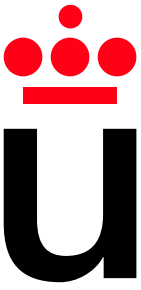

Universidad
Rey Juan Carlos

Escuela Técnica Superior
de Ingeniería Informática

Grado en Ingeniería de la Ciberseguridad

Curso 2023-2024

Trabajo Fin de Grado

EVOLUCIÓN, IMPORTANCIA ACTUAL DE LA INTELIGENCIA DE FUENTES ABIERTAS Y APLICACIÓN A HUELLAS DIGITALES

Autor: Óscar Lozano Pérez
Tutor: Javier Yuste Moure
Cotutor: Raúl Martín Santamaría

©2024 Óscar Lozano Pérez
Algunos derechos reservados

Este documento se distribuye bajo la licencia "Atribución-CompartirIgual 4.0 Internacional" de Creative Commons, disponible en: https://creativecommons.org/licenses/by-sa/4.0/deed.es

Resumen

La Inteligencia de Fuentes Abiertas es un concepto de vital importancia en el ámbito actual de la ciberinteligencia. Bajo este término se recogen todas aquellas fuentes de información que están disponibles de forma pública y que pueden aprovecharse en investigaciones que busquen generar inteligencia ya que, actualmente, existen numerosas fuentes abiertas de información que son fácilmente accesibles gracias a la extensión de internet.

Este documento presenta una introducción al concepto de Inteligencia de Fuentes Abiertas para así profundizar en la tipología actual de fuentes de información, características de los procesos de inteligencia y sus diferentes etapas, además de diferentes técnicas de obtención de información. Esto permite establecer una metodología para realizar una investigación utilizando fuentes abiertas.

A continuación, se aplicará la metodología en un caso práctico particular, en el que se definirá el concepto de "huella digital" y se investigará la información disponible en fuentes abiertas acerca de una persona pública: el rector de una universidad pública madrileña. Tras llevar a cabo la investigación siguiendo la metodología establecida, se ofrecen una serie de recomendaciones a partir de la información recopilada con la finalidad de reducir la exposición del sujeto a usuarios malintencionados.

Finalmente, se definen conclusiones y trabajos a futuro teniendo en cuenta el estudio realizado sobre la Inteligencia de Fuentes Abiertas y el desarrollo del caso práctico mencionado.

Palabras clave:

- Fuentes de información
- Ciberinteligencia
- Técnicas OSINT
- Huella digital
- Internet

Abstract

Open Source Intelligence constitutes a concept of great value in the present field of cyberintelligence. This term refers to every information source publicly available, which could be used in investigations in order to generate intelligence, as the number of easily accesible open information sources has considerably increased due to the internet's extent.

This paper presents an introduction to the concept of Open Source Intelligence as a means to examine the current categorisation on different information sources, intelligence processes' characteristics and their phases, along with different information gathering techniques. This is synthesized in a established methodology to conduct an investigation using open sources.

Next, this methodology will be applied in a real case, which will define the term "digital footprint" and research any information available in open sources about a public figure: a public university chancellor from Madrid. Following the results of the investigation, a series of recommendations are included in order to mitigate the level of exposure of the subject to malicious users.

Lastly, conclusions and related future proyects are defined considering the proposed analysis on Open Source Intelligence and the development of the investigation about the subject's digital footprint.

Keywords:

- Information sources
- Cyberintelligence
- OSINT Techniques
- Digital footprint
- Internet

Índice de contenidos

Índice de tablas

Índice de figuras

Lista de Acrónimos

CIA *Central Intelligence Agency*. 5

CNI Centro Nacional de Inteligencia. 14

HUMINT *Human Intelligence*. 17

IMINT *Image Intelligence*. 17

MASINT *Measurement And Signature Intelligence*. 17

OSINT *Open Source Intelligence*. 1, 17, 43

SIGINT *Signal Intelligence*. 17

SOCMINT *Social Media Intelligence*. 18

TECHINT *Technology Intelligence*. 17

1 Introducción

Hoy en día, existe una amplia interconexión de personas gracias a la *World Wide Web*, fenómeno que no ha parado de expandirse desde finales del siglo pasado. La globalización actual permite al usuario promedio valerse de herramientas automatizadas para acceder a todo tipo de información, y, a su vez, los expertos en inteligencia han necesitado adaptarse a los tiempos que corren para optimizar sus investigaciones. En estos ámbitos profesionales, acogidos dentro de la ciberseguridad, se aplican las técnicas más sofisticadas para explorar las fuentes abiertas presentes en internet.

En este documento se recoge la memoria del Proyecto de Fin de Grado del autor, que, en línea con las asignaturas impartidas en la Ingeniería de la Ciberseguridad, abarca uno de los temas más amplios dentro de la ciberinteligencia. Esta primera sección se divide en diferentes apartados: la Subsección 1.1 especifica el contexto y el alcance del trabajo, la Subsección 1.2 expone la motivación e importancia del mismo y la Subsección 1.3 resume la historia de la inteligencia de fuentes abiertas.

1.1. Contexto y alcance

Este trabajo se centra en el concepto de Inteligencia de Fuentes Abiertas (OSINT)[2] junto con las causas y consecuencias de los cambios que han tenido lugar a lo largo del crecimiento de internet en lo referido a las técnicas OSINT. Las fuentes abiertas las constituyen todas las fuentes de información disponibles de forma pública, y dado el gran número de fuentes presentes en internet, así

como la envergadura de los datos que aportan, se requiere el uso de técnicas que se encarguen de explorar y obtener la información disponible, para así cumplir distintos objetivos de inteligencia [3].

Las técnicas OSINT son utilizadas en cualquier circunstancia que requiera de una investigación, sea a un individuo o a una entidad, por lo que resultan prácticas para cualquier usuario que considere necesario obtener conocimientos sobre alguna persona u organización.

Actualmente, existen muchos tipos de agentes investigadores cuyas funciones requieren el uso de técnicas OSINT y que, gracias a su conocimiento técnico, algunos son capaces de desarrollar y mantener herramientas que ayudan a aplicarlas de forma más eficiente. Los expertos en inteligencia de hoy en día pueden alcanzar una gran diversidad de metas gracias a la estricta recopilación de información en línea; por ello, muchos de estos expertos pertenecen al moderno ámbito de la ciberinteligencia y son capaces de aplicar estas habilidades en una serie de contextos diferentes, véase:

- Trazado de huellas digitales, que consiste en la compilación e interrelación de datos personales o de interés de un usuario para conocer sus hábitos, relaciones sociales o vivencias a través de la información disponible en internet [4]. Su valor reside en investigaciones de individuos y, en casos extremos, la localización de personas desaparecidas[1].

- Procesos de inteligencia de amenazas, definiendo este ámbito como la obtención y tratamiento de información sobre actores maliciosos con el ánimo de aplicar efectivas medidas preventivas y reactivas [5]. Especialmente útil en entidades empresariales o estatales que buscan perfilar grupos de cibercriminales y estimar riesgos relativos a los mismos, normalmente gracias a la subcontratación de compañías[2] expertas en este aspecto.

- Utilidad en pruebas de penetración o *pentesting*, donde se evalúa la seguridad de una empresa, infraestructura o aplicación a partir de simular un ciberataque real cuyo objetivo es ganar acceso a la entidad a partir del compromiso del activo [6]. Las técnicas OSINT suelen utilizarse en fases previas a la realización del ciberataque para recolectar información sobre la organización o el software en cuestión y poder comenzar a dar forma al proceso[3].

- Otros ejemplos de aplicación actuales, algo apartados de los ámbitos más técnicos de la ciberseguridad, son la verificación de antecedentes u obtención de pruebas legales contra cibercriminales, procedimientos realizados con fre-

[1] Como ejemplo, los proyectos de *Trace Labs*: https://www.tracelabs.org/

[2] Ejemplo: https://www.threatintelligence.com/about-us

[3] Para más información: https://www.eccouncil.org/cybersecurity-exchange/penetration-testing/penetration-testing-open-source-intelligence-osint/

cuencia por peritos judiciales informáticos[4] y otros agentes dedicados a la investigación.

A lo largo de esta memoria se abordarán diferentes temas relativos a las técnicas OSINT, ordenados en diferentes secciones. En la Sección 2 se especificará un objetivo principal acompañado de una serie de objetivos secundarios para continuar con la Sección 3, que mostrará el estado del arte de los procesos de inteligencia formales remarcando las características clave de estos. Tras ello, la Sección 4 estará dedicada a la aplicación de lo desarrollado en un caso práctico relacionado con la enumeración anterior, utilizando técnicas OSINT en una investigación real con herramientas que facilitarán la búsqueda de información. Finalmente, quedarán por escrito las conclusiones alcanzadas en la Sección 5, donde se hablará de la aplicación, eficacia y versatilidad de las fuentes abiertas, y se incluirán directrices para posibles trabajos futuros que complementen lo expuesto en este documento.

1.2. Motivación

Desde hace décadas, muchas organizaciones de importancia, ya sean gobiernos o empresas, hacen uso de las fuentes abiertas para alcanzar sus objetivos [7]: ya sea para llevar a cabo planificaciones alineadas con sus procesos de negocio o por aplicar la información de estas fuentes en investigaciones más enfocadas. En el contexto descrito anteriormente, han surgido factores que dotan a las fuentes abiertas de gran potencial; para principios de Abril de este año, ya se reporta un número de usuarios en internet equivalente al 62.6 % de la población mundial, un billón de historias en Facebook publicadas al día, más de cuatro billones de "me gusta" en Instagram al día y, de media, 6000 *tweets* cada segundo [8]. Dejando de lado las redes sociales, internet actualmente da disponibilidad a más de un billón de páginas web, y un 71 % de las empresas poseen una [9], ya sea para presentar su negocio o para hacerlo efectivo a través de tiendas en línea, una plataforma digital para pedidos o para reserva de cita previa; todo ello, habitualmente, mediante el trato de datos personales de sus clientes a través de la gestión de cuentas de usuario, factor de interés para investigaciones a través de la red.

Al conocer estos datos, es sencillo imaginar casos investigados por autoridades policiales en los que, gracias a datos o publicaciones encontradas en internet, se llega a la identificación de criminales y posterior arresto, como se explica en este artículo de *Police Journal* [10]. Como también explica el artículo, a la hora de obtener evidencias judiciales puede resultar complicado añadir una fuente abierta de información al deber cumplir con determinados principios, pero es innegable

[4]Para referencia, ver proyectos realizados por el usuario P. Duchement en https://pduchement.org

que no sólo es posible, sino que han estado dándose casos en las últimas dos décadas: "Entonces, ¿puede la inteligencia ser tratada como una prueba? La respuesta es sencilla: sí, siempre que la inteligencia cumpla con los requisitos relevantes para ese tipo de evidencia. Además de las exclusiones forales específicas, [...] no hay razón por la que este material no sea admitido como prueba. Existen, aun así, consideraciones adicionales a tratar por los investigadores antes de presentar dicho material."

Por ello, es posible mencionar alguna incidencia real cuya investigación no hubiera podido avanzar de no ser por las fuentes abiertas. Como primer ejemplo, fue posible identificar y arrestar a una mujer que prendió fuego a dos coches policiales en una protesta a favor de George Floyd [11] a través de indagar en Etsy sobre la ropa que vestía y, posteriormente, encontrando su perfil en LinkedIn. Como segundo ejemplo, puede hacerse referencia a un caso que, aunque no inició ningún procedimiento judicial, fue capaz de exponer, hace 5 años, a los perpetradores de una estafa basada en hacerse pasar por equipo de soporte técnico de Norton [12]. Este último ejemplo muestra la efectividad de los agentes que son capaces de indagar en este tipo de fraudes, pero también, sus limitaciones a la hora de detener a estos cibercriminales, ya que estos actúan bajo la legislación de otro país.

A raíz de lo expuesto, resulta relevante llevar a cabo un estudio en el que quede reflejado el impacto de valerse de las fuentes abiertas actuales para fabricar inteligencia, teniendo en cuenta el trasfondo histórico y el estado del arte de estos procedimientos.

1.3. Historia breve

Existen muchos artículos en distintas revistas profesionales sobre los antecedentes de la disciplina OSINT según investigadores veteranos. Basándonos en uno de estos artículos [13], las primeras apariciones formalizadas de técnicas OSINT surgieron alrededor del comienzo de la década de los años 40, con la creación de dos grupos especializados; el Servicio de Monitorización de la BBC, en Inglaterra, 1939, y el Servicio de Monitorización de Emisiones Extranjeras (*Foreign Broadcast Monitoring Service*, FBMS) en Estados Unidos, 1941. La organización británica, de carácter más civil, no tardó en desarrollar el servicio comercial denominado como el "Compendio de Emisiones Globales"(*Summary of World Broadcasts*, SWB), que se dedicaba a registrar mensajes retransmitidos a lo largo del mundo sin importar su origen. Por otro lado, el FBMS nació de un grupo de investigación en la universidad de Princeton, y cobró gran importancia tras el ataque a Pearl Harbor por parte de la aviación bélica japonesa. Seis años después de su surgimiento, el FBMS pasó a formar parte de la Agencia Central de Inteligencia bajo el nombre de Servicio de Inteligencia de Emisiones Extranjeras (FBIS).

Siguiendo lo explicado en esta misma fuente, entre 1947 y 1948 ambas organizaciones empezaron a cooperar, lo cual fue un factor determinante en las capacidades de la inteligencia estadounidense frente a la soviética. Más adelante, en el contexto de la Guerra Fría, las fuentes abiertas eran la principal forma de obtener información relevante para los servicios de inteligencia, y así, estar al tanto de la situación actual con respecto a los adversarios del momento: intenciones políticas, competencias militares, riesgos prioritarios... Por esta razón, las fuentes abiertas acabaron por ser el primer recurso en procesos de inteligencia, que aportaban "las piezas en el marco del puzle"[13].

Por aquel entonces había una gran falta de interconexión en esta época del mundo en comparación a la existente en la época actual. Como indica un artículo de *The Intelligencer* [14], una revista oficial estadounidense, gran parte del desempeño de los agentes investigadores se valían de las fuentes abiertas mediante la recolección y disección de declaraciones de figuras gubernamentales de países ajenos, artículos, e informes de científicos extranjeros con respecto a sus avances tecnológicos. Esto permitía deducir, o como mínimo, intuir, las capacidades nucleares que iba desarrollando cada nación y sus relaciones diplomáticas. La CIA no sólo se encargaba de monitorizar de esta forma a Rusia, sino también a China y países europeos que les fueran destacables. Dentro del proceso de inteligencia que tenía lugar en la organización, cualquier tipo de información a la que pudieran acceder era bienvenida. De ahí la importancia de las fuentes abiertas, y del uso de técnicas que permitan usarlas a su favor.

Podemos afirmar que, antes de la expansión de internet tal y como lo conocemos, las primeras etapas del proceso de inteligencia de fuentes abiertas se veían obstaculizadas por barreras físicas y culturales, como la distancia entre un territorio nacional y otro o el idioma, por lo que resultaba más tedioso realizar un tratamiento de la información satisfactorio. Esto es debido a que dicha información habitualmente se obtenía de la prensa nacional y diferentes medios de comunicación tradicionales presentes por aquel entonces, según situación particular de cada país [15].

Aunque estuviesen disponible de forma pública, esto sólo significaba que los medios responsables de publicar informativos y boletines únicamente la redactarían en papel y mantendrían su disponibilidad, en una zona del mundo o medio informativo concreto, hasta que fuera archivada y hubiese que dejar paso a nuevas noticias, artículos y piezas de literatura científica. De manera previa a la gran interconexión digital, observamos estas dificultades en las fuentes abiertas de la época, situación que se vería mantenida, aunque no sin progresar, hasta principios del milenio siguiente, como podemos ver en la tipología clásica de estas fuentes. [16].

Por lo descrito en el párrafo anterior, también es necesario hablar del concepto "literatura gris", información escrita originalmente en privado en un entorno empresarial, gubernamental o académico que mantuvo o mantiene una disponi-

bilidad limitada, ya sea por la escasez de existencias, porque el material no es conocido, o porque el acceso al mismo está regulado de alguna forma. Algunos ejemplos son cuadernos de bitácora personales, borradores de estudios o investigaciones, documentos no oficiales de estado o procedimientos internos de cualquier entidad.

Según la historia ofrecida por el *OSINT Report 2010* [13], a finales de los años 80 la milicia norteamericana comenzó a emplear por primera vez en el mundo el término OSINT, al argumentar y exponer la necesidad de una reforma en los procesos de inteligencia dada la naturaleza dinámica de los requisitos relacionados con la obtención de la información del país y, sobre todo, de la creación de tácticas militares efectivas. En los años venideros, la Comunidad de Inteligencia de esta nación comenzó a volverse cada vez más consciente del potencial presente en la gran cantidad de información accesible gracias a las fuentes abiertas.

Gracias al mismo artículo sabemos que, en paralelo, la OTAN comenzó a valorar este aspecto en los procesos de inteligencia e incluso redactó diferentes guías y manuales para tratar de definir un marco para el uso de OSINT. Una vez comenzó el nuevo milenio, la Unión Europea creó el *European Media Monitor* (EMM) en 2002 como principal plataforma OSINT, gracias al Centro Común de Investigación (*Joint Research Centre*, JRC), que también estaba a cargo de otras herramientas y proyectos relacionados con las fuentes abiertas.

Resulta crucial tratar las aportaciones de la OTAN con la publicación, a finales de 2001, del *NATO OSINT Handbook*. De manera consecutiva, también se finalizaron otras piezas fundamentales de la documentación que definiría el uso de técnicas OSINT en aquella época: *NATO OSINT Reader* e *Intelligence Exploitation of the Internet*. A continuación, echaremos un vistazo a cada una de estas guías, al ser una forma sencilla de definir claras diferencias entre las fuentes abiertas existentes a finales del milenio pasado y a principios del actual, junto con las formas de aprovecharlas.

Siguiendo un orden cronológico, el *NATO OSINT Handbook* [17] fue de los primeros documentos que reconoce el potencial de las fuentes abiertas en el papel que desempeñan dentro de procesos de inteligencia, y expone la necesidad, en aquel momento, de encauzar a la OTAN en el camino de implementar estas fuentes dentro de esta rama. Fragmentos del informe como "[...] es virtualmente imposible mantener una colección viable de materiales de fuentes abiertas que den cobertura a toda la información necesaria a la vez. El foco debería estar en la recolección de fuentes, no de información." representan el cambio de perspectiva incipiente en comparación con cómo, hasta entonces, se relacionaban los agentes investigadores con sus fuentes de información. Desde un punto de vista más amplio, esta guía aportó definiciones que se usan a día de hoy, una clasificación de las fuentes abiertas existentes en el internet del momento, una jerarquía del software útil para la recolección y el procesado de información de fuentes abiertas, la primera definición del ciclo de OSINT y conceptos utilizados por el mismo, y métodos para

analizar fuentes abiertas en internet, mantenerse anónimo y agrupar los diferentes tipos de fuentes abiertas. Algunas de las definiciones mencionadas se desarrollan en la sección 3.4.

El próximo reporte, *NATO OSINT Reader* [18], se difundió en Febrero de 2002 para acompañar al *NATO OSINT Handbook* mediante una recolección de artículos y materiales de referencia creados por expertos en todo el mundo, tanto para expandir el contexto como para enumerar todos los antecedentes relevantes al ámbito OSINT. El documento comienza valiéndose de informes para tratar la historia del concepto OSINT, los tipos de fuentes que abarca, y su importancia a la hora de aplicar estrategias de recolección en todas las fuentes alcanzables en aquel momento. Esta primera parte también define con exactitud el concepto, obstáculos y valor de la literatura gris, junto con otros dos tipos de literatura más comunes: la blanca, conformada por estudios, noticias y libros identificables y accesibles hasta cierto punto; y la efímera, correspondiente a objetos con un período de vida diminuto, como podrían ser horarios impresos o notificaciones ocasionales imprimidas. Después, *NATO OSINT Reader* continúa con artículos centrados en la aplicación de la inteligencia de fuentes abiertas en casos específicos. Los más relevantes fueron: la novedosa captación global de imágenes de calidad vía satélite; puntos de vista internacionales de OSINT, que incluyen la evolución de la comunidad de inteligencia estadounidense para expandir el uso de fuentes abiertas usando, entre otras fuentes, la información proporcionada por el FBIS; y la privatización de la inteligencia en naciones europeas para una coordinación entre estos servicios de cada país a nivel continental. Por tanto, el valor del documento se basa en el compendio de ensayos por parte de diferentes agentes de referencia de sus respectivas organizaciones gubernamentales.

La última publicación fue *Intelligence Exploitation of the Internet* [19] en Octubre de 2002, más enfocada en las técnicas OSINT nacidas del internet existente a principios de siglo. Por aquel entonces, como los investigadores no tenían por qué tener destreza con equipos informáticos o conocer el funcionamiento de la enrutación de estos equipos a nivel internacional, este informe dedica el primero de sus cinco capítulos a explicar qué es la *World Wide Web* y cómo navegar, qué eran los grupos de noticias (*newsgroups*), cómo usar listados de correos electrónicos o *email lists*, y qué es un "chat", ya implementado en algunas páginas web. Con respecto a los siguientes capítulos, estos representan las diferentes fases dadas en un ciclo de inteligencia de la época: dirección, obtención, procesamiento y diseminación. Estas etapas son, a grandes rasgos, muy similares a las que describiremos en el apartado 3.2, y se describen muy detalladamente desde un punto de vista militar: el apartado de dirección trata la especificación de una misión y su ramificación en objetivos; el apartado de obtención habla sobre la planificación de las fuentes a consultar en internet, el refinamiento de búsquedas en navegadores y sus funcionalidades, motores de búsqueda y operadores para las consultas que realizan, y el concepto de *deep web*, páginas web de difícil acceso o inaccesibles usando sólamente buscadores; el capítulo de procesamiento indica

las características a analizar de cada fuente o información obtenida planteando preguntas concretas sobre lo encontrado o mediante herramientas como *whois* y *traceroute*; y por último, la diseminación habla sobre cómo compartir la inteligencia generada en un reporte, el uso de Outlook y el criterio a aplicar a la hora de decidir el nivel de secretismo a mantener.

Los documentos citados anteriormente conforman el punto de partida de la explotación de internet para encontrar fuentes abiertas. A partir de estas guías, fue posible afrontar las posibilidades que ofrecía este fenómeno y comenzar a normalizar las propiedades dinámicas de este tipo de OSINT más digital entre gobiernos acostumbrados a métodos tradicionales para ganar inteligencia u obtener información sobre asuntos externos.

En el caso de los Estados Unidos, el desarrollo de organizaciones gubernamentales que aplicaran procesos OSINT se vio tremendamente afectado por el incidente 11 de Septiembre [13]. La Comisión 11-S (*9/11 Comission*, también conocida como *National Commission on Terrorist Attacks Upon the United States*) realizó una serie de recomendaciones que, en parte, abarcaban las capacidades de inteligencia del gobierno de dicha nación, valiéndose del concepto de comunidad de inteligencia (o *Intelligence Community*) y estableciendo una jerarquía sin precedentes en este ámbito. A partir del documento que crearon, el *9/11 Comission Report*, surgió la idea de crear una entidad dedicada exclusivamente a la inteligencia de fuentes abiertas [20].

Por lo descrito anteriormente, a finales de 2005 la oficina del Director de Inteligencia Nacional anunció la creación del Centro de Fuentes Abiertas (*Open Source Center*, OSC) [21] a partir del FBIS, y, por ello, bajo el mando de la CIA. Cabe destacar que, continuando con las tareas propias del FBIS a partir de una decisión tomada en 1947, esta nueva rama de la CIA seguiría recolectando reportes y generando artículos traducidos de países extranjeros para proporcionar información accesible públicamente. Aunque sus servicios cesaron el último día de 2013, la página web del *World News Connection* crecía diariamente con este tipo de materiales, usados muy frecuentemente en ensayos e investigaciones producidos en la época [22].

Antes de que acabara 2015, el OSC se convirtió en el *Open Source Enterprise*, absorbiendo la Dirección de Innovación Digital de la CIA el 1 de Octubre como parte del plan de modernización que estaba llevando a cabo dicha agencia de inteligencia. En este caso, esta actualización no tuvo ningún cambio significativo en la estructura de la entidad o en sus funciones, y se mantiene, a día de hoy, como el organismo gubernamental de referencia en el ámbito OSINT [23].

Acercándonos más al presente, es posible analizar la situación actual a la que se ha llegado con la normalización del continuo uso de redes sociales fijándonos en el documento *"Defining Second Generation OSINT for the Defense Enterprise"* [24] emitido por *RAND Corporation* en 2018, una empresa financiada por

departamentos del gobierno estadounidense dedicada a investigaciones relativas a políticas públicas. El grado de influencia de este fenómeno mundial es claro: "El Internet y el auge de las redes sociales han hecho (el ámbito) OSINT mucho más complejo en términos de fuentes y de técnicas. Esta transformación es tan significativa que este informe defiende que debería verse como OSINT de segunda generación"[24].

El reporte citado profundiza en los aspectos más notables y novedosos que han surgido en las fuertes abiertos los últimos años a medida que la cantidad de usuarios activos de forma simultánea en internet aumentaba. Teniendo en cuenta lo escrito en párrafos anteriores, podemos resaltar varias de estas características.

- *Escisión de las fuentes abiertas en diferentes subtipos.* Al surgir tantos servicios diferentes a lo largo de la historia de internet, la situación actual nos lleva a encontrarnos con diferentes tipos de fuentes abiertas en las que indagar. Por lo general, tenemos: **contenido generado en medios de comunicación**, **literatura gris** y **contenido en redes sociales**, tanto **extensible** como **limitado** según la cantidad de texto involucrada en dicho contenido.

- *Masificación de contenido en internet.* De manera relativa a la clasificación de la información según su complejidad y componentes, la gran cantidad de contenido que se genera en pequeños intervalos de tiempo nos permite obtener datos que, de no ser por la globalización de Internet, serían imposibles de obtener. El ejemplo más sencillo es la asimilación de una gran cantidad de publicaciones realizadas en una red social extensa (como *tweets* en Twitter realizados en un rango geopolítico determinado en un específico rango de tiempo) para conocer el impacto social que ha podido tener algún evento localizado.

- *Surgimiento de diferentes técnicas de análisis de datos en redes sociales.* En línea con el punto anterior, no sólo es importante valorar la gran cantidad de información que ahora se puede tratar, sino también de qué forma se va a manejar. A consecuencia de esto, aparecen distintos procedimientos que permiten sacar conclusiones a partir de estos datos; análisis léxico o basado en palabras clave (*keywords*), análisis de opinión (*stance or sentiment analysis*), modelos de red de usuarios (*social network analysis*) y el consecuente análisis de grafos, y el *geotagging* presente en publicaciones. Estas técnicas, algunas impulsadas por el aprendizaje automático, son las bases en las que se basan las nuevas herramientas para investigaciones OSINT.

2 Objetivos

Para guiar este trabajo, en la presente sección se enumeran una serie de objetivos a cumplir durante su desarrollo. En la Subsección 2.1 se describe el objetivo principal, y en la Subsección 2.2 se denotan objetivos secundarios.

2.1. Objetivo principal

El objetivo principal de este documento es revisar la metodología establecida en el área de inteligencia para la realización de investigaciones en fuentes abiertas y aplicarla en un caso práctico que consistirá en una investigación real.

2.2. Objetivos secundarios

Para cumplir el objetivo principal enunciado en la sección anterior, se establecen los siguientes objetivos secundarios:

- Hacer un estudio sobre el estado del arte en relación a las fuentes abiertas de información y las técnicas OSINT.
- Analizar y sintetizar la metodología de investigación en fuentes abiertas para generar inteligencia.

- Identificar un caso real, similar a los ejemplificados en la Subsección 1.1, en el que las técnicas OSINT sean de utilidad a la hora de realizar una investigación.
- Aplicar la metodología establecida para llevar a cabo la investigación en el caso real identificado.
- Mediante la información obtenida durante la investigación, reflejar hallazgos relevantes derivados de dicha información para cumplir con determinados requisitos.
- Denotar una serie de recomendaciones para mitigar la exposición del sujeto investigado.
- A partir del trabajo realizado, establecer conclusiones y trabajos a futuro relevantes.

3

Estado del Arte

En esta sección se hará una revisión del estado actual de los procesos de inteligencia, que aplican investigaciones en diferentes fuentes para procesar la información necesaria.

Se van a desarrollar las siguientes subsecciones: en la Subsección 3.1, se introducen y explican este tipo de procesos; en la Subsección 3.2, se especifican las diferentes fases en las cuales tiene lugar el ciclo de inteligencia; en la Subsección 3.3, se tratan los diferentes tipos de fuentes de información a tratar en una investigación; en la Subsección 3.4, se definen los diferentes estados por los que pasa la información para especificar la síntesis llevada a cabo durante el proceso; por último, en la Subsección 3.5, se habla sobre la situación y el trato de conclusiones alcanzadas de una investigación, pues estas requieren cierto nivel de confidencialidad y discreción.

3.1. Procesos de Inteligencia

Como ya se ha expuesto, desde el siglo anterior las naciones que mantienen una constante diplomacia entre países y/o forman parte de asociaciones internacionales activas, mantienen instituciones dedicadas al ámbito de inteligencia, es decir; procesan información de la que extraer conocimiento, para así poder realizar decisiones en mayor beneficio según los objetivos particulares de cada estado. Esta última finalidad suele ser la más habitual y, aunque ha habido muchas definiciones de inteligencia, en las más aceptadas existen dos factores determinantes

que identifican este tipo de tratamiento de la información como tal: el proceso en sí mismo se compone en etapas de recolección, preprocesamiento, disección, evaluación e interpretación (o semejantes y en similar orden), y la finalidad del mismo es estar al corriente del contexto en el que se encuentra tanto uno mismo como un agente externo por extensión, como puede ser un país extranjero [25].

Por esto, se puede deducir que el proceso de inteligencia se representa como uno circular, que suele comenzar con una planificación y dirección para estipular qué conocimientos se buscan obtener, y terminar con la generación de conclusiones e interrelación de hechos que se transmite en un entorno donde existen individuos capaces de interiorizar dichas conclusiones. A partir de aquí, la siguiente iteración de este tipo de procesos se realizará basándose en estos resultados obtenidos anteriormente, de ahí que se utilice el término **ciclo de inteligencia**.

3.2. Las etapas del ciclo de inteligencia

En esta subsección describiremos tanto los componentes que conforman el ciclo de inteligencia como los sucesos previos y posteriores al mismo que se dan habitualmente, y que van provocando nuevas iteraciones de este ciclo según la necesidad de "refinar" la inteligencia obtenida y hallar respuestas más detalladas o un mayor número de respuestas [26].

Cabe destacar que la posterior descripción de este proceso no está basada en ningún estándar ni tipo de consenso; incluso algunas organizaciones estatales de inteligencia difieren en el número de etapas que presenta dicho ciclo, como es el caso del CNI [27] frente a la CIA [28], pues la entidad española de inteligencia aúna en una sola etapa el procesado y el análisis de los datos, denominada elaboración. Aunque estos sean los hechos, es factible concluir que toda corporación estructurada dedicada oficial o profesionalmente a la inteligencia sigue un proceso en forma de ciclo, como se ha indicado.

Por lo explicado anteriormente, a continuación se expone una enumeración específica para cada parte de este ciclo, cada una siempre presente, de una forma u otra, en el proceso de inteligencia de cada entidad. El ciclo se representa en el diagrama de actividad de la Figura 3.1, donde se expone cada fase junto con las actividades de cada persona incluida en el proceso debe llevar a cabo, formándose dos roles, **decisor** e **investigador**.

- **Planificación**. De manera preliminar a las etapas más técnicas, lo primero es que el decisor presente los requisitos encontrados ante los analistas o investigadores, de forma que se establezcan unas prioridades y se esquematice cómo se va a proceder en la siguiente etapa y qué se espera encontrar, en caso de que pueda darse dicha estimación. Adicionalmente, puede aportarse

más contexto a la investigación, especificarse el método de proporcionar el producto de este proceso (vía oral, por escrito o por medios digitales) y, especialmente, el plazo o plazos para satisfacer los requerimientos. A partir de este punto, el investigador comenzará a planificar la investigación a realizar basándose en las necesidades del decisor y la información a obtener y procesar, teniendo en cuenta las técnicas y capacidades a su disposición.

- **Recolección**. Una vez se tiene claro cuándo, cómo y qué se necesita, se pasa a la siguiente fase, que consiste en la explotación de las fuentes de información disponibles para la obtención de información, normalmente en forma de datos. Esto requiere identificar qué fuentes de información están disponibles, y a raíz de esto, lo más común es utilizar fuentes abiertas. Por tanto, aquí entran en juego las técnicas OSINT, que trataremos en secciones posteriores, y uno de los retos actuales en la realización de este tipo de investigaciones; el tratamiento de un gran volumen de datos. Como lo más común es que haya grandes cantidades de información disponible, lo óptimo suele ser priorizar calidad a cantidad, ya que, a más datos, más tiempo llevarán las siguientes etapas.

  Suele ser conveniente comunicarse en ciertos puntos de esta etapa con el decisor, quien planteó los requerimientos, para ser capaces de tantear la utilidad de los datos que se vayan recolectando y evitar seguir líneas de investigación que difieran demasiado de la planificación ya realizada.

- **Filtrado y procesamiento**. Cuando se considera que no se puede sacar más provecho de las fuentes de información de las que se dispone, hay que examinar los datos extraídos y juzgar tanto la fiabilidad del origen como su relevancia. Esta etapa también acoge el procesamiento de estos datos de manera que se siga un formato común que facilite la ejecución de la siguiente etapa.

  En este punto del ciclo de inteligencia pueden surgir nuevos requerimientos, generados por las características de los datos o de la investigación que los propicia.

- **Integración y análisis**. En esta fase, los analistas se valdrán de toda la información recolectada y procesada para tratar de convertir dicha información a un estado más complejo y elaborado, que conformará la inteligencia requerida según se vayan interpretando, interrelacionando, combinando y detallando datos y hechos que, en un principio y de forma aislada, puedan aparentar no tener relación entre sí.

  La experiencia y capacidad de los analistas es clave para llevar a cabo la integración mencionada, ya que resulta complicado llevarla a cabo de forma satisfactoria y depende de las etapas anteriores.

- **Producción y difusión**. Finalmente, las conclusiones alcanzadas que forman la inteligencia como producto final del proceso se documentan y se trans-

miten al decisor por la vía deseada, que debe de contar con los mecanismos de seguridad adecuados para no comprometer la confidencialidad, aspecto que también trataremos más adelante.

Tras esto, el decisor expondrá su punto de vista y valorará hasta qué punto la inteligencia generada cumple los requerimientos especificados en un principio. Por ello, una nueva iteración del ciclo de inteligencia puede ocurrir según lo que el decisor vea conveniente si es que surgen requerimientos diferentes. Esto quiere decir que esta nueva iteración puede aportar inteligencia de mejor calidad sobre unos requerimientos mejor definidos gracias a la primera iteración, o aportar inteligencia sobre un nuevo ámbito relacionado con el anteriormente investigado, con requerimientos ajustados a este nuevo ámbito de interés.

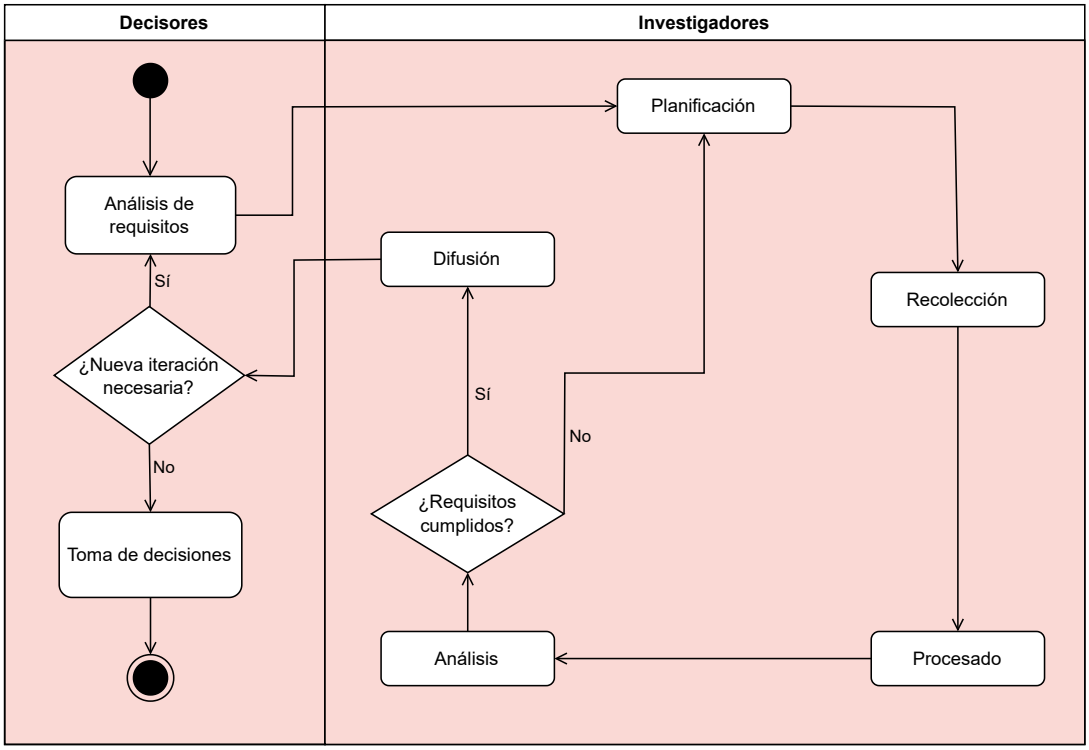

Figura 3.1: Diagrama de actividad (UML) del Ciclo de Inteligencia detallado.

Por último, es importante tener en cuenta que cada una de estas etapas no siempre tienen por qué seguirse de forma lineal, pues según vaya considerando el decisor en función de nuevos datos que vaya observando o recibiendo, los requisitos pueden adaptarse a esta nueva información y por ello podría ser necesario, por ejemplo, una segunda explotación de fuentes de información o una reconsideración de la información que está en proceso de análisis. Es responsabilidad del analista familiarizarse con estas etapas y tratar de cumplimentar los requerimientos de la mejor forma posible [26].

3.3. Tipos de fuentes de información

Situándonos en la etapa de *recolección*, el analista o investigador debe valerse de fuentes por las que sacar datos, hechos o información, y cada una puede presentar diferentes características y complejidades. Como se ha descrito en capítulos anteriores, en la actualidad resulta muy sencillo consultar fuentes abiertas de información y procesar las noticias y comunicados que encontremos de manera que se realice OSINT, pero según el contexto de la investigación a cumplimentar, cabe mencionar una serie de tipos de fuentes de información o inteligencia que, a menudo, suele complementar adecuadamente los datos que aportan las técnicas OSINT, al permitir acceso a información que no tiene por qué estar disponible para todos los usuarios.

A continuación, se presentan en la Tabla 3.1 los tipos de inteligencia, según la fuente de información en la que se base, que un investigador puede aplicar según la situación en la que se encuentre a la hora de efectuar la etapa de recolección [29].

La tabla expone tres tipos generales de fuentes de información o inteligencia: OSINT, explicado en la propia tabla; HUMINT, inteligencia creada a partir de interactuar con otras personas y según la información que poseen; y TECHINT, fuentes de información formadas gracias a dispositivos, redes o máquinas que pueden aportar diferentes tipos de datos, como se menciona en la tabla.

OSINT		La información es extraída de medios públicos y accesibles por cualquiera, sea de forma gratuita o no. Esto incluye reportajes, artículos, declaraciones, multimedia online, páginas web, redes sociales, etc.
HUMINT	White Agent	La información se obtiene a partir de relaciones interpersonales no perjudiciales sin realizar ningún tipo de acto ilegal en ningún momento.
	Black Agent	La información se obtiene a partir de espionaje, robo de secretos, infiltraciones u otras acciones ilegales o inmorales que perjudican a algún individuo o entidad.
TECHINT	IMINT	La información es obtenida gracias al análisis de imágenes o vídeo producido por cámaras de seguridad, satélite, planos aéreos, multimedia online, etc.
	SIGINT	La información es obtenida a raíz de la recopilación de señales electromagnéticas, de radio o de radar que conforman las telecomunicaciones modernas y que permiten extraer datos relevantes.
	MASINT	La información es obtenida a partir de otros métodos que no encajan en las dos anteriores definiciones, pero que resultan eficaces a la hora de hallar medidas interesantes, números identificadores, documentación personal y otras características distintivas pertenecientes a individuos o entidades clave.

Tabla 3.1: Tipos de fuentes de información [1].

La identificación de qué tipo de fuentes de inteligencia serán las más útiles a la hora de recoger información queda a disposición del investigador, quien, habitualmente, es capaz de observar estas diferentes posibilidades a medida que se va completando la etapa de *recolección*, aunque cabe destacar que no es necesaria una diversidad en la tipología de las fuentes para que dicha fase, así como el ciclo de inteligencia al completo, se realice de forma satisfactoria. Asimismo, la clasificación de las fuentes puede llegar a solaparse a menudo; IMINT, por ejemplo, podemos encontrar de forma abundante en redes sociales. En los últimos años, este último fenómeno capaz de interconectar a los usuarios de forma masiva y hacer que una gran mayoría sea consumidor de generosas cantidades de contenido publicado online ha llegado a desarrollarse tanto que muchos lo consideran una fuente de información específica dentro de las fuentes OSINT, denominado SOCMINT[30].

3.4. Estados de la información durante el ciclo

Es interesante observar, en primera persona, cómo va tomando forma el producto que se colecta, trata e interrelaciona a medida que se van sucediendo las etapas descritas con anterioridad. El ya mencionado NATO OSINT HANDBOOK [17] definió una serie de categorizaciones en los que se puede encontrar la información a medida que el ciclo de inteligencia tiene lugar, pues esta pasará por varios estados según su complejidad y grado de elaboración. De forma resumida, estos estados son los siguientes.

- **Open Source Data (OSD)**. Lo conforman los datos directa o textualmente sacados de las fuentes abiertas de información.
- **Open Source Information (OSIF)**. Se compone de una agregación de datos relacionados entre sí. El OSD es la materia prima de este estado de la información, que representa hechos o conclusiones básicas, a las que se ha llegado a partir del filtrado y la validación de dichos datos.
- **Open Source Intelligence (OSINT)**. Es la información deliberadamente descubierta, destilada y diseminada que ha sido generada para responder una "pregunta específica" (o, en el caso del ciclo de inteligencia, un requerimiento). Es el resultado de la aplicación de técnicas OSINT sobre dichas fuentes abiertas, y por ello, es a su vez producto final del proceso que nos atañe.
- **Validated Open Source Intelligence (OSINT-V)**. Será la información a la que se le puede atribuir un grado de certeza suficientemente elevado. Esto suele conseguirse mediante el respaldo de dicha información con otros datos obtenidos a partir de fuentes que no se consideran abiertas, o bien

gracias a la validación de esta inteligencia de una fuente que se considera de una confianza prácticamente incuestionable, como un superior o una organización de autoridad.

3.5. La importancia del secretismo

Dada la finalidad del ciclo de inteligencia y la necesidad presentada en los requerimientos, en muchas ocasiones la inteligencia generada precisa de mantenerse en privado, ya sea por las características de la organización a la que pertenecen los decisores o por salvaguardar la utilidad de la inteligencia generada en el proceso. Esto, si bien es común en investigaciones más clásicas cuyos analistas o agentes se valen de métodos confidenciales o clandestinos para obtener su información y confirmar la veracidad de la misma de manera igualmente opaca, también se aplica a la creación de inteligencia actual, cuyas fuentes son, en mayoría, abiertas, como ya se ha descrito [31].

En general, lo mejor para el proceso de inteligencia es mantener un nivel de secretismo tal que garantice la facilidad en la toma de decisiones que debe hacerse en la organización o entidad, una vez han recibido el "producto final" del proceso. Por ello, las entidades que se valen de estos procesos a menudo suelen defender la inteligencia que obtienen de manera similar a otro tipo de información confidencial que puedan manejar, como secretos de estado, de empresa, o proyectos.

Para poner un ejemplo sobre la confidencialidad relativa a la inteligencia generada en el ciclo, podemos observar la designación que se le otorga a la misma según la Comunidad de Inteligencia presente en Estados Unidos según el control que se tiene; existe la información privada, desconocida por la comunidad, la información compartida en los límites de dicha comunidad, e información publicada y difundida en medios accesibles públicamente [32]. Hemos de tener en cuenta que la comunidad de inteligencia estará compuesta de analistas y personas dedicadas a la investigación dentro del gobierno de esta nación; habrá equivalencias con respecto a este *paradigma del secretismo* en otros contextos a la hora de asignar niveles de confidencialidad, como en el intercambio de información entre entidades privadas (por ejemplo, el protocolo semáforo del FIRST, en el que los colores correspondientes a los semáforos se aplican a documentos internos de las organizaciones para representar qué individuos tienen permitido acceso a los mismos) [33].

4

Caso práctico

A lo largo de esta sección se va a aplicar la metodología especificada en puntos anteriores dentro de un caso práctico en el que se llevará a cabo una investigación real. Dicha investigación tendrá como objetivo general crear inteligencia gracias a aplicar el proceso correspondiente. Para ello, el caso se dividirá en diferentes apartados: primero, se hablará sobre el contexto del individuo, entidad o suceso a investigar; después, se definirán unos requerimientos siguiendo la función del decisor, y así dar paso a la investigación; en la investigación, se diferenciarán las fases de planificación, recolección, procesado y análisis, teniendo en cuenta que pueden darse diferentes iteraciones por parte del investigador antes de alcanzar la etapa de difusión; finalmente, se expondrán los resultados que recibirá el decisor, la inteligencia generada gracias al investigador. Por esto último, la etapa de difusión quedará reflejada en el subtítulo "Resultados", aunque exponer el desarrollo de la investigación forma parte de una fase de divulgación, como se especificará más adelante en dicho subtítulo.

Es importante tener en cuenta que esta investigación puede darse en contextos diferentes. Para no limitar las necesidades del decisor, y por tanto los requerimientos a especificar, van a mencionarse una serie de situaciones en las que podría ocurrir el caso dentro de su introducción. Sin embargo, con el fin de darle una aplicación a los resultados obtenidos y al tratarse de un caso real, el último subtítulo se ajustará a una situación en específico en la que la investigación se utilizará para enumerar recomendaciones que mitigarán los riesgos inherentes de tener una huella digital extensa, por lo que podría considerarse el producto de una toma de decisiones por parte del decisor. Estos riesgos se explican en los últimos párrafos de la Subsección 4.4 como respuesta a un requerimiento.

4.1. Caso práctico: Huella Digital

El caso práctico consistirá en la comprobación y descripción de la huella digital de un usuario, lo que permitirá obtener información sobre la persona detrás de las cuentas de usuario que podemos encontrar en internet, a partir de los datos que encontremos en artículos, redes sociales y otras páginas web en función del tipo de perfil se esté investigando. Gracias a esto, es posible hacerse una idea de la reputación de un usuario, el nivel de privacidad al que acostumbra, posibles vulnerabilidades en sus cuentas, personalidad, hábitos, vida social, salud, etc. [34] Este concepto se ha matizado brevemente en la Subsección 1.1.

En concreto, se va a realizar un resumen de la huella digital de un rector de una de las universidades de renombre de Madrid, por lo que se trata de una figura pública. Existen diferentes potenciales usuarios que podrían beneficiarse de obtener cierto nivel de inteligencia sobre la vida de un rector, como individuos pertenecientes a la prensa, alumnos proactivos con ideologías extremistas o resentimientos, entidades empresariales que busquen contratar a algún familiar, o cibercriminales con distintas motivaciones que busquen socavar su reputación o lucrarse económicamente.

Para mantener la privacidad del individuo investigado (de ahora en adelante, el sujeto) se han anonimizado todos los datos personales obtenidos en la investigación y, por ello, los datos que aparezcan en subsecciones consecuentes hacen referencia a personas que no existen.

4.2. Requerimientos

Para comenzar, y según la finalidad de la toma de decisiones de la que se encarga el decisor de forma posterior, los requerimientos definidos que pueden darse se limitan a asegurarse de que la información obtenida del sujeto sea de valor. Por lo tanto, una serie general de requisitos definidos sería la siguiente:

- Comprobar el grado de presencia en internet habitual del rector.
- Indagar sobre su reputación y opinión pública.
- Hallar todo tipo de datos personales disponibles.
- Evaluar el nivel de exposición a usuarios malintencionados.

4.3. Investigación

4.3.1. Planificación

En este caso y una vez se tienen los requerimientos, para el investigador resulta sencillo determinar el objetivo principal: obtener la máxima cantidad de información fiable posible sobre el sujeto. Al ser una figura pública, hay diferentes formas de hacerlo.

La más básica consiste en el uso de buscadores que permitan encontrar al rector de la universidad según artículos de periódicos digitales, cuentas en redes sociales y contenido en línea para verificar la identidad del sujeto. Al ser una universidad reconocida, no se prevee ningún obstáculo más allá de la gran cantidad de información que se espera encontrar, como es propio de la búsqueda en fuentes abiertas.

Después, habrá que realizar un procesado de las fuentes para tener en cuenta sólo las fuentes clave, las que más datos aporten según su cantidad y fiabilidad. A más fuentes se consideren desde un principio, mayor filtrado habrá que llevar a cabo, pero las fuentes que alcancen la fase de análisis serán de mayor calidad.

Posteriormente, se analizará en profundidad las fuentes seleccionadas para hacer valer la información disponible. Esto no sólo permite empezar a maquetar algunas conclusiones, si no, también, identificar más bases de las que partir para continuar en *nuevas iteraciones* y así ganar más información, como se mostrará más adelante.

Finalmente, bien se llegue al punto en el cual sea posible dar respuesta a los requerimientos de forma satisfactoria, o bien se agote la información disponible en fuentes abiertas y no se considere procedente el uso de otros tipos de fuentes, se pasará a la etapa de difusión para exponer los resultados y las conclusiones de la investigación.

4.3.2. Recolección

Como se ha explicado, por el momento en esta etapa se harán búsquedas en internet. Una consulta como "rector universidad Alfonso X el Sabio"[1] nos facilitará páginas oficiales de la propia universidad, que certifican quién, exactamente, es el gobernador de la universidad en cuestión. Acto seguido, y antes de proceder con las siguientes etapas, se continúa con datos simples como este para hacer una recolección más extensa.

[1]Se recuerda al lector que los datos mostrados en este documento no corresponden con la investigación realizada.

Búsquedas como, por ejemplo, "Ezequiel Ortega Ronda", devuelven una serie de enlaces que serán útiles más adelante; cuentas de redes sociales, páginas web donde se enumeran los proyectos y *papers* donde ha participado el sujeto así como formación y experiencia, y diferentes artículos de noticias en caso de que los hubiere. Adicionalmente, ayudarán a encontrar fuentes más específicas el uso de comillas dobles y "Google Dorks"[2], filtros asignables a búsquedas en google que, por ejemplo, ayudan a encontrar *links* a archivos de un sólo tipo o que dirigen a páginas de un único dominio.

Una vez más, a medida que se encuentran datos básicos o simples, se van utilizando en nuevas búsquedas. A su vez, todas las cuentas de redes sociales que parezcan relevantes también se guardan para la etapa siguiente.

4.3.3. Procesado

Partiendo de las fuentes halladas, en esta etapa deben de filtrarse todas aquellas que son ajenas al sujeto, ya sea por encontrar algún dato que asegure que la cuenta o la persona que trata la fuente es alguien ajeno al rector o por no existir ningún dato que lo relacione con el mismo.

Como es de esperar, dentro de las páginas web recolectadas se encuentran muchas cuentas de personas que comparten nombre y primer apellido con el sujeto (o incluso el segundo), pero cuya imagen, formación o puesto actual no encajan con el perfil que ya conocemos. Por tanto, existen muchos Ezequiel Ortega que realmente no tienen nada que ver con el rector, y debe de tenerse en cuenta que vamos a encontrarnos familiares de estos otros usuarios que podrían parecer, en primera instancia, familiares de nuestro sujeto.

Esta casuística se puede superar una vez se investigan las fuentes que sí nos aportan información verídica. Al ser una figura pública, es sencillo encontrar las redes sociales de Instagram, Twitter, Facebook e incluso Youtube del sujeto, al tener la imagen del mismo y exponerla en los contenidos de estas cuentas. Lo mismo ocurre con noticias y otro tipo de documentos oficiales con diferentes datos personales que puedan aparecer. Incluso si no tratan con datos personales, es conveniente buscar otras fuentes de información relativas a la prensa y comprobar el grado de correlación de entre ellas.

4.3.4. Análisis

Al haber realizado un filtrado, es momento de hallar la máxima cantidad de información posible de las fuentes identificadas como fiables y útiles. Por ello,

[2]Breve artículo sobre qué es Google Dorking: https://nordvpn.com/es/blog/google-hacks/

toca investigar todos los datos presentes posibles en cuentas de redes sociales, noticias de interés, y demás documentos.

En primera instancia, encontramos información más accesible; la formación y experiencia del rector, es decir, dónde se sacó su grado universitario junto con su doctorado, en qué universidades ha impartido clase y en qué momento empieza en su puesto de máxima autoridad gobernante de la universidad. También debe tenerse en cuenta que estos datos no sólo sirven para conocer al sujeto, sino también para tener un contexto temporal de cuándo se dieron estos sucesos y poder encontrar más información relativa a los mismos, directa o indirectamente. En este caso y para desgracia del rector, parece ser que hubo polémicas desde el momento de su primera candidatura relacionada con su experiencia laboral y la ley de incompatibilidades. Esto fue a raíz de entrar en un puesto, según el sujeto, no remunerado, en una empresa privada, a la vez que trabajaba en la universidad que nos atañe.

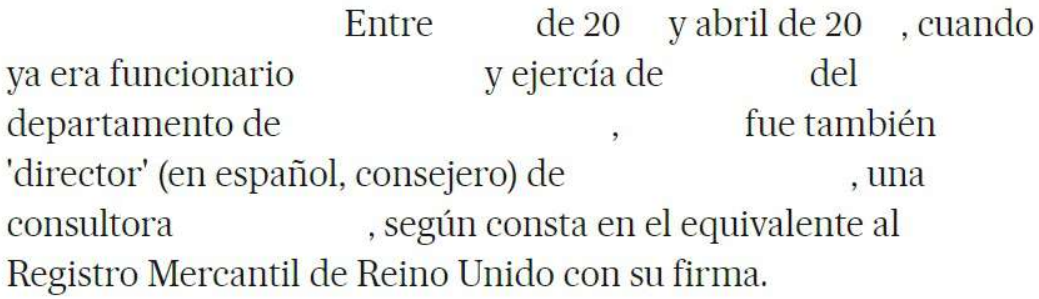
Entre de 20 y abril de 20 , cuando
ya era funcionario y ejercía de del
departamento de , fue también
'director' (en español, consejero) de , una
consultora , según consta en el equivalente al
Registro Mercantil de Reino Unido con su firma.

Figura 4.1: Fragmento censurado de la noticia que desarrolla la supuesta infracción de la ley de incompatibilidades.

En la Figura 4.1, se muestra un ejemplo de nueva información que servirá como base de la que partir en una nueva iteración, en la que se compruebe si es posible tener acceso a este documento del "Registro Mercantil del Reino Unido". No obstante, el propio artículo de periódico incluye una captura de dicho documento con muy baja resolución y parcialmente censurada, en la que aun así se exponen datos personales del sujeto como la fecha de su nacimiento.

Este artículo periodístico no es el único en contra del rector. Existen más polémicas reflejadas en noticias basadas en pruebas que, sin profundizar de más al no tratar sobre la vida personal del sujeto, podrían considerarse el origen de una reputación manchada. La posibilidad de contrastar esta idea constituye una razón más para realizar, como mínimo, una segunda iteración. Además, en este propio artículo se hace mención a la cuenta de LinkedIn del sujeto la cual, por el momento, no ha sido posible encontrar, algo muy poco común teniendo en cuenta su posición en una universidad, por lo que se tendrá también en cuenta en la siguiente iteración.

En lo que respecta a redes sociales, las características de cada una de ellas hacen factible obtener datos distintos según las publicaciones dentro de cada perfil del sujeto. Inicialmente, las cuentas encontradas en Facebook, Twitter, Instagram y Youtube parecen estar dedicadas a la divulgación de noticias de actualidad desde su posición de rector universitario, junto con otros comunicados formales. Sin embargo, según lo explicado en párrafos anteriores, estas cuentas públicas se pueden utilizar como indicador directo de su reputación, ya sea actual o a lo largo de cierto período de tiempo.

Siguiendo con el análisis más específico de estas fuentes, se observa que el canal de Youtube es el que menos información aporta; simplemente hay una serie de vídeos, de breve duración, subidos el primer día de la última campaña electoral para hablar de su programa e intenciones. Ninguno de estos vídeos posee comentarios, y la información del canal sólo es útil para confirmar sus cuentas oficiales de Instagram y Twitter, dado que, aunque se incluye una página web oficial que también se usó durante la campaña, actualmente el dominio indicado no conecta con ningún servidor.

La cuenta de Facebook del sujeto, si bien es cierto que posee un gran número de publicaciones, todas se hicieron o bien durante su primera campaña electoral, o bien durante el cuatrimestre posterior a la misma. Se encuentran un par de publicaciones interesantes, como alguna en la que incluye una captura de pantalla de un *tweet* sobre una noticia en la que se observa su proveedor de internet, u otra en la que comparte un enlace a una entrevista cuyo titular remarcaba su respuesta a las acusaciones de la segunda principal candidata al puesto. Este perfil de Facebook no comparte de forma pública nada más, y en las interacciones con sus publicaciones parece que sólo se encuentra personal de la universidad o interesados en la misma, aparentemente lejos de la vida privada del sujeto. Por desgracia, al ser una cuenta tan accesible, sí que presenta algún que otro comentario atacando al rector por alguna polémica relativa a la política del país.

En cambio, su cuenta de Twitter sí que ha tenido más uso a lo largo del tiempo, pero ningún *tweet* desde hace más de un año. Continuando el análisis sobre el contenido de la misma, y prestando atención sobre las fechas que corresponden a épocas de elecciones, las respuestas que reciben los *tweets* del rector confirman el impacto reputacional que han tenido las polémicas. Aunque Twitter expone que la cuenta tiene más de una década de antigüedad, llama la atención la inexistencia de ningún tipo de contenido hasta 2019[3]. No hay más datos destacables, pues el rector utiliza esta cuenta, al menos actualmente, con la misma finalidad que usó en su momento la cuenta de Facebook. Sin embargo, en la siguiente fase de recolección se ha de tener en cuenta la posibilidad de usar filtros disponibles en la búsqueda avanzada de Twitter[4], pues resultará sencillo encontrar más información

[3]Con la intención de proteger la identidad del sujeto, se mencionan estos datos con deliberada inexactitud.

[4]Entrada de blog que detalla estos filtros, de usabilidad similar a los ya nombrados Google

sobre controversias pasadas (nivel de gravedad de la misma y otros puntos de vista) u otras noticias que sólo se hayan divulgado por esta red social.

Concluyendo el análisis preliminar en redes sociales, la cuenta de Instagram del rector presenta características previsibles según lo analizado. Su contenido corresponde con el publicado en su cuenta de Twitter, en marco temporal y contenido, al centrarse en las campañas electorales por las que ha pasado el sujeto para promover el voto y difundir iniciativas y estadísticas positivas de la universidad. Sin embargo, el apartado de publicaciones en las que ha sido etiquetada la cuenta hace factible comprobar que la cuenta realmente parece haberse creado en 2019, al no haber publicaciones que hagan referencia a la misma antes de este año. En lo que a controversia respecta, este apartado también enseña alguna foto, subida recientemente, que ataja temas relativos a despidos de profesores y malversación de presupuestos. Esto reafirma la necesidad de seguir investigando para tener un punto de vista más informado con respecto a la proporción de usuarios que estén en contra de la posición del sujeto como rector, junto con sus motivos.

De manera resumida, la necesidad de realizar como mínimo una iteración más se hace cada vez más evidente. Más allá de las razones ya señaladas, el análisis del perfil de Instagram revela un conjunto de cuentas en las listas de seguidos y seguidores que, en principio, pertenecen a personas que comparten su dos apellidos ("Ortega Ronda"). Esto puede suponer un descubrimiento inicial sobre su familia que podrá expandirse una vez se profundice la investigación con la repetición de las fases ya aplicadas.

Antes de dar por terminada esta etapa, todavía queda hablar de los proyectos y documentos de divulgación que el rector ha llevado a cabo o contribuido a lo largo de su carrera como personal de investigación. Aunque pueda parecer algo trivial en realidad puede ser de ayuda, tanto para tener en cuenta sus hitos personales como para saber con qué otras personas ha tenido relación en algún momento, y así tener más información con la que trabajar en caso de llegar a requerir más fuentes a valorar.

La Figura 4.2 muestra una serie de apartados que resumen la actividad profesional y laboral del rector, junto con un compendio de sus investigaciones que incluye referencias a las mismas y una enumeración de los autores en cada una. Al ser una cantidad de documentos considerable, podría hacerse un recuento de las participaciones de cada autor distinto e identificar qué investigadores se repiten más, pues el sujeto podría tener una relación más personal con estos usuarios.

Dorks: https://www.tweetbinder.com/blog/twitter-advanced-search/

Figura 4.2: Desplegables en la página web del perfil del sujeto en orcid.org, censurada parcialmente.

4.3.5. Segunda iteración

Estos últimos apartados no constituyen una etapa más, sino la concatenación de un conjunto de ellas que se dan para obtener resultados de mayor calidad y ganar inteligencia sobre la vida del sujeto, no sólo información. Por esta razón, aunque se van a aplicar de nuevo las etapas del ciclo de inteligencia que ya se han definido, esta aplicación no se expondrá de forma tan remarcada, ya que se basará en (y quedará limitada por) la información encontrada en una primera iteración.

En la fase inicial de la segunda iteración se va a **planificar** qué fuentes van a utilizarse, según lo mencionado en el anterior análisis. Ha de tenerse en cuenta: la existencia de casos tratados por la prensa que dañan la reputación del sujeto y que se basan en pruebas como su cuenta de LinkedIn o los registros oficiales de Reino Unido; la búsqueda necesaria de menciones del rector en Twitter para comprobar comentarios y acusaciones de otros usuarios; y el haber hallado cuentas de potenciales familiares en Instagram, quienes indican sus nombres y apellidos. Una vez enumerados estos factores, la investigación queda dividida por dos finalidades principales que requerirán desarrollos apartados. Por un lado, habrá de tratarse con la información sobre casos controversiales, y por otro, con la información relativa a miembros de su familia, lo que da lugar a fases más complejas. Es por esto que primero se va a exponer la investigación del primer fenómeno, y después el desarrollo del segundo.

Comenzando con la **recolección** de las noticias importantes, lo más sencillo es buscar en Twitter qué personas se han dirigido al rector en la primera temporada electoral en la que participó, una vez se lleve a cabo una búsqueda más general en internet. La anterior fase de recolección y de procesado parece indicar que su nombre empezó a cobrar más popularidad una vez se presentó a candidato a rector, y no sólo a raíz de este hecho, sino también a causa de los artículos de prensa que surgieron. El artículo de la Figura 4.1 es una de las primeras noticias

polémicas publicadas de forma previa a las elecciones. A medida que se acercaba el recuento de votos, se fueron publicando más artículos dada la situación en la que se estaba eligiendo al nuevo rector; el catedrático en el puesto de rector de por aquel entonces ya sufría de una fama negativa, encontrándose nuestro sujeto en una situación con alta probabilidad de ser su relevo. Esto pareció provocar que cierto medio de comunicación abordara determinados asuntos que evidencian malas prácticas por parte del sujeto y exponía su candidatura como una fabricada a través de una estrecha relación personal con el antiguo rector, la cual quedó evidenciada por otros artículos hasta cierto punto. El número de noticias de este medio de comunicación en contra del rector, publicados en internet, lo identifican como el principal origen, entre otros, de una reputación dañada. A causa de esto, pocos días antes del cierre de plazo para presentar candidatura a rector se presentó un segundo candidato como oposición principal del sujeto.

Adicionalmente, estos hechos se reflejan en una serie de *tweets* accesibles de forma pública. Antes de continuar, es interesante matizar que el sujeto se valía de una cuenta secundaria, actualmente borrada, para realizar su primera campaña electoral en Twitter. Esto significa que, de no haber realizado una investigación previa (en este caso, una primera iteración), no es posible encontrar la cantidad de comentarios dirigidos a su persona durante un período de tanta crispación. A través de una simple consulta, es posible encontrar algunos usuarios opuestos a las intenciones del sujeto de alcanzar la posición de rector. La Figura 4.3 muestra una publicación clave que no sólo ha permitido identificar la cuenta de Twitter del segundo candidato, sino también la solidaridad con el mismo de una figura opuesta al sujeto.

Figura 4.3: Captura de la consulta realizada y sus resultados. Censurada parcialmente y datos falsificados.

Con consultas similares es posible comprobar los *tweets* del segundo candidato por aquel entonces dirigiéndose al actual rector, en los que comparte algún enlace en referencia a una noticia controversial del medio de comunicación ya mencionado o incluso muestra una captura de pantalla de la Junta Electoral concediéndole

un debate con el sujeto, el cual ignoró. Finalmente, el rector salió elegido con minoría de votos en total, pero con mayoría de ellos entre los grupos cuyo voto tiene mayor ponderación.

Esta etapa de **recolección** ha sido más extensa al haberse tratado de un contexto más alejado del presente y tener una casuística más complicada, ya que las fuentes tratadas comparten información basándose en un contexto previo. Para llegar a explicar lo anterior ha sido necesario ir encauzando la recopilación mediante un **procesado** que permitiese al investigador filtrar cualquier hecho divulgado que se aleje demasiado de los requisitos que se han definido. A la vez, también se han realizado más consultas relativas a la cuenta de LinkedIn y el registro oficial de Reino Unido en las que igualmente ha sido necesario aplicar un filtrado, pues las publicaciones de actualidad en esta red social mencionan al sujeto sin etiquetarle y no se sabía si en un principio este tipo de documentos son accesibles por cualquier usuario. Tras indagar, se ha encontrado una cuenta de Linkedin[5] y, además, se ha hallado una página web bajo el dominio find-and-update.company-information.service.gov.uk que expone la aplicación del sujeto a la empresa indicada en la Figura 4.1, ambas fuentes a analizar a continuación.

Para terminar esta primera parte de la segunda iteración con la etapa de **análisis**, se tratará todo lo explicado anteriormente. Siguiendo con al recién descubierta cuenta de LinkedIn, esta no tiene foto de perfil, foto de fondo, segundo apellido, descripción o actividades[6], pero sí una biografía en la que el sujeto indica su posición de rector en la universidad correcta y una serie de datos en el apartado de información y experiencia que encaja con lo que ya se conocía en la iteración previa. Adicionalmente y como prueba de mayor valor, se sabe que varios de sus contactos son profesores de la universidad a la que pertenece.

Sobre la página web del gobierno británico que muestra la aplicación del sujeto al puesto de director de cierta empresa, en esta se expone información que incluso el artículo original se encargó de censurar, indicada en la Figura 4.4.

Observando la dirección indicada, que lleva a una urbanización en la mitad norte de la comunidad, se considera bastante probable que sea la correcta y que el rector continúe viviendo allí. Este dato podrá ser útil en la segunda parte de esta iteración.

Por último, para tener un punto de vista al día de la reputación del sujeto, se va a analizar si las cuentas encontradas en la fase de recolección siguen publicando *tweets* en contra del mismo, incluyendo su propia cuenta. Esto ayudará a encontrar polémicas más actuales, para hacer una medición aproximada sobre la frecuencia con la que recibe comentarios negativos. Por ir en orden de valor

[5]Gracias a la búsqueda concreta "UAX Ezequiel Ortega Ronda site:linkedin.com".

[6]Las características del perfil prueban lo difícil que resulta encontrar la cuenta del rector, posible causa de que no se le etiquete nunca. También es posible que el sujeto tenga desactivada esta la opción de forma intencional en los ajustes de su cuenta.

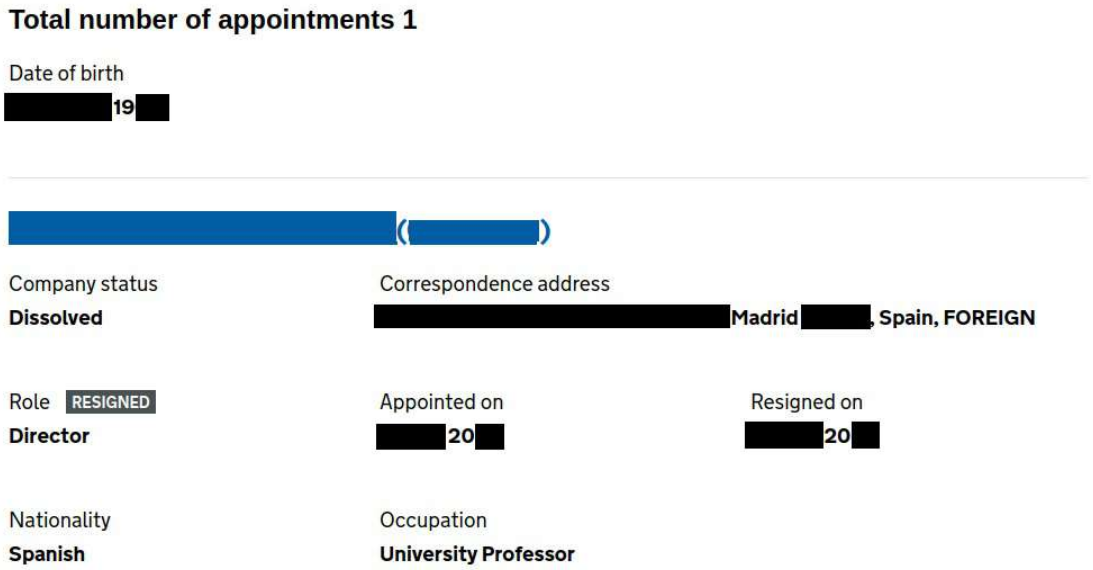

Figura 4.4: Datos expuestos públicamente en la página web del gobierno británico.

aportado a la descripción de su huella digital, puede decirse que, desde 2023 inclusive, el rector ha sufrido cada año una polémica diferente que han incitado malas palabras sobre su persona, pero no a raíz de un odio injustificado (como suele ocurrir en redes sociales de cualquier índole con figuras públicas) sino a partir de sus decisiones. Con el fin de no comprometer el ejercicio de anonimización que se ha implementado en este primer caso, no se hará mención de ningún tema controversial de actualidad, aunque se ha de denotar que Twitter es la red social que refleja, como mínimo, cierto grado de repulsión hacia el rector en mayor parte por el alumnado. En la última campaña electoral, el usuario que publicó el *tweet* mostrado en la Figura 4.3 respondió a una de las cuentas dedicada a tratar este tipo de temas, que expuso un caso de malversación de fondos. Este hilo de Twitter encontrado gracias a la respuesta de este usuario, trata otro caso más, en contra de la imagen del rector, en la que se alega que está involucrada su esposa. Al mostrar documentos oficiales, también se enseña el nombre de la mujer, por lo que es posible llevar a cabo una investigación sobre si realmente mantiene esta relación con el sujeto. Una vez más, esto es razón para realizar una iteración más.

Hasta aquí concluye la primera parte de la segunda iteración y comienza el desarrollo de la segunda parte, que trata sobre los familiares del sujeto gracias a las cuentas observadas en su Instagram y se ha llevado a cabo de forma paralela.

Durante la fase de **recolección**, se han hecho consultas en motores de búsqueda usando el nombre y apellidos de las personas que parecían tener relación con el sujeto según lo descubierto anteriormente. En esta línea de investigación es más simple separar esta etapa de la siguiente; antes de filtrar, se encuentran principal-

mente cuentas de Facebook de posibles parientes del rector gracias a la igualdad de apellidos, aunque también se ha de tener en cuenta las propias cuentas de Instagram que han permitido su identificación. Esto es gracias a poder ver, en muchas de estas cuentas, a los amigos añadidos en Facebook, que coinciden con los seguidos/seguidores en las cuentas de Instagram en el caso de los hermanos del sujeto en los que es posible comprobar esto, al mantener sus cuentas públicas. Adicionalmente, en las listas de amigos de Facebook de los supuestos hermanos también se hallan un par de perfiles de "Ezequiel Ortega" que no habían aparecido en la iteración anterior, y que, en un principio, parecen pertenecer al sujeto en cuestión. También se descubren más posibles parientes ostentando el apellido "Ortega", "Ronda" o ambos.

El **procesado** de las cuentas encontradas no resulta difícil de llevar a cabo. Muchas de las cuentas demuestran ser relevantes gracias a compartir su foto de perfil con su cuenta de Instagram relacionada con el rector, y a pertenecer a la red de amigos o seguir a otras cuentas relacionadas con el sujeto. Esto quiere decir que ha sido sencillo dejar de considerar a usuarios que realmente son ajenos al sujeto al no tener relación con absolutamente ninguna de las personas que se considera que sí la tienen. En otras investigaciones, especialmente en las que el sujeto no es una figura pública o en las que este utiliza configuraciones de privacidad efectivas, pueden surgir muchos más obstáculos o incluso ser imposible llegar a filtrar de forma tan satisfactoria.

Alcanzada la etapa de **análisis**, obtener información que nos pueda propiciar estas cuentas secundarias del rector es prioritario. Desde un principio no sólo se confirma al sujeto como el propietario gracias a las imágenes publicadas, sino también a causa del puesto de trabajo indicado. Una de las cuentas mantiene pública su lista de amigos y en ella aparece uno de los hermanos, mientras que la otra cuenta tiene publicaciones antiguas más personales y, especialmente, una prueba de parentesco con otros dos usuarios, que no fue posible terminar de descartar dado que sus cuentas de Instagram presentaban relación con otros hermanos.

Como puede verse en la Figura 4.5, esto sería prueba suficiente para asignar parentesco a Raúl, Fernando y Elisa[7], lo que aumenta significativamente el alcance de nuestra fase de análisis al poder usar un punto de vista más informado sobre estos usuarios. Por suerte, tanto el tío del sujeto como su prima mantienen una lista de amigos pública, lo que ayuda a identificar más miembros de la familia. El procedimiento a aplicar es idéntico a las dos fases anteriores, pero con extensión mucho más reducida al estar limitándonos a trazar la huella digital del sujeto. Asimismo, aunque Raúl no sigue a su hermano por Instagram ni mantiene una lista pública de amigos, ya se le encontró en esta red social al seguir a otro hermano de Ezequiel, teniendo ambos cuenta pública.

[7]Estos nombres falsos se utilizarán más adelante para mostrar parte de la huella digital anonimizada en el apartado de resultados.

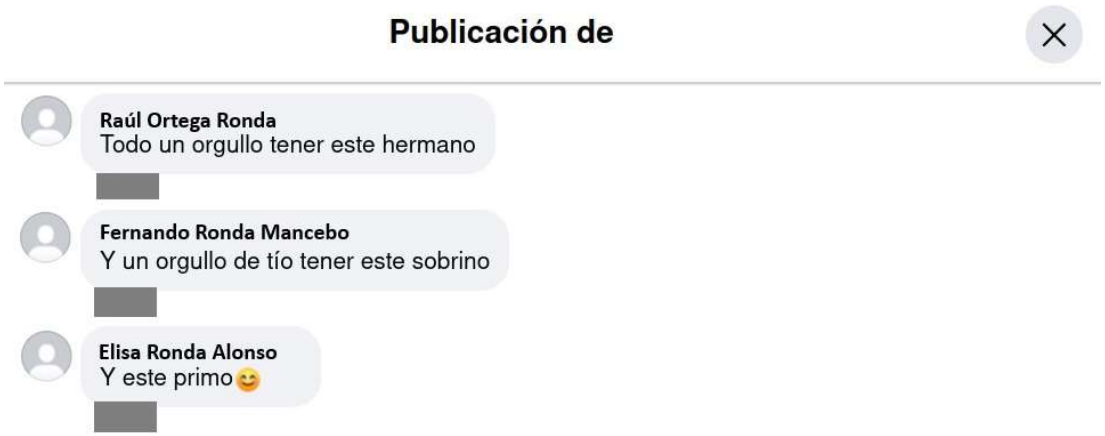

Figura 4.5: Comentarios encontrados en una publicación del sujeto, con fotos de perfil editadas y nombres falsos asignados.

Durante el análisis de la familia del sujeto, la información en las cuentas de Facebook también nos permiten encontrar cónyuges de los hermanos a través de las listas de amigos. Al ser una red social en la que los apellidos son públicos, se identifican como mínimo dos hermanos políticos gracias a haber encontrado cuentas de usuarios más jóvenes que poseen uno de los dos apellidos del rector, es decir, sobrinos del sujeto. La Figura 4.6 ilustra de forma más sencilla este hallazgo, que ha sido posible replicar en parte con otro hermano.

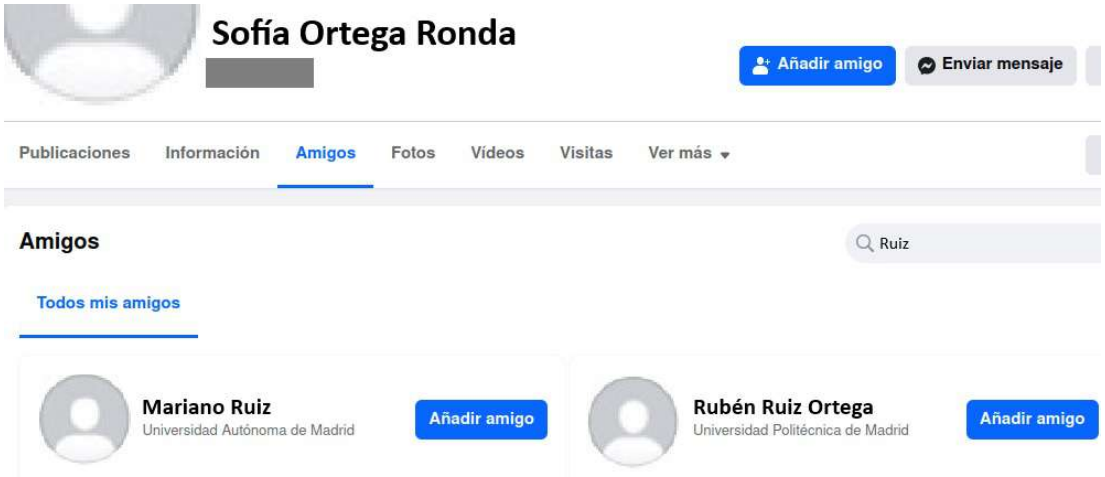

Figura 4.6: Lista de amigos de una hermana del sujeto en la que se enseña un posible hijo y marido.

Aunque la correspondencia de los apellidos es demasiado precisa para ser coincidencia, comprobar si verdaderamente existe una relación entre estos usuarios no es complicado. El sobrino es amigo del sujeto en Facebook, y el cónyuge de

Sofía tiene publicadas fotos de un mismo viaje que encaja con las publicaciones de su mujer y el marco temporal del mismo. Este es el tipo de averiguaciones que nos permiten ir esbozando el árbol familiar del sujeto valiéndonos sólo de fuentes abiertas y, aunque se ha continuado la investigación con otros familiares, no se ha procedido de forma distinta a lo ya expuesto, por lo que no se incluirá en esta memoria. Sin embargo, todas las relaciones encontradas en redes sociales se muestran en la Figura A.1 dentro del apéndice al final de esta memoria, con nombres anonimizados.

La complejidad del grafo en la Figura A.1 en cantidad de aristas, junto con los apellidos coincidentes, es prueba suficiente para dar por segura la identificación de una serie de individuos como familia del rector. En el apartado de resultados, la Figura A.2 representa todos los individuos de la familia del rector identificados junto con las fuentes que han hecho esto posible, utilizando los nombres ya asignados.

Por último, se ha encontrado una cuenta secundaria en Instagram del rector que no ha podido analizarse, al ser privada. Aun así, ha sido de ayuda para filtrar qué cuentas tienen relación directa con el sujeto y qué cuentas no, ya que aparece en la lista de seguidos/seguidores de los perfiles de Instagram de interés. Además, uno de los familiares del sujeto, que aparece en una de las publicaciones de una de sus cuentas de Facebook secundarias, indica que procede del área de Madrid coincidente con la dirección descubierta en Figura 4.4; aunque se desconoce si es su hermano menor o su hijo, la antigüedad de la foto adjuntada en la publicación hace mucho más probable una relación fraternal.

Para dar por finalizada la segunda iteración, se debe considerar si existe alguna otra fuente que no se haya consultado todavía para complementar toda la información adquirida. Al haber encontrado tantas cuentas del sujeto, una iteración consecuente podría dedicarse a indagar sobre posibles archivos generados por brechas de datos en las redes sociales que ha utilizado el rector. Estas brechas o *leaks* a menudo recogen muchos datos privados de los perfiles que incluyen datos personales y, normalmente, suelen estar disponibles en el "internet profundo" o canales privados de aplicaciones de mensajería como Discord o Telegram[8].

4.3.6. Tercera iteración

Esta tercera iteración se desarrollará de la misma forma que la iteración anterior, partiendo de una primera etapa de **planificación** que enumere los factores que se quedaron pendientes por tratar y la línea de actuación a seguir para continuar la investigación en base a estos datos. En principio, el punto de partida

[8]Según la definición escrita en la Sección 1.1, estas fuentes son abiertas siempre que cualquiera pueda acceder a ellas. Por ello, algunos canales privados también son potenciales fuentes abiertas en caso de que el acceso al mismo no esté lo suficientemente restringido.

en este caso es de menor complejidad pero compuesto también de dos factores independientes; utilizar el nombre de la supuesta mujer del sujeto para encontrar más información, y la búsqueda de *leaks* que sean accesibles y proporcionen información adicional sobre las cuentas del rector.

En la fase de **recolección** se procederá de forma similar con la esposa del sujeto, realizando consultas en motores de búsqueda que permita tener disponibles un gran número de fuentes a las cuales habrá que aplicar un **procesado** y descartar las que aporten información ajena a este usuario. También, se buscarán eventos recientes en los que se hayan robado datos de las cuentas de las redes sociales[9] que se han tenido en cuenta durante la investigación, así como enlaces de dominio *onion*[10] y otros recursos obtenidos en eventos y conferencias de carácter público. La etapa de *procesado* también es importante en esta segunda parte de la recolección, pues muchas fuentes encontradas no estarán relacionadas con archivos generados a partir de brechas de datos.

Para explicar de forma resumida el desarrollo de estas fases, a la esposa del sujeto se le ha asignado deliberadamente el nombre "María del Mar Manzano Manzano" para exponer algunas de las dificultades nuevas que se han encontrado en la búsqueda por internet. Un nombre así puede escribirse de varias formas (como "Marimar" o simplemente "Mar") y esta característica, junto al segundo apellido repetido, puede ser un obstáculo en el momento de encontrar fuentes de información. Aun así, no se han encontrado cuentas en redes sociales de la usuaria más allá del LinkedIn, donde se presenta con un diminutivo equivalente a los ejemplos expuestos. Todas las fuentes halladas son documentos de carácter público relativos a su posición dentro de la universidad, que corresponde con la que dirige el sujeto. Por suerte, también se ha encontrado un *post* de hace más de una década en un blog de un instituto. Con respecto a los *leaks*, se ha utilizado una invitación a un canal de Telegram que se mostró en un evento[11] en el que se incluía un ejemplo de un canal privado que difundía archivos masivos con datos de usuarios de páginas web y redes sociales. Excluyendo brechas de datos no relevantes, se han obtenido las suficientes como para obtener más datos personales de cuentas del sujeto en algunas de sus redes sociales principales, como se mostrará en la siguiente etapa.

Por último, la etapa de **análisis** de la tercera iteración abordará la información encontrada en las fuentes relacionadas con la posible esposa del sujeto y en los

[9]Página web que permite comprobar si tu correo se encuentra en algún *leak* y, por tanto, informa sobre la ocurrencia de dicha brecha de datos: https://haveibeenpwned.com

[10]Este dominio es accesible mediante conexión a la red anónima Tor, y las páginas web alojadas en estos servidores no pueden encontrarse mediante un motor de búsqueda, al menos no fácilmente. Más información en su página oficial: https://support.torproject.org/es/

[11]Las charlas organizadas por Cryptored, disponibles en su página web (https://www.criptored.es/criptoredtalks/index.html) se celebraron por primera vez el 14 de Junio de 2024 de forma gratuita en el campus de Fuenlabrada en la URJC. Allí, se compartió un enlace en una de las exposiciones.

archivos de las brechas de datos. El perfil de LinkedIn de la usuaria en cuestión es muy similar al correspondiente al sujeto; sólo incluye su puesto de trabajo actual cuánto tiempo lleva trabajando en el mismo, sin añadir siquiera formación o más experiencia. Es posible saber que se trata de la misma persona dada esta información, pues, en las páginas oficiales de la universidad, se muestra tanto su nombre completo como correo corporativo, así como la imagen de la usuaria en diferentes vídeos formativos sobre la infraestructura digital disponible para los alumnos. Gracias a esto, en la publicación en el blog mencionada anteriormente se la identifica a la perfección al haber incluido una foto suya. Esta publicación es importante, dado que la docente habla sobre su descendencia junto con la edad de esta pero no menciona a quien debería ser su marido, el sujeto. Por tanto, con la información actual no es posible afirmar que se haya identificado a la mujer del rector por falta de pruebas.

En lo que a los *leaks* respecta, los archivos de interés recuperados de tamaño considerable[12] no contenían información de las cuentas del sujeto, a excepción de uno que enumeraba cuentas de Facebook. En este archivo, se mostraba el número de teléfono de cada cuenta enumerada, junto con toda la información disponible públicamente, en nuestro caso ya conocida. Esto quiere decir que se ha encontrado el número de teléfono móvil del sujeto y, dado que la cuenta de Facebook al que pertenece es una de las cuentas encontradas en la segunda iteración, es muy probable que se trate de su número personal. La forma más fácil de comprobarlo es mediante Whatsapp, pues basta con añadir el número a contactos para revelar su perfil. Como se muestra en la Figura 4.7, su foto de perfil nos confirma que se trata del número del rector, aunque se ha censurado de forma deliberada esta imagen para no mostrar su rostro.

Antes de pasar a las conclusiones, destacar que se ha decidido no considerar más factores para continuar en más iteraciones dada la extensión de este caso práctico. Sin embargo, investigaciones realizadas con un contexto específico en el que el decisor tiene un objetivo claro suelen valerse de más tipos de fuentes y estrategias con las que obtener información, que pueden implicar la creación de perfiles falsos que intenten hacerse pasar por personas que no existen o, directamente, la interacción con personas relevantes que puedan suponer una fuente de información de calidad[13].

[12]Ha sido necesario el uso de Notepad++ para buscar texto en estos archivos, al tratarse con tamaños del orden de varios gigabytes.

[13]Se recuerda al lector la Subsección 3.3, donde se expone una clasificación de fuentes general. Los tipos de fuentes indicados son independientes de las fuentes abiertas; por ejemplo, fuentes HUMINT no tienen por qué ser OSINT.

Figura 4.7: Perfil en Whatsapp del sujeto, con una foto de él como única información disponible.

4.4. Resultados

En este apartado, la parte técnica de la investigación queda finalizada y se procede a realizar una **difusión** de las conclusiones alcanzadas según los requerimientos definidos. Parte de este proceso de divulgación la conforma el desarrollo de la investigación en esta memoria, y la anonimización del sujeto es una protección a su privacidad añadida por las características de este documento al estar disponible al público[14]. Teniendo en cuenta esto, se van a repasar los requisitos y asignar el conjunto de información que se ha obtenido a cada uno de ellos, para mostrar el valor que ha tenido la investigación.

El primer requisito consistía en comprobar el **grado de presencia habitual en internet** del rector. Esto puede aportar información útil sobre la implicación del sujeto en sus redes sociales, generando contenido que es susceptible de análisis, como se ha expuesto en la memoria.

Como el significado de este requerimiento yace en la publicación de contenido en línea del rector, se tienen en cuenta los diferentes perfiles que se han hallado en la investigación junto con el uso que se ha observado de los mismos:

- **Perfiles dedicados a campañas electorales**. Estas cuentas sólo han sido utilizadas por el sujeto durante el período electoral, para promover una buena imagen y su programa de gestión universitaria. Aquí se incluye Youtube

[14] Al tratarse de una figura pública, resulta imposible proteger al completo su identidad sin perjudicar la transparencia de la investigación.

y una cuenta borrada de Twitter, de la que persisten *tweets* en los que se etiqueta.

- **Perfiles profesionales**. Son las cuentas que, más allá de utilizarse en períodos electorales, también publica otro tipo de eventos y noticias de actualidad relacionadas con la universidad que el rector considera relevantes. Estas son las cuentas activas de Instagram, Twitter y Facebook, junto con la cuenta inactiva de LinkedIn.

- **Perfiles personales**. Cuentas que el sujeto no expone o no relaciona con su vida profesional de forma directa. En este ámbito se encuentra su cuenta secundaria de Instagram, junto con sus dos cuentas adicionales de Facebook.

El factor más destacable es la ausencia de actividad en cuanto a la publicación de contenido. Siendo la cuenta de Twitter la que más publicaciones posee y de más actualidad, desde esta se tuiteó por última vez hace más de año y medio. Los otros perfiles profesionales contienen un repunte de actividad muy notorio en período de campaña electoral, aunque también se utilicen como vía de anunciar noticias y eventos. Según el análisis realizado, se prevé que el rector vuelva a publicar activamente, una vez más, cuando el cargo de rector vuelva a someterse a elecciones.

El segundo requisito trataba de indagar sobre su **reputación y opinión pública**. Por no comprometer el grado de anonimización que se ha establecido en la memoria, no es posible incluir el desarrollo de las fuentes de prensa que originan una reputación manchada o que reflejan algunas de las decisiones del rector que no han sido bien vistas por los alumnos de la universidad. Más allá de esto, además se han tenido en cuenta los comentarios dirigidos al sujeto por redes sociales, cuya cantidad determina cierto nivel de negatividad de la opinión pública.

Todos los artículos que hacen referencias negativas a la figura del rector de la universidad, y por tanto, conformen parcialmente el origen de una mala imagen, se reparten a lo largo de los años en los que el sujeto ha ocupado esta posición administrativa e, incluso, antes de haber sido elegido rector. Más allá de esto, ciertas decisiones que afectaron a los alumnos y para las que no se ha tenido en cuenta la opinión de sus máximos representantes también han perjudicado la reputación del sujeto. Por todo lo hallado, filtrado y analizado, es posible suponer que existen personas en contra del rector mayormente entre el alumnado de la universidad, pero también, de forma mucho más minoritaria, entre ciertos docentes.

El siguiente requisito radicaba en la obtención de **todo tipos de datos personales** disponibles. Una vez más, el grado de anonimización que se desea mantener en este documento dificulta la especificación de detalles, pero sí es posible

comentar que se ha obtenido su nombre completo, tres correos electrónicos utilizados en diferentes cuentas profesionales (incluyendo su correo personal institucional de dominio universitario), su fecha de nacimiento, su número de teléfono, y una probable dirección de su domicilio. Adicionalmente, y mediante las consultas habituales explicadas en el apartado anterior, también se encontró su DNI en un archivo PDF disponible públicamente. Las fuentes abiertas también nos permiten enumerar una linea temporal básica en la que se muestre su formación y su experiencia.

En relación a este requisito, también es importante añadir la información encontrada relativa a su vida personal. La Figura A.2, incluida en el apéndice al final del documento, muestra un diagrama en forma de árbol representativo de todos los individuos pertenecientes a la familia cercana al sujeto gracias a las fuentes abiertas, indicando también las redes sociales en las que se han encontrado cuentas de dichos usuarios. También se utilizan los nombres asignados en el desarrollo de la investigación y, en caso de no tener pruebas suficientes para conocer la relación del sujeto con alguno de los usuarios, se indica con un signo de interrogación.

Como puede observarse en la figura mencionada, la información sobre la vida del rector es bastante completa. Dependiendo de los requisitos de la investigación, podría haberse enfocado a obtener una huella digital detallada de los familiares del sujeto y así intentar expandir este diagrama, aunque sería un proceso más complicado al no haber figuras públicas emparentadas con el rector, en principio.

Toda esta información está directamente relacionada con el último requisito, donde se pide evaluar el **nivel de exposición a usuarios malintencionados**. En este caso, el término "usuarios malintencionados" engloba a algunos agentes ya mencionados en la introducción al caso práctico de forma previa a la definición de requerimientos, como alumnos proactivos en contra del rector o cibercriminales que intenten engañar al sujeto ya sea para obtener más información o beneficio económico. Estos usuarios maliciosos aprovecharán las fuentes encontradas a lo largo de nuestra investigación para generar inteligencia que les permita aumentar sus probabilidades de éxito en la toma de decisiones.

Para hacer un reporte que aporte valor a este punto definido por el decisor genérico que se ha establecido en el caso práctico, es necesario cambiar el punto de vista con el que se estima la utilidad de la información obtenida en la investigación. Este nuevo enfoque sugiere que los datos personales del sujeto, al ser una figura pública, conforman una superficie de exposición cuya principal consecuencia son la materialización de riesgos causados por amenazas ejemplificadas en el apartado anterior. El hecho de ser una figura pública siempre conlleva la exposición de algunos de tus datos personales básicos como el nombre, los apellidos, la formación o el puesto de trabajo, pero estos riesgos quedan potenciados por factores que tratan los tres requerimientos anteriores; la actividad que mantenga el sujeto en redes sociales, su reputación, y los datos personales adicionales

disponibles públicamente.

Por tanto, a mayor superficie de exposición, mayor riesgo de sufrir algún daño directo o indirecto de un usuario malicioso. En el caso del rector, su superficie de exposición quedaría definida no sólo por los datos personales básicos, en los que podríamos incluir los correos electrónicos encontrados ya que son formales, sino también por su fecha de nacimiento, DNI, teléfono personal, posible dirección, y miembros de su familia identificados, en respectivo orden de sensibilidad de los datos. Afortunadamente, el rector no es demasiado activo con sus cuentas profesionales en redes, por lo que es más difícil que su reputación sufra más daños o que aumenten las probabilidades de exponer más información personal en función de lo mostrado en las publicaciones. Sin embargo, según lo analizado en este documento relacionado con su imagen, el rector no posee una opinión pública favorable, lo que hace más probable que haya más usuarios descontentos con el rector y busquen realizar este tipo de investigaciones para usar estos datos en su contra, es decir, sacar provecho de la superficie de exposición.

Finalmente, evaluar este tipo de riesgos conduce a pensar en las diferentes formas en las que se le podría dar un mal uso a los datos expuestos. Alguien podría investigar la dirección encontrada y descubrir, de forma inmoral o ilegal, si es verídica y de qué tipo de domicilio se trata. Se podría, también, llegar a extremos de monitorizar no sólo sus redes sociales y eventos públicos en los que haga aparición el rector, sino también las cuentas de su familia, algo bastante peligroso en caso de que las publicaciones de sus familiares puedan llegar a indicar, por ejemplo, que el sujeto está en período de vacaciones, lejos de su casa; esta información es de interés para delincuentes que les pueda interesar realizar un robo en el supuesto domicilio del rector.

Más allá del mal uso que se le podría dar a su DNI o número de teléfono, hoy en día resulta fácil crearse un perfil falso en cualquier red social que se haga pasar por una persona inexistente con imágenes y datos totalmente imposibles de probar como ficticios. Estos perfiles falsos quizá llegarían a tener éxito infiltrándose en las cuentas privadas del rector (o, una vez más, sus familiares), accediendo a fuentes con información más delicada. Una amenaza capaz de ganar inteligencia del modo mostrado en este caso práctico es, a su vez, capaz de poner en peligro los ahorros económicos, reputación o incluso integridad del sujeto con mucha mayor facilidad que a cualquier otra figura pública con una superficie de exposición más reducida.

Por lo expuesto en el anterior párrafo en respuesta al último requisito, es interesante plantear la situación en la que un decisor busca examinar la huella digital del rector para aconsejar al sujeto y mitigar los riesgos explicados. Normalmente, los usuarios no tienen en cuenta la superficie de exposición descrita anteriormente ni las formas en las que podría aprovecharse por otro usuario malicioso, algo que fomenta esta exposición. Por ello, podría resultar de ayuda informar de estos resultados, así como de las contramedidas que es posible

implementar para aumentar la privacidad de los datos disponibles públicamente.

4.5. Recomendaciones y mitigaciones

Con la finalidad de reducir la superficie de exposición, a continuación se indica una serie de directrices relacionadas con los resultados obtenidos en la investigación para prevenir o dificultar una recolección tan eficaz de información personal y mitigar los riesgos que conlleva dicha exposición. Para ello, se van a utilizar las respuestas dadas a los tres primeros requisitos definidos en la investigación, factores clave que definen hasta qué punto es accesible la vida personal del sujeto sin formar parte de la misma.

Desde un punto de vista de la ciberseguridad, el **grado de presencia habitual en internet** del rector es adecuado, sin compartir demasiados detalles de sus actividades actuales o recientes y manteniendo un uso responsable de sus redes sociales, algo que no resulta incompatible con mantener un uso profesional. Por tanto, no se necesita una recomendación relativa al contenido generado por el sujeto en internet.

Algo similar ocurre con su **reputación**; al no tener un control directo sobre este factor más allá de las acciones que pueda llevar a cabo el propio sujeto fuera de internet, las únicas medidas a tomar desde el punto de vista de la seguridad informática se encuentran dentro del litigio y denuncia de comentarios despectivos o pertenecientes a posibles campañas de desinformación. Esto es algo más difícil de denunciar si se trata de artículos o noticias por los derechos a la libertad de prensa y de expresión, sin embargo, muchos usuarios que atacan al rector en redes sociales a raíz de las polémicas sí que podrían reportarse, como mínimo, a la propia red social, pues cada una de ellas mantiene una serie de reglas y políticas[15].

No obstante, el factor que conforman los *datos personales* del sujeto es la principal causa de una superficie de exposición tan mejorable. Como se ha mostrado en el desarrollo de la investigación, varias páginas web públicamente accesibles mantienen varios datos personales con diferente nivel de sensibilidad; desde un artículo de periódico donde se muestra su fecha de nacimiento, hasta una página oficial de un registro gubernamental británico donde incluye una dirección de un posible domicilio, dejando de lado el número de teléfono personal contenido en un *leak* de Facebook. Además, un análisis en redes sociales ha permitido identificar a más de 15 familiares, como se ha esquematizado en la Figura A.2. Existen diferentes mitigaciones para limitar la exposición de datos personales directamente relacionados con el rector, y contramedidas a tener en cuenta para limitar el

[15]Reglas, políticas y *guidelines* de Twitter, donde es más común encontrar comentarios negativos: https://help.x.com/es/rules-and-policies

descubrimiento de familiares.

En lo que a datos personales respecta, es necesario investigar si el artículo de periódico mencionado está incurriendo en alguna ilegalidad por no haber censurado lo suficiente el documento que muestran, donde se enseña su fecha de nacimiento. Asimismo, el propio registro gubernamental británico indica en su página web que no enseñan públicamente las direcciones de domicilio o las fechas de nacimiento[16], algo que no parece cumplirse en el caso del sujeto, por lo que se podría contactar por vías oficiales con los administradores de la página oficial para evitar la exposición de estos datos personales. Sobre el teléfono móvil personal, las medidas a implementar no podrían evitar su difusión por canales privados o el internet profundo, pero sí sería posible prevenir, por ejemplo, la comprobación de su número por Whatsapp. La aplicación provee al usuario de ajustes de privacidad[17] que permiten ocultar la foto de perfil y mostrarla únicamente a contactos que tenga el sujeto almacenados en su teléfono.

Estos ajustes de privacidad son importantes para otras redes sociales. Una de las cuentas personales del rector en Facebook mantiene públicos los amigos añadidos, algo a mantener en privado como ocurre con las otras dos cuentas de Facebook. Además, para limitar las posibilidades de identificar familiares, también habría que tener en cuenta las relaciones en redes sociales señaladas en la Figura A.1, que a su vez permiten identificar cuentas personales del sujeto.

Siguiendo con lo mostrado en el grafo resulta sencillo aportar recomendaciones que ayudan a reducir la superficie de exposición, pues el sujeto ya realiza una buena práctica a la hora de mantener una imagen pública; mantener una cuenta profesional y una cuenta personal por separado en redes sociales. El problema es la falta de optimización en el momento mantener esta separación, ya que la cuenta profesional almacena relaciones con cuentas de sus familiares, ya sea en sus listas de seguidores en el caso de Instagram o en la lista de amigos de sus familiares en el caso de Facebook, quienes no son difíciles de encontrar una vez se conoce la cuenta de Instagram de cada uno de ellos. Por tanto, se propone una modificación de los seguidores y de las cuentas seguidas en Instagram desde la cuenta profesional del sujeto basándonos en el uso de la cuenta secundaria para mantener estas relaciones, lo que ayudará a imposibilitar la enumeración de familiares. Además, en lo que al uso de Facebook se refiere, el mantener dos cuentas personales aumenta la superficie de exposición. Por lo tanto, se propone el borrado de una de ellas para juntar su contenido y amigos con la otra, de forma que se proteja su vida social sin verse perjudicada. Aplicando estas modificaciones, el grafo representativo de la relación con su familiares por redes sociales se vería modificado, resultando en la Figura A.3.

Como se puede observar, la figura pública del rector queda más separada de

[16]https://www.gov.uk/government/organisations/companies-house/about/personal-information-charter

[17]https://faq.whatsapp.com/3307102709559968/

las cuentas personales de sus familiares, pero se mantienen ciertas relaciones hacia las cuentas personales del sujeto al mostrarse en cuentas públicas diferentes del mismo, es decir; los perfiles de Instagram y Facebook personales del rector seguirán apareciendo en las listas de seguidos y seguidores y listas de amigos de los familiares que las compartan de forma pública. Aunque esto no es algo preocupante, ya que no habría forma de llegar a estos perfiles de los familiares del rector con estas modificaciones, es cierto que es información que sigue disponible de forma pública, en fuentes abiertas, por lo que podría no ser del interés del sujeto. En función de las preferencias del rector, este podría comunicarse con los familiares que no se dediquen a subir contenido en línea para un público determinado o que no les importe pasar a tener una cuenta privada para implementar las medidas de privacidad ya mencionadas anteriormente.

5

Conclusiones y trabajos futuros

5.1. Conclusiones

A lo largo de esta memoria se ha establecido el concepto de fuentes abiertas de inteligencia y técnicas OSINT para mostrar su aplicación en investigaciones siguiendo una metodología actualizada por expertos y organizaciones dedicadas a la generación de inteligencia. En una primera sección se ha introducido este concepto, junto con los ámbitos donde suele utilizarse, su relevancia a día de hoy y una historia resumida de las técnicas OSINT desde que surgió el concepto en el siglo XX. Junto con unos objetivos definidos en la segunda sección, de los que se hablará más adelante, esto ha permitido estudiar el estado del arte de los procesos de inteligencia en una tercera sección para poder buscar, valorar y aprovechar las numerosas fuentes abiertas disponibles hoy en día, realizar un tratamiento eficaz de la información y crear un reporte adecuado a los requisitos de una investigación. La cuarta sección ha permitido poner en práctica esta metodología en un caso real, enfocado en la proyección de la huella digital de una figura pública.

En esta última sección se valorará la investigación desarrollada en el caso práctico mencionado, cuya extensión ha hecho posible la demostración de todo lo ilustrado en las secciones más teóricas: se ha hecho uso de una gran variedad de fuentes abiertas que han aportado diferentes tipos de datos e información; esta información se ha utilizado para generar inteligencia, a modo de respuesta de los requerimientos especificados en la investigación; la inteligencia obtenida, al tratarse de una huella digital, ha propiciado una evaluación sobre el grado

de exposición de la vida del sujeto y de los consecuentes riesgos; por último, gracias a todo este proceso, se han identificado contramedidas para reducir esta exposición y mantener una privacidad más apropiada para la seguridad del sujeto investigado. Esta serie de resultados no habría sido tan detallada de no haber seguido la metodología estudiada, que ha orientado toda la investigación y cuyas características son apreciables en todo el caso práctico.

Esta valoración queda completada una vez se repasan los objetivos definidos en la Sección 2. Teniendo en cuenta el objetivo principal, los apartados desarrollados en este documento alcanzan cada uno de los objetivos secundarios que constituyen al principal. La Sección 3 muestra el **estado del arte en relación a las fuentes abiertas y técnicas OSINT** y, a su vez, sirve de **análisis y síntesis de la metodología de investigación para generar inteligencia**. En la Sección 4, tras **identificar un caso real en el que las técnicas OSINT son de utilidad**, se ha **aplicado la metodología** para facilitar una serie de resultados que constituyen **hallazgos relevantes que cumplen con los requisitos definidos** para la investigación. Estos resultados han demostrado la necesidad de enumerar **recomendaciones que mitiguen la exposición del sujeto**. Por tanto, es posible describir una serie de **conclusiones y trabajos a futuro** tras el desarrollo del caso práctico, no sólo gracias al cumplimiento de los objetivos enumerados sino, también, al resultado satisfactorio[1] de la investigación llevada a cabo.

Una vez comentado el valor del proyecto llevado a cabo, se ha llegado a diferentes conclusiones sobre el uso de técnicas OSINT en investigaciones a partir de la gran interconexión e intercambio de información existente en la actualidad gracias a internet. Al haber utilizado únicamente fuentes abiertas, la característica más llamativa del caso práctico es la accesibilidad total a los datos e información encontrados a medida que ha ido teniendo lugar, a pesar de seguir un proceso de inteligencia equivalente a otras investigaciones que, como se ha expuesto en la Subsección 3.3, no tienen por qué valerse de fuentes abiertas. Esto permite afirmar que la investigación expuesta en este documento es replicable por cualquier usuario que tenga intenciones de trazar la huella digital del sujeto, lo que puede interpretarse de forma errónea como un ejercicio ejecutable de forma eficaz por cualquier persona independientemente de su formación o experiencia. A continuación, se enumerará una sucesión de puntos que clarifican esta posible impresión ante el lector, y a su vez, exponen tanto dificultades como ventajas desde el punto de vista del investigador o analista.

- **Gran cantidad y variedad de fuentes disponibles.** Como se ha puesto

[1] Los archivos con los resultados sin anonimizar del caso práctico se han almacenado en un repositorio privado de GitHub mantenido por el autor de la memoria. Si el lector lo desea, puede escribir un correo a o.lozano.2020@alumnos.urjc.es especificando cuenta de GitHub (apodo o correo electrónico) y motivo de acceso para recibir una invitación. Para velar por la privacidad del sujeto, no se concederá acceso sin motivo justificado.

de manifiesto en la Subsección 1.2, hoy en día existe una enorme cantidad de información disponible en línea, no sólo por los artículos que se publican y el contenido en redes sociales, sino también por los perfiles que se crean y contienen datos personales de cada usuario, reales o no. Gran parte de la aplicación de técnicas OSINT son necesarias para recolectar esta información, pero el propio investigador es el encargado de llevar a cabo correctamente las etapas de procesado que ayuden a filtrar todas las fuentes que no sean de utilidad para obtener inteligencia, como se ha desarrollado en las iteraciones del caso práctico en la Sección 4. La utilidad de estas fuentes no es particularidad fácil de valorar y, en muchas ocasiones, el valor de estas se hace visible únicamente en fases posteriores de la investigación, lo que añade complejidad a estos procesos, como se desarrolla en un punto posterior. Además, la variedad de fuentes abiertas es un concepto también observado en el caso práctico y a complementar en la Subsección 5.2; internet no sólo ofrece lo observable mediante consultas a motores de búsqueda convencionales, sino también conectividad a las redes Tor y, en otras investigaciones que resulte de interés, información de enrutación o registros DNS que aporten datos sobre el *hosting* de páginas web relevantes e incluso acceso a dispositivos con direcciones IP públicas gracias a buscadores como Shodan[2]. Sacar beneficio de estas posibilidades manteniendo un nivel de privacidad requiere de determinadas capacidades por parte del investigador y, como se ha demostrado, esta gran cantidad de información es, a la vez, tanto un obstáculo como una virtud de las fuentes abiertas.

- **Particularidades de herramientas[3] diseñadas para técnicas OSINT.** En relación con el punto anterior, existen herramientas que pueden facilitar el progreso de una investigación automatizando diferentes técnicas OSINT, y una gran mayoría de ellas se aprovechan de funcionalidades programables[4] que permiten realizar consultas de forma más cómoda y adquirir información más ordenadamente. Sin embargo, el uso de estas herramientas conllevan determinadas desventajas que se presentan con el tiempo. El mantenimiento de las mismas, al basarse muchas en consultar a servicios disponibles en redes sociales o en motores de búsqueda, debe de ser constante y adaptable, tarea ardua para los individuos autores de estas herramientas. Como se expone en el libro de Michael Bazzell sobre OSINT[35], la disponibilidad de estos programas informáticos fluctúa con el paso de los meses y para el investigador es importante entender cómo funciona cada una de las herramientas que tiene a su disposición; no sólo es cuestión de tener a mano reemplazos (sencillos de encontrar en gran parte de ocasiones al haber un

[2]https://www.shodan.io

[3]La compilación más famosa de herramientas se encuentra en https://osintframework.com/, aunque hay muchas otras que incluyen herramientas instalables con más funcionalidades.

[4]Se está haciendo referencia a las APIs o *Application Program Interfaces* disponible habitualmente en aplicaciones web. Más información sobre este concepto: https://www.redhat.com/en/topics/api/what-are-application-programming-interfaces

catálogo tan amplio de aplicaciones en este ámbito), sino también de ser capaz de actuar de forma independiente a las mismas.

- **Complejidad de una investigación dirigida.** Un caso real como el mostrado en la Sección 4 siempre comenzará con unos requerimientos a lograr por parte del investigador. Por lo tanto, lo más común es que las investigaciones se tornen más enrevesadas al paso que se va iterando para generar inteligencia de calidad, como se ha demostrado tanto en el desarrollo del caso práctico como en las figuras mostradas en el Apéndice A. Si bien esto puede parecer algo negativo, esta característica es algo inherente al propio proceso de inteligencia, pues una obtención satisfactoria de información que permita llevarlo a cabo sólo lo será si hay suficientes datos conectados entre sí, al ser lo que les aporta relevancia. Un ejemplo de esto no sólo son los grafos de las figuras contenidas en el apéndice señalado, sino también, las etapas de análisis de la Sección 4, donde se explican las relaciones que tienen las fuentes tratadas con el sujeto y, a raíz de su entendimiento y valoración, es posible encontrar nuevas necesidades para una iteración consecuente. Este hecho comprobable prueba una proporcionalidad directa entre la complejidad de una investigación y el valor que aporta a la toma de decisiones que sustentará posteriormente.

- **Responsabilidad sobre la inteligencia generada.** En línea con las capacidades requeridas por parte del individuo investigador, la investigación propiamente dicha también supone un ejercicio de responsabilidad, al tratarse información sensible a diferentes niveles y estar guiada por requisitos definidos. No sólo se debe de tener en cuenta la discreción del analista para no comprometer el nivel de secretismo impuesto por el decisor, concepto desarrollado en la Subsección 3.5, sino, además, ser capaz de comprometerse con la finalidad de la investigación y no usar los resultados en beneficio propio, mucho menos si conlleva el perjuicio de los sujetos relativos a la misma.

5.2. Trabajos futuros

En esta última subsección se hablará sobre posibles formas de complementar el proyecto reflejado en esta memoria y que no ha sido posible realizar para no extender de más el desarrollo de la investigación. Aunque este documento se ha enfocado en huellas digitales, otras investigaciones podrían orientarse a obtener información sobre sucesos u organizaciones en vez de individuos, conformando casos prácticos de diferente índole y que requerirían técnicas OSINT distintas a demostrar y aplicar. No obstante, es factible continuar con la identificación de posibles nuevas iteraciones en la investigación seleccionada para probar la importancia y eficacia de la metodología estudiada en la Sección 3; para ello,

debe tenerse en cuenta qué ha podido faltar a la hora de investigar o, también, qué técnicas adicionales pueden aplicarse para detallar mejor los resultados.

Según lo anterior, se identifican dos principales vías por las que completar la investigación:

- **Extender la búsqueda de información en el "internet profundo".** Como se ha podido observar en la tercera iteración de la investigación, lo único que se ha encontrado en el internet profundo ha sido a raíz de un canal privado de Telegram. Finalmente no se ha utilizado la red Tor ni ningún dominio *onion* del que extraer información de las cuentas del sujeto, pues esta búsqueda podría requerir los recursos equivalentes a una única investigación por separado al no partir de ninguna base con la que buscar fuentes accesibles que proporcionen archivos generados por brechas de datos. Lo único que se encontraron fueron páginas web que pedían pagos en criptomonedas[5] para poder realizar consultas en bases de datos privadas para obtener datos de cuentas bloqueadas, y aunque estas fuentes seguirían considerándose abiertas al estar accesibles a cualquiera, no se han tenido en cuenta por falta de confianza. Por lo tanto, sería viable continuar con la investigación por este medio.

- **Usar *sock puppets accounts*[6] en redes sociales.** Una parte importante del trazado de la huella digital realizado en la memoria ha sido el uso de redes sociales como fuentes abiertas para ganar información tanto de la reputación del sujeto como de su vida personal. Muchas de las relaciones encontradas se representan correctamente en la Figura A.1, pero, como ya se ha comentado, esto ha sido gracias a la disponibilidad pública de las listas de seguidos, seguidores o amigos de las cuentas según a la red social que correspondan. De ahí la necesidad de que el investigador trate de acceder, de forma totalmente legal, a las cuentas privadas que tienen relación con el sujeto o, directamente, a las cuentas personales del sujeto que mantienen el nivel de privacidad adecuado, como ya se indicó en la Subsección 4.5. En este punto entran las "cuentas marioneta", perfiles que utilizan datos personales fabricados (nombre, apellidos, imágenes, publicaciones, seguidores...) y que pueden servir para dar la impresión de que pertenecen a un usuario real. Al no estar usando datos personales pertenecientes a otra persona, el investigador no está incurriendo en ninguna ilegalidad, y cabe la posibilidad de que el sujeto (o los usuarios cercanos a él) acepten su solicitud de

[5] En este tipo de sitios web normalmente se piden pagos en Bitcoin para mantener anonimizados a los usuarios que formen parte de la transacción. Más información sobre esta criptomoneda en su página oficial: https://bitcoin.org/es/

[6] En español, "cuentas marionetas". Aquí una publicación en el blog de Maltego, la herramienta utilizada para crear los grafos expuestos en el Apéndice A, para más información: https://www.maltego.com/blog/how-to-use-sock-puppet-accounts-to-gather-social-media-intelligence/

seguimiento o de amistad según los rasgos del usuario que la cuenta falsa trata de representar. Por tanto, se tendría acceso a fuentes de información que no estaban disponibles y que, aunque no se consideren fuentes abiertas, permitirían completar los resultados de la investigación con información adicional.

Aunque en el caso práctico expuesto en este documento no ha coincidido con la siguiente posibilidad, puede ocurrir que el sujeto del que se está repasando la huella digital sea activo en alguna de las cuentas de sus redes sociales y cree publicaciones en línea de manera asidua. Por resultar interesante, en esta sección se añade una mención a herramientas de monitorización que ofrecen servicios de generación de alertas o almacenado automático de publicaciones que el sujeto vaya realizando en sus redes y otros cambios en su perfil. Dicha monitorización puede extenderse a cuentas relacionadas con el sujeto y así posibilitar un seguimiento automático de los movimientos del mismo. A falta de incluir aplicaciones que establezcan este servicio de monitorización, comentar la posibilidad de crear estos programas desde cero o programar estas funcionalidades para que, por ejemplo, se ejecuten desde un bot en Discord[7], una plataforma de mensajería instantánea.

[7]El desarrollo de bots en Discord para servir funcionalidades ajenas a la aplicación no es algo desconocido. Discord es una plataforma muy utilizada actualmente, y noticias como la siguiente demuestran el potencial de los bots en esta: https://www.lisanews.org/actualidad/ciberdelincuentes-utilizan-un-malware-espia-en-discord-a-traves-de-emojis/

Bibliografía

[1] P. Casanovas, "Cyber warfare and organised crime. a regulatory model and meta-model for open source intelligence (OSINT)," in *Ethics and Policies for Cyber Operations: A NATO Cooperative Cyber Defence Centre of Excellence Initiative*, M. Taddeo and L. Glorioso, Eds. Springer International Publishing, pp. 139–167. [Online]. Disponible en: https://doi.org/10.1007/978-3-319-45300-2_9 (Fecha de último acceso: 2024-04-19)

[2] L. Block, "The long history of OSINT," *Journal of Intelligence History*, pp. 1–15, Jun. 2023. [Online]. Disponible en: https://www.tandfonline.com/doi/full/10.1080/16161262.2023.2224091 (Fecha de último acceso: 2024-06-05)

[3] M. Bazzell, "Introduction," in *OSINT Techniques: Resources for Uncovering Online Information*, 2023.

[4] S. D. Weaver and M. Gahegan, "Constructing, visualizing, and analyzing a digital footprint," vol. 97, no. 3, pp. 324–350. [Online]. Disponible en: https://www.tandfonline.com/doi/full/10.1111/j.1931-0846.2007.tb00509.x (Fecha de último acceso: 2024-06-28)

[5] V. Mavroeidis and S. Bromander, "Cyber threat intelligence model: An evaluation of taxonomies, sharing standards, and ontologies within cyber threat intelligence," in *2017 European Intelligence and Security Informatics Conference (EISIC)*. IEEE, pp. 91–98. [Online]. Disponible en: http://ieeexplore.ieee.org/document/8240774/ (Fecha de último acceso: 2024-06-28)

[6] R. Messier, *Penetration Testing Basics*. Apress, accedido el 28-06-2024. [Online]. Disponible en: http://link.springer.com/10.1007/978-1-4842-1857-0 (Fecha de último acceso: 2024-06-28)

[7] J. J. McGonagle and C. M. Vella, "What is competitive intelligence and why should you care about it?" in *Proactive Intelligence: The Successful Executive's Guide to Intelligence*, J. J. McGonagle and C. M. Vella, Eds. Springer, pp. 9–19. [Online]. Disponible en: https://doi.org/10.1007/978-1-4471-2742-0_2 (Fecha de último acceso: 2024-07-01)

[8] Global social media statistics. [Online]. Disponible en: https://datareportal.com/social-media-users (Fecha de último acceso: 2024-07-01)

[9] K. Haan. Top website statistics for 2024. Section: Software. [Online]. Disponible en: https://www.forbes.com/advisor/business/software/website-statistics/ (Fecha de último acceso: 2024-07-01)

[10] F. Sampson, "Intelligent evidence: Using open source intelligence (OSINT) in criminal proceedings," vol. 90, no. 1, pp. 55–69. [Online]. Disponible en: http://journals.sagepub.com/doi/10.1177/0032258X16671031 (Fecha de último acceso: 2024-07-01)

[11] A. B. C. News. Philly woman accused of torching police cars during protest tracked down through etsy, LinkedIn. [Online]. Disponible en: https://abcnews.go.com/US/

philly-woman-accused-torching-police-cars-protest-tracked/story?id=71325821 (Fecha de último acceso: 2024-07-05)

[12] MwOsint. Unravelling the norton scam. [Online]. Disponible en: https://keyfindings.blog/2019/08/28/unravelling-the-norton-scam/ (Fecha de último acceso: 2024-07-05)

[13] F. Schaurer and J. Störger, "The Evolution of Open Source Intelligence (OSINT)," *AFIO's The Intelligencer*, 2013.

[14] R. A. Norton, "Guide to Open Source Intelligence," *AFIO's The Intelligencer*, 2011.

[15] C. Burke, "Freeing knowledge, telling secrets: Open source intelligence and development," *CEWCES Research Papers*, Jan. 2007.

[16] R. A. Best, "Open source intelligence (OSINT): Issues for congress," pp. 6–7.

[17] W. F. Kernan, "NATO OPEN SOURCE INTELLIGENCE HANDBOOK." [Online]. Disponible en: https://ia801403.us.archive.org/32/items/NATOOSINTHandbookV1.2/NATO%20OSINT%20Handbook%20v1.2%20-%20Jan%202002.pdf (Fecha de último acceso: 2024-04-11)

[18] A. Supreme Allied Commander and E. Supreme Allied Command, "NATO OSINT reader." [Online]. Disponible en: https://cyberwar.nl/d/NATO%20OSINT%20Reader%20FINAL%20Oct2002.pdf (Fecha de último acceso: 2024-06-12)

[19] "Intelligence exploitation of the internet." [Online]. Disponible en: https://nsarchive2.gwu.edu/NSAEBB/NSAEBB436/docs/EBB-005.pdf (Fecha de último acceso: 2024-01-24)

[20] . Comission, "The 9/11 commission report," pp. 407–415.

[21] Office of the director of national intelligence. [Online]. Disponible en: https://web.archive.org/web/20060623072458/http://dni.gov/press_releases/20051108_release.htm (Fecha de último acceso: 2024-06-03)

[22] S. Aftergood. CIA cuts off public access to its translated news reports. [Online]. Disponible en: https://fas.org/publication/fbis-wnc/ (Fecha de último acceso: 2024-06-09)

[23] ——. Open source center (OSC) becomes open source enterprise (OSE). [Online]. Disponible en: https://fas.org/publication/osc-ose/ (Fecha de último acceso: 2024-06-09)

[24] H. Williams and I. Blum, *Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise.* RAND Corporation. [Online]. Disponible en: https://www.rand.org/pubs/research_reports/RR1964.html (Fecha de último acceso: 2024-06-09)

[25] M. Warner, "Wanted: A definition of ïntelligence"," vol. 46, no. 3. [Online]. Disponible en: https://www.cia.gov/resources/csi/static/Wanted-Definition-of-Intel.pdf

[26] C. Seisdedos and V. Aguilera, "OSINT disciplina de inteligencia," in *Open Source INTelligence (OSINT) Investigar personas e Identidades en Internet*, 1st ed., pp. 15–22.

[27] L. Santos. La inteligencia. [Online]. Disponible en: https://www.cni.es/la-inteligencia (Fecha de último acceso: 2024-04-07)

[28] The intelligence cycle — central intelligence agency. [Online]. Disponible en: https://web.archive.org/web/20200508151219/https://www.cia.gov/kids-page/6-12th-grade/who-we-are-what-we-do/the-intelligence-cycle.html (Fecha de último acceso: 2024-04-10)

[29] The work of a nation - the intelligence cycle. [Online]. Disponible en: https://web.archive.org/web/20140313061711/https://www.cia.gov/library/publications/additional-publications/the-work-of-a-nation/work-of-the-cia.html (Fecha de último acceso: 2024-04-18)

[30] E. Șușnea and A. Iftene, "The Significance of Online Monitoring Activities for the Social Media Intelligence," Jul. 2018. [Online]. Disponible en: https://ibn.idsi.md/sites/default/files/imag_file/230-240.pdf (Fecha de último acceso: 2024-06-20)

[31] B. H. Miller, "Open source intelligence (OSINT): An oxymoron?" vol. 31, no. 4, pp. 702–719. [Online]. Disponible en: https://www.tandfonline.com/doi/full/10.1080/08850607.2018.1492826 (Fecha de último acceso: 2024-05-29)

[32] S. Stottlemyre, "The united states intelligence community, secrecy and the 'steele dossier': Reconceptualizing the intelligence process," vol. 4, pp. 11–27.

[33] Traffic light protocol (TLP). [Online]. Disponible en: https://www.first.org/tlp (Fecha de último acceso: 2024-05-29)

[34] What is digital footprint? | check and protect your digital footprint. [Online]. Disponible en: https://www.malwarebytes.com/cybersecurity/basics/digital-footprint (Fecha de último acceso: 2024-07-09)

[35] M. Bazzell, *OSINT Techniques: Resources for Uncovering Online Information.* (independently published), 2023.

Apéndices

A

Figuras complementarias

A.1. Grafos relativos a la Sección 4

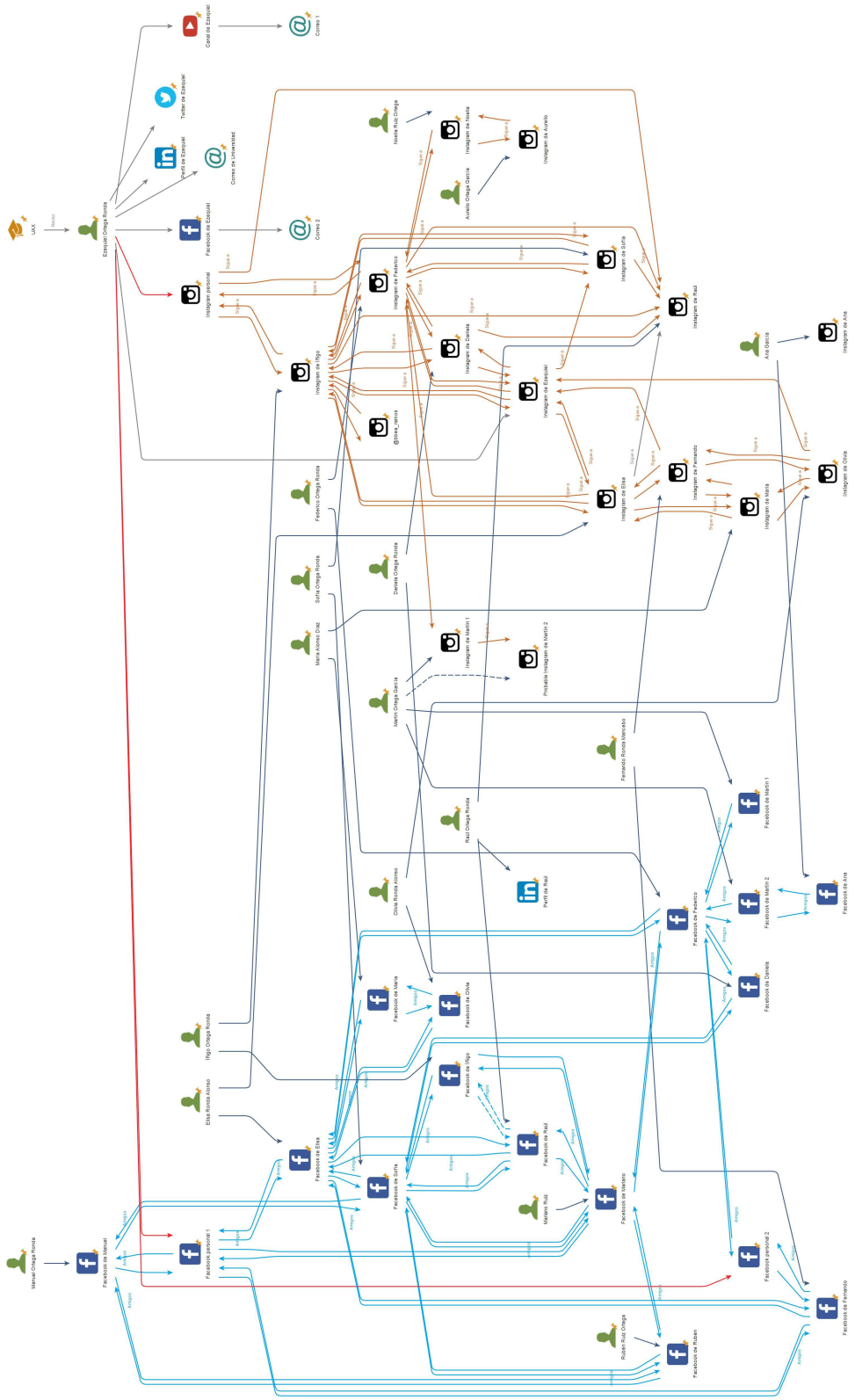

Figura A.1: Grafo relacional entre las cuentas en redes sociales de la familia "Ortega Ronda". Las flechas rojas representan cuentas personales. Referenciado en la Subsección 4.3.5.

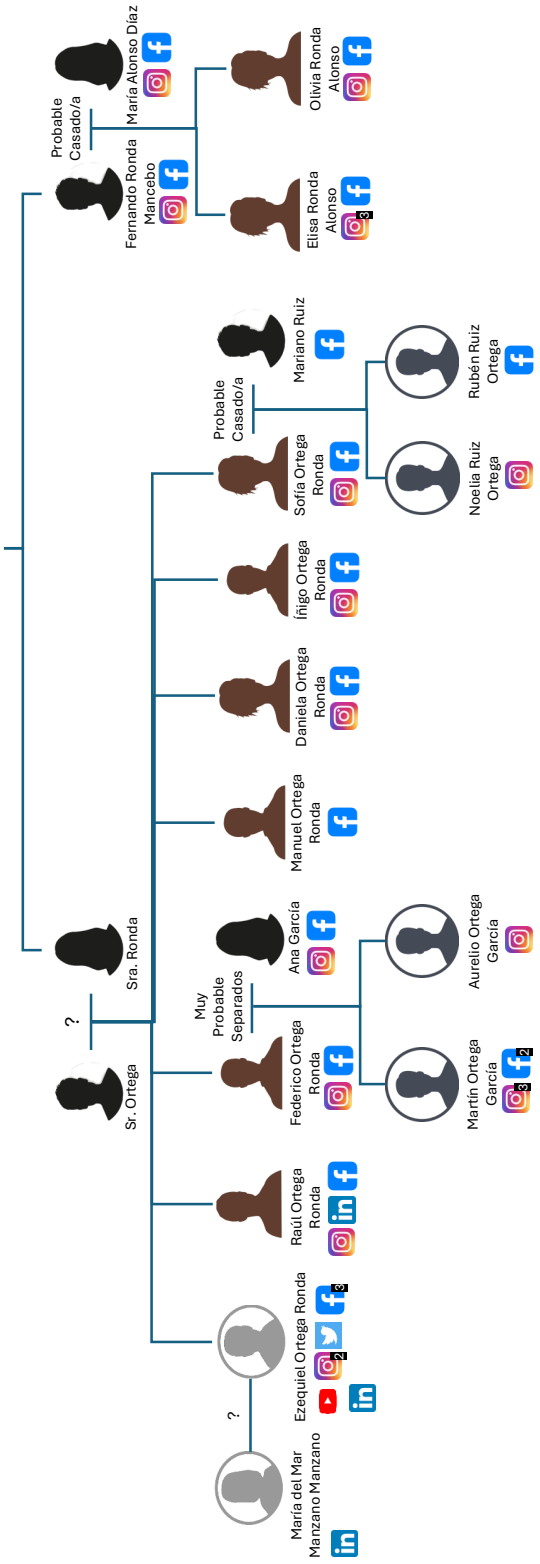

Figura A.2: Diagrama de árbol representativo de la familia descubierta del rector, "Ezequiel Ortega Ronda". Referenciado en la Subsección 4.4.

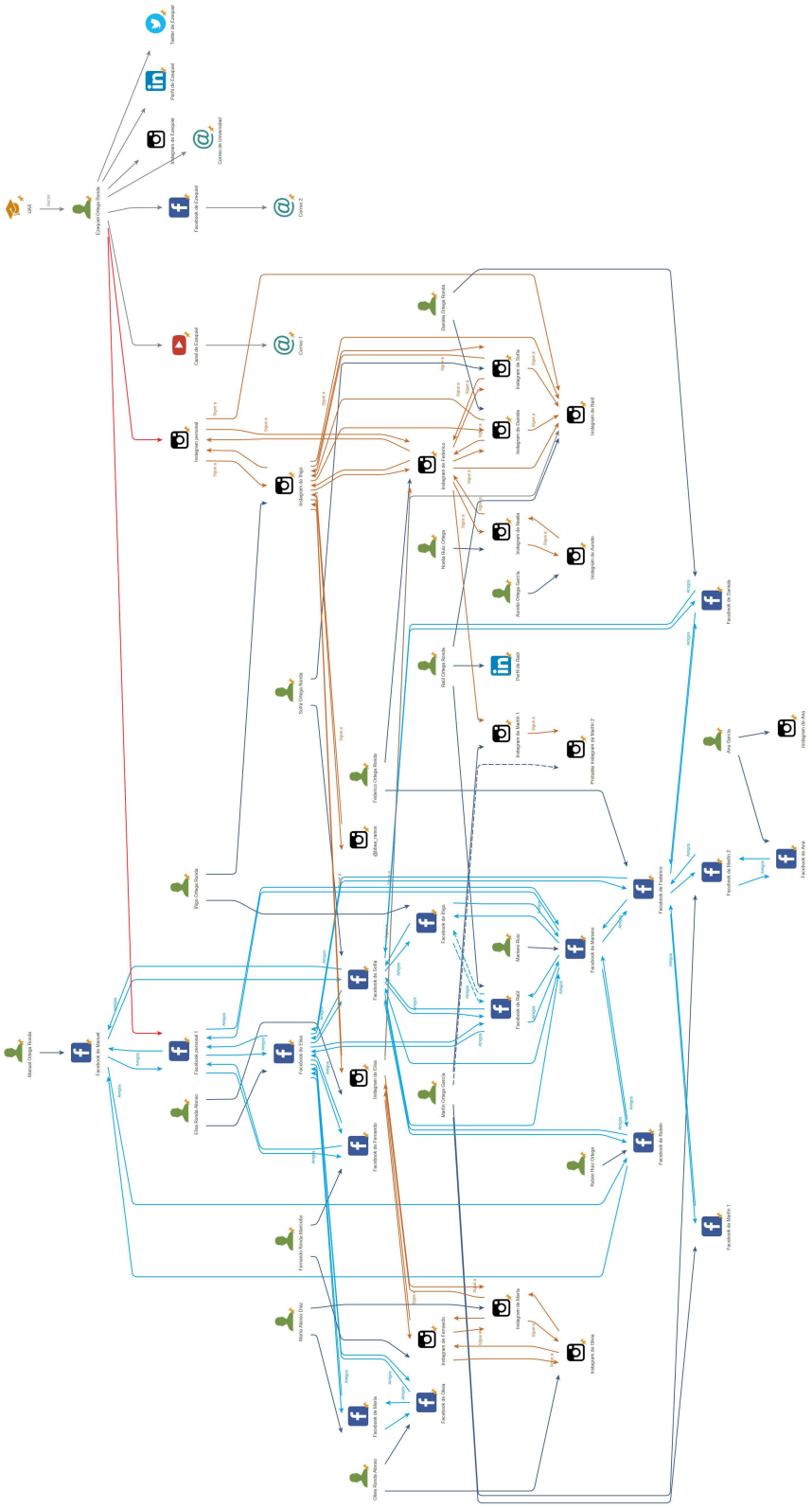

Figura A.3: Grafo relacional modificado entre las cuentas en redes sociales de la familia "Ortega Ronda". Las flechas rojas representan cuentas personales. Referenciado en la Subsección 4.5.