Las redes sociales como fuente de datos para un Observatorio Regional de Ingeniería en Sistemas de Información e Informática. Oportunidades y limitaciones técnicas, éticas y legales.

Abbatemarco Martín, Brizuela Luisina, Cervino Alejo, Riva Fabiana María
Departamento de Ingeniería en Sistemas de Información
Facultad Regional Rosario
Universidad Tecnológica Nacional
E. Zeballos 1342, 2000 Rosario, Argentina
abbatemarco.martin@gmail.com, luisina.brizuela@yahoo.com,
alejocervino@hotmail.com, fabianamriva@gmail.com

Abstract

El presente trabajo pretende establecer la factibilidad de la recolección y el análisis sistematizado de datos que permita contar con información relacionada al desenvolvimiento en el mercado laboral de estudiantes y graduados de la carrera de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional - Facultad Regional Rosario, para la futura construcción de un mapa interactivo que posibilite conocer su ubicación geográfica y el ámbito de ejercicio profesional. Se analizarán dos redes sociales específicas, Facebook y LinkedIn, destacando cuestiones técnicas de autenticación y autorización para el acceso a los datos y las características de las interfaces (APIs) que las mismas proveen para este fin. Por otro lado se discutirá acerca de las cuestiones éticas y legales que se derivan del uso masivo de datos.

El análisis realizado nos permitirá establecer, finalmente, una serie de oportunidades, limitaciones y alternativas que tendremos en cuenta para el cumplimiento de los objetivos mencionados.

Palabras clave: Redes Sociales - Observatorio - Análisis Masivo de Datos

1. Introducción

Actualmente vivimos en una época regida por la tecnología que, con sus incesantes cambios y progresos, ha traído un inimaginable incremento en el caudal de datos que circulan por la red. Cualquier persona conectada a Internet se convierte hoy, intencionalmente o no, en una generadora de contenido; principalmente a través de la continua interacción en redes sociales. Estas redes crecen y se multiplican a pasos agigantados, ofreciendo diversos atractivos y funcionalidades que buscan persuadir y atraer a usuarios alrededor del mundo. En este sentido, Manovich (2011) señala que *"el ascenso de las redes sociales junto con el progreso de las herramientas computacionales que permiten el procesamiento masivo de datos hace posible un nuevo enfoque para el estudio de los seres humanos y la sociedad"*[1].

Ahora bien, siendo tan grande la cantidad de datos personales almacenados en estas plataformas, surge la importancia de establecer si las mismas brindan la oportunidad de aportar información sustantiva a algunos de los objetivos del proyecto en el cual el presente trabajo se encuentra enmarcado: Observatorio Regional de Desarrollo de la Ingeniería en Sistemas de Información e Informática (IISI.d.r.O.). Particularmente, los objetivos a los que apunta este trabajo son *"la recolección y el análisis sistematizado de datos que permita contar con información relacionada al desenvolvimiento en el mercado laboral de estudiantes y graduados de nuestra carrera y la construcción de un mapa interactivo que permita conocer su ubicación geográfica y el ámbito de ejercicio profesional"*. Por otro lado, aparece también la necesidad del planteo y el intento de responder a ciertas cuestiones que hacen al uso específico de los datos: ¿es posible utilizar o manipular los datos personales y supuestamente públicos de los usuarios de las distintas redes socia-

les?, y de ser así, ¿hasta qué punto es legal hacerlo en Argentina?, ¿qué posición asumen las corporaciones a cargo de estas redes al respecto?.

Sin lugar a dudas las respuestas a estos interrogantes no son únicas y dependen fundamentalmente de la perspectiva con la que se los afronte. En la concepción de que las redes sociales pueden servir como una constante fuente de información no estructurada para IISI.d.r.O., será primordial para nosotros, en una primera etapa, focalizarnos en el análisis del acceso a información disponible sobre estudiantes y graduados de la carrera de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional - Facultad Regional Rosario (ISI-UTN-FRRo) existentes en redes como LinkedIn y Facebook. Nos interesará estudiar tanto la eficacia y eficiencia como el grado de accesibilidad y privacidad en estas redes para la recolección sistematizada y análisis de datos relacionados al desenvolvimiento en el mercado laboral de estudiantes y graduados.

Diversos autores analizados ([2, 3, 4]) han tratado las problemáticas que surgen en torno al análisis de redes sociales, sobre todo en lo que respecta a las limitaciones tanto espaciales como temporales derivadas de ser éste un campo muy reciente, poco maduro y en constante crecimiento. En este sentido, señalan que es extremadamente difícil conciliar, sin perder el enfoque, todos los conceptos, metodologías y herramientas utilizadas para la recolección y análisis de datos junto con los resultados derivados de la aplicación de los mismos y sumando además las implicancias éticas y morales de las actividades realizadas durante toda la investigación.

Para responder a los interrogantes planteados, este documento resume el trabajo realizado para acceder a datos almacenados en redes sociales factibles de ser utilizados en IISI.d.r.O. analizando, en su desarrollo, las cuestiones referidas a aspectos éticos y legales. Se dejan para futuros trabajos la aplicación de las metodologías y herramientas técnicas disponibles para la obtención, registro y armado del mapa interactivo.

2. Métodos de acceso a redes sociales

Las redes sociales son hoy en día una valiosa fuente de información para organizaciones gubernamentales, empresas e investigadores. A través de ellas se pueden obtener enormes cantidades de datos, casi en tiempo real, asociados al comportamiento individual y grupal de las personas en todos los ámbitos de su vida diaria. Esto hace que cada vez más investigadores deseen acceder a estas fuentes de forma fácil, rápida y segura. Como indica Bright (2014), este interés no despertó hasta hace algunos años y se dio principalmente por dos causas: mientras que por un lado se produjo la expansión del uso de Internet en todo el mundo, por el otro surgieron redes sociales masivas con un alto grado de penetración en la sociedad [5]. Claro ejemplo de ello es Facebook, que es hoy la red social con más usuarios registrados, contando con más de 1650 millones de usuarios activos al mes [6].

A partir de este marcado crecimiento y el fuerte interés en el aprovechamiento de estas constantemente actualizadas fuentes de datos, las plataformas han abierto algunas de sus puertas a un público determinado a través de sus respectivas Interfaces de Programación de Aplicaciones (API por sus siglas en inglés - Application Programming Interface). Existen además métodos alternativos para recabar información pública disponible en las redes sociales de forma automatizada, evitando hacer uso de las APIs oficiales provistas, como son los métodos de *scraping* (del verbo inglés *scrape* que significa raspar).

En pocas palabras, Glez-Peña et al. (2014) definen al scraping como un proceso por el cual un software, conocido como *web robot*, imita la navegación tradicional de un humano en diversas páginas de Internet, primero analizando gramaticalmente su contenido y luego encontrando, extrayendo y estructurando automáticamente todos los datos de interés de forma sistematizada [7]. Sin embargo, coincidimos con Bruns (2013) en que dichos métodos suelen ser muy poco prácticos, principalmente porque fallan a la hora de recolectar datos que no están expuestos en la web y son únicamente accesibles, en teoría, a través de las APIs [2]. Por otra parte, toda práctica de *scraping* y otras similares suelen ser fuertemente desalentadas, sino prohibidas, en los términos y condiciones de uso de las plataformas sociales. En consecuencia, acotaremos nuestro análisis al acceso a datos a través de las APIs oficiales existentes en las respectivas plataformas.

En esencia, las APIs podrían pensarse como simples intermediarios entre programador y aplicación; el intermediario recibe peticiones del programador y, si ellas son válidas y cuentan con los permisos necesarios, retorna los datos requeridos. Debido a que cada interfaz tiene un funcionamiento particular, suelen estar acompañadas por una detallada documentación donde se encuentran todas las funcionalidades disponibles, cómo debe ser usada y qué formatos puede recibir como entrada o retornar como salida [8].

Ahora bien, desde el punto de vista del usuario, ¿cuáles son los aspectos que hacen buena a una API?, ¿qué factores hay que tener en cuenta para su comparación? De acuerdo con Bloch (2006), una buena API debe ser principalmente fácil de usar y aprender, como así también apropiada para la audiencia a la que se dirige. Adicionalmente, se espera que sean lo suficientemente potentes para satisfacer todos los requerimientos de sus usuarios sin problemas[9]. Una parte fundamental de cualquier API es su documentación; a veces los servicios ofrecidos son tan variados y complejos que, más allá de todo el potencial que la API pueda poseer, será imposible de utilizar sin las correctas indicaciones. Deberá tenerse en cuenta entonces, para la evaluación y comparación, todo lo relativo a la estructura, especificidad e inteligibilidad de la documentación asociada.

2.1. Autenticación y autorización en APIs

Todos los métodos y herramientas incluidos en las APIs que analizaremos tienen en común dos instancias previas fundamentales: autenticación y autorización. Para llevar adelante estas actividades, hoy en día existe un estándar ampliamente difundido e implementado en diferentes aplicaciones y plataformas a nivel mundial. Este estándar, llamado *OAuth*, se encuentra en su versión 2.0 y fue co-creado por una amplia comunidad, que incluyó en su momento a representantes de Google, Twitter, Yahoo y Facebook, entre otros. Como bien destaca su web institucional, *OAuth* se puede definir como un protocolo abierto que permite la autorización simple, segura y estándar de aplicaciones web, aplicaciones para dispositivos móviles y de escritorio. Leiba (2012) agrega que el protocolo *OAuth* nos permite establecer estándares para la gestión de identidades, proveyendo una solución simple a todas las problemáticas que existen en compartir nuestros usuarios y contraseñas [10]. Mientras tanto, Hardt (2012) sintetiza su funcionamiento diciendo que en lugar de usar las credenciales del dueño de los recursos protegidos que se quieren acceder, a través de OAuth 2.0 el usuario obtiene un token de acceso (*access token*) que es generado por un servidor de autorizaciones, junto con la aprobación del dueño del recurso. El token de acceso es simplemente una cadena de caracteres que indica, entre otros atributos, el alcance y tiempo de vida de los permisos otorgados [11].

Se analizará luego, en forma particular para cada API, de qué manera se obtienen los respectivos tokens de acceso y qué características específicas tienen en cada caso. Estos tokens serán nuestros "boletos de entrada" para acceder y obtener los datos personales que residen dentro de los perfiles de los distintos usuarios de las redes sociales bajo análisis.

3. Facebook

Actualmente la empresa liderada por Mark Zuckerberg cuenta principalmente con tres interfaces de programación de aplicaciones bien diferenciadas: Atlas API, Marketing API y Graph API. Ya que Atlas y Marketing están orientadas a otro tipo de usos como marketing dirigido, propaganda y gestión de campañas publicitarias, nuestro análisis estará totalmente enfocado en la última mencionada, es decir, Graph API.

En palabras de Facebook: *"API Graph es la principal forma de introducir y extraer datos en la plataforma de Facebook. Se trata de una API basada en HTML de bajo nivel que se puede utilizar para consultar datos, publicar nuevas historias, administrar anuncios, subir fotos y muchas otras tareas que se pueden requerir en una aplicación"*.

La evolución de la API Graph ha sido constante; pasó por sucesivas e incrementales versiones hasta llegar a la actual, v2.6 presentada el 12 de Abril del 2016. A pesar de promover fuertemente la actualización de todas las aplicaciones que hagan uso de su Graph API a la versión más moderna, todavía se da soporte a versiones anteriores. Respecto al tiempo de duración de dichas versiones, Facebook mantiene una política firme y precisa: se garantiza que todos los componentes del *"core"* de cada versión operarán sin cambio alguno por lo menos dos años más a partir del lanzamiento de la versión subsiguiente.

Conocer las políticas referidas a la actualización de versiones es de particular interés para aquellos desarrolladores que precisen saber cuándo utilizar una u otra versión. En este marco es que la documentación cuenta con una sección exclusiva dedicada a mantener informados a todos los desarrolladores alrededor del mundo sobre las distintas versiones de la plataforma, los registros de cambios (*changelogs*) y minuciosas guías de actualizaciones, que incluyen tutoriales paso a paso para migrar de una versión a otra. Todo esto permite a los programadores obtener, de manera rápida y directa, toda la información necesaria para anticiparse a posibles obsolescencias de características y funcionalidades de la interfaz [12].

Para todos aquellos desarrolladores principiantes o

simplemente entusiastas con deseos de realizar algunas pruebas simples y rápidas sobre la API Graph para comenzar a conocer sus capacidades y limitaciones, Facebook ofrece una herramienta extremadamente útil, fácil de aprender y utilizar llamada *Graph API Explorer*. Para comenzar a realizar toda clase de peticiones a través de la herramienta sólo se necesita una cuenta en Facebook y conocimientos básicos del formato de solicitudes HTTP.

Cabe destacar que existen un sin número de soluciones de terceros similares a Graph API Explorer, como la consola de la empresa Apigee o bien la aplicación de escritorio Facepager [13], la cual tiene la ventaja de ser distribuida con licencia de tipo MIT. Asimismo, posee la capacidad de guardar los datos públicos recolectados en una base de datos SQLite y exportarlos en formato CSV. Las dos últimas herramientas nombradas junto con Graph API Explorer (todas de libre acceso), han servido de gran ayuda como puntapié inicial para sumergirnos en los aspectos técnicos de nuestro trabajo.

Más allá de la herramienta que se utilice para la búsqueda y extracción de datos desde la API Graph provista por Facebook, será siempre condición necesaria la generación de tokens de acceso a través del estándar OAuth. En caso de no incluirse estos tokens como parámetros *access_token* en la llamada, la API devolverá un error de tipo *OAuthException* con el mensaje: *"An active access token must be used to query information about the current user"*. Existen diferentes métodos para obtener los identificadores, dependiendo del tipo de token que se necesite.

Facebook propone tres tipos de tokens de acceso: de usuario, de aplicación y de página. A pesar de esta variedad, encontramos en este punto el primer escollo en el camino al cumplimiento del objetivo planteado: siempre necesitaremos de un usuario logueado para consultar y extraer datos a través de la API, es decir, no hay modo de obtener información de Facebook de forma completamente anónima.

El token de acceso de usuario es el más común, ya que es aquel que se emplea cuando se requiere leer, modificar o escribir datos de un perfil concreto de Facebook. Puede ser adquirido a través de un formulario de inicio de sesión y precisa que el perfil haya concedido permisos para la lectura y modificación de datos. De no tener un usuario logueado sería imposible obtener los tokens de acceso. Los permisos cumplen un rol fundamental, ya que sumados a la configuración de privacidad del perfil definen exactamente qué información estará disponible a través de la API Graph y qué datos permanecerán ocultos. Existe un amplio rango de permisos, no obstante sólo algunos de ellos podrían ser de interés en nuestro trabajo: *user_friends, user_hometown, user_location, user_work_history, user_education_history y user_birthday*.

Como mencionamos, también existen identificadores de acceso de aplicación y de página. Mientras que los primeros tienen la limitación de no permitir leer datos de usuarios y utilizarlos en una aplicación, los tokens de página sólo son de utilidad para administrar páginas de Facebook, algo que no concierne a nuestra investigación.

3.1. Estructura de los datos en Facebook

Para trabajar con la API Graph, hay tres conceptos muy importantes que se deben conocer, y que hacen referencia a cómo se compone la información de los perfiles de Facebook. Estamos hablando de los nodos, los perímetros o *edges* y los campos o *fields*. Mientras que los nodos son definidos por Facebook diciendo que son básicamente "cosas" como un usuario, una foto, una página o un comentario, se puede entender a los perímetros como conexiones entre dichos nodos: los comentarios de una foto, o las fotos de una página. Sumado a ello, todos los nodos poseen información inherente que los caracteriza; como ser el nombre de una página o el lugar de residencia de un usuario. Dicha información es lo que Facebook llama fields. Todos los nodos y los perímetros en la API Graph se pueden leer simplemente con una solicitud GET HTTP enviada a la API en `graph.facebook.com`. Aunque no es estrictamente necesario, es una buena práctica hacer referencia a la versión de la API que se quiere consultar, por lo que en los casos de ejemplo haremos todas las solicitudes a `graph.facebook.com/v2.6/`. Es relevante saber que, sea cual sea el perímetro o nodo que se consulte, la API siempre responderá en el formato estándar JSON [1]. Para el análisis que queremos llevar a cabo, alcanzará con la utilización de las herramientas ya introducidas anteriormente (consola de Apigee, Explorer de Facebook o Facepager), las cuales nos dan la ventaja de generar automáticamente los tokens de acceso necesarios para todas las solicitudes.

Hasta el momento, las herramientas consideradas fueron de gran utilidad para la investigación. No obstante, a partir de la evaluación realizada, entendemos

[1] Puede consultarse [14] para conocer más sobre este formato de intercambio de datos

que estas no poseen el potencial ni las capacidades necesarias para lograr nuestro objetivo planteado. Al momento de realizar el mapa interactivo, solucionaremos esta problemática haciendo uso de la gran variedad de kits provisto por Facebook para el desarrollo de aplicaciones, tanto oficiales como de terceros, mejor conocidos como SDKs.

4. LinkedIn

Es ahora el turno de poner bajo análisis las posibilidades técnicas y problemáticas que aparecen alrededor de la extracción de datos de esta popular red social orientada a conectar estudiantes, profesionales y empresas de todo el mundo. A grandes rasgos, LinkedIn ofrece una única API a los desarrolladores, la cual llaman REST API (`https://api.linkedin.com/v1/`). Con los radicales cambios anunciados en Febrero del 2015, que fueron implementados a partir de mediados de Mayo de ese año, LinkedIn limitó en gran medida las posibilidades de acceso y extracción de datos de su plataforma para los desarrolladores comunes, dando una mayor prioridad a las empresas y programadores asociados (también llamados *partners*) a su "Partner Program". En este punto encontramos algunas similitudes en cómo Facebook y LinkedIn se relacionan y comunican con desarrolladores externos. Para que existan transiciones rápidas y prolijas entre versiones, ambas empresas brindan tutoriales paso a paso y *changelogs* que permiten conocer funcionalidades nuevas y deprecadas, aunque la estructura e inteligibilidad de la documentación en Facebook sobrepasa a LinkedIn. Además, la primera posee claras políticas respecto a la duración de cada versión, mientras que LinkedIn, con sólo tres meses de previo aviso, deshabilitó en su momento una gran cantidad de puntos de acceso a la REST API (algunos ampliamente utilizados como era Connections API), siendo tres los únicos restantes: Profile API (/v1/people/$\sim$), Share API (/v1/people/$\sim$/shares) y Companies API (/v1/companies/id); notar que el símbolo $\sim$ puede utilizarse como reemplazo del ID del usuario de LinkedIn que obtuvo los tokens de acceso. En consecuencia, para aquellos desarrolladores no inscriptos en el Partner Program han quedado fuera del alcance servicios anteriormente ofrecidos como la gestión de grupos, conexiones y búsqueda de personas o puestos de trabajo. Puesto que nuestra intención es acceder a información en los perfiles de los usuarios, dejaremos de lado Share API y Companies API, para centralizar nuestro análisis exclusivamente sobre Profile API y los datos a partir de ella extraíbles.

Como se aclara en la documentación al respecto, para que una aplicación logre acceder a datos de miembros[2] y/o actuar en su nombre, se debe atravesar el proceso de autenticación y autorización ya citado anteriormente. Al igual que Facebook, LinkedIn hace uso de las ventajas de OAuth 2.0 para hacer este proceso lo más simple y transparente posible. En este sentido, la empresa ofrece un didáctico tutorial paso a paso sobre cómo lograr obtener los permisos de usuario necesarios y los tokens de acceso correspondientes para realizar solicitudes a la API. Al parecer LinkedIn se preocupa en gran medida de la seguridad en las aplicaciones que acceden a su plataforma, y en consecuencia ofrece en la documentación adjunta al tutorial un listado de buenas prácticas para garantizarla en el proceso de acceso y extracción de datos de cualquier aplicación. Sin embargo, no será necesario para el análisis que queremos llevar a cabo la inmersión en conceptos de seguridad tratados en dicha documentación, pues haremos uso una vez más de herramientas de terceros que nos facilitarán la obtención de los tokens. A pesar de que LinkedIn no posee una herramienta propia como era el caso de Graph API Explorer de Facebook, sí recomienda (y enlaza desde su página) la consola de la empresa Apigee. Una vez más, para utilizarla, necesitaremos ingresar una cuenta de LinkedIn válida, ya que de otra forma no es posible obtener los tokens de acceso que se precisan. Vemos entonces que encontramos aquí la misma limitación que en Facebook, en el sentido de que no es posible la extracción de datos anónima de la API, sino que siempre será primordial la existencia de un usuario logueado que haya otorgado los permisos correspondientes.

En la suposición de que no se cuenta con una membresía en el Partner Program de LinkedIn (el cual permite un acceso privilegiado a la API) sólo se puede obtener lo que se llama información básica de perfil, habiendo conseguido en primera instancia el permiso *r_basicprofile*. A diferencia de Graph API Explorer, la consola de Apigee no permite gestionar los diferentes permisos existentes, sino que son configurados por defecto en la herramienta y son otorgados al momento de loguearse con la cuenta de LinkedIn. Para el caso que se quiera acceder a la API a través de una aplicación propia, en primer lugar el desarrollador deberá registrar en LinkedIn, con una cuenta real, una aplicación. Luego los permisos del usua-

[2] En LinkedIn, miembros y usuarios se consideran sinónimos.

rio que haga uso de ella serán obtenidos en el segundo paso de autenticación (ver tutorial mencionado), haciendo uso del parámetro *scope* en la solicitud HTTP para obtener un código de autorización. Volviendo a la información básica de perfil, la misma contiene algunos campos que podrían ser de interés para nuestro trabajo: *location, industry, positions*. La consulta que permite obtener todos estos datos de la API es *https://api.linkedin.com/v1/people/(user-id):(location,industry,positions)?format=json*. Con la información de dichos campos ya podríamos obtener dónde se ubica la persona, en qué tipo de industria trabaja y qué posición ocupa. Es más, el campo *positions* es definido como un objeto que posee campos propios, como son: *title, start-date, end-date, is-current, company*. Es decir, además de la información anterior, con estos nuevos datos obtendríamos el título del cargo ocupado y la compañía, si dicho cargo es el actual o no, y cuándo se inició o desistió la persona en el cargo. El gran obstáculo para este caso está en la obtención de los ID de usuarios que sean alumnos o graduados de ISI-UTN-FRRo.

5. Oportunidades y limitaciones técnicas

Volviendo a analizar nuestros objetivos, vemos que se desprenden de lo estudiado diferentes oportunidades y limitaciones. Comenzando con las oportunidades, a partir de los permisos mencionados sería posible acceder a datos de gran valor para nuestro proyecto.

En Facebook, utilizando el historial educativo podríamos filtrar aquellos estudiantes o graduados de ISI-UTN-FRRo, para luego conocer su historia laboral, como así también el lugar de residencia o la fecha y lugar de nacimiento. Potencialmente podríamos obtener la lista de amigos, para continuar explorando en sus redes otras personas que hayan estudiado en la Facultad. También sería factible pensar como alternativa utilizar como punto de referencia de la investigación la página en Facebook de nuestra facultad, sin embargo es allí donde encontramos las primeras limitaciones.

Mientras que sí se puede navegar hacia la página en cuestión desde un usuario que la haya colocado en su perfil como Educación, lamentablemente no es posible hacer el proceso inverso. Es decir, no existe la posibilidad desde la página de Facebook UTN-FRRo conocer todas aquellas personas que la hayan designado como su lugar de Educación. Además, todos aquellos alumnos y ex-alumnos que no hubiesen completado totalmente su perfil en Facebook, indicando historia laboral y educativa con el campo Ingeniería en Sistemas en UTN-FRRo, quedarían excluidos del análisis. Por otro lado, existe la posibilidad de conocer todas aquellas personas que colocaron "Me Gusta" en la página en cuestión una vez que se obtuvo el ID de la página de la Facultad, pero lamentablemente dicha lista de personas no puede obtenerse a través de la API, sino que sólo puede verse ingresando a `www.facebook.com/search/116602285100347/likers?ref=about`. Cuando usuarios de Facebook reportaron como un bug[3] la imposibilidad de recuperar la lista completa de fans de una página, el equipo de Facebook respondió aclarando que dicha situación no es un bug, sino que es algo que no puede hacerse por el propio diseño de la API.

Cabe señalar en este punto que no es nuestra intención contactarnos con cada uno de los graduados o actuales estudiantes para que utilicen una aplicación con su cuenta de Facebook, y así poder acceder a los datos que buscamos. El proyecto en el cual está enmarcado este trabajo ya tuvo experiencias poco fructíferas para contactar a dichas personas a través de encuestas y formularios, por lo que pretendemos, de forma sistematizada, obtener los datos públicos sin necesidad de la interacción con los usuarios. Aunque parezca algo radical, podemos argumentar a nuestro favor que esta práctica es realizada por otros investigadores que trabajan con grandes conjuntos de datos (datasets) alrededor del mundo. Weller y Kinder-Kurlanda (2014) presentan testimonios de investigadores involucrados en diferentes proyectos de extracción y análisis de datos de redes sociales, entre los que destacan que las prácticas estándar utilizadas en otros campos, como pedir el consentimiento de participantes, son ya imposibles de llevar a cabo cuando se trabaja con grandes datasets. En consecuencia, plantean que *"trabajar con datos recolectados de redes sociales constituye un nuevo contexto para la ética y metodologías de la investigación"* [15]. Manteniendo al margen por el momento los alcances éticos, morales y legales, y retomando el análisis puramente técnico, parece que el único camino restante es recurrir a nuestras propias cuentas de Facebook para buscar todos aquellos perfiles, pertenecientes a nuestras listas de amigos, que sean alumnos o graduados de Ingeniería en Sistemas en la UTN-FRRo.

[3] El bug reportado (y ya cerrado) junto con la respuesta del equipo de Facebook puede seguirse en https://developers.facebook.com/bugs/147185208750426

Para ello, en primer lugar puede obtenerse la lista de amigos del usuario logueado a través de la solicitud GET *me/friends*, y luego debe adjuntarse a dicha solicitud los parámetros *fields=birthday,education,hometown,location,work*. Hecho esto, deberá filtrarse todos aquellos usuarios obtenidos donde el campo *school*, ubicado dentro del objeto *education*, corresponda a la página de Facebook de la UTN-FRRo, como se ve en la Figura 1.


```
{
  "education": [
    {
      "concentration": [
        {
          "id": "186554774718639",
          "name": "Ing. en Sistemas"
        }
      ],
      "school": {
        "id": "116602285100347",
        "name": "UTN Facultad Regional Rosario"
      },
      "type": "College",
      "id": "3251301929475"
    }
  ],
  "id": "                "
}
```


Figura 1. Respuesta de la API Graph.

En cuanto a LinkedIn, el conjunto de datos resultante que es posible de obtener aparentaría ser, a primera vista, muy prometedor. Sin embargo, sumado a la dificultad existente para obtener los IDs de personas de ISI-UTN-FRRo a través de la API, hay un faltante que afecta en gran medida a la viabilidad de nuestro proyecto: la lista de contactos del usuario logueado, que nos permitiría navegar entre redes de usuarios de LinkedIn.

En su momento, la REST API incluía la llamada Connections API, que permitía acceder a toda la red de personas en la plataforma, navegando a través de las conexiones entre contactos. Aunque antiguamente estas conexiones del usuario podían obtenerse a través de la solicitud https://api.linkedin.com/v1/people/(user-id)/connections?format=json, hoy la API responde a dicha solicitud con el mensaje de la Figura 2.


```
{
  "errorCode": 0,
  "message": "Access to connections denied",
  "requestId": "           ",
  "status": 403,
  "timestamp":
}
```


Figura 2. Respuesta de REST API a solicitud de lista connections.

Según Wagner (2015), existen dos razones que explican que se diera de baja una sección tan importante y utilizada hasta ese momento de la API. Por un lado, el abuso en el uso de la funcionalidad para hacer *spamming* y por otro, una razón competitiva, y hasta económica, consistente en el uso de la plataforma por parte de terceros para obtener las conexiones entre usuarios [16].

Lamentablemente, en LinkedIn tampoco existe hoy en día la opción de buscar aquellos miembros de la plataforma que concurren a ISI-UTN-FRRo para obtener sus IDs de usuario. La API que permitía hacer uso del buscador a través de la solicitud https://api.linkedin.com/v1/people-search, se encuentra hoy cerrada para desarrolladores comunes, pudiéndose ver la respuesta obtenida al consultarla en la Figura 3.


```
{
  "errorCode": 0,
  "message": "Access to people search denied.",
  "requestId": "           ",
  "status": 403,
  "timestamp":
}
```


Figura 3. Respuesta de REST API a solicitud people-search.

El no tener acceso a la lista de contactos ni poder realizar el rastreo de personas a través del buscador de LinkedIn parecería enfrentarnos a un callejón sin salida. Sin embargo, todavía queda una alternativa aludida pero no explorada; nos referimos al nombrado "Partner Program". En términos generales, sólo son aceptadas en este programa *"aquellas aplicaciones que provean valor a miembros, desarrolladores y a LinkedIn"*. Asociarse a LinkedIn a través de este acuerdo otorga al desarrollador permisos especiales para acceder a funcionalidades adicionales de la API, logrando un acceso mucho más amplio y profundo a la plataforma. Esto se debe a que se puede manejar otra clase de permisos, como es *r_fullprofile*. Con él podría conocerse información más amplia y variada sobre un usuario, entre ellas, skills, publicaciones, cursos, premios obtenidos, etcétera.

Para aplicar, la persona a cuya cuenta estará asociada la aplicación deberá completar un minucioso formulario, en el cual se deben especificar objetivos, métricas y audiencia de la aplicación, como así también una descripción de la información que se planea extraer, datos personales y de la compañía. A pesar de solicitar esa amplia variedad de requisitos, en ningún

lugar se establecen los criterios de elegibilidad utilizados por el comité evaluador para aceptar una nueva membresía al Parter Program. Se explican en los Términos de Uso de la REST API criterios para el uso general de la API, donde se aclara que deben aplicar al Partner Program aquellas aplicaciones que no cumplan con alguno de ellos. A saber:

- Las bases fundamentales de la aplicación no dependen del acceso a la API.
- La aplicación no debe tener más de 250.000 miembros en toda su vida, hacer más de 500.000 llamadas diarias a la API ni hacer más de 500.000 búsquedas de personas en todo su ciclo de vida.
- La aplicación no busca atraer a actuales o potenciales clientes que paguen por los servicios y productos de LinkedIn o a personas participantes en actividades relativas a dichos productos.

Aunque la aplicación al Partner Program es teóricamente una alternativa factible, al analizar el tipo de empresas que han logrado ser aceptadas nos encontramos con corporaciones de la talla de Microsoft, Apple, Samsung, medios digitales como TechCrunch y Recode, o bien consultoras de recursos humanos estadounidenses. En ningún caso se mencionan universidades (ni siquiera privadas) en las listas de miembros asociados. Esto nos lleva a pensar que la aceptación de una aplicación como la que pretendemos construir es muy poco probable, sobre todo si tenemos en consideración el hecho de que LinkedIn cuenta actualmente con una funcionalidad exclusivamente orientada a universidades: las llamadas *university pages* o páginas de universidades. Las mismas, según lo expresado por la empresa, *"permiten interactuar con millones de antiguos, actuales y futuros estudiantes, utilizar herramientas de comunicación segmentada y conseguir información más detallada sobre las trayectorias de los antiguos alumnos"*. Puede obtenerse, en la sección de FAQs[17], mayor información respecto a funcionalidades, creación y administración de este tipo de páginas. Estudiando la correspondiente a la Universidad Tecnológica Nacional, hallamos que pese a revelar información de interés sobre actuales y ex-alumnos relativa a la ubicación, compañía en que trabajan, y aptitudes (skills), no logra satisfacer nuestros requerimientos por las siguientes razones:

- Al tratarse de una página de la UTN a nivel nacional, en ella se agregan datos de estudiantes de todas las facultades regionales y todas las carreras. A pesar de existir la posibilidad de filtrar por rama y estudios realizados, no resulta de utilidad para relevar la situación particular del alumnado de cada Regional en particular.
- Las estadísticas no pueden extraerse sistemáticamente de LinkedIn a través de la REST API para trabajar con ellas y realizar el filtrado con herramientas propias.
- Sólo se computan en las estadísticas aquellos usuarios que hayan colocado en el campo Universidad de su Educación a la Universidad Tecnológica Nacional.

El último punto toma especial relevancia considerando que la UTN-FRRo (al igual que el resto de las regionales) figura en LinkedIn como Compañía. Esta situación provoca que muchos ex-alumnos, actuales docentes e investigadores de la facultad, no hayan colocado en sus perfiles a la UTN como lugar de educación, sino simplemente a la UTN-FRRo como lugar de trabajo.

La creación de una página de universidad para la regional Rosario quedará descartada dado que dicha iniciativa no es recomendada por LinkedIn. Respecto a este tema, declara: *"no recomendamos crear páginas independientes para facultades individuales dentro de la misma universidad. [...] Al haber menos páginas de universidad, los datos y de antiguos alumnos se agregan para comprender mejor las trayectorias profesionales de antiguos alumnos"*.

6. Aspectos éticos y legales

Más allá de las cuestiones técnicas que hemos analizado hasta aquí, indudablemente existen aristas morales, éticas y legales que debemos tener en consideración. La masificación del uso de las redes sociales ha hecho que cada vez más información personal esté disponible en ellas, en general provista por los mismos usuarios que las utilizan. Es por ello que Madden (2012) destaca que las cuestiones relativas a la privacidad de dicha información es un tema que está tomando más y más relevancia, tanto entre las empresas dueñas de los datos como en los propios usuarios generadores de contenido [18]. De acuerdo con esto, todas las grandes corporaciones tienen publicados en sus sitios una serie de obligaciones que deben asumir aquellas personas que se muestren interesadas en el uso y extracción de la información disponible

en las plataformas sociales. Las APIs provistas tanto por LinkedIn como Facebook tienen sus "Términos de Uso", que dan una enumeración precisa de todas las regulaciones existentes relativas al uso de la herramienta.

A la hora de analizar las cuestiones que abordaremos en esta sección, debemos realizar una primera división de la información existente en las redes que, más allá de resultar obvia, es de gran importancia puesto que produce el punto separación inicial de dos grupos con matices bien diferenciados. Esta división consiste en separar la información entre pública y privada. Con pública nos referiremos a la información de los perfiles personales que está disponible para cualquier internauta que decida visitarlo; mientras que por privada entenderemos a toda aquella información a la cual sólo pueden acceder ciertos usuarios que posean los permisos necesarios para visualizarla. Esta definición es apoyada por autores como Zimmer (2010), cuando remarca que *"la información de un estudiante no debería ser considerada objetivamente pública o privada, sino que debería ser considerada pública o privada desde la perspectiva particular de la persona que descarga los datos del estudiante"* [19].

En las redes bajo análisis, las personas son dueñas de configurar el carácter de la información de su perfil personal. No obstante, existen ciertos campos que son considerados como públicos por la red social que los almacena, sin importar la opinión del usuario al respecto. En el caso de Facebook, existe una sección dentro del Servicio de Ayuda en donde se define el concepto de *perfil público*. Este perfil incluye los siguientes campos del perfil de un usuario: nombre, sexo, identificador de usuario, foto de perfil, foto de portada y redes; conjuntamente con esto, Facebook aclara que tanto el rango de edad y el idioma como así el país del usuario, también son considerados públicos.

Por otra parte, LinkedIn ofrece al usuario un mayor control de los campos que conforman su perfil. Si bien en la sección de ayuda se deja asentado claramente que toda la información del perfil es de carácter público, también se manifiesta que la información de contacto (como los campos correo electrónico y dirección física) y la lista de contactos del usuario sólo es visible por los contactos de 1er grado del usuario, a menos que éste decida no compartir dicha información. En adición a esto, LinkedIn permite modificar la visibilidad tanto del perfil del usuario como de su actividad reciente y publicaciones. Para ello, se define de manera homóloga el concepto de *perfil público*. A diferencia de Facebook, el perfil público de LinkedIn se refiere al perfil del usuario que aparecerá como resultado en motores de búsqueda como Google, Yahoo!, etc. Este perfil puede mostrarse también a miembros de LinkedIn que posean el correo electrónico del usuario o que hayan tenido reuniones con el mismo. Resumiendo, la diferencia principal entre los distintos perfiles públicos es que el definido por LinkedIn permite ser configurado por el usuario para que sea él quien decida qué campos se mostrarán, dando incluso la posibilidad de hacer que el perfil no sea visible para nadie.

Definidos ya los conceptos de información pública y privada, y de perfil público tanto para Facebook como LinkedIn, procederemos a desglosar las cuestiones legales respectivas al uso y manipulación de datos personales y a la privacidad de los mismos. Finalmente nos enfocaremos en la arista ética y moral estos procesos, en donde entraremos en cuestiones polémicas como la transgresión virtual de la privacidad.

6.1. Condiciones de Uso

Antes de comenzar, resulta oportuno aclarar que una persona que acepte las Condiciones de Uso de cualquier producto o servicio, en este caso el servicio ofrecido por la empresa a cargo de la red social, está suscribiendo a un acuerdo jurídicamente vinculante. Es decir, las condiciones de uso son un documento de carácter legal de suma importancia, por lo que repudiamos la liviandad con la que se lo considera por parte del colectivo internauta en general. Sin más, comenzamos a analizar las condiciones de uso de las redes sociales en las que nos enfocamos hasta el momento.

El documento que posee las condiciones de uso de Facebook, llamado Declaración de Derechos y Responsabilidades (DDR), tiene su origen en los Principios de Facebook[4] y rige la relación de la empresa con los usuarios y todos aquellos que interactúan con Facebook. En este documento se deja en claro que al utilizar o acceder a cualquier servicio ofrecido, el usuario muestra su conformidad con las políticas y condiciones de uso de la empresa. Asimismo, LinkedIn también posee un documento con igual finalidad, denominado simplemente Condiciones de Uso. En él se engloban tanto los usuarios que utilizan el servicio de LinkedIn como los que manejan los demás servicios de la empresa (SlideShare, Pulse, etc.). En adición a

[4]Nos referimos a los derechos y responsabilidades de aquellos que conforman el servicio de Facebook y que fueron establecidos por la misma empresa.

estos documentos, cada empresa posee una serie de políticas referidas a tópicos tales como políticas de publicidad o políticas de copyright, las cuales se cercioran de cubrir cada aspecto de los servicios ofrecidos. Concretamente, existe una política análoga para ambas empresas, la política de datos en Facebook y la política de privacidad de LinkedIn, que resulta de gran interés y utilidad para el análisis de la factibilidad legal del proceso de recopilación sistematizada de datos. Sin embargo, realizar dicho análisis implicaría sumergirnos en aspectos legales muy específicos, corriendo el riesgo de perder el enfoque de este trabajo. Por consiguiente, la evaluación de estas políticas será postergada para un trabajo futuro, enfocado de manera más precisa a las cuestiones legales específicas.

Volviendo a los puntos que se desarrollan dentro de la DDR de Facebook, existen dos de ellos que nos conciernen de sobremanera. Estos son el punto 3.2: *"No recopilarás información o contenido de otros usuarios ni accederás a Facebook utilizando medios automáticos (como bots de recolección, robots, spiders o scrapers) sin nuestro permiso previo"*; y el punto 5.7: *"Si obtienes información de los usuarios, deberás obtener su consentimiento previo, dejar claro que eres tú (y no Facebook) quien recopila la información y publicar una política de privacidad que explique qué datos recopilas y cómo los usarás"*.

De forma similar al primer punto expuesto de la DDR, LinkedIn indica en el punto 8.2 de sus Condiciones de Uso que, al ser aceptadas, el usuario asume la responsabilidad de no realizar diversas acciones, entre las cuales se encuentran:

- Utilizar el método de scraping o copiar perfiles e información de otras personas a través de cualquier medio (incluidos los crawlers, los plugins de navegación y complementos, y cualquier otra tecnología o programas manuales).
- Recabar, utilizar, copiar o transferir cualquier información obtenida de LinkedIn sin el consentimiento de LinkedIn.
- Compartir o revelar información de otras personas sin su consentimiento expreso.

Estas normativas resultan clave a la hora de realizar una recopilación sistematizada de datos puesto que regulan y restringen este proceso. Sin lugar a dudas, las restricciones pretenden conseguir que las personas interesadas en recoger datos de Facebook o LinkedIn lo hagan mediante las APIs que la empresa provee o, en su defecto, mediante el uso de técnicas y/o herramientas que posean la aprobación de las empresas a priori (como Apigee o Facepager) y así evitar conflictos legales referidos a la transgresión de la privacidad o a la sustracción ilegal de datos.

6.2. Términos de Uso de las APIs

A pesar de todas las limitaciones técnicas que hemos encontramos para la extracción de datos en las plataformas provistas por LinkedIn y Facebook, trataremos a continuación las consideraciones legales a tener en cuenta por cualquier desarrollador para el tratamiento, distribución y registro de los datos extraídos. El total de las consideraciones pueden encontrarse en detalle en los Términos de Uso de la API, por lo que sólo ahondaremos aquí en aquellas que conciernen exclusivamente a nuestros objetivos.

6.2.1. Facebook - Graph API. Al brindar más de una API a los desarrolladores, Facebook posee una sección única, llamada *Política de la Plataforma*, en donde se exponen las distintas normas que rigen la plataforma de desarrollo de aplicaciones de Facebook en general. Dichas normas regulan el accionar de desarrolladores al momento de construir una aplicación y las características que ella debe poseer para ser aprobada por Facebook.

Una de las normas más ricas, y quizás la de mayor importancia para nuestro análisis, es la llamada *"Dar a los usuarios el control"*. En ella yacen cuatro puntos referidos a la recolección de datos personales de los usuarios y la privacidad de los mismos. En primer lugar, los puntos 2.4 y 2.5 indican de manera conjunta que cualquier aplicación realizada por un usuario en la plataforma que provee Facebook debe poseer una política de privacidad propia, sumada a la política de la empresa misma. Además de eso, se señala que en el caso de que la aplicación cumpla con ambas políticas de privacidad, el desarrollador puede hacer uso de la información de la cuenta[5] del usuario dentro de la aplicación, o fuera de ésta si posee el consentimiento expreso del usuario. Esto supone una posibilidad para la futura concreción del objetivo de construir el mapa interactivo, puesto que la información a la que se puede acceder mediante este medio resulta suficiente para ubicar a los respectivos graduados o estudiantes, aunque existe una fuerte limitación en la obtención del permiso para utilizar sus datos.

[5] Incluye: nombre, dirección de correo electrónico, sexo, fecha de nacimiento, ciudad actual y URL de la foto del perfil

En segundo lugar, el punto 2.9 señala dos cuestiones de suma relevancia. Se indica que si un usuario solicita al desarrollador la total eliminación de sus datos, éste debe cumplir con el pedido a menos que haya una ley, reglamento o acuerdo aparte suscrito con Facebook que lo obligue a lo contrario. En adición a esto, se advierte que el desarrollador puede conservar los datos globales si no es posible deducir o crear información que identifique a una persona en particular a partir de estos. Si bien este punto pretende brindarle cierto nivel de control de sus datos al usuario, dicho control no es pleno pues existe la posibilidad de que el desarrollador se encuentre obligado a conservar los datos por algún documento judicial vinculante o que éste los conserve argumentando la imposibilidad de identificación unívoca del usuario demandante.

El último punto de interés de la norma es el 2.11, el cual indica que Facebook puede también recabar información del usuario a partir de la aplicación del desarrollador. Textualmente: *"Obtén el consentimiento adecuado de las personas antes de utilizar cualquier tecnología de Facebook que nos permita a nosotros recolectar y procesar datos sobre ellas [...]. Existen terceros, incluido Facebook, que pueden utilizar cookies, web beacons y otros métodos de almacenamiento para recolectar o recibir información de tus sitios web, aplicaciones o donde sea en Internet [...]"*. Aquí podemos apreciar que los datos del usuario no son tan privados como él mismo creería, pues no sólo el desarrollador posee acceso a los mismos, sino que Facebook y cualquier tercero asociado tienen capacidad y derecho de acceso. Por ende, podemos pensar que cualquier usuario consciente de esto reflexionaría el hecho de utilizar una aplicación que permita tanto al desarrollador como a Facebook y diversos terceros la recolección de sus datos personales para un uso desconocido y/o no consensuado por el mismo. Ergo, esto presenta una limitación crucial para cualquier desarrollador ávido de recolectar datos de buena fe para un uso fructífero, como es el caso de los autores del presente trabajo.

Otra norma que resulta trascendente para nuestro análisis es la llamada *"Proteger los Datos"*. Como su nombre lo indica, cubre diversos aspectos referidos a la protección de los datos que circulan dentro del uso de una herramienta creada en la plataforma de desarrollo de Facebook. Tal es el intento de proteger estos datos que, en el punto 3.1, se informa de manera inmediata al desarrollador que los datos deben ser protegidos de accesos o usos no autorizados. Luego, el punto 3.2 expone que un desarrollador puede mostrar datos del usuario, obtenidos mediante un token de acceso, únicamente en los dispositivos que estén asociados a dicho token. Esto quiere decir que si un desarrollador obtiene los datos de un usuario mediante una aplicación, él sólo podrá mostrarlos en la aplicación en cuestión.

Por último resulta pertinente considerar el punto 3.13, donde se subraya que en caso de que el desarrollador deje de utilizar la plataforma, deberá borrar de inmediato todos los datos personales de los distintos usuarios, sin el previo consentimiento de ellos. También se aclara que el desarrollador puede conservar la información básica sobre las cuentas que utilizaron la herramienta, si el desarrollador previamente mostró su política de privacidad dentro de la herramienta[6].

6.2.2. LinkedIn - REST API Anteriormente se indicó que, técnicamente, no era factible la utilización de esta API para la recopilación sistematizada de los datos necesarios para crear el mapa interactivo propuesto, por lo que es probable que el lector considere que un análisis de las condiciones de uso de la API resulte fútil. Sin embargo, debemos discrepar con este pensamiento dado que, en caso de aplicar de manera satisfactoria para convertirse en partner de LinkedIn, un desarrollador encontraría en REST API una herramienta con características notorias y un gran potencial.

Las condiciones de uso de REST API se dividen, al igual que las condiciones de uso de la plataforma de desarrollo de Facebook, en puntos que abarcan diferentes aspectos referidos al uso de la API. Durante el desarrollo previo de este trabajo, ya se consideraron dos puntos con implicancias sobre nuestro análisis. En primer lugar, se habló sobre el punto 1.4, el cual expone los criterios de elegibilidad que determinan si se debe o no aplicar al Partner Program de LinkedIn. Por otro lado también se hizo referencia, aunque indirectamente, a los puntos 2.1 y 2.2, los cuales indican que el desarrollador debe poseer una cuenta real en LinkedIn con la cual registrar la aplicación. Sólo luego de este proceso de registro la aplicación obtendrá las credenciales de acceso requeridas para utilizar la REST API.

Una limitación crucial que hemos encontrado dentro de las condiciones de uso de la API reside en el punto 3.3.2, perteneciente a la sección *"Uso de la API y contenido de LinkedIn"*. En él se aclara que no está permitido combinar contenido recopilado de

[6]Conforme lo indica el punto 2.6, el cual exige al desarrollador que incluya la URL de la política de privacidad dentro de su aplicación

LinkedIn con información recabada de otras fuentes, de tal manera que un usuario final no pueda atribuir el contenido original a LinkedIn. Esto resulta contraproducente puesto que, en el caso de que los datos que se obtengan de LinkedIn sean escasos, no se podrán complementar con datos obtenidos de otras fuentes a menos de que se esclarezca qué campos fueron extraídos de LinkedIn inicialmente.

Los puntos 4.4, 4.6 y 4.7, englobados dentro de la sección *"Almacenamiento de contenido"*, exponen varias cuestiones de interés. En primer lugar, el punto 4.4 LinkedIn declara que un desarrollador sólo puede almacenar datos del perfil de un usuario si éste brindó previamente su consentimiento para hacerlo. Para lograr esto, el desarrollador debe hacer saber al usuario, de manera clara y transparente, cuáles de sus datos serán almacenado. Se informa en este punto además que el almacenamiento de datos de usuarios se permite bajo el solo propósito de beneficiarlo mediante la manipulación de los datos recolectados; por ejemplo, utilizar su información profesional para evaluar la postulación a un puesto de trabajo disponible.

Por otra parte, los puntos 4.6 y 4.7 refieren a la eliminación de los datos almacenados, y si se recuerda el apartado 2.9 de la política de Facebook, se puede observar que ambas empresas adoptan el mismo enfoque en esta cuestión. En caso de que el usuario borre su cuenta de LinkedIn o que la empresa indique que la aplicación transgrede los términos de uso de la API, LinkedIn exige al desarrollador la completa eliminación de cualquier dato que la empresa le haya provisto. Sin embargo, en el punto 4.7 se remarca la siguiente salvedad: *"[. . .] excepto que hacer esto cause la violación de alguna ley u obligación impuesta por una autoridad gubernamental"*.

Finalmente se considerará la sección *"No dañar o engañar a los miembros de LinkedIn"*. En ella, LinkedIn exige al desarrollador la inclusión de un acuerdo de política de privacidad en la solicitud de consentimiento, el cual deberá ser fácilmente identificable y/o localizable cuando el usuario de LinkedIn accediera a la aplicación. Además, las políticas de privacidad deben cumplir con los estándares legales y describir con precisión la recopilación, uso, almacenamiento e intercambio de datos. Obtenida la aceptación del usuario, no se permitirá cualquier solicitud de campos a la API que no hayan sido especificados de manera clara, precisa y transparente en las políticas.

6.3. Perspectiva ética

Indudablemente existe una gran falta de estándares éticos y marcos de trabajo regulatorios en el campo de la investigación en redes sociales. Entendemos que al ser esta un área relativamente reciente, todavía no ha transcurrido el tiempo necesario para el desarrollo de protocolos éticos robustos que puedan adaptarse al sinnúmero de diversas tecnologías que han aparecido en los últimos años. Esto lleva a que, la mayoría de las veces, cada investigador en particular tome una perspectiva distinta respecto a las implicancias éticas que tiene su trabajo (si es que siquiera las tiene en consideración) y las formas de sobreponerse a las limitaciones que de los dilemas éticos surgen. En las entrevistas que Weller y Kinder-Kurlanda (2014) llevan adelante con distintos investigadores de redes sociales para conocer el enfoque ético con que trabajan, encuentran que las prácticas de investigación en este campo pueden diferir fundamentalmente en hipótesis y asunciones generales, como ser la diferencia entre datos públicos y privados, o bien variar en cuestiones más sutiles como son las técnicas de anonimización de datos aplicadas [15].

Es evidente que, ya que no estaríamos pidiendo el consentimiento de los usuarios para la utilización de sus datos, será imperativo anonimizar todos los datos mostrados en el mapa interactivo. Es decir, en ningún momento quedará explícita la asociación entre un punto geográfico y la identidad de una persona en particular, sino que simplemente serán muestras estadísticas generales de un área geográfica; por ejemplo, cantidad de estudiantes y/o graduados de ISI-UTN-FRRo trabajando en cierta empresa de la ciudad. El proceso de anonimización toma aún más relevancia si pretendemos que el conjunto de datos obtenidos de nuestra investigación sea abierto a la comunidad. De finalmente decidirnos por esta alternativa, será primordial que todo aquel investigador con deseos de acceder al dataset, firme un documento destinado a regular los términos y condiciones de uso del mismo. Deberán también establecerse los mecanismos necesarios para la monitorización y control del cumplimiento del acuerdo firmado, pues como plantea Zimmer (2010), *"mientras que requerir términos de uso es ciertamente un paso positivo, sin su correcta aplicación podrían tener limitado éxito en prevenir cualquier uso de los datos potencialmente invasivo de la privacidad"* [19].

Podría pensarse en primera instancia que la anonimización de los datos extraídos puede ser alcanzada

con eliminar aquellas variables típicas que identifican a una persona, como son el nombre y apellido. Sin embargo, el proceso a seguir para efectivamente anonimizar un conjunto de datos es mucho más complejo y difícil de abarcar completamente. Un caso ejemplar es el que presenta Zimmer (2010), donde en un principio se recolectó manualmente y publicó, sin consentimiento alguno, información de perfiles de Facebook de estudiantes de distintas universidades estadounidenses. A pesar de que se creyó este conjunto de datos completamente anonimizado luego de ciertos procesos de limpieza, a la semana de su publicación todas las personas involucradas habían sido reconocidas unívocamente [19]. Con esto queremos dejar en claro que la identificación de una persona no se basa estrictamente en datos personales, sino que a veces está relacionada con datos aparentemente triviales; ¿cuánto tardaría en reconocerse a una persona si es la única que trabaja en cierta empresa, vive en una ciudad particular del extranjero y se graduó de ISI-UTN-FRRo en un año determinado?, ¿querrá esa persona que se den a conocer sus datos en nuestro, aparentemente inofensivo, mapa informativo?.

En este sentido, es nuestra opinión que si bien los datos que los usuarios colocan en sus perfiles de redes sociales pueden ser públicos, eso no quiere decir que como investigadores tengamos total libertad para darles el uso que queramos. Es más, existirán casos donde se extraiga información de usuarios a la que sólo tendremos permiso para acceder por pertenecer a su lista de amigos, pero tal vez dichos datos no sean públicos para cualquier persona; ¿no estaríamos cruzando barreras éticas relativas a la privacidad del usuario si sus datos son fuente para estadísticas en un conjunto de datos público?. Definitivamente esta pregunta no es posible de responder en forma generalizada, por lo que llegado el momento será imperativo evaluar para cada caso particular qué datos publicaremos en el mapa y qué datos permanecerán ocultos y protegidos.

7. Conclusiones

Es indudable que las redes sociales están cambiando radicalmente los paradigmas establecidos para el estudio del comportamiento humano. Es por ello que el interés de investigadores para incorporar las redes sociales a sus estudios no ha parado de crecer, como así tampoco lo ha hecho la preocupación por la privacidad y protección de los datos en ellas generados y almacenados. Aunque hoy en día es muy variado el número de herramientas, APIs y SDKs que permiten la extracción de datos en tiempo real, se observó que más allá de la plataforma que se ponga bajo análisis, todas toman posiciones similares a la hora de poner límites al acceso y la utilización de datos que aparentan ser públicos. Muchos de los límites impuestos resultan una amenaza para la concreción del mapa planteado, siendo el principal obstáculo que deberemos afrontar la obtención del consentimiento del usuario. Además, la eficaz anonimización del conjunto de datos recolectado junto con la redacción de términos y condiciones de uso del mismo serán primordiales si en el futuro se abren a la comunidad los resultados obtenidos de la aplicación de las técnicas analizadas.

Por otro lado, el análisis técnico realizado nos permite concluir que aunque LinkedIn ofrece en su plataforma datos más precisos, estructurados y completos, su extracción se hace extremadamente dificultosa pues se requieren permisos especiales a los cuales sólo grandes corporaciones asociadas a LinkedIn han tenido acceso. Siendo Facebook la única fuente de datos restante de las estudiadas, encontramos que su plataforma brinda un conjunto de datos más genérico, incompleto, y poco orientado al perfil profesional de estudiantes y graduados, donde el principal faltante radica en la información relativa a skills técnicos de los sujetos. Sin embargo, manipulando este conjunto de datos y realizando un análisis pormenorizado enfocado a los distintos estudiantes y profesionales encontrados dentro de Facebook, será posible recopilar información suficiente para la correcta confección del mapa.

En resumen, entendemos que, aunque restringida, existe la factibilidad tanto técnica como legal para que una persona que publicó datos en redes sociales acabe en un mapa informativo donde se publique su historial académico y profesional. Entonces, el dilema ético al cual nos enfrentaremos de aquí en adelante, y cuya respuesta dependerá de cada investigador dada la total ausencia de estándares éticos, se resume en, ¿es correcto hacer que ello suceda?.

Reconocimientos

Este trabajo está enmarcado en el PID Tutorado: Observatorio Regional de Desarrollo de la Ingeniería en Sistemas de Información e Informática (II-SI.d.r.O), homologado por la Universidad Tecnológica Nacional (código TOTUNAV0004307) Vigencia: 1/4/2016 al 31/03/2019 de la Universidad Tecnológica Nacional - Facultad Regional Rosario, cuyo ob-

jetivo principal es *"el diseño, construcción e implementación de una plataforma tecnológica integrada y abierta que recopile, analice y administre información sustantiva en torno al desarrollo y evolución de las Tecnologías de Información y Comunicaciones, Software y Servicios Informáticos (TIC-SSI) y su aporte a las cadenas productivas transversales, para atender a las necesidades de los diferentes sectores que conforman el Triángulo de Sábato (Universidad-Estado-Industria)"*.

Se agradece la colaboración a todos los integrantes y adscriptos al proyecto de investigación que han participado en la búsqueda bibliográfica y en la ejecución de las pruebas de las APIs evaluadas en este trabajo.

Referencias

[1] L. Manovich, "Trending: The promises and the challenges of big social data," *Debates in the digital humanities*, vol. 2, pp. 460–475, 2011.

[2] A. Bruns, "Faster than the speed of print: Reconciling 'big data' social media analysis and academic scholarship," *First Monday*, vol. 18, no. 10, 2013.

[3] K. Weller y K. Kinder-Kurlanda, "Uncovering the challenges in collection, sharing and documentation: The hidden data of social media research?," 2015.

[4] A. Vercelli, "Repensando las regulaciones de internet. análisis de las tensiones políticas entre no-regular y re-regular la red-de-redes," *Chasqui. Revista Latinoamericana de Comunicación*, no. 129, pp. 95–112, 2016.

[5] J. Bright, G. B. department for work, y pensions, *The Use of Social Media for Research and Analysis: A Feasibility Study*. DWP ad hoc research report, Corporate Document Services, 2014.

[6] E. Rosenfeld, "Facebook shatters wall street estimates, proposes new share structure." http://www.cnbc.com/2016/04/27/facebook-reports-first-quarter-earnings.html, 2016. Accedido 12/07/2016.

[7] D. Glez-Peña, A. Lourenço, H. López-Fernández, M. Reboiro-Jato, y F. Fdez-Riverola, "Web scraping technologies in an api world," *Briefings in bioinformatics*, vol. 15, no. 5, pp. 788–797, 2014.

[8] M. Patterson, "What is an api, and why does it matter?." http://sproutsocial.com/insights/what-is-an-api/, 2015. Accedido 03/07/2016.

[9] J. Bloch, "How to design a good api and why it matters," in *Companion to the 21st ACM SIGPLAN symposium on Object-oriented programming systems, languages, and applications*, pp. 506–507, ACM, 2006.

[10] B. Leiba, "Oauth web authorization protocol," *IEEE Internet Computing*, vol. 16, no. 1, p. 74, 2012.

[11] D. Hardt, "The oauth 2.0 authorization framework," 2012.

[12] Facebook, "Desarrollo de aplicaciones de facebook." https://developers.facebook.com/docs/apps, 2016. Accedido 26/07/2016.

[13] T. Keyling y J. Jünger, "Facepager. an application for generic data retrieval through apis," 2016. Código fuente y distribuciones disponibles en https://github.com/strohne/Facepager/.

[14] T. Bray, "The javascript object notation (json) data interchange format," 2014.

[15] K. Weller y K. Kinder-Kurlanda, "I love thinking about ethics!. perspectives on ethics in social media research," 2014. Presentation at Internet Research (IR15), Daegu, South Korea, 22.-24.10.2014.

[16] K. Wagner, "Linkedin is sharing less with developers." http://recode.net/2015/5/12/11562548/linkedin-is-sharing-less-with-developers, 2015. Accedido 20/07/2016.

[17] LinkedIn, "Páginas de universidad: preguntas frecuentes para los administradores." www.linkedin.com/help/linkedin/answer/38582, 2014. Accedido 20/07/2016.

[18] M. Madden, "Privacy management on social media sites," *Pew Internet Report*, pp. 1–20, 2012.

[19] M. Zimmer, "But the data is already public: on the ethics of research in facebook," *Ethics and information technology*, vol. 12, no. 4, pp. 313–325, 2010.