

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.

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

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.

«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.

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.

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.

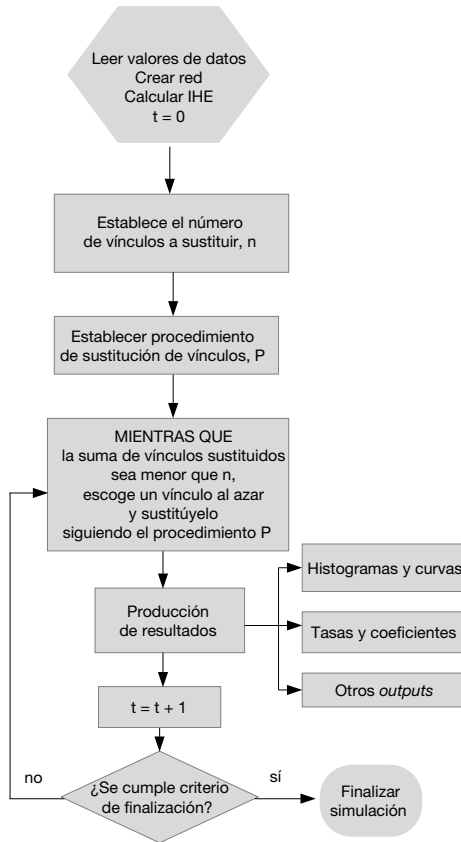
TABLA 1. Características básicas del componente gigante de la red de alumnos de IES San Borondón

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-

critado 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 desciende 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
ln-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(*)	f_i (**)
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 ($p < 0,05$).

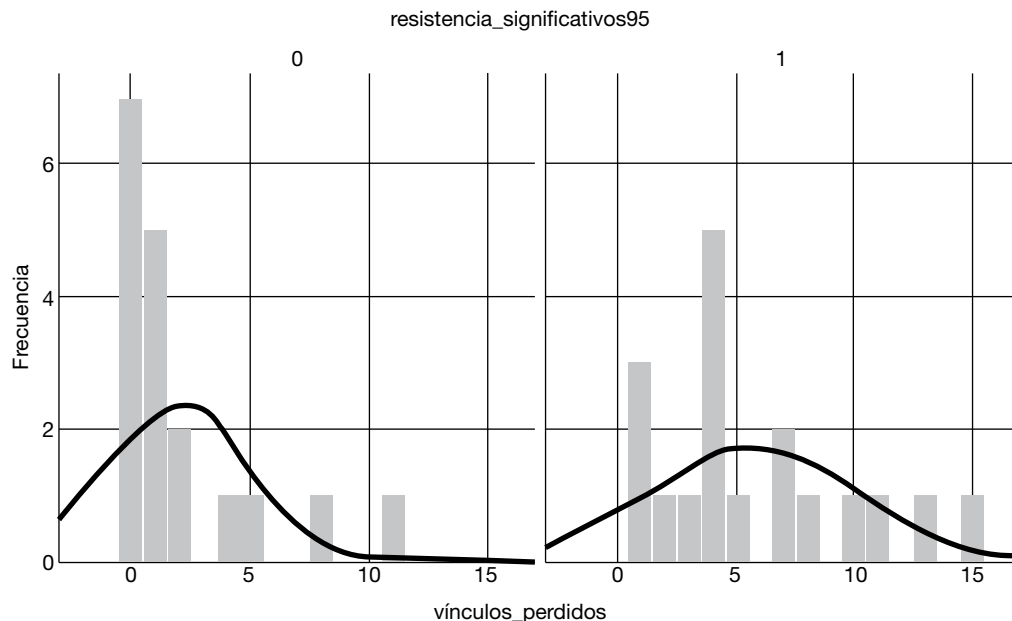
Fuente: Elaboración propia.

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 permanecen

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

GRÁFICO 1. Distribuciones del número de vínculos reasignados



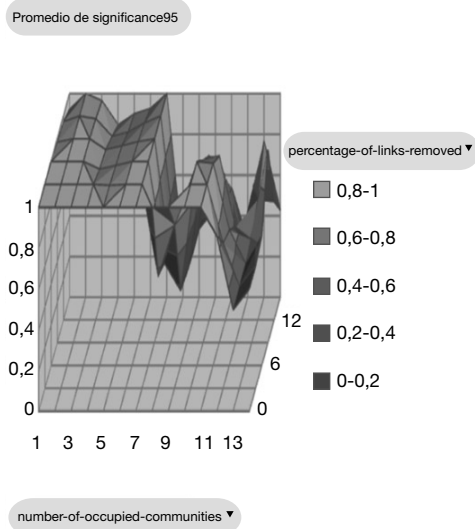
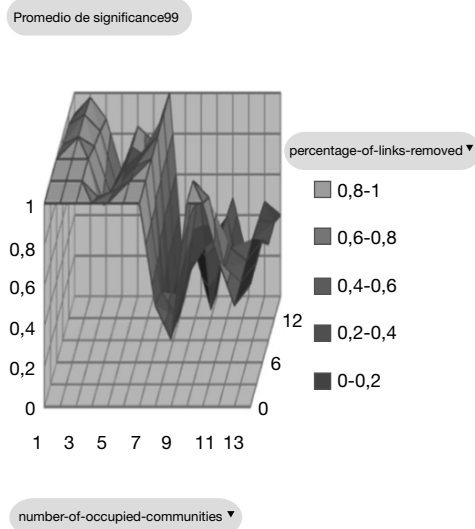
Nota: El panel 0 muestra la distribución del número de vínculos homofílicos perdidos en el caso de los índices que «no» permanecen significativos, mientras que el panel 1 muestra la misma distribución para el caso de los índices que «sí» permanecen significativos.

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.

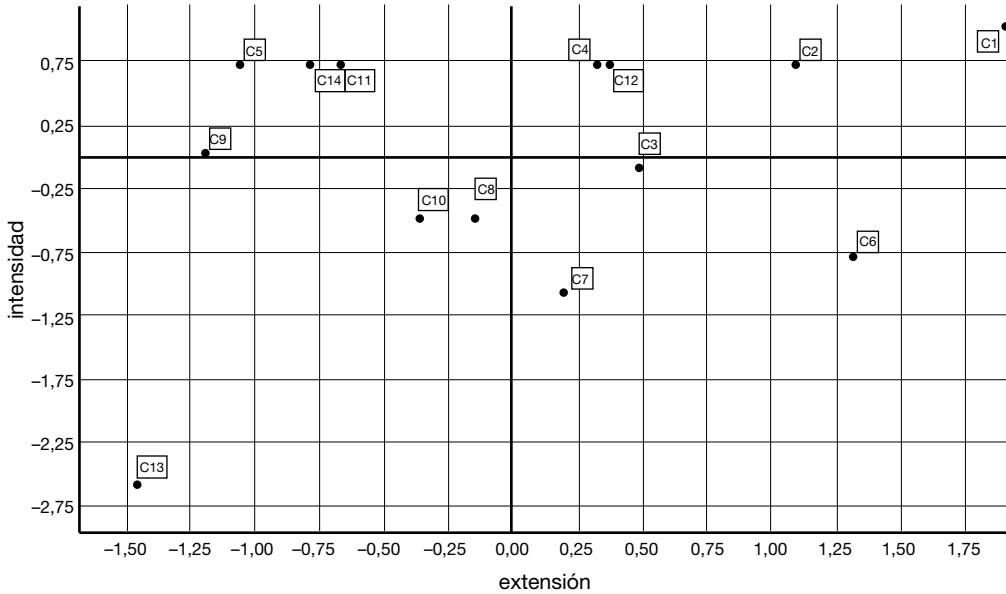
El diagrama 2 muestra la disposición de las catorce comunidades en un espacio cartesiano en el que los ejes de ordenadas y abscisas representan sus puntuaciones tipificadas en cada uno de los índices. Se aprecia nítidamente que cuatro comunida-

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 participan 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

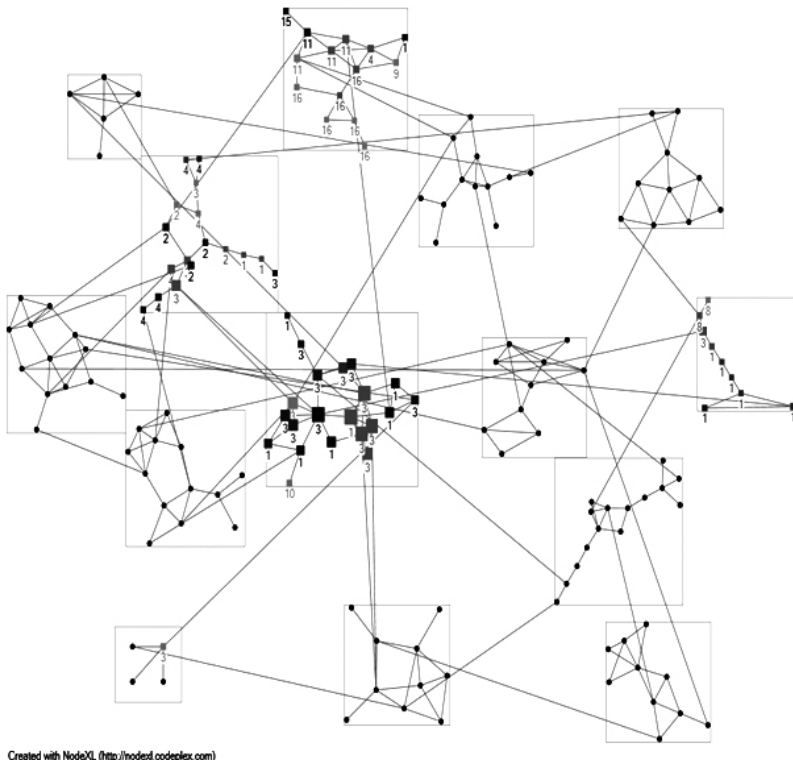
ATRIBUTO	Frecuencia en el núcleo homofílico (N = 59)	Frecuencia en el componente gigante (N = 163)
MUNICIPIO1	20,9%	52,00%
MUNICIPIO2	30,6%	16,00%
GÉNERO HOMBRE	52,0%	56,00%
GÉNERO MUJER	48,0%	44,00%
EDAD 16	30,5%	24,50%
EDAD 17	45,8%	53,30%
EDAD 18	13,6%	14,11%
EDAD 19	6,8%	3,70%
RESIDENCIA (*)	61,3%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

(*) Tiene plaza en la residencia de estudiantes.

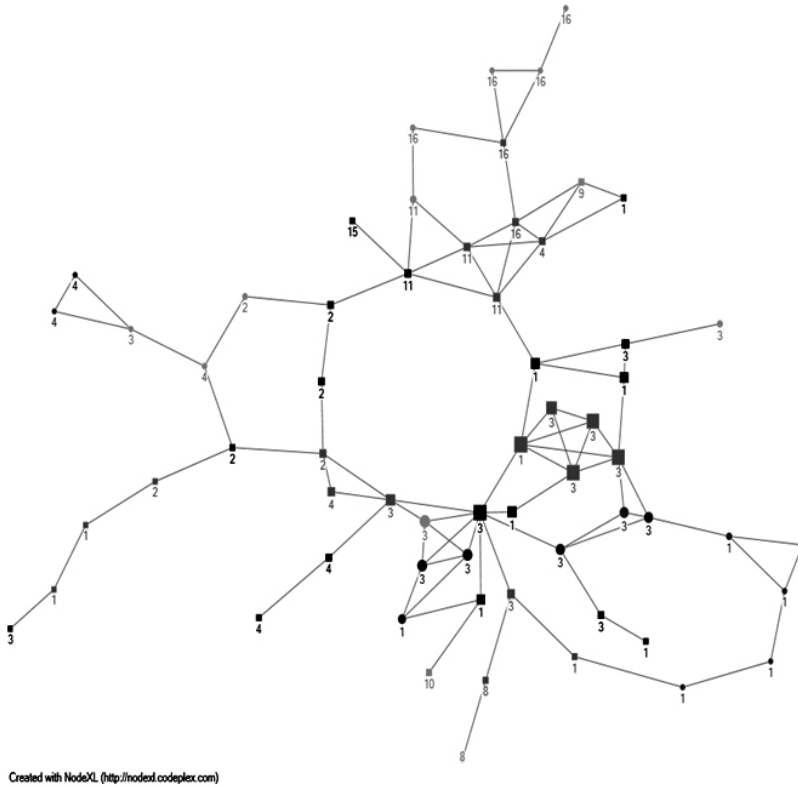
(**) Cursa el bachillerato de ciencia y tecnología.

(***) 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

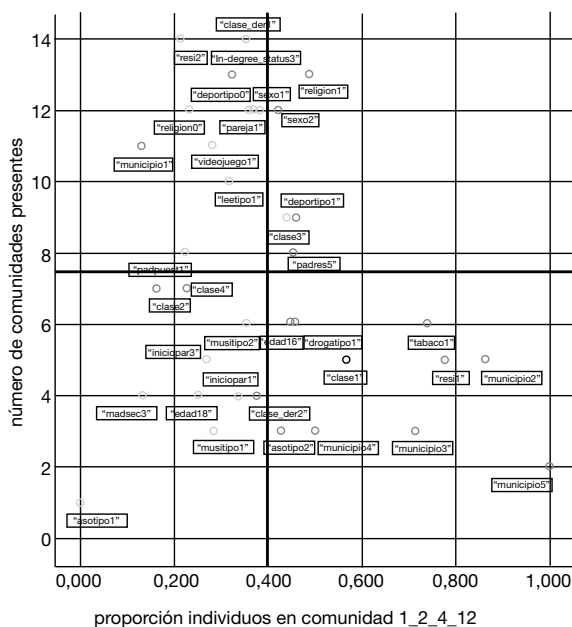
¹³ Debido a la orografía de la isla, el municipio donde se encuentra el IES San Borondón dispone de una residencia para los estudiantes procedentes de otros municipios.

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.

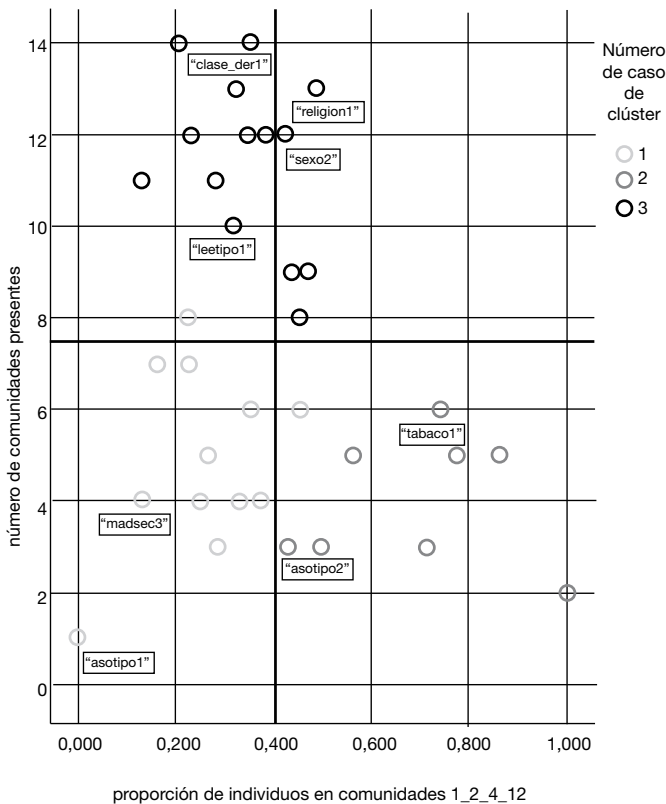
Entre los índices que sí son resistentes encontramos principalmente tres casos: en primer lugar, índices que son resultado de un foco de atracción para los individuos, como puede ser la clase en la que estudian, la residencia de estudiantes o el club de fútbol. En segundo lugar, índices que son el resultado 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), «religion1» (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), «iniciar3» (individuos que tienen pareja desde

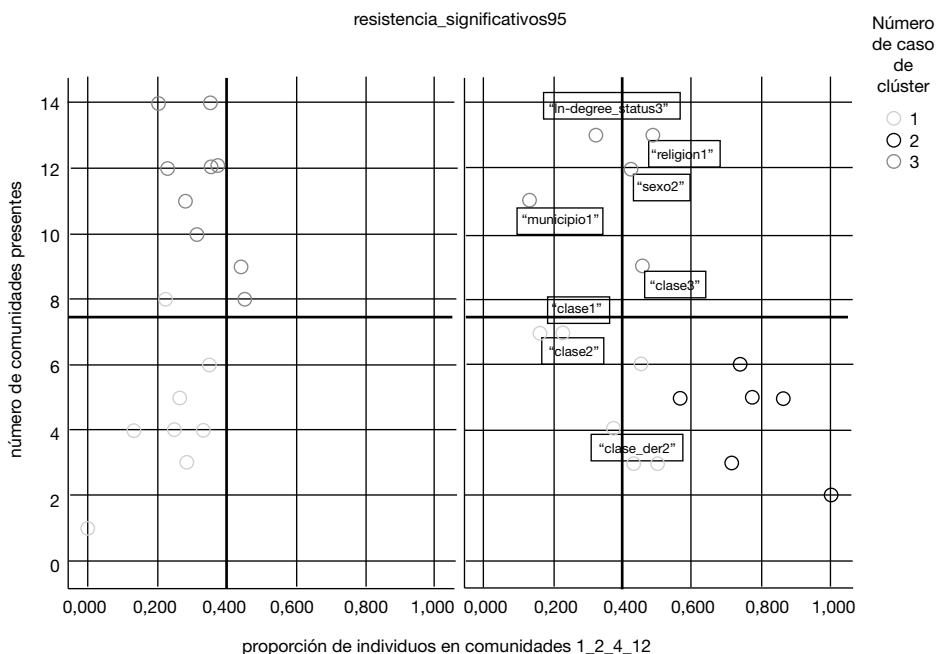
hace más de un año), y «madsec3» (madre trabajadora del sector de la educación y servicios). Aunque la mayoría de los índices de este clúster no resisten el procedimiento RRL, algunos de ellos sí.

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-

da 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 «indegree_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 homófila en un parámetro implica necesariamente asociación heterófila 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 homófila entre individuos que cursan, digamos, primero de ciencias naturales, produce necesariamente asociaciones heterófilas 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 pop	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**
NUMPARS (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 = n° 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, $v_{i?}$, del conjunto de vínculos homofílicos de i respecto al atributo A .
6. Elimina $v_{i?}$.
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.

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

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

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

Key words

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

Palabras clave

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

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 indexes and their statistical significance and then simulated small alterations in the distribution of links. The results clearly show that some homophily indexes 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.

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.

Citation

Linares Martínez, Francisco; Miguel Quesada, Francisco J. and Kohl, Mona (2022). "Resilient Homophily Patterns in Youth Friendship Networks: A Case Study Using a Computer Simulation Experiment". *Revista Española de Investigaciones Sociológicas*, 177: 43-68. (doi: 10.5477/cis/reis.177.43)

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

INTRODUCTION¹

The concept of homophily was coined by Robert K. Merton in his seminal article on² friendship relations co-authored with Paul Lazarsfeld (Lazarsfeld and Merton, 1954). Merton justified the need to introduce a new term into the sociological vocabulary by arguing that there was no word in the English language to concisely refer to friendships between people “of the same kind”. Today, the most common definition is the one provided in a much-cited literature review article: “the principle that a contact between similar people occurs at a higher rate than among dissimilar people” (McPherson, Smith-Lovin and Cook, 2001: 416). As will be seen below, there is some ambiguity or confusion in the use of the term, despite its apparent clarity, since some academics associate it with a person’s preference to hold relationships with similar ones, while others use it to denote an empirical regularity, a pattern of collective behaviour (the frequency of contacts between similar individuals) which may be the result of various social mechanisms. The second meaning is more closely related to Lazarsfeld and Merton’s original use of the term and the most obvious one in McPherson, Smith-Lovin and Cook’s definition. It will be the meaning adopted in this paper.

Simply put, there is homophily regarding an attribute in a network of relationships between individuals if the proportion of links between individuals who have that attribute is higher than the proportion of individuals with that attribute in that population. Thus, for instance, if a link between two Catholic individuals is more likely than the propor-

tion of individuals who profess that religion, there are fewer inter-group links and more intra-group links than would be expected in the absence of social mechanisms inducing the structuring of social relations, which implies non-random behavior. The aim of this paper is to study a very specific problem in the identification of this phenomenon, namely, the extent to which the homophily indexes that can be calculated in a network of individuals are resistant to small alterations in their links, given that some links disappear, and new ones are created in the course of social life.

To address this question, a simulation model was built to replicate data from a network of students in a secondary school (coded as IES San Borondón). Once statistically significant homophily indexes were identified, a “virtual” experiment was carried out in which the real links were recursively replaced by randomly chosen links, up to 15% of the total number of homophilic links. As a result of this procedure, all indexes decreased but some nevertheless remained statistically significant, while others were no longer significant. This immediately raises the question of why some homophily patterns are resilient to small disturbances in network structure while others are not. The general hypothesis is that the explanation for why certain patterns of homophily are resilient and others are not rests on how individuals are distributed across the different communities which make up the network as a whole. To our knowledge, no systematic study of this issue exists in the literature.

From a methodological point of view, tackling the problem requires moving through different conceptual levels or planes of reality (individuals, communities, and network) where information is coded into three different databases.

At the most basic level, the units of analysis are the individuals. A questionnaire was used to collect information on their typical attributes of the units of analysis are the in-

¹ This research has received funding from projects CSO2015-6474-R (MINECO) and PID2019-107589GB-I00 (MICIN).

² The article was published under both names, but was divided into two parts: the substantive part, authored by Merton, and the methodological part, written by Lazarsfeld.

dividuals (e.g., age, gender, religion, etc.). A name generator was used in the same questionnaire to collect relational information.

The network of friendships was reconstructed using the information obtained from the questionnaire. This is a *small world* network, made up of communities or groups³ of individuals who, in turn, have some links that *bridge* with other communities. The second database was made up of the properties of the communities, which are supra-individual units.

Finally, since the analysis computed all possible homophily indexes in the network, a third database was constructed that included the indexes and the data associated with each of them (such as: number of individuals, communities to which they belonged, number of links, etc.). This provided an exhaustive description of the patterns of homophilic behaviour observable in the network as a whole.

The study is structured as follows. The next section contains a theoretical review with special emphasis on the explanatory mechanisms of homophily patterns. Methodological details are then described, including a brief outline of the agent-based simulation model (ABM) built specifically to address the research question; its operation is also illustrated. The results of the analysis of three different types of data are then presented: firstly, homophily indexes, identifying those that remained significant from those that did not; secondly, the groups of individuals (communities) that were most frequently associated with the homophily indexes that remained significant; and thirdly, the characteristics of the homophily indexes that remained resilient. The article ends with a discussion of these results and some conclusions.

³ We use the term group or community interchangeably, leaving the use of the term *cluster* for the statistical technique named *cluster analysis*.

INBREEDING HOMOPHILY MECHANISMS

As McPherson, Smith-Lovin and Cook pointed out in the literature review cited in the introduction, empirical research into the phenomenon became significantly popular from the 1970s onwards, largely stimulated by Peter Blau's theory of social structure and by the development of network analysis. However, it has not always been suitably recognised that the concept had a dual status in Blau's (1977) study, which may have led to some confusion in some scholarly works.

On the one hand, Blau used the concept of homophily as a basic assumption concerning individuals' preferences. If individuals *i* and *j* share a quality, they will be interested in creating a friendship bond if given the opportunity. That is, the bond of friendship is explained by that love (*philia*) of equals (*homo*). In this case, homophily is a mechanism that operates at the individual level and helps to explain certain patterns of social relations. This is, for example, the notion that Shalizi and Thomas (2011) used to examine the problem of the distinction between homophily and social contagion.

On the other hand, the term homophily is also used in Blau's theory to mean the proportion of individuals in a category of a given parameter of the social structure who maintain relationships with individuals in the same category⁴. As Blau showed, this proportion would depend on the relative weight of each category in the population as a whole. If one group is larger than another, all other things being equal, its members will have fewer opportunities for heterophilic relationships than those from the smaller group. Therefore, in this second use, the terms homophily/heter-

⁴ In Peter Blau's theory, social structure was conceived as an intersection of parameters (nominal or graded) that reflect the relationships that individuals have with each other.

ophily are linked to certain population characteristics related to the degree of social cohesion (Lozares and Verd, 2011), not to a type of individual motivation.

The degree of homophily resulting from the opportunities created by the quantitative distribution of the population is termed baseline homophily. But when the links between individuals with a certain trait exceed the proportion of individuals with that trait in the population, i.e., when such relationships occur more frequently than those offered by the opportunity set, this fact is termed inbreeding homophily. In this case, there must be psycho-social mechanisms that make intra-group relations more frequent than expected and inter-group relations less frequent than expected. In this paper, the term homophily is used to mean the tendency (observed in a network of individuals) for contacts between individuals who have a similar trait or characteristic to occur more frequently than with individuals who are dissimilar in terms of that characteristic, irrespective of the particular mechanism causing such in-breeding.

The mechanisms that produce in-breeding homophily, in turn, belong to two types of families. The first is the family of mechanisms based on the structure of relationships that facilitate the maintenance of social contacts. Such opportunities are provided by *social foci* of interaction (Feld, 1981, 1982) such as organisations, and by the social networks created in everyday life. The second family of mechanisms relates primarily to individual decision-making that involves the creation, maintenance, and dissolution of social ties.

This second type of mechanisms involve psychosocial reinforcement processes. As Lazarsfeld and Merton pointed out in their analysis of *value homophily*, two strangers with similar values are likely to form a bond of friendship if they have the opportunity to meet on a regular basis (this is connected with the meaning of the proverb “Birds of

a feather flock together”). It is also possible that two individuals may be motivated to change their values precisely as a result of their friendship, in a give-and-take that progressively smooths out initial discrepancies. This second process falls into the category of ‘social influence’ or *peer influence* (Cohen, 1977). A third possibility is that an individual may acquire one or more traits of another individual with whom they have a bond through some imitation mechanism; in this case influence is not reciprocal, so we call it “social contagion”, although the outcome is difficult to distinguish empirically from the previous case. Finally, the preference for similarity or homophilic preference (which, as noted, is often confused with the concept of homophily itself) is possibly the mechanism that operates most frequently in cases of *status homophily*, regarding variables such as gender, ethnicity, educational status, age, etc.

Another question related to the two families of mechanisms briefly presented is which of them is more prominent in explaining the empirically observed patterns of homophily. While no definitive answer can be given to this question in the current state of the art, some evidence suggests that an important part of the explanation lies in structural elements such as the existence of social environments, organisations, etc. that attract individuals with similar characteristics (McPherson and Smith-Lovin, 1987; Moody, 2001; Kossinets and Watts, 2009). Compared to these structural elements which shape individuals’ opportunities to find similar others, psychosocial mechanisms operating at the individual level seem to play a relatively minor role in explaining patterns of relationships, although evidence is inconclusive.

METHODOLOGY

The paper is based on data from a previous study (Linares and Kohl, 2017) in

which a questionnaire was distributed to 194 post-compulsory secondary school students in a secondary school in the Canary Islands Region, which we called IES San Borondón⁵. The survey contained three modules: 1) socio-demographic characteristics, 2) leisure and consumption habits, 3) friendship relationships. In this third module, the students were asked to provide the names of up to four “friends with whom you talk about your problems” (Marsden, 1987). They were also asked to give the name of the person with whom they had a romantic relationship, if any. The result is a network with a giant component that included 163 nodes. The characteristics of this network are shown in Table 1.

An initial analysis showed varying degrees of homophily⁶ consistent with the findings in the literature (Kandel, 1978; McPherson, Miller and Smith-Lovin, 1986, 1987; Moody, 2001; Shrum, Cheek and MacD. Hunter, 1988), considering variables such as gender, age, smoking/non-smoking, religion, etc. although not all indexes were statistically significant⁷. The value

of the 86 calculated homophily indexes and their level of statistical significance are shown in Appendix 1. However, as noted in the introduction, the focus of this paper was not on the statistical significance *per se*, but on the resilience of significant homophily indexes in light of the small disturbances that inevitably occur in a social network due to the dynamics of link creation and dissolution.

To address this issue, an agent-based simulation model (ABM) was built using the Netlogo platform (Wilenski and Rand, 2015)⁸. This model, described in simple terms in Diagram 1, contained a set of modules that sequentially performed the following operations:

1. Importing the survey data of the students from IES San Borondón. This allowed for the *in silico* replication of the real subjects, who were “transformed” into virtual agents with the same attributes and relationships as those reported in the survey.
2. Calculating the Coleman homophily index (CHI) for each of the agents’ attributes and its statistical significance.
3. Random rewiring of links (RRL) for each of the agents’ attributes, followed by calculating of the new CHI and its statistical significance.

⁵ Fieldwork was conducted between 25 February 2015 and 1 March 2015. The questionnaire was distributed in the classrooms to all students present, who accounted for 67% of the total study population.

⁶ There are several options for measuring the degree of homophily (Bojanowski and Corten, 2014). We use the Coleman Homophily Index (CHI) (Coleman, 1957), which compares the proportion of actually existing homophilic relationships with what would be expected if the relationships between individuals were established randomly. The CHI ranges between +1 and -1, with the value 0 representing the *baseline homophily* level, which relates exclusively to the relative size of the subpopulations, and therefore means that there is no social mechanism inducing the choices of individuals.

⁷ Statistical significance was measured using a test specifically designed for the CHI by Signori and O’Shea (1965), who addressed the problem of finding the probability that a given value of the homophily index, or a larger value, can be obtained assuming that there is no relationship between the attributes of the node from which the link departs (the individual mentioned) and the node which receives it (the individual mentioned);

finding the parameters μ and σ which allow the value to be standardised according to a normal distribution. In what follows, the analysis focuses exclusively on the statistically significant indexes.

⁸ Appendix 2 provides a brief description of this type of model.

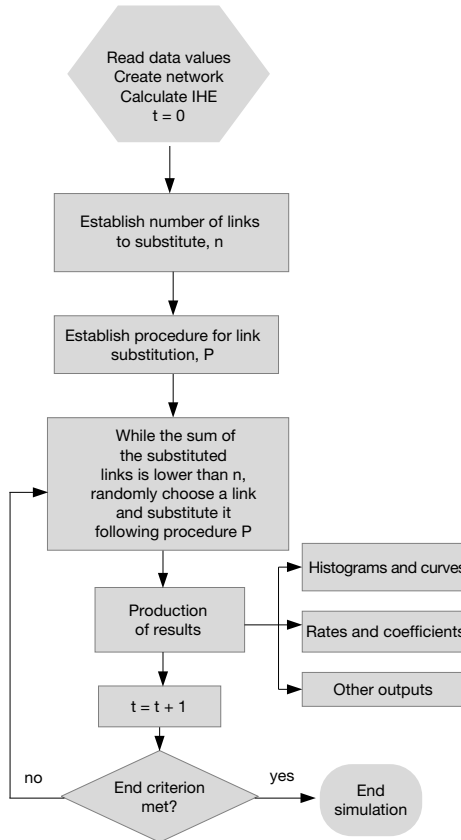
TABLE 1. Basic characteristics of the giant component of the student network from IES San Borondón

No. of nodes	No. of links	Average number of links	Maximum number of links	Density	Average geodesic distance	Max geodesic distance	No. of communities (*)
163	275	3.35	9	0.021	5.92	14	14

(*) Number of communities found using the Clauset-Newman-Moore algorithm.

Source: Own elaboration.

DIAGRAM 1. Model flowchart



Source: Own elaboration.

The RRL procedure⁹ involved the computer randomly choosing one of the homophilic links and replacing it with a new link; it then immediately recalculated the homophily in-

dex and reassessed its statistical significance. This operation was repeated until a number equivalent to 15% of the homophilic links in the selected category were replaced. By substituting one link for another in each case, the basic properties of the network, such as density, mean geodesic distance,

⁹ See Appendix 3 for a detailed description of the RRL procedure.

etc., remained virtually identical, since these measures were insensitive to these small disturbances; however, some of the homophily indexes were no longer statistically significant.

This procedure was repeated using different algorithms for link substitution, so 780 cases of artificially manipulated networks were available for each of the attributes under analysis at the end of the simulation. Restricting the analysis to the 36 attributes (shown in tables 3a and 3b, in the next section) whose homophily indexes were positive and statistically significant prior to any manipulation gave a final population of 29,640 cases.

As an example, Tables 2a and 2b show the results of the simulations for the categories “smoker” and “male”, respectively. All numbers in the table show the averages of the 780 simulations. The first column shows the percentage of homophilic links replaced (%HLR) related to the total homophilic links of that category. For each of its values (0, 3, 6, 9, 12, 15), the rest

of the columns indicate the percentage of homophilic links replaced “related to the total links on the network” (% “total links rewired”, TLR), the value of the Coleman homophily index (CHI), the difference between the original value of the CHI and the new value (“drop value”, DV), the relative frequency with which the index remained significant with probabilities of 95 % and 99% and, finally, the number of communities in which individuals with the quality of “man” or “smoker” were distributed, a number that remained constant in the simulations, since the model did not recompute the communities in the network.

The information contained in Table 2a shows that smokers were concentrated in six communities and that their initial homophily index had a statistically significant value of 0.329. As the simulation model randomly replaced links, this value dropped to 0.222 once 15% of the homophilic links (which constituted 1.97% of the total links in the network) had been replaced, a value that always remained significant.

TABLE 2A. *Simulation results for the “smoker” category*

%HLR	%TLR	CHI	DV	P95	P99	Cs
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

Notes: %RHL = percentage of homophilic links replaced out of the total number of homophilic links in that category; %TLR = percentage of homophilic links replaced out of the total number of links in the network; CHI = Coleman homophily index; DV = “drop value” or difference between the original CHI value and the new value; P95 = frequency with which the index remains significant (p-value < 0.05); P99 = frequency with which the index remains significant (p-value < 0.01); Cs= number of communities in which individuals with the “smoker” quality were distributed.

Source: Own elaboration.

In the case of Table 2b, for the “male” category, the individuals were distributed

into twelve groups and the value of the initial homophily index was 0.331, similar to

that of smokers and equally significant. However, once the random replacement process started, the value dropped to 0.076

and the percentage of times it remained significant at the 0.05 and 0.01 levels was 0.03% and 0.00%, respectively.

TABLE 2B. Simulation results for the "male" category

%HLR	%TLR	CHI	DV	P95	P99	Cs
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

Notes: %RHL = percentage of homophilic links replaced out of the total number of homophilic links in that category; %TLR = percentage of substituted homophilic links replaced out of the total number of links in the network; CHI = Coleman homophily index; DV = "drop value" or difference between the original CHI value and the new value; P95 = frequency with which the index remains significant (p-value < 0.05); P99 = frequency with which the index remains significant (p-value < 0.01); Cs= number of communities in which individuals with the "male" quality were distributed.

Source: Own elaboration.

RESULTS

Relevance of the numbers of communities in which individuals were distributed

Tables 3a and 3b show the averages of the 780 homophily index values calculated for each attribute after the RRL process had been completed.¹⁰ Table 3a shows those attributes for which the index value remained significant in all simulation rounds (i.e., the probability of the index remaining significant was equal to 1), which therefore corresponded to resilient homophily patterns. Table 3b shows those attributes where the probability of the index remaining significant was less than 1. In this table, a wide variety of possibilities can be seen, ranging from indexes that never remained significant

("Baccalaureate student" (class_der1), "father employed in the public administration" (fathempl1) and "has a partner" (partner1)), to indexes that were very likely to remain significant ("belongs to the NSG music association" (assoctype1), "started a relationship less than five months ago" (startrel1) and "mother working in the education and social services sector" (mothsec3).

The first step in the analysis was to rule out the possibility that the differences between the indexes in Table 3a and those in Table 3b was simply due to the absolute number of homophilic links in each case. Thus, since the random replacement of links necessarily decreases the value of the CHI (due to the fact that new links have a very high probability of not being homophilic links), it could be the case that the number of replaced links and the probability of the CHI remaining significant were negatively associated.

¹⁰ For reasons of space, the variable codes appear in the tables. Their labels can be found in Appendix 1.

TABLE 3A. *Resistant homophily indexes*

Category	CHI(*)
assotype2	0.238
class_der2	0.467
class1	0.278
class2	0.059
class3	0.364
class4	0.313
drugtype1	0.158
age16	0.282
In-degree_status3	0.256
municipality1	0.208
municipality2	0.271
municipality3	0.209
municipality4	0.207
municipality5	0.195
religion1	0.145
gender2	0.202
tobacco1	0.222
dorm1	0.546

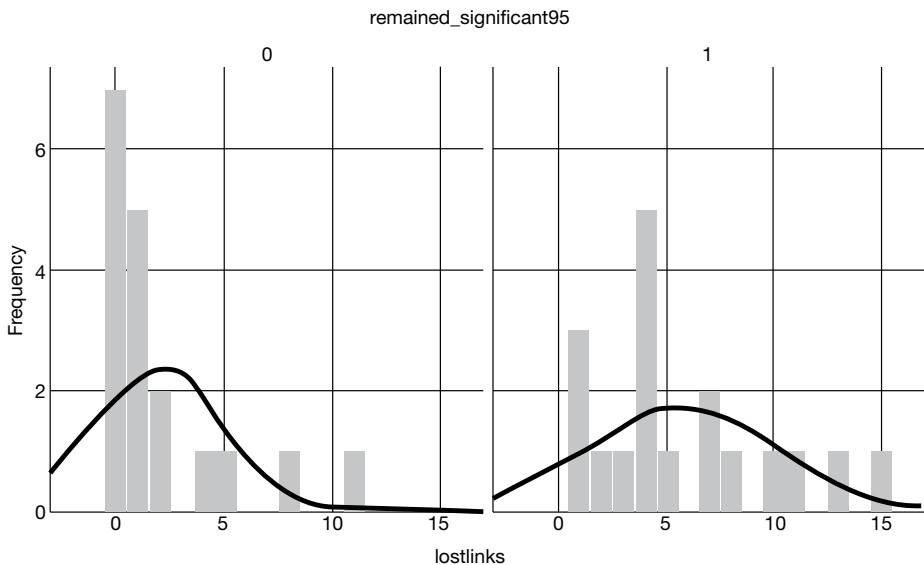
(*) Average CHI after RRL procedure.
 Source: Own elaboration.

TABLE 3B. *Resistant homophily indexes*

Category	CHI(*)	f_i (**)
assotype1	0.123	0.92
class_der1	-0.010	0.00
sporttype0	-0.050	0.01
sporttype1	0.022	0.01
age18	0.067	0.57
startrel1	0.118	0.91
startrel3	0.069	0.40
readtype1	0.081	0.44
mothsec3	0.127	0.91
musictype1	0.124	0.91
musictype2	0.090	0.72
fathempl1	0.010	0.00
fathsec5	0.066	0.46
rel1	0.021	0.00
religion0	0.090	0.48
gender1	0.076	0.03
Videogames1	0.066	0.08
dorm2	0.315	0.84

(*) Average CHI after RRL procedure.
 (**) Relative frequency of statistically significant index ($p < 0.05$).
 Source: Own elaboration.

FIGURE 1. *Distribution of the number of reallocated links*



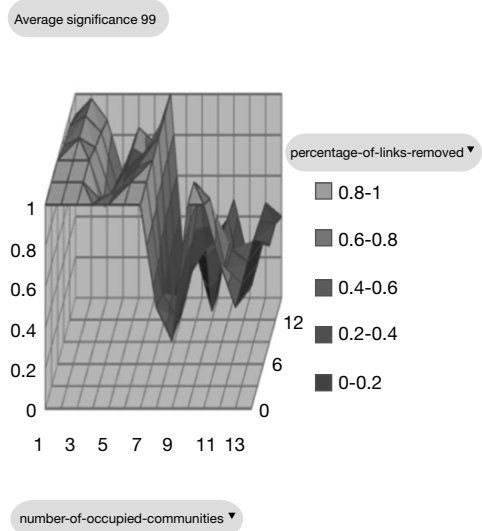
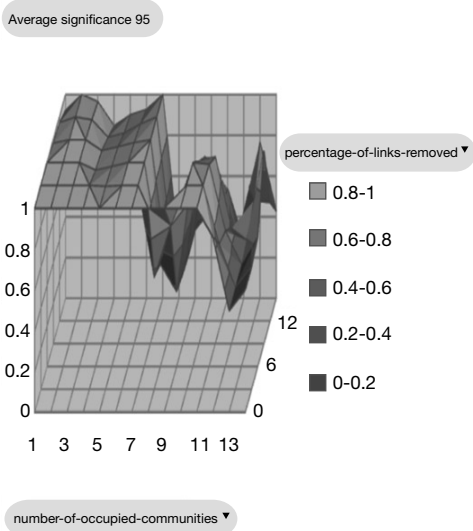
Note: Panel 0 shows the distribution of the number of missing homophilic links in the case of the indexes where “no” remained significant, while panel 1 shows the same distribution for the case of the indexes where “yes” remained significant.
 Source: Own elaboration.

As can be seen in Figure 1, this was not the case. The right panel (1) shows the histogram of the missing links for the indexes that remained significant after the RRL procedure. It can easily be seen that about half of the indexes withstood the

loss of more than six links, up to a maximum of fifteen. The left-hand panel (0) shows that most of the indexes that did not remain significant did not withstand the loss of more than a few links, or even a single link.

FIGURE 2A. Probability that the CHI would remain significant ($p < 0.05$)

FIGURE 2B. Probability that the CHI would remain significant ($p < 0.01$)



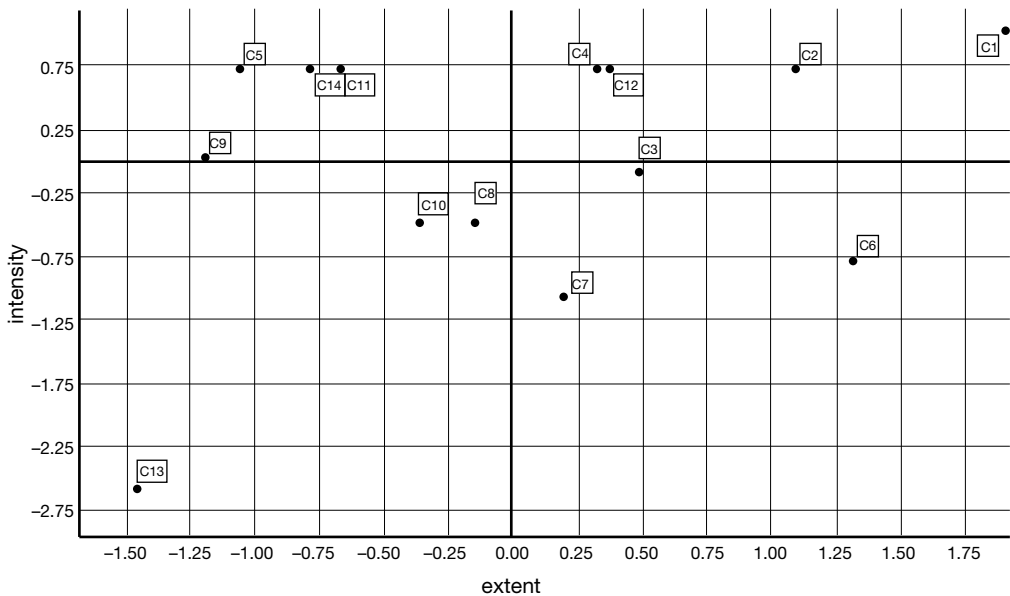
Note: Changes in the probability that the index would remain significant (y-axis) according to the number of communities to which individuals belong (x-axis) and the percentage of homophilic links replaced (z-axis).

Source: Own elaboration.

As shown in Figures 2a and 2b, there was a clear relationship between the probability that the CHI would remain significant (y-axis values) with respect to the number of communities¹¹ among which individuals with the corresponding attribute were distributed (x-axis values), relatively

independent of the percentage of links replaced using the RRL procedure (z-axis values). Thus, the probability that the CHI would remain significant decreased very significantly if the number of communities was greater than seven. For example, “men” were distributed into twelve groups and “smokers” into six groups. In the second case, the homophily index was resilient to the RRL procedure, while in the first case the probability of remaining significant started to drop significantly after 9% of the links were replaced.

¹¹ A community is a subset of the population of individuals in which the density of relationships is higher than in the network as a whole. Community identification algorithms assign each individual in the network to a single group.

DIAGRAM 2. Dispersion of communities by “extent” and “intensity”

Source: Own elaboration.

Identification of the most frequent communities in the resilient homophily indexes

The evidence shown in the previous section raises the question of whether all communities will be equally important in the “production” of resilient indexes. Two new indexes that we called “extent” and “intensity” were constructed with the purpose of measuring the degree to which each of the fourteen communities identified by the Clauset-Newman-Moore algorithm contributed to resilient or non-resilient CHIs. The term “contribution” relative to a given community C_i was used to denote that at least a fraction of the individuals displaying the attribute whose CHI was being calculated belonged to that community. These indexes were defined as follows:

- The “intensity index” of a community, IIC_i , measured the probability that the CHI indexes to which C_i contributed would remain significant.

- The “extent index” of a community, EIC_i , measured the proportion of CHI indexes that remained significant where C_i contributed, relative to the proportion of CHI indexes that remained significant where C_i did not contribute¹².

Diagram 2 shows the arrangement of the fourteen communities in a Cartesian space where the horizontal and vertical axes represent their scores for IIC_i and EIC_i . Four communities (numbers 1, 2, 4 and 12) scored above average in both dimensions, deviating by approximately 0.75 standard deviations from the mean value for the intensity index, and by 0.25 to 1.80 standard deviations from the mean value of the extent index. Moreover, as can be seen in Table 4, the scores of both indexes were strongly correlated with some community characteristics: there was a strong negative correlation of both indexes

¹² A detailed account of the construction of the extent index can be found in Appendix 3.

with link density (the ratio of existing links to possible links) and a positive correlation with both the average distance and the maximum geodesic distance between two nodes.

These correlations suggest that certain topological features of the communities could make it easier for the indexes in which they participate to remain significant.

TABLE 4. *Correlations of extent and intensity indexes with the characteristics of the communities*

	Link density	Average geodetic distance	Maximum geodetic distance
EIC_i	-0.772	0.771	0.668
IIC_i	-0.614	0.592	0.573

Source: Own elaboration.

TABLE 5. *Some salient attributes of the homophilic core*

ATTRIBUTE	Frequency in the homophilic core (N = 59)	Frequency in the giant component (N = 163)
MUNICIPALITY1	20.9%	52.00%
MUNICIPALITY2	30.6%	16.00%
GENDER MALE	52.0%	56.00%
GENDER FEMALE	48.0%	44.00%
AGE 16	30.5%	24.50%
AGE 17	45.8%	53.30%
AGE 18	13.6%	14.11%
AGE 19	6.8%	3.70%
DORMITORY (*)	61.3%	31.00%
B. SCIENCE (**)	36.0%	41.00%
MODULES (***)	24.0%	17.00%

(*) Had a place in the student dormitory.

(**) Studying the Science and Technology strand of the Baccalaureate.

(***) Studying a Vocational Training Module.

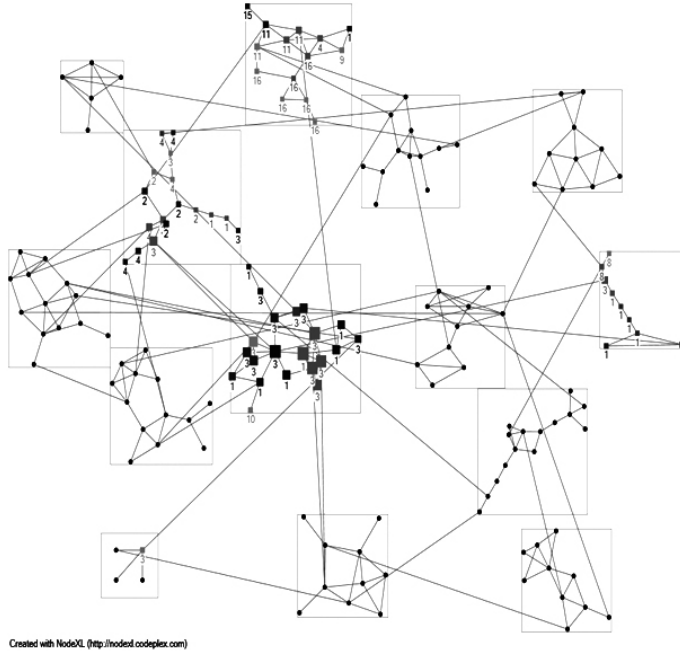
Source: Own elaboration.

Table 5 shows some differential characteristics of the individuals that formed part of the communities indicated, which were called the “homophilic core”, the disposition of which in the giant component of the network is shown in Figure 1a (Figure 1b shows exclusively the nodes belonging to the homophilic core). Of particular importance was the higher proportion of students who had a place in the school dormitory, correlated with a lower propor-

tion of individuals coming from municipality 1 (where IES San Borondón¹³ is located), suggesting that part of the resilient homophily patterns were linked to the relational web generated by the cohabitation of young people in the student dormitory.

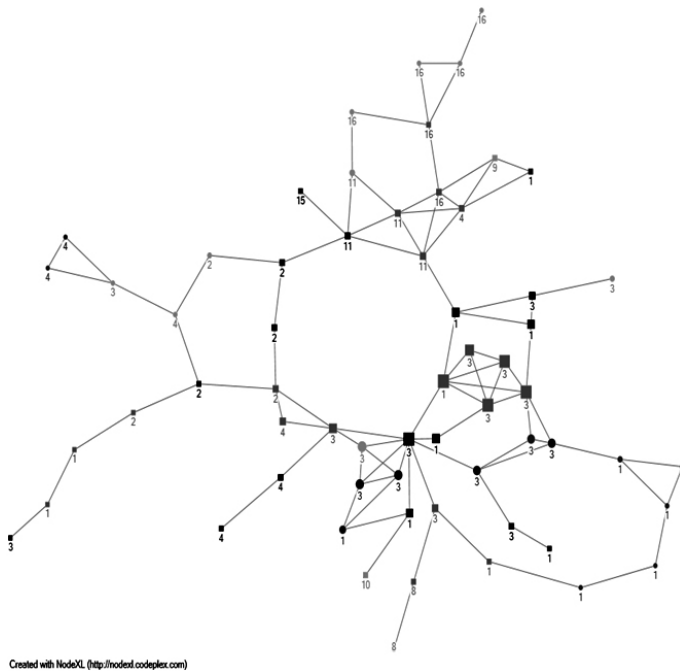
¹³ Due to the orographic features of the island, the municipality where IES San Borondón is located had a dormitory for students coming from other municipalities.

NETWORK 1A. *Giant component of IES San Borondón*



Source: Own elaboration.

NETWORK 1B. *Homophilic core*



Source: Own elaboration.

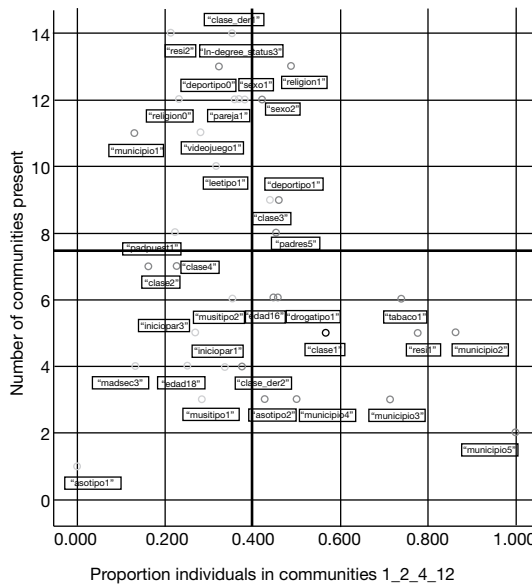
Resilience of homophily patterns

The previous sections have shown two findings: (1) when individuals with the attribute under study were concentrated in seven or fewer communities, homophily indexes were more likely to remain statistically significant after the RRL process, and (2) communities 1, 2, 4 and 12 participated in a higher proportion in indexes that remained significant and, at the same time, their participation increased the probability of an index remaining significant.

Figure 3 shows the arrangement of homophily indexes in a Cartesian space defined by the variable “proportion of individuals in communities 1, 2, 4 and 12” (x-axis)

and the variable “number of communities present” (y-axis). For example, the point “municipality5” at the bottom right represents the homophily index among students coming from that municipality of San Borondón; these students were distributed into two groups and 100% of them belonged to communities 1, 2, 4 and 12. The point “dorm2” in the upper left corner represents the homophily index for the individuals who did not have a place in the student dormitory, who were distributed into fourteen groups, and only 20% of them belonged to communities 1, 2, 4 and 12. The lines represent the means of the variables, and most of the indexes that remained significant appear in the lower-right quadrant, as expected.

FIGURE 3. *Distribution of homophily indexes*



Source: Own elaboration.

The resilient indexes can be classified in three main categories. Firstly, indexes that resulted from a focal point for individuals, such as grade’s classroom of which they are part, the student dormitory, or the football club. Secondly, indexes that were the

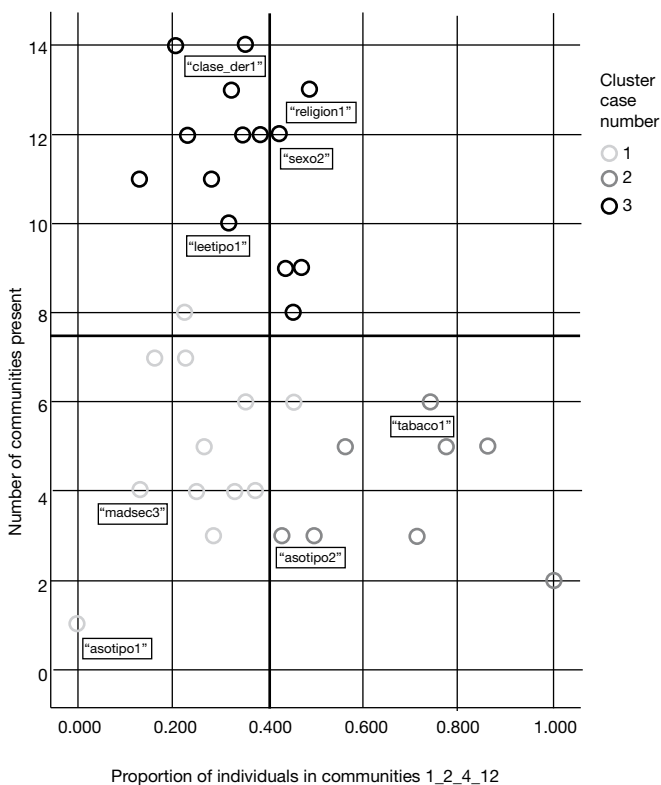
result of pre-existing relationships (i.e., all those linked to the municipalities of origin), which clearly illustrated the concept of status homophily. Thirdly, indexes reflecting behaviour susceptible to social contagion, such as tobacco and marijuana use, illus-

trated the concept of value homophily. The full list can be found in Table 3a above.

Figure 4 shows the arrangement of the homophily indexes in the same Cartesian space as above. In this case, the indexes are classified into three clusters identified by the k-means algorithm. It is interesting to note that the algorithm has classified in the same cluster #3 most of the homophily indexes that corresponded to features that were widely distributed in the network (eight or more com-

munities), and that, precisely because of this, individuals belonging to communities 1, 2, 4 and 12 represented a smaller percentage of the total. Examples of these indexes are “class_der1” (individuals who doing their baccalaureate), “religion1” (individuals who define themselves as Catholics), “video game1” (video game users), “readtype1” (readers) or “gender2” (female individuals). The vast majority of the elements in this cluster #3 did not withstand the RRL procedure.

FIGURE 4. Clustering of homophily indexes (k-means algorithm)



Source: Own elaboration.

Cluster #1 consisted of a heterogeneous set of indexes that corresponded to characteristics that were not widely distributed among the population, in which the percentage of individuals belonging to communi-

ties 1, 2, 4 and 12 were relatively low. Examples are: “musictype2” (individuals who like pop music), “startrel3” (individuals who have been in a relationship for more than a year), and “mothsec3” (mothers working in the ed-

ucation and service sector). Although most of the indexes in this cluster did not withstand the RRL procedure, some of them did.

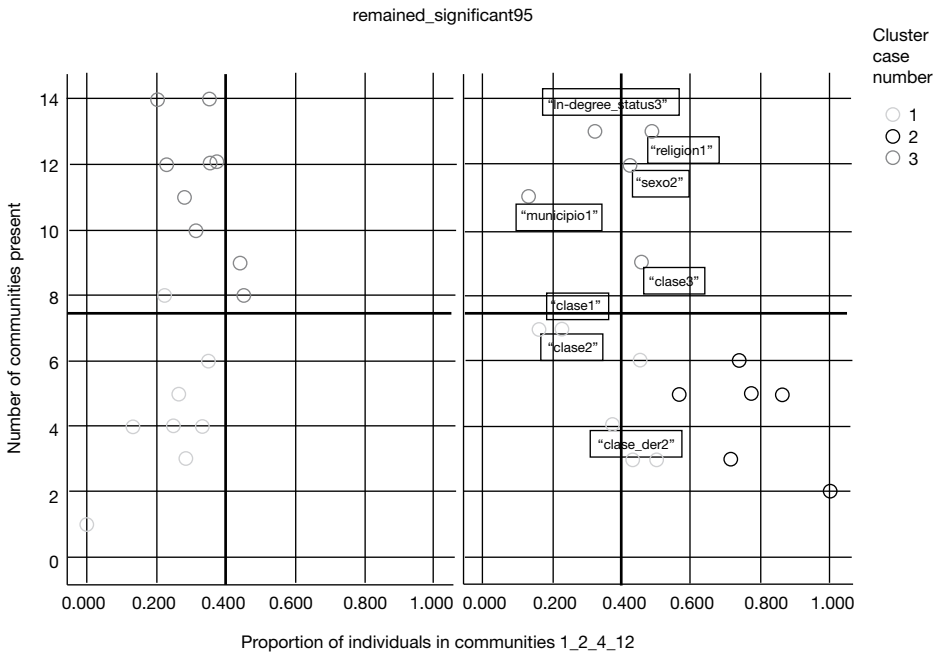
Finally, all elements of cluster #2 (characteristics not widely distributed but with a high degree of presence of communities 1, 2, 4 and 12) corresponded to indexes that remained significant, such as “tobacco1” (individuals who smoked), “assoctipe2” (individuals in the football club), “class1” (individuals in the first year of the science and technology baccalaureate) or “dorm” (individuals who lived in student accommodation).

To complete the description, Figure 5 shows the information in Figure 4 divided into two panels: the one on the right (1) represents

the indexes that withstood the RRL procedure and the one on the left (0) those that did not. This figure clearly shows the “deviant” cases, i.e., the indexes which resisted the RRL procedure, despite not being in the lower-right quadrant.

Looking at the right-hand panel, three cases from cluster #1 are in the lower-left quadrant: “class2” (students in the 2nd year science and technology baccalaureate), “class4” (students in the 2nd year social sciences and humanities baccalaureate) and “class_der2” (students taking vocational training modules); these are part of the group of resilient homophily indexes related to a given focal point.

FIGURE 5. “Deviant” cases



Source: Own elaboration.

We can also see five indexes classified in cluster #3 in the top half. Of these, “municipality1” and “class3” (individuals in the first year of the social sciences and humani-

ties baccalaureate) belonged to the above-mentioned types of resilient indexes too. The presence in this panel of “in-degree_status3” (individuals mentioned three times

or more) was a result consistent with similar findings in the literature (Maggio and Gari, 2012) regarding the opportunities provided by the structure of social networks: people with more relationships also relate more to each other. Finally, the resilience of “religion1” (Catholic religion) certainly seems to be an anomalous case; as was “gender2” (female individual), whose characteristics were very similar to those of “gender1” (male individual), which, in contrast, was not a resilient index.

DISCUSSION AND CONCLUSIONS

This paper has addressed the issue of what makes a pattern of homophily robust, so it can therefore be considered a stable characteristic of a network of individuals. A robust or resilient pattern has been understood to be one that produces a homophily index that remains statistically significant even when the network of relationships undergoes a number of disturbances which, for practical purposes, has been set at the random replacement of 15% of the homophilic links.

The evidence suggests that the distribution of individuals across communities in the network is a key factor, albeit in a counter-intuitive sense: there appears to be an inverse relationship between the number of communities in which individuals are distributed and the likelihood that the homophily index remains significant after applying the RRL procedure. Thus, most of the characteristics widely distributed in the population did not give rise to resilient homophily indexes (with some notable exceptions, such as the “female” attribute). This inverse relationship appears to be due to the fact that a few communities (namely, numbers 1, 2, 4 and 12 in our case study) contributed disproportionately to the production of resilient homophily

patterns. Paradoxically, this involves that the indexes that remained significant did not provide information on the network as a whole, but on some of the communities within it; and, consequently, the statement “there was homophily in the whole of the network N regarding attribute A” is a generalisation that is not true without further qualification.

The second result is that the most important communities in the production of resilient homophily patterns seem to have specific topological features, such as greater spread and lower density than the average. However, given the small number of communities in the network analysed, this result requires replication of the study with other networks to be confirmed or discarded.

With regard to the mechanisms explaining homophilic patterns, the results are consistent with studies that have underlined the importance of focal points of activity and the opportunity structure of networks, given that most of the resilient indexes corresponded to these cases. A significant number (but not all) of characteristics that can be labelled as status homophily, as well as some characteristics susceptible to social contagion, also showed resilient homophily indexes.

However, most homophily indexes for attributes associated with students’ parental status were not robust; a result that is actually consistent with Peter M. Blau’s conceptual framework, since when parameters are intersecting, a homophilic association in one parameter necessarily implies a heterophilic association in another. Thus, for example, for the homophily index of students whose parents worked in the service sector to be resilient, the school-year-based association would have to be non-resilient. In other words, homophilic association between individuals in, say, the first year of natural sciences necessar-

ily produced heterophilic associations between individuals with parents of a different occupational status.

The robustness of the results of this study obviously depends on the possibility of replicating them with other populations of individuals and with other algorithms for both community search and link replacement. However, the idea that a pattern of behaviour will be resilient if it is concentrated in certain communities in the network has a theoretical significance that transcends the particular case we have analysed. It is reasonable to suggest that this may be a generalisable result for the 'small world' class of networks. If this were the case, the analysis procedure designed for this research would be useful for a wide number of empirical domains, some belonging to other disciplines such as economics, anthropology or ecology, where the objects of analysis could be expected to be embedded in *small world* networks.

BIBLIOGRAPHY

- Aral, Sinan; Muchnik, Lev and 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 and 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 and 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 and Duncan, Watts (2009). "Origins of Homophily in an Evolving Social Network". *American Journal of Sociology*, 115(2): 405-50.
- Lazarsfeld, Paul F. and Merton, Robert K. (1954). "Friendship as a Social Process: A Substantive and Methodological Analysis". In: 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 and 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 and 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 and Smith-Lovin, Lynn (1986). "Sex Segregation in Voluntary Associations". *American Sociological Review*, 51: 61-79.
- McPherson, Miller and 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 and 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. and 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. and MacD. Hunter, Sandra (1988). "Friendship in School: Gender and Racial Homophily". *Sociology of Education*, 61: 227-239.
- Signorile, Vito and O'Shea, Robert M. (1965). "A Test of Significance for the Homophily Index". *American Journal of Sociology*, 70(4): 467-470.
- Wilensky, Uri and Rand, William (2015). *An Introduction to Agent-based Modeling*. Cambridge, Massachusetts: The MIT Press.

RECEPTION: May 27, 2020

REVIEW: November 12, 2020

ACCEPTANCE: December 23, 2020

APPENDIX 1. LIST OF VARIABLES AND CATEGORIES

Variable	Category	Description	N	CHI
ASSOCTYPE (membership of an association)	0	Does not belong to an association	117	0.128
	1	NSG Music Association	9	0.170**
	2	Unión Deportiva G	11	0.305**
CLASS_DER (secondary education option)	1	Taking the baccalaureate	133	0.608**
	2	Taking vocational training modules	21	0.612**
CLASS (year and type of baccalaureate)	1	1st year Science and Technology baccalaureate	30	0.401**
	2	2nd year Science and Technology baccalaureate TC	38	0.676**
	3	1st year Social Sciences and Humanities baccalaureate	39	0.540**
	4	2nd year Social Sciences and Humanities baccalaureate	26	0.438**
SPORTTYPE (Does sport)	0	Does not do sport	69	0.153**
	1	Plays football	37	0.164**
	2	Goes to the gym	14	0.192**
DRUGTYPE (drug use)	0	No drug use	131	0.090
	1	Takes drugs	29	0.240**
AGE	16	16 years old	40	0.419**
	17	17 years old	87	0.234**
	18	18 years old	23	0.127**
FREQALC (alcohol use)	0	No alcohol use	33	0.089
	1	Drinks alcohol only at parties	122	0.398**
FREQDRUG (frequency of drug use)	0	Never	131	0.090
	1	Only uses drugs at parties	18	0.169**
	2	Drug use also at other times	12	0.195**
FREQSMOK (frequency of tobacco use)	0	Non-smoker	133	0.264**
	1	Only smokes at parties	18	0.057
	2	Also smokes at other times	11	0.414**
GENDER	1	Male	91	0.340**
	2	Female	72	0.544**
WEEKENDCURF (curfew set by parents)	1	Weekend curfew	29	-0.021
	2	No weekend curfew	130	0.113
STARTSMOK (curfew set by parents) (started smoking)	0	Non-smoker	133	0.264**
	1	Started before the age of 15	11	0.212**
	2	Started at age 15 or older	11	0.016

APPENDIX 1. LIST OF VARIABLES AND CATEGORIES (CONTINUATION)

Variable	Category	Description	N	CHI
STARTREL (start of the current relationship)	1	Started a relationship less than 5 months ago	17	0.203**
	2	Started a relationship between 5 months and a year ago	18	0.042
	3	Started a relationship over a year ago	25	0.164**
IN-DEGREE_STATUS (number of mentions received in the questionnaire)	1	Not mentioned or mentioned once in the questionnaire	57	-0.730**
	2	Mentioned twice or three times (2=Avg) in the questionnaire	67	0.046
	3	Mentioned more than 3 times in the questionnaire	39	0.352**
RDGTYPE (fond of reading)	1	Is a keen reader	50	0.153**
	2	Does not read	113	-0.032
MOTHSEC (mother's work sector)	1	Agriculture, livestock, fisheries	1	NC
	2	Hospitality and tourism	33	0.078
	3	Education and social services	26	0.206**
	4	Business	15	-0.086
	5	Construction	0	NC
	6	Health	11	0.156**
MOTHJOB (mother's job)	1	Public administration employee	37	0.045
	2	Company employee	53	0.095*
	3	Owner of a company or business	18	0.064
MUSICTYPE (preferred type of music)	1	Mentions <i>reggaeton</i> music	18	0.211**
	2	Mentions pop music	28	0.206**
MUNICIPALITY (municipality of origin)	1	municipality SS	85	0.425**
	2	municipality VG	26	0.443**
	3	municipality AG	9	0.299**
	4	municipality HE	14	0.288**
	5	municipality AL	15	0.270**
	6	municipality VH	14	0.305**
NUMPART (number of past love relationships)	0	0 relationships in the last 18 months (not counting the current one)	62	0.012
	1	1 relationship in the last 18 months (not counting the current one)	42	0.070
	2	2 relationships in the last 18 months (not counting the current one)	21	0.025
	3	3 or more relationships in the last 18 months (not counting the current one)	16	-0.250

APPENDIX 1. LIST OF VARIABLES AND CATEGORIES (CONTINUATION)

Variable	Category	Description	N	CHI
FATHSEC (father's work sector) ²	1	Agriculture, livestock, fisheries	11	0.027
	2	Hospitality and tourism	24	-0.155
	3	Education and social services	14	0.193**
	4	Business	7	0.019
	5	Plumbing, electricity, construction	32	0.176**
	6	Health	1	NC
FATHJOB (father's job)	1	Public administration employee	29	0.085*
	2	Company employee	55	0.067
	3	Owner of a company or business	28	0.040
PARRELIG (Parents' religious beliefs)	0	Parents with no religious beliefs	24	-0.033
	1	At least one Catholic parent	122	-0.012
	2	At least one parent who professes a religion other than Catholic	14	-1.000**
ALLOWTYPE (allowance received from parents)	1	Receives a weekly allowance from parents	31	-0.018
	2	Does not receive a weekly allowance from parents	132	-0.026
RELATIONSHIP (current relationship)	1	Is currently in a relationship	63	0.139*
	2	Is not currently in a relationship	96	0.084
STUDENT DORMITORY (living in student accommodation)	1	Living in the student dormitory	51	0.706**
	2	Not living in the student dormitory	112	0.635**
Religion (religious beliefs)	0	No religious beliefs	65	0.314**
	1	Has Catholic beliefs	80	0.319**
	2	Has beliefs other than Catholic	16	0.000**
SMOKING (tobacco use)	1	Smoker	30	0.329**
	2	Non-smoker	133	0.429**
VIDEOGAME (fond of videogames)	1	Is a fan of video games	56	0.177**
	2	Is not a fan of video games	108	0.267**

Notes: N = number of individuals; CHI = Coleman homophily index; NC = Not Computable; (*) = Significant ($p < 0.05$); (**) = Significant ($p < 0.01$).

Source: Own elaboration.

APPENDIX 2. AGENT-BASED MODELS

There is a broad group of techniques for programming models and running simulations. The technique used in this article is agent-based modelling, which is different from other techniques of the same type that are also used in the social sciences, such as system dynamics.

The use of this type of model involves writing a sequence of instructions detailing the variables that characterise the system (in the case of this study, homophily indexes), the characteristics of the agents (the attributes of the real individuals), and the rules by which certain attributes of the agents change (the replacement of links with other agents) and, in turn, the characteristics of the system (the new homophily indexes).

The computer executes the set rules recursively until the end condition of the simulation is met. Each simulation is repeated N times, manipulating various parameters to obtain a 'population' of cases that is sufficiently diverse to perform statistical sensitivity analyses. A more detailed explanation can be found in Linares (2018a and 2018b).

APPENDIX 3. PSEUDOCODE OF RRL PROCEDURE

1. Set the number of links to be replaced, $N = 0.15$ times total homophilic links between individuals with attribute A.
2. Set the number of substituted links $M = 0$.
3. As long as $M < N$, repeat steps 4 to 11.
4. Randomly choose an individual i from the set of individuals with homophilic links with respect to attribute A.
5. Randomly choose a link, $v_{i?}$ from the set of homophilic links of i with respect to attribute A.
6. Delete $v_{i?}$.
7. According to procedure P_i (P_i belongs to the set of procedures P for the creation of new links), choose an individual j ($j \neq i$) from the set of individuals with homophilic links with respect to attribute A.
8. Create link v_{ij} .
9. Calculate the value of the CHI.
10. Calculate the statistical significance of the CHI.
11. Set $M = M + 1$.

APPENDIX 4. DEFINITION AND CALCULATION OF THE “EXTENT” INDEX

For every C_i there is a number N_i of indexes in which it participated and a number M_i in which it did not participate. In turn, as can be seen in Table 6, both N_i and M_i are the result of the sum of the number of indexes that remained significant, n_1 or m_1 , and the number of indexes that did not remain significant after the RRL procedure, n_2 or m_2 . In addition, $n_1 + m_1$ must be the number of indexes that remained significant and $n_2 + m_2$ must be the number of indexes that did not remain significant.

Let p_i be the proportion of indexes that remained significant relative to the total number of indexes in which C_i participates, that is, n_1 / N_i , and let q_i be the proportion

of indexes that remained significant relative to the total number of indexes in which C_i did not participate, m_1 / M_i , then the extension index is given by the following equation:

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

TABLE 6. *Distribution of homophily indexes*

	C_i contributed	C_i did not contribute
Remained significant	n_1	m_1
Did not remain significant	n_2	m_2
TOTAL	N_i	M_i

Source: Own elaboration.