de optimización local de los resultados obtenidos por otros métodos, de forma que no se varíe el número de mercados de partida, en este caso aplicándolo a la delimitación de la Comunidad Valenciana realizada por el método de las TTWA.

### 2.3.3. Análisis de estabilidad

Por último, para analizar el grado de indeterminismo de las soluciones producidas, en términos de estabilidad en su asignación a un mercado entre diferentes ejecuciones, se calcula la frecuencia de asignación de cada UT a la UT cabecera (la de mayor tamaño del mercado) más frecuente.

### 2.3.4. Experimentación, resultados y conclusiones

Para facilitar la comparación, el caso de estudio en la experimentación de esta publicación vuelve a ser la Comunidad Valenciana. Los resultados ofrecidos por el método con diferentes parámetros y función objetivo siguen mejorando notablemente los resultados obtenidos con el método de las TTWA.

La optimización local aplicada sobre la delimitación de las TTWA también mejora su evaluación, pero la limitación, intencionada, del número de mercados no permite acercarse a los valores alcanzados por la delimitación no restringida del algoritmo evolutivo. Eso refuerza la idea de que el número de mercados identificados por ese método no es el más adecuado.

En cuanto a la estabilidad de los resultados, se comprueba que la mayor parte de las UT que constituyen el territorio analizado, así como la mayoría de la población activa en las mismas, es asignada en la mayoría de casos siempre a la misma cabecera, lo que indica una baja variabilidad de los resultados, aunque los cambios en las fronteras entre diferentes ejecuciones pueden parecer a simple vista de mayor calado que lo observado

al contabilizar la población presente en los municipios más “inestables”, de poca población en la mayoría de los casos. Los histogramas de frecuencias que se muestran en las fig. 9 y 10 en el apéndice C (p. 18 de la publicación original) ilustran estos datos.

## 2.4. A Memetic Algorithm for the Delineation of Local Labour Markets (PPSN 2008)

Este trabajo supone la primera tentativa de aplicación de algoritmos meméticos a la delimitación de MLT.

Existían dos motivaciones iniciales para incluir este tipo de rutinas en el algoritmo. La primera era reducir la variabilidad de los resultados, dado que la rutina de optimización tiene una fuerte componente de exploración local. La segunda es reducir el número de soluciones inválidas que producían los operadores, con el objetivo de aumentar el rendimiento global del algoritmo, por un lado, aprovechando la información valiosa que puedan contener las soluciones inválidas y, por otro, reduciendo el tiempo dedicado a producir soluciones inválidas.

De nuevo, los operadores genéticos, la codificación de las soluciones empleada y las funciones objetivo son las mismas que en la anterior publicación, por lo que no se describirán aquí. La restricción de vecindad funcional también se aplicó en este trabajo (véase la sección 1.6.3.2).

### 2.4.1. Algoritmo memético

Partiendo del trabajo publicado en FLÓREZ-REVUELTA *et al.* (2008), se añade una heurística de reparación de individuos inválidos y otra de optimización local. La primera se basa en la desintegración de mercados inválidos escogidos al azar y la redistribución óptima de sus UT, hasta que todos los mercados son válidos. La rutina de optimización selecciona mercados al azar para reasignar de forma óptima una de sus UT, aceptando el cambio si mejora la evaluación previa. Pueden consultarse las descripciones completas en el apéndice D (pp. 1016 y 1017 en la publicación original).

1. Generar al azar una población inicial de  $N_p - 1$  soluciones, además de la solución trivial (todo el territorio en un único mercado).
2. Evaluar la validez y calidad de las soluciones y ordenarlas.
3. Mientras no haya cambio en la mejor solución durante  $N_i$  iteraciones:
  - 3.1. Generar  $N_r$  nuevas soluciones por recombinación.
  - 3.2. Evaluar la validez de las soluciones, aplicar la rutina de reparación a las inválidas, aplicar la rutina de optimización local y evaluar su calidad final.
  - 3.3. Generar  $N_m$  nuevas soluciones por mutación
  - 3.4. Evaluar la validez de las soluciones, aplicar la rutina de reparación a las inválidas, aplicar la rutina de optimización local y evaluar su calidad final.
  - 3.5. Ordenar la población por calidad.
  - 3.6. Seleccionar las  $N_p$  mejores soluciones para formar la población de la siguiente generación.

Figura 2.2: Pseudocódigo del AE (PPSN 2006)

Con la incorporación de estas dos subrutinas se modifica ligeramente la estructura del algoritmo, quedando como se describe en la fig. 2.2.

Los mecanismos de selección de los operadores a aplicar, de los individuos con los que operar y de los que formarán la generación siguiente son los mismos que en las publicaciones anteriores.

#### 2.4.2. Experimentación, resultados y conclusiones

A pesar del potencial aumento de eficiencia y eficacia que se espera de estas técnicas, los resultados sólo consiguen mejorar ligeramente los mejores resultados de la versión inicial, no memética, y como contrapartida sufre un alto aumento del coste computacional (tanto si se emplea sólo la reparación, sólo la optimización o ambas). Esto no implica que reparar individuos o realizar optimizaciones locales sean una mala estrategia para el PDAF, dado

que los resultados pueden ser debidos a una implementación ineficiente de las subrutinas (en especial, las llamadas a reparaciones dentro de la optimización local suponen un alto coste computacional que, al parecer, no se justifica), a un planteamiento inadecuado para los operadores ya presentes o, más probablemente, dado que se aprecia un alto número de soluciones repetidas, a que se requiera un ajuste más cuidadoso de los parámetros (de forma que, p. ej., se aplique menos veces o en momentos más adecuados, o detecte antes cuando no está siendo eficiente). Ello hace que continuar con esta línea de investigación sea una de las propuestas incluidas en el apartado de conclusiones de esta tesis.

## 2.5. Adaptive Evolutionary Algorithms for the Delineation of Local Labour Markets (CEC 2009)

El ajuste dinámico de los parámetros ha sido ampliamente utilizado cuando se requería configurar un gran número de parámetros para evitar las desventajas de utilizar valores estáticos preestablecidos (LOBO *et al.*, 2007). Este trabajo explora esa línea de mejora del rendimiento del algoritmo evolutivo, la de la adaptación y autoadaptación de las probabilidades de aplicación de los diferentes operadores genéticos. La primera alternativa, adaptación, consiste en la variación de las probabilidades de aplicación de los operadores (las mismas para todos los individuos) en función del desempeño de los propios operadores. En la segunda, autoadaptación, las probabilidades de aplicación de los operadores se incluyen dentro de los cromosomas, de forma que individuos distintos pueden tener probabilidades de aplicación de los operadores diferentes.

Otro cambio destacable en esta publicación con respecto a las anteriores se produce en la estructura del algoritmo evolutivo: ya no hay dos fases independientes en las que se aplica, en primer lugar, una cantidad determinada de cruces y, después, otra de mutaciones. En su lugar se incluye una única fase que puede aplicar por igual cruces y mutaciones. Además, como condición de terminación se emplea un límite de tiempo fijo en lugar de una cantidad de generaciones consecutivas sin mejorar la mejor solución.

También es conveniente resaltar que los parámetros del algoritmo evolutivo (en este caso número de padres en la población inicial y número de nuevos individuos) se reducen mucho con respecto a las publicaciones anteriores: la población inicial tiene sólo diez padres (frente a los cien usados con anterioridad) y se crea un único nuevo individuo en cada generación (en lugar de los anteriores 170, 10 por cada operador).

La codificación de soluciones, funciones objetivo y restricciones son las de la anterior publicación.

### 2.5.1. Propuestas adaptativas y autoadaptativa

Prestando atención a la alta variabilidad en el éxito de los operadores (en términos de su porcentaje de aparición en el historial del mejor individuo, la interacción aportada con cada aplicación o su coste computacional), resulta adecuado tratar de aplicar más veces los operadores que más mejora aportan y hacerlo más en las etapas de la evolución en que mejores resultados ofrecen. Por ejemplo, por la experiencia acumulada en los anteriores trabajos, sabemos que las operaciones de división van perdiendo eficiencia a medida que el número de mercados en las soluciones es mayor. Así mismo, no resulta útil aplicar uniones en las primeras etapas. Estas sencillas reglas no lo son tanto para el resto de operadores, que muestran un comportamiento más irregular, no sólo entre ejecuciones con distintos parámetros o territorios, debido a la compleja interacción que ocurre entre los efectos de cada operador (la presencia de cada operador influye indirectamente en la eficiencia de los demás). No resulta sencillo, por tanto, establecer las reglas que permitirían ajustar mejor esas probabilidades. Por lo tanto se hace necesaria una manera automática de ajustar dinámicamente las probabilidades de aplicación.

En este artículo se presentan tres alternativas de ajuste dinámico de los parámetros, descritas resumidamente a continuación:

1. **Autoadaptación con recompensa fija:** Los individuos generados heredan el vector de probabilidades de su padre (el del mejor de ellos

en caso de que sea una recombinación). La probabilidad de aplicación del operador que generó el individuo se incrementa en función del puesto que alcance el individuo generado en la población, disminuyendo proporcionalmente la de los demás. Tras esto, para evitar que la probabilidad de algún operador descienda hasta 0 y no vuelva a ser utilizado, se perturban todas las probabilidades con ruido blanco.

2. **Adaptación con recompensa según la mejora acumulada:** En esta alternativa se incluye un vector global con una celda por cada operador. En cada celda se acumulan los incrementos de evaluación sobre el (mejor) padre. Las probabilidades de aplicación de los operadores son actualizadas periódicamente en función de los valores de ese vector, que a su vez son reducidas proporcionalmente con el transcurso de las generaciones para evitar que los operadores que acumularon más incremento de la evaluación en etapas anteriores pero que ya no están siendo rentables sigan recibiendo atención. También en este caso se aplica ruido blanco al vector de probabilidades tras cada generación.
3. **Adaptación con recompensa según la mejora sobre el tiempo:** Esta alternativa es equivalente a la anterior cambiando únicamente la forma en que se incrementa el vector de mejora acumulada: en este caso se añaden el cociente entre el incremento de evaluación y el tiempo requerido por el operador. De este modo, se premia a los operadores que presentan una mayor tasa temporal de mejora de las soluciones.

En los tres casos, las probabilidades iniciales son fijadas aleatoriamente en la población de partida. Para una descripción detallada de estas técnicas véase el apéndice E, p. 2358 en la publicación original.

### 2.5.2. Experimentación, resultados y conclusiones

Para analizar el éxito de esta propuesta se realizó una serie de pruebas comparando los resultados ofrecidos por el algoritmo de probabilidades uniformes con los obtenidos al aplicar cada una de las tres técnicas de adaptación presentadas.

Los resultados muestran que cualquiera de esas tres alternativas aumenta el valor medio de las evaluaciones de las soluciones (lo que indica un

descenso de la variabilidad en los resultados y una mejora de los resultados en general), pero al mismo tiempo ninguna de las tres alcanza el valor máximo ofrecido por el algoritmo evolutivo con probabilidades uniformes. Al comparar entre las tres técnicas de ajuste dinámico, la que ofrece un mejor valor medio es la tercera (recompensa según mejora sobre tiempo) y la que da un mejor valor máximo (tanto en evaluación como en número de mercados) es la segunda (recompensa según mejora). En cualquier caso, ninguna de las cuatro versiones presenta una ventaja estadísticamente significativa sobre las demás.

Las técnicas de ajuste dinámico sí consiguen acelerar la evolución global al compararla con la técnica con probabilidades fijas, como se aprecia en la fig. 4 del apéndice E, p. 2359 de la publicación original.

Con la aplicación de estas técnicas se esperaba conseguir más aumento de la eficiencia global del algoritmo de la obtenida. Una posible explicación de que esto no ocurriera se halla en la cantidad y variedad de los operadores presentes en el algoritmo: ninguna de las medidas de rendimiento probadas permite una estimación del éxito de los operadores de forma supeditada a las relaciones entre los operadores o, en otras palabras, el hecho de que un operador sea más efectivo que otros en un momento dado no depende sólo de él mismo sino también de los efectos previos introducidos por los demás operadores.

Como principal aporte de esta publicación se tiene la incorporación al algoritmo de unas técnicas que permiten un ajuste de los parámetros de los operadores que no requieren conocimiento del problema en cuestión ni del caso concreto de experimentación, además de la aceleración del proceso de búsqueda en sus etapas iniciales.

## **2.6. Los mercados locales de trabajo españoles. Una aplicación del nuevo procedimiento británico (Thomson-Civitas 2010)**

El trabajo presentado en esta publicación tenía como objeto la divulgación de algunos de los resultados obtenidos en el proyecto ref. SEJ2007-

67767-C04-02 del Plan Nacional de I+D+i financiado por el Ministerio de Ciencia e Innovación y FEDER.

En este caso no se emplean algoritmos evolutivos, sino la versión más reciente del método de la TTWA (ONS *et al.*, 2008), reseñado en el capítulo 1, para delimitar los MLT en España. A tal efecto se presenta un análisis exploratorio de los parámetros (umbrales mínimos de autonomía y tamaño) más adecuados para la delimitación de MLT en España, dado que no existen valores “naturales” para esos parámetros y que valores distintos pueden ser más apropiados para determinados territorios o usos de las delimitaciones.

### 2.6.1. Resolución de discontinuidades

Como se señaló en el capítulo 1, la contigüidad de los mercados identificados es uno de los requerimientos del problema, que en métodos como el de las TTWA se resuelve manualmente en una etapa posterior para no constreñir el análisis ofrecido por el método. Para asegurar que las delimitaciones producto del método de las TTWA cumplan la restricción de continuidad, se presenta un procedimiento de resolución y optimización voraz de discontinuidades, similar en su estructura al método ERA descrito en la sección 1.3.1.

### 2.6.2. Evaluación de calidad

Uno de los aspectos menos desarrollados en la literatura sobre la delimitación de AF es el de la cuantificación de la medida en que resultan apropiadas (CASADO-DÍAZ y COOMBES, 2005). En este trabajo se presentan esquemáticamente varias alternativas para abordar este asunto (aparte de las que ya se estaban utilizando previamente, el número de mercados y la autonomía global):

1. autonomía mínima de oferta y de demanda de cada municipio en su mercado, presente en las restricciones del problema;

2. autonomía local intramercado, calculada como la proporción de la suma de oferta y demanda sobre el total agregado, que sería una alternativa a la medida anterior;
3. cociente entre la suma de la demanda y la oferta externa e interna (sin considerar los flujos reflexivos), para tomar en consideración explícita el grado de relación con el exterior de un municipio con respecto a la relación con el mercado al que ha sido asignado; y por último
4. un índice de bondad de asignación similar al utilizado por WATTS (2009) para detectar UT que no están asignadas de forma óptima en la delimitación actual, que a su vez es una variación del índice presentado por FENG (2008).

Sobre el último indicador se construyó otro a nivel de mercado, el porcentaje de UT que están asignadas a su mercado óptimo, al siguiente o a los demás (véase tabla 3 del apéndice F, p. 301 de la publicación original).

El desarrollo de estos indicadores permite generar una batería de medidas que pueden ser muy útiles a la hora de evaluar cualquier procedimiento de delimitación de MLT. Esto incluye, lógicamente, los AEA que constituyen el núcleo de esta tesis. Se trata, además, de indicadores que podrían dar lugar a nuevas funciones objetivo que respondieran de manera más adecuada a la naturaleza del problema abordado.

### 2.6.3. Experimentación, resultados y conclusiones

Los resultados del análisis exploratorio ponen de manifiesto que las fronteras administrativas municipales no son representativas de un MLT en la mayoría de casos debido a que no alcanzan unos mínimos de autonomía, tamaño o ambos. Las discrepancias entre los municipios españoles en términos de superficie y densidad de población son enormes y mucho mayores que las observadas en las unidades geográficas de base que se emplean en otros países, especialmente el Reino Unido. Eso complica la elección de unos parámetros que produzcan una delimitación en la que se consiga mantener independientes los municipios o grupos de los mismos que sean lo bastante

autónomos y, a la vez, dé lugar a mercados cuyo rango de tamaños permita la movilidad cotidiana. Por todo ello se presentan finalmente cuatro alternativas de delimitación de MLT para España en función de los correspondientes juegos de parámetros.

El método de resolución de discontinuidades permitió generar delimitaciones continuas, que cumplieran todas las restricciones del problema, siguiendo un procedimiento formalizado y determinista que no requiere de intervención manual. Sin embargo, el estudio de los resultados de su aplicación a numerosas delimitaciones con discontinuidades mostró que en una gran parte de los casos los resultados finales diferían notablemente de los mapas iniciales con discontinuidades. Esto llevó a la conclusión de que la introducción de la restricción de contigüidad a lo largo del proceso de delimitación podía ser conveniente, dado que en cualquier caso la resolución de las discontinuidades en una etapa final alteraba en gran medida el resultado del análisis previo. Esta conclusión permitió mejorar el diseño del AEA en el sentido que se describe en la sección 2.7.

La análisis de los indicadores de calidad propuestos y su visualización sobre las delimitaciones geográficas ayudan a caracterizarlas y a identificar gráficamente municipios fronterizos conflictivos y mercados cuya existencia pueda ser fruto de un mal ajuste de los parámetros. Este análisis pone de manifiesto que, por un lado, es muy probable que el algoritmo utilizado no sea capaz de cumplir completamente su objetivo (“delimitar tantos MLT como sea posible sujetos a los criterios estadísticos establecidos”, ONS *et al.*, 2008, p. 8), a pesar de ser uno de los más elaborados de entre los aplicados para la delimitación de AF de uso oficial. Esta conclusión se apoya en los resultados obtenidos con los algoritmos evolutivos recogidos en las publicaciones anteriores. Por otro, en última instancia cualquiera de estos procedimientos de delimitación se enfrenta al hecho de que el problema a resolver es muy complejo, con objetivos en conflicto y sin soluciones triviales, situación que se agrava con el creciente grado de interrelación que se produce hoy en día entre UT cercanas, de manera especialmente nítida en las áreas de carácter más urbano.

## 2.7. Grouping genetic operators for the delineation of functional areas based on spatial interaction (ESWA 2012)

A pesar de los avances logrados y presentados en las publicaciones anteriores referidas a algoritmos evolutivos y operadores de agrupación, aún no se había logrado el rendimiento suficiente como para resolver problemas de tamaño más grande, como era el caso de España, sin tener que recurrir a técnicas subóptimas.

El desarrollo de los operadores de estos trabajos siguió un esquema de abajo arriba en el que se introducían primero pocos operadores, se estudiaba al detalle la evolución de las soluciones y los resultados finales para identificar puntos de mejora, y se incorporaban después nuevos operadores o modificaciones y nuevas versiones de los operadores anteriores. Siguiendo este procedimiento se alcanzó un número relativamente alto de operadores (5 cruces y 12 mutaciones, ya resumidos en las secciones anteriores) Tras esto se procedió a tratar de eliminar los operadores que menos aportasen al proceso de optimización, pero la complejidad de las relaciones entre los diferentes operadores implicaba que eliminar cualquiera de ellos podía (según el problema concreto y la función objetivo empleada) empeorar los resultados, por lo que resultaba mejor mantener todos los operadores en el método seleccionándolos con probabilidad uniforme o aplicando una técnica adaptativa.

Con esa motivación y empleando todo el conocimiento acumulado durante las investigaciones y desarrollos previos, se decidió en el trabajo de esta publicación reconsiderar todas las implicaciones de las restricciones del problema (tanto las que ya se consideraban en el trabajo inicial como las que otros autores descartaban en sus aplicaciones de algoritmos de *clustering*, especialmente la restricción de contigüidad y sus efectos sobre la complejidad de los operadores) y del uso de las mutaciones completamente azarosas, óptimas y estocásticas, para diseñar un nuevo conjunto de operadores desde cero y de arriba abajo.

Así, como ya se señaló en la introducción (sección 1.5.2), se siguió la premisa de que todos los operadores fueran estocásticos, descartando tanto

el pleno azar como la asignación óptima según algún criterio (como ya se ha explicado a lo largo del primer capítulo, no existe un criterio concreto que pueda guiar el proceso de búsqueda con seguridad y en cualquier situación hacia la solución óptima) y que mantuviesen el cumplimiento de las restricciones del problema durante toda su operación salvo en los casos en que no fuera posible, para reducir el número de individuos inválidos.

La estructura del algoritmo es la misma que en la anterior publicación referida a algoritmos evolutivos (CEC 2009), en donde no se hace distinción entre cruces y mutaciones, con la única diferencia de que se redujo la presión de selección de la población siguiente, cambiando la estrategia de truncado por la de selección proporcional al *ranking* con elitismo para el mejor. De esta forma se favorece la exploración global frente a la explotación local.

### 2.7.1. Aglomeración jerárquica estocástica

Para la creación de la población inicial —una colección de puntos distribuidos al azar para comenzar el proceso de búsqueda— se utiliza un algoritmo de aglomeración jerárquica estocástica (AJE) similar a los procedimientos de *clustering* jerárquico reseñados en la introducción (sección 1.3), con la diferencia de que éstos son deterministas y no estocásticos. Este procedimiento de aglomeración, partiendo de una determinada partición de un conjunto (p. ej., la máxima división), escoge iterativamente un grupo de entre los inválidos (por torneo a tres sobre grado de invalidez) y lo combina con otro escogido entre los adyacentes a éste que tengan interacción no nula (con probabilidad proporcional a la interacción), hasta que todos los mercados cumplan con las restricciones del problema. También se utiliza como una subrutina en tres de los operadores de mutación, así como en el procedimiento de reparación en el operador de cruce (véase la descripción completa en el apéndice G, sección 4.2).

### 2.7.2. Operadores genéticos de agrupación

En esta sección se incluye, junto con un breve resumen de los operadores desarrollados, la argumentación que fundamenta este desarrollo. Como

se ve discute en las conclusiones, el resultado fue un gran avance en esta investigación.

En la mayoría de las aplicaciones de AEA citadas en la sección 1.4 se emplea un único operador de mutación, que normalmente se complementa con alguna heurística reparadora, y se reportan resultados que mejoran o igualan los generados por otras técnicas de optimización aplicadas al problema en cuestión. En la experimentación con los operadores ad hoc para el problema de la delimitación de MLT desarrollados en las publicaciones previas, que no era rigurosa ni completa debido a la limitación de tiempo de ejecución, se encontró que la eliminación de cualquiera de los doce operadores de mutación no daba como resultado un aumento significativo del rendimiento y que en la mayoría de casos se producía una reducción de la calidad de las soluciones o un aumento del coste computacional requerido por el algoritmo para alcanzar la misma calidad. En cualquier caso, esto contrasta con la propuesta de Falkenauer de utilizar sólo tres operadores generales, así como con las aplicaciones concretas mencionadas, en las que una única operación de conjuntos es suficiente para explorar y explotar todo el espacio de búsqueda. Por otra parte, no podemos suponer a priori que los operadores de esos trabajos fueran suficientes en el problema que se trata en esta tesis.

Finalmente, se ha tenido en cuenta esta experiencia durante el desarrollo del nuevo conjunto de operadores de mutación y se ha optado por tratar de cubrir desde un principio todas las posibles operaciones básicas a realizar sobre un conjunto de grupos disjuntos (que son considerados como los operandos), con el objetivo de asegurar que el algoritmo podrá realizar una exploración y explotación cuando menos efectiva en un tiempo razonable para cualquier posible configuración del problema.

De este modo, se han diseñado nueve operadores de mutación, contemplando operaciones unarias (como pueden ser la división de un grupo), binarias (como la unión de dos grupos o el intercambio de elementos entre ellos) y n-arias y, dentro de éstas, relaciones uno a muchos (como que un grupo reparta, o absorba, elementos entre sus grupos vecinos) y muchos a muchos (reasignación de elementos escogidos al azar en diferentes grupos). A continuación se muestra la lista completa de operadores desarrollados:

1. partición de un grupo (creación de nuevos grupos a través de la división de un grupo) mediante una heurística voraz (véase sección 2.7.1), lo que puede conducir a una alteración de los grupos existentes o también a la creación de otros nuevos;
2. unión de dos grupos contiguos y relacionados;
3. partición a través de una heurística voraz (véase sección 2.7.1) de la unión de dos grupos relacionados (modifica dos grupos y se pueden crear nuevos grupos);
4. creación de un nuevo grupo, tomando una semilla tomada al azar de entre las fronteras en un mercado también al azar y asignando elementos desde grupos adyacentes hasta que se cumplen las restricciones o no es posible continuar (crea un nuevo grupo y modifica al menos otro más);
5. desmembramiento de un grupo y asignación de sus elementos a los grupos adyacentes relacionados (elimina un grupo y modifica al menos uno más);
6. exclusión de una serie de elementos de un grupo y reasignación de éstos entre los grupos adyacentes relacionados (modifica dos o más grupos);
7. inclusión en un grupo de elementos adyacentes relacionados, operación inversa a la expulsión (modifica dos o más grupos);
8. segregación de un subconjunto de elementos de un grupo y reasignación de éstos a un grupo adyacente relacionado (modifica dos grupos);
9. anexión, por parte de un grupo, de una serie de elementos adyacentes relacionados pertenecientes a otro grupo, operación inversa a la segregación (modifica dos grupos);
10. reasignación de elementos fronterizos escogidos al azar a grupos adyacentes relacionados (modifica dos o más grupos); e
11. intercambio de elementos entre dos grupos adyacentes relacionados (modifica dos grupos).

Las descripciones detalladas de estos operadores pueden consultarse en la sección 4.5 del apéndice G (pág. 6760–6763 en la numeración original).

Como operador unario sólo se ha considerado la división (toma un grupo y trata de dividirlo en dos o más). Como operadores binarios se han contemplado la unión (se descarta la unión n-aria), la anexión/segregación y el intercambio simultáneo entre dos grupos. En las relaciones uno a muchos se considera lo que hemos llamado inclusión y exclusión (a un grupo se le asignan o extraen elementos de forma individual desde o hacia otros grupos), la destrucción de un grupo (desintegrar un grupo y reasignar sus elementos a los grupos que permanecen) y la creación de un nuevo grupo (tomando elementos de otros grupos hasta que conformen uno válido). Y, por último, como operación de muchos a muchos consideramos la reasignación de múltiples elementos de forma individual, tomados de y asignados a diferentes grupos. Esto hace un total de diez operadores de mutación. Además, se consideró la combinación de la unión y la división en una sola operación (binaria) en la que se escogen dos grupos para unir y se aplica sobre el grupo unido la división.

En la aplicación de alguno de estos operadores puede ocurrir que el resultado sea inválido, debido a las fuertes restricciones del problema (p. ej., mover un elemento de un grupo a otro podría violar la condición de contigüidad o de autonomía y tamaño mínimos de uno o ambos grupos implicados). Como ya se señaló en la sección 1.5.1, en estos casos se sugiere aplicar una heurística que repare los grupos inválidos, aunque los resultados de esta técnica en uno de nuestros trabajos previos (PPSN 2008, apéndice D) no justificaban su aplicación.

Para reducir la cantidad de operaciones infructuosas y, con ello, acelerar la convergencia al óptimo, se ha procurado en todas las mutaciones que nunca se realicen cambios que violen alguna restricción fuerte —p. ej., el operador de unión nunca unirá grupos no adyacentes, dado que la restricción de la contigüidad no se respetaría excepto con el uso de disruptivos operadores de reparación que muy probablemente devolverán resultados peores que los previos a la unión—, a menos que sea posible resolver la violación con poco esfuerzo computacional antes de terminar la operación o si los posibles beneficios que traería la proporción de soluciones válidas compensan el tiempo perdido en calcular las soluciones inválidas —p. ej., la creación de un nuevo grupo a partir de un elemento actuando como semilla que absorbe elementos adyacentes de otros grupos implicará hacer algunos pasos

a través de soluciones no válidas (aquellas en que el grupo creciente aún no ha alcanzado el tamaño mínimo) antes de llegar a una válida (y probablemente mejor); aunque también podría llevar a soluciones inviables si el proceso queda atrapado en un punto en que no se puede completar el grupo creciente sin convertir en inválido alguno de los grupos colindantes que le están cediendo elementos—. Por lo tanto, no todos los cambios no válidos están prohibidos, ya que algunas operaciones “arriesgadas” pueden conducir a soluciones válidas muy útiles para el proceso de optimización a pesar de producir individuos inválidos en un alto porcentaje de sus aplicaciones.

La mutación de reasignación múltiple es el operador más parecido a la mutación estándar en algoritmos genéticos, en que se seleccionan elementos (celdas del vector) al azar y se cambian sus valores al azar, sin considerar relación alguna entre esos elementos (como la pertenencia a un mismo grupo o la existencia o no de flujos que los unan). Es también, a simple vista, la mutación más exploratoria, ya que puede afectar a todos los elementos en una misma operación; también a muy pocos, incluso sólo a uno. Se podría argüir, entonces, que esa operación sería suficiente para emular, tras muchas aplicaciones, a cualquiera de los demás operadores propuestos. Sin embargo, la fase de experimentación mostró que éste no era el caso: el operador de mutación estándar en este problema es extremadamente ineficiente, presumiblemente debido a que la casuística del problema se torna intratable cuando el número de elementos (UT) asciende lo suficiente de tal forma que, al considerar todas las posibles reasignaciones de cada elemento sin restricción alguna, la proporción de reasignaciones válidas sobre el total es muy pequeña. En consecuencia, el operador no genera ninguna solución válida en un tiempo razonable y el proceso de búsqueda se hace inviable si no se eliminan las restricciones durante el mismo (permitiendo soluciones no válidas en la población); esto último es, a su vez, contraproducente, dado que permite al algoritmo dirigir la búsqueda hacia zonas del espacio en que no hay ninguna solución válida pudiendo quedar estancado. El operador de reasignación múltiple, que sí considera las restricciones y cuya tasa de soluciones válidas generadas es mucho mayor, tampoco consigue realizar por sí solo una exploración completa.

Se hace necesario, pues, considerar otros operadores o incluir en los mismos conocimiento sobre el problema, tanto en cuanto a su naturaleza de partición como en cuanto a sus objetivos y requerimientos.

Sin embargo, se ha de tener la precaución de no informar demasiado las operaciones. Los trabajos previos pusieron de manifiesto que ninguna heurística es óptima siempre. Por ello tratamos de encontrar un equilibrio a la hora de diseñar los operadores, procurando permitir siempre realizar los casos extremos y tomar las decisiones que parezcan (según la heurística escogida) menos prometedoras, sin dejar de favorecer esas operaciones supuestamente mejores.

En todas las descripciones de esta sección dos áreas/UT se consideran relacionadas si tienen un valor mayor que 0 en la ecuación (1.15) (existe cierta interacción entre ambas), dos áreas son adyacentes si al menos una de las UT del primero es adyacente a una UT del segundo, y se llama UT fronterizas a aquellas dentro de un área que son adyacentes a, como mínimo, una UT de un área diferente. Los términos grupo, área y mercado se usan como si fueran equivalentes, al igual que elemento y UT.

Para facilitar el diseño e implementación de los operadores, éstos se basan en heurísticas estocásticas de la forma “seleccionar un mercado/UT de <atributo> elevado/medio/bajo” y “seleccionar un mercado/UT con alta/media/baja <relación> con otro mercado/UT”. Estas heurísticas han de permitir dirigir el proceso de búsqueda hacia regiones prometedoras manteniendo la capacidad de los operadores de realizar una búsqueda aleatoria que pueda producir cualquier posible cambio y alcanzar cualquier región del espacio de búsqueda. En el PDAF, <atributo> puede ser la autonomía mínima (ecuación (1.19)) o cualquier otra forma de autonomía, grado de validez (ecuación (1.21)) y tamaño (en términos de flujos de interacción). A su vez, <relación> puede ser cualquiera de los índices de interacción que se vayan a emplear (en nuestro caso eq.(1.15)). En cuanto a los métodos de selección estocástica, se ha empleado selección por torneo a tres en los casos en que se haya de escoger entre un gran número de mercados o UT del territorio, porque se calcula con rapidez y no ejerce demasiada presión de selección, y selección con probabilidad proporcional a la relación cuando se trata de identificar mercados o UT relacionados con otros, ya que el número de posibles relaciones será con frecuencia mucho menor al número total (incluso menor que tres) y en esos casos el torneo a tres ejercería demasiada presión de selección, cuando es preferible un comportamiento más aleatorio.

Obviamente, estas heurísticas son dependientes del problema concreto y en otros casos serán de interés otros atributos o relaciones, aunque la forma general de las heurísticas será la misma.

### 2.7.2.1. Operador de cruce basado en grupos

En este trabajo se ha diseñado un único operador de cruce de agrupación (a diferencia de las publicaciones anteriores) para trabajar sobre la representación de codificación grupo-número, siguiendo la estructura descrita por FALKENAUER (1998), común a muchas aplicaciones de AEA (DATTA *et al.*, 2008; JAMES *et al.*, 2007a,b; LEWIS y PAECHTER, 2006; NOKTEHDAN *et al.*, 2010; PANKRATZ, 2005).

A uno de los dos individuos cuya información se va a cruzar le denominaremos el padre donante y al otro, el padre receptor. El operador procede como sigue:

1. Crea un nuevo individuo, hijo, copia exacta del padre receptor.
2. Selecciona al azar una proporción de entre 1 y el 66 % de los grupos del padre donante y los sobrescribe sobre el hijo, cambiándoles la codificación para que no coincida con ningún grupo presente inicialmente en el hijo.
3. Se comprueba que todos los grupos sean continuos, dividiendo en sus partes continuas aquellos que no lo sean.
4. Se aplica la heurística AJE sobre todos los grupos del hijo para reparar los grupos del padre receptor que hayan dejado de cumplir las restricciones del problema.
5. Terminar con éxito (la solución será evaluada, pudiera ser inválida).

La necesidad de esta heurística es difícilmente cuestionable dado que una inmensa mayoría de cruces resultará a priori en grupos inválidos, y en los AE es la operación de cruce la única que permite un intercambio de información entre soluciones, factor importante en el éxito de estos algoritmos poblacionales.

### 2.7.3. Resultados y comparación con propuestas anteriores

En esta publicación se presenta una comparativa de resultados enfrentando el conjunto de operadores desarrollado en ella con el que se desarrolló en las publicaciones anteriores. Se utilizan dos casos de territorios reales, con datos de movilidad cotidiana derivados del Censo de Población de 2001 (INSTITUTO NACIONAL DE ESTADÍSTICA, 2005): (a) la Comunidad Valenciana, con 541 UT, que es el mismo caso de estudio empleado en las publicaciones anteriores relativas a AE y que permite comparar el rendimiento y eficacia de la propuesta evolutiva más reciente frente a las anteriores (se emplean los mismos parámetros que en las demás publicaciones), y (b) el territorio completo de España, con 8104 UT, que hasta el momento se ha resuelto únicamente mediante la aplicación de métodos voraces y permite evaluar el rendimiento de la nueva propuesta en problemas de tamaño superior, frecuentes en este contexto cuando no se recurre a técnicas de preprocesamiento que agreguen las UT iniciales siguiendo alguna heurística voraz para reducir su número.

Como ya se ha indicado anteriormente, las propuestas evolutivas anteriores resultaban demasiado lentas para resolver problemas de dimensión similar en tiempo razonable, incluso ejecutando en paralelo sobre decenas de máquinas (computadores personales), necesitando varios minutos de procesamiento por cada mutación. Como sabemos que casos de un tamaño muy inferior, como el de la Comunidad Valenciana, necesitan millones de generaciones para alcanzar la convergencia, podemos estimar que el caso de España requeriría años de procesamiento.

Para permitir la comparación con los procedimientos oficiales relevantes también se han incluido los resultados de la aplicación del algoritmo de las TTWA (COOMBES *et al.*, 1986) y su nueva versión (ONS *et al.*, 2008) empleada en la última regionalización oficial en Reino Unido, con los mismos parámetros.

#### 2.7.3.1. Comunidad Valenciana

La tabla 1 del apéndice G (p. 6764 de la publicación original) muestra la evaluación de los resultados de las dos versiones del método de las

TTWA (COOMBES *et al.*, 1986; ONS *et al.*, 2008) sin resolución de discontinuidades, de la versión no adaptativa y una de las adaptativas probadas en FLÓREZ-REVUELTA *et al.* (2009). Los parámetros de todos los algoritmos evolutivos son los mismos. El nuevo conjunto de operadores ofrece una mejora significativa con respecto a los resultados obtenidos anteriormente, aun cuando se les permita ejecutar sólo el 10 % del tiempo dado al procedimiento anterior. De hecho, el valor promedio de la función objetivo y del número de mercados delimitados con ejecuciones de tres minutos es mayor que la mejor solución obtenida con el algoritmo anterior en ejecuciones diez veces más largas; además, la mejor solución con el nuevo método incrementa en un 5,5 % la función objetivo del mejor resultado anterior.

La fig. 3 del apéndice G (p. 6764 de la publicación original) ilustra cómo los nuevos operadores propician una evolución más rápida tanto en número de generaciones necesarias para alcanzar la convergencia como en incremento absoluto de la función objetivo por generación.

En la tabla 2 del apéndice G se muestran algunos estadísticos del comportamiento de los nuevos operadores. En comparación con los valores de las mismas estadísticas para los operadores de la propuesta publicada en FLÓREZ-REVUELTA *et al.* (2009) los que aquí se proponen producen un porcentaje mucho mayor de soluciones válidas y que mejoran a sus padres. El número de operaciones completadas por tiempo (sobre el mismo *hardware*) resultó ser similar, por lo que no se aprecia que la inclusión de la restricción de contigüidad haya acelerado las operaciones. Se concluye, pues, que la principal causa del incremento de rendimiento apreciado en los nuevos operadores, en comparación con los de propuestas anteriores, estriba en la reducción de la tasa de individuos inválidos que se consigue gracias a la consideración en los operadores de las restricciones y no en la aceleración de las operaciones.

Se recomienda al lector acudir a la sección 5.1 del apéndice G (p. 6763 de la publicación original) para una discusión ampliada de estos resultados.

### **2.7.3.2. Resolviendo problemas grandes: España**

Como se observa en la tabla 3 del apéndice G, el nuevo conjunto de los operadores propuesto en este trabajo, aplicado a grandes problemas como es

el caso del territorio de España, es capaz de producir en un tiempo razonable muy buenos resultados, comparado con métodos institucionales. Esto no era posible haciendo uso del primer conjunto de operadores desarrollado y presentado en las publicaciones previas: en las pruebas de aplicación de los algoritmos en FLÓREZ-REVUELTA *et al.* (2009) sobre el caso español, tras una hora de cómputo sólo se habían completado siete mutaciones de las que sólo dos produjeron cambios válidos, mientras que la nueva propuesta completó más de medio millón de generaciones con más de trescientas mil soluciones válidas en el mismo plazo. En este caso sí se nota una fuerte mejora de rendimiento frente a los operadores que no consideran la restricción de contigüidad.

Con respecto al comportamiento de los operadores en este segundo ejercicio, la tabla 4 del apéndice G (p. 6765 de la publicación original) muestra las mismas estadísticas utilizadas en la tabla 2. Hay diferencias notables en el rendimiento en comparación con las estadísticas obtenidas de la Comunidad Valenciana: p. ej., el incremento de adecuación acumulado por segundo contribuido por el operador R es mucho menor en este caso, mientras que la mutación de inclusión mejora mucho en ese aspecto.

#### 2.7.4. Conclusiones

El método presentado en esta publicación supone un gran avance en la delimitación de MLT mediante técnicas de optimización y es en cierto sentido el remate de la investigación realizada en esta tesis, ya que ofrece resultados para problemas reales típicos que superan en calidad a ninguno de los publicados hasta la fecha.

Cabe señalar que las soluciones aportadas por los métodos TTWA y las versiones anteriores del genético tienen por lo general cierto número de regiones discontinuas, por lo que se hace necesario un paso adicional de refinamiento. Como se argumenta en CASADO-DÍAZ *et al.* (2010), trabajo ya revisado en la sección anterior y que se incluye en apéndice F, este proceso tiene como resultado una reducción del valor de adecuación alcanzado por la solución discontinua, incluso cuando se emplea una técnica de optimización local embebida en el procedimiento. La nueva propuesta considera en todo

momento esta restricción sin ver restringida su capacidad exploratoria del espacio de búsqueda, como demuestran los buenos resultados obtenidos.

Durante la experimentación se comprobó que, para el problema y función objetivo escogidos, la inclusión del operador de unión y división combinadas (R) ofrecía mejores resultados y hacía innecesario el uso de los correspondientes operadores básicos. Esto indica que la inclusión de técnicas adaptativas que ajusten dinámicamente las probabilidades de los operadores podría dar buenos resultados.

Los principales aportes de este trabajo se pueden resumir en dos: por una parte, el desarrollo de un nuevo conjunto de operadores orientados a objetos, generales y de fácil adaptación para su aplicación en problemas de delimitación de AF, capaces de producir resultados mucho mejores que los de los métodos utilizados en este campo hasta la fecha, y, por otra, la confirmación de que la inclusión de la restricción de contigüidad en este tipo de problemas es una práctica beneficiosa cuando se resuelven mediante técnicas de optimización combinatoria que, por su propia naturaleza, garantizan una exploración adecuada del espacio de soluciones, a diferencia de lo que ocurre con los métodos deterministas empleados en este ámbito tradicionalmente.

## Capítulo 3

# Conclusiones

En este capítulo se resumen las conclusiones alcanzadas a lo largo del desarrollo de esta tesis y se cierra con un apartado dedicado a reseñar algunas líneas de investigación que podrían extender el trabajo hecho hasta ahora.

### 3.1. Discusión

La delimitación de áreas funcionales es un problema muy relevante en las ciencias regionales, entre otras. Trazar las fronteras adecuadas que definen áreas altamente autónomas formadas por unidades territoriales muy cohesionadas influye en gran medida en la calidad de la formulación de políticas a nivel local en ámbitos como son el de la vivienda, el transporte o los mercados de trabajo.

En esta tesis se propone un algoritmo evolutivo instrumentado con un nuevo conjunto de operadores de agrupación cuidadosamente diseñados para afrontar problemas de delimitación de mercados laborales y otros relacionados. Entre nuestros principios o guías de diseño hemos incluido: producir el menor número posible de soluciones no válidas para reducir el espacio de búsqueda; incorporar conocimiento estocástico (pero no óptimo) en los operadores para acelerar la convergencia sin impedir una exploración exhaustiva; y considerar cualquier restricción del problema en el funcionamiento de

los operadores, así como en el proceso de optimización conjunto. Algunos de estos principios fueron descartados en las primeras versiones de este trabajo, en particular la inclusión de la restricción de contigüidad en las AF delimitadas, puesto que varios autores argumentaban que su aplicación podría dar lugar a una exploración insuficiente del espacio de búsqueda y, por lo tanto, mantener los resultados lejos de la solución óptima —p. ej., no se utiliza en el método de las TTWA (COOMBES *et al.*, 1986; ONS *et al.*, 2008)—. Además, la exclusión de esta condición en los primeros trabajos de esta investigación no impidió que los operadores desarrollados fueran capaces de producir soluciones continuas casi en su totalidad.

Sin embargo, los resultados presentados en MARTÍNEZ-BERNABEU *et al.* (2012) contradicen estas asunciones, ya que se superan ampliamente los valores alcanzados en diferentes indicadores de calidad (número de mercados identificados, autonomía media de los mercados, autonomía global de la delimitación y, obviamente, en la función objetivo), tanto en términos de la mejor solución encontrada como en sus valores medios, con respecto a los resultados obtenidos con los métodos tradicionalmente empleados y con los anteriores algoritmos evolutivos desarrollados en esta investigación.

Se pensaba que esta estrategia reduciría la complejidad de muchas subrutinas de uso frecuente en el método en cualquier caso, acelerando notablemente el cálculo en comparación con el funcionamiento de los mismos operadores sin la restricción. Sin embargo, la experimentación demostró que los cambios han tenido un impacto leve en ese aspecto cuando se trata de problemas de tamaño pequeño, como es el caso de la Comunitat Valenciana, con 541 UT, donde el coste de aplicación de ambos conjuntos de operadores es similar en cuanto al rendimiento en operaciones por unidad de tiempo (si se midiera como el incremento de la función objetivo por unidad de tiempo el nuevo método consigue una evolución notablemente más rápida y eficiente). Así, el principal impulso en el rendimiento en esos casos se debe al mayor porcentaje de soluciones válidas producidas por los operadores, y a su mayor variedad.

En problemas más grandes los operadores propuestos sí tienen un rendimiento significativamente superior al de los operadores que no incluyen la contigüidad geográfica en su lógica, por lo que no se puede descartar el

impacto de la restricción de la contigüidad en la complejidad del método. La experimentación mostró un rendimiento muy superior para el caso de España, con 8108 UT, en el que el primer conjunto de operadores no era capaz de dar una solución en un plazo razonable. Estos resultados parecen indicar que la introducción del requisito de contigüidad en la lógica de los operadores (no sólo en la evaluación de las soluciones) no ha limitado perjudicialmente la capacidad de búsqueda de los nuevos operadores, sino que puede haber reducido notablemente la complejidad de éstos mejorando su rendimiento. Ello sería posible gracias a la drástica reducción del número de UT o áreas que se han de considerar en las operaciones, limitándose a las adyacentes geográficamente, que suelen ser menos que las conectadas funcionalmente, y, también, al aumento de la tasa de operaciones válidas.

Por otra parte, las regionalizaciones resultantes de este nuevo método son, obviamente, continuas sin la necesidad de aplicar un procedimiento final para corregir discontinuidades (como el propuesto en CASADO-DÍAZ *et al.*, 2010) que podría volver inútil parte del esfuerzo de producir una solución buena pero discontinua. En resumen, la inclusión de esta restricción ha dado lugar únicamente a efectos positivos, reduciendo considerablemente el espacio de búsqueda sin excluir a buenas soluciones o dificultar la exploración y limitando la necesidad de ajustes manuales o no formalizados en los resultados de las delimitaciones.

También se ha considerado la posibilidad de que las soluciones inválidas permanezcan en la población y se crucen con los padres válidos, pero ello implicaría cambios importantes en el método general y en los propios operadores, ya que todos ellos están diseñados bajo la premisa de que cualquier mercado en una solución cumple a priori con las restricciones —y ello permite evitar muchas costosas comprobaciones—. Además, las soluciones inválidas (siempre en términos de cumplimiento de las restricciones) podrían tener mayor éxito reproductivo que las válidas, lo que llevaría a un estancamiento en una población completa de buenas soluciones (según la función objetivo) que no son aceptables (según las restricciones), si no se penaliza de alguna manera. Por lo tanto la adecuación de esa idea merece una consideración más amplia que excede los propósitos de este estudio.

Algunos de los operadores del primer conjunto diseñado producían cambios totalmente aleatorios, otros empleaban heurísticas voraces (que persi-

guen cambios óptimos localmente pero que no necesariamente conducen a la solución óptima global) y otros eran informados (probabilidad proporcional a algún atributo o relación). El diseño de estos operadores estuvo guiado por la búsqueda de soluciones a dificultades concretas en los procesos de exploración o explotación que se observaban al analizar los resultados que se conseguían durante el desarrollo del algoritmo. Globalmente considerados, todos ellos tienen una función ad hoc, especializados en resolver el problema al que se aplican, y con un desarrollo *bottom-up*. En la propuesta final, que supone el cuerpo principal de este trabajo, se ha seguido un esquema de diseño *top-down* más sistemático, esbozando primero todas las operaciones lógicas sobre conjuntos disjuntos que pudieran ser útiles en un proceso de búsqueda combinatoria, tanto en la exploración como en la explotación, en el contexto del particionado. Además, todos los operadores se han hecho informados sin perder su naturaleza estocástica. Por lo tanto, es probable que el rendimiento del método sea más alto también debido a una estrategia de búsqueda más versátil.

Una breve experimentación inicial no ha demostrado que la reducción del conjunto de operadores afecte positivamente al rendimiento del método, con la excepción de las mutaciones de unión y partición, a las que es preferible descartar cuando se emplea la mutación R que combina ambas. Una batería de pruebas más amplia es necesaria para determinar esta cuestión con seguridad.

A pesar de estar diseñado para el problema de la delimitación de MLT, el nuevo conjunto de operadores propuesto es directamente aplicable a la delimitación de otras áreas funcionales como los mercados de vivienda (BROWN y HINCKS, 2008) y las áreas de transporte (KRYGSMAN *et al.*, 2009), sujeto a la disponibilidad de las matrices de interacción necesarias (como los flujos de movilidad, las migraciones, los flujos de mercancías y pasajeros, los movimientos de los teléfonos móviles entre las células, etc.). Los operadores propuestos son simples y generales y el método completo es capaz de resolver problemas reales de gran tamaño. La adaptación de la propuesta presentada, mediante la adecuada elección de unas restricciones y función objetivo y el ajuste de las heurísticas de los operadores, podría permitir resolver muchos otros problemas de partición de grafos o redes —como multiplicación de matrices dispersas, balanceado de carga, elaboración de calendarios, enrutamiento de vehículos, coloreado de mapas, etc.

Para concluir, cabe señalar que el método desarrollado en esta tesis supone un nuevo enfoque para la delimitación de MLT, que mejora de manera sustancial los resultados de los métodos tradicionales. El hecho de que expertos en esta área estén comenzando a citar algunas de las publicaciones aquí resumidas y a emplear en sus trabajos el mismo enfoque que se presenta en esta tesis (WATTS, 2009; FARMER y FOTHERINGHAM, 2011; OBERST, 2011) nos indica que esta propuesta ha abierto una línea de investigación importante.

### **3.2. Objetivos de trabajo futuros**

Como primera línea de ampliación, se están explorando los modelos paralelos para algoritmos evolutivos, no solo para acelerar los cálculos sobre arquitecturas paralelas sino para mejorar la eficacia del proceso lógico de búsqueda, mediante la creación de subpoblaciones. En la actualidad ya se han obtenido resultados de un modelo celular, que mejoran notablemente los presentados hasta la fecha, y ya se está preparando una publicación en la que se presentará este trabajo.

Por otro lado, se aplicará el método desarrollado a otros problemas relacionados que comparten la mayoría de los criterios y restricciones, en concreto la delimitación de mercados locales de vivienda (sobre flujos migratorios) y económicos (sobre flujos de bienes/servicios). También se plantea realizar las delimitaciones de MLT en ámbitos territoriales más amplios, como sería el caso de los países de la UE. Estos resultados serán de interés en el campo de las ciencias regionales y urbanas.

En la literatura reciente, varios autores han llegado a la conclusión de que las áreas funcionales económicas se deben delimitar con más información que la movilidad cotidiana (COMMUNITIES y GOVERNMENT, 2010), combinando en el análisis flujos de distintos tipos (p. ej. bienes y servicios, migraciones, desplazamientos por ocio y llamadas de teléfono). Se abre por tanto otra línea de investigación, pudiendo desarrollar bien un método que combine la información de los diferentes tipos de flujos en una única matriz de interacción como entrada del método ya desarrollado, bien un método

que combine los resultados del método aplicado a cada una de las matrices de interacción iniciales, o bien la modificación de los operadores para que consideren al mismo tiempo la información de las diferentes matrices. Combinando las opciones primera y segunda, se plantea la superposición de fronteras de múltiples soluciones, es decir, contar para cada par de UT la cantidad de áreas en que están juntos en diferentes soluciones (se crea así un mapa de frecuencias, ya considerado en FLÓREZ-REVUELTA *et al.*, 2008), y utilizar la matriz resultante (en la que dos UT tendrán una fuerte relación si suelen formar parte de la misma área) como entrada de otra regionalización (COOMBES, 2000).

Será interesante también la aplicación del método desarrollado a problemas de otros campos no relacionados y que están recibiendo mucha atención en la actualidad, como son la modularización de *software* y la detección de comunidades en redes sociales y biológicas. Los objetivos a optimizar en estos problemas son equivalentes a los tratados en este trabajo, por lo que la adaptación del conjunto de operadores propuesto no será costosa. El tamaño de alguno de estos problemas puede llegar a ser muy elevados (cientos de miles de nodos y millones de aristas en los datos de interacción), por lo que serían un buen marco de experimentación para la mejora de los operadores.

Si se comprobara que el método es eficaz y eficiente en la suficiente cantidad de problemas se podría considerar el desarrollo de un paquete de *software* configurable que permita a investigadores su aplicación a problemas afines. En este caso sería interesante realizar la integración de la aplicación con algún *middleware* para computación GRID, que permita sacar partido de las altas capacidades de paralelización de este tipo de técnicas de búsqueda poblacional.

En ese sentido, otra línea de trabajo práctica consiste en la incorporación dentro del *software* de delimitación de una interfaz con un sistema de información geográfica (SIG), que permita el mapeado y visualización de los resultados, su evolución y los diferentes indicadores de calidad a nivel de UT y de área, para reducir drásticamente el tiempo requerido por el usuario para la presentación y análisis de los resultados.

Con el interés de aumentar el rendimiento del método en su conjunto, se está trabajando también en una versión adaptativa que aproveche el hecho

de que cada operador es más útil en etapas específicas de la búsqueda, o bajo ciertas condiciones de la solución sobre la que se aplica el operador. Como ocurrió con nuestras aproximaciones previas, los datos sobre la eficiencia variable de los operadores apuntan a un posible aumento en el rendimiento de todo el método mediante un correcto ajuste o preferiblemente la (auto) adaptación de las probabilidades de aplicación de cada operador en función de sus resultados o de la etapa de la evolución, de modo que cada operador se utilice cuando su contribución al proceso de optimización puede ser mayor. La experiencia previa con estas técnicas (FLÓREZ-REVUELTA *et al.*, 2009) nos muestra que el único indicador a tener en cuenta en tales procesos es el de la calidad final alcanzada y el tiempo requerido para ello, y que no podemos basarnos en otros indicadores parciales como son el porcentaje de individuos válidos o que mejoran al padre. Dado que, en última instancia, sólo es determinante el resultado final obtenido por el conjunto de operadores tras completar su evolución, parece a priori que no será inmediato encontrar una forma de adaptación.

Para concluir, una línea inmediata de trabajo sería, lógicamente, la consideración de funciones objetivo alternativas, en función de las necesidades e intereses de los distintos grupos de usuarios de los procedimientos desarrollados. En esta línea, una versión multiobjetivo del procedimiento será de especial interés, por su capacidad de conciliar todos los indicadores de calidad en un conjunto de buenas soluciones (las soluciones no dominadas, el frente de Pareto) que se presentan a los usuarios y que cubren todo el espectro de posibles equilibrios entre los diferentes objetivos que se han de optimizar. De esta forma se podría facilitar el ajuste de parámetros del FADP o incluso eliminar su necesidad, además de ayudar al usuario en la elección de la mejor solución. Funciones objetivo alternativas a las ya desarrolladas podrían evaluar la cohesión (a través, p. ej., de diferentes indicadores de uso en Teoría de Grafos), compacidad geográfica y/o la homogeneidad de tamaño, por nombrar alguna de las variables posibles.



# Bibliografía

- ALBA, Enrique, TALBI, El-Ghazali, LUQUE, Gabriel y MELAB, Nouredine (2005). «Metaheuristics and Parallelism». En ALBA, Enrique (ed.) *Parallel Metaheuristics*, Wiley Series on Parallel and Distributed Computing, Wiley.
- ALONSO LOGROÑO, María Pilar, BEAMONTE SAN AGUSTÍN, María Asunción, GARGALLO VALERO, María Pilar y SALVADOR FIGUERAS, Manuel (2008). *Mercados locales de trabajo en Aragón*. Zaragoza: FUNDEAR.
- ANDERSEN, Anne K. (2000). *Commuting Areas in Denmark*. Inf. téc., AKF Forlaget, Copenhagen.
- (2002). «Are Commuting Areas Relevant for the Delimitation of Administrative Regions in Denmark?» *Regional Studies*, **36(8)**: 833–844.
- BAUMANN, Johann H., FISCHER, Manfred M. y SCHUBERT, Uwe (1983). «A Multiregional Labour Supply Model for Austria: The Effects of Different Regionalisations in Multiregional Labour Market Modelling». *Papers in Regional Science*, **52(1)**: 53–83.
- BLUM, Christian, ROLI, Andrea y ALBA, Enrique (2005). «An Introduction to Metaheuristic Techniques». En ALBA, Enrique (ed.) *Parallel Metaheuristics*, Wiley Series on Parallel and Distributed Computing, Wiley.
- BRANDES, Ulrik, DELLING, Daniel, GAERTLER, Marco, GÖRKE, Robert, HOEFER, Martin, NIKOLOSKI, Zoran y WAGNER, Dorothea (2007). «On Finding Graph Clusterings with Maximum Modularity». En BRANDS-TÄDT, Andreas, KRATSCH, Dieter y MÜLLER, Haiko (eds.) *Graph-Theoretic Concepts in Computer Science, Lecture Notes in Computer Science*, 4769, 121–132, Springer Berlin / Heidelberg.

- BROWN, Evelyn C. y SUMICHRAST, Robert T. (2005). «Evaluating Performance Advantages of Grouping Genetic Algorithms». *Engineering Applications of Artificial Intelligence*, **18(1)**: 1–12.
- BROWN, Peter J. B. y HINCKS, Stephen (2008). «A Framework for Housing Market Area Delineation: Principles and Application». *Urban Studies*, **45(1)**: 2225–2247.
- CASADO-DÍAZ, José Manuel (2000). «Local Labour Market Areas in Spain: A Case Study». *Regional Studies*, **34(9)**: 843–856.
- CASADO-DÍAZ, José Manuel y COOMBES, Mike G. (2005). «The Delineation of 21st Century Local Labour Market Areas (LLMAs)». En *Proceedings of the 8th Nectar Conference*, Las Palmas de Gran Canaria, Spain. Unpublished.
- CASADO-DÍAZ, José Manuel, MARTÍNEZ-BERNABEU, Lucas y FLÓREZ-REVUELTA, Francisco (2010). «Los mercados locales de trabajo españoles. Una aplicación del nuevo procedimiento británico». En ALBERTOS, J. M. y FERIA, J. M. (eds.) *La ciudad real en España: procesos urbanos y metropolitanos en la primera década del siglo XXI*, 275–313, Madrid: Thomson-Civitas.
- CASADO-DÍAZ, José Manuel y COOMBES, Mike (2011). «The Delineation of 21st Century Local Labour Market Areas: A Critical Review and a Research Agenda». *Boletín de la Asociación de Geógrafos Españoles*, **(57)**: 7–32 y 393–396.
- COMMUNITIES y GOVERNMENT, Local (2010). *Functional Economic Market Areas: An Economic Note*. Inf. téc., Communities and Local Government, London.
- COOMBES, Mike G. (2000). «Defining Locality Boundaries with Synthetic Data». *Environment and Planning A*, **32**: 1499–1518.
- (2002). *Travel to Work Areas and the 2001 Census, Report to the Office of National Statistics*. Inf. téc., Centre for Urban and Regional Development Studies, Newcastle upon Tyne.
- COOMBES, Mike G., GREEN, Anne E. y OPENSHAW, Stan (1986). «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(10)**: 943–953.
- CÖRVERS, Frank, HENSEN, Maud y BONGAERTS, Dion (2009). «Delimitation and Coherence of Functional and Administrative Regions». *Regional Studies*, **43(1)**: 19–31.
- DATTA, D., FIGUEIRA, J. R., FONSECA, C. M. y TAVARES-PEREIRA, F. (2008). «Graph Partitioning Through a Multi-Objective Evolutionary Algorithm: A Preliminary Study». En *Genetic and Evolutionary Computation Conference, GECCO'08*, Atlanta, USA: ACM.
- DE JONG, Kenneth A. (2006). *Evolutionary Computation: a Unified Approach*. Bradford Books, Cambridge, MA: MIT Press.
- DE SOLLA PRICE, Derek J. (1965). «Networks of Scientific Papers». *Science*, **149**: 510–515.
- DEMIRKOL, Ilker, ERSOY, Cem, ÇAĞLAYAN, Mehmet Ufuk y DELİÇ, Hakan (2001). «Location Area Planning in Cellular Networks Using Simulated Annealing». En *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, 1, 13–20, IEEE.
- (2004). «Location Area Planning and Cell-to-Switch Assignment in Cellular Networks». *IEEE Transactions on Wireless Communications*, **3**: 880–890.
- DORIGO, M. (1992). *Optimization, Learning and Natural Algorithms*. Tesis Doctoral, Politecnico di Milano, Italia.
- EUROSTAT y COOMBES, M. G. (1992). *Étude sur les zones d'emploi*. Inf. Téc. E/LOC/20, Office for Official Publications of the European Communities, Luxembourg.
- FALKENAUER, E. (1998). *Genetic Algorithms and Grouping Problems*. New York: John Wiley & Sons.
- FARMER, Carson J.Q. y FOTHERINGHAM, A. Stewart (2011). «Network-Based Functional Regions». *Environment and Planning A*, **43(11)**: 2723–2741.

- FENG, Zhiqiang (2008). «Fuzziness of Travel to Work Areas». *Regional Studies*, **43(5)**: 707–720.
- FERIA, José María, SUSINO, Joaquin, CASADO-DÍAZ, José Manuel *et al.* (2005). *Movilidad por razón de trabajo en Andalucía, 2001*. Inf. téc., Instituto de Estadística de Andalucía, Sevilla.
- FLÓREZ-REVUELTA, Francisco, CASADO-DÍAZ, José Manuel y MARTÍNEZ-BERNABEU, Lucas (2008). «An Evolutionary Approach to the Delineation of Functional Areas Based on Travel-to-work Flows». *International Journal of Automation and Computing*, **1(5)**: 10–21.
- (2009). «Adaptive Evolutionary Algorithms for the Delineation of Local Labour Markets». En *IEEE Congress on Evolutionary Computation*, 2354–2360. CEC2009, Trondheim.
- FLÖRKEMEIER, H. (2002). *Functional Regions and the Measurement of Economic Integration*. Discussion Paper No. 01/02, Albert-Ludwigs-Universität Freiburg.
- FORTUNATO, Santo (2010). «Community Detection in Graphs». *Physics Reports*, **486**: 75–174.
- FRANCIS, Richard L. y WHITE, John A. (1974). *Facility Layout and Location: an Analytical Approach*. Prentice-Hall.
- GAREY, Michael R. y JOHNSON, David S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York, NY, USA: W. H. Freeman & Co.
- GAREY, Michael R., JOHNSON, David S. y STOCKMEYER, L. (1976). «Some Simplified NP-Complete Graph Problems». *Theoretical Computer Science*, **1(3)**: 237–267.
- GHOSH, Tamal, DAN, Pranab K., SENGUPTA, Sourav y CHATTOPADHYAY, Manojit (2010). «Genetic Rule Based Techniques in Cellular Manufacturing (1992-2010): A Systematic Survey». *International Journal of Engineering, Science and Technology*, **2(5)**: 198–215.
- GLOVER, Fred (1977). «Heuristics for Integer Programming Using Surrogate Constraints». *Decision Sciences*, **8(1)**: 156–166.

- GOLDBERG, David E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Professional.
- GUIMERÀ, Roger y NUNES AMARAL, Luis A. (2005). «Functional Cartography of Complex Metabolic Networks». *Nature*, **433(026113)**: 895–900.
- HARMAN, Mark, MANSOURI, S. Afshin y ZHANG, Yuanyuan (2009). *Search Based Software Engineering: A Comprehensive Analysis and Review of Trends Techniques and Applications*. Inf. Téc. TR-09-03, King's College London, Washington, D.C.
- HARMAN, Mark, SWIFT, Stephen y MAHDAVI, K. (2005). «An Empirical Study of the Robustness of Two Module Clustering Fitness Functions». En *Proceedings of the 2005 Conference on Genetic and Evolutionary Computation (GECCO '05)*, 1029–1036, ACM.
- HOLLAND, John H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI, USA: University of Michigan Press.
- HRUSCHKA, Eduardo Raul, CAMPELLO, Ricardo J. G. B., FREITAS, Alex A. y DE CARVALHO, André C. Ponce Leon F. (2009). «A survey of evolutionary algorithms for clustering». *Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews*, **39(2)**: 133–155.
- INSTITUTO NACIONAL DE ESTADÍSTICA (2005). *Censo de Población 2001*. Inf. téc., Instituto Nacional de Estadística. <http://www.ine.es>.
- ISTAT (1997). «I sistema locali del lavoro 1991». *Argomenti*, **(10)**.
- (2005). *I Sistemi Locali Del Lavoro 2001*. Inf. téc., Istituto Nazionale di Statistica. <http://www.istat.it>.
- JAMES, Tabitha L., BROWN, Evelyn C. y KEELING, Kellie B. (2007a). «A Hybrid Grouping Genetic Algorithm for the Cell Formation Problem». *Computers & Operations Research*, **34(7)**: 2059–2079.
- JAMES, Tabitha L., VROBLEFSKI, Mark y NOTTINGHAM, Quinton (2007b). «A Hybrid Grouping Genetic Algorithm for the Registration Area Planning Problem». *Computer Communications*, **30(10)**: 2180–2190.
- JIANG, Hsinyi (2006). «Can the Genetic Algorithm Be a Good Tool for Software Engineering Searching Problems?» En *Proceedings of the 30th*

*Annual International Computer Software and Applications Conference - Volume 02*, 362–366, Washington, DC, USA: IEEE Computer Society.

JOHNSON, Kenneth P. y KORT, John R. (2004). «2004 Redefinition of the BEA Economic Areas». *Surveys of Current Business*, **84(11)**: 68–75.

JONES, Donald R. y BELTRAMO, Mark A. (1991). «Solving Partitioning Problems with Genetic Algorithms». En BELEW, Richard K. y BOOKER, Lashon B. (eds.) *Proceedings of the Fourth International Conference on Genetic Algorithms*, 442–449, San Mateo, CA: Morgan Kaufmann.

KARLSSON, C. y OLSSON, M. (2006). «The Identification of Functional Regions: Theory, Methods and Applications». *The Annals of Regional Science*, **40(1)**: 1–18.

KELLAM, Paul, LIU, Xiaohui, MARTIN, Nigel, ORENKO, Christine, SWIFT, Stephen y TUCKER, Allan (2002). «A Framework for Modelling Virus Gene Expression Data». *Intelligent Data Analysis*, **6(3)**: 267–279.

KING, J. R. y NAKORNCHAI, V. (1982). «Machine-Component Group Formation in Group Technology: Review and Extension». *International Journal of Production Research*, **20(2)**: 117–133.

KIRKPATRICK, S., GELATT, C. D. y VECCHI, M. P. (1983). «Optimization by Simulated Annealing». *Science*, **220(4598)**: 671–680.

KRASNOGOR, Natalio y GUSTAFSON, Steven (2004). «A Study on the Use of “Self-generation” in Memetic Algorithms». *Natural Computing*, **3(1)**: 53–76.

KRYGSMAN, S., DE JONG, T. y NEL, J. (2009). «Functional Transport Regions in South Africa: An Examination of National Commuter Data». En *Proceedings of the 28th Southern African Transport Conference*, Document Transformation Technologies.

KUSIAK, Andrew (1987). «The Generalized Group Technology Concept». *International Journal of Production Research*, **25(4)**: 561–569.

LANCICHINETTI, Andrea y FORTUNATO, Santo (2009). «Community Detection Algorithms: A Comparative Analysis». *Physical Review E*, **80**.

- LEWIS, Rhydian y PAECHTER, Ben (2006). «Finding Feasible Timetables Using Group-based Operators». *IEEE Transactions on Evolutionary Computation*, **11(3)**: 397–413.
- LIMBOURG, S. y JOURQUIN, B. (2010). «Location and Market Area of Railroad Terminals». En *Proceedings of ORBEL 24*, SOGESCI-BVWB. ORBEL, Liege, 28–29 January 2010.
- LOBO, F.G., LIMA, C.F. y MICHALEWICZ, Z. (2007). *Parameter Setting in Evolutionary Algorithms*, 54. Springer Verlag.
- LUQUE, Gabriel, ALBA, Enrique y DORRONSORO, Bernabé (2005). «Parallel Genetic Algorithms». En ALBA, Enrique (ed.) *Parallel Metaheuristics*, Wiley Series on Parallel and Distributed Computing, Wiley.
- MA, Kang-Rae y BANISTER, David (2006). «Excess Commuting: A Critical Review». *Transport Reviews*, **26(6)**: 749–767.
- MANCORIDIS, S., MITCHELL, Brian S., CHEN, Y. y GANSNER, E. R. (1999). «Bunch: A Clustering Tool for the Recovery and Maintenance of Software System Structures». En *Proceedings of the IEEE International Conference on Software Maintenance (ICSM '99)*, 50–59, IEEE Computer Society Press.
- MANCORIDIS, S., MITCHELL, Brian S. y RORRES, C. (1998). «Using Automatic Clustering to Produce High-Level System Organizations of Source Code». En *Proceedings of the 6th International Workshop on Program Comprehension (IWPC '98)*, 45–53, IEEE Computer Society Press.
- MANTERE, Timo y ALANDER, Jarmo T. (2005). «Evolutionary Software Engineering, a Review». *Applied Soft Computing*, **5(3)**: 315–331.
- MARTÍNEZ-BERNABEU, Lucas, FLÓREZ-REVUELTA, Francisco y CASADO-DÍAZ, José Manuel (2012). «Grouping Genetic Operators for the Delineation of Functional Areas Based on Spatial Interaction». *Expert Systems with Applications*.
- MASSER, Ian y BROWN, Peter J. B. (1975). «Hierarchical Aggregation Procedures for Interaction Data». *Environment and Planning A*, **7(5)**: 509–523.

- MELLER, Russell D. y GAU, Kai-Yin (1996). «The Facility Layout Problem: Recent and Emerging Trends and Perspectives». *Journal of Manufacturing Systems*, **15**(5).
- MITCHEL, William y WATTS, Martin (2010). «Identifying Functional Regions in Australia Using Hierarchical Aggregation Techniques». *Geographical Research*, **48**(1): 24–41.
- MITCHELL, Brian S. (2002). *A Heuristic Search Approach to Solving the Software Clustering Problem*. Tesis Doctoral, Drexel University, Philadelphia, PA.
- MOSCATO, Pablo y COTTA, Carlos (2010). «A Modern Introduction to Memetic Algorithms». En GENDREAU, Michel y POTVIN, Jean-Yves (eds.) *Handbook of Metaheuristics, International Series in Operations Research and Management Science*, 146, 141–183, Springer US.
- NEL, J.H., KRYGSMAN, S.C. y DE JONG, T. (2008). «The Identification of Possible Future Provincial Boundaries for South Africa Based on an Intramax Analysis of Journey-to-Work Data». *ORiON*, **24**(2): 131–156.
- NEWELL, James O. y PERRY, Martin (2005). «Explaining Continuity in New Zealand's Local Labour Market Areas 1991 to 2001». *Australasian Journal of Regional Studies*, **11**(2): 155–174.
- NEWMAN, M. E. J. y GIRVAN, M. (2004). «Finding and Evaluating Community Structure in Networks». *Physical Review E*, **69**.
- NOKTEHDAN, Azadeh, KARIMI, Behrooz y KASHAN, Ali Husseinzadeh (2010). «A Differential Evolution Algorithm for the Manufacturing Cell Formation Problem Using Group Based Operators». *Expert Systems With Applications*, **37**: 4822–4829.
- OBERST, C.A. (2011). *A Methodology Approach to Delineate Functional Economic Market Areas: With an Iterative Three-Step Spatial Clustering Procedure*. CAWM discussion paper 55, Centrum für Angewandte Wirtschaftsforschung Münster.
- OCDE (2002). *Redefining Territories: The Functional Regions*. Inf. téc., Organización para la Cooperación y el Desarrollo Económicos, Paris.

- ONS, COOMBES, Mike G. y BOND, S. (2008). *Travel-To-Work Areas: the 2007 Review*. Inf. téc., Office for National Statistics, London.
- OPENSHAW, Stan (1984). *The Modifiable Areal Unit Problem*, 38. Norwich: Geo Books.
- PANKRATZ, Giselher (2005). «A Grouping Genetic Algorithm for the Pickup and Delivery Problem with Time Windows». *Operations Research Spectrum*, **27**: 21–41.
- PAPAIIOANNOU, Grammatoula y WILSON, John M. (2010). «The Evolution of Cell Formation Problem Methodologies Based on Recent Studies (1997–2008): Review and Directions for Future Research». *European Journal of Operational Research*, **206**: 509–521.
- PAPPS, Kerry L. y NEWELL, James O. (2002). *Identifying Functional Labour Market Areas in New Zealand: A Reconnaissance Study Using Travel-to-Work Data*. IZA Discussion Papers 443, Institute for the Study of Labor (IZA).
- RADCLIFFE, Nicholas J. (1991). «Forma Analysis and Random Respectful Recombination». En *Proceedings of the Fourth International Conference on Genetic Algorithms*, Morgan Kaufmann.
- RÄIHÄ, O. (2009). *A Survey on Search-Based Software Design*. Inf. Téc. D-2009-1, Department of Computer Sciences, University of Tampere.
- RASHEED, Khaled, NI, Xiao y VATTAM, Swaroop (2003). «Methods for Using Surrogate Models to Speed Up Genetic Algorithm Optimization: Informed Operators and Genetic Engineering». En JIN, Y. (ed.) *Knowledge Incorporation in Evolutionary Computation, Studies in Fuzziness and Soft Computing*, 167, 103–122, Springer-Verlag.
- RELA, Leo (2004). *Evolutionary Computing in Search-based Software Engineering*. Proyecto Fin de Carrera, Lappeenranta University of Technology.
- ROSVALL, Martin y BERGSTROM, Carl T. (2008). «Maps of Random Walks on Complex Networks Reveal Community Structure». *Proceedings of the National Academy of Sciences*, **105**(4): 1118–1123.
- ROTA, Gian Carlo (1964). «The Number of Partitions of a Set». *The American Mathematical Monthly*, **71**(5): 498–504.

- ROY, John R. y THILL, Jean-Claude (2004). «Spatial Interaction Modelling». *Papers in Regional Science*, **83**: 339–361.
- SANDOU, Guillaume, FONT, St'ephane, TEBBANI, Sihem, HIRET, Arnaud y MONDON, Christian (2008). «Enhanced Genetic Algorithm with Guarantee of Feasibility for the Unit Commitment Problem». En MONMARCHÉ, N. et al. (ed.) *Proceedings of the Evolution artificielle, 8th international conference on Artificial evolution*, EA'07, 291–302, Berlin, Heidelberg: Springer-Verlag.
- SCB (1992). *Local Labour Markets and Employment Regions, a New Regional Division of Labour Market Areas in Sweden Grounded on Commuting Statistics, Labour Information 1991:7*. Inf. téc., SCB (Statistics Sweden), Stockholm.
- SHIBATA, Naoki, KAJIKAWA, Yuya, TAKEDA, Yoshiyuki y MATSUSHIMA, Katsumori (2008). «Detecting Emerging Research Fronts Based on Topological Measures in Citation Networks of Scientific Publications». *Technovation*, **28(11)**: 758–775.
- TOLBERT, Charles M. y SIZER, Molly K. (1987). *Labor Market Areas for the United States*. Staff Report AGES870721, Economic Research Service, U.S. Dept. Agr., Washington, D.C.
- (1996). *U.S. Commuting Zones and Labor Market Areas: A 1990 Update*. Staff Report AGES-9614, Rural Economy Division, Economic Research Service, U.S. Dept. Agr., Washington, D.C.
- TOMASSINI, Marco (2005). *Spatially Structured Evolutionary Algorithms*. Artificial Evolution in Space and Time, Secaucus, NJ, USA: Springer-Verlag New York, Inc.
- TUCKER, Allan, SWIFT, Stephen y LIU, Xiaohui (2001). «Variable Grouping in Multivariate Time Series Via Correlation». *IEEE Transactions on Systems, Man, and Cybernetics. Part B: Cybernetics*, **31(2)**: 235–245.
- VROBLEFSKI, Mark y BROWN, Evelyn C. (2006). «A Grouping Genetic Algorithm for Registration Area Planning». *Omega*, **34(3)**: 220–230.
- ŠTRANCAR, Janez, KOKLIČ, Tilen, ARSOV, Zoran, FILIPIČ, Bogdan, STOPAR, David y HEMMINGA, Marcus A. (2005). «Spin Label EPR-Based

- Characterization of Biosystem Complexity». *Journal of Chemical Information and Modeling*, **45(2)**: 394–406.
- WANG, Tsan-Pin, HWANG, Shu-Yuen y CHIEN-CHAO, Tseng (1998). «Registration Area Planning for PCS Networks Using Genetic Algorithms». *Vehicular Technology, IEEE Transactions on*, **47(3)**: 987–995.
- WATTS, Martin (2004). «Local Labour Markets in New South Wales: Fact Or Fiction?» En CARLSON, E (ed.) (ed.) *A Future that Works*, 461–472, Callaghan, N.S.W: University of New Castle, Centre of Full Employment and Equity.
- (2009). «Rules Versus Hierarchy: An Application of Fuzzy Set Theory to the Assessment of Spatial Grouping Techniques». En ET AL., M. Kolehmainen (ed.) *Proceedings of the 9th international conference on Adaptive and natural computing algorithms*, ICANNGA'09, 517–526, Berlin, Heidelberg: Springer-Verlag. Kuopio, Finland.
- WELSH ASSEMBLY GOVERNMENT (2006). *Local Housing Market Assessment Guide*. Inf. téc., Welsh Assembly Government.
- WOLPERT, D. H. y MACREADY, W. G. (1997). «No Free Lunch Theorems for Optimization». *Evolutionary Computation, IEEE Transactions on*, **1(1)**: 67–82.

