Facultad de Derecho

TÉCNICAS DE ANÁLISIS DE REDES PARA LA DETECCIÓN DE FRAUDE

Autor: Filipp Andrianov
Director: Lucía Barcos Redín

MADRID | Junio 2024

RESUMEN

Este trabajo explora cómo las técnicas de análisis de redes pueden emplearse para investigar y descubrir una red de esquemas de evasión fiscal que involucran a empresas *off-shore*, con un enfoque particular en los Papeles de Panamá. Aprovechando métodos avanzados de análisis de redes, la investigación busca identificar a los actores clave, descubrir conexiones ocultas y trazar las complejas relaciones entre individuos y entidades implicadas en estas actividades. Los Papeles de Panamá, una masiva filtración de documentos financieros, revelaron el uso extensivo de empresas *off-shore* por parte de individuos y corporaciones ricas para evadir impuestos y blanquear el dinero. Este trabajo demuestra cómo el análisis de redes puede visualizar y analizar efectivamente estas redes, permitiendo una comprensión integral de los aspectos estructurales y funcionales de los esquemas de evasión fiscal. Los hallazgos se destinan a subrayar el potencial del análisis de redes para ser utilizado como herramienta para investigaciones forenses y ejecución regulatoria, contribuyendo a estrategias más efectivas para combatir los delitos financieros y mejorar la transparencia en el sistema financiero global.

PALABRAS CLAVE

Análisis de redes, aprendizaje automático, detección de fraude, fraude fiscal, Papeles de Panamá.

ABSTRACT

This dissertation explores how network analysis techniques can be used to investigate and unravel the intricate web of tax evasion schemes involving off-shore companies, with a particular focus on the Panama Papers. By leveraging advanced network analysis methods, the research aims to identify key players, uncover hidden connections, and map out the complex relationships between individuals and entities implicated in these activities. The Panama Papers, a massive leak of financial documents, revealed the extensive use of off-shore companies by wealthy individuals and corporations to evade taxes and launder money. This paper demonstrates how network analysis can effectively visualize and analyze these networks, providing a comprehensive understanding of the structural and functional aspects of tax evasion schemes. The findings are aimed to highlight the potential of network analysis to be used as a

tool for forensic investigations and regulatory enforcement, contributing to more effective strategies to combat financial crimes and enhance transparency in the global financial system.

KEY WORDS

Network analysis, machine learning, fraud detection, fiscal fraud, Panama Papers.

ÍNDICE

LISTADO DE ABREVIATURAS 5

CAPÍTULO 1. INTRODUCCIÓN 6

1.1 CONTEXTUALIZACIÓN Y JUSTIFICACIÓN DEL TRABAJO 6

1.2 OBJETIVOS GENERALES Y ESPECÍFICOS DEL TRABAJO 8

1.3 METODOLOGÍA 10

1.4 ESTRUCTURA DEL TRABAJO 11

CAPÍTULO 2. MARCO TEÓRICO I: APRENDIZAJE AUTOMÁTICO Y ANÁLISIS DE REDES 13

2.1 QUÉ ES EL MACHINE LEARNING 13

2.2 TÉCNICAS DE ANÁLISIS DE REDES 15

CAPÍTULO 3. MARCO TEÓRICO II: DELITOS FISCALES Y LA TRAMA DE LOS PAPELES DE PANAMÁ 18

3.1 NOCIONES BÁSICAS SOBRE LOS DELITOS ECONÓMICOS 18

3.1.1 Consideraciones iniciales 18

3.1.2 Delito fiscal 19

3.1.3 Evasión y elusión fiscal 20

3.1.4 Blanqueo de capitales 22

3.2 LA TRAMA DE LOS PAPELES DE PANAMÁ 23

3.2.1 Breve mención al concepto de paraíso fiscal 23

3.2.2 Contextualización 24

3.2.3 Repercusiones 25

CAPÍTULO 4. ANÁLISIS DE LA RED DE LOS PAPELES DE PANAMÁ 26

4.1 FUENTES DE DATOS EMPLEADOS. PRE-PROCESAMIENTO 26

4.1.1 Descripción de la base de datos 26

4.1.2 Pre-procesamiento de datos 27

4.2 TÉCNICAS EMPLEADAS E INTERPRETACIÓN DE RESULTADOS 28

4.2.1 Medidas de centralidad 28

4.2.2 Componentes (comunidades) de la red 32

4.2.3 Asertividad y homofilia. Comunidades 33

4.2.4 Otras métricas 34

4.2.5 Estudio de vecindad 35

4.3 ESTUDIO PARTICULAR DE LA IMPLICACIÓN DE ENTIDADES ESPAÑOLAS EN LA TRAMA 37

CAPÍTULO 5. CONCLUSIONES GENERALES Y PARTICULARES 41

DECLARACIÓN RESPECTO AL USO DE LA INTELIGENCIA ARTIFICIAL 44

ANEXOS **45**
ANEXO I: Código fuente de R 45
BIBLIOGRAFÍA **56**

LISTADO DE ABREVIATURAS

AEAT: Agencia Estatal de Administración Tributaria

CP: Código Penal

IA: Inteligencia Artificial

ICIJ: *International Consortium of Investigative Journalists*

LGT: Ley General Tributaria

ML: *Machine Learning*

CAPÍTULO 1. INTRODUCCIÓN

1.1 CONTEXTUALIZACIÓN Y JUSTIFICACIÓN DEL TRABAJO

El mundo está caracterizado por la imparable digitalización y la interconexión global, las actividades económicas han experimentado una transformación sin precedentes, dando lugar a un crecimiento nunca antes visto, pero también a nuevas formas de fraude económico en paralelo. El fraude fiscal da lugar a la disminución de los ingresos del Estado y de Administraciones regionales, lo cual conduce a que tengan menos recursos para financiar los servicios públicos que ofrecen. De igual modo, una creciente tendencia en defraudar a las arcas públicas conduce a la pérdida de confianza de los ciudadanos en el sistema y en los poderes públicos

Estos desafíos emergentes requieren soluciones igualmente avanzadas, donde el aprendizaje automático se ofrece como una herramienta crucial para la lucha contra estas actividades delictivas.

La detección y contención de fraude económico ha evolucionado considerablemente en las últimas décadas, desde enfoques manuales reactivos hasta sistemas automatizados proactivos. Sin embargo, tanto la sofisticación como el dinamismo de las estrategias fraudulentas requiere de herramientas que sean capaces de anticipar y neutralizar incluso intentos –o, por lo menos, casos de menor magnitud antes de que se expandan– de fraude. En este contexto, y teniendo en cuenta cómo el *machine learning* ha evolucionado en los últimos años, y cómo ya es capaz de ofrecer soluciones que facilitan el trabajo del ser humano en distintos campos como la medicina, la educación o el marketing, procede considerar la aplicación de sus técnicas en este complejo laberinto del delito económico de fraude fiscal. Y, en concreto, son las técnicas de análisis de redes las que pueden servir a tal fin, puesto que las transacciones económicas pueden ser modeladas como redes complejas, con patrones y links ocultos entre las entidades participantes. En consecuencia, eso permitiría su rastreo efectivo que, en última instancia, podría ayudar a las autoridades a situarse ante una potencial presencia de actividad fraudulenta.

El análisis de redes permite la visualización y cuantificación de la estructura y de la dinámica de las relaciones entre los actores involucrados en el sistema económico. Esta visión

holística hace que comportamientos anómalos y agrupaciones sospechosas, que pasan desapercibidas si se emplean técnicas más tradicionales, puedan ser identificadas de forma eficaz y preventiva. Por añadidura, el análisis de redes se podría incluso combinar con otras técnicas de aprendizaje automático para abundar en el estudio de una red. Estos modelos pueden ser entrenados para identificar con precisión los distintos patrones de fraude no destacables a simple vista en grandes conjuntos de datos.

No obstante, existen ciertos límites y desafíos a la hora de implementar las técnicas concernientes ante un conjunto de datos que recoja un caso de fraude económico. Por ejemplo, la calidad de los datos, la interpretación de los modelos y la necesidad de adaptarse continuamente a nuevas formas de fraude. Asimismo, la ética y la privacidad también pueden ser considerados como obstáculos que deben ser tomados en cuenta al manejar información altamente sensible.

Sentadas las consideraciones anteriores, conviene ahora abordar brevemente la relevancia práctica de llevar a cabo estudios del uso de técnicas de análisis de redes para la detección de fraude económico.

Así, en el marco económico, nótese que, por ejemplo, el Reino Unido estima unas pérdidas que oscilan entre 31 y 48 mil millones de libras esterlinas cada año, esto es, el equivalente de unas pérdidas del 0.5%-5% de su gasto en fraude (*International Public Sector Fraud Forum*, 2020). Australia, por su parte, estima unas pérdidas financieras debido al fraude de entre 5 y 25 mil millones de dólares australianos al año. En lo que refiere a España, los datos del 2021 muestran que Hacienda Pública dejó de ingresar 662 millones de euros sólo por fraude del IVA (De la Cruz, 2023). Como último dato que se pretende aportar en este aspecto, se estima que el 10% del PIB mundial se retiene en paraísos fiscales. Concretamente, países como Estados Unidos han estimado una pérdida anual de recaudación fiscal de unos 30.000-40.000 millones de dólares, debido a las actividades *off-shore* (Gould y Rablen, 2020).

Como se puede apreciar, las pérdidas por fraude fiscal en los distintos países desarrollados adquieren una importancia mayor y, en la práctica, pueden suponer, a simple vista, unas consecuencias devastadoras para sus economías. Por ejemplo, que los poderes públicos no dispongan de fondos suficientes para afrontar el gasto en mejorar y seguir prestando determinados servicios públicos, o que, directamente, y en consecuencia de lo anterior, tengan que repercutir estas pérdidas en los contribuyentes a través del aumento de tipos impositivos para paliar esos *gaps* que se crean por culpa del fraude.

A nivel social, la adopción de tecnologías avanzadas para la detección de fraudes, como el *machine learning* o el análisis de redes, contribuiría a la seguridad y la confianza del ecosistema financiero y a las transacciones digitales. Al detectar y prevenir esta actividad delictiva, se protege a los consumidores de pérdidas económicas y a las empresas de posibles perjuicios en este mismo aspecto. Al mismo tiempo, se contribuye al fortalecimiento de la integridad de los sistemas financieros. Los organismos de investigación y las entidades reguladoras verían en el análisis de redes para la detección de fraudes una solución eficaz contra las cada vez más sofisticadas estrategias que emplean los delincuentes fiscales. La capacidad de aprendizaje y adaptación continua de los modelos del análisis de redes, o del *machine learning*, les serviría para mantenerse actualizados ante la eventual aparición de nuevas formas de fraude con nuevos patrones.

En el panorama penal y procesal, los procesos también se beneficiarían de la eficiencia y la efectividad del análisis de redes. Ello conduciría inevitablemente a la optimización de los recursos humanos de Juzgados y Tribunales, así como del tiempo dedicado a la instrucción de las causas.

En resumen, la necesidad de abordar las vicisitudes del fraude económico, infiriendo, asimismo, las distintas soluciones desde el punto de vista tecnológico al respecto, es evidente. A partir de la misma, las implicaciones prácticas de adoptar técnicas de análisis de redes para la detección de fraude adquieren, por ende, una relevancia todavía mayor, pues permitiría una mayor seguridad económica, así como la efectiva protección del consumidor, eficiencia operativa de las entidades y la eficacia de las autoridades responsables de combatir el fraude. Se trata, en definitiva, de una cuestión estratégica que agilizaría los procesos concernientes y serviría de apoyo a los poderes públicos en la detección y prevención de fraude.

1.2 OBJETIVOS GENERALES Y ESPECÍFICOS DEL TRABAJO

En lo referente a los objetivos generales de este trabajo, se trata, en primera instancia, de investigar el rol de las técnicas de análisis de redes en la detección de fraude. Es un objetivo centrado en comprender cómo pueden ser utilizadas para identificar y analizar patrones de fraude que permanecen ocultas en un conjunto de datos de elevado tamaño, destacando su importancia y aplicabilidad en la detección de actividades fraudulentas.

En segundo lugar, el estudio se ha puesto como objetivo determinar la eficacia de las métricas en cuestión en lo que refiere a la detección de distintos tipos de fraude. En concreto, se trata de estudiar su capacidad para identificar patrones, indicios, o cualesquiera otros detalles necesarios que permitan esclarecer los hechos, relaciones y vínculos que deliberadamente se ocultan por las entidades integrantes de la red de la trama.

Seguidamente, se trata de explorar las limitaciones y desafíos de estas herramientas. En este sentido, se pretende identificar las barreras y obstáculos que enfrentan las técnicas de análisis de redes ante el caso práctico propuesto, incluyendo problemas de precisión, escalabilidad y adaptabilidad a nuevos tipos de fraude.

Por lo que respecta a los objetivos específicos, teniendo en cuenta el caso práctico concreto que será objeto de estudio, se buscará explorar el desempeño de la herramienta de análisis de redes frente a grandes volúmenes de datos. Este proceso permitirá identificar problemas de rendimiento y posibles soluciones al respecto, pero siempre manteniendo la mirada puesta en la optimización y la precisión del análisis. En este sentido, y en conexión con lo propuesto, se estudiará la posibilidad y, en su caso, la procedencia de la estrategia de segmentación de las bases de datos en subconjuntos más manejables (*subsets*, o subgrafos, en el lenguaje correspondiente al campo de análisis de redes), lo cual, en teoría, mejoraría la velocidad y la precisión del análisis, sin comprometer la calidad de los resultados obtenidos. Esta estrategia se plantea como una solución potencial para manejar la complejidad y el tamaño abrumador de los datos, permitiendo un procesamiento más eficiente, pero procurando que, incluso en este caso, los resultados sean concluyentes y representativos, extrapolables al panorama general de la red íntegra original.

A mayor abundamiento, también se trata de identificar patrones específicos dentro de la trama de los Papeles de Panamá, buscando comportamientos o conexiones entre entidades que sugieran prácticas fraudulentas. Este enfoque no sólo apunta a revelar las estrategias de fraude que emplean los integrantes de la red, sino también a entender las dinámicas subyacentes dentro de estos complejos esquemas.

Otro objetivo relevante que se propone es la creación de una red gráfica que ilustre las conexiones entre entidades, personas y otros elementos relevantes, revelando la estructura y las dinámicas de la trama. Este modelo de red busca ofrecer una comprensión más profunda sobre cómo interactúan las partes involucradas, facilitando la identificación de vínculos clave y patrones de comportamiento fraudulentos.

En siguiente lugar, se sugiere un análisis comparativo entre diferentes métricas de análisis de redes para determinar cuál puede revelar información esencial. Este proceso de evaluación permitirá determinar las fortalezas y debilidades de cada técnica que se estudie, pero también identificará cuál proporciona los resultados más precisos y eficientes.

A partir del análisis de redes, se busca descubrir información nueva y obtener *insights* significativos sobre la trama de fraude, lo cual contribuiría a una comprensión más profunda del fenómeno. Este enfoque insiste en la importancia de interpretar los datos de manera que se revelen conclusiones y patrones no evidentes en un primer momento.

Por su lado, también se buscará la posibilidad de extrapolar los hallazgos específicos del estudio de los Papeles de Panamá al ámbito general del fraude, aplicando, por tanto, las lecciones aprendidas y los patrones identificados a la detección de fraude en otros contextos.

Finalmente, se abordará el análisis específico de la participación de entidades españolas en la trama, buscando comprender su rol y las posibles conexiones con el resto de los integrantes de la red. Este análisis se centra en identificar patrones de comportamiento o conexiones que puedan ser relevantes para la comprensión global de la trama y sus implicaciones.

1.3 METODOLOGÍA

La elaboración de este trabajo sigue un enfoque metodológico estructurado, sistemático y detallado, enfocado en aprovechar al máximo los recursos bibliográficos y datos disponibles, así como en la aplicación de técnicas avanzadas de análisis.

El primer paso, una vez concretado el tema central sobre el que versa el trabajo, fue la revisión de literatura esencial relacionada con los temas principales del trabajo: aprendizaje automático, análisis de redes, la trama de los Papeles de Panamá delitos económicos, o el fraude fiscal, entre otros. Este paso es fundamental para proporcionar una base teórica sólida, sobre la cual construir el estudio y aplicar los conocimientos adquiridos en el análisis subsiguiente.

Seguidamente, el proceso se centra en la aplicación práctica de las técnicas en lenguaje de programación R. Este esfuerzo se apoya en una amplia bibliografía para garantizar una comprensión completa de las herramientas y técnicas que se utilizarán.

El siguiente desafío que tenía que ser superado era el referente a la descarga y validación de datos. A tal respecto, era necesario acceder a la base de datos de los *Panama* Papers de la ICIJ, disponible al público en formato .csv. Este paso implica una validación preliminar de los datos, revisándolos meticulosamente para confirmar su fiabilidad y relevancia para el análisis. Una vez hecha esta comprobación, los datos se sometieron al pre-procesamiento, caracterizado, sobre todo, por la integración de varias tablas .csv entre sí para un análisis más profundo, así como por la limpieza de datos y variables irrelevantes o vacíos.

El análisis se realiza principalmente utilizando el lenguaje de programación R, elegido por su amplio soporte para análisis estadístico y de redes. En concreto, se programa a través de la interfaz R-Studio, empleando las últimas versiones estables disponibles tanto de ésta como del lenguaje en sí y de sus paquetes.

Con los datos ya preparados, se procede a la exploración de las técnicas de análisis de redes para examinar las conexiones y patrones dentro de la base de datos de los *Panama Papers*.

Finalmente, el estudio se enfoca en identificar patrones significativos, construir una red detallada de entidades y relaciones, y extraer *insights* valiosos sobre la trama de los *Panama Papers*. Además, buscamos extender estos hallazgos al ámbito general del fraude, incluyendo la implicación de entidades españolas en la trama.

Realizado el estudio práctico con las métricas y herramientas correspondientes, nos dispusimos a plasmar los resultados en este trabajo, realizando las observaciones y conclusiones pertinentes.

1.4 ESTRUCTURA DEL TRABAJO

El presente trabajo consta de un total de cinco capítulos ordenados sistemáticamente. El Capítulo 1 se dedica íntegramente a la introducción del trabajo, pues es aquí donde se realiza una contextualización sobre la situación que se vive en relación con los delitos fiscales llevados a cabo a escala mundial y la necesidad de aprovecharse de las diferentes herramientas de las nuevas tecnologías a fin de detectar y prevenir con mayor brevedad la comisión de tales ilícitos. Asimismo, este Capítulo introductorio también recoge los objetivos que se persigue alcanzar con el estudio realizado en este trabajo. Finalmente, se describe la metodología que se ha

seguido a la hora de realizar el presente estudio, además de la estructura del trabajo, punto éste que estamos tratando en el presente apartado.

Los Capítulos 2 y 3 pretenden darle al lector todo el contexto necesario a fin de comprender el problema a tratar y el estudio que se va a realizar. En este sentido, el Capítulo 2 introduce sucintamente los conceptos relacionados con el análisis de redes y el aprendizaje automático, tan íntimamente ligados que, en ocasiones, incluso se solapan entre sí. Se describen, asimismo, los casos de uso y las distintas métricas que caracterizan aquellas técnicas, con especial énfasis en las que se usarán luego en este trabajo.

Por su parte, el Capítulo 3 nos pone en contexto alejado de lo tecnológico. Trata los delitos económicos, fiscales, y analiza brevemente la trama de los Papeles de Panamá, que es sobre la que se realizará el análisis a modo ejemplificativo para darle contenido a este trabajo.

El Capítulo 4 corresponde al desarrollo práctico del trabajo en sí. En primer lugar, describimos la base de datos con la que vamos a trabajar, así como la depuración y el pre-procesamiento de datos que hemos realizado previamente. Seguidamente, se describe todo el proceso realizado, los análisis hechos, así como los resultados y la interpretación que les otorgamos. Todos los *insights* y conclusiones particulares que el análisis de redes nos ha permitido poner de manifiesto también se subrayan en este capítulo, en línea con los objetivos específicos marcados en el capítulo introductorio.

Finalmente, el Capítulo 5 analiza las conclusiones generales que se han alcanzado, vinculándolas con los objetivos que se han establecidos en el Capítulo 1.

CAPÍTULO 2. MARCO TEÓRICO I: APRENDIZAJE AUTOMÁTICO Y ANÁLISIS DE REDES

2.1 QUÉ ES EL MACHINE LEARNING

El *Machine Learning* (ML), o aprendizaje automático en castellano, se puede definir como el estudio de algoritmos y modelos que los sistemas informáticos emplean a fin de realizar tareas específicas sin estar explícitamente programadas para ello (Mahesh, 2018). De esta definición se puede inferir que las técnicas de ML sirven para ayudar al ser humano a interpretar un conjunto de datos (*dataset*) de un elevado tamaño con mayor facilidad. Las técnicas de aprendizaje automático, como su nombre indica, lo que hacen es ejercitarse en base a los datos recibidos en el pasado, para así poder mejorar su capacidad predictiva en las tareas subsiguientes. Así, pues, la utilidad principal de ML es que puedan aprender y, a continuación, analizar los datos sin estar expresamente programados para una tarea concreta.

El aprendizaje automático está basado en la estadística y probabilidad (Hastie et al., 2009) y optimización (Boyd y Vandenberghe, 2004).

Existen diversas técnicas de aprendizaje de *Machine Leaning*, orientadas a la solución de un problema concreto según su naturaleza. Las dos categorías más destacables son el aprendizaje supervisado y el no supervisado (Sandoval, 2017). Dentro del primero se engloban las técnicas de regresión y clasificación, y se consideran aprendizaje supervisado debido a que el diseñador de la misma proporciona a la máquina tanto las características (las preguntas, o la "X") como las etiquetas (las respuestas, o la "Y") que espera recibir del análisis que va a realizar el ordenador. Por ejemplo, ante un estudio de clasificación, el diseñador le proporcionará a la máquina un conjunto de *dataset* concreto (la "X"), donde figurará el grupo al que pertenece cada una de las observaciones (esto es, la "Y"). Así, el mecanismo se entrenará en base a dicho conjunto de datos y, a partir de este momento, podrá predecir la pertenencia de una nueva observación a un grupo u otro en base a las características y valores que toma la misma en las diferentes variables.

Por otro lado, el aprendizaje no supervisado solamente recibe las características (la "X"), y será la propia máquina la que tenga que definir las etiquetas (la "Y"). A modo ejemplificativo, una técnica de *clustering* (o agrupación), de aprendizaje no supervisado, lo que

busca es crear conjuntos de grupos de observaciones que tengan características similares y que, precisamente por ello, han de pertenecer a una misma categoría. Pero esta vez será la máquina la que tenga que predecir dónde encaja cada observación, puesto que el *dataset* inicial no dispone de esa información.

En este sentido, resulta menester poner de relieve el hecho de que, a lo largo del desarrollo de ML, matemáticos y científicos han ido descubriendo diversas técnicas de este campo de la ciencia, cada uno de los cuales se destina a resolver un problema concreto. Por ejemplo, el método de *k-means* sirve para definir *clusters* (o grupos) de datos o entidades, en línea con lo expuesto *supra*, mientras que el *support vector machine* (SVM) sirve para clasificar de forma binaria (aprendizaje supervisado) las observaciones dadas, pero también funciona como una medida de regresión.

Por añadidura, cabe mencionar que las técnicas del aprendizaje automático no se limitan a esta clasificación binaria que se ha expuesto hasta ahora, sino que existen más categorías. En particular, se suele incluir una tercera: el aprendizaje reforzado (*reinforcement learning*), el cual se caracteriza por tener una retroalimentación (*feedback*) evaluativa, pero no existen señales supervisadas como tal.

La mayoría de los autores coincide en que el aprendizaje reforzado es secuencial y tiene una visión hacia el futuro lejano en cuanto a predicción de resultados, pues tiene en cuenta la recompensa acumulativa a largo plazo (por todo, Yuxi Li, 2018). La máquina aprende de la interacción que realiza con el entorno, percibiendo el estado actual del mismo, y se ejercita para tomar una determinada acción ante una situación concreta a fin de maximizar la señal numérica de recompensa –que sirve de indicador del acierto de la acción elegida– a la que se acaba de aludir. En otras palabras, en base al estado actual que el agente de la técnica de aprendizaje reforzado percibe, va explorando las distintas acciones que puede tomar para cambiar dicho estado, recibiendo señales de refuerzo de vuelta que le describen la transición realizada. Seguidamente, analiza dichas señales para entender cuál ha sido la mejor acción para "*obtener una política deseada*" (Fonseca-Reyna et al., 2018).

Por todo lo anterior, se entiende que las técnicas del *reinforcement learning* guardan estrecha relación con la neurociencia y la psicología (Sutton y Barto, 2018).

2.2 TÉCNICAS DE ANÁLISIS DE REDES

El análisis de redes es un campo de estudio utilizado para estudiar las relaciones e interacciones entre distintos nodos dentro de un sistema de redes complejo en base a la teoría de los grafos (Otte & Rousseau, 2002). Según dicha teoría, un grafo es un conjunto de objetos (llamados vértices o nodos) que están conectados entre sí. Los grafos pueden ser dirigidos o no. En el primer caso, todas las aristas están dirigidas de un vértice a otro, y dicho grafo a veces adopta el nombre dígrafo, o red dirigida. En cambio, un grafo en el que las aristas son bidireccionales es uno no dirigido (por todo, Nykamp, s.f.).

En el ámbito de los delitos de fraude fiscal, el empleo de las técnicas de *network analysis* deviene imprescindible debido a su capacidad de revelación de patrones *a priori* ocultos en el sistema, habilidad esta que resulta de difícil cumplimentación por el ser humano.

El *network analysis* se aprovecha del uso de algoritmos matemáticos y estadísticos la consecución de su fin. Las redes se componen de nodos (entidades individuales que pueden ser personas físicas, jurídicas, cuentas bancarias, o similares) y aristas (los vínculos existentes entre dichos nodos, que se generan por las relaciones que puede haber entre ellos, por ejemplo, por realización de transacciones u otros negocios). Según (Borgatti et al., 2013), el análisis de redes se enfoca en las estructuras de conexión y cómo las mismas influyen en el comportamiento y la dinámica de los nodos.

A su vez, existen diferentes tipos de técnicas de análisis de redes, ente las que se pueden destacar algunas de las siguientes. En primera instancia, las métricas de centralidad, encargados de evaluar la importancia de un nodo dentro de la red. Existen varias medidas de centralidad, como la centralidad de grado (mide el número de conexiones que tiene un nodo), centralidad de intermediación (*betweenness*, que determina cuántas veces un nodo actúa como intermediario en la red) y centralidad de cercanía (*closeness*, que estudia la distancia de un nodo a los integrantes de la red restantes), entre otras. Estas medidas ayudan a identificar nodos clave que pueden ser influyentes o críticos dentro de la red (Freeman, 1978).

En segundo lugar, la clusterización, que tiene como fin la identificación de grupos dentro de la red dada la cercana conexión que puede existir entre los nodos integrantes de cada uno de dichos *clusters*. Las medidas de clusterización, como el coeficiente de Watts y Strogatz (1998), permiten entender la cohesión y la estructura interna de las redes.

Seguidamente, la técnica de detección de comunidades, que es muy similar a la de agrupación de la que se acaba de hablar, con la diferencia de que, en este caso, son técnicas diferentes a los de la clusterización. Destaca, en este sentido, el método de *Louvain*, *Girvan-Newman* (Blondel et al., 2008 y Girvan & Newman, 2002) o el *Walktrap*, que identifica las comunidades mediante desplazamientos aleatorios por el grafo (Pons & Latapy, 2005).

Finalmente, el análisis de camino, que estudia las rutas más cortas y las más frecuentes entre los nodos de una red, y permite sacar conclusiones en relación con la eficiencia de las conexiones existentes en dicha red (Dijkstra, 1959).

Expuestos los caracteres y las distintas técnicas del *network analysis*, se puede pensar en algunos supuestos de aplicaciones prácticas de estos métodos: estudio de redes sociales, estudios de relaciones biológicas (por ejemplo, el ADN), identificación de redes de organizaciones criminales y la subsecuente prevención de comisión de delitos por aquellos y, finalmente, la detección de fraude fiscal en transacciones financieras. Esto último es el caso de los Papeles de Panamá que se estudia en el presente trabajo. Por tanto, conviene ahora desarrollar brevemente algunas ideas acerca del empleo de las técnicas del análisis de redes en la detección de fraude. En este contexto, desentrañar una red tan compleja con interacciones entre millones de nodos no es una tarea fácil, pero por eso mismo la tecnología y, particularmente, el aprendizaje automático ha sido desarrollado: para simplificar tareas mundanas cuya consecución supondría un esfuerzo (y tiempo) muy elevado al ser humano.

El análisis de redes permitiría identificar nodos clave en la red gracias al empleo de medidas de centralidad explicadas anteriormente. Se trata de individuos o entidades que juegan un rol central en la red de fraude. Estos nodos críticos pueden actuar como intermediarios o, directamente, como coordinadores de actividades ilícitas (Papachristos, 2009). Por añadidura, las técnicas de detección de comunidades permitirían detectar grupos relevantes cuyos nodos actúan estrechamente en los esquemas de fraude. El nexo podría estar en las transacciones frecuentes entre los integrantes, por ejemplo, o en patrones de comportamiento conjunto realizado por varios de ellos a la vez.

Por su lado, el análisis de caminos puede identificar patrones de transacciones sospechosas y anormales, por ejemplo, si se está ante la presencia de las llamadas cuentas *off-shore* o movimientos de las mismas cantidades realizados de manera circular para esconderlas (Ferrara et al., 2014). Finalmente, no se puede dejar atrás la visualización simplificada de las

redes, que permite al ojo humano entender las relaciones escondidas que pueden existir en la red (Wasserman & Faust, 1994).

CAPÍTULO 3. MARCO TEÓRICO II: DELITOS FISCALES Y LA TRAMA DE LOS PAPELES DE PANAMÁ

3.1 NOCIONES BÁSICAS SOBRE LOS DELITOS ECONÓMICOS

3.1.1 Consideraciones iniciales

Los agentes involucrados en la trama de los Papeles de Panamá pretenden la comisión de delitos fiscales a fin de obtener ventajas económicas para las empresas e individuos concernientes. Por ende, en este capítulo resulta menester detenerse en los conceptos de delitos económicos y fiscales.

En el contexto de una economía globalizada y la creciente complejidad de las actividades financieras, la comprensión de los delitos económicos se ha vuelto esencial para académicos, profesionales del derecho, reguladores y la ciudadanía en su conjunto. Los delitos económicos, también conocidos como delitos de cuello blanco, abarcan una amplia gama de conductas ilícitas que tienen en común el uso indebido de la confianza y la manipulación de sistemas financieros y comerciales para obtener beneficios económicos ilegítimos.

Los delitos económicos pueden definirse como "*aquellos actos ilegales cometidos por un individuo o una organización, caracterizados por el engaño y la violación de la confianza, cuyo principal objetivo es obtener una ventaja financiera o evitar la pérdida económica mediante actividades no éticas, ilegales o inmorales*" (Gottschalk, 2010).

Los delitos económicos tienen un impacto profundo y multifacético en la sociedad. A nivel macroeconómico, pueden desestabilizar los mercados financieros, mermar la confianza en las instituciones y reducir la recaudación fiscal, lo cual tendría sus lógicas consecuencias: aumento de la presión fiscal, regulación más intensa e intervencionista de los mercados, más prohibiciones, etc. Son numerosos los estudios que han evidenciado que los delitos económicos pueden provocar estos efectos devastadores en las economías nacionales y globales. Por ejemplo, se subraya que el blanqueo de capitales y otros delitos financieros pueden conducir a la inevitable inestabilidad en los mercados y aumentar el riesgo en el sector financiero (Levi y Reuter, 2006).

A nivel microeconómico, estos delitos pueden afectar a familias consumidoras y empresas, sobre todo causando pérdidas significativas. La victimización corporativa y el fraude empresarial pueden tener consecuencias severas para las empresas afectadas, incluidas pérdidas económicas directas, costes legales y una disminución en la confianza de los inversores, lo cual reduce la financiación de la empresa y provoca, a su vez, más desventajas para la entidad. Según Button, Lewis y Tapley (2014), los efectos del fraude corporativo pueden suponer también una erosión de la confianza entre las partes interesadas y un deterioro de la moral interna dentro de las organizaciones.

La relevancia de estos delitos en el entorno actual se ve acentuada por la estrecha conexión existente entre las economías mundiales. Las acciones ilícitas en un país con cierto grado de participación e intervención en el mundo (por ejemplo, EE. UU., Alemania, China…) tendrán repercusiones a nivel internacional. Ello acentúa aún más la necesidad de una coordinación global en la lucha contra los delitos económicos. La globalización y el avance tecnológico han facilitado la comisión de estas ilegalidades a una escala sin precedentes, lo que exige respuestas coordinadas y estrategias de prevención innovadoras (Sikka y Willmott, 2010).

3.1.2 Delito fiscal

El delito fiscal consiste en defraudar a la Hacienda Pública por un montante superior al fijado por la Ley. Se defrauda, fundamentalmente, omitiendo ingresos y dejando de abonar la cuota fiscal correspondiente. Lo anterior describe el elemento subjetivo –esto es, la actuación que tienda deliberadamente a evitar el abono de las cantidades que el sujeto debe a Hacienda–, que ha de estar presente en la conducta del sujeto para que pueda ser considerado delito, aparte de que se trate de cantidades que marque la Ley, requisito referido *supra*. Se trata de una conducta punitiva de gravedad superior que, por ende, reside en el ámbito penal por encima del incumplimiento de obligaciones fiscales menores, sancionables en el Ordenamiento jurídico español en el orden administrativo.

El artículo 305 del Código Penal (CP) dispone que aquel sujeto que, por acción u omisión, defraude a la Hacienda Pública mediante (i) elusión del pago de tributos, (ii) obtención indebida de devoluciones, (iii) disfrute de beneficios fiscales indebidos, será castigado con pena de prisión que oscila entre uno y cinco años, y con multa del tanto al séxtuplo de la cuantía

mínima defraudada que marca el precepto, que es de 120.000€ (ciento veinte mil euros). Por tanto, cualquier fraude cuyo montante no supere dicha cifra supondrá la comisión de una infracción administrativa y quedará extramuro al derecho penal.

Por su lado, el artículo 305 bis CP establece penas agravadas (con prisión de entre dos y seis años) en los casos de cuantías más elevadas (concretamente, las que excedan de 600.000€, seiscientos mil euros), así como si el delito se ha cometido por una organización criminal, o si se han empleado intermediarios (por ejemplo, sociedades o instrumentos ubicados en paraísos fiscales, o a otras personas físicas) a fin de dificultar la determinación de la identidad del responsable, de la cuantía defraudada o del patrimonio total de aquel.

Para combatir las distintas formas de delitos fiscales, los poderes públicos han de disponer de los datos veraces de los ingresos y del patrimonio de los contribuyentes, así como los vínculos que pueden existir entre los ciudadanos, puesto que en estas actividades ilegales se emplean todos los esfuerzos para ocultarlos. Así, por ejemplo, una vez descubierta la relación que existe entre el contribuyente y sus activos, se puede determinar si los mismos han sido colocados en un paraíso fiscal. Del mismo modo, una persona que intenta aumentar su patrimonio a través de estas actividades ilícitas, típicamente utiliza técnicas de fraude para realizar transacciones que carecen de sentido desde el punto de vista financiero, o directamente oculta sus beneficios o los vínculos que pueda tener con propiedades adquiridas de manera artificiosa o ilegal. Asimismo, se crean complejas redes de empresas a fin de ocultar a quién pertenece en realidad un determinado activo. La misión de los poderes públicos consiste, a raíz de todo lo anterior, de desenlazar estos vínculos y hallar la información que se oculta detrás de estos complejos esquemas (González y Mateos, 2018). En este sentido, tradicionalmente se solía cruzar y verificar la información que proporcionaba el contribuyente con la que compartían terceros para identificar posibles ingresos o activos no declarados (González y Mateos, 2018).

3.1.3 Evasión y elusión fiscal

Los fenómenos de evasión y elusión fiscal tienen un impacto significativo en la recaudación y equidad fiscal de cualquier sistema tributario. Estos conceptos tienen diferencias fundamentales tanto en su naturaleza como en sus implicaciones legales, aunque, desafortunadamente, se utilizan indistintamente en el lenguaje cotidiano.

Por un lado, la evasión fiscal es se refiere al delito fiscal referente al impago de los impuestos debidos por medio de la omisión u ocultación de ingresos, deducción de gastos falsos, o manipulación de la contabilidad, al y como ya se ha comentado en el apartado anterior. En términos generales, la evasión fiscal implica una vulneración deliberada de la norma tributaria, y se tipifica como delito en la legislación interna de los países. La evasión fiscal implica ocultación y engaño, constituyendo un auténtico delito contra la Hacienda Pública, y en línea con lo ya expuesto, conlleva implicaciones penales y administrativas serias (Slemrod y Yitzhaki, 2002).

En lo que a la elusión fiscal se refiere, se trata de una utilización (o, si se quiere, maquinación) de estructuras y estrategias fiscales que tiene como lógico corolario la reducción de la carga tributaria, que se realiza, no obstante, en el marco de los límites establecidos por la Ley. Por ello, la elusión no implica una vulneración directa de las normas fiscales, sino más bien el aprovechamiento de lagunas legales y ambigüedades en la legislación tributaria para disminuir la carga tributaria que realmente corresponde al contribuyente. En otras palabras, la elusión fiscal implica una planificación fiscal agresiva en la que los contribuyentes buscan beneficios fiscales que, aunque técnicamente legales, pueden contravenir el espíritu de la norma (Avi-Yonah, 2006). Por ende, son merecedores de un reproche ético desde la sociedad.

Sentado lo anterior, procede analizar ahora la regulación legal de los anteriores fenómenos. En este sentido, y como se ha expuesto antes, la evasión fiscal constituye un delito. Las penas por fraude fiscal están establecidas en los artículos 305 y 305 bis del Código Penal a los que ya se ha hecho alusión anteriormente. Además, la Ley General Tributaria (LGT, en lo que sigue) otorga a la Agencia Tributaria (en adelante, AEAT) competencias en materia de investigación, auditoria, así como potestad sancionadora en el marco de la evasión fiscal.

Por lo que respecta la elusión fiscal, la regulación es más primitiva dado que, como ya se ha dicho, no constituye un delito. No obstante, la LGT establece pautas específicas para prevenir la manipulación, principalmente mediante el concepto de "fraude legal" y "abuso legal". Concretamente, el artículo 13 LGT permite que las autoridades fiscales recalifiquen las operaciones que, aunque sean legales formalmente, se consideren realizadas con la intención de eludir la aplicación de la norma tributaria correspondiente al negocio realmente pretendido por los contribuyentes. Asimismo, también prohíbe la simulación o el fraude de Ley, y las autoridades tratarán las operaciones financieras llevadas a cabo acorde a su verdadera naturaleza, en atención a lo que efectivamente se haya realizado (artículo 15 LGT).

Asimismo, la Directiva 2016/1164 de la Unión Europea es de aplicación a las sociedades, no individuos, y establece normas contra prácticas de elusión fiscal que distorsionan el funcionamiento del mercado interior, previendo un catálogo de medidas como la limitación de la deducibilidad de intereses, normas sobre sociedades extranjeras controladas, y disposiciones contra asimetrías híbridos –estás últimas imponen a la sociedad la obligación de deducirse un gasto una sola vez por el mismo hecho en caso de que se dé una doble deducción en varios territorios, *ex* artículo 9.1 de la Directiva–.

3.1.4 Blanqueo de capitales

El artículo 1.2 de la Ley 10/2010, de 28 de abril, de prevención del blanqueo de capitales y de la financiación del terrorismo, recoge las actividades que tendrán la consideración de blanqueo de capitales, destacando, entre otras, la conversión de bienes provenientes de actividades delictivas para encubrir tal origen ilícito, o la adquisición de bienes procedentes de actividades ilegales.

Para proporcionar más contexto, el precepto añade que *"se entenderá por bienes procedentes de una actividad delictiva todo tipo de activos cuya adquisición o posesión tenga su origen en un delito, […], así como los documentos o instrumentos jurídicos […], que acrediten la propiedad de dichos activos o un derecho sobre los mismos, con inclusión de la cuota defraudada en el caso de los delitos contra la Hacienda Pública".*

Son dos las conclusiones que se pueden inferir del precepto aducido. En primera instancia, que el blanqueo de capitales hace referencia al proceso de disimulación de los orígenes ilícitos de fondos obtenidos a través de actividades delictivas, a fin de que aparenten legales de cara al público o al mercado. En segunda instancia, se puede deducir que son, fundamentalmente, tres las fases de este proceso: (i) inserción de los fondos ilícitos en el sistema financiero, (ii) realización de movimientos de fondos por medio de transacciones financieras que resultan complicadas de ser rastreadas, (iii) reintroducción de los fondos ya "lavados" en el mercado o en la economía. Es precisamente el concepto "lavados" el que hace referencia a la aparente licitud de estos fondos después de haber llevado a cabo los movimientos necesarios para ello en la segunda etapa descrita.

Nótese que el blanqueo de capitales no constituye un delito *per se*, sino que, para ello, hay que ponerlo en contexto con lo dispuesto en el Código Penal ya comentado con anterioridad en materia de delitos fiscales. En este sentido, hay que fijarse en el artículo 301 CP, cuyo apartado primero empieza diciendo: "*El que adquiera, posea, utilice, convierta, o transmita bienes, sabiendo que éstos tienen su origen en una actividad delictiva, [...]*". Es decir, que la actividad ilícita del blanqueo de capitales constituirá delito siempre y cuando recaigan sobre bienes que procedan de alguna actividad delictiva. Dicho de otro modo, habrá delito de blanqueo de capitales cuando haya un delito fiscal previo.

La anterior conclusión sobre la comisión de un delito fiscal previo constituye uno de los debates más ardidos en el ámbito doctrinal. Parte de la doctrina considera que, en efecto, el delito fiscal opera como un antecedente del blanqueo de capitales, mientras que, de otro lado, se sostiene que de ningún modo puede haber blanqueo sobre la comisión de un delito fiscal previo. Esta segunda postura utiliza el argumento de que el defraudador realmente no obtiene nada del delito fiscal, puesto que los bienes que son objeto de la cuota tributaria ya se hallan en su patrimonio. Esto es, que el delito fiscal no origina nada que no existiera antes en el haber del defraudador. Pero la Ley del 2010 parece despejar las dudas. Asimismo, la Fiscalía, la AEAT y el Tribunal Supremo siguen esta misma línea de considerar el delito fiscal como antecedente al blanqueo de capitales (Blanco Cordero, 2011).

3.2 LA TRAMA DE LOS PAPELES DE PANAMÁ

3.2.1 Breve mención al concepto de paraíso fiscal

El acudir a paraísos fiscales puede llegar a suponer una alternativa de corrupción que refuerza la fuga de capitales y la evasión fiscal puesto que los propietarios de las entidades constituidas de forma "extraterritorial" no declaran sus ingresos a sus respectivas autoridades fiscales en el territorio en el que realmente se encuentran (Domínguez et al., 2019). Un paraíso fiscal se caracteriza por tener bajos tipos impositivos. Y una colocación *off-shore* hace referencia a la instalación de las empresas en dichos paraísos fiscales.

Asimismo, se aprovechan de la confidencialidad y privacidad de la información y las cuentas contables (o la falta de convenios relativos a cesión de información) para acumular capital financiero sin residir en el país extraterritorial (Chfonseristensen, 2011). Así, resulta más difícil la identificación de los verdaderos propietarios de estas entidades, que designan a otros individuos que actúan en el tráfico como directores de las mismas cuando, en realidad, no lo son (Jalan y Vaidyanathan, 2017).

Uno de los principales caracteres de los paraísos fiscales es que normalmente son países pequeños, con una población inferior a 1 millón de habitantes. Así, en torno al 15% de los países son paraísos fiscales (Dharmapala y Hines, 2009a). Según la información recopilada de *Fortune 500* en 2014, 358 empresas (71,6%) tenían, como mínimo, 7.622 filiales en paraísos fiscales (Akamah et al., 2016).

Algunas de las observaciones que se han hecho son como siguen. En primer lugar, según Aziz y Thornton (2015), las empresas privadas se benefician de una mayor reducción fiscal gracias a los paraísos fiscales en comparación con las empresas públicas. Seguidamente, el instalarse en el *off-shore* supone una minoración importante de las bases imponibles de las empresas en los países que no son considerados paraísos, ya que la cantidad recaudada disminuye (Bovi y Cerqueti, 2014). Asimismo, los paraísos fiscales promueven la desmoralización de las empresas honestas, que pierden la fe en el sistema (Domínguez et al., 2019). Finalmente, la evasión fiscal en los paraísos provoca un aumento en los costes de regulación y control debido al necesario refuerzo por parte de las autoridades de países no considerados paraísos a la hora de perseguir y prevenir estas prácticas (Domínguez et al., 2019).

3.2.2 Contextualización

Los Papeles de Panamá hacen referencia a una filtración de unos 11,5 millones de documentos, destapada por el *International Consortium of Investigative Journalists* (en lo que sigue, *ICIJ*). Se trata de archivos de naturaleza tanto financiera como legal confidenciales correspondientes al despacho de abogados panameño *Mossack Fonseca*, que se convirtió en una de las mayores entidades proveedoras de servicios financieros *off-shore* en todo el mundo (Joaristi et al., 2019). Los documentos revelados incluyen información sobre los negocios *off-shore* ilícitos llevados a cabo por empresas, particulares y personalidades políticas: Ministros,

Jefes de Gobierno y Estado de varios países, o diputados, entre otros. Se pueden destacar, entre otros, el artista Jackie Chan, el futbolista Lionel Messi, el presidente ruso Vladímir Putin, o el expresidente de Ucrania Petro Poroshenko. Sin olvidarnos de las grandes empresas como son el Santander, BBVA, UBS, o Société General, entre otros (Obermaier et al., 2016).

La red es tan extensa que abarca en torno a 200 países conectados a estructuras *off-shore*. Detrás de este escándalo fraudulento se han detectado operaciones ilegales de blanqueo de dinero, evasión y fraude fiscales de los que se ha hablado *supra* (Domínguez et al., 2019). Con el paso del tiempo, Panamá empezó a convertirse en un paraíso fiscal al que acudían un creciente número de empresas e individuos con fines delictivos, dejando de percibirse este como paraíso fiscal seguro o fiable (Dharmapala y Hines, 2009).

3.2.3 Repercusiones

El economista Gabriel Zucman calculó en (2015), que en torno al 8% de la riqueza financiera mundial está oculta en los paraísos fiscales. También estimó que las pérdidas fiscales ascienden a unos 200.000 millones de dólares al año.

La revelación de los datos de la trama condujo a la dimisión del primer ministro de Islandia, una presión social sin precedentes ejercida hacia el ex-primer ministro del Reino Unido, David Cameron que, a pesar de todo, no dimitió (The Objective, 2016).

En España, concretamente, la repercusión más significativa del escándalo fue, sin duda, la dimisión de José Manuel Soria, el entonces ministro de Industria, Energía y Turismo, después de no haber podido aclarar a los medios y a la ciudadanía su situación y relación con la trama. Concretamente, presentó su dimisión el 15 de abril de 2016 (El País, 2016).

Asimismo, el Gobierno español anunció el 5 de abril de 2016, que la AEAT investigaría las cuentas opacas abiertas en Panamá de oficio (EFE, 2015). Concretamente, se estima que La Agencia Tributaria investigó a unos 244 contribuyentes, lo que resultó en la recaudación de 142 millones de euros en impuestos no declarados previamente. Por añadidura, se iniciaron más de 100 casos penales como consecuencia de las revelaciones conseguidas con dichas investigaciones (La Sexta, 2021).

CAPÍTULO 4. ANÁLISIS DE LA RED DE LOS PAPELES DE PANAMÁ

4.1 FUENTES DE DATOS EMPLEADOS. PRE-PROCESAMIENTO

4.1.1 Descripción de la base de datos

El entramado se construye a partir de las relaciones que aparecen en la base de datos *ICIJ Offshoring and Panama Papers* (*Offshore Leaks Database*, 2018) y permite detectar patrones de asociación entre los actores que, en términos de técnicas de análisis de redes, se llamarían nodos. Nótese que la base de datos recoge no sólo a los agentes involucrados en la trama de los Papeles de Panamá, sino también de otras –por ejemplo, Papeles de Pandora– que ha investigado y destapado el consorcio.

La red, a su vez, se representa según los países y regiones geográficas concernientes, lo cual permite a los investigadores descubrir la estructura de las conexiones mundiales de deslocalización para las principales naciones en términos de ocurrencia en la base de datos. Asimismo, la base de datos describe los vínculos existentes entre los actores principales, así como la naturaleza de dichos actores (Domínguez -et al., 2019).

La base de datos entera se divide a su vez en las siguientes tablas: (i) *Entities*, (ii) *Intermediaries*, (iii) *Officers*, (iv) *Addresses*, (v) *Relationships (Edges)*. La tabla *Entities* recoge a las 213.634 empresas creadas en paraísos fiscales integrantes de la trama de Panamá. Los intermediarios son los despachos (o cualquier otro tipo de interviniente) que ofrecen servicios de *off-shoring*, y en la base de datos figuran concretamente 14.110 los que se han visto involucrados en los Papeles de Panamá. Los *Officers*, por su lado, son personas físicas o jurídicas que desempeñan un rol destacado en la entidad *off-shore*. Pueden ser, por ejemplo, accionistas mayoritarios, consejeros, directores, o cualquier otro cargo similar. Hay 238.402 agentes involucrados en esta trama en particular. Finalmente, la tabla correspondiente a los *Relationships*, y contiene los vínculos existentes entre los nodos de las tablas anteriormente aducidas. Estos vínculos pueden deberse a varios motivos, a saber:

- Si un nodo es agente (*officer*) de otro, puede deberse a que aquel es el director, accionista, secretario o beneficiario del nodo receptor (*entity*). Estos datos los

extrajo el ICIJ y no figuran en la base de datos, pero un ejemplo de una conclusión como esta podría ser un documento de la escritura de constitución de la sociedad (o, sus estatutos) en el que figura tanto la entidad como su agente (en este caso, podría ser el accionista mayoritario) (Hunter & Lyon, 2016).

- Los intermediarios juegan un papel esencial en la trama ya que son los que crean las estructuras *off-shore*, asesorando a las entidades que se someten a esta deslocalización. Por tanto, puede tratarse de bufetes de abogados que asesoran estas operaciones, o bancos que financian las constituciones societarias, entre otros (ICIJ, s.f.).

Respecto de lo anterior, se utiliza la información de la tabla *relationships* (*rel_type*) para obtener la vinculación entre los nodos que figuran en la variable *node_id_start* y *node_id_end*.

4.1.2 Pre-procesamiento de datos

Previamente a la realización del análisis pertinente, hemos tenido que depurar los datos de la siguiente manera. En primera instancia, se han filtrado únicamente aquellas observaciones correspondientes a los '*Panama Papers*' (variable de *sourceID* en las tablas de *entities, officers, intermediarias* y *relationships*), recortando significativamente las bases de datos iniciales.

Seguidamente, nos hemos centrado en la tabla *relationships*, que es la que recoge los vínculos entre los nodos integrantes de la red. Basándonos en las columnas *node_id_start* y *node_id_end*, hemos extraído del resto de las tablas los nombres correspondientes a los identificadores para trabajar directamente con ellos. Lo mismo se ha hecho con la variable *countries* para cada uno de ellos, guardándolos en las variables *start_countries* y *end_countries*. Finalmente, se han descartado variables innecesarias en la tabla: *sourceID, status, start_date, end_date*, manteniendo una variable tan esencial como es *rel_type*, que representa el tipo de vínculo que guardan los dos nodos entre sí. Como consecuencia de todo lo anterior, hemos pasado de 3.336.913 observaciones en esta tabla a tan sólo 674.102.

En este punto, ya se procede a la creación del grafo no dirigido. Para permitir un mejor análisis que no sobrecargue excesivamente la máquina, se eliminan del grafo todas aquellas conexiones que tengan un grado de centralidad inferior a 5.

4.2 TÉCNICAS EMPLEADAS E INTERPRETACIÓN DE RESULTADOS

A modo de apunte breve, las definiciones o descripciones de las distintas técnicas (funciones de R) empleadas a continuación se han extraído de la página web de *igraph* (Csardi & Neputz, 2006). Se desarrollan con observaciones propias. El paquete *igraph* en cuestión es el que permite realizar análisis de redes con los grafos.

4.2.1 Medidas de centralidad

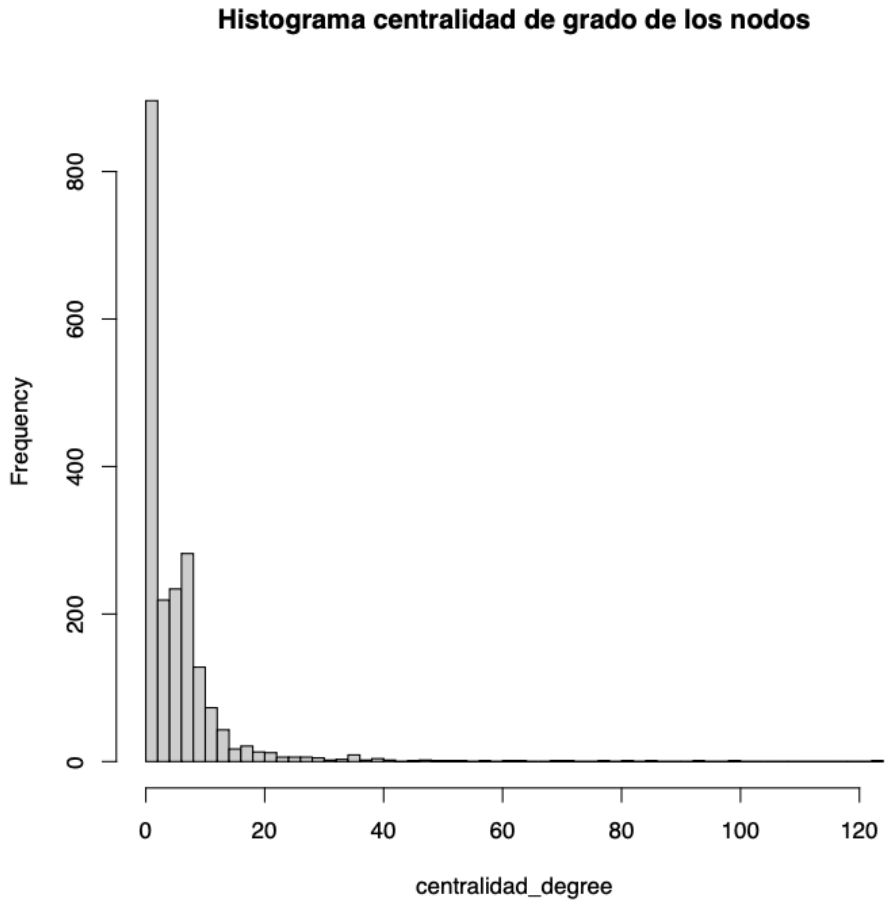
En primera instancia, se han empleado las medidas de centralidad sobre un subgrafo reducido de la red principal para estudiar la importancia de los nodos. Para ello, se han ordenado los nodos según su grado de centralidad de manera descendiente para quedarse con los primeros 2.000 más significativos, que se guardan en una variable que luego sirve de entrada para un subgrafo, que ha sido objeto de aplicación de las medidas de centralidad.

Centralidad de grado (*degree*). Esta métrica calcula el número de arcos que tenga cada nodo, esto es, conexiones directas con otros. Por tanto, esto significa que los nodos con más vértices serán los más influyentes en la red.

Pues bien, estos han resultado ser *OrbusNeich Medical Company, HCP Africa Limited, INGELSA LTD., ValuAccess Asia Limited* y *BHP Billiton (UK) Limited*. Curiosamente, el agente de todas estas entidades mencionadas es el despacho Mossack, y todas estaban radicadas en las Islas Vírgenes Británicas.

A continuación, se visualiza en la Figura 1 un histograma que representa el número total de nodos (frecuencia) por cada grado de centralidad en el subgrafo elaborado para realizar el análisis. Hay que tener en cuenta que existen nodos ("*hubs*") que tienen más conexiones de las que figuran en el eje equis del gráfico:

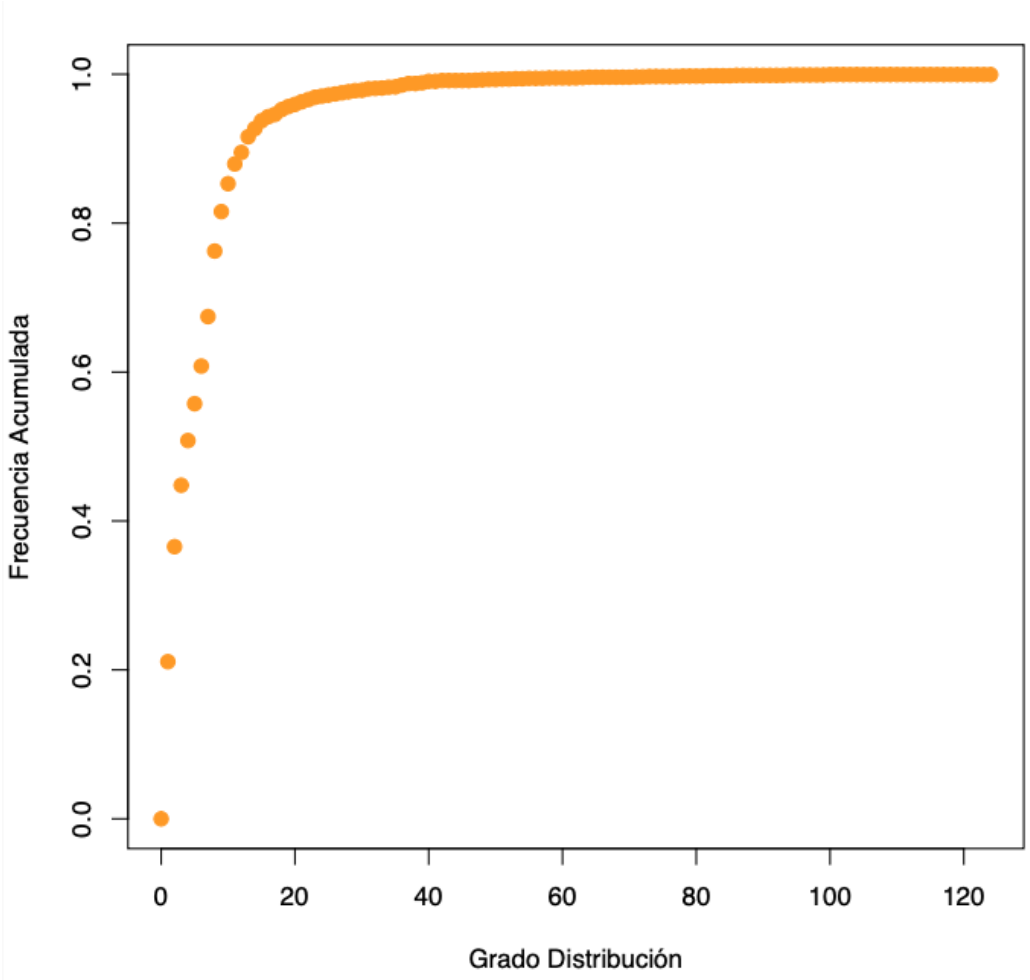
Figura 1. Histograma representando la frecuencia de la centralidad de grado de los nodos



Fuente: elaboración propia

Por añadidura, en la Figura 2 se muestra un gráfico de distribución de frecuencia acumulada del grado del componente principal que se estudia. En línea con lo ya visto anteriormente, conviene puntualizar que hay nodos que tienen más de las 120 conexiones que figura en el extremo derecho del eje equis. Como la curva se eleva rápidamente al inicio, ello indica que hay un número importante de nodos que tienen pocas vinculaciones con otros miembros de la red. Dado que luego se nivela la curva, aproximándose al 1,0 de frecuencia acumulada lentamente, se confirma lo antes expuesto, pero en sentido contrario: hay pocos nodos que tengan un grado elevado. De igual forma, se entiende que, alrededor del 90% de los nodos tienen menos de 20 vecinos:

Figura 2. Gráfico de distribución de frecuencia acumulada de grado de los nodos



Fuente: elaboración propia

Centralidad de cercanía (*closeness*). En otro orden de cosas, esta métrica calcula la media de la distancia de los caminos cortos que van de un nodo a sus respectivas conexiones, y permite ver cómo de rápido un nodo puede acceder al resto de la red. Destacan, entre otras, *Holding Facilities, Inc., METAXAS, SPYRO A., MCG S.A.* (que es, por ejemplo, un intermediario que tiene más de 200 conexiones en toda la red), *DYNA MANAGEMENT INC.,* y *MOSSFON TRUST CORPORATION*, adquiriendo esta última entidad alta significancia en la red, puesto que 'Mossfon' es la abreviación de *Mossack Fonseca*, intermediario pilar de la trama. Es un intermediario constituido en Panamá, que tiene más de 260 conexiones en toda la red. Todas las anteriores tienen un grado de cercanía de 1.

A modo de apunte en relación con lo anterior, si bien el número de conexiones puede parecer bajo, hay que tener en cuenta que Mossack, a su vez, tenía varias filiales constituidas por todo el mundo, con lo cual esta no es la única.

Centralidad de intermediación (*betweenness*). Mide cuántas veces un nodo actúa de mediador, esto es, que se sitúa en el medio entre otros nodos en la red. En este sentido, destacamos la entidad *MOSSFON SUBSCRIBERS LTD.*, otra filiar de Mossack, con un grado de intermediación de 0.00432, *PHATSHOEANE HENNEY ATTORNEYS* (tiene una red de vecindad de más de 400 entidades a su alrededor, por ejemplo)*, HONEYBOURNE HOLDINGS LIMITED* (algunos de cuyos agentes, como es el caso de CMS, que también figuran en la lista de los nodos con más centralidad de intermediación)*, SEWARD LIMITED, o CHESTERFIELD GROUP LIMITED*, con grados de *betweenness* que varían entre 0.00425 y 0.00338.

Esta medida de centralidad permite medir el control que ostentan los nodos que cuentan con un elevado valor de intermediación sobre la red, en general. Nótese que, aunque las entidades mencionadas sólo mantienen conexiones con, por ejemplo, 400 ó 14 nodos distintos, en una red, recordemos, de más de 670.000 relaciones entre los nodos, se trata igualmente de entidades con una gran influencia sobre aquella, lo cual les otorga más representatividad, puesto que sus nodos vecinos, a su vez, también tienen vínculos con otros nodos, y así sucesivamente. Lo anterior es de lo que se encarga el estudio de las comunidades (vecindad) que se realiza *infra*, pero también la medida de centralidad de cercanía permite respaldar la conclusión anterior.

Centralidad del autovalor (*eigenvector*). Mide la importancia del nodo en la red sin dejar de lado la relevancia de otros nodos integrantes del grafo que le rodean. El autovalor dominante de la matriz de adyacencia del grafo es 57,97, un valor elevado que indica, por tanto, que la red tiene nodos con alto grado de influencia sobre la misma. Es decir, que los nodos están bien conectados, y que las conexiones son fuertes. En definitiva, se trata de un grupo cohesionado.

En lo referente a los valores individuales más altos del autovalor, las entidades y personalidades que más destacan son *ANUBIS PROPERTIES CORP.,* cuyo agente es *Mossack*, *Meyer Joseph Nigri* que, a su vez, ostenta el rol de beneficiario de la empresa *Anubis*, *TECNISA ENGENHARIA E COMERCIO LTDA.*, que es accionista de *Anubis, Calistoga Corporation o Mount Eagle S.A.*, siendo aquella accionista de esta última.

Los elevados autovalores de estos nodos (por ejemplo, *Anubis* tiene un autovalor de 1, *Tecnisa* un 0,93 y *Meyer Joseph*, un 0,36) suponen que estos no sólo adquieren suma relevancia en la red, sino que también mantienen conexiones directas con nodos igualmente importantes en el *network*.

4.2.2 Componentes (comunidades) de la red

Las comunidades pueden definirse como subconjuntos de nodos dentro de un grafo que tienen conexiones más densas entre ellas en comparación con los vínculos que tengan con el resto de los vértices del grafo (Radicchi et al., 2004). Se emplea la función *components()* que permite determinar nodos entre los que existen esos caminos estrechos para poder extraer tales subgrafos.

Posteriormente, se centra el estudio en el componente con la mayor membresía, esto es, el que más conexiones entre los nodos tenga. Se emplea el método de *cluster walktrap* que pretende buscar subgrafos densamente conectados dentro de un grafo general utilizado el método de los *random walks*. Luego, a fin de permitir una visualización mejor y más interpretable, el análisis se centra en los primeros 3 *clusters* de una comunidad que cuenta con 302 miembros. Resulta que las comunidades más pobladas son como siguen:

- Comunidad #15: 28 miembros, donde se integran, por ejemplo, 4 filiales de *CANNON*.
- Comunidad #37: 17 miembros, algunos de los cuales son *BHP BILLITON (UK)*, varias filiales de *ESCOM*, y tres beneficiarios portugueses de las empresas *Diamonds Limited*, igualmente integrantes de la comunidad. A su vez, *BHP* ostenta participaciones en todas estas entidades.
- Comunidad #29: 16 miembros. Este componente integra, sobre todo, a las filiales de *Hoegh Capital*, así como a beneficiarios y accionistas de aquellas, que son *Thomas Hoegh*, *VIND SWEDEN AB* o *GORAN ENTERPRISES*, siendo este último accionista de *HCP Africa Limited*, una empresa que también aparece en la comunidad y que ya hemos tenido la ocasión de conocer al analizar las medidas de centralidad.

Como apunte extra, en el código fuente de R adjuntado en el Anexo I, se prevé la posibilidad de crear un dendrograma que permite visualizar, en este caso, los tres núcleos más importantes del subcomponente que se ha creado a efectos de este estudio. Las comunidades #15, #29 y #37 salen reflejadas en el mismo.

4.2.3 Asertividad y homofilia. Comunidades

La asertividad u homofilia (*assortativity*) es una métrica que estudia la tendencia de los nodos de conectarse con otros con los que guarden cierta similitud en algunas variables. En el paquete *igraph* del lenguaje R, se prevén dos funciones: *assortativity_degree()* y *assortativity_nominal()*. La primera calcula la homofilia basado en el grado de centralidad de los nodos, considerando todos los atributos por igual.

Mientras, el segundo método basa el cálculo de la asertividad en algún atributo categórico del grafo, de ahí que la función recibe dos paramentos en vez de uno. En cualquier de los dos casos, los valores que puede tomar la asertividad pueden ser entre –1 y +1, ambos inclusive. Un valor de homofilia del 0 sería indicativo de una ausencia de asertividad. En tal caso, los nodos se conectarían con otros indistintamente, sin atender al valor de un atributo concreto, en el caso de asertividad nominal.

Por una parte, la asociación de grado del grafo general es de –0.105, lo cual es indicativo de que los nodos que tengan un *degree* elevado (esto es, que tienen muchas conexiones con otros nodos) suelen conectarse con nodos con pocos enlaces. Esto tiene sentido debido a la estructura de la red de los Papeles de Panamá: la mayoría de veces, una entidad puede tener muchas aristas dirigidas a sus agentes entendido en sentido amplio, esto es, accionistas, directores, beneficiarios, etc. Estos últimos, a su vez, no suelen tener conexiones más allá de estas entidades principales en las que participan.

De otro lado, la función de homofilia nominal se ha aplicado a tres variables categóricas del grafo. Primero, a las comunidades identificadas con el método *cluster walktrap*, que arroja un resultado de 0,9937, hecho que incida casi una obviedad: las comunidades de vecinos que guardan estrechos vínculos entre sí, evidentemente van a tener una homofilia elevada.

En segundo lugar, a la variable de *countries* de los nodos; el resultado ha sido de 0,415, lo que muestra que, normalmente, los nodos de un mismo país se suelen vincular entre sí con mayor frecuencia que con nodos de otras jurisdicciones.

Finalmente, se ha estudiado la asertividad del subgrafo en función del tipo del nodo (entidad, agente o intermediario), y el resultado de –0,542 es claro indicador de que los nodos suelen conectarse con otros de un tipo distinto. A modo ejemplificativo, es mucho más frecuente que una entidad se vincule con un agente o un intermediario, en vez de otra entidad. Lo anterior

cobra todo el sentido del mundo considerando que los nodos correspondientes a los Papeles de Panamá en la base de datos, únicamente pueden tener relaciones de agente o intermediario. El restante –0,5 que falta para que se trate de una heterofilia completa se debe a que la base de datos cuenta con relaciones de las direcciones de los nodos (*address*).

4.2.4 Otras métricas

El presente estudio se ha nutrido de otras métricas que nos permite utilizar el paquete de *igraph* de R a fin de investigar con mayor abundamiento la base de datos de los Papeles de Panamá.

Transitividad. Mide el grado de agrupamiento de los nodos en triángulos. El resultado de 0,098 demuestra que los nodos no suelen formar triadas y, por tanto, hay una baja cohesión local, con lo cual hay una baja densidad de los triángulos.

Diámetro. Esta medida arroja la mayor distancia geodésica entre cualquier par de nodos en el grafo. En este caso, el valor del diámetro es 20 si interpretamos el grafo como no dirigido. Esta cifra, algo elevada, tiene sentido, puesto que se trata de una red con más de 600.000 vinculaciones, y de ninguna forma todos los nodos del grafo están interconectados entre sí, sino que hay núcleos más o menos densos, con un mayor o menos número de miembros. Por tanto, lógicamente se trata de un gráfico disperso.

En línea con lo anterior, también se ha determinado el listado de los nodos que integran el camino más largo, esto es, el diámetro del grafo, en términos de distancia geodésica. La cadena, que figura en el orden que tal cual aparece a continuación, es como sigue:

> AEON MANAGEMENT ESTABLISHMENT, EUROPEAN AMERICAN SECURITIES INC., HANSARD LIMITED, CANNON INTERNATIONAL LIMITED, CANNON ASSET MANAGEMENT LTD., Wallis Property Holdings Limited, Spread Nominees Limited, HONEYBOURNE HOLDINGS LIMITED, CMS LIMITED, SEWARD LIMITED, INTERCON LIMITED, BAYGROVE LIMITED,CORPORATE NOMINEES LIMITED, INTERNATIONAL HYDRO INVESTMENTS LIMITED, MOSSFON NOMINEES LIMITED, DRAGON TRUST COMPANY LIMITED, CHELTENHAM LIMITED, Anchor Management Services

Limited, TELFORD SECURITIES LIMITED, HERALD TRUST COMPANY LIMITED, ESTUARY LIMITED.

Hemos subrayado los nodos que ya han aparecido en los análisis anteriormente realizados en este trabajo (u otros filiales con un nombre parecido).

Distancia media. La función *mean_distance()* estudia la conectividad del grafo, en general. La métrica en sí calcula la distancia promedia entre todos los pares de nodos de la red. Este resultado ha sido de 5,69. El valor en cuestión confirma la conclusión que se lleva a cabo empleando la métrica del diámetro, que la red no está tan densamente conectada y que, en promedio, para llegar de un nodo a otro, se necesitan 5,69 aristas (o pasos).

4.2.5 Estudio de vecindad

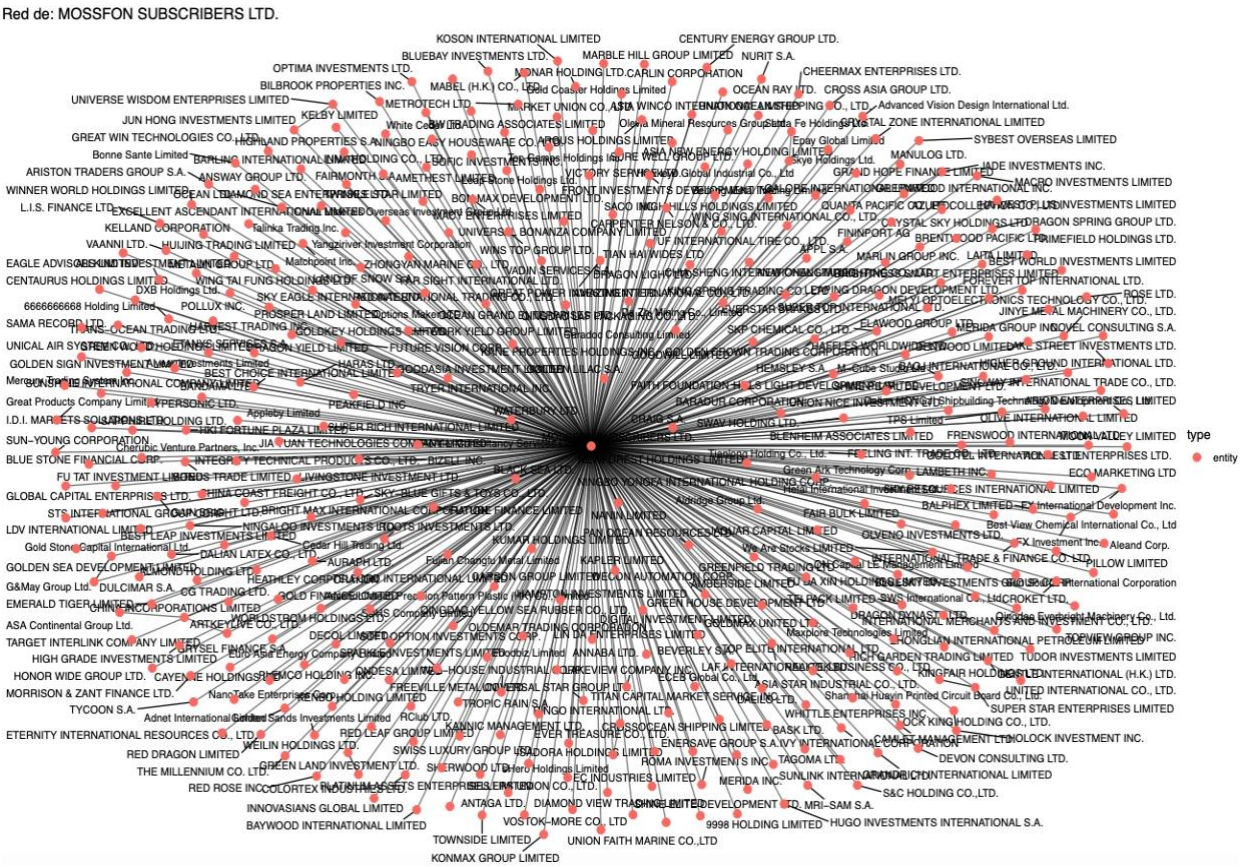
Se ha elaborado una lista con los 10 nodos que más vecinos tienen en el grafo original. No resulta nada sorprendente encontrar entre los primeros tres las entidades filiales de Mossack Fonseca:

- MOSSFON SUBSCRIBERS LTD.: 394 vecinos
- MOSSFON MANAGERS LTD.: 362 vecinos
- MOSSACK FONSECA & CO. (PERU) CORP.: 325 vecinos
- OFFSHORE BUSINESS CONSULTANT (INT'L) LIMITED: 289 vecinos
- ORION HOUSE SERVICES (HK) LIMITED: 256 vecinos
- CONSULCO INTERNATIONAL LIMITED: 189 vecinos
- PHATSHOEANE HENNEY ATTORNEYS: 182 vecinos
- MAYO NOMINEES LIMITED: 163 vecinos
- MAYO SECRETARIES LIMITED: 156 vecinos
- LEGAL CONSULTING SERVICES LIMITED: 154 vecinos

La observación que procede realizar tras visualizar esta lista es como sigue: en los grafos no dirigidos, las métricas de vecindad coincidirán con los resultados de las medidas de centralidad.

En cualquier caso, la Figura 3 muestra la densidad de la comunidad correspondiente a *Mossfon Subscribers*, que es la entidad que más vínculos tiene en este *dataset*:

Figura 3. Vínculos de la entidad MOSSFON SUBSCRIBERS LTD.



Fuente: elaboración propia

El hecho de encontrarse con entidades con alto número de vecinos en una red es algo inusual pues, como hemos visto antes, la mayoría de los nodos tienen un grado de conexiones relativamente bajo. Por ende, ello ya es indicativo de presencia de un comportamiento anómalo, posiblemente fraudulento en esta red. La anterior conclusión adquiere más peso cuando, analizando estas entidades, nos damos cuenta de que las primeras tres son filiales pertenecientes a un mismo agente.

Por las características de una red como esta, la eliminación de estos tres nodos significativos que hemos seleccionado en la práctica puede suponer un aumento de vulnerabilidad importante, pues los nodos que, a su vez, quedaban vinculados a estas entidades borradas, quedan expuestos al ser despojados de sus vínculos.

Lo expuesto se puede comprobar empleado métricas ya utilizadas a lo largo de este trabajo. Primero, los tres nodos tienen un grado de centralidad superior a la media más dos desviaciones típicas de todo el conjunto de este tipo de centralidad. Pero, de modo adicional, se ha realizado una simulación de eliminación de estos tres nodos de la red, procediendo, posteriormente, a la comparación de las métricas esenciales:

- Número de componentes: aumenta de 5.233 a 5.650, lo cual sugiere que la red ha quedado más fragmentada porque ahora hay más comunidades. Esto hace que devenga más vulnerable y menos cohesionada.
- Componente más grande: baja de 13.383 a 12.172, por tanto, la conectividad del grafo se ha visto disminuida.
- Diámetro: sube de 28 a 32, por ende, los caminos obligatorios entre los nodos ahora son más largos, sobre todo si se quiere alcanzar algún nodo distante o más alejado.
- Distancia media: aumenta de 9,33 a 10,88, por tanto, en línea con la argumentación sentada en sede de la métrica del diámetro, los nodos, en media y *ceteris paribus*, quedan más separados entre sí, lo que aumenta, insistimos, la vulnerabilidad de la red.

De cualquier modo, vemos necesario volver a insistir en el hecho de que la compañía *Mossack Fonseca* tiene muchas filiales que figuran en la base de datos. Lo anterior ha sido una simple demostración de cómo aumenta la fragilidad de la red de la trama con la eliminación de tan sólo tres entidades de este despacho. El efecto sería devastador si se eliminan todas las entidades dependientes de *Mossack*.

4.3 ESTUDIO PARTICULAR DE LA IMPLICACIÓN DE ENTIDADES ESPAÑOLAS EN LA TRAMA

A mayor abundamiento, hemos decidido poner a prueba las métricas y herramientas del análisis de redes a la hora de estudiar desde un enfoque más concreto, centrándonos, en este caso, en las entidades y agentes españoles involucrados en la trama.

Pre-procesamiento. Antes de nada, conviene aclarar brevemente que nos hemos decantado por el uso de la variable *countries* de las tablas de datos en vez de

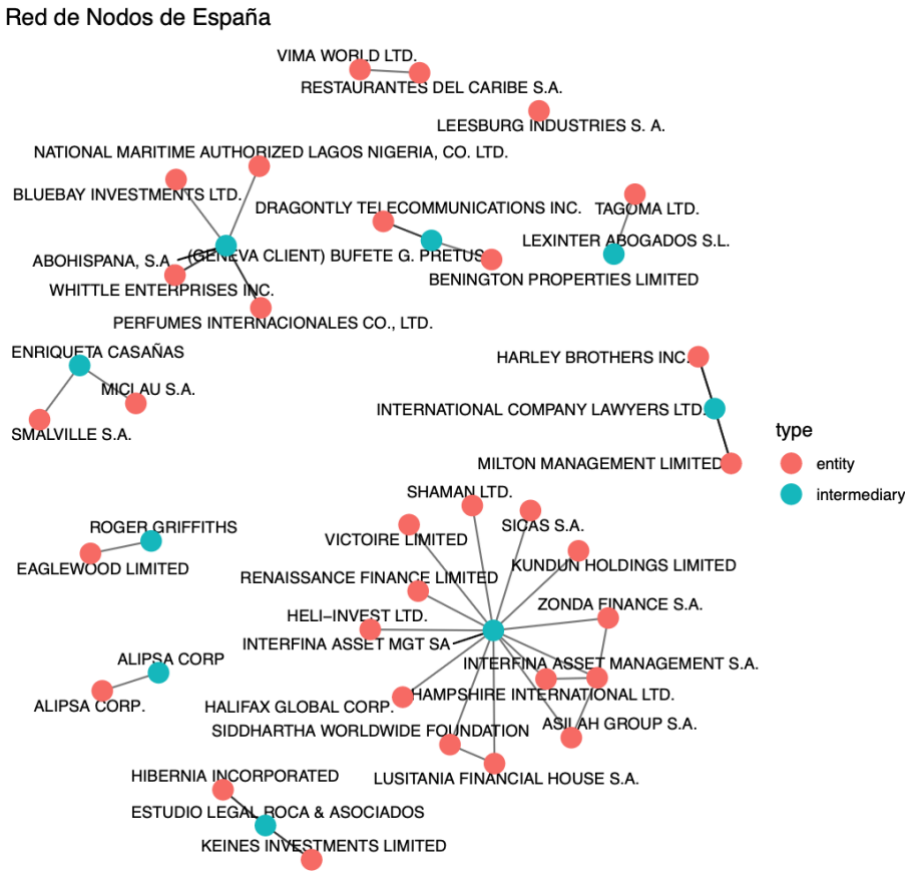
jurisdiction_description puesto que este último es donde radica, por ejemplo, una entidad. Y, dado que se trata de una trama de delitos fiscales cometidos en el *off-shore*, es lógico que casi todos los nodos van a tener su jurisdicción en paraísos fiscales. Simplemente basta con ejecutar la consulta de *unique(entities$jurisdiction_description)*, que devuelve valores como Samoa, Isle Of Man, Jersey, Panamá, Seychelles, Islas Vírgenes Británicas, o Belice. Mientras, la variable *countries* hace referencia al verdadero "punto original" del nodo, o ubicación de la sede principal desde la que se ejerce el control efectivo sobre esa entidad, si se quiere.

Medidas de centralidad. Las entidades con el mayor grado de centralidad (13, 6 y 6, respectivamente), respaldado por su índice de cercanía de 1,0, y un *betweenness* igualmente considerable, son *INTERFINA ASSET MGT SA, ABOHISPANA, S.A.* e *INTERNATIONAL COMPANY LAWYERS, LTD*.

Homofilia. Se aprecia heterofilia manifiesta en variables como tipo de nodo (–0,725) y en la asertividad de grado (–0.385). Mientras, la homofilia en la comunidad es un 1,0 perfecto.

Relaciones entre nodos españoles. En la Figura 4 a continuación se puede observar una red relativamente simple. Sólo vienen incluidos aquellos nodos españoles que, a su vez, tienen conexiones con otras entidades o intermediarios españoles:

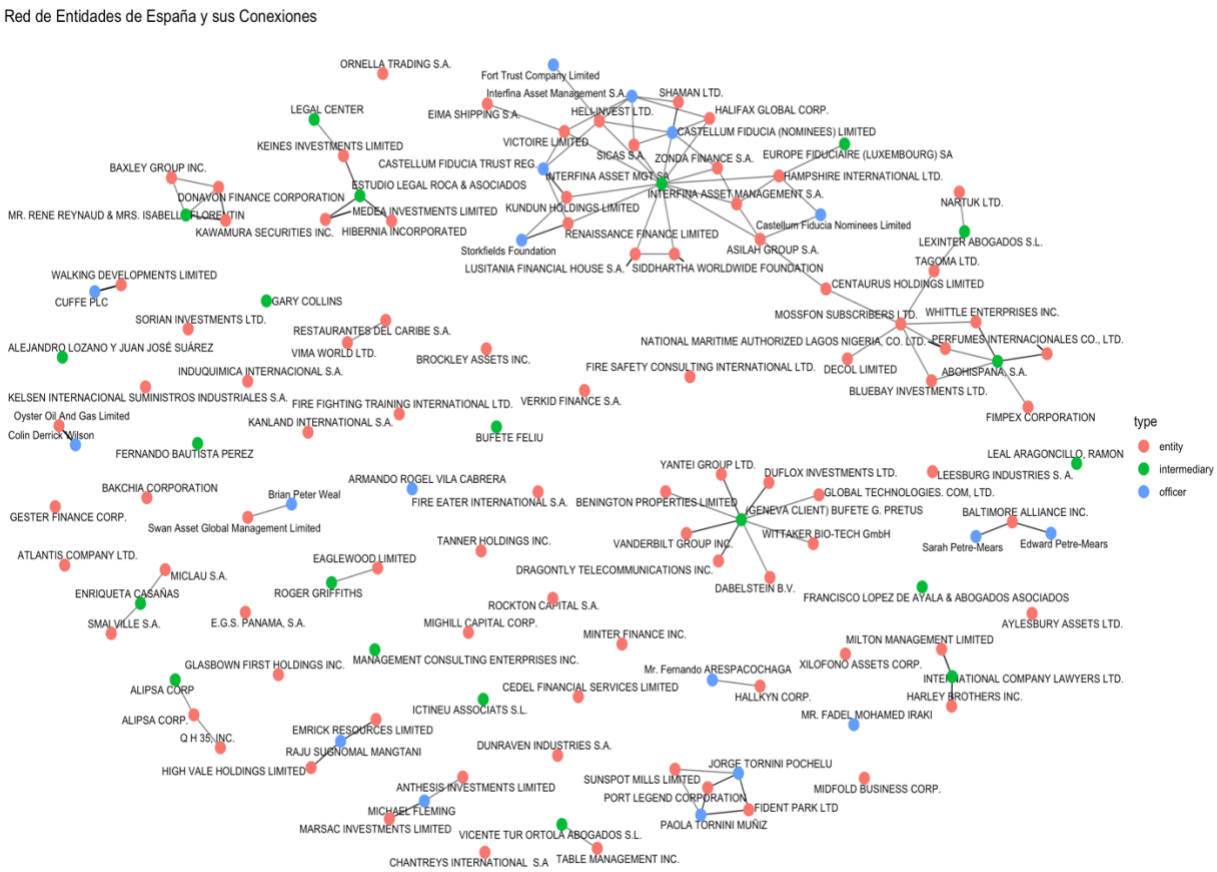
Figura 4. Red de Nodos de España



Fuente: elaboración propia

Relaciones con el resto de la red. La representación anterior confirma lo ya concluido con las medidas de centralidad: los nodos españoles más influyentes en la red son, en efecto, *INTERFINA* y *ABOHISPANA*, que son intermediarios de varias entidades con control mayoritariamente español, aunque algunas radican en otras jurisdicciones, de ahí que sus formas sociales se ajusten a los Ordenamientos jurídicos de otros países (véase, por ejemplo, *Bluebay Investments Ltd.*). En cualquier caso, la Figura 5 muestra las relaciones existentes de los nodos españoles con el resto del conjunto de la red.

Figura 5. Conexiones de los nodos españoles con la red global



Fuente: elaboración propia

Como se puede observar, las entidades españolas no parecen estar muy involucradas en la red, por lo general, puesto que no tienen un número elevado de conexiones y, en caso de tenerlas, no es con nodos que tengan una representatividad o importancia elevada según las medidas de centralidad u otras métricas relevantes en la red. De hecho, el autovalor de los tres nodos españoles más significativos según la centralidad de grado es 0. Por tanto, las entidades españolas no tienen mucha influencia en la red.

La misma conclusión se alcanza con las restantes medidas de centralidad: el máximo grado de centralidad de una entidad española es 13, seguido de 6, cuando el más alto de todo el grafo, el de *Mossfon*, es de 394.

CAPÍTULO 5. CONCLUSIONES GENERALES Y PARTICULARES

5.1 CONCLUSIONES GENERALES

Las técnicas de análisis de redes han demostrado ser métodos apropiados para detectar patrones de fraude y de las relaciones ocultos en conjuntos de datos masivos. La capacidad de visualizar las conexiones y relaciones entre entidades permite identificar nodos altamente conectados (*hubs*) y patrones de comportamiento sospechosos que podrían no ser evidentes mediante el empleo de métodos tradicionales. En el caso de los Papeles de Panamá, estas técnicas ayudaron a descubrir conexiones significativas entre entidades y personas implicadas en la trama.

Por su parte, las métricas de análisis de redes, como las centralidades o el estudio de la homofilia de los nodos, han demostrado ser apropiados para detectar patrones en la red, resaltando nodos y conexiones clave en las tramas. Por ejemplo, la centralidad de grado identifica los nodos más conectados que podrían actuar como intermediarios en estos esquemas fraudulentos, mientras que la centralidad de *betweenness* ayuda a destacar los nodos que controlan la información que atraviesa la red. También la centralidad del autovector, que permite determinar nodos que ejercen una influencia manifiesta en la red, no sólo por el peso que tienen ellos mismos, sino porque se relacionan con otros igualmente importantes.

A pesar de su utilidad, también hemos podido identificar inconvenientes varios. Por ejemplo, que la precisión de las métricas puede verse afectada por la calidad y la completitud de los datos. Un pre-procesamiento mal hecho condiciona la realización del estudio posterior, y las funciones de análisis de redes del paquete *igraph* no advierten de ello.

También destaca el problema de la escalabilidad, que se ha visto en el presente trabajo, cuando hemos tenido que crear un grafo con un número de datos reducidos para realizar el análisis y evitar excesivos tiempos de ejecución, por ejemplo. Asimismo, en ocasiones hay entidades que son filiales de una empresa matriz y que figuran por separado en la base de datos, lo cual abre la puerta a futuras líneas de investigación que tendrían esto en cuenta, realizando un esfuerzo extra a la hora de llevar a cabo el pre-procesamiento de los datos, juntando estas entidades en una única observación, para, posteriormente, estudiar su influencia como un solo

nodo en la red. En cualquier caso, la necesidad de segmentar grandes bases de datos en subgrafos más manejables deviene imperativo para un análisis eficiente.

5.2 CONCLUSIONES PARTICULARES

El análisis de subgrafos permite manejar grandes volúmenes de datos de manera efectiva. Al segmentar la red completa en grafos más reducidos, por ejemplo, eliminando todos aquellos nodos con un grado de centralidad igual o inferior a 5, fue posible realizar análisis detallados sin comprometer mucho el rendimiento. Esto demuestra que, con una adecuada segmentación, las técnicas de análisis de redes pueden aplicarse a grandes conjuntos de datos de manera eficiente, y los resultados son igualmente extrapolables a la red completa porque se mantienen los nodos clave de la trama.

Por añadidura, el estudio ha permitido identificar patrones específicos de comportamiento y conexión entre entidades que sugieren prácticas fraudulentas. Los *hubs* y los nodos con alta centralidad de *betweenness* mostraron ser puntos críticos en la red, a menudo conectando varias entidades sospechosas entre sí. El análisis de las comunidades permite visualizar núcleos influyentes en la red constituidos por varias entidades y personalidades.

Asimismo, al identificar los nodos clave de la trama, y realizar una simulación de eliminación de estos de la red, nos podemos dar cuenta de que, todas las métricas relevantes arrojan peores resultados de cohesión de la red. Este análisis, por tanto, es importante porque ayuda a identificar nodos realmente influyentes en la red.

Lo que se echa en falta en este trabajo es una visualización del panorama general de la red, limitándonos, por motivos de rendimiento del ordenador, a realizar visualizaciones de estudios particulares concretos. Lo anterior, no obstante, sigue siendo útil si se realiza un análisis segmentado de la trama, que es lo que hemos hecho, en gran parte.

En otro orden de cosas, hemos observado que todas las métricas empleadas arrojan resultados que pueden someterse a una extensa interpretación y, por tanto, ayudan a sacar conclusiones relevantes en relación con la identificación de patrones de relaciones y fraudes cometidos. Cada una está destinada a analizar una cuestión en particular, y no hemos observado

ninguna cuya aplicación haya carecido de sentido, salvo en casos de obviedad que se han realizado adrede y se han señalado en este trabajo.

Concluimos también, que los hallazgos del estudio de los Papeles de Panamá pueden ser extrapolados a otras redes de fraude o tramas similares. Si bien, cada grafo que se constituye adopta unas características particulares, realizando un correcto ajuste del código fuente utilizado a una base de datos distinta permitirá al investigador sacar *insights* valiosos de cualquier modo. Las técnicas y métricas utilizadas son aplicables a cualquier red compleja donde haya que identificar patrones de comportamiento y conexiones sospechosas.

Finalmente, también hemos visto que es posible estudiar el involucramiento de entidades de un área geográfico o país concreto, como ha sido el caso en este estudio, al centrar la última parte del trabajo en el análisis de redes hecho sobre las entidades españolas. Hemos podido concluir, por ejemplo, que estas no tienen una influencia determinante en la red global. Por añadidura, no son muchas y tienen grados de centralidad muy bajos, lo cual indica que no se relacionan intensamente con la red en su conjunto.

DECLARACIÓN RESPECTO AL USO DE LA INTELIGENCIA ARTIFICIAL

Por la presente, yo, Filipp Andrianov Andrianov, estudiante de 5º E3-Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "Técnicas de Análisis de Redes para la Detección de Fraude", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Brainstorming de ideas de investigación:** Utilizado para idear y esbozar posibles áreas de investigación.
2. **Referencias:** Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
3. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
4. **Interpretador de código:** Para realizar análisis de datos preliminares.
5. **Generador de problemas de ejemplo:** Para ilustrar conceptos y técnicas.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para qué se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 21 de junio 2024

Firma: Filipp Andrianov

ANEXOS

ANEXO I: Código fuente de R

```
install.packages("ggraph")
install.packages("pbapply")
install.packages("igraph")
install.packages("tidyverse")
install.packages("tidygraph")
install.packages("data.table")
install.packages("tm")
install.packages("slam")

library(ggraph)
library(pbapply)
library(igraph)
library(tidyverse)
library(tidygraph)
library(data.table)
library(tm)
library(slam)

#-----------------------------------------------------#

#Ubicados en el folder [full-oldb copy]
relationships <- read_csv("relationships.csv")
entities <- read_csv("nodes-entities.csv")
intermediaries <- read_csv("nodes-intermediaries.csv")
officers <- read_csv("nodes-officers.csv")

entities <- entities %>% filter(sourceID == 'Panama Papers')
intermediaries <- intermediaries %>% filter(sourceID == 'Panama Papers')
```

```
officers <- officers %>% filter(sourceID == 'Panama Papers')
relationships <- relationships %>% filter(sourceID == 'Panama Papers')



#-----------------------------------------------------#

# Unimos para incluir la jurisdiccion (pais) en la tabla de las relaciones
# Con esto, evitamos tener que crear una base de datos SQL para hacer el preprocesamiento y lo realizamos aquí directamente
# Crear una lista con todas las entidades, intermediaries y officers
entidades <- entities %>%  select(node_id, name, countries) %>% mutate(type = "entity")

intermediarios <- intermediaries %>% select(node_id, name, countries) %>%  mutate(type = "intermediary")

agentes <- officers %>%  select(node_id, name, countries) %>% mutate(type = "officer")

nodos_total <- bind_rows(entidades, intermediarios, agentes)

relationships <- relationships %>%
  left_join(nodos_total %>% select(node_id, start_countries = countries, start_name = name), by = c("node_id_start" = "node_id")) %>%
  left_join(nodos_total %>% select(node_id, end_countries = countries, end_name = name), by = c("node_id_end" = "node_id"))

relationships <- relationships %>% select(-sourceID, -status, -start_date, -end_date)

#Comprobamos
head(relationships)

#Creamos el grafo a partir de la tabla modificada arriba y, para poder trabajar mejor con el dataset, recortamos el grafo y creamos subred
```

```
g <- graph_from_data_frame(d = relationships %>% select(start_name, end_name), directed = F)
g <- delete_vertices(g, V(g)[degree(g) <= 5])



# En el grafo, añadimos los atributos de los nodos, que son oficiales, interns y entities, conforme se expone en la parte teorica del trabajo
V(g)$type <- nodos_total$type[match(V(g)$name, nodos_total$name)]
V(g)$countries <- nodos_total$countries[match(V(g)$name, nodos_total$name)]

g <- delete_vertices(g, V(g)[is.na(V(g)$countries)])



#-----------------------------------------------------#

# Para no sobrecargar el ordenador con el analisis que se realiza luego (metricas de centralidad, etc)
# Lo que hacemos ahora es 1) encontrar los nodos más conectados entre sí (la función del degree nos ayuda a ver cuantos arcos tiene un nodo)
# Y con el parametro decreasing los ordenamos en orden descendiente (esto es, que figuren los nodos que más conexiones reciben)
#Y nos quedamos solo con 1000 nodos del subgrafo
nodos_mas_conectados <- names(sort(degree(g), decreasing = T))[1:2000]

componente_para_centralidad <- induced_subgraph(g, V(g)$name %in% nodos_mas_conectados)

# A continuación, hacemos estudio de las centralidades (de grado, de cercania, eigen y de betweenness)

centralidad_grado <- degree(componente_para_centralidad)
centralidad_cercania <- closeness(componente_para_centralidad, normalized = T)
centralidad_eigen <- eigen_centrality(componente_para_centralidad, directed = F)$vector
```

```
centralidad_eigen_valor <- eigen_centrality(componente_para_centralidad, directed = F)$value
centralidad_betweenness <- betweenness(componente_para_centralidad, directed = F, normalized = T, cutoff = -1)
centralidad_betweenness <- unlist(centralidad_betweenness) #convertirlo a vector

print(head(sort(centralidad_grado, decreasing = T), 10))
print(head(sort(centralidad_cercania, decreasing = T), 10))
print(head(sort(centralidad_eigen, decreasing = T), 15))
print(centralidad_eigen_valor)
print(head(sort(centralidad_betweenness, decreasing = T), 20))

centralidad_degree <- degree(componente_para_centralidad, mode="all")
pdf("histograma_centralidad.pdf")
hist(centralidad_degree, breaks=80, main="Histograma centralidad de grado de los nodos")
dev.off()

deg <- degree(componente_para_centralidad, mode ='all')
deg.dist <- degree_distribution(componente_para_centralidad, cumulative=T, mode="all")
pdf("grado_distribucion.pdf")
plot(x=0:max(deg),y=1-deg.dist, pch=19, cex=1.2, col="orange", xlab="Grado Distribución", ylab="Frecuencia Acumulada")
dev.off()



#-----------------------------------------------------#

#Siguiente asunto. Creamos un dendrograma. Nos permitirá estudiar, desde otra perspectiva, la conexión entre las comunidades dentro del
#componente concreto.
#Antes, para ello empleamos la función de components() que nos permite determinar nodos entre los que existen caminos para poder extraer subgrafos
#con dichos nodos (componentes conectados)
```

```
#Vamos a centrarnos en el componente con la mayor membresía, esto es, el que más conexiones entre los nodos tenga
componentes <- components(g)
componente_principal <- induced_subgraph(g, which(componentes$membership == which.max(componentes$csize)))
nodos_mas_conectados <- names(sort(degree(componente_principal), decreasing = T))[1:500]

componente_para_dendrograma <- induced_subgraph(componente_principal, V(componente_principal)$name %in% nodos_mas_conectados)
vcount(componente_para_dendrograma)



comunidades_walktrap <- cluster_walktrap(componente_para_dendrograma)
top_clusteres <- sort(table(membership(comunidades_walktrap)), decreasing = T)[1:3]
nodos_seleccionados <- unlist(lapply(names(top_clusteres), function(x) {
  which(membership(comunidades_walktrap) == as.numeric(x)) }))

subgrafo_nodos_seleccionados <- induced_subgraph(componente_para_dendrograma, nodos_seleccionados)
comunidades_subgrafo <- cluster_walktrap(subgrafo_nodos_seleccionados)
dendrograma_walktrap <- as.dendrogram(comunidades_subgrafo)
par(cex = 0.5)
pdf("dendrograma_comunidades_walkstrap.png")
plot(dendrograma_walktrap, main = "Dendrograma Comunidad Principal")
dev.off()
par(cex = 1

print(length(comunidades_walktrap))
print(sort(sizes(comunidades_walktrap), decreasing = T))

c15 <- V(componente_para_dendrograma)[comunidades_walktrap$membership == 15]
c37 <- V(componente_para_dendrograma)[comunidades_walktrap$membership == 37]
```

```
c29 <- V(componente_para_dendrograma)[comunidades_walktrap$membership == 29]
print(V(componente_para_dendrograma)$name[c15])
print(V(componente_para_dendrograma)$name[c37])
print(V(componente_para_dendrograma)$name[c29])

#-----------------------------------------------------#


comunidades <- cluster_walktrap(componente_para_centralidad)
V(componente_para_centralidad)$community <- comunidades$membership

assortativity_degree(g)
assortativity_nominal(componente_para_centralidad,
as.factor(V(componente_para_centralidad)$community))
assortativity_nominal(componente_para_centralidad,
as.factor(V(componente_para_centralidad)$countries))
assortativity_nominal(componente_para_centralidad,
as.factor(V(componente_para_centralidad)$type))

#-----------------------------------------------------#

transitivity(componente_para_centralidad, type = "global")

head(sort(transitivity(componente_para_centralidad, type = "local"), decreasing=T),10)

diameter(componente_para_centralidad, directed=F, weights=NA)

get_diameter(componente_para_centralidad, directed=F)

mean_distance(componente_para_centralidad, directed=F)


#-----------------------------------------------------#
```

```
#Estudio de neighbors del grafo general: quien más tiene y el top 10

vecinos_top <- function(g) {
  node_neighbors <- sapply(V(g), function(x) length(neighbors(g, x)))
  indices <- order(node_neighbors, decreasing = T)[1:10]
  nodos <- V(g)[indices]
  count <- node_neighbors[indices]
  return(data.frame(name = V(g)$name[nodos], neighbors = count))
}
vecinos_top(g)


nodo_mossfon <- neighbors(g, V(g)[name == "MOSSFON SUBSCRIBERS LTD."])
subgrafo_nodos <- c(V(g)[name == "MOSSFON SUBSCRIBERS LTD."], nodo_mossfon)
mossfon_subgrafo <- induced_subgraph(g, subgrafo_nodos)

pdf("mossfon_red.pdf")
ggraph(mossfon_subgrafo, layout = "fr") +
  geom_edge_link(aes(edge_alpha = 0.5), show.legend = FALSE) +
  geom_node_text(aes( label = name ), repel = TRUE, size = 3) +
  geom_node_point(aes( color = type), size = 3) +
  theme_void() +
  labs(title = paste("Red de: MOSSFON SUBSCRIBERS LTD."))
dev.off()


mossfon1 <- "MOSSFON SUBSCRIBERS LTD."
mossfon2 <- "MOSSFON MANAGERS LTD."
mossack <- "MOSSACK FONSECA & CO. (PERU) CORP."

centr_g <- degree(g, mode = "all")
```

```
degree_media <- mean(centr_g)
degree_desv <- sd(centr_g)
mossfon1 > (degree_media + 2*degree_desv)
mossfon2 > (degree_media + 2*degree_desv)
mossack > (degree_media + 2*degree_desv)



g_sin_mossfon1 <- delete_vertices(g, V(g)[name == mossfon1])
g_sin_mossfon2 <- delete_vertices(g_sin_mossfon1, V(g_sin_mossfon1)[name == mossfon2])
g_sin_mossfon3 <- delete_vertices(g_sin_mossfon2, V(g_sin_mossfon2)[name == mossack])
g_sin_mossfon <- g_sin_mossfon3



componentes_uno <- components(g)
mas_grande_antes <- max(componentes_uno$csize)
diametro_antes <- diameter(g, directed = F)
media_antes <- mean_distance(g, directed = F)



componentes_dos <- components(g_sin_mossfon)
mas_grande_despues <- max(componentes_dos$csize)
diametro_despues <- diameter(g_sin_mossfon, directed = F)
media_despues <- mean_distance(g_sin_mossfon, directed = F)



data.frame(
 Metricas = c("N Componentes", "Componente Mayor", "Diametro", "Distancia Media"),
 Antes = c(length(componentes_uno$csize), mas_grande_antes, diametro_antes, media_antes),
 Despues = c(length(componentes_dos$csize), mas_grande_despues, diametro_despues, media_despues)
```

```
)



#-----------------------------------------------------#



#Estudio España
#Usamos countries en vez de jurisdictions para centrarnos en lo que realmente radica desde España (nodo, control...)
spain_nodos <- V(g)[V(g)$countries == "Spain"]
spain_grafo <- induced_subgraph(g, spain_nodos)

cent_grado_spain <- degree(spain_grafo, mode = "all")
cent_closeness_spain <- closeness(spain_grafo, mode = "all", normalized = TRUE)
centrality_betweenness_spain <- betweenness(spain_grafo, directed = FALSE, normalized = TRUE)

centralidad_spain <- data.frame(
  name = V(spain_grafo)$name,
  degree = cent_grado_spain,
  closeness = cent_closeness_spain,
  betweenness = centrality_betweenness_spain
)

centralidad_spain %>% arrange(desc(degree)) %>% head(10)

components_spain <- components(spain_grafo)
components_spain$csize
V(spain_grafo)$community <- components_spain$membership
assortativity_nominal(spain_grafo, as.factor(V(spain_grafo)$community))

spain_neighbors <- unlist(neighborhood(g, order = 1, nodes = spain_nodos))
spain_grafo_expanded <- induced_subgraph(g, spain_neighbors)
```

```
ggraph(spain_grafo_expanded, layout = "fr") +
  geom_edge_link(aes(edge_alpha = 0.5), show.legend = F) +
  geom_node_text(aes(label = name), repel = T, size = 3) +
  geom_node_point(aes(color = type), size = 4) +
  theme_void() +
  labs(title = "Red de Entidades de España y sus Conexiones")



# Tres primeros nodos clasificados según centralidad
top_invol <- centralidad_spain %>% arrange(desc(degree)) %>% head(5)
top_invol$type <- V(g)$type[match(top_invol$name, V(g)$name)]
top_invol

vecinos_listar <- function(graph, node_name) {
  neighbors(graph, V(graph)[name == node_name])$name
}

top_invol$vecinos <- lapply(top_invol$name, function(node_name)
get_neighbors(spain_grafo, node_name))
top_invol
top_invol$vecinos

crear_subgrafo_vecinos <- function(graph, node_name) {
  vecinosid <- neighbors(graph, V(graph)[name == node_name])
  subgraph(graph, c(V(graph)[name == node_name], vecinosid))
}

for (node_name in top_invol$name) {
  subgrafo <- crear_subgrafo_vecinos(spain_grafo, node_name)
  p <- ggraph(subgrafo, layout = "fr") +
    geom_node_point(aes(color = type), size = 8) +
    geom_edge_link(aes(edge_alpha = 0.5), show.legend = F) +
```

```
    geom_node_text(aes(label = name), repel = T) +
    theme_void() +
    labs(title = paste("Vecinos del Nodo:", node_name))
  print(p)
}
```

BIBLIOGRAFÍA

Avi-Yonah, R. S. (2006). The three goals of taxation. *Tax Law Review, 60*(1), 1-28.

Batta, M. (2020). Machine learning algorithms - A review. *International Journal of Science and Research (IJSR), 9*(1).

Blanco Cordero, I. (2011). El delito fiscal como actividad delictiva previa del blanqueo de capitales. *Revista Electrónica de Ciencia Penal y Criminología, 13*-01.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), P10008.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2018). Analyzing social networks (2nd ed.). *SAGE Publications*.

Boyd, S., & Vandenberghe, L. (2004). Convex optimization. *Cambridge University Press*.

Consejo Económico y Social (CES). (2021). *Informe distribución de la renta en España: Desigualdad, cambios estructurales y ciclos*.

Csardi, G., & Nepusz, T. (2006). *The igraph software package for complex network research. InterJournal, Complex Systems, 1695*. Retrieved from https://igraph.org/r/

Dharmapala, D., & Hines, J. R., Jr. (2009). Which countries become tax havens? *Journal of Public Economics, 93*(9-10), 1058-1068.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik, 1*, 269-271.

Dominguez, D., Pantoja, O., Pico, P., Mateos, M., Alonso-Almeida, M. M., & González, M. (2020). Panama Papers' offshoring network behavior. *Heliyon, 6*(12).

Ferrara, E., De Meo, P., Fiumara, G., & Baumgartner, R. (2014). Detecting criminal organizations in mobile phone networks. *Expert Systems with Applications, 41*(13), 5733-5750.

Fonseca-Reyna, Y. C., Martínez-Jiménez, Y., & Nowé, A. (2018). Aprendizaje reforzado aplicado a la programación de tareas bajo condiciones reales. *Ingeniería Industrial, 39*(1), La Habana, ene.-abr.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks, 1*(3), 215-239.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences, 99*(12), 7821-7826.

González, I., & Mateos Caballero, A. (2018). Social network analysis tools in the fight against fiscal fraud and money laundering. In *15th International Conference on Modeling Decisions for Artificial Intelligence (MDAI 2018)* (pp. 226-237).

Gottschalk, P. (2010). Categories of financial crime. *Journal of Financial Crime, 17*(4), 416-435.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

International Consortium of Investigative Journalists. (s.f.). Explore the Panama Papers Key Figures. *ICIJ*.

International Public Sector Fraud Forum. (2020). *Guide to understanding the total impact of fraud*.

La Razón. (2023). España dejó de ingresar 662 millones de euros por fraude de IVA en 2021. *La Razón*.

Lateef, I. (2021). Machine learning techniques for network analysis. *New Jersey Institute of Technology*.

Li, Y. (2018). Deep reinforcement learning: An overview. *arXiv*.

Nykamp D.Q. (s.f.). Directed graph definition. *Math Insight*.

Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science, 28*(6), 441-453.

Papachristos, A. V. (2009). Murder by structure: Dominance relations and the social structure of gang homicide. *American Journal of Sociology, 115*(1), 74-128.

Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks. *Lecture Notes in Computer Science, 3733*, 284-293.

Sandoval, L. L. J. (2017). Machine learning algorithms for analysis and data prediction. In *2017 IEEE 37th Central America and Panama Convention (CONCAPAN XXXVII)* (pp. 1-5).

Slemrod, J., & Yitzhaki, S. (2002). Tax avoidance, evasion, and administration. In A. J. Auerbach & M. Feldstein (Eds.), *Handbook of Public Economics* (Vol. 3, pp. 1423-1470). Elsevier.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440-442.

Wikipedia. (2023, June 21). Repercusión de los Panama Papers en España. *Wikipedia, la enciclopedia libre*.

Zhuhadar, L., & Ciampa, M. (2019). Leveraging learning innovations in cognitive computing with massive data sets: Using the offshore Panama papers leak to discover patterns. *Computers in Human Behavior, 92*, 507-518.