

Patrones de homofilia resilientes en redes de amistad juvenil: estudio de caso mediante un experimento de simulación computacional

*Resilient Homophily Patterns in Youth Friendship Networks:
A Case Study Using a Computer Simulation Experiment*

Francisco Linares Martínez, Francisco J. Miguel Quesada y Mona Kohl

Palabras clave

Homofilia
• Índice de homofilia de Coleman
• Redes de amistad
• Resiliencia
• Simulación basada en agentes
• Sociología computacional

Key words

Homophily
• Coleman's
Homophily Index
• Friendship Networks
• Resilience
• Agent-Based
Simulation
• Computational
Sociology

Resumen

El presente trabajo aborda la cuestión de cómo conocer si los patrones de homofilia hallados en una red social son resilientes ante pequeñas perturbaciones que pueden producirse en dicha red. Para ello se han replicado con un modelo basado en agentes los datos de una encuesta realizada a los estudiantes de un instituto de enseñanza secundaria de las islas Canarias. Dicho modelo calcula los índices de homofilia y su significatividad estadística para posteriormente proceder a la simulación de pequeñas alteraciones en la distribución de los vínculos. Los resultados muestran claramente que algunos índices de homofilia resisten dichas perturbaciones y otros no. La evidencia hallada apunta a que la distribución de los individuos entre las comunidades que configuran la red es un factor clave que explica que ciertos patrones de relaciones sean más resilientes que otros.

Abstract

This paper deals with how to recognise if the patterns of homophily found in a social network are resilient to small disturbances that may occur in that network. Data from a survey of students in a secondary school in the Canary Islands were replicated using an agent-based model. The model calculated homophily indices and their statistical significance and then simulated small alterations in the distribution of links. The results clearly show that some homophily indices resist these kinds of perturbations and others do not. Evidence suggests that the distribution of individuals across the social network communities is a key factor in explaining why certain patterns of relationships are more resilient than others.

Cómo citar

Linares Martínez, Francisco; Miguel Quesada, Francisco J. y Kohl, Mona (2022). «Patrones de homofilia resilientes en redes de amistad juvenil: estudio de caso mediante un experimento de simulación computacional». *Revista Española de Investigaciones Sociológicas*, 177: 43-68. (doi: 10.5477/cis/reis.177.43)

La versión en inglés de este artículo puede consultarse en <http://reis.cis.es>

Francisco Linares Martínez: Universidad de La Laguna | flinares@ull.es

Francisco J. Miguel Quesada: Universitat Autònoma de Barcelona | Miguel.Quesada@uab.cat

Mona Kohl: Atos Consulting (Canarias) (México) | mona.kohl@atos.net

INTRODUCCIÓN¹

El concepto de homofilia fue acuñado por Robert K. Merton en el artículo seminal escrito junto con Paul Lazarsfeld (Lazarsfeld y Merton, 1954), cuyo objeto son las relaciones de amistad². Merton justifica allí la necesidad de introducir un nuevo término en el vocabulario sociológico argumentando que no existe en inglés una palabra para referirse concisamente a la amistad entre personas «*of the same kind*». A día de hoy la definición más corriente es la presentada en un citadísimo artículo de revisión de la literatura: «El principio de que un contacto entre personas similares ocurre con mayor probabilidad que un contacto entre personas distintas» (McPherson, Smith-Lovin y Cook, 2001: 416). Como se verá más adelante, a pesar de su aparente claridad, existe cierta ambigüedad o confusión en el uso del término, ya que en unos trabajos académicos se vincula a una preferencia de los sujetos a mantener relaciones con otros similares mientras que en otros se usa para denotar una regularidad empírica, un patrón de comportamiento colectivo (la frecuencia de contactos entre similares) que puede ser el resultado de diversos mecanismos sociales. Este segundo sentido es el que se corresponde más claramente con el uso original que Lazarsfeld y Merton hacen del término; es también el sentido más obvio en la definición de McPherson, Smith-Lovin y Cook, y es igualmente el significado que se dará al término en el presente trabajo.

De manera simplificada podemos decir que en una red de relaciones entre individuos existe homofilia respecto a un atributo si la proporción de vínculos entre individuos que muestran ese atributo es superior a la propor-

ción de individuos con ese atributo en dicha población. Así, si un lazo entre dos individuos católicos es más probable que la proporción de individuos de esa religión, ello implica un comportamiento no azaroso: hay menos lazos entregrupos y más lazos intragrupo de los que cabría esperar en ausencia de mecanismos sociales que induzcan la estructuración de relaciones sociales. El presente trabajo tiene como objeto estudiar un problema muy conciso en la identificación de este fenómeno, a saber, en qué medida los índices de homofilia que pueden calcularse en una red de individuos son resistentes a pequeñas alteraciones en sus vínculos, dado que en el transcurso de la vida social algunos lazos desaparecen y otros nuevos se crean.

Para abordar esta cuestión se ha construido un modelo de simulación que replica los datos de una red de estudiantes de un instituto de enseñanza secundaria (codificado como IES San Borondón). Una vez identificados los índices de homofilia que son estadísticamente significativos, se ha procedido a una experimentación «virtual» en la que recursivamente se sustituyen vínculos reales por vínculos elegidos al azar, hasta un 15% del total de los lazos homofílicos. Como resultado de este procedimiento todos los índices disminuyeron pero algunos permanecieron, no obstante, siendo estadísticamente significativos mientras que otros dejaron de serlo. Esto suscita de forma inmediata la cuestión de por qué algunos patrones de homofilia son resilientes ante pequeñas perturbaciones en la estructura de la red mientras que otros no. La hipótesis general que se sostiene es que la explicación de que ciertos patrones de homofilia sean resilientes y otros no descansa en cómo los individuos están distribuidos en las distintas comunidades en las que se descompone el conjunto de la red. Hasta donde sabemos, no existe un precedente de un estudio sistemático de esta cuestión en la literatura especializada.

Desde el punto de vista metodológico, abordar el problema requiere, como se

¹ Este trabajo se ha beneficiado de los fondos de los proyectos CSO2015-6474-R (MINECO) y PID2019-107589GB-I00 (MICIN).

² El artículo está firmado por ambos, pero se divide en dos partes: la sustantiva, cuya autoría es de Merton, y la metodológica, escrita por Lazarsfeld.

verá, desplazarse a través de distintos niveles conceptuales o planos de la realidad (individuos, comunidades y red) cuya información es codificada en tres bases de datos distintas:

En el nivel más básico las unidades de análisis son los sujetos encuestados. De estas unidades se recoge, a través de un cuestionario, información de ciertos atributos típicos (como, por ejemplo, edad, género, religión, etc.), así como información de carácter relacional a través de un generador de nombres en el mismo cuestionario.

Con la información obtenida del cuestionario es reconstruida la red de relaciones de amistad. Esta red es del tipo «mundo pequeño» (*small world*), configurada por comunidades o grupos³ de individuos que, a su vez, tienen algunos vínculos que realizan el papel de «puente» (*bridge*) con otras comunidades. La segunda base de datos está constituida por las propiedades de las comunidades, que son unidades supraindividuales.

Finalmente, dado que en el análisis se computan todos los índices de homofilia posibles en la red, una tercera base de datos es la del conjunto de índices y los datos asociados a cada uno de ellos (tales como: número de individuos, comunidades a las que pertenecen, número de lazos, etc.), lo que constituye una descripción exhaustiva de los patrones de comportamiento homofílico observables en el conjunto de la red.

En adelante el trabajo se estructura de la siguiente manera: en el siguiente apartado se realiza una revisión teórica en la que se presta especial atención a los mecanismos explicativos de los patrones de homofilia. Posteriormente se ofrecen los detalles metodológicos, se describe brevemente el modelo de simulación ABM construido

específicamente para abordar la pregunta de investigación y se ilustra su funcionamiento. De forma seguida se procede a exponer los resultados del análisis de tres tipos de datos distintos: en primer lugar, los índices de homofilia, distinguiendo los que permanecen significativos de los que no; en segundo lugar, las agrupaciones de individuos (comunidades) que están asociadas con mayor frecuencia a índices de homofilia que permanecen significativos; y, en tercer lugar, las características de los índices de homofilia que permanecen resistentes. El artículo finaliza con una discusión de estos resultados y algunas conclusiones.

MECANISMOS DE LA HOMOFILIA ENDOGÁMICA

Como señalan McPherson, Smith-Lovin y Cook en la revisión de la literatura citada en la introducción, la investigación empírica del fenómeno entró en auge a partir de la década de los setenta del pasado siglo, en buena medida estimulada por la teoría de la estructura social de Peter Blau, así como por el desarrollo del análisis de redes. Empero no siempre se ha reconocido correctamente que en la obra de Blau (1977) el concepto tiene un doble estatus, lo que puede haber producido cierta confusión en algunos trabajos académicos.

Por una parte, Blau lo usa como una asunción básica relativa a las preferencias de los individuos. En este caso, si los individuos *i* y *j* comparten una cualidad tendrán interés en crear un lazo de amistad si se da la oportunidad. Es decir, el lazo de amistad se explica por ese amor (*philia*) por los iguales (*homo*). En este caso, la homofilia es un mecanismo que opera a nivel individual y que contribuye a explicar ciertas pautas de relaciones sociales. Esta es, por ejemplo, la noción que Shalizi y Thomas (2011) usan para examinar el problema de la distinción entre homofilia y contagio social.

³ Empleamos el término grupo o comunidad indistintamente, dejando el uso del término *clúster* para la técnica estadística *cluster analysis*.

Sin embargo, por otra parte, el término homofilia también se usa en la teoría de Blau para significar la proporción de individuos en una categoría de un parámetro dado de la estructura social que mantiene relaciones con individuos de esa misma categoría⁴. Como muestra Blau, esta proporción dependerá del peso relativo de cada categoría en el conjunto de la población, dado que si un grupo es más grande que otro, *ceteris paribus*, sus miembros tendrán menos oportunidades para mantener relaciones heterofílicas que los del grupo menor. Por lo tanto, en este segundo uso, los términos homofilia/heterofilia están vinculados a ciertas características de la población relativas al grado de cohesión social (Lozares y Verd, 2011), no a un tipo de motivación individual.

El grado de homofilia fruto de las oportunidades creadas por la distribución cuantitativa de la población es denominado «homofilia base» (*baseline homophily*). Cuando los lazos entre individuos con un cierto rasgo exceden la proporción de individuos con ese rasgo en la población, es decir, cuando dichas relaciones ocurren con más frecuencia que las que ofrecen el conjunto de oportunidades (*opportunity set*), entonces hablamos de «homofilia endogámica» (*imbreeding homophily*). En este segundo caso deben existir mecanismos psicosociales que hacen más frecuente de lo esperado las relaciones intragrupo y menos frecuente de lo posible las relaciones intergrupos. En el presente trabajo se usa el término homofilia para designar la tendencia, observada en una red de individuos, a que los contactos entre individuos similares en cierto rasgo o característica ocurran con mayor frecuencia que con individuos distintos en esa característica, con independencia de cuál sea el mecanismo concreto que causa dicha endogamia.

⁴ En la obra de Peter M. Blau la estructura social se concibe como una intersección de parámetros (nominales o graduados) que reflejan las relaciones que los individuos mantienen entre sí.

Los mecanismos que producen homofilia endogámica, a su vez, pertenecen a dos tipos de familias. La primera es la familia de mecanismos basados en la estructura de las oportunidades que facilitan mantener contactos sociales. Dichas oportunidades son ofrecidas por «focos sociales» (*social foci*) de interacción (Feld, 1981, 1982), como las organizaciones, y por las redes sociales creadas en la vida cotidiana. La segunda familia de mecanismos se relaciona principalmente con la toma de decisiones a nivel individual que conllevan la creación, mantenimiento y disolución de lazos sociales.

Dentro del segundo tipo de mecanismos se encuentran los procesos psicosociales de refuerzo. Así, como señalan Lazarsfeld y Merton en su análisis de la «homofilia de valores» (*value homophily*), dos desconocidos con valores similares probablemente crearán un lazo de amistad si tienen la oportunidad de encontrarse de forma regular (esto se correspondería con el refrán castellano «Dios los cría y ellos se juntan»). También existe la posibilidad de que dos individuos puedan verse motivados a modificar sus valores precisamente como resultado de su amistad, en un toma y daca que de forma progresiva va limando las divergencias iniciales (lo que se correspondería con el refrán castellano «dos que duermen en el mismo colchón se vuelven de la misma condición»). Este segundo proceso entra en la categoría de «influencia social» o «influencia de los pares» (*peer influence*) (Cohen, 1977). Una tercera posibilidad es que un individuo adquiera a través de algún mecanismo de imitación uno o varios rasgos de otro individuo con el que mantiene un lazo; en este caso la influencia no es recíproca y por ello lo denominamos «contagio social», aunque el resultado es difícil de distinguir empíricamente del caso anterior. Finalmente, la preferencia por lo similar o preferencia homofílica (que, como se señaló, suele confundirse con el propio concepto de homofilia) es posiblemente el mecanismo que opera con más frecuencia en casos de

«homofilia de estatus» (*status homophily*), en variables tales como el género, la etnia, el estatus educacional, la edad, etc.

Una cuestión adicional, con relación a las dos familias de mecanismos brevemente presentadas es cuál de ellas es más prominente en la explicación de las pautas de homofilia observadas empíricamente. Aunque se trata de una cuestión a la que, en el estado actual del arte, no se le puede dar una respuesta definitiva, sí se puede no obstante señalar que se ha hallado evidencia que apunta a que una parte importante de la explicación reside en elementos estructurales, como la existencia de entornos sociales, organizaciones, etc., cuya existencia atrae a individuos con características similares (McPherson y Smith-Lovin, 1987; Moody, 2001; Kossinets y Watts, 2009). En comparación con estos elementos estructurales, que moldean las oportunidades de los individuos para encontrar a otros similares, los mecanismos psicosociales que operan a un nivel individual parecen jugar un papel relativamente menor en la explicación de los patrones de relaciones, si bien la evidencia no es concluyente.

METODOLOGÍA

El trabajo se basa en los datos de un estudio previo (Linares y Kohl, 2017) en el que se distribuyó un cuestionario a 194 estudiantes de enseñanza secundaria posobligatoria de un instituto de la comunidad de las islas Canarias, que denominamos IES San Borondón⁵. La encuesta contenía tres módulos: 1) características sociodemográficas, 2) hábitos de ocio y consumo, 3) relaciones de amistad. En este tercer módulo, se pedía al estudiante que facilitara los nombres de hasta cuatro «ami-

gos con los que hablas de tus problemas» (Marsden, 1987). Igualmente se le preguntaba por el nombre de la persona con la que mantuviera, si era el caso, una relación amorosa. El resultado es una red con un componente gigante de 163 nodos, que tiene las características que se muestran en la tabla 1.

Un primer análisis mostró diversos grados de homofilia⁶ consistentes con los hallazgos de la literatura especializada (Kandel, 1978; McPherson, Miller y Smith-Lovin, 1986, 1987; Moody, 2001; Shrum, Cheek y MacD. Hunter, 1988), atendiendo a variables como el género, la edad, el hábito de fumar, la religión, etc., si bien no todos los índices resultan estadísticamente significativos⁷. El valor de los 86 índices de homofilia calculados y su nivel de significación estadística se muestran en el Apéndice 1. No obstante, como se señaló en la introducción, el foco de atención de este trabajo no es la significatividad estadística *per se*, sino la resiliencia que muestran los índices de homofilia significativos a las pequeñas perturbaciones que inevitablemente se producen en una red social debido a las dinámicas de creación y disolución de vínculos.

⁶ Existen diversas opciones para medir el grado de homofilia (Bojanowski y Corten, 2014). En el presente trabajo se ha optado por el índice de homofilia de Coleman (Coleman, 1957) que compara la proporción de relaciones homofílicas realmente existentes con la que cabría esperar si las relaciones entre individuos se establecieran de manera aleatoria. El IHC varía entre +1 y -1, representando el valor 0 el nivel de homofilia base (*baseline homophily*), que responde exclusivamente al tamaño relativo de las subpoblaciones, y que, por tanto, implica que no existe ningún mecanismo social induciendo las elecciones de los individuos.

⁷ La significatividad estadística se ha medido usando el test de Signorile, V. y O'Shea, R. M (1965) específicamente diseñado para el IHC. Signorile y O'Shea abordan el problema de hallar la probabilidad de que un valor dado del índice de homofilia, o un valor mayor, pueda obtenerse asumiendo que no existe relación entre las características del nodo del que sale el vínculo (el individuo que menciona) y el nodo al que llega (el individuo mencionado); hallando los parámetros μ y σ que permiten estandarizar el valor asumiendo el supuesto de una distribución normal. En lo que sigue el análisis se centra exclusivamente en los índices con significatividad estadística.

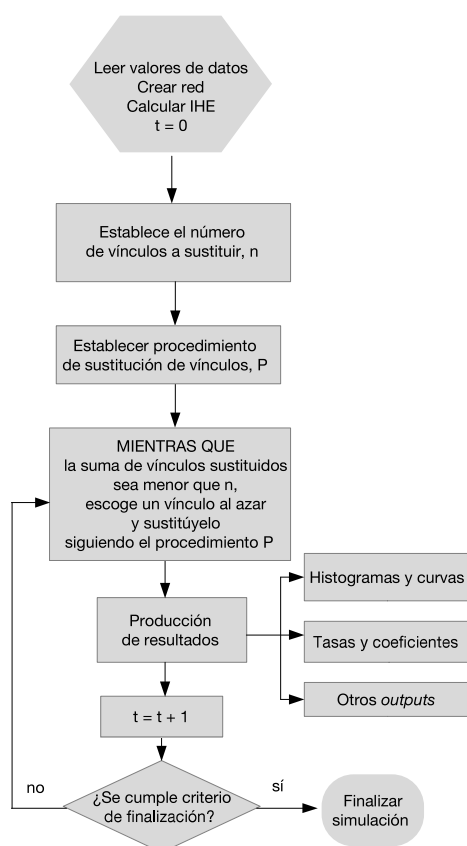
⁵ El trabajo de campo fue realizado entre el 25 de febrero de 2015 y el 1 de marzo de 2015. El cuestionario se distribuyó en las aulas a todos los estudiantes presentes, que constituyen el 67% del total de la población objeto de estudio.

TABLA 1. Características básicas del componente gigante de la red de alumnos de IES San Borondón

N.º de nodos	N.º de vínculos	Promedio de vínculos	Máximo n.º de vínculos	Densidad	Distancia geodésica promedio	Máx. distancia geodésica	N.º de comunidades (*)
163	275	3,35	9	0,021	5,92	14	14

(*) N.º de comunidades halladas por el algoritmo Clauset-Newman-Moore.

Fuente: Elaboración propia.

DIAGRAMA 1. Diagrama de flujo del modelo

Fuente: Elaboración propia.

Para abordar esta cuestión se ha construido un modelo de simulación basado en agentes (ABM) con la plataforma Netlogo (Wilenski y Rand, 2015)⁸. Este modelo, des-

crito de forma sencilla en el diagrama 1, contiene un conjunto de módulos que realizan de forma secuencial las siguientes operaciones:

⁸ El Apéndice 2 ofrece una breve descripción de este tipo de modelos.

1. Importación de los datos de la encuesta de los estudiantes de IES San Boron-

dón. Esta importación permite replicar *in silico* a los sujetos reales, que son «transformados» en agentes virtuales con los mismos atributos y las mismas relaciones que los reportados en la encuesta.

2. Para cada uno de los atributos de los agentes, calcular el índice de homofilia de Coleman (en adelante IHC) y la significatividad estadística del mismo, así como hallar la distribución de los agentes que muestran ese atributo entre las distintas comunidades de la red.
3. Para cada uno de los atributos de los agentes, proceder a una sustitución aleatoria de vínculos (*random rewiring of links*, en adelante RRL), al cálculo del nuevo IHC y al de su significatividad estadística.

El procedimiento RRL⁹ hace que la computadora escoja aleatoriamente uno de los vínculos homofílicos y lo sustituya por un vínculo nuevo; inmediatamente después calcula otra vez el índice de homofilia y valora de nuevo su significatividad estadística. Esta operación se repite hasta la sustitución de un número equivalente al 15% de los vínculos homofílicos de la categoría seleccionada. Al sustituir en cada caso un vínculo por otro las propiedades básicas de la red, como la densidad, la distancia geodésica media, etc., permanecen prácticamente idénticas puesto que tales medidas son insensibles a estas pequeñas perturbaciones; sin embargo, algunos de los índices de homofilia dejan de ser estadísticamente significativos.

Tal procedimiento se repite empleando distintos algoritmos para la sustitución de los vínculos de forma que, al final de la simulación, para cada uno de los atributos

sometidos a análisis se dispone de 780 casos de redes artificialmente manipuladas. Restringiendo el análisis a los 36 atributos (mostrados en las tablas 3a y 3b, en la siguiente sección) cuyos índices de homofilia, previos a cualquier manipulación, eran positivos y estadísticamente significativos, la población final es de 29.640 casos.

A título de ejemplo, las tablas 2a y 2b muestran los resultados de las simulaciones para las categorías «hombre» y «fumador». Todos los números contenidos en la tabla muestran los promedios de las 780 simulaciones. La primera columna indica el porcentaje de vínculos homofílicos sustituidos (%HLR) respecto al total de vínculos homofílicos de esa categoría; para cada uno de sus valores (0, 3, 6, 9, 12, 15), el resto de las columnas muestran el porcentaje de vínculos homofílicos sustituidos «respecto al total de vínculos de la red» (% *total links rewired*, TLR), el valor del índice de homofilia de Coleman (IHC), la diferencia entre el valor original del IHC y el valor nuevo (*drop value*, DV), la frecuencia relativa con la que el índice permanece significativo con probabilidades del 95% y del 99% y, finalmente, el número de comunidades en los que se distribuyen los individuos con la cualidad «hombre» o «fumador», número que permanece constante en las simulaciones puesto que el modelo no recomputa las comunidades de la red.

La información contenida en la tabla 2a indica que los fumadores están concentrados en seis grupos y que su índice de homofilia inicial tiene un valor estadísticamente significativo de 0,329. Conforme el modelo de simulación procede a la sustitución aleatoria de vínculos, este valor descendiendo hasta 0,222 una vez se han sustituido un 15% de los vínculos homofílicos (que constituyen un 1,97% de los vínculos totales de la red), valor que permanece siendo significativo.

⁹ Véase Apéndice 3 para una descripción detallada del procedimiento RRL.

TABLA 2A. Resultados de la simulación para la categoría «fumador»

%HLR	%TLR	IHC	DV	P95	P99	CC
0	0,00	0,329	0,000	1,00	1,00	6
3	0,39	0,307	-0,022	1,00	1,00	6
6	0,79	0,283	-0,047	1,00	1,00	6
9	1,18	0,260	-0,059	1,00	1,00	6
12	1,57	0,239	-0,091	1,00	1,00	6
15	1,97	0,222	-0,107	1,00	1,00	6

Notas: %HLR = porcentaje de vínculos homofílicos sustituidos respecto al total de vínculos homofílicos de esa categoría; %TLR = porcentaje de vínculos homofílicos sustituidos respecto al total de vínculos de la red; IHC = valor del índice de homofilia de Coleman; DV = *drop value* o diferencia entre el valor original del IHC y el valor nuevo; P95 = frecuencia con la que el índice permanece significativo (p-valor < 0,05); P99 = frecuencia con la que el índice permanece significativo (p-valor < 0,01); CC = número de comunidades en las que se distribuyen los individuos con la cualidad «fumador».

Fuente: Elaboración propia.

En el caso de la tabla 2b, para la categoría «hombre», los individuos están distribuidos en doce grupos y el valor del índice de homofilia inicial es 0,331, similar al de los fumadores e igualmente significativo.

No obstante, una vez se procede al proceso de sustitución aleatoria, el valor cae a 0,076 y el porcentaje de veces que permanece significativo con niveles de 0,05 y de 0,01 es 0,03% y 0,00% respectivamente.

TABLA 2B. Resultados de la simulación para la categoría «hombre»

%HLR	%TLR	IHC	DV	P95	P99	CC
0	0,00	0,331	0,000	1,00	1,00	12
3	1,31	0,274	-0,058	1,00	1,00	12
6	2,49	0,223	-0,108	1,00	0,98	12
9	3,81	0,171	-0,160	0,96	0,21	12
12	4,99	0,121	-0,210	0,51	0,01	12
15	6,30	0,076	-0,255	0,03	0,00	12

Notas: %HLR = porcentaje de vínculos homofílicos sustituidos respecto al total de vínculos homofílicos de esa categoría; %TLR = porcentaje de vínculos homofílicos sustituidos respecto al total de vínculos de la red; IHC = valor del índice de homofilia de Coleman; DV = *drop value* o diferencia entre el valor original del IHC y el valor nuevo; P95 = frecuencia con la que el índice permanece significativo (p-valor < 0,05); P99 = frecuencia con la que el índice permanece significativo (p-valor < 0,01); CC = número de comunidades en las que se distribuyen los individuos con la cualidad «hombre».

Fuente: Elaboración propia.

RESULTADOS

Índices de homofilia resistentes y no resistentes: importancia del número de comunidades en las que se distribuyen los individuos

Las tablas 3a y 3b muestran los promedios de los 780 valores de los índices de homofi-

lia calculados para cada atributo una vez el proceso RRL ha finalizado¹⁰. En la tabla 3a aparecen aquellos atributos en los que el valor del índice permanece significativo en todas las rondas de simulación (es decir, la

¹⁰ Por razones de espacio en las tablas aparecen los códigos de las variables. Las etiquetas de las mismas pueden hallarse en el Apéndice 1.

probabilidad de que el índice permanezca significativo es igual a 1) y que, por tanto, se corresponden con patrones de homofilia resilientes. En la tabla 3b aparecen aquellos atributos en los que la probabilidad de que el índice permanezca significativo es menor que 1. En esta segunda tabla puede observarse un amplio espectro de posibilidades, que abarca desde índices que nunca permanecen significativos, «estudiante de bachiller» (clase_der1), «padre empleado de la Administración pública» (padpuest1) y «tiene pareja» (pareja1), hasta índices que pueden permanecer significativos con una probabilidad muy alta, «pertenece a la asociación musical NSG» (asotipo1), «inició una relación de pareja hace menos de cinco meses» (iniciopar1) y «madre trabajando en el sector de la educación y los servicios sociales» (madsec3).

TABLA 3A. Índices de homofilia resistentes

Categoría	IHC(*)
asotipo2	0,238
clase_der2	0,467
clase1	0,278
clase2	0,059
clase3	0,364
clase4	0,313
drogatipo1	0,158
edad16	0,282
In-degree_status3	0,256
municipio1	0,208
municipio2	0,271
municipio3	0,209
municipio4	0,207
municipio5	0,195
religión1	0,145
género2	0,202
tabaco1	0,222
resi1	0,546

(*) Promedio del IHC tras el procedimiento RRL.

Fuente: Elaboración propia.

El primer paso en el análisis es descartar la posibilidad de que la diferencia entre la

evolución de los índices de la tabla 3a y los de la tabla 3b obedezca simplemente al número absoluto de vínculos homofílicos existentes en cada caso. Así, dado que la sustitución aleatoria de vínculos necesariamente hace decrecer el valor del IHC (pues los vínculos nuevos, con muy alta probabilidad, no serán vínculos homofílicos), podría darse el caso de que el número de vínculos sustituidos y la probabilidad de que el IHC permanezca significativo estén negativamente asociados.

TABLA 3B. Índices de homofilia no resistentes

Categoría	IHC(*)	fi (**)
asotipo1	0,123	0,92
clase_der1	-0,010	0,00
deportipo0	-0,050	0,01
deportipo1	0,022	0,01
edad18	0,067	0,57
iniciopar1	0,118	0,91
iniciopar3	0,069	0,40
leetipo1	0,081	0,44
madsec3	0,127	0,91
musitipo1	0,124	0,91
musitipo2	0,090	0,72
padpuest1	0,010	0,00
padsec5	0,066	0,46
pareja1	0,021	0,00
religión0	0,090	0,48
género1	0,076	0,03
videojuego1	0,066	0,08
resi2	0,315	0,84

(*) Promedio del IHC tras el procedimiento RRL.

(**) Frecuencia relativa de índice estadísticamente significativo ($p < 0,05$).

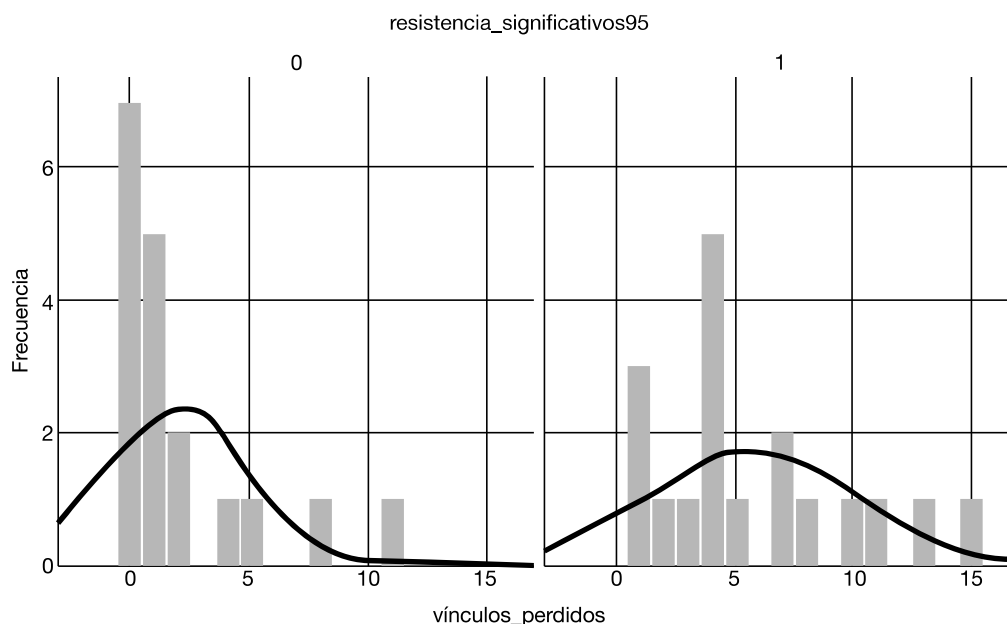
Fuente: Elaboración propia.

Como puede apreciarse en el gráfico 1, este no es el caso. El panel de la derecha (1) muestra el histograma de los vínculos perdidos por los índices que permanecen siendo significativos tras el procedimiento RRL. Puede apreciarse fácilmente que aproximadamente la mitad de los índices resisten la pérdida de más de seis

vínculos, hasta un máximo de quince. En el panel de la izquierda (0) se aprecia que la mayoría de los índices que no permane-

cen significativos no resisten la pérdida de más de unos pocos vínculos, o incluso un solo vínculo.

GRÁFICO 1. Distribuciones del número de vínculos reasignados



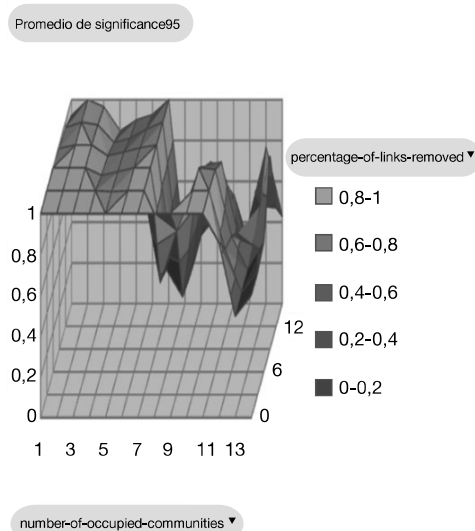
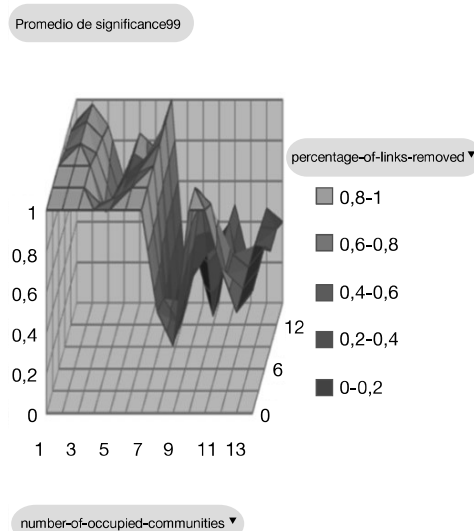
Nota: El panel 0 muestra la distribución del número de vínculos homofílicos perdidos en el caso de los índices que «no» permanecen significativos, mientras que el panel 1 muestra la misma distribución para el caso de los índices que «sí» permanecen significativos.

Fuente: Elaboración propia.

Por otra parte, como se muestra en los gráficos 2a y 2b, sí existe una clara relación entre la probabilidad de que el IHC permanezca significativo (valores del eje y) respecto al número de comunidades¹¹ entre los que están distribuidos los individuos con el atributo correspondiente (valores del eje x), con relativa independencia del por-

centaje de vínculos sustituidos por el procedimiento RRL (valores del eje z). Así, la probabilidad de que el IHC permanezca significativo disminuye de forma muy evidente si el número de comunidades es superior a siete. A título de ejemplo, recuérdese que los «hombres» se encuentran repartidos en doce grupos y los «fumadores» en seis. En el segundo caso el índice de homofilia resiste el procedimiento RRL, mientras que en el primero la probabilidad de que permanezca significativo comienza a caer sensiblemente a partir de la sustitución del 9% de los vínculos.

¹¹ Una comunidad es un subconjunto de la población de individuos en el que la densidad de relaciones es mayor que en el conjunto de la red. Los algoritmos de identificación de comunidades asignan cada individuo de la red a un único grupo.

GRÁFICO 2A. Probabilidad de que el IHC permanezca significativo ($p < 0,05$)**GRÁFICO 2B.** Probabilidad de que el IHC permanezca significativo ($p < 0,01$)

Nota: Evolución de la probabilidad de que el índice permanezca significativo (eje y) en función del número de comunidades a las que pertenecen los individuos (eje x) y del porcentaje de vínculos homofílicos sustituidos (eje z).

Fuente: Elaboración propia.

Identificación de las comunidades más frecuentes en los índices de homofilia resistentes

La evidencia mostrada en el apartado anterior suscita el interrogante de si todas las comunidades tendrán la misma importancia en la «producción» de índices resistentes. A tal efecto hemos construido dos índices que denominamos de «extensión» y de «intensidad», cuya finalidad es medir en qué grado cada una de las catorce comunidades identificadas por el algoritmo Clauset-Newman-Moore contribuyen a los IHC resistentes o no resistentes. El término «contribución», relativo a una comunidad dada C_i , se emplea para denotar que al menos una fracción de los individuos que muestran el atributo cuyo IHC se calcula pertenece a dicha comunidad. Estos índices se definen de la siguiente manera:

- El «índice de intensidad» de una comunidad, IIC_i , mide la probabilidad de que los índices IHC a los que contribuye C_i permanezcan significativos.
- El «índice de extensión» de una comunidad, IEC_i , mide la proporción de índices IHC que permanecen significativos en los que C_i contribuye, poniéndola en relación con la proporción de índices IHC que permanecen significativos en los que C_i no contribuye¹².

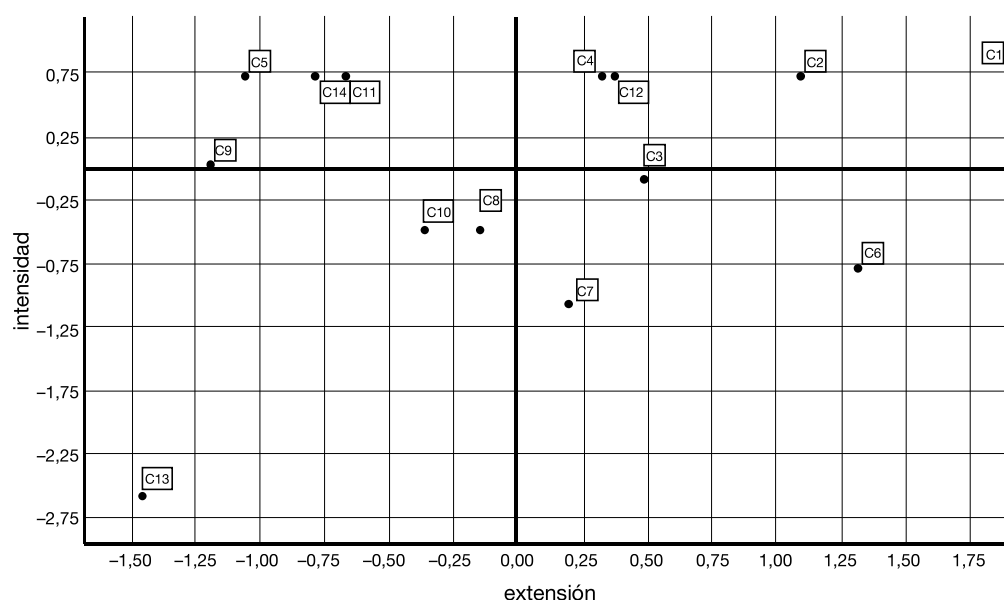
El diagrama 2 muestra la disposición de las catorce comunidades en un espacio cartesiano en el que los ejes de ordenadas y abscisas representan sus puntuaciones tipificadas en cada uno de los índices. Se aprecia nítidamente que cuatro comunida-

¹² Una exposición detallada de la construcción del índice de extensión puede hallarse en el Apéndice 3.

des (los números 1, 2, 4 y 12) puntúan por encima de la media en ambas dimensiones, desviándose aproximadamente 0,75 desviaciones típicas del valor medio del índice de intensidad, y entre 0,25 a 1,80 desviaciones típicas del valor medio del índice de extensión. A su vez, como puede apreciarse en la tabla 4, las puntuaciones de ambos índices están fuertemente correlacionadas con algunas características de las

comunidades: existe una fuerte correlación negativa de ambos índices con la densidad de lazos (la relación entre lazos existentes y lazos posibles) y positiva tanto con la distancia promedio como con la máxima distancia geodésica entre dos nodos. Estas correlaciones sugieren que ciertos rasgos topológicos de las comunidades podrían facilitar que los índices en los que estas participen permanezcan significativos.

DIAGRAMA 2. *Dispersión de las comunidades según «extensión» e «intensidad»*



Fuente: Elaboración propia.

TABLA 4. *Correlaciones entre índices de extensión e intensidad y características de las comunidades*

	Densidad de lazos	Distancia geodésica promedio	Distancia geodésica máxima
IEC_i	-0,772	0,771	0,668
IIC_i	-0,614	0,592	0,573

Fuente: Elaboración propia.

TABLA 5. Algunos atributos destacados del núcleo homofílico

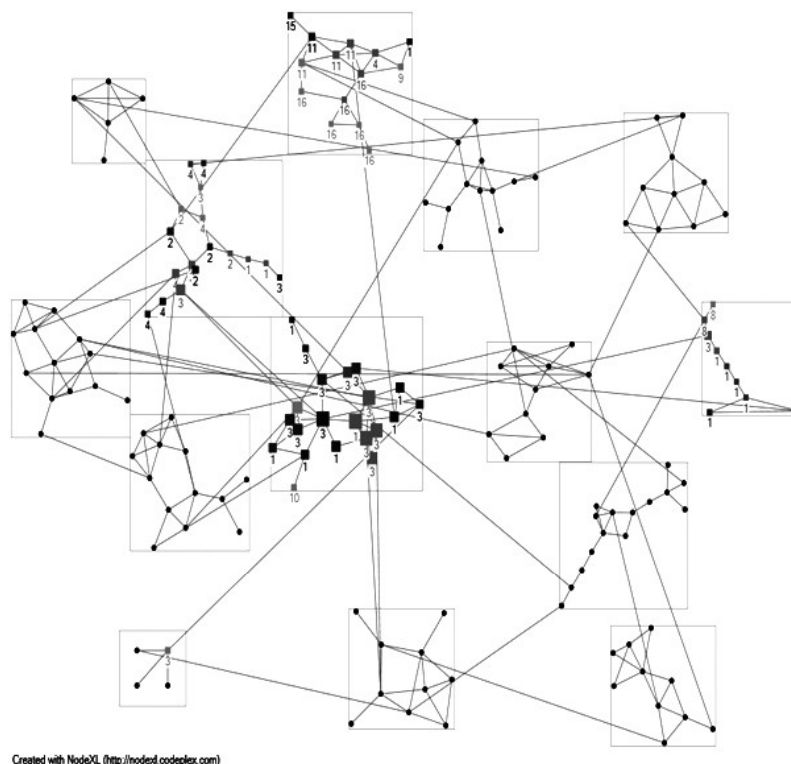
ATRIBUTO	Frecuencia en el núcleo homofílico (N = 59)	Frecuencia en el componente gigante (N = 163)
MUNICIPIO1	20,9%	52,00%
MUNICIPIO2	30,6%	16,00%
GÉNERO HOMBRE	52,0%	56,00%
GÉNERO MUJER	48,0%	44,00%
EDAD 16	30,5%	24,50%
EDAD 17	45,8%	53,30%
EDAD 18	13,6%	14,11%
EDAD 19	6,8%	3,70%
RESIDENCIA (*)	61,3%	31,00%
B. CIENCIAS (**)	36,0%	41,00%
MÓDULOS (***)	24,0%	17,00%

(*) Tiene plaza en la residencia de estudiantes.

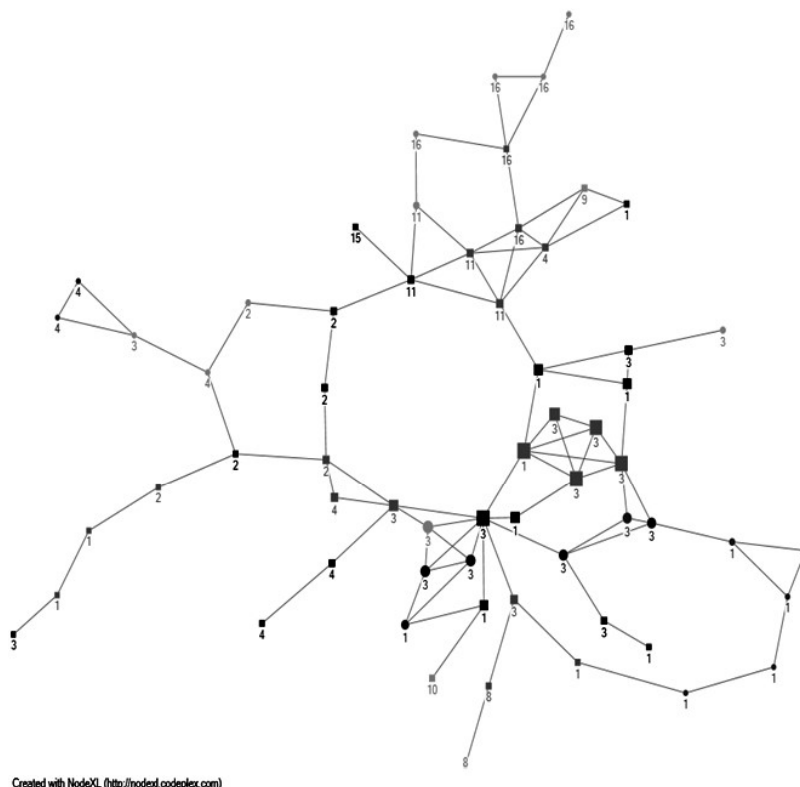
(**) Cursa el bachillerato de ciencia y tecnología.

(***) Cursa un módulo de formación profesional.

Fuente: Elaboración propia.

GRAFO 1A. Componente gigante del IES San Borondón

Fuente: Elaboración propia.

GRAFO 1B. *Núcleo homofílico*

Fuente: Elaboración propia.

En la tabla 5 se muestran algunas características diferenciales de los individuos que forman parte de las comunidades señaladas, que denominamos «núcleo homofílico», cuya disposición en el componente gigante de la red se visualiza en el grafo 1a (por otra parte, el grafo 1b muestra exclusivamente los nodos pertenecientes al núcleo homofílico). De especial relevancia es la mayor proporción de estudiantes que tienen plaza en la residencia escolar, correlativa con una menor proporción de los individuos de procedencia del municipio 1 (la localidad donde se encuentra el IES San Borondón¹³), lo que

sugiere que una parte de los patrones de homofilia resilientes están vinculados al tejido relacional que se genera en la convivencia de los jóvenes en la residencia de estudiantes.

Resiliencia de los patrones de homofilia

En los apartados anteriores hemos mostrado una doble evidencia: 1) cuando los individuos con el atributo objeto de estudio están concentrados en siete o menos comunidades, los índices de homofilia tienen mayor probabilidad de permanecer estadísticamente significativos después del proceso de RRL, y 2) las comunidades 1, 2, 4 y 12 participan en una mayor proporción en índices que permanecen significativos y, al

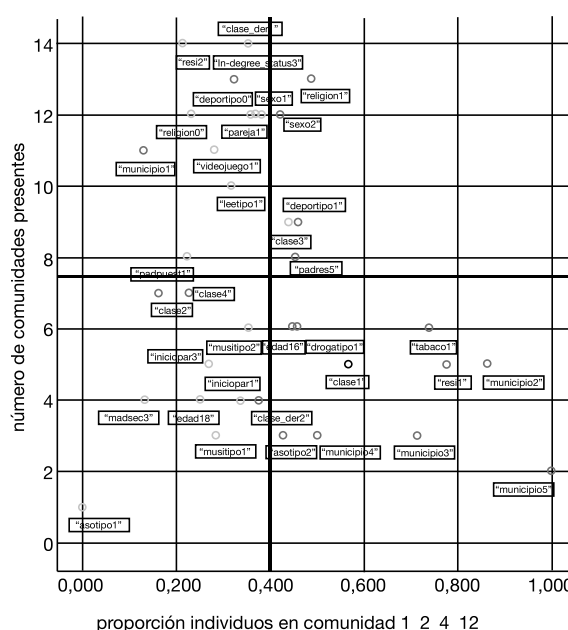
¹³ Debido a la orografía de la isla, el municipio donde se encuentra el IES San Borondón dispone de una residencia para los estudiantes procedentes de otros municipios.

mismo tiempo, su participación incrementa la probabilidad de que un índice permanezca significativo.

El gráfico 3 muestra la disposición de los índices de homofilia en un espacio cartesiano definido por la variable «proporción de individuos en las comunidades 1, 2, 4 y 12» (eje de abscisas) y la variable «número de comunidades presentes» (eje de ordenadas). Así, por ejemplo, el punto «municipio5» en el extremo inferior derecha representa el índice de homofilia entre los estudiantes que proceden de ese municipio de San Borondón; dichos estudiantes es-

tán distribuidos en dos grupos y el 100% de ellos pertenece al conjunto de las comunidades 1, 2, 4 y 12. Por otra parte, el punto «resi2», en el extremo superior izquierda, representa el índice de homofilia de los individuos que no pertenecen a la residencia de estudiantes, los cuales están distribuidos en catorce grupos, y solo el 20% de ellos pertenece al conjunto de las comunidades 1, 2, 4 y 12. Las rectas representan las medias de las variables, y la mayoría de los índices que permanecen significativos aparecen, como cabía esperar, en el cuadrante inferior-derecho.

GRÁFICO 3. Distribución de los índices de homofilia



Fuente: Elaboración propia.

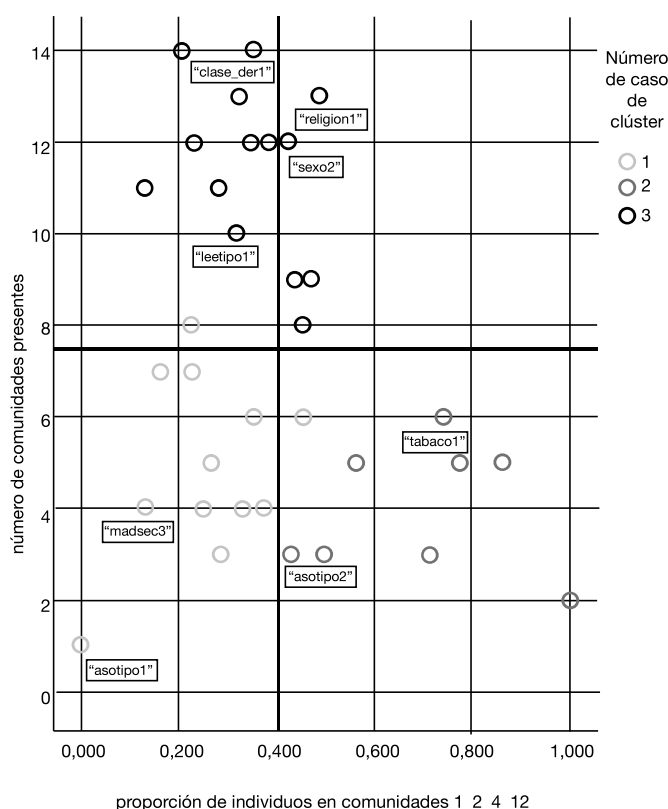
Entre los índices que sí son resistentes encontramos principalmente tres casos: en primer lugar, índices que son resultado de un foco de atracción para los individuos, como puede ser la clase en la que estudian, la residencia de estudiantes o el club de fútbol. En segundo lugar, índices que son el resultado de relaciones de vecindad preexistentes (es

decir, todos los vinculados a los municipios de origen), lo que ilustra muy bien el concepto de homofilia de estatus. Finalmente, índices que reflejan comportamiento susceptible de contagio social, como el consumo de tabaco y marihuana, que ilustran el concepto de homofilia de valores. La lista completa puede hallarse en la tabla 3a, arriba.

El gráfico 4 muestra la disposición de los índices de homofilia en el mismo espacio cartesiano que el anterior. En este caso, los índices aparecen clasificados en tres clústeres identificados por el algoritmo de las k-medias. Es interesante señalar que el algoritmo ha clasificado en el mismo clúster #3 a la mayoría de los índices de homofilia que se corresponden con características que están ampliamente distribuidas en la red (ocho o más comunidades) y que,

precisamente por ello, los individuos que pertenecen a las comunidades 1, 2, 4 y 12 representan un porcentaje menor del total. Ejemplos de estos índices son «clase_der1» (individuos que cursan bachillerato), «religión1» (individuos que se definen como católicos), «videojuego1» (usuarios de videojuegos) «leetipo1» (lectores) o «género2» (individuos mujeres). La inmensa mayoría de los elementos de este clúster #3 no resisten el procedimiento RRL.

GRÁFICO 4. Clusterización de los índices de homofilia (algoritmo de las k-medias)



Fuente: Elaboración propia.

Por otra parte, el clúster #1 está constituido por un conjunto heterogéneo de índices que se corresponden con características que no están ampliamente distribuidas entre la población y en las que el porcen-

taje de individuos que pertenecen a las comunidades 1, 2, 4 y 12 es relativamente bajo. Ejemplos son: «musitipo2» (individuos a los que les gusta la música pop), «inicio-par3» (individuos que tienen pareja desde

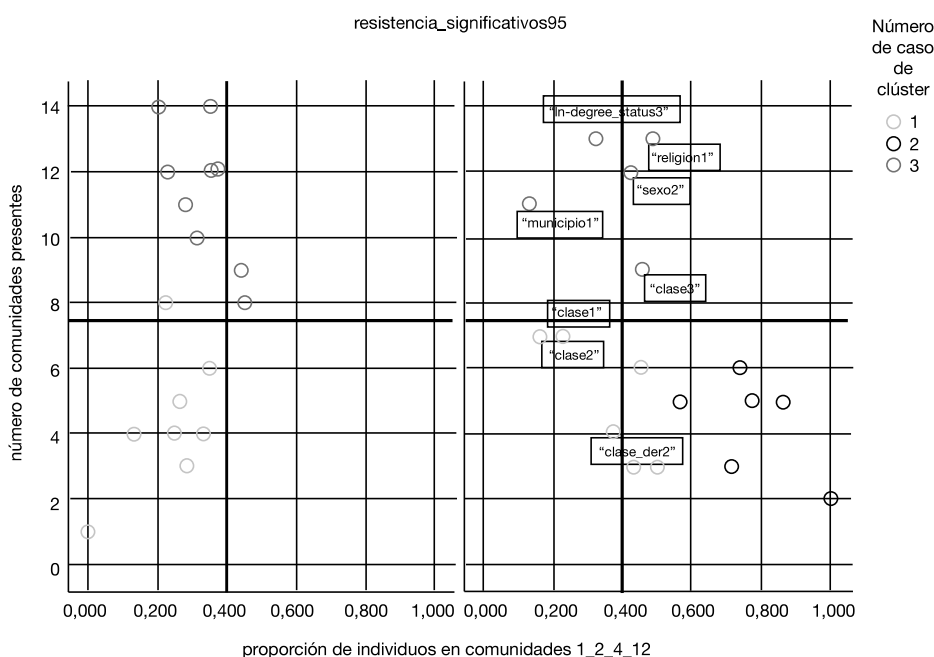
hace más de un año), y «madsec3» (madre trabajadora del sector de la educación y servicios). Aunque la mayoría de los índices de este clúster no resisten el procedimiento RRL, algunos de ellos sí.

Finalmente, todos los elementos del clúster #2 (características no ampliamente distribuidas pero con un alto grado de presencia de las comunidades 1, 2, 4 y 12) se corresponden con índices que permanecen significativos, tales como «tabaco1» (individuos fumadores), «asotipo2» (individuos del club de fútbol), «clase1» (individuos de primer curso del bachiller de ciencia y tecnología) o «resi» (individuos con plaza en la residencia de estudiantes).

Para concluir la descripción, el gráfico 5 muestra la información del gráfico 4 divi-

dida en dos paneles: el de la derecha (1) representa los índices que resisten el procedimiento RRL y el de la izquierda (0) los que no. Este gráfico nos permite visualizar claramente los casos «desviados», es decir, los índices que, aunque no se hallan en el cuadrante inferior-derecha, resisten el procedimiento RRL. Fijándonos en el panel de la derecha observamos, por una parte, tres casos del clúster #1 que se encuentran en el cuadrante inferior-izquierda: «clase2» (alumnos de 2.º de bachillerato de ciencias y tecnología), «clase4» (alumnos de 2.º de bachillerato de ciencias sociales y humanidades) y «clase_der2» (alumnos que cursan módulos de formación profesional); estos forman parte del grupo de índices de homofilia resistentes que responden a la existencia de un foco de actividad.

GRÁFICO 5. Casos «desviados»



Fuente: Elaboración propia.

Por otra parte, observamos cinco índices clasificados en el clúster #3 (en la mitad superior). De ellos «municipio1» y «clase3» (alumnos de 1.º de bachillerato de ciencias sociales y humanidades) pertenecen igualmente a los tipos de índices resistentes ya citados. La presencia en este panel de «in-degree_status3» (individuos que reciben tres o más menciones) es un resultado consistente con hallazgos similares de la literatura especializada (Maggio y Gari, 2012), relativos a las oportunidades ofrecidas por la estructura de las redes sociales: las personas con más relaciones también se relacionan más entre ellas. Finalmente, la resistencia de «religión1» (religión católica) parece un caso ciertamente anómalo, como lo es «género2» (individuo mujer), cuyas características son muy similares a las de «género1» (individuo hombre) que, sin embargo, no es un índice resistente.

DISCUSIÓN Y CONCLUSIONES

En este trabajo se ha tratado de abordar la pregunta de qué es lo que hace que un patrón de homofilia sea robusto y, por lo tanto, pueda considerarse como una característica estable de una red de individuos. Hemos entendido que un patrón robusto o resiliente es aquel que produce un índice de homofilia que permanece estadísticamente significativo aun cuando la red de relaciones sufra un número de perturbaciones que, a efectos prácticos, se ha fijado en la sustitución aleatoria de un 15% de los vínculos homofílicos.

La evidencia hallada apunta a que la distribución de los individuos entre las distintas comunidades de la red es un factor clave, si bien en un sentido contraintuitivo: parece existir una relación inversa entre el número de comunidades en las que se hallan distribuidos los individuos y la probabilidad de que el índice de homofilia permanezca significativo tras aplicar el

procedimiento RRL. Así, la mayoría de las características ampliamente distribuidas en la población no dan lugar a índices de homofilia resistentes (con algunas excepciones notables, como la característica «mujer»). Esta relación inversa parece deberse a que unas pocas comunidades (en concreto, los números 1, 2, 4 y 12 en nuestro estudio de caso) contribuyen de manera desproporcionada a la producción de los patrones de homofilia resilientes. Esto implica, paradójicamente, que los índices que permanecen significativos no proporcionan información sobre el conjunto de la red, sino sobre algunas de las comunidades que la constituyen y que, por tanto, la afirmación «en el conjunto de la red R existe homofilia con relación al atributo A » es una generalización que no es cierta sin más cualificación.

Un segundo resultado es que las comunidades más relevantes en la producción de patrones de homofilia resilientes parecen tener rasgos topológicos específicos, como son mayor amplitud y menor densidad que la media. No obstante, dado que el número de comunidades en la red analizada es reducido, este resultado requiere la replicación del estudio con otras redes para ser confirmado o descartado.

Con relación a los mecanismos que explican los patrones de homofilia, los resultados son consistentes con los estudios que subrayan la importancia de los focos de actividad y de la estructura de oportunidades de las redes, dado que la mayoría de los índices resistentes se corresponden con estos casos. Un importante número de características (aunque no todas) que pueden etiquetarse como homofilia de estatus, así como algunas características susceptibles de contagio social, muestran índices de homofilia resistentes.

No obstante, no son resistentes la mayoría de los índices de homofilia de características asociadas al estatus de los padres de los estudiantes; resultado que, por otra parte, es consistente con el marco concep-

tual de Peter M. Blau, ya que cuando existen parámetros secantes, la asociación homofílica en un parámetro implica necesariamente asociación heterofílica en otro. Así, por ejemplo, para que el índice de homofilia de estudiantes cuyos padres trabajan en el sector servicios fuera resistente tendría que no serlo la asociación basada en el curso. Dicho de otra manera, la asociación homofílica entre individuos que cursan, digamos, primero de ciencias naturales, produce necesariamente asociaciones heterofílicas entre individuos con padres de distinto estatus ocupacional.

La robustez de los resultados de este trabajo depende, obviamente, de la posibilidad de replicarlos con otras poblaciones de individuos y con otros algoritmos tanto de búsqueda de comunidades como de sustitución de vínculos. Con todo, la idea de que un patrón de comportamiento será resiliente si se encuentra concentrado en ciertas comunidades de la red tiene un sentido teórico que trasciende el caso particular que hemos analizado. Es razonable sospechar que puede tratarse de un resultado generalizable a la clase de redes tipo «mundo pequeño». Si esto fuera así, el procedimiento de análisis diseñado para esta investigación sería útil para un amplio número de ámbitos empíricos, algunos pertenecientes a otras disciplinas, como la economía, la antropología o la ecología, en las que cabe esperar que los objetos de análisis estén incrustados en redes *small world*.

BIBLIOGRAFÍA

- Aral, Sinan; Muchnik, Lev y Sundarajan, Arun (2009). «Distinguishing Influence-Based Contagion from Homophily Driven Diffusion in Dynamic Networks». *PNAS*, 16(51).
- Blau, Peter M. (1977). *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. New York: Free Press.
- Bojanowski, Michel y Corten, Rense (2014). «Measuring Segregation in Social Networks». *Social Networks*, 39: 14-32.
- Cohen, Jere M. (1977). «Sources of Peer Group Heterogeneity». *Sociology of Education*, 50: 227-241.
- Coleman, James S. (1957). «Relational Analysis: the Study of Social Organization with Survey Methods». *Human Organization*, 17(4): 28-36.
- DiMaggio, Paul y Garip, Filiz (2012). «Networks Effects in Social Inequality». *Annual Review of Sociology*, 38: 93-118.
- Feld, Scott (1981). «The Focused Organization of Organizational Ties». *American Journal of Sociology*, 86: 1015-1035.
- Feld, Scott (1982). «Structural Determinants of Similarity among Associates». *American Sociological Review*, 47: 797-801.
- Kandel, Denise B. (1978). «Homophily, Selection and Socialization in Adolescent Friendship». *American Journal of Sociology*, 84(2): 427-436.
- Kossinets, Gueorgi y Duncan, Watts (2009). «Origins of Homophily in an Evolving Social Network». *American Journal of Sociology*, 115(2): 405-50.
- Lazarsfeld, Paul F. y Merton, Robert K. (1954). «Friendship as a Social Process: A Substantive and Methodological Analysis». En: Berger, M. (ed.). *Freedom and Control in Modern Society*, pp. 18-66. New York: Van Nostrand.
- Linares, Francisco (2018a). *Sociología y teoría social analíticas: la ciencia de las consecuencias inintencionadas de la acción*. Madrid: Alianza Editorial.
- Linares, Francisco (2018b). «Agent Based Models and the Science of Unintended Consequences of Social Action»/«Los modelos basados en agentes y la ciencia de las consecuencias inintencionadas de la acción». *Revista Española de Investigaciones Sociológicas*, 162: 21-37.
- Linares, Francisco y Kohl, Mona (2017). «Social Networks and Homophily Patterns among Post-Secondary Students in San Borondón». *I Encuentro de Sociología Analítica y Migraciones*. Universidad de A Coruña.
- Lozares, Carlos y Verd, Joan M. (2011). «De la homofilia a la cohesión social y viceversa». *Redes-Revista Hispana para el Análisis de Redes*, 20(2): 29-50.
- Marsden, Peter V. (1987). «Core Diffusion Networks among Americans». *American Sociological Review*, 52: 122-131.
- McPherson, Miller y Smith-Lovin, Lynn (1986). «Sex Segregation in Voluntary Associations». *American Sociological Review*, 51: 61-79.

- McPherson, Miller y Smith-Lovin, Lynn (1987). «Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups». *American Sociological Review*, 55: 370-379.
- McPherson, Miller; Smith-Lovin, Lynn y Cook, James M. (2001). «Birds of a Feather: Homophily in Social Networks». *Annual Review of Sociology*, 27: 415-444.
- Moody, James (2001). «Race School Integration, and Friendship Segregation in America». *American Journal of Sociology*, 107(3): 679-716.
- Shalizi, Cosma R. y Thomas, Andrew C. (2011). «Homophily and Contagion are Generically Confounded in Observational Social Network Studies». *Sociological Methods and Research*, 40(2): 211-239.
- Shrum, Wesley; Cheek, Neil H. Jr. y MacD. Hunter, Sandra (1988). «Friendship in School: Gender and Racial Homophily». *Sociology of Education*, 61: 227-239.
- Signorile, Vito y O'Shea, Robert M. (1965). «A Test of Significance for the Homophily Index». *American Journal of Sociology*, 70(4): 467-470.
- Wilensky, Uri y Rand, William (2015). *An Introduction to Agent-based Modeling*. Cambridge, Massachusetts: The MIT Press.

RECEPCIÓN: 27/05/2020

REVISIÓN: 12/11/2020

APROBACIÓN: 23/12/2020

APÉNDICE 1. LISTADO DE VARIABLES Y CATEGORÍAS

Variable	Categoría	Descripción	N	IHC
ASOTIPO (pertenencia a una asociación)	0	No pertenece a una asociación	117	0,128
	1	Asociación musical NSG	9	0,170**
	2	Unión Deportiva G	11	0,305**
CLASE_DER (opción de enseñanza secundaria)	1	Cursa bachillerato	133	0,608**
	2	Cursa módulos FP	21	0,612**
CLASE (curso de bachiller)	1	1.º bachillerato CC. TT.	30	0,401**
	2	2.º bachiller CC. TT.	38	0,676**
	3	1.º bachillerato CC. SS.	39	0,540**
	4	2.º bachiller CC. SS.	26	0,438**
DEPORTIPO (práctica de deporte)	0	No hace deporte	69	0,153**
	1	Juega al fútbol	37	0,164**
	2	Acude al gimnasio	14	0,192**
DROGATIPO (consumo de drogas)	0	No consume drogas	131	0,090
	1	Sí consume drogas	29	0,240**
EDAD	16	16 años	40	0,419**
	17	17 años	87	0,234**
	18	18 años	23	0,127**
FRECALC (consumo de alcohol)	0	No consume alcohol	33	0,089
	1	Consume alcohol solo en fiestas	122	0,398**
FRECDROG (frecuencia de consumo de drogas)	0	Nunca	131	0,090
	1	Consume drogas solo en fiestas	18	0,169**
	2	Consumo también en otras ocasiones	12	0,195**
FRECTAB (frecuencia de consumo de tabaco)	0	No consume tabaco	133	0,264**
	1	Consume tabaco solo en fiestas	18	0,057
	2	Consume también en otras ocasiones	11	0,414**
GÉNERO	1	Masculino	91	0,340**
	2	Femenino	72	0,544**
HORFINDE (horarios establecidos por los padres)	1	Tienen horarios de fines de semana	29	-0,021
	2	No tienen horarios los fines de semana	130	0,113
INICIOTAB (inicio del consumo de tabaco)	0	No es fumador	133	0,264**
	1	Se inició antes de los 15 años	11	0,212**
	2	Se inició con 15 años o mayor	11	0,016

APÉNDICE 1. LISTADO DE VARIABLES Y CATEGORÍAS (CONTINUACIÓN)

Variable	Categoría	Descripción	N	IHC
INICIOPAREJA (inicio de la relación de pareja actual)	1	Inició una relación de pareja hace menos de 5 meses	17	0,203**
	2	Inició una relación de pareja hace entre 5 meses y un año	18	0,042
	3	Inició una relación de pareja hace un año o más	25	0,164**
IN-DEGREE_STATUS (número de menciones recibidas en el cuestionario)	1	Ha recibido 0 o 1 menciones en el cuestionario	57	-0,730**
	2	Ha recibido 2 o 3 menciones (2 = mediana) en el cuestionario	67	0,046
	3	Ha recibido más de 3 menciones en el cuestionario	39	0,352**
LEETIPO (afición a la lectura)	1	Sí lee	50	0,153**
	2	No lee	113	-0,032
MADESEC (sector de trabajo de la madre)	1	Agricultura, ganadería, pesca	1	NC
	2	Hostelería y turismo	33	0,078
	3	Educación y servicios sociales	26	0,206**
	4	Comercio	15	-0,086
	5	Construcción	0	NC
	6	Sanidad	11	0,156**
MADPUEST (puesto de trabajo de la madre)	1	Empleada en Administración pública	37	0,045
	2	Empleada en empresa	53	0,095*
	3	Propietaria de una empresa o negocio	18	0,064
MUSITIPO (tipo de música a la que es aficionado)	1	Menciona la música <i>reggaeton</i>	18	0,211**
	2	Menciona la música <i>pop</i>	28	0,206**
MUNICIPIO (municipio de origen)	1	municipio SS	85	0,425**
	2	municipio VG	26	0,443**
	3	municipio AG	9	0,299**
	4	municipio HE	14	0,288**
	5	municipio AL	15	0,270**
	6	municipio VH	14	0,305**
NUNPARS (número de relaciones de pareja pasadas)	0	0 parejas en los últimos 18 meses (sin contar la actual)	62	0,012
	1	1 parejas en los últimos 18 meses (sin contar la actual)	42	0,070
	2	2 parejas en los últimos 18 meses (sin contar la actual)	21	0,025
	3	3 o más parejas en los últimos 18 meses (sin contar la actual)	16	-0,250

APÉNDICE 1. LISTADO DE VARIABLES Y CATEGORÍAS (CONTINUACIÓN)

Variable	Categoría	Descripción	N	IHC
PADSEC (sector de trabajo del padre)	1	Agricultura, ganadería, pesca	11	0,027
	2	Hostelería y turismo	24	-0,155
	3	Educación y servicios sociales	14	0,193**
	4	Comercio	7	0,019
	5	Fontanería, electricidad, construcción	32	0,176**
	6	Sanidad	1	NC
PADPUEST (puesto de trabajo del padre)	1	Empleado de Administración pública	29	0,085*
	2	Empleado de empresa	55	0,067
	3	Propietario de una empresa o negocio	28	0,040
PADRELI (creencias religiosas de los padres)	0	Padres sin creencias religiosas	24	-0,033
	1	Al menos un progenitor católico	122	-0,012
	2	Al menos un progenitor de otra religión	14	-1,000**
PAGATIPO (asignación recibida de los padres)	1	Sí recibe una paga semanal de sus padres	31	-0,018
	2	No recibe una paga semanal de sus padres	132	-0,026
PAREJA (relación actual de pareja)	1	Sí tiene pareja en la actualidad	63	0,139*
	2	No tiene pareja en la actualidad	96	0,084
RESI (plaza en la residencia de estudiantes)	1	Tiene plaza en la residencia de estudiantes	51	0,706**
	2	No tiene plaza en la residencia de estudiantes	112	0,635**
RELIGIÓN (creencias religiosas)	0	No tiene creencias religiosas	65	0,314**
	1	Tiene creencias católicas	80	0,319**
	2	Tiene otras creencias distintas de la religión católica	16	0,000**
TABACO (consumo de tabaco)	1	Consume tabaco	30	0,329**
	2	No consume tabaco	133	0,429**
VIDEOJUEGO (afición a los videojuegos)	1	Es aficionado a los videojuegos	56	0,177**
	2	No es aficionado a los videojuegos	108	0,267**

Notas: N = nº de individuos; IHC = índice de homofilia de Coleman; NC = No Computable; (*) = Significativo ($p < 0,05$); (**) = Significativo ($p < 0,01$).

Fuente: Elaboración propia.

APÉNDICE 2. SOBRE LOS MODELOS BASADOS EN AGENTES

Existe una amplia familia de técnicas para programar modelos y ejecutar simulaciones. La técnica que se emplea en este artículo es la de modelos basados en agentes, diferentes de otras técnicas de la misma familia que han tenido también repercusión en las ciencias sociales, como los de dinámica de sistemas.

La elaboración de este tipo de modelos consiste en redactar una secuencia de instrucciones que detallan las variables que caracterizan el sistema (en el caso de este trabajo, los índices de homofilia), las características de los agentes (los atributos de los individuos reales) y las reglas por las que cambian ciertos atributos de los agentes (la sustitución de los vínculos con otros agentes) y, a su vez, las características del sistema (los nuevos índices de homofilia).

La computadora ejecuta las reglas establecidas de forma recursiva hasta que se cumple la condición de finalización de la simulación. Cada simulación es repetida un número N de veces, manipulando diversos parámetros para obtener una «población» de casos con diversidad suficiente para realizar análisis estadísticos de sensibilidad. Una explicación más detallada puede hallarse en Linares (2018a, 2018b).

APÉNDICE 3. PSEUDOCÓDIGO DEL PROCEDIMIENTO RRL

1. Establece el número de vínculos a sustituir, $N = 0,15$ multiplicado por el total de vínculos homofílicos entre individuos con el atributo A .
2. Establece el número de vínculos sustituidos $M = 0$.
3. Mientras que $M < N$ repite los pasos 4 a 11.
4. Elige al azar un individuo i entre el conjunto de individuos con vínculos homofílicos respecto al atributo A .
5. Elige al azar un vínculo, v_{ij} , del conjunto de vínculos homofílicos de i respecto al atributo A .
6. Elimina v_{ij} .
7. Elige según el procedimiento P_i (P_i pertenece al conjunto de procedimientos P para la creación de nuevos vínculos) un individuo j ($j \neq i$) del conjunto de individuos con vínculos homofílicos respecto al atributo A .
8. Crea el vínculo v_{ij} .
9. Calcula el valor del índice IHC.
10. Calcula la significatividad estadística del valor del IHC.
11. Establece $M = M + 1$.

APÉNDICE 4. DEFINICIÓN Y CÁLCULO DEL «ÍNDICE DE “EXTENSIÓN”»

Para toda C_i hay un número N_i de índices en los que esta participa y un número M_i en los que no participa. A su vez, como puede observarse en la tabla 6, tanto N_i como M_i son el resultado de la suma del número de índices que permanecen significativos, n_1 o m_1 , y el número de índices que no permanecen significativos tras el procedimiento RRL, n_2 o m_2 . Por otra parte, necesariamente $n_1 + m_1$ es el número de índices que permanecen significativos y $n_2 + m_2$ es el número de índices que no permanecen significativos.

Sea p_i la proporción de índices que permanecen significativos respecto al total de índices en los que C_i participa, es decir, n_1 / N_i ,

y sea q_i la proporción de índices que permanecen significativos respecto al total de índices en los que C_i no participa, m_1 / M_i , entonces el índice de extensión viene dado por la siguiente ecuación:

$$IEC_i = \frac{p_i}{p_i + q_i}$$

TABLA 6. Distribución de los índices de homofilia

	C_i contribuye	C_i no contribuye
Permanecen significativos	n_1	m_1
No permanecen significativos	n_2	m_2
TOTALES	N_i	M_i

Fuente: Elaboración propia.