

DT

005

2014

# Segmentación **laboral** y análisis de **clusters**

Una aplicación al mercado de trabajo  
cualificado andaluz

Fernando Núñez Hernández

Ana Rodríguez Martín

Universidad de Sevilla

Carlos Usabiaga Ibáñez

Universidad Pablo de Olavide



JUNTA DE ANDALUCÍA

Centro de Estudios Andaluces

**CONSEJERÍA DE LA PRESIDENCIA**

**Segmentación laboral y análisis de clusters. Una aplicación al mercado de trabajo cualificado andaluz**

**Labour segmentation and cluster analysis. An application to the skilled Andalusian labour market**

RESUMEN:

Este trabajo aborda la aplicación de una metodología de «clusters» a la información disponible sobre las colocaciones producidas en la economía andaluza entre enero de 2007 y diciembre de 2010 en el ámbito de los trabajadores con mayor cualificación, que son los pertenecientes a los grupos de ocupación 1 y 2 dentro de la Clasificación Nacional de Ocupaciones (CNO-94). Dicha metodología permite dividir al mercado de trabajo en clusters (o mercados de trabajo locales homogéneos) mediante el agrupamiento de segmentos laborales definidos según el municipio, el grupo de ocupación y el sector de actividad tanto del puesto cubierto como del trabajador que lo cubre, y también según el sexo y la edad del trabajador. Esta segmentación del mercado de trabajo se ajusta a la forma en que describen el emparejamiento los modelos teóricos actuales sobre el desempleo de búsqueda. Los desempleados que pertenecen a un determinado cluster tienen mayores opciones de encontrar empleo si buscan dentro de su cluster, ya que dicho cluster se basa precisamente en las experiencias exitosas de empleo de otros trabajadores con similares características. Pensamos que esta metodología puede constituir una útil herramienta de información para la intermediación laboral, por cuanto permite proporcionar a cada buscador de empleo información precisa e individualizada sobre su entorno laboral más propicio.

**Palabras clave:** Emparejamiento Laboral; Segmentación Laboral; Trabajadores Cualificados; Análisis de Clusters; Políticas Activas del Mercado de Trabajo.

**Códigos JEL:** J63, J64.

ABSTRACT:

This paper addresses the application of a cluster methodology to available information on the hiring process observed in the Andalusian economy between January 2007 and December 2010 in the area of workers with high qualifications, which are those belonging to groups of occupation 1 and 2 within the National Classification of Occupations (CNO-94). This methodology allows dividing the labour market into clusters (or homogeneous local labour markets) by means of the clustering of labour segments defined by the municipality, the occupational group and sector of activity of both the vacancy and the worker, and additionally by sex and age of the worker. This segmentation of the labour market adjusts to current theoretical models of search unemployment. Unemployed workers belonging to a particular cluster are more prone to find employment if they search inside their cluster, since this cluster is based precisely on the successful experiences of employment of other workers with similar characteristics. We think that this methodology can be a useful information tool for labour intermediation, because it can provide each job seeker with detailed and precise information on the most suitable employment environment.

**Key Words:** Labour Matching; Labour Segmentation; High Skilled Workers; Cluster Analysis; Active Labour Market Policies.

**JEL Codes:** J63, J64.

# Segmentación **laboral** y análisis de *clusters*

Una aplicación al mercado de trabajo  
cualificado andaluz



DT

005  
2014

# Segmentación **laboral** y análisis de *clusters*

Una aplicación al mercado de trabajo cualificado  
andaluz

Fernando Núñez Hernández

Ana Rodríguez Martín

Universidad de Sevilla

Carlos Usabiaga Ibáñez

Universidad Pablo de Olavide



Centro de Estudios Andaluces  
CONSEJERÍA DE LA PRESIDENCIA

Colección Documentos de Trabajo: 005/2014

Edita:  
Fundación Pública Andaluza Centro de Estudios Andaluces,  
Consejería de la Presidencia, Junta de Andalucía

© Del texto: sus autores  
© De la edición:  
Fundación Pública Andaluza Centro de Estudios Andaluces  
Bailén, 50 — 41001 Sevilla  
Tel.: 955 055 210  
Fax: 955 055 211  
[www.centrodeestudiosandaluces.es](http://www.centrodeestudiosandaluces.es)

Octubre de 2014

ISSN: en trámite

<b>1.</b>	<b>INTRODUCCIÓN .....</b>	<b>9</b>
<b>2.</b>	<b>EVOLUCIÓN DEL MERCADO DE TRABAJO EN ESPAÑA Y ANDALUCÍA.....</b>	<b>13</b>
<b>3.</b>	<b>LOS DATOS EMPLEADOS.....</b>	<b>17</b>
<b>4.</b>	<b>METODOLOGÍA PARA ANALIZAR LA INFORMACIÓN SOBRE EL EMPLEO .....</b>	<b>19</b>
4.1.	Segmentación y propensión al emparejamiento .....	19
4.2.	Agrupaciones de segmentos en <i>clusters</i> .....	22
<b>5.</b>	<b>UNA APLICACIÓN A LOS TRABAJADORES ANDALUCES CUALIFICADOS.....</b>	<b>25</b>
5.1.	Obtención de 2.000 <i>clusters</i> de segmentos de trabajador .....	28
5.2.	Obtención de 1.000 <i>clusters</i> de segmentos de trabajador .....	34
5.3.	Obtención de 500 <i>clusters</i> de segmentos de trabajador.....	39
<b>6.</b>	<b>CONCLUSIONES.....</b>	<b>45</b>
<b>7.</b>	<b>BIBLIOGRAFÍA .....</b>	<b>49</b>



# 1. Introducción<sup>1</sup>

**A**ctualmente, la preocupación que experimenta un trabajador desempleado por estar en situación de desempleo se ha visto acentuada por la falta de puestos de trabajo que ha generado la persistente crisis económica en España. En este contexto, analizar y comprender la distribución de las colocaciones que se han producido recientemente en el mercado de trabajo puede suponer una ayuda para aquellos trabajadores que buscan empleo, ya que les puede proveer de información acerca de cómo otros trabajadores de parecidas características han conseguido tener éxito en su proceso de búsqueda de empleo.

Este trabajo se sustenta en información sobre las colocaciones de trabajadores con estudios superiores que han tenido lugar en el territorio andaluz en el período 2007-2010. Es decir, consideramos a trabajadores cualificados que residen, al comienzo de su búsqueda de empleo, en algún municipio andaluz. Los trabajadores cualificados suelen mostrar una alta predisposición relativa a la movilidad laboral; además, pueden experimentar episodios laborales de sobre-educación, por lo que son un colectivo interesante desde el punto de vista del emparejamiento laboral.

Los mercados laborales modernos se caracterizan por la existencia de grandes flujos de puestos y de trabajadores que se mueven entre los estados de actividad e inactividad.

---

1 Agradecemos al Centro de Estudios Andaluces su apoyo a esta línea de investigación (Proyecto PRY 010/10), y a Pablo Álvarez de Toledo sus sugerencias de mejora del trabajo.

Un elemento clave en la modelización de los flujos del mercado de trabajo es la función de emparejamiento agregada, que representa la tecnología de intercambio entre los trabajadores que buscan empleo y las empresas que buscan a trabajadores, que pueden finalmente dar lugar a emparejamientos productivos. La idea clave es que un proceso de intercambio complejo, como el que se da en el mercado de trabajo, puede ser englobado en una función con buenas propiedades matemáticas que señala el número de colocaciones que se forman en cada período en función, básicamente, de los recursos que empresas y trabajadores dedican a la búsqueda de una pareja laboral. Variaciones en el número de colocaciones, dados unos recursos o *inputs*, reflejan cambios en la intensidad de las fricciones que afectan al mercado de trabajo; fricciones tales como la falta de buena información sobre la existencia de vacantes o de trabajadores adecuados, heterogeneidades, congestión en determinados segmentos laborales, escasa movilidad del factor trabajo, poca eficacia de la intermediación laboral, etc. Cuando las fricciones laborales son más fuertes, el mercado de trabajo se vuelve menos eficiente en el emparejamiento, y eso puede haber sucedido en la economía andaluza en los últimos años. La base teórica de este trabajo se asienta en los modelos que realizan un análisis del mercado laboral desde una perspectiva de los flujos que en él acontecen. En estos modelos, la existencia de heterogeneidades y de segmentación juega un papel fundamental a la hora de explicar el proceso de emparejamiento laboral<sup>2</sup>.

En este trabajo se aborda el concepto de segmentación del mercado de trabajo desde una perspectiva empírica. En concreto, las variables descriptivas de la colocación contenidas en nuestros datos nos han permitido segmentar el mercado de trabajo andaluz. Los trabajadores que buscan empleo han sido agrupados en «segmentos de trabajador», que son definidos por las siguientes características del buscador: sexo, edad, municipio de residencia, grupo de ocupación y sector de actividad. Por otro lado, los puestos de trabajo ofrecidos por las empresas se van a agrupar en «segmentos de puesto», cuyas características son el grupo de ocupación, el sector de actividad y el municipio donde se encuentra el centro de trabajo de la vacante. La unión de ambos segmentos, cada vez que se produce una colocación, forma lo que llamamos un «segmento conjunto». La forma en que se distribuyen las colocaciones de los segmentos de trabajador en los diferentes segmentos de puesto nos va a permitir cuantificar la similitud que existe entre dos segmentos de trabajador cualesquiera; esta medida de similitud se usará para aplicar una metodología de agrupamiento a los diferentes segmentos de trabajador. La idea es agrupar a los segmentos que son más similares en grupos o *clusters* que puedan ser considerados como mercados de trabajo locales.

---

2 Véanse en esta línea los trabajos de Devine y Kiefer (1991), Mortensen y Pissarides (1999), Pissarides (2000, 2008), Petrongolo y Pissarides (2001), Shimer (2007) y Yashiv (2007).

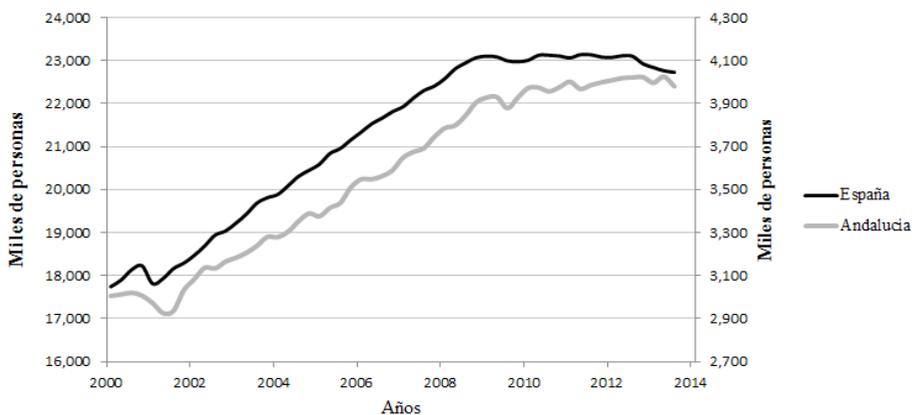
Nuestro trabajo se estructura como sigue. En la sección 2 se ofrece una perspectiva general de la evolución de los mercados de trabajo español y andaluz en los últimos años. En la sección 3 se realiza una breve descripción de los datos que se han empleado en la parte empírica del trabajo. Por su parte, en la sección 4 se establece el marco empírico del trabajo, dentro del cual se puede realizar un análisis de conglomerados o *clusters*. En la sección 5 se aplica dicha metodología de *clusters* a una base de datos de colocaciones de trabajadores titulados. Analizaremos los resultados, comentando los grupos o *clusters* más significativos que se han formado en el proceso de agrupamiento y los flujos de colocaciones principales que se han producido dentro de cada *cluster*. Todo este análisis permitirá generar un «mapa» de *clusters* (o mercados de trabajo locales) de trabajadores con formación, *clusters* que coexisten en el territorio andaluz. Esta información puede suponer una de las hojas de ruta a seguir por un trabajador titulado que comience una búsqueda de empleo en Andalucía. Finalmente, la sección 6 expone las principales conclusiones del trabajo.



## 2. Evolución del mercado de trabajo en España y Andalucía

**E**l mercado de trabajo andaluz representa una parte importante del panorama laboral nacional, ya que la población activa andaluza supone cerca de una quinta parte de la población activa española. La tasa de desempleo se situaba en el conjunto nacional en un 25,73 % en el cuarto trimestre de 2013, mientras que en la comunidad autónoma andaluza alcanzaba un 36,26 %. Aunque nuestro análisis posterior se concentra en el período 2007-2010, a continuación vamos a analizar la evolución en España y en Andalucía en el período 2000-2013 de las principales variables laborales que determinan el emparejamiento: población activa, número de desempleados y de vacantes y flujo de colocaciones.

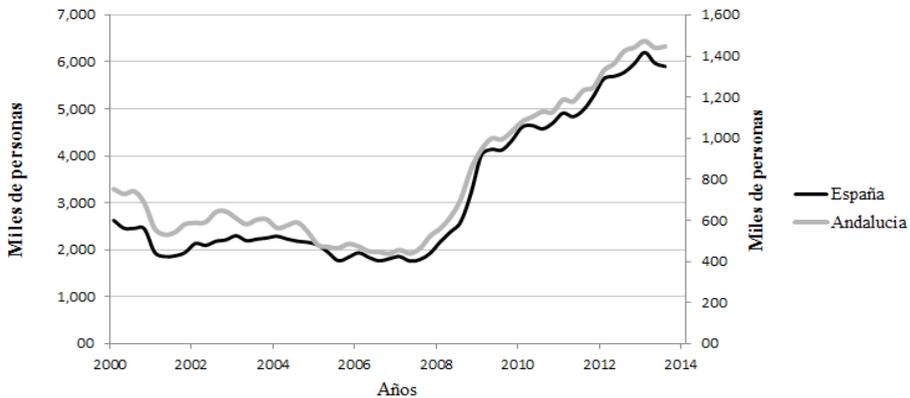
Figura 1. Evolución de la población activa (2000-2013)



Fuente: elaboración propia a partir de datos de la EPA.

En primer lugar, se muestra la evolución de la población activa en España (eje izquierdo); se observa un aumento progresivo de la misma desde el año 2001 hasta el 2009, momento en el cual se estanca, pudiendo ser debido a la entrada del país en crisis económica en el año 2008. Este hecho ha podido generar en España cierto desánimo sobre las personas que han perdido su empleo o en aquellas personas que se han incorporado por primera vez al mercado de trabajo y que por dicha crisis han experimentado que la búsqueda de empleo que han llevado a cabo no ha dado sus frutos. También ha podido producirse un efecto «trabajador añadido». En torno al año 2012 comienza incluso una tendencia negativa, descendiendo el número de personas pertenecientes a la población activa; pudiendo ser debido a que el desánimo mencionado en la búsqueda de empleo se haya trasladado geográficamente al extranjero: por un lado, jóvenes sin opciones de un puesto de trabajo en España y, por otro lado, inmigrantes que llegaron en periodo de bonanza económica y que vuelven a sus lugares de origen o se van a otros países. La evolución de la población activa en Andalucía se ha trazado en el eje derecho y se observa que la tendencia en la comunidad andaluza ha sido similar a la tendencia nacional y ha supuesto algo menos de una quinta parte de la población activa de España aproximadamente.

Figura 2. Evolución del número de personas en paro (2000-2013)

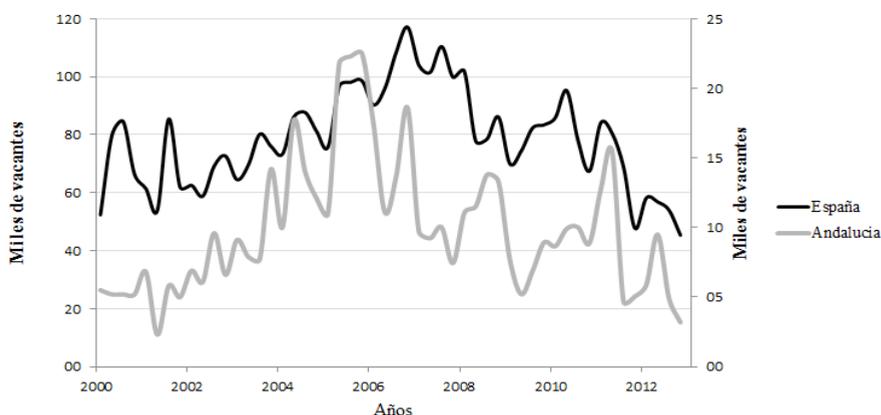


Fuente: elaboración propia a partir de datos de la EPA.

En este gráfico, se observa con claridad la drástica subida del paro que se produjo cuando España se vio envuelta en la crisis económica en el año 2008; además, esta tendencia creciente se extiende hasta el año 2013, cuando se atisba una ligera mejoría, ya que el número de parados ha sufrido un ligero descenso en los últimos meses. Por su parte, la evolución en la comunidad autónoma andaluza (representada en el eje

derecho) ha seguido la misma tendencia, suponiendo en torno a una cuarta parte de los desempleados españoles. El valor máximo que ha alcanzado el número de parados en España ha resultado ser de 6.278.200 desempleados, mientras que en la comunidad autónoma andaluza esta cifra se ha situado en 1.481.700 desempleados.

Figura 3. Evolución del número de vacantes (2000-2013)

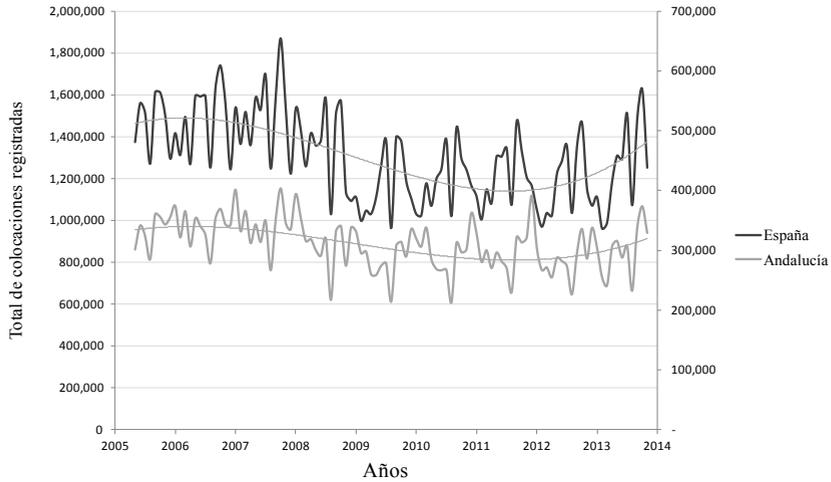


Fuente: elaboración propia a partir de datos de la *Encuesta de Coyuntura Laboral*<sup>3</sup>.

La tendencia del número de vacantes en España es aproximadamente inversa a la mostrada en el gráfico anterior, ya que el aumento del número de personas desempleadas se ha podido ver influenciado, entre otras causas, por la disminución de puestos vacantes en el país. Se observa que la evolución ha sido ascendente hasta el tercer trimestre de 2007, donde se llega a un máximo de 110.000 vacantes disponibles en España, y desde entonces el número de vacantes ha ido disminuyendo hasta alcanzar un nivel de 46.000 vacantes, cifra menor que los puestos de trabajo disponibles al comienzo del año 2000. En Andalucía, representada en el eje derecho, el curso de las vacantes ha sido muy parecido al nacional, aunque se diferencia del anterior en el pico de vacantes, ya que se produce en el cuarto trimestre de 2005, es decir, dos años antes. Se puede observar que ambos trazados no son continuos, sino que se producen grandes variaciones de un trimestre a otro consecutivo.

3 La metodología de la *Encuesta de Coyuntura Laboral* indica que la serie de vacantes se circunscribe a las actividades económicas de la industria, la construcción y los servicios, excluyéndose de este último sector los subsectores de Administración Pública, defensa y seguridad social obligatoria, y organismos extraterritoriales y organizaciones religiosas.

Figura 4. Evolución del número de colocaciones (2005-2013)



Fuente: elaboración propia a partir de los datos del Servicio Público de Empleo Estatal.

Por último, se representa la evolución en el tiempo del número de colocaciones en España y Andalucía, así como las tendencias cuadráticas de ambas series (Andalucía nuevamente se representa en el eje de la derecha). En ambas zonas se puede observar una tendencia negativa desde mayo de 2006 hasta mediados de 2011, momento en el cual las colocaciones empezaron a aumentar, siguiéndose esta tendencia positiva hasta el último año representado. Las contrataciones en Andalucía se han mantenido en torno a las 310.000 al año.

### 3. Los datos empleados

**L**os datos empleados en este trabajo proceden de una muestra de colocaciones de trabajadores titulados producidas en el territorio andaluz durante el período 2007-2010; se trata de una muestra de flujo de salida —flujo de salida en el sentido de que, en cada colocación, el buscador sale de su estatus de actividad (normalmente del desempleo) hacia el empleo—. La base de datos está formada por microdatos, por lo que su grado de detalle es bastante mayor que si tratásemos con datos agregados<sup>4</sup>. Para cada colocación observada tenemos información sobre las características de las dos partes que se emparejan: el trabajador que buscaba empleo y el puesto que estaba vacante. Las variables que se van a utilizar en nuestro análisis, inherentes al buscador, son edad, sexo, provincia y municipio en que reside, grupo de ocupación y sector de actividad donde desarrolla su búsqueda. En relación al puesto de trabajo que se cubre, se han considerado el municipio del centro de trabajo de la vacante, el sector de actividad y el grupo de ocupación. Estas variables están codificadas de acuerdo con las clasificaciones ofrecidas por el INE. De esta forma, el sector de actividad se ha codificado según la Clasificación Nacional de Actividades Económicas (CNAE-93)<sup>5</sup>, el grupo de ocupación haciendo uso de la Clasificación Nacional de Ocupaciones (CNO-94), y los municipios y provincias haciendo uso de la relación de códigos de municipios y provincias.

---

4 Un análisis agregado sobre el emparejamiento laboral en la economía española puede verse en Álvarez de Toledo *et al.* (2008); a nivel andaluz, véase por ejemplo Álvarez de Toledo *et al.* (2011).

5 Para las colocaciones producidas a partir de marzo de 2009 ha sido necesario combinar las clasificaciones CNAE-93 y CNAE-09 en lo que se refiere a las variables de sector de actividad del trabajador y del puesto.



## 4. Metodología para analizar la información sobre el empleo

### 4.1. Segmentación y propensión al emparejamiento

En este estudio, se parte de un mercado laboral en el que el trabajo es un bien heterogéneo y en el que la información relativa a los puestos de trabajo vacantes no llega por igual a todos los trabajadores, ni la información sobre los solicitantes de empleo se reparte de igual forma entre todas las empresas. La falta de información perfecta sobre candidatos y vacantes idóneos para cada buscador hace que el periodo de tiempo necesario para que se produzca un contacto empresa-trabajador sea mayor, y la existencia de heterogeneidad en las habilidades y preferencias inherentes a los trabajadores y en las que son requeridas por las empresas determinan que no todo contacto empresa-trabajador, una vez que se produce, acabe en colocación. El modelo de Pissarides (2000) sintetiza todo este proceso de búsqueda y emparejamiento mediante la definición de una función agregada de emparejamiento, que muestra el número de colocaciones que se crean en un instante de tiempo en un mercado de trabajo en función del número de trabajadores que buscan empleo (principalmente desempleados), del número de vacantes que ofrecen las empresas y de otras variables que pueden influir en dicho proceso de emparejamiento. El grado de heterogeneidad que subyace en esta función y, por tanto, en el mercado de trabajo, recibe el nombre de desajuste laboral

o *mismatch*<sup>6</sup>. Si en el mercado de trabajo no existiese ningún *mismatch*, las vacantes y los trabajadores se encontrarían de forma inmediata, no existiendo una función de emparejamiento como tal.

En los últimos años, se ha realizado un notable esfuerzo por comprender y explicitar la *tecnología* que subyace en la *caja negra* que es la función de emparejamiento. Modelos como el de *mismatch*, islas, bolas en urnas, taxis, colas, stock flujo, etc., han explorado diferentes tipos de heterogeneidad, extendiendo la teoría de búsqueda y emparejamiento laboral hacia modelos más microfundamentados<sup>7</sup> que han segmentado, de alguna forma, el mercado de trabajo en diferentes partes (mercados de trabajo locales, islas, colas, emparejamientos productivos y no productivos, trabajadores y puestos *stock* (antiguos) o flujo (nuevos), etc.). En general, cada parte considerada ha sido tratada como homogénea. En esta línea de la segmentación laboral se engloba nuestro estudio.

La principal idea de este trabajo es que la falta de homogeneidad en el mercado de trabajo provoca la segmentación del mismo; es decir, características tales como la cualificación, localización, edad, etc., hacen que determinados puestos de trabajo sólo sean adecuados para determinados trabajadores. Por ejemplo, puede haber personas que sólo realizan la búsqueda de empleo enfocada hacia puestos de trabajo con una determinada localización y para una determinada ocupación o actividad, y prefieren esperar a conseguir un empleo de este tipo antes que cambiar los parámetros de búsqueda; en nuestra terminología: prefieren no moverse de su segmento laboral.

Siguiendo la metodología de segmentación propuesta por Álvarez de Toledo *et al.* (2013, 2014), vamos a definir, a partir de los datos de colocaciones de los que disponemos, segmentos laborales atendiendo a las características del puesto y del trabajador que forman cada colocación. El objetivo del trabajo es encontrar la relación o proximidad que existe entre los distintos segmentos definidos, pudiendo formar agrupamientos o *clusters* de los mismos, que bien podrían ser considerados como mercados de trabajo locales. La segmentación, de acuerdo con nuestros datos, va a consistir en distinguir segmentos tanto del lado de los trabajadores como del lado de los puestos vacantes. Los segmentos de trabajador estarán formados por las variables municipio

---

6 El término *mismatch* hace referencia, por ejemplo, al desajuste existente en términos de cualificaciones y de localización espacial entre los trabajadores y los puestos de trabajo.

7 Sobre el microfundamento de la función de emparejamiento, véanse por ejemplo los *surveys* de Petrongolo y Pissarides (2001) y Núñez y Usabiaga (2007) —este último con algunos apuntes para España—.

de residencia, grupo de ocupación, sector de actividad, sexo y edad<sup>8</sup>; todas ellas relativas al trabajador contratado. Los segmentos de puesto van a tener la información relativa al municipio del centro de trabajo, grupo de ocupación y sector de actividad del puesto vacante cubierto. Por último, la combinación o unión de ambos segmentos para cada colocación dará lugar a lo que llamaremos el segmento conjunto de la colocación.

En general, se van a obtener  $n$  segmentos de trabajador y  $m$  segmentos de puesto, por lo que se van a generar  $n \times m$  segmentos conjuntos. En un periodo de tiempo  $t$ , el número de colocaciones que se han producido para cada segmento conjunto viene dado por  $M_{ijt}$ , donde  $i$  y  $j$  representan las características correspondientes al segmento de trabajador y de puesto respectivamente. El número total de colocaciones de nuestra muestra ( $M_t$ ) es la suma, para todos los segmentos conjuntos, de los emparejamientos que se han producido en el intervalo de tiempo  $t$ ; también podemos obtener por agregación el número de colocaciones en el segmento de trabajador  $i$  ( $M_{it}$ ) y en el de puesto  $j$  ( $M_{jt}$ ):

$$M_t = \sum_{i=1}^n \sum_{j=1}^m M_{ijt} ; M_{it} = \sum_{j=1}^m M_{ijt} ; M_{jt} = \sum_{i=1}^n M_{ijt} \quad (1)$$

A partir de estos cálculos, se pueden obtener las probabilidades reales (usando la base de datos) de que se produzca un emparejamiento en un segmento conjunto, en un segmento de trabajador y en un segmento de puesto determinados:

$$p_{ijt} = \frac{M_{ijt}}{M_t} ; p_{it} = \frac{M_{it}}{M_t} = \sum_{j=1}^m p_{ijt} ; p_{jt} = \frac{M_{jt}}{M_t} = \sum_{i=1}^n p_{ijt} \quad (2)$$

Consideramos que el mercado laboral no presenta segmentación si la probabilidad de que un emparejamiento se produzca en el segmento de trabajador  $i$  no afecta a la probabilidad de que se produzca un emparejamiento en el segmento de puesto  $j$ , y viceversa; es decir, si ambos eventos son independientes. En este caso, la probabilidad de que se produzca una colocación en un segmento conjunto  $ij$  vendría dada por:

$$\widehat{P}_{ijt} = p_{it} p_{jt} \quad (3)$$

---

8 En los trabajos de Álvarez de Toledo *et al.* (2013, 2014) el segmento del trabajador consta de tres características: municipio, ocupación y actividad. En este trabajo dicho segmento está formado por esas características y dos más, sexo y edad, que también pueden dar lugar a segmentación laboral. Por otro lado, en nuestros cálculos, la edad se ha discretizado en intervalos de 5 años, para evitar que tengamos un segmento diferente para cada año de edad.

Por el contrario, si la probabilidad de que un emparejamiento se produzca en el grupo de trabajador  $i$  afecta a la probabilidad de que un emparejamiento se produzca en el grupo de puesto  $j$  o viceversa (por ejemplo, si el municipio, la ocupación o el sector de actividad de un puesto determinado afecta a su probabilidad de emparejamiento con un trabajador correspondiente a un municipio, ocupación y sector de actividad determinados), se considera que el mercado de trabajo está segmentado, y será diferente, en general, de  $p_{ijt}$ . En un caso extremo de segmentación, cada segmento de trabajador se emparejaría sólo con un segmento de puesto en particular, lo cual daría lugar a mercados de trabajo locales *aislados*.

La propensión al emparejamiento entre un segmento de trabajador  $i$  y un segmento de puesto  $j$  en un periodo de tiempo  $t$  ( $pe_{ijt}$ ) se puede medir como:

$$pe_{ijt} = p_{ijt} / \widehat{p}_{ijt} \quad (4)$$

tomando  $pe_{ijt}$  el valor 1 en mercados no segmentados. En mercados con segmentación, esta variable será mayor que 1 si los trabajadores del segmento  $i$  se emparejan con los puestos del segmento  $j$  con mayor frecuencia o intensidad de la que se observaría si el mercado estuviera no segmentado, y viceversa.

## 4.2. Agrupaciones de segmentos en *clusters*

Tras la segmentación del mercado propuesta, tenemos nuestros datos estructurados en segmentos de trabajadores, de puestos y de colocaciones, según las características ya mencionadas. La gran cantidad de segmentos generados hace difícil una interpretación sintética de los datos de colocaciones, por lo que, haciendo uso de la metodología de *clusters*, se puede reducir el número de segmentos a un número más reducido de grupos (o mercados de trabajo locales) basándonos en la similitud existente entre ellos. A partir de este proceso de agrupación, obtenemos una mejor visión global de la estructura del mercado de trabajo, ya que lo analizamos como si estuviese compuesto por un conjunto de mercados de trabajo locales. El período objeto de análisis, años 2007 a 2010, ha sido tratado como un único intervalo temporal —es decir, sin fraccionar anualmente los flujos de colocaciones observados—; esto impide analizar la posible dinámica temporal de los *clusters* identificados, pero, a cambio, permite aplicar nuestra metodología sobre un volumen de información mayor —tenemos más colocaciones dentro de cada segmento conjunto que si el análisis fuera, por ejemplo, anual—, por lo que el análisis gana en robustez.

La metodología de *clusters* o agrupamiento está basada en una medida de similitud. En nuestro caso, la similitud entre dos segmentos de trabajador se define por la forma en que *envían* colocaciones a los segmentos de puesto; es decir, dos segmentos de trabajador son más similares entre sí cuanto más se parece la forma en la que se distribuyen sus emparejamientos con los distintos segmentos de puesto. Siguiendo este enfoque, se define la similitud,  $sw_{i_1-i_2}$ , entre cada pareja de segmentos de trabajador  $i_1-i_2$  como la superposición o porcentaje de coincidencia de sus respectivas distribuciones. Este índice de superposición se puede obtener a partir de la siguiente suma de valores mínimos:

$$sw_{i_1-i_2} = \sum_{j=1}^m \min(pe_{i_1j}, pe_{i_2j}) = \sum_{j=1}^m p_j \min(pe_{i_1j}, pe_{i_2j}) \quad (5)$$

donde, por ejemplo,  $pe_{i_1-j_m} p_{j_m} = p_{j_m} (p_{i_1-j_m} / p_{i_1} p_{j_m}) = M_{i_1-j_m} / M_{i_1}$  supone el porcentaje de colocaciones que representa el segmento conjunto  $i_1-j_m$  en todas las colocaciones generadas por el segmento de trabajador  $i_1$  con los diferentes segmentos de puesto.

El valor de la similitud tomará un valor mínimo de cero, si el segmento de trabajador  $i_1$  no envía ninguna colocación a los segmentos de puesto a los que envía colocaciones el segmento de trabajador  $i_2$ , y tomará un valor máximo de uno, si las distribuciones de las colocaciones de ambos segmentos de trabajador son idénticas. Finalmente, se podría definir de forma paralela la similitud entre dos segmentos de puesto, pero en este trabajo nos vamos a centrar en la similitud de los segmentos de trabajadores<sup>9</sup>.

La metodología de *clusters* empleada se basa en una medida de similitud entre segmentos, pero el programa informático que hemos utilizado (Stata 11), trabaja con medidas de disimilitud (no con similitudes), por lo que, una vez obtenidas las superposiciones o similitudes con el procedimiento descrito, se ha procedido a calcular la disimilitud entre cada pareja de segmentos de trabajador como 1 menos el valor de su similitud. El proceso de agrupamiento sigue un método jerárquico, en el que se van agrupando, sucesivamente, los segmentos de trabajador con menor disimilitud entre sí en *clusters*, y los *clusters* entre sí para formar grupos mayores (que contienen cada vez un mayor número de segmentos de trabajadores). Es decir, se comienza considerando cada segmento de trabajador como un grupo o *cluster* separado, seguidamente se agrupan los dos segmentos de trabajador entre los que existe una menor disimilitud, y así sucesivamente, pudiendo continuarse el proceso hasta la fusión total de todos los segmentos en un único *cluster*.

<sup>9</sup> Para un análisis por el lado de la oferta de puestos, véase el trabajo de Álvarez de Toledo *et al.* (2013).

Análiticamente, la disimilitud entre dos *clusters* se calcula con la siguiente expresión:

$$D(A, B) = \frac{1}{N_A \cdot N_B} \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} d(s_i, s_j) \quad s_i \in A, s_j \in B \quad (6)$$

donde  $s_i$  y  $s_j$  son segmentos de trabajador del *cluster*  $A$  y  $B$  respectivamente,  $N_A$  y  $N_B$  son el número de segmentos de cada *cluster*, y  $d(s_i, s_j)$  es la disimilitud entre el segmento  $s_i$  del *cluster*  $A$  y el segmento  $s_j$  del *cluster*  $B$ . El proceso de agrupamiento de dos grupos cualesquiera se basa, por tanto, en el promedio de la disimilitud de ambos grupos.

Este método permite llegar a un número de *clusters* especificado, ya que se puede interrumpir el proceso cuando se alcanza dicho número. Cuanto menor sea el número de *clusters* formados, es decir, cuanto más cerca estemos de llegar a un único *cluster*, mayor será el número de segmentos dentro de cada *cluster* y, por tanto, mayor será el grado medio de heterogeneidad o disimilitud interior. Los resultados del proceso de agrupamiento pueden mostrarse gráficamente mediante un *dendrograma*, figura en la que se observa cómo se van formando grupos cada vez mayores y más distantes entre sí, aunque cuando el número de elementos a fusionar es muy elevado, esta figura pierde claridad, siendo mejor recurrir a los datos numéricos del propio proceso de agrupamiento para entender el proceso.

## 5. Una aplicación a los trabajadores andaluces cualificados

**E**n este trabajo se estudia a aquellos trabajadores (buscadores de empleo) cuyo grupo de ocupación pertenece a los dos primeros grupos de la CNO-94, que son los de mayor formación:

1. Dirección de las empresas y de las administraciones públicas;
2. Técnicos y profesionales científicos e intelectuales.

Asimismo, hemos eliminado de nuestra base de datos aquellas colocaciones en las que alguna característica del segmento conjunto al que pertenece es desconocida, ya que pueden generar superposiciones entre segmentos no reales. Por ejemplo, si no conocemos la variable *municipio* en un segmento de puesto y resulta que en ese segmento se producen varias colocaciones, cuando se realice la superposición de dos segmentos de trabajador que envíen colocaciones a dicho segmento de puesto, la superposición va a ser mayor que cero, pero en realidad el segmento de puesto puede corresponder realmente a dos o más segmentos diferentes cuya variable *municipio* sea distinta en ellos, lo cual desconocemos; si eso fuera así, la superposición entre los dos segmentos de trabajador anteriores no tendría por qué ser mayor que cero. Una vez eliminadas las colocaciones donde el grupo de ocupación del trabajador es menor a 3.000 y las correspondientes a segmentos conjuntos incompletos, el número de colocaciones de nuestra base de datos asciende a 358.298.

Los segmentos de trabajador se generan por la unión de las variables sexo, tramo de edad, municipio de residencia, ocupación y actividad; en el caso de segmentos de puesto, las variables que se usan son municipio del centro de trabajo, ocupación y actividad. Por último, los segmentos conjuntos están formados por todas las variables anteriores y salen de la unión del segmento de trabajador y del segmento de puesto que ha dado lugar a cada colocación. Para cada segmento (de trabajador, de puesto y conjunto) calculamos su frecuencia: el número total de colocaciones que se han producido en él. Una vez que se tiene este dato, la base de datos puede quedar expresada de manera que quede una fila por segmento conjunto diferente —en lugar de una fila por colocación—, y sobre dicha base de datos compactada se va a producir el cálculo de las propensiones al emparejamiento entre segmentos de trabajador y segmentos de puesto. Nuestra base de datos cuenta con 226.763 segmentos conjuntos, formados por 102.373 segmentos de trabajador y 61.232 segmentos de puesto.

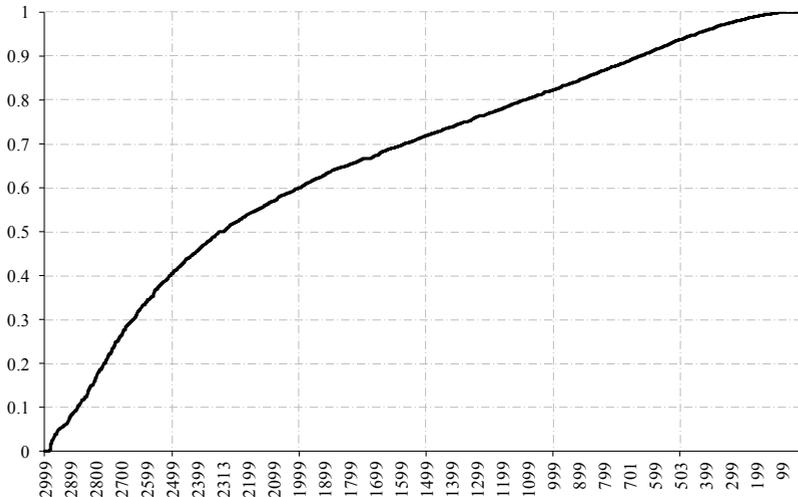
Al aplicar nuestra metodología, hemos encontrado dos problemas fundamentales: la baja frecuencia de ciertos segmentos de trabajador, lo cual los hace estadísticamente poco representativos, y el excesivo tamaño de la matriz de frecuencias conjuntas necesaria para calcular la similitud existente entre cualquier pareja de segmentos de trabajador; dicha matriz tiene por filas segmentos de trabajador y por columnas segmentos de puesto. Para evitar estos dos problemas, hemos optado por reducir las dimensiones de dicha matriz de frecuencias conjuntas de la siguiente forma: sólo se va a trabajar con aquellos segmentos de trabajador cuya frecuencia sea mayor o igual a 16, dando lugar esta operación a que nos quedemos sólo con 3.000 segmentos de trabajador (los más representativos). Además, no se han tenido en cuenta aquellos segmentos de puesto que sólo han recibido una colocación, ya que, de acuerdo con la ecuación (5), no van a aportar valor alguno al cálculo de las superposiciones entre parejas cualesquiera de segmentos de trabajador: por ejemplo, si calculamos la superposición entre los segmentos de trabajador  $i_1$  e  $i_2$  ( $sw_{i_1-i_2}$ ), el mínimo entre los valores  $pe_{i_1-j}$  y  $pe_{i_2-j}$  siempre va a ser cero, ya que una de las dos propensiones será nula (el segmento de puesto  $j$  sólo ha recibido una única colocación de alguno de los dos segmentos de trabajador, pero no del otro, y con el segmento de trabajador con el que no tiene colocación la  $pe$  será nula). A partir de estas operaciones, la base de datos queda reducida a 37.178 segmentos conjuntos, que contienen 3.000 segmentos de trabajador y 10.291 segmentos de puesto; el total de colocaciones asciende a 101.660.

Es importante resaltar en este punto que en este trabajo se han superado dos dificultades que se habían presentado en los trabajos anteriores de Álvarez de Toledo *et al.* (2013, 2014). Por un lado, la magnitud del problema que se ha conseguido abarcar es

mayor que en esos trabajos, ya que, en este caso, se ha aplicado la metodología a 3.000 segmentos de trabajador y 10.921 segmentos de puesto, frente a los algo más de 1.500 segmentos de trabajador y de segmentos de puesto empleados en dichos trabajos. Por otro lado, en los estudios anteriores se había utilizado como medida de similitud entre dos segmentos de trabajador la correlación lineal de sus colocaciones con los diferentes segmentos de puesto. En este trabajo, sin embargo, se ha calculado la similitud entre dos segmentos de trabajador a partir de la superposición de sus respectivas distribuciones, que es una medida más ajustada en teoría al concepto de similitud. Además, la superposición entre dos segmentos (o uno menos la superposición) toma valores entre 0 y 1, permitiendo una interpretación más clara de la similitud que la correlación, que puede llegar a ser negativa.

El siguiente gráfico muestra cómo evoluciona la disimilitud cuando el número de *clusters* formados es cada vez menor (partiendo de 3.000 segmentos de trabajador). Esta disimilitud corresponde a la altura en el dendrograma de la última agrupación que se ha generado en el proceso de *cluster* que se ha llevado a cabo. Para nuestro análisis, se ha decidido realizar tres cortes en el proceso de la agrupación: cuando se han formado 2.000 *clusters*, cuando se han formado 1.000 y cuando quedan sólo 500.

Figura 5. Evolución de la disimilitud



Fuente: elaboración propia.

En el gráfico se observa que la disimilitud es cercana a 0,6 cuando hay 2.000 *clusters*. Es decir, la última unión que se ha formado (de las 1.000 acontecidas) lo hace con una disimilitud —ecuación (6)— cercana a 0,6; todas las fusiones anteriores se han dado con un grado de disimilitud inferior. La disimilitud es cercana a 0,8 cuando sólo hay 1.000 *clusters*, y se sitúa por encima de 0,9 cuando sólo quedan 500 *clusters*.

En lo que resta de trabajo vamos a analizar los *clusters* con mayor número de colocaciones cuando se ha llegado al número de grupos citado anteriormente (2.000, 1.000 y 500 grupos); *clusters* donde predominan las colocaciones de enfermería y las relacionadas con el sector de la educación. Asimismo, comentaremos algunos *clusters* adicionales que presentan unas características algo diferentes a lo generalmente observado en los *clusters* anteriores, por cuanto no se corresponden con ocupaciones relacionadas con la sanidad y la educación; los llamaremos *clusters peculiares* (o no habituales).

En un análisis preliminar de los resultados obtenidos, se puede observar que los principales criterios de unión de segmentos de trabajador han sido el sexo —véase Usabiega (2014)— y la localización geográfica; en concreto, cada *cluster* está formado, mayoritariamente, por segmentos de trabajador correspondientes a mujeres de una misma provincia. En cuanto a la edad de los trabajadores, destacan aquellas comprendidas entre 20 y 35 años. Finalmente, como hemos comentado anteriormente, se observa que los *clusters* más importantes se mueven en el entorno de los sectores de educación y sanidad.

A continuación se realiza un estudio más detallado de los tres cortes producidos en el proceso de agrupamiento. Cada segmento de trabajador (de los 3.000 de partida) se ha identificado con un número que representa la ubicación del mismo en la base del dendrograma que genera el programa ejecutado. Esta variable indica, en cierta forma, la similitud que presentan dos segmentos distintos, ya que cuanto más próximos estén en la base del dendrograma antes se producirá la unión de los mismos; ya sea la unión de los propios segmentos o de las agrupaciones de las que formen parte.

## **5.1. Obtención de 2.000 *clusters* de segmentos de trabajador**

Cuando se ha detenido el proceso de agrupamiento en 2.000 grupos, se ha observado que la disimilitud de la última unión de *clusters* ha estado en torno a 0,6. En la tabla 1 se observa que el *cluster* con mayor número de colocaciones es el 1.746 (número asignado por el programa, que tiene que ver con su ubicación dentro del dendrograma), el cual está formado por 55 segmentos de trabajador, que han dado lugar a 3.110

colocaciones. La ubicación geográfica de los segmentos de trabajador está localizada en la provincia de Málaga, con un mayor número de colocaciones en la capital, aunque también se observa un buen número de ellas en los municipios de Fuengirola, Marbella y Mijas —figura 6—. Las personas que han dado lugar a estas contrataciones son, principalmente, mujeres, con una edad comprendida entre 25 y 35 años; el grupo de ocupación mayoritario es «enfermeros», que representa el 97,1 % de las colocaciones en este *cluster*. Además de la información descriptiva contenida en las tablas de esta sección, nuestros datos nos permiten realizar un análisis de la dirección de los flujos de colocaciones observada dentro de cada *cluster*, lo cual nos permite vislumbrar los segmentos de puesto que han recibido colocaciones del *cluster* en cuestión. En este sentido, en este primer *cluster*, podemos destacar que el 78 % de las contrataciones ha tenido lugar en el segmento de puesto situado en Marbella y para la ocupación de «enfermeros».

El siguiente *cluster* se identifica con el número 1.611, y está compuesto por 32 segmentos de trabajador, los cuales han sumado 1.666 colocaciones. Este *cluster* se localiza principalmente en la provincia de Almería, aunque también engloba algunos municipios de la provincia de Granada. Es un *cluster* parecido al anterior, ya que las contrataciones que se observan son mayoritariamente de mujeres (un 86,1 %), en una horquilla de edad de 20 a 30 años, con grupo de ocupación de enfermería y actividades sanitarias y de servicios sociales. Analizando los flujos de colocaciones que generan los segmentos de trabajador del *cluster*, observamos que el flujo más habitual es el de trabajadores que se han dirigido hacia el segmento de puesto con municipio El Ejido y grupo de ocupación «enfermería»; este segmento de puesto capta el 68 % de las colocaciones dentro del *cluster*. El tercer grupo es el 332, que aporta 1.278 colocaciones, que provienen de 10 segmentos de trabajador, los cuales están situados en Sevilla capital. El género femenino sigue siendo mayoritario, ya que está presente en el 70 % de las colocaciones. En un 83,4 % de las colocaciones, la ocupación del trabajador es la de profesores de secundaria. En cuanto al análisis de la dirección de las colocaciones, observamos que estas se han producido principalmente dentro de la ciudad de Sevilla y con el mismo grupo de ocupación (profesores de secundaria) y sector de actividad (educación) que los segmentos de trabajador de origen. El *cluster* 605 es el más heterogéneo de los cinco que se están analizando dentro de los 2.000 grupos, ya que la proporción de hombres y mujeres está más igualada (41 %-59 %), así como la proporción entre los distintos intervalos de edad considerados; además, las ocupaciones de los trabajadores son totalmente distintas a las de los otros cuatro *clusters*: los grupos «técnicos de empresas y actividades turísticas» y «gerentes de pequeñas empresas de restauración» suponen un porcentaje del 50 % de las colocaciones. El total de colocaciones de este *cluster* es de 1.191, colocaciones que han tenido lugar a partir de 24

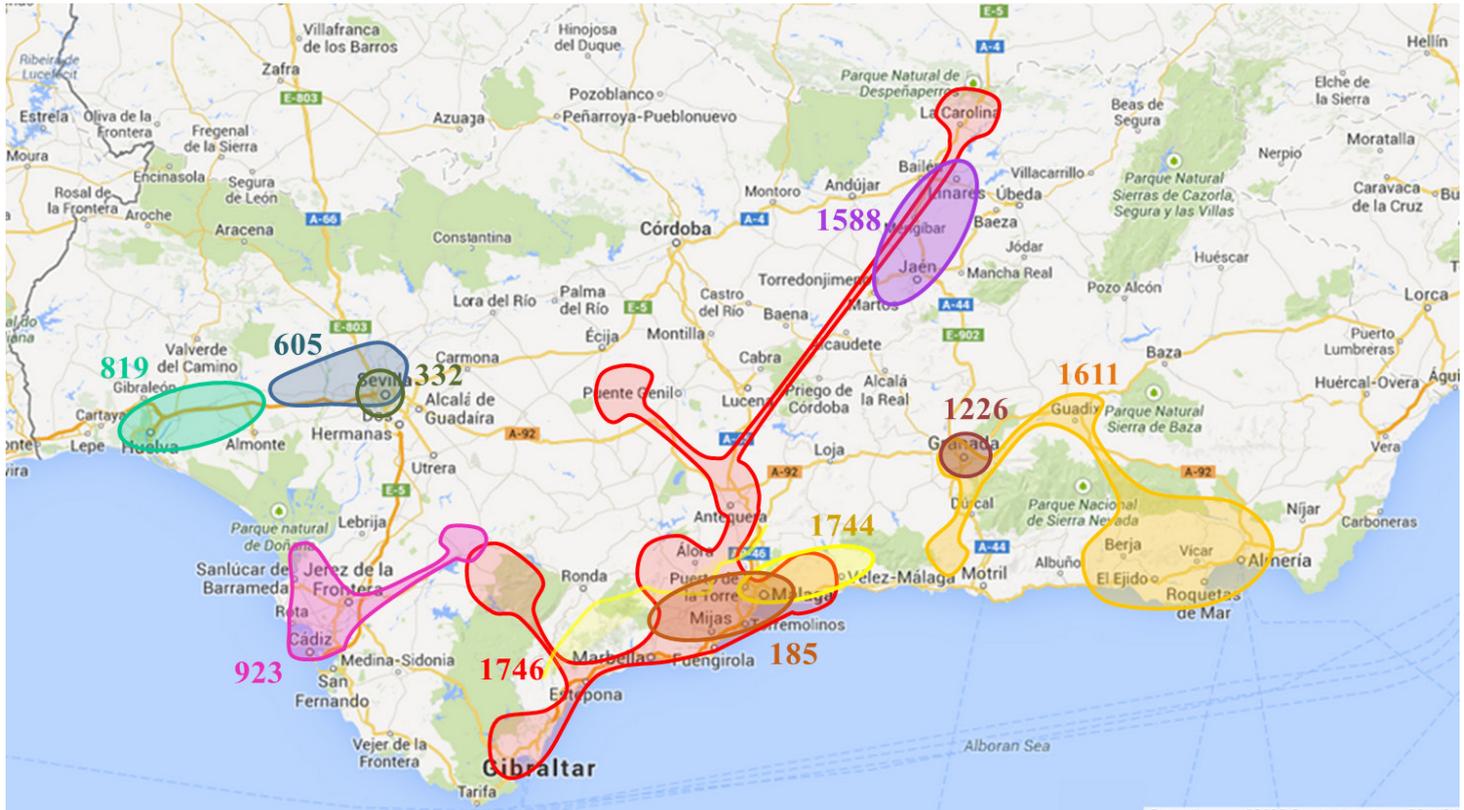
segmentos de trabajador. En este *cluster* más del 90 % de los solicitantes residen en la capital andaluza y sus contrataciones se produjeron mayoritariamente en esta misma ciudad, en el grupo de ocupación camareros, lo cual representa un posible problema de sobre-educación si tenemos en cuenta que los grupos de ocupación de los trabajadores en origen eran de alta cualificación. El quinto *cluster*, en cuanto a número de colocaciones, es el número 1.744 y está formado por 11 segmentos de trabajador y 1.102 colocaciones. El lugar de residencia del 97 % de los trabajadores de este *cluster* se encuentra en Málaga capital, el 71 % son mujeres, y la mayoría son jóvenes, de 20 a 25 años de edad. Asimismo, el grupo de ocupación mayoritario es el de enfermería. La localización de los puestos de trabajo que ocuparon se reparte, fundamentalmente, entre Málaga y Marbella, y la mayor parte en el grupo de ocupación que solicitaban (enfermeros).

Como se ha comentado antes, se han incorporado cinco *clusters* más, que presentan unas características distintas a las descritas anteriormente. En concreto, en la tabla 1 el primer *cluster* «peculiar» es el 923, cuenta con 17 segmentos de trabajador y, aunque también destaca la actividad de educación, los grupos de ocupación que predominan son el de «graduados sociales y asimilados» y el de «arquitectos técnicos». El 46 % de las personas que han sido contratadas (481 colocaciones) se encuentran en un intervalo de edad de 30 a 35 años y los trabajadores (70,1 % de ellos son mujeres) se localizan principalmente en la provincia de Cádiz, el 47 % de ellos en la capital, seguido del 18,7 % en Jerez de la Frontera. Analizando los flujos, se observa que casi el 80 % de los centros de trabajo se encuentran en Cádiz capital, destacando la actividad de educación y la ocupación «otro profesorado técnico de formación profesional». La siguiente agrupación que se muestra se identifica con el número 819 y se ubica en Huelva (60,1 %). El tramo de edad que predomina entre los trabajadores es de 25 a 30 años, siendo el 90,4 % de género femenino. En este caso, la actividad que más destaca es la de educación, y en una proporción parecida también se observa la agricultura. En cuanto a los grupos de ocupación, predominan los «diplomados en educación social» y «arquitectos técnicos». En este grupo, gran parte de las colocaciones han tenido lugar en la capital onubense, destacando la actividad de educación y la ocupación «otro profesorado técnico de formación profesional». En el *cluster* 1.588, los grupos de ocupación que se observan son «psicólogos», «biólogos, botánicos, zoólogos y asimilados» y «diplomados en educación social», en similar proporción; en cuanto a los municipios de residencia de los buscadores de empleo, son Jaén y Linares, entre otros pueblos de la provincia, los que destacan en similar proporción. Por su parte, los sectores de actividad que destacan son «administración pública» y «actividades sanitarias y servicios sociales». Esta última actividad es la que predomina en las colocaciones que se agrupan en este *cluster*, destacando la ocupación «auxiliar de enfermería» y la ubicación de

los centros de trabajo en la ciudad de Jaén. Esta agrupación está formada por 144 colocaciones, que han tenido lugar a partir de seis segmentos de trabajador. El siguiente *cluster* (1.226) está formado por cuatro segmentos de trabajador que han generado 135 colocaciones y se localiza en la ciudad de Granada, lugar donde residían los trabajadores en el momento de la demanda. Todas ellas son mujeres, y más del 80 % con edad comprendida entre 25 y 30 años. Las ocupaciones que destacan, ambas con más de un 30 %, son «archiveros y conservadores de museos» y «otros diversos profesionales del derecho», y las dos actividades que se observan en este *cluster* («otras actividades empresariales» y «comercio minorista») se presentan en una proporción muy parecida. En este *cluster* se vuelve a observar un posible problema de sobre-educación, ya que en los puestos de trabajo que se han generado, principalmente en Granada capital, predomina la ocupación «camareros» y la actividad «comercio minorista». Por último, el *cluster* 185 se localiza en las localidades de Coín y Málaga, y está formado por 117 colocaciones, que se han producido en cinco segmentos de trabajador. De los trabajadores de esta agrupación, el 59,8 % son hombres, especialmente entre 25 y 30 años, con ocupación mayoritaria de «actores y directores» (más del 50 %) y en «actividades de ocio». Se ha observado que en los principales segmentos receptores destaca el pueblo de Coín, como lugar de los centros de trabajo, y básicamente con la misma actividad y ocupación que se registraron en las demandas de empleo.



Figura 6. Mapa de los clusters en la división de 2.000 grupos



Fuente: elaboración propia utilizando Google Maps.

## 5.2. Obtención de 1.000 *clusters* de segmentos de trabajador

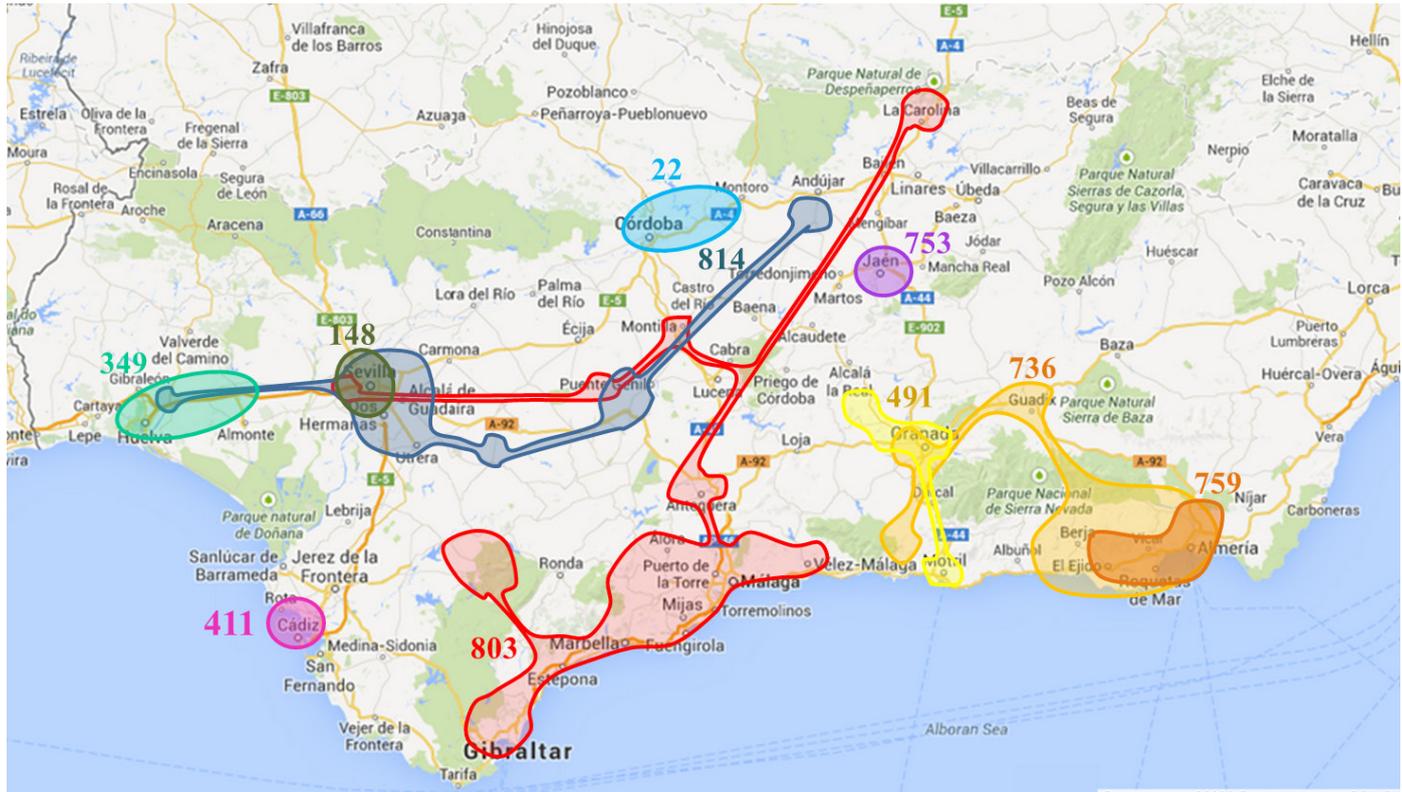
Nuestro análisis se plantea ahora detener el proceso de agrupamiento cuando la base de datos cuenta con 1.000 *clusters* —tabla 2 y figura 7—, habiéndose realizado la última agrupación con una disimilitud en torno a 0,8. Los cinco *clusters* con mayor número de colocaciones reúnen a segmentos de trabajador dentro de los grupos de ocupación de enfermeros y profesores de secundaria principalmente, y cada grupo se localiza en su mayor parte en una sola provincia. Más detalladamente, el *cluster* 803 cuenta con 4.297 colocaciones, que han tenido lugar a partir de 70 segmentos de trabajador diferentes. El 90,8 % de los trabajadores residían en la provincia de Málaga, un 84,6 % son de género femenino, y las edades más habituales son las comprendidas entre 20 y 35 años. El grupo de ocupación más frecuente es el de enfermería. Analizando la dirección de los flujos, se ha observado que el 65 % de las colocaciones ha tenido lugar en Marbella, y de forma secundaria en la capital malagueña, y más del 90 % han sido colocaciones en el mismo grupo de ocupación que el solicitado por los trabajadores («enfermeros»). El segundo *cluster*, con identificador 148, está compuesto por 37 segmentos de trabajador, los cuales han generado 2.581 colocaciones. Los trabajadores residen, en el momento de la búsqueda de empleo, en la provincia de Sevilla, y prácticamente la totalidad en su capital. El 74 % son mujeres, predominando las edades comprendidas entre 25 y 35 años, y el grupo de ocupación de profesores de primaria (67,3 %), seguido de filólogos, intérpretes y traductores; el sector de actividad que destaca es la educación. La mayor parte de las contrataciones han tenido lugar en la ciudad de residencia de los trabajadores en búsqueda de empleo (el 79% se concentran en Sevilla), así como en el mismo sector de actividad que constaba en su solicitud, principalmente educación; aunque el grupo de ocupación que más destaca en los segmentos de puesto es el de profesor de primaria, también aparecen otros, como profesor de universidad o animadores. El *cluster* 491, formado por 45 segmentos de trabajador y 1.788 colocaciones, se localiza en la provincia de Granada, más concretamente en la capital, aunque también en diversos municipios como La Zubia e Íllora. Es una agrupación parecida a la anterior en cuanto al porcentaje de mujeres (74,2 %) y al grupo de ocupación más frecuente, profesor de secundaria, esta vez seguida por psicólogos (12,4 %); por su parte, el sector de actividad que más predomina es educación. En cuanto a la dirección de los flujos de colocaciones, la mayoría de estas personas han encontrado trabajo en su ciudad de residencia, Granada, dentro del sector de actividad de la educación y en el grupo de ocupación de «profesionales de la enseñanza». Por su parte, los *clusters* cuarto y quinto, tienen en común el grupo de ocupación predominante entre sus trabajadores, que es el de «enfermeros». El *cluster* 814 se ubica geográficamente en la provincia de Sevilla, principalmente en la capital, aunque también destaca Dos Hermanas, con un 13,2 % de las colocaciones; un 77 % eran mujeres, y las edades más

observadas se corresponden con el tramo de 20 a 35 años; el 52 % de los trabajadores han encontrado un puesto de trabajo en la capital andaluza y con la misma ocupación que deseaban, la de enfermeros. El número de segmentos de trabajador de este grupo es de 28 y el total de colocaciones que se han producido en este *cluster* ha sido de 1.711. Por otro lado, el *cluster* 736 engloba 33 segmentos de trabajador, que han generado un total de 1.685 contrataciones; el 86 % de ellas han tenido lugar por parte de mujeres, y el tramo de edad de 20 a 30 años representa más del 90 % de las colocaciones del *cluster*. Este *cluster* está localizado en la provincia de Almería, aunque también se han unido segmentos de trabajador cuyo municipio pertenece a la provincia de Granada; la localidad más destacada es Almería capital, seguida de El Ejido. Observando los flujos, se aprecia que la mayoría de las contrataciones han tenido lugar en El Ejido, en el grupo de ocupación «enfermeros».

En relación a los *clusters* «peculiares», el primero que se ha considerado es el número 22, que cuenta con 624 colocaciones, generadas por 25 segmentos de trabajador. Es un grupo que se localiza en la provincia de Córdoba, casi en la totalidad en su capital (96,5 %); el 72,9 % de las personas que lo componen son de género femenino y con edades comprendidas en su mayoría entre 25 y 40 años. En cuanto a los grupos de ocupación, es un *cluster* bastante heterogéneo, ya que no hay ninguno que destaque especialmente; entre otras, se encuentran las ocupaciones de «psicólogos», «agentes de igualdad de oportunidades para la mujer y otros diplomados en trabajo social», «profesionales en recursos humanos», «técnicos de empresas y actividades turísticas» y «otros profesionales de la enseñanza»; en cambio, sí se observa un sector de actividad predominante, el de educación, seguido por el de «actividades asociativas». En el análisis de los flujos, se ha observado que las contrataciones se han producido mayoritariamente en la ciudad de Córdoba, predominando la ocupación «otros profesionales de la enseñanza» y la actividad «educación». El siguiente *cluster*, cuyo identificador es el 759, compuesto por ocho segmentos de trabajador, que han dado lugar a 256 colocaciones, se ubica principalmente en Almería capital, además de en otras tres localidades de la misma provincia; los trabajadores que buscaban empleo se encontraban en una franja de edad de entre 20 y 30 años, y un 76,6 % eran mujeres. Las ocupaciones predominantes son «químicos», «ingenieros técnicos con especialidades agrícolas y forestales» y «profesores de primaria», y el sector de actividad que destaca es «otras actividades empresariales». En cuanto a los flujos, la ciudad que predomina como lugar de los centros de trabajo donde se han generado las contrataciones es Almería, aunque también se han observado otros municipios, pero con un porcentaje de colocaciones mucho menor. Se registraron un 65 % de los puestos de trabajo con la actividad «otras actividades empresariales» y un 46 % en la ocupación «guías». El *cluster* 349, cuya ubicación geográfica se sitúa en la provincia de Huelva, cuenta con ocho segmentos

de trabajador y 232 colocaciones generadas, mayoritariamente por personas de sexo femenino (80,6 %) y más de la mitad con edades comprendidas entre 25 y 30 años. El sector de actividad que más se ha observado es el de educación (44 %), seguido por la agricultura (28,9 %); y en cuanto a grupo de ocupación «diplomados en educación social» (40,1 %), seguido de «arquitectos técnicos» (21,6 %). En cuanto a los puestos de trabajo que se han generado en este *cluster*, destacan aquellos con centro de trabajo en la ciudad de Huelva, con ocupación «otro profesorado técnico de formación profesional» y en el sector de actividad «educación». El cuarto *cluster* «peculiar» (411) se sitúa en la ciudad de Cádiz y está formado básicamente por mujeres (78 %), de 25 a 30 años de edad (83,5 %). Destaca el grupo de ocupación «filólogos, intérpretes y traductores», y el sector de actividad que predomina es «otras actividades empresariales», aunque también aparecen «educación» y «comercio minorista», estos últimos observados en un 27,5 % y un 22 % respectivamente en las 182 colocaciones que se han generado en cuatro segmentos de trabajador. Los flujos se han dirigido básicamente hacia segmentos de puesto situados en la ciudad de Cádiz (67 %), predominando el sector de actividad «educación» y el grupo de ocupación «telegrafistas» (42 %), seguido de «taquígrafos y mecanógrafos» y «profesores de secundaria» (ambos con un porcentaje en torno al 10 %). Por último, el *cluster* 713 se localiza en la ciudad de Jaén. La totalidad de los segmentos de trabajador (cinco segmentos) están formados por mujeres con la ocupación de «psicólogos», destacando las edades comprendidas entre 25 y 30 años (52,9 %). El sector de actividad que predomina es el de «actividades sanitarias y servicios sociales», ya que se ha observado en el 72,9 % de colocaciones. En el análisis de flujos, se ha observado que la mayor parte de las contrataciones, 155 en este *cluster*, han tenido lugar en la capital jienense, destacando la actividad «actividades sanitarias y servicios sociales», entre otras como «educación», «actividades asociativas» y «otras actividades empresariales»; y, en cuanto a los grupos de ocupación, el que más se observa, con un 24 %, es «psicólogos», entre otros, como «profesionales en recursos humanos», «educadores sociales» y «animadores».



Figura 7. Mapa de los *clusters* en la división de 1.000 grupos

Fuente: elaboración propia utilizando Google Maps.

### 5.3. Obtención de 500 *clusters* de segmentos de trabajador

Por último, se ha realizado un análisis de los *clusters* que se han formado cuando se ha detenido el proceso de agrupamiento en un total de 500 *clusters*, dando lugar la última agrupación a una disimilitud o altura en el dendrograma de 0,94 —véase la figura 5—. En estos grupos, la horquilla de edad es más amplia que en los anteriores casos, aunque se sigue observando que en torno al 80 % de los trabajadores son mujeres. El grupo con más número de colocaciones, un total de 9.160, se compone de 183 segmentos de trabajador localizados fundamentalmente en la provincia de Sevilla, y más concretamente en la capital. La mayor parte de las colocaciones corresponden a trabajadores con edad de 25 a 35 años y con ocupaciones de «profesor de secundaria», «psicólogos», «especialistas en métodos didácticos», «profesores de primaria» y «profesores de infantil». Destaca el sector de actividad de educación, seguido de «otras actividades empresariales». Los puestos cubiertos se encuentran principalmente en la ciudad de Sevilla, presentando diversas ocupaciones, como «profesores» o «animadores». En el segundo *cluster*, el 251 (134 segmentos de trabajador), también destaca la actividad de educación y las ocupaciones de profesor de secundaria y primaria. Las edades de las personas que han sido contratadas (4.885 colocaciones) van desde los 20 a los 50 años, y los trabajadores se localizan principalmente en la provincia de Málaga, el 93 % de ellos en la capital. Las colocaciones se han dirigido hacia vacantes situadas sobre todo en la ciudad de Málaga y en el 51 % de los casos el sector de actividad del puesto de trabajo ha sido educación. Los grupos de ocupación de las vacantes han sido más diversos, destacando «profesores de secundaria» y «otros profesionales de la enseñanza». El *cluster* 340 está formado por 72 segmentos de trabajador y 4.414 colocaciones. El 88 % del *cluster* se localiza en la provincia de Málaga, destacando la capital, aunque en torno a un 8,5 % de las colocaciones (en cada caso) se observan en los municipios de Fuengirola, Marbella y Mijas. El 85 % son mujeres, siendo la ocupación mayoritaria la de «enfermeros». En cuanto a los segmentos de puesto, el 64 % de las contrataciones han tenido lugar en Marbella, y el 16 % en la ciudad de Málaga; la ocupación principal ha sido la de «enfermeros». El cuarto *cluster*, identificado con el número 53, es más heterogéneo en cuanto a las ocupaciones de los segmentos de trabajador que lo componen (un total de 62 segmentos), ya que destacan «abogados», «especialistas en métodos didácticos» y «graduados sociales y asimilados», pero aparecen 25 grupos de ocupación más; el sector de actividad que predomina es el de «otras actividades empresariales». La ubicación geográfica de este *cluster* se sitúa en la provincia de Sevilla, y más concretamente en la capital. Del total de colocaciones de este grupo (2.679 colocaciones), el 84 % corresponde a mujeres; y las edades predominantes van desde los 20 a los 40 años. En este grupo, también se observa un posible problema de sobreeducación, ya que las ocupaciones a las que se han dirigido, la mayor parte en la ciu-

dad de Sevilla, han sido de «telefonistas», «taquígrafos y mecanógrafos» y «auxiliares administrativos con tareas de atención al público»; ocupaciones que poco tienen que ver con las más frecuentemente ofertadas por los trabajadores. El *cluster* 336 es muy parecido a otros descritos anteriormente. Está formado por trabajadores del grupo de ocupación de enfermeros. El 91 % de las colocaciones corresponden a mujeres, destacando los trabajadores cuya edad se encuentra entre 20 y 30 años. El *cluster* se localiza en la provincia de Cádiz, y los municipios que destacan son Cádiz, Jerez de la Frontera, San Fernando, El Puerto de Santa María y Sanlúcar de Barrameda. Este *cluster* está formado por 45 segmentos de trabajador, que han generado un total de 2.467 colocaciones. Desde el punto de vista de las vacantes cubiertas, estas colocaciones han tenido lugar en los municipios citados anteriormente, y en un 90 % de ellas la ocupación del puesto fue la de enfermeros.

A continuación se comentan los *clusters* que se han añadido a esta tabla y que no se encuentran entre los grupos con mayor número de colocaciones en esta parada que se ha efectuado en el proceso de agrupamiento. El grupo 124 está formado por 15 segmentos de trabajador, en los que se han creado 476 colocaciones, y se sitúa en la capital onubense. Todos los trabajadores de este *cluster* son mujeres, predominando el tramo de edad de 25 a 35 años, y el 95,4 % de las colocaciones ha sido en el grupo de ocupación «diplomados en educación social»; además, los sectores de actividad que destacan son «actividades sanitarias y servicios sociales», «otras actividades empresariales» y «actividades asociativas». Analizando los flujos, se observa que las contrataciones han tenido lugar, en la mayor parte de los casos, en la ciudad de residencia de los buscadores de empleo, Huelva, y en diversas actividades entre las que destacan «actividades sanitarias y servicios sociales» y «actividades asociativas»; además, alrededor del 30 % de los trabajadores han encontrado empleo en el mismo grupo de ocupación que solicitaban, «diplomados en educación social»; el resto lo ha encontrado en otras ocupaciones diferentes, como «educadores sociales» o «auxiliares de enfermería». El *cluster* identificado con el número 16 se localiza en la provincia de Córdoba, y el 94,1 % de las colocaciones que lo componen (273 colocaciones, en ocho segmentos de trabajador) se sitúan en la capital cordobesa. Este mismo porcentaje es la proporción de mujeres en esta agrupación, prevaleciendo, con un 50,1 %, el tramo de edad de 25 a 30 años, aunque la edad de los trabajadores de este *cluster* no sobrepasa los 35 años. La ocupación que más se ha observado es «técnicos de empresas y actividades turísticas» (49,5 %), y le sigue «biólogos, botánicos, zoólogos y asimilados» (16,5 %); y el único sector de actividad de este grupo es «otras actividades empresariales». En cuanto a la dirección de los flujos de trabajadores, más del 70 % ha encontrado un puesto de trabajo en su localidad de residencia, Córdoba. Los grupos de ocupación que más predominan son «guías» y «dependientes», entre otros muchos, y destaca el sector de ac-

tividad que solicitaban, que era «otras actividades empresariales»; aunque los empleos se han registrados en diversas actividades. El siguiente *cluster* aparece en la tabla 3 con el identificador 292 y se ubica en la provincia de Jaén, siendo el lugar de residencia del 68,1 % de los trabajadores en el momento de la demanda de empleo la capital de la misma. El grupo de ocupación que más destaca es «biólogos, botánicos, zoólogos y asimilados» (34,8 %), aunque también se observa, con un 19,8 %, la ocupación «graduados sociales y asimilados»; en cuanto al sector de actividad, predominan «otras actividades empresariales», «actividades asociativas» y «educación», con porcentajes de un orden de magnitud parecido. En cuanto a los segmentos de puesto hacia los que se han dirigido los trabajadores, predominan aquellos ubicados en la ciudad de Jaén, destacando la actividad «educación» y el grupo de ocupación «otro profesorado técnico de formación profesional». En cuanto al *cluster* 317, se ha observado que las características del mismo son idénticas a uno de los *clusters* analizados anteriormente; en concreto, corresponde al mismo *cluster* que se había formado cuando se detuvo el proceso de agrupación en 1.000 grupos y que se ha identificado con el número 759 en la tabla 2. Resulta interesante comprobar que un *cluster* ya formado anteriormente no ha crecido cuando se ha seguido realizando el proceso de agrupamiento, al no encontrarse otros segmentos de trabajador similares a los que forman esta agrupación; por lo que podría tratarse de un mercado de trabajo local situado en la provincia almeriense, y formado principalmente por trabajadores de los grupos de ocupación «químicos», «ingenieros técnicos en especialidades agrícolas y forestales» y «profesores de primaria», y en el que destaca el sector de actividad «otras actividades empresariales». Por último, el *cluster* 217, localizado en la ciudad de Granada, está formado por trabajadores con edades comprendidas entre 25 y 30 años (78,5 %), repartidos en seis segmentos de trabajadores y que han dado lugar a 228 colocaciones. Las ocupaciones que predominan son «arquitectos técnicos», «archiveros y conservadores de museos» y «otros profesionales del derecho» (cada una con un porcentaje superior al 20 %), y los sectores de actividad registrados son «comercio minorista», «otras actividades empresariales» y «construcción». En cuanto a la dirección de los flujos, se ha observado que la mayor parte de los puestos de trabajo registrados pertenecientes a este *cluster* se han producido en la ciudad de Granada y en el sector de actividad «comercio minorista»; además, se vuelve a observar un posible problema de sobre-educación, ya que un 60 % de las contrataciones ha tenido lugar en el grupo de ocupación «camareros», siguiéndole en número la ocupación «dependientes», aunque en un porcentaje bastante menor.



Figura 8. Mapa de los clusters en la división de 500 grupos



Fuente: elaboración propia utilizando Google Maps.



## 6. Conclusiones

**E**l presente trabajo se basa en la idea de que la heterogeneidad de trabajadores y de puestos vacantes que existe en el mercado de trabajo da lugar a que este funcione de forma segmentada. Tal segmentación consiste básicamente en que los trabajadores de un grupo o segmento específico (definido por la ubicación geográfica y la cualificación de sus trabajadores) tienen una mayor probabilidad de emparejamiento con determinados grupos o segmentos de puesto. En relación con esta idea de segmentación, podemos definir dos variables del mercado de trabajo que se pueden medir de forma empírica y que están relacionadas entre sí: la «propensión al emparejamiento» entre un segmento de trabajador y otro de puesto, y el «grado de similitud» entre dos segmentos de trabajador cualesquiera. La propensión al emparejamiento entre un segmento de trabajador y un segmento de puesto determinados se obtiene como el cociente entre la probabilidad muestral de observar una colocación en el segmento conjunto que forman y la probabilidad de que se produzca dicha colocación en el caso de que el proceso de emparejamiento fuera un fenómeno totalmente aleatorio. Una propensión superior a 1 indicaría que la conexión entre el segmento del trabajador y el del puesto es más fuerte que la que se observaría en un escenario aleatorio de colocaciones, y una propensión inferior a 1 apuntaría hacia lo contrario.

Bajo este enfoque metodológico, este trabajo propone un estudio del mercado de trabajo andaluz de los trabajadores titulados. El análisis se ha basado en la formación de agrupaciones o *clusters* de segmentos de trabajador, creados a partir de una base de datos de colocaciones de trabajadores titulados; colocaciones producidas en la eco-

nomía andaluza durante el período 2007-2010. Dichos segmentos se han definido a partir de las siguientes características del trabajador en el momento de iniciar su búsqueda de empleo: municipio de residencia, sexo, edad, grupo de ocupación y sector de actividad. A partir de aquí, hemos aplicado una metodología de agrupamiento que permite unir a los segmentos de trabajador que más se parecen, para formar grupos o *clusters* mayores (relativamente homogéneos); intentándose identificar de esta forma los mercados de trabajos locales. Todo este proceso de fusiones puede ser seguido a través de un diagrama, que recibe el nombre de dendrograma. De acuerdo con esta metodología, las agrupaciones de segmentos se tienen que realizar en base a una medida de similitud (o de disimilitud, dependiendo del método matemático de agrupamiento), que en nuestro trabajo viene dada por la superposición de las distribuciones de cada pareja de segmentos de trabajador; es decir, dos segmentos de trabajador son más similares cuanto más parecidas sean las distribuciones de sus colocaciones con los diferentes segmentos de puesto con los que se relacionan.

Este trabajo sigue la línea metodológica de los análisis de Álvarez de Toledo *et al.* (2013, 2014), y consigue introducir algunas mejoras en la metodología de *clusters* respecto a dichos trabajos, mejoras que radican principalmente en intentar salvar problemas computacionales propios de esta metodología. Por un lado, se ha calculado la medida de similitud entre parejas de segmentos de trabajador a partir de las superposiciones de sus distribuciones, y no a partir de las correlaciones de sus frecuencias o colocaciones, como se hace en esos trabajos. Por otro lado, el estudio se ha llevado a cabo a partir de un número mayor de segmentos de trabajador y de segmentos de puesto: 3.000 segmentos de trabajador y 10.291 segmentos de puesto, frente a los 1.500 segmentos de trabajador y 1.500 segmentos de puesto, aproximadamente, de los trabajos mencionados.

El proceso de agrupamiento, que se inicia con 3.000 segmentos de trabajador, se ha detenido para su análisis en tres momentos distintos, cuando la suma restante de grupos era 2.000, 1.000 y 500 grupos —cuanto menor es el número de grupos restante, menor es la similitud promedio entre sus elementos (segmentos de trabajador)—. Para el análisis de los resultados obtenidos en cada corte, se han confeccionado unas tablas y unos mapas en los que se han descrito los cinco grupos o *clusters* con mayor número de colocaciones (*clusters* «principales») y otros cinco *clusters* que resultan peculiares, porque sus características son diferentes a las de los *clusters* «principales» en cuanto a la ocupación y ubicación. Se ha observado que los *clusters* «principales» se sitúan, por lo general, dentro de una misma provincia, están formados mayoritariamente por mujeres, el intervalo de edad que más destaca es el de 25 a 30 años y los grupos de ocupación predominantes son los de enfermeros y los relacionados con la educación.

A diferencia de los *clusters* generados en el estudio de Álvarez de Toledo *et al.* (2013), que se refieren a todos los trabajadores (no sólo a los titulados), los *clusters* obtenidos en este trabajo presentan formas más alargadas —son *clusters* donde pueden observarse ciertos municipios alejados de su núcleo principal—. Se podría concluir que los trabajadores titulados, concentrados en nuestro análisis fundamentalmente en los sectores de sanidad y educación, presentan una mayor movilidad que los no titulados, pudiendo ser debido esto a que las bolsas públicas de empleo en dichos sectores, en cierta medida, empujan a los trabajadores a aceptar un empleo en municipios alejados del lugar inicial de residencia de dichos trabajadores.

Además de analizar las características de los trabajadores que se engloban dentro de cada *cluster*, se ha realizado un análisis de los flujos de colocaciones que han tenido lugar dentro de cada *cluster*, desde los segmentos de trabajador hacia los distintos segmentos de puesto; de esta forma, se han podido identificar las características de los puestos de trabajo ocupados por los trabajadores de cada *cluster*. Este análisis de los flujos nos ha llevado a identificar ciertos *clusters* que presentan un posible problema de sobre-educación, ya que las vacantes que se han observado en estas agrupaciones presentaban unas ocupaciones de menor cualificación que las poseídas por los trabajadores que las han ocupado.

La forma en que se analiza la información sobre las transiciones al empleo en este trabajo puede ser útil como una posible «hoja de ruta» para aquellos trabajadores con estudios superiores que se encuentran en proceso de búsqueda de empleo. Identificando en qué *cluster* se encuentra el segmento de trabajador al que pertenece el desempleado (o buscador) y realizando un análisis de los flujos de colocaciones que se han producido dentro de dicho *cluster*, se puede proporcionar información al individuo sobre qué itinerarios de búsqueda han seguido en el pasado otros trabajadores de su perfil que han conseguido emplearse. Proporcionar este tipo de información detallada a los trabajadores en búsqueda de empleo puede suponer una útil política activa del mercado de trabajo, de bajo coste económico. Una posible continuación de este trabajo, en la medida en que se pueda disponer de un intervalo temporal más extenso —pensemos por ejemplo en un ciclo económico completo—, podría consistir en investigar si la división del período total disponible en intervalos temporales menores permitiría captar alguna dinámica de interés en los *clusters* obtenidos, de manera que se pudieran identificar, por ejemplo, *clusters* emergentes («nuevos yacimientos de empleo») y *clusters* que tienden a perder relevancia.



## 7. Bibliografía

- ÁLVAREZ DE TOLEDO, P.; NÚÑEZ, F. y USABIAGA, C. (2008): «La Función de Emparejamiento y el Mercado de Trabajo Español», *Revista de Economía Aplicada*, 16 (48), pp. 5-35.
- (2011): «An Empirical Analysis of the Matching Process in Andalusian Public Employment Agencies», *Hacienda Pública Española*, 198, 3/2011, pp. 67-102.
- (2013): «Análisis “Cluster” de los Flujos Laborales Andaluces», *Revista de Estudios Regionales*, 97, pp. 195-221.
- (2014): «An Empirical Approach on Labour Segmentation. Applications with Individual Duration Data», *Economic Modelling*, 36, pp. 252-267.
- DEVINE, T. J. y KIEFER, N. M. (1991): *Empirical Labor Economics: The Search Approach*, Oxford: Oxford University Press.
- MORTENSEN, D. T. y PISSARIDES, C. A. (1999): «New Developments in Models of Search in the Labor Market», en ASHENFELTER, O. y CARD, D. (eds.): *Handbook of Labor Economics 3B*, Amsterdam: North-Holland, pp. 2567-2627.
- NÚÑEZ, F. y USABIAGA, C. (2007): *La Curva de Beveridge y la Función de Emparejamiento: Revisión de sus Fundamentos Teóricos y de la Literatura Empírica, con Especial Énfasis en el Caso Español*, Sevilla: Centro de Estudios Andaluces, Factoría de Ideas.
- PETRONGOLO, B. y PISSARIDES, C. A. (2001): «Looking into the Black Box: A Survey of the Matching Function», *Journal of Economic Literature*, 39 (2), pp. 390-431.

PISSARIDES, C. A. (2000): *Equilibrium Unemployment Theory*, Cambridge (Mass.): The MIT Press.

PISSARIDES, C. A. (2008): «EconomicDynamics Interviews Christopher Pissarides on the Matching Function», *EconomicDynamics Newsletter & Mailing List*, 10 (1).

USABIAGA, C. (2014) (coord.): *La Segregación Laboral de la Mujer Andaluza. Indicadores, Consecuencias y Medidas hacia la Igualdad de Género*, Sevilla: Centro de Estudios Andaluces.

SHIMER, R. (2007): «Mismatch», *American Economic Review*, 97 (4), pp. 1074-1101.

YASHIV, E. (2007): «Labor Search and Matching in Macroeconomics», *European Economic Review*, 51 (8), pp. 1859-1895.

**DT**

# DEPT



Centro de Estudios Andaluces  
**CONSEJERÍA DE LA PRESIDENCIA**