

Uso de herramientas informáticas para analizar la autoría e influencias mutuas de los libros del Nuevo Testamento griego

Manuel Alfonseca
UNIVERSIDAD AUTÓNOMA
MADRID

RESUMEN Se ha utilizado la herramienta informática AC para comparar los textos griegos del Nuevo Testamento. La herramienta permite detectar relaciones de semejanza que pueden proceder de influencia, plagio o coautoría. Como experimento de control, se ha utilizado el mismo procedimiento para comparar libros de autores conocidos, obteniéndose resultados coherentes. El uso de esta herramienta permite obtener conclusiones razonables sobre los textos del N.T.

PALABRAS CLAVE Nuevo Testamento griego, agrupamiento, medidas de semejanza, influencia, coautoría.

SUMMARY *The AC computer tool has been used to compare the New Testament texts. This tool makes it possible to detect similarity relations which can proceed from influence, plagiarism or common authorship. As a control experiment, the same procedure has been used to compare a set of books written by known authors, which has given appropriate results. The use of this tool leads to several reasonable conclusions about the texts of the New Testament.*

KEYWORDS *Greek New Testament, clustering, similarity measures, influence, common authorship.*

I. INTRODUCCIÓN

Empiezo por reconocer que no soy experto en el Nuevo Testamento. Sé que las discusiones y teorías sobre quiénes fueron los autores de sus 27 libros y cómo se influyeron unos a otros duran desde hace siglos y constituyen

todo un campo de estudio en el que no puedo competir con los expertos. Sin embargo, como profesional de la Informática, pienso que mi carácter de *no experto* puede ser beneficioso, pues permite ver hasta dónde se puede llegar partiendo de cero (o casi) con el uso de algunas de las herramientas informáticas más modernas.

En primer lugar, voy a resumir lo que quizá pueda considerarse el sentir común, un conjunto mínimo de las ideas más aceptadas por los expertos bíblicos sobre la autoría de los libros del Nuevo Testamento.

- Los tres evangelios sinópticos probablemente se influyeron unos a otros, directa (teoría de las dos fuentes) o indirectamente (teoría de la fuente Q). El orden de influencia no está claro, aunque muchos eruditos consideran a Marcos el más próximo a la fuente original (si la hay).
- El evangelio de Juan es claramente diferente de los sinópticos, aunque también presenta algunas influencias.
- El autor del Libro de los Hechos es el mismo que el del evangelio de Lucas.
- De las 13 epístolas atribuidas a San Pablo, todos los expertos están de acuerdo en que Pablo fue autor de las cuatro siguientes: Corintios 1 y 2; Romanos; Gálatas. La mayor parte de los eruditos le atribuyen tres más: Filemón; Filipenses; Tesalonicenses 1. Otras tres no alcanzan el mismo consenso (las epístolas Deutero-Paulinas: Tesalonicenses 2; Colosenses; Efesios). Algunos creen, por ejemplo, que la epístola a los Efesios fue escrita por un discípulo de San Pablo que usó como apoyo principal la epístola a los Colosenses. Las tres últimas, las epístolas pastorales (Timoteo 1 y 2 y Tito) son atribuidas por bastantes eruditos a un discípulo de San Pablo que podría haber incluido párrafos escritos por su maestro.
- La epístola a los Hebreos, antiguamente asignada a San Pablo, hoy es adscrita de forma unánime a un autor diferente.
- De las dos epístolas de San Pedro, algunos eruditos piensan que la primera la escribió el propio Pedro, aunque otros lo dudan. Sin embargo, en general se piensa que la segunda procede de un autor diferente.
- De las tres epístolas de San Juan, la autoría de la segunda y la tercera se ha puesto en duda desde hace mucho tiempo. Algunos eruditos modernos piensan que las tres fueron escritas por el mismo autor (sea éste San Juan, o no); otros creen que la primera fue escrita por un autor diferente al de las otras dos.

- Hay consenso en que las epístolas de Santiago y San Judas fueron escritas por autores distintos de los de las demás epístolas y de los otros libros del Nuevo Testamento.
- Finalmente, el libro del Apocalipsis se ha adscrito tradicionalmente a San Juan, junto con el evangelio y las tres epístolas. Los eruditos modernos están divididos al respecto.

Desde hace más de 40 años, la informática se viene utilizando como herramienta auxiliar del análisis de textos. El intento más antiguo se remonta a los años sesenta y se basa en el análisis estadístico léxico¹, que analiza la frecuencia de aparición de las palabras. Después vino el desarrollo de técnicas centradas en el contenido, como el análisis factorial de correspondencias², y más tarde las que se apoyan en las características sintácticas³ y semánticas de los textos objeto del análisis. Todas estas técnicas han proporcionado resultados muy válidos y sugerentes.

Este artículo describe el uso de la herramienta informática AC para detectar semejanzas y relaciones de autoría o de influencia entre los libros de la versión griega del Nuevo Testamento. Dicha herramienta no sabe nada de las opiniones previas de los eruditos, y tampoco se apoya en la frecuencia de uso de las palabras, ni en las propiedades léxicas, sintácticas o semánticas de los textos. De hecho, la herramienta ignora que los objetos que está comparando son textos en griego, y se limita a comparar cadenas de bits, a pesar de lo cual es capaz de sacar conclusiones razonables. Cada texto se considera como una simple cadena de bits (16 bits por cada carácter griego y 8 bits por cada carácter latino). La herramienta se limita a medir el grado de semejanza (distancia) entre ambas cadenas. Al utilizar una técnica diferente de las que usualmente se emplean, pienso que este estudio podría arrojar alguna luz independiente sobre los problemas que acabamos de comentar y quizá abrir una nueva línea de investigación.

1 C. MULLER, *Initiation à la statistique linguistique* (Larousse, Paris 1968); en alemán: *Einführung in die Sprachstatistik* (Hueber, München 1972).

2 J.-P. BENZÉCRI, *L'Analyse des Données. Volume II. L'Analyse des Correspondances* (Dunod, Paris 1973).

3 Cf. I. CARBAJOSA, *El valor del análisis sintáctico para discernir entre griego original y griego de traducción en la Biblia. Status quaestionis*, en J. M. GARCÍA PÉREZ (ed.), *Rastreado los orígenes. Lengua y exégesis en el Nuevo Testamento. En memoria del profesor Mons. Mariano Herranz Marco*, *Studia Semitica Novi Testamenti* 17 (Madrid 2011) 33-54.

La segunda sección del artículo explica los procedimientos informáticos utilizados. La tercera los aplica a un caso de control (un conjunto de libros escritos por autores conocidos), para asegurarnos de la eficacia del método. La cuarta detalla los experimentos realizados sobre los libros del Nuevo Testamento. Finalmente se presentan algunas conclusiones y líneas de trabajo futuro.

II. DESCRIPCIÓN DEL MÉTODO Y LA HERRAMIENTA INFORMÁTICA UTILIZADOS

En muchas ramas de la ciencia es importante disponer de herramientas que permitan comparar objetos (archivos) diversos, tales como textos, series de medidas, música, cadenas de nucleótidos o datos de muchos otros tipos. La disciplina que estudia de forma teórica estas comparaciones se llama teoría de agrupamientos (*clustering*, en inglés), porque tiene por objeto clasificar los objetos de que se trate de acuerdo con ciertos criterios, en función de propiedades comunes que compartan. Para realizar el agrupamiento se utilizan medidas de distancia o cercanía entre los objetos.

Existen métricas universales de semejanza que es posible aplicar para medir distancias entre archivos cualesquiera. Una de ellas, la distancia de información normalizada (DIN)⁴, ha demostrado ser mejor para este propósito que cualquier otra métrica computable. Dicha distancia se calcula en función de la complejidad de Kolmogorov condicional (CKC), cuya definición, propuesta por Kolmogorov, se remonta a los años sesenta. Desgraciadamente, ni CKC ni DIN son funciones computables.

Para medir la semejanza entre dos objetos, “x” e “y”, parece natural pensar que ambos serán tanto más similares cuanto mayor sea la parte de “x” que está incluida en “y”, y viceversa. De hecho, “x” e “y” serán iguales si todo “x” está incluido en “y”, y todo “y” está incluido en “x”.

Un compresor es una herramienta informática que convierte un archivo en otro equivalente, buscando repeticiones y aprovechándolas para describir el archivo original con un tamaño más pequeño. Supongamos que tenemos dos objetos o archivos, “x” e “y”, que deseamos comprimir, obteniendo dos nuevos archivos, $C(x)$ y $C(y)$, cuyos tamaños serán en principio más pequeños

4 M. LI – P. B. M. VITANYI, *An Introduction to Kolmogorov Complexity and its Applications* (Springer-Verlag, New York 1997).

que los de “x” e “y”, puesto que los hemos comprimido. A continuación se utiliza el mismo compresor para comprimir la concatenación de los dos objetos, obteniendo el archivo $C(xy)$. Si su tamaño es sólo un poco mayor que el de $C(x)$, esto significa que gran parte de la información que contiene el objeto “y” está también contenida en el objeto “x”, pues el compresor ha sido capaz de utilizar la información de “x” para comprimir mucho más “y”. Por el contrario, si “x” e “y” son muy diferentes, el compresor casi no encontrará semejanzas entre las dos partes y el tamaño de $C(xy)$ será próximo a la suma de los tamaños de $C(x)$ y $C(y)$.

Este procedimiento fue formalizado por Cilibrasi and Vitanyi⁵ en el concepto de *distancia de compresión normalizada* (NCD, por sus siglas en inglés), que utiliza compresores para medir la semejanza (o distancia) entre los objetos. Dicha distancia puede entonces utilizarse para agrupar esos objetos. Se ha demostrado que la NCD es una estimación computable de la DIN. Se trata de una distancia de semejanza normalizada casi universal, aplicable para el agrupamiento de objetos de cualquier tipo.

La idea es muy poderosa, pues puede aplicarse por igual a objetos tales como música⁶, textos o secuencias genéticas. No es necesario recurrir a características específicas de los objetos que se va a agrupar, como la sintaxis o el significado de las palabras en el caso de textos, o las notas musicales concretas en el caso de música, pues sólo se trabaja con cadenas de bits. Lo único que se necesita para calcular la distancia de un objeto “p” a otro “q”, es medir hasta qué punto un compresor es capaz de reducir el tamaño de “q” al concatenarlo con “p”, y viceversa.

La distancia de compresión normalizada se define así:

$$NCD(p, q) = \frac{C(pq) - \min [C(p), C(q)]}{\max [C(p), C(q)]}$$

donde “pq” es la concatenación de los archivos u objetos “p” y “q” (considerados como cadenas de bits), y $C(x)$ denota la longitud de la cadena “x”, una vez comprimida. Es decir, se toma el objeto “p” y se le comprime con un com-

5 R. CILIBRASI – P. VITANYI, “Clustering by compression”, *IEEE Transactions on Information Theory* Vol.51:4 (2005) 1523-1545.

6 R. CILIBRASI – P. VITANYI, “Algorithmic clustering of music”, *Proceedings of the Fourth International Conference on Web Delivering of Music* (IEEE Computer Society, 2004) 49-67.

presor determinado; se mide el tamaño del archivo comprimido: $C(p)$. Se hace lo mismo con el objeto “q”, y se obtiene $C(q)$. Finalmente se comprime la concatenación de los dos archivos “p” y “q”, y se mide el tamaño resultante $C(pq)$. Se resta de $C(pq)$ el menor de los números $C(p)$ y $C(q)$, y se divide por el mayor de los dos. El resultado es la distancia de compresión normalizada.

Para dos objetos idénticos, $C(p)=C(q)\approx C(pq)$, pues el compresor debería ser capaz de detectar esa igualdad y comprimir “pq” casi con el mismo tamaño que “p” o que “q”. En tal caso, la distancia resultante será muy próxima a 0. En cambio, si los dos objetos son totalmente diferentes, $C(pq)$ será más o menos igual a $C(p)+C(q)$, con lo que la distancia resultará ser aproximadamente igual a 1. Este método de medida nos proporciona, por lo tanto, distancias en una escala entre 0 y 1, donde 0 corresponde a la identidad y 1 a la máxima separación entre los objetos.

El concepto de la NCD ha servido de base para construir una herramienta informática llamada AC⁷, que se está aplicando con éxito para detectar plagios en los trabajos prácticos que entregan los estudiantes (en este caso, los textos que se comparan son programas escritos en un lenguaje de ordenador). AC genera diversos resultados útiles, que ayudan a tomar decisiones sobre la proximidad o independencia de los objetos que se comparan:

- Un dendrograma⁸ de dichos objetos, que se genera por medio de métodos de agrupamiento jerárquico, utilizando los procedimientos clásicos de la teoría de agrupamientos.
- Un grafo⁹ en el que los objetos a comparar están unidos por arcos a los que se asocia un peso (o longitud del arco), que representa la distancia (NCD) entre los dos objetos situados en los extremos del arco. Puesto que el objetivo inicial de la herramienta AC es la detección de plagio entre los estudiantes, este grafo (que llamaremos de distancia mínima) no muestra todas las relaciones posibles entre todos los nodos, sino sólo las más cortas (las que indican cuál es la fuente más probable). Por ello, la ausencia de conexión entre dos nodos concretos no significa que no

7 M. FREIRE – M. CEBRIÁN – E. DEL ROSAL, *Uncovering plagiarism networks*, arXiv:cs/0703136v7 [cs.IT] (2007).

8 Representación gráfica de datos en forma de árbol, que permite apreciar relaciones de agrupación entre los datos y entre grupos de ellos.

9 Un dibujo que representa un conjunto de objetos (nodos) unidos por enlaces (arcos) que representan relaciones entre los elementos del conjunto.

estén conectados entre sí, sino que existe otro nodo más cercano a ambos. Antes de construir el grafo, la matriz de distancias puede podarse, eliminando todas las que son superiores a un valor dado.

Para calcular la NCD, AC ofrece una selección de compresores diferentes. En estos experimentos he utilizado Ppmz¹⁰ cuando hay que comparar textos completos, porque este compresor no tiene limitación de tamaño, aunque es más lento. Cuando se comparan trozos más pequeños se ha utilizado Zip, que tiene un buffer de trabajo de 32 kbytes, lo que limita el tamaño máximo de los objetos a comparar a menos de 16 kbytes¹¹.

Al utilizar la herramienta AC para comparar dos textos, debemos tener en cuenta que las semejanzas pueden deberse a dos causas diferentes:

- Influencia: el autor de un texto puede haber utilizado como apoyo o punto de partida otro texto escrito por un autor diferente (esto ocurre, por ejemplo, en los plagios, pero también puede deberse a razones honradas).
- Coautoría: si ambos textos fueron escritos por el mismo autor, las semejanzas de estilo y las preferencias en el uso de palabras pueden ser detectables mediante herramientas apropiadas.

A priori, el uso de la herramienta AC no permite distinguir entre las dos razones, por lo que en cada caso concreto será preciso utilizar información adicional para deducir cuál es la causa de la semejanza. Por ejemplo, al comparar los archivos entregados por los estudiantes, sabemos que cada uno viene firmado por autores diferentes, por lo que cualquier semejanza importante despierta de forma automática sospechas de plagio. En el caso del Nuevo Testamento, las posibilidades son mucho más amplias.

10 Ppmz es un algoritmo de compresión de datos que utiliza el método PPM (Prediction by Partial Match) para predecir los datos siguientes a partir de los anteriores y, de ese modo, comprimirlos.

11 M. CEBRIÁN – M. ALFONSECA – A. ORTEGA, "Common Pitfalls Using Normalized Compression Distance: What to Watch Out for in a Compressor", *Communications in Information and Systems*, Vol. 5:4 (2005) 367-384.

III. CASO DE CONTROL

Se ha llevado a cabo un experimento de control, para ver qué resultados da AC cuando se conocen los autores de los libros a comparar.

Se eligió la siguiente lista de 25 libros, todos ellos escritos en inglés durante el siglo XIX por 10 autores diferentes (8 mujeres y 2 hombres)¹²:

- Las seis novelas de Jane Austen: *Emma* [Emma], *Mansfield Park* [Mans], *Pride and prejudice* [Pride], *Persuasion* [Persu], *Sense and sensibility* [Sense], y *The Northanger Abbey* [Abbey].
- Seis libros de Charles Dickens: una novela larga (*A tale of two cities* [2cities]) y cinco historias más breves, que pertenecen a su ciclo de cuentos de Navidad: *A Christmas carol* [Carol], *The cricket of the hearth* [Cricket], *The chimes* [Chimes], *The battle of life* [Battle], y *The haunted man and the ghost's bargain* [Ghost].
- Cuatro novelas de Elizabeth Gaskell: *North and South* [North], *Cranford* [Cranf], *My lady Ludlow* [Ludlow], y *Cousin Phyllis* [Phyllis]. La primera es bastante más larga que las otras tres.
- Tres libros de George Eliot. Uno de ellos (*Silas Marner* [Silas]) es una novela; los otros dos son ensayos: *Impressions of Theophrastus Such* [Teoph] y una colección de 10 ensayos más breves [Essays].
- *Wuthering Heights* de Emily Brontë [Wuther].
- *Agnes Grey* de Anne Brontë [Agnes].
- *Frankenstein* de Mary Shelley [Franken].
- *Patricia* de Emilia Elliott [Patri].
- *The warden* de Anthony Trollope [Warden].
- *A vindication of the rights of woman* de Mary Wollstonecraft [Woman].

La figura 1 muestra el dendrograma producido tras el análisis de estos 25 textos por la herramienta AC. Es curioso observar lo siguiente:

- Las seis novelas de Jane Austen quedan agrupadas y claramente separadas del resto de los libros.
- Las cinco historias breves de Dickens quedan estrechamente agrupadas, pero su novela larga queda aparte.

¹² Al lado de cada libro se indica entre corchetes el nombre resumido con que aparece en las figuras que acompañan el texto.

- Los tres libros de ensayos (dos de George Eliot y uno de Mary Wollstonecraft) aparecen agrupados. Es curioso que la única novela que queda ligada a ellos sea *Frankenstein de Mary Shelley*, hija de Mary Wollstonecraft.
- Las cuatro novelas de Elizabeth Gaskell no están agrupadas (pero véase más abajo).

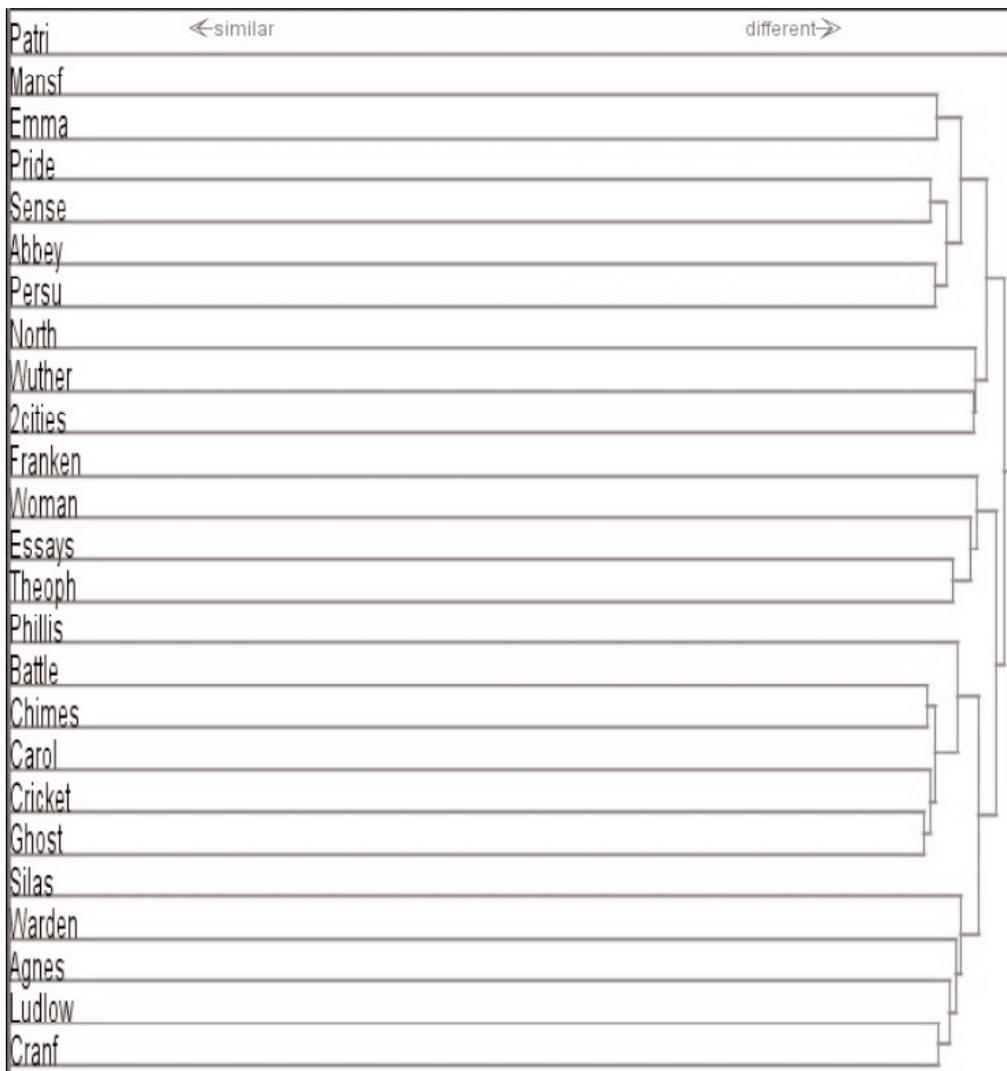


Fig. 1: Dendrograma de los 25 libros de control generados por la herramienta AC.

La figura 2 muestra el grafo de distancia mínima sin realizar ninguna poda. Es interesante notar lo siguiente:

- Las seis novelas de Jane Austen forman un subgrafo casi conexo. Nótese que la ausencia de enlace entre dos nodos (como *Emma* y *Persuasion*) no significa que la herramienta no haya detectado semejanzas entre ellas, sino que existen otras novelas de la autora más próximas a ellas. Observando el dendrograma, se ve que la más próxima a *Emma* es *Mansfield Park*, mientras la más próxima a *Persuasion* es *The Northanger Abbey*. De hecho, las dos primeras están relativamente alejadas de las otras cuatro, aunque siempre están más próximas a ellas que a cualquiera de los otros 19 libros de control.
- Las cinco historias cortas de Dickens forman un subgrafo fuertemente conexo, pero la novela larga está bastante lejos de ellas.
- Las cuatro novelas de Elizabeth Gaskell quedan ligadas linealmente entre sí, aunque no de forma tan clara como en los dos casos anteriores, pues también presentan enlaces con libros de otros autores (Trollope, Dickens, George Eliot, Emilia Elliott y Anne Brontë).
- Los tres libros de George Eliot están ligados, aunque dos de ellos son ensayos y el tercero es una novela.
- Los seis libros de autores diferentes quedan relegados a posiciones periféricas. Sin embargo, es interesante notar que el ensayo de Mary Wollstonecraft queda unido a los otros dos ensayos, y que las dos novelas de las hermanas Brontë también están ligadas entre sí.

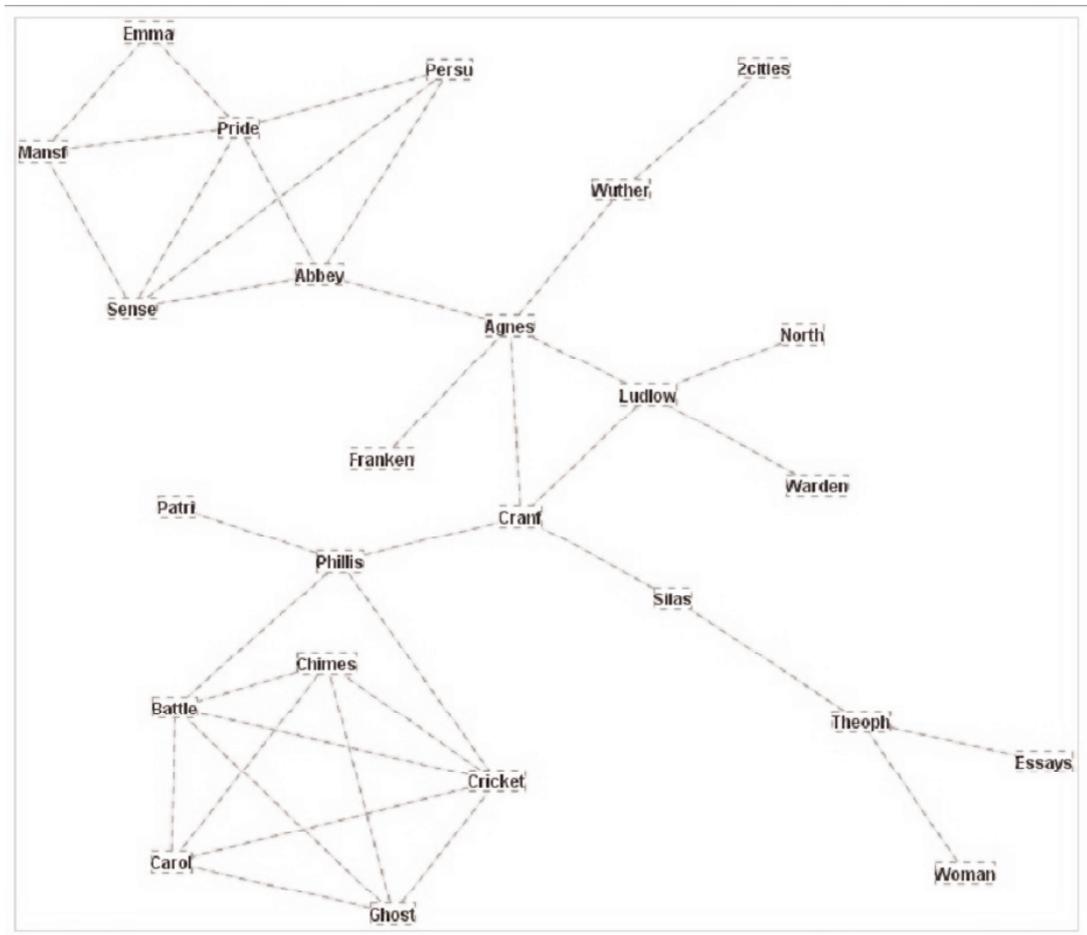


Fig. 2: Grafo de distancia mínima (sin poda) entre los 25 libros de control, obtenido por la herramienta AC.

La figura 3 muestra el resultado de podar el grafo de la figura 2 eliminando todos los arcos mayores que 0,93. El grafo queda reducido a tres secciones disjuntas que sólo contienen 14 libros (los otros 11 están separados de cualquiera de los demás por distancias mayores que ese valor). Una de las secciones incluye las seis novelas de Jane Austen; otra sección comprende las cinco historias de Navidad de Dickens; la tercera incluye dos de las novelas de Elizabeth Gaskell y la de Anne Brontë.

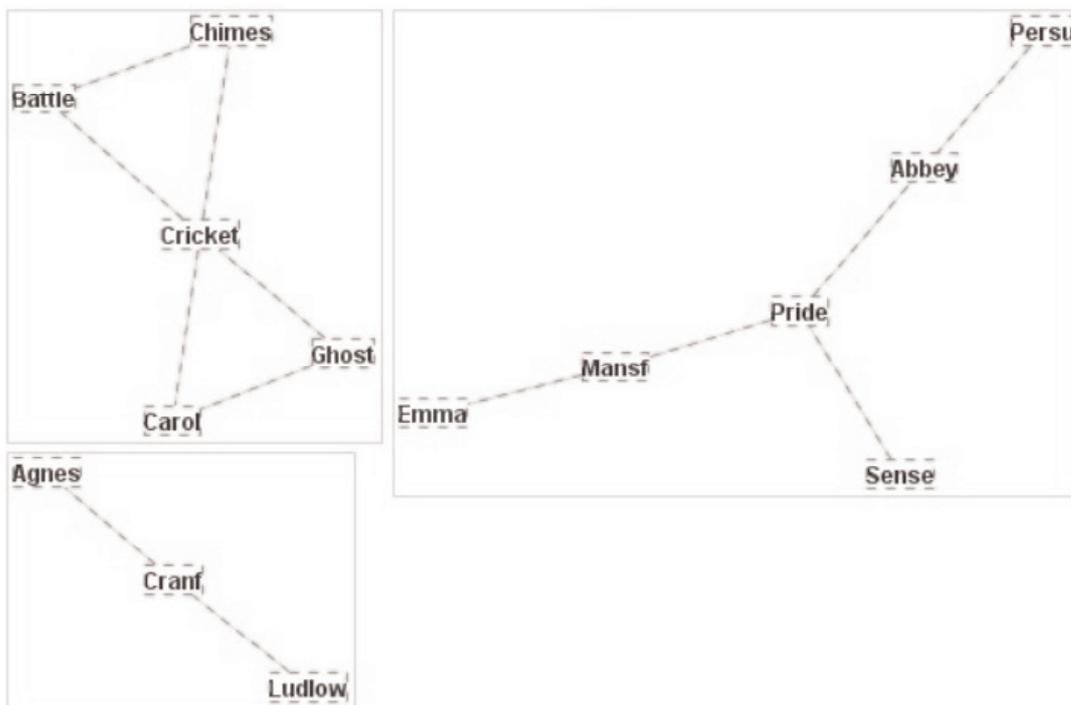


Fig. 3: Resultado de podar todas las ramas con distancias mayores que 0.93 en la fig.2.

En conclusión: El experimento de control puede considerarse un éxito. La herramienta AC ha sido capaz de detectar relaciones de autoría común, con la excepción de la novela larga de Dickens, que ha quedado lejos de sus historias breves. Esto podría deberse a dos razones: o bien la longitud de los libros es un punto a tener en cuenta, o quizá Dickens utilizaba un estilo diferente en sus historias cortas y en sus novelas largas.

IV. EL NUEVO TESTAMENTO GRIEGO

1. PRIMER EXPERIMENTO CON LOS LIBROS DEL NT:

ANÁLISIS DE LOS LIBROS COMPLETOS

En el primer experimento proporcioné a la herramienta AC los 27 libros completos del Nuevo Testamento griego. Los nombres que asigné a los archivos recuerdan (aunque no son idénticos) a las abreviaturas usualmente utilizadas para los libros, tal como se indica en la tabla 1. La figura 4 presenta el dendrograma obtenido por la herramienta al comparar todos los libros entre sí.

Tabla 1. Correspondencia entre los nombres de los archivos y las abreviaturas usuales de los libros del Nuevo Testamento.

Mateo	Mt	Marcos	Mc	Lucas	Lc	Juan	Jn	HCH	Hch	ROM	Ro
COR1	1Co	COR2	2Co	GAL	Gl	EFE	Ef	FIP	Flp	COL	Col
TS1	1Ts	TS2	2Ts	TM1	1Ti	TM2	2Ti	TIT	Tit	FIM	Flm
HEB	Heb	SAN	Stg	P1	1P	P2	2P	J1	1Jn	J2	2Jn
J3	3Jn	JUD	Jud	AP	Ap						

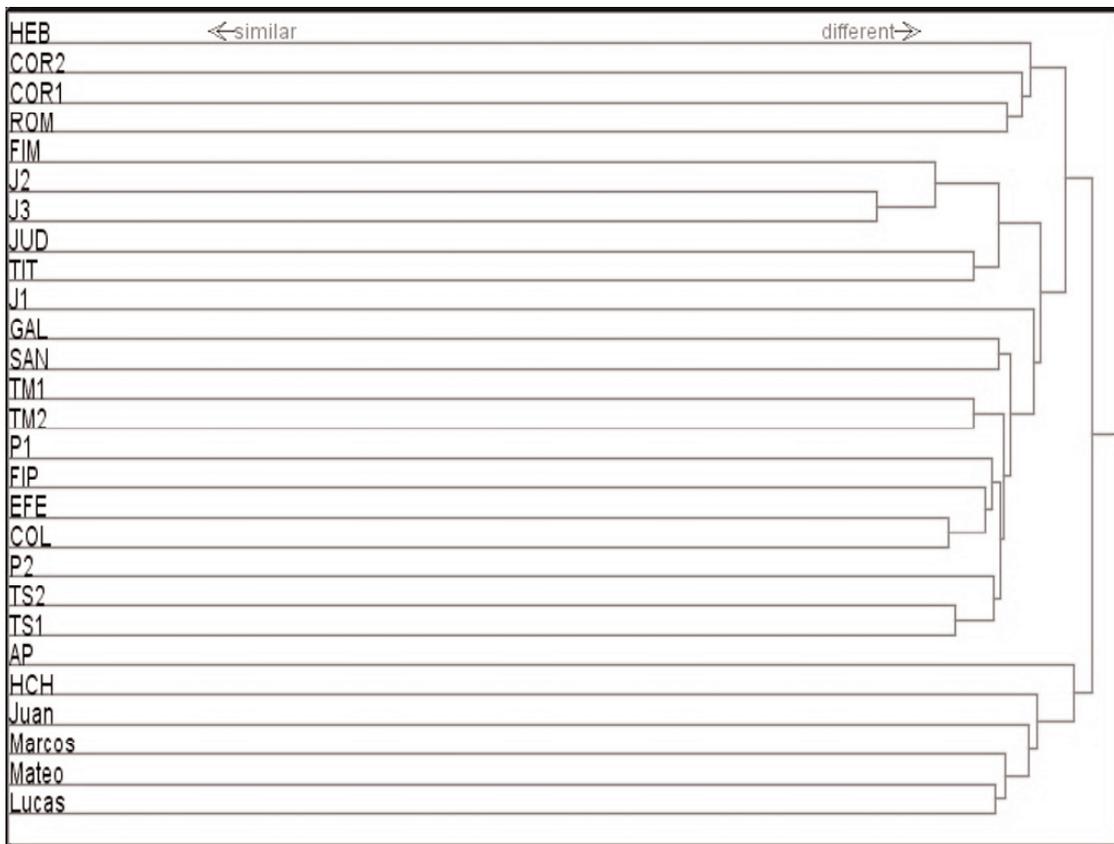


Fig. 4: Dendrograma de los 27 libros del Nuevo Testamento, generado por la herramienta AC.

Los principales puntos de interés que se pueden extraer de este dendrograma son los siguientes:

- Tres de las cuatro epístolas indudablemente paulinas quedan agrupadas con la epístola a los Hebreos.

- Las epístolas de Juan 2 y 3 aparecen muy próximas entre sí, pero lejos de Juan 1.
- Pedro 1 y 2 están lejos una de otra, pero no demasiado.
- Los tres evangelios sinópticos quedan muy cerca unos de otros, y se agrupan con el evangelio de Juan y con los Hechos.
- Tesalonicenses 1 y 2 quedan agrupadas.
- Timoteo 1 y 2 también están agrupadas.
- La herramienta detecta una relación estrecha entre las epístolas a los Efesios y los Colosenses.

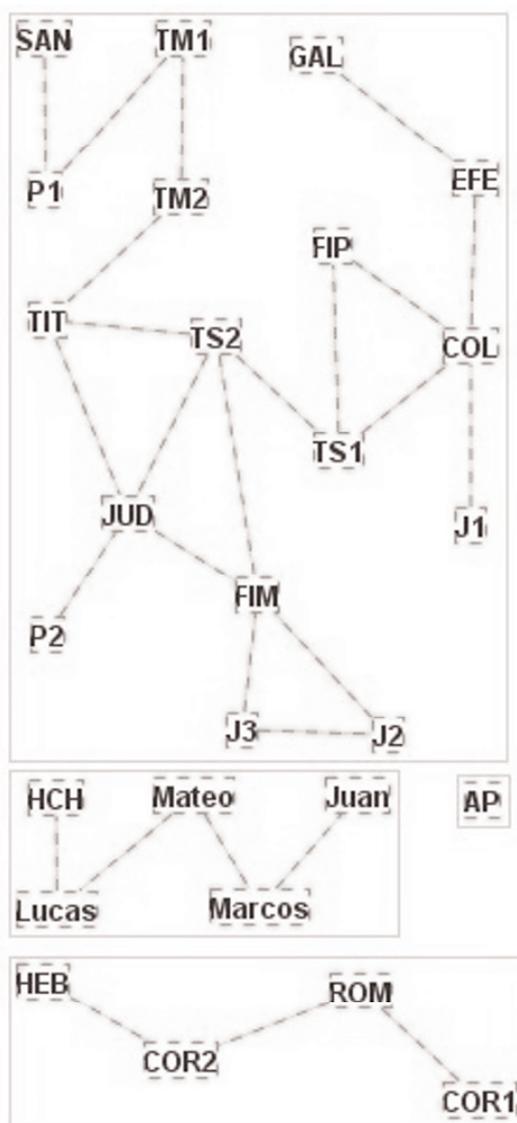


Fig. 5: Grafo asociado al dendrograma de la figura 4, tras podarlo.

La figura 5 muestra un grafo obtenido del dendrograma si se eliminan todas las relaciones de distancia mayor que 0,95. Los 27 libros del NT quedan agrupados en cuatro grupos: uno que incluye las cuatro epístolas mencionadas en el primer punto arriba; el segundo con todas las epístolas restantes; el tercero con los cuatro evangelios más los Hechos; en el cuarto, el Apocalipsis se ha quedado solo. Obsérvese que las diez epístolas atribuidas a San Pablo están ligadas entre sí por una cadena continua, desde GAL hasta TM1, con FIM (Filemón) en una rama lateral.

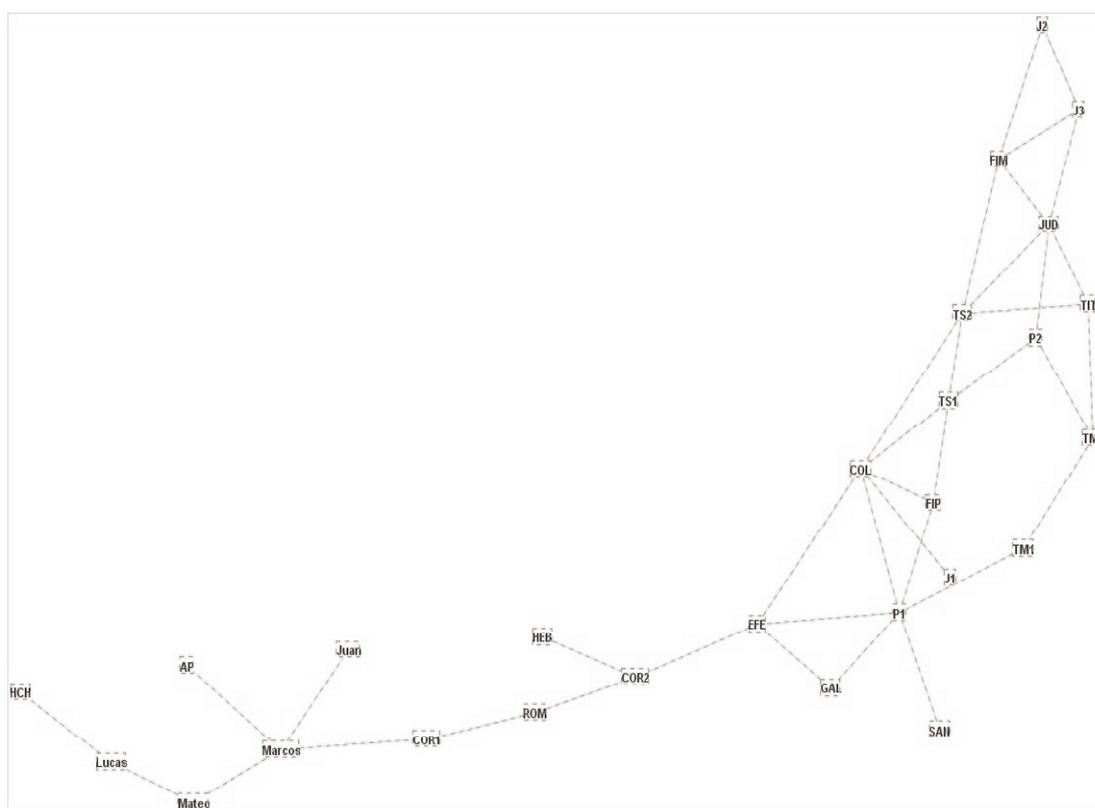


Fig. 6: Grafo de distancia mínima sin podar entre los 27 libros del NT.

La figura 6 muestra el grafo de distancia mínima sin podar generado por AC. Obsérvese los siguientes detalles:

- El libro más próximo a los Hechos es el evangelio de Lucas.
- Aunque el Apocalipsis está lejos de todos los demás libros, el más próximo resulta ser el evangelio de Marcos.

- Los tres sinópticos están juntos.
- Las cuatro epístolas paulinas indudables (COR1 y 2, ROM y GAL) están próximas unas a otras, aunque Efesios se interpone entre GAL y COR2.
- Las trece epístolas tradicionalmente atribuidas a San Pablo están ligadas entre sí. Once de ellas forman una cadena: COR1-ROM-COR2-EFE-COL-FIP-TS1-TS2-TIT-TM2-TM1, con GAL y FIM como ramas laterales.
- Queda clara la relación estrecha entre Efesios y Colosenses.
- P1 y P2 están alejadas entre sí, como también J1 de J2-J3.
- Hebreos aparece como rama lateral, más próxima a COR2.

Las figuras 7 a 9 muestran los grafos obtenidos mediante podas sucesivas.

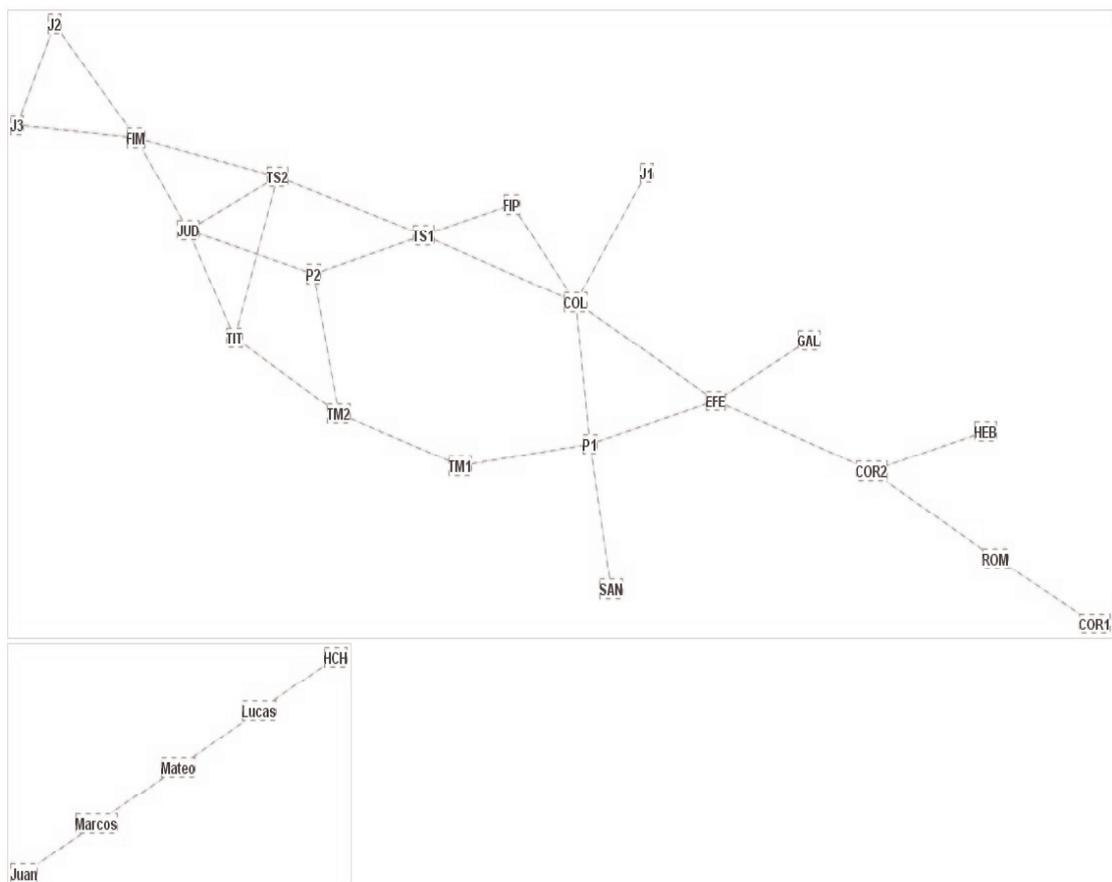


Fig. 7: Grafo de distancia mínima podado a 0,94 para los libros del NT. Hechos queda agrupado con los cuatro evangelios en un subgrafo separado. Apocalipsis se queda fuera.

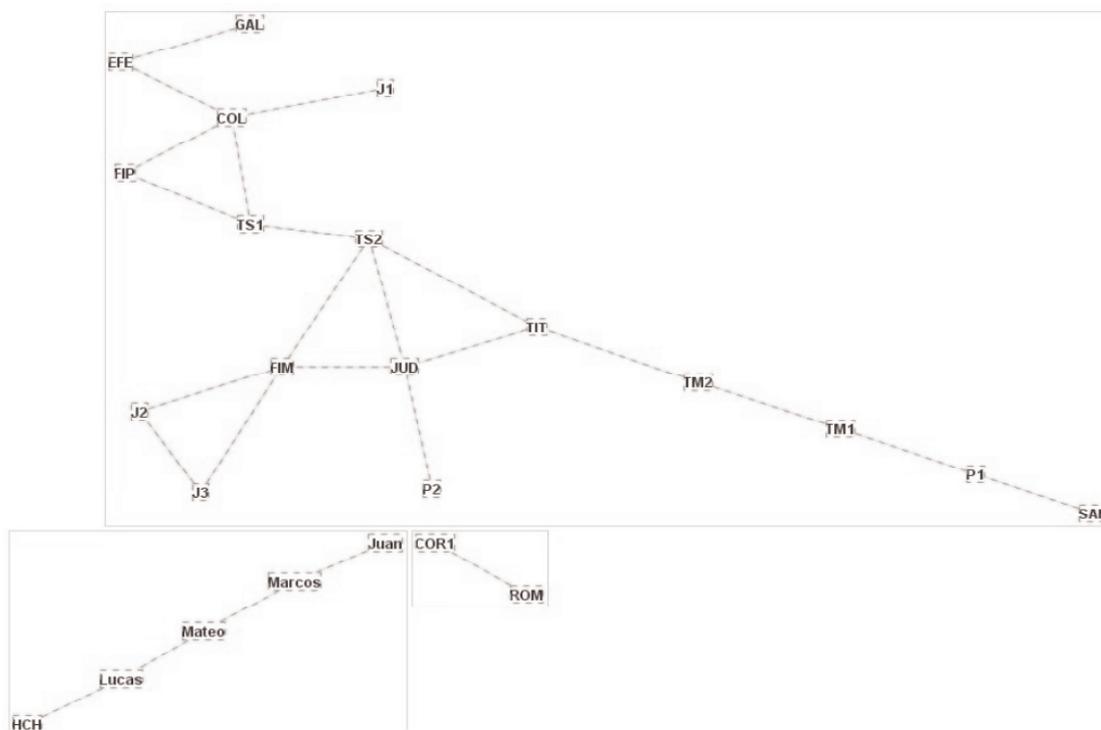


Fig. 8: Grafo de distancia mínima podado a 0,91 para los libros del NT. COR1 y ROM quedan agrupados en un subgrafo separado. COR2 y HEB se quedan fuera.

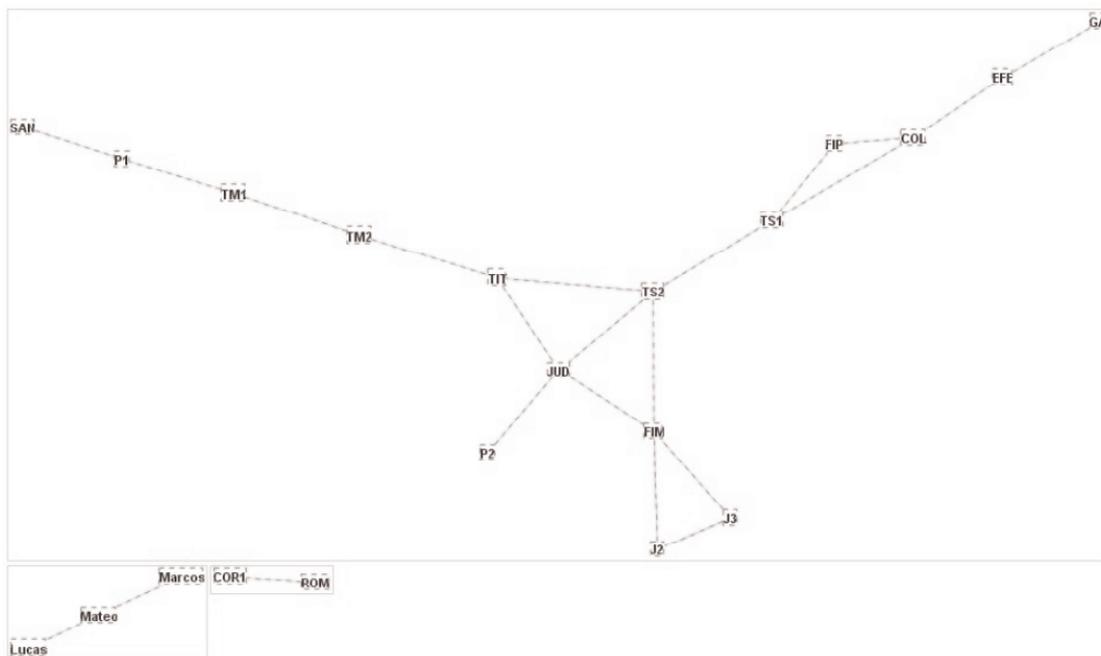


Fig. 9: Grafo de distancia mínima podado a 0,90 para los libros del NT. Los tres sinópticos se quedan solos en un subgrafo. Hechos y el evangelio de Juan se quedan fuera, junto con la primera epístola de San Juan.

2. SEGUNDO EXPERIMENTO CON LOS LIBROS DEL NT: ANÁLISIS DE LOS LIBROS DIVIDIDOS EN TROZOS DE LA MISMA LONGITUD

En el segundo experimento, los 27 libros del NT griego se dividieron en secciones de la misma longitud (15 kbytes), con objeto de reducir al mínimo cualquier efecto debido a las diferencias de tamaño de los objetos sometidos a la comparación por la herramienta AC. Para interpretar los resultados, hay que tener en cuenta que cada carácter griego ocupa 2 bytes, mientras que los caracteres latinos (espacios en blanco y algunos más) ocupan un solo byte, de modo que cada fragmento contiene aproximadamente unos 7680 caracteres. Por supuesto, la última sección de cada libro tiene una longitud diferente, pues sólo incluye los caracteres restantes del libro. La tabla 2 muestra el número de secciones para cada libro y el tamaño de la última sección.

En este caso el número de componentes manejado (120) hace que no sea práctico presentar dendrogramas y grafos globales. La figura 10 muestra un subgrafo derivado del dendrograma que se obtiene del análisis, que incluye la mayor parte de las secciones de los cuatro evangelios. Obsérvese que las piezas del evangelio de San Juan aparecen conectadas entre sí y separadas de los otros tres evangelios, excepto por la penúltima parte (JUAN11¹³), que está ligada con la penúltima parte de Mateo (MATEO14), y a través de ésta a las penúltimas partes de Marcos y Lucas. (MARCOS9 y LUCAS15). Esto es comprensible, pues estas partes de los cuatro evangelios describen la Pasión de Cristo, así que hablan del mismo tema y se parecen entre sí más que otras.

13 Recuérdese que JUAN11 no representa el capítulo 11 del Evangelio de San Juan, sino la pieza número 11 obtenida al dividir el Evangelio en partes iguales de 15 kbytes.

Tabla 2: Número de secciones en que se divide cada libro del NT.
Todas las secciones excepto la última ocupan exactamente 15Kbytes.

Libro	Nr. de secciones	Tamaño de la última
AP (Apocalipsis)	8	8Kb
COL (Colosenses)	2	5Kb
COR1 (1 Corintios)	6	6Kb
COR2 (2 Corintios)	4	10Kb
EFE (Efesios)	2	15Kb
FIM (Filemón)	1	4Kb
FIP (Filipenses)	2	5Kb
GAL (Gálatas)	2	13Kb
HCH (Hechos)	16	6Kb
J1 (1 Juan)	2	9Kb
J2 (2 Juan)	1	3Kb
J3 (3 Juan)	1	3Kb
JUAN (Evangelio de Juan)	12	13Kb
JUD (Judas)	1	7Kb
LUCAS (Lucas)	16	10Kb
MARCOS (Marcos)	10	3Kb
MATEO (Mateo)	15	10Kb
P1 (1 Pedro)	2	7Kb
P2 (2 Pedro)	1	15Kb
ROM (Romanos)	6	10Kb
SAN (Santiago)	2	7Kb
TIT (Tito)	1	9Kb
TM1 (1 Timoteo)	2	7Kb
TM2 (2 Timoteo)	2	1Kb
TS1 (1 Tesalonicenses)	2	4Kb
TS2 (2 Tesalonicenses)	1	10Kb

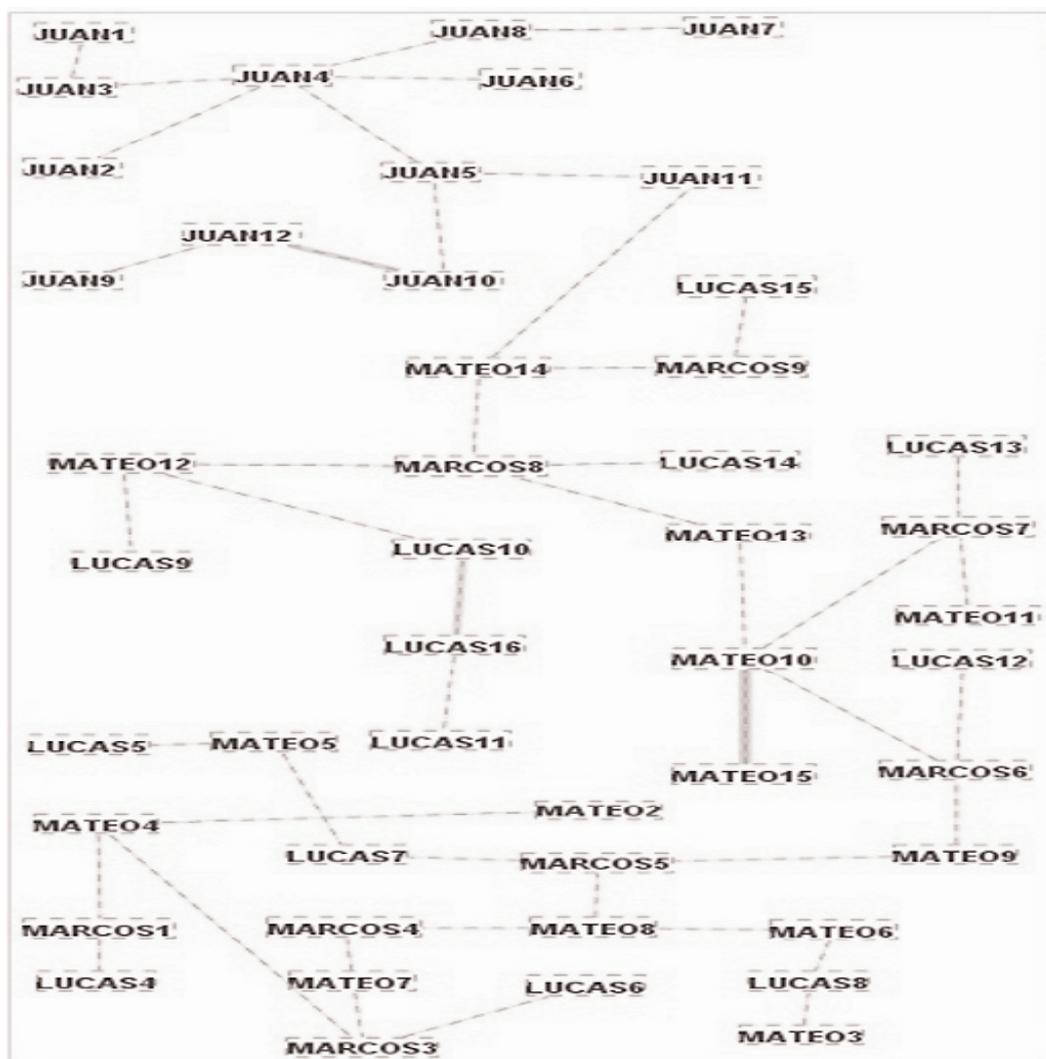


Fig. 10: Agrupamiento asociado al dendrograma para algunas de las 120 piezas de los libros del NT.

Es fácil ver que en los evangelios sinópticos la semejanza de tema domina sobre la autoría. La columna vertebral que los relaciona alterna entre los evangelios de Mateo y Marcos así: MARCOS9 - MATEO14 - MARCOS8 - MATEO13 - MATEO10 - MARCOS6 - MATEO9 - MARCOS5 - MATEO8 - MARCOS4 - MATEO7 - MARCOS3 - MATEO4 - MARCOS1, con las distintas secciones de Lucas y las restantes de Mateo y Marcos como ramas laterales. Esto podría indicar que los evangelios de Marcos y Mateo serían los dos primeros, mientras el de Lucas deriva a veces de Mateo y a veces de Marcos. La tabla 3 muestra la equivalencia en capítulos y versículos de las secciones de 15 Kbyte de los cuatro evangelios.

Tabla 3: Capítulos y versículos contenidos en cada sección de los cuatro evangelios.

Sección	Capítulo- versículos	Sección	Capítulo- versículos	Sección	Capítulo-ver- sículos	Sección	Capítulo-ver- sículos
MARCOS1	1:1 - 2:27	MATEO1	1:1 - 4:1	LUCAS1	1:1 - 2:4	JUAN1	1:1 - 3:3
MARCOS2	2:27 - 4:41	MATEO2	4:1 - 6:1	LUCAS2	2:4 - 3:23	JUAN2	3:3 - 4:44
MARCOS3	5:1 - 6:30	MATEO3	6:1 - 8:9	LUCAS3	3:23 - 5:17	JUAN3	4:44 - 6:19
MARCOS4	6:30 - 8:11	MATEO4	8:9 - 10:8	LUCAS4	5:17 - 6:45	JUAN4	6:19 - 7:26
MARCOS5	8:11 - 9:43	MATEO5	10:8-12:11	LUCAS5	6:45 - 8:13	JUAN5	7:26 - 8:49
MARCOS6	9:43-11:15	MATEO6	12:1-13:30	LUCAS6	8:13 - 9:22	JUAN6	8:49-10:25
MARCOS7	11:15-13:9	MATEO7	13:30-15:10	LUCAS7	9:22-10:26	JUAN7	10:25 - 12:5
MARCOS8	13:9-14:48	MATEO8	15:10-17:15	LUCAS8	10:26-11:53	JUAN8	12:5-13:31
MARCOS9	14:49-16:9	MATEO9	17:15-19:21	LUCAS9	11:53-13:12	JUAN9	13:31-16:11
MARCOS10	16:9-16:20	MATEO10	19:21-21:28	LUCAS10	13:12-15:10	JUAN10	16:11-17:22
		MATEO11	21:28-23:16	LUCAS11	15:10-17:18	JUAN11	17:23-19:6
		MATEO12	23:16-25:1	LUCAS12	17:18-19:17	JUAN12	19:6-21:25
		MATEO13	25:1-26:35	LUCAS13	19:17-21:4		
		MATEO14	26:35-27:35	LUCAS14	21:4-22:48		
		MATEO15	27:35-28:20	LUCAS15	22:48-24:4		
				LUCAS16	24:4-24:53		

La figura 11 muestra un segundo subgrafo derivado del mismo dendrograma. Todas las secciones del Apocalipsis se agrupan entre sí. Otro grupo conecta cuatro de las secciones de Hebreos con la segunda epístola de Pedro, y ésta con una sección de la segunda epístola a Timoteo. Finalmente, varias secciones de las epístolas indudables de San Pablo (Gálatas y 2ª Corintios enteras, más cuatro secciones de Romanos y tres de 1ª Corintios) aparecen juntas en un solo *cluster*¹⁴.

La figura 12 muestra cinco *clusters* más. Dos de ellos contienen las secciones que faltaban de 1ª Corintios y Romanos. Otro vuelve a mostrar la relación especial entre Efesios y Colosenses. Otro agrupa la mayor parte de las secciones de Hechos, y muestra que dos de ellas (HCH5 y HCH8) están próximas a dos secciones del evangelio de Lucas (LUCAS3 y LUCAS2). El

14 La palabra inglesa *cluster* (que puede traducirse por *agrupamiento*) se utiliza usualmente en el argot de las tecnologías de proceso de información para referirse a los grupos en que pueden clasificarse diversos objetos.

último *cluster* agrupa Tesalonicenses 1 y 2, Filipenses y Filemón con la segunda y tercera epístolas de Juan, la pieza que faltaba de Hebreos, un par de secciones de Marcos y una de Timoteo 2.

Finalmente, la figura 13 muestra los dos *clusters* que faltan: uno que aísla la primera epístola de Juan de todo lo demás, y otro que une Timoteo 1, Tito, Pedro 1, Santiago y Judas.

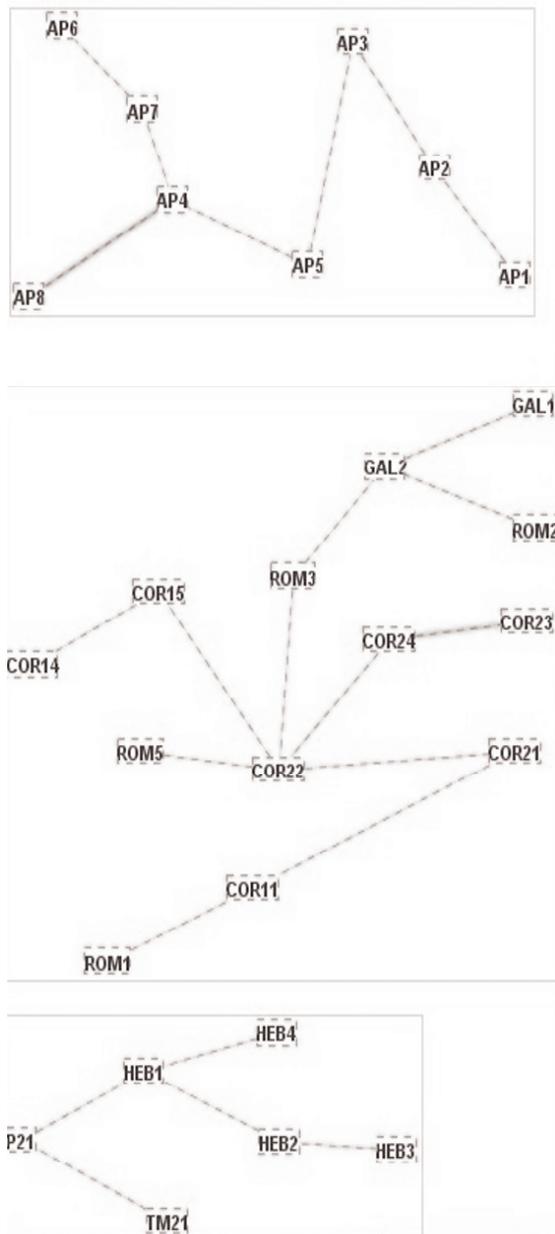


Fig. 11: Tres clusters más asociados al dendrograma.

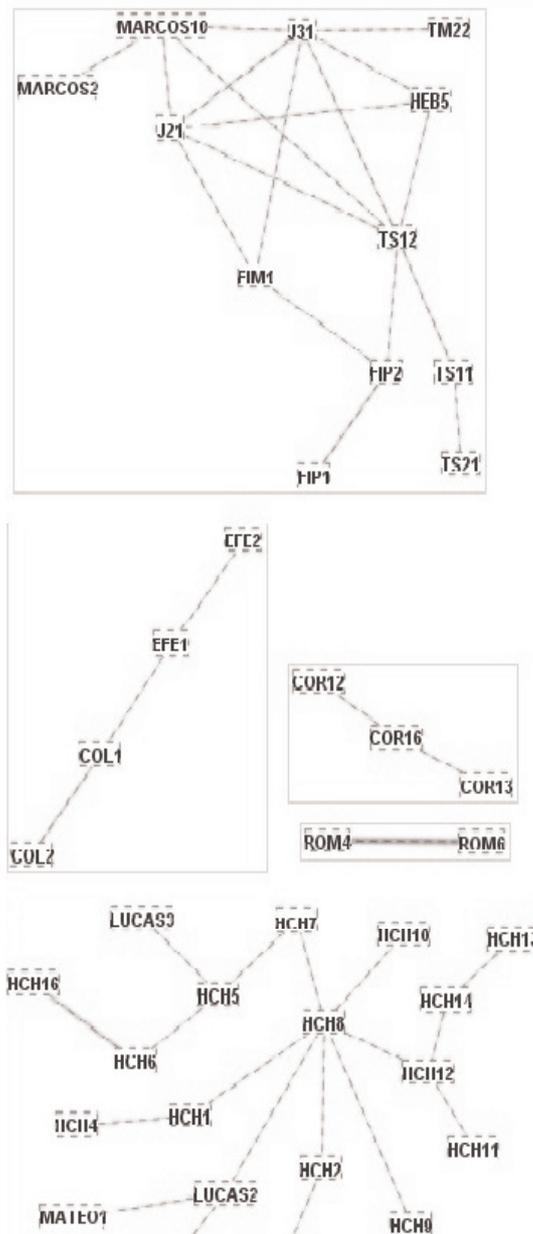


Fig. 12: Cinco clusters más asociados al dendrograma.

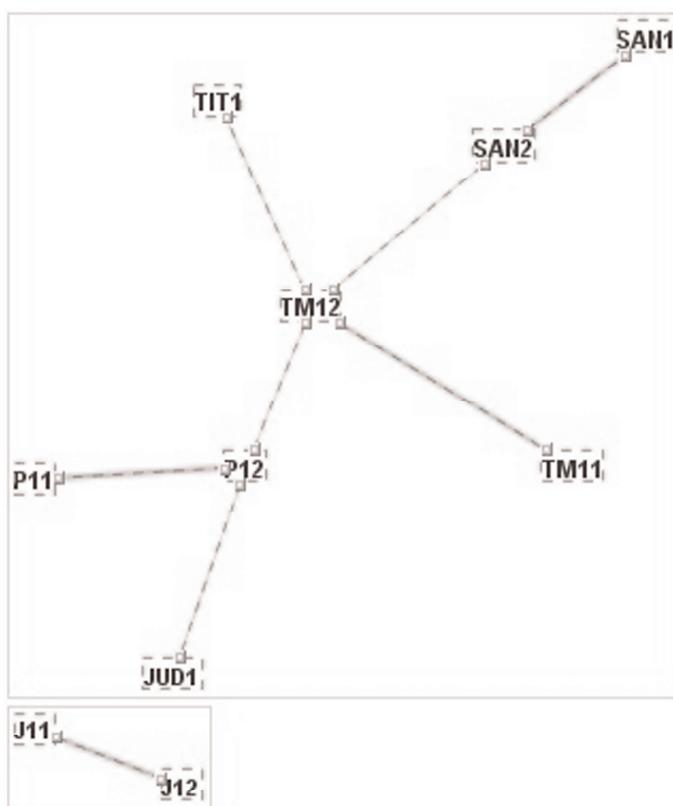


Fig. 13: Dos clusters más asociados al dendrograma.

3. TERCER EXPERIMENTO CON LOS LIBROS DEL NT: ANÁLISIS DE LOS EVANGELIOS DIVIDIDOS EN CAPÍTULOS

En el tercer experimento se dividieron los evangelios en sus respectivos capítulos (un total de 89) y se analizaron con la herramienta AC, obteniendo el dendrograma que muestra la figura 14 y los grafos de las figuras 15 y 16.

- Aparecen buenos ejemplos de casos en que la semejanza de contenido domina sobre la autoría, como en los subgrafos triangulares Mt4-Mc1-Lc4; Mt19-Mc10-Lc18; Mt21-Mc12-Lc20; Mt22-Mc12-Lc20; Mt24-Mc13-Lc21; Mt26-Mc14-Lc22, todo lo cual está de acuerdo con las relaciones cruzadas que suelen señalar muchas traducciones actuales de los evangelios.
- También se observan casos donde la autoría domina, como en casi todo el evangelio de San Juan, con la excepción de la Pasión (Jn18-19); o en el subgrafo Lc1-Lc2-Lc24-Lc4; o en todas las ramas que ligan dos capítulos diferentes del mismo evangelio.



Fig. 14: Dendrograma de los capítulos de los cuatro evangelios.

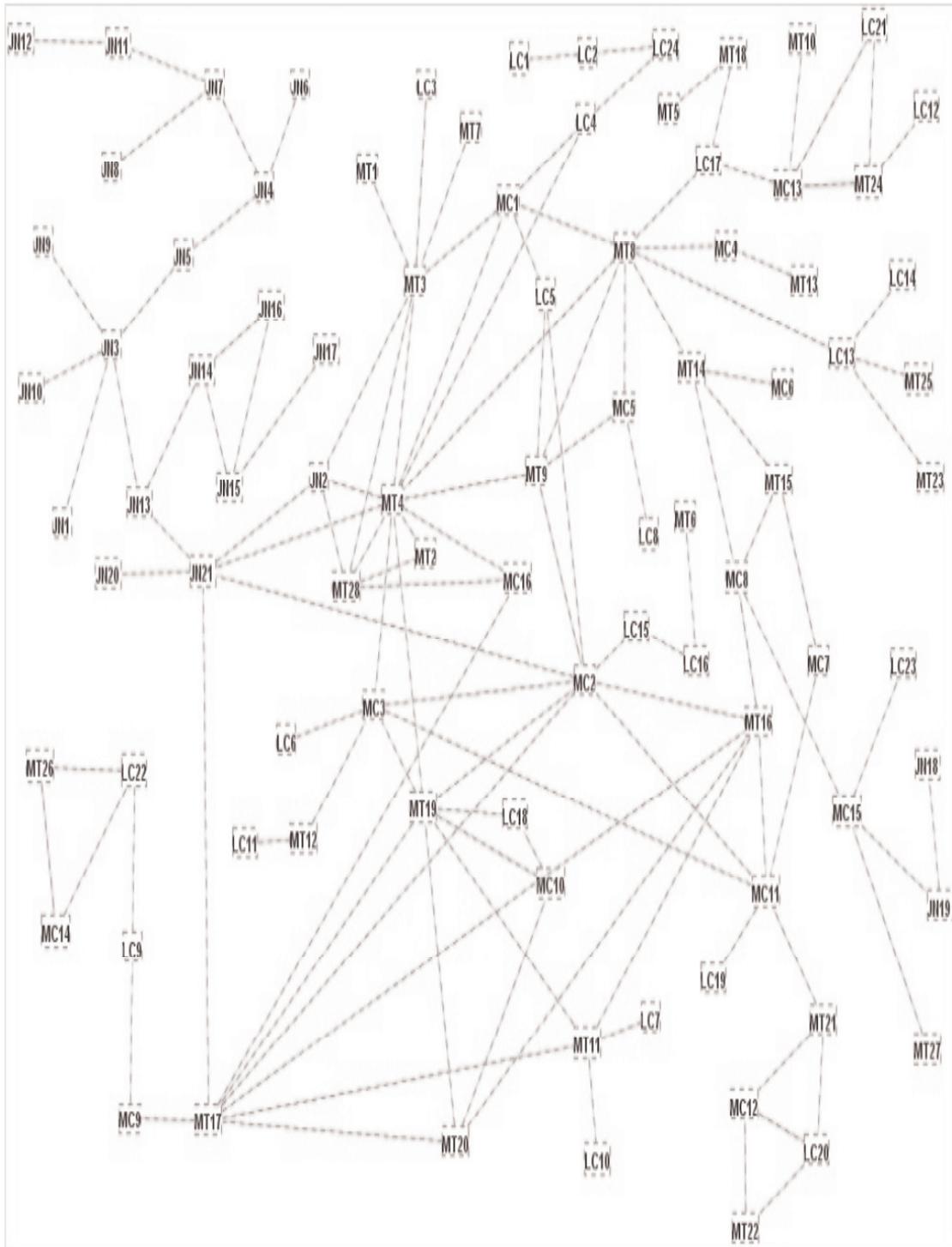


Fig. 15: Grafo de distancia mínima para los capítulos de los evangelios.

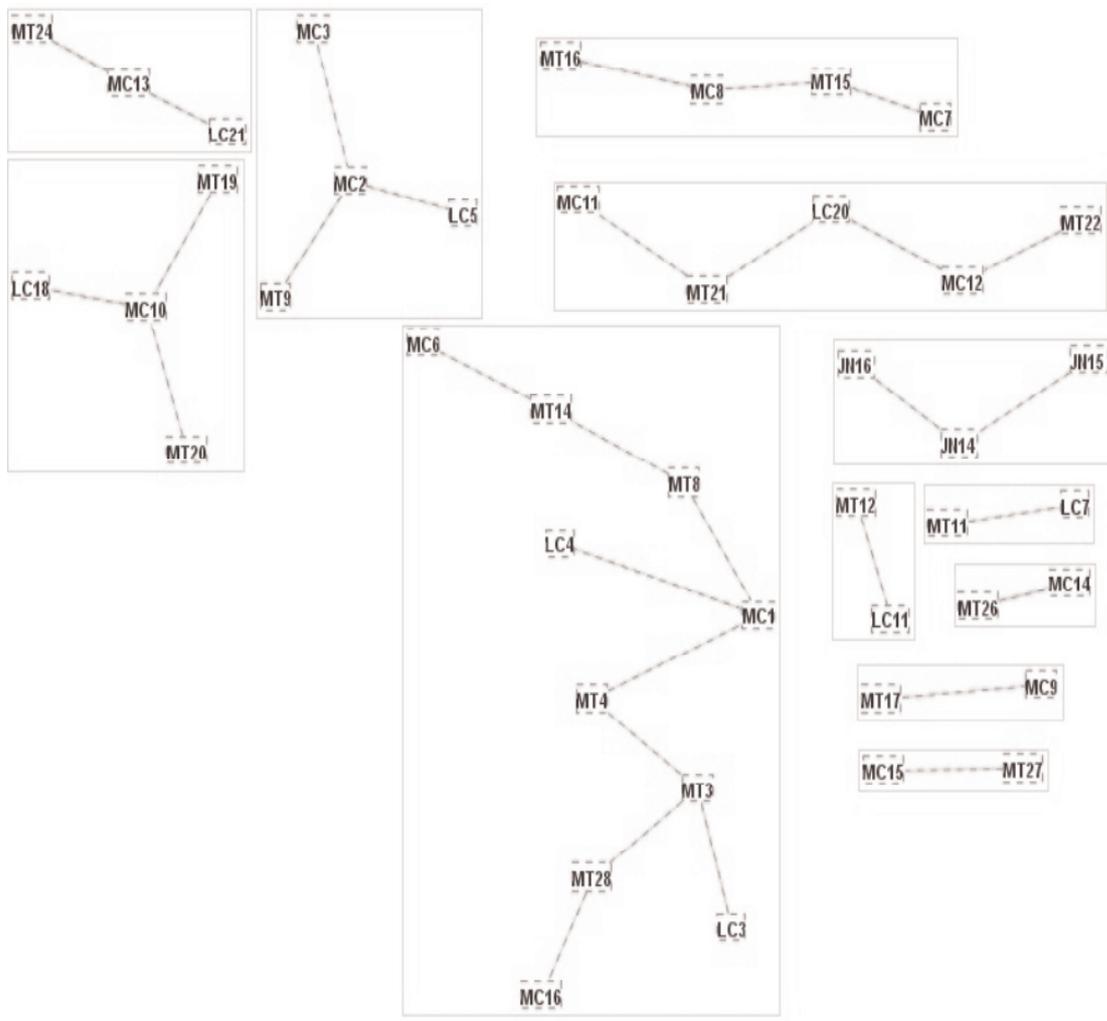


Fig. 16: Grafo de distancia mínima podado a distancia 0,79 para los capítulos de los evangelios.

V. CONCLUSIONES Y TRABAJO FUTURO

De este conjunto de experimentos sería posible extraer las siguientes conclusiones tentativas:

- Cronológicamente, Marcos y Mateo podrían ser los dos primeros evangelios, pues Lucas se habría basado (directa o indirectamente) en ambos. Juan está ligado a los sinópticos principalmente por su descripción de la pasión. Los Hechos están en relación más directa con el evangelio de

Lucas, lo que puede explicarse, como es usual, porque los dos fueron escritos por el mismo autor.

- Parece existir una relación estrecha entre Tesalonicenses 1 y 2, por lo que sería razonable separar 2Ts de las otras dos epístolas Deutero-Paulinas (Colosenses y Efesios), con las que usualmente se asocia, y agruparla con 1Ts, junto con Filipenses y Filemón. Estas cuatro epístolas siempre aparecen conectadas en los experimentos.
- Colosenses y Efesios quedan estrechamente conectadas en todas las pruebas. Se confirma la sospecha de que la primera pudo ser utilizada como fuente para la segunda.
- La primera epístola de San Juan siempre está claramente separada de las otras dos, lo que sugiere que quizá pertenezcan a autores diferentes.
- Se puede llegar a la misma conclusión respecto a las dos epístolas de San Pedro.
- Dependiendo del experimento concreto realizado, Hebreos aparece conectada a diversas epístolas, incluso de autores diferentes. Probablemente tiene su propio autor.

Entre los trabajos futuros que se pueden abordar podemos mencionar los siguientes:

- Analizar con cuidado los resultados descritos en la sección 3.3 para establecer todas las relaciones plausibles de semejanza de contenido o de autoría entre los distintos capítulos de los evangelios, y validarlas con las relaciones cruzadas que señalan las diversas versiones publicadas.
- Realizar nuevos experimentos con los textos del Nuevo Testamento.
- Integrar estos análisis con los realizados con otras herramientas para tratar de separar las tres posibles causas de semejanza entre textos (plagio, influencia o coautoría).