



**ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA
UNIVERSIDAD DE SEVILLA**



**DEPARTAMENTO DE ORGANIZACIÓN INDUSTRIAL Y
ADMINISTRACIÓN DE EMPRESAS I**

PROYECTO FIN DE CARRERA

**SEGMENTACIÓN LABORAL Y ANÁLISIS DE *CLUSTERS*
CON DATOS INDIVIDUALES. UNA APLICACIÓN AL
MERCADO DE TRABAJO ANDALUZ DE LOS TITULADOS.**

Autora: ANA RODRÍGUEZ MARTÍN

Tutor: FERNANDO NÚÑEZ HERNÁNDEZ

SEVILLA, MARZO 2014

Índice

1. Introducción.....	3
2. Mercado de trabajo, heterogeneidades y segmentación.	5
2.1. Análisis clásico del mercado de trabajo.....	11
2.2. Un modelo del mercado de trabajo basado en los flujos. El modelo de Pissarides.	17
2.2.1. Las políticas del mercado de trabajo.	18
3. Los datos empleados.....	24
4. Metodología para analizar la información sobre el empleo.	25
4.1. Segmentación y propensión al emparejamiento.....	25
4.2. Agrupaciones de segmentos o <i>clusters</i>	27
5. Programación mediante el software STATA de nuestra metodología.	29
5.1. Segmentos laborales y propensión al emparejamiento.....	29
5.2. Cálculo de las superposiciones entre segmentos de trabajador.....	31
5.3. Obtención de <i>clusters</i> de empleo.	35
6. Principales resultados.	38
6.1. Obtención de 2.000 <i>clusters</i> de segmentos de trabajador.....	40
6.2. Obtención de 1.000 <i>clusters</i> de segmentos de trabajador.....	46
6.3. Obtención de 500 <i>clusters</i> de segmentos de trabajador.....	52
7. Conclusiones.....	58
Bibliografía.....	61

1. Introducción.

Actualmente 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 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 parecidos a él han conseguido tener éxito en su proceso de búsqueda de empleo.

Este proyecto parte de una base de datos que contiene 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 sufrir 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. Un elemento clave en el modelado 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 resumido en una función con buenas propiedades matemáticas que da 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 los intermediarios, 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 proyecto 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.

Nuestro proyecto se estructura como sigue. En la sección 2 se ofrece una perspectiva general de la evolución del mercado laboral andaluz en los últimos años. Además, se exponen dos modelos teóricos sobre el funcionamiento del mercado de trabajo: el modelo clásico, basado en variables de nivel (o stock), y un modelo basado en los flujos que se producen en el mercado de trabajo, cuyo principal exponente es el Nobel de Economía Christopher Pissarides. En la sección 3 se realiza una descripción de los datos que se han empleado en la parte empírica del proyecto.

Por su parte, en la sección 4 se desarrolla la metodología empírica que vamos a aplicar al análisis de los datos de colocaciones que poseemos. Las variables descriptivas de la colocación contenidas en nuestra base de 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.

En la sección 5 se procede a programar la metodología de agrupamiento o “*cluster*”, descrita anteriormente, mediante el software econométrico STATA. Una vez que el programa se ha ejecutado sobre nuestros datos, es en la sección 6 donde se analizan los resultados, comentando los grupos o *clusters* más significativos que se han formado en el proceso de agrupamiento. Asimismo, se analiza en esta sección el flujo de colocaciones que se ha producido dentro de cada *cluster*. Todo este análisis permite generar un “mapa” de *clusters* o mercados de trabajo locales que coexisten en el territorio andaluz; información que puede suponer una hoja de ruta a seguir por un trabajador titulado que comience una búsqueda de empleo en Andalucía. Finalmente, la sección 7 expone las principales conclusiones del proyecto.

1. Mercado de trabajo, heterogeneidades y segmentación.

En el mercado de trabajo concurren la oferta y la demanda de trabajo. La oferta de trabajo viene determinada por la población activa, que está formada a su vez por todos los ocupados y todos los desempleados de la economía. La demanda de trabajo, por su parte, la determinan las empresas con el número de puestos vacantes que ponen a disposición de los trabajadores. Ambas fuerzas del mercado dan lugar a la generación de un flujo de colocaciones cada período.

La medición del desempleo¹ en los países occidentales es realizada por las oficinas estadísticas nacionales. En el caso particular de España, es el Instituto Nacional de Estadística (INE) el que se encarga de realizar una encuesta sobre la población activa (EPA), la cual se rige por unos estándares internacionales consensuados en la Organización Internacional del Trabajo (OIT). Según las respuestas aportadas por los encuestados, la población en edad de trabajar puede clasificarse en:

* *Ocupados*: Son las personas de 16 o más años con trabajo por cuenta ajena o por cuenta propia que han estado trabajando de forma continua, esporádica u ocasional, al menos una hora, en la semana de referencia (aquella en la que se pasa la encuesta) a cambio de una retribución o ganancia familiar en metálico o en especie. Además de

¹ Para una visión general del desempleo, véase el trabajo Mankiw (2007).

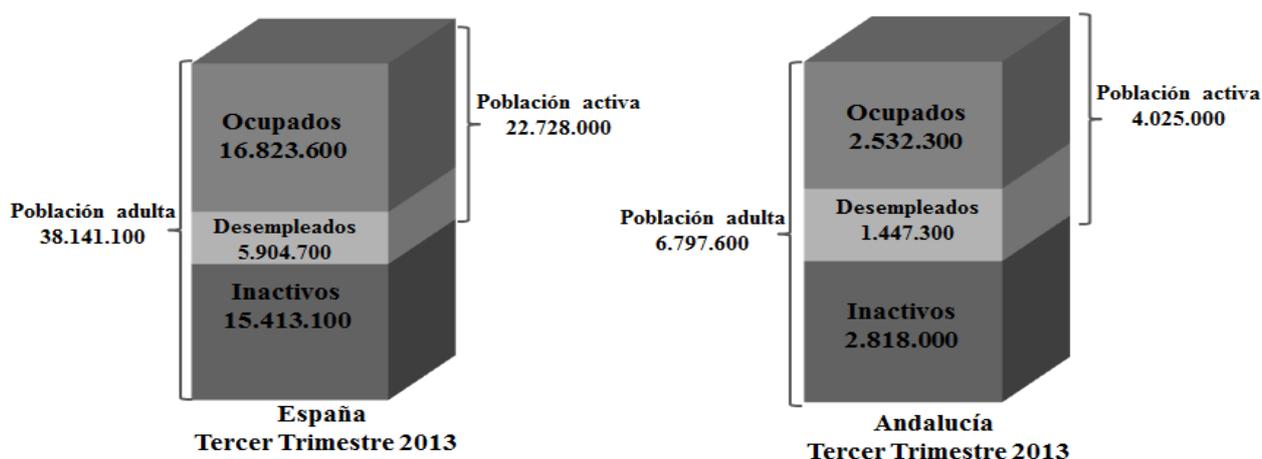
aquellas personas que teniendo un trabajo no han asistido al mismo por vacaciones, enfermedad u otras circunstancias similares.

* *Desempleados*: La definición formal de parado que usa el INE es la definición internacional de la OIT²: personas de 16 o más años, que se hallen sin trabajo durante la semana de referencia, disponibles para trabajar en las dos semanas siguientes a la de referencia y en búsqueda activa de trabajo durante las cuatro semanas que preceden a la de referencia. También se considera desempleado a aquellos trabajadores que hayan encontrado empleo y empiecen a trabajar en un período de tres meses como máximo.

* *Inactivos*: Son aquellas personas de 16 o más años que no se identifican con los anteriores casos, como personas que se dedican a las labores de su hogar, los estudiantes a tiempo completo, los jubilados u otros colectivos que no estén buscando trabajo.

Se puede decir que todas aquellas personas de 16 o más años se pueden identificar con una de las tres categorías mencionadas, además forman el conjunto de la población adulta. Para obtener una visión general de la situación de desempleo en la que se encuentran España y la comunidad autónoma andaluza, se han representado los stocks de cada una de las categorías descritas empleando los datos aportados por el INE en el tercer trimestre de 2013.

Figura 1: Stocks de ocupados, desempleados e inactivos en España y Andalucía.



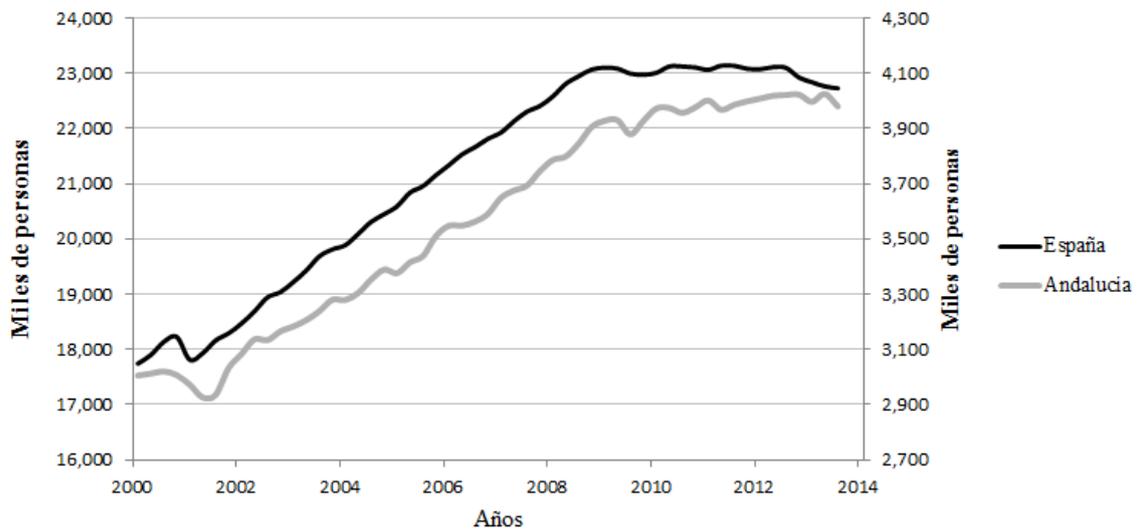
Fuente: Elaboración propia a partir de los datos de la EPA.

² Reglamento N° 1897/2000 de la Comisión de 7 de septiembre de 2000.

Se observa que el mercado de trabajo andaluz representa una parte importante del panorama laboral nacional, ya que la población activa andaluza es más de una quinta parte de la población activa española. La tasa de desempleo, que se calcula a partir de estas variables, como el cociente del número de desempleados entre la población activa, se sitúa en el conjunto nacional en un 25,98%, mientras que en la comunidad autónoma andaluza alcanza un 35,96%.

A continuación, vamos a analizar la evolución de la población activa, el número de vacantes y los parados, así como la evolución de las colocaciones que ha tenido lugar en España y Andalucía en el período 2000-2013.

Figura 2: Evolución de la población activa: 2000-2013.

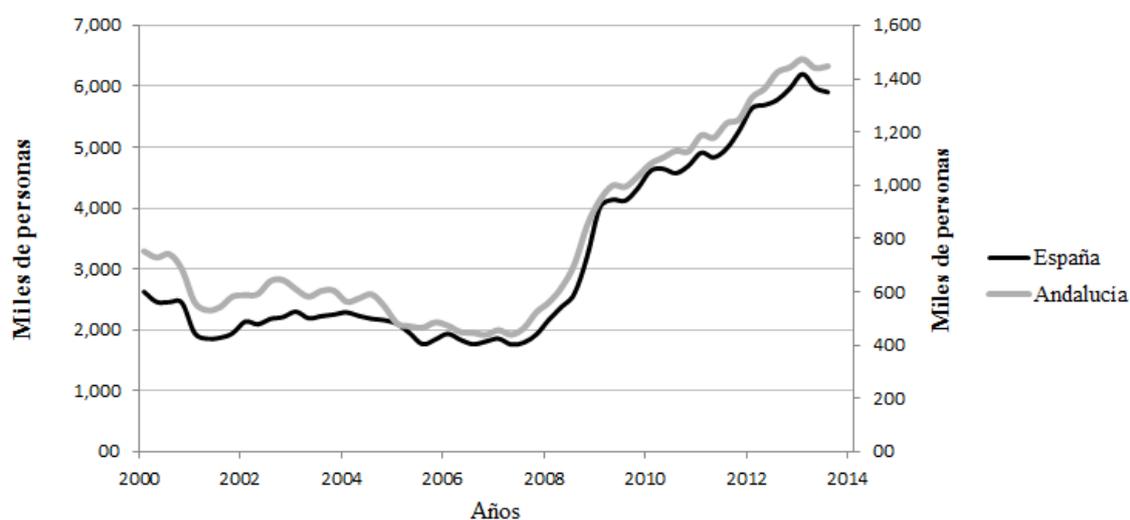


Fuente: Elaboración propia a partir de datos de la EPA.

En primer lugar se muestra la evolución de la población activa en España; se observa un aumento progresivo de la misma desde el año 2001 hasta el año 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 a las personas que han perdido su empleo o aquellas personas que se han incorporado por primera vez al mercado de trabajo y que por dicha crisis ha motivado que la búsqueda de empleo que han llevado a cabo no ha dado sus frutos. En torno al año 2012 comienza una tendencia negativa, descendiendo el número de personas pertenecientes a la

población activa, pudiendo ser debido al desánimo mencionado que 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, los 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 una quinta parte de la población activa de España aproximadamente, como también se observó en los stocks –véase figura 1–.

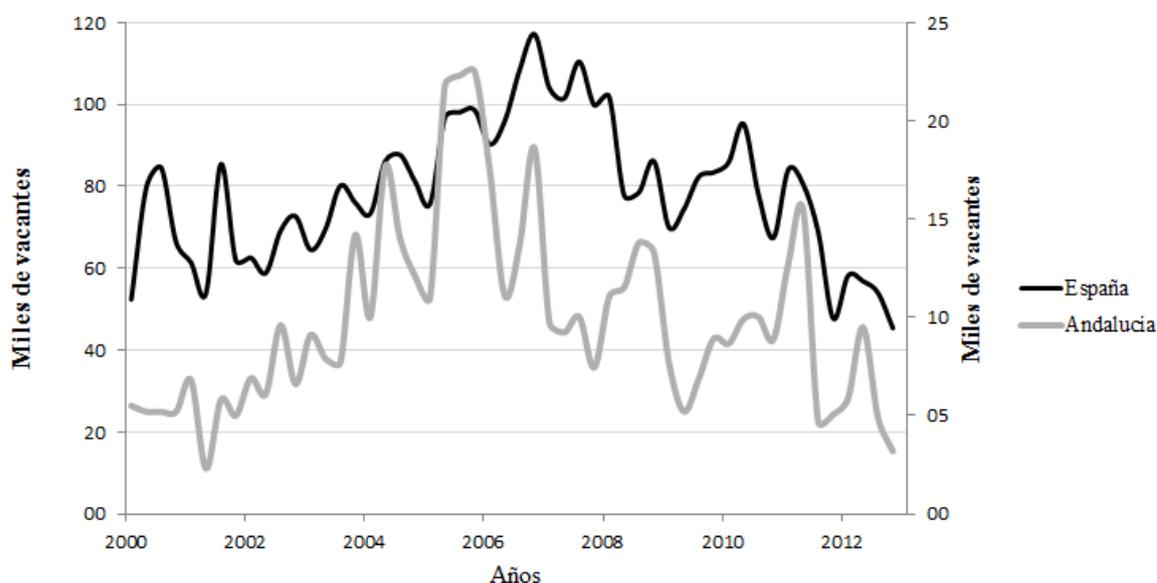
Figura 3: Evolución del número de personas en paro: 2000-2013.



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 6.202.700 desempleados, mientras que en la comunidad autónoma andaluza esta cifra se ha situado en 1.473.700 desempleados.

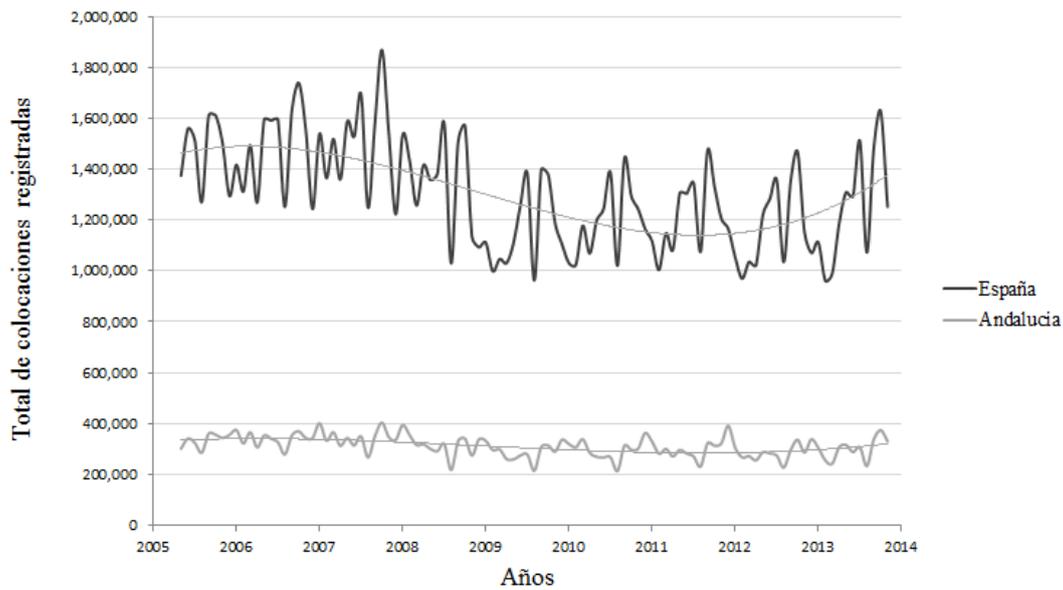
Figura 4: Evolución del número de vacantes: 2000-2013.



Fuente: Elaboración propia a partir de datos de la *Encuesta de Coyuntura Laboral*.

La tendencia del número de vacantes en España es aproximadamente inversa a la mostrada en el gráfico anterior –véase la figura 3–, 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 un nivel de 46.000 vacantes, cifra menor que los puestos de trabajo disponibles al comienzo del año 2000. En Andalucía, 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 del 2005, es decir, dos años antes. Como se observa, ambos trazados no son continuos, sino que se producen grandes variaciones de un trimestre a otro consecutivo.

Figura 5: 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. Para España se puede observar una tendencia negativa desde mayo del 2006 hasta mediados del 2011, momento en el cual las colocaciones empezaron a aumentar, siguiéndose esta tendencia positiva hasta el último año representado. En la figura se aprecian oscilaciones bastante pronunciadas entre dos trimestres consecutivos, por lo que se ha optado por representar, además, una línea de tendencia, para observar con claridad el nivel medio de las colocaciones registradas en las oficinas públicas de empleo. Por su parte, pese a no verse bien reflejada la tendencia de las colocaciones (no se ha representado en el eje derecho por claridad de los trazados), Andalucía ha seguido una evolución paralela a la llevada a cabo por las colocaciones en el territorio español. Las contrataciones en Andalucía se han mantenido en torno a las 300.000.

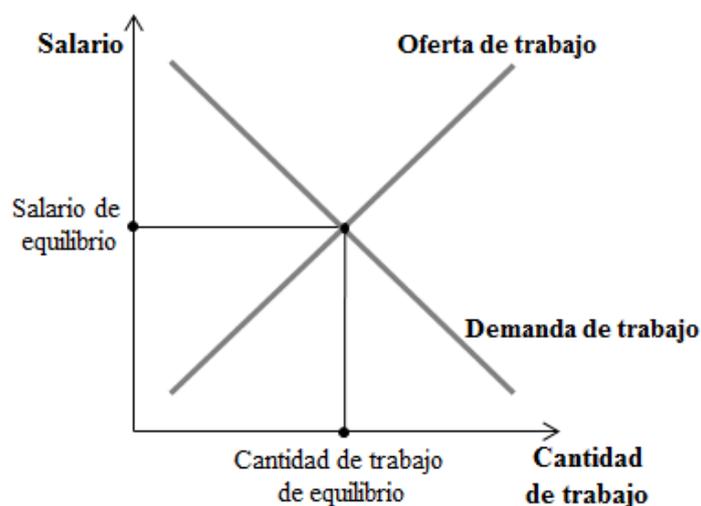
Una vez que hemos descrito el escenario empírico en que nos movemos, pasamos a analizar el mercado de trabajo desde la perspectiva teórica; en este sentido, podemos distinguir dos enfoques teóricos del mercado de trabajo: el clásico, que está basado, principalmente, en los stocks de individuos ocupados, desempleados e inactivos, y un enfoque más moderno, que analiza el mercado, principalmente, desde la perspectiva de los flujos que en él acontecen; este segundo enfoque va a ser la referencia teórica fundamental del proyecto.

2.1. Análisis clásico del mercado de trabajo.

La visión neoclásica del mercado de trabajo se basa en la Teoría del Equilibrio General Walrasiano, introducida por Léon Walras en 1874. Esta teoría intenta dar explicación al hecho de que se podía llegar a un orden social en una economía de mercado a partir de agentes de mercado que se mueven por sus propios intereses³. Esto sucedería a nivel laboral en un mercado de trabajo en competencia perfecta en el que la información entre desempleados y empleadores es perfecta y el trabajo es un bien homogéneo⁴.

En el modelo de la competencia perfecta, se representan la oferta y la demanda de trabajo como sendas funciones en un espacio bidimensional de salario real frente a cantidad de trabajo. La oferta es una función con pendiente positiva: a salarios altos, el número de personas que se ofrecen para trabajar con esta retribución es mayor que si perciben menos dinero. Las empresas son las que determinan la demanda de trabajo, la cual se representa por una función de pendiente negativa, ya que cuando el mercado sitúa el salario a un nivel bajo, las empresas están dispuestas a contratar a más personas; por el contrario, a salarios altos, la cantidad de trabajo que demandan es menor.

Figura 6: Modelo de competencia perfecta.



Fuente: Elaboración propia.

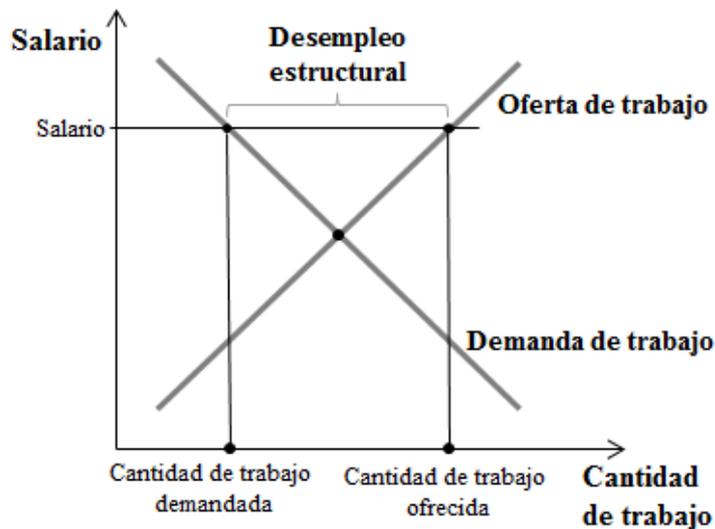
³ Esta idea fue introducida por Adam Smith en su obra *La Riqueza de las Naciones*, 1776.

⁴ Las deficiencias de la Teoría del Equilibrio General Walrasiano para responder a la pregunta de Smith se analizan en el ensayo de Frasser (2008).

El salario debe ajustarse hasta que oferta y demanda se igualen, lo cual sucede en el punto de equilibrio, donde no hay exceso de oferta ni de demanda de trabajo. Dicho nivel salarial es llamado el salario de equilibrio.

Pero en el mercado de trabajo pueden existir varias causas que impiden que en el mercado se alcance el equilibrio; en concreto, puede haber factores que dan lugar a la existencia de un exceso de oferta de trabajo, donde no hay suficientes puestos para todos los trabajadores que buscan empleo. Este exceso de oferta de trabajo es denominado “desempleo estructural” y es provocado, en definitiva, porque el salario que hay en el mercado de trabajo es superior al salario de equilibrio.

Figura 7: Modelo de competencia perfecta con exceso de oferta.



Fuente: Elaboración propia.

Las razones fundamentales por las que el salario vigente en el mercado de trabajo se puede situar por encima del de equilibrio son:

- a) La legislación sobre el salario mínimo

El Estado fija el salario más bajo que pueden pagar los empresarios a los trabajadores, pero en ocasiones dicho salario podría ser superior al propio de equilibrio. Esto podría suceder con los trabajadores menos cualificados y con menor experiencia. La oferta de este tipo de trabajadores es alta, mientras que las empresas no ofertan

muchos puestos de trabajo para estas personas con menor formación, por lo que el salario de equilibrio puede llegar a ser realmente bajo. Que el Estado garantice un salario mínimo beneficia a estos trabajadores, los cuales podrían estar dispuestos a aceptar un salario menor, pero a cambio provoca desempleo en el colectivo.

b) Los sindicatos y la negociación colectiva

Los sindicatos son agrupaciones de trabajadores cuyo objetivo es la obtención de unos fines comunes y la defensa de sus intereses. Estas asociaciones negocian con las empresas en cuestiones relativas a los salarios y a las condiciones laborales. A este proceso se le conoce como negociación colectiva.

La forma principal de presionar a los empresarios es mediante la retirada organizada de trabajo. A este poder de los sindicatos se le denomina huelga, y provoca un descenso en la producción, las ventas y los beneficios, por lo que los empresarios se ven obligados a aceptar el pago de sueldos más elevados con respecto a los salarios que pagarían si no existiesen los sindicatos.

No existe unanimidad de opiniones sobre la bondad o no de la actividad de los sindicatos sobre la economía. Por un lado, los salarios son mayores que el salario de equilibrio, lo cual es una causa de desempleo; al subir el salario, las empresas desean contratar menos mano de obra. Por tanto, se ven perjudicados los trabajadores que han perdido su empleo, y solo se ven beneficiados los que siguen conservando su trabajo, con un salario mayor que el existente antes de la negociación. Esto supone una crítica a estas asociaciones, ya que aumentar el desempleo implica que la cantidad de trabajo disponible aumenta; esto puede provocar un desplazamiento de la función de oferta de trabajo hacia la derecha en otros mercados laborales no sindicados, por lo que se reducen los salarios de parte de la economía.

Por otro lado, los defensores de los sindicatos sostienen que son importantes para ayudar a las empresas a conocer y tratar con más eficacia las preocupaciones de sus trabajadores, y proporcionar así unas condiciones de trabajo lo más beneficiosas posibles para ambas partes.

c) Los salarios de eficiencia.

A las empresas les interesa que la productividad de sus trabajadores sea alta, para que su producción, sus ventas y sus beneficios sean mayores. Una mayor productividad del trabajador puede lograrse a través del pago de unos salarios más altos, lo cual puede deberse a varios motivos:

c.1) La salud de los trabajadores.

En los países en desarrollo el salario de equilibrio es más bajo que en los países desarrollados, debido a que la formación de los trabajadores es menor, los recursos con los que trabajan son inferiores, y a que existe indefensión jurídica por parte de los trabajadores. Con unos salarios demasiado bajos, los empleados no pueden llevar una dieta nutritiva, debido a su bajo poder adquisitivo, por lo que la salud de estas personas puede verse afectada, así como su productividad, ya que sus condiciones físicas no son las más idóneas para el desarrollo normal de su actividad. El incentivo para pagar salarios más altos es el de conseguir una dieta adecuada para los trabajadores, así como otros mínimos de salubridad, etc., lo que debe mejorar significativamente su productividad.

c.2) La rotación de los trabajadores.

Los trabajadores pueden tener varios motivos que les lleven a cambiar de empleo. Ante la situación de incertidumbre del mercado laboral actual, es lógico que los empleados piensen en un cambio de trabajo ante la posible pérdida de su empleo, al observar que sus empresas no tienen los beneficios esperados. Otra causa es la posible mejora salarial y de las condiciones del nuevo puesto de trabajo, así como un mayor desarrollo profesional trabajando en algún proyecto que suponga un paso adelante en la carrera profesional del trabajador.

A las empresas no les interesa que haya muchos cambios en su plantilla, ya que ello implica unos gastos adicionales asociados al coste de contratar y formar a nuevo personal, y una pérdida de productividad al inicio de los nuevos contratos, por lo que están dispuestos a pagar sueldos más elevados para que la rotación sea menor.

c.3) El esfuerzo de los trabajadores.

Los trabajadores deciden en cierta medida el esfuerzo que realizan en su puesto de trabajo, y las empresas controlan estos esfuerzos despidiendo a aquellos trabajadores que eluden su responsabilidad. Pero la supervisión del empresario no es perfecta y tiene un coste asociado, por lo que unos salarios más elevados pueden ser la forma de motivar a los empleados para realizar con más eficacia su trabajo, ya que sus sueldos están por encima del salario de equilibrio; si son sorprendidos sin esforzarse son despedidos y tienen más que perder en términos salariales. Este razonamiento resulta más convincente en mercados con un cierto desempleo, ya que en el caso de no haber exceso de oferta de trabajo, los trabajadores tendrían menos razones para esforzarse, ya que si pierden su puesto actual, podrían conseguir otro al salario de equilibrio.

c.4) Calidad de los trabajadores.

Las empresas tratan de tener en su plantilla a los trabajadores más competentes y con más formación. Por tanto, ofrecer puestos de trabajo con salarios más altos, llamaría la atención de este tipo de trabajadores, que solicitarían estas vacantes; es decir, no sólo acudirán malos candidatos a interesarse por el puesto, sino también buenos candidatos. Una segunda tarea del empresario será la de identificar al buen trabajador de entre todos los candidatos existentes.

Por tanto, podemos concluir que pagar un salario más alto que el salario de equilibrio puede dar lugar a una ganancia de productividad, aunque desgraciadamente, todos estos argumentos no son muy tenidos en cuenta por las empresas españolas y andaluzas. Al igual que en los casos anteriores de salario superior al mínimo y salario sindical, este mayor nivel salarial provoca que la oferta de trabajo sea superior a la demanda de trabajo, generando desempleo; pero la diferencia con respecto a las anteriores causas de desempleo reside en que ahora son las propias empresas las interesadas en pagar un salario más alto.

d) El desempleo de búsqueda o friccional.

Aunque el modelo walrasiano contempla la idea del pleno empleo cuando el mercado se sitúa en el punto de equilibrio, sin que operen las causas anteriores, la realidad del mercado de trabajo es que siempre hay personas que se encuentran desocupadas, incluso cuando el mercado se encuentre situado en el punto de equilibrio. Esto se debe al hecho de que los trabajadores son heterogéneos entre sí y a que lo mismo sucede con los puestos vacantes; rompemos, por tanto, el supuesto walrasiano de trabajadores y puestos homogéneos. La causa de este desempleo “friccional” es la llamada “búsqueda de empleo” o proceso no inmediato de emparejamiento entre los trabajadores y las vacantes por el que los trabajadores encuentran el empleo que mejor se adecúa a sus preferencias y aptitudes. En el equilibrio del mercado se cumple la siguiente igualdad entre oferta (L^s) y demanda (L^d) de trabajo:

$$L^s = N + U = N + V = L^d \quad (1)$$

siendo N el número de empleados, U el de desempleados y V el volumen de puestos de trabajo vacantes. Simplificando en ambos términos de la igualdad, tenemos que en el equilibrio se cumple $U = V$. Hay puestos de trabajo para todos los trabajadores demandantes de empleo, pero al existir un periodo de búsqueda de empleo, se genera dicho desempleo friccional⁵.

El modelo walrasiano no es suficiente para explicar los mercados de trabajo reales, donde el intercambio de información entre trabajadores y empresas no es perfecto y en el que el trabajo no es un bien homogéneo. Por estas razones, vamos a analizar un nuevo modelo que analiza el mercado laboral desde la perspectiva de flujos, e introduce la influencia de las heterogeneidades⁶ y el tiempo de búsqueda en el proceso de emparejamiento entre trabajadores y puestos de trabajo.

⁵ Imagínese un mercado de trabajo que estuviera generando vacantes en el norte del país y donde la mayor parte de los desempleados se encuentran en el sur, buscando dichos puestos.

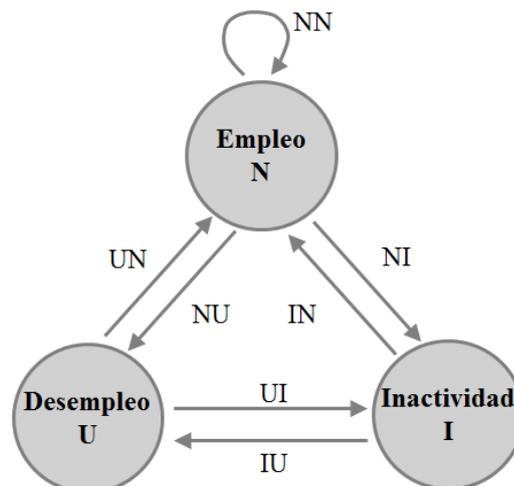
⁶ Véanse los trabajos de Petrongolo *et al.* (2001) y Yashiv (2007), en los que se realiza un análisis de las heterogeneidades en el mercado de trabajo.

2.2. Un modelo del mercado de trabajo basado en los flujos. El modelo de Pissarides.

El análisis de los flujos en el mercado de trabajo ha recibido atención en las últimas décadas debido, en primer lugar, a los estudios empíricos realizados por Davis y Haltiwanger (1992), que analizan los flujos del mercado laboral americano, y en segundo lugar, a una nueva generación de modelos teóricos, entre los que destaca el modelo de Blanchard y Diamond (1989) y el modelo de Pissarides (2000); ambos modelos tratan de determinar el nivel de desempleo de equilibrio y de analizar el impacto de las instituciones laborales en el mercado de trabajo.

En el mercado de trabajo existe un continuo movimiento de trabajadores entre las tres situaciones en las que se puede clasificar la población mayor de 16 años (empleo, desempleo e inactividad), y realizar un análisis de los flujos del mercado de trabajo da una idea más real de lo que ocurre dentro de él. La tasa de desempleo es una medida del porcentaje de desempleados que existe en un mercado de trabajo, pero una misma tasa de desempleo puede manifestar dos perspectivas muy diferentes: un mercado de trabajo dinámico, en el que se producen numerosas entradas y salidas del desempleo; o un mercado de trabajo inactivo, en el que el número de desempleados apenas varía porque no hay movimiento del desempleo al empleo ni viceversa. En la siguiente figura se esquematizan los flujos que se producen en el mercado laboral.

Figura 8: Flujos en el mercado laboral.



Fuente: Elaboración propia según las categorías laborales.

Los flujos de entrada al empleo están formados por trabajadores ocupados que cambian de empleo, desempleados que encuentran trabajo y por desempleados desanimados que no realizan búsqueda de trabajo activa, es decir, son individuos que se encuentran en la situación de inactividad según la EPA, pero que si se les ofreciera directamente un empleo, lo aceptarían.

Por otro lado, los flujos de salida del empleo se componen de personas que transitan hacia el desempleo o hacia la inactividad. Los movimientos empleo-desempleo se deben a diversas causas: despido, vencimiento de contrato y abandono voluntario para buscar otro trabajo, para conseguir una mejora salarial o un mayor desarrollo profesional. Los movimientos empleo-inactividad tampoco son despreciables: jubilaciones, incapacidades, desánimo, etc.

Los flujos brutos del mercado de trabajo pueden estimarse, aunque no es una tarea sencilla, ya que no todos los flujos se miden empíricamente. Para los flujos que sí se miden existen fuentes de información diversas con diferentes metodologías, como los datos aportados por el Servicio Público de Empleo Estatal (SPEE-INEM) o los resultados ofrecidos por la EPA (INE)⁷.

2.2.1. Las políticas del mercado de trabajo.

En este proyecto, 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 sobre candidatos y vacantes idóneos para cada buscador hace que el periodo de tiempo necesario para que se dé 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. Imaginemos, por ejemplo, que

⁷ Puede ver en los trabajos de Estrada *et al.* (2002) y García *et al.* (2007) información relativa a los flujos de trabajadores en España así como una metodología propuesta para el cálculo de los flujos brutos que se producen entre las distintas categorías laborales a partir de datos de la EPA y del INEM.

se crea un nuevo puesto de trabajo y que la empresa que lo ofrece la publica sólo a nivel local; entonces, aquellas personas que busquen empleo en una ubicación geográfica diferente no tienen accesible dicha información de la existencia de la vacante; además, incluso si la empresa amplía su área de búsqueda y algunos trabajadores solicitan entonces dicho puesto, puede que no cumplan los requisitos del mismo o que las condiciones de trabajo no les interesen, lo que supondría seguir el proceso de búsqueda por ambas partes.

Otro aspecto importante de la búsqueda de empleo es que lleva asociados unos costes que afectan tanto a las personas que buscan trabajo –por el transporte hacia el lugar de las entrevistas de trabajo, adquirir la formación necesaria para poder solicitar puestos de mayor rango o con unos requisitos establecidos (por ejemplo, el nivel de idiomas), etc.–, como a las empresas –por la realización de entrevistas, por la formación de los candidatos una vez seleccionados, etc.– A estos costes hay que añadir el propio tiempo de búsqueda que unos y otros tienen que asumir. La existencia de tales costes otorga un cierto poder de negociación a ambas partes: por un lado, la empresa puede ofrecer un salario más bajo al trabajador, que éste podría aceptar si no quiere tener que invertir más tiempo y dinero en la búsqueda de empleo; por otro lado, el trabajador sabe que si la empresa lo rechaza tendrá que asumir los costes de la búsqueda de un nuevo trabajador, y esto puede llevarle a pedir un salario algo mayor. El que tenga mayor poder de negociación de los dos será el que tenga una mayor influencia en el salario finalmente pactado.

El modelo de Pissarides (2000) sintetiza todo este proceso de búsqueda 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 posibles 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*”; este término hace referencia al desajuste existente entre las cualificaciones y de localización espacial de trabajadores y puestos de trabajo, existiendo más tipos de desajustes. Si en el mercado de trabajo no existiese *mismatch*,

las vacantes y los trabajadores se encontrarían de forma inmediata, no existiendo una función de emparejamiento como tal.

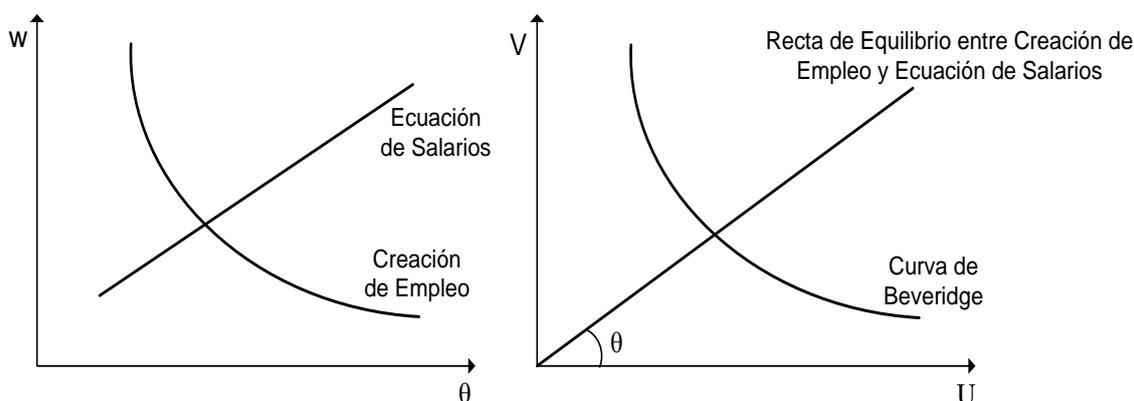
La función de emparejamiento es posiblemente el elemento central del modelo del desempleo de equilibrio de Pissarides (2000), el cual constituye un buen punto de partida para analizar el efecto de las políticas de empleo sobre determinadas variables del mercado de trabajo. En concreto, dicho modelo contiene tres ecuaciones fundamentales que permiten determinar las variables desempleo (U), vacantes (V), ratio vacantes-desempleo (θ) y salario real (w); estas ecuaciones son la ecuación de salarios, la ecuación de creación de empleo y la relación teórica de equilibrio entre vacantes y desempleo, conocida con el nombre de curva de Beveridge, relación donde aparece implícita la función de emparejamiento, ya que en cada punto de la curva de Beveridge se cumple la igualdad entre los flujos de entrada y salida al desempleo, y dicho flujo de salida se puede aproximar a través de dicha función.

La ecuación de salarios se representa en el espacio (w, θ) con pendiente positiva: cuando el ratio de vacantes-desempleo aumenta, la tensión del mercado de trabajo aumenta en el sentido de que las empresas compiten por mano de obra relativamente escasa, por lo que el salario tenderá a subir; los trabajadores tienen poder de negociación. Por su parte, la ecuación de creación de empleo tiene pendiente negativa en dicho espacio (w, θ): a menores salarios reales, mayor será el deseo de las empresas de contratar mano de obra, es decir, mayor será la oferta de vacantes por cada trabajador desempleado. En cuanto a la curva de Beveridge, ésta se representa en el espacio (V, U) con pendiente negativa: cuando en el mercado de trabajo aumentan las vacantes, aumenta también el número de emparejamientos o flujo de salida del desempleo, cayendo el nivel de desempleo y aumentando el nivel de empleo dada una población activa; al aumentar el empleo aumenta también el flujo de entrada al desempleo –definido como un porcentaje del volumen de empleo–, el cual finalmente vuelve a igualarse con el flujo de salida pero para un nivel de desempleo menor. Finalmente, el resultado del equilibrio entre las ecuaciones de salarios y de creación de empleo puede

ser representado en el espacio (V, U) como una recta de pendiente θ que pasa por el origen⁸.

El modelo se puede resumir en dos gráficos: el primero muestra el equilibrio en el espacio (w, θ) a través de la interacción de la condición de creación de empleo y de la ecuación de salarios; el segundo representa la curva de Beveridge y la condición de creación de empleo en el espacio (V, U) .

Figura 9: Representación gráfica del modelo.



Fuente: Pissarides (2000).

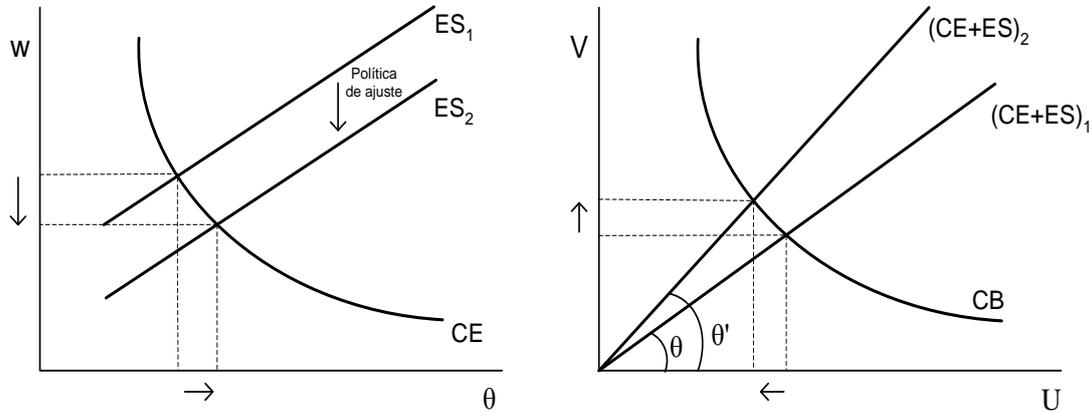
Una de las utilidades de este modelo es que permite analizar de forma gráfica el efecto de diversas políticas de empleo, tales como las políticas de moderación salarial, las de fomento del empleo o las de intermediación laboral. Dichas políticas se apuntan en el trabajo de Álvarez de Toledo *et al.* (2013). En esta sección, sin embargo, ampliaremos, desde el punto de vista gráfico, el análisis desarrollado en dicho artículo.

La figura 10 muestra el efecto de una política de moderación salarial –como, por ejemplo, la existencia de una reforma laboral que reste poder de negociación a los trabajadores–. Si dicha política es efectiva, hará caer a la ecuación de salarios –se produce una caída del salario dado un determinado grado de escasez de vacantes–; la existencia de salarios más bajos provocará un movimiento a lo largo de la curva de creación de empleo hacia mayores niveles de vacantes ofrecidas por trabajador desempleado. En la parte derecha de la figura, la recta CE+ES rotaría en el sentido

⁸ Una versión más detallada del modelo se puede encontrar en el texto de Pissarides (2000) –véase también al respecto Núñez y Usabiaga (2007)–.

contrario a las agujas del reloj, lo que conduciría finalmente a un mayor nivel de puestos vacantes y un menor nivel de desempleo.

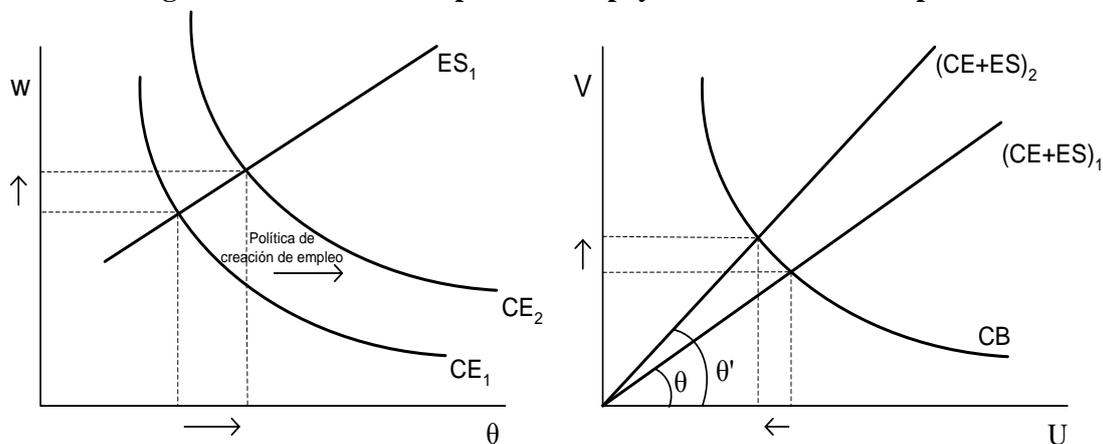
Figura 10: Efecto de una política de moderación salarial.



Fuente: Elaboración propia a partir del modelo de Pissarides (2000).

La figura 11 analiza el efecto de las políticas de apoyo a la creación de empleo –como, por ejemplo, las políticas de localización industrial–. Políticas de este tipo deben mover a la ecuación de creación de empleo hacia la derecha, dando lugar a mayores valores del ratio vacantes-desempleo para cualquier salario establecido. La congestión que se genera entre las empresas a la hora de contratar a trabajadores, ahora más escasos, dará lugar también a salarios mayores en la nueva situación de equilibrio. En la parte derecha de la figura, la recta CE+ES girará hacia la izquierda, lo que conduce finalmente a un mayor número de vacantes y un menor número de trabajadores en desempleo.

Figura 11: Efecto de una política de apoyo a la creación de empleo.

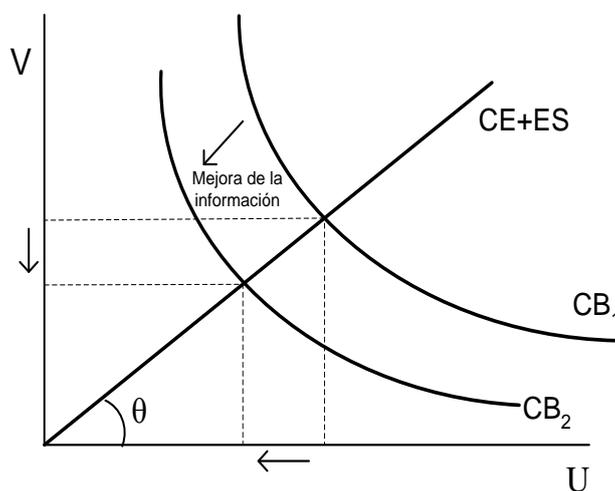


Fuente: Elaboración propia a partir del modelo de Pissarides (2000).

Finalmente, la figura 12 analiza el efecto de introducir una mejora de la información sobre el empleo que necesitan empresas y trabajadores; mejora hacia la que va orientada la herramienta empírica que desarrollamos en este proyecto, en las siguientes secciones.

La mejora de la información actúa en el modelo de Pissarides a través de la curva de Beveridge, y más concretamente a través del flujo de salida del desempleo que se encuentra implícito en dicha curva⁹. Cuando mejora la información que necesitan los trabajadores y las empresas para conseguir una colocación se produce una mejora en la eficiencia del proceso de emparejamiento laboral, ya que se consigue generar un mayor número de colocaciones por período con el mismo número inicial de buscadores y vacantes, por lo que descienden los niveles de desempleo y de puestos vacantes –es como si se hubiera producido un avance tecnológico en el proceso de emparejamiento–; este hecho se refleja gráficamente en la figura 12 a través de un movimiento de la curva de Beveridge hacia el origen, con el correspondiente movimiento descendente a lo largo de la recta CE+ES.

Figura 12: Efecto de una mejora de la información sobre el empleo.



Fuente: Elaboración propia a partir del modelo de Pissarides (2000).

⁹ La búsqueda de microfundamento para una función como ésta ha sido algo habitual en la literatura. Una panorámica sobre la función de emparejamiento y su microfundamento se puede encontrar en los trabajos de Petrongolo *et al.* (2001) y de Núñez *et al.* (2007) –este último más centrado en la economía española–.

3. Los datos empleados.

La base de datos empleada en este proyecto recoge una muestra de colocaciones producidas en el territorio andaluz durante el período 2007-2010; por tanto, 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–. Para cada colocación observada tenemos información sobre las características de las dos partes que se emparejan: el trabajador que buscaba empleo y el puesto pendiente de cubrir.

La base de datos está formada por datos individuales, por lo que su grado de detalle es bastante mayor que si tratásemos con datos agregados¹⁰. 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 siguiendo las propias clasificaciones creadas por el INE, de esta forma, el sector de ocupación se ha codificado según la Clasificación Nacional de Actividades Económicas (CNAE), 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 municipios, provincias, comunidades y ciudades autónomas y sus códigos. Por otro lado, se ha considerado necesaria la discretización de la variable edad, en tramos de 5 años; de esta forma, evitamos tener una fila distinta por cada año de edad. En concreto, los intervalos que se han usado son:

Tabla 1: Intervalos de edad.

Intervalos	Valor variable	Intervalos	Valor variable
16 < edad < 20	20	45 < edad < 50	50
20 < edad < 25	25	50 < edad < 55	55
25 < edad < 30	30	55 < edad < 60	60
30 < edad < 35	35	60 < edad < 65	65
35 < edad < 40	40	65 < edad < 70	70
40 < edad < 45	45	edad > 70	+70

Fuente: Elaboración propia.

¹⁰ 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).

4. Metodología para analizar la información sobre el empleo.

4.1. Segmentación y propensión al emparejamiento.

La falta de homogeneidad en el mercado de trabajo provoca la segmentación del mismo; es decir, características tales como 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 a 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; prefieren no moverse de su segmento laboral.

A partir de los datos de colocaciones de que disponemos, vamos a definir segmentos laborales atendiendo a las características del puesto y del trabajador que forman cada colocación. El objetivo del proyecto es encontrar la relación o proximidad que existe entre los distintos segmentos definidos, pudiendo formar agrupamientos o *clusters* de los mismos, que podrían ser considerados 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 de residencia, grupo de ocupación, sector de actividad, sexo y edad¹¹ relativas al trabajador contratado, estas dos últimas variables se han incorporado al análisis de la segmentación, lo cual no se había realizado en trabajos previos. 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.

¹¹ En nuestros cálculos, la edad se ha discretizado en intervalos de 5 años, para evitar que tengamos un segmento diferente por cada año de edad.

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}):

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

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:

$$p_{ijt} = \frac{M_{ijt}}{M_t} ; p_{it} = \frac{M_{it}}{M_t} = \sum_{j=1}^m p_{ijt} ; p_{jt} = \frac{M_{jt}}{M_t} = \sum_{i=1}^n p_{ijt} \quad (3)$$

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 produjese una colocación en un segmento conjunto ij vendría dada por:

$$\widehat{P}_{ijt} = p_{it} p_{jt} \quad (4)$$

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 \widehat{p}_{ijt} 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:

$$pe_{ijt} = p_{ijt} / \widehat{p}_{ijt} \quad (5)$$

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 o *clusters*.

Tras el análisis de la segmentación del mercado, 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:

$$sw_{i_1-i_2} = \sum_{j=1}^m \min(pe_{i_1j} p_j, pe_{i_2j} p_j) = \sum_{j=1}^m p_j \min(pe_{i_1j}, pe_{i_2j}) \quad (6)$$

donde, por ejemplo, $pe_{i_1-jm} p_{jm} = p_{jm} (p_{i_1-jm} / p_{i_1} p_{jm}) = M_{i_1-jm} / M_{i_1}$ 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 proyecto nos vamos a centrar en la similitud de los segmentos de trabajadores¹².

El proceso de agrupamiento sigue un método jerárquico, en el que se van agrupando, sucesivamente, los segmentos de trabajador más similares 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

¹² Para un análisis por el lado de la oferta de puestos, véase el trabajo de Álvarez de Toledo *et al.* (2013).

segmento de trabajador como un grupo o *cluster* separado, seguidamente se agrupan los dos segmentos de trabajador entre los que existe mayor similitud, y así sucesivamente, pudiendo continuarse el proceso hasta la fusión total de todos los segmentos en un único *cluster*. 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. Programación mediante el software STATA de nuestra metodología.

Para aplicar la metodología de *clusters* a la base de datos de la que se dispone, se ha creado un fichero de programación que se divide en tres bloques, que se explican a continuación.

5.1. Segmentos laborales y propensión al emparejamiento.

En este proyecto 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, y 2.- Técnicos y profesionales científicos e intelectuales. Asimismo, también se eliminan de la base de datos aquellas colocaciones en las que alguna de sus dimensiones (las características del trabajador y del puesto) es desconocida. Es cierto que estamos eliminando información acerca de las colocaciones totales que se han producido en el intervalo de tiempo que estamos considerando, pero también es cierto que dichos segmentos conjuntos no conocidos totalmente 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 distinta de 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 la base de datos asciende a 358.298.

Seguidamente se van a crear los tres tipos de segmentos con los que vamos a trabajar, concatenando las dimensiones de cada uno. Para los segmentos de trabajador habrá que unir 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 conjunto calculamos su frecuencia: el número total de colocaciones que se han producido en él. Una vez que se tiene este dato, se pueden eliminar segmentos conjuntos de manera que en nuestra base de datos quede una única fila por segmento conjunto diferente. Con esta operación, la base de datos ha cambiado: anteriormente cada fila correspondía a una colocación, en cambio ahora hay tantas filas como segmentos conjuntos distintos. Esta transformación va a facilitar el cálculo de las propensiones al emparejamiento entre segmentos. 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 la metodología, hemos encontrado dos problemas: la baja frecuencia de ciertos segmentos de trabajador, lo cual los hace estadísticamente poco representativos, y el excesivo tamaño de la matriz de frecuencias cruzadas 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 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 3000 segmentos de trabajador. Además, no se han tenido en cuenta aquellos segmentos de puesto que sólo han recibido una colocación, ya que, de acuerdo con la ecuación (6), 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 cuenta con 37.178 segmentos conjuntos, formados por 3.000 segmentos de trabajador y 10.291 segmentos de puesto, y el total de colocaciones consideradas asciende a 101.660.

Antes de alcanzar el objetivo del cálculo de las superposiciones, tenemos que realizar unos cálculos intermedios para la obtención de las propensiones al emparejamiento, pe , que se calculan usando la expresión (5). Hay que obtener previamente el número total de colocaciones, tanto para cada segmento conjunto, M_{ijt} , como para cada segmento de trabajador, M_{it} , y de puesto, M_{jt} , que en el fichero de programación se corresponden con las variables $frec_conj$, $frec_dem$ y $frec_puesto$. A continuación, se pueden calcular las probabilidades reales de emparejamiento en un segmento conjunto de acuerdo a las expresiones (3), que permiten calcular la probabilidad aleatoria de emparejamiento, y con ambas se obtienen las propensiones al emparejamiento. Con todo este proceso, la base de datos ya está preparada para el cálculo de las superposiciones (o similitudes) a partir de dichas propensiones.

5.2. Cálculo de las superposiciones entre segmentos de trabajador.

Para trabajar con las variables de segmentos en formato numérico, y no en el formato “string” o de texto con que se generan, se va a generar un identificador

numérico para cada segmento de trabajador (id_dem) y para cada segmento de puesto (id_puesto); cuando sea necesario se volverá a la nomenclatura “string” inicial.

Para el cálculo de las superposiciones entre segmentos de trabajador, partimos de la expresión (6):

$$sw_{i1-i2} = \sum_{j=1}^m \min(pe_{i1j} p_j, pe_{i2j} p_j) = \sum_{j=1}^m \min(\text{sumando_sw}_{i1j}, \text{sumando_sw}_{i2j})$$

Primero, necesitamos calcular, para cada segmento conjunto, el producto entre la pe del segmento de trabajador y de puesto que forman dicho segmento conjunto y la probabilidad de que se produzcan colocaciones en el segmento de puesto (p_j); ésta es la variable que vamos a llamar $sumando_sw$. A continuación, ordenamos la base de datos, de forma que los segmentos de trabajador (id_dem) que van a un determinado segmento de puesto (id_puesto) queden ordenados de menor a mayor $sumando_sw$. De esta forma, el valor de $sumando_sw$ del segmento de trabajador que se encuentra en primer lugar será tomado como sumando para el cálculo de las superposiciones de dicho segmento con el resto de segmentos que están debajo de él. Para ilustrar este nuevo orden, se representa un extracto de la base de datos donde varios segmentos de trabajador envían colocaciones a un mismo segmento de puesto. Como se observa, están ordenados por la variable $sumando_sw$ de menor a mayor; de esta forma, el sumando correspondiente en la superposición entre el segmento de trabajador 1412 con cualquiera de los que se encuentran por debajo de él, enviando ambos segmentos colocaciones al segmento de puesto 180, va a tener el valor de 0,04.

Tabla 2: Ejemplo de la ordenación de la base de datos.

id_dem	id_puesto	sumando_sw
1412	180	0.04
2564	180	0.0434783
2105	180	0.0454545
129	180	0.0555556
905	180	0.0606061
888	180	0.0625

Fuente: Elaboración propia.

A continuación, se amplía la base de datos de tal forma que, dentro de cada segmento de puesto, cada segmento de trabajador va a tener tantas filas como segmentos de trabajador tenga por debajo, los cuales tienen un *sumando_sw* mayor. Además, se crea la variable *id_dem_par*, que va a capturar los *id_dem* de aquellos segmentos de trabajador cuyo sumando va a ser mayor que el del segmento anotado en *id_dem*. Para cada pareja de segmentos de trabajador (*id_dem*, *id_dem_par*), el *sumando_sw* de *id_dem* (y no el de *id_dem_par*) es el que debe intervenir en la suma de mínimos que da lugar al indicador de la superposición entre dichos segmentos de trabajador; en este indicador, cada sumando se corresponde a un segmento de puesto donde ambos segmentos de trabajador han enviado colocaciones¹³. Apoyándonos en la tabla 3, se puede observar que el primer segmento del trabajador debe aparecer cinco veces, ya que su sumando es menor que los sumandos de los otros cinco segmentos que envían colocaciones al mismo segmento de puesto, y así sucesivamente.

Tabla 3: Ejemplo de la ampliación de la base de datos.

<i>id_dem</i>	<i>id_dem_par</i>	<i>id_puesto</i>	<i>sumando_sw</i>
1412	2564	180	0.04
1412	2105	180	0.04
1412	905	180	0.04
1412	888	180	0.04
1412	129	180	0.04
2564	2105	180	0.0434783
2564	905	180	0.0434783
2564	888	180	0.0434783
2564	129	180	0.0434783
2105	905	180	0.0454545
2105	888	180	0.0454545
2105	129	180	0.0454545
129	905	180	0.0555556
129	888	180	0.0555556
905	888	180	0.0606061
888		180	0.0625

Fuente: Elaboración propia.

¹³ Cuando uno de los segmentos de trabajador envíe colocaciones a un segmento de puesto determinado, pero el otro no, el sumando correspondiente en el cálculo de la superposición entre ambos será simplemente cero.

Seguidamente, hay que acumular la variable *sumando_sw* para cada pareja de segmentos de trabajador, con el fin de obtener su superposición. Nuestros cálculos determinan que en algunos segmentos de puesto un segmento de trabajador determinado, el 1412 por ejemplo, quede ubicado en la variable *id_dem*, mientras que el otro segmento de trabajador, el 2564 por ejemplo, queda situado por tanto en *id_dem_par*; sin embargo, esto puede suceder al contrario si analizamos otro segmento de puesto cualquiera al que ambos segmentos de trabajador también envían colocaciones. Para acumular correctamente la variable *sumando_sw* para cada pareja de segmentos de trabajador con independencia de su ubicación en *id_dem* e *id_dem_par* dentro de cada segmento de puesto, hemos creado las variables auxiliares *id_dem1* e *id_dem2*, que almacenan, respectivamente, los valores mínimo y máximo de cada pareja (*id_dem*, *id_dem_par*) a lo largo de la base de datos. Una vez creadas dichas variables, se ordena la base de datos según ellas y se procede a acumular la variable *sumando_sw* para cada pareja (*id_dem1*, *id_dem2*). De esta forma, obtenemos por fin la variable de superposición o similitud entre cada pareja de segmentos de trabajador, para nuestros 3.000 segmentos de trabajador.

Se ha comentado anteriormente que la metodología de *clusters* usada en este trabajo se basa en una medida de similitud entre segmentos, pero el programa informático que hemos utilizado (STATA¹⁴) 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 superposición.

Todas las disimilitudes calculadas se pueden introducir en una matriz simétrica y cuadrada de disimilitudes, en la que tanto las filas como las columnas corresponden a los segmentos de trabajador, por lo que su dimensión va a ser de 3.000x3.000. Ya que cada celda corresponde a la disimilitud de un segmento del trabajador fila con otro segmento del trabajador columna, la diagonal principal de la matriz será nula: la disimilitud de un segmento consigo mismo es cero. Todas las operaciones anteriores

¹⁴ Para conocer la programación de sentencias en este software informático, se propone el manual de Hamilton (2009).

han ido dirigidas a la obtención de esta matriz, que es la que va a permitir al programa informático realizar los agrupamientos de los 3.000 segmentos de trabajadores.

Es importante resaltar en este punto que en este proyecto se han superado dos dificultades que se habían presentado en los trabajos anteriores que nos han servido de referencia principal¹⁵. 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 proyecto, 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.

5.3. Obtención de *clusters* de empleo.

Una vez que se ha obtenido la información de disimilitudes necesaria para realizar el proceso de agrupamiento, éste va a comenzar uniendo en un mismo grupo o *cluster* los dos segmentos de trabajador que son menos distintos entre sí, es decir, aquellos cuya disimilitud sea menor, y progresivamente seguirá formando *clusters* cada vez mayores, hasta poder llegar a un único *cluster*. Analíticamente, la disimilitud entre dos *clusters* se calcula con la siguiente expresión:

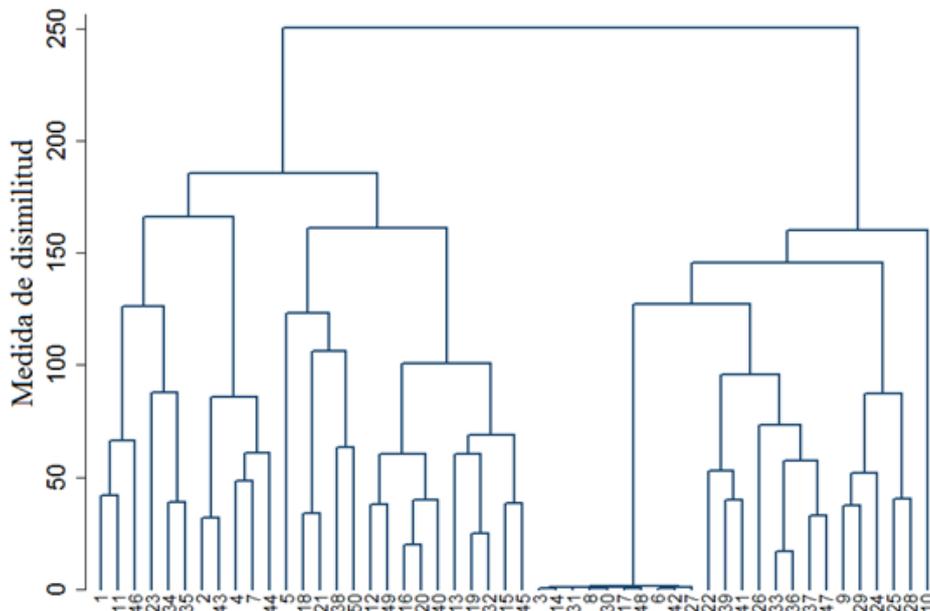
$$D(A, B) = \frac{1}{N_A \cdot N_B} \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} d(s_i, s_j) \quad s_i \in A, s_j \in B \quad (7)$$

¹⁵ Véanse a este respecto los trabajos de Álvarez de Toledo *et al.* (2013) y Álvarez de Toledo *et al.* (2014).

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 parar el proceso de agrupamiento en un número determinado de grupos, así como detenerlo en el momento en el que la disimilitud de la última agrupación que se produce sea la especificada. Además, los resultados pueden representarse gráficamente mediante un dendrograma, que es un gráfico en forma de árbol que muestra la secuencia de agrupamiento y la disimilitud que tienen los grupos formados. Sobre el dendrograma, se puede observar como los grupos son cada vez más escasos y más distantes entre sí. En el caso de un número elevado de segmentos iniciales, como es nuestro caso, la lectura de la figura se hace complicada, siendo necesario analizar los resultados numéricos que proporciona el software informático. Éste crea tres variables asociadas al proceso de agrupamiento, que denomina: *cluster_id*, *cluster_ord* y *cluster_hgt*. A modo de ejemplo, se muestra un gráfico del propio software informático, ya que en este proyecto nos encontramos en el caso de un número elevado de *clusters* y el software informático no permite realizar esta figura.

Figura 13: Ejemplo de dendrograma.



Fuente: Ayuda del programa STATA.

Analizando la primera de las variables, asigna un identificador a los segmentos de partida con el orden en que aparecen antes de la agrupación, y si ordenamos la base de datos a partir de esta variable, se observa que *cluster_ord* indica el lugar que ocupan los distintos segmentos en la base del dendrograma, y que *cluster_hgt* equivale a la altura o disimilitud con que se han unido dos segmentos o los grupos a los que estos ya pertenecían. A continuación, se muestra un extracto de la tabla del ejemplo anterior ordenada por la variable *cluster_id*; puede observarse cómo se sitúan las observaciones en la base del dendrograma según la variable *cluster_ord* –véase figura 13–.

Tabla 4: Ejemplo de las variables generadas en el agrupamiento.

<i>cluster_id</i>	<i>cluster_ord</i>	<i>cluster_hgt</i>
1	1	42.0
2	11	66.5
3	46	126.5
4	23	88.0
5	34	38.7
6	35	165.9
7	2	32.1
8	43	85.5
9	4	48.3
10	7	60.6
11	44	185.7
12	5	123.4
13	18	33.6
14	21	106.1
15	38	63.1

Fuente: Elaboración propia a partir de datos del programa STATA.

El programa también genera unas variables llamadas *grupo#*, siendo # un número cualquiera, que permiten controlar paso por paso el proceso de agrupamiento. Obtenemos tantas variables de grupo como segmentos de partida tengamos; en nuestro caso, obtenemos 3.000 variables que van desde *grupo1* a *grupo3000*. Cada variable *grupo#* muestra los grupos de segmentos que se han formado hasta la etapa # de agrupamiento, pudiendo pasar que, en una determinada etapa de agrupamiento, se forme más de un grupo nuevo. Ordenando la base de datos usando la variable *grupo3000*, ahora es *cluster_id* la que recorre la base del dendrograma de izquierda a derecha. Comparando las columnas *grupo3000* y *grupo2999*, por ejemplo, se detecta una

agrupación cuando en la segunda columna (*grupo2999*) aparece un número repetido; en estas dos filas donde se da la repetición, *cluster_id* indica qué dos segmentos (o los *clusters* a los que pertenecen) se han unido en un nuevo grupo, y *grupo3000* indica el lugar en la base del dendrograma que ocupan dichos segmentos. En el ejemplo siguiente puede verse más claramente.

Tabla 5: Identificación de la primera unión en el proceso de agrupamiento.

<i>segmento_dem</i>	<i>cluster_id</i>	<i>cluster_hgt</i>	<i>grupo2999</i>	<i>grupo3000</i>
1 30 41091 2512 690	375	0.41176474	1076	1076
1 40 41039 2912 1	710	0.69613338	1077	1077
1 45 41091 2912 530	798	0.68965518	1078	1078
1 55 41091 2512 530	856	0.12500006	1078	1079
2 30 41092 2433 84	2081	0.57777774	1079	1080
2 35 41069 2911 530	2452	0.7257926	1080	1081
1 45 41021 2399 530	784	0.95238096	1081	1082

Fuente: Elaboración propia

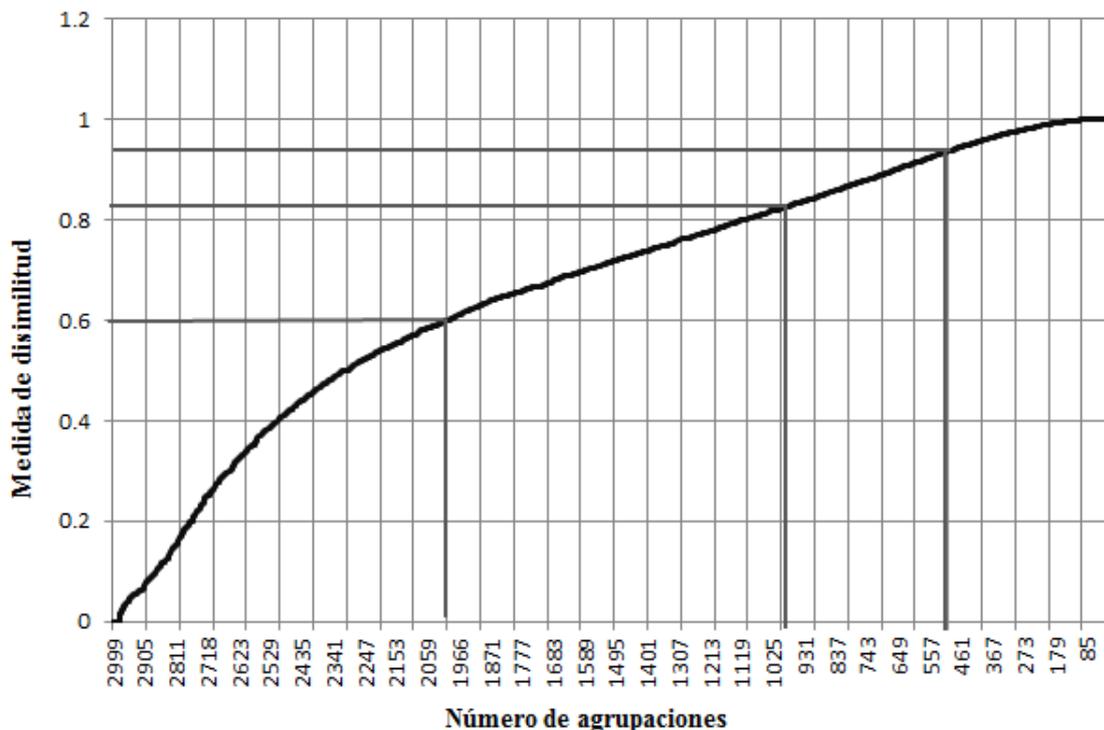
A modo de ejemplo, y haciendo uso de la tabla 5, se va a analizar la primera unión que se produce al realizar el proceso de agrupamiento. Puede observarse que la variable *grupo2999* se repite en el número 1078, por lo que los segmentos de trabajador que se unen son aquellos cuyos identificadores son 798 y 856, que son los correspondientes a la variable *cluster_id*, y la variable *grupo3000* indica la posición que estos segmentos ocupan en la base del dendrograma, empezando por la izquierda del mismo. Por su parte, *segmento_dem* está formado por la codificación de las cinco variables de las que está compuesto, que por orden corresponden a: sexo, edad, municipio de residencia, grupo de ocupación y sector de actividad.

6. Principales resultados.

A partir de los cálculos realizados anteriormente, se ha dibujado un gráfico en el que se observa cómo evoluciona la disimilitud cuando el número de grupos formados es cada vez menor. 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. Se ha decidido realizar tres cortes en el desarrollo de la agrupación, resultando dichas rupturas

en 2.000, 1.000 y 500 grupos, pudiendo observarse la disimilitud de la última unión que se produjo en los correspondientes cortes.

Figura 14: Evolución de la disimilitud.



Fuente: Elaboración propia.

Además, se han elaborado unas tablas que recogen los *clusters*, con mayor número de colocaciones cuando se ha llegado al número de grupos citado anteriormente (2000, 1000 y 500 grupos). En un análisis preliminar de los resultados obtenidos, se puede observar que los principales criterios de unión han sido el sexo 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 principalmente, y en cuanto a los grupos de ocupación de estos segmentos, predominan dos: las colocaciones de enfermería y relacionadas con la educación. Además, en cada tabla, se han analizado cinco *clusters* adicionales que presentan unas características algo diferentes a lo generalmente observado en el resto de *clusters*, es decir, no se corresponden con ocupaciones relacionadas con la sanidad y la educación; los llamamos *clusters* “peculiares” (o no habituales). A continuación se realiza un estudio más

detallado de cada tabla. Además, se han generado unas figuras, una para cada parada realizada en el proceso de agrupamiento, que recogen la información relativa a la ubicación geográfica de los *clusters* en el mapa de Andalucía. Cada *cluster* se ha identificado con su correspondiente variable grupo#, que para cada corte habrá sido *grupo2000*, *grupo1000* y *grupo500*, la cual representa la ubicación en la base del dendrograma de dichas agrupaciones. Esta variable indica, de cierta forma, la similitud que presentan dos *clusters* distintos, ya que cuanto más próximos están en la base del dendrograma antes se producirá la unión de los mismos, ya sea la unión de los propios *cluster*, o de las agrupaciones de las que forman parte.

6.1. 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 –véase la figura 14–. El grupo 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. 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*.

Al margen 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 a 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 éstas 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 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 ocupaciones de mayor 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 6 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”.

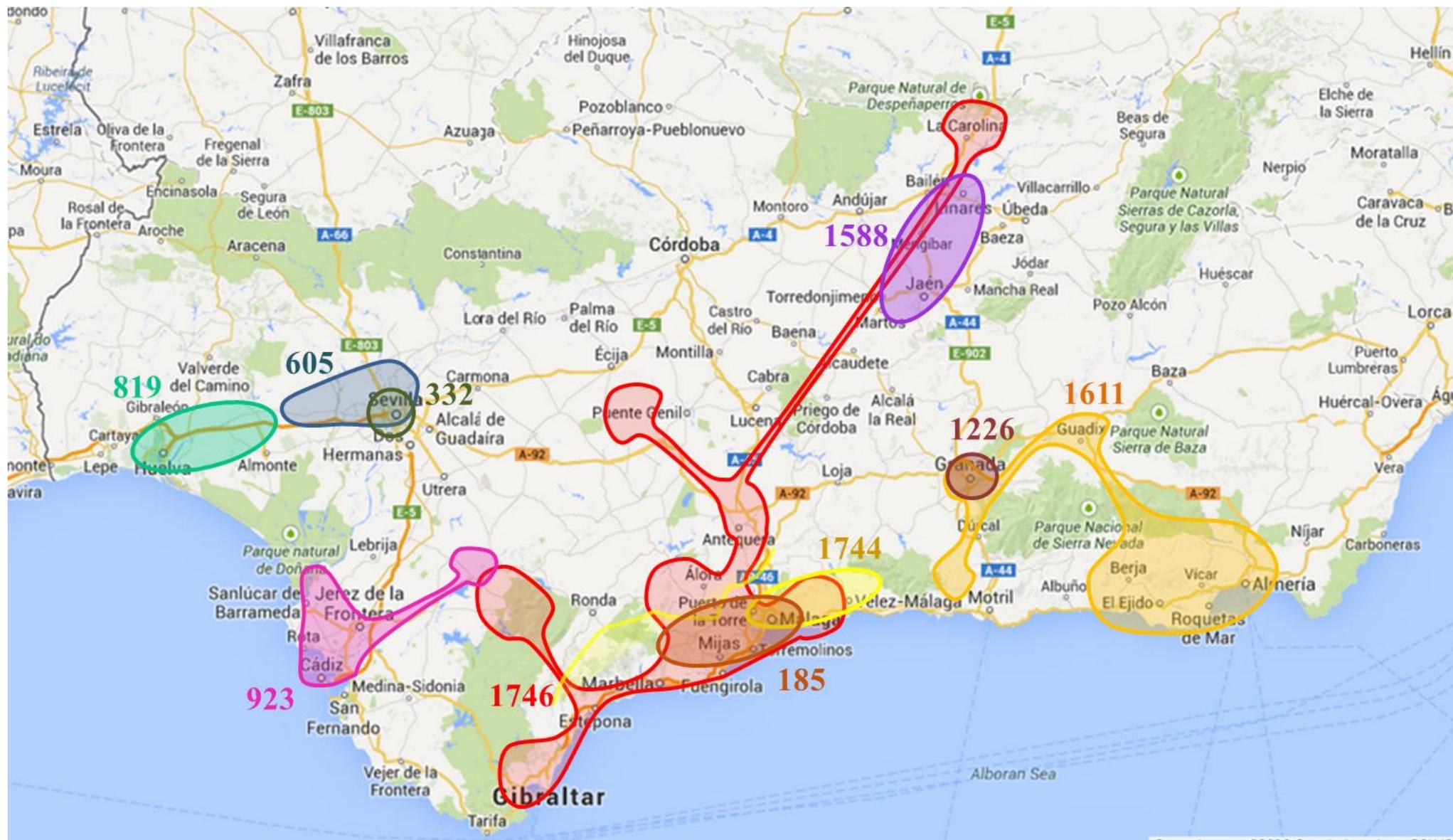
En el *cluster* 1.588, los grupos de ocupación que se observan son “psicólogos”, “biólogos, botánicos, zoólogos y asimilados” y “diplomados en educación social”, en similar proporción, en cuanto a los municipios de residencia de los buscadores de empleo, son Jaén y Linares, entre otros pueblos de la provincia, las 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*, con la ocupación “auxiliar de enfermería” y ubicándose los centros de trabajo en la ciudad de Jaén. Esta agrupación está formada por 144 colocaciones, que han tenido lugar a partir de 6 segmentos de trabajador.

El siguiente *cluster* (1.226) está formado por 4 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* se presentan en una proporción muy parecida, y son “otras actividades empresariales” y “comercio minorista”. 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, predominan 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 5 segmentos de trabajador. De los trabajadores de esta agrupación, el 59,8% son hombres, especialmente de 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 15: Mapa de los *clusters* con mayor número de colocaciones en la división de 2000 grupos.



Fuente: Elaboraci n propia utilizando Google Maps.

6.2. Obtención de 1.000 *clusters* de segmentos de trabajador.

Se plantea ahora detener el proceso de agrupamiento cuando la base de datos cuenta con 1.000 grupos, habiéndose realizado la última agrupación con una disimilitud en torno a 0,8 –véase la figura 14–. 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 misma 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 Ilora. 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 misma ciudad de residencia, Granada, dentro del sector de actividad de la educación y en el grupo de ocupación de profesionales de la enseñanza.

El cuarto y quinto *cluster*, 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 el pueblo de 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 es 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 8 segmentos de trabajador, que han dado lugar a 256 colocaciones, se ubica principalmente en Almería capital, además de en otras 3 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 8 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 4 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%).

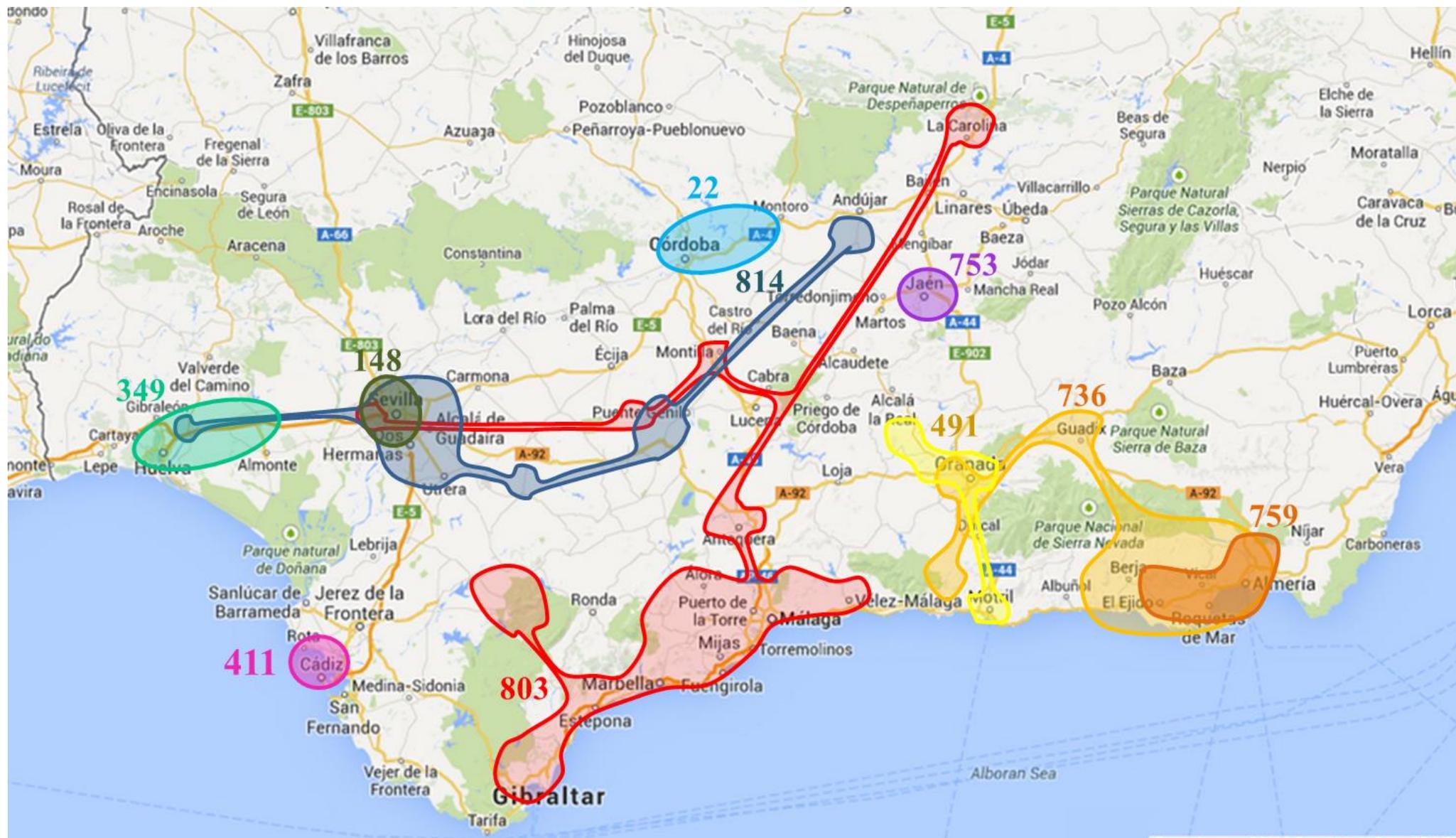
Por último, el *cluster* 713 se localiza en la ciudad de Jaén. La totalidad de los segmentos de trabajador (5 segmentos) están formados por mujeres con la ocupación de “psicólogos”, destacando las edades comprendidas entre 25 y 30 años (52,9%). El sector de actividad que predomina es el de “actividades 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”.

Tabla 7: División de los titulados en 1000 grupos

Posición del grupo	Colocaciones	Segmentos de trabajo	provincia	% provincia	sexo	% sexo	edad	% edad	provincia de pertenencia	municipio	% municipio	ocupación	% ocupación	actividad	% actividad	
803	4297	70	Málaga	90.8	Mujer	84.6	25-30	53.7	Málaga	Málaga	43.1	Enfermeros	98.1	Actividades sanitarias y servicios sociales	86.5	
			Cádiz	6.6	Hombre	15.4	20-25	34.4	Málaga	Fuengirola	9.5	Compositores, músicos y cantantes	0.8	Administración pública	4.9	
			Sevilla	1.3			30-35	10.3	Málaga	Marbella	8.7	Veterinarios	0.7	Otras actividades empresariales	4.8	
			Jaen	0.7			35-40	1.6	Málaga	Mijas	8.4	Profesores de educación especial	0.5	Comercio minorista	2.8	
			Córdoba	0.6					Málaga	Estepona	6.8			Hostelería	0.9	
									Cádiz	Algeciras	2.6					
									Málaga	Torremolinos	2.3					
										Otros(17)	18.7					
148	2581	37	Sevilla	100.0	Mujer	74.3	25-30	35.7	Sevilla	Sevilla	98.8	Profesores de secundaria	67.3	Educación	71.6	
					Hombre	25.7	30-35	33.0	Sevilla	Mairena del Aljarafe	0.6	Filólogos, intérpretes y traductores	17.4	Otras actividades empresariales	20.9	
							35-40	16.2	Sevilla	Dos Hermanas	0.5	Biólogos, botánicos, zoólogos y asimilados	6.0	Comercio minorista	3.5	
							40-45	8.2				Escritores, periodistas y asimilados	3.4	Hostelería	2.1	
							20-25	5.2				Otros profesionales de la enseñanza	2.8	Actividades de ocio (lúdicas)	1.9	
							45-50	1.7				Químicos	1.5			
												Profesores de universidades y otros centros de enseñanza superior	0.9			
												Profesionales en recursos humanos	0.8			
491	1788	45	Granada	100.0	Mujer	74.2	25-30	26.6	Granada	Granada	91.2	Profesores de secundaria	41.2	Educación	85.8	
					Hombre	25.8	30-35	24.1	Granada	La Zubia	2.4	Psicólogos	12.4	Construcción	6.6	
							35-40	20.4	Granada	Illora	1.5	Filólogos, intérpretes y traductores	8.5	Otras actividades empresariales	4.5	
							40-45	11.0	Granada	Ogijares	1.2	Arquitectos técnicos	6.6	Administración pública	1.2	
							20-25	8.1	Granada	Chauchina	1.1	Otros profesionales de la enseñanza	5.5	Fabricación de material electrónico, equipo, aparatos e instrumentos	1.1	
							50-55	6.7	Granada	Huétor Vega	1.1	Biólogos, botánicos, zoólogos y asimilados	4.6	Comercio minorista	0.8	
							45-50	3.3	Granada	Canes de la Vega	0.9	Especialistas en métodos didácticos y pedagógicos	3.9			
									Granada	Motril	0.7	Otros(12)	17.3			
814	1711	28	Sevilla	97.3	Mujer	77.3	20-25	48.9	Sevilla	Sevilla	70.3	Enfermeros	100.0	Actividades sanitarias y servicios sociales	79.8	
			Jaen	1.4	Hombre	22.7	25-30	41.6	Sevilla	Dos Hermanas	13.2			Otras actividades empresariales	8.9	
			Huelva	1.3			30-35	9.6	Sevilla	Los Palacios y Villafraña	3.3			Administración pública	7.9	
									Sevilla	San Juan de Aznalfarache	2.1			Agricultura	1.4	
									Sevilla	Alcalá de Guadaíra	1.6			Comercio mayorista	1.0	
										Otros(8)	1.4			Hostelería	0.9	
736	1685	33	Almería	91.4	Mujer	86.2	20-25	51.4	Almería	Almería	49.4	Enfermeros	95.4	Actividades sanitarias y servicios sociales	83.1	
			Granada	8.6	Hombre	13.8	25-30	40.5	Almería	El Ejido	12.8	Profesores de primaria	2.5	Administración pública	6.5	
							30-35	6.0	Almería	Roquetas de Mar	5.8	Diplomados en educación social	2.1	Comercio minorista	3.8	
							40-45	2.1	Almería	Vícar	5.6			Hostelería	2.5	
									Almería	Berja	4.8			Otras actividades empresariales	1.5	
									Almería	Huécija	3.7			Comercio mayorista	1.4	
										Otros(8)	17.9			Educación	1.1	
22	624	25	Córdoba	100.0	Mujer	72.9	30-35	33.8	Córdoba	Córdoba	96.5	Psicólogos	16.0	Educación	58.2	
					Hombre	27.1	25-30	33.2	Córdoba	Villafraña de Córdoba	3.5	Agentes de igualdad de oportunidades para la mujer y otros diplomados en trabajo social	15.2	Actividades asociativas	12.8	
							35-40	22.4				Profesionales en recursos humanos	13.1	Otras actividades empresariales	8.8	
							40-45	4.2				Técnicos de empresas y actividades turísticas	13.0	Actividades sanitarias y servicios sociales	7.7	
							50-55	3.8				Otros profesionales de la enseñanza	12.2	Administración pública	6.9	
							45-50	2.6				Veterinarios	8.0	Construcción	3.0	
												Ingenieros técnicos mecánicos	6.3	Comercio mayorista	2.6	
												Graduados sociales y asimilados	6.3			
												Otros profesionales de nivel medio de informática	4.0			
												Ayudantes de biblioteca y asimilados	3.4			
									Profesores de secundaria	2.6						
759	256	8	Almería	100.0	Mujer	76.6	25-30	52.7	Almería	Almería	72.7	Químicos	27.3	Otras actividades empresariales	55.1	
					Hombre	23.4	20-25	27.3	Almería	Roquetas de Mar	12.1	Ingenieros técnicos en especialidades agrícolas y forestales	23.4	Comercio mayorista	19.9	
									Almería	El Ejido	9.0	Profesores de primaria	19.9	Hostelería	18.4	
									Almería	Huércal de Almería	6.3	Dirección de departamentos de comercialización y ventas	9.0	Agencias de viaje y actividades anexas al transporte	6.6	
349	232	8	Huelva	100.0	Mujer	80.6	25-30	55.2	Huelva	Huelva	69.4	Diplomados en educación social	40.1	Educación	44.0	
					Hombre	19.4	35-40	14.2	Huelva	Bollullos Par del Condado	14.2	Arquitectos técnicos	21.6	Agricultura	28.9	
							50-55	12.1	Huelva	Moguer	8.6	Otros ingenieros superiores (excepto agropecuarios)	17.7	Administración pública	18.5	
							30-35	11.2	Huelva	Lucena del Puerto	7.8	Otros profesionales de la enseñanza	12.1	Actividades asociativas	8.6	
							40-45	7.3				Ingenieros técnicos en especialidades agrícolas y forestales	8.6			
411	182	4	Cádiz	100.0	Mujer	78.0	25-30	83.5	Cádiz	Cádiz	100.0	Filólogos, intérpretes y traductores	78.0	Otras actividades empresariales	50.5	
					Hombre	22.0	30-35	16.5				Filósofos, historiadores y profesionales en ciencias políticas	22.0	Educación	27.5	
713	155	5	Jaén	100.0	Mujer	100.0	25-30	52.9	Jaén	Jaén	100.0	Psicólogos	100.0	Actividades sanitarias y servicios sociales	72.9	
					Hombre	0.0	30-35	36.1						Otras actividades empresariales	16.1	
							35-40	11.0						Actividades asociativas	11.0	

Fuente: Elaboración propia.

Figura 16: Mapa de los clusters con mayor número de colocaciones en división de 1000 grupos.



Fuente: Elaboración propia utilizando Google Maps.

6.3. 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 14–. 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 más 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 principales 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 ciudad 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 frecuentes 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 que 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 misma 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 8 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 actividad que solicitaban, que era “otras actividades empresariales”; aunque los empleos se han registrados en diversas actividades.

El siguiente *cluster* aparece en la tabla 8 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 1000 grupos y que se ha identificado con el número 759 en la tabla 7. Resulta interesante comprobar como 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 6 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.

Tabla 8: División de los titulados en 500 grupos

Posición del grupo	Colocaciones	Segmentos de trabajo	provincia	% provincia	sexo	% sexo	edad	% edad	provincia de pertenencia	municipio	% municipio	ocupación	% ocupación	actividad	% actividad
53	9160	183	Sevilla	99.5	Mujer	81.3	25-30	39.5	Sevilla	Sevilla	94.3	Profesores de secundaria	21.6	Educación	56.5
			Granada	0.2	Hombre	18.7	30-35	29.8	Sevilla	Dos Hermanas	2.6	Psicólogos	14.3	Otras actividades empresariales	27.4
			Córdoba	0.2			20-25	14.0		Otros(12)	3.1	Profesores de primaria	12.9	Actividades sanitarias y servicios sociales	3.4
			Almería	0.1			35-40	10.0				Especialistas en métodos didácticos y pedagógicos	12.1	Hostelería	3.2
							40-45	4.8				Profesores de infantil	10.1	Actividades de ocio (lúdicas)	2.9
251	4885	134	Málaga	100.0	Mujer	80.9	25-30	33.9	Málaga	Málaga	93.0	Profesores de secundaria	24.8	Educación	60.1
					Hombre	19.1	30-35	23.4	Málaga	Vélez-Málaga	1.4	Profesores de primaria	21.8	Otras actividades empresariales	18.8
							35-40	13.6	Málaga	Rincón de la Victoria	1.2	Otros profesionales de la enseñanza	9.5	Comercio minorista	8.1
							20-25	13.6	Málaga	Alhaurín de la Torre	1.2	Filólogos, intérpretes y traductores	8.9	Hostelería	4.5
							40-45	9.4	Málaga	Cártama	0.9	Biólogos, botánicos, zoólogos y asimilados	6.1	Actividades asociativas	4.2
340	4414	72	Málaga	88.4	Mujer	85.0	25-30	52.3	Málaga	Málaga	42.0	Enfermeros	98.1	Actividades sanitarias y servicios sociales	86.9
			Cádiz	6.4	Hombre	15.0	20-25	36.2	Málaga	Fuengirola	9.2	Compositores, músicos y cantantes	0.8	Administración pública	4.8
			Jaen	3.3			30-35	10.0	Málaga	Marbella	8.4	Veterinarios	0.6	Otras actividades empresariales	4.7
			Sevilla	1.3			35-40	1.5	Málaga	Mijas	8.2	Profesores de educación especial	0.5	Comercio minorista	2.7
			Córdoba	0.6					Málaga	Estepona	6.6		0.5	Hostelería	0.9
56	2679	82	Sevilla	99.3	Mujer	84.3	25-30	36.6	Sevilla	Sevilla	88.9	Abogados	22.7	Otras actividades empresariales	68.2
			Málaga	0.7	Hombre	15.7	30-35	33.1	Sevilla	Arahal	2.2	Especialistas en métodos didácticos y pedagógicos	11.3	Comercio minorista	6.8
							35-40	16.8	Sevilla	Utrera	1.6	Graduados sociales y asimilados	9.6	Administración pública	5.5
							20-25	7.5	Sevilla	Mairena del Alcor	1.3	Filólogos, intérpretes y traductores	5.8	Actividades de ocio (lúdicas)	3.3
							40-45	4.0	Sevilla	Alcalá de Guadaíra	1.1	Técnicos de empresas y actividades turísticas	5.3	Hostelería	3.2
336	2467	45	Cádiz	99.3	Mujer	91.6	25-30	49.5	Cádiz	Cádiz	21.3	Enfermeros	100.0	Actividades sanitarias y servicios sociales	91.0
			Córdoba	0.7	Hombre	8.4	20-25	36.9	Cádiz	Jerez de la Frontera	21.0	Otras actividades empresariales	4.5		
							30-35	11.2	Cádiz	San Fernando	15.1	Seguros y planes de pensiones, excepto seguridad social obligatoria	1.8		
							35-40	1.7	Cádiz	El Puerto de Santa María	12.0	Comercio minorista	1.4		
							45-50	0.7	Cádiz	Sanlúcar de Barrameda	9.9	Transporte terrestre	0.6		
124	476	15	Huelva	100.0	Mujer	100.0	25-30	37.4	Huelva	Huelva	100.0	Diplomados en educación social	95.4	Actividades sanitarias y servicios sociales	34.0
					Hombre	0.0	30-35	37.4				Profesores de educación especial	4.6	Otras actividades empresariales	21.4
							20-25	16.6					4.6	Actividades asociativas	19.5
							35-40	8.6					4.6	Educación	11.1
													4.6	Administración pública	7.8
16	273	8	Córdoba	100.0	Mujer	94.1	25-30	50.2	Córdoba	Córdoba	94.1	Técnicos de empresas y actividades turísticas	49.5	Otras actividades empresariales	100.0
					Hombre	5.9	20-25	33.3	Córdoba	Posadas	5.9	Biólogos, botánicos, zoólogos y asimilados	16.5		
							30-35	16.5				Otros profesionales de la enseñanza	12.8		
												Escritores, periodistas y asimilados	11.4		
												Abogados	9.9		
292	270	10	Jaén	100.0	Mujer	87.4	25-30	45.6	Jaén	Jaén	68.1	Biólogos, botánicos, zoólogos y asimilados	34.8	Otras actividades empresariales	34.8
					Hombre	12.6	30-35	34.1	Jaén	Alcalá la Real	10.0	Graduados sociales y asimilados	19.6	Actividades asociativas	22.2
							35-40	20.4	Jaén	Torreblascopedro	8.1	Enfermeros	10.0	Educación	21.9
									Jaén	Pegalajar	7.8	Profesores de universidades y otros centros de enseñanza superior	8.1	Construcción	7.4
									Jaén	Torreperogil	5.9	Químicos	7.8	Actividades de ocio (lúdicas)	7.0
317	256	8	Almería	100.0	Mujer	76.6	25-30	52.7	Almería	Almería	72.7	Químicos	27.3	Otras actividades empresariales	55.1
					Hombre	23.4	20-25	27.3	Almería	Roquetas de Mar	12.1	Ingenieros técnicos en especialidades agrícolas y forestales	23.4	Comercio mayorista	19.9
							30-35	19.9	Almería	El Ejido	9.0	Profesores de primaria	19.9	Hostelería	18.4
									Almería	Huércal de Almería	6.3	Dirección de departamentos de comercialización y ventas	9.0	Agencias de viaje y actividades anexas al transporte	6.6
												Profesores de secundaria	7.4		
217	228	6	Granada	100.0	Mujer	71.5	25-30	78.5	Granada	Granada	100.0	Arquitectos técnicos	28.5	Comercio minorista	39.9
					Hombre	28.5	20-25	21.5				Arquitectos técnicos	22.4	Otras actividades empresariales	31.6
												Otros diversos profesionales del derecho	20.6	Construcción	28.5
												Agentes de igualdad de oportunidades para la mujer y otros diplomados en trabajo social	12.3		
												Profesionales en contabilidad	9.2		
									Arquitectos, urbanistas e ingenieros planificadores de tráfico	7.0					

Fuente: Elaboración propia.

Figura 17: Mapa de los clusters con mayor número de colocaciones en división de 500 grupos.



Fuente: Elaboración propia utilizando Google Maps.

7. Conclusiones.

El presente proyecto 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 éste 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 grupo de trabajador y otro de puesto, y el “grado de similitud” entre dos grupos 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 dé dicha colocación en 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 proyecto propone un estudio del mercado de trabajo andaluz de los titulados superiores, que se basa en el empleo de datos individuales sobre colocaciones. 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 economía 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 los segmentos de trabajador que más se parecen, para formar grupos o *clusters* mayores (y relativamente homogéneos); es decir, para definir mercados de trabajos locales. Todo este proceso de fusiones puede ser seguido a través de un diagrama en forma de árbol invertido 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, que en nuestro proyecto 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 introduce algunas mejoras en la metodología de *clusters* respecto a los dos trabajos que han sido nuestra principal referencia: Álvarez de Toledo *et al.* (2013) y Álvarez de Toledo *et al.* (2014); las mejoras obtenidas 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 o ubicación predominante en cada *cluster*. Se ha observado que los *clusters* “principales” se sitúan, por lo general, dentro de una misma provincia, están formados mayoritariamente por mujeres, y que destacan el intervalo de edad de 25 a 30 años y los grupos de ocupación de enfermeros y los relacionados con la educación.

A diferencia de los *clusters* generados en el estudio de Álvarez de Toledo *et al.* (2013), los *clusters* obtenidos en este proyecto presentan formas más alargadas –son *clusters* donde pueden observarse ciertos municipios alejados de su núcleo principal–. Se podría concluir

entonces que los trabajadores titulados, concentrados 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 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 “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 a los trabajadores en búsqueda de empleo puede suponer una política activa del mercado de trabajo, ya que se propone el desarrollo de este tipo de metodología para obtener un mercado de trabajo en el que se reduzca el tiempo de búsqueda por parte de los trabajadores.

Una posible continuación para este proyecto, 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 y *clusters* que tienden a desaparecer.

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.

Álvarez de Toledo, P., Núñez, F. y Usabiaga, C. (2011): “An Empirical Analysis of the Matching Process in Andalusian Public Employment Agencies”, *Hacienda Pública Española*, 198.

Álvarez de Toledo, P., Núñez, F. y Usabiaga, C. (2013): “Análisis “*Clusters*” de los Flujos Laborales Andaluces”, *Revista de Estudios Regionales*, **97**, págs. 195 – 221.

Álvarez de Toledo, P., Núñez, F. y Usabiaga, C. (2014): “An Empirical Approach on Labour Segmentation. Applications with Individual Duration Data”, *Economic Modelling*, **36**, págs. 252 – 267.

Blanchard, O.J. (2008): *Macroeconomía*, Pearson, Prentice-Hall, 4ª Ed.

Blanchard, O.J. y Diamond, P. (1989): “The Beveridge Curve”, *Brookings Papers on Economic Activity*, **1**, págs. 1 - 76.

Davis, S.J. y Haltiwanger, J. (1992): “Gross Job Creation, Gross Job Destruction, and Employment Reallocation”, *The Quarterly Journal of Economics*, **107** (3), págs. 819-863.

Estrada, A., García, P. e Izquierdo, M. (2002): “Los Flujos de Trabajadores en España: El Impacto del Empleo Temporal”, *Banco de España*, Documento de Trabajo nº 0206.

Frasser, C.C. (2008): "Equilibrio General Walrasiano: Método y Limitaciones", *Mundo Económico y Empresarial*, **6**, págs. 71-80.

García, J.J. (2003): *El Mercado de Trabajo en España desde la Perspectiva de Flujos*, Proyecto Fin de Carrera, Universidad de Sevilla, Departamento de Organización Industrial y Gestión de Empresas.

García, P. e Izquierdo, M. (2007): “La Evolución Reciente de los Flujos del Mercado Español”, *Boletín Económico del Banco de España*, Marzo, págs. 47 – 56.

Hamilton, L.C. (2009): *Statistics with STATA*, Brooks/Cole, Cengage Learning.

Mankiw, N.G. (2007): *Principios de Ecomomía*, Thomsom, 4º Ed.

Petrongolo, B. y Pissarides, C.A. (2001): “Looking into the Black Box: A Survey of the Matching Function”, *Journal of Economic Review*, **97** (4), págs. 1074-1101.

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

Yashiv, E. (2007): *Labor Search and Matching in Macroeconomics*, *European Economic Review*, 51, págs. 1859 – 1895.

Referencias electrónicas

Web del Instituto Nacional de Estadística: www.ine.es

Web de indicadores estadísticos de Andalucía:

<http://www.juntadeandalucia.es/institutodeestadisticaycartografia/bd/indea/index.jsp?tema=35>